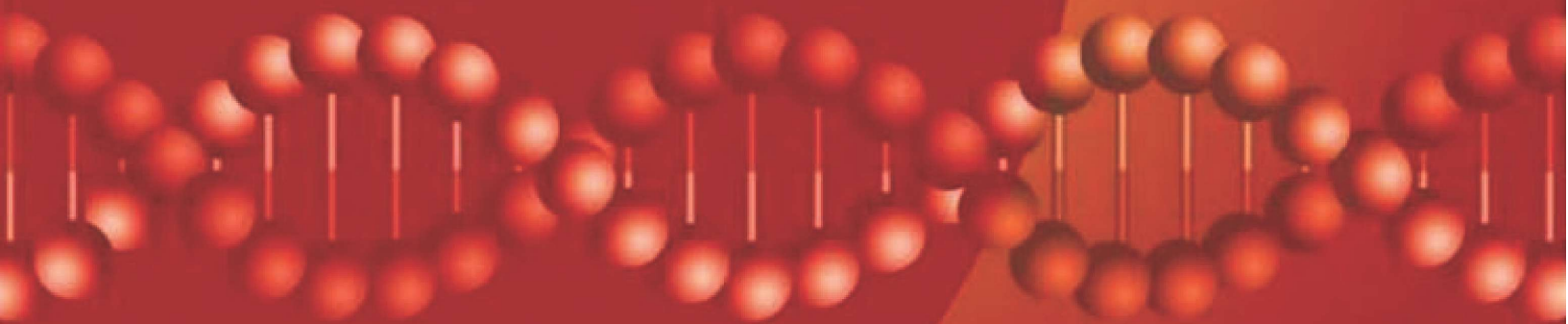


William J. Lennarz
M. Daniel Lane

Encyclopedia of

Biological **Chemistry**



Second Edition



ENCYCLOPEDIA OF
BIOLOGICAL CHEMISTRY

SECOND EDITION

ENCYCLOPEDIA OF BIOLOGICAL CHEMISTRY

EDITORS-IN-CHIEF

WILLIAM J. LENNARZ

Stony Brook University,

Stony Brook, NY, USA

M. DANIEL LANE

Johns Hopkins University School of Medicine,

Baltimore, MD, USA

VOLUME 1

A–Dis

VOLUME 2

DNA–L

VOLUME 3

M–Q

VOLUME 4

R–Z



AMSTERDAM BOSTON HEIDELBERG LONDON NEW YORK OXFORD
PARIS SAN DIEGO SAN FRANCISCO SINGAPORE SYDNEY TOKYO
Academic Press is an imprint of Elsevier



ACADEMIC
PRESS

Academic Press is an imprint of Elsevier
32 Jamestown Road, London NW1 7BY, UK
30 Corporate Drive, Suite 400, Burlington MA 01803, USA
525 B Street, Suite 1900, San Diego, CA 92101-4495, USA

First edition 2004
Second edition 2013

Copyright © 2013 Elsevier, Inc. All rights reserved.

The following articles are US Government works in the public domain and are not subject to copyright:

Adipocyte Fat Mobilization: The Regulatory Roles of Perilipin 1, Adipose Triglyceride Lipase, and Hormone-Sensitive Lipase
Arf Family
Calcium Signaling: Calmodulin-Dependent Phosphatase
Cyclic AMP Receptors of *Dictyostelium*
Cyclic Nucleotide Phosphodiesterases
Cytokines
DNA Mismatch Repair in Disease and Aging
DNA Polymerase β , Eukaryotic
Endocytosis
Eukaryotic Chemotaxis
Gs Family of Heterotrimeric G Proteins
Inositol Phosphate Kinases and Phosphatases
Phospholipase C
Prions of Yeast and Fungi: Proteins Acting as Genes
Riboswitches
Structure and Regulation of Pyruvate Dehydrogenases

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher.

Permissions may be sought from Elsevier's Science & Technology Rights department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com.

Alternatively you can submit your request online by visiting the Elsevier website at <http://elsevier.com/locate/permissions> and selecting *Obtaining permission to use Elsevier material*.

Notice

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

ISBN: 978-0-12-378630-2

For information on all Elsevier publications
visit our website at store.elsevier.com

Printed and bound in the United States of America

13 14 15 16 17 10 9 8 7 6 5 4 3 2

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER BOOK AID International Sabre Foundation

Editorial: Jo Collett, Cari Owen
Production: Justin Taylor

EDITORS-IN-CHIEF

M. Daniel Lane, Ph.D. is Distinguished Service Professor of Biological Chemistry at the Johns Hopkins Medical School. He received B.S. and M.S. degrees from Iowa State University, a Ph.D. degree from the University of Illinois and an honorary Doctor of Humane Letters degree, from Iowa State University. He did a Senior Postdoctoral Fellowship with Professor Feodor Lynen at the Max-Planck Institute Für Zellchemie in Munich. Following faculty positions at Virginia Tech and New York University School of Medicine, he joined (in 1970) the faculty at the Johns Hopkins Medical School where he served as DeLamar Professor and Director of the Department of Biological Chemistry from 1978 to 1997. He was elected to membership in the National Academy of Sciences and the American Academy of Arts and Sciences (1987) and as a Fellow of the American Society of Nutritional Sciences (1996). He received the Mead Johnson Award from the American Society for Nutritional Sciences and the William C. Rose Award from the American Society for Biochemistry and Molecular Biology and served as President of the American Society of Biochemistry and Molecular Biology. He served on numerous editorial boards including the Journal of Biological Chemistry and the Annual Reviews of Biochemistry. With William Lennarz he co-edited the first edition (2004) of the Encyclopedia of Biological Chemistry. Currently he is Associate Editor for Biochemical and Biophysical Research Communications. His early work focused on enzymatic CO₂ fixation reactions, notably the mechanisms by which the B-vitamin, biotin, functions in carboxylases. His research on acetyl-CoA carboxylase, key regulatory enzyme of fatty acid synthesis, led him to his present interests in the basic mechanisms of lipogenesis, adipogenesis and the hypothalamic control of energy balance.

William J. Lennarz received his B.S. in Chemistry from Pennsylvania State University and a Ph.D. in Organic Chemistry from the University of Illinois. Subsequently he carried out postdoctoral work at Harvard with Konrad Bloch on fatty acid biosynthesis. In 1962 he was appointed Assistant Professor at Johns Hopkins in the Department of Physiological Chemistry. After promotion to Associate Professor in 1967, and full Professor in 1971, he remained at Hopkins until 1983. At that time, he was appointed Robert A. Welch Professor and Chair of the Department of Biochemistry and Molecular Biology at the University of Texas Cancer Center, M.D. Anderson Hospital. In 1989 he became a Leading Professor and Chair of the Department of Biochemistry and Cell Biology at SUNY at Stony Brook. In 1990 he founded and became Director of the Institute for Cell and Developmental Biology at Stony Brook.

Dr. Lennarz has served on many national and international committees. He has served as President of the Biochemistry Chairman's Organization, President of the American Society for Biochemistry and Molecular Biology, and President of the Society for Glycobiology. He was a member of the Executive Committee of the International Union of Biochemistry and Molecular Biology for almost a decade.

He has presented special lectures at the University of Notre Dame, the NIH, the University of West Virginia, Johns Hopkins University, Florida State University, the University of California at San Diego, the University of Arkansas, Indiana University, and the Medical College of Virginia.

He is a member of the National Academy of Sciences. The focus of his early work was on lipids and bacterial cell surfaces. More recent efforts have been in the structure, biosynthesis, and function of cell surface glycoproteins. The biosynthesis studies were initially carried out in liver and oviduct, but these efforts now are focused in yeast. The functional studies have concentrated on the role of cell surface glycoproteins in fertilization and early development in the sea urchin and, more recently, the frog. He served as Distinguished Professor and Chair of his department. He now is Distinguished Professor Emeritus.

EDITORIAL BOARD

Wolfgang Baumeister was born on November 22, 1946 in Wesseling near Cologne. Study of biology Münster and Bonn University, doctorate Düsseldorf University (1973), Habilitation in biophysics Düsseldorf University (1978), Heisenberg Scholarship Cambridge/England, head of a group (C3 level) at the Max Planck Institute for Biochemistry (1981), Adjunct Professor Düsseldorf University (1982), Adjunct Professor Chemistry Dept. TU Munich (1987), Director and Scientific Member at the Max Planck Institute of Biochemistry (since 1988), Honorary Professor, Physics Dept., TU Munich (2000). He was made a fellow with the German Academy of Sciences Leopoldina in 2001 and was elected as a Foreign Associate with the US National Academy of Sciences in 2010.

Wolfgang Baumeister has been honored with numerous awards for his research including the Otto Warburg Medal of the German Society for Biochemistry and Molecular Biology (1998), the prize of the Geneva-based Louis-Jeantet Foundation for Medicine (2003), the Ernst Schering Prize (2006), the Schleiden Medal of the German Academy of Sciences Leopoldina (2005), and the Harvey Prize in Science and Technology of the Technion, Haifa (2005).

Ernesto Carafoli was born in 1932, in Italy. He gained his M.D. in 1957 at the University of Modena, Italy; "Abilitation" (Libera Docenza) in General Pathology (1965) and in Biochemistry (1968); Fogarty International Post-doctoral Fellow in the Dept. of Physiological Chemistry of the Johns Hopkins University, Baltimore, MD, USA (1963–1965); Visiting Lecturer in the same department 1968–1969; Assistant Professor, then Associate Professor of General Pathology in the University of Modena School of Medicine, (1959 to 1972); Professor of General Pathology, University of Padova School of Medicine (Italy) (1973); Professor of Biochemistry, Swiss Federal Institute of Technology (ETH) (Zurich, Switzerland), (1973 to 1998); Chairman of the Dept. of Biochemistry of the Swiss Federal Institute of Technology (ETH) in 1978 and 1987 to 1991; Professor of Biochemistry, University of Padova, School of Medicine, Italy (since 1990). From 1971 to 1991 Visiting Professor for various periods in several Italian Universities, at the University of Nairobi (Kenya), at the Universidad Nacional Autonoma of Mexico, Mexico City (Mexico), at the Universidad Central de Venezuela, Caracas (Venezuela), and at Case Western Reserve University, Cleveland (OH, USA). He has been Fogarty Scholar in Residence at the NIH from 1985 to 1988. Co-founder of the Venetian Institute of Molecular Medicine (Padova, Italy), of which he was the first Scientific Director (2000 to 2005). He has received numerous awards and honors including five Doctor Honoris Causa degrees. He is Member of several Academies, including the Johns Hopkins Society of Scholars, and he has been made "Grande Ufficiale" of the Order of Merit of the Republic of Italy, 2006.

Dr. Carafoli has been a member of a number of professional Societies, including the Swiss Biochemical Society, the Biochemical Society, the American Society for Cell Biology, the American Society of Biological Chemistry and Molecular Biology (Honorary Member), the Society of General Physiology, the Biophysical Society, the Italian Society of Biochemistry. He has been a member of the Editorial Board of numerous journals in the area of Biochemistry, Biophysics, and Cell Biology. He has been co-organizer of about 50 International Congresses and Symposia, and of about 25 Advanced Courses, (held in a dozen countries on behalf of FEBS, ICRO/UNESCO, the Gulbenkian Foundation, and WHO). He has delivered lectures (including a number of Plenary and Keynote lectures) at about 350 International Congresses, Symposia, Colloquia. He has written over 500 articles in refereed journals on topics of muscle biochemistry, membrane biochemistry, mitochondrial bioenergetics, membrane transport of ions (calcium specially by pumps) regulation of calcium metabolism, and over 100 book chapters and 70 invited review articles. He has edited about 20 books on topics related to his work.

Pierre A. Coulombe is the E.V. McCollum Professor and Chair of the Department of Biochemistry and Molecular Biology at the Bloomberg School of Public Health at Johns Hopkins University in Baltimore, Maryland. He holds several joint appointments in the School of Medicine and the same institution. Dr. Coulombe obtained a Ph.D. degree from Université de Montréal in Canada and underwent postdoctoral training at the University of Chicago. At that time, he and Dr. Elaine Fuchs discovered that mutations in keratins elicit epithelial fragility in transgenic mice and in individuals suffering from inherited blistering diseases. He continues to explore the properties and function of keratin genes and proteins in the setting of tissue repair and cancer.

Nancy L. Craig is an Investigator with the Howard Hughes Medical Institute, as well as a Professor of Molecular Biology and Genetics at the Johns Hopkins University School of Medicine, where her work is focused on the molecular mechanisms by which transposable elements move and how they can be exploited for genome engineering. After receiving a bachelor's degree in biology and chemistry from Bryn Mawr College, she did graduate work on cellular responses to DNA damage in *E. coli* with Jeffrey Roberts at Cornell University, leading to her Ph.D. degree in biochemistry. As a postdoctoral fellow at the National Institutes of Health, she studied the mechanism of site-specific recombination of bacteriophage lambda with Howard Nash. Before joining HHMI, she was Associate Professor of Microbiology and Immunology and of Biochemistry and Biophysics at the University of California, San Francisco, where she began her research on Tn7 transposition. Dr. Craig is a Member of the National Academy of Sciences and the American Academy of Arts and Sciences. She is also a Fellow of the American Association for the Advancement of Science and the American Academy of Microbiology.

Joel Moss, M.D., Ph.D. is Deputy Chief of the Cardiovascular and Pulmonary Branch, National Heart, Lung, and Blood Institute at the National Institutes of Health, Bethesda, Maryland, USA. He graduated from Brandeis University in 1967, Summa Cum Laude with Honors in Chemistry. In 1972, he received M.D. and Ph.D. degrees from New York University School of Medicine, with his dissertation in the Department of Biochemistry under the mentorship of Dr. M. Daniel Lane. Following an internship and residency in medicine at the John Hopkins Hospital, he completed a post-doctoral fellowship with Dr. Martha Vaughan and pulmonary fellowship with Dr. Ronald Crystal in the National Heart, Lung, and Blood Institute (NHLBI) at the National Institutes of Health. Dr. Moss has coauthored over 600 scientific papers on basic and clinical research, edited or co-authored several books, and is a co-inventor of biotechnology patents. Dr. Moss was a member of the NHLBI Institutional Review Board 1988 to 2006, and its Chair 1995 to 2006. In this regard, he has coauthored a textbook with Dr. Evan DeRenzo entitled: "Writing Clinical Research Protocols: Ethical Considerations". Dr. Moss has received multiple awards, including the Passano Foundation Young Investigator Award (1981), the AFCR Young Investigator Award (1981), and the LAM Foundation Award (1999, 2010). He is a Fellow of the American College of Chest Physicians. Dr. Moss is a member of Phi Beta Kappa, Alpha Omega Alpha, the American Society for Biochemistry and Molecular Biology, Association of American Physicians, American Thoracic Society, and the American Society for Clinical Investigation (ASCI), of which he has served as a Councilor and Vice President.

In his clinical research, Dr. Moss has focused on destructive lung diseases (e.g., lymphangioleiomyomatosis, cystic fibrosis), with primary emphasis on the roles of infection/inflammation and susceptibility/modifier genes on disease progression and severity. His LAM research centers on the abnormal smooth muscle-like cells (LAM cells) that proliferate in the lungs, lymphatics, and kidneys, and are responsible for cystic lung destruction and abnormalities in the lymphatic system. Subjects of his basic research include guanine nucleotide-binding regulatory proteins and post-translational modification of proteins by ADP-ribosylation.

Carole Parent obtained her B.Pharm. and M.Sc. degrees from the University of Montréal in Canada and her Ph.D. in Pharmacy from the University of Illinois at Chicago. She then pursued postdoctoral studies in the laboratory of Dr. Peter Devreotes at Johns Hopkins. In 2000, she joined the Laboratory of Cellular and Molecular Biology at the Center for Cancer Research (NCI, NIH) where she is now Senior Investigator and Deputy Chief. Using three complementary model systems, her research interests are centered on understanding how cells detect and respond to external chemotactic signals and, in particular, how the relay of chemotactic signals from cell-to-cell impacts single and group cell migration.

CONTRIBUTORS

A Abe
University of Michigan, Ann Arbor, MI, USA

P Ädelroth
Stockholm University, Stockholm, Sweden

A Adessi
University of Florence, Florence, Italy

P Agre
*Johns Hopkins Bloomberg School of Public Health,
Baltimore, MD, USA*

S Akira
*World Premier International Immunology Frontier
Research Center, Osaka, Japan; Osaka University, Osaka,
Japan*

E Alani
Cornell University, Ithaca, NY, USA

S Al-Attar
Wageningen University, Wageningen, The Netherlands

R W Alexander
Wake Forest University, Winston-Salem, NC, USA

W Ali
Stony Brook University, Stony Brook, NY, USA

A Aliverti
Università degli Studi di Milano, Milan, Italy

I S Ambudkar
*National Institute of Dental and Craniofacial Research,
Bethesda, MD, USA*

A Amcheslavsky
University of California, Irvine, CA, USA

R A Anderson
University of Wisconsin, Madison, WI, USA

M-L Andréola
*UMR 5234 CNRS, Université Victor Segalen
Bordeaux 2, Bordeaux, France*

E Antony
*Washington University School of Medicine, St. Louis,
MO, USA*

E Arnold
Rutgers University, Piscataway, NJ, USA

E-M Aro
University of Turku, Turku, Finland

M Ascoli
*The University of Iowa Carver College of Medicine, Iowa
City, IA, USA*

J A Askari
University of Manchester, Manchester, UK

J Atkinson
Brock University, St. Catharines, ON, Canada

D S Auld
Harvard Medical School, Boston, MA, USA

A Bacolla
The University of Texas, Smithville, TX, USA

H Bading
University of Heidelberg, Heidelberg, Germany

S F Bairstow
Baxter Healthcare Corporation, Round Lake, IL, USA

T O Baldwin
University of Arizona, Tucson, AZ, USA

S W Ballinger
*University of Alabama at Birmingham, Birmingham,
AL, USA*

D P Ballou
University of Michigan, Ann Arbor, MI, USA

J R Bamburg
Colorado State University, Fort Collins, CO, USA

V Bandarian
*University of Michigan, Ann Arbor, MI, USA; University
of Arizona, Tucson, AZ, USA*

D Baniulis
*Lithuanian Research Centre for Agriculture and Forestry,
Kaunas, Lithuania*

O Bannach
*Heinrich-Heine-Universität Düsseldorf, Düsseldorf,
Germany*

J Barber
Imperial College, London, UK

K Barbieri
University of Minnesota, Minneapolis, MN, USA

B Barker
Thomas Jefferson University, Philadelphia, PA, USA

E J Bartlett
University of Sussex, Brighton, UK

C E Barton
Vanderbilt University, Nashville, TN, USA

R Bassel-Duby
*University of Texas Southwestern Medical Center, Dallas,
TX, USA*

H Baum
King's College London, London, UK

J W Baynes
University of South Carolina, Columbia, SC, USA

W A Beard
*National Institute of Environmental Health Sciences,
National Institutes of Health, Research Triangle Park,
NC, USA*

D H Bechhofer
Mount Sinai School of Medicine, New York, NY, USA

S J Benkovic
*The Pennsylvania State University, University Park, PA,
USA*

R A O Bennett
*Stony Brook University School of Medicine, Stony Brook,
NY, USA*

J L Benovic
Thomas Jefferson University, Philadelphia, PA, USA

P Bernardi
University of Padua, Padua, Italy

R Bernasconi
*Institute for Research in Biomedicine, Bellinzona,
Switzerland*

R Bernhardt
Saarland University, Saarbrücken, Germany

K M Bernot
*Johns Hopkins University, Bloomberg School of Public
Health, Baltimore, MD, USA*

M J Berridge
The Babraham Institute, Cambridge, UK

J M Berrisford
*Medical Research Council Mitochondrial Biology Unit,
Cambridge, UK*

D M Bers
University of California at Davis, Davis, CA, USA

A Bertoli
University of Padua, Padua, Italy

L Bettendorff
University of Liège, Liège, Belgium

U Beuers
University of Amsterdam, Amsterdam, The Netherlands

P J Beuning
Northeastern University, Boston, MA, USA

A S Bhagwat
Wayne State University, Detroit, MI, USA

S Bheemanaik
Indian Institute of Science, Bangalore, India

A Bianchi
University of Sussex, Brighton, UK

M Biel
Ludwig-Maximilians-Universität, Munich, Germany

A Bindoli
University of Padua, Padua, Italy

J A Birchler
University of Missouri, Columbia, MO, USA

H Birkedal-Hansen
National Institutes of Health, Bethesda, MD, USA

S M Blackman
*Johns Hopkins University School of Medicine, Baltimore,
MD, USA*

S C Blanchard
Weill Cornell Medical College, New York, NY, USA

R E Blankenship
Washington University in St. Louis, St. Louis, MO, USA

L B Bloom
University of Florida, Gainesville, FL, USA

A Böck
University of Munich, Munich, Germany

K L Bogan
The University of Iowa, Iowa City, IA, USA

R Borghese
University of Bologna, Bologna, Italy

S Borukhov
*University of Medicine and Dentistry of New Jersey,
School of Osteopathic Medicine, Stratford, NJ, USA*

J D Boughter Jr.
*University of Tennessee Health Science Center, Memphis,
TN, USA*

F Bouillaud
*Centre National de la Recherche Scientifique and
Université Paris Descartes, Paris, France*

A Boussac
*Centre National de la Recherche Scientifique, Gif-sur-
Yvette, France*

A Boveris
University of Buenos Aires, Buenos Aires, Argentina

N G Bowery
*University of Birmingham Medical School, Birmingham,
UK*

A Bracher
*Max Planck Institute of Biochemistry, Martinsried,
Germany*

R A Bradshaw
University of California, Irvine, CA, USA

L Brandi
University of Camerino, Camerino, Italy

U Brandt
Goethe University, Frankfurt am Main, Germany

C Brenner
The University of Iowa, Iowa City, IA, USA

M D Breyer
Vanderbilt University, Nashville, TN, USA

R M Breyer
Vanderbilt University, Nashville, TN, USA

M Brini
University of Padua, Padua, Italy

H L Brockman
University of Minnesota, Austin, MN, USA

S J J Brouns
Wageningen University, Wageningen, The Netherlands

D A Brown
State University of New York, Stony Brook, NY, USA

E M Brown
Harvard Medical School, Boston, MA, USA

R B Brown
*University of Nebraska Medical Center, Omaha, NE,
USA*

M J Brownstein
Azevan Pharmaceuticals, Bethlehem, PA, USA

D A Bryant
*The Pennsylvania State University, University Park, PA,
USA*

P Brzezinski
Stockholm University, Stockholm, Sweden

M W Bunce
*The Children's Hospital of Philadelphia, Philadelphia,
PA, USA*

R M Burch
Naurex, Inc., Morris, CT, USA

P M J Burgers
*Washington University School of Medicine, St. Louis,
MO, USA*

S J W Busby
The University of Birmingham, Birmingham, UK

M Butala
University of Ljubljana, Ljubljana, Slovenia

D J Buttle
University of Sheffield, Sheffield, UK

D B Bylund
*University of Nebraska Medical Center, Omaha, NE,
USA*

E Cadenas
University of Southern California, Los Angeles, CA, USA

M D Cahalan
University of California, Irvine, CA, USA

R Cammack
King's College London, London, UK

M Campeol
*University of Padua and Venetian Institute of Molecular
Medicine, Padua, Italy*

J Capecchi
Tufts University School of Medicine, Boston, MA, USA

E Carafoli
University of Padua, Padua, Italy

J J Caramelo
Fundación Instituto Leloir, Buenos Aires, Argentina

G Carignani
University of Padua, Padua, Italy

G M Carman
Rutgers University, New Brunswick, NJ, USA

D D Carson
Rice University, Houston, TX, USA

B D Carter
Vanderbilt University Medical School, Nashville, TN, USA

P W Caton
Queen Mary University of London, London, UK

W A Catterall
University of Washington, Seattle, WA, USA

J J Champoux
University of Washington, Seattle, WA, USA

B Chance
University of Pennsylvania School of Medicine, Philadelphia, PA, USA

M G Chandler
CNRS, Toulouse, France

A Chapman-Smith
The University of Adelaide, Adelaide, SA, Australia

D J Chen
University of Texas Southwestern Medical Center, Dallas, TX, USA

X Cheng
Emory University School of Medicine, Atlanta, GA, USA

I Chesnokov
University of Alabama, Birmingham, AL, USA

T C Cheung
University of California, San Diego, CA, USA

S Christakos
University of Medicine and Dentistry of New Jersey, Newark, NJ, USA

K Christian
Johns Hopkins University School of Medicine, Baltimore, MD, USA

D T Chuang
University of Texas Southwestern Medical Center, Dallas, TX, USA

J T Chun
Stazione Zoologica Anton Dohrn, Villa Comunale, Napoli, Italy

A Ciechanover
Israel Institute of Technology, Haifa, Israel

M Clagett-Dame
University of Wisconsin–Madison, Madison, WI, USA

S D Clarke
City University of New York, Brooklyn, NY, USA

D A Clayton
Howard Hughes Medical Institute, Ashburn, VA, USA

W W Cleland
University of Wisconsin, Madison, WI, USA

S D Clouse
North Carolina State University, Raleigh, NC, USA

R J Colbran
Vanderbilt University, Nashville, TN, USA

K J Colley
University of Illinois, Chicago, IL, USA

M Colombini
University of Maryland, College Park, MD, USA

R C Conaway
Stowers Institute for Medical Research, Kansas City, MO, USA

J W Conaway
Stowers Institute for Medical Research, Kansas City, MO, USA

R D Cone
Vanderbilt University Medical Center, Nashville, TN, USA

P J Conn
Vanderbilt University Medical Center, Nashville, TN, USA

A C Conner
Warwick University, Warwick, UK

D W Cooke
Johns Hopkins University School of Medicine, Baltimore, MD, USA

J A Cooper
Fred Hutchinson Cancer Research Center, Seattle, WA, USA

B S Cooperman
University of Pennsylvania, Philadelphia, PA, USA

D Corbeil
Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

A H Corbett
Emory University School of Medicine, Atlanta, GA, USA

J A Corbin
University of Colorado, Boulder, CO, USA

J D Corbin
Vanderbilt University School of Medicine, Nashville, TN, USA

J L Corden
Johns Hopkins University School of Medicine, Baltimore, MD, USA

P A Coulombe
Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD, USA

P Cramer
Ludwig-Maximilians-Universität München, Munich, Germany

W A Cramer
Purdue University, West Lafayette, IN, USA

N Cremona
Case Western Reserve University School of Medicine, Cleveland, OH, USA

L Cynober
Paris Descartes University, Paris, France

M P Czech
University of Massachusetts Medical School, Worcester, MA, USA

I Dach
Aarhus University, Aarhus, Denmark

H A Dailey
University of Georgia, Athens, GA, USA

K T Dalen
University of Oslo, Oslo, Norway

V M Darley-Usmar
University of Alabama at Birmingham, Birmingham, AL, USA

K Das
Rutgers University, Piscataway, NJ, USA

S C Daubner
Texas A&M University, College Station, TX, USA

E W Davie
University of Washington, Seattle, WA, USA

M de Caestecker
Vanderbilt University, Nashville, TN, USA

B De Hertogh
Facultés Universitaires Notre-Dame de la Paix, Namur, Belgium

R De Philippis
University of Florence, Florence, Italy

M Debidia
University of Cambridge, Cambridge, UK

E Degerman
Lund University, Lund, Sweden

J A Della-Maria
University of Maryland School of Medicine, Baltimore, MD, USA

L J Dell'Italia
University of Alabama at Birmingham, Birmingham, AL, USA

H F DeLuca
University of Wisconsin–Madison, Madison, WI, USA

B Demple
Stony Brook University School of Medicine, Stony Brook, NY, USA

C J Der
University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

R J Deschenes
College of Medicine, Tampa, FL, USA

S Despa
University of California at Davis, Davis, CA, USA

P N Devreotes
Johns Hopkins University School of Medicine, Baltimore, MD, USA

N Dharmaraj
Rice University, Houston, TX, USA

D C Di Giammartino
Columbia University, New York, NY, USA

F Di Lisa
University of Padua, Padua, Italy

G Dialynas
University of Iowa, Iowa City, IA, USA

C L Dieckmann
University of Arizona, Tucson, AZ, USA

A M Diehl
Duke University Medical Center, Durham, NC, USA

S DiMauro
Columbia University, New York, NY, USA

A A DiSpirito
Iowa State University, Ames, IA, USA

J E Dixon
University of California, San Diego, CA, USA

A J Doherty
University of Sussex, Brighton, UK

J G Donaldson
National Institutes of Health, Bethesda, MD, USA

C J Dorman
Trinity College Dublin, Dublin, Republic of Ireland

D-W Doug Chung
Institute for Integrative Genome Biology, University of California Riverside, Riverside, CA, USA

R N Douglas
University of Missouri, Columbia, MO, USA

D R Dowd
Case Western Reserve University, Cleveland, OH, USA

S F Dowdy
University of California, La Jolla, CA, USA

C P Downes
University of Dundee, Dundee, UK

A C Drohat
University of Maryland School of Medicine, Baltimore, MD, USA

D T F Dryden
University of Edinburgh, Edinburgh, UK

G R Dubyak
Case Western Reserve University, Cleveland, OH, USA

B M Dunn
University of Florida College of Medicine, Gainesville, FL, USA

R L Dusek
Stanford University School of Medicine, Stanford, CA, USA

N Dyson
Massachusetts General Hospital Cancer Center, Boston, MA, USA

B A Eipper
University of Connecticut Health Center, Farmington, CT, USA

C P Elco
Cleveland Clinic Foundation, Cleveland, OH, USA; Case Western Reserve University, Cleveland, OH, USA

S C R Elgin
Washington University, St. Louis, MO, USA

S D Emr
Cornell University, Ithaca, NY, USA

S J Enna
University of Kansas Medical Center, Kansas City, KS, USA

B Entsch
University of Michigan, Ann Arbor, MI, USA

V Ermer
German Centre for Neurodegenerative Diseases, DZNE, Bonn, Germany

K Erreger
Vanderbilt University, Nashville, TN, USA

J D Esko
University of California, San Diego, La Jolla, CA, USA

P J Espenshade
Johns Hopkins University School of Medicine, Baltimore, MD, USA

J Everse
Texas Tech University Health Sciences Center, Lubbock, TX, USA

J H Exton
Vanderbilt University, Nashville, TN, USA

J J Falke
University of Colorado, Boulder, CO, USA

B H Falkenburger
University of Washington School of Medicine, Seattle, WA, USA

G-H Fan
VA Medical Center, Nashville, TN, USA; Vanderbilt University School of Medicine, Nashville, TN, USA

O Fedorova
Yale University, New Haven, CT, USA

A R Ferré-D'Amaré
National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA

G C Ferreira
University of South Florida, Tampa, FL, USA

C Ficicioglu
University of Pennsylvania School of Medicine, Philadelphia, PA, USA

C M Field
Harvard Medical School, Boston, MA, USA

A Finazzi Agrò
University of Rome, Rome, Italy

P F Fitzpatrick
University of Texas Health Science Center, San Antonio, TX, USA

G Floris
University of Cagliari, Monserrato (CA), Italy

M Forgac
Tufts University School of Medicine, Boston, MA, USA

S L Forsburg
University of Southern California, Los Angeles, CA, USA

F Förster
Max Planck Institute of Biochemistry, Martinsried, Germany

J A Foster
Boston University School of Medicine, Boston, MA, USA

S H Francis
Vanderbilt University School of Medicine, Nashville, TN, USA

H H Freeze
Sanford – Burnham Medical Research Institute, La Jolla, CA, USA

P A Frey
University of Wisconsin, Madison, WI, USA

I Fridovich
Duke University Medical Center, Durham, NC, USA

E C Friedberg
University of Texas Southwestern Medical Center, Dallas, TX, USA

A D Friedman
Johns Hopkins University School of Medicine, Baltimore, MD, USA

M A Frohman
Stony Brook University, Stony Brook, NY, USA

M Frolov
Massachusetts General Hospital Cancer Center, Boston, MA, USA

P Fromme
Arizona State University, Tempe, AZ, USA

P Fuchs
University of Vienna, Vienna, Austria

N Furnham
European Bioinformatics Institute, Cambridge, UK

A Furukohri
Nara Institute of Science and Technology, Ikoma, Japan

E J Gallagher
Mount Sinai Medical Center, New York, NY, USA

A Galli
Vanderbilt University, Nashville, TN, USA

J S Garavelli
University of Delaware, Newark, DE, USA

T J Gardella
Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

B Geiger
The Weizmann Institute of Science, Rehovot, Israel

A A Genazzani
University of Cambridge, Cambridge, UK

J Geng
University of Michigan, Ann Arbor, MI, USA

M L Genova
Alma Mater Studiorum – Università di Bologna, Bologna, Italy

A C Gentry
Vanderbilt University School of Medicine, Nashville, TN, USA

T W Gettys
Pennington Biomedical Research Center, Baton Rouge, LA, USA

M L Ghirardi
Bioscience Center, National Renewable Energy Laboratory (NREL), Golden, CO, USA

M Giacomello
University of Padua and Venetian Institute of Molecular Medicine, Padua, Italy

G M Giacometti
Università di Padua, Padua, Italy

H F Gilbert
Baylor College of Medicine, Houston, TX, USA

T D Gilmore
Boston University, Boston, MA, USA

S I Girgis
Imperial College London, London, UK

M H Glickman
Israel Institute of Technology, Haifa, Israel

A Goffeau
Université Catholique de Louvain, Louvain-la-Neuve, Belgium

J H Golbeck
The Pennsylvania State University, University Park, PA, USA

A L Goldberg
Harvard Medical School, Boston, MA, USA

D P Goldenberg
University of Utah, Salt Lake City, UT, USA

B L Golden
Purdue University, West Lafayette, IN, USA

L E Goldfinger
Temple University School of Medicine, Philadelphia, PA, USA

L E Goldsmith
Duke University Medical Center, Durham, NC, USA

E A Golemis
Fox Chase Cancer Center, Philadelphia, PA, USA

P Gollnick
University at Buffalo, Buffalo, NY, USA

N Goosen
Leiden University, Leiden, The Netherlands

J I Gordon
Washington University School of Medicine, St. Louis, MO, USA

S J Gould
Johns Hopkins University School of Medicine, Baltimore, MD, USA

E Gracheva
Medros Inc., St. Louis, MO, USA

P Graham
Rutgers University, Piscataway, NJ, USA

K J Green
Northwestern University Feinberg School of Medicine, Chicago, IL, USA

K J Gregory
Vanderbilt University Medical Center, Nashville, TN, USA

S T Gregory
Brown University, Providence, RI, USA

P J Gresch
Vanderbilt University-School of Medicine, Nashville, TN, USA

J Groenendyk
University of Alberta, Edmonton, AB, Canada

E L Gross
Ohio State University, Columbus, OH, USA

I Grotjohann
Arizona State University, Tempe, AZ, USA

C O Gualerzi
University of Camerino, Camerino, Italy

J-L Guan
University of Michigan, Ann Arbor, MI, USA

L Guan
Texas Tech University Health Sciences Center, Lubbock, TX, USA

J S Gutkind
National Institute of Dental and Craniofacial Research, Bethesda, MD, USA

C S Guy
St. Jude Children's Research Hospital, Memphis, TN, USA

S Hakomori
University of Washington, Seattle, WA, USA

F Haleem
University of Medicine and Dentistry of New Jersey, Newark, NJ, USA

S E Halford
University of Bristol, Bristol, UK

R S Haltiwanger
Stony Brook University, Stony Brook, NY, USA

G-S Han
Rutgers University, New Brunswick, NJ, USA

S K Hanks
Vanderbilt University School of Medicine, Nashville, TN, USA

K K Hansen
University of Calgary, Calgary, AB, Canada

R W Hanson
Case Western Reserve University School of Medicine, Cleveland, OH, USA

J C Harr
Johns Hopkins University School of Medicine, Baltimore, MD, USA

R A Harris
Indiana University School of Medicine, Indianapolis, IN, USA

G W Hart
Johns Hopkins University, Baltimore, MD, USA

P E Hart
University College London, London, UK

F U Hartl
Max Planck Institute of Biochemistry, Martinsried, Germany

S S Hasan
Purdue University, West Lafayette, IN, USA

H Hashimoto
Emory University School of Medicine, Atlanta, GA, USA

S Hata
Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan

Y Hatori
Johns Hopkins University, Baltimore, MD, USA

H Haviv
Weizmann Institute of Science, Rehovoth, Israel

L A Hawver
Northeastern University, Boston, MA, USA

O Hayaishi
Osaka Bioscience Institute, Osaka, Japan

D L Hay
Auckland University, Auckland, New Zealand

A J R Heck
Biomolecular Mass Spectrometry and Proteomics, Utrecht University, Utrecht, The Netherlands; Netherlands Proteomics Center, Utrecht, The Netherlands

R L Heinrikson
The Pharmacia Corporation, Kalamazoo, MI, USA

C U T Hellen
State University of New York Downstate Medical Center, Brooklyn, NY, USA

J D Helmann
Cornell University, Ithaca, NY, USA

W M Henne
Cornell University, Ithaca, NY, USA

D Hereld
National Institute on Alcohol Abuse and Alcoholism, Rockville, MD, USA

J M Herrmann
University of Kaiserslautern, Kaiserslautern, Germany

W-D Heyer
University of California, Davis, Davis, CA, USA

I D Hickson
University of Oxford, Oxford, UK

H N Higgs
Dartmouth Medical School, Hanover, NH, USA

J D Hildebrandt
Medical University of South Carolina, Charleston, SC, USA

B Hille
University of Washington School of Medicine, Seattle, WA, USA

N Hirokawa
University of Tokyo, Tokyo, Japan

C B Hirschberg
Boston University School of Dental Medicine, Boston, MA, USA

C Hisatsune
RIKEN Brain Science Institute, Saitama, Japan

B T Hoang
CNRS, Toulouse, France

P G Hogan
La Jolla Institute for Allergy and Immunology, La Jolla, CA, USA

M Hogarth
Burnet Institute, Prahran, VIC, Australia

M D Hollenberg
University of Calgary, Calgary, AB, Canada

M A Hollingsworth
University of Nebraska Medical Center, Omaha, NE, USA

N M Hollingsworth
Stony Brook University, Stony Brook, NY, USA

K Holmbeck
National Institutes of Health, Bethesda, MD, USA

M J Holness
Queen Mary University of London, London, UK

E-J E Hong
University of Massachusetts Medical School, Worcester, MA, USA

A B Hooper
University of Minnesota, St. Paul, MN, USA

C L Hoppel
Case Western Reserve University School of Medicine, Cleveland, OH, USA; Veterans Affairs Medical Center, Cleveland, OH, USA

P Horton
University of Sheffield, Sheffield, UK

K M Howard
University of Nevada Las Vegas School of Dental Medicine, Las Vegas, NV, USA

P Hsieh
National Institutes of Health, Bethesda, MD, USA

T-S Hsieh
Duke University Medical Center, Durham, NC, USA

A Hudmon
Indiana University School of Medicine, Indianapolis, IN, USA

S H Hughes
National Institutes of Health, Frederick, MD, USA

M J Humphries
University of Manchester, Manchester, UK

N Hunter
Howard Hughes Medical Institute, University of California Davis, Davis, CA, USA

W B Huttner
Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

T Inagami
Vanderbilt University, Nashville, TN, USA

G Inesi
California Pacific Medical Center Research Institute, San Francisco, CA, USA

T S Ingebritsen
Iowa State University, Ames, IA, USA

G Ira
Baylor College of Medicine, Houston, TX, USA

M Isalan
Centre for Genomic Regulation (CRG), UPF, Barcelona, Spain

G H Jacobs
University of California, Santa Barbara, CA, USA

B Jagannathan
The Pennsylvania State University, University Park, PA, USA

C Janetopoulos
Vanderbilt University, Nashville, TN, USA

S Jansson
University of Umeå, Umeå, Sweden

J B Jensen
University of Washington School of Medicine, Seattle, WA, USA

J H Jo
National Renewable Energy Laboratory, Golden, CO, USA

K A Johnson
University of Texas, Austin, TX, USA

J C R Jones
Northwestern University Feinberg School of Medicine, Chicago, IL, USA

L L Jones
St. Jude Children's Research Hospital, Memphis, TN, USA

M M Jore
Wageningen University, Wageningen, The Netherlands

S Joshi
University of California San Diego, La Jolla, CA, USA

C M Joyce
Yale University, New Haven, CT, USA

S-H Jun
The Pennsylvania State University, University Park, PA, USA

M S Jurica
University of California Santa Cruz, Santa Cruz, CA, USA

H R Kaback
University of California, Los Angeles, Los Angeles, CA, USA

J M Kaguni
Michigan State University, East Lansing, MI, USA

Y Kamimura
Johns Hopkins University School of Medicine, Baltimore, MD, USA

E Kanervo
University of Turku, Turku, Finland

D S Kang
Thomas Jefferson University, Philadelphia, PA, USA

M M Kaplan
Tufts University School of Medicine, Boston, MA, USA

S J D Karlish
Weizmann Institute of Science, Rehovoth, Israel

A Kazlauskas
Harvard Medical School, Boston, MA, USA

P J Keely
University of Wisconsin, Madison, WI, USA

J Keller
Israelitic Hospital, Hamburg, Germany

J B Kelley
University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

L M Kelman
Montgomery College, Germantown, MD, USA

Z Kelman
Institute for Bioscience and Biotechnology Research, Rockville, MD, USA; University of Maryland, College Park, MD, USA

R S Kenchappa
Vanderbilt University Medical School, Nashville, TN, USA

J Kerner
Case Western Reserve University School of Medicine, Cleveland, OH, USA

A R Kimmel
National Institutes of Health, Bethesda, MD, USA

C B Klee
National Institutes of Health, Bethesda, MD, USA

K R Kligys
Northwestern University Feinberg School of Medicine, Chicago, IL, USA

J P Klinman
University of California, Berkeley, CA, USA

D J Klionsky
University of Michigan, Ann Arbor, MI, USA

R A Kohanski
Johns Hopkins University, Baltimore, MD, USA

A A Komar
*Cleveland State University, Cleveland, OH, USA;
Case Western Reserve University, Cleveland, OH, USA*

J B Konopka
Stony Brook University, Stony Brook, NY, USA

M Korc
Dartmouth Medical School, Hanover, NH, USA

M Kornmann
University of Ulm, Ulm, Germany

S I Korsching
University of Cologne, Cologne, Germany

M V Kovalenko
Harvard Medical School, Boston, MA, USA

J Krebs
*Max Planck Institute for Biophysical Chemistry,
Göttingen, Germany*

G Kreil
University of Vienna, Vienna, Austria

R H Kretsinger
University of Virginia, Charlottesville, VA, USA

K N Kreuzer
Duke University Medical Center, Durham, NC, USA

F P Kuhajda
*Johns Hopkins University School of Medicine, Baltimore,
MD, USA*

J D Kulman
University of Washington, Seattle, WA, USA

H Kumar
*Indian Institute of Science Education and Research
(IISER), Bhopal, India; World Premier International
Immunology Frontier Research Center, Osaka, Japan;
Osaka University, Osaka, Japan*

M M Kwatra
Duke University Medical Center, Durham, NC, USA

B J Lamarche
The Ohio State University, Columbus, OH, USA

C R D Lancaster
Saarland University, Homburg (Saar), Germany

B R Landau
Case Western Reserve University, Cleveland, OH, USA

M D Lane
*Johns Hopkins University School of Medicine, Baltimore,
MD, USA*

A Lange
Emory University School of Medicine, Atlanta, GA, USA

R E Larson
University of Sao Paulo, Sao Paulo, Brazil

R Latorre
Universidad de Valparaíso, Valparaíso, Chile

B Laube
Technical University of Darmstadt, Darmstadt, Germany

P Y Law
University of Minnesota, Minneapolis, MN, USA

J B Lawrence
*University of Massachusetts Medical School, Worcester,
MA, USA*

C Lee
Cornell University, Ithaca, NY, USA

C-H Lee
*Johns Hopkins University, Bloomberg School of Public
Health, Baltimore, MD, USA*

R T Lee
Johns Hopkins University, Baltimore, MD, USA

Y C Lee
Johns Hopkins University, Baltimore, MD, USA

R C Leegood
University of Sheffield, Sheffield, UK

N Lehman
Portland State University, Portland, OR, USA

A R Lehmann
University of Sussex, Brighton, UK

M A Lehrman
*University of Texas Southwestern Medical Center, Dallas,
TX, USA*

G Lenaz
*Alma Mater Studiorum – Università di Bologna, Bologna,
Italy*

K G Le Roch
*Institute for Integrative Genome Biology, University of
California Riverside, Riverside, CA, USA*

D LeRoith
Mount Sinai Medical Center, New York, NY, USA

N R Leslie
University of Dundee, Dundee, UK

S J Levine
National Institutes of Health, Bethesda, MD, USA

L Leybaert
Ghent University, Ghent, Belgium

G-D Li
University of California, Los Angeles, CA, USA

G-M Li
University of Kentucky Medical Center, Lexington, KY, USA

S Li
National Institutes of Health, Bethesda, MD, USA

Z Li
Florida Atlantic University, Boca Raton, FL, USA

M Lieb
University of Southern California, Los Angeles, CA, USA

R K H Liem
Columbia University College of Physicians and Surgeons, New York, NY, USA

M Lima
University of South Carolina, Columbia, SC, USA

M E Linder
Cornell University, Ithaca, NY, USA

A J Lindsay
University College Cork, Cork, Republic of Ireland

H Lis
Weizmann Institute of Science, Rehovot, Israel

S Litvak
UMR 5234 CNRS, Université Victor Segalen Bordeaux 2, Bordeaux, France

K Liu
Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

X Liu
National Institute of Dental and Craniofacial Research, Bethesda, MD, USA

M Löffler
Philipps-University, Marburg, Germany

H H Loh
University of Minnesota, Minneapolis, MN, USA

T M Lohman
Washington University School of Medicine, St. Louis, MO, USA

D Lohr
Arizona State University, Tempe, AZ, USA

E London
Stony Brook University, Stony Brook, NY, USA

C Londres
National Institutes of Health, Bethesda, MD, USA

G D Lopaschuk
University of Alberta, Edmonton, AB, Canada

J Lowery
Harvard School of Dental Medicine, Boston, MA, USA

H B Lowman
NGM Biopharmaceuticals, Inc., San Francisco, CA, USA

M Lowndes
Stanford University, Stanford, CA, USA

W Lubitz
Max-Planck-Institut für Bioanorganische Chemie, Mülheim an der Ruhr, Germany

A-L Lu
University of Maryland, Baltimore, MD, USA

L-S Lu
Texas Heart Institute at St. Luke's Episcopal Hospital, Houston, TX, USA

A N Lupas
Max Planck Institute for Developmental Biology, Tübingen, Germany

S Lutsenko
Johns Hopkins University, Baltimore, MD, USA

T Lynagh
Technical University of Darmstadt, Darmstadt, Germany

K Lynch
University of Pennsylvania School of Medicine, Philadelphia, PA, USA

J Lytton
University of Calgary, Calgary, AB, Canada

P N MacDonald
Case Western Reserve University, Cleveland, OH, USA

S R Mackey
St. Ambrose University, Davenport, IA, USA

S R MacLellan
University of New Brunswick, Fredericton, NB, Canada

A S Maddox
Université de Montréal, Montréal, QC, Canada

L J Mady
University of Medicine and Dentistry of New Jersey, Newark, NJ, USA

L Magnusson
National Renewable Energy Laboratory, Golden, CO,
USA

L Maillette de Buy Wenniger
University of Amsterdam, Amsterdam, The Netherlands

R E Mains
University of Connecticut Health Center, Farmington,
CT, USA

M Maiorino
University of Padua, Padua, Italy

A Maiti
University of Maryland School of Medicine, Baltimore,
MD, USA

S Maiti
IISER Kolkata, Mohanpur, India

H Maki
Nara Institute of Science and Technology, Ikoma, Japan

L Mancini
Pistoia, Italy

P-C Maness
National Renewable Energy Laboratory, Golden, CO,
USA

V C Manganiello
National Institutes of Health, Bethesda, MD, USA

J L Manley
Columbia University, New York, NY, USA

C A Mannella
New York State Department of Health, Albany, NY, USA

D Manor
Western Reserve University, Cleveland, OH, USA

M G Marinus
University of Massachusetts Medical School, Worcester,
MA, USA

M A Marletta
University of California, Berkeley, CA, USA

J Martin
Max Planck Institute for Developmental Biology,
Tübingen, Germany

M R Marzahn
University of Florida, Gainesville, FL, USA

A B Mason
Yale University, New Haven, CT, USA

I Matic
INSERM U1001, Université Paris Descartes, Paris,
France

F M Matschinsky
University of Pennsylvania School of Medicine,
Philadelphia, PA, USA

A Mattevi
University of Pavia, Pavia, Italy

K S Matthews
Rice University, Houston, TX, USA

R G Matthews
University of Michigan, Ann Arbor, MI, USA

H J G Matthies
Vanderbilt University, Nashville, TN, USA

M W McCaffrey
University College Cork, Cork, Republic of Ireland

C P McCann
University of Maryland College Park, College Park, MD,
USA

D R McCaslin
University of Wisconsin, Madison, WI, USA

R R McCully
Johns Hopkins University School of Medicine, Baltimore,
MD, USA

A E Medlock
University of Georgia, Athens, GA, USA

G Meissner
University of North Carolina, Chapel Hill, NC, USA

G B Melikyan
Emory University School of Medicine, Atlanta, GA, USA;
Children's Healthcare of Atlanta, Atlanta, GA, USA

G Melino
Leicester University, Leicester, UK; University of Rome
"Tor Vergata", Biochemistry IDI-IRCCS laboratory,
Rome, Italy

A K Menon
Weill Cornell Medical College, New York, NY, USA

R I Menz
Flinders University, Adelaide, SA, Australia

W C Merrick
Cleveland State University, Cleveland, OH, USA; Case
Western Reserve University, Cleveland, OH, USA

A H Merrill Jr.
Georgia Institute of Technology, Atlanta, GA, USA

E K Michaelis
University of Kansas, Lawrence, KS, USA

M Michalak
University of Alberta, Edmonton, AB, Canada

H Michel
Max Planck Institute of Biophysics, Frankfurt, Germany

R L Miesfeld
University of Arizona, Tucson, AZ, USA

K Mikoshiba
RIKEN Brain Science Institute, Saitama, Japan

B A Miller
*Pennsylvania State University College of Medicine,
Hershey, PA, USA*

J L S Milne
National Institutes of Health, Bethesda, MD, USA

G-L Ming
*Johns Hopkins University School of Medicine, Baltimore,
MD, USA*

F Minier
François-Rabelais University, Tours, France

M Molinari
*Ecole Polytechnique Fédérale de Lausanne, Lausanne,
Switzerland*

A L Moore
University of Sussex, Brighton, UK

M S Moore
Baylor College of Medicine, Houston, TX, USA

S Mora
University of Liverpool, Liverpool, UK

N Moradi-Bidhendi
*Barts and The London School of Medicine and Dentistry,
London, UK*

T H Moran
*Johns Hopkins University School of Medicine, Baltimore,
MD, USA*

F J Morera
*University of California, Los Angeles, Los Angeles, CA,
USA*

S Morimoto
*Kyushu University Graduate School of Medicine,
Fukuoka, Japan*

T Morosinotto
Università di Padova, Padova, Italy

J S Mort
Shriners Hospital for Children, Montréal, QC, Canada

G J Morton
University of Washington, Seattle, WA, USA

G Mosedale
University of Oxford, Oxford, UK

J Moss
*Cardiovascular and Pulmonary Branch of the National
Heart, Lung, and Blood Institute in Bethesda, Bethesda,
MD, USA*

L Motta-Mena
*University of Pennsylvania School of Medicine,
Philadelphia, PA, USA*

R D Mullins
University of California, San Francisco, CA, USA

S D Munger
*University of Maryland School of Medicine, Baltimore,
MD, USA*

H Münzberg
*Pennington Biomedical Research Center, Baton Rouge,
LA, USA*

K S Murakami
*The Pennsylvania State University, University Park, PA,
USA*

G Murphy
University of Cambridge, Cambridge, UK

R C Murphy
University of Colorado Medical School, Denver, CO, USA

U M N Murthy
*Bioscience Center, National Renewable Energy
Laboratory (NREL), Golden, CO, USA*

H Nagase
University of Oxford, Oxford, UK

M T Nakamura
*University of Illinois at Urbana–Champaign, Urbana, IL,
USA*

J L Napoli
University of California, Berkeley, CA, USA

J R Naranjo
Centro Nacional de Biotecnología – CSIC, Madrid, Spain

A Narasimha
University of California, La Jolla, CA, USA

N M Nathanson
University of Washington, Seattle, WA, USA

W J Nelson
Stanford University, Stanford, CA, USA

W Neupert
*Max-Planck-Institute of Biochemistry, Munich-
Martinsried, Germany*

K A Neve
*Portland Veterans Affairs Medical Center, Portland, OR,
USA*

A C Newton
University of California, San Diego, CA, USA

B G Ng
Sanford – Burnham Medical Research Institute, La Jolla, CA, USA

D T W Ng
Temasek Life Sciences Laboratory and Department of Biological Sciences, National University of Singapore, Singapore

D G Nicholls
Buck Institute for Age Research, Novato, CA, USA

D W Nicholson
Merck Frosst Centre for Therapeutic Research, Pointe Claire – Dorval, QC, Canada

M J Niciu
University of Connecticut Health Center, Farmington, CT, USA

D A Nicoll
David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

P Nicotera
German Centre for Neurodegenerative Diseases, DZNE, Bonn, Germany

P Nissen
Aarhus University, Aarhus, Denmark

E Nogales
University of California, Berkeley, CA, USA

B Ochoa
University of the Basque Country UPV/EHU, Bilbao, Spain

M O'Donnell
Rockefeller University, Howard Hughes Medical Institute, New York, NY, USA

S Offermanns
Max-Planck-Institute for Heart and Lung Research, Bad Nauheim, Germany

H Ogata
Max-Planck-Institut für Bioanorganische Chemie, Mülheim an der Ruhr, Germany

R Ohi
Vanderbilt University Medical Center, Nashville, TN, USA

H Oh-oka
Osaka University, Osaka, Japan

I Ohtsuki
The Jikei University School of Medicine, Tokyo, Japan

R W Olsen
University of California, Los Angeles, CA, USA

E N Olson
University of Texas Southwestern Medical Center, Dallas, TX, USA

J M Olson
University of Massachusetts, Amherst, MA, USA

S K Olson
University of California, San Diego, La Jolla, CA, USA

Y Ono
Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan

N Osheroff
Vanderbilt University School of Medicine, Nashville, TN, USA

M Ottolia
David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

O E Owen
Southern Illinois University School of Medicine, Springfield, IL, USA

G Pacheco-Rodriguez
Cardiovascular and Pulmonary Branch of the National Heart, Lung, and Blood Institute in Bethesda, Bethesda, MD, USA

D J Pagliarini
University of California, San Diego, CA, USA

F Palmieri
University of Bari, Bari, Italy

V Pandini
Università degli Studi di Milano, Milan, Italy

J Pantel
Centre de Psychiatrie et Neuroscience, Paris, France

C A Parent
Center for Cancer Research, Bethesda, MD, USA

B Paria
National Institute of Dental and Craniofacial Research, Bethesda, MD, USA

V Parissi
UMR 5234 CNRS, Université Victor Segalen Bordeaux 2, Bordeaux, France

R Parker
Cornell University, Ithaca, NY, USA

A J Parodi
Fundación Instituto Leloir, Buenos Aires, Argentina

J M Pascal
Thomas Jefferson University, Philadelphia, PA, USA

B M Paschal
University of Virginia, Charlottesville, VA, USA

S S Patel
*UMDNJ-Robert Wood Johnson Medical School,
Piscataway, NJ, USA*

M J Paul
Rothamsted Research, Harpenden, UK

T Pederson
*University of Massachusetts Medical School, Worcester,
MA, USA*

K H Pedone
*University of North Carolina at Chapel Hill, Chapel Hill,
NC, USA*

J K M Penn
Harvard University, Cambridge, MA, USA

A Penna
University of California, Irvine, CA, USA

E Perez
Signum Biosciences, Monmouth Junction, NJ, USA

S Peri
Fox Chase Cancer Center, Philadelphia, PA, USA

J Pessin
Albert Einstein College of Medicine, Bronx, NY, USA

T V Pestova
Moscow State University, Moscow, Russia

P S Petersen
University of Copenhagen, Copenhagen, Denmark

C L Peterson
*University of Massachusetts Medical School, Worcester,
MA, USA*

M Petkovich
Queen's University, Kingston, ON, Canada

K K Pfister
University of Virginia, Charlottesville, VA, USA

K D Philipson
*David Geffen School of Medicine at UCLA, Los Angeles,
CA, USA*

J A Pietenpol
Vanderbilt University, Nashville, TN, USA

D Piomelli
University of California, Irvine, CA, USA

R K Plemper
*Emory University School of Medicine, Atlanta, GA, USA;
Children's Healthcare of Atlanta, Atlanta, GA, USA*

A J Plys
Cornell University, Ithaca, NY, USA

S W Polyak
The University of Adelaide, Adelaide, SA, Australia

J L Pomerantz
*Johns Hopkins University School of Medicine, Baltimore,
MD, USA*

K Potter
*Case Western Reserve University School of Medicine,
Cleveland, OH, USA*

L R Potter
University of Minnesota, Minneapolis, MN, USA

D R Poyner
Aston University, Birmingham, UK

D H Price
University of Iowa, Iowa City, IA, USA

D J Prockop
*Tulane University Health Sciences Center, New Orleans,
LA, USA*

E R Prossnitz
University of New Mexico, Albuquerque, NM, USA

C G Proud
University of Southampton, Southampton, UK

T Püsl
Klinikum Augsburg, Augsburg, Germany

X Qi
University of California, Irvine, Irvine, CA, USA

L A Quilliam
*Indiana University School of Medicine, Indianapolis, IN,
USA*

D M Raben
*Johns Hopkins University School of Medicine, Baltimore,
MD, USA*

R Ramachandran
University of Calgary, Calgary, AB, Canada

D Raman
*Vanderbilt University School of Medicine, Nashville, TN,
USA*

A Rao
*La Jolla Institute for Allergy and Immunology, La Jolla,
CA, USA; Children's Hospital Boston, Boston, MA, USA*

D N Rao
Indian Institute of Science, Bangalore, India

F Rappaport
Centre National de la Recherche Scientifique, Paris, France

N Rascio
University of Padua, Padua, Italy

V Rauhamäki
University of Helsinki, Helsinki, Finland

M Rechsteiner
University of Utah, Salt Lake City, UT, USA

K L Reddy
Johns Hopkins University School of Medicine, Baltimore, MD, USA

M D Resh
Memorial Sloan-Kettering Cancer Center, New York, NY, USA

H Reuter
University of Bern, Bern, Switzerland

P R Rich
University College London, London, UK

D J Richardson
University of East Anglia, Norwich, UK

A Richmond
VA Medical Center, Nashville, TN, USA; Vanderbilt University School of Medicine, Nashville, TN, USA

D Ricquier
Centre National de la Recherche Scientifique and Université Paris Descartes, Paris, France

A J Ridley
King's College, London, UK

D Riesner
Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

M P Rigobello
University of Padua, Padua, Italy

M Rivas
Centro Nacional de Biotecnología – CSIC, Madrid, Spain

R Rizzuto
University of Padua, Padua, Italy

P J Roach
Indiana University School of Medicine, Indianapolis, IN, USA

G A Roberts
University of Edinburgh, Edinburgh, UK

T M Roberts
Florida State University, Tallahassee, FL, USA

D N Robinson
Johns Hopkins University School of Medicine, Baltimore, MD, USA

F L Robinson
University of California, San Diego, CA, USA

A J Robison
Mt. Sinai School of Medicine, New York, NY, USA

J-D Rochaix
University of Geneva, Geneva, Switzerland

R S Rock
The University of Chicago, Chicago, IL, USA

N C Rockwell
University of California at Davis, Davis, CA, USA

S Rodriguez
Johns Hopkins University School of Medicine, Baltimore, MD, USA

D Ron
University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Cambridge, UK

R J Rose
Biomolecular Mass Spectrometry and Proteomics, Utrecht University, Utrecht, The Netherlands; Netherlands Proteomics Center, Utrecht, The Netherlands

J P Roth
Johns Hopkins University, Baltimore, MD, USA

G A Roybal
University of California San Francisco, San Francisco, CA, USA

R B Rucker
University of California, Davis, CA, USA

E I Rugarli
University of Cologne, Cologne, Germany

R R Russell III
Yale University School of Medicine, New Haven, CT, USA

K Rutherford
University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

J Rydström
University of Gothenburg, Gothenburg, Sweden

G Sachs
University of California, Los Angeles, CA, USA; VA Greater Los Angeles Healthcare System, Los Angeles, CA, USA

G H Sack Jr.
Johns Hopkins University, Baltimore, MD, USA

D B Sacks
National Institutes of Health, Bethesda, MD, USA

A Sahasranaman
Carnegie Mellon University, Pittsburgh, PA, USA

J Sai
Vanderbilt University School of Medicine, Nashville, TN, USA

K Sakabe
Johns Hopkins University, Baltimore, MD, USA

E Sakata
Max Planck Institute of Biochemistry, Martinsried, Germany

H K Salz
Case Western Reserve University, Cleveland, OH, USA

C E Samuel
University of California, Santa Barbara, CA, USA

M J Sanderson
University of Massachusetts Medical School, Worcester, MA, USA

S Sandmeyer
University of California, Irvine, Irvine, CA, USA

T J Santangelo
Ohio State University, Columbus, OH, USA

L Santella
Stazione Zoologica Anton Dohrn, Villa Comunale, Napoli, Italy

C Saunders
Vanderbilt University, Nashville, TN, USA

G Sawers
John Innes Centre, Norwich, UK

L A Sazanov
Medical Research Council Mitochondrial Biology Unit, Cambridge, UK

P Y Scaraffia
University of Arizona, Tucson, AZ, USA

G Schäfer
University of Lübeck, Lübeck, Germany

A H V Schapira
University College London, London, UK

G Schatz
Swiss Science and Technology Council, Bern, Switzerland

P Schedl
Princeton University, Princeton, NJ, USA

H Scheer
Universität München, Munich, Germany

P E Scherer
University of Texas Southwestern Medical Center, Dallas, TX, USA

U Schlattner
University Joseph Fourier and Inserm, Grenoble, France

T M Schmeing
McGill University, Montreal, QC, Canada

T Schmülling
Free University of Berlin, Berlin, Germany

A Schonbrunn
University of Texas Health Science Center – Houston, Houston, TX, USA

V L Schramm
Yeshiva University, Bronx, NY, USA

T A Schroer
Johns Hopkins University, Baltimore, MD, USA

H Schulman
Allosteros Therapeutics, Sunnyvale, CA, USA

C J Schultz
University of Adelaide, Glen Osmond, SA, Australia

S G Schultz
University of Texas Medical School, Houston, TX, USA

H Schulz
City University of New York, New York, NY, USA

M W Schwartz
University of Washington, Seattle, WA, USA

L Scorrano
University of Geneva, Geneva, Switzerland

D L Segaloff
The University of Iowa Carver College of Medicine, Iowa City, IA, USA

I H Segel
University of California, Davis, CA, USA

M Seibert
National Renewable Energy Laboratory, Golden, CO, USA

F Sekiya
National Institutes of Health, Bethesda, MD, USA

G L Semenza
Johns Hopkins University School of Medicine, Baltimore, MD, USA

D Sen
*UMDNJ-Robert Wood Johnson Medical School,
Piscataway, NJ, USA*

G C Sen
*Cleveland Clinic Foundation, Cleveland, OH, USA; Case
Western Reserve University, Cleveland, OH, USA*

I Serebriiskii
Fox Chase Cancer Center, Philadelphia, PA, USA

B Shane
University of California, Berkeley, CA, USA

N Sharon
Weizmann Institute of Science, Rehovot, Israel

J A Shayman
University of Michigan, Ann Arbor, MI, USA

S B Shears
NIEHS, Research Triangle Park, NC, USA

M Shepherd
University of Kent, Canterbury, UK

W Shi
University of California, La Jolla, CA, USA

S Shikano
University of Illinois, Chicago, IL, USA

K Shimabukuro
Florida State University, Tallahassee, FL, USA

J M Shin
*University of California, Los Angeles, CA, USA; VA
Greater Los Angeles Healthcare System, Los Angeles,
CA, USA*

D Shore
University of Geneva, Geneva, Switzerland

T J Shuttleworth
*University of Rochester Medical Center, Rochester, NY,
USA*

J N Siedow
Duke University, Durham, NC, USA

E Sigel
University of Bern, Bern, Switzerland

I Simpson
Penn State College of Medicine, Hershey, PA, USA

N Singh
*Massachusetts General Hospital Cancer Center, Boston,
MA, USA*

S Sirpiö
University of Turku, Turku, Finland

D G Skalnik
Indiana University, Indianapolis, IN, USA

L A Sklar
University of New Mexico, Albuquerque, NM, USA

C W Slayman
Yale University, New Haven, CT, USA

P P Slonimski
University Pierre et Marie Curie, Paris, France

J V Small
*IMBA-Institute of Molecular Biotechnology of the
Austrian Academy of Sciences, Vienna, Austria*

K P Smith
*University of Massachusetts Medical School, Worcester,
MA, USA*

W L Smith
*University of Michigan Medical School, Ann Arbor, MI,
USA*

M C M Smith
University of Aberdeen, Aberdeen, UK

T E Smithgall
*University of Pittsburgh School of Medicine, Pittsburgh,
PA, USA*

T Sobolik-Delmaire
*Vanderbilt University School of Medicine, Nashville, TN,
USA*

H Song
*Johns Hopkins University School of Medicine, Baltimore,
MD, USA*

E D Sonnenburg
*Washington University School of Medicine, St. Louis,
MO, USA*

K L Soole
Flinders University, Adelaide, SA, Australia

M C Sorgato
University of Padua, Padua, Italy

H Sorimachi
*Tokyo Metropolitan Institute of Medical Science, Tokyo,
Japan*

R Sousa
*University of Texas Health Science Center, San Antonio,
TX, USA*

J Sowa
Baylor College of Medicine, Houston, TX, USA

M M Spiering
*The Pennsylvania State University, University Park, PA,
USA*

K Srinivasan
Vanderbilt University, Nashville, TN, USA

V Srivastava
Johns Hopkins University School of Medicine, Baltimore, MD, USA

M Stammes
University of Iowa, Iowa City, IA, USA

C J Stefan
Cornell University, Ithaca, NY, USA

F M Steinberg
University of California, Davis, CA, USA

J B Stock
Princeton University, Princeton, NJ, USA

D J Stuehr
The Lerner Research Institute, Cleveland, OH, USA

S Subramani
University of California San Diego, La Jolla, CA, USA

M C Sugden
Queen Mary University of London, London, UK

B A Sullivan
Duke University, Durham, NC, USA

K Sun
University of Texas Southwestern Medical Center, Dallas, TX, USA

J W Suttie
University of Wisconsin–Madison, Wisconsin, WI, USA

T Suzuki
RIKEN Advanced Science Institute, Wako, Japan

A Svitin
University of Alabama, Birmingham, AL, USA

T Svitkina
University of Pennsylvania, Philadelphia, PA, USA

L Swint-Kruse
University of Kansas Medical Center, Kansas City, KS, USA

I Szabò
University of Padua, Padua, Italy

A L Szymczak-Workman
St. Jude Children's Research Hospital, Memphis, TN, USA

H Taegtmeier
The University of Texas Medical School at Houston, Houston, TX, USA

J A Tainer
The Scripps Research Institute, La Jolla, CA, USA

R Takemura
National College of Nursing, Tokyo, Japan

O Takeuchi
World Premier International Immunology Frontier Research Center, Osaka, Japan; Osaka University, Osaka, Japan

B Tandler
Veterans Affairs Medical Center, Cleveland, OH, USA

R Taussig
University of Texas Southwestern Medical Center, Dallas, TX, USA

D J Taylor
Cleveland State University, Cleveland, OH, USA; Case Western Reserve University, Cleveland, OH, USA

H Teramoto
National Institute of Dental and Craniofacial Research, Bethesda, MD, USA

J Tfelt-Hansen
Copenhagen University Hospital, Copenhagen, Denmark

S Thammannagowda
National Renewable Energy Laboratory, Golden, CO, USA

G Thibault
Temasek Life Sciences Laboratory, National University of Singapore, Singapore

K A Thomas
Drew University, Madison, NJ, USA

J W Thorner
University of California, Berkeley, CA, USA

J M Thornton
European Bioinformatics Institute, Cambridge, UK

G J Tigyi
University of Tennessee Health Sciences Center Memphis, Memphis, TN, USA

J Titus
University of Massachusetts, Amherst, MA, USA

M A Titus
University of Minnesota, Minneapolis, MN, USA

V Todorović
Northwestern University Feinberg School of Medicine, Chicago, IL, USA

M Tokarska-Schlattner
University Joseph Fourier and Inserm, Grenoble, France

A Toker
Harvard Medical School, Boston, MA, USA

- T J Tolbert
Indiana University, Bloomington, IN, USA
- A E Tomkinson
University of Maryland School of Medicine, Baltimore, MD, USA
- M D Toney
University of California, Davis, CA, USA
- M K Topham
University of Utah, Salt Lake City, UT, USA
- S Tornaletti
University of Florida, Gainesville, FL, USA
- C Toyoshima
The University of Tokyo, Tokyo, Japan
- B L Trumpower
Dartmouth Medical School, Hanover, NH, USA
- M-D Tsai
The Ohio State University, Columbus, OH, USA
- S Tsukita
Osaka University, Osaka, Japan
- J L Tubbs
The Scripps Research Institute, La Jolla, CA, USA
- H Tuppy
University of Vienna, Vienna, Austria
- B K Tye
Cornell University, Ithaca, NY, USA
- A Tzagoloff
Columbia University, New York, NY, USA
- G Unden
Johannes-Gutenberg-Universität Mainz, Mainz, Germany
- F Ursini
University of Padua, Padua, Italy
- J R Ussher
University of Alberta, Edmonton, AB, Canada
- K Uyeda
University of Texas Southwestern Medical Center & Veteran Affairs Medical Center, Dallas, TX, USA
- J van der Oost
Wageningen University, Wageningen, The Netherlands
- G D Van Duyne
University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA
- B Van Houten
University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; Hillman Cancer Center, Pittsburgh, PA, USA
- E Van Schaftingen
Université Catholique de Louvain, Brussels, Belgium
- R J M Van Spanning
Free University of Amsterdam, Amsterdam, The Netherlands
- S J Vannucci
Weill Cornell Medical College of Cornell University, New York, NY, USA
- A Varki
University of California, San Diego, CA, USA
- K M Vasquez
The University of Texas, Smithville, TX, USA
- M Vaughan
Cardiovascular and Pulmonary Branch of the National Heart, Lung, and Blood Institute in Bethesda, Bethesda, MD, USA
- D L Vaux
La Trobe University, Melbourne, VIC, Australia
- R L Veech
National Institutes of Health, Rockville, MD, USA
- D A A Vignali
St. Jude Children's Research Hospital, Memphis, TN, USA
- P Villalonga
Universitat de les Illes Balears, Palma, Spain
- J V Virbasius
University of Massachusetts Medical School, Worcester, MA, USA
- M W Vitalini
University of Iowa, Iowa City, IA, USA
- D R Voelker
National Jewish Health, Denver, CO, USA
- P Wadsworth
University of Massachusetts, Amherst, MA, USA
- C J Waechter
University of Kentucky College of Medicine, Lexington, KY, USA
- C Waldsich
MFPL, University of Vienna, Vienna, Austria
- J E Walker
The Medical Research Council Mitochondrial Biology Unit, Cambridge, UK
- D C Wallace
University of Pennsylvania, Philadelphia, PA, USA
- T Wallimann
Institute of Cell Biology, ETH Zürich, Switzerland

L L Wallrath

University of Iowa, Iowa City, IA, USA

J Y J Wang

University of California, San Diego, La Jolla, CA, USA

Z-Q Wang

Kent State University Tuscarawas, New Philadelphia, OH, USA

R D Ward

Baylor College of Medicine, Houston, TX, USA

C F Ware

Sanford-Burnham Medical Research Institute, La Jolla, CA, USA

B A Warner

The Pennsylvania State University, University Park, PA, USA

P A Watkins

Johns Hopkins University School of Medicine, Baltimore, MD, USA

Z Wei

Johns Hopkins University School of Medicine, Baltimore, MD, USA

N L Weigel

Baylor College of Medicine, Houston, TX, USA

L S Weinstein

National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA

M D Weitzman

Children's Hospital of Philadelphia, Philadelphia, PA, USA

R D Wells

The Texas Medical Center, Houston, TX, USA

E R Westra

Wageningen University, Wageningen, The Netherlands

E Weterings

Nevada Cancer Institute, Las Vegas, NV, USA

R Wetzel

University of Tennessee, Knoxville, TN, USA

R J White

Beatson Institute for Cancer Research, Glasgow, UK

D G Whitehouse

University of Sussex, Brighton, UK

J Whitelegge

University of California, Los Angeles, CA, USA

G Wiche

University of Vienna, Vienna, Austria

R B Wickner

National Institutes of Health, Bethesda, MD, USA

M Wikström

University of Helsinki, Helsinki, Finland

R J P Williams

University of Oxford, Oxford, UK

S H Wilson

National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA

P Wins

University of Liège, Liège, Belgium

J A Wise

Case Western Reserve University School of Medicine, Cleveland, OH, USA

M J Wolfgang

Johns Hopkins University School of Medicine, Baltimore, MD, USA

C-H Wong

The Scripps Research Institute, La Jolla, CA, USA

G W Wong

Johns Hopkins University School of Medicine, Baltimore, MD, USA

C L Woodcock

University of Massachusetts, Amherst, MA, USA

J R Woodgett

Samuel Lunenfeld Research Institute, Toronto, ON, Canada

J L Woolford Jr.

Carnegie Mellon University, Pittsburgh, PA, USA

C A Worby

University of California, San Diego, CA, USA

L Wordeman

University of Washington, Seattle, WA, USA

G Wright

Vanderbilt University, Nashville, TN, USA

W D Wright

University of California, Davis, Davis, CA, USA

E Yamashita

Osaka University, Yamada-oka Suita, Japan

H Yamashita

Okayama Prefectural University, Okayama, Japan

W Yang

Medical University of South Carolina, Charleston, SC, USA

T M Yankee
University of Kansas Medical Center, Kansas City, KS, USA

N Y Yao
*Rockefeller University, Howard Hughes Medical Institute,
New York, NY, USA*

K-W Yau
*Johns Hopkins University School of Medicine, Baltimore,
MD, USA*

J Ye
*University of Texas Southwestern Medical Center, Dallas,
TX, USA*

E T H Yeh
*The University of Texas MD Anderson Cancer Center and
Texas Heart Institute at St. Luke's Episcopal Hospital,
Houston, TX, USA*

T J Yen
Fox Chase Cancer Center, Philadelphia, PA, USA

A V Yeromin
University of California, Irvine, CA, USA

M Yudkoff
*University of Pennsylvania School of Medicine,
Philadelphia, PA, USA*

C Zaelzer
Montreal General Hospital, Montreal, QC, Canada

E Zameitat
Philipps-University, Marburg, Germany

E Zamir
*Max Planck Institute of Molecular Physiology, Dortmund,
Germany*

G Zanetti
Università degli Studi di Milano, Milan, Italy

D Zannoni
University of Bologna, Bologna, Italy

B R Zelickson
*University of Alabama at Birmingham, Birmingham,
AL, USA*

D Žgur-Bertok
University of Ljubljana, Ljubljana, Slovenia

J Zhang
*University of Alabama at Birmingham, Birmingham,
AL, USA*

L Zhao
The Ohio State University, Columbus, OH, USA

D O Zharkov
*SB RAS Institute of Chemical Biology and Fundamental
Medicine, Novosibirsk, Russia*

Y Zhou
*La Jolla Institute for Allergy and Immunology, La Jolla,
CA, USA*

N Zhukovskaya
Thomas Jefferson University, Philadelphia, PA, USA

V Zickermann
Goethe University, Frankfurt am Main, Germany

M Zoratti
University of Padua, Padua, Italy

HOW TO USE THE ENCYCLOPEDIA

The *Encyclopedia of Biological Chemistry* is a comprehensive and authoritative study covering the field of biochemistry. It includes over 500 articles on various aspects of this subject, contained in four volumes. Each article provides a focused description of the given topic, intended to inform a broad range of readers, ranging from students, to research professionals, and interested others. Each article serves as a comprehensive overview of a given area, providing both breadth of coverage for students, and depth of coverage for research professionals.

All articles in the encyclopedia are arranged alphabetically by subject in the table of contents. The index is located in volume 4. Article titles generally begin with the key word or phrase indicating the topic.

Each article contains a glossary, cross-references to other related encyclopedia articles, and further readings and relevant websites for additional information. The glossary contains terms that may be unfamiliar to the reader, with each term defined in the context of its use in that article. Thus, a term may

appear in the glossary for another article defined in a slightly different manner or with a subtle nuance specific to that article. For clarity, we have allowed these differences in definition to remain so that each article is fully understandable on its own.

Each article has been cross-referenced to other related articles in the encyclopedia at the close of each article. We encourage readers to use the cross-references to locate other encyclopedia articles that will provide more detailed information about a subject.

The further readings include recent secondary sources to aid the reader in locating more detailed or technical information. Review articles and research articles that are considered of primary importance to the understanding of a given subject area are also listed. These suggested readings are not intended to provide a full reference listing of all material covered in the context of a given article, but are provided as next steps for a reader looking for additional information.

PREFACE

Biological Chemistry encompasses the “*chemistry of life*”, more specifically the chemistry of the compounds and processes that constitute living organisms. The ultimate goal is to understand and define biology at a mechanistic level. This was aptly stated in an historical treatise on the founding of the *Journal of Biological Chemistry*, where John Edsall quoted a statement in a letter from J. L. Loeb (in Berkeley), “The future of biology lies with those who attack its problems from a chemical point of view.” What was an emerging field in 1900 with its origins in physiology, nutrition, and chemistry has now broadened to include numerous other fields including mechanistic enzymology, molecular biology, structural biology, cell biology, genomics, proteomics, metabolomics, bioinformatics, and others, that were not defined as discrete fields in that era.

Modern biochemistry (biological chemistry) really began with the accidental discovery, by Eduard Buchner in 1897, that a cell-free yeast extract could carry out fermentation of glucose to alcohol and CO₂ *in the absence of intact cells*. He named the dissolved substance responsible for this process zymase, the substance(s) we now refer to as enzymes. Importantly, Buchner recognized the significance of his discovery. This ended the dogma of the time, perpetuated by Louis Pasteur, the concept of *vitalism* that fermentation (and presumably other complex biological phenomena) required the action of intact cells. Thus, serendipity and a prepared mind ushered in a new period of discovery. It became possible to dissect complex physiological processes and to study them with preparations free of the constraints of intact cells. Once a metabolic pathway/process was established, it became possible to purify the enzymes, cofactors, and substrates involved, to reconstitute the process with purified components and to characterize the components chemically. What followed was an information explosion in the field of biochemistry and progression through a series of trends, each “in vogue” in its time. The identification of the dietary essentials, the hunt for the vitamins/cofactors, the hormones, identification of metabolic pathways and the enzymes involved, oxidative phosphorylation, protein synthesis, molecular biology — each developed as a primary focus.

The need to associate chemistry with function came early and was evident in the naming of Departments and Journals. Over time names changed from Agricultural Chemistry to Physiological Chemistry to Biochemistry to Biological Chemistry. An example is the Department of Biochemistry at the University of Wisconsin, which began in 1883 as the Department of Agricultural Chemistry.

Where are we headed? We have reached the point where the borders of these areas have become blurred. What constitutes cell biology, molecular biology, genetics, developmental biology, physiology, immunology — ultimately reduces to chemistry. To understand these processes we must know what the molecules are and understand how they interact, i.e. the basic chemistry. That is what the *Encyclopedia of Biological Chemistry* is about.

The breadth of content of this encyclopedia aims to cover major topics of modern biochemistry, each authored by an expert in the area. We feel that the coverage is broad and we have been inclusive in choice of topics. The encyclopedia is a reference work containing over 500 articles with nearly 900 authors and coauthors. Each article/topic covers an important area of the field that reflects the point of view of the authors. Together the articles cover virtually every aspect of biology for which we have “mechanistic” information. For those who wish to probe more deeply into a topic, references to review articles are included at the end of each article. The editorial board that made decisions on coverage consists of eight members, each an expert representing a major area in the field of biochemistry. An effort was to make coverage as complete as possible. The content is pitched at a level we hope interpretable to interested individuals with some background in chemistry and biology. It is intended for such individuals rather than specialists with extensive scientific backgrounds in specific areas. It is aimed at the generalist as opposed to the specialist.

CONTENTS

<i>Editors-in-Chief</i>	<i>v</i>
<i>Editorial Board</i>	<i>vii</i>
<i>Contributors</i>	<i>ix</i>
<i>How to Use the Encyclopedia</i>	<i>xxxiii</i>
<i>Preface</i>	<i>xxxv</i>

VOLUME 1

AAA-ATPases <i>J Martin and A N Lupas</i>	1
ABC Transporters <i>A Goffeau and B De Hertogh</i>	7
Actin Assembly/Disassembly <i>H N Higgs</i>	12
Actin-Capping and -Severing Proteins <i>S Maiti and J R Bamburg</i>	18
Actin Organization <i>T Svitkina</i>	27
Actin-Related Proteins <i>R D Mullins</i>	36
Adenylyl Cyclases <i>R Taussig</i>	42
Adipocyte Fat Mobilization: The Regulatory Roles of Perilipin 1, Adipose Triglyceride Lipase, and Hormone-Sensitive Lipase <i>K T Dalen, C Londos, and A R Kimmel</i>	47
Adipogenesis <i>M D Lane</i>	52
Adrenergic Receptors <i>D B Bylund</i>	57
Affinity Tags for Protein Purification <i>J J Falke and J A Corbin</i>	61
Algal Hydrogen Production <i>U M N Murthy and M L Ghirardi</i>	66

Allosteric Regulation <i>B S Cooperman</i>	71
Alternative Splicing <i>K Lynch and L Motta-Mena</i>	75
Alternative Splicing: Regulation of <i>Drosophila melanogaster</i> Sex Determination <i>J K M Penn, P Graham, P Schedl, and H K Salz</i>	81
Amine Oxidases <i>G Floris and A Finazzi Agrò</i>	87
Amino Acid Metabolism <i>L Cynober</i>	91
Aminopeptidases <i>R A Bradshaw</i>	97
Amyloid <i>R Wetzel</i>	100
Anaplerosis <i>R R Russell III and H Taegtmeyer</i>	105
Angiotensin Receptors <i>T Inagami</i>	111
Anorexigenic and Orexigenic Gut Peptides <i>P S Petersen and T H Moran</i>	115
Aquaporin Water Channels <i>K Liu and P Agre</i>	120
Arachidonic Acid Regulated Calcium Channel <i>T J Shuttleworth</i>	127
Arf Family <i>G Pacheco-Rodriguez, J Moss, and M Vaughan</i>	134
Aspartic Proteases <i>B M Dunn</i>	137
ATP in Plant Mitochondria: Substrates, Inhibitors, and Uncouplers <i>K L Soole and R I Menz</i>	141
ATP Synthesis: Mitochondrial Cyanide-Resistant Terminal Oxidases <i>J N Siedow</i>	145
Autophagy in Fungi and Mammals <i>J Geng, D J Klionsky, and J-L Guan</i>	149
B ₁₂ -Containing Enzymes <i>V Bandarian and R G Matthews</i>	155
B-Cell Antigen Receptor <i>T M Yankee</i>	161
Bcl2 Family: Cell Death Enhancers and Inhibitors <i>D L Vaux</i>	165
Bile Salts <i>L Maillette de Buy Wenniger, T Pusch, and U Beuers</i>	167
Biochemistry: Niacin/NAD(P) <i>K L Bogan and C Brenner</i>	172

Biochemistry of Development: Striated Muscle <i>R Bassel-Duby and E N Olson</i>	179
Biochemistry of Hematopoiesis <i>A D Friedman</i>	187
Biochemistry of Liver Regeneration <i>B Ochoa and A M Diehl</i>	190
Biochemistry of Neurogenesis <i>K Christian, H Song, and G-L Ming</i>	196
Biochemistry of Thiamine and Thiamine Phosphate Compounds <i>L Bettendorff and P Wins</i>	202
Bioenergetics and the Mitochondrial Genome <i>D C Wallace</i>	210
Bioenergetics: General Definition of Principles <i>D G Nicholls</i>	217
Biotin <i>S W Polyak and A Chapman-Smith</i>	221
Biotinylation of Proteins <i>R A Kohanski</i>	226
Bmp Signaling and Vascular Disease <i>J Lowery and M de Caestecker</i>	229
Bradykinin Receptors <i>R M Burch</i>	240
Branched-Chain Amino Acids <i>D T Chuang</i>	244
Brassinosteroids <i>S D Clouse</i>	250
Cadherin-Mediated Cell–Cell Adhesion <i>M Lowndes and W J Nelson</i>	255
Cadherin Signaling <i>D B Sacks</i>	261
Calcitonin Gene-Related Peptide and Adrenomedullin Receptors <i>D L Hay, A C Conner, and D R Poyner</i>	265
Calcitonin Receptor <i>S I Girgis, N Moradi-Bidhendi, and L Mancini</i>	270
Calcium-Binding Proteins: Cytosolic (Annexins, Gelsolins, and C ₂ -Domain Proteins) <i>J Krebs</i>	275
Calcium, Biological Fitness of <i>R J P Williams</i>	281
Calcium-Buffering Proteins: Calbindin <i>L J Mady, F Haleem, and S Christakos</i>	284
Calcium-Buffering Proteins: ER Luminal Proteins <i>J Groenendyk and M Michalak</i>	290
Calcium/Calmodulin-Dependent Protein Kinase II <i>A Hudmon and H Schulman</i>	297

Calcium/Calmodulin-Dependent Protein Kinases <i>A J Robison and R J Colbran</i>	304
Calcium in the Regulation of Gene Expression <i>M Rivas and J R Naranjo</i>	310
Calcium-Modulated Proteins (EF-Hand) <i>R H Kretsinger</i>	316
Calcium Oscillations <i>M Brini</i>	322
Calcium-Sensing Receptor (CaSR) <i>J Tfelt-Hansen and E M Brown</i>	326
Calcium Signaling by Cyclic ADP-Ribose and NAADP <i>L Santella and J T Chun</i>	331
Calcium Signaling: Calmodulin-Dependent Phosphatase <i>C B Klee and S Li</i>	337
Calcium Signaling: NO Synthase <i>Z-Q Wang and D J Stuehr</i>	342
Calcium Transport in Mitochondria <i>R Rizzuto and M Brini</i>	347
Calpain <i>H Sorimachi, S Hata, and Y Ono</i>	353
Carbohydrate Chains: Enzymatic and Chemical Synthesis <i>T J Tolbert and C-H Wong</i>	362
Carbohydrate Metabolism in the Central Nervous System <i>I Simpson and S J Vannucci</i>	368
Carbohydrate-Responsive Element Binding Protein <i>K Uyeda and H Yamashita</i>	374
Cardiac Excitation–Contraction Coupling <i>D M Bers and S Despa</i>	379
Carnitine and β -Oxidation <i>J Kerner and C L Hoppel</i>	384
Caspases and Cell Death <i>D W Nicholson and G Melino</i>	388
Catalysis by RNA, Structural Themes: Group I Introns <i>B L Golden</i>	397
Cell Cycle: Control of Entry and Progression through S Phase <i>S L Forsburg</i>	401
Cell-Cycle Controls in G ₁ and G ₀ <i>A Narasimha, W Shi, and S F Dowdy</i>	406
Cell Cycle: DNA Damage Checkpoints <i>M D Weitzman and J Y J Wang</i>	410
Cell Cycle: Mitotic Checkpoint <i>T J Yen</i>	417
Cell Death by Apoptosis and Necrosis <i>V Ermer and P Nicotera</i>	423

Cell–Matrix Interactions <i>J A Askari and M J Humphries</i>	426
Cell Migration <i>J V Small</i>	430
Cell Migration within Three-Dimensional Matrices <i>P J Keely</i>	436
Centromeres <i>B A Sullivan</i>	446
c- <i>fes</i> Proto-Oncogene <i>T E Smithgall</i>	451
Chaperonins <i>A Bracher and F U Hartl</i>	456
Chemiluminescence and Bioluminescence <i>T O Baldwin</i>	461
Chemiosmotic Theory <i>P R Rich</i>	467
Chemistry of Alzheimer Disease <i>G H Sack Jr.</i>	473
Chemokine Receptors <i>T Sobolik-Delmaire, D Raman, J Sai, G-H Fan, and A Richmond</i>	480
Chemolithotrophy <i>A B Hooper and A A DiSpirito</i>	486
Chemotactic Peptide/Complement Receptors <i>E R Prossnitz and L A Sklar</i>	493
Chlorophylls and Carotenoids <i>H Scheer</i>	498
Chloroplasts <i>N Rascio</i>	506
Chloroplast Signaling and Redox Control <i>J-D Rochaix</i>	511
Cholesterol Synthesis and Regulation <i>P J Espenshade</i>	516
Chromatin: Methyl-CpG-Binding Proteins <i>D G Skalnik</i>	521
Chromatin: Nucleosome Positioning – the GAL Promoter <i>D Lohr</i>	525
Chromatin: Physical Organization <i>C L Woodcock</i>	530
Chromosome Organization and Structure, Overview <i>E Gracheva and S C R Elgin</i>	535
Coenzyme A <i>M D Lane</i>	539
Collagenases <i>K Holmbeck and H Birkedal-Hansen</i>	542

Collagens <i>D J Prockop</i>	545
Color Vision <i>G H Jacobs</i>	550
Conservative Site-Specific Recombination <i>M C M Smith</i>	555
Cyclic AMP Receptors of <i>Dictyostelium</i> <i>D Hereld, Y Kamimura, and P N Devreotes</i>	562
Cyclic GMP Phosphodiesterases <i>S H Francis and J D Corbin</i>	567
Cyclic Nucleotide-Dependent Protein Kinases <i>S H Francis and J D Corbin</i>	574
Cyclic Nucleotide Phosphodiesterases <i>V C Manganiello and E Degerman</i>	580
Cyclic Nucleotide-Regulated Cation Channels <i>M Biel</i>	586
Cysteine Proteases <i>D J Buttle and J S Mort</i>	589
Cytochrome <i>bc</i> ₁ Complex (Respiratory Chain Complex III) <i>B L Trumpower</i>	593
Cytochrome <i>c</i> <i>H Tuppy and G Kreil</i>	599
Cytochrome Oxidases, Bacterial <i>P Brzezinski and P Ädelroth</i>	602
Cytochrome P-450 <i>R Bernhardt</i>	607
Cytokines <i>S J Levine</i>	613
Cytokinesis <i>V Srivastava and D N Robinson</i>	622
Cytokinin <i>T Schmölling</i>	627
Cytoskeletal Motors: General Principles <i>R S Rock</i>	632
Desmosomes and Hemidesmosomes <i>V Todorović, K R Kligys, R L Dusek, J C R Jones, and K J Green</i>	636
Detergent Properties <i>D R McCaslin</i>	644
Diabetes <i>S M Blackman and D W Cooke</i>	649
Diacylglycerol Kinases and Phosphatidic Acid Phosphatases <i>M K Topham</i>	659
Disulfide Bond Formation <i>H F Gilbert</i>	664

VOLUME 2

DNA Base Excision Repair Pathways <i>R A O Bennett and B Dimple</i>	1
DNA Damage: Alkylation <i>J L Tubbs and J A Tainer</i>	9
DNA Glycosylases: Mechanisms <i>A C Drohat and A Maiti</i>	16
DNA Helicases: Hexameric Enzyme Action <i>D Sen and S S Patel</i>	21
DNA Ligases: Mechanism and Functions <i>A E Tomkinson and J A Della-Maria</i>	28
DNA Ligases: Structures <i>J M Pascal</i>	33
DNA Methyltransferases: Eubacterial GATC <i>M G Marinus</i>	40
DNA Mismatch Repair and Homologous Recombination <i>I Matic</i>	45
DNA Mismatch Repair and the DNA Damage Response <i>G-M Li</i>	51
DNA Mismatch Repair in Bacteria <i>A-L Lu</i>	54
DNA Mismatch Repair in Disease and Aging <i>P Hsieh</i>	60
DNA Mismatch Repair in Mammals <i>A J Phys and E Alani</i>	66
DNA Modification, Restriction Endonucleases: Type I Enzymes <i>D T F Dryden and G A Roberts</i>	72
DNA Oxidation <i>D O Zharkov</i>	77
DNA Polymerase β , Eukaryotic <i>W A Beard and S H Wilson</i>	82
DNA Polymerase I, Bacterial <i>C M Joyce</i>	87
DNA Polymerase III, Bacterial <i>H Maki and A Furukohri</i>	92
DNA Polymerases: Kinetics and Mechanism <i>K A Johnson</i>	96
DNA Polymerases: Reverse Transcriptase Integrase, and Retrovirus Replication <i>M-L Andréola, V Parissi, and S Litvak</i>	101
DNA Replication: Eukaryotic Origins and the Origin Recognition Complex <i>I Chesnokov and A Svitin</i>	108
DNA Replication Fork, Bacterial <i>M M Spiering and S J Benkovic</i>	114

DNA Replication Fork, Eukaryotic <i>L M Kelman and Z Kelman</i>	118
DNA Replication: Initiation in Bacteria <i>J M Kaguni</i>	121
DNA Replication, Mitochondrial <i>D A Clayton</i>	126
DNA Restriction and Modification: Type II Enzymes <i>S E Halford</i>	130
DNA Restriction and Modification: Type III Enzymes <i>D N Rao and S Bheemanaik</i>	136
DNA Secondary Structure <i>A Bacolla, K M Vasquez, and R D Wells</i>	142
DNA Sequence Recognition by Proteins <i>K Rutherford and G D Van Duyne</i>	149
DNA Supercoiling <i>T-S Hsieh</i>	154
DNA Topoisomerases: Type I <i>J J Champoux</i>	157
DNA Topoisomerases: Type II <i>A C Gentry and N Osheroff</i>	163
Dopamine Receptors <i>K A Neve</i>	169
Dynactin <i>T A Schroer</i>	174
Dynamic Behavior of Photosystem II Light Harvesting <i>P Horton</i>	178
Dynein <i>K K Pfister</i>	184
Eicosanoid Receptors <i>R M Breyer and M D Breyer</i>	188
Elastin <i>J A Foster</i>	192
Endocannabinoids <i>D Piomelli</i>	194
Endocytosis <i>J G Donaldson</i>	197
Endoplasmic Reticulum-Associated Protein Degradation <i>R Bernasconi and M Molinari</i>	200
Energy Transduction in Anaerobic Bacteria <i>G Unden</i>	204
Enzyme Inhibitors <i>V L Schramm</i>	210
Enzyme Kinetics <i>I H Segel</i>	216

Enzyme Reaction Mechanisms: Stereochemistry <i>M-D Tsai, L Zhao, and B J Lamarche</i>	221
ER/SR Calcium Pump: Function <i>G Inesi</i>	226
ER/SR Calcium Pump: Structure <i>C Toyoshima</i>	231
Eukaryotic Chemotaxis <i>C P McCann and C A Parent</i>	236
Eukaryotic DNA Polymerase α <i>C Lee and B K Tye</i>	241
Eukaryotic DNA Polymerase δ <i>P M J Burgers</i>	245
Eukaryotic Protein Biosynthesis: The Elongation Cycle <i>A A Komar, D J Taylor, and W C Merrick</i>	249
Evolving Concepts of Leptin <i>H Münzberg and T W Gettys</i>	256
Exosomes and Microvesicles <i>S J Gould</i>	262
FAK Family <i>S K Hanks</i>	265
F-ATPases <i>J E Walker</i>	269
Fatty Acid Metabolism and Cancer <i>F P Kuhajda</i>	275
Fatty Acid Oxidation <i>H Schulz</i>	281
Fatty Acid Structure and Synthesis <i>S D Clarke and M T Nakamura</i>	285
Fatty Acyl-CoA Synthetases <i>P A Watkins</i>	290
Ferredoxin <i>G Zanetti and V Pandini</i>	296
Ferredoxin-NADP ⁺ Reductase <i>G Zanetti and A Aliverti</i>	299
Fibroblast Growth Factor Receptors and Cancer-Associated Perturbations <i>M Kornmann and M Korc</i>	303
Flavins <i>B Entsch and D P Ballou</i>	309
Flippases <i>C J Waechter</i>	314
Focal Adhesions and Related Integrin Contacts <i>E Zamir and B Geiger</i>	318
Folate and Vitamin B ₁₂ Function <i>B Shane</i>	324

Free Oligosaccharides: Formation and Degradation (Free Oligosaccharides Related to N-Linked Glycans) <i>T Suzuki</i>	329
Friedreich's Ataxia <i>P E Hart and A H V Schapira</i>	332
F _X , F _A , and F _B Iron-Sulfur Clusters in Type I Photosynthetic Reaction Centers <i>B Jagannathan and J H Golbeck</i>	335
G ₁₂ /G ₁₃ Family <i>S Offermanns</i>	343
GABA _A Receptor <i>R W Olsen and G-D Li</i>	347
GABA _B Receptor <i>S J Enna and N G Bowerly</i>	351
Gastrointestinal Digestion and Absorption <i>J Keller</i>	354
Gene Expression in Bacterial Systems: The <i>trp</i> Operon and Attenuation <i>P Gollnick</i>	360
Gene Expression in Eukaryotes: RNA Polymerase II Structure <i>P Cramer</i>	365
Genome-Wide Analysis of Gene Expression <i>D-W Doug Chung and K G Le Roch</i>	369
Giant Mitochondria (Megamitochondria) <i>B Tandler and C L Hoppel</i>	375
Gi Family of Heterotrimeric G Proteins <i>M E Linder</i>	377
Gluconeogenesis <i>R W Hanson and O E Owen</i>	381
Glucose/Sugar Transport in Bacteria <i>L Guan and H R Kaback</i>	387
Glucose/Sugar Transport in Mammals <i>S Mora and J Pessin</i>	391
Glutamate Receptors, Metabotropic <i>K J Gregory and P J Conn</i>	395
Glutathione Peroxidases <i>F Ursini and M Maiorino</i>	399
Glycation <i>M Lima and J W Baynes</i>	405
Glycerolipid Structure, Function, and Synthesis in Eukaryotes <i>D R Voelker</i>	412
Glycine Receptors <i>T Lynagh and B Laube</i>	419
Glycogen Metabolism <i>P J Roach</i>	425
Glycogen Storage Diseases <i>G H Sack Jr.</i>	428

Glycogen Synthase Kinase-3 <i>J R Woodgett</i>	434
Glycolipid-Dependent Adhesion Processes <i>S Hakomori</i>	439
Glycolysis Overview <i>R A Harris</i>	443
Glycoprotein Folding and Processing Reactions <i>J J Caramelo and A J Parodi</i>	448
Glycoprotein-Mediated Cell Interactions, Nonmucin-Type O-Linked <i>R S Haltiwanger</i>	453
Glycoproteins, N-Linked <i>M A Lehrman</i>	457
Glycoproteins, Plant <i>C J Schultz</i>	465
Glycosylation, Congenital Disorders of <i>B G Ng and H H Freeze</i>	469
Glycosylphosphatidylinositol Anchors <i>A K Menon</i>	476
Golgi Complex <i>M Stamnes</i>	479
G-Protein-Coupled Receptor Kinases and Arrestins <i>N Zhukovskaya, B Barker, D S Kang, and J L Benovic</i>	483
G Protein Signaling Regulators <i>J H Exton</i>	492
G _q Family <i>W Yang and J D Hildebrandt</i>	496
Green Bacteria: Chlorophyll Biosynthesis, Light-Harvesting, Reaction Centers, and Electron Transport <i>D A Bryant</i>	501
Green Bacteria: Secondary Electron Donor (Cytochromes) <i>H Oh-oka and R E Blankenship</i>	510
Green Bacteria: The Light-Harvesting Chlorosome <i>J M Olson</i>	513
Gs Family of Heterotrimeric G Proteins <i>L S Weinstein</i>	517
Heat/Stress Responses <i>G Thibault and D T W Ng</i>	522
Hematopoietin Receptors <i>B A Miller</i>	526
Heme Proteins <i>J Everse</i>	532
Heme Synthesis <i>G C Ferreira</i>	539
Hexokinase/Glucokinase <i>E Van Schaftingen</i>	543

HIV-1 Reverse Transcriptase Structures <i>K Das, E Arnold, and S H Hughes</i>	548
HIV Protease <i>B M Dunn</i>	554
Homologous Recombination in Meiosis <i>N M Hollingsworth</i>	557
Hydrogenases, Structure and Function <i>W Lubitz and H Ogata</i>	562
Immunoglobulin (Fc) Receptors <i>M Hogarth</i>	568
Indicators of Intracellular Calcium <i>M Giacomello and M Campeol</i>	574
Inositol Phosphate Kinases and Phosphatases <i>S B Shears</i>	580
Inositol Trisphosphate and Calcium Signaling <i>M J Berridge</i>	583
Insect Biochemistry/Hormones <i>P Y Scaraffia and R L Miesfeld</i>	590
Insulin- and Glucagon-Secreting Cells of the Pancreas <i>F M Matschinsky</i>	596
Insulin Mechanisms/Metabolic Actions <i>E J Gallagher and D LeRoith</i>	602
Integrin Signaling <i>L E Goldfinger</i>	608
Intercellular Ca^{2+} Waves: Mechanisms of Initiation and Propagation <i>L Leybaert and M J Sanderson</i>	613
Interferon Receptors <i>C P Elco and G C Sen</i>	619
Intermediate Filament Linker Proteins: Plectin and BPAG1 <i>P Fuchs and G Wiche</i>	624
Intermediate Filaments <i>K M Bernot and P A Coulombe</i>	631
Intracellular Calcium Channels: NAADP ⁺ -Modulated <i>A A Genazzani and M Debidia</i>	637
Intracellular Calcium Waves <i>J T Chun and L Santella</i>	640
Ion Channel Protein Superfamily <i>W A Catterall</i>	648
IP ₃ Receptors <i>K Mikoshiba and C Hisatsune</i>	653
Iron-Sulfur Proteins <i>R Cammack</i>	657
Keratins and the Skin <i>P A Coulombe, K M Bernot, and C-H Lee</i>	665

Kinesins as Microtubule Disassembly Enzymes <i>R Ohi and L Wordeman</i>	672
Kinesin Superfamily Proteins <i>N Hirokawa and R Takemura</i>	679
Kinetic Isotope Effects <i>J P Roth and J P Klinman</i>	688
<i>lac</i> Operon <i>L Swint-Kruse and K S Matthews</i>	694
Lectins <i>N Sharon and H Lis</i>	701
LexA Regulatory System <i>M Butala, D Žgur-Bertok, and S J W Busby</i>	706
Ligand-Operated Membrane Channels: Calcium (Glutamate) <i>E K Michaelis</i>	712
Ligand-Operated Membrane Channels: GABA <i>F Minier and E Sigel</i>	722
Light-Harvesting Complex I and II: Pigments and Proteins <i>S Jansson</i>	726
Lipases <i>H L Brockman</i>	729
Lipid Bilayer Structure <i>E London</i>	733
Lipid Modification of Proteins: Targeting to Membranes <i>M D Resh</i>	736
Lipid Rafts <i>D A Brown</i>	741
Lipid Signaling and Ion Channels <i>J B Jensen, B H Falkenburger, and B Hille</i>	745
Live Cell Imaging of Nuclear Dynamics <i>J C Harr and K L Reddy</i>	749
Low-Barrier Hydrogen Bonds <i>P A Frey</i>	756
Luft's Disease <i>S DiMauro</i>	760
Lysophospholipid Receptors <i>G J Tigyi</i>	762
VOLUME 3	
Major Sperm Protein and Sperm Locomotion <i>K Shimabukuro and T M Roberts</i>	1
Mammalian DNA Methyltransferase Structural Themes <i>X Cheng and H Hashimoto</i>	6
Mass Spectrometry of Native Complexes <i>R J Rose and A J R Heck</i>	11

Meiosis <i>N Hunter</i>	17
Melanocortin System <i>J Pantel and R D Cone</i>	24
Membrane-Associated Energy Transduction in Bacteria and Archaea <i>G Schäfer</i>	28
Membrane Fusion <i>R K Plumper and G B Melikyan</i>	36
Membrane Transporters: Na ⁺ /Ca ²⁺ Exchangers <i>J Lytton</i>	44
Membrane Transport, General Concepts <i>S G Schultz</i>	49
Messenger RNA Degradation in Bacteria <i>D H Bechhofer</i>	52
Messenger RNA Processing in Eukaryotes <i>K Potter, N Cremona, and J A Wise</i>	59
Metabolic Control during Ischemia of the Heart <i>J R Ussher and G D Lopaschuk</i>	65
Metabolic Roles of Adiponectin <i>K Sun and P E Scherer</i>	72
Metabolic Roles of the Orexigenic and Anorexigenic Neuropeptides <i>G J Morton and M W Schwartz</i>	76
Metabolite Channeling: Creatine Kinase Microcompartments <i>U Schlattner, M Tokarska-Schlattner, and T Wallimann</i>	80
Metalloproteases <i>D S Auld</i>	86
Metalloproteinases, Matrix <i>H Nagase and G Murphy</i>	90
MicroRNAs in Eukaryotes <i>R N Douglas and J A Birchler</i>	98
Microtubule-Associated Proteins <i>N Hirokawa and R Takemura</i>	103
Mitochondria in Myocardial Ischemia <i>F Di Lisa</i>	108
Mitochondrial Auto-Antibodies <i>H Baum</i>	113
Mitochondrial Calcium Transport: Historical Aspects <i>E Carafoli</i>	118
Mitochondrial Channels <i>M C Sorgato and A Bertoli</i>	127
Mitochondrial DNA <i>G Schatz</i>	132
Mitochondrial Dynamics <i>L Scorrano</i>	135

Mitochondrial Genes and Their Expression: Yeast <i>P P Slonimski and G Carignani</i>	140
Mitochondrial Membranes, Structural Organization <i>C A Mannella</i>	145
Mitochondrial Metabolite Carrier Protein Family <i>F Palmieri</i>	149
Mitochondrial Outer Membrane and the VDAC Channel <i>M Colombini</i>	159
Mitochondrial Permeability Transition Pore <i>P Bernardi</i>	162
Mitochondria: Source and Target of Free Radicals <i>A Boveris and E Cadenas</i>	168
Mitogen-Activated Protein Kinase Family <i>H Teramoto and J S Gutkind</i>	176
Mitosis <i>P Wadsworth and J Titus</i>	181
Monoamine Oxidase <i>A Mattevi</i>	185
mRNA Polyadenylation in Eukaryotes <i>J L Manley and D C Di Giammartino</i>	188
mTOR and its Downstream Targets <i>C G Proud</i>	194
Mucin Family of Glycoproteins <i>R B Brown and M A Hollingsworth</i>	200
Mucins in Embryo Implantation <i>N Dharmaraj and D D Carson</i>	205
Multidrug Resistance Membrane Proteins <i>N C Rockwell</i>	210
Muscarinic Acetylcholine Receptors <i>N M Nathanson</i>	215
Myosin Motors <i>M A Titus and R E Larson</i>	218
Natriuretic Peptides, Their Receptors and Therapeutic Applications <i>K Barbieri and L R Potter</i>	225
Neoglycoproteins <i>Y C Lee and R T Lee</i>	229
Neuronal Intermediate Filaments <i>R K H Liem</i>	233
Neurotransmitter Transporters <i>K Erreger, H J G Matthies, A Galli, and C Saunders</i>	238
Neurotrophin Receptor Signaling <i>R S Kenchappa and B D Carter</i>	241
Nicotinamide Nucleotide Transhydrogenase <i>J Rydström</i>	247

Nitric Oxide Signaling	253
<i>M A Marletta</i>	
N-Linked Glycan-Processing Glucosidases and Mannosidases	257
<i>T Suzuki</i>	
Nonhexameric SF1 DNA Helicases/Translocases	261
<i>E Antony and T M Lohman</i>	
Nonhomologous End Joining in Bacteria	266
<i>E J Bartlett and A J Doherty</i>	
Nonhomologous End Joining in Eukaryotes	269
<i>E Weterings and D J Chen</i>	
Nonhomologous Recombination: Bacterial Transposons	275
<i>B T Hoang and M G Chandler</i>	
Nonhomologous Recombination: Retrotransposons	283
<i>X Qi and S Sandmeyer</i>	
Nuclear Calcium in Synapse-to-Nucleus Communication: Common Regulator of Persistent Adaptations in the Nervous System	292
<i>H Bading</i>	
Nuclear Compartmentalization	296
<i>K P Smith and J B Lawrence</i>	
Nuclear Factor Kappa B	302
<i>T D Gilmore</i>	
Nuclear Genes Involved in Mitochondrial Function and Biogenesis	306
<i>A Tzagoloff and C L Dieckmann</i>	
Nuclear Lamina	310
<i>B M Paschal and J B Kelley</i>	
Nuclear Organization, Chromatin Structure, and Gene Silencing	314
<i>M W Vitalini, G Dialynas, L L Wallrath, and S R Mackey</i>	
Nuclear Pores and Nuclear Import/Export	318
<i>A Lange and A H Corbett</i>	
Nucleolus, Overview	324
<i>T Pederson</i>	
Nucleotide Excision Repair, Bacterial: The UvrABCD System	328
<i>B Van Houten</i>	
Nucleotide Excision Repair: Biology	337
<i>E C Friedberg</i>	
Nucleotide Excision Repair in Eukaryotes	341
<i>N Goosen</i>	
Olfactory Receptors	345
<i>S I Korsching</i>	
O-Linked GlcNAc Biosynthesis and Function	350
<i>K Sakabe and G W Hart</i>	
Opioid Receptors	354
<i>P Y Law and H H Loh</i>	

ORAI Store-Operated Calcium Channel <i>Y Zhou, P G Hogan, and A Rao</i>	359
Organization of the Bacterial Nucleoid <i>C J Dorman</i>	367
Oxygenases <i>O Hayaishi</i>	371
P2Y Receptors <i>G R Dubyak</i>	375
p53 Family <i>C E Barton and J A Pietsenpol</i>	379
Parathyroid Hormone/Parathyroid Hormone-Related Protein Receptor <i>T J Gardella</i>	384
Pentose Phosphate (Hexose Mono Phosphate) Pathway <i>B R Landau</i>	391
Peptide Amidation <i>M J Niciu, R E Mains, and B A Eipper</i>	396
Periplasmic Electron-Transport Systems in Bacteria <i>D J Richardson, G Sawers, and R J M Van Spanning</i>	400
Peroxidase Biochemistry and Redox Signaling <i>A Bindoli and M P Rigobello</i>	407
Peroxisomal Metabolism <i>S J Gould</i>	413
Peroxisome Proliferator-Activated Receptors <i>M C Sugden, P W Caton, and M J Holness</i>	418
Peroxisomes <i>S Joshi and S Subramani</i>	425
Phage Display for Protein Binding <i>H B Lowman</i>	431
Phagocytosis <i>G Wright, K Srinivasan, and C Janetopoulos</i>	437
Pheromone Receptors (Yeast) <i>J B Konopka and J W Thorner</i>	441
Phosphatidylinositol-3-Phosphate <i>J V Virbasius and M P Czech</i>	447
Phosphatidylinositol Bisphosphate and Trisphosphate <i>A Toker</i>	451
Phosphofructokinase-2/Fructose Bisphosphatase-2 <i>D M Raben</i>	456
Phosphoinositide 4- and 5-Kinases and Phosphatases <i>S F Bairstow, M W Bunce, and R A Anderson</i>	459
Phosphoinositide-Dependent Protein Kinases <i>N R Leslie and C P Downes</i>	463
Phospholipase C <i>F Sekiya</i>	467

Phospholipase D	472
<i>W Ali and M A Frohman</i>	
Phospholipid Synthesis in Yeast	478
<i>G-S Han and G M Carman</i>	
Photoinhibition and Photoprotection in Plants, Algae, and Cyanobacteria	482
<i>G M Giacometti and T Morosinotto</i>	
Photoreceptors	488
<i>K-W Yau</i>	
Photosynthesis	492
<i>R C Leegood</i>	
Photosynthetic Carbon Dioxide Fixation	497
<i>M J Paul</i>	
Photosystem I	503
<i>I Grotjohann and P Fromme</i>	
Photosystem II: Assembly and Turnover of the D1 Protein	508
<i>S Sirpiö, E Kanervo, and E-M Aro</i>	
Photosystem II: Redox and Protein Components	512
<i>J Barber</i>	
Photosystem II: Water Oxidation, Overview	520
<i>F Rappaport and A Boussac</i>	
Plasma-Membrane Calcium Pump: Structure and Function	525
<i>M Brini</i>	
Plastocyanin	530
<i>E L Gross</i>	
Platelet-Activating Factor Receptor	533
<i>K M Howard</i>	
Platelet-Derived Growth Factor Receptor Family	538
<i>M V Kovalenko and A Kazlauskas</i>	
Porphyrin Metabolism	544
<i>M Shepherd, A E Medlock, and H A Dailey</i>	
Potassium Channels in the Inner Membrane of Mitochondria in Various Organisms: From Unicellular Eukaryotes to Higher Plants and Mammals	550
<i>I Szabò and M Zoratti</i>	
Pre-tRNA and Pre-rRNA Processing in Bacteria	554
<i>Z Li</i>	
Primary Biliary Cirrhosis	561
<i>M M Kaplan</i>	
Prions of Yeast and Fungi: Proteins Acting as Genes	566
<i>R B Wickner</i>	
Prions Overview	571
<i>O Bannach and D Riesner</i>	
Processivity Clamps in DNA Replication	576
<i>N Y Yao and M O'Donnell</i>	

Prostaglandins and Leukotrienes <i>W L Smith and R C Murphy</i>	581
Proteases in Blood Clotting <i>J D Kulman and E W Davie</i>	585
Proteasomes, Overview <i>M Rechsteiner</i>	590
26S Proteasome: Structure and Function <i>F Förster and E Sakata</i>	595
Proteinase-Activated Receptors <i>R Ramachandran, K K Hansen, and M D Hollenberg</i>	601
Protein Carboxyl Esterification <i>E Perez and J B Stock</i>	607
Protein Data Resources <i>N Furnham, J M Thornton, and J S Garavelli</i>	611
Protein Degradation <i>A L Goldberg</i>	617
Protein Folding and Assembly <i>D P Goldenberg</i>	625
Protein Import into Mitochondria <i>J M Herrmann and W Neupert</i>	632
Protein Kinase C Family <i>A C Newton</i>	637
Protein N-Myristoylation <i>E D Sonnenburg and J I Gordon</i>	641
Protein Palmitoylation <i>R J Deschenes</i>	645
Protein Tyrosine Phosphatases <i>D J Pagliarini, F L Robinson, C A Worby, and J E Dixon</i>	648
Proteoglycans <i>S K Olson and J D Esko</i>	654
Proton Pumping in the Respiratory Chain <i>V Rauhamäki and M Wikström</i>	661
Pteridines <i>S C Daubner and P F Fitzpatrick</i>	666
P-Type Pumps: Copper Pump <i>Y Hatori and S Lutsenko</i>	670
P-Type Pumps: H ⁺ /K ⁺ Pump <i>J M Shin and G Sachs</i>	676
P-Type Pumps: Na ⁺ ,K ⁺ -ATPase <i>H Haviv and S J D Karlish</i>	681
P-Type Pumps: Plasma-Membrane H ⁺ Pumps <i>A B Mason and C W Slayman</i>	688
Purple Bacteria: Electron Acceptors and Donors <i>A Adessi and R De Philippis</i>	693

Purple Bacteria: Photosynthetic Reaction Centers <i>C R D Lancaster</i>	700
Pyridoxal Phosphate <i>M D Toney</i>	708
Pyrimidine Biosynthesis and Degradation (Catabolism) <i>M Löffler and E Zameitat</i>	712
Pyruvate Kinase <i>K Uyeda</i>	719
Quinones <i>G Lenaz and M L Genova</i>	722
VOLUME 4	
Rab Family <i>M W McCaffrey and A J Lindsay</i>	1
Ran GTPase <i>M S Moore</i>	7
Ras Family <i>L A Quilliam</i>	12
Reactive Oxygen and Nitrogen Species: Interactions with Mitochondria and Pathophysiology <i>B R Zelikson, S W Ballinger, L J Dell'Italia, J Zhang, and V M Darley-Usmar</i>	17
Recombination-Dependent DNA Replication <i>K N Kreuzer</i>	23
Recombination: DNA Helicases and Nucleases <i>J Sowa and G Ira</i>	29
Recombination: DNA-Strand Transferases <i>W D Wright and W-D Heyer</i>	35
RecQ Helicase Systems <i>G Mosedale and I D Hickson</i>	43
Regulated Intramembrane Proteolysis <i>J Ye</i>	50
Regulation of Body Weight by Malonyl-CoA in the CNS <i>S Rodriguez and M J Wolfgang</i>	56
Regulation of Chromatin Dynamics <i>E-J E Hong and C L Peterson</i>	59
Regulation of Gene Transcription by Hypoxia-Inducible Factor 1 <i>G L Semenza</i>	67
Renewable Hydrogen from Biomass <i>S Thammannagowda, L Magnusson, J H Jo, P-C Maness, and M Seibert</i>	72
Repair of G–T Mismatches by <i>Escherichia coli</i> Vsr and Eukaryotic DNA Glycosylases <i>M Lieb and A S Bhagwat</i>	76
Respiration in Phototrophic Microorganisms <i>R Borghese and D Zannoni</i>	79
Respiratory Chain and ATP Synthase <i>D G Whitehouse and A L Moore</i>	83

Respiratory Chain Complex I <i>U Brandt and V Zickermann</i>	87
Respiratory Chain Complex II and Succinate:Quinone Oxidoreductases <i>C R D Lancaster</i>	92
Respiratory Chain Complex IV <i>H Michel</i>	98
Retinoblastoma Protein (pRB) <i>N Singh, M Frolov, and N Dyson</i>	104
Retinoic Acid Receptors <i>M Petkovich</i>	108
Rho GTPases and Actin Cytoskeleton Dynamics <i>P Villalonga and A J Ridley</i>	111
Ribosome Assembly <i>A Sahasranaman and J L Woolford Jr.</i>	116
Ribosome Regulation by EF-G and EF-Tu <i>S T Gregory</i>	122
Ribosome Structure <i>T M Schmeing</i>	128
Riboswitches <i>A R Ferré-D'Amaré</i>	136
Ribozymes and Evolution <i>N Lehman</i>	142
Ribozyme Structural Elements: Group II Introns and the Spliceosome <i>O Fedorova and C Waldsich</i>	147
RNA Editing <i>C E Samuel</i>	154
RNA Polymerase I and RNA Polymerase III in Eukaryotes <i>R J White</i>	159
RNA Polymerase II and Its General Transcription Factors <i>J L Corden</i>	162
RNA Polymerase II Elongation Control in Eukaryotes <i>D H Price</i>	164
RNA Polymerase Reaction in Bacteria <i>S-H Jun, B A Warner, and K S Murakami</i>	167
RNA Polymerase Structure, Bacterial <i>S Borukhov</i>	173
RNA Splicing <i>M S Jurica and G A Roybal</i>	185
Roles of Micro-RNAs in Metabolism <i>Z Wei and G W Wong</i>	191
Ryanodine Receptor Calcium Ion Channels <i>G Meissner</i>	195
Secretases <i>R L Heinrikson</i>	199

Secretory Pathway <i>S Shikano and K J Colley</i>	203
Selenoprotein Synthesis <i>A Böck</i>	210
Septins and Cytokinesis <i>A S Maddox and C M Field</i>	214
Serine/Threonine Phosphatases <i>T S Ingebritsen</i>	219
Serotonin Receptor Signaling <i>P J Gresch</i>	224
Siglecs <i>A Varki</i>	229
Sigma Factors <i>J D Helmann and S R MacLellan</i>	232
Sliding Clamps in DNA Replication: <i>Escherichia coli</i> β -Clamp and PCNA Structure <i>M R Marzahn and L B Bloom</i>	236
Small GTPases <i>K H Pedone and C J Der</i>	242
Small RNAs in Bacteria <i>E R Westra, M M Jore, S Al-Attar, S J J Brouns, and J van der Oost</i>	249
Sodium–Calcium Exchanger: Structural Aspects <i>M Ottolia, D A Nicoll, and K D Philipson</i>	256
Somatostatin Receptors <i>A Schonbrunn</i>	262
Spastic Paraplegia <i>E I Rugarli</i>	267
Spectrophotometric Assays <i>B Chance</i>	273
Sphingolipid Biosynthesis <i>A H Merrill Jr.</i>	281
Sphingolipid Catabolism <i>A Abe and J A Shayman</i>	287
Src Family of Protein Tyrosine Kinases <i>J A Cooper</i>	293
Starvation <i>O E Owen and R W Hanson</i>	298
Steroid/Thyroid Hormone Receptors <i>R D Ward and N L Weigel</i>	308
Store-Operated Calcium Channels <i>A Amcheslavsky, A V Yeromin, A Penna, and M D Cahalan</i>	314
Structure and Regulation of Pyruvate Dehydrogenases <i>J L S Milne</i>	321
Structure–Function of the Cytochrome b_6f Complex of Oxygenic Photosynthesis <i>W A Cramer, E Yamashita, D Baniulis, J Whitelegge, and S S Hasan</i>	329

Structure of P-Type Adenosine Triphosphatases <i>I Dach and P Nissen</i>	335
Structure of Respiratory Complex I <i>J M Berrisford and L A Sazanov</i>	341
Substrate Binding, Catalysis, and Product Release <i>W W Cleland</i>	347
Sugar Nucleotide Transporters <i>C B Hirschberg</i>	350
Superoxide Dismutase <i>I Fridovich</i>	352
T7 RNA Polymerase <i>R Sousa</i>	355
Tachykinin/Substance P/Neurokinin-1 Receptors <i>L E Goldsmith and M M Kwatra</i>	360
Taste Receptors <i>J D Boughter Jr. and S D Munger</i>	366
T-Cell Antigen Receptor <i>A L Szymczak-Workman, L L Jones, C S Guy, and D A A Vignali</i>	369
T-Cell Receptor Signaling to NF- κ B <i>R R McCully and J L Pomerantz</i>	376
Telomeres: Maintenance and Replication <i>A Bianchi and D Shore</i>	380
Thyroid-Stimulating Hormone/Luteinizing Hormone/Follicle-Stimulating Hormone Receptors <i>D L Segaloff and M Ascoli</i>	387
Tight Junctions <i>S Tsukita</i>	392
Toll-Like Receptors <i>H Kumar, O Takeuchi, and S Akira</i>	396
Transcription-Coupled DNA Repair Overview <i>S Tornaletti</i>	402
Transcription Termination <i>T J Santangelo</i>	407
Transient Receptor Potential Channels <i>I S Ambudkar, X Liu, and B Paria</i>	412
Translation Elongation in Bacteria <i>S C Blanchard</i>	418
Translation Initiation in Bacteria: Factors and Mechanisms <i>L Brandi and C O Gualerzi</i>	425
Translation Initiation in Eukaryotes: Factors and Mechanisms <i>T V Pestova and C U T Hellen</i>	432
Tricarboxylic Acid Cycle <i>R L Veech</i>	436
tRNA Synthetases <i>R W Alexander</i>	441

Troponin <i>I Ohtsuki and S Morimoto</i>	445
Tubulin and Its Isoforms <i>E Nogales</i>	450
Tumor Necrosis Factor Receptors <i>T C Cheung and C F Ware</i>	454
Two-Hybrid Protein-Protein Interactions <i>I Serebriiskii, S Peri, and E A Golemis</i>	460
Tyrosine Sulfation <i>D Corbeil and W B Huttner</i>	463
Ubiquitin-Like Proteins <i>L-S Lu and E T H Yeh</i>	467
Ubiquitin System <i>A Ciechanover and M H Glickman</i>	473
UmuC D Lesion Bypass DNA Polymerase V <i>L A Hawver and P J Beuning</i>	477
Uncoupling Proteins <i>D Ricquier and F Bouillaud</i>	482
Unfolded Protein Responses <i>D Ron</i>	488
Urea Cycle: Disease Aspects <i>C Ficicioglu and M Yudkoff</i>	494
Vacuoles <i>C J Stefan, W M Henne, and S D Emr</i>	502
Vascular Endothelial Growth Factor Receptors <i>K A Thomas</i>	509
Vasopressin/Oxytocin Receptor Family <i>M J Brownstein</i>	515
V-ATPases <i>J Capecchi and M Forgac</i>	520
Vitamin A (Retinoids) <i>J L Napoli</i>	525
Vitamin C <i>F M Steinberg and R B Rucker</i>	530
Vitamin D <i>H F DeLuca and M Clagett-Dame</i>	535
Vitamin D Receptor <i>D R Dowd and P N MacDonald</i>	540
Vitamin E <i>J Atkinson, D Manor, and R Parker</i>	545
Vitamin K: Biochemistry, Metabolism, and Nutritional Aspects <i>J W Suttie</i>	551
Voltage-Dependent K ⁺ Channels <i>R Latorre, F J Morera, and C Zaelzer</i>	555

Voltage-Gated Ca^{2+} Channels <i>H Reuter</i>	560
Voltage-Gated Sodium Channels: Structure, Function, and Pathophysiology <i>W A Catterall</i>	564
von Hippel–Lindau (VHL) Protein <i>R C Conaway and J W Conaway</i>	570
XPV Polymerase and the Bypass of Ultraviolet DNA Damage <i>A R Lehmann</i>	572
Zinc Fingers <i>M Isalan</i>	575
<i>Index</i>	581

CONTENTS BY SUBJECT AREA

Bioenergetics

ABC Transporters
Algal Hydrogen Production
Amine Oxidases
Anaplerosis
Arachidonic Acid Regulated Calcium Channel
ATP in Plant Mitochondria: Substrates, Inhibitors, and Uncouplers
ATP Synthesis: Mitochondrial Cyanide-Resistant Terminal Oxidases
Bioenergetics and the Mitochondrial Genome
Bioenergetics: General Definition of Principles
Calcium-Binding Proteins: Cytosolic (Annexins, Gelsolins, and C₂-Domain Proteins)
Calcium, Biological Fitness of
Calcium-Buffering Proteins: Calbindin
Calcium-Buffering Proteins: ER Luminal Proteins
Calcium/Calmodulin-Dependent Protein Kinase II
Calcium in the Regulation of Gene Expression
Calcium-Modulated Proteins (EF-Hand)
Calcium Oscillations
Calcium-Sensing Receptor (CaSR)
Calcium Signaling by Cyclic ADP-Ribose and NAADP
Calcium Signaling: Calmodulin-Dependent Phosphatase
Calcium Signaling: NO Synthase
Calcium Transport in Mitochondria
Cardiac Excitation–Contraction Coupling
Cell Death by Apoptosis and Necrosis
Chemiosmotic Theory
Chemolithotrophy
Chlorophylls and Carotenoids
Chloroplasts
Chloroplast Signaling and Redox Control
Cytochrome *bc*₁ Complex (Respiratory Chain Complex III)
Cytochrome *c*
Cytochrome Oxidases, Bacterial
Cytochrome P-450
Dynamic Behavior of Photosystem II Light Harvesting
Energy Transduction in Anaerobic Bacteria

ER/SR Calcium Pump: Function
ER/SR Calcium Pump: Structure
F-ATPases
Ferredoxin
Ferredoxin-NADP⁺ Reductase
Friedreich's Ataxia
F_X, F_A, and F_B Iron–Sulfur Clusters in Type I Photosynthetic Reaction Centers
Giant Mitochondria (Megamitochondria)
Glutathione Peroxidases
Green Bacteria: Chlorophyll Biosynthesis, Light-Harvesting, Reaction Centers, and Electron Transport
Green Bacteria: Secondary Electron Donor (Cytochromes)
Green Bacteria: The Light-Harvesting Chlorosome
Heme Synthesis
Hydrogenases, Structure and Function
Indicators of Intracellular Calcium
Inositol Trisphosphate and Calcium Signaling
Intercellular Ca²⁺ Waves: Mechanisms of Initiation and Propagation
Intracellular Calcium Channels: NAADP⁺-Modulated
Intracellular Calcium Waves
IP₃ Receptors
Iron–Sulfur Proteins
Ligand-Operated Membrane Channels: Calcium (Glutamate)
Ligand-Operated Membrane Channels: GABA
Light-Harvesting Complex I and II: Pigments and Proteins
Lipid Signaling and Ion Channels
Luft's Disease
Membrane-Associated Energy Transduction in Bacteria and Archaea
Membrane Transporters: Na⁺/Ca²⁺ Exchangers
Membrane Transport, General Concepts
Metabolite Channeling: Creatine Kinase Microcompartments
Mitochondria in Myocardial Ischemia
Mitochondria: Source and Target of Free Radicals

Mitochondrial Auto-Antibodies
 Mitochondrial Calcium Transport: Historical Aspects
 Mitochondrial Channels
 Mitochondrial DNA
 Mitochondrial Dynamics
 Mitochondrial Genes and Their Expression: Yeast
 Mitochondrial Membranes, Structural Organization
 Mitochondrial Metabolite Carrier Protein Family
 Mitochondrial Outer Membrane and the VDAC Channel
 Mitochondrial Permeability Transition Pore
 Monoamine Oxidase
 Nicotinamide Nucleotide Transhydrogenase
 Nuclear Calcium in Synapse-to-Nucleus Communication: Common Regulator of Persistent Adaptations in the Nervous System
 Nuclear Genes Involved in Mitochondrial Function and Biogenesis
 ORAI Store-Operated Calcium Channel
 Oxygenases
 Periplasmic Electron-Transport Systems in Bacteria
 Peroxidase Biochemistry and Redox Signaling
 Phosphatidylinositol-3-Phosphate
 Photoinhibition and Photoprotection in Plants, Algae, and Cyanobacteria
 Photosystem I
 Photosystem II: Assembly and Turnover of the D1 Protein
 Photosystem II: Redox and Protein Components
 Photosystem II: Water Oxidation, Overview
 Plasma-Membrane Calcium Pump: Structure and Function
 Plastocyanin
 Protein Import into Mitochondria
 Proton Pumping in the Respiratory Chain
 P-Type Pumps: Copper Pump
 P-Type Pumps: H^+/K^+ Pump
 P-Type Pumps: Na^+, K^+ -ATPase
 P-Type Pumps: Plasma-Membrane H^+ Pumps
 Purple Bacteria: Electron Acceptors and Donors
 Purple Bacteria: Photosynthetic Reaction Centers
 Pyrimidine Biosynthesis and Degradation (Catabolism)
 Quinones
 Reactive Oxygen and Nitrogen Species: Interactions with Mitochondria and Pathophysiology
 Renewable Hydrogen from Biomass
 Respiration in Phototrophic Microorganisms
 Respiratory Chain and ATP Synthase
 Respiratory Chain Complex I
 Respiratory Chain Complex II and Succinate: Quinone Oxidoreductases
 Respiratory Chain Complex IV
 Ryanodine Receptor Calcium Ion Channels
 Sodium–Calcium Exchanger: Structural Aspects

Spastic Paraplegia
 Spectrophotometric Assays
 Store-Operated Calcium Channels
 Structure–Function of the Cytochrome b_6f Complex of Oxygenic Photosynthesis
 Structure of Respiratory Complex I
 Superoxide Dismutase
 Transient Receptor Potential Channels
 Troponin
 Uncoupling Proteins
 V-ATPases
 Voltage-Dependent K^+ Channels
 Voltage-Gated Ca^{2+} Channels
 Voltage-Gated Sodium Channels: Structure, Function, and Pathophysiology

Cell Architecture and Function

Actin Assembly/Disassembly
 Actin-Capping and -Severing Proteins
 Actin Organization
 Actin-Related Proteins
 Autophagy in Fungi and Mammals
 Bcl2 Family: Cell Death Enhancers and Inhibitors
 Cadherin-Mediated Cell–Cell Adhesion
 Caspases and Cell Death
 Cell Cycle: Control of Entry and Progression through S Phase
 Cell-Cycle Controls in G_1 and G_0
 Cell Cycle: DNA Damage Checkpoints
 Cell Cycle: Mitotic Checkpoint
 Cell Migration
 Cell Migration within Three-Dimensional Matrices
 Centromeres
 Chromosome Organization and Structure, Overview
 Cytokinesis
 Cytoskeletal Motors: General Principles
 Desmosomes and Hemidesmosomes
 Dynactin
 Dynein
 Endoplasmic Reticulum-Associated Protein Degradation
 Eukaryotic Chemotaxis
 Exosomes and Microvesicles
 Focal Adhesions and Related Integrin Contacts
 Golgi Complex
 Heat/Stress Responses
 Intermediate Filament Linker Proteins: Plectin and BPAG1
 Intermediate Filaments
 Keratins and the Skin
 Kinesins as Microtubule Disassembly Enzymes
 Kinesin Superfamily Proteins
 Live Cell Imaging of Nuclear Dynamics
 Major Sperm Protein and Sperm Locomotion

Meiosis
 Metalloproteinases, Matrix
 Microtubule-Associated Proteins
 Mitosis
 Myosin Motors
 Neuronal Intermediate Filaments
 Nuclear Compartmentalization
 Nuclear Lamina
 Nuclear Pores and Nuclear Import/Export
 Olfactory Receptors
 Peroxisomes
 Phagocytosis
 Potassium Channels in the Inner Membrane of
 Mitochondria in Various Organisms: From
 Unicellular Eukaryotes to Higher Plants and
 Mammals
 26S Proteasome: Structure and Function
 Rho GTPases and Actin Cytoskeleton Dynamics
 RNA Splicing
 Septins and Cytokinesis
 Tight Junctions
 Tubulin and Its Isoforms
 Unfolded Protein Responses
 Vacuoles

Lipids Carbohydrates Membranes and Membrane Proteins

Carbohydrate Chains: Enzymatic and Chemical
 Synthesis
 Cell–Matrix Interactions
 Detergent Properties
 Endocytosis
 Flippases
 Free Oligosaccharides: Formation and Degradation
 (Free Oligosaccharides Related to N-Linked
 Glycans)
 Glycation
 Glycerolipid Structure, Function, and Synthesis in
 Eukaryotes
 Glycolipid-Dependent Adhesion Processes
 Glycoprotein Folding and Processing Reactions
 Glycoprotein-Mediated Cell Interactions, Nonmucin-
 Type O-Linked
 Glycoproteins, N-Linked
 Glycoproteins, Plant
 Glycosylation, Congenital Disorders of
 Glycosylphosphatidylinositol Anchors
 Ion Channel Protein Superfamily
 Lectins
 Lipases
 Lipid Bilayer Structure
 Lipid Rafts
 Membrane Fusion
 Mucin Family of Glycoproteins
 Mucins in Embryo Implantation

Multidrug Resistance Membrane Proteins
 Neoglycoproteins
 N-Linked Glycan-Processing Glucosidases and
 Mannosidases
 O-Linked GlcNAc Biosynthesis and Function
 Phospholipid Synthesis in Yeast
 Prostaglandins and Leukotrienes
 Proteoglycans
 Secretory Pathway
 Siglecs
 Sphingolipid Biosynthesis
 Sphingolipid Catabolism
 Sugar Nucleotide Transporters

Metabolism Vitamins and Hormones

Adipocyte Fat Mobilization: The Regulatory Roles of
 Perilipin 1, Adipose Triglyceride Lipase, and
 Hormone-Sensitive Lipase
 Adipogenesis
 Amino Acid Metabolism
 Anorexigenic and Orexigenic Gut Peptides
 Aquaporin Water Channels
 Bile Salts
 Biochemistry: Niacin/NAD(P)
 Biochemistry of Development: Striated Muscle
 Biochemistry of Hematopoiesis
 Biochemistry of Liver Regeneration
 Biochemistry of Neurogenesis
 Biochemistry of Thiamine and Thiamine Phosphate
 Compounds
 Branched-Chain Amino Acids
 Carbohydrate Metabolism in the Central Nervous
 System
 Carbohydrate-Responsive Element Binding Protein
 Carnitine and β -Oxidation
 Chemistry of Alzheimer Disease
 Cholesterol Synthesis and Regulation
 Coenzyme A
 Color Vision
 Diabetes
 Fatty Acid Metabolism and Cancer
 Fatty Acid Oxidation
 Fatty Acid Structure and Synthesis
 Fatty Acyl-CoA Synthetases
 Folate and Vitamin B₁₂ Function
 Gastrointestinal Digestion and Absorption
 Gluconeogenesis
 Glucose/Sugar Transport in Bacteria
 Glucose/Sugar Transport in Mammals
 Glycogen Metabolism
 Glycogen Storage Diseases
 Glycolysis Overview
 Hexokinase/Glucokinase
 Insect Biochemistry/Hormones
 Insulin- and Glucagon-Secreting Cells of the Pancreas

Insulin Mechanisms/Metabolic Actions
Metabolic Control during Ischemia of the Heart
Metabolic Roles of Adiponectin
Metabolic Roles of the Orexigenic and Anorexigenic Neuropeptides
Pentose Phosphate (Hexose MonoPhosphate) Pathway
Peroxisomal Metabolism
Phosphofructokinase-2/Fructose Bisphosphatase-2
Photosynthesis
Photosynthetic Carbon Dioxide Fixation
Porphyrin Metabolism
Primary Biliary Cirrhosis
Pyruvate Kinase
Regulation of Body Weight by Malonyl-CoA in the CNS
Regulation of Gene Transcription by Hypoxia-Inducible Factor 1
Roles of Micro-RNAs in Metabolism
Starvation
Structure and Regulation of Pyruvate Dehydrogenases
T-Cell Receptor Signaling to NF- κ B
Tricarboxylic Acid Cycle
Urea Cycle: Disease Aspects
Vitamin A (Retinoids)
Vitamin C
Vitamin D
Vitamin E
Vitamin K: Biochemistry, Metabolism, and Nutritional Aspects

Molecular Biology

Alternative Splicing
Alternative Splicing: Regulation of *Drosophila melanogaster* Sex Determination
Catalysis by RNA, Structural Themes: Group I Introns
Chromatin: Methyl-CpG-Binding Proteins
Chromatin: Nucleosome Positioning – the GAL Promoter
Chromatin: Physical Organization
Conservative Site-Specific Recombination
DNA Base Excision Repair Pathways
DNA Damage: Alkylation
DNA Glycosylases: Mechanisms
DNA Helicases: Hexameric Enzyme Action
DNA Ligases: Mechanism and Functions
DNA Ligases: Structures
DNA Methyltransferases: Eubacterial GATC
DNA Mismatch Repair and Homologous Recombination
DNA Mismatch Repair and the DNA Damage Response
DNA Mismatch Repair in Bacteria
DNA Mismatch Repair in Disease and Aging

DNA Mismatch Repair in Mammals
DNA Modification, Restriction Endonucleases: Type I Enzymes
DNA Oxidation
DNA Polymerase β , Eukaryotic
DNA Polymerase I, Bacterial
DNA Polymerase III, Bacterial
DNA Polymerases: Kinetics and Mechanism
DNA Polymerases: Reverse Transcriptase Integrase, and Retrovirus Replication
DNA Replication: Eukaryotic Origins and the Origin Recognition Complex
DNA Replication Fork, Bacterial
DNA Replication Fork, Eukaryotic
DNA Replication: Initiation in Bacteria
DNA Replication, Mitochondrial
DNA Restriction and Modification: Type II Enzymes
DNA Restriction and Modification: Type III Enzymes
DNA Secondary Structure
DNA Sequence Recognition by Proteins
DNA Supercoiling
DNA Topoisomerases: Type I
DNA Topoisomerases: Type II
Eukaryotic DNA Polymerase α
Eukaryotic DNA Polymerase δ
Eukaryotic Protein Biosynthesis: The Elongation Cycle
Gene Expression in Bacterial Systems: The *trp* Operon and Attenuation
Gene Expression in Eukaryotes: RNA Polymerase II Structure
Genome-Wide Analysis of Gene Expression
HIV-1 Reverse Transcriptase Structures
Homologous Recombination in Meiosis
lac Operon
LexA Regulatory System
Mammalian DNA Methyltransferase Structural Themes
Messenger RNA Degradation in Bacteria
Messenger RNA Processing in Eukaryotes
MicroRNAs in Eukaryotes
mRNA Polyadenylation in Eukaryotes
Nonhexameric SF1 DNA Helicases/Translocases
Nonhomologous End Joining in Bacteria
Nonhomologous End Joining in Eukaryotes
Nonhomologous Recombination: Bacterial Transposons
Nonhomologous Recombination: Retrotransposons
Nuclear Organization, Chromatin Structure, and Gene Silencing
Nucleolus, Overview
Nucleotide Excision Repair, Bacterial: The UvrABCD System
Nucleotide Excision Repair: Biology
Nucleotide Excision Repair in Eukaryotes

Organization of the Bacterial Nucleoid
 Pre-tRNA and Pre-rRNA Processing in Bacteria
 Prions of Yeast and Fungi: Proteins Acting as Genes
 Processivity Clamps in DNA Replication
 Recombination-Dependent DNA Replication
 Recombination: DNA Helicases and Nucleases
 Recombination: DNA-Strand Transferases
 RecQ Helicase Systems
 Regulation of Chromatin Dynamics
 Repair of G–T Mismatches by *Escherichia coli* Vsr and
 Eukaryotic DNA Glycosylases
 Ribosome Assembly
 Ribosome Regulation by EF-G and EF-Tu
 Ribosome Structure
 Riboswitches
 Ribozymes and Evolution
 Ribozyme Structural Elements: Group II Introns and
 the Spliceosome
 RNA Editing
 RNA Polymerase I and RNA Polymerase III in
 Eukaryotes
 RNA Polymerase II and Its General Transcription
 Factors
 RNA Polymerase II Elongation Control in Eukaryotes
 RNA Polymerase Reaction in Bacteria
 RNA Polymerase Structure, Bacterial
 Sigma Factors
 Sliding Clamps in DNA Replication: *Escherichia coli*
 β -Clamp and PCNA Structure
 Small RNAs in Bacteria
 T7 RNA Polymerase
 Telomeres: Maintenance and Replication
 Transcription-Coupled DNA Repair Overview
 Transcription Termination
 Translation Elongation in Bacteria
 Translation Initiation in Bacteria: Factors and
 Mechanisms
 Translation Initiation in Eukaryotes: Factors and
 Mechanisms
 tRNA Synthetases
 UmuC D Lesion Bypass DNA Polymerase V
 XPV Polymerase and the Bypass of Ultraviolet DNA
 Damage

Protein/Enzyme Structure Function and Degradation

AAA-ATPases
 Affinity Tags for Protein Purification
 Allosteric Regulation
 Aminopeptidases
 Amyloid
 Aspartic Proteases
 B₁₂-Containing Enzymes
 Biotin
 Biotinylation of Proteins
 Calpain

Chaperonins
 Chemiluminescence and Bioluminescence
 Collagenases
 Collagens
 Cysteine Proteases
 Disulfide Bond Formation
 Elastin
 Enzyme Inhibitors
 Enzyme Kinetics
 Enzyme Reaction Mechanisms: Stereochemistry
 Flavins
 Heme Proteins
 HIV Protease
 Kinetic Isotope Effects
 Lipid Modification of Proteins: Targeting to
 Membranes
 Low-Barrier Hydrogen Bonds
 Mass Spectrometry of Native Complexes
 Metalloproteases
 Peptide Amidation
 Phage Display for Protein Binding
 Prions Overview
 Proteases in Blood Clotting
 Proteasomes, Overview
 Protein Carboxyl Esterification
 Protein Data Resources
 Protein Degradation
 Protein Folding and Assembly
 Protein N-Myristoylation
 Protein Palmitoylation
 Pteridines
 Pyridoxal Phosphate
 Regulated Intramembrane Proteolysis
 Secretases
 Selenoprotein Synthesis
 Substrate Binding, Catalysis, and Product Release
 Two-Hybrid Protein–Protein Interactions
 Tyrosine Sulfation
 Ubiquitin-Like Proteins
 Ubiquitin System
 Zinc Fingers

Signaling

Adenylyl Cyclases
 Adrenergic Receptors
 Angiotensin Receptors
 Arf Family
 B-Cell Antigen Receptor
 Bmp Signaling and Vascular Disease
 Bradykinin Receptors
 Brassinosteroids
 Cadherin Signaling
 Calcitonin Gene-Related Peptide and
 Adrenomedullin Receptors
 Calcitonin Receptor

- Calcium/Calmodulin-Dependent Protein Kinases
- c-fes* Proto-Oncogene
- Chemokine Receptors
- Chemotactic Peptide/Complement Receptors
- Cyclic AMP Receptors of *Dictyostelium*
- Cyclic GMP Phosphodiesterases
- Cyclic Nucleotide-Dependent Protein Kinases
- Cyclic Nucleotide Phosphodiesterases
- Cyclic Nucleotide-Regulated Cation Channels
- Cytokines
- Cytokinin
- Diacylglycerol Kinases and Phosphatidic Acid Phosphatases
- Dopamine Receptors
- Eicosanoid Receptors
- Endocannabinoids
- Evolving Concepts of Leptin
- FAK Family
- Fibroblast Growth Factor Receptors and Cancer-Associated Perturbations
- G₁₂/G₁₃ Family
- GABA_A Receptor
- GABA_B Receptor
- Gi Family of Heterotrimeric G Proteins
- Glutamate Receptors, Metabotropic
- Glycine Receptors
- Glycogen Synthase Kinase-3
- G-Protein-Coupled Receptor Kinases and Arrestins
- G Protein Signaling Regulators
- G_q Family
- Gs Family of Heterotrimeric G Proteins
- Hematopoietin Receptors
- Immunoglobulin (Fc) Receptors
- Inositol Phosphate Kinases and Phosphatases
- Integrin Signaling
- Interferon Receptors
- Lysophospholipid Receptors
- Melanocortin System
- Mitogen-Activated Protein Kinase Family
- mTOR and its Downstream Targets
- Muscarinic Acetylcholine Receptors
- Natriuretic Peptides, Their Receptors and Therapeutic Applications
- Neurotransmitter Transporters
- Neurotrophin Receptor Signaling
- Nitric Oxide Signaling
- Nuclear Factor Kappa B
- Opioid Receptors
- P2Y Receptors
- p53 Family
- Parathyroid Hormone/Parathyroid Hormone-Related Protein Receptor
- Peroxisome Proliferator-Activated Receptors
- Pheromone Receptors (Yeast)
- Phosphatidylinositol Bisphosphate and Trisphosphate
- Phosphoinositide 4- and 5-Kinases and Phosphatases
- Phosphoinositide-Dependent Protein Kinases
- Phospholipase C
- Phospholipase D
- Photoreceptors
- Platelet-Activating Factor Receptor
- Platelet-Derived Growth Factor Receptor Family
- Proteinase-Activated Receptors
- Protein Kinase C Family
- Protein Tyrosine Phosphatases
- Rab Family
- Ran GTPase
- Ras Family
- Retinoblastoma Protein (pRB)
- Retinoic Acid Receptors
- Serine/Threonine Phosphatases
- Serotonin Receptor Signaling
- Small GTPases
- Somatostatin Receptors
- Src Family of Protein Tyrosine Kinases
- Steroid/Thyroid Hormone Receptors
- Tachykinin/Substance P/Neurokinin-1 Receptors
- Taste Receptors
- T-Cell Antigen Receptor
- Thyroid-Stimulating Hormone/Luteinizing Hormone/Follicle-Stimulating Hormone Receptors
- Toll-Like Receptors
- Tumor Necrosis Factor Receptors
- Vascular Endothelial Growth Factor Receptors
- Vasopressin/Oxytocin Receptor Family
- Vitamin D Receptor
- von Hippel–Lindau (VHL) Protein

ENCYCLOPEDIA OF **BIOLOGICAL CHEMISTRY**

SECOND EDITION

Volume 1

A–Dis

AAA-ATPases

J Martin and A N Lupas, Max Planck Institute for Developmental Biology, Tübingen, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

AAA proteins A protein family of oligomeric ATPases with diverse cellular activities.

AAA+ proteins A superfamily that encompasses, among other families, the AAA proteins.

Arginine finger A conserved arginine residue in the SRH that points into the nucleotide-binding pocket.

Clade A group of organisms or molecules, whose members share homologous features derived from a common ancestor.

Cluster analysis A data analysis tool for solving classification problems. Organisms or molecules are sorted into clusters, so that the degree of association is strong between members of the same and weak between members

of different clusters. Each cluster thus describes the class to which its members belong.

P-loop NTPases Nucleotidases that are characterized by the presence of a loop element that is associated with the phosphates of bound nucleotides.

Second region of homology (SRH) A hallmark of the AAA family. It is a conserved sequence region containing sensor 1 residue and arginine finger.

Sensor 1 and 2 Conserved amino acid residues in AAA+ proteins that contact bound nucleotide; sensor 1 is a polar residue, sensor 2 typically arginine.

Walker A and B motifs Consensus sequence elements that are characteristic of nucleotide-binding folds.

AAA Proteins

Complex cellular events commonly depend on the activity of molecular machines that efficiently couple enzymatic and regulatory functions within a multiprotein assembly. As these functions require energy, adenosine triphosphatases (ATPases) are often found as integral components in these assemblies. All ATPases harness energy gained from breaking the γ -phosphate bond of ATP. An essential and expanding subset in the areas of protein remodeling and quality control comprises proteins of the AAA family.

Structure and Mechanism of Action

AAA proteins, which belong to the superfamily of P-loop nucleoside triphosphatases (NTPases), share the common feature of a highly conserved, extended nucleotidase domain containing 200–250 residues. This so-called AAA domain is formed by two subdomains, an amino-terminal nucleotide-binding subdomain and a smaller carboxy-terminal helical subdomain. Comparison to the structures of other AAA+ proteins reveals that this subdomain arrangement represents the common structural core of the AAA+ superfamily (see later).

The nucleotide-binding subdomain contains two classical Walker A (GX4GKT) and Walker B (HyDE) motifs, which are involved in binding of the triphosphate moiety of the

nucleotide and coordination of a Mg^{2+} -ion, which is important for subsequent hydrolysis of ATP. It is a RecA-like, three-layered $\alpha\beta\gamma$ structure with a central five-stranded, parallel β sheet. The property that separates this fold from that of all other P-loop NTPases is the insertion of a β strand which results in the addition of a conserved polar residue to the active site. The residue has been named sensor-1 for the fact that it contacts the γ -phosphate of the bound nucleotide and is thought to discriminate between ATP and ADP. Indeed, mutagenesis of this residue highlights its critical role in cooperative ATP hydrolysis. The so-called second region of homology (SRH), in which sensor-1 lies, is the hallmark of the AAA family and makes it distinguishable from the larger and more diverse superfamily of AAA+ proteins.

The carboxy-terminal subdomain is largely helical and shows substantially greater structural variation, although it invariably occupies a similar position, diagonally above the base of the bound nucleotide. A recurring feature of this subdomain is a residue (typically arginine), called sensor-2 for its ability to interact with the nucleotide. Depending on the subtype of AAA proteins, either one or two AAA domains (D1 and D2) are present. In some proteins with two domains, both are evolutionarily well conserved (like in Cdc48p/p97); in others, either the D2 domain (like in Pex1p and Pex6p) or the D1 domain (in Sec18p/NSF) is better conserved.

AAA proteins assemble into oligomers, which form ring-shaped structures with a rather narrow central pore. These

oligomers are typically hexameric, although in a few cases, heptameric forms have also been proposed. Depending on the respective protein, the inter-subunit contacts that mediate assembly can lie either in the N-domain (see later) or the AAA domain or both. Some members with two AAA domains have dedicated one domain for the purpose of maintaining the structural stability and integrity of the oligomeric complex, such as D1 in p97/Cdc48p or D2 in Sec18/NSF. These domains are degenerate and have lost their ability to hydrolyze ATP. In addition to the physiologically active hexameric form, some AAA proteins, like katanin, do exist also as dimers, which assemble into rings in the presence of substrate proteins or regulatory cofactors. Further oligomerization to dodecameric complexes was also noticed for some proteins, but the physiological relevance of dodecamers remains unclear.

Several crystal structures of AAA proteins have been determined, among them the complete structure of p97, an ATPase with two canonical AAA domains (Figure 1). These structures have shown that in the hexameric conformation, the ATP-binding site is positioned at the interface between subunits, such that the SRH bridges the space between nucleotide-binding pockets of adjacent subunits. Whereas the sensor-1 residue of one subunit points into its own pocket, the conserved arginine residue (the so-called arginine finger) points into the next pocket in the ring (Figure 2). This observation has suggested a mechanism for concerted nucleotide hydrolysis, in which loss of the γ -phosphate in one binding site is conveyed to the next active site via rigid-body motion of the helix connecting the sensor-1 and arginine finger residues. It also provides an explanation for the high degree of sequence conservation in the SRH. Interestingly, the meiotic clade of the AAA family and some minor clades (see later), are characterized by a two-residue deletion in the SRH immediately preceding the arginine finger.

Amino-terminally to the AAA domains are non-ATPase domains (N-domains) which are involved in substrate binding and recognition. Divergence in the function of the various family members occurs to a large part from the wide variety of N-domains that are found. These domains act as tool heads interacting with substrates, either directly or through adaptor molecules. Upon ATP binding and hydrolysis, AAA proteins undergo conformational changes in the AAA domains that are registered in the N-domains. The result is a repositioning of the N-domains in the ring relative to each other and to the AAA domains. Motions in the N-domains can be transmitted as mechanical force to a bound substrate, resulting, for example, in remodeling or dissociation of a protein complex, or in the complete unfolding of a polypeptide chain and its subsequent translocation through the central pore of the AAA protein ring. In a sense, AAA proteins use the AAA module like a chemo-mechanical converter (motor) powered by ATP hydrolysis to energize such a mechanical force. The conserved features of AAA domains imply a common mechanism for the operation of the motor. Yet, small differences in nucleotide-sensing residues, in the domains attached to the motor or in the connecting regions between these parts, may determine the direction and timing of the forces that are exerted. They may also

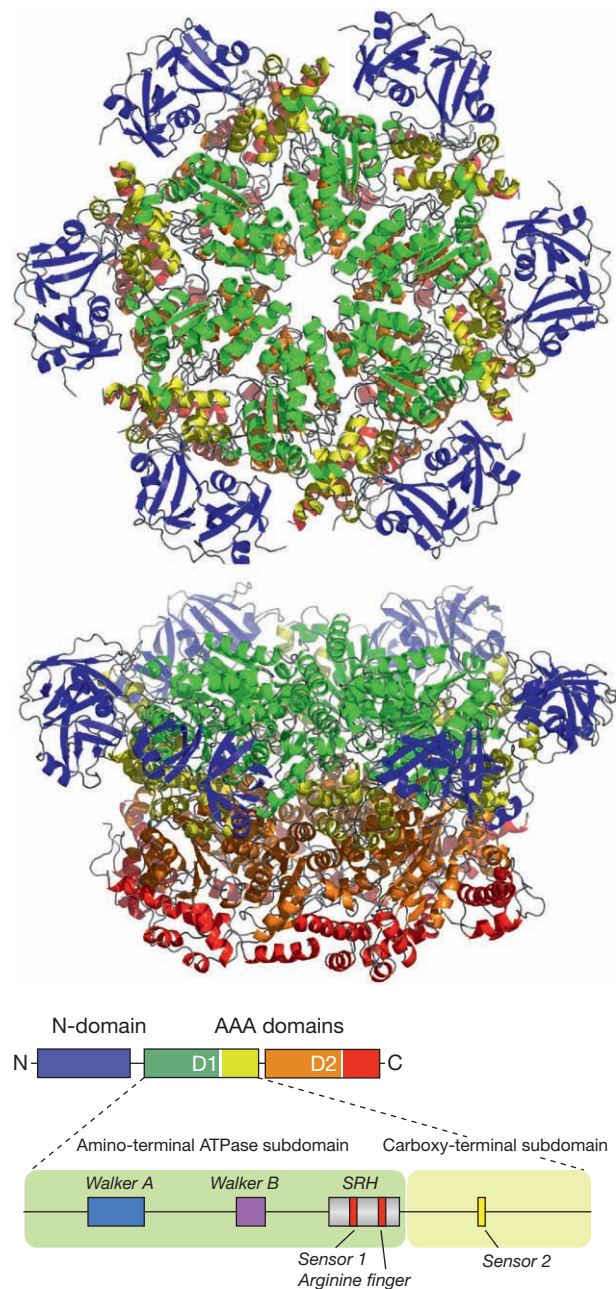


Figure 1 Crystal structure of a representative AAA protein with two ATPase domains, mouse p97. The domains are colored in their succession from amino- to carboxy terminus: N-domain blue, D1 P-loop subdomain green, D1 helical subdomain yellow, D2 P-loop subdomain orange, D2 helical subdomain red. The top panel shows a top view and the middle panel a side view of the complex. The bottom panel illustrates schematically the succession of N, D1, and D2 domains in the polypeptide chain, as well as an enlargement of one of the D domains with the salient sequence motifs labeled.

affect substrate specificity of these proteins. Sequence divergence in the AAA domains, acquisition of different N-domains, followed by gene duplication in case of proteins with two AAA domains, have thus produced an array of biological activities from a common ancestral protein.

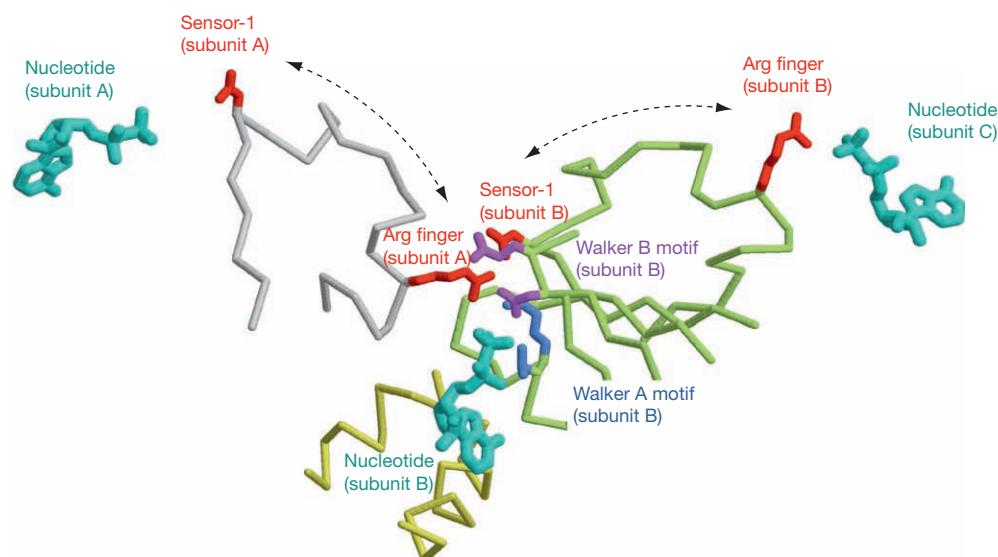


Figure 2 Enlarged, cut-away view of three consecutive nucleotide-binding sites in the p97 ring. Only the structural elements needed to illustrate the communication between active sites are shown. The residues participating in nucleotide binding and hydrolysis are labeled and color coded as in **Figure 1**. (Note that most AAA proteins, including p97 depicted here, lack a sensor-2 arginine.)

Table 1 Examples of AAA proteins (upper part) and AAA+ proteins (lower part) involved in cell biological processes

Protein	Function	Cellular location
PAN, ARC, Rpt1-Rpt6	Accessory factors to the proteasome in archaea, bacteria, and eukaryotes; unfolding and translocation of proteins	Cytosol
Pex1 and Pex6	Dislocation of PTS receptor from the peroxisomal membrane for recycling in the cytosol	Peroxisome and cytosol
NSF	Disassembly of SNARE complexes for vesicular transport	Cytosol, vesicles, Golgi, and endosomes
Vps4	Disassembly of ESCRT-III complex during budding into multivesicular bodies; related function in archaea?	Multivesicular body and endosomes in eukaryotes; cytosol in archaea
Cdc48p/p97	Retranslocation of misfolded proteins during ERAD, nuclear envelope, and Golgi reassembly after mitosis; archaeal VAT homologs with possible function in protein quality control	Cytosol, endoplasmic reticulum, Golgi
FtsH	Metalloprotease; degradation of misfolded membrane proteins	Inner membrane of bacteria and endosymbiotic organelles
i-AAA and m-AAA proteases	Maturation of mitochondrial proteins and mitochondrial protein quality control	Mitochondrial inner membrane
Katanin	Microtubule severing	Centrosome and spindle
BCS1	Formation of Rieske iron–sulfur cluster proteins	Mitochondrial inner membrane
Dynein	Minus-end-directed microtubule-based motor protein	Microtubules, spindle, vesicles
Lon	Protease; degradation of short-lived and misfolded proteins	Organelles and bacterial cytosol
Clp/Hsp100	Protein quality control in conjunction with proteases, disaggregation, and unfolding of proteins	Cytosol
DnaA	Initiation of chromosomal replication in bacteria, promotion of local DNA unwinding within the replication origin	Bacterial cytosol
RuvB clamp loader	Branch migration helicase in DNA recombination	Bacterial cytosol
Bchl	Component of magnesium chelatase complex involved in chlorophyll synthesis	Plastids and bacterial cytosol

Classification and Evolution: AAA and AAA+ (Super)Families

The classical AAA family is part of the AAA+ superfamily of ATPases that includes also a number of more distantly related cellular regulators. This superfamily has an expanded activity spectrum compared to the core AAA family (**Table 1**). AAA+

proteins are involved in protein degradation, membrane fusion, DNA replication, microtubule dynamics, intracellular transport, flagellar and ciliary beating, and disassembly of protein complexes and protein aggregates. Notably, some proteins like DnaA and RuvB interact with DNA. In their case, mechanical force results in the unwinding of the nucleic acid rather than unfolding of a protein.

The classification of AAA proteins is based on the analysis of their AAA domains. Several approaches have been used to compute phylogenies for AAA proteins, differing in the data available at the time (i.e., sequenced genomes) and in the inclusion of degenerate, inactive, and/or rapidly evolving sequences. More recently, cluster analysis has been applied toward the same goal. A cluster map of the AAA+ superfamily yields a well defined and compact group for AAA proteins, whose substructure can be further analyzed either by clustering at more stringent cutoffs or by classical phylogenetic methods. This approach outlines six major clades of AAA domains, namely proteasome subunits, metalloproteases, domains D1 and D2 of ATPases with two AAA domains, the meiotic group with MSP1/katanin/spastin, and BCS1 and its homologs, as well as a number of minor clades (Figure 3). Although deep branching, two of these minor clades are located next to major clades: ARC (AAA ATPase forming ring-shaped complexes)

next to the D1 and D2 clade, and a group of methanogenic sequences next to the metalloproteases, named AMA for its occurrence in *Archaeoglobus fulgidus* and methanogenic archaea. Most of the minor clades close to the root have gapped SRH regions and are similar to the AAA+ proteins in the positions of the conserved arginine finger residues, which may indicate an ancestral trait of these sequences. Clear clustering of the eukaryotic clades and predominance of prokaryotic sequences in deep-branching clades suggest that the AAA family had already reached most of its diversity before the three domains of life separated.

As mentioned above, the N-domains of the AAA proteins are crucial for substrate specificity and consequently for the cellular function of the respective AAA protein. A cluster analysis of N-domains resolves these sequences into 20 groups. This implies that the wide variety of biological functions of these proteins originates primarily from divergence in their

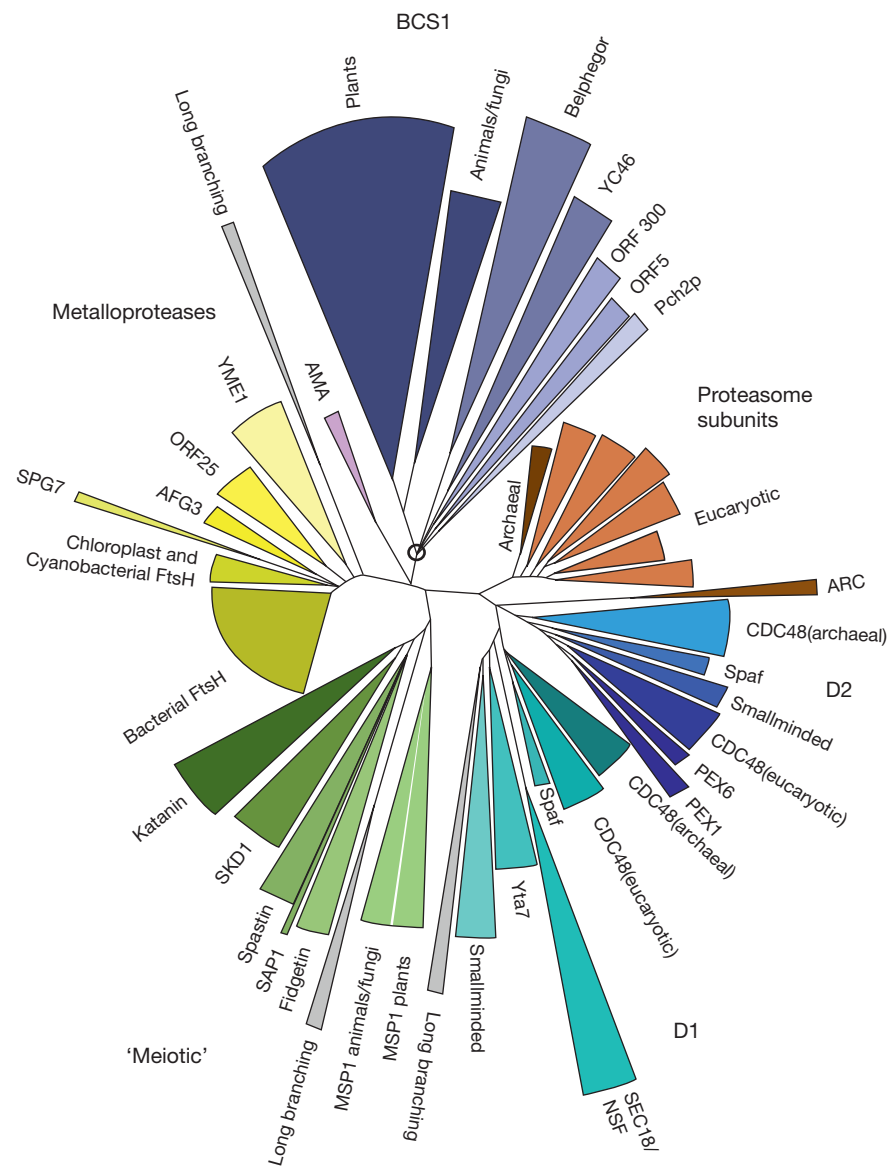


Figure 3 Phylogenetic tree of AAA proteins.

N-domains. Some of these differences were acquired after the division into plants and other eukaryotes (fungi and animals) as observed in the BCS1 clade. Other family members show high divergence in their N-domains although their AAA domains are of obvious monophyletic origin, suggesting an evolutionary exchange or recruitment of N-domains (like smallminded and Yta7 in the D1 clade). Interestingly, there are also N-domain homologies between the CDC48p/p97 group and the deep-branching AMA group as well as between ARC and proteasomal ATPases. Sequence analysis of N-domains of AMA proteins using PSI-Blast led to the assumption that these proteins adopt a structure similar to the β -clam fold found in N-domains of the CDC48p/p97 group. The homology between ARC and PAN proteins was proposed earlier mainly because of the presence of a coiled-coil region at the very end of the N-domains and of a similar genomic context (proteasome loci). Subsequent crystallographic and biochemical characterization of ARC indicated that it possesses indeed structural similarity to PAN. This similarity extends to function; ARC interacts with bacterial proteasomes and is thought to funnel unfolded substrates into the proteasomal ring chamber for proteolysis (see later for more details). This example demonstrates the usefulness of phylogenetic trees for gaining insight into possible biological functions of as yet uncharacterized members of the AAA protein family.

Cellular Functions of AAA Proteins

Members of the AAA family are found in all three organismal kingdoms and they are essential for many cellular processes (see Table 1). To showcase the functional diversity of this protein family, the following sections provide three important examples of well-characterized AAA proteins.

Mediating-Protein Degradation by the Proteasome

The proteasome forms the core of the protein quality control system in archaea and eukaryotes and also occurs in one bacterial lineage, the Actinobacteria. AAA proteins are destined as accessory factors of proteasomes, as they can use the energy from ATP hydrolysis to unfold potential protease substrates and transfer them into the interior of the associated proteasomal cylinder for degradation. The AAA proteins in the proteasome subunit clade form homo-oligomeric complexes in archaea and hetero-oligomeric complexes in eukaryotes. In eukaryotes, six distinct monomers form a hexameric ring that acts as an unfoldase and de-ubiquitinates tagged proteins prior to threading them through the proteasome for digestion. It appears that the different AAA proteins within a proteasome aid in the recognition and degradation of different subsets of proteins. The N-domains of the archaeal proteasomal ATPase proteasome-activating nucleotidase (PAN) and of its actinobacterial homolog, ARC or Mpa, form hexameric rings, whose identical subunits consist of an amino-terminal coiled coil and a carboxy-terminal OB domain. In ARC-N, the OB domains are duplicated and form separate rings. Detached from the ATPase domains, PAN-N and ARC-N can act as chaperones, demonstrating their general ability to interact with denatured polypeptides. It has been suggested that, in the context of the entire AAA protein, concerted radial

motions of the coiled coils relative to the OB rings are crucial for unfolding and/or translocation of the bound polypeptide chain through the pore of the ring. The process that determines which proteins eventually become substrates for the unfoldases and subsequently for the proteasome is still unknown. At least for a range of proteins, however, dedicated coupling factors do exist. Similar to eukaryotic ubiquitin, which tags proteins for degradation, small tagging factors have been found in archaea (SAMP) and actinomycetes (Pup). Pup has been shown to interact with the coiled coils of ARC, thereby recruiting any conjugated protein to the unfoldase.

Cdc48p/p97 is perhaps the best-studied AAA protein. It has also been implicated in proteasome-dependent protein degradation, but belongs to a different clade than the above-described proteins (see Figure 3). In eukaryotes, misfolded secretory proteins are exported from the endoplasmic reticulum (ER) and degraded by the ER-associated degradation pathway (ERAD). Nonfunctional membrane and luminal proteins are extracted from the ER and degraded in the cytosol by proteasomes. Substrate retrotranslocation and extraction is assisted by Cdc48p in complex with its adapter proteins Ufd1p and Npl4p on the cytosolic side of the membrane. Once it emerges on the cytosolic side, the substrate is ubiquitinated by ER-based E2 and E3 conjugating enzymes before degradation by the proteasome. The ancestral function of Cdc48p must lie somewhere else, though. Highly conserved homologs (named VAT) are found in archaea, in some organisms even as multiple variants. VATs may play a role in archaeal protein quality control, possibly in conjunction with the proteasome. Their ability to handle misfolded proteins could have made them well suited to adapt to new functions during the evolution of the eukaryotic cell.

Membrane Fusion and Vesicle Transport

Several AAA proteins seem to have a central role in reshaping cellular membranes. These include the originally defined AAA proteins with two ATPase domains in the D1 and D2 clade, namely Sec18/NSF, Pex1 and Pex6, YTA7, SPAF, and smallminded. Vesicular transport requires cargo selection and vesicle production through a budding process. The subsequent vesicle transport and targeting process concludes with specific membrane fusion of the vesicle to its target membrane. Early molecular analysis of vesicular trafficking between Golgi cisternae identified an essential *N*-ethyl-maleimide-sensitive factor (NSF). It functions as a soluble NSF attachment protein (SNAP) receptor (SNARE) dissociation factor. Aided by SNAPs, it binds to SNARE complexes and utilizes the energy of ATP hydrolysis to disassemble them, thus facilitating SNARE recycling.

A related process is mediated by vacuolar protein sorting 4 (Vps4), a protein involved in endosomal transport through multivesicular bodies. These bodies are endosomal compartments in eukaryotes where ubiquitinated membrane proteins are sorted into specific vesicles. This process involves the sequential action of three multiprotein complexes, ESCRT I–III (ESCRT standing for endosomal sorting complexes required for transport). Vps4p is anchored to the endosomal membrane and has been shown to catalyze the dissociation of ESCRT complexes. Assembly of Vps4p itself is assisted by the conserved Vta1p protein, which regulates its oligomerization

status and ATPase activity. Bioinformatic analysis has shown Vps4p and ESCRT homologs in archaea, pointing to the existence of a rudimentary mechanism for vesicle transport and membrane invagination. As archaea do not have elaborate intracellular membrane systems, the exact cellular role of these membrane-sculpting processes remains to be discovered.

Vps4 belongs to the meiotic clade of the AAA family, which derives its name from its member katanin, a microtubule severing protein implicated in mitosis and meiosis. Spastin, another member of this group, is involved in microtubule disassembly.

AAA Proteins As Membrane-Bound Metalloproteases

Proteins in the metalloprotease group are distinguished by a unique domain structure, with an extracellular N-domain, connected by a transmembrane helix to an AAA domain and a carboxy-terminal metalloprotease domain. They are found in the inner membranes of bacteria, where they degrade both soluble and membrane proteins, and in endosymbiont-derived organelles, where they primarily degrade unassembled subunits of membrane complexes. In the latter, they occur as two distinct complexes, one of which has an inverted orientation in the membrane.

The paradigm for the metalloprotease group is FtsH, the only essential protease of *Escherichia coli*. It forms a homo-hexamer, which, at least in some species, is further complexed with an oligomer of the periplasmic, membrane-bound modulating factor HflKC. FtsH is a processive endopeptidase and degrades certain misassembled membrane proteins, as well as a set of short-lived proteins, thereby enabling cellular regulation at the level of protein stability. Although the details of substrate recognition are largely unknown, the ATPase module is thought to be crucial for the ability of FtsH to dislocate membrane protein substrates out of the membrane and to translocate them to the metalloprotease domain. FtsH, thus, turns out to be an intriguing case of an AAA protein, where translocase activity and proteolytic function are integrated as successive modules into one polypeptide chain.

Conclusions

The function of several AAA family members has been well delineated by now and protein unfolding and disassembly emerges as a common mechanistic theme. An increasing number of high-resolution structural studies have provided a detailed picture of the conserved structure of the AAA domain. How these domains assemble into functional oligomers and change their conformation during ATP hydrolysis to generate movement and mechanical force has been studied less extensively. Which are the approaches likely to advance our knowledge in this area? High-resolution structures of oligomeric complexes in different nucleotide-bound states, and eventually

bound to substrates and/or cofactors, will enable a better molecular understanding of the inner workings of AAA proteins. To determine the precise function and mechanism of action of a particular ATPase, reconstitution of its reaction *in vitro* with purified components remains an essential step. Results from such studies would also shed more light on the fact that there is no strict correlation between AAA protein subtypes and clades on the one side, and a specific function and activity on the other side. This suggests that early in the evolution of AAA, there was a small number of subtypes that subsequently expanded and adapted to allow the processing of a wide variety of targets.

Another aspect gaining importance concerns the nature of disease mutations in AAA proteins, of which there is a growing number. Interestingly, some autosomal dominant diseases, such as hereditary spastic paraplegia (spastin defect), and some autosomal recessive diseases, such as mitochondrial complex III deficiency (BCS1 defect), are both caused by mutations in the AAA domain that are thought to affect ATP binding and hydrolysis; it is not clear whether for autosomal dominant diseases, the mutations act as dominant negatives. A thorough dissection of the generic AAA ATPase reaction cycle would, therefore, be beneficial for this research area as well.

See also: [Protein/Enzyme Structure Function and Degradation: Chaperonins](#); [Protein Degradation](#); [Protein Folding and Assembly](#).

Further Reading

- DeLaBarre B and Brunger AT (2003) Complete structure of p97/valosin containing protein reveals communication between nucleotide domains. *Nature Structural Biology* 10: 856–863.
- Erzberger JP and Berger JM (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annual Review of Biophysics and Biomolecular Structure* 35: 93–114.
- Frickey T and Lupas AN (2004) Phylogenetic analysis of AAA proteins. *Journal of Structural Biology* 146: 2–10.
- Halawani D and Latterich M (2007) p97: The cell's molecular purgatory? *Molecular Cell* 22: 713–717.
- Hanson PI and Whiteheart SW (2005) AAA+ proteins: Have engine, will work. *Nature Reviews Molecular Cell Biology* 6: 519–529.
- Lupas AN and Martin J (2002) AAA proteins. *Current Opinion in Structural Biology* 12: 746–753.
- Pye VE, Dreveny I, Briggs LC, et al. (2006) Going through the motions: The ATPase cycle of p97. *Special Issue: AAA+ ATPases. Journal of Structural Biology* 156: 12–28.
- Smith DM, Benaroudj N, and Goldberg A (2006) Proteasomes and their associated ATPases: A destructive combination. *Special Issue: AAA+ ATPases. Journal of Structural Biology* 156: 72–83.
- Suno R, Niwa H, Tsuchiya D, et al. (2008) Structure of the whole cytosolic region of ATP-dependent protease FtsH. *Molecular Cell* 22: 575–585.
- Tucker PA and Sallai L (2007) The AAA+ superfamily – a myriad of motions. *Current Opinion in Structural Biology* 17: 641–652.
- White SR and Lauring B (2007) AAA+ ATPases: Achieving diversity of function with conserved machinery. *Traffic* 8: 1657–1667.

ABC Transporters

A Goffeau, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

B De Hertogh, Facultés Universitaires Notre-Dame de la Paix, Namur, Belgium

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by André Goffeau, Benoît De Hertogh, and Philippe V. Baret, volume 1, pp. 1–5, © 2004, Elsevier Inc.

Glossary

ABC1 Specific small amino acid sequence which is a specific signature of ABC proteins.

C-, D-, H-, P-, and Q-loop Loops are located at the junction of α -helices and are represented by conserved amino acid residue in one letter code: C is cysteine, D is aspartate, H is histidine, P is proline, and Q is asparagine.

NBD Nucleotide-binding domain of about 100 amino acids which is characterized by a Walker A and Walker B motif.

Pdr5p The major yeast ABC transporter involved in pleiotropic drug resistance including the azole-antifungals.

P-gp The major ABC transporter (glycoprotein) involved in mammalian multidrug resistance

and which is often activated during cancer chemotherapy.

Transmembrane span (TMS) and transmembrane domain (TMD) The existence of the transmembrane span is predicted by frequency of hydrophobic residues in a reading window of about 17 amino acids. In ABC transporters, the transmembrane domains usually comprise six contiguous transmembrane spans.

Transporter classification (TC) Classification of over 1000 transporter families, developed by Milton Saier based on a combination of mechanistic and phylogenetic criteria.

Walker A and B Small consensus of amino acid sequences involved in ATP binding.

The ABCs constitute the largest family of proteins. They are present in all living species from *Archaea* to *Homo sapiens*. They make up to 4% of the full genome complement of bacteria such as *Escherichia coli* or *Bacillus subtilis*. Each eukaryote genome contains several dozens of ABC proteins. They are recognized by a consensus nucleotide-binding domain (NBD) of approximately 100 amino acids which comprises a series of structural and sequence motifs illustrated in **Figure 1**: the Walker A motif, the ABC1 signature, the Walker B motif, the Helical domain, and the P, Q-, C-, D-, and H-loops which are structural junctions between α -helices marked by a conserved proline, asparagine, cysteine, aspartate, or histidine residue. The ABC proteins catalyze a wide variety of physiological functions, many but not all of which are related to transport. We wish to describe the major physiological and biochemical functions as well as the structural properties of some of the best-known ABC transporters using as examples the multidrug exporters Sav1866 from *Staphylococcus aureus*, Pdr5p from *Saccharomyces cerevisiae*, and P-gp from *H. sapiens*.

Topology

Each ABC transporter molecule contains, or is associated to, one or two cytoplasmic adenosine triphosphate (ATP)-binding domains named NBDs and one or two transmembrane domains (TMDs). Each TMD comprises usually six transmembrane spans (TMSs) separated by the three extracytoplasmic loops ECL1, ECL2, and ECL3 and the two intracytoplasmic loops ICL2 and ICL3. Association of one TMD to one NBD results in a half-size ABC transporter, which functions as homo- or heterodimers so that the minimal functional organization of an ABC transporter is considered to be TMD–NBD/TMD–NBD or NBD–TMD/NBD–TMD. In bacteria, two NBDs

often associate with two TMDs either as four single subunits encoded within the same operon or in various combinations of fused subunits (**Figure 2**). Association of other proteins may occur. The most prominent associated bacterial protein is the periplasmic solute-binding receptor, which in Gram-negative bacteria is found in the periplasm, and in Gram-positive bacteria is often a lipoprotein bound to the external membrane surface via electrostatic interactions. The three domains of the bacterial ABC uptake transporters, namely the periplasmic-binding receptor, the cytoplasmic NBD, and the membrane TMD, are believed to have arisen from a common ancestral ABC transporter in which these three proteins were already present. However, during evolution the sequence of the periplasmic solute-binding receptors diverged more rapidly than that of the TMDs while that of NBDs remained the least divergent. Thus, all NBDs are homologous, but this is not true for the TMDs or the receptors.

In eukaryotes, two TMDs and two NBDs are often, but not always, associated in one single molecule called ‘full-sized ABC transporter’. The topological relation between NBDs and TMDs is variable (**Figure 3**); additional TM spans occur in some systems as well as extracytoplasmic domains of presumed regulatory function.

Phylogeny

The different families of ABC proteins transport a wide variety of substrates against their concentration gradient using the energy of ATP hydrolysis carried out by NBD. In bacteria, the transported substrates are either imported in or exported out of the cell. In eukaryotes, only extracytoplasmic exporters transporting substrates either out of the cell or into organelles are known. Within the ABC transporters, a total of 173

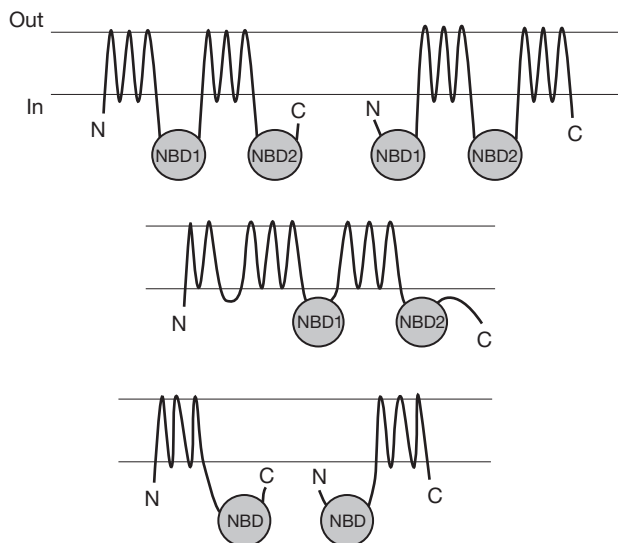


Figure 1 Example of topological relations between nucleotide-binding domains (NBDs) and transmembrane domains (TMDs) in full-sized and half-sized ABC transporters.

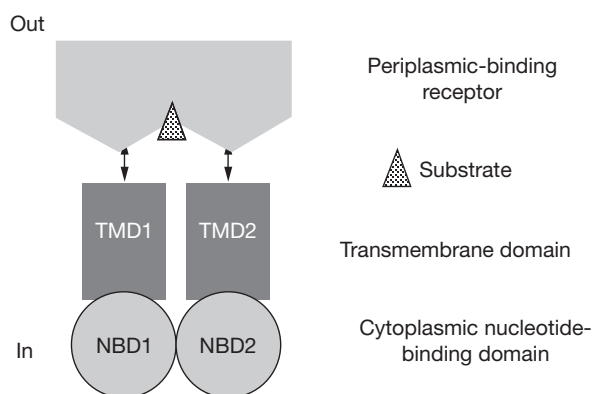


Figure 2 Example of protein subunits constituting bacterial ABC influx transporters.

phylogenetic families have been identified so far and classified within the transporter classification (TC) system developed by Milton Saier in San Diego. These families generally correlate with substrate specificity and comprise both efflux and influx transporters.

Since 1995, the *in silico* analysis of hundreds of full genome sequences has allowed the identification of thousands of new ABC proteins. For instance, within the 49 ABC proteins identified in the human genome, five major families of transporters have been identified and named ABCA (or ABC1), ABCB (or MDR including the famous P-gp described below), ABCC (or MRP), ABCD (or ALD), and ABCG (or WHITE). This classification overlaps with the TC.

S. cerevisiae genome contains 32 ABC proteins among which 22 are associated to TMDs. The largest yeast ABC family is Pdrp identified in 1997 by Anabelle Decottignies and André Goffeau and shown later to be present in all fungi and plants.



Figure 3 Reconstructed structure of the purified yeast Pdr5p reconstituted in lipid-containing particles and observed by cryo electron microscopy.

This family is not detected in the animal kingdom. Conversely, the large human and mouse ABCA family identified in 1994 by Giovanna Chimini is not represented in fungal genomes.

Function and Diseases

The ABC transporters from Archaea, Gram-negative, and Gram-positive plasma membranes are widely different in their organization and composition. The variety of substrates transported: sugars, amino acids, lipids, ions, polysaccharides, peptides, proteins, toxins, drugs, antibiotics, xenobiotics, and other metabolites is reflected by the divergence of the periplasmic sensor and that of the TMD which control the specificity and the direction (efflux or influx) of transport.

Even if all eukaryotic ABC transporters are effluxers that comprise each a TMD fused to NBD, some of them are not directly involved in moving substrates. For instance, in the cystic fibrosis transmembrane regulator (CFTR) and in the sulfonyleurea receptor (SUR), the hydrolysis of ATP appears to be linked to the regulation of opening and closing of ion channels carried by the ABC protein itself or other proteins. The presence of NBD and TMD in all ABC transporters, however, suggests that a basic coupling mechanism exists for efflux and influx whatever the transported substrate. However, distantly related proteins utilize an NBD to drive diverse nontransport processes such as DNA repair or protein elongation or regulation of RNase activities.

One strong impetus for the study of mammalian ABC transporters is their involvement in human diseases. Many mendelian diseases and complex genetic disorders are caused by ABC transporters including cystic fibrosis, adrenoleukodystrophy, Stargardt disease, Tangier disease, immune deficiencies, progressive familial intrahepatic cholestasis, Dublin-Johnson syndrome, Pseudoxanthoma elasticum, persistent hyperinsulinemic hypoglycemia of infancy due to focal adenomatous hyperplasia, X-linked sideroblastosis and anemia, age-related macular degeneration, familial hypoapoproteinemia, fundus flavimaculatus, retinitis pigmentosa, and cone rod dystrophy. Frequent cases of resistance to plant, fungal, or mammalian pathogens and to chemotherapeutic treatments of cancers often result from induction or activation mutations of pleiotropic drug resistance (Pdrp), multiple drug resistance (MDR), or multidrug resistance-related protein (MRP) transporters. Basic

studies of human, plants, or pathogen ABC transporters would greatly benefit from heterologous expression of human ABC transporter genes in yeast or other cells but this technology is not fully satisfactory yet. Meanwhile, knockout technology in the mouse allows physiological analysis of the mammalian transporters.

Structure and Biochemical Mechanism

In 2010, over a dozen high-resolution X-ray structures of functional ABC transporters are available and have contributed to the emergence of a consensus view of the catalytic mechanism of transport. Recent analyses at the electron microscopy level of purified bacterial (BmrA from *B. subtilis*) and a fungal (Pdr5p from *S. cerevisiae*) drug efflux ABC transporter came to a remarkably coherent set of conclusions. In both cases, the basic structural unit seems to comprise four joining NBDs that correspond to two full-size Pdr5p or four half-size BmrA. The NBD are related to the TMDs through four distinct stalks. Each NBD is oriented at a fixed 90° angle relative to its neighbor NBDs. This raises the possibility of concerted rotation movements of the NBDs implying a certain flexibility of the stalks. No intramolecular or no intramembrane pores were observed even though there is room (or chamber) between the four stalks that join together at their NBD tips (Figure 4).

The nature of the stalks, TMD and NBD, was further identified by the analysis of the well-resolved structure of complete dimeric ABC bacterial and eukaryote transporters such as those obtained for the phospholipid flippase MsbA from *E. coli* and *Staphylococcus typhimurum*, the vitamin B12 importer BtuCD from *E. coli*, the importer of metal-chelate H11470/1 from *Haemophilus influenzae*, the drug exporter Sav1866 from *S. aureus*, and the mammalian multidrug exporter P-gp.

The folds of the membrane and stalk domains were found to be dissimilar in the different structures analyzed so far. This may not be surprising taking into account the different species analyzed, the different cocrystallized ligands, the different numbers of TMS, and the different functions of the proteins analyzed. In particular, the communication between the NBD and the TMD was found to be variable and, for instance, is carried out either through a long and complex intracytoplasmic loop (ICL) in MsbA, through a short L-shaped ICL between

the transmembrane spans 6 and 7 in BtuCD, or through the ICL3 and ICL 6 in Sav1866.

A clear contribution of the X-ray structures is the observation that two molecules of ATP are bound at the interface of the two NBDs. Each nucleotide-binding site comprises a Walker A motif from one monomer and the ABC1 C motif from the other monomer in a head-to-tail arrangement of the two interacting NBD monomers.

As the folds of the various ABC transporters are highly variable we focus on some basic structural properties of the bacterial Sav1866 and the mammalian P-gp multidrug drug exporters which have been reported, respectively, by Kaspar Locher in 2006 and Goffrey Chang in 2009.

The homodimeric Sav1866 transporter has two identical subunits, each comprising a NBD fused to a TMD made of six transmembrane α -helices. It exports a wide series of unrelated drugs such as the anticancer compounds doxorubicin and vinblastin. Two coupling α -helices, located in ICL3 and ICL6, interact with grooves from the two NBDs. This structural motif is conserved in all ABC transporters but its amino acid sequence is not conserved. Individual TMS from the two TMDs embrace each other and form two symmetric membrane bundles each made of six mixed TMS comprising two membrane α -helices from one TMD and four membrane α -helices from the other TMD. A central cavity was identified between the TMDs. As it faces outward, it was interpreted to represent an open drug extrusion pocket.

P-gp is a full size NBD–TMD–NBD–TMD transporter which is responsible for multiple drug resistance of mouse and human cell and to which has been attributed failure of chemo-treated cancer patients. The mouse P-gp was expressed in the yeast *Pichia pastoris*, purified and crystallized. This 3.8 Å structure is a nucleotide-free, inward conformation arranged in two symmetric halves each comprising an NBD and a TM bundle. Each TM bundle comprises six mixed TMS in which spans from the two TMDs embrace each other. A large drug-binding pocket of about 6000 cubic Å located at the intersection of the two TM bundles is open to both the inner membrane leaflet and the cytoplasm. A total of 73 amino acids' aromatic and hydrophobic residues facing this cavity was accessible to solvent and involved in the overlapping binding of three different ABC inhibitors. Only two amino acid residues are common to the three binding sites. Two portals are wide open to the inner

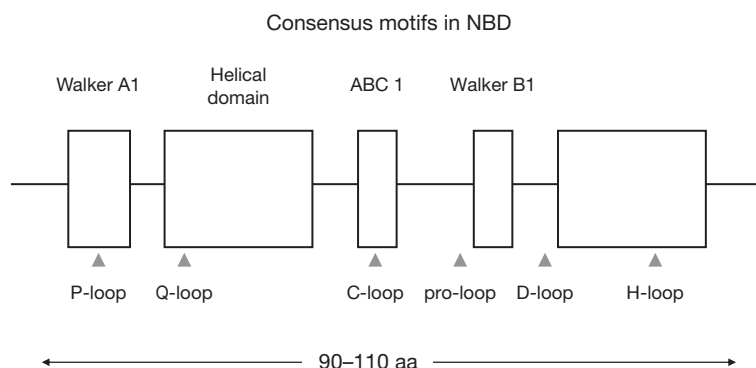


Figure 4 The consensus adenosine triphosphate (ATP)-binding region of a typical ABC protein is made of approximately 100 amino acids (aa), including the Walker A motif, the ABC1 signature, the Walker B motif, the helical domain, and the P-, Q-, C-, D-, and H-loops (see section Glossary).

phospholipid bilayer and allow entry of hydrophobic drugs which penetrate the membrane by passive diffusion.

Transport Mechanism

Recent structural observations have led to the disposal of the early hypothetic mechanism of alternating ATP-binding sites during the drug transport cycle. Instead, a consensus emerges favoring a rather simple alternating outward-inward access mechanism by which binding and hydrolysis of ATP by the cytoplasmic NBD controls the binding and translocation of the ligand through the TMD (Figure 5). Ligand binding to a high-affinity pocket formed by the TMDs induces a conformational change in the NBDs resulting in a higher affinity for ATP. Two molecules of ATP each binding to both NBDs induce a close dimeric conformation which promotes the extrusion of the ligand. ATP hydrolysis triggers the opening of the NBD dimer. Finally, phosphate and ADP are released and the ligand-free NBD dimer conformation is restored.

This two-state alternating access mechanism is shared not only by eukaryotic drug exporters but also by bacterial importers. This is shown by analyses of the type 1 bacterial importers such as the maltose (MalFGK), histidine (HisPQM), and methionine (MetNid/I) importers in which low-affinity substrate-binding proteins mediate the delivery of substrates to the ATP-bound form and also of the type 2 bacterial importers, such as the vitamin B12 (BtuCD) and metal-chelate (Hi1470/1) importers which involve a binding protein of high affinity for the substrate even in the absence of ATP. The delivery of substrates via a binding protein introduces an additional conformational switch that is of appreciable structural and mechanistic divergences between type 1 and type 2 importers.

The multiple conformation changes occurring during the ATP-driven substrate transport from prokaryotic and eukaryotic transporters are far from being unraveled; however, at present, it is best to regard the importers and the exporters as having partly distinct molecular mechanisms. Moreover, the

MDR family of mammalian ABC exporters may have mechanisms partially different from that of the Pdrp family of plant and fungal multidrug transporters. Obviously, the resolution of more complex mechanisms such as those involved in the gated chloride flux of CFTR, the regulation of the potassium channel of SUR1, or the loading of antigenic peptide by TAP1 will have to wait until more structural and biochemical information become available.

Substrate Specificity of the Multidrug Exporters and Reversal Agents

One of the most intriguing contemporary biochemical problems is the characterization of the interactions of the TMDs with their transported substrate. The most extraordinary feature in this context is the apparent lack of specificity of the yeast Pdr5p and human Pgp that transports hundreds of different chemicals, apparently contradicting the famous key/slot concept described in all textbooks. It is now recognized that at least three partly overlapping drug-binding sites operate in MDR, MRP, and Pdrp multidrug transporters and that drug binding depends on the hydrogen-bond acceptor properties, presence of aromatic rings, and hydrophobicity properties as well as on the surface volume of the substrates.

Since 1980, tremendous efforts have been made to identify inhibitors that can reverse MDR in cancer cells or Pdrp in fungi. Over 30 compounds with cabalistic names and structures have been shown to inhibit the cancer-related efflux pumps ABCB1 (MDR), ABCC1 (MRP), or ABCG2 (BCRP), or the fungal Pdrp transporters.

Such inhibitors may act directly by acting as pseudosubstrate, as competitive inhibitor of ATP binding or as a noncompetitive inhibitor acting at remote sites. They may also act indirectly on aspects of metabolism that affect efflux. The variety of identified drug pump inhibitors indicates that the development of efficient and specific inhibition of the anticancer and antifungal pumps should be possible. However, no

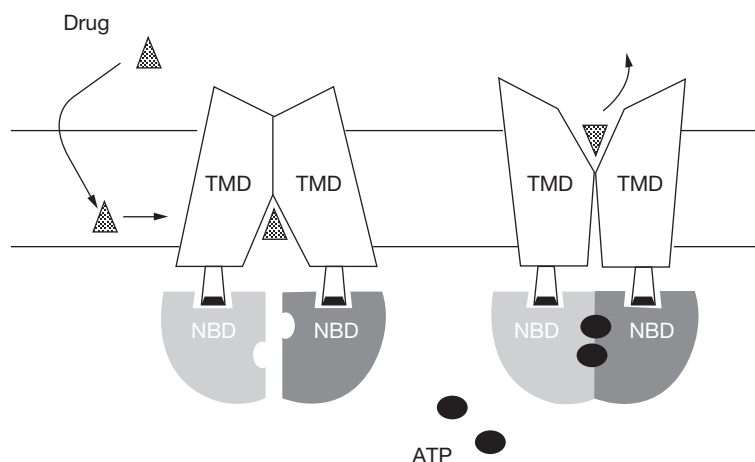


Figure 5 The flip-flop hypothesis for drug transport. Coupling between TMDs and NBDs is assured through conserved α -helices located in ICLs. The inward conformation of the TM bundles binds drugs in its cavity which is open both to the cytoplasm and to the inner phospholipid leaflet of the membrane. Upon binding of ATP, the NBDs dimerize and set the outward conformation in which the drug-binding site is exposed to the external milieu. ATP hydrolysis releases drugs, disrupts NBD dimerization, and resets the inward conformation.

ABC pump inhibitor is yet used clinically. Obviously, many pump inhibitors, such as the antimalarial quinine derivatives or the immunodepressor FK506 are known to act on diverse physiological functions. Moreover, ABC drug pumps seem to be involved in important physiological functions not related to drug efflux such as the development, differentiation, and maturation of immune cells. This may explain the multiple secondary effects of the inhibitors. One may expect that until the molecular structure and mechanisms of drug efflux are fully understood, the development of anti-MDR or Pdrp compounds will remain inefficient and merely based on brute screening. The ball is clearly in the hand of scientists studying basic mechanisms. However, some peculiar clinical treatments have been considered. For instance, low concentrations of isobuprofen, which is a potent anti-inflammatory drug, inhibits azole efflux from *Candida albicans*, and has been proposed for combination treatment of fungal infections with fluconazole.

In addition, alternative strategies such as regulating expression of drug transporters or uptake of anticancer drugs are also being considered.

Conclusion

The next frontiers in the study of ABC transporters are challenging. Many aspects of the evolutionary history of this large ubiquitous family are still to be unraveled. Better heterologous overexpression systems have to be developed to allow further biochemical and mechanistic studies. Additional atomic structures have to be produced to identify all conformation changes during the ATP-driven transport cycle. The diverse mechanisms

of ABC transporters in bacteria, fungi, plants, and animals have to be sorted out. Specific inhibitors of ABC drug importers and exporters have to be screened for. The physiological mechanisms of ABC-linked diseases have to be further studied in mouse knockouts. Systems prone to specific inhibition of ABC transporters expression by interfering RNA have to be explored. Genetic therapy of the ABC diseases has to be developed.

See also: [Lipids Carbohydrates Membranes and Membrane Proteins: Multidrug Resistance Membrane Proteins.](#)

Further Reading

- Aller SG, Yu J, Ward A, et al. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* 323: 1718–1722.
- Cannon RD, Holmes AR, Mason AB, et al. (2009) Efflux-mediated antifungal drug resistance. *Clinical Microbiology Reviews* 22: 291–321.
- Dean M (2002) *The Human ATP-Binding Cassette (ABC) Transporter Superfamily*, Monograph. Bethesda, MD: NCBI, National Library of Medicine (US).
- Decottignies A and Goffeau A (1997) Complete inventory of the yeast ABC proteins. *Nature Genetics* 15: 137–145.
- Gottesman MM and Ambudkar SV (2001) Overview: ABC transporters and human disease. *Journal of Bioenergetics and Biomembranes* 33: 453–458.
- Gutmann DA, Ward A, Urbatsch IL, Chang G, and van Veen W (2010) Understanding polyspecificity of multidrug ABC transporters: Closing in on the gaps in ABCB1. *Trends in Biochemical Sciences* 35: 36–42.
- Jones PM and George AM (2004) The ABC transporter structure and mechanism: Perspective on recent research. *Cellular and Molecular Life Sciences* 61: 682–699.
- Locher K (2009) Structure and mechanism of ATP-binding cassette transporters. *Philosophical Transactions of the Royal Society B Biological Sciences* 364: 239–245.

Actin Assembly/Disassembly

H N Higgs, Dartmouth Medical School, Hanover, NH, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Capping protein In theory, any protein that binds either the barbed or pointed end of actin filament and prevents monomer addition to that end. In practice, used to mean barbed-end-capping protein, and often heterodimeric capping protein in particular.

Critical concentration The concentration of monomers in a solution of actin at equilibrium between monomers and

filaments. Under polymerizing conditions, the critical concentration with both the barbed and pointed ends free is $\sim 0.1 \mu\text{M}$. When the barbed end is capped, the critical concentration rises to $0.5\text{--}0.7 \mu\text{M}$.

Sequestering protein Protein that binds actin monomer and prevents addition to either end of the filament. Neither profilin nor cofilin are sequestering proteins.

Actin is a monomeric protein that polymerizes into helical filaments. Apart from its role in muscle cells as a scaffold for myosin-based contraction, actin's function often depends on its ability to assemble into filaments from monomers rapidly and to disassemble equally rapidly. Actin alone can polymerize *in vitro*, but both the kinetics and equilibria of polymerization are controlled in cells by specific actin-binding proteins that serve to modify the assembly/disassembly cycle inherent to actin itself. Some actin-binding proteins of particular importance are sequestering proteins, profilin, capping protein, and ADF/cofilin.

Actin Structure

Actin Monomer Structures

Actin is a 43-kDa (375-amino-acid) globular monomer, which binds a nucleotide (adenosine triphosphate (ATP) or adenosine diphosphate (ADP) in cells) in a deep cleft between two halves of the protein (Figure 1(a)). The affinity of actin for nucleotide is greatly increased by divalent cation. Due to its relative cytosolic abundance (hundreds of micromolar), Mg^{2+} is the main actin-bound divalent cation in cells, but other cations, especially Ca^{2+} , are used for special purposes *in vitro*.

Actin Isoforms and Model Systems

Mammals contain six highly conserved actin isoforms, with at least 93% amino acid identity (Table 1). Actin from nonmammals is similarly conserved, with budding yeast 88% and *Acanthamoeba castellanii* 95% identical to human nonmuscle β -actin. Bacteria contain several proteins with some properties similar to actin, MreB, ParM, and Mbl.

Due to its ease of isolation, much of the biochemical data on actin assembly/disassembly are derived from studies of vertebrate muscle actin. Seminal studies on actin biochemistry were begun by Straub and Feuer during World War II using rabbit muscle actin. The sole actin isoform of budding yeast has also been a model, due to the combination of biochemistry and genetics in this system. Of the bacterial actin homologs, ParM is the most well characterized biochemically.

The highly similar sequences of actins from diverse eukaryotic species suggest that results in model systems will be close approximations for most actins. However, specific biochemical properties can vary between actins, such as affinities for some actin-binding proteins.

Individual cells can possess multiple actin isoforms simultaneously. Mammalian nonmuscle cells often contain both β - and γ -actin. The reasons behind this diversity are not well understood, but differential subcellular localization has been observed. Some unicellular organisms possess multiple actin isoforms. For instance, *Dictyostelium* has ~ 20 actin genes.

Actin Filaments

Actin filaments are right-handed two-stranded helices, with ~ 13 monomers per turn (Figure 1(b)). Each monomer adds $\sim 2.7 \text{ \AA}$ to the filament, so there are ~ 370 monomers per micron. All monomers are oriented in the same direction along the helix, making the filament polar. Due to the filament's arrow-like appearance when coated with myosin, the two ends are often called 'barbed' and 'pointed', respectively. The nucleotide-binding cleft faces the pointed end. Some refer to the barbed and pointed ends as '+' and '-', respectively.

Actin Dynamics

General Concepts in Actin Assembly/Disassembly

Under polymerizing conditions, actin monomers spontaneously polymerize into filaments, without need of other proteins. Polymerization from monomers can be divided into two processes: nucleation, whereby actin monomers assemble to form a trimeric nucleus, and elongation, whereby additional monomers add to this nucleus (Figure 2). Nucleation is highly unfavorable, with equilibrium dissociation constants in the millimolar range. Consequently, a major control point for cellular actin assembly is enhancement of nucleation rate. Elongation occurs rapidly at the barbed end (diffusion-limited on-rate) and 10-fold more slowly at the pointed end. For reasons explained later, pointed-end elongation rarely occurs in cells, and we will only consider barbed-end elongation in this article. While both

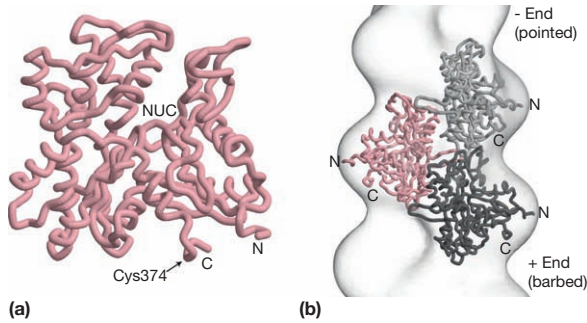


Figure 1 Structure of actin monomers and filaments. (a) Ribbon diagram of actin monomer backbone, oriented with the surface at the pointed end facing upward. NUC, nucleotide/divalent cation-binding pocket. N and C termini are indicated, as well as cysteine374, which is labeled by pyrene-iodoacetimide. (b) Atomic models of three actin monomers fit into a three-dimensional reconstruction of an actin filament obtained by electron microscopy and image analysis. (a, b) Modified from images provided by Dorit Hanein and Niels Volkmann.

Table 1 Actin isoforms

Species	Isoform	Identity ^a (%)	GI #
Human	Nonmuscle β	100	113270
Human	Nonmuscle γ	98	113278
Human	Skeletal muscle	93	113287
Human	Cardiac	94	113272
Human	Smooth muscle α	94	113266
Human	Smooth muscle γ	93	113279
Rabbit	Skeletal muscle	93	71611
Chicken	Skeletal muscle	93	71613
<i>Drosophila</i>	Muscle 88 F	95	2113792
Budding yeast	Sole isoform	88	2262056
<i>Acanthamoeba</i> ^b		95	71630

^aTo human nonmuscle β .

^bNumber of actin isoforms unknown.

Mg-ATP and Mg-ADP monomers can nucleate and elongate, Mg-ATP monomers do both faster. Upon adding to filaments, monomers are referred to as 'subunits'.

Monomers hydrolyze ATP extremely slowly. Upon adding to a filament, the hydrolysis rate increases to $\sim 0.3 \text{ s}^{-1}$. With a reasonable concentration of ATP-actin monomers, ATP hydrolysis occurs more slowly than elongation, which occurs at $10 \mu\text{M}^{-1} \text{ s}^{-1}$. Thus, a 'cap' of ATP-subunits builds up at the barbed end. Furthermore, the inorganic phosphate (Pi) hydrolysis product remains bound to the subunit for a considerable time before dissociating, while the ADP product remains tightly bound while the subunit remains on the filament. At steady state, three regions exist on an actin filament: ATP-subunits at the barbed end; ADP-Pi-subunits in the middle; and ADP-subunits toward the pointed end. As discussed later, Pi release is crucial to actin disassembly.

Subunits disassociate from both ends of a filament and do so more rapidly from the barbed end than from the pointed end. ADP-subunits release from the barbed end faster than ATP-subunits. Once released from the filament, the monomer slowly exchanges its ADP with ATP, assuming an excess of ATP in solution (which is the cellular situation).

Thus, actin assembly/disassembly occurs in a cycle (Figure 2). Nucleation occurs slowly, but filaments elongate rapidly once formed. Since barbed-end elongation is 10-fold faster than pointed-end elongation, and since ATP hydrolysis occurs slowly, ADP-subunits are at higher concentration toward the pointed end. Subunits dissociate from both ends and do so faster from the barbed end in both the ATP- and ADP-bound states. However, since monomer association is much more rapid at the barbed end, more net growth occurs at this end. Released ADP-monomers exchange ATP for ADP slowly (half-time $\sim 50 \text{ s}$).

Through this cycle, equilibrium between assembly and disassembly is reached, in which the concentrations of actin in filaments and as monomers become constant. When ATP is in constant supply (and assuming an excess of Mg^{2+} over Ca^{2+}), the equilibrium monomer concentration is $\sim 0.1 \mu\text{M}$ under polymerizing conditions. This monomer concentration, the 'critical concentration', is always found in solution at equilibrium

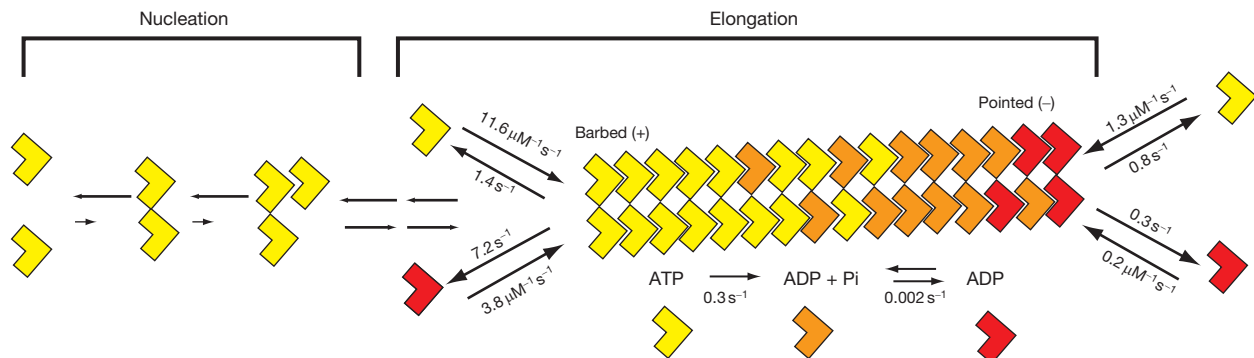


Figure 2 Assembly/disassembly of actin alone. Nucleation is highly unfavorable, signified by the large back arrows in the dimerization and trimerization steps. Once a tetramer is formed, the filament is more stable and monomers add to and disassemble from each end with the indicated kinetic constants (Pollard TD (1986) Rate constants for the reaction of ATP- and ADP-actin with the ends of actin filaments. *Journal of Cell Biology* 103: 2747–2754). Similar constants were determined more recently by a second method (Kuhn JR and Pollard TD (2005) Real-time measurements of actin filament polymerization by total internal reflection fluorescence microscopy *Biophysics Journal* 88: 1387–1402). Subunits hydrolyze bound ATP slowly upon addition to filaments, and release the Pi product even more slowly, creating sectors of the filament enriched in ATP-actin (yellow), ADP-Pi-actin (orange), and ADP-actin (red). Upon release from the filament, the ADP on the monomer exchanges slowly with ATP.

under these conditions. The concentration of actin in filaments is the excess above $0.1\ \mu\text{M}$. In other words, given $4\ \mu\text{M}$ actin under polymerizing conditions, $3.9\ \mu\text{M}$ will be in filaments and $0.1\ \mu\text{M}$ as monomers. With $100\ \mu\text{M}$ actin, $99.9\ \mu\text{M}$ is in filaments and, $0.1\ \mu\text{M}$ as monomers. At or below $0.1\ \mu\text{M}$, there are no filaments. Even at equilibrium, monomers constantly add to and release from both ends.

Polymerization Conditions

Three buffer conditions strongly affect actin's propensity to polymerize: ionic strength, the divalent cation present, and pH. Increased ionic strength increases elongation rate. The reason for this effect is not well understood but thought to be due to masking of charge repulsion between the highly anionic actin monomers. The ionic effect reaches a maximum at around the equivalent of $50\ \text{mM}$ KCl, and higher concentrations slow polymerization, probably due to inhibition of specific inter-monomer charge interactions in the filament. As for divalent cation, Mg-actin is better than Ca-actin at polymerizing, due to differences in monomer conformation. For pH, higher pH inhibits polymerization, likely due to increasing

the net negative charge of actin, hence increasing charge repulsion between monomers.

Effects of Actin-Binding Proteins

Actin-binding proteins are widely expressed in eukaryotes, generally at concentrations that exert a significant effect on cellular actin assembly/disassembly. Each protein alters one or more aspects of actin's assembly/disassembly cycle. The basic biochemical functions of these proteins are discussed in this section and illustrated in Figure 3, while their coordinated cellular effects are described in the next section.

Sequestering Proteins

Sequestering proteins, present in most eukaryotic cells, bind actin monomers and prevent their addition to either end of filaments (Figure 3). Metazoan cells often contain the small protein, thymosin ($5\ \text{kDa}$), the most well-characterized isoform being thymosin $\beta 4$. Yeasts do not have thymosins, but do express twinfilin, which can sequester. Twinfilin is present in mammals and other organisms as well.

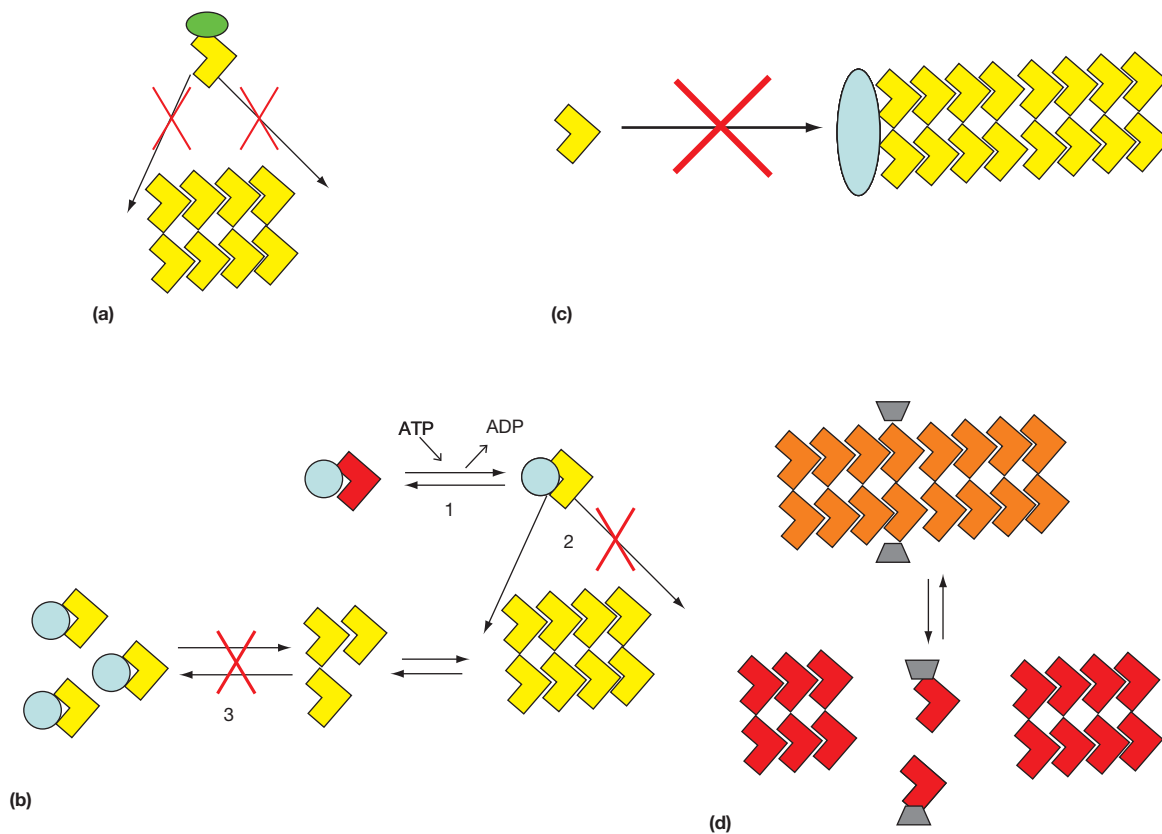


Figure 3 Effects of actin-binding proteins on actin dynamics. (a) Sequestering proteins (green oval) bind actin monomers and prevent addition to either end of the filament. (b) Profilin (light blue circle) binds monomers and has three effects: (1) acceleration of nucleotide exchange on monomers; (2) prevention of monomer addition to pointed ends; and (3) inhibition of spontaneous nucleation. (c) Capping proteins (light blue oval) bind filament barbed ends and prevent monomer addition. (d) ADF/cofilin (gray trapezoid) binds ADP-subunits on filaments and promotes P_i release from other subunits. ADF/cofilin also severs filaments. These two activities result in accelerated filament depolymerization. Actin colors: yellow = ATP, orange = ADP- P_i , red = ADP.

Profilin

Profilin is a 13-kDa protein that binds actin monomers, preferring ATP-actin over ADP-actin about fivefold. Profilin binding has three consequences for actin (Figure 3). First, profilin strongly inhibits spontaneous nucleation. Second, profilin prevents monomer addition to pointed ends. Third, profilin serves as a nucleotide exchange factor for actin, increasing exchange of ATP or ADP at least fivefold, though by some accounts its effect is much greater. A common misconception is that profilin is a sequestering protein. The profilin-bound actin monomer can add to barbed ends at a rate comparable to that of free monomer. An additional feature of profilin is that it binds stretches of polyproline found in many cytoskeletal proteins (e.g., vasodilator-stimulated phosphoprotein (VASP), formins, WASp/Scar/WAVE family proteins, and vinculin), binding such proteins and an actin monomer simultaneously. In the case of formin proteins, which bind filament-barbed ends, profilin-actin binding can accelerate barbed-end elongation rate dramatically by increasing the efficiency of actin monomer addition to barbed ends.

Capping Proteins

Capping protein is a heterodimer of 30-kDa subunits that binds filament barbed ends, preventing monomer addition (Figure 3). With barbed ends capped, the critical concentration of actin shifts to 0.7 μM , the critical concentration of the pointed end. Capping protein's high affinity ($K_d < 1 \text{ nM}$) and slow off-rate (half-life of capped barbed end is 30 min) cause a stable block to barbed-end elongation. Members of the gelsolin family also cap barbed ends, in addition to severing filaments.

ADF/Cofilin

ADF/cofilin is a 15-kDa protein that binds both filaments and monomers, with a 40-fold preference for the ADP-bound state over ATP- or ADP-Pi states in both cases. Mammals contain two isoforms, ADF and cofilin. ADF/cofilin binding to filaments is cooperative and accelerates Pi release from other subunits on the filament at least 20-fold. ADF/cofilin also severs filaments. ADF/cofilin is often referred to as a depolymerization factor, because its binding to filaments accelerates depolymerization. The mechanism of depolymerization acceleration probably involves both acceleration of Pi release (increasing subunit off-rate from filaments) and severing (creating more ends for depolymerization). However, ADF/cofilin does not prevent polymerization, as ADF/cofilin-bound monomers can still add to filaments.

Cellular Aspects of Actin Assembly/Disassembly

In mammalian cells, actin concentration is in the 50–200 μM range. Since the critical concentration is 0.1 μM , one would expect the vast majority of actin to be in filaments. However, the ratio of polymerized:unpolymerized actin is often around 1:1 in unactivated cells, while activated cells can reach close to

100% polymerized actin. The 1:1 ratio is maintained largely by the proteins discussed. Many mammalian cells contain more thymosin than actin, as well as a high concentration of profilin (e.g., neutrophils have $\sim 200 \mu\text{M}$ actin, 400 μM thymosin, and 100 μM profilin). Despite its lower concentration, profilin binds ATP-actin more tightly than thymosin (K_d of 0.2 μM vs. $> 10 \mu\text{M}$), so that monomers distribute between the two proteins. The rapid off-rates of monomers from these proteins (at or above five per second) allow rapid equilibration. Almost all monomers are bound to one protein or the other, leaving $\sim 0.5 \mu\text{M}$ free actin monomer in cells. This low free monomer concentration virtually eliminates spontaneous nucleation in cytoplasm. Another consequence of these two proteins is that pointed-end elongation is effectively suppressed, since thymosin prevents elongation from both ends and profilin prevents pointed-end elongation. For this reason, cellular filaments only elongate from barbed ends (except under specific conditions in muscle cells).

Cellular actin filaments are assembled by specific nucleation factors. Examples of such factors are Arp2/3 complex and formins. Both nucleation factors are tightly regulated. Thus, cells control the when and where of filament formation.

Mammalian cells contain $\sim 1 \mu\text{M}$ capping protein, and its high affinity for barbed ends means that new filaments are capped rapidly ($\sim 1 \text{ s}$ after formation). Capping protein is inhibited by bound polyphosphoinositides, which can promote cellular filament elongation. Some proteins can also inhibit capping protein by competing for binding, specifically VASP and formins. By controlling capping, cells control filament elongation.

The mechanisms by which cellular filaments disassemble are not understood fully, since off-rates from pointed ends cannot explain the rapidity of disassembly in many circumstances. Filament networks hundreds of nanometers thick can disassemble in seconds. ADF/cofilin clearly has a central role in this process, probably both by catalyzing Pi release from filaments and by severing filaments to create new ends for depolymerization. There is, however, evidence that ADF/cofilin may contribute to filament assembly upon initiation of cell motility, perhaps by severing to create new ends capable of elongating and supporting Arp2/3 complex-based nucleation.

The coordinated action of these proteins results in a cycle of assembly/disassembly in cells (Figure 4). Monomers shift from thymosin to profilin, then add to the barbed end. The filament-bound monomer hydrolyzes its ATP and retains ADP and Pi. ADF/cofilin binding to the filament accelerates Pi release and causes filament severing. Monomers dissociate from the pointed end and remain bound to ADF/cofilin. Monomeric ADP-actin exchanges between ADF/cofilin, thymosin, and profilin. Even though ADF/cofilin binds more tightly than profilin to ADP-actin, profilin is at higher concentration (100 vs. $\sim 20 \mu\text{M}$), and the rapid off-rate from ADF/cofilin enables equilibration. Profilin catalyzes recharging of ATP onto its bound monomer, enabling re-addition of that monomer to the barbed end. Capping protein terminates barbed-end elongation except under specific circumstances.

The abilities of some proteins, such as formins and VASP, to bind at or near the barbed end have interesting consequences on actin dynamics. These proteins can compete with capping protein for barbed-end binding. In the absence of these

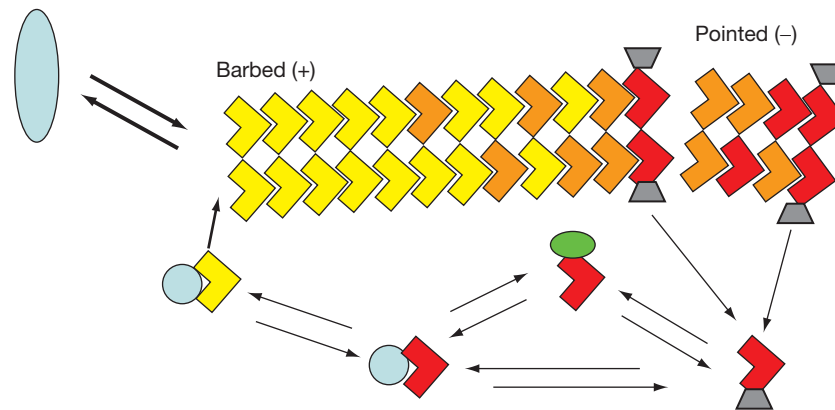


Figure 4 Concerted effects of actin-binding proteins in the assembly/disassembly cycle. Profilin-bound ATP-actin adds to the filament barbed end, then hydrolyzes its ATP slowly. ADF/cofilin binding accelerates Pi release along the filament and severs the filament. ADF/cofilin-bound ADP-subunits release from filament ends. Profilin, thymosin, and ADF/cofilin compete for binding ADP-monomers. Profilin-bound ADP-monomers exchange ADP for ATP, enhancing their ability to assemble onto barbed ends. Actin colors: yellow = ATP, orange = ADP-Pi, red = ADP.

proteins, capping protein terminates barbed-end elongation within seconds of filament nucleation, resulting in very short filaments. Barbed-end binding by formins or VASP, both of which generally allow barbed-end elongation, enables filaments to continue elongating. Capping protein inhibition is important for assembly of longer actin filaments, such as those in filopodia or microvilli.

The above description is probably accurate for most eukaryotes, including mammals, other vertebrates and multicellular animals, and protozoa. Fungi such as *Saccharomyces cerevisiae* (budding yeast) are somewhat different, containing lower actin and profilin concentrations (5–10 μM for each) and no thymosin. A yeast-sequestering protein, twinfilin, is present at unknown concentration. Unlike in vertebrates, most yeast actin (>90%) is polymerized. Yeasts contain ADF/cofilin, capping protein, Arp2/3 complex, and formins.

Actin assembly/disassembly in plants is poorly understood. Plants contain all the discussed proteins except thymosin. Plant profilin does not catalyze nucleotide exchange.

Other proteins, less well characterized at present, may play significant roles in actin assembly/disassembly. The Srv2/CAP complex may accelerate ADF/cofilin-mediated filament depolymerization and monomer recycling. Aip1 may associate with ADF/cofilin and cap barbed ends.

Reagents Used in Actin Research

Small Molecules Affecting Actin Dynamics

Table 2 presents a catalog of the reagents commonly used in actin research. While this table is simplistic, the effects of some reagents are not, and care must be taken to avoid artifacts. An example is the cytochalasins, a group of compounds commonly used to depolymerize cellular actin. This effect is attributed to the barbed-end-capping activity of cytochalasins. With barbed ends capped, filaments depolymerize from pointed ends since profilin and thymosin prohibit pointed-end elongation. However, at higher concentrations, cytochalasins can have other effects, such as causing ATP hydrolysis on actin monomers through nonproductive dimer formation. Use of

Table 2 Small molecule reagents used in actin research

Reagent	Biochemical effect	Cellular effect	Cell permeant?
Latrunculin A	Sequesters monomer	Depolymerization	Yes
Cytochalasins	Cap barbed ends	Depolymerization	Yes
Swinholide A	Sequesters dimers, severs filaments	Depolymerization	Yes
Phalloidin	Binds filaments	Prevent depolymerization	No
Jasplakinolide	Binds filaments	Prevent depolymerization	Yes

concentrations below 1 μM largely suppresses this secondary effect. In addition, cytochalasin B binds and inhibits a plasma membrane glucose transporter. Despite this property, cytochalasin B is sometimes used in preference to cytochalasin D, because the 10-fold lower affinity of cytochalasin B for barbed ends enables more rapid wash-out during cellular recovery experiments.

Latrunculins (isomers A and B) can cause a similar cellular effect to cytochalasins (depolymerizing filaments), but by a different mechanism (sequestering monomers). In neither case do these compounds actively depolymerize filaments. What they do is block polymerization, leaving only depolymerization to occur by the normal cellular mechanisms.

Swinholide A binds/severs filaments, and binds/sequesters dimers, again causing cellular actin depolymerization. This compound has not been widely used in actin cell biology.

Two compounds – phalloidin and jasplakinolide – stabilize filaments. Phalloidin is much better characterized. In the presence of phalloidin, the off-rate from barbed and pointed ends drops to near zero, and the critical concentration essentially becomes zero. Phalloidin also slows Pi release significantly. Phalloidin can be labeled with many compounds, including fluorophores, enabling localization of actin filaments in permeabilized cells. Isothiocyanate derivative of rhodamine

(TRITC) – phalloidin is especially useful, since its fluorescence increases 20-fold upon binding filaments. Since phalloidin is not cell permeable, jasplakinolide is used for live cell studies. For either compound, concentrations equal to those of actin are necessary to block depolymerization fully, because they bind stoichiometrically along the filament.

See also: **Cell Architecture and Function:** Actin-Capping and -Severing Proteins; Actin-Related Proteins; Cell Migration; Focal Adhesions and Related Integrin Contacts; **Signaling:** Phosphatidylinositol Bisphosphate and Trisphosphate.

Further Reading

- Amos LA, van den Ent F, and Lowe J (2004) Structural/functional homology between the bacterial and eukaryotic cytoskeletons. *Current Opinion in Cell Biology* 16: 24–31.
- Carlier MF and Pantaloni D (1997) Control of actin dynamics in cell motility. *Journal of Molecular Biology* 269: 459–467.
- Condeelis J (2001) How is actin polymerization nucleated *in vivo*? *Trends Cell Biology* 11: 288–293.
- Cooper JA and Schafer DA (2000) Control of actin assembly and disassembly at filament ends. *Current Opinion in Cell Biology* 12: 97–103.
- Goddette DW and Frieden C (1986) Actin polymerization. The mechanism of action of cytochalasin D. *Journal of Biological Chemistry* 261: 15974–15980.
- Kreis T and Vale R (eds.) (1999) *Guidebook to Cytoskeletal and Motor Proteins*, 2nd edn. New York: Oxford University Press.
- Kuhn JR and Pollard TD (2005) Real-time measurements of actin filament polymerization by total internal reflection fluorescence microscopy. *Biophysics Journal* 88: 1387–1402.
- MacLean-Fletcher S and Pollard TD (1980) Identification of a factor in conventional muscle actin preparations which inhibits actin filament self-association. *Biochemical and Biophysical Research Communications* 96: 18–27.
- Pollard TD (1986) Rate constants for the reaction of ATP- and ADP-actin with the ends of actin filaments. *Journal of Cell Biology* 103: 2747–2754.
- Pollard TD (1984) Purification of ADP-actin. *Journal of Cell Biology* 99: 769–777.
- Pollard TD, Blanchoin L, and Mullins RD (2000) Molecular mechanisms controlling actin filament dynamics in nonmuscle cells. *Annual Review of Biophysics and Biomolecular Structure* 29: 545–576.
- Spudich JA and Watt S (1971) The regulation of rabbit skeletal muscle contraction. *Journal of Biological Chemistry* 246: 4866–4871.
- Straub FB and Feuer G (1950) Adenosine triphosphate. The functional group of actin. *Biochimica et Biophysica Acta* 4: 455–470.

Actin-Capping and -Severing Proteins

S Maiti, IISER Kolkata, Mohanpur, India

J R Bamburg, Colorado State University, Fort Collins, CO, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Actin depolymerizing factor (ADF)/cofilin Related actin-binding proteins that increase the turnover of actin filaments *in vivo* through severing. The ADF-H domain is used to build larger versions of related proteins that can sever and cap filaments.

Arp2/3 complex A complex of two actin-related proteins, Arp2 and Arp3, with five other proteins that can nucleate the growth of new actin filaments and cause branching of filaments at the leading edge of a migrating cell.

Enabled/vasodilator-stimulated phosphoprotein (Ena/VASP) A family of actin-binding proteins that bind to but do not cap the growing barbed end of actin filaments, reducing branching and aiding in growth of long filaments.

F-actin The dynamic filamentous cytoskeletal structures, composed of subunits of actin monomer (G-actin), which organize the cytoplasm to maintain the shape of eukaryotic cells.

Filament capping The binding of a protein at the end of an actin filament to stop filament growth or depolymerization.

Filament severing The process whereby a protein binds along the side of an actin filament and severs the filament into two smaller pieces.

Filament treadmilling The dynamic process in which actin filaments elongate at their barbed (plus) end by addition of subunits and at the same time disassemble at their pointed (minus) end with loss of subunits.

Filopodia Finger-like protrusions with sensory capability that extend from the cell membrane and contain a bundle of parallel cross-linked actin filaments with uniform polarity.

Lamellipodia Flat web-shaped membrane protrusions containing highly branched actin filaments, seen at the leading edge of many migrating cells.

Wiscott–Aldrich syndrome protein (WASP) A family of proteins involved in activation of the Arp2/3 complex and actin-filament nucleation.

Overview

Actin is one of the major components of the cytoskeleton, a network of protein filaments that organizes the cytoplasm of eukaryotic cells. Actin filaments (F-actin), in the form of a right-handed double-stranded helix, are formed by assembly of globular (G-actin) monomer subunits in a head-to-tail orientation that gives the filaments a molecular polarity. Based on the orientation of an arrowhead pattern on F-actin decorated with a fragment of the muscle motor protein myosin, one end of the filament is called the barbed end and the other the pointed end. Spatial and temporal regulation of actin-filament assembly, disassembly, and stability in cells drives most aspects of cell motility, cell shape, and the interaction of a cell with its neighbors or surrounding matrix. These multifaceted cellular functions of actin are tightly regulated by more than 100 actin-binding proteins (ABPs) that interact with either G-actin or F-actin. Actin-capping and -severing proteins are special classes of ABP that regulate the assembly dynamics of actin filaments. Actin-capping proteins are classified into two groups, the barbed-end-capping proteins and the pointed-end-capping proteins. Depending on their binding affinity and off-rate constants, capping proteins slow down or stop the growth and/or depolymerization of filaments at the end to which they bind. Their interaction with actin may promote nucleation of new filaments, increase branching of filaments, and/or enhance filament depolymerization. Actin-severing proteins bind to the side of actin filaments and enhance fragmentation. Some severing proteins also become barbed-end-capping proteins (e.g., gelsolin), whereas others (e.g., actin depolymerizing

factor (ADF)/cofilin) can provide both free barbed and pointed ends that can influence actin dynamics.

Actin

Actin is a major cytoskeletal protein virtually in all eukaryotic cells, often comprising 2–5% of total cellular protein (total cellular concentration 50–200 μM). In muscle cells where it was first identified, F-actin is the major component of the thin filaments of the contractile apparatus where the myosin motor head groups of the thick filament walk along the actin thin filaments in an energy-dependent process, causing contraction. Mammals have six actin genes, each expressing a different isoform, but the structure of the 42.5-kDa actin monomer is highly conserved across phyla. G-actin has four quasi-subdomains, at the center of which is bound an essential adenine nucleotide, either adenosine triphosphate (ATP) or adenosine diphosphate (ADP). As is the case with most molecular self-assembly systems, actin assembly is driven by a positive entropy change that arises from the displacement of water as the subunits come together. Following assembly, bound ATP is rapidly hydrolyzed to ADP and inorganic phosphate (P_i), whose dissociation to yield ADP-actin is about 100 times slower than hydrolysis. Nucleotide hydrolysis provides the energy input required for actin filaments to maintain equilibrium-binding constants for monomers that differ at the barbed and pointed ends. The concentration of monomer required to maintain the filament end is called the critical concentration, and is lower for the barbed end than for the pointed end. Thus, as monomer

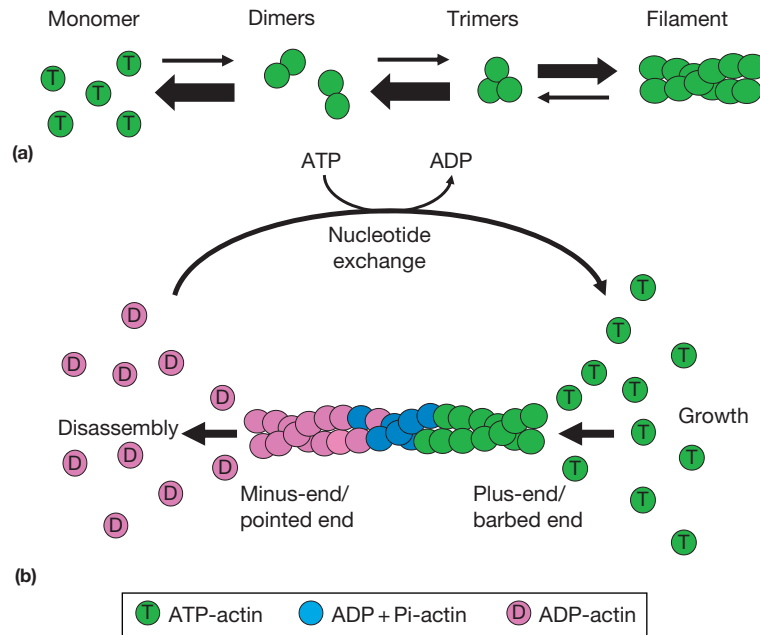


Figure 1 (a) Slow nucleation of actin-filament formation *in vitro* results from the instability of actin dimers and trimers. Spontaneous nucleation of actin from monomer does not occur *in vivo* due to sequestering of monomer. When actin monomer is plentiful, filaments can elongate from both ends, but grow faster from the barbed end. (b) Dynamics of actin filaments at steady state demonstrating treadmilling. ATP-actin adds onto the barbed end, nucleotide hydrolysis occurs rapidly, and the inorganic phosphate dissociates slowly leaving ADP-actin subunits to dissociate from the pointed end. Nucleotide exchange of ATP for ADP occurs on the monomer, giving rise to more ATP-actin for filament growth.

assembles into filaments, the monomer concentration decreases and the critical concentration of the pointed end is reached first. Monomers still add to the barbed end, leading to monomer loss from the pointed end. This phenomenon is called 'treadmilling'. At steady state, there is consumption of energy through ATP hydrolysis that drives the treadmilling of the actin subunits (Figure 1). In some cells this may account for up to 50% of the total ATP consumption. Tropomyosin (Tm), a short helical filamentous ABP, binds along the side of actin filaments and covers five to seven actin subunits. Binding of Tm usually aids in the formation of stable actin filaments, in part because some Tms can inhibit the severing and depolymerizing activity of other ABPs. However, there are more than 40 different isoforms of Tms formed by alternative splicing of the four Tm genes, and some of these might enhance actin dynamics by maintaining filaments in a conformation that enhances the binding of severing proteins. Dimers of actin are not stable and tend to dissociate, but, with the addition of another monomer, the trimeric actin acts as a nucleus for filament formation (Figure 1). In the cell, spontaneous nucleation of filaments is kept at minimum by the presence of actin monomer sequestering proteins. Thus, nucleation is a highly regulated event and may be initiated in the cell by three mechanisms: (1) the activation of certain actin-binding proteins that help form trimeric nuclei; (2) the removal of capping proteins from the filament barbed ends; and (3) the severing of existing filaments to provide new barbed ends for filament elongation. The sites of nucleation and directionality of actin-filament growth are important for the development of cell polarity and cell migration and are regulated, in part, by the localized activity of capping and severing proteins.

Actin Dynamics at the Leading Edge

Various cell membrane protrusions and inversions, especially lamellipodia, filopodia (microspikes), phagocytic cups, and upward ruffles, are distinct morphologies of cells and are important for cell function, motility, and organismal development. These leading-edge phenomena are based on actin-filament reorganization. In extending lamellipodia and filopodia, polarized arrays of actin filaments turn over very rapidly, with filament barbed ends facing toward the plasma membrane. The pointed ends of the filaments rapidly depolymerize at the rear of these arrays. In extending lamellipodia, the actin-filament growth is regulated mainly by formin, enabled/vasodilator-stimulated phosphoprotein (Ena/VASP), capping protein, Arp2/3 complex, and actin-filament cross-linking proteins. The Arp2/3 complex causes branching of growing actin filaments, and the barbed-end-capping protein limits the growth of newly assembled filaments to maintain short, rigid filaments that can serve to push the membrane forward in the broad lamellipodium. Ena/VASP proteins and formins bind to the barbed ends of growing actin filaments and inhibit the binding of capping protein and the Arp2/3 complex, leading to the formation of longer and unbranched actin filaments, such as those found in the finger-like filopodial extensions (microspikes). This antagonism in barbed-end binding between formins and Ena/VASP on the one hand, or Arp2/3 complex and capping protein on the other, can help explain the often rapid conversion between lamellipodial and filopodial membrane protrusions seen at the leading edge of many migrating cells, especially the growing tip of neurons, called the growth cone (Figure 2). The remodeling of

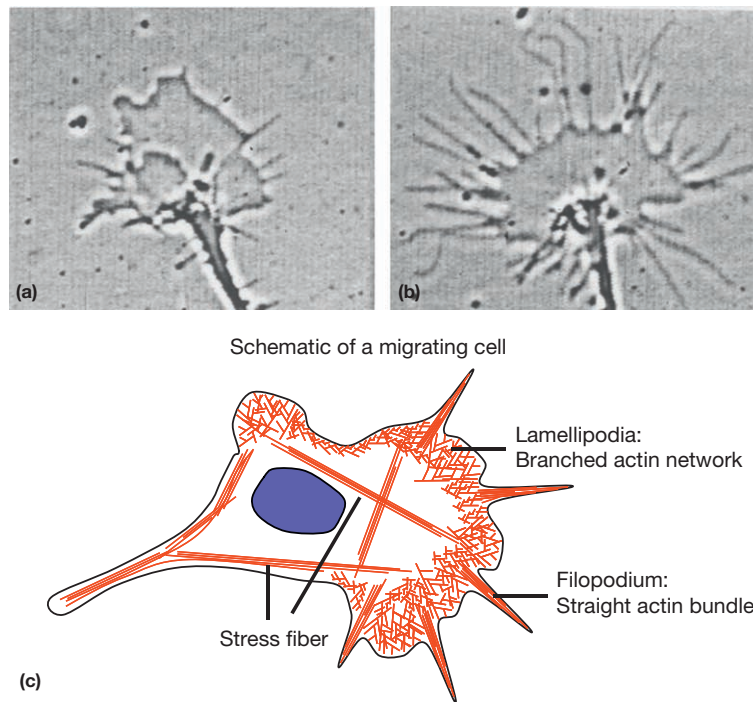


Figure 2 (a) A typical dorsal root ganglion neuronal growth cone. Phase-contrast picture on left shows growth cone exhibiting lamellipodial growth. (b) Within a few minutes after reducing the osmolarity of the growth medium, the growth cone exhibits many long and active filopodia. Panel width = 44 μm . (c) Schematic of a migrating cell showing filopodia with straight long actin-filament bundles, lamellipodia with a branched actin network, and the cell body with long actin-filament bundles linked to sites of adhesion. The bundles are referred to as stress fibers in nonmotile cells but as graded polarity actin bundles in highly migratory cells because of the changes in orientation of the filaments within the bundles in going from the front to the rear of the cell.

the actin cytoskeleton for this conversion also requires actin-severing proteins, such as ADF/cofilin.

Pointed-End Binding and Branching

Arp2 and Arp3 are two actin-related proteins that form part of a seven-subunit complex known as the Arp2/3 complex, ubiquitous in eukaryotes. The Arp2/3 complex is required for generating the 70° branched actin filaments observed at the leading edge of expanding cell membranes in lamellipodia. There are at least two mechanisms for generating this type of filament array. The Arp2/3 complex may capture the pointed ends of actin filaments created by severing pre-existing filaments and then associate with the side of another pre-existing filament to create a branch point. Alternatively, Arp2/3 complex may bind at the side of a growing filament to create a branch point. The Arp2/3 complex requires activation to become a potent nucleator of actin-filament growth. Activation can occur through the binding of members of the Wiskott-Aldrich syndrome protein (WASP) family, some of which are activated by signaling through G-protein-coupled receptors and others by Cdc42, a Rho family GTPase. The Arp2/3 complex can also be activated by other mechanisms. The active Arp2/3 complex can bind an actin monomer and initiate the nucleation of a new filament at an angle of ~70° to the mother filament. This branched actin-filament network, called the dendritic brush, generates the force required for cell protrusion at the leading edge (Figure 3).

To maintain proper stiffness and the best angle for membrane expansion, the filaments are not allowed to grow very long before their barbed ends are capped by capping protein. The branching of filaments is inhibited by the binding of at least some forms of Tm, which prevent Arp2/3 binding, and branching is also inhibited by proteins in the Ena/VASP family.

Evolutionary and Structural Relationships between Some Severing and Barbed-End-Capping Proteins

The present sequence, structural, and biochemical data suggest that many of the actin-binding proteins have evolved through gene duplication or mutation of DNA sequences encoding a small number of protein motifs. For example, proteins in the ubiquitous eukaryotic ADF/cofilin family are derived from a single 15-kDa module (the ADF-H domain), can interact with both G- and F-actin, and have F-actin-severing activity but do not cap filaments. Proteins containing one, two, three, and six copies of the ADF-H domain have evolved for regulating actin dynamics and will be discussed separately later. Although ADF/cofilins and some of the ADF-H domains within the six-ADF-H-domain family members show little or no sequence homology, structurally they display similarity on overall folding and secondary/tertiary structural elements, suggesting that both divergent and convergent evolutionary mechanisms have given rise to these structurally related molecules. Another common actin-binding motif is found in proteins containing the

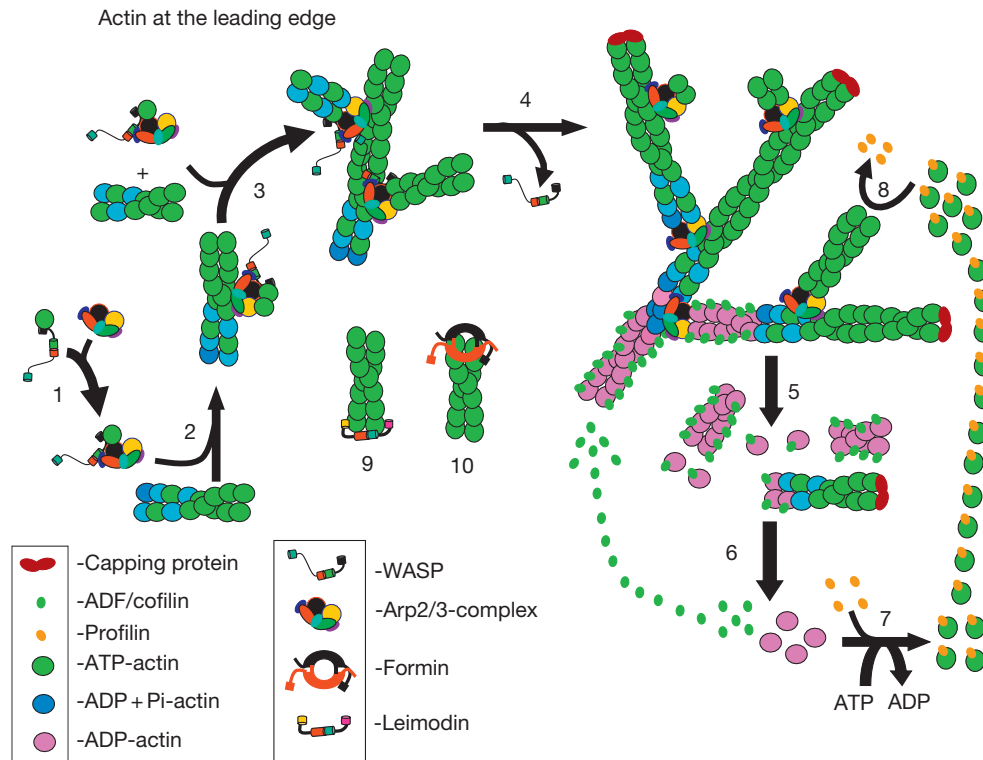


Figure 3 Schematic of actin-filament dynamics at the leading edge of a cell. (1) Activation of the Arp2/3 complex by WASP can cause *de novo* nucleation of filaments, but active Arp2/3 complex may also be captured by pointed ends of existing filaments (3). (2–4) Arp2/3 complex nucleates from the side of actin filaments or, when bound to pointed end of actin filament, is captured along the side of a filament to create the 70° branches. Further growth of barbed ends with rapid capping and nucleation of new branches leads to complex networks of actin filaments (4). (5) ADF/cofilin-mediated severing and depolymerization of the filament arrays occurs when subunits become ADP-actin. Release of ADF/cofilin (6) and enhanced nucleotide exchange by profilin and Srv2/CAP (7) provide profilin-actin complex that serves as a monomer pool for filament growth (8), especially during formin-mediated elongation (10). Leimodin (also cordonbleu and others) nucleate actin filaments, allowing growth at the barbed end. They remain associated with the pointed as a pointed-end-capping protein (9), whereas formin nucleates and moves with the growing barbed end to form straight actin filaments (10).

WASP homology domain 2 (WH2), which evolved as a versatile, short (17–27 amino acid), flexible α -helical motif that can bind actin by a hydrophobic interaction. Proteins with this motif have many diverse functions with regard to actin, including nucleation, capping, severing, and monomer sequestering activity.

Proteins with ADF-H Domains

The Single-Domain ADF/Cofilin Family

Cell motility and morphogenesis require high rates of actin-filament turnover, ~100–150-fold faster than occurs for purified actin *in vitro*. The ADF/cofilins are 13–19-kDa proteins containing a single ADF-H domain that are required for this high rate of actin-filament dynamics. They have been isolated or identified by genomic or complementary DNA (cDNA) analysis from more than 40 different animals, higher plants, fungi, yeasts, and protists. The lethality of the loss of the ADF/cofilin gene in organisms that express only a single gene for this family (e.g., yeast) demonstrates the essential nature of these proteins. Mammals have three separate genes for these proteins: the ubiquitous cofilin 1, ADF, and cofilin-2, which are found mostly in skeletal muscle and also in low amounts in other

tissues. Within any given mammal, the three proteins are ~70% identical. In an ADF-null mouse, cofilin 1 expression is up-regulated to compensate for the loss of ADF and the animals are viable. The only major defect observed is that the mice become blind due to corneal thickening, even though cofilin 1 is up-regulated in the cornea. This finding, along with differences in the quantitative behavior of the three proteins *in vitro*, suggests that each isoform has some unique tissue-specific functions. Many ADF/cofilins have nuclear localization sequences, but the protein is usually cytosolic. However, it can be targeted to the nucleus, along with cytoplasmic actin, under certain conditions. Its nuclear transport may simply be a mechanism for uptake of actin, which has many nuclear functions in chromatin remodeling and transcription. However, in plants, nuclear ADF has a transcriptional role.

ADF/cofilins enhance filament turnover (dynamize filaments) by severing filaments in a concentration-dependent manner. At low ratios with F-actin subunits, cofilin has been shown to continuously sever filaments. At higher ratios, cofilin binds cooperatively along the filament, stabilizing a twisted form of the F-actin, which leads to rapid severing and also stabilization of the shorter cofilin-saturated pieces of F-actin. If there is a low concentration of available actin monomer, severing of F-actin provides more pointed ends from which

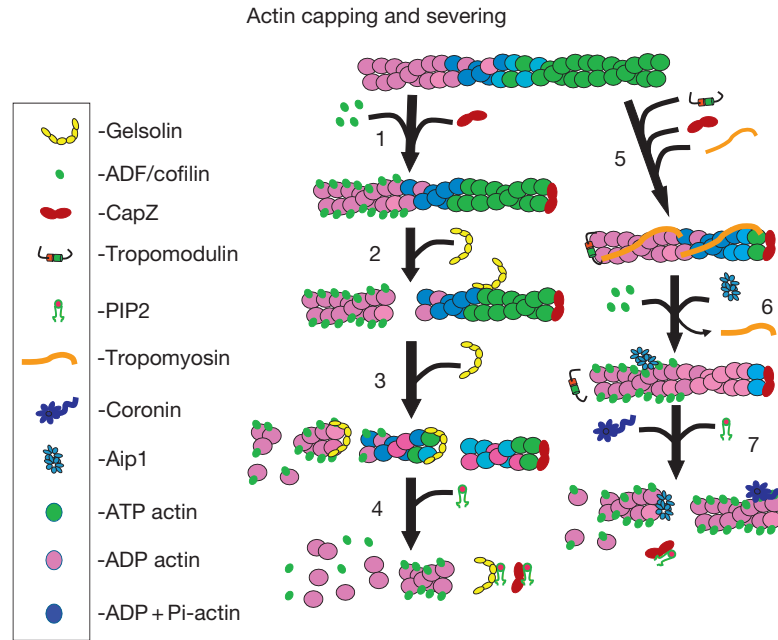


Figure 4 Diagrams showing actin-filament capping and severing by different types of molecules. ADF/cofilin: (1) binds cooperatively and with highest affinity to ADP-actin subunits and (2) severs actin filaments. (5) Stable filaments are generated by tropomodulin binding to pointed ends, CapZ binding to barbed ends, and tropomyosin binding along filament length. ADF/cofilin competes for F-actin binding with some tropomyosins (6). Cooperative activity between ADF/cofilin and Aip1 can lead to faster severing of tropomyosin-decorated filament. Coronin also aids ADF/cofilin to sever ADP-actin filaments (7). (4, 7) PIP₂ regulates the binding of many capping proteins, dissociating both gelsolin and CapZ from barbed ends of actin filaments. (2, 3) Actin-filament capping and severing by gelsolin.

dissociation of subunits can occur leading to more rapid filament depolymerization. If the environment contains assembly-competent actin, the severed filaments can promote filament nucleation and elongation. ADF/cofilins bind ADP-actin subunits or monomer with higher affinity than they bind to ATP-actin. Thus, within an actin filament, ADF/cofilin binding is preferentially toward the pointed end, the desired region for filament turnover in cells (Figures 3 and 4). However, cofilin has a stronger nucleating activity than ADF, resulting in ADF having stronger monomer sequestering activity *in vivo*. If F-actin is severed by ADF/cofilin at the leading edge of cells where active Arp2/3 complex is present, the Arp2/3 complex can capture the pointed ends of the newly severed filaments and bind them along the side of existing filaments to create the highly branched network found in lamellipodia.

ADF/cofilin activity is regulated by pH-dependent phosphatidylinositol 4,5-bisphosphate (PIP₂) binding as well as by pH effects on the ADF/cofilin-actin filaments resulting in greater depolymerization at pH > 7.3, whereas pH < 7.1 favors F-actin binding. ADF/cofilin-mediated F-actin severing is inhibited by some of the larger isoforms of tropomyosin (e.g., TM5_{NM1}), which can stabilize F-actin in a conformation that is not conducive to ADF/cofilin binding. However, there may be one or more isoforms of tropomyosin (e.g., TMBr3) that can enhance the dynamizing activity of ADF/cofilin. Actin binding, and thus ADF/cofilin activity, is inhibited by phosphorylation at a highly conserved serine residue (the encoded Ser 3 in mammalian ADF and cofilin, which is Ser 2 in the demethionated and N-acetylated processed proteins). Two related kinases, the LIM (LIM domain first discovered in proteins Lin11, Isl-1 & Mec-3) kinases and TES (named for testes in which they were first discovered) kinases, each with two members, phosphorylate ADF/cofilins at

this site. The regulated removal of the phosphate from Ser3 (activation) is accomplished by specific phosphatases in the slingshot or chronophin families, although more general phosphatases may sometimes play a role. Both the kinase and phosphatase pathways are regulated by upstream signaling events through which extracellular signals affect actin dynamics. Although phospho-cofilin is inactive in binding either G- or F-actin, it is an activator of phospholipase D1, an important protein in generating membrane phosphatidic acid, a requirement in some cells for the establishment of polarized migration through interaction with the dedicator of cytokinesis (DOCK) family of proteins, which are guanine nucleotide exchange factors for generating the active form of the small GTPase Rac1. Because Rac1 is an activator of the p21 kinase (Pak), which activates LIM kinase to phosphorylate cofilin, cofilin is both an upstream activator and a downstream effector of Rac1 signaling.

Mammalian cofilin has four free sulfhydryl groups. The actin-binding activity of cofilin is altered by the oxidation of these cysteine residues to form either one or two internal disulfides. Cofilin with a single internal disulfide formed between Cys39 and Cys80 is still able to bind F-actin but does not enhance dynamics of the actin, whereas cofilin with two disulfides (Cys39–Cys80 and Cys139–Cys147) does not bind to actin but rather is targeted to mitochondria where it promotes mitochondrial-dependent apoptosis through cytochrome *c* release.

The Two-ADF-H-Domain Twinfilin Family

Twinfilin (Twf), a 40-kDa protein with two ADF-H domains connected by a small linker region, is a multifunctional actin-binding protein, also found across the phyla. Mammals have

two Twf isoforms, whereas only one isoform occurs from yeast to flies. Twf binds tightly with ADP actin and this activity is regulated by PIP₂. Yeast Twf at neutral pH sequesters actin monomer to prevent spontaneous actin-filament assembly, but under more acidic conditions (pH < 6.0) it can also sever actin filaments. In budding yeast, Twf binds to capping protein, which is necessary for Twf to localize to the cortical actin cytoskeleton. Mouse Twf1 caps filament barbed ends, but with highest affinity only to those made up of ADP actin, such as one finds on severed filaments.

The Three-ADF-H-Domain Proteins: CapG, Severin, and Fragmin

CapG, severin, and fragmin are gelsolin family proteins containing three ADF-H segments that have Ca²⁺-dependent barbed-end-capping and -severing activity. The first segment (S1) binds to G-actin and has the capping function, whereas segments 2 and 3 (S2, S3) bind to F-actin. CapG is a 39-kDa barbed-end-capping protein, the name reflecting its capping function and structural similarity to gelsolin, though only half of its size. CapG is highly expressed in macrophages and was initially known as macrophage capping protein, gCap39 or Mbh1. CapG has a structure similar to severin and fragmin, but it does not sever actin filaments as do severin and fragmin. CapG caps barbed ends of actin filaments at micromolar concentrations of Ca²⁺, but it dissociates from barbed ends when Ca²⁺ levels return to their normal nanomolar levels, whereas gelsolin remains bound to the barbed ends under the same conditions. CapG is important for the actin-based motility of macrophages. PIP₂ binds to CapG, thereby inhibiting its binding to F-actin. The actin monomer sequestering protein profilin is also able to dissociate CapG from actin-filament barbed ends when profilin is present at high concentration (>10 μM).

CapG is also found within the nucleus as well as the cytoplasm, but its nuclear function is unknown. Although CapG lacks a nuclear export signal, which is present in severin and fragmin, CapG does not contain any known nuclear localization signal. Thus, the method by which it is translocated into the nucleus is also unknown. CapG is also found associated with the mother centriole in interphase cells, with the mitotic spindle during mitosis, and in the midbody ring during abscission. Localization to these structures suggests that CapG may work as a mediator between the actin cytoskeleton and microtubules, and may have a greater role in motile processes than previously appreciated. Severin and fragmin, found in the slime molds *Dictyostelium* and *Physarum*, respectively, are structurally similar to CapG, but they also bind to the side of and sever F-actin as well as cap the barbed ends of the severed filaments. They have two Ca²⁺-binding sites and can be thought of as small versions of the six-ADF-H-domain protein, gelsolin.

The Six-ADF-H-Domain Proteins: Gelsolin, Villin, and Adseverin/Scinderin

Gelsolin, villin, and adseverin/scinderin contain six ADF-H repeats and presumably arose through duplication of the gene for the ADF-H three domain proteins. These proteins, often referred to as the gelsolin family, also have F-actin barbed-end-capping and -severing activity. The N-terminal

half of these proteins, with three repeats similar to CapG, severin, and fragmin, is able to bind two actin monomers and sever actin filaments. The C-terminal half is able to bind to a single actin monomer at elevated Ca²⁺ concentration.

Gelsolin, with a mass of 80 kDa, is expressed in all eukaryotes. The name arose from its ability to rapidly dissolve an actin gel and convert it into a soluble form. The N-terminal half of gelsolin contains two Ca²⁺-binding sites, one of which is high affinity, in the submicromolar range. Gelsolin severs actin filaments at Ca²⁺ concentrations >10⁻⁵ M. The severing mechanism probably involves the binding of gelsolin along the side of the filament and intercalation between actin subunits during the normal thermal-induced flexing of the actin. After severing filaments, gelsolin remains bound to the barbed end of one new filament (Figure 4) and produces a new pointed end. At Ca²⁺ concentrations in the submicromolar range, gelsolin acts to cap barbed ends of actin filaments. An extracellular (secreted) isoform of gelsolin with molecular mass 83 kDa, called serum gelsolin, circulates in blood where it severs and caps actin filaments released from lysed cells. The resulting monomers of actin are then sequestered by vitamin D-binding protein, also a potent actin monomer-binding protein, and ultimately removed from circulation by liver. This mechanism prevents accumulation of actin filaments in the blood, which could potentially block circulation in capillaries.

Gelsolin binding to F-actin is inhibited by tropomyosin in two ways. Tropomyosin binding to F-actin induces a cooperative conformational change in actin filaments resulting in dissociation of gelsolin. Second, tropomyosin can directly bind gelsolin and thereby dissociate gelsolin from actin-gelsolin complex. Dissociation of gelsolin by tropomyosin from a gelsolin-capped actin filament leads to the formation of long filaments from smaller gelsolin-actin complexes, even from gelsolin-actin heterodimers, heterotrimers, and heterotetramers. PIP₂ can bind directly to gelsolin and dissociate gelsolin that is bound to the barbed end of an actin filament, yielding a free barbed end for growth. Gelsolin can be proteolytically cleaved by caspase-3 during apoptosis. The short amino-terminal (residues 1–352) cleavage fragment does not need Ca²⁺ for severing existing actin filaments and contributes to the progression of apoptosis.

Villin, a gelsolin family member with molecular mass of 92.5 kDa, controls actin assembly in the microvilli of epithelial cells. Villin has the unique property of being able to bundle actin filaments through an 8-kDa F-actin-binding module known as the villin headpiece. The activity of villin is highly regulated by Ca²⁺. At submicromolar Ca²⁺ concentrations, villin bundles actin filaments, whereas at micromolar Ca²⁺ concentrations it caps barbed ends of actin filaments and at >100 μM Ca²⁺ it severs F-actin. As with gelsolin, the activity of villin is regulated by both PIP₂ and Ca²⁺. In addition, villin is also regulated by phosphorylation on tyrosine residues in intestinal epithelial cells, a requirement for regulating the interaction of actin filaments with apical ion channels. Deletion of the villin gene in mice leads to accumulation of actin filaments within the epithelial cells making them prone to injury.

Adseverin, also known as Scinderin, is only expressed in secretory cells. It has slightly different temperature and Ca²⁺ sensitivity from gelsolin, but its functions appear to be very similar. During cell stimulation by Ca²⁺, adseverin is activated and disassembles cortical F-actin to allow secretory vesicle

exocytosis. As with gelsolin, adseverin binds to and is regulated by PIP₂.

Other Severing, Severing-Enhancing, or Dynamizing Proteins

Molecules Interacting with CasL

Molecules interacting with CasL (Micals) are multidomain cytosolic proteins found in both vertebrates and invertebrates that bind to the semaphorin 1a receptor, plexin A. Three vertebrate genes encoding Micals have been identified. Micals have a calponin homology (CH) domain in their N-terminal half that is involved in actin-filament binding, a polyproline-rich region that binds the SH3 domain of the focal adhesion docking protein CasL, and a C-terminal ezrin/radixin/moesin (ERM) domain needed for linking F-actin to plexin A. However, of particular interest is the N-terminal enzymatic flavoprotein monooxygenase domain that provides a redox-regulated depolymerization of actin filaments that will even depolymerize fascin-bundled F-actin in a reduced nicotinamide adenine dinucleotide phosphate (NADPH)-dependent manner. Micals regulate *Drosophila* bristle and neuronal growth cone morphology and mediate semaphorin-induced changes in cell shape. Micals are unique actin-severing proteins in that they are regulated directly through redox signaling.

Actin-Interacting Protein 1

Actin-interacting protein 1 (Aip1), an actin-filament-capping protein with a mass of ~65 kDa, has 8–14 repeats of a structural motif, often terminating in a tryptophan (W)-aspartic acid (D) dipeptide, called a WD domain. Each WD domain forms a four-stranded antiparallel β -sheet (or blade), six or seven of which comprise a β -propeller, a structure found in many different proteins with diverse functions. Aip1, a ubiquitous eukaryotic protein, binds to F-actin weakly or not at all in the absence of ADF/cofilin, but caps the barbed ends of filaments severed and saturated by ADF/cofilin. Aip1 inhibits annealing of the ADF/cofilin-severed filaments, maintaining a higher filament number, and blocks the growth of the filament at barbed ends. It also enhances a more complete depolymerization of ADF/cofilin-bound filaments to generate actin monomers. Importantly, coordinate activity of Aip1 and ADF/cofilin can sever tropomyosin-decorated actin filaments. In the nematode *Caenorhabditis elegans*, which expresses two forms of ADF/cofilin, the ubiquitous unc60A and the muscle-specific unc60B, only the weaker F-actin-binding/severing unc60B is able to form a high-affinity complex with Aip-1 (unc78). Thus, Aip1 enhances the apparent severing activity of the weaker ADF/cofilins and gives this family of proteins the ability to function much like the severing and capping proteins discussed earlier.

Coronins

Coronins (Crns), WD repeat-containing actin-binding proteins widely expressed in eukaryotes, are localized to the leading edge of migrating cells, a region of highly dynamic actin.

Crns are involved in cell motility, phagocytosis, cytokinesis, and endocytosis. Lower eukaryotes express only a single isoform of Crns, whereas higher eukaryotes express multiple isoforms (up to 12 in vertebrates). All Crns have WD repeat domains at their N-terminal region, forming a β -propeller structure that is involved in both F-actin binding and, for a few isoforms (Crn1 and Crn3), membrane binding. Crns have a variable middle region and a C-terminal domain with a coiled-coil structure important for oligomerization and F-actin binding. Crn1 can regulate (inhibit) the activity of the Arp2/3 complex by direct binding, the only known direct inhibitor of Arp2/3 complex *in vitro*. *In vivo*, deletion of Crn isoforms 1, 4, and 11 leads to decreased filament turnover. In budding yeast, Crn induces actin-filament depolymerization at a high rate, working in collaboration with ADF/cofilin and Aip1. Crn further blocks the weak cofilin activity on newly synthesized ATP-actin filaments, but synergizes with cofilin and Aip1 for disassembly of older filaments composed primarily of ADP-actin (Figure 4).

Formins

Formins are actin-filament nucleators that bind to the barbed end of actin filaments. All formins have signature domains called the formin homology domains 1 and 2 (FH1 and FH2), the latter of which dimerizes and nucleates actin filament (Figure 3). There are more than 15 different formins present in humans with diverse and versatile functions. Some formins, such as the diaphanous-related formins, mDia1 and mDia2, are autoinhibited through intramolecular interactions involving an N-terminal diaphanous inhibitory domain (DID) interacting with a C-terminal diaphanous autoinhibitory domain (DAD, which has sequence similarity to the WH2 domain). These proteins are usually activated by the binding of Rho GTPase at an N-terminal Rho-binding domain, releasing the autoinhibition. The diaphanous-related formins, including mDia1-3, Daam1-2, and Bni1, do not sever filaments but rather help elongate filaments by recruitment of profilin-actin complexes to the growing barbed ends. However, the formin-like inverted formin 2 (IFN2) and FRL1 have both actin-severing and -nucleating activities. INF2 and FRL1 have a putative DAD domain with WH2 similarity, which plays a critical role in filament disassembly.

Cyclase-Associated Protein

Although its abbreviation might suggest an actin-capping function, cyclase-associated protein (CAP; aka Srv2) actually does not interact with F-actin. It was named for its association with adenylyl cyclase and is found in all eukaryotes. Its major function is in recycling actin subunits from ADF/cofilin-depolymerized actin. When ADF/cofilin depolymerizes actin, the binding of ADF/cofilin to ADP-actin inhibits the nucleotide exchange. CAP forms a multimeric protein complex with ADF/cofilin, actin, and profilin to exchange ATP for ADP on actin and recycle ATP-actin. The C-terminal half has a WH2 domain and binds monomeric actin. It also has two proline-rich regions through which it binds profilin. The N-terminal half also has one monomer actin-binding site.

Capping Proteins

In general, capping proteins bind weakly to actin monomer but often bind strongly to dimer and higher oligomers (F-actin). Thus, pointed-end-capping proteins can stabilize the dimer and enhance nucleation of filament growth. They have high affinity for the filament ends to which they bind and they usually bind in a 1:1 stoichiometry with the filament (Figures 3 and 4). The main function of capping is to restrict polymerization and depolymerization, thereby controlling filament length. As mentioned earlier, limiting the growth of actin filaments at the leading edge of migrating cells is important for actin-assembly-driven membrane protrusion (lamellipodial growth). Pointed-end-capping proteins are important for filament nucleation, branching, regulating filament turnover, and cross-linking of actin filaments to other components of the cytoskeleton.

Barbed-End-Capping Proteins

CapZ

The most abundant barbed-end-capping protein is CapZ (Figures 3 and 4). In striated muscle cells, CapZ binds to actin filaments at the Z-disk, where CapZ also interacts with α -actinin, an actin cross-linking protein, to anchor the thin filaments from the neighboring sarcomere, the contractile unit of skeletal muscle. Nonsarcomeric isoforms of CapZ are localized to sites of membrane-actin contact. CapZ is a heterodimer composed of α - and β -subunits, 32 and 31 kDa, respectively. Each subunit has two isoforms. Both subunits are structurally similar, though they lack sequence similarity, but assemble tightly to form a dimer with pseudo-twofold rotational symmetry. The C-terminal ends of each subunit are flexible, amphipathic α -helices necessary for actin binding. This flexibility explains how CapZ can tightly bind two actin subunits, or the barbed end of a double-stranded actin filament, with subnanomolar affinity. Cells have CapZ in a near 1:1 molar ratio with actin filaments and 1:100 to 1:150 with respect to total actin subunits. *In vivo*, PIP₂ binds CapZ directly to inhibit its capping activity. The generation of PIP₂ in platelets in response to thrombin uncaps actin-filament barbed ends, allowing localized filament growth needed for the shape changes of platelets associated with blood clot retraction. V-1/myotrophin, CKIP1, CD2AP, CARMIL, and myosin1-linker bind directly to CapZ to inhibit its capping activity. VASP, a member of the Ena/VASP family of proteins, also functions as an antagonist for CapZ as well as for other barbed-end cappers, such as gelsolin. Formin proteins have high affinity for barbed ends of actin filaments and also act as competitors for CapZ binding, enabling filament growth.

Tensin

Tensin, a large multidomain focal adhesion molecule with a mass of ~200 kDa, interacts with actin and several signaling molecules through Src homology 2 (SH2) and phosphotyrosine-binding (PTB) domains. Tensin caps the barbed ends of actin filaments within the focal adhesions of fibroblasts and muscle, and links the actin cytoskeleton to the focal adhesion through multiple interactions with other

adhesion molecules. Tensin is very susceptible to proteolysis, especially by the calcium-activated calpain-II. Full-length recombinant tensin shows complete capping activity, whereas insertin, most likely a proteolytic fragment of tensin, only partially caps F-actin.

Pointed-End-Capping Proteins

Tropomodulin

Tropomodulin (Tmod) is a unique pointed-end-capping protein first discovered as an actin-tropomyosin-binding protein (Figure 4). The name tropomodulin was given because it binds actin filaments tightly in the presence of tropomyosin. Human skeletal muscle Tmod has a mass of about 40 kDa. Four isoforms of Tmod have been identified, one of which is found mainly in erythrocytes (Tmod1), one in brain (Tmod2), one in skeletal muscle (Tmod4), and one (Tmod 3) which is widely distributed. Tmod caps the pointed end of actin filaments and in skeletal muscle plays an important role in the assembly of the sarcomere where it is present in a 1:1 molar ratio with respect to the thin filaments. The N-terminal half of Tmod is rod shaped and binds to the N-terminal part of the two tropomyosins that lie along the opposite sides of actin in thin filaments. The C-terminal half of Tmod is responsible for capping the pointed end of F-actin. Although Tmods 1 and 4 bind to actin filaments tightly, they do not bind to actin monomer. However, Tmod 3 is able to sequester actin monomers. At low concentrations, Tmod3 increases actin polymerization by nucleating filaments, but at high concentrations it decreases actin-filament polymerization, both by capping the pointed end and by sequestering monomer. Binding affinity of all Tmods to the pointed end of F-actin is significantly higher in the presence of muscle tropomyosin. Tmods from nonmuscle cells bind with differing affinities to F-actin containing different isoforms of tropomyosin, suggesting that it might play a role in sorting Tm isoforms to different filament populations and in modulating filament turnover through regulating the dynamics at filament pointed ends.

Leimodin

A 64-kDa protein, named leimodin (Lmod), has three isoforms. Lmods 1 and 3 are found in smooth muscle, and Lmod 2 is found in cardiac and skeletal muscle. Like Tmod, Lmod also caps filament pointed ends and acts as a potent nucleator of actin-filament formation (Figure 3). Lmod and Tmod have a similar N-terminal actin-binding domain and a similar leucine-rich region containing another actin-binding domain. However, Lmod has only a single tropomyosin-binding site in the N-terminal region, whereas Tmod has two, and Lmod has an extended C-terminus with a polyproline region and an actin-binding WH2 domain that are not present in Tmod. The three actin-binding domains of Lmod are able to form stable trimeric actin nuclei and allow filament elongation at the barbed end while Lmod remains bound at the pointed end. Lmod recruits muscle tropomyosin to stabilize the filaments that become part of the sarcomere. Indeed, the recruitment of different tropomyosin isoforms to specific families of actin filaments may depend, in part, upon the nature of the filament nucleation complex.

Spectrin/band 4.1

Spectrin/band 4.1 acts as an F-actin pointed-end-capping complex. These proteins are expressed in a wide variety of tissues in multiple isoforms. Spectrin/band 4.1-actin complex was first isolated from red blood cell (erythrocyte) membranes, where actin filaments are short and of uniform length. Both ends of these filaments are capped, the pointed end with spectrin/band 4.1, but the nature of the barbed-end-capping protein is not known. Spectrin, a member of a large calponin homology repeat family of proteins, binds to some transmembrane proteins and/or their cytoplasmic-associated linkers (e.g., ankyrin) as well as to actin, forming a cortical cytoskeleton in erythrocytes that can restore cell shape after distortion by passage through small capillaries. Band 4.1 belongs to a superfamily of proteins having a FERM domain (FERM, four point one, ezrin, radixin, and moesin) and bind to several other proteins.

Other Nucleating/Capping Proteins

As mentioned earlier, proteins containing the WH2 domain have many diverse functions with regard to actin. Except for cofilin, Tmod, and formins, other well-characterized nucleators of actin assembly, including Lmod discussed above, use the WH2 domain for interaction with actin. These WH2-containing proteins include Spire, Cordonbleu (Cobl), VopL/VopF, and TARP. A common architecture found in Spire, Cobl, and VopL/VopF consists of tandem WH2 motifs that bind multiple actin subunits, either in a longitudinal array

representing one strand of an actin filament that can dimerize to form a nucleation complex (Spire) or in a trimeric complex with subunits representing both actin protofilaments (Cobl). These complexes remain associated with the pointed ends of growing filaments.

See also: [Cell Architecture and Function: Actin Assembly/Disassembly; Actin-Related Proteins; Rho GTPases and Actin Cytoskeleton Dynamics.](#)

Further Reading

- Bernstein BW and Bamberg JR (2010) ADF/cofilin: A functional node in cell biology. *Trends in Cell Biology* 20: 187–195.
- Campellone KG and Welch MD (2010) A nucleator arm race: Cellular control of actin assembly. *Nature Reviews Molecular Cell Biology* 4: 237–251.
- Carlier MF, Le Clainche C, Wiesner S, and Pantaloni D (2003) Actin-based motility: From molecules to movement. *BioEssays* 25: 336–345.
- Cooper JA and Sept D (2008) New insight into mechanism and regulation of actin capping protein. *International Review of Cell and Molecular Biology* 267: 183–206.
- Dominguez R (2009) Actin filament nucleation and elongation factors – structure–function relationships. *Critical Reviews in Biochemistry and Molecular Biology* 44: 351–366.
- Ono S (2007) Mechanism of depolymerization and severing of actin filament and its significance in cytoskeletal dynamics. *International Review of Cytology* 258: 1–82.
- Pollard TD and Cooper JA (2008) Actin, a central player in cell shape and movement. *Science* 326: 1208–1212.
- Puius YA, Mahoney NM, and Almo SC (1998) The modular structure of actin-regulatory proteins. *Current Opinion in Cell Biology* 10: 23–34.

Actin Organization

T Svitkina, University of Pennsylvania, Philadelphia, PA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Actin-associated proteins Proteins that bind monomeric actin, actin filaments, or both, thereby altering respective actin properties. They are also known as actin-binding proteins (ABPs).

Barbed end/pointed end Opposite ends of intrinsically polar actin filaments with the barbed end (plus end) exhibiting more rapid dynamics than the pointed end (minus end).

Filopodia Thin, finger-like protrusions extending from the cell surface and responsible for probing the surrounding environment and reaching distant targets. They usually contain tight bundles of actin filaments.

Focal adhesions Cellular structures responsible for strong adhesion of cell to extracellular matrix. They are composed of a large set of signaling and adhesion molecules and often associated with actin stress fibers.

Lamellipodia Broad, flat, and thin cellular protrusions primarily responsible for the leading edge advance. They contain a dynamic network of branched actin filaments.

Myosin Large family of motor proteins capable of binding and moving along actin filaments. Individual myosins can be specialized for particular functions, such as contraction or transport of cargo.

Nucleation Initial slow phase of polymerization that involves thermodynamically unfavorable clustering of first few monomers. This cluster, or nucleus, then starts fast polymerization by incorporating additional monomers.

Sarcomere Repeating sections of myofibrils with a characteristic pattern of overlapping, contractile thick filaments (myosin) and thin filaments (actin) responsible for muscle contraction in skeletal or cardiac muscles.

Stress fibers Mixed bundles of actin and myosin-II filaments together with additional actin-associated proteins found in some nonmuscle cells. These fibers are usually attached to the extracellular matrix through focal adhesions and are responsible for generation of contractile force during cell migration and tissue morphogenesis.

Treadmilling Phenomenon of actin dynamics at which actin monomers incorporate at the barbed end of the actin filament and dissociate from its pointed end simultaneously. Treadmilling is possible due to lower critical concentration of actin addition to the barbed end than the pointed end and occurs when the ambient concentration of monomers is intermediate between two values.

Introduction

Actin is an abundant cytoplasmic protein that is able to assemble into polar filaments consisting of two strands of monomers twisted around each other. In cells, actin filaments undergo constant polymerization and depolymerization. Moreover, they usually function as an ensemble. These key features of actin behavior are regulated by numerous accessory proteins. Dynamic behavior of actin filaments is critical for their functions; therefore, all aspects of actin polymerization and depolymerization, such as nucleation, elongation, and disassembly, are tightly controlled in time and space by special proteins. Accessory proteins also help actin filaments to form higher-order structures, such as various networks and bundles, which differ in their properties, functions, and intracellular localization. Actin filaments frequently act in association with cellular membranes, which requires additional proteins. A collection of actin filaments with their associated proteins is usually referred to as the 'actin cytoskeleton'.

The actin cytoskeleton is the major force-generating machinery in the cell, which can produce pushing, pulling, and resistance forces. Actin filaments also serve as tracks for molecular motors of the myosin superfamily that mediate cargo delivery. The generation of pushing (or protrusive) forces involves coordinated polymerization of multiple actin filaments. Typical examples of actin polymerization-driven protrusion are the extension

of the cell plasma membrane during cell migration or shape change, and propulsion of membrane-enclosed organelles. Pulling (or contractile) forces are produced by the actin cytoskeleton in cooperation with myosin motors. The classic example of such activity is muscle contraction. Resistance forces are important for cell shape maintenance, for example, in biconcave red blood cells or in cylindrical cellular appendages, such as microvilli and stereocilia. The generation of resistance forces largely depends on activity of actin filament cross-linking proteins.

Actin Organization for Protrusion

General Concept

The actin cytoskeleton generates pushing force using energy of actin polymerization. The directionality of force originates from the structural polarity of actin filaments, in which one end (barbed or plus end) undergoes faster dynamics than the other (pointed or minus end). The filament polarity is further enhanced by time-dependent hydrolysis of the actin-bound adenosine triphosphate (ATP) after monomer incorporation into the filament. Due to structural and biochemical differences, the barbed end of a filament can grow while the pointed end shrinks, a process called 'treadmilling' (Figure 1). The intrinsic treadmilling of actin filaments is greatly augmented in cells by regulatory proteins, primarily by profilin and actin

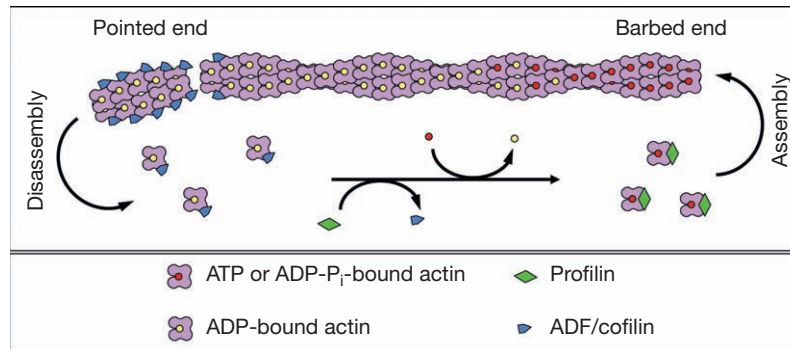


Figure 1 Actin filament treadmilling. Polymerization of the actin filament preferentially occurs at the barbed end from ATP-actin/profilin complexes, which dissociate releasing profilin after incorporation of the actin monomer into the filament. Polymerization-triggered ATP hydrolysis and subsequent phosphate release in assembled actin subunits make filaments more susceptible to depolymerization and increase their affinity for ADF/cofilin. ADF/cofilin severs filaments promoting their depolymerization. Released actin subunits are recycled when profilin competes off ADF/cofilin and promotes nucleotide exchange in the actin monomer.

depolymerizing factor (ADF)/cofilin. Profilin binds the barbed end surface of actin monomers precluding their interactions with pointed ends of monomers or polymers. This directs filament elongation toward barbed ends and prevents spontaneous filament nucleation. ADF/cofilin binds laterally to actin filaments and causes their severing, thus exposing additional ends susceptible to depolymerization. Preferential binding of ADF/cofilin to adenosine diphosphate (ADP)-containing actin subunits leads to accelerated depolymerization of the older parts of the filament located closer to the pointed end.

Intrinsic actin filament treadmilling lies at the core of the concept of pushing force generation. In cells, pushing actin filaments are uniformly oriented with their barbed ends toward the load, commonly, the plasma membrane or a membrane organelle. While elongating barbed ends push on the load, disassembly occurs closer to the pointed ends, releasing monomers for recycling. Although this concept is well conserved among cells and across biological processes, the specific design of actin polymerization machinery varies. In general, individual actin filaments are too weak to produce large forces. Therefore, they usually act as a cohort of interconnected filaments, which allows them to achieve necessary stiffness. The main strategies of coordinating polymerization of multiple filaments are the formation of parallel bundles and branched networks.

Protrusion of Actin Filament Bundles

The first example of membrane protrusion driven by elongating actin filament bundles was the acrosome reaction in echinoderm sperm, which was also the original discovery of the actin polymerization-dependent force generation. An initial contact of a sperm cell with an egg leads to explosive extension of a long acrosomal process from the sperm head, which helps the sperm to penetrate the egg jelly. The extension of the process correlates with the growth of an internal actin filament bundle consisting of actin filaments uniformly oriented with their barbed ends toward the tip. Polymerization stops when the available pool of monomers is used up.

The molecular machinery of treadmilling actin bundles is best studied in filopodia, which are needle-like protrusions at the leading edge of migrating cells (Figure 2). Individual

actin filaments in the filopodial bundles are uniformly oriented with barbed ends toward the tip and span the entire length of filopodia. During protrusion, actin subunits are incorporated at the tips, move away from the tip as a part of the filament lattice, and are released at the rear of filopodia. Slow intrinsic treadmilling of actin filaments is accelerated in filopodia by proteins controlling elongation, cross-linking, interaction with the membrane, and disassembly of actin filaments. Elongation is assisted by two distinct families of barbed end-binding proteins, formins and Ena/vasodilator-stimulated phosphoprotein (VASP) proteins. Although unrelated genetically or biochemically, both classes of proteins allow for incorporation of new actin subunits without dissociating from the barbed end. Elongation is accelerated because these proteins bind profilin, through which they recruit multiple profilin-actin complexes and increase the local concentration of actin subunits near the barbed end. In addition, both formins and Ena/VASP proteins prevent binding of capping proteins to the barbed ends, which would terminate elongation, and keep the elongating ends near the membrane to increase efficiency of pushing. Filopodia elongation is also promoted by myosin X, which likely delivers actin subunits and/or regulators of polymerization to filopodial tips using its motor activity. The intrinsically slow dissociation of actin subunits from the pointed ends cannot account for maintaining the steady state of fast protruding filopodia. Two actin filament severing proteins, ADF/cofilin and gelsolin, are believed to facilitate depolymerization by increasing the number of exposed pointed ends. Myosin II, which breaks filaments by applying a pulling force on them, also contributes to this process. Cross-linking of actin filaments in filopodia increases the stiffness of the bundle, thus allowing it to overcome the membrane resistance and prevent buckling during protrusion. The major cross-linker in filopodia is fascin, a bivalent monomeric protein. Lateral binding of filopodial bundles to the plasma membrane helps maintain the shape of filopodia. It is likely mediated by proteins of the ezrin-radixin-moesin (ERM) family, which bind both actin filaments and membrane components, phospholipids and transmembrane proteins. Specific membrane-deforming proteins, such as IRSp53, also play positive roles in filopodia protrusion by stabilizing the

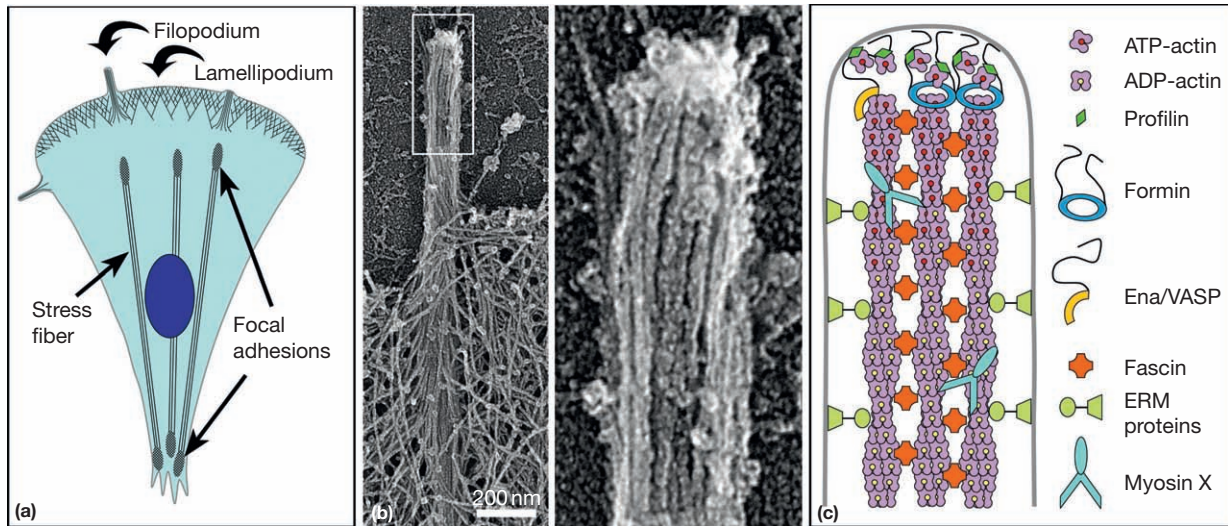


Figure 2 Actin organization in filopodia. (a) Typical actin structures in a migrating fibroblast-like cell. Protrusive organelles, lamellipodia and filopodia, are located at the cell leading edge. (b) Platinum replica transmission electron microscopy of a filopodium at the leading edge of a mouse melanoma cell. Boxed region is enlarged in the right panel and shows long parallel actin filaments. (c) Molecular organization of filopodia. Actin filaments are oriented with barbed ends toward the filopodial tip. They are bundled by fascin and laterally attached to the plasma membrane by ERM proteins. Formin and/or Ena/VASP proteins associate with the barbed ends, promote their elongation, and anchor to the membrane at the tip. Unconventional myosin X moves along actin filaments and accumulates at the filopodial tips (not shown).

tubular shape of the filopodial membrane and by indirectly linking it to the actin cytoskeleton.

Protrusion of Branched Networks

The assembly of branched actin filament networks provides pushing force for multiple cellular events, the best studied of which are protrusion of lamellipodia at the leading edge of migrating cells (Figure 3) and rocketing motility of certain intracellular pathogens, such as *Listeria monocytogenes*. In addition, branched networks promote various endocytic events, motility and biogenesis of intracellular organelles, and the formation of diverse cell–cell junctions including neuronal and immune synapses.

Protrusion of branched networks involves continuous nucleation of new actin filaments at the side of pre-existing filaments in the process called ‘dendritic nucleation’. The nascent filaments elongate and produce force while they are short and stiff. Their elongation is then terminated by barbed end capping, while new filaments are nucleated to maintain protrusion. Being anchored to pre-existing filaments and, consequently, to the entire branched network, nascent filaments can efficiently transform the energy of polymerization into useful work, rather than push themselves away from the load. Filaments are nucleated at a defined angle of 70° , which allows the network to expand or reorient itself by simply adjusting the rates and sites of nucleation and capping. At steady state, nucleation of new filaments is balanced by capping of old filaments at equivalent rate. The recycling of actin monomers is mediated by disassembly of the network through a combination of debranching and severing followed by depolymerization of filament fragments. Branched networks disassemble throughout their volume except for the very front, where actin filaments are enriched in ATP-loaded subunits. Thus, the branched network as a whole, but not individual filaments,

undergoes treadmilling by assembling at the front and, on average, disassembling at the rear.

The assembly of branched networks of actin filaments requires several accessory molecules. A heteromeric complex of seven proteins, the Arp2/3 complex, is responsible for key events in this process: it nucleates filaments, links them to the pre-existing filaments at the defined angle of 70° , and caps the pointed end of the new filament to prevent its premature depolymerization. The intrinsically inactive Arp2/3 complex is activated by nucleation-promoting factors, primarily by the Wiskott–Aldrich syndrome protein (WASP) family proteins, which are recruited to the working interface, such as the plasma membrane or a particle. This allows the nucleation to occur, and therefore force to be generated, right where it is needed. Barbed end capping is mediated by the heterodimeric capping protein, which tightly binds the barbed end and prevents its dynamics. ADF/cofilin plays a key role in depolymerization of branched networks. The activated Arp2/3 complex, capping protein, and ADF/cofilin represent the minimal molecular toolbox required for steady-state treadmilling of branched networks. Many other molecules improve the performance of this machinery in a living cell. Thus, profilin prevents spontaneous nucleation and targets monomers to the barbed ends, formins and Ena/VASP proteins transiently protect elongating ends from capping and processively anchor them to the membrane, the actin-binding protein cortactin stabilizes branches, the cross-linkers filamin A and α -actinin consolidate the whole network, and numerous molecules regulate all these proteins.

Actin Organization for Contraction

General Concept

Pulling forces are generated by the actin cytoskeleton in cooperation with myosin motors. Myosins are a large superfamily

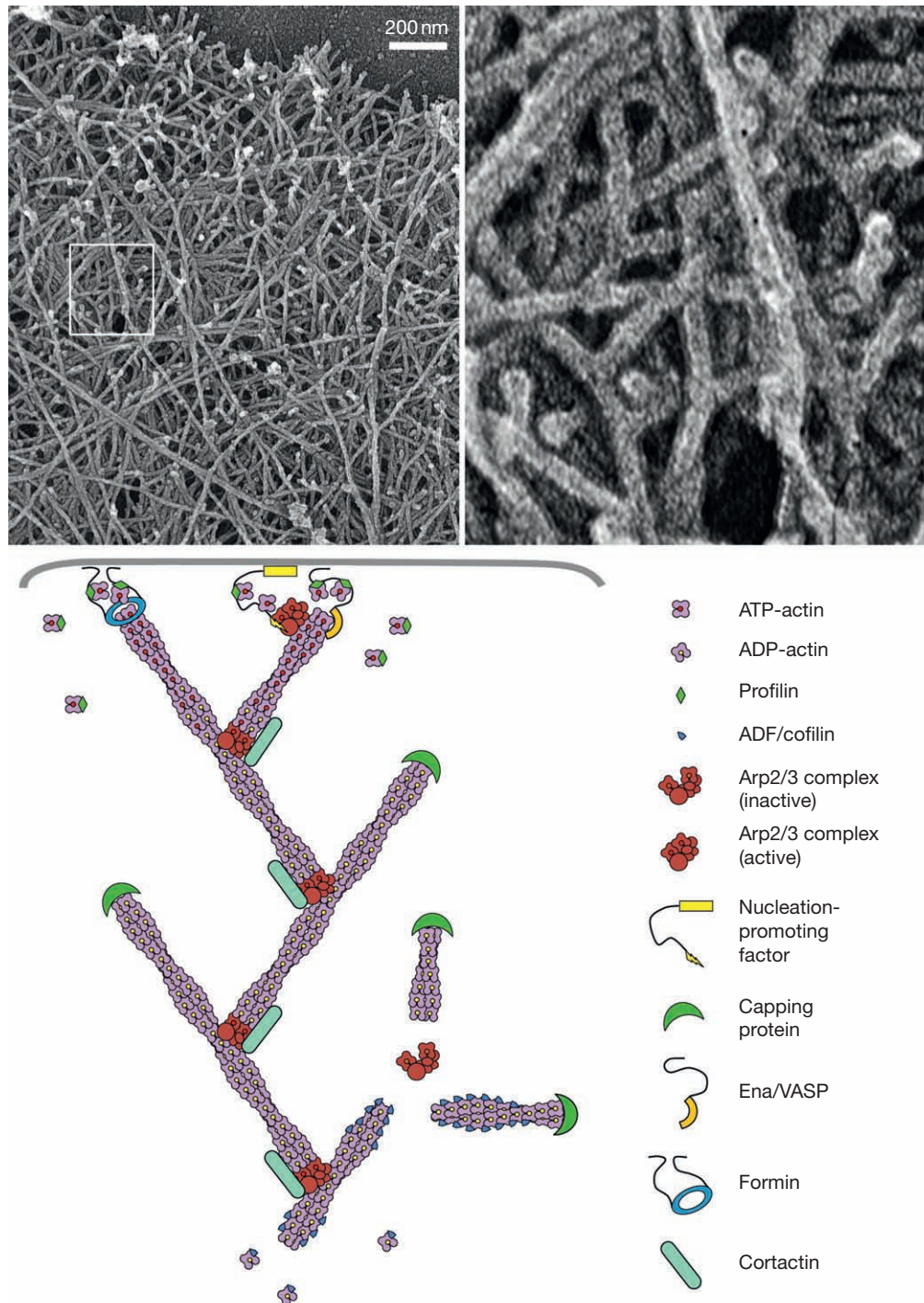


Figure 3 Actin organization in lamellipodia. (Top) Platinum replica transmission electron microscopy of a lamellipodium at the leading edge of a fish epidermal keratocyte. Boxed region is enlarged in the right panel and shows branched actin filaments. (Bottom) Molecular organization of lamellipodia. The Arp2/3 complex is cooperatively activated by a membrane-targeted nucleation-promoting factor and a pre-existing actin filament. Upon activation, it nucleates a new actin filament at the side of the pre-existing filament and remains associated with the branch point. Branch points are further stabilized by cortactin. The nascent filament elongates with its barbed end oriented toward the membrane. Its elongation is promoted by formins and/or Ena/VASP proteins. After a period of elongation, the barbed end is capped by capping protein and lags behind the protruding leading edge. Disassembly at the rear of the lamellipodial network occurs through dissociation of branches and ADF/cofilin-mediated severing.

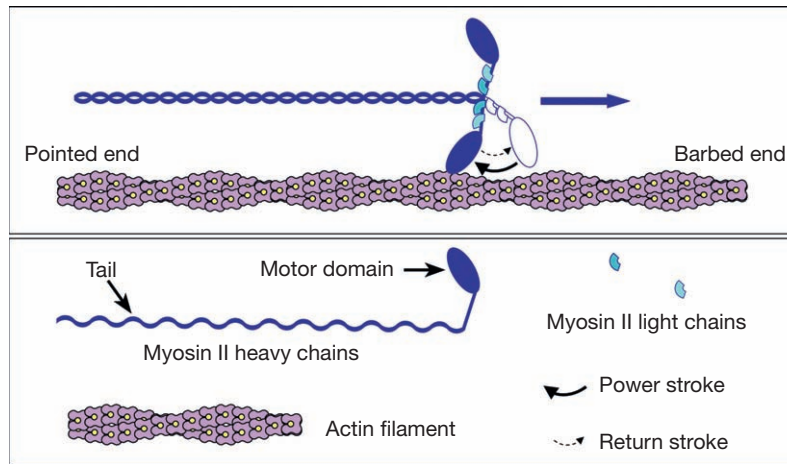


Figure 4 Myosin II motor activity. Myosin II hexamer consisting of two heavy chains and two pairs of light chains moves toward the barbed end of the actin filament by swinging its motor domain in an ATP-dependent manner.

of motor proteins, which are able to convert chemical energy of ATP into mechanical work (Figure 4). The myosin molecule consists of one or two heavy chains containing a motor domain and a cargo-binding tail, and a variable number of light chains. Cycles of ATP binding, hydrolysis, and release of hydrolytic products by the motor domain are coupled to specific and sequential changes in the myosin conformation and affinity for actin filaments, which allow myosin to move and translocate cargo along actin filaments, or to pull on the actin filament. The majority of myosins, with the notable exception of myosin VI, move toward the barbed ends of the actin filament.

Class II myosins are specifically designed to mediate cell contraction. In contrast to other myosins, myosin II polymerizes into bipolar filaments through antiparallel interactions between their coil-coiled tail domains, so that the motor domains are located at both filament ends. When presented to actin filaments of opposite polarity, bipolar filaments cause contraction by pulling on actin filaments from both sides in opposite directions. This concept of actin–myosin-mediated contraction, the sliding filament model, was proposed independently by Hugh Huxley and Andrew Huxley in 1954 based on their studies of skeletal muscle contraction. Other cells utilize the same principle of mutual sliding of actin and myosin II filaments, but in a less ordered and more dynamic fashion.

In contrast to actin polymerization-driven processes, actin filaments in contractile systems should be relatively stable to provide reliable tracks for myosin motors, while actin dynamics is needed primarily for renewal of aged or damaged filaments. Stabilization of actin filaments is mediated by capping proteins blocking dynamics at the ends, and by tropomyosins stabilizing filaments along their length. Tropomyosin is a long dimeric coiled-coil protein that sits in the groove between two strands and binds seven consecutive subunits in the actin filament, without interfering with myosin movement. Correct spatial organization of actin filaments is also crucially important for the functionality of contractile systems. It is controlled by cross-linking proteins, especially by α -actinin, and by system-specific proteins anchoring the actin filaments to the load.

Striated Muscle Contraction

The cellular contractile machinery in striated (skeletal and cardiac) muscles appears to have evolved to perfection. It consists of highly ordered arrays of actin and myosin filaments efficiently generating contractile force. The minimal functional unit in striated muscles is the sarcomere (Figure 5). A multinuclear muscle cell (muscle fiber) contains multiple longitudinally connected sarcomeres separated by Z-lines. A set of actin (or thin) filaments of uniform length and polarity is attached to the Z-line on each side of the sarcomere. The barbed ends of actin filaments are embedded into the Z-line and capped by heterodimeric capping protein, which in muscle is called 'CapZ'. Pointed ends are located close to the middle of the sarcomere and are capped by the pointed end capper, tropomodulin. Tropomyosin additionally stabilizes actin filaments along their length. Another protein laterally bound to actin filaments, nebulin, is thought to control actin filament length. The regular spacing between actin filaments is defined by α -actinin, located at Z-lines. Bipolar filaments composed of the skeletal muscle myosin II (thick filaments) are suspended in the middle of the sarcomere and their ends interdigitate with actin filaments in both halves of the sarcomere. The regular arrangement of myosin II filaments is maintained by proteins of the M-line, a structure in the exact center of the sarcomere, and by the giant protein titin, which extends from the M-line to the Z-line along each myosin filament. The motor domains of myosin filaments interact with actin filaments, forming cross-bridges structurally recognizable by electron microscopy. Translocation of myosin motor domains toward the barbed ends of actin filaments on each side of the sarcomere shortens individual sarcomeres which, given their serial organization, contracts the entire cell.

Regulation of muscle contraction is primarily mediated by Ca^{2+} . When an action potential arrives from the motor neuron, it opens calcium channels in the membrane of the sarcoplasmic reticulum through a series of intermediate events and causes a sudden increase in the Ca^{2+} concentration in the muscle fiber. A calcium-sensing regulatory element in the sarcomere is the troponin complex consisting of troponins I, T,

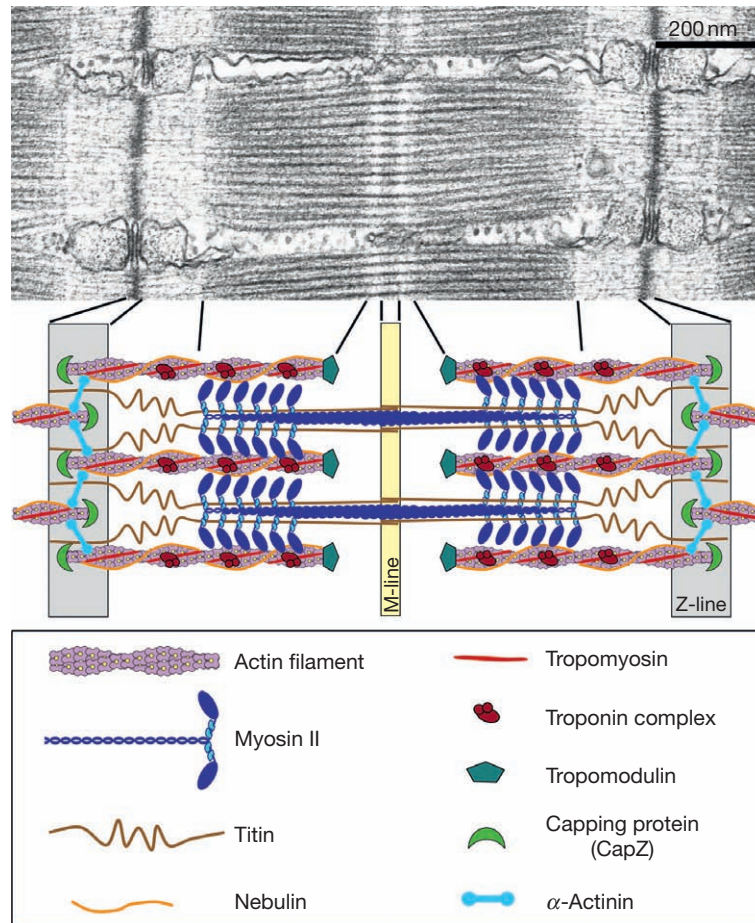


Figure 5 Actin organization in striated muscles. (Top) Thin section transmission electron microscopy image of frog sartorius muscle shows sarcomeric organization of muscle fibers. One full sarcomere is present in the image. (Bottom) Molecular organization of the sarcomere. Black lines connect corresponding zones in the micrograph and the diagram. Top image: Courtesy of Dr. Clara Franzini-Armstrong (University of Pennsylvania).

and C. The troponin complex binds both actin and tropomyosin and regulates the positioning of tropomyosin on actin filaments in a Ca^{2+} -dependent manner. In the absence of Ca^{2+} , tropomyosin moves out of its normal position in the groove on the actin filament and blocks the myosin–actin interaction. In the presence of Ca^{2+} , when the inhibitory interaction between actin, troponin, and tropomyosin is released, tropomyosin moves back into the groove, allowing myosin to move.

Smooth Muscle Contraction

The spatial organization and the mechanism of contraction of smooth muscle cells are still poorly understood. Smooth muscle cells do not have sarcomeric organization and appear to combine features of striated muscle fibers and nonmuscle cells. Instead of Z-lines, barbed ends of actin filaments in smooth muscle cells associate with α -actinin-rich dense bodies scattered through the cytoplasm or bound to the plasma membrane and interconnected by intermediate filaments. Smooth muscle isoforms of myosin II polymerize into side-polar filaments, instead of bipolar filaments, of variable lengths.

Side-polar myosin II filaments lack a bare zone in the middle, so that myosin heads project from the filament shaft all along the length, but change polarity on each side of the filament. The actin and myosin filaments in smooth muscle cells are only roughly aligned with the longitudinal axis of these spindle-shaped cells. Sliding of myosin II filaments along actin filaments causes contraction due to extensively interconnected filamentous network in the cell.

The regulation of smooth muscle contraction relies on phosphorylation-dependent changes of myosin II activity, rather than on troponin-dependent repositioning of tropomyosin, although both are regulated by Ca^{2+} . Smooth muscle myosin II is active when one of its associated light chains, the regulatory light chain, is phosphorylated. Phosphorylation is mediated primarily by myosin light chain kinase (MLCK), which is activated by Ca^{2+} . When light chains are not phosphorylated, the smooth muscle myosin has very low actin-activated adenosine triphosphatase (ATPase) activity in its motor domains. Also, it cannot form filaments under this condition, because its tail bends and interacts with the motor domains. Phosphorylation of the regulatory light chain causes unfolding of the molecule, activation of the ATPase activity,

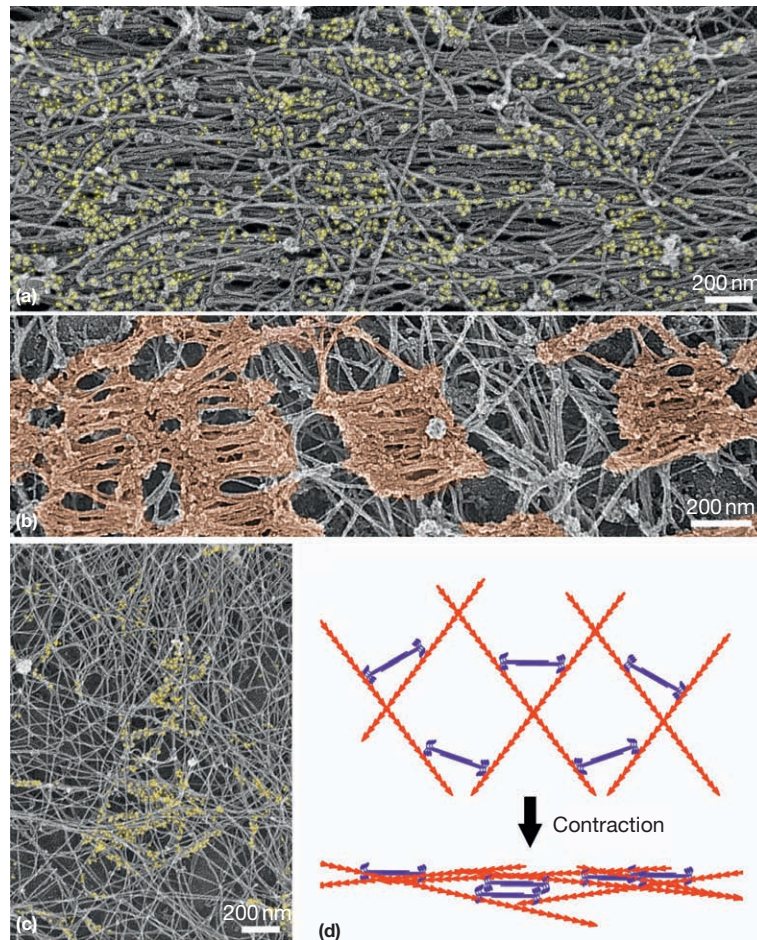


Figure 6 Contractile actin structures in nonmuscle cells. (a, b) Stress fibers in rat fibroblasts (platinum replica transmission electron microscopy). (a) Immunolabeling of myosin II with colloidal gold (colored in yellow) shows semi-periodic distribution of myosin II along the actin filament bundle. (b) Dissolution of actin filaments by gelsolin treatment reveals sets of aligned bipolar myosin II filaments (colored in brown) separated by irregular intervals. (c, d) Contractile actin-myosin II network in lamella of a fish epidermal keratocyte. (c) Immunogold staining of myosin II shows linear sets of gold particles (colored in yellow) representing myosin filaments among the network of actin filaments. Orientation of actin and myosin filaments changes from the cell's front (top) to rear (bottom) reflecting network contraction during keratocyte migration. (d) Network contraction model. Actin filaments (orange) form a diagonal network with barbed ends to the front (top). Myosin bipolar filaments (purple) that initially are randomly bound to actin filaments move toward barbed ends causing network contraction and mutual orientation of actin and myosin filaments. At the next step, the newly formed bundle will contract longitudinally.

formation of myosin II filaments, and muscle contraction. Conversely, relaxation of smooth muscles requires low calcium and phosphatase activity.

Contraction of Nonmuscle Cells

Many nonmuscle cells are able to generate substantial contractile forces that are used for morphogenesis, cytokinesis, rearrangement of extracellular matrix, and formation of cell–cell and cell–matrix junctions. For contraction, nonmuscle cells use mixed bundles and networks of actin and myosin filaments. Nonmuscle myosin II forms very short bipolar filaments ($\sim 0.3 \mu\text{m}$) and is regulated similar to smooth muscle myosin II. However, many more kinases, besides MLCK, can phosphorylate nonmuscle myosin II on the regulatory light chain. Activation of these kinases occurs in response to various

signals, with the most important ones being Ca^{2+} and Rho family guanosine triphosphatases (GTPases).

Contractile bundles of nonmuscle cells, or stress fibers (Figures 2, 6(a), and 6(b)), are best studied in fibroblasts, which are mesenchymal cells engaged in secreting and remodeling the extracellular matrix. Stress fibers consist of aligned actin and myosin filaments and are attached to the extracellular matrix through transmembrane integrin receptors at special attachment sites called 'focal adhesions'. Focal adhesions are very complex structures that contain α -actinin among many other structural and signaling proteins. At focal adhesions, actin filaments are uniformly oriented with their barbed ends toward the plasma membrane, but they are not capped, as in sarcomeres, and can elongate. Away from focal adhesions, actin filaments in stress fibers have mixed polarity. Myosin bipolar filaments are arranged in a semi-periodic fashion along the length of the stress fiber, where they colocalize with

tropomyosin, whereas intervening areas are enriched in α -actinin. This organization is distantly related to that seen in striated muscle sarcomeres, although in a considerably less ordered fashion, given that myosin filaments are not perfectly aligned across the stress fiber and are unevenly spaced along the stress fiber. Actin filaments also have variable lengths and may pass through several myosin-rich regions. Contraction in stress fibers, like in other contractile cells, involves mutual sliding of actin and myosin filaments. Productive contraction likely occurs when myosin motors interact with nearby actin filaments of appropriate orientation.

In order to mediate contraction, actin and myosin filaments do not have to be aligned, but may be arranged into a network, in which individual filaments are oriented at a range of angles (Figures 6(c) and 6(d)). If actin filaments are attached at their barbed ends, as is usually the case in cells, movement of two ends of a myosin filament toward randomly oriented barbed ends produces mutual alignment of actin and myosin filaments between the major attachment sites, and fosters a network contraction in the perpendicular direction. This mechanism is responsible, for example, for the translocation of the cell body in fish epidermal keratocytes, and for initial formation of stress fibers in fibroblasts. It also likely plays a role in other less understood contractile events in nonmuscle cells.

structures except for that needed for their renewal. Association of these actin ensembles with the plasma membrane is required to translate their mechanical properties into a specific cell shape.

Actin filament cross-linking proteins use two or more actin-binding sites to bring actin filaments together. Some cross-linkers, such as fascin, fimbrin (I-plastin), and espin, have two actin-binding sites within one polypeptide. Others, such as α -actinin and filamin, readily form homodimers with one actin-binding site per subunit. Spectrin is a long flexible heterotetramer with two actin-binding sites at the opposite ends of the molecule. Myosin II filaments can also function as multivalent cross-linkers when they do not use their motor activity, or use it in an isometric mode. The distance between actin-binding sites along with the flexibility of spacers determines the geometry and mechanical properties of resulting actin filament ensembles. Thus, the short spacing between actin-binding sites in fascin, espin, or fimbrin produces stiff and tightly packed bundles of actin filaments. The longer spacer present in α -actinin keeps cross-linked actin filaments farther apart to allow for the movement of myosin filaments, but it is still narrow enough to align filaments into bundles. Cross-linkers with significantly distant actin-binding sites, such as spectrin and filamin, foster the formation of loose elastic networks.

Actin Organization for Shape Determination

General Concept

Cells extensively use the actin cytoskeleton to define their shape and resist environmental stresses. Although protrusive and contractile forces and junctions with other cells and matrices are intrinsic parts of the cell shape determination by the actin cytoskeleton, some actin structures are designed just to resist external forces. Individual actin filaments are relatively weak and too unstable for this function, but they form strong and stable ensembles when interconnected by cross-linking proteins. Actin dynamics is also usually suppressed in such

Shape Determination by Elastic Networks

The best-studied example of cell shape determination by elastic networks is the actin–spectrin membrane skeleton in red blood cells, which is responsible for the characteristic biconcave shape of these cells and their ability to withstand significant shear stresses in the circulation (Figure 7). Spectrin, rather than actin itself, is the major structural component of this network. Long and flexible spectrin tetramers form a triangular mesh, in which very short actin filaments (~ 13 monomers long) join ends of up to six spectrin molecules at each node forming the so-called junctional complex. Each actin filament is stabilized

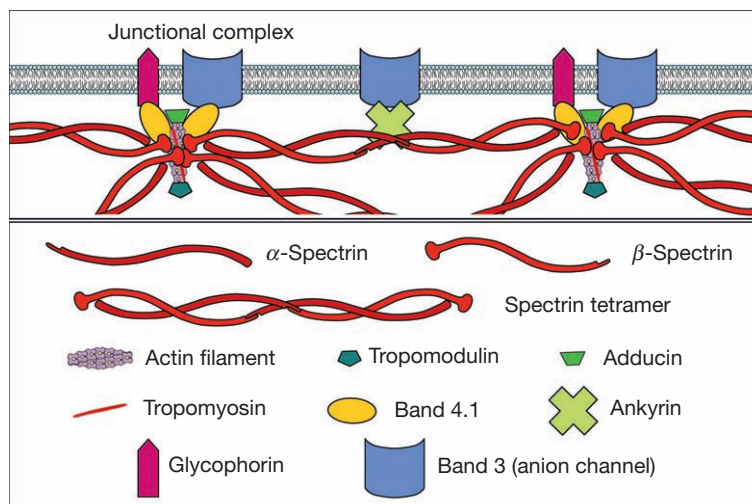


Figure 7 Membrane skeleton in red blood cells. The actin–spectrin network is attached to the plasma membrane through two major anchors, ankyrin and band 4.1, which associate with transmembrane proteins, glycophorin and band 3.

along its length by tropomyosin and is capped by tropomodulin and adducin at the pointed and barbed ends, respectively. Connection of the actin-spectrin mesh to the membrane is mediated by several anchors, the most important of which are band 4.1 and ankyrin. Band 4.1 binds both actin and spectrin at the junctional complex and stabilizes their interaction. Simultaneously it interacts with the plasma membrane through membrane phospholipids and transmembrane proteins, band 3 (anion exchanger) and glycophorin. Ankyrin binds the middle of the spectrin tetramer and links it to band 3 in the plasma membrane. Similar actin-spectrin networks also function in nonerythroid cell types, where they use many of the same, or homologous, proteins to maintain the membrane integrity and define its physical properties.

Shape Determination by Stiff Bundles

Stiff actin filament bundles play important roles in the elaboration of cylindrical cell surface protrusions, the best-understood examples of which being microvilli in the intestinal epithelium and stereocilia in the hair cells in the inner ear.

Microvilli are nonmotile finger-like protrusions from the apical surface of epithelial cells that function to increase the cell surface area and the efficiency of absorption. Microvilli have uniform length (1–2 μm) and thickness (~ 200 nm) and are tightly packed next to each other forming a special compartment of the cell called the 'brush border'. The shape of microvilli is maintained by an internal bundle of ~ 20 – 30 parallel actin filaments. The bundle is structurally similar to that occurring in filopodia, but exhibits approximately threefold slower dynamics. The actin filament barbed ends located at the tips of microvilli are embedded into a material of unknown composition, which likely regulates their dynamics. The main cross-linkers in the bundle are fimbrin and villin, a gelsolin-family severing protein with bundling activity. The actin bundle is laterally attached to the plasma membrane by monomeric myosin Ia, which binds actin filaments by its motor domain and acidic phospholipids in the plasma membrane by its tail. Myosin Ia is fully functional as a motor and is believed to regulate membrane tension around microvilli. Microvilli are also present on the surface of other cell types, such as other epithelial cells, lymphocytes, and various sensory cells. However, the main cross-linkers in sensory cell microvilli are fimbrin and espin, whereas villin is specific for epithelial cells.

Stereocilia are specialized mechanosensory microvilli involved in hearing and the sense of balance. They differ from microvilli by their size, shape, dynamics, and molecular composition. Stereocilia lengths vary from 1 to 100 μm , but are tightly controlled to match the frequency of sound waves to be detected. Stereocilia have ~ 10 -fold more actin filaments in their core bundle than microvilli. They also undergo ~ 100 -fold slower dynamics. Actin filaments are cross-linked by fimbrin

and espin, as in sensory, but not intestinal, microvilli. In contrast to fully cylindrical microvilli, stereocilia taper toward the bottom, which allows them to pivot around the base in response to vibration while remaining straight. Unconventional myosins Ia, Ic, IIIa, VI, VIIa, and XVa, have been functionally implicated in hearing by contributing to stereocilia structure. Myosins IIIa and XVa localize to stereocilia tips, like myosin X in filopodia, and contribute to length determination in a complementary fashion. Myosins Ia, Ic, and VIIa are present along the length of the stereocilia, similar to myosin Ia in microvilli, and may also regulate membrane tension. Myosin VI is present at the tapered base of stereocilia, and is thought to deliver actin regulators to these sites.

See also: [Bioenergetics: Troponin](#); [Cell Architecture and Function: Actin Assembly/Disassembly](#); [Actin-Capping and -Severing Proteins](#); [Actin-Related Proteins](#); [Cell Migration](#); [Cytoskeletal Motors: General Principles](#); [Focal Adhesions and Related Integrin Contacts](#); [Myosin Motors](#); [Rho GTPases and Actin Cytoskeleton Dynamics](#).

Further Reading

- Bugyi B, Le Clainche C, Romet-Lemonne G, and Carlier MF (2008) How do *in vitro* reconstituted actin-based motility assays provide insight into *in vivo* behavior? *FEBS Letters* 582: 2086–2092.
- Campellone KG and Welch MD (2010) A nucleator arms race: Cellular control of actin assembly. *Nature Reviews. Molecular Cell Biology* 11: 237–251.
- Chhabra ES and Higgs HN (2007) The many faces of actin: Matching assembly factors with cellular structures. *Nature Cell Biology* 9: 1110–1121.
- Gunst SJ and Zhang W (2008) Actin cytoskeletal dynamics in smooth muscle: A new paradigm for the regulation of smooth muscle contraction. *American Journal of Physiology – Cell Physiology* 295: C576–C587.
- Kee AJ, Gunning PW, and Hardeman EC (2009) Diverse roles of the actin cytoskeleton in striated muscle. *Journal of Muscle Research and Cell Motility* 30: 187–197.
- Manor U and Kachar B (2008) Dynamic length regulation of sensory stereocilia. *Seminars in Cell and Development Biology* 19: 502–510.
- Martin AC (2010) Pulsation and stabilization: Contractile forces that underlie morphogenesis. *Development Biology* 341: 114–125.
- Mattila PK and Lappalainen P (2008) Filopodia: Molecular architecture and cellular functions. *Nature Reviews. Molecular Cell Biology* 9: 446–454.
- Mellor H (2010) The role of formins in filopodia formation. *Biochimica et Biophysica Acta* 1803: 191–200.
- Mogilner A and Keren K (2009) The shape of motile cells. *Current Biology* 19: R762–R771.
- Mohandas N and Gallagher PG (2008) Red cell membrane: Past, present, and future. *Blood* 112: 3939–3948.
- Pollard TD and Borisy GG (2003) Cellular motility driven by assembly and disassembly of actin filaments. *Cell* 112: 453–465.
- Tilney LG and DeRosier DJ (2005) How to make a curved *Drosophila* bristle using straight actin bundles. *Proceedings of the National Academy of Sciences of the United States of America* 102: 18785–18792.
- Vicente-Manzanares M, Ma X, Adelstein RS, and Horwitz AR (2009) Non-muscle myosin II takes centre stage in cell adhesion and migration. *Nature Reviews. Molecular Cell Biology* 10: 778–790.
- Zigmond SH (2004) Beginning and ending an actin filament: Control at the barbed end. *Current Topics in Development Biology* 63: 145–188.

Actin-Related Proteins

R D Mullins, University of California, San Francisco, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 27–33, © 2004, Elsevier Inc.

Glossary

Actin Ubiquitous, eukaryotic cytoskeletal protein. Actin monomers assemble into filaments that are used to organize the intracellular space and to define the shape and mechanical properties of eukaryotic cells.

Adenosine triphosphate (ATP) It is the major energy currency of living cells. The removal or hydrolysis of phosphates from ATP releases energy that individual proteins can use to do work.

Chromatin DNA that has been packaged and compacted by specialized proteins, called histones. The packaged DNA must be partially unpackaged (by chromatin remodeling complexes) to allow transcription of genes.

Cytoskeleton A network of protein polymers that organize the intracellular space; transport intracellular cargos;

determine cell shape; and drive cell locomotion.

The three most important eukaryotic cytoskeletal polymers are actin filaments, microtubules, and intermediate filaments.

Eukaryote A cell with a nucleus and other membrane-bound organelles.

Histones DNA-packaging proteins. In compacted chromatin, DNA is wound around histones like thread around a spool.

Primary sequence The sequence of amino acids that comprise a protein. Usually given from amino to carboxy terminus.

Prokaryote A simple cell, such as a bacterium, that lacks a nucleus and other intracellular compartments.

Actin-related proteins (Arps) are a class of proteins found in all eukaryotes and many species of bacteria and archaea. Arps are defined by their degree of similarity to actin (conventional actin), a ubiquitous, eukaryotic cytoskeletal protein. All Arps are built around a common structural fold. This actin fold binds and hydrolyzes adenosine triphosphate (ATP). In most Arps, ATP hydrolysis is converted into a conformational change that regulates assembly and/or function of multiprotein complexes. Conventional actin assembles into two-stranded helical filaments that form structural scaffolds used to organize the intracellular space and to drive cell division, shape change, and cell locomotion. Eukaryotic Arps fall into 10 groups (Figure 1 and Table 1) that can be loosely grouped into two categories: (1) cytoskeleton-associated Arps (Arp1, Arp2, Arp3, and Arp10) that regulate the assembly and function of the actin and microtubule cytoskeletons; and (2) chromatin-associated Arps (Arp4–Arp9) that regulate the structure and function of eukaryotic chromatin. Prokaryotic Arps were discovered only recently and their functions are not well understood. The best-characterized prokaryotic Arps (MreB, Mbl, and the plasmid-encoded protein, ParM) polymerize into filaments required for proper cell shape and DNA segregation. The primary sequences of prokaryotic Arps are highly divergent from those of the eukaryotic Arps but we include them in this discussion based on the similarity of their biochemical activities and cellular functions to conventional actin and the eukaryotic Arps. The presence of Arps in chromatin-remodeling complexes and the importance of prokaryotic Arps in DNA segregation suggest that their role in the handling of DNA is as ancient and well conserved as their more well-known cytoskeletal functions.

The Actin Fold

There is little or no conservation of surface-exposed residues among Arps. This reflects the fact that the proteins in this family have different binding partners and cellular functions (Figure 1 and Table 1). All Arps, however, share a common structural fold. The core of the protein is formed from two globular domains (domain I and domain II) connected by a flexible hinge (Figure 2(b)). The interface between the two domains forms an ATP-binding pocket. Five highly conserved sequences – three ATP-binding loops and two connecting helices – line this pocket and define the actin fold. These sequences are an adenosine-binding loop (adenosine), two phosphate-binding loops (phosphate 1 and phosphate 2), and two connecting domains (connect 1 and connect 2). These sequences are conserved in actin, the Arps, and all other members of the actin superfamily (which includes proteins like hexokinase and HSP70) and they are always arranged in the following order: phosphate 1, connect 1, phosphate 2, adenosine, and connect 2 (Figure 2(a)). The ATP-binding loops are distributed between the two domains of the protein. Phosphate 1 is located in domain I and both phosphate 2 and adenosine are in domain II (Figure 2(b)). When all three loops interact with ATP the two domains are tightly associated. Hydrolysis of the gamma phosphate of ATP weakens binding to the phosphate 1 domain and causes the two domains of the protein to swing apart around the hinge formed by connect 1 and connect 2. Complete loss of nucleotide results in further opening of the protein around the hinge. In some members of the actin superfamily the binding of ATP to a nucleotide-free protein causes the domains to rotate by 30° around the hinge region (Figure 2(c)).

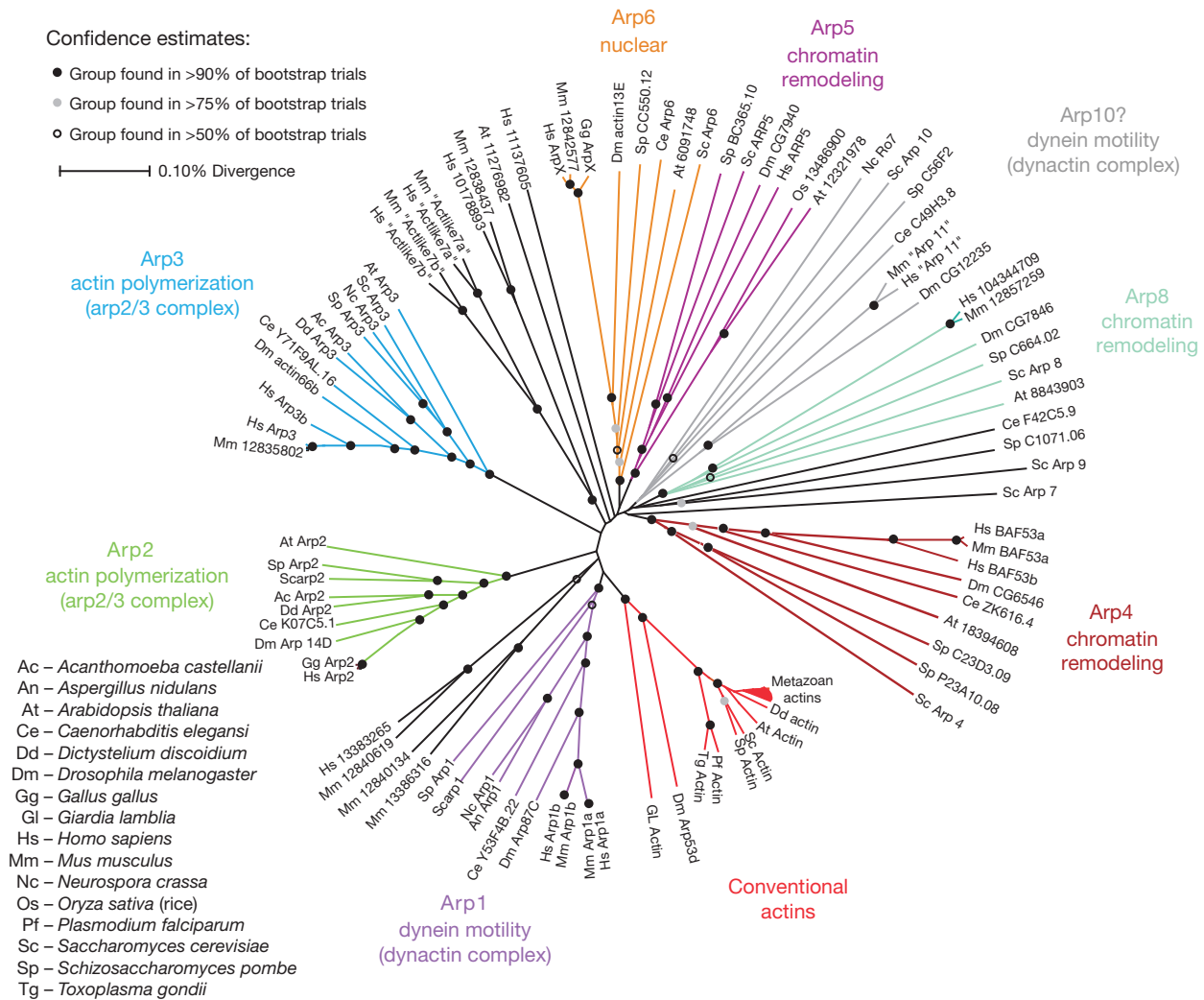


Figure 1 The eukaryotic Arp family is composed of 10, highly conserved members. Members fall broadly into two categories: cytoskeletal Arps (conventional actin, Arp1, Arp2, Arp3, and Arp10) and nuclear Arps (Arp4, Arp5, Arp6, Arp7, Arp8, and Arp9). The diagram is an unrooted phylogenetic tree of the actin superfamily. Reprinted from Goodson HV and Hawse WF (2002) Molecular evolution of the actin family. *Journal of Cell Sciences* 115(Pt 13): 2619–2622. Calculated by Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, and Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25(24): 4876–4882.

The shift from closed to open conformation is used by Arps to regulate their interactions with other proteins.

Prokaryotic Arps

An actin cytoskeleton is thought to be a prerequisite for the establishment and maintenance of intracellular compartments and organelles and until recently actin was thought to be a strictly eukaryotic protein. Recent work, however, has identified actin-like proteins in prokaryotes. The first two prokaryotic actin-like proteins, MreB and Mbl, were discovered in *Bacillus subtilis*. These proteins have very limited primary sequence similarity to conventional actin but, remarkably, they fold into three-dimensional structures very similar to actin and, like conventional actin, they polymerize into two-stranded filaments in a manner that depends on ATP and salt. At present

not much is known about the cellular roles of MreB and Mbl, but they do appear to form filaments *in vivo* that are required for proper positioning of the cell-wall synthesis machinery and proper shape determination of rod-shaped bacteria. In addition, MreB and Mbl are required for proper segregation of bacterial DNA during cell division. These are the types of functions performed by polymeric cytoskeletal elements in eukaryotes so it is fair to say that MreB and Mbl form a bona fide bacterial cytoskeleton.

Another remarkable prokaryotic actin-like protein is encoded by the *parM* locus of the low-copy drug-resistance plasmid, R100. The R100 plasmid is stably maintained at one copy per cell in bacterial populations because it accurately segregates one copy each into daughter cells during each cell division. Three plasmid loci are required for this accurate segregation: *ParR*, *parM*, and *ParC*. *ParR* encodes a DNA-binding protein that binds tightly to the *ParC* locus. *parM* encodes an actin-like

Table 1 Cellular localization, function and multiprotein complexes of eukaryotic Arps

Arp	Sequence similarity to actin	Cellular localization	Function	Multiprotein complexes
Actin	100% ident. 100% simil.	Cell cortex, leading edge, stress fibers, endosomes, vesicles, nucleus	Cell shape and motility, intracellular trafficking, chromatin remodeling (?), histone acetylation (?)	Actin filaments, INO80, BAF, and NuA4 complexes
Arp1	45% ident. 68% simil.	Golgi apparatus, transport vesicles, centrosomes, cell cortex	Dynein-driven intracellular movements	Dynactin complex
Arp2	45% ident. 69% simil.	Cell cortex, leading edge, some motile endosomes	Actin filament nucleation and cross-linking	Arp2/3 complex
Arp3	39% ident. 60% simil.	Cell cortex, leading edge, some motile endosomes	Actin filament nucleation and crosslinking	Arp2/3 complex
Arp4	30% ident. 53% simil.	Nucleus	Histone acetylation, chromatin remodeling, control of transcription	NuA4 and INO80 complexes
Arp5	26% ident. 51% simil.	Nucleus	Chromatin remodeling, control of transcription	INO80 complex
Arp6	24% ident. 46% simil.	Nucleus	Unknown	Unknown
Arp7	22% ident. 44% simil.	Nucleus	Chromatin remodeling, control of transcription	Swi/Snf and RSC complexes
Arp8	21% ident. 44% simil.	Nucleus	Chromatin remodeling, control of transcription	INO8 complex
Arp9	17% ident. 40% simil.	Nucleus	Chromatin remodeling, control of transcription	Swi/Snf and RSC complexes
Arp10	17% ident. 38% simil.	Golgi apparatus, transport vesicles, centrosomes, cell cortex	Dynein-driven intracellular movements	Dynactin complex

protein that assembles into two-stranded filaments in an ATP-dependent manner. The ParR protein bound to the ParC locus appears to nucleate formation of *parM* filaments, which drive plasmid segregation. This simple three-piece system may form a structure functionally equivalent to a mitotic spindle and may represent the minimal system for segregating DNA. Unlike conventional actin filaments in which ATP hydrolysis regulates interaction with depolymerizing factors (**Figure 3(b)**), ATP binding and hydrolysis directly control assembly and disassembly of ParM filaments. ATP-bound monomers polymerize into filaments. Within the filament, monomers hydrolyze ATP to adenosine diphosphate (ADP). ADP-bound filaments then rapidly disassemble (**Figure 3(a)**).

Eukaryotic Arps

Cytoskeletal Arps

We briefly discuss conventional actin, which is well described in a separate article on actin dynamics in this encyclopedia. Arp1 and Arp10 are part of a dynein-activating (dynactin) complex that regulates activity of the microtubule motor protein dynein. Arp2 and Arp3 are subunits of the Arp2/3 complex – a protein complex that nucleates formation of new actin filaments in response to upstream signaling events (see **Table 1**).

Conventional actin

Actin is one of the most abundant and highly conserved proteins in eukaryotes. In motile cells, actin is typically the most abundant cellular protein and is present at concentrations of 100–200 μM (or 5–10% of total cellular protein). Protozoans,

such as budding and fission yeast and slime molds, generally express one isoform of actin. Metazoans typically express multiple, tissue-specific isoforms. Humans express at least six isoforms of actin: two striated muscle isoforms (α -skeletal and α -cardiac), two smooth muscle isoforms (α -smooth muscle and α -enteric), and two widely expressed cytoplasmic isoforms (β -cytoplasmic and γ -cytoplasmic). The α -skeletal and α -cardiac striated muscle isoforms are expressed specifically in muscle tissues where they form arrays of thin filaments that interdigitate with thick filaments composed of the motor protein myosin. Sliding of the thick and thin filaments past each other drives contraction of the muscle tissue. The thin filaments in muscle tissues are very stable structures whose architecture does not change much over time. The smooth muscle actins form less organized filament arrays involved in contraction of vascular and enteric smooth muscle. The cytoplasmic β - and γ -isoforms are the most widely expressed and form much more dynamic cytoskeletal structures. The β -isoform is ubiquitously expressed and is responsible for building the contractile ring that divides cells at the end of mitosis and for most actin-based intracellular traffic. The assembly of β -cytoplasmic actin also drives cell spreading and the amoeboid, or crawling, motility of many cell types, including those of the nervous and immune systems. The β -isoform is most closely related to protozoan actins. The γ -cytoplasmic isoform is expressed in many different cell types but it is much less abundant than β -cytoplasmic actin. The specific function of γ -actin is unknown.

The defining property of conventional actin is its self-assembly into helical polymeric filaments. Monomeric actin binds ATP with high (nanomolar) affinity but does not hydrolyze it. The transition from monomer to polymer

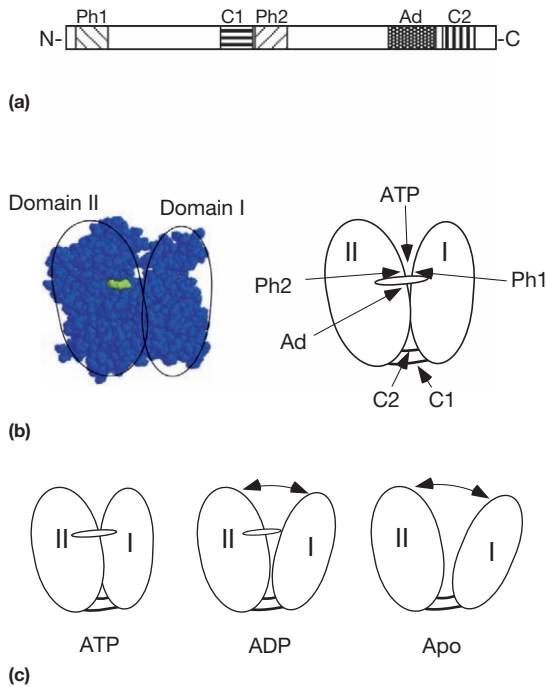


Figure 2 Organization and structure of Arps and the conformational changes driven by ATP binding and hydrolysis. (a) Organization of conserved sequences that define the ATP-binding pocket of the actin fold. Ph1 and Ph2: phosphate-binding loops. C1 and C2: connecting domains. Ad: adenosine-binding domain. The order of these sequences is the same in the primary amino acid sequence of all actin-related proteins. (b) Atomic structure and domain organization of an actin monomer. The atomic structure of actin (left) is shown with a bound ATP molecule (green) and with ellipses identifying the two large domains – domain I and domain II – that define the ATP-binding pocket. The phosphate- and adenosine-binding sequences line the interface between domain I and domain II and interact with the ATP molecule (right). The connecting sequences (C1 and C2) link the two protein domains and act as a hinge. (c) Nucleotide-dependent conformational changes in the actin fold. ATP binding holds the domain I and domain II tightly together (ATP, left). Hydrolysis of the γ -phosphate of ATP to form ADP weakens the interaction with one of the phosphate-binding loops (Ph1) and causes the cleft between the domains to open (ADP, center). Loss of nucleotide from the binding pocket causes the cleft to open even further (Apo, right). Actin-related proteins use this conformational change to regulate their interactions with other proteins.

stimulates the hydrolysis of ATP. Hydrolysis is not required for polymerization but rather for depolymerization. Cleavage of the γ -phosphate and its subsequent dissociation trigger a conformational change that enables binding of actin-depolymerizing factors such as ADF or cofilin (Figure 3(b)).

Arp1

Arp1 is most closely related to conventional actin and, like actin, it assembles into a filamentous polymer. Arp1 forms part of a large (1.2 MDa) protein complex that modulates the activity of the microtubule motor protein dynein. Within the dynactin complex, ~10 Arp1 molecules assemble to form a short (37 nm), actin-like filament. This short filament functions as a scaffold that connects dynein motors to cargo vesicles. The Arp1 filaments bind an isoform of the actin-binding

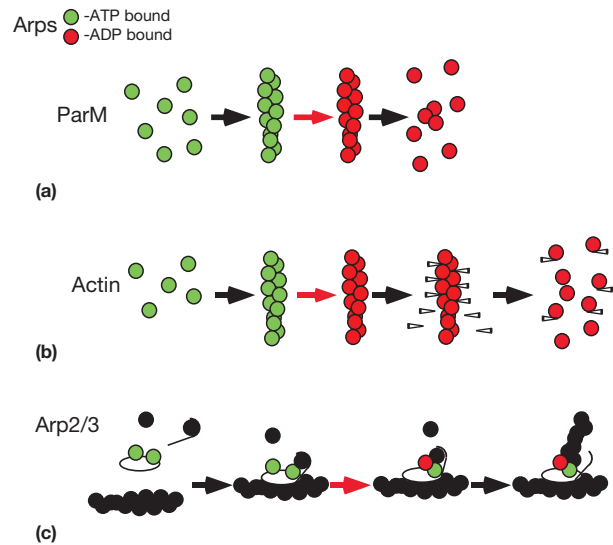


Figure 3 Regulation of actin and Arp function by ATP hydrolysis. The relationship between ATP hydrolysis and protein–protein interactions is best understood in the cytoskeletal Arps. (a) When bound to ATP the plasmid-encoded Arp, ParM (green circles), polymerizes into actin-like filaments. Polymerization induces hydrolysis of bound ATP (red arrow); and the ADP ParM filaments (red circles) undergo a conformational change that causes them to rapidly fall apart. (b) Conventional actin and Arp1 (green circles) polymerize into similar two-stranded helical filaments. Polymerization induces ATP hydrolysis on both proteins (red arrow) but the filaments do not spontaneously disassemble. In the case of conventional actin, a conformational change in the ADP filament structure (red circles) promotes binding of actin depolymerization factors (white wedges), which drive filament disassembly. (c) As part of the Arp2/3 complex, Arp2 and Arp3 bind ATP (green circles are ATP-bound Arp2 and Arp3) and nucleate formation of new actin filaments. The nucleation reaction requires monomeric actin and a pre-formed actin filament (black circles) and an Arp2/3 activator (black hook). Assembly of these cofactors induces ATP hydrolysis on Arp2 and promotes a conformational change in the Arp2/3 complex that converts it into a filament nucleus.

protein spectrin that is specific for the Golgi apparatus and for Golgi-derived vesicles. Arp1 and Golgi spectrin are thought to assemble into a membrane-associated meshwork that surrounds vesicles like a cargo net and allows dynein motors to move them around inside the cell.

In vitro, purified Arp1 polymerizes into short filaments indistinguishable from conventional actin filaments. Arp1 hydrolyzes ATP upon polymerization but this does not have a dramatic effect on filament stability *in vitro*. Like actin, Arp1 may hydrolyze ATP to modulate its affinity for an accessory factor required for filament disassembly (Figure 3(b)).

Arp2 and Arp3

Arp2 and Arp3 are both subunits of a seven-member protein complex, called the 'Arp2/3 complex' that is an essential regulator of the actin cytoskeleton. The Arp2/3 complex nucleates formation of new actin filaments in response to activation of intracellular signaling systems. It also cross-links actin filaments into mechanically rigid networks so that, in general, it functions to convert intracellular signals into cytoskeletal

structures. Within the complex, the Arp2 and Arp3 subunits are thought to form a heterodimer with a surface that mimics the fast-growing, barbed end of an actin filament. This surface forms a template to initiate actin filament formation. In the crystal structure of the complex, Arp2 and Arp3 are separated by a large (40 Å) cleft and formation of an actin-like Arp2–Arp3 dimer would require a significant conformational change. This conformational change may require ATP hydrolysis on one of the Arps (**Figure 3(c)**). Binding of ATP to Arp2/3 is required for actin nucleation and neither ADP nor nonhydrolyzable ATP analogs support Arp2/3-dependent actin filament nucleation. Further study is required to understand the connection between ATP hydrolysis and conformational changes on the Arp2/3 complex.

Chromatin-Associated Arps

Several Arps are associated with eukaryotic chromatin and with chromatin-remodeling complexes. In budding yeast, Arp4, Arp5, Arp7, Arp8, and Arp9 are subunits of ATP-dependent chromatin remodeling complexes. Very little is known about the molecular mechanisms of their activity, but presumably adenosine triphosphatase (ATPase)-driven conformational changes in the Arp subunits regulate the overall structure and activity of these complexes. All of these complexes contain two or four Arp subunits so it is possible that they function as heterodimers and that their interaction is regulated in a manner similar to that of the Arp2/3 complex described earlier.

Arp4

Arp4 is involved in chromatin remodeling and transcriptional activation. In budding yeast, Arp4 is a subunit of both the NuA4 histone acetyl-transferase complex and the INO80 chromatin-remodeling complex. The NuA4 complex acetylates histones H4 and H2A. Acetylation increases accessibility of the DNA within the nucleosome and recruits other chromatin remodeling enzymes to the region. Human cells contain a similar histone acetyl-transferase complex, called 'Tip60'. This complex also contains an Arp4-type subunit in addition to at least one monomer of conventional actin. The INO80 complex contains Arp4 as well as Arp5, Arp8, and conventional actin. The precise role of Arp4 in these complexes is unknown but yeast Arp4 mutants exhibit phenotypes similar to mutants missing the catalytic acetylase subunit of NuA4, suggesting that it plays an integral role in the function of the NuA4 complex. Arp4 has also been shown to bind directly to histones. Its ATPase activity has not been well studied but Arp4 may use ATP binding and hydrolysis to regulate the interaction between chromatin-remodeling complexes and DNA-bound histones.

Arp5 and Arp8

Arp5 and Arp8 (along with Arp4 and conventional actin) are members of the INO80 chromatin-remodeling complex. INO80 is a large (2 MDa) protein complex that plays a role in transcription of several important yeast genes including *pho5*, an essential regulator of phosphate metabolism. The total number of yeast genes that require INO80 activity for transcription is not known. Arp8 has been shown to bind directly to histones H3 and H4 so it may be directly involved in moving histones around on the DNA. Arp8 is also required for the incorporation of both

Arp4 and conventional actin into the complex so it may play an important structural or scaffolding role.

Arp7 and Arp9

Arp7 and Arp9 associate tightly to form a heterodimer and dimerization appears to be important to their cellular function. In budding yeast, the Arp7/Arp9 dimer is an essential part of both the Swi/Snf and RSC chromatin-remodeling complexes. Like the INO80 complex, these large complexes drive DNA and histone rearrangements necessary for transcription. Swi/Snf activity is required for transcription of ~5% of yeast genes while RSC activity is more globally important. As with the other chromatin-associated Arps, the precise functions of Arp7 and Arp9 are unknown. Loss of the RSC complex is lethal. The loss of either Arp7 or Arp9 causes severe growth defects *in vivo* but does not significantly impair RSC activity *in vitro*. So *in vivo* the Arp7/Arp9 dimer may act to correctly target or regulate the RSC complex.

Chromatin-associated conventional actin

The large majority of cellular actin is cytoplasmic and generally actin is considered as a cytoskeletal protein whose function is limited to the cytoplasm. In many species, however, chromatin-remodeling complexes contain one or more tightly associated monomers of conventional actin. In addition to Arp4, the NuA4 complex contains a subunit of conventional actin. Also, the yeast INO80 complex and the mammalian Brg- or hBrm-associated factor (BAF) complex (a chromatin-remodeling complex similar to the budding yeast swi/snf) contain tightly associated actin monomers. The role of conventional actin in chromatin structure and remodeling is mysterious. It has been speculated that actin assembly in the nucleus could drive large-scale changes in chromatin architecture in the way that actin assembly in the cytoplasm drives large-scale plasma membrane movements. It is also possible that short actin filaments form an essential scaffold for the assembly and function of chromatin organizing factors.

Arps of unknown function

We know almost nothing about the function of Arp6. In budding yeast Arp6 is localized to the nucleus so it may be part of a chromatin-remodeling complex or it may play a role in determining nuclear architecture.

See also: **Cell Architecture and Function:** Actin Assembly/Disassembly; Actin-Capping and -Severing Proteins; **Molecular Biology:** Chromatin: Physical Organization; Nuclear Organization, Chromatin Structure, and Gene Silencing; Regulation of Chromatin Dynamics.

Further Reading

- Bork P, Sander C, and Valencia A (1992) An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. *Proceedings of the National Academy of Sciences of the United States of America* 89: 7290–7294.
- Boyer LA and Peterson CL (2000) Actin-related proteins (Arps): Conformational switches for chromatin-remodeling machines? *Bioessays* 22(7): 666–672.

- Carballido-Lopez R and Errington J (2003) A dynamic bacterial cytoskeleton. *Trends in Cell Biology* 13(11): 577–583.
- Goodson HV and Hawse WF (2002) Molecular evolution of the actin family. *Journal of Cell Sciences* 115(Pt 13): 2619–2622.
- Herman IM (1993) Actin isoforms. *Current Opinion in Cell Biology* 5(1): 48–55.
- Olave IA, Reck-Peterson SL, and Crabtree GR (2002) Nuclear actin and actin-related proteins in chromatin remodeling. *Annual Review of Biochemistry* 71: 755–781.
- Pollard T, Blanchoin L, and Mullins RD (2000) Molecular mechanisms controlling actin filament dynamics in nonmuscle cells. *Annual Review of Biophysics and Biomolecular Structure* 29: 545–576.
- Sterner DE and Berger SL (2000) Acetylation of histones and transcription-related factors. *Microbiology and Molecular Biology Reviews* 64(2): 435–459.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, and Higgins DG (1997) *Nucleic Acids Research* 25(24): 4876–4882.

Adenylyl Cyclases

R Taussig, University of Texas Southwestern Medical Center, Dallas, TX, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 40–45, © 2004, Elsevier Inc.

Glossary

G protein A heterotrimeric protein composed of an α -, β -, and γ -subunit that binds and hydrolyzes guanosine triphosphate (GTP) and transduces signals from outside to the inside of the cell.

Receptor A protein that resides in the plasma membrane and specifically recognizes an extracellular signal and upon binding of this signal, transmits information to the cell interior.

Second messenger A small molecule that is synthesized by a cellular protein and conveys the information within the cell that an extracellular signal has been received.

Signal transduction The process of transmitting information from the outside to the inside of a cell, resulting in a change in cellular behavior.

Adenylyl cyclases are enzymes that synthesize the intracellular second messenger, cyclic adenosine monophosphate (cAMP), which in turn triggers a cascade of biochemical changes that regulate a number of important cellular processes. These cAMP-regulated cellular processes play an important role in the control of a number of metabolic processes including the regulation of blood glucose homeostasis, learning and memory, and cell growth.

The Hormone-Regulated Adenylyl Cyclase System

In order to function and survive, cells must respond to a multitude of signals present in their environment. These include chemical messengers released from local or distant parts of the body (such as circulating hormones and neurotransmitters) and external stimuli (such as light and odorants). Typically, these chemical messengers do not cross the cell membrane, but rather transduce their signal to the interior of the cell utilizing a hand-off mechanism occurring at the plasma membrane. In this process, the stimulus or first messenger acts at the membrane to regulate the activity of an enzyme that synthesizes (or degrades) an intracellular second messenger that can in turn regulate downstream intracellular metabolic enzymes (Figure 1). cAMP, which is synthesized by adenylyl cyclases, was the first of these second messengers identified, and this mechanism has subsequently shown to be utilized as a common biological strategy to regulate cellular behaviors.

Regulation of intracellular cAMP concentrations is principally controlled at the level of its synthesis, through the hormonal regulation of adenylyl cyclase, the enzyme responsible for the conversion of adenosine triphosphate (ATP) into cAMP. The adenylyl cyclase system is comprised of three components: heptahelical G-protein-coupled receptors for a variety of hormones, neurotransmitters, and odorants; heterotrimeric G proteins; and the catalytic entity itself. Nine isoforms of membrane-bound adenylyl cyclases and a single soluble form have been identified in man. The nine-membrane-bound isoforms are subjected to hormonal regulation that is mediated by G proteins of the G_s , G_i subclasses. Most importantly, adenylyl

cyclase isoforms have the ability to integrate the multitude of hormonal inputs relayed by these pathways to ultimately control intracellular cAMP levels.

Discovery

The identification of cAMP resulted from studies aimed at elucidating the biochemical mechanism underlying the breakdown of glycogen in liver cells that is triggered by the hormone glucagon. In his Nobel-prize-winning discoveries, Earl Sutherland demonstrated that the process involved two distinct stages, a binding of the hormone at the membrane that activated adenylyl cyclase and subsequently generated the second messenger cAMP, and the activation within the cell of the downstream enzymatic process. The later discoveries that the receptor and the adenylyl cyclase were two distinct entities, and the subsequent identification of a third component (the intervening heterotrimeric G protein) have provided our modern day model of the hormone-regulated adenylyl cyclase system. Almost three decades following the discovery of cAMP, the genetic material encoding the first adenylyl cyclase was isolated, which further exposed the complexity of this signaling system.

Members of the Family

Our current knowledge of the structure and function of the mammalian adenylyl cyclase family has reached a new understanding following the application of molecular approaches. Ten distinct genes encoding the soluble and membrane-bound adenylyl cyclases have been identified (Table 1). These genes are broadly distributed on different chromosomes throughout the human genome. Further diversity of adenylyl cyclases is provided by alternatively spliced messenger RNAs derived from transcripts of some of the family members.

These 10 genes are differentially expressed in various tissues in mammalian species. All isoforms of adenylyl cyclase appear to be expressed in the brain; however, expression of each isoform is limited to specific regions of the central nervous

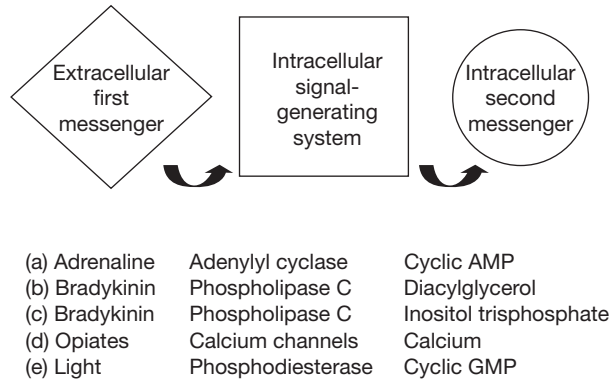


Figure 1 Schematic diagram of second messenger systems. The diagram above outlines the progression of cell signaling from the extracellular signal (first messenger) to the generation of intracellular second messengers. The lettered subheadings below outline specific examples of intracellular second messengers, their extracellular signals, and the cellular proteins that mediate the changes in the levels of the second messengers. Arrows indicate the direction of the flow of information.

system. Additionally, some adenylyl cyclase appears to be rather widely distributed in most tissues of the body (e.g., types 4, 6, and 7). All adenylyl cyclases have roughly the same size and contain between 1064 and 1610 amino acid residues. The deduced amino acid sequence of all identified adenylyl cyclase isoforms predicts a complex topology within the membrane (**Figure 2**). The noted exception is the soluble adenylyl cyclase (SAC), which is localized to the intracellular compartment and exhibits regulation quite distinct from the membrane-bound forms. This isoform is more related to an ancestral form found in bacteria than to the other types.

Based on the similarities of their amino acid sequences, and patterns of regulation by G protein pathways, the membrane-bound adenylyl cyclases can be grouped into subfamilies. The first group, comprised of the types 1 and 8, are characterized by their stimulation by calmodulin. The types 2, 4, and 7 adenylyl cyclases form the second group; characteristic of these isoforms is their stimulation by G protein $\beta\gamma$ -subunits. Types 5 and 6, which are the two most related isoforms constitute the last group and are characterized by their inhibition by calcium. The type 9 is the largest and most diverse of the adenylyl cyclase isoforms, and like the type 3 enzyme, does not belong to any of the three groups outlined above.

Table 1 Localization and expression of the adenylyl cyclase gene family

<i>AC isoform</i>	<i>Length (amino acids)</i>	<i>Location (chromosome)</i>	<i>Sites of expression (tissue)</i>
AC1	1135	7	Brain and adrenal medulla
AC2	1090	5	Brain, skeletal muscle, lung, and heart
AC3	1144	2	Olfactory epithelium, and brain
AC4	1064	14	Widely distributed
AC5	1184	3	Heart, brain, kidney, liver, lung, and uterus
AC6	1165	12	Ubiquitous
AC7	1099	16	Ubiquitous (highly expressed in brain)
AC8	1248	8	Brain and lung, (some in testes, adrenal, uterus, and heart)
AC9	1353	16	Brain, skeletal muscle, lung, and heart
SAC	1610	1	Testes

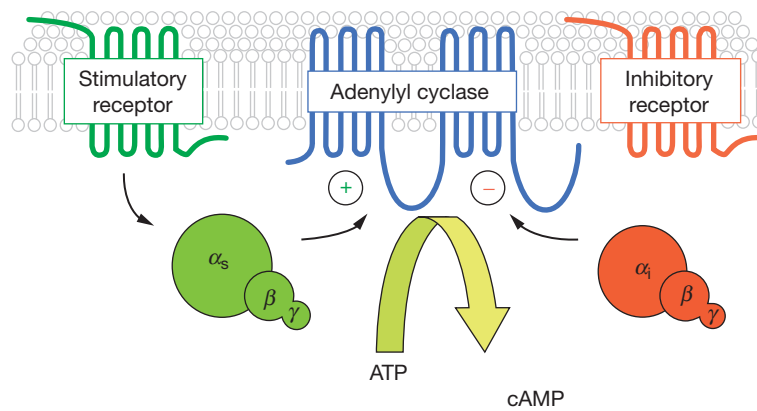


Figure 2 Schematic view of plasma membrane components that participate in the hormone-regulated adenylyl cyclase system. A prototypic adenylyl cyclase is shown. G-protein-coupled receptors that enhance or inhibit adenylyl cyclase activity are shown in green and red, respectively. All receptors of this family have a similar predicted structure and span the plasma membrane seven times. These receptors couple to their corresponding heterotrimeric G proteins (composed of α -, β -, and γ -subunits); G_s (green) and G_i (red) couple to stimulatory and inhibitory receptors, respectively. Adenylyl cyclase (shown in blue) is predicted to span the plasma membrane 12 times, and projects its active site toward the cytoplasmic side of the membrane. The result of extracellular messengers binding to the receptors results in the regulation of the conversion of ATP into cAMP by the enzyme adenylyl cyclase.

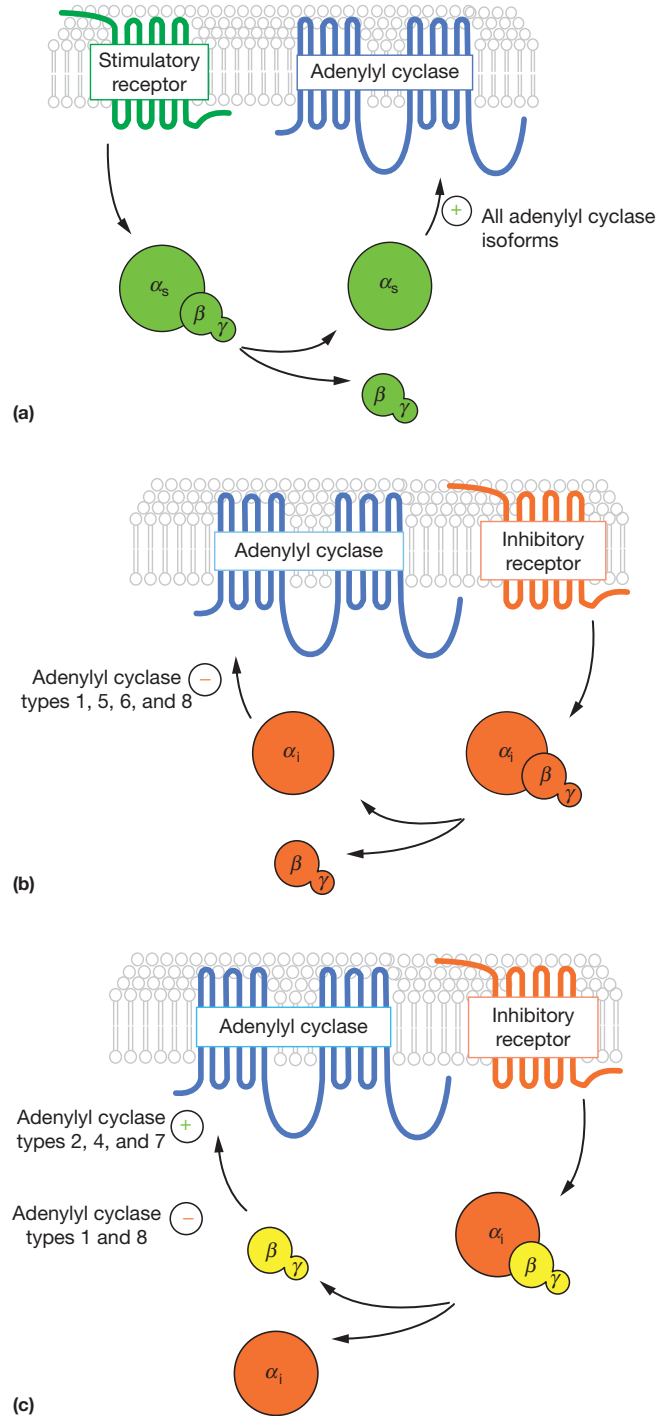


Figure 3 Schematic representation of the isoform-specific regulation of adenylyl cyclases by G protein subunits. (a) Regulation of adenylyl cyclases by $G_s\alpha$. Upon activation of stimulatory receptors, the heterotrimeric G protein, G_s undergoes subunit dissociation whereby the $\beta\gamma$ -subunits are released from the α -subunit. The α -subunit subsequently binds to the adenylyl cyclase and increases its catalytic activity, resulting in an increase in intracellular cAMP levels (+). (b) Regulation of adenylyl cyclases by $G_i\alpha$. Upon activation of inhibitory receptors, the heterotrimeric G protein, G_i undergoes subunit dissociation whereby the $\beta\gamma$ -subunits are released from the α -subunit. The α -subunit subsequently binds to the adenylyl cyclase and decreases its catalytic activity, resulting in a decrease in intracellular cAMP levels (-). Specific isoforms of adenylyl cyclase that are responsive to G_i inhibition are noted. (c) Dual regulation of adenylyl cyclases by G-protein $\beta\gamma$ -subunits. $\beta\gamma$ -subunits released from G_i heterotrimers can either positively (+) or negatively (-) regulate adenylyl cyclases. Specific isoforms of adenylyl cyclase that are responsive to these two types of regulation are noted.

Regulation of Adenylyl Cyclase Activity by G Proteins and Calcium

There are multiple classes of heterotrimeric G proteins that regulate adenylyl cyclases, either in a stimulatory (G_s family) or in an inhibitory (G_i family) manner. The two G protein families are normally coupled to distinct receptor subtypes. The $\beta\gamma$ -subunits also regulate adenylyl cyclase activity, but in a manner specific to an adenylyl cyclase isoform. Additionally, calcium ions are very strong modulators of some isoforms of adenylyl cyclase; thus, G proteins that regulate calcium entry through voltage-dependent Ca^{2+} channels may also regulate adenylyl cyclase activity.

Hormonal activation of AC occurs primarily through receptors coupled to the stimulatory G protein G_s . G_s is the most widely distributed activator of all mammalian membrane-bound AC isoforms. All membrane-bound forms of adenylyl cyclases are regulated by G_s through the interaction of adenylyl cyclase with the $G\alpha$ -subunit that is released from the $\beta\gamma$ -subunits upon activation by a hormone-bound stimulating receptor (Figure 3(a)).

Members of the G_i family inhibit adenylyl cyclase activity through the binding of the alpha subunit to the cyclase, but this regulation exhibits selectivity for a subset of adenylyl cyclase isoforms (Figure 3(b)). All three G_i family members (G_{i1} , G_{i2} , and G_{i3}) inhibit types 1, 5, 6, and 8; the remaining cyclase isoforms are insensitive to this inhibitory input. Interestingly, this mode of inhibition is not through direct competition with G_s binding to the cyclase, and a number of experimental approaches demonstrate that $G_{i\alpha}$ exerts its effects at a site, symmetrical to the G_s binding site, and located on the side opposite the cyclase molecule.

As illustrated in Figure 3(c), activation of inhibitory receptors leads to the release of G protein $\beta\gamma$ -subunits which can exert profound regulation of adenylyl cyclase activity. Like

regulation by the G_i α -subunit, this regulation is isoform specific. However, an added complication is that the $\beta\gamma$ -subunits can either activate or inhibit the activity – types 2, 4, and 7 are activated whereas types 1 and 8 are inhibited.

Regulation of adenylyl cyclase activity by calcium is still more complex (Figure 4). At physiological intracellular calcium concentrations the types 3, 5, and 6 adenylyl cyclases are inhibited by calcium. For the types 5 and 6, this appears to be due to a direct effect of calcium on the adenylyl cyclase. Stimulation of adenylyl cyclase activity by calcium has been demonstrated for the types 1 and 8 isoforms, but this form of regulation requires calmodulin, a calcium-binding protein that regulates the activity of many cellular enzymes. Indeed, one of these calmodulin-regulated enzymes is a protein kinase, that mediates the inhibitory effects that were observed for the type 3 adenylyl cyclase.

Physiological Roles of Adenylyl Cyclases

The hormonal activation of adenylyl cyclases and the subsequent regulation of intracellular cAMP impacts on a wide array of cellular processes (Table 2). These roles include diverse processes underlying regulation of blood sugar levels, regulation of heart function, water retention as well as higher brain functions of learning and memory. However, while much is known at the biochemical level regarding the regulation of the many adenylyl cyclase isoforms, the precise physiological roles of each isoform are less well known. A further complication is that each cell of our body expresses more than one isoform of adenylyl cyclase. Nevertheless, investigators have begun to examine the specific roles of each adenylyl cyclase isoform through the use of mouse strains that are genetically altered by the removal of a particular adenylyl cyclase gene from the mouse genome. Such studies emphasize the involvement of

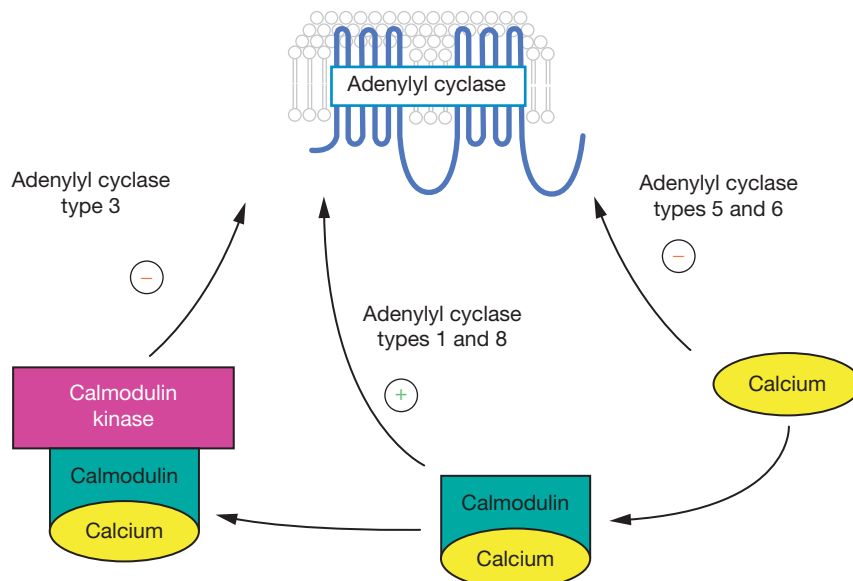


Figure 4 Schematic representation of the isoform-specific regulation of adenylyl cyclases by calcium. Increases in intracellular calcium can regulate the activities of adenylyl cyclases by at least three mechanisms. Calcium can directly bind to and inhibit AC 5 and 6. Calcium bound to calmodulin can bind directly to and activate AC 1 and 8, or can bind and activate a calmodulin-regulated kinase that can subsequently inhibit AC3.

Table 2 Examples of physiological effects mediated by adenylyl cyclases

<i>Extracellular signal (hormone or chemical)</i>	<i>Tissue</i>	<i>Isoform (if known)</i>	<i>Effect</i>
Adrenaline	Heart	AC5 and 6	Increased rate and force of contraction
Adrenaline glucagon	Liver	Unknown	Breakdown of glucagon into glucose
Neurotransmitters (e.g., serotonin and dopamine)	Brain	Many	Learning, memory, and synaptic transmission
Chemical odorants	Olfactory epithelium	AC3	Odorant sensation
Antidiuretic hormone	Kidney	Unknown	Regulation of water retention
Prostaglandin	Arteries	AC3	Inhibition of smooth muscle cell proliferation
Bicarbonate	Testes	SAC	Fertilization (sperm acrosome reaction)

cAMP-signaling pathways in pattern formation of the brain and provide definitive evidence for roles of the types 1 and 8 adenylyl cyclases in higher brain function. Similar approaches have revealed a role for the type 3 cyclase in regulation of smooth muscle proliferation in the aorta and, recently, a role of the type 5 cyclase in heart function.

A number of studies have uncovered both sporadic and inherited mutations in components of the adenylyl cyclase system associated, or in some cases, shown to be causal to certain human diseases. Mutations causing constitutively active receptors resulting in elevated intracellular cAMP levels due to constant stimulation of adenylyl cyclase via G_{sz} have been found in patients with familial male precocious puberty/testotoxicosis (LH receptor), hyperfunctioning thyroid adenomas and nonautoimmune autosomal dominant hyperthyroidism (thyroid-stimulating hormone receptor), and Jansen-type metaphyseal chondrodysplasia (parathyroid hormone receptor). Similarly, diseases associated with constitutively activated G proteins (G_{sz}) were also identified in patients with endocrine tumors, McCune–Albright syndrome, and testotoxicosis. However, to date no mutations associated with these diseases have been found in the adenylyl cyclase component of this system, but these investigations are ongoing. Here again, a major

complication of these studies is the presence of multiple adenylyl cyclase isoforms that are present in each of the cell types.

See also: **Signaling:** Adrenergic Receptors; Cyclic GMP Phosphodiesterases; Cyclic Nucleotide Phosphodiesterases; Gi Family of Heterotrimeric G Proteins; Gs Family of Heterotrimeric G Proteins; Phospholipase C.

Further Reading

- Hanoune H and Defer N (2001) Regulation and role of adenylyl cyclase isoforms. *Annual Review of Pharmacology and Toxicology* 41: 145–174.
- Johnston CA and Watts VJ (2003) Sensitization of adenylate cyclase: A general mechanism of neuroadaptation to persistent activation of G $\alpha(i/o)$ -coupled receptors? *Life Sciences* 73: 2913–2925.
- Sunahara RK and Taussig R (2002) Isoforms of mammalian adenylyl cyclase: Multiplicities of signaling. *Molecular Interventions* 2: 168–184.
- Tesmer JJ and Sprang SR (1998) The structure, catalytic mechanism and regulation of adenylyl cyclase. *Current Opinion in Structural Biology* 8: 713–719.
- Wuttke MS, Buck J, and Levin LR (2001) Bicarbonate-regulated soluble adenylyl cyclase. *Journal of the Pancreas* 2: 154–158.

Adipocyte Fat Mobilization: The Regulatory Roles of Perilipin 1, Adipose Triglyceride Lipase, and Hormone-Sensitive Lipase

K T Dalen, University of Oslo, Oslo, Norway

C Londos[†] and A R Kimmel, National Institutes of Health, Bethesda, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Adipocytes Mammalian cells that store most of the whole body energy reserves in triacylglycerides (TAGs), commonly referred to as fat cells.

Adipocyte triglyceride lipase (ATGL/PNPLA2) An enzyme that catalyzes the rate-limiting step in TAG hydrolysis in adipocytes.

Comparative gene identification-58 (CGI-58) or α/β -hydrolase domain containing 5 (ABHD5)
A protein required to maximally activate ATGL.

Hormone-sensitive lipase (HSL) A hormonally regulated enzyme primarily responsible for catalyzing diacylglyceride hydrolysis within adipocytes.

Lipid droplets (LDs) Organelles with a hydrophobic inner core of neutral lipids, which serve as energy reserves and metabolic precursors.

Lipolysis The degradation of TAGs into free fatty acids and glycerol.

Perilipin 1 (Plin1) The most abundant Plin in adipocytes.

Perilipin protein family Plin proteins that share common amino acid sequence motifs and target LD surfaces.

Introduction

Nature's solution for packaging energy reserves in organisms as diverse as plants and mammals is to sequester energy-rich fatty acids within intracellular neutral lipid storage droplets in the form of triacylglycerides (TAGs) or cholesteryl esters (CEs). The great survival value for this type of packaging is attributed to the high density of energy in fatty acids and the relatively low requirement of water for their storage. If humans were to carry the energy equivalent of fat as carbohydrate, the additional required water would render them immobile. Typically, intracellular neutral lipid storage droplets are coated with specific targeting classes of proteins; proteins involved in lipid metabolism may also be found in tight association. We discuss the functional interactions among these protein and their cooperative roles in the activation and suppression of lipolysis, TAG hydrolysis, in mammalian adipocytes.

Lipid Droplets

The first reports describing the existence of lipid droplets (LDs) within cells dates to the invention of the light microscope. Such LDs were long considered simple reservoirs of energy, and garnered little attention. Studies of proteins binding to lipids in humans initiated with the discovery of lipoproteins, circulating lipids in plasma. Later, oil bodies, a terminology used in plants, were described together with their coat proteins, the oleosins. The oleosins protect the deposited lipids during desiccation of plant seeds and tapeta. In animal cells, the LD coat proteins help sequester lipids from the cytosol and regulate their access to lipases.

Intracellular LD-Binding Proteins of Animal Cells

The filament protein Vimentin was the first protein found to associate closely with intracellular LDs in animal cells. Soon thereafter, Perilipin was identified at the LD surface in rat adipocytes. Perilipin has no obvious sequence relationship to vimentin or the oleosins, but is closely related to other animal proteins, which were subsequently found to also localize with intracellular LDs (Figure 1(a), Table 1). Based upon the sequence similarities among these proteins, their shared subcellular localization properties, their similar gene structure organizations, and the multiple names attributed to several of these, a universal nomenclature has been adopted for this gene family, with Perilipin/Plin chosen as the root name and symbol (Table 1). Perilipin is now termed perilipin 1, Plin1.

Genes that encode species-related proteins have also been identified in evolutionary diverse organisms, including *Drosophila melanogaster* and *Dictyostelium discoideum* (Figure 1(a)). These genes are related to the mammalian counterparts not only in their protein-coding sequences, but also in their overall genomic structural organizations, indicating that these genes derive from a gene family of ancient origin. Similar to the mammalian Plins, the *Drosophila* and *Dictyostelium* proteins localize to LDs endogenously or when expressed in mammalian cells. All of these proteins are presumed to share specific structural motifs that determine their unique intracellular LD targeting ability, but the exact nature of this motif is still to be experimentally determined.

Perilipin 1, the Major Adipocyte Perilipin

The most extensively studied perilipin is perilipin 1, Plin1. *Plin1* is a single-copy gene in mice and humans that encodes multiple messenger RNAs (mRNAs) and protein forms (Figure 1(b)). The *Plin1a*, *b*, *c*, *d* mRNAs encode proteins with identical

[†]Deceased.

amino-termini, but distinct carboxyl-termini. All of the Plin1 proteins localize to the surface of neutral lipid deposits (Figure 2), but it is not clear if Plin1 binds at the surface or, like oleosins, is embedded within the phospholipid monolayer, which forms the boundary between the aqueous cytosol and the hydrophobic inner lipid core of the droplet.

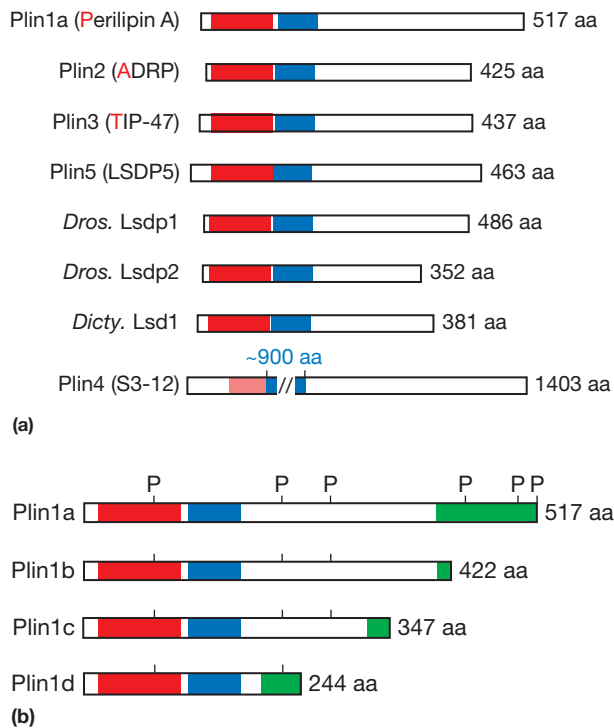


Figure 1 (a) Schematic of the major Plin family members. There are five Plin family members in mice and humans, Plins 1–5; alternative names are indicated. The PAT domain name derives from the founding members of the family, Perilipin, ADRP, and TIP47. Only the major Plin1 variant is shown (see (b)). Related lipid storage droplet (LSD) proteins have been identified in *Drosophila*, *Dictyostelium*, and other species. Red boxes indicate the N-terminal PAT domain shared by all members of the Plin family; in Plin4, this region is highly diverged. Blue boxes indicate the 11-mer repeat region also shared by members of the family; in Plin4, this region is more extended. (b) Schematic of Perilipin 1 isoforms and PKA phosphorylation sites. *Perilipin 1* (*Plin1*) is a single-copy gene that encodes multiple mRNAs as the result of differential splicing and poly(A) addition sites. The four Plin1 proteins are diagramed and the sites for phosphorylation by PKA indicated. Red boxes indicate the N-terminal PAT domain shared by all members of the Plin family. Blue boxes indicate the 11-mer repeat region also shared by members of the Plin family. Green boxes indicate regions unique to each Plin1 protein isoform.

Plin1 is preferentially expressed in adipocytes and steroidogenic cells. Only Plin1a and b are expressed in adipocytes, with Plin1a being more abundant. Plin1c and Plin1d are exclusively expressed in steroidogenic cells. Both Plin1a and 1b are hyperphosphorylated upon elevation of cyclic adenosine monophosphate (cAMP) during lipolytic stimulation of adipocytes. Plin1a has six consensus protein kinase A (PKA) phosphorylation sites (Figure 1(b)); the three carboxyl-terminal sites are absent from Plin1b. It is not known if all of the PKA sites in Plin1 are phosphorylated or if phosphorylation occurs in a specific order. As lipolysis in adipocytes is activated by a PKA-mediated process, and given the specific association of Plin1 with adipocyte LDs, it was early predicted that Plin1 would play an essential role in the regulation of adipocyte lipolysis. Functional analyses in cell culture systems and in mice lacking expression of all Plin1 forms confirm this early speculation.

The Dual Nature for Plin1 Regulation of Lipolysis in Adipocytes

Adipocytes store most of the whole body energy depots in TAG-containing LDs. To maintain these energy reserves, yet allow their release under restricted conditions, lipolysis in adipocytes is subject to complex regulation. Depending on its PKA-dependent phosphorylation state, Plin1 exhibits two opposing functions in the regulation of lipolysis in adipocytes. Nonphosphorylated Plin1 appears to shield TAGs within LDs from hydrolysis by lipases. By contrast, when phosphorylated by PKA, Plin1 facilitates lipase-mediated TAG hydrolysis. The protective function of Plin1 is best illustrated by the *Plin1*^{−/−} mouse that exhibits a ~70% decrease in adipose tissue without a change in gross animal weight or in caloric intake. This reduction in adipose tissue mass is attributed to the constitutively elevated rate of basal lipolysis in adipocytes lacking Plin1. Thus, Plin1 is essential for mice to retain their TAG depots. These mice also reveal a critical role for Plin1 in stimulated lipolysis. In the absence of Plin1, adipocytes fail to respond maximally to lipolytic stimuli. In comparison to *Plin1*^{+/+} wild-type (WT) cells, PKA-activated *Plin1*^{−/−} adipocytes have attenuated lipolysis. This dual nature for Plin1 function has been mechanistically confirmed by expressing WT and phospho-site mutant forms of Plin1a and Plin1b in fibroblastic cells which do not normally express Plin1.

Dissecting the Adipocyte Lipolysis Model

For many years, the prevailing model for adipocyte lipolysis was that hormone-sensitive lipase (HSL) was the rate-limiting

Table 1 The human/mouse Perilipin (Plin) protein family

Approved human symbol	Approved name	Previous aliases	Primary tissue protein expression
PLIN1	Perilipin 1	Perilipin	Adipose, steroidogenic cells
PLIN2	Perilipin 2	ADRP, ADFP, adipophilin	Ubiquitous enriched in liver, muscle, fibroblasts, not adipose
PLIN3	Perilipin 3	TIP47, pp17	Ubiquitous
PLIN4	Perilipin 4	S3-12	Adipose, heart, glycolytic muscle
PLIN5	Perilipin 5	LSDP5, PAT1, OXPAT, MLDP	Heart, oxidative muscle, brown fat

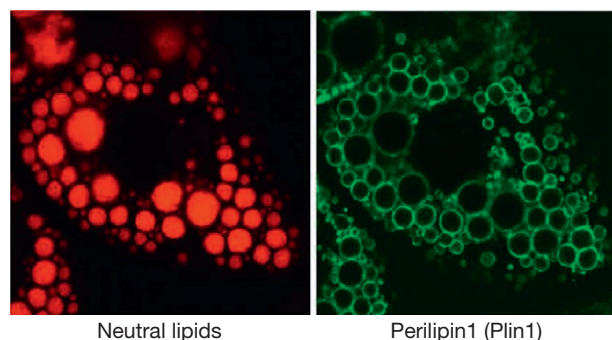


Figure 2 Co-localization of Plin1 with lipid droplets of adipocytes. Isolated rat adipocytes were stained for neutral lipids with Nile red and immunostained for Perilipin1 protein with α -Plin1. Plin1 localizes to the surface of intracellular LDs.

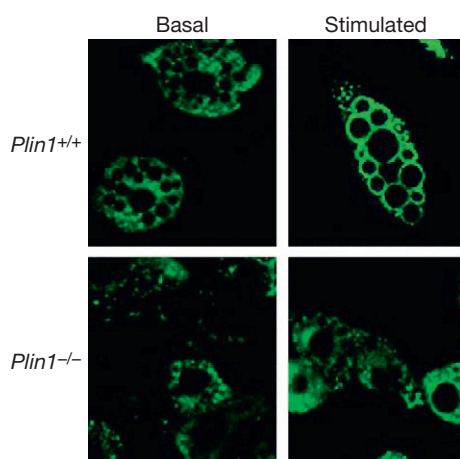


Figure 3 Stimulated migration of HSL from the cytosol to lipid droplet surfaces requires Perilipin 1. Adipocytes were isolated from *Plin1*^{+/+} or *Plin1*^{-/-} mice. Cells were maintained in a basal state or lipolytically stimulated with isoproterenol. Subcellular localization of hormone-sensitive lipase (HSL) was determined by immunostaining. HSL is cytosolic in basal cells. Upon stimulation, HSL redistributes to the surface of LDs in adipocytes of *Plin1*^{+/+} mice but not of *Plin1*^{-/-} mice.

enzyme and that lipolytic rates *in vivo* were regulated exclusively by the PKA-dependent enzymatic activation of HSL. However, this mechanism was questioned because phosphorylation of HSL *in vitro* elicits only a doubling of lipase activity, whereas activation of cellular PKA *in vivo* increases lipolysis 50–100-fold. Further, whereas HSL is localized in the cytosol of quiescent adipocytes, the enzyme is absent in the cytosol of homogenates of stimulated cells, but is associated with LDs (Figure 3). It was further demonstrated that PKA activation led to phosphorylation-dependent translocation of HSL from the cytosol to the surface of LDs.

An important breakthrough in the understanding of HSL translocation came from studies of adipocytes isolated from *Plin1*^{-/-} animals. In the absence of Plin1, HSL cannot translocate from the cytosol to LD surfaces (Figure 3). This finding established HSL translocation, and not phosphorylation *per se*, as a key reaction for stimulated lipolysis in adipocytes and revealed the central role of Plin1 in this process. Translocation

has been reconstituted in fibroblast cells expressing WT and mutant forms of Plin1 and HSL. PKA-dependent phosphorylation of WT HSL and WT Plin1 mediates the translocation process; phospho-HSL does not translocate efficiently in the absence of Plin1 or in the presence of Plin1 variants that cannot be phosphorylated. Similarly, nonphosphorylated HSL does not translocate to LDs even in the presence of phospho-Plin1. This model of translocation of lipases from the cytosol to the LD surface upon PKA activation forms the basis for subsequent work and identification of additional regulatory components in adipocyte lipolysis.

The Current Adipose Lipolysis Model

Based on the above model (see Figure 3), mice lacking HSL posed a surprising phenotype. Instead of having a presumed increase in adipocyte TAG levels, adipocytes from mice lacking HSL were smaller, with less TAG but increased diacylglyceride (DAG) in comparison to WT controls. Although HSL has broad substrate specificity and is capable of hydrolyzing TAG, DAG, monoacylglyceride (MAG), CE, and retinyl ester, TAG is the least favorable substrate. Subsequently, another lipase, adipose triglyceride lipase (ATGL), was identified in adipocytes (and other cells) that exhibited substantially higher substrate specificity for TAG. Loss of ATGL reduced TAG hydrolysis in adipocytes and promoted abnormal accumulation of TAG in many other cells. Indeed, ATGL is a widely expressed enzyme that is the more critical for rate-limiting regulation of lipolysis in multiple tissues. Maximal activity of ATGL requires specific interaction with the α/β -hydrolase domain containing protein CGI-58/ABDH5.

The significance of ATGL and CGI-58 in lipid metabolism is underscored by recent observations that autosomal recessive mutations in either human gene are causative for neutral lipid storage diseases. CGI-58/ABDH5 deficiency in humans is the basis of Chanarin–Dorfman syndrome (CDS), characterized by systemic accumulation of TAG, hepatic steatosis, and ichthyosis. Mutations in ATGL/PNPLA2 are associated with NLSMD, a neutral lipid storage disease with myopathy, and a distinct pattern of TAG accumulation compared with that of CDS. Although CGI-58 is a definitive positive regulator of ATGL, these phenotypic differences suggest that the actions of CGI-58 may not be limited to ATGL and that CGI-58 may regulate lipolytic activity in tissues more broadly than does ATGL. In addition, neither disease is characterized by gross obesity, indicating a central role for HSL, in addition to ATGL, in the regulation of fat mobilization in adipocytes. It is also likely that ATGL activity is subject to negative regulation and recent data implicate G₀/G₁ Switch Gene 2 (G0S2) as a potential inhibitory molecule that can suppress ATGL lipase activity.

Clearly, the activity of the lipolytic enzymes and adipose lipolysis is tightly regulated. There are multiple regulatory cellular inputs, with PKA-mediated signaling being a highly significant pathway. Adipocytes sense body energy reserves in response to circulating hormones. Some act directly on G-protein-coupled receptors to suppress or stimulate adenylyl cyclase activity, with a consequent effect on downstream cAMP production and activity of PKA.

The current model for adipose PKA-mediated lipolysis is depicted in **Figure 4**. In quiescent cells adenylyl cyclase is functionally inhibited (see **Figure 4(a)**); cAMP levels are low, PKA is minimally active, and Plin1 is unphosphorylated. Under these conditions, Plin1 serves as a functional barrier for HSL and ATGL to access TAGs within the inner core of the intracellular LDs in adipocytes. CGI-58 is bound to the LD surface, in complex with the carboxyl-terminal region of unphosphorylated Plin1, which may sequester CGI-58 from interaction with ATGL.

Upon hormonal stimulation and adenylyl cyclase activation (see **Figure 4(b)**), cAMP levels rise, leading to the activation of PKA and the phosphorylation of both Plin1 and HSL. Phosphorylation of Plin1 facilitates binding of HSL to the LD surface. In addition, phosphorylation of HSL by PKA is also required for translocation of HSL from the cytosol to the LDs.

Although phosphorylated Plin1 may serve as a docking site for phosphorylated HSL, this is not fully clarified. Unphosphorylated Plin1 may sterically restrict access of HSL and other lipases from LD surfaces. Phosphorylation of Plin1 also regulates ATGL. In quiescent cells, unphosphorylated Plin1 is complexed with CGI-58 at the LD surface, but phosphorylation of Plin1 is coordinated with the dissociation of CGI-58. This renders CGI-58 accessible for interaction with and activation of ATGL. Released CGI-58 may bind to ATGL in the cytosol and the activated complex may then bind LDs; alternatively, released CGI-58 may remain bound to the LD surface and recruit ATGL to the LD. There are data to support both mechanisms.

In summary, Plin1 serves as a functional barrier to lipase access to TAGs in quiescent cells, but PKA stimulation initiates a cascade of events that ultimately activate lipolysis. Both Plin1

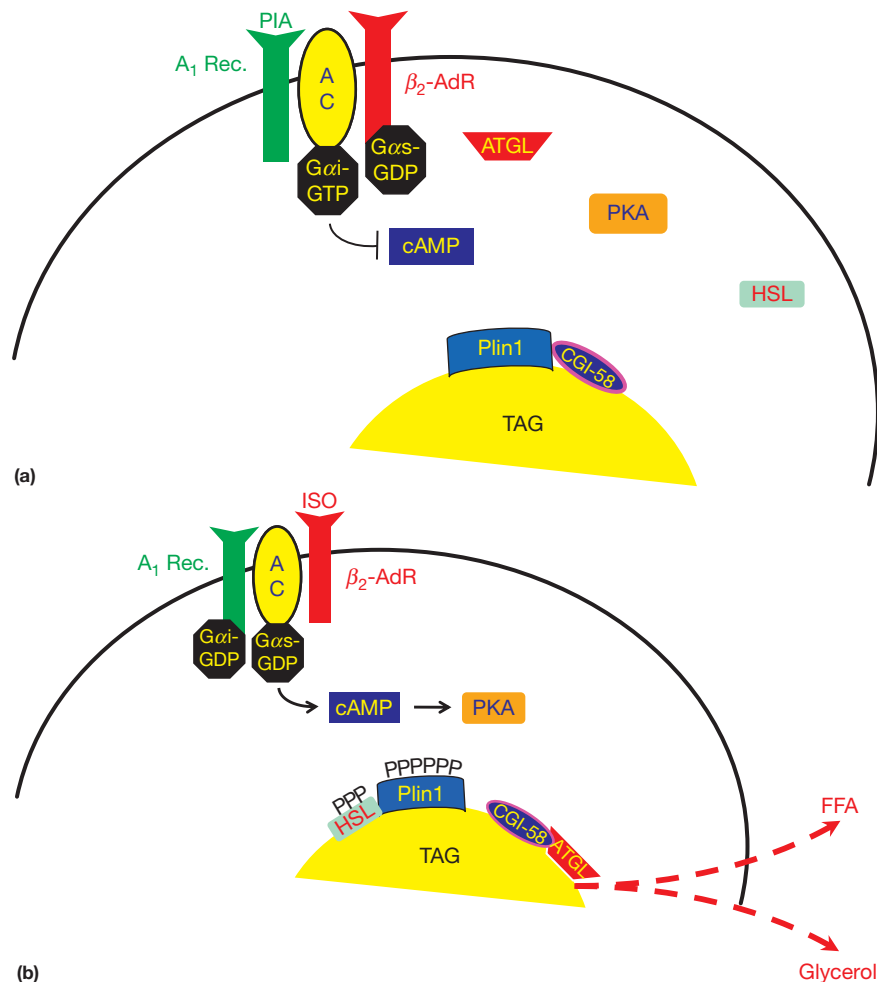


Figure 4 Model for PKA-mediated lipolytic activation in adipocytes. (a) Quiescent cells: adenylyl cyclase (AC) is inactive through coupling to inhibitory $G\alpha$ subunits. Here, the adenosine analog phenylisopropyladenosine (Pia) activates adenosine type 1 receptors, which promotes formation of the $G\alpha i$ -GTP inhibitory complex. Intracellular cAMP levels are suppressed and protein kinase A (PKA) is inactive. At the LD surface, Plin1 (Perlipin 1) interacts with CGI-58; hormone-sensitive lipase (HSL) and adipose triacylglyceride lipase (ATGL) are primarily cytosolic. (b) Activated cells: adenylyl cyclase (AC) is activated through interaction with stimulatory $G\alpha$ subunits. Here, the β_2 -adrenergic receptor agonist isoproterenol (Iso) promotes formation of the $G\alpha s$ -GTP stimulatory complex. Intracellular cAMP levels rise and protein kinase A (PKA) is activated. In turn, PKA phosphorylates both Plin1 and HSL. At the LD surface, phospho-Plin1 recruits phospho-HSL and CGI-58 is released. CGI-58 couples with and activates ATGL at the LD surface. Accordingly, triacylglycerides become hydrolyzed to free fatty acids (FFA) and glycerol and are secreted.

and HSL are phosphorylated by PKA; functionally activated ATGL and HSL translocate to LD surfaces, and lipolytic rates are stimulated 50–100-fold above basal levels. Thus, in *Plin1*^{-/-} adipocytes, basal rates of lipolysis are elevated due to loss of barrier function, but activated lipolysis is attenuated due to loss of lipase recruitment.

The Function of Other Plins in the Regulation of Lipid Metabolism in Various Tissues and Organisms

The shared structural organization of the Plins suggests that all Plin family members may regulate LD function. Each of the various mammalian Plins have distinct expression patterns and are likely to regulate lipolysis and, potentially other processes, in a distinct manner. Plin2, although usually absent in differentiated adipocytes, is the predominant protein found at LD surfaces of *Plin1*^{-/-} adipocytes. Certainly, the appearance of Plin2, in addition to the loss of Plin1, may contribute to the lipolytic defects in these cells.

A general role for the Plins in lipid metabolism is emerging. Plins 2, 3, and 5, can partially protect TAG from lipases; in their absence, fatty acids are released at an elevated level. The increased release of fatty acids from LDs interferes with cellular insulin signaling and can induce insulin resistance in cultured liver cells. Thus, the Plins may sequester TAG from lipase action and help maintain insulin sensitivity. Recent data in *Drosophila* extend these observations. *Drosophila* lacking one of their two Plin-related proteins exhibit a lean phenotype, while overexpression promotes an increase in TAG stores, phenotypes consistent with a general protective function for the Plins in lipid storage.

Acknowledgment

This work was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.

See also: **Bioenergetics:** P-Type Pumps; Copper Pump; **Signaling:** Adenylyl Cyclases; Gi Family of Heterotrimeric G Proteins; Gs Family of Heterotrimeric G Proteins.

Further Reading

Bell M, Wang H, Chen H, et al. (2008) Consequences of lipid droplet coat protein downregulation in liver cells: Abnormal lipid droplet metabolism and induction of insulin resistance. *Diabetes* 57: 2037–2045.

- Bickel PE, Tansey JT, and Welte MA (2009) PAT proteins, an ancient family of lipid droplet proteins that regulate cellular lipid stores. *Biochimica et Biophysica Acta* 1791: 419–440.
- Blanchette-Mackie EJ, Dwyer NK, et al. (1995) Perilipin is located on the surface layer of intracellular lipid droplets in adipocytes. *Journal of Lipid Research* 36: 1211–1226.
- Brasaemle DL (2007) Thematic review series: Adipocyte biology. The perilipin family of structural lipid droplet proteins: Stabilization of lipid droplets and control of lipolysis. *Journal of Lipid Research* 48: 2547–2559.
- Brasaemle DL, Barber T, Wolins NE, et al. (1997) Adipose differentiation-related protein is an ubiquitously expressed lipid storage droplet-associated protein. *Journal of Lipid Research* 38: 2249–2263.
- Dalen KT, Dahl T, Holter E, et al. (2007) LSDP5 is a PAT protein specifically expressed in fatty acid oxidizing tissues. *Biochimica et Biophysica Acta* 1771: 210–227.
- Grönke S, Beller M, Fellert S, et al. (2003) Control of fat storage by a *Drosophila* PAT domain protein. *Current Biology* 13: 603–606.
- Greenberg AS, Egan JJ, Wek SA, et al. (1991) Isolation of cDNAs for perilipins A and B: Sequence and expression of lipid droplet-associated proteins of adipocytes. *Proceedings of the National Academy of Sciences of the United States of America* 90: 12035–12039.
- Kimmel AR, Brasaemle DL, McAndrews-Hill M, Szalryd C, and Londos C (2010) Adoption of PERILIPIN as a unifying nomenclature for the mammalian PAT-family of intracellular lipid storage droplet proteins. *Journal of Lipid Research* 51: 468–471.
- Lefèvre C, Jobard F, Caux F, et al. (2001) Mutations in CGI-58, the gene encoding a new protein of the esterase/lipase/thioesterase subfamily, in Chanarin–Dorfman syndrome. *American Journal of Human Genetics* 69: 1002–1012.
- Lu X, Gruia-Gray J, Copeland NG, et al. (2001) The murine *Perilipin* gene: The lipid droplet-associated Perilipins derive from tissue-specific, mRNA splice variants and define a gene family of ancient origin. *Mammalian Genome* 12: 741–749.
- Miura S, Gan J-W, Brzostowski J, et al. (2002) Functional conservation for lipid storage droplet association among Perilipin, ADRP, and TIP47 (PAT)-related proteins in mammals, *Drosophila* and *Dictyostelium*. *Journal of Biological Chemistry* 277: 32253–32257.
- Osuga J, Ishibashi S, Oka T, et al. (2000) Targeted disruption of hormone-sensitive lipase results in male sterility and adipocyte hypertrophy, but not in obesity. *Proceedings of the National Academy of Sciences of the United States of America* 97: 787–792.
- Radner FP, Streith IE, Schoiswohl G, et al. (2010) Growth retardation, impaired triacylglycerol catabolism, hepatic steatosis, and lethal skin barrier defect in mice lacking comparative gene identification-58 (CGI-58). *Journal of Biological Chemistry* 285: 7300–7311.
- Szalryd-Woodle C, Xu G, Tansey JT, et al. (2003) Perilipin is essential for the translocation of hormone-sensitive lipase during lipolytic activation. *Journal of Cell Biology* 161: 1093–1103.
- Tansey JT, Huml AM, Vogt R, et al. (2003) Functional studies on native and mutated forms of Perilipins: A role in protein kinase A-mediated lipolysis of triacylglycerols in Chinese hamster ovary cells. *Journal of Biological Chemistry* 278: 8401–8406.
- Tansey J, Szalryd-Woodle C, Gruia-Gray J, et al. (2001) Perilipin ablation results in a lean mouse with aberrant adipocyte lipolysis, enhanced leptin production, and resistance to diet-induced obesity. *Proceedings of the National Academy of Sciences of the United States of America* 98: 6494–6499.
- Zechner R, Kienesberger PC, Haemmerle G, Zimmermann R, and Lass A (2009) Adipose triglyceride lipase and the lipolytic catabolism of cellular fat stores. *Journal of Lipid Research* 50: 3–21.
- Zimmermann R, Strauss JG, Haemmerle G, et al. (2004) Fat mobilization in adipose tissue is promoted by adipose triglyceride lipase. *Science* 306: 1383–1386.

Adipogenesis

M D Lane, Johns Hopkins School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Clone/mitotic clonal expansion A clone is a group of cells that has arisen from a single cell and which have the identical genome.

Commitment The process that mesenchymal stem cells (MSCs) undergo to give rise to cells of a single lineage.

Hyperplasia An increase in cell number.

Hypertrophy An increase in cell size/content.

Mesenchymal stem cells (MSCs) Pluripotent cells arising from the embryonic mesoderm that have the capacity to commit/differentiate into cells of the adipocyte, muscle, cartilage, or bone lineages.

Introduction

Adipogenesis is the process by which fat-laden cells, that is, adipocytes, develop and accumulate as adipose tissue at various sites in the body, both as subcutaneous fat and as depots. The major roles of adipocytes are to store energy as fat during periods when energy intake exceeds expenditure and to mobilize this stored fuel when energy expenditure exceeds intake. Adipocytes arise both during late embryonic development and in the mature animal under conditions that promote obesity. Obesity – the excessive accumulation of body fat – is a consequence of persistent energy intake that exceeds energy expenditure. This condition gives rise both to hyperplasia (an increase in number) and to hypertrophy (an increase in size) of tissue adipocytes (Figure 1). An increasingly sedentary lifestyle coupled with an energy-rich diet has contributed to a high frequency of obesity and its attendant health problems, including type 2 diabetes.

Adipocytes also function in the control of metabolism through the secretion of paracrine and endocrine hormones/factors (e.g., leptin and adiponectin) that affect feeding behavior, metabolism, insulin sensitivity and secretion, and reproductive and immune functions. As the largest energy reserve in the body, adipose tissue has a major impact on energy flux, plasma lipid levels, and rates of glucose uptake. An understanding of the development program and its regulation is required to comprehend the roles of adipocytes.

Origin of Adipocytes

Adipocytes are derived from a pluripotent stem cell population that has the capacity both to self-renew and to differentiate into various cell types. While the fertilized egg is totipotent and can give rise to all adult cell types, mesenchymal stem cells (MSCs) are more restricted in their developmental potential, that is, they are pluripotent. Such cells can commit and differentiate into a limited number of cell types including adipocytes, osteocytes, chondrocytes, and myocytes. MSCs are found in adipose tissue, bone marrow, and muscle tissue. Recent evidence indicates that MSCs enter the bloodstream, presumably to populate and replenish this population in these sites. MSCs allow the replacement of missing or damaged cells following injury and, in expanding adipose tissue, provide a

reservoir of adipocyte progenitors that can proliferate and differentiate into mature adipocytes.

MSCs from adipose tissue can readily be isolated and this population expanded, making them an attractive source of MSCs for therapeutic use. Furthermore, recent evidence indicates that adult stem cells may have greater developmental plasticity than previously thought. It has been reported that MSCs are capable of committing and differentiating into non-mesodermal cell types; however, this transdifferentiation remains controversial. A greater understanding of the genetic mechanisms that orchestrate development may allow recapitulation of developmental processes *ex vivo* by directing totipotent or pluripotent stem cells to adopt specific fates for cell-based therapies. In addition, a better understanding of the mechanisms that regulate differentiation of terminally differentiated cell types, such as osteoblasts or adipocytes, may yield therapeutic interventions in pathological conditions such as osteoporosis or obesity.

The Adipocyte Development Program

Cell Culture and *In Vivo* Models

Much of what is known concerning adipocyte development has been derived from studies with cell-culture models. Two basic model systems have been employed: (1) multipotent stem cell lines that have not yet undergone commitment to the adipocyte lineage, and (2) preadipocyte cell lines that have undergone commitment to the adipocyte lineage and can be induced to terminally differentiate into adipocytes in cell culture (or when implanted subcutaneously into athymic mice). Pluripotent MSCs can be induced to undergo commitment in cell culture into several distinct cell lineages including the adipoblasts, myoblasts, osteoblasts, and chondroblasts. Once committed, preadipocytes can be induced to differentiate into adipocytes, but not into other cell types. For convenience, the following discussion of adipogenesis will be divided into two sections, commitment and differentiation.

Commitment

The MSC cell line most widely used to study commitment to the adipocyte lineage is the C3H10T1/2 line (referred to as the 10T1/2 line). This cell line was isolated from 14- to 17-day C3H mouse embryos and displays fibroblastic morphology

in culture. These cells are functionally similar to MSCs of mesodermal origin since they have the capacity (upon brief exposure to 4-azacytidine, a DNA methyl transferase inhibitor) to convert into a limited number of cell types that (when induced to differentiate) display the morphological and biochemical characteristics of adipocytes, myocytes, chondrocytes, or osteocytes. The conversion to each of these cell types is heritable as clonal lines can be isolated that retain these specific phenotypes without further treatment with 4-azacytidine. It is believed that this phenotypic alteration results from hypomethylation of a small number (one or two) of regulatory genes involved in the commitment process. By following this procedure, MyoD – a transcription factor that converts the 10T1/2 cells to myoblasts – was discovered. Use of this strategy with uncommitted 10T1/2 stem cells gave rise to several committed preadipocyte cell lines that have been extensively studied. Unlike the 10T1/2 progenitor line, these lines, for example, an A33 preadipocyte line derived from 10T1/2 cells treated with 4-azacytidine, undergo differentiation into adipocytes when subjected to the adipocyte differentiation protocol (see below).

The 10T1/2 cells also respond to treatment with bone morphogenetic proteins (BMPs), members of the transforming growth factor- β superfamily. BMPs are known morphogens that control embryonic patterning, most notably of the mesoderm from which MSCs are derived. BMP2 and BMP4 have been shown to play a role in the control of MSC commitment, promoting adipose, cartilage, or bone lineages, depending on culture conditions. Thus, treatment of 10T1/2 cells with BMP4 induces virtually complete commitment and subsequent differentiation to the adipocyte lineage. Evidence indicates that BMP4 is a morphogen that acts in the adipocyte commitment pathway. An outline of the steps thought to be involved in commitment of MSCs to the adipocyte lineage is shown in Figure 2.

The Wnt pathway

Wnts have been shown to act upstream of BMP4. They are secreted glycoproteins that serve as ligands for members of the frizzled receptor family. Wnts appear to function at two points in the adipogenic pathway – both in the commitment

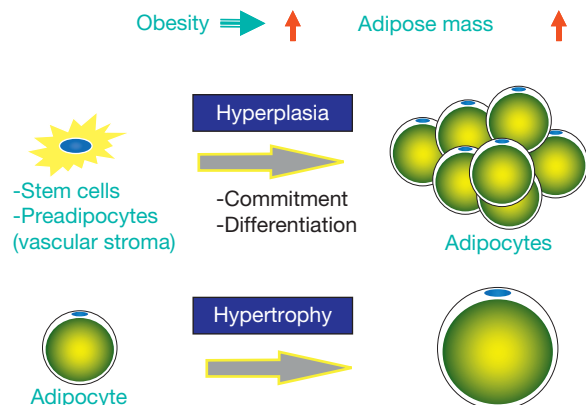


Figure 1 Adipocyte hyperplasia and hypertrophy are responsible for the increased adiposity/obesity that occurs when energy intake exceeds expenditure.

and differentiation phases. Active Wnt signaling is known to suppress the activity of GSK3 β , a protein kinase in a complex of several other proteins (axin, APC, and β -catenin), by blocking phosphorylation of the β -catenin. This action stabilizes β -catenin and facilitates its translocation into the nucleus where it acts as co-activator for several transcription factors of lymphoid-enhancing factor/T-cell factor (Lef/Tcf) family. During commitment, Wnt signaling is activated and initiates the pathway illustrated in Figure 2, which includes activating the expression of BMP4.

In contrast, active Wnt signaling suppresses differentiation. Thus, for differentiation to occur Wnt signaling is inactivated. In the absence of Wnt signaling, GSK3 β is active leading to the phosphorylation of β -catenin targeting it for destruction by the proteasomal system. Thus, inactivation of Wnt is permissive for adipogenic differentiation. Since active Wnt signaling promotes the activation of β -catenin and blocks expression of C/EBP α and PPAR γ , Wnt signaling must be inactivated for differentiation to proceed. It should be noted that active Wnt promotes osteogenesis and myogenesis – a branch point in development from mesodermal pluripotent MSCs (see Figure 2).

Comparison of differential gene expression patterns between 10T1/2 stem cells and A33 preadipocytes by gene chip analysis indicates that activation of a small set of genes during the 10T1/2 cells-to-preadipocyte transition is responsible for acquisition of the preadipocyte phenotype. Together, these observations indicate that Wnt signaling plays a dual activating role early in adipocyte development and an inhibitory role later during

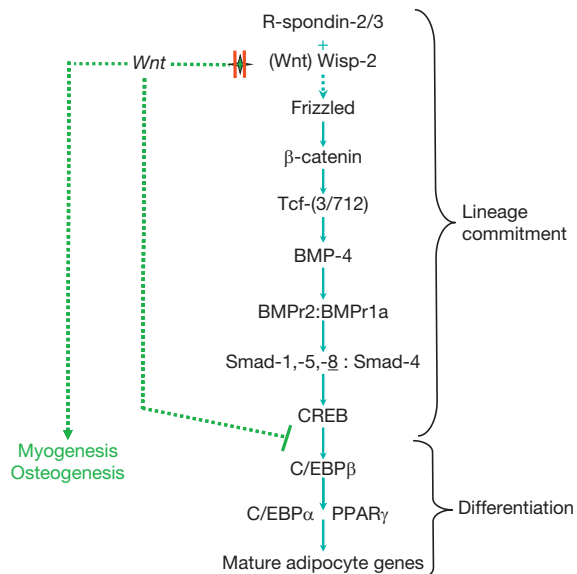


Figure 2 Intermediates in the commitment pathway by which mesenchymal stem cells (MSCs) undergo conversion into the preadipocytes. The probable role of Wnt in stem cell commitment to the adipocyte lineage is illustrated. It is visualized that R-spondins activate the Wnt/ β -catenin signaling cascade that leads to expression of BMP-4. BMP-4, acting via Smad signaling and an as-yet unidentified downstream intermediary (possibly CREB), induces the expression of C/EBP β , transcriptional activator of C/EBP α and PPAR γ . C/EBP α and PPAR γ act as pleiotropic activators of genes whose expression defines the terminally differentiated adipocyte phenotype.

the terminal differentiation phase. Thus, Wnt activity must be suppressed during the differentiation phase of adipogenesis.

Differentiation

Research with established preadipocyte cell lines has provided much of our knowledge concerning adipocyte differentiation. These homogeneous cell populations can be induced in cell culture to synchronously enter the stably differentiated state. The subcutaneous implantation of 3T3-F442A preadipocytes into athymic mice resulted in the development of an adipose depot that is indistinguishable from endogenous adipose tissue. Thus, this use of the cells provides a test for authenticity of findings made in cell culture. The mouse 3T3-L1 and 3T3-F442A lines have been the best-characterized preadipocyte cell lines. These lines were selected from disaggregated mouse embryos for their ability to accumulate cytoplasmic fat and represent faithful models of preadipocyte differentiation process *in vivo*.

Induction of differentiation

The protocol (referred to as MDI) most often used to induce adipocyte differentiation is to treat preadipocytes with a combination of cyclic adenosine monophosphate (cAMP) (or an agent such as methylisobutylxanthine that increases the level of cellular cAMP) and a glucocorticoid (or dexamethasone, a synthetic glucocorticoid mimic), and a high concentration of insulin (as an insulin-like growth factor 1 (IGF-1) mimic) in the presence of fetal bovine serum (FBS). Prior to differentiation, preadipocyte cell lines have a fibroblastic morphology similar to that of preadipose cells in the stroma of adipose tissue. Upon induction of differentiation, the cells lose their fibroblastic character, assume a rounded-up appearance, and acquire the morphological and biochemical characteristics of adipocytes.

Mitotic clonal expansion

In cell culture, the induced preadipocytes become cell-density inhibited and arrest at the G0/G1 cell-cycle boundary. Growth arrest is a prerequisite for preadipocyte differentiation. The growth-arrested cells then synchronously reenter the cell cycle, undergo two to three rounds of mitosis (mitotic clonal expansion), exit the cell cycle, and terminally differentiate. It is believed that DNA replication and the associated changes in chromatin structure that occur during the mitotic clonal expansion process increase the accessibility of *cis*-elements to transacting (transcription) factors, which activate or de-repress transcription of numerous adipocyte gene(s) that produce the adipocyte phenotype.

Terminal differentiation

As the cells complete clonal expansion, they enter a unique growth-arrested stage of the cell cycle, that is, GD, which is permissive for subsequent differentiation. With 3T3-L1 preadipocytes, this process is characterized by changes in expression of cyclin-dependent kinase inhibitors p18, p21, and p27. Even after differentiation has been initiated, the preadipocytes retain the ability to dedifferentiate and reenter mitosis. However, once the cells have exited GD, they become committed to terminal differentiation. Terminal differentiation is maintained, at least in part, through the expression of C/EBP α and PPAR γ . C/EBP α and PPAR γ act synergistically to activate transcription of genes that produce the adipocyte phenotype (Figure 3). The promoter of the C/EBP α gene possesses a C/EBP regulatory element to which C/EBP α can bind and thereby autoactivate its expression and maintain the differentiated state. C/EBP α also has anti-mitotic activity; its over-expression leads to cell cycle arrest, which terminates mitotic clonal expansion. This anti-mitotic activity is through direct interaction with the cyclin-dependent kinases, cdk2 and cdk4. Like

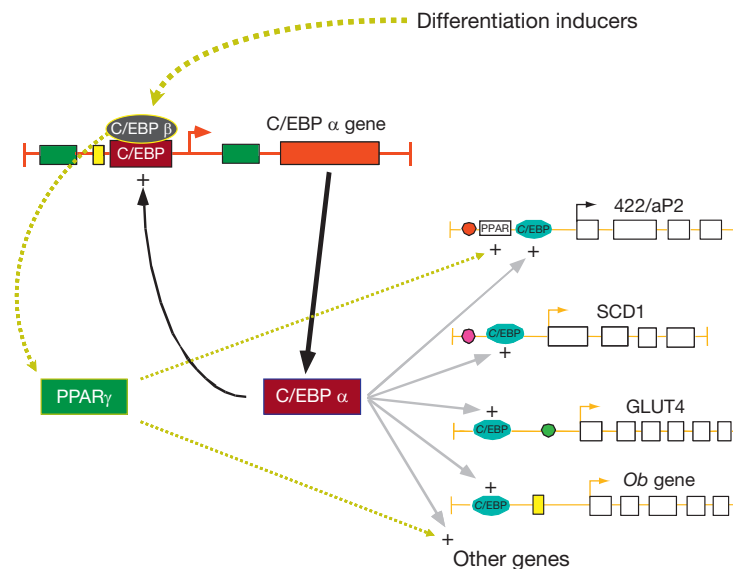


Figure 3 Transcription cascade initiated by C/EBP β , which leads to the expression of the adipogenic genes. C/EBP, CCAAT/enhancer binding protein; PPAR, peroxisome proliferator-activated receptor; 422/aP2, adipocyte-specific fatty acid binding protein; SCD1, stearoyl-CoA desaturase; GLUT4, the insulin-responsive glucose transporter; Ob, gene that encodes leptin.

C/EBP α , PPAR γ has been observed to induce growth arrest in certain cells.

Transcriptional control

After treating 3T3-L1 preadipocytes with a cocktail of differentiation inducers (i.e., MDI, see above), the cells synchronously reenter the cell cycle and begin to express a cascade of transcription factors that ultimately produce the adipocyte phenotype. The sequence of events in this transcriptional cascade is summarized in **Figure 4**. It should be noted that these events were identified primarily in studies with 3T3-L1 and -F442A cells, which have the advantage of proceeding through the entire differentiation program in a synchronous manner, allowing a clear delineation of each step. The transcription factors are listed below in their order of appearance/function in the differentiation program.

cAMP regulatory element binding protein

cAMP activates protein kinase A, which, in turn, rapidly (≤ 5 min) phosphorylates/activates the nuclear transcription factor cAMP regulatory element binding protein (CREB). One of the preadipocyte differentiation inducers, that is, cAMP, acts rapidly in the nucleus to phosphorylate/activate CREB. Activated CREB then binds at two sites in the proximal promoter of C/EBP β gene, thereby triggering its expression. This event triggers the transcriptional cascade that initiates differentiation.

C/EBP β

C/EBP β is an early transcriptional activator of preadipocyte differentiation and is expressed within 4 h of induction of differentiation. Although C/EBP β is expressed immediately (≤ 20 min) after induction of differentiation, it lacks DNA-binding activity, which is acquired much later in the program. Acquisition of DNA binding occurs at ~ 16 h induction, coincident with entry of the cells into the S-phase at the onset of

mitotic clonal expansion. Concomitant with acquisition of DNA-binding activity, C/EBP β becomes localized to centromeres through C/EBP consensus binding sites in centromeric satellite DNA. As indicated above, the transcriptional activation of C/EBP β is regulated by CREB. The C/EBP β gene possesses two CREB-like *cis* regulatory binding sites in its proximal promoter that are necessary for transcriptional activation. The phosphorylation/activation of CREB is closely correlated with the expression of C/EBP β .

Phosphorylation is an important posttranslational modification of C/EBP β that leads to the acquisition of DNA-binding function as preadipocytes traverse to the G1-S checkpoint at the onset of mitotic clonal expansion. C/EBP β is phosphorylated sequentially, first by mitogen-activated protein (MAP) kinase and then much later by GSK3 β . Phosphorylation on Thr188 by MAP kinase occurs within 4 h of induction of differentiation and is required for mitotic clonal expansion; C/EBP β DNA-binding activity and terminal differentiation require additional phosphorylations on Thr179 and Ser184. Phosphorylation at the latter sites by GSK3 β occurs between ~ 16 h after induction. Hyperphosphorylation of C/EBP β at these three sites activates its DNA-binding activity, and although the mechanism by which this occurs is not understood, it is likely that a phosphorylation-induced conformational change is involved.

C/EBP homologous protein

C/EBP homologous protein (CHOP-10) is a member of the C/EBP family that exerts a suppressive effect early in the differentiation program. CHOP-10 is expressed by growth-arrested preadipocytes, but is downregulated when C/EBP β acquires DNA-binding activity. Thus, during preadipocyte differentiation, CHOP-10 heterodimerizes with C/EBP β and prevents acquisition of DNA binding to C/EBP regulatory elements. It appears that the interaction of CHOP-10 with C/EBP β provides a fail-safe mechanism that prevents the acquisition of

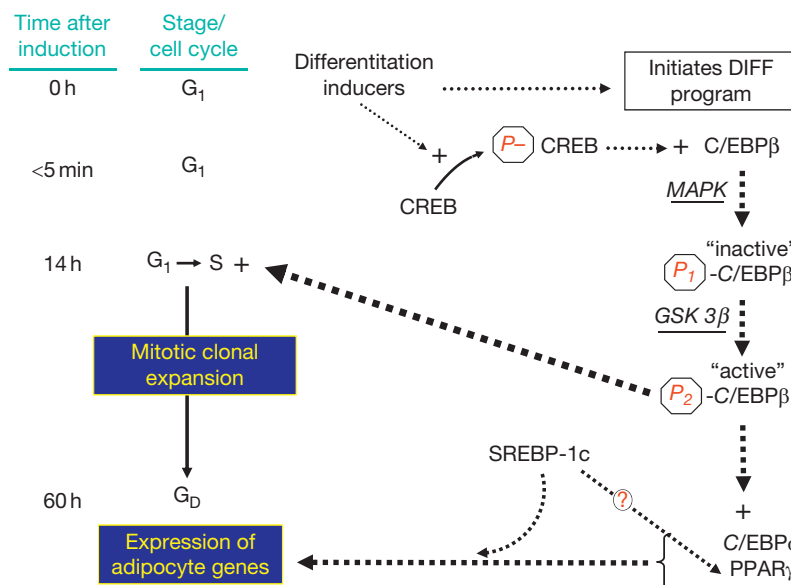


Figure 4 Key events in the adipocyte differentiation pathway. P, phosphoryl group; C/EBP, CCAAT/enhancer binding protein; CREB, cAMP regulatory element binding protein; GSK-3 β , glycogen synthase kinase-3 β ; MAPK, mitogen-activated protein kinase; PPAR, peroxisome proliferator-activated receptor; SREBP, sterol regulatory element binding protein-1c; G1 and S, G1- and S-phases of the cell cycle, respectively.

DNA-binding activity by C/EBP β until preadipocytes have entered mitotic clonal expansion.

C/EBP α

C/EBP α is a pleiotropic transcriptional activator of adipocyte-specific genes. Upon acquisition of DNA-binding activity, C/EBP β binds to the proximal promoter and activates the expression of C/EBP α . The promoters of numerous adipocyte genes possess C/EBP regulatory consensus sequences that support transactivation by C/EBP α . Why is there redundancy in the expression of the C/EBP gene family (notably C/EBP β and C/EBP α) during adipogenesis? This may be because C/EBP α has anti-mitotic activity. Thus, the expression of C/EBP α early during adipogenesis prevents preadipocytes from entering mitotic clonal expansion, a requirement for subsequent differentiation. C/EBP α and PPAR γ (see below) are coordinately expressed late in the differentiation program and, together, these transcription factors activate the coordinate expression of the family of genes that give rise to the adipocyte phenotype.

Peroxisome proliferator-activated receptor- γ

Peroxisome proliferator-activated receptor- γ (PPAR γ) is a member of the nuclear hormone receptor family, many of which are activated by ligands that control their activity. PPAR γ is a central regulator in adipogenesis. Like C/EBP α , PPAR γ is responsible for activating a number of genes involved in fatty acid metabolism. C/EBP α and PPAR γ are expressed coordinately during adipocyte differentiation and act together to activate the expression of most of the genes that encode proteins that give rise to the adipocyte phenotype.

Sterol regulatory binding protein

Sterol regulatory binding proteins (SREBPs) are transcription factors that regulate the transcription of numerous genes of cholesterol and fatty acid metabolism (Horton et al., 2002). Of the three members of the SREBP family, only SREBP-1c activates transcription of genes required for fatty acid synthesis, which are required for acquisition of the adipocyte phenotype. What triggers the expression/activation of SREBP-1c during the preadipocyte differentiation program is uncertain; however, it is known that insulin (or IGF1) – one of the differentiation inducers – activates its expression.

Overview

An encapsulation of the adipocyte differentiation pathway is shown in **Figure 4**. The differentiation process is initiated by the rapid phosphorylation of CREB, which activates the expression of C/EBP β via cAMP-responsive regulatory elements within the proximal promoter of the C/EBP β gene. At this point, C/EBP β lacks DNA-binding activity. However, after a lengthy lag period and reentry of the cell cycle C/EBP β undergoes phosphorylation by MAP kinase and later by GSK-3 β . This dual phosphorylation provokes acquisition of DNA-binding activity by C/EBP β , which allows C/EBP β to bind to and transcriptionally activate the genes encoding two key adipogenic transcription factors, that is, C/EBP α and PPAR γ . The coordinate expression of C/EBP α and PPAR γ then acts together to activate the expression of a large group of adipocyte genes. These transcription factors, along with SREBP-1c, activate the production of and maintain the adipocyte phenotype as illustrated in **Figure 4**.

See also: **Metabolism Vitamins and Hormones: Fatty Acid Structure and Synthesis.**

Further Reading

- Bowers RR and Lane MD (2007) A role for bone morphogenetic protein-4 in adipocyte development. *Cell Cycle* 6: 385–389.
- Li X, Kim J-W, Grønberg M, et al. (2007) Role of CDK2 in the sequential phosphorylation/activation of C/EBP β during adipocyte differentiation. *Proceedings of the National Academy of Sciences of the United States of America* 104: 11597–11602.
- Otto TC and Lane MD (2005) Adipose development: From stem cell to adipocyte. *Critical Reviews in Biochemistry and Molecular Biology* 40: 229–242.
- Tang Q-Q, Grønberg M, Huang H, et al. (2005) Sequential phosphorylation of C/EBP β during adipogenesis: Role of MAP kinase and GSK3 β . *Proceedings of the National Academy of Sciences of the United States of America* 102: 9766–9771.
- Tang Q-Q, Otto TC, and Lane MD (2003) CCAAT/Enhancer binding protein β is required for mitotic clonal expansion and adipocyte differentiation. *Proceedings of the National Academy of Sciences of the United States of America* 100: 850–855.
- Tang Q-Q, Otto TC, and Lane MD (2003) Mitotic clonal expansion: A synchronous process required for adipogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 100: 44–49.
- Zhang J-W, Klemm DJ, Vinson C, and Lane MD (2003) Transcriptional regulation of CCAAT/enhancer binding protein β gene during adipogenesis: Role of CREB. *Journal of Biological Chemistry* 279: 4471–4478.

Adrenergic Receptors

D B Bylund, University of Nebraska Medical Center, Omaha, NE, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Agonist Compound that binds to a receptor and activates it, thus causing a biological response.

Antagonist Compound that binds to a receptor but does not activate it. It can block or inhibit the activation caused by an agonist.

Desensitization Decrease in response of a tissue to a neurotransmitter or agonist drug following repeated administration.

Downregulation Decrease in the number or density of receptor-binding sites following chronic agonist treatment.

Partial agonist Agonist that has less than full efficacy in activating its receptor.

Polymorphism Variability in DNA sequence that occurs with an allele frequency of greater than 1% in the population.

Adrenergic receptors (adrenoceptors) mediate the central and peripheral actions of the neurotransmitter norepinephrine (noradrenaline) and the hormone epinephrine (adrenaline); they are widely distributed throughout the body. There are three major adrenergic receptor types: alpha-1, alpha-2, and beta. Each of these three receptor types is further divided into three subtypes. Adrenergic receptors are seven transmembrane receptors, which consist of a single polypeptide chain with seven hydrophobic regions that form alpha helical structures that span or transverse the membrane. Because the mechanism of action of adrenergic receptors includes the activation of guanine nucleotide regulatory binding proteins (G proteins), they are also called G-protein-coupled receptors.

Epinephrine and Norepinephrine

Norepinephrine (noradrenaline) is a neurotransmitter in both the peripheral and central nervous systems. Epinephrine (adrenaline) is a hormone released from the adrenal gland. Norepinephrine and epinephrine are catecholamines, because they have both the catechol moiety (two hydroxyl groups on a benzene ring) and an amine (NH_2). Both of these catecholamine messengers play important roles in the regulation of diverse physiological systems by acting through adrenergic receptors. Stimulation of adrenergic receptors by catecholamines, released in response to activation of the sympathetic autonomic nervous system, results in a variety of effects such as increased heart rate, regulation of vascular tone, and bronchodilatation. In the central nervous system, adrenergic receptors are involved in many functions, including memory, learning, alertness, and the response to stress.

Norepinephrine is synthesized starting with the amino acid tyrosine, which is obtained from the diet and can also be synthesized from phenylalanine. Tyrosine is converted to dihydroxyphenylalanine (DOPA) by the enzyme tyrosine hydroxylase, and DOPA in turn is converted to dopamine. Dopamine is then converted to norepinephrine by the enzyme dopamine beta-hydroxylase. In the adrenal medulla and in a few brain

regions, norepinephrine is converted to epinephrine by the enzyme phenylethanolamine N-methyltransferase. The major mechanism by which the effects of norepinephrine are terminated is reuptake back into the nerve terminal by a high-affinity norepinephrine transporter. Epinephrine, as well as norepinephrine, are metabolized to inactive products. Norepinephrine is metabolized by the enzymes monoamine oxidase (MAO) and catechol-O-methyltransferase (COMT) to 3-methoxy-4-hydroxyphenylglycol (MHPG) and 3-methoxy-4-hydroxymandelic acid (VMA). The major metabolite found in the blood and urine is MHPG. Epinephrine is similarly metabolized by MAO and COMT to VMA.

Classification and Mechanism of Action of Adrenergic Receptors

Adrenergic receptors were originally divided into two major types, alpha and beta, based on their pharmacological characteristics (i.e., rank order potency of agonists). Subsequently, the beta adrenergic receptors were subdivided into beta-1 and beta-2 subtypes; more recently, a beta-3 subtype has been defined. The alpha adrenergic receptors were first subdivided into postsynaptic (alpha-1) and presynaptic (alpha-2) subtypes. After it was realized that not all alpha receptors with alpha-2 pharmacological characteristics were presynaptic, the pharmacological definition was used. The current classification scheme is based on three major types: alpha-1, alpha-2, and beta. Each of these three receptor types is further divided into three subtypes as shown in [Figure 1](#): alpha-1A, alpha-1B, alpha-1D; alpha-2A, alpha-2B, alpha-2C; and beta-1, beta-2, beta-3.

The binding of an agonist to an adrenergic receptor induces (or stabilizes) a conformational change that allows the receptor to interact with and activate a G protein. The activated receptor facilitates the exchange of guanidine diphosphate for guanidine triphosphate, leading to the dissociation of the α and $\beta\gamma$ subunits of the G protein, which in turn stimulate or inhibit the activity of various effectors. It is important to note

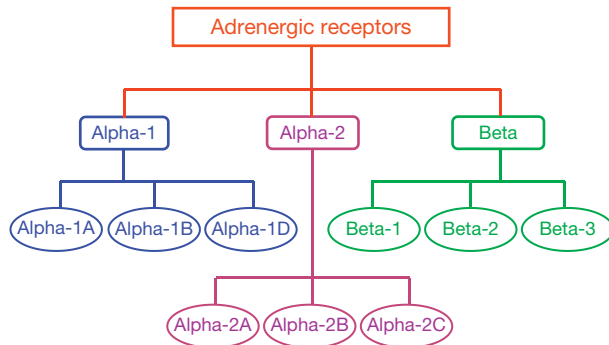


Figure 1 The classification scheme for adrenergic receptors.

that each of the three types of adrenergic receptors couples to a distinct class of G proteins: alpha-1 to G_q ; alpha-2 to $G_{i/o}$; and beta to G_s . In addition to G proteins, adrenergic receptors interact with other signaling proteins and pathways such as those involving tyrosine kinases.

Alpha-1 Adrenergic Receptors

Three genetic and four pharmacological alpha-1 adrenergic receptor subtypes have been defined. The alpha-1A and alpha-1B subtypes were initially defined based on their differential affinity for adrenergic agents, such as WB4101 and phentolamine, and on their differential sensitivities to the site-directed alkylating agent chloroethylclonidine. The alpha-1B subtype was subsequently cloned from the hamster and the alpha-1A was cloned from bovine brain, although it was originally called the alpha-1c adrenergic receptor. A third subtype, the alpha-1D adrenergic receptor, was subsequently cloned from the rat cerebral cortex, although this clone was originally called the alpha-1a subtype by some investigators. A fourth pharmacological subtype, the alpha-1L, has been identified in vascular tissues from several species, but it may represent a conformational state of the alpha-1A receptor. The current classification scheme includes the alpha-1A, the alpha-1B, and the alpha-1D, but there is no alpha-1 C (Figure 1).

Pharmacological and Molecular Characteristics of Alpha-1 Adrenergic Receptors

In addition to norepinephrine and epinephrine, alpha-1 receptors are activated by various agonists such as phenylephrine (neosynephrine) and methoxamine (Vasoxyl). These agonists are relatively selective for alpha-1 receptors and have low affinity for alpha-2 and beta receptors. By contrast, they have similar affinities for the three alpha-1 subtypes and are thus nonsubtype selectively agonists. Similarly, antagonists including prazosin (Minipress) and tamsulosin (Flomax) are relatively selective for alpha-1 receptors and block alpha-2 and beta receptors only at high concentrations. Several other antagonists such as phentolamine (Regitine) and phenoxybenzamine (Dibenzylamine) block both alpha-1 and alpha-2 adrenergic receptors with similar affinities. Alpha-1A selective

antagonists include 5-methylurapidil and niguldipine, whereas cirazoline appears to be a selective alpha-1A agonist.

The alpha-1 adrenergic receptors are single polypeptide chains of 446–572 amino acid residues that span the membrane 7 times, with the amino terminal being extracellular and the carboxy terminal intracellular. Thus, there are three intracellular loops and three extracellular loops. In contrast to the alpha-2 receptors, but similar to the beta receptors, the alpha-1 receptors have a long carboxy terminal tail (137–179 amino acid residues) and a short third intracellular loop (68–73 amino acid residues). The amino terminal of the alpha-1A and alpha-1B subtypes have three (alpha-1A) or four (alpha-1B) consensus sites for N-linked glycosylation. The carboxy terminal tails of all three subtypes are potentially palmitoylated, thus anchoring the tail to the membrane and forming a small fourth intracellular loop. The carboxy terminal tails also have multiple sites of phosphorylation that are thought to be important in the desensitization, recycling, and downregulation of the receptor.

The human alpha-1 adrenergic receptor genes consist of two exons and a single large intron of at least 20 kb in the region corresponding to the sixth transmembrane domain. No splice variants are known for the alpha-1B and alpha-1D subtypes. By contrast, at least 10 splice variants of human alpha-1A subtype have been reported, but only four produce full-length receptors.

Alpha-2 Adrenergic Receptors

Three genetic and four pharmacological alpha-2 adrenergic receptor subtypes have also been defined (Figure 1). The alpha-2A and alpha-2B subtypes were initially defined based on their differential affinity for adrenergic agents such as prazosin and oxymetazoline. These subtypes were subsequently cloned from human, rat, and mouse. A third subtype, alpha-2 C, was originally identified in an opossum kidney cell line and has also been cloned from several species. A fourth pharmacological subtype, the alpha-2D, has been identified in the rat, mouse, and cow. This pharmacological subtype is a species ortholog of the human alpha-2A subtype, and thus it is not a separate genetic subtype.

Pharmacological and Molecular Characteristics of Alpha-2 Adrenergic Receptors

In addition to norepinephrine and epinephrine, alpha-2 receptors are activated by clonidine (Catapres) and brimonidine (Alphagan). These agonists are relatively selective for alpha-2 receptors and have lower affinity at alpha-1 and beta receptors. Similarly, the antagonist yohimbine is relatively selective for alpha-2 receptors and blocks alpha-1 and beta receptors only at higher concentrations. Antagonists that are at least somewhat selective for one of the alpha-2 subtypes include BRL44408 for the alpha-2A, prazosin and ARC-239 for the alpha-2B (note, however, that these two agents have much higher affinities for alpha-1 receptors), and rauwolscine for the alpha-2 C subtype. Oxymetazoline is a partial agonist that has a higher affinity for the alpha-2A subtype when compared to the alpha-2B and alpha-2 C subtypes.

The alpha-2 adrenergic receptors are single polypeptide chains of 450–462 amino acid residues. In contrast to the alpha-1 and beta receptors, the alpha-2 receptors tend to have long third intracellular loops (148–179 amino acid residues) and a short carboxy terminal tail (20–21 amino acid residues). The amino terminal of the alpha-2A and alpha-2C subtypes has two consensus sites for N-linked glycosylation, and the carboxy terminal tails of all three subtypes are potentially palmitoylated. The third intracellular loops have multiple sites of phosphorylation, which are thought to be important in the desensitization, recycling, and downregulation of the receptor. The alpha-2 adrenergic receptor genes do not contain introns, and thus there are no splice variants.

Beta Adrenergic Receptors

Three beta adrenergic receptor subtypes have been identified. The beta-1 adrenergic receptor, the dominant receptor in heart and adipose tissue, is equally sensitive to epinephrine and norepinephrine, whereas the beta-2 adrenergic receptor, responsible for relaxation of vascular, uterine, and airway smooth muscle, is less sensitive to norepinephrine as compared to epinephrine. The beta-3 receptor is insensitive to the commonly used beta-adrenergic receptor antagonists and was previously referred to as the 'atypical' beta adrenergic receptor. A beta-4 receptor has been postulated; however, definitive evidence of its existence is lacking, and it is now thought to be a 'state' of the beta-1 adrenergic receptor.

Pharmacological and Molecular Characteristics of Beta Adrenergic Receptors

Isoproterenol (Isuprel) is the prototypic nonsubtype selective beta agonist that has no activity at alpha-1 and alpha-2 receptors except at high concentrations. Epinephrine is 10–100-fold more potent at the beta-2 receptor as compared to the beta-1 subtype, whereas norepinephrine is more potent than epinephrine at the beta-3 subtype. Many beta-2 selective agonists, such as terbutaline (Brethine) and salmeterol (Serevent), have been developed for the treatment of asthma. Due to their subtype selectivity, they have a lower incidence of side effects mediated by the beta-1 receptor. Propranolol (Inderal) is the prototypic nonsubtype selective beta antagonist that has equal affinities at the beta-1 and beta-2 subtypes. Other nonselective beta adrenergic antagonists include timolol (Blocadren), pindolol (Visken, which is actually a weak partial agonist), and carvedilol (Coreg), which is also an alpha-1 antagonist. Several beta-1 selective antagonists have been developed, such as metoprolol (Lopressor) and esmolol (Brevibloc).

The beta adrenergic receptors are single polypeptide chains of 408–477. In contrast to the alpha-2 receptors, but similar to the alpha-1 receptors, the beta receptors tend to have longer carboxy terminal tails (61–97 amino acid residues) and shorter third intracellular loops (54–80 amino acids). The amino terminal of the beta receptors have one or two consensus sites for N-linked glycosylation, and the carboxy terminal tails are potentially palmitoylated. The carboxy terminal tails also have multiple sites of phosphorylation, which are thought to

be important in the desensitization, recycling, and downregulation of the receptor. The crystal structure of modified beta-2 adrenergic receptors has been determined.

The beta-1 and beta-2 adrenergic receptor genes do not contain introns; thus, they have no splice variants. By contrast, the beta-3 receptor has one intron, resulting in two splice variants. However, no functional differences have been found between the two splice variants.

Regulation of Adrenergic Receptors

The processes involved in homologous desensitization and downregulation have been extensively investigated for the beta-2 adrenergic receptor. The other adrenergic receptors, as well as many other G-protein-coupled receptors, appear to behave in a similar manner. Initial uncoupling of the beta-2 receptor from the G protein after agonist binding is mediated by phosphorylation of specific residues in the carboxyl tail of the receptor. The phosphorylated beta-2 receptor serves as a substrate for the binding of β -arrestin, which not only uncouples the receptor from the signal transduction process but also serves as an adapter protein that mediates the binding of additional signaling proteins and entry into the internalization pathway. The mechanisms of beta-2 adrenergic receptor downregulation appear to involve both an increase in the rate of degradation of the receptor as well as a decrease in the levels of beta receptor messenger RNA (mRNA).

Adrenergic Receptor Signal Transduction Pathways

The alpha-1 adrenergic receptors activate the $G_{q/11}$ family of G proteins leading to the dissociation of the α and $\beta\gamma$ subunits and the subsequent stimulation of the enzyme phospholipase C. This enzyme hydrolyzes phosphatidylinositol 1,2-bisphosphate in the membrane producing inositol trisphosphate (IP_3) and diacylglycerol. These molecules act as second messengers mediating intracellular Ca^{2+} release via the IP_3 receptor and activating protein kinase C. Other signaling pathways that have also been shown to be activated by alpha-1 receptors include Ca^{2+} influx via voltage-dependent and -independent calcium channels, arachidonic acid release, and activation of phospholipase A_2 , phospholipase D, and mitogen-activated protein kinase.

The alpha-2 adrenergic receptors activate the $G_{i/o}$ family of G proteins and alter (classically inhibit) the activity of the enzyme adenylyl cyclase, which, in turn, decreases the concentration of the second messenger cyclic adenosine monophosphate (AMP). In addition, the stimulation of alpha-2 receptors can regulate several other effector systems including the activation of K^+ channels, inhibition or activation of Ca^{2+} channels, and activation of phospholipase A_2 , phospholipase C, and Na^+/H^+ exchange.

The beta adrenergic receptors activate the G_s family of G proteins and activate adenylyl cyclase, thus increasing cyclic AMP concentrations. Beta adrenergic receptors interact with many other signaling proteins, including the phosphoprotein EBP50 (ezrin-radixin-moesin-binding phosphoprotein-50), the Na^+/H^+ exchanger regulatory factor, and with CNrasGEF.

Adrenergic Receptor Polymorphisms

Polymorphisms have been identified in some of the alpha-2 and beta adrenergic receptor subtypes, which may have important clinical implications. A common polymorphism has been identified in the third intracellular loop of the alpha-2B receptor, which consists of a deletion of three glutamate residues (301–303) that results in a loss of short-term agonist-induced desensitization. This deletion is a risk factor for acute coronary events, but not hypertension. A common polymorphism has been identified in the third intracellular loop of alpha-2 C subtype, which consists of a deletion of four amino acid residues (322–325).

The gene encoding the human beta-1 adrenergic receptor is quite polymorphic with 18 single-nucleotide polymorphisms, 7 of which cause amino acid substitutions. A total of 13 polymorphisms in the beta-2 adrenergic receptor gene and its transcriptional regulator upstream peptide has been identified. Three closely linked polymorphisms, two coding regions at amino acid positions 16 and 27 and one in the upstream peptide, are common in the general Caucasian population. The glycine-16 receptor exhibits enhanced downregulation *in vitro* after agonist exposure. By contrast, arginine-16 receptors are more resistant to downregulation. Some studies have suggested a relationship among these polymorphisms, airway responsiveness (e.g., asthma), and the responsiveness to beta adrenergic agonists.

A tryptophan-64 to arginine polymorphism has been identified in the beta-3 adrenergic receptor. The allele frequency is approximately 30% in the Japanese population, higher in Pima Indians, and lower in Caucasians. Type 2 diabetic patients with this mutation showed a significantly younger onset age of diabetes and an increased tendency to obesity, hyperinsulinemia, and hypertension.

See also: **Metabolism** Vitamins and Hormones: Diabetes; **Signaling:** Adenylyl Cyclases; Dopamine Receptors; G-Protein-Coupled Receptor Kinases and Arrestins; Inositol Phosphate Kinases and Phosphatases; Phosphatidylinositol Bisphosphate and Trisphosphate; Phospholipase C; Phospholipase D.

Further Reading

- Audet M and Bouvier M (2008) Insights into signaling from the beta-2 adrenergic receptor structure. *Nature Chemical Biology* 4: 397–403.
- Bylund DB (1988) Subtypes of α_2 -adrenoceptors: Pharmacological and molecular biological evidence converge. *Trends in Pharmacological Sciences* 9: 356–361.
- Bylund DB, Eikenberg DC, Hieble JP, et al. (1994) International Union of Pharmacology nomenclature of adrenoceptors. *Pharmacological Reviews* 46: 121–136.
- Cooper JR, Bloom FE, and Roth RH (2003) *The Biochemical Basis of Neuropsychopharmacology*, 8th edn. New York, NY: Oxford University Press.
- Crassous PA, Denis C, Paris H, and Senard JM (2007) Interest of alpha2-adrenergic agonists and antagonists in clinical practice: Background, facts and perspectives. *Current Topics in Medicinal Chemistry* 7: 187–194.
- Docherty JR (2010) Subtypes of functional alpha-1 adrenoceptors. *Cellular and Molecular Life Sciences* 67: 405–417.
- Gyires K, Zadori ZS, Torok T, and Matyus P (2009) Alpha-2 adrenoceptor subtypes-mediated physiological, pharmacological actions. *Neurochemistry International* 55: 447–453.
- Hein P and Michel MC (2007) Signal transduction and regulation: Are all alpha1-adrenergic receptor subtypes created equal? *Biochemical Pharmacology* 73: 1097–1106.
- Hieble JP, Bylund DB, Clarke DE, et al. (1995) International Union of Pharmacology X. Recommendation for nomenclature of alpha-1 adrenoceptors: Consensus update. *Pharmacological Reviews* 47: 267–270.
- Knaus AE, Muthig V, Schickinger S, et al. (2007) Alpha-2 adrenoceptor subtypes: Unexpected functions for receptors and ligands derived from gene-targeted mouse models. *Neurochemistry International* 51: 277–281.
- Liggett SB (2010) Pharmacogenomics of beta-1 adrenergic receptor polymorphisms in heart failure. *Heart Failure Clinics* 6: 27–33.
- Perez DM (2007) Structure–function of alpha-1 adrenergic receptors. *Biochemical Pharmacology* 73: 1051–1062.
- Rosenbaum DM, Rasmussen SG, and Kobilka BK (2009) The structure and function of G protein-coupled receptors. *Nature* 459: 356–363.

Affinity Tags for Protein Purification

J J Falke and J A Corbin, University of Colorado, Boulder, CO, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 57–63, © 2004, Elsevier Inc.

Glossary

Epitope tag A short polypeptide fused to a protein of interest so that it will be recognized as a high-affinity binding target by a specific, well-characterized antibody.

Fusion protein A polypeptide made from a recombinant gene consisting of two or more gene fragments fused together.

Immobilized metal-affinity chromatography

(IMAC) A type of affinity chromatography based on the specific interaction between a metal chelate stationary phase and a metal-binding peptide fused to a protein of interest.

Applications of Affinity Tags

Affinity tags have proved to be tremendously effective tools for a wide variety of applications, and they are now incorporated as a standard feature to reduce the number of steps in purification protocols developed for recombinant proteins. In addition, affinity tags have become indispensable in the immobilization of proteins for display on a surface, where the tag ensures that the fusion protein of interest is oriented with its functional regions exposed to docking with other macromolecules. Affinity tags are also used to detect and quantitate target proteins and to analyze protein–protein or protein–ligand interactions. Related technologies are developing rapidly, including bioreactors for multistep enzymatic reactions and bioadsorbents for extraction or degradation of toxic contaminants.

Construction of a Fusion Protein

An affinity-tagged fusion protein typically consists of a single polypeptide chain with one or more affinity tags coupled to the C- or N-terminus of the target protein or inserted into a loop region. The coupling is via a peptide linker, usually up to 15 amino acids in length. If desired, the linker can contain a specific protease site for affinity tag removal. The use of multiple tags is increasingly common and endows the target protein with two or more unique molecular handles.

Selection Criteria for an Affinity Tag

Construction of a fusion protein begins with the selection of an affinity tag that is appropriate for the protein of interest (see below). The selection criteria include the size of the affinity tag, the organism that will be used to express the fusion protein, conditions for the immobilization or purification of the fusion protein, and the influence of the affinity tag on the structural and functional characterization of the target protein.

Affinity-Fusion Gene Constructs

To incorporate the selected affinity tag, a suitable expression plasmid vector is chosen and recombinant DNA techniques are

used to insert the target gene next to the affinity tag gene. Numerous vectors are commercially available that code for a selected affinity tag and a linker peptide, the latter often containing an imbedded protease recognition sequence. In addition, an inducible promoter is present to enhance and regulate fusion protein expression, and a multiple cloning site region is included to facilitate insertion of the target gene with the fusion site at the C- or N-terminus as desired. In some cases, the coding region for a signal peptide is included to promote the secretion of the fusion protein into a specific cellular compartment or into the extracellular medium. Newly developed vectors fuse two different tags to each terminus of a target protein, thereby allowing two-step isolation schemes with enhanced purity and ensuring that only fully intact forms of the fusion protein are isolated.

Linker Peptide Composition

The composition of the linker peptide can have a substantial impact on the function and utility of the fusion protein; therefore, a number of factors are considered during linker design. Typically, the linker is at least 5–10 residues long to allow free tumbling of the fusion and target proteins relative to one another, thereby preventing unwanted steric hindrance to binding events. For linkers containing a proteolytic cleavage site (typically for enterokinase, factor Xa, thrombin, or tobacco etch virus (TEV)), inclusion of 5–10 flanking amino acids at both ends of the site ensures adequate accessibility to protease. Finally, the linker amino acid composition must be selected to minimize susceptibility to host organism proteases.

Overview of Fusion Protein Production and Immobilization

Purification of a recombinant fusion protein begins with the introduction of the expression plasmid into the host cell, followed by cell growth and induction of the fusion promoter. The host cells are lysed and the resulting lysate, containing the affinity-tagged protein, is incubated with a solid-phase support to which the binding partner is coupled. Following binding of the tag to its immobilized partner, the support is washed

to remove unbound components, yielding specific isolation of the tagged protein. If desired, the tagged protein can be released by either disrupting its interaction with the binding partner or by proteolysis of the linker region. Isolation and immobilization procedures often need to be optimized for each new affinity-tagged protein and expression host.

Features and Types of Affinity Tag Systems

The large number of affinity tags currently available offers a diverse spectrum of biochemical properties that can be exploited for protein immobilization in a variety of contexts. Affinity tags range in size from short peptides less than 1 kDa to proteins as large as 120 kDa, and they can be classified into general categories based on their binding partner interaction such as protein–ligand, polyamino acid–matrix, antigen–antibody, and protein–protein. New types of tags based on novel binding partner interactions continue to emerge at a rapid pace.

When choosing an affinity tag for a specific application, several features are considered. A subset of affinity tags possesses the useful property of binding to their partner even under denaturing conditions. The ability to immobilize a tagged protein under denaturing conditions is a great advantage when the fusion protein is expressed in an insoluble or non-native form. Other tags utilize a binding partner interaction that is easily disrupted under mild conditions, thereby facilitating the recovery of target protein with full biological activity. Affinity tag size can also impact target protein function and structural integrity. Small tags generally have less impact on protein structure and function and often need not be removed, while large tags can be perturbing and thus are typically removed prior to structural and functional studies. On the other hand, certain large tags can significantly enhance the solubility of a fusion protein expressed at high levels; therefore, they are appropriate choices for the isolation of poorly soluble target proteins. Finally, when a target protein is toxic to the expression host, affinity tags that promote the aggregation of fusion protein into insoluble aggregates known as inclusion bodies are used. Key features of representative affinity tags are discussed below, and [Table 1](#) presents a more comprehensive list.

Protein–Ligand Interaction Affinity Tags

Enzymes and small-molecule-binding proteins are designed to bind their ligands with high specificity; thus, many highly effective affinity tag systems are based on ligand-binding interactions. One of these is the widely utilized glutathione S-transferase (GST) tag (26 kDa) that binds with high specificity and low affinity ($K_D \sim 180 \mu\text{M}$) to its substrate glutathione. Although the affinity of the GST tag for its ligand is low, commercially available immobilized glutathione matrices provide high local ligand concentrations that ensure adequate retention of the fusion protein. GST-tagged proteins are typically isolated from crude cell lysates by their interaction with immobilized glutathione on beads and then eluted by competitive displacement with free, reduced glutathione. The low ligand affinity enables elution under mild, nondenaturing conditions.

The GST tag often increases fusion protein solubility and stability, and it is ideally suited for overexpression of recombinant proteins in their native state. Often, the GST tag is proteolytically removed at the end of the purification due to its large size and tendency to form dimers. In other applications, the GST tag is retained and used to couple the target protein to a bead or surface for biochemical or biophysical studies.

Another family of affinity tags that relies on protein–ligand interactions are the cellulose-binding domains (CelBDs). CelBDs are small, highly stable protein modules consisting of between 33 and 180 amino acids (3–20 kDa) that bind to different forms of cellulose or chitin with a broad range of affinities ($K_D \sim 0.01\text{--}400 \mu\text{M}$). CelBD fusions can be constructed with affinity tags inserted internally or linked to either the C- or N-terminus of the protein of interest. Over 180 different CelBDs have been identified, providing a spectrum of polysaccharide specificities, binding affinities, and targeting properties. Certain CelBDs bind cellulose essentially irreversibly, making them ideal for protein immobilization, while other CelBDs exhibit easily reversible binding. Denaturing conditions or proteolysis are often needed to elute irreversible CelBDs, while reversible CelBDs can be eluted under mild conditions such as the use of a competitive ligand (cellulose or ethylene glycol) or desorption with water. Finally, CelBDs can be selected that target the fusion protein for secretion or inclusion body formation.

The 51-amino-acid chitin-binding domain (ChiBD) is derived from *Bacillus circulans* chitinase and exhibits high-affinity, essentially irreversible binding to its ligand chitin, a natural carbohydrate polymer. ChiBD fusion proteins are immobilized under physiological conditions by adsorption to chitin coupled to a solid-phase material. ChiBD fusion proteins often have an intein element incorporated into the linker region in place of a proteolytic recognition sequence. When remobilization of the fusion protein is desired, the intein element is activated and undergoes self-cleavage, resulting in the release of the target protein, while the intein–ChiBD region is retained on the solid phase. Similar inteins will probably soon be incorporated into the linkers of other affinity tags. Several other protein–ligand affinity tag systems are currently in use, each having unique advantages. For example, the 40-kDa maltose-binding protein (MalBP) is a large tag that often increases fusion protein solubility. MalBP binds to cross-linked amylose with moderate affinity (K_D in micromolar range), permitting MalBP-tagged fusion proteins to be immobilized in a reversible fashion. The full-length MalBP tag directs a fusion protein to the oxidizing *Escherichia coli* periplasm, where MalBP is a native protein and where the fusion protein may benefit from disulfide bond formation. Alternatively, removal of the MalBP leader peptide yields retention of the fusion protein in the cytosol, where the greater volume can allow higher expression levels.

Polyamino Acid–Matrix Interaction Affinity Tags

Immobilized-metal affinity chromatography (IMAC) is a common technique used to purify recombinant proteins fused to a short peptide affinity tag. IMAC, which can be carried out under denaturing or nondenaturing conditions, relies on the interaction between multiple electron donors on the affinity

Table 1 Commonly used affinity tags and their features

Category of tag	Name of tag	Number of residues in tag	Size of tag (kDa)	Location of tag	Immobile phase-binding partner	Nondenaturing		Expression host ^a
						Immobilization conditions	Affinity tag elution conditions (eluent)	
Protein–ligand interactions	Glutathione S-transferase	211	26	N-term, C-term	Glutathione	Yes	Yes (10 mM Reduced glutathione)	B, Y, I, M
	Cellulose-binding domains	27–189	3–20	N-term, C-term, Internal	Cellulose and other cellulotics	Yes	Yes (water) or no (guanidinium HCL)	B, Y, I, M
	Maltose-binding protein	396	40	N-term, C-term	Cross-linked amylose	Yes	Yes (10 mM maltose)	B
	Chitin-binding domain	51	5.6	N-term, C-term	Chitin	Yes	No	B
Polyamino acid–matrix interactions	Polyhistidine	6–12	0.8–1.7	N-term, C-term, Internal	Metal chelate	Yes or no (denaturants)	Yes (imidazole) or No (low pH)	B, Y, I, M, P
	Polyarginine	5–15	0.8–2.4	C-term	Cation-exchange resin	Yes	Yes (NaCl gradient at alkaline pH)	B, P
Antibody–epitope interactions	Hemagglutinin (HA)	9	1.1	N-term, C-term	IgG	Yes	No (low pH)	Y, M
	T7	11	1.1	N-term	IgG	Yes	No (low pH)	B, Y, M, P, I
	Synthetic protein A (Z domain)	58	7	N-term	IgG	Yes	No (low pH)	B, Y, I, M, P
Protein–protein interactions	FLAG	8	1.0	N-term, C-term	mAb M1 and mAb M2	Yes	Yes (EDTA)	B, Y, I, M, P
	c-myc	11	1.2	N-term, C-term	IgG	Yes	Yes (free c-myc epitope)	B, Y, I, M, P
	Biotin	~20–100	~2–11	N-term, C-term	Modified avidin or streptavidin	Yes	Yes (20 mM biotin)	B, I, M
	Calmodulin-binding peptide	26	3	N-term, C-term	Calmodulin	Yes (+Ca ²⁺)	Yes (+EGTA)	B
	Strep-tag II	8	1.1	N-term, C-term, internal	Strep-tactin	Yes	Yes (desthiobiotin)	B, Y, I, M, P
	Streptavidin-binding peptide	38	4	C-term	Streptavidin	Yes	Yes (biotin)	B
	Thioredoxin	109	11.7	N-term, C-term	Phenylarsine oxide-agarose	Yes	Yes (β -mercaptoethanol)	B
	His-patch thioredoxin	109	11.7	N-term, C-term	Metal chelate	Yes or no	Yes (imidazole) or no (low pH)	B
	S-tag	15	1.8	N-term, C-term	S-protein	Yes	Yes (2 M sodium thiocyanate)	B, I, M

^aB, bacteria; Y, yeast; I, insect; M, mammalian; P, plant.

tag with a transition metal ion (Co^{2+} , Ni^{2+} , Cu^{2+} , Zn^{2+}) chelated to a solid-phase support. The affinity tag is typically polyhistidine, ranging 6–12 residues in length fused to the N- or C-terminus of the target, where 6-His is most common and the electron donor is the histidine imidazole ring. Recently, an affinity tag based on a natural peptide derived from the N-terminus of chicken lactate dehydrogenase has also been utilized. This HAT-tag contains six histidine residues interspersed within a 19-residue polypeptide (KDHLI HNVHK EEHAH AHNK) that binds to specifically Co^{2+} -carboxymethyl-aspartate and has a lower net charge than polyhistidine tags. The most widely employed metal chelator is iminodiacetic acid (IDA), while nitrilotriacetic acid (NTA) and carboxymethylated aspartic acid (trade name TALON) are also popular. The chelator is generally covalently coupled to polymer beads, or ferromagnetic beads for magnetic isolation. Adsorption of IMAC-tagged proteins is normally performed at neutral to slightly basic pH to ensure that the histidine imidazole groups are not protonated. Mild elution conditions include ligand exchange with imidazole, extraction of the metal ion by a strong chelator like ethylene diamine tetraacetic acid (EDTA), or proteolytic elution. Low pH will also elute, but it can sometimes denature the target protein. IMAC is not recommended for target proteins possessing a metal center, since the metal can be stripped by the binding partner chelators.

The Arg-tag is another example of a small polyamino acid affinity tag, in this case designed to raise the isoelectric point of the fusion protein to enhance its binding to a cation-exchange matrix. Mild elution conditions are generally a NaCl gradient at alkaline pH. This tag also binds to flat mica sheets, which may enable a variety of new applications.

Antibody–Epitope Interaction Affinity Tags

Recombinant DNA technology has been employed to create peptide epitopes that bind well-characterized antibodies. This process is known as epitope tagging and is complementary to a more traditional approach where a novel antibody is generated for an existing epitope. Both approaches have been exploited to develop affinity tag systems based on the binding interaction between a peptide epitope and its specific antibody partner. In some cases, several antibodies with different properties are available for a given epitope. Proteins fused to peptide epitope affinity tags can be captured by immunoaffinity interactions via immobilized monoclonal antibodies. Typical epitope tags range from 6 to 30 residues in length, are highly charged, and have little effect on protein structure–function. They can be fused at either the C- or N-terminus, or even inserted within the protein of interest. Epitope tag expression vectors are commercially available for mammalian, insect, yeast, and bacterial host cells and provide a variety of tags, including c-myc, FLAG, HA, T7, V5, vesicular stomatitis virus-G (VSV-G), recA, Protein C, Protein A, and Protein Z. Some of these are optimized for immobilization while others are primarily used for protein detection. The importance of epitope tags is illustrated by their many applications including analysis of *in vivo* protein expression, subcellular localization of gene products, determination of protein topology, cellular trafficking studies, and the investigation of protein–protein interactions. They are also important in protein purification and immobilization.

Antibody-binding partners are covalently coupled to agarose chromatography resins or magnetic glass beads. Immobilized antibodies tend to be less stable than many other affinity-binding partners due to their need for native structure. Elution under mild native conditions can be achieved by competitive displacement with free epitope, by proteolysis, or by exposure of the Ca^{2+} -dependent FLAG-tag to EDTA.

Protein–Protein Interaction Affinity Tags

Several high-affinity and high-specificity protein–protein interactions identified in biological systems have been adapted for use as affinity tags. *In vivo* biotinylation enzymatically couples biotin, a small-molecule vitamin, to a protein acceptor domain. The resulting modified protein provides an interesting example of a natural affinity tag. A number of well-characterized protein sequences are enzymatically biotinylated *in vivo*. These natural peptides of approximately 100 residues have been used as fusion tags that direct *in vivo* biotinylation of a fusion protein in a site-specific manner. In addition, shorter synthetic peptides (approximately 20 residues) known as biotin acceptor peptides (BAPs) have been developed for use as biotin affinity tags.

The high-affinity, specific interaction between biotin and either avidin or streptavidin ($K_D \sim 10^{-15}$ M) serves as the basis for immobilization. Biotin-tagged fusion proteins are typically captured by binding to monomeric avidin or streptavidin coupled to a chromatography matrix or other surface. Elution with free biotin requires harsh conditions that can be detrimental to the target protein, but proteolytic elution can be carried out under mild conditions. Alternatively, the biotin-tag fusion can be immobilized to modified forms of avidin or streptavidin that exhibit lower biotin affinity. Elution can then be conducted under nondenaturing conditions by competitive displacement with free biotin or alkaline pH.

Other examples of protein–protein affinity tags include the Strep-tag, the calmodulin-binding peptide (CalBP), and the S-tag. The Strep-tag is a nine-amino-acid peptide that was developed as an artificial ligand for streptavidin. Further refinements to these molecules yielded the eight-amino-acid peptide Strep-tag II and a mutant form of streptavidin called Strep-Tactin capable of moderate-affinity binding ($K_D \sim 1 \mu\text{M}$). Strep-tag fusion proteins are bound to streptavidin under physiological conditions and are efficiently displaced by free biotin or a biotin derivative. The mild conditions used for binding and elution enable Strep-fusions to be utilized in their native state. This small, stable, and nonperturbing tag is ideal for many applications in which a small tag is required, and for applications where detection of the fusion protein on a Western blot or by enzyme-linked immunosorbent assay (ELISA) is desired. Moreover, as the Strep-tag is not itself a metalloprotein and its elution does not require metal chelators, it is well suited for metalloprotein applications.

The 26-amino-acid CalBP was derived from the C-terminus of skeletal muscle light-chain kinase. The CalBP binds to calmodulin with high affinity ($K_D \sim 1 \text{ nM}$) in the presence of low CaCl_2 concentrations ($\geq 0.2 \text{ mM}$). The binding partner, calmodulin, is commercially available linked to chromatography resins that are used to immobilize CalBP fusions. CalBP fusion proteins are eluted under mild conditions by a Ca^{2+} chelating agent such as ethylene glycol tetraacetic acid (EGTA). CalBP

fusions are primarily expressed in *E. coli* that lacks calmodulin family members and thus possesses no sequences evolved to bind to calmodulin. Eukaryotic cells are not recommended for expression of CalBP-tagged proteins due to their large number of endogenous proteins (~30) that bind to calmodulin.

The S-tag protein fusion system consists of the S-peptide and its binding partner, the S-protein, both derived from RNAase A. The S-tag binds to the S-protein with moderate affinity ($K_D \sim 100$ nM), resulting in a strong interaction that is influenced by pH, temperature, and ionic strength. A unique feature of the S-tag system is that ribonucleolytic activity is restored when the S-peptide is bound to the 103-amino-acid S-protein. This enzymatic activity is exploited to measure the molar concentration of the S-tag fusion protein down to 20 fM in a highly sensitive and rapid assay. S-protein-based reagents have been developed to probe sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) blots for S-tag fusions employing colorimetric or chemiluminescent detection of the target protein down to nanogram quantities.

See also: **Protein/Enzyme Structure Function and Degradation: Two-Hybrid Protein–Protein Interactions.**

Further Reading

- Bornhorst J and Falke J (2000) Purification of proteins using polyhistidine affinity tags. *Methods in Enzymology* 326: 245–254.
- Jarvic J and Telmer C (1998) Epitope tagging. *Annual Review of Genetics* 32: 601–618.
- Nilsson J, Ståhl S, Lundeberg J, Uhlén M, and Per-Åke N (1997) Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. *Protein Expression and Purification* 11: 1–16.
- Sheibani N (1999) Prokaryotic gene fusion expression systems and their use in structural and functional studies of proteins. *Preparative Biochemistry and Biotechnology* 29: 77–90.
- Terpe K (2003) Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology* 60: 523–533.

Algal Hydrogen Production

U M N Murthy and M L Ghirardi, Bioscience Center, National Renewable Energy Laboratory (NREL), Golden, CO, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Catalytic H-cluster The catalytic core of the [FeFe]-hydrogenase (HydA), termed the 'H-cluster', exists as a [4Fe-4S] subcluster linked by a cysteine thiolate to a modified 2Fe subcluster with unique nonprotein ligands.

Fuel cells An electrochemical cell that converts a source fuel into an electrical current. It generates electricity inside a cell through reactions between a fuel and an oxidant, triggered in the presence of an electrolyte.

Photobioreactor A transparent reactor that allows growth of cells under illumination; this term mostly refers to closed systems that have no direct exchange of gases and contaminants with the environment, as opposed to open ponds.

Radical SAM proteins A novel protein superfamily with over 600 members that catalyze diverse reactions, including unusual methylations, isomerization, sulfur insertion, ring formation, anaerobic oxidation, and protein radical formation.

Introduction

Fossil fuels have been utilized as a source of energy for many centuries, resulting in congestion, pollution, and growing environmental stress. Global oil supplies are rapidly being depleted and production levels are increasingly being reported to be close to their peak. More importantly, zero CO₂ emission fuels are becoming increasingly important due to the constraints of global climate change. Therefore, the development of new systems to produce these fuels is one of the greatest challenges facing our society. Solar energy is the most abundant and accessible renewable energy source available for future sustainable production of fuel and electricity. For effective use of solar energy, it is important to develop more cost-effective systems with the improved ability to convert solar energy into chemical energy. In this context, hydrogen (H₂) is considered to be a promising, clean energy carrier for future technologies. The fact that the burning of H₂ produces only H₂O increases its attractiveness, particularly in light of recent advances in hydrogen fuel cell technology. The direct photoproduction of H₂ by cyanobacteria and green algae is an alternative to utilizing the reductant generated from photosynthetic water reduction for the fixation of CO₂ into carbon-containing molecules.

Hydrogen metabolism is primarily the domain of bacteria and microalgae. Within these groups, it involves many taxonomically diverse species, a variety of enzymes, and several metabolic pathways and processes. Green algae and cyanobacteria are photoautotrophic organisms; they can grow under only sunlight and CO₂, without organic sources of carbon. It has been known for more than 60 years that, under anaerobic conditions, unicellular green algae can metabolize H₂, either by uptaking it and using it as an electron donor in CO₂-fixation, or by evolving H₂ in the light. Indeed, hydrogenase enzymes that either uptake or evolve H₂ have been found in many green algae. This article briefly examines green algal hydrogen metabolism and how it can be adapted for commercial H₂-production processes.

Green Algal Pathways for H₂ Production

The unicellular alga *Chlamydomonas reinhardtii* belongs to the group of eukaryotic green algae and lives in freshwater or in moist places. In addition to photosynthesis, unicellular green algae are able to perform many biotechnologically interesting metabolic reactions, such as fermentation and hydrogen photoproduction. The starting point for all biological solar-driven H₂-production methods is the process of photosynthesis (see [Figure 1](#)). Oxygenic photosynthesis involves a sequential chain of reactions that include light absorption, charge separation, electron transport, and dark CO₂ fixation. Oxygenic organisms harness solar energy to extract electrons from H₂O, which are required for CO₂ fixation. Electrons generated from H₂O molecules at photosystem II (PSII) are transferred to the cytochrome *b6/f* complex (*b6f*) and subsequently delivered to photosystem I (PSI). Through PSI, the electrons are transferred to ferredoxin (Fd) (a 2Fe-2S protein). Normally, Fd shuttles electrons to the enzyme ferredoxin-nicotinamide adenine dinucleotide phosphate (NADP)-oxidoreductase (FNR) that reduces NADP⁺ to NADPH, the direct source of reductant to convert CO₂ into carbohydrates in the Calvin-Benson cycle. Concomitantly, protons (H⁺) are carried from the stromal to the lumenal side of the thylakoids, creating a proton gradient across the thylakoid membrane, which is utilized by the enzyme adenosine triphosphate (ATP) synthase to generate the ATP that is also required for CO₂ fixation. Under anaerobic conditions, reduced Fd efficiently binds to an [FeFe]-hydrogenase and donates electrons to its catalytic site, known as the 'H-cluster' (HC). The function of the hydrogenase is to combine protons (H⁺) and electrons (e⁻) to form and release molecular H₂. Photosynthetic reducing power can thus be partitioned between at least two pathways: CO₂ reduction (under aerobic conditions) and H₂ production (under anaerobiosis). It is important to point out that CO₂ reduction requires ATP, whereas H₂ photoproduction does not. Instead, the latter is downregulated in the absence of ATP production (see [section H₂ Production: Issues and Challenges](#)). As expected, removal

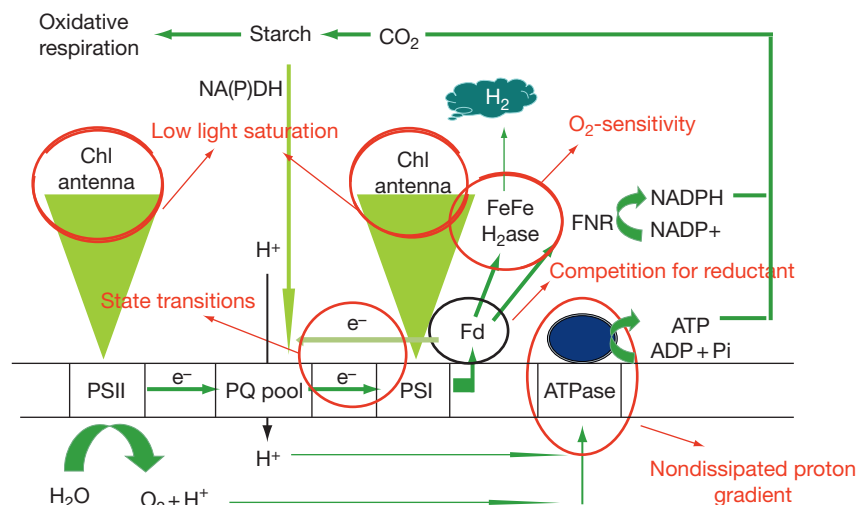


Figure 1 Photosynthetic electron transport pathways in the green alga, *Chlamydomonas reinhardtii*. Photosynthesis extracts electron from H₂O and via the electron transport chain delivers them to the [FeFe]-hydrogenase where H₂ is generated. The red circles denote the major barriers limiting technological use of algae for photobiological H₂ production. Adapted from Ghirardi ML and Mohanty P (2010) Oxygenic hydrogen photoproduction – current status of the technology. *Current Science* 98: 499–507, with permission from Current Science.

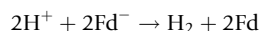
of CO₂ from the headspace enhances light-driven H₂-evolution, indicating competition for electrons between the CO₂-fixation and the H₂-evolution processes.

Oxygenic H₂ photoproduction results in the simultaneous generation of H₂ and O₂ in an ~2:1 ratio as long as O₂ is continuously removed from the headspace. In the absence of active removal of O₂, this mechanism can operate for only 30–90s since O₂ is a powerful inhibitor of the [FeFe]-hydrogenase reaction (and it is also a suppressor of [FeFe]-hydrogenase gene transcription, which is discussed in detail in later sections).

An alternate source of electrons for H₂ photoevolution is endogenous substrate degradation. This pathway is dependent on the reduction of the plastoquinone (PQ) pool by NADH (released from the initial step of glycolysis), in a reaction catalyzed by the PQ/NAD(P)H-oxidoreductase. Light absorption by PSI and ensuing electron transport elevates the redox potential of these electrons to the redox equivalent of ferredoxin, thus permitting the generation of molecular H₂.

Enzymes for Algal Photohydrogen Production

Hydrogenases are enzymes capable of producing or uptaking molecular hydrogen. Hydrogenases are classified into three main groups, according to the cofactor(s) they contain in their catalytic site, as iron–iron [FeFe], nickel–iron [NiFe], or Fe-only hydrogenase. Algal hydrogenases belong to the class of [FeFe]-hydrogenases, they contain only Fe and S in their catalytic site, and are typically involved in H₂ production rather than H₂ oxidation. They catalyze the reversible reduction of protons to H₂:



[FeFe]-hydrogenases are small, monomeric enzymes of 45–49-kDa size containing a novel type of [Fe–S] cluster commonly referred to as the ‘H-cluster’ at their catalytic center. The HC comprises a conventional [4Fe–S] complex bridged by

the sulfur atom from a cysteine residue to a unique binuclear iron–sulfur subcluster. Three other cysteines coordinate the [4Fe4S]-center to the protein in all known [FeFe]-hydrogenases. Except for the bridging cysteine, the iron atoms of the 2Fe2S center are coordinated to carbon monoxide (CO) and cyanide (CN[−]) ligands, which are responsible for stabilizing the reduced form of the enzyme, and by a dithiomethylamine group linking the two S atoms. One of the CO groups is found in a position bridging both iron atoms when the enzyme is in an oxidized state, but its orientation has been proposed to change toward the distal Fe when the enzyme is in the reduced state. Highly conserved residues comprising several hydrophobic amino acids are thought to be involved in the formation of H₂ channels, thus connecting the catalytic site to the protein surface.

In *C. reinhardtii*, a monomeric [FeFe]-hydrogenase of 48 kDa (named HYDA1) with high specific activity was first isolated and biochemically characterized. More recent research revealed the presence and expression of a second [FeFe]-hydrogenase (HYDA2) in *C. reinhardtii* that is 74% similar and 68% identical to HYDA1. The transcripts for both hydrogenases, HYDA1 and HYDA2, are selectively expressed upon dark anaerobic incubation, suggesting regulation of gene expression by O₂. HYDA1, and most probably HYDA2, directly interact with reduced ferredoxin. The nucleus-encoded algal hydrogenases are synthesized in the cytosol as precursor proteins but the mature proteins are localized in the chloroplast stroma. A transit peptide domain that routes the [FeFe]-hydrogenases from the cytoplasm across the chloroplast envelope and into the chloroplast stroma has been identified at the N-terminal region of the enzyme.

[FeFe]-Hydrogenases Are Sensitive to O₂

The metallo-cluster of the [FeFe]-hydrogenases is very sensitive to O₂. This observation, together with the fact that hydrogenases are located in the chloroplast, where PSII releases O₂, causes H₂

production rates to be usually low. In algal cells, O_2 can adversely affect enzymatic activity by deactivating previously assembled [FeFe]-hydrogenases, and the isolated enzyme competes with CO for the same binding site in the HC of [FeFe]-hydrogenases. Oxygen inactivation is thought to occur by the direct interaction of O_2 with the [2Fe-2S] center on catalytic HC. In contrast to O_2 inactivation, inhibition by CO is largely but not entirely reversible. The nature of the final oxygen species bound to the inactivated enzyme is not known. However, it is known that diffusion of O_2 gas is necessary for the inactivation of the enzyme, since the catalytic site of the enzyme is embedded in the protein structure.

Melis and his coworkers in 2000 demonstrated a novel strategy to sustain green algal H_2 production in the light. The O_2 sensitivity of H_2 metabolism was bypassed in a two-stage process, in which O_2 -evolution and H_2 -evolution reactions were temporarily separated. In the first stage, normal photosynthesis, CO_2 -fixation, and release of O_2 upon oxidation of H_2O enabled green algae to survive. In the second stage, the authors selectively and partially inactivated the O_2 -evolving activity of PSII by incubating algal cultures in a sulfur-depleted medium. Since respiratory O_2 consumption is not affected by sulfur deprivation, inactivation of O_2 evolution facilitated the transition from the aerobic to an anaerobic state in a sealed photobioreactor. The establishment of anaerobiosis in the photobioreactor induced the expression of the two [FeFe]-hydrogenases in algal cells and inhibited the expression of ribulose-1,5-bisphosphate carboxylase oxygenase (Rubisco). This resulted in redirection of the flow of electrons from the photosynthetic electron transport chain in the chloroplast away from carbon fixation and towards proton reduction. As a consequence, sulfur-deprived algae were able to produce H_2 for a total of 3–4 days.

[FeFe]-Hydrogenase Maturation

Biosynthesis of an active [FeFe]-hydrogenase is a complex process. [FeFe]-hydrogenase maturation proteins were initially discovered in the green alga *C. reinhardtii*. Posewitz et al. showed that two novel radical S-adenosyl-methionine (radical SAM) proteins, HYDEF and HYDG, are required for *C. reinhardtii* [FeFe]-hydrogenase activity. Using a chemochromic sensor, a *hydef* mutant that was unable to produce hydrogen was isolated from a *C. reinhardtii* mutagenesis library. Moreover, the *HYDEF* gene was found to be adjacent to the *HYDG* gene, and both were shown to be present in all other [FeFe]-hydrogenase-containing organisms, although their role was unknown at that time. In the absence of a functional *HYDEF*, both hydrogenase structural genes were induced but no hydrogenase activity was detected in the mutant. Complementation of the mutant with genomic DNA containing a functional copy of the *HYDEF* gene restored hydrogenase activity in the algae. Later on, King et al. showed that the heterologous expression of active algal [FeFe]-hydrogenase in the bacterium, *Escherichia coli*, could be accomplished by co-transforming the *HYDA1* structural gene along with both the *HYDEF* and *HYDG* genes, and that all these genes were required and sufficient for the expression of a functional [FeFe]-hydrogenase in this bacterium. The algal *HYDEF* gene encodes for two unique protein

segments, HYDE and HYDF, that are exclusively found in other organisms containing [FeFe]-hydrogenase but encoded as two separate genes, *HydE* and *HydF*. The HydE proteins are highly homologous to the biotin synthase radical SAM proteins, and HYDF shares homology with the large family of nucleoside triphosphate (NTP)-binding proteins.

The specific role of HydE, HydF, and HydG proteins in [FeFe]-hydrogenase maturation is not clearly understood yet. Peters et al. proposed that the maturation proteins use common amino acids or other metabolites to synthesize a precursor to the [FeFe]-hydrogenase active site, which is then transferred to the structural enzyme. HydE and HydG act on HydF to assemble an active-site precursor containing all the nonprotein ligands of the binuclear Fe subcluster, which is subsequently transferred to the hydrogenase. It seems that HydE is probably responsible for the synthesis of CO and CN^- ligands and HydG for the dithiomethylamine ligand. The dithiomethylamine-bridged cluster is then transferred as a whole to HydF to generate a complete, functional HC for insertion to the [FeFe]-hydrogenase apoprotein.

The Interaction of [FeFe]-Hydrogenases with Ferredoxin

In green algae, H_2 production couples to the photosynthetic electron transport chain via ferredoxin. Ferredoxins are small soluble [Fe-S] proteins which can be found in every phylogenetic group. They are mainly low-potential electron carriers involved in various metabolic pathways having diverse redox partners. Ferredoxin distributes photosynthetic electrons not only to hydrogenase but also to other electron-dependent enzymes, the most important of which is ferredoxin-NAD(P)-reductase (FNR) for CO_2 fixation.

There are six ferredoxin genes in *C. reinhardtii*. The physiological regulation and metabolic function of each of the six ferredoxins has not yet been established, although it is assumed that both hydrogenases may associate with the photosynthetic electron transfer ferredoxin (PetF, also called FDX1) which is involved in photosynthesis and it is a typical plant-type [2Fe-2S] ferredoxin. Recently, Winkler et al. characterized the kinetics of electron transfer processes between HydA1 and PetF. Using an *in vitro* system by reconstituting part of the photosynthetic electron transport chain consisting of plastocyanin, PSI, PetF, and [FeFe]-hydrogenase, they demonstrated that residue K396 of HydA1 is crucial for a successful binding and electron transfer between PetF and HydA1. However, other studies by Terauchi et al. found that each of the six *FDX* genes is differentially regulated in response to changes in nutrient supply and/or anoxia. In this regard, more detailed understanding of the hydrogenase-ferredoxin association mechanism and kinetics is necessary for engineering hydrogenase to divert a greater electron flux to the H_2 production pathway and thus to optimize photobiological H_2 production.

H_2 Production: Issues and Challenges

Photobiological H_2 production by *C. reinhardtii* has become an important research field because of its potential to be applied

in renewable energy production. To realize the economic potential of producing hydrogen photobiologically, several challenges must be overcome. Foremost, cellular metabolism and the basic biochemistry supporting this process must be understood. Following are the major issues and challenges that need to be addressed.

O₂ Sensitivity

The simultaneous photoproduction of H₂ and O₂ from H₂O in green algae depends on O₂ tolerance of the reversible hydrogenase enzyme. When oxygenic photosynthesis is active, and unless strict measures are taken to remove O₂, green algal H₂-production activity is transient owing to the rapid inactivation of the reversible hydrogenase by photosynthetic O₂. Sustained hydrogen production will only be possible when the hydrogenase enzyme remains active under O₂-evolving conditions. Different strategies for development of O₂ tolerant hydrogenases have included molecular engineering of the hydrogenases by either random or site-directed mutagenesis to remove O₂ sensitivity. The latter has mostly focused on physical restriction of O₂ access catalytic site by reducing the O₂ diffusion rate or shielding the active catalytic site. Although the results are encouraging from a scientific perspective in understanding enzyme structure and function, none of these approaches has yielded biotechnologically significant hydrogen-producing strains yet.

An alternative approach to sustain H₂ production in the light is to use sulfur deprivation to partially inactivate O₂ evolution, as discussed above. The novel application of this process has led to significant and sustainable light-dependent release of H₂ by algal cultures, although at low-conversion efficiencies due to limited PSII activity. Several approaches have been examined to increase the yield of hydrogen in *C. reinhardtii* under sulfur-deprived condition, including optimization of the light and pH conditions in the photobioreactors optimization of the medium composition and synchronization of cell division (see Figure 2). Kruse et al. reported a significant increase in the rate and duration of H₂ photoproduction in sulfur-deprived mutants that are starch over-accumulators and blocked in state transitions. Another advance related to the sulfur-deprived process came with the recent discovery of a mutant affected in sulfate permease activity, which is required for transport of sulfate into the chloroplast. This mutant is a potential candidate for H₂ photoproduction without the need to totally remove sulfate from the growth medium.

The reported rates of H₂ production by sulfur-deprived cultures is still far below the maximum potential rate of H₂ photoproduction for an algal system, mainly due to partial inactivation of PSII. Furthermore, the system is difficult to scale up and maintain. These challenges, which are related partly to the photobioreactor design, might be addressed by immobilizing the green algae. The immobilization technique effectively separates cells from the liquid phase, significantly increases the cell density and, as a consequence, allows for more efficient light utilization on a per area basis. Laurinavichene, et al. in 2006 demonstrated that sulfur-deprived, immobilized cells of a nonmotile mutant of *C. reinhardtii* on glass fibers significantly increase the duration of H₂ production. More recently, Kosourov and Seibert, demonstrated that



Figure 2 Photograph of the photobioreactor used for H₂ production from *Chlamydomonas reinhardtii* culture showing probes used for monitoring and the controlling the system. Hydrogen bubbles emanate toward the surface of the culture and are collected in a liquid-accumulating bottle (not shown).

immobilization of sulfur/phosphorous-deprived algae in thin Ca²⁺-alginate film matrices has extended H₂ photoproduction for more than 150 h and also increased the stability of H₂ photoproduction activity in the presence of atmospheric O₂. This work shows potential for further significant improvements in both the rate and efficiency of H₂ production under sulfur-deprived conditions.

Reduced Antenna Size

Photosynthesis and H₂ production in unicellular green algae can in principle operate with a nearly 100% photon utilization efficiency, making them potentially efficient biocatalysts for the generation of H₂ from sunlight and water. However, green algal cultures show poor light utilization efficiency under direct sun light and can waste up to 80% of the absorbed irradiance. The reason for this is that green algal photosynthesis normally saturates at about one-fifth of full sunlight intensity or less. This optical property of the cells would further lower the productivity of a commercial hydrogen production. Enhanced hydrogen production may be achieved by engineering the antenna size to suppress fluorescence and heat dissipation that causes a reduction on antenna efficiency.

Competition for Reductants

Competition from the CO₂-fixation pathway for electrons from H₂O is a major concern for sustained H₂ production, as discussed earlier in the article. This concern is based on the higher affinity of ferredoxin for FNR than hydrogenase, which could result in substantial drainage of electrons away from H₂ production. Therefore, improved H₂ production depends on interaction of [FeFe]-hydrogenases with the electron-carrying protein ferredoxin. Agapakis et al. fused the hydrogenase from *C. acetobutylicum* to ferredoxin from spinach or from

C. acetobutylicum using flexible protein linkers of various lengths and showed a threefold increase in H₂ production by a nonphotosynthetic, recombinant *E. coli* system. Moreover, chimeras between ferredoxin and various ferredoxin reductases have been shown to be functional *in vitro*, with improved electron transfer rates, presumably due to the increased local concentration of reduced ferredoxin. This approach may provide new targets for metabolic engineering of H₂ production in the future.

Downregulation of Electron Transport by the Proton Gradient

Under sulfur-deprived conditions, very little CO₂ is fixed into glucose, mainly because of the decrease in Rubisco levels. These sulfur-deprived, anaerobic cultures use very little ATP, as a consequence, the proton gradient generated by photosynthetic electron transport is not dissipated. This causes a decrease in the rate of electron transport as demonstrated by the acceleration in the rate of H₂ production observed upon addition of uncouplers. Although, mutants defective in one or other ATPase subunits have been generated, their uncoupling properties are yet to be studied. These mutants may be able to circumvent the cellular regulatory process that downregulate PSII activity at the level of the plastoquinone pool and limit H₂ production.

State Transition

Oxygenic photosynthetic organisms adapt to varying energetic requirements by transitioning between linear and cyclic electron transport, through the so-called state transitions. If the ATP requirement is higher than what is provided by linear electron transport alone, the cultures transition to state 2, where cyclic electron transport predominates. Anaerobiosis leads to the establishment of state 2 and, as a result, additional limitations are placed on the rates of H₂ photoproduction. The relevance of the state 2 limitation on H₂ photoproduction by sulfur-deprived culture was recently demonstrated by Kruse et al., who used a chlorophyll fluorescence video-imaging screening procedure to identify mutants of *C. reinhardtii* blocked in state 1. They isolated one mutant, *stm6*, with rates of H₂ photoproduction 13 times higher than its wild-type strain, and with extended H₂ photoproduction activity under sulfur deprivation. These results underscore the importance of molecular engineering to address this particular limitation to algal H₂ production technologies.

Conclusion

Biological hydrogen production is not only a challenging but also promising biotechnology area, with the potential to help solve environmental and energy-related problems. The application of biological H₂ production as a mature technology, however, depends on research advances to address improvements in H₂-production rates and efficiency through a combination of genetically engineering microorganisms and developing bioreactors. Significant progress is being made in understanding

the molecular regulation of the genes encoding hydrogenases and accessory proteins needed for assembly of hydrogenases, as well as in identifying and addressing the major barriers to a future commercial process. With increasing interest from the scientific community and the public in clean and renewable energy sources, rapid progress could bring this technology to fruition in the not-so-long future. Moreover, the possibilities and challenges within synthetic biology should be further explored for creating green algae with a high potential for H₂ production.

See also: [Bioenergetics: Ferredoxin; Hydrogenases, Structure and Function; Renewable Hydrogen from Biomass.](#)

Further Reading

- Agapakis CM, Ducat DC, Boyle PM, et al. (2010) Insulation of a synthetic hydrogen metabolism circuit in bacteria. *Journal of Biological Engineering* 4: 3.
- Ghirardi ML (2006) Hydrogen production by photosynthetic green algae. *Indian Journal of Biochemistry and Biophysics* 43: 201–210.
- Ghirardi ML, Dubini A, Yu JP, and Maness PC (2009) Photobiological hydrogen-producing systems. *Chemical Society Reviews* 38: 52–61.
- Ghirardi ML and Mohanty P (2010) Oxygenic hydrogen photoproduction – current status of the technology. *Current Science* 98: 499–507.
- Ghirardi ML, Posewitz MC, Maness PC, et al. (2007) Hydrogenases and hydrogen photoproduction in oxygenic photosynthetic organisms. *Annual Review of Plant Biology* 58: 71–91.
- Ghirardi ML, Zhang JP, Lee JW, et al. (2000) Microalgae: A green source of renewable H₂. *Trends in Biotechnology* 18: 506–511.
- King PW, Posewitz MC, Ghirardi ML, and Seibert M (2006) Functional studies of [FeFe] hydrogenase maturation in an *Escherichia coli* biosynthetic system. *Journal of Bacteriology* 188: 2163–2172.
- Kruse O, Rupprecht J, Bader KP, et al. (2005) Improved photobiological H₂ production in engineered green algal cells. *Journal of Biological Chemistry* 280: 34170–34177.
- Laurinavichene TV, Fedorov AS, Ghirardi ML, Seibert M, and Tsygankov AA (2006) Demonstration of sustained hydrogen photoproduction by immobilized, sulfur-deprived *Chlamydomonas reinhardtii* cells. *International Journal of Hydrogen Energy* 31: 659–667.
- Melis A and Happe T (2001) Hydrogen production. Green algae as a source of energy. *Plant Physiology* 127: 740–748.
- Melis A, Seibert M, and Ghirardi ML (2007) Hydrogen fuel production by transgenic microalgae. *Transgenic Microalgae as Green Cell Factories* 616: 108–121.
- Melis A, Zhang LP, Forestier M, Ghirardi ML, and Seibert M (2000) Sustained photobiological hydrogen gas production upon reversible inactivation of oxygen evolution in the green alga *Chlamydomonas reinhardtii*. *Plant Physiology* 122: 127–135.
- Mus F, Dubini A, Seibert M, Posewitz MC, and Grossman AR (2007) Anaerobic acclimation in *Chlamydomonas reinhardtii* – anoxic gene expression, hydrogenase induction, and metabolic pathways. *Journal of Biological Chemistry* 282: 25475–25486.
- Peters JW, Szilagyi RK, Naumov A, and Douglas T (2006) A radical solution for the biosynthesis of the H-cluster of hydrogenase. *FEBS Letters* 580: 363–367.
- Posewitz MC, King PW, Smolinski SL, et al. (2004) Discovery of two novel radical S-adenosylmethionine proteins required for the assembly of an active [Fe] hydrogenase. *Journal of Biological Chemistry* 279: 25711–25720.
- Prince RC and Khesghi HS (2005) The photobiological production of hydrogen: Potential efficiency and effectiveness as a renewable fuel. *Critical Reviews in Microbiology* 31: 19–31.
- Stripp ST and Happe T (2009) How algae produce hydrogen—news from the photosynthetic hydrogenase. *Dalton Transactions* 45: 9960–9969.
- Terauchi AM, Lu SF, Zaffagnini M, et al. (2009) Pattern of expression and substrate specificity of chloroplast ferredoxins from *Chlamydomonas reinhardtii*. *Journal of Biological Chemistry* 284: 25867–25878.
- Winkler M, Kuhlert S, Hippler M, and Happe T (2009) Characterization of the key step for light-driven hydrogen evolution in green algae. *Journal of Biological Chemistry* 284: 36620–36627.

Allosteric Regulation

B S Cooperman, University of Pennsylvania, Philadelphia, PA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Allosteric effector A molecule binding to a protein site other than the active site, which can either increase or decrease the activity of the protein.

Heterotropic cooperativity The effect of an allosteric effector on protein activity toward a substrate.

Homotropic cooperativity The effect of initial substrate binding on subsequent substrate binding.

KNF and MWC models Thermodynamic models that account for the cooperativity displayed by allosteric enzymes.

Allosterism and Cooperativity

Allosteric regulation has been found to be extensive in proteins, particularly in enzymes at key branch points of metabolism and in receptors that must be sensitive to small variation in signals. Although a monomeric protein having one subunit can display allosteric regulation, the great majority of proteins regulated in this manner have multiple subunits, with changes in activity arising from changes in subunit-subunit contacts. A characteristic feature of these regulatory proteins is the occurrence of cooperative interactions for both the substrate and the regulatory ligand. This property renders their function dependent upon threshold concentrations of ligand.

Cooperative Binding of O₂ to Hemoglobin

Cooperativity may be defined as any process in which an initial event affects subsequent similar events. It was initially identified with the sigmoid plot for the binding of four molecules of O₂ to hemoglobin (Hb) (Figure 1), a pseudo-tetrameric protein that gives red blood cells their color. A similar cooperativity of substrate binding occurs in many allosteric enzymes, leading to sigmoid plots of enzyme activity versus substrate concentration. Hb has the subunit composition $\alpha_2\beta_2$, in which the α - and β -subunits are nearly identical to one another, with each containing a heme group to which O₂ binds. The sigmoid plot was explained by the concept that the first molecule of O₂ bound makes it easier for subsequent molecules to bind, and so is an example of positive cooperativity. In fact, if the Hb-binding curve is fitted to four successive binding constants, the affinity for the fourth O₂ bound is calculated to be 100–1000 times as high as the first O₂ bound. In contrast, O₂ binding to myoglobin (Mb), a monomeric protein found in vertebrate muscle with no possibility for site-site interaction, follows a rectangular hyperbola plot.

An essential feature of the positive cooperativity, shown in Figure 1 for Hb, is that it sharpens the responsiveness of a system to a change in substrate (or effector) concentration. Thus, to go from 10% to 90%, O₂ saturation of Mb requires an 81-fold change in O₂ concentration, whereas the corresponding change for O₂ saturation of Hb requires only a fourfold change. Such responsiveness is obviously desirable in a protein whose activity must be highly regulated.

The Hill n , a Measure of Cooperativity

The Hill plot linearizes saturation data over the major portion of the saturation curve (typically 10–90% saturation) yielding a slope, called the Hill n and denoted n_H , which provides a convenient empirical measure of cooperativity. If we define the saturation function Y_S as the fraction of all binding sites containing a bound ligand (i.e., $0 \leq Y_S \leq 1$), then the Hill plot is obtained by plotting $\log(Y_S/(1-Y_S))$ versus $\log[S]$ (when enzyme activity is measured, $\log(v/(V_{\max}-v))$ vs. $\log[S]$ is plotted instead). The value of n_H is 1 for a noncooperative saturation function (e.g., O₂ binding to Mb). Positive cooperativity is defined by an $n_H > 1$, with an upper limit of n , the number of identical subunits, for a cooperative saturation function. For a tetramer, an n_H equal to its upper limit of 4 would correspond to the situation where there were only two protein species in solution, one with no ligands bound and one with ligands bound to all four sites. For Hb, a strongly cooperative tetramer, the n_H value is near 3 under physiological conditions. Some allosteric proteins have $n_H < 1$, that is, have negative cooperativity. In contrast to positive cooperativity, negative cooperativity gives a less sensitive response over a broad range of stimulus and might be important, for example, in the response of growth or general metabolism to changes in hormone concentration.

One caveat to the use of n_H values evaluated from enzyme activity studies is that hysteresis or ligand-induced slow transitions that have the same timescale as the catalytic cycle can give rise to apparent cooperativity, even for monomeric enzymes. In addition, n_H values of <1 , whether determined by binding or enzyme activity studies, can arise from sample heterogeneity, rather than from true negative cooperativity.

Two-State Models to Explain Cooperativity in Allosteric Proteins

Two models were proposed in the 1960s to explain not only the cooperative binding of O₂ to Hb, and of substrate binding to allosteric proteins in general, but also the effects of allosteric effector molecules on substrate binding. Each of the models posits that the identical subunits within the protein can exist in two states, T and R , with the R state having higher activity for substrate. Here, some definitions are in order before proceeding further. O₂ binding to hemoglobin is an example of a

positive homotropic effect, since the initial binding of O_2 increases the affinity for subsequent O_2 molecules. On the other hand, addition of the allosteric inhibitor 2,3-diphosphoglycerate reduces O_2 affinity for Hb, providing an example of a negative heterotropic effect. In a similar fashion, an allosteric activator which increases substrate affinity gives rise to a positive heterotropic effect, whereas a negative homotropic effect is seen when initial binding of substrate decreases the affinity for subsequent substrate molecules.

The first model, proposed by Monod, Wyman, and Changeux in 1965, and known as the MWC (or symmetry) model hypothesizes that: (1) allosteric regulatory proteins, in general, are oligomers made up of a finite number of identical subunits that occupy equivalent positions and, as a consequence, possess at least one axis of rotational symmetry; (2) the allosteric oligomers can exist in two freely interconvertible and discrete conformational states (T or R) that differ in the energy of their intersubunit interactions, but in which molecular symmetry is conserved, so that all subunits are either in the T state or in the R state; (3) in the absence of ligand, the pre-existing conformational equilibrium is characterized by an allosteric constant $L = T_n/R_n$, where n is equal to the number of identical subunits; and (4) ligand affinities for the active and allosteric sites carried by the oligomers may differ between the two states, allowing ligand binding to preferentially stabilize the state for which it exhibits a higher affinity. Such modulation of the conformational equilibrium by ligand-binding suffices to generate positive homotropic effects and both positive and negative heterotropic effects.

The second model, proposed by Koshland, Nemethy, and Filmer in 1966, and known as the KNF (or sequential) model, hypothesizes that in the absence of ligand the protein exists in

a single state, that ligand-binding induces a conformational change only in the subunit to which it binds, and that cooperative interactions arise through the influence of such conformational changes on intersubunit interaction. The KNF model embodies the notion of induced fit, whereby the binding of substrate to an active site causes conformational changes of active site residues that are necessary for the protein's function. Since conformational change only occurs on ligand binding, partially saturated protein contains a mixture of R and T states, so that the symmetry of the oligomeric protein is not preserved during the binding process.

Both the MWC and KNF models are limiting cases of the more general scheme of a two-state model, as shown for the case of substrate binding to a tetrameric protein in Figure 2. In the MWC model, only species corresponding to the left-most (T_nS_i , $i = 0-n$) and right-most (R_nS_i , $i = 0-n$) columns are considered, because it is posited that hybrid states containing both R - and T -subunits are inherently unstable and do not accumulate. In contrast, in the KNF model, only species lying along the diagonal connecting T_n to R_nS_n (i.e., $T_{n-i}R_iS_i$, $i = 0-n$) are considered, because of the requirement that conformational change only results from substrate binding.

Either limiting model generates the sigmoidal saturation curve for substrate binding to an allosteric protein using just three parameters. For the MWC model these are L , the allosteric constant, and K_T and K_R , the dissociation constants for S binding to the T and R states, respectively. In the KNF model the three parameters are $K_S K_t$ (as a product), K_{RT} , and K_{RR} , where K_S is the binding constant for substrate to the R state, K_t is the equilibrium constant for changing a single, isolated, subunit from the T to the R conformation ($= [R]/[T]$), and K_{RT} and K_{RR} measure the relative strength of associations of R - T and R - R intersubunit contacts, respectively, versus the strength of association of T - T intersubunit contact. In the two models, allosteric effectors exert their heterotropic effects by altering the apparent

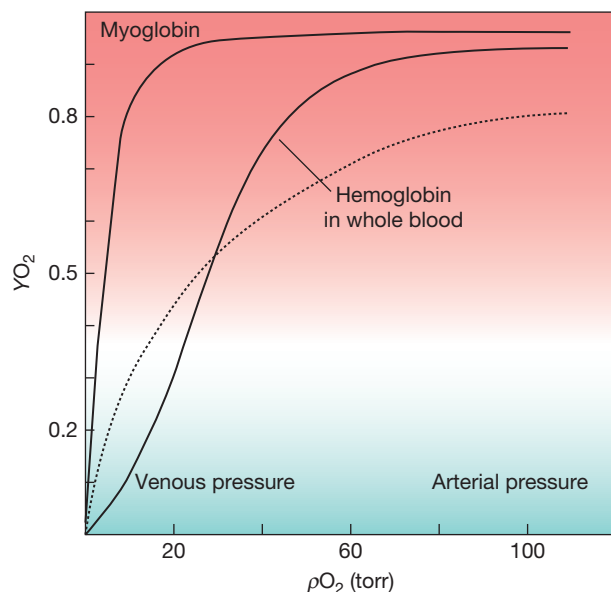


Figure 1 The O_2 -binding curves for Hb (cooperative) and Mb (noncooperative). The dashed line represents a noncooperative curve for O_2 binding having the same value as Hb for the O_2 pressure (pO_2) required for half-saturation ($YO_2 = 0.5$). Modified from Voet D, Voet JG, and Pratt CW (2002) *Fundamentals of Biochemistry*. New York, NY: Wiley.

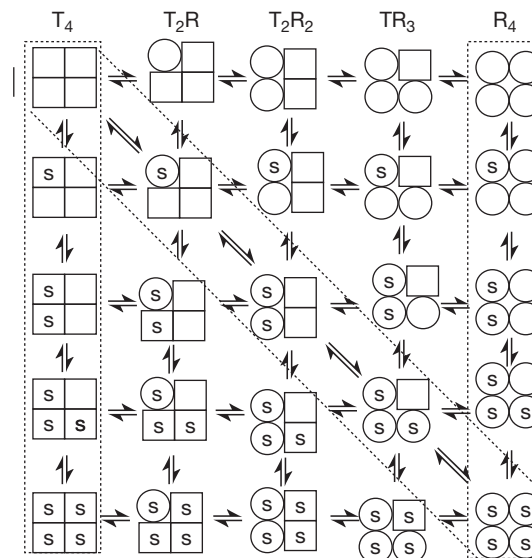


Figure 2 The general two-state model for substrate binding to an allosteric tetrameric protein. The columns on the left and right show the species included in the MWC model. The diagonal from upper left to lower right shows the species included in the KNF model.

conformational equilibrium constant (L' or K'_l) in each model, with inhibitors, which bind T state preferentially, increasing L' or decreasing K'_l and activators, which bind R state preferentially, decreasing L' or increasing K'_l .

Because of their relative simplicities, both the MWC and KNF models provide useful frameworks for the analysis of allosteric proteins, although, in detail, each enzyme studied may show deviations. In some cases, one model may be preferred over the other. Thus, the MWC model does not account for negative homotropic effects ($n_H < 1$), whereas such effects can arise in the KNF model when K_{RT} is substantially greater than either K_{TT} or K_{RR} . On the other hand, the MWC predicts that the state function R , which measures the fraction of all subunits in the R state, will, in general, increase more rapidly as a function of substrate concentration than the saturation function Y_S , since the binding of one substrate molecule can shift all the subunits in T_n to the R_n conformation. In contrast, the KNF model demands that R is always equal to Y_S . Experimental examples have been found for both negative homotropic cooperativity and for $R \neq Y_S$. Detailed examination of structural or energetic changes on partial ligand binding has been used to differentiate between the two models.

Extensions and Modifications of Two-State Models

When ligand binding induces a change in the oligomeric state of the protein, cooperativity can result as a direct consequence of ligand binding preferentially to either the associated or dissociated form. If, for example, substrate binding favors a higher state of oligomerization, then higher enzyme concentrations will enhance activity but depress cooperativity, and, at constant enzyme concentration, allosteric ligands that favor association will be activators while those that favor dissociation will be inhibitors. The equations to account for this type of behavior are similar in form to those generated by either the MWC or KNF models, but with the addition of terms that are dependent on protein concentration.

Thus far, we have considered allosteric proteins in which the substrate and effector molecules have different affinities for the R and T states. These are the so-called K (binding constant) systems. In pure K allosteric enzymes, the T and R states have identical $V_{\max}$ values. However, allosteric enzymes can also be regulated by V (velocity) systems. In pure V allosteric enzymes, the R and T states have identical affinity for substrate but different $V_{\max}$ values. In both K and V enzymes, an effector functions by binding preferentially to the R or T state. For K systems, such binding results in altered affinity and cooperativity of substrate binding. In V systems, effector binding results in altered $V_{\max}$ values, and there is no cooperativity in substrate binding.

Structures of Allosteric Proteins

High-resolution structures have been determined for a number of allosteric enzymes in both the R and T states, permitting some generalizations to be made and allowing critical consideration of the MWC and KNF models. One common finding is that ligand-binding sites, both active site and regulatory, are located at subunit interfaces. Such placement provides an exquisite

means of communicating cooperative and/or allosteric effects between subunits in an oligomeric enzyme, since sites at the interface are likely to respond to the changes in subunit interactions that are critical for allosteric regulation. For similar reasons, a second common location for ligand binding is at the interface of two domains within the same subunit. Placement of the active site at either interface has as a consequence that even modest conformational changes can cause significant changes in the size and shape of the active site, thus altering the protein's activity. Allosteric binding sites also tend to be located in low-stability regions of the protein. The binding of an effector stabilizes both the region itself and specific contacts the region makes with adjacent regions, requiring the movements of polypeptide backbone and side chains, and, in particular, the alteration of salt bridges and hydrogen bonds at subunit interfaces. These movements provide a mechanism for transmission of signals to the active site and to other allosteric sites over distances of tens of angstroms. Other common features are: (1) small but critical movements at the active site; (2) the rotation of subunits with respect to one another around a cyclic axis; (3) only two modes of subunit:subunit docking, consistent with the MWC model of concerted transition; and (4) a more highly constrained and extensive subunit interface in the less active T -state than in the R -state.

Examples of Allosteric Proteins other than Hemoglobin

Aspartate Transcarbamoylase

Aspartate transcarbamoylase (ATCase) catalyzes a key step of pyrimidine biosynthesis, the condensation of carbamoyl phosphate with aspartate to form N -carbamoylaspartate. The *Escherichia coli* enzyme has been extensively studied. Cytidine triphosphate (CTP) is an allosteric inhibitor representing a classic case of feedback inhibition whereby the end product of a biosynthetic pathway inhibits an enzyme catalyzing a reaction at the beginning of the pathway. Adenosine triphosphate (ATP) is an allosteric activator, and together CTP and ATP act on ATCase to coordinate the rates of purine and pyrimidine nucleotide biosynthesis. The enzyme has the subunit composition c_6r_6 , where c and r are catalytic and regulatory subunits, respectively. The c subunits are arranged as two c_3 's, which are complexed with three r_2 's. In the absence of r subunits, the c subunits are catalytically active, and are unaffected by ATP or CTP, which bind only to the r subunit. Crystal structures have shown that the c_6r_6 holoenzyme exists in two conformations, with CTP preferentially binding to the inactive T state and ATP to the active R state. Interestingly, ATP and CTP bind competitively to the same allosteric site in the r subunit. CTP binding induces a contraction in r_2 , which in turn, via interactions at the r - c interface, results in the c_3 's moving together by 0.5 Å. This movement causes a perturbation of the active site and a loss in catalytic activity. In contrast, ATP binding results in the c_3 's moving apart by 0.4 Å.

Ribonucleotide Reductase

Ribonucleotide reductases (RRs) form a family of allosterically regulated enzymes that catalyze the conversion of ribonucleotides to deoxyribonucleotides and are essential for *de novo* DNA

biosynthesis. Their allosteric regulation is designed to match the flux of the four deoxynucleotides produced with the base composition of the organism's DNA. Class Ia RRs are the most widespread in nature. They accept the four common nucleoside diphosphates (NDPs) as substrates, with enzymatic activity dependent upon the formation of a heterocomplex between subunit R1, which contains the active site and three allosteric sites (the s-, a-, and h-sites), and subunit R2, which contains a stable tyrosyl-free radical that is necessary for NDP reduction at the active site. The substrate specificity of RR is determined by the allosteric ligand occupying the s-site. ATP and deoxy adenosine triphosphate (dATP) stimulate the reduction of CDP and uridine diphosphate (UDP), deoxythymidine triphosphate (dTTP) stimulates the reduction of guanosine diphosphate (GDP), and dGTP stimulates the reduction of adenosine diphosphate (ADP). The s-site is located at the interface between R1 monomers, such that effector binding drives formation of R1₂. The regulation of total enzyme activity is controlled by ATP and dATP, chiefly at the level of changes in oligomerization state. Both bind to the a-site, located at the interface between two R1₂'s, driving formation of R1₄, which exists in two conformations, R1_{4a} and R1_{4b}, with the latter predominating at equilibrium, while only ATP binds to the h-site, which drives formation of R1₆. Only the R2₂ complexes of R1₂, R1_{4a}, and R1₆ are enzymatically active. dATP is a universal inhibitor of RR activity due to its induction of R1_{4b} formation, whereas ATP is a universal activator because it induces R1₆ formation. Class II RRs, which are only found in bacteria, are considerably simpler, comprising only a monomeric protein containing an active site, an s-site, and a B₁₂ cofactor. Interestingly, the active site and s-site have the same relative orientations as in class Ia, with the s-site located at an interdomain interface.

GroEL

E. coli GroEL is an example of a chaperonin, which mediates protein folding in an MgATP-dependent manner. GroEL consists of 14 identical subunits that form two heptameric rings. Cooperativity in ATP binding and hydrolysis by chaperonins reflects the switching of rings between protein-binding and protein-release states and is important for regulation of their reaction cycle. GroEL displays two levels of allostery: one within each ring and the second between the two rings. In the first level of allostery, each heptameric ring is in equilibrium between *T* and *R* states that interconvert in a concerted manner, in accordance with the MWC model. In the absence of ligands, GroEL is predominantly in the *T*₇*T*₇ state. In the presence of low concentrations of ATP (<100 μM), the equilibrium is shifted toward the *T*₇*R*₇ state, displaying positive cooperativity in ATP binding and hydrolysis. A further shift in the equilibrium from the *T*₇*R*₇ state to the *R*₇*R*₇ state ($L2 = [RR]/[TR]$) takes place only at higher ATP concentrations because of inter-ring negative cooperativity. This second level of allostery is better described by the KNF model.

Cell-Surface Receptors Involved in Signal Transduction

Membrane receptors for transmitters, peptides, and pharmacological agents are central to signal transduction. They selectively recognize chemical effectors (neuronal or hormonal) and allosterically transduce binding recognition into biological action through the activation of ligand-gated ion channels (LGICs) and/or G-protein-coupled receptors (GPCRs). Various features of membrane receptors for neurotransmitters are well accommodated by the MWC model. These receptor proteins are typically heterooligomers and exhibit transmembrane polarity. In general, the regulatory site to which the neurotransmitter binds is exposed to the synaptic side of the membrane while the biologically active site is either a transmembrane ion channel, a G-protein-binding site, or a kinase-catalytic site facing the cytoplasm. Interactions between the two classes of sites are mediated by a transmembrane allosteric transition. Signal transduction or activation is mediated by a concerted cooperative transition between a silent resting state and an active state. Agonists stabilize the active state and competitive antagonists the silent state, and partial agonists may bind nonexclusively to both. These receptors can also undergo a cascade of slower, discrete allosteric transitions, which include refractory regulatory states that result in the desensitization or potentiation of the physiological response.

See also: **Bioenergetics:** Ligand-Operated Membrane Channels: Calcium (Glutamate); Ligand-Operated Membrane Channels: GABA; Pyrimidine Biosynthesis and Degradation (Catabolism); **Protein/Enzyme Structure Function and Degradation:** Chaperonins.

Further Reading

- Changeux J-P and Edelstein SJ (1998) Allosteric receptors after 30 years. *Neuron* 21: 959–980.
- Cooperman BS and Kashlan OB (2003) A comprehensive model for the allosteric regulation of class in ribonucleotide reductases. *Advances in Enzyme Regulation* 43: 167–182.
- Eaton WAE, Henry ER, Hofrichter J, and Mozzarelli A (1999) Is cooperative oxygen binding by hemoglobin really understood?. *Nature Structural Biology* 6: 351–358.
- Fersht A (1999) *Structure and Mechanism in Protein Science*. New York, NY: WH Freeman.
- Horovitz A (2001) Review: Allostery in chaperonins. *Journal of Structural Biology* 135: 104–114.
- Koshland DE, Jr., Nemethy G, and Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5: 365–385.
- Luque I and Freire E (2000) Structural stability of binding sites: Consequences for binding affinity and allosteric effects. *Proteins: Structure, Function, and Genetics* 41(supplement 4): 63–71.
- Monod J, Wyman J, and Changeux JP (1965) On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology* 12: 88–118.
- Neet KE (1995) Cooperativity in enzyme function: Equilibrium and kinetic aspects. *Methods of Enzymology* 249: 519–567.
- Voet D and Voet JG (2004) *Biochemistry*, 3rd edn., vol. 1. New York, NY: Wiley.
- Voet D, Voet JG, and Pratt CW (2002) *Fundamentals of Biochemistry*. New York, NY: Wiley.

Alternative Splicing

K Lynch and L Motta-Mena, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Alternative exon A sequence of a gene that is included in some, but not all, mature mRNAs derived from that gene.

Alternative splicing Differential inclusion or exclusion of exons in a pre-mRNA substrate, which results in two or more mature transcripts being produced from a single pre-mRNA.

Constitutive exon A sequence of a gene that is included in all final mature mRNAs.

Pre-mRNA splicing The catalytic joining of exons and corresponding removal of intervening sequences (introns) from a pre-mRNA to produce a final mRNA.

Spliceosome A macromolecular ribonucleoprotein machine that catalyzes the removal of introns and ligation of exons in a pre-mRNA molecule.

Introduction

In higher eukaryotes a single genome codes for the information required to produce many different cell types, and to arrange these cells into specific tissues and organs that will eventually make up a complete organism. Strikingly, in humans this process is accomplished with only ~25 000 protein-coding genes, which is a relatively small number if one considers that the small mustard plant *Arabidopsis thaliana* has close to ~21 000 protein-coding genes. To amplify the information content of their seemingly limited genomes, higher eukaryotes have developed several mechanisms that allow them to generate the required biological complexity. Some of these include the genetic swapping of protein domains, posttranslational protein modifications, small RNA regulation, and alternative splicing.

Alternative splicing is the differential inclusion or exclusion of exons in a precursor messenger RNA transcript. Based on the number of genes impacted (~95% of humans genes), alternative splicing is one of the most extensively used mechanisms for generating proteomic diversity and cellular complexity. In this article, we describe the predominant classes of *cis*-acting sequences and RNA-binding proteins that regulate splicing decisions, and outline the basic mechanisms used by these splicing effector proteins to determine splicing patterns.

Splicing of Pre-Messenger RNAs

Most protein-coding genes in eukaryotes are composed of coding segments (exons) that are interspersed among noncoding sequence (introns) along the DNA. Both exons and introns are transcribed to produce a pre-mRNA, which is then spliced to remove intronic sequences and join exons to create a mature mRNA molecule (Figure 1). In a pre-mRNA the exon/intron boundaries are marked by specific, conserved sequences called splice sites (ss). These sequences are specifically recognized by the splicing machinery and direct the precise removal of introns and joining together of exons to produce a functional protein-coding transcript. The 5'-ss marks the exon/intron junction at the 5'-end of the intron and includes a conserved guanosine-uridine (GU) dinucleotide (Figure 1(a)). At the other end of the intron the 3'-ss has three conserved sequence

elements. The first element, going from 5' to 3' of the intron, is the branch point sequence (BPS), which contains a reactive adenosine that participates in the catalysis of splicing. The second element is a pyrimidine-rich sequence called the polypyrimidine tract (PPT) that is 10–40 nucleotides upstream of the third element, a conserved adenosine-guanosine (AG) dinucleotide at the extreme 3'-end of the intron (Figure 1(a)).

Bona fide splice sites must be recognized and paired from within a multitude of similar sequences in a given pre-mRNA molecule. This critical process is carried out by a highly specialized, RNA-based macromolecular enzyme known as the spliceosome. It is composed of five small nuclear RNP (snRNP) complexes (U1, U2, U4/U6, and U5), all of which consist of a uridine-rich snRNA (or two in the case of U4/U6) and multiple common and complex-specific proteins. Importantly, the spliceosome is not a preformed enzyme but instead forms through the assembly of the snRNP complexes on the pre-mRNA in a step-wise pathway that involves several intermediate complexes (E-, A-, B-, and C-complex; Figure 1(b)). The final catalytically active version of the spliceosome is ~1–2 MDa in size and consists of over 100 proteins. Although spliceosome assembly is commonly described as occurring across an intron, in higher eukaryotes where introns can be as many as hundreds of thousands of nucleotides long, spliceosome components are thought to first assemble across an exon (which are no more than ~50–250 nucleotides). However, cross-exon interactions must eventually transition into cross-intron interactions in order for catalysis to occur.

Types of Alternative Splicing Patterns

Although some exons in an RNA messenger transcript are constitutively spliced, that is, they are always included in the final message, the splicing of other exons is regulated (Figure 2). The most common type of regulated splicing event is what is called a cassette exon, which is an exon that is sometimes included and sometimes excluded from the mature mRNA. In some cases, cassette exons are mutually exclusive in which case only one exon from a group of two or more variants is selected. There are also instances where alternate 5' or 3' splice sites, located in tandem, can be used to

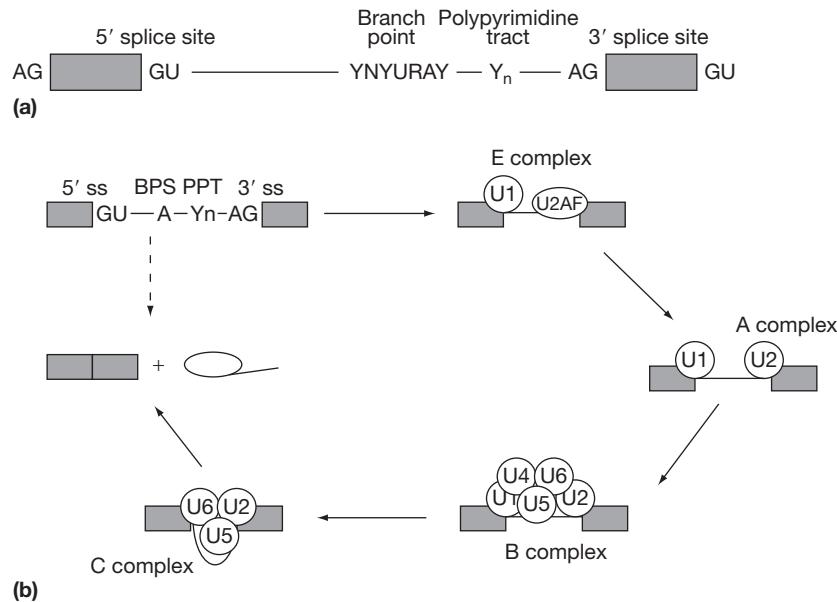


Figure 1 The spliceosome catalyzes the splicing of pre-mRNAs. (a) The exon and intron boundaries in a pre-mRNA are defined by conserved sequence elements: the 5' splice site, the BPS, a pyrimidine-rich track, and the 3' splice site. Y is a pyrimidine, R is a purine, and N is any nucleotide. (b) The stepwise assembly of the spliceosomal snRNPs (circles) during the removal of an intron from a pre-mRNA containing two exons (gray boxes) is shown. Assembly begins with the addition of the U1 snRNP to the 5'-ss and the U2AF (U2 snRNP auxiliary factor) to the BPS and PPT to yield the E complex. This is followed by the ATP-dependent binding of the U2 snRNP to the BPS, which displaces U2AF and leads to A complex formation. Subsequent recruitment of the U4/U6/U5 tri-snRNP results in the formation of a B complex. At this point, the snRNPs undergo extensive remodeling that results in the loss of the U1 and U4 snRNPs and the formation of catalytically active spliceosome (C complex).

produce different mRNA isoforms that differ in the length of a particular exon. By combining the use of alternative promoters and alternative splicing, the 5'-end of a transcript can be altered. Similarly, coordinate regulation of polyadenylation and splicing can alter the 3'-end of a transcript. Finally, in some cases the failure to remove a particular intron (i.e., intron retention event) in an mRNA can occur.

As shown in **Figure 2**, each of these altered splicing patterns has the potential to alter the coding sequence of the mRNA to give rise to a protein of distinct sequence and function. Frequently, individual protein domains are encoded by discrete exons, such that functionality can be determined by modular inclusion or exclusion of a particular domain-encoding exon. Alternatively, differential use of an exon, or inclusion of an intron may result in the introduction of a premature stop codon. In such cases, the mRNA or protein may be destabilized, such that alternative splicing becomes a mechanism to regulate the absolute level of protein expression. Finally, alternative splicing that changes the 5'- or 3'-UTR (UTR, untranslated region) of a message can determine the presence or absence of sequences that regulate message stability or translation. Thus, even changes in noncoding regions of an mRNA can ultimately control the expression or function of the resulting protein.

Auxiliary Sequence Elements and Proteins That Control Splice Site Recognition

In metazoan organisms, the majority of introns are large and contain many pseudo-splicing signals. Thus, pre-mRNA substrates usually contain additional regulatory sequence

elements that aid in the recognition and selection of legitimate splice sites and the repression of pseudo-splice sites. These *cis*-acting sequences are found in exons as well as in introns and can have positive or negative effects on spliceosome assembly. Sequence elements that activate splice-site usage and promote spliceosome assembly are called splicing enhancers. By contrast, sequences that block splice-site usage and inhibit assembly of the splicing machinery are described as splicing silencers. In a few cases, such regulatory sequences have been shown to act by inducing particular secondary structures in the pre-mRNA. However, the majority of regulatory sequences exert their effects by acting as binding sites for *trans*-acting protein factors, which in turn engage the snRNP subunits in a way that promotes or, in the case of negative regulators, inhibits assembly on neighboring splice sites (**Figure 3**).

Exonic splicing enhancers (ESEs) are often bound by members of the SR (serine-arginine) protein family. SR proteins are defined as containing one or more RNA recognition motif-type RNA-binding motifs at their N-terminus and a C-terminal domain that consists of Arg-Ser dipeptide repeats (RS domain) which interact with various components of the splicing machinery. ESEs are found within most constitutive exons in addition to many regulated exons. Therefore, SR protein binding to ESEs is essential for the selection and pairing of most genuine splice sites in both constitutive and alternative splicing. Splicing enhancers can also be located in introns (so-called intronic splicing enhancers or ISEs). ISEs do not act primarily through SR proteins, but rather serve as binding sites for a diverse array of RNA-binding proteins.

Splicing silencer sequences can similarly be located in either exons (exonic splicing silencer, ESS) or introns (intronic

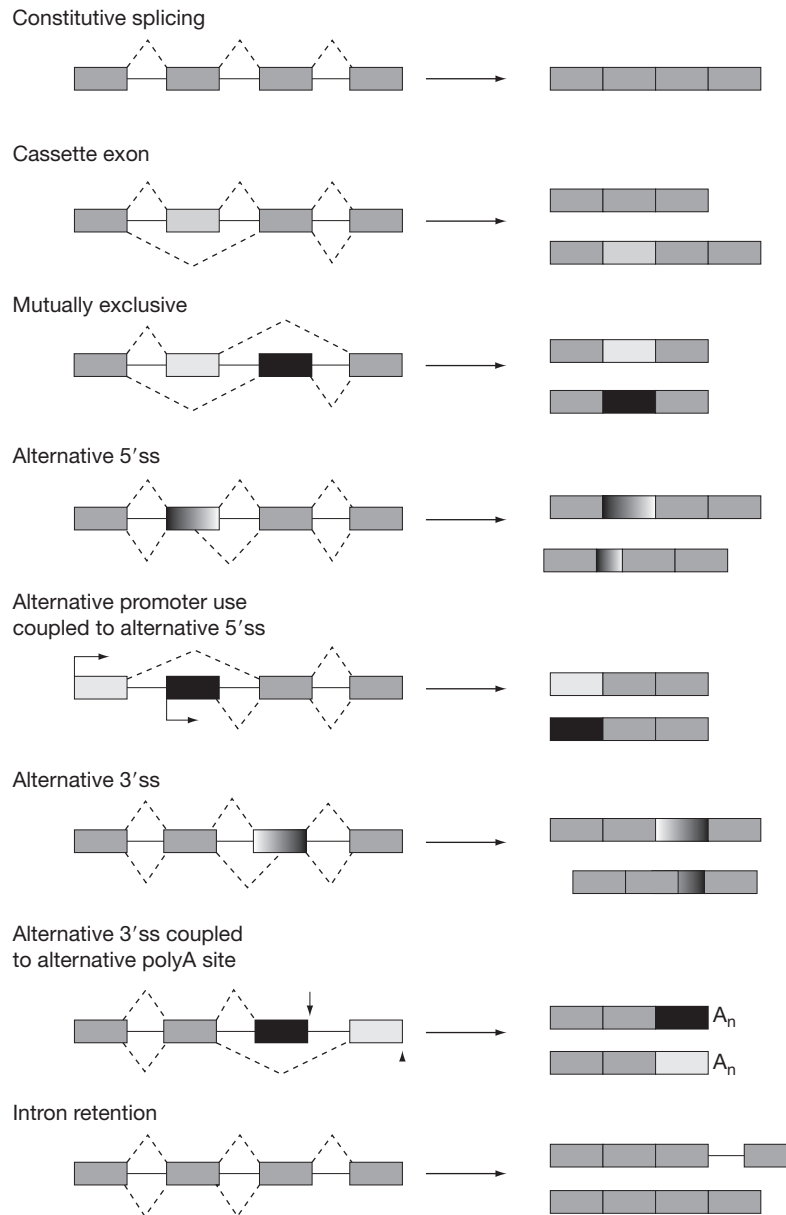


Figure 2 Patterns of pre-mRNA alternative splicing. The most common patterns of alternative splicing are shown along with alternate products. Cases in which alternative promoter sites or alternative polyA sites are coupled with alternative splicing are indicated. Exons are depicted as boxes and introns as lines; dashed lines indicate the splice sites used in the splicing reaction. Constitutive exons are shown in gray and regulated exons are shown as light gray or black boxes.

splicing silencer, ISS). These negative regulatory elements are commonly recognized by members of the hnRNP family of RNA-binding proteins, a loosely defined family of structurally diverse RNA-binding proteins that share overlapping functions in RNA splicing, packaging, export, and stability. However, despite the above generalization that hnRNPs function as repressors and SR proteins function as activators of splicing, the activity of these two families of proteins, in some cases, depends on the location (exonic or intronic) of their binding site relative to the regulated exon. Other variables – such as the affinity of a protein for its binding site, its proximity to canonical splice sites, and the strength of the regulated splice sites – can also influence the exact function of a splicing regulatory

protein. For example, it has been demonstrated that hnRNP protein L (hnRNP L) represses exons flanked by strong splice sites but enhances those flanked by weak splice sites.

Mechanisms of Alternative Splicing Regulation

Fundamentally, alternative splicing occurs through the enhancing or silencing of exons or splice sites by modulating the assembly of the spliceosome on a pre-mRNA. Ultimately, the decision to include or exclude an exon into the final mRNA is based on the combination and/or integration of both the synergistic and antagonistic forces between groups of protein

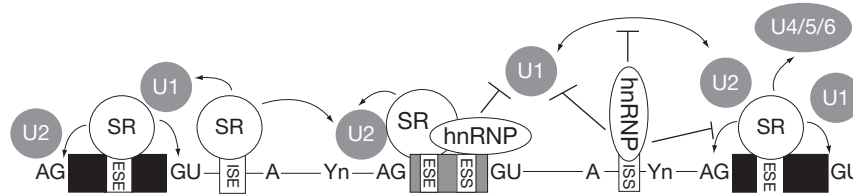


Figure 3 *cis*-Acting regulatory elements and their cognate binding proteins control pre-mRNA splicing. Exonic and intronic splicing enhancer (ESE and ISE) elements are typically bound by SR proteins. SR proteins can promote and/or stabilize the interactions of the U1 snRNP to the 5'-ss and of the U2 snRNP to the 3'-ss across the exon or the intron. Also, ESE-bound SR proteins can recruit the U4/U6/U5 tri-snRNP. Exonic and intronic splicing silencers (ESSs and ISSs) are negative regulatory elements that are commonly recognized by hnRNP proteins. hnRNPs can occlude or antagonize the binding of the spliceosomal subunits to canonical splicing signals; also, binding of hnRNPs can antagonize the binding or the enhancing activity of SR proteins. Constitutive and alternative exons are depicted as black and gray boxes, respectively. Intronic regions (lines) are designated by the splice sites, where GU and AG marks the 5'-ss and 3'-ss, respectively, A is the branch site adenosine, and Yn is the polypyrimidine tract.

regulators and between protein regulators and the spliceosomal subunits. Such combinatorial regulation of splicing makes sense because the majority of the RNA–RNA, RNA–protein, and protein–protein interactions in the spliceosome are weak (i.e., low-binding affinity). Having these types of flexible intermolecular interactions enables the spliceosome to be highly responsive to regulation. The physiological significance of this flexibility is evident in the fact that lower eukaryotes such as budding yeast have much more precision in the RNA–RNA interactions that direct assembly of the spliceosome than is observed in metazoans, and correspondingly have little, if any, alternative splicing. Therefore, the inherent suboptimal nature of spliceosome assembly in metazoans allows for the richness of regulation that generates the diversity of protein expression required of complex organisms.

Both SR proteins and hnRNPs draw on a variety of strategies to direct the selection and pairing of specific splice sites and regulate splicing patterns (Figure 3). In the simplest case, the binding of hnRNPs to the pre-mRNA sterically blocks the interaction of spliceosomal subunits to an overlapping or adjacent splice site. hnRNP proteins can also function as silencers by competitively inhibiting the binding of an SR protein to a target exon. Conversely, in the best-characterized examples of SR protein function, the RS domain interacts with protein subunits of the splicing machinery so as to recruit or stabilize the binding of the spliceosome components to nearby splice sites.

In the simplest mechanisms described above, regulatory decisions occur during the early stages of spliceosome assembly (i.e., during the initial binding of the U1 and U2 snRNPs to the pre-mRNA, Figure 1). However, regulatory proteins such as SRs or hnRNPs can also influence later steps in spliceosome assembly (after the stabilization of U1 and U2 snRNP binding). Examples of such late assembly regulation include hyperstabilizing abortive intermediates, blocking of exon pairing, inhibition or enhancement of tri-snRNP (U4/5/6) recruitment, and even altering the location of RNA cleavage during catalysis (Figure 3). Indeed, as all steps in spliceosome assembly and catalysis involve molecular interactions and rearrangements, regulation of this process can potentially occur by manipulating the efficiency of any of these processes. It is also important to remember that virtually all pre-mRNAs contain multiple exons; therefore, there is inherent competition between flanking exons or splice sites. It follows then that strengthening or weakening a particular splice site, even marginally, can completely alter the composition of the final

mature mRNA by flipping the balance of splice site competition just enough to drive assembly down a favored pathway.

Additional Determinants That Contribute to Splicing Control

Splicing regulation does not simply lead to the static generation of multiple proteins from a single gene. More often than not, alternative exons are regulated in a cell-type- and developmental-stage-dependent manner, or, in some cases, their splicing is controlled in response to specific extracellular signals and signaling cascades. This allows cells to fine-tune protein expression in accordance with cell function and/or in response to changes in the extracellular environment.

Although SR and hnRNP proteins are ubiquitously expressed, there is some variability in the level of expression of individual family members across different cell types. As the balance of SR and hnRNP function often plays a key role in the relative activity of enhancer and silencer sequence elements, deviation in the relative expression levels of these proteins in a particular cell type can alter splicing patterns. A number of tissue- or developmental stage-specific hnRNP-like splicing regulators have also been identified in recent years. These include neuronal-specific proteins (e.g., Nova-1/2 and nPTB), muscle-enriched proteins (e.g., Fox, muscleblind-like (MBNL), and CUGBP and ETR3-like factor (CELF)), and the epithelial-specific proteins (e.g., ESRP-1/2). Such proteins have been shown to control the tissue-specific splicing and expression of large families of genes that are critical to the function of their respective cell types.

The activity of splicing regulatory proteins can also be regulated by posttranscriptional modifications. Signaling pathways including the MAP kinase cascade, growth factor signaling, neuronal depolarization, and many others have been shown to result in altered phosphorylation of various splicing factors. Changing phosphorylation status of SR or hnRNP proteins has, in turn, been demonstrated to alter activity and/or intracellular localization (e.g., nuclear vs. cytoplasmic localization) of these proteins and thus to alter the splicing of their target pre-mRNAs. Many hnRNP proteins are also predicted to undergo methylation, sumoylation, and acetylation among other possible posttranslational modifications. Such additional modifications of hnRNPs could potentially also influence their activity as splicing factors, although concrete examples of such are less well documented.

Finally, it is also worth noting that splicing patterns can be influenced by additional steps in mRNA biogenesis. Most notably, the splicing of alternative exons has been shown in some instances to be regulated by interactions between splicing factors and RNA polymerase II during active transcription. Variables such as changes in promoter activity, the presence or absence of a particular transcription factor, or the rate of transcriptional elongation can thus influence the pattern in which a gene is spliced.

Alternative Splicing of the CD45 Gene

An excellent model system to illustrate many of the mechanisms of alternative splicing described above, as well as the physiologic significance of this mode of regulation, is the human CD45 gene. CD45 encodes a tyrosine phosphatase protein that is expressed on the surface of lymphocyte cells and plays central roles in antigen receptor signal transduction and lymphocyte development and maturation. The function of the CD45 protein in these processes depends heavily on the regulation of the expression of its different splice variant mRNAs. In T cells, the three variable exons in CD45 (exons 4, 5, and 6) are partially repressed under quiescent or resting conditions, which gives rise to five different mRNA isoforms. When T cells encounter antigen and become activated, the three variable exons are hyper-repressed resulting in increased levels of the lower molecular weight isoform that lacks the three variable exons. Importantly, this switch in isoform expression upon cellular activation affects the enzymatic activity of the CD45 protein. The larger isoforms, which contain the

region encoded by the variable exons, exist predominantly as monomers and are catalytically active. In comparison, the smaller isoform, which lacks the variable exons, dimerizes more efficiently resulting in steric inhibition of the catalytic site.

The sequences and proteins that control CD45 splicing have been well studied and exemplify many of the concepts described in the sections above (Figure 4). The skipping of the three CD45 variable exons is mediated by a conserved ESS element that is located within each of the exons. Mutation or deletion of the ESS from each of the regulated exons abolishes exon skipping under both resting and stimulated conditions. For exons 4 and 6 the ESS motif is contiguous. By contrast, in exon 5 the ESS sequence is bisected by an ESE element (Figure 4(a)).

The protein hnRNP L is the primary ESS-binding protein in resting T cells and mediates the partial repression of each of the variable exons which is observed under resting conditions. Interestingly, though bound to a similar sequence in each exon, the mechanism by which hnRNP L functions to repress at least two of the CD45 exons is distinct (Figure 4(a)). The 3' splice site flanking CD45 exon 5 is particularly weak (poor match to consensus); thus, inclusion of this exon is completely dependent on the presence of the SR protein SF2/ASF bound to the ESE, which promotes association of the U2 snRNP. In exon 5, hnRNP L causes skipping by binding to the ESS in such a way that blocks the association of SF2/ASF with the intervening ESE, thereby preventing SR-dependent recruitment of the U2 snRNP. By contrast, in CD45 exon 4 the binding of hnRNP L to the ESS hyperstabilizes the already-efficient binding of the U1 and U2 snRNPs to the splice sites flanking the exon in such a way to inhibit their ability to participate in the cross-intron interactions necessary for catalysis.

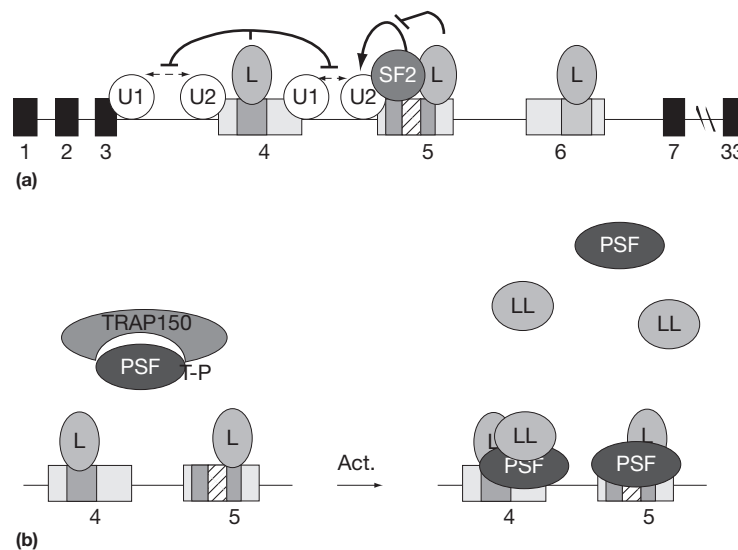


Figure 4 Signal-induced alternative splicing of the CD45 variable exons. (a) In resting T cells, skipping of the three variable exons in CD45 (exons 4, 5, and 6) is mediated by the binding of hnRNP L to an exonic splicing silencer (ESS, gray box) embedded within each of the variable exons. hnRNP L represses exon 4 by hyperstabilizing the binding of the U1 and U2 snRNPs across the exon, thus preventing the cross-intron interactions necessary for catalysis. On exon 5, hnRNP L directly competes with SF2/ASF bound to an enhancer element (ESE, hatched box), inhibiting the ability of SF2/ASF to recruit the U2 snRNP to the upstream 3'-ss. Given the similar arrangement of the ESS in exons 4 and 6, we predict that mechanism of hnRNP L on exon 6 will likely mirror that of exon 4; however, this has not yet been confirmed experimentally. Constitutive CD45 exons are shown as black boxes. (b) Under resting conditions, PSF is phosphorylated at Thr 687 (T-P) and interacts with TRAP150. The interaction of PSF with TRAP150 prevents PSF from binding to the ESS in CD45 exons 4 and 5. Upon T-cell activation (Act.), the phosphorylation of PSF is reduced and it is now able to bind and further repress exons 4 and 5. Furthermore, expression of hnRNP LL is increased upon activation and participates in the repression of exon 4.

In response to T-cell activation, two proteins, hnRNP L-like (hnRNP LL) and the hnRNP-related protein PTB-associated splicing factor (PSF), are recruited to the ESS in CD45 exon 4 and function in combination with hnRNP L to achieve maximal exon repression. PSF also participates together with hnRNP L in the stimulation-induced exon silencing of CD45 exon 5; however, hnRNP LL has little to no effect on the regulation of CD45 exon 5 (**Figure 4(b)**). This differential protein involvement upon activation underscores the fact that small sequence variations between seemingly homologous regulatory elements can have profound consequences on the proteins through which they function. Therefore, sequence-based predictions of regulatory factors and mechanism should be viewed with caution.

The differential recruitment of hnRNP LL to exon 4 under resting and activated conditions is due, at least in part, to an increase in hnRNP LL protein levels upon stimulation, consistent with existence of cell-fate-specific regulatory factors as described above. Interestingly, however, the expression of PSF remains unchanged between resting and stimulated conditions. Instead, the recruitment of PSF to the CD45 pre-mRNA is controlled by the phosphorylation at a specific site in the protein. In resting cells, Thr 687 is phosphorylated which enhances the interaction of PSF with another protein called TRAP150. The interaction of TRAP150 with PSF prevents PSF from binding to the CD45 pre-mRNA (**Figure 4(b)**). Upon T-cell activation, the Thr687 phosphorylation of PSF is reduced and so the interaction of PSF with TRAP150 is also reduced. Unphosphorylated PSF is thus free to bind to the splicing regulatory elements in the CD45 pre-mRNA to repress variable exon inclusion. It should be noted, however, that precisely how the recruitment of PSF to exon 5, or PSF and hnRNP LL to exon 4, alters the mechanism of repression upon cellular stimulation is not entirely known.

The physiologic significance of CD45 alternative splicing is underscored by the fact that naturally occurring mutations in the human CD45 gene which disrupt splicing sequences have been implicated in the susceptibility to autoimmune diseases

and viral infections. One polymorphism in particular that disrupts the activity of ESS in exon 4 and results in increased inclusion of this exon is associated with susceptibility to the autoimmune disorder multiple sclerosis. In conclusion, CD45 is one of many examples of genes whose regulation at the level of splicing strongly determines the activity of the encoded protein and where splicing misregulation can cause disease. Therefore, understanding how alternative splicing can be controlled, and mapping the regulatory sequences and proteins that determine the splicing pattern of a particular gene, is essential for deciphering the precise expression of a protein and for predicting the consequence of genetic mutations.

See also: **Molecular Biology:** Alternative Splicing: Regulation of *Drosophila melanogaster* Sex Determination; Genome-Wide Analysis of Gene Expression; Messenger RNA Processing in Eukaryotes; mRNA Polyadenylation in Eukaryotes; Ribozyme Structural Elements: Group II Introns and the Spliceosome.

Further Reading

- Chen M and Manley JL (2009) Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches. *Nature Reviews Molecular and Cellular Biology* 10: 741–754.
- Cooper TA, Wan L, and Dreyfuss G (2009) RNA and disease. *Cell* 136: 777–793.
- Heyd F and Lynch KW (2010) Phosphorylation-dependent regulation of PSF by GSK3 controls CD45 alternative splicing. *Molecular Cell* 40: 126–137.
- Licatalosi DD and Darnell RB (2010) RNA processing and its regulation: Global insights into biological networks. *Nature Reviews Genetics* 11: 75–87.
- Long JC and Caceres JF (2009) The SR protein family of splicing factors: Master regulators of gene expression. *Biochemistry Journal* 417: 15–27.
- Martinez-Contreras R, Cloutier P, Shkreta L, et al. (2007) HnRNP proteins and splicing control. *Advances in Experimental Medicine and Biology* 623: 123–147.
- Motta-Mena LB, Heyd F, and Lynch KW (2010) Context-dependent regulatory mechanism of the splicing factor hnRNP L. *Molecular Cell* 37: 223–234.
- Nilsen TW and Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463: 457–463.
- Wahl MC, Will CL, and Luhrmann R (2009) The spliceosome: Design principles of a dynamic RNP machine. *Cell* 136: 701–718.

Alternative Splicing: Regulation of *Drosophila melanogaster* Sex Determination

J K M Penn, Harvard University, Cambridge, MA, USA

P Graham, Rutgers University, Piscataway, NJ, USA

P Schedl, Princeton University, Princeton, NJ, USA

H K Salz, Case Western Reserve University, Cleveland, OH, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Jill K.M. Penn, Patricia Graham, and Paul Schedl, volume 1, pp. 78–84, © 2004, Elsevier Inc.

Glossary

Polypyrimidine tract Pyrimidine-rich sequence near the 3' splice site.

Snf Fly homolog of the mammalian U1 and U2 snRNP proteins U1A and U2B''.

snRNP Protein–snRNA complexes.

U1-70K A protein component of the snRNP U1.

U2AF A splicing factor that recognizes the 3' polypyrimidine tract and is composed of two proteins, U2AF38 and U2AF50.

Alternative splicing provides a general mechanism for producing multiple products from a single gene that can have distinct biological activities. Alternative splicing can also be used to direct development by producing tissue-, stage-, or even sex-specific messenger RNAs (mRNAs). This is the case in *Drosophila* where the master-switch gene *Sex-lethal* (*Sxl*) uses sex-specific alternative splicing as a mechanism for remembering sexual identity, and for orchestrating sexual development. The use of alternative splicing for sex determination is not unique to the *Drosophilids*, but instead is widespread among insects. Interestingly, in some of the insect species that also use this mode of regulation for sex determination, the *transformer* gene, which plays a subordinate role to (*Sxl*) in *Drosophila* sex determination, functions as the master regulatory switch for memory and morphogenesis.

The Sex Determination Hierarchy in *Drosophila melanogaster*

Sexual Choice, Memory, and Differentiation

In *Drosophila*, the choice of sexual identity is made in early embryogenesis by the primary sex-determination signal, the X chromosome to autosome (X/A) ratio. The function of the chromosomal counting system is to set the activity state of the master regulatory switch gene *Sxl*. An X/A ratio of 1.0 specifies female identity by turning *Sxl* on, while an X/A ratio of 0.5 specifies male identity by keeping *Sxl* off (Figure 1). More recent studies on the transcriptional regulatory steps involved in the initial activation of *Sxl* argue that the process is more akin to an X chromosome counting system than a system that measures the X/A ratio. Once *Sxl* is activated, it maintains female identity through a positive autoregulatory feedback loop in which *Sxl* proteins promote their own expression. In addition to providing a mechanism for remembering the female determined state, *Sxl* orchestrates female development by directing the expression of several regulatory cascades in the female mode, and by sex-specifically modulating the activity of signaling pathways like *hedgehog* and *Notch*. In 1X/2A animals, male identity is remembered by the absence of *Sxl* protein, while male development is

specified by the default expression of these same regulatory cascades. While the initial choice of sexual identity is controlled at the level of transcription, many of the key regulatory steps for both memory and differentiation depend upon sex-specific alternative splicing and/or translational regulation.

Details of the Hierarchy

Although *Sxl* is on in females and off in males, it is transcribed in both sexes from a promoter, *Sxl-Pm*, which is active from the cellular blastoderm stage onward. The critical difference between the sexes is that all *Sxl* mRNAs in males have an additional exon, exon 3, which is spliced out in females (Figure 2(a)). This male-specific exon has multiple in-frame stop codons that prematurely truncate the *Sxl* open reading frame which begins upstream in exon 2. Consequently, male mRNAs do not encode functional proteins. By contrast, female *Sxl* mRNAs encode proteins that have two RNA recognition motif (RRM)-type RNA-binding domains. These proteins positively autoregulate their own synthesis by directing the splicing machinery to join exon 2 to exon 4, skipping the male-specific exon 3 (Figure 2(a)).

Since female splicing only occurs in the presence of *Sxl* protein, the X/A ratio must activate the *Sxl* autoregulatory feedback loop in female embryos by a mechanism that bypasses this requirement. This is accomplished by turning on a special promoter, *Sxl-Pe*, in pre-cellular blastoderm 2X/2A but not 1X/2A embryos. The *Sxl* proteins encoded by the *Sxl-Pe* mRNAs set the feedback loop in motion in 2X/2A embryos by directing the female-specific splicing of the first *Sxl-Pm* transcripts. Since *Sxl-Pe* is not activated in 1X/2A embryos, transcripts generated by *Sxl-Pm* are spliced in the default male pattern.

The primary regulatory target for *Sxl* in somatic sexual differentiation is *transformer* (*tra*). The second exon of the *tra* pre-mRNA has two 3' splice sites, the upstream default splice site and the downstream female-specific site (Figure 2(b)). *Sxl* turns *tra* on in females by promoting splicing to the downstream 3' splice site of exon 2. This generates mRNAs encoding functional proteins. When *Sxl* is absent, as in males, the default 3' splice site is used. Because there is an in-frame stop codon

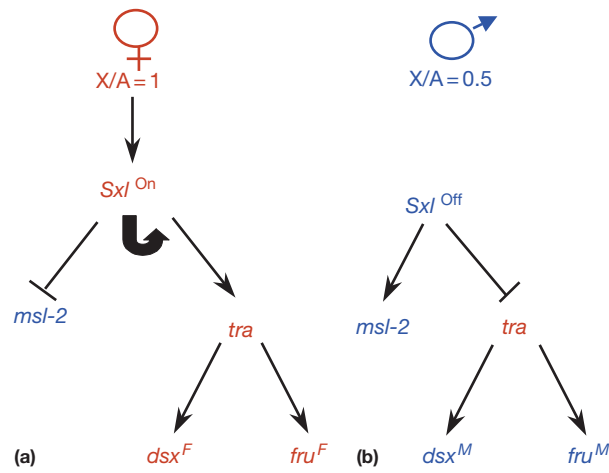


Figure 1 Regulatory hierarchy in *Drosophila* sex determination: (a) female and (b) male.

just downstream of the default 3' splice site, male *tra* mRNAs encode truncated nonfunctional polypeptides.

Like *Sxl*, *tra* promotes female differentiation by regulating splicing. The two known targets, *double-sex* (*dsx*) and *fruitless* (*fru*), encode transcription factors. In the case of *dsx*, *tra* controls the production of sex-specific isoforms that have a common N-terminal DNA-binding domain, but different C-terminal activation domains. In females *Tra* promotes the joining of the 5' splice site of exon 3 to the 3' splice site of exon 4 (**Figure 2(d)**). In males, where *Tra* is absent, splicing is to the default 3' splice site of exon 5. While *Tra* regulates *dsx* splicing by controlling the use of alternative 3' splice sites, it regulates *fru* splicing by controlling alternative 5' splice sites. As illustrated in **Figure 2(e)**, the default upstream 5' splice site in exon 2 is used in males while the splicing machinery skips this 5' splice site in females using instead a 5' splice site located 1500 bp further downstream. Although both male and female mRNAs are predicted to encode proteins, the female-specific mRNA does not appear to be translated.

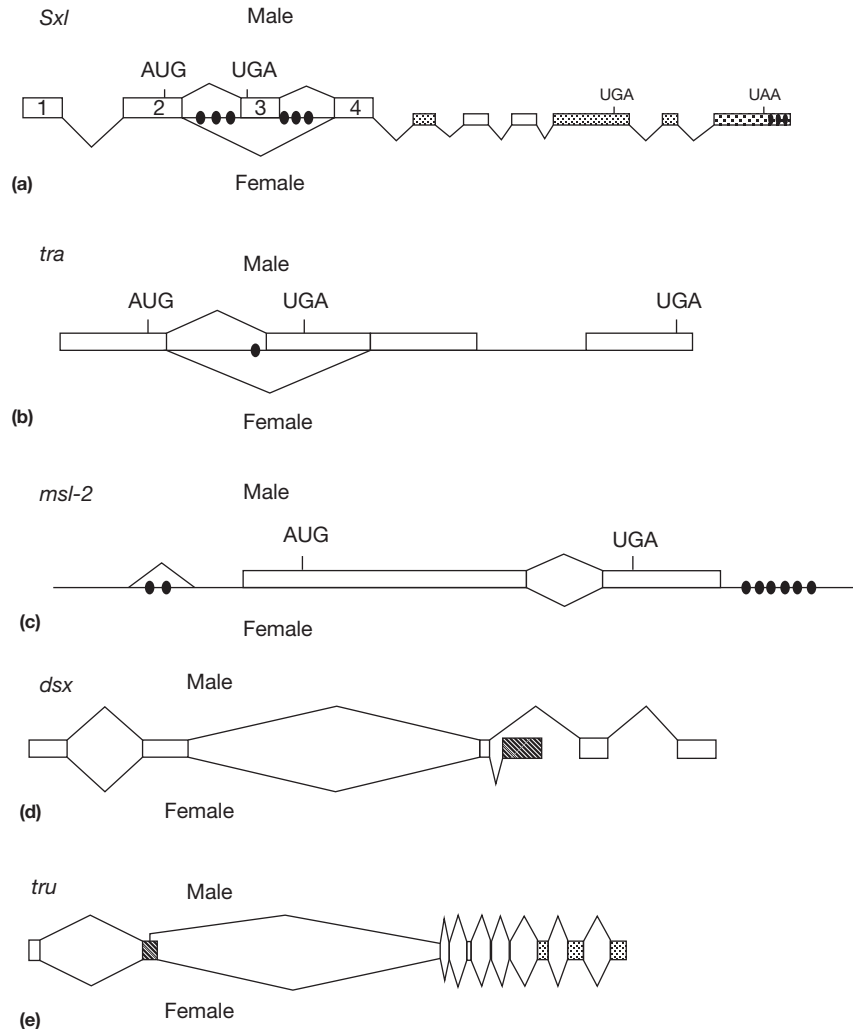


Figure 2 Exon organization and splicing patterns of alternatively spliced mRNAs encoded by *Drosophila* sex-determination genes: (a) *Sxl*; (b) *tra*; (c) *msl-2*; (d) *dsx*; and (e) *fru*. Female-specific splicing of *Sxl*, *tra*, and *msl-2* pre-mRNAs is regulated by *Sxl*, while female-specific splicing of *dsx* and *fru* is regulated by *Tra* and *Tra-2*.

Sxl also has *tra*-independent functions in female development. One of these is to downregulate *Notch*. This is thought to occur at the level of translation through *Sxl* binding to recognition sites in the untranslated regions (UTRs) of the *Notch* mRNA. Another is to augment *hedgehog* signaling. *Sxl* is found associated with cytoplasmic components of the *hedgehog* pathway and is thought to increase signaling by stabilizing the Cubitus interruptus transcription factor.

In addition to controlling somatic sexual differentiation, *Sxl* regulates dosage compensation. To equalize the dose of X-linked gene products in the two sexes, X chromosome transcription is upregulated in males by the MSL dosage compensation system. *Sxl* turns off the dosage compensation system in females by negatively regulating *msl-2* by blocking alternative splicing of an intron in the 5' UTR and by inhibiting the translation (Figure 2(c)).

Alternative Splicing Regulation by *Sxl*

tra: A Simple Blockage Model?

The regulation of *tra* splicing by *Sxl* involves a choice between alternative 3' splice sites (Figure 2). In principle, *Sxl* could promote the use of the downstream female-specific 3' splice site by preventing the splicing machinery from using the upstream default 3' splice site. Alternatively, it could activate the downstream 3' splice site. Sonsnowski et al. tested these models by deleting the alternative 3' splice sites. Deletion of the default 3' splice site led to use of the downstream female-specific 3' splice site not only in females, but also in males. This finding is inconsistent with a simple activation model since in this model *Sxl* would be required to promote the use of the downstream site even when the upstream site is missing. Deletion of the downstream 3' splice site did not interfere with splicing in males, but led to an increased amount of unspliced mRNA in females. This result is consistent with a simple blocking model because when the female-specific 3' splice site is absent, *Sxl* would still inhibit the use of the default 3' splice site, leading to an increase in unspliced mRNA.

Biochemical studies using a heterologous HeLa nuclear extract suggest that *Sxl* blocks the use of the default 3' splice site by competing with the generic splicing factor U2AF. U2AF binds to the polypyrimidine tract upstream of 3' splice site and facilitates spliceosome assembly by recruiting U2 small nuclear ribonucleoprotein (snRNP). Although mammalian U2AF recognizes both of the *tra* polypyrimidine tracts, it binds much more avidly to the default tract. The default tract is unusual in that it contains two canonical *Sxl*-binding sites, which are long poly U runs. Mutations in the *tra* poly U runs that eliminate *Sxl* binding *in vitro*, prevent *Sxl* regulation *in vivo*. Thus, the classic model for *tra* splicing is that the *Sxl* outcompetes U2AF for binding to the polypyrimidine tract of the default 3' splice site, forcing U2AF to bind to and activate the weaker, downstream 3' splice site.

Since a protein consisting of only the two *Sxl* RRM domains can bind to the default polypyrimidine tract with higher affinity than mammalian U2AF, this competition model predicts that only the RNA-binding domain of the *Sxl* should be necessary to direct female-specific splicing of *tra*. Indeed, this is the

case in *in vitro* splicing reactions. However, a *Sxl* protein that lacks the first 40 amino acids, but retains the RNA-binding domain is unable to regulate *tra* splicing *in vivo*. One explanation for this discrepancy is that *Sxl* binding of *Sxl* to the *tra* poly U runs must be stabilized *in vivo* by protein:protein interactions involving the N-terminal domain. However, this reasoning would not explain why a protein consisting of the N-terminal 100 amino acids of *Sxl* fused to β -galactosidase (N- β gal) is able to weakly promote female-specific splicing of *tra* when expressed in males. This N- β gal fusion protein does not have the two *Sxl* RNA-binding domains and consequently cannot directly compete with U2AF for binding to the polypyrimidine tract of the default 3' splice site. Taken together, these findings suggest that in addition to competing with U2AF for binding to the default polypyrimidine tract, *Sxl* may play a more active role in directing the splicing machinery to the downstream female-specific 3' splice site. As described below, *Sxl* is known to associate with a number of components of the splicing machinery *in vivo* including U2AF. Potentially such interactions could provide a mechanism for stabilizing U2AF binding to the female polypyrimidine tract and the consequent assembly of a functional spliceosome.

Sxl Autoregulation

Although *Sxl* autoregulation is also thought to involve a blockage mechanism, it is more complicated than *tra* because *Sxl* has to regulate the use of both the 3' and 5' splice sites of the male exon (Figure 2(a)). Like *tra*, the polypyrimidine tract at the 3' splice site of the *Sxl* male exon has a poly U run that is recognized by *Sxl*. However, since mutations in this poly U run have no effect on the splicing of *Sxl* pre-mRNAs in either tissue culture cells or transgenic animals, one can rule out a mechanism in which *Sxl* promotes female splicing by binding to the polypyrimidine tract of the male exon and occluding U2AF or other generic splicing factors.

In fact, splicing regulation *in vivo* does not seem to pivot on controlling the use of the male exon 3' splice site. When the male exon 5' splice site is deleted, placing the 3' splice sites of the male exon and exon 4 in direct competition, the male exon 3' splice site is skipped in *both* sexes and exon 2 is spliced exclusively to exon 4. A different result is obtained when the 5' splice sites of exon 2 and the male exon are placed in competition by deleting the male exon 3' splice site. In males, a new male exon is generated using a cryptic 3' splice site in the second intron and the normal male exon 5' splice site. Regulation of the deletion transcripts is also normal in females where *Sxl* directs the splicing of exon 2 to exon 4, skipping the cryptic male exon.

One mechanism for enhancing the role of the 5' splice site of the male exon is to weaken the male exon 3' splice site. The male exon is unusual in that it has two 3' splice site AGs, located 18 nucleotides apart, that compete with each other and reduce the efficiency of splicing. Utilization of the male exon in a heterologous context can be improved by mutating the upstream AG so that it can no longer be used. In the context of the regulated splice (exon 2-3-4), inactivating the upstream AG compromises the ability of *Sxl* to block the inclusion of the male exon presumably because the male exon 3' splice site now

functions much more efficiently. Lallena et al. have shown that the downstream AG is recognized by U2AF while the upstream AG is recognized by the splicing factor SPF45. As would be predicted if SPF45 is important for utilization of the upstream AG, reducing the activity of SPF45 in tissue culture cells or in flies substantially increases the utilization of the downstream 3' AG. The consequent improvement in efficiency of splicing of exon 2 to the male exon compromises the ability of Sxl to block the inclusion of the male exon.

How does Sxl regulate the male exon 3' and 5' splice sites? The critical targets for Sxl regulation are a series of poly U runs located about 200 bp away from the male exon in the upstream and downstream introns. Deletion of the upstream or downstream poly U runs greatly compromises but does not completely abolish regulation by Sxl, while regulation is eliminated when both are deleted. Since the critical poly U runs are distant from the regulated splice sites, it is thought Sxl exerts its effect on male exon splicing indirectly through interactions with other components of the splicing machinery. Consistent with this idea, a number of proteins have been found in complexes with Sxl, and mutations in the genes encoding these proteins show genetic interactions with Sxl. The best characterized of these is *sans-fille* (*snf*) which encodes the *Drosophila* homolog of the mammalian U1 and U2 snRNP proteins U1A and U2B'. Although recombinant Sxl and Snf interact directly with each other *in vitro*, the immunoprecipitable complexes between Snf and Sxl seen in nuclear extracts are RNase sensitive. Cline et al. proposed an snRNP-independent function for Snf in regulating Sxl splicing because they found that increasing the levels of *snf*, but not other genes encoding U1 or U2 snRNP proteins, could promote Sxl autoregulation under conditions of limiting Sxl protein. Supporting this idea, recent studies have shown that Snf can bind not only to the U1 and U2 small nuclear RNAs (snRNAs), but also to regions of the Sxl intron that have double poly U runs. However, this is probably not the whole story as genetic experiments by Nagengast et al. on mutants that specifically affect the association of Snf with the U1 and U2 snRNA show female lethal interactions with Sxl are observed for *snf* mutants defective in binding to the U1snRNP, but not the U2 snRNP. The same study also showed that Sxl can form an RNA-independent complex with the U1 snRNP protein U170K, and genetic interactions indicate that this protein:protein interaction is important for autoregulation.

The association of Sxl with two U1 snRNP proteins suggests that Sxl may prevent splicing between the male exon and exon 4 by blocking the assembly of a functional U1 pre-spliceosome complex at male exon 5' splice site. However, this would not explain how Sxl prevents exon 2 from splicing to the male exon 3' splice site(s). An understanding of the role of Sxl in this context is suggested from studies done with U2AF. In nuclear extracts, Sxl can form a complex with the U2AF proteins, U2AF50 and U2AF38, and Sxl mutations interact genetically with U2AF38. Nagengast et al. suggest that one mechanism Sxl might use to block splicing between exon 2 and the male exon would be through interactions with the U2AF heterodimer. Since U2AF recognizes the downstream AG in the male exon 3' splice site, some other mechanism might be required to block the use of the upstream AG. Since SPF45 is known to interact with U2AF, one possibility is that Sxl interaction with

U2AF at the downstream AG brings it into indirect contact with SPF45 at the upstream AG.

Genetic and molecular studies indicate that genes encoding other splicing factors play an important role in Sxl splicing regulation. One of these is *female lethal on the 2nd*, (*fl(2)d*). *fl(2)d* is homologous to the human spliceosomal *wilm's tumor associated protein* (*wtap*). Like other factors critical for Sxl splicing, *fl(2)d* shows synergistic female lethal interactions with mutations in Sxl, *snf*, and U2AF. For *snf*, *fl(2)d* showed genetic interactions not only with the *snf* mutant defective in U1 snRNP association, but also with the mutation defective in U2 association. Consistent with these genetic interactions Penn et al. found that like Snf, Fl(2)d interacts directly with Sxl protein *in vitro*, while in nuclear extracts Fl(2)d is in complexes with Sxl, Snf, U2AF50, U2AF38, and U1-70K. With the exception of Snf bound to U2 snRNP, all of these proteins are associated with spliceosomal complexes that are present at early steps in the splicing reaction prior to the first transesterification reaction. By contrast, complexes between Fl(2)d and proteins present like the U5 snRNP protein U5-40K and SKIP that are present later in the splicing reaction after the association of the U4/U6,U5 tri-snRNP with the spliceosome are not detected. Likewise, while Sxl was associated with U2AF⁵⁰, U2AF³⁸, and U1-70K, it was not found in complexes with either U5-40K or SKIP.

Taken together these findings would argue that splicing regulation by Sxl takes place at an early step in the splicing pathway, most likely prior to the first transesterification reaction. This idea is supported by studies on a Snf-interacting protein partner of Snf (PPS). As would be expected for splicing factors involved in Sxl autoregulation, mutations in *pps* shows female lethal interactions with Sxl and with other known splicing factors such as *fl(2)d* and U2AF38. In addition, PPS is found in complexes with the U1 snRNP, Sxl, and unspliced Sxl RNAs. However, PPS does not seem to be a typical splicing factor. Rather it seems to be a component of the transcriptional apparatus, perhaps functioning to coordinate transcription with alternative splicing. Rather it seems to be a component of the transcriptional apparatus, perhaps functioning to coordinate transcription with alternative splicing. Consistent with this hypothesis, recent studies have shown that the coupling between transcription and splicing of the regulated splice (exons 2–4: [Figure 2\(a\)](#)) in Sxl pre-mRNAs differs in males and females. Splicing of constitutively spliced exons typically occurs as the pre-mRNA is being transcribed and is completed before the RNA is released and polyadenylated. This is what is observed for Sxl pre-mRNA splicing in males – the regulated splice (exon2–exon3–exon4) is processed co-transcriptionally just like the constitutively spliced exons elsewhere in the Sxl pre-mRNA. In females, splicing of the constitutively spliced exons is also co-transcriptional; however, the splicing of the regulated splice (exon2–exon4: [Figure 2\(a\)](#)) is uncoupled from transcription and does not take place until after the transcript has been released from the Sxl locus. This uncoupling of transcription and splicing when pre-mRNAs are spliced in an alternative or regulated pattern rather than in the default pattern appears to be a general phenomenon. Moreover, in the case of Sxl it would explain why translation factors such as eIF4E and PABP have implicated in helping Sxl direct the female-specific splicing of Sxl and also *msl-2* pre-mRNAs.

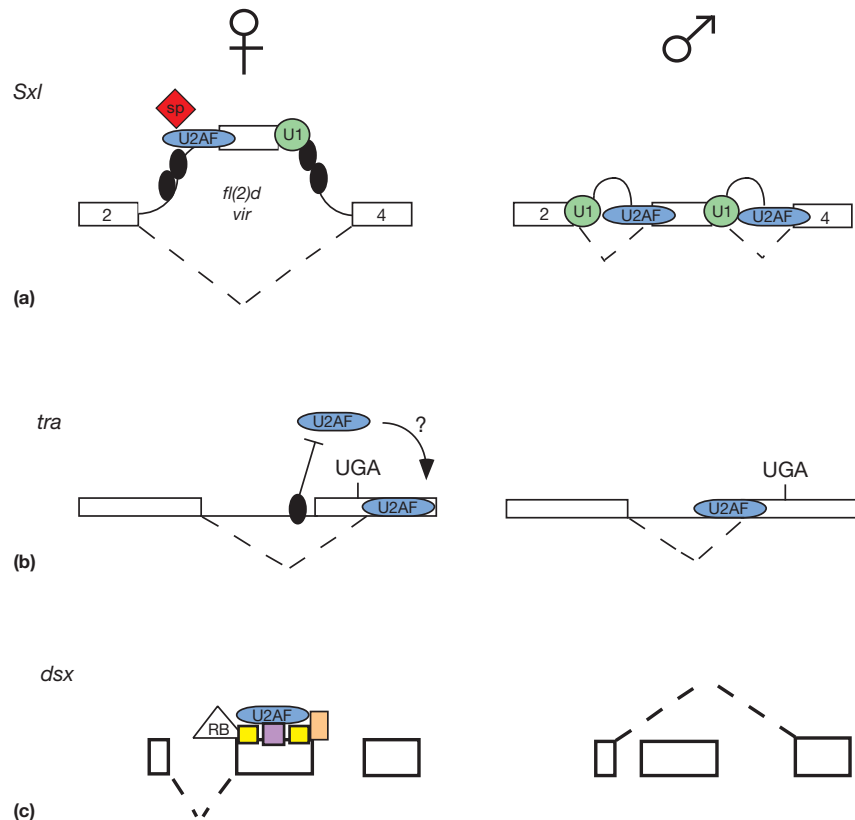


Figure 3 Splicing factors recognize splicing signals differently in females and males. (a) In this model for *Sxl* autoregulation, *Sxl* prevents inclusion of the male exon in the spliced mRNA in females by sequestering splicing complexes associated with the male exon splice sites. (b) In this model for *tra* regulation, *Sxl* blocks the association of U2AF with the default 3' splice site, forcing it to associate with the female 3' splicing site. (c) Tra (purple), Tra-2 (yellow), SR proteins (tan), and Rbp1 bind to sites in the female-specific *dsx* exon, recruiting U2AF to the female 3' splice site.

Alternative Splicing Regulation by Tra

Tra Protein

Tra is an arginine/serine domain containing splicing regulator (SR) protein. SR proteins generally function to activate splice sites, but there are examples of SR proteins that are able to block splice sites instead. In *Drosophila*, female-specific Tra functions with sex-nonspecific Tra-2 to activate splice sites in *dsx* and *fru* pre-mRNAs.

Regulation of *dsx* and *fru*

In males, the default splicing machinery joins the 5' splice site of *dsx* exon 3 to the 3' splice site of exon 5. When Tra is present in females it promotes splicing of exon 3 to exon 4 by activating the 3' splice site of exon 4. Exon 4 contains six 13-nucleotide repeat elements (REs) and one purine rich element (PRE) (located between RE #5 and RE #6) that are necessary for efficient use of the female-specific splice site. When these elements are deleted, *dsx* is spliced in the male pattern irrespective of whether Tra is present. Conversely, when these elements are placed downstream of a heterologous 3' splice site they can activate the splice site when Tra is present. *In vitro* experiments indicate that Tra and Tra-2, together with

the SR protein Rbp1 associate with the 13-nucleotide REs, while the SR proteins dSRp30 and b52/dSRp55 function at the PRE. It is thought Tra and Tra-2 plus the various SR proteins form an enhancer complex in exon 4. Since Tra and Tra-2 interact with U2AF38, this suggests a model in which the enhancer complexes assembled in exon 4 recruit U2AF38 and subsequently U2AF50 to the 3' splice site.

This model is complicated by the fact that the *fru* transcript contains three similar Tra/Tra-2 binding sites within exon 2, which are necessary for activating a female-specific 5' splice site. This is difficult to reconcile with the U2AF recruitment model because recruiting U2AF is not useful for activating a 5' splice site. Moreover, when the six REs and the PRE within exon 4 of *dsx* are replaced with the Tra/Tra-2 binding sites found in the *fru* transcript, activation of the female-specific 3' splice site still occurs. One explanation for this result is that recruitment of general splicing regulators by Tra, Tra-2, and other SR proteins might be context dependent. Which component of the splicing machinery is recruited may be determined by additional proteins specific to each transcript that bind to sites other than the REs and/or PREs.

In addition to functioning as a cofactor in female-specific *dsx* and *fru* splicing, *tra-2* is required in male testes for normal spermatogenesis. Interestingly, there are testes-specific *tra-2* mRNAs in which the splicing of one of the introns, M1, is

blocked by the Tra-2 protein. This results in the production of a nonfunctional *tra-2* mRNA. Recent studies by Qi et al. have shown that Tra-2 together with the Rbp1 protein interacts with recognition sequences in the M1 intron to block splicing. This contrasts with the activation of splicing by Tra-2 and Rbp1 seen for *dsx*. Tra-2 is normally expressed at very high levels in the testes compared to other tissues, and it appears that this negative autoregulatory activity of Tra-2 provides a mechanism for preventing further accumulation.

Alternative Splicing and Sex Determination in Other Species

Sex-lethal is produced in a sex-specific manner within the genus *Drosophila* but in every other dipteran examined outside of this genus, *Sex-lethal* is not expressed sex specifically, nor does it appear to have a role in sex determination. By contrast, the downstream genes seem to have maintained their role in sex determination. At the bottom of the *Drosophila* sex-determination hierarchy is *dsx*. This gene is widely distributed in the animal kingdom and in insects is differentially spliced in males and females. As in the *Drosophilids*, the default *dsx* splicing pattern is male in the Queensland fruit fly, the Phorid fly, and the housefly, while the female pattern is orchestrated by Tra and Tra-2 through target sequences for these factors in the female-specific exon. While the splicing of *dsx* mRNAs is also sex-specifically regulated in *Bombxy*, the default pattern is female and not male, and splicing regulation does not depend upon either Tra or Tra-2. Instead, it appears to be mediated by a KH domain protein, P-element somatic inhibitor (PSI) protein, which binds to sites in one of the female-specific exons and blocks the inclusion of the exon in male mRNAs.

In insects that have *tra* and *tra-2* but not *Sxl*, the *tra* gene seems to act as the master regulator of sex determination. This was first shown for *Ceratitidis Capitata tra*. *Cetra* is the target for initial sex-determination signal and once activated it not only directs the female splicing of *dsx*, but also maintains its own expression by promoting the female splicing of its own mRNAs. Similar findings have been reported for the housefly, *Musca*. *Mdtra* and *Mdtra2* are required for the female-specific splicing of *Mdtra* mRNAs and ectopic expression of a female *Mdtra* mRNA is sufficient to induce female splicing of message from the endogenous *Mdtra* gene. As in *Drosophila*, the male and female *Cetra* and *Mdtra2* differ by the presence or absence, respectively, of a translation termination signal in male-specific exon sequences just downstream of the translation start site.

However, the splicing patterns that generate the male- and female-specific exon sequence are complex and in the case of the housefly it is argued that male splicing is also likely to require male-specific factors and these male factors may be relevant to the mechanism of pathway choice.

Summary

Drosophila sex determination is a powerful and practical system in which to examine the mechanisms of alternative splicing and within the sex-determination hierarchy are examples of different mechanisms for regulating alternative splicing. *Sxl* autoregulation is controlled by blocking the 3' and 5' splice sites of an exon, while female-specific *tra* expression is achieved by blocking the stronger 3' splice site (Figure 3). Splice site activation is used for female-specific *dsx* and *fru* expression. However, a 3' splice site is activated in *dsx* whereas a 5' splice site is activated in *fru*. Each of the different alternative splicing events are unique, but some of the proteins regulating these splicing reactions are shared.

See also: [Molecular Biology: Alternative Splicing](#).

Further Reading

- Cline TW (1999) Functioning of the *Drosophila* integral U1/U2 protein *Snf* independent of U1 and U2 small nuclear ribonucleoprotein particles is revealed by *snf+* gene dosage effects. *Proceedings of the National Academy of Sciences of the United States of America* 96: 14451–14458.
- Cline TW and Meyer BJ (1996) *Vive la différence: Males vs females in flies vs worms*. *Annual Review of Genetics* 30: 637–702.
- Hediger M, Henggeler C, Meier N, et al. (2010) Molecular characterization of the key switch F provides a basis for understanding the rapid divergence of the sex-determining pathway in the housefly. *Genetics* 184: 155–170.
- Johnson ML, Nagengast A, and Salz HK (2010) PPS, a large multidomain protein, functions with *Sex-lethal* to regulate alternative splicing in *Drosophila*. *Plos Genetics* 6: 1–11.
- Lallena MJ, Chalmers KJ, Llamazares S, Lamond AI, and Valcarcel J (2002) Splicing regulation at the second catalytic step by *Sex-lethal* involves 3' splice site recognition by SPF45. *Cell* 109: 285–296.
- Penn KM, Graham P, Deshpande G, et al. (2008) Functioning of the *Drosophila* Wilms' tumor 1-associated protein (WTAP) homolog, FI(2)d, in *Sex-lethal*-dependent alternative splicing. *Genetics* 178: 737–748.
- Qi J, Su S, and Mattox W (2007) The doublesex splicing enhancer components Tra2 and Rbp1 also repress splicing through an intronic silencer. *Molecular Cellular Biology* 27: 699–708.
- Sosnowski BA, Belote JM, and McKeown M (1989) Sex-specific alternative splicing of RNA from the *transformer* gene results from sequence-dependent splice site blockage. *Cell* 58: 449–459.

Amine Oxidases

G Floris, University of Cagliari, Monserrato (CA), Italy

A Finazzi Agrò, University of Rome, Rome, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Amines Hydrocarbon compounds bearing an amine group. They are called primary, secondary, or tertiary when nitrogen binds two, one, or zero hydrogen atoms.

Oxidases Enzymes that oxidize substrates using molecular oxygen.

Polyamines Hydrocarbon compounds bearing both primary and secondary amino groups

(e.g., spermine and spermidine) strongly interacting with nucleic acids involved in many important cellular functions.

Primary amines Metabolically derived from amino acids by decarboxylation. Primary amines often show potent pharmacological or hormonal activity (e.g., histamine, serotonin, GABA, and noradrenaline).

Classification

The oxidative deamination of mono-, di-, and polyamines is catalyzed by a number of enzymes called 'amine oxidases' (AOs) that exhibit different patterns of substrate specificity and inhibitor sensitivity and also differ in their action mechanism. AOs are divided into two main categories, depending on the nature of the cofactor involved. Monoamine oxidases (MAOs) and polyamine oxidases (PAOs) belong to the first class and utilize flavin adenine dinucleotide (FAD) as the redox cofactor (**Figure 1(a)**). The second class is represented by enzymes having a tightly bound Cu^{II} ion and a carbonyl-type moiety identified as 6-hydroxydopa quinone (2,4,5-trihydroxyphenethylamine quinone; TPQ or TOPA) at the active site (**Figure 1(b)**).

Distribution, Reaction Mechanism, and Physiological Roles

Monoamine Oxidases

Two forms of MAOs (MAOs, amine: oxygen oxidoreductase (deaminating, flavin-containing); EC 1.4.3.4) have been described, namely MAO A and MAO B, the former made by 527 amino acid residues and the latter by 520, with an overall 70% identity. Both enzymes have a FAD covalently bound through its C8 α -position to a cysteine (Cys406 in MAO A and Cys397 in MAO B) side chain (**Figure 1(a)**). Both forms of MAOs have been crystallized from many different sources.

MAOs are ubiquitous in cells of most mammalian species, the notable exception being erythrocytes, and are tightly associated with the mitochondrial outer membrane. The distribution of MAOs has been mainly studied in brain: MAO A is found in catecholaminergic neurons, whereas MAO B is abundant in serotonergic and histaminergic neurons and glial cells. Some tissues mainly contain MAO A (human placenta and bovine thyroid), while other tissues contain predominantly MAO B (human platelets and bovine liver and kidney).

MAOs oxidize primary amino groups of arylalkyl amines to form an imine intermediate with the concomitant reduction of flavin to FADH_2 . The imine is then hydrolyzed to the

corresponding aldehyde and ammonia, and FADH_2 is oxidized back to FAD by oxygen with the formation of hydrogen peroxide. MAO A oxidizes dopamine, noradrenaline, and serotonin whereas MAO B preferentially oxidizes benzylamine and phenylethylamine. Adrenaline, kynuramine, tyramine, and tryptamine are substrates for both enzymes.

MAO A and B have been implicated in apoptosis, immunosuppression, cytotoxicity, cell growth, and proliferation. These enzymes play a protective role in the body by preventing the entry of amines at renal and intestinal level or by oxidizing them in blood. Liver MAOs are important for the control of blood pressure by regulating the amines level. For instance, MAO A can oxidize circulating serotonin thus preventing its effects on heart and vascular system. Both enzymes have important roles in the metabolism of neurotransmitters and other biogenic amines in the brain and are implicated in a large number of neurological and psychiatric disorders. As a matter of fact, MAO A inhibitors are used as antidepressants whereas MAO B inhibitors have an important role in the treatment of Parkinson's disease.

A further important role of MAOs is recently emerging, that is, their role in the demethylation of histone lysines and arginines, an essential process in the regulation of gene expression and genomic imprinting.

Polyamine Oxidases

PAOs (N^1 -acetylspermidine:oxidoreductase (deaminating); EC 1.5.3.11) have been classified as part of flavoprotein superfamily but in this case FAD is noncovalently bound to the protein. PAOs are monomeric enzymes with a molecular mass ranging from 53 kDa to 63 kDa. PAOs are found in vertebrates and plants. Enzymes with characteristics similar to those of vertebrates' enzymes are present in yeasts and amoebae, whereas bacteria and protozoans contain enzymes resembling those of plants. Plant PAOs have been isolated and characterized, particularly from the Gramineae oat, maize, barley, wheat, and rye. The enzyme from maize seedlings has been crystallized.

Much alike the catalytic mechanism of MAOs, the oxidation of polyamines can be divided into two half-reactions: (1) flavin

reduction upon polyamine oxidation followed by (2) flavin reoxidation by molecular oxygen. PAOs catalyze the oxidation of polyamines' secondary amino groups yielding different products depending on the organism considered. Mammalian

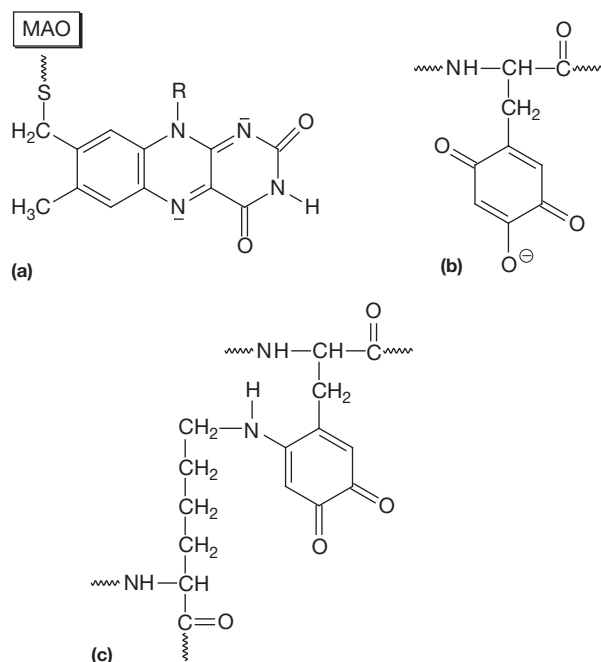


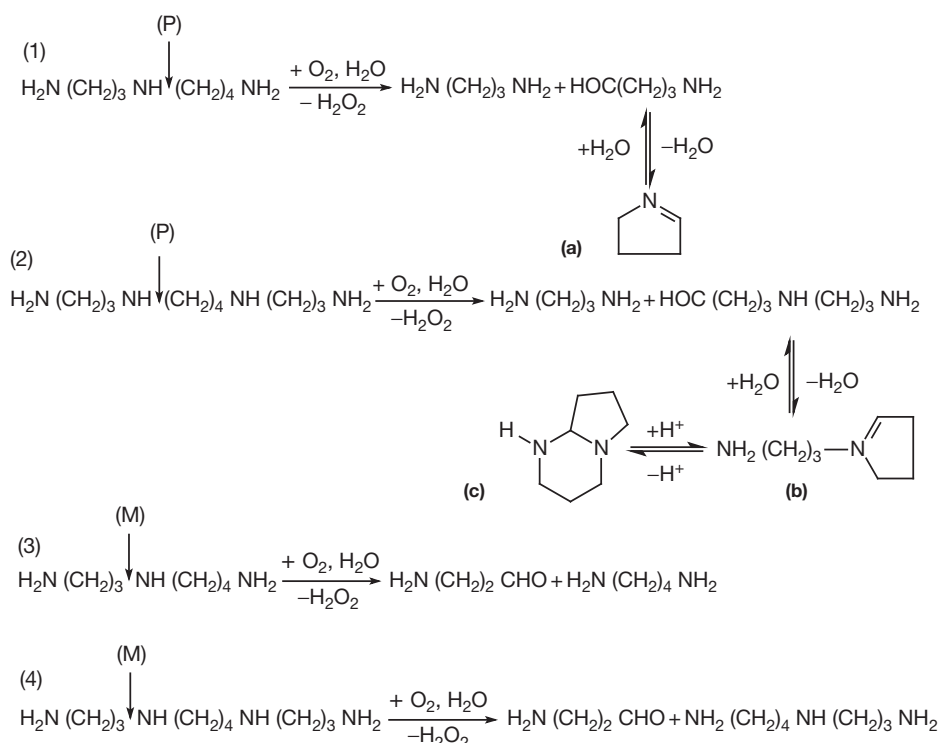
Figure 1 Structure of the (a) FAD covalently bound in MAOs; (b) the oxidized form of TPQ; and (c) lysyl tyrosylquinone.

PAOs preferentially oxidize acetyl spermine and acetyl spermidine: spermidine and putrescine are the respective reaction products, together with 3-aminopropanal and hydrogen peroxide. Oxidation of spermine by plant PAOs gives 1,3-diaminopropane, hydrogen peroxide, and 1-(3-aminopropyl)-4-aminobutanal. The latter spontaneously cyclizes to 1-(3-aminopropyl)pyrrolinium that undergoes further spontaneous rearrangement to 1,5-diazobicyclo[4.3.0.]nonane (**Scheme 1**). The oxidation of spermidine by plant PAOs gives 1,3-diaminopropane, hydrogen peroxide, and 4-aminobutanal that yields 1-pyrroline by spontaneous cyclization. At variance with MAOs and copper/topaquinone (TPQ) AOs, the oxidation of polyamines by PAOs does not release ammonia.

PAOs play an important role in the regulation of intracellular polyamines level, and seem to be important for homeostasis and cell survival.

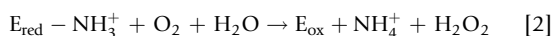
Copper Amine Oxidases

Copper amine oxidases (amine: oxygen oxidoreductase (deaminating; copper containing); EC 1.4.3.6) (CAOs) are homodimers in which each subunit (M_w 70–90 kDa) contains a tightly bound type II copper ion and a TPQ. TPQ is derived from a posttranslational modification of a tyrosyl residue, found in a strictly conserved amino acid consensus sequence asn-tyr-asg/glu (NYD/E). The pink color of CAOs stems from an intense absorption band around 500 nm due to the oxidized TPQ. CAOs catalyze the oxidative deamination of primary amino groups of mono-, di-, and polyamines, abstracting two electrons from amines and transferring them to molecular oxygen, to form the corresponding aldehyde, ammonia, and



Scheme 1 Spermidine and spermine oxidation catalyzed by plant (P) and mammalian (M) PAOs. Spermidine (1, 3); Spermine (2, 4); (a) 1-pyrroline; (b) 1-(3-aminopropyl)-pyrrolinium; and (c) 1,5-diazobicyclo[4.3.0.]nonane. Arrows indicate the cleavage of polyamines.

hydrogen peroxide. Again, the catalytic mechanism can be divided into two half-reactions: (1) enzyme reduction at the quinone moiety ($\text{TPQ} \rightarrow \text{TPQH}_2$) by substrate followed by (2) reoxidation by molecular oxygen



The group of CAOs includes the following:

1. Diamine oxidases (DAOs), ubiquitous enzymes occurring in prokaryotes as well as eukaryotes (bacteria, fungi, plants, and animals). Several DAOs have been crystallized from bacteria, from the yeast *Hansenula polymorpha*, and from pea seedlings. Plant DAOs from various species have been purified to homogeneity and characterized, the best known and studied being those from lentil (*Lens esculenta*) and pea (*Pisum sativum*) and from latex of the shrub *Euphorbia characias*. In mammals, the best-known enzymes are those from pig kidney and intestine and human placenta. DAOs are mainly active on short aliphatic diamines, such as putrescine (1,4 diaminobutane) and cadaverine (1,5 diaminopentane).

The physiological role of DAOs is multifaceted, because this class includes several enzymes with different localizations and substrates. Bacteria and yeasts can utilize amines as nitrogen and carbon sources through the reaction with AO. Plant DAOs are implicated in wound healing, in detoxification, and in cell growth by regulating the intracellular di- and polyamine levels, and the aldehyde products might have a key role in the biosynthesis of some alkaloids. The function of DAOs in mammals is even more diverse and elusive and a plethora of physiological functions, sometimes opposite, have been described. AO activity is found in many tissues, the highest levels being in decidual cells of placenta, in kidney tubular epithelial cells, and in intestinal epithelial cells. These localizations may suggest a barrier function for DAOs to prevent the entry of diamines and polyamines in the body. Furthermore, these enzymes may keep at bay the endogenous histamine, which otherwise can cause several pathological conditions such as allergy, peptic ulcer, and anaphylactic reactions. Several observations point to a relationship between AO activity and growth, both in normal and tumor tissues, that is, in cell proliferation and differentiation. DAOs have also been proposed as immune response modulators. It has been demonstrated that human placental DAO is identical to the amiloride-binding protein and thus in some way involved in the regulation of epithelial ion transport. In analogy with MAOs, also DAOs have been proposed to be involved in demethylation of histones, in particular of methylated arginines. In this case, their importance might be not only the modification of histones but also the production of hydrogen peroxide just at nuclear level.

2. The mammalian extracellular soluble and intracellular tissue-bound AOs are able to metabolize mono-, di-, and polyamines, even though *in vitro* they are preferentially active toward nonphysiological amines such as benzylamine. Extracellular AOs are generally called either plasma and serum AOs or benzylamine oxidases (BzAOs). The better-known BzAOs are those from bovine, swine, and equine plasma.

The enzyme from bovine plasma has been crystallized. Tissue-bound AOs are often indicated somewhat misleadingly as semicarbazide-sensitive amine oxidases (SSAOs), since all CAOs, at variance with FAD-dependent enzymes, are inhibited by semicarbazide and other carbonyl reagents. Intracellular SSAOs are widely distributed: smooth muscle cells of vascular tissue of all mammalian species are good sources of these enzymes, which have been also detected in uterus, ureter and vas deferens, in ox dental pulp, in human umbilical artery, in rat adipocytes, and chondrocytes.

Plasma AOs may be considered key enzymes in cell growth and differentiation processes and, thus, in tumor development and growth regulation. Moreover, the increased BzAO activity in blood serum of pregnant women supports the possibility that this enzyme has a protective role against the polyamines released from the fetoplacental unit. SSAOs seem to play an important role in vascular wall. Moreover, these enzymes might be involved in the regulation of glucose metabolism (perhaps via the H_2O_2 produced) and in the regulation of leukocyte trafficking in endothelial cells. SSAOs have also been shown to be a new class of DNA-binding proteins: in the presence of polyamines, they bind DNA and oxidize DNA-bound polyamines. A structural similarity between SSAOs and VAP-1, a protein involved in cellular adhesion, has been observed and a human SSAO/VAP-1 protein has been crystallized.

3. The mammalian extracellular matrix-bound lysyl oxidase [EC 1.4.3.13] (LOX), a 30-kDa protein, differs from other members of the group because a lysyl tyrosyl quinone (LTQ) rather than TPQ is the redox cofactor (Figure 1(c)). LTQ is formed by oxidation of a conserved tyrosine residue and subsequent intramolecular cross-linking with the ϵ -amino group of a conserved lysine. LOX catalyzes the oxidative deamination of lysyl residues in elastin and collagen allowing the formation of cross-links essential for the structure of these proteins. The genetic or acquired decrease of LOX activity is accompanied by severe pathological conditions such as tendon laxity and formation of aneurisms in arteries.

A LOX from *Pichia pastoris* has been crystallized and characterized. This enzyme, cloned and overexpressed, shows TPQ as the redox cofactor. *Pichia pastoris* LOX displays catalytic activity toward tropoelastin (the un-cross-linked form of elastin), and lysine and lysine-containing peptides are the preferred substrates for this enzyme. In this respect, it resembles the LTQ-containing mammalian AOs.

Concluding Remarks

In the past few years, evidence has accumulated about the physiological relevance of hydrogen peroxide, which is generated in the catabolism of mono-, di-, and polyamines by all AOs. Hydrogen peroxide, which is always formed in the reactions catalyzed by AOs, is more and more considered either a crucial reagent for important biochemical processes or a signal and a defense molecule, rather than a noxious waste product.

This reactive oxygen species is toxic at high concentration. However, at lower concentrations it regulates the cell number

during embryonic development as well as the proliferation and adhesive properties of endothelial and smooth muscle cells. Hydrogen peroxide appears to be involved in the impairment of cell growth and proliferation, in the regulation of many genes and transcription factors, and in the transduction of cellular signals in many living species. In plants, hydrogen peroxide might be utilized by peroxidases in basic physiological events, such as development, polymerization of lignin and suberin precursors, and catabolism of indoleacetic acid, in response to wounding or to pathogen invasion.

See also: **Bioenergetics: Quinones.**

Further Reading

- Binda C, Coda A, Angelini R, et al. (1998) Crystallization and preliminary X-ray analysis of polyamine oxidase from *Zea mays* L. *Acta Crystallographica Section D* 54: 1429–1431.
- Ciccone DN, Su H, Hevi S, et al. (2009) KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* 461: 415–419.
- Dandriofosse G (ed.) (2009) *Biological Aspects of Biogenic Amines, Polyamines and Conjugates*. Trivandrum, India: Transworld Research Network.
- De Colibus L, Li M, Binda C, et al. (2005) Three-dimensional structure of human monoamine oxidase A (MAO A): Relation to the structures of rat MAO A and human MAO B. *Proceedings of the National Academy of Sciences of the United States of America* 102: 12684–12689.
- Duff AP, Cohen AE, Ellis PJ, et al. (2003) The crystal structure of *Pichia pastoris* lysyl oxidase. *Biochemistry* 42: 15148–15157.
- Floris G and Mondovi B (eds.) (2009) *Copper Amine Oxidases*. Boca Raton, London NY: CRC Press, Taylor & Francis Group.
- Jalkanen S and Salmi M (2001) Cell surface monoamine oxidases: Enzymes in search of a function. New EMBO member's review. *The EMBO Journal* 20: 3893–3901.
- Janes SM, Mu D, Wemmer D, et al. (1990) A new redox cofactor in eukaryotic enzymes: 6-hydroxydopa at the active site of bovine serum amine oxidase. *Science* 248: 981–987.
- Mure M, Mills SA, and Klinman JP (2002) Current topics. Catalytic mechanism of the Topa quinone containing copper amine oxidases. *Biochemistry* 41: 9269–9278.
- Shi Y, Lan F, Matson C, et al. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119: 941–953.

Amino Acid Metabolism

L Cynober, Paris Descartes University, Paris, France

© 2013 Elsevier Inc. All rights reserved.

Glossary

Ammoniogenesis *De novo* ammonia synthesis from amino acids. The main precursor is glutamine and this process occurs mainly in the kidney. Ammoniogenesis plays a key role in acid–base homeostasis.

Enterocyte A cell that ensures the transport of nutrients from the gut lumen to the bloodstream and protects the internal milieu from invasion by bacteria and others.

Essential amino acid An amino acid that cannot be synthesized by humans; hence, their provision is strictly dependent upon alimentation.

Gluconeogenesis Glucose synthesis from nonglucidic precursors. Main substrates are amino acids (AAs, mainly

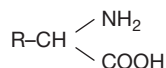
alanine, glutamine, and proline), lactate, pyruvate, and glycerol. Gluconeogenesis occurs during fasting mainly in the liver and also in the kidney.

Ketogenesis Synthesis of ketone bodies. Fatty acids are the main substrates. Some AAs are also involved, especially leucine. Ketogenesis occurs specifically in the liver.

Ureagenesis Synthesis of urea from ammonia or from ammonia derived from amino acids. This allows removal of amino acid in excess and/or the incorporation of the carbon moiety of AAs into glucose (i.e., gluconeogenesis). Ureagenesis is located in the liver.

Structure, Functions, and Classification of Amino Acids

Amino acids (AAs) are defined as organic molecules possessing an amino moiety located at α -position to a carboxylic group. AAs therefore have the following general formula:



where R may be an aliphatic or a cyclic structure.

Note that proline (PRO) is considered as an AA even though its $-\text{NH}_2$ group is part of a heterocycle. Also, taurine is considered as an AA even though it has a sulfur moiety instead of an amino group in the α -position.

AAs can be classified in three different ways (Table 1):

Chemical classification. This classification is based on structure and identifies different chemical families of AAs: aliphatic (subdivided into several subgroups – see Table 1 for details), aromatic, heterocyclic, etc.

Metabolic classification. Although in theory most AAs can be precursors of glucose, *in vivo*, because most AAs preferentially use other metabolic pathways, only alanine (ALA), glutamine (GLN), and, to a lesser extent, PRO and glycine (GLY) contribute significantly to gluconeogenesis.

Only AAs engaged in protein bonds are presented. Note that there are numerous other AAs: (1) intermediary metabolites (ornithine, citrulline (CIT)) and (2) post-transcriptionally formed (hydroxyproline, γ -carboxyGLU acid) (a) in italics: essential AAs; (b) in bold type: gluconeogenic AAs; and (c) underlined: ketogenic AAs.

Certain AAs can be precursors of ketone bodies (ketogenic AAs). Again, *in vivo*, only leucine (LEU) contributes significantly to ketogenesis.

Finally, some AAs can be both glucogenic and ketogenic: isoleucine (ILE) and phenylalanine (PHE).

Nutritional classification. This is based on what AAs the human body can or cannot synthesize. The former are

named nonessential AAs (NEAAs) and the latter essential AAs (EAAs). There are nine EAAs (Table 1). Note that some NEAAs can become EAAs in specific situations where synthesis does not match requirements (e.g., arginine (ARG) during growth, tyrosine (TYR) during renal failure, or GLN in trauma patients).

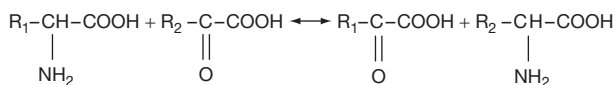
General Reaction of AA Metabolism

There is a considerable number of metabolic reactions involving AAs. The general reactions (i.d. concerning a significant number of AAs) relate to the (1) amino group(s), (2) the carboxylic group, and (3) the carbon structure.

Reactions Involving the Amino Group

Transamination reaction

This is a quite general reaction, present in all tissues. This is both a synthesis and a catabolic reaction since it is the reversible transfer of an amino group from an AA (AAR1) (becoming a keto acid) to a keto acid (becoming an AAR2):



It should be noted that, for one given AA, a transamination reaction may be a synthesis way in one organ (e.g., ALA in muscles) and a catabolic way in another (e.g., ALA in the liver).

The transamination enzymes possess a prosthetic group: pyridoxal phosphate which binds the AA during the reaction. So the reaction follows a ping-pong-type mechanism:

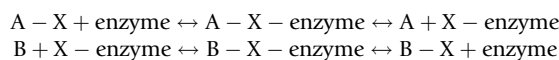


Table 1 Classifications of amino acids

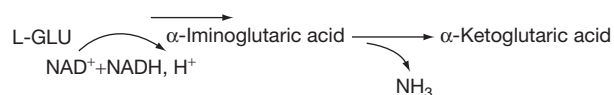
<i>Aliphatic</i>	<i>Aromatic</i>	<i>Dibasic</i>	<i>Diacid</i>	<i>Heterocyclic</i>	<i>Imino acid</i>
● Short-chain Glycine Alanine ● Alcohol Threonine Serine ● Branched-chain Leucine Isoleucine Valine ● Sulfur Methionine Cysteine	Phenylalanine Tyrosine	Arginine Histidine Lysine	Glutamate Glutamine Aspartate Asparagine	Tryptophan	Proline

where A–X and B–X are AAs, A and B are keto acids, and X is an amino group.

Oxidative deamination

The best example of oxidative deamination is the L-glutamate dehydrogenase.

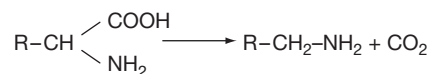
The direction in this reaction depends on the rate of use of the product of the reaction:



- degradation of GLU into the intestine (to fuel the Krebs cycle) and in the kidney (for ammoniogenesis) and
- synthesis of GLU into perivenous hepatocytes (for resynthesis of GLN).

Decarboxylation Reactions

These reactions generate amines:



the most important ones include putrescine (from ornithine (ORN)), histamine (from histidine (HIS)), and tyramine (from TYR).

Metabolism of the Carbon Skeleton

The metabolism is specific for each AA but metabolites converge toward a limited number of end products (e.g., intermediary of the Krebs cycle, glucose, and ketone bodies).

Intestinal Absorption of AAs

Digestion of proteins (beyond the scope of this article) releases a mixture of free AAs and short-chain peptides. Nitrogen is absorbed mainly in the jejunum and the ileum in the form of

free AAs and di- and tripeptides. It is now well recognized that the latter have a kinetic advantage for uptake. To date, two peptide transporters have been cloned, PEPT-1 and PEPT-2.

AAs are taken up by systems that are rather different from those found in other cells and also from those located at the basolateral side of enterocytes. The groups of transporters are relatively specific to:

- neutral AAs,
- imino acids,
- dibasic AAs + cysteine (CYS), and
- dicarboxylic AAs.

Intestinal Metabolism

It was long thought that the gut had a single function, namely, taking up AAs from the lumen for transport in the bloodstream.

In fact, the intestine avidly consumes some AAs for its energy requirements: GLU and GLN are used to the same extent as glucose by enterocytes, producing α-ketoglutarate, which is oxidized in the Krebs cycle. After a balanced meal, at least 30% of GLU + GLN is used by enterocytes, and as a consequence ammonia is released in the portal vein. Interestingly, when GLN is taken up at the luminal side of enterocytes, its uptake at the basolateral side decreases. The reverse is true in the post-absorptive state, so that the supply of GLN to enterocytes remains roughly constant.

Also, the turnover of enterocytes is very rapid, causing a strong requirement for purine and pyrimidine precursors (e.g., GLN).

Cellular Transport of AAs

Once AAs appear in the circulation, cellular transport is a critical step in AA metabolism, since it is a prerequisite for any further metabolism. In certain cases this step may even be rate limiting or rate controlling for metabolic pathways.

The systems of transport have long been classified according to their preferred substrates, their dependency toward sodium, their sensitivity to pH, and their ability to transport nonmetabolizable analogs. Based on these criteria, a large number of transport systems have been identified. The most

ubiquitous systems are system A (A for ALA-preferring), system ASC (ALA-, serine (SER-) and CYS-preferring), system L (LEU-preferring), γ + (ARG being the main substrate), N (for the transport of GLN and HIS), etc. With progress in molecular biology methods, a number of transporters have been cloned, allowing a new classification. Not surprisingly, the number of transport systems was found to be greater. For example, for ARG, four transporters have been cloned (CAT-1, CAT-2A, CAT-2B, and CAT-3) with different cell localizations, properties, and affinities for the different cationic AAs. The organs that contribute most to AA metabolism are the liver and muscle.

Liver Metabolism

The liver holds a central place in the metabolism of AAs because it is responsible for the synthesis of most circulating proteins, transforms AAs as energy sources (i.e., glucose and ketone bodies) for other tissues, and eliminates surplus nitrogen.

AAs supplied via the portal vein after a meal are heavily metabolized (50%). The carbon chain is oxidized or forms glucose, which is saved as glycogen, while the N-moiety (ies) is (are) removed in ureagenesis.

The reason for this process is that whereas the intestinal absorption of AAs is not limiting (95–99% of nitrogen is absorbed in a large range of protein intake), the brain must be protected against excessive AA exposure because several of them are neurotoxic or are the precursors of potent neuromediators (see Table 2). Thus, the liver acts as a filter, limiting the amount of AAs released in the general circulation. Notable exceptions are the branched-chain AAs (BCAAs; valine (VAL), ILE, LEU), which are almost nonmetabolized in the liver: BCAAs form 22% of AAs in food proteins but almost 50% of AAs reaching the general circulation.

The biochemical explanation for this effect is that hepatocytes do not contain BCAA transaminase (BCAA-T), which mediates the first step of BCAA metabolism. The physiological reason may be related to the fact that LEU plays a critical role in the stimulation of muscle protein synthesis in the post-prandial phase.

In the post-absorptive state, with glycogen exhaustion, gluconeogenesis increases to maintain glucose homeostasis. As stated above, ALA is the main gluconeogenic substrate among AAs: in healthy humans starved for 8 h, 30% of perfused ALA is metabolized into glucose, and, in these conditions, 11%

of glucose formed by the liver comes from ALA. When the starvation time is increased, and in situations of hyperglucagonemia (e.g., during response to injury such as burn, trauma, or sepsis), the contribution of ALA to gluconeogenesis increases like those of other AAs, in particular GLN. In these situations, most AAs taken up by the liver come from the muscles.

During starvation, another metabolic pathway is activated: ketogenesis is mainly supported by free fatty acid metabolism, but one AA, LEU, also contributes. This may seem paradoxical because, as stated above, hepatocytes do not express BCAA-T. However, hepatocytes express all the other enzymes required for ketone body (KB) production, in particular the highly regulated enzyme branched-chain ketoacid-dehydrogenase (BCKA-dh) and so can metabolize branched-chain keto acids (BCKAs) coming from muscles. Thus, as described below, the production of KB from LEU is a typical example of the importance of inter-organ exchanges in AA metabolism.

Muscle Metabolism

Muscles contain 50% of the proteins in the human body and the largest pool of free AAs (i.e., 87 g in a male weighing 70 kg as compared to the plasma pool only 1.2 g). Therefore, muscles are often considered as a reserve of AAs, either directly (i.e., as free AAs) or indirectly (i.e., in the form of proteins). This is not exact because every protein has a function (note: whereas glycogen is only a reserve of glucose and has no other function). Therefore, in catabolic states or starvation, loss of proteins is accompanied with a loss of functions.

After a meal, arterio-venous measurements indicate that all AAs are taken up by the muscles to support protein synthesis.

At the post-absorptive state, all but one (i.e., GLU) are released by the muscles. It is noteworthy that ALA+GLN together represent 60% of the total AAs released by muscle, whereas they form less than 20% of protein content. This underlies the fact that most ALA and GLN come from *de novo* synthesis in the muscle. The carbon chain required for ALA synthesis comes from anaerobic degradation of glucose, and the amino moiety comes from transamination with GLU. The resulting α -KG is retransaminated by BCAA-T releasing BCKAs from BCAAs (Figure 1). Ultimately, 18% of the glucose taken up by muscle is metabolized into ALA and 12.5% of the nitrogen contained in circulating ALA comes from LEU. For GLN, the carbon chain comes from GLU, and GLU comes both from transamination and from the bloodstream. The ammonia required for the amidation of GLU comes from oxidative deamination of AAs and from the degradation of purine bases (Figure 1).

Table 2 Some amino acid functions

Function	Amino acid	Example
Constituent of proteins	20 AAs	
Hormone precursor	Phenylalanine	Tyroxine
	Tyrosine	Catecholamines
	Tryptophan	Serotonin
Mediators	Glutamate	Glutamate, GABA
	Glutamine	Glutamine
	Arginine	Nitric oxide
Binding of calcium	Glutamate	γ -Carboxyglutamate
Collagen structure	Proline	Hydroxyproline

GABA, γ -aminobutyric acid.

Inter-Organ Exchanges

Every tissue or organ possesses enzymatic equipment that is specific both qualitatively and quantitatively. Correspondingly, each tissue plays a specific role in nitrogen homeostasis. This explains the importance of inter-organ exchanges, each organ contributing as a provider of those AAs that others are unable (or not sufficiently able) to produce. The inter-organ exchanges highly depend on the feeding state (i.e., post-prandial vs. post-absorptive).

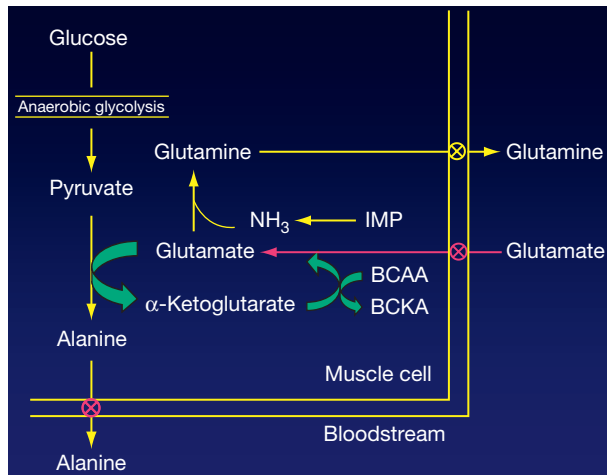


Figure 1 *De novo* synthesis of alanine and glutamine in muscles at the post-absorptive state. IMP, inosin monophosphate; BCAA, branched-chain amino acid; BCKA, branched-chain keto acid.

At the post-prandial state, as mentioned above, there is major AA splanchnic sequestration. AAs appearing in the general circulation are taken up by peripheral tissues, especially muscles.

At the post-absorptive state, the body must use its stores to generate the energy required. Lipolysis and glycogenolysis contribute to this process. However, the glycogen stores are limited and as fasting progresses, the glucose supply becomes more and more dependent on gluconeogenesis. This process implies the transfer to the liver of gluconeogenic AAs such as ALA and GLN. As described earlier in this article, the nitrogen moiety ultimately comes from BCAAs. This fact has two consequences:

1. Since BCAAs are EAAs (i.e., no possibility of synthesis in humans), their sole source is protein breakdown. Hence, gluconeogenesis in the liver results in net protein breakdown in muscle.
2. BCKAs resulting from BCAA transamination are poorly metabolized in muscles (except in severe catabolic situations where BCKA-dh is activated by pro-inflammatory cytokines) and therefore released in the bloodstream. They are then taken up by the liver, where two of them (α -ketoisocaproate and α -ketomethylvalerate) contribute to ketogenesis.

Gluconeogenic AAs taken up by the liver are readily transformed into glucose and this is favored by increased glucagon levels and decreased insulin secretion. Of course, synthesis of one molecule of glucose from one molecule of AA requires the removal of the N-residues. This is the job of ureagenesis. This process has an important consequence: each molecule of glucose produced in this pathway leads to the irreversible loss of (at least) one N-residue from net muscle proteolysis. This explains why in conditions of prolonged food restriction or when energy demand increases (e.g., in injury), a characteristic sign of metabolic adaptation is loss of muscle mass.

Glucose produced by the liver is taken up by glucose-dependent tissues. In the muscle, part of glucose is used anaerobically, leading to further generation of ALA, producing an ALA-glucose-ALA cycle also called the Cahill cycle.

At the whole-body level, this cycle makes no sense energetically because anaerobic glucose consumption generates little adenosine triphosphate (ATP) and gluconeogenesis consumes at least the same amount of ATP. However, reasoning at the tissue level casts a different light: in the post-absorptive state the liver is rich in energy owing to β -oxidation of free fatty acids from lipolysis in adipose tissue, whereas muscle is somewhat depleted in energy. Therefore, the Cahill cycle corresponds ultimately to the transfer of energy from an organ that is energy rich (the liver) to a tissue that demands energy (muscle) at the cost of net protein breakdown.

Final Products of AA Metabolism, Elimination of Surplus Nitrogen

N-metabolic products of AAs are mostly excreted in urine. Only an average 7 mg kg^{-1} body weight in males and 8 mg kg^{-1} in females is removed daily in other ways: shed skin (5 mg kg^{-1}), nasal secretions, shed hair, menstrual losses, etc.

Elimination of nitrogen cannot be carried out directly as ammonia because this substance is toxic for the central nervous system at plasma concentrations above $50 \mu\text{mol l}^{-1}$. Yet the equivalent of a mole of NH_3 has to be cleared everyday. For this reason, in humans, the main form of nitrogen elimination is urea, a water-soluble nontoxic compound synthesized by the liver.

However, not all the ammonia must be converted into urea, because ammonia plays a key role in acid-base homeostasis, in particular in the kidney.

Ureagenesis

The urea cycle is partly cytoplasmic and partly mitochondrial. Only the liver possesses all the enzymes required to synthesize urea from ammonia, and this pathway is strictly located in periportal hepatocytes.

Five enzymes are involved: carbamoylphosphate synthase (CPS), ornithine carbamoyltransferase (OCT), argininosuccinate synthase, argininosuccinate lyase, and arginase.

Ureagenesis is governed by three different types of regulations:

1. Regulation by the availability of precursors: the rate of flux of AAs toward the liver is a key regulator of ureagenesis; the more AAs the liver takes up, the higher is the rate of ureagenesis. AAs may come from food in the post-prandial phase or from muscles in the post-absorptive phase. Conversely, during prolonged starvation, urea production declines simply because the muscle efflux of AAs decreases. All the AAs are not equally ureogenic: GLN, ALA, and ARG make the most contribution.
 - GLN is hydrolyzed into GLU and ammonia by glutaminase present in large amounts in periportal hepatocytes. This reaction forms a ureagenesis amplification loop because the product of the reaction (ammonia) has the unusual property of activating glutaminase. The accumulation of GLU, the other product of the reaction, allows the synthesis of N-acetylglutamate, the regulatory role of which is described next.
 - ALA is transaminated into PYR and in parallel α -ketoglutarate gives rise to GLU. This GLU forms a

pool distinct from those described above: a second transamination reaction turns oxaloacetate into aspartate, which provides the second nitrogen donor in ureagenesis.

- ARG is very ureagenic because (1) it is the direct precursor of urea and (2) it plays a key role in the allosteric regulation of ureagenesis.
2. **Allosteric regulation: role of *N*-acetylglutamate.** This substance plays a key role in ureagenesis regulation because it is the allosteric regulator of CPS, the enzyme controlling the entry of ammonia into the cycle. *N*-acetylglutamate synthesis is catalyzed by *N*-acetylglutamate synthase, which is strongly activated by ARG. Hence, the flux of substrates (GLN, NH_3 , and ARG) and the allosteric regulation of CPS act synergistically to modulate ureagenesis both upstream and downstream.

When protein intake is low or null, the availability of precursors also plays a key role. In this situation, there is a derepression of arginase and OCT in the enterocytes and since, in adults (see below the specific situation of babies), the gut does not express ASS and ASL, this organ releases CIT instead of ARG in the portal vein. Of major importance CIT is not taken up by the liver. Hence, decrease in ARG availability for the liver slows down ureagenesis. Of note, CIT is largely extracted by the kidneys which exhibit ASS and ASL activities but not those of other urea cycle enzymes. Thus, the kidney releases ARG which is required by other tissues. Taken as a whole, this forms the ARG–CIT–ARG inter-organ cycle.

3. **Hormonal regulation.** This regulation operates at two levels:
 - On substrate availability: cortisol increases proteolysis and muscle efflux of AAs; glucagon promotes their transport into hepatocytes and further metabolism into ureagenesis and gluconeogenesis.
 - On the activity of enzymes, especially ASS and ASL.

Ammoniogenesis

This pathway is located in kidney tubular cells and is 80% supported by GLN. The first step is mediated by type I phosphodependent glutaminase, an enzyme activated by acidosis; GLU can then be transaminated into α -ketoglutarate or deaminated by GLU dehydrogenase. This last reaction is strongly activated by acidosis, further increasing the flux of NH_3 . NH_3 passes freely into the lumen, where it combines with protons to form the ammonium ion, which cannot return to cells.

The one-way flux of NH_3 means that during acidosis (i.e., high amount of H^+ in the lumen) the intra-cellular NH_3 level is low, derepressing GLU-dh. In turn, this favors GLU metabolism and low GLU derepresses glutaminase. Hence, it appears that ammoniogenesis is an adaptative pathway that plays a fundamental role in metabolic acidosis.

Relationship between Ureagenesis and Ammoniogenesis: A Contribution to Acid–Base Homeostasis

The observations described above underline the unique role of GLN as a donor of nitrogen for both ureagenesis and ammoniogenesis. However, heavy consumption of GLN in

ureagenesis is not compatible with an increased demand by the kidney in acidosis. A balance between these two almost exclusive processes is achieved, thanks to an anatomical detail – the liver contains two different hepatocyte populations:

- Perportal hepatocytes (93% of the total), which possess a glutaminase activity and enzymes of the urea cycle.
- Perivenous hepatocytes, which form only 7% of the total but have a metabolic activity 100 times higher. These cells possess GLN synthase activity. Hence, catabolism and synthesis of GLN are two processes that occur simultaneously in the liver, but at a different rate according to the situation.

Physiologically, most of the GLN from the portal vein is taken up by perportal hepatocytes, and the liver is a net consumer of GLN. In acidosis, hepatic glutaminase is inhibited (note: acidosis activates kidney glutaminase and inhibits the liver isoform). Consequently, GLN remains available for amino acid metabolism into the kidney.

Hormonal Control of Amino Acid Metabolism

Physiologically as well as in disease, hormones play a key role in the control of AA metabolism, with a balance between anabolic and catabolic hormones.

Anabolic Hormones

1. **Insulin.** Insulin exerts actions at every level of AA metabolism at the post-prandial state:
 - (a) It increases the cell transport of numerous AAs, especially in muscle and liver.
 - (b) It favors net protein anabolism by increasing protein synthesis and by decreasing protein breakdown. Insulin effect is re-enforced by LEU. Both are regulating positively the mammalian target of rapamycin (mTOR) signaling pathway. Upstream to mTOR, insulin action is dependent of AKT, whereas LEU is not.
 - (c) It decreases gluconeogenesis both by decreasing the availability of precursors and by inhibiting key enzymes of this pathway.
2. **Growth hormone (GH).** GH stimulates protein synthesis and decreases protein breakdown both at the whole body level and in muscles. However, at the basal state, the effects of GH on protein metabolism are likely of minor biological importance. Conversely, during fasting (where GH is the only anabolic hormone to increase), the anabolic effects of GH are consistent, certainly as a counter-regulatory agent in order to avoid excessive protein loss.

Catabolic Hormones

1. **Glucagon.** Like insulin, glucagon activates the A system of AA transport, but unlike insulin, glucagon favors the use of AAs in gluconeogenesis. In addition, glucagon favors proteolysis (through macro-autophagy) in the liver.
2. **Cortisol.** Cortisol induces hyper-amino-acidemia because although it increases hepatic, intestinal and renal uptake of AAs, it increases their muscle release even more strongly.

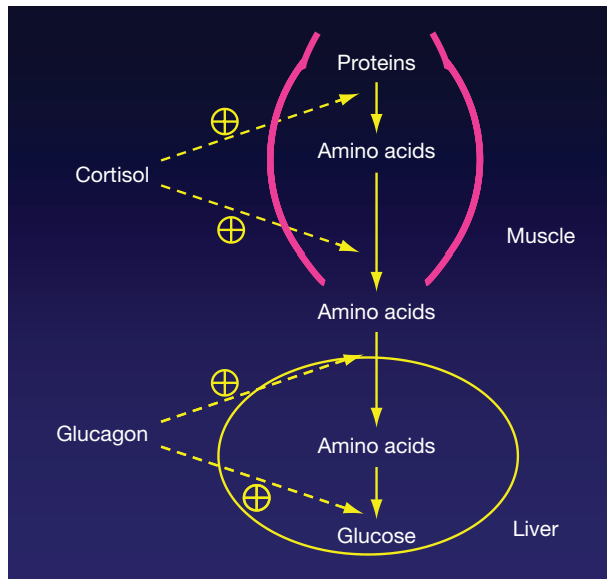


Figure 2 Cortisol and glucagon act synergistically to drive AAs from muscle to the liver.

In addition, cortisol favors net protein breakdown. Hence in a stress situation, cortisol and glucagon have a synergistic action (**Figure 2**) leading to a unidirectional flux of nitrogen from the muscle to the liver.

3. **Cytokines.** Physiologically, their role is minor except with the advance in age (see below). However, in disease, pro-inflammatory cytokines (e.g., tumor necrosis factor α , interleukins 1 and 6) are overproduced and act synergistically with glucagon and cortisol on AA metabolism.

Some Specificities of Amino Acid Metabolism along the Life Cycle

Infants. After birth, the enterocytes express ASS and ASL and therefore the gut releases ARG and not CIT. The reason is that the milk does not contain a sufficient amount of ARG to cover requirements.

Elderly. With aging there is an increased sequestration of amino acids into the splanchnic area after a protein meal and therefore a decrease in availability of amino acids at the periphery to stimulate muscle protein synthesis. In addition, the anabolic signal of insulin and LEU is blunted at the muscle level. One possible explanation is the low-grade inflammation which develops with aging: the increase in pro-inflammatory cytokine secretion may be involved in metabolic alterations described just above. This explains the occurrence of sarcopenia (a decrease in muscle mass and functions) in the elderly.

See also: **Metabolism Vitamins and Hormones:** Carnitine and β -Oxidation; Gluconeogenesis; Glycogen Metabolism; Insulin- and Glucagon-Secreting Cells of the Pancreas; **Signaling:** Steroid/Thyroid Hormone Receptors.

Further Reading

- Curis E, Nicolis I, Moinard C, et al. (2005) Almost all about citrulline in mammals. *Amino Acids* 29: 177–205.
- Cynober L (ed.) (2004) *Metabolic and Therapeutic Aspects of Amino Acids in Clinical Nutrition*, p. 746. Boca Raton, FL: CRC Press.
- Cynober LA (2002) Plasma amino acid levels with a note on membrane transport: Characteristics, regulation, and metabolic significance. *Nutrition* 18: 761–766.
- Felig P (1975) Amino acid metabolism in man. *Annual Review in Biochemistry* 44: 933–955.
- Harper AE, Miller RH, and Block KP (1984) Branched-chain amino acid metabolism. *Annual Review of Nutrition* 4: 409–454.
- Haussinger D (1990) Nitrogen metabolism in liver: Structural and functional organization and physiological relevance. *Biochemical Journal* 267: 281–290.
- Husson A, Brasse-Lagnel C, Fairand A, et al. (2003) Argininosuccinate synthetase from the urea cycle to the citrulline-NO cycle. *European Journal of Biochemistry* 270: 1887–1899.
- Meijer AJ, Lamers WH, and Chamuleau RA (1990) Nitrogen metabolism and ornithine cycle function. *Physiological Review* 70: 701–748.
- Millward DJ (1990) The hormonal control of protein turnover. *Clinical Nutrition* 9: 115–126.
- Møller N and Jørgensen JOL (2009) Effects of growth hormone on glucose, lipid and protein metabolism in human subjects. *Endocrine Reviews* 30: 152–177.

Aminopeptidases

R A Bradshaw, University of California, Irvine, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 96–98, © 2004, Elsevier Inc.

Glossary

Anti-angiogenesis The process by which blood vessel formation is inhibited. Disruption of this activity is an effective way to prevent tissue proliferation, as is encountered in tumor growth.

Carboxypeptidase One of two major classes of exopeptidases. They cleave protein and peptide substrates sequentially from the carboxyl terminal end.

Endopeptidase One of two major classes of proteolytic enzymes (the other being exopeptidases) that cleave peptide and protein substrates at internal peptide bonds.

N-End rule pathway An intracellular pathway in which selected N-termini of proteins are recognized by a specific part of the ubiquitin tagging machinery of the cell, leading to their degradation via proteosomal cleavage.

Open reading frames (ORFs) Genome sequences that can be continuously interpreted in an unbroken protein sequence. They generally, but not always, correspond to true structural genes.

Proteosome An intracellular suborganelle (or protein machine) composed of multiple subunits organized in a stack of four seven-membered rings. The subunits of the inner rings are proteolytic enzymes that degrade target proteins into short peptides, which are then either further broken down into free amino acids by exopeptidases action or (in the immune system) presented as cell-surface antigens to elicit an antibody response. When associated with additional subunits that recognize appropriately marked proteins, it is the principal entity responsible for intracellular protein turnover.

Zinc-finger domains Short amino acid sequences that contain four appropriately spaced residues capable of binding a zinc ion through their side chains. This structure, usually predictable from sequence alignments, is often involved in binding to both protein and nucleic acid partners.

General Description

Properties

Aminopeptidases occur in both soluble and membrane-bound forms and can be found in various cellular compartments as well as in the extracellular environment. The majority are metalloenzymes; that is, they minimally require a metal cofactor at the catalytic site for activity, which is usually zinc ion but can be a number of other metal ions, including Fe^{2+} , Mn^{2+} , and Co^{2+} . There are subclasses containing one and two metal ions. In some cases, the physiologically relevant metal is uncertain. There are a few aminopeptidases, particularly of the dipeptidyl- and tripeptidyl-peptidase type, that are classified as serine or cysteine proteases. Aminopeptidases can be processive, meaning that they will continue to degrade the substrate until they reach an unfavorable residue (or combination of residues), and nonprocessive. The latter are usually highly specific for an amino acid type and do not further degrade the substrate after the initial susceptible residue is removed.

Functions

The physiologic functions of aminopeptidases can be divided into three main categories: processing/maturation, activation, and degradation. The enzymes involved in the first two areas usually are highly specific, are nonprocessive, and tailor the N-terminus of proteins or peptides either co- or posttranslationally to induce activity or to allow for subsequent modifications that will, in turn, affect activity. Degradation (including

inactivation) can also use specific enzymes, particularly where individual bioactive peptides are targeted, but most often involves nonspecific enzymes that participate in protein turnover by the reduction of small peptides, arising from proteosomal cleavage of targeted proteins, to amino acids or by the extracellular degradation of peptides arising from a number of sources.

N-Terminal Cotranslational Processing

One of the best-understood functions of aminopeptidases is their role in N-terminal cotranslational processing. Protein synthesis that is directed by the genetic material is universally initiated by the amino acid methionine (in prokaryotes, N-formyl methionine), but the majority of the protein mass in organisms does not reflect this event; that is, most proteins (at least those commonly studied to date) do not have an N-terminal methionine residue. Extracellular proteins that are exported through the endoplasmic reticulum and proteins imported into mitochondria (as well as related organelles such as chloroplasts) lose their respective signal peptides (usually 20–25 amino acids), including the initiator methionine, through the action of specific signal peptidases, which are endopeptidases. The removal of the initiator methionine from intracellular proteins is accomplished through the action of a specific class of aminopeptidases (clan MG of the metallo-peptidases) designated methionine aminopeptidases (MetAPs). All living organisms apparently have at least one form of this

enzyme, and they require this activity for vitality. There are several reasons for this: (1) methionine is a relatively scarce amino acid, and failure to return a large percentage of the methionine used in the initiation process for reuse (in a variety of activities) leads to a starvation condition that is ultimately lethal; (2) many proteins are subsequently modified on the α -amino group of the newly exposed residue (but can also occur on methionine residues that are not removed) that in many cases is a requirement for their further function; and (3) a few proteins use the newly exposed penultimate residue as a part of their active structure. The reason that methionine is not removed from all N termini is apparently due to its role as a protecting group against degradation by the N-end rule pathway.

MetAP Specificity

The substrate profile of the MetAPs, despite other distinguishing physical characteristics, is highly conserved over all living organisms, suggesting that it is a very ancient enzyme. Briefly, susceptible substrates have the seven smallest amino acids (glycine, alanine, serine, threonine, cysteine, proline, and valine) in the adjacent (penultimate) position to the methionine. In yeast, a little over one-half of the open reading frames (ORFs) are predicted to code for proteins with these N-terminal sequences, but in terms of mass the percentage of soluble proteins is much higher, perhaps as high as 80%. Because essentially all exported and transmembrane proteins (which may account for as much as one-third of the total protein in a cell) and most of the proteins imported into the mitochondrion also lose their N-termini through cleavage of their signal peptides, which are in turn degraded to amino acids, there is a very high degree of recyclization of initiator methionine.

Properties of MetAPs

Structural organization

The basic catalytic domain, first defined by X-ray analyses of the *Escherichia coli* enzyme, is 30 kDa in mass and contains two metal ions at the active site. It has an internal symmetry (described as a pita bread fold) suggestive of an early gene duplication event. This structure appears, with only minor variations, to be found in all eubacteria and represents the type 1 isoform. Archaeobacteria contain a different but related (homologous) form of MetAP, principally characterized by an insertion of approximately 65 residues that forms an extra highly helical domain on the surface of the protein. It is designated type 2. Eukaryotes contain both types and, in addition, each contains an amino terminal extension. The N-terminal segment of the type 1 enzyme contains zinc-finger domains that are thought to act as a tether to tie this enzyme to the ribosome. The corresponding segment of the eukaryotic type 2 proteins is marked by extended stretches of polyacidic and basic amino acids. The function of this domain is unknown. The overall organization of the MetAP family is shown in Figure 1.

Metal use

These enzymes were initially characterized as using two Co^{2+} , based on early experiments with the eubacterial MetAPs. Subsequently, it was shown that multiple metals could be used in

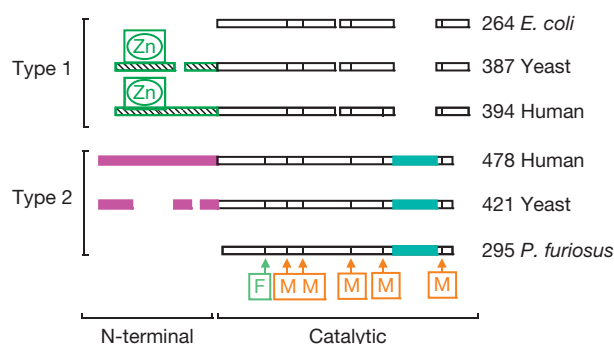


Figure 1 Schematic presentation of the structural organization of the MetAP family. Separate domains are indicated by color (magenta, N-terminal extension characteristic of MetAP2s; brown, N-terminal extension with putative Zn-binding domains characteristic of MetAP1s; light blue, catalytic domain insert characteristic of MetAP2s) and are presented approximately to scale. Major deletions (and insertions) deduced from sequence and structural comparisons are indicated as gaps. M denotes the site of a metal ligand; F indicates the location of the histidine modified in type 2 enzymes. The number of residues for each protein is given in the column at the right. Reproduced from Bradshaw RA and Yi E (2002) Methionine aminopeptidase and angiogenesis. *Essays in Biochemistry* 38: 65–78.

many of the forms, particularly the *E. coli* enzyme, which is active with Fe^{2+} , Zn^{2+} , Ni^{2+} , and Mn^{2+} , in addition to Co^{2+} . It is possible that, depending on availability, the eubacterial enzymes may use different metals. In recent studies in eukaryotes, Mn^{2+} has emerged as the preferred candidate for the type 2 enzyme, whereas Zn^{2+} remains the most likely metal cofactor for the type 1 enzymes under physiological conditions. Although early studies suggested that the MetAPs used a bivalent metal structure, they can clearly function with only a single metal ion present. In all cases, the metal ions are bound through five amino acid side chains and these are well preserved in all the isoforms in both prokaryotes and eukaryotes.

Roles of the MetAP Isoforms

The substrate specificity of all of the isoforms is generally the same as just described and is heavily predicated on the nature of the penultimate residue. Substrate length, in *in vitro* studies, does have some effect but it is unclear whether this is of significance *in vivo*. Similarly, different metal ions could also affect substrate selection, but this has not been systematically demonstrated. Null mutations (manipulations that prevent the expression of a gene) in yeast demonstrate some measure of redundancy because cells will survive with one or the other of the isoforms but not when both are eliminated. Deletion of the single gene in prokaryotes is also lethal. Nonetheless, there is both direct and indirect evidence to support the view that the two isoforms have different cellular functions. MetAP1 is thought to function as the main processing enzyme and be physically associated with the ribosome in a position to hydrolyze the methionine from germane nascent chains during protein synthesis. MetAP2 is thought to be a soluble enzyme and to probably provide secondary processing for substrates that

improperly escape the action of MetAP1. However, it is clearly involved in other activities as well. This is illustrated by a class of irreversible chemical inhibitors that are highly selective for MetAP2, which cause cell-cycle arrest in endothelial cells (but not cell death) resulting in anti-angiogenesis. These potential drugs are being refined for use in treating tumors. Presumably, this inhibition results from the failure of MetAP2 to process a select protein or subset of proteins involved in mitosis of these cells, but their nature is unknown. Downstream sequences probably are responsible for rendering these select targets susceptible to MetAP2 but not MetAP1. MetAP2 also functions to inhibit the phosphorylation of eIF2 α and thereby promote translation. This activity is entirely distinct from its catalytic one and is not connected to its cell-cycle functions because the inhibited protein is still fully functional as a phosphorylation inhibitor. It is thus one of many proteins known to have dual (and often unrelated) functions.

See also: **Bioenergetics:** Protein Import into Mitochondria; **Metabolism Vitamins and Hormones:** Amino Acid Metabolism; **Protein/Enzyme Structure Function and Degradation:** Metalloproteases; Proteasomes, Overview; Zinc Fingers.

Further Reading

- Barrett AJ, Rawlings ND, and Woessner JF (eds.) (1998) *Handbook of Proteolytic Enzymes*. London: Academic Press.
- Bradshaw RA, Hope CJ, Yi E, and Walker KW (2001) Co- and posttranslational processing: The removal of methionine. *Enzymes* 22: 387–420.
- Bradshaw RA and Yi E (2002) Methionine aminopeptidase and angiogenesis. *Essays in Biochemistry* 38: 65–78.
- Lowther WT and Matthews BW (2000) Structure and function of the methionine aminopeptidases. *Biochimica et Biophysica Acta* 1477: 157–167.
- Taylor A (ed.) (1996) *Aminopeptidases*. Austin, TX: RG Landes.

Amyloid

R Wetzel, University of Tennessee, Knoxville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 99–104, © 2004, Elsevier Inc.

Glossary

Folding motif The three-dimensional (3-D) shape pattern exhibited by a folded protein, consisting of the secondary structural units (e.g., α -helix and β -strands) accessed by different elements of the primary sequence, and how those secondary structural units interact with one another in space. Although there are tens of thousands of protein sequences coded within the human genome, it is believed that the 3-D structures of these proteins will ultimately be described by perhaps only a few hundred folding motifs.

Glial cells Non-neuronal, accessory cells in the brain that are responsible for maintaining brain structure and development and for fighting infection.

Membrane depolarization Loss of chemical potential across a membrane.

Molecular chaperones Enzymatic protein assemblies that are responsible for detecting, reversing, and, where necessary, eliminating failed products of protein folding, such as misfolded and aggregated proteins.

Nuclear magnetic resonance A type of spectroscopy in which the unique covalent and noncovalent environments

of atoms in molecules in solution can be assessed, allowing for construction of a high-resolution structural model of the 3-D relationships between atoms.

Ubiquitin–proteasome system A complex pathway of enzymes responsible for marking and destroying proteins that are no longer required by the cell due to their being obsolete or defective. Such proteins are first marked by the attachment of the small protein ubiquitin, at which point they are recognized and destroyed by the proteasome, a large protein machine consisting of proteases and other enzyme activities.

X-ray crystallography A technique for determining at high resolution the spatial relationships between atoms in a molecule in the solid state, by the detailed diffraction pattern generated when a molecular crystal is exposed to X-ray beams.

X-ray fiber diffraction An intermediate-resolution version of X-ray diffraction analysis, not requiring a crystalline form, in which repeated patterns of a structure can be recognized from a limited diffraction pattern.

Amyloid is an aggregated protein structure consisting of unbranched microscopic fibrils often found in dense tissue deposits and associated with a variety of human diseases, including a number of significant neurodegenerative disorders. In its broadest usage, the term amyloid does not pertain to a specific protein molecule or sequence, but rather to a general folding pattern, or folding motif, that appears to be accessible to many, if not all, polypeptide chains. Although the three-dimensional structures of these proteins in their native states can vary enormously, the abnormal amyloid structures that they form exhibit a characteristic folding pattern, called a cross- β structure, that differs from that of the native state structure.

Introduction

Amyloid formation thus involves a protein misfolding reaction. Amyloid fibrils are generally very stable and quite insoluble in native, aqueous buffer. In a way, amyloid is a foreign substance composed of self-proteins, but it is not easily recognized and removed as a foreign substance by the immune system, for reasons that are not well understood. In some disease states involving amyloid deposition, the amyloid deposit is directly involved in the disease mechanism, whereas in other cases its mechanistic role is less clear. In some microorganisms, specific proteins and protein domains appear to

have evolved for the purpose of making amyloid fibrils that play an important, positive role in the cell. In most cases, however, amyloid is a pathogenic structure, formed by accident under conditions of molecular, cellular, or organismic stress, from proteins that evolved to fold and function in quite different structural states. The recognition that proteins not known to be involved in amyloid disease can be induced to form amyloid fibrils, through exposure to certain nonnative conditions *in vitro*, has led to a picture of the amyloid motif as a general default structure in protein folding accessible to many, if not all, polypeptide sequences. Notwithstanding the predominant usage described above, the word 'amyloid' is also occasionally found as part of the names of specific proteins. These proteins tend to be associated in some way with the amyloid phenomenon, either as proteins capable of making amyloid fibrils (such as serum amyloid A, islet amyloid polypeptide (IAPP), and β -amyloid (also known as A β)) or as proteins that interact with amyloid fibrils (such as serum amyloid P component).

History

Waxy masses found at autopsy in the liver and spleen have been observed in humans since the seventeenth century. In 1854, Virchow reported that an iodine stain thought to be

specific for starch generated a positive test on such material and therefore classified these structures as amyloid (starch-like). Although 5 years later Friedreich and Kekule demonstrated that these deposits contained negligible carbohydrate, the name amyloid has been retained. In 1922, Bennhold introduced the use of the dye Congo red to stain tissue amyloid deposits. After the dye binds, it exhibits an apple-green birefringence when examined by polarized microscopy that sharply distinguishes amyloid from other tissue components (Figure 1). Ever since, Congo red has been the primary method by which pathologists identify amyloid deposits in tissue samples. In 1927, Divry showed that the cores of senile plaques, the cortical structures that Alzheimer linked with early-onset senile dementia in his classic 1906 paper, exhibit the Congo red birefringence characteristic of amyloid. Amyloid plaques are recognized as one of the pathological hallmarks of Alzheimer's disease (AD) and it is believed that some aggregated form of the main protein component of the amyloid fibrils in these plaques plays a direct role in AD etiology. Although it had long been recognized that the main chemical component of amyloid fibrils is protein, it was not until the pioneering work of the Glenner and Benditt groups in the early 1970s that it was realized that amyloid fibrils are predominantly composed of specific protein sequences rather than a broad mixture of protein molecules. Glenner showed that amyloid extracted from primary amyloidosis patients contained the immunoglobulin light chain molecule, and Benditt found that amyloid from tissues of chronic inflammation patients consisted of a new protein called amyloid A. Other clinically distinct forms of amyloid were soon thereafter shown to have different characteristic protein constituents. In the mid-1980s, Glenner isolated and characterized by amino acid sequencing the previously unknown peptide $A\beta$ as the principal component of AD amyloid fibrils. By the end of the twentieth century, many naturally occurring amyloid proteins had been characterized and the list continues to grow with the development of more sensitive means of analysis.

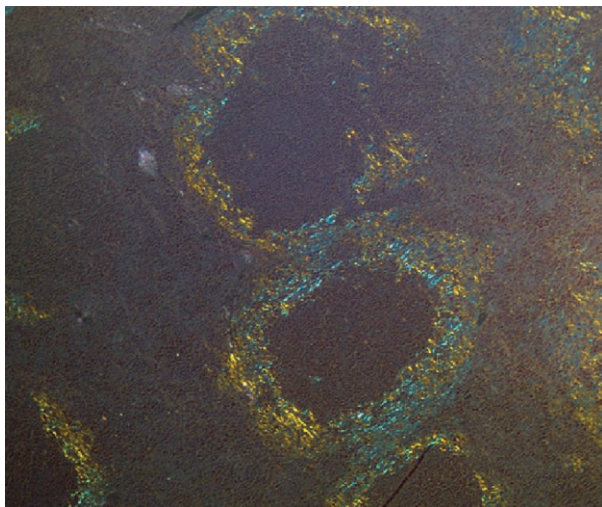


Figure 1 Rings of amyloid detected by Congo red staining and polarized microscopy in the spleen of a mouse with serum amyloid A amyloidosis. Courtesy of professor Jonathan Wall, University of Tennessee.

Amyloid Diseases

Range of Amyloid-Related Diseases

Table 1 lists some of the over 30 human disease states in which amyloid fibrils have been observed, as characterized by the principal organ involved and the principal protein component of each. Amyloid diseases can occur both in the brain and in the rest of the body. The systemic amyloid diseases tend, as a general rule, to involve large deposits of amyloid that in extreme cases dramatically increase the size of the affected organ. In these cases, the toxicity of the amyloid deposit may be largely due to the mass of deposited material and its ability to disrupt normal tissue structure and function. In some peripheral amyloid diseases, the amyloid deposits are localized to a particular tissue. Such is the case in pancreatic amyloid, which is found in over 90% of adult-onset (type 2) diabetes patients (despite this high incidence, the pathogenic role of amyloid deposition in diabetes is not yet known). By contrast, in other amyloidoses, protein deposits can be found in a variety of tissues. Amyloid is also observed in a number of brain diseases. In AD, extracellular amyloid deposits are found in the cerebral cortex, usually embedded with fragments of neuronal processes and surrounded by activated glial cells. In Huntington's disease, amyloid deposits are found in both the cytoplasm and the nucleoplasm of particular neurons in the cortex and elsewhere. In Parkinson's disease, large neuronal deposits called Lewy bodies are rich in the protein α -synuclein. In each of these brain diseases, the mechanistic role of protein deposits has not yet been established.

Table 1 Major polypeptide components of pathogenic amyloid deposits in humans

<i>Polypeptide</i>	<i>Major disease states</i>
Transthyretin	Heart, kidney, and peripheral neuropathy
Serum amyloid A	Kidney and peripheral neuropathy
Immunoglobulin light chain	Kidney and heart
Immunoglobulin heavy chain	Spleen
β_2 -Microglobulin	Carpal tunnel syndrome and osteoarthritis
Lysozyme	Non-naturopathic visceral amyloid
Islet amyloid polypeptide	Diabetic pancreatic islet cells
Fibrinogen α -chain	Kidney
Apolipoprotein A1	Peripheral neuropathy and liver
Atrial natriuretic peptide	Heart
Amyloid β -protein ($A\beta$)	Brain (Alzheimer's disease and cerebral amyloid angiopathy)
α -Synuclein	Brain (Parkinson's disease)
Huntingtin polyglutamine sequence	Brain (Huntington's disease)
Prion protein (PrP)	Brain (Creutzfeldt-Jakob disease and mad cow disease)
Cystatin C	Brain (cerebral amyloid angiopathy)
Gelsolin	Brain (cerebral amyloid angiopathy)
ABri	Brain (familial British dementia)

Physiological Factors in Amyloid Formation

Although much remains to be learned about the events that trigger amyloid deposition, it is clear that a variety of factors are involved. Some well-established factors are molecular aspects of the proteins themselves. Since the rates of aggregation reactions, such as amyloid formation, depend on concentration, dramatic increases in the amount of the amyloidogenic precursor protein can initiate amyloid formation. This is the case for the protein serum amyloid A, which increases in concentration in response to infection and, when persistently elevated, can deposit as amyloid. Similarly, the reason patients with Down's syndrome invariably develop AD may be because the extra chromosome associated with the disease contains the gene for the amyloid β -protein precursor (APP), the protein from which the main plaque component A β is generated. The presence of a third active copy of the APP gene in cells results in increased levels of A β , which in turn serves to initiate, or nucleate, fibril formation. One of the unsolved mysteries of amyloid disease is how nucleation of amyloid growth occurs in most disease states. When studied *in vitro*, proteins responsible for amyloid disease often require extremely high, nonphysiological concentrations for nucleation and growth of amyloid fibrils. This indicates that there are unknown specific structures and/or environments in the body that facilitate nucleation. Amyloid plaques typically contain other proteins in addition to the main fibril component and one role of some of these proteins may be to contribute to amyloid formation by helping to stabilize the fibrils and protect them from normal degradative mechanisms. Amyloid diseases are often age related, leading to the speculation that sporadic amyloid deposition may be a not uncommon occurrence that is held in check in the young and healthy but which advances when normal surveillance mechanisms break down in the aged. This may account for the finding that, in the inherited amyloidosis associated with mutant forms of amyloidogenic precursor proteins, the diseases generally occur later in life, despite the fact that the proteins are expressed from birth.

Prions and Amyloid Enhancement Factor

In one group of amyloid-related brain diseases, a group including Creutzfeldt-Jakob disease, mad cow disease, and the sheep disease scrapie, the conditions can be transmitted by ingestion of prion particles, misfolded versions of a normal cellular protein, from the nervous system tissue of a previous victim. Although it is not firmly established that the actual prion particle or state itself is an amyloid fibril, the ability of prion infectivity levels to be amplified in the body bears an intriguing resemblance to the ability of amyloid fibrils to amplify themselves from a pool of precursor protein, through seeding. All mammals express a version of the prion precursor protein (PrP), which has a normal function and does not normally lead to disease. Except in rare instances of sporadic and genetic forms of the disease, prion-associated diseases are observed only when the animal has been exposed to exogenously supplied prion particles, which appear to act *in vivo* as seeds or templates for the propagation of new prions from the normal PrP pool. This scenario is reminiscent of the phenomenon that serum amyloid A amyloidosis can be accelerated in

experimental animals by administration of an agent called amyloid enhancement factor (AEF). AEF extracted from the amyloidotic spleen of a mouse and injected intravenously into new mice can induce the rapid development of amyloid deposits compared to untreated controls. Biochemical studies are consistent with AEF being essentially composed of amyloid fibrils, although the presence of an important minor component is difficult to rule out. A possible strong connection between the AEF and prion phenomena has been suggested by experiments showing that AEF, as well as pure amyloid fibrils created *in vitro*, can induce amyloid disease in mice when placed in their drinking water. Thus, although most amyloid diseases are not considered to be transmissible, the results in this experimental system suggest that various amyloid fibrils might be capable of behaving as prions in some circumstances.

Toxicity of Amyloid Fibrils

Although in some peripheral conditions, amyloid can cause life-threatening disease by accumulating in such high mass that normal tissue structure and function are disrupted, in other cases, particularly in the neurodegenerative diseases, the accumulated mass of amyloid is very low compared to the surrounding cell mass. In these cases, the mechanism(s) by which amyloid fibrils cause cell death or cell dystrophy is not at all clear. Whether there is a uniform toxic mechanism in all diseases or different mechanisms for different diseases is also unknown. Toxicity mechanisms under investigation include the following: (1) collateral damage caused by immune responses to an amyloid deposit; (2) membrane depolarization resulting from channels created by amyloid fibril assembly intermediates inserted into membranes; (3) recruitment of other proteins into growing aggregates, which has the effect of denying the cell the activity of the recruited protein(s); and (4) disruption or overwhelming of the normal cellular apparatus for breakdown and elimination of misfolded proteins, such as the ubiquitin-proteasome system and the molecular chaperones.

Evolved Amyloid

As far as is known, most protein sequences in evolution were selected for their abilities to efficiently access functional, folded states, while being either unselected, or selected against, with respect to the ability to make amyloid fibrils. Thus, in most cases where amyloid forms, the amyloid can be viewed as an accidental, environmentally induced structure never intended by nature. There are several examples, however, of proteins whose ability to make amyloid fibrils in microbial cells is beneficial to the organism. Yeast prions, for example, function through their ability to form and seed amyloid fibrils and in doing so they modulate the levels of the soluble form of the prion protein in the cell, the fluctuations in which play metabolic and ultimately regulatory roles. In another example, a complex system of proteins in *Escherichia coli* is responsible for producing an extracellular amyloid fibril that plays a role in cell adhesion. The existence of functional, beneficial amyloid might be viewed as an example of nature's ability to exploit to advantage the novel properties of the products of evolution.

Amyloid Proteins

Range of Amyloid Proteins

Amyloid fibrils composed of different proteins share a number of common structural features, despite the fact that these proteins, in their native states, vary widely in structure, cellular locations, and properties. For example, the amyloid protein transthyretin is a tetramer of subunits with a total molecular weight of 55 kDa. In contrast, small peptides of approximately 40 amino acid residues, such as A β and IAPP, and fragments of IAPP as short as pentapeptides, are also capable of amyloid fibril formation. Proteins can grow into amyloid fibrils regardless of the nature of their normal folding patterns. In their native states, α -synuclein exhibits no stable structure in solution, transthyretin is dominated by β -sheet, lysozyme is a mixed α -helix/ β -sheet protein, and serum amyloid A is essentially fully α -helical, yet all are capable of forming amyloid fibrils. Most of the peripheral amyloidoses involve secreted proteins, whereas the nine different proteins responsible for nine expanded polyglutamine diseases – including Huntington's disease – are cellular proteins (which, in turn, are found in a variety of subcellular locations). Most amyloid proteins contain significant percentages of hydrophobic residues, but polyglutamine consists entirely of the relatively polar amino acid glutamine. One of the few patterns shared among amyloidogenic protein regions is the infrequent occurrence of proline residues, which strongly disfavor β -sheet – the dominant secondary structural feature of the amyloid fibril.

Molecular Factors in Amyloid Formation

Until the early 1990s, the unusually insoluble, fibrous, quasi-crystalline amyloid fibril was generally viewed as a structural form in which the subunit building block was essentially the native state of the amyloidogenic precursor protein, in a model analogous to structural models for aggregated sickle cell hemoglobin. A significant advance in the understanding of amyloidosis was the demonstration in the mid-1990s that amyloid formation is strongly enhanced when the folded state of a protein is destabilized by mutation or by the solution environment, indicating that fibril structure involves non-native states. These studies were later extended to show that denaturing conditions can induce amyloid formation even in some proteins not known to be pathogenic. It is clear that misfolding, that is, the rearrangement of the complex folded shape of a protein molecule, is central to amyloid formation. Thus, protein stability – the resistance of the folded conformation to misfolding/unfolding – is an important factor in determining susceptibility to amyloid formation. Extreme environments in the body, such as acidic cell compartments, may, in some cases, facilitate protein unfolding and therefore promote amyloid formation. Proteolytic removal of a portion of a protein by an endogenous protease can also be a destabilizing factor leading to amyloid formation. Notably, mutations that alter the primary structure of proteins can affect protein stability. In fact, many of the amyloid diseases involve amino acid substitutions in an amyloid precursor protein. The most common example is familial amyloidosis, caused by a variety of single point mutations in the protein transthyretin. In addition,

noninheritable, specific amino acid changes in immunoglobulin light chains are linked to a high risk of amyloidosis. Amino acid changes can contribute to the efficiency of amyloid formation not only by modifying the stability of the native state of the precursor protein, but also by modifying the stability of the amyloid fibril itself; an example of the latter is the ability of proline residues to destabilize amyloid structure.

Amyloid Assembly

The manner in which amyloid fibrils grow can be considered in terms of the kinetics of the process or the structures of assembly intermediates. Kinetically, amyloid fibril formation *in vitro* typically exhibits a lag time of hours to days, followed by a rapid growth phase. In many such cases, the lag phase can be abbreviated or eliminated if a small amount of fibril is provided as seed at the start of the reaction. Such behavior has been interpreted as evidence that fibril growth takes place by nucleated growth polymerization, a mechanism of colloidal assembly in which the initiation of aggregate growth depends on the sporadic generation of a highly unstable, highly organized nucleus state. However, it is far from clear how amyloid growth is initiated *in vivo* and it is possible, even likely, that other molecules or structures are involved, which would make the *in vivo* mechanism more akin to what in physics is referred to as heterogeneous nucleation. *In vitro*, after the nucleation (or seeding) phase has occurred, growth continues predominantly via the elongation of existing fibrils and it has been argued that this is a more realistic view of how amyloid develops *in vivo*. A number of assembly intermediates have been detected when amyloid growth is studied *in vitro*. Whether or not any of these can be considered to be true kinetic nuclei, some of these structures may have relevance *in vivo*. Amyloid growth from monomeric proteins seems to proceed via the formation of successively larger, more complex structures, from small globular aggregates to short curvilinear filaments to isolated straight and unbranched fibrils to bundles of fibrils. The early assembly intermediates in this progression appear to be more cytotoxic than the mature fibrils, leading to the speculation that assembly intermediates are the real culprits in disease pathogenesis. Little is known, however, about whether and how these intermediates might be formed in the body and how they might kill cells.

Amyloid Structure

Amyloid ultrastructure as revealed by electron microscopy (Figure 2) or atomic force microscopy shows fibrils to be relatively straight and unbranched, with diameters in the range of 80–160 Å. Fibrils have a characteristic twist, being composed of two to six protofilaments of diameter 30–40 Å. X-ray fiber diffraction studies reveal that fibrils and their constituent protofilaments are rich in a type of β -sheet structure known as cross- β . In this conformational structure type, the extended chain elements of the β -sheet are perpendicular to the fibril axis and the hydrogen bonds between the strands are parallel to the fibril axis. Beyond this level of resolution, little is known about the detailed protein folding of the amyloid motif because of imposing technical barriers to their study by the

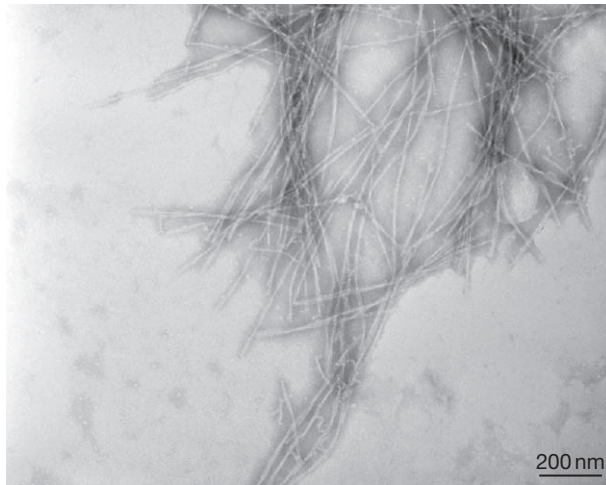


Figure 2 An electron micrograph of an amyloid fibril grown in the laboratory from the Alzheimer's disease peptide A β . Courtesy of professors Indu Kheterpal and John Dunlap, University of Tennessee.

standard methods of X-ray crystallography and solution-phase nuclear magnetic resonance spectroscopy. Nevertheless, a number of structural models have been proposed and important insights continue to be obtained through lower resolution studies. The structure of the amyloid folding motif is important for a number of reasons, only one of which is their association with a variety of human diseases. It is clear that misfolded, aggregated proteins are formed routinely in the cell due to fundamental inefficiencies in the protein-folding process; understanding amyloid structure may help elucidate how the

cell manages to recognize and eliminate the products of folding mishaps. Mapping amyloid structure and assembly will also contribute to the understanding of this previously ignored aspect of the protein-folding reaction – unproductive side products. Finally, better structural information may allow a rational design approach to the development of anti-amyloid therapeutics.

See also: **Bioenergetics:** Cell Death by Apoptosis and Necrosis; **Molecular Biology:** Prions of Yeast and Fungi: Proteins Acting as Genes; **Protein/Enzyme Structure Function and Degradation:** Prions Overview; Proteasomes, Overview; Ubiquitin System.

Further Reading

- Dobson CM (2001) The structural basis of protein folding and its links with human disease. *Philosophical Transactions of the Royal Society B: Biological Sciences* 356: 133–145.
- Falk RH, Comenzo RL, and Skinner M (1997) The systemic amyloidoses. *New England Journal of Medicine* 337: 898–909.
- Kelly JW (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Current Opinion in Structural Biology* 8: 101–106.
- Lindquist S (1997) Mad cows meet psi-chotic yeast: The expansion of the prion hypothesis. *Cell* 89: 495–498.
- Martin JB (1999) Molecular basis of the neurodegenerative disorders. *New England Journal of Medicine* 340: 1970–1980.
- Selkoe DJ (1999) Translating cell biology into therapeutic advances in Alzheimer's disease. *Nature* 399: A23–A31.
- Sipe JD and Cohen AS (2000) Review: History of the amyloid fibril. *Journal of Structural Biology* 130: 88–98.
- Wetzel R (1994) Mutations and off-pathway aggregation. *Trends in Biotechnology* 12: 193–198.

Anaplerosis

R R Russell III, Yale University School of Medicine, New Haven, CT, USA

H Taegtmeier, The University of Texas Medical School at Houston, Houston, TX, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Anaplerosis The entry of substrates into the citric acid cycle as intermediates other than acetyl-CoA, thereby increasing the citric acid cycle pool size.

Carboxylation The introduction of a carboxyl group into a compound.

β -Oxidation The oxidative metabolism of fatty acids through a cycle of reactions that removes successive two-carbon units (acetyl-CoA) from the fatty acid.

The word anaplerosis (from the Greek, meaning to fill up) was coined in the 1960s by Sir Hans Kornberg to describe pathways that replenish intermediates in metabolic cycles. A metabolic cycle transforms a substrate with little or no change in the concentration of the cycle's constituents, the central process in energy substrate metabolism. A classic example for a metabolic cycle is the citric acid cycle. Here anaplerotic pathways mediate the entry of metabolites into the cycle as compounds other than acetyl-CoA and thereby replenish the pool size of the citric acid cycle. Other anaplerotic pathways are present in eukaryotic and prokaryotic organisms; however, this article focuses on anaplerosis as it relates to the citric acid cycle in mammalian tissues and the importance of anaplerosis in the maintenance of organ function.

Cycles have evolved because in a competitive environment the chances for survival are greatest if resources are used in an optimal way. For example, energy transfer in heart muscle is linked to a series of cycles, beginning with the circulation and ending with actin-myosin cross-bridge formation (Figure 1). Steady-state concentrations of metabolic intermediates in a cycle depend on their rates of synthesis and degradation. In other words, substrates entering the citric acid cycle as citrate (via the condensation of acetyl-CoA with oxaloacetate) do not cause a net change in the citric acid cycle pool size because the two carbons of acetyl-CoA are ultimately lost as CO₂ in the isocitrate and the α -ketoglutarate dehydrogenase reactions, respectively. In contrast, anaplerosis supplies compounds to the citric acid cycle through reactions other than those catalyzed by citrate synthase. These compounds replenish carbon intermediates lost primarily as amino acids (glutamate or aspartate) or as oxaloacetate.

The role of anaplerosis in the citric acid cycle is threefold. First, as intermediates are drained away during biosynthetic processes such as gluconeogenesis, glyceroconeogenesis, and amino acid synthesis, anaplerotic pathways replace them, assuring sufficient carbon flux for citrate synthesis and, hence, allow citric acid cycle flux to proceed unimpaired. These aspects of anaplerosis are especially important in the liver and renal cortex. Second, anaplerosis facilitates energy transfer in organs with high rates of energy turnover, such as the heart and skeletal muscle, especially under situations such as ischemia in which normal functioning of the citric acid cycle is impaired. Third, anaplerosis also regulates whole-body

metabolic homeostasis through the regulation of insulin secretion from the β -islet cells of the pancreas. We now focus on the mitochondrial anaplerotic pathways.

Anaplerotic Pathways

Substrates for Anaplerosis

The known anaplerotic pathways for the citric acid cycle are depicted in the schematic drawing of Figure 2. As a key component of all living cells, the citric acid cycle is highly organized and responds instantly to changes in its environment. The citric acid cycle is tied into a distributive system of energy transformation reactions (Figure 1) and flux through the anaplerotic pathways. The primary substrates for anaplerosis are lactate and alanine as precursors of pyruvate, although glutamate can be a physiologically significant anaplerotic substrate as well. In addition, the branched chain amino acids valine and isoleucine, as well as odd-chain fatty acids and their metabolic end product, propionate, serve as anaplerotic substrates.

Entry at Malate and Oxaloacetate

In the heart the major anaplerotic substrate in the heart is pyruvate. Although the majority of pyruvate entering the mitochondria is oxidatively decarboxylated (and most likely by CO₂ arising from pyruvate decarboxylation) and enters the citric acid cycle as acetyl-CoA, a portion is also carboxylated and enters the four-carbon (C-4) pool of the citric acid cycle. This conversion is catalyzed by either pyruvate carboxylase (PC, E.C. 6.4.1.1), forming oxaloacetate or by nicotinamide adenine dinucleotide phosphate (NADP)-dependent malic enzyme (E.C. 1.1.1.40), forming malate. While the NADP-dependent malic enzyme reaction is reversible, the reaction in the direction of malate production is slow and unlikely to play a significant role in anaplerosis. Another provider of C-4 intermediates in the citric acid cycle is aspartate, which undergoes transamination with α -ketoglutarate via aspartic aminotransferase to form oxaloacetate and glutamate. In contrast to the carboxylation of pyruvate, the entry of aspartate as oxaloacetate is balanced by the loss of α -ketoglutarate from the citric acid cycle, resulting in the homeostasis of citric acid cycle intermediates.

Entry via α -Ketoglutarate

Mitochondrial α -ketoglutarate can be readily formed from glutamate via transamination with pyruvate (forming alanine) or oxaloacetate (forming aspartate). The transamination of glutamate with oxaloacetate by glutamate oxaloacetate transaminase (E.C. 6.2.1.1) results in a gain of a C-5 intermediate at the same time that a C-4 intermediate is lost from the citric acid cycle. Of importance is that the mitochondrial pools of α -ketoglutarate and glutamate are in equilibrium and therefore labeling patterns of glutamate are identical to those for α -ketoglutarate. As discussed below, this relationship between α -ketoglutarate and glutamate can be exploited in measuring citric acid cycle flux, relative rates of contribution of substrates to the citric acid cycle, and rates of anaplerosis. Furthermore, glutamate can be converted to α -ketoglutarate and NH_4^+ by glutamate dehydrogenase, which

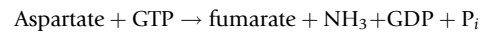
is present in a variety of tissues, including the liver, kidney, brain, heart, and skeletal muscle.

Entry via Succinyl-CoA

The citric acid cycle intermediate succinyl-CoA plays an important role in fatty acid and amino acid metabolism because it is the entry point of odd-chain fatty acids, propionate, and the branched chain amino acids valine and isoleucine into the citric acid cycle. Substrate entry as α -ketoglutarate or succinyl-CoA, in contrast to other anaplerotic pathways, is associated with the generation of energy-rich phosphates via a substrate-level phosphorylation. The reaction catalyzed by succinyl-CoA synthetase (E.C. 6.2.1.4) normally favors the conversion of succinyl-CoA to succinate and leads to substrate level phosphorylation of GDP to GTP. This energy-providing pathway becomes important under anaerobic conditions when ATP generation by oxidative phosphorylation is inhibited, such as in the setting of ischemia.

Entry via Fumarate

In addition to the amino acids phenylalanine and tyrosine entering as fumarate, the purine nucleotide cycle enriches the citric acid cycle based on the net reaction



Several workers have demonstrated that flux through the purine nucleotide cycle increases in skeletal muscle during intense exercise. This increased flux has two effects: first, it can maintain the pool of adenine nucleotides, and second, it can increase the citric acid cycle pool size via anaplerosis. Both of these effects are expected to improve energy metabolism in exercising muscle.

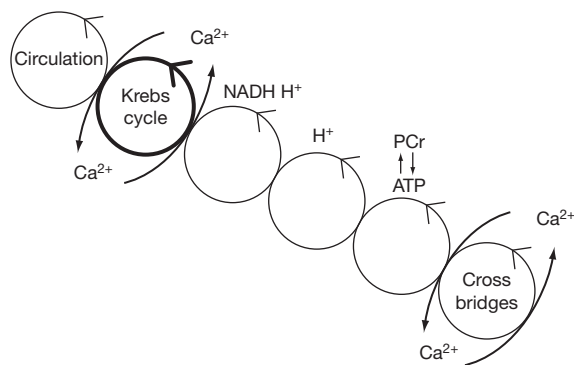


Figure 1 Cycles of energy transfer in heart muscle. The central role of the Krebs cycle is highlighted. See text for details.

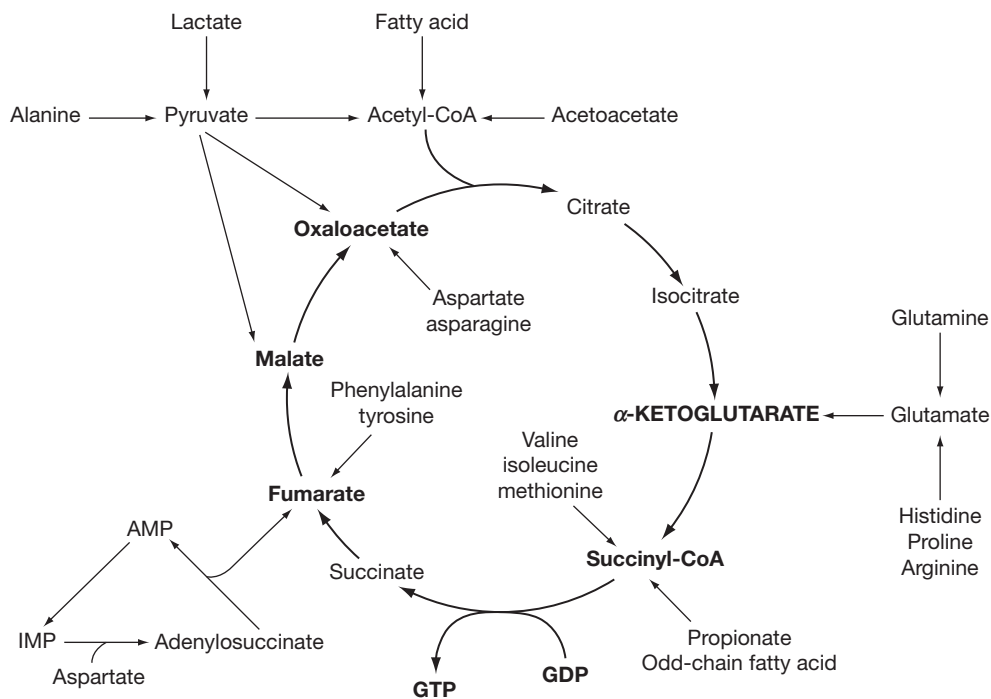


Figure 2 Summary of anaplerotic pathways leading to the citric acid cycle. See text for details.

Exit of Cycle Intermediates: Balancing Anaplerosis

Under steady-state conditions, substrates entering the citric acid cycle via anaplerotic pathways are balanced by removing an equivalent amount of citric acid cycle intermediates through pathways that maintain a constant citric acid cycle pool size. These pathways have been termed cataplerotic, although the term is somewhat an oxymoron; instead, we prefer the more appropriate term drainage. Drainage pathways generally involve removing the citric acid cycle of cycle intermediates oxaloacetate (as aspartate via transamination, or phosphoenolpyruvate via decarboxylation by phosphoenolpyruvate carboxykinase (E.C. 4.1.1.32)), citrate via export from the mitochondrial matrix, or α -ketoglutarate via transamination to glutamate. While these drainage pathways are usually thought of as means to balance anaplerotic pathways, they play critical roles in renal gluconeogenesis and enterocyte energy production from glutamine (see below).

Measuring Rates of Anaplerosis

Changes in Citric Acid Cycle Pool Size

There are two principal methods for assessing anaplerosis based on changes in citric acid cycle pool size. The amount of citric acid cycle pool intermediates can be measured enzymatically or resolved with high-performance liquid chromatography. While both methods can reliably assess changes in the citric acid pool size, they do not provide insight into the pathways involved in anaplerosis nor do they determine relative rates of citric acid cycle enrichment from anaplerotic substrates. Furthermore, because anaplerosis is usually balanced by the exit of intermediates, changes in citric acid pool size are generally negligible.

Simply determining changes in citric acid cycle pool size provides no information on rates of anaplerotic flux or on specific anaplerotic reactions. Using substrates labeled with tracer ^{14}C and measuring incorporation into specific positions in citric acid cycle intermediates is more revealing. This method can determine rates of anaplerosis and is able to characterize the pathways involved in the process. However, the method is also laborious and prone to error. It requires sequential enzymatic degradation of intermediates quantifying the release of the ^{14}C label. This method has generally been replaced by ^{13}C -NMR spectroscopy or mass spectroscopy (NMR, nuclear magnetic resonance).

^{13}C -NMR/Mass Spectroscopy

^{13}C -NMR spectroscopy is used to assess quantitatively the relative contributions of various substrates to the citric acid cycle. This method determines the anaplerotic enrichment of individual carbon positions in citric acid cycle intermediates, similar to the radiotracer method discussed before, but without the need for digesting the carbon skeleton. ^{13}C -labeled compounds enter the carbon skeleton of citric acid cycle intermediates at specific positions through anaplerotic pathways or through dilution of ^{13}C by entry of unlabeled anaplerotic substrates. The enrichment of ^{13}C at those positions is then assessed and rates of anaplerotic flux are determined relative to rates of citric acid flux (a/c).

The method is illustrated in Figure 3 in which the enrichment of the carbon skeleton of α -ketoglutarate (and therefore glutamate, which is present in sufficient concentrations to be measurable by NMR spectroscopy) at the C-3 and C-4 positions is used to quantify anaplerosis. Label in these positions is generated from randomization of the C-4 carbon in the first turn of the citric cycle. Therefore, the sum of the ^{13}C enrichments of the C-2 and C-3 carbons will be equal to the enrichment of the C-4 carbon of glutamate. In contrast, with anaplerotic activity, the labeled C-2 and C-3 carbons will be diluted by ^{12}C carbons arising from unlabeled anaplerotic substrates, and the sum of the enrichments of the labeled C-2 and C-3 carbons will be less than the enrichments of the C-4 carbon of glutamate. For example, using $[2\text{-}^{13}\text{C}]\text{acetyl-CoA}$ as a substrate and conditions where there is no anaplerotic flux, analysis of the enrichment of C-2, C-3, and C-4 positions of glutamate reveals that $a/c = 0$, indicating no anaplerotic flux. Further refinements to the NMR-based methods take into account the recycling of carbons in pyruvate (a major source of anaplerotic substrate) derived from the citric acid cycle to improve the accuracy of measurements of anaplerotic flux.

Expression and Activity of Proteins Regulating Anaplerosis

The enzyme pyruvate carboxylase was discovered half a century ago, and over the last decade there has been much progress in understanding its structure and function. Using methods outlined above, studies have focused on changes in enrichment of citric acid cycle intermediates to assess the activity of anaplerotic pathways and explain changes in flux in terms of mass action and allosteric regulation of key enzymes. While this is most likely the case with acute changes in anaplerotic flux, changes in the expression and activity of enzymes responsible for anaplerosis must be taken into account when studying chronic processes such as pressure-overload-induced hypertrophy, heart failure, or diabetes. In streptozocin-treated diabetic rats, transcript levels of anaplerotic enzymes in heart and skeletal muscle are downregulated (Figure 4). Furthermore, mice with germline deletion of peroxisome proliferator-activated receptor- α show an increase in pyruvate carboxylase mRNA expression. Changes in transcript levels are paralleled by similar changes in protein expression. Although, the *in vitro* activities of the anaplerotic enzymes involved in the carboxylation of substrates (e.g., pyruvate carboxylase) can be measured by determining the incorporation of ^{14}C from $\text{H}^{14}\text{CO}_3^{-1}$, it is not known whether changes in expression correlate with changes in activity. However, in pressure overload-induced left ventricular hypertrophy in the rat, compensatory increases in anaplerotic flux into the citric acid cycle are accompanied by increases in transcript levels of malic enzyme.

Changes in Anaplerosis in Response to Environmental Stress: Workload, Nutritional Status, and Disease

Ultimately, any metabolic process has functional consequences, and anaplerotic pathways are no exception. Anaplerotic pathways play important roles in regulating a wide variety of organ responses to conditions of metabolic stress ranging from exercise to inborn errors of metabolism.

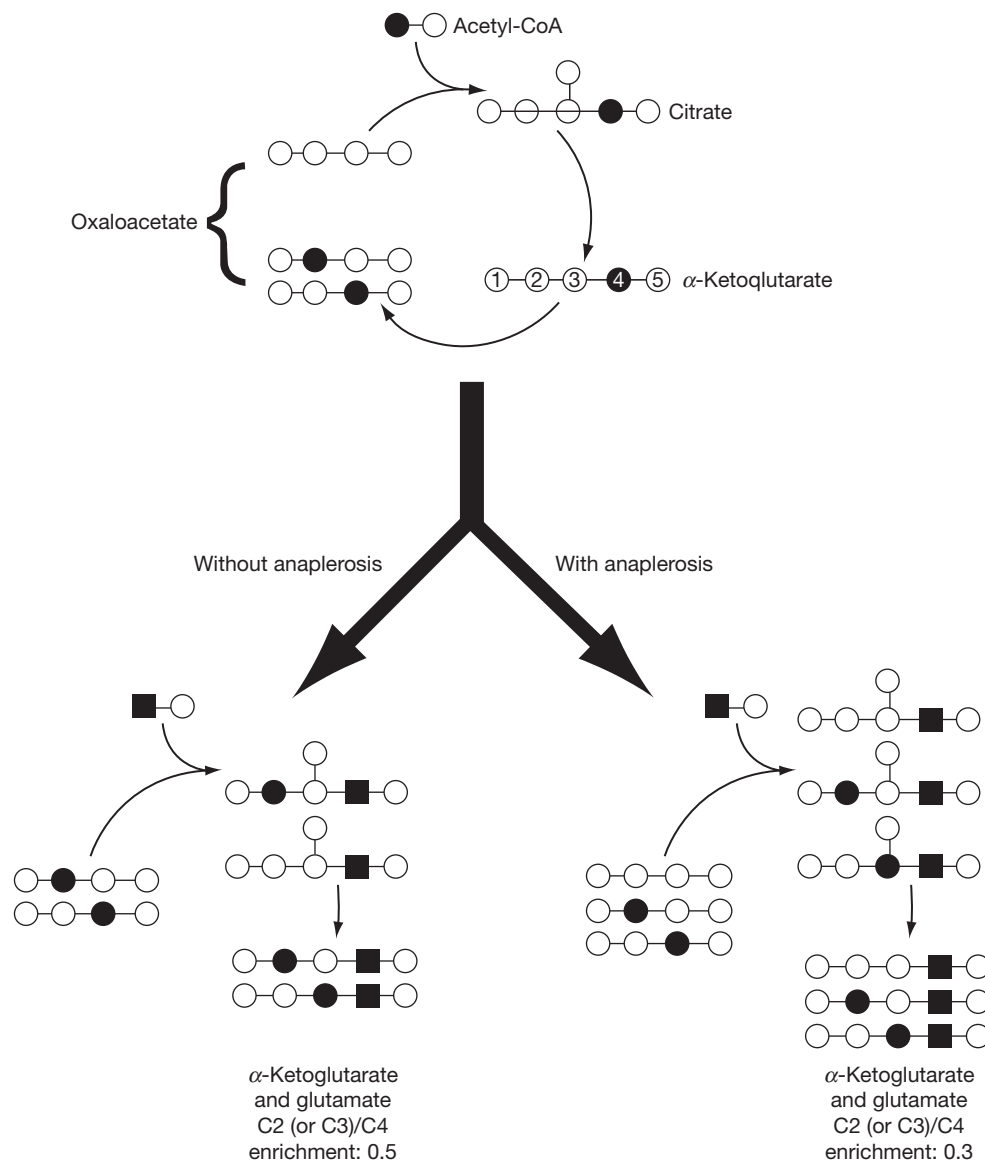


Figure 3 Fundamental concept of assessing anaplerotic flux using ^{13}C -NMR spectroscopy. The black circles represent ^{13}C that enters the citric acid cycle pool via the first turn of the citric acid cycle. The black squares represent ^{13}C that enters the citric acid cycle pool via the second turn of the citric acid cycle. From Malloy C, Sherry A, and Jeffrey F (1988) Evaluation of carbon flux and substrate selection through alternate pathways involving the citric acid cycle of the heart by ^{13}C -NMR spectroscopy. *Journal of Biological Chemistry* 263: 6964–6971.

Anaplerosis in Skeletal Muscle during Exercise

The transition from rest to moderate or intense exercise is associated with large increases in skeletal muscle adenosine triphosphate (ATP) turnover, which must be supported by up to 100-fold increases in citric acid cycle flux. Although citric acid cycle intermediates increase only three- to fourfold, these relatively small increases in the citric acid cycle pool size (via anaplerosis) can reflect dramatic increases in citric acid cycle flux. It has been suggested that the increases allow skeletal muscle to adapt to the energetic demands of exercise, but interestingly, there is no appreciable change in citric acid cycle pool size at lower levels of exercise. The majority of changes in individual intermediates occur in the second span of the citric acid cycle (i.e., from succinate to oxaloacetate)

because the primary source of enrichment during acute exercise appears to be flux through the reaction catalyzed by alanine aminotransferase (E.C. 2.6.1.2). This reaction results in entry of glutamate as α -ketoglutarate; however, pyruvate carboxylase and malic enzyme may contribute minor amounts to citric acid cycle enrichment in exercising muscle.

Glutamine Metabolism by the Small Intestine

Glutamine is a source of energy for a number of specialized tissues, including the small intestine. When glutamine is taken up in the cell, it enters the citric acid cycle as α -ketoglutarate and leaves the cycle through oxidation as CO_2 . More specifically, α -ketoglutarate is converted to malate by citric acid cycle reactions and malate may be transported out of the

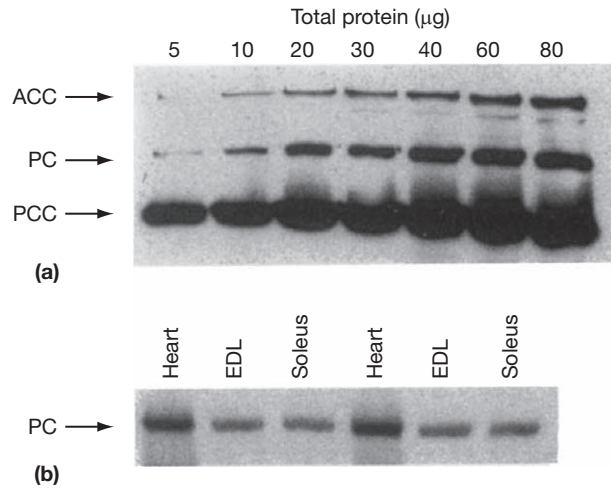


Figure 4 Expression of the biotin-containing carboxylases, acetyl-CoA (ACC), pyruvate carboxylase (PC), and propionyl-CoA carboxylase (PCC), in heart muscle based on streptavidin blotting (a) and pyruvate carboxylase expression in rat heart and skeletal muscle (b). Reprinted from Gibala MJ, Young ME, and Taegtmeyer H (2000) Anaplerosis of the citric acid cycle: Role in energy metabolism of heart and skeletal muscle. *Acta Physiologica Scandinavica* 168: 657–665.

mitochondria. In the cytosol, malate is converted to oxaloacetate. The malate–aspartate shuttle is bidirectional and can also transport electrons from extramitochondrial reduced nicotinamide dinucleotide (NADH) into the mitochondria or from intramitochondrial NADH to the cytosol.

Renal Ammonia Formation during Starvation

During starvation, when protein breakdown, renal gluconeogenesis and hepatic ketogenesis increase. Concomitantly, the rate of renal ammoniogenesis also increases. Ammonia is generated from amino acids, including glutamine, released by skeletal muscle. Glutamine is converted to α -ketoglutarate in renal cells. Alpha-ketoglutarate is metabolized to malate. Malate is transported out of the mitochondria, oxidized to oxaloacetate and to phosphoenolpyruvate, and ultimately used for gluconeogenesis. Although different from glutamine oxidation by enterocytes, this series of reactions also reflects the balance that is generally observed between anaplerotic and drainage pathways.

Ischemia and Substrate Level Phosphorylation

Early studies assessing cardiac metabolism by measuring arteriovenous differences in amino acid concentrations revealed that in patients with coronary artery disease and myocardial ischemia, the heart avidly extracts glutamate and releases alanine. This finding led to the hypothesis that the anaplerotic substrate glutamate enters the citric acid cycle as α -ketoglutarate via transamination with pyruvate (thereby forming the alanine that is released). Subsequently, the α -ketoglutarate is metabolized to succinate with the concomitant substrate level phosphorylation of guanosine diphosphate (GDP) to form guanosine triphosphate (GTP) (Figure 5). In this way, a span of the citric acid cycle can generate high-energy phosphates in the absence of sufficient oxygen for full citric acid cycle

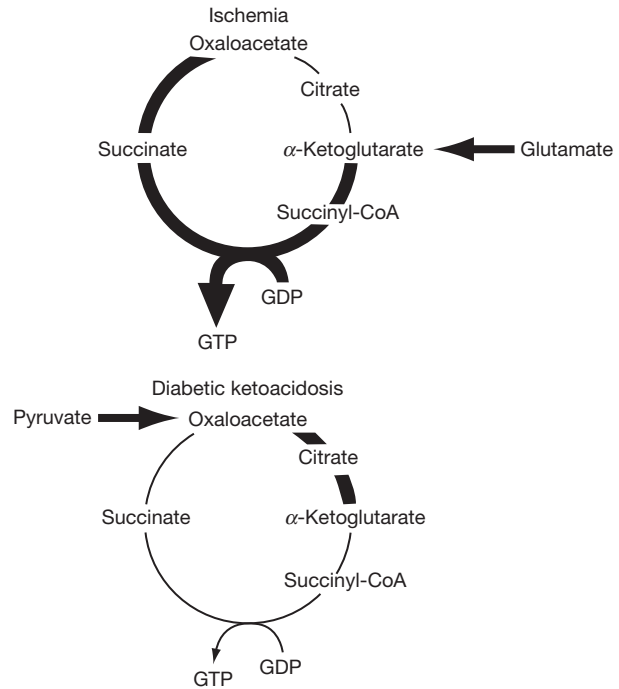


Figure 5 Role of anaplerosis in improving myocardial energetics in the setting of ischemia (upper panel) or diabetic ketoacidosis (lower panel). Under both conditions, loss of cofactors ($\text{NAD}^+/\text{FAD}^+$ for ischemia, CoASH for diabetic ketoacidosis) inhibits full citric acid cycle activity. Anaplerotic pathways allow the citric acid cycle to work in spans, thereby increasing the production of high-energy phosphates. Modified from Taegtmeyer H and Passmore JM (1985) Defective energy metabolism of the heart in diabetes. *Lancet* 1: 139–141.

operation. Recent ^{13}C -NMR spectroscopic studies have demonstrated that during myocardial ischemia, the relative contribution of anaplerosis to citric acid flux may increase sevenfold compared to nonischemic conditions. Translational research based on the concept of this anaplerotic pathway has led to the development of glutamate-enriched solutions that increase the provision of anaerobic energy for the heart during coronary artery bypass surgery.

Insulin Secretion, Diabetes and Ketone Body Metabolism

Insulin secretion occurs through two mechanisms. The first is a process that involves closing of the cell-surface ATP-sensitive potassium channels in response to increases in the circulating glucose concentrations, which stimulates exocytosis of insulin-containing vesicles from the β -islet cells through an increase in the cytosolic calcium concentration. The second mechanism is dependent on pyruvate carboxylase, which is highly expressed in β -islet cells; it has been estimated that 35–45% of pyruvate enters the citric acid cycle through this anaplerotic pathway in β -islet cells. Inhibition of pyruvate carboxylase with phenylacetic acid decreases glucose-stimulated insulin release from β -islet cells. Furthermore, there is evidence that pyruvate carboxylase plays an important role in the early stages of type 2 diabetes. Specifically, in Zucker fatty rats with insulin resistance, the hyperfunctioning β -islet cells increase insulin production in part through increases in pyruvate carboxylase activity.

Citric acid cycle pool size increases in hearts of rats with experimentally induced diabetes, suggesting enrichment by anaplerotic pathways. We have suggested that an increase in anaplerotic flux, which primarily occurs through pyruvate carboxylation (via malic enzyme), plays an important role in maintaining flux through the second span of the citric acid cycle. Acutely, the metabolic derangement of ketoacidosis that occurs with diabetes inhibits flux through α -ketoglutarate dehydrogenase by sequestration of coenzyme A (CoASH). This phenomenon is associated with contractile dysfunction of the heart that can be readily reversed by the addition of glucose, lactate, or pyruvate (all of which are anaplerotic substrates). The effects of pyruvate are mediated by enrichment of malate in the citric acid cycle pool, which occurs by carboxylation of pyruvate to form malate and oxaloacetate through the actions of malic enzyme and pyruvate carboxylase, respectively (Figure 5). The citric acid cycle is thereby able to operate once again in a span that can generate reducing equivalents to support oxidative phosphorylation of adenosine diphosphate to form the ATP necessary to drive the contractile machinery of the heart.

Long-Chain Fatty Acid Oxidation Defects and Myopathies

The inability to oxidize long-chain fatty acids due to deficiencies in activity of carnitine palmitoyltransferase-1 (E.C. 2.3.1.21) or the enzymes involved in β -oxidation is associated with contractile dysfunction due to skeletal and heart muscle damage. One proposed mechanism responsible for the decrease in contractile activity is decreased citric acid cycle flux due to loss of intermediates from damaged myocytes. Based on this hypothesis, one clinical study has treated patients with defects in long-chain fatty acid oxidation with odd-chain triglycerides (which can increase the citric acid cycle pool size by entering as succinyl-CoA) and tested the hypothesis that an increase in citric acid cycle may improve muscle function. As proof of principle, the study has demonstrated beneficial effects such as reversing left ventricular dysfunction, decreasing muscle breakdown, and decreasing weakness. However, the effects of medium-length odd-chain fatty acids as potential anaplerotic substrates for the treatment of other forms of heart failure have not yet been assessed.

Pyruvate carboxylase deficiency is a rare autosomal recessive disease that affects a variety of organs, including the liver, kidney, brain, skeletal muscle, heart, and adipose tissue. The disorder is characterized by impaired gluconeogenesis, lactic acidosis, and, depending on the severity of the disease, neurologic manifestations ranging from minimal neurologic abnormalities to profound psychomotor retardation. In the most severe form of the disease, patients die within the first few weeks of life. Nutritional treatment of the disease with anaplerotic odd-chain triglycerides, such as triheptanoin, or aspartate has been shown to prolong life.

Summary and Perspective

Efficient energy transfer in the mammalian cell is linked to a series of moiety-conserved cycles, including the citric acid

cycle. Changes in the cell's environment lead to depletion and replenishment (anaplerosis) of citric acid cycle intermediates. The multiple pathways of anaplerosis utilize a variety of substrates, including carbohydrates, odd-chain fatty acids, and amino acids. These pathways reflect a system of redundancy that is important for both function and survival of the cell.

See also: **Bioenergetics:** Mitochondrial Metabolite Carrier Protein Family; **Metabolism Vitamins and Hormones:** Amino Acid Metabolism; Branched-Chain Amino Acids.

Further Reading

- Baldwin JE and Krebs HA (1981) The evolution of metabolic cycles. *Nature* 291: 381–382.
- Cohen DM and Bergman RN (1997) Improved estimation of anaplerosis in heart using ^{13}C NMR. *American Journal of Physiology* 273: E1228–E1242.
- Gibala MJ, Young ME, and Taegtmeyer H (2000) Anaplerosis of the citric acid cycle: Role in energy metabolism of heart and skeletal muscle. *Acta Physiologica Scandinavica* 168: 657–665.
- Gunawardana SC, Liu YJ, Macdonald MJ, Straub SG, and Sharp GW (2004) Anaplerotic input is sufficient to induce time-dependent potentiation of insulin release in rat pancreatic islets. *American Journal of Physiology – Endocrinology and Metabolism* 287: E828–E833.
- Kornberg H (1966) Anaplerotic sequences and their role in metabolism. *Essays in Biochemistry* 2: 1–31.
- Krebs HA and Johnson WA (1937) The role of citric acid in intermediate metabolism in animal tissues. *Enzymologia* 4: 148–156.
- Malloy C, Sherry A, and Jeffrey F (1988) Evaluation of carbon flux and substrate selection through alternate pathways involving the citric acid cycle of the heart by ^{13}C NMR spectroscopy. *Journal of Biological Chemistry* 263: 6964–6971.
- Martini WZ, Stanley WC, Huang H, et al. (2003) Quantitative assessment of anaplerosis from propionate in pig heart *in vivo*. *American Journal of Physiology* 284: E351–E356.
- Mochel F, DeLonlay P, Touati G, et al. (2005) Pyruvate carboxylase deficiency: Clinical and biochemical response to anaplerotic diet therapy. *Molecular Genetics and Metabolism* 84: 305–312.
- Owen OE, Kalhan SC, and Hanson RW (2002) The key role of anaplerosis and cataplerosis for citric acid cycle function. *Journal of Biological Chemistry* 277(34): 30409–30412.
- Panchal AR, Comte B, Huang H, et al. (2000) Partitioning of pyruvate between oxidation and anaplerosis in swine hearts. *American Journal of Physiology* 279(5): H2390–H2398.
- Pound KM, Sorokina N, Ballal K, et al. (2009) Substrate–enzyme competition attenuates upregulated anaplerotic flux through malic enzyme in hypertrophied rat heart and restores triacylglyceride content: Attenuating upregulated anaplerosis in hypertrophy. *Circulation Research* 104: 805–812.
- Roe CR, Sweetman L, Roe DS, David F, and Brunengraber H (2002) Treatment of cardiomyopathy and rhabdomyolysis in long-chain fat oxidation disorders using an anaplerotic odd-chain triglyceride. *Journal of Clinical Investigation* 110: 259–269.
- Russell R and Taegtmeyer H (1991) Changes in citric acid cycle flux and anaplerosis antedate the functional decline in isolated rat hearts utilizing acetoacetate. *Journal of Clinical Investigation* 87: 384–390.
- Russell R and Taegtmeyer H (1991) Pyruvate carboxylation prevents the decline in contractile function of rat hearts oxidizing acetoacetate. *American Journal of Physiology* 261: H1756–H1762.
- Sorokina N, O'Donnell JM, McKinney RD, et al. (2007) Recruitment of compensatory pathways to sustain oxidative flux with reduced carnitine palmitoyltransferase I activity characterizes inefficiency in energy metabolism in hypertrophied hearts. *Circulation* 115: 2033–2041.

Angiotensin Receptors

T Inagami, Vanderbilt University, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 111–115, © 2004, Elsevier Inc.

Glossary

Angiotensin II Peptide hormone derived from angiotensinogen.

Cardiovascular effects Effects on the heart and blood vessels.

Receptor Hormone-binding protein.

Receptor signal Reaction emitted from the receptor upon hormone binding.

Angiotensin (Ang), initially considered as a major pressor substance, is now recognized to mediate numerous physiological and pathophysiological functions. Over the past 70 years, four different active forms of angiotensins, Ang II, Ang III, Ang IV, and Ang (1–7), have been identified. The complexity of their action was multiplied by subtypes of their respective receptors, AT_{1A}, AT_{1B}, and AT₂ for Ang II, AT₄ for Ang IV, and Ang (1–7) receptors (Table 1). Each subtype of the receptors shows more than one signaling mechanism. This article presents an overview of the formation and the structure of angiotensins, the receptors of Angs, and the representative signaling pathways, which play very colorful regulatory functions covering wide areas such as vasoconstriction, vasorelaxation cardiovascular hypertrophy and remodeling, atherosclerosis, thrombosis, stimulation of mineralocorticoid synthesis, reactive oxygen species (ROS) formation and release, facilitation of sympathetic outflow, renal electrolyte metabolism, control of central nervous system in water-drinking behavior, blood-pressure regulation, memory retention, growth-inhibition apoptosis, tissue differentiation, arachidonic acid and prostaglandin formation, and regulation of insulin signals.

Angiotensins (Ang) Structure and Formation

Angiotensin I

All angiotensins are derived from the amino terminus of the ~65 kDa prohormone angiotensinogen, via a series of proteolytic cleavage by a variety of proteinases and peptidases. The enzyme that initiates this process is renin. This aspartyl protease is active in pH 6–7, reacts exclusively with angiotensinogen, and cleaves only one singular leucyl peptide bond of the prohormone between residues 10 and 11 generating the decapeptide Ang I (Figure 1). Ang I is hormonally inactive; thus, it is prohormone as it has no specific receptor to transmit its signal. The physiological significance of Ang I is that it appears to confer a high degree of specificity to the angiotensin-generating system in plasma or in the intracellular system in which there are numerous proteins.

Angiotensin II

The inactive Ang I is activated to active Ang II by the metallo-endopeptidase, angiotensin I converting enzyme (ACE), which

is identical to kininase II. This enzyme was the target for the first most successful drugs for the treatment of renin-dependent hypertension, with a minimum of side effects and complications. Starting with captopril, a series of long-acting ACE inhibitors were synthesized and were found to almost completely block Ang II formation. Ang II has two subtypes of receptors, AT₁ and AT₂. The heptapeptide Ang III is the product of amino terminal cleavage of Ang II by aminopeptidases. Although Ang III is prominent in the neuronal system, it shares receptors, AT₁ and AT₂, with Ang II. The amino-acid sequence of rat AT₁ and AT₂ are shown in Figure 2.

Ang (1–7)

Ang (1–7) can be formed mainly from Ang I by the successive removal of the C-terminal Leu by ACE2 then the dipeptide Phe-His by Ace. It can be formed from Ang II by ACE 2 (Figure 1). It has been shown to activate phospholipase A₂ to release arachidonic acid from phospholipids leading to the formation of prostaglandins.

Ang IV

Ang IV is the hexapeptide Ang (3–8). It was reported to stimulate the production of plasminogen activator inhibitor to stabilize blood clots and to promote atherosclerosis. Other roles in the brain have also been reported.

Whereas receptors for Ang II, Ang III, and Ang (1–7) are G-protein-coupled receptors, the receptor isolated and cloned from the adrenal membranes is a single transmembrane type.

Angiotensin Receptor

Ang I-Binding Protein

Despite the ubiquitous distribution of Ang I-binding proteins, no physiological function was found for it.

Ang II Receptors

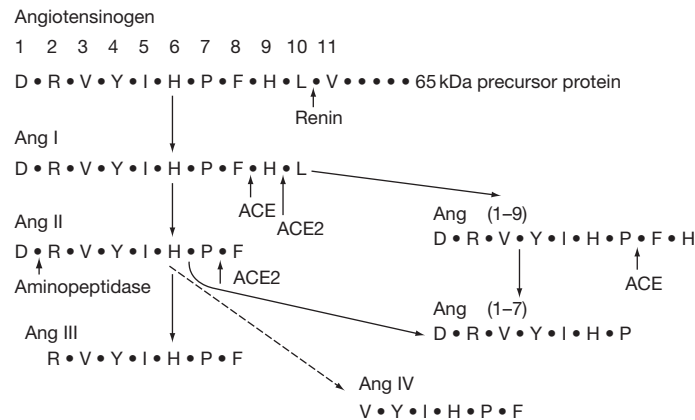
Receptor subtypes

Synthesis of a series of variants of Ang II revealed the presence of an Ang II receptor, confirming the earlier finding that there are two types of Ang II receptors, one resistant to dithiothreitol

Table 1 Angiotensin

Angiotensins	Receptors	G-proteins coupled
Ang II	AT _{1a}	G _{q/11} , G _i , G _{12/13}
	AT _{1b}	G _{q/11} , G _i , G _{12/13}
	AT ₂	G _i
Ang III	The same as Ang II	The same as AT ₁ and AT ₂
Ang (1–7)	<i>Mas</i> oncogene product	?
Ang IV	AT ₄ /IRAM	Single transmembrane receptor

and the other sensitive to the treatment. Inactive variants of Ang II were synthesized to obtain Ang II receptor antagonists. These inhibitor peptides also showed the importance of Arg², Tyr⁴, and Arg⁸ for type 1 Ang II receptors. Particularly, replacement of Phe⁸ to Ile, Thr, or Ala (saralisine) greatly reduced Ang II receptor signals, such as Ca²⁺ response, although binding is not significantly affected. However, these peptidic analogs were partial agonists. These peptidic inhibitors did not distinguish two different receptors, AT₁ and AT₂. It was only after the syntheses of nonpeptide inhibitors that indications were obtained for the presence of at least two receptor subtypes for Ang II, which differ in biochemical and physiological responses, localization, and ontogeny.

**Figure 1** Angiotensinogen and generation.

AT _{1A}		MALN	SSAEDGIKRI	QDDCPKAGRH	SYIFVM	IPTL	34
AT ₂	MKDNFSFAAT	SRNITSSLPF	DNLNATGTNE	SAFNCSHKPA	DKHLE	AIPVL	50
AT _{1A}	YSIIFVVGIF	GNSLVVIVII	FYMKL KTVAS	VFLNLALAD	LCFLLTLPLW		84
AT ₂	YIMIFVIGFA	VNIVVVSFLC	CQKGP KKVSS	IYIFNLAVAD	LLLLLATLPLW		100
TM-1							
AT _{1A}	AVYTAMEYRW	PFGNHLCKIA	SASV TFNLYA	SVFLLTCLSI	DRYLAIVHPM		134
AT ₂	ATYYSRYDWD	LFGPVMCKVF	GSFLTLMFA	SIFFITCMSV	DRYQSVIYPF		150
TM-3							
AT _{1A}	KSRLRRTMLV	AKVTCIIIWL	MAGLASLPAV	IHRNVYFIEN	TNITV CAFHY		184
AT ₂	LSQRRNP-WQ	ASYVPLVWC	MACLSSLPTF	YF RDVRTIEY	LGVNA CIMAF		199
TM-4							
AT _{1A}	ESRNSTLPIG	LGLT- KNILG	FLFPFLIILT	SYTLIWKALK	KAYEIQKNKP		233
AT ₂	PPEKYAQWSA	GIALMKNILG	FIIPILFIAT	CYFGIRKHL	KTNSYGNRI		249
TM-5							
AT _{1A}	RNDIFRIIM	AIVLFFFFSW	VPHQIFTFLD	VLIQLGVIHD	CKISDIVDTA		283
AT ₂	TRDQVLKMAA	AVVLAFFIICW	LFFHVL TFLD	ALTWMGIINS	CEVIAVIDLA		299
TM-6							
AT _{1A}	MPITICIAFY	NNCLNPLFYG	FLGKKFKKYF	LQLLKYIPPK	AKSHSSLSTK		333
AT ₂	LPFAILLGFT	NSCVNPFLYC	FVGNRFQOKL	RSVFRVPITW	LQGKRETMSC		349
TM-7							
AT _{1A}	MSTLSYRPSD	NMS SAKKPA	SCFEVE				359
AT ₂	RKSSSLREMD	TFVS					363

Figure 2 Rat angiotensin receptor AT_{1A} and AT₂.

Direct proof for the two major receptor subtypes, AT₁ and AT₂, were provided by expression cloning of their cDNAs from rat and bovine tissues. Rodents were found to have three genes: AT_{1a}, AT_{1b}, and AT₂. It is likely that birds, reptiles, and amphibians may also have more than one type of receptor. Rat or mouse murine AT_{1a} and AT_{1b} share a 98% amino acid sequence identity through their respective genes and are located in different chromosomes. They can be distinguished only by Northern blot analysis or *in situ* hybridization of 3' noncoding sequences, where a sizeable difference occurs in nucleotide sequences. In contrast, AT₁ and AT₂ share only 34% amino acid sequence homology. Noteworthy is the localization of the AT₂ gene in the X chromosome throughout species.

AT₁ receptor signaling

Most of the signal responses considered earlier as Ang II actions seem to be mediated by AT₁. These responses are smooth muscle contraction and hypertension, mineralocorticoid synthesis, tyrosine hydroxylase gene expression in adrenal and central and peripheral adrenergic facilitation, stimulation of cellular hypertrophy, mitogenesis and migration, plasminogen activator synthesis, stimulation of fibrosis, renal sodium retentions, centrally regulated dipsogenesis, hypertension, and generation of ROS. The major AT₁ signaling pathways are mediated by the Gq/11-mediated phospholipase β_1 activation via cellular [Ca²⁺] increase and protein kinase activation. However, AT₁ can also couple to Gi and G_{12/13}.

It is now well accepted that AT₁ can also activate pathways that involve tyrosine kinase activation, such as epidermal growth factor (EGF) receptor through Ca²⁺-stimulated heparin-binding EGF, and consequent activation of Ras, ERK 1/2, which further results in p70^{S6K} and phosphatidyl inositol-3-kinase and activation of immediate early gene pathways. AT₁ also activates the JAK-STAT system, Src family kinases, tyrosine phosphatase SHP-2, PLC- δ , JNK, and p38 MAP kinase. Accumulating lines of evidence seem to support the idea that some of these tyrosine kinase pathways do not involve G-proteins, based on results of mutagenesis studies that eliminate key residues required for interaction of cytoplasmic loops of AT₁.

Importantly, many of the tyrosine kinase activation steps are activated by the ROS, such as superoxide anion and hydrogen peroxide. The formation of superoxide anion is mediated by NAD(P)H oxidase, which is activated by Ang II. Important recent additions to the knowledge of Ang II-mediated smooth muscle contraction are the Rho-Rho kinase system. This information is of particular significance for the mechanism of hypertension. On the other hand, cardiovascular hypertrophy and remodeling of resistance vessels are other important aspects for cardiac failure and diminished vascular compliance and hypertension.

AT₂ receptor signaling

Studies on AT₂ signaling, particularly using cultured cells, were not straightforward because of the rapid disappearance of AT₂ receptor during repeated passage of cells and exposure to growth medium. Interpretation of results was also difficult because of rapid ontogenic change in AT₂ expression, particularly postnatal decline and increase under stress, during tissue repair or inflammation. *In vivo* studies eliminate many of these difficulties. Thus, *in vivo* studies of loss of function of AT₂ in the

kidney, vasculature, heart, and brain, and particularly adrenal and colon, where AT₂ remains expressed in adults, are producing substantive results.

In vitro studies on AT₂ receptor signals

Although AT₂ (the AT₂ receptor) is labile in cells in growth-culture medium, some cells, such as PC12w, N1E115, and R3T3, maintain AT₂ expression, but not AT₁, over several generations. Ang II stimulation resulted in growth inhibition via either MAP kinase phosphatase or the tyrosine phosphatase SHP1 even inducing apoptosis, but not all AT₂-expressing tissues or cells undergo apoptosis. Ovarian atretic cells express AT₂ prominently, yet Ang II treatment failed to induce apoptosis. AT₂ expressed in neuronal cells activate Ser/Thr phosphatase PP2A via a Gi-coupled mechanism and was reported to open delayed rectifier K⁺ channel, a hyperpolarizing mechanism. Its growth inhibitory mechanism on endothelial cells was shown to involve the following sequence of events:

Lowering pH → release of cytoplasmic kallikrein
→ generation of bradykinin → activation of NOS

This can explain relaxation or dilation of resistance arteries *ex vivo* by AT₂ and its reversal by its specific synthetic inhibitor PD123319 or PD123177, and relaxation by AT₂ partial agonists CGP24112. Although AT₂ was shown to bind specifically with G α_2 or G α_3 , not all of these growth-suppressing or cGMP-mediated signals were inhibited by pertussis toxin.

In vivo studies on AT₂ signals

AT₂ gene-deleted mice and AT₂ inhibitor-treated mice were shown to slowly retain Na and show impaired natriuresis and markedly diminished urinary cGMP, presumably by the mechanism analogous to endothelial cells discussed above. These observations seem to indicate that the signals mediated presumably by inhibition of phosphatases work in the direction opposite to the growth-stimulating signal of AT₁. However, paradoxical observation exists in which AT₁ and AT₂ function in parallel. An increasing number of publications reported that chronic infusion of the AT₂-specific blocker PD123319 was able to inhibit chronic Ang II-induced aortic smooth-muscle hypertrophy and fibrosis despite continuously elevated blood pressure. Despite a short half-lifetime of PD123319 *in vivo*, chronic infusion by osmotic mini-pump suppressed the Ang II-infused aortic hypertrophy. AT₂ gene-deleted mice also showed marked resistance against left ventricular hypertrophy induced by pressure overload or chronic Ang II infusion. Another example of parallel phenotypic effect of AT₁ and AT₂ was found in stimulation of tyrosine hydroxylase by both AT₁ and AT₂ in adrenal chromaffin cells *in vivo*.

Since all of the paradoxical results were obtained under well-controlled conditions, signaling mechanisms in specific tissues were investigated. It was found that the AT₂ binds a zinc finger protein and serves as a transporter to deliver the zinc finger protein to the nucleus, which activated transcription of phosphatidyl inositol 3-kinase p85 α subunit (p85 α PI3K). This triggers a subsequent cardiac hypertrophic pathway. These results indicate that the AT₂-mediated cardiac hypertrophy uses signals distinct from those which are activated by AT₁. The parallel phenotypic results were due to tissue

specific expression of the zinc finger protein in the heart but not in the kidney, where the AT_2 signal depends on the kallikrein–NOS–cGMP mechanism. In contrast, the cardiac tissue uses a more direct activation mechanism of nuclear transcription of p85 α PI3K. The p85PI3K activation has been shown to induce a variety of cardiac hypertrophy mechanisms.

Major Physiological Roles of AT_1 and AT_2

Targeted gene deletion of AT_{1A} , AT_{1B} , or angiotensinogen in mice showed that angiotensin II has a profound effect on the maintenance of blood pressure; dual deletion of AT_{1A} and AT_{1B} resulted in hypotension by 45–50 mmHg and no pressor response to Ang II infusion. AT_{1B} was responsible for only 10% of the pressor effect. Interestingly, these animals showed profound defects in the renal papillary and medullary formation suggesting defects in Na reabsorption, an idea in agreement with the polyurea and increased Na excretion. These animals were stunted and it was difficult to keep them alive without daily saline injection. AT_2 gene deletion did not cause overt physical deficiency. However, closer examination revealed that AT_2 -deficient mice show high frequency (28% penetrance) of uretero-renal pelvis ligation that is closely analogous to a human neonatal disease called Kakkut syndrome. Connection to the ureter of the renal pelvis is the last step of the kidney morphogenesis and AT_2 concentration is seen in the renal pelvis at this stage in a normal fetus.

Recently, Srivastava's group reported a very intriguing observation in which they identified families with severe mental retardation and traced their gene defect to the absence or 'mis-sense' mutation of AT_2 gene on the X chromosome. Thus, AT_2 function is not only limited to the renal, adrenal, and cardiovascular system, but brain function also depends on AT_2 . Indeed, AT_2 gene-deficient mice show a pattern of fearful behavior.

Angiotensin (1–7)

Functions

Ferrario and his associate reported physiological functions of Ang (1–7). Since the heptapeptide seemed to be generated by one step of carboxypeptidase, it was not clear whether it was a part of Ang II function. However, it was shown that Ang (1–7) is generated from Ang I by an ACE2 and ACE as in the updated [Figure 1](#). Since the affinity of the heptapeptide binding to the brain plasma membrane was in the nanomolar range, it seemed to have a physiological significance. Its specific action was shown as arachidonate release and prostaglandin synthesis. Its systemic presence and function were also demonstrated. Decisive evidence of Ang (1–7) as a hormonal peptide came as the receptor was identified.

Ang (1–7) receptor

In 1988, Hunley and colleagues presented the hypothesis that *Mas* oncogene product was the Ang II receptor. This hypothesis failed to meet the test of many investigators. AT_1 and AT_2 were cloned soon afterward as unmistakable Ang II receptors. However, the *Mas* oncogene story made a full circle recently when Santos, Walther, and their collaborators showed that the *Mas* oncogene product expressed in COS7 cells binds Ang (1–7) with a nanomolar K_D value and releases

prostaglandins in an abstract form. It has a structural feature of a seven-transmembrane-domain receptor. Further details are awaited.

Ang IV

Harding and colleagues showed that the hexapeptide Ang IV binds to receptors in a variety of tissues and increases blood flow. Recently, Xie, Mendelsohn, and their collaborators purified Ang IV receptor from adrenal plasma membrane and succeeded in cloning the cDNA. They found that it is identical with insulin-regulated amino peptidase (IRAP) with a single transmembrane domain. Its possible interaction with insulin is of potential interest as an interface between angiotensin and insulin. However, Ang IV is not the only ligand that binds the extracellular domain. LVV hemorphin, a segment of hemoglobin, was also reported to be a binding ligand. AT_4 is expressed heavily in the hippocampus and signals to stimulate acetylcholine. Its potential role in memory retention is of great interest in relation to a possible regulatory role of this aminopeptidase. Ang II also was shown to be involved in the regulation of long-term potentiation (LTP). Thus, the roles of angiotensin are not limited to the control of blood pressure, water drinking, and dipsogenesis in the hypothalamus and brain stem. AT_4 expressed in hippocampus, AT_2 expressed in central amygdala nucleus, and Ang II stimulation of LTP, point to the roles of angiotensins in normal cortical function.

AT_3

We reported non- AT_1 –non- AT_2 Ang II receptors in neuro 2A cells were expressed at room temperature, but not at 37 °C. Therefore, we consider it as a likely product of mycoplasma that infected the neuro 2A cell line we used. Thus, at present, AT_3 is not of mammalian origin.

See also: **Signaling: Src Family of Protein Tyrosine Kinases.**

Further Reading

- Berk BC and Corson MA (1997) Angiotensin II signal transduction in vascular smooth muscle, role of tyrosine kinase. *Circulation Research* 80: 607–616.
- deGasparo M, Catt KJ, Inagami T, and Harding J (2000) The angiotensin II receptors. *Pharmacological Reviews* 52: 639–672.
- Eguchi S and Inagami T (2000) Signal transduction of angiotensin II type 1 receptor through receptor tyrosine kinase. *Regulatory Peptides* 91: 13–20.
- Ferrario CM and Iyar SM (1998) Angiotensin (1–7): A bioactive fragment of the renin–angiotensin system. *Regulatory Peptides* 78: 13–18.
- Gallinat S, Busche S, Raizada MK, and Sumners C (2000) The angiotensin II type 2 receptor: An enigma with multiple variations. *American Journal of Physiology* 278: E357–E374.
- Horiuchi M, Akishita M, and Dzau VJ (1999) Recent progress in angiotensin II type 2 receptor research in the cardiovascular system. *Hypertension* 33: 613–621.
- Inagami T (1999) Molecular biology and signaling of angiotensin receptors: An overview. *Journal of the American Society of Nephrology* 10: 52–57.
- Matsubara H (1998) Pathophysiological roles of angiotensin II type 2 receptor in cardiovascular and renal diseases. *Circulation Research* 83: 1182–1191.
- Timmermans PBMWM, Wong PC, Chiu AT, et al. (1993) Angiotensin receptor and angiotensin II antagonists. *Pharmacological Reviews* 45: 205–251.
- Touyz RM and Schiffrin E (2000) Signal transduction mechanisms mediating the physiological and pathological actions of angiotensin II in vascular smooth muscle cells. *Pharmacological Reviews* 52: 639–672.
- Wright JW, Krebs LT, Stobb JW, and Harding JW (1995) The angiotensin IV system: Functional implication. *Frontiers in Neuroendocrinology* 16: 23–52.

Anorexigenic and Orexigenic Gut Peptides

P S Petersen, University of Copenhagen, Copenhagen, Denmark

T H Moran, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Anorexigenic Feeding inhibitory.

Cephalic Actions stimulated by the site, smell, and taste of food prior to contact with gastrointestinal sites.

Orexigenic Feeding stimulatory.

Satiation Negative feedback process stimulated by ingested food to inhibit eating.

Satiety State of satisfaction following ingestion.

Introduction

There are many inputs from the periphery to the brain providing information regarding the energy status of the body. Energy intake is registered through taste receptors, via distension of the stomach, and through nutrient-induced alterations in the release of gut hormones that potentially affect energy intake, either acutely or over the longer term. These gut hormones generally act in the hindbrain through either vagal afferents or directly, and by altering neuronal activity in the hypothalamic appetite centers. The first gut hormones were discovered more than 50 years ago; yet, the function and potential of these peptides in modulating energy balance is not completely clear, and the picture gets more complex as new hormones are discovered. In this article is provided a brief introduction to the most thoroughly studied gut hormones in relation to satiety and food intake.

Anorexigenic Gut Hormones

Cholecystokinin

Cholecystokinin (CCK) is a peptide produced throughout the gastrointestinal tract, but especially in the enteroendocrine I cells of the duodenum and jejunum. It is secreted in response to the presence of nutrients in the gut lumen. Fatty acids containing more than a 12-carbon chain are particularly effective in stimulating CCK secretion. The anorexic effect of CCK, which was first discovered by Gibbs and colleagues in 1972, has been the most-studied feeding effect of any gut peptide and in many ways has become the model against which the effects of other gut peptides are measured. Exogenous CCK administration results in dose-dependent reductions in food intake across multiple species, including man, and in multiple testing situations. CCK-induced reductions in food intake can occur without reported side effects. Meal-contingent CCK administration results in reductions in meal size and repeated meal-contingent administration continues to reduce meal size, but results in a compensatory increase in meal frequency such that overall food intake is unchanged.

Roles for CCK1 (also called CCK A receptors) in the pyloric sphincter and on vagal afferent fibers in the satiety action of CCK have been suggested. These receptors are also widely distributed in brain areas such as the nucleus tractus solitarius area postrema (AP), hypothalamic ventromedial nucleus and dorsomedial nucleus sites at which central CCK may also affect food intake.

In rats and nonhuman primates, antagonism of the CCK1 receptor blocks the satiating effect of exogenous CCK, and, when given alone, CCK1 receptor antagonists increase meal size and in some situations, overall food intake. Results of assessments of the satiety actions of endogenous CCK in human studies have been mixed. CCK doses that lead to a plasma CCK peak of similar magnitude as seen with meal-stimulated release have had mixed effects on food intake, but have resulted in reliable changes in reports of decreased hunger and increased fullness. Another approach clearly supports CCK's role as a satiety factor in humans. Intraduodenal perfusions of long-chain fatty acids lead to reduced calorie intake and increase in plasma CCK concentration. Administration of the CCK1 receptor antagonist lorglumide intravenously at the same time as the intraduodenal infusion abolishes the satiety effect. Further implicating physiological relevance of CCK in satiety is the impaired meal stimulation of CCK secretion in bulimic patients.

The importance of the CCK1 receptor is also seen in the Otsuka Long Evans Tokushima Fatty (OLETF) rat, which completely lacks this receptor. Early on, these rats develop obesity, hyperglycemia, and impaired glucose regulation that progresses into noninsulin-dependent diabetes and finally into severe insulin-dependent diabetes. The same phenotype is not seen in the CCK1 receptor knockout (KO) mice, indicating species differences or that an additional impairment may be implicated in the phenotype seen in the OLETF rats.

Peptide YY

Peptide YY (PYY) is a member of the family of peptides that includes neuropeptide Y (NPY) and pancreatic polypeptide (PP). There is about 70% homology among the three peptides and the peptide sequences are highly conserved between species. Furthermore, they have the same tertiary structure, namely the characteristic PP fold consisting of an extended pyroproline helix and an α -helix held together by β -turns. All three peptides have been associated with food intake.

PYY is expressed in the enteroendocrine L-cells primarily in the ileum and colon. These cells also express glucagon-like peptide-1 (GLP-1) and oxyntomodulin (OXM). PYY binds with high affinity to all Y-receptor subtypes but PYY is cleaved by dipeptidyl peptidase yielding the circulating acid form of PYY referred to as PYY₃₋₃₆. PYY₃₋₃₆ binds with high affinity to Y₂ but with lower affinity to Y₁ and Y₅ receptors.

The plasma concentration of PYY₃₋₃₆ increases shortly after nutrient intake and even before nutrients reach the PYY-producing cells. The plasma levels peak around 90 min after meal initiations, and in contrast to the more short-term acting peptides, such as CCK, the levels stay elevated for up to 6 h.

Intravenous administration of PYY₃₋₃₆ to normal weight and obese human subjects has effects on appetite resulting in up to 30% decrease in food intake. The reduction in calorie intake is accompanied by a reduction in subjective hunger scores. After the infusion is stopped, the anorectic effect has been observed to last for up to 12 h, even after plasma PYY₃₋₃₆ returns to normal levels.

The decrease in food intake that is seen when administering PYY peripherally is considered to be mediated both through delayed gastric emptying and by changes in central feeding circuits. In the arcuate nucleus of the hypothalamus, PYY₃₋₃₆ leads to inhibition of the orexigenic AgRP/NPY-neurons by binding to presynaptic Y₂ receptors.

The actions of endogenous PYY in inhibiting feeding remain to be completely determined. When infusing PYY intravenously in human subjects, in amounts resulting in peak concentrations comparable to the physiological peak seen after meal initiation, no effect on food intake and satiety has been obtained. Furthermore, although Y₂ antagonists have been shown to block the feeding inhibitory actions of exogenous PYY, they have not been shown to increase food intake when given alone. This does not rule out that PYY is an anorectic hormone, but it may imply that the physiological effect of PYY should not be considered alone. GLP and PYY are cosecreted following meal initiation. Co-administering these peptides in low concentrations has an effect on food intake, which is not seen when giving either peptide separately at the same dose. This additive effect has been observed in both rodents and humans.

Apolipoprotein A-IV

Apolipoprotein A-IV (APO-IV) is primarily synthesized in the small intestine, where secretion is stimulated by lipid absorption. Plasma APO-IV level peaks 15 min after fat ingestion and the level remains elevated for about 3 h. In rodents, there is a circadian fluctuation of plasma APO-IV determined by the pattern of food intake with highest levels in the beginning of the dark phase.

During prolonged ingestion of a high-fat diet, APO-IV levels increase in both human and rat. APO-IV is also produced within the hypothalamus, and intracerebroventricular (ICV) administration of APO-IV antibody initiates food intake, leading to the suggestion that APO-IV affects food intake by inhibiting meal initiation rather than affecting meal size.

Supporting a physiological satiety role of APO-IV is the dose-dependent decrease in food intake seen when administering the protein both through the central and the peripheral route. This effect is seen with doses in the physiological range.

Incretins

The incretins are a group of anorexigenic hormones defined by their ability to stimulate glucose-dependent insulin secretion and by the fact that they are released in response to nutrient

intake. This phenomenon was originally discovered through the observation that more insulin is secreted after oral ingestion of glucose than after intravenous glucose injection. Therefore, something in the gut was thought to be involved. The two most important hormones in this group are GLP-1 and gastric inhibitory polypeptide (GIP).

Glucagon-Like Peptide-1

There are several tissue-specific cleavage products of the proglucagon mRNA. In the intestine and brainstem, the cleavage results in OXM, glicentin, GLP-1, and GLP-2 whereas in the pancreatic α -cells the major product is the glucagon hormone. The differential production of the gene products is regulated by tissue/cell-specific expression of prohormone convertases 1 and 2.

GLP-1 is a potent incretin. Administration of the hormone leads to decreased glucose levels and augmented glucose-stimulated insulin release in diabetic patients. GLP-1 directly stimulates insulin release and increases β -cell glucose sensitivity both by binding to GLP-1 receptors on the β -cells and by indirect neural activation.

The active forms of GLP-1 consist of 29–30 amino acids either as the amidated form GLP-1 (7–36) amide or as the glycine-extended form GLP-1 (7–37), which is the most abundant form in humans. GLP-1 is secreted throughout the gastrointestinal tract from L-cells predominantly in the distal part of the jejunum, ileum, and the colon. Basal fasting human GLP-1 plasma levels are around 5–10 pM. The level increases in response to meal ingestion, and the first increase is observed within 15 min of meal initiation and peaks within 60–90 min. After peaking, the level quickly returns to basal. The magnitude of the rise in plasma GLP-1 follows the amount of calories ingested and the composition of the meal. The protease dipeptidyl peptidase-IV (DPP-IV) cleaves peptides that have an alanine at position 2. This leads to very short half-lives of systemic active GLP-1, GLP-2, and OXM of less than 7 min.

A direct central role of GLP-1 in feeding control is plausible as GLP-1 binds to several brain areas involved in food intake, and injection of GLP-1 directly in the paraventricular nucleus (PVN) has been shown to decrease food intake in rats. The central binding and activation could be mediated through activation of GLP-1 production in the neurons within the hypothalamus.

Several studies indicate that systemic GLP-1 acts to affect food intake through peripheral GLP-1 receptors. An anorexic effect is seen with peripheral administration of the combined GLP-1-albumin protein that does not cross the blood–brain barrier (marketed as Liraglutide). The effect of peripheral GLP-1 could be mediated via the vagal nerve as GLP-1 increases firing of vagal afferents and this increased firing is inhibited by exendin 9–39. In addition, high expression of GLP-1 receptors in the nodose ganglion and vagal terminals supports vagal signaling.

Studies using the competitive GLP-1 receptor antagonist exendin 9–39 confirm that the effect of GLP-1 on feeding may be partly central and partly peripheral mediated. Systemic administration of exendin 9–39 prevents the anorexic effect of systemic GLP-1. The same is the case for the central route where ICV injection of exendin 9–39 inhibits the anorexic effect of ICV GLP-1. However, central exendin 9–39 does not prevent the anorexic effect of systemic GLP-1, supporting distinct pathways. Disruptions of vagal signaling have been shown to

decrease the anorexic effect of the peripheral GLP-1, further supporting dual mechanisms.

Exogenous delivery of GLP-1 induces satiety and decreases food intake in normal weight, obese, and diabetic humans without any side effects. This effect is sustained with lower physiologically relevant doses. The anorectic effect of the hormone is abolished by administration of exendin 9–39, which will even lead to increased food intake under certain circumstances. These data suggest that GLP-1 is a physiological regulator of satiety and food intake.

The short half-life of GLP limits the therapeutic potential of the hormone itself. Effort has therefore been put into developing long-acting GLP-1 agonist, and this enables to not only decrease glucose levels in diabetic patients but also decrease food intake and body weight. Liraglutide has been shown to decrease body weight when administered daily in humans and the first longer-lasting GLP-1 analog (exendin-4) is available in the market for diabetes treatment (Exenatide).

Oxyntomodulin

OXM consists of 37 amino acids including the entire sequence of glucagon with the addition of a basic C-terminal extension, the so-called 8-amino-acid spacer peptide-1. OXM got its name because the first effect seen of the peptide was the inhibition of the oxyntic glands of the stomach. OXM is co-released postprandially with GLP-1 and PYY from intestinal L-cells. An anorectic effect of OXM was discovered when the hormone was peripherally administered to rodents, and this effect has been reproduced in humans. The effect is blocked by prior injections of the GLP-1 receptor antagonist exendin 9–39 and the effect is lost in GLP-1 receptor KO mice.

Similar to GLP-1, OXM is an incretin although not as potent as GIP and GLP-1. This correlates with the lower affinity of OXM for the GLP-1 receptor. In contrast, the anorectic properties of OXM are as potent as those of GLP-1, indicating that separate mechanisms are involved in the incretin and feeding effects of the peptide. Part of the anorexic effect of OXM is explained by 20–40% decrease in ghrelin levels that may be measured after OXM administration in both rodents and humans. Moreover, increased energy expenditure following administration contributes to the weight loss seen in humans.

Central pathways could be involved in the anorexic effects of OXM as it increases release of anorexic neuropeptide α -melanocyte-stimulating hormone from hypothalamic explants, and direct injections of OXM into the PVN affects food intake. This action is likely to occur through the GLP-1 receptor as exendin 9–39 administration into the arcuate nucleus has been shown to prevent the OXM-mediated decrease in energy intake.

The physiological plasma level of OXM is increased in gastric bypass patients. This may be part of an overall gastric hormonal shift, which could contribute to the fast weight loss and improved glucose regulation seen in these patients.

Gastric Inhibitory Polypeptide

GIP is a 42-amino-acid gastrointestinal hormone. It is a member of the incretin family mediating increases of insulin secretion prior to the postabsorptive nutrient-induced increase in blood glucose. GIP is secreted primarily from the stomach and

the duodenum K cells in response to ingestion. The release follows the rate of intake and absorption of, especially, fat and glucose. GIP is also a substrate for DPP-IV cleavage leading to a very short half-life of the active peptide of only 5–7 min. Besides the incretin effect of GIP, it has also been shown to enhance fatty acid synthesis and triglyceride storage in adipocytes. Administration of GIP does not have an acute effect on food intake, nor does the KO of the receptor. However, GIP receptor KO mice are protected against diet-induced obesity and have reduced fat mass. Together with data demonstrating that GIP KO mice have increased energy expenditure, this suggests a physiological role of GIP in maintaining long-term energy homeostasis.

Pancreatic Peptides

Pancreatic Polypeptide

PP is secreted by specific cells in the pancreatic islets. The amount of peptide released is dependent upon calorie intake and the composition of the meal. Peripheral infusion of PP reduces food intake only after 2 h and the effect is prolonged over the following 24 h in humans. The same effect is seen after peripheral administration in rodents, while central administration increases food intake. These contradictory effects, which are also seen with PYY, may be caused by differential stimulation of the Y-receptor subtypes. PP does not seem to cross the blood-brain barrier and the anorexic effect is believed to be largely mediated through vagal or circumventricular organ sites. After peripheral administration, autoradiography studies have shown accumulation of PP in AP. This area has a high expression of Y_4 receptors, suggesting that the anorectic effect of PP is mediated by this receptor. Similar to the Y_2 receptors, the presynaptic Y_4 receptors in the brain have an autoinhibitory effect on NPY neurons. A role for the vagus is also supported by studies showing diminished anorexic effect of peripherally administered PP in vagotomized animals, and further by a high concentration of Y_4 receptors in the dorsal vagal complex, the site of central vagal afferent terminations. The role of endogenous PP in feeding control is yet to be investigated.

Glucagon

The most prominent role of glucagon is to stimulate glycogenolysis and gluconeogenesis to maintain euglycemia during fast or in the mode of rapid energy expenditure. However, glucagon is also released from the pancreatic α -cells when food is ingested. The meal-related glucagon release is brief and initiated within the first few minutes of meal initiation. This is most likely cephalic stimulated. The rise in systemic glucagon is modest but the portal-vein glucagon concentration can be relatively high without change in overall systemic levels.

Exogenous glucagon reduces food intake most efficiently and is dose dependent when administered through hepatic portal infusion. These infusions decrease meal size without affecting overall meal number in rodents and humans, and do so at physiological concentrations. The feeding effect is prevented by afferent vagotomy. Intraperitoneal and hepatic portal injection of a specific polyclonal glucagon antibody increases meal size and hepatic glucose output suggesting a

role for endogenous glucagon in feeding control. This effect is not seen when injected in the vena cava, suggesting that the feeding effect of glucagon is mediated in the liver.

Insulin

The most prominent function of insulin is mediating glucose uptake in peripheral tissues and thereby maintaining euglycemia. The level of insulin is low during fasting. During and after a meal, insulin is released in a biphasic manner. Insulin also acts as an adiposity factor with basal levels being correlated with body fat. Insulin is transported across the blood–brain barrier and there is a high concentration of insulin receptors in the arcuate nucleus of the hypothalamus. In addition, insulin increases the brain's sensitivity to other satiety factors and thereby relates body adiposity to food intake. ICV insulin injection reduces food intake in a dose-dependent manner and insulin antibodies have the opposite effect leading to increased food intake.

Amylin

Amylin is cosecreted with insulin from the pancreatic β -cells. Similar to insulin, amylin levels are low during fasting and increase quickly during eating, in proportion to the meal size. The overall level of amylin is also proportional to the amount of body fat. Exogenous systemic delivery of amylin, prior to a meal, decreases meal size in rats. This is mediated through activation of amylin-binding sites in the AP, which are accessible through the relatively permeable blood–brain barrier of this circumventricular organ. Evidence for an AP site of action also derives from lesions studies demonstrating that AP lesions prevent the anorexic effect of exogenous systemic amylin. Actions in other brain areas are also implicated as low-dose ICV amylin produces potent long-lasting decrease in feeding. Supporting a physiological role for amylin in feeding control are data demonstrating increase in food intake in response to amylin receptor antagonism.

Enterostatin

Procolipase is synthesized and released from the pancreas in response to lipid intake. In the intestinal lumen, procolipase is cleaved to active colipase and enterostatin. Enterostatin is a 41-amino-acid protein and part of the sequence is highly conserved in mammal species. Plasma enterostatin increases significantly after ingestion of a lipid-containing meal. Exogenous administration of enterostatin reduces food intake and especially the intake of fat or high-fat foods. An iv dose of enterostatin giving a similar rise in plasma enterostatin to that found after ingesting a high-fat meal inhibits food intake, thereby supporting a physiological function. An inhibition of food intake is also obtained when administering the peptide ICV.

Orexigenic Gut Hormone

Ghrelin

Ghrelin is a 28-amino-acid enteric peptide. It is a natural ligand for the 7-trans-membrane growth hormone secretagog

receptor (GHS-R 1a), also known as the ghrelin receptor. The acylation of the third residue is necessary for ghrelin to activate the ghrelin receptor and for the peptide to cross the blood–brain barrier. The acylation is catalyzed by the gastric enzyme ghrelin O-acyltransferase. Eighty to ninety percent of circulating ghrelin is in the nonacylated form, which is not able to bind to and activate GHS-R 1a, and therefore has no effect on growth hormone release and food intake.

Ghrelin is mainly secreted from X/A-like cells in the oxyntic glands of the gastric fundal mucosa, but it is also present in the bowel, especially the small intestine with decreasing density of ghrelin-positive cells from the stomach throughout the gut.

There are diurnal changes in ghrelin levels in humans and rodents. During fasting, levels are high with a peak just before meal initiation and then a decline after food intake, mediated by the presence of nutrients in the gut. The postprandial suppression of ghrelin is proportional to the load of ingested calories. Ghrelin levels not only change in response to acute changes in nutritional status but also are affected by chronic metabolic changes (obesity, etc.). Fasting plasma ghrelin levels are lower in obese subjects compared to normal weight subjects, while subjects with low body mass index (BMI) (such as patients suffering from anorexia nervosa or cachexia) have elevated ghrelin secretion. These abnormalities in ghrelin secretion are reversed upon normalization of body weight. The only known exception to this negative correlation between ghrelin level and BMI is in the Prader–Willi syndrome, where obesity is correlated with hypersecretion of ghrelin.

Currently, ghrelin is the only peripheral hormone known to stimulate food intake. Peripheral administration of ghrelin increases feeding in rodents and humans, and continuous administration increases body weight, not only through hyperphagia but also through increased lipogenesis and decreased energy expenditure.

Ghrelin's feeding stimulatory actions are thought to depend on ghrelin binding to hypothalamic NPY/AgRP neurons and increasing the activity of these orexigenic neurons by a direct presynaptic effect. This binding simultaneously has an inhibitory effect on proopiomelanocortin neurons mediated by increased frequency of spontaneous GABA-ergic inhibitory postsynaptic currents, leading to a decrease in the spontaneous activity of these anorexigenic neurons. Ghrelin can access the brain via areas not protected by the blood–brain barrier and by transport across the barrier. Furthermore, a lesser production of ghrelin within the hypothalamus is possible, as mRNA has been detected there.

There is no clear evidence showing that the orexigenic effect of ghrelin is vagally mediated. However, the ghrelin receptor is expressed in the nodose ganglion and transport to the nerve terminal has been described, indicating that vagal afferents may be involved in mediating some of ghrelin's other functions.

Chronic administration of ghrelin-receptor antagonists leads to a reduction in food intake and a reduction of body weight in normal weight and diet-induced obese rodents supporting a role for endogenous ghrelin in feeding control. In ghrelin and ghrelin-receptor KO mice, there is a shift in

preferred energy substrate toward fat, and these mice are resistant to early onset diet-induced obesity.

Conclusion and Perspective

Evolutionarily, there is a high basal drive for seeking and eating food, which makes sense in an environment where food is scarce. In our modern society, however, where high-energy food is abundant and lifestyles are more sedentary, this drive may contribute to the obesity epidemic seen worldwide. Globally, 400 million people were considered to be obese in 2006, based on their BMI. Obesity has severe health implications and treating these is very expensive. Therefore, there is great interest in exploring the appetite-regulating system and developing potential therapeutic agents to decrease food intake and body weight. Although much is still to be learned about the gut–brain communication, we now know that gut hormones are fundamental in physiological body-weight regulation. Great efforts are therefore being applied to fully understanding the interacting mechanisms relaying information from the gut to the brain. Limited or short-lasting decrease in food intake and body weight has been seen with many gut hormones or related agonist and antagonist compounds. The effects have been improved by chemically increasing the bioavailability of the compounds. However, more recent studies show greater effect by combining the effects of several hormones, a strategy that is already being applied and appears to be promising for the development of effective gut-hormone-derived weight-reducing therapeutics.

See also: **Metabolism Vitamins and Hormones:** Diabetes; Insulin- and Glucagon-Secreting Cells of the Pancreas; Metabolic Roles of the Orexigenic and Anorexigenic Neuropeptides; **Signaling:** Melanocortin System.

Further Reading

- Baggio LL and Drucker DJ (2004) Clinical endocrinology and metabolism. Glucagon-like peptide-1 and glucagon-like peptide-2: Best practice and research. *Clinical Endocrinology and Metabolism* 18: 531–554.
- Banks WA (2008) The blood–brain barrier: Connecting the gut and the brain. *Regulatory Peptides* 149: 11–14.
- Beglinger C and Degen L (2006) Gastrointestinal satiety signals in humans: Physiologic roles for GLP-1 and PYY? *Physiology and Behavior* 89: 460–464.
- Geary N (2004) Endocrine controls of eating: CCK, leptin, and ghrelin. *Physiology and Behavior* 81: 719–733.
- Moran TH (2006) Neural and hormonal controls of food intake and satiety. In: Johnson LR (ed.) *Physiology of the Gastrointestinal Tract*, pp. 877–894. San Diego, CA: Academic Press.
- Nauck MA (2009) Unraveling the science of incretin biology. *American Journal of Medicine* 122: S3–S10.
- Neary MT and Batterham RL (2009) Gut hormones: Implications for the treatment of obesity. *Pharmacology and Therapeutics* 124: 44–56.
- Tso P and Liu M (2004) Apolipoprotein A-IV, food intake, and obesity. *Physiology and Behavior* 83: 631–643.
- Vincent RP, Ashrafian H, and le Roux CW (2008) Mechanisms of disease: The role of gastrointestinal hormones in appetite and obesity. *Nature Clinical Practice. Gastroenterology and Hepatology* 5: 268–277.
- Williams DL (2009) Finding the sweet spot: Peripheral versus central glucagon-like peptide 1 action in feeding and glucose homeostasis. *Endocrinology* 150: 2997–3001 (mini-review).

Aquaporin Water Channels

K Liu and P Agre, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Activation energy The minimum amount of energy required (for water molecules to come across aquaporin channels).

Hydropathy analysis A technique which uses a scale (containing some sort of free energy-derived value for each amino acid) windowing to determine the hydrophobic nature of a protein sequence.

Immunogold electron microscopy A form of electron microscopy in which gold particles are used to label

antibodies so that the protein of interest can be localized by localizing gold particles.

Loss-of-function mutations The change in the DNA sequence of a gene with gene product having less or no function.

Phosphorylation The addition of a phosphate group (PO_4^{3-}) to an amino acid of a protein.

Water homeostasis A state of water equilibrium.

Introduction

Every living organism is primarily water, and approximately 70% of the human body mass is also water. Thus, the distribution and movement of water through tissues are fundamental to life. The surface of our brains is protected by cerebrospinal fluid. The orbits of our eyes are filled with aqueous humor, and the exterior surface by a thin film of tears. We sweat during exercise to maintain normal body temperature. Every day our kidneys absorb 99% of water from the glomerular filtrate, concentrating urine, and thereby protecting us from dehydration.

Water can pass across biological membranes (lipid bilayers) by simple diffusion relatively slowly. However, over the past several decades, multiple research groups have reported that membranes of certain tissues have very rapid water permeability, such as mammalian red cells and amphibian bladder. No pharmacologic inhibitors of diffusion are known. However, an important experiment in 1970 established that water transport into red cells is inhibited by mercurials and restored by chemical reducing reagents. These observations indicated that water-specific channels must be proteins with free sulfhydryls and characteristically low Arrhenius activation energy (Figure 1).

Discovery of Aquaporin-1

Although sought by multiple groups of investigators, water-transporting molecules remained elusive for decades, and most scientists were dubious that they even existed. Interest in water channels dramatically rose after the publication of the functional description of a protein now known as aquaporin-1 (AQP1). By using *Xenopus laevis* oocytes to express AQP1, dramatic explosions were viewed microscopically after transferring oocytes to hypotonic solutions (Figure 2), the first demonstration of molecular water channels.

AQP1-mediated water permeability is distinct from simple diffusion. First, water movement through AQP1 has higher capacity and lower activation energy ($E_a < 5 \text{ kcal mol}^{-1}$) than diffusion

($E_a > 10 \text{ kcal mol}^{-1}$). Second, water but not protons (H_3O^+) pass through the AQP1 channel pore. Subsequent experiments revealed that water transport by AQP1 is inhibited by mercurials, is restored by reducing agents, such as β -mercaptoethanol, and lacks measurable membrane currents – all features consistent with the previous predictions of water channels.

Following discovery of AQP1, multiple research groups joined the quest and isolated homologous proteins from plant, animal, and microbial tissues with high water transport, and the field grew rapidly.

Structure of AQP1

Hydropathy analysis of the deduced AQP1 protein sequence revealed a novel topology—two tandem repeats, each formed by three transmembrane domains, oriented at 180° with respect to each other in the lipid bilayer. In this unique symmetry, loops B and E each carries a signature asparagine–proline–alanine (NPA) motif highly conserved within the AQP family (Figure 3, upper panel). Cysteine 189 (Cys-189), proximal to the second NPA motif, is the mercurial inhibition site. By mutagenesis it was established that six bilayer-spanning domains surround a central pore containing loop B, dipping into the bilayer from the cytoplasmic surface, and loop E, dipping into the bilayer from the extracellular surface (Figure 3, lower panel). Resembling the ancient timepiece, this model was termed as the ‘hourglass’ and was later confirmed by high-resolution structural analysis.

Efforts were directed to solving the AQP1 structure because of its fundamental role in physiology. Structures of functional human AQP1 were first achieved using electron crystallography by the Engel and Fujiyoshi groups at $3.8\text{-}\text{\AA}$ resolution. The structures of *Escherichia coli* glycerol facilitator (GlpF) and bovine AQP1 were subsequently determined at $2.2\text{-}\text{\AA}$ resolution by X-ray crystallography. Although present as a tetramer in the cell membrane, each AQP1 monomer contains its own pore and works independently. This organization is distinct from most ion channels, where four subunits surround a central pore. Four residues, Phe-56, His-180, Cys-189 (the

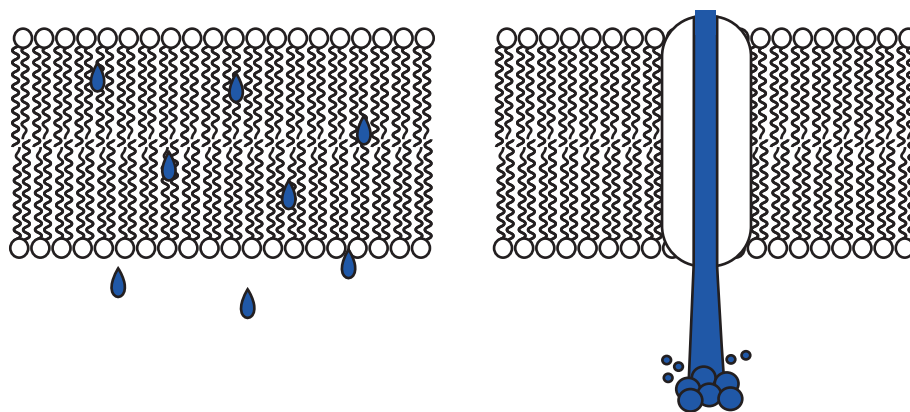


Figure 1 Transmembrane water permeation through a lipid bilayer by simple diffusion (left) compared to permeation through a water channel (right).

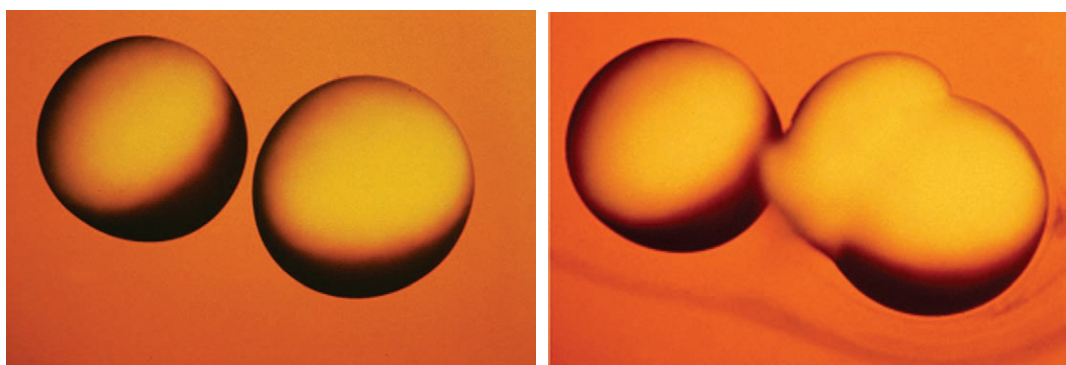


Figure 2 Functional expression of AQP1 water channel in *X. laevis* oocytes. (Left panel) A control oocyte (left) and an AQP1-expressing oocyte (right) in isotonic solution. (Right panel) The AQP1-expressing oocyte explodes after transferred to hypotonic buffer. The control oocyte has swollen negligibly, whereas the AQP1-expressing oocyte has exploded due to the fast water influx along the osmolality gradient. From Preston GM, Carroll TP, Guggino WB, and Agre P (1992) Appearance of water channels in *Xenopus* oocytes expressing red cell CHIP28 protein. *Science* 256: 385–387.

mercury sensitive residue), and Arg-195, form the 2.8-Å constriction of the pore. The narrowest part of GlpF is about 1 Å wider than AQP1, providing sufficient space for the larger glycerol molecule to pass through.

These structures also shed light upon the mechanism of proton (H_3O^+) exclusion. In solution, both H_2O and H_3O^+ exist. Water molecules linked to each other by hydrogen bonds allow protons to jump rapidly along a column of water, referred to as a 'proton wire'. Arg-195, the highly conserved residue in the aquaporin family, provides a positive charge at the narrowest pore area of the channel. His-180 becomes protonated at lower pH and provides a second positive charge. These two residues, together with the two asparagines residues in NPA motifs (Asn-76 and Asn-192 in human AQP1), effectively prevent proton passage by electrostatic repulsion and water dipole reorientation (Figure 4). The structure of AQP1 is ideally suited to its biological functions. This mechanism allows human kidneys to reabsorb water while excreting acid – necessary to maintain hydration while preventing systemic acidosis. Moreover, the membrane potential in plants is derived from the pH gradient that exists in spite of rapid transmembrane water movements.

Physiological Functions of AQP1

Approximately 180 l of plasma are filtered by our kidneys every day to eliminate toxic wastes and maintain the appropriate blood pH and ionic composition. By light microscopy and immunoelectron microscopy, AQP1 was visualized in renal proximal convoluted tubules and descending thin limbs of the loop of Henle, segments with constitutively high water permeability. In contrast, the ascending segments are permeable to ions but not water and lack AQP1 proteins. Furthermore, AQP1 protein is present only in the apical and basolateral plasma membranes of these tissues and not in intracellular membranes. These studies provide clear evidence for the physiological and pathological roles of AQP1 – constitutively transporting large amounts of water from tubular lumen to the interstitium and then into the vascular space with the driving force provided by small, standing osmotic gradients created by solute transporters (Figure 5). AQP1 was also found in other tissues with important secretory roles such as nonpigmented epithelium in anterior compartment of eye (aqueous humor), cholangiocytes (bile), and capillary endothelium in many organs including bronchial circulation of lung.

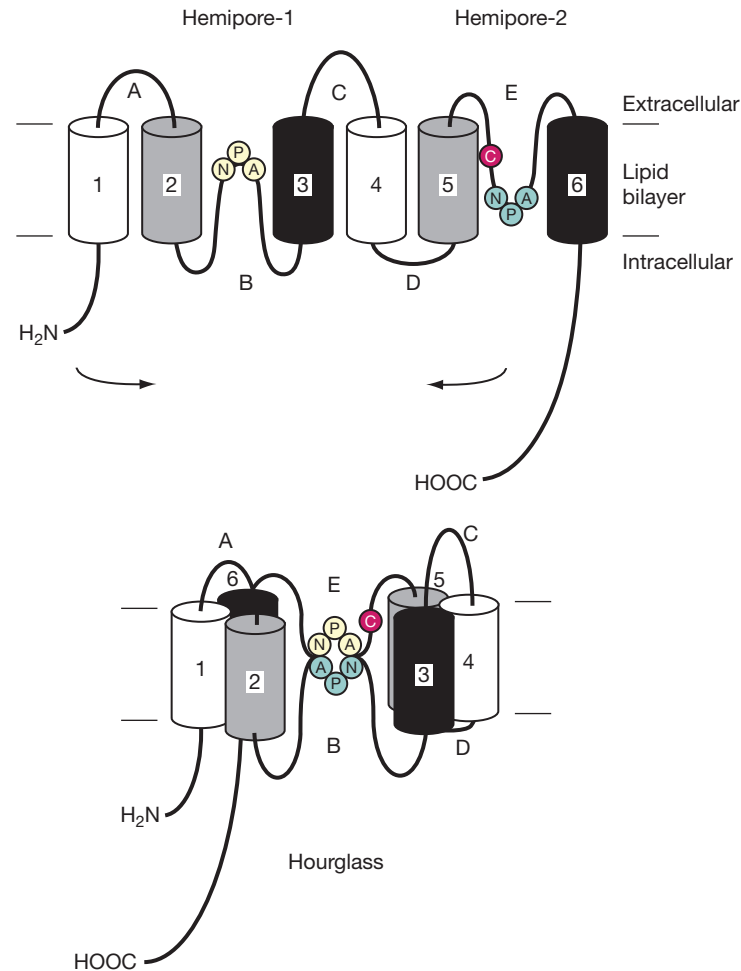


Figure 3 Membrane topology of AQP1 (upper panel) and the hourglass model (lower panel). Reproduced from Agre P, King LS, Yasui M, et al. (2002) Aquaporin water channels – from atomic structure to clinical medicine. *Journal of Physiology* 542: 3–16, with permission.

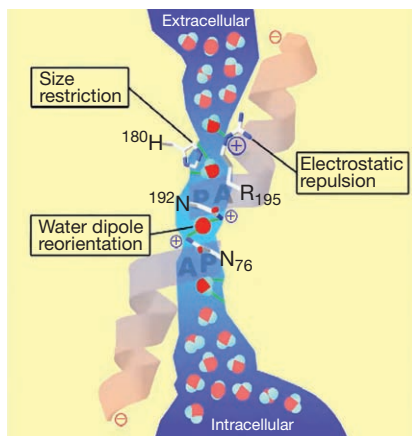


Figure 4 Schematic of sagittal cross section of AQP1 at the pore region. Reproduced from Kozono D, Yasui M, King LS, and Agre P (2002) Aquaporin water channels: Atomic structure and molecular dynamics meet clinical medicine. *Journal of Clinical Investigation* 109: 1395–1399, with permission.

Human AQP1 gene was mapped on the short arm of chromosome 7 by fluorescence *in situ* hybridization where it coincides with the Colton blood group antigens. The Colton A and Colton B antigens correspond to a polymorphism (alanine–valine) in an extracellular domain of AQP1 protein. Colton-null is extremely rare, with only six kindreds documented by the International Blood Group Reference Laboratory. The Colton-null individuals were found to be homozygous for knockout mutations in the AQP1 gene, and their red-cell membranes lacked AQP1 protein and exhibited marked reductions in red-cell osmotic water permeability.

In a clinical renal function study, two unrelated AQP1-null individuals showed no obvious defects under baseline conditions; however, both exhibited markedly reduced renal concentration after overnight thirsting ($450 \text{ mosmol kg}^{-1}$ in AQP1-null individuals vs. $1000 \text{ mosmol kg}^{-1}$ in normals). Although they are extremely rare, individuals with defects in AQP1 would be at risk for life-threatening clinical problems when they are dehydrated due to environmental causes or other illnesses.

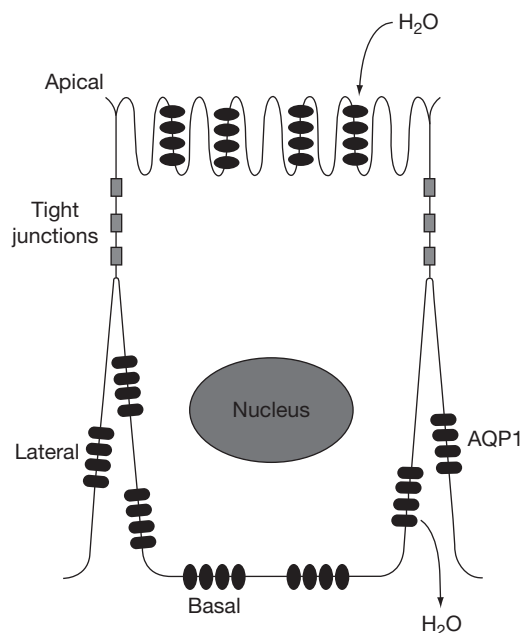


Figure 5 Constitutive transcellular water movements by AQP1 through a single cell in renal proximal tubule epithelium. Reproduced from Nielsen S, Smith BL, Christensen EI, Knepper MA, and Agre P (1993) CHIP28 water channels are localized in constitutively water-permeable segments of the nephron. *Journal of Cell Biology* 120: 371–383, with permission.

In a follow-up study of AQP1 functions in vascular endothelium, high-resolution computed tomography was applied to scan the lungs before and after saline infusion. Normal individuals had a significant increase in the airway wall thickness with early peribronchiolar edema and moderate pulmonary congestion. Interestingly, none of the AQP1-null individuals had increased bronchiolar wall thickness. This study showed that AQP1-null individuals have significant reduction in vascular water permeability, indicating the critical role of AQP1 in lung. While yet unproven, such AQP1-null individuals may be at risk of pulmonary dysfunction due to reduced absorption of amniotic fluid at birth.

AQP2 – Vasopressin-Regulated Water Channel in Renal Collecting Ducts

Water transport occurring in renal collecting ducts (CDs) is regulated by the antidiuretic hormone arginine vasopressin (AVP). The lack of AQP1 in CDs predicted the existence of another aquaporin family member, AQP2, which is predominantly expressed in CD principal cells. In the absence of AVP, AQP2 is largely restricted to intracellular vesicles, and dilute urine is released (the diuretic state). In the presence of AVP, AQP2 is redistributed to the apical membrane, accompanied by increased water permeability in CDs with corresponding release of concentrated urine (the antidiuretic state). After vasopressin is removed, AQP2 is re-internalized and the CD water permeability returns to baseline. Humans go from the diuretic state to the antidiuretic state and back to the diuretic state as water intake changes.

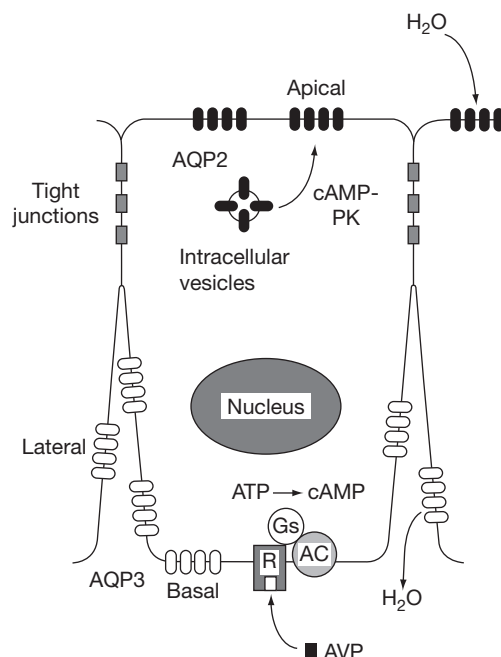


Figure 6 Vasopressin-regulated AQP2 translocation and transcellular water movements. Reproduced from Nielsen S and Agre P (1995) The aquaporin family of water channels in kidney. *Kidney International* 48: 1057–1068, with permission.

Subsequently, researchers found that AQP2 is regulated by exocytosis. Briefly, AVP binds to vasopressin 2 (V2) receptor at the basolateral membrane of CD principal cells, activating a cyclic adenosine monophosphate (cAMP)-dependent phosphorylation cascade. The phosphorylation at Ser-256 and probably other residues in AQP2 results in movement to the apical plasma membrane by exocytosis. AQP3 and AQP4 were later shown to reside on the basolateral membranes of CD principal cells. Thus, water from the primary urine enters the apical membrane of principal cells through AQP2 and exits from the basolateral membranes through AQP3 or 4, following the transcellular pathway (Figure 6). Certain genetic mutations in AQP2 have been reported to cause nephrogenic diabetes insipidus, where the kidney fails to concentrate urine in response to vasopressin and patients show severe clinical symptoms. Even milder forms of renal defective urine concentration, such as bedwetting by small children, are believed to represent temporary deficiencies of AQP2 that occur during growth.

Other Aquaporin Family Members

Aquaporin proteins have been identified in many diverse forms of life. At present, 13 homologs (AQP0–AQP12) have been identified in mammals, with a wide tissue and organ distribution and transport specificities (summarized in Table 1). The physiological roles of AQP0–AQP9 are relatively well elucidated; however, little is known about the functions of AQP10, AQP11, and AQP12. According to protein sequences and substrate specificity, most mammalian aquaporins can be classified into two subgroups, classical aquaporins (water-

Table 1 The predominant distribution and permeability of mammalian aquaporins

	Predominant distribution	Permeability and other roles
AQP0	Lens	Water (low), cell adhesion
AQP1	Red cells, kidney, lung, vascular endothelium, brain, eye	Water (high)
AQP2	Kidney, vas deferens	Water (high), regulated by exocytosis
AQP3	Kidney, skin, lung, eye, colon	Water (high), glycerol (high), urea (moderate)
AQP4	Brain, muscle, kidney, lung, stomach, small intestine	Water (high)
AQP5	Salivary gland, lacrimal gland, sweat gland, lung, cornea	Water (high)
AQP6	Kidney	Water (low), anions
AQP7	Adipose tissue, kidney, testis	Water (high), glycerol (high), urea (high), arsenite
AQP8	Testis, kidney, liver, pancreas, small intestine, colon	Water (high)
AQP9	Liver, leukocytes, brain, testis	Water (low), glycerol (high), urea (high), arsenite
AQP10	Small intestine	Water (low)
AQP11	Brain, testis, liver, kidney, heart	Unknown
AQP12	Pancreas	Unknown

Reproduced from King L, Kozono D, and Agre P (2004) From structure to disease: The evolving tale of aquaporin biology. *Nature Reviews. Molecular Cell Biology* 5(9): 687–698 (review), with permission.

specific) and aquaglyceroporins (permeable to glycerol or other small, uncharged solutes).

It should be emphasized that the functional roles of sequence-related proteins may be more complex. As discussed below, AQP0 may have a second role as a cell adhesion molecule, and AQP6 is an anion-selective channel. The evolutionary relationships of human AQP1–10 are shown in Figure 7. Lacking the signature motif, NPA (pore-lining NPA), as well as certain highly conserved residues, AQP11 and AQP12 appear to be more distantly related to the other homologs.

AQP0, initially called major intrinsic protein (MIP), is expressed only in fiber cells of lens. In an AQP0 structure revealed by electron microscopy of two-dimensional (2D) crystals, every subunit of a tetramer in one lipid layer interacts with two subunits of a tetramer in the adjacent layer. As a result, the pore is too narrow for water passage. This conformation explains why AQP0 has fairly low water permeability, and also demonstrates its structural role in cell–cell adhesion. Two structurally significant AQP0 amino acid substitutions were identified in two kindreds with congenital cataracts of dominant inheritance, suggesting that AQP0 is a risk factor for cataract development.

AQP4 is expressed in multiple organs, including brain. By immunogold electron microscopy, AQP4 has been shown to be highly concentrated in astroglial endfeet membranes, forming microcrystalline lattices referred to as ‘square arrays’. A C-terminal PDZ-binding (PDZ, postsynaptic density protein –

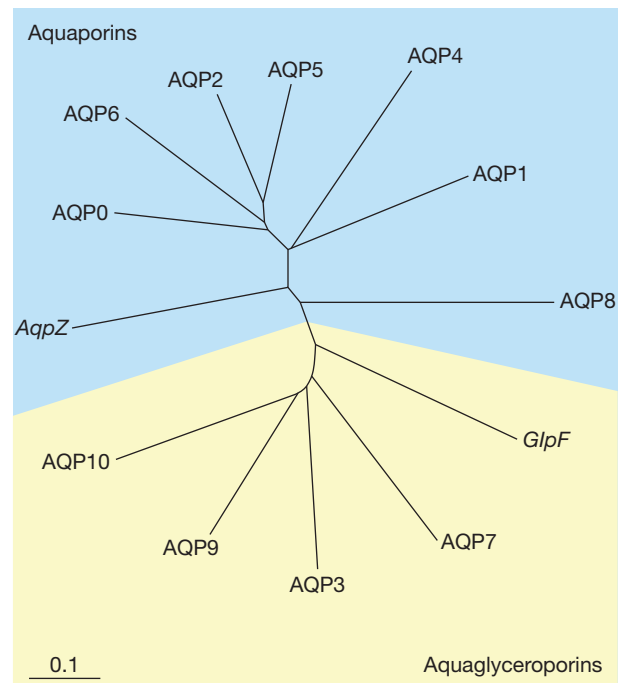


Figure 7 The evolutionary relationships of human aquaporins and aquaglyceroporins with two *E. coli* proteins, AqpZ and GlpF. Reproduced from Agre P, Bonhivers M, and Borgnia M (1998) The aquaporins, blueprints for cellular plumbing systems. *Journal of Biological Chemistry* 273: 14659–14662, with permission.

Drosophila disk large tumor suppressor – zonula occludens-1 protein) motif stabilizes the perivascular localization of AQP4 in brain through binding to a structural protein complex. Results from AQP4 localization and biological function studies, together with analysis of AQP4-null mice, suggest that AQP4 may permit water to cross the blood–brain barrier that separates parenchyma and vascular space. The importance of AQP4 in human cerebral edema is expected.

AQP5 resides in secretory glands that facilitate the release of large amounts of body fluids, including saliva, tears, and airway surface liquid. AQP5-null mice have diminished ability to salivate and sweat. In a subset of humans with Sjögren’s syndrome, apparent mislocalization of AQP5 coincides with dry eyes and dry mouth, clinical hallmarks of the disorder. Failure of AQP5 to traffic to the apical surface membranes of the glandular epithelium may contribute to decreased lacrimation in these patients.

Although similar to AQP2 in terms of protein sequence, AQP6 has unique subcellular distribution and distinct biophysical functions. AQP6 is localized on intracellular vesicles in renal alpha intercalated cells and apparently transports anions, particularly nitrate, when activated by low pH. Mutagenesis studies suggest that an asparagine residue, Asn-60, at the contact point between the second and fifth transmembrane domains in AQP6 may function as a teeterboard which is needed for rapid structural oscillations during anion permeation. While still unproven, AQP6 may play a role in acid–base equilibrium.

The aquaglyceroporins, AQP3, AQP7, and AQP9, are permeated by water and other small solutes such as glycerol and urea.

AQP3 may contribute to skin integrity, and this protein is heavily expressed in basal layer of skin epithelium. Evidence indicates that expression of AQP3 is increased during wound healing and decreased during aging of sun-exposed skin. Important roles are being established for AQP7 in adipocytes and AQP9 in liver. During fasting, triglycerides in adipocytes are degraded into glycerol and fatty acids. Glycerol released from fat cells through AQP7 on plasma membrane enters hepatocytes through AQP9 on plasma membrane. Working together, AQP7 and AQP9 form an axis for energy transfer and maintenance of normal blood glucose levels even during starvation.

Surprisingly, AQP7 and AQP9 were also found to be permeated by heavy metals, including arsenite, providing a possible route for excretion of the extremely toxic arsenicals. Considering its expression in leukocytes, AQP9 could be a pharmacological pathway for the selective toxicity of arsenite which is used as an agent to treat promyelocytic leukemia.

Unlike human red cells, which express AQP3, the major glycerol channel in mouse red cells is AQP9, and AQP9-null mice infected with malaria parasite *Plasmodium berghei* are slightly protected during the initial phase of infection compared with wild-type mice. These results suggest that the AQP9-mediated transport pathway may contribute to the virulence of intra-erythrocytic stages of malarial infection.

An aquaglyceroporin homolog, PbAQP, exists in rodent malaria parasite *Plasmodium berghei*. The parasite has to utilize host glycerol for its own lipid biosynthesis and membrane biogenesis. Furthermore, the parasite also faces rapid and significant volume change in host red blood cells. As the only channel for water and glycerol passage, PbAQP contributes to the parasite's high proliferation rate and the ability to evade physiological stresses. Targeted disruption of PbAQP has shown that mice infected with PbAQP-null parasites survived longer than those infected by wild-type parasites, suggesting that PbAQP plays an important role in the blood-stage development of the parasite.

Functions of AQP10, 11, and 12 are not well understood. Human AQP10 may be a pseudogene and the number of paralogs is being debated. Analysis of mice with targeted disruption of *Aqp10* and *Aqp12* has not revealed their physiological functions. Interestingly, the AQP11-null mouse has been reported to sustain fatal polycystic kidneys, while the pathophysiological mechanism is not clear yet.

Unlike animals, plants cannot effectively evade unfavorable climatic conditions. Therefore, a fine-tuned adaptation mechanism including water balance is fundamental for plant physiology. Plants contain a large number of aquaporins (38 in *Arabidopsis*), with a variety of functions in root water uptake, reproduction, or photosynthesis. Plant aquaporins conduct water, and some may also conduct NH_3 , CO_2 , and H_2O_2 . Some plant aquaporins, including the spinach protein SoPIP2;1, have been reported to exhibit gating in response to changes in pH and phosphorylation.

Methanothermobacter marburgensis, a methanogenic archaeon thriving in anaerobic environments even at 65 °C, also carries an aquaporin, aquaporin-M (AQPM), which has been proposed to facilitate transport of hydrogen sulfide *in vivo*. Its existence demonstrates the ubiquity of aquaporins in nature and the crucial importance of this family.

Several fungal aquaporins have also been studied. *Saccharomyces cerevisiae* expresses two, AQY1 and AQY2, and

Candida albicans expresses one, AQY1. Yeast aquaporins from wild-type organisms all facilitate water transport when expressed in oocytes, whereas sequences obtained from laboratory strains carry loss-of-function mutations. Functional AQY2 protein from *Saccharomyces chevalieri* was expressed in *aqy1* and *aqy2* double-null yeast and this caused a reduction in cell surface hydrophobicity, suggesting that fungal aquaporins may promote metabolic advantages or affect virulence by enhancing cell dispersion.

Eight aquaporin genes (*aqp1–8*) are present in the genome of the important flatworm *Caenorhabditis elegans* genome. These aquaporins are expressed at multiple sites (such as intestine, excretory cells, and hypodermis) that play important roles in osmoregulation. Some members have water and glycerol transport activities when measured in *Xenopus* oocytes. However, worms with targeted disruption in one or more aquaporin genes did not show obvious functional clues to their roles, so the specific physiological roles of aquaporins in *C. elegans* await further analysis.

Insects have also been reported to contain members of the aquaporin family – usually at the sites where fluid transport is high, such as digestive tract of sap-sucking *Cicadella viridis*, silk glands of silkworms, and salivary glands of blood-feeding dog ticks. Moreover, the *Drosophila* neurogenic protein, big brain (Bib), has sequence similarity to aquaporins and has been described to exhibit novel cellular functions in endosome maturation and notch receptor trafficking.

Summary

Water is the major component of all living organisms, and water homeostasis is crucial to life, especially at times of environmental or physiological stress. The 2003 severe heat wave in Europe lasted 2 weeks. Public health officials reported tens of thousands of deaths from severe dehydration and related clinical complications. In many instances, it is likely that the victims were unable to adequately compensate by renal concentration and sweat mechanisms. Thus, aquaporin physiological pathways may have been overwhelmed. Since the discovery of AQP1 water channel, this field has expanded dramatically. Researchers from several countries are now studying aquaporins in multiple types of organisms. Besides water, glycerol, and urea permeation, additional roles for aquaporins have been proposed as structural membrane support, heavy metal transport, and even gas transport. Although many questions remain unanswered, knowledge obtained from studies of aquaporins provides improved understanding of the pathophysiology of water homeostasis and may in the future provide clinical benefits in treating a wide range of diseases.

See also: [Lipids Carbohydrates Membranes and Membrane Proteins: Ion Channel Protein Superfamily](#).

Further Reading

Carbrey JM, Bonhivers M, Boeke JD, and Agre P (2001) Aquaporins in *Saccharomyces*: Characterization of a second functional water channel protein.

- Proceedings of the National Academy of Sciences of the United States of America* 98: 1000–1005.
- Carbrey JM, Gorelick-Feldman DA, Kozono D, et al. (2003) Aquaglyceroporin AQP9: Solute permeation and metabolic control of expression in liver. *Proceedings of the National Academy of Sciences of the United States of America* 100: 2945–2950.
- Carbrey JM, Song L, Zhou Y, et al. (2009) Reduced arsenic clearance and increased toxicity in aquaglyceroporin-9-null mice. *Proceedings of the National Academy of Sciences of the United States of America* 106: 15956–15960.
- Deen PM, Verdijk MA, Knoers NV, et al. (1994) Requirement of human renal water channel aquaporin-2 for vasopressin-dependent concentration of urine. *Science* 264: 92–95.
- Fu D, Libson A, Miercke LJ, et al. (2000) Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* 290: 481–486.
- King LS, Choi M, Fernandez PC, Cartron JP, and Agre P (2001) Defective urinary concentrating ability due to a complete deficiency of aquaporin-1. *New England Journal of Medicine* 345: 175–179.
- King LS, Nielsen S, Agre P, and Brown RH (2002) Decreased pulmonary vascular permeability in aquaporin-1-null humans. *Proceedings of the National Academy of Sciences of the United States of America* 99: 1059–1063.
- Liu K, Kozono D, Kato Y, et al. (2005) Conversion of aquaporin 6 from an anion channel to a water-selective channel by a single amino acid substitution. *Proceedings of the National Academy of Sciences of the United States of America* 102: 2192–2197.
- Liu Y, Promeneur D, Rojek A, et al. (2007) Aquaporin 9 is the major pathway for glycerol uptake by mouse erythrocytes, with implications for malarial virulence. *Proceedings of the National Academy of Sciences of the United States of America* 104: 12560–12564.
- Maeda N, Funahashi T, Hibuse T, et al. (2004) Adaptation to fasting by glycerol transport through aquaporin 7 in adipose tissue. *Proceedings of the National Academy of Sciences of the United States of America* 101: 17801–17806.
- Murata K, Mitsuoka K, Hirai T, et al. (2000) Structural determinants of water permeation through aquaporin-1. *Nature* 407: 599–605.
- Neely JD, Amiry-Moghaddam M, Ottersen OP, et al. (2001) Syntrophin-dependent expression and localization of Aquaporin-4 water channel protein. *Proceedings of the National Academy of Sciences of the United States of America* 98: 14108–14113.
- Nielsen S, Chou CL, Marples D, et al. (1995) Vasopressin increases water permeability of kidney collecting duct by inducing translocation of aquaporin-CD water channels to plasma membrane. *Proceedings of the National Academy of Sciences of the United States of America* 92: 1013–1017.
- Nielsen S, Frokiaer J, Marples D, et al. (2002) Aquaporins in the kidney: From molecules to medicine. *Physiological Reviews* 82: 205–244.
- Preston GM, Carroll TP, Guggino WB, and Agre P (1992) Appearance of water channels in *Xenopus* oocytes expressing red cell CHIP28 protein. *Science* 256: 385–387.
- Preston GM, Smith BL, Zeidel ML, Moulds JJ, and Agre P (1994) Mutations in aquaporin-1 in phenotypically normal humans without functional CHIP water channels. *Science* 265: 1585–1587.
- Promeneur D, Liu Y, Maciel J, et al. (2007) Aquaglyceroporin PbAQP during intraerythrocytic development of the malaria parasite *Plasmodium berghei*. *Proceedings of the National Academy of Sciences of the United States of America* 104: 2211–2216.
- Walz T, Smith BL, Zeidel ML, Engel A, and Agre P (1994) Biologically active two-dimensional crystals of aquaporin CHIP. *Journal of Biological Chemistry* 269: 1583–1586.
- Yasui M, Hazama A, Kwon TH, et al. (1999) Rapid gating and anion permeability of an intracellular aquaporin. *Nature* 402: 184–187.

Arachidonic Acid Regulated Calcium Channel

T J Shuttleworth, University of Rochester Medical Center, Rochester, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

AKAP (A-kinase anchoring protein) A family of proteins that binds to the regulatory subunits of the cyclic AMP-activated kinase protein kinase A, to provide a scaffold enabling the targeting of various other proteins (e.g., kinases, phosphodiesterases, phosphatases, etc.) in close proximity to PKA, whose activity they regulate.

ARC channel (arachidonic acid regulated Ca^{2+} channel) A small conductance, highly Ca^{2+} -selective channel that is specifically activated by low concentrations (2–8 μM) of arachidonic acid in a manner that is entirely independent of any store depletion (cf. CRAC channels).

CRAC channel (Ca^{2+} release-activated Ca^{2+} channel) Another small conductance, highly Ca^{2+} -selective channel that is specifically activated following depletion of intracellular Ca^{2+} stores.

Orai A group of three recently identified novel proteins (Orai1–3) that are the pore-forming components of both CRAC and ARC channels.

STIM1 (stromal interacting molecule 1) A protein that was originally identified as a cell-surface protein that played a key role in interactions between hematopoietic cells and cells in the bone marrow stroma, and as a growth suppressor of certain tumors. More recently, it was identified as playing an essential role in the activation of store-operated Ca^{2+} entry via its ability to sense falling Ca^{2+} concentrations within the intracellular stores and to subsequently activate the plasma-membrane store-operated Ca^{2+} channels. Subsequent studies have revealed that the pool of STIM1 that constitutively resides in the plasma membrane plays an equally critical role in the activation of the store-independent ARC channels.

TRPC channels (canonical TRP channels) A subgroup of the mammalian transient receptor potential (TRP) channels that is most closely related to the TRP channels of *Drosophila*. They are essentially nonselective cation channels that are activated under various conditions of receptor activation, some of which are claimed to be store-operated and to involve STIM1, although this remains controversial.

Different Modes of Ca^{2+} Entry and the Discovery of the ARC Channels

The receptor-activated entry of Ca^{2+} in nonexcitable cells plays a key role in the generation and modulation of the resulting intracellular Ca^{2+} signals that are known to be responsible for a wide variety of critical cellular responses, including cell growth and proliferation, differentiation, movement, secretion, metabolism, and cell death. Our understanding of the underlying mechanisms and pathways involved in such entry began with the work of Putney in the mid-1980s and his development of the concept of capacitative, or store-operated, Ca^{2+} entry, where activation of the entry of Ca^{2+} is strictly a consequence of the depletion of intracellular Ca^{2+} stores. Demonstration of the presence of such store-operated entry pathways in almost all cell types examined has led to the widely held belief that this specific mode of Ca^{2+} entry represents the major, if not the only, mode of such receptor-activated entry in nonexcitable cells. However, examination of the literature reveals that direct evidence demonstrating a principal role for such pathways in the various receptor-activated Ca^{2+} signals and responses seen in these different cell types is limited. More specifically, although various conductances have been proposed to provide the route for store-operated Ca^{2+} entry in different cells types, the most thoroughly characterized store-operated Ca^{2+} channel remains the Ca^{2+} -release activated Ca^{2+} (CRAC) channel. This conductance was originally identified in T lymphocytes and in mast cells, and here it is clear that the activity of the CRAC channels is critical for the physiological activation of these cells. However, examination of the properties of this channel and its activation

raises several questions as to its ability to function as the exclusive pathway of agonist-activated Ca^{2+} entry in other cell types. These include a requirement for a rather profound depletion of intracellular stores before any activation begins, the subsequent relatively slow rate of activation, and evidence indicating that such activation occurs in an essentially all-or-none fashion. Together, these features are difficult to reconcile with the predicted requirement of a finely regulated, graded Ca^{2+} response in order to accurately coordinate the appropriate cellular response to any particular increase in the Ca^{2+} signal. On a more general level, it should also be remembered that depletion of the intracellular Ca^{2+} stores can be induced by a variety of different events, many of which are unrelated to the normal, physiologically relevant, agonist-activated signaling pathways. For example, store depletion can result from an inhibition of the sarcoplasmic and endoplasmic calcium (SERCA) pump such as may occur during ischemia, or an increase in passive leak pathways such as those provided by the presenilins, as well as exposure to oxidants and other cellular stressors. Finally, the absolute requirement for a sustained depletion of intracellular Ca^{2+} stores in order to activate these conductances is potentially problematic, as such depletion can seriously impact various vital cell processes, the most obvious being the correct folding and processing of secretory and membrane proteins. Together, consideration of these issues suggests that, although store-operated modes of Ca^{2+} entry can be demonstrated in almost all cell types, their principal role might lie in some process other than normal signaling in response to agonist stimulation. Indeed, given their exclusive dependence on the depletion of intracellular Ca^{2+} stores for activation, the most

likely primary role for such store-operated entry pathways would seem to be the maintenance of adequate levels of Ca^{2+} in the ER.

These arguments eventually led to the search for possible alternate pathways of agonist-activated Ca^{2+} entry whose activity might be more directly or specifically connected to agonist-induced receptor activation. The result of one such search was the discovery of an entry pathway whose activation is entirely unaffected by store depletion and is instead dependent on the agonist-activated, receptor-mediated generation of arachidonic acid. Initial evidence in support of such a pathway included the demonstration that addition of low concentrations ($<8 \mu\text{M}$) of exogenous arachidonic acid resulted in an entry of Ca^{2+} without detectable release from stores, that the generation of arachidonic acid was increased following addition of low physiologically relevant concentrations of agonists to cells, and that the pharmacological inhibition of such agonist-activated arachidonic acid generation resulted in the inhibition of the associated entry of Ca^{2+} . Finally, it was shown that the activation of the entry of Ca^{2+} was an effect of arachidonic acid itself, and not any of its metabolites. Subsequently, much as the CRAC channel was eventually identified as the archetypal store-operated Ca^{2+} -selective conductance, the specific conductance responsible for this store-independent, arachidonic acid-activated Ca^{2+} entry was identified and named the arachidonic acid regulated Ca^{2+} (ARC) channel. In the approximately 10 years since their first identification, considerable details of these channels have been revealed, including their biophysical and pharmacological profiles, specificity for arachidonic acid, occurrence in different cell types, and, most recently, the molecular basis of their composition and regulation. Moreover, there are now clear examples of situations where these ARC channels have been shown to play unique, critical roles in the physiological responses of various cell types.

Characteristics of ARC Channels

Much like the CRAC channels, ARC channels represent a relatively small (with typical whole-cell currents of approximately $0.5\text{--}1.5 \text{ pA pF}^{-1}$ in most cell types), highly Ca^{2+} -selective conductances. Consistent with this high selectivity for Ca^{2+} , their current-voltage relationship reveals marked inward rectification and a reversal potential in excess of $+60 \text{ mV}$ (Figure 1). Similarly, like most Ca^{2+} channels, they are blocked by La^{3+} and by Gd^{3+} with complete inhibition achieved by $50\text{--}100 \mu\text{M}$ and $5 \mu\text{M}$, respectively. However, unlike the CRAC channels, they are not affected by moderate acidification of the external medium, and they are unaffected by the widely used, but relatively nonspecific, drug 2-aminoethoxydiphenyl borate (2-APB). In addition, the ARC channels do not share the clear fast inactivation during brief pulses to negative potentials that are a characteristic of CRAC channels (Figure 1).

Activation of the ARC channels occurs at the same low concentrations of exogenous arachidonic acid ($2\text{--}8 \mu\text{M}$) as was seen to induce store-independent Ca^{2+} entry in the earlier studies (see above). Importantly, such concentrations are clearly physiologically relevant as they are similar to values for the K_d 's of the cytosolic enzymes that metabolize arachidonic acid. Moreover, effective and specific activation of the conductance by these low concentrations is a critical property

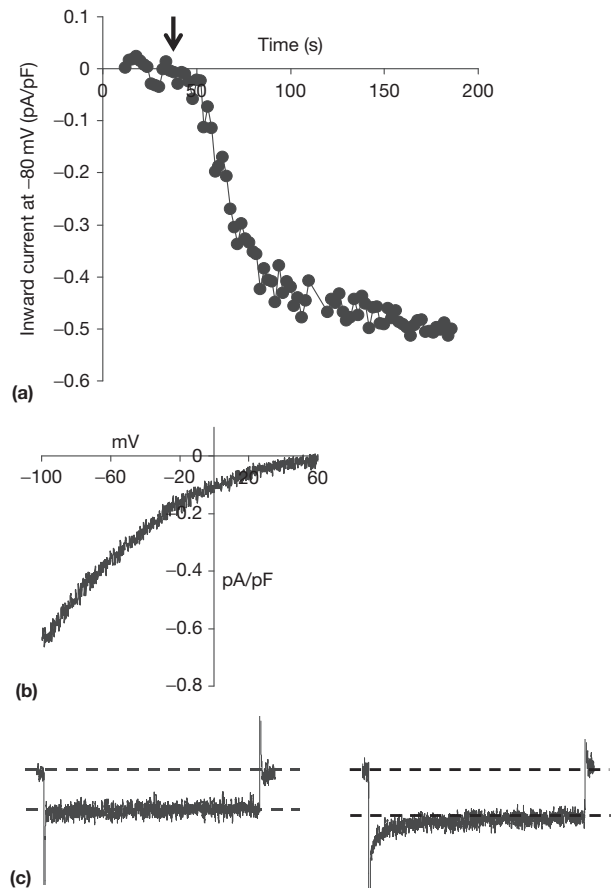


Figure 1 Basic features of currents through endogenous ARC channels. (a) Whole-cell patch-clamp recording of inward current density at -80 mV in a HEK293 cell. Free Ca^{2+} concentration in the external (bath) solution was 10 mM Ca^{2+} and in the internal (pipette) solution was 100 nM . At the point indicated (arrow), $8 \mu\text{M}$ arachidonic acid was added to the external bath. (b) Representative current-voltage relationship of the arachidonic acid-activated current showing the marked inward rectification and very positive reversal potential typical of a highly Ca^{2+} -selective conductance. (c) Comparison of the current traces through the ARC channels (left) and through the corresponding store-operated CRAC channels (right) present in the same cells during brief (250 ms) pulses to -120 mV . Note that the fast-inactivation characteristic of currents through CRAC channels is absent in the ARC channel currents.

as it is known that arachidonic acid at higher concentrations (e.g., $>10\text{--}20 \mu\text{M}$) can induce significant effects on the integrity of the cell membrane, resulting in the permeation of even relatively large, normally impermeant, ions such as NMDG^+ . Activation of the ARC channels also shows a high specificity for arachidonic acid, as both saturated fatty acids (e.g., palmitic acid) and mono-unsaturated fatty acids (e.g., oleic acid) are completely without effect. Even among other polyunsaturated fatty acids, only linoleic acid shows a small effect and this probably arises from the fact that this fatty acid is a metabolic precursor for arachidonic acid. In addition, and in contrast to certain members of the transient receptor potential channel (TRPC) family of channels, the ARC channels are entirely unaffected by diacylglycerol and its analogs. Finally, experiments using the nonpermeable arachidonic acid analog arachidonyl-coenzyme

A demonstrated that the activation of the channel by arachidonic acid is exclusively via an action of the fatty acid at the inner face of the membrane.

Since their initial discovery and characterization almost 10 years ago, ARC channels (or, at least, conductances displaying all the key features of these channels) have been described in a wide variety of different cell types, including many common cell lines (HEK293, HeLa, COS, and RBL), SY6Y neuroblastoma cells, K562 erythroleukemia cells, as well as the exocrine parotid and pancreatic acinar cells, and the endocrine pancreatic β cells. In many of these cell types, the ARC channels coexist with store-operated channels, including the biophysically very similar CRAC channels. Nevertheless, it can be demonstrated that these two coexisting channels represent entirely distinct entities. Thus, using the whole-cell patch clamp mode in HEK293 cells, it has been shown that maximal activation of the CRAC channels by store depletion has no effect on the magnitude of the subsequently activated ARC channel currents induced by addition of exogenous arachidonic acid – in other words, at least under whole-cell patch clamp conditions, the two currents are strictly additive in the same individual cell.

Reciprocal Regulation of ARC Channels and CRAC Channels

As described above, under whole-cell patch clamp conditions, currents through the ARC channels and coexisting store-operated CRAC channels are strictly additive. However, it is important to note that such additivity of the two currents does not appear to occur in the intact cell. The reason for this difference is that, in the whole-cell patch clamp configuration, the presence of buffers in the pipette solution clamps the intracellular Ca^{2+} concentration at resting levels while, in the normal intact cell situation, activation of store-operated Ca^{2+} entry typically results in a sustained elevation of intracellular Ca^{2+} . Studies have shown that this sustained elevation in intracellular Ca^{2+} results in the activation of the Ca^{2+} -calmodulin-dependent phosphatase calcineurin and this, in turn, induces an inhibition of ARC channel activity. Consequently, in the intact cell, activation of the store-operated channels (e.g., CRAC channels) will result in an inhibition of the ARC channels, a process that has been described as the 'reciprocal regulation of Ca^{2+} entry'.

The discovery that the phosphatase calcineurin inhibits the ARC channels suggested that the activity of these channels must be dependent, in some way, on phosphorylation. Subsequent studies showed that the specific kinase responsible was the cyclic AMP-dependent protein kinase A (PKA). Thus, the currents through the ARC channels measured in HEK293 cells maintained under normal culture conditions following addition of low concentrations of arachidonic acid are inhibited by preincubation with the cell-permeable PKA-specific inhibitory peptide myristoylated PKI but are unaffected by inhibition of either protein kinase C (with calphostin) or protein kinase G (with KT5823). Correspondingly, the addition of forskolin, particularly after induction of the calcineurin-mediated dephosphorylation, resulted in enhanced arachidonic acid-activated ARC channel currents. In addition, it has been shown that this action of PKA in modulating ARC channel activity involves a member of the A-kinase anchoring protein (AKAP) family. For example,

studies utilizing a cell-permeable stearylated version of a synthetic peptide that disrupts the binding of PKA to AKAPs (st-Ht31 peptide) both inhibited normal ARC channel currents and blocked the above effects of forskolin on the currents.

Molecular Nature (and Activation) of ARC Channels

Although the CRAC channel had been originally described and characterized in the early 1990s, the molecular nature of its structure and regulation were unknown until the recent identification of stromal interacting molecule 1 (STIM1) as the sensor of store depletion, and the discovery of Orai1 as the molecule that forms the CRAC channel itself. These studies revealed that STIM1 located in the membrane of the endoplasmic reticulum (ER) sensed the depletion of Ca^{2+} in these stores via a luminal Ca^{2+} -binding EF-hand motif present in the N-terminal region of the protein. Loss of Ca^{2+} from this motif results in the oligomerization of STIM1 and its translocation to regions of the ER that lie adjacent to the plasma membrane. Here, it interacts with Orai1 resulting in the activation of the CRAC channels.

Given this clearly demonstrated role of STIM1 as the ER Ca^{2+} sensor in the activation of the CRAC channels, it would be expected that STIM1 would be unlikely to play any role in the activation of the store-independent ARC channels. However, experiments showed that ARC channel currents were increased by overexpression of STIM1 and reduced by siRNA knock-down of the endogenous protein, in a manner that was indistinguishable from that seen with the CRAC channels present in the same cells. Reconciliation of these unexpected results began with the realization that the STIM1 protein is not exclusively restricted to the ER membrane. In fact, it was originally identified as a cell surface (i.e., plasma-membrane) protein present in the stromal cells of the bone marrow, where it was involved in interactions between these cells and hematopoietic cells (specifically pre-B lymphoid cells). Studies have shown that the pool of STIM1 that is resident in the plasma membrane is relatively small (approximately 10–25% depending on cell type), and its delivery to this location is strictly dependent on N-linked glycosylation at two sites within the N-terminal region of the protein (Asn131 and Asn171). To examine the effect of specifically eliminating this cell-surface pool of STIM1, cells treated with an siRNA targeting the endogenous STIM1 were transfected with an siRNA-resistant STIM1 construct in which these two glycosylation sites were mutated (N131Q and N171Q) – thereby effectively replacing the native STIM1 with the mutant version. Examination of both ARC and CRAC currents in these cells revealed that while CRAC channel currents were unaffected, the ARC channel currents in the same cells were profoundly inhibited (Figure 2). These data clearly established that it is the pool of STIM1 resident in the plasma membrane that is essential for activation of the ARC channels. As yet, exactly how the plasma membrane STIM1 acts to regulate the activation of the ARC channels is unknown. However, it is important to note that, in the case of this plasma membrane pool of STIM1, the Ca^{2+} -binding N-terminal EF-hand motif will lie in the extracellular environment. Given the high and relatively stable Ca^{2+} concentrations of this environment, it is clear that the EF-hand of this pool of STIM1 would always, under normal circumstances, be in the Ca^{2+} -bound state.

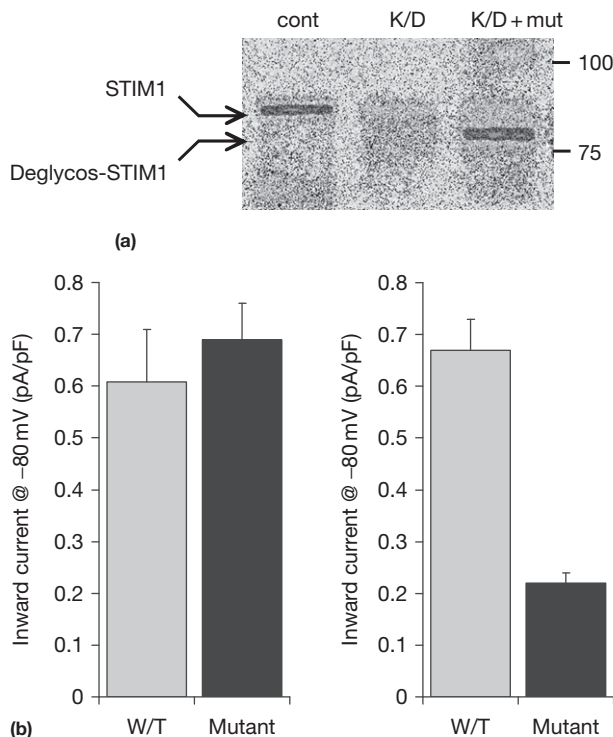


Figure 2 Expression of STIM1 in which the glycosylation sites that are essential for delivery of the protein to the plasma membrane are mutated reveals that activation of ARC channels is exclusively dependent on the pool of STIM1 that constitutively resides in the plasma membrane. (a) Western blot of STIM1 in HEK293 cells, showing the levels of total STIM1 in untransfected cells (cont), in cells transfected with an STIM1 siRNA construct (K/D), and in siRNA-treated cells following transfection with an siRNA-resistant STIM1 in which the two N-glycosylation sites have been mutated (N131Q and N171Q) (K/D + mut). (b) The effects of expressing siRNA-resistant versions of either wild-type STIM1 (gray bars) or the glycosylation-mutant STIM1 (black bars) in STIM1 siRNA-treated cells. Currents through the store-operated CRAC currents (left panel) are unaffected by the expression of the glycosylation-mutant STIM1, while similar expression of the glycosylation-mutant STIM1 results in a profound inhibition of ARC channel currents (right panel). Reproduced from Mignen O, Thompson JL, and Shuttleworth TJ (2007) STIM1 regulates Ca^{2+} entry via arachidonate-regulated Ca^{2+} -selective (ARC) channels without store-depletion or translocation to the plasma membrane. *Journal of Physiology* 579: 703–715.

This stands in marked contrast to the critical role of the loss of Ca^{2+} from the luminal EF-hand of STIM1 in the ER in the regulation of the store-operated CRAC channels. Consequently, whatever the mechanism by which STIM1 regulates the ARC channels, it must be very different from the way it activates the CRAC channels.

The demonstration that STIM1 is required for the activation of both CRAC and ARC channels, although entirely distinct pools of this protein are responsible, and that both channels display very similar biophysical properties raised the possibility that, as had already been demonstrated for the CRAC channels, Orai proteins may form the ARC channel. Three distinct, but closely related Orai proteins (Orai1–3) are expressed in mammals, and while the CRAC channel is formed exclusively from Orai1, studies have demonstrated that both Orai1 and

Orai3 are required for the formation of the functional ARC channel. Thus, overexpression of Orai1 in cells stably expressing STIM1 results, not only in markedly increased CRAC channel currents but also in similar increases in ARC channel currents. Correspondingly, expression of the dominant-negative (E106Q) Orai1 mutant effectively eliminates both CRAC and ARC channel currents. However, expression of the corresponding dominant-negative E81Q Orai3 mutant essentially eliminates currents through the ARC channels and has no effect on the CRAC channel currents in the same cells.

ARC Channel Structure

Recent studies, utilizing a variety of different approaches, have demonstrated that the functional CRAC channel is a tetrameric assembly of Orai1 subunits. Given this, and the above data demonstrating the essential requirement for both Orai1 and Orai3 subunits in the functional ARC channel, the most likely structure for the channel would seem to be some form of heterotetrameric assembly of Orai1 and Orai3 subunits. However, examination of the conductances resulting from each of the possible tetrameric combinations of Orai1 and Orai3 subunits expressed in the form of pre-assembled concatenated constructs revealed unexpected results. Thus, while expression of some of these constructs resulted in significant arachidonic acid-activated, highly Ca^{2+} -selective currents, they also gave rise to similarly increased highly Ca^{2+} -selective currents that were activated by store depletion. Such a mixed selectivity in their mode of activation is entirely inconsistent with the native channels and, instead, suggests that such constructs generate some form of hybrid CRAC/ARC channel. Subsequently, additional experiments revealed that the functional ARC channel is, in fact, a pentameric structure formed from three Orai1 subunits and two Orai3 subunits (Figure 3). Expression of pre-assembled concatenated pentameric constructs with this composition resulted in large arachidonic acid-activated Ca^{2+} -selective currents without affecting the corresponding store-operated currents (Figure 3). Critically, the arachidonic acid-activated currents resulting in the expression of these pentameric constructs displayed all the key biophysical and pharmacological features of the endogenous ARC channels, including a high sensitivity to low concentrations (2–4 μM) of exogenous arachidonic acid, insensitivity to 2APB, and even an exclusive dependence on the plasma-membrane pool of STIM1 for their activation (Figure 4). As noted above, this demonstration that the functional ARC channel was a pentameric structure was unexpected, given that it had previously been shown that the functional CRAC channel was a tetramer. It remains to be determined precisely what role, if any, this fundamental difference in the overall structure of the two channels plays in the determination of their individual specific properties and distinct modes of activation.

What Do ARC Channels Do?

The fact that the ARC channels coexist with store-operated conductances in the same cells, and their biophysical properties are generally so similar to the store-operated CRAC channels, raises the obvious question of what their specific role

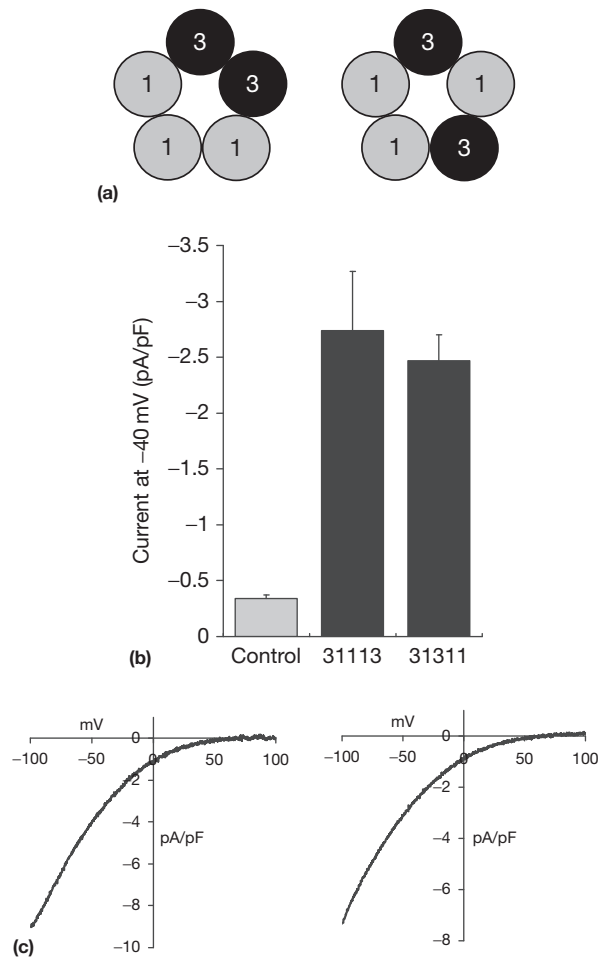


Figure 3 The functional ARC channel pore is comprised of a heteropentamer of three Orai subunits and two Orai3 subunits. (a) Cartoon illustrating the structural organization of the two Orai heteropentamers that have been shown to duplicate all the features of the endogenous ARC channels. (b) Inward currents, measured at -40 mV, activated by addition of $8 \mu\text{M}$ exogenous arachidonic acid in untransfected cells stably expressing STIM1 (control) are increased 8–10-fold following expression of either the 31113 or 31311 pentamer in the form of pre-assembled concatenated constructs. (c) Representative current–voltage relationships of the arachidonic acid-activated currents in STIM1-stable cells expressing either the concatenated 31113 pentamer (left) or the 31311 pentamer (right). Both display the marked inward rectification and very positive reversal potential seen in the endogenous ARC channels.

might be. The most likely answer was that their role is related, in some way, to their unique store-independent mode of activation and this has turned out to be the case. Thus, it has been demonstrated in various cell types that the ARC channels are specifically activated by the low, physiologically relevant, concentrations of agonists that typically produce Ca^{2+} signals in the form of repeated oscillations of various frequencies. Studies have shown that such oscillatory Ca^{2+} signals are associated with, at most, only a minimal (2–5%) depletion of the intracellular Ca^{2+} stores – levels that would not be capable of inducing activation of the store-operated conductances (see above). Consequently, the available evidence indicates that,

under these conditions, it is the ARC channels that provide the predominant pathway for Ca^{2+} entry. Although occasional arguments to the contrary have been proposed, these have generally relied on pharmacological approaches which are severely limited by the questionable specificity and selectivity of the drugs currently available (e.g., 2-APB), or have involved the application of arachidonic acid at concentrations that are known to induce obvious nonspecific effects on membrane integrity. Clearly, definitive identification of the specific Ca^{2+} entry pathways operating during such oscillatory responses can only come from studies involving the direct electrophysiological analysis of the activated conductance(s). Using this criterion, studies in HEK293 cells, parotid acinar cells, and in pancreatic acinar cells have all confirmed that the only Ca^{2+} entry channel that can be demonstrated to be active under such conditions is the store-independent ARC channel. Such analysis is particularly clear in the parotid and pancreatic acinar cells, as the biophysical properties of the store-operated conductances in these cells are entirely distinct from those of the ARC channels – they are relatively nonselective conductances (most likely reflecting the activity of TRPC channels), displaying relatively linear current–voltage relationships and with obvious outward currents at positive potentials. Moreover, in the case of the pancreatic acinar cells, the precise physiologically relevant concentrations of cholecystokinin, a major endogenous activator of secretion, are known ($10\text{--}20 \text{ pM}$), and studies have demonstrated that the only detectable Ca^{2+} entry conductance activated at these concentrations is the ARC channel.

A recent report has identified a more specific, and potentially highly important, role of the ARC channels as the key player in the release of insulin from pancreatic β cells. Thus, insulin secretagogues such as acetylcholine and cholecystokinin are known to induce the generation of arachidonic acid which, in turn, triggers insulin secretion from islets at low glucose levels and enhances the secretion of insulin mediated by depolarization. The recent study revealed that this action of arachidonic acid did not involve voltage-gated Ca^{2+} channels and instead specifically resulted from the activation of ARC channels in these cells. The ARC channels therefore represent a critical, and previously unsuspected, component of the overall secretagog-induced release of insulin in pancreatic β cells.

Conclusions and Implications

The demonstrated existence of the ARC channels in a range of different cell types, and the store-independent agonist-activated Ca^{2+} entry pathway that they provide, has several important implications. The most obvious and important is that it can no longer be assumed that simply demonstrating that store-operated Ca^{2+} entry is present in a cell (e.g., by addition of thapsigargin) implies that it is the only pathway for the agonist-activated entry of Ca^{2+} entry in that cell. Determination of which specific Ca^{2+} entry pathway is operating under the relevant conditions can only reliably be achieved by the direct measurement, and biophysical analysis, of the conductance(s) active under those conditions. The necessity of such specific identification of the relevant Ca^{2+} entry pathway is only emphasized by the fact that, as discussed above, the ARC channels have been shown to be present in a wide variety

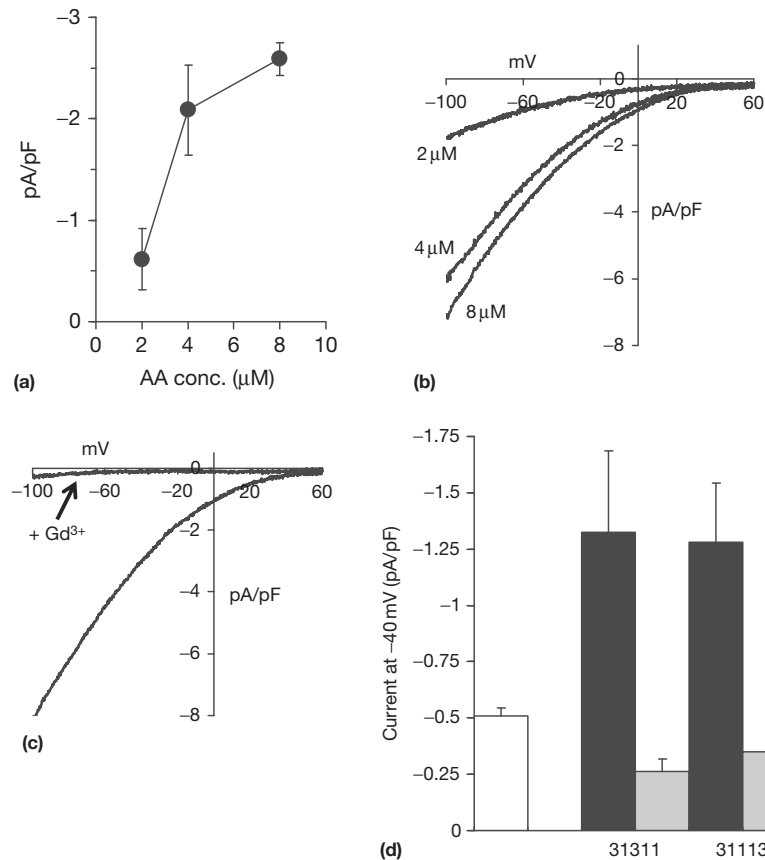


Figure 4 The features of the currents induced by expression of the appropriate Ora1/Orai3 heteropentamers precisely duplicate those of the endogenous ARC channels. (a) Activation of the currents induced following expression of the concatenated 31311 pentamer in STIM1-stable cells occurs at the same low concentrations of exogenous arachidonic acid as that reported for the endogenous ARC channels. Shown are the mean $\pm$ SE of the inward currents at -40 mV activated by the indicated concentrations arachidonic acid. (b) The resulting current–voltage relationships at all concentrations of arachidonic acid display marked inward rectification and very positive ($> +60$ mV) reversal potentials. (c) The arachidonic acid-activated currents in STIM1-stable cells expressing the concatenated 31311 heteropentamer are completely blocked by $5 \mu\text{M}$ gadolinium (Gd^{3+}). (d) Activation of the increased arachidonic acid-induced currents seen on expression of the 31113 and 31311 pentamers is entirely dependent on STIM1 in the plasma membrane. Shown are inward current density at -40 mV in STIM1-stable cells treated with siRNA to STIM1, and transfected with a wild-type STIM1 (white bar), together with the same siRNA-treated cells expressing either the 31311 or 31113 concatenated pentamer, as indicated, after transfection with either an siRNA-resistant wild-type STIM1 (black bars) or with an siRNA-resistant glycosylation mutant STIM1 (gray bars). Reproduced from Mignen O, Thompson JL, and Shuttleworth TJ (2009) The molecular architecture of the arachidonate-regulated Ca^{2+} -selective ARC channel is a pentameric assembly of Orai1 and Orai3 subunits. *Journal of Physiology* 587: 4181–4197.

of different cell types, and their potential to provide a unique, store-independent, route for agonist-activated Ca^{2+} entry in such cells is clear.

An additional point to note is that the simple demonstration of a dependence on STIM1 and/or Orai1 for any particular cellular response or effect is not sufficient to definitively conclude that such a response involves store-operated Ca^{2+} entry, or the activity of CRAC channels, as both STIM1 and Orai1 proteins are also essential for the activity of the ARC channels. Consequently, proposals to target these proteins therapeutically as a means of specifically affecting CRAC channel activity are not valid. In addition, it is clear that caution must be applied to studies involving the simple overexpression of individual components of the overall ARC channel complex and its regulators. Thus, we already know that the ARC channel requires both Orai1 and Orai3 subunits for its formation, but the ability of these subunits to assemble in the appropriate

manner would likely be profoundly impacted by the relative levels of expression of either of these components – resulting channels of unknown composition and/or stoichiometry. In addition to Orai1 and Orai3, the ARC channels also require adequate STIM1 in the plasma membrane for their activation. Limitations in any one of these components would be expected to have significant impact on overall channel activity. Moreover, as discussed above, studies have shown that ARC channel activity is also dependent on a balance between a PKA-dependent phosphorylation which requires an AKAP and a calcineurin-mediated dephosphorylation. Consequently, factors that influence the balance between these opposing actions would likely profoundly impact the resulting entry of Ca^{2+} through the channels, with the obvious effects on the resulting agonist-induced Ca^{2+} signals generated in the cell.

As discussed above, the specific properties and features of the ARC channels suggest that they are uniquely suited to

provide a critical role in agonist-activated Ca^{2+} entry, particularly under conditions where the sustained depletion of intracellular Ca^{2+} stores is either unlikely or inappropriate. Consistent with this, the activity of these channels has been shown to play a key role in important physiological responses of diverse cell types, including the insulin-secreting β cells, and the exocrine cells of the parotid and pancreas. Undoubtedly, as more groups begin to study these channels, additional roles will be revealed.

See also: [Bioenergetics: ORAI Store-Operated Calcium Channel; Store-Operated Calcium Channels.](#)

Further Reading

- Dziadek MA and Johnstone LS (2007) Biochemical properties and cellular localisation of STIM proteins. *Cell Calcium* 42: 123–132.
- Manji SSM, Parker NJ, Williams RT, et al. (2000) STIM1: A novel phosphoprotein located at the cell surface. *Biochimica et Biophysica Acta* 1481: 147–155.
- Mignen O and Shuttleworth TJ (2000) I_{ARC} , a novel arachidonate-regulated, noncapacitative Ca^{2+} entry channel. *Journal of Biological Chemistry* 275: 9114–9119.
- Mignen O, Thompson JL, and Shuttleworth TJ (2003) Ca^{2+} selectivity and fatty acid specificity of the noncapacitative, arachidonate-regulated Ca^{2+} (ARC) channels. *Journal of Biological Chemistry* 278: 10174–10181.
- Mignen O, Thompson JL, and Shuttleworth TJ (2005) Arachidonate-regulated Ca^{2+} -selective (ARC) channel activity is modulated by phosphorylation and involves an A-kinase anchoring protein. *Journal of Physiology* 567: 787–798.
- Mignen O, Thompson JL, and Shuttleworth TJ (2007) STIM1 regulates Ca^{2+} entry via arachidonate-regulated Ca^{2+} -selective (ARC) channels without store-depletion or translocation to the plasma membrane. *Journal of Physiology* 579: 703–715.
- Mignen O, Thompson JL, and Shuttleworth TJ (2008) Both Orai1 and Orai3 are essential components of the arachidonate-regulated Ca^{2+} -selective ARC channels. *Journal of Physiology* 586: 185–195.
- Mignen O, Thompson JL, and Shuttleworth TJ (2009) The molecular architecture of the arachidonate-regulated Ca^{2+} -selective ARC channel is a pentameric assembly of Orai1 and Orai3 subunits. *Journal of Physiology* 587: 4181–4197.
- Shuttleworth TJ (1999) What drives calcium entry during $[\text{Ca}^{2+}]_i$ oscillations? Challenging the capacitative model. *Cell Calcium* 25: 237–246.
- Shuttleworth TJ (2009) Arachidonic acid, ARC channels, and Orai proteins. *Cell Calcium* 45: 602–610.
- Shuttleworth TJ, Thompson JL, and Mignen O (2004) ARC channels – a novel pathway for receptor-activated calcium entry. *Physiology* 19: 355–361.
- Shuttleworth TJ, Thompson JL, and Mignen O (2007) STIM1 and the noncapacitative ARC channels. *Cell Calcium* 42: 183–191.
- Wolf BA, Turk J, Sherman WR, and McDaniel ML (1986) Intracellular Ca^{2+} mobilization by arachidonic acid. Comparison with myo-inositol 1,4,5-trisphosphate in isolated pancreatic islets. *Journal of Biological Chemistry* 261: 3501–3511.
- Yeung-Yam-Wah V, Lee AK, Tse FW, and Tse A (2010) Arachidonic acid stimulates extracellular Ca^{2+} entry in rat pancreatic beta cells via activation of the noncapacitative arachidonate-regulated Ca^{2+} (ARC) channels. *Cell Calcium* 47: 77–83.

Arf Family

G Pacheco-Rodriguez, J Moss, and M Vaughan, Cardiovascular and Pulmonary Branch of the National Heart, Lung, and Blood Institute in Bethesda, Bethesda, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Domain Element of protein molecular structure, with or without known function, which is conserved among different proteins and/or organisms.

GAP Guanine nucleoside triphosphatase (GTPase)-activating protein that accelerates hydrolysis of GTPase-bound guanine nucleoside triphosphate (GTP) to generate inactive GTPase-guanine nucleoside diphosphate (GDP).

GEP Guanine nucleotide-exchange protein or factor (GEF) that accelerates replacement of GTPase-bound GDP by GTP.

GTPase Protein that catalyzes hydrolysis of GTP to yield inorganic phosphate and GDP.

Vesicle Membrane-enclosed intracellular structure, 100–200 nm in diameter, with or without visible coat structure, that can be a vehicle for the transport of molecules to support many kinds of cell functions, for example, secretion, endocytosis, migration, and proliferation.

Molecular Characteristics of Arfs

Arfs or adenosine diphosphate (ADP)-ribosylation factors were first identified, purified, and characterized because of their ability to enhance the cholera toxin-catalyzed ADP-ribosylation of G α s, the guanine nucleoside triphosphate (GTP)-binding protein activator of adenyl cyclase. Arf action in cells requires its controlled alternation between the inactive guanine nucleoside diphosphate (GDP)-bound and active GTP-bound forms. This property enables Arfs to regulate enzyme activities (e.g., phospholipase D (PLD) and phosphatidylinositol 4-phosphate 5-kinase) and to participate in the assembly, organization, and function of multimolecular complexes that affect or regulate dynamic processes of cytoskeletal structure and intracellular transport.

Mammalian (human) ARFs are grouped into three classes, with ARF1, 2, and 3 in class I, ARF4 and 5 in class II, and ARF6 in class III, based on molecular size, amino acid sequence, and gene structure. Yeast *Saccharomyces cerevisiae* has one class III and two class I Arfs. The Human Genome Nomenclature Committee described the Arf family as comprising Arfs, Arls, ARD1 (Trim23), and Sar1. Amino acid sequences of Arfs are highly conserved among species with >65% identity among human, yeast, and *Giardia* Arfs. All Arfs contain the specific sequences and amino acids that serve in GTP binding and hydrolysis. One of the hallmarks of these molecules is the presence of the Rossman fold, a β - α - β structure responsible for nucleotide binding. The availability of high-resolution structures of Arfs permits identification of other regions, such as the internal β -sheet, that distinguish Arfs from other subfamilies of GTPases. Arf structure also determines the specificity of its interactions with other molecules, and thereby its function. Both Arf and SAR1 protein structures differ from those of all other c. 20-kDa GTPases in the presence of an N-terminal extension, which in Sar1 proteins lacks the N-myristoylated glycine present in all Arfs. Fatty acid modification plus the additional amphipathic helix at the N-terminus facilitates membrane association of Arf proteins via interaction with phospholipids as well as proteins. Recently, N-terminal acylation of Arls, which contributes to their interactions with membranous structures, was shown.

ARF-Related Proteins

In both structure and function, Arfs resemble the several subfamilies of c. 20-kDa Ras-like GTPases that include Rho, Rac, Rab, and Rap. Each of these proteins, like the Arfs, has special, specific roles in cells. Many of their actions, effects, and regulatory mechanisms are similar, and may seem to overlap, perhaps because all of these GTPases appear to function by interacting with numerous other molecules, simultaneously or sequentially, to modify their activities in a manner dependent on whether GDP or GTP is bound. Dimerization of ARF1 has been described but differs from that of proteins activated by nucleotide-dependent dimerization (GADs).

Among the Arf-related proteins, are Arf-like proteins or Arls that appear to be more diverse in function and, as a group, differ more in structure than do the Arfs. Sar1 proteins are most similar to Arfs in both function and structure, and are, therefore, included in the Arf-family. The 181-amino-acid sequences of human ARF1 and ARL1 are 56% identical. Other Arls differ much more, and some Arls have important intranuclear roles (e.g., in nucleolus). There is also an ARP (ARF-related protein) and the intriguing 64-kDa ARD1 (ARF domain protein 1), with an 18-kDa Arf domain at the C-terminus. Structurally, ARD1 is a member of the tripartite motif (TRIM) family that exhibits E3 ubiquitin ligase activity and is apparently auto-ubiquitinated *in vitro* after addition of purified recombinant E1, an appropriate E2, ubiquitin, and adenosine triphosphate.

Arf Actions

With GTP bound, all Arfs are activators of the cholera toxin ADP-ribosyltransferase, which is secreted by *Vibrio cholerae*. They also activate the very similar *Escherichia coli* heat-labile enterotoxin (LT) that causes travelers' diarrhea. Allosteric activation of the cholera toxin A subunit (CTA), which is the ADP-ribosyltransferase enzyme, has been repeatedly shown *in vitro*. The effect of Arfs on CTA action in cells, however, is less clear. The capacity to activate CTA and LT distinguishes

Arfs from other GTPases, including the Arf-like Arl proteins. The biochemical and/or catalytic behavior of Arfs can be markedly influenced by specific phospholipids.

Arf-GTP can also activate mammalian PLD and phosphatidylinositol kinases that act on phospholipids critical in cell signaling. PLD-catalyzed hydrolysis of phosphatidylcholine produces phosphatidic acid, a molecule that alters the physical properties of membranes and also binds specific receptors to initiate signaling. Phosphorylation of phosphatidylinositol 4-phosphate by an Arf-activated kinase likewise contributes to localized modification of membrane composition and, thus, the molecular interactions that can be vitally important in properly controlling cell functions.

The Arf molecule differs dramatically in shape depending on whether GDP or GTP is bound. Arf activation by GTP binding results from movement of the N-terminal helix and its N-myristoylated glycine away from a position close to the Arf molecule surface, where it rests when GDP is bound. This alters, of course, its interaction with other molecules and enables Arf-GTP to bind membranes more effectively than does Arf-GDP (Figure 1). Membrane-bound active Arf-GTP then recruits adaptor and coat proteins to initiate vesicle formation and cargo loading. Arf function in vesicular trafficking probably involves interactions with phospholipids, as well as with proteins, whether the result is the activation of an enzyme, for example, PLD, or initiation of vesicle formation

when membrane-bound Arf-GTP begins assembly of proteins and lipids that will become vesicle coat and content. Among three types of coated vesicles, the Coat Protein Complex I (COPI)- and clathrin-coated structures require, perhaps most often, class I Arfs.

Regulators of ARF Activity

Activation of ARF requires the replacement of bound GDP by GTP, which without specific stimulus occurs very slowly, and is regulated by guanine nucleotide-exchange proteins, or GEPs, that accelerate GDP release and thereby GTP binding (Figure 1). Turn-off (Arf inactivation) results from hydrolysis of bound GTP to generate Arf-GDP, after interactions with a GTPase-activating protein or GAP (Figure 1). Several Arf GEP molecules have multiple interactions with proteins of diverse functions, not all of which have been related to their critical roles as regulatory GTPases in vesicular trafficking pathways. All GEP molecules contain an ~190-amino-acid Sec7 domain, which is responsible for specificity of Arf binding and acceleration of release of Arf-bound GDP that is rapidly followed by binding of GTP (with an intracellular concentration higher than that of GDP).

GEP families are distinguished by overall molecular size and structures of functional sequences outside of the Sec7 domain, as well as by sensitivity of Arf-activation to inhibition by brefeldin A (BFA). BFA is a fungal fatty acid metabolite first used experimentally as an inhibitor of viral replication, and later shown to inhibit protein secretion via specific pathways of intracellular vesicular trafficking. BFA-inhibited GEPs include mammalian BIG1, BIG2, BIG3, and GBF1, as well as the yeast Sec7, Gea1, Gea2, and the *Arabidopsis* GNOM. The structures of molecular complexes formed by ARF GEPs and Arfs with and without BFA and the unusual complex of ARF-GDP associated with BFA, and fostered the idea of a specific GEP effectively locked by the additional interactions with BFA. Resulting uncompetitive BFA inhibition, which suggested development of novel therapies, is reversed by decreasing BFA concentration and release of complex-bound BFA (Figure 1).

Among several known BFA-insensitive GEPs, the family of four human c. 47-kDa cytohesins appears to be the largest. These molecules contain a central Sec7 domain responsible for Arf activation, followed by a pleckstrin homolog (PH) domain, which binds phosphatidylinositol phosphates with notable specificity, contributing to intracellular localization. Cytohesins are of interest as cell-extracellular matrix (ECM) adhesion molecules and regulators of cell signaling. Cytohesin-1 had been identified and cloned because of its function in adhesion of lymphocytes to the extracellular matrix (hence the name) before its role in ARF activation was recognized. The existence of many more GEPs, for example, than the six mammalian (five human) ARFs, is surely critical for accomplishing specific and precise receipt, regulation, integration, and transduction of diverse signals from multiple sources at sites throughout the cell, as well as the adjacent extracellular milieu. Complexity of GEP structures emphasizes the importance of accurate, timely, and spatially limited coordination of cell functions.

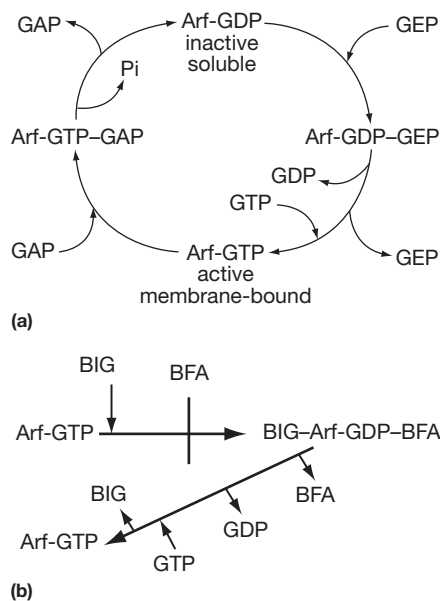


Figure 1 (a) Cyclic Arf activation and inactivation catalyzed, respectively, by GEPs and GAPs. Active, GTP-bound Arf, the product of GEP-catalyzed replacement of bound GDP with GTP, associates with membranes, where it recruits adaptor and coat proteins. GAP-accelerated hydrolysis of Arf-bound GTP yields inactive Arf-GDP, which has a lower affinity for membranes and dissociates more readily. ARF6-GDP differs in structure from other Arf-GDPs sufficiently that it tends to be membrane-associated throughout the GTPase cycle. This difference also distinguishes it from class I and II Arfs in GEP interactions. (b) In the BFA-sensitive GEPs, BIG1, BIG2, and GBF1, specific amino acids within the Sec7 domain allow them to form inactive ternary complexes with BFA when Arf-GDP is bound.

Unlike some other ARF GEPs, FBOX8 contains an E3 ligase domain that uses ARF6 as substrate. Exchange Factor for Arf 6 (EFA6) and GEP100/Brefeldin-A-resistant ArfGEP (BRAG2) also contribute to the regulation of ARF6 guanine nucleotide binding, affecting molecular adhesion, nuclear metabolism, and exocytosis in eukaryotic cells. Regulation of Arf inactivation similarly involves multiple families of GAPs that are, such as GEPs, necessary for integrity of Arf GTPase cycle. These proteins contain characteristic zinc-finger elements that accelerate hydrolysis of ARF-bound GTP via a critical arginine residue. The distinctive GAPs, like GEP molecules, contain multiple domains specialized for different functions and interactions, for example, with cytoskeletal structures. Three types of Arf GAPs, (1) GAPs, (2) G-protein-coupled receptor (GPCR)-kinase-interacting proteins (GITs), and (3) Arf-GAP with Src homology 3, ankyrin repeat and PH domains (ASAPs), are recognized. Multiple, individual domains of the ASAP molecules including PH, anchoring, proline-rich, and SH3, enable them to localize at appropriate intracellular sites and to interact specifically with numerous additional molecules to achieve diverse effects on function.

See also: **Cell Architecture and Function:** Rho GTPases and Actin Cytoskeleton Dynamics; **Signaling:** Phosphatidylinositol Bisphosphate and Trisphosphate; Phospholipase D; Ran GTPase; Small GTPases.

Further Reading

- Bonifacino JS and Lippincott-Schwartz J (2003) Coat proteins: Shaping membrane transport. *Nature Reviews Molecular Cell Biology* 4: 409–414.
- Bui QT, Golinelli-Cohen MP, and Jackson CL (2009) Large Arf1 guanine nucleotide exchange factors: Evolution, domain structure, and roles in membrane trafficking and human disease. *Molecular Genetics and Genomics* 282: 329–350.
- Gillingham AK and Munro S (2007) The small G proteins of the Arf family and their regulators. *Annual Review of Cell and Developmental Biology* 23: 579–611.
- Inoue H and Randazzo PA (2007) Arf GAPs and their interacting proteins. *Traffic* 8: 1465–1475.
- Jackson CL and Casanova JE (2000) Turning on ARF: The Sec7 family of guanine-nucleotide-exchange factors. *Trends in Cell Biology* 10: 60–67.
- Kahn RA (2009) Toward a model for Arf GTPases as regulators of traffic at the Golgi. *FEBS Letters* 583: 3872–3879.
- Kahn RA, Cherfills J, Elias M, et al. (2006) Nomenclature for the human Arf family of GTP-binding proteins: ARF, ARL, and SAR proteins. *Journal of Cell Biology* 172: 645–650.
- Kolanus W (2007) Guanine nucleotide exchange factors of the cytohesin family and their roles in signal transduction. *Immunol Reviews* 218: 102–113.
- Moss J and Vaughan M (1998) Molecules in the ARF orbit. *Journal of Biological Chemistry* 273: 21431–21437.
- Ogasawara M, Kim SC, Adamik R, et al. (2000) Similarities in function and gene structure of cytohesin-4 and cytohesin-1, guanine nucleotide-exchange proteins for ADP-ribosylation factors. *Journal of Biological Chemistry* 275: 3221–3230.
- Pacheco-Rodriguez G, Morinaga N, Noda M, Moss J, and Vaughan M (2003) Activation of cholera toxin and *E. coli* heat-labile enterotoxin (LT) by ARF. In: Kahn RA (ed.) *Small GTPases*, vol. 1, ch. 10, pp. 209–222. Dordrecht: Kluwer.
- Pelham HR and Rothman JE (2000) The debate about transport in the Golgi – two sides of the same coin? *Cell* 102: 713–719.
- Pucadyil TJ and Schmid SL (2009) Conserved functions of membrane active GTPases in coated vesicle formation. *Science* 325: 1217–1220.
- Randazzo PA, Nie Z, Miura K, and Hsu VW (2000) Molecular aspects of the cellular activities of ADP-ribosylation factors. *Science STKE* 21: 1–15.

Aspartic Proteases

B M Dunn, University of Florida College of Medicine, Gainesville, FL, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 123–127, © 2004, Elsevier Inc.

Glossary

Active site A portion of a protein where the catalytic amino acids or associated groups are located and where a binding site for substrate exists.

Chemotherapy Use of compounds to kill disease-causing cells in the human body.

Hemoglobin The oxygen-carrying molecule of the bloodstream found in red blood cells.

Inhibitor A molecule that binds to the active site of an enzyme to block the catalytic activity.

Proteolytic enzyme A protein that has the ability to digest other proteins by cleaving peptide bonds to produce smaller fragments.

Substrate A molecule acted upon by an enzyme to cause a chemical change.

Three-dimensional structure The detailed arrangement of atoms in space for a molecule, usually determined by X-ray crystallography.

Transition-state analog A molecule with a chemical structure that resembles the geometry of the state between a substrate and the products of a reaction.

Zymogen A precursor form of an enzyme, usually larger in size due to the addition of extra amino acid residues; it is also known as a proenzyme.

The Spectrum of Aspartic Proteases in Biology

Historically, aspartic peptidases were first found in the digestive tract of many animals. Pepsin is found in the stomach of most animals with an acidic phase of digestion. A related enzyme, gastricsin, is also present in the stomach and, in some animals, it is the dominant enzyme. Chymosin, a similar enzyme from the fourth stomach of the calf, is responsible for clotting milk when the young animal suckles from its mother. This helps retain the semisolid form of milk proteins in the digestive track for further processing. The presence of an activity was inferred in ancient times when milk was placed in a sack made from the stomach of an animal so that the milk could be transported on long journeys. The travelers found the milk had clotted and realized that something from the stomach lining was causing this to happen. Chymosin is now used to initiate the production of cheese. In some countries, farmers discovered that certain flowering plants also contained a substance that caused milk to clot and these flowers are used in what is termed 'artisanal cheese making'. Other plant enzymes may be involved in digestion of stored seed protein to provide nourishment for a growing seedling. In the world of fungi, many aspartic peptidases have been found. For example, the common baker's yeast, *Saccharomyces cerevisiae*, contains several enzymes in an intracellular organelle called the food vacuole; one of these enzymes is yeast proteinase A.

Due to the progress of genome sequencing projects from 1990 to the present, sequence information is now available for 354 members of the aspartic protease family. They can be related by sequence identities as seen in [Figure 1](#). Representative examples are given for enzymes from animals, plants, fungi, protozoa, insects, and worms. Furthermore, within one organism, different members of the family play different roles. As an example, humans have two enzymes in the

stomach that are involved in the digestion of protein in the diet. In addition, an enzyme in the bloodstream, renin, is involved in regulation of blood pressure by converting a precursor molecule, angiotensinogen to angiotensin I. Subsequent conversion of angiotensin I to angiotensin II has a large effect upon blood pressure, as the latter molecule causes constriction of blood vessels and increases in blood pressure. Within cells, several aspartic proteases are found to play important roles. Cathepsin D in the lysosomes of cells acts to activate other enzymes, prohormones, and growth factors. Cathepsin E, which is found mainly in cells of the lymphoid system, seems to have a function in the response of cells to immunological stimuli. Memapsin, also known as beta-site APP cleaving enzyme (BACE), is involved in the conversion of the β -amyloid precursor protein to a form that can aggregate and cause fibrils that are found in Alzheimer's disease. Several enzymes known as napsins have recently been discovered in humans. Their function is unknown at this point in time. In the genomes of the plant, *Arabidopsis thaliana*, the nematode, *Caenorhabditis elegans*, and the fruit fly, *Drosophila melanogaster*, a similar pattern of multiple potential functions has emerged, as multiple sequences with identities to the aspartic protease family have also been identified.

Structure

Pepsin from the stomach of the pig was one of the first proteins to be crystallized; however, the three-dimensional (3D) structure of three fungal enzymes was reported before the structure of pig pepsin was solved. Since the 1970s several hundred structures have been solved by X-ray diffraction. At this point in time, we can state with certainty that all proteins that have sequence homology of over 40% with the aspartic protease family will have the same overall structure.

Overall Structure

The structure of a typical aspartic protease is shown in **Figure 2** in several views. The structure can be divided in two ways. First, there is a twofold axis of symmetry down the middle of the protein as seen in **Figure 2(a)**. The first 165 amino acids of the protein (the N-terminal domain) are an independent folding unit that is mostly composed of β -strands and very little α -helical structure. The last 165 amino acids (the C-terminal domain) are also an independent folding unit with exactly the same folding pattern. The active site cleft is located between the two domains and on one side of the molecule. Each domain contributes to the surface of the cleft, which either binds substrates for cleavage or acts as inhibitors that block cleavage. The two halves of the enzymes are believed to have arisen by a gene duplication event before divergence into the different forms shown in **Figure 1**. A second way to analyze

the structure has the same division as described above, and also considers the β -sheet shown at the bottom of **Figure 2(b)** as an independent unit comprised of β -strands from each of the two domains. In either analysis, all the known structures are highly similar in structure. Some variation is seen in the relative orientation of the two (or three) domains with respect to each other. In addition, some variability is provided by the length of some of the β -hairpin loops that impinge on the active-site cleft. This provides diversity in the active site-binding characteristics of each enzyme, which could be involved in the specific function of each member of the family.

Catalytic Residues

Each of the two domains (N-terminal and C-terminal) provides one aspartic acid; these two amino acids are located in

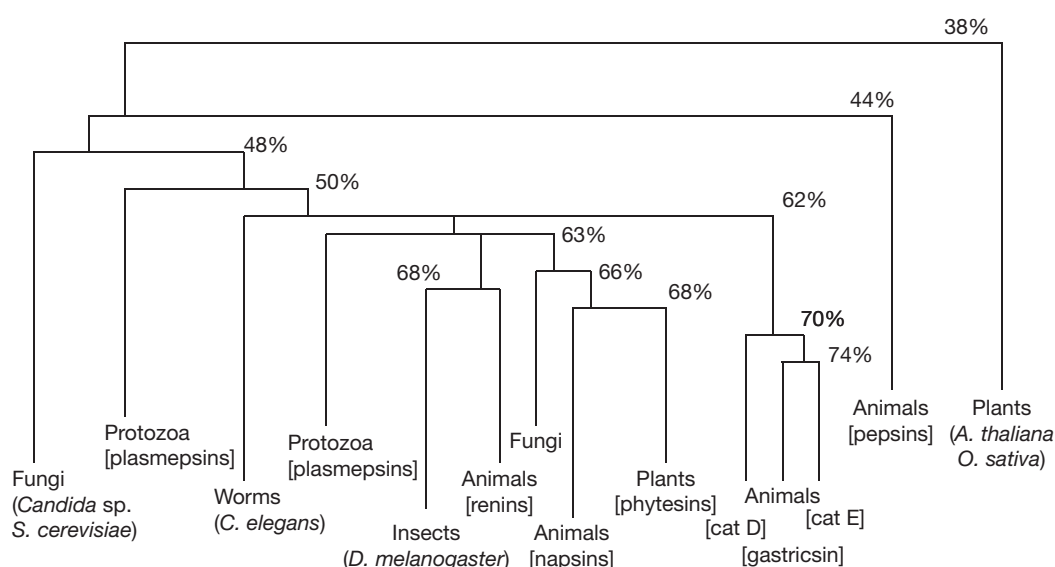


Figure 1 Diagrammatic representation of the relationship among several major classes of aspartic proteinases. The number at the top of each division point is the percentage identity between all species below the line. Organism names are given in italics inside parentheses and specific enzyme names are given inside square brackets. Note that different organisms are seen in different points in the diagram. This is due to the fact that different enzymes within one species play different functional roles. This figure was prepared using the program Pymol.

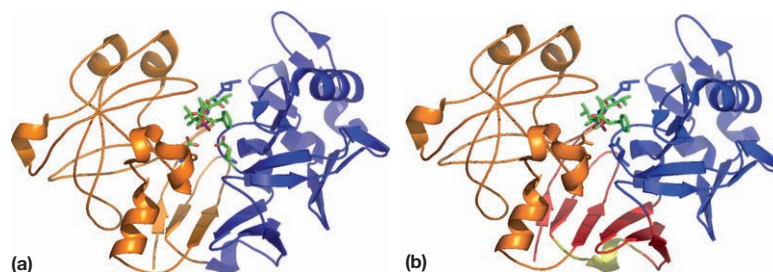


Figure 2 Representation of the 3D structure of a typical aspartic protease. The flat arrows represent β -strand structure and the spiral shapes represent α -helices. These elements are from the backbone only and do not show the complete structure of the protein. (a) View of the enzyme looking end-on into the active site cleft at the top. In this representation, the N-terminal domain is colored orange and the C-terminal domain is colored blue. The two catalytic aspartic acid residues are shown in sticks at the bottom of the active site cleft. Above these amino acids one can see an inhibitor molecule bound in the active site. (b) Same orientation as (a), with the N-terminal domain colored orange, the C-terminal domain colored blue, and the β -sheet at the bottom of the molecule colored red. A small connecting loop between the first three strands of the red β -sheet and the second three strands is colored yellow. This figure was prepared using the program Pymol.

the 3D structure at the bottom of the active site cleft; the carboxylic acid functional group of each of the aspartic acid side chains is within 3–5 Å. In the enzyme without bound ligands, a water molecule is hydrogen bonded to the two aspartic acid residues and is the nucleophile that attacks the peptide bond of a substrate held within the active site cleft.

Flap Structure

Another significant feature of all aspartic proteases is a β -hairpin that hangs over the active site cleft. This is frequently referred to as the flap and movement of this is important in the binding of substrates and inhibitors in the active site. Amino acid differences on the flap, when comparing different members of the family, are important in the different activities exhibited by the enzymes.

Proenzyme Structure

Most aspartic proteases are synthesized within cells as inactive precursors known as zymogens or proenzymes. In all known cases, this involves an N-terminal extension of between 35 and 125 amino acids. In cases where the 3D structure has been determined for some of the proenzyme forms, the extra amino acid sequence acts to either block the active site cleft or to pull the two domains apart from one another; in either case, this prevents activity. For many cases, conversion from the proenzyme to the mature enzyme form is self-catalyzed and occurs when the proenzyme moves from a neutral pH condition to a compartment of lower pH. Removal of the N-terminal extension results in a decrease in molecular weight of between 4 and 15 kDa. In a few known cases, the conversion requires a second enzyme.

Diseases

Efforts to understand the structure and function of aspartic proteases have been stimulated by the discovery that many disease processes involve the activity of these enzymes.

Hypertension

As mentioned above, renin is involved in a cascade of reactions that leads to the elevation of blood pressure. While this is a normal function that can increase blood pressure in times of activity of an individual, when the levels of renin activity are slightly out of balance, the effects can be dangerous to health. Studies in animals have shown conclusively that giving a selective renin inhibitor, that is, a compound that will bind to and block the activity of renin without affecting the other aspartic proteases of the human body, has a dramatic effect on blood pressure levels. This area of investigation was extremely important to the study of the whole field of aspartic proteases, as it allowed scientists to develop the tools and strategies to develop selective inhibitors. This has paid dividends in the development of antiviral inhibitors as well as in programs searching for inhibitors against many other family members.

Cancer

High levels of cathepsin D have been observed in areas surrounding tumors in both humans and animals. This has led to the hypothesis that cathepsin D is involved in the invasiveness of cancer cells. However, cathepsin D activity is normally highest at pH values well below that of tissue. Cathepsin D exhibits its activity in the lysosome, where the pH level is believed to be less than 6. In addition, cathepsin D is released when cells break open, so that the excess levels seen in necrotic tissues may be due to cellular decay. Furthermore, in some conditions, excess levels of cathepsin D are due to overloads of the secretory pathway so that the proenzyme form of cathepsin D is released rather than the mature form. Cathepsin D or the proenzyme form may act as a growth factor to stimulate the growth of cancerous cells.

Acquired Immunodeficiency Syndrome

An aspartic protease is part of the causative agent for acquired immunodeficiency syndrome (AIDS), the human immunodeficiency virus (HIV). This protease acts to cut apart a viral polyprotein so that new virus particles can assemble. The HIV protease was the second target against which new antiviral drugs were developed. These compounds have had a major impact on therapy for AIDS.

Malaria

The malaria parasite, *Plasmodium falciparum*, lives during part of its complicated life cycle within human erythrocytes where it degrades hemoglobin for two reasons: one, to provide nutrients in the form of free amino acids for new protein synthesis, and, two, to create space for growth of the parasitic cell. Digestion of hemoglobin is initiated by the action of aspartic proteases. Following the sequencing of the *P. falciparum* genome, it is now known that 10 different aspartic protease genes are present. Four of these are expressed and the proteins are located within a special organelle, the food vacuole. This organelle is equivalent to the lysosome of mammalian cells and a similar vacuole of *S. cerevisiae*. Complete digestion of hemoglobin requires the coordinated activity of cysteine proteases and metallo-proteases as well as the aspartic proteases. Inhibitors of aspartic proteases are able to kill the parasite in culture studies.

Fungal Infections

A number of fungi are known to cause diseases in humans and plants. For example, *Candida albicans* is a common commensal organism, living within the human body but kept in check by a healthy immune system. In situations where the immune system is compromised, due to chemotherapy for cancer or immune suppression to avoid transplant rejection or due to suppression by diseases such as AIDS, *C. albicans* can become a life-threatening systemic infection. The fungus is also known to have a number of aspartic proteases, including several that are secreted into the extracellular environment surrounding the fungal cells. These enzymes are believed to assist in providing nutrients to the fungus by digesting proteins found in the

surrounding environment. Some of these secreted enzymes may work to aid in the invasiveness of the fungus as the infection spreads.

Alzheimer's Disease

The newest target for drug discovery is β -secretase or memapsin. This enzyme acts to cleave the β -amyloid precursor protein at a specific bond, generating a fragment that can become, with further modification by the so-called gamma secretase, a peptide that forms amyloid fibrils. This condition leads to neurofibrillary plaques in the brain, which are a hallmark of Alzheimer's disease. Due to the unique specificity of memapsin, development of specific inhibitors is possible and new therapies for this condition may develop in the next few years.

Inhibitors

Naturally Occurring Inhibitors

Unlike the serine protease family, naturally occurring inhibitors of aspartic proteases are relatively rare. A few have been described in detail and their structural information obtained. By investigating the mechanism of inhibition of these protein inhibitors, it is hoped that new insights into inhibitor design will be found.

Pepstatin

Culture filtrates of the organism *Streptomyces* have been found to contain many interesting compounds. Pepstatin, discovered around 1970, was found to strongly inhibit pepsin and other members of the aspartic protease family. This compound is a relatively nonspecific inhibitor, as it seems to block most members of the aspartic protease class. Pepstatin contains an unusual amino acid, 3-hydroxy-4-amino-6-methylheptanoic acid. The 3-OH group binds to the two catalytic aspartic acids to form a transition-state analog, which provides very tight binding. Because all aspartic proteases have the identical catalytic mechanism, pepstatin is an excellent inhibitor. This compound has been used in many studies of the 3D structure of members of the aspartic protease family and thus provides convenient comparisons between these enzymes. Many synthetic inhibitors, mentioned below, utilize the same concept in the design of selective inhibitors.

PI-3 pepsin inhibitor of Ascaris

The nematode, *Ascaris lumbricoides*, lives within the human digestive tract for a significant part of the life cycle. The digestive tract contains many proteolytic enzymes, including aspartic proteases. The nematode produces several proteins, each of which is able to bind to and inhibit a specific host enzyme. Several inhibitors have been purified to homogeneity, including PI-3. From a purification of the aspartic protease inhibitors, the third fraction had the highest activity in assays of the ability to inhibit pepsin, thus leading to the designation, PI-3. PI-3 is a

protein of 149 amino acid residues. It inhibits pepsins and gastricsins from humans and pigs, and has been shown to inhibit cathepsin E from humans. The structure of PI-3 in complex with porcine pepsin has been determined and it was found that the inhibitor is a rigid protein, which makes contact with pepsin at two points, using a total of about 13 out of the 149 residues. The structure of the inhibitor does not change when interacting with pepsin.

IA-3 yeast protease inhibitor of S. cerevisiae

The yeast *S. cerevisiae* synthesizes a 68-residue protein called IA-3, which is a strong and selective inhibitor of yeast proteinase A. The 3D structure of the complex between the two proteins shows that the inhibitor forms a near-perfect α -helix between amino acids 2 and 34 when it binds to the enzyme, while the inhibitor, when free in solution, has little or no organized structure. The IA-3 inhibitor does not inhibit any other member of the aspartic protease class that have been tested so far and, in fact, is cleaved by several of the other enzymes. The transition from an unstructured protein to a tightly folded one is a unique aspect of this inhibitor. This is an example of an intrinsically unstructured protein.

Synthetic Inhibitors

Due to the importance of aspartic proteases in several disease processes, strong efforts have gone into the design and testing of compounds that could become new drugs for the treatment of disease. In many cases, the observations that pepstatin was a potent inhibitor gave the initial clue to inhibitor design. Thus, the presence of a hydroxyl group within the structure of a new compound is frequently coupled with the variation in the amino acid sequence (for peptide-like compounds), or in the structural groups (for nonpeptide compounds) to create a selective inhibitor. Other variations on the central hydroxyl group include replacements for the peptide bond such as $-\text{CH}_2\text{NH}-$, $-\text{PO}_2-\text{CH}_2-$, and $-\text{CHOH}-\text{CHOH}-$. These replacements have proved to be effective in the development of new inhibitors.

See also: Protein/Enzyme Structure Function and Degradation: Aminopeptidases; Amyloid; HIV Protease; Metalloproteases; Secretases; Signaling: Angiotensin Receptors.

Further Reading

Dunn BM (1999) *Proteases of Infectious Agents*. San Diego, CA: Academic Press.
 Dunn BM (2002) Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chemical Reviews* 102: 4431–4458.
 James MNG (1998) *Aspartic Proteinases: Retroviral and Cellular Enzymes*. New York, NY: Plenum.

Relevant Website

<http://merops.sanger.ac.uk> – MEROPS: The Peptidase Database.

ATP in Plant Mitochondria: Substrates, Inhibitors, and Uncouplers

K L Soole and R I Menz, Flinders University, Adelaide, SA, Australia

© 2013 Elsevier Inc. All rights reserved.

Glossary

Alternative oxidase An ubiquinol oxidase located in the inner mitochondrial membrane, which accepts electrons from reduced ubiquinone and reduces oxygen to water. During this process, no protons are translocated across the inner mitochondrial membrane.

Nonphosphorylating pathway A route of electron transfer with no concomitant translocation of protons across the inner mitochondrial membrane; hence, this route of electron flow does not contribute to the proton-motive force.

Nonphosphorylating reduced nicotinamide adenine dinucleotide phosphate (NAD(P)H) dehydrogenase(s) Enzymes that accept electrons from NADH or NADPH and reduce ubiquinone, a mobile lipophilic electron carrier in the inner mitochondrial membrane with no consequent proton translocation across the membrane.

Proton-motive force Refers to the proton gradient that is established across the inner mitochondrial membrane during electron transfer through complexes I, II, and IV, often referred to as $\Delta\mu_{H^+}$ or pmf.

Adenosine triphosphate (ATP) is the energy carrier of the cell and in eukaryotic cells is synthesized via photosynthesis and respiration. Within respiration, there is a low level of ATP synthesis associated with glycolysis in the cytoplasm; however, the majority of ATP is synthesized via oxidative phosphorylation which occurs within mitochondria, specifically via the operation of an electron transport chain (ETC) in the inner mitochondrial membrane. In mammals, flux through the respiratory pathway is tightly regulated by the ATP/adenosine diphosphate (ADP) ratio or adenylate energy charge of the cell. Plants, unlike mammals, must synthesize all of their cellular components. In plants, the presence of a nonphosphorylating pathway in the mitochondrial ETC and an uncoupler protein in the inner membrane overcomes this restriction by adenylate control. In plants, respiration not only is important for energy production but also has a major role in biosynthesis and anabolic reactions, in programmed cell death, and, more recently, has been implicated as stress response signaling sensor.

Mitochondrial ETC in Plants

Sucrose and starch are converted to metabolites that feed into the early steps of glycolysis which occurs in the cytoplasm. The end products of glycolysis in plants can be either pyruvate or malate, the latter being formed by the action of phosphoenol pyruvate carboxylase and malate dehydrogenase, which together bypass pyruvate kinase. Once within the mitochondrial matrix, pyruvate is metabolized by pyruvate dehydrogenase and the enzymes of the citric acid cycle. Being a substrate for mitochondrial malate dehydrogenase, malate can either feed into the citric acid cycle or be metabolized to pyruvate via nicotinamide adenine dinucleotide (NAD)-malic enzyme. There must be a balance between these processes, because of the poor forward equilibrium of malate dehydrogenase, the accumulation of oxaloacetate would prevent further malate metabolism. The net result of these reactions is the production of NADH and FADH₂ (via succinate dehydrogenase). During a

turn of the citric acid cycle, one ATP is produced directly by substrate-level phosphorylation, not guanosine triphosphate (GTP) as in mammalian mitochondria. Matrix NADH is oxidized by the ETC and more ATP is produced via oxidative phosphorylation.

Oxidative phosphorylation occurs via the interaction of large lipoprotein complexes of the ETC and smaller mobile electron carriers found in the inner mitochondrial membrane (Figure 1). NADH is oxidized by complex I, which donates its electrons to a mobile lipophilic electron carrier, ubiquinone. Complex II or succinate dehydrogenase also donates electrons to the ubiquinone pool. Reduced ubiquinone (or ubiquinol) is oxidized by complex III via the Q-cycle and reduces the mobile peripheral protein cytochrome *c*. By interacting with complex IV (cytochrome oxidase), the electrons carried by cytochrome *c* are donated to the terminal electron acceptor, oxygen. During these electron-transfer events, protons are translocated from the matrix side of the inner membrane to the intermembrane space at complexes I, III, and IV, thus establishing the proton-motive force (pmf or $\Delta\mu_{H^+}$). If ADP is bound to the F₀F₁-ATPase in the inner membrane, then protons pass through this complex and the energy within the pmf is used to synthesize ATP. Once ATP is synthesized, it is then exported out of the matrix in exchange for another ADP via the adenine nucleotide translocase. Thus, the flow of electrons and, hence, oxygen consumption are tightly under control of cellular ADP levels. This is called acceptor or adenylate control.

In plant mitochondria, additional protein complexes are found associated with this ETC. They are distinct from complexes I to V, in that they participate in the transfer of electrons from NADH to oxygen, but do not contribute to the pmf. These enzymes are the alternative or nonphosphorylating reduced nicotinamide adenine dinucleotide phosphate (NAD(P)H) dehydrogenases (NDE and NDI), which donate electrons to the ubiquinone pool and the alternative oxidase (AOX), which accepts electrons from reduced ubiquinone (Figure 1).

Evidence for the alternative nonphosphorylating NAD(P)H dehydrogenases comes from the ability of isolated plant

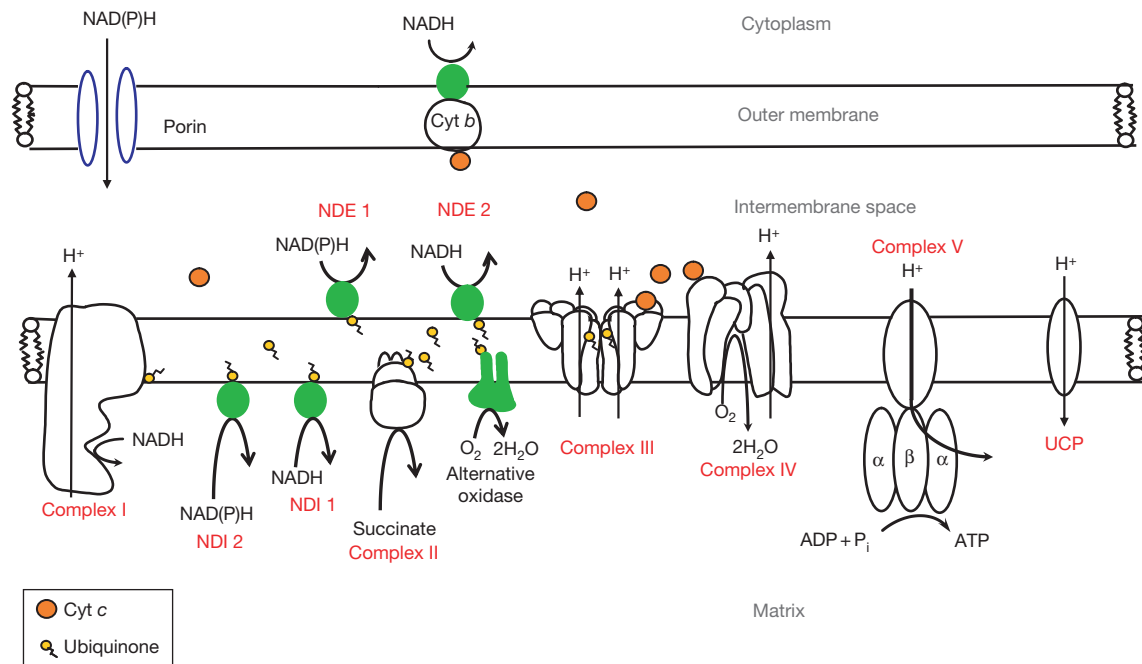


Figure 1 A schematic representation of the routes of NAD(P)H oxidation and the electron transport chain (ETC) of plant mitochondria. NDI refers to internal matrix-facing nonphosphorylating NAD(P)H dehydrogenase, and NDE refers to external cytosolic nonphosphorylating NAD(P)H dehydrogenase. UCP refers to the uncoupling protein. Reproduced from Rasmusson AR, Soole KL, and Elthon TE (2004) Alternative NAD(P)H dehydrogenases in plant mitochondria. *Annual Review of Plant Biology* 54: 23–39.

mitochondria to oxidize externally supplied NAD(P)H, which cannot pass through the inner membrane. The external cytosolic-facing NAD(P)H dehydrogenase (NDE) feeds directly into the ubiquinone pool, as external NAD(P)H oxidation has an ADP/O ratio equivalent to succinate of less than 2.0. Matrix NADH oxidation in plants has a component which is insensitive to inhibition by the complex I inhibitor, rotenone. This oxidation is catalyzed by internal inner mitochondrial membrane transdehydrogenases (NDI). The rotenone-insensitive alternative matrix NADH oxidation has an ADP/O ratio also similar to succinate indicating that NDI also feeds electrons into the ubiquinone pool, bypassing proton pumping at complex I. The AOX is a cyanide and antimycin A-resistant ubiquinol oxidase that catalyzes the reduction of oxygen to water with electrons from ubiquinol and, thus, bypasses the proton pumps at complexes III and IV. Sequence analysis of the genes putatively encoding NADH(P)H dehydrogenases in various plant species has separated this gene family into three subfamilies, NDA, NDB, and NDC, based on their homology to the yeast and bacterial gene sequences. There is functional evidence that links NDA to NDI activity and the NDB gene family to NDE activities. It is as yet unknown whether NDC encodes an NDI- or NDE-like activity.

Thus, it is possible to have electron transport/transfer with no ATP synthesis in plant mitochondria. Therefore, due to the bypasses that exist around key regulatory sites in glycolysis, plant respiration can be totally independent of respiratory control, which gives plants great metabolic flexibility. This flexibility, which results in the loss of adenylate control of respiration by the energy status of the cell, can theoretically

allow the release of biosynthetic intermediates from the respiratory pathway, independent of energy charge of the cell. This role of the nonphosphorylating pathway is hypothesized and has not been clearly demonstrated *in vivo*.

Substrates for ATP Synthesis

The plant mitochondrial ETC can oxidize matrix NADH and NADPH to a much lesser extent. The source of this NADH is from pyruvate dehydrogenase and the citric acid cycle, and plants have the full complement of these enzymes in addition to other unique enzymes. The presence of matrix pools of NADPH is indicative of the mitochondria's anabolic role. NADP-dependent enzymes involved in folate and thymidylate synthesis such as NADP-dihydrofolate reductase and methylene tetrahydrofolate dehydrogenase have been found in plant mitochondria and their continued operation requires the turnover of the NADPH generated via NDI2. Apart from NAD-dependent isocitrate dehydrogenase, there is also another NADP-dependent form in the matrix; however, its role in plant metabolism is not yet clear. Additionally, via the action of a soluble transhydrogenase detected recently in pea leaf matrix, the mitochondrial oxidation of any strictly NAD-linked substrates such as pyruvate would also be able to produce NADPH.

Another unique enzyme found in plant mitochondria is glycine decarboxylase, which generates NADH in the matrix, is part of photorespiratory cycle and is important in photosynthetic metabolism. It is accepted that the mitochondrial

ETC operates in the light, and that ATP synthesis in the light is not exclusively generated from photosynthesis. Evidence indicates that glycine oxidation is linked to a nonphosphorylating NADH dehydrogenase, such as NDI. Recently, it has been shown that NDI is regulated at the transcriptional level by light, at least in potato leaf, suggesting a link between nonphosphorylating enzymes and photorespiratory metabolism.

Plant mitochondria also have the capacity to oxidize cytosolic NADH and NADPH via the enzymes NDE1 and NDE2. Thus, ATP synthesis can also occur from the recycling of cytosolic NAD(P)H.

Inhibitors of ATP Synthesis

There are a multitude of inhibitors that act on specific components of the ETC and the use of many of these has been invaluable in elucidating the composition of the respiratory chain in different species (Table 1).

The two most common and specific complex I inhibitors are rotenone and piericidin A, both inhibitors block electron flow between the final iron-sulfur center and ubiquinone. However, they display different inhibition kinetics and it is postulated that the transfer of electrons to quinone may involve more than one quinone-binding site, similar to those observed in complex III. Many other inhibitors of mammalian complex I have been discovered, such as aurachin A and B, thiagazole, phenoxan, and aureothin, and many are likely to inhibit plant complex I; however, their efficacy and specificity have not been demonstrated.

Table 1 Specific inhibitors of plant respiratory enzymes

Respiratory protein	Inhibitor	Site/mode of action
Complex I	Rotenone	Blocks between FeS
	Piericidin A	Cluster N2 to quinone
Complex II	Malonate	Competitive substrate
	Thenoyltrifluoroacetone (TTFA)	Blocks FeS cluster S3
Complex III	Antimycin A	Quinone binding (N side)
	Myxothiazol	Quinone binding (P side)
	Stigmatellin	Quinone binding (P side)
Complex IV	Cyanide	Competitive inhibitors of
	Azide	O ₂ -binding site
	Nitric oxide	
Complex V	Oligomycin	Binds to OSCP-subunit
	Venturicidin	Proton translocation (c-subunit)
NDE1 (external NAD(P)H)	DPI	
	Calcium chelators	
NDE2 (external NADH)	Calcium chelators	
NDI1 (internal NADH)		
NDI2 (internal NAD(P)H)	Calcium chelator	
	DPI	
Alternative oxidase	<i>n</i> -Propylgallate	
	SHAM	

Relatively few inhibitors of plant complex II have been described (Table 1), the most widely used is the competitive substrate inhibitor malonate; however, several new complex II inhibitors that are potential fungicides or pesticides have been reported recently.

The two most significant specific complex III inhibitors are antimycin A and myxothiazol, which are specific to the N-side and P-side quinone-binding sites, respectively, and have been important for developing models of the Q-cycle which is involved in proton translocation by complex III.

Complex IV is inhibited by a variety of competitive inhibitors of the oxygen-binding site, such as cyanide and azide. The most interesting of these is the rapid and reversible inhibition by nitric oxide, which may play an important role in regulation of oxidative phosphorylation.

There are several chemical inhibitors of ATP synthase (Table 1), the most prominent being oligomycin, which is a specific inhibitor of F-type ATP synthases found in a variety of organisms. Apart from the chemical inhibitors, there are also inhibitor proteins, which play a role in regulating the activity *in vivo*. The most characterized is the mammalian inhibitor protein (IF1), whose conformation and inhibitory activity change in response to pH. Proteins with low homology to IF1 have been found in plants, and although they can inhibit ATP synthetase activity, they do not appear to be regulated by pH. More recently, a class of phosphoserine/phosphothreonine-binding proteins, called the 14-3-3's, have been identified in plant mitochondria and shown to regulate the ATP synthetase activity.

There are several inhibitors of the alternative NAD(P)H dehydrogenases, while a few show differential selectivity between the various alternative NAD(P)H dehydrogenases many of these such as platanetin, 7-iodo-acridone-4-carboxylic acid, dicumarol, hydroxyflavaones, and sulfhydryl reagents can inhibit all of the alternative NAD(P)H dehydrogenases. Interestingly, many of these were originally reported as specific inhibitors of particular alternative NAD(P)H activities. The most effective selective inhibitors are calcium chelators such as egtazic acid, which are potent inhibitors of the external and NADPH-utilizing enzymes, which have a calcium requirement, and diphenyleneiodonium, which is more effective at inhibiting NADPH-utilizing enzymes compared to NADH-utilizing enzymes (Table 1). Salicylhydroxamic acid (SHAM) and *n*-propylgallate are the most prominent inhibitors of the AOX, the latter being more specific as SHAM can also effect other enzymes of oxidative metabolism such as peroxidases, lipoxygenase, and xanthine oxidases.

Uncouplers of ATP Synthesis

Chemical Uncouplers

Oxidative phosphorylation in plants is sensitive to chemical uncouplers such as dinitro phenol and carbonyl cyanide *p*-(trifluoromethoxy)phenylhydrazone (FCCP) which dissipate the $\Delta\mu_{H^+}$ by carrying proteins across the inner membrane, thus equilibrating the proton gradient. In heterotrophically grown tobacco-suspensions cells, cellular respiration is increased by the addition of the uncoupler, FCCP; therefore, under these conditions respiration is substrate limited.

Nonphosphorylating Respiration

Plants have the capacity to uncouple respiration and ATP synthesis using naturally occurring enzymes such as the alternative NDE and NDI, and AOX. Therefore, the level of coupled respiration will be dependent on the level and activity of these alternative ETC enzymes.

Alternative NAD(P)H dehydrogenases

In plants, complex I has a much better affinity for NADH than NDI. This suggests that the latter route of electron flow will only be active when the matrix concentration of NADH is high. This has been clearly demonstrated in tissues such as potato, which lose NAD from their mitochondrial matrix during long-term storage. In NAD-depleted mitochondria, the rate of electron flow through NDI was markedly reduced, that is, malate oxidation in isolated mitochondria was totally sensitive to rotenone; however, this could be overcome if the mitochondria were reloaded with NAD from the bathing media. NAD enters via a specific transporter on the inner membrane. It is not clear if this level of regulation occurs *in vivo*. One of the NDI enzymes that uses NADPH as a substrate is calcium dependent and could be regulated via matrix calcium levels. One question is whether these alternative pathways operate as overflow for complex I or operate simultaneously during respiration *in vivo*. In a mutant where expression of one of the NDI enzymes was eliminated, the total respiration was reduced, which suggests that NDI contributes to respiration in presence of ADP along with complex I. Both the NDE enzymes are dependent on calcium for maximal activity. It has been suggested that electron flow through the NDE can be regulated by alteration of the local calcium concentration which can be facilitated by polyamines that occur naturally in plant cells, for example, spermine and spermidine.

Alternative oxidase

For many years, it was thought that the AOX acted as an overflow, only being used when there was an excess of reduced ubiquinone and the cytochrome pathway (via complexes III and IV) was saturated. It is now known that this is not the case, and that AOX and cytochrome pathway can operate simultaneously. Thus, the level of ATP synthesis will rely on the regulation of AOX. AOX can be regulated at both the transcriptional and posttranslational level. AOX exists as a monomer or dimer, with the monomeric form being most active. Further, AOX is activated by organic acids such as pyruvate. Pyruvate is the end product of glycolysis and, thus, this activation can act as a positive feedforward mechanism of control when glycolytic flux is high. AOX is induced upon inhibition of the complexes III and/or IV and also under various environmental stresses. It is thought that the expression of AOX is an attempt by the plant to have a flexible control of ATP synthesis and to maintain growth rate homeo-

stasis. It is also thought to be part of the cell's defense against oxidative stress, as increased flux through the ETC would prevent the accumulation of high-energy electrons in the respiratory pathway, which could react with oxygen to form harmful, reactive oxygen species. More recently, reverse genetics experiments with transgenic plants under- and overexpressing AOX in plants have led to the proposal that a core role of AOX may be to modulate the strength of a stress-signaling pathway from the mitochondrion that controls cellular responses to stress. Thus, AOX could be acting to provide a degree of signaling homeostasis from the mitochondrion, influencing cell-wide stress response.

Uncoupling protein

In addition to AOX, plants also have uncoupling protein (UCP), which mediates proton reuptake across the inner membrane (Figure 1). This activity is activated by free fatty acids and inhibited by pyridine nucleotides. UCP is active during respiration in the presence of excess ADP when the free fatty acid concentration is $\sim 4 \mu\text{M}$, and, thus, can divert energy from oxidative phosphorylation, impacting on the capacity for ATP synthesis. There is a reciprocal regulation between UCP and AOX as free fatty acids inhibit AOX activity, similarly 4-hydroxy-2-nonenal, a product of lipid oxidation, has been shown to inhibit AOX but stimulate UCP. However, a precise understanding of the integration and regulation of nonphosphorylating respiratory pathway, UCP and ATP synthase, remains a major research challenge.

See also: [Bioenergetics: ATP Synthesis: Mitochondrial Cyanide-Resistant Terminal Oxidases; Cytochrome *c*; Cytochrome Oxidases, Bacterial; Respiratory Chain and Adenosine Triphosphate Synthase; Uncoupling Proteins.](#)

Further Reading

- Albury MS, Elliott C, and Moore AL (2009) Towards a structural elucidation of the alternative oxidase in plants. *Plant Physiology* 137: 316–327.
- Hong S and Pedersen PL (2008) ATP synthase and the actions of inhibitors utilized to study its roles in human health, disease, and other scientific areas. *Microbiology and Molecular Biology Reviews* 72: 590–641.
- Jarmuszkiewicz W (2001) Uncoupling proteins in mitochondria of plants and some microorganisms. *Acta Biochimica Polonica* 48(1): 145–155.
- Joseph-Horne T, Hollomon DW, and Wood PM (2001) Fungal respiration: A fusion of standard and alternative components. *Biochimica et Biophysica Acta* 1504: 179–195.
- Moore AL, Albury MS, Crichton P, and Affourtit C (2002) Function of the alternative oxidase: Is it still a scavenger? *Trends in Plant Science* 7(11): 478–481.
- Rasmussen AR, Soole KL, and Elthon TE (2004) Alternative NAD(P)H dehydrogenases in plant mitochondria. *Annual Review of Plant Biology* 54: 23–39.
- Scheffler IE (2001) Mitochondria make a come back. *Advanced Drug Delivery Reviews* 49: 3–36.
- Vanlerberghe GC, Cvetkovska M, and Wang J (2009) Is the maintenance of homeostatic mitochondrial signalling during stress a physiological role for alternative oxidase? *Plant Physiology* 137: 392–406.

ATP Synthesis: Mitochondrial Cyanide-Resistant Terminal Oxidases

J N Siedow, Duke University, Durham, NC, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Aerobic respiration The biochemical process whereby reduced organic compounds are completely oxidized in three stages, glycolysis, the tricarboxylic acid (TCA) cycle, and the mitochondrial electron transfer chain with the free energy released used to drive the synthesis of adenosine triphosphate (ATP).

Alternative pathway The mitochondrial electron transfer pathway that goes via a cyanide-resistant, alternative oxidase (AOX) and transfers electrons from the ubiquinone pool to molecular oxygen without storing any of the released free energy in the form of a proton gradient.

Cytochrome (cyt) pathway The mitochondrial electron transfer pathway that goes from the ubiquinone pool to molecular oxygen via complex III and complex IV (cytochrome *c* oxidase) and is coupled at each complex to

the transport of protons across the membrane to produce a proton gradient that subsequently drives the synthesis of ATP.

Diiron carboxylate protein Any member of a large family of proteins that contain a diiron active site formed by a four-helix bundle, two helices of which provide ligands to the two iron atoms via an E-X-X-H sequence motif and two of which each provide a single carboxylate ligand (i.e., Glu or Asp) to the irons. Family members include the R2 subunit of ribonucleotide reductase and the hydroxylase subunit of methane monooxygenase.

Reactive oxygen species (ROS) Any of several highly reactive chemical species that can be formed following the reduction of molecular oxygen (O_2), including superoxide anion (O_2^-), hydrogen peroxide (H_2O_2) and hydroxyl radical ($OH^\cdot$).

The Uniqueness of Plant Mitochondria

Plant mitochondria, like those of all other eukaryotes, contain a standard set of protein complexes that make up the electron transport chain in the inner mitochondrial membrane, complexes I–IV (Figure 1). This electron transfer chain functions as the third stage of aerobic respiration, whereby reductants reduced nicotinamide adenine dinucleotide (NADH) and reduced flavin adenine dinucleotide ($FADH_2$) generated during operation of the tricarboxylic acid (TCA) cycle in the mitochondrial matrix are oxidized and the resulting electrons transferred to molecular oxygen. The free energy released during this electron transfer is used to translocate protons across the inner membrane from the mitochondrial matrix to the intermembrane space at complexes I, III, and IV, forming an electrochemical proton gradient across the inner membrane. Another inner membrane protein complex, complex V (adenosine triphosphate (ATP) synthase), provides a path for protons to move back into the matrix and, in the process, couples the release of the energy stored in the proton gradient to the synthesis of ATP from adenosine diphosphate (ADP) and P_i . The terminal reaction of the electron transfer chain is the reduction of O_2 to H_2O by complex IV (cytochrome *c* oxidase, COX). This standard electron transfer pathway is frequently termed the 'cytochrome' (cyt) pathway because of the cytochromes present in complexes III and IV and cytochrome *c*. Specific inhibitors of each of the mitochondrial complexes are known. For example, rotenone inhibits complex I, antimycin A inhibits complex III, and cyanide inhibits COX.

In addition to the standard cyt pathway, mitochondria in plants contain a variety of additional electron transport proteins that impart flexibility to the system because they can operate without being coupled to ATP synthesis (Figure 1). On the electron input side of the ubiquinone pool are two sets of reduced nicotinamide adenine dinucleotide (phosphate)

(NAD(P)H) dehydrogenases. One pair of dehydrogenases is found on the outside of the inner mitochondrial membrane, allowing mitochondria to oxidize cytoplasmic NADH or NADPH, respectively. A second set exists on the inner surface of the inner membrane and oxidizes matrix derived NAD(P)H. These dehydrogenases do not contribute to proton gradient formation during electron transfer and therefore are able to operate unconstrained by it, unlike proton-pumping components such as complex I whose activity will be impaired as the size of the proton gradient increases. These alternative dehydrogenases are also insensitive to complex I inhibitors such as rotenone. Electrons derived from the oxidation of NAD(P)H by these dehydrogenases are transferred to the common pool of ubiquinone that also serves as the electron sink for complexes I and II (Figure 1).

Cyanide-Resistant Respiration

Plant mitochondria (and those of many other organisms cf. below) also have an additional electron transport branch downstream of the ubiquinone pool due to the presence of a second terminal oxidase in addition to COX. This alternative oxidase (AOX) is a ubiquinol:oxygen oxidoreductase that accepts electrons from reduced ubiquinol and subsequently reduces O_2 to H_2O (Figure 1). AOX activity is resistant to cyanide and other traditional inhibitors of COX but is specifically inhibited by salicylhydroxamic acid (SHAM) and *n*-alkyl gallate, neither of which affects the cyt pathway. An important feature distinguishes the two oxidase pathways. Electron flux downstream of the ubiquinone pool through the cyt pathway drives proton gradient formation at complexes III and IV, but no proton gradient formation occurs during operation of the AOX pathway. Thus, AOX would appear to be a wasteful enzyme from an energetic standpoint, eliminating a primary

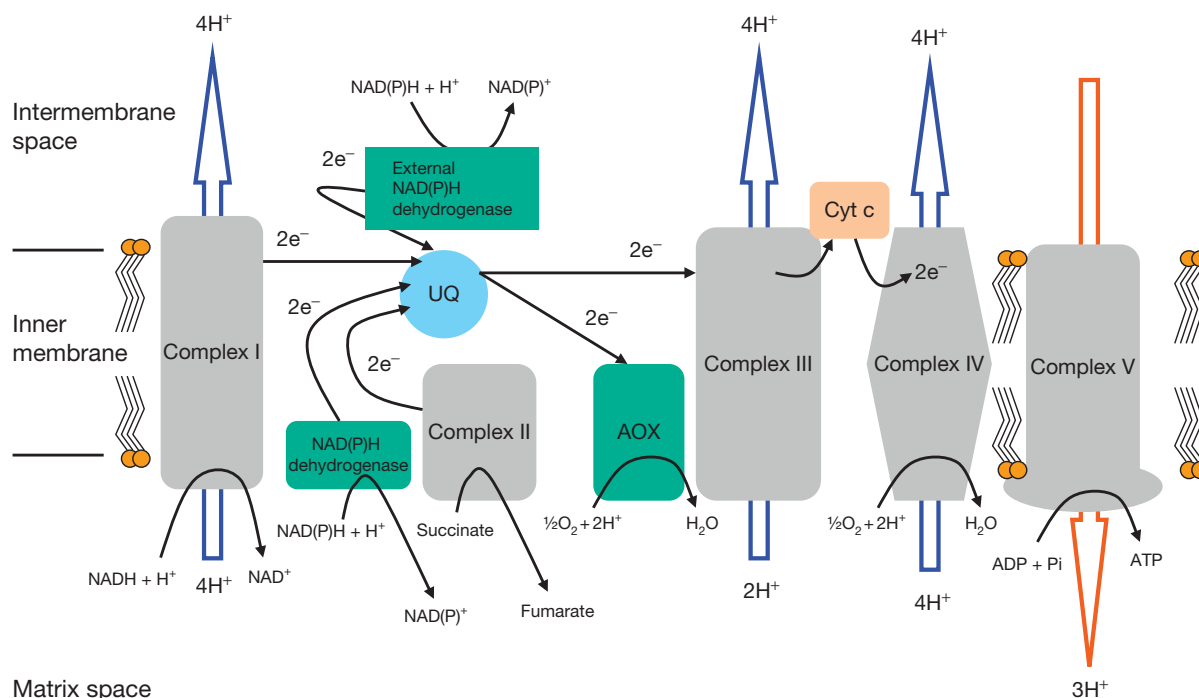


Figure 1 Schematic representation of the plant mitochondrial electron transfer chain.

mitochondrial function during aerobic respiration, coupling the oxidation of the reductants generated during the operation of glycolysis and the TCA cycle to the synthesis of ATP.

Historically, AOX has been associated with all higher plants, most algae, many fungi, and some protozoa. It was also long felt to be the case that no prokaryotes or metazoans (animals) contained this oxidase. More recently, the advent of large-scale genome sequencing has led to the discovery of the AOX protein in an ever-expanding number of organisms, including many proteobacteria and at least eight phyla of metazoans. At present, arthropods and vertebrates appear not to have AOX, and it also has yet to be found in Archaeobacteria. Among protozoa, AOX is noteworthy for its important role in the bloodstream form of the organism responsible for African sleeping sickness, *Trypanosoma brucei*.

Contemporary understanding of AOX can be traced to the development of an antibody against the plant AOX in 1989. The availability of this antibody led to the first isolation of a plant AOX gene and now hundreds of AOX sequences across a wide range of organisms (cf. above) are known. In eukaryotes, AOX is nuclear-encoded, and in plants, fungi, and protozoa, it often occurs as a small gene family (e.g., five genes in *Arabidopsis* and three in soybean), which produces proteins from 32 to 35 kDa (~285 amino acids). More recently, characterization of the gene associated with a mutation leading to a defect in chloroplast biogenesis, IMMUTANS, identified a chloroplast protein having sequence similarity to AOX. IMMUTANS also functions as a quinol oxidase, leading to its characterization as a plastid terminal oxidase (PTOX). It has been shown to participate in a desaturation step during carotenoid biosynthesis in chloroplasts. As prokaryotic genome sequencing has progressed, AOX homologs of unknown function have now been identified in many α -proteobacteria and PTOX homologs have

been reported in several cyanobacteria, respectively. Phylogenetic analyses of the bacterial and eukaryotic sequences suggest that AOX and PTOX arose from a common prokaryotic ancestor, but the two activities diverged prior to the endosymbiotic events that gave rise to mitochondria and chloroplasts.

AOX Structure

AOX is an integral inner membrane protein, but early attempts to purify and characterize it were hampered by the instability of the protein following its solubilization from the membrane and the lack of any apparent useful spectroscopic features to follow, including no visible (>350 nm) absorption spectrum or electron paramagnetic resonance (EPR) spectral features in either its fully oxidized or reduced states. The observation that iron was required for induction of AOX activity in the fungus *Hansenula anomala* originally led to the suggestion that the active site contained iron. AOX sequence analyses later revealed the presence of conserved iron-binding motifs found in members of the family of diiron carboxylate proteins that include the R2 subunit of ribonucleotide reductase and the hydroxylase subunit of methane monooxygenase. This led to the initial development of a model of the AOX active site that contained a four-helical bundle forming a hydroxo-bridged diiron binding site. The suggestion of a hydroxo-bridged iron center accounted for the absence of any visible absorbance, analogous to the diiron center in methane monooxygenase. Subsequently, an EPR signal associated with a mixed valence Fe (II)/Fe(III) hydroxo-bridged binuclear iron center was reported for an *Arabidopsis thaliana* AOX expressed in *Escherichia coli*, providing direct spectral evidence for the diiron carboxylate nature of the AOX and, by analogy, the PTOX active site.

Over time, the AOX structural model has been refined. As currently envisioned (**Figure 2(a)**), the C-terminal two-thirds of the AOX protein comprise a four-helical bundle that makes up the diiron binding scaffold. Two short hydrophobic helical regions anchor the protein in the inner leaflet of the bilayer, with the bulk of the protein protruding into the mitochondrial matrix. The C-terminal region is relatively conserved among all AOXs, showing 85–90% sequence similarity among plants and over 55% similarity between plants and fungi.

While the C-terminal two-thirds of the AOX protein is responsible for forming the catalytic diiron site, the structure of the N-terminal third of the protein is unknown. When plant and fungal amino acid sequences in this region are compared, there is considerable sequence conservation within each organismal type but marked sequence variability between them. Within the N-terminal third of the plant AOX, there is a highly conserved block of about 38 amino acids that includes a

regulatory cysteine (cf. below). Fungal sequences initially align with plant sequences right after this block. From that point on, the plant and fungal sequences are highly similar with two exceptions: (1) an insertion of 20–25 amino acids unique to fungi just before the first of the four iron-binding helical regions and (2) an extension of the fungal C-terminus, 20–50 amino acids beyond that found in plants. AOX sequences in protozoa are more similar to the fungal AOX than they are to that of plants.

In plants, chemical cross-linking studies have established that AOX exists in the membrane as a dimer consisting of two AOX monomers (**Figure 2(b)**). However, the fundamental unit of catalytic activity based on radiation-inactivation analysis is the monomer. Consistent with the latter observation, AOXs dimers are not universal; cross-linking studies with AOX in fungi and protozoa have established that they both exist in the membrane as monomers.

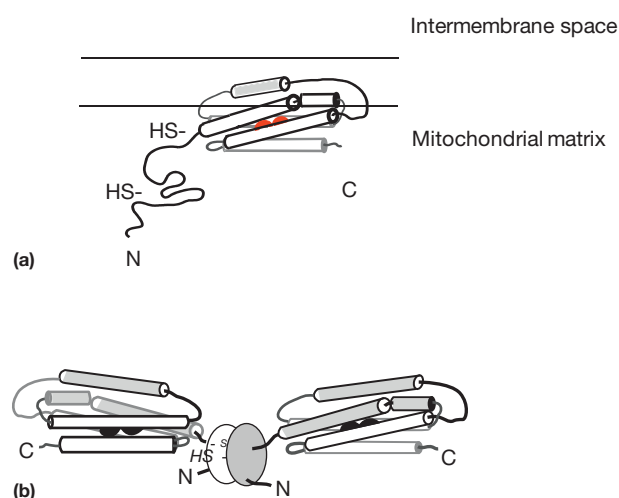


Figure 2 Diagrammatic representation of the structure of AOX. (a) AOX monomer, (b) dimeric AOX structure found in plants. The two red spheres in (a) represent the diiron center (shown in black in (b)). The inner mitochondrial membrane is delimited by the set of parallel horizontal lines.

Biochemical Regulation of AOX Activity

While the N-terminal third of the AOX protein is less well characterized structurally than the remainder of the protein, this region has been found to confer regulatory features on the plant AOX. The plant AOX homodimer can exist in two forms, an oxidized state where the two monomeric subunits are covalently linked by an intermolecular disulfide bond and a reduced state where the AOX remains dimeric but the disulfide is reduced to its component sulfhydryls (**Figure 3**). In the oxidized state, AOX is inactive. When the disulfide is reduced, AOX remains inactive but now has the ability to become activated in the presence of an α -keto acid, such as pyruvate (**Figure 3**). In the absence of α -keto acids, AOX has essentially no activity. AOX is activated by α -keto acids through reaction with a cysteine residue to form a thiohemiacetal. Site-directed mutagenesis of AOX has established that a single cysteine residue (termed CysI, the more N-terminal of two highly conserved cysteines located in the N-terminal third of the protein (**Figure 2(a)**)) is responsible for both the inactivating disulfide bond and interaction with α -keto acids to form the thiohemiacetal that activates AOX (**Figure 3**). CysI appears in the middle

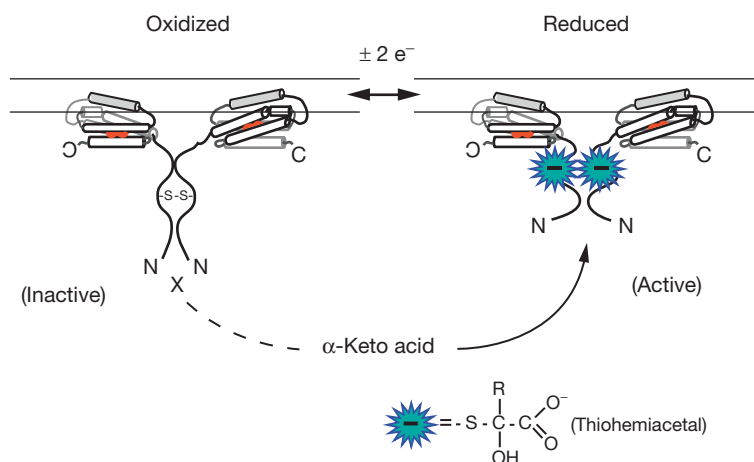


Figure 3 Biochemical regulation of plant AOX by the combined effects of reduction/oxidation of the redox-active sulfhydryl/disulfide bond and the reaction with α -keto acids to form an activating thiohemiacetal.

of the conserved block of 38 amino acids found in plant AOXs. Given the close proximity of the two regulatory CysIs on adjacent AOX monomers, it appears that AOX activation following α -keto acid binding results from a conformational change brought about by charge repulsion between the resulting two proximal negative charges (Figure 3).

The regulatory features outlined above provide a facile mechanism for activating AOX when electron flow through the standard cyt pathway is restricted. Under such conditions, electron transfer would slow down leading to a buildup of reductant in the mitochondrial matrix. The more reduced environment in the matrix would bring about the reduction of the AOX regulatory disulfide, possibly via reduced thioredoxin, which is found in the plant mitochondrial matrix. As electron transfer becomes limiting, TCA cycle activity would become restricted causing pyruvate concentrations to increase, activating the AOX through a feed-forward regulatory system.

Biochemical regulation of AOX activity in fungi and protozoa is not as well characterized as in plants, but purine nucleotides, particularly guanosine monophosphate (GMP), adenosine monophosphate (AMP), and ADP, are known to markedly stimulate AOX activity in fungi and protozoa. No stimulation of AOX activity by α -keto acids has been seen in fungi.

Physiological Role(s) of AOX

The fact that the alternative pathway wastes much of the free energy released during aerobic respiration must be related to its metabolic role. AOX has been found in all higher plants but there is only one instance where its role is clearly established, thermogenesis during flowering in certain members of the plant family Araceae. Just prior to pollen release, high levels of mitochondria found in club-like structures on the floral inflorescence undergo very high rates of respiration, heating up the inflorescence and volatilizing aromatic compounds that serve to attract pollinating insects. Mitochondria isolated from this thermogenic tissue have extremely high levels of AOX protein, allowing the high rates of respiration without having to synthesize and use comparable amounts of ATP.

Most plant tissues do not respire at rates sufficient to support even modest thermogenesis, so that they cannot represent its primary function in the vast majority of plant tissues. Because AOX is not involved in proton gradient generation, its activity is independent of the size of the gradient, unlike the cyt pathway, whose activity will be impaired as the proton gradient increases. Therefore, the AOX is in a position to oxidize mitochondrial and cytoplasmic reductant in excess of that needed to maintain ATP synthesis under normal metabolic circumstances, or in response to abnormal conditions when either ATP synthesis or the cyt pathway is impaired. Recognition of this fact originally served as the basis for the suggestion that the alternative pathway only operated when the activity of the cyt pathway was saturated, the so-called 'energy overflow' hypothesis. Subsequent studies have established that the alternative pathway can compete with the cyt pathway for electrons

from the ubiquinone pool when the cyt pathway is not saturated. This eliminates the overflow hypothesis in its purest form, but not from the standpoint of the alternative pathway helping to regulate mitochondrial electron flow and, as such, maintain mitochondrial redox homeostasis.

One clear outcome of the ability of the alternative pathway to use excess mitochondrial reductant, resulting from either increased input into the ubiquinone pool via the respiratory dehydrogenases or decreased output from the ubiquinone pool via the cyt pathway, is amelioration of the formation of reactive oxygen species (ROS; superoxide, hydrogen peroxide, and hydroxyl radical) that results from over-reduction of the ubiquinone pool. Operation of the alternative pathway has been shown to decrease both the reduction state of the ubiquinone pool in roots and the formation of ROS in plant culture cells. The ability of AOX to decrease mitochondrial ROS formation is the basis of a hypothesis, suggesting that AOX plays a role in plant responses to environmental stresses. In several circumstances where increased mitochondrial ROS levels are predicted, including chilling stress, post-hypoxia-induced reperfusion injury, and reduced Pi availability, AOX gene expression and protein levels are observed to increase markedly.

Similarly, enhanced AOX gene expression and protein formation have been found to accompany oxidative stresses in fungi, including treatment with hydrogen peroxide. A survey of the appearance of the alternative pathway among yeasts found that it was only present in so-called 'Crabtree negative' yeast, which lack the presence of alcoholic fermentation as an option to the cyt pathway consistent with the concept of the alternative pathway serving as a mechanism for eliminating excess reducing equivalents in these organisms.

See also: **Bioenergetics:** ATP in Plant Mitochondria: Substrates, Inhibitors, and Uncouplers; Structure of P-Type Adenosine Triphosphatases; V-ATPases; **Protein/Enzyme Structure Function and Degradation:** AAA-ATPases.

Further Reading

- Albury MS, Elliott C, and Moore AL (2009) Towards a structural elucidation of the alternative oxidase in plants. *Plant Physiology* 137: 316–327.
- Berthold DA and Stenmark P (2003) Membrane-bound diiron carboxylate proteins. *Annual Review of Plant Biology* 54: 497–517.
- Finnegan PM, Soole KL, and Umbach AL (2004) Alternative mitochondrial electron transport proteins in higher plants. In: Day DA, Millar AH, and Whelan J (eds.) *Advances in Photosynthesis and Respiration*, pp. 163–230. Dordrecht: Kluwer.
- McDonald AE (2008) Alternative oxidase: An inter-kingdom perspective on the function and regulation of this broadly distributed 'cyanide-resistant' terminal oxidase. *Functional Plant Biology* 35: 535–552.
- Moller IM (2001) Plant mitochondria and oxidative stress: Electron transport, NADPH turnover and metabolism of reactive oxygen species. *Annual Review of Plant Physiology and Plant Molecular Biology* 52: 561–591.
- Moore AL and Siedow JN (1991) The regulation and nature of the cyanide-resistant alternative oxidase of plant mitochondria. *Biochimica et Biophysica Acta* 1059: 121–140.
- Peltier G and Cournac L (2002) Chlororespiration. *Annual Review of Plant Biology* 53: 523–550.
- Siedow JN and Umbach AL (2000) The mitochondrial cyanide-resistant oxidase: Structural conservation amid regulatory diversity. *Biochimica et Biophysica Acta* 1459: 423–439.

Autophagy in Fungi and Mammals

J Geng, D J Klionsky, and J-L Guan, University of Michigan, Ann Arbor, MI, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Daniel J. Klionsky and Ju Guan, volume 1, pp. 138–143, © 2004, Elsevier Inc.

Glossary

Autophagy All processes in which intracellular material is degraded within the lysosome/vacuole followed by recycling of the macromolecular constituents.

Cytoplasm-to-vacuole targeting A biosynthetic pathway that transports resident hydrolases to the vacuole through a specific autophagy-like process.

Lysosome/vacuole A degradative organelle that compartmentalizes a range of hydrolytic enzymes and maintains a reduced pH.

Macroautophagy An autophagic process involving the formation of a double- or multiple-membrane cytosolic vesicle of non-lysosomal/vacuolar origin.

Microautophagy An autophagic process involving direct uptake of cytosol, inclusions, and organelles at the lysosome/vacuole by protrusion, invagination, or septation of the sequestering organelle membrane.

Mitophagy A selective type of autophagy involving the sequestration and degradation of mitochondria.

Pexophagy A selective type of autophagy involving the sequestration and degradation of peroxisomes.

Autophagy in Mammals

Macroautophagy in Mammals

The phenomenon of macroautophagy has been observed in virtually all eukaryotic cell types. Mammalian hepatic tissues and cell cultures have been extensively studied, where macroautophagy accounts for the majority of macromolecular recycling once it is induced by hormonal or other signals. During starvation-induced macroautophagy, the content of the sequestering vesicles, the autophagosomes, is indistinguishable from the surrounding cytoplasm, and often includes recognizable structures, such as mitochondria and ribosomes, suggesting that the sequestration is nonselective. There also exists evidence for selective sequestration of certain structures, such as peroxisomes and mitochondria, particularly when these organelles are specifically induced to proliferate before the onset of autophagy. Based on morphological and immunocytochemical studies, macroautophagy proceeds through distinct steps of autophagosome formation and maturation (Figure 1).

Formation of autophagosomes

The autophagosomes are distinct from most other intracellular vesicles in that double or multiple delimiting membranes are employed. In addition, formation of these vesicles topologically converts the sequestered cytoplasm to a luminal/extracellular space. Although the membrane source of the sequestering vesicle has been extensively investigated, the origin is still not known, and, in fact, it may be derived from several organelles. The autophagosomal membranes are extremely poor in proteins, indicating that they do not have protein profiles typical of most endomembranes, which also makes it difficult to determine their origin; however, various studies implicate the endoplasmic reticulum, the Golgi complex, and mitochondria as potentially providing lipids. The initial sequestering membrane is termed the 'phagophore', and this expands into an autophagosome by membrane addition.

Maturation of the phagophore

Maturation of phagophores into autophagosomes is thought to occur in a stepwise manner. That is, the autophagosome likely does not form in a single step from a pre-existing membrane, such as a cisterna of the endoplasmic reticulum (ER). Rather, additional phospholipids may be added by vesicular fusion or by direct flow from a source membrane(s). Upon completion, which is marked by closure of the phagophore membrane, the cytosolic vesicle becomes a double-membrane autophagosome. The autophagosomes fuse with lysosomes to acquire degradative enzymes, and the resulting structure is termed an 'autolysosome'. As a result of fusion, the inner autophagosomal membrane is exposed to the lysosome lumen, where it is broken down to allow access of the sequestered cytoplasm to the lysosomal hydrolases. The cytoplasmic cargo is subsequently degraded and recycled for use by the cell; the breakdown products are released into the cytosol through lysosomal membrane permeases. A population of autophagosomes may fuse with endosomes to give rise to 'amphisomes', an intermediate compartment named for its dual role in both macroautophagy and endocytosis. The amphisome ultimately fuses with a lysosome to allow degradation of its cargo.

Regulation of macroautophagy in mammals

The regulatory mechanism of macroautophagy is rather complex. Hormones and metabolites may activate or inhibit autophagy in different tissues. The intracellular signaling pathway has been studied mainly by using reagents that can alter the rates of autophagic protein degradation. It is very likely that the cellular responses to these reagents are pleiotropic. Various proteins have been identified that play a role in regulating macroautophagy, including a phosphatidylinositol 3-kinase (PtdIns3K), heterotrimeric G proteins, protein kinases, and phosphatases. Inhibition of membrane fusion and cytoskeletal functions affect the final stages of cargo delivery. Table 1 lists some of the factors and chemicals that affect macroautophagy.

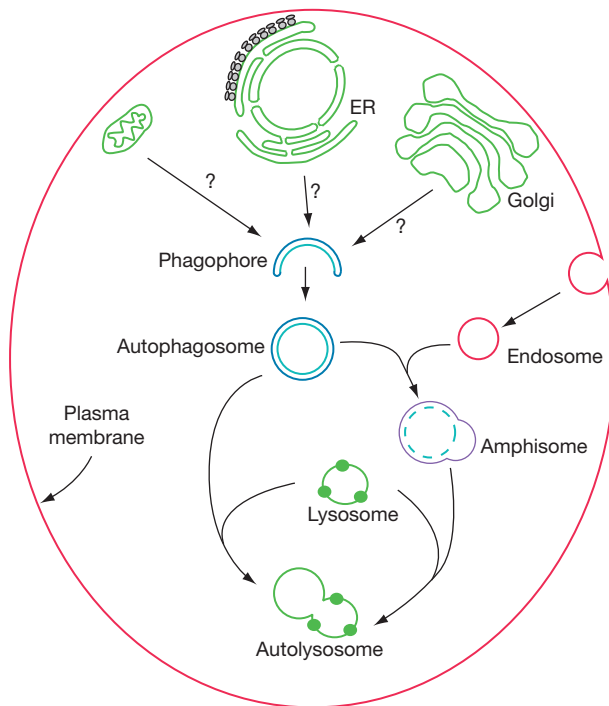


Figure 1 Macroautophagy in mammals. The initial sequestering membrane is termed the 'phagophore'. Upon closure, the resulting double-membrane cytosolic vesicle is called an 'autophagosome'. The origin of the phagophore or autophagosomal membrane is not known, but may include the endoplasmic reticulum, the Golgi complex, and/or the mitochondria. The autophagosome may fuse with an endosome to form an amphisome. The autophagosome or amphisome ultimately fuses with a lysosome, allowing breakdown of the inner vesicle (in the case of the autophagosome), followed by degradation and recycling of the vesicle contents.

Table 1 Inhibitors of macroautophagy in mammals

Compounds	Effects on autophagy	Mechanism/targets
<i>Physiological effectors</i>		
Amino acids	Decrease	Charged tRNA and surface receptors
Anabolic hormones (insulin, IGF, and EGF)	Decrease	Signaling
Catabolic factors/hormones (cyclic AMP and glucocorticoids)	Increase	Signaling
<i>Pharmacological agents</i>		
Bafilomycin A ₁	Decrease	Vacuolar-ATPase
Chloroquine	Decrease	Lysosomal pH
Cycloheximide	Decrease	Ribosome
Cytochalasin B	Decrease	Microfilaments
Colchicine and vinblastine	Decrease	Microtubules
Lithium	Increase	Inositol monophosphatase
Okadaic acid	Decrease	Serine/threonine protein phosphatases
Rapamycin	Increase	TOR complex
Wortmannin, LY294002, 3-methyladenine	Decrease	Phosphatidylinositol 3-kinase class III

Physiological functions of macroautophagy in mammals

One of the primary functions of macroautophagy is the supply of macromolecules and energy as an adaptive mechanism under starvation conditions in mammals and other organisms. In response to different forms of metabolic stress, such as nutrient deprivation, growth-factor depletion, or hypoxia, activation of autophagy allows degradation of bulk cellular contents to generate free amino acids and other essential blocks of cellular metabolism to be recycled and used for maintaining survival of the organisms. Although occurring at a relatively low level under normal conditions, basal (or constitutive) autophagy is also recognized to play an important role in cellular homeostasis by functioning in the removal of surplus or damaged organelles and in the degradation of misfolded protein aggregates that are too large for the ubiquitin-proteasome system.

Recent studies using gene knockouts in the mouse system provide strong evidence that macroautophagy plays an essential role in the survival of mammals through its functions during both starvation and normal conditions. Mice with knockouts of either *Atg5* or *Atg7*, both encoding essential components of mammalian macroautophagy, die shortly after birth, probably due to their inability to activate macroautophagy in response to the starvation period in neonates. Deletion of FIP200, a component of the ULKs-ATG13-FIP200 complex that is important in the formation of phagophores, leads to embryonic lethality caused by developmental defects in the heart and liver, where autophagy may play a more crucial role due to the high metabolic rate in these organs in late embryogenesis. Tissue-specific knockout of different macroautophagy genes in hepatocytes, neurons, or cardiomyocytes results in the abnormal accumulation of ubiquitinated protein aggregates in inclusion bodies associated with cellular degeneration, suggesting a role for basal macroautophagy in protein quality control in cellular homeostasis and survival.

Microautophagy in Mammals

Microautophagy refers to the direct import of cytoplasmic materials by lysosomal membrane protrusions or invaginations. This process is less well characterized in terms of its mechanisms and its contribution to overall protein degradation. In mammals, lysosomes with appendages or intralysosomal vesicles have been observed *in vivo*. Isolated lysosomes show the ability to uptake and degrade proteins or incorporate electron-dense markers into intralysosomal vesicles. Microautophagy is also induced by environmental cues, such as nutrient deprivation.

Autophagy in Fungi

Macroautophagy in Yeast

Budding yeast as a model for macroautophagy

Ever since the early 1990s, autophagy has been studied using the budding yeast *Saccharomyces cerevisiae* as a model system. Nitrogen or carbon starvation induces the formation of cytoplasmic autophagosomes that have the characteristic double-delimiting membranes. Similar to the findings in mammalian cells, yeast autophagosomal membranes are also protein poor, as revealed

by freeze-fracture electron-microscopy studies. Once formed in the cytoplasm, the outer autophagosomal membrane will fuse with the vacuole (equivalent to the mammalian lysosome) membrane, leaving the inner vesicle, called the 'autophagic body', in the vacuolar lumen. The autophagic bodies are subsequently degraded by vacuolar hydrolases and the degradation products will be released to the cytosol for reuse (Figure 2).

The cytoplasm-to-vacuole targeting pathway and its overlap with macroautophagy

The cytoplasm-to-vacuole targeting (Cvt) pathway was discovered as the biosynthetic targeting pathway of a vacuolar resident enzyme, aminopeptidase I (Ape1). Precursor Ape1 (prApe1) is synthesized in the cytosol. It rapidly assembles into a dodecamer and remains in an oligomerized form during its course of transport into the vacuole (Figure 2). Cleavage of the N-terminal propeptide of prApe1 by vacuolar proteases gives rise to the mature, active form of the protein (mApe1).

Several screens were carried out to identify mutants (*atg*) defective in macroautophagy, the Cvt pathway, and the selective degradation of organelles. Genetic analyses revealed that these mutants overlap. In the case of the Cvt pathway, for example, morphological and biochemical studies confirm that prApe1 is transported to the vacuole using a vesicular mode similar to that of macroautophagy. Precursor Ape1 exists in the cytoplasm as a large structure that can be detected by

electron microscopy. This structure, which consists of multiple prApe1 dodecamers bound to receptor proteins, is called the 'Cvt complex'. Under vegetative growth conditions, the Cvt complex is specifically enwrapped by double-membrane Cvt vesicles. When autophagy is induced, the Cvt complex becomes a preferential cargo of autophagosomes due to the function of the prApe1 receptor. Thus, the Cvt pathway can be regarded as a specific form of autophagy. Another hydrolase, α -mannosidase, is also targeted to the vacuole via the Cvt and autophagy pathways. The parallel studies of the Cvt pathway complement the analysis of autophagy by providing a specific cargo and the opportunity to study selective uptake mechanisms of autophagy.

Molecular mechanism of the Cvt and autophagy pathways in yeast

The molecular machinery of autophagy and the Cvt pathway was identified by cloning the genes that complement the mutants defective in these pathways. Additional components were found by other approaches, including two-hybrid studies. A unified nomenclature was adopted, which uses *ATG* (autophagy-related) to name all the identified genes. Most of the *Atg* proteins are required for both autophagy and the Cvt pathway. Table 2 lists the *Atg* proteins and their proposed roles in these pathways. They can be divided into at least three basic subcategories: (1) *Atg8* and *Atg12* conjugation systems; (2) *Atg9* and its cycling

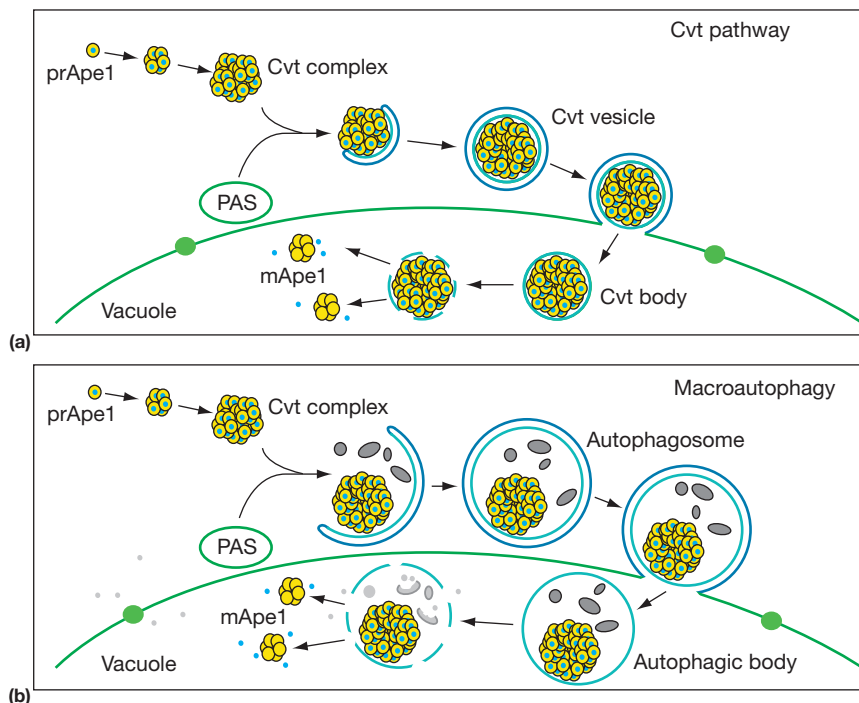


Figure 2 Cvt pathway and macroautophagy in yeast. (a) The Cvt pathway. Precursor Ape1 (prApe1) forms a dodecamer that assembles into a higher order Cvt complex. Recruitment of the Cvt complex to the PAS leads to the formation of a Cvt vesicle. The completed Cvt vesicle fuses with the vacuolar membrane and releases the inner vesicle (Cvt body) into the vacuole lumen. Breakdown of the Cvt body allows maturation of prApe1 (to mApe1) by removal of a propeptide. (b) Macroautophagy. Under starvation conditions, the PAS is activated by specific signals, resulting in the formation of autophagosomes. The Cvt complex is enclosed along with bulk cytoplasm inside autophagosomes. The inclusion of the Cvt complex is a specific process resulting from the action of receptor and adaptor proteins. Fusion between autophagosomes and the vacuole membrane and the subsequent breakdown of the inner membrane (autophagic body) release the cytoplasmic contents into the vacuole lumen for degradation and recycling or, in the case of the Cvt complex, maturation of precursor aminopeptidase I.

Table 2 Proteins involved in macroautophagy and the Cvt pathway

Protein	Description of function	Cvt ^a	Atg
Atg1	A serine/threonine kinase involved in induction of autophagy; required for Atg9 retrieval from the PAS to peripheral sites; a component of the Atg1-Atg13-Atg17-Atg29-Atg31 complex	–	–
Atg2	Required for Atg9 retrieval from the PAS to peripheral sites	–	–
Atg3	An E2-like enzyme involved in Atg8–PE conjugation	–	–
Atg4	A cysteine protease that cleaves the C-terminal arginine of Atg8 prior to its lipidation and mediates the deconjugation of Atg8–PE	–	–
Atg5	A component of the Atg12–Atg5-Atg16 complex that has an E3-like function in Atg8 lipidation	–	–
Atg6	A component of two PtdIns3K complexes involved in autophagy (complex I) and the Vps pathway (complex II)	–	–
Atg7	An E1-like enzyme involved in the formation of Atg8–PE and Atg12–Atg5 conjugates	–	–
Atg8	A ubiquitin-like protein conjugated to PE; involved in phagophore expansion and usually used as an autophagosome marker	–	–
Atg9	An integral membrane protein that cycles between the PAS and multiple peripheral sites	–	–
Atg10	An E2-like enzyme involved in the formation of Atg12–Atg5 conjugates	–	–
Atg11	A protein involved in cargo recognition and PAS assembly during specific autophagy; mediates Atg9 transport to the PAS during specific autophagy	–	+
Atg12	A ubiquitin-like protein covalently conjugated to Atg5	–	–
Atg13	A component of the Atg1-Atg13-Atg17-Atg29-Atg31 complex; dephosphorylated during starvation conditions	–	–
Atg14	A component of PtdIns3K complex I involved in autophagy	–	–
Atg15	A putative vacuolar lipase required for the lysis of autophagic and Cvt bodies	–	–
Atg16	A component of the Atg12–Atg5-Atg16 complex; determines the localization of the complex and the site of Atg8 lipidation	–	–
Atg17	A scaffold protein essential for PAS assembly during starvation; a component of the Atg1-Atg13-Atg17-Atg29-Atg31 complex that regulates Atg1 kinase activity	+	–
Atg18	A PtdIns(3)phosphate binding protein essential for vesicle formation in autophagy and the Cvt pathway; required for Atg9 retrieval from the PAS to peripheral sites	–	–
Atg19	A cargo receptor in the Cvt pathway; directs the cargo complex to the PAS by interaction with Atg11	–	+
Atg20	A protein with a PX domain that binds to PtdIns(3)phosphate	–	+
Atg21	A PtdIns(3)phosphate binding protein required for the Cvt pathway	–	Partial
Atg22	An amino acid permease on the vacuolar membrane that releases amino acids generated by vacuolar degradation back to the cytosol	+	+
Atg23	A protein partially colocalized with Atg9 and essential for its transport to the PAS	–	Partial
Atg24	A protein with a PX domain that binds to PtdIns(3)phosphate	–	+
Atg25	Not identified in <i>Sacharomyces cerevisiae</i> ; required for pexophagy in <i>Hansenula polymorpha</i>	N.A.	N.A.
Atg26	A protein with a GRAM domain required for pexophagy in <i>Pichia pastoris</i> , but not in <i>S. cerevisiae</i>	+	+
Atg27	An integral membrane protein partially colocalized with Atg9 and essential for its transport to the PAS	–	Partial
Atg28	A coiled-coil protein required for pexophagy in <i>P. pastoris</i> ; not identified in <i>S. cerevisiae</i>	N.A.	N.A.
Atg29	A component of the Atg1-Atg13-Atg17-Atg29-Atg31 complex	+	–
Atg30	Not identified in <i>S. cerevisiae</i> ; required for pexophagy in <i>P. pastoris</i>	N.A.	N.A.
Atg31	A component of the Atg1-Atg13-Atg17-Atg29-Atg31 complex	+	–
Atg32	An integral membrane protein localized on the mitochondria outer membrane; mediates cargo recognition in mitophagy by interaction with Atg11	+	+
Atg33	A protein required for mitophagy	+	+
Atg34	A cargo receptor for α -mannosidase during starvation	+	+
Atg35	A protein required for micropexophagy in <i>P. pastoris</i>	+	+

^aA '–' or a '+' indicates that the protein is required or not required for the Cvt or autophagy pathway, respectively (i.e., the pathway does not, or does function in the absence of this protein).

system; and (3) the PtdIns3K complex and a series of phosphatidylinositol-binding proteins. The molecular mechanism for the Cvt pathway and macroautophagy is briefly outlined below.

Phagophore assembly site and de novo vesicle formation

In vivo fluorescence microscopy studies have colocalized the autophagosomal membrane marker protein Atg8 with several other Atg proteins at a perivacuolar site. This site is physiologically functional and plays a pivotal role in autophagosome formation; therefore, it has been termed the phagophore assembly site (PAS). Localized on the PAS are two conjugates, Atg12 covalently linked to Atg5, and Atg8 covalently attached to phosphatidylethanolamine (Atg8–PE). These conjugates are both essential in the generation of autophagosomes and Cvt vesicles. Formation of the Atg12–Atg5 and Atg8–PE conjugates

involves ubiquitin-like cascades. Recruitment of Atg12–Atg5 and Atg8 to the PAS depends on the function of the transmembrane protein Atg9 and the autophagy-specific PtdIns3K complex (there are two PtdIns3K complexes in yeast), underlying the key role of specific lipids in autophagosome formation. Atg9 cycles between the PAS and multiple peripheral sites, which may represent the donor membranes source(s) that allow expansion of the phagophore.

Regulation of autophagy

The Cvt pathway operates under vegetative conditions, whereas autophagy is induced by starvation. Atg1 is localized at the PAS and is important in signaling Cvt vesicle or autophagosome formation, possibly via its differential association with other proteins such as Atg11 and Atg13. In particular, the

Atg1-Atg13 interaction is regulated by the target of rapamycin (TOR) kinase. Inhibition of TOR activity results in the induction of autophagy. TOR also regulates autophagy at the transcriptional level. Besides TOR, there are other signaling pathways involved in autophagy regulation such as the Ras/cyclic adenosine monophosphate (Ras/cAMP)-dependent protein kinase A (PKA) pathway.

Mechanism for cargo selection in the Cvt pathway and autophagy

The Cvt complex is selectively sequestered by Cvt vesicles or autophagosomes. Atg19, a structural component of the Cvt complex, acts as a receptor and mediates the recruitment of the oligomerized prApe1 cargo and its recognition by the PAS through successive protein-protein interactions with Atg11 and Atg8. This may provide a prototype for other selective autophagic targeting pathways.

Fusion and breakdown

The fusion between the Cvt vesicles or autophagosomes and the vacuole membrane is similar to vacuole homotypic fusion and requires docking/tethering factors in addition to SNARE and Rab proteins. Breakdown of the inner vesicles requires the action of specific components such as the putative Atg15 lipase in addition to other vacuolar hydrolases. Free amino acids generated by vacuolar protein degradation are recycled into the cytosol via vacuolar permeases such as Atg22.

Microautophagy and Selective Organelle Degradation in Yeasts

Nonselective microautophagy

The morphological features of microautophagy have been observed in yeasts. Microautophagic vacuole invagination was also reconstituted *in vitro*. Interestingly, the molecular components of macroautophagy, the Atg proteins, were implicated to function at least partly in microautophagy as well, although in the case of nonselective microautophagy this may be indirect; microautophagy may function in part to maintain vacuole-membrane homeostasis by removing membrane following fusion of vesicles, including those of macroautophagy, with the vacuole-limiting membrane.

Pexophagy: Selective autophagy of peroxisomes in methylotrophic yeasts

Selective degradation of peroxisomes by autophagic mechanisms has been best demonstrated in methylotrophic yeasts. Peroxisomes are necessary for utilizing methanol and are induced when methanol is the sole carbon source. Upon shift of culture medium to preferred carbon sources, superfluous peroxisomes are targeted to the vacuole via either a macro- or microautophagic mode, depending on nutrient conditions. The selective degradation of peroxisomes is termed 'pexophagy'. The proteins involved in selective macroautophagy of peroxisomes (macropexophagy), and even those required for selective microautophagy of peroxisomes (micropexophagy), are generally found to be homologous to the Atg proteins. This latter finding suggests that macroautophagy and selective microautophagy may be more closely related at the molecular level than they appear to be morphologically.

Mitophagy: Selective degradation of mitochondria via autophagy

Mitochondria are critical for supplying energy in eukaryotic cells, but they also generate reactive oxygen species during catabolism, which can potentially harm cells. Mitophagy plays an important role in the elimination of damaged or excess mitochondria, depending on the energy requirement. In *S. cerevisiae*, mitophagy can be induced by nitrogen starvation or treatment with the TOR inhibitor rapamycin after pre-culturing in nonfermentable media, which induce mitochondrial proliferation. The autophagy machinery is essential for mitophagy. In addition to the known ATG genes, genetic screening has identified additional genes required specifically for mitophagy (e.g., ATG32 and ATG33). Further analyses will be needed to characterize the specific roles of the corresponding gene products in mitophagy.

Autophagy, from Yeast to Mammals

Most of the Atg proteins are conserved in higher eukaryotes, suggesting that the mechanism for autophagy is similar in yeasts and mammals. Future studies with mammalian cells may further benefit from the molecular model provided by yeast systems. However, there are also some Atg proteins that are unique to higher eukaryotes, perhaps reflecting the increased roles for macroautophagy in these organisms. For example, abnormalities in autophagy are connected with various human pathophysiologicals, such as neurodegeneration and certain types of cancer. The importance of autophagy in developmental or pathological macromolecular turnover in higher eukaryotes promises to be an exciting area of future research due to its potential therapeutic implications.

See also: [Cell Architecture and Function: Vacuoles](#); [Protein/Enzyme Structure Function and Degradation: Phage Display for Protein Binding](#); [Protein Degradation](#).

Further Reading

- Cecconi F and Levine B (2008) The role of autophagy in mammalian development: Cell makeover rather than cell death. *Developmental Cell* 15: 344–357.
- Dunn WA Jr., Cregg JM, Kiel JAKW, et al. (2005) Pexophagy: The selective autophagy of peroxisomes. *Autophagy* 1: 75–83.
- Geng J and Klionsky DJ (2008) The Atg8 and Atg12 ubiquitin-like conjugation systems in macroautophagy. *EMBO Reports* 9: 859–864.
- He C and Klionsky DJ (2009) Regulation mechanisms and signaling pathways of autophagy. *Annual Review of Genetics* 43: 67–93.
- Kanki T and Klionsky DJ (2010) The molecular mechanism of mitochondria autophagy in yeast. *Molecular Microbiology* 75: 795–800.
- Kon M and Cuervo AM (2010) Chaperone-mediated autophagy in health and disease. *FEBS Letters* 584: 1399–1404.
- Levine B and Klionsky DJ (2004) Development by self-digestion: Molecular mechanisms and biological functions of autophagy. *Developmental Cell* 6: 463–477.
- Levine B and Kroemer G (2008) Autophagy in the pathogenesis of disease. *Cell* 132: 27–42.
- Lynch-Day MA and Klionsky DJ (2010) The Cvt pathway as a model for selective autophagy. *FEBS Letters* 584: 1359–1366.
- Meijer AJ and Codogno P (2006) Signalling and autophagy regulation in health, aging and disease. *Molecular Aspects of Medicine* 27: 411–425.

- Mizushima N, Levine B, Cuervo AM, and Klionsky DJ (2008) Autophagy fights disease through cellular self-digestion. *Nature* 451: 1069–1075.
- Münz C (2009) Enhancing immunity through autophagy. *Annual Review of Immunology* 27: 423–449.
- Reggiori F (2006) Membrane origin for autophagy. *Current Topics in Developmental Biology* 74: 1–30.
- Rubinshtein DC, DiFiglia M, Heintz N, et al. (2005) Autophagy and its possible roles in nervous system diseases, damage and repair. *Autophagy* 1: 11–22.
- Seglen PO, Berg TO, Blankson H, et al. (1996) Structural aspects of autophagy. *Advances in Experimental Medicine and Biology* 389: 103–111.
- Virgin HW and Levine B (2009) Autophagy genes in immunity. *Nature Immunology* 10: 461–470.
- Xie Z and Klionsky DJ (2007) Autophagosome formation: Core machinery and adaptations. *Nature Cell Biology* 9: 1102–1109.
- Yen W-L and Klionsky DJ (2008) How to live long and prosper: Autophagy, mitochondria, and aging. *Physiology* 23: 248–262.

B₁₂-Containing Enzymes

V Bandarian, University of Michigan, Ann Arbor, MI, USA; University of Arizona, Tucson, AZ, USA

R G Matthews, University of Michigan, Ann Arbor, MI, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 145–151, © 2004, Elsevier Inc.

Glossary

B₁₂ A complex organometallic cofactor, also called cobalamin, that is used by the enzymes that catalyze group transfer or radical-mediated rearrangement reactions.

Cofactor A compound that is noncovalently associated with an enzyme and is required for the catalytic activity.

Radical A species that contains an unpaired electron.

The Cobalamin Cofactor

The central feature of the cobalamin cofactor (**Figure 1(a)**) is a cobalt atom with octahedral geometry, which is centered in the corrin macrocycle by coordination to pyrrole nitrogens from the corrin in the four equatorial coordination positions of the metal ion. A nitrogen atom that is donated by a dimethylbenzimidazole moiety, which in turn is ligated to the corrin ring, coordinates the cobalt from the lower axial position. In some eubacteria and Archaea, the dimethylbenzimidazole substituent of the cobalamin is replaced by other compounds (e.g., *p*-cresol); such cofactors are referred to as corrinoids. The macrocycle is further adorned by methyl, acetamide, and propionamide side chains. The upper axial position of the cobalt ion is occupied by an alkyl substituent specific to the type of reaction catalyzed by the protein.

The upper face of the cobalamin is where the chemistry takes place in all cobalamin-dependent enzymes. The chemical versatility of cobalamin and corrinoids lies in the carbon–cobalt bond of the cofactor, which is susceptible to cleavage by heterolytic or homolytic pathways (see **Figure 1(b)**). The upper coordination position of the cobalamin, in the enzymes that catalyze methyl group transfer, is transiently occupied by a methyl group en route from a donor molecule to an acceptor molecule. In these enzymes, the cobalamin serves both as a nucleophile, accepting the methyl group from the donor, and as a leaving group, donating the methyl group to an acceptor molecule. By contrast, enzymes that catalyze radical-mediated transformations require that the cobalamin has 5'-deoxyadenosine coordinated to the cobalt on the upper face of the corrin. In these enzymes, the organometallic bond between the 5'-methylene of the deoxyadenosyl moiety and the cobalt is severed homolytically to generate a 5'-deoxyadenosyl radical, which initiates the radical cascades that lead to the formation of product.

The identity of the ligand that occupies the lower axial coordination position of protein-bound cobalamin may differ significantly from that in solution. In some proteins, in addition to the network of interactions with the side chains of the corrin, the imidazole side chain of a histidine residue substitutes for the dimethylbenzimidazole of the cofactor and donates a nitrogen ligand to the cobalt. This latter form of binding is referred to as a base-off binding mode. Most of the proteins that bind the cobalamin in the base-off mode also contain a DxHxxG sequence that contains the histidine residue that donates the lower axial imidazole ligand. In other proteins, binding of cobalamin to the protein backbone is accomplished by extensive interactions with the corrin and its side chains, and the dimethylbenzimidazole remains coordinated to the cobalt atom. This is referred to as base-on binding of the cofactor.

The cobalamin cofactor can exist in three oxidation states, and each form of the cobalamin exhibits distinct ultraviolet (UV)–visible spectra that allow one to follow the course of the catalytic cycle. For instance, the cobalt in alkyl cobalamins, such as adenosylcobalamin or methylcobalamin, is in the +3 oxidation state (formally, the alkyl ligand is considered to be the anion R[−]). Homolysis of the cobalamin to generate a deoxyadenosyl radical is accompanied by the formation of a reduced cobalamin, cob(II)alamin, in which the cobalt is in the +2 oxidation state, and reformation of the C–Co bond of the cofactor oxidizes the cobalt to the +3 state. Therefore, adenosylcobalamin-dependent proteins cycle between the +3 and +2 oxidation states during the course of the catalytic cycle. By contrast, upon nucleophilic displacement of methyl group from the cobalamin during group transfer chemistry, the cobalamin cycles between the +3 oxidation state in the methylcobalamin form and the +1 state upon loss of the alkyl group. Remethylation of the cobalamin reoxidizes the cobalamin to the +3 oxidation state.

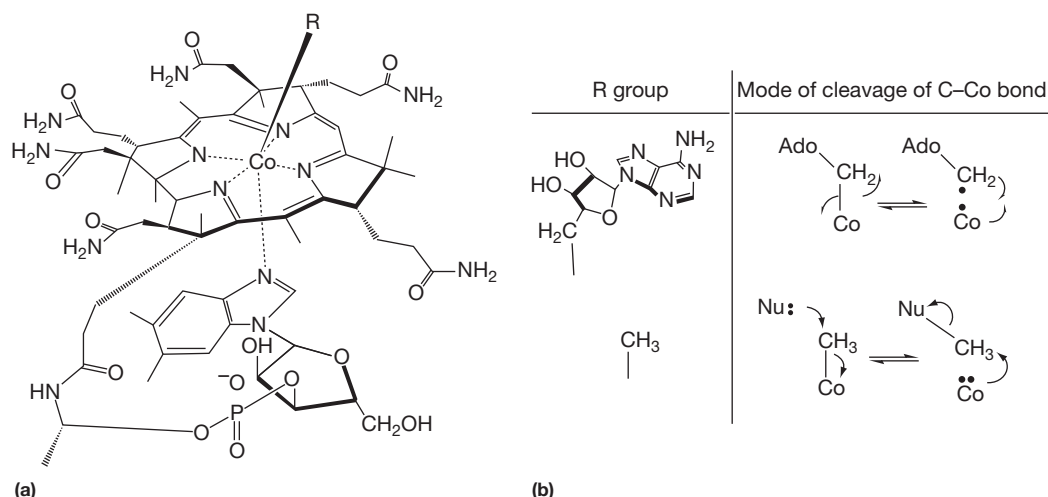


Figure 1 Structure and reactivity of alkylcobalamins: (a) the upper coordination position of cobalamin can be occupied by 5'-deoxyadenosine or a methyl group; (b) homolysis of the C–Co bond in adenosylcobalamin-dependent enzymes generates a 5'-deoxyadenosyl radical and cob(II)alamin. By comparison, the cofactor is transiently methylated in enzymes that catalyze group transfer reactions.

In this article, we summarize the reactions that are catalyzed by cobalamin-requiring enzymes, first highlighting the reactions in which cobalamin serves as a source of the highly reactive 5'-deoxyadenosyl radical and then discussing the reactions that involve methyl transfer chemistry.

Reactions Catalyzed by Cobalamin-Dependent Enzymes

Radical-Mediated Rearrangement Reactions

The enzymes that catalyze radical-mediated transformation exploit the inherent weakness of the carbon–cobalt bond of adenosylcobalamin (bond dissociation energy $\sim 30 \text{ kcal mol}^{-1}$) to form the highly reactive 5'-deoxyadenosyl radical. Binding of the cobalamin to these enzymes accelerates the rate of homolysis of the C–Co bond $\geq 10^{12}$ -fold. Homolysis of the carbon–cobalt bond is triggered by the presence of the substrate or by an allosteric effector. Presumably, the binding energy derived from substrate/effector–enzyme interactions is essential to elicit the conformational changes that are required to facilitate homolysis of the relatively weak organometallic bond. The 5'-deoxyadenosyl radical is a primary radical and readily abstracts a hydrogen atom from the substrate or from a residue on the protein. Therefore, a hallmark of catalysis by the enzymes in this group is formation of radical intermediates. **Figure 2** shows the general catalytic cycles for these enzymes.

Adenosylcobalamin-dependent enzymes can be divided into four groups based on the details of the catalytic transformations (**Table 1**). The reactions catalyzed by enzymes in each group are discussed next.

Migration and Elimination Reactions

The enzymes in this group participate in the fermentation of short-chain organic compounds such as ethanolamine, glycerol, 1,2-propanediol, and 1,2-ethanediol. The reactions catalyzed by these enzymes are formally the interchange of a

hydrogen atom on one carbon with a group X ($=\text{OH}$ or NH_3) at the adjacent position. The catalytic mechanism of these enzymes is quite similar to the mechanism of the carbon skeleton mutase enzymes discussed later. These two groups, in fact, differ only in that the mutases retain the rearranged substituent.

The catalytic mechanisms of ethanolamine ammonia-lyase and diol dehydratase have been studied extensively and a generalized mechanism for these enzymes is shown in **Figure 2(a)**. As with all cobalamin-dependent enzymes, the carbon–cobalt bond of the cofactor is homolyzed to generate 5'-deoxyadenosyl radical and cob(II)alamin in the first step of the catalytic reaction. The deoxyadenosyl radical abstracts a hydrogen atom from the substrate (S–H) to generate a substrate-like radical ($\text{S}\cdot$) $\sim 9\text{--}12 \text{ \AA}$ away from the cob(II)alamin. The rearrangement of $\text{S}\cdot$ to a product-like radical ($\text{P}\cdot$) followed by return of a hydrogen atom from the 5'-deoxyadenosine results in the formation of the enol form of the product (P–H). The deoxyadenosyl radical recombines with cob(II)alamin to regenerate adenosylcobalamin. The initial *gem*-diol or *gem*-amino alcohol that is formed in the course of these transformations subsequently eliminates water or ammonia, respectively, to generate the corresponding aldehydes.

Diol and glycerol dehydratases have the best-characterized structures in this group. Few sequence similarities exist between the diol and glycerol dehydratases and ethanolamine ammonia-lyase, despite the similarities in the reactions that they catalyze. Nevertheless, members of these groups of enzymes appear to retain coordination of the cobalamin to dimethylbenzimidazole.

Diol dehydratase and glycerol dehydratase are rapidly inactivated in the course of their catalytic cycle by side reactions that lead the formation of tightly bound inactive cofactor at the active site of these enzymes. The gene clusters that code for the structural genes for these proteins also contain open reading frames whose products have been shown to catalyze the adenosine triphosphate (ATP)- and Mg^{2+} -dependent exchange of inactive cofactor with cofactor from solution. The generality of the reactivation mechanism remains to be determined.

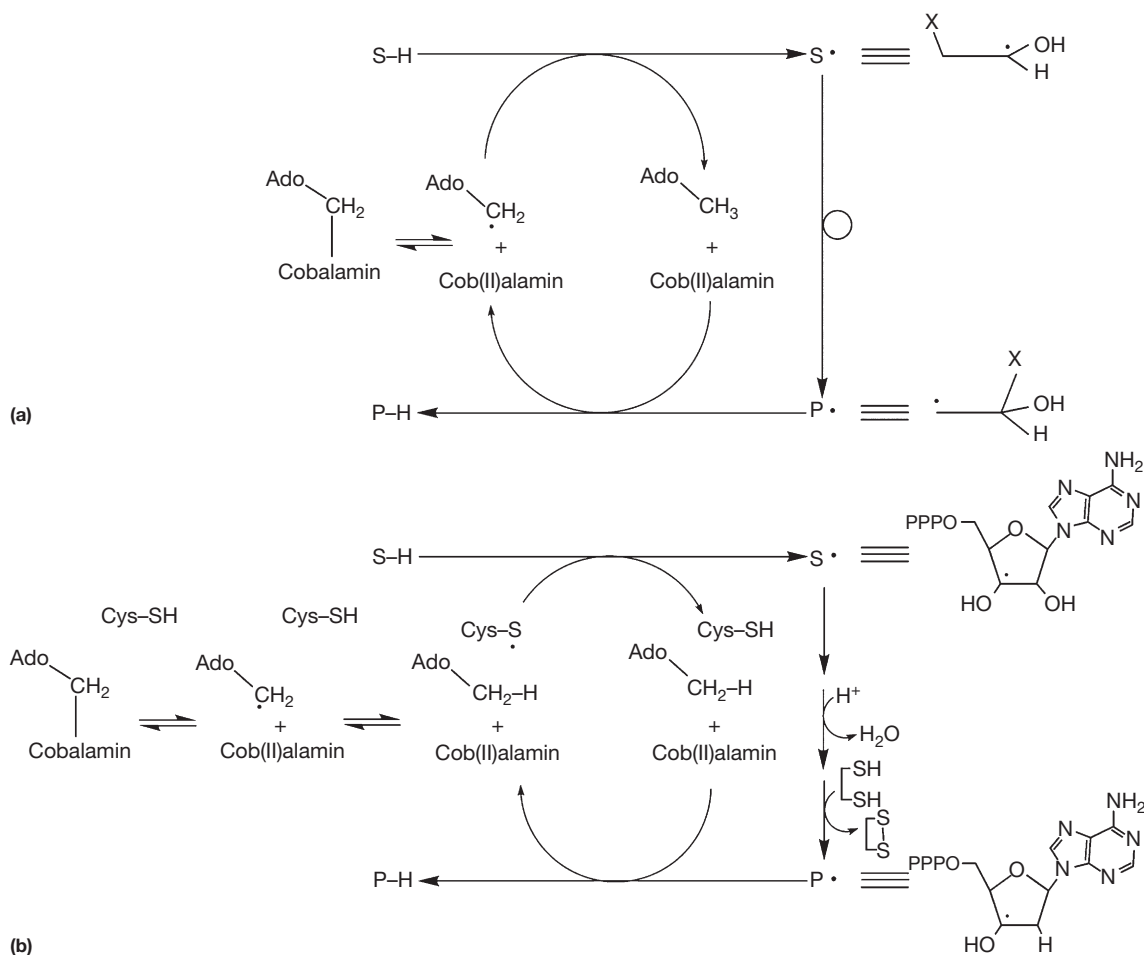


Figure 2 Generalized reaction cycles for enzymes that catalyze radical-mediated rearrangements. (a) The catalytic cycles for enzymes that catalyze rearrangements and eliminations, carbon skeleton rearrangements, and PLP-dependent aminomutase reactions. The structures on the right are examples of the intermediates that would be expected in the conversion of ethanolamine acetaldehyde and ammonia by ethanolamine ammonia-lyase. (b) The catalytic cycle of ribonucleotide triphosphate reductase differs from the cycle in (a) in that the 5'-deoxyadenosyl radical abstracts a hydrogen atom from an active site cysteine thiol to generate a thiyl radical that subsequently generates a substrate-based radical species. The structures on the right show some of the reaction intermediates that form in the course of conversion of ATP to dATP.

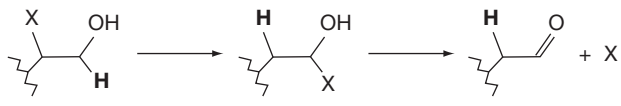
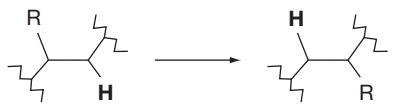
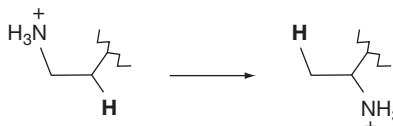
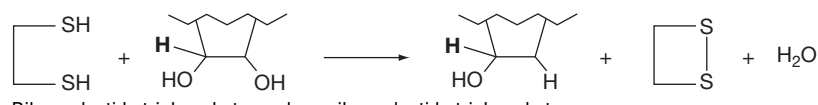
Carbon Skeleton Rearrangement Reactions

As with the enzymes that catalyze rearrangement and elimination, the enzymes in this group catalyze the interchange of a hydrogen atom on one carbon atom with an alkyl group on the adjacent carbon atom (see Table 1). The reactions catalyzed by these enzymes differ from the others in several respects. First, the migrating alkyl group is not eliminated from the product. Second, the mutase reactions are reversible. Two significant structural differences have also been noted between these enzymes. First, spectroscopic and structural studies have shown that cobalamin binds in the base-off conformation and that the histidine residue that donates the imidazole ligand is conserved in the members of the group. Second, the substrate-based radicals that are formed in the course of the reaction reside 6 Å away from the cob(II)alamin. However, despite these differences, the 5'-deoxyadenosyl radical is involved in generating a substrate-based radical and the catalytic cycle of these enzymes resembles that of the enzymes that catalyze rearrangement and elimination reactions (see Figure 2(a)).

Pyridoxal 5'-Phosphate-Dependent Aminomutase Reactions

Pyridoxal 5'-phosphate (PLP)-dependent aminomutases catalyze the interchange of a hydrogen atom with an amino group on the adjacent carbon atom. These enzymes (see Table 1) have been purified from several strict anaerobes that catabolize lysine or ornithine to organic acids and ammonia. The salient features of the generic aminomutase mechanism are the formation of a Schiff base linkage between the substrate and PLP and the use of 5'-deoxyadenosyl radical to initiate catalysis. In these enzymes, the PLP may assist in the intramolecular migration of the amino group. In addition to the presence of PLP, these enzymes differ from the enzymes that catalyze migration and elimination reactions in that aminomutases catalyze a reversible interchange of a hydrogen atom with the amino group on the adjacent carbon atom and the cofactor binds in a base-off configuration. Despite these differences, the catalytic cycle shown in Figure 2(a) applies to these enzymes as well.

Table 1 Adenosylcobalamin-dependent radical-mediated transformations

Group 1	Rearrangement/elimination reactions	
	Diol dehydratase	1,2-propanediol → propionaldehyde + water
		Ethylene glycol → acetaldehyde + water
	Glycerol dehydratase	Glycerol → 3-hydroxypropionaldehyde
	Ethanolamine ammonialyase	Ethanolamine → acetaldehyde + ammonia
Group 2	Carbon skeleton mutases	
	Methylmalonyl-CoA mutase	(2R)-methylmalonyl-CoA ⇌ succinyl-CoA
	Isobutyryl-CoA mutase	Isobutyryl-CoA ⇌ <i>n</i> -butyryl-CoA
	Glutamate mutase	<i>S</i> -glutamate ⇌ (2 <i>S</i> ,3 <i>S</i>)-3-methylaspartate
	2-Methyleneglutarate mutase	2-methyleneglutarate ⇌ (R)-3-methylitaconate
Group 3	PLP-dependent aminomutases	
	Lysine 5,6-aminomutase	D-lysine ⇌ 2,5-diaminohexanoic acid
		L-β-lysine ⇌ 3,5-diaminohexanoic acid
	D-ornithine 4,5-aminomutase	D-ornithine ⇌ (2 <i>R</i> ,4 <i>S</i>)-diaminopentanoate
Group 4	Ribonucleotide reduction	
	Ribonucleotide triphosphate reductase	Ribonucleotide triphosphate → deoxyribonucleotide triphosphate

Ribonucleotide Triphosphate Reductase Reaction

Ribonucleotide reductases catalyze the conversion of nucleotides to deoxynucleotides in all organisms. The prominent role of these enzymes in nucleotide metabolism has made them attractive targets of antiviral and antitumor therapies. Ribonucleotide reductase can synthesize all four DNA bases from the corresponding nucleosides; however, the activity of the enzymes from all sources is allosterically regulated by the ratio of the deoxyribonucleotides and nucleotides. Four classes of reductases have been described. Although the overall reactions catalyzed by all of these enzymes are identical, they differ in two respects. First, some use nucleoside diphosphates as substrates, whereas others prefer nucleotide triphosphates. Second, these enzymes differ in the source of the species that initiates the radical turnover cascade. In fact, ribonucleotide reductases have been categorized into four classes based on the free-radical initiators. Only one of these, the class II ribonucleotide triphosphate reductase, is adenosylcobalamin dependent. However, the catalytic transformations that ensue following the formation of the substrate radical are remarkably similar. The ribonucleotide triphosphate-specific enzyme from *Lactobacillus leichmannii* is the best-characterized example of adenosylcobalamin-dependent reductases.

The mechanistic details of the ribonucleotide reduction differ significantly from the paradigms that have been discussed in

the previous sections for the group migration and elimination, aminomutase, and carbon skeleton mutase enzymes. First, ribonucleotide reductases contain a pair of active-site cysteine residues that are oxidized in the course of the reaction and must undergo reduction between each catalytic cycle. Second, a cysteine on the protein forms a thiyl radical that is directly involved in the reaction. **Figure 2(b)** shows a general catalytic cycle for the adenosylcobalamin-dependent ribonucleotide reductase, highlighting the differences between this enzyme and other radical-mediated rearrangement reactions. The carbon–cobalt bond of the cofactor is homolyzed by the protein in the presence of allosteric effector; the resulting 5'-deoxyadenosyl radical abstracts a hydrogen atom from an active-site cysteine residue to generate a thiyl radical, which in turn abstracts a hydrogen atom from C3' of the substrate. In the course of reduction of the substrate to the product, a pair of active-site thiols is oxidized and the active-site thiyl radical is reformed. Reformation of the cofactor and reduction of the active-site thiols prepare the active site for subsequent turnover.

Methyl Transfer Reactions

Enzymes that catalyze cobalamin-dependent methyl transfer reactions use the potential for heterolytic cleavage of the carbon–cobalt bond of the cofactor (see **Figure 1(b)**). The cofactor

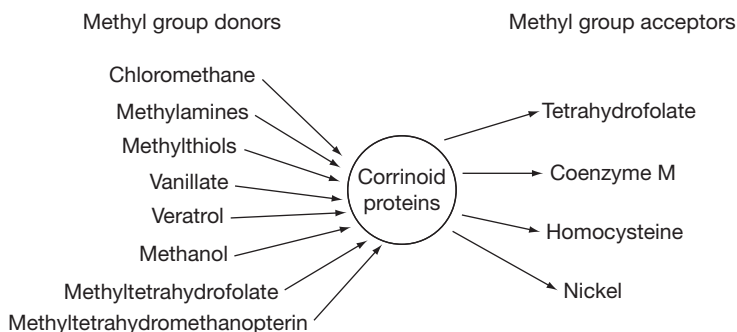


Figure 3 Diversity of methyl group donors and acceptors in cobalamin-dependent methyltransferase enzymes.

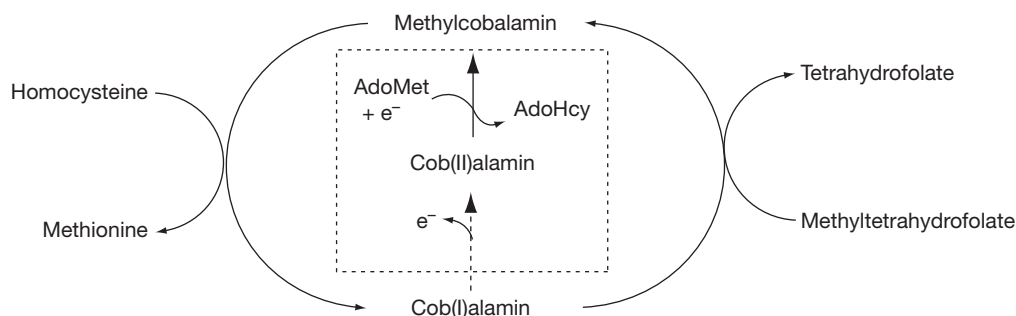


Figure 4 Catalytic cycle of cobalamin-dependent methionine synthase. The dashed box shows the reactivation of the oxidized cofactor by reductive methylation.

in these enzymes serves both as a nucleophile and as a leaving group, undergoing transient methylation in the course of the reaction. Although the identities of the donor and acceptor molecules vary among the family of cobalamin-dependent methyl transferase enzymes (**Figure 3**), the overall catalytic cycles of these enzymes are similar. In this section, we consider methionine synthase as a prototype of the cobalamin-dependent methyltransferases, highlighting the similarities and differences among the members of the methyl transferase family of enzymes.

Cobalamin-Dependent Methionine Synthase

Cobalamin-dependent methionine synthase is found in mammals and in *Caenorhabditis elegans*, but not in insects or in plants, which neither produce nor transport cobalamin. It is also found in many prokaryotes, including *Escherichia coli*, but not in Archaea. The enzyme catalyzes the transfer of a methyl group from the tertiary amine methyltetrahydrofolate to the thiol of homocysteine, and the cobalamin cofactor serves as an intermediary in the methyl transfer. A catalytic cycle for methionine synthase is shown in **Figure 4**.

Methionine synthase is a large monomeric protein (1227 amino acid residues in the enzyme from *E. coli*) containing four discrete modules. The N-terminal module binds and activates homocysteine; the thiol of homocysteine coordinates to a catalytically essential zinc metal ion in this module. The next module in the sequence binds and activates methyltetrahydrofolate. Each of these substrate-binding modules communicates with the third module in the sequence, the B₁₂-binding

module, to transfer methyl groups to and from the base-off cofactor. The C-terminal module is required for the activation of methionine synthase when the cob(I)alamin cofactor becomes oxidized during turnover (1 in 2000 turnovers results in oxidation *in vitro*). Activation of the inactive cob(II)alamin form of the cofactor requires both a methyl donor and an electron donor. The methyl donor is S-adenosyl-methionine. In *E. coli* the electron donor is reduced flavodoxin, and in mammals it is methionine synthase reductase, a protein containing a domain with homology to flavodoxin. To allow methyl transfer during reductive activation, the C-terminal domain must also access the cobalamin; a structure of a fragment of methionine synthase in this conformation has recently been determined.

Methyl Transferases Involved in Methanogenesis and Acetogenesis

Although cobalamin-dependent methionine synthase is the only B₁₂-dependent methyl transferase in eubacteria and eukaryotes, a large number of enzymes catalyzing similar methyl transfers that employ corrinoid cofactors have been identified in Archaea and acetogenic eubacteria. The diversity in the methyl group donor-acceptor pairs is illustrated in **Figure 3**. In general, in methanogens, corrinoid-dependent enzymes are involved in the formation of methylcoenzyme M, the methyl thioether that provides the methyl group destined to form methane in methanogenesis. The methyl donors in these reactions include methyltetrahydromethanopterin, which is

chemically similar to methyltetrahydrofolate. The methyl group of methyltetrahydromethanopterin is formed by the reduction of carbon dioxide using reducing equivalents derived from molecular hydrogen. Some methanogens can also use simple compounds such as methylamines, methanol, or methylsulfides as the source of methyl groups for methanogenesis. In contrast to cobalamin-dependent methionine synthase, in which a single protein catalyzes methyl transfer from methyltetrahydrofolate to cobalamin and from methylcobalamin to homocysteine, most of these methanogenic methyl transferase reactions require three separate proteins, a corrinoid-containing protein, a methyltransferase that transfers methyl groups from the donor molecule to the corrinoid protein, and a second methyltransferase that transfers methyl groups from the methylcorrinoid protein to the acceptor.

Eubacterial acetogenesis also involves corrinoid-dependent methyl transferases. In these organisms, carbon dioxide and hydrogen are used to generate acetyl coenzyme A (CoA). The methyl group of acetyl CoA is generated by reduction of carbon dioxide using reducing equivalents from molecular hydrogen, as in methanogens, and is transferred to a corrinoid cofactor that serves as the direct methyl donor to the enzyme responsible for the synthesis of acetyl CoA from a methyl group and carbon dioxide. Additional acetogenic substrates include vanillate, veratrol, and halogenated aryl or alkene compounds. The precise nature of the chemistry that is promoted by cobalamin with the last substrates remains to be elucidated.

See also: **Molecular Biology:** DNA Methyltransferases: Eubacterial GATC; **Protein/Enzyme Structure Function and Degradation:** Pyridoxal Phosphate.

Further Reading

- Banerjee R (ed.) (1999) *Chemistry and Biochemistry of B₁₂*. New York, NY: Wiley.
- Dolphin D (ed.) (1982) *B₁₂*, vol. 2. New York, NY: Wiley.
- Hay BP and Finke RG (1987) Thermolysis of the Co–C bond in adenosylcobalamin. 3. Quantification of the axial base effect in adenosylcobalamin by the synthesis and thermolysis of axial base-free adenosylcobinamide: Insights into the energetics of enzyme-assisted cobalt–carbon bond homolysis. *Journal of the American Chemical Society* 109: 8012–8018.
- Licht S and Stubbe J (1999) Mechanistic investigations of ribonucleotide reductases. In: Poulter CD (ed.) *Comprehensive Natural Products Chemistry, Enzyme and Enzyme Mechanisms, Proteins, and Aspects of NO Chemistry*, vol. 5, pp. 163–204. Oxford: Elsevier.
- Marsh EN and Drennan CL (2001) Adenosylcobalamin-dependent isomerases: New insights into structure and mechanism. *Current Opinion in Chemical Biology* 5: 499–505.
- Sintchak MD, Arjara G, Kellogg BA, Stubbe J, and Drennan CL (2002) The crystal structure of class II ribonucleotide reductase reveals how an allosterically regulated monomer mimics a dimer. *Nature Structural Biology* 9: 293–300.
- Thauer RK (1998) Biochemistry of methanogenesis: A tribute to Marjory Stephenson. 1998 Marjory Stephenson prize lecture. *Microbiology* 144: 2377–2406.
- Toraya T (2000) The structure and the mechanism of action of coenzyme B₁₂-dependent diol dehydratases. *Journal of Molecular Catalysis B* 10: 87–106.

B-Cell Antigen Receptor

T M Yankee, University of Kansas Medical Center, Kansas City, KS, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Antibody A protein that is produced in response to immune challenge and binds specifically to a particular antigen.

Antigen A molecule that specifically binds an antibody.

Avidity The sum of the binding strengths of all points of interaction between an antigen and an immunoglobulin.

Humoral immune response The antibody-mediated response to a specific antigen.

Immunoglobulin A protein complex with a characteristic structure of heavy and light chains.

Isotype A class of immunoglobulin as determined by the constant region.

Somatic hypermutation The process by which mutations are introduced into the heavy- and light-chain genes.

B-cell functions in the immune system are dependent on their ability to recognize foreign antigens and to discriminate between foreign antigens and self-antigens. The complex of proteins responsible for the identification of foreign antigens is the B-cell antigen receptor (BCR) complex. Signals originating with the BCR can lead to a variety of cell fates, depending on the developmental stage of the B cell and the concentration and avidity of the antigen. B-cell precursors in the bone marrow require BCR signals for survival. Once a competent signaling complex is formed, the BCR is tested to ensure that it does not react with self-antigens. If the BCR is ligated by self-antigens, it must either change its specificity or undergo cell death. Later in B-cell development, basal BCR signaling is required for the survival and homeostatic maintenance of the B-cell pool. BCR stimulation of mature B cells initiates an immune response, characterized by the proliferation and differentiation of B cells into either antibody-producing plasma cells or memory B cells. Hence, a competent BCR that recognizes foreign antigens and not self-antigens is critical for the development and maintenance of B cells and for the initiation of humoral immune responses.

Structure of the BCR

The BCR complex contains a membrane-bound immunoglobulin (mIg). mIg can be one of five isotypes: IgM, IgD, IgA, IgG, or IgE. The isotype of mIg expressed on a given B cell varies with the stage of development and activation. Immature and transitional B cells express mIgM, whereas mature resting B cells express mostly mIgD and some mIgM. Cells that have been activated undergo a process called 'isotype class-switching' and may then express mIgG, mIgA, or mIgE in addition to soluble IgG, IgA, or IgE. The process of isotype class-switching results in the activation of other hematopoietic cells. Each molecule of mIgM contains two heavy-chain (H) and two light-chain (L) proteins, that is H2L2 (Figure 1). Disulfide bonds link the heavy chains to one another and each heavy chain is disulfide bonded to a light chain. At the membrane proximal end of the heavy chain is a transmembrane domain followed by a short intracellular domain. mIgM and mIgD contain only three intracellular residues that anchor the

protein in the membrane. Other classes of mIg contain slightly longer intracellular domains and appear to mediate functions only present in B cells that express mIgG, mIgA, or mIgE.

The short intracellular domains of mIgM and mIgD alone cannot generate a signal; mIg-associated accessory proteins fulfill this function. There are two such accessory molecules, Ig α (also called CD79a) and Ig β (CD79b). The interaction between mIg and Ig α /Ig β depends on the transmembrane domains of each protein. The transmembrane domains of mIg, Ig α , and Ig β are α -helices and contain both polar and nonpolar regions. The polar regions are critical for protein-protein interactions among the BCR components. mIg contains polar residues on both sides of the α -helix. One side is conserved among all Ig isotypes and the other side is isotype specific. The residues that are isotype specific allow for oligomerization of the BCR complex. The degree of oligomerization varies with each Ig isotype. The side of the molecule that is conserved among all isotypes participates in the association between mIg and Ig α /Ig β .

Like mIg, the transmembrane domain of Ig α contains polar residues on both sides of the α -helix. By contrast, Ig β contains polar residues on only one side of the α -helix. Thus, a favored model for the structure of the BCR complex is that Ig α bridges mIg and Ig β (Figure 1). According to this model, when the BCR complex is assembled, the polar residues within the transmembrane domains are masked by protein-protein interactions. This is essential for the surface expression and stability of the BCR.

BCR Diversity

In order to contribute to protective immunity to a range of pathogens, B cells must be able to recognize a vast array of antigens. The diversity of the Ig repertoire is accomplished using three factors: combinatorial diversity, junctional diversity, and somatic hypermutation. Combinatorial diversity occurs through the recombination of sets of heavy- and light-chain gene segments. Three gene segments, variable (VH), diversity (DH), and joining (JH), recombine to form the variable region of the heavy chain. There are many VH gene segments, DH gene segments, and JH gene segments. Thus, sets of

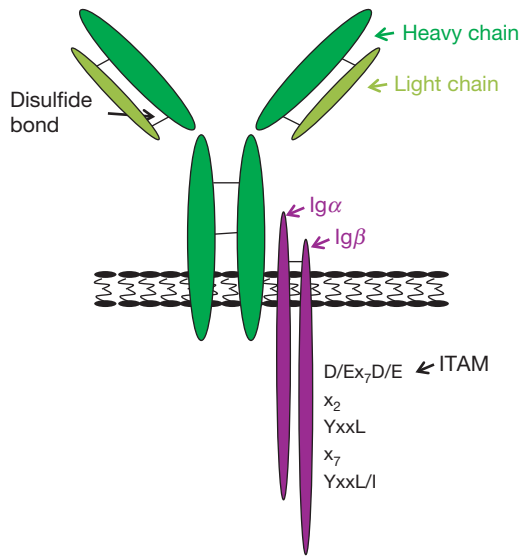


Figure 1 The structure of the BCR. The BCR consists of two heavy chains disulfide bonded to one another. Each heavy chain is disulfide bonded to a light chain. Ig α and Ig β are also disulfide linked. mIg and Ig α /Ig β are noncovalently coupled. Also shown is the ITAM motif on Ig β ; Ig α and Ig β each have one ITAM motif.

different segments recombine to form a range of different VDJH genes. The light chain undergoes a similar process of recombination. To generate the light chain, two gene loci, κ and λ , undergo recombination. Here, VL and JL gene segments recombine and result in VJL. The combination of heavy-chain rearrangement and light-chain rearrangement results in substantial BCR diversity.

Junctional diversity also arises from the recombination of the gene segments. When a VH gene segment recombines, nucleotides may be added or subtracted. This results in the addition or loss of amino acids or a shift in the reading frame of the new gene. The third mechanism by which diversity is introduced to the BCR repertoire is through somatic hypermutation (discussed later).

BCR Signaling

Ligation of the BCR complex triggers a series of events that ultimately affects the fate of the cell. A B cell stimulated through the antigen receptor may proliferate, differentiate, or undergo apoptosis. B cells are also very efficient antigen-presenting cells. BCR signaling is essential for antigen internalization and processing after the BCR binds and captures antigen.

Antibody–Antigen Interactions

Antigens interact with the variable regions of mIg. Within the variable region, there are framework regions interspersed with three hypervariable regions, the areas that contain the most variability. The framework regions, folded as β -sheets, provide much of the antibody structure. The hypervariable regions are

located on one edge of the β -sheets. To provide even greater repertoire diversity, the hypervariable regions of the heavy and light chain are within close proximity to one another and together create the antigen-binding site. Haptens and other small antigens bind the antibody in the grooves between the hypervariable regions of the heavy and light chains. Interactions between the BCR and larger antigens may extend into the framework region. The interactions between the antibodies and antigens are noncovalent in nature. These interactions may consist of electrostatic interactions, van der Waals interactions, hydrophobic interactions, and hydrogen bonding.

Early Signaling Events

Ig α and Ig β each contains tyrosine residues that are phosphorylated following BCR engagement. Two of the tyrosine phosphorylation sites on Ig α and Ig β are located within immunoreceptor tyrosine-based activated motifs (ITAMs) (Figure 1). These regions contain six conserved residues in the specific configuration, D/ExxxxxxD/ExYxxLxxxxxxYxxL/I. The spacing between the tyrosine residues in the ITAM results in these residues being positioned on the same side of an α -helix, facilitating the interactions with downstream SH2-domain-containing signaling molecules. (SH2 domains are motifs that bind phosphotyrosine and the surrounding residues.)

In addition to the ITAM tyrosine residues, Ig α and Ig β contain other residues that become phosphorylated. Non-ITAM tyrosine residues may also recruit SH2-domain-containing signaling proteins. Serine/threonine phosphorylation of Ig α may negatively regulate phosphorylation of the ITAM tyrosine residues and, therefore, may negatively regulate downstream signaling.

Three families of protein tyrosine kinases (PTKs) play key roles in the initiation of BCR signaling ([Figure 2](#)). These kinases are Src-family PTKs, such as Lyn; the Syk/ZAP-70 family PTKs, such as Syk; and the Tec-family PTKs, such as Btk. Following BCR ligation, Lyn becomes activated and phosphorylates the ITAM tyrosines. The phosphorylated ITAM becomes the docking sites for Syk, which has two tandem SH2 domains. The binding of the SH2 domains of Syk to the phospho-ITAM contributes to its activation. Syk can also become activated in the absence of Src-family PTKs. A small amount of Syk is constitutively associated with the BCR complex. Cross-linking the BCR leads to the clustering of Syk, which then may transautophosphorylate and become activated. Activation of Syk and Lyn, then, leads to the activation of Btk. The combination of Syk, Lyn, and Btk activity is necessary for optimal signaling through the BCR.

Downstream Signaling Events

The activation of Syk-, Src-, and Tec-family PTKs trigger a number of downstream signaling pathways (Figure 2). For example, phosphatidylinositol 3'-kinase (PI3-K) and phospholipase C- γ 2 (PLC- γ 2) are two enzymes that generate second messengers in BCR signaling. PI3-K is a lipid kinase important for the recruitment and activation of PH-domain-containing proteins. (PH domains are motifs that bind phospholipids.)

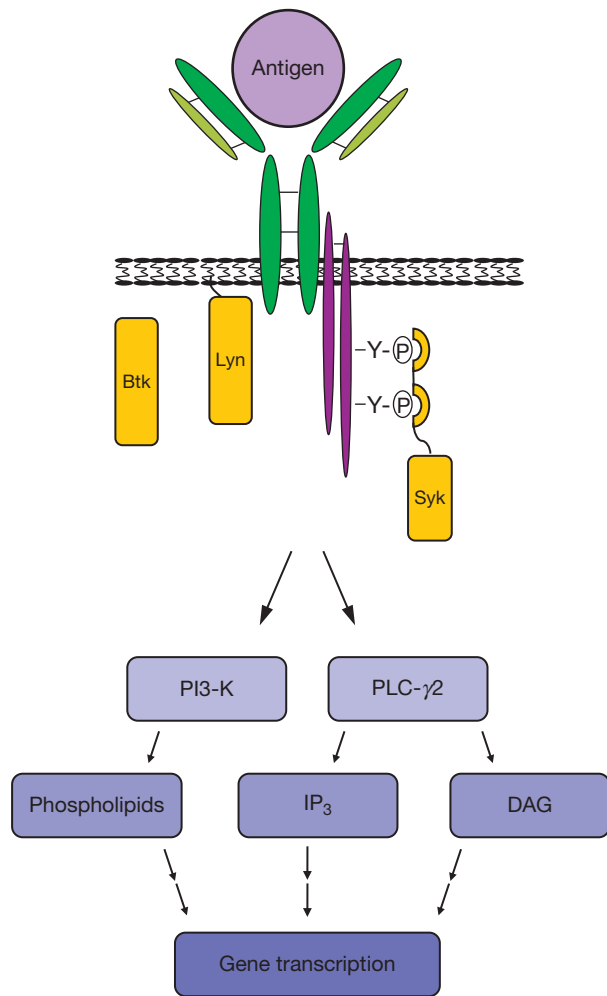


Figure 2 BCR-mediated signaling events. Following antigen binding to the BCR complex, tyrosine residues on Igα and Igβ become phosphorylated and bound by the tandem SH2 domains of the Syk. Lyn is also activated. The activation of these two PTKs trigger many downstream signaling cascades, including the activation of PI3-K and PLC-γ2.

PLC-γ2 generates inositol-1,4,5-triphosphate (IP₃) and diacylglycerol (DAG), second messengers critical for calcium influx and protein kinase C activation. The activation of these signaling pathways, along with many other pathways, leads to gene transcription and cell-fate decisions, such as proliferation, apoptosis, or differentiation.

In addition to gene transcription, BCR signaling is also critical for antigen internalization and processing. When the BCR binds an antigen, the BCR complex is internalized. The antigen is then processed and presented on the cell surface in order to activate antigen-specific T cells.

The BCR and the Immune Response

When antigen and other immune cells activate a B cell, a complex series of events takes place that can result in proliferation, somatic hypermutation, isotype class-switching, and

differentiation. The BCR on mature naïve B cells may be able to recognize more than one antigen. This multispecific nature of the BCR results in relatively low affinity to any specific antigen. Somatic hypermutation is a process by which high-affinity BCRs are generated. Under some conditions following BCR ligation, a somatic hypermutation program is induced. During somatic hypermutation, mutations are introduced into the V segment of the heavy- and light-chain genes. After somatic hypermutation, B cells expressing BCRs with high affinity are selected for further expansion. In this way, the immune system creates B cells that have produced high-affinity antibodies tailor-made for specific pathogens.

The other major changes in the BCR following B-cell activation include isotype class-switching. The genomic organization of mIg is the recombined variable region followed by the constant genes of μ , δ , $\gamma 3$, $\gamma 1$, $\gamma 2b$, $\gamma 2a$, and α in mice and μ , δ , $\gamma 3$, $\gamma 1$, $\alpha 1$, $\gamma 2$, $\gamma 4$, and $\alpha 2$ in humans. Following the activation of the B cell through certain key receptors such as CD40, DNA recombination can occur so that the VH joins a CH segment of another isotype. This process is a guided repetitive DNA sequence preceding each constant domain known as a 'switch region'. The precise mechanism of class-switching is unknown, but probably involves DNA recombination between homologous repeats in the switch regions of different constant genes.

The result of somatic hypermutation and class-switching is the generation of highly specific antibodies of an isotype that generates an effective immune response. In particular, the constant regions of the antibody bind specific receptors on other immune cells. For example, the constant region of IgGγ1 and IgGγ3 bind Fcγ receptors on macrophages and neutrophils leading to phagocytosis of the antigen, whereas the constant region of IgE binds Fc receptors on mast cells and basophils leading to the secretion of inflammatory agents.

Membrane-bound forms of IgG, IgA, and IgE may function differently than mIgM and mIgD. Cross-linking mIgG leads to the proliferation and antibody secretion of mIgG-expressing B cells. Although mIgG and mIgM activate similar downstream signaling pathways, mIgG cross-linking triggers a much more robust response than does mIgM cross-linking. This may be due to the fact that mIgG is resistant to downregulation by certain B-cell coreceptors, such as CD22. The mechanism for this resistance may be related to differences in the intracellular domains between mIgM and mIgG.

Summary

The BCR complex is a critical component of the humoral immune response. The BCR repertoire must be diverse enough to recognize an incredible number of foreign antigens, but must also avoid targeting self-antigens. Ligation of the BCR triggers a complex series of events that include the activation of PTKs, the phosphorylation of Igα and Igβ, and the initiation of multiple signaling cascades. These signaling pathways lead to cell-fate decisions, antigen presentation, and an immune response.

See also: [Signaling: Diacylglycerol Kinases and Phosphatidic Acid Phosphatases; Inositol Phosphate Kinases and Phosphatases;](#)

Phosphatidylinositol Bisphosphate and Trisphosphate; Protein Tyrosine Phosphatases; Src Family of Protein Tyrosine Kinases.

Further Reading

Cyster J (1997) Signaling thresholds and interclonal competition in preimmune B-cell selection. *Immunological Reviews* 156: 87–101.

Janeway C, Travers P, Walport M, and Schlomchick M (2001) *Immunobiology: The Immune System in Health and Disease*. New York, NY: Garland Publishing.

Matsuuchi L and Gold M (2001) New views on BCR structure and function. *Current Opinion in Immunology* 13: 270–277.

Melchers F, ten Boekel E, Seidl T, et al. (2000) Repertoire selection by pre-B-cell receptors and B-cell receptors, and genetic control of B-cell development from immature to mature B-cells. *Immunological Reviews* 175: 33–46.

Niir H and Clark E (2002) Regulation of B-cell fate by antigen receptor signals. *Nature Reviews Immunology* 2: 945–956.

Reth M (1992) Antigen receptors on B-lymphocytes. *Annual Review of Immunology* 10: 97–121.

Yankee Y and Clark E (2000) Signaling through the B-cell antigen receptor in developing B-cells. *Reviews in Immunogenetics* 2: 185–203.

Bcl2 Family: Cell Death Enhancers and Inhibitors

D L Vaux, La Trobe University, Melbourne, VIC, Australia

© 2013 Elsevier Inc. All rights reserved.

Glossary

Apoptosis A physiological form of cell death with a characteristic morphology whose mechanism is shared by metazoans.

Bcl2 family proteins Proteins that have a structural similarity to the apoptosis inhibitory protein Bcl2.

BH domains Bcl2 homology domains; there are four, BH1–4.

BH3-only proteins A group of pro-apoptotic proteins that have the BH3 domain but no other BH domains.

Caspases Cysteine proteases with aspartic acid specificity that are related to the *Caenorhabditis elegans* caspase CED-3 that is essential for programmed cell death in the worm.

ced genes Genes that implement programmed cell death in the worm *C. elegans*.

Apoptosis, the physiological cell death mechanism used by metazoans to remove unwanted cells, is controlled by a family of pro- and anti-apoptotic proteins that bear varying degrees of similarity to a protein called Bcl2. Some members of this family, such as Bcl2 itself, stop cells from dying, whereas other members of this family, such as Bax and Bak, are part of the mechanism that cells use to kill themselves.

Identification of Bcl2 as a Cell Death Inhibitor

A gene at the site of chromosomal translocations associated with the common lymphoid cancer, follicular lymphoma, was designated Bcl2 for B-cell leukemia/lymphoma gene number 2. Rather than promoting cell growth and proliferation like most cancer genes, Bcl2 does not stimulate cells to divide but prevents them from activating their endogenous apoptotic cell suicide mechanism. The association of activation of the apoptosis inhibitor Bcl2 with lymphoma was the first evidence that cell death is required to avoid the development of cancer.

Genetics of Cell Death in *Caenorhabditis elegans*

Although Bcl2 was the first component of the apoptosis mechanism to be recognized at the molecular level in any organism, genetic analysis of programmed cell death in the nematode *Caenorhabditis elegans* revealed that developmental cell death in worms is controlled and implemented by a specific genetic program. Programmed cell death in *C. elegans* fails to occur in *egl-1*, *ced-4*, and *ced-3* loss-of-function mutants, as well as in *ced-9* gain-of-function mutants, implying that *egl-1*, *ced-4*, and *ced-3* encode killer proteins, and that *ced-9* encodes a survival protein. The expression of human Bcl2 in the worm, and the cloning and sequencing of *ced-9*, showed that they are structurally and functionally homologous. In the worm, CED-9 prevents cell death by preventing the adaptor CED-4 from activating the protease CED-3, and in mammalian cells Bcl2 stops apoptosis by indirectly preventing activation of CED-3-like proteases, termed caspases.

Three Classes of Bcl2 Family Proteins

In vertebrates, apoptosis is regulated by three classes of Bcl2-like proteins that directly interact with each other. Members of this family all share regions of similarity termed Bcl2 homology (BH) domains. The three classes are (1) anti-apoptotic proteins such as Bcl2 itself, Bclx, and Mcl1; (2) pro-apoptotic proteins such as Bax and Bak; and (3) the pro-apoptotic BH3-only proteins, such as Puma, Noxa, Bad, Bim, and Bid, which only have a single BH domain.

The BH3-only proteins are the triggers of the death mechanism; Bcl2 and the anti-apoptotic proteins are like safety catches; and Bax and Bak can be thought of as the bullets.

When activated, Bax and Bak can form structures in the outer membrane of the mitochondria that allow cytochrome *c* to escape into the cytosol, where it binds to the CED-4 homolog Apaf-1, which, in turn, activates the protease caspase-9. Although Bax and Bak can activate spontaneously, this can be triggered by binding of BH3-only proteins, and is inhibited by binding of anti-apoptotic Bcl2 family proteins. Adding to the complexity of regulation of cell death, BH3-only proteins can also be bound by anti-apoptotic Bcl2 family members (Figure 1).

Interactions between Bcl2 family proteins determine whether and when a cell kills itself. For example, when the DNA in a cell is damaged, such as by ultraviolet (UV) light in sunshine, the cell detects the damage and increases levels of a protein called p53. In the nucleus, p53 turns on a number of genes, including the gene for the BH3-only protein Puma. Puma can cause activation of Bax and Bak by binding to them, as well as by binding to Bcl2, Bclx, and Mcl1, thereby preventing their inhibition of Bax and Bak. Either way, Bax and Bak activate and multimerize, to cause the outer mitochondrial membrane to become permeable, releasing cytochrome *c*, which activates Apaf-1 and caspases in the cytosol, culminating in cell death.

Therapeutic Implications

The initial cause of the malignancy follicular lymphoma is a chromosomal translocation that causes excess production of Bcl2 in a lymphocyte. Because of this, the cell and its progeny

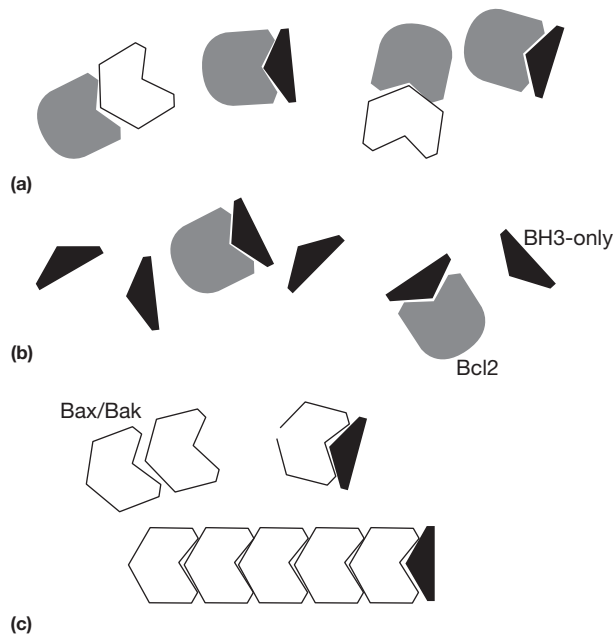


Figure 1 Interactions between Bcl2 family members. (a) In healthy cells, BH3-only proteins such as Bim, Bad, Bid, Noxa, Puma (black), and Bax and Bak (white) are kept inactive by anti-apoptotic proteins such as Bcl2 (gray). When a cell receives a signal to die (b), levels of BH3-only proteins increase, and levels of anti-apoptotic proteins decrease. This frees Bax and Bak to spontaneously multimerize in the mitochondrial outer membrane (c), and multimerization is also facilitated by direct binding of BH3-only proteins. Multimerized Bax and Bak somehow allow cytochrome *c* to exit from the mitochondria to the cytosol, where it can bind to and activate Apaf-1 and caspases.

cannot kill themselves, and accumulate, eventually becoming malignant. The discovery that Bcl2 could be inhibited by BH3-only proteins prompted development of artificial BH3-only mimetic compounds that could be used to treat lymphoma. A compound called ABT-263 developed by Abbott is a Bcl2 antagonist that causes death of tumor cells by binding to Bcl2. It is currently in phase 2 clinical trials in humans.

See also: **Bioenergetics:** Cytochrome *c*; **Cell Architecture and Function:** Caspases and Cell Death; **Protein/Enzyme Structure Function and Degradation:** Cysteine Proteases.

Further Reading

- Adams JM and Cory S (2001) Life-or-death decisions by the Bcl-2 protein family. *Trends in Biochemical Sciences* 26: 61–66.
- Huang DCS and Strasser A (2000) BH3-only proteins – essential initiators of apoptotic cell death. *Cell* 103: 839–842.
- Lindsten T, Ross AJ, King A, et al. (2000) The combined functions of proapoptotic Bcl-2 family members Bak and Bax are essential for normal development of multiple tissues. *Molecular Cell* 6: 1389–1399.
- Metzstein MM, Stanfield GM, and Horvitz HR (1998) Genetics of programmed cell death in *C. elegans* – past, present and future. *Trends in Genetics* 14: 410–416.
- Oltersdorf T, Elmore SW, Shoemaker AR, et al. (2005) An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* 435: 677–681.
- Sattler M, Liang H, Nettesheim D, et al. (1997) Structure of Bcl-x(l)-Bak peptide complex – recognition between regulators of apoptosis. *Science* 275: 983–986.
- Vaux DL, Cory S, and Adams JM (1988) Bcl-2 gene promotes haemopoietic cell survival and cooperates with c-myc to immortalize pre-B cells. *Nature* 335: 440–442.
- Vaux DL, Weissman IL, and Kim SK (1992) Prevention of programmed cell death in *Caenorhabditis elegans* by human bcl-2. *Science* 258: 1955–1957.

Bile Salts

L Maillette de Buy Wenniger, University of Amsterdam, Amsterdam, The Netherlands

T Pusch, Klinikum Augsburg, Augsburg, Germany

U Beuers, University of Amsterdam, Amsterdam, The Netherlands

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Ulrich Beuers and Thomas Pusch, volume 1, pp. 159–163, © 2004, Elsevier Inc.

Glossary

Bile acid Protonated bile salt.

Bile salt Organic anion with a steroid nucleus formed by enzymatic conversion of cholesterol (major human bile salts: cholate, chenodeoxycholate, deoxycholate, lithocholate, and ursodeoxycholate).

Bile salt transporters Membrane carrier proteins mediating the transport of bile salts across cell membranes. The Na⁺-taurocholate-cotransporting polypeptide (NTCP) at the basolateral hepatocyte membrane, the bile salt export pump (BSEP) at the apical hepatocyte membrane, and the apical sodium bile salt transporter (ASBT) at the apical membrane of ileocytes and cholangiocytes represent key carriers for the maintenance of an effective enterohepatic circulation of bile salts.

Cholestasis Impairment of bile formation and flow, caused by hepatocellular or cholangiocellular

secretory failure, intrahepatic or extrahepatic bile duct obstruction, or a combination of these factors, and leading to a retention of bile constituents in liver, serum, and other organs.

Enterohepatic circulation Effective recycling of bile salts from the intestine back to the liver, where bile salts can be resecreted into bile and intestine to fulfill their various functions as metabolic regulators and surfactants.

Nuclear hormone receptors Ligand-activated transcription factors regulating gene expression by interaction with specific DNA sequences. Bile salts have so far been identified to act as physiological ligands of the farnesoid X receptor (FXR), the pregnane X receptor (PXR), and the vitamin D receptor (VDR), thereby modulating their own metabolism and transport.

Introduction

Bile salts play a crucial role in hepatobiliary and intestinal homeostasis and digestion. The liver synthesizes primary bile salts from cholesterol. Secondary and tertiary bile salts with a similar molecular structure derive their individual properties from further chemical alterations by bacterial enzymes and the liver during their enterohepatic circulation.

Under physiological conditions, bile salts can solubilize lipophilic molecules in the intestine to enable their uptake from the lumen. Above a critical micellar concentration, molecular self-aggregation occurs, and micelles are formed. Internalization of hydrophobic molecules such as cholesterol, phospholipids, and monoglycerides into these mixed bile salt micelles promotes the absorption of dietary lipids and fat-soluble vitamins in the small intestine.

In addition to their function as a surfactant, bile salts are potent signaling molecules in the liver and intestine. In the small intestine, they play a key role in the defense against ascending microbes by farnesoid X-receptor (FXR)-dependent mechanisms and modulate hepatobiliary bile formation by FXR-controlled ileal release of the peptide hormone fibroblast growth factor 19 (FGF19). FGF19 regulates hepatocellular bile salt synthesis via its receptor, FGFR4, on hepatocyte membranes. Bile salts directly modulate their hepatocellular uptake, synthesis, and biliary secretion (and alternative basolateral release in cholestasis) at the transcriptional level via activation of nuclear receptors such as FXR and at the posttranscriptional level via modulation of cytosolic signaling cascades. This modulation

affects the synthesis and trafficking of bile salt transporters to their target membranes and thereby modifies hepatobiliary secretory capacity. Recent evidence indicates that bile salts also function as potent endocrine and paracrine messenger molecules within and outside the liver that affect a range of metabolic processes. These mechanisms are, in part, mediated via the bile salt membrane receptor, G-protein-coupled bile salt receptor, TGR5, which is expressed in various cell types except hepatocytes.

At high hepatocellular concentrations, hydrophobic bile salts may induce cholestasis, apoptosis, and necrosis. Bile salts with anticholestatic and antiapoptotic properties may be applied in clinical practice to modify the circulating bile salt pool in an effort to minimize bile toxicity while optimizing hepatobiliary function.

Biochemistry of Bile Salts

Bile salts consist of a steroid nucleus with its hydroxyl or other substituents and an aliphatic side chain of variable length. This steroid nucleus is similar to the core of steroid hormones and is an evolutionary strongly conserved structure.

In higher vertebrates and other modern animals, the majority of bile salts has a backbone of 24 carbon atoms. The bile salt pool of evolutionary predecessors is often dominated by molecules with 25–27 carbon atoms.

Protonation of bile salts changes them to their acid form. The salt form is more prevalent at a physiological pH (due to a pK_a of 4–5 for glycine-conjugated and unconjugated bile salts

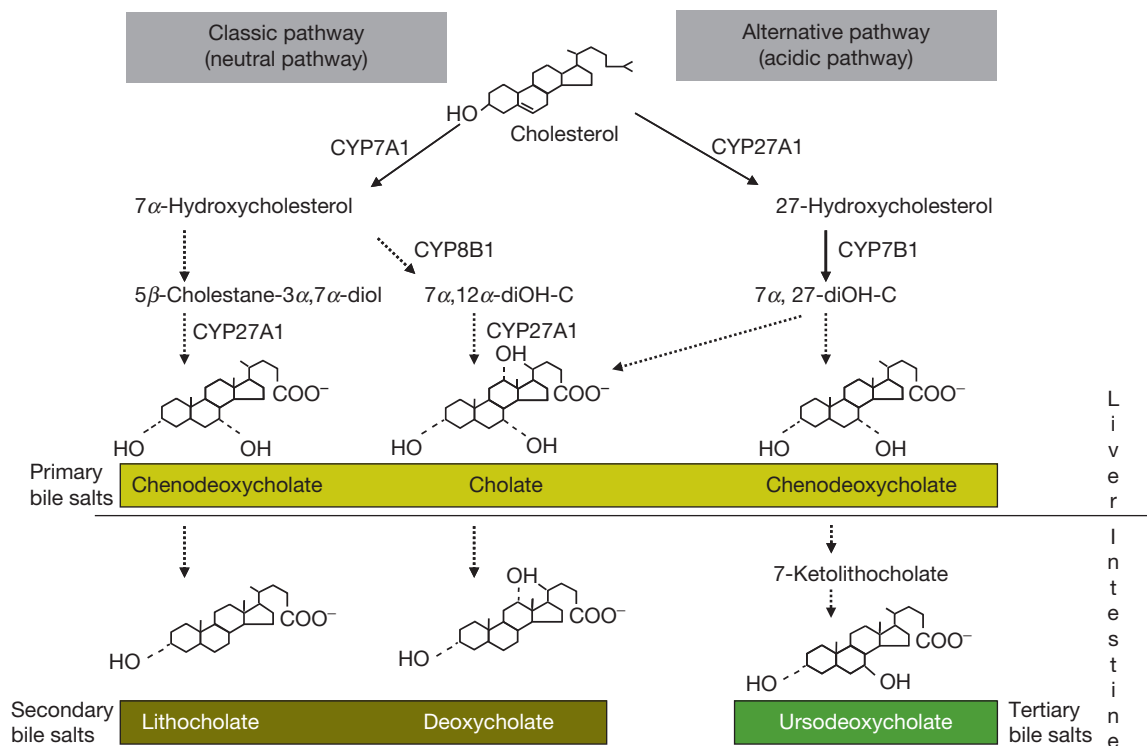


Figure 1 Major bile salt biosynthetic pathways in the liver. The initial and rate-determining step of the classic (neutral) pathway is the conversion of cholesterol to 7 α -hydroxycholesterol by microsomal cholesterol 7 α -hydroxylase (CYP7A1) leading to formation of cholate and chenodeoxycholate. The initial step of the alternative (acidic) pathway is the conversion of cholesterol to 27-hydroxycholesterol by mitochondrial sterol 27-hydroxylase (CYP27A1) leading to formation mainly of chenodeoxycholate. These primary bile salts undergo further structural modifications by bacterial enzymes during enterohepatic cycling resulting in the formation of the secondary bile salts deoxycholate and lithocholate, and, by bacterial and hepatic enzymes, the tertiary bile salt ursodeoxycholate. Only major biosynthetic steps are shown. CYP7B1, oxysterol 7 α -hydroxylase; CYP8B1, sterol 12 α -hydroxylase; 7 α ,27-diOH-C, 7 α ,27-dihydroxycholesterol; 7 α ,12 α -diOH-C, 7 α ,12 α -dihydroxy-4-cholestene-3-one.

and a pK_a of about 1 for taurine-conjugated bile salts). Yet, the term 'bile acid' is often used in literature.

Synthesis of Bile Salts

Bile salts are mainly synthesized in the pericentral hepatocytes from cholesterol by different pathways, which involve up to 17 enzymes located in the endoplasmatic reticulum, mitochondria, cytosol, and peroxisomes (Figure 1).

The rate-limiting step of the classical bile salt biosynthetic pathway, also known as the neutral pathway, is 7 α -hydroxylation of cholesterol by a microsomal cytochrome P-450 monooxygenase (CYP7A1).

The alternative or acidic pathway is a second important route for bile salt synthesis, and is initiated by the mitochondrial enzyme sterol 27-hydroxylase (CYP27A1), which supports the conversion of cholesterol to both 27-hydroxycholesterol and 3 β -hydroxy-5-cholestenoic acid. CYP27A1 is present in some peripheral tissues, such as macrophages and vascular epithelium, and may thus add to the production of intermediate metabolites. For the final conversion to bile salts, they are transported to the liver where the full range of enzymes required for bile salt synthesis is present. Recently, functional bile salt formation was reported in the ovarian follicle, but the relevance of this finding remains to be explored.

The relative contribution of CYP7A1 and CYP27A1 to overall bile salt synthesis is unclear, but CYP7A1 has been estimated to account for 75% (mice) to 90–95% (human) of total bile salt synthesis under physiological conditions. Cholesterol 25-hydroxylase and cholesterol 24-hydroxylase may also initiate bile salt synthesis *in vitro*, but their contribution to total bile salt production may be limited.

Bile salt synthesis is highly regulated and subject to feedforward and feedback control whereby bile salts downregulate their own synthesis and oxysterols upregulate bile salt synthesis, mainly by regulating CYP7A1 gene transcription (Figure 2). Promoter analyses of the CYP7A1 gene have identified two highly conserved bile acid response elements and are binding sites for nuclear hormone receptors.

Nuclear hormone receptors are ligand-activated transcription factors with a strongly conserved DNA-binding domain (DBD) in the N-terminal region and a moderately conserved ligand-binding domain (LBD) in the C-terminal region. Upon ligand binding to the LBD, conformational changes in the nuclear receptors allow dissociation of corepressors and recruitment of coactivator proteins to activate the expression of target genes. The nuclear hormone receptor superfamily includes receptors for steroid and thyroid hormones, retinoids and vitamin A and D, as well as the bile-acid receptors FXR and pregnane X receptor (PXR).

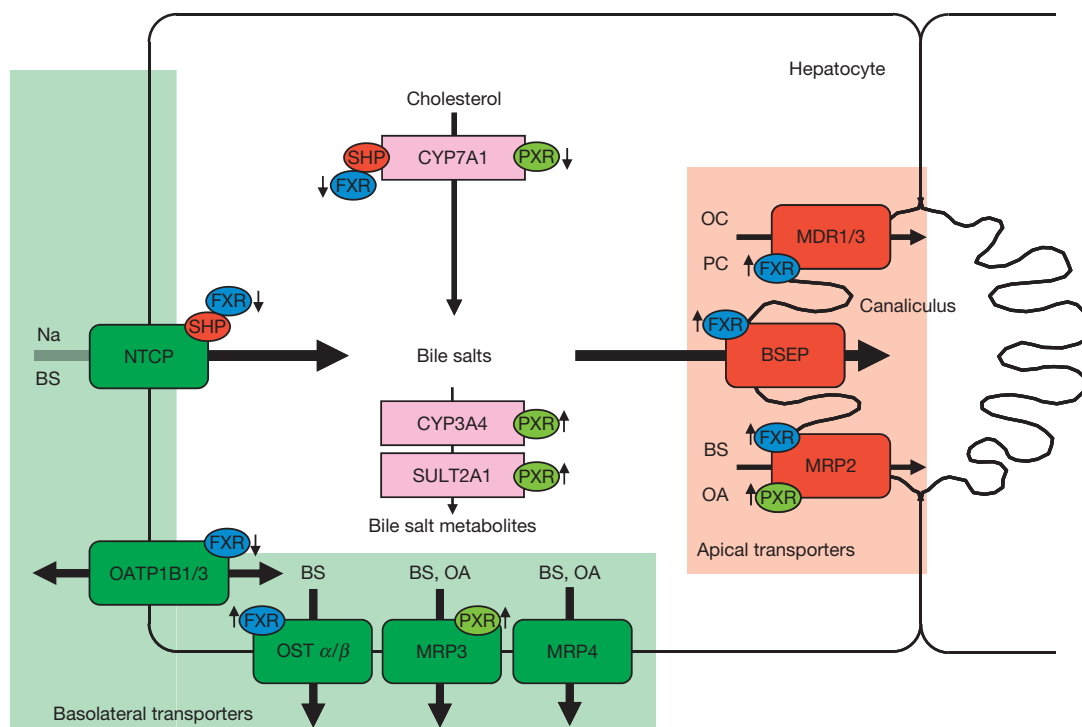


Figure 2 Regulation of bile salt transport and metabolism by nuclear bile salt receptors in hepatocytes. The transport of bile salts is vectorial in the liver. Bile salts are taken up by the NTCP and OATP1B1/OATP1B3 in the sinusoidal membrane of hepatocytes and are excreted into bile mainly via BSEP in the canalicular membrane. Some bile salt metabolites are secreted into bile via the conjugate export pump (MRP2). FXR acts as a bile salt sensor and counteracts intracellular bile salt accumulation by inhibiting bile salt uptake and synthesis via upregulation of small heterodimer partner-1 (SHP-1) and transcriptional repression of CYP7A1 and NTCP genes. FXR also stimulates bile salt secretion by upregulating expression of apical BSEP and MRP2, and basolateral OST α/β . PXR represses bile salt synthesis by downregulation of transcription of CYP7A1 and stimulates bile salt detoxification by inducing expression of biotransformation phase 1 (e.g., CYP3A4) and phase 2 (e.g., SULT2A1) enzymes. PXR also stimulates the expression of the conjugate export pump, MRP2. 7 α -OH-C, 7 α -hydroxycholesterol ($\rightarrow$ activation; $\dashv$ repression).

Chenodeoxycholate (CDC) and cholate (C) are the endogenous ligands of FXR. In the liver, bile salt activation of FXR represses CYP7A1 transcription by an indirect mechanism involving nuclear receptors small heterodimer partner (SHP) and liver receptor homolog 1 (LRH1), or by a SHP-independent mechanism via FGF-19 and its receptor FGFR4. Activation of PXR, a promiscuous bile salt receptor, also results in repression of CYP7A1. The liver X receptor α (LXR α), abundantly expressed in the liver and activated by oxysterol ligands, has been shown to mediate the feedforward cholesterol catabolism to bile salts by upregulation of CYP7A1.

Bile salts can activate nuclear hormone receptor-independent pathways, such as protein kinase C (PKC) and mitogen-activated protein kinase (MAPK) signaling pathways. Activation of the G-protein-coupled bile salt receptor TGR5 induces an increase in intracellular cyclic adenosine monophosphate (cAMP), which may be an important mechanism in the regulation of energy homeostasis by bile salts and was shown to have an immunomodulatory effect on Kupffer cells.

Chemical Properties of Bile Salts

The end products of the complex biosynthetic steps from cholesterol to bile salts in humans are the two major primary bile salts, C (3 α ,7 α ,12 α -trihydroxy-5 β -cholanoate) and CDC (3 α ,7 α -dihydroxy-5 β -cholanoate). A final modification is the

conjugation of the terminal side-chain carboxylic acid with glycine or taurine by bile acid coenzyme A:amino acid N-acyltransferase (BAAT).

These (conjugated) primary bile salts undergo further modifications during enterohepatic cycling by bacterial enzymes in the distal intestine. Dehydroxylation at C-7 of C or CDC forms the secondary bile salts, deoxycholate (DC; 3 α ,12 α -dihydroxy-5 β -cholanoate) and lithocholate (LC; 3 α -hydroxy-5 β -cholanoate). Epimerization of the hydroxy group at the same C-7 position of CDC via the intermediate 7-ketolithocholate forms the tertiary bile salt ursodeoxycholate (UDC; 3 α ,7 β -dihydroxy-5 β -cholanoate).

The chemical properties of the different bile salts, such as their hydrophobicity, are determined by their molecular structure and conjugation. UDC is the most hydrophilic bile salt, followed by C, CDC, DC, and LC. Taurine-conjugated bile salts are more hydrophilic than glycine-conjugated ones, which are, in turn, more hydrophilic than free species.

The Bile Salt Pool

Bile is a mixture of 3–10 different species of bile salts, inorganic compounds such as ions and bicarbonate, and organic molecules such as bilirubin, phospholipids, cholesterol, adenosine nucleotides, albumin, and immunoglobulins.

Bile salts secreted into bile by hepatocytes originate from both their synthesis pathways as well as from their enterohepatic circulation. More than 95% of intestinal conjugated bile salts are actively reabsorbed, predominantly from the distal ileum by an apical sodium-dependent bile salt transporter (ASBT). After basolateral release into the mesenteric blood via the enterocyte organic solute transporter (OST), the OST α /OST β heterodimer, the bile salts, in part bound to albumin, reach the liver via the portal vein, to be reabsorbed and resecreted by the hepatocyte. The normal bile salt pool (about 3–4 g in humans) recirculates 5–10 times per day. As much as 5–10% of circulating bile salts are lost daily via the feces and are replaced by newly synthesized analogs.

Hepatic Transport of Bile Salts

The hepatocyte plays a key role in the vectorial transport of bile salts from the portal vein to the bile canaliculus, with distinct bile salt transport systems at its basolateral (sinusoidal) and apical (canalicular) plasma membranes (Figure 2). Functional impairment of bile salt transport at any level of the hepatocyte may be an important cause of cholestasis, a syndrome characterized by a reduction in bile flow and retention of biliary constituents in serum, liver, and other organs leading to biochemical, morphological, and clinical alterations.

Physiological Hepatocellular Bile Salt Transport

Bile salt uptake into the hepatocyte is predominantly mediated by the Na⁺-taurocholate-cotransporting polypeptide (NTCP), and, to a lesser extent, by the organic anion transporting protein (OATP) family. Transporters in the basolateral membrane, such as multidrug resistance-associated protein 4 (MRP4), can transfer conjugated bile salts back to the sinusoidal blood.

The exact nature of intrahepatocellular bile salt trafficking remains elusive. Bile salt-binding proteins are most likely involved in the rapid bile salt movement across the cell, and vesicular transport may also play a role.

Under physiological conditions, bile salt levels in the hepatocyte remain low. Efficient secretion of bile salts at the canalicular membrane is mainly achieved by the adenosine triphosphate (ATP)-dependent bile salt export pump (BSEP), which transports monovalent bile salts against a steep gradient (up to 1000-fold) across the apical membrane into the canaliculus. Mutations in this member of the ATP-binding cassette (ABC) transporter family may lead to progressive familial intrahepatic cholestasis type 2 (PFIC 2) in childhood, a severe disorder with markedly elevated serum bile salt levels and low content of bile salts in the bile. The conjugate export pump (MRP2) is another ABC transporter capable of transporting divalent sulfated and glucuronidated bile salts.

Active transport of bile salts across the canalicular membrane represents the rate-limiting step in the overall transport from blood to bile. It is under both short-term and long-term regulation by posttranscriptional and transcriptional mechanisms, respectively. On the transcriptional level, hepatic transporters have been shown to be regulated by nuclear receptors, such as FXR, which enhances BSEP expression in the

canalicular membrane of the hepatocyte. Bile salt-induced modulation of intracellular signaling cascades, including cytosolic free calcium [Ca²⁺]_i, isoforms of PKC and protein kinase A (PKA), mitogen-activated protein (MAP) kinase, and phosphatidylinositol 3-kinase (PI3K), contributes to bile salt-specific short-term regulation of bile secretion by the hepatocyte.

Cholestasis

In experimental models of cholestasis as well as in human cholestatic disorders, most of the hepatic uptake transporters are downregulated. Biliary secretion is further impaired by the retrieval of canalicular transporters from the plasma membrane and their storage in pericanalicular vesicles. Adaptive induction of basolateral bile salt exporter expression such as MRP3, MRP4, and OST α / β may protect the hepatocyte against rising intracellular concentrations of bile salts and other organic anions.

Bile Salts in Therapy

UDC is the major bile salt in black bear's bile, and has been used in Chinese traditional medicine in the form of dried black bear's bile for the treatment of icteric liver diseases for over 1000 years. UDC is normally present in human bile at a low concentration of 1–3% of total bile salts, but can comprise 40–50% of biliary bile salts in cholestatic patients treated with 13–15 mg per kg UDC per day. Its mechanisms of action are in part unraveled and include detoxification of bile, and stimulation of (impaired) hepatocellular and cholangiocellular secretion mainly by posttranscriptional mechanisms in hepatocytes and cholangiocytes, as well as antiapoptotic effects.

UDC is the established first-line treatment of primary biliary cirrhosis (PBC) and intrahepatic cholestasis of pregnancy (ICP). In PBC, UDC may normalize life expectancy in two-thirds of patients, may prolong transplant-free survival, delay histologic progression to fibrosis and cirrhosis, lower rates of complications of late-stage liver disease, and markedly improve serum liver tests. In ICP, UDC improves pruritus and serum liver tests in mothers and may prolong time to delivery and alleviate signs of fetal stress during pregnancy.

UDC is also widely used in other cholestatic disorders including primary sclerosing cholangitis (PSC), cystic fibrosis-associated liver disease (CFALD), or prolonged drug-induced cholestatic liver disorders due to its anticholestatic properties, although its long-term efficacy is not proven yet for these cholestatic conditions.

Cholate is effectively used for bile salt supplementation therapy in children with rare inborn errors of bile salt synthesis.

Increasing insight into the role of bile salts as important regulators of metabolic homeostasis, digestion, and intestinal defense has led to a steep increase in research efforts in the field of bile salt therapy. Several promising novel bile salt-derived compounds are at present under evaluation for therapeutic use in patients with cholestatic disorders, nonalcoholic fatty liver disease, as well as disorders of lipid metabolism.

See also: **Metabolism Vitamins and Hormones:** Cholesterol Synthesis and Regulation; Primary Biliary Cirrhosis; **Signaling:** Vitamin D Receptor.

Further Reading

Beuers U (2006) Drug insight: Mechanisms and sites of action of ursodeoxycholic acid in cholestasis. *Nature Clinical Practice Gastroenterology and Hepatology* 3: 318–328.

Boyer JL (2009) It's all about bile. *Hepatology* 49: 711–723.

Dawson PA, Lan T, and Rao A (2009) Bile acid transporters. *Journal of Lipid Research* 50: 2340–2357.

EASL (2009) EASL clinical practice guidelines: Management of cholestatic liver diseases. *Journal of Hepatology* 51: 237–267.

Hofmann AF (2009) Bile acids: Trying to understand their chemistry and biology with the hope of helping patients. *Hepatology* 49: 1403–1418.

Paumgartner G and Puhl T (2008) Medical treatment of cholestatic liver disease. *Clinics in Liver Disease* 12: 53–80.

Zollner G and Trauner M (2009) Nuclear receptors as therapeutic targets in cholestatic liver diseases. *British Journal of Pharmacology* 156: 7–27.

Biochemistry: Niacin/NAD(P)

K L Bogan and C Brenner, The University of Iowa, Iowa City, IA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Nicotinoproteins NAD^+ - or NADP^+ -dependent oxidoreductases that bind the coenzymes tightly as prosthetic groups.

Oxidoreductase A coenzyme-dependent hydride transfer enzyme.

Pellagra Niacin-deficient nutritional condition.

Poly(ADPribose) polymerase NAD^+ -dependent enzyme that forms protein linked or unlinked chains of ADPribose.

Sirtuin NAD^+ -dependent protein lysine deacetylase related to yeast Sir2.

Nicotinamide Adenine Dinucleotide History and Structure

Nicotinamide adenine dinucleotide (NAD^+) was at the center of some of the greatest discoveries in the biological sciences in the early twentieth century. Originally termed cozymase or codehydrogenase I and later diphosphopyridine nucleotide (DPN), the activity was described in 1905 as a component of yeast extracts that accelerated cell-free alcoholic fermentation. Arthur Harden and William Young discovered that glycolysis proceeded slowly until a heat-stable and dialyzable cozymase fraction was added, which we now know contained NAD^+ , adenosine triphosphate (ATP), and Mg^{2+} . The NAD^+ component was later purified by Harden and Hans von Euler. In 1936, the structure of NAD^+ was determined independently by Otto Warburg and von Euler, and a role in oxidative metabolism was defined. Codehydrogenase II was found to be a triphosphopyridine dinucleotide, that is, nicotinamide adenine dinucleotide phosphate (NADP^+). Two salvageable precursors of NAD^+ , nicotinic acid (NA) and nicotinamide (Nam), were identified by Conrad Elvehjem in a search for non protein fractions from the liver that would reverse black spots on the tongues of dogs, an induced nutritional deficiency similar to pellagra, which was an epidemic in the American South in the early 1900s. Later, the basis for both *de novo* synthesis of NAD^+ from amino acids and salvage synthesis of NAD^+ was worked out. NA and Nam were collectively termed 'niacins'. The broad use of niacin supplementation has virtually eliminated pellagra.

NAD^+ is so termed because it consists of a Nam nucleotide, that is, nicotinamide riboside 5'-monophosphate, joined to the phosphate of an adenine nucleotide, that is., adenosine 5'-monophosphate. NADP^+ is formed when a phosphate group is added to the 2' position of adenosine (Figure 1). The glycosidic linkages between bases and ribosyl moieties are β in both nucleotides. The hydride-accepting moiety is the Nam base such that the plus signs on NAD^+ and NADP^+ refer to the oxidized, hydride-accepting forms. The overall charge of these molecules is negative because of the phosphates.

Salvage and De Novo Pathways

All dividing cells either form NAD^+ *de novo* from an amino acid and/or resynthesize NAD^+ from NA, Nam, and/or

nicotinamide riboside (NR). These water-soluble vitamins are NAD^+ breakdown products and metabolites that are transported systemically; these are available in the diet (Figure 2(a)).

NA is utilized by the Preiss-Handler pathway, which involves three enzymatic steps through nicotinic acid mononucleotide (NaMN) and nicotinic acid adenine dinucleotide (NaAD). In vertebrates, Nam is utilized in two enzymatic steps through nicotinamide mononucleotide (NMN). Since Nam can be converted into NA in many bacteria by a nicotinamidase not encoded in vertebrate genomes, Nam entry into the Preiss-Handler pathway in vertebrates is thought to depend on bacterial nicotinamidase in the gut. NR can be converted into two steps to NAD^+ through NMN, or can be converted into NAD^+ via splitting the nucleoside followed by Nam salvage. Specific transporters have been identified for the uptake of NA and NR, but not Nam or nicotinic acid riboside (NAR), an NAD^+ metabolite that can also be utilized by yeast at high doses.

Most organisms synthesize NAD^+ from either tryptophan or aspartic acid. *De novo* synthesis nearly always proceeds through quinolinic acid (QA) and NaMN, at which point NAD^+ is produced by the last two enzymes of the Preiss-Handler pathway (Figure 3).

The phosphorylated forms of NAD^+ are also generated by specific enzyme activities (Figure 2(b)). NADP^+ is generated from NAD^+ by NAD^+ kinase (NADK). In yeast, a mitochondrial NADH kinase generates NADPH from NADH. In addition, in bacteria and in animal mitochondria, the proton translocating, membrane-bound NADPH transhydrogenase interconverts NADPH and NAD^+ into NADP^+ and NADH. Thus, NADK activity is a gatekeeper of both phosphorylated forms of NAD^+ , and NADPH transhydrogenase allows a cell to maintain a redox balance in which NAD^+ levels exceed NADH levels, while NADPH levels exceed those of NADP^+ .

NAD^+ and NADP^+ in Redox Metabolism

Electron transfers that occur in oxidation-reduction (redox) reactions are essential to metabolism and life in all organisms. Redox reactions are catalyzed by a large family of enzymes called 'oxidoreductases'. The electron-donating species in a redox reaction is referred to as the reducing agent, or reductant, and the electron-accepting species is the oxidizing agent, or

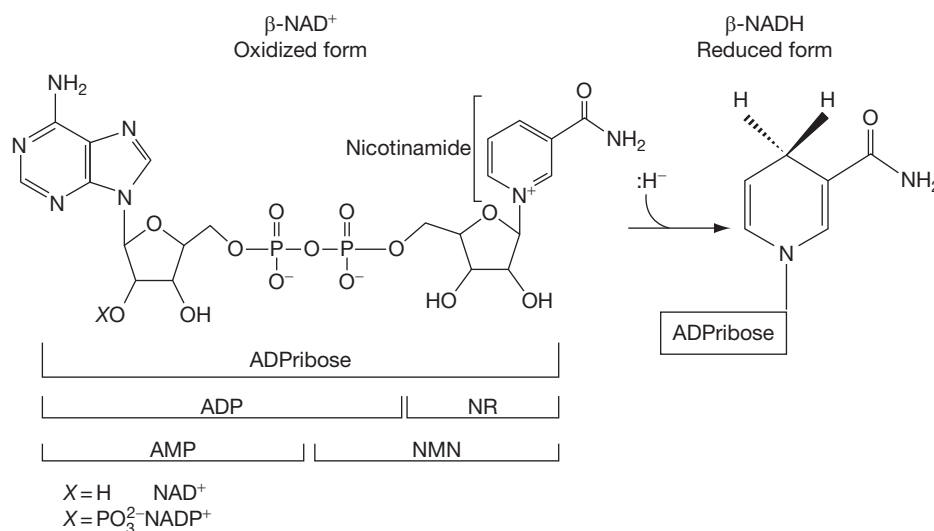


Figure 1 NAD^+/NADH and $\text{NADP}^+/\text{NADPH}$ structure. NAD^+ is a dinucleotide consisting of a nicotinamide nucleotide, that is, nicotinamide riboside 5'-monophosphate (NMN) joined to the phosphate of an adenine nucleotide, that is, adenosine 5'-monophosphate (AMP). In NAD^+ , the X in the 2' position of the adenosine is an H, whereas in NADP^+ , a phosphate group is in the X position. The glycosidic linkages between bases and ribosyl moieties are β in both nucleotides. The Nam base is the hydride-accepting moiety, such that the plus signs on NAD^+ and NADP^+ refer to the oxidized, hydride-accepting forms.

oxidant. In biological redox reactions, the movement of electrons often occurs concomitant with the loss of hydrogen. This species is referred to as a hydride ion (:H^-).

In NAD^+ -dependent dehydrogenation reactions, NAD^+ and NADP^+ undergo reduction of the nicotinamide ring to form NADH and NADPH , respectively (Figure 1). In such reactions, a hydride ion is transferred to the carbon at the fourth position of the nicotinamide ring of NAD^+ (Figure 1). This transfer of hydride can happen in two ways. In A-type oxidoreductases, the hydrogen is transferred from above the plane of the nicotinamide ring to the front of the nicotinamide ring (A side), while B-type oxidoreductases transfer of the atom occurs from below to the backside of the nicotinamide ring (B side). This depends on how NAD^+ is oriented within the nucleotide-binding domain of the enzyme.

Oxidoreductases bind the NAD(P)^+ dinucleotide in a structural motif termed as a Rossmann fold (Figure 4(a)). The core Rossmann fold consists of one nucleotide-binding domain constructed from at least three β -strands flanked by two α -helices ($\beta\alpha\beta\alpha$), though some oxidoreductases have additional β -strands. There are four important motifs found within the Rossmann fold: a conserved positively charged residue (Arg or Lys) at the beginning of the first β -strand, a conserved negatively charged residue (Glu or Asp) at the end of the second β -strand, a phosphate-binding sequence –GXGXXG– and six conserved positions occupied by small hydrophobic amino acids.

Three of four conserved motifs within the Rossmann fold have well-defined nucleotide-binding, structural, or catalytic functions. The conserved positively charged residue (Arg or Lys) found at the beginning of the first β -strand is not well understood. It is thought to participate in stabilizing interactions with nearby β -strands, though these are not fully characterized. The conserved negatively charged residue at the carboxyl terminus of the second β -strand is used to bind the

2'-hydroxyl group of the adenine nucleotide. This site also helps to distinguish enzymes that bind NAD^+ from those that bind NADP^+ , as this residue is uncharged in enzymes that bind NADP^+ in order to accommodate the 2'-phosphate that takes the place of the hydroxyl. In NADP^+ -binding enzymes, the phosphate interacts with a nearby arginine residue instead. The phosphate-binding glycine-rich sequence (GXGXXG) resides in a loop between the first α -helix and β -strand. The first and second glycines are thought to allow important turns of the main chain in this loop. These turns promote an interaction between the main chain and the diphosphate bridge of NAD^+ to occur (Figure 4(b)). The third glycine promotes tight packing in the nucleotide-binding domain that prohibits NADP^+ from binding. NADP^+ -binding enzymes have a larger amino acid (alanine, serine, or proline) in the place of the glycine, which disrupts the close packing and allows NADP^+ to bind. Finally, the hydrophobic core, which contains six small hydrophobic amino acids, is necessary for maintaining the proper packing of the β -strands with respect to the α -helices. While these motifs form the proper tertiary structure and interactions to specify binding of NAD^+ versus NADP^+ , some enzymes, such as aldose reductase, glucose-6-phosphate dehydrogenase, and methylenetetrahydrofolate reductase, are exceptions to the rule, and bind both NAD^+ and NADP^+ .

For the majority of enzymes that do show a high level of specificity for either NAD^+ or NADP^+ , the preference often reflects the distinct metabolic role of the enzyme. NAD^+ is typically utilized in catabolic pathways in which energy is liberated from the oxidation of substrate molecules, whereas NADP^+ is most commonly utilized in anabolic reactions such as fatty acid synthesis and photosynthesis. The ratios of oxidized to reduced forms of NAD^+ and NADP^+ reflect these activities as well. In most tissues, in which it has been measured, the ratio of NAD^+ to NADH is high, providing ample NAD^+ to

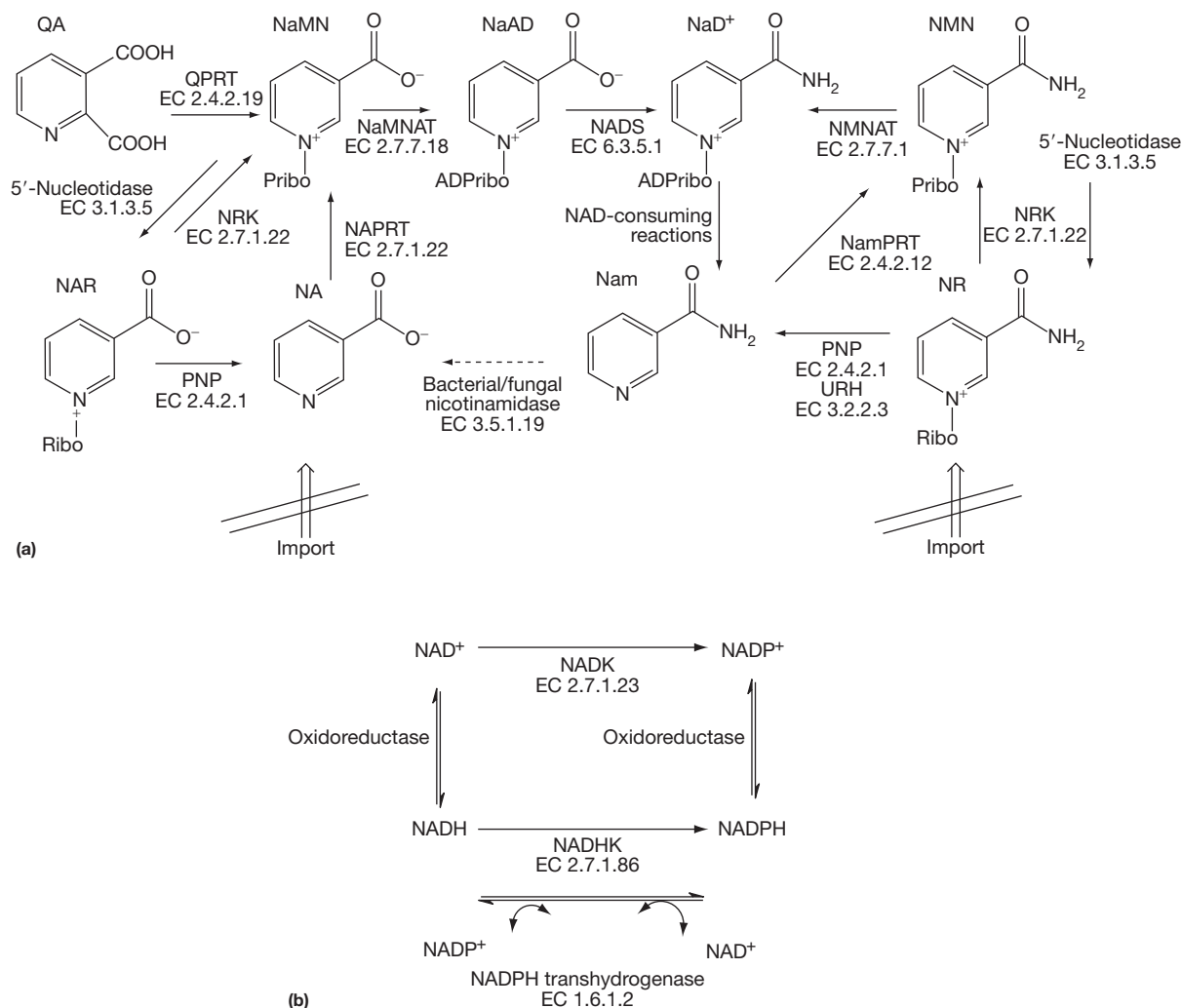


Figure 2 NAD⁺ and NADP⁺ generation. (a) Cells have evolved pathways to salvage the pyridine ring of NAD⁺ from precursors NA, Nam, and NR, each of which is produced in the cell as metabolites and obtained in the diet as vitamins. Nicotinic acid (NA) is utilized by what is termed as the Preiss–Handler pathway, which involves three enzymatic steps. First, NA is converted into nicotinic acid mononucleotide (NaMN) by the nicotinic acid phosphoribosyltransferase enzyme (NAPRT). NaMN is subsequently converted into nicotinic acid adenine dinucleotide (NaAD), also called desamido NAD⁺, by nicotinamide mononucleotide adenyl transferase enzymes (NMNAT). Finally, NaAD is converted into NAD⁺ by the NAD⁺ synthetase enzyme (NADS). In mammals, salvaged Nam is utilized by a nicotinamide phosphoribosyltransferase enzyme (NamPRT). NamPRT catalyzes the addition of a phosphoribose moiety onto Nam forming nicotinamide mononucleotide (NMN). NMN is then converted into NAD⁺ by the NMNAT enzymes. Nam utilization differs among species. In some bacteria and fungi, Nam is converted into NA by a nicotinamidase enzyme (dotted line), after which it is utilized through the Preiss–Handler pathway to form NAD⁺. NR can be utilized through two distinct pathways. In an Nrk-dependent pathway, it is phosphorylated by nicotinamide riboside kinases (NRKs) to form NMN, which is converted into NAD⁺ by NMNAT enzymes. Alternatively, NR is utilized by purine nucleoside phosphorylase (PNP), which converts NR into Nam. Specific transporters have been identified that are responsible for the uptake of NA and NR. Another NAD⁺ precursor, that is not efficiently imported, but that can be produced intracellularly, is nicotinic acid riboside (NAR). NAR is phosphorylated by NRK to form NaMN, and is hydrolyzed to NA by PNP. Data in yeast *Saccharomyces cerevisiae* indicate that NR and NAR are generated by the activities of NMN/NaMN 5′-nucleotidases. (b) NADP⁺ and NADPH are generated from NAD⁺ and NADH by kinases (NADK and NADHK). NADPH transhydrogenase interconverts NADPH and NAD⁺ into NADP⁺ and NADH.

accept hydride ions from substrates, whereas the level of NADP⁺ is often significantly lower than NADPH promoting hydride transfer to substrate molecules from NADPH, as occurs in anabolic pathways. Likewise, the organelles in which these cofactors are primarily used are also distinct and reflective of their different activities. NAD⁺ is most often utilized in the mitochondria for oxidation of fuel molecules (pyruvate, fatty

acids, and α -keto acids), while NADPH serves as a reductant for fatty acid synthesis in the cytosol.

Hydride transfer from substrate to NAD⁺ has three prerequisites. First, the substrate must have high electron density on the carbon that donates the hydride. This is often ensured by interactions of the substrate with active site residues in redox enzymes. Similarly, the C4 carbon of the Nam ring of NAD⁺,

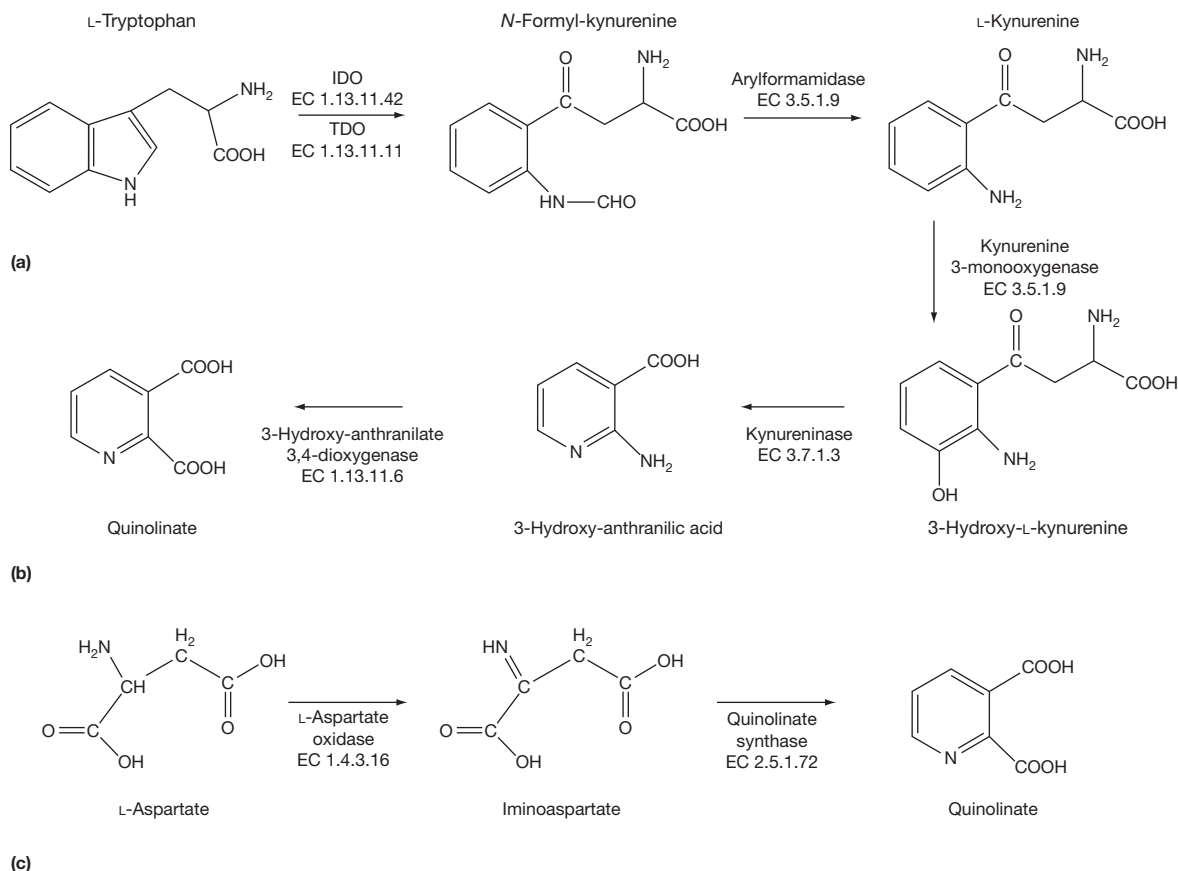


Figure 3 *De novo* biosynthetic pathways. (a) Present in most eukaryotic systems, the *de novo* pathway from tryptophan synthesizes quinolinic acid (QA) in five enzymatic steps. In the first step, tryptophan is converted into *N*-formyl-kynurenine by either indoleamine 2,3 deoxygenase (IDO) or tryptophan 2,3-dioxygenase (TDO). Formylkynurenine is subsequently converted into L-kynurenine by kynurenine formamidase (KFase). Kynurenine-3-hydrolase (K3H) produces 3-hydroxykynurenine from L-kynurenine, which is then a substrate of kynureninase (Kyase) to produce 3-hydroxyanthranilic acid. 3-Hydroxyanthranilate 3,4-dioxygenase (3HAO) then converts 3-hydroxyanthranilic acid into 2-amino-3-carboxy-muconate semialdehyde, which is converted into quinolinic acid by spontaneous cyclization. (b) Present in most prokaryotic systems, the *de novo* pathway from aspartic acid synthesizes QA in two enzymatic steps. Aspartate is converted into iminoaspartate by L-aspartate oxidase (AO), which is subsequently converted into QA by quinolinate synthase (QS).

which accepts the hydride ion, must have sufficiently low electron density to accept the hydride ion. Finally, the Nam moiety of NAD^+ must be positioned close enough to the hydride-donating species on the substrate to accept it. This is accomplished by the specific binding of NAD^+ in the Rossmann fold. For example, the active site of lactate dehydrogenase (LDH), an A-type oxidoreductase, utilizes an active site residue, histidine-195, as a general base. In catalysis, the histidine residue removes a proton from the hydroxyl group of the alcohol substrate promoting the hydride transfer from the hydroxyl group to the C4 position of the Nam moiety (Figure 5). Importantly, these mechanisms both avoid actual formation of the hydride ion, which is highly reactive, and cannot exist in water.

NAD^+ as a Prosthetic Group for Enzymes

In contrast to the reversible binding behavior of NAD^+ -coenzymes in redox reactions, these cofactors can also function as tightly bound prosthetic groups. In such enzymes, NAD^+ and NADP^+ do not dissociate with the release of the product,

but stay tightly bound. Enzymes that catalyze reactions using a prosthetic group often have higher turnover rates than enzymes that allow the coenzyme to dissociate by virtue of eliminating slow association and dissociation steps. In addition, enzymes that bind and sequester a cofactor in the active redox state are rendered insensitive to regulation due to the redox state of the cell. This is due to the enzyme's ability to return the bound coenzyme to its original active redox state after catalysis so that it can perform subsequent reactions. Enzymes accomplish this in several ways. First, the enzyme can participate in alternating oxidation and reduction reactions with different substrates. Alternatively, the bound coenzyme can be reoxidized or reduced by a nonsubstrate molecule, such as ferredoxin. In other reactions, the coenzyme can transiently oxidize or reduce a substrate, but return to the original redox state prior to product release. Finally, in some reactions, the coenzyme can accept only a pair of electrons and not the entire hydride ion. In this way, the bound coenzyme can promote catalysis without full oxidation or reduction occurring, eliminating the regeneration requirement. Enzymes that bind NAD^+ and NADP^+ as prosthetic groups are termed

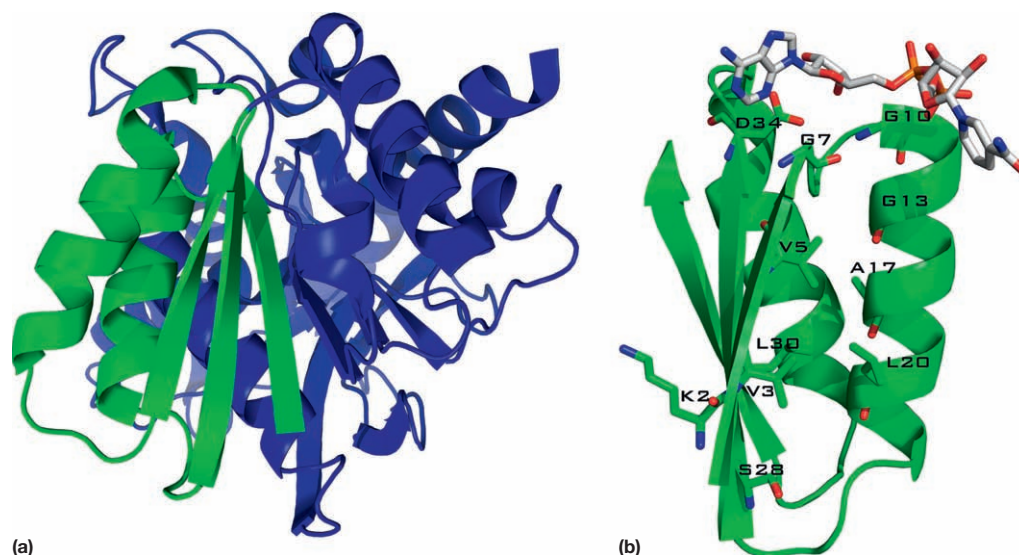


Figure 4 The structure of lactate dehydrogenase (LDH). (a) The structure of LDH (PDB entry 2JC9) highlighting the Rossmann fold in green. (b) A close-up of the Rossmann fold illustrating the critical residues and the interactions with NAD^+ .

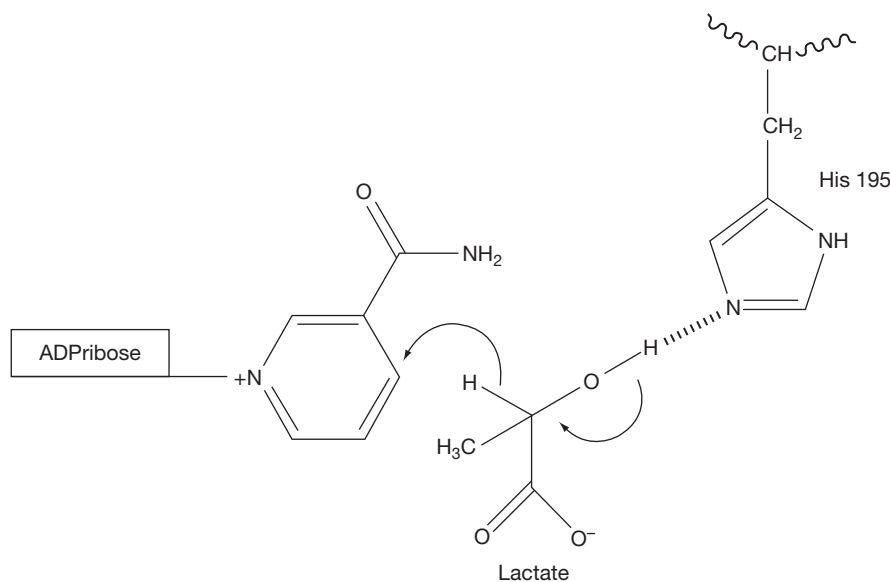


Figure 5 Dehydrogenation mechanism of LDH. His 195 acts as a general base to promote hydride transfer to the nicotinamide ring.

'nicotinoproteins', which belong to a superfamily of metallo alcohol dehydrogenases.

Nonredox Roles

In addition to its coenzymatic functions, NAD^+ is a consumed substrate of multiple families of enzymes (Figure 6). Generally termed as glycohydrolases, these enzymes cleave the glycosidic bond between the nicotinamide and adenosine diphosphate (ADP)ribose moieties of NAD^+ .

One class of NAD^+ -consuming enzymes possesses ADP-ribose transferase (ART) and poly(ADP-ribose) polymerase (PARP)

activities, which transfer and/or polymerize NAD^+ -derived ADP-ribose to target molecules. PARP activity, which is considered one of the major NAD^+ -consuming activities within the cell, is upregulated in response to genome damage. Individual members of the PARP family have been shown to mediate diverse cellular responses including inflammation, intracellular trafficking, and cell death.

A second class of NAD^+ -consuming enzymes consists of cyclic (ADP-ribosyl) synthases, which are largely extracellular enzymes that consume NAD^+ to produce and hydrolyze cyclic-ADP-ribose. These activities are particularly important for calcium (Ca^{2+}) regulation within the cell, as cyclic-ADP-ribose is a potent Ca^{2+} -mobilizing second messenger.

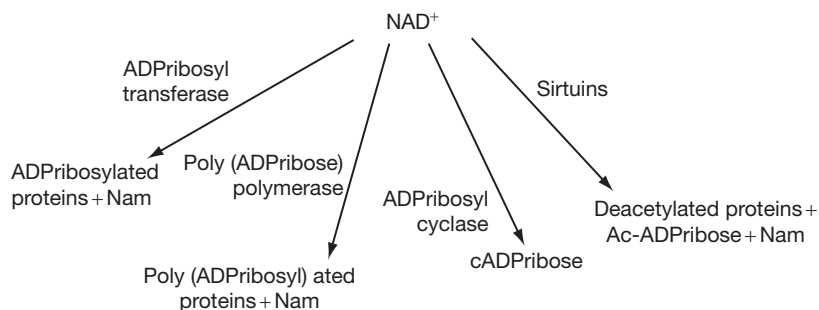


Figure 6 Nonredox roles of NAD^+ . Enzymes that cleave the glycosidic bond between the nicotinamide and ADP-ribose moieties of NAD^+ are termed glycohydrolases. These include ADP-ribose transferase (ART) and poly(ADP-ribose) polymerase (PARP) activities, which attach NAD^+ -derived ADP-ribose to target molecules, and cyclic ADP-ribose synthases, which consume NAD^+ to produce and hydrolyze cyclic-ADP-ribose. In addition, sirtuins catalyze NAD^+ -dependent protein lysine deacetylation by conversion of a lysine-acetylated peptide to a deacetylated peptide with production of 2'-O-acetyl-ADP-ribose and Nam as products.

A third class of NAD^+ -consuming enzymes is sirtuins, the NAD^+ -dependent protein lysine deacetylases related to yeast Sir2. These enzymes convert a lysine-acetylated peptide into a deacetylated peptide with the production of 2'-O-acetyl-ADP-ribose and Nam as products. This process has been implicated in gene regulation, enzyme activity, and longevity.

NAD^+ Alterations in Aging

Calorie restriction (CR) is a powerful means to extend life span in model organisms. In yeast, Sir2 plays a role in mediating the effect of CR on extension of life span. Since NAD^+ is involved in glucose utilization and is an essential Sir2 substrate, its metabolism is one potential subsystem that may be altered by CR. Among the proposed mechanisms, it is conceivable that the NAD^+/NADH ratio, the NAD^+/Nam ratio, altered flux, and/or increased net synthesis may regulate sirtuin activities.

PARP activities may also influence the aging process in an NAD^+ -dependent manner. In particular, the DNA-damage surveillance activity of PARP has been proposed as a critical factor to reduce genotoxic stress, maintain genomic stability, and promote cellular survival. PARP family enzymes, tankyrase-1 and tankyrase-2, are also involved in the maintenance of telomere length, which has been implicated in the regulation of aging. Roles in aging and telomere maintenance are underscored by the observation that PARP-1, tankyrase-1, and tankyrase-2 interact with the Werner syndrome protein, which is altered in progeric humans.

Human Health and Disease

Widespread use of niacins has largely eliminated pellagra, whereas high-dose NA is widely used for dyslipidemia and high-dose Nam has been tested for efficacy in stroke and other conditions. Since high-dose NA causes a flushing response and high-dose Nam may inhibit sirtuins and other NAD^+ -consuming enzymes, NR is being tested for effectiveness in the

therapy of neurodegenerative diseases, such as Alzheimer's and Parkinson's disease.

Many of the enzymes that use NAD^+ either as a co-enzyme or as a substrate are considered as potential targets for drug design. Although redox enzymes bind NAD^+ in a conserved binding pocket, enzyme-specific inhibitors based on the structure of NAD^+ have been developed. For example, the nucleoside analog tiazofurin, which is converted in the cell into a toxic analog of NAD^+ , tiazofurin adenine dinucleotide, is a potent prodrug inhibitor of IMP dehydrogenase (IMPDH), an enzyme important in *de novo* purine synthesis that might be targeted in malignancies, viral infection, and/or for the purposes of immunosuppression.

NAD^+ biosynthetic pathways have also been targeted, especially for antibiotic development. The differences in NAD^+ biosynthetic pathways and enzymes between humans and lower organisms make the development of specific inhibitors to block NAD^+ synthesis in infectious organisms a plausible path to antibiotics. So far, drugs of this nature have been used to target *Mycobacterium tuberculosis*, the tuberculosis-causing bacterium. Isonicotinylhydrazine is a Nam-related anti-tuberculosis agent that is metabolized to form inhibitors of redox enzymes. It is anticipated that additional agents will be developed against genetically validated enzymes, potentially using structure-based drug design and/or high-throughput screening.

See also: [Bioenergetics: Bioenergetics: General Definition of Principles; Calcium Signaling by Cyclic ADP-Ribose and NAADP; Nicotinamide Nucleotide Transhydrogenase.](#)

Further Reading

- Bellamacina CR (1996) The nicotinamide dinucleotide binding motif: A comparison of nucleotide binding proteins. *FASEB Journal* 10: 1257–1269.
- Bogan KL and Brenner C (2008) Nicotinic acid, nicotinamide and nicotinamide riboside: A molecular evaluation of NAD^+ precursor vitamins in human nutrition. *Annual Review of Nutrition* 28: 115–130.
- Bogan KL, Evans C, Belenky P, et al. (2009) Identification of Isn1 and Sdt1 as glucose- and vitamin-regulated nicotinamide mononucleotide and nicotinic acid adenine mononucleotide 5'-nucleotidases responsible for production of

- nicotinamide riboside and nicotinic acid riboside. *Journal of Biological Chemistry* 284: 34861–34869.
- Gazzaniga F, Stebbins R, Chang SZ, McPeck MA, and Brenner C (2009) Microbial NAD metabolism: Lessons from comparative genomics. *Microbiology and Molecular Biology Reviews* 73: 529–541.
- Longo VD (2009) Linking sirtuins, IGF-I signaling and starvation. *Experimental Gerontology* 44: 70–74.
- Magni G, Amici A, Emanuelli M, et al. (2004) Enzymology of NAD⁺ homeostasis in man. *Cellular and Molecular Life Sciences* 61: 19–34.
- Nelson DL and Cox MM (2008) Biological oxidation–reduction reactions. In: *Lehninger Principles of Biochemistry*, 5th edn., chapter 13, pp. 512–526. New York, NY: WH Freeman.
- Oppenheimer NJ (2010) NAD⁺ and NADP⁺ as prosthetic groups for enzymes. In: *Encyclopedia of Life Sciences*, doi:10.1002/9780470015902.a000637.pub2. Chichester: Wiley.
- Ziegler M (2001) New functions of a long-known molecule: Emerging roles of NAD in cellular signaling. *European Journal of Biochemistry* 267: 1550–1564.

Biochemistry of Development: Striated Muscle

R Bassel-Duby and E N Olson, University of Texas Southwestern Medical Center, Dallas, TX, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Cardiogenesis Development of the heart during embryonic development.

MicroRNA (miRNA) Single-stranded RNA molecules about 21–23 nucleotides in length that act as negative regulators of gene expression by interacting with specific mRNA targets and inhibiting translation or promoting mRNA degradation.

Myogenesis The formation of muscle fibers and skeletal muscle during embryonic development.

Sarcomere The basic contractile unit of the muscle cell that is arranged in distinct bands to give muscle its striated appearance.

Satellite cells A resident skeletal muscle stem cell population located in a quiescent state under the basal lamina of skeletal muscle fibers that upon muscle injury are activated to proliferate, differentiate, and repair damaged muscle.

Somites Spherical masses of primordial muscle tissue that occur at regular intervals along the length of the embryo that form skeletal muscle in the vertebrate limb and trunk.

Transcription factor A protein that binds to specific DNA sequences and activates or represses gene expression.

Introduction

Cardiac and skeletal muscles are defined as striated muscles because under the microscope they appear striped or striated in appearance due to the arrangements of the sarcomeres in distinct bands (Figure 1). The sarcomere is the basic contractile unit of the muscle cell and is comprised of bands termed A, M, I, and Z lines. The A band comprises the thick filaments of myosin and proteins that bind myosin. The M line is the middle part of the A band and the I band is composed of thin actin filaments and proteins that bind actin. The Z line, situated in the middle of the I band, contains the greatest number of sarcomeric proteins and functions as a scaffold that links the sarcomeric contractile units in series. It is the Z line that defines the lateral boundaries of the sarcomere. Although the sarcomere appears to be in an organized static state, it is well recognized that sarcomeric proteins have a dynamic distribution within the cell and can communicate signals between the sarcomere and the nucleus to participate in the response of the muscle cell to changes in the structure or function of the contractile machinery. Indeed, the importance of the cytoarchitecture of the sarcomere is underscored by the observation that mutations of sarcomeric proteins are associated with a large percentage of cardiac and skeletal muscle myopathies in humans.

This article provides an overview of the biochemistry of striated muscle development, highlighting the differences between cardiac and skeletal muscles. Emphasis is given to embryonic cells adopting specific cell fates and how these cells then differentiate to become cardiac or skeletal muscle. In particular, many of the transcription factors and microRNAs (miRNAs) that regulate the programs controlling cell differentiation and morphogenesis during development are described. Of note, the processes involved in muscle development are evolutionarily conserved across diverse organisms. This enabled many of the myogenic regulatory processes to be identified in fruit flies, zebrafish, chick, mouse, and humans, and validated by performing gain- and loss-of-function experiments

in vivo and *in vitro*. It is the integration and collaboration of developmental biology and molecular biology research studies that have systematically defined the factors involved in determining the regulatory processes necessary for striated muscle development.

Development of Skeletal Muscle

Myogenesis

Although both cardiac and skeletal muscles are striated muscles, they originate and develop from distinct cell populations. Skeletal muscle is derived from somites of the embryo, which are the spherical masses of primordial muscle tissue that occur at regular intervals along the length of the embryo (Figure 2). At the onset of myogenesis in the somite, progenitor cells delaminate from the dermomyotome to locate to the underlying myotome. The dermomyotome gives rise to all skeletal muscle of the trunk and limbs. As myogenesis proceeds, primitive cells become myoblasts, which are cells committed to become skeletal muscle. Myoblasts proliferate, migrate, and aggregate into clusters, and fuse to become multi-nucleated myotubes. Further lengthening occurs by fusion of more myoblasts with the ends of the myotubes. Primary myotubes, the first formed myotubes, separate into clusters and are surrounded by the basal lamina. More myoblasts aggregate beneath the basal lamina and fuse to form secondary myotubes. With further development, the primary myotube begins to look like mature skeletal muscle with the nuclei being forced from the center to the periphery as the myofiber is filled with contractile proteins. Secondary myotubes also begin to mature into fibers so that the primary and secondary myotubes are contained within a single basement membrane. As myogenesis proceeds, skeletal muscle fibers develop into slow twitch or fast twitch, based on the type of nerve innervation and spatiotemporal expression of regulatory factors. It is the heterogeneity of muscle fibers within the muscle bed that provides the differences in the physiological and metabolic parameters of the muscles, enabling different muscle groups to

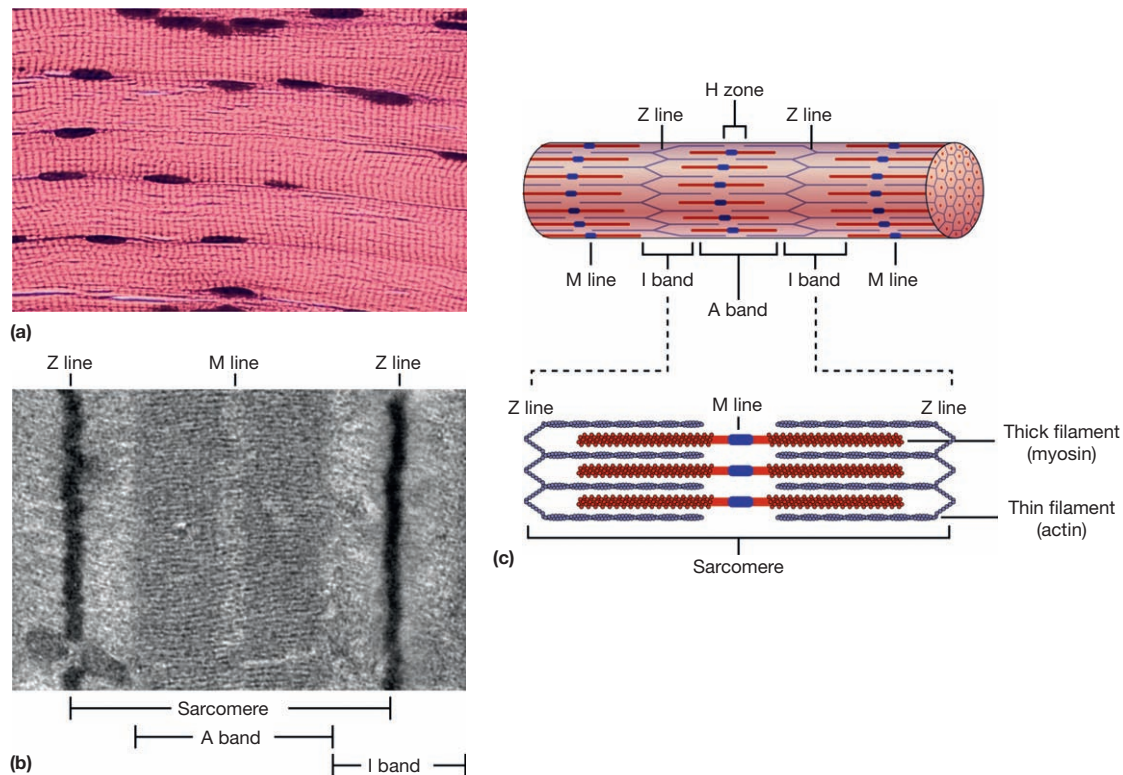


Figure 1 Striated muscle. (a) Paraffin sections of skeletal muscle stained with hematoxylin and eosin showing longitudinal muscle fibers depicting cross-striations appearance due to the arrangement of the sarcomeres in distinct bands. (b) Electron microscopy image of a sarcomere. (c) Illustration of striated muscle. (a) Image by Melanie Weiler. (b) Image by Dr. Christopher Gilpin. (c) Image by José Cabrera.

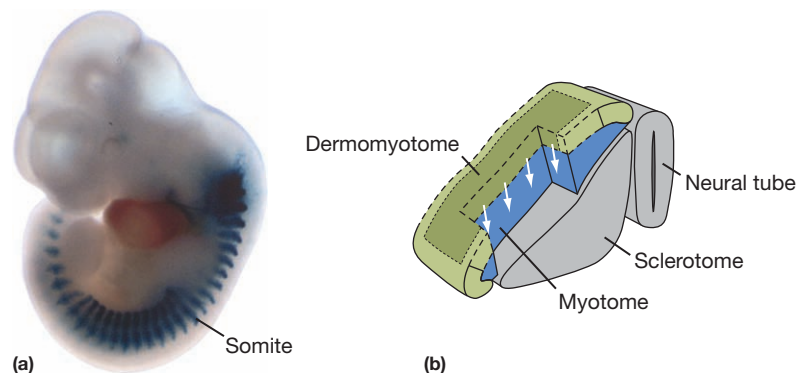


Figure 2 Development of skeletal muscle. (a) Mouse embryo with somites highlighted in blue by β -galactosidase staining. (b) Illustration of the sections of a somite. Reproduced from Figure 1 in Buckingham M and Vincent SD (2009) Distinct and dynamic myogenic populations in the vertebrate embryo. *Current Opinion in Genetics and Development* 19(5): 444–453, with permission from Elsevier.

provide a variety of functional properties in response to environmental demands.

Skeletal muscle throughout the body is not derived from the same cell population. The trunk and limb muscles are derived from the somite but the head and neck muscles (termed the branchiomeric muscles) are derived from the mesodermal core of the branchial arches and from the prechordial mesoderm. Of interest, recent studies using molecular biology tools have shown a close relationship between the origin of head skeletal muscle and heart muscle. The term ‘cardio-craniofacial

morphogenetic field’ reflects the intimate developmental relationship between the head, face, and heart.

The Role of Transcription Factors in Skeletal Muscle Development

Skeletal muscle development is controlled by a network of evolutionarily conserved regulatory factors that specify myogenic progenitor cells and direct them to muscle differentiation. Some of the regulatory factors that are critical for the

determination and terminal differentiation are restricted to skeletal muscle, such as the four myogenic regulatory factors (MRFs): MyoD, Myf5, myogenin, and MRF4 (Figure 3). Other skeletal muscle regulatory factors, such as myocyte enhancer factor 2 (MEF2), are more broadly expressed in a variety of tissues, such as cardiac and smooth muscle, T cells, and brain. It is the cooperative interactions of the MRFs with MEF2 proteins in the context of various signaling pathways that regulate myogenesis of skeletal muscle.

MRFs were discovered over two decades ago and have the unique distinction of converting a variety of cell types to myoblasts. MRFs are DNA-binding proteins that contain a basic helix–loop–helix (bHLH) domain and bind to an E-box consensus sequence (CANNTG) which is present in the promoters and enhancers of many muscle-specific genes. *In vitro*, the MRFs appear indistinguishable; however, *in vivo*, primarily due to different spatiotemporal expression patterns, they have distinct roles. During embryogenesis Myf5 is expressed in the dorsal–medial somites, which give rise to the trunk and intercostal muscles, followed by MyoD expression in the dorsal–lateral somites, which give rise to the body wall and limb muscles. Myogenin plays a critical role in terminal differentiation of myoblasts and MRF4 has a complex temporal expression pattern during embryogenesis, suggesting roles in both muscle determination and terminal differentiation. It is of interest that although *in vitro* MRFs can substitute for one another, *in vivo*, genetic engineering studies showed that *myogenin* inserted into the *myf5* locus was not able to completely

replace the function of Myf5 in muscle development, suggesting that each gene has a unique function *in vivo*.

MEF2 factors alone do not possess myogenic activity but, in combination with the MRF transcription factors, drive and amplify the myogenic differentiation program. There are four MEF2 genes (*Mef2a*, *b*, *c*, and *d*) in vertebrates and they bind to the consensus DNA sequence YTA(A/T)₄TAR which is present in many muscle-specific gene promoter and enhancer regions. In addition to regulating numerous muscle structural genes, MEF2 proteins regulate the expression of the MRF genes, such as myogenin, thereby providing a positive feed-forward loop that perpetuates and amplifies the decision to differentiate. Further, MEF2 interacts with additional transcription factors that are required for proper muscle development. Similar to MRFs, MEF2 is expressed in the somites. However, unlike the restrictive expression pattern of the MRF proteins in skeletal muscle, MEF2 proteins are expressed in many cell types, including neurons, chondrocytes, and cardiac, skeletal, and smooth muscles. Of interest, the expression of MEF2 in all these various cell types occurs concomitantly with the activation of the differentiation programs in different cell types. The activity of MEF2 is tightly governed by class IIa histone deacetylases (HDACs), which directly bind and promote the formation of multiprotein repressive complexes on MEF2-dependent genes. It is the balance between the transcription-activating functions of MEF2 and the repressive functions of class IIa HDACs that dictates tissue development.

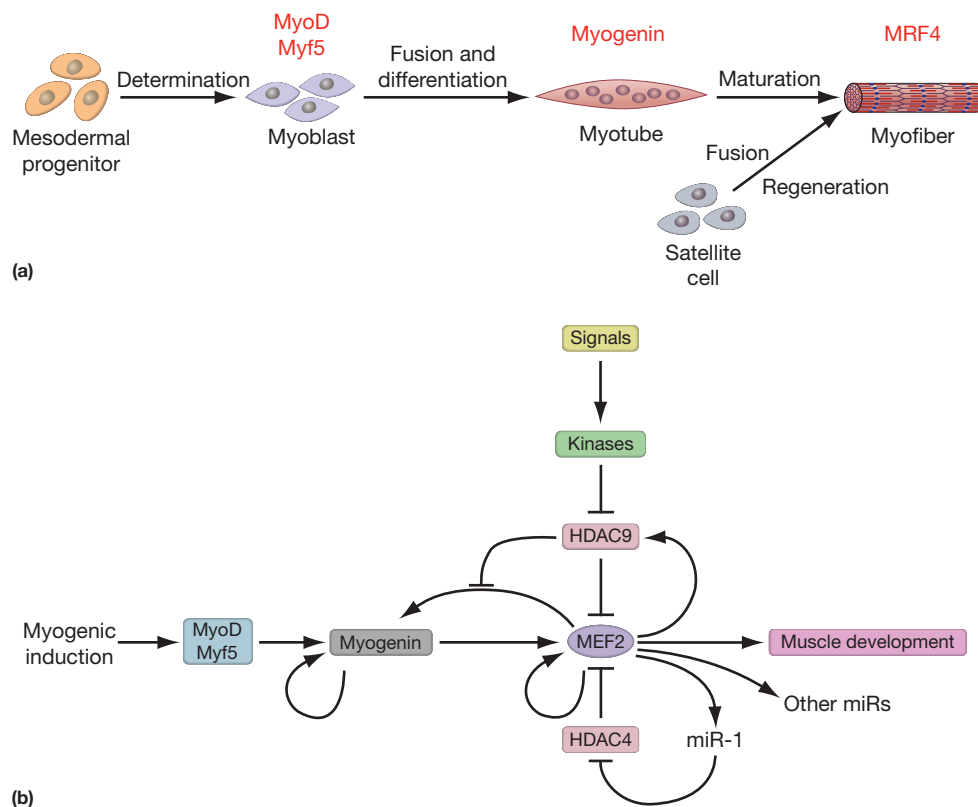


Figure 3 Pathways during skeletal muscle development. (a) Stages of myogenesis. (b) Regulatory pathways during myogenesis. Image by José Cabrera.

Pax3 and Pax7 are transcription factors that belong to the conserved Pax gene family and consist of a paired domain that confers sequence-specific binding to DNA, an octapeptide and a homeodomain. Pax3 and Pax7 are expressed in distinct areas of the somite and play a key role during myogenesis in regulating myogenic cell fate. Pax3 is transcribed throughout the somite before becoming restricted to the dermomyotome and Pax7 is expressed in the central domain of the dermomyotome. During myogenesis, Pax3 regulates the MRFs by directly activating the *Myf5* gene, and eventually forming the trunk and limb skeletal muscles. Pax3 and Pax7 not only regulate the entry of progenitor cells into the myogenic program but also control critical aspects of cell survival, proliferation, and migration. In addition, Pax3 and Pax7 also mark a reserve population of myogenic progenitor cells, called satellite cells (see below for more information), which are involved in regeneration.

In humans, translocation of the PAX3 and PAX7 gene to the *FKHR* locus on chromosome 13 produces a skeletal muscle tumor, called alveolar rhabdomyosarcoma. The gene fusions give rise to an in-frame PAX-FKHR chimeric transcription factor, in which the PAX DNA-binding domains are fused to the transcriptional activation domain of FKHR. PAX-FKHR has been shown to drive the generation of discrete nucleated cells from differentiated myofibers and thereby cause neoplastic transformation of skeletal muscle.

Although the branchiomic muscles of the head, like the skeletal muscles of the rest of the body, require MRFs for differentiation, many of the transcription factors involved in specifying myogenic progenitors in the head are just beginning to be discovered and appear to be distinct from those in the trunk. Myogenesis in the head does not depend on Pax3 and Pax7 in the same way as the trunk and limb muscles. Pax3 is not expressed in head muscles, and Pax7, although present in the myogenic cells of the branchial arches that give rise to facial muscles, is expressed after *Myf5* and *MyoD*. Other transcription factors, *Pitx2*, the T-box gene *Tbx1*, and the bHLH genes *capsulin/Pod-1/epicardin/Tcf21* and *MyoR/musculin* encode transcription factors that play roles in specifying progenitor cells that will give rise to branchiomic muscles of the head. *Pitx2* is expressed well before the onset of myogenic progression in the first branchial arch mesodermal core and is essential for the formation of first branchial arch-derived muscle groups. *Pitx2* contributes to the network that specifies pre-myogenic progenitors for extraocular and mastication muscles. *MyoR/musculin* and *capsulin/Pod-1/epicardin*, are transcription factors expressed in specific facial muscle precursors. *MyoR* and *capsulin* are transiently expressed in migratory paraxial mesodermal cells in the branchial arches that appear to represent precursors of the muscles of mastication.

Myostatin, a TGF- β family member, is a secreted protein that acts as a negative regulator of skeletal muscle mass. During embryogenesis, myostatin is expressed by cells in the myotome and in developing skeletal muscle and acts to regulate the final number of muscle fibers that are formed. Myostatin signaling acts on embryonic muscle progenitor cells (*Pax7⁺/Myf5⁺*) that have entered the muscle masses to limit their proliferation by activating p21 and *MyoD* expression. Myostatin affects the balance between proliferation and differentiation of embryonic muscle progenitors by enhancing

their terminal differentiation. Of note, a toddler born with a genetic mutation in the *myostatin* allele developed bulging arm and leg muscles twice the size of other kids his age. This shows that a genetic mutation in myostatin, a negative regulator of skeletal muscle mass, causes a human to naturally build muscle.

The Role of miRNAs in Skeletal Muscle Development

Recent evidence supports a role for miRNAs as integral components of the regulatory circuitry for muscle development. miRNAs are small RNAs that are transcribed as larger transcripts (pri-miRNAs) that are processed by Droscha into stem-loop precursors (pre-miRNAs). The pre-miRNAs are transported from the nucleus to the cytoplasm and subsequently cleaved by Dicer to yield a 20–22-nucleotide mature miRNA. miRNAs act as negative regulators of gene expression by interacting with the 3'-untranslated regions (UTRs) of specific mRNA targets and inhibiting translation or promoting mRNA degradation. The primary determinant of binding specificity to complementary target mRNAs is determined by Watson-Crick base pairing of nucleotides 2–8 at the 5'-end of the miRNA, referred to as the 'seed sequence'. However, nucleotides outside of this region also influence mRNA repression, as does the secondary structure of the surrounding regions of the mRNA 3'-UTR sequence, the number and spacing of target sites, and the location of the miRNA target site relative to the open reading frame. While some miRNAs may exert their effects through pronounced repression of a relatively small number of mRNA targets, more commonly, miRNAs regulate dozens or hundreds of targets by relatively subtle changes in expression. Computational-experimental analyses suggest that approximately 30% of human mRNAs are miRNA targets. It is most likely the summation of the changes in multiple mRNAs that causes the phenotypic effects of miRNAs.

The skeletal muscle miRNAs, miR-1, miR-133, and miR-206, regulate skeletal myoblast proliferation and differentiation. miR-1 and miR-133, both expressed in striated muscle, are generated from a common bicistronic transcript. It is believed that the miR-1/133 locus was evolutionarily duplicated twice to generate the miR-1-2/133a-1 locus, also expressed in striated muscle and the miR-206/133b locus, which is only expressed in skeletal muscle.

The importance of miRNA processing for muscle development and function was demonstrated when skeletal muscle-specific deletion of the Dicer allele resulted in perinatal lethality due to skeletal muscle hypoplasia. As myogenesis progresses from the myoblast stage to the myotube stage, the level of the muscle-specific miR-133 increases and miR-133 represses expression of serum response factor (SRF), a positive regulator of miR-1/133 expression and repressor of myoblast proliferation, thereby establishing negative feedback loops within muscle cell lineages. MEF2 establishes an additional level of myogenic regulation by regulating the expression of miR-1 and miR-133. miR-1 represses expression of HDAC4, a repressor of MEF2 activity, thereby establishing a negative feedback loop to modulate miR-1 and miR-133 expression and promoting myoblast differentiation. Thus, miR-1 upregulation during differentiation is a mechanism to reduce HDAC4 expression and to potentiate MEF2 pro-myogenic activity. miR-206, generated

from a bicistronic transcript that also encodes miR-133b, is exclusively expressed in skeletal muscle. miR-206 is expressed during myogenesis and is involved in the formation of differentiated skeletal muscle cells.

Of interest, a genome scan done on Texel sheep, renowned for their exceptional meatiness, showed a mutation in the myostatin gene that created an illegitimate miRNA target site. The spontaneous mutation of a G-to-A in the 3'-UTR creates a target site for miR-1 and miR-206 resulting in translational inhibition of the *myostatin* gene and dramatic muscularity in the Texel strain of sheep. This enhancement in muscle mass mimics the phenotype seen in mice and humans with mutations in the *myostatin* gene.

Development of Cardiac Muscle

Cardiogenesis

Heart development begins when mesodermal progenitor cells within a bilaterally symmetrical region of the embryo, known as the cardiac crescent or primary heart field, adopt a cardiac fate in response to inductive cues from surrounding tissues. Soon thereafter, these cells converge along the midline of the embryo to form a linear heart tube. Generation of the mature multi-chambered heart from the linear heart tube involves a complex series of events that include rightward looping and diversification of cardiac cell types, balloon-like growth and septation to form the cardiac chambers, and connection to the inflow and outflow tract vasculature (Figure 4).

In the developing heart, there are two heart fields, termed the primary and the secondary heart fields. This designation is based on both the timing of entry of the precursor cells into the heart and the timing of differentiation. The primary heart field, which generates the linear heart tube, serves as the source of precursors of the left ventricle and atrial chambers. The secondary heart field, a population of cardiac precursor cells, is derived from a region of the splanchnic mesoderm, medial to and distinct from the primary heart field that makes up the cardiac crescent. Cells from the secondary heart field are added to the anterior region of the heart tube at the onset of looping and give rise to the outflow tract and right ventricle. Cells from the secondary heart field, in addition to giving rise to the heart, also contribute to the branchiomeric muscles of the head that

are involved in operating the jaw, facial expression, feeding, and breathing.

The outermost surface of the heart is covered by the epicardium. The epicardium develops as a monolayered, squamous epithelial sheet. During the process of cardiac looping, cells of the epithelial proepicardium migrate anteriorly from the proepicardial organ at the posterior end of the heart to cover the surface of the myocardium to form the epicardium, which persists through development and adulthood. It was initially thought that the epicardium was a mere protection to the heart; however, recent new findings have shown that the epicardial cells are important as the origin of part of the coronary system and cardiac interstitial/fibrous cells. A subpopulation of cells from the epicardium undergoes an epithelial-mesenchymal transformation, infiltrating the heart and giving rise to smooth muscle, vascular endothelium, and cardiac fibroblasts of the coronary vasculature. In addition to serving as a source of progenitor cells and signals that are critical for normal heart development, recent studies revealed an unexpected role for the epicardium in myocardial repair following injury.

The Role of Transcription Factors in Cardiac Muscle Development

In contrast to skeletal muscle in which any member of the MRF family can convert a variety of cell types to skeletal muscle myoblasts, there is no single transcription factor that can transform cells to cardiac cells *in vitro*. However, over the decades, studies of heart development have revealed an evolutionarily conserved gene regulatory network consisting of functional interconnections between myogenic transcription factors, their downstream target genes, and upstream signaling pathways that direct cardiac cell fate, myocyte differentiation, and cardiac morphogenesis. In fact, the fruit fly (*Drosophila melanogaster*) has provided a powerful model for delineating the foundation of the cardiac regulatory network. The *Nkx2.5* gene encoding a transcription factor considered to be a cardiac specifier was initially discovered in the fruit fly, in which it was found to be required for a heart development and called *tinman*.

In vertebrate embryos, cardiac precursor cells are specified by signals from adjacent noncardiogenic tissues. Signals from

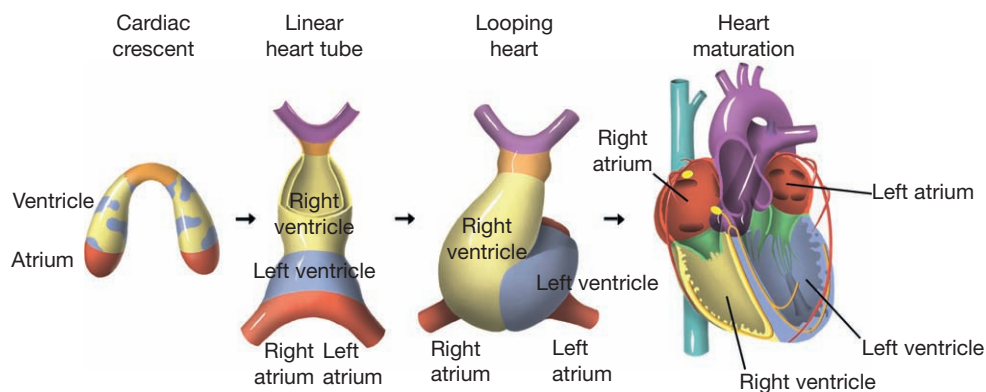


Figure 4 Illustration of the stages of cardiogenesis. Image by Ryan Carre.

the transforming growth factor (TGF)- β superfamily (bone morphogenetic protein (BMP)2/4, TGF- β s, and activin) and the fibroblast growth factor (FGF) family (especially FGF2, 4, 8) are secreted from surrounding tissues to promote the specification of myocardial fate and induce cardiogenesis. In addition, there is evidence that noncanonical Wnt and Sonic hedgehog (Shh) signaling facilitate myocardial induction.

The cardiac progenitors of the primary heart field express both BMP2/4 and Nkx2.5. Nkx2.5, the first cardiac marker gene to be expressed, occupies the center of the transcriptional circuitry, which confers commitment and determination of the heart field precursor cells to a myocardial fate. Nkx2.5 is involved in activation of other transcription factors such as SRF, GATA4/6, Mef2c, Tbx5, and Hand1/2, which impact the transcription of genes involved in the myocardial contractile apparatus and sarcomeric proteins (Figure 5).

The activity of SRF, a transcription factor that binds DNA at a CArG box, is controlled to a large extent by its interaction with tissue-specific regulatory cofactors. There are a large number of SRF cofactors, including GATA4 and Nkx2.5, which form complexes with each other and with SRF. Myocardin, expressed concomitantly with Nkx2.5, is a potent activator of SRF. Remarkably, mis-expression of myocardin can activate cardiac genes even in spinal cord neurons and can convert fibroblasts into smooth muscle cells. Thus, under certain conditions, myocardin acts as a master regulator of cardiovascular cell fates.

The GATA family of zinc-finger-containing genes has been shown to be involved in the progression of differentiation of the mesodermal progenitor cells to a cardiac fate. In the heart, GATA4 is expressed in cardiomyocytes and their mesodermal precursors, as well as in the endocardium and the epicardium. GATA4 binds to and regulates expression of a number of myocardium-expressed genes. GATA factors have been shown to be involved in the migration of the cardiac progenitor cells to form the heart tube and in controlling cell survival in cardiomyocytes. In humans, haploinsufficiency of GATA4 can result in cardiac septal defects.

MEF is expressed in cardiac cells and regulates other transcription factors necessary for heart formation. MEF2 autoregulates its expression and activates a myriad of cardiac-specific genes. MEF2 also controls the timing and magnitude of muscle

gene expression *in vivo* by recruiting HDACs, which act as transcriptional repressors. The interaction of HDACs with MEF2 is a key determinant of heart size from embryonic life to adulthood; mutations in these genes leads to a spectrum of cardiac defects that include ventricular hypoplasia, ventricular-septal defects, and cardiac sudden death.

Members of the T-box transcription factor family are characterized by the presence of a highly conserved 180-amino acid, sequence-specific, DNA-binding domain termed the T-box. T-box proteins function as transcriptional activators or repressors depending on the cell context and have been shown to be crucial to heart development. At least seven members of the T-box gene family are expressed in the embryonic heart in humans and vertebrate models, namely Tbx1-5, Tbx18, and Tbx20, and these genes show overlapping expression patterns in the first and second heart precursor lineages, the myocardium, the endocardium and valves, the conduction system, and the epicardium. Of interest, several human congenital anomaly syndromes have been linked to T-box gene haploinsufficiency.

An important insight into the genetic circuitry of ventricular development was provided by the discovery of two chamber-restricted bHLH transcription factors, Hand1 (also known as eHand/Thing-1/Hxt) and Hand2 (also known as dHand/Hed), which are expressed in the left and right ventricular chambers, respectively. Hand1 expression demarcates the primary heart field, which gives rise to the left ventricle. Hand2 is expressed throughout the linear heart tube before becoming restricted predominantly to the right ventricular chamber. Mice lacking Hand2 fail to form a right ventricle and mice lacking both Hand1 and Hand2 form neither ventricle. These mutant mice were the first to demonstrate that a single gene mutation could result in the ablation of an entire segment of the heart. Clinicians have long known that congenital heart defects in humans typically affect specific structures of the heart, such as the right or left ventricles, leaving the rest of the heart intact. The lack of ventricles in HDAC1/2 mutant mice supports the notion that the heart is assembled in a modular fashion in which each compartment is governed by unique genetic programs. In humans, mutations in the *HAND* genes are associated with hypoplasia of the heart, the most severe form of congenital heart disease.

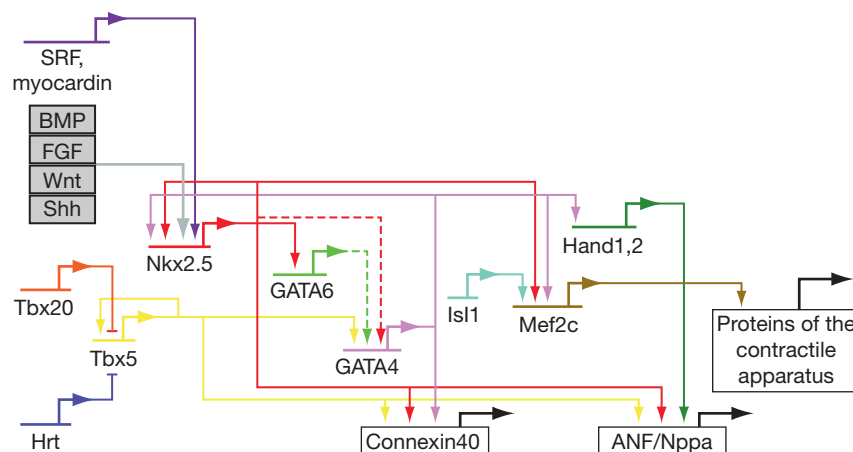


Figure 5 Regulatory pathways during cardiogenesis. Modified from Muñoz-Chápuli R and Pérez-Pomares JM (2010) Cardiogenesis: An embryological perspective. *Journal of Cardiovascular Translational Research* 3: 37–48, with permission from Springer.

Another family of bHLH transcription factors with important functions in cardiovascular development is the hairy-related transcription factor (Hrt) family, also known as Hey, Hesr, CHF, and Herp factors. During embryogenesis, Hrt genes show highly specific expression patterns that demarcate regions of the developing heart, vasculature, pharyngeal arches, and somites. Hrt proteins function as transcriptional repressors downstream of Notch signaling, which regulates binary cell fate decisions during development. Upon activation by ligands, such as Delta and Jagged on the surfaces of adjacent cells, the intracellular domain of the Notch receptor is cleaved and translocated to the nucleus where it cooperates with CSL/RBP-Jk to stimulate transcription of Hrt genes. Hrt expression then fulfills a negative feedback loop to modulate further Notch signaling, thereby resulting in a transient or periodic signal. In addition to binding to an E-box DNA sequence motif, Hrt proteins interact with GATA factors and repress GATA-dependent transcription and cardiomyocyte hypertrophy.

MyoR and capsulin are related bHLH proteins that are co-expressed in branchiomic muscles and the primary heart field. Both genes also show unique expression domains in trunk skeletal muscle precursors and in derivatives of the lateral and intermediate mesoderm. Capsulin also marks the proepicardial organ and developing epicardium and, after birth, is expressed in a population of cells associated with the coronary vasculature. MyoR and capsulin mutant mice have revealed these genes to play redundant roles in specification of branchiomic muscles and show MyoR and capsulin as unique transcriptional regulators for the development of specific head muscles.

The secondary heart field builds the right ventricle and the outflow tract including its myocardial, smooth muscle, and endothelial investments, as well as parts of the atria. Islet1 (Isl1), a LIM-homeobox transcription factor is expressed in the secondary heart field and has been defined as a marker for a cardiac progenitor cell lineage that is capable of differentiating into all three major cell types of the heart: cardiomyocytes, smooth muscle, and endothelial cell lineages. More importantly, isolated progenitor cells expressing Isl1 are capable of self-renewal and expansion before differentiation into the three major cell types in the heart. These Isl1-positive cells serve as a cardiac stem cell population and offer novel opportunities for human regenerative cardiovascular medicine.

Most recently, it was reported that a combination of three developmental transcription factors, Gata4, Mef2c, and Tbx5, rapidly and efficiently reprograms dermal fibroblasts into differentiated cardiomyocyte-like cells. Although no single master regulator has been identified to direct cardiac reprogramming, this report shows that functional cardiomyocytes can be directly reprogrammed from differentiated somatic cells by defined factors. This discovery provides new prospects for an easy, accessible source of cardiomyocytes for regenerative approaches to combat heart disease.

The Role of miRNAs in Cardiac Muscle Development

Similar to skeletal muscle development, miRNAs are integral components of the regulatory circuitry for cardiac muscle development. Cardiac deletion of Dicer, the enzyme that cleaves pre-miRNAs to form mature miRNAs, results in embryonic lethality implying that miRNAs are essential for normal

heart development. miR-1, miR-133, and miR-208 are exclusively expressed in striated muscle, both skeletal muscle and heart. miR-206 is expressed in skeletal muscle and not in cardiac muscle. miR-1, a conserved miRNA ranging from fruit flies to humans, is expressed in striated muscle and is implicated in heart development. miR-1 and miR-133 are both expressed in the heart and are generated from a common bicistronic transcript.

There is an intricate relationship between miRNAs and transcription factors. miRNA expression is regulated by transcription factors. For example, the miR-1/133a locus is regulated by SRF and MEF2. These are the transcription factors that activate genes involved in muscle development and function, highlighting the integration of miRNA and mRNA regulatory networks.

Both miR-1 and miR-133 control cardiac fate; however, they have different target mRNAs and exert antagonistic effects on differentiation of muscle lineages. miR-1 promotes differentiation of embryonic stem cells toward a cardiac fate, whereas miR-133 inhibits differentiation into cardiac muscle. Overexpression of miR-1 in embryonic cardiomyocytes results in decreased proliferation of cardiomyocytes, a phenotype that has been attributed to a decrease in the expression of Hand2, a transcription factor that promotes cardiomyocyte proliferation. Other functions have been attributed to miR-1 and miR-133a, such as regulation of apoptosis in cardiomyocytes; miR-1 and miR-133a have also been shown to have functions within the vertebrate cardiac conduction system. miR-1 also modulates numerous ion channels involved in cardiac conduction.

Cardiac contraction is dependent on myosin heavy chain (MHC) proteins which are located in the sarcomere. There is an intimate relationship between MHC and miRNAs. The three muscle-specific myosin genes, Myh6 (α -MHC), Myh7 (β -MHC), and Myh7b, encode a family of intronic miRNAs (miR-208a, miR-208b, and miR-499), called MyomiRs, which control muscle myosin content and remodeling, myofiber identity, and muscle performance. The functions of the MyomiRs in the human heart are still being examined. In addition to regulation of cardiomyocyte development and function, recent reports have revealed roles of miRNAs in vascular development. The cardiovascular-specific miRNAs, miR-143 and miR-145, play key roles in modulating vascular smooth muscle cells.

MiRNAs are not only required for normal heart development and function but also involved in cardiac remodeling and disease. Multiple studies have shown an expression profile of distinct stress-responsive miRNAs during cardiac hypertrophy and heart failure. In fact, it has been reported that expression of individual miRNAs causes or alleviates disease. These studies have led the way to the concept of pharmacological manipulations of miRNAs as therapeutic reagents. AntagomiRs are cholesterol-modified antisense oligonucleotides that can be delivered *in vivo* and have been shown to inhibit miRNAs. By contrast, miRNA mimics are oligonucleotides that contain a mature miRNA sequence and can function as miRNAs to repress endogenous target genes. Since miRNAs control numerous facets of cardiac biology (including myocyte growth and survival, contractility, energy metabolism, fibrosis, and angiogenesis), miRNA-based therapeutics represents a potentially powerful approach to treat cardiovascular and skeletal muscle diseases.

Regeneration, Repair, and Striated Muscle Stem Cells

Skeletal Muscle Stem Cells

Adult skeletal muscle has the ability to regenerate and repair after injury. This is primarily due to a resident muscle stem cell population, called satellite cells. During skeletal muscle development, a progenitor cell population that expresses Pax3 and Pax7 arises in the center of the somite. As myogenesis progresses, Pax3 expression is decreased and the progenitor cell population, which expresses Pax7, occupies a satellite position around myofibers. Each muscle contains a pool of resident muscle stem cells that can either differentiate into muscle fibers or remain as proliferating progenitors.

In adult skeletal muscle, satellite cells reside as quiescent cells located under the basal lamina of skeletal muscle fibers. Following muscle injury, satellite cells are activated, move outside of the basal lamina, co-express Pax7 and MyoD, and differentiate into proliferating myoblasts that also express Myf5. Subsequently, they decrease Pax7 expression, express myogenin, and fuse to form multinucleated myofibers to repair damaged skeletal muscle.

A fraction of the Pax7 and MyoD expressing myoblasts will lose MyoD, Myf5, and myogenin expression, and stop dividing. It is believed that these cells become quiescent satellite cells and self-renew the satellite cell population. This indicates that satellite cells repair damaged muscle and self-renew, thus establishing them as skeletal muscle stem cells.

As skeletal muscle throughout the body is not derived from the same cell population, the same holds true for the satellite cells. The progenitor cells expressing Pax3 and Pax7 in the central region of the somite (central dermomyotome) give rise to satellite cells of the trunk. By contrast, the limb muscle satellite cells derive from the Pax3-expressing muscle progenitor cells that migrate from another region of the somite (hypaxial dermomyotome).

MiRNAs are emerging as regulators of skeletal muscle stem cells. For example, miRNA-27b (miR-27b) is expressed during myogenesis in the myotome and in activated satellite cells of adult muscle and targets the 3'-UTR of Pax3 mRNA. Recent studies showed that overexpression of a miR-27b in the embryo leads to downregulation of Pax3, resulting in interference with progenitor cell migration and in premature myogenic differentiation. AntagomiRs of miR-27b allow for continued Pax3 expression leading to more proliferation and a delay in the onset of differentiation.

Cardiac Stem Cells

Unlike skeletal muscle, it was considered dogma that the heart did not have resident cells that divide. However, recently it has been shown that the heart does have a cardiac stem cell population. There are different pools of cardiac stem cells and they are classified by their surface markers (c-kit, Sca-1, Islet 1, etc.). These cells are believed to comprise about 2% of the total heart cells and contribute to a natural turnover of cardiomyocytes.

In response to injury, such as a heart attack (myocardial infarction), cardiac stem cells are activated to form myocytes, which can restore some dead tissue and can contribute to normalization of cardiac function. However, unlike satellite cells, cardiac cells cannot fully repair the damaged heart and provide only a partial regenerative response.

The epicardium, the outermost surface of the heart, is another source of adult cardiac progenitor cells. It plays a role in the formation of the coronary vasculature during development and retains a regenerative potential in the adult. Understanding the transcription factors and miRNAs that regulate cardiac stem cells is under intense investigation. This information is key to the goal of using patient-specific cardiac stem cell therapy to regenerate and repair an injured heart with the intention of combating heart disease.

Conclusion

As highlighted in this article, skeletal and cardiac muscles use different evolutionarily conserved networks of transcription factors and miRNAs to regulate the programs controlling cell differentiation and morphogenesis during development. Due to the evolutionary conservation across diverse organisms, many of the regulatory processes were identified in a variety of organisms ranging from fruit flies to humans. Continued integration and collaboration among developmental and molecular biologists will reveal additional genetic networks involved in striated muscle development. These new discoveries will not only provide new insights into the general principles of organogenesis but also provide new approaches to facilitate therapies for cardiac and skeletal muscle diseases.

See also: **Bioenergetics:** Mitochondria in Myocardial Ischemia; **Cell Architecture and Function:** Myosin Motors; **Metabolism Vitamins and Hormones:** Metabolic Control during Ischemia of the Heart; **Molecular Biology:** MicroRNAs in Eukaryotes.

Further Reading

- Bollini S, Smart N, and Riley PR (2010) Resident cardiac progenitor cells: At the heart of regeneration. *Journal of Molecular and Cellular Cardiology*, doi:10.1016/j.yjmcc.2010.07.006.
- Braun T and Gautel M (2011) Transcriptional mechanisms regulating skeletal muscle differentiation, growth and homeostasis. *Nature Reviews Molecular Cell Biology* 12: 349–361.
- Buckingham M and Relaix F (2007) The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. *Annual Review of Cell and Developmental Biology* 23: 645–673.
- Dyer LA and Kirby M (2009) The role of secondary heart field in cardiac development. *Developmental Biology* 336: 137–144.
- Eisenberg I, Alexander MS, and Kunkel LM (2009) miRNAs in normal and diseased skeletal muscle. *Journal of Cellular and Molecular Medicine* 13: 2–11.
- Ieda M, Fu JD, Delgado-Olguin P, et al. (2010) Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* 142: 375–386.
- Kuang S and Rudnicki MA (2008) The emerging biology of satellite cells and their therapeutic potential. *Trends in Molecular Medicine* 14: 82–91.
- Liu N and Olson EN (2010) MicroRNA regulatory networks in cardiovascular development. *Developmental Cell* 18: 510–525.
- Olson EN (2006) Gene regulatory networks in the evolution and development of the heart. *Science* 313: 1922–1927.
- Potthoff MJ and Olson EN (2007) MEF2: A central regulator of diverse developmental programs. *Development* 134: 4131–4140.
- Sanger JM and Sanger JW (2008) The dynamic Z bands of striated muscle cells. *Science Signaling* 1: pe37.
- Tajbakhsh S (2009) Skeletal muscle stem cells in developmental versus regenerative myogenesis. *Journal of Internal Medicine* 266: 372–389.
- Tapscott SJ (2005) The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* 132: 2685–2695.

Biochemistry of Hematopoiesis

A D Friedman, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Leucine zipper An amphoteric α -helical protein motif that includes several leucines on the hydrophobic surface that mediates coiled-coil protein dimerization.

Stem cell A cell capable of symmetric divisions giving rise to two additional stem cells or of asymmetric divisions giving rise to a new stem cell and a more committed progenitor.

Transcription factor A protein that binds to DNA in order to activate or repress gene expression.

Zinc finger A protein motif that complexes zinc ion via cysteine and histidine residues to form a structure that mediates DNA binding when present within transcription factors.

Developmental Hematopoiesis

Hematopoietic cells develop from the mesodermal germ layer, with embryonic hematopoiesis beginning in close association with endothelial cells within blood islands in the yolk sac. A primitive hemangioblast gives rise to endothelial cells and primitive hematopoietic progenitors at this early stage of development. Additional blood vessels throughout the embryo develop from a separate mesodermally derived angioblast. Embryonic hemoglobin predominates in primitive erythrocytes, which retain their nuclei. Primitive macrophages and neutrophils also develop, but the presence of pluripotent hematopoietic stem cells (HSCs) capable of giving rise to all the adult hematopoietic and lymphoid lineages awaits the development of definitive hematopoiesis. As development proceeds, a second wave of hematopoiesis initiates from a specialized hemogenic endothelium in the dorsal aorta within the aorta–gonad–mesonephros region of the embryo and in placental vessels. Both the transcription factors SCL, a member of the basic region helix–loop–helix or bHLH family, and GATA-2, a zinc finger protein, are required for the emergence of primitive hemangioblasts. SCL is also required for the formation of definitive hemogenic endothelium, and the transcription factor RUNX1 is required for subsequent emergence of HSC (Figure 1). Multiple endothelial promoters are regulated by cooperation between the FoxoC2 and ETV2, Ets family members, suggesting that these factors direct the endothelial fate of hemangioblasts or angioblasts. HSCs initially populate the fetal liver and ultimately the bone marrow. Red cells derived from HSCs lack nuclei, initially express fetal hemoglobin, and ultimately express adult hemoglobin consisting of α - and β -globin chains.

Hematopoietic Stem Cells

A diagram depicting key regulators required for development of adult hematopoietic lineages from the HSC is shown in Figure 2. Neither SCL nor GATA-2 is required for maintenance of HSC. In contrast, the absence of either TEL, an Ets family member, or the c-Myb or MLL transcription factors leads to a marked defect in HSC activity, whereas the absence of C/EBP α ,

a basic-region leucine zipper transcription factor, leads to HSC expansion, reflecting the ability of C/EBP α to inhibit cell cycle progression. At any given time, most HSCs remain in a quiescent state in marrow osteoblast or endothelial niches and enter cell cycle as needed throughout life to provide mature blood elements. As with all true stem cells, HSCs are capable of giving rise either to additional HSCs or to mature blood progenitors. Two key mediators of an early developmental decision made by HSCs are GATA-1, a zinc finger transcription factor, and PU.1, an Ets family protein. GATA-1 specifies formation of the megakaryocyte–erythroid progenitor (MEP), while PU.1 directs formation of the lymphoid–myeloid progenitor (LMP). GATA-1 and PU.1 each activate their own promoters to provide positive feedback for lineage specification, and GATA-1 and PU.1 antagonize each other's activities via direct protein interaction to prevent reversal of one or the other developmental decision.

Erythropoiesis and Megakaryopoiesis

The MEP commits to the erythroid lineage under the direction of GATA-1 in combination with EKLF, a zinc-finger protein, with additional contributions from SCL and NF-E2, a bZIP factor. Alternatively, MEPs give rise to the megakaryocyte lineage, directed by RUNX1 and members of the Ets family, including Fli-1 and ERG. GATA-ETS composite DNA elements help regulate several megakaryocytic genes. Of note, RUNX1 is absent from the erythroid lineage within hematopoiesis and directly interacts with GATA-1 to also induce megakaryocytic genes. FOG-1 is a transcriptional coregulator that cooperates with GATA-1 to activate erythroid genes and to repress megakaryocytic genes during erythropoiesis, while having the opposite effect on these genes during megakaryopoiesis.

Lymphopoiesis

The LMP gives rise to a common lymphoid progenitor (CLP) under the direction of Ikaros, a zinc finger protein. The CLP further matures along the B-lymphoid lineage under the direction of EBF, E2A, Pax-5, and PU.1. EBF binds DNA via a novel

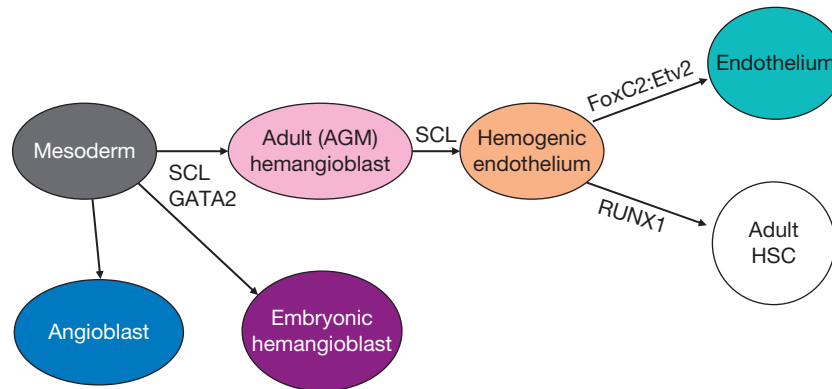


Figure 1 Diagram depicting transcription factors and the cellular pathways leading to development of adult hematopoietic stem cells. AGM, aorta-gonad-mesonephros; HSC, hematopoietic stem cell.

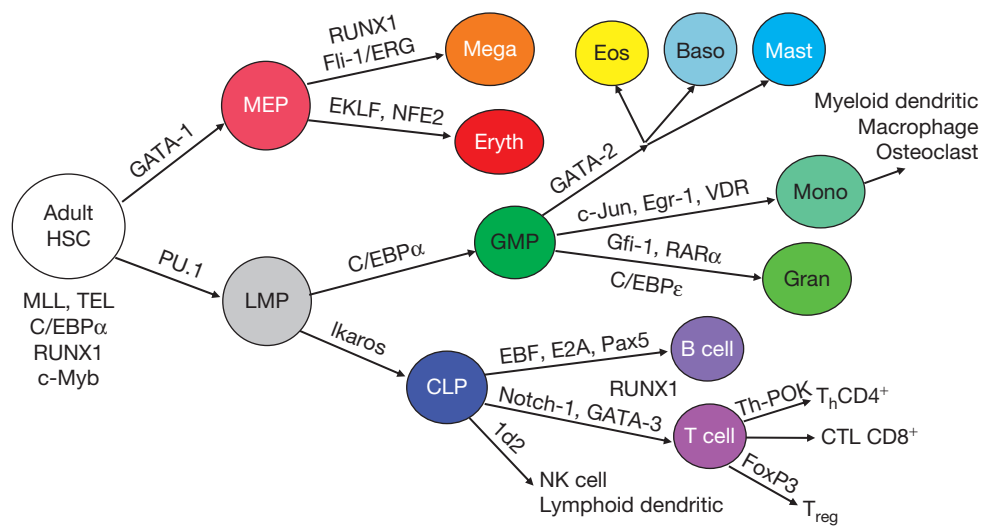


Figure 2 Diagram depicting transcription factors and the cellular pathways leading to development of the adult hematopoietic lineages. CLP, common lymphoid progenitor; GMP, granulocyte-monocyte progenitor; HSC, hematopoietic stem cell; LMP, lymphoid-myeloid progenitor; MEP, megakaryocyte-erythroid progenitor; NK cell, natural killer cell; VDR, vitamin D receptor.

zinc-binding motif, E2A is a bHLH protein, and Pax-5 contacts DNA via a paired-box homeodomain. Low levels of PU.1 are required for B-lymphoid development, whereas monocytic maturation requires increased PU.1 levels. The CLP also gives rise to the T-lymphoid lineage under direction of Notch-1 and GATA-3. Notch-1 simultaneously suppresses E2A activity to inhibit B-cell development. RUNX1 is also required for optimizing B- or T-lymphoid development. RUNX1 favors the formation of $CD4^+CD8^+$ cytotoxic T cells through repression of the gene encoding the Th-POK transcription factor as well as the CD4 gene, and RUNX1 favors formation of $CD4^+CD25^+$ regulatory T cells through induction of FoxP3. Deletion of FLT3, a receptor tyrosine kinase, disrupts B lymphopoiesis in murine models, indicating a role for specific receptor signals in specification of this lineage. Natural killer cells and lymphoid dendritic cells also trace their origins to the CLP, dependent on expression of Id2, an antagonist of bHLH proteins such as E2A.

Monopoiesis and Granulopoiesis

The LMP gives rise to the granulocyte-monocyte progenitor (GMP) under direction of C/EBP α , and C/EBP α simultaneously suppresses lymphoid commitment through inhibition of Pax-5 and Notch-1 activities. C/EBP α induces PU.1 gene transcription to provide the increased levels required for monocyte lineage development. C/EBP α also forms complexes with c-Jun or c-Fos via their leucine zipper domains to generate heterodimeric complexes that favor monopoiesis. In contrast, C/EBP α homodimers may favor granulopoiesis. The zinc finger protein Gfi-1 helps direct granulopoiesis, and cross-inhibition between Gfi-1 and Egr-1 fixes monocyte versus granulocyte lineage specification. Retinoic acid receptor α (RAR α) complexes with RXR to help direct granulopoiesis, whereas vitamin D receptor binds RXR to contribute to monopoiesis. C/EBP β can assist C/EBP α during stress granulopoiesis and cooperates with nuclear factor

kappa B (NF- κ B) in mature monocytes/macrophages to direct expression of inflammatory mediators, and NF- κ B also contributes to the formation of monocytic and granulocytic progenitors. A third C/EBP protein, C/EBP ϵ , directs terminal phases of granulopoiesis. A micro-RNA, miR-223, is induced by C/EBP α to contribute to granulopoiesis via down-modulation of NFI-A. G-CSF activates STAT3 and the SHP2 tyrosine phosphatase, whereas M-CSF specifically activates PLC- γ and induces increased ERK activation, and small molecule inhibition of SHP2 impedes granulopoiesis, whereas inhibition of ERK reduces monopoiesis *in vitro*, suggesting a role for cytokine signal transduction in myeloid lineage specification. Immature monocytes give rise to mature macrophages, or alternatively to osteoclasts in response to M-CSF and RANKL or myeloid dendritic cells in response to IL-4 and GM-CSF.

Eosinophils, Basophils, and Mast Cells

The GMP also gives rise to eosinophils, basophils, and mast cells, dependent upon the order of transcription factor expression. When GATA-2 increases in GMP expressing C/EBP α , commitment to the eosinophil lineage occurs. Expression of GATA-2 combined with reduction of C/EBP α generates basophil/mast cell progenitors, and a subsequent increase in C/EBP α produces basophils, whereas further reduction of C/EBP α generates mast cells.

See also: **Protein/Enzyme Structure Function and Degradation:** Zinc Fingers; **Signaling:** B-Cell Antigen Receptor; Hematopoietin Receptors; Nuclear Factor Kappa B; T-Cell Antigen Receptor; Vitamin D Receptor.

Further Reading

- Friedman AD (2007) Transcriptional control of granulocyte and monocyte development. *Oncogene* 26: 6816–6828.
- Friedman AD (2009) Cell cycle and developmental control of hematopoiesis by Runx1. *Journal of Cellular Physiology* 219: 520–524.
- Goldfarb AN (2007) Transcriptional control of megakaryocyte development. *Oncogene* 26: 6795–6802.
- Iwasaki H and Akashi K (2007) Hematopoietic developmental pathways: On a cellular basis. *Oncogene* 26: 6687–6696.
- Kim S-I and Bresnick EH (2007) Transcriptional control of erythropoiesis: Emerging mechanisms and principles. *Oncogene* 26: 6777–6794.
- Lancrin C, Sroczynska P, Stephenson C, et al. (2009) The haemangioblast generates haematopoietic cells through a hemogenic endothelium stage. *Nature* 457: 892–895.
- Laslo P, Pongubala JMR, Lancki DW, and Singh H (2008) Gene regulatory networks directing myeloid and lymphoid cell fates within the immune system. *Seminars in Immunology* 20: 228–235.
- Paz-Priel I, Ghosal AK, Kowalski J, and Friedman AD (2009) C/EBP α or C/EBP α oncoproteins regulate the intrinsic and extrinsic apoptotic pathways by direct interaction with NF- κ B p50 bound to the bcl-2 and FLIP gene promoters. *Leukemia* 23: 365–374.
- Pina C and Enver T (2007) Differential contributions of haematopoietic stem cells to foetal and adult hematopoiesis: Insights from functional analysis of transcriptional regulators. *Oncogene* 26: 6750–6765.

Biochemistry of Liver Regeneration

B Ochoa, University of the Basque Country UPV/EHU, Bilbao, Spain

A M Diehl, Duke University Medical Center, Durham, NC, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Hedgehog ligands Proteins, such as Indian hedgehog and Sonic hedgehog, whose transient production is necessary for normal liver regeneration and survival but whose prolonged production promotes cirrhosis and neoplasia.

Hedgehog pathway Morphogenic signaling pathway that orchestrates tissue construction during fetal development and is silenced during hepatocyte differentiation and in healthy adult liver, but it is required for normal adult liver regeneration after chronic epithelial injury.

Hepatic steatosis Transient accumulation of lipids within hepatocytes that occurs during liver regeneration as a result of discrete alterations in the oxidation, transport mechanisms, and signaling pathways of fatty acids.

Liver regeneration Highly orchestrated process involving multiple intra- and intercellular pathways and signals that reconstructs the entire adult liver to restore optimal tissue-specific functions.

Liver regeneration phases An artificial temporal classification of related cellular and regulatory events required to accomplish liver regeneration. Acquisition of replicative competence of cells and extracellular matrix degradation occur in the priming phase; expansion of the cell population occurs in the proliferation phase; and cell growth is suppressed and extracellular matrix synthesized to terminate regeneration at a set point.

Matrix remodeling Sequential degradation and formation of extracellular matrix components to restore cell-to-matrix relationships among the various liver cell populations.

Stromal cells Stellate cells, endothelial cells, and immune cells, such as specialized macrophages and lymphocytes, that accompany epithelial cells (hepatocytes and cholangiocytes) in the liver and that need to be replaced in order to restore normal liver function.

Stromal elements Stromal cells, vasculature, and biliary conduits that support the viability and function of mature hepatocytes and cholangiocytes.

Introduction

Liver regeneration describes the process that reconstructs damaged adult liver to restore optimal tissue-specific functions. Adult livers become damaged as a result of normal day-to-day wear and tear, as well as by superimposed stresses imposed by various insults, such as trauma, exposure to endogenous or environmental toxins, infectious agents and/or immune system-mediated attack, disturbed perfusion, and processes that obstruct bile flow. Most of the research about liver regeneration has focused on mechanisms that regulate the proliferative activity of mature hepatocytes after 70% partial hepatectomy (PH). However, it is important to recognize that successful regeneration requires much more than the mere replacement of one type of liver epithelial cell (i.e., hepatocytes), because the healthy adult liver comprises multiple cell types. In addition to mature liver epithelial cells (hepatocytes and cholangiocytes) and their progenitors, there are different types of stromal cells. The latter include hepatic stellate cells, endothelial cells, and various types of immune cells, such as macrophages and lymphocytes. Liver regeneration after PH requires replenishing all of these cells, and reestablishing normal cell-to-cell and cell-to-matrix relationships among various liver cell populations so that the tissue can resume its normal biosynthetic, metabolic, and excretory functions. Given that the goal of liver regeneration is to reconstruct the healthy tissue, a brief review of normal hepatic morphology is justified.

Healthy Adult Liver Architecture

The adult liver is configured to optimize exposure of hepatocytes to blood and to provide efficient egress of lipid-soluble products from hepatocytes into the biliary tree for eventual delivery into the intestinal lumen. Unlike other organs, the liver has two sources of afferent blood supply, which are carried into the hilum of the organ via the portal vein and hepatic artery. Each of these vessels arborizes to give rise to progressively smaller branches that terminate in multiple small portal triads, which also house small intralobular branches of the biliary tree. From these portal triads, portal venous and hepatic arterial blood flows into a network of sinusoids that are lined by fenestrated endothelial cells. Only a small space (dubbed the space of Disse) separates the sinusoidal endothelial cells from the basal surfaces of the hepatocytes that are oriented along the sinusoids because neither the fenestrated endothelia nor the hepatocytes elaborate a typical basement membrane. The sinusoids gradually coalesce into larger vessels that become the terminal branches of the hepatic vein, which drains venous blood from the liver into the inferior vena cava. Thus, blood flows from portal triads toward terminal hepatic venules, bathing hepatocytes in factors derived from both the systemic and the mesenteric circulations. Bile flows in the opposite direction within bile canaliculi that are lined by the apical membranes of hepatocytes. Bile originates from hepatocytes that rim terminal hepatic venules and becomes progressively enriched with products that are released from the apical domains of hepatocytes that are closer to the

portal triads. As bile canaliculi approach the portal triads, their lumens become lined by cholangiocytes that form the intralobular bile ducts.

Although hepatocytes are large, and, thus, the most conspicuous cell type that lines hepatic sinusoids, other cells also localize in these areas. Hepatic stellate cells (which store retinoids and other lipids) reside in the space of Disse and the sinusoids themselves are enriched with various types of immune cells, particularly hepatic macrophages (dubbed Kupffer cells) and natural killer T cells (NKT cells; specialized T lymphocytes that respond to glycolipid antigens that are presented by CD1 molecules on the surfaces of other resident liver cells). Precursors of hepatic immune cells (Kupffer cells and lymphocytes) and sinusoidal endothelial cells are presumed to be recruited to the liver from extrahepatic sources (e.g., bone marrow and lymphoid tissues), and then to differentiate *in situ*, generating replacement cells for immune and endothelial cells that turnover during both health and disease states. The origin of replacement hepatic stellate cells in adult livers is unknown (see below) (Figure 1).

The portal tracts (which harbor the intralobular branches of the portal vein, hepatic artery, and the bile duct) also include the cells and stroma that comprise the walls of these conduits. Hence, portal tracts are enriched with fibroblastic cells, smooth muscle cells, and fibrous matrix. In addition, they contain occasional neurons and glial cells that form the peri-vascular and peri-ductular neural plexus. The portal regions are also enriched with progenitor populations for the liver epithelial compartments, including bipotent progenitors that are capable of differentiating into either hepatocytes or cholangiocytes.

Some of these progenitors are oriented along the bile canaliculi nearest portal triads (dubbed canals of Hering). Finally, the entire external surface of the liver is encapsulated in a serosal lining that is comprised of mesothelial cells and associated stromal elements.

Recent studies of developing fetal livers suggest that cells that reside immediately beneath the liver capsule may be precursors of hepatic stellate cells in developing livers. However, it remains to be determined whether or not stellate cells originate from subcapsular progenitors in adult livers. Adult stellate cell populations, themselves, exhibit considerable phenotypic plasticity, expressing stem/progenitor, neural, and adipocyte markers when quiescent, and efficiently transitioning into myofibroblastic cells during many types of liver injury. In addition, cultured stellate cells can be induced to express markers of other liver cell types, including endothelial cells, cholangiocytes, and hepatocytes. These findings have prompted speculation that stellate cells may be capable of functioning as multipotent progenitors under certain circumstances. At present, however, definitive proof of this concept is lacking. In any case, it is evident that healthy adult livers comprise multiple different types of cells (including progenitors), that many of these cells retain considerable phenotypic plasticity, and that different cell types interact to maintain the normal structure and function of the liver.

Liver Regeneration after PH

Resection of 70% of healthy adult liver (PH) triggers regenerative mechanisms that result in complete restoration of liver

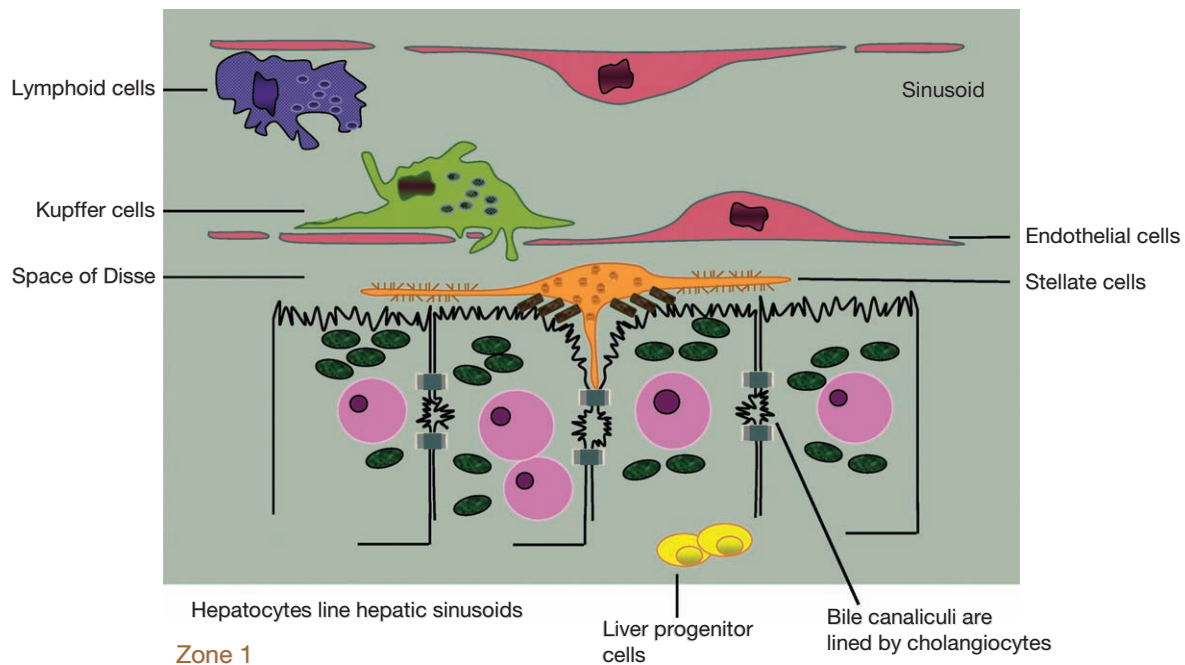


Figure 1 The cellular complexity of the liver. Two major epithelial cell types constitute the liver: hepatocytes and cholangiocytes. Some liver diseases significantly impair the ability of hepatocytes to replenish the organ, and liver progenitor cells become activated. From the periportal zone of the liver (zone 1), liver progenitor cells or oval cells, due to their oval shape nucleus, proliferate, infiltrate along the liver plate, and differentiate into hepatocytes and cholangiocytes to restore liver function and mass. Bile canaliculi are connected with bile ducts through the canal of Hering, which is the niche of the liver progenitor cells. Kupffer cells, endothelial cells, stellate cells, and myofibroblastic cells are also resident liver cells.

mass and function within a week or so in adult mice and rats, and within months in adult humans. Although this process involves hyperplasia of residual mature hepatocytes, other cell types that were removed during the resection must also be replaced in order to restore the flow of blood and bile within the mass of newly formed hepatocytes. Without the latter, full recovery of liver-specific metabolic, biosynthetic, and detoxification functions would not be possible. Therefore, liver regeneration after PH requires reconstruction of stromal elements (e.g., hepatic stellate cells and resident immune cells), vasculature, and biliary conduits that support the viability and function of mature hepatocytes. Unfortunately, however, unlike mechanisms regulating hepatocyte proliferation (which have been extensively researched), mechanisms for replacing other types of liver cells and for reestablishing appropriate cell–cell interactions among various liver cell populations after PH have not been studied much and, thus, are poorly understood. We postulated that regeneration of the adult liver after PH might recapitulate certain aspects of fetal liver development because both situations require global construction of a complex tissue. Before discussing evidence for and against the concept that adult liver regeneration recapitulates fetal development, a brief review of pertinent aspects of fetal liver development is necessary.

Liver Morphogenesis during Fetal Development

During embryogenesis, the liver bud is formed via a process that requires paracrine signaling and cell migration between the developing heart (mesoderm), septum transversum mesenchyme (mesoderm and ectoderm), and the ventral endoderm. This occurs between days E8.5 and E9.5 in mice. Bone-marrow-derived cells subsequently infiltrate the liver bud and provide signals that are critical for further growth of the developing liver. Hence, liver morphogenesis involves movement of, and interactions among, cells that originated in the mesoderm, ectoderm, and the endoderm. A detailed review of the signaling mechanisms that orchestrate this process is beyond the scope of this article, but they are beautifully summarized elsewhere.

Suffice it to say that work in fetal explants showed that fibroblast growth factors (FGFs) and bone morphogenetic proteins (BMPs) released from the developing heart field are required for hepatic specification of the ventral endoderm. Further, it was demonstrated that hepatic specification was accompanied by induction of Sonic hedgehog (Shh) expression in endodermal cells, and that the latter did not occur unless FGF was provided. Those studies, therefore, suggested that Hedgehog (Hh) pathway activation was among the earliest events that occurred during construction of the fetal liver. Due to the pleiotropic actions of the Hh pathway during development, and functional redundancies among different Hh ligands and Hh-regulated transcription factors, the precise role(s) of Hh signaling in hepatic morphogenesis has (have), however, been difficult to demonstrate. Nevertheless, the following data support a role for the Hh pathway in fetal liver morphogenesis. First, pluripotent embryonic stem cells, endoderm-committed multipotent progenitors, and some committed

hepatic progenitors from fetal livers produce and respond to Hh ligands. Second, FoxA1 and FoxA2 (transcription factors which have been proved to be essential for hepatic specification of endodermal progenitors) are Hh target genes. Third, Hh signaling promotes viability and growth of Hh-responsive fetal liver progenitors. Fourth, the Hh pathway is known to regulate growth and differentiation of endothelial progenitors, thymic-derived lymphocyte progenitors, and bone-marrow-derived hematopoietic progenitors (all of which infiltrate the developing liver bud at some point during fetal life). Such cells also generally produce Hh ligands. Fifth, activation of Hh signaling promotes migration in multiple types of Hh-responsive cells (Figure 2).

In summary, many of the cell types which migrate into and accumulate within the fetal liver are capable of producing and/or responding to Hh ligands, and these factors regulate important cell-fate decisions, as well as the net growth, of those Hh-responsive cell populations. It seems, however, that the Hh pathway is gradually repressed during hepatocyte differentiation, because mature hepatocytes in healthy adult livers neither produce, nor respond to, Hh ligands. Moreover, although many other types of resident cells in healthy adult livers are capable of producing and/or responding to Hh ligands, only rare cells in and around portal tracts exhibit appreciable Hh pathway activity in uninjured adult livers (see below). The mechanisms that silence Hh signaling during hepatocyte differentiation, as well as those that more generally repress Hh pathway activation in Hh-responsive populations of other resident liver cells, remain to be discovered.

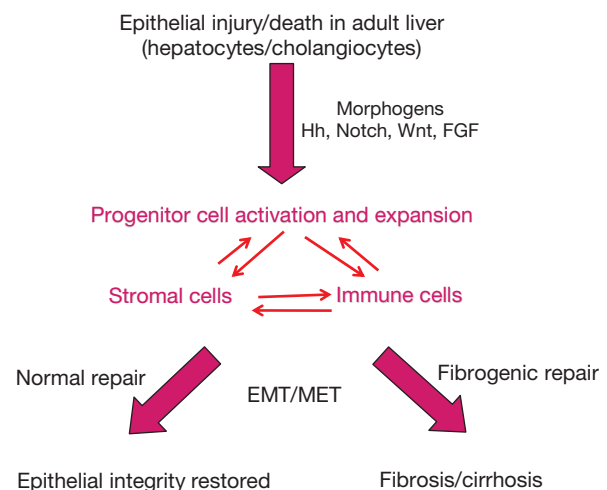


Figure 2 Normal or fibrogenic repair of the liver. Unlike other organs, the liver is capable of efficiently reconstructing the entire tissue without scarring (i.e., regenerating) following acute resection of most of its mass. This suggests that adult livers may be particularly adept at reactivating, and appropriately modulating, morphogenic mechanisms that orchestrate tissue construction during fetal development. However, sometimes liver injury leads to progressive fibrosis and architectural distortion (i.e., cirrhosis), rather than effective regeneration. Whether repair results in regeneration or cirrhosis may depend upon the success with which developmental morphogenic mechanisms are reactivated and then repressed.

Reactivation of the Hh Pathway in Regenerating Livers after PH

Because many types of cells (not merely mature hepatocytes) must be replaced after PH, and, as mentioned above, most of these are derivatives of Hh-responsive progenitors, we examined the possibility that the Hh pathway might become reactivated in the remnant liver following PH. Healthy adult mice were subjected to PH and sacrificed at various time points ranging from 6 h to 10 days post-PH so that Hh pathway activity could be examined in whole liver and primary liver cells that were harvested from regenerating liver remnants. PH was followed by an immediate and sustained reduction in hepatic expression of Hh interacting protein (Hhip), a soluble inhibitor of Hh ligands, and sequential increases in messenger RNA (mRNA) and protein levels of the Hh ligands, Indian hedgehog (Ihh), and Shh. Subsequently, expression of the Hh-regulated transcription factors, Gli1 and Gli2, also increased, followed by significant upregulation of well-established Gli-target genes, such as soluble frizzled related peptide (sFrp)-1. These findings indicated that PH resulted in Hh pathway activation in regenerating livers.

Immunohistochemistry and cell isolation studies proved that the hepatocyte and bile duct populations became particularly enriched with Hh-responsive cells (i.e., cells that expressed Hh-regulated transcription factors) around the times when these cells were maximally proliferative following PH. Moreover, treating mice with cyclopamine, a specific inhibitor of Hh signal transduction, significantly inhibited normal post-PH increases in both hepatocyte and cholangiocyte proliferation. Cyclopamine treatment also dramatically reduced post-PH survival in mice during the initial 72h after PH. When regenerating hepatocytes that had been harvested from mice 24h after PH were treated *in vitro* with cyclopamine, Hh signaling and proliferation were also reduced, proving that Hh pathway activation directly contributed to increases in hepatocyte proliferation that normally follow PH. Given that healthy mature hepatocytes are not Hh-responsive, this finding was curious.

One potential explanation was that Hh pathway activation promoted the outgrowth of hepatic progenitors that quickly differentiated and then proliferated. This concept is contradictory to current dogma (which posits that post-PH liver regeneration results from replication of mature hepatocytes in the liver remnant), but supported by several lines of evidence from our recent studies. First, cyclopamine treatment also inhibited cholangiocyte proliferation following PH, and both hepatocytes and cholangiocytes are the progeny of bi-potent hepatic progenitors. Second, populations of regenerating hepatocytes that were harvested 24h after PH were found to be enriched with cells that expressed progenitor markers. Third, PH resulted in a striking upregulation of progenitor-associated gene expression in whole liver tissue, causing, for example, a 40-fold induction of Fn14 (a tumor necrosis factor (TNF) superfamily receptor that is said to specifically mark bi-potent hepatic progenitors in liver) and more than 160-fold induction of alpha-fetoprotein (AFP), a well-established marker of immature liver epithelial cells. Fourth, immunohistochemistry for several other progenitor markers confirmed that PH stimulated accumulation of liver progenitors in regenerating livers. Fifth,

cyclopamine treatment significantly inhibited normal post-PH induction of FGF-inducible 14-kDa protein (Fn14), AFP, and several other progenitor markers in mice. Sixth, it also virtually eliminated progenitor accumulation following PH as assessed by immunohistochemistry. Similar approaches were used to prove that Hh pathway activation also plays a significant role in stromal cell responses that follow PH, including transient expansion of myofibroblast populations and hepatic matrix remodeling. Together, these data support the concept that normal regenerative responses in adult livers involve transient reactivation of mechanisms, such as Hh, which are likely to regulate liver morphogenesis during fetal development.

Sustained Hh Pathway Activation during Repair of Chronic Liver Injury

Healthy adult livers typically respond to acute resection of large amounts of their mass (e.g., PH) by regenerating successfully. Paradoxically, however, repetitive insults that merely accelerate the rate of death in subpopulations of liver cells (e.g., mature hepatocytes and/or cholangiocytes) often lead to scarring and hepatic architectural distortion (i.e., cirrhosis). Differences in the outcomes of PH and chronic liver injury suggest that entirely different repair mechanisms may be involved. Surprisingly, however, there is definite evidence of Hh pathway reactivation in all types of chronic liver injuries that have been examined to date in either rodents or humans, including injury caused by toxins (alcoholic- and obesity-related (nonalcoholic) fatty liver diseases), biliary tract damage (due to mechanical biliary obstruction, autoimmune disorders such as primary biliary cirrhosis and sclerosing cholangitis, or congenital factors, including biliary atresia, Alagille syndrome, and progressive familial intrahepatic cholestasis), and infections (hepatitis B and hepatitis C). Increased Hh pathway activity has also been demonstrated in many primary hepatocellular carcinomas and cholangiocarcinomas (Figure 3).

Recent research has demonstrated that many types of adult liver cells are capable of producing Hh ligands, including hepatocytes, cholangiocytes, hepatic progenitors, hepatic stellate cells/myofibroblasts, liver endothelial cell and certain types of liver lymphocytes (e.g., natural killer T cells NKT cells). Stimuli for Hh ligand production have been identified, and include growth regulatory cytokines (e.g., epidermal growth factor, platelet-derived growth factor, and transforming growth factor beta (TGF- β)) and pro-apoptotic stress (e.g., disruption of nuclear factor kappa beta signaling). The common downstream mediator of both types of ligand-inducing stimuli may be the phosphoinositide 3-kinase (PI3K)/AKT pathway, because direct adenovirally mediated manipulation of AKT activity alters Hh ligand production. It has also been demonstrated that ligand-producing liver cells release biologically active Hh ligands in membrane particles that have features of exosomes. Moreover, Hh-containing exosomes have been purified from the plasma and bile of rats with liver injury due to chronic biliary obstruction. Thus, although additional research is needed to clarify the molecular mechanisms controlling liver cell production and release of Hh ligands, current knowledge provides a plausible explanation that may

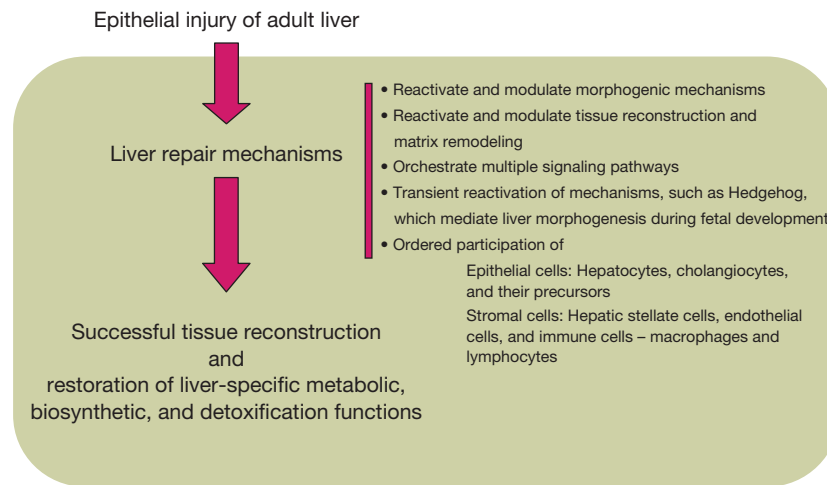


Figure 3 Reconstitution of the organ after liver injury or tissue loss involves replacement of target cells, and complex remodeling processes resulting in tissue architecture reconstruction. Hepatocytes are not terminally differentiated cells, but cells situated in the G_0 phase that can undergo proliferation upon appropriate stimulation. Hepatic stem cells and the other liver cells seem to be involved in this response. Hepatocyte proliferation is accomplished by a sequence of three phases: initiation, rendering cells in a state of replicative competence; proliferation, where expansion of the cell population occurs; and a termination phase, where cell growth is suppressed. Liver regeneration is regulated by ordered action of a plethora of transcription factors, growth factors, and cytokines, the significance of which has partly been defined by the use of animal models with target gene deletions. Hepatocyte regeneration is accompanied by a complex matrix remodeling, with a transient breakdown of the lobular architecture. Much less is known of the regenerative replacement of bile ducts, blood vessels, and hepatic stellate cells.

account for differences in the outcomes of acute PH and chronic liver injury. PH provides a temporally circumscribed stimulus for Hh ligand production, while the repetitive death of mature liver epithelial cells during chronic liver injury continuously drives Hh ligand production, resulting in sustained induction of Hh signaling in Hh-responsive cells.

Indeed, interrogation of chronically injured livers has revealed that many types of liver cells are capable of responding to Hh ligands, including immature hepatocytes, cholangiocytes, hepatic progenitors, myofibroblastic hepatic stellate cells, activated liver endothelial cells, and NKT cells. Hh ligands exert important effects on the fate of such Hh-responsive cells. For example, Hh pathway activation enhances the viability and growth of cholangiocytes, hepatic progenitors, myofibroblastic hepatic stellate cells, and NKT cells. It also activates liver sinusoidal endothelial cells and upregulates cholangiocyte production of chemokines for immune cells (monocyte/macrophages, fibrocytes, neutrophils, and B and T lymphocytes, including NKT cells). In cholangiocyte and hepatic stellate cell populations, induction of Hh signaling promotes cells to acquire a more mesenchymal, migratory phenotype. This process is mediated via Hh-dependent induction of several factors (e.g., bone morphogenetic protein-7, inhibitor of differentiation-2, and snail) that promote epithelial-to-mesenchymal transitions by repressing expression of E-cadherin and other junctional complex components. Hh pathway activation also stimulates NKT cells to express fibrogenic cytokines, such as interleukins 4 and 13. Thus, the net outcomes of protracted Hh signaling in chronically injured livers include virtually all of the responses that are known to occur during the pathogenesis of liver cirrhosis, including

fibrogenesis, inflammatory cell accumulation, vascular remodeling, and the outgrowth of liver progenitor populations that may harbor cancer stem cells for primary liver cancers.

The concept that both liver cirrhosis and carcinogenesis result from overactivation of conserved signaling pathways that are necessary for fetal liver development and adult liver regeneration is somewhat heretical, but supported by several lines of experimental evidence. First, both Hh pathway activation and increased fibrogenesis generally precede the emergence of cirrhosis and liver cancer by months to years in mice and humans with chronic liver injury. Second, the level of Hh pathway activity in injured livers tightly parallels the severity of liver fibrosis in rodents and humans, with the greatest accumulation of Hh-responsive cells occurring in cirrhosis. Third, cirrhosis is the single greatest risk factor for primary liver cancer in humans. Fourth, the capsules of human hepatocellular carcinomas are enriched with Hh-responsive cells that express markers of cancer stem/progenitor cells, and Hh-responsive cells are scattered throughout the stroma of primary liver cancers. Fifth, Hh signaling antagonists inhibit the growth of cultured liver myofibroblasts and liver cancer cell lines. Sixth, therapeutic interventions that successfully interrupt liver injury are accompanied by progressive downregulation of Hh signaling and regression of liver fibrosis, with eventual silencing of Hh pathway activity once healthy liver architecture has been recovered. Seventh, transgenic mice that are unable to efficiently silence Hh signaling after Hh pathway activation exhibit excessive liver fibrosis and expansion of hepatic progenitor populations when subjected to various liver injuries. These mice are also susceptible to spontaneous development of Hh-responsive malignancies. Work is being done in animal

models to determine if mice with ongoing liver injury survive long-term blockade of Hh signaling and, if so, whether or not this reduces subsequent liver fibrosis and cancer formation.

Summary

Recent evidence that transient activation of the Hh pathway is critical for normal liver regeneration and survival after PH complements other data which suggest that prolonged Hh pathway activation promotes cirrhosis and neoplasia. Together, these findings support the hypothesis that successful (as opposed to dysmorphic/fibrotic) repair of adult liver injury recapitulates aspects of fetal liver development, and is likely to be determined by the balance of factors that promote induction versus silencing of Hh signaling. Indeed, in non-liver cell types, it has already been shown that the Hh pathway interacts with other morphogenic signaling mechanisms that are known to regulate liver development, growth, and repair, including Wnt and TGF- β . Moreover, significant roles for those other pathways in the pathogenesis of liver cancer and cirrhosis are widely accepted: excessive Wnt signaling is thought to promote hepatocarcinogenesis and dysregulated TGF- β signaling is believed to provide a major stimulus for liver cirrhosis. Therefore, further research is needed to clarify cell-autonomous and paracrine mechanisms that modulate the activities and interactions of developmental morphogens in adult livers. The resultant knowledge will likely advance the prevention and treatment of cirrhosis and liver cancer, two of the most dreaded consequences of chronic liver injury. In addition, it may help to explain why healthy adult livers retain the capacity for efficient regeneration, but other vital organs do not.

See also: **Cell Architecture and Function:** Cell Cycle: Control of Entry and Progression through S Phase; Cell Cycle: Mitotic Checkpoint; Cell-Cycle Controls in G₁ and G₀; Mitosis; **Lipids** **Carbohydrates** **Membranes and Membrane Proteins:** Cell–Matrix Interactions; **Metabolism Vitamins and Hormones:** Bile Salts; Cholesterol Synthesis and Regulation; Fatty Acid Oxidation; Fatty Acid Structure and Synthesis; Primary Biliary Cirrhosis; Urea Cycle: Disease Aspects.

Further Reading

- Acloque H, Adams MS, Fishwick K, Bronner-Fraser M, and Nieto MA (2009) Epithelial-mesenchymal transitions: The importance of changing cell state in development and disease. *Journal of Clinical Investigation* 119: 1438–1439.
- Duncan AW, Dorrell C, and Grompe M (2009) Stem cells and liver regeneration. *Gastroenterology* 137: 466–481.
- Fausto N, Campbell JS, and Riehle KJ (2006) Liver regeneration. *Hepatology* 43: S45–S53.
- Hirose Y, Itoh T, and Miyajima A (2009) Hedgehog signal activation coordinates proliferation and differentiation of fetal liver progenitor cells. *Experimental Cell Research* 315: 2648–2657.
- Jiang J and Hui CC (2008) Hedgehog signaling in development and cancer. *Developmental Cell* 15: 801–812.
- Michalopoulos GK (2010) Liver regeneration after partial hepatectomy: Critical analysis of mechanistic dilemmas. *American Journal of Pathology* 176: 2–13.
- Minguez B, Tovar V, Chiang D, Villanueva A, and Llovet JM (2009) Pathogenesis of hepatocellular carcinoma and molecular therapies. *Current Opinions in Gastroenterology* 25: 186–194.
- Ochoa B, Syn WK, Delgado I, et al. (2010) Hedgehog signaling is critical for normal liver regeneration after partial hepatectomy in mice. *Hepatology* 51(5): 1712–1723.
- Omenetti A and Diehl AM (2008) The adventures of sonic hedgehog in development and repair. II. Sonic hedgehog and liver development, inflammation, and cancer. *American Journal of Physiology: Gastrointestinal and Liver Physiology* 294: G595–G598.
- Varjosalo M and Taipale J (2008) Hedgehog: Functions and mechanisms. *Genes and Development* 22: 2454–2472.

Biochemistry of Neurogenesis

K Christian, H Song, and G-L Ming, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Adult neural stem cells Self-renewing multipotent precursor cells that can generate new neurons, astrocytes, and oligodendrocytes of the adult nervous system.

Dentate gyrus A subregion of the hippocampal formation involved in learning and memory and one of the few sites of neurogenesis in the adult mammalian brain.

Neurogenesis Process of generating functional neurons from stem/progenitor cells. It includes individually

regulated stages of proliferation, fate specification, migration, morphogenesis, and synaptic integration.

Neurogenic niche Microenvironment surrounding stem cells and their progeny which contains a host of extrinsic factors regulating adult neurogenesis.

Subgranular zone Proliferative region that lies adjacent to the granule cell layer in the dentate gyrus and is the locus of neural stem cells and progenitors that give rise to adult born dentate granule neurons.

Introduction

Neurogenesis describes the multistage cellular developmental process that begins with the first step toward proliferation of multipotent neural stem cells and culminates in the integration of functional new neurons into the local circuitry. Historically considered to be a singular series of events that occurs exclusively during embryonic development, cumulative evidence over the past several decades has confirmed that neurogenesis is a constitutive process in the mammalian brain. New neurons are born and incorporated into the existing neuronal framework throughout the life span of all mammals examined. Although neurodevelopmental processes are recapitulated in the mature brain, there are important differences between embryonic and adult neurogenesis. These differences can be broadly categorized along two dimensions – spatial and temporal. At the systems level, adult neurogenesis is spatially restricted and observed primarily in discrete regions of the central nervous system – the subgranular zone (SGZ), giving rise to glutamatergic neurons in the dentate gyrus of the hippocampus and the subventricular zone (SVZ), giving rise to at least two types of interneurons and a small percentage of glutamatergic neurons in the olfactory bulb. At the cellular level, it is informative to delineate between cell-autonomous and non-cell-autonomous factors that regulate neurodevelopment. Local microenvironments, or the neurogenic niche, can significantly impact neurogenesis, both quantitatively and qualitatively. Environmental conditions vary profoundly from embryogenesis to the adult brain. In the adult brain, essentially all steps of neurogenesis are impacted by diffusible and activity-driven signaling molecules operating within a functionally mature network. Temporally, adult neurogenesis is protracted compared to embryonic neurogenesis – individual developmental processes take longer in the adult brain. Overarching temporal factors also play a role in distinguishing embryonic and adult neurogenesis in that expression levels of functionally relevant proteins are often dynamically regulated throughout life.

Neurodevelopment occurs over the lifetime of an individual organism, first in a global process that generates numerous

cell types comprising the mammalian nervous system, and then in a local and prolonged process that generates specific subsets of neurons. Biochemical mediators of neurogenesis are thus spatially and temporally regulated. In addition, each process that contributes to the trajectory of neurodevelopment is governed by its own biochemistry. In this article, we focus on the major biochemical pathways that are involved in different stages of neurogenesis, describing the influential factors within key functional groups broadly divided between extrinsic and intrinsic factors. The spatial restriction and temporal expansion of adult neurogenesis make it an ideal model system to investigate neurodevelopment as a whole. Here we describe the current knowledge regarding the molecular mechanisms underlying postnatal neurogenesis in the dentate gyrus of the hippocampus.

Stages of Neurogenesis

Neurogenesis consists of sequential stages that define the progress of newly born neurons from neural stem cells to synaptic integration (Figure 1). Using adult hippocampal neurogenesis as our prototype, we will discuss biochemical mediators with respect to five stages of neurogenesis. (1) Proliferation of putative neural stem cells that reside in the SGZ of the hippocampus giving rise to transient amplifying cells. The primary neural stem/progenitor cells are radial glial-like cells that express nestin, glial fibrillary acidic protein (GFAP), and sex-determining region Y (SRY)-box 2 (Sox2). (2) Differentiation of highly proliferative progenitors into immature neurons. Immunohistochemical markers distinguish at least four cell types during this differentiation stage based on sequential and overlapping expression of nestin, T-brain gene-2 (Tbr-2), and doublecortin (DCX). All are GFAP negative and the first two subtypes, chronologically, express nestin. The intermediate subtype expresses nestin and Tbr2, while the late-stage, proliferative neuroblasts that are poised to migrate are nestin negative, Tbr2 negative, and positive for both DCX and the polysialylated embryonic form of neural cell adhesion

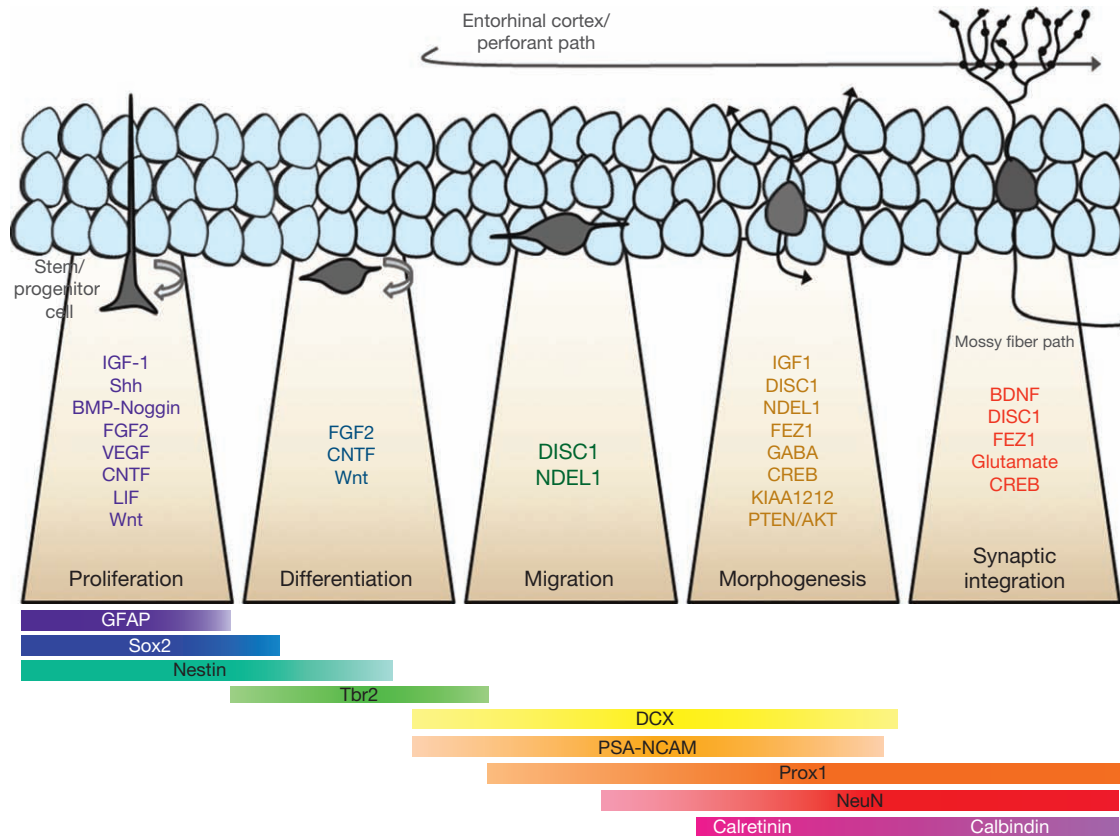


Figure 1 Stages of neurogenesis in the adult dentate gyrus. Simplified schematic of the trajectory of neurodevelopment from the adult neural stem cell to functional integration within the hippocampal circuitry. Below are the stage-specific markers that can be used to birth date and identify newly born neurons with the hippocampus.

molecule (PSA-NCAM). (3) Migration of immature, but post-mitotic neurons, expressing dentate granule cell marker Prox-1 to the granule cell layer of the dentate gyrus. (4) Axonal and dendritic morphogenesis of immature neurons. At this stage, neurons begin to express calretinin and the neuron-specific neuronal marker (NeuN). (5) Synaptic integration of newly born neurons. Mature granule cells are marked by the expression of calbindin rather than calretinin and the complex structure of dendritic and axonal processes are established within the hippocampal network in this final stage of neurogenesis.

In addition to the stage-specific markers described above, thymidine analogs such as bromodeoxyuridine (BrdU) can be incorporated into dividing cells during the S phase of the cell cycle, thereby establishing a means of birth dating newly born neurons through immunohistochemical analysis. Similarly, Ki67 marks all active phases of the cell cycle and histological quantification of this protein can timestamp levels of proliferation in the postnatal brain. All of these markers can be used individually or in combination to delineate the temporal phases of neurogenesis and track the progress of neural development from initial division of neural stem cells to the stable integration and functional viability of adult-born neurons. Using these markers in conjunction with targeted genetic disruption and manipulation of endogenous protein levels, much progress has been made recently in identifying the critical factors that regulate different aspects of adult neurogenesis.

Extrinsic Factors

Developing neurons are highly influenced by the local environment, which is itself subject to a wide variety of dynamic factors that reflect both cyclic variation and acute perturbations in the systemic balance of the organism. Many exogenous factors influence the rate of neurogenesis and the survival and integration of newly born neurons. Generally, there is a profound correspondence between factors that promote neurogenesis and those that promote the overall health of the organism as a whole. Exercise, sleep, sex, environmental enrichment, and learning are all conducive to neurogenesis. Conversely, stress, drugs of abuse, aging, and depression decrease neurogenesis in the adult brain. Contextual factors, such as genetic background and sexually divergent response patterns, add to the complexity of the dynamic neurogenic milieu. For instance, stress-induced glucocorticoids have a sexually dimorphic effect on hippocampal neurogenesis wherein increased levels of corticosterone in male rodents decrease cell proliferation and survival, but have minimal effect in females.

Understanding how these factors work in a synergistic or competitive manner at the molecular level to regulate neuronal development in the adult brain is an important focus of current research. Little is currently known about the relative strength of extrinsic modulators and whether there is a hierarchical relationship between competing influences. For example, focal

trauma may exert a dominant effect on neurogenesis but in the normal environment, it is not well understood how dynamically regulated factors interact to determine the neurogenic capacity of the local niche – for example, the net effect of temporally coincident exercise (a positive regulator) and elevated stress levels (a negative regulator). In addition, how extrinsic signals are integrated within the receiving cell is less well understood both at the transduction cross-talk and epigenetic levels. Nonetheless, an important first step in understanding the complex environment has been the identification of key components that have a consistent effect on specific stages of neurogenesis, which are described below along with known interactions.

Growth Factors, Cytokines, and Neurotrophins

Given the canonical role of neurotrophic factors in regulating cellular development, it is not surprising that many of the secreted proteins in this class have a demonstrated impact on multiple neurogenic processes. Predominantly, growth-related factors appear to exert influence over the early stages of neurogenesis, specifically serving to enhance stem cell self-renewal, as well as proliferation and differentiation of neural progenitors. We will review in brief some of the most extensively studied factors in this group that have established roles in neurogenesis.

Insulin-like growth factor 1 (IGF1) acts through at least two pathways to stimulate proliferation, neuronal differentiation, morphogenic development, and survival of newly born cells. Primarily synthesized in the liver, serum IGF1 also operates in a paracrine/autocrine manner in the brain and indirect actions of IGF1 include promotion of angiogenesis, which enriches the neurogenic niche. Dose-dependent proliferative effects of IGF1 on cultured hippocampal neural progenitors require activation of the mitogen-activated protein kinase (MAPK) pathway. Ras/Raf/MAPK activation may itself be initiated through distinct mechanisms mediated by the IGF1 receptor (IGF1R) including Son of Sevenless (SOS) or G-protein activation, culminating in transcriptional regulation through nuclear translocation of phosphorylated extracellular signal-regulated kinase (ERK)-1. Operating through the phosphoinositide 3-kinase (PI3K)-AKT (also known as protein kinase B) pathway, IGF1 signaling mediates the exercise-induced increase in survival of newly born neurons via promotion of anabolic activity and glucose utilization, and the inhibition of glycogen synthase kinase 3 β (GSK3 β).

In cultures of hippocampal neural stem cells, fibroblast growth factor (FGF-2) has been shown to regulate all aspects of self-renewal. Acting through the ERK1/2 pathway, FGF-2 enhances proliferation and suppresses spontaneous differentiation. Through a separate mechanism, FGF-2 works to preserve the multipotency that is characteristic of neural stem cells via activation of the phospholipase C gamma 1 (PLC γ 1) pathway. Knockdown of PLC γ 1 function predisposes stem cells toward differentiation of glial phenotypes suggesting an endogenous role of FGF-2-mediated signaling in preserving the possibility of neuronal differentiation *in vivo*. *In vivo* studies have demonstrated a reduction in granule cell population numbers in FGF2-deficient mice and an increase in hippocampal cell number following exogenous administration of FGF-2.

A critical determinant of the qualitative impact of the extrinsic neurogenic environment is the proximal vasculature. Vascular endothelial growth factor (VEGF) is a potent regulator of angiogenesis but also has a direct proliferative effect on progenitors and may also be involved in survival of newborn cells directly. While VEGF is minimally important in maintaining basal levels of adult neurogenesis, it plays an essential role in the enhanced neurogenesis observed following exposure of adult rodents to an enriched environment. It is also critically involved in the increased proliferation of neurons in the SGZ following antidepressant treatment. Flk-1, a VEGF receptor tyrosine kinase, is expressed in both neural progenitors and epithelial cells in the hippocampus and is the likely substrate for the initiation of the VEGF signaling pathway, although the downstream effectors remain largely unknown.

Brain-derived neurotrophic factor (BDNF) exerts its role in later phases of adult neurogenesis. It has a clearly established role in mediating synaptic plasticity and has been implicated in numerous higher-order processes, including learning and memory. In the hippocampus, BDNF may promote the selective stabilization of synapses recruited through activity-dependent mechanisms that reflect the experience of the organism. Experiments with hippocampus-specific deletion of BDNF suggest that this factor primarily affects neuronal survival in basal conditions, with little effect on proliferation or neuronal differentiation. However, intact BDNF signaling appears to partly mediate exercise and environmental enrichment effects on proliferation and stimulation of dendritic development. Downstream pathways that mediate the effects of BDNF through its cognate receptor TrkB include the MAPK/ERK pathway and the PI3K pathway, demonstrating a mechanistic convergence with other neurotrophic factors.

The cytokines, ciliary neurotrophic factor (CNTF) and leukemia inhibitory factor (LIF), are associated with common signal-transducing components, gp130 and LIF receptor β (LIFR β), and were originally believed to have similar effects in enhancing proliferation of progenitor cells in the SGZ. Recent data using knockout mice have shown that a loss of function of CNTF results in a decrease in the available pool of neural stem cells and that the mechanism of action appears to be mediated through the janus kinase (JAK)/signal transducer and activator of transcription 3 (STAT3) pathway. Impaired CNTF function mostly affects proliferation of primary progenitor/stem cell and also leads to the reduction of the population of newly born neurons. Unlike FGF-2, CNTF biases differentiation toward glial phenotypes. Functionally, CNTF appears to have a similar effect on proliferation in the SVZ but likely has a minimal effect during embryonic neurodevelopment and indeed, expression of this cytokine is restricted to the postnatal brain. LIF knockout mice did not display a phenotype in analysis of stem cell self-renewal, proliferation, or number of newly born neurons. Evidence instead suggests that this cytokine may be preferentially involved in the enhancement of neurogenesis in response to injury or ischemia, particularly during the perinatal period, when there is an absence of CNTF expression.

Bone morphogenic proteins (BMPs) inhibit the proliferation of neural stem cells and their action is antagonized by the endogenous expression of Noggin. Overexpression of Noggin

in transgenic mice increases the number of GFAP-positive neural stem cells, indicating that a homeostatic balance in the proliferative capacity of the neural progenitors in the hippocampus is at least partially maintained through a BMP–Noggin interaction. BMP signaling is also decreased by exercise in mice and is another putative mechanism for exercise-induced enhancement of neurogenesis. These actions are likely mediated by Smad signaling, but MAPK inhibition may also be involved.

In the adult brain, the embryonic morphogen sonic hedgehog (Shh) has a sustained influence on the proliferation of neuronal progenitors. Patched (Ptc), the Shh receptor, is present in granule cells and neural progenitors in the adult dentate gyrus. Transcriptional regulation of Shh targets is at least partly mediated through the release of Ptc inhibition of its constitutively associated transmembrane protein, Smoothened (Smo), which in turn activates members of the Gli (glioma-associated oncogene) family of transcription factors.

Wingless-type MMTV integration site-3 (Wnt3) is expressed in the neurogenic niche of the adult hippocampus and regulates proliferation and neuronal differentiation. The canonical Wnt pathway involves inhibition of GSK3 β and a subsequent increase of cytoplasmic β -catenin which translocates to the nucleus to regulate transcription of target genes such as NeuroD1. Noncanonical pathways operate through the direct activation of Jun-N-terminal kinase (JNK), or protein kinase C (PKC) and calcium-calmodulin-dependent protein kinase (CaMK) II, via a rise in intracellular calcium. Secreted frizzled-related protein 3 (sFRP3), an endogenous inhibitor of Wnt, is highly expressed in the dentate gyrus of the adult hippocampus and may regulate adult neurogenesis.

Neurotransmitters

A classical inhibitory neurotransmitter, gamma aminobutyric acid (GABA), actually depolarizes newly born neurons, as opposed to the hyperpolarizing role it plays in mature neurons. Mechanistically, the effect of GABA signaling is, at least in part, determined by the relative expression of Na–K–Cl cotransporter isoform 1 (NKCC1) and K–Cl cotransporter 2 (KCC2) transporters. These transporters determine levels of intracellular chloride and are dynamically regulated over the course of neuronal maturation. Current evidence suggests that the transcription factor, cAMP response element-binding protein (CREB), is a downstream effector of depolarizing GABA. Glutamatergic effects on neurogenesis appear to be highly dynamic from early proliferative stages to terminal synaptic integration and depend critically on the expression of receptor subtypes during the various stages. Glutamate signaling plays a critical role in the activity-dependent promotion of cell survival and synaptic development, an effect that could be mediated in part through a glutamatergic-mediated enhancement of BDNF. Other neurotransmitters have been implicated in select stages of neurogenesis, although there is much left to learn about the mechanisms which determine the developmental response of neurons to these signaling molecules. For example, norepinephrine- β_3 receptor signaling appears to activate precursor and stem cells in the SGZ. Dopaminergic signaling appears to have an opposite effect in that depletion of dopamine inputs to the hippocampus activates stem/progenitor cells.

Intrinsic Factors

Ultimately, most extrinsic modulators of neurogenesis rely on intracellular signaling pathways, but there are also specific proteins and molecular complexes that have been shown to have a cell-autonomous role in regulating neurodevelopment. Protein expression levels that are intrinsically regulated can determine the capacity of the developing neuron to survive and integrate into the existing circuitry. In addition, the ability of a proliferating neuronal precursor or newly born neuron to respond to extrinsic cues can be influenced by the genetically determined integrity of key signaling pathways. Here we review the most prominent and highly convergent pathways that regulate neurogenesis, selectively illustrated in [Figure 2](#).

Scaffold Proteins and Cell-Autonomous Signaling Complexes

One of the most prominent scaffold proteins implicated in neurogenesis is disrupted-in-schizophrenia 1 (DISC1). Several recent studies have shown that DISC1 is at the core of a multifunctional signaling complex, which affects nearly every stage of neurogenesis in the adult brain. Through direct interaction, DISC1 inhibits GSK3 β , which in turn stabilizes β -catenin via suppression of phosphorylation. GSK3 β activation and β -catenin stabilization inhibit proliferation of neural progenitors. Therefore, DISC1 effectively acts to disinhibit this negative regulation of proliferation in the adult dentate gyrus. Loss of function of the DISC1 protein results in decreased proliferation through modulation of cell cycle exit, which can be rescued by targeted inhibition of GSK3 β . Emerging evidence suggests that activation of the GSK3 β / β -catenin is a highly convergent pathway mediating the downstream effects of several extrinsic modulators of neurogenesis, including the neurotrophic factors IGF1, VEGF, FGF, Wnt3, as well as circulating glucocorticoids. In adult neurogenesis, retroviral-mediated suppression of DISC1 confers several distinct phenotypes of accelerated growth and overextended migration, suggesting that the endogenous role of DISC1 is to constrain intermediate and late-stage neurodevelopment. A central mediator of the DISC1 effect on postnatal neurogenesis is the AKT-mammalian target of rapamycin (mTOR) signaling pathway. DISC1 directly inhibits KIAA1212 (also known as Girdin), an activator of the AKT-mTOR pathway. As the investigation has progressed into the mechanisms that underlie these phenotypes, it has become clear that two additional signaling pathways contribute to specific facets of this developmental hypertrophy. Migratory defects and the appearance of ectopic dendrites are regulated by a direct interaction between DISC1 and nude nuclear distribution gene E homolog-like 1 (NDEL1). Somatic hypertrophy and excessive arborization of dendrites are the result of a fasciculation and elongation protein zeta-1 (FEZ1)–DISC1 interaction. Downstream mediators of these complementary DISC1-dependent pathways, however, have yet to be described in full.

Transcriptional Regulation and Epigenetic Factors

Transcription factors can control expression of multiple genes and proteins and thus may exert a pervasive effect on intracellular signaling cascades and the neurogenic niche. One of the most

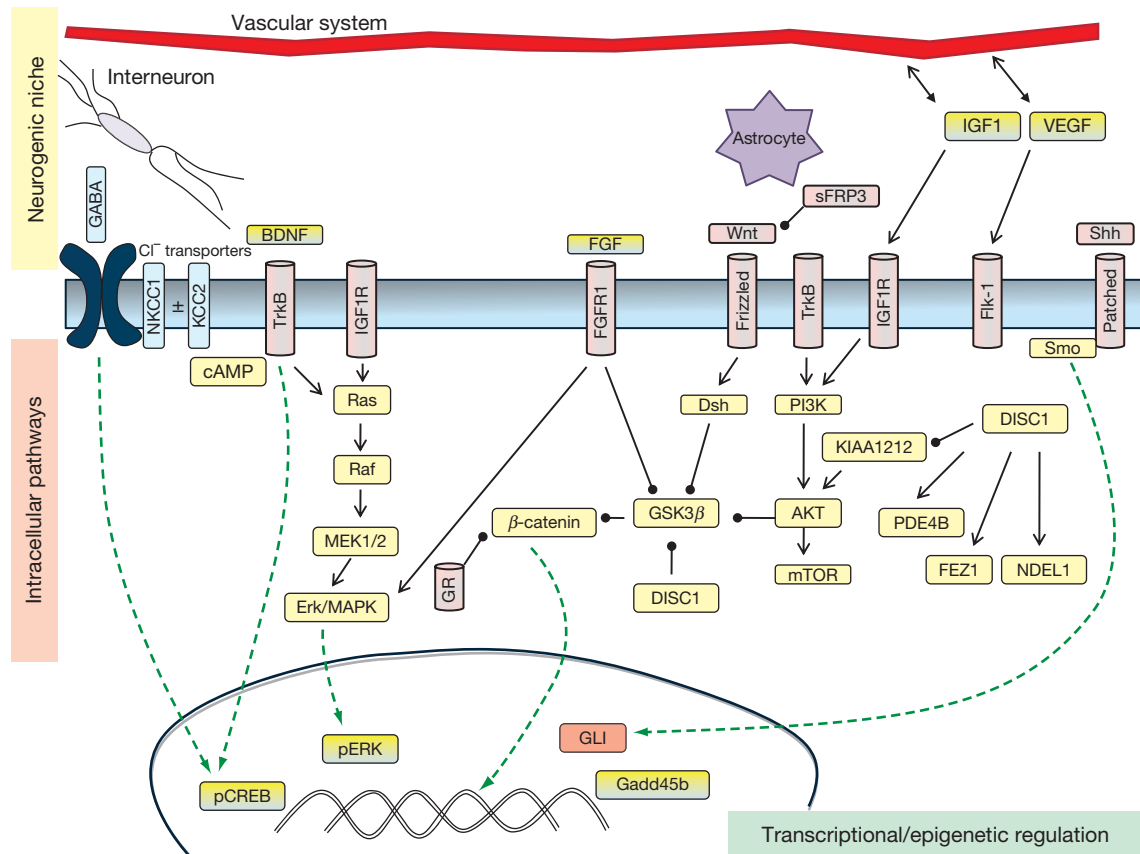


Figure 2 Extrinsic and intrinsic factors regulating adult neurogenesis in the dentate gyrus. Highly simplified schematic illustrating some of the convergent pathways and prominent molecular regulators of neurodevelopment – target transcriptional and epigenetic modulators are shown in the lower center of the diagram.

prominent regulators of transcription to be implicated in neurogenesis is CREB. Although recent data have called into question the proliferative effects of CREB activation in adult neurogenesis, the phosphorylated form of CREB clearly supports the survival and morphological development of newborn neurons. In another example of the integral interactions between many of these neurogenic factors, phosphorylation of CREB may be a crucial effector signal in transducing the GABA-mediated depolarizing driving force on neurodevelopment.

Recent data have suggested a novel mechanism underlying how transient activity can generate long-lasting effects on neurogenesis. In the adult hippocampus of mice, growth arrest and DNA damage inducible 45-beta (Gadd45b) can be induced by synchronous activation of local networks following electroconvulsive treatment (ECT), as well as by physiological patterns of activation following increased exercise on a running wheel. Gadd45b, in turn, was shown to be required for region-specific DNA methylation following ECT. Intriguingly, activity-induced, Gadd45b-dependent, targeted DNA methylation was observed in the regulatory regions of the established modulators of neurogenesis, *Bdnf* and *Fgf1b*. Knockout of Gadd45b markedly reduces activity-induced increase in progenitor proliferation and dendritic growth. These results suggest a sequence of events wherein transient activation of cells in the neurogenic microenvironment can lead to an epigenetic modification cascade resulting in the persistent secretion of specific niche factors.

Cell-Type Specific Factors

In the adult brain, neurogenesis is largely restricted to a few neuronal and glial populations within the hippocampus and olfactory bulb. During embryonic neurogenesis, however, the entire spectrum of neuronal phenotypes must be generated. Distinct neuronal populations are identified by a unique combination of factors, including morphology, structural integration within local circuitry, and chemical and electrical transmission properties. Very little is currently known about the signaling pathways that underlie differentiation of specific neuronal types during adult neurogenesis. In order to fully realize the potential of stem cell therapy in the treatment of neurodegenerative disease, it is clear that much work needs to be done to understand the molecular mechanisms responsible for the differentiation and growth of specific neuronal populations.

Conclusion

Despite the complexity of the numerous signaling pathways and interacting biochemical mediators outlined above, this is by no means an exhaustive account of the molecular regulation of neurogenesis. We have focused on the most influential functional groups at present, highlighting only the most prominent factors in each. However, there are likely many key

cellular and molecular processes that remain to be identified and many explanatory gaps yet to fill, such as, signal integration hierarchy. Importantly, there may be characteristic modulators of specific neurogenic stages that operate in either, or both, a cell-autonomous or nonautonomous fashion and that a functional categorization of critical factors may expedite our progress toward a more complete understanding of neurogenesis. As with any investigation of a complexly regulated process, the goal should be to develop a coherent and simplified working model that accounts for the current data and provides a comprehensive framework in which to interpret new findings and generate novel hypotheses. With neurogenesis, the potential clinical benefits that could arise from such a model are profound. Increasing our understanding of the mechanisms underlying this constitutive phenomenon represents one of the most promising avenues toward exploiting the potential of stem cells to prevent and reverse the devastating consequences of neural pathology arising from disease and injury.

See also: **Cell Architecture and Function:** Cell-Cycle Controls in G_1 and G_0 ; Cell Cycle: Control of Entry and Progression through S Phase; **Signaling:** Fibroblast Growth Factor Receptors and Cancer-Associated Perturbations; Mitogen-Activated Protein Kinase Family;

mTOR and its Downstream Targets; Neurotrophin Receptor Signaling; Vascular Endothelial Growth Factor Receptors.

Further Reading

- Altman J and Das GD (1965) Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. *Journal of Comparative Neurology* 124(3): 319–335.
- Alvarez-Buylla A and Lim DA (2004) For the long run: Maintaining germinal niches in the adult brain. *Neuron* 41(5): 683–686.
- Balu DT and Lucki I (2009) Adult hippocampal neurogenesis: Regulation, functional implications, and contribution to disease pathology. *Neuroscience and Biobehavioral Reviews* 33(3): 232–252.
- Eriksson PS, Perfilieva E, Björk-Eriksson T, et al. (1998) Neurogenesis in the adult human hippocampus. *Nature Medicine* 4(11): 1313–1317.
- Kempermann G, Jessberger S, Steiner B, and Kronenberg G (2004) Milestones of neuronal development in the adult hippocampus. *Trends in Neurosciences* 27(8): 447–452.
- Lledo PM, Alonso M, and Grubb MS (2006) Adult neurogenesis and functional plasticity in neuronal circuits. *Nature Reviews Neuroscience* 7(3): 179–193.
- Ma DK, Bonaguidi MA, Ming G-L, and Song H (2009) Adult neural stem cells in the mammalian central nervous system. *Cell Research* 19(6): 672–682.
- Ming GL and Song H (2005) Adult neurogenesis in the mammalian central nervous system. *Annual Review of Neuroscience* 28: 223–250.
- Schmidt HD and Duman RS (2007) The role of neurotrophic factors in adult hippocampal neurogenesis, antidepressant treatments and animal models of depressive-like behavior. *Behavioural Pharmacology* 18(5–6): 391–418.
- Zhao C, Deng W, and Gage FH (2008) Mechanisms and functional implications of adult neurogenesis. *Cell* 132(4): 645–660.

Biochemistry of Thiamine and Thiamine Phosphate Compounds

L Bettendorff and P Wins, University of Liège, Liège, Belgium

© 2013 Elsevier Inc. All rights reserved.

Glossary

Ataxia A neurological condition characterized by severe lack of movement coordination. It is a nonspecific manifestation due to lesions in the brain regions involved in movement coordination such as the cerebellum and the vestibular system.

Blood–brain barrier (BBB) The layers of epithelial cells separating the brain parenchyma and the cerebrospinal fluid from the blood. The BBB results from tight junctions between endothelial cells forming the walls of the capillaries. Therefore, only molecules having specific carriers (glucose, vitamins, and salts) or lipid-soluble molecules (O_2 , CO_2 , alcohol, psychoactive drugs, etc.) may cross this barrier. The BBB is essential for the protection and homeostasis of neurons and glia.

Encephalopathy A global dysfunction of the brain and altered mental state. Encephalopathies can have many different causes such as metabolic stress (due to inhibition of enzymes necessary for energy metabolism, mitochondrial dysfunction, or liver failure), infectious agents (prions, bacteria, and viruses), and toxins (drugs and industrial chemicals).

Excitotoxicity A pathological process during which receptors for excitatory amino acid neurotransmitters, in particular the *N*-methyl-D-aspartic acid (NMDA)-type glutamate receptors, are stimulated above normal levels,

leading to increased entry of Ca^{2+} into neurons. Overstimulation of various Ca^{2+} -dependent processes may then lead to neuronal death. This occurs in ischemia, neurodegenerative diseases such as Alzheimer's and Parkinson's diseases, and Wernicke–Korsakoff syndrome.

Nutritional downshift When bacteria are transferred from a rich medium (optimal for growth) to a medium lacking one or several nutrient categories (phosphates, nitrogen, carbon, amino acids, etc.), they respond by producing signaling molecules (alarmones) redirecting the cell metabolism to cope with the induced starvation stress.

Polyneuritis or peripheral neuropathy A damage of the nerves of the peripheral nervous system caused by trauma (compression), metabolic illness (vitamin deficiencies and diabetes mellitus), genetic illness (Friedreich's ataxia and Charcot–Marie–Tooth syndrome), toxins (heavy metals), or inflammatory disease (Guillain–Barré syndrome). This generally translates into weakness of the limbs (legs and feet) and sensory modifications.

Ylide (or ylid) A molecule wearing opposite charges on adjacent atoms. The negative charge is always on a carbon and it is therefore a particular kind of carbanion. Ylides often appear as reactive intermediates in organic reactions.

Introduction

Thiamine (thiamin, vitamin B1), also formerly called aneurin (anti-neuritic factor), is a water-soluble vitamin of the B group. It is an indispensable molecule for intermediate metabolism in all life forms. This is mainly because its diphosphorylated derivative (thiamine pyrophosphate or diphosphate (ThDP); see Figure 1) is required as a cofactor for central metabolic reactions such as the oxidative decarboxylation of pyruvate or oxoglutarate – key steps in energy metabolism. Thiamine is synthesized in microorganisms and plants; however, animals, including humans, have to rely on exogenous dietary sources.

As can be expected from the fundamental role of ThDP in metabolism, thiamine deficiency in animals has very harmful consequences, especially on the nervous system. This is, in part, because nerve cells heavily rely on glucose oxidation for their energy needs, and also because thiamine absorption is rather slow in these cells.

In humans, chronic nutritional thiamine deficiency is responsible for beriberi, a disease causing polyneuritis and paralysis. Beriberi was a major health problem in East Asian countries until the beginning of the twentieth century, because the major food source there was polished rice, which contains

very little thiamine. Nowadays, the most common disease linked to thiamine deficiency is the Wernicke–Korsakoff syndrome, an encephalopathy generally associated with chronic alcoholism. In developed countries, Korsakoff's psychosis not only remains the third cause of dementia, but subclinical thiamine deficiency is probably more common than generally thought, especially in elderly people and patients with diabetes or acquired immune deficiency syndrome (AIDS).

Chemical Properties of Thiamine

The structure of thiamine (Figure 1) is quite unusual as it consists of a pyrimidine moiety (2-methyl-4-amino-pyrimidine) and a thiazolium moiety [5-(2-hydroxyethyl)-4-methylthiazolium] connected through a methylene bridge. The aromatic heterocycle thiazole is rarely found in biological molecules. In cells, the hydroxyethyl group is generally esterified to form phosphorylated derivatives, the most widespread being the cofactor ThDP.

Thiamine and its phosphorylated derivatives are very hydrophilic molecules: 1 g of thiamine chloride hydrochloride ($C_{12}H_{17}ClN_4OS \cdot HCl$, molecular mass $337.27 \text{ g mol}^{-1}$) dissolves in $\sim 1 \text{ ml}$ of water. Thiamine is poorly soluble in organic solvents, including ethanol. It is heat-labile and

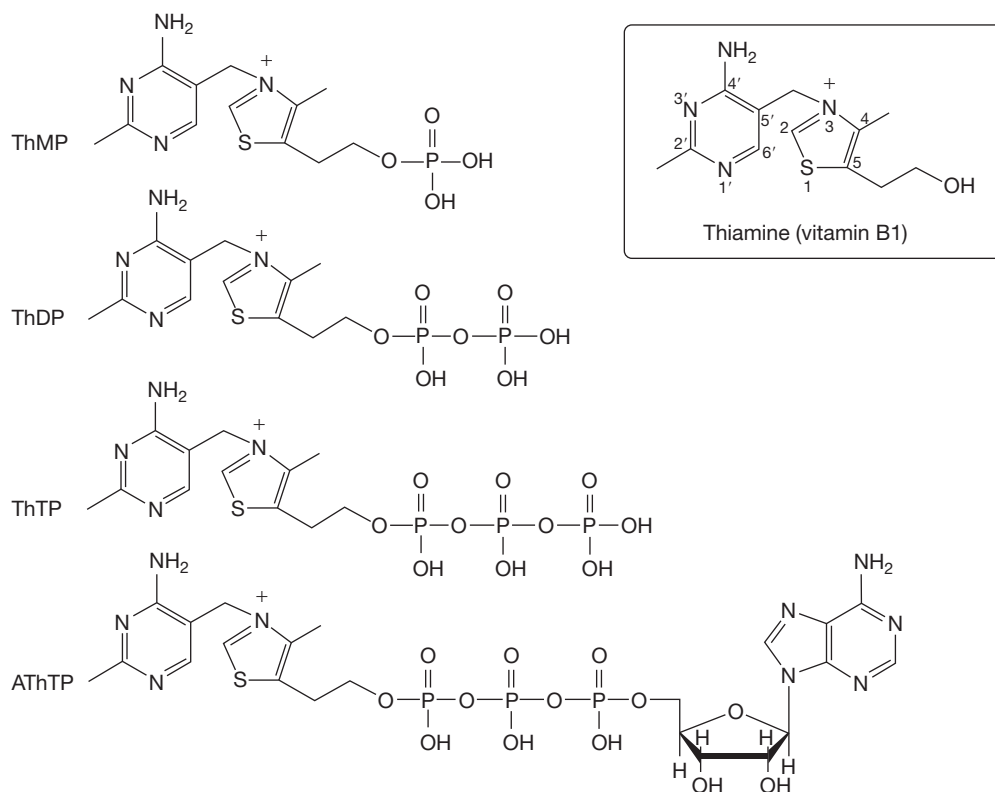


Figure 1 Natural thiamine derivatives. AthTP, adenosine thiamine triphosphate; ThDP, thiamine diphosphate; ThMP, thiamine monophosphate; ThTP, thiamine triphosphate.

partially lost by cooking. Aqueous solutions are most stable between pH 2 and 4. At pH < 4, thiamine is dicationic, as the N-1' of the pyrimidine ring (but not the -NH_2) is protonated. In alkaline solution, thiamine is unstable: in the pH range 8–10, hydroxyl attack on C-2 of the thiazolium ring leads to the relatively slow formation of a pseudobase intermediate (compound II in Figure 2) and the subsequent fast opening of the thiazolium ring. This yields a thiolate form (III) that can lose a water molecule to form the so-called yellow form (IV).

Note that, even in relatively strong alkaline medium, the ylide carbanion (which plays a key role in catalysis by ThDP-dependent enzymes, see below) is never observed in aqueous solutions of thiamine, because the estimated pK_a of the CH in question is too high (in the range 14–20). Nonetheless, a relative lability of the C-2 proton is demonstrated by the observation that it can be rapidly exchanged with deuterium.

Thiamine Biosynthesis and Riboswitches

Thiamine biosynthesis occurs in prokaryotes, fungi, and plants through complex pathways. The thiazole and pyrimidine moieties are synthesized separately before assembling to form thiamine monophosphate (ThMP), which is either hydrolyzed to thiamine or phosphorylated to ThDP.

Thiamine-degrading enzymes (thiaminases) exist in many organisms. Thiaminase I (EC 2.5.1.2), a pyrimidine transferase, is present not only in microorganisms, but also in some higher multicellular eukaryotes such as fern, shellfish, and fish. Consumption of raw parts of these organisms by animals or man can lead to serious thiamine deficiency syndromes, such as cortical necrosis in ruminants and beriberi in humans. Thiaminase II (EC 3.5.99.2), a hydrolase, might not be involved in the breakdown of thiamine as previously thought, but rather in the hydrolysis of aminopyrimidine to hydroxypyrimidine, a building block for the biosynthesis of thiamine by some bacteria.

Thiamine biosynthesis is regulated by metabolite-sensing messenger RNAs (mRNAs), called 'riboswitches'. These small RNA molecules are part of an mRNA molecule, typically at the 5'-untranslated region, coding for a protein involved in a biosynthetic pathway. Riboswitches have been identified for vitamins, amino acids, and nucleotides. The binding of such small molecules leads to a change in the conformation of the mRNA, affecting protein expression. The ThDP riboswitch (Thi-box) is found in many bacteria, fungi, and plants, and is the most widespread riboswitch known today. It acts by various mechanisms. In *Escherichia coli*, ThDP binding to the Thi-box leads to the formation of a terminator hairpin, halting the transcription of several thiamine biosynthetic enzymes. In fungi and plants, the Thi-box is involved in the control of RNA splicing rather than transcription or translation. The ThDP riboswitch is the only one

so far described in eukaryotes. The existence of riboswitches, in addition to ribozymes, demonstrates that macromolecules other than proteins are able to bind small molecules with high specificity, supporting the RNA world hypothesis.

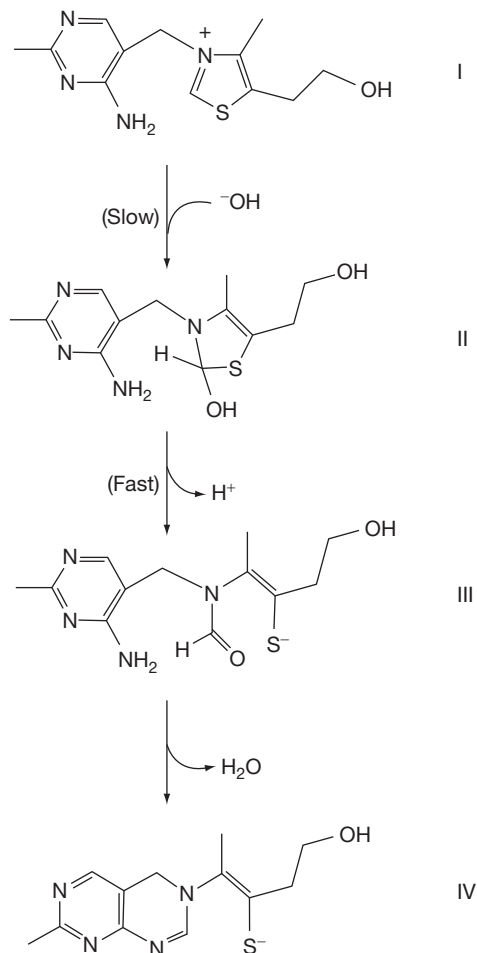


Figure 2 Conversion of thiamine to thiolate forms in alkaline solution. I, thiamine (cationic thiazolium form); II, unstable intermediate (pseudobase); III, thiolate form; IV, yellow weakly basic anion.

Determination of Thiamine Derivatives and Their Distribution in Various Living Organisms

Early methods for the determination of thiamine in biological or pharmaceutical preparations included paper, thin-layer, or liquid chromatography. Nowadays, mainly high-performance liquid chromatography (HPLC), after derivatization of thiamine compounds to the corresponding highly fluorescent thiochrome derivatives, is used because of its sensitivity, selectivity, and resolution (Figure 3). All thiochrome derivatives have approximately the same excitation optimum at 360 nm and the same emission maximum at 430–435 nm.

In most eukaryotic tissues, the cofactor ThDP is the most abundant thiamine compound. Free thiamine and ThMP generally amount to 5–15% of total thiamine, while ThTP, adenosine thiamine triphosphate (AThTP), and adenosine thiamine diphosphate (AThDP) are even less abundant (but exceptions exist, see below). In animal cells, such as neurons, ThDP is found in highest amounts in mitochondria, where it is mostly bound to pyruvate and oxoglutarate dehydrogenase complexes. Smaller amounts are found in the cytosol (partly free and partly bound to transketolase (TK)). In liver and skeletal muscle, however, ThDP is mostly cytosolic. ThDP is also present in peroxisomes, where it is bound to 2-hydroxyacyl-CoA ligase. The triphosphorylated derivatives ThTP and AThTP are generally minor compounds in animal tissues, but ThTP has been consistently found in most organisms, from bacteria to mammals. The subcellular distribution of ThTP in animal cells varies with the tissue studied. In rodent brain, it is essentially localized in mitochondria, while it is mainly cytoplasmic in skeletal muscle. These differences are probably related to different mechanisms of ThTP synthesis (see below). In *E. coli*, ThTP is not found under optimal growth conditions, but it can accumulate significantly (>20% of total thiamine) during amino acid starvation.

AThTP is found in several tissues (mammalian liver and heart, roots of higher plants, etc.), but it is not as widespread as ThTP. In *E. coli*, it is produced in response to carbon starvation. As to AThDP, it was only found in low amounts in rodent liver and in *E. coli*.

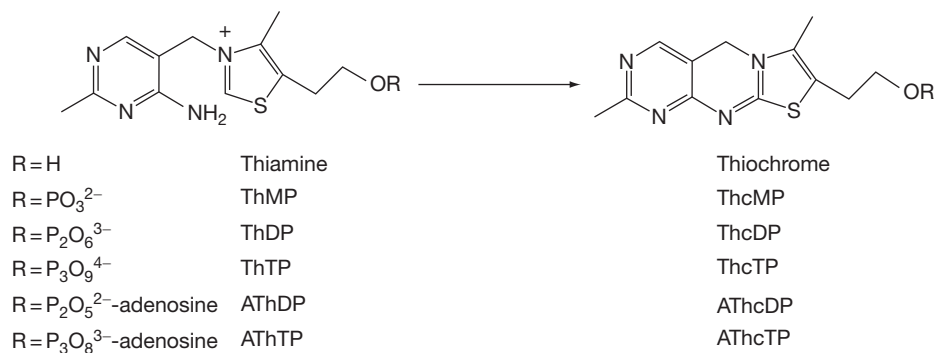


Figure 3 Conversion of thiamine compounds to the corresponding thiochrome derivatives by oxidation in alkaline medium by potassium ferricyanide or cyanogen bromide. AThcDP, adenosine thiochrome diphosphate; AThcTP, adenosine thiochrome triphosphate; ThcDP, thiochrome diphosphate; ThcMP, thiochrome monophosphate; ThcTP, thiochrome triphosphate; AThDP, adenosine thiamine diphosphate; AThTP, adenosine thiamine triphosphate; ThDP, thiamine diphosphate; ThMP, thiamine monophosphate; ThTP, thiamine triphosphate.

Thiamine Transport

Thiamine is found in most food, but whole grains, liver, and yeast are particularly rich sources. The Reference Daily Intake is 1.4 mg, but depends on carbohydrate intake. Large excess oral intake does not seem to have adverse effects. However, allergic reactions (anaphylactic shock) are a rare complication of intravenous thiamine administration.

Thiamine phosphate esters taken up from the food are hydrolyzed to thiamine in the small intestine, transported as thiamine or ThMP across the brush-border membrane to the portal system, and distributed to the various tissues. There is no true thiamine storage reservoir in the body, though, because of its size and high ThDP content, the liver contains 10–20% of total body thiamine. In case of reduced thiamine intake, it can be released from the liver and transported via the blood flow to other tissues (in particular the brain). Excess thiamine and some of its metabolites (2-methyl-4-amino-5-pyrimidine carboxylic acid and 4-methyl-thiazole-5-acetic acid) are excreted in the urine. Three kinds of cell membrane thiamine transporters have been described. They all belong to the *SLC19A* solute carrier gene family. *SLC19A1* encodes a reduced folate transporter that may also transport ThMP and ThDP, but not thiamine. THTR-1 (product of the *SLC19A2* gene) and hTHTR-2 (product of the *SLC19A3* gene) are thiamine/ H^+ antiporters, rather ubiquitously expressed in mammalian tissues.

Thiamine transport across cell membranes, particularly the blood–brain barrier, is a relatively slow process. Furthermore, plasma levels of thiamine are relatively low in humans as compared to other mammalian species. Both factors probably contribute to the fact that marginal thiamine deficiency is more common than initially thought of. In order to overcome severe deficiencies, several thiamine precursors (i.e., thiamine disulfides and benfotiamine) with a higher bioavailability were

developed. These molecules are generally lipophilic, easily crossing the membranes, and their administration results in a rapid and strong increase in plasma thiamine levels.

Metabolism of Thiamine Phosphates

Thiamine phosphate derivatives can be readily interconverted, but the enzymes involved in these reactions remain poorly characterized (Figure 4). Only two of them, the essential thiamine pyrophosphokinase and the 25-kDa thiamine triphosphatase, have been characterized at the molecular level.

Synthesis of ThDP

Thiamine taken up by cells is converted to ThDP by a ubiquitous cytosolic adenosine triphosphate (ATP): thiamine diphosphokinase or thiamine pyrophosphokinase (ThDPK or TPK; EC 2.7.6.2). The K_m for thiamine is in the low micromolar range, while the K_m for ATP is high ($\sim 10^{-2}$ M). The equilibrium of the reaction is not in favor of ThDP formation. Therefore, it must be shifted to the right, either by binding of ThDP to apoenzymes or through conversion of adenosine monophosphate (AMP) to adenosine diphosphate (ADP) by the adenylate kinase reaction (in muscle cytoplasm).

Synthesis of ThTP

The mechanism of ThTP synthesis and its physiological role are very controversial. The first mechanism proposed, a reaction catalyzed by a soluble ThDP-kinase: $ATP + ThDP \rightleftharpoons ADP + ThTP$, was never substantiated. ThTP can be synthesized by adenylate kinases (EC 2.7.4.3) according to the reaction $ADP + ThDP \rightleftharpoons AMP + ThTP$. The rate of reaction

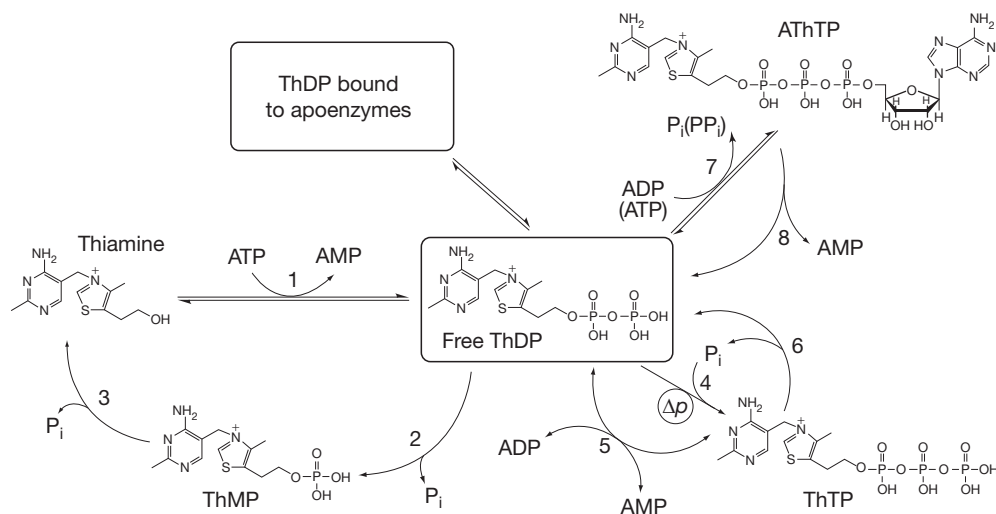


Figure 4 Interconversion of thiamine derivatives in a typical human cell. 1, thiamine diphosphokinase (ThDK); 2, thiamine diphosphatases (nonspecific); 3, thiamine monophosphatases (nonspecific); 4, mitochondrial ThTP synthase; 5, cytosolic adenylate kinase; 6, cytosolic 25-kDa ThTPase; 7, ThDP:ADP adenyl transferase; 8, AThTP hydrolase (hypothetical). ADP, adenosine diphosphate; AMP, adenosine monophosphate; AThTP, adenosine thiamine triphosphate; ATP, adenosine triphosphate; P_i , inorganic phosphate; PP_i , inorganic pyrophosphate; ThDP, thiamine diphosphate; ThMP, thiamine monophosphate; ThTP, thiamine triphosphate. Modified from Bettendorff L and Wins P (2009) Thiamine diphosphate in biological chemistry: New aspects of thiamine metabolism, especially triphosphate derivatives acting other than as cofactors. *FEBS Journal* 276: 2917–2925.

is 10^6 – 10^7 times slower than the synthesis of ATP according to the reaction $2 \text{ ADP} \rightleftharpoons \text{AMP} + \text{ATP}$. Thus, synthesis of ThTP by this mechanism is significant only in tissues where adenylate kinase is very abundant, such as skeletal muscle, electric organs, or red blood cells. ThTP may be the predominant thiamine compound when the 25-kDa ThTPase is absent, such as in electric organs or chicken skeletal muscle. In brain mitochondria, ThTP is synthesized according to the reaction $\text{ThDP} + \text{P}_i \rightleftharpoons \text{ThTP}$, coupled to the respiratory chain through a chemiosmotic mechanism.

Synthesis of AThTP

AThTP can be synthesized by a soluble ThDP adenylyl transferase, according to the reaction $\text{ADP (ATP)} + \text{ThDP} \rightleftharpoons \text{P}_i (\text{PP}_i) + \text{AThTP}$. Although both ATP and ADP may be the phosphate donor for this reaction, ADP is probably more physiologic as AThTP is synthesized under conditions of energy stress in *E. coli*. This bacterial enzyme has been partially characterized.

Hydrolysis of ThDP and ThMP

Several phosphatases may hydrolyze ThDP, but until now no specific thiamine diphosphatase (ThDPase) or thiamine monophosphatase (ThMPase) has been characterized at the molecular level. ThDPases studied so far also hydrolyze nucleoside diphosphates. ThDPase activity has been used as a Golgi marker. ThMPase activity, a classical marker of small-diameter dorsal root ganglia neurons, is probably identical to the transmembrane isoform of prostatic acid phosphatase. This enzyme, in fact an ecto-5'-nucleotidase, hydrolyzes adenosine monophosphate to adenosine, which can then activate A_1 -adenosine receptors, suppressing pain.

Hydrolysis of ThTP

Membrane-associated and cytosolic thiamine triphosphatases (ThTPase; EC 3.6.1.28) have been described. The membrane-associated ThTPase could not be purified so far. In contrast, a soluble ThTPase, purified from mammalian tissues, has a nearly absolute specificity for ThTP. It is a 25-kDa protein with no homology with other known mammalian proteins. Its mRNA is widely expressed in human tissues, though at relatively low levels. Orthologs are found in nearly all animals, except birds.

Thiamine-Binding Proteins

Thiamine-binding proteins have been described in many animal and plant tissues, but none were characterized at the molecular level. They exist in eggs or plant seeds, probably as thiamine storage proteins.

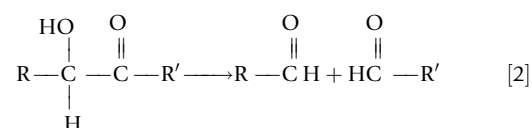
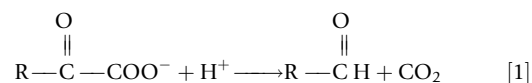
Cofactor Role of ThDP

ThDP is an indispensable cofactor for energy metabolism in virtually all cell types. In animals, it is involved in several important enzyme reactions in both carbohydrate and amino

acid catabolism (Figure 5). TK (EC 2.2.1.1), a cytosolic enzyme, catalyzes a key step in the pentose phosphate pathway. In mitochondria, ThDP is required for the oxidative decarboxylation of pyruvate and 2-oxoglutarate. Pyruvate dehydrogenase (PDHC; EC 1.2.4.1) and oxoglutarate dehydrogenase (OGDHC; EC 1.2.4.2) complexes catalyze two essential steps of the Krebs cycle. ThDP is also required for the catabolism of branched-chain amino acids, being the coenzyme of branched-chain 2-oxo acid dehydrogenase (BCOAD; EC 1.2.4.4) in mitochondria. In peroxisomes, it is the coenzyme for 2-hydroxyacyl-CoA lyase (HACL1). In yeasts, ThDP is a cofactor for pyruvate decarboxylase (EC 4.1.1.1) catalyzing the anaerobic decarboxylation of pyruvate to acetaldehyde in alcoholic fermentation. ThDP is also the cofactor for several additional enzymes in plants and prokaryotes, such as acetolactate synthase (EC 4.1.3.18) or pyruvate oxidase (EC 1.2.3.3).

Catalytic Mechanism Involving Thiamine Diphosphate

ThDP acts as catalyst for the decarboxylation of 2-oxo acids (α -cleavage; eqn [1]) and the cleavage (or formation) of α -hydroxyketones (eqn [2])



There is no simple acid–base-catalyzed mechanism for such reactions, hence the need for a special coenzyme.

It is now well established that the first committed step in ThDP-dependent enzyme reactions is the deprotonation of the thiazolium C-2 and attack of the resulting carbanion on a substrate carbonyl. As evidenced by the easy exchange of the C-2 proton with deuterium, this CH group is more acidic than most $=\text{CH}-$ groups. Proton removal from C-2 is assisted by the adjacent positive charge ($=\text{N}^+$). In addition, the aminopyridinium moiety of ThDP, presumably involving the imino tautomer, acts as a general base. Finally, a strictly conserved glutamate residue of the apoenzyme forms a hydrogen bond with the weakly basic N-1' of the aminopyridinium, also facilitating the dissociation of the H^+ in C-2 position of the thiazolium.

As already suggested by Breslow over 50 years ago, the thiazolium dipolar ion (or ylide) formed by this dissociation is the key intermediate in all ThDP-dependent enzyme reactions. Note that the diphosphate group plays no role in catalysis and is only required for tight binding of ThDP to the apoenzymes. The anionic carbon of the thiazolium ring readily reacts with substrates by addition to the carbonyl group, forming a covalent adduct. As a typical example, we shall consider the mechanism of pyruvate decarboxylation that occurs as the first step in the conversion of pyruvate to acetyl-CoA by the PDH complex. This step is catalyzed by the E1 subunits. As

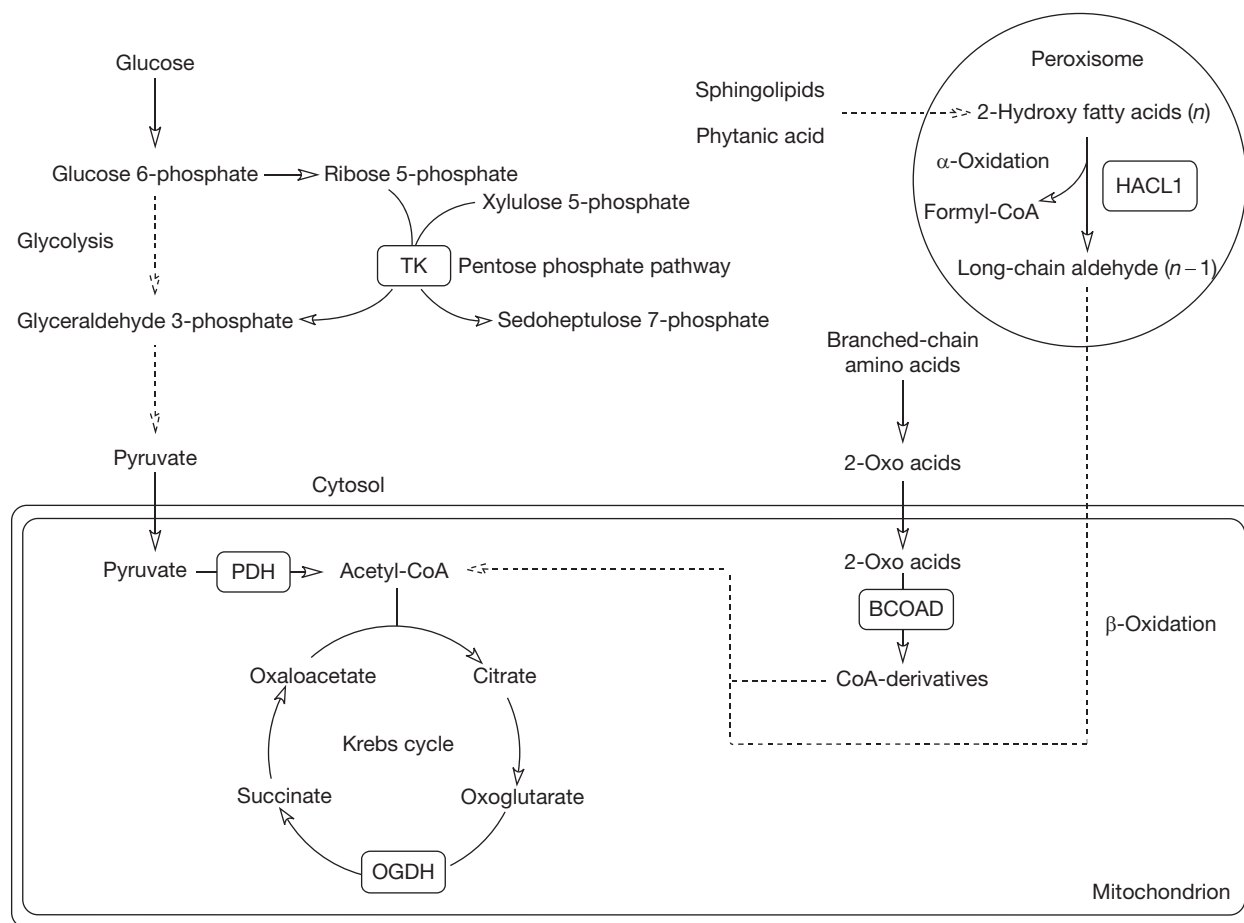


Figure 5 Enzyme reactions requiring ThDP as a cofactor in an animal cell. BCOAD, branched-chain 2-oxo acid dehydrogenase; HAC11, 2-hydroxyacyl-CoA lyase; OGDH, oxoglutarate dehydrogenase; PDH, pyruvate dehydrogenase; TK, transketolase.

shown in **Figure 6**, the first adduct formed is α -lactyl-ThDP. In this complex, the $=N^+$ group of the thiazolium acts as an electron sink to stabilize the formation of the negative charge, which is necessary for decarboxylation. After the release of CO_2 , the activated aldehyde attached to ThDP is oxidized to form an acetyl group and concomitantly transferred to lipoyl groups attached to the E2 subunits, while the ylide is regenerated (**Figure 6**).

One of best-studied ThDP-dependent enzymes is *E. coli* PDH. Its mechanism can be considered as the paradigm for the oxidative decarboxylation of 2-oxo acids. This multisubunit complex catalyzes the conversion of pyruvate to acetyl-CoA, which can then enter the Krebs cycle:



E. coli PDH is a large complex (~2390 kDa) composed of multiple copies of three distinct subunits: 24 subunits of PDH (E1, EC 1.2.4.1), 24 subunits of dihydrolipoyl S-acetyltransferase (E2, EC 2.3.1.12), and 12 subunits of dihydrolipoyl dehydrogenase (E3, EC 1.8.1.4). Five different coenzymes (ThDP, lipoic acid, coenzyme A, FAD, and NAD^+) are involved in the reaction catalyzed by the complex. ThDP is the cofactor for E1, the enzyme catalyzing the decarboxylation

of pyruvate, and the transfer of the acetyl moiety to the lipoyl cofactor attached of E2. The structure and mechanism of the OGDH complex are quite similar to those of the PDH complex.

ThDP is bound to a conserved amino acid sequence GDG-X₂₆-NN via a Mg^{2+} cation. This consensus sequence (ThDP-binding motif) is conserved in all ThDP-dependent enzymes.

Non-Cofactor Roles of Thiamine Derivatives

Non-cofactor roles of thiamine and its phosphate derivatives have been proposed for many years, following the work of Alexander von Murlalt in the 1930s and 1940s. This was based on the observation, made independently by several investigators, that electrical stimulation of isolated nerves led to release of thiamine. It was speculated that this release might result from a dephosphorylation shift of phosphorylated derivatives. This drove interest to ThTP that exists in small amounts in nervous tissues, and was sometimes considered as a neuroactive form of thiamine. However, interest in this compound rapidly faded, as early results were questioned, because reliable detection proved to be difficult. In the early 1980s, estimation of ThTP content in brain by HPLC showed that it was less abundant than previously thought, and an earlier hypothesis

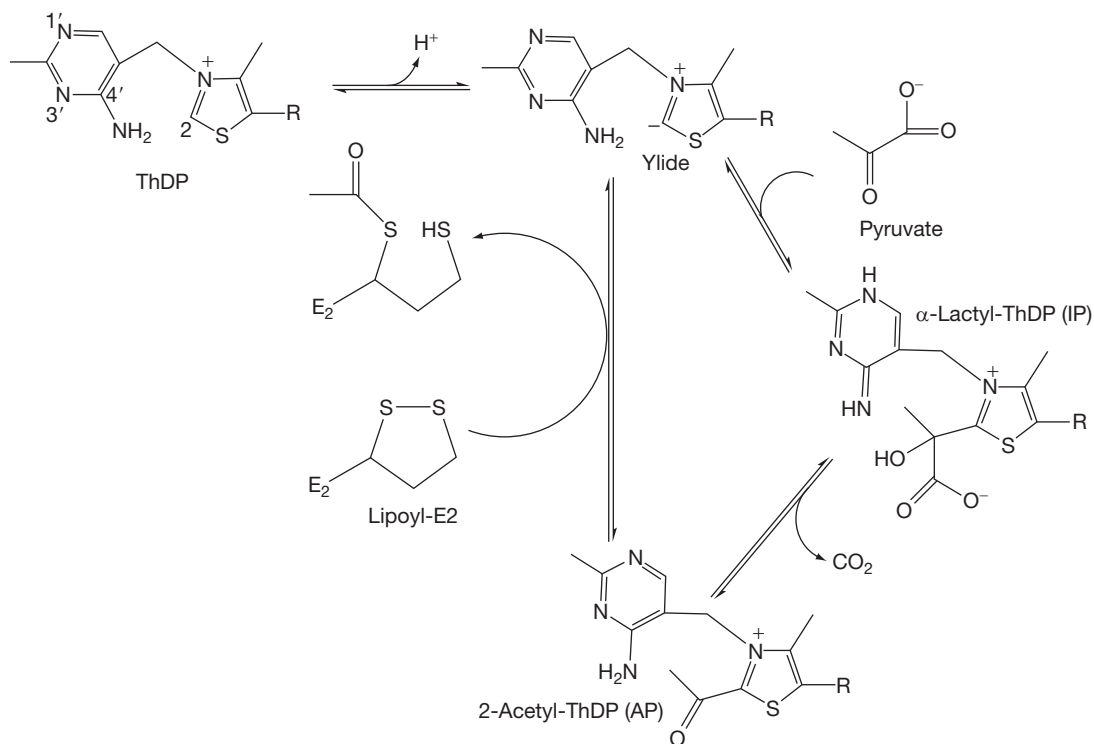


Figure 6 Catalytic mechanism for the first step (pyruvate decarboxylation) of the reaction catalyzed by the PDH complex. Pyruvate decarboxylation is catalyzed by the E1 subunits of the complex that bind ThDP. AP, 4'-aminopyrimidinium IP, 1', 4'-iminopyrimidinium; ThDP, thiamine diphosphate; $R = C_2H_4OP_2O_6^{3-}$.

concerning its possible role in nerve excitation could not be substantiated. However, the discovery of the specific 25-kDa ThTPase with high catalytic efficiency in mammalian tissues, as well as the finding that ThTP can phosphorylate some proteins in brain and electric tissue still suggests that ThTP may play some physiological role in brain and other animal tissues. In *E. coli*, it was recently found that ThTP accumulates in large amounts in response to amino acid starvation, suggesting a specific role in the adaptation of bacteria to nutritional downshifts. The discovery of an adenylated thiamine triphosphate derivative (ATHTP or thiaminylated ATP; **Figure 1**) in 2007 came as a surprise. ATHTP accumulates in *E. coli* during carbon starvation. This compound, like ThTP, might thus be a kind of alarmone, involved in emergency signaling when the cells undergo some forms of metabolic stress. In eukaryotic organisms, the possible biological role(s) of triphosphorylated derivatives (ATHTP and ThTP) remains unknown.

Human Diseases Related to Thiamine Deficiency

Ill-equilibrated diet or problems in absorption can lead to thiamine deficiency. Humans appear to have relatively low plasma and cellular thiamine levels, and this may be the reason why humans are particularly sensitive to thiamine deficiency. Severe chronic or acute deficiency leads to two main diseases, beriberi and Wernicke–Korsakoff syndrome. However, in addition to these classical disorders, marginal thiamine deficiency is probably quite common, particularly in high-risk groups such as elderly

people, lactating women, diabetic, and AIDS patients. People with Alzheimer's disease also have lower cerebral ThDP levels and impaired 2-oxoglutarate dehydrogenase activity.

Beriberi

Beriberi is a polyneuritic syndrome, mainly affecting the lower limbs. Symptoms of dry beriberi include weakness and pain in the limbs, weight loss, and partial paralysis. In wet beriberi, there are also cardiovascular symptoms, including cardiac insufficiency and enlargement, with tendency to edema. A low TK activity of red cells is a reliable diagnostic indicator of severe thiamine deficiency. Beriberi is the classical manifestation of chronic dietary deficiency in vitamin B1. Before the discovery of thiamine, it was a major health problem in East Asian countries, where polished rice (thiamine is mainly present in the husk of rice) was the staple food.

Wernicke–Korsakoff Syndrome

Although severe dietary thiamine deficiency has become rare in Western countries, a special form of deficiency associated with chronic alcoholism remains frequent. This is related to the poor nutritional status of many alcoholics and also to the fact that alcohol interferes to some extent with thiamine absorption and its intracellular conversion to ThDP.

The acute manifestation of this disorder is Wernicke's encephalopathy, a state of confusion with neurological symptoms such as ophthalmoplegia, nystagmus, and ataxia. In the absence of

prompt therapy with high-dose thiamine, coma and death occur rapidly. In many cases, recovery will not be complete and irreversible sequelae, due to specific brain lesions, will be observed. These lesions often result in a state of dementia called Korsakoff's psychosis. This is an amnesic syndrome, most often including a loss of anterograde memory. This may be accompanied by confusion, apathy, disorientation, and fabrications. Surprisingly, the brain lesions are rather specific, mainly affecting thalamic nuclei and mammillary bodies, while (in contrast to Alzheimer's disease) the cortex is largely spared. Brain damage in Wernicke–Korsakoff syndrome is assumed to be related to impairment of oxidative energy metabolism, neuronal death arising from a combination of excitotoxicity, oxidative stress, and inflammatory processes (the latter may primarily affect endothelial cells of the microvessels rather than neurons). However, the reasons for the selective vulnerability of some brain regions remain elusive and are the subject of active investigation.

Defects in Thiamine Transport and in ThDP-Dependent Enzymes

Mutations in the genes coding thiamine transporters or ThDP-dependent enzymes may lead to specific diseases in humans.

Thiamine-Responsive Megaloblastic Anemia

Thiamine-responsive megaloblastic anemia is characterized by megaloblastic anemia, sensorineural hearing loss, and diabetes mellitus. It is caused by mutations in the THTR-1 high-affinity transporter. Relief of the symptoms is generally obtained by high-dose thiamine treatment.

Amish Lethal Microcephaly

The children affected by Amish lethal microcephaly are born with a very small head and an underdeveloped brain, and they do not live beyond six months. The disease results from a point mutation in the mitochondrial ThDP transporter gene *SLC25A19*, leading to its inability to transport the cofactor to the mitochondrial matrix. This causes the inactivation of PDH and OGDH complexes, and impairment of oxidative metabolism.

Leigh Disease (Subacute Necrotizing Encephalomyelopathy)

Leigh disease is a rare heterogeneous neurodegenerative disorder of early childhood, involving focal, symmetric necrotic brain lesions, leading to death generally before the age of five. It is clinically and genetically heterogeneous, presenting central and peripheral nervous system abnormalities that result from deficits in respiratory chain mitochondrial complexes (I, II, III, IV, and V) or PDH. In a small percentage of patients (<10%), high-dose thiamine may have beneficial effects.

Maple Syrup Urine Disease

Maple syrup disease (branched-chain ketoaciduria) is an autosomal recessive disorder caused by mutations in branched-chain 2-oxo acid dehydrogenase. The toxic effect is due to the accumulation of the branched-chain amino acids (leucine, isoleucine, and valine) and their 2-oxo acid degradation

products. If not treated by special diets, the patients will develop brain damage after birth, leading to death. High-dose thiamine therapy may be efficient, in particular when the E1 component of branched-chain 2-oxo acid dehydrogenase is affected.

Conclusion

It was long considered that the only reason for the existence of thiamine was to form the cofactor ThDP, an argument also based on the observation that thiazole, a heterocycle absolutely required for catalysis, is rarely observed in biological molecules. Although it remains true that the main functions of this vitamin are linked to catalysis by ThDP, the existence of triphosphorylated derivatives, such as ThTP and ATHTP, whose levels are highly specifically regulated, strongly suggest that these compounds play a biological role, at least in bacteria. Thiamine compounds form a family of molecules that, with respect to their degree of phosphorylation, is reminiscent of nucleotides (though thiamine, lacking a sugar moiety, is not a nucleotide). This suggests that in addition to the classical cofactor role of ThDP, some thiamine derivatives have non-cofactor roles.

See also: [Metabolism Vitamins and Hormones: Pentose Phosphate \(Hexose MonoPhosphate\) Pathway; Structure and Regulation of Pyruvate Dehydrogenases; Tricarboxylic Acid Cycle.](#)

Further Reading

- Bettendorff L (1994) Thiamine in excitable tissues: Reflections on a non-cofactor role. *Metabolic Brain Disease* 9: 183–209.
- Bettendorff L and Wins P (2009) Thiamin diphosphate in biological chemistry: New aspects of thiamin metabolism, especially triphosphate derivatives acting other than as cofactors. *FEBS Journal* 276: 2917–2925.
- Bettendorff L, Wirtzfeld B, Makarchikov AF, et al. (2007) Discovery of a natural thiamine adenine nucleotide. *Nature Chemical Biology* 3: 211–212.
- Carpenter KJ (2000) *Beriberi, White Rice, and Vitamin B: A Disease, a Cause, and a Cure*. Berkeley, CA: University of California Press.
- Gangolf M, Wins P, Thiry M, El Moualij B, and Bettendorff L (2010) Thiamine triphosphate synthesis in the rat brain is mitochondrial and coupled to the respiratory chain. *Journal of Biological Chemistry* 285: 583–594.
- Gibson GE and Blass JP (2007) Thiamine-dependent processes and treatment strategies in neurodegeneration. *Antioxidants and Redox Signaling* 9: 1605–1619.
- Harper C (2009) The neuropathology of alcohol-related brain damage. *Alcohol and Alcoholism* 44: 136–140.
- Hazell AS and Butterworth RF (2009) Update of cell damage mechanisms in thiamine deficiency: Focus on oxidative stress, excitotoxicity and inflammation. *Alcohol and Alcoholism* 44: 141–147.
- Jordan J and Patel MS (2004) *Thiamine. Catalytic Mechanisms in Normal and Disease States*. New York, NY: Dekker.
- Jurgenson CT, Begley TP, and Ealick SE (2009) The structural and biochemical foundations of thiamin biosynthesis. *Annual Review of Biochemistry* 78: 569–603.
- Kawasaki T (1992) Vitamin B1: Thiamine. In: De Leenheer AP, Lambert WE, and Nelis HJ (eds.) *Modern Chromatographic Analysis of Vitamins*, vol. 60, pp. 319–354. New York, NY: Dekker.
- Kluger R and Tittmann K (2008) Thiamin diphosphate catalysis: Enzymic and nonenzymic covalent intermediates. *Chemical Reviews* 108: 1797–1833.
- Lakaye B, Wirtzfeld B, Wins P, Grisar T, and Bettendorff L (2004) Thiamine triphosphate, a new signal required for optimal growth of *Escherichia coli* during amino acid starvation. *Journal of Biological Chemistry* 279: 17142–17147.
- McCandless DW (2010) *Thiamine Deficiency and Associated Clinical Disorders*, Contemporary Clinical Neuroscience, 192pp. Totowa, NJ: Humana Press.
- Nemeria NS, Chakraborty S, Balakrishnan A, and Jordan F (2009) Reaction mechanisms of thiamin diphosphate enzymes: Defining states of ionization and tautomerization of the cofactor at individual steps. *FEBS Journal* 276: 2432–2446.

Bioenergetics and the Mitochondrial Genome

D C Wallace, University of Pennsylvania, Philadelphia, PA, USA

© 2013 Elsevier Inc. All rights reserved.

It is becoming increasingly clear that our understanding the biology of the mitochondrion and the genetics of the mitochondrial genome will be central to our determining the genetic and pathophysiological basis of a wide variety of complex human diseases. This is because to be alive requires structure (anatomy), the flow of energy (the vital force), and the information necessary to maintain the anatomical and energetic systems of the body. While all of the information for the anatomy of humans and animals is encoded by the chromosomal and thus Mendelian-inherited nuclear DNA (nDNA), the most important elements of the energy-generating system, oxidative phosphorylation (OXPHOS), are encoded in the non-Mendelian mitochondrial DNA (mtDNA). Therefore, most human diseases will involve an interaction between anatomy and energy, and hence their genetics will involve a mixture of both Mendelian and non-Mendelian genes.

Mitochondrial Origin and Function

The mitochondria are symbiotic bacteria that reside within the cytoplasm of virtually all eukaryotic cells (cells with a nucleus). The mitochondrion originated when an α -Proteobacteria, a *Rickettsia*-like oxidative bacterium, formed a symbiotic relationship with the proto-nucleus-cytosol organism about 2–2.5 billion years ago. Hence, each human cell is composed of two life forms which were originally co-equal microorganisms. As the symbiosis matured, one organism became specialized in encoding structural information and gave rise to the nucleus-cytosol of the modern human cell. The other became specialized in energy production and gave rise to the modern mitochondrion. As this functional specialization matured over the first 1.2 billion years, the proto-mitochondrion gave up all of its genomic information for maintaining its structure to the proto-nucleus-cytosolic organism, ultimately surrendering more than 99% of its genome. However, the mtDNA retained the core information for the electrical components of OXPHOS, literally the wiring diagram of the mitochondrial power plants.

From its origins up to the present day, virtually all energy flux through the eukaryotic cell goes through the mitochondria and is controlled by the mtDNA. With the advent of multicellularity, ever-increasing cellular and later multicellular structural complexities were elaborated resulting in the continual addition of new anatomical information to the nDNA. Since considerable energy is required to replicate, transcribe, and translate genes, as the number of nDNA genes increased in the cell, the number of mitochondria per cell also needed to increase to provide the requisite energy. Today the human cell has a massive nucleus encompassing two copies of approximately 23000 genes. To generate sufficient energy for these genes, the human cell generally requires hundreds of mitochondria and thousands of mtDNAs.

The human mtDNA is a 16569 nucleotide (nt) double-stranded, circular molecule. This mtDNA is replicated and transcribed inside the mitochondrion via mitochondria-specific DNA and RNA polymerases. Each mtDNA encodes 13 OXPHOS proteins plus the 12S and 16S RNAs and 22 transfer RNAs (tRNAs) necessary for mitochondrial protein synthesis to express these genes. The 12S and 16S ribosomal RNAs (rRNAs) form the scaffold of the bacteria-like mitochondrial ribosomes which initiate translation with a formyl-methionine and are sensitive to bacterial ribosomal inhibitor, chloramphenicol. This latter sensitivity is why prolonged treatment with chloramphenicol is toxic to humans.

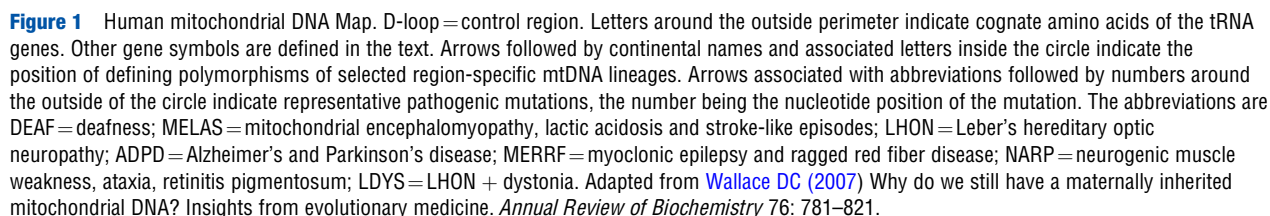
When the mtDNA gene number became sufficiently reduced, the genetic code of the fungal-animal mtDNA lineage began to drift. As a result, today, all members of the fungal-animal lineage use the opal stop codon to encode tryptophan. This change rendered the mtDNA transcripts unintelligible to the nuclear-cytosol, and the further transfer of functional genes from the mtDNA to the nDNA was blocked.

The 13 mtDNA OXPHOS polypeptides include seven (ND1, 2, 3, 4L, 4, 5, and 6) of the 45 polypeptides that constitute OXPHOS complex I; one (cytochrome *b*, *cytb*) of the 11 proteins of complex III; three (COI, II, and III) of the 13 subunits of complex IV; and two (ATP6 and 8) of the about 17 polypeptides of complex V (Figure 1).

OXPHOS (Figure 2) burns hydrogen derived from carbohydrates and fats with oxygen to generate water (H_2O). Carbohydrates are processed by cytosolic glycolysis to pyruvate. The pyruvate then enters the mitochondrion and, in the process, is converted to acetyl-coenzyme A (acetyl-CoA). Acetyl-CoA enters the mitochondrial tricarboxylic acid (TCA) cycle, a central purpose of which is to strip the reducing equivalents (electrons) from the carbon atoms and transfer them either to NAD^+ (nicotinamide adenine dinucleotide, oxidized form) to generate NADH (NAD, reduced form) or to FAD (flavin adenine dinucleotide, oxidized form) to give FADH₂ (FAD, reduced form). Acetyl-CoA and reducing equivalents are also generated within the mitochondrial matrix from fatty acids via the β -oxidation pathway.

The electrons from $NADH + H^+$ are oxidized by a series of OXPHOS complexes called the electron transport chain (ETC) arrayed within the mitochondrial inner membrane. The terminal electron acceptor is oxygen (O_2) and results in water (H_2O). The ETC collects electrons from NADH through complex I (NADH dehydrogenase) or from succinate in the TCA cycle through complex II (succinate dehydrogenase). These and other dehydrogenases transfer the electrons to the shared lipid electron carrier, ubiquinone (coenzyme Q_{10} or CoQ). CoQ can accept two electrons, one to generate ubiquinol (CoQH) and the second to produce ubiquinol (CoQH₂). From CoQH₂ the electrons are transferred to complex III, then cytochrome *c*, then complex IV, and finally to $\frac{1}{2} O_2$ to give H_2O .

The energy that is released as the electrons traverse complexes I, III, and IV, though not through complex II, is used to pump



The energy stored in ΔP can be used to drive many functions. For example, it influences cellular pH (H^+). It is used to transport calcium ions (Ca^{2+}) from the cytosol through carriers in the mitochondrial inner membrane into the mitochondrial matrix, thus regulating cellular Ca^{2+} concentrations. It can also be used to drive the condensation of adenosine diphosphate and phosphate ($ADP + P_i$) to make adenosine triphosphate (ATP). This is achieved by a fifth mitochondrial inner-membrane OXPHOS complex, the ATP synthase, or complex V. Complex V couples the flow of protons back into the mitochondrial matrix through an enzyme proton channel with the condensation of $ADP + P_i$ to give ATP. Mitochondrial ATP is then exchanged across the mitochondrial inner membrane for

The efficiency by which ΔP is converted to ATP is known as the coupling efficiency. Tightly coupled OXPHOS generates the maximum ATP and minimum heat, while loosely coupled OXPHOS requires more calories to produce the same amount of ATP, the excess calories generating additional heat. Whether calories go through an intermediate of ATP or are converted directly to heat, ultimately, each calorie burned produces the same amount of heat. In endotherms like humans this heat generation is regulated to maintain a stable body temperature.

As a by-product of OXPHOS, the mitochondria generate most of the reactive oxygen species (ROS or oxygen radicals) produced within eukaryotic cells. This results from the misdirection of electrons from the early stages of the ETC surrounding ubiquinone directly to O_2 to generate superoxide anion ($O_2^{\cdot -}$). OXPHOS complex I generates $O_2^{\cdot -}$ in the matrix and this $O_2^{\cdot -}$ is converted to hydrogen peroxide (H_2O_2) by the matrix Mn superoxide dismutase. Complex III produces $O_2^{\cdot -}$ in the inter-membrane space and this $O_2^{\cdot -}$ is detoxified by Cu/ZnSOD located within the inter-membrane space and the cytosol. In the presence of reduced transition metals H_2O_2 can be

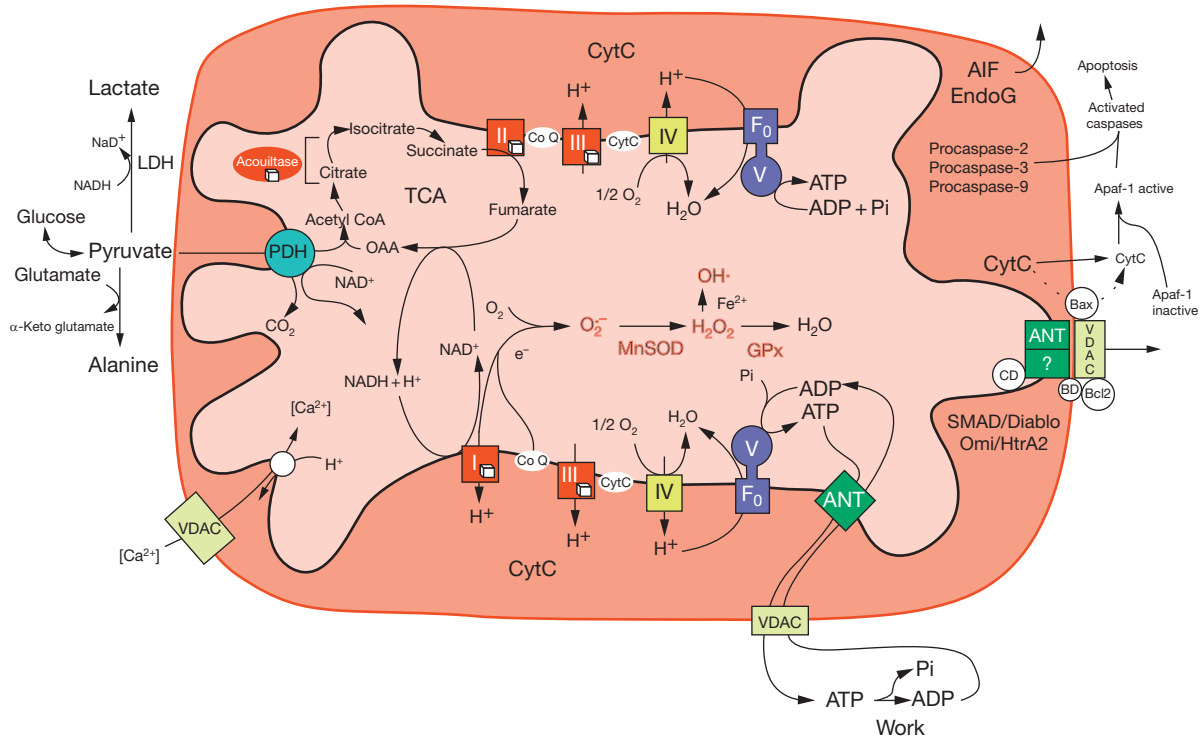


Figure 2 Five features of mitochondrial metabolism relevant to the pathophysiology of disease: (1) energy production by oxidative phosphorylation (OXPHOS), (2) reactive oxygen species (ROS) generated by OXPHOS as a signaling system and when excessive as a toxic by-product, (3) regulation of mitochondrial and cellular oxidation–reduction status, (4) regulation of cellular and mitochondrial Ca^{2+} concentrations, and (5) regulation of apoptosis through activation of the mitochondrial permeability transition pore (mtPTP). I, II, III, IV, and V = OXPHOS complexes I–V; ADP or ATP = adenosine di- or triphosphate; ANT = adenine nucleotide translocator, CytC = cytochrome *c*; GPx = glutathione peroxidase-1; LDH = lactate dehydrogenase; MnSOD = manganese superoxide dismutase or SOD2; NADH = reduce nicotinamide adenine dinucleotide; TCA = tricarboxylic acid cycle; VDAC = voltage-dependent anion channel. Adapted from Wallace DC (2007) Why do we still have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine. *Annual Review of Biochemistry* 76: 781–821.

reduced by single electron transfer to hydroxyl radical ($\cdot\text{OH}$) which is the most toxic ROS. Consequently, the cell attempts to minimize H_2O_2 by its conversion to water (H_2O) via a two-electron transfer mediated by glutathione peroxidase (GPx1) or conversion to water and oxygen by catalase (Figure 2).

Glutathione peroxidase reduces H_2O_2 to H_2O by using electrons donated from reduced glutathione (2GSH) to generate in oxidized glutathione (GSSG). GSSG is reduced again via glutathione reductase using electrons derived from $\text{NADPH} + \text{H}^+$. $\text{NADPH} + \text{H}^+$ is used since $\text{NADH} + \text{H}^+$ has insufficient reducing power to reduce GSSG. Therefore, to transfer the electrons from $\text{NADH} + \text{H}^+$ to NADP^+ additional energy is required. This energy is collected from the electrochemical gradient (ΔP) by the mitochondrial inner membrane protein, the nicotinamide nucleotide transhydrogenase. Hence, under normal health circumstances, mitochondrial ROS is carefully regulated based on electron flux (redox) and thus energy flux.

Mitochondrially generated ROS has a critical signaling function permitting the nucleus to monitor the energetics of the mitochondrion. However, if OXPHOS is inhibited by a block in one of the OXPHOS complexes or there is an excess of calories relative to demand for ΔP causing ΔP to become maximized, then the ETC stalls, all of the electron carriers become reduced, and the excess electrons are transferred as single electrons to O_2 accelerating ROS production. Excessive ROS can overwhelm the

antioxidant defense systems and persist long enough to damage mitochondrial proteins, lipids, and nucleic acids. This increased oxidative stress can ultimately compromise mitochondrial energy production and destroy the mtDNA's capacity to maintain OXPHOS function resulting in energetic failure.

Cells with faulty mitochondria are removed by programmed cell death (apoptosis) through the activation of the mitochondrial permeability transition pore (mtPTP). The mtPTP spans the mitochondrial inner and outer membrane. While the structure of the mtPTP remains unclear, proteins that have been implicated in its function include the outer-membrane VDAC, the inner-membrane ANT, the pro- and anti-apoptotic Bcl2-Bax gene family proteins, cyclophilin D, and the mitochondrial transporter protein (TSPO also known as the peripheral benzodiazepine receptor). This complex senses changes in mitochondrial ΔP , adenine nucleotides, ROS, and Ca^{2+} , and when these factors get sufficiently out of balance, the mtPTP is activated and opens a channel between the cytosol and the mitochondrial matrix. This depolarizes ΔP , frequently causing the mitochondrion to swell, and produces a large channel in the outer membrane. The perforation of the mitochondria outer membrane permits proteins stored in the mitochondrial inter-membrane space to escape into the cytosol. These proteins include cytochrome *c* which activates cytosolic Apaf-1 to convert pro-caspases 9 to an active protease.

Caspase 9 then cleaves and activates caspases 1 and 3. Caspases 1 and 3 and others then degrade all of the proteins of the cell. Apoptosis initiating factor (AIF) and caspase activated DNase (CAD) are also released and transported to the nucleus where they degrade the nDNA (Figure 2).

MtPTP-induced apoptosis means that cells with faulty mitochondria are removed from the tissue so that the energetic failure of the cell does not compromise the function of the remaining cells of the tissue and organ. However, post-mitotic tissues have a finite number of cells. As more cells are lost due to mitochondrial-induced apoptosis, the tissue cellularity declines. Ultimately, insufficient cells remain in the tissue to maintain normal tissue function and symptoms result. Generally, we are born with sufficient extra cells that we continue to function past reproductive age. However, there is no evolutionary advantage for maintaining the body beyond reproduction so eventually the tissues run low on cells and the individual dies. Thus, the pathophysiology of degenerative diseases and aging is the result of age-related mitochondrial failure, but the symptoms of old age are the product of the secondary loss of tissue cellularity.

Mitochondrially induced apoptosis also serves a second function. Many of the mitochondrial proteins are highly homologous to those of other bacteria. If the mitochondria were released into the circulation, they could induce an immunological response. Hence, apoptotic degradation of the cell and its mitochondria avoid the release of the mitochondrial antigens into the circulation.

Mitochondrial Genetics and Epigenomics

The nonchromosomal nature of the mtDNA means that it is not inherited according to the Mendel laws. The human mtDNA is inherited exclusively from the mother, the father's mtDNA being destroyed at fertilization. Since the mtDNA is bathed in ROS, it is prone to mutation. Therefore, individual mtDNAs within a cell can become mutated creating a mixture of mutant and normal mtDNAs, a state known as heteroplasmy. When a heteroplasmic cell undergoes cytokinesis, the mutant and normal mtDNAs are stochastically distributed between the daughter cells, a process that occurs during both mitosis and meiosis. This replicative segregation can generate a continuous array of genotypes from pure normal to pure mutant. The higher the proportion of the mutant mtDNAs, the lower the number of functional enzyme complexes available to sustain OXPHOS. Ultimately, the energy output of the mitochondria falls below the minimal energy requirements for the cell, resulting in cellular dysfunction, the bioenergetic threshold.

The high mtDNA mutation rate and critical function of the mtDNA genes should result in a high frequency of deleterious mtDNA mutations. If this were the case, the genetic load resulting from mtDNA mutations would compromise the robustness of the species, ultimately leading to extinction. Such a genetic catastrophe is prevented in mammals by an inter-ovarian selection system that eliminates those proto-oocytes with severe mitochondrial defects prior to fertilization. This pre-fertilization selection system mitigates the genetic load of the high mutation rate while at the same time permitting a wide range of potentially adaptive mutations to be

regularly introduced into the species to produce genetic diversity. Since the mtDNA controls the energetics of the cell, this system provides the genetic diversity needed for a species to genetically adapt to regional differences in the energetic environment within a species' niche.

While nDNA mutations in mitochondrial genes could also permit adaptation to environmental change, mutations in the nDNA can only be selected for or against following fertilization and the development of the individual. Since most structural gene mutations are deleterious, the nDNA mutation rate must be kept low. Hence, regional adaptive polymorphisms in the nDNA are relatively rare, consequently the mtDNA mutations are the primary genetic variation for intra-species adaptation to regional and long-term environmental changes.

The over one thousand nDNA-encoded mitochondrial genes dispersed across the chromosomes also play an essential role in human adaptation, but not through mutation. Many environmental changes are both short term and reversible. This includes changes in the energetic environment that occur daily, monthly, annually, and over recent generations. Since genetic mutations are irreversible, longer-term reversible adaptive changes in energetics are achieved by changes in the expression of the nDNA-encoded mitochondrial genes via the epigenome through modification of chromatin structure. Shorter-term reversible changes are achieved through energetic modification of the proteins of the cellular signal transduction pathways.

Both the chromatin and the signal transduction pathways are modulated by protein and DNA modifications using mitochondrially derived high-energy intermediates, including ATP, acetyl-CoA, and S-adenosylmethionine (SAM). The histones that package the nDNA into nucleosomes have positively charged amino terminal tails. When unmodified, the positive charges of the histone tails bind to the negative charges of the DNA sugar-phosphate backbone shutting down transcription and replication. However, when calories are in excess, then intra-cellular ATP, acetyl-CoA, and SAM levels are high. These compounds phosphorylate, acetylate, or methylate the histone tails neutralizing the positive charge. This releases the grip of the histones on the DNA and opens up the chromatin for transcription and replication. Hence, when energy is flowing through the mitochondria, the mitochondrial high-energy products are prevalent and the mitochondrial high-energy intermediates activate nDNA gene expression for growth and reproduction. When energy flow subsides, mitochondria high-energy intermediates decline, histone modification declines, and nDNA transcription shuts down. The same logic can be generalized to the regulation of the cellular signal transduction pathways that are mediated by phosphorylation and acetylation.

Therefore, bioenergetics permits three levels of adaptation: adaptation to reversible environmental changes occurring over days to generation through energetic modulation of epigenomic regulation of nDNA mitochondrial gene expression, adaptation to stable regional variation within the species' niche over thousands of years through high-frequency mtDNA mutations, and adaptation to different energy reservoirs (eating nuts vs. insects for Darwin's finches) through nDNA gene mutations over millions of years that change the species' anatomy (the beaks of finches) resulting in speciation. This evolutionary perspective now permits interpretation of the array of disease-causing mitochondrial dysfunctions.

Disease-Causing mtDNA Mutations

A wide range of pathogenic mtDNA mutations have been described over the past 20 years. The first inherited pathogenic mtDNA mutations were base substitution mutations. These included the mtDNA missense mutation in the ND4 gene at nt 11778 that converted the arginine at codon 340 to histidine and causes Leber's hereditary optic neuropathy, a form of sudden-onset, mid-life, blindness, and the tRNA^{Lys} nt 8344 base substitution protein synthesis mutation that causes myoclonic epilepsy and ragged red fiber disease. The first spontaneous mtDNA disease-causing mutations were deletions associated with mitochondrial myopathy. Deletions in the mtDNA are now known to cause chronic progressive external ophthalmoplegia, the Kearns–Sayre syndrome, or the Pearson's marrow pancreas syndrome. Since these discoveries, a plethora of pathogenic mutations have been described in the mtDNA.

Mitochondrial defects primarily affect the more energetic tissues including central nervous system, heart, muscle, kidney, and endocrine systems. While the frequency of deleterious mtDNA mutations in the human population is high, in fact only the milder of all possible deleterious mutations generally appear in the population due to the filtering function of the intra-ovarian selection system. A catalogued identified mtDNA pathogenic base substitution mutations is available at the MITOWEB website.

The high mtDNA mutation rate, paired with intra-ovarian selection, permits the continuous introduction of new mtDNA nucleotide substitutions into the population. These range from deleterious to advantageous, depending on the environmental and genetic context in which the mtDNA mutations arise. Mutations that are beneficial in a particular environment will preferentially reproduce and become enriched in the region. Since the mtDNA is exclusively maternally inherited, and can only change by sequential mutations along radiating maternal lineages, this results in the development of regional groups of related mtDNA haplotypes derived from the mtDNA in which the original adaptive mutation first appeared. Such regional groups of related haplotypes are called haplogroups.

By defining the regional mtDNA haplogroups and deducing the sequence of mutation events that connect the haplogroups, it has been possible to reconstruct the history of woman (Figure 3). These studies have revealed that humans arose in Africa about 150 000–200 000 years before present (YBP). Sub-Saharan African mtDNAs radiate into four African-specific lineages: L0, L1, L2, and L3, the most ancient mtDNA lineages being found among the Khoisan Bushman. About 65 000 YBP, L3 gave rise to two new lineages in northeastern Africa, designated M and N. Only M and N left Africa to colonize all of Eurasia, with one group migrating along the southeastern Asian coast to colonize Australia. In Europe, lineage N gave rise to the European-specific lineages H, I, J, K, T, U, V, W, and X. In Asia, both M and N radiated to generate a

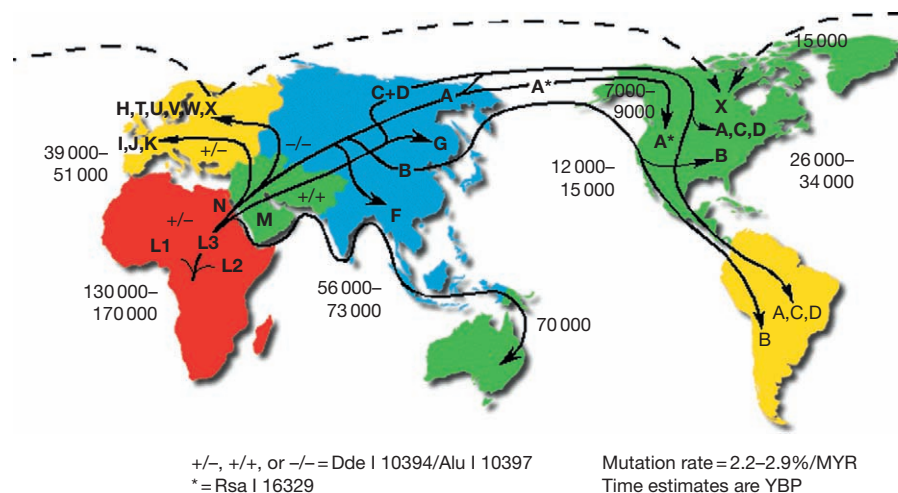


Figure 3 Global radiation of mtDNA types showing the regional association of mtDNA types. Capital letters represent major region groups of related haplotypes (haplogroups). In sub-Saharan Africa, haplogroup L0, which arose between 150 000 and 200 000 years before present (YBP), radiated into haplogroups L1, L2, and L3. At about 65 000 YBP haplogroup L3 gave rise to haplogroups M and N in northeastern Africa. M and N then left Africa in two directions. A southeastern migration moved through Southeast Asia, ultimately giving rise to Australian mtDNAs. A northeastern migration colonized Eurasia, with haplogroup N giving rise to the European mtDNA haplogroups H, I, J, K, T, U, V, W, and X and to the Asian haplogroups A, B, F, etc. Haplogroup M radiated into Asia to give rise to Asian haplogroups C, D, G, etc. Haplogroups A, C, and D became enriched in northeastern Siberia and crossed the Bering land bridge about 20 000 YBP. European-like haplogroup X arrived in Central North America about 15 000 YBP and haplogroup B migrated along the Asian and Bering coasts into southern North America, Central America, and northern South America about 12 000–15 000 YBP. X and B mixed with the haplogroups A, C, D mtDNA to generate the Amerind paleo-Indian population. About 7000–9000 YBP derivatives of haplogroup A crossed the Bering Strait to generate the Na-Dene, and around 3000–5000 YBP derivatives of predominantly haplogroup D but also A and C crossed the Bering Strait to give rise to the Eskimos and Aleuts. The +/–, +/+, –/– represent important M and N diagnostic restriction sites. The ages were calculated using the empirically determined mtDNA sequence evolution rate of 2.2–2.9%/million years (MYR). Adapted from Wallace DC (2007) Why do we still have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine. *Annual Review of Biochemistry* 76: 781–821.

plethora of Asian-specific mtDNA lineages, including A, B, and F from N and C, D, and G from M. Among all of the Asian mtDNA lineages, only A, C, and D successfully occupied arctic Chukotka in northeastern Siberia. When the Bering land bridge appeared, individuals bearing these mtDNAs moved to the Americas about 20 000 YBP. A subsequent coastal migration brought lineage B to mix with A, C, and D about 12 000–15 000 YBP, and another migration across the arctic brought the European haplogroup X to the Great Lakes region of North America about 15 000 YBP. The sum of these migrations generated the Paleo-Indians. A migration from around the Sea of Okhotsk about 7000–9000 YBP brought a modified A to find the Na-Dene. Finally, recent migrations bringing predominantly D but also A and C gave rise to the Eskimos and Aleuts.

Since these regional haplogroups permitted individuals to adapt to local bioenergetic environments, a change in the local environment may result in a mismatch between the mtDNA genotype and the environment, predisposing the individuals to disease. The migration of indigenous people out of their traditional homelands and cultural globalization of the Northwestern high-fat European diet have contributed to an increasing mismatch between mitochondrial genotype and bioenergetic environment. This results in a marked rise in predisposition to a wide range of metabolic and degenerative diseases, including diabetes and obesity, Alzheimer's and Parkinson's disease, cardiovascular disease, predisposition to blindness, sensitivity to infections, tolerance of trauma, etc.

Further exacerbation of the negative effects of the mismatch between mtDNA genes and environment is the rise in toxic industrial and agriculture chemical toxins in the environment, many of which are potent inhibitors of mitochondrial function. These factors interact to contribute to the rapid rise in complex diseases in the industrialized countries.

Disease Causing nDNA-Mitochondrial Gene Mutations

Pathogenic mutations in nDNA-encoded mitochondrial genes fall into two broad categories: mutations that affect the mitochondrial metabolic functions and mutations that affect the biogenesis of the mitochondrion. Deleterious mutations in the structural genes can result in a wide range of clinical symptoms. However, the most well-recognized clinical syndrome associated with function defects in OXPHOS is Leigh syndrome. Leigh syndrome typically affects children that had previously seemed relatively normal. At some point the children experience a rise in temperature and then progressively lose intellectual and motor milestones in association with lesions in the basal ganglia of the brain. In many cases, the children ultimately die. Leigh syndrome has been associated with mutations that inactivate polypeptides for the structure and assembly of key energy-generating enzymes, including pyruvate dehydrogenase and complexes I, II, III, IV, and V. Severe mtDNA mutations can also result in Leigh syndrome.

In addition to structural gene defects, nDNA mutations have been observed that affect mitochondrial biogenesis and turnover. Pathogenic mutations can occur in the mtDNA replication genes mtDNA polymerase γ (POLG) and Twinkle helicase resulting in multiple mtDNA deletions and causing mitochondrial myopathy and neurodegenerative disease,

among other representations. Mutations in the mitochondrial nucleotide metabolism genes for thymidine kinase 2 and deoxyguanosine kinase result in mtDNA depletion and cause various syndromes depending on the organ experiencing the mtDNA depletion, while mutations in the thymidine phosphorylase result in both mtDNA deletions and depletion and cause myoneurogastrointestinal encephalopathy. Mutations in mitochondrial fission and fusion genes such as Opa1 cause optic atrophy and in Mfn2 cause Charcot-Marie-Tooth 2A2 peripheral neuropathy. Finally, mutations in the mitochondrial autophagy (mitophagy) turn over genes PINK1 and parkin cause familial forms of Parkinson's disease.

Diseases Caused by Defects in the Epigenome

While it is commonly asserted that epigenomic changes link the environment to the etiology of complex disease, the environmental factors that modulate the epigenome and the pathophysiological consequences of alterations in the epigenome remain ill defined. The realization that virtually all epigenomic and signal transduction pathway modifications are mediated by reactions using high-energy intermediates implicates energy flux as central to epigenomic regulation and hence to the interaction between the environment and the genome. This role of energy flow in modulating the epigenome and the signal transduction pathways is mediated in two ways. In the first, the mitochondrially generated ATP, acetyl-CoA, and SAM modify and regulate the proteins of chromatin, signal transduction pathways, and metabolism. In the second, the mitochondria regulate cellular oxidation–reduction (redox) status of the cell through the $\text{NAD(P)}^+/\text{NAD(P)H}+\text{H}^+$ ratios, ROS production, and alteration of protein thiol/disulfide equilibrium.

The HIF-1 transcription factor family, which mediates the cellular response to hypoxia, provides an example of the importance of ROS and redox in regulating cellular functions. When activated, HIF-1 upregulates glycolysis and stimulates tissue vascularization. Concurrently, it downregulates mitochondrial function in four ways: it modifies the oxygen interaction of complex IV by switching from the COX4-1 isoform to the COX4-2 isoform, inactivates pyruvate dehydrogenase by phosphorylation, inhibits mitochondrial biogenesis by suppression of c-myc, and stimulates mitochondrial degradation by mitophagy. HIF-1 activity is increased by the stabilization of HIF-1 α , which normally turned over as rapidly as it is synthesized. However, in hypoxia and with increased mitochondrial ROS and succinate production, the enzyme that turns over HIF-1 α , prolylhydroxylase domain protein 2, is inactivated and HIF-1 α is stabilized.

HIF-1 α interacts with HIF-1 β resulting in the active transcription factor. However, once generated, HIF-1 is further modulated through the oxidation–reduction of its thiol/disulfide-coupled amino acids by the mitochondrial and cytosolic redox status which is directly linked to energy flow through the mitochondrion.

Evidence that the epigenome is regulating cell energy metabolism has been found both for cancer and in imprinting diseases. In hepatomas, changes in methylation activate the hexokinase II gene. Hexokinase II then binds to mitochondrial VDAC and coops mitochondrial ATP to drive glycolysis.

In Wilm's tumor and Beckwith–Weideman syndrome (chromosome 11q15.5), epigenomic changes activate the insulin-like growth factor 2 pathway which regulates OXPHOS and glycolysis. In the imprinting disease, Angelman syndrome, due to modifications in the chromosome15q11–13 region, the brain hippocampal cells in the mouse model have been found to have shrunken mitochondria and reduced complex II–III activity. Mutations in the *LMNA*, gene whose protein is concentrated on the inside of the nuclear envelope and is also present in the nucleoplasm, can produce all of the same clinical phenotypes as seen in mitochondrial disease (myopathy and cardiomyopathy, peripheral neuropathy, diabetes and metabolic syndrome, and premature aging). Cells of patients harboring *LMNA* mutations associated with diabetes and metabolic syndrome have been reported to have mitochondrial defects.

Conclusions

For the past 450 years, Western medicine has focused on the anatomical features of disease, and for the past 125 years on genetic traits that behaved according to the Mendel laws. These two features were internally consistent in that all of the anatomical genes are also Mendelian. Belief that these two perspectives encompassed all relevant phenomena has led to two erroneous corollaries: that tissue-specific symptoms are due to tissue-specific defects and that familial traits that do not Mendelize must be environmental. However, life requires anatomy, energy, and information, and the role of energy failure and the non-Mendelian energy genetics in human disease has been essentially overlooked. It is now clear that systemic defects in energy metabolism can cause tissue-specific symptoms, energy diseases can be inherited in a non-Mendelian fashion, and the nongenetic regulation of the nDNA-encoded energy genes based on environmental energy availability and use have a major effect on human characteristics and predisposition to metabolic and degenerative diseases, cancer and aging. These realizations indicate that most of the common metabolic and degenerative diseases are not anatomical but energetic and that these diseases have appeared genetically complex because their inheritance has been interpreted using the Mendelian laws which are only partially applicable.

See also: **Bioenergetics:** Calcium, Biological Fitness of; Cytochrome *bc₁* Complex (Respiratory Chain Complex III); F-ATPases;

Glutathione Peroxidases; Luft's Disease; Membrane Transport, General Concepts; Mitochondrial Calcium Transport: Historical Aspects; Mitochondrial Channels; Mitochondrial DNA; Mitochondrial Dynamics; Mitochondrial Membranes, Structural Organization; Mitochondrial Metabolite Carrier Protein Family; Mitochondrial Outer Membrane and the VDAC Channel; Mitochondrial Permeability Transition Pore; Mitochondria: Source and Target of Free Radicals; Nicotinamide Nucleotide Transhydrogenase; Nuclear Genes Involved in Mitochondrial Function and Biogenesis; Oxygenases; Quinones; Respiratory Chain and Adenosine Triphosphate Synthase; Respiratory Chain Complex I; Respiratory Chain Complex II and Succinate: Quinone Oxidoreductases; Respiratory Chain Complex IV; Spectrophotometric Assays; Structure of Respiratory Complex I; Superoxide Dismutase; V-ATPases.

Further Reading

- Coskun PE, Wyrembak J, Derbereva O, et al. (2010) Systemic mitochondrial dysfunction and the etiology of Alzheimer's disease and Down syndrome dementia. *Journal of Alzheimer's Disease* 20(Suppl 2): S293–S310.
- Lane N and Martin W (2010) The energetics of genome complexity. *Nature* 467(7318): 929–934.
- Wallace DC (1997) Mitochondrial DNA in aging and disease. *Scientific American* 277(2): 40–47.
- Wallace DC (1999) Mitochondrial diseases in man and mouse. *Science* 283(5407): 1482–1488.
- Wallace DC (2005) A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: A dawn for evolutionary medicine. *Annual Review of Genetics* 39: 359–407.
- Wallace DC (2005) Mitochondria and cancer: Warburg addressed. *Cold Spring Harbor Symposia on Quantitative Biology* 70: 363–374.
- Wallace DC (2007) Why do we still have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine. *Annual Review of Biochemistry* 76: 781–821.
- Wallace DC (2010) Bioenergetics, the origins of complexity, and the ascent of man. *Proceedings of the National Academy of Sciences of the United States of America* 107(Suppl 2): 8947–8953.
- Wallace DC and Fan W (2010) Energetics, epigenetics, mitochondrial genetics. *Mitochondrion* 10(1): 12–31.
- Wallace DC, Fan W, and Procaccio V (2010) Mitochondrial energetics and therapeutics. *Annual Review of Pathology* 5: 297–348.
- Wallace DC, Lott MT, and Procaccio V (2006) Mitochondrial genes in degenerative diseases, cancer and aging. In: Rimoin DL, Connor JM, Pyeritz RE, and Korf BR (eds.) *Emery and Rimoin's Principles and Practice of Medical Genetics*, 5th edn., vol. 1, pp. 194–298. London: Churchill Livingstone.

Relevant Website

<http://www.mitomap.org> — MITOWEB: A Human Mitochondrial Genome Database.

Bioenergetics: General Definition of Principles

D G Nicholls, Buck Institute for Age Research, Novato, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 170–173, © 2004, Elsevier Inc.

Glossary

Electrochemical gradient The driving force inherent in an ion gradient across a membrane, made up of two components: the membrane potential and the ion concentration gradient.

Gibbs free energy A thermodynamic term that expresses the extent to which a process is displaced from equilibrium and hence defines its capacity to do work.

Proton circuit The closed cycle comprising protons pumped across, for example, the mitochondrial inner

membrane and reentering via the adenosine triphosphate synthase or a leak pathway.

Proton conductance The conductance of a membrane to protons calculated by applying Ohm's law to the proton circuit.

Redox couple A reduced/oxidized pair, such as nicotinamide adenine dinucleotide hydrogenase/nicotinamide adenine dinucleotide.

Redox potential A thermodynamic measure of the tendency of a redox couple to gain or lose electrons.

Bioenergetics is the study of the molecular mechanisms by which energy, made available by catabolic metabolic pathways or by light capture in photosynthesis, is transformed and made available for cellular processes of growth, motility, and survival. Traditionally, the field has largely focused on studies of bacteria, mitochondria, and chloroplasts, although the principles, largely involving the generation and utilization of ion gradients across membranes, are much more generally applicable.

Thermodynamics

Two factors are required for a biochemical reaction to occur spontaneously: the process should involve an increase in the entropy of the system and its surroundings and there should be a mechanism, usually enzymatic, to lower the energy barrier sufficiently to allow the transformation to occur at a significant rate. The entropy difference is the realm of thermodynamics, whereas the energy barrier helps to define the rate of the process.

The simplest thermodynamics deals with the limiting case of isolated systems that exchange neither material nor energy across their boundaries and, therefore, have no influence on their surroundings. Closed systems exchange energy but not material with their surroundings, whereas real biological systems are open, that is, exchanging both energy and material (substrates and products). The treatment of open systems is complex and requires nonequilibrium thermodynamics; however, the equilibrium thermodynamics of closed systems can be used to obtain important information, such as the conditions under which a given reaction would be at equilibrium and the extent to which the actual reaction in the cell is displaced from equilibrium. Although equilibrium thermodynamics gives no information as to the rate of a reaction, it can unequivocally exclude any process that is energetically impossible.

Gibbs Free Energy

The entropic driving force is equally valid for closed and isolated systems; however, analysis is complicated by the

effects of the energy transfer across the boundaries of the system affecting the entropy of the surroundings. The Gibbs free energy change, ΔG , is a function that, under conditions of constant temperature and pressure, corrects for the entropy change in the surroundings and enables equilibrium to be calculated only using parameters for the system itself. Gibbs energy changes are used not only to study a reaction in solution, but also to quantify reduction–oxidation (redox) reactions occurring, for example, in the mitochondrial respiratory chain, to calculate the available energy in an ion gradient, and to compare the energy available from absorption of light quanta in photosynthesis.

A basic understanding of bioenergetics is considerably helped by the use of simple analogies. Figure 1 shows schematically the content of Gibbs energy (G) in a closed system when the reaction $A = B$ is displaced on either side of equilibrium. The slope of the parabola at any point represents the Gibbs energy change (ΔG). The bottom of the parabola represents the equilibrium condition: the content of Gibbs free energy is at a minimum and the tangent to the slope, ΔG , is zero. Equilibrium is the lowest energy state and a small conversion of substrate to product at equilibrium causes no free energy change. The left side of the parabola represents reaction conditions that have not yet proceeded as far as equilibrium. The slope ΔG is negative, which indicates that the reaction can proceed spontaneously (as long as a pathway exists). The further from equilibrium, the greater is the slope; that is, the Gibbs energy change becomes more negative and more energy is available from a conversion of substrate to product at those concentrations. Beyond the equilibrium point, the slope is uphill. Reactions cannot proceed beyond equilibrium without some independent energy input.

The Gibbs free energy change can be quantified as a function of the displacement from equilibrium

$$\Delta G(\text{kJ mol}^{-1}) = -2.3RT \log_{10} \left(\frac{K}{\Gamma} \right) \quad [1]$$

where R is the gas constant, T is the absolute temperature, K is the equilibrium constant under the conditions of the

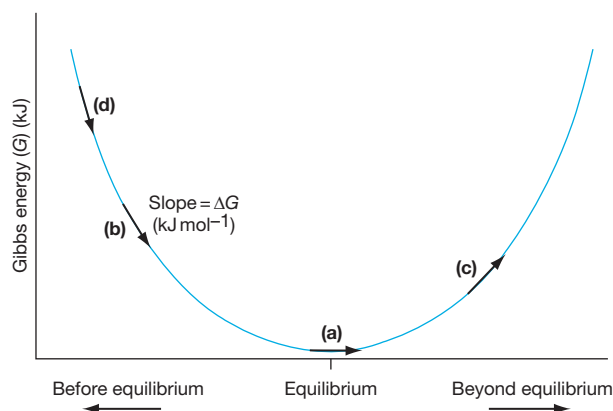


Figure 1 Schematic representation of Gibbs energy in relation to displacement from equilibrium. (a) Gibbs energy (G) is at a minimum at equilibrium and Gibbs energy change (ΔG), given by the slope of the parabola, is zero. (b) Before equilibrium, the slope (ΔG) is negative and the reaction can proceed spontaneously. (c) Beyond equilibrium, the slope is uphill and an additional energy input would be required. (d) The further the reaction is displaced from equilibrium, the greater is the slope; that is, ΔG becomes more negative. Adapted from Nicholls DG and Ferguson SJ (2002) *Bioenergetics* 3. London: Academic Press.

determination, and Γ is the mass-action ratio of the concentrations of substrates and products measured in the cell or assay mixture. Put more simply, ΔG changes by approximately 5.7 kJ mol^{-1} for each 10-fold displacement from equilibrium. Note that this equation corresponds to the more familiar form that involves standard free energy, for example,

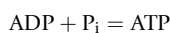
$$\Delta G = \Delta G^0 + 2.3RT \log_{10} \left(\frac{[C]^c [D]^d}{[A]^a [B]^b} \right) \quad [2]$$

but is more logical and symmetrical, emphasizing displacement from equilibrium, rather than the frequently misused abstract concept of standard free energy.

Adenosine Triphosphate

Adenosine triphosphate (ATP) synthesis from adenosine diphosphate (ADP) and phosphate (P_i) and the reverse hydrolysis reaction are the dominant reactions in the distribution and utilization, respectively, of Gibbs free energy in the cell. The ATP synthesis reaction is held up to 10^{10} -fold beyond its equilibrium and thus requires an input of Gibbs free energy of approximately $10 \times 5.7 = 57 \text{ kJ mol}^{-1}$. Conversely, the hydrolysis of 1 mol of ATP can yield up to 57 kJ mol^{-1} . There is no magic property associated with ATP (such as the myth of the high-energy bond) that fits it for this universal role, except that the equilibrium constant of the reaction is such that a sufficient concentration of ADP remains in a functioning cell to bind to the ATP synthase (see below).

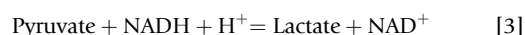
The ATP synthesis reaction is usually expressed in shorthand form, ignoring water, ionization, pH, and chelating ions such as Mg^{2+} ; that is,



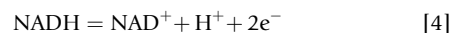
This may only be used as the basis for thermodynamic calculations if the equilibrium constant is known under exactly identical conditions, in which case the common factors cancel out of the ratio Γ/K in eqn [1].

Redox Potentials

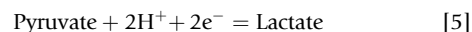
Redox reactions are those that involve coordinate oxidation of one substrate and the reduction of another. An example of a redox reaction occurring in the cytoplasm is that catalyzed by lactate dehydrogenase:



Redox reactions can be considered hypothetically as the sum of two linked half-reactions:



and



All biological redox reactions can be considered to involve primary electron transfer, although in many cases the increased negative charge on the reduced component leads to the subsequent binding of one or more protons; thus, ubiquinone reduction to ubiquinol in the respiratory chain involves the addition of two electrons followed by two protons, whereas NAD^+ reduction to NADH requires the addition of two electrons followed by one proton (eqn [5]).

A reduced/oxidized pair, such as nicotinamide adenine dinucleotide hydrogenase/nicotinamide adenine dinucleotide (NADH/NAD^+), is a redox couple and the equilibrium of the half-reaction can be defined in terms of its half-reduction redox potential (the potential at which the concentrations of oxidized and reduced species are equal). A redox couple with a more negative redox potential will tend to oxidize one with a more positive potential. Zero is defined as the potential of a $\frac{1}{2}\text{H}_2/\text{H}^+$ couple at pH 0 when H_2 is 1 atm. Important half-reduction potentials include those for $\text{NADPH}/\text{NADP}^+$ and NADH/NAD^+ (both -320 mV), fumarate/succinate ($+30 \text{ mV}$), ubiquinol/ubiquinone ($+60 \text{ mV}$), and reduced/oxidized cytochrome c ($+220 \text{ mV}$). The important $\text{H}_2\text{O}/\frac{1}{2}\text{O}_2$ couple has a redox potential of $+820 \text{ mV}$ when O_2 is present at 1 atm and water is 55 M.

Redox potentials are not constants, but vary with the state of reduction of the couple. The redox potential of a one-electron reduction couple becomes 60 mV more negative for each 10-fold increase in the reduced/oxidized ratio. Thus, although the standard hydrogen electrode at pH 0 has a redox potential of zero, at pH 7 when the proton concentration is only 10^{-7} M , the redox potential is -420 mV . Two electron reductions change by 30 mV per decade; thus, the NADH/NAD^+ couple in mitochondria is typically 10% reduced and, therefore, has a redox potential of approximately $-320 + 30 = -290 \text{ mV}$, whereas the $\text{NADPH}/\text{NADP}^+$ pool is at least 90% reduced and so operates at approximately -350 mV .

Although the historical origin of redox potentials (in electrochemistry) may obscure the relationship, redox changes are governed by Gibbs free energy principles; indeed, the ΔG

associated with the difference in redox potential between two couples can be simply calculated as

$$\Delta G = -nF\Delta E_h \quad [6]$$

where n is the number of electrons transferred and F is the Faraday constant ($0.965 \text{ kJ mol}^{-1} \text{ mV}^{-1}$). As a rough rule of thumb, a mole of electrons falling through a potential of 1 V makes 100 kJ of Gibbs free energy available, and 2 mol of electrons passing from NADH to O_2 through the respiratory chain corresponds to a ΔG of -224 kJ . Note that this is compatible with the formation of approximately 3 mol of ATP.

Ion Electrochemical Potential Differences

The concept of Gibbs free energy as displacement from equilibrium can be extended to gradients of ions and solutes across biological membranes. Two components must be considered: that due to a difference in concentration between the two aqueous compartments and (if the species is charged) that due to the difference in electrical potential between the compartments (usually termed the membrane potential, not to be confused with any surface charge on the membrane itself). The Gibbs free energy is simply the sum of the two components (hence electrochemical):

$$\Delta G(\text{kJ mol}^{-1}) = -mF\Delta\psi + 2.3RT\log_{10} \frac{[X^{m+}]_B}{[X^{m+}]_A} \quad [7]$$

where m is the charge on the ion, $\Delta\psi$ is the membrane potential in millivolts, and X^{m+} is the concentration of the ion in compartment A or B.

The ion of most interest in bioenergetics is the proton. The equation for the proton electrochemical gradient ($\Delta\mu_{\text{H}^+}$) is rather simple, due to the logarithmic nature of pH:

$$\Delta\mu_{\text{H}^+}(\text{kJ mol}^{-1}) = -F\Delta\psi + 2.3RT\Delta\text{pH} \quad [8]$$

The mitochondrial convention is that ΔpH is the pH outside minus that in the matrix and is usually negative.

The proton electrochemical gradient is usually expressed in units of voltage. This proton-motive force, Δp , is defined as

$$\Delta p (\text{mV}) = -(\Delta\mu_{\text{H}^+})/F = \Delta\psi - 60\Delta\text{pH} \quad [9]$$

Photons

The Gibbs free energy available from a single photon is equal to Planck's constant times frequency, that is, $h\nu$, where ν is the frequency of the radiation. A mole (Avogadro's number) of photons corresponds to a ΔG of $120\,000 \lambda \text{ kJ mol}^{-1}$, where λ is the wavelength in nanometers. For 600-nm red light, this corresponds to 200 kJ mol^{-1} . Note that the absorption of such a photon would theoretically be capable of lowering (making more negative) the potential of an electron by 2 V.

The Thermodynamics of Bioenergetic Interconversions

The bioenergetic processes occurring in mitochondria involve the sequential conversion of redox energy into proton-motive force and of proton-motive force into the Gibbs free energy of the ATP pool. As each is based on Gibbs free energy, the interconversion is very simple. Consider a respiratory chain complex that transfers

two electrons through a redox span of $\Delta E_h \text{ mV}$. If the complex pumps n protons against a proton-motive force of $\Delta p \text{ mV}$, equilibrium would be reached when $n\Delta p = 2\Delta E_h$. For the complex to work in the forward direction, therefore, $n\Delta p > 2\Delta E_h$. Roughly speaking, the forward rate of the complex is proportional to the disequilibrium for small displacements, which is why respiration accelerates when the proton gradient is lowered.

The equilibrium condition for the ATP synthase, where Δp is expressed in kilojoules per mole and Δp is expressed in millivolts, requires the Faraday constant, $\Delta G_{\text{ATP}} = mF\Delta p$, where m represents the number of protons required to generate 1 mol of ATP. As the ATP synthase is reversible, it can operate in the forward direction (ATP synthesis) when $\Delta G_{\text{ATP}} < mF\Delta p$ and in the reverse direction (ATP hydrolysis) when $\Delta G_{\text{ATP}} > mF\Delta p$.

Mitochondrial Respiration Rate: The Proton Circuit

A simple but immensely powerful aid to understanding the relationships between proton-motive force, respiration, and proton reentry pathways is to use a simple electrical analogy

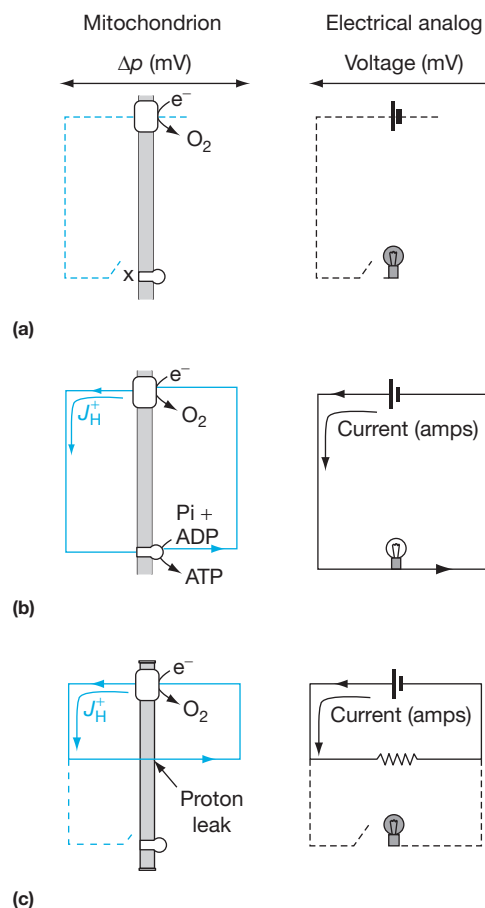


Figure 2 The analogy between the proton circuit and a simple electrical circuit. (a) Open circuit. Zero current (no respiration), potential (Δp) maximal. (b) Circuit completed, current flows (respiration). Useful work done (ATP synthesized). Potential (Δp) less than maximal. (c) Short-circuit introduced; energy dissipated, potential low, respiration maximal. Adapted from Nicholls DG and Ferguson SJ (2002) *Bioenergetics 3*. London: Academic Press.

(Figure 2). Because there is a fixed stoichiometry between electron flow (and hence respiration) and proton translocation, and also because protons must reenter the mitochondrial matrix at exactly the same rate that they are pumped out, the proton current, J_{H^+} , circulating around the proton circuit, will equal the rate of respiration times the H^+/O (protons pumped divided by oxygen utilized) stoichiometry of the respiratory chain.

Because the current and potential (Δp) in the proton circuit are known, Ohm's law can be used to calculate the resistance (or its reciprocal, i.e., the conductance, $C_{\text{M}}\text{H}^+$) of the inner membrane to proton reentry:

$$C_{\text{M}}\text{H}^+ = J_{\text{H}^+} / \Delta p \quad [10]$$

Two major components of proton conductance are that which occurs through the ATP synthase, which is tightly coupled to the rate of ATP synthesis, and an endogenous proton leak possessed by all mitochondria. One specialized tissue, brown adipose tissue, possesses an additional controllable

proton leak catalyzed by an uncoupling protein that allows respiration, and heat production, to occur without concomitant ATP synthesis.

See also: **Bioenergetics:** ATP in Plant Mitochondria: Substrates, Inhibitors, and Uncouplers; Chloroplast Signaling and Redox Control; Energy Transduction in Anaerobic Bacteria; Membrane-Associated Energy Transduction in Bacteria and Archaea.

Further Reading

- Cramer WA and Soriano GM (2003) Thermodynamics of energy transduction in biological membranes. In: *Biophysics Textbook Online*, available at <http://www.biophysics.org/btol/>. Biophysical Society.
- Nelson DL, Cox MM, and Lehninger AL (2000) *Principles of Biochemistry*, 3rd edn. New York: Worth.
- Nicholls DG and Ferguson SJ (2002) *Bioenergetics 3*. London: Academic Press.

Biotin

S W Polyak and A Chapman-Smith, The University of Adelaide, Adelaide, SA, Australia

© 2013 Elsevier Inc. All rights reserved.

This article is a reproduced from the previous edition, volume 1, pp. 174–178, © 2004, Elsevier Inc.

Glossary

Biocytin Biotinylated lysine which is formed between the carboxyl group of biotin and the epsilon amino group on the side chain of lysine.

Biotin A small, water-soluble vitamin belonging to the B-complex group; it is also known as vitamin H.

Biotinidase A mammalian enzyme that can hydrolyze biotin from either biocytin or short

biotin-containing peptides, thus facilitating recycling of the vitamin.

Biotin protein ligase The enzyme [E.C. 6.3.4.15] responsible for covalent attachment of a biotin moiety onto a biotin-enzyme; it is also known as the biotin-inducible repressor (BirA) in some prokaryotes and holocarboxylase synthetase in mammalian cells.

Historical Perspective

At the end of the nineteenth century, it was observed that extracts from liver and meat could serve as effective treatments for skin lesions induced by a diet of raw egg. Earlier studies had identified a heat-stable factor in the same extracts that was also required for yeast growth. In 1935, Kogl and Tannis isolated this factor from egg yolk and called it biotin. The essential growth factor present in the meat extracts, which had become known as vitamin H (*Haut*, 'skin' in German), was found to be identical to biotin. Later, it was discovered that avidin, a biotin-binding glycoprotein, was the toxic component of raw eggs. The structure of biotin was determined in 1942 and confirmed by chemical synthesis and X-ray analysis.

During the 1950s, it became evident that biotin is involved as a cofactor for a class of enzymes that catalyze a variety of carboxyl transfer reactions. It was first demonstrated that β -methylcrotonyl-coenzyme A (CoA) carboxylase, a biotin-requiring enzyme that catalyzes a step in the degradation of leucine, was capable of carboxylating free biotin. It soon became apparent that, in the cell, biotin functions as a cofactor only when bound to protein and that biotin is covalently attached to the ϵ -amino group of a specific lysine residue of propionyl CoA carboxylase. Subsequently, it was demonstrated for other biotin enzymes that the biotin prosthetic group is attached in the same way.

Biotin

Biotin, a member of the water-soluble B-complex group of vitamins, is synthesized by higher plants and most fungi and bacteria. Humans, other mammals, and birds cannot synthesize biotin *de novo* and therefore must obtain this essential micronutrient from material synthesized by intestinal microflora and from dietary sources. In mammals, absorption of biotin occurs in the small intestine. Other important sites that play a role in normal biotin physiology include the liver (the principal site of biotin utilization), kidney (the site of reabsorption of filtered biotin), and placenta (the site of

transport of biotin from the maternal circulation to the developing embryo). Biotin obtained from dietary sources exists in both free and protein-bound forms. Protein-bound biotin is digested by gastrointestinal proteases and peptidases to biocytin and biotin-containing short peptides. Biotin present in these compounds is recycled through the action of biotinidase to release free biotin.

Biotin Enzymes

Biotin enzymes are a family of enzymes ubiquitously found throughout the nature. It appears that all organisms contain acetyl CoA carboxylase, reflecting the essential role this enzyme plays in the synthesis of fatty acids required for membrane biogenesis. While some organisms, such as bacteria, possess only one biotin enzyme, other organisms have multiple biotin enzymes. For example, mammalian cells express pyruvate carboxylase, propionyl-CoA carboxylase, and methylcrotonyl-CoA carboxylase, in addition to two isoforms of acetyl-CoA carboxylase. These enzymes catalyze key metabolic reactions in gluconeogenesis and amino acid metabolism, in addition to fatty acid metabolism.

The members of this enzyme family utilize biotin as a mobile carboxyl carrier in a conserved reaction mechanism that requires three domains: a biotin carboxylase domain, a transcarboxylase (TC) domain, and a biotin domain containing the covalently attached biotin moiety. As shown in [Figure 1](#), these enzymes catalyze reactions in two partial reactions carried out at spatially separate sites on the protein where biotin itself carries a carboxyl group between the two sites. Biotin enzymes can be divided into three classes depending on the nature of the original donor and final carboxyl acceptor. These are the carboxylases (class I), decarboxylases (class II), and TCs (class III). All eukaryotic enzymes belong to class I, while prokaryotes contain enzymes from all the three classes.

The mechanism of biotin carboxylation in the class I enzymes has been extensively investigated. Experiments on propionyl CoA carboxylase and pyruvate carboxylase showed that the

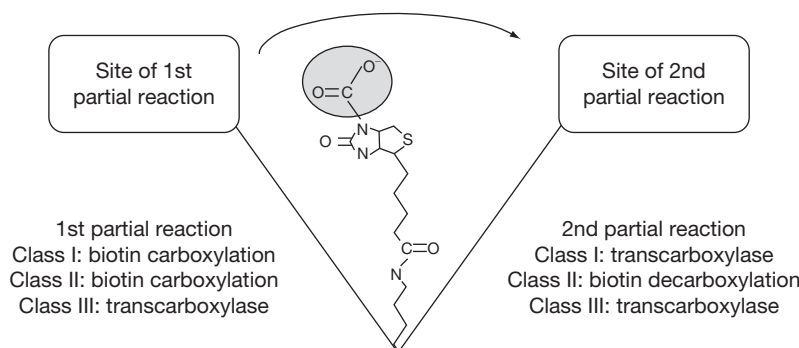
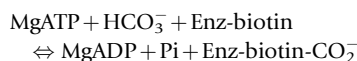


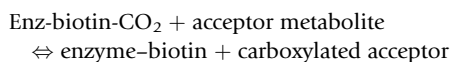
Figure 1 Biotin-mediated carboxyl transfer between reaction subsites of biotin enzymes. The biotin group attached to the target lysine is shown located in the cleft of the biotin enzyme. The carboxylate anion, complexed to the biotin, is shown in the shaded circle. The arrow represents movement of the mobile cofactor between the two partial reaction subsites during catalysis. The reactions performed at the separate subsites are indicated for the three classes of biotin enzymes.

source of CO_2 in the reaction is bicarbonate and the reaction is dependent upon the presence of adenosine triphosphate (ATP):



At the first partial reaction site, biotin is carboxylated at the 1'-nitrogen probably after the formation of a very labile enzyme-bound carboxyphosphate intermediate that donates the activated carboxyl group to biotin. Although no direct evidence of this reaction mechanism has been reported, several observations suggest this as the most likely scheme. Most notable are the similarities between the biotin enzymes and carbamoyl phosphate synthetase, an enzyme known to utilize a carboxyphosphate intermediate.

Subsequently, in the second partial reaction, the transfer of the carboxyl group from carboxybiotin to the acceptor molecule occurs. It has been proposed that the decarboxylation of carboxybiotin proceeds along a pathway where CO_2 and the enolate of biotin are formed. Following protonation, carboxylated biotin can again be formed at the first reaction site under physiological conditions:



The class II decarboxylases differ from the class I carboxylases in that the reactions are ATP independent. The decarboxylases of anaerobic prokaryotes all catalyze decarboxylation of a specific β -keto acid or acyl-CoA coupled to sodium ion export against a concentration gradient. Thus, this class of enzyme serves as an important energy transducer that does not require ATP. There is only one known member belonging to the class III biotin enzymes: *Propionibacterium freudenreichii* subsp. *shermanii* TC. TC catalyzes the formation of methylmalonyl-CoA from propionyl-CoA using oxaloacetate as a carboxyl donor, and it exists as a multimeric complex composed of 30 polypeptides of three different types.

Biotin Domains

The biotin prosthetic group is covalently attached via the ϵ -amino group of one specific lysine residue within the biotin

enzymes. This biotin-accepting lysine is found in a tetrapeptide sequence, Ala-Met-Lys-Met, which is extremely conserved among all biotin enzymes (Figure 2). In addition, the primary structure surrounding the target lysine residue shows a high degree of homology between a wide range of enzymes and species (Figure 2). Deletion studies have revealed that 30–35 amino acid residues on either side of biocytin are necessary to specify biotinylation. Collectively, these structures, able to incorporate biotin *in vivo*, are termed biotin domains. The removal of determinants necessary to define the structure of a biotin domain by truncation or mutation results in a molecule that is unable to be biotinylated.

The biotin domain participates in a number of heterologous protein-protein interactions in the cell. First, it serves as a substrate in the biotinylation reaction. Here, the biotin prosthetic group is covalently attached to the biotin domain through the enzymatic action of biotin protein ligase (BPL). Subtle conformational changes to the biotin domain that occur upon biotinylation are thought to signal dissociation of the two proteins and yield the biotinylated product. The biotinylated or holo biotin domain is then free to interact with each of the two partial reaction sites in the carboxylase, shuttling carboxyl groups between substrates in the enzyme complex.

The formation of multimeric protein complexes, characteristic of all biotin enzymes, is not necessary for substrate recognition by BPLs. Enzymatic biotinylation experiments, performed using the biotin accepting subunit of the *P. shermanii* TC or the biotin carboxyl carrier protein (BCCP) of *Escherichia coli* acetyl CoA carboxylase, have shown that these domains function equally well as BPL substrates as do the intact or multimeric protein complex. Therefore, the information required for association with BPL is present within the structured biotin domain. BPLs from various sources have been found to recognize and biotinylate acceptor proteins from very different sources.

In 1966, McAllister and Coon first showed that extracts containing BPLs from rabbit liver, yeast, and *P. shermanii* were able to activate enzyme substrates from rabbit and bacteria via the attachment of the biotin prosthetic group. Evidence of cross-species recognition *in vivo* was demonstrated when the 1.3S subunit of TC was recombinantly expressed in *E. coli* and shown to be a substrate for the *E. coli* BPL. Truncation analysis of the 1.3S TC subunit revealed that the minimum amount of

BCCP <i>E. coli</i>	<u>HIVRSP</u> ■ <u>VMGTFY</u> ■ <u>RTP</u>	SPDAKAFIEV■GQKVN	VGDTLCIVEAMKMMN	■ <u>QIEADKSG</u> ■ <u>TVKAILV</u>	■ <u>ESQQPVEF</u> ■ <u>DEPLVVIE</u>
<i>P. aerug.</i> ACC	NVVRSP■ <u>VMGTFY</u> ■ <u>RAA</u>	SPTSANFVEVGQSVK	KGDILCIVEAMKMMN	HIEAEVSGTIESILV	ENGQPVEFDQPLFTIV
Rat PCC	SVLRSPKPGV■VVAVS	-----VKPGDMVA	EGQEICVIEAMKMN	SMTAGKMGKVLVHC	KAGDTVGEGDLLVEL-
Human PCC	SVLRSPMPGV■VVAVS	-----VKPGDAVA	EGQEICVIEAMKMN	SMTAGKTGTVKSVHC	QAGDTVGEGDLLVEL-
Rat PC	GQIGAPMPGKVIDIK	-----VVAGAKVA	KGQPLCVLSAMKMET	VVTSPMEGTVRKVHV	TKDMTLEGDDLILEIE
Human PC	GQIGAPMPGKVIDIK	-----VVAGAKVA	KGQPLCVLSAMKMET	VVTSPMEGTVRKVHV	TKDMTLEGDDLILEIE
Chicken ACC	SILRSPSAGKLIQYV	-----VEDGGHVF	AGQCFAEIEVMKMVM	TLTAGESGCIHYVKR	P-GAVLDPGCVIAKLQ
Human ACC	SVMRSPSAGKLIQYI	-----VEDGGHVL	AGQCYAEIEVMKMVM	TLTAVESGCIQYVKR	P-GAALDPGCVLAKMQ
Yeast ACC	TQLKTPSPGKLVKFL	-----VENGEHI I	KGQPYAEIEVMKQM	PLVSQENGIVQLLKQ	P-GSTIVAGDIMAIMT
<i>P. sherm.</i> TC	<u>GEIPAPLAGT</u> ■ <u>VS</u> ■ <u>SKIL</u>	-----VKEGDTVK	AGQTVLVLEAMKMET	EINAPT■ <u>DGKVEKVLV</u>	KERDAVQGGQGLIKIG

Figure 2 Sequence alignment of the biotin domains of biotin carboxylases from diverse organisms. Residues forming β -strands in the three-dimensional structures of *Escherichia coli* biotin carboxyl carrier protein (BCCP) and *Propionibacterium freudenreichii* subsp. *shermanii* transcarboxylase (TC) are underlined, and hydrophobic core residues are indicated by (•). The biotinylated lysine is marked (♦). Shading indicates residues very highly conserved in all biotin domains for which sequence data are available. Reproduced from Chapman-Smith A and Cronan JE, Jr. (1999) The enzymatic biotinylation of proteins: A post-translational modification of exceptional specificity. *Trends in Biochemical Sciences* 24: 359–363.

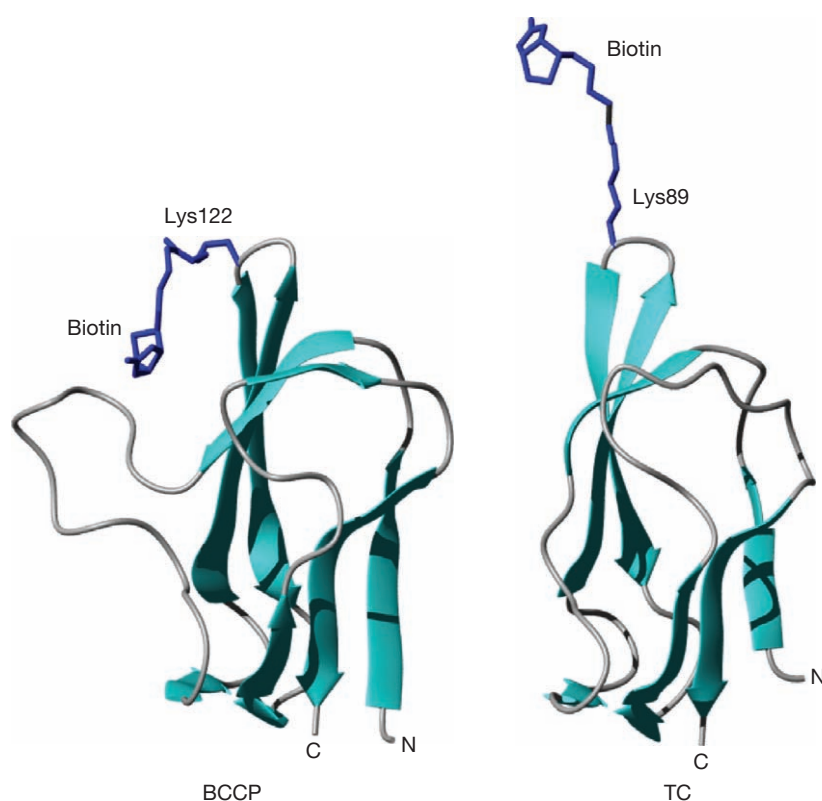


Figure 3 Three-dimensional structures of two biotin domains. The structures of the biotin domains from the biotin carboxyl carrier protein (BCCP) of (left) *Escherichia coli* acetyl CoA carboxylase and (right) the 1.3S subunit of *Propionibacterium freudenreichii* subsp. *shermanii* transcarboxylase (TC) have been determined. The holo forms of the two proteins with the biotin moiety specifically attached to the target lysine residues at position 122 and 89, respectively, are depicted. Solid arrows represent the β -strands in the fold. The amino (N) and carboxyl (C) termini of the domain are indicated. The molecules have been orientated to highlight the interaction of biotin with the thumb structure in BCCP.

information required to specify biotinylation was present in the 75 C-terminal amino acid residues. This minimal peptide, fused to β -galactosidase, was biotinylated *in vivo* by the BPL from *E. coli*. Similarly, the C-terminal 87-amino-acid residue of BCCP (BCCP-87) was shown to be a stable protein that can function as a substrate for *E. coli*, insect, yeast, and human BPLs. Together, these studies suggest that all biotin domains fold into an essentially common tertiary structure recognized by all members of the BPL enzyme family.

Biotin Domain Structure

The structures of two biotin domains have been determined: that of the *E. coli* BCCP-87 (Figure 3) and the 1.3S subunit of *P. shermanii* TC. These proteins are structurally related to the lipoyl domains of 2-oxo acid dehydrogenase multienzyme complexes, which also undergo an analogous posttranslational modification. These domains form a flattened β -barrel

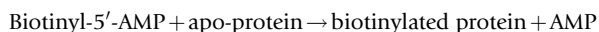
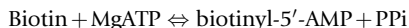
structure comprising two four-stranded β -sheets with the N- and C-terminal residues close together at one end of the structure. At the other end of the molecule, the biotinyl- or lipoyl-accepting lysine resides on a highly exposed, tight hairpin loop between $\beta 4$ and $\beta 5$ strands. The structure of the domain is stabilized by a core of hydrophobic residues, which are important structural determinants. Conserved glycine residues (Figure 2) occupy β -turns linking the β -strands.

BCCP-87 contains a seven-amino-acid insertion common to prokaryotic acetyl-CoA carboxylases but not present in other biotin domains. This region of the peptide adopts a thumb structure between the $\beta 2$ and $\beta 3$ strands and, interestingly, forms direct contacts with the biotin moiety in both the crystal and solution structures. The significance of this interaction is not understood, but structural studies on the 1.3S subunit of TC, which is a thumbless biotin domain, have revealed that biotin also interacts with the protein. Residues in the N-terminal region of the TC subunit, independent of the biotin domain, functionally compensate for the thumb structure by binding to biotin only when the cofactor is in its carboxylated state. Therefore, it is possible that the mechanism employed by the biotin enzymes may involve noncovalent interactions between the protein and the prosthetic group.

Biotin Ligase

Reaction Mechanism

All BPLs catalyze biotinylation through the same reaction mechanism:



In the first partial reaction, BPLs bind biotin and ATP in an ordered manner (depending on the species) to form an enzyme:biotinyl-5'-AMP complex or holo-BPL. At this stage, the carboxyl group of inert biotin is activated by the addition of an adenylate group forming the intermediate biotinyl-5'-AMP. In the second partial reaction, holo-BPL complexes with an apo-biotin domain. Nucleophilic attack upon the activated biotin by the amine group of the biotin-accepting lysine residue results in the transfer of biotin from the adenylate onto the biotin domain with AMP acting as the leaving group. This two-step process is analogous to the tRNA synthetase-catalyzed linkage of amino acids to the 3' hydroxyl group of tRNAs.

E. coli Biotin Ligase, BirA

The best-characterized BPL is the 35.3-kDa protein BirA from *E. coli*. This protein is a bifunctional molecule which, in addition to BPL activity, acts as the transcriptional repressor of the biotin biosynthetic operon. By combining the regulation of biotin synthesis and metabolism through the actions of a single protein, bacteria have evolved an exceptional mechanism for maintaining biotin homeostasis. Recent structural studies of BirA, together with enzymatic and genetic studies, have provided powerful insights into BirA action. The crystal structure of BirA shows that the protein is composed of three domains: a DNA-binding N-terminal domain, a central catalytic domain, and a C-terminal domain (Figure 4). The central

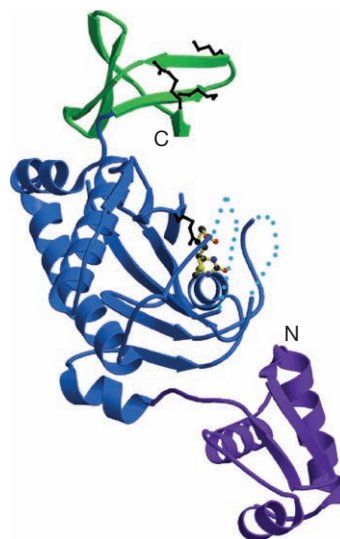


Figure 4 The three-dimensional structure of BirA. The diagram shows the structure of *Escherichia coli* biotin protein ligase, BirA, determined by X-ray crystallography. This figure highlights the three functional domains in the bifunctional protein. The N-terminal domain (N), shown in purple, contains a helix-loop-helix fold to facilitate DNA binding. The central catalytic domain (shown in blue) is depicted with biotin (ball and stick representation) bound at the active site. Dots represent the unstructured loops around the active site that become ordered during the reaction. The function of the C-terminal (C) domain (shown in green) is not fully understood. Modified from Chapman-Smith A, Mulhern TD, Whelan F, Cronan JE, Jr., and Wallace JC (2001) The C-terminal domain of biotin protein ligase from *E. coli* is required for catalytic activity. *Protein Science* 10: 2608–2617.

domain contains several poorly defined loops that become more ordered upon binding of biotin or biotinyl-5'-AMP. Ordering of these loops in the complex results in the generation of an extended β -sheet, which is required for the interaction with other proteins. Thus, when complexed with biotinyl-5'-AMP, BirA is competent to participate in one of the two competing interactions: either with apo-BCCP or with the 40-base-pair operator sequence within the biotin synthesis operon. In the presence of apo-BCCP, monomeric BirA preferentially interacts with this protein substrate and catalyzes biotinylation. In the absence of apo-BCCP, the BirA complex forms the homodimer structure required for DNA binding and thus functions as a transcriptional repressor. Therefore, when *E. coli* needs to produce additional active acetyl-CoA carboxylase, its biotin biosynthetic pathway can be derepressed by the expression of apo-BCCP, thereby allowing the bacteria to balance its nutritional requirements.

See also: **Protein/Enzyme Structure Function and Degradation: Biotinylation of Proteins.**

Further Reading

Athappilly FK and Hendrickson WA (1995) Structure of the biotinyl domain of acetyl-coenzyme A carboxylase determined by MAD phasing. *Structure* 3: 1407–1419.

- Attwood PV (1995) The structure and the mechanism of action of pyruvate carboxylase. *International Journal of Biochemistry and Cell Biology* 27(3): 231–249.
- Attwood PV and Wallace JC (2002) Chemical and catalytic mechanisms of carboxyl transfer reactions in biotin-dependent enzymes. *Accounts of Chemical Research* 35(2): 113–120.
- Chapman-Smith A and Cronan JE, Jr. (1999) The enzymatic biotinylation of proteins: A post-translational modification of exceptional specificity. *Trends in Biochemical Sciences* 24: 359–363.
- Chapman-Smith A, Mulhern TD, Whelan F, Cronan JE, Jr. and Wallace JC (2001) The C-terminal domain of biotin protein ligase from *E. coli* is required of catalytic activity. *Protein Science* 10: 2608–2617.
- McMahon RJ (2002) Biotin in metabolism and molecular biology. *Annual Review of Nutrition* 22: 221–239.
- Pacheco-Alvarez D, Solorzano-Vargas RS, and Del Rio AL (2002) Biotin in metabolism and its relationship to human disease. *Archives of Medical Research* 33(5): 439–447.
- Reddy DV, Shenoy BC, Carey PR, and Sonnichsen FD (2000) High resolution solution structure of the 1.3S subunit of transcarboxylase from *Propionibacterium shermanii*. *Biochemistry* 39: 2509–2516.
- Roberts EL, Shu N, Howard MJ, et al. (1999) Solution structures of Apo and Holo biotinyl domains from acetyl coenzyme A carboxylase of *Escherichia coli* determined by triple-resonance nuclear magnetic resonance spectroscopy. *Biochemistry* 38: 5045–5053.
- Weaver L, Kwon K, Beckett D, and Matthews BW (2001) Competing protein: Protein interactions are proposed to control the biological switch of the *E. coli* biotin repressor. *Protein Science* 10: 2618–2622.
- Wilson KP, Shewchuk LM, Brennan RG, Otsuka AJ, and Matthews BW (1992) *Escherichia coli* biotin holoenzyme synthetase *bio* repressor crystal structure delineates the biotin- and DNA-binding domains. *Proceedings of the National Academy of Sciences of the United States of America* 89: 9257–9261.

Biotinylation of Proteins

R A Kohanski, Johns Hopkins University, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 179–181, © 2004, Elsevier Inc.

Glossary

Avidin A protein that binds avidly to biotin ($K_d < 10^{-15}$ M) and is found in egg whites. A related bacterial protein is streptavidin, from *Streptomyces avidinii*.

Biocytin A biotin derivative formed between biotin and lysine. It is formed by the catabolism of carboxylases, and is

itself degraded to biotin and lysine by the enzyme biotinidase.

Biotin A water-soluble vitamin that is used by carboxylases for transfer of CO₂ groups in biosynthesis. It is bound tightly by the proteins avidin and streptavidin.

The biotinylation of proteins is the covalent coupling of biotin to an amino acid or carbohydrate moiety of the protein. Biotinylation occurs in a specific group of proteins known as carboxylases. These are enzymes that are important in several metabolic pathways, including amino acid metabolism, fatty acid biosynthesis, and gluconeogenesis. Each of these carboxylases is biotinylated on a single lysyl residue through the action of a biotin:protein ligase (BPL), and are found in all organisms (bacteria, plants, and animals). Biotinylation of proteins is also done by *in vitro* chemical synthesis, whereby biotin can be covalently coupled to any reactive functional group on the protein, and is not necessarily restricted to a single amino acid residue.

Analytical and Clinical Applications

Chemically biotinylated proteins are important tools in cell biology and biochemistry, and have clinical applications. The uses of chemically biotinylated proteins have evolved from the strong interaction between biotin (and its derivatives) and avidin or streptavidin. The medical importance of biotinylated proteins is twofold: (1) Life-threatening diseases result from failure to produce adequate levels of endogenous biotinylated proteins. These diseases are particularly serious in newborns and early childhood. (2) The diagnosis and treatment of tumors has been greatly facilitated by chemically biotinylated proteins. As an analytical reagent in cell biology and biochemistry, chemical biotinylation of proteins is in widespread use not only primarily for protein detection and assay, but also for purification. These are considered in greater detail in this article.

Biosynthesis and Degradation of Natively Biotinylated Proteins

The biotinylated proteins native to cells are the biotin-dependent carboxylases: acetyl CoA carboxylase, beta-methylcrotonyl CoA carboxylase, propionyl CoA carboxylase, and pyruvate carboxylase. The first is a cytosolic protein and the latter three are in the mitochondria. In bacteria, the biotinylated protein is biotin carboxyl carrier protein which is a subunit of

bacterial carboxylases. The biosynthesis of naturally biotinylated proteins occurs in two steps. Biotin is activated by reaction with adenosine triphosphate, releasing pyrophosphate and biotinyl 5'-adenosine monophosphate. The hydrolysis of pyrophosphate makes this step essentially irreversible under physiological conditions. The transfer of biotin to an ϵ -amino group of a specific lysyl residue in the acceptor protein is catalyzed by holocarboxylase synthetase (HS) in eukaryotes and BPL in bacteria. The specific lysyl residue that is biotinylated is found within a highly conserved structural motif of the apoprotein – the biotin domain – that ranges in length from 67 to 85 amino acids. The biotin domain is an independently folding motif with a core structure of eight β -strands forming two apposing antiparallel β -sheets. Conservation of the catalytic portions of the HS/BPL enzymes and of the biotin domains are so high across evolution that bacterial BPL can recognize and biotinylate mammalian biotin domains, and mammalian HS can recognize and biotinylate bacterial biotin domains. The consensus amino acid sequence surrounding the acceptor lysyl residue is Glu-Ala-Met-Lys-Met, but only the Met-Lys-Met residues are invariant. The conservation of the methionyl residues appears to be dictated by evolutionary pressure to retain a catalytically active carboxylase rather than a recognition site for the synthetase, since Thr or Val can be substituted for these Met residues with little loss of biotinylation but more significant losses in catalytic efficiency of the biotinylated product. In general, it appears that conservation of the β -sheet structure of the biotin domain is as critical for recognition by the ligase as the consensus sequence, but the latter is governed by the need to conserve optimal enzymatic activity in the biotinylated proteins. Thus, the two aspects controlling specificity of protein biotinylation are: first, recognition by the ligase and second, enzymatic activity of the biotinylated protein. Biotinylated proteins are degraded by cellular proteases, and the biotinylated lysyl residue (biocytin) is the product of that degradation. Biotinidase is the enzyme that cleaves biocytin to biotin and lysine.

Medical Importance of Biotinylated Proteins

Diseases caused by deficiencies in the biotin-dependent carboxylases fall into two general categories based on whether biotin

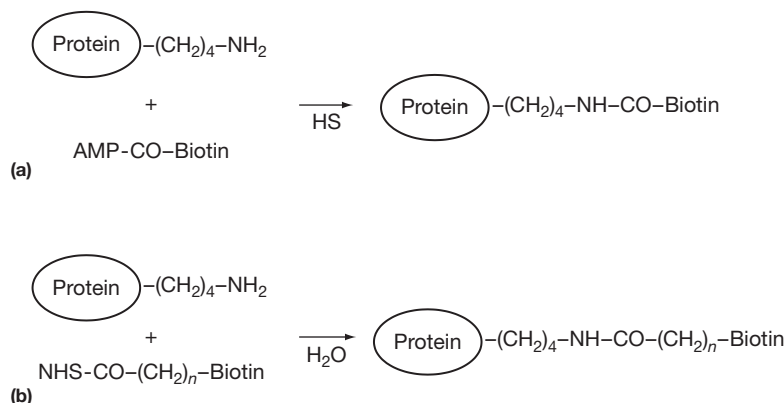


Figure 1 Biotinylation of proteins by enzymatic and chemical reactions. (a) The transfer of biotin from AMP to the ϵ -amino group of a lysyl residue in a protein, catalyzed by holocarboxylase synthetase (HS) forming an amide linkage. (b) The transfer of biotin, with a hydrocarbon spacer of 'n' methylene groups from an N-hydroxysuccinimide ester (NHS) to an ϵ -amino group of a protein in water.

administration is an effective treatment. Those not treatable with biotin are due to genetic lesions in the biotin-dependent carboxylases (OMIM 200350, 210200, 232050, 232000, and 266150 which are described at the National Institutes of Health website), whereas those treatable with biotin are the result of genetic lesions in the synthetase (multiple carboxylase deficiency; OMIM 253270) or biotinidase (OMIM 25360). Clinically, the synthetase and biotinidase deficiencies are distinguished by the early (up to 3 months postpartum) versus late onset of symptoms, respectively. Mutations in holocarboxylase generally cause a raised Michaelis constant for biotin. Thus, the treatment with pharmacological doses of biotin is probably successful because the enzyme only requires higher concentrations of biotin for efficient biotinylation of the apoprotein carboxylases. The success of biotin administration when biotinidase is lost indicates the biological importance of recycling biotin from biocytin to maintain sufficient levels of carboxylase activity. The other clinically significant aspect of biotinidase is its ability to degrade radiolabeled biotin derivatives, which are used in conjunction with biotinylated antibodies to image and/or treat tumors.

Chemical Synthesis of Biotinylated Proteins

The usual goal of protein biotinylation is to detect the modified protein with avidin, where the avidin is itself linked to a reporter group (a radionuclide, chromophore, fluorophore, or an enzyme). The detection of biotinylated proteins has been important in Western blots in order to determine the expression of proteins in diverse cells exposed to different stimuli or growth conditions, for the immunocytochemical and immunohistochemical localization of proteins within subcellular compartments or in tissues, and for the identification and tracking of surface-biotinylated proteins. In addition, the immobilized avidin (and derivatives) can be used for affinity capture and purification of biotinylated proteins, or in the capture of receptors using biotinylated ligands. Biotinylation of proteins can be achieved through covalent attachment of a biotin derivative to any freely reactive functional group on a protein. The most commonly used sites include primary amines

(α - or ϵ -amino groups of the N-terminal or lysyl residues, respectively, as illustrated in Figure 1(b)), thiol groups (of cysteinyl residues), and carbohydrate groups. The reactive derivatives of biotin for these reactions use N-hydroxysuccinimide esters, iodoacetyl, or disulfide derivatives of biotin, and hydrazide derivatives of biotin, respectively. The critical factor in the useful biotinylation of proteins is allowing avidin to bind the biotinyl group after conjugation. This requires sufficient distance between the surface of the biotinylated protein and avidin, since biotin is bound in a fairly deep pocket in avidin. This problem was solved by Klaus Hofmann, who found that an ϵ -aminocaproyl group offered a sufficiently long spacer group for this purpose, as illustrated in Figure 1. The use of sulfo-N-hydroxysuccinimidyl-esters (sulfo-NHS esters) for biotin derivitization introduced by James Staros further simplified the use of biotin derivatives because it increased the water solubility and amine specificity of the reactive ester. Importantly, the sulfo-NHS esters cannot cross cell membranes, and this permits the selective labeling of cell-surface proteins through biotinylation of their extracellular domains.

Clinical Application of Biotinylated Antibodies

Clinically, biotinylated antibodies are in current use for the detection and treatment of tumors, with success in many cases. The multistep procedure can be summarized as follows: biotinylated antibodies against a tumor-specific antigen are introduced into circulation. Binding to the antigen concentrates the biotinylated antibody at the site of the tumor. Avidin or streptavidin is introduced, which binds the pretargeted antibody. Because these are tetramers, at least one biotin-binding site is expected to remain unoccupied in the avidin:biotinylated antibody complex. A biotin derivative that chelates a radionuclide is then introduced, which is trapped by the open biotin-binding sites on avidin at the site of the tumor. This protocol can be used to visualize the tumor *in vivo* by immunoscintigraphy or positron emission tomography. With the appropriate radionuclide, radiotherapy of the tumor is also achievable. In such protocols, circulating biotinidase plays a counterproductive role, since it can and does break down biocytin

analogs. Thus, knowledge of the biotin cycle of endogenously biotinylated proteins has been important in the clinical application of chemically synthesized biotinylated proteins.

See also: **Protein/Enzyme Structure Function and Degradation:** Biotin.

Further Reading

Boerman OC, van Schaijk FG, Oyen WJG, and Corstens FHM (2003) Pretargeted radioimmunotherapy of cancer: Progress step by step. *Journal of Nuclear Medicine* 124: 400–411.

Lunain EJ (2001) Biotinylation of proteins in solution and on cell surfaces. In: Coligan JE, Dunn JE, Ploegh HL, Speicher DW, and Wingfield, PT (eds.) *Current Protocols in Protein Science*, 1st edn., vol. 3, supplement 6, pp. 3.6.1–3.6.15. New York, NY: Wiley.

Wilchek M and Bayer EA (eds.) (1990) *Methods in Enzymology: Avidin–Biotin Technology*, vol. 184. San Diego, CA: Academic Press.

Wilchek M and Bayer EA (eds.) (1999) *Biomolecular Engineering (Strept) Avidin–Biotin System*, vol. 16. San Diego, CA: Academic Press.

Wolfin B (2001) Disorders of biotin metabolism. In: Scriver CR, Beaudet AL, Sly WS, and Valle D (eds.) *The Metabolic and Molecular Bases of Inherited Disease*, 8th edn., vol. 3, pp. 3935–3962. New York, NY: McGraw-Hill.

Relevant Website

<http://www.ncbi.nlm.nih.gov> – National Center for Biotechnology Information: Online Mendelian Inheritance in Man (OMIM).

Bmp Signaling and Vascular Disease

J Lowery, Harvard School of Dental Medicine, Boston, MA, USA
M de Caestecker, Vanderbilt University, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Bone morphogenetic proteins A family of secreted ligands originally identified by their ability to induce ectopic bone formation which have since been shown to have diverse functions during embryonic development and tissue homeostasis.

Clade A group of genes that are related to a common ancestral gene.

Co-receptors A cell-surface receptor that binds to a signaling molecule (BMP ligand or receptor) in addition to

the primary receptors (type I and II receptors) to promote ligand recognition and/or receptor activation.

Ortholog Homologous genes that are found in different species.

SMAD A family of intracellular proteins that mediate canonical signaling from TGF- β superfamily of receptors to the nucleus. SMAD proteins are the mammalian orthologs of the *Caenorhabditis elegans* protein, SMA, and the *Drosophila* protein, Mothers Against Decapentaplegic, MAD (the name SMAD is a combination of SMA and MAD).

Bmp Signaling

Bmp Family of Ligands

Bone morphogenetic proteins (Bmps) are secreted transforming growth factor-beta (TGF- β) superfamily ligands first identified in extracts from bone matrix that could induce ectopic bone formation when implanted subcutaneously in rats. It is now known that Bmps play essential roles in the development of nearly all vertebrate organs, including the embryonic vasculature. Within the TGF- β superfamily, Bmp ligands (some of which are also referred to as growth and differentiation factors (Gdfs)) can be organized into clades based on sequence similarity to conserved molecules in primitive organisms (Figure 1). This helps to understand differences in receptor binding and cellular responses between Bmps and other ligands within this family. On this basis, 15 *bona fide* mammalian Bmp/Gdf ligands have been identified and shown to activate typical Bmp-dependent responses. Some of these responses are context dependent. For example, the more distantly related Gdf15 group of ligands can not only activate Bmp-dependent responses but also act as inhibitors of Bmp signaling in a context-dependent fashion. Other ligands such as Gdf8, Gdf11, Gdf1, and Gdf3 are misnamed since they activate pathways more typical of the TGF- β /activin ligands, while others such as Bmp3 (and inhibin) should be classified as Bmp antagonists, since they act by sequestering inactive Bmp receptor signaling complexes.

Regulation of Bmp ligand expression

Secretion of Bmp ligands is regulated at a variety of levels. Transcription of *Bmp2*, *Bmp4*, *Bmp5*, and *Gdf6* messenger RNA (mRNA) is regulated by long-range, modular tissue-specific *cis*-regulatory elements within their genomic loci. RNA expression is also regulated by gene-specific posttranscriptional regulation of mRNA stability. Secretion of mature Bmps is dependent on correct intracellular cleavage of pro-domains. This is regulated in a tissue-specific manner, indicating that secretion of biologically active ligands is regulated independently of their mRNA expression. Unlike TGF- β s, the cleaved pro-domain of most Bmps dissociates as the mature ligand

is secreted. However, Bmp7 and Bmp9 are secreted as stable complexes with their cleaved pro-domains which regulate extracellular matrix interactions, or independently activate Bmp receptors. These regulatory mechanisms have implications for studying the role of Bmp ligands in the vasculature, and reinforce the need to evaluate mature protein expression where possible (although there is a dearth of reliable antibodies that can detect tissue expression of mature Bmp ligands *in vivo*).

Extracellular Regulation of Bmp Ligands

Most Bmp ligands are secreted as biologically active dimers. These regulate local cellular responses, but a number of biologically active ligands are also detectable at nanogram per milliliter concentrations in the circulation (including Bmp4, Bmp6, and Bmp9). Given their close association with endothelial cells (ECs), these may play a role in maintaining vascular homeostasis. A number of Bmp antagonists and other non-receptor-binding proteins regulate long-range effects of Bmp ligands. These include Bmp antagonists Noggin, Twisted gastrulation (Tsg), Follistatin, and members of the Chordin and DAN family of proteins. In addition, membrane-associated heparan sulfite proteoglycans and components of the extracellular matrix, including type IV collagen, fibrillin and matrix Gla-protein, interact with and modify specific Bmp ligand activity in a context- and concentration-dependent manner. The mechanisms by which these extracellular factors regulate Bmp signaling have been extensively investigated during embryonic development, where spatiotemporal control of Bmp gradients plays a critical role in controlling developmental patterning. Many of these proteins are also expressed in the adult, and are likely to play a role in regulating Bmps expressed in the vasculature. However, there are little data on the extracellular regulation of Bmps in the adult.

These proteins inhibit Bmp activity by sequestering ligands away from their receptors, reducing free movement in the extracellular space and/or promoting endocytic uptake and degradation of the ligands. Ligands bind with varying affinity to these inhibitors. For example, Noggin binds to Bmp2,

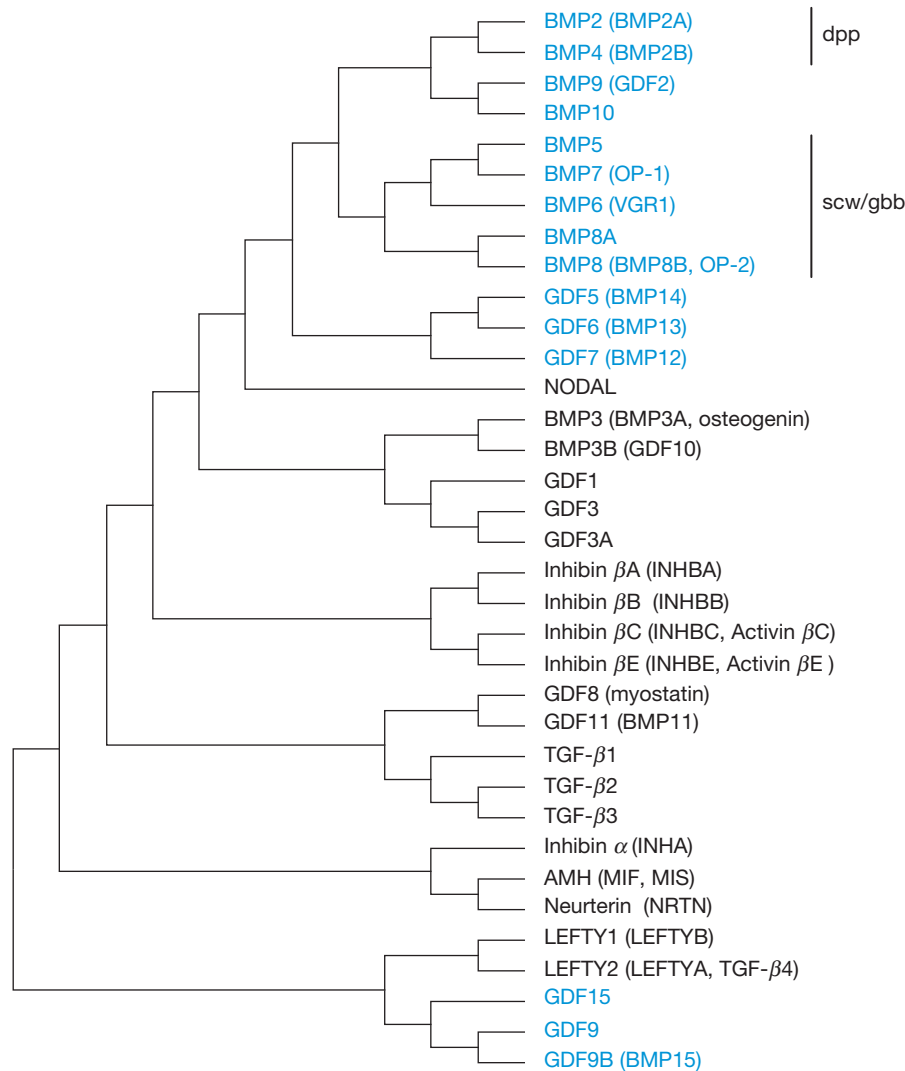


Figure 1 Phylogenetic comparison of TGF- β superfamily ligands based upon full-length human protein sequences. Sequence similarity to *Drosophila* Bmp orthologs is indicated. Alternative names are listed in parentheses. Ligands in blue are *bona fide* Bmps that have been shown to activate Bmp Smads. Reproduced from Lowery JW and de Caestecker MP (2010) BMP signaling in vascular development and disease. *Cytokine and Growth Factor Reviews* 21: 287–298.

Bmp4, Bmp7, and Gdf5 with high affinity but does not inhibit Bmp9 or 10 activity at all. Chordin binds to Bmp2, Bmp4, and Bmp7 but does not interact with other Bmp family members. These effects are more complex in the intact organism where different binding proteins interact. For example, Tsg forms a tri-molecular Chordin/Bmp/Tsg complex that inhibits Bmp4 by stabilizing Chordin–Bmp interactions. However, Tsg also enhances Bmp signaling by promoting cleavage of Chordin from the Chordin–Bmp complex by the protease Bmp1/Tolloid. (Bmp1 was identified as part of the original screen for bone matrix-derived extracts with bone-inducing activity, but is not a member of the Bmp family.) This sets the stage to establish a gradient of Bmp activity where high concentrations of Chordin antagonize Bmp activity, while relative excess of Tsg releases Bmp from the Chordin complex and activates of Bmp signaling. Bmper/Cv2 also inhibits Bmp4 at high concentrations (promoting endocytic uptake and degradation), but activates Bmp signaling at low concentrations (facilitating

ligand interaction with receptors). This phenomenon whereby Bmp-binding proteins can activate Bmp signaling in a context- and concentration-dependent fashion has implications for investigators exploring the mechanisms by which Bmps have different short- and long-range effects on the vasculature.

Bmp Receptors

Signal transduction is initiated by binding to receptors and formation of heterotetrameric complexes comprising combinations of type 1 (activin-like kinases (Alk) 1/2/3 and 6) and type 2 Bmp receptors (Bmpr2, Actr2a, and Actr2b). Both types of receptors have extracellular ligand binding, transmembrane and cytoplasmic serine/threonine kinase domains. Proximity of the type 2 receptor kinase domain with the type 1 receptor occurs upon ligand-induced oligomerization and promotes phosphorylation of serine and threonine residues in the GS box N-terminal to the kinase domain of the type 1 receptor.

This alters conformation of the type 1 receptor, activating the kinase by allowing adenosine triphosphate (ATP) to enter the kinase pocket. Bmpr2 is unique in having a long C-terminal tail that may regulate Bmp-dependent responses by promoting degradation of the Smad ubiquitin ligase, Smurf1. Tail domain interactions with cyclic guanoside monophosphate-dependent kinase 1 and LIM kinase 1 may also enhance downstream responses to Bmps. In general, however, the activated kinase domain of type 1 Bmp receptors phosphorylates and activates a subclass of signaling intermediates, the Bmp-activated Smads (Smad1, Smad5, and Smad8). This distinguishes the TGF- β /activin-activated type 1 receptors (Alk4, Alk5, and Alk7) which activate a different class of Smads (Smad2 and Smad3).

Bmp ligand–receptor interactions

Bmp type 1 receptors are divided into two groups based on sequence similarities: Alk3/6 and Alk1/2. These have distinct ligand-binding properties (Table 1) but show promiscuity in binding to different members of the TGF- β superfamily. Alk3 and Alk6 interact with and are activated by Bmp2 and 4 equally, while Bmp5–8 group ligands principally activate Alk2. Bmps also have different affinities for type 1 and type 2 receptors. Bmp2 and 4 bind to Alk3 and 6 but only recruit type 2 receptors, Bmpr2, Actr2a, and/or Actr2b once bound to these type 1 receptors. In contrast, Bmp5–8 group ligands bind with high affinity to Actr2a and Actr2b and subsequently recruit type 1 receptors into active signaling complexes. Bmp9 behaves like Bmp2/4 and Bmp5–8 in different cells: it binds directly to Alk1 in ECs but only interacts with Alk2 in the presence of type 2 receptors in other cell types. Some Bmps, such as Gdf9, interact with Bmpr2, and also recruit the TGF- β receptor Alk5 into active signaling complexes. Moreover, TGF- β can recruit Alk1, Alk2, and Alk3 into heteromeric complexes with the type 1 TGF- β receptor Alk5. This accounts for the coincident activation of Smad1/5/8 and Smad2/3 by TGF- β in a variety of cell types. Different TGF- β ligands may also exert opposing effects by signaling through the same receptors. Bmp9 also activates Alk1 in ECs, but, unlike TGF- β /Alk1, this sustained Smad1/5/8 signal exerts anti-angiogenic effects in ECs. In addition, while Bmpr2 only interacts with Bmp family members, Actr2a and Actr2b interact with activins, Gdf8 and Gdf11, which recruit TGF- β /activin type 1 receptors into signaling complexes. This inhibits Bmp responses by sequestering type 2 receptors that also interact with Bmp ligands. Some TGF- β family ligands, including Bmp3 and inhibin, also inhibit Bmp signaling by sequestering type 2 receptors without activating downstream signals. Furthermore, Gdf9b(Bmp15), which activates Bmp-Smads in granulosa cells, inhibits Smad phosphorylation in other cell types by forming complexes with Bmp4 and likely inhibiting receptor binding. This context-dependent inhibitory function is shared by Bmps that lack a cysteine residue required for ligand dimerization, including Gdf9 and Gdf3. Bmp ligands can also be secreted as hetero-dimers; presenting dual ligand-receptor binding sites that promote higher receptor binding affinities and/or recruit more complex assemblies of Bmp receptors. Studies in Zebrafish embryos show that Bmp2 and 7 form heterodimers when over-expressed at physiologically relevant concentrations. These recruit and activate heteromeric complexes of type 1 Bmp receptors Alk3 and Alk8 (Alk8 is the Zebrafish orthologue of Alk2). It remains to be established

Table 1 Demonstrated type I and type II receptor interactions for *bona fide* Bmp ligands

	Receptor	Ligand
Type I	Alk1 (Acvr11)	Bmp9 (Gdf2)
		Bmp10
	Alk2 (Acvr1)	Bmp6 (Vgr1)
		Bmp7 (OP-1)
	Alk3 (Bmpr1A)	Bmp9 (Gdf2)
		Bmp2 (Bmp2A)
		Bmp4 (Bmp2B)
		Bmp5
		Bmp6 (Vgr1)
		Bmp7 (OP-1)
		Bmp10
		Gdf5 (Bmp14)
		Gdf6 (Bmp13)
		Gdf9
Type II	Alk6 (Bmpr1B)	Bmp2 (Bmp2A)
		Bmp4 (Bmp2B)
		Bmp6 (Vgr1)
		Bmp7 (OP-1)
		Bmp10
		Gdf5 (Bmp14)
		Gdf6 (Bmp13)
		Gdf9
		Gdf9B (Bmp15)
		Bmp2 (Bmp2A)
Type II	Bmpr2	Bmp4 (Bmp2B)
		Bmp6 (Vgr1)
		Bmp7 (OP-1)
		Bmp9 (Gdf2)
		Bmp10
		Gdf5 (Bmp14)
		Gdf6 (Bmp13)
		Gdf9
		Gdf9B (Bmp15)
		Bmp2 (Bmp2A)
Type II	Actr2A (Acvr2A)	Bmp4 (Bmp2B)
		Bmp6 (Vgr1)
		Bmp7 (OP-1)
		Bmp9 (Gdf2)
		Gdf9
		Bmp2 (Bmp2A)
		Bmp4 (Bmp2B)
		Bmp6 (Vgr1)
		Bmp7 (OP-1)
		Bmp9 (Gdf2)
Type II	Actr2B (Acvr2B)	Bmp2 (Bmp2A)
		Bmp4 (Bmp2B)
		Bmp6 (Vgr1)
		Bmp9 (Gdf2)
		Bmp2 (Bmp2A)
		Bmp4 (Bmp2B)
		Bmp6 (Vgr1)
		Bmp9 (Gdf2)
		Bmp2 (Bmp2A)
		Bmp4 (Bmp2B)

Alternative names are listed in parentheses.

whether Bmps are actually secreted as heteromeric ligands endogenously (due to lack of effective reagents). However, these experiments provide insight into the diversity of receptor usage by Bmps *in vivo*.

Bmp Co-Receptors

A number of membrane-associated co-receptors have been identified, including the repulsive guidance molecule (RGM) family, betaglycan, and endoglin, which enhance Bmp signaling by modifying Bmp ligand–receptor interactions (Table 2).

Table 2 Co-receptors of BMP signaling

Co-receptor	Partner
Endoglin (CD105)	Bmp2 (Bmp2A) Bmp7 (OP-1) Bmp9 (Gdf2)
TGF β R3 (betaglycan)	Bmp2 (Bmp2A) Bmp4 (Bmp2B) Bmp7 (OP-1) Gdf5 (Bmp14)
RGMa	Bmp2 (Bmp2A) Bmp4 (Bmp2B) Gdf7 (Bmp12)
RGMb (DRAGON)	Bmp2 (Bmp2A) Bmp4 (Bmp2B) Gdf7 (Bmp12)
RGMc (hemojuvelin)	Bmp2 (Bmp2A) Bmp4 (Bmp2B) Bmp6 (Vgr1) Gdf7 (Bmp12)
Bambi (Nma)	Alk1 Alk3 Alk6 Actr2
TrkC	Bmpr2
Ror2	Alk6 (Bmpr1B)

Alternative names are listed in parentheses.

In addition, the type 1 TGF- β family pseudo-receptor Bambi/Nma interacts with a number of type 1 and 2 Bmp receptors, and is thought to inhibit Bmp signaling by interfering with the assembly of activated receptor signaling complexes. However, *Bambi* knockout mice do not have any obvious developmental defects, suggesting that these inhibitory effects are likely to be minor *in vivo*.

Betaglycan/TGF β R3 and endoglin/CD105 are related transmembrane proteoglycans that also modify Bmp-receptor interactions. Both have large extracellular domains and short cytoplasmic tails lacking kinase activity. Endoglin and betaglycan undergo proteolytic cleavage, with the soluble extracellular domain being released into the extracellular space and antagonizing the effects of the membrane-associated co-receptors. Soluble endoglin and betaglycan are detected in the circulation both normally and in disease states.

Betaglycan

Betaglycan is ubiquitously expressed and is the major cell surface TGF- β -binding protein, and also interacts with Bmp2, Bmp4, Bmp7, Gdf5, and inhibin. The classical mechanism by which betaglycan enhances Bmp signaling is by increasing Bmp ligand binding to Alk3 and Alk6. However, betaglycan also inhibits Bmp2 and Bmp7 signaling by sequestering Bmpr2 and Actr2 into inactive complexes containing inhibin, and enhances Bmp signaling through its ability to regulate trafficking and cell localization of interacting receptors. For example, through its interaction with the scaffolding protein, β -arrestin2, betaglycan binds to and promotes internalization of Alk6, thereby selectively enhancing Alk6-dependent signaling.

Endoglin

Endoglin is dominantly expressed in ECs, but, like betaglycan, it interacts with other TGF- β family ligands, including Activin A, Bmp2, Bmp7, and Bmp9. Most of these interactions are dependent on the presence of the respective ligand binding TGF- β superfamily receptors, with the exception of Bmp9, with which it can bind directly. These interactions are required for TGF- β to activate Alk1-Smad1/5 but inhibit Alk5 signaling in ECs. Endoglin also enhances Bmp7 and Bmp9-dependent Smad1/5 responses, although the mechanism by which this occurs is largely unknown. Endoglin does not enhance nor is it required to generate type 1/type 2 receptor complexes. However, Alk5 phosphorylates the cytoplasmic tail of endoglin and this is required for endoglin to activate both TGF- β and Bmp9-dependent Alk1 signaling in ECs, suggesting that the cytoplasmic tail plays an essential role in activating Alk1-signaling. This may be of significance in vascular disease since there is increased expression of S-endoglin, a splice variant that lacks this cytoplasmic tail, in senescent ECs, and transgenic over-expression of S-endoglin promotes endothelial dysfunction.

RGM/DRAGON family of Bmp co-receptors

The RGM GPI-anchored proteins RGMa, RGMb/Dragon, and RGMc/HJV, originally identified as regulators of axonal guidance, are widely expressed outside the nervous system and bind to a number of Bmps. All three RGMs interact with Bmp2, Bmp4, and Gdf7 and enhance Smad1/5 signaling. RGMc undergoes proteolytic cleavage, releasing a soluble form that inhibits the membrane-associated co-receptor. Unlike RGMa, membrane-associated RGMc activates Bmp6-dependent responses and plays an essential role in regulating Bmp6-dependent control of iron metabolism *in vivo*. The mechanism by which RGMs enhance Bmp signaling is unclear. RGMa increases binding of Bmp2 to Alk3 and Actr2a, while both RGMa and RGMc increase utilization of Actr2a by Bmp2 and Bmp4 in the presence of Bmpr2. This suggests that RGMs modify the composition of activated receptor complexes by altering binding affinities of Bmp ligands for their receptors.

Smad Signaling

GS box phosphorylation of the type 1 receptor kinase increases affinity of the type 1 receptor for its Smad substrate and promotes phosphorylation of a conserved SXS motif at the C-terminus of receptor activated Smads (RA-Smads). This requires trafficking into early endosomes where the type 1 receptor encounters SMAD anchor for receptor activation (SARA) and Hgs (which recruit Smad2 to Alk5), or endofin (which recruits Smad1 to Alk3 and Alk6). Phosphorylation of RA-Smads destabilizes RA-Smad-SARA/endofin interactions, exposing nuclear import sequences in the Smad N-terminus. SXS motif phosphorylation also increases RA-Smads affinity for Smad4, giving rise to heteromeric RA-Smad/Smad4 complexes that translocate to the nucleus and activate transcriptional responses. Smad4 is required for most, but not all, Smad-dependent responses. Signaling specificity is determined by sequences in the L45 loop of the type 1 receptors. Invariant sequences in the L45 loop of Alk4/Alk5 and Alk7 interact with conserved sequences in the C-terminus of Smad2/3, while distinct L45 loop sequences in Alk3/6,

and Alk 1/2, interact with conserved residues in Smad1/5/8. This paradigm holds true in many contexts. However, recent studies indicate that TGF- β -induced activation of Smad1/5/8 is not always dependent on Bmp type 1 receptors, but can result from recruitment of Smad1/5/8 to activated TGF β R2/Alk5 complexes. Conversely, while the mechanisms remain unclear, both Bmp2 and 9 can phosphorylate and activate Smad1/5/8 and Smad2. This indicates that while the two-pathway model of Smad signaling is widely applicable, the Bmp versus TGF- β signaling paradigm is not as straightforward as originally thought.

Regulation of Smad signaling

The duration and intensity of Smad signaling are regulated at a number of different levels. Caveolar-mediated endocytosis and association with inhibitory Smads (I-Smads: Smad6 and Smad7) promote proteosomal degradation of Bmp and TGF- β receptor complexes and RA-Smads by Smad ubiquitin ligases Smurf1 and Smurf2. I-Smads are up-regulated by TGF- β and Bmps, thus providing feedback inhibition. Smad6 inhibits Bmp signaling by selectively interacting with Alk1, Alk2, Alk3, and Alk6, while Smad7 lacks type 1 receptor specificity and inhibits both Bmp and TGF- β responses. Smad signaling is also regulated by type 1 receptor de-phosphorylation. For example, PP1 is recruited to activated receptors by SARA and endofin, promoting receptor de-phosphorylation. In the nucleus, the phosphatases Ppm1A and Scp1-3 de-phosphorylate RA-Smads. This dissociates RA-Smad/Smad4 complexes, exposing nuclear export signals and enables recycling of Smad4 and RA-Smads to the cytoplasm. Other posttranslational modifications, including sumoylation, mono-ubiquitination, and acetylation, regulate nucleocytoplasmic shuttling of TGF- β -activated Smad2/3, but it is unclear whether they regulate Bmp signaling. In addition, a number of kinases regulate Smad activity by phosphorylating the linker and N-terminal domains of RA-Smads. Epidermal growth factor (EGF)- and fibroblast growth factor-induced extracellular signal regulated kinase (ERK) mitogen-activated protein kinase (MAPK) activation and ultraviolet (UV)-induced activation of p38 MAPK and JUN N-terminal kinase (JNK) promote phosphorylation of serine residues in the Smad1 linker. This inhibits Smad1 activity by preventing nuclear import and promoting Smurf1-dependent Smad degradation. Some of these phosphorylation events activate Smad signaling: Smad2 phosphorylation induced by hepatocyte growth factor and EGF stimulates Smad activity, while serotonin phosphorylates and activates Bmp-dependent Smad1/5 in vascular smooth muscle cells (VSMCs). Moreover, linker phosphorylation of C-terminally phosphorylated Smad1 by cyclin-dependent kinases CDK8 and CDK9 recruits YAP1 to the Smad1/Smad4 complex and enhances transcription. On this basis, sequential phosphorylation of Smads provides a mechanism by which different signaling pathways can limit (or enhance) the duration and intensity of Smad signaling.

Smad-dependent transcriptional responses

Once in the nucleus, the RA-Smad/Smad4 complexes activate transcriptional responses directly by binding to Smad-binding elements (SBE), or indirectly through interactions with DNA-binding transcription factors. DNA binding affinity is relatively weak, so recruitment of Smad complexes often depends on

synergistic interactions with other transcription factors. Given the diversity of transcriptional binding partners, this allows for cross-talk with a variety of signaling pathways. For example, Bmp-dependent target gene activation by Smad1/Smad4 is activated by β -catenin (mediator of Wnt responses) when recruited to Lef1 DNA-binding transcription factor. Similarly, Smad1 interaction with Notch intracellular domain (NICD) recruits Bmp-Smads to the RBP-Jk binding site (mediator of Notch responses) in the Herp2 promotor, thereby synergistically activating a Notch target gene in ECs.

In addition to these direct effects of Smad proteins on gene transactivation responses, there is evidence that Smad proteins can activate transcription by regulating histone modifications and reorganizing chromatin structure. *In vitro* transcription studies indicate that Smad-dependent transcription only occurs on chromatin templates. Moreover, Smads acetylate histones by recruiting histone acetyl transferases, CBP and p300, and repress transcription by recruiting repressor complexes with histone deacetylase (HDAC) activity. For example, Bmp-Smads interact with Evi-1 which recruits an HDAC complex to transcriptional targets. Bmp-Smads also recruit the co-repressor Ski to promoters, disrupting Smad interaction with CBP and p300 and recruiting a repressor complex containing HDAC1. Since expression of Smad-binding repressors is regulated, RA-Smads have the capacity to activate or repress transcription depending on availability of these co-repressor molecules.

Nontranscriptional Smad-dependent responses

Recent studies indicate that Bmp and TGF- β activated Smads regulate microRNA processing by directly interacting with Drosha complexes. Bmp signaling has also been shown to regulate transcription of a variety of microRNAs. However, unlike these Smad-dependent transcriptional responses, regulation of microRNA processing is Smad4-independent. This pathway is of importance since microRNA processing is required for their nuclear export and function, and there is evidence that a single mature microRNA can regulate expression of hundreds of proteins. This mechanism has recently been shown to play an essential role in regulating Bmp-dependent VSMC differentiation.

Smad-Independent Signaling

Different limbs of the MAP kinase (ERK, p38 MAPK, and JNK), PI3 kinase/AKT, and PKC signaling pathways, and a variety of Rho-GTPases (direct regulators of cytoskeletal dynamics), can also be activated by Bmps in a context-dependent fashion. Some of these responses require Smad-dependent transcription. Others occur rapidly, are Smad-independent, and result from signaling directly downstream of Bmp receptors. This is supported by the observation that Dorsomorphin, a small molecule inhibitor of Bmp type 1 receptor kinases, inhibits Smad1/5 activation but has no effect on Bmp-induced p38 MAPK phosphorylation in VSMCs. Activation of Smad versus non-Smad-signaling is dependent on the order of Bmp receptor assembly. Over-expression studies indicate that, unlike TGF- β receptors, Bmp receptors can form heteromeric complexes in the absence of ligand. These preformed receptor complexes (PRCs) activate Smad-dependent responses following treatment with Bmp ligands, while Bmp-induced signaling

complexes (BISCs) preferentially activate non-Smad-dependent responses (p38 MAPK in these studies). Moreover, deletion of the C-terminal tail of Bmpr2 interferes with formation of PRCs, suggesting that the Bmpr2 tail may play a role in promoting Bmp-Smad signaling. The mechanism by which these complexes promote selective Smad versus non-Smad-dependent responses is unclear. Furthermore, given the paucity of reagents to study endogenous Bmp receptor behavior, it is unknown whether PRCs occur endogenously, or whether they represent an artifact of protein over-expression.

Links between Smad-independent responses and Bmp receptors are poorly understood. Some are dependent on which type 1 and type 2 receptors are engaged. For example, Bmp-dependent activation of the Rho GTPase Cdc42 regulates cytoskeletal dynamics by phosphorylation of cofilin (promoting actin polymerization). This is dependent on interaction of Limk1 (a cofilin kinase) with the C-terminal tail of Bmpr2. In addition, Bmp7-dependent activation of PI3 kinase in cultured monocytes is dependent on expression of Bmpr2 and Actr2a but not Actr2b. One of the better-characterized Smad-independent pathways is mediated by the MAPK activators TAK1 and TAB1. TAK1 and TAB1 are required to mediate TGF- β and Bmp-dependent activation of p38 MAPK and JNK, and associate with Alk3 through an adaptor protein XIAP. However, TAK1 has also been shown to interact with Bmp-Smads and induces activating SXS phosphorylation of Smad1/5/8 (but not TGF- β activated Smad2/3). This indicates that TAK1 has distinct functions in regulating both branches of the Bmp signaling pathway, and suggests an alternative mechanism by which Bmp signaling activates Smads.

Bmp Signaling in Vascular Development and Disease

Bmp ligands, inhibitors, receptors, and downstream signaling intermediates show distinct patterns of regulation during vascular development and disease. However, given the complex nature of this signaling pathway, definitive functional data based on *in vivo* genetic data are essential to understand the functional implication of these changes. The second part of this article outlines the results of these studies from a molecular rather than a disease-oriented perspective, focusing on selective molecular components of the Bmp signaling pathway that have been properly evaluated *in vivo*. We draw parallels between defects in different vascular beds, stages of development, and adult disease resulting from mutations in different components of the Bmp signaling pathway. We have summarized the relevant vascular phenotypes of humans and mice carrying defined mutations of Bmp pathway components in [Table 3](#).

Bmp Ligands

A number of Bmps have pro- or anti-angiogenic effects, and are up-regulated at sites of atherosclerosis and vascular injury. This suggests that Bmps could play a role in regulating normal vascular homeostasis and disease-associated vascular pathology. Despite these observations, the functional roles of these ligands in the intact vasculature are largely unknown.

Bmp2 and Bmp4

Bmp2 and Bmp4 are up-regulated in ECs at sites of systemic vascular injury and in atherosclerotic plaques. Moreover, Bmp2 is up-regulated in vessel walls in two different rat models of systemic hypertension. The effects of these changes in Bmp expression are impossible to predict based on the results of *in vitro*, cell culture studies alone. For example, while it has been shown that Bmp2 and Bmp4 inhibit serum and platelet-derived growth factor-induced proliferation of aortic VSMCs, others have shown that the same ligands enhance proliferation of aortic VSMCs. Furthermore, while Bmp2 and Bmp4 increase expression of VSMC differentiation markers in human aortic VSMCs, others have shown that Bmp2 (but not Bmp4) down-regulates expression of the same differentiation markers in rat aortic SMCs. These contrasting findings reflect the heterogeneity of cultured vascular cells and the effects of passage number on cell differentiation. For these reasons, and given the complex and integrative nature of the Bmp signaling pathway *in vivo*, we believe that functional effects of these ligands can only properly be evaluated in the intact organism.

There are no data on the effects of genetic manipulation of Bmp ligand expression on the systemic vasculature *in vivo*. However, over-expression of Bmp2 induced using an adenoviral delivery system reduces neointimal proliferation following carotid artery balloon injury in rats. This suggests that Bmp2 may have protective effects in atherosclerotic plaque formation. Conversely, continuous subcutaneous infusion of Bmp4 in mice promotes systemic hypertension, and is associated with impaired EC-dependent vasodilatation in aortic ring preparations. The latter effects are dependent on Bmp4 induction of reactive oxygen species (ROS) in ECs, and are consistent with the observation that exogenous Bmp2 induces ROS and inhibits EC-dependent vasodilatation in rat carotid artery rings. Moreover, both Bmp2 and Bmp4 increase EC adhesiveness to circulating inflammatory cells in an ROS-dependent fashion. These findings suggest that Bmps not only influence vascular cell proliferation and/or differentiation but also influence vascular reactivity and inflammatory responses. These effects may be dependent on the vascular bed, since Bmp4 has been shown to inhibit EC-dependent vasodilatation in carotid and coronary but not pulmonary artery preparations.

Pulmonary Bmp2 and Bmp4 expressions are strongly up-regulated in a model of chronic, hypoxia-induced pulmonary hypertension (PH). This model is characterized by sustained increases in pulmonary vascular resistance associated with increased peripheral pulmonary vascular remodeling. Bmp4 displays a wide expression pattern in alveolar and bronchial epithelium, while Bmp2 is largely expressed in the peripheral resistance vessels of the pulmonary vasculature following exposure to hypoxia. These vessels are the primary site of pulmonary vascular remodeling and resistance in hypoxic PH, suggesting that local expression of Bmp2 and Bmp4 may have different effects on the pulmonary vasculature in hypoxic PH. However, *in vitro* studies provide conflicting reports about the effects of Bmps on pulmonary vascular cells. For example, Bmp4 inhibits proliferation and promotes apoptosis in proximal pulmonary artery VSMCs, but increases proliferation and inhibits apoptosis in VSMCs derived from more peripheral vessels. Moreover, Bmp4 promotes proliferation in early-passage mouse proximal

Table 3 Vascular phenotypes of Bmp mutations in humans (all capital letters) and mice

Genetic modification	Het/homozygous	Vascular phenotype	Survival
<i>Bmp2</i> null	Heterozygous	Susceptible to hypoxic PH	Viable
<i>Bmp4</i> null-LacZ	Heterozygous	Less severe hypoxic PH, impaired vascular remodeling	Viable
<i>Bmpr2</i> null	Heterozygous	Susceptible to PH, defective vascular remodeling, abnormal vascular tone	Viable
<i>Bmpr2</i> hypomorph (Δ Ex2)	Heterozygous	Susceptible to hypoxic PH; endothelial dysfunction and abnormal vascular tone	Viable
<i>Bmpr2</i> knockdown	90% knockdown by RNAi	Hemorrhage; vascular dysplasia; impaired VSMC recruitment	Viable
<i>Bmpr2</i> conditional	Pulmonary EC	Spontaneous PH, vascular inflammation	Viable
<i>BMPR2</i> (mis-sense, nonsense)	Heterozygous	HPAH, HPAH with HHT	Viable
<i>Alk1</i> null	Homozygous	Angiogenesis defects, impaired VSMC recruitment	Lethal (E10.5)
	Heterozygous	Models HHT	Viable
<i>Alk1</i> conditional	EC-specific	Vascular dysplasia	Viable
<i>ALK1</i> (mis-sense, nonsense)	Heterozygous	HHT2, HPAH with and without HHT2	Viable
<i>Alk3</i> conditional	Mesoderm deletion	Hemorrhage, cardiac defects, impaired VSMC recruitment	Lethal (E10.5)
	SMC (embryo)	Cardiac defects, hemorrhage, impaired vascular remodeling	Lethal (E11.5)
	SMC (adult)	Impaired vascular remodeling	Viable
<i>Endoglin</i> null	Homozygous	Angiogenesis defects	Lethal (E10.5)
	Heterozygous	Models HHT, spontaneous PH, abnormal vascular tone	Viable
<i>ENDOGLIN</i> (mis-sense, nonsense)	Heterozygous	HHT1, HPAH with HHT1	Viable
<i>Smad1</i> null	Homozygous	Defects in fusion of the chorioallantoic membrane and (minor) yolk sac angiogenesis	Lethal (E9.5)
<i>SMAD1</i> (mis-sense)	Heterozygous	HHT	Viable
<i>Smad4</i> conditional	EC	Cardiac and angiogenesis defects, VSMC recruitment	Lethal (E10.5)
<i>SMAD4</i> (mis-sense, nonsense)	Heterozygous	HHT (with and without JP)	Viable
<i>Smad5</i> null	Homozygous	Cardiac and angiogenesis defects	Lethal (E9.5)
<i>Smad6</i> null	Homozygous	Cardiac defects, vascular calcification, systemic hypertension	Some viable
<i>Smad8</i> null-LacZ	Homozygous	Spontaneous pulmonary vascular remodeling	Viable
<i>SMAD8</i> (nonsense)	Heterozygous	HPAH	Viable
<i>Tak1</i> null	Homozygous	Angiogenesis/cardiac defects	Lethal (E10.5)

All mutations are global unless noted otherwise.

Embryonic lethality is indicated in parenthesis as day post-coitum.

PH, pulmonary hypertension; VSMC, vascular smooth muscle cell; EC, endothelial cell; HPAH, heritable pulmonary arterial hypertension; HHT, hereditary hemorrhagic telangiectasia; VSMC, vascular smooth muscle cell; JP, juvenile polyposis.

pulmonary VSMCs but has the opposite effect on late-passage VSMCs. Given the complexity and heterogeneity of these *in vitro* responses therefore, we have taken a genetic approach to evaluate the functional role of these ligands in the intact vasculature.

Heterozygous null *Bmp2* and *Bmp4* mice show a blunted *Bmp2* and *Bmp4* response to hypoxia. However, while *Bmp2*-deficient mice have increased susceptibility to hypoxic PH, *Bmp4* deficiency partially protects against the development of hypoxic PH. In addition, while *Bmp4*-deficient mice have reduced pulmonary VSMC proliferation and peripheral vessel remodeling, there is no change in VSMC proliferation or peripheral vessel remodeling in hypoxic *Bmp2*-deficient mice. This suggests that hypoxia-induced *Bmp4* expression has distinct effects on pulmonary VSMC proliferation and remodeling while *Bmp2* does not. Moreover, hypoxic *Bmp2* but not *Bmp4*-deficient mice have reduced pulmonary endothelial nitric oxide synthase (eNOS) expression and activity (phospho-S239 VASP expression), suggesting that hypoxia-induced *Bmp2* may reduce pulmonary vascular resistance by increasing EC production of nitric oxide in small intrapulmonary resistance vessels. However, since both *Bmp2* and *Bmp4* induce eNOS expression in ECs, these findings suggest that the opposing effects of *Bmp2* and

Bmp4 in hypoxic PH result from their dominant effects on different cell types (*Bmp4* on VSMCs and *Bmp2* on ECs). Given their distinct spatial expression patterns, opposing effects of *Bmp2* and *Bmp4* are likely to result from differences in access of these ligands to distinct vascular cells types in the hypoxic lung.

Circulating Bmp ligands

In addition to local tissue production of Bmp ligands, high (ng ml⁻¹) concentrations of biologically active *Bmp4*, *Bmp6*, and *Bmp9* have been found in human serum. These biologically active ligands would have direct access to the systemic and pulmonary vascular beds, suggesting that they could play a role in maintaining normal vascular homeostasis. Unlike *Bmp4* and *Bmp6*, which have pro-angiogenic effects on cultured ECs, exogenous *Bmp9* is a potent inhibitor of angiogenic responses at concentrations found in the circulation. This suggests that *Bmp9* may play a unique role as a circulating vascular quiescence factor. Interestingly, unlike *Bmp4* (and probably *Bmp6*), *Bmp9* is not inhibited by Noggin, suggesting that selective antagonism by Noggin could determine the net pro-versus anti-angiogenic effects of circulating Bmps in vascular

beds. Despite this, the functional role of circulating Bmp9 (or other Bmp ligands) *in vivo* is unknown. Moreover, recent studies have shown that when used in combination, Bmp9 and TGF- β may act in a cooperative manner to enhance angiogenic responses. These studies emphasize the importance of establishing the functional role of these Bmp-dependent responses in the context of the intact vasculature *in vivo*.

Extracellular Regulators of Bmp Signaling

A number of studies have demonstrated changes in expression of extracellular modulators of Bmp signaling associated with vascular injury and disease. These Bmp regulators have been shown to modify vascular cell function, but there are limited data on their role in regulating vascular structure and/or function *in vivo*. Furthermore, while many of these factors have been shown to modify Bmp signaling, they often have other functional properties, so that changes in their expression are not necessarily associated with alterations in Bmp signaling.

Gremlin

The Bmp antagonist Gremlin promotes aortic VSMC proliferation and migration (141), and is up-regulated in the systemic vasculature of patients with uremia (which is associated with extensive vascular calcification), in a carotid artery injury model and in alveolar hypoxia (associated with the development of PH). However, Gremlin is dynamically regulated and its effects should be interpreted in the context of other components of the Bmp signaling pathway. For example, Gremlin expression in the vascular media is up-regulated in a rat model of vascular calcification, but this is also associated with increased expression of Bmp2. Furthermore, in the carotid artery injury model an early decrease is followed by a prolonged increase in Gremlin expression. These kinetic changes in Gremlin expression correlate inversely with the temporal expression of Bmp4 in another model of carotid artery injury. Similarly, other studies have shown that Bmp4 is co-expressed with Bmp antagonists Noggin, Follistatin, and Matrix Gla protein (MGP) at sites of oscillatory shear stress in the systemic vasculature. These studies support the hypothesis that the primary function of these Bmp antagonists is to provide a temporal limit on Bmp signaling in the context of vascular injury. This underscores the importance of developing integrated approaches to evaluate gene/protein regulation and function in the Bmp signaling pathway in vascular disease.

Bmper/Cv2

Bmper/Cv2 also regulates vascular development. Knockdown of Bmper/Cv2 in embryonic zebrafish results in a defect in vascular patterning. In mammalian cells, Bmper/Cv2 is expressed in ECs from the heart, lung, and skin, and modulates Bmp4-dependent angiogenic responses. Low-level Bmper/Cv2 expression promotes EC sprouting and migration *in vitro*, while higher-level Bmper/Cv2 represses these responses. However, functional evidence that Bmper/Cv2 regulates mammalian vascular homeostasis *in vivo* is lacking.

Matrix Gla-protein

MGP is an extracellular matrix component produced by VSMCs that inhibits or activates Bmp2 and Bmp4 in a

concentration-dependent fashion. MGP-deficient mice have widespread vascular calcification and stenosis of peripheral pulmonary vessels is observed in some patients with Keutel syndrome, which is genetically linked to germ line MGP mutations. There are no data on the effects of these mutations on Bmp signaling. However, over-expression of MGP in the lung (MGP bacterial artificial chromosome transgenic mice) is associated with reduced Bmp-dependent Smad signaling along with reduced pulmonary vascular branching. This suggests that vascular effects of MGP may be mediated through inhibition of Bmp signaling.

Bmp Receptors and Co-Receptors

Definitive evidence that the Bmp receptors play a role in maintaining normal vascular homeostasis and promote vascular disease in humans has come from genetic analyses of patients with rare systemic and pulmonary vascular diseases. Genetic studies of the same receptor mutations in mice phenocopy the corresponding vascular diseases and provide insight into the mechanism by which these receptors regulate embryonic vascular development and homeostasis in the adult. Moreover, there is increasing evidence from these studies that defects in Bmp receptor signaling not only regulate vascular structure but also have potent effects on vascular tone. This section will summarize data on the dominant receptors and co-receptors that have been identified and studied genetically in mice and humans (summarized in Table 3).

Bmpr2

More than 70% of patients with rare, autosomal dominant inherited hereditary pulmonary arterial hypertension (HPAH) inherit mutations at the *BMPR2* locus. In addition, ~20% of patients with apparently idiopathic (nonhereditary) PAH also carry *BMPR2* mutations. These mutations occur throughout the *BMPR2* open reading frame indicating that they are likely to cause loss of function and/or dominant negative effects. Moreover, given that the pathology of HPAH is identical to other forms of PAH, these findings suggest that defects in Bmpr2 signaling may also participate in pathogenesis of other forms of PH not associated with heritable *BMPR2* mutations. This is supported by the observation that components of the Bmp signaling pathway are down-regulated in the lungs of patients with nonhereditary forms of PAH. Despite this, there is no clear consensus as to how Bmpr2 signaling affects the pulmonary vasculature. Moreover, while there is no evidence that HPAH patients are predisposed to develop systemic hypertension or vascular disease, it is completely unknown why *BMPR2* mutations have dominant effects on the pulmonary versus systemic vasculature.

Studies in mice with genetic mutations at the *Bmpr2* locus provide insight into the mechanisms by which *BMPR2* mutations give rise to pulmonary vascular disease in HPAH. Using ribonucleic acid interference (RNAi) in the germline to reduce Bmpr2 expression to ~10% of normal leads to massive intestinal hemorrhage, vascular dysplasia, and impaired VSMC recruitment to pulmonary vessels. This indicates that Bmpr2 plays a role in maintaining normal systemic and pulmonary vascular stability and structure, and is similar to the vascular phenotype seen in mice carrying *Endoglin* and *Alk1* mutations

(see below). Moreover, ~30% of mice with conditional deletion of *Bmpr2* in ECs develop spontaneous PH. This is associated with focal infiltration of inflammatory cells around distal pulmonary arteries, indicating that loss of *Bmpr2* expression in ECs gives rise to local vascular inflammation. Similar effects have been described in another mouse model in which a mutant form of *Bmpr2* is over-expressed in VSMCs. Since these mice also develop spontaneous PH, it is possible that *Bmpr2*-dependent inflammation within the pulmonary vasculature plays a role in the development of PH. This is consistent with the observation that pulmonary vascular lesions in the patients with PAH often contain infiltrating mononuclear cells and lymphocytes. However, the effect of mononuclear cell depletion on the development of PH has yet to be determined. Furthermore, these findings are hard to reconcile with the fact that *Bmp2* and *Bmp4* have pro-inflammatory effects on the systemic vasculature. Since *Bmpr2* is the dominant type 2 receptor required for *Bmp2* and *Bmp4* signaling, this suggests that *Bmpr2* signaling may have opposite, anti- versus pro-inflammatory effects on the pulmonary versus systemic vasculature, respectively. An alternative explanation is that any disturbance in the balance of Bmp signaling (either increased or decreased) enhances pro-inflammatory responses in systemic and pulmonary vasculature. Careful analysis of the systemic vasculature in mice with disruption of *Bmpr2* signaling in ECs would address this question.

Studies in mice carrying heterozygous germ line mutations at the *Bmpr2* locus also provide insight into the pathobiology of HPAH. Heterozygous germ line null and hypomorphic *Bmpr2* mutant mice (*Bmpr2 Δ Ex2*, in which there is an in-frame deletion of Exon 2) do not develop spontaneous PH, but have increased susceptibility to PH in response to inflammatory mediators and serotonin, or chronic hypoxia, respectively. However, none of these models develops alliterative vascular remodeling that is more typical of human HPAH. In fact, if anything, heterozygous null *Bmpr2* mutant mice have reduced pulmonary vascular remodeling in response to hypoxia, suggesting that increased susceptibility to PH does not result from structural alterations in the pulmonary vasculature. It has been argued that this failure to recapitulate the severity of human PAH is a significant limitation of mouse models of PH. However, another possibility is that changes observed in the pulmonary vasculature of mice carrying different *Bmpr2* mutations are indicative of early, pre-clinical events that occur in the pulmonary vasculature of patients with HPAH. For this reason, the observation that heterozygous null and hypomorphic *Bmpr2* mutant mouse pulmonary vasculature show increased pulmonary vasoconstriction in response to different agonists is of particular interest. Moreover, hypomorphic *Bmpr2* mutant mice have a marked defect in EC-dependent pulmonary vasodilatation associated with reduced pulmonary vascular eNOS expression, suggesting that *Bmpr2* mutations may have a dominant effect on EC function. This is consistent with the observation that *Bmpr2* is dominantly expressed in EC versus VSMC compartments of the pulmonary vasculature. Moreover, *Bmp2* (ligand) heterozygous mutant mice which show increased susceptibility to hypoxic PH also have reduced pulmonary vascular eNOS expression and activity. In addition, mice carrying heterozygous germ line mutations at the *Endoglin* locus have

abnormal pulmonary and systemic vascular reactivity and decreased eNOS activity (see below). Taken together, these findings suggest that aberrant Bmp signaling associated with *BMPR2* mutations in patients with HPAH not only regulates vascular cell proliferation, remodeling, and inflammation, but also promotes abnormal pulmonary vascular reactivity in the patients.

Alk1

Inactivating mutations at the *ALK1* locus have been identified in patients with hereditary hemorrhagic telangiectasia type 2 (HHT2). HHT2 is an autosomal dominant heritable disease characterized by vascular dysplasia that results in abnormal, dilated blood vessels and arterio-venous malformations. A functional role for *Alk1* in this disease has been confirmed by gene targeting in mice. Homozygous *Alk1* null mice die embryonically with widespread arteriovenous malformations resulting from defective pericyte recruitment to the primitive vasculature. Similar vascular defects are seen in *Bmpr2*-deficient (germ line RNAi knockdown) and *Endoglin*-null mice, suggesting that all three receptors play a role in inducing or maintaining embryonic vascular stability. Aging adult heterozygous *Alk1* mutant mice also develop systemic vascular malformations that phenocopy the lesions seen in human HHT, and deletion of *Alk1* in ECs alone is sufficient to phenocopy HHT. The same vascular malformations are seen in heterozygous *Endoglin* mutant mice (see below). Furthermore, a subset of patients with HHT2 carrying *ALK1* mutations develop PAH, suggesting that *ALK1* and *BMPR2* mutations have common effects on the pulmonary vasculature. Taken together, these findings suggest that *Alk1* participates with *Bmpr2* and *endoglin* in regulating systemic and pulmonary vascular homeostasis in the adult.

Alk3

ALK3 has not been genetically linked to human disease but has been functionally implicated in vascular biology from genetic studies in mice. *Alk3* is ubiquitously expressed during development and homozygous deletion in mice results in embryonic lethality due to gastrulation defects. However, conditional *Alk3* deletion in embryonic vascular progenitor cells results in widespread vascular abnormalities in vessel maturation and remodeling. Moreover, patchy conditional deletion of *Alk3* in VSMCs reduced the ability of pulmonary VSMCs to remodel in response to chronic hypoxia. These findings suggest that *Alk3* expression in VSMCs regulates Bmp-dependent vascular remodeling. Stabilization of systemic and pulmonary vasculature may be mediated by *Alk3*-dependent repression of VEGF. Furthermore, since *Bmp4*-deficient mice have defective pulmonary vascular remodeling and VSMC proliferation in response to chronic hypoxia, these findings also suggest that *Bmp4* may be signaling through *Alk3* in the lung vasculature to promote VSMC proliferation and vascular remodeling.

Endoglin

Genetic studies show that patients with HHT1 inherit heterozygous mutations at the *ENDOGLIN* (*ENG*) locus. Moreover, there are case reports of families with HHT1 and PAH that have *ENG* mutations, suggesting that *endoglin* may play a common role in regulating both systemic and pulmonary vascular

homeostasis. The vascular defects in patients with HHT1 are indistinguishable from those seen in patients with HHT2 who inherited heterozygous mutations at the *Alk1* locus. Heterozygous *Eng* mutant mice phenocopy HHT. Recent studies have also shown that these mice also develop spontaneous PH. These findings provide further evidence that *Alk1*, endoglin, and *Bmpr2* play a cooperative role in regulating both systemic and pulmonary vascular homeostasis. There is evidence that *Eng* heterozygous mutant mice also have abnormalities in systemic and pulmonary vascular tone, and that these effects are mediated by ROS-dependent uncoupling of eNOS in ECs. Moreover, the finding that heterozygous null *Bmpr2*, and *Bmpr2* hypomorphic mutant mice also have abnormalities in pulmonary vascular tone associated with decreased activity and/or expression of eNOS, suggests that endoglin-dependent regulation of eNOS may be mediated through a *Bmpr2*-dependent pathway. These findings provide further support for the hypothesis that Bmp signaling not only regulates structural changes in the systemic and pulmonary vasculature, but also plays an important role in maintaining normal vascular tone.

Defects in endoglin signaling are also reported in patients with pregnancy-associated hypertension resulting from pre-eclampsia (a condition resulting from destabilization and leakage of maternal vascular ECs). Soluble endoglin is detected in the sera of women with pre-eclampsia. Moreover, increased levels of placental derived soluble endoglin and soluble vascular endothelial growth factor receptor 1 (Vegfr1) correlate with severity of pre-eclampsia, and adenoviral over-expression studies indicate that soluble endoglin and soluble Vegfr1 act in a cooperative manner to induce severe pre-eclampsia in pregnant rats. The mechanism by which soluble endoglin mediates these effects is unknown, but could result from scavenging Bmp9, which may act as a circulating endothelial quiescence factor. In addition to the association with pre-eclampsia, endoglin may also play a role in the development of aging associated hypertension. Two splice variants of endoglin have been described, long (L-) and short (S-) endoglin, which differ in their intracellular tails. The ratio of S/L endoglin increases with aging in mice, while transgenic over-expression of S-endoglin in ECs leads to systemic hypertension associated with decreased eNOS activity. Since L-endoglin promotes *Alk1* and S-endoglin promotes *Alk5* signaling in ECs, alterations in the ratio of the two isoforms may modify vascular function by influencing the balance of *Alk1/5* signaling in ECs.

Downstream Signaling

Bmp signaling activates both Smad-dependent and Smad-independent responses in a variety of cultured vascular cell types. While the functional roles of some of these responses have been defined by genetic studies in mice, many of these effects are poorly defined and their roles in mediating Bmp-dependent effects in vascular development and disease are largely unknown.

Bmp-activated Smads

Biochemical studies tend to lump together Bmp-activated Smad1, Smad5, and Smad8 as a single response. However, genetic studies indicate that different Smads have nonredundant functions at different stages of embryonic development.

Germ line *Smad5*-null mice die early in gestation with defects in vascular organization and pericyte recruitment to the primitive extra-embryonic (yolk sac) and embryonic vasculature. These defects are reminiscent of the angiogenesis defects seen in *Bmpr2* knockdown and *Eng* and *Alk1*-null embryos, suggesting Smad5 may be functioning in the same pathway. However, mice in which Smad5 is conditionally deleted in ECs slightly later in development are viable and do not have defects in the embryonic or adult vasculature, indicating that there is functional compensation for loss of Smad5 expression later in development. In contrast to the *Smad5*-null mouse, germ line *Smad1*-null mutants have prominent defects in chorioallantoic membrane fusion, but only minor defects (possible secondary) in the yolk sack vasculature. Moreover, *Smad8* mutant mice are viable, indicating that Smad1 and Smad8 play redundant roles in early and late embryonic vascular development, respectively. Interestingly, aging *Smad8*-null mice develop spontaneous remodeling of the intrapulmonary vasculature reminiscent of the obliterative pulmonary vasculopathy seen in patients with HPAH. Furthermore, Smad8 is highly expressed in the pulmonary vasculature, and an inactivating, nonsense mutation in *SMAD8* has recently been reported in one idiopathic PAH patient. Together, these findings indicate that Smad8 plays a selective role in regulating adult pulmonary vascular homeostasis, and suggest that defective Smad8 signaling could mediate *Bmpr2*-dependent effects in HPAH.

Smad4

In humans, *SMAD4* mutations have been linked to patients with combined HHT and juvenile polyposis (JV/HHT). Moreover, while *Smad4*-deficient mice display early embryonic lethality due to defective gastrulation, conditional deletion of *Smad4* in ECs leads to defects in the extra-embryonic and embryonic vasculature that resemble those seen in *Alk1* and *Eng* knockout mice. Taken together, these findings suggest that Bmp-dependent Smad5/4 signaling mediates early angiogenesis in the developing embryo. Genetic linkage to HHT suggests that Smad4 also regulates vascular homeostasis in adults. However, Smad4 is a common mediator of Bmp and TGF- β activated Smad signaling, so the relative contribution of Smad4 in mediating Bmp-dependent effects in the vasculature is unknown.

Smad6

Smad6 is widely expressed in the cardiovascular system in the embryo, and in the adult Smad6 expression is largely restricted to the vascular ECs. While a significant proportion of *Smad6* mutant mice die due to cardiac septation defects, those that survive develop systemic hypertension, enhanced vascular contractility, and impaired NO-dependent vasodilatation of aortic rings. These findings are reminiscent of the effects of exogenous Bmp2 and Bmp4 on the systemic vasculature, but are difficult to reconcile with the known functional properties of Smad6 as a selective inhibitor of Bmp-dependent Smad signaling. This is consistent with the observation that these Bmp2- and Bmp4-dependent effects are mediated through Smad-independent ROS-mediated mechanisms, and that the effects of Smad6 on the systemic vasculature may be independent of its effects on Bmp-activated Smad signaling.

Smad-independent signaling

Bmp ligands activate a range of different Smad-independent transcriptional and nontranscriptional pathways in cultured vascular cells. However, the functional role of these responses in mediating Bmp-dependent effects in the intact vasculature is largely unknown. An exception to this is the downstream mediator of Bmp-dependent MAPK signaling, Tak1. *Tak1*-deficient mice have abnormal extra-embryonic and embryonic vasculature resulting from defective VSMC recruitment to the remodeling vessels. This phenotype is identical to the *Alk1* and *Eng*-null mice, suggesting that Tak1 is mediating signaling downstream of the Alk1/Eng signaling complex. Moreover, over-expression of Tak1 rescues the vascular phenotype resulting from *Alk1* deficiency in zebrafish embryos. However, it is unclear whether these effects result from Tak1-dependent activation of p38 MAPK and JNK, or result from the effects of Tak1-dependent activation of Bmp activated Smads. This illustrates the complexity of understanding the true nature of Bmp signaling in the intact organism, and the challenges ahead.

Concluding Remarks

The functional roles of Bmp signaling defects in human vascular diseases remain unclear. However, genetic studies in mice indicate that different components of the Bmp signaling pathway stabilize EC/VSMC interactions in the developing vasculature, suggesting that this pathway may also be required to maintain vascular stability in the adult. Functional studies in adult mice also indicate that Bmp signaling plays a role in regulating vascular reactivity, EC function, and inflammation. New discoveries are revealing novel roles for Bmp signaling,

such as regulating microRNAs. Additionally, given that expression of Bmp ligands is often controlled by long-range regulatory elements, dysfunctional signaling may be involved in human disease not attributable to mutation at the gene locus. Moreover, with the identification of small-molecule inhibitors (and potentially activators) of Bmp signaling using high content screening, new possibilities for modulation *in vivo* will assist in future discovery and therapeutics.

See also: **Bioenergetics: Iron–Sulfur Proteins.**

Further Reading

- Bernabeu C, Lopez-Novoa JM, and Quintanilla M (2009) The emerging role of TGF-beta superfamily coreceptors in cancer. *Biochimica et Biophysica Acta* 1792: 954–973.
- Corradini E, Babbitt JL, and Lin HY (2009) The RGM/DRAGON family of BMP co-receptors. *Cytokine and Growth Factor Reviews* 20: 389–398.
- Lonn P, Moren A, Raja E, Dahl M, and Moustakas A (2009) Regulating the stability of TGFbeta receptors and Smads. *Cell Research* 19: 21–35.
- Moustakas A and Heldin CH (2009) The regulation of TGFbeta signal transduction. *Development* 136: 3699–3714.
- Ross S and Hill CS (2008) How the Smads regulate transcription. *International Journal of Biochemistry and Cell Biology* 40: 383–408.
- Shi Y and Massague J (2003) Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* 113: 685–700.
- Umulis D, O'Connor MB, and Blair SS (2009) The extracellular regulation of bone morphogenetic protein signaling. *Development* 136: 3715–3728.
- Wrighton KH, Lin X, and Feng XH (2009) Phospho-control of TGF-beta superfamily signaling. *Cell Research* 19: 8–20.
- Zhang YE (2009) Non-Smad pathways in TGF-beta signaling. *Cell Research* 19: 128–139.
- Zhao GQ (2003) Consequences of knocking out BMP signaling in the mouse. *Genesis* 35: 43–56.

Bradykinin Receptors

R M Burch, Naurex, Inc., Morris, CT, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Angiotensin Peptide produced by the action of the enzyme renin on the protein angiotensinogen.

G protein Trimeric complex of α , β , and γ subunits. Activation of a G-protein-coupled receptor results in dissociation of the G α subunit, which in turn activates signaling proteins such as adenylyl cyclase, certain ion channels, and phospholipases. The activity of the G α subunit is terminated by endogenous GTPase activity, followed by reassociation of the G α with G $\beta\gamma$ subunit.

Ischemia Loss of blood flow with resultant metabolic insufficiency of a tissue bed.

Myocardial infarction Death of heart muscle following prolonged, complete ischemia.

Prostaglandin Autocoid (local hormone) produced by the action of cyclooxygenase 1 or cyclooxygenase 2 on arachidonic acid. A variety of prostaglandins may be produced, each with specific biological actions elicited by binding to cell-surface receptors.

Single nucleotide polymorphism Difference at a specific place in a DNA sequence of a single nucleotide. The two gene variants are alleles.

Overview

Most G-protein agonists are hormones or neurotransmitters synthesized intracellularly and stored in secretory granules prior to release. By contrast, bradykinin is a member of the kallikrein-kininogen-kinin system, a complex of two substrates (kininogens) activated by two enzymes (kallikreins) to produce four inflammatory mediators (kinins) that bind to two bradykinin receptors (Figure 1).

The substrates, high-molecular-weight (HMW) kininogen and low-molecular-weight (LMW) kininogen, are α_2 macroglobulins produced by alternative splicing of the same mRNA. HMW kininogen circulates in blood in a 1:1 complex with plasma prekallikrein. Upon tissue injury, factor XII, a component of the coagulation cascade, cleaves plasma prekallikrein to activate it. Plasma kallikrein acts on HMW kininogen to release bradykinin (plasma kallikrein also activates LMW kininogen to produce bradykinin). Tissue kallikrein is synthesized by many types of epithelial cells and secretory cells. Thus, unlike plasma kallikrein, it acts locally. Tissue kallikrein acts only on LMW kininogen and cleaves it one residue to the amino terminus of the site of cleavage of plasma kallikrein to produce a decapeptide, kallidin (or Lys-bradykinin). Bradykinin and kallidin are equipotent agonists of bradykinin B2 receptors and are usually referred to generically as bradykinin or kinins. It should be noted that bradykinin and kallidin are chemically heterogeneous since the proline residues within the peptides may be partially hydroxylated.

Kinins are metabolized further by a variety of enzymes, most of which inactivate them. The major inactivating enzyme is kininase II, especially prevalent in the lungs and in vascular beds, and its activity is largely responsible for the very short half-life of the kinins, about 15 s. Kininase II removes the carboxyterminal dipeptide from both bradykinin and kallidin. Kininase II is also known as an angiotensin-converting enzyme because it converts angiotensin I into the vasoconstrictor angiotensin II. Of importance to the present discussion, another enzyme, kininase I, removes the carboxyterminal Arg from bradykinin and kallidin to produce

the other two inflammatory mediators, the bradykinin B1 receptor agonists desArg⁹-bradykinin and desArg¹⁰-kallidin, respectively. Like the agonists of B2 receptors, bradykinin and kallidin, the desArg metabolites have similar potency as agonists of B1 receptors.

Nearly all tissues express B2 receptors, and bradykinin-induced activation of B2 receptors is implicated in many disease states. B1 receptors are expressed in only a few tissues under normal conditions and only in very small numbers. Several disease states are associated with rapid induction of B1 receptors in specific tissues.

B1 Receptor

The existence of a B1 receptor activated by what had been considered an inactive bradykinin metabolite, des-Arg⁹-bradykinin, was a provocative proposal. This receptor was not found in normal tissues, but was found to be induced during incubation of tissues *in vitro*. Identification of a synthetic B1 receptor antagonist, desArg⁹-Leu⁸-bradykinin, added credence to the existence of a physiologic receptor. Later, induction of B1 receptors was found pharmacologically to be associated with specific disease processes *in vivo*, and in 1994, the existence of the B1 receptor was conclusively confirmed with the cloning of the receptor protein.

B1 receptors have been implicated in hyperalgesia, plasma extravasation, white blood cell activation, and accumulation, and in the control of blood pressure. However, it seems certain that complete understanding of the roles of this inducible receptor in pathophysiology is not at hand. B1 receptors from several species have been cloned. They are typical G-protein-coupled receptors in sequence. The human receptor has a predicted sequence of 353 amino acids and is only 36% homologous to B2 receptors. The B1 receptor activates phosphatidylinositol-specific phospholipase C and possibly phospholipase A₂.

The B1 receptor gene is composed of three exons. The entire coding region for the receptor is contained within the third

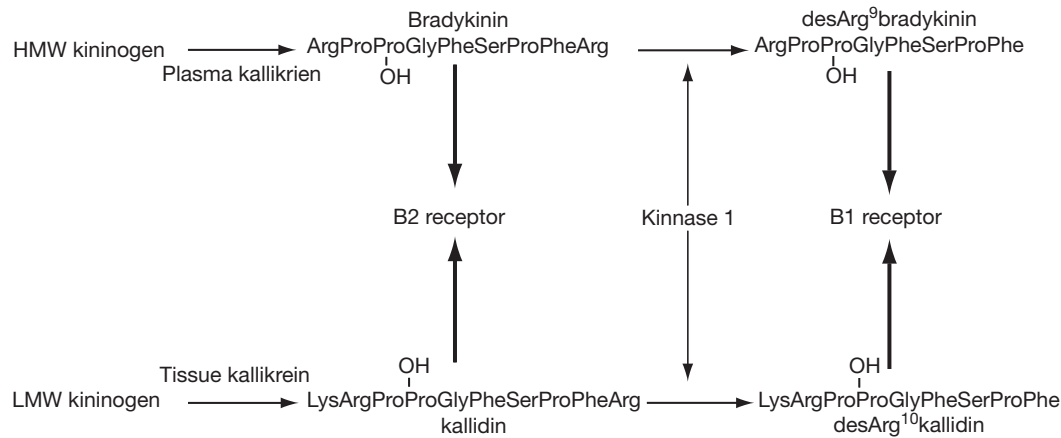


Figure 1 Formation of endogenous bradykinin receptor ligands. Bradykinin and kallidin bind to B2 receptors, whereas desArg⁹-bradykinin and desArg¹⁰-kallidin bind to B1 receptors. Both bradykinin and kallidin may contain some proline residues that are hydroxylated. If hydroxylated, the most common residue is Pro³ in bradykinin and Pro⁴ in kallidin (shown hydroxylated in the figure).

exon. A variety of polymorphisms have been identified. One, a G/C single-nucleotide polymorphism in the promoter region has been associated with disease. Expression of the C allele is higher than the G allele, and patients with the G allele have greater incidence of inflammatory bowel disease and end-stage renal disease.

B1 Receptor Regulation

Most G-protein-coupled receptors are internalized quickly after the binding of their ligands. B1 receptors, however, are not internalized following binding of desArg-kinins. This results in very prolonged activation of signal transduction, which translates into rather few receptors per cell being necessary for biological activity when compared to many other G-protein-coupled receptors. Ligand-binding studies in a variety of tissues reveal only hundreds to a few thousand B1 receptors per cell, whereas most G-protein-coupled receptors, including B2 receptors, are found in tens of thousands per cell.

In contrast to prolonged phosphatidylinositol turnover, B1 receptor-mediated arachidonic acid release and prostaglandin synthesis are short-lived, which is similar to B2 receptors. The mechanism for this differential desensitization is unknown, but it may be linked to sequestered pools of arachidonate-containing phospholipids or phosphorylation of an intracellular binding site for a transduction protein.

B1 receptors have been shown by immunoprecipitation studies to couple to G_{zq} and G_{zi}. These G-proteins commonly couple receptors to phosphatidylinositol-specific phospholipase C and elevation of intracellular free calcium activity. Activation of B1 receptors leads to the elevation of intracellular free calcium activity by increasing calcium entry into the cell; this is in contrast to the B2 receptor, which acts primarily to release bound intracellular calcium.

Regulation of B1 Receptor Induction

Unlike many G-protein-coupled receptors, the primary regulation of the B1 receptor number appears to occur by

transcriptional induction. Biological studies have found that B1 receptor expression is induced by a variety of cytokines including endotoxin, interleukin 1 β , and tumor necrosis factor. The specific domains affected by these stimulating ligands have not been identified, but activation of B1 receptor transcription has been correlated to the activation of the transcription factor nuclear factor-kappa B (NF- κ B).

The regulation of the B1 receptor gene has also been shown to be regulated by stabilization of mRNA. Interleukin 1 treatment of cells has been demonstrated to double the half-life of B1 receptor mRNA. In addition, protein synthesis inhibitors increase B1 receptor gene stability, suggesting the existence of a short-lived protein that causes destabilization. This protein may interact with the 3' untranslated region, since experimentally altering the 3' untranslated region results in B1 receptor messenger RNAs (mRNAs) with varying half-lives.

B1 Receptor Knockout Animals

B1 receptor knockout mice have been reported to develop normally and to have normal blood pressure. By contrast, when inflammatory stimuli are applied, dramatic reductions in accumulation and apoptosis of neutrophils have been reported, as well as hypoalgesia; this is consistent with pharmacological evidence presented in the past.

B2 Receptors

B2 receptors mediate bronchoconstriction, local blood flow regulation, hypotension, acute inflammatory reactions, pain, and hyperalgesia. B2 receptors from several species have been cloned. Like B1 receptors, they are typical G-protein-coupled receptors in sequence. The human receptor has a predicted sequence of 364 amino acids. Both B1 and B2 receptors are found on chromosome 14 in humans, only about 12 kb apart. Like B1 receptors, B2 receptors activate phosphatidylinositol-specific phospholipase C. In addition, in most tissues, B2 receptor activation results in the production of prostaglandins and other arachidonic acid metabolites.

The B2 receptor gene is composed of three exons, of which exon 2 and exon 3 provide the coding region for the receptor. Bradykinin has been implicated in hypertension for decades. More than 80 single-nucleotide polymorphisms have been identified for the B2 receptor. The promoter region for the B2 receptor contains a single-nucleotide polymorphism, T/C. The C allele has been demonstrated to be an independent risk factor for essential hypertension in several ethnic groups. Several other alleles have been reported and were found to be differentially associated with several disease states. Expansion of these observations promises the potential to define the role of bradykinin in a variety of physiologic and pathophysiologic states.

B2 Receptor Regulation

In contrast to the B1 receptor, B2 receptor number does not seem to be significantly regulated by inducible gene expression. Instead, B2 receptors are constitutively expressed, as is the case for many G-protein-coupled receptors. Activation of the B2 receptor by binding of bradykinin results in rapid internalization of the receptor protein by endocytosis, occurring within a few minutes. Internalization results in cessation of the biological activity of the receptor. Thus, activation of a particular B2 receptor results in a transient increase in intracellular calcium activity and transient prostaglandin release. Receptors are recycled intracellularly, with stripping of bradykinin and the return of a reactivated B2 receptor to the cell surface where it may be reactivated. Most G-protein-coupled receptors are internalized following attachment to arrestin and dynamin-dependent clathrin-coated vesicles. By contrast, the B2 receptor becomes associated with caveolae.

B2 Signal Transduction

B2 receptors have been demonstrated to activate all of the signal transduction mechanisms described for G-protein-coupled receptors (Figure 2).

Any given cell type supports only a few of the transduction pathways, and these are primarily controlled by the specific G proteins expressed by the given cell type. B2 receptors have been shown by immunoprecipitation studies to couple to $G_{\alpha q}$, which mediates interaction with phosphatidylinositol-specific phospholipase C. B2 receptors have also been shown by immunoprecipitation to couple to $G_{\alpha i}$. G-protein depletion studies have suggested that $G_{\alpha i}(2)$ and $G_{\alpha i}(3)$ couple B2 receptors to arachidonic acid release, presumably through the activation of phospholipase A_2 . In contrast to B1 receptors, activation of B2 receptors results in increased intracellular calcium activity from the release of calcium from intracellular stores.

B2 Receptor Knockout Animals

Studies with B2 receptor knockout mice have shown that the animals develop normally. However, one study found that when animals are fed a high-salt diet, severe hypertension occurs. It has also been reported that the renin-angiotensin system was abnormal (the kallikrein-kinin system and the renin-angiotensin system are antagonistic hemodynamic regulatory systems that play important roles in blood pressure

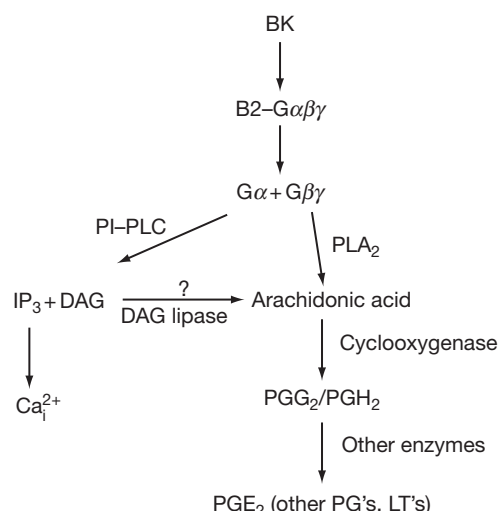


Figure 2 Signal transduction elicited by B2 receptors: BK, bradykinin or kallidin; PLA₂, phospholipase A₂; IP₃, inositol trisphosphate; DAG, diacylglycerol; PG, prostaglandin; LT, leukotriene. Not shown is phospholipase D, which, in some cells, provides a shuttle for arachidonic acid to appropriate substrates for hormone-sensitive phospholipase A₂. In most cells, bradykinin does not directly alter cAMP concentration; often, this occurs as a consequence of PGE₂ binding to its receptor.

homeostasis) and abnormal renal development occurred. A variety of cardiac defects have also been reported as well as chronically elevated heart rate.

Other Bradykinin Receptors

Many pharmacologic studies using bradykinin analogs have yielded data suggestive of additional bradykinin receptors, or at least heterogeneity of bradykinin receptors. For example, in smooth muscle preparations, there is evidence for expression of different B2-like receptors expressed by the smooth muscle and the neuronal endings in the smooth muscle. Similarly, B3 receptors have been reported in trachea, and B4 and B5 receptors have been reported in opossum esophagus. A criticism of all of these studies is the use of rather weak bradykinin analogs, with dissociation constants of several nanomolars, compared to pM dissociation constant for bradykinin itself. No genetic evidence for additional receptors beyond B1 and B2 has been obtained to date.

Bradykinin Receptor Antagonists in Clinical Medicine

While highly potent peptide antagonists for B1 receptors have been described, none to date have been evaluated in clinical trials. By contrast, several peptide antagonists for B2 receptors have been evaluated as anti-inflammatory agents, including NPC 567 and Hoe 140. In addition, several nonpeptide antagonists of B2 receptors have been described. Hoe 140 (icatibant) was approved in Europe 2009 for the treatment of hereditary angioedema. Icatibant has also been found to reduce airways hyperresponsiveness and eosinophilia following antigen challenge in patients allergic to grass pollen antigen. CP-1027 (Bradycor) has been found to reduce intracranial pressure following traumatic brain injury.

An issue that has slowed clinical development of such antagonists is the observation that bradykinin appears to play an important role in coronary artery function. It has been suggested that chronic antagonism of B2 receptors in patients with coronary artery disease may have deleterious effects during ischemic events, perhaps even triggering myocardial infarction.

See also: [Lipids](#) [Carbohydrates](#) [Membranes and Membrane Proteins](#); [Prostaglandins and Leukotrienes](#); [Signaling](#): [Angiotensin Receptors](#); [G-Protein-Coupled Receptor Kinases and Arrestins](#); [Inositol Phosphate Kinases and Phosphatases](#).

Futher Reading

Austin CE, Faussner A, Robinson HE, et al. (1997) Stable expression of the human kinin B1 receptor in Chinese hamster ovary cells. *Journal of Biological Chemistry* 272: 11420–11425.

- Bork K, Frank J, Grundt B, et al. (2007) Treatment of acute edema attacks in hereditary angioedema with a bradykinin receptor-2 antagonist (Icatibant). *Allergy and Clinical Immunology* 119: 1497–1503.
- Burch RM (ed.) (1991) *Bradykinin Antagonists. Basic and Clinical Research*. New York, NY: Dekker.
- Burch RM and Kyle DJ (1992) Recent developments in the understanding of bradykinin receptors. *Life Sciences* 50: 829–838.
- Gras J (2009) Icatibant for hereditary angioedema. *Drugs Today (Barc)* 45: 855–864.
- Marceau F and Bachvarov DR (1998) Kinin receptors. *Clinical Reviews in Allergy and Immunology* 16: 385–401.
- Marmarou A, Nichols J, Burgess J, et al. (1999) Effects of the bradykinin antagonist Bradycor (deltibant, CP-1027) in severe traumatic brain injury: Results of a multi-center, randomized, placebo-controlled trial. American Brain Injury Consortium Study Group. *Journal of Neurotrauma* 16: 431–444.
- Prado GN, Taylor L, Zhou X, et al. (2002) Mechanisms regulating the expression, self-maintenance, and signaling-function of the bradykinin B2 and B1 receptors. *Journal of Cell Physiology* 193: 275–286.
- Regoli D and Barabe J (1980) Pharmacology of bradykinin and related kinins. *Pharmacological Reviews* 32: 1–46.

Branched-Chain Amino Acids

D T Chuang, University of Texas Southwestern Medical Center, Dallas, TX, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Activity state The ratio of the actual (partially de-phosphorylated) activity to the maximal (fully de-phosphorylated) activity of the mammalian mitochondrial branched-chain α -ketoacid complex (BCKDC). The activity state represents the percentage of the total BCKDC that is active in a tissue under defined physiological conditions.

Ketoacidosis A clinical condition associated with the accumulation of the BCKDs derived from branched-chain amino acids (BCAAs) leucine, isoleucine, and valine present in dietary proteins. This pathological state is presented by patients with the maple syrup urine disease caused by congenital defects in the BCKDC.

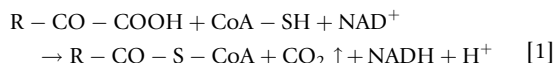
Metabolon The association of several related metabolic enzymes to form a supramolecular structure. The advantage of such a supramolecular assembly is to channel substrates between two juxtaposed steps in a metabolic pathway for maximal efficiency in an appropriate intracellular location.

Oxidative decarboxylation The removal of a CO₂ group from a molecule by a decarboxylase/dehydrogenase with a concomitant transfer of electrons from the substrate to a biological electron acceptor. For the branched-chain α -ketoacids (BCKDs), this reaction is catalyzed by the BCKDC coupled with the reduction of nicotinamide adenine dinucleotide (NAD⁺).

Phosphorylation The posttranslational modification of a protein or enzyme molecule by a specific kinase, in which the γ -phosphoryl group from ATP is incorporated into a serine, threonine, or tyrosine residue of the protein/enzyme as a means to regulate its biological activity.

Transamination A reversible enzymatic reaction that converts an amino acid into a corresponding α -ketoacid. For BCAAs, this reaction is carried out by the distinct isozymes of branched-chain aminotransferase both in the cytoplasm and in the mitochondrion.

The branched-chain α -ketoacids (BCKAs) comprise α -ketoisocaproate (KIC), α -keto- β -methylisovalerate (KMV), and α -ketoisovalerate (KIV) derived from branched-chain amino acids (BCAAs) leucine, isoleucine, and valine, respectively, through reversible transamination. The oxidative decarboxylation of these three BCKAs is catalyzed by a single branched-chain α -ketoacid dehydrogenase complex (BCKDC) giving rise to the corresponding branched-chain acyl-coenzymes (acyl-CoAs) (reaction [1]):



In patients with heritable maple syrup urine disease (MSUD), activity of the BCKDC is deficient. This metabolic block results in the accumulation of BCAAs and BCKAs. Clinical manifestations include neonatal or late onset of often-fatal ketoacidosis, encephalopathy, and other acute and chronic neurological dysfunction, as well as mental retardation in survivors. A distinct phenotype is the strong maple syrup odor in the urine of patients, and hence the name of this metabolic disease. There are currently five clinical phenotypes associated with MSUD: classic, intermediate, intermittent, thiamin-responsive, and E3-deficient. The prevalence of MSUD is one in 185 000 live births worldwide.

Degradative Pathways of BCAAs

The oxidation of BCAAs leucine, isoleucine, and valine begins with the transport of these amino acids into cells through the Na⁺-independent L transporter on the plasma membrane. In the cell, the BCAAs undergo the first transamination step

catalyzed by BCAA aminotransferases (BCATs), which are in either the cytosolic or mitochondrial compartment, to produce BCKAs (Figure 1). The second step, that is, the oxidative decarboxylation of BCKAs catalyzed by the mammalian BCKDC, occurs exclusively in the mitochondrion. The reaction products of KIC, KMV, and KIV are isovaleryl-CoA, α -methylbutyryl-CoA, and isobutyryl-CoA, respectively. These branched-chain acyl-CoAs subsequently undergo the third step, that is, dehydrogenation by specific acyl-CoA dehydrogenases. The dehydrogenation of isovaleryl-CoA is catalyzed by isovaleryl-CoA dehydrogenase, and the dehydrogenation of α -methylbutyryl-CoA and isobutyryl-CoA is carried out by α -methyl branched-chain acyl-CoA dehydrogenase. After these three steps, the degradative pathways for each of BCKAs diverge. Leucine yields acetyl-CoA and acetoacetate as end products, and is therefore a ketogenic amino acid. Valine produces succinyl-CoA, and is accordingly glucogenic. Succinyl-CoA enters the Krebs cycle, and is eventually converted to glucose through gluconeogenesis. Isoleucine is both ketogenic and glucogenic since it is metabolized to acetyl-CoA and succinyl-CoA.

Interorgan Relationships

Oxidation of BCAAs involves extensive interplay of metabolites between muscle and liver in the rat. This is due to the nonuniform distribution of BCATs and the BCKDC between organs and tissues. In rat skeletal muscle, BCAT activity is high, but BCKDC activity is low. A reverse situation exists in rat liver with respect to levels of the two enzyme activities. Both rat liver and kidney degrade BCKAs at high rates, but in hepatectomized rats

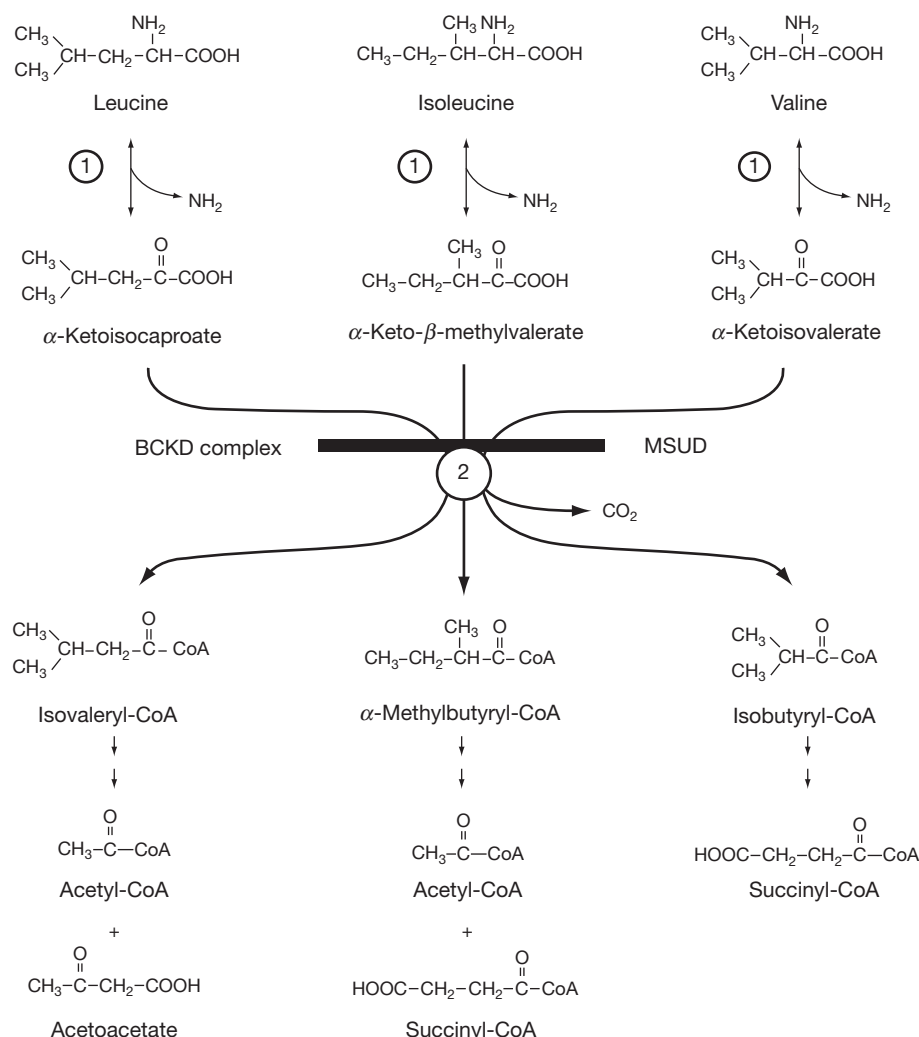


Figure 1 Catabolic pathways for the branched-chain amino acids (BCAAs) leucine, isoleucine, and valine. The first three common reactions are catalyzed by the following enzymes: reversible transamination by BCAA aminotransferases (BCATs) (1), oxidative decarboxylation of BCKAs and esterification of CoASH by the single mitochondrial branched-chain α -ketoacid dehydrogenase (BCKD) complex (2), and dehydrogenation by isovaleryl-CoA dehydrogenase or α -methyl branched-chain acyl-CoA dehydrogenase. After these steps, the degradation pathway for each amino acid diverges. As end products, leucine yields acetyl-CoA and acetoacetic acid; isoleucine produces acetyl-CoA and succinyl-CoA; and valine is converted exclusively to succinyl-CoA. The solid horizontal bar (center) depicts the metabolic block imposed by maple syrup urine disease (MSUD) mutations in the BCKDC.

leucine oxidation is decreased. This led to the prevailing view, based on the rat model, that the primary role of muscle is transamination of BCAAs, which provides the major source of circulating BCKAs. These are then transported to liver, kidney, and heart, where they are oxidized. However, the interorgan shuttling of BCAA metabolites may not occur in humans. Studies have shown that BCKDC activity in the human liver is similar to that in skeletal muscle, but the human liver exhibits twice as high BCAT activity as human skeletal muscle. This finding suggests that human liver is capable of oxidizing BCAAs *in situ* and does not depend on BCAT activity in skeletal muscle for the conversion of BCAAs to BCKAs. As skeletal muscle comprises 40% of the body mass, this tissue is likely the major site for BCAA oxidation in humans.

Regulation of BCAA and BCKA Oxidation

The oxidation of BCAAs and BCKAs is tightly regulated and exhibits tissue-specific patterns. The plasma concentrations of BCAAs and BCKAs are elevated in starvation and in clinical conditions such as diabetes mellitus, obesity, and MSUD. Paradoxically, BCAA oxidation is accelerated in muscles from fasted and diabetic rats. Epinephrine (10^{-5} – 10^{-6} M) and glucagon (2×10^{-8} to 5×10^{-9} M) were shown to also stimulate BCAA oxidation in the heart and hemidiaphragms of rats. Insulin decreases the oxidation of BCKAs in striated muscle of fed rats, whereas the same hormone increases oxidative decarboxylation of the ketoacids during starvation. Clofibrate administration augments BCAA oxidation in muscle, but inhibits their

degradation in liver. Carnitine, ketone bodies, hexanoate, and octanoate increase the oxidation of leucine by skeletal muscle, whereas pyruvate and decanoate exert inhibitory effects. The metabolism of KIC in isolated rat hepatocytes is inhibited by fatty acids, KIV, KMV, and pyruvate. It is now known that the regulation of BCAA and BCKA oxidation is largely mediated through the reversible phosphorylation of the BCKDC (see the section Dietary and Hormonal Regulation of the BCKDC).

Leucine as a Nutrition and Endocrine Signal

Among BCAAs, leucine is recognized as a nutrition signal that promotes protein synthesis via the activation of the mammalian target of rapamycin (mTOR) complex 1 and its downstream targets. KIC derived from leucine, on the other hand, inhibits protein degradation. Based on these properties, BCAA supplementation has been clinically used for the treatment of liver disease, sepsis, and burns. Moreover, leucine or BCAA supplements in diets reduce body fat mass and body weight and augment glucose oxidation in rodents. These findings suggest that leucine supplements may be beneficial to controlling obesity and type 2 diabetes. The significance of leucine signaling was demonstrated in a mitochondrial BCAT (BCATm) knockout model. The BCATm^{-/-} mice showed elevated plasma BCAA levels and significantly reduced fat and body mass. The energy expenditure in the animal model is markedly increased with improved glucose and insulin tolerance despite increased food intake. The increased energy expenditure and the resulting weight loss are explained by a futile cycle depicted in Figure 2. The cycle is implemented by increased protein synthesis promoted by leucine, which is coupled with the opposing accelerated proteolysis resulting from the relief of KIC-mediated inhibition. These findings strongly suggest that leucine signaling plays an important role in reducing fat and body mass through increased energy expenditure and insulin sensitivity.

In pancreatic islet β -cells, leucine is a known insulin secretagogue that acutely stimulates insulin secretion. Long-term treatment with leucine improves impaired insulin secretion from the islet cells of diabetic patients with a better glycemic control. Leucine pool in the islet serves as both an energy source and an allosteric activator of glutamate dehydrogenase (GDH), which results in increased glutaminolysis. Leucine-mediated glutaminolysis is shown to be critical for interprandial insulin release when blood glucose levels are low.

The Macromolecular Organization of the BCKDC

The mammalian BCKDC is a member of the highly conserved α -ketoacid dehydrogenase complexes with similar structure and function, comprising pyruvate dehydrogenase complex (PDC), α -ketoglutarate dehydrogenase complex (α -KGDC), and the BCKDC. The mammalian BCKD multienzyme complex contains three catalytic components: a heterotetrameric ($\alpha_2\beta_2$) BCKA decarboxylase or E1, a homo-24 meric dihydrolipoyl transacylase or E2, and a homodimeric dihydrolipoamide dehydrogenase or E3. The E1 and E2 components are specific for the BCKDC, whereas the E3 component is common

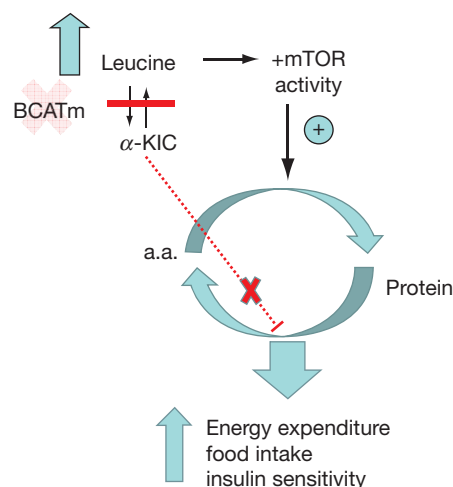


Figure 2 An energy-wasting futile cycle in mitochondrial BCAT (BCATm) knockout mice. The deletion of BCATm results in an increased leucine level, which promotes protein synthesis through the mTOR kinase-mediated signaling pathway. Reciprocally, the reduced pool size of α -ketoisocaproate (α -KIC) in BCATm-deficient mice removes the inhibition of protein degradation by α -KIC, leading to accelerated proteolysis. The increased protein synthesis accompanied by an accelerated protein degradation results in an energy-wasting futile cycle. The increased circulating leucine concentration also enhances insulin sensitivity associated with reduced body fat resulting from increased energy expenditure. Reproduced from Fried SK and Watford M (2007) Leucine weight with a futile cycle. *Cell Metabolism* 6: 155–156, with permission from Elsevier.

among the three α -ketoacid dehydrogenase complexes. In addition, the mammalian BCKDC contains two regulatory enzymes: the specific kinase and the specific phosphatase that regulate activity of the BCKDC through the phosphorylation (inactivation)/dephosphorylation (activation) cycle. The six subunits with their cofactors and prosthetic groups that make up the mammalian BCKDC are shown in Table 1. Except the phosphatase, three-dimensional structures for E1, E3, and the three E2 domains including the 24-meric cubic core of the mammalian BCKDC have been determined. The BCKDC is organized around the cubic E2 core, to which 12 copies of E1, up to six copies of E3, and unknown numbers of the kinase and the phosphatase are attached through ionic interactions (Figure 3). The molecular mass of the BCKD multienzyme complex is estimated to be 4×10^6 Da. The purified bovine E1–E2 subcomplex has a sedimentation coefficient ($S_{20,w}$) of 40 S. The E3 component has low affinity for the E2 core, and was mostly lost during purification of the mammalian BCKDC.

The reaction steps catalyzed by the three enzyme components are also shown in Figure 3. The E1 component catalyzes a thiamin diphosphate (ThDP)-mediated decarboxylation of the α -ketoacids and the subsequent reduction of the lipoyl moiety covalently bound to E2. The reduced lipoyl moiety and the lipoyl domain serve as a swinging arm to transfer the acyl group from E1 to CoA giving rise to acyl-CoA. Finally, the E3-component with tightly bound flavin adenine dinucleotide (FAD) reoxidizes the dihydrolipoyl residue of E2 with NAD^+ as the ultimate electron acceptor. The net or overall reaction is the production of a branched-chain acyl-CoA, CO_2 , and

Table 1 Component enzymes and subunit composition of the mammalian branched-chain α -ketoacid dehydrogenase complex

Component	Molecular mass (Da)	Prosthetic group (P) and cofactor (C)
BCKA decarboxylase (E1)	1.7×10^5 ($\alpha_2 \beta_2$)	ThDP (C)
α -Subunit	46 500	Mg ²⁺ (C)
β -Subunit	37 200	K ⁺
Dihydrolipoyl transacylase (E2)	1.1×10^6 (α_{24})	Lipoic acid (P)
Subunit	46 518 ^a	
Dihydrolipoyl dehydrogenase (E3)	1.1×10^5 (α_2)	FAD (C)
Subunit	55 000	
BCKD kinase	86 000 (α_2)	Mg ²⁺ (C)
Subunit	43 000	
BCKD phosphatase (PP2Cm)	38 000	Mn ²⁺ (C)
Subunit	38 000	

BCKA, branched-chain α -ketoacid; ThDP, thiamin diphosphate.

^aCalculated from the amino acid composition deduced from a bovine E2 cDNA. The E2 subunit migrates anomalously as a 52-kDa species in sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

Data from Mendelian Inheritance in Man catalog (OMIM).

NADH from a α -ketoacid, CoA, and NAD⁺ (Figure 3 and reaction [1]).

Dietary and Hormonal Regulation of the BCKDC

The regulation of BCKA oxidation was shown to be at the BCKDC step by using inhibitors such as oleate and palmitoyl carnitine. The mammalian BCKDC is negatively regulated by the reversible phosphorylation of the Ser-292 and Ser-302 residues in the E1 α subunit. The activity state of the BCKDC in different tissues is the ratio of the actual activity of a partially de-phosphorylated form to the maximal activity of the fully de-phosphorylated form. Thus, the activity state represents the percentage of the total BCKDC that is active in a given tissue or cell type. Studies showed that the activity state of the BCKDC in human tissues is 40% in heart, 26% in skeletal muscle, 28% in liver, and only 14% in kidney. Rats fed with low-protein diets show a reduction in the activity state, but not the amount, of the BCKDC in hepatocytes, compared to chow-fed rats. The decrease in activity state inversely correlates with an increase in the amounts of the BCKD kinase protein and mRNA in rat liver. Starvation, exercise, and diabetes stimulate the activity of the BCKDC in skeletal muscle by increasing the

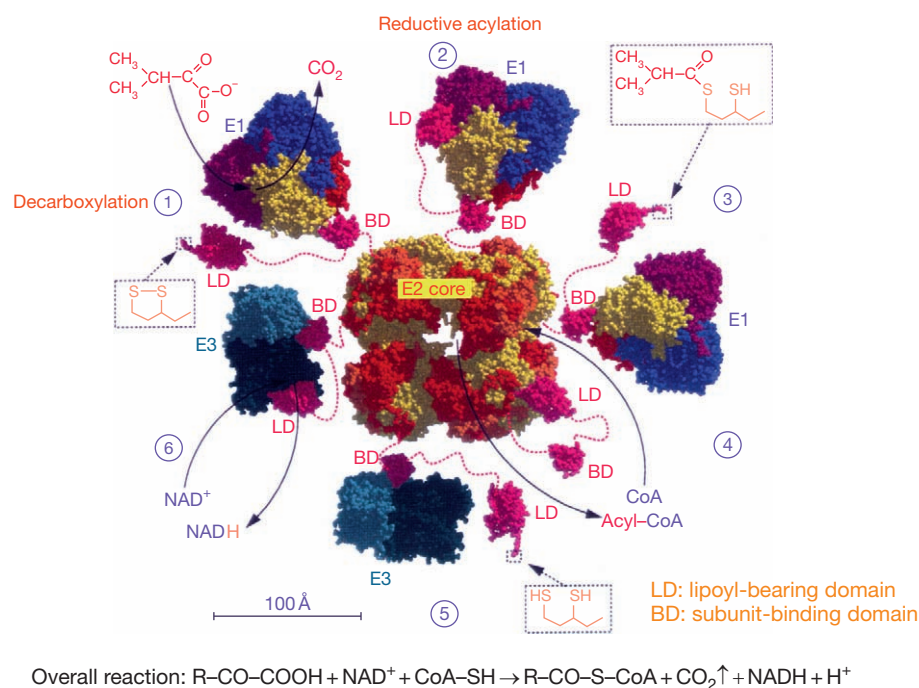


Figure 3 Model for structural organization and individual component reactions of the branched-chain α -ketoacid complex (BCKDC). The macromolecular structure (4×10^6 Da in size) is organized about a cubic transacylase (E2) core (based on the structure of *Azotobacter* pyruvate dehydrogenase E2), to which a decarboxylase (E1) (based upon *Pseudomonas* branched-chain α -ketoacid (BCKD) E1 structure) and a dehydrogenase (E3) (according to the structure of *Azotobacter* pyruvate dehydrogenase E3) are attached through ionic interactions. E2 of the BCKDC contains 24 identical subunits with each polypeptide made up of three folded domains – lipoyl (LD), E1/E3-binding (BD), and the E2 core domains – that are linked by flexible regions (represented by dotted lines). E1 $\alpha_2\beta_2$ heterotetramers or E3 homodimers are attached to BD. The BCKD kinase and BCKD phosphatase (not shown) bind to LD. E1 catalyzes the thiamin diphosphate (ThDP)-mediated oxidative decarboxylation of α -ketoacids. The ThDP-hydroxyacyl moiety is transferred to a reduced lipoyl prosthetic group (in the box) on LD. The flexible LD carries S-acyldihydrolipoamide to the active site in the E2 core to generate acyl-CoA. The reduced lipoyl moiety on LD is oxidized by E3 on BD with the concomitant reduction of nicotinamide adenine dinucleotide (NAD⁺). The sum of the above component reactions is the oxidative decarboxylation of BCKAs (reaction [1] in the text). Reproduced from Aevarsson A, Seger K, Turley S, Sokatch JR, and Hol WG (1999) Crystal structure of 2-oxoisovalerate dehydrogenase and the architecture of 2-oxo acid dehydrogenase multienzyme complexes. *Nature Structural Biology* 6: 785–792, with permission from Nature.

proportion of active or de-phosphorylated enzyme. The activation of BCKDC in exercise is caused by the dissociation of BCKD kinase from the BCKDC, which results in the inactivation of BCKD kinase. Glucocorticoids and acidification also significantly increase the activity state of the BCKDC in pig kidney cells expressing the hormone receptor. Glucocorticoids causes a marked reduction in the BCKD kinase mRNA level in rat hepatoma cells and rat liver, which explains the increased activity state of BCKDC.

The pivotal role of reversible phosphorylation in downregulating BCAA oxidation was corroborated by findings from a BCKD kinase-deficient mouse model. BCKDC activity is greatly enhanced in most tissues, which correlates with markedly lower BCAA concentrations. Neurological abnormalities are prevalent in hindlimb flexion, and notably apleptic seizures occur after 6–7 months of age. The phenotype of the BCKD kinase knockout mice establishes that the tight control of BCAA oxidation is critical for normal growth and neurological functions. Conversely, the PP2Cm phosphatase has been identified as the BCKD phosphatase that binds to the BCKDC and specifically dephosphorylates the phosphoserine residue at position 293 of the E1 protein. PP2Cm-deficient mice manifested defects in BCAA oxidation and significantly produced a metabolic phenotype similar to the intermediate or intermittent MSUD. These outcomes support PP2Cm as the endogenous BCKD phosphatase, which is essential for the upregulation of BCKDC activity. The results also suggest that defects in the BCKD phosphatase may be responsible for a subset of MSUD.

The Supramolecular BCAA Metabolon

The concept that metabolic enzymes associate to form a supramolecular structure was developed over 50 years ago, which was termed 'metabolon'. The advantage of such a supramolecular assembly is to channel substrates between two juxtaposed steps in a metabolic pathway for maximal efficiency in an appropriate intracellular location. Recent studies showed that the BCATm is associated with the E1 component of human BCKDC to produce the BCAA metabolon. The dissociation constant (K_d) for the BCAA metabolon is 2.8 μM as determined by isothermal calorimetry titration. When coupled with BCATm, the k_{cat} values for the E1-catalyzed decarboxylation of BCKAs are enhanced by 12-fold. Mutations of the BCATm protein in the catalytically important CXXC center or of the E1 protein in the phosphorylation loop prevent the supramolecular complex formation. The results suggest that the formation of the BCAA metabolon can be influenced by the redox state in the mitochondrial matrix. The BCAA metabolon was further shown to contain GDH1 and some other proteins. Kinetic results suggest that GDH1 facilitates regeneration of the BCATm form that binds to the E1 protein of BCKDC. This promotes the metabolon formation, BCAA oxidation, and the cycling of nitrogen through GDH1.

Brain Neuropathology of BCKAs

The BCAAs, in particular leucine, are rapidly transported into the brain and actively metabolized. It was proposed that a leucine–glutamate cycle plays an important role in

maintaining a steady supply of glutamate, a major excitatory neurotransmitter for interneuronal communication. The BCAAs are nitrogen donors for glutamate synthesis in astrocytes, a major site of BCAA transamination, which is transferred to α -ketoglutarate to yield glutamate. This amino acid is, in turn, converted into glutamine.¹⁵ N-labeled BCAA studies showed that at least one-third of the amino groups of brain glutamate are derived from BCAAs. The BCKAs may be released from astrocytes to the extracellular fluid and taken up by neurons. Although neurons can oxidize KIC, it is preferentially re-aminated to leucine. The flux of the reverse transamination in rat cortical synaptosomes is several times greater than the rate of nitrogen transfer from leucine to glutamate. This is in contrast to the flux in astrocytes which is mainly for glutamate synthesis. In the leucine oxidative pathway, acetoacetyl-CoA is generated for ketone synthesis or for cleavage to acetyl-CoA to enter the tricarboxylic acid cycle.

In the classic form of MSUD, the excess BCAAs and BCKAs are likely to interfere with the neuronal and astrocytic metabolism. BCKAs accumulated in MSUD compromise brain energy metabolism by blocking the respiratory chain and inhibiting creatine kinase activity. Perfusion by microdialysis with leucine and KIV creates a microenvironment similar to that found in MSUD. The infusion of leucine resulted in an increase of large neutral amino acids in the extracellular space, thereby a decreased concentration in neurons. The infusion of KIC caused an 11-fold increase in leucine and two- to threefold increase in other large neutral amino acids in the extracellular space. This pattern is consistent with active transamination of KIC to leucine. These changes could affect biosynthesis of serotonin and catecholamines, and alter the homeostasis of the leucine/glutamate cycle and the glutamate/glutamine cycle in brain.

See also: [Metabolism Vitamins and Hormones: Amino Acid Metabolism; Biochemistry of Thiamine and Thiamine Phosphate Compounds; Structure and Regulation of Pyruvate Dehydrogenases; Tricarboxylic Acid Cycle; Signaling: mTOR and its Downstream Targets.](#)

Further Reading

- Aevarsson A, Seger K, Turley S, Sokatch JR, and Hol WG (1999) Crystal structure of 2-oxoisovalerate dehydrogenase and the architecture of 2-oxo acid dehydrogenase multienzyme complexes. *Nature Structural Biology* 6: 785–792.
- Avruch J, Long X, Ortiz-Vega S, et al. (2009) Amino acid regulation of TOR complex 1. *American Journal of Physiology – Endocrinology and Metabolism* 296: E592–E602.
- Chuang DT and Shih VE (2001) Maple syrup urine disease (branched-chain ketoaciduria). In: Scriver CR, Beaudet AL, Sly WS, and Valle D (eds.) *The Metabolic and Molecular Bases of Inherited Disease*, 8th edn., pp. 1971–2006. New York, NY: McGraw-Hill.
- Fried SK and Watford M (2007) Leucine weight with a futile cycle. *Cell Metabolism* 6: 155–156.
- Harper AE, Miller RH, and Block KP (1984) Branched-chain amino acid metabolism. *Annual Review of Nutrition* 4: 409–454.
- Harris RA, Hawes JW, Popov KM, et al. (1997) Studies on the regulation of the mitochondrial α -ketoacid dehydrogenase complexes and their kinases. *Advances in Enzyme Regulation* 37: 271–293.
- Holecek M (2010) Three targets of branched-chain amino acid supplementation in the treatment of liver disease. *Nutrition* 26: 482–490.

- Islam MM, Wallin R, Wynn RM, et al. (2007) A novel branched-chain amino acid metabolon. Protein–protein interactions in a supramolecular complex. *Journal of Biological Chemistry* 282: 11893–11903.
- Joshi MA, Jeoung NH, Obayashi M, et al. (2006) Impaired growth and neurological abnormalities in branched-chain alpha-keto acid dehydrogenase kinase-deficient mice. *Biochemical Journal* 400: 153–162.
- Lu G, Sun H, She P, et al. (2009) Protein phosphatase 2Cm is a critical regulator of branched-chain amino acid catabolism in mice and cultured cells. *Journal of Clinical Investigation* 119: 1678–1687.
- Lynch CJ, Halle B, Fujii H, et al. (2003) Potential role of leucine metabolism in the leucine-signaling pathway involving mTOR. *American Journal of Physiology – Endocrinology and Metabolism* 285: E854–E863.
- Perham RN (2000) Swinging arms and swinging domains in multifunctional enzymes: Catalytic machines for multistep reactions. *Annual Review of Biochemistry* 69: 961–1004.
- She P, Reid TM, Bronson SK, et al. (2007) Disruption of BCATm in mice leads to increased energy expenditure associated with the activation of a futile protein turnover cycle. *Cell Metabolism* 6: 181–194.
- Stipanuk MH (2007) Leucine and protein synthesis: mTOR and beyond. *Nutrition Reviews* 65: 122–129.
- Yang J, Chi Y, Burkhardt BR, Guan Y, and Wolf BA (2010) Leucine metabolism in regulation of insulin secretion from pancreatic beta cells. *Nutrition Reviews* 68: 270–279.
- Yudkoff M (1997) Brain metabolism of branched-chain amino acids. *Glia* 21: 92–98.

Brassinosteroids

S D Clouse, North Carolina State University, Raleigh, NC, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Brassinolide The most active brassinosteroid. It affects cell elongation and division at nanomolar levels. A unique polyhydroxylated steroid with a C-6, C-7 lactone.

Gas chromatography–mass spectrometry with selected ion monitoring (GC–MS–SIM) Technique used with

deuterated standards to determine endogenous levels of brassinosteroids in plants.

Receptor kinase A signal-transduction protein with multiple domains, including an extracellular ligand-binding domain, a membrane-anchoring region, and an intracellular kinase. Numerous receptor kinases occur in various cell types in both plants and animals.

Brassinosteroids (BRs) are endogenous plant growth-promoting hormones that function at low concentrations to influence cellular expansion and proliferation, while interacting with other plant hormones and environmental factors to regulate the overall form and function of the plant. BRs are found throughout the plant kingdom in seeds, pollen, and young vegetative tissue, and the examination of the phenotype of mutants affected in BR biosynthesis or signaling provides genetic evidence that BRs are essential for normal organ elongation, vascular differentiation, male fertility, timing of senescence, and leaf development. BRs also participate in the regulation of plant responses to light, flowering time, and seed germination. They are unique among plant hormones in their close structural similarity to vertebrate and invertebrate steroid hormones, which have well-known roles in regulating embryonic and post-embryonic development and adult homeostasis in mammals and insects.

BR Structure and Natural Occurrence

BRs have been identified throughout the plant kingdom as naturally occurring compounds in monocot and dicot angiosperms, gymnosperms, algae, and pteridophytes. At least 40 free BRs and four conjugates have been rigorously characterized from plant tissue by a variety of biochemical approaches.

Discovery and Chemical Structure of Brassinolide

The discovery of brassinolide, the most active naturally occurring BR currently identified, was preceded by three decades of experiments at the United States Department of Agriculture (USDA), in which organic extracts of pollen from over 60 species were applied to a variety of crop plants to identify new compounds with growth-promoting properties. An extract from *Brassica napus* pollen was extremely potent in promoting cell elongation and division in bean second internodes and enhanced overall growth of radishes, leafy vegetables, and potatoes when young seedlings were sprayed in greenhouse experiments. USDA researchers identified the active component of the *B. napus* extract in 1979 and named the novel compound brassinolide. It was determined by single-crystal X-ray analysis to be a polyhydroxylated derivative of 5α -cholestane, namely

(22R,23R,24S)- $2\alpha,3\alpha,22,23$ -tetrahydroxy-24-methyl-B-homo-7-oxa- 5α -cholestan-6-one (Figure 1). Parallel work in a number of Japanese laboratories also identified BRs from a variety of plant tissues.

Other BRs differ from brassinolide by variations at C-2 and C-3, the presence of a ketone or de-oxo function instead of a lactone at C-6, various substitutions at C-24, and the stereochemistry of the hydroxyl groups in the side chain. Known BR conjugates include glycosylated, myristylated, and laurylated derivatives of the hydroxyls at C-2 and C-3 or in the side chain. Many BRs are biosynthetic precursors or metabolic products of brassinolide, although some of these BRs are believed to have independent biological activity in specific plants.

Endogenous Levels of BRs in the Plant Kingdom

Endogenous levels of BRs vary according to plant organ type, tissue age, and species, with pollen and immature seeds containing the highest levels, followed by young growing shoots. Pollen and immature seeds generally show BR levels of $1\text{--}100\text{ ng g}^{-1}\text{ fw}$; shoots and leaves typically have lower amounts, $0.01\text{--}0.1\text{ ng g}^{-1}\text{ fw}$. BRs are isolated from plant tissues by organic solvent extraction followed by reverse-phase high-performance liquid chromatography. Accurate quantification of endogenous levels involves spiking with deuterated forms of the BR of interest followed by gas chromatography–mass spectrometry with selected ion monitoring (GC–MS–SIM) analysis of derivatized BR samples.

BR Biosynthesis

The BR biosynthetic pathway can be divided into general sterol synthesis (cycloartenol to campesterol) and the BR-specific pathway from campesterol to brassinolide. Details of the pathway have been obtained by following the fate of labeled intermediates fed to cell-suspension cultures and whole plants, and by measuring endogenous sterol and BR levels in biosynthetic mutants blocked at various steps of the pathway.

Mevalonic Acid to Cycloartenol

Like animal steroids, BRs are products of the isoprenoid biosynthetic pathway beginning with acetyl-CoA and proceeding through intermediates such as mevalonate, isopentenyl

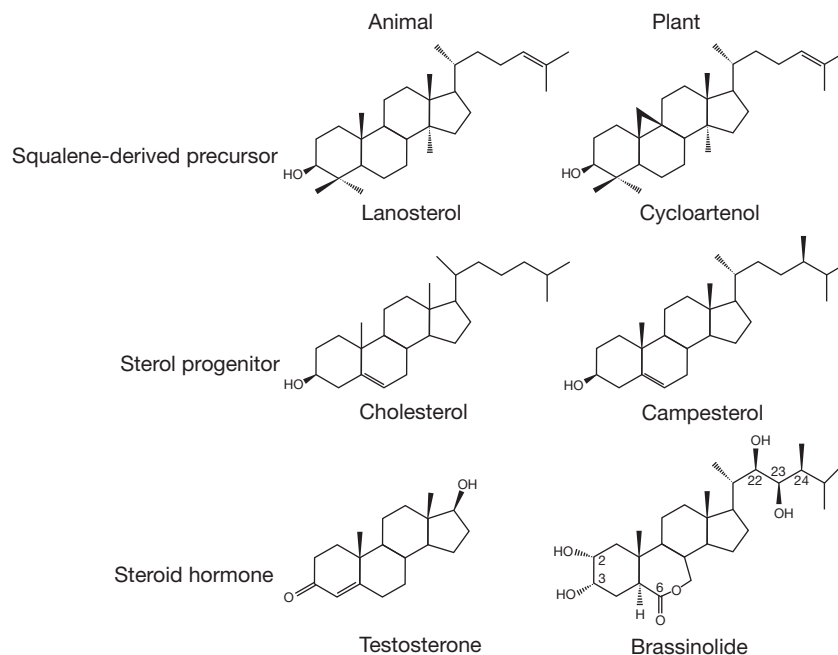


Figure 1 Structure of brassinolide and its sterol progenitor is compared with the animal steroid hormone testosterone and its sterol progenitor.

pyrophosphate, farnesyl pyrophosphate, and squalene-2,3-epoxide. It appears that the biosynthetic steps from mevalonate to squalene-2,3-epoxide are conserved among the phyla, but in animals and fungi squalene-2,3-epoxide is converted to lanosterol, the precursor of cholesterol, whereas in plants it is generally converted to cycloartenol, the progenitor of the plant sterols campesterol, stigmasterol, and sitosterol (Figure 1).

Cycloartenol to Campesterol

The conversion of cycloartenol to campesterol begins with C-24 alkylation of the side chain catalyzed by the enzyme sterol methyl transferase 1 (SMT1) in the presence of S-adenosylmethionine. The product, 24-methylenecycloartenol, is converted to 24-methylenelophenol in a series of steps including a C-14 sterol reductase and a Δ^7 - Δ^8 sterol isomerase, encoded in *Arabidopsis* by the *FAKEL* and *HYDRA1* genes, respectively. Mutants in the *SMT1*, *FAKEL*, and *HYDRA1* genes show dwarfism and severe defects in embryogenesis and vascular development that are not rescued by BR treatment, possibly because the early sterol pathway may produce a non-BR signaling molecule that is critical for these developmental processes.

The conversion of 24-methylenelophenol to campesterol involves a Δ^7 -C-5-desaturase, encoded by *DWARF7* (*DWF7*); a Δ^7 -sterol reductase, encoded by *DWARF5* (*DWF5*); and a C-24 sterol reductase, encoded by *DWARF1* (*DWF1*). *Arabidopsis* *dwf7*, *dwf5*, and *dwf1* mutants are intermediate dwarfs with altered vascular development and have reduced BR levels. These developmental defects can be rescued by treatment with exogenous BRs.

Campesterol to Brassinolide

In the classical pathway of BR biosynthesis, four reactions lead from campesterol to campestanol via reduction of the C-5

double bond. One of these steps requires a 5α -sterol reductase (encoded by the *DE-ETIOLATED2*, *DET2*, gene) that is an ortholog of the mammalian enzyme that catalyzes the reduced nicotinamide adenine dinucleotide phosphate (NADPH)-dependent reduction of testosterone to dihydrotestosterone. From campestanol, two biosynthetic routes are possible: early C-6 oxidation, in which a ketone is introduced at C-6 before the hydroxylation of the side chain, and late C-6 oxidation, in which the side chain is modified first and the ketone is introduced in the penultimate step. Many plants, including *Arabidopsis*, have both pathways functional, whereas others, such as tomato, appear to use only late C-6 oxidation. Hydroxylation of the side chain in either pathway occurs via successive steps involving several cytochrome P450 steroid hydroxylases, including *DWARF4* (*DWF4*). Multiple pathways from campesterol to brassinolide may occur in the same plant and brassinolide biosynthesis appears to be more of a metabolic grid than two independent linear pathways of early and late C-6 oxidation. Recent evidence also suggests that C-22 hydroxylation of campesterol by *DWF4* most likely occurs before the C-5 reduction catalyzed by *DET2*. The most direct biosynthetic route from campesterol to brassinolide, based on current experimental evidence, is shown in Figure 2.

The *det2* and *dwf4* mutants of *Arabidopsis* are extremely dwarf in stature; have dark-green, rounded, and downward-curling leaves; and exhibit a prolonged life span with reduced fertility and altered vascular development. In the dark, these mutants show some of the features of light-grown plants, including shortened hypocotyls and open cotyledons. All these phenotypic alterations can be rescued to wild type by exogenous application of brassinolide or BR intermediates downstream of the biosynthetic block caused by the mutant. Altering *DWF4* gene expression in *Arabidopsis* and rice results in altered plant growth and development, providing further evidence of the importance of regulating BR biosynthesis for normal plant growth.

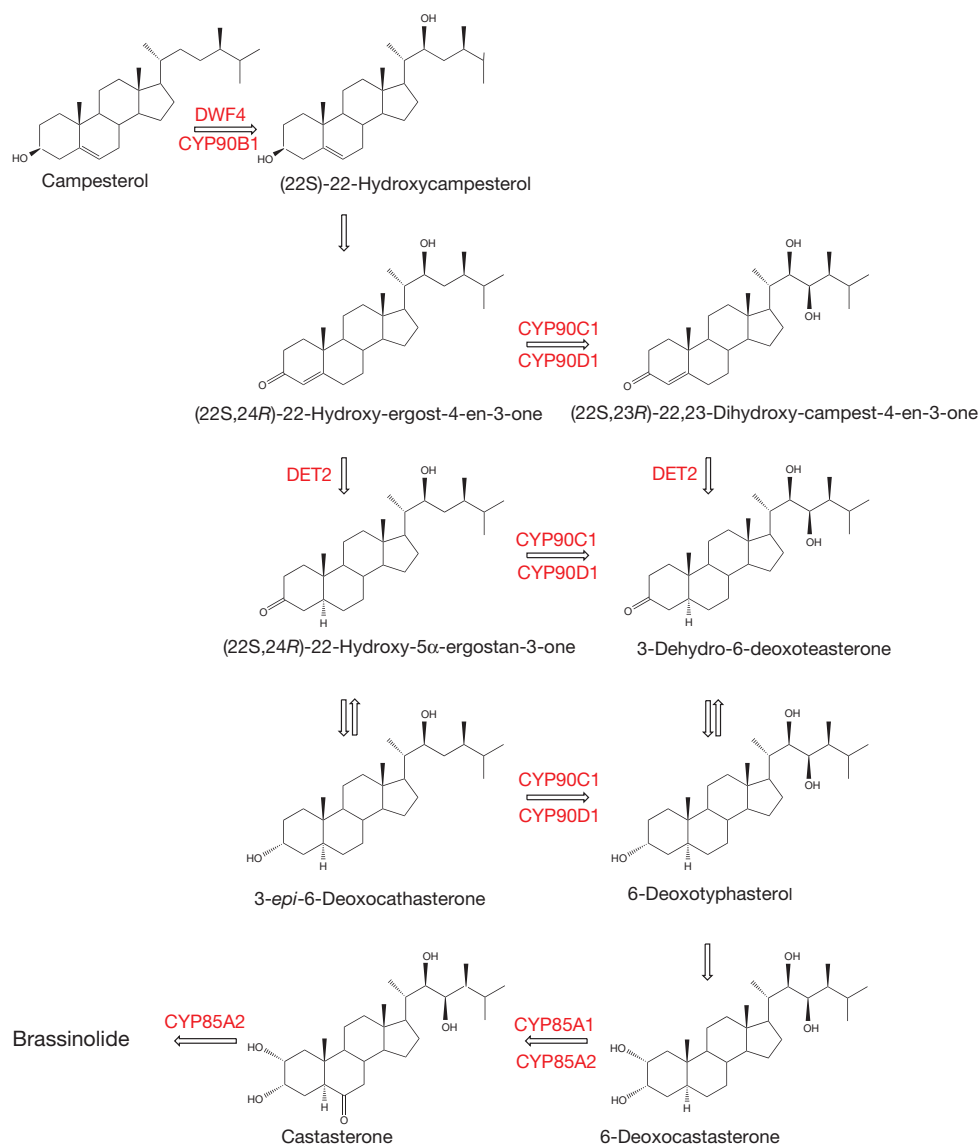


Figure 2 Biosynthetic pathway leading from the plant sterol campesterol to brassinolide. From campestanol, multiple alternative pathways exist and several points of interconnection are possible between the pathways. The most direct biosynthetic route based on recent experimental evidence is shown. DET2 is a steroid 5- α reductase and various CYP proteins are cytochrome P450 enzymes.

BR Physiology

BRs have a dramatic positive effect on stem elongation, promoting epicotyl, hypocotyl, and peduncle elongation in dicots and enhancing coleoptile and mesocotyl growth in monocots. Exogenous BRs also stimulate tracheary element differentiation and numerous studies *in vivo* suggest that endogenous BRs are essential for normal vascular development. In many systems, BRs increase rates of cell division, particularly under conditions of limiting auxin and cytokinin. BRs also promote seed germination, accelerate senescence, cause hyperpolarization of membranes, stimulate ATPase activity, and alter the orientation of cortical microtubules. In addition to effects on growth, BRs also mediate the effects of abiotic and biotic stresses, including salt and drought stress, temperature extremes, and pathogen attack.

BR Signal Transduction

Recently, it has been shown that animal steroids can be recognized by cell-surface receptors. However, the classic signaling pathway for these hormones involves binding the steroid to an intracellular receptor consisting of a variable N-terminal domain, a highly conserved DNA-binding domain with two zinc fingers, and a multifunctional domain that mediates ligand binding, dimerization, and ligand-dependent transcriptional activation. The completed genome sequence of *Arabidopsis thaliana* indicates that plants do not contain members of this superfamily of intracellular steroid receptors, suggesting that cell-surface recognition is the primary, if not only, form of plant steroid perception. Thus, plant and animal steroid hormones share many structural and functional features, but differ in their primary signaling pathways.

Receptor Kinases and BR Perception

The best-characterized component of BR signal transduction is *BRASSINOSTEROIDINSENSITIVE 1 (BRI1)*, a single genetic locus in *Arabidopsis* encoding a leucine-rich repeat (LRR) receptor kinase. Homologous genes have been identified in rice, tomato, and pea. Numerous *bri1* mutant alleles have been identified by various genetic screens, most of which exhibit the extreme dwarfism and other phenotypic characteristics of severe BR-deficient mutants. In contrast to the biosynthetic mutants, *bri1* mutants cannot be rescued by BR treatment, consistent with their role in signal transduction.

The BRI1 protein consists of a putative signal peptide followed by the extracellular domain proper including a leucine zipper and 25 LRRs that are flanked by short sequences containing paired cysteines. A 70-amino-acid island that is critical

for biological function is embedded between repeats 21 and 22. Downstream of the extracellular domain lies a short hydrophobic single-pass transmembrane domain, followed by the juxtamembrane region and the cytoplasmic kinase domain. Binding studies with radiolabeled brassinolide and microsomal fractions of wild-type, mutant, and transgenic *Arabidopsis* plants clearly showed that BRI1 is an essential component of the BR-receptor complex. Subsequent experiments using photo affinity-labeled BRs demonstrated that BR binds directly to BRI1 without an intermediate accessory protein. The 70-amino-acid island domain in conjunction with the adjacent C-terminal LRR was necessary and sufficient for BL binding, thus defining a novel 94-amino-acid plant steroid-binding motif.

The mechanism of action of many animal receptor kinases involves ligand-mediated homo- or heterodimerization of the

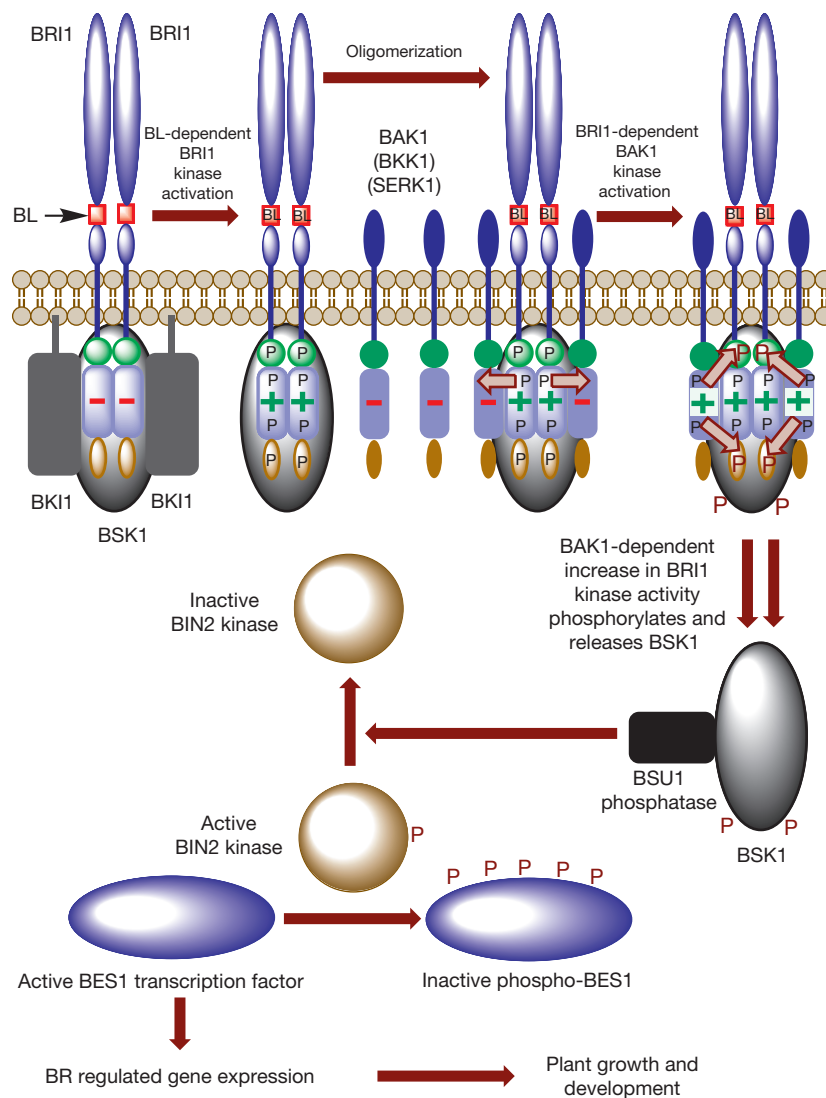


Figure 3 A model for BR signal transduction in *Arabidopsis*. BR binds to the BRI1 receptor kinase, activating the intracellular kinase domain, which releases the negative regulator BK1 and allows the BRI1/BAK1 heterotetramer to form. Sequential transphosphorylation of the BRI1/BAK1 pair results in downstream phosphorylation of BSK1, which then interacts with and activates the BSU1 phosphatase. BSU1 removes Tyr-200 of the negative regulatory kinase, BIN2, inactivating it and allowing the unphosphorylated forms of the BES1/BZR1 transcription factors to accumulate in the nucleus where they regulate the expression of genes associated with the BR response.

receptor followed by autophosphorylation of the intracellular kinase domain. This activation of the kinase results in the recognition and phosphorylation of downstream signaling components, leading ultimately to the regulation of specific gene expression. Plant receptor kinases appear to follow the same general paradigm of receptor kinase action, and mutational analysis shows that both a functional extracellular domain and an active cytoplasmic kinase are essential for BRI1 function. Identification of BRI1 phosphorylation sites by mass spectrometry followed by a functional analysis of their biochemical and biological significance, clearly demonstrated that phosphorylation plays a key role in BRI1 function.

Recent evidence also confirms that oligomerization with another LRR receptor kinase plays an important role in BR signaling. BAK1 (BRI1 ASSOCIATED RECEPTOR KINASE 1) shares similar structural organization to BRI1 except that it has only five LRRs in its extracellular domain and lacks the 70-amino-acid island of BRI1. BAK1 is expressed in all tissues of the plant, similar to the global expression pattern of BRI1, and confocal laser microscopy of transgenic plants expressing fusion proteins shows plasma membrane as well as endosomal membrane localization for both BRI1 and BAK1. Direct physical interaction between BRI1 and BAK1 was confirmed in yeast cells, plant protoplasts, and whole *Arabidopsis* plants by a variety of biochemical approaches. Moreover, genetic studies strongly suggest that BAK1 plays an important role in BR signaling. A detailed dissection of the BRI1/BAK1 interaction resulted in the sequential transphosphorylation model of BR signaling, in which BRI1 controls signaling specificity by direct BR binding followed by downstream substrate phosphorylation. The co-receptor BAK1 is then activated by BRI1-dependent transphosphorylation and subsequently enhances signaling output through reciprocal BRI1 transphosphorylation. This model suggests both conservation and distinct differences between the molecular mechanisms regulating phosphorylation-dependent receptor kinase activation in plants and animals.

Downstream Components

Besides BAK1, several other BRI1-interacting proteins have been identified with important functions in BR signal transduction. BKI1 (BRI1 KINASE INHIBITOR 1) is a negative regulator of BR signaling which binds to BRI1 and is membrane associated in the absence of BR. BR binding to BRI1 causes dissociation of BKI1 from the membrane, which releases repression of the pathway by allowing BRI1 and BAK1 to hetero-oligomerize to initiate BR signaling. Another membrane-associated BRI1-interacting protein, BSK1 (BR SIGNALING KINASE 1), is a positive regulator of BR signaling which belongs to the cytoplasmic receptor-like kinase (RLCK) subfamily containing an N-terminal kinase domain and a C-terminal tetratricopeptide motif. BR-dependent phosphorylation of BSK1 by BRI1 allows

BSK1 to interact with another positive regulator of BR signaling, BSU1 (BRI1 SUPPRESSOR 1), a protein phosphatase that propagates the BR signal downstream by dephosphorylating and in activating a downstream negative regulator of BR signaling, the BIN2 (BRASSINOSTEROID INSENSITIVE 2) kinase.

BIN2, which is homologous to insect and mammalian shaggy-like kinases involved in the Wntless/Wnt-signaling pathways, acts by phosphorylating two related transcription factors, BZR1 and BES1. Phosphorylation by BIN2 targets BZR1 and BES1 for proteasome-mediated degradation, inhibits binding to promoters of BR-regulated genes, and promotes binding to 14–3–3 proteins resulting in cytoplasmic retention of the transcription factors. In the presence of BR, BIN2 activity is inhibited by BSU1 dephosphorylation on Tyr-200 and the nonphosphorylated forms of BZR1 and BES1 accumulate in the nucleus where they interact with other transcription factors to regulate the expression of specific genes involved in the BR response. Microarray analyses have identified hundreds of BR-regulated genes including those encoding wall-modifying proteins involved in cell expansion, genes regulating cell division, and numerous transcription factors including those also regulated by the plant growth-promoting hormone, auxin. BR signal transduction has become one of the best-studied plant-signaling pathways involving receptor kinase action. Our current knowledge of BR signal transduction is summarized in [Figure 3](#).

See also: **Metabolism Vitamins and Hormones:** Cholesterol Synthesis and Regulation; **Signaling:** G-Protein-Coupled Receptor Kinases and Arrestins; Protein Tyrosine Phosphatases; Vascular Endothelial Growth Factor Receptors.

Further Reading

- Bishop GJ (2007) Refining the plant steroid hormone biosynthesis pathway. *Trends in Plant Science* 12: 377–380.
- Chinchilla D, Shan L, He P, de Vries S, and Kemmerling B (2009) One for all: The receptor-associated kinase BAK1. *Trends in Plant Science* 14: 535–541.
- Clouse SD and Sasse JM (1998) Brassinosteroids: Essential regulators of plant growth and development. *Annual Review of Plant Physiology and Plant Molecular Biology* 49: 427–451.
- Fujioka S and Yokota T (2003) Biosynthesis and metabolism of brassinosteroids. *Annual Review of Plant Biology* 54: 137–164.
- Gendron JM and Wang ZY (2007) Multiple mechanisms modulate brassinosteroid signaling. *Current Opinion in Plant Biology* 10: 436–441.
- Kim TW and Wang ZY (2010) Brassinosteroid signal transduction from receptor kinases to transcription factors. *Annual Review of Plant Biology* 61: 681–704.
- Sakurai A, Yokota T, and Clouse SD (eds.) (1999) *Brassinosteroids: Steroidal Plant Hormones*. Tokyo: Springer.
- Vert G (2008) Plant signaling: Brassinosteroids, immunity and effectors are BAK! *Current Biology* 18: R963–R965.
- Vert G, Nemhauser JL, Geldner N, Hong F, and Chory J (2005) Molecular mechanisms of steroid hormone signaling in plants. *Annual Review of Cell and Developmental Biology* 21: 177–201.

Cadherin-Mediated Cell–Cell Adhesion

M Lowndes and W J Nelson, Stanford University, Stanford, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Frauke Drees and W. James Nelson, volume 1, pp. 205–211, © 2004, Elsevier Inc.

Glossary

Adaptor proteins Cytosolic proteins that link membrane proteins to signaling and cytoskeletal complexes and allow for regulation of physiological responses. In the example of cell-adhesion proteins, adaptor proteins (catenins) provide the linkage between cell adhesion transmembrane proteins and the cytoskeleton.

Cell–cell adhesion Process of establishment and maintenance of contacts between adjacent cells in tissues and organs of multicellular organisms.

Cell–cell adhesion molecules Transmembrane glycoproteins that mediate cell–cell binding. There are three major protein families: the cadherin superfamily, the immunoglobulin superfamily, and selectins.

Cell–cell junctions Specialized regions on the cell surface through which cells are joined to each other. Tight junctions in epithelial cell layers form a ribbon-like seal between compartments. Gap junctions are protein-lined channels between two cells that allow diffusion of small molecules from one cell to the next. Adherens junctions and desmosomes are dense protein plaques connected to cytoskeletal networks that link two adjacent cells.

Cytoskeleton Network of fibrous proteins found in the cytosol of cells that provides structural support for the cell, determines cell shape and motility, and allows directional movement of organelles and vesicles inside the cell. There are three major classes of cytoskeletal filaments: actin microfilaments, microtubules, and intermediate filaments.

Cadherin Superfamily

Cell–cell adhesion complexes are essential for the organization and maintenance of complex tissues in multicellular organisms. Proteins that participate in cell–cell adhesion include the cadherin superfamily, the immunoglobulin superfamily, and selectins. The cadherin superfamily comprises transmembrane glycoproteins found in all metazoa, and shares a conserved extracellular domain that contains many calcium-binding domains. The four major subfamilies include classical cadherins, desmosomal cadherins, protocadherins, and atypical cadherins.

Classical Cadherins (type I and II)

The classical cadherins are characterized in terms of structure, mechanism of adhesion, and function. They generally have five extracellular repeats (EC 1–5), a single transmembrane domain, and an intracellular domain that binds the adaptor proteins β -catenin, γ -catenin, p120, and hankai (Figure 1). β -Catenin binds α -catenin that directly or indirectly interacts with the actin cytoskeleton. Classical cadherins are classified as type I or type II depending on the adaptor proteins that bind to the cytoplasmic domain.

While expressed broadly in all tissues, some classical cadherin subtypes are more abundant in specific tissue types than in others (e.g., epithelial, E-cadherin; nervous, N-cadherin; placenta, and P-cadherin), but, generally, multiple forms are expressed in different ratios in the same tissue. Classical cadherins mostly form homophilic bonds, such as E-cadherin with E-cadherin on the adjacent cell, but heterophilic interactions also occur. Classical cadherins cluster along the plasma membrane, forming an adhesion complex known as the ‘adherens junction’ (AJ). Nectins, a class of immunoglobulin G (IgG) superfamily cell–cell adhesion proteins, localize there as well and contribute to the formation and maintenance of the cell–cell adhesion complex. Classical cadherins are spatially and temporally regulated throughout development and are crucial for the formation of cell layers and complex tissue structures.

Desmosomal Cadherins

Desmosomal cadherins are the major components of the desmosome, a dense adhesion complex required for tissues to withstand mechanical stress. There are two types of desmosomal cadherins with multiple isoforms and splice variants: desmogleins (1–4) and desmocollins (1–3; a and b). Desmosomal cadherins are thought to form heterophilic bonds between desmoglein and desmocollin, but there is also evidence for

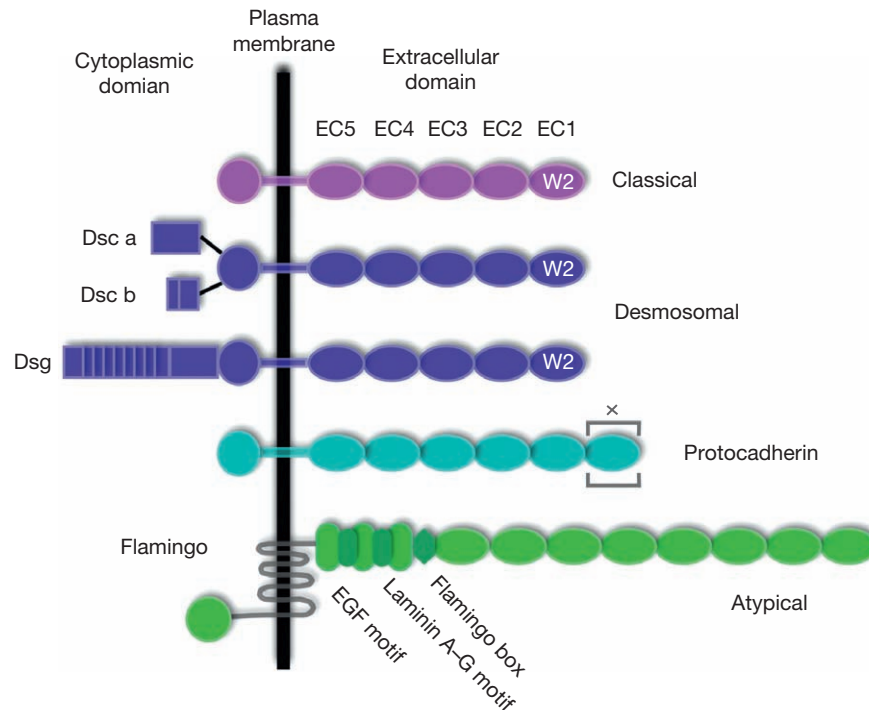


Figure 1 Members of the cadherin superfamily. Classical and desmosomal cadherins contain EC1–5 domains with a tryptophan at position 2 (W2). Whereas desmosomal cadherins have a similar extracellular domain to classical cadherins, their cytoplasmic domain differs. There are two types of desmosomal cadherins: desmoglein (Dsg) and two splice variants of desmocollin (Dsc a,b). Protocadherins, such as the β subtype, have varying numbers of cadherin domains, marked by [X]. Atypical cadherins contain cadherin domains but share no other commonalities with other cadherin subtypes. One example, flamingo, names after its Flamingo Box, is a seven-pass transmembrane protein.

homophilic interactions. Apart from the type 2 isoforms, desmosomal cadherins are only expressed in stratified epithelia, in a differentiation-specific manner within the layers of the epidermis. Similar to classical cadherins, desmosomal cadherins have five EC repeats, a single transmembrane domain, and an intracellular domain that binds to adaptor proteins (Figure 1). The structural organization of desmosomes is well defined. Plakoglobin and plakophilin bind to the cytoplasmic domain of cadherins, and desmoplakin links the complex to the intermediate filament cytoskeleton. The desmosomal complex is present in most higher metazoa, with the exception of *Drosophila* and *Caenorhabditis elegans*. While the classical cadherin-containing AJ is typically established first upon cell–cell contact and during development, the desmosome is essential for maintaining strong intercellular adhesion and resisting shearing stress across a tissue.

Protocadherins and Atypical Cadherins

The largest cadherin subfamily is the protocadherins, which consists of more than 50 subtypes. Three exons encode a ‘constant’ region, which is family specific and cytoplasmic. The other part of the protein is encoded by a single large ‘variable’ exon. The combination of the intracellular ‘constant’ and ‘variable’ extracellular domains provides potentially thousands of combinations, similar to immunoglobulin proteins. Protocadherins are most commonly found in the nervous system and their diversity may contribute to nervous system development and organization. While there is evidence for homophilic binding, the regulation and strength of protocadherins is most likely

different from classical cadherins. Heterophilic binding may prove more important for protocadherin interactions. A few binding partners have been identified, among them Fyn kinase.

Atypical cadherins contain one or more cadherin-like extracellular repeat domains but have no other resemblance to the other cadherins. The atypical subfamily includes Fat-like, seven-pass transmembrane (flamingo), and the Dcad102F-like cadherins. They do not appear to participate in cell–cell adhesion but rather in different signaling pathways, including planar cell polarity (PCP), Dpp and Hippo pathways. A summary of the structural differences between the subfamilies is shown in Figure 1.

Regulation of Cadherin Adhesion

Cadherin proteins have two major structural and functional domains: the extracellular domain, which binds to another cadherin on an adjacent cell; and the cytoplasmic domain, which binds cytosolic adaptor proteins that link the complex to the cytoskeleton. The exact mechanism how cadherins achieve cell–cell adhesion is still an area of much research. Most knowledge comes from studies on the classical and desmosomal cadherins, and therefore this article will focus on them.

Role of the Extracellular Domain

Extracellular domain structure

The extracellular domain of cadherins is characterized by the ECs. Three calcium molecules bind successively between the repeats, giving the protein a curved structure and rigidifying it.

A specific binding structure of these tandem repeats has been hypothesized based on the recent crystal structure of the entire ectodomain of C-cadherin (classical cadherin). The structure suggests the importance of the tryptophan at position two (W2) in the EC1 domain, which is conserved in both classical and desmosomal cadherins (Figure 1). This residue contributes to the formation of a strand-swapped dimer, connecting two opposing cadherin EC1 domains in a *trans* configuration. Mutations in the W2 residue result in lethality in zebrafish (termed the *glass onion* mutant) and substitution with alanine (W2A) results in loss of Ca^{2+} -dependent, cadherin-mediated cell–cell adhesion. There is also the possibility of a lateral *cis* dimer between cadherins on the same cell, although it is likely to be close to the same interface as the *trans* interaction. Additional crystal-structure studies reveal a second configuration, termed the ‘X dimer’ that is mediated by residues near the EC1–EC2 calcium-binding sites and is thought to be an intermediate that facilitates the final EC1–EC1 *trans*-dimer.

In contrast to the AJ, desmosomes have an electron-dense midline between the intercellular space and the plasma membrane. The appearance of a midline appears to correlate with a hyper-adhesive and Ca^{2+} -independent state of desmosome adhesion. Desmosomes in embryonic tissue and wounded epithelia are Ca^{2+} dependent and most often do not form such a midline. Three-dimensional reconstructions based on electron-microscopy-tomography images and X-ray-crystal structures of C-cadherin have been used to examine the structure of the extracellular intermolecular interface. Two possible explanations have been proposed: desmosomal cadherins undergo a locking mechanism that creates a periodic arrangement, trapping Ca^{2+} , and resulting in strong extracellular Ca^{2+} -independent adhesive properties, or the desmosomal cadherins form a series of densely tangled knots with no regular structure. A crystal structure has not yet been obtained for desmosomal cadherins.

Mechanisms of cadherin specificity in cell sorting

Initial experiments conducted on classical cadherins demonstrated that cells expressing specific cadherin subtypes segregate into homotypic aggregates when mixed. During development of certain tissues, expression levels of cadherin subtypes switch, indicating that specific combinations of cadherins are important for cell–cell adhesion and proper development. However, additional studies using cadherin-coated beads or cadherin-transfected cells in aggregation assays show promiscuity in cadherin-subtype interactions, suggesting that, for at least a subclass of cadherins, homotypic binding is not as strict as previously thought. Several factors, such as the level of expression on the cell surface and the level of shear force during mixing, affects adhesion specificity. Affinity differences between cadherins can explain the dependence on shear forces.

By contrast, desmosomal cadherins are traditionally thought to form heterotypic adhesion (e.g., desmoglein–desmocollin). Both desmocollin and desmoglein are necessary for strong cell–cell adhesion but the specific interactions of these cadherins remains unclear. Biochemical studies with purified desmosomal cadherins show heterophilic and homophilic interactions, assessed using Biacore and single-molecule atomic-force microscopy. A more comprehensive analysis of desmosomal cadherin interactions will need to be conducted with all possible combinations under the same experimental conditions.

Role of the Cytoplasmic Domain

Linkage of cadherin to adaptor proteins

The cytoplasmic domain of classical and desmosomal cadherins recruits cytoplasmic adaptor proteins that belong to the catenin family: β -catenin, plakoglobin (γ -catenin), and p120. Typically, classical cadherins bind β -catenin, which, in turn, binds another catenin family member, α -catenin. α -Catenin is a monomer or homodimer, and *in vitro* reconstitution studies show that monomeric and dimeric α -catenin have different binding specificities. Monomeric α -catenin binds to β -catenin but the α -catenin homodimer cannot and instead, the homodimer has a higher affinity for F-actin.

p120 and hakai bind directly to the juxtamembrane domain of classical cadherins and controls cadherin turnover. Plakoglobin, a homolog of β -catenin, can bind to type I classical cadherins but is most often outcompeted by β -catenin due to weaker affinity for the cadherin cytoplasmic domain. Desmosomal cadherins, on the other hand, bind plakoglobin (not β -catenin) and catenin-related proteins called ‘plakophilins’ (Pkp1–3). The cytoplasmic domain of the desmosomal cadherins bind to plakoglobin and plakophilin binds to a plakin protein called ‘desmoplakin’. Plakin proteins are cytolinkers, and in the case of desmoplakin, link the desmosome complex to intermediate filaments. The cadherin complexes are depicted in Figure 2.

Linkage to the cytoskeleton

During cell–cell adhesion, cadherin–catenin complexes cluster together and cell–cell adhesion strengthens. How classical cadherin complexes are linked to the actin cytoskeleton is currently an area of debate. Actin near the complex is bundled; α -catenin homodimers most likely participate in the bundling of actin filaments, and also inhibit the Arp2/3 complex and actin branching. Mechanosensing studies show an important link between the complex and the cytoskeleton, but the mode remains unclear. Most likely, more proteins are involved in the regulation of the actin cytoskeleton by the cadherin complex.

Linkage between the cytoskeleton and desmosomes has been investigated using desmoplakin mutants. Deletion of the N-terminal domain of desmoplakin results in desmosomal complexes that have lost their connection to intermediate filaments, and, as a result, desmosomal and classical cadherin complexes intermix. The crystal structure of desmoplakin indicates a possible intermediate filament-binding domain. Intermediate filaments are less dynamic than actin and are thought to add stability and rigidity to the desmosomal complex, as well as provide physical and spatial separation from other adhesion complexes.

Adhesion complex turnover

Cadherin complex assembly and disassembly is a highly regulated process and can be regulated through phosphorylation of cadherin or β -catenin on serine/threonine and tyrosine residues. Phosphorylation affects their binding affinity for one another. Serine/threonine phosphorylation stabilizes the cadherin/catenin complex, thereby strengthening adhesion. Conversely, tyrosine phosphorylation of either cadherin or β -catenin destabilizes their bond and weakens adhesion. Although the mechanism is unclear, p120 also seems to participate in the regulation of cadherin endocytosis.

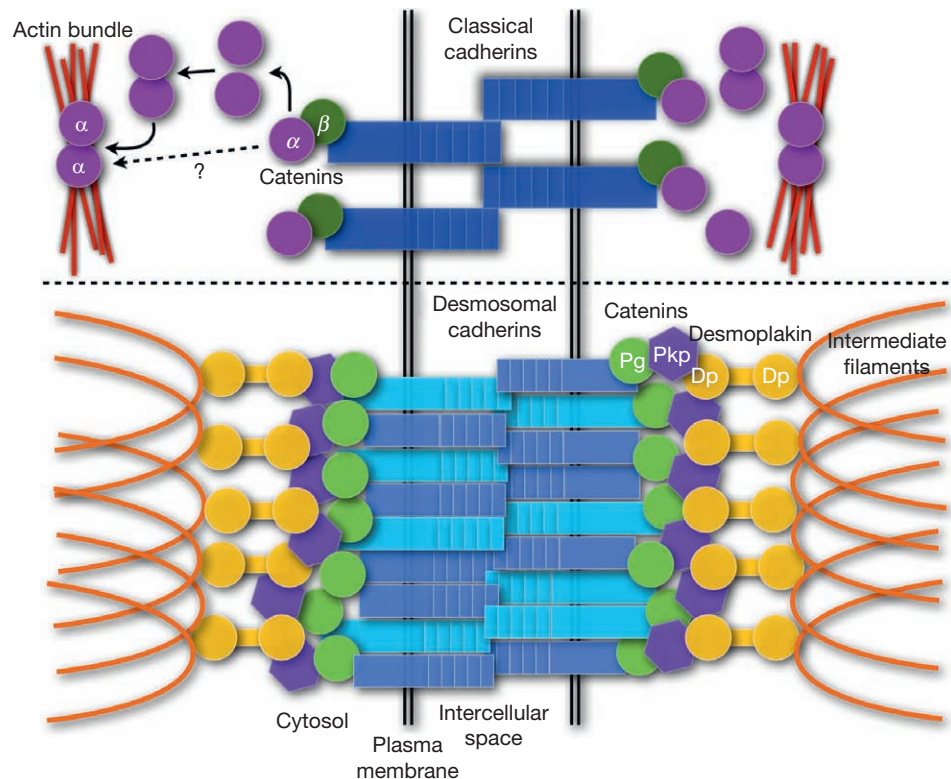


Figure 2 The adherens junction and desmosome. Top: Classical cadherins bind to one another across intracellular space (between two adjacent cells) and assemble into a complex known as the adherens junction. The cytoplasmic side of cadherins binds to β -catenin (β). β -Catenin then binds α -catenin (α), creating a ternary complex at the membrane. α -Catenin can form a homodimer, which does not bind β -catenin, but has a high affinity for F-actin and promotes actin bundling. How the α -catenin monomer and homodimer are regulated is not known. Below: Two types of desmosomal cadherins bind one another to assemble the desmosome, a dense adhesion complex. The cytoplasmic domain of desmosomal cadherins binds the catenins, plakoglobin (Pg), and plakophilin (Pkp). Desmoplakin (Dp) facilitates the binding of the complex to intermediate filaments.

Desmosomal turnover is less well understood. Within the epidermis, desmosomal cadherin subtypes have specific expression patterns. The change in distribution could be due to exchange of molecules or replacement of existing desmosomes as cells move up layers in the epidermis. In addition, the organization of desmosomes is disrupted in wounded epithelia, possibly due to protein kinase C, making the complex more extracellular- Ca^{2+} dependent and increasing its turnover rate. Whole desmosomal complexes can also be internalized as a faster means of weakening adhesion.

Functions in signaling

A number of the adaptor proteins bound to adhesion complexes also participate in signaling events. β -Catenin, plakoglobin, and p120 function as coactivators for the lymphoid enhancer-binding factor/Tcf (LEF/Tcf) family of transcription cofactors, downstream of Wnt signaling. Plakophilins are described as multifunctional scaffolds for signaling proteins at the plasma membrane. For example, plakophilin-2 is implicated in signaling pathways that regulate junction assembly, Rho-mediated actin remodeling, and transcriptional activity. It remains unclear, if the different roles for these catenin proteins are independent or regulated with respect to one another.

Role of Cadherins in Development and Disease

Cell and Tissue Morphogenesis

As discussed above, classical cadherins can sort cells based on their expression levels; it is thought that during development cell populations sort out from each other using a similar mechanism. Most of the information about the morphogenic roles of different cadherins is based on genetic deletions in mice ([Table 1](#)). Knockout of the E-cadherin gene produces the most severe phenotype with early embryonic lethality at the time of implantation due to lack of formation of the trophoblast and blastocyst cavity. Although E-cadherin is essential for development of the blastocyst, knockout of N-cadherin leads to heart defects and malformation of the neural tube, resulting in embryonic lethality at days 9–10. The developmental defects are attributed to loss of adhesion, but the distinction between the classical cadherin subtypes suggests a role for cadherins in determining cell fate as well. Similarly, loss of classical cadherin-associated adaptor proteins also causes embryonic lethality, although β -catenin knockout embryos survive implantation unlike those in which α E-catenin is deleted. This difference is due to replacement of β -catenin by plakoglobin and the loss of Wnt signaling after implantation, resulting in a failure in mesoderm formation.

Table 1 Effects of gene deletions of selected cadherins and adhesion components on mouse development

<i>Gene/protein</i>	<i>Subfamily</i>	<i>Gene defect</i>	<i>Phenotype</i>
<i>E-cadherin</i>	Classical	Knockout, deletion of extracellular Ca^{2+} -binding and adhesion motif	Embryonic lethal at time of implantation, embryonic compaction occurs, probably due to maternally deposited E-cadherin, but embryos fail to form trophectoderm or blastocyst cavity.
<i>N-cadherin</i>	Classical	Knockout	Embryonic lethal E9–10, major heart defects, and malformed neural tube and somites.
β -catenin	AJ component	Knockout	Embryonic lethal E7–9, at day 7 cells detach from ectodermal cell layer, the three germ layers and amniotic folds fail to form, and epithelial organization of ectoderm is completely lost.
α -E-catenin	AJ component	Gene trap insertion deleting carboxyterminal portion	Embryonic lethal at time of implantation, failure to form the trophoblast epithelium
Dsc-1	Desmosomal	Knockout	Normal at birth, epidermal fragility (flaking) and barrier defects, hyperproliferation and abnormal differentiation of epidermis, and hair loss.
Dsc-3	Desmosomal	Knockout	Lethal before implantation.
Dsg-2	Desmosomal	Knockout	Embryonic lethal shortly after implantation. Decreased embryonic-stem-cell proliferation.
Dsg-3	Desmosomal	Knockout	Normal at birth, disintegration of epidermis (acantholysis), hair loss, and runting (probably due to erosions in oral mucosa).
<i>Plakoglobin</i>	Desmosome Component	Knockout	Embryonic lethal E10.5–15 due to heart failure. Some genetic backgrounds have delayed embryonic lethality with epidermal blistering due to abnormal desmosome structure and loss of IF association.
<i>Desmoplakin</i>	Desmosome Component	Knockout	Embryonic lethal E6.5 due to defects in extra-embryonic tissues and failure of egg-cylinder expansion, abnormal desmosomes. Reduction in number and size of desmosomes and abnormal structure.

Desmosomes also play an important role during development, although certain desmosomal cadherin isoforms appear to be less essential. Deletion of some desmosomal cadherins leads to early lethality due to heart defects, whereas other knockouts produce viable offspring with skin abnormalities. Although deletion of desmocollin-3 causes pre-implantation lethality, certain mutations in the gene result in skin defects in postnatal pups. Knockout of desmoglein-2 also causes death shortly after implantation, most likely due to decreased embryonic-stem-cell proliferation.

It should be noted that cell–cell adhesion is crucial for normal development because cadherins also play an important role in the functional polarization of epithelial cells. Structurally, the formation of cadherin-mediated adhesion complexes and tight junctions, another cell–cell junction, establishes a cell monolayer (often a simple epithelium) that separates two biological compartments. The epithelial monolayer comprises cells with structurally and functionally different plasma-membrane domains (termed ‘apical’ and ‘basolateral’), and creates a barrier that maintains homeostasis by regulating the movement of proteins, solutes, and ions across and between cells.

As mentioned previously, atypical cadherins, such as flamingo, establish a different kind of polarity (PCP) across the cell populations, for example, in producing left/right asymmetry in the ommatidia of the *Drosophila* eye (Figure 1).

Consequences of Disruption of Cadherin Function: Cancer and Disease States

A hallmark of cancer is the loss of epithelial phenotypes, such as cell–cell adhesion along with changes in cytoskeletal organization. These changes are accompanied by an increase in

expression of proteins most common in mesenchymal cells, such as vimentin. This switch, known as the epithelial-to-mesenchymal transition (EMT), positively correlates with increased tumor invasiveness. In order for metastasis to occur, primary tumor cells must egress from the original tumor mass and migrate to other sites within the body. EMT also includes a switch in the subtype of classical cadherin, in which E-cadherin levels decrease whereas N-cadherin levels increase. This switch, combined with only partial loss of junctional components, may facilitate the transition from mesenchymal- back to epithelial-type cells (MET), and, ultimately the progression from a micro-metastases to full metastases.

Remodeling of desmosomes may also play an important role in tumor growth, invasion, and metastasis. Introduction of desmosomal proteins into invasive cells induces aggregation and decreases their invasiveness. Similar to AJ proteins, complete loss of all desmosomal proteins does not occur in metastatic tumors. Currently, however, it is unclear what part desmosomes play in EMT, metastasis, and MET.

Mutations in desmosomal components have been identified as the cause of a number of hereditary diseases. Several heart defects, such as arrhythmogenic right-ventricular cardiomyopathy (ARVC), result from mutations in plakoglobin, desmoplakin, or Pkp2. By contrast, mutations in Pkp1, Dsg1, Dsg4, or specific domains of desmoplakin cause defects in skin development. Skin defects manifest as skin fragility, blistering, or epidermal thickening, depending on the mutation. The variation in skin and heart defects suggests that each desmosomal protein has a cell-type-specific role in cell differentiation and tissue patterning. To date, no human mutations have been identified in desmocollins. In addition, an autoimmune disease called ‘pemphigus’, targets desmosomal cadherins and

causes loss of cell–cell adhesion in the epidermis and mucus membranes. The area affected depends on the isoform targeted by the autoantibodies and results in severe blistering.

In summary, cadherin-mediated cell–cell adhesion is crucial for normal development, tissue organization and establishing cell polarity. Abnormalities in cadherins or cadherin-binding proteins result in disease states, such as cancer. Many questions still remain regarding how adhesion complexes are regulated and their roles in cytoskeletal rearrangement.

See also: [Bioenergetics: Algal Hydrogen Production](#); [Lipids Carbohydrates Membranes and Membrane Proteins: Carbohydrate Chains: Enzymatic and Chemical Synthesis](#).

Further Reading

- Bass-Zubek A, Godsel LM, Delmar M, et al. (2009) Plakophilins: multifunctional scaffolds for adhesion and signaling. *Current Opinion in Cell Biology* 21(5): 708–716.
- Chidgey M and Dawson C (2007) Desmosomes: A role in cancer? *British Journal of Cancer* 96(12): 1783–1787.
- Delva E, Tucker D, Kowalczyk AP, et al. (2009) The desmosome. *Cold Spring Harbor Perspectives in Biology* 1(2): a002543.
- Desai B, Harmon R, Green KJ, et al. (2009) Desmosomes at a glance. *Journal of Cell Science* 122(Pt 24): 4401–4407.
- Frank M and Kemler R (2002) Protocadherins. *Current Opinion in Cell Biology* 14(5): 557–562.
- Franke WW (2009) Discovering the molecular components of intercellular junctions – a historical view. *Cold Spring Harbor Perspectives in Biology* 1(3): a003061.
- Garrod D and Chidgey M (2008) Desmosome structure, composition and function. *Biochimica et Biophysica Acta* 1778(3): 572–587.
- Halbleib JM and Nelson WJ (2006) Cadherins in development: Cell adhesion, sorting, and tissue morphogenesis. *Genes and Development* 20(23): 3199–3214.
- Huber O (2003) Structure and function of desmosomal proteins and their role in development and disease. *Cellular and Molecular Life Sciences* 60(9): 1872–1890.
- Kitajima Y (2002) Mechanisms of desmosome assembly and disassembly. *Clinical Experimental Dermatology* 27(8): 684–690.
- Patel SD, Chen CP, Bahna F, et al. (2003) Cadherin-mediated cell–cell adhesion: Sticking together as a family. *Current Opinion Structural Biology* 13(6): 690–698.
- Shapiro L and Weis WI (2009) Structure and biochemistry of cadherins and catenins. *Cold Spring Harbor Perspectives in Biology* 1(3): a003053.

Cadherin Signaling

D B Sacks, National Institutes of Health, Bethesda, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Adherens junction Cadherin-dependent adhesive structures, linked to the actin cytoskeleton, that mediate attachment of cells to one another.

Cadherins Family of homophilic adhesion molecules that mediate Ca^{2+} -dependent cell–cell adhesion.

Catenins Family of proteins comprising α -, β -, γ -, and p120 catenin, that attach to the cytoplasmic tails of cadherins.

Homophilic The preference of a molecule for interactions with identical molecules.

Rho family GTPases Family of regulatory proteins that act as molecular switches, alternating between active and inactive forms.

The Cadherin Family

The cadherin superfamily comprises several subfamilies, including classic cadherins, atypical cadherins, desmocollins, desmogleins, protocadherins, Fat, and Flamingo cadherins.

Classic Cadherins

Classic cadherins, which comprise 26 members in humans, like E-, N-, and P-cadherin, are components of adherens junctions between cells. The most highly characterized is E-cadherin, which is predominantly expressed in epithelium. The classic cadherins are single-span transmembrane proteins with a highly conserved carboxy-terminal cytoplasmic domain and five extracellular domains, termed cadherin repeats. Ca^{2+} binding to the extracellular domain induces homophilic adhesion with cadherin molecules on adjacent cells. The cytoplasmic tail of cadherin associates with the actin cytoskeleton via proteins called catenins. Cadherin binds β -catenin or γ -catenin (also termed plakoglobin), which in turn bind to α -catenin (Figure 1). The interaction of α -catenin with the actin cytoskeleton is thought to strengthen the intercellular adherens junction. Another catenin, termed p120^{cas}, also binds to the cadherin tail in the cytoplasm. The closely related type II cadherins (e.g., VE-cadherin) have similar properties.

Desmosomal Cadherins

Desmocollins and desmogleins are components of desmosomes, which are sites of cell–cell adhesion in tissues subjected to mechanical strain (e.g., epidermis and the myocardium). The overall structure of these desmosomal cadherins resembles that of the classical cadherins, but their cytoplasmic tails are linked to the intermediate filaments of the cytoskeleton.

Protocadherins

The protocadherin family members have as many as seven extracellular Ca^{2+} -binding domains, a transmembrane region, and divergent cytoplasmic domains. They appear to participate in development and may have a particularly important role in the central nervous system.

Other Cadherins

Other cadherins range from those with seven transmembrane segments (such as Flamingo) to T-cadherin, which lacks both transmembrane and cytoplasmic regions.

Functions of Cadherin Signaling

Originally described exclusively as cell adhesion proteins, cadherins have been shown to influence multiple aspects of cell behavior. Indeed, cadherins can transmit signals across the cellular membrane into the interior to modulate cell function. While many members of the cadherin family are likely to share this capacity, to date only members of the classic cadherin family have been studied in detail and they are the focus of this article.

Signaling Triggered by Cell–Cell Contact

Signaling can be triggered when cadherins attach to other cadherins on adjacent cells, allowing information about the external environment to be conveyed into the cell. For example, most normal cells grown in culture dishes will reproduce until they form a single continuous layer. A process known as ‘contact inhibition’ then prevents further cell division. Cadherin–cadherin junctions between cells produce a signal that leads to contact inhibition and renders the cells insensitive to the stimulatory effects of growth factors. Signals generated by interactions between cadherins are also implicated in the formation of desmosomal cell junctions, nerve outgrowth, establishment of cell polarity, cytoskeletal organization, tissue formation, cell differentiation, cell movement, and regulation of gene expression.

Influence of Cadherin Level on Signaling

The amount of cadherin present on the cell surface may influence the signals generated. For example, expression of high levels of E-cadherin can inhibit cell motility and cell division and induce cell polarity. These effects have led to the description of E-cadherin as a tumor suppressor. In fact, loss of functional E-cadherin is often associated with epithelial tumor

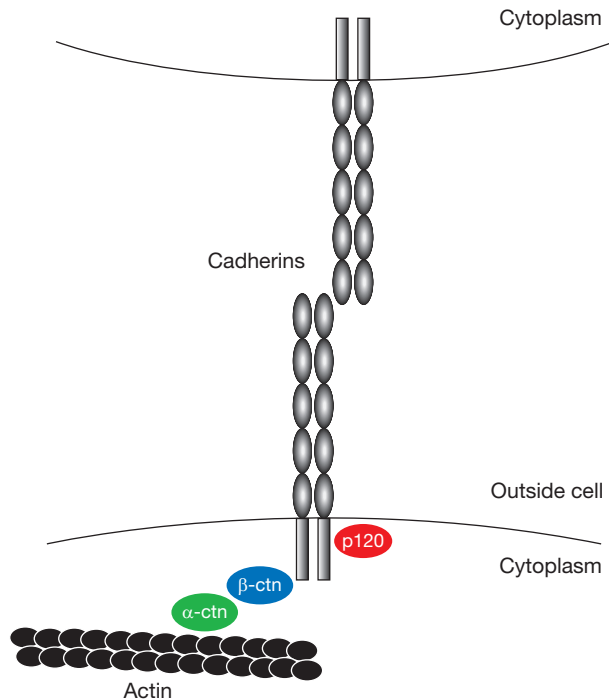


Figure 1 Schematic model of a cadherin-mediated adherens junction. Cadherins on adjacent cells interact with one another to provide intercellular adhesion. The terminal portion of the cytoplasmic tail of cadherin binds to β -catenin (β -ctn) or γ -catenin (not shown), which in turn binds to α -catenin (α -ctn). The interaction of α -catenin with the actin cytoskeleton stabilizes the adherens junction. Another catenin, p120^{ctn} (p120), stabilizes cadherin–catenin complexes by interacting with the cytoplasmic portion of the cadherin tail near the cell membrane.

progression and invasion. Reduced E-cadherin expression is mediated by mutation, aberrant internalization, or transcriptional regulation. Regulation of E-cadherin gene transcription alters the total amount of E-cadherin protein in cells during morphogenic processes such as hair follicle development. The level of E-cadherin on the cell surface also can be specifically regulated. These changes are likely to modulate not only the adhesive properties of the cell, but also the signals that the cell receives through contacts with other cells. In some cases, alterations in the level of cadherin may regulate cellular function even in the absence of cadherin–cadherin interactions between cells.

Different Signaling Properties of Various Cadherins

It is notable that the different types of cadherins induce different changes in cell behavior. For example, while E-cadherin usually reduces the motility of a cell, the presence of N-cadherin or cadherin-11 has the opposite effect, enhancing the ability of cells to crawl over surfaces and invade tissue. It is clear that the various cadherins trigger different signaling pathways. These differences are likely to play a crucial role in the regulated changes in cell behavior that occur during early animal development (embryogenesis). A striking example of this sort of change occurs when cells undergo epithelial–mesenchymal transitions. This process occurs during tissue remodeling when relatively sedentary epithelial cells with stable adherens

junctions transform into more motile and invasive mesenchymal cells in order to migrate and form new tissue structures. These events are frequently accompanied by a switch in cadherin expression, for example, from E-cadherin to N-cadherin. Similar changes occur during malignant transformation. Most human cancers arise from epithelial cells. These cells normally express E-cadherin. In many cases, expression of functional E-cadherin is lost during malignant transformation, and anomalous expression of N-cadherin or cadherin-11 occurs. These changes may contribute to the increased growth rate, and increased migratory and invasive capacity of cancer cells.

Inside-Out Cadherin Signaling

It is also apparent that the adhesiveness of cell-surface cadherins can be modulated from within the cell in the absence of changes in the level of cadherin expression. Therefore, cadherins are able to transduce both outside-in and inside-out signals. Feedback between these two types of signals is crucial to allow the correct formation and disassembly of cadherin–cadherin junctions during morphogenesis and for the continuous remodeling of adherens junctions that occurs even in sedentary cells.

Mechanisms of Cadherin Signaling

Considerable progress has been made in describing cadherin signaling. Classic cadherins interact with several signaling molecules, but the molecular mechanisms that constitute the signaling pathways are not completely understood. Three major signaling pathways have been described.

Rho Family GTPase Signaling

An important target of cadherin signaling is the cytoskeleton. This dynamic scaffold of rod-like actin must be rearranged to allow cells to form junctions with other cells and for cells to move. Members of the Rho family of small guanosine triphosphatases (GTPases) are vital regulators of these events and they are activated in response to the formation of cadherin-mediated cell junctions. Changes in the activity of the Rho family GTPases also lead to alterations in the adhesive capacity of cadherins. Thus, communication between cadherin and Rho GTPases is bidirectional. In general, GTPases serve as switches that are on when bound to GTP and off when the bound GTP is hydrolyzed to guanosine diphosphate (GDP). The best-characterized members of the Rho family are Cdc42, Rac1, and RhoA.

Rac1 is activated rapidly in response to cadherin binding (Figure 2). Active Rac1 stimulates remodeling of the actin cytoskeleton that leads to the formation of cell membrane protrusions known as ‘lamellipodia’. This change in morphology may help extend the contact zone between cells, enhancing formation of stable adherens junctions.

Cdc42 induces the formation of filopodia or microspikes, small finger-like projections from the cell, that are needed for the first steps of adherens junction formation by cadherins. Cdc42 is activated upon the assembly of E-cadherin-mediated cell junctions (Figure 2). This mechanism perhaps allows

E-cadherin to initiate signals that in turn further stimulate adherens junction formation. Cdc42 also activates the so-called partitioning defective (PAR)/atypical protein-kinase (aPKC) complex and thereby stimulates cell polarity and tight junction formation.

Activated Cdc42 or Rac1 can also influence the function of a protein known as 'IQGAP1'. High levels of IQGAP1 in a cell decrease E-cadherin-mediated adhesion, perhaps by binding β -catenin, thus displacing α -catenin from the E-cadherin- β -catenin complex. Activated Cdc42 or Rac1 can bind to IQGAP1 and counteract this activity, thereby increasing the adhesive capacity of cadherins.

Activation of RhoA leads to the formation of stress fibers. Active RhoA influences E-cadherin function and appears necessary to promote the formation of contractile actin bundles for maturation of adherens junctions. Moreover, cadherin signals can modulate the activity of RhoA. For example, RhoA activity is decreased as cells become confluent. Thus, cadherin can regulate the actin cytoskeleton by multiple mechanisms.

The transition from the off to the on state of GTPases can be brought about by guanine nucleotide exchange factors (GEFs) that cause the displacement of GDP from the GTPase, and replacement with GTP. Conversely, GTPase activating proteins (GAPs) enhance the conversion of GTP to GDP, so inactivating the GTPase. It is likely that cadherin-mediated adhesion leads to the activation of GEFs, or perhaps the inactivation of GAPs, for Cdc42 and Rac1. RhoA may be regulated by direct interactions with p120^{cas}. When not bound to cadherin, p120^{cas} inhibits activation of RhoA.

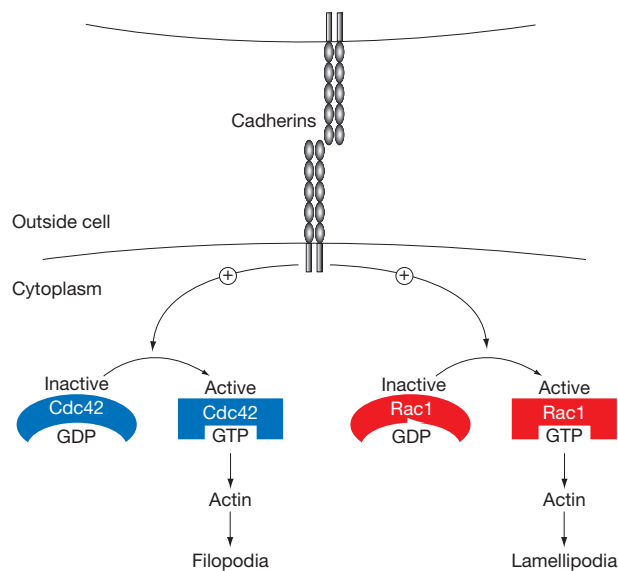


Figure 2 Cadherin-mediated adhesion activates Rho family GTPases. The formation of cadherin contacts between cells leads to the activation of Cdc42 and Rac1 (i.e., generation of GTP-bound forms). In the active state, Cdc42 and Rac1 cause alterations in the actin cytoskeleton that lead to the generation of finger-like filopodia and membrane ruffles known as 'lamellipodia', respectively. These events lead to changes in both the ability of cells to adhere to their surroundings and cell migration, and can increase the stability of the cadherin-mediated junctions themselves.

Wnt/ β -Catenin Signaling

In addition to linking the cytoplasmic tails of cadherins to the actin cytoskeleton (see Figure 1), β -catenin is an important component of a signaling pathway involved in both normal development and cancer. Normally, β -catenin molecules not attached to cadherins are degraded in the cytoplasm. When cells are exposed to an extracellular growth factor known as Wnt, β -catenin is protected from degradation and accumulates in the nucleus (Figure 3). There, in association with the T-cell factor/lymphoid enhancer factor (TCF/LEF) family of DNA-binding proteins, β -catenin activates the transcription of genes that stimulate cell proliferation. Thus, the extent to which β -catenin is soaked up by binding to cadherins can potentially modulate the signal generated by activation of the Wnt pathway. High expression levels of E-cadherin can substantially reduce the amount of free β -catenin and lower the cellular response to Wnt. In the absence of Wnt, increased adherens junction formation recruits β -catenin, decreasing the nuclear pool of β -catenin and attenuating β -catenin-mediated transcription. Conversely, disruption of cell-cell adhesion may

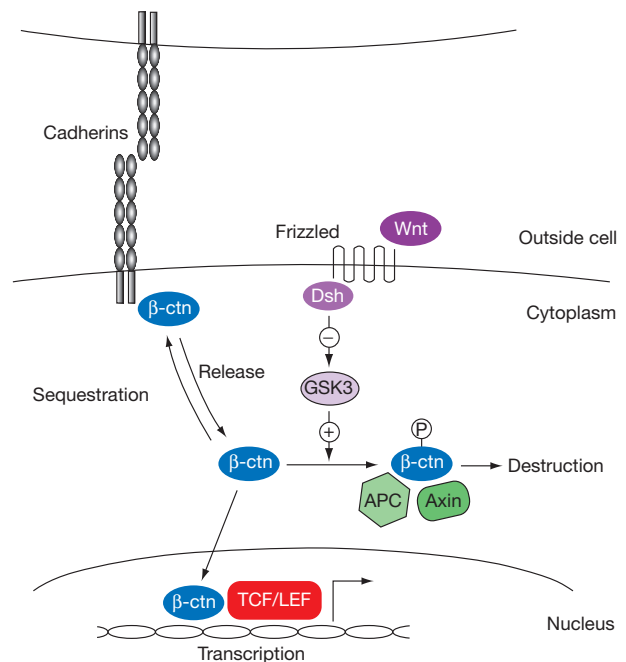


Figure 3 Cadherins modulate Wnt/ β -catenin signaling. In addition to its role as a component of adherens junctions, β -catenin (β -ctn) increases the activity of the TCF/LEF transcription factor in the nucleus. The extent to which β -catenin accumulates in the nucleus depends in part upon the level of free β -catenin in the cytoplasm. The amount of free β -catenin is normally low due to its phosphorylation by glycogen synthase kinase-3 (GSK3), binding to adenomatous polyposis coli (APC) and Axin proteins, and subsequent destruction (P indicates the added phosphate group). GSK3 activity is decreased by Dishevelled (Dsh) when the growth factor Wnt binds to its receptor frizzled on the cell surface. Therefore, Wnt can reduce degradation of β -catenin and increase the amount of β -catenin in the nucleus, thereby promoting transcription. By contrast, the formation of cadherin-cadherin junctions, or an increase in the amount of cadherin, can lead to the sequestration of β -catenin due to its interaction with cadherin tails. In this way, cadherins are thought to decrease the amount of nuclear β -catenin and act in opposition to Wnt signals.

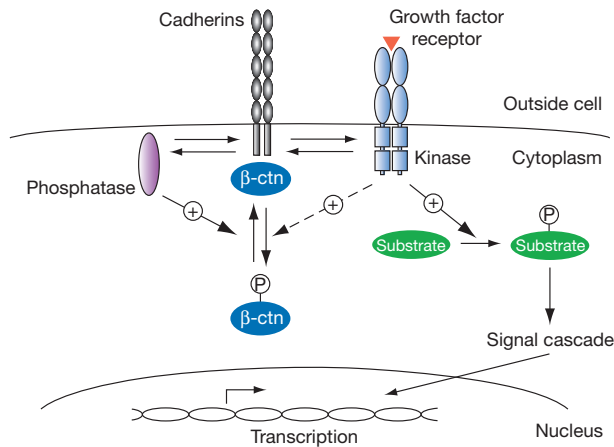


Figure 4 Cadherin cross-talk with kinases and phosphatases. Cadherins interact with several growth factor receptors that have intrinsic kinase activity. By this mechanism, cadherins can modulate the signaling cascades triggered by growth factors that lead, for example, to the activation of gene transcription. In addition, activation of receptor kinases can promote the phosphorylation on tyrosine residues of β -catenin (dotted line; P indicates the added phosphate group), its dissociation from cadherin tails, and disassembly of the adherens junction (not shown). Cadherins are also found in a complex with a number of protein phosphatases that can dephosphorylate β -catenin. In this way, cadherins can modulate the reversible phosphorylation of adherens junction components that alter the activity of cadherin itself.

release β -catenin, thereby enhancing transcription. It is noteworthy that mutations of proteins that regulate turnover of β -catenin have been found in several human cancers.

Phosphorylation and Dephosphorylation

The phosphorylation of proteins (i.e., addition of phosphate groups) by kinase enzymes is a common mechanism for regulating numerous cellular processes. Although the cadherins themselves do not have kinase activity, several examples of cross-talk have been observed between cadherins and growth factor receptors that contain intrinsic tyrosine kinases (Figure 4). For example, N-cadherin interacts with the fibroblast growth factor receptor FGFR1 and can boost the phosphorylation cascade triggered by binding of fibroblast growth factor-2 to FGFR1. Analogous interactions may occur between E-cadherin and the epidermal growth factor receptor, as well as VE-cadherin and the vascular endothelial growth factor receptor-2. In this way, cadherins can modulate changes in the survival, proliferation, and motility of cells in response to a variety of growth factors.

Conversely, phosphorylation can modulate inside-out signals to regulate cadherin activity. The association between growth factor receptor tyrosine kinases and cadherins allows growth factors to promote the phosphorylation of several adherens junction proteins (**Figure 4**). A number of non-receptor kinases, including Src, also can phosphorylate junction components. Phosphorylation of β -catenin on tyrosine residues is believed to induce dissociation of β -catenin from cadherins and decrease adherens junction stability. Phosphorylation of the cadherins themselves, and of γ -catenin and p120^{cas}, also modulates adherens junction assembly. In particular, the interaction of p120^{cas} with the cytoplasmic domain of cadherins appears to be an important regulator of cadherin activity, perhaps by attenuating internalization (by endocytosis) or destruction of cadherins. Cadherins are also found in a complex with selected protein phosphatases. These enzymes, which are activated by cadherins, dephosphorylate (i.e., remove phosphate from) β -catenin, allowing reversible regulation of cadherin adhesive capacity (**Figure 4**).

See also: **Cell Architecture and Function:** Cadherin-Mediated Cell-Cell Adhesion; Rho GTPases and Actin Cytoskeleton Dynamics.

Further Reading

- Ben-Ze'ev A, Shutman M, and Zhurinsky J (2000) The integration of cell adhesion with gene expression: The role of beta-catenin. *Experimental Cell Research* 261: 75–82.
- Bracke ME, Van Roy FM, and Mareel MM (1996) The E-cadherin/catenin complex in invasion and metastasis. *Current Topics in Microbiology and Immunology* 213: 123–161.
- Braga VM and Yap AS (2005) The challenges of abundance: Epithelial junctions and small GTPase signalling. *Current Opinion in Cell Biology* 17: 466–474.
- Christofori G and Semb H (1999) The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. *Trends in Biochemical Sciences* 24: 73–76.
- Clevers H (2006) Wnt/beta-catenin signaling in development and disease. *Cell* 127: 469–480.
- Gumbiner BM (2005) Regulation of cadherin-mediated adhesion in morphogenesis. *Nature Reviews Molecular Cell Biology* 6: 622–634.
- Harris TJ and Tepass U (2010) Adherens junctions: From molecules to morphogenesis. *Nature Reviews Molecular Cell Biology* 11: 502–514.
- Hartsock A and Nelson WJ (2008) Adherens and tight junctions: Structure, function and connections to the actin cytoskeleton. *Biochimica et Biophysica Acta* 1778: 660–669.
- Mosesson Y, Mills GB, and Yarden Y (2008) Derailed endocytosis: An emerging feature of cancer. *Nature Reviews Cancer* 8: 835–850.
- Nelson WJ and Nusse R (2004) Convergence of Wnt, beta-catenin, and cadherin pathways. *Science* 303: 1483–1487.
- Perez-Moreno M, Jamora C, and Fuchs E (2003) Sticky business: Orchestrating cellular signals at adherens junctions. *Cell* 112: 535–548.
- Polakis P (2000) Wnt signaling and cancer. *Genes and Development* 14: 1837–1851.

Calcitonin Gene-Related Peptide and Adrenomedullin Receptors

D L Hay, Auckland University, Auckland, New Zealand

A C Conner, Warwick University, Warwick, UK

D R Poyner, Aston University, Birmingham, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Agonist A drug that activates a receptor.

Antagonist A drug that blocks the action of an agonist at a receptor.

Glycosylation The attachment of sugars to a protein.

pA₂ The negative logarithm of the antagonist concentration that causes a twofold shift in the agonist dose–response curve. It is a measure of antagonist affinity.

Calcitonin gene-related peptide (CGRP) and adrenomedullin are abundant, related peptides that share many biological actions. Most prominently, both are very potent vasodilators. CGRP is a neurotransmitter in sensory neurons and adrenomedullin is a hormone that is locally released from tissues including vascular endothelium and smooth muscle. They form a peptide family whose major members are CGRP, calcitonin, adrenomedullin 2/intermedin, and amylin. For some time, there was considerable controversy about the nature of the receptors for this family. It has now been established that they define a new paradigm among G-protein-coupled receptors (GPCRs), in that they are heteromers of a seven-transmembrane-spanning protein that resembles a classical GPCR, the calcitonin receptor (CTR)-like receptor (CLR) and a member of the single-transmembrane-spanning receptor activity-modifying protein (RAMP) family. The CGRP receptor comprises CLR/RAMP1, whereas there are two adrenomedullin receptors: CLR/RAMP2 (AM₁ receptor) and CLR/RAMP3 (AM₂ receptor). RAMPs also interact with other GPCRs.

The Pharmacology of CGRP and Adrenomedullin

The Structure of CGRP and Adrenomedullin

The amino acid sequence identity of CGRP and adrenomedullin (Figure 1) is low but the secondary structure of these peptides is conserved. For both peptides, an amino-terminal ring is important for receptor activation; peptides that lack this structure (CGRP_{8–37} and adrenomedullin_{22–52}) are antagonists.

The Pharmacology of Endogenous CGRP and Adrenomedullin Receptors

CGRP receptor-selective drugs and their pharmacological properties

For CGRP, the most widely used antagonist has been CGRP_{8–37}. This has only limited selectivity. A number of high-affinity nonpeptide antagonists have been developed such as BIBN4096BS and MK0974 (Figure 2), which show marked selectivity for primate over rodent CGRP receptors and have very little affinity for adrenomedullin or other related receptors. It was formerly suggested that CGRP receptors should

be divided into two subtypes: CGRP₁ receptors with a high affinity for CGRP_{8–37} (pA₂ > 7) and CGRP₂ receptors with a lower affinity for this antagonist. It is now clear that only the CGRP₁ receptor is a bone-fide receptor for CGRP and this is now simply called the 'CGRP receptor'. The CGRP₂ receptor is now considered to be a mixture of adrenomedullin (AM₂) and amylin receptors.

The development of nonpeptide CGRP antagonists has been driven by the desire to develop new therapies for migraine. There is good evidence that many of the symptoms of migraine are caused by the release of CGRP by the trigeminal nerve. Both BIBN4096BS (Olcgepant) and MK0974 (Telcagepant) have been evaluated in clinical trials. BIBN4096BS is well tolerated but its use is likely to be limited as it is not orally bioavailable. In phase III clinical trials, MK0974 showed comparable efficacy to the triptan antimigraine drug zolmitriptan but with fewer cardiovascular side effects. However, there is evidence that chronic treatment with MK0974 can cause elevations in serum transaminases, placing some limitations on its use. Nonetheless, the availability of these drugs may open the way to both treatments for migraine and also their evaluation in other conditions where CGRP may play a role, such as hot flushes in postmenopausal women.

Adrenomedullin receptor-selective drugs and their pharmacological properties

The only adrenomedullin antagonist that has been widely used is adrenomedullin_{22–52}. This is not particularly potent (pA₂ ~ 7) and does not antagonize all adrenomedullin-mediated responses. It shows a slight selectivity for AM₁ receptors over the AM₂ receptors but not of a sufficient magnitude to make this a useful agent for defining pharmacology.

Drugs that act against both CGRP and adrenomedullin receptors

Agents that interact solely with CLR, which is common to all the receptors, are likely to block the effects of both peptides. Experimental compounds have been described with these properties; for example, a compound prepared by Merck and also probably SB (+) 273779 (Figure 2). These bind to the receptors with micromolar affinities, around 1000-fold less tightly than the nonpeptide CGRP antagonists.



Figure 1 A comparison of the structure of human CGRP and adrenomedullin_{15–52}. Note the disulfide bond in both peptides, indicated by a bar over the sequences. The first 14 amino acids of adrenomedullin are not required for biological activity and are not shown.

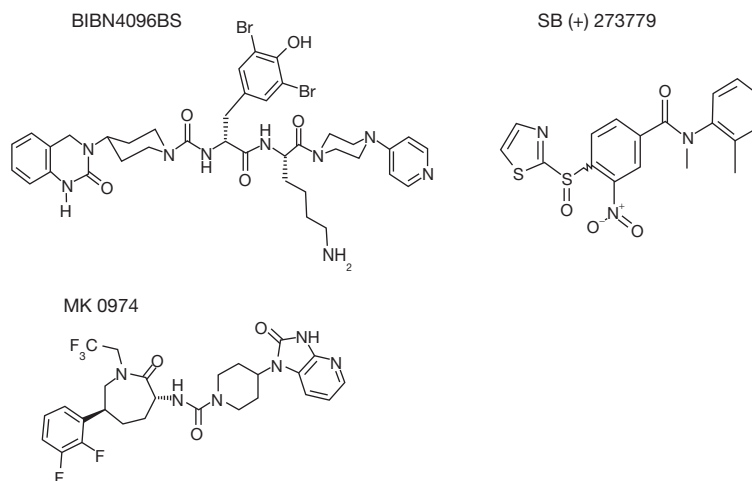


Figure 2 Structures of BIBN4096BS, MK0974, and SB (+) 273779.

The Molecular Structure of CGRP and Adrenomedullin Receptors

CLR and RAMPs

CGRP and adrenomedullin receptors display a very unusual molecular architecture. Both are composed of a common seven-transmembrane-spanning protein called CLR (CL receptor, often shortened to CLR, and previously known as CRLR; **Figure 3**). This is a member of the B-family of GPCRs (related to other receptors, such as those for secretin and calcitonin). Human CLR is 461 amino acids long and has 55.5% sequence identity with the human CTR. By itself, CLR does not function as a receptor for any peptide. However, when associated in a complex with a RAMP, it functions as a high-affinity receptor for either CGRP or adrenomedullin. The RAMPs are a family of three proteins (**Figure 4**), each with an extracellular amino terminus of approximately 100 amino acids, a single transmembrane section, and a very short carboxyl terminus of approximately 10 amino acids. Crystal structures are available for the extracellular domain of RAMP1 (protein database structure 2YX8) and CLRI RAMP1 (protein database structure 3N75).

CLR/RAMP1; the CGRP Receptor

CLR–RAMP1 association is intracellular and this facilitates their transport to the cell surface. There, the complex functions as a CGRP receptor, with a high affinity for CGRP, CGRP_{8–37}, and nonpeptide antagonists such as BIBN4096BS and

MK0974. The complex binds adrenomedullin with an approximately 10-fold lower affinity than CGRP; adrenomedullin 2/intermedin is roughly equipotent with adrenomedullin at this receptor.

Little is known about the mechanism of ligand binding and receptor activation. CGRP can cross-link to both CLR and RAMP1, showing that both components create the peptide-binding site. However, it is not clear whether CGRP has specific contacts with both proteins or whether RAMP1 indirectly contributes to the ligand-binding site by modifying the structure of CLR. It is likely that CGRP makes contacts with both the extracellular domain of CLR/RAMP1 and the membrane–extracellular loop interfaces of CLR. The selectivity of the nonpeptide antagonists is due to a tryptophan residue found at position 74 in RAMP1; this is responsible for their selectivity at primate over rodent CGRP receptors and contributes to their selectivity over adrenomedullin receptors.

CLR/RAMP2 and CLR/RAMP3: The Adrenomedullin Receptors

The heteromers formed by RAMP2 and RAMP3 with CLR produce adrenomedullin receptors. Their formation is very similar to that of the CLR/RAMP1 complex; in the absence of CLR, they are not expressed at the cell surface. It was initially thought that RAMP2 might work by modulating the glycosylation state of CLR. However, it is now thought that adrenomedullin and CGRP both bind to the fully glycosylated form of CLR. As with the CLR/RAMP1 complex, there is little detailed information

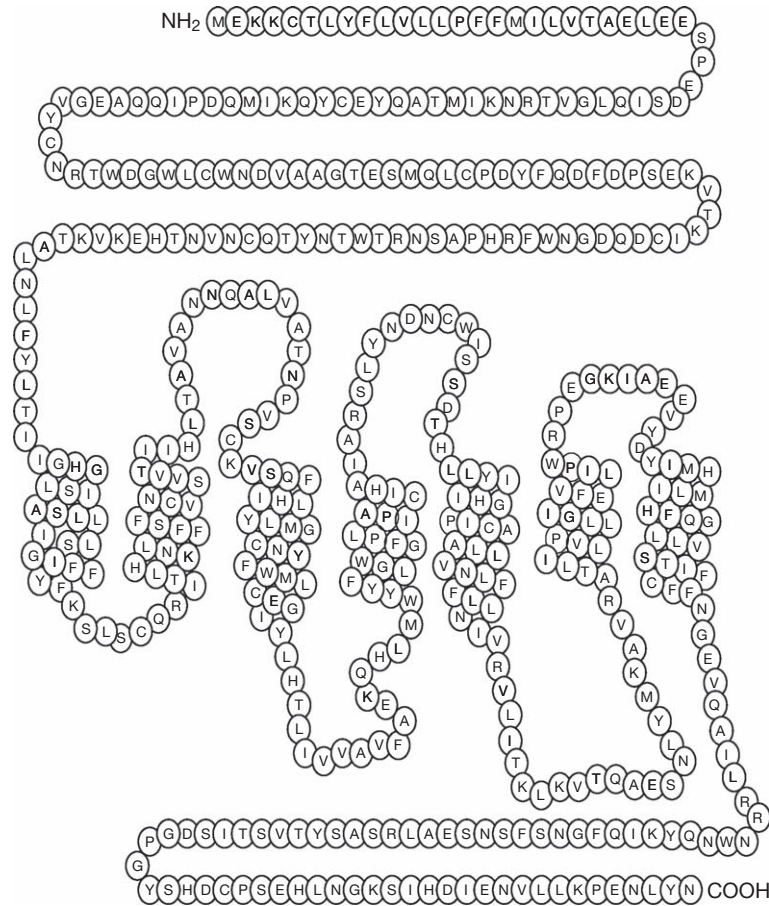
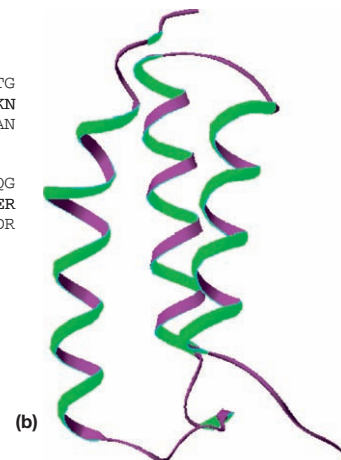


Figure 3 The structure of human CLR, showing the postulated seven transmembrane helices. The residues that are likely glycosylation sites on the amino terminus are underlined.

RAMP3	-----METGALRRPQLPLLLLLCG-----GCPRAGCCNETG
RAMP2	<u>MA</u> SLRVERAGGPRLPRTVRGRPAAVRLLLLLGAVLNPHLEALQPLPTTGTGPGSEGGTVKN
RAMP1	-----MARALCRLPRRGLWLLLAHH-----LFMTTACQEAN
	* ***
RAMP3	MLERL-PLCGKAFADMMGKVDVWKWCNLSFIVVYESFTNCTEMEANVVGCVWPNPLAQG
RAMP2	YETAV-QFCWNHYKDQMDPIEK-DWCWAMISRPYSTLRDCLHFAELFDLGFNPPLAER
RAMP1	YGALLRELCLTQFQVDEAVGETLWCDWGRITRSYRELADCTWHMAEKLGCFWPNAEVDR
	* * ** * * * **
RAMP3	FITGIHRQFFSNCTVDRVHLEDPPDEVLIPLIVIPVVLTVAMAGLVVWRSKRTDTLL
RAMP2	IIFETHQIHFA NC SLVQPTFSDDPEDVLLAMIAPICLIPFLITLVVWRSKDSEAQA
RAMP1	FFLAVHGRYFRSCPISGRAVRDPPGSILYPFIVVPITVTLLVTALVVWQSKRTEGIV
	* * * *** * * * **** **

(a)

TM domain



(b)

Figure 4 (a) Alignment of human RAMPs 1, 2, and 3. *Single, fully conserved residue. Underlined italic residues at the start of the sequence represent signal peptides. Underlined boldface residues are potential glycosylation sites. (b) Structure of the extracellular domain of RAMP1.

on mechanisms of adrenomedullin binding to either the CLR/RAMP2 or the CLR/RAMP3 complexes, although glutamic acid at position 74 in RAMP3 appears to be important. Both complexes exhibit higher affinity for adrenomedullin than CGRP. The CLR/RAMP2 complex is designated as the AM₁ receptor and the CLR/

RAMP3 complex as the AM₂ receptor. The AM₁ receptor has high affinities for adrenomedullin and adrenomedullin₂₂₋₅₂ but low affinity for CGRP. The AM₂ receptor has high affinity for adrenomedullin and a modest affinity for CGRP. Adrenomedullin 2/intermedin can stimulate both AM₁ and AM₂ receptors.

Cellular Signaling and Receptor Component Protein

The CLR/RAMP complexes are all coupled to the stimulation of adenylate cyclase via the G protein known as G_s . However, as is the case with many related GPCRs, coupling to G_i/G_o and the G_q family of G proteins (leading to changes in ion channel activity and elevation of intracellular calcium) have also been reported. The receptors also associate with β -arrestin, with the potential for activation of further downstream signaling pathways.

A novel protein called receptor component protein (RCP) is reportedly essential for coupling the CGRP and at least the AM_1 receptors to stimulation of G_s . Immunoprecipitation studies suggest that RCP is physically associated with CLR. It has been reported that inhibiting RCP expression disrupts signaling mediated via CGRP and AM_1 receptors. RCP is very widely distributed in cell lines but a broader role has yet to be demonstrated.

Other Receptors that Respond to CGRP and Adrenomedullin

As noted above, amylin receptors can also respond to CGRP. Amylin receptors are complexes between the CTR and RAMPs. The AMY_1 receptor (CTR/RAMP1) has the highest affinity for CGRP, followed by the AMY_3 receptor (CTR/RAMP3). The AMY_2 receptor (CTR/RAMP2) has only a weak affinity for CGRP.

The Distribution and Physiology of CGRP and Adrenomedullin Receptors

The Distribution of Binding Sites for CGRP and Adrenomedullin

CGRP-binding sites are present in both the central nervous system (CNS) and peripheral tissues. In some of the earlier literature, there was confusion due to cross-labeling of what would now be considered amylin receptors. Specific CGRP receptors show the potency order: CGRP > adrenomedullin > amylin. In the rat CNS, the highest densities of CGRP binding are in the nucleus accumbens, caudate putamen, amygdaloid body, pontine nuclei, cerebellum, spinal cord, and inferior olive. In the periphery, CGRP receptors are particularly associated with the cardiovascular system, on blood vessels and in the atria. There are high densities of these receptors in the spleen. However, receptors are also found on nonvascular cells, for example, in the vas deferens and secretory cells in the gastrointestinal tract. In tissues, adrenomedullin receptors show a high level of specificity for adrenomedullin over CGRP or other peptides. Binding in the CNS is highest in the spinal cord, but moderate to low levels are seen in many other brain regions. In the periphery, adrenomedullin binding is closely associated with the cardiovascular system, with very high levels in the heart and lungs.

Distribution of RAMPs and CLR

The most reliable information on the distribution of CLR and RAMPs comes from messenger RNA (mRNA) measurements; antibodies to these proteins are often of uncertain specificity. Generally, there is fair agreement between the distribution of RAMPs and CLR and binding sites for CGRP and adrenomedullin. Often, RAMP2 appears to be more abundant than

RAMP1, with RAMP3 being expressed at the lowest levels. However, there are many exceptions; RAMP1 is the most abundant transcript in the brain and the pancreas and RAMP3 predominates in the liver. There are some anomalies; the nucleus accumbens has strong CGRP binding and RAMP1 expression but little CLR (the binding sites are probably amylin receptors). By contrast, the frontal cortex has very little binding for CGRP, adrenomedullin, or amylin, but has high levels of RAMP expression. RAMPs can associate with other receptors in addition to those for calcitonin and CLR, and so it is not surprising that there should be some variable correlation between the expression of these components.

Changes in Expression of CGRP and Adrenomedullin Receptors

There is good evidence for changes in CGRP and adrenomedullin receptors during various (patho)physiological processes. For example, adjuvant-induced arthritis in rats leads to a decrease in CGRP binding in the spinal cord. Changes in CLR and RAMP expression have also been noted in disease. In a mouse model of sepsis, CLR mRNA and RAMP2 mRNA show large decreases but RAMP3 expression increases in the lung. In a rat model of ischemic heart failure, RAMP2 expression is increased in the ventricles.

The Pathophysiology of CGRP and Adrenomedullin Receptors

Changes in receptor expression suggest that both CGRP and adrenomedullin receptors are of importance during normal physiological processes and in disease. Knockout models for CLR, RAMP2, and adrenomedullin produce similar effects, where homozygotes die *in utero*, showing gross defects in their lymphatic and cardiovascular systems. Heterozygote knockout mice are viable but have reduced fertility; in some, but not all models, there is an increase in blood pressure. Overexpression of RAMP2, while enhancing sensitivity to the vasodilatory effects of adrenomedullin and protecting against vascular hypertension, does not alter resting blood pressure. Thus, adrenomedullin appears to be vital for the development and operation of the cardiovascular and lymphatic systems and failures can lead to infertility but it may not be important for normal control of blood pressure. Overexpression of CLR causes an increase in intraocular pressure, an effect due to enhanced expression of AM_1 receptors in the papillary sphincter muscle. The overall phenotype resembled that of acute angle glaucoma in man.

RAMP3 knockout animals have very different phenotypes to RAMP2 mice; the homozygotes survive to adulthood but gain weight less than controls. Thus, AM_1 and AM_2 receptors have different functions.

Studies on α CGRP-deficient mice have mixed effects, although some workers have noted increased blood pressure, decreased pain sensitivity, and reduced hypersensitivity to antigens in airways. Homozygous RAMP1 knockout animals have elevated blood pressure and an enhanced inflammatory response to lipopolysaccharide. Overexpression of RAMP1 in the trigeminal nerve increases sensitivity to CGRP, as manifested by enhanced plasma extravasation and release of substance P. The data are consistent with roles for CGRP in the

cardiovascular system and also as an inflammatory mediator (with both pro- and anti-inflammatory actions, depending on the cellular context), involved in pain perception and the activation of cells of the immune system.

In summary, the evidence from genetic manipulation of CGRP and adrenomedullin receptor components is that this changes receptor sensitivity and can give phenotypes consistent with the involvement of these receptors in human disease. Further, changes in the receptors are observed in diseases and disease models. Thus, there are grounds for thinking that stimulation or inhibition of CGRP and adrenomedullin receptors may be important in both chronic and acute conditions involving, for example, angiogenesis, lymphangiogenesis, and inflammation.

See also: [Signaling: Calcitonin Receptor; G-Protein-Coupled Receptor Kinases and Arrestins.](#)

Further Reading

Brain SD and Grant AD (2004) Vascular actions of calcitonin gene-related peptide and adrenomedullin. *Physiological Reviews* 84: 903–934.

- Conner AC, Hay DL, and Poyner DR (2009) CGRP receptor antagonists: Design and screening. *Expert Opinion in Drug Discovery* 4: 1253–1265.
- Dunworth WP and Caron KM (2009) G protein-coupled receptors as potential drug targets for lymphangiogenesis and lymphatic vascular diseases. *Arteriosclerosis, Thrombosis, and Vascular Biology* 29: 650–656.
- Hay DL and Dickerson IM (eds.) (2010) *The Calcitonin Gene-Related Peptide Family: Form, Function and Future Perspectives*. Dordrecht: Springer.
- Hay DL, Poyner DR, and Quirion R (2008) International union of pharmacology. LXIX. Status of the calcitonin gene-related peptide subtype 2 receptor. *Pharmacological Reviews* 60: 143–145.
- Kusano S, Kukimoto-Niino M, Akasaka R, et al. (2008) Crystal structure of the human receptor activity-modifying protein 1 extracellular domain. *Protein Science* 17: 1907–1914.
- Poyner DR, Sexton PM, Marshall I, et al. (2002) Foord, International Union of Pharmacology XXXII. The mammalian CGRP, adrenomedullin, amylin and calcitonin receptors. *Pharmacological Reviews* 52: 233–246.
- Sexton PM, Poyner DR, Simms J, Christopoulos A, and Hay DL (2009) Modulating receptor function through RAMPs: Can they represent drug targets in themselves? *Drug Discovery Today* 14: 413–419.
- ter Haar E, Koth CM, Abdul-Manan N, et al. (2010) Crystal structure of the ectodomain complex of the CGRP receptor, a class-B GPCR, reveals the site of drug antagonism. *Structure* 18: 1083–1093.
- Villalón CM and Olesen J (2009) The role of CGRP in the pathophysiology of migraine and efficacy of CGRP receptor antagonists as acute antimigraine drugs. *Pharmacology and Therapeutics* 124: 309–323.

Calcitonin Receptor

S I Girgis, Imperial College London, London, UK

N Moradi-Bidhendi, Barts and The London School of Medicine and Dentistry, London, UK

L Mancini, Pistoia, Italy

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Samia I. Girgis, Niloufar Moradi-Bidhendi, Lucia Mancini, and Iain MacIntyre, volume 1, pp. 217–220, © 2004, Elsevier Inc.

Glossary

Affinity The strength with which a chemical messenger binds to its receptor.

Calcitonin (CT) A small peptide hormone and a potent inhibitor of bone resorption.

Downregulation A decrease in receptor number resulting from prolonged exposure to high concentration of the message.

G proteins A family of plasma membrane regulatory proteins that many activated receptors interact with and alter.

Osteoclast A large multinucleated bone cell that can destroy bone.

Phospholipase C Enzyme that catalyzes the breakdown of a plasma membrane phospholipid known as

phosphatidyl-inositol bisphosphate (PIP₂) to inositol triphosphate (IP₃) and diacylglycerol (DAG).

Protein kinases Enzymes that phosphorylate other proteins by transferring to them a phosphate group from adenosine triphosphate. Introduction of a phosphate group alters the activity of the protein, often itself an enzyme.

Receptor modulation A molecular mechanism, whereby cells/tissues could change their responsiveness to different peptides.

Receptor-activity-modifying proteins (RAMPs) Single transmembrane-domain proteins that alter the phenotype of the calcitonin receptor, calcitonin receptor-like receptor and other G-protein-coupled receptors.

Second messengers Substances that serve as the relay from the plasma membrane to the biochemical machinery inside the cell.

Calcitonin Peptide Family

This article is dedicated to the memory of Professor Iain MacIntyre, who was the senior author of CT receptor chapter of the 1st edition of the *Encyclopedia of Biological Chemistry* in 2004. The calcitonin peptide family comprises seven known members (calcitonin (CT), amylin (AMY), two calcitonin gene-related peptides (CGRP α and CGRP β), adrenomedullin (AM), intermedin (IMD), and calcitonin receptor-stimulating peptide (CRSP)). These peptides display a weak homology at the primary structure but are similar in their secondary structure. They share a ring structure, formed by a disulfide bridge at or near the N-terminus, and have C-terminal amides, which are essential for full biological activity (Figure 1). They have some overlapping biological effects due to cross-reactivity at each other's receptors.

Calcitonin

Calcitonin (CT) is a small peptide hormone secreted mainly by special type of cells within the thyroid gland in mammals, called C cells. These cells originate from the neural crest and then migrate during embryonic development to be localized mainly in the thyroid gland in mammals. However, in submammalian vertebrates these cells remain as a separate gland, termed the ultimobranchial body. CT is a potent inhibitor of bone resorption, acting via receptor-mediated inhibition of osteoclast function. In addition to its effect on bone, CT has a wide range of biological activities, including analgesic effect, suppression

of gastric acid secretion, and appetite. On the kidney, CT enhances the excretion of sodium, calcium, and phosphate.

CT was first identified as a calcium-lowering factor in 1961. The calcium-lowering effect of CT is mainly due to direct inhibition of osteoclast activity, and, to a lesser extent, is due to increasing calcium excretion by the kidney. The potency of CTs from different species greatly varies, with submammalian CTs (e.g., salmon CT and eel CT) being the most potent. Because of its potent inhibitory effect on osteoclastic bone resorption, CT is used in the treatment of conditions associated with increased bone resorption such as Paget's disease of bone, osteoporosis, and hypercalcemia of malignancy. Medullary thyroid carcinoma is a tumor of the C cells of the thyroid gland that produce and secrete abnormally large amounts of CT. Measurement levels of CT in the blood of these patients is useful in the diagnosis and monitoring of this condition.

Calcitonin Gene-Related Peptides

Calcitonin gene-related peptides (CGRPs) are highly conserved neuropeptides among various species. There are two forms, CGRP α and CGRP β , which differ in only three amino acids in humans and in one amino acid in rats. Each peptide is the product of a separate gene. In human, both genes are localized on the short arm of chromosome 11. CGRP α is the product of the calcitonin gene, referred to as Calc I or CT/CGRP α gene. One feature of this gene is that alternative splicing of the primary RNA transcript leads to the differential of either CT mRNA in the thyroid C cells or alternatively to CGRP α messenger RNA (mRNA) in neural tissues. CGRP α is abundantly

Human CGRP α	ACDTATCVTHRLAGLLSRSGGVVKNNFVPTNVGSKAF-amide
Human CGRP β	ACNTATCVTHRLAGLLSRSGGMVKS NFVPTNVGSKAF-amide
Human AMY	KCNTATCATQRLANFLVHSSNNFGA ILSSTNVGSNTY-amide
Human ADM 1	YRQSMNNFQGLRSFGCRFGTCTVQKLAHQIYQFTDKDKDNVYAPRSK I SPQGY-amide
Human ADM 2	TQAQLLR SVGCVLGT CQVQNL SHRLWQLMGPAGRQDSAPVDPSSPHS Y-amide
Porcine CRSP-1	SCNTATCMTHRLVGLLSRSGSMVRSNLLPTKMGFKVF-amide
Human CT	CGNLSTCMLGTYTQDFNKFHTFPQTAIGVGAP-amide

Figure 1 The CT peptide family. The peptides share a six-/seven-amino-acid ring structure, formed by a disulphide bridge at or close to the N terminus. The peptides also have C-terminal amides, which are essential for full biological activity. Adapted from Girgis SI and MacIntyre I (2002), Calcitonin Gene-related Peptide (CGRP). In: Creghton TE (ed.) *Encyclopedia of Molecular Medicine*, 5 vols. New York, NY: Wiley, with permission from John Wiley and Sons.

and widely distributed neuropeptide throughout the central and peripheral nervous systems. It is present most abundantly in sensory neurons, where it is mostly co-localized with substance P (SP). It is found in motoneurons and neurons of the autonomic nervous system. It is also abundant in the posterior horn of the spinal cord suggesting a role in the transmission of sensory impulses. CGRP-containing nerve fibers are widely distributed to most organ system, and are particularly abundant around arteries. By contrast, CGRP β is localized primarily in the gut, in sites including the enteric nerves and the pituitary gland.

CGRP is one of the most potent vasodilator substances identified to date, with a potency ~ 10 -fold greater than prostaglandin and 100–1000 times greater than other classic vasodilators. Both forms of CGRP have similarly potent biological activity in terms of vasodilator activity.

The distribution of CGRP-containing nerves throughout the vascular system and its potent vasodilatory effects suggests an important role for CGRP in the modulation of vascular tone and in the regulation of regional blood flow. Further evidence of the physiological importance of CGRP α came from gene-knockout mice. Mice deficient in CGRP α demonstrated elevated peripheral vascular resistance and high blood pressure caused by increased peripheral sympathetic activity.

Amylin

AMY is a peptide hormone that is produced principally by β cells of the pancreas (cells that secrete insulin). It is co-localized with insulin within the same secretory granules, and co-secreted with insulin in response to a high blood glucose, for example, after a meal. AMY is also found in much smaller quantities in the gut. It opposes the metabolic actions of insulin in skeletal muscle. AMY is also a potent inhibitor of gastric emptying and of food intake. In the kidney, AMY enhances renin activity. It appeared to have a role in the development of the kidney. Although less potent than CT, AMY also inhibits the activity and formation of osteoclasts. AMY-deficient mice (by deletion of the gene encoding AMY) develop a low bone mass phenotype mimicking osteoporosis, which is solely due to an increase in bone resorption. AMY is now considered a crucial physiological inhibitor of bone resorption.

Adrenomedullin

This regulatory peptide was first isolated from human pheochromocytoma (a tumor of the adrenal medulla) in 1993. Unlike CGRP, AM is produced primarily by non-neural tissues especially vascular endothelial and smooth muscle cells. It is widely expressed mainly in areas where microvascular vessels exist (e.g., lungs, brain, kidneys, and bone). It is also found in the adrenal medullae and in a variety of cells including fibroblasts and macrophages.

As CGRP, AM has multiple biological effects including potent vasodilation. It acts directly on the renal, pulmonary, cerebral, mesenteric, and systemic circulation including the vascular supply of the skeleton. In the kidney, AM enhances the excretion of sodium. AM also acts as a local regulatory factor that can influence cell growth and development. Studies of gene-knockout mice revealed the physiological importance of AM. Deletion of the AM gene proved to be lethal beyond mid-gestation due to defective development of blood vessels. AM $\pm$ heterozygous mice showed high blood pressure and susceptibility to tissue injury. Thus, AM is essential for normal embryonic development, regulation of blood pressure, and tissue protection against injury.

Intermedin or AM2

This peptide is predominantly expressed in the hypothalamus where it is co-localized with arginine vasopressin, in the pituitary gland (in the intermediate lobe, hence the name intermedin), and in the kidney. Its distribution is consistent with a physiological role in the central regulation of thirst, fluid, and electrolyte balance. IMD is a potent systemic and pulmonary vasodilator. It influences regional blood flow and enhances cardiac contractility.

Calcitonin Receptor-Stimulating Peptide 1

CRSP-1 is expressed and secreted in the hypothalamus, mid-brain, and pituitary. It has a high amino acid sequence similarity with CGRP and participates in physiological regulation in peripheral tissue and central nervous system (CNS). It inhibits food intake, pain sensation, and locomotive activity. It has been identified in mammals but not specifically in humans.

Receptors for the Calcitonin Peptide Family

Receptors for the whole CT peptide family can be reconstituted by two GPCRs (the CTR and the calcitonin receptor-like receptor CLR) and three RAMPs.

The CT Receptor

Distribution

CTRs are widely distributed. They are particularly numerous in osteoclasts, where there are approximately 1.3 million receptors per cell. They are found in smaller amounts in the CNS, in areas involved in the control of appetite, pain perception, and of lactation, for example, the hypothalamus and the pituitary. CTRs are also found in peripheral tissues, for example, in cells of the distal nephron of the kidney, testes, placenta, lung, prostate, and lymphocytes. Receptors for CT have been identified in some human cancer cells, including those of lung, breast, bone, and prostate.

Protein structure

The CTR was initially cloned from a pig kidney cell line in 1991. Analysis of the predicted 482 amino acid sequences demonstrated seven hydrophobic regions that could generate transmembrane (TM)-spanning domains. The CTR is a member of a subset of the G-protein-coupled receptors (GPCRs) family termed GPCR_β. Members of this family typically recognize regulatory peptides, including parathyroid hormone (PTH/PTH-rp), vasoactive intestinal polypeptide (VIP), glucagons, growth hormone-releasing hormone, gastrin inhibitory polypeptide, and secretin. This family of receptors is characterized by seven TM segments connected by three extracellular and three intracellular loops (Figure 2). They all have an extracellular amino (N)-terminal sequence and an intra-cellular carboxy (C)-terminal sequence. The extended extracellular N-terminal region contains multiple potential glycosylation sites and conserved cysteine

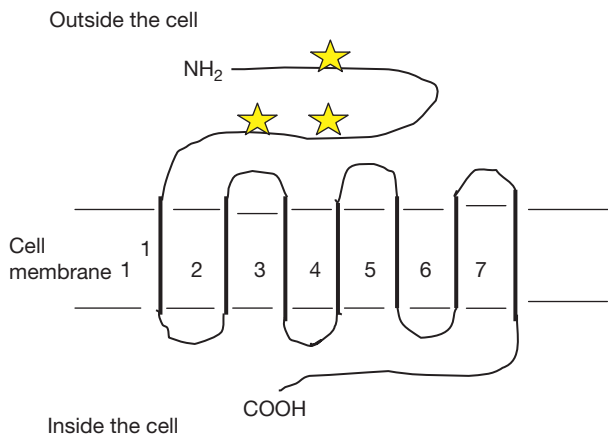


Figure 2 Schematic representation of the porcine CTR. Seven transmembrane regions span the cell membrane, and are connected by three extracellular and three intracellular loops. The amino-terminal extracellular loop contains three glycosylation sites (indicated by stars) and is involved with CT recognition and binding. The intracellular domain is involved with G-protein coupling. Reproduced from Girgis SI, Moraldi-Bidhendi N, Mancini L, and MacIntyre I (2004) *Encyclopedia of Biological Chemistry*, 1st edn. London: Elsevier, with permission from Elsevier.

residues. Glycosylation is important for recognition and high-affinity binding of CT. While the TM domain sequences are conserved (40–60% identical), the N-terminal domains are generally less than 25% identical. The N-terminal domain fulfills the role for ligand binding and receptor specificity.

Multiple forms

The human CTR gene is localized to chromosome 7. The CTR gene has a complex structural organization with several CTR protein isoforms derived from alternative splicing of transcripts from a single gene. These isoforms are functionally distinct in terms of ligand-binding specificity, signal transduction pathway utilization, and tissue distribution. In humans, at least five splice variants have been found, whereas only two splice variants have been identified in rodents (C1a and C1b), which differ, by the presence or absence of an additional 37 amino acids, in the second extracellular domain. The C1a is more widely distributed and is the predominant isoform found in osteoclasts. C1b isoform is primarily localized in the CNS.

Signaling

The CTR is coupled to multiple signal pathways through interaction with different members of the heterotrimeric G-protein family. These are a family of plasma membrane regulatory proteins that interact with and alter activated receptors. The main signal pathways are as follows.

The adenylate cyclase/cAMP/protein kinase A. The binding of CT to its receptor induces conformational change in the receptor that allows it to bind to an adjacent membrane G-protein, known as G_s (the subscript s denotes stimulation). The binding causes the G_s protein to activate the membrane enzyme called 'adenylate cyclase'. The activated adenylate cyclase, whose catalytic site is on the cytosolic site of the plasma membrane, then catalyzes the conversion of cytosolic ATP to cyclic 3', 5' adenosine monophosphate (cAMP). Cyclic AMP can diffuse throughout the cell to trigger the sequence of events leading to the cell's ultimate response to CT.

The phosphoinositol-dependent phospholipase C (PLC) pathway. The pathway which results in both an increase in intracellular Ca²⁺ and protein kinase C (PKC) activation via G_q protein. Both cAMP and intracellular Ca²⁺ are important second messengers for mediating the actions of CT on the osteoclast.

While adenylate cyclase, PLC, and PLD are established effectors for CTR, recently CTR-mediated activation of the mitogen-activated protein kinase (MAPK) pathway has been described. MAPKs are a group of serine/threonine protein kinases that are activated by several distinct classes of cell surface receptors, including G-protein coupled receptors and tyrosine kinases.

Regulation

Regulation of the level or affinity of cell-surface receptors is a key component in the response to either endogenous or administered agents. The CTR is subject to both homologous (CT-induced) and heterologous regulation. Although CT effectively inhibits osteoclast-mediated bone resorption after acute administration, continuous exposure to CT results in a loss of responsiveness due to receptor downregulation. Continuous treatment of osteoclast-like cells with CT results in the decrease in steady state levels of CTR mRNA and downregulation of CTR binding. It is also known that glucocorticoids stimulate CTR expression.

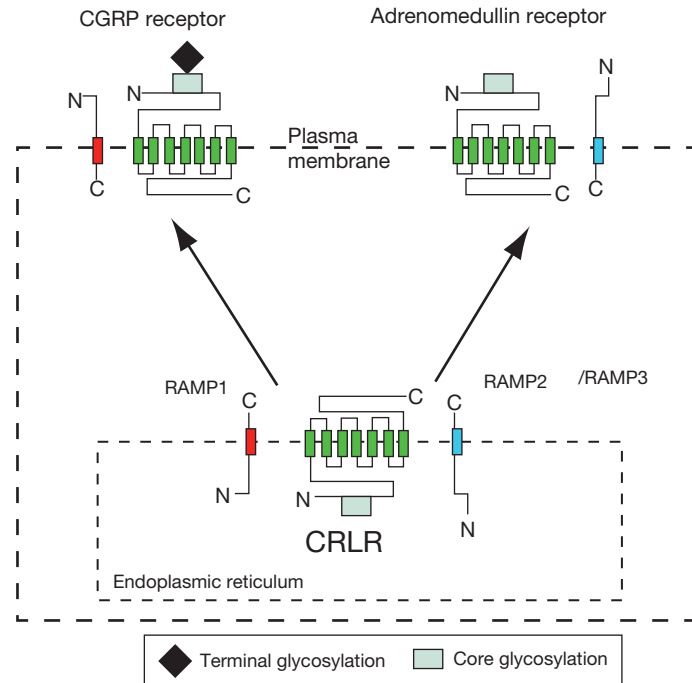


Figure 3 The role of RAMPs 1 and 2 and CLR (CRLR) in generating CGRP or ADM receptors. Adapted from McLatchie LM, Fraser NJ, Main MJ, et al. (1998) RAMPs regulate the transport and ligand specificity of the calcitonin receptor-like receptor. *Nature (London)* 393: 333–339, with permission from Nature Publishing Group.

CTR-Like Receptor

The rat CLR was first identified in 1993. The receptor has a significant sequence homology (55%) with the corresponding CTR and share similarities in general structure and length (belongs to the GPCR_β family of receptors). Historically, this protein has been abbreviated as CRLR; however, it will be referred to in this article as CLR. Initially, it was considered an ‘orphan’ receptor. The important breakthrough came in 1998 when an expression cloning approach demonstrated that CLR required a single-cell membrane protein called receptor activity modifying protein, RAMP1, to function as a CGRP receptor (Figure 3). RAMP2 and RAMP3 enable the same CLR to behave as an AM receptor (Table 1). CLR alone does not bind with high affinity to any of the members of the CT peptide family.

Immunohistochemical analysis revealed the presence of the CRL in the endothelium of all blood vessels including large and small arteries, veins, capillaries, in heart muscle cells, and the endocardium. The CLR is ubiquitously expressed in a variety of tissues, being most prevalent in the brain, and in high densities in the pulmonary and cardiovascular system (especially in blood vessels). It is also expressed in the adrenal and pituitary glands, in the pancreas in the islets of Langerhans, kidney, and bone.

Receptor-Activity-Modifying Proteins

The RAMP family of proteins comprises three members: RAMP1, RAMP2, and RAMP3. They belong to the family of type I single trans-membrane proteins. They have a large

Table 1 Nomenclature of receptors resulting from the association of individual RAMPs with either CTR or CLR

Molecular units	Receptor name
CTR + RAMP1	AMY1
CTR + RAMP2	AMY2
CTR + RAMP3	AMY3
CLR + RAMP1	CGRP
CLR + RAMP2	AM1
CLR + RAMP3	AM2

Amylin (AMY) receptors are formed by the association of the calcitonin receptor (CTR) with one of the three receptor-activity modifying proteins (RAMPs), whereas CGRP and adrenomedullin (AM) receptors are formed by the association of the calcitonin receptor-like receptor (CLR) with one of the RAMPs. From Poyner DR, Sexton PM, Marshall I, et al. (2002) International Union of Pharmacology. XXXII. The mammalian calcitonin gene-related peptides, adrenomedullin, amylin, and calcitonin receptors. *Pharmacological Reviews* 54: 233–246; Hay DL, Poyner DR, and Quirion R (2008) International Union of Pharmacology. LXIX. Status of calcitonin gene-related peptide subtype 2 receptor. *Pharmacological Reviews* 60: 143–145.

extracellular amino terminal domain, a single TM spanning domain and a short, ~10 amino acids, cytosolic domain. The three RAMPs share a similar basic structure but have only about 30% identity in the amino acid sequence. One important function of RAMPs is to control the glycosylation of the CLR and thus its transportation to the cell membrane. For example, RAMP1 acts to modify the glycosylation of CLR, enhance the transport of the receptor protein to the cell surface and potentially contribute to the receptor phenotype. RAMP1-transported CLR is a CGRP receptor. RAMP2 and 3 also enhance the transport of CLR to the cell surface, but do not

alter the pattern of glycosylation of the receptor. RAMP2 or 3-CLR is an ADM receptor.

RAMP1 is widely expressed in the brain, spinal cord, gastrointestinal tract, adrenal gland, perivascular nerve, and smooth muscles of the arteries. RAMP2 is found in macrovascular and microvascular endothelial cells, vascular smooth muscle, and cardiomyocytes. RAMP3 is less extensively distributed: high levels are found in kidney, lung, spleen, and thymus.

See also: [Signaling: Calcitonin Gene-Related Peptide and Adrenomedullin Receptors.](#)

Further Reading

- Bell D and McDermot BJ (2008) Intermedin (adrenomedullin-2): A novel counter-regulatory peptide in the cardiovascular and renal systems. *British Journal of Pharmacology* 153: S247–S262.
- Brain SD and Grant AD (2004) Vascular actions of calcitonin gene-related peptide and adrenomedullin. *Physiological Reviews* 84: 903–934.
- Chang CL, Roh J, and Hsu SYT (2004) Intermedin, a novel calcitonin family peptide that exists in teleosts as well as in mammals: A comparison with other calcitonin/intermedin family peptides in vertebrates. *Peptides* 25: 1633–1642.
- Christopoulos G, Perry KJ, Morfis M, et al. (1999) Multiple amylin receptors arise from receptor activity – modifying protein interaction with the calcitonin receptor gene product. *American Society for Pharmacology and Experimental Therapeutics* 56: 235–242.
- Galson DL and Goldring SR (2002) Structure and molecular biology of the calcitonin receptor. In: Bilezikian JP, Raisz LG, and Rodan GA (eds.) *Principle of Bone Biology*, 2nd edn., vol. 1, pp. 603–617. San Diego, CA: Academic Press.
- Girgis SI and MacIntyre I (2002) Calcitonin gene-related peptide (CGRP). In: Creghton TE (ed.) *Encyclopedia of Molecular Medicine*, 5 vols. New York, NY: Wiley.
- Girgis SI, Moraldi-Bidhendi N, Mancini L, and MacIntyre I (2004) *Encyclopedia of Biological Chemistry*, 2nd edn. London: Elsevier.
- Granhölm S, Lundberg P, and Lerner UH (2008) Expression of the calcitonin receptor, calcitonin receptor-like receptor and receptor activity modifying proteins during osteoclast differentiation. *Journal of Cell Biology* 104: 920–933.
- Hay DL, Poyner DR, and Quirion R (2008) International Union of Pharmacology. LXIX. Status of Calcitonin gene-related peptide subtype 2 receptor. *Pharmacological Reviews* 60: 143–145.
- Ishimitsu T, Ono H, Minami J, and Matsuka H (2006) Pathophysiologic and therapeutic implications of adrenomedullin in cardiovascular disorders. *Pharmacology and Therapeutics* 111: 909–927.
- Katafuchi T, Yasue H, Osaki T, and Minamino N (2009) Calcitonin receptor-stimulating peptide: Its evolutionary and functional relationship with calcitonin/calcitonin gene-related peptide based on gene structure. *Peptides* 30: 1753–1762.
- Lerner UH (2006) Deletions of genes encoding calcitonin/ α -CGRP, amylin and calcitonin receptor have given new and unexpected insights into the function of calcitonin receptors and calcitonin receptor-like receptors in bone. *Journal of Musculoskeletal and Neuronal Interactions* 6: 87–95.
- McLachlan LM, Fraser NJ, Main MJ, et al. (1998) RAMPs regulate the transport and ligand specificity of the calcitonin receptor-like receptor. *Nature (London)* 393: 333–339.
- Naot D and Cornish J (2008) The role of peptides and receptors of the calcitonin family in the regulation of bone metabolism. *Bone* 43: 813–818.
- Pondel M (2000) Calcitonin and calcitonin receptors: Bone and beyond. *International Journal of Experimental Pathology* 81: 405–422.
- Poyner DR, Sexton PM, Marshall I, et al. (2002) International Union of Pharmacology. XXXII. The mammalian calcitonin gene-related peptides, adrenomedullin, amylin, and calcitonin receptors. *Pharmacological Reviews* 54: 233–246.
- Sexton PM, Findlay DM, and Martin TJ (1999) Calcitonin. *Current Medicinal Chemistry* 6: 1067–1093.

Calcium-Binding Proteins: Cytosolic (Annexins, Gelsolins, and C₂-Domain Proteins)

J Krebs, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

Actin A eukaryotic protein that has the capability to form thin helical filaments.

Alternative splicing Alternate usage of particular exons to create isoforms of a given protein.

Amphipathic proteins Proteins composed of helices containing both hydrophobic and hydrophilic amino acid residues.

Apoptosis Apoptosis (or programmed cell death) is derived from a Greek word describing the shedding of leaves from trees. During apoptosis, the cell responds to specific physiological or developmental signals undergoing a regulated, well-programmed series of events which will lead to its death and its removal from the organism.

Cytoskeleton Complement of actin filaments, microtubules, and intermediate filaments forming a network in the cytoplasm.

EF-hand proteins Term coined by R.H. Kretsinger to describe the helix-loop-helix calcium-binding domains of specific proteins. The highly conserved motif (first described on the basis of the crystal structure of parvalbumin containing six helices, A–F) in which certain amino acids are invariant consists of two helices enclosing the Ca²⁺-binding

loop. As a model, the forefinger and the thumb of the right hand can resemble the two helices (e.g., E and F of the second Ca²⁺-binding domain of parvalbumin) and the bent midfinger the enclosed loop, hence the EF-hand.

Exon Segments of a eukaryotic gene preserved in the mature messenger RNA.

Intron Segments of a gene transcribed into the precursor RNA but excised by RNA splicing before the mature RNA is exported from the nucleus into the cytoplasm.

Posttranslational modification Enzymatic modification such as acetylation, phosphorylation, myristoylation, or ubiquitination of proteins regulate their activity, topology, or degradation.

Second messenger Intracellular small molecules (e.g., cyclic nucleotides or inositol polyphosphates) or ions such as Ca²⁺ or gases such as NO, indispensable for the transduction of signals converting extracellular stimuli – e.g., raised by hormones (=primary messengers) – into intracellular responses.

Splicing Removal of introns from the precursor messenger RNA and joining of exons in mature messenger RNA; positions of boundaries between introns and exons are often conserved in homologous proteins of different species.

Calcium is one of the most common elements on the Earth, and it is the fifth most abundant element of the human body. Inside cells, calcium is a very versatile second messenger involved in the regulation of a variety of different cellular processes. This pivotal role of calcium is made possible due to the binding of Ca²⁺ to a great variety of different calcium-binding proteins. Next to the well-studied family of the so-called EF-hand containing proteins, a number of other calcium-binding proteins became known in recent years. These proteins are composed of a number of repeat units containing a variety of different Ca²⁺-binding sites. The families of annexins, gelsolins, and C₂-domain containing proteins described in this article bind to different membranous or cytoskeletal structures in a Ca²⁺-dependent manner, and are involved in a number of different cellular functions.

Annexins

The family of intracellular Ca²⁺-binding proteins called ‘annexins’ are soluble, amphipathic proteins that bind to membranes containing negatively charged phospholipids in a Ca²⁺-dependent manner. They are called annexins since they bring or hold together different cellular structures, in particular membranes. The close to 200 different annexin proteins

known to date are widespread in the animal and plant kingdoms, and have been claimed to be involved in a variety of cellular functions such as interaction with the cytoskeleton, membrane fusion, anticoagulation, signal transduction, or phospholipase inhibition. The primary structure of the annexins contains four or eight (annexin A6) conserved repeat units of ~75 amino acids in length. This protein core consists of five alpha helices, which are separated by intervening sequences of variable length. These repeat units probably originated from gene duplications, a view which is corroborated by the evolutionary conservation of the intron–exon boundary positions. Most annexins containing four repeat units are comprised of 12–15 exons resulting in a quite variable amino-terminal part of the different annexins. Most alternative splicing sites have been located within exons encoding the variable amino-terminal parts of annexins. Since these parts are also the locations containing motifs for binding partners or for post-translational modifications, alternative splicing may contribute to the regulation of annexin function.

After the first determination of an annexin crystal structure (annexin A5) by Huber and his co-workers, a number of additional annexin structures from different sources (including those from lower eukaryotes and plants) have been reported. From these structures, it can be concluded that the conserved repeat units are packed into an α -helical disk, as first described for

annexin A5, which is almost entirely α -helical. It consists of five α -helices bundled into a right-handed super-helix. On the basis of this structure it was proposed that annexin A5 functions as a calcium channel, and some experiments using an *in vitro* reconstituted system seem to support a voltage-gated mechanism. This property is connected with the structure of the annexin core, which is shared by most annexins. Meanwhile, for most annexins a Ca^{2+} channel activity has been demonstrated in artificial bilayer systems, but never *in vivo*; therefore, the physiological relevance of these observations is questionable.

In contrast to the EF-hand containing Ca^{2+} -binding proteins, the ligands coordinating calcium in the annexins are not adjacent in sequence. Several calcium-binding sites seem to exist in annexins, two invariably in repeats II and IV, one in either repeat I or III, but annexins may contain as many as 10–12 Ca^{2+} -binding sites along the membrane-binding surface of the protein. The sites with the highest Ca^{2+} affinity are structurally related to the Ca^{2+} - and phospholipid-binding site of phospholipase A₂. The calcium ion is heptacoordinated with ligands organized in a pentagonal bipyramidal arrangement with ligating oxygens provided mainly from peptide carbonyls and water molecules, together with a single side-chain oxygen from a distant part of the sequence. Replacement of this latter capping residue, which is usually acidic, with alanine precludes Ca^{2+} binding at the site. Once bound to membranes, many annexins oligomerize to form highly ordered two-dimensional arrays that have been shown to strongly influence *in vitro* membrane properties, for example, increased rigidity.

An interesting feature is the role of Trp185 in the Ca^{2+} -binding site of repeat III of annexin A5 (see Figure 1). As suggested by Seaton and others, a Ca^{2+} -dependent exposure of Trp185, which is buried within the protein core in the absence of calcium, may facilitate the interaction between annexin A5 and the phospholipid bilayer. This view has been supported by fluorescence data and by mutagenesis experiments in which replacement of Trp185 by alanine decreased the annexin A5 membrane-binding affinity.

Annexins have long been known as targets for post-translational modifications. In fact, annexin A2 was originally isolated as a major substrate of the *src*-encoded protein kinase, and annexin A1 was long known to be phosphorylated by the tyrosine kinase activity of the epidermal growth factor (EGF) receptor, suggesting that annexins A1 and A2 could be used as

coupling factors between growth factor receptors and their cellular targets, since these phosphorylations alter the Ca^{2+} -dependent binding of annexin A1 and A2 to membranes. In addition to the phosphorylation by different tyrosine kinases, also serine/threonine kinases have been identified to phosphorylate annexins. As a result, annexin phosphorylation often leads to an altered susceptibility toward proteolysis. On the other hand, phosphorylation of annexin A2 by different kinases interferes with its ability to form stable heterotetrameric complexes with p11, a member of the calcium binding S100 family (S100A10), since these phosphorylation sites are located – like in the other annexins – within the variable amino-terminal domain, which comprises also binding motifs to interact with other proteins. S100A10 itself is not able to bind Ca^{2+} due to mutations critical for Ca^{2+} binding. Based on the crystal structure solved by Rety et al., it was suggested that S100A10 is in a kind of Ca^{2+} -activated conformation even in the absence of Ca^{2+} which enables S100A10 to form a heterotetrameric complex with annexin A2 as reported by Rety et al. who also solved the structure of the heterotetrameric complex between S100A10 and the N-terminal peptide of annexin A2 representing the S100A10 binding site. The Ca^{2+} sensitivity of the annexin A2/S100A10 complex in phospholipid binding is increased by three orders of magnitude as compared to free annexin A2. Next to the interaction between annexin A2 and S100A10 (p11), two other complexes between an annexin and an EF-hand containing protein of the S100 family may exist, that is, interactions between annexin A11 and S100A6 and annexin A1 and S100A11, respectively, but in contrast to annexin A2/S100A10 the existence of the latter complexes *in vivo* has not been described to date.

In spite of extended studies in many laboratories over the years, the precise function of individual annexins is not yet understood. The accumulated data suggest that due to the many ways that annexins interact with different membranes in a Ca^{2+} -dependent manner, annexins may participate in Ca^{2+} signaling as effectors, mediators, or even as regulators, but to clarify their biological activities unambiguously, this has to await further experiments. However, *in vivo* experiments from knockout and transgenic mice are beginning to shed some light on annexin function.

A number of annexinopathies have been described in the literature, for example, connected with the anti-phospholipid syndrome, diabetes, or with different forms of heart failures. A direct connection between an annexin anomaly and a human disease seems to occur in acute promyelocytic leukemia (APL), a subset of acute myelocytic leukemia (AML). It is characteristic for cells from patients with APL that annexin A2 is upregulated. These cells can be characterized by a disruption of the retinoic acid receptor alpha gene due to a fusion with the PML tumor-suppressor gene product followed by a translocation which causes a fibrinolytic phenotype. It is interesting to note that this phenotype can be induced in cells overexpressing annexin A2 and inhibited by annexin A2-specific antibodies in promyelocytic cells showing the specific translocation mentioned before. These observations indicate that dysregulated overexpression of annexin A2 may cause the fibrinolytic complications of APL. In another recent study, annexin A1 was shown to regulate transforming growth factor (TGF)- β signaling to promote the formation of metastatic cells in basal-like breast

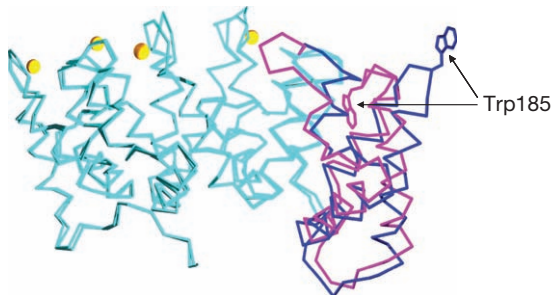


Figure 1 Superposition of α -carbon backbones of annexin A5 in the Ca^{2+} -free and the Ca^{2+} -bound form. Domains 1, 2, and 4 are light green, both in the Ca^{2+} -free and in the Ca^{2+} -bound form. Domain 3 is shown in magenta (Ca^{2+} free, trp185 buried) and in blue (Ca^{2+} bound, trp185 exposed). Courtesy: Barbara Seaton.

cancer (BLBC). In a proteomic analysis, De Graauw et al. in 2010 identified annexin A1 as a possible candidate to regulate a cell morphology switch to change epithelial cells to a more migratory phenotype. In a detailed study, the authors provided evidence that high expression of annexin A1 is associated with the BLBC phenotype of breast cancer, whereas depletion of annexin A1 prevented the migratory phenotype associated with decreased TGF- β signaling. Using annexin A1 as a marker in tissue microarrays, BLBC patients could be unambiguously discriminated from other breast cancer patients indicating that annexin A1 is functionally involved in BLBC progression and metastasis formation.

Gelsolin

Gelsolin belongs to the superfamily of actin-binding proteins (ABPs) expressed in all eukaryotes. It is a multifunctional protein, binding actin in a Ca²⁺-dependent manner, and is composed of up to six repeats of 120–150 amino acids. To the gelsolin family belong proteins such as villin, adseverin, CapG, filil, and others. Villin, for example, is a six gelsolin-domain ABP regulating actin assembly in microvilli, whereas CapG is a smaller, three-domain gelsolin analog that responds to Ca²⁺ and phosphatidylinositol-(4,5)-diphosphate under conditions where gelsolin is ineffective.

Recently, the crystal structure of the C-terminal region of adseverin, containing six domains (A1–A6) homologous to the six gelsolin domains (G1–G6), was solved by the group of Robinson. Biochemical studies have indicated that gelsolin contacts actin through two sites on G4 and G6 whereas adseverin interacts with actin by using a site on A5. By a detailed comparison between the Ca²⁺-bound structures of G4–G6 from gelsolin with A4–A6 from adseverin, a high conservation in calcium and actin binding could be detected which is not surprising due to the 60% sequence identity between the two proteins. Therefore, these structural data indicate that adseverin and gelsolin respond to calcium and bind actin monomers in a similar manner.

Alternative transcription initiation and selective RNA processing produces two isoforms of gelsolin from the same gene. One isoform of 80 kDa is located intracellular, whereas the other extracellular, slightly larger (83-kDa) protein is derived by alternative splicing, which results in a short amino-terminal elongation of the protein. The intracellular gelsolin is involved in cell motility regulating actin function, whereas extracellular gelsolin can act as an actin-scavenging system to prevent the polymerization of actin released after cell death.

Gelsolin consists of six repeat units (G1–6). The units are organized in two clusters of similar architecture, and are connected by a flexible linker of ~50 amino acids that may be cleaved by caspase-3, thereby cleaving the actin-binding domain of gelsolin from its calcium-binding domain. Caspase-3 cleavage and separation of subdomains of gelsolin coincide with plasma membrane blebbing, one of the characteristics of apoptosis. Apoptosis could also be induced by the overexpression of the amino-terminal half of gelsolin, whereas neutrophils of gelsolin-null mice have a delayed onset of apoptosis.

The structure of gelsolin in the absence of Ca²⁺ has been determined at 2.5 Å resolution by Robinson and his

co-workers (Figure 2). The amino-terminal half can bind to two actin monomers independent of calcium, whereas the carboxy-terminal half binds a single actin in a Ca²⁺-dependent manner.

A detailed mechanism for the regulation of gelsolin activity by Ca²⁺ has recently been proposed. This was made possible by comparing the Ca²⁺-free structure of nonactive gelsolin (Figure 2) with the recently solved structure of the active form of Ca²⁺-bound gelsolin complexed to actin (see Figure 3). In this latter structure consisting of the C-terminal half of gelsolin, that is, G4–6, bound to monomeric actin, two classes of Ca²⁺-binding sites have been identified: type-1 sites bind calcium in a coordination sphere shared by actin and gelsolin, whereas type-2 sites are identified only in gelsolin, that is, located within G5 and G6. Due to conservation of the ligating residues within all gelsolin repeats, type-2 Ca²⁺-binding sites should exist in all six domains of gelsolin. In total, gelsolin should be able to bind eight calcium ions in the active gelsolin–actin complex, that is, two in type-1 sites and six in type 2. The latter are proposed to facilitate structural rearrangements within gelsolin as part of the activation and actin-binding process.

In the absence of Ca²⁺, the six repeat units of gelsolin provide a very compact globular structure, thus blocking the actin-binding helices of the appropriate subdomains. Therefore, in a first step, this compact globular arrangement of the six subdomains has to be opened up. As suggested by H. Choe and co-workers in 2002, binding of Ca²⁺ first to G6 should induce a conformational rearrangement in which G6 is flipped over G5 to tear apart the continuous β -sheet core of G4 and G6, thereby unmasking the actin-binding site on G4, and hence permit binding to actin strands (see Figure 3). Similar events should also occur in the N-terminal half of gelsolin, that is, Ca²⁺ binding to G3, flipping over G2 to open up the actin-binding site of G1. Finally, to tighten up the gelsolin–actin complex, Ca²⁺ binds to the two type-1 sites shared between actin and gelsolin, which might explain the very high affinity of gelsolin to actin (K_d 50 nM).

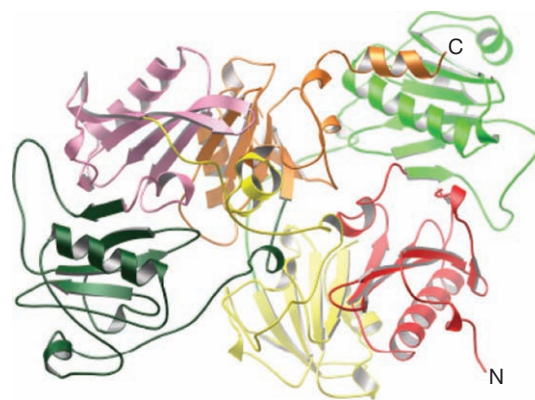


Figure 2 A ribbon diagram of the structure of gelsolin in the absence of calcium. The following color code for the different segments (G1–6) has been used: G1 (containing the amino terminus), red; G2, light green; G3, yellow; G4, magenta; G5, dark green; and G6 (containing the carboxy terminus), gold. Reproduced from Burtnick LD, Koepf EK, Grimes J, et al. (1997) The crystal structure of plasma gelsolin: Implications for actin severing, capping, and nucleation. *Cell* 90: 661–670, with permission from Elsevier.

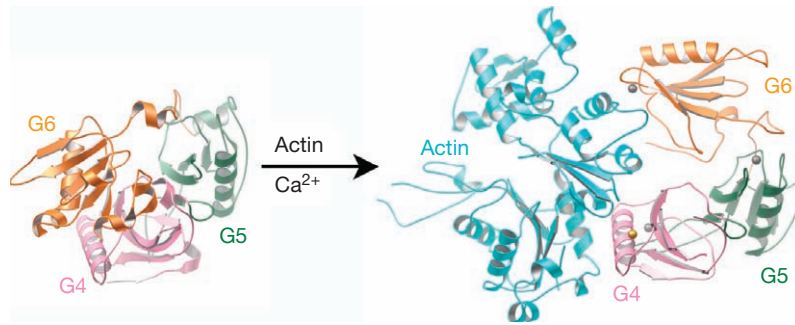


Figure 3 Ribbon diagram of gelsolin domains G4–G6 demonstrating the influence of binding Ca^{2+} and actin (shown in cyan) on the conformational changes of the protein. The same color code as in [Figure 2](#) is used for domains G4–G6. The left-hand panel shows G4–G6 in a Ca^{2+} -free conformation (see [Figure 2](#)), whereas the right-hand panel depicts the actin and Ca^{2+} -bound form of G4–6. Reproduced from Choe H, Burtnick LD, Mejillano M, et al. (2002) The calcium activation of gelsolin: Insights from the 3 Å structure of the G4–G6/actin complex. *Journal of Molecular Biology* 324: 691–702, with permission from Elsevier.

In a recent study, Nag et al. in 2009 made the interesting observation that G2 and G6 seem to be cooperative with respect to Ca^{2+} binding to gelsolin, and thus be responsible for the exposure of the actin-binding site on G2. By a detailed analysis it could be shown that due to mutations of the two conserved Ca^{2+} -coordinating residues Asp and Glu to Asn and Gln in G2 a destabilization of the gelsolin structure could be observed, resulting in a sustained exposure of a protease cleavage site. In this respect it is of interest that the hereditary disease known as familial amyloidosis Finnish-type (FAF) is based on a single mutation within the Ca^{2+} -binding site in G2 of gelsolin in which Asp187 is mutated to Asn or Tyr, thereby destabilizing the gelsolin structure through loss of Ca^{2+} coordination. This makes the protein susceptible to protease cleavage in the Golgi compartment.

Gelsolin-null mice show normal embryonic development, but suffer subtle changes. This suggests the need of gelsolin for cell motility during processes such as hemostasis leading to reduced platelet function, inflammation leading to delayed neutrophil migration, or during wound healing, which would lead to reduced movement of fibroblasts. In this context, it is interesting to note that it was demonstrated that during development of zebrafish embryos, gelsolin is required for dorso-ventral patterning. Inhibition of gelsolin expression that starts to be expressed already at the two-cell stage resulted in ventralized phenotypes, some of which lacked brain or eyes. These phenotypes could be rescued by injecting zebrafish gelsolin messenger RNA (mRNA) or even by injecting human gelsolin protein. These data indicate that gelsolin may have at least two separate functions: a structural role for the cytoskeletal/cell motility apparatus and a regulatory role during development.

C₂-Domain Proteins

Interaction of proteins, intracellularly or extracellularly, often occurs via different binding modules or domains such as SH2, SH3, WW, PDZ, or C₂-domains. These modules are formed by folding domains consisting of 100–150 residues with different binding properties, that is, SH2-domains interact with phosphotyrosine-containing sequences, SH3- and WW-domains with proline-rich sequences, PDZ-domains with C-terminal

sequences to link multiprotein complexes to the cytoskeleton, and C₂-domains with phospholipids, some in a Ca^{2+} -dependent, some in a Ca^{2+} -independent manner. C₂-domains consist of about 120–130 amino acid residues. They were first identified in protein kinase C. More than 100 C₂-domain containing proteins can now be identified in current data banks.

Most proteins containing C₂-domains either are linked to signal transduction pathways or are involved in membrane traffic. The former proteins either generate lipid second messengers (e.g., phospholipase A2, phospholipase C, or phosphatidylinositol-3-kinase), phosphorylate proteins (e.g., protein kinase C), or ligate ubiquitin (e.g., Nedd4). Those proteins involved in membrane traffic include, for instance, the Rab-binding proteins rabphilin and RIM, which are involved in regulating the exocytosis of secretory vesicles. However, the best-characterized protein of this category is synaptotagmin, which will be discussed in more detail.

Synaptotagmin I is a transmembrane protein belonging to a family of more than 10 members which contain two C₂-domains, C₂A and C₂B, comprising most of the cytoplasmic region of synaptotagmins. The C₂-domains contain Ca^{2+} -binding domains of non-EF-hand character. Synaptotagmin I is found in synaptic vesicles, and is believed to act as the major Ca^{2+} -sensor of exocytosis and neurotransmitter release. Important in this respect is the Ca^{2+} -dependent binding of synaptotagmin to syntaxin, a member of the SNARE family of membrane proteins, involved in vesicle transport mechanisms.

C₂-domains have been suggested to be responsible for binding to membranes in response to Ca^{2+} . Sudhof and his co-workers determined the first structure of a C₂-domain, the C₂A-domain of synaptotagmin I (see [Figure 4](#)). The C₂-domain consists of a β -sandwich of two four-stranded β -sheets ([Figure 4\(a\)](#)). The eight β -strands are connected by three loops at the top that bind Ca^{2+} in the form of a cluster, primarily through oxygen of aspartate side chains serving as bidentate ligands, and by four loops at the bottom lacking Ca^{2+} -binding sites. In contrast to Ca^{2+} binding to EF-hand type of sites, which causes substantial conformational changes, Ca^{2+} binding to sites of C₂-domains leads to structural stabilization rather than backbone rearrangements.

Structural determination of other C₂-domains revealed similar designs, but significant differences in the topology of the

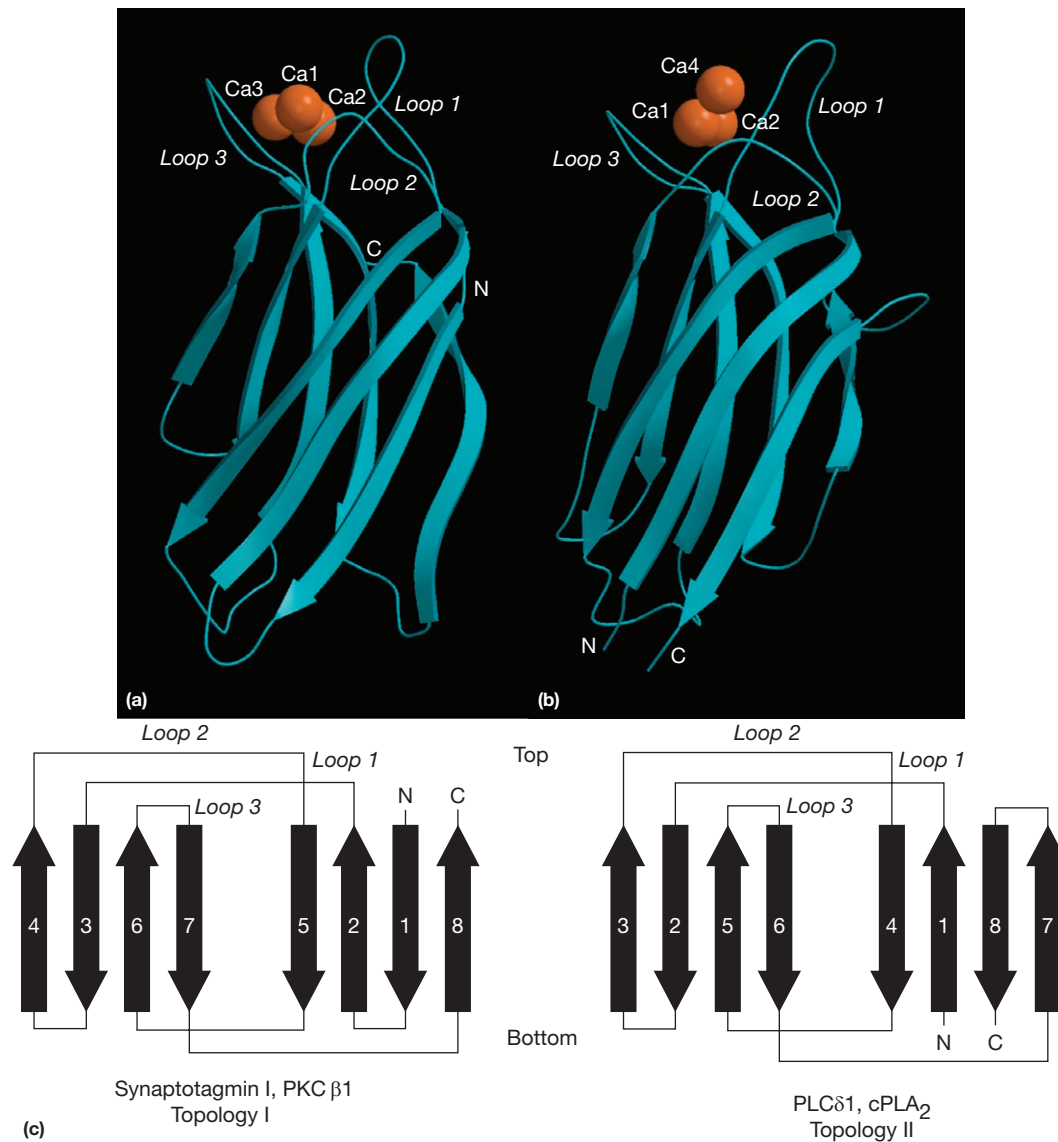


Figure 4 Ribbon diagrams of the structures of the C₂A-domain of synaptotagmin I (a) and of the C₂-domain of phospholipase Cδ1 (b). (c) A schematic drawing of the β-strand topologies of the two structures. The locations of the Ca²⁺ clusters (orange) are indicated. Reproduced from Rizo J and Südhof TC (1998) C₂-domains, structure and function of a universal Ca²⁺-binding domain. *Journal of Biological Chemistry* 273: 15879–15882, with permission of the American Society for Biochemistry & Molecular Biology.

arrangement of the β-strands (see Figure 4(c)). It could be shown that the structure of the C₂-domain of protein kinase Cβ is very similar to synaptotagmin I (topology I), whereas the topological arrangement of the structure of the C₂-domain of phospholipase Cδ is significantly different (topology II) (see Figure 4(c)). Phospholipase Cδ contains several protein modules including EF-hand Ca²⁺-binding domains, which provides evidence for the existence of proteins containing EF-hand and non-EF-hand Ca²⁺-binding sites.

An interesting finding concerning important functional differences between Ca²⁺-binding sites of C₂A- and C₂B-domains of synaptotagmin has recently been reported. It has been shown that by mutating an aspartate essential for Ca²⁺ binding to the C₂A-domain into an asparagine, this domain lost its Ca²⁺-dependent binding to phospholipids or to syntaxin.

However, introducing such a mutated synaptotagmin into the germline of *Drosophila* lacking synaptotagmin I that were severely impaired in neurotransmitter release, the mutated synaptotagmin I could fully rescue this defect, indicating that Ca²⁺ binding to the C₂A-domain of synaptotagmin is not essential for neurotransmitter release. However, by replacing an aspartate for asparagine essential for Ca²⁺ binding in the C₂B-domain, this mutated protein could not rescue such a defect in a similar set of experiments suggesting a significant functional difference in Ca²⁺ binding between the C₂A- and C₂B-domains of synaptotagmin I.

In another C₂-domain containing protein, Nedd4, it was demonstrated that binding of Ca²⁺ to the C₂-domain was responsible not only for the localization of the protein but also for part of its function. Nedd4, the neuronal precursor

cell-expressed developmentally downregulated four protein, is a multimodular ubiquitin protein ligase. In humans, there are two Nedd4 genes located on chromosomes 15 and 18, respectively. This enzyme is involved in controlling the turnover of membrane proteins. Four groups of Nedd4 substrates have been identified to date: membrane transporters and ion channels, membrane receptors like EGFR, tumor suppressors, and endocytic regulatory proteins. It was shown that localization of Nedd4 to the apical region of polarized epithelial cells was dependent on the Ca²⁺ binding to its C₂-domain. Here, an important target of Nedd4 is the Na⁺-channel (ENaC), which plays a critical role in Na⁺ homeostasis of epithelial cells. Interestingly, the WW-domain of Nedd4, but not the C₂-domain was required to downregulate EnaC. On the other hand, it was shown that Nedd4 lacking its C₂-domain was still able to ligate ubiquitin to appropriate membrane protein targets such as Gap1, the general amino acid permease, but subsequent internalization by endocytosis, a prerequisite for the downregulation of membrane receptor proteins, was impaired. Recently, Wang et al. (2010) reported that the C₂ domain of Nedd4 is the auto-inhibitory domain for its ubiquitin ligase activity. By binding of Ca²⁺ to the C₂ domain the ubiquitin ligase activity of Nedd4 is activated *in vitro* and *in vivo*. On the other hand, if the C₂ domain is deleted Nedd4 is constitutively activated.

See also: **Bioenergetics:** Calcium-Modulated Proteins (EF-Hand); **Cell Architecture and Function:** Actin-Capping and -Severing Proteins; **Molecular Biology:** Catalysis by RNA, Structural Themes: Group I Introns; **Signaling:** Phospholipase C.

Further Reading

Burtnick LD, Koepf EK, Grimes J, et al. (1997) The crystal structure of plasma gelsolin: Implications for actin severing, capping, and nucleation. *Cell* 90: 661–670.

- Choe H, Burtnick LD, Mejillano M, et al. (2002) The calcium activation of gelsolin: Insights from the 3 Å structure of the G4–G6/actin complex. *Journal of Molecular Biology* 324: 691–702.
- Chumrarnsilpa, Lee WL, Nag S, et al. (2009) The crystal structure of the C-terminus of adseverin reveals the actin-binding interface. *Proceedings of the National Academy of Sciences of the United States of America* 106: 13719–13724.
- Concha NO, Head JF, Kaetzel MA, Dedman JR, and Seaton BA (1993) Rat annexin V crystal structure: Ca²⁺-induced conformational changes. *Science* 261: 1321–1324.
- de Graauw M, van Miltenburg MH, Schmidt MK, et al. (2010) Annexin A1 regulates TGF- β signalling and promotes metastasis formation of basal-like breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* 107: 6340–6345.
- Dos Remedios CG, Chhabra D, Kekic M, et al. (2003) Actin binding proteins: Regulation of cytoskeletal microfilaments. *Physiological Reviews* 83: 433–473.
- Gerke V and Moss SE (2002) Annexins: From structure to function. *Physiological Reviews* 82: 331–371.
- Hayes MJ, Longbottom RE, Evans MA, and Moss SE (2007) Annexinopathies. In: Carafoli E and Brini M (eds.) *Calcium Signalling and Disease*, pp. 1–28. Berlin: Springer.
- Huber R, Römisch J, and Paques EP (1990) The crystal and molecular structure of human annexin V, an anticoagulant calcium, membrane-binding protein. *EMBO Journal* 9: 3867–3874.
- Nag S, Ma Q, Wang H, et al. (2009) Ca²⁺-binding by domain 2 plays a critical role in the activation and stabilisation of gelsolin. *Proceedings of the National Academy of Sciences of the United States of America* 106: 13713–13718.
- Pallanck L (2003) A tale of two C₂-domains. *TINS* 26: 2–4.
- Réty S, Sopkova J, Renouard M, et al. (1999) The crystal structure of a complex of p11 with the annexin II N-terminal peptide. *Nature Structural and Molecular Biology* 6: 89–95.
- Rizo J and Südhof TC (1998) C₂-domains, structure and function of a universal Ca²⁺-binding domain. *Journal of Biological Chemistry* 273: 15879–15882.
- Spinardi L and Witke W (2007) Gelsolin and diseases. In: Carafoli E and Brini M (eds.) *Calcium Signalling and Disease*, pp. 55–69. Berlin: Springer.
- Swairjo MA and Seaton BA (1994) Annexin structure and membrane interactions: A molecular perspective. *Annual Review of Biophysics and Biomolecular Structure* 23: 193–213.
- Wang J, Peng Q, Lin Q, et al. (2010) Calcium activates Nedd4 E3 ubiquitin ligases by releasing the C2 domain-mediated auto-inhibition. *Journal of Biological Chemistry* 285: 12279–12288.

Calcium, Biological Fitness of

R J P Williams, University of Oxford, Oxford, UK

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 270–273, © 2004, Elsevier Inc.

Glossary

Biological minerals Shells and bones.

Calcium complexes Soluble combinations of calcium with ligands.

Calcium ion properties Ion size, ionization potential, and fast exchange.

Fitness The Darwinian concept of being valuable for a biological function.

Messenger function of calcium Calcium moving between internal compartments of a cell or between external and internal zones.

Organelles Mitochondria or chloroplasts. They differ from vesicles in having residual DNA/RNA synthetic systems, although both organelles and vesicles are compartments separated by membranes from the cytoplasm.

Calcium has a special cation, Ca^{2+} , in cellular activity. It is rejected by all cell cytoplasm either to the outside of cells or to the inside of resides. From these stores it is released by events outside eukaryote cells to the cytoplasm and organelles where it activates many processes. Thus, calcium is the dominant messenger, relaying information concerning the environment to activities such as contraction of muscles, release of digestive juices from the pancreas, to hormone release from glands. It is also the major cation in organizing external protein structures and in the formation of biological minerals – shells and bones.

Calcium: The Element and Its Ion, Ca^{2+}

Calcium, atomic number 20, lies in group 2 in the third row of the periodic table and before the first row of transition metals. It resides below magnesium and above strontium and then barium in its group. It has low ionization potentials, giving a rather large calcium ion, Ca^{2+} (radius $1.0\text{\AA}$), which is a rather poor Lewis acid and hence a somewhat poor acid catalyst. However, it can bind quite strongly to certain anion centers. Within its group, the chemistry of the Ca^{2+} ion has close resemblance to that of the larger strontium ion, Sr^{2+} , but differs markedly from the smaller magnesium ion, Mg^{2+} , which is a better Lewis acid.

Calcium, like magnesium, is an abundant element in the universe and on Earth, and its salts are quite soluble in water. Hence, calcium is available in the sea, 10^{-2}M , compared to magnesium at $5 \times 10^{-2}\text{M}$, whereas in freshwater it, like magnesium, is often as high as 10^{-3}M . It is against these background levels in water, especially in the sea, that life evolved.

The calcium ion has three major mass isotopes, 40, 42, and 44, which can be used to follow the reactions of its ions. It can also be followed in biological systems through association with fluorescent organic dyestuffs.

The Aqueous Calcium Ion

The free aqueous calcium ion has a variable fluctuating number of water molecules around it, exceeding six but not

exceeding nine. This loose hydration sphere is common to all cations of a radius greater than $0.8\text{\AA}$. In this way, the calcium ion differs from the magnesium ion, radius $0.6\text{\AA}$, which has a fixed, rather rigid coordination of six water molecules. The differences in structure are reflected in the rates of exchange of the water molecules from their respective coordination spheres, that of magnesium being rather slow, 10^6s^{-1} and that of calcium being close to the diffusion limit of 10^9s^{-1} .

Insoluble Salts

The binding of calcium, a large ion, is less than that of magnesium when the incoming anion is small (e.g., fluoride or hydroxide) and these magnesium salts tend to be the more insoluble. However, larger groups, which include the vast majority of anions (e.g., carbonate, sulfate, and phosphate anions), are more readily accommodated around the larger calcium ion. Hence, we observe that the vast majority of calcium salts are more insoluble than the corresponding magnesium salts. This becomes a feature in organisms, where we find calcium carbonate in shells and calcium phosphate in bones, whereas in plants calcium oxalate is often precipitated. Aberrant precipitation causes stones in animal organs. Among minerals, calcium and magnesium often occur together, as, for example, in dolomite.

Soluble Complex Ions and Their Binding Strengths

Calcium ions do not bind strongly to simple anions in water at pH 7. Thus, complexes with carbonate, sulfate, hydroxide, and phosphate are not of great importance in organisms. However, binding to ligands with several donor atoms can be strong because despite their bulkiness they can be fitted around the large calcium ion. By contrast, the Mg^{2+} ion fails to bind strongly, so these chelating ligands selectively bind calcium relative to magnesium by at least a 1000-fold. Typical examples are proteins both inside the cells and in external circulating fluids. The use of the calcium buffer (ethylene tetra-acetate) depends on this quite strong binding to calcium but weak binding to magnesium at pH 7.

Structures of Calcium Salts and Complexes

The structures of many calcium salts have been determined, and the extremes of their variation are shown even within the allotropic forms of its carbonates. Whereas calcite has a six-coordinate lattice, aragonite has a nine-coordinate structure. These salts have no water of crystallization and both are found in shells. More frequently, when calcium ions are bound to bulky chelating organic ligands in solution, calcium has seven near-neighbors, but the distances from the metal ion to the nearest ligand atom are very irregular, from 2.2 to 3.0 Å. Bound to saccharides or proteins, the ion has this structure, but it fluctuates. It is usual for one or two water molecules to remain bound. The binding ligand atoms are invariably oxygen from such centers as oxyanions, carboxylates, alcohols, carbonyls, and ethers.

Rates of Exchange

The rate of loss of water molecules from around Ca^{2+} , 10^9 s^{-1} , is the rate at which ligands can bind, k_{on} . The rate of leaving is related to the ligand-binding constant, $K = k_{\text{on}}/k_{\text{off}}$, so for $K = 10^6 \text{ M}^{-1}$ the leaving rate is 10^3 s^{-1} . This binding constant is close to that found for several proteins inside cells. The combination of such a protein (e.g., calmodulin) with calcium then provides a response time of 10^3 s^{-1} , as is seen in fast muscle cells. It is the rates of exchange together with the selective binding constants that make calcium an incomparable messenger in aqueous solution.

Biological Salts and Ligand Complexes

The major insoluble calcium salts found in organisms are the fluoride (in krill), carbonates (in shells in at least three allotropic forms), phosphates (in bone), and oxalates (in many plants). There are variable amounts of other metal ions and anions in these salts, so their exact nature is far from simple; for example, bone, hydroxy apatite, does not have an ideal stoichiometric formula and easily incorporates other elements including magnesium and fluoride. The apatite structure is somewhat different in bones of different organisms and even in the teeth and bones of one species (e.g., humans). Calcite is also found in the ears of animals, with a balance function.

In cellular complexes, calcium is bound, seven coordinate, to oxygen (O^-) donor groups of proteins and polysaccharides through both side- and main-chain atoms (e.g., $-\text{OH}$, $>\text{CO}$, CO_2^- ether, and SO_4^{2-}). There is little or no binding to DNA or RNA in cells, in marked contrast to the essential function of magnesium in the structures of these nucleotides. The selective binding arises from the differences in concentration and binding ability.

Concentrations and Selectivity of Binding in Organisms

The ability of calcium ions to precipitate many organic anions including DNA required that one primary necessity of the original life forms, prokaryotes, was the rejection of calcium. Consequently, all organisms hold Ca^{2+} concentrations at

10^{-6} M or below in their cytoplasm using calcium-outward pumps. Magnesium ions do not precipitate organic anions and are found at 10^{-3} M in all cell cytoplasm. Hence, for Ca^{2+} to bind selectively inside cells, its binding constant must be close to 10^6 M^{-1} , whereas that of Mg^{2+} has to be less than 10^3 M^{-1} . This selectivity between the two ions is found for many cytoplasmic messenger proteins, for example, calmodulin, troponin, and S-100. Meanwhile, Ca^{2+} fails to bind to molecules such as adenosine triphosphate (ATP) and DNA, $K = 10^4$, but Mg^{2+} can bind with the same binding constant. Outside cells, both Mg^{2+} and Ca^{2+} are $>10^{-3} \text{ M}$ and in eukaryotes the cell vesicles hold Ca^{2+} at this concentration also. Ca^{2+} can again bind selectively if its binding constant to proteins or now saccharides is $\geq 10^3 \text{ M}^{-1}$, whereas that of Mg^{2+} is less than 10^2 M^{-1} , because both ions are present outside the cells at 10^{-3} M . These constants are found for some coat, blood-clotting, and immune-response proteins and for some vesicle proteins. Note that in higher organisms the free Ca^{2+} concentration in external fluids and vesicles is extremely well controlled and in the extracellular fluid of humans is nearly in equilibrium with the precipitated bone. The further requirement for a messenger activity, that the rate of exchange to the internal proteins be fast, at least 10^3 s^{-1} for calcium, is also met.

Consider now a cytoplasmic concentration in a resting cell of Ca^{2+} of 10^{-7} M and an external or vesicle concentration of 10^{-3} M . Any influx of Ca^{2+} to the cytoplasm from outside through channels will increase Ca^{2+} to 10^{-6} M , which will trigger activity through cytoplasmic protein binding. On closing the channel, the outward pumps will restore rest conditions quickly. The calcium ion is, therefore, able to act as a pulsed messenger. Hence, the combination of sufficiency of binding, of selectivity, of concentration control, and of rates of on/off reactions gives optimal messenger character to this elementary ion. No other ion or molecule could be better.

Overall Functional Fitness of Calcium Ions in Evolution

The properties of the calcium ion, its fast on/off reactions reflecting the structural flexibility of its coordination sphere and its quite strong binding to selected organic ligands, have made it an extremely valuable partner to organic chemicals in almost all forms of life. In the most primitive forms of life, this value is seen not only in protective external constructs (walls) and mineralization, but also in the use of bound calcium in external digestive catalytic enzymes. These functions are found in all other organisms, but the most striking feature of the use of calcium appeared at the time the eukaryotes developed from the prokaryotes. At that time, the problem of coordinating activities of different compartments (vesicles) and the need for an increasing recognition of the environment became critical for the survival of the large eukaryote cells. Cooperation between compartments and sensing required the rapid transfer of information between cellular compartments and from the outside environment, and there is no better possible mode of communication in cellular biology than that which can be derived from the happy coincidental values of the high availability with concentration control and the response properties of the calcium ion. Hence, it was an evolutionary requirement

Table 1 Calcium-controlled events in cells

Activity	Controlled events or systems
Photosynthesis	Dioxygen release
Oxidative phosphorylation	Dehydrogenases
Receptor responses	Nerve synapse; IP ₃ -linked reactions
Contractile devices	Muscle triggering (actomyosin); cell filament controls (tubulins)
Phosphorylation	Activation of kinases, e.g., in fertilization
Metabolism	Numerous enzymes inside cells
Membrane/filament organization	Annexin-like proteins modulate tension
Cell division	S-100 proteins, immune system
Cell death (apoptosis)	Internal proteases
Hormone/transmitter release	Homeostasis
Binding to membranes	C2 domains of enzymes
Cross-linking	Outside cells
Enzyme activation	Outside cells; in membranes

that led to a messenger system in which the calcium ion was the best of all available carriers (Table 1). The protein receptors in the cytoplasm acted as mechanical triggers of action, changing conformation upon binding calcium. They include all muscle responses (troponins), fertilizations, many metabolic responses (calmodulins), genetic activations (possibly S-100 and other proteins), and immune responses (calcineurin). The necessary refinement of the homeostasis of the calcium concentrations increased all the way up to humans and included the control over both the internal cytoplasmic and the external circulating fluid concentrations. In the external fluids, there arose various novel functions that used the calcium ion, including further immune responses and blood-clotting controls in addition to digestive actions. Finally, we must remember that the messenger system from outside to inside the cell was augmented by the further responses, releases due to calcium pulses of chemicals, which caused release from internal vesicles and organelles to both the cytoplasm and the external fluids. Thus, the pulsed calcium ion input to the mitochondria and chloroplasts stimulates the release of energy (ATP) to the cytoplasm, and the release of this ion from vesicles to the external fluids stimulates the release not only of enzymes (e.g., for digestion, general to many organisms) but also of hormones and synaptic messengers, leading to overall homeostatic control. Hence, calcium signaling is central to a huge variety of communication modes leading up to and including synaptic action in the brain. In this homeostasis, we must not forget the buffer action of the bone. The overall fitness of calcium for a multitude of functions is very clear (see Figure 1).

See also: [Bioenergetics: Calcium Oscillations; Calcium Signaling: Calmodulin-Dependent Phosphatase; Calcium Transport in Mitochondria; ER/SR Calcium Pump: Function; ER/SR Calcium Pump: Structure; Ligand-Operated Membrane Channels: Calcium \(Glutamate\).](#)

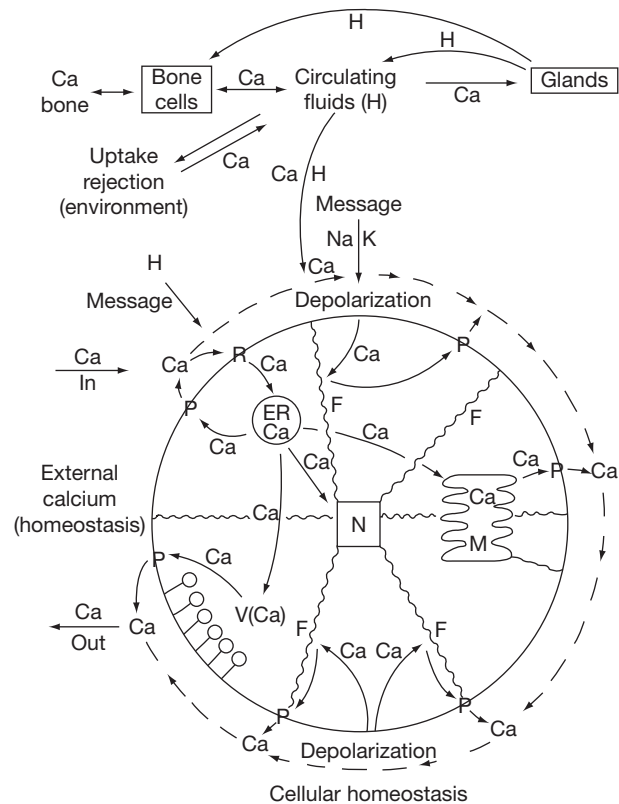


Figure 1 The generalized way in which calcium ions act as a messenger in advanced animal cells. The top of the figure shows the connection between the calcium in the external fluids, bone, and glands and uptake and rejection systems (e.g., the epithelial cells of digestion and the rejection cells of the kidney); H is a hormone. Calcium ions from the circulating fluids together with hormones and Na/K nerve messages activate a large range of cells. The calcium entry is controlled by channels R responding to perturbations. On entering the cell, calcium can release further calcium from vesicles such as the endoplasmic reticulum (ER). The calcium in the cytoplasm can act directly on filaments (F), on mitochondria (M) (or chloroplasts), through transcription factors on the nucleus (N), and on vesicles (V). The vesicles act to release transmitters, hormones, or enzymes. The calcium is removed by pumps (P). The bottom of the diagram shows the effect of general depolarization rather than specific channel activation. This allows calcium entry by electrically operated channels, as, for example, at nerve synapses of the heart.

Further Reading

- Carafoli E and Klee C (eds.) (1999) *Calcium as a Cellular Regulator*. New York: Oxford University Press.
- Carafoli E and Krebs J (eds.) (2000) *Calcium Homeostasis*. Berlin: Springer.
- Frausto da Silva JJR and Williams RJP (2001) *Chapters 10 and 20. The Biological Chemistry of the Elements*. Oxford: Oxford University Press.
- Poche R (ed.) (2000) *Calcium: The Molecular Basis of Calcium Action in Biology and Medicine*. Dordrecht: Kluwer.
- Vogel HJ (ed.) (2002) In: *Calcium-Binding Protocols*, vols. 1–2. Totowa, NJ: Humana Press.

Calcium-Buffering Proteins: Calbindin

L J Mady, F Haleem, and S Christakos, University of Medicine and Dentistry of New Jersey, Newark, NJ, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Willi Hunziker and Igor Bendik, volume 1, pp. 221–225, © 2004, Elsevier Inc.

Glossary

Apoptosis Form of programmed cell death mediated by intracellular proteolytic enzymes known as caspases.

Calbindin D_{9K} Calcium-binding protein of approximately 9000 molecular weight with two calcium binding domains, present in highest concentration in mammalian intestine.

Calbindin D_{28K} Calcium-binding protein of approximately 28 000 molecular weight with four calcium-binding sites, present in highest concentration in avian intestine and in avian and mammalian kidney, brain, and pancreas.

Knock-out mice Genetically engineered mice in which one or more genes have been inactivated through targeted deletions or mutations.

Vitamin D A secosteroid that is synthesized in the skin from 7-dehydrocholesterol by ultraviolet irradiation. Vitamin D is biologically inert and must undergo two successive hydroxylations in the liver and kidney to become the biologically active 1,25-dihydroxyvitamin D₃.

1,25-Dihydroxyvitamin D₃ [1,25(OH)₂D₃] Biologically active metabolite of vitamin D.

Introduction

The calbindins, calbindin-D_{9K} and calbindin-D_{28K}, belong to a family of high-affinity calcium-binding proteins that contains more than 300 members and are characterized by the EF hand structural motif (an octahedral structure consisting of two α -helices connected by a sequence of 12–14 amino acids that constitute a high-affinity calcium-binding site). Calbindin was the first identified target of vitamin D action. Calbindin-D_{28K}, a protein of approximately 28 000 molecular weight which has four high-affinity binding sites for calcium, is present in highest concentrations in avian intestine and in avian and mammalian kidney, brain, and pancreas. Calbindin-D_{28K} is highly conserved in evolution. Calbindin-D_{9K}, a protein of approximately 9000 molecular weight which has two calcium-binding domains, is present in highest concentrations in mammalian intestine and, unlike calbindin-D_{28K}, is not evolutionarily conserved. It is present only in mammals. There is no amino acid sequence homology between calbindin-D_{9K} and calbindin-D_{28K}. The characterization, distribution, functional significance, and regulation of these proteins as well as recent findings using calbindin knock-out mice are reviewed.

Calbindin-D_{28K}

Characterization and Tissue Distribution

Calbindin-D_{28K}, previously known as the vitamin D-dependent Ca²⁺-binding protein (CaBP), was initially discovered in avian intestine by Wasserman and Taylor in 1966. In 1985, the protein became officially known as calbindin-D_{28K}. Through examination of the protein's distribution in tissues and colocalization with the vitamin D receptor (VDR), significant advances have been made in our understanding of the vitamin D endocrine system.

Chicken and mammalian calbindin-D_{28K} are single-chain polypeptides containing 261 amino acid residues, have a

molecular weight of approximately 28 000 based on migration on sodium dodecyl sulfate-polyacrylamide gels, and contain blocking groups at the amino terminus. Calbindin-D_{28K} is highly conserved in evolution, with 98% homology in mammalian calbindin-D_{28K} sequences and 79% homology between mammalian and chicken sequences. Conservation of calbindin-D_{28K} sequences suggests its essential physiological role in mediating intracellular calcium-dependent processes.

Calbindin-D_{28K}, an acidic and heat-stable protein, is a member of the EF-hand protein family that binds Ca²⁺ with high affinity ($K_d = 10^{-8}$ – 10^{-6}). The EF-hand domain is a 29–30-amino-acid octahedral structure consisting of two α helices separated by a highly conserved sequence of 12 amino acids. The 12-amino-acid loop contains side chain oxygens essential for binding the divalent Ca²⁺ cation. Though calbindin-D_{28K} contains six EF-hand domains, only four actively bind Ca²⁺ (Figure 1).

The human calbindin-D_{28K} gene (symbol: *CALB1*) is located on chromosome 8, spans 24 kb, and consists of 11 exons.

Since its discovery in avian intestine, calbindin-D_{28K} has been identified in many other tissues including kidney and bone and in non-serum Ca²⁺-regulating tissues such as pancreas, testes, and brain. Table 1 summarizes the distribution of calbindin-D_{28K} in a variety of tissues and species.

Proposed Functional Significance

Intestine

Calbindin-D_{28K} in avian intestine (as well as calbindin-D_{9K} in mammalian intestine) are markedly induced by 1,25(OH)₂D₃ and are localized in the cytoplasm of absorptive cells, supporting a proposed role in intestinal calcium absorption (also see section on calbindin-D_{9K}). However, calbindin may also act in the intestine by buffering calcium, protecting against calcium-mediated cell death during 1,25(OH)₂D₃-dependent intestinal calcium transport.

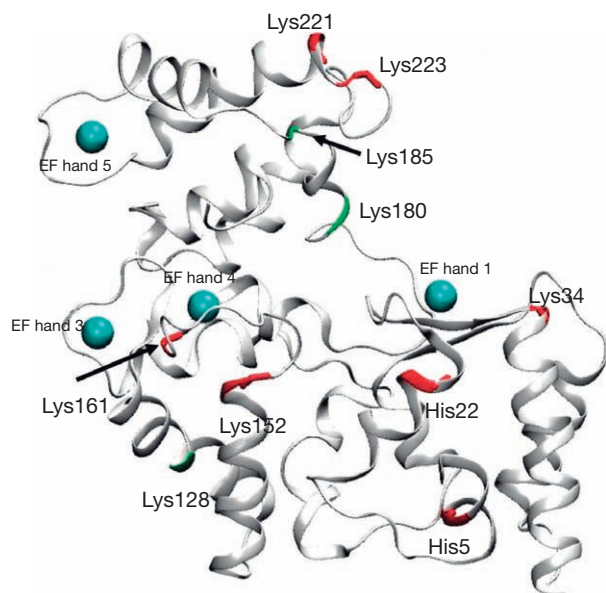


Figure 1 Computational model of disulfide-bonded calbindin-D_{28K}. Calbindin-D_{28K} is made up of six EF hand domains, four of which bind calcium. Calcium-binding EF hands are shown, with calcium ions shown as spheres. Mass spectrometry and nuclear magnetic resonance experiments have shown that EF hands 1, 3, 4, and 5 bind calcium and that calcium ions are sequentially bound with EF hand 1 first, followed by EF hand 4 and 5 and lastly EF hand 3. Reproduced from figure 7 in Hobbs CA (2009) Structural characterization of the conformational change in calbindin-D_{28K} upon calcium binding using differential surface modification analyzed by mass spectrometry. *Biochemistry* 48(36): 8603–8314, with permission from ACS.

Table 1 Distribution of calbindin-D_{28K}

Avian intestine
Avian, reptilian, amphibian, and mammalian kidney
Hen egg-shell gland (uterus)
Mouse reproductive tissues (uterus, oviduct, ovary)
Avian and mammalian beta cells of the pancreas
Alpha cells of the rat pancreas
Rat and chick growth cartilage
Ameloblasts and osteoblasts of rodent teeth; mouse osteoblasts
Brain (avian, reptilian, amphibian, molluscan, fish, and mammalian brain)

Kidney

In mammals, chickens, and reptiles, renal calbindin-D_{28K} is found exclusively in the cytosol, and to a lesser extent the nucleus, of distal convoluted tubule and connecting tubule cells of the distal nephron. It has been shown that the VDR is also predominantly found in the distal nephron. In the distal nephron, 1,25(OH)₂D₃ enhances PTH-mediated uptake of Ca²⁺ at the apical membrane of the distal tubule cell. Although 1,25(OH)₂D₃ has been reported to induce calbindin-D_{28K}, little is known regarding the exact role of vitamin D-inducible calbindin in Ca²⁺ transport in the distal nephron. In a process similar to intestinal Ca²⁺ absorption, renal transcellular Ca²⁺ transport involves Ca²⁺ entry through an apical Ca²⁺ channel, diffusion of Ca²⁺ through the cell interior, and active Ca²⁺ extrusion from the basolateral membrane through a

Ca²⁺-dependent adenosine triphosphatase (ATPase). As has been suggested for intestinal calbindin, renal calbindin-D_{28K} may act to shuttle Ca²⁺ through the cell interior to the basolateral membrane and to buffer Ca²⁺ and protect against Ca²⁺-mediated cell death.

The transient receptor potential vanilloid-subtype 5 (TRPV5) is an epithelial Ca²⁺ channel that acts as a gatekeeper of apical Ca²⁺ entry during active reabsorption. TRPV5 and calbindin-D_{28K} are co-expressed in the distal nephron and are co-regulated by 1,25(OH)₂D₃ and dietary Ca²⁺. Under conditions of low intracellular Ca²⁺, calbindin-D_{28K} directly interacts with TRPV5 at the apical membrane and has been shown to modulate the activity of the epithelial Ca²⁺ channel. These findings suggest an additional role for renal calbindin-D_{28K} as a modulator of calcium influx.

Bone

In rats and chicks, calbindin-D_{28K} has been found in chondrocytes of growth plate cartilage suggesting a possible role for calbindin-D_{28K} in cartilage calcification and chondrocyte maturation and differentiation. Calbindin-D_{28K} has also been localized to highly specialized cells of rodent teeth, including ameloblasts, odontoblasts, and osteoblasts. Through Ca²⁺ handling, calbindin-D_{28K} may be involved in the mineralization of tissues, enamel, dentine, and bone.

Through inhibition of caspase-3, a key mediator of apoptosis, calbindin-D_{28K} has also been shown to protect against tumor necrosis factor (TNF)-induced cell death in osteoblastic cells and glucocorticoid-induced cell death in osteoblastic and osteocytic cells. The ability of calbindin-D_{28K} to inhibit caspase-3 is independent of its Ca²⁺-binding characteristics. As a natural non-oncogenic inhibitor of caspase-3, calbindin-D_{28K} may be an important target in the therapeutic prevention of bone cell degeneration.

Pancreas

Calbindin-D_{28K} has been identified in human pancreatic islet cells in addition to chick insulin-producing beta cells and alpha and beta cells of the rat pancreas. It has been suggested that calbindin-D_{28K} may modulate depolarization-stimulated insulin release by regulating intracellular Ca²⁺. Further studies have indicated that calbindin-D_{28K}, by acting as a Ca²⁺ buffer, can prevent Ca²⁺-mediated mitochondrial damage and subsequent free radical formation. In this way, calbindin-D_{28K} protects against cytokine-mediated destruction of beta cells. These findings have important implications in the treatment of type 1 diabetes and the prevention of autoimmune destruction of pancreatic beta cells. Calbindin-D_{28K} has also been shown to regulate Ca²⁺ influx in pancreatic beta cells via L-type voltage-dependent Ca²⁺ channels. In rat pancreatic beta cells, the alpha₁ subunit of the L-type channel can interact with calbindin-D_{28K} and L-type channels show enhanced sensitivity to Ca²⁺-dependent inactivation in the presence of calbindin-D_{28K}. Together, these findings indicate that pancreatic calbindin-D_{28K} is involved in intracellular Ca²⁺ homeostasis as well as modulation of Ca²⁺ influx.

Reproductive tissues

Calbindin-D_{28K} has been localized in spermatogonia and spermatocytes of the seminiferous tubules and some interstitial

Leydig cells in chick and rat testes. The involvement of calbindin-D_{28K} in spermatogenesis has been suggested.

In addition, calbindin-D_{28K} is also localized in the tubular gland cells of the chick shell gland which are involved in Ca²⁺ secretion during formation of the egg shell.

In female mice (but not rats), calbindin-D_{28K} is present in reproductive tissues including the endometrium and glandular epithelium of the uterus, the oviduct epithelium, and primary follicles of the ovary. Calbindin is not regulated by 1,25(OH)₂D₃ in these tissues. Estradiol has been shown to regulate calbindin-D_{28K} in the uterus. Although the exact role of uterine calbindin-D_{28K} is not known, it has been suggested that calbindin-D_{28K} may act to facilitate transcellular Ca²⁺ transport in epithelial cells of the uterus.

Nervous tissue

In a wide variety of species, including mammals, avians, reptiles, amphibians, fish, and mollusks, calbindin-D_{28K} in the brain is vitamin D independent and is found in most neuronal cell groups and in some epithelial cells lining the ventricular system. Neurons containing calbindin-D_{28K} are found in layers 2–4 of the cerebral cortex, the hippocampus, the cerebellum (specifically in Purkinje cells) as well as the hypothalamus, amygdala, pyriform region, and thalamus. Furthermore, calbindin-D_{28K} is found in specific neuronal sensory cells of the visual and auditory systems. These cells include cochlear and vestibular hair cells of the inner ear, avian basilar papilla, photoreceptor cells (cone but not rod cells) of avian and mammalian retina, and in pineal transducers (conelike, modified photoreceptor cells (pinealocytes)).

It has been suggested that neuronal calbindin, by buffering Ca²⁺, can limit Ca²⁺ influx and regulate intracellular Ca²⁺ responses to physiological stimuli. Through maintenance of Ca²⁺ homeostasis, calbindin-D_{28K} can protect neurons against calcium-mediated neurotoxicity and modulate synaptic transmission. Studies examining the effect of calbindin-D_{28K} overexpression, either by introduction of exogenous calbindin-D_{28K} via the whole-cell patch-clamp method or by transfection and overexpression of calbindin-D_{28K} have demonstrated that calbindin-D_{28K} can modify Ca²⁺ signals to alter synaptic plasticity and neuronal firing patterns.

The role of calbindin-D_{28K} in influencing physiological responses is further suggested in studies examining the expression of calbindin-D_{28K} in mouse suprachiasmatic nucleus (SCN). In mammals, the SCN is the fundamental regulator of circadian function and entrainment. In the mouse SCN, calbindin-D_{28K} is markedly expressed during development and is co-expressed in adult melanopsin-producing retinal ganglion cells. In calbindin-D_{28K}-null mutant mice significant abnormalities are observed in circadian locomotor rhythmicity and entrainment to light/dark cycles, though response to acute light remains normal. These findings suggest a role for calbindin-D_{28K} in circadian organization and light transduction.

Studies of ischemic injury, seizure activity, and chronic neurodegeneration (Alzheimer's, Huntington's, and Parkinson's diseases) have suggested a correlation between decreased neuronal calbindin-D_{28K} and neurodegeneration. With reduced levels of calbindin, neurons may be unequipped to buffer excess intracellular Ca²⁺, resulting in calcium-mediated neuronal damage and cell death. In primary cultures of neuronal

cells and in neuronal cell lines transfected with the calbindin-D_{28K} gene, calbindin-D_{28K} has been shown to increase the survival of cells injured in response to hypoglycemia and amyotrophic lateral sclerosis IgG-mediated cytotoxicity. In rats, it has been shown that pretreatment with a calbindin-D_{28K} fusion protein (protein transduction domain (PTD)-calbindin-D_{28K}) attenuates neuronal insults induced by ischemia and reperfusion. PTDs are small peptides that serve to deliver larger molecules into a variety of cells, including those of the blood-brain barrier, without employing traditional endocytic mechanisms. PTD-fused proteins have been shown to retain their biologic activity and thus represent a potentially novel approach for the treatment of neuropathies, including ischemic injury.

Calbindin-D_{28K} has also been shown to prevent apoptosis in neural cells expressing mutant presenilin 1 (PS-1), which is causally linked to about 50% of cases in early-onset familial Alzheimer's disease. Amyloid β peptide (A β) comprises the major component of plaques in Alzheimer's disease and has been reported to induce neuronal damage in cells sensitized to apoptosis by mutant PS-1 via oxidative stress mechanisms and disruption of calcium balance. Calbindin-D_{28K} protects against the proapoptotic action of mutant PS-1 by restoring intracellular Ca²⁺ homeostasis, thereby maintaining mitochondrial integrity and function.

Regulation

Regulation by 1,25(OH)₂D₃

1,25(OH)₂D₃ is known to induce calbindin-D_{28K} in the avian intestine and kidney and in mammalian kidney. Only modest responses in the activities of the chicken and mouse calbindin-D_{28K} promoters to 1,25(OH)₂D₃ have been observed. Such findings reflect the small transcriptional response of the calbindin-D_{28K} gene to 1,25(OH)₂D₃. However, calbindin-D_{28K} messenger RNA (mRNA) is largely induced long after 1,25(OH)₂D₃ treatment, suggesting mostly posttranscriptional mechanisms for the accumulation of calbindin-D_{28K} mRNA. These findings illustrate the complexity of 1,25(OH)₂D₃ action on calbindin-D_{28K} and suggest that a model beyond the conventional hormone receptor-transcriptional activation model is needed to understand calbindin-D_{28K} regulation.

Regulation by other steroids and factors

Studies have shown that calbindin-D_{28K} gene expression may be regulated by factors other than 1,25(OH)₂D₃. Vitamin D-treated chicks that are given glucocorticoids show decreased levels of calbindin-D_{28K} mRNA and protein in intestine and inhibition of intestinal Ca²⁺ absorption (mammalian calbindin-D_{9K} in intestine is also downregulated by glucocorticoids; see section in calbindin-D_{9K}). Alterations in dietary Ca²⁺ and phosphorus have also been shown to modulate avian intestinal and renal calbindin-D_{28K} gene expression.

Studies have suggested that gender-specific hormones may also modulate calbindin-D_{28K} expression. As previously discussed, calbindin-D_{28K} is found in the avian egg-shell gland and in the reproductive tissues of female mice. In *in vivo* experiments examining the avian egg-shell gland, estradiol is shown to induce calbindin-D_{28K}. By contrast, in female mice, calbindin-D_{28K} gene expression is downregulated by estradiol in the uterus. The differing effects of estradiol in chick and mouse suggest

species-specific regulation of calbindin-D_{28K} by estradiol. Gender differences have been reported to influence renal handling of Ca²⁺ as male mice display greater urinary Ca²⁺ loss than female mice. Male mice also show significantly decreased levels of TRPV5 and calbindin-D_{28K} compared to female mice. Furthermore, androgen-deficient male mice show increased renal TRPV5 and calbindin-D_{28K} mRNA and protein. This increase is suppressed by testosterone treatment and is not accompanied by significant differences in serum estrogen, parathyroid hormone, or 1,25(OH)₂D₃ levels. These findings suggest that the inhibitory actions of androgens on TRPV5 and calbindin-D_{28K} may, in part, explain gender differences in renal Ca²⁺ reabsorption.

In the central nervous system, where calbindin-D_{28K} is detected in significant amounts, calbindin-D_{28K} regulation is independent of 1,25(OH)₂D₃. In contrast to the action of 1,25(OH)₂D₃ on intestinal and renal calbindin-D_{28K} gene expression, 1,25(OH)₂D₃ does not affect levels of neuronal calbindin-D_{28K}. On the other hand, TNFs and neurotropic factors including neurotrophin 3 (NT-3), brain-derived neurotrophic factor (BDNF), and fibroblast growth factor (FGF), have been shown to induce calbindin-D_{28K} in rat hippocampal cultures. The induction of calbindin-D_{28K} by these factors may be, in part, a mechanism against cytotoxic injury, as neurotropic factors have been shown to protect against excitotoxic neuronal damage. The cell-specific regulation of neuronal calbindin-D_{28K} by a variety of factors is further supported by studies showing induction of calbindin-D_{28K} by corticosterone in the CA1 region of rat hippocampus, retinoic acid in medulloblastoma cells that pose a neuronal phenotype, insulin-like growth factor I (IGF-I) in cultured Purkinje cells, and by IFG1 and insulin in rat embryonic neuronal cells.

Though calcium-binding proteins were the first identified targets of 1,25(OH)₂D₃ action, studies examining the complexity in calbindin-D_{28K} regulation illustrate that its utility extends beyond the traditional paradigm of the vitamin D endocrine system.

Calbindin-D_{28K} Knockout Mice

The phenotype of calbindin-D_{28K}-null mutant mice is impaired motor coordination associated with alterations in Ca²⁺ signaling in Purkinje cells. Calbindin-D_{28K}(-/-) mice show normal levels of serum-ionized calcium, phosphorus, and PTH, and normal urinary calcium and phosphorus. Histological examination reveals normal bone structure. In order to further determine the *in vivo* functional significance of mammalian calbindin-D_{28K} VDR/calbindin-D_{28K} double KO mice were generated. Phenotypic characteristics of mice deficient only in VDR include secondary hyperparathyroidism, rickets, and osteomalacia, and renal calbindin-D_{9K} is reduced to 10% of typical levels (thus, in these mice, calbindin-D_{9K} would not compensate for the loss of calbindin-D_{28K}). Compared to VDR KO mice, VDR/calbindin-D_{28K} double KO mice experience a higher loss of urinary calcium and as a result, more pronounced effects of calcium dysregulation including severe hyperparathyroidism, greater bone resorption, lower body mass index (BMD), and increased osteoid accumulation. For both VDR KO and double KO mice, a high Ca²⁺ diet can restore normal levels of serum ionized and urinary calcium. However, unlike in VDR KO mice, high calcium supply cannot

completely rescue the skeletal abnormalities in double KO mice. These findings provide *in vivo* evidence for the importance of calbindin-D_{28K} in maintaining calcium homeostasis.

Studies were also done comparing TRPV5(-/-), calbindin-D_{28K}(-/-), and TRPV5/calbindin-D_{28K} double KO mice. Double KO mice show hypercalciuria similar to TRPV5(-/-) mice, though calbindin-D_{28K}(-/-) mice do not experience comparable urinary Ca²⁺ loss. TRPV5(-/-) and double KO mice also exhibit similar increases in renal calbindin-D_{9K} expression and upregulation of duodenal calbindin-D_{9K} and TRPV6, resulting in increased intestinal Ca²⁺ absorption (to compensate, at least in part, for the urinary Ca²⁺ loss). Calbindin-D_{28K}(-/-) mice do not exhibit alterations in intestinal Ca²⁺ absorption and expression of calbindin-D_{9K} and TRPV6 remain unchanged. The similarities seen in TRPV5(-/-) and TRPV5/calbindin-D_{28K} double KO mice provide evidence that TRPV5-facilitated Ca²⁺ entry in the distal convoluted tubule may be the rate-limiting step in active Ca²⁺ reabsorption. Loss of calbindin-D_{28K}, on the other hand, may be compensated by upregulation of calbindin-D_{9K}.

Calbindin-D_{28K} Summary

Calbindin-D_{28K} is highly conserved in evolution, suggesting a fundamental role for calbindin-D_{28K} in mediating intracellular calcium-dependent processes. Calbindin-D_{28K} is regulated by 1,25(OH)₂D₃ in avian intestine and in avian and mammalian kidney. Calbindin-D_{28K} is present in numerous other tissues including brain, pancreas, and uterus and is regulated by factors other than 1,25(OH)₂D₃. Evidence suggests that calbindin-D_{28K} is involved in protection against apoptotic cell death by buffering calcium and by inhibiting caspase-3. Additional functions of calbindin-D_{28K} include maintenance of intracellular calcium homeostasis and modulation of calcium influx.

Calbindin-D_{9K}

Characterization and Tissue Distribution

Kallfelz, Taylor, and Wasserman first identified a vitamin D regulated 9000-molecular-weight calcium-binding protein in rat intestine in 1967 (which was later called 'calbindin-D_{9K}'). It is a cytosolic calcium-binding protein found only in mammals. Calbindin-D_{9K} contains 79 amino acid residues (9016 molecular weight based on amino acid sequence), is acidic, and heat stable. There is 78% and 74% homology of the human calbindin-D_{9K} sequence to that of the rat and mouse respectively. Calbindin-D_{9K} contains two EF hand domains of 29 and 31 amino acids. Each domain binds one calcium. It is related to the S100 family of calcium-binding proteins. The human calbindin-D_{9K} gene (symbol: *CALB3*) spans 5.5 kb and consists of three exons. The chromosomal localization of the human calbindin-D_{9K} gene is assigned to Xp22.2.

Calbindin-D_{9K} is most abundant in mammalian intestine. It is found in lower concentrations in mouse kidney and neonatal rat kidney (in the distal nephron), placenta of rats and mice, in rat uterus, rat lung and in ameloblasts and osteoblasts of rodent teeth. Table 2 summarizes the distribution of calbindin-D_{9K} in a variety of tissues and species.

Table 2 Distribution of calbindin-D_{9k}

Mammalian intestine
Mouse and neonatal rat kidney
Rat and mouse yolk sac
Rat uterus
Rat and mouse placenta
Rat growth cartilage
Ameloblasts and osteoblasts of rodent teeth
Rat lung

Proposed Functional Significance

Intestine and calbindin-D_{9k} KO mice

Similar to studies of calbindin-D_{28k} in avian intestine, it is thought that intestinal calbindin-D_{9k}, which is markedly induced by 1,25(OH)₂D₃, acts to facilitate the diffusion of calcium through the cell interior toward the basolateral membrane. However, recent studies using calbindin-D_{9k} KO mice have shown that calbindin-D_{9k} is not required for 1,25(OH)₂D₃-induced active intestinal calcium transport. Active intestinal calcium transport is similarly induced in response to a low calcium diet or 1,25(OH)₂D₃ treatment in both calbindin-D_{9k} KO mice and wild-type mice. In addition, calbindin-D_{9k} KO mice are able to maintain normal serum calcium levels regardless of age or sex. These findings suggest that calbindin-D_{9k} may be compensated for by another factor or that calbindin-D_{9k} may have another role in intestine, for example, as a modulator of the activity of TPRV6 (the vitamin D inducible epithelial calcium channel present in the intestine) and/or as an intracellular calcium buffer to prevent toxic levels of calcium from accumulating in the intestine.

Reproductive tissues

Calbindin-D_{9k} is present in placenta of rats and mice and in rat endometrium and myometrium. Since calbindin-D_{9k} in placenta increases at the end of gestation when there is increased calcium need by the fetus, it has been suggested that calbindin-D_{9k} may have a role in transporting calcium from mother to fetus. The presence of calbindin-D_{9k} in myometrium suggests the involvement of calbindin in modulation of intracellular calcium which may alter the frequency and strength of uterine contractions. Since calbindin-D_{9k} KO mice reproduce similar to wild-type mice and calbindin-D_{9k} KO pups are indistinguishable from wild-type pups, the absence of calbindin-D_{9k} in placenta and myometrium may be compensated for by another factor.

Regulation

Similar to avian intestinal calbindin-D_{28k}, intestinal calbindin-D_{9k} is regulated by 1,25(OH)₂D₃ by a small rapid transcriptional response followed by a posttranscriptional effect. Besides VDR, it has been shown that the intestine-specific transcription factor caudal homeobox-2 (Cdx-2) is also important for the transcription of the calbindin-D_{9k} gene.

In addition to 1,25(OH)₂D₃, glucocorticoids also regulate intestinal calbindin-D_{9k}. *In vivo* treatment of mice with glucocorticoids reduces the expression of duodenal calbindin-D_{9k}

mRNA. The reduction in calbindin-D_{9k} mRNA is associated with diminished intestinal calcium absorption. It remains to be determined whether the *in vivo* effects of glucocorticoids on intestinal calbindin-D_{9k} are due to a direct effect of glucocorticoids on the calbindin-D_{9k} gene.

In the rat uterus, calbindin-D_{9k} is induced by estradiol but is unaffected by 1,25(OH)₂D₃ treatment. An estrogen responsive element that mediates the transcriptional regulation of the calbindin-D_{9k} gene has been identified, indicating a direct effect of estradiol on the calbindin-D_{9k} gene.

Calbindin-D_{9k} Summary

Calbindin-D_{9k} is present only in mammals. It is not homologous to calbindin-D_{28k} but rather belongs to the S100 family of calcium-binding proteins. Calbindin-D_{9k} is regulated by 1,25(OH)₂D₃ in intestine and kidney. Although calbindin-D_{9k} is present in other tissues (uterus, placenta, and bone), it has a more restricted tissue distribution compared to calbindin-D_{28k}. Studies in calbindin-D_{9k} KO mice suggest that calbindin-D_{9k} may be compensated for by another factor yet to be identified.

Conclusion

Calbindin-D_{28k} and calbindin-D_{9k} can no longer be viewed exclusively as vitamin D-dependent targets. Regulatory control of calbindin is complex and involves factors other than 1,25(OH)₂D₃, providing evidence that calbindins play a role beyond the traditional model of the vitamin D endocrine system. Calbindin-D_{28k} and calbindin-D_{9k} are found in variety of tissues and species (see [Tables 1](#) and [2](#)) and demonstrate many different tissue-specific functions. Recent studies have provided new insight into the mechanism of action of the calbindins, including the maintenance of intracellular calcium homeostasis, modulation of calcium influx, and protection against apoptosis.

See also: [Bioenergetics: Calcium-Modulated Proteins \(EF-Hand\); Metabolism Vitamins and Hormones: Vitamin D.](#)

Further Reading

- Akhter S, Kutzova GD, Christakos S, and DeLuca HF (2007) Calbindin D9k is not required for 1,25-dihydroxyvitamin D₃-mediated Ca²⁺ absorption in small intestine. *Archives of Biochemistry and Biophysics* 460: 227–232.
- Benn BS, Ajibade D, Porta A, et al. (2008) Active intestinal calcium transport in the absence of transient receptor potential vanilloid type 6 and calbindin-D_{9k}. *Endocrinology* 149: 3196–3205.
- Fan Y, Shi L, Gu Y, et al. (2007) Pretreatment with PTD-calbindin D 28k alleviates rat brain injury induced by ischemia and reperfusion. *Journal of Cerebral Blood Flow and Metabolism* 27: 719–728.
- Gkika D, Hsu YJ, van der Kemp AW, et al. (2006) Critical role of the epithelial Ca²⁺ channel TRPV5 in active Ca²⁺ reabsorption as revealed by TRPV5/calbindin-D_{28k} knockout mice. *Journal of the American Society of Nephrology* 17: 3020–3027.
- Ho BK, Alexianu ME, Colom LV, et al. (1996) Expression of calbindin-D_{28k} in motoneuron hybrid cells after retroviral infection with calbindin-D_{28k} cDNA prevents amyotrophic lateral sclerosis IgG-mediated cytotoxicity. *Proceedings of the National Academy of Sciences of the United States of America* 93: 6796–6801.

- Hsu YJ, Dimke H, Schoeber JP, et al. (2010) Testosterone increases urinary calcium excretion and inhibits expression of renal calcium transport proteins. *Kidney International* 277: 601–608.
- Kreiner L, Christel CJ, Benveniste M, Schwaller B, and Lee A (2010) Compensatory regulation of $\text{Ca}_v2.1$ Ca^{2+} channels in cerebellar Purkinje neurons lacking parvalbumin and calbindin D-28k. *Journal of Neurophysiology* 103: 371–381.
- Kriegsfeld LJ, Mei DF, Yan L, et al. (2008) Targeted mutation of the calbindin D28K gene disrupts circadian rhythmicity and entrainment. *European Journal of Neuroscience* 27: 2907–2921.
- Kutuzova GD, Akhter S, Christakos S, et al. (2006) Calbindin D_{9k} knockout mice are indistinguishable from wild-type mice in phenotype and serum calcium level. *Proceedings of the National Academy of Sciences of the United States of America* 103: 12377–12381.
- Lee D, Obukhov AG, Shen Q, et al. (2006) Calbindin-D28k decreases L-type calcium channel activity and modulates intracellular calcium homeostasis in response to K^{+} depolarization in a rat beta cell line RINr1046-38. *Cell Calcium* 39: 475–485.
- Liu Y, Porta A, Peng X, et al. (2004) Prevention of glucocorticoid-induced apoptosis in osteocytes and osteoblasts by calbindin-D28k. *Journal of Bone and Mineral Research* 19: 479–490.
- Schwaller B (2009) The continuing disappearance of “pure” Ca^{2+} buffers. *Cell and Molecular Life Sciences* 66: 275–300.
- Zheng W, Xie Y, Li G, et al. (2004) Critical role of calbindin-D28k in calcium homeostasis revealed by mice lacking both vitamin D receptor and calbindin-D28k. *Journal of Biological Chemistry* 279: 52406–52413.

Calcium-Buffering Proteins: ER Luminal Proteins

J Groenendyk and M Michalak, University of Alberta, Edmonton, AB, Canada

© 2013 Elsevier Inc. All rights reserved.

Glossary

Calreticulin Multifunctional Ca^{2+} binding/buffering endoplasmic reticulum (ER) resident chaperone. The protein is responsible for the buffering of over 50% of ER luminal Ca^{2+} and for assisting in the folding of newly synthesized glycoproteins.

Chaperones Molecules that bind to misfolded proteins and assist in their folding and posttranslational modification.

Endoplasmic reticulum (ER) Intracellular organelle involved in many cellular functions including modulation of Ca^{2+} homeostasis, protein synthesis and modification, and lipid synthesis.

KDEL This is a sequence in the amino acid structure of a protein which keeps it from secreting from the ER. It also targets proteins from other locations (such as the cytoplasm) to the ER. Proteins can only leave the ER after this sequence has been cleaved off. The KDEL sequence is responsible for retrieval of ER luminal proteins from the Golgi apparatus. The abbreviation KDEL is formed by the corresponding letters to each amino acid. This letter system was defined by the IUPAC and IUBMB in 1983, and is as follows: K, lysine; D, aspartic acid; E, glutamic acid; L, leucine.

Protein disulfide isomerase Belonging to a family of proteins containing thioredoxin, they modify and catalyze the formation of disulfide bonds in newly synthesized proteins.

Ca^{2+} buffering endoplasmic reticulum (ER) proteins are ER resident proteins that bind Ca^{2+} with a high capacity. They play an important role in the maintenance of resting free Ca^{2+} concentration in the lumen of ER and subsequently in the modulation of cellular Ca^{2+} homeostasis. This is critical because Ca^{2+} is a universal intracellular signaling molecule, which controls numerous developmental and cellular pathways. Changes in Ca^{2+} concentration affect processes as diverse as learning and memory, secretion, contraction and relaxation, membrane excitability, cell motility, cytosolic and mitochondrial metabolism, protein and lipid synthesis, cell cycle, apoptosis, organellar trafficking, and protein folding and quality control. Therefore, it is of the utmost importance that Ca^{2+} homeostasis is tightly regulated in all cells. This is accomplished by a variety of Ca^{2+} transporting, binding, and buffering proteins, which are found in many cellular compartments.

The ER and Ca^{2+} Homeostasis

The ER is one of the most important intracellular Ca^{2+} buffering and storage organelles and performs a vital function during Ca^{2+} signaling with communication to the plasma membrane Ca^{2+} channels, controlling cytosolic Ca^{2+} concentration and fluctuations. Moreover, persistent oscillation of ER luminal Ca^{2+} concentration functions as a signaling event for several ER functions, including protein and lipid synthesis. In response to a variety of external stimuli, Ca^{2+} is released from the lumen of the ER into the cytosol, via the inositol-1,4,5-triphosphate receptor (InsP_3R) and/or the ryanodine receptor (RyR) Ca^{2+} channels. Some Ca^{2+} also enters the cytosol from the extracellular environment, via Ca^{2+} channels in the plasma membrane. Most of the released Ca^{2+} is taken up back into the lumen of the ER via the sarcoplasmic-endoplasmic-reticulum Ca^{2+} -adenosine triphosphatase (ATPase) (SERCA). Some of

the Ca^{2+} is also removed from the cytosol by the plasma membrane Ca^{2+} -ATPase and the Na^+ - Ca^{2+} exchanger. Within the ER, Ca^{2+} is bound to and buffered by high concentrations of luminal Ca^{2+} -binding proteins. A decrease in ER Ca^{2+} levels leads to interrupted trafficking from the ER, delayed transport of molecules across the nuclear pore, and repressed chaperone function. The resting free cytosolic Ca^{2+} concentration has to be sustained at an extremely low level (sub-micromole), so that only tiny fluctuations (microdomains) in local Ca^{2+} levels can trigger a signaling event. This is achieved by actively pumping Ca^{2+} from the cytosol, either into the ER or out of the cell. Extracellular Ca^{2+} concentration is in excess of 2 mM, two orders of magnitude higher than free cytosolic Ca^{2+} concentration, which is about 100 nM, while the free ER Ca^{2+} concentration is approximately 200 μM , with total ER Ca^{2+} concentration up to 1 mM. Ca^{2+} homeostasis and signaling are maintained by controlling Ca^{2+} release from the ER by the InsP_3R and RyR , whereas the ER stores are refilled by the SERCA. Ca^{2+} is also transported out of the cytosol into the extracellular fluid via the sodium- Ca^{2+} exchanger (NCX) and the plasma membrane Ca^{2+} -ATPase (PMCA). Ca^{2+} -binding and buffering proteins play an important role in the translation, amplification, and protraction of Ca^{2+} signals, performing essential negative and positive feedback roles during Ca^{2+} trafficking and ultimately facilitating cells to generate rapid changes in local Ca^{2+} levels. The Ca^{2+} present in the ER stores serves as a source of easily releasable Ca^{2+} , and is also important as a regulator of a number of ER enzymes and proteins, including feedback regulation of the InsP_3R , the RyR , and SERCA. Coordination of ER luminal Ca^{2+} release and extracellular Ca^{2+} influx into the cytosol is critical as sustained cytosolic Ca^{2+} levels are cytotoxic and lead to apoptosis. Overall, it is clear that the ER may be defined as a multifunctional organelle able to detect and integrate incoming signals, modulate its own luminal dynamics, and generate output signals in response to environmental changes.

Ca²⁺ Buffering in the ER Lumen

The lumen of the ER contains spatially and temporally localized Ca²⁺ microdomains involved in downstream Ca²⁺ signaling. This is characterized by the heterogeneity in the localization of Ca²⁺ channels, including the InsP₃R, as well as Ca²⁺ buffering proteins, such as calsequestrin, in the lumen of the ER. This was discovered by using a luminal Ca²⁺ sensor such as ER targeted recombinant aequorin as well as electron microscopy, demonstrating an electron-dense network of calsequestrin in the terminal ends of the cisternae. Free luminal concentrations of Ca²⁺ range from 5 mM to 100 μM with spatial organization of intracellular Ca²⁺ appearing to depend on the strategic distribution of ER Ca²⁺ stores. These may be specific domains that trigger Ca²⁺ puffs, spikes, waves, and oscillations, involved in specific cellular functions including excitation-contraction and Ca²⁺ signaling. For example, the clustering of IP₃R in the membrane of the ER is suggested to regulate the propagation of Ca²⁺ waves with global repositioning of InsP₃R observed during long-term agonist stimulation in a smooth muscle cell line. In addition, activation of neurotrophils triggers whole cell shape change accompanied by dynamic relocation of ER Ca²⁺ stores. Recent work demonstrates that protein-protein interactions are chiefly responsible for the clustering of the InsP₃R in rat basophilic leukemia cells. In total, the lumen of the ER contains 1–3 mM Ca²⁺, a concentration similar to that in the extracellular space. A small portion of this ER luminal Ca²⁺ is not buffered by Ca²⁺ binding to proteins and is considered free. Resting free Ca²⁺ concentrations in the ER vary from 100 to 400 μM, while those in the cytosol are in the nanomolar range. The actual resting free Ca²⁺ concentration in the ER is determined by the balance between the rates of Ca²⁺ uptake and Ca²⁺ leak and by the extent of buffering by the resident Ca²⁺-binding proteins.

Most of the Ca²⁺ in the lumen of the ER is bound to the Ca²⁺-binding molecular chaperones that are resident there. These proteins maintain specific concentration gradients of Ca²⁺ in the ER and localize Ca²⁺ to the sites of release. They also directly affect both the Ca²⁺ storage capacity of the ER and the free Ca²⁺ concentration under resting conditions and after cellular stimulation. Table 1 shows a list of ER proteins that are involved in buffering Ca²⁺ in the ER lumen. There are two major classes of ER Ca²⁺-binding proteins. Class I, the larger of the two, represents proteins which bind Ca²⁺ with relatively high capacity and low affinity. These proteins contain a large proportion of acidic amino acid residues that are involved in the high-capacity and low-affinity Ca²⁺ binding. Their Ca²⁺-binding capacity varies from as little as 2 mol of Ca²⁺ per mole of protein to as much as 50 mol of Ca²⁺ per mole of protein. The high-capacity Ca²⁺-binding sites have a relatively low affinity for Ca²⁺ ($K_d = 1\text{--}2\text{ mM}$), which parallels the relatively high content of Ca²⁺ in the ER lumen. These Ca²⁺-binding proteins are responsible for Ca²⁺ storage in the lumen of the ER/sarcoplasmic reticulum (SR), and they effectively determine the capacity of these intracellular organelles for Ca²⁺. In contrast, class II represents Ca²⁺-binding proteins in the lumen of the ER that bind Ca²⁺ with relatively high affinity and low capacity. The function of these high-affinity-binding sites is

Table 1 Ca²⁺-binding proteins of the endoplasmic reticulum lumen

Protein name	Ca ²⁺ binding (mole/mole of protein)	Gene knockout mouse phenotype
<i>Class I proteins</i> (high-capacity and low-affinity Ca ²⁺ proteins)		
Calreticulin	25	Embryonic lethal E13.5
Grp 94/endoplasmic	10	Embryonic lethal E7.5
Grp 78/BiP	Binds Ca ²⁺	Embryonic lethal E3.5
PDI family		
PDI	20	Not investigated
ERp72	12	Not investigated
ERCalcistorin/PDI	20	Not investigated
P5	Binds Ca ²⁺	Not investigated
ERp57	Not investigated	Embryonic lethal E13.5
ERp44	Not investigated	Not investigated
ERp29	Not investigated	Not investigated
Calsequestrin	50	Not lethal – disrupted Ca ²⁺ homeostasis in muscle
<i>Class I/II protein</i> (contains EF-hand Ca ²⁺ -binding motif but binds with low affinity)		
Stim1	$K_d = 0.6\text{ mM}$ (low affinity)	Neonatal lethality
<i>Class II proteins</i> (contain EF-hand Ca ²⁺ -binding motifs and bind 6–7 mol of Ca ²⁺ /mole of protein with high affinity)		
Reticulocalbin	High affinity	Not investigated
Cab45	High affinity	Not investigated
Calumenin	High affinity	Not investigated
ERC55	High affinity	Not investigated
Crocalbin/CBP-50	High affinity	Not investigated

not obvious, given the high concentrations of Ca²⁺ in the lumen of the ER. It is thought that they might play a structural role, or that they might be responsible for the regulation/maintenance of specific protein-protein interactions. Several of the class II proteins contain a classic EF-hand Ca²⁺-binding site, previously identified in many cytosolic proteins that bind Ca²⁺ with high affinity.

A characteristic common to proteins in both classes is that they contain an N-terminal signal sequence that targets them to the lumen of the ER. They also terminate with a KDEL (K, lysine; D, aspartic acid; E, glutamic acid; L, leucine) like amino acid sequence that is responsible for their continual retrieval back to the ER lumen. Finally, most of the proteins resident in the lumen of the ER are multifunctional, regardless of their Ca²⁺-binding properties. They are involved in maintaining Ca²⁺ homeostasis and they assist in both protein folding and assembly, and lipid synthesis and transport.

Class I Ca²⁺-Binding Proteins in the ER Lumen

The class I ER luminal proteins are those that bind Ca²⁺ with high capacity and low affinity. In most of these proteins, specific regions of the amino acid sequence contain a high proportion of acidic residues, and it is these acidic residues

that are involved in the high-capacity Ca^{2+} binding. The following ER proteins belong to class I: calreticulin, glucose-regulated protein (Grp94), immunoglobulin binding protein (BiP), the protein disulfide isomerase (PDI) family, and calsequestrin.

Calreticulin

Calreticulin (46-kDa) is a major Ca^{2+} storage protein in the ER lumen, which contains both high-affinity/low-capacity and low-affinity/high-capacity Ca^{2+} -binding sites. Specifically, it binds up to 25 mol of Ca^{2+} per mole of protein with $K_d = 1 \text{ mM}$ (low affinity) and 1 mol of Ca^{2+} per mole of protein with a $K_d = 1 \text{ }\mu\text{M}$ (high affinity). Calreticulin, one of the most extensively studied ER luminal proteins, is responsible for binding approximately 50% of the Ca^{2+} stored in the lumen of the ER. As observed for other ER luminal proteins, the expression of calreticulin is upregulated by stress, Ca^{2+} depletion, and metabolic starvation. The over-expression of calreticulin in the ER significantly affects the ER Ca^{2+} storage capacity, Ca^{2+} uptake via SERCA, and release via InsP_3 receptor, indicating an important role for calreticulin in ER Ca^{2+} buffering.

Calreticulin is composed of three structural and functional domains. The N-domain has the most conserved amino acid sequence and binds both Zn^{2+} and adenosine triphosphate (ATP). In conjunction with the central, P-domain of the protein, it enables the chaperone function of calreticulin. The central, P-domain of the protein is very rich in proline residues and contains several repeated amino acid sequences. It forms a highly unusual hairpin-like structure that involves the entire proline-rich region, and it is stabilized by a three-strand anti-parallel B-sheet that is formed by the amino acid repeat sequences. The P-domain binds Ca^{2+} with high affinity and low capacity. It also interacts with other ER chaperones and their substrates. The carboxyl-terminal C-domain of calreticulin contains a large number of acidic amino acid residues, and it is responsible for the high-capacity Ca^{2+} -binding behavior of calreticulin.

In addition to binding and buffering Ca^{2+} in the lumen of ER, calreticulin also plays a significant role as a molecular chaperone with a specificity for glycosylated proteins. Specifically, calreticulin binds N-linked, monoglucosylated carbohydrates on newly synthesized proteins. In conjunction with calnexin, an integral membrane protein and also a chaperone in the ER, calreticulin facilitates protein folding and oligomerization, and it supports quality control in protein folding processes. As components of the calreticulin/calnexin cycle, calreticulin might prefer secretory proteins and calnexin integral membrane proteins. Calreticulin and calnexin both form heterodimeric complexes with ERp57 and other chaperones. The formation of these chaperone-chaperone and chaperone-substrate interactions is regulated by changes in Ca^{2+} concentration in the lumen of the ER.

A deficiency in calreticulin in mice is embryonic lethal at E13.5 due to cardiac malformation with this phenotype rescued by targeted over-expression of calcineurin in the heart. Calreticulin is highly expressed in the heart during embryonic development and significantly downregulated in mature heart. Recent evidence points to a role for calreticulin even early

during embryogenesis, during implantation with high expression from E1 to E4. In addition, embryonic stem cells isolated from calreticulin-deficient mice are unable to form beating cardiomyocytes, but have an increased capacity to differentiate into adipocytes, a lineage that has an inverse relationship to cardiomyocytes. Mouse embryonic fibroblasts deficient in calreticulin have disrupted intracellular Ca^{2+} signaling, partially due to the buffering capacity of calreticulin, but possibly also due to the fact that plasma membrane receptors that are part of the Ca^{2+} signaling pathway are not functional, such as the bradykinin receptor and the thrombin receptor, both G-protein-coupled receptors. The Ca^{2+} signaling capacity of calreticulin, both chaperone function and buffering ability, appears to be critical during embryonic development.

Grp94

Grp94 (94-kDa) is another major Ca^{2+} -binding protein, and one of the most abundant proteins in the lumen of the ER. It binds 8–10 mol of Ca^{2+} per mole of protein with low affinity ($K_d = \sim 600 \text{ }\mu\text{M}$). Like calreticulin, Grp94 contains an acidic C-terminal region that comprises the low-affinity Ca^{2+} -binding site. Because of their relatively high capacity for Ca^{2+} binding and their abundance, Grp94 and calreticulin are the most significant Ca^{2+} -binding/buffering proteins in the ER lumen. The expression of Grp94 is upregulated upon glucose starvation, hence its name (glucose-regulated protein). It binds ATP and requires ATP hydrolysis for chaperone function but binds to limited numbers of nascent proteins, most of them advanced folding intermediates or misfolded proteins that have been previously associated with other chaperones; it has a relatively low affinity for early folding intermediates. The N-terminal portion of Grp94 contains one or more high-affinity Ca^{2+} -binding sites which regulates the peptide binding activity of the chaperone. Grp94 is similar in conformation to cytosolic Hsp90 in the absence of nucleotide. It appears that Grp94 is a chaperone specific for advanced intermediates in protein biosynthesis and that it works downstream of, or in conjunction with, other chaperones. Ca^{2+} binding to Grp94 may play a role in the *in vivo* interactions of the protein with other chaperones and with its substrates.

Grp94 deficiency in mice, similar to calreticulin, is lethal *in utero* at day E7.5 due to failure to develop mesoderm. Grp94-deficient embryonic stem cells are unable to differentiate into cardiac, smooth, or skeletal muscle, but have enhanced adipogenesis, similar to calreticulin which may be due to both the chaperone function of Grp94 and possibly the Ca^{2+} buffering capacity. In fact, Grp94 is responsible for chaperoning insulin growth factor II (IGFII), an important factor that is necessary for muscle differentiation, but perhaps may be providing a substantial amount of Ca^{2+} necessary during differentiation for activation of transcription factors.

BiP

BiP is a 78-kDa ER chaperone that was initially identified to bind immunoglobulin heavy chain binding protein (IgH) in B cells. It is a monomeric protein with two distinct functional domains, an ATP-binding domain and a peptide-binding domain. While it has a relatively low capacity for binding

Ca^{2+} (1–2 mol of Ca^{2+} per mole of protein), it contributes possibly as much as 25% of the total Ca^{2+} storage capacity of the ER, and its elevated expression causes an appreciable increase in ER Ca^{2+} storage capacity. BiP assists in the folding of newly synthesized polypeptides by binding to exposed hydrophobic side chains and subsequently coordinating the formation of their correct tertiary and quaternary structure. BiP binds ATP and has high ATPase activity essential for its chaperone function. Its association with nascent polypeptides is stabilized by the high concentrations of Ca^{2+} in the ER lumen and likely involves Ca^{2+} binding. The expression of BiP is upregulated by a variety of stress conditions, including ER Ca^{2+} depletion. During ER stress resulting from Ca^{2+} depletion, BiP senses misfolded protein accumulation and in combination with three other ER proteins, ATF6, IRE1, and PERK, comprises the unfolded protein response. IRE1, PERK, and ATF6 are activated, followed by downregulation in protein production, upregulation of chaperone production, and activation of ER-associated degradation (ERAD). In addition, BiP performs other functions including as a permeability barrier by physically blocking the translocon pore and ratcheting the nascent protein through the pore.

BiP is highly expressed during embryonic differentiation; however, BiP-deficient embryos that are completely devoid of BiP lead to peri-implantation lethality at E3.5. BiP-deficient blastocysts demonstrate a general proliferation defect and an increase in apoptosis of the iodinated contrast media with an inability to generate ES cells. During early development, BiP activity during ER stress may be necessary for implantation and differentiation with BiP essential for embryonic cell growth and pluripotent cell survival. As BiP performs an essential role in the ER as a molecular chaperone as well as a Ca^{2+} buffering protein, BiP chaperone function may be essential for the processing and secretion of growth factors necessary for differentiation, with the Ca^{2+} buffering function necessary for the activation of transcription factors upregulated during the early stages of implantation and differentiation.

The PDI Family of Proteins

Several members of the PDI family are well characterized. Each is folded similarly to thioredoxin, with a typical sequence of α -helices and β -strands. PDI proteins also contain one or more active sites (typically two), again similar to that of thioredoxin. The main function of the PDI family of proteins is to catalyze the oxidation of disulfide bonds and to isomerize incorrectly formed disulfide bonds in newly synthesized polypeptides. Recent studies indicate that some members of this family (PDI, ERp72, and ERcalcistorin/PDI) also contribute to Ca^{2+} buffering in the ER lumen.

PDI

PDI is a ubiquitously expressed multifunctional protein found in the ER. It comprises approximately 0.8% of total cellular protein and attains near millimolar concentrations in the ER lumen of some tissues. During protein folding, PDI catalyzes the formation of native disulfide bonds and disulfide bond rearrangement. PDI is a 58-kDa Ca^{2+} -binding chaperone that binds 19 mol of Ca^{2+} per mol of protein with low affinity ($K_d=2\text{--}5\text{ mM}$). The C-terminus of the protein contains

numerous pairs of acidic residues that form the low affinity, high capacity Ca^{2+} binding sites. PDI is an abundant and versatile redox enzyme that has oxidase, reductase and isomerase activity and interacts with a variety of substrates and is widely distributed in the ER lumen. Its major function is to catalyze disulfide bond formation in newly synthesized proteins and, importantly, high concentrations of Ca^{2+} augment this activity. PDI contains a typical thioredoxin active site, which contains two cysteines separated by two other amino acid residues (the CXXC motif). PDI also functions as a molecular chaperone by binding polypeptides and preventing protein aggregation, and the combination of its redox and chaperone activities is what makes PDI an effective and important enzyme in oxidative protein folding.

PDI knockdown using small-interfering RNA (siRNA) determined that PDI has a protective effect on endothelial cells under both normoxia and hypoxia conditions. To date, no one has been able to generate a PDI-deficient mouse model, possibly due to the critical role of PDI during cell growth, proliferation, and differentiation.

ERp72

ERp72 is a 72-kDa member of the PDI family that binds 12 mol of Ca^{2+} per mol protein with low affinity. Its amino acid sequence includes an acidic C-terminus that is involved in the high-capacity Ca^{2+} binding. ERp72 contains three thioredoxin-like active sites and has PDI-like activity. The expression of ERp72 is induced under stress conditions, including ER Ca^{2+} depletion. A mouse deficient in ERp72 has not to date been generated.

ERcalcistorin/PDI

ERcalcistorin/PDI, a 58-kDa protein, also has high Ca^{2+} -binding capacity and low affinity. ERcalcistorin/PDI binds 23 mol of Ca^{2+} per mole of protein with low affinity ($K_d\sim 1\text{ mM}$). Like other class I ER Ca^{2+} -binding proteins, the high-capacity Ca^{2+} -binding sites in ERcalcistorin/PDI are localized to a C-terminal stretch of acidic amino acid residues. As noted previously for members of the PDI family, high concentrations of Ca^{2+} augment the protein's isomerase activity. ERcalcistorin/PDI has 55% amino acid sequence identity to PDI, and it shows PDI-like chaperone activity.

Other PDI family members

The PDI family of oxidoreductases contains at least 17 family members and are localized to the ER of mammalian cells. They are thought to catalyze disulfide formation to aid folding or to regulate protein function; however, little is known about their individual functions. While information regarding their Ca^{2+} -binding properties is limited, it has been established that they play a role in protein folding and posttranslational modification. P5, a 44-kDa polypeptide also known as CaBP1, contains two internal thioredoxin-like domains and a C-terminal KDEL retention/retrieval signal. While it does bind Ca^{2+} , it is unclear what role it plays in Ca^{2+} buffering in the ER lumen. Protein disulfide isomerase-related (PDIR) has the three CXXC motifs typically found in proteins within the PDI superfamily that are responsible for their oxidoreductase activity. PDIR is preferentially expressed in cells that actively secrete proteins, and its expression is stress-inducible. PDip is a PDI-like protein

expressed specifically in the pancreas. ERdj5 is an ER localized oxidoreductase expressed in salivary glands. Mice deficient in ERdj5 are viable and healthy but have accelerated ER stress response in the salivary gland. Protein disulfide isomerase-like, testis expressed (PDILT) is a testis-specific PDI with a nonconventional SXXC active motif that participates in disulfide-dependent interactions in the ER. EndoPDI is preferentially expressed in endothelial cells and contains three thioredoxin-fold domains. Erp19 contains a CXXC active site and exists as a dimer. Interestingly, Erp44 and Erp28/29, two other ER localized proteins of the PDI family, each contains a single thioredoxin-fold domain but no thioredoxin-like CXXC active site and may be involved in the processing of secretory proteins within the ER. Each PDI family member appears to have specific substrates that they target and this may help explain the apparent redundancy within the family.

Erp57 and Erp44

Two PDI-like proteins, Erp57 and Erp44, are also involved in the regulation of intraluminal Ca^{2+} homeostasis via an interaction with the Ca^{2+} pump SERCA2b and Ca^{2+} channel InsP3R located in the ER membrane. Both proteins appear to be regulated via an interaction of a PDI family member with a specific luminal loop containing conserved cysteine residues. Extrinsic expression of Erp57 in conjunction with SERCA2b results in inhibited Ca^{2+} transport across the ER membrane and is regulated in a Ca^{2+} -dependent manner through an interaction with calreticulin. When ER Ca^{2+} concentration is high, SERCA2b recruits calreticulin and Erp57, resulting in a feedback loop inhibition of Ca^{2+} uptake. Erp57, a 57-kDa protein, carries out disulfide bond exchange in conjunction with calreticulin or calnexin. Ca^{2+} binding to this protein has not been investigated. When ER Ca^{2+} concentration is low, Ca^{2+} release through the InsP3R is blocked by the PDI family member Erp44 interacting with the luminal conserved cysteines. It may be that interaction of these Ca^{2+} channels with these two family members is regulated by redox conditions within the ER, and is dependent on ER Ca^{2+} concentration.

Due to the large number of oxidoreductases in the lumen of the ER, it was surprising to note that Erp57 is embryonic lethal at E13.5. Targeted Erp57 deficiency in CD19⁺ cells demonstrate impaired assembly of the major histocompatibility complex class I peptide-loading complex, demonstrating a role for Erp57 in immune response. Similar to other molecular chaperones, Erp57 is highly expressed during the very early stages of embryogenesis, including formation of the blastocyst. Erp57 regulation of Stat3 signaling during these early stages may be responsible for the early lethality observed in Erp57-deficient mice.

Calsequestrin

Calsequestrin is a high-capacity Ca^{2+} -binding protein found in the SR of cardiac and skeletal muscle. It binds 40–50 mol of Ca^{2+} per mole of protein with low affinity ($K_d \sim 1\text{--}2\text{ mM}$). Small quantities of calsequestrin are also present in the ER of smooth muscle and the cerebellum. There are two isoforms of calsequestrin that are encoded by distinct genes: a 55-kDa cardiac form and a 63-kDa skeletal muscle form. These proteins have very similar amino acid sequences with a C-terminal

stretch containing over 100 acidic amino acids that are involved in the high-capacity Ca^{2+} binding. Calsequestrin is localized to the terminal cisternae of the SR, where it acts as a Ca^{2+} buffer. It lowers the concentration of free Ca^{2+} inside the SR, thereby decreasing the concentration gradient against which the Ca^{2+} -ATPase must work as it transports Ca^{2+} from the cytosol back into the SR. During the initiation of muscle contraction, Ca^{2+} is released from the terminal cisternae of the SR into the cytosol via the ryanodine receptor. Calsequestrin, in addition to being a reversible buffering protein, also functions as an intra-SR Ca^{2+} sensor and mediates the activity of the Ca^{2+} channel. Ca^{2+} -induced conformational changes in calsequestrin affect the junctional face of these membrane proteins and this, in turn, regulates the opening and closing of the Ca^{2+} channel. Intraluminal Ca^{2+} exerts a two-way modulation in cardiac muscle: increase in SR Ca^{2+} promotes channel opening and depletion triggers channel closing. Interestingly, the crystal structure of calsequestrin shows that it contains three similar domains that contain a thioredoxin-like region. The thioredoxin-like fold may, therefore, be a common structural feature of ER/SR luminal proteins.

Mice deficient in skeletal calsequestrin are viable but exhibit reduced levels of releasable Ca^{2+} and contractile sensitivity, increased spontaneous mortality and susceptibility to heat- and anesthetic-induced sudden death. Muscle from skeletal calsequestrin-deficient mice demonstrates abnormal sarcomere morphology with disruption in orientation of T-tubules, double the number of mitochondria and double the density of Ca^{2+} release channels. Most interestingly, calsequestrin-deficient muscle is fatigue resistant with 100% higher force generated after 2 min of continuous stimulation as compared to wild-type controls. A mouse model deficient in cardiac calsequestrin is also viable and displays normal cardiac contractility, left-ventricular size, and unchanged electrocardiogram (ECG) parameters, but, with a significantly slower heart rate and under stress conditions, exhibits cardiac arrhythmias. The Ca^{2+} buffering capacity of calsequestrin appears to be an integral part of the contractile apparatus in muscle.

Class I/II Ca^{2+} -Binding Protein in the ER Lumen

Stim1 is an unusual protein in that it is a membrane protein that binds Ca^{2+} within the lumen of the ER but utilizes a noncanonical EF-hand motif, generating a low-affinity (K_d is between 0.2 and 0.6 mM) Ca^{2+} -binding site. Stim1 is therefore neither part of class I Ca^{2+} -binding proteins or class II Ca^{2+} -binding proteins, but somewhere in between, performing an important role during intracellular Ca^{2+} signaling and involved in ER Ca^{2+} store refilling. After the release of ER Ca^{2+} stores into the cytosol, the cell has a recovery mechanism to reduce the cytosolic Ca^{2+} levels and refill the ER Ca^{2+} stores. This recovery involves the SERCA transporter in the ER membrane as well as Ca^{2+} transporters located at the plasma membrane, which are involved in a specific part of this Ca^{2+} refilling, called store operated Ca^{2+} entry (SOCE). SOCE is a system that was identified a number of years ago, but the actual mechanism has only recently been elucidated. SOCE is an example of coordination between Ca^{2+} signals originating from the ER luminal store and the extracellular

media. Simply, upon ER Ca^{2+} release via InsP_3 , an ER membrane protein, Stim1 senses decreased ER luminal Ca^{2+} and transmits this signal to the plasma membrane where it interacts with a membrane protein, Orai1, responsible for influx of Ca^{2+} .

Stim1

Stim1 is a 685-amino acid, type I transmembrane protein, contains two Ca^{2+} -binding EF hands, one putative disulfide linkage, several glycosylation sites, and a sterile alpha motif (SAM) domain in the luminal domain, as well as cytosolic localized coiled-coil domains, and serine-, proline-, and lysine-rich regions. Stim1 contains two EF-hand motifs, one that binds Ca^{2+} and one that is hidden and is not capable of binding Ca^{2+} . Via the EF-hand Ca^{2+} -binding motif, Stim1 appears to be the initial sensor of decreased ER Ca^{2+} levels, leading to increased Stim1 oligomerization and accumulation near the plasma membrane. The K_d of the EF-hand domain of STIM1 is between 0.2 and 0.6 mM, consistent with ER Ca^{2+} level fluctuations. This K_d is more than 10–100-fold higher than the cytosolic EF-hand Ca^{2+} -binding protein, calmodulin, with Stim1 specially designed to sense changes in the relatively high Ca^{2+} concentration of the ER compartment. Ca^{2+} signaling can trigger dynamic changes of the ER with movement of specialized domains of the ER, specifically movement of the transmembrane protein Stim1, involved in sensing luminal Ca^{2+} concentration, to a sub-plasma membrane location where it is in direct contact with a Ca^{2+} pump, Orai1. This interaction stimulates direct Ca^{2+} influx from the extracellular media, leading to increased cytosolic Ca^{2+} levels. Ca^{2+} is taken up into the ER via the SERCA pump and is used to fill the ER Ca^{2+} stores, enabling the cell to be ready for the next round of Ca^{2+} signaling. Release of ER Ca^{2+} does not result in a uniform increase in all areas of the cytosol, but generates areas of high concentration, closely associated with the compact region between the ER and plasma membrane where Ca^{2+} channels on the plasma membrane open, called Ca^{2+} microdomains. The complex of Ca^{2+} signaling proteins in these microdomains may account for the specific characteristics of individual Ca^{2+} signals. It is possible that the ER-plasma membrane punctae containing Stim1 and Orai1 may perhaps produce an SOCE microdomain.

Stim1-deficient mice present with perinatal and early post-natal lethality and growth retardation. Cells isolated from Stim1-deficient mice demonstrate abnormal cellular physiology and decreased platelet aggregation; thapsigargin-induced Ca^{2+} store release is reduced by 60%, with subsequent SOCE almost completely absent. Stim1 is important for ER Ca^{2+} store refilling, a necessary part of intracellular Ca^{2+} signaling.

Class II Ca^{2+} -Binding Proteins in the ER Lumen

Class II Ca^{2+} -binding proteins in the ER lumen are those that bind Ca^{2+} with high affinity ($K_d = 1 \mu\text{M}$ or less). Much less is known about this class of Ca^{2+} -binding proteins, which includes reticulocalbin, Cab45/stromal cell derived factor 4, calumenin, ERC-55/TCBP-49/E6BP, crocalbin/CBP-50, and calreticulin. The structure, function, and Ca^{2+} -binding

properties of calreticulin have been discussed previously with the class I ER proteins. The class II family of proteins are localized in the cytosol, in various parts of the secretory pathway, secreted to the extracellular space or on the cell surface and may play important roles in the secretory pathway. The emerging functions appear to be highly diverse. With the exception of calreticulin, they all contain six or seven EF-hand type, high-affinity Ca^{2+} -binding sites, and they are considered to be members of the reticulocalbin family or CREC family. This family is a new subset of the EF-hand superfamily of proteins. The function of the high-affinity Ca^{2+} -binding sites in these proteins is not clear, but they may be involved in Ca^{2+} -dependent processes. They may also play an important structural role in the proteins.

Reticulocalbin is a 44-kDa protein that contains six EF-hand motifs. The difference between the amino acid sequences of these domains and EF-hand consensus sequences suggests that some of the domains may have lost their Ca^{2+} -binding capability. Cab45 is a 45-kDa, ubiquitously expressed protein. It contains six EF-hand Ca^{2+} -binding motifs. In contrast to reticulocalbin and ERC-55, which are soluble components of the ER, Cab45 is localized to the Golgi. Cab45 is the first Ca^{2+} -binding protein that has been localized to the lumen of the Golgi, a post-ER compartment. Calumenin is a 39-kDa member of the reticulocalbin family. It is highly homologous to other family members, including reticulocalbin, Cab45, and ERC55. It is localized to the lumen of the ER and has a KDEL-like (HDEF) ER retrieval signal. Calumenin is linked to the inhibition of several proteins in the ER or SR membrane, the vitamin K (1) 2,3-epoxide reductase, the gamma-carboxylase, the RyR, and the Ca^{2+} -transporting ATPase, SERCA. Calumenin has also recently been shown to influence the synthesis of certain coagulation factors in conjunction with thrombospondin1. ERC-55 (55-kDa) comprises an amino-terminal signal sequence followed by six copies of the EF-hand Ca^{2+} -binding motif.

Interestingly, none of the class II CREC family members have been used to generate a deficient mouse model system to date, potentially due to the diverse function and expression as well as inherent nature of Ca^{2+} signaling.

The Functional Significance of Ca^{2+} Buffering in the ER

There are many Ca^{2+} -binding proteins in the ER lumen that both bind and buffer Ca^{2+} and that also assist in protein folding and posttranslational modification. Ca^{2+} buffering within the ER lumen is an essential component of proper cellular Ca^{2+} homeostasis. Furthermore, it has a profound effect on other ER functions including protein and lipid synthesis and stress response. For example, reduction of Ca^{2+} concentration in the ER leads to the accumulation of misfolded proteins, increased expression of ER chaperones, and ER to nucleus/ER to plasma membrane signaling, which inhibits protein synthesis and facilitates protein degradation. Overall, disruption of ER Ca^{2+} homeostasis has profound effects on intracellular communication and cell growth, resulting in organellar diseases, and it has detrimental effects at cellular and systemic levels. This is readily demonstrated by the varied but acute phenotypes associated with Ca^{2+} -binding proteins as seen in mouse model systems. These Ca^{2+} buffering

protein-deficient mouse models provide insight into Ca^{2+} signaling regulation and protein function. By targeting a specific gene and therefore protein to be deficient in the mouse and observing the phenotype, researchers can infer probable function. The broader application of these mice may be transmitted into other areas of biomedical research. Ca^{2+} -binding proteins play an important role in the translation, amplification, and prolonging Ca^{2+} signals, performing essential negative and positive feedback roles in Ca^{2+} trafficking and ultimately facilitating cells to generate rapid changes in local Ca^{2+} levels. The Ca^{2+} present in the ER stores serves as a source of easily releasable Ca^{2+} , and is also important as a regulator of a number of ER enzymes and proteins, including feedback regulation of the IP_3R , the RyR , and SERCA . Coordination of ER luminal Ca^{2+} release and extracellular Ca^{2+} influx into the cytosol is critical as sustained cytosolic Ca^{2+} levels are cytotoxic and lead to apoptosis. Overall, it is clear that the ER may be defined as a multifunctional organelle able to detect and integrate incoming signals, modulate its own luminal dynamics,

and generate output signals in response to environmental changes.

See also: **Bioenergetics:** Calcium-Binding Proteins: Cytosolic (Annexins, Gelsolins, and C_2 -Domain Proteins); Calcium-Modulated Proteins (EF-Hand).

Further Reading

- Baumann O and Walz B (2001) Endoplasmic reticulum of animal cells and its organization into structural and functional domains. *International Review of Cytology* 205: 149–214.
- Eggleton P and Michalak M (2003) *Calreticulin in Health and Disease*. Austin, TX: R. G. Landes Co.
- Molinari M and Helenius A (2000) Chaperone selection during glycoprotein translocation into the endoplasmic reticulum. *Science* 288: 331–333.
- Noiva R (1999) Protein disulfide isomerase: The multifunctional redox chaperone of the endoplasmic reticulum. *Seminars in Cell and Development Biology* 10: 481–493.

Calcium/Calmodulin-Dependent Protein Kinase II

A Hudmon, Indiana University School of Medicine, Indianapolis, IN, USA

H Schulman, Allosteros Therapeutics, Sunnyvale, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Ca²⁺ transient A temporary rise in the intracellular levels of free Ca²⁺ (≈ 100 nM to 1 μM) induced by a variety of extracellular signals (hormones, neurotransmitters, and depolarization) that may vary in its localization, duration, frequency, and amplitude.

Hippocampus Part of the limbic system of the brain that is involved in spatial learning; because of its laminar architecture, it is routinely used for electrophysiological studies designed to evaluate synaptic connectivity and plasticity.

Long-term potentiation A long-lasting increase in synaptic strength that could function to strengthen synaptic connectivity; it is considered to represent a cellular correlate of learning and memory.

Morris water maze An apparatus that is routinely used to assess spatial learning and memory in rodents. In this task, a mouse is placed in a large tub of an opaque liquid and assayed both for its ability to find a visible platform as well as a submerged platform – tasks that evaluate the general function and ability to learn the task as well as the mouse's spatial learning, respectively.

Pre- and postsynaptic elements of neurons Components of typical neuronal synapse: the presynaptic terminal releases a neurotransmitter that diffuses across a synaptic cleft to bind and activate ligand-gated channels along the postsynaptic specialization.

Pseudosubstrate A component of the autoregulatory domain that binds to the substrate-binding pocket on the catalytic domain to prevent kinase activity in the absence of calmodulin.

Calcium/calmodulin-dependent protein kinase II (CaMKII) is a multifunctional serine/threonine (Ser/Thr) protein kinase involved with a diverse array of cellular functions. From metabolism and cytoskeletal organization to gene expression and the cell cycle, CaMKII phosphorylation of key substrates transduces a transient rise in free calcium (Ca²⁺) induced by electrical activity, neurotransmitters, hormones, and other signaling molecules into diverse cellular effects. CaMKII is particularly important in the heart and nervous system, where genetics (knockouts, knockins, and transgenics) has been employed to delineate its key functions, including excitation–contraction coupling in heart and learning/memory in brain. In neurons, CaMKII is strategically localized in both pre- and postsynaptic compartments, where it regulates Ca²⁺-dependent processes central to neuronal communication, including neurotransmitter synthesis and release and ion flux through voltage- and ligand-gated channels. In cardiomyocytes, CaMKII functions to tune gene expression and the Ca²⁺ signaling machinery in response to myocyte activity and hormones. CaMKII possesses autoregulatory properties that may enable it to function as a 'molecular switch' to translate the calcium signaling associated with cellular activity into graded levels of kinase activity. Autophosphorylation can be considered as a form of molecular memory, as it generates a form of persistent activity (autonomous) that is temporally uncoupled from the Ca²⁺ signal. This autoregulation also enables CaMKII to function as a Ca²⁺ spike integrator – the ability to generate graded levels of autonomous activity that is differentially sensitive to the frequency of Ca²⁺ transients. These emergent properties underlie its function as a cognitive kinase with both short- and long-term consequences on complex behaviors, such as learning and memory. In heart, this may enable it to decode heart rate and coordinate

physiological changes, such as the 'fight or flight' response, by modulation of many substrates involved in Ca²⁺ signaling and homeostasis.

Neuronal CaMKII: Localization and Substrates

CaMKII is highly concentrated in the nervous system. At roughly 1–2% of the total protein in the brain, it is the predominant Ca²⁺/calmodulin (Ca²⁺/CaM)-regulated kinase in the central nervous system. CaMKII has a diverse intracellular localization, including the postsynaptic density (PSD), a cytoskeletal specialization replete with signaling molecules (e.g., ligand- and voltage-gated receptors, kinases, phosphatases, and anchoring and structural proteins) that lie immediately across the presynaptic terminal below the postsynaptic membrane. Biochemical cascades at the PSD regulated by CaMKII are likely central to the function of this highly specialized signaling network for the following reasons: (1) CaMKII constitutes a significant fraction of the total PSD protein, (2) it is recruited to the PSD in an activity-dependent manner, (3) it anchors to multiple PSD proteins (e.g., densin-180, α-actinin, and N-methyl-D-aspartate (NMDA) receptors), and (4) it regulates the activity of multiple PSD proteins through phosphorylation. One prominent PSD substrate of CaMKII is the GluR1 subunit of the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA)-subtype of the glutamate receptor. CaMKII phosphorylation of a specific serine (Ser⁸³¹) in this channel increases its conductance by favoring the open probability of its high conductance state. CaMKII also regulates the trafficking and postsynaptic insertion of AMPA receptors. Although both effects could contribute to synaptic

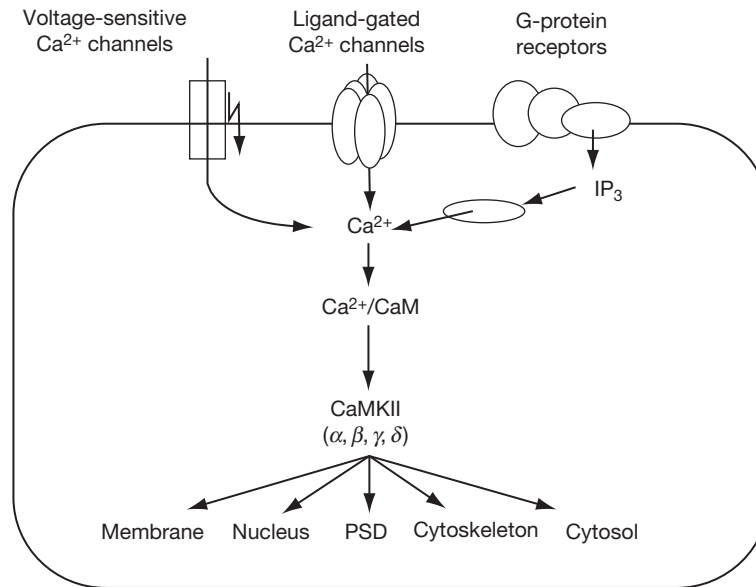


Figure 1 A variety of extracellular signals (e.g., depolarization, neurotransmitters, and hormones) function to elevate intracellular Ca^{2+} levels through the action of ligand- and voltage-gated channels and IP_3 -operated intracellular stores. Activation of CaMKII (α , β , γ , and δ isoforms) throughout the neuron via Ca^{2+} /CaM leads to phosphorylation of substrates at the membrane, cytoskeleton, postsynaptic density (PSD), nucleus, and in the cytosol. Adapted from Hudmon A and Schulman H (2002) Neuronal Ca^{2+} /calmodulin-dependent protein kinase II: The role of structure and autoregulation in cellular function. *Annual Review of Biochemistry* 71: 473–510.

plasticity, the high concentration of CaMKII in the PSD also suggests that this protein plays a structural function at the synapse. For example, CaMKII functions as a scaffold to bind and recruit the ubiquitin proteasome to dendritic spines.

Presynaptic CaMKII regulates neurotransmitter synthesis and neurotransmitter vesicle availability and release by its phosphorylation of vesicle proteins (e.g., Synapsin I) cytoskeletal regulating proteins (e.g., MAP2), membrane channels (e.g., calcium channels), and other enzymes (e.g., tyrosine hydroxylase). CaMKII is also present in the nucleus, where it phosphorylates key substrates (e.g., CREB) involved in the regulation of gene transcription and long-term changes in neuronal function. Thus, CaMKII is strategically positioned throughout the neuron to transduce and coordinate a host of calcium-mobilizing stimuli into the appropriate alterations in neuronal function (see Figure 1).

Activation, Autoregulation, and Structure of CaMKII

CaMKII is composed of a family of highly conserved isoforms (α , β , γ , and δ) that are encoded by unique genes and differentially expressed throughout the body (see Figure 1). Although the δ and γ isoforms are ubiquitous, the α and β isoforms are highly expressed in the nervous system. The molecular weights of these subunits range from 54 to 72 kDa, with α isoform being the smallest subunit at 54 kDa. This size difference is due primarily to domain insertions in the region between the regulatory and association domains (i.e., variable domains) because each of these genes encodes protein products possessing a catalytic, regulatory, and association domain. A prototypical α -subunit of CaMKII, illustrating the relative positions of

these domains, is shown in Figure 2. Like all protein kinases, the catalytic domain serves to bind substrates and catalyze the phosphotransferase reaction from Mg^{2+} /adenosine triphosphate (Mg^{2+} /ATP) to the protein substrate.

CaMKII is coupled to the Ca^{2+} stimulus via CaM, a small dumbbell-shaped protein that, following Ca^{2+} binding, is competent to bind and activate CaMKII and other target proteins. The autoregulatory domain is composed of three components: a region that contains the Thr286 autophosphorylation site, an adjacent domain that anchors the regulatory domain to the catalytic surface, and a C-terminal region that contains the CaM recognition sequence. Autophosphorylation of Thr286 permits CaM to bind into the middle region (presumably to Phe293) to enable high-affinity binding (CaM trapping) of CaM to the autoregulatory domain. The association domain produces the association of kinase subunits into a multimeric complex, termed the 'holoenzyme'.

A computer-assisted reconstruction of α -CaMKII obtained from images using transmission electron microscopy indicated that the catalytic heads extend from the gear-shaped central core in a dodecameric (12 subunits) arrangement (Figure 2). Although the CaMKII holoenzyme has yet to be resolved at the atomic resolution, crystal structures of the catalytic and association domains have been independently resolved. The oligomerization domain functions to bring catalytic subunits within the holoenzyme into close approximation for a trans-autophosphorylation event (Figure 2, lower right). The regulatory domain blocks substrate access to the catalytic surface in the nonactivated state; however, following Ca^{2+} /CaM binding, the regulatory domain "peels" back to expose the catalytic cleft and substrate binding pocket for kinase activity (Figure 2, lower left).

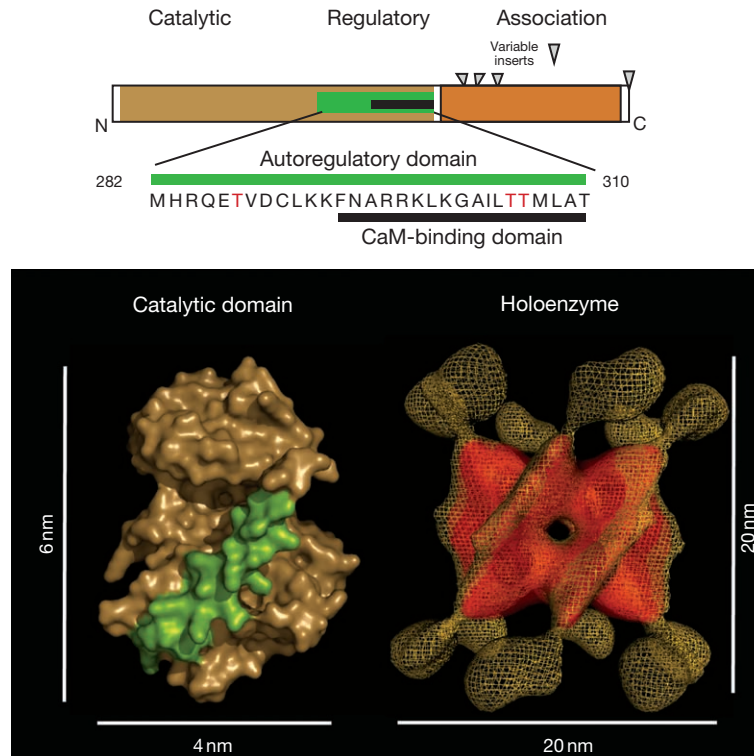


Figure 2 Linear diagram of α -CaMKII illustrating the catalytic, regulatory, and association domains. Like all kinases, the catalytic domain is composed of both an ATP and a protein–substrate binding region. The autoregulatory region is a trifunctional module (shown in green), composed of a region that contains the Thr286 autophosphorylation site, a region that serves to anchor the regulatory domain to the catalytic surface, and a region that contains the calmodulin recognition motif. The association domain drives subunit multimerization into a holoenzyme (orange). A computer-assisted reconstruction of α -CaMKII obtained from images produced using transmission electron microscopy indicates a dodecameric structure, with 12 catalytic heads radiating out from the top and bottom hexameric rings of association domains (orange) at the center. The crystal structure of a single catalytic subunit (1–315) shows the autoinhibitory domain (green) positioned across the catalytic surface (gold). Adapted from Kolodziej SJ, Hudmon A, Waxham MN, and Stoops JK (2000) Three-dimensional reconstructions of calcium/calmodulin-dependent (CaM) kinase II and truncated CaM kinase II reveal a unique organization for its structural core and functional domains. *Journal of Biological Chemistry* 275: 14354–14359; and from Rosenberg RS, Deindl S, Sung R-J, Nairn AC, and Kuriyan J (2005) Structure of the autoinhibited kinase domain of CaMKII and SAXS analysis of the holoenzyme. *Cell* 123, 849–860.

The isoform composition of the holoenzyme may endow CaMKII with the properties and regulation unique to each isoform. For example, the β isoform possesses an F-actin binding domain that targets this isoform to cytoskeletal structures, where it participates in regulating the sprouting and stability of neuronal dendrites.

The Structure/Function of CaMKII Autophosphorylation

In addition to phosphorylating protein substrates throughout the cell (Figure 1), CaMKII is a substrate for its own catalytic activity via autophosphorylation. Following Ca^{2+} /CaM activation, autophosphorylation may occur at Thr²⁸⁶ (Thr²⁸⁷ of the β , γ , and δ isoforms), at the N-terminal end of the regulatory domain (see Figure 2). Autophosphorylation at Thr²⁸⁶ is not necessary for enzyme activation. Unlike many protein kinases, the ‘activation loop’ of CaMKII requires no phosphorylation for catalytic competence. The autophosphorylation reaction is dominated by an intraholoenzyme intersubunit reaction that requires CaM bound to the subunit acting as kinase as well as the subunit behaving as substrate (Figure 3). Three functional

consequences associated with Thr²⁸⁶ autophosphorylation are: (1) the generation of autonomous activity, (2) an increase in the binding affinity of CaM (CaM trapping), and (3) exposure of a site on the catalytic domain that functions as a binding site to anchor CaMKII to its substrates.

Autonomous Activity

Autonomous activity (Ca^{2+} /CaM-independent activity) has intrigued neurobiologists and enzymologists since its discovery 25 years ago in the marine mollusk, *Aplysia californica* – a well-characterized animal model for studying the molecular mechanisms of learning and memory. Autonomous activity is the residual activity of the autophosphorylated subunit after CaM dissociation. At the structural level, the mechanism of autonomous activity is likely attributed to the negative charges associated with phospho-Thr²⁸⁶ preventing reassociation and inactivation of the catalytic domain by the autoregulatory domain, following CaM dissociation. As indicated below, oxidation of two methionine residues near this critical Thr residue similarly generate autonomous activity.

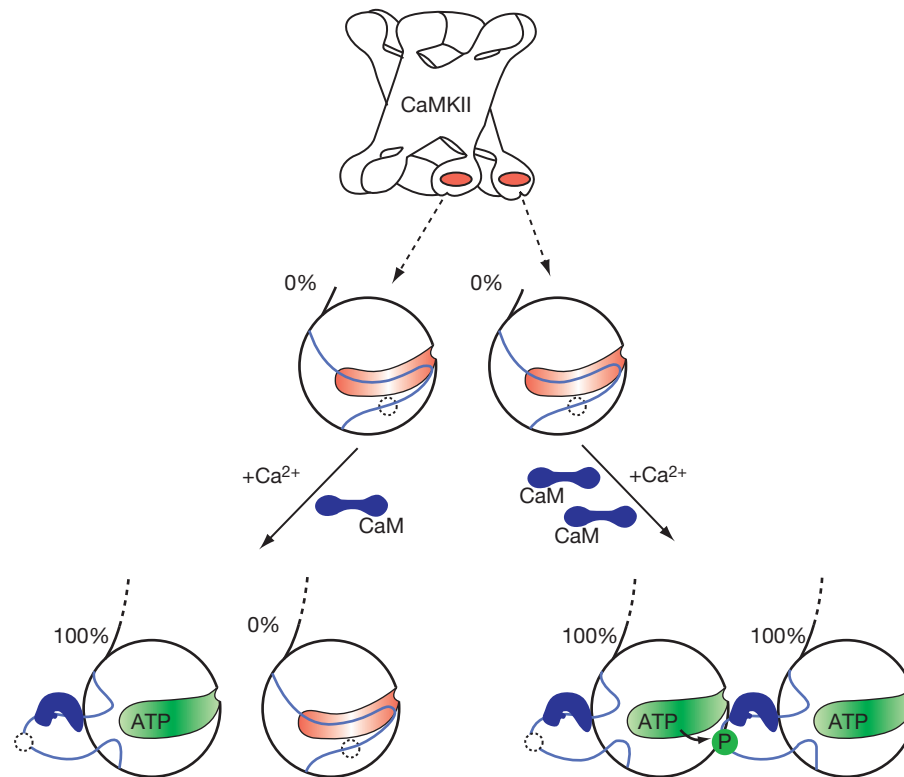


Figure 3 Autophosphorylation requires coincident Ca^{2+} /CaM binding between neighboring subunits within CaMKII. Catalytic/regulatory regions are depicted from subunits within the same holoenzyme (i.e., intraholoenzyme) and CaM is depicted as the purple dumbbell. Ca^{2+} /CaM binding displaces the autoinhibitory domain to activate subunits within the holoenzyme. Coincident binding of Ca^{2+} /CaM exposes the autonomy site (Thr²⁸⁶), represented as a dashed circle, of one subunit for phosphorylation by an active neighboring subunit in an intersubunit reaction. Adapted from Hudmon A and Schulman H (2002) Neuronal Ca^{2+} /calmodulin-dependent protein kinase II: The role of structure and autoregulation in cellular function. *Annual Review of Biochemistry* 71: 473–510.

CaM Trapping

Among the various CaM target proteins reported, CaMKII possesses a relatively low binding affinity for CaM ($K_d \sim 15$ nM). However, following autophosphorylation, the CaM binding affinity is increased by over a 1000-fold ($K_d \sim 20$ pM), thus attaining one of the highest CaM binding affinities. The autophosphorylation of Thr²⁸⁶ is thought to indirectly produce CaM trapping because a peptide designed from the Ca^{2+} /CaM-binding domain of CaMKII (290–309) possesses high-affinity Ca^{2+} /CaM binding, even though Thr²⁸⁶ is not contained within the peptide sequence (see Figure 2). This is mechanistically consistent with steric constraints or an inaccessibility of the binding site for Ca^{2+} /CaM in the non-Thr²⁸⁶ autophosphorylated state. In neurons, autophosphorylation and subsequent CaM trapping appear to regulate the translocation and dissociation of CaMKII from the PSD. Priming or prior neuronal stimulation is believed to enhance CaMKII autophosphorylation, a process that may trigger long-term PSD association.

Substrate Anchoring

The subcellular targeting of multifunctional kinases (e.g., CaMKII, cyclic adenosine monophosphate (cAMP)-dependent protein kinase, and protein kinase C) is thought to be integral

to substrate specificity and the kinetics of the phosphorylation reaction. Ca^{2+} /CaM binding and autophosphorylation of Thr²⁸⁶ expose a binding site on CaMKII to permit anchor proteins and substrates to directly associate with the enzyme. Protein targeting may also afford new modes of CaMKII regulation. For example, CaMKII targeting to the NR2B subunit of the NMDA-type glutamate receptor requires Ca^{2+} /CaM to induce binding. However, once the interaction is formed, the bound enzyme remains active ‘even’ in the absence of Ca^{2+} /CaM and Thr²⁸⁶ autophosphorylation. Binding of the NR2B subunit to the catalytic domain may act as a wedge to prevent the autoregulatory domain from reforming its contacts subsequent to CaM dissociation. A number of new CaMKII interactors utilizing a similar binding mechanism have been described, including dopamine receptors, connexins, accessory subunits for ion channels, and the proteasome. In addition, autophosphorylation associated with ischemia and epilepsy leads to a new site, Thr²⁵³, which appears to enhance CaMKII interaction with PSD-binding proteins.

Autophosphorylation at Thr³⁰⁵/Thr³⁰⁶

Ca^{2+} /CaM-independent autophosphorylation occurs at amino acids Thr³⁰⁵/Thr³⁰⁶ within the CaM-binding domain. Often referred to as ‘inhibitory autophosphorylation’, it is initiated

once $\text{Ca}^{2+}/\text{CaM}$ dissociates from a kinase phosphorylated at Thr²⁸⁶. Autophosphorylation within the CaM-binding domain (Thr³⁰⁵ or Thr³⁰⁶) blocks rebinding of CaM to this subunit. This autophosphorylation also occurs in the basal state (auto-inhibited), by an intrasubunit mechanism. Basal autophosphorylation preferentially occurs at Thr³⁰⁶, and blocks $\text{Ca}^{2+}/\text{CaM}$ binding. Thus, autophosphorylation is a complex process involving multiple sites that can function to uncouple CaMKII from requiring $\text{Ca}^{2+}/\text{CaM}$ for activity as well as preventing $\text{Ca}^{2+}/\text{CaM}$ from binding to activate (or reactivate) the kinase.

Autophosphorylation as an Index of Neuronal Activity

CaMKII is a major mediator of cellular stimuli related to elevation in intracellular free Ca^{2+} . Autophosphorylation of CaMKII in response to these stimuli not only potentiates its activity beyond the Ca^{2+} transient, but also translates the temporal dynamics (e.g., frequency) of Ca^{2+} signaling. *In vitro*, CaMKII subjected to $\text{Ca}^{2+}/\text{CaM}$ pulsed at varying frequencies generates distinct graded levels of autonomous activity *in vitro*. The level of CaMKII autonomy or level of phospho-Thr²⁸⁶ in response to cell stimuli is commonly

used to monitor its activity. Although the molecular mechanisms of this frequency detection remains to be fully elucidated, factors identified in computer simulations important for this process include the effective cooperativity in the Thr²⁸⁶ autophosphorylation reaction and the availability of CaM for this reaction. In cells, CaM is thought to be largely bound to a variety of cellular proteins and presumably not available to maximally activate all of its targets. CaMKII activation and autophosphorylation are particularly sensitive to limiting CaM levels, as its relatively low affinity for CaM would predict that a transient rise in Ca^{2+} would lead to only a few of the subunits in each holoenzyme being activated (Figure 4). In addition, structural data suggest that cooperativity may also be generated when the catalytic surface of an activated subunit captures the regulatory domain of non-activated subunits to facilitate CaM binding to the captured subunit. This process would potentiate CaM binding and subsequent autophosphorylation of adjacent subunits within the holoenzyme. Because autophosphorylation further prolongs CaM association, a feed-forward process ensues at a threshold level of autophosphorylation to produce the non linear Ca^{2+} spike frequency function for CaMKII activation/autophosphorylation as illustrated in Figure 4.

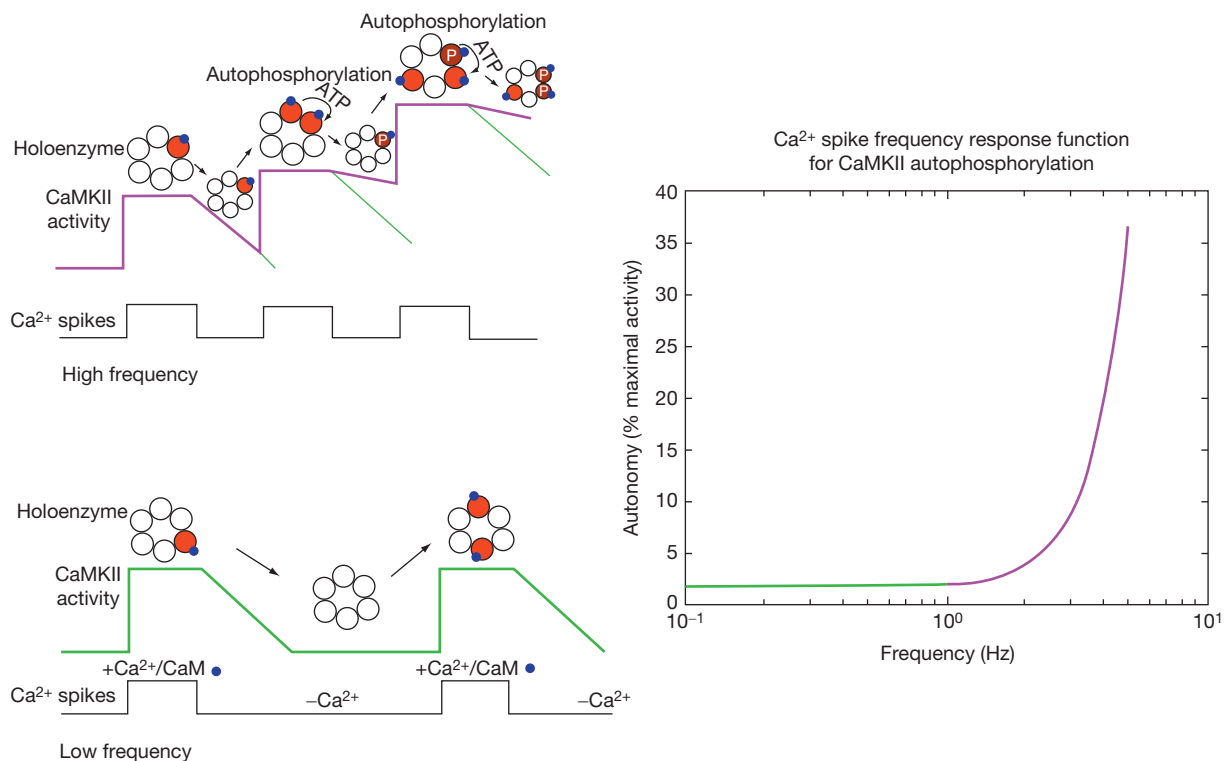


Figure 4 (Left) The role of coincident CaM binding, autophosphorylation, and CaM trapping in Ca^{2+} spike-frequency detection. Inactive subunits within CaMKII (holoenzymes are shown as 6-mers for simplicity) are represented by open circles. $\text{Ca}^{2+}/\text{CaM}$ (blue circle) binding activates a given subunit (shown in red) and coincident $\text{Ca}^{2+}/\text{CaM}$ binding results in autophosphorylation and CaM trapping ('P' inside a filled dark red circle). During low-frequency Ca^{2+} spikes, CaM completely dissociates between Ca^{2+} spikes, effectively producing 'naïve' CaMKII subunits at each inner spike interval. However, at high Ca^{2+} spike frequencies, CaM does not completely dissociate between Ca^{2+} spikes; this increases the probability that coincident CaM binding, autophosphorylation, and CaM trapping may occur – a process that effectively increases the probability that a neighboring subunit will also bind $\text{Ca}^{2+}/\text{CaM}$ during successive spikes to produce the nonlinear increase in autonomous activity illustrated by a computer simulation encompassing the biophysical and enzymatic characteristics of CaMKII (Right). Adapted from Hudmon A and Schulman H (2002) Structure–function of the multifunctional Ca^{2+} /calmodulin-dependent protein kinase II. *Biochemical Journal* 364: 593–611.

CaMKII in Heart

The cardiac action potential stimulates CaMKII with each heartbeat, which then performs the function of a master regulator of multiple downstream physiological responses across a range of heart rates and timescales. These include Ca^{2+} homeostasis, excitation–contraction coupling, arrhythmia and automaticity, and cardiomyocyte hypertrophy. The kinase is concentrated along the Z band in ventricular myocytes in association with its substrates (L-type Ca^{2+} channel (LTCC) and the ryanodine receptor), where it is poised to regulate Ca^{2+} influx into the cell as well as release from intracellular stores. CaMKII mediates an important feed-forward effect relevant for both normal physiology and pathophysiology. Ca^{2+} entry via the LTCC activates the kinase, which phosphorylates the channel to facilitate subsequent Ca^{2+} current by shifting channel gating to frequent long openings. This frequency-dependent regulation of cardiac contraction functions to enhance blood flow through the heart despite the shorter duration of the cardiac cycle at high heart rates (i.e., force–frequency relationship or the ‘treppe’ effect). This and other actions are consistent with proarrhythmic electrical remodeling in myocardial hypertrophy and heart failure, actions that are enhanced in transgenic animals overexpressing CaMKII and are blocked in transgenic animals expressing a peptide inhibitor of CaMKII.

CaMKII responds to hormonal as well as electrical stimulation of the heart. Paradoxically, the β -adrenergic receptor-mediated stimulation of the heart, which is a classical example of cAMP acting on protein kinase A (PKA) and on a cyclic nucleotide-gated ion channel, also leads to increases in Ca^{2+} and CaMKII activation. In the sinoatrial node that paces heart rate, β -adrenergic stimulation leads to a faster action potential that is the basis of the ‘fight or flight’ response. In mice, this β -adrenergic response is primarily mediated by CaMKII, as it is retained in animals lacking the cyclic nucleotide-gated ion channel and is blocked by inhibition of CaMKII. The pro-oxidant angiotensin II pathway in the heart demonstrates a novel modulation of CaMKII by reactive oxygen species (ROS). ROS leads to oxidation of two methionines Met²⁸¹/Met²⁸² with consequences similar to autophosphorylation of Thr²⁸⁷ of the cardiac CaMKII δ isoform. Oxidation disables the autoinhibitory domain so that the kinase remains active after dissociation of Ca^{2+} /CaM. It usurps a mechanism used by autophosphorylation that links oxidative stress mediated by angiotensin II and ischemia with CaMKII activation and resulting downstream effects, including myocardial apoptosis.

CaMKII in Long-Term Potentiation and Learning and Memory

Neuronal information in the brain is propagated and encoded through synaptic activity, and multiple lines of evidence indicate that CaMKII is critical for long-lasting changes in synaptic efficacy, such as long-term potentiation (LTP). LTP is a long-lasting enhancement in the strength of neurotransmission that is believed to represent a cellular correlate of learning. Synaptic stimulation protocols that produce LTP in the hippocampus also produce long-lasting changes in the activation and autophosphorylation of CaMKII. Importantly, peptide inhibitors of

CaMKII inhibit LTP induction, whereas viral introduction of a constitutively active form of CaMKII occludes LTP generated by the high-frequency stimulation protocols typically used to induce LTP. The activity and autophosphorylation of CaMKII are also important for learning and memory *in vivo*.

Genetic approaches in flies, worms, and mice have greatly advanced our understanding of the molecular mechanisms involved in behavior and learning. Although we will focus on the role of CaMKII in mammalian learning, disrupting the activity of CaMKII induces behavioral or cognitive defects in ‘all’ of these organisms. The genetic deletion in the α -isoform of CaMKII was a pioneering study aimed at understanding the role of CaMKII in mammalian learning and memory. These CaMKII knockout mice displayed deficits in hippocampal LTP and possessed severe learning impairments in spatial learning tasks, such as the Morris water maze. These knockout mice directly implicated CaMKII and hippocampal LTP in mammalian learning and memory.

The use of knockin genetics permitted further dissection of the role of CaMKII in learning by altering autoregulatory properties without ablating kinase activity. A knockin mouse in which Thr²⁸⁶ (α -isoform) was mutated to Ala²⁸⁶ to prevent autophosphorylation displayed deficient LTP and spatial learning, directly implicating Thr²⁸⁶ autophosphorylation in synaptic plasticity and learning. Transgenic mice in which Thr²⁸⁶ was replaced with a charged residue (e.g., Asp²⁸⁶) in order to mimic an autophosphorylated state of the enzyme display enhanced LTP at low levels of transgene expression and deficient LTP at higher levels, suggesting complex interactions between the level of activated CaMKII and its function in regulating synaptic strength. Autophosphorylation within the CaM-binding domain (Thr³⁰⁵/Thr³⁰⁶) has also been implicated in regulating CaMKII targeting and function. Knockin mice that cannot undergo inhibitory autophosphorylation (Ala³⁰⁵/Ala³⁰⁶) have increased levels of CaMKII in the PSD and a lower threshold for LTP. Mimicking inhibitory autophosphorylation (Asp³⁰⁵/Asp³⁰⁶) lowers the CaMKII content in the PSD and blocks LTP induction and hippocampal learning. Finally, disrupting the normal dendritic mRNA targeting of α -CaMKII impairs LTP and learning, suggesting that local translation and delivery of CaMKII to its site of action at the synapse is important for synaptic plasticity learning.

The Tao of CaMKII

The function of CaMKII in synaptic plasticity and learning and memory appears to be tightly linked to its unique structure and autoregulation. The multimeric structure of CaMKII permits different isoforms to associate into a macromolecular signaling complex to provide specific targeting information essential for its subcellular localization as well as an intersubunit autophosphorylation reaction. Key autoregulatory features following autophosphorylation include the ability to: (1) reduce its Ca^{2+} /CaM dependence following dissociation of CaM (autonomous activity), (2) increase its affinity for Ca^{2+} /CaM by over 1000-fold (CaM trapping), (3) enable activity-dependent substrate targeting, and (4) enable it to function as a Ca^{2+} spike frequency detector. Thus, the designation of a kinase as a cognitive kinase (i.e., based on its requirement in learning

and memory) possesses a twofold meaning for CaMKII. Because of this enzyme's complex autoregulation and autophosphorylation, CaMKII activity may directly function as a molecular representation of synaptic activity: a molecular memory that is expressed as a persistent form of activity temporarily uncoupled from the Ca^{2+} stimulus. In brain as in other organs, the role of CaMKII in transducing cellular signals into the complex changes associated with its many functions appears to be an intimate dance between its autoregulation and its capacity to phosphorylate key protein substrates integral to cell function.

See also: **Bioenergetics: Voltage-Gated Ca^{2+} Channels; Signaling: Protein Kinase C Family.**

Further Reading

Couchonnal LF and Anderson ME (2008) The role of calmodulin kinase II in myocardial physiology and disease. *Physiology* 23: 151–159.

- Griffith LC (1997) *Drosophila melanogaster* as a model system for the study of the function of calcium/calmodulin-dependent protein kinase II in synaptic plasticity. *Invertebrate Neuroscience* 3: 93–103.
- Hudmon A and Schulman H (2002) Neuronal Ca^{2+} /calmodulin-dependent protein kinase II: The role of structure and autoregulation in cellular function. *Annual Review of Biochemistry* 71: 473–510.
- Kennedy MB (2000) Signal-processing machines at the post-synaptic density. *Science* 290: 243–257.
- Lisman J, Schulman H, and Cline H (2002) The molecular basis of CaMKII function in synaptic and behavioral memory. *Nature Reviews Neuroscience* 3: 175–190.
- Miller S, Yasuda M, Coats JK, et al. (2002) Disruption of dendritic translation of CaMKII α impairs stabilization of synaptic plasticity and memory consolidation. *Neuron* 36: 507–519.
- Rellos P, Pike ACW, Niesen FH, et al. (2010) Structure of the CaMKII δ /calmodulin complex reveals the molecular mechanism of CaMKII kinase activation. *PLoS Biology* 8: e1000426.
- Shen K, Teruel MN, Connor JH, Shenolikar S, and Meyer T (2000) Molecular memory by reversible translocation of calcium/calmodulin-dependent protein kinase II. *Nature* 3: 881–886.
- Silva AJ (2003) Molecular and cellular cognitive studies of the role of synaptic plasticity in learning and memory. *Journal of Neurobiology* 54: 224–327.
- Wayman GA, Lee YS, Tokumitsu H, Silva AJ, and Soderling TR (2008) Calmodulin-kinases: Modulators of neuronal development and plasticity. *Neuron* 59: 914–931.

Calcium/Calmodulin-Dependent Protein Kinases

A J Robison, Mt. Sinai School of Medicine, New York, NY, USA

R J Colbran, Vanderbilt University, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Autoinhibitory domain (AID) The region of a Ca^{2+} /calmodulin-dependent protein kinase that binds and inhibits the catalytic domain. Inhibition is relieved by interaction with calmodulin and, in some cases, is prevented by subsequent autophosphorylation.

Calmodulin A small cytosolic protein that binds four Ca^{2+} ions and subsequently associates with and modulates many important proteins, including some protein kinases.

Catalytic domain The region of an enzyme responsible for activity; in protein kinases, this domain transfers the γ -phosphate from adenosine triphosphate (ATP) to the target residue.

Protein kinase An enzyme that transfers the terminal (γ) phosphate from ATP to a serine, threonine, or tyrosine residue of a substrate protein.

Protein phosphatase An enzyme that hydrolyzes a phosphate group from a serine, threonine, or tyrosine residue of a phosphoprotein substrate.

The Ca^{2+} /calmodulin-dependent protein kinase enzyme family is activated by increases in intracellular calcium ions (Ca^{2+}) and transfers phosphates from adenosine triphosphate (ATP) to specific serine or threonine residues in other proteins. Family members vary in expression level, subcellular location, and substrate specificity. They are essential for signal transduction in all cells and modulate a variety of physiological processes including learning and memory, metabolism, gene transcription, messenger RNA (mRNA) translation, and muscle contraction.

Ca^{2+} Signaling

Modulation of the concentration of Ca^{2+} is used as an intracellular signal to allow hormones and neurotransmitters to control numerous processes in virtually all cell types. In general, cells have a resting Ca^{2+} concentration of 50–100 nM, but this concentration can spike above 1 μM in an activated cell. Interestingly, these changes in Ca^{2+} concentration can activate extremely disparate signaling cascades not only in different cell types but also in different areas of the same cell. For instance, Ca^{2+} elevation in the nucleus of a neuron activates gene transcription, whereas a similar elevation at the synapse causes neurotransmitter release. These responses are possible because of the hundreds of different molecules that are directly or indirectly responsive to Ca^{2+} , the numerous mechanisms for increasing Ca^{2+} , and the variation in duration, frequency, amplitude, and localization of Ca^{2+} spikes. Ca^{2+} can enter the cell through channels in the membrane or be released from internal stores. The resulting Ca^{2+} spikes can occur on the millisecond level in neurons or be separated by days in the mitotic cell, and can modulate diverse enzymes and/or alterations in the shape of structural molecules. Many of the intracellular effects of Ca^{2+} are mediated by calmodulin, a small protein that undergoes dramatic conformational changes when bound to Ca^{2+} ions. Among the many enzymes

activated by Ca^{2+} /calmodulin (**Figure 1**), members of the Ca^{2+} /calmodulin-dependent protein kinase (CaMK) family mediate diverse physiological responses in most, if not all, cell types.

Protein kinases are enzymes that catalyze the addition of phosphates onto specific serine, threonine, or tyrosine residues in target proteins, thereby regulating them in some way. Substrate proteins include other enzymes, gene transcription factors, scaffolding proteins, receptors, and many other cellular proteins. Thus, kinases can regulate such diverse cellular events as intercellular signaling, cell motility and shape, cell cycle, metabolic pathways, and gene transcription. In addition, kinases themselves may be regulated by intracellular signals, including Ca^{2+} acting via calmodulin. Most cells contain multiple representatives from the diverse family of CaMKs (**Figure 2**). Generally, CaMKs contain an autoinhibitory domain (AID) that interferes with catalysis by blocking or disrupting the catalytic site. This inhibition is relieved by the binding of Ca^{2+} /calmodulin to a domain that overlaps the AID (**Figure 2**), thus activating the kinase.

CaMKs phosphorylate specific serine or threonine residues in substrate proteins based, in part, on the surrounding amino acid sequence. Each of the kinases favors target serines or threonines in specific consensus sequences, although their substrate specificities can overlap. For instance, CaMKII and CaMKIV recognize the consensus sequences $\text{Hyd-X-Arg-NB-X-Ser/Thr-Hyd}$ and $\text{Hyd-X-Arg-X-X-Ser/Thr-X}$ (where Hyd is a hydrophobic residue, NB is a nonbasic residue, and X is any residue), respectively. However, some proteins containing sites that conform to both of these canonical sequences are phosphorylated by the two kinases with similar affinity. In addition, the catalytic domains of most CaMKs contain acidic residues that are important for interaction with the basic residues in their specific substrates (exceptions are noted below). Based on their substrate specificity, CaMKs can be divided into two general groups: those with narrow substrate specificity and those with broad substrate specificity.

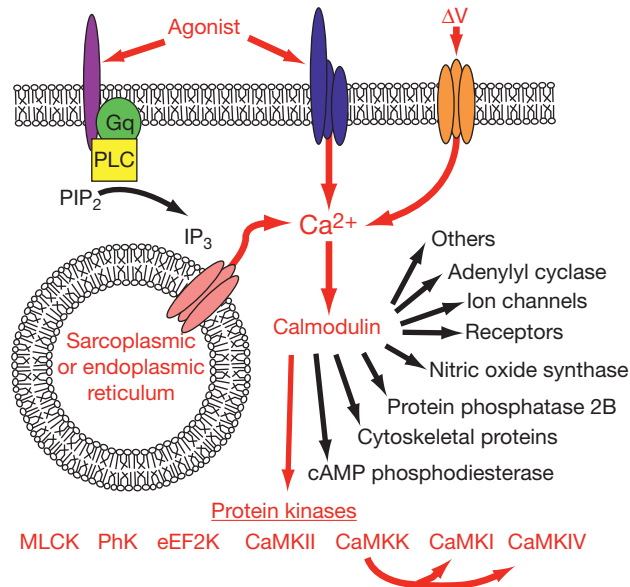


Figure 1 Modulation and effects of intracellular Ca^{2+} . Ca^{2+} enters cells via ligand-gated (blue) or voltage-gated (orange) Ca^{2+} channels in the plasma membrane or can be released from sarcoplasmic or endoplasmic reticulum stores via inositol 1,4,5-triphosphate (IP_3) receptors and/or ryanodine receptors (not shown). IP_3 is generated from phosphatidyl-inositol 4,5-bisphosphate (PIP_2) by phospholipase C (PLC), which is activated by ligand binding to specific G_q -protein-coupled receptors (magenta) in the plasma membrane. Cytosolic Ca^{2+} is rapidly extruded from the cell or taken up by the intracellular stores (not shown). The combined actions of these pathways result in spatially localized and transient changes in intracellular Ca^{2+} . The ubiquitous Ca^{2+} -binding protein calmodulin senses the changes in Ca^{2+} and mediates the activation of multiple intracellular enzymes and other processes, including the family of CaMKs.

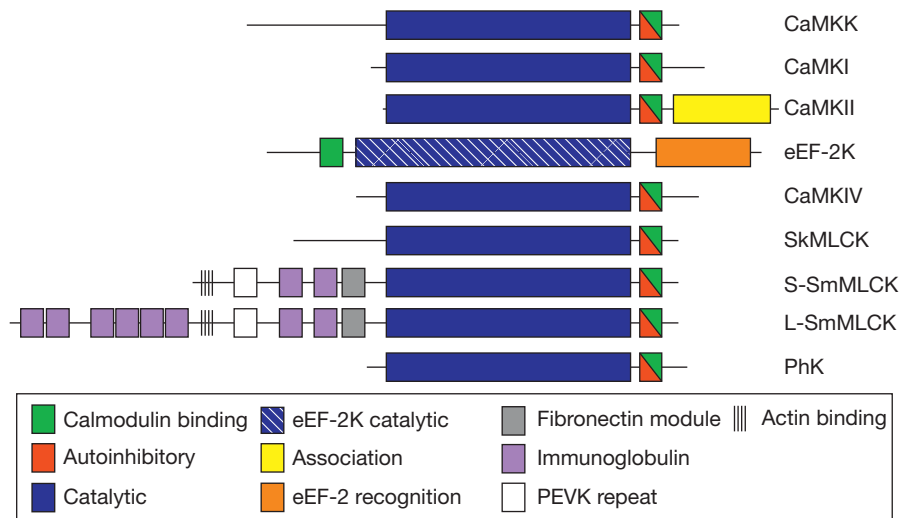


Figure 2 Schematic representation of CaMK domain structures. The key depicts various domains. The kinase domain of eEF-2K, formerly known as CaMKIII, bears no homology to those of the other CaMKs and thus is represented with white hatching. Most family members have multiple isoforms (not shown).

Narrow-Specificity Ca^{2+} /Calmodulin-Dependent Protein Kinases

Myosin Light Chain Kinase

Myosin is the only known *in vivo* substrate of myosin light chain kinase (MLCK), which is found in all vertebrate muscle cells and many nonmuscle tissues. Distinct genes encode three MLCK isoforms. Skeletal muscle MLCK (SkMLCK) is found

only in skeletal muscle and contains a catalytic core, an AID, a calmodulin-binding domain, and a short N-terminal extension. Cardiac MLCK (cMLCK) is expressed exclusively in the heart, has a structure similar to SkMLCK, and is essential for heart development and proper function. The smooth muscle MLCK (SmMLCK), which is expressed in multiple tissues, also contains an actin-binding domain, a fibronectin domain, a PEVK (region of protein rich in proline, glutamate, valine,

and lysine residues) repeat, and a number of immunoglobulin domains that can be varied by alternative splicing. MLCK is normally activated by Ca^{2+} release from the sarcoplasmic reticulum (SR) or endoplasmic reticulum and phosphorylates the regulatory light chain of the molecular motor myosin II, which is sufficient to initiate contraction in smooth muscle or to potentiate the force and speed of contraction in skeletal and cardiac muscle. Although all forms of MLCK are tightly regulated by calmodulin binding, SmMLCK can be further regulated by phosphorylation. Many kinases, including CaMKII, cyclic AMP-dependent protein kinase (PKA), cyclic GMP-dependent protein kinase, p21-activated kinase, and protein kinase C, can phosphorylate SmMLCK *in vitro* at two serine residues in the calmodulin-binding domain and thus decrease the affinity for calmodulin. In addition, two sites outside the catalytic and regulatory regions of SmMLCK can be phosphorylated by members of the mitogen-activated protein kinase family, resulting in an increase in $V_{\max}$ but no change in K_{CaM} . SmMLCK, but not SkMLCK, is targeted to actin/myosin filaments through interaction with actin subunits.

Phosphorylase Kinase

Phosphorylase kinase (PhK) phosphorylates and activates glycogen phosphorylase, thus accelerating glycogen degradation. Splice variations and gene duplications result in multiple isoforms, many of which are tissue specific. Isoforms of PhK are found in muscle, brain, heart, liver, intestines, and many other cell types. The highly abundant liver and muscle enzymes contribute to blood-glucose homeostasis and provide an energy source to facilitate muscle contraction, respectively. The catalytic γ -subunit associates with α , β , and δ regulatory subunits to form a 1.3-MDa holoenzyme that contains four copies of each subunit. The three regulatory subunits inhibit the phosphotransferase activity of the catalytic subunit. The δ -subunit is in fact calmodulin, which, unusually, is stably associated with the holoenzyme even at low Ca^{2+} concentrations. Ca^{2+} binding to δ partially activates PhK and phosphorylation of the α - and β -subunits by PKA potentiates activation. Binding of additional Ca^{2+} /calmodulin molecules, or Ca^{2+} -bound troponin C, to the holoenzyme maximally activates PhK. In muscle, PhK associates with a protein termed protein targeted to glycogen (PTG), which may facilitate phosphorylation of glycogen phosphorylase *b*, which is also associated with glycogen particles. PhK also interacts with other proteins in a tissue-specific manner, such as tubulin in rat brain lysates, and is found in membrane preparations from various tissue types.

Eukaryotic Elongation Factor 2 Kinase

The ubiquitous 90 kDa eukaryotic elongation factor 2 kinase (eEF-2 K), also known as CaMKIII, phosphorylates Thr-56 in eEF-2, a protein elongation factor that promotes ribosomal translocation along mRNA during translation. Interestingly, the catalytic domain of eEF-2 K bears no homology to conventional kinases and is a member of the α -kinase family. Since eEF-2 is inactivated by Thr-56 phosphorylation, eEF-2 K mediates a global reduction of protein synthesis following Ca^{2+}

elevation. Paradoxically, this global inhibition of protein synthesis can enhance the translation of specific mRNAs that are inefficiently translated under normal conditions. The eEF-2 K contains a calmodulin-binding domain and an eEF-2 recognition domain. Regulation of eEF-2 K is both subtle and complex. In the presence of Ca^{2+} and calmodulin, autophosphorylation of eEF-2 K occurs at several serine residues and confers a partially Ca^{2+} -independent activity. In addition, phosphorylation by a variety of kinases *in vitro* can both up- and downregulate eEF-2 K activity.

Phosphorylation by PKA confers Ca^{2+} independence. Thus, stimulation of β -adrenergic receptors and subsequent cyclic adenosine monophosphate (AMP) elevation causes eEF-2 K activation and inhibition of protein synthesis in cultured mammalian cells. In addition, studies suggest that slight drops in cellular pH (such as those caused by hypoxia and ischemia) cause activation of eEF-2 K. In combination, these data suggest that this kinase may not be a simple, direct link between Ca^{2+} and translation, but rather may function to integrate multiple intracellular signals.

Broad-Specificity Ca^{2+} /Calmodulin-Dependent Protein Kinases

Ca^{2+} /Calmodulin-Dependent Protein Kinase I

Ca^{2+} /calmodulin-dependent protein kinase I (CaMKI) is a ubiquitous cytoplasmic enzyme enriched in brain, liver, and intestine. Three isoforms of the kinase, α , β , and γ , have been reported and are products of separate genes, each of which may be alternatively spliced. CaMKI is an approximately 42 kDa monomer and its catalytic domain shows sequence and structural homologies with those of other serine/threonine kinases. A kinase originally termed CaMKV has been shown to be a splice variant of CaMKI. Goldberg et al. crystallized CaMKI, demonstrating that the AID binds and distorts the ATP-binding pocket in the basal state. Binding of Ca^{2+} /calmodulin presumably disrupts these interactions. A threonine residue in the catalytic domain (activation loop) of CaMKI is then phosphorylated by an upstream kinase, Ca^{2+} /calmodulin-dependent protein kinase kinase (CaMKK), resulting in a 10- to 20-fold increase in CaMKI catalytic activity without conferring Ca^{2+} /calmodulin independence. However, this phosphorylation can occur only when both CaMKK and CaMKI are bound by Ca^{2+} /calmodulin, perhaps permitting synergism in kinase activation. CaMKI phosphorylates synapsin I, cAMP response element binding protein (CREB) *in vitro*, and CaMKI activates the extracellular signal-regulated kinase (ERK) MAP kinase pathway and CREB-dependent gene transcription. A recent study found that CaMKI phosphorylates and regulates the guanine nucleotide exchange factor (GEF) β isoform of the PAK-interacting exchange factor (β PIX) to modulate the F-actin cytoskeleton. The activation of CaMKI was recently shown to enhance growth cone motility and the development of neuronal processes (axons, dendrites, and spines/synapses), in part via its actions on ERK, CREB, and β PIX.

Ca^{2+} /Calmodulin-Dependent Protein Kinase II

CaMKII is a ubiquitous kinase that has been implicated in the regulation of diverse physiological processes, such as gene

transcription, neuronal plasticity, exocytosis, and metabolism. Four separate genes encode the α -, β -, γ -, and δ -isoforms of CaMKII. Each isoform is subjected to alternative mRNA splicing, resulting in over 20 known variants of the kinase, many of which have specific cellular and tissue distributions. In addition to highly conserved catalytic, AID, and calmodulin-binding domains, most CaMKII isoforms contain a C-terminal association domain that facilitates the formation of a dodecameric holoenzyme, and may contain multiple isoforms or splice variants. An X-ray crystal structure of the core of the CaMKII α holoenzyme was recently reported, revealing a stacked pair of hexameric rings. The holoenzyme structure of CaMKII facilitates complex regulatory properties. Each subunit is independently activated by Ca^{2+} /calmodulin binding. When adjacent subunits in a holoenzyme bind Ca^{2+} /calmodulin, there is a trans-autophosphorylation at Thr-286, resulting in Ca^{2+} independence and also a 1000-fold increase in affinity for calmodulin. The kinase can be inactivated only by the actions of cellular protein phosphatases. Additional autophosphorylation at Thr-305 and Thr-306 in the calmodulin-binding domain occurs in the absence of Ca^{2+} /calmodulin. Thr-305/Thr-306 phosphorylation interferes with calmodulin binding, blocking Ca^{2+} /calmodulin-dependent activation. Recently reported X-ray crystal structures of a monomeric catalytic/regulatory domain fragment of CaMKII α provide insights into these dynamic regulatory properties.

The number of subunits in a holoenzyme that become active and autophosphorylated at Thr-286 is related to the frequency and amplitude of Ca^{2+} spikes. This unique capacity to decode complex and dynamic Ca^{2+} signals has been proposed to play an important role in prolonged signaling after transient changes in Ca^{2+} , such as may be involved in learning and memory. Consistent with this idea, transgenic mice in which the CaMKII α gene has been mutated such that the kinase cannot autophosphorylate at Thr-286 or at Thr-305/Thr-306 are deficient in spatial learning tasks.

CaMKII can be found in the cytosol, in the nucleus, and associated with various organelles or membranes, depending on the cell type examined. The identity of CaMKII splice variants expressed is believed to play a role in modulating subcellular localization, although few specific examples have been documented. For example, one variant of CaMKII δ contains a nuclear localization sequence and has been shown to regulate gene transcription in cardiac myocytes. Variable subcellular localization can also be attributed in part to the large number of proteins with which CaMKII may associate, which we have termed CaMKII-associated proteins (CaMKAPs). These include several neuronal proteins (e.g., N-methyl-D-aspartate (NMDA) receptor, α -actinin, synapsin I, and densin-180) and contractile/motor proteins (e.g., F-actin and myosin V). Binding of CaMKII to these proteins may position the kinase close to Ca^{2+} sources, and/or close to specific substrates, and may modulate its availability to cellular phosphatases, thereby affecting the rate of inactivation. However, recent studies suggest that some biological roles of full-length CaMKII isoforms may not require catalytic activity, but may instead depend on the ability to serve as a scaffold to assemble multiprotein complexes, perhaps explaining the very high expression levels of CaMKII in some cells. In this regard, it is interesting that one splice variant of the CaMKII α gene (α KAP) lacks a catalytic domain and functions as a

CaMKAP in muscle by forming hetero-multimers with the conventional isoforms through an intact association domain that are targeted to the SR by a unique N-terminal hydrophobic region.

At last count, CaMKII phosphorylates at least 50 distinct proteins *in vitro* and many of these are likely to be physiologically relevant. However, relatively few of these have been formally shown to be phosphorylated in intact cells under physiologically relevant conditions. One of the best characterized substrates is the neuronal α -amino-3-hydroxy-5-methylisoxazole-4-propionate (AMPA)-type glutamate receptor; phosphorylation of this receptor at Ser-831 by CaMKII plays an important role in the regulation of synaptic transmission. CaMKII can also phosphorylate numerous other proteins in neuronal dendrites to regulate excitability, dendritic morphology, and local mRNA translation. In neuronal presynaptic terminals, phosphorylation of synapsin I by CaMKII modulates synaptic vesicle movement. In the heart, CaMKII plays a critical role in feedback regulation of Ca^{2+} dynamics by regulating Ca^{2+} entry via L-type voltage-gated Ca^{2+} channels, by phosphorylating ryanodine receptors to regulate SR Ca^{2+} release, by phosphorylating phospholamban at Thr-17, thereby regulating SR Ca^{2+} uptake via the Ca^{2+} -ATPase, and by phosphorylating histone deacetylase to modulate gene transcription. Recent work has shown that CaMKII inhibition protects cardiomyocytes from developing multiple forms of heart disease.

Ca²⁺/Calmodulin-Dependent Protein Kinase IV

α CaMKIV is a monomeric kinase expressed largely in neuronal tissues, testes, and T cells, whereas the β CaMKIV splice variant is differentially expressed in the cerebellum during development. Like CaMKI, CaMKIV is initially activated by calmodulin binding and is further activated when phosphorylated by CaMKK in its activation loop. However, combined with a subsequent N-terminal autophosphorylation, the activation by CaMKK eventually leads to Ca^{2+} -independent activity of CaMKIV. An additional autophosphorylation at Ser-332 of the calmodulin-binding domain prevents further Ca^{2+} /calmodulin binding in a manner analogous to CaMKII autophosphorylation at Thr-305 and Thr-306. CaMKIV contains a nuclear localization sequence and it is thought to be responsible for Ca^{2+} -dependent phosphorylation of a variety of transcription factors (e.g., CREB, serum response factor, and myocyte enhancer factor 2 (MEF2)) and regulatory factors, such as the CREB-binding protein. In the cytosol, CaMKIV phosphorylates oncoprotein 18, preventing its association with tubulin. Protein phosphatase 2A (PP2A), a ubiquitous multimeric protein phosphatase, associates with CaMKIV in a manner antagonized by Ca^{2+} /CaM, dephosphorylating and deactivating it. This calcium-dependent self-regulating complex may be a prototype for phosphatase-kinase cross-talk in cell signaling. Multiple studies have exploited CaMKIV knockout mice to demonstrate a key role in fear-conditioned memory formation, though recent studies have suggested that other genes may compensate for the loss of CaMKIV in these mice.

Ca²⁺/Calmodulin-Dependent Protein Kinase

CaMKK, originally characterized as an activating factor from brain extracts, can dramatically increase the activity of CaMKI

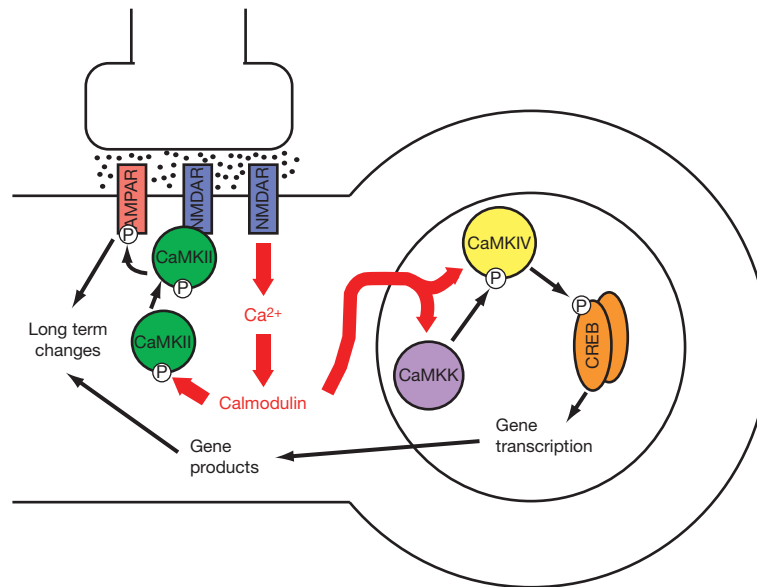


Figure 3 Modulation of neuronal function by activation of CaMKII and CaMKIV. Glutamate release from presynaptic terminals activates NMDA- (blue) and AMPA- (red) type glutamate receptors on the postsynaptic neuron. The NMDA receptor (NMDAR) is a ligand-gated Ca^{2+} channel. The elevated local intracellular Ca^{2+} in the dendrites activates CaMKII, which phosphorylates AMPA-type glutamate receptors, potentiating synaptic transmission. Ca^{2+} elevation via NMDA- and voltage-gated Ca^{2+} channels (not shown) also signals to the nucleus to activate CaMKK and CaMKIV, driving the transcription of specific genes. The dual CaMKII and CaMKIV signals combine to promote long-term changes in synaptic function. AMPAR, AMPA receptor.

and CaMKIV when incubated in the presence of Ca^{2+} /calmodulin, Mg^{2+} , and ATP. Two isoforms (α and β) of CaMKK have been cloned and their domain organization and function are similar to those of other Ca^{2+} /calmodulin-dependent kinases. However, CaMKK is unique in its method of substrate recognition, lacking acidic residues to recognize basic residues near its preferred phosphorylation sites. Instead, this kinase contains an arginine- and proline-rich (RP) insert that is important for phosphorylation of CaMKI and CaMKIV, and Ca^{2+} /calmodulin must be bound to both CaMKK and the substrate (CaMKI or CaMKIV) for phosphorylation to occur. CaMKK also phosphorylates protein kinase B both *in vitro* and in cultured cells, although this does not require the RP domain, and the AMP-dependent kinase (AMPK), which regulates multiple pathways underlying food intake and metabolic balance in several cell types. Regulation of CaMKK also occurs via other kinases, such as PKA, which phosphorylates CaMKK in its calmodulin-binding domain, resulting in the inhibition of Ca^{2+} /calmodulin-induced activity. Recently, experiments using knockout mice have revealed that CaMKK α plays a role in contextual memory consolidation, while CaMKK β is required for spatial memory consolidation. Surprisingly, these results are true only for male mice, as hippocampal function is unperturbed in female mutants.

Summary

Repetitive transient changes in Ca^{2+} concentration in every mammalian cell are sensed by a number of different Ca^{2+} /calmodulin-dependent protein kinases. Although many of these kinases share a great deal of functional homology, there are diverse regulatory nuances. Moreover, some of these

kinases phosphorylate a number of different substrates, whereas others have only one principal substrate. The specific response of individual cells to changes in intracellular Ca^{2+} concentration likely depends on the subset and specific isoforms of Ca^{2+} /calmodulin-dependent kinases expressed in that cell. For example, in hippocampal neurons, synaptic activity increases Ca^{2+} levels in dendritic regions. In the short term, dendritic Ca^{2+} locally activates CaMKII, which phosphorylates preexisting synaptic glutamate receptors. However, over a longer time frame, Ca^{2+} signals the activation of nuclear CaMKIV and thus the transcription of specific genes. It is the combined actions of CaMKII and CaMKIV that result in long-lasting changes in the synaptic response of these neurons that may underlie learning and memory (Figure 3). However, much work remains to be performed in identifying substrates and physiological roles of the various Ca^{2+} /calmodulin-dependent protein kinases in other cells and tissues.

See also: **Bioenergetics:** Calcium/Calmodulin-Dependent Protein Kinase II; **Mitochondrial Membranes, Structural Organization;** **Signaling:** Protein Kinase C Family.

Further Reading

- Berridge MJ, Lipp P, and Bootman MD (2000) The versatility and universality of calcium signaling. *Nature Reviews* 1: 11–20.
- Brushia RJ and Walsh DA (1999) Phosphorylase kinase: The complexity of its regulation is reflected in the complexity of its structure. *Frontiers in Bioscience* 4: 618–641.
- Colomer J and Means AR (2007) Physiological roles of the Ca^{2+} /CaM-dependent protein kinase cascade in health and disease. *Subcellular Biochemistry* 45: 169–214.

- Goldberg J, Nairn AC, and Kuriyan J (1996) Structural basis for the autoinhibition of calcium/calmodulin-dependent protein kinase I. *Cell* 84: 875–887.
- Hook SS and Means A (2001) Ca^{2+} /CaM-dependent kinases: From activation to function. *Annual Review of Pharmacology and Toxicology* 41: 471–505.
- Hudmon A and Schulman H (2002) Neuronal Ca^{2+} /calmodulin-dependent protein kinase II: The role of structure and autoregulation in cellular function. *Annual Review of Biochemistry* 71: 473–510.
- Kamm KE and Stull JT (2001) Dedicated myosin light chain kinases with diverse cellular functions. *Journal of Biological Chemistry* 276: 4527–4530.
- Lisman J, Schulman H, and Cline H (2002) The molecular basis of CaMKII function in synaptic and behavioural memory. *Nature Reviews. Neuroscience* 3: 175–190.
- Proud CG (2006) Regulation of protein synthesis by insulin. *Biochemical Society Transactions* 35: 213–216.
- Rosenberg O, Deindl S, Sung R, Nairn A, and Kuriyan J (2005) Structure of the autoinhibited kinase domain of CaMKII and SAXS analysis of the holoenzyme. *Cell* 123: 849–860.
- Wayman G, Lee Y, Tokumitsu H, Silva A, and Soderling T (2008) Calmodulin-kinases: Modulators of neuronal development and plasticity. *Neuron* 59: 914–931.

Calcium in the Regulation of Gene Expression

M Rivas and J R Naranjo, Centro Nacional de Biotecnología – CSIC, Madrid, Spain

© 2013 Elsevier Inc. All rights reserved.

Glossary

Histone acetyltransferases (HATs) Enzymes that acetylate lysine amino acids on histone proteins by transferring an acetyl group from acetyl CoA.

Histone deacetylases (HDACs) Enzymes that remove acetyl groups from acetylated lysine amino acids on histones.

Protein kinase An enzyme that transfers a phosphate group from adenosine triphosphate (ATP)

to a serine, threonine, or tyrosine residue of a substrate protein.

Protein phosphatase An enzyme that removes a phosphate group from a serine, threonine, or tyrosine residue of a phosphorylated protein substrate.

Transcription factors Proteins that bind directly to sequence-specific DNA regions, thereby controlling the transcriptional activity of their target genes.

Introduction

The control of gene expression has evolved to respond to the environmental and intracellular cues so that cells can adjust their growth patterns and activities, and ultimately make life or death decisions. Cells are exposed to both intrinsic and extrinsic signals in the form of chemical and/or physical factors that are presented to them from either side of the plasma membrane. These signals are converted into intracellular messengers through the actions of receptors, ion channels, and/or effector enzymes. Among these messengers, Ca^{2+} is a central player in many cellular signal-transduction cascades that modulate gene transcription. Due to space limitations, we include only reviews in the field.

Ca^{2+} Homeostasis

The resting concentration of Ca^{2+} in the cytoplasm is normally kept in the range of 10–100 nM. To accomplish this, Ca^{2+} is actively pumped from the cytosol to the extracellular space and into intracellular organelles, mainly the sarco/endoplasmic reticulum (S/ER) and the mitochondria. Ca^{2+} signaling occurs when the cell is stimulated to release Ca^{2+} from intracellular stores and/or when calcium enters the cell through plasma membrane ion channels. Cytosolic Ca^{2+} signals can be evoked by chemical stimuli (hormones, growth factors, and toxins), environmental changes (changes in pH or temperature), mechanical deformation, and depolarization. Elevations of intracellular free Ca^{2+} are generally short-lived (milliseconds to several minutes) due to an elaborated network of calcium buffers, pumps, channels, and exchangers in the plasma membrane, as well as membranes of intracellular Ca^{2+} -storage organelles. As an example, plasma membrane Ca^{2+} -ATPases pump Ca^{2+} from the cytoplasm to the extracellular space, while S/ER Ca^{2+} -ATPases (SERCA) transfer Ca^{2+} from the cytoplasm into the S/ER lumen. In addition, cytosolic Ca^{2+} -binding proteins (e.g., calbindin, parvalbumin, and calretinin) contribute to buffer-free Ca^{2+} , while luminal Ca^{2+} -binding proteins (e.g., calreticulin and calsequestrin) enhance the storage capacity of S/ER for Ca^{2+} .

Different types of Ca^{2+} channels mediate release of Ca^{2+} from intracellular stores, of which inositol(1,4,5)-trisphosphate

receptors (InsP3Rs) and ryanodine receptors (RyRs) are the best characterized. Depletion of Ca^{2+} from the endoplasmic reticulum leads to Ca^{2+} entry from outside the cell by activation of 'store operated channels'. The molecular mechanism involved in the communication between the ER and the plasma membranes has recently been disclosed. At the ER, the STIM1 protein senses the depletion of intracellular Ca^{2+} stores and communicates this state to the plasma membrane. The mechanism involves an orchestrated series of events, including the aggregation into macromolecular complexes with the plasma membrane protein Orai, as well as with other molecules such as the TRPC1 channel, SERCA and the microtubule end tracking protein EB1, to induce activation of Ca^{2+} influx.

Changes in cytosolic-free Ca^{2+} underlie a multitude of basic cellular functions ranging from extremely rapid, such as neurosecretion and skeletal muscle contraction, to delayed responses, such as cellular differentiation and mitogenesis. It is well known that different cell types use various elements from a broad Ca^{2+} signaling 'toolkit' and that the characteristics of the global Ca^{2+} signals, and their physiological effects, can vary considerably. Downstream the Ca^{2+} signal, different signaling cascades, including the Ras/mitogen-activated protein kinase (MAPK) pathway, calcium/calmodulin (CaM)-dependent protein kinases, the phosphatase calcineurin, and Rac GTPases, convey the information to mediate changes in gene expression by stimulating the translocation of transcription factors from the cytosol to the nucleus, by causing the translocation and/or activation of enzymes that regulate the activity of nuclear transcription factors or the structure of chromatin.

The Nuclear Calcium Signal

For many years, it has been accepted that the propagating cytosolic Ca^{2+} wave enters the nucleus through the nuclear pores of the nuclear envelope (NE) by passive diffusion. Although undisputable, this mechanism, however, is not exclusive and recent studies have highlighted the existence of an independent nuclear Ca^{2+} compartment nourished by Ca^{2+} signals generated locally.

Like in the cytosol, a nuclear Ca^{2+} toolkit contributes to the genesis of the nuclear Ca^{2+} signal, though it is still not well understood how the nuclear Ca^{2+} signal is generated in the

nucleoplasm and whether the nuclear signal is autonomous or somehow influenced by cytosolic cues. To date, the most widely implicated mechanism for nuclear Ca^{2+} release is via the activation of InsP3 receptors, which could occur independently of cytosolic messengers by mechanisms parallel to those described at the plasma membrane. Activation of phospholipase C (PLC) enzymes at the nuclear membrane leads to the hydrolysis of the membrane phospholipid phosphatidyl inositol-bisphosphate to form inositol 1,4,5-trisphosphate (IP3) and diacylglycerol (DAG). DAG activates the protein kinase C (PKC) enzyme, while IP3 diffuses to bind to InsP3 receptors and release Ca^{2+} . In the nucleus, InsP3 receptors are present in the NE and trigger Ca^{2+} release from the NE lumen directly into the nucleoplasm. InsP3 receptors are located also in small vesicular Ca^{2+} stores within chromatin, which could represent nuclear InsP3-sensitive Ca^{2+} stores. Other receptors such as RyR and nicotinic acid adenine dinucleotide phosphate receptors are also present in the NE in various cell types and regulate Ca^{2+} signaling in the nucleus.

The mechanism by which nucleoplasmic Ca^{2+} signals decline back to prestimulation levels has not been entirely clarified. It is known, however, that nuclear Ca^{2+} signals persist considerably longer in the nucleoplasm than in the cytosol. Reduced nuclear Ca^{2+} -buffering capacity in comparison to the cytosol and the absence of nuclear Ca^{2+} ATPases could, in part, explain the slower decay of the nuclear Ca^{2+} signal. Nuclear Ca^{2+} transients, nevertheless, appear to be reversed by passive diffusion through nuclear pore complexes and by uptake into the NE by sodium-calcium exchangers expressed on the inner face of the NE.

Nuclear Mediators of Calcium Signaling

Calcium Sensors

Elevation of cytosolic Ca^{2+} is translated into functional changes in cellular activity by specific sensors that bind Ca^{2+} with different affinities in different subcellular compartments. CaM, S-100 proteins, transcriptional repressor downstream regulatory element antagonist modulator (DREAM), and calreticulin, thus, interact with various nuclear targets in a Ca^{2+} -dependent manner to regulate the activity of transcription factors or enzymes that influence chromatin structure to eventually modify gene expression.

CaM is a ubiquitous protein, although constitutively present in the nucleus. Increase in cytosolic Ca^{2+} enhances CaM transport to the nuclear compartment. Binding of Ca^{2+} produces a conformational change in CaM that promotes interactions of the Ca^{2+} /CaM complex to numerous target proteins. The list of CaM effectors includes the CaM-dependent adenylyl cyclases, phosphodiesterases, protein kinases (CaMK), and the protein phosphatase calcineurin. CaM also regulates the activities of the plasma-membrane Ca^{2+} pump, various ion channels, RyR, and isoforms of the InsP3 receptors, as well as the nuclear import and transcriptional activity of transcription factors such as Sox 9 and several basic/helix-loop-helix (bHLH) proteins. CaM, thus, participates in signaling pathways that regulate many crucial processes such as growth, proliferation, and movement.

Calreticulin is a molecular chaperone located in storage compartments that binds Ca^{2+} with low affinity but high capacity. Although predominantly within the ER, it is also present in

the nucleus where it interacts with the DNA-binding domain of several nuclear receptors, including glucocorticoid, androgen, retinoic acid, and vitamin D receptors. As a result of these interactions, calreticulin inhibits nuclear receptor-mediated transcription. In addition, calreticulin mediates Ca^{2+} -dependent nuclear export of glucocorticoid receptors reducing their availability to mediate transcription.

The S100 protein family comprises at least 25 members, forming the largest group of EF-hand-signaling proteins in humans. They are expressed in many tissues; however, each S100 protein is frequently restricted to one particular tissue or cell type. S100 proteins can form homo- and heterodimers and interact with a variety of target proteins usually in a Ca^{2+} -dependent manner. S100 proteins have been reported to modulate the activity of bHLH and p53 transcription factors and are involved in many biological processes such as cell growth and motility, cell-cycle regulation, differentiation, and cell survival.

Protein Kinases

Activation of kinases and phosphatases constitutes the best-described mechanism by which changes in intracellular Ca^{2+} concentration can affect gene expression.

The CaMK kinase (CaMKI, CaMKII, and CaMKIV) family members contain a catalytic domain, an autoregulatory domain, and an overlapping CaM-binding domain. Binding to Ca^{2+} /CaM triggers kinase activity by facilitating a conformational change that removes the autoregulatory domain from the catalytic pocket to enable substrate access. CaMKs are particularly abundant in brain where they play critical roles in neuronal development, plasticity, and behavior. CaMKs have been associated with activity-dependent direct phosphorylation of several transcription factors, including cyclic adenosine monophosphate (cAMP) response element-binding protein (CREB), CREB-binding protein (CBP), and E-twenty six-1 transcription factor (Ets1), as well as with the nuclear translocation of nuclear factor kappa-light-chain-enhancer of activated B cells (NF κ B) and silencing mediator of retinoic acid and thyroid hormone receptors (SMRTs).

PKC family members can be divided into four structurally and functionally distinct subgroups specified by their divergent regulatory domains. The conventional PKC isoforms (α , β I, β II, and γ) are activated by a combination of DAG and phospholipid binding to their conserved region 1 (C1) domains and Ca^{2+} -dependent phospholipid binding to their C2 domains. The novel PKCs (δ , ϵ , η , and θ), the atypical PKCs (ζ , λ /I, and μ) and the PKN subfamily members (PKN1, PKN2, and PKN3) do not depend on Ca^{2+} or diacylglycerol for activation.

Elevated concentrations of intracellular Ca^{2+} activate conventional PKC isoforms, recruiting the enzyme to the inner leaflet of the plasma membrane or the NE within minutes. This process has been associated with changes in gene expression by two different mechanisms. The first involves phosphorylation of kinase of the inhibitory kappa B subunit (I κ B) followed by nuclear translocation of NF κ B which activates cytokine gene expression and T-cell proliferation. The second mechanism relates to increased nuclear activity of histone acetyltransferases (HATs) and has been described in epidermal cells where Ca^{2+} induces epidermal cell differentiation.

Protein Phosphatases

As important as phosphorylation events, Ca^{2+} -dependent dephosphorylation of specific residues in key proteins within a given transcriptional network also induces drastic changes in the activity of the network. Interestingly, activity-dependent phosphorylation and dephosphorylation are, in some occasions, coupled events that act on a particular protein or even on a particular phosphoresidue. In those cases, the specificities of the calcium signal to activate the two events alternatively require a fine-tuning that is not always well understood.

Calcineurin is a Ca^{2+} /CaM-activated serine/threonine phosphatase, with a catalytic A subunit that binds CaM and a regulatory B subunit that binds calcium. Calcineurin conveys cytosolic Ca^{2+} signals to the nucleus, regulating the activity of at least three important transcription pathways involving nuclear factor of activated T-cell (NFAT), CREB, and myocyte enhancer factor-2 (MEF-2) proteins, which control responses in the immune, nervous, and cardiovascular systems. Calcineurin may also regulate with gene expression by changing the concentration of intracellular-free calcium by dephosphorylating phospholamban, the endogenous regulator of the SERCA2a calcium pump, ryanodine receptors, as well as the IP3R1. Calcineurin has a particularly high affinity for Ca^{2+} /CaM and is at least one order of magnitude more sensitive to Ca^{2+} /CaM than is CaMKII. As a consequence, calcineurin responds to sustained, low-amplitude calcium transients, whereas transient, high-amplitude calcium spikes preferentially activate CaMKII.

PP1 (calcium/calcineurin-dependent type 1 serine/threonine protein phosphatase) contains a highly conserved catalytic subunit and one or more regulatory subunits. PP1 activity is modulated by the interaction with auxiliary proteins that target the enzyme to different subcellular compartments, conferring substrate specificity. A subset of these auxiliary proteins are inhibitory proteins that block the catalytic activity of PP1. These include PP1 inhibitors 1 and 2, dopamine and cAMP-regulated phosphoprotein 32kDa (DARPP-32), as well as protein phosphatase 1 nuclear-targeting subunit (PNUTS) and nuclear inhibitor of protein phosphatase 1 (NIPP-1), two inhibitory subunits of PP1 that direct the nuclear targeting and retention of the enzyme, regulating PP1 nuclear activity. Binding to PP1 and inhibition of its activity are regulated by protein kinase A (PKA) phosphorylation, while calcineurin dephosphorylates inhibitor-1 blocking its activity, which results in PP1 activation. A major role of PP1 activity in the nucleus is to dephosphorylate the transcription factor CREB and terminate CREB-dependent gene expression.

Final Effectors of Nuclear Calcium

Transcription Factors

Nuclear factor of activated T-cells

The NFAT family of transcription factors, NFAT1 (NFATp, NFATc2), NFAT2 (NFATc, NFATc1), NFAT3 (NFATc4), NFAT4 (NFATx, NFATc3), and NFAT5 (TonEBP), are proteins that are evolutionary related to the Rel/NF κ B family. Originally characterized in immune cells, it is now known that all isoforms are ubiquitously expressed and appear to have relatively redundant functions. NFATs induce the expression of genes

important in cellular processes such as development, activation of lymphocytes, differentiation of cardiac muscle cells, and cancer progression.

A key rate-limiting event in the activation of NFATs is a rise in intracellular calcium concentrations. Ca^{2+} /CaM activates calcineurin that induces NFAT nuclear translocation and transcriptional activation. NFAT5 does not possess a calcineurin-binding site, and, therefore, is insensitive to calcium and calcineurin.

Phosphorylation by maintenance kinases such as dual-specificity tyrosine phosphorylation-regulated kinase 2 (DYRK2) and casein kinase 1 prevents NFAT nuclear translocation. By contrast, phosphorylation by export kinases such as DYRK1 and glycogen synthase kinase 3 promotes NFAT export from the nucleus (Figure 1). The process as a whole is extremely rapid, giving NFAT proteins the remarkable ability to sense dynamic changes in intracellular Ca^{2+} levels and frequencies of Ca^{2+} oscillations. In addition to phosphorylation, other posttranslational modifications, such as sumoylation and ubiquitination, have been reported to regulate the cytoplasmic-nuclear trafficking of some NFAT family members.

The MEF-2 family

The MEF-2 family of transcription factors (MEF-2 A–D) includes calcium-dependent regulators of cell division, differentiation, and death. MEF2 has been studied most extensively in muscle cells, though MEF2 is also an important calcium-dependent regulator of neuronal and immune cell differentiation and function. Unlike NFAT, MEF2 is constitutively bound to its cognate DNA-binding elements independently of Ca^{2+} . In resting conditions, MEF2 recruits transcriptional repressors, including Cabin1, MEF2-interacting transcriptional repressor, and histone deacetylases (HDACs) to repress gene expression. So far, two different mechanisms have been described by which MEF2 repression can be relieved. First, after Ca^{2+} influx, the co-repressors are removed from MEF2 by binding to activated CaM, and MEF2 binds instead to co-activators (e.g., p300), resulting in the transcription of target genes. Second, phosphorylation of HDACs by CaMK I and IV creates a phosphodomain that binds to 14–3–3 proteins, mediating the shuttling of HDACs from the nucleus to the cytosol, thus freeing MEF2 to be phosphorylated by MAP kinases, which leads to the full transcriptional activity of MEF2. Activation of nonrepressed MEF-2 proteins is also a Ca^{2+} -dependent process, for which two major mechanisms have been proposed. First, as shown in cerebellar neurons, calcium influx triggers activation of MKK6-p38 MAPK resulting in the phosphorylation of Ser-387 located in the transactivation domain of MEF-2 C. Second, the Ca^{2+} /CaM-dependent phosphatase calcineurin activates MEF-2 through a posttranscriptional mechanism that either increases its DNA-binding activity in cerebellar granules or triggers interaction of MEF-2 and NFAT proteins and recruitment of CBP to these complexes in T lymphocytes. Ca^{2+} /CaM-dependent signaling, thus, prevents formation of the MEF-2/HDAC complexes, induces nuclear export of class II HDACs, activates MEF-2-dependent transcription, and, as a result, stimulates cytokine expression, myogenesis, and neuronal survival (Figure 1).

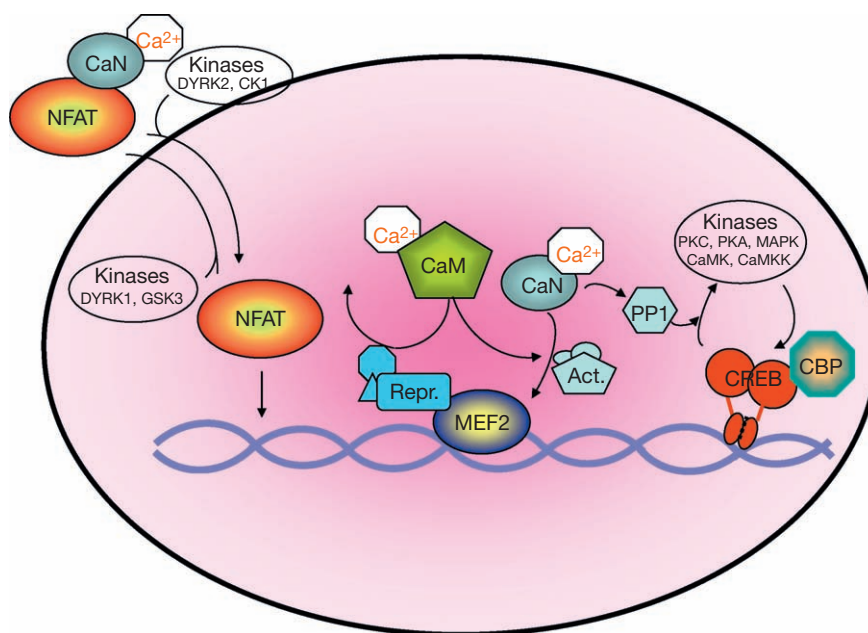


Figure 1 Calcium-modulated transcription factors. Nuclear-cytosolic translocation of NFAT is regulated by calcineurin (CaN), ‘maintenance’ kinases, and ‘export’ kinases. In resting conditions, MEF2 is bound to DNA forming a complex with transcriptional repressors (Repr.), while calmodulin (CaM) promotes removal of repressors and binding of transcriptional activators (Act.) together with CaN. CREB transcriptional activity is regulated by Ca-dependent and cAMP-dependent kinases as well as phosphorylation-independent events.

Nuclear factor kappa-light-chain-enhancer of activated B cells

NFκB is expressed in nearly all cell types and controls the expression of genes involved in a broad range of biological processes such as immune, inflammatory, and stress responses or neuronal survival. NFκB is a transcription factor composed of diverse combinations of p50, RelA/p65, c-Rel, RelB, and p52. NFκB is normally found in the cytoplasm as an heterodimer retained by the interaction with an inhibitory kappa B (IκB) molecule that masks its nuclear localization signal. Activation of the NFκB signaling cascade results in specific IκB hyperphosphorylation by two IκB kinases, IKKα and IKKβ, and subsequent degradation of IκB by the proteasome complex. This releases NFκB, allowing its nuclear translocation. In addition, more recent studies in T lymphocytes and neurons identified CaMKII as an activator of NFκB.

The cAMP response element-binding protein

CREB is a basic leucine zipper (bZIP) transcription factor activated by both cAMP and Ca^{2+} signals. CREB is ubiquitously expressed and regulates the expression of a wide variety of genes. In the nervous system, where it has been most widely studied, CREB is involved in neuronal survival and synaptic plasticity. Several kinases, including PKA, PKC, MAPKs (ERK and p38), CaMK, and CaMKK phosphorylate CREB at Ser-133 in the transactivation domain, known as the kinase-inducible domain (KID). Phospho-KID facilitates recruitment of the co-activators CBP and p300, which enhances CRE-dependent transcription. CBP/p300 activity is as well modulated by several kinases, including CaMKI and the interaction with the Ca^{2+} -responsive transactivator. The Ca^{2+} signal does not need to propagate to the nucleus to induce phosphorylation of Ser133 on CREB; however, an increase in the level of nuclear Ca^{2+} is

required for CREB-mediated gene transcription to occur. This is because additional Ca^{2+} -dependent phosphorylation of CREB and its co-activators is required.

Calcineurin exerts either stimulatory or inhibitory effects on CREB-dependent gene expression in different cellular systems. In cultured rat hippocampal neurons, calcineurin activation of PP1 limits the duration of CREB phosphorylation and, thus, restricts the transcriptional responses of CREB activation. By contrast, inhibition of calcineurin in non-neuronal cells, such as pancreatic islet cells or chromaffin cells, can block CREB-dependent gene expression (Figure 1).

Phosphorylation of CREB is necessary, but not sufficient to stimulate CREB-dependent transcriptional activity. Moreover, different phosphorylation-independent mechanisms regulate CREB activity. CREB-regulated transcription co-activators (Crtcs) also known as transducers of regulated CREB activity (TORC) transducers of regulated CREB activity proteins, mediate calcium- and cAMP-dependent, phosphorylation-independent activation of CREB. Crtc interacts with the bZIP DNA-binding/dimerization domain of CREB enhancing its transcriptional activity. The pathway involves calcineurin and the repressor salt-inducible kinase. Also, in line with the phosphorylation-independent regulation of CREB activity, a Ca^{2+} -dependent interaction between transcriptional repressor DREAM and CREB or phosphoCREB regulates the accessibility of CBP to the KID domain in CREB, and so compromises CRE-dependent transcription.

The downstream regulatory element antagonist modulator

DREAM binds as a tetramer to specific sites in the DNA (DRE sites) and represses transcription of specific genes in a Ca^{2+} -dependent manner. The DRE sequence was defined as the element responsible for basal and induced expression of the

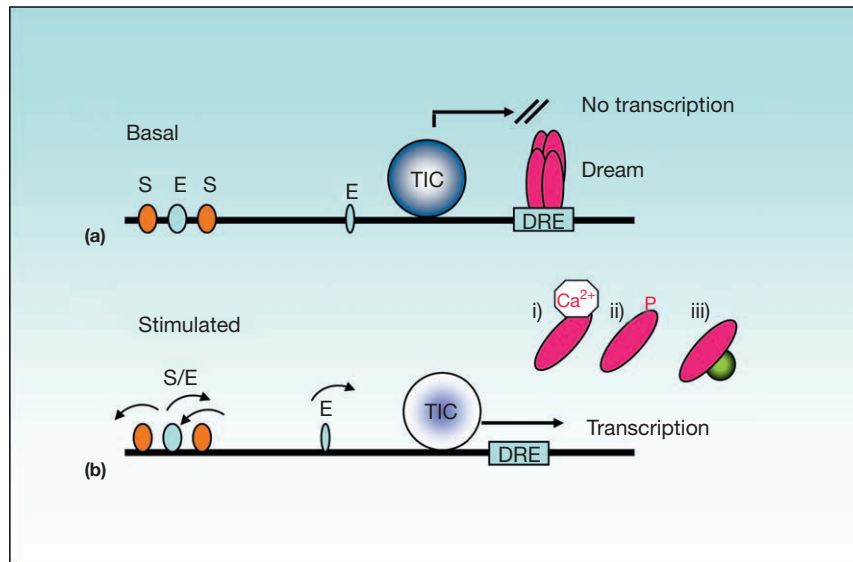


Figure 2 DREAM-mediated transcriptional repression. (a) Basal repression of DREAM target genes after binding of DREAM tetramers to the DRE site. (b) Stimulated derepression of DREAM-mediated repression by different stimuli including: (i) Ca^{2+} , (ii) DREAM phosphorylation, and (iii) the interaction of DREAM with other nucleoproteins such as αCREM . Once DREAM is detached from the DRE site, transcription of a given target gene will proceed driven by the specific combination of enhancers (E) and silencers (S) present in the promoter region. Adapted from Mellström B, Savignac M, Gomez-Villafuertes R, and Naranjo JR (2008) Ca^{2+} -operated transcriptional networks: Molecular mechanisms and *in vivo* models. *Physiological Reviews* 88: 421–449.

prodynorphin gene. More recent studies, however, have shown DRE sites and DREAM regulation of other targets including a number of immediate early transcription factors such as c-fos, c-jun, and ICER. Furthermore, binding of DREAM to DNA is regulated also by phosphorylation through PI3K and interaction with other nucleoproteins such as phosphoCREM, while the interaction with CREB or TTF-1 allows DREAM to regulate the transcription independently of DRE sites (Figure 2).

The bHLH family

The bHLH superfamily of transcription factors (that include MyoD, Myf5, and myogenin, among others) is essential for numerous developmental processes, including neurogenesis, hematopoiesis, and myogenesis as well as tumor cell invasion and metastasis. Based on tissue distribution, dimerization capacities, and DNA-binding specificities, bHLH proteins have been subdivided into seven classes. In general, bHLH proteins form homo- or heterodimers through the HLH domain and bind to regulatory E-box sequences in the DNA. Upon binding to DNA, most class I, II, and III bHLH dimers activate transcription through the recruitment of cofactors containing HAT activity, such as CBP/p300, p300/CBP-associated factor (PCAF) or the interaction with chromatin-remodeling factors of the mating-type switching and sucrose non-fermenting chromatin remodeler complexes (SWI-SNF) class. Conversely, transcriptional repression by bHLH proteins is related to the capacity to recruit nuclear complexes containing HDAC activity or to block HAT activity.

The regulation of bHLH proteins occurs at many levels, including differential oligomerization and DNA-binding specificities, interaction with negatively acting HLH proteins, phosphorylation, and other posttranslational modifications. Ca^{2+} /CaM regulates DNA-binding activity of bHLH proteins

by interacting with their DNA-binding domains or interfering with PKC-mediated phosphorylation. Moreover, binding of Ca^{2+} /CaM or S100 proteins to bHLH proteins may disrupt their interactions with components of the transcriptional machinery.

The calcium channel-associated transcriptional regulator

Cleavage of the COOH-terminal end of voltage-gated calcium channels (VGCCs) is a well-known phenomenon, and cleaved fragments have been implicated in regulating channel activity by reducing Ca^{2+} influx. Interestingly, the COOH-terminal intracellular fragment of the Cav1.2 subunit of the L-type VGCC, named calcium channel-associated transcriptional regulator (CCAT), was shown to localize in the nuclei of GABAergic inhibitory interneurons in the cortex and hippocampus and to regulate transcription of selective target genes (GABA, gamma aminobutyric acid). Importantly, the nuclear pool of CCAT is rapidly extruded in response to calcium entry following neuronal membrane depolarization. Transcriptional regulation by CCAT does not seem to involve direct interaction with DNA but rather involves interaction with other nucleoproteins, such as p54(nrb)/NonO, nuclear RNA- and DNA-binding protein of 54 kDa, also known as non-POU domain-containing octamer-binding protein.

Ca^{2+} -Dependent Changes in Chromatin Structure

Ca^{2+} -dependent enzymes not only modify the activity of specific transcriptional networks but also act on chromatin-associated proteins, histones (H) and high-mobility group (HMG) proteins, to evoke selective modifications in the nucleosome. The spatial relationship between DNA and the histone core of the nucleosome determines the accessibility of

transcription factors to the DNA and the transcriptional status of surrounding genes. The spatial association between DNA and histones is largely governed by posttranslational modifications of histones, such as acetylation, phosphorylation, methylation, and adenosine diphosphate-ribosylation.

Acetylation and deacetylation, linked predominantly to activation and repression, respectively, are to date probably the most extensively studied posttranslational histone modifications. Histone acetylation is regulated by a balance between HAT and HDAC activities. Several mechanisms are known to regulate these enzymes, including changes in subcellular localization, ubiquitylation-mediated degradation, changes in interactions with regulatory partners or cofactors, and Ca^{2+} /CaM- and CaMK-dependent phosphorylation.

Phosphorylation is a well-characterized posttranslational modification that induces activity-dependent changes in chromatin structure. Ca^{2+} increase induces the specific phosphorylation of histone H3, but not other histones, and of HMG protein-17 allowing gene transcription. Several Ca^{2+} -dependent kinases phosphorylate histone H3 and specific transcription factors to coordinately activate transcriptional networks. The system is reset by PP1, which has been implicated in H3 dephosphorylation.

Concluding Remarks

Recent years have witnessed an eruption of new calcium-sensitive proteins and calcium-dependent pathways that regulate gene expression and cellular activity. Remarkably, a significant progress has been made in unraveling Ca^{2+} -signaling systems in the nucleus. Yet, mechanisms that regulate Ca^{2+} release from nuclear pools as well as Ca^{2+} passage and regulation of nuclear pore complexes are open fields that require further investigation.

See also: **Molecular Biology:** Gene Expression in Eukaryotes: RNA Polymerase II Structure; Genome-Wide Analysis of Gene Expression.

Further Reading

- Aramburu J, Rao A, and Klee CB (2000) Calcineurin: From structure to function. *Current Topics in Cellular Regulation* 36: 237–295.
- Berridge MJ, Lipp P, and Bootman MD (2000) The versatility and universality of calcium signalling. *Nature Reviews Molecular Cell Biology* 1: 11–21.
- Bootman MD, Fearnley C, Smyrniak I, MacDonald F, and Roderick HL (2009) An update on nuclear calcium signalling. *Journal of Cell Science* 122: 2337–2350.
- Chin D and Means AR (2000) Calmodulin: A prototypical calcium sensor. *Trends in Cell Biology* 10: 322–328.
- Clapham DE (2007) Calcium signaling. *Cell* 131: 1047–1058.
- Gomez-Ospina N, Tsuruta F, Barreto-Chang O, Hu L, and Dolmetsch R (2006) The C terminus of the L-type voltage-gated calcium channel $\text{Ca}_v1.2$ encodes a transcription factor. *Cell* 127: 591–606.
- Israel A (2000) The IKK complex: An integrator of all signals that activate NF- κ B? *Trends in Cell Biology* 10: 129–133.
- Mayr B and Montminy M (2001) Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nature Reviews Molecular Cell Biology* 2: 599–609.
- McKinsey TA, Zhang CL, and Olson EN (2002) MEF2: A calcium-dependent regulator of cell division, differentiation and death. *Trends in Biochemical Sciences* 27: 40–47.
- Meffert MK, Chang JM, Wiltgen BJ, Fanselow MS, and Baltimore D (2003) NF- κ B functions in synaptic signaling and behavior. *Nature Neuroscience* 6: 1072–1078.
- Mellert HS and McMahon SB (2009) Biochemical pathways that regulate acetyltransferase and deacetylase activity in mammalian cells. *Trends in Biochemical Sciences* 34: 571–578.
- Mellstrom B, Savignac M, Gomez-Villafuertes R, and Naranjo JR (2008) Ca^{2+} -operated transcriptional networks: Molecular mechanisms and *in vivo* models. *Physiological Reviews* 88: 421–449.
- Oh-hora M and Rao A (2009) The calcium/NFAT pathway: Role in development and function of regulatory T cells. *Microbes and Infection* 11: 612–619.
- Rosse C, Linch M, Kermorgant S, et al. (2010) PKC and the control of localized signal dynamics. *Nature Reviews Molecular Cell Biology* 11: 103–112.
- Santamaria-Kisiel L, Rintala-Dempsey AC, and Shaw GS (2006) Calcium-dependent and -independent interactions of the S100 protein family. *Biochemical Journal* 396: 201–214.
- Vaca L (2010) SOC1C: The store-operated calcium influx complex. *Cell Calcium* 47: 199–209.
- Wayman GA, Lee YS, Tokumitsu H, Silva AJ, and Soderling TR (2008) Calmodulin-kinases: Modulators of neuronal development and plasticity. *Neuron* 59: 914–931.

Calcium-Modulated Proteins (EF-Hand)

R H Kretsinger, University of Virginia, Charlottesville, VA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 294–299, © 2004, Elsevier Inc.

Glossary

Calcium-modulated protein Calcium-binding protein found in the cytosol whose structure changes with binding and release of Ca^{2+} ions associated with a pulse of messenger calcium.

Calmodulin Protein consisting of two pairs of EF-hands. It is found in nearly all cells of all eukaryotes and interacts with over 40 different target enzymes or structural proteins.

Chimera Monomeric protein, consisting of two, or more, domain families.

Congruence Two (or more) subfamilies, each having two or more homologous domains, are congruent if the common precursor to all of these subfamilies had the full complement of domains and all of the subfamilies are

related by duplication of entire, encoding gene. Congruent proteins (subfamilies) are characterized by having all their domains 1 more similar to one another than to other domains within the same proteins; correspondingly all domains 2, etc.

EF-hand Protein domain of 30 residues (amino acids) consisting of an α -helix (E), a loop around a Ca^{2+} ion, and a second α -helix (F).

Ortholog Two homologous domains are orthologs if they were generated by speciation of an ancestral organism. If the two domains, in the same or in different species, are related by gene duplication, they are paralogs.

Transduction Process of changing energy and/or information from one modality to another.

Introduction

Many different proteins – extracellular, membrane, and intracellular – bind calcium, more or less selectively. Those calcium-binding proteins in the cytosol or bound to membranes facing the cytosol are inferred to be calcium modulated. That is, when the cell is quiescent, the concentration of the free Ca^{2+} ion is less than 10^{-7} M ($\text{pCa} > 7$) and the calcium-modulated protein is in the apo- or magnesi-form. Following stimulus, the concentration of calcium rises ($\text{pCa} < 5.5$) and the protein binds calcium. The attendant change in structure is involved in the transduction of the information of a pulse or wave of Ca^{2+} ions to an ultimate target enzyme or structure. Most of these calcium-modulated proteins contain from two to 12 copies of the EF-hand domain. There are other proteins in the cytosol that bind calcium and appear to be modulated by calcium; these include proteins that contain one or several C2 domains, such as protein kinases C or synaptotagmin, the annexins, as well as calcium pumps and channels.

Calcium Coordination

Many of the functional characteristics of calcium and of calcium-modulated proteins can be rationalized from the geometry of calcium coordination.

Pentagonal Bipyramid

In proteins, the Ca^{2+} ion (atomic radius 0.99 Å) is usually coordinated by seven oxygen atoms in an approximate pentagonal bipyramidal conformation at average Ca–O distance 2.3 ± 0.3 Å; the oxygen atoms have some lateral flexibility. The Mg^{2+} ion (atomic radius 0.65 Å) is usually bound by six oxygen atoms at the vertices of an octahedron with Mg–O

distance 2.0 Å; these oxygens are in close van der Waals contact with one another. Although many small molecules bind magnesium with greater affinity than they bind calcium, most intra- and extra-cellular proteins bind calcium with much greater affinity than they bind magnesium.

Kinetics of Calcium and Magnesium Binding

The dissociation constant is the ratio of the off rate to the on rate: $K_d(\text{M}) = k_{\text{off}}(\text{s}^{-1})/k_{\text{on}}(\text{M}^{-1}\text{s}^{-1})$. The rate limiting dehydration of $\text{Ca}(\text{H}_2\text{O})_7^{2+}$ is fast, $\sim 10^{8.0} \text{ s}^{-1}$; while that of $\text{Mg}(\text{H}_2\text{O})_6^{2+}$ is slow, $\sim 10^{4.6} \text{ s}^{-1}$. This difference reflects the loose pentagonal bipyramidal versus the tight octahedral packing of the oxygen ligands. The cation must be (partially) dehydrated before it can bind to the protein. These rates are important for modeling the flux of Ca^{2+} ions through the cytosol and the attendant binding of proteins. The increase in affinity of most proteins for calcium relative to magnesium derives primarily from this difference in k_{on} .

Temporal Buffering

It is intriguing that most EF-hand proteins bind calcium $\sim 10,000$ times more strongly than they bind magnesium. This reflects strong evolutionary selection. The cytosolic concentration of the free Ca^{2+} ion is $\sim 10^{-7.2}$ M and that of the Mg^{2+} ion is $\sim 10^{-2.8}$ M; by contrast, both pCa_{out} and pMg_{out} are ~ 2.8 . This means that a cytosolic protein, such as parvalbumin, that binds calcium with high affinity, for example, $\text{p}K_d(\text{Ca}) \sim 8.0$ will also bind magnesium with medium affinity, for example, $\text{p}K_d(\text{Mg}) \sim 4.1$. In the resting cell, it will be primarily in the magnesi-state. By contrast, lower-affinity sites, such as EF-hands 1 and 2 of troponin C have lower affinities for both divalent cations, $\text{p}K_d(\text{Ca}) \sim 6.5$ and $\text{p}K_d(\text{Mg}) \sim 2.3$,

and are apo in the resting state. This leads to the counter intuitive situation in which a pulse of messenger calcium first binds to the weaker apo-sites in domains 1 and 2 of troponin C, misleadingly referred to as calcium specific. The Ca^{2+} ion, whose concentration during the pulse exceeds 10^{-6} M, can bind to the higher-affinity parvalbumin only after the Mg^{2+} ion diffuses off its strong sites. Parvalbumin, with higher affinity for divalent cations, binds calcium after the weaker troponin C and thereby helps to relax the muscle and prevent tetany. Such temporal buffering surely plays a significant role in the propagation and transduction of calcium waves observed in many types of cell types. The information encoded in the frequencies, durations, and amplitudes of calcium waves and spikes might be decoded by a corresponding matching of k_{off} and k_{on} rates for calcium and for magnesium in calcium-modulated proteins.

EF-Hand Containing Proteins

The structure of the EF-hand provides insight into its evolution and function.

Structure of the EF-Hand

The canonical EF-hand (**Figure 1**) consists of α -helix E (forefinger, residues 1–10), a loop around the Ca^{2+} ion (clenched middle finger, 10–21), and α -helix F (thumb, 19–29). Residue 1 is often Glu; the insides of the helices (palmer surfaces) usually have hydrophobic residues that contact the insides of

the other EF-hand of the pair (**Figure 2**). The side chains of the five residues, approximating the vertices of an octahedron (X, residue 10; Y, 12; Z, 14; –X, 18; and –Z, 21), provide oxygen atoms to coordinate the Ca^{2+} ion; residue 16 at –Y bonds to calcium with its carbonyl oxygen. The positions of these ligands within the loop are often referred to as 1, 3, 5, 7, 9, and 12. The Ca^{2+} ion is actually seven coordinate in a pentagonal bipyramid with major axis: X, –X. There are five oxygen atoms in the Y,Z plane since the –Z ligand (residue 21, usually Glu) coordinates Ca^{2+} with both oxygen atoms of its carboxylate group. A Gly at 15 permits a tight bend; residue 17 has a hydrophobic side chain that attaches the loop to the hydrophobic core of the pair of EF-hands.

Several variations to this canonical calcium coordination scheme have been inferred from amino acid sequences and confirmed in crystal structures of other EF-hand proteins. Nearly one-third of all known EF-hands do not bind calcium; those with no indels (insertions or deletions) have a nonoxygen containing side chain substituted at position 10, 12, 14, 18, or 21. Other EF-hands have indels, most notable is EF-hand 1 of the S-100 group subfamilies (**Table 1**), in which several carbonyl oxygens, instead of oxygens of side chains, coordinate the Ca^{2+} ion.

Evolution of the EF-Hand Family

All members of a protein homolog family are inferred to have evolved from a common precursor domain in a single ancestral

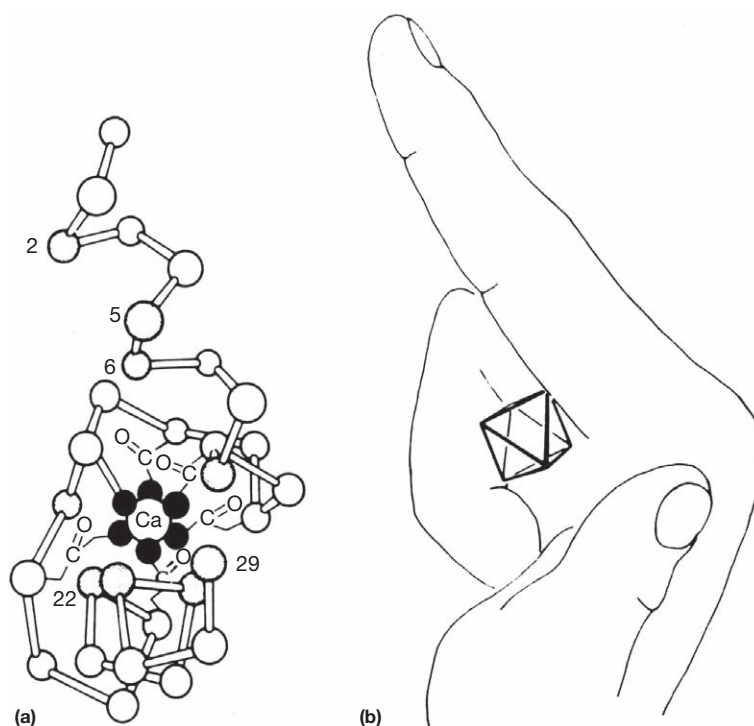


Figure 1 The canonical EF-hand. (a) The spheres represent α -carbons. Residues 2, 5, 6, and 9 of helix E and residues 22, 25, 26, and (often 29) of helix F are usually hydrophobic and face the hydrophobic surface of the other EF-hand of the pair. Five residues – 10 (X), 12 (Y), 14(Z), 18 (–X), and 21 (–Z) – coordinate the Ca^{2+} ion and an oxygen atom from a side chain. Residue 21 is usually Glu and contributes both oxygens of its carboxylate to the pentagonal bipyramid coordination of calcium. Residue 16 (–Y) coordinates with its carbonyl oxygen. (b) Helix E, the loop around the Ca^{2+} ion, and helix F are represented by the forefinger, clenched middle finger, and thumb.

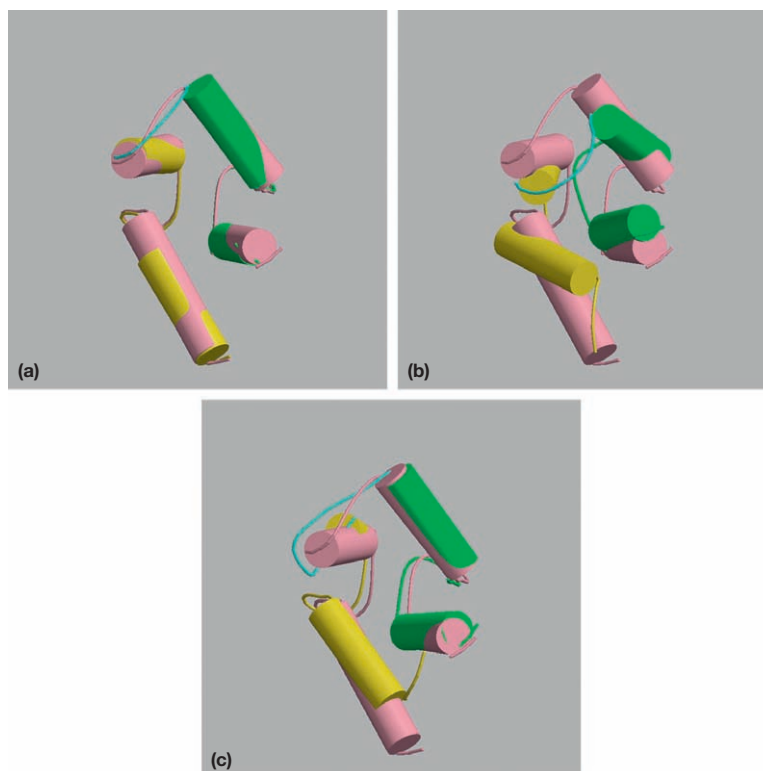


Figure 2 Conformations of pairs of EF-hands. Calmodulin consists of two pairs (1 and 2 and 3 and 4) of EF-hands. A flexible tether, eight residues long, connects helix F2 to helix E3. In the crystal structure, F2, the tether, and E3 form a single continuous helix. Calmodulin is widely distributed and interacts with at least 30 different target proteins. Furthermore, the four EF-hands of calmodulin are all canonical; hence, it serves as a reference point for many evaluations and comparisons. (a) The backbone of the pair calmodulin 3 and 4 in its dicalci-form is shown as a cylinder where it is α -helical and as a strand elsewhere. It is viewed down its approximate twofold axis and is shown in light salmon color; the calcium-binding loops are far from the viewer. Dicalci-parvalbumin CD and EF (CD, yellow; linker, light blue; EF green) is superimposed on dicalci-calmodulin 3 and 4; their conformations are very similar. (b) Apo-calmodulin 3 and 4 (yellow, blue, green) is superimposed on dicalci-calmodulin 3 and 4 (salmon); the binding of calcium opens the cleft between the two helices of each EF-hand and between the two EF-hands of the pair. This is a necessary step permitting calmodulin to interact with its numerous targets. (c) Dicalci-calmodulin 3 and 4 (yellow, blue, green) as complexed with the target peptide from myosin light chain kinase (not shown) is superimposed on dicalci-calmodulin 3 and 4 (salmon). Both the 3 and 4 and the 1 and 2 pairs of calmodulin undergo little additional change in conformation to bind their targets. The binding of calcium facilitates binding of calcium and vice versa. Courtesy of Hiroshi Kawasaki, Yokohama City University, Yokohama, Japan.

organism. The most parsimonious interpretation is that all of the EF-hand domains listed in [Table 1](#) are homologs; however, the statistics of sequence alignment are weak since the domains are only 30 residues long. We employ additional criteria, such as the fact that nearly all EF-hand proteins occur in pairs, so-called EF-lobe ([Figure 2](#)), and that in most EF-hand proteins at least one of the two EF-hands binds calcium. Many (portions of) domains have been suggested to resemble the EF-hand. They are not included in [Table 1](#) unless they have passed a hidden Markov model test based on unambiguous EF-hands of known structure. Most EF-hand-containing proteins have been found in eukaryotes; however, there are several examples in eubacteria and viruses. EF-hand proteins, for example, calmodulin, have been found in all eukaryotes subject to thorough investigation. This distribution might reflect an origin in the bacterium that gave rise to eukaryotic cells; other bacteria may have lost their EF-hands. Or, the precursor EF-hand may have arisen in an early eukaryote and its encoding gene transferred to a few bacteria by some sort of transduction or transformation.

Distribution of EF-Hand Domains

EF-hand proteins contain two to 12 tandem copies of the EF-hand domain. EF-hands occur in pairs, EF-lobe, and are related by an approximate twofold axis of rotation ([Figure 2](#)). Although about one-third of all EF-hands are known or inferred not to bind calcium, usually at least one EF-hand domain in any protein does bind calcium with $pK_a(\text{Ca}) \sim 7.0$. The proteins, such as the ubiquitous and archetypical calmodulin, are in the apo- and/or magnesi-form prior to simulation of the cell; following a rise in $[\text{Ca}^{2+}]_{\text{cytosol}}$, the competent EF-hand(s) binds calcium with attendant change in relative orientation of its helices E and F as well as a change in the relationship of the two EF-hands of the EF-lobe. If the protein is chimeric with a non-EF-hand catalytic domain(s), the change in conformation of the EF-hand region activates the enzyme. If the EF-hand protein itself is not catalytic, the change in conformation causes the EF-hand protein to activate a target enzyme or structural protein.

The characteristics of 73 distinct EF-hand subfamilies are summarized in [Table 1](#). The functions of only 33 of the 73 are

Table 1 Groups of subfamilies of EF-hand proteins

	<i>A</i>	<i>P</i>	<i>F</i>	<i>S</i>	<i>Fn/St</i>	<i>Chimer</i>
<i>CTER, four EF-hands</i>						
Calmodulin	A	P	F	S	+ X	—
Troponin C	A				+ X	—
Essential light chain, myosin	A		F		+ X	—
Regulatory light chain, myosin	A		F	S	+ X	—
Troponin, nonvertebrate (TPNV)	A				+ —	—
Caltractin (centrin)	A	P	F	S	— N	—
Squidulin	A				— —	—
Cal1 gene product	A				— —	—
Calmodulin-like leaf prot (CLAT)		P			— —	C
pCAST gene product		P			— —	C
<i>CPR, four EF-hands</i>						
Calcineurin B	A	F			+ X	—
p22	A				— —	—
Recoverin (visinin)	A				+ X	—
Calcineurin B like sensor (SOS3)		P			+ X	—
Calsenilin	A				— —	C
DREAM(down. reg. el. antag. mod.)	A				+ —	C
Ca/calmodulin protein kinase		P			+ —	C
<i>Penta-EF hand subfamily</i>						
Grancalcin	A				+ X	—
ALG-2 (apoptosis linked gene 2)	A	P	F	S	+ X	—
Peflin	A	P	F	S	+ —	C
Calpain	A				+ X	C
Sorcin	A				+ X	C
<i>S100</i>						
S100	A				+ X	—
Calbindin D _{9k}	A				+ X	—
P26olf (dicalcin)	A				— N	—
Profilaggrin, trichohyalin	A				+ —	C
<i>Proteins with eight and twelve EF-hands</i>						
<i>L. pictus</i> SPEC-resembling prot.	A				— —	—
EF12	A				— —	C
<i>Proteins with six EF-hands</i>						
Calbindin D _{28k}	A				+ X	C
Calretinin	A				+ —	C
CREC subfamily (reticulocalbin)	A				— —	C
EPS15 homology domain	A				— —	C
<i>P. falciparum</i> surface protein	A				— —	C
DNA supercoil factor (<i>Bombyx</i>)	A				— —	C
<i>Proteins with four EF-hands</i>						
Aequorin	A				+ X	—
Calcyphosine	A				+ X	—
<i>S. purpuratus</i> ectodermal CaBP	A				— —	—
SARCoPlasm CaBP	A				— X	—
Calcium vector protein	A				— —	—
Calcium integrin bind. prot. 1	A				— X	—
Hra32		P			— —	—
CBP1-8, <i>Dictyostelium</i>				S	— —	—
EFH5				S	— —	—
Calflagin (1 F8 TB17 calcimedin)				S	— —	—
Phospholipase C	A		F		+ X	C
Protein phosphatase, <i>Drosophila</i>	A				+ —	C
PM129 clone, <i>A. thaliana</i>		P			— —	C
LAV1		F			— —	C
<i>Tetrahymena pyriformis</i> (TCBP)			S		— —	C
Spasmin			S		+ —	C
<i>P. falciparum</i> protein kinase			S		+ —	C
Surface protein (<i>Plasmodium</i>)			S		— —	C

(Continued)

Table 1 (Continued)

	<i>A</i>	<i>P</i>	<i>F</i>	<i>S</i>	<i>Fn/St</i>	<i>Chimer</i>
Parvalbumin (EF's 2, 3, 4)	A				+ X	—
<i>Proteins with two EF-hands</i>						
CBL, proto oncogene product	A				+ X	—
Calmodulin related gene product	A				— —	—
Groovin	A				— —	—
Ca and integrin binding protein	A				— —	—
Calsensin	A				— —	—
Phl p 7 (polcalcin)		P			— X	—
α -Actinin	A	F			+ —	C
Fimbrin	A	F			+ —	C
α -Spectrin & α -fodrin	A				+ —	C
Diacylglycerol kinase	A				+ —	C
Glycerol-3-P dehydrogenase	A				+ —	C
Ryanodine receptor protein	A				+ —	C
BM-40 (osteonectin, SPARC)	A				— X	C
Polycystin-2 (<i>PKD2</i> gene prod.)	A				— —	C
Ras guanyl nucl. releas. prot.	A				— —	C
Nucleobindin	A				— —	C
Allograft inflammatory factor-1	A				— —	C
<i>Four and six EF-hand proteins in bacteria and viruses</i>						
Calerythrin, <i>Saccharopolyspora</i>					— N	—
MSV097, <i>Entomopoxvirinae</i> , 4 EF					— —	—
Calsymin, <i>Rhizobium etli</i> (6 EF)					— —	C

known. Nine of these are enzymes and have been shown or inferred to be activated by the binding of calcium. Many, but certainly not all, of the remaining 64 such as calmodulin, function in information transduction pathways. However, others, such as calbindin D_{9k}, probably facilitate the diffusion of calcium through the cytosol; parvalbumin appears to function as a temporal buffer. Thirty-six, including the nine enzymes, of the 73 subfamilies are chimeric. In addition to their EF-hands, they contain other domains of different evolutionary origin and different conformation. Crystal structures are available for 23 of the subfamilies; nuclear magnetic resonance (NMR) structures are available for three additional subfamilies. Members of the penta-EF-hand subfamily (calpain, sorcin, peflin, grancalcin, and ALG-2) consist of two canonical EF-lobes (domains 1 and 2 and 3 and 4). These proteins form homo- or hetero-dimers by pairing their fifth EF-hands to form a canonical EF-lobe.

It is not unusual for a basic protein domain to find many uses, often spliced together with other domains; however, the EF-hand is one of the most widely distributed domains in eukaryotes, perhaps reflecting the range and subtlety of calcium signaling. The downstream regulation element antagonist modulator (DREAM) upon binding calcium dissociates from a DNA-binding regulatory element that otherwise functions as a gene silencer. This might provide a precedent for long-term potentiation.

Most classifications of proteins recognize the domain, for example, the EF-hand, as the fundamental unit of both evolution and structure. All homologous domains are assumed to have evolved from a single precursor domain in an ancestral organism. All domains in a single homolog family have similar

tertiary structures and, within the limits of statistical deviation, recognizably similar amino acid sequences. Exceptions may include short linker regions between domains and intrinsically disordered regions, both of which often contain tandem repeats of oligo-peptides or of single residues. Any given domain belongs to one domain family; by this definition, a domain cannot belong to two families. Many proteins consist of multiple domains; some are chimeric with domains from different families spliced together, reflecting gene fusion. Each domain in a multidomain protein has its own evolutionary history.

If two homologous domains within two proteins or within a single, multidomain protein were generated by gene duplication (followed by gene fusion in the multidomain protein), they are paralogs. If two proteins in two organisms were generated by speciation, they are orthologs. Two homologous proteins within a single species are, by definition, paralogs. Two proteins, one each from two different species, may be either paralogs or orthologs; the burden of proof lies with the advocate of orthology. A dendrogram relating orthologs should reflect the phylogeny of the host species, and, within statistical error, be identical to other dendrograms based on families of other orthologs from those same species.

A similar analysis can be applied to groups of proteins composed of several homologous domains, for example, the four EF-hands of calmodulin, troponin C, essential and regulatory light chains (CTER) from several species. All domains 1 cluster together, near the cluster of domains 3; all domains 2 cluster together, near the cluster of domains 4. This implies that all CTERs (and other proteins included in this group) evolved from a single, four EF-hand precursor by gene duplication and fusion with subsequent speciation. All of the subfamilies within

CTER are congruent. Further, this four-domain precursor evolved by gene duplication and fusion from a two-domain ODD, EVEN precursor. All EF-hands, with exceptions to be noted, occur as ODD, EVEN EF-lobes. Probably all EF-lobes evolved from a single EF-lobe, itself the result of the duplication of a single ur-EF-hand in an ancestor to all eukaryotes.

The characteristics of groups of subfamilies of EF-hands are summarized in [Table 1](#). Four groups of subfamilies – CTER (with 10 subfamilies), CPR (7), penta (5), and S100 (4) – show significant congruence, the basis for their classifications. The remaining five groups – eight and 12 EF-hand domains (2 subfamilies), six (6), four (19), two (17), and bacteria and virus (3) – reflect the number of domains or a noneukaryote host. The evolutionary relationships of the subfamilies within these groups are problematic. Parvalbumin is listed with the four EF-hand group because it probably evolved from a four-domain precursor, with the first (ODD) domain having been deleted. The host organism(s) are indicated as animal, plant, fungi, and/or protist (single cell). Also indicated are whether a function is known (+/-), whether a crystal or NMR structure is available (X,N/-), and whether the protein is chimeric (C/-) or consists solely of EF-hands.

For several of these 73 subfamilies, with few examples and limited functional information, the decision whether ‘to lump or to split’ is somewhat arbitrary. Thirty-six of the 72 are chimeric. In a few cases, this assignment depends on whether the non-EF-hand regions are considered to be canonical domains or tandem repeat linkers or intrinsically disordered regions.

Calmodulin

Calmodulin is probably present in all cells of all eukaryotes. It consists of four EF-hands and its amino acid sequence is highly conserved. The second α -helix of domain 2 (F2), an eight-residue linker, and the first α -helix of domain 3 (E3) forms a single continuous α -helix ~28 residues long, thereby giving calmodulin a dumbbell shape as seen both in crystals and in solution. Upon binding calcium, the relative orientations in both EF-lobes change, thereby exposing hydrophobic regions and permitting the interaction of calmodulin with over 40 different target enzymes and structural proteins. The structures of calmodulin complexed with α -helical regions of several targets, such as myosin light chain kinase (MLCK),

reveal what appears to be a general pattern for the interactions of four-domain EF-hand proteins, such as calmodulin, troponin C, and the essential and regulatory light chains of myosin, with their respective targets. MLCK is self-inhibited by its own peptide, 796–813; limited proteolysis removes this peptide. The 1–795 MLCK is then constitutively active. *In vivo*, calmodulin binds calcium, undergoes a change in conformation, binds α -helix, 796–813, of MLCK, and thereby removes the self-inhibition of MLCK. Calmodulin assumes near spherical shapes when complexed with its various targets, as opposed to the elongated dumbbell shape of uncomplexed calmodulin. The eight-residue linker is flexible, permitting calmodulin to assume a broad range of conformations. Several of these conformations match the many targets of calmodulin.

See also: [Bioenergetics: Calcium-Binding Proteins: Cytosolic \(Annexins, Gelsolins, and C₂-Domain Proteins\); ER/SR Calcium Pump: Structure.](#)

Further Reading

- Berridge MJ (2006) Calcium microdomains: Organization and function. *Cell Calcium* 40: 405–412.
- Gerke V, Creutz CE, and Moss SE (2005) Annexins: Linking Ca²⁺ signaling to membrane dynamics. *Nature Reviews: Molecular and Cellular Biology* 6: 449–461.
- Hofer AM (2005) Another dimension to calcium signaling: A look at extracellular calcium. *Journal of Cell Science* 118: 855–862.
- Kitaura Y, Satoh H, Takahashi H, Shibata H, and Maki M (2002) Both ALG-2 and peflin, penta-EF-hand (PEF) proteins, are stabilized by dimerization through their fifth EF-hand regions. *Archives of Biochemistry and Biophysics* 399: 12–18.
- Nakayama S, Kawasaki H, and Kretsinger RH (2000) Evolution of EF-hand proteins. *Topics in Biological Inorganic Chemistry* 3: 29–58.
- Permyakov EA and Kretsinger RH (2010) *Calcium Binding Proteins*. New York, NY: Wiley.
- Schaub MC and Heizmann CW (2008) Calcium, troponin, calmodulin, S100 proteins: From myocardial basics to new therapeutic strategies. *Biochemical and Biophysical Research Communications* 369: 247–264.
- Südhof TC (2002) Synaptotagmins: Why so many? *Journal of Biological Chemistry* 277: 7629–7632.
- Toyoshima C and Nomura H (2002) Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* 418: 605–611.

Relevant Website

<http://calcium.sci.yokohama-cu.ac.jp> – Symmetry alignment of EF-hand.

Calcium Oscillations

M Brini, University of Padua, Padua, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

G protein Protein associated to specific hormone receptor of the plasma membrane and responsible for the conversion of extracellular signals in intracellular messages.

Inositol (1,4,5)-trisphosphate (InsP₃) Diffusible cytosolic messenger which induces the release of Ca²⁺ from intracellular stores (the endo/sarcoplasmic reticulum) by opening InsP₃-sensitive Ca²⁺ channels.

Membrane potential Electrical gradient across the plasma membrane (and other membranes as well).

Neurotransmitters Chemical messengers by which neurons communicate with each other.

Nicotinamide adenine dinucleotide (NAD(P)H) Acceptor of the reducing equivalent produced in the oxidation of most metabolites. It is a small organic molecule, which participates as a coenzyme in enzymatic reactions.

General Considerations

It has been known since the 1950s that the periodic opening of plasma membrane Ca²⁺ channels, such as induced, for example, by the rhythmic changes of the plasma membrane potential of the heart or by burst of action potential in neurons, can produce fluctuations in cytosolic [Ca²⁺]. The idea that [Ca²⁺]_c oscillations may also occur in nonexcitable cells has emerged much later, thanks to seminal observations by Cobbold and co-workers in mid-1980s on the fertilization of oocytes and on hormone-stimulated hepatocytes. Surprisingly, these oscillations were not dependent on the periodic opening and closing of plasma membrane Ca²⁺ channels, but rather on cycles of Ca²⁺ release and uptake from the intracellular compartment sensitive to the second messengers inositol (1,4,5)-trisphosphate (InsP₃), which is the endoplasmic reticulum (ER). Later, Ca²⁺ oscillations were observed in other animals as well as plant cells, even if many of these cells do not have an obvious oscillatory biological function.

A question that immediately arises is the reason for the transmission of the Ca²⁺ signal by oscillations rather than by a stationary change in Ca²⁺ concentration. It is generally assumed that the oscillatory behavior has physiological advantages. This is so because the elevation of Ca²⁺ concentration over a long period is known to be lethal to the cell: the oscillatory behavior would prevent this negative effect, while still permitting optimal activation of Ca²⁺ sensing enzymes. The oscillatory regime would also prevent long-lasting receptor desensitization, thus increasing the sensitivity of sensing enzymes to Ca²⁺, that is, [Ca²⁺] would be permitted to periodically exceed the threshold for enzyme activation, even its level would remain below the threshold. Another advantage is the wide range of temporal and spatial aspects that oscillations can assume, allowing different signals to be transmitted within cells and from cell to cell.

Usually, Ca²⁺ oscillations are characterized by rather constant amplitude and by a variable frequency, which ranges from 5 to 60 seconds depending on the cell type, and on the nature and the strength of the stimulus (Figure 1). These characteristics have led to the concept of frequency-modulated Ca²⁺ signaling, which reflects the ability of cells to interpret changes in the frequency of the oscillations.

Mechanism of [Ca²⁺]_c Oscillations

Although the mechanism of Ca²⁺ oscillation has received much experimental and theoretical attention, it is still largely debated. A first distinction considers the origin of the Ca²⁺ spikes, for which two major mechanisms exist: one is linked to the periodic opening of plasma membrane channels and the other to cycles of Ca²⁺ release and uptake from intracellular stores (Figure 2). However, this distinction is incomplete since very often the contributions of intracellular Ca²⁺ release and of extracellular Ca²⁺ influx cannot be clearly separated. An obvious example is that of cardiac myocytes, where the periodic contraction of the cells is directly promoted by the release of Ca²⁺ from the sarcoplasmic reticulum (SR), which is triggered by localized influx of Ca²⁺ through plasma membrane voltage-operated Ca²⁺ channels (VOCs), a process known as Ca²⁺-induced Ca²⁺ release (CIRC).

In excitable cells such as neurons, heart, and neuroendocrine cells, the transient [Ca²⁺] elevation is due to Ca²⁺ entry through VOCs or receptor-operated Ca²⁺ channels (ROCs) activated in response to neurotransmitters. In these cells, the oscillatory behavior is secondary to the oscillatory nature of the trigger itself. For example, cortical and cerebellar neurons in primary cultures exhibit spontaneous oscillations, which occur in synchrony with adjacent neurons and are dependent on synaptic activity, that is, changes of membrane potential.

In nonexcitable cells the predominant mechanism of [Ca²⁺]_c elevation is through the activation of plasma membrane receptors coupled to G proteins and the phosphoinositide pathway. Activation of membrane phospholipase C (PLC) generates the second-messenger InsP₃ from phosphatidylinositol 4,5-bisphosphate, mobilizing Ca²⁺ from stores, associated with the ER or SR. The pattern of the Ca²⁺ signals is often complex but InsP₃-linked agonists frequently give rise to repetitive [Ca²⁺]_c transients.

Other molecules (i.e., cyclic adenosine diphosphate (ADP)-ribose (cADPR) and nicotinic acid adenine dinucleotide phosphate (NAADP)) have recently also been shown to mobilize Ca²⁺ from internal stores generating Ca²⁺ oscillations. Studies using Ca²⁺ indicator dyes targeted to the lumen of intracellular stores of intact cells have shown that agonist-induced cytosolic Ca²⁺ oscillations are accompanied by inverse oscillations of stored Ca²⁺.

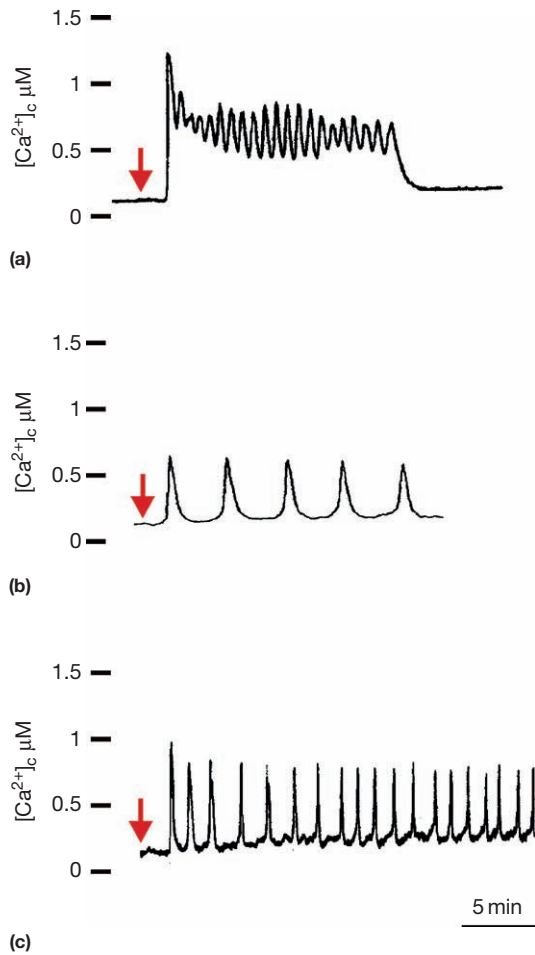


Figure 1 Examples of different oscillatory pathways: (a) parotid gland; (b) hepatocyte; and (c) endothelial cell. The pattern of oscillations may vary with respect to amplitude and period. The red arrow indicates agonist stimulation.

Recent works suggest at least two possible mechanisms for the generation of oscillatory Ca^{2+} signals: either an oscillatory production of InsP_3 or an oscillatory inactivation of InsP_3 receptors. Both mechanisms appear to operate in different cell types, the common denominator being the positive and negative feedback by Ca^{2+} on the release system; for example, in hormone-stimulated hepatocytes, oscillations are driven by the cycling of the InsP_3 channels between a fully open and a largely closed state, rather than by oscillations in InsP_3 . Similarly, in pancreatic acinar cells, the pulsatile calcium release does not depend on fluctuations in InsP_3 concentration, but in kidney epithelial cells spatiotemporal changes in the concentration of InsP_3 appear to be synchronous with Ca^{2+} oscillations, and intracellular InsP_3 waves accompany Ca^{2+} waves.

The fluctuations in InsP_3 may be controlled by Ca^{2+} itself through the regulation of PLC activity or through regulatory proteins which act directly on G proteins, thus affecting the downstream InsP_3 production. Alternatively, the activity of the InsP_3 -sensitive Ca^{2+} channels may be controlled by Ca^{2+} itself at both the cytosolic and the luminal side, and, possibly, by kinase-mediated phosphorylation/dephosphorylation cycles. Thus, low concentrations of cytosolic Ca^{2+} increase the open

probability of the channels, whereas higher concentrations favor their closure, generating the oscillatory Ca^{2+} -signaling pattern.

Another intriguing aspect that has recently emerged is that it is clear that influx of Ca^{2+} across the plasma membrane also plays a critical role in maintaining Ca^{2+} oscillations and in delivering their signal to the correct cellular locus. As ER Ca^{2+} levels drop (as a consequence of physiological receptor activation), Ca^{2+} enters through highly specific store-operated channels (SOCs) in the plasma membrane that are controlled by stromal-interacting molecule (STIM) proteins, sensors of ER luminal Ca^{2+} levels. STIM1 protein is shown to translocate cyclically in and out of ER-plasma membrane junctions during each Ca^{2+} oscillatory spike, thus activating Orai1 channels and Ca^{2+} entry. The elimination of STIM1 or Orai1 channels suppresses the oscillations.

Role and Functional Significance of $[\text{Ca}^{2+}]_c$ Oscillations

Ca^{2+} oscillations have been implicated in the control of different cell processes, including oocyte activation and fertilization, growth-cone migration, growth-cone tuning, axonal growth of cortical neurons, neuronal-cell migration, development of neurotransmitter phenotypes, formation of nodules in plant root hairs, development of muscle, release of cytokines from renal epithelial cells, and disassembly of adhesive structures during cell migration (Figure 3).

Numerous experimental protocols modulate the amplitude and the frequency of Ca^{2+} spikes in living cells. In general, cells choose the shape of Ca^{2+} signals according to the type of process that must be activated: single Ca^{2+} transients are used to activate processes such as muscle contraction, and repetitive signals are used when information must be relayed over long time periods, as in fertilization and in the triggering of the developmental program. Furthermore, oscillations can be directed to discrete subcellular domains to generate large local $[\text{Ca}^{2+}]_c$ increases required to couple $[\text{Ca}^{2+}]_c$ to secretion and to mitochondrial metabolism. They can also propagate through entire cells, or even through coupled cells to coordinate the activities in intact tissues and organs.

The most obvious mechanism for the regulating of specific functions would seem to be modulation of the amplitude of Ca^{2+} signals. However, the efficiency of the activating stimulus is frequently related to the frequency rather than to the amplitude of the $[\text{Ca}^{2+}]_c$ oscillations.

A particularly illustrative example of decoding of the Ca^{2+} signal through the frequency of the oscillations is the Ca^{2+} -regulated gene expression.

Dolmetsch and colleagues have recently reported that Ca^{2+} oscillations in T-lymphocytes reduce the effective Ca^{2+} -threshold for activating some transcription factors, more effectively than the sustained Ca^{2+} increase. The efficacy is encoded by the oscillation frequency: rapid oscillations stimulate three transcription factors (nuclear factor of activated T-cells (NF-AT), Oct/OAP, and nuclear factor kappaB (NF- κ B)), while lower frequency oscillations only activate NF- κ B. Thus, Ca^{2+} -modulated gene transcription is a frequency-sensitive process. Another elegant example of the correlation between the frequency of the $[\text{Ca}^{2+}]_c$ spikes and cellular function is the

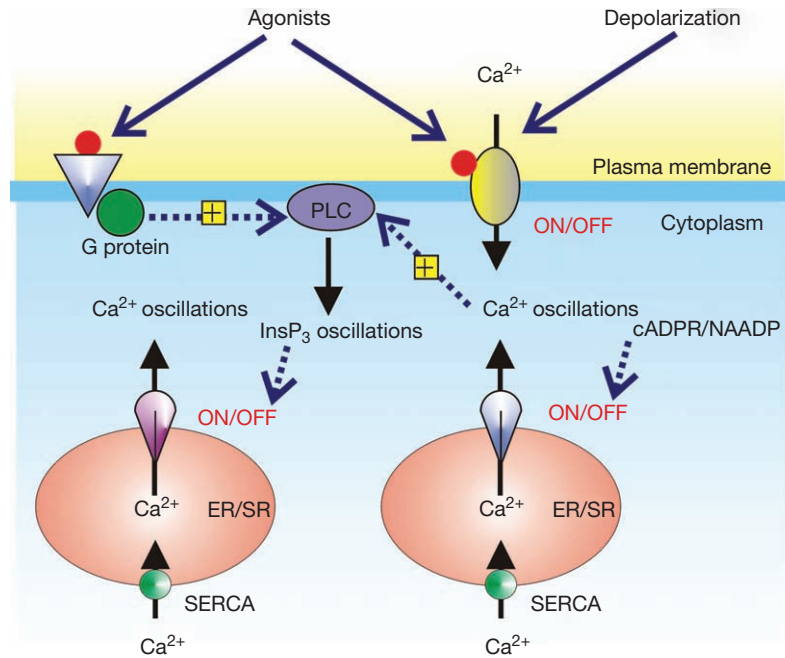


Figure 2 Signaling pathways that participate in the generation of Ca^{2+} oscillations. Three models are proposed. In excitable cells, Ca^{2+} oscillations are evoked by rhythmic oscillatory changes in membrane potential. In nonexcitable cells they are induced either by the oscillatory production of InsP_3 or following the oscillatory inactivation of InsP_3 receptors (ON/OFF). ER/SR, endoplasmic/sarcoplasmic reticulum; PLC, phospholipase C; SERCA, sarco/endoplasmic reticulum calcium ATPase; InsP_3 , inositol 1,4,5-trisphosphate; cADPR, cyclic ADP-ribose; NAADP, nicotinic acid adenine dinucleotide phosphate.

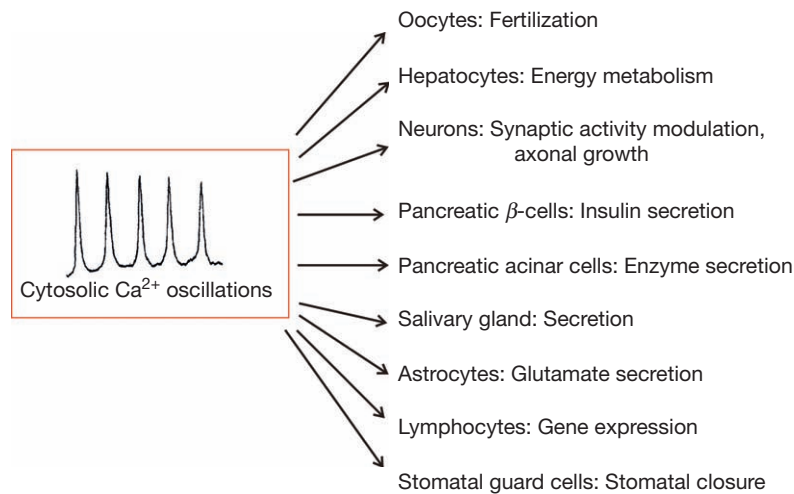


Figure 3 Scheme of the main cellular events modulated by the oscillatory behavior of Ca^{2+} signal.

translation of repetitive cytosolic Ca^{2+} spiking through mitochondrial Ca^{2+} oscillations in primary cultures of rat hepatocytes in the final and sustained elevation of nicotinamide adenine dinucleotide (NAD(P)H) production. This coupling allows the coordination of mitochondrial adenosine triphosphate (ATP) production, linked to the sustained activation of mitochondrial Ca^{2+} -sensitive dehydrogenases, with the level of energy demand, at the same time avoiding the deleterious effects of the prolonged $[\text{Ca}^{2+}]$ increases.

In polarized cells, such as those of pancreatic acini, the receptor-evoked Ca^{2+} signal initiates at the apical pole,

generating agonist-specific Ca^{2+} oscillation patterns. Acetylcholine stimulation evokes repetitive local spikes, which remain confined to the secretory apical pole. Cholecystokinin instead induces local Ca^{2+} spikes that end in longer-lasting Ca^{2+} transients that spread to the whole cell.

Last but not least, in considering the functional significance of Ca^{2+} oscillations it is necessary to empathize their role in the central nervous system. In neurons, the sequential openings of plasma membrane Ca^{2+} channels lead to rhythmic Ca^{2+} spikes that, in turn, are linked to synaptic communication within and among neuronal circuitries, that is, by inducing neurotransmitter

release. A peculiar role has been recently attributed to the spontaneous and synchronous Ca^{2+} oscillations that occur independently from synaptic activity in developing neurons, such as those in rat visual cortex. It was suggested that these oscillations could propagate through different cells and determine a common developmental pattern of cells from the same domain.

A sophisticated example of the role of Ca^{2+} oscillation in the coordination of the activity of neuronal circuits has been described in glial astrocytes, a cell type of the central nervous system, which plays a number of important functions. Very recently, it has been proposed that they work as detectors of synaptic activity: by changing the frequency of Ca^{2+} oscillations evoked by the synaptic release of glutamate these astrocytes acquire the ability to discriminate between different levels and patterns of synaptic activity.

How Are $[\text{Ca}^{2+}]_c$ Oscillations Decoded?

Although the frequency of Ca^{2+} oscillations seems critical for the induction of selective cellular functions, the mechanisms by which the cell decodes the oscillatory signal remain unclear.

In 1998 De Koninck and Schulman showed that the oscillation frequency modulates the activity of Ca^{2+} -calmodulin kinase II (CaMKII) *in vitro*. This enzyme is tuned to respond optimally to Ca^{2+} oscillations with high frequency. The multimeric structure of the enzyme, coupled to its autophosphorylation properties, has led to the suggestion that it can act as a frequency decoder of $[\text{Ca}^{2+}]_c$ oscillations. The enzyme is now considered of special importance to the nervous system, where it regulates the process of memory formation and storage.

The advent of the green fluorescent protein (GFP) technology has shown that protein kinase C (PKC), one of the most important enzymes in cell signaling, undergoes an oscillatory plasma membrane association in phase with receptor-mediated oscillations in $[\text{Ca}^{2+}]_c$. Each $[\text{Ca}^{2+}]_c$ oscillation induces the

transient membrane association of PKC leading to a transient burst of PKC substrate phosphorylation. It was suggested that the detachment of PKC from the membrane, which occurs as Ca^{2+} returns to the basal level between oscillations, results in a pause of PKC activity, allowing substrate dephosphorylation. In this respect, the Ca^{2+} -mediated regulation of PKC differs from that of CaMKII, since the latter does not fully deactivate between spikes as the frequency of $[\text{Ca}^{2+}]_c$ oscillations increases.

The matter of decoding Ca^{2+} oscillations however is even further complex since emerging evidence highlights the possibility that Ca^{2+} oscillations that are identical in amplitude and frequency should differentially activate downstream targets depending on spatial location of the underlying Ca^{2+} channels: only oscillations with accompanying Ca^{2+} entry through SOCs triggered gene expression, and Ca^{2+} entry is needed to sustain Ca^{2+} oscillations.

See also: [Bioenergetics: Calcium in the Regulation of Gene Expression](#); [Calcium Signaling: Calmodulin-Dependent Phosphatase](#); [ORAI Store-Operated Calcium Channel](#).

Further Reading

- Berridge MJ (1990) Calcium oscillations. *Journal of Biological Chemistry* 265: 9583–9586.
- Berridge MJ (2007) Inositol trisphosphate and calcium oscillations. *Biochemical Society Symposia* 74: 1–7.
- Bird GS, Hwang SY, Smyth JT, et al. (2009) STIM1 is a calcium sensor specialized for digital signalling. *Current Biology* 19: 1724–1729.
- Di Capite J, Ng SW, and Parekh AB (2009) Decoding of cytoplasmic Ca^{2+} oscillations through the spatial signature drives gene expression. *Current Biology* 19: 853–858.
- Petersen OH, Petersen CC, and Kasai H (1994) Calcium and hormone action. *Annual Review of Physiology* 56: 297–319.
- Thomas AP, Bird GS, Hajnóczky G, Robb-Gaspers LD, and Putney JW, Jr. (1996) Spatial and temporal aspects of cellular calcium signalling. *FASEB Journal* 10: 1505–1517.

Calcium-Sensing Receptor (CaSR)

J Tfelt-Hansen, Copenhagen University Hospital, Copenhagen, Denmark

E M Brown, Harvard Medical School, Boston, MA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Autosomal dominant hypoparathyroidism

(ADH) A generally benign form of hypocalcemia caused by gain-of-function mutations in the CaSR.

Calcilytic An inhibitor of the CaSR.

Calcimimetic An allosteric activator of the CaSR.

Calcium-sensing receptor (CaSR) A G-protein-coupled receptor maintaining the near constancy of extracellular calcium.

Familial hypocalciuric hypercalcemia (FHH) A benign form of hypercalcemia caused by loss-of-function mutations in the CaSR.

Neonatal severe primary hyperparathyroidism

(NSHPT) An autosomal recessive disease usually caused by loss-of-function mutations of both alleles of the CaSR.

Background

The body maintains the level of extracellular free (i.e., unbound) calcium ($\text{Ca}^{2+}_{\text{o}}$) within a narrow range (1.1–1.3 mM) because both high and low levels of $\text{Ca}^{2+}_{\text{o}}$ are dangerous and can be life threatening. Very small changes in $\text{Ca}^{2+}_{\text{o}}$, on the order of a few percent, lead to immediate physiological responses that restore the level of $\text{Ca}^{2+}_{\text{o}}$ to normal. These fast responses are essential because rapid changes in $\text{Ca}^{2+}_{\text{o}}$ are more dangerous than those that develop more slowly. The calcium-sensing receptor (CaSR), which was cloned more than a decade ago, plays a central role in this delicate homeostatic system. The CaSR acts like a thermostat, but instead of measuring changes in temperature it measures alterations in the level of $\text{Ca}^{2+}_{\text{o}}$, thereby functioning as a calcistat. This means that the CaSR tells the chief cells of the parathyroid glands the exact level of $\text{Ca}^{2+}_{\text{o}}$. The CaSR is a seven-transmembrane-domain (TMD) receptor that is coupled to G proteins – a so-called G-protein-coupled receptor (GPCR). The importance of the CaSR is illustrated by the diseases caused by naturally occurring mutations in it that lead to either loss of function or gain of function. Heterozygous (one allele) loss-of-function mutations give rise to a generally benign condition called ‘familial hypocalciuric hypercalcemia (FHH)’, also called ‘familial benign hypocalciuric hypercalcemia (FBHH)’, in which most affected individuals have asymptomatic hypercalcemia. The homozygous variant of inactivating CaSR mutations, in contrast, is a much more severe, potentially fatal disease if left untreated; it is called ‘neonatal severe primary hyperparathyroidism (NSHPT)’. The mouse model of FHH (e.g., heterozygous knockout of the CaSR gene) has a phenotype comparable to that of the human condition, suggesting that the number of receptors on the cell surface is the determining factor for the clinical expression of the disease. Gain-of-function mutations give rise to autosomal-dominant hypoparathyroidism (ADH). Most individuals with this condition have asymptomatic hypocalcemia. Although the principal role for the CaSR is in calcium homeostasis, other functions of

the receptor have been elucidated in the wide variety of cell types expressing it. The CaSR can regulate the cell cycle, ion channels, peptide secretion, and diverse other functions. This article provides a broad overview of the CaSR’s biochemical features, a discussion of its functions in normal and pathophysiological states, and a discussion of new drugs that have been developed that target the CaSR.

Biochemical Features of the CaSR

The CaSR is a seven-TMD GPCR whose gene is located on chromosome 3q13. Other receptors in the superfamily of GPCRs that are related structurally to the CaSR include the gamma aminobutyric acid type B (GABA_{B}) and metabotropic glutamate receptors (mGluRs), which recognize GABA and glutamate, respectively, as their ligands, as well as pheromone, taste, and odorant (in fish) receptors.

Structure

The human CaSR has a very large N-terminal extracellular domain (ECD) of 612 amino acids and has a bilobed Venus flytrap structure (VFT) (Figure 1). The ECD contains key binding sites for calcium, although there is at least one additional site(s) in the receptor’s central core of 250 amino acids, including the TMDs and/or associated extracellular loops (ECLs). Binding of Ca^{2+} (the CaSR’s principal physiological ligand) to the ECD is thought to close the VFT, in association with additional conformational changes in the ECD and/or TMDs that activate signal transduction. The binding of $\text{Ca}^{2+}_{\text{o}}$ is very weak compared to the affinities of many other receptors for their ligands. These weak forces between the ECD and $\text{Ca}^{2+}_{\text{o}}$ are crucial for the receptor to act as a calcistat. The TMDs also serve as the target for newly developed drugs that act as allosteric activators (calcimimetics) or inhibitors (calcilytics) of the CaSR (discussed later). The intracellular carboxy (C)-terminus of the receptor consists of 216 amino acids and has important roles, along with the CaSR’s intracellular loops

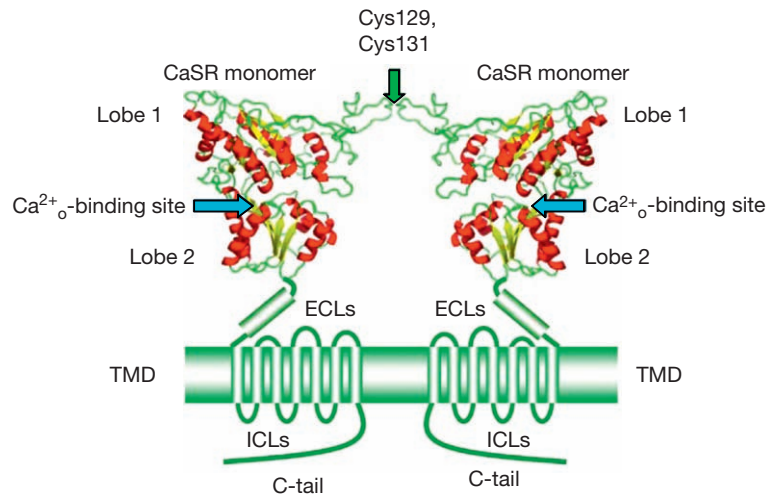


Figure 1 Representation of the predicted structure of the bilobed dimeric, 612 amino acid ECDs of the human extracellular Ca^{2+} -sensing receptor based on the known structure of the ECDs of several mGluRs. Also shown schematically in the ECDs are the positions of the putative Ca^{2+} -binding sites in the crevice between lobes 1 and 2. Cys 129 and Cys 131 indicate the locations of the disulfide bonds between the flexible loop 2 of each ECD. Shown in simplified form in the bottom half of the picture are the receptor's TMDs, extra (ECLs)- and intracellular loops (ICLs), and intracellular C-tail (C-tail). This research was originally published in Huang Y, Zhou Y, Yang W, et al. (2007) Identification and dissection of Ca^{2+} -binding sites in the extracellular domain of Ca^{2+} -sensing receptor. *Journal of Biological Chemistry* 282(26): 19000–19010. © The American Society for Biochemistry and Molecular Biology.

(ICLs), in the transduction of the Ca^{2+} signal to inside the cell (see below).

N-Glycosylation with carbohydrates is crucial for cell surface expression

The CaSR has several N-linked glycosylation sites that are important for its expression on the cell surface. Western blotting of the CaSR shows several bands between 120 and 200 kDa. A minor band at 120 kDa represents the nonglycosylated CaSR, whereas major bands at about 140 and 160 kDa correspond to the immature, high mannose and mature, fully N-glycosylated forms of the CaSR, respectively. If the degree of glycosylation of the receptor is reduced to a significant extent by site-directed mutagenesis of its N-linked glycosylation sites, less of the receptor reaches the cell surface, thereby indirectly reducing the activity of the receptor, although the intrinsic activity of the receptor is not thought to be directly linked to glycosylation.

CaSR functions as a dimer on the cell surface

Additional immunoreactive bands observed on Western blot analysis of the CaSR at >200 kDa represent CaSR dimers or even larger species. The fully glycosylated, ~320-kDa dimer is the major form of CaSR on the cell surface and is thought to be the active form of the receptor. Two monomers of the CaSR are linked covalently by two disulfide bonds involving cysteine residues 129 and 131 as well as by noncovalent, hydrophobic interactions. The disulfide bonds constrain the receptor in its inactive form; disrupting them by site-directed mutagenesis or naturally occurring mutations produces a receptor that responds to lower than normal levels of Ca^{2+} . Moreover, heterodimerization of a normal CaSR monomer with a monomer harboring an FHH mutation can interfere via a dominant negative mechanism with the function of the wild-type CaSR

monomer, documenting that there are functional interactions between the monomers of the dimeric CaSR. The CaSR can also heterodimerize with the GABA_B receptor, which modulates the function of both heterodimeric partners.

Amino acids in ICLs are pivotal for activation of intracellular signaling by the CaSR

Amino acids in ICLs 2 and 3 are key players in determining selective coupling of the CaSR to G proteins. For example, two amino acid residues within ICL 2 and eight residues in ICL 3 have been shown to be important for the G-protein-dependent activation of phospholipase C (PLC) – one of the major intracellular signaling pathways used by the CaSR.

The C-tail of the CaSR is important for its function

The amino acid sequence of the proximal portion of the C-tail of the CaSR is highly conserved among species, suggesting its functional importance. In addition, naturally occurring mutations in the C-tail have been shown to cause the diseases mentioned earlier (e.g., FHH and ADH). Furthermore, certain truncation and deletion mutations of the receptor's C-tail upregulate the expression of the receptor on the cell surface, thereby causing gain of function. Other truncations cause loss of function, demonstrating that there are discrete elements within the C-tail that modulate both the cell surface expression and the function of the receptor. The CaSR has several binding partners that interact with the CaSR's C-tail, thereby affecting the function of one or both of the two partners. These include filamin-A, a large actin-binding protein that can serve as a scaffold protein (e.g., for mitogen-activated protein kinases (MAPK)), dorfin, which likely regulates proteasomal degradation of the receptor, receptor activity modifying proteins (RAMPs), and potassium channels.

Functionally important mutations in the CaSR as experiments in nature

Approximately 200 naturally occurring mutations have so far been described that cause either a loss of function of the CaSR in FHH and NSHPT or a gain of function in ADH. Deletion, nonsense, insertion, missense, truncations, and splice-site mutations have been described throughout the whole CaSR gene. They have not only proven the physiological importance of the CaSR in Ca^{2+} homeostasis, but also provided an important source of information about the structure–function relationships of the receptor, as noted previously. There are several polymorphisms present within the CaSR's C-tail and elsewhere. Available evidence suggests that they may contribute to the variation in the range of serum calcium concentrations among normal individuals and modify the severity of diseases of calcium homeostasis, such as primary and secondary hyperparathyroidism. Inactivating or activating antibodies to the CaSR causes conditions phenotypically similar to FHH and ADH, respectively.

Agonists of the CaSR

As previously mentioned, the CaSR can detect very small changes in Ca^{2+}_o , on the order of a few percent. As noted above, calcium interacts with the ECD of the CaSR. This has been established elegantly through the creation of a chimeric receptor containing the CaSR's ECD and the TMDs and C-tail of an mGluR. The chimeric receptor responds to Ca^{2+}_o but not to glutamate. A key Ca^{2+} -binding site likely resides within the crevice between the two lobes of each monomeric VFT. The CaSR's Hill coefficient is typically 3–4, implying the presence of additional binding sites for Ca^{2+}_o , yet to be definitively identified; the steepness of the physiological response of the parathyroid glands to Ca^{2+} is even more extreme. As a result, suppression of parathyroid hormone (PTH) release *in vitro*, and *in vitro* occurs over a very narrow range of Ca^{2+} , between about 0.75–1.0 mM and 1.5–2 mM. Unlike many receptors, the CaSR is relatively resistant to desensitization; this optimizes its capacity, especially in the parathyroid gland, for continuous surveillance of extracellular Ca^{2+} levels. Although Ca^{2+}_o is the major ligand for the CaSR, a variety of other agonists have been identified. Mg^{2+} (another biologically important divalent cation whose extracellular level is maintained within a narrow range similar to, albeit slightly lower than, that of Ca^{2+}) has been shown, like Ca^{2+}_o , to inhibit PTH secretion. Although this effect seems to occur at supraphysiological levels, magnesium's action on the CaSR in the kidney, where the local concentration of Mg^{2+} can be higher, may regulate its renal reabsorption and contribute to Mg^{2+}_o homeostasis.

Other well-established agonists acting on the CaSR's ECD are gadolinium, neomycin, and spermine; all of which exhibit positive cooperativity with regard to their interactions with calcium in activating the receptor. These three agonists, as well as Mg^{2+}_o , are defined as type 1 agonists, which bind to the ECD and activate the CaSR regardless of whether Ca^{2+}_o is present. Type 2 agonists, by contrast, are allosteric activators that require calcium to exert their effects. Certain L-amino acids, particularly aromatic amino acids, have been shown to be type 2 agonists, binding to a site within the CaSR's ECD

close to that of Ca^{2+} . It is possible that they function in this manner in the gut, where the CaSR is widely distributed, enabling the CaSR to act, in effect, as a nutrient receptor. The CaSR also senses ionic strength, as illustrated by the fact that exposing dispersed bovine parathyroid cells to a 40-mM increment in NaCl shifts the EC_{50} for high Ca^{2+}_o -evoked inhibition of PTH release by at least 0.5 mM. Ionic strength may exert physiologically relevant actions on the CaSR in select regions of the gastrointestinal (GI) tract and kidney, where the levels of ionic strength can vary greatly.

Intracellular Signaling Pathways Utilized by the CaSR

CaSR agonists activate phospholipases C, A_2 , and D in parathyroid as well as in CaSR-transfected but not in nontransfected human embryonic kidney (HEK293) cells. The high Ca^{2+}_o -evoked, transient rise in the cytosolic calcium concentration (Ca^{2+}_i) in bovine parathyroid and CaSR-transfected HEK293 cells is mediated by activation of PLC and the resultant inositol triphosphate (IP_3)-mediated release of Ca^{2+} from intracellular stores. In parathyroid and CaSR-transfected HEK293 cells, the activation of PLC is insensitive to pertussis toxin and involves $\text{G}_{\alpha_{q/11}}$. The CaSR also stimulates the p42/44 and p38MAPKs as well as the c-Jun-N-terminal kinase (JNK) by pathways that involve protein kinase C and/or cytoplasmic tyrosine kinases (e.g., c-Src) and inhibit adenylate cyclase via G_{α_i} in parathyroid and some kidney cells. In other cells, the CaSR inhibits the accumulation of cyclic adenosine monophosphate (cAMP) by increasing Ca^{2+}_i , which then inhibits the activity of a Ca^{2+} -inhibitable isoform of adenylate cyclase. Additional CaSR-linked signaling pathways include stimulation of ceramide generation and activation of Rho kinase, phosphatidylinositol 3-kinase (Akt), and phosphatidylinositol 4-kinase.

Feedback regulation of the CaSR by protein kinase C

The CaSR has two protein kinase A (PKA) and five protein kinase C (PKC) phosphorylation sites within its intracellular domains. Phosphorylation of these PKC sites, especially the one located at threonine 888, inhibits the receptor's coupling to PLC, one of the CaSR's principal intracellular signaling pathways. This serves as a negative feedback loop because PKC is activated downstream of PLC. Phosphorylation of the PKA sites has only minor impact on the receptor's function.

The Importance of the CaSR in Calcium Homeostasis

The level of calcium in the blood is regulated by three calcitropic hormones: PTH, calcitonin (CT), and $1,25(\text{OH})_2$ vitamin D_3 . There is an inverse sigmoidal relationship between PTH, a key calcium-elevating hormone, and Ca^{2+}_o (Figure 2), whereas there is a positive relationship between Ca^{2+}_o and CT, the calcium-lowering hormone, both of which are mediated by the CaSR. These features confer on the CaSR the role of the gatekeeper to normal calcium homeostasis. The additional Ca^{2+}_o -elevating hormone, $1,25(\text{OH})_2$ vitamin D_3 , like PTH, displays an inverse relationship with Ca^{2+}_o , which is likely also CaSR mediated. In addition to the effects of the

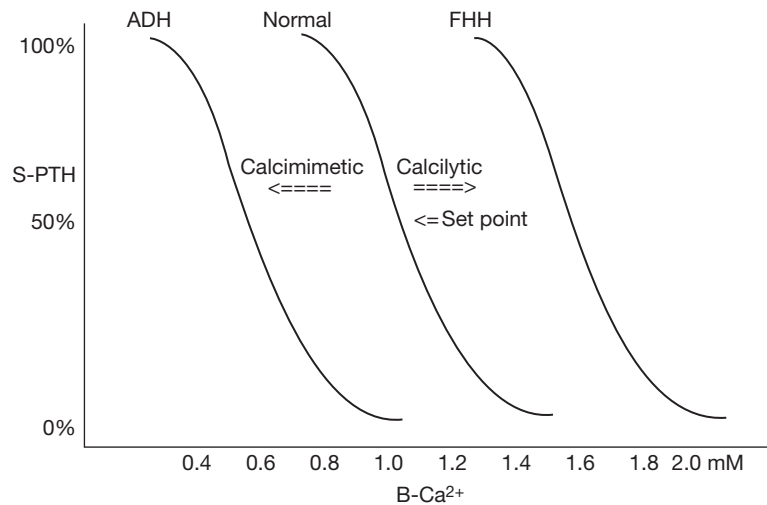


Figure 2 Sigmoidal relationship between the B-Ca²⁺ and serum PTH (S-PTH) levels in normal individuals (eucalcemia), in autosomal dominant hypoparathyroidism (ADH) (hypocalcemia), and in familial hypocalciuric hypercalcemia (FHH). The arrows represent the effects of calcimimetics (shifting the set point to the left) and calcilytics (shifting the set point to the right). Set point is defined as the level of calcium that produces one-half of the maximal inhibition of PTH secretion. Reproduced from Tfelt-Hansen J, Schwarz P, Brown EM, and Chattopadhyaya N (2003) *Frontiers in Bioscience* 8: S377–S390.

CaSR on calciotropic hormone production and secretion, the receptor also directly modulates the function of target tissues for PTH, especially in the kidney and perhaps in the bone and intestine. In this way, Ca²⁺_o functions as a first messenger, serving as the body's major Ca²⁺_o-lowering hormone (described later).

Role of the CaSR in the Defense against Hypercalcemia

The responses of the CaSR to hypercalcemia provide an instructive example of its role in calcium homeostasis.

Parathyroid

An increase in Ca²⁺_o is sensed by the chief cells in the parathyroid gland, which within seconds lowers the rate of PTH secretion. Additional CaSR-mediated effects of hypercalcemia on parathyroid function are inhibition of PTH gene expression, suppression of parathyroid proliferation, and upregulation of the vitamin D receptor (VDR), which mediates the inhibitory actions on 1,25(OH)₂D₃ on parathyroid function. The phosphaturic hormone, fibroblast growth factor-23 (FGF-23), is another, recently described, negative regulator of parathyroid function, although its importance in Ca²⁺_o homeostasis is not fully elucidated. The impact of the hypercalcemia-evoked reduction in PTH secretion and the resultant direct and indirect actions of Ca²⁺_o on cellular function occur in several target tissues.

Bone

The high Ca²⁺_o-elicited lowering of PTH decreases the rate of release of calcium ions from bone – an action mediated through its action on the PTH receptor (PTHr) on the osteoblast. The production of 1,25(OH)₂D₃ is likewise inhibited by hypercalcemia – and this fall in 1,25(OH)₂D₃ also reduces bone resorption. High Ca²⁺_o, by itself, also exerts negative

effects on osteoclast formation and function and increases osteoblast activity, thereby reducing the net release of calcium from bone. Taken together, the effects of PTH, 1,25(OH)₂D₃, and Ca²⁺_o on bone all promote lowering of the level of blood calcium.

Kidney and intestine

The high Ca²⁺_o-elicited decrease in the circulating level of PTH also promotes rapid excretion of calcium by the kidney. This action is thought to take place through a reduction in PTH-stimulated calcium reabsorption in the thick ascending limb of the loop of Henle as well as in the distal convoluted tubule. However, high Ca²⁺_o also directly inhibits renal tubular calcium reabsorption by virtue of the presence of the CaSR on the basolateral surface of the epithelial cells of the thick ascending limb, where the receptor antagonizes the stimulatory action of PTH on calcium reabsorption. Both the high Ca²⁺_o-elicited decrease in PTH and the direct inhibitory action of high Ca²⁺_o on the 1-hydroxylation of 25-hydroxyvitamin D₃ also inhibit the production of 1,25(OH)₂D₃, thereby reducing the absorption of Ca²⁺_o by the intestine. All these effects of hypercalcemia on the kidney, and indirectly on the intestine, reduce the level of Ca²⁺_o toward normal. When combined with the effects of a reduction in PTH and an increase in Ca²⁺_o on bone, the Ca²⁺_o homeostatic system provides an elegant mechanism for maintaining near constancy of Ca²⁺_o.

Nonhomeostatic Functions of the CaSR

Although the physiological roles of the CaSR in cells that do not participate in mineral ion homeostasis still remain to be fully defined, studies of the CaSR's functions in these diverse cell types have widened the range of functions that the CaSR is known to regulate. These CaSR-modulated processes include (1) regulating the secretion of peptides, for example,

adrenocorticotrophic hormone (ACTH), gastrin, insulin, growth hormone, rennin, and parathyroid hormone-related protein (PTHrP); (2) modulating ion channel/transporter activity, for example, aquaporin-2 water channels, nonselective cation channels, voltage-dependent Ca^{2+} channels, and calcium-activated potassium (K^+) and other K^+ channels; (3) controlling gene expression, for example, increasing expression of the vitamin D receptor and the CaSR itself; (4) stimulating the proliferation of ovarian surface epithelial cells and fibroblasts, and inhibiting that of parathyroid cells, keratinocytes, and colonic cells; (5) promoting the differentiation of keratinocytes, goblet cells, and mammary epithelial cells; (6) regulating the apoptosis of fibroblasts, HEK-293 cells stably transfected with the CaSR, and prostate cancer cells; (7) enhancing the chemotaxis of preosteoblastic cells and macrophages; and (8) promoting cellular adhesion (i.e., of hematopoietic stem cells to their bone marrow niche). Thus, Ca^{2+}_o may serve as an extracellular messenger regulating diverse cellular functions, not only in cells directly involved in Ca^{2+}_o homeostasis but also in the variety of additional cell types expressing the CaSR that play no apparent role in this process. In addition, increasing evidence implicates the CaSR in diseases of various non-homeostatic tissues, such as in suppressing the development of colon and breast cancer, inhibiting vascular calcification and atherosclerosis in CKD, and modulating the risk of pancreatitis.

Pharmaceutical Aspects

The CaSR has become a drug target as a consequence of its central role in calcium homeostasis and the evidence given earlier that it contributes to hyper- and hypocalcemic states. There are two classes of drugs acting on the CaSR. One group activates the receptor – these are allosteric activators and are called ‘calcimimetics’. The other group inhibits the effect of Ca^{2+}_o on the CaSR – these agents are called ‘calcilytics’.

Calcimimetics

Calcimimetics are type 2 activators that bind to the receptor’s TMDs and potentiate the action of Ca^{2+}_o on the CaSR. NPS R-568 and cinacalcet are two such agents; they shift the curve for Ca^{2+}_o -regulated changes in Ca^{2+}_i and PTH release to the left without affecting the maximal or minimal responses (Figure 2). Clinical trials have shown that cinacalcet can effectively lower serum PTH concentration in patients with stage 5 chronic kidney disease (CKD) (i.e., those receiving dialysis

treatment), and it has received regulatory approval for use in this setting. Cinacalcet is also approved for the treatment of hypercalcemia in patients with parathyroid carcinoma. Cinacalcet normalizes serum calcium concentration in the majority of patients with mild primary hyperparathyroidism. It has been approved for use in this setting in Europe but not in the US. The drug has been utilized in various other clinical settings but it has not been approved for any of these.

Calcilytics

This group of agents has the opposite effect, inhibiting the receptor and thereby stimulating PTH secretion (Figure 2). This class of drugs could potentially be useful in the treatment of osteoporosis by providing brief pulses of endogenous PTH secretion. The resultant transient increase in PTH mimics the elevation in PTH caused by once daily administration of exogenous PTH by subcutaneous injection. The latter is known to exert anabolic actions on bone and is approved by the US FDA for use in the treatment of osteoporosis. Clinical trials are underway investigating the safety and efficacy of various short-acting calcilytics as a treatment for osteoporosis.

See also: **Bioenergetics:** Calcium, Biological Fitness of; **Signaling:** Mitogen-Activated Protein Kinase Family; Phospholipase C; Phospholipase D; Protein Kinase C Family.

Further Reading

- Brown EM (2008) Biology of the extracellular Ca^{2+} -sensing receptor. In: Bilezikian JP, Raisz LG, and Martin TJ (eds.) *Principles of Bone Biology*, 2nd edn., pp. 533–554. San Diego, CA: Elsevier.
- Brown EM, Gamba G, Riccardi D, et al. (1993) Cloning and characterization of an extracellular Ca^{2+} -sensing receptor from bovine parathyroid. *Nature* 366: 575–580.
- Egbuna O and Brown EM (2008) Hypercalcaemic and hypocalcaemic conditions due to calcium-sensing receptor mutations. *Best Practice and Research. Clinical Rheumatology* 22: 129–148.
- Huang C and Miller RT (2010) Novel Ca receptor signaling pathways for control of renal ion transport. *Current Opinion in Nephrology and Hypertension* 19: 106–112.
- Messa P, Alfieri C, and Brezzi B (2008) Cinacalcet: Pharmacological and clinical aspects. *Expert Opinion on Drug Metabolism and Toxicology* 4: 1551–1560.
- Peterlik M, Grant WB, and Cross HS (2009) Calcium, vitamin D and cancer. *Anticancer Research* 29: 3687–3698.

Relevant Website

<http://www.casrdb.mcgill.ca> – CASRdb: Calcium-Sensing Receptor Database.

Calcium Signaling by Cyclic ADP-Ribose and NAADP

L Santella and J T Chun, Stazione Zoologica Anton Dohrn, Villa Comunale, Napoli, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Acrosome reaction A Ca^{2+} -mediated process occurring in the sperm upon its interaction with the egg envelopes. In echinoderms, following sperm receptor activation, the globular actin in the acrosome vesicle polymerizes to form filaments that extend outside the sperm head to reach the egg surface. In the reaction, lytic enzymes are liberated by the sperm to facilitate its penetration through the egg envelopes.

Cortical reaction A massive exocytosis of the cortical granules (acidic secretory vesicles) located beneath the egg plasma membrane. Shortly after egg activation by the sperm, these granules fuse with the plasma membrane and discharge their contents into the perivitelline space

beneath the vitelline coat. The latter becomes altered, permitting the formation and elevation of the fertilization envelope.

Ectoenzyme Enzyme located in the plasma membrane with the active site facing the extracellular medium.

Inositol 1,4,5-trisphosphate (InsP_3), cyclic ADP-ribose (cADPr), nicotinic acid adenine dinucleotide phosphate (NAADP) Second messengers that liberate Ca^{2+} from intracellular organelles.

Intracellular calcium-releasing receptors Proteinaceous structures in the membranes of organelles that bind second messengers to activate a channel in the structure that lets through Ca^{2+} .

Ca^{2+} -Releasing Activity of Nicotinamide Adenine Dinucleotide and Nicotinamide Adenine Dinucleotide Phosphate

The development of a cell-free system using sea urchin egg homogenates and purified microsomes derived from the endoplasmic reticulum (ER) that possessed adenosine triphosphate (ATP)-dependent Ca^{2+} sequestration showed that sea urchin eggs possess systems for releasing Ca^{2+} from internal stores different from those activated by 1,4,5-inositol trisphosphate (InsP_3). Since it was already known that the concentrations of pyridine nucleotides dramatically changed after fertilization, a series of phosphorylated molecules were assayed to determine whether they could release Ca^{2+} . It was thus found that, in addition to InsP_3 , the oxidized form of nicotinamide adenine dinucleotide (NAD^+) and its phosphorylated form (NADP^+) were effective in releasing Ca^{2+} from the sea urchin homogenate system. While InsP_3 and NADP^+ released Ca^{2+} immediately upon addition, the release induced by NAD^+ showed a lag period of 1–2 min. These results suggested the enzymatic conversion of NAD^+ to a true Ca^{2+} -releasing metabolite. The metabolite was indeed purified, crystallized, and characterized as a cyclized adenosine diphosphate (ADP)-ribose (cADPr) which is synthesized from NAD^+ by the ADP-ribosyl cyclase (ARC), a soluble 29-kDa protein first purified from the sea snail *Aplysia californica*. ARC was prominently localized on granules and vesicles of immature oocytes in the *Aplysia* ovotestis: its crystallized structure showed extensive sequence similarity to human CD38, a type II cell surface transmembrane glycoprotein involved in the regulation of lymphocyte adhesion to endothelial cells, and to CD157 (BST-1), another related molecule with similar enzyme activities. Both CD38 and CD157 are ecto-enzymes, the catalytic domains of which are located outside the cell. They have both ARC and cADPr hydrolase activities, that is, they synthesize cADPr from NAD^+ and hydrolyze it to ADP-ribose.

The localization of the active site of CD38 at the ectocellular domain raised a number of questions on the mechanism of the CD38/cADPr system. The ectocellular synthesis of cADPr whose

only known function, that is, calcium mobilization, is necessarily intracellular was apparently paradoxical. Somehow, cADPr could gain access to the interior of the cells. Extracellular cADPr, either produced by CD38 or directly added to the medium, elicited an enhanced intracellular Ca^{2+} response in several cell types including fibroblasts. Thus, the possibility that CD38, in addition to generating cADPr, could actively transfer it across the plasma membrane into the cytosol is plausible.

The initial studies on the receptor function of CD38 showed that it not only mediated cell adhesion but also initiated transmembrane signaling in response to the interaction with its ligand CD31 or with agonistic CD38 antibodies, which resulted in tyrosine phosphorylation of intracellular proteins such as extracellular signal-regulated kinase (ERK) and the c-cbl proto-oncogene product (p120c-cbl) and in the upregulation of cytokine production. That various transmembrane signaling events were mediated by CD38 was demonstrated with CD38 knockout mice. In these mice, the ARC activity was dramatically decreased and, as a result, cADPr levels were decreased in all tissues examined with the exception of heart and brain. In pancreatic cells of the knockout mice, acetylcholine failed to trigger Ca^{2+} spikes since the agonist failed to activate cADPr production. CD38-deficient mice were also more susceptible to bacterial infections, as the failure to produce cADPr eventually prevented the Ca^{2+} -dependent migration of neutrophils to the infection sites. CD38 knockout female and male mice also showed marked defects in maternal nurturing and social behavior, respectively, suggesting that CD38 may be important in neurodevelopmental disorders.

Three isoforms of ARC have been cloned and characterized in sea urchin. One of them (SpARC1) is active in the lumen of the ER, indicating once again that messenger production may take place outside the cytosol. In contrast to SpARC1, heterologously expressed SpARC2 is localized to the plasma membrane as a glycosylphosphatidylinositol (GPI)-anchored plasma membrane protein. It has been speculated that SpARC2 could become internalized into the acidic endosomal system in which the base-exchange reaction is favored. As a result, SpARC2 could

preferentially catalyze the reaction that generates another second messenger, nicotinic acid adenine phosphate (NAADP), from a different substrate, NADP^+ . Conversely, the neutral luminal environment of SpARC1 would favor the cyclization reaction, thus predominantly promoting the synthesis of cADPr from NAD^+ . It must be mentioned that electron microscopy and immunogold labeling studies in sea urchin eggs have instead indicated that, in addition to the ectocellular isoform (ARC α), two other isoforms ARC β and ARC γ reside within the lumen of the acidic cortical granules that are responsible for the elevation of the fertilization envelope following egg activation.

Increasing evidence shows that CD38 may also reside in different subcellular compartments of many invertebrate and mammalian tissues. CD38 has been detected in the inner membrane of the nuclear envelope as a nonmembranous form of the enzyme that is linked to proteins involved in nuclear processes.

Analogues of cADPr have been synthesized, the first having amino (8-NH₂), azido (8-N₃), or bromo (8-Br) groups in the adenine ring. All three are competitive antagonists of cADPr-induced Ca^{2+} release in sea urchin egg homogenates. Being more potent, 8-NH₂-cADPr has become the reagent of choice to antagonize the effects of cADPr. As mentioned, unlike NAD^+ , NADP^+ induced Ca^{2+} release from egg homogenates without delay: in principle, then, no enzymatic conversion of NADP^+ appeared to be required. However, a simple alkaline treatment of NADP^+ greatly increased its activity to release Ca^{2+} , indicating that what is responsible for Ca^{2+} liberation might not be NADP^+ itself but its metabolic derivative. It was indeed shown that both CD38 and the *Aplysia* cyclase can take NADP^+ as a substrate and substitute its nicotinamide group with nicotinic acid to produce NAADP. Microinjection of NAADP into live eggs of sea urchin induced transient elevation of intracellular Ca^{2+} and triggered exocytosis of the cortical granules as InsP_3 would do. A unique property of the NAADP-sensitive Ca^{2+} release mechanism was its insensitivity to the antagonists of all the intracellular Ca^{2+} release channels ever known. However, subthreshold concentration of NAADP that does not itself release Ca^{2+} in sea urchin eggs prevented Ca^{2+} release in response to a second addition of NAADP. Likewise, extremely high concentration of NAADP does not release Ca^{2+} , suggesting that NAADP may bind to its receptor irreversibly. Another important point is that the NAADP-mediated Ca^{2+} release was not regulated by Ca^{2+} . However, it can still prime Ca^{2+} -induced Ca^{2+} release mechanisms through other Ca^{2+} channels. Early findings on such a cross-talk between NAADP and other intracellular Ca^{2+} channels were reported in pancreatic acinar cells, a popular cell model in the study of specific Ca^{2+} signals. In these cells, both neurotransmitter acetylcholine and the hormone cholecystokinin induce Ca^{2+} oscillations. However, those produced by the neurotransmitter are mediated by InsP_3 and Ca^{2+} , whereas those produced by the hormone are linked to the generation of cADPr and NAADP. It has been shown that NAADP acts at very low concentration and sensitizes cADPr and InsP_3 receptors.

cADPr- and NAADP-dependent Ca^{2+} Stores

Early evidence that the cADPr- and the NAADP-sensitive Ca^{2+} stores were physically distinct and separable from that of InsP_3 came from cell fractionation and cytoplasmic organelle

stratification of sea urchin eggs. Global photoactivation of the caged NAADP, injected into the stratified eggs, induced a localized pattern of Ca^{2+} . The NAADP-induced release occurred mainly in the pole distal to the nucleus, while the stores sensitive to cADPr localized to the nuclear pole and co-purified with ER markers in Percoll gradients. Uniform photo-activation of caged InsP_3 produced a pattern of Ca^{2+} mobilization that was similar to that of caged cADPr, but clearly distinct from that of NAADP. The identity of cytoplasmic organelle from which NAADP mobilized Ca^{2+} , which was evidently different from the ER, remained unknown for many years. Then, it was found that incubation of intact sea urchin eggs with glycyl-L-phenylalanine 2-naphthylamide (GPN), a substrate of the lysosomal exopeptidase cathepsin C which causes the loss of lysosome integrity, reduced the responsiveness to NAADP. The suggestion that NAADP could specifically mediate Ca^{2+} release from the acidic lysosomal compartment was further substantiated by the requirement of acidic environment for ARC to produce NAADP.

The implication of the lysosome studies is interesting, but the exact nature of the stores targeted by NAADP is still controversial. Several studies have, for instance, suggested that the ryanodine receptors (RyRs) in the ER could be a possible alternative candidate in a sense that the nuclei isolated from pancreatic cells are deprived of acidic organelles but still release Ca^{2+} from the nuclear envelope (that forms a continuity with the ER) when challenged with NAADP. In permeabilized pancreatic acinar cells, InsP_3 , cADPr, and NAADP can all release Ca^{2+} from the fraction of the ER located in the basal region, as well as from an acidic store in the secretory granular region at the opposite pole of the cell. It was claimed that the action of NAADP is dependent on ryanodine receptors operational in the ER. Thus, the NAADP-induced release from the acidic stores could be due to the extension of the ER into the regions containing the acidic granules. Further evidence suggesting the ER as the Ca^{2+} store targeted by NAADP comes from studies where NAADP releases Ca^{2+} from a thapsigargin-sensitive store, and from membranes incorporated into planar lipid bilayers. In the latter case, low concentrations of NAADP activated single-channel current with the characteristics of type 2 RyRs.

The extremely low level at which NAADP acts has complicated the measurements of its changes upon stimulation. Recent technological advances have made possible to detect both the basal levels of NAADP and its increases following cell stimulation in many cell types, including insulin-secreting pancreatic β -cells, human myometrical cells, and Jurkat cells. In sea urchin sperm, NAADP is rapidly synthesized upon exposure to the egg jelly coat and activates cation channels localized in the acrosome, a lysosome-related organelle involved in the formation of the acrosomal process.

cADPr-Mediated Ca^{2+} Signaling in Cells

The sea urchin eggs have been a popular system for studying Ca^{2+} mobilization since the intracellular Ca^{2+} stores express multiple Ca^{2+} release channels creating the fertilization Ca^{2+} wave. At sperm-egg interaction, a fast Ca^{2+} influx into the egg occurs as a result of the plasma membrane depolarization induced by the sperm (the cortical flash), which is followed

by an intracellular Ca^{2+} wave that propagates to the antipode of the egg. The rapid and transient increase in the intracellular concentration of cyclic guanosine monophosphate (cGMP) at fertilization, together with the finding that cGMP could mobilize Ca^{2+} in sea urchin eggs and homogenates supplemented with NAD^+ , suggested that ARC is activated by a cGMP-dependent protein kinase to produce cADPr, which in turn activates ryanodine receptor Ca^{2+} channels. This signaling pathway was further supported by the finding that nitric oxide (NO) could also evoke Ca^{2+} signals in sea urchin eggs through the activation of guanylyl cyclase. However, later work has suggested that the NO/cGMP/cADPr pathway may serve as a 'regulator' of the later phases of the Ca^{2+} transient at fertilization (see below), rather than as an 'activator' of the Ca^{2+} response. Thus, cADPr appears to play roles in both initiating and sustaining the Ca^{2+} response induced by the sperm.

It has also been shown that cADPr is generated in human pancreatic β -cells by glucose stimulation to mobilize Ca^{2+} from the ER in the process of insulin secretion. Interestingly, cADPr does not bind directly to the type 2 RyR expressed in rat pancreatic islets but acts on it through a mediator such as the FK506-binding protein 12.6. The binding of cADPr to FKBP12.6 causes the dissociation of the latter from the receptor, thereby increasing its Ca^{2+} -releasing channel activity. Hence, it was suggested that cADPr may serve as an endogenous modulator of Ca^{2+} -induced Ca^{2+} release (CICR) which is a key property of RyRs.

cADPr and NAADP at Fertilization

The fertilization of eggs from invertebrates to mammals is accompanied by single or repetitive Ca^{2+} transients. The transient increase of Ca^{2+} propagates across the egg as a Ca^{2+} wave, which is necessary for the egg. As mentioned, the global Ca^{2+} wave is preceded by the cortical flash which is produced by Ca^{2+} influx through voltage-gated channels activated by sperm-induced depolarization of the egg plasma membrane (Figure 1). The proposal that the InsP_3 pathway is responsible for the generation or propagation of Ca^{2+} transients has gained general support on a number of findings: microinjection of InsP_3 into eggs induces a Ca^{2+} elevation accompanied by the increase of phosphoinositide metabolism immediately after fertilization. In eggs preloaded with heparin, a potent competitive inhibitor of the InsP_3 -induced Ca^{2+} release, the Ca^{2+} transient of fertilization was delayed and reduced in both velocity and amplitude. Furthermore, intracellular injection of the tandem Src Homology 2 (SH2) domains of bovine

phospholipase $\text{C}\gamma$ (PLC γ) produced a concentration-dependent inhibition of the Ca^{2+} release. In addition, cADPr is as effective as InsP_3 in liberating Ca^{2+} from the intracellular stores of sea urchin eggs. Sea urchin microsomes respond to cADPr even after their desensitization to a high concentration of InsP_3 , demonstrating that a distinct cADPr-linked Ca^{2+} mobilization mechanism exists in these eggs. The cADPr-sensitive Ca^{2+} release is independent of InsP_3 , as it was insensitive to heparin. The cADPr-sensitive Ca^{2+} release mechanism critically depends on calmodulin (CaM). Sea urchin microsomes do not respond to cADPr even at concentrations as high as 100 μM , but CaM restores their sensitivity to nanomolar concentrations of cADPr. At fertilization, cADPr thus acts synergistically with InsP_3 to sustain the intracellular Ca^{2+} wave. It has recently been shown that cADPr is also synthesized in the lumen of the acidic exocytotic vesicles (cortical granules) located beneath the plasma membrane, where it provides local Ca^{2+} signals to drive exocytosis. On the other hand, cADPr shuttled into cytosol through a separate cADPr transporter contributes to the formation of a global Ca^{2+} wave.

As discussed above, NO is also involved in the Ca^{2+} swings during the initial and later phases of the sperm-induced Ca^{2+} response. NO is produced in the sperm undergoing acrosome reaction and in the eggs during fertilization. NO would stimulate the production of cGMP and activates ARC to produce cADPr. However, the precise temporal relationship between the rises of NO and the Ca^{2+} transient was not fully resolved. As the enzyme synthesizing NO is also activated by Ca^{2+} , it appeared that the main role of the NO/cGMP/cADPr pathway was to sustain the later phase of Ca^{2+} transient during the fertilization of sea urchin eggs. However, the fertilization-induced increase in cGMP was also observed to precede the explosive Ca^{2+} elevation in other species of sea urchin eggs. Hence, the controversy remains to be resolved, although it is possible that cGMP production could be localized strictly around the sperm entry site and induce a local Ca^{2+} increase by the production of cADPr.

As mentioned, the RyR complex is the target of cADPr. RyR antagonists such as ruthenium red, Mg^{2+} , and procaine were shown to inhibit the cADPr-induced Ca^{2+} release, while the agonists of the CICR such as caffeine, Ca^{2+} , and Sr^{2+} potentiated the effect. In mouse eggs, activation of the cADPr/ryanodine pathway induced the conversion of the zona pellucida glycoprotein ZP2 to its postfertilization form ZP2f. However, the treatment of the eggs that would block the release of Ca^{2+} from the ryanodine-sensitive store had no effect on egg activation following fertilization. Thus, the release of Ca^{2+} from RyRs is not essential for sperm-induced egg activation.

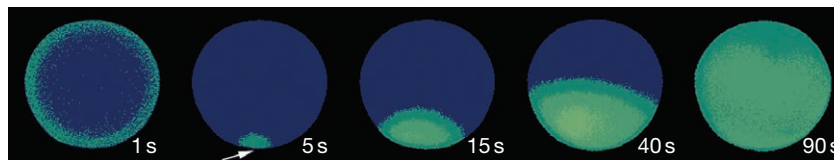


Figure 1 Spatio-temporal changes of the fertilization-induced Ca^{2+} signals in mature eggs of starfish. The first image shows the Ca^{2+} release in the cortex of the cell (cortical flash). This is followed by the traversing Ca^{2+} wave that spreads from the site of sperm-egg interaction (arrow) to the cytoplasm of the opposite side of the egg. Ca^{2+} changes were monitored with the Ca^{2+} indicator, calcium green dextran 10 000 MW. The Ca^{2+} increase was monitored by employing a cooled CCD camera, and the relative fluorescence images were presented at different time points after sperm-egg contact.

It nevertheless has an important role: it modulates the recharge stage of the ER Ca^{2+} store. In *Xenopus* oocytes lacking the RyR system, photoliberation of caged InsP_3 following the injection of the nonmetabolizable analog 3-deaza-cADPr strongly potentiated the Ca^{2+} response.

Microinjection of NAADP into live sea urchin eggs also induces a large transient increase of intracellular Ca^{2+} . The injected egg, however, became desensitized to NAADP, that is, they failed to respond to a second injection of NAADP. The Ca^{2+} release induced by NAADP is kinetically faster than that induced by cADPr or InsP_3 and insensitive to 8-NH₂cADPr or heparin. Thus, the NAADP-linked mechanism is independent of cADPr or InsP_3 . Conversely, in microsomes desensitized by subthreshold concentrations of NAADP, both cADPr and InsP_3 can still induce Ca^{2+} release.

Ca^{2+} Response to cADPr and NAADP in Starfish Eggs

Starfish oocytes have contributed important information to the understanding of the Ca^{2+} signaling induced by cADPr and NAADP. Being arrested at prophase I of meiosis, immature oocytes of starfish are characterized by the presence a large nucleus (up to 60 μm in diameter). With the use of the hormone, 1-methyladenine (1-MA), the spatiotemporal pattern of cADPr- and NAADP-promoted Ca^{2+} release can be studied during the maturation process.

The induction of the maturation process by 1-MA induces profound changes in the oocytes along with the increased sensitivity of the Ca^{2+} stores to InsP_3 . This is more evident in the perinuclear area near the animal hemisphere, as ultraviolet (UV) irradiation of fully diffused caged InsP_3 starts to release Ca^{2+} specifically in these areas before the Ca^{2+} wave propagates to the entire oocyte along the animal/vegetal axis.

Maturation also enhances the eggs' Ca^{2+} response to cADPr. The photoactivation of the same amount of caged cADPr induces a higher Ca^{2+} increase in the eggs matured with 1-MA. The Ca^{2+} signal starts at one or two circumscribed sites in the cortex, and expands to a cortical flash and then spreads centripetally to the remainder of the cell. The cortical flash is probably linked to Ca^{2+} influx from the extracellular medium, as it is blocked when the cADPr was uncaged in the eggs washed several times in CaFSW. The cADPr-induced Ca^{2+} responses thus closely resemble those induced by the sperm. However, the physiological role of cADPr-dependent Ca^{2+} signaling at fertilization is still questionable. Although *Astropecten arancia-cus* eggs respond to microinjected cADPr with Ca^{2+} release, starfish eggs of other species, for example, *Asterina pectinifera*, did not display the same response. Furthermore, pre-injection of 8-NH₂cADPr into mature eggs blocks the cADPr-induced Ca^{2+} increase but not the one induced by the sperm, indicating that this second messenger may not play an essential role in shaping the Ca^{2+} waves at fertilization.

Determination of the precise spatial distribution of the NAADP-induced Ca^{2+} signals in starfish eggs also contributed to the understanding of the mechanisms underlying the sperm-induced Ca^{2+} increase at fertilization. The photoliberation of caged NAADP into immature and mature eggs induces a cortical Ca^{2+} increase that spreads throughout the cytoplasm as a wave due to the amplification of the signals by the InsP_3 and ryanodine receptors through the CICR mechanism (Figure 2(a)). The uncaging of NAADP in CFSW failed to promote the Ca^{2+} response. Suggesting its involvement of Ca^{2+} influx, the NAADP-induced Ca^{2+} signal was shown to be inhibited by SK&F 96365, an inhibitor of receptor-operated calcium entry, and by the L-type calcium channel blocker verapamil. The NAADP-elicited Ca^{2+} influx is the result of the activation of a cationic membrane current requiring both an initial

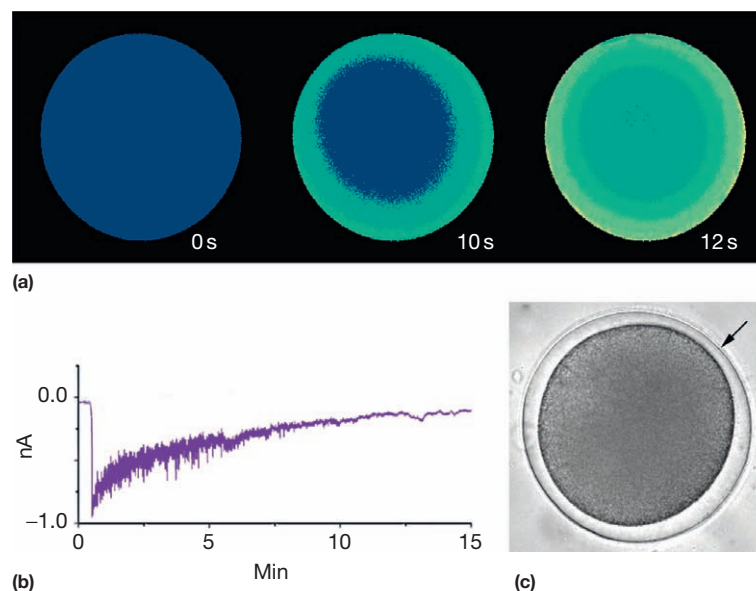


Figure 2 Ca^{2+} influx induced by NAADP in matured eggs of starfish. (a) The global photoliberation of pre-injected caged NAADP induces a Ca^{2+} response that occurs in the cortical region of the egg and spreads subsequently to the center of the cell. (b) The NAADP-elicited Ca^{2+} influx is associated with a membrane current. Removal of extracellular Ca^{2+} abolishes the current, whereas the replacement of external Na^+ with an equimolar amount of choline does not affect it. (c) The NAADP-linked Ca^{2+} release promotes the elevation of the vitelline envelope (arrow).

submembrane Ca^{2+} pulse and NAADP (Figure 2(b)). It was also suggested that NAADP might trigger a highly localized cortical Ca^{2+} increase from a thapsigargin-sensitive store that would be tightly coupled to the NAADP-dependent membrane channels, raising the possibility that NAADP could play an important role in the initiation of the Ca^{2+} response at fertilization. The finding that the changes in membrane potential induced by the sperm were similar to that induced by NAADP supports the suggestion. This triggering role has also been supported by experiments on immature oocytes from which the nucleus had been removed before the induction of maturation by 1-MA. The removal of the nucleus does not affect the initial cortical Ca^{2+} response by NAADP or by the sperm but blocks the globalization of the Ca^{2+} wave which is due to InsP_3 . Subsequent work has shown that NAADP also activates Ca^{2+} influx in sea urchin eggs. The desensitization of the eggs by pretreatment with NAADP strongly reduced the cortical flash and the subsequent generation of the Ca^{2+} wave at fertilization, indicating that NAADP is involved in sperm-induced Ca^{2+} signaling and activation of the eggs. The source of NAADP that acts on the egg activation remains obscure, but the presence of a large amount (5 μM) of NAADP sea urchin sperm suggests a possibility of direct transfer from the sperm. In sea urchin eggs, NAADP has been proposed to mobilize Ca^{2+} from lysosome-like acidic stores, but the phenomenon was not confirmed in starfish eggs. As echinoderm, the cortical distribution of lysosomes in starfish eggs (Figure 3) was similar and suggestive of their involvement in the NAADP-elicited Ca^{2+} release. However, drugs disrupting the acidic compartments failed to affect

the NAADP-induced depolarization and cortical Ca^{2+} release, indicating that lysosomes are probably not the NAADP target in starfish eggs.

Identification of TPCs as NAADP Receptors

The search for the molecular identity of NAADP receptors has been, and still is, difficult. The putative NAADP receptors have been solubilized from sea urchin homogenates and were shown to bind their ligand with appropriate affinity and selectivity. NAADP binding was not affected by luminal or cytosolic Ca^{2+} concentrations and showed little dependence on pH. Furthermore, the binding of NAADP was apparently irreversible, as no displacement of bound ^{32}P -NAADP by unlabeled NAADP was observed. The binding depends on K^+ and is not observed when the latter was replaced by Na^+ in the buffer. Following the finding that NAADP targets lysosome-like acidic stores, it was suggested that the mammalian ion channel activated by NAADP to produce a local Ca^{2+} burst from lysosomes could be the transient receptor potential-mucolipin 1 channel (TRP-ML1), which was found to be associated with lysosomes in coronary arterial myocytes. Consistent with this suggestion, a TRP-ML1 antibody blocked NAADP-sensitive Ca^{2+} release from lysosome membranes, purified from the rat liver and coronary arterial myocytes. Silencing TRP-ML1 expression also diminished lysosome-associated Ca^{2+} release in the myocytes. On the other hand, other evidence has been provided that NAADP mobilizes Ca^{2+} from acidic organelle stores through the activation of a family of ion channels known as TPCs, which are new members of the voltage-gated cation channel superfamily. Comparative genomic analysis indicates the presence of up to three family members of these channels in sea urchin and most vertebrate species. In plants, it was suggested that TPC plays a critical role in regulating Ca^{2+} -linked physiological processes essential for resistance and adaptation to environmental stresses. The TPC protein contains 12 putative transmembrane segments that can be divided into two six-transmembrane homologous domains, each containing a pore loop. The cytoplasmic side of the endolysosomal membranes contains the amino and carboxyl termini as well as the large intermediate loop. The molecular cloning of all three sea urchin TPC isoforms that are targeted to acidic vesicles has supported the proposal of NAADP-dependent calcium release from acidic calcium stores. First, NAADP binds TPC2 at high- and low-affinity sites and thereby induces calcium release and homologous desensitization, as was expected from the studies of the then unknown endogenous NAADP receptors. Second, NAADP-evoked calcium signals via TPC2 are ablated by silencing TPC2, and by depletion of acidic calcium stores with bafilomycin A1. Interestingly, NAADP-induced Ca^{2+} signals are biphasic in their pattern. The initial phase of Ca^{2+} release from lysosomes via TPC2 channel is subsequently amplified by CICR from the ER. By contrast, Ca^{2+} release via endosome-expressed TPC1 induces only a spatially restricted quantal vesicular Ca^{2+} release that is not amplified by CICR. In line with this, TPC1 in *Arabidopsis thaliana* is also localized in the vacuolar membrane of guard cells to form a Ca^{2+} release channel. It plays an important role in regulating Ca^{2+} -induced stomatal closure, which is necessary to reduce water loss in response to a variety of external and

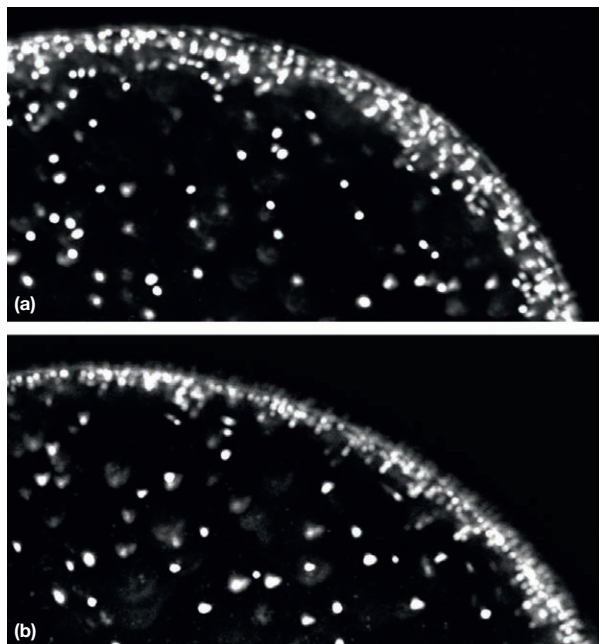


Figure 3 Distribution of acidic organelles labeled by the injection of Lyso-Tracker Red in starfish eggs. (a) The lysosomal dye visualizes a thick layer of vesicles in the cortical domain of immature starfish oocytes (not treated with 1-MA). (b) One hour after hormonal stimulation, the acidic organelles crowded into a thinner layer closer to the plasma membrane, indicating their translocation during the maturation process.

internal stimuli such as light, CO₂, drought, pathogen attack, and plant hormones. Anti-TPC antibodies and organelle markers have shown that TPCs are expressed intracellularly, and specifically targeted to lysosomal membranes. Overexpression of TPC2 gave rise to much higher biphasic Ca²⁺ release in response to uncaged NAADP in the cells that would normally respond with a short-lived small Ca²⁺ signal. This observation is also consistent with the idea that lysosomes play a role in mediating the NAADP-dependent Ca²⁺ release.

Conclusions

The nature of the Ca²⁺-releasing functions of cADPr and NAADP has been characterized, in major part, from the studies on the eggs and fertilization. For that reason, much emphasis was placed on the sperm-egg interaction in this article. However, the two messengers are likely to have more physiological significance in other cell types. For instance, it has been shown that NAADP takes temporal primacy over the well-known InsP₃ route during the Ca²⁺ response induced by activators of mammalian pancreatic acinar cells. In the sperm-egg system, the initial and localized Ca²⁺ response might be induced by NAADP through the initial activation of plasma membrane cationic channels in both the sperm and the egg at fertilization.

The initial local response is followed by the globalization of the Ca²⁺ signal through the activation of specific receptors of InsP₃ and cADPr in the egg. This response is dramatically amplified by the CICR process. Hence, the cross-talk between the signaling systems involving different second messengers may contribute to diversify the ways to fine-tune the Ca²⁺ response in cells.

See also: **Bioenergetics: Intercellular Ca²⁺ Waves: Mechanisms of Initiation and Propagation; Intracellular Calcium Waves.**

Further Reading

- Calcraft PJ, Ruas M, Pan Z, et al. (2009) NAADP mobilizes calcium from acidic organelles through two pore channels. *Nature* 459: 596–600.
- Galione A and Petersen OH (2005) The NAADP receptor: New receptors or new regulation? *Molecular Interventions* 5: 73–79.
- Glick DL, Hellmich MR, Beushausen S, et al. (1991) Primary structure of a molluscan egg-specific NADase, a second-messenger enzyme. *Cell Regulation* 2: 211–218.
- Guse AH and Lee HC (2008) NAADP: A universal Ca²⁺ trigger. *Science Signaling* 1: re10.
- Lee HC (ed.) (2002) *Cyclic ADP-Ribose and NAADP: Structure, Metabolism and Functions*. Dordrecht: Kluwer.
- Moccia F, Lim D, Nusco GA, Ercolano E, and Santella L (2003) NAADP activates a Ca²⁺ current that is dependent on F-actin cytoskeleton. *FASEB Journal* 17: 1907–1909.
- Santella L (2005) NAADP: A new second messenger comes of age. *Molecular Interventions* 5: 70–72.

Calcium Signaling: Calmodulin-Dependent Phosphatase

C B Klee and S Li, National Institutes of Health, Bethesda, MD, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Claude Klee, Hao Ren, and Shipeng Li, volume 1, pp. 241–245, © 2004, Elsevier Inc.

Glossary

A-kinase anchoring protein (AKAP)79/150 It constitutes a scaffold of cAMP-dependent protein kinase, protein kinase C, and calcineurin. The colocalization of these three enzymes with their respective substrates insures the coordination of their phosphorylation pathways.

Calmodulin A unique member of a class of calcium-binding proteins that acts as sensors of changes in intracellular concentration of calcium induced by external signals. It modulates the activity of a large number of enzymes involved in the regulation of cellular processes particularly those involved in

protein phosphorylation, and in cyclic adenosine monophosphate (cAMP) synthesis and degradation.

EF-Hand Structural motif composed of 30 contiguous amino acids forming two helices joined by a Ca^{2+} -binding loop.

Immunosuppressive drugs (FK506 and cyclosporin A (CsA)) Fungal natural products used to prevent organ rejection after organ transplant operation and treatment of autoimmune diseases.

Regulators of calcineurin (RCANs) A family of endogenous protein inhibitors of calcineurin expressed in different tissues and organisms. They modulate calcineurin activity acting as feedback inhibitors.

Subunit Structure and Isoforms

Calcineurin is a heterodimer of a 58–64-kDa catalytic subunit, calcineurin A (CnA), tightly bound, even in the presence of ethylene glycol tetraacetic acid (EGTA), to a 19-kDa Ca^{2+} -binding regulatory subunit, calcineurin B (CnB). This subunit structure, unique among protein phosphatases, is conserved in all eukaryotes. Three mammalian isoforms (α , β , and γ) of CnA have been identified at the messenger RNA (mRNA) level. The corresponding human genes (PPP3CA, PPP3CB, and PPP3CC) are located on human chromosomes 4, 10, and 8 respectively. Additional isoforms, products of alternative splicing, have been detected at the mRNA level. Two mammalian isoforms of CnB, CnB1 and CnB2 (expressed essentially in testis), are the products of two genes (PPP3R1) located on chromosome 2 and (PPP3R2) located on chromosome 9. Although expressed in most, if not all, tissues, calcineurin is particularly abundant in the brain where the α -isoform of CnA is predominant. The β -isoform is broadly distributed while the γ -isoform is expressed essentially in the testis. Calcineurin has been identified in all eukaryotes, including fungi, and the budding yeast *Saccharomyces cerevisiae*. In plants, no calcineurin A has been identified, and calcineurin B-like proteins (CBLs) specifically interact with a family of CBL-interacting kinases (CIPKs).

Structure

The domain structure of calcineurin is equally well conserved. In all species, the catalytic domain located in the N-terminal of two-thirds of CnA, is followed by a CnB-binding domain. The C-terminal regulatory domain contains a CaM-binding and an autoinhibitory domain. In the absence of CaM, the native protein has a very low activity. The catalytic and CnB-binding domains, severed from the regulatory domain by limited proteolysis, constitute the Ca^{2+} -independent truncated form of

calcineurin routinely used in overexpression experiments. CnB is an EF-hand Ca^{2+} -binding protein of the CaM family. It binds four moles of Ca^{2+} , two with high affinity ($k_d < 10^{-7}\text{M}$) and two with moderate affinity in the micromolar range. The N-terminal glycine of CnB is myristoylated. This perfectly conserved posttranslational modification of CnB, apparently not involved in membrane association or required for enzymatic activity, may serve as a stabilizing structural element.

The crystal structure of the recombinant α -isoform of human calcineurin expressed in *Escherichia coli* is illustrated in **Figure 1(a)**. The architecture of the catalytic domain is similar to those of protein phosphatases γ -1 and γ -2A. It consists of a β -sandwich surrounded on one side by seven α -helices and on the other side by a mixed α/β structure. The catalytic site contains an Fe^{2+} – Zn^{2+} dinuclear metal center. The catalytic domain extends into a five-turn amphipathic α -helix whose top face is covered by a 33-Å groove formed by the N- and C-terminal lobes and the C-terminal strand of CnB. A short helical segment blocks the catalytic center. With the exception of this segment, the C-terminal regulatory domain, not visible in the electron density map, is flexible and highly sensitive to proteolytic attack in the absence of CaM. The crystal structure of the proteolytic derivative of bovine calcineurin, lacking the regulatory domain, bound to FK506/FK506-Binding Protein12 (FKBP12), illustrated in **Figure 1(b)**, is similar to that of the full-length protein. The bottom face of the CnB-binding helix of CnA and CnB forms the site of interaction with the FKBP–FK506 and CsA–CyPA complexes. The key role of CnB in forming the drug-binding site provides a molecular basis for the specificity of FK506 and CsA as calcineurin inhibitors.

Enzymatic Properties

Ca^{2+} /CaM-Dependent Stimulation

Calcineurin is a protein phosphatase with a binuclear iron–zinc active center. The purified enzyme, partially depleted of

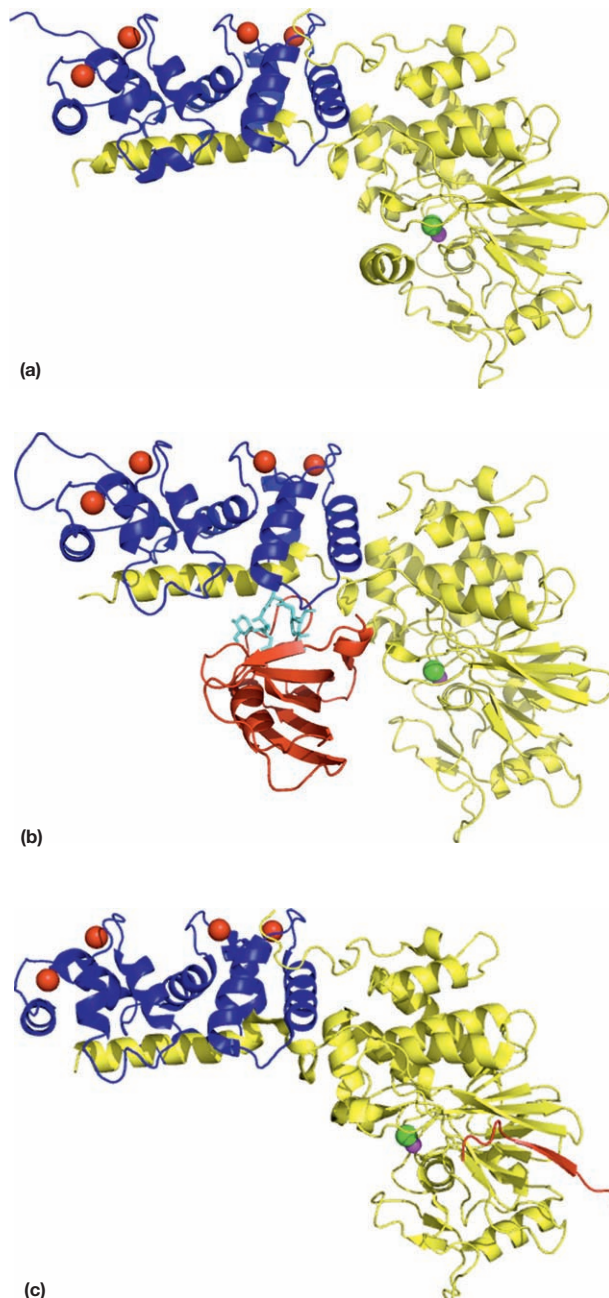


Figure 1 Ribbon representation of the crystal structure of calcineurin. (a) Human recombinant α -calcineurin (PDB code 1AUI). (b) Truncated calcineurin complexed with FKBP12-FK506 (PDB code 1TCO). (c) Human recombinant bovine α -calcineurin complexed with PVIVIT peptide (PDB code 2P6B). CnA is shown in yellow, CnB in blue. Iron and zinc are shown as green and magenta spheres respectively. The four Ca²⁺ bound to CnB sites are shown as red spheres. FKBP12 and PVIVIT peptide are shown in red. FK506 is shown in ball and stick representation. (a) From Kissinger CR, Parge HE, Knighton DR, et al. (1995) Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex. *Nature* 378: 641–644. (b) From Griffith JP, Kim JL, Kim EE, et al. (1995) X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex. *Cell* 82: 507–522. (c) Li H, Zhang L, Rao A, et al. (2007) Structure of calcineurin in complex with PVIVIT peptide: Portrait of a low-affinity signalling interaction. *Journal of Molecular Biology* 369: 1296–1306.

its metal cofactors, is activated by exogenous metal ions (0.1 mM Mn²⁺ or more than 6 mM Mg²⁺). Purified calcineurin, which has retained stoichiometric amounts of Fe³⁺ and Zn²⁺, is inactive and not activated by exogenous metal ions. Activation requires treatment under anaerobic conditions with the reducing agent, ascorbate which converts Fe³⁺ to Fe²⁺. Its phosphatase activity does not require exogenous metal ions. Regardless of its metal content, the phosphatase activity of calcineurin is dependent on the presence of Ca²⁺ and CaM. A small activation is observed upon addition of Ca²⁺ ($K_{act} = 0.5 \mu\text{M}$). Addition of the CaM results in a 50–100-fold increase of the V_{max} without affecting the value of the K_m . The highly cooperative stimulation of the enzyme by CaM (Hill coefficient of 2.5–3) allows calcineurin to respond to narrow Ca²⁺ thresholds following cell stimulation. Because of its high affinity for CaM ($K_{act} \leq 10^{-10} \text{ M}$), the activation of calcineurin in response to a Ca²⁺ signal can precede the activation of all known CaM-regulated kinases.

Enzymatic Assays

A 19-residue synthetic peptide containing the phosphorylation site of the RII subunit of cyclic adenosine monophosphate (cAMP)-dependent protein kinase (PKA) and *p*-nitrophenyl-phosphate are routinely used to measure the phosphatase activity of calcineurin. In crude tissue extracts, calcineurin activity is distinguished from that of other neutral serine-threonine protein phosphatases by its (1) Ca²⁺ dependence; (2) inhibition by CaM-binding peptides and inhibitors; (3) resistance to the inhibitors of protein phosphatases –1 and –2A (Inhibitor-1, DARPP32, Inhibitor-2, okadaic acid, microcystin, and calyculin); and (4) specific inhibition of its protein phosphatase activity by FK506 (but not rapamycin) and CsA in the presence of saturating amounts of their respective binding proteins, FKBP12 and CypA. The *p*-nitrophenyl-phosphate activity of calcineurin, stimulated by FK506 and CsA, cannot be used to measure calcineurin activity in crude extracts. Unlike protein phosphatase-2 C, which requires Mg²⁺ for activity, calcineurin that has retained its natural metal cofactors does not require Mg²⁺.

Substrate Specificity

Calcineurin is a dual protein phosphatase. It dephosphorylates both phosphoserine/threonine and phosphotyrosine peptides. Peptides with a basic residue on the N-terminal side of the phosphorylated amino acid and lacking acidic residue of the C-terminal side are preferentially dephosphorylated by calcineurin. In addition to the recognition of the sequence surrounding the phosphorylated residues, the substrate specificity of calcineurin also depends on the presence of a calcineurin-binding motif, LxVP, which was previously identified in the sequence of a major substrate of calcineurin, the transcription factor, nuclear factor of activated T cells (NFAT). This motif is also present in most other calcineurin substrates. Its presence greatly decreases the K_m values, and LxVP peptides act as competitive inhibitors of calcineurin toward most of its substrates. The presence of an additional aromatic residue at the N-terminus greatly decreases the K_i value.

The calcineurin-binding site of this motif has recently been shown to overlap with the binding site of the immunosuppressive drugs bound to their binding proteins. The transcription factor NFAT also contains another calcineurin-binding motif, PxlIT, responsible for its Ca^{2+} -dependent and phosphorylation-independent anchoring to calcineurin. The anchoring of NFAT allows its dephosphorylation, despite its low intracellular concentration, and the nuclear cotranslocation of calcineurin and NFAT. The crystal structure of human recombinant-truncated calcineurin complexed with the peptide ligand, PVI-VIT, closely related to the docking sequence PxlIT of NFAT3 and NFAT4 anchoring NFAT to calcineurin is shown in [Figure 1 \(c\)](#). PVI-VIT adds as a β strand along the edge of the calcineurin constituted by the strands β 11 and β 14, away from the catalytic center. This peptide prevents NFAT activation but does inhibit calcineurin-mediated dephosphorylation of the RII peptide.

Regulation

Ca^{2+} Regulation

Two structurally similar but functionally different Ca^{2+} -binding proteins, CaM and CnB, mediate the Ca^{2+} regulation of calcineurin. CnB is tightly bound to CnA even in the absence of Ca^{2+} and plays both structural and functional roles. It is required to confer enzymatic activity to the catalytic subunit. The recombinant α - and β -isoforms of CnA expressed in bacteria and SF9 cells require CnB for activity. Ca^{2+} binding to CnB is required for the interaction of calcineurin with the immunosuppressive drugs bound to their respective binding proteins.

Ca^{2+} -dependent binding of CaM to calcineurin induces a conformational change of the regulatory domain accompanied by the displacement of the autoinhibitory domain. The resulting exposure of the catalytic center is sufficient to activate the *p*-nitrophenylphosphatase activity of calcineurin but not its protein phosphatase activity.

Activation of the latter requires CaM binding to the CaM-binding domain to prevent inhibition of calcineurin by the CaM-binding domain. Prolonged exposure of the catalytic center facilitates the oxidation of Fe^{2+} and results in inactivation of calcineurin that is prevented by superoxide dismutase and reversed by ascorbate. Thus, CaM can both activate and inactivate calcineurin depending on the length of the Ca^{2+} signal. The CaM-dependent oxidation of Fe^{2+} and inactivation of calcineurin is a potential mechanism to regulate its activity during oxidative stress.

Endogenous Regulatory Proteins

Endogenous regulatory proteins, whose expression varies from tissue to tissue, modulate the activity of calcineurin *in vivo*. A family of 22–24-kDa proteins, regulators of calcineurin (RCANs), identified in both yeast and mammalian cells acts as feedback inhibitors of calcineurin. The mammalian proteins are identical to proteins encoded in the Down's syndrome critical region 1 (DSCR1) gene (ZAK1-4, DSCRIL1, and DSCRIL). The expression of these proteins is upregulated by a calcineurin/NFAT-dependent mechanism. RCANs inhibit both calcineurin expression and activity. These proteins can interact with calcineurin through multiple interaction sites. Two sites

are located in exon 7 of RCAN, a PxlIT motif similar to the PxlIT motif found in NFAT and an ELHA motif. Another highly conserved serine–proline motif (FLSPP) is similar to sequences found in calcineurin substrates. The interaction of RCANs with calcineurin does not require, but is facilitated by, the presence of Ca^{2+} and CaM. Phosphorylation of RCANs by GSK3 prevents their inhibitory activity suggesting that *in vivo*, depending on their phosphorylation state, RCANs could act as activators as well as inhibitors of calcineurin.

A 240-kDa nuclear protein, Cain/Cabin1, has been identified as a noncompetitive inhibitor of calcineurin. Binding of Cabin1 to calcineurin requires Ca^{2+} and protein kinase C (PKC) activation and is inhibited by FK506/FKBP, suggesting that it binds at the drug-interacting site. A basic domain in the C-terminus of Cabin1 has been identified as the calcineurin-binding site. A third group of proteins, including the PKA scaffold protein, AKAP79 in brain, and calsarcins in skeletal and cardiac muscles may serve to anchor calcineurin to its site of action.

Functions

Regulation of Gene Expression

The elucidation of the Ca^{2+} /calcineurin/NFAT transduction pathway leading to T-cell activation opened the way to uncover the important role of calcineurin in the regulation of gene expression in many other cell types. As illustrated in [Figure 2](#), in T cells, the release of Ca^{2+} from the inositol triphosphate (IP_3)-sensitive stores, following occupancy of the T-cell receptor, activates calcineurin which dephosphorylates the transcription factor NFAT. Dephosphorylation of NFAT is accompanied by the exposure of a nuclear localization signal and translocation of NFAT to the nucleus. Calcineurin, anchored to NFAT through an anchoring domain (PxlIT), is translocated to the nucleus with NFAT. Dephosphorylation of NFAT results not only in its nuclear translocation but also in enhancement of DNA binding and transcriptional activity that is also dependent on the activation of the Ras/PKC pathway. Nuclear export depends on NFAT rephosphorylation upon removal of Ca^{2+} and calcineurin inactivation. Several kinases, including glycogen synthase kinase -3 (GSK3) and casein kinase 1, have been implicated in this process.

Four NFAT isoforms (NFAT1–4) are expressed in many tissues. These isoforms share a highly conserved N-terminal regulatory domain, composed of two calcineurin-binding motifs, multiple phosphorylation sites, and a nuclear localization signal. The two calcineurin-binding motifs bind calcineurin with an affinity in the micro-molar range. In skeletal muscle, calcineurin-mediated dephosphorylation of the transcription factors, NFAT-3 and the myogenic enhancing factor (MEF2), play a crucial role in the switch from fast to slow oxidative fibers induced by sustained tonic contraction. In the heart, the calcineurin-mediated activation of NFAT3 and MEF2D, in conjunction with the cardiac-specific transcription factor GATA4, has been implicated in the development of cardiac hypertrophy in response to stress. Transgenic mice lacking the β -isoform of calcineurin (the predominant isoform in heart) have impaired ability to develop cardiac hypertrophy in response to hypertrophic agonists while severe hypertrophy

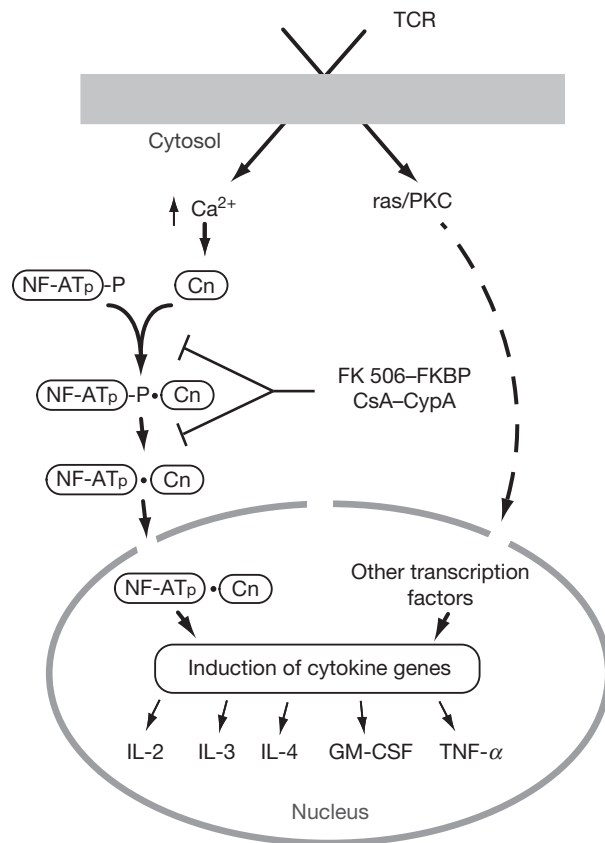


Figure 2 Calcineurin-mediated signal transduction from the T-cell receptor to the nucleus leading to the induction of the cytokine genes (IL-2, -3, -4), granulocyte-macrophage stimulating factor (GM-CSF), and tumor necrosis factor (TNF- α). Gene activation requires the concomitant activation of the ras/PKC pathway. Reproduced from Klee C, Wang X, and Ren H (1999) Calcium-regulated protein dephosphorylation. In: Carafoli E and Klee C (eds.) *Calcium as a Cellular Regulator*, pp. 344–370. New York: Oxford University Press.

and cardiac heart failure is induced by overexpression of the constitutively active calcineurin. The calcineurin/NFAT pathway is essential for the development of heart valves and the vascular developmental pattern during embryogenesis. In the brain, Ca^{2+} homeostasis is itself regulated by calcineurin. Activation of NFAT increases the expression of the IP_3 receptors, the plasma membrane Ca^{2+} pumps, and the Ca^{2+} exchanger. The increasing number of genes whose expression is induced by NFAT indicates a potential role of calcineurin in the regulation of gene expression during cell differentiation and apoptosis.

The importance of calcineurin in the regulation of these cellular processes and its involvement in the pathogenesis of many diseases based on overexpression and knockout models is well documented. To fully assess the contribution of calcineurin in the transduction of so many diverse signals, we must understand how different pathways interact with each other. In budding yeast, the calcineurin-mediated dephosphorylation of a transcription factor (Crz1p/Tcn1p) plays a major role in the regulation of genes involved in ion homeostasis. The mechanism is similar to NFAT-mediated transcriptional control in

mammalian cells. In yeast, genetic analysis adds strong support to the physiological significance of this calcineurin signaling pathway in response to exposure to abnormal concentrations of ions or prolonged exposure to mating pheromones. It not only illustrates the important role of calcineurin in the regulation of gene expression, but also reveals the participation of many other factors in this complex signaling pathway.

Neuronal Functions

Calcineurin is 1% of the total protein and broadly distributed in brain. There are few neuronal functions which are not modulated by calcineurin. A major role of calcineurin in brain is to trigger a protein phosphatase cascade initiated by the dephosphorylation of two other major substrates of calcineurin, the endogenous inhibitors of protein phosphatase-1, Inhibitor-1, and dopamine and cAMP-regulated neuronal phosphoprotein of 32 kDa (DARPP32). This cascade counteracts the stimulatory effects of cAMP and Ca^{2+} -regulated kinases. It explains the antagonistic effects of Ca^{2+} release induced by glutamate binding to the NMDA receptor and dopamine binding to some dopamine receptors in striatal neurons. Whether it activates protein phosphatase-1 or dephosphorylates an increasing number of specific substrates, calcineurin has been implicated in the regulation of neuronal processes as diverse as the expression and activity of ion channels, the synthesis of nitric oxide, the release of neurotransmitters, synaptic vesicle recycling, and neurite outgrowth. Calcineurin inhibitors and transgenic animals lacking the calcineurin genes or overexpressing the constitutive form of calcineurin have been widely used to study the involvement of calcineurin in the storage of long and short memory. How calcineurin and other protein kinases and phosphatases affect learning and memory remains one of the most challenging problems in neurobiology.

See also: **Bioenergetics:** Calcium/Calmodulin-Dependent Protein Kinase II; **Signaling:** Calcium/Calmodulin-Dependent Protein Kinases; Protein Kinase C Family.

Further Reading

- Aramburu J, Rao A, and Klee CB (2000) Calcineurin: From structure to function. *Current Topics in Cellular Regulation* 36: 237–295.
- Bito H, Deisseroth K, and Tsien RW (1996) CREB phosphorylation and dephosphorylation: A Ca^{2+} - and stimulus duration-dependent switch for hippocampal gene expression. *Cell* 87: 1203–1214.
- Crabtree GR and Schreiber SL (2009) SnapShot: Ca^{2+} -calcineurin-NFAT signaling. *Cell* 138: 210.
- Cyert MS (2001) Genetic analysis of calmodulin and its targets in *Saccharomyces cerevisiae*. *Annual Review Genetics* 35: 647–672.
- Davies KJ, Ermak G, Rothermel BA, et al. (2007) Renaming the DSCR1/Adapt78 gene family as RCAN: Regulators of calcineurin. *The Federation of American Societies for Experimental Biology Journal* 21: 3023–3028.
- Fienberg AA, Hiroi N, Mermelstein PG, et al. (1998) DARPP-32: Regulator of the efficacy of dopaminergic neurotransmission. *Science* 281: 838–842.
- Ghosh MC, Wang X, Li S, and Klee CB (2003) Regulation of calcineurin by oxidative stress. In: Klumpp S and Kriegstein J (eds.) *Protein Phosphatases. Methods in Enzymology* 366: pp. 289–304. Academic Press.
- Griffith JP, Kim JL, Kim EE, et al. (1995) X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex. *Cell* 82: 507–522.

- Hashimoto K and Kudla J (2011) Calcium decoding mechanisms in plants. *Biochimie* (2011), 93(12): 2054–2059, <http://dx.doi.org/10.1016/j.biochi.2011.05.019>.
- Heineke L and Molkentin JD (2006) Regulation of cardiac hypertrophy by intracellular signaling pathways. *Nature Review Molecular Cell Biology* 7: 589–600.
- Kao SC, Wu H, Xie J, et al. (2009) Calcineurin/NFAT signaling is required for neuregulin-regulated Schwann cell differentiation. *Science* 323: 651–654.
- Kissinger CR, Parge HE, Knighton DR, et al. (1995) Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex. *Nature* 378: 641–644.
- Li H, Zhang L, Rao A, et al. (2007) Structure of calcineurin in complex with PVIVIT peptide: Portrait of a low-affinity signalling interaction. *Journal of Molecular Biology* 369: 1296–1306.
- Malleret G, Haditsch U, Genoux D, et al. (2001) Inducible and reversible enhancement of learning, memory, and long-term potentiation by genetic inhibition of calcineurin. *Cell* 104: 675–686.
- Oh-hora M and Rao A (2008) Calcium signaling in lymphocytes. *Current Opinion in Immunology* 20: 250–258.
- Oliveria SF, Dell'Acqua ML, and Sather WA (2007) AKAP79/150 anchoring of calcineurin controls neuronal L-type Ca^{2+} channel activity and nuclear signaling. *Neuron* 55: 261–275.
- Rodriguez A, Roy J, Martinez-Martinez S, et al. (2009) A conserved docking surface on calcineurin mediates interaction with substrates and immunosuppressants. *Molecular Cell* 33: 616–626.
- Takeuchi K, Roehrl MH, Sun ZY, and Wagner G (2007) Structure of the calcineurin-NFAT complex: Defining a T cell activation switch using solution NMR and crystal coordinates. *Structure* 15: 587–597.
- Wang H, Du Y, Xiang, et al. (2008) A renewed model of CNA regulation involving its C-terminal regulatory domain and CaM. *Biochemistry* 47: 4461–4468.
- Williams AH, Liu N, van Rooij E, and Olson EN (2009) MicroRNA control of muscle development and disease. *Current Opinion in Cell Biology* 21: 461–469.

Calcium Signaling: NO Synthase

Z-Q Wang, Kent State University Tuscarawas, New Philadelphia, OH, USA

D J Stuehr, The Lerner Research Institute, Cleveland, OH, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Calmodulin Calcium-binding protein that binds to target enzymes in a Ca^{2+} -dependent manner and activates their function.

Nitric oxide Diatomic free radical that is generated from L-arginine by the nitric oxide (NO) synthases and is widely active in numerous processes in biology.

Reductase domain Flavoprotein domain of NO synthases that contains bound flavin adenine dinucleotide and flavin mononucleotide and is responsible for transferring electrons from nicotinamide adenine dinucleotide phosphate to the heme group in NO synthase.

Nitric oxide (NO) is an important signal and effector molecule in animal physiology. NO is generated from L-arginine by the NO synthases (NOSs). Three NOSs have been characterized in animals: neuronal NOS (nNOS, type I), cytokine-inducible NOS (iNOS, type II), and endothelial NOS (eNOS, type III). Importantly, the activities of all three depend on their binding calmodulin (CaM). CaM binds reversibly to eNOS and nNOS in response to elevated Ca^{2+} concentrations and so their activities are regulated by intracellular Ca^{2+} . By contrast, iNOS binds CaM independent of Ca^{2+} concentrations and is therefore continuously active after it is expressed. Because it is unusual for a redox enzyme like NOS to be controlled by a Ca^{2+} -binding protein, the NOS–CaM interaction has provided new insight into some aspects of protein structure and function.

NO Synthases

To become active, each NOS enzyme must assemble into a dimer of two subunits. Each subunit in the NOS dimer contains an N-terminal oxygenase domain that binds iron protoporphyrin IX (heme), the cofactor 6R-tetrahydrobiopterin (H_4B), and the substrate Arg, and a C-terminal reductase domain that binds flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD), and nicotinamide adenine dinucleotide phosphate (NADPH). Between oxygenase and reductase domains, there is a 20–25-amino-acid CaM-binding motif (Figure 1).

NO synthesis depends on electron transfer to the heme, and this process only occurs when CaM is bound. In a CaM-bound NOS dimer, the electrons from NADPH load into the FAD and FMN groups in one subunit, and then transfer from the FMN to the heme that is located in the partner subunit (Figure 1). In general, CaM activates electron transfer in NOS at two points: electron transfer into the flavins (intradomain) and between the FMN and heme (interdomain). Intradomain transfer is associated with enhanced catalysis by the NOS reductase domain, as measured by rates of NADPH-dependent ferricyanide or cytochrome *c* (Cyt *c*) reduction. CaM activation of the interdomain

electron transfer in NOS results in reduction of the ferric heme iron, which is a prerequisite for oxygen binding to the heme and for NO synthesis.

Calmodulin Interactions with Ca^{2+} and with NOS

CaM is a small Ca^{2+} -binding protein (molecular weight 17 kDa) consisting of two similar globular lobes connected by a central linker region. Each lobe contains two E–F hands which include an N-terminal helix (E helix), a centrally located Ca^{2+} coordinating loop, and a C-terminal helix (F helix) (domains of CaM are generally labeled as 1–4 from N to C terminal). Upon Ca^{2+} binding, CaM is able to bind to and activate more than 30 target enzymes, enabling it to regulate numerous second messengers and cell functions. Binding and full activation of target proteins by CaM typically requires occupancy of all four Ca^{2+} -binding sites in CaM. The carboxy-terminal lobe contains two high-affinity Ca^{2+} -binding sites, while the amino-terminal lobe contains two sites with lower Ca^{2+} affinity. Conformational changes occur in CaM after it binds Ca^{2+} : hydrophobic surface residues become exposed in order to form crucial van der Waals interactions with the hydrophobic face of the target peptide recognition site.

CaM binds to each NOS subunit in a 1:1 stoichiometry with reasonably high affinity. Binding affinity studies with peptides that correspond to NOS CaM recognition sequences show that the three NOSs display different affinities toward CaM with the general order being cytokine-inducible NOS (iNOS) $\gg$ neuronal NOS (nNOS) $\approx$ endothelial NOS (eNOS). The iNOS peptide binds well to CaM both in the presence and absence of Ca^{2+} , but eNOS and nNOS peptides bind to the CaM only in the presence of Ca^{2+} .

Each lobe of CaM binds to NOS independently and displays different affinities toward NOS. The CaM–nNOS interaction has been investigated using CaM mutants and plant CaM proteins. nNOS displays a high degree of structural specificity toward CaM domains 1, 3, and possibly 4 regarding activation of heme reduction and NO synthesis. The CaM–nNOS interactions are particularly important for electron transfer and involve the latch region of CaM (formed by CaM domains 1 and 3) and a methionine

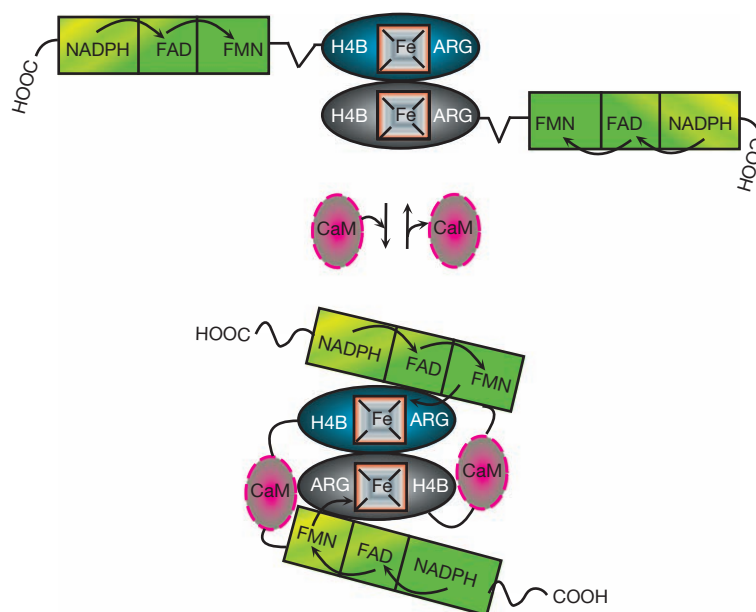


Figure 1 CaM-induced electron transfer in nNOS. CaM binding in response to elevated Ca^{2+} concentrations enables electron transfer between FMN and the heme located in adjacent subunits of the NOS dimer.

residue in domain 4. These regions of CaM appear to interact with as yet unknown regions on nNOS that are distinct from its canonical CaM-binding sequence. The charge and length of the central linker are also found to affect the binding and activation of iNOS through electrostatic interaction in the central linker and hydrophobic interactions in the globular domain of CaM, which seems responsible for its tight association to the iNOS. Ca^{2+} dissociation from NOS-bound CaM occurs in two sequential steps: rapid Ca^{2+} dissociation from the N-terminal lobe occurs first and corresponds with inactivation of NO synthesis, followed by a slower Ca^{2+} dissociation from the C-terminal lobe, which leads to the dissociation of CaM from NOS. Conformational changes occur in the NOS CaM-binding peptides when they bind CaM. Typically, the peptides acquire an α -helical conformation upon interaction with Ca^{2+} bound CaM. A CaM-induced conformational change in the NOS reductase domain also occurs as indicated by changes in protein and flavin fluorescence and by a change in the trypsin proteolysis pattern.

A crystallographic structure of Ca^{2+} -loaded CaM bound to a 20-residue peptide corresponding to the eNOS CaM recognition sequence is available. The structure revealed that the α -helical eNOS peptide binds to CaM in an antiparallel orientation through extensive hydrophobic interactions: the N-terminal and C-terminal lobes of CaM wrap around the bound peptide, interacting with the C- and N-terminal halves of the peptide, respectively. Specific CaM residues in the latch region and domain 4 were positioned in a manner consistent with their importance as determined by mutagenesis studies. The crystal structure also suggested a basis for tighter CaM binding in iNOS: because iNOS contains a greater number of hydrophobic residues within its corresponding CaM recognition sequence, it was argued that these would support more extensive van der Waals contacts with CaM and minimize unfavorable solvent exposure of hydrophobic residues in order to favor tighter association between CaM and the iNOS CaM recognition sequence.

The crystal structure of a complex between Ca^{2+} /CaM and the FMN domain of human iNOS is also available. Four different conformations were identified in the structure of the complex, which demonstrates the flexible nature of CaM and FMN domain-domain interactions. A salt bridge formed between CaM and the FMN domain is apparent in the structure and may be important for transducing the effect of CaM binding on NOS functions.

Ca^{2+} /CaM Regulation of NOS

How Ca^{2+} /CaM regulates NO synthesis and heme reduction in NOS is not entirely clear. CaM binding does not appear to change the thermodynamic driving force for electron transfer in NOS and instead may control the electron transfer primarily via structural rearrangement. CaM binding likely induces a relatively large conformational rearrangement which then enables the FMN subdomain to get close enough to the NOS oxygenase domain heme to make electron transfer between them more efficient.

Some cellular proteins that bind to NOS can regulate its response to Ca^{2+} /CaM. These include caveolins, dynamin, the bradykinin β -2 receptor, and heat shock protein-90. In addition, several different structural elements that are present in NOS regulate its interaction and response to Ca^{2+} /CaM. These include the canonical CaM-binding site in each NOS and at least four other structural elements within the reductase domain (Table 1, Figure 2).

Initially, it was imagined that the CaM interaction with each NOS isoform might depend solely on its CaM recognition sequence. However, when the CaM-binding sequences of both eNOS and nNOS were replaced with the corresponding sequence from iNOS, enzyme still required added Ca^{2+} for full CaM binding and activity. Other deletion studies have

Table 1 NOS structural elements involved in its Ca^{2+} /CaM response

<i>NOS structural element</i>	<i>Role</i>
CaM binding site	Binds CaM to NOS; is inhibitory in the absence of CaM
Autoinhibitory loop	Impacts Ca^{2+} concentration response, represses NOS activity in the absence of CaM, also required for full CaM response
C-terminal extension	Impacts Ca^{2+} concentration response, represses NOS activities in the absence of CaM, also required for full CaM response
Amino acid K842 in the autoinhibitory loop of nNOS	Impacts Ca^{2+} concentration response, helps to repress NOS activities in the absence of CaM, also stabilizes closed conformation of NOS reductase domain
R1400 in C-terminal tail of nNOS	Helps to repress NOS activities in the absence of CaM through ionic interaction with 2-phosphate group of bound NADP(H).
Amino acid 484–726 in the FMN subdomain of iNOS	Required for Ca^{2+} -independent CaM-binding to iNOS
Acidic amino acid (E762, E789, E790, D813, E816, E819) on the surface of FMN subdomain of nNOS	Helps to repress NOS activities in the absence of CaM, also required for CaM to relieve the repression
Amino acid R1229 in the FAD subdomain of nNOS	Helps to repress NOS activities in the absence of CaM, also required for CaM to fully relieve the repression
CD2 loop in reductase domain of eNOS	Impacts Ca^{2+} concentration response, represses NOS activities in the absence of CaM
Phe1395 in FNR subdomain of nNOS	Helps to repress NOS activities in absence of CaM, also required for CaM to fully relieve the repression
Phosphorylation sites in eNOS S1179	Phosphorylation impacts Ca^{2+} concentration response and increases enzyme activity
T497	Dephosphorylation increases enzyme activity
S635	Phosphorylation may impact enzyme activity
S617	Phosphorylation impacts Ca^{2+} concentration response, no change in enzyme activity

identified short regions within the iNOS and eNOS reductase domains that modulate the NOS Ca^{2+} requirement and CaM affinity in a positive or negative manner (1). Finally, there are several amino acids that can undergo phosphorylation within the NOS reductase domains and CaM-binding sites (Figure 2). Phosphorylation at these individual or combined sites in NOS alters the CaM response with regard to its Ca^{2+} concentration

requirement and may also alter the maximal level of enzyme activation that is achieved.

Autoinhibitory Loop

One regulatory element that is unique to NOS is the autoinhibitory loop insert (Figure 2). Constitutive NOSs (e.g., eNOS and nNOS) contain a 40–50-amino-acid insert in their FMN-binding module that is not shared with iNOS or with other related flavoproteins. Similar autoinhibitory loops have been reported in a number of CaM-dependent enzymes, including CaM-dependent protein kinase II, smooth muscle myosin light chain kinase, and calcineurin. Peptides based on the eNOS insert sequence inhibited NOS enzymatic activity by altering their CaM binding. For instance, deletion of the insert in eNOS or nNOS decreased their EC_{50} for Ca^{2+} , confirming that the insert was inhibitory (i.e., causes the enzyme to require a higher Ca^{2+} concentration to enable its CaM response). Deleting the autoinhibitory peptide also enables some NO synthesis in the absence of CaM and increases the intrinsic catalytic activity of CaM-bound eNOS. Thus, the insert contributes to the Ca^{2+} dependence of the constitutive NOS and also regulates the electron transfer between its FMN and heme centers.

How the autoinhibitory insert participates in Ca^{2+} /CaM regulation is still unclear. One model has the insert docking on a site in NOSs that impedes both CaM binding and the structural rearrangement that is expected to be required for enzymatic activation. Upon binding, the Ca^{2+} /CaM, a conformational change that displaces the autoinhibitory insert from its docking site would then relieve the inhibitory effect and facilitate activation of the enzyme. A two-stage model seems reasonable to explain the Ca^{2+} /CaM activation in NOSs: CaM binding is largely determined by the recognition sequence and some other interactions with the enzyme. However, activation of the enzyme is fully dependent on a movement of the autoinhibitory insert, which could only be affected by the Ca^{2+} -saturated CaM. The conformational change induced by Ca^{2+} /CaM is passed on to the autoinhibitory insert, and this in turn presumably enables the enzyme reductase and oxygenase domains to adopt conformations that lead to productive electron transfer between the FMN and heme redox centers.

C-Terminal Extension

Another regulatory element that is unique to NOS is the C-terminal extension. The protein sequences of NOS reductase domains closely resemble that of NADPH-cytochrome P450 reductase (CPR). However, all NOS isoforms have an additional 20–40 residue extension in their C terminus, forming a tail that is absent in CPR. Deletion of 33 or 42 C-terminal residues from nNOS or eNOS speeds electron transfer into the flavins, implying that the C-terminal extension is a negative regulator. Another NOS mutant with a partial deletion of the C-terminal extension ($\Delta 27$ eNOS) exhibited a lower Ca^{2+} concentration response and an increase in some of its catalytic activities. Importantly, there is an Akt-dependent phosphorylation site located in the C-terminal tail of the constitutive NOS enzymes that modulates their regulation by Ca^{2+} /CaM (Figure 2). Phosphorylation at this site relieves some of the repression attributed to the C-terminal tail in the absence of CaM. It also lowers the Ca^{2+}

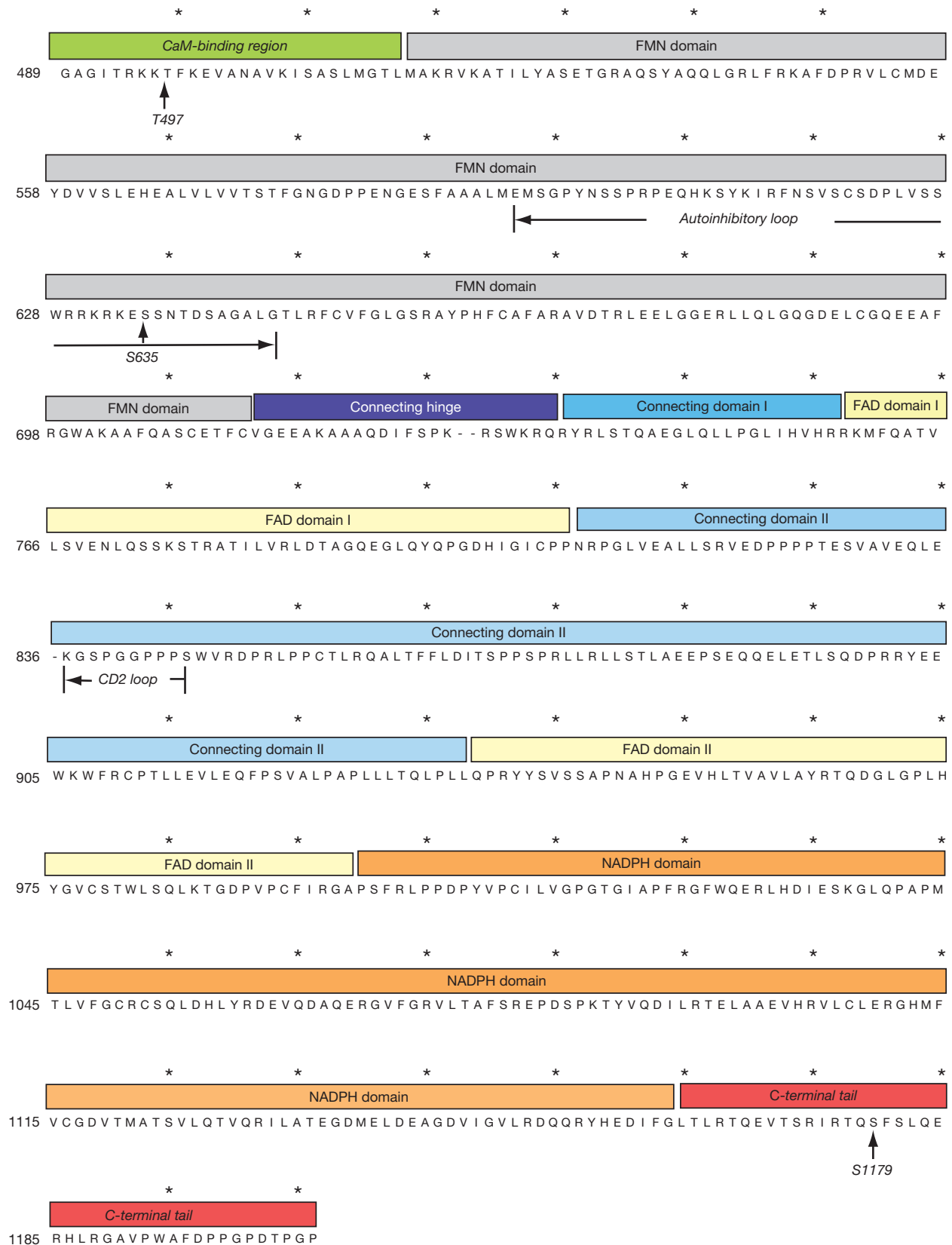


Figure 2 Sequence of eNOS–CaM binding site and reductase domain. Colored rectangles indicate subdomains that make up the reductase domain. Structural elements and phosphorylation sites that impact Ca^{2+} /CaM regulation of NOS are noted in italics.

concentration that is required for CaM binding and increases the NO synthesis activity of eNOS (but not of nNOS). When the autoinhibitory loop is present (as in nNOS and eNOS), the C-terminal tail is also needed to enable full activation in response to Ca^{2+} /CaM. This indicates that it functions as a positive regulatory element under this circumstance.

Ca^{2+} /CaM regulation of NOS may involve the above two regulatory elements in the following manner. In the absence of CaM, the C-terminal extension helps to cover the NADPH-FAD and FMN subdomain interfaces, and this inhibits (but does not completely prevent) the FMN subdomain from transferring electrons to redox partners like Cyt *c* or the NOS heme. This repression by the C-terminal tail requires that the enzyme NADPH binding site be occupied. Under this condition, the autoinhibitory insert is positioned in a way that antagonizes CaM binding and prevents the FMN subdomain from interacting productively with the NOS oxygenase domain. Upon CaM binding, the autoinhibitory loop swings away and this conformational change enables interactions between the FMN subdomain and the oxygenase domain that are productive for electron transfer to the heme. Specific interactions between the C-terminal extension and the autoinhibitory insert are likely to occur within the CaM-bound NOS. Such interactions between the two regulatory elements are likely to enable full activation of electron transfer by Ca^{2+} /CaM. Thus, the autoinhibitory loop and C-terminal tail repress enzyme function in the absence of CaM, and help to positively modulate function when Ca^{2+} /CaM is bound. Other structural elements that are present in the NOS reductase domain (1) along with protein phosphorylation events also help to regulate Ca^{2+} /CaM binding and to tune its activation of enzyme catalysis.

See also: **Bioenergetics:** Calcium/Calmodulin-Dependent Protein Kinase II; Cytochrome *c*; Heme Synthesis; Ligand-Operated Membrane Channels: Calcium (Glutamate); **Protein/Enzyme Structure**

Function and Degradation: Flavins; Heme Proteins; Pteridines; **Signaling:** Nitric Oxide Signaling.

Further Reading

- Abu-Soud HM and Stuehr DJ (1993) Nitric oxide synthases reveal a role for calmodulin in controlling electron transfer. *Proceedings of the National Academy of Sciences of the United States of America* 90: 10769–10772.
- Aoyagi M, Arvai AS, Tainer JA, and Getzoff ED (2003) Structural basis for endothelial nitric oxide synthase binding to calmodulin. *EMBO Journal* 22: 766–775.
- Feng C and Tollin G (2009) Regulation of interdomain electron transfer in the NOS output state for NO production. *Dalton Transactions* 6692–6700.
- Hemmings B and Mayer B (1998) Enzymology of nitric oxide synthases. *Methods in Molecular Biology* 100: 1–32.
- James P, Vorherr T, and Carafoli E (1995) Calmodulin-binding domains: Just two faced or multi-faceted? *Trends in Biochemical Sciences* 20: 38–42.
- Li H and Poulos TL (2005) Structure–function studies on nitric oxide synthases. *Journal of Inorganic Biochemistry* 99: 293–305.
- Nishida CR, Knudsen G, Straub W, and Ortiz de Montellano PR (2002) Electron supply and catalytic oxidation of nitrogen by cytochrome P450 and nitric oxide synthase. *Drug Metabolism Reviews* 34: 479–501.
- Persechini A, White HD, and Gansz KJ (1996) Different mechanisms for Ca^{2+} dissociation from complexes of calmodulin with nitric oxide synthase or myosin light chain kinase. *Journal of Biological Chemistry* 271: 62–67.
- Roman LJ, Martasek P, and Masters BS (2002) Intrinsic and extrinsic modulation of nitric oxide synthase activity. *Chemical Reviews* 102: 1179–1190.
- Roman LJ and Masters BS (2006) Electron transfer by neuronal nitric-oxide synthase is regulated by concerted interaction of calmodulin and two intrinsic regulatory elements. *Journal of Biological Chemistry* 281: 23111–23118.
- Spratt DE, Israel OK, Taiakina V, and Guillemette JG (2008) Regulation of mammalian nitric oxide synthases by electrostatic interactions in the linker region of calmodulin. *Biochimica et Biophysica Acta* 1784: 2065–2070.
- Stuehr DJ, Tejero J, and Haque MM (2009) Structural and mechanistic aspects of flavoproteins: Electron transfer through the nitric oxide synthase flavoprotein domain. *FEBS Journal* 276: 3959–3974.
- Welland A, Garnaud PE, Kitamura M, Miles CS, and Daff S (2008) Importance of the domain–domain interface to the catalytic action of the NO synthase reductase domain. *Biochemistry* 47: 9771–9780.
- Xia C, Misra I, Iyanagi T, and Kim JJ (2009) Regulation of interdomain interactions by calmodulin in inducible nitric-oxide synthase. *Journal of Biological Chemistry* 284: 30708–30717.

Calcium Transport in Mitochondria

R Rizzuto and M Brini, University of Padua, Padua, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Adenosine triphosphate (ATP) The biochemical molecule utilized by the cells as energy source.

Apoptosis A type of cell death, characterized by specific morphological changes (i.e., chromatin condensation, DNA damage, and shrinkage of the cells) and utilized by organisms to regulate the growth and the development of tissues, organs, etc. Also called 'programmed cell death'.

Caspases Protein family responsible for degradation of specific proteins that contain recognizing sequences.

Their activation leads to signals of programmed cell death.

G protein Protein associated to specific hormone receptor of the plasma membrane and responsible for the conversion of extracellular signals into intracellular messages.

Respiratory chain A series of multisubunit proteins located in the inner mitochondrial membrane responsible for the electron and proton transfer necessary to generate the electrochemical gradient used for ATP production.

The Origins and the Fundamental Principles

Studies of Ca^{2+} transport in isolated mitochondria began in the 1960s when evidence was provided that isolated, respiring mitochondria accumulate Ca^{2+} in the presence of inorganic phosphate. Ca^{2+} uptake was inhibited by respiratory chain blockers but not by the adenosine triphosphate (ATP) synthesis inhibitor oligomycin, and, during Ca^{2+} uptake, no adenosine diphosphate (ADP) phosphorylation took place. In the presence of ATP, no respiratory chain was necessary. Thus, the process of Ca^{2+} uptake was regarded as an alternative pathway to ADP phosphorylation for the use of the respiratory chain energy. In this scenario mitochondria were thought as large stores of Ca^{2+} , which could be accumulated together with inorganic phosphate and precipitate in the matrix, leaving its free Ca^{2+} concentration virtually unchanged.

The general picture became clearer when the chemiosmotic theory solved the mechanism allowing mitochondria to store the energy obtained from metabolite breakdown and couple it to ATP synthesis. Indeed, the concept that the respiratory chain, by pumping protons across the ion-impermeable inner membrane, establishes a gradient ($\Delta\mu_{\text{H}^+}$), composed of an electrical potential ($\Delta\psi$) and a concentration ratio (ΔpH), according to the Nernst equation, $\Delta\mu_{\text{H}^+} = zF\Delta\psi + RT \ln[\text{H}^+]_{\text{i}}/[\text{H}^+]_{\text{o}}$, has major implications for Ca^{2+} transport and distribution. In actively respiring mitochondria, considering the buffering capacity mostly provided by weak acids, the gradient is supposed to be maintained mostly in the form of electrical gradient across the inner membrane (~ 180 mV). This implies a strong thermodynamic force in favor of the accumulation of cations (in the case of a divalent cation, such as Ca^{2+} , thermodynamic equilibrium is attained when in the matrix the concentration is approximately 10^6 higher than in the intermembrane space). Based on these simple considerations, large attention was drawn in the 1960s and 1970s on the capacity of mitochondria to accumulate Ca^{2+} , and on the biochemical and thermodynamic properties of this transport. Through the contribution of numerous laboratories and a large body of experimental work carried out in isolated, respiring mitochondria, these organelles were shown to possess separate

accumulation and release pathways for the cation, with defined characteristics. Accumulation was shown to depend on the activity of an electrogenic uniporter that transports Ca^{2+} down the electrical gradient established by the respiratory chain. Inhibition of the respiratory chain or collapse of the electrical gradient (e.g., by the use of a protonophore) abolishes the capacity of mitochondria to accumulate Ca^{2+} . Ca^{2+} uptake is also directly inhibited by the compound Ruthenium Red (or its recently identified subcomponent Ru360) or lanthanides. Unfortunately, none of the inhibitors is highly specific; thus, no selective tool has aided the purification of the transporter. Through the years, many attempts have been made to identify the uniporter and some putative candidates have been proposed but, only very recently, this key transporter has been truly identified. As to the release pathways, two sets of exchangers have been shown to extrude Ca^{2+} from mitochondria: a $\text{Na}^+/\text{Ca}^{2+}$ exchanger, mostly expressed in excitable cells (muscle and brain), and a $\text{H}^+/\text{Ca}^{2+}$ exchanger that represents the prevailing route in most other tissues. Both exchangers (albeit with different K_{d}) are inhibited by verapamil, diltiazem, and other Ca^{2+} channel blockers. Very recent evidence indicates that $\text{Na}^+/\text{Ca}^{2+}\text{-Li}^+$ exchanger (NCLX), the single mammalian member of the phylogenetically ancestral Ca^{2+} /cation exchanger family, initially characterized as a novel plasma membrane $\text{Na}^+/\text{Ca}^{2+}$ exchanger that catalyzes Na^+ - or Li^+ -dependent Ca^{2+} transport, is the long elusive mitochondrial $\text{Na}^+/\text{Ca}^{2+}$ exchanger.

More recently, much attention has been drawn to the potential role of a large-conductance channel, commonly referred to the permeability transition pore (PTP). It is supposed to be a multiprotein complex (the putatively essential components being the voltage-dependent anion channel (VDAC) of the outer membrane, the adenine nucleotide transporter of the inner membrane and cyclophilin D) activated by various pathophysiological conditions (e.g., Ca^{2+} increases in the mitochondrial matrix and oxidation of critical cysteines). As to the (patho)physiological role, it is most likely involved in the swelling and fragmentation of the mitochondrial network that underlies the release of caspase cofactors from mitochondria. Its molecular identity is still unknown.

The Middle Ages

In the 1980s, mitochondrial participation in intracellular Ca^{2+} homeostasis rapidly lost support, as the focus in Ca^{2+} signaling was largely diverted to the endoplasmic reticulum (ER). In these years, the chain of events that couples the stimulation of plasma membrane receptors to the induction of a rise in cytoplasmic Ca^{2+} concentration was clarified, and it became clear that the ER, and not mitochondria, were the main site of action. Indeed, stimulation of receptors coupled, through a G_q protein, to the activation of phospholipase C induces the hydrolysis of the plasma membrane phosphatidylinositol-4,5-bisphosphate into diacylglycerol and inositol-1,4,5-trisphosphate (IP3). IP3 diffuses into the cytosol and interacts with Ca^{2+} -permeant channels of the ER membrane, causing their opening and the release of Ca^{2+} into the cytoplasm. The ER (and its specialized version of muscle cells, the sarcoplasmic reticulum) was shown to be endowed with a molecular repertoire allowing it to act as a highly reactive intracellular Ca^{2+} store: pumps (that accumulate Ca^{2+} at the expense of ATP hydrolysis and maintain a standing gradient between the lumen and the cytosol), low affinity Ca^{2+} -binding proteins (that increase the net amount of cation that can be accumulated, but promptly make it available when the concentration in the lumen decreases), and channels (opening, with different mechanisms, after stimulation of plasma membrane receptors).

What was left for mitochondria? In principle, given the large thermodynamic force for Ca^{2+} accumulation, they could take up most of the Ca^{2+} released by the ER. In the same years, however, the availability of powerful and versatile Ca^{2+} indicators, the intracellularly trappable fluorescent dyes (e.g., quin2 and fura-2), allowed us to carefully estimate the cytosolic Ca^{2+} concentration ($[\text{Ca}^{2+}]_c$) of living cells and further reduced the possible roles of mitochondria. Indeed, these indicators showed that both at rest (0.1 μM) and upon stimulation (1–3 μM), the $[\text{Ca}^{2+}]_c$ is well below the affinity of the mitochondrial uniporter, and thus very little Ca^{2+} is supposed to be accumulated in mitochondria during the brief pulse of a physiological stimulation. Thus, the general consensus incorporating this series of information was that mitochondria acted as a low-affinity sink that could accumulate large Ca^{2+} loads, but only if the cell was exposed to challenges (most likely pathological events) that induced sustained, large $[\text{Ca}^{2+}]_c$ increases. In other words, mitochondria were mostly intended as salvage mechanisms against deleterious Ca^{2+} overloads.

The Renaissance

In this situation, the only opportunity for mitochondria to obtain credit in the Ca^{2+} field was to directly demonstrate in intact, living cells changes in mitochondrial $[\text{Ca}^{2+}]$ occurring in physiological conditions. This became possible in the early 1990s, when novel experimental tools allowed us to specifically measure the Ca^{2+} concentration in the mitochondrial matrix ($[\text{Ca}^{2+}]_m$): positively charged fluorescent indicators, that are largely accumulated in the mitochondria and molecularly engineered chimeras of the Ca^{2+} -sensitive photoprotein

aequorin, that include mitochondrial targeting sequences and are thus exclusively localized in the mitochondrial matrix. Using these probes, it was possible to demonstrate that, in a variety of cell systems (ranging from epithelial cells to skeletal and cardiac myocytes, from hepatocytes to neurons), the $[\text{Ca}^{2+}]_c$ rises evoked by physiological stimulations are always paralleled by rapid $[\text{Ca}^{2+}]_m$ increases that reach values well above those of the bulk cytosol (up to $\sim 500 \mu\text{M}$ in chromaffin cells). The obvious discrepancy between this prompt response and the low affinity of the Ca^{2+} uniporter was reconciled by the demonstration that mitochondria are exposed to microdomains of high $[\text{Ca}^{2+}]$ that largely exceed the values reported in the bulk cytosol and meet the low affinity of the uniporter. This is achieved through a close interaction between the mitochondria and the ER, the intracellular Ca^{2+} store, which could be directly demonstrated using targeted chimeras of a fluorescent recombinant protein (green fluorescent protein (GFP)) and a high-resolution imaging system. A consequence of this morphological arrangement is the capacity of mitochondria to sense the microenvironment at the mouth of the IP3-sensitive channel (and/or the plasma membrane Ca^{2+} channels), and thus the high $[\text{Ca}^{2+}]_c$ generated by their opening upon cell stimulation (Figure 1).

One (or Many) Roles?

What is the functional significance of the re-evaluated mitochondrial Ca^{2+} transport? An obvious answer again stems from biochemical work by Denton, McCormack, and Hansford in the 1960s, who could demonstrate that three key metabolic enzymes (the pyruvate, α -ketoglutarate, and isocitrate dehydrogenases) are activated by Ca^{2+} , by different mechanisms: in the case of pyruvate dehydrogenase through a Ca^{2+} -dependent dephosphorylation step, in the latter two cases through the direct binding of Ca^{2+} to the enzyme complex. Thus, the extension of a Ca^{2+} signal originated in the cytoplasm to the mitochondria would serve the purpose of transmitting an activatory signal to the energy powerhouse of the cell; thus, in conjunction to the triggering of energy-consuming processes in the cytosol (contraction, secretion, etc.) mitochondrial dehydrogenases are stimulated, thus adapting aerobic metabolism to the increased needs of an active cell. Using the luciferase-based probe for ATP, it was possible to demonstrate that a rise in mitochondrial ATP levels parallels the $[\text{Ca}^{2+}]_m$ increase evoked by cell stimulation and strictly depends on the $[\text{Ca}^{2+}]_m$ rise; if the latter is prevented by the use of Ca^{2+} chelators, such as 1,2-bis(O-aminophenoxy)ethane-*N,N,N',N'*-tetraacetic acid (BAPTA), also the [ATP] rise does not occur.

However, recent work by different groups has clarified that the role of mitochondrial Ca^{2+} uptake is not limited to the control of organelle function, but has a direct impact on the Ca^{2+} signals evoked by agonist stimulation in the cytosol. Two different mechanisms concur in this effect. The first occurs in the microdomains where mitochondria and ER get in close contact. Here, the efficiency of mitochondrial Ca^{2+} accumulation accounts for the rapid clearing of the high $[\text{Ca}^{2+}]$ at the mouth of the release channel of the ER and thus reduces the (positive or negative) feedback effect of the cation on the channel itself. Such a mechanism has been shown in a

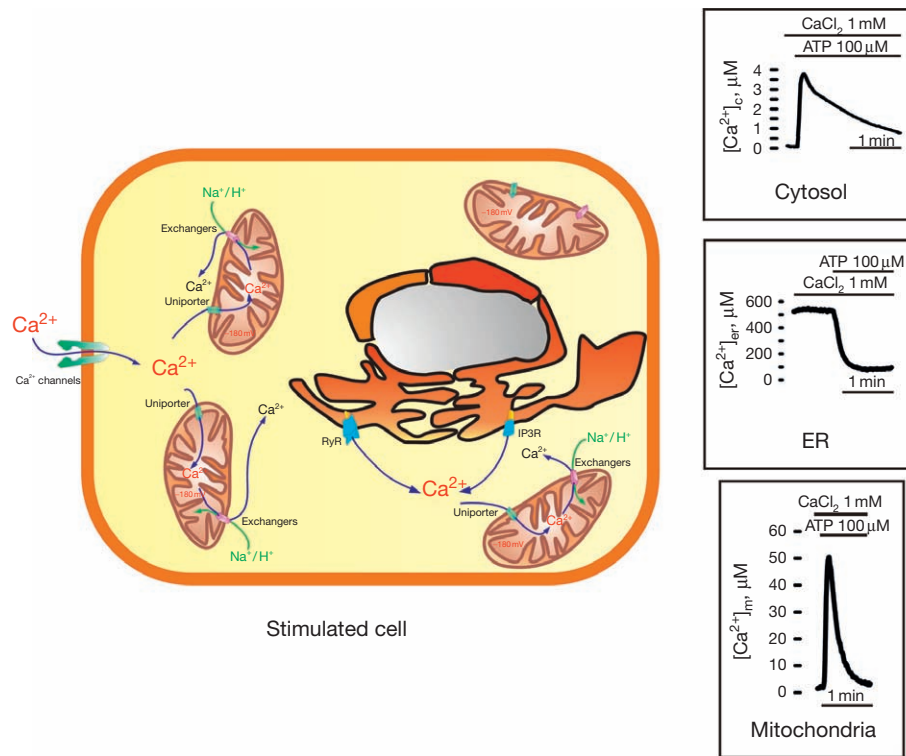


Figure 1 Mitochondrial Ca^{2+} transport. Cell stimulation induces the opening of plasma membrane Ca^{2+} channels and ER Ca^{2+} channels (IP3R and RyR), which, in turn, generates hot spots of Ca^{2+} concentration. Mitochondria located in the Ca^{2+} channels proximity could rapidly take up Ca^{2+} since their low affinity uniporter could experience high local Ca^{2+} concentrations. The three traces represent measurements of $[\text{Ca}^{2+}]$ in the three cellular compartments shown, performed with suitably targeted aequorin.

variety of experimental systems ranging from *Xenopus* oocytes (in which the diffusion properties of Ca^{2+} waves correlate with the energization state of mitochondria) to mammalian cells such as hepatocytes and glial cells, where it was demonstrated that the kinetics of Ca^{2+} release from the ER, and thus the spatio-temporal properties of the $[\text{Ca}^{2+}]_c$ rise are influenced by the process of mitochondrial Ca^{2+} uptake.

The second mechanism by which mitochondrial Ca^{2+} uptake affects cytosolic Ca^{2+} signals has been demonstrated in pancreatic acinar cells that are endowed with a defined polarized morphology and the occurrence of cellular Ca^{2+} signals with different cellular locations and physiological consequences. Namely, at lower doses of agonist (e.g., colecystokinin), the $[\text{Ca}^{2+}]_c$ rise is restricted to the apical pole and causes granule secretion; at higher doses, the $[\text{Ca}^{2+}]_c$ rise extends to the basal pole (where the nucleus is located), thus causing activation of gene expression and long-term alterations of cell function. The restriction mechanism was shown to be dependent on Ca^{2+} uptake by mitochondria clustered below the apical region that act as a firewall preventing the spreading of the $[\text{Ca}^{2+}]_c$ rise. When the mitochondrial belt is overwhelmed (e.g., upon intense stimulation, or when the Ca^{2+} uptake capacity of mitochondria is experimentally impaired), then the Ca^{2+} signal can freely diffuse to the rest of the cell.

Finally, in the past years a more dangerous role for mitochondrial Ca^{2+} uptake has emerged. Work from various labs has revealed that the alteration of the Ca^{2+} signal reaching the

mitochondria and/or the combined action of apoptotic agents or pathological conditions (e.g., oxidative stress) can induce a profound alteration of organelle structure and function. As a consequence, proteins normally retained in the organelle (such as an important component of the respiratory chain, cytochrome *c*, as well as newly discovered proteins, such as apoptosis-inducing factor and Smac/Diablo) are released into the cytoplasm, where they activate effector caspases and drive cells to apoptotic cell death. Different effects have been described: in hepatocytes, upon treatment with suboptimal doses of the lipid mediator of apoptosis ceramide, the repetitive spiking of cytoplasmic Ca^{2+} concentration ($[\text{Ca}^{2+}]_c$) rather than triggering the activation of matrix dehydrogenases (and thus the stimulation of aerobic metabolism) causes the opening of the PTP, the swelling of mitochondria, and the release of cytochrome *c*. Thus, a mechanism of coincidence detection of physiological agonists and apoptotic stimuli allows the specific induction of the apoptotic program (the mitochondria acting as the site where this differential decoding is operated). In addition, apoptotic stimuli can induce the impairment of the Ca^{2+} -clearing mechanism of the cell (the plasma membrane Ca^{2+} -ATPase), thus potentiating Ca^{2+} loading. In this scenario, it can be easily understood how the partial depletion of Ca^{2+} from the ER, induced by the oncogene Bcl-2, has a protective effect toward some apoptotic stimuli (those involving mitochondria in the signaling pathway). Both in neurons and in hepatic cells, it was observed that apoptotic signals cause mitochondrial Ca^{2+} overload and thus the apoptotic

Physiological conditions

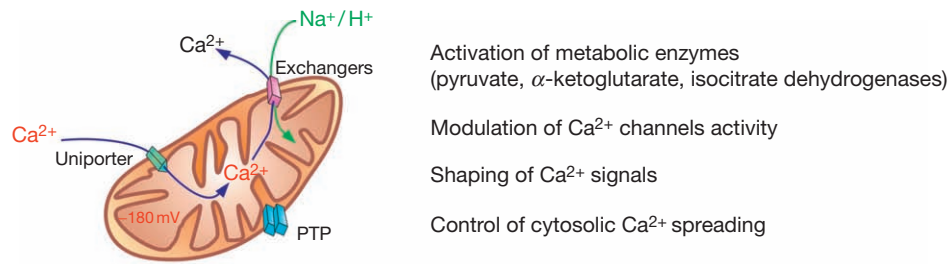
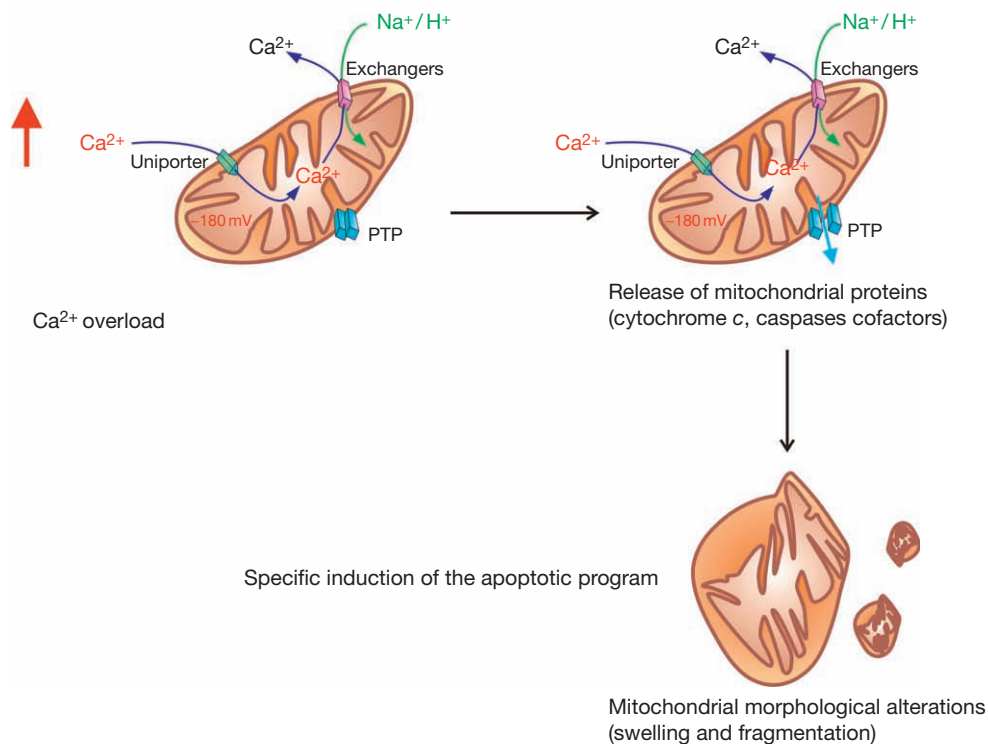
Apoptotic agents and/or alteration of the Ca²⁺ signal

Figure 2 Physiological and pathological role of mitochondrial Ca²⁺ signaling.

morphological and functional alterations. A general scheme of the physiological role of mitochondrial Ca²⁺ signaling, and the potential alterations occurring in pathological conditions (modifying the final outcome) is summarized in [Figure 2](#).

The Molecular Machinery

The Outer Membrane Channels

Mitochondrial Ca²⁺ effectors are mostly located within the matrix, and Ca²⁺ elevation in the latter compartment requires the crossing of the outer membrane (OMM), known to be highly permeable to ions and small solutes, and the ion-impermeable inner membrane. While great interest was placed

on the identification of the transporters of the inner membrane, the elucidation of the permeation properties and the correlation with organelle structure and function, recent data have however also pointed to an important role of the fine-tuning of the outer membrane permeability. Indeed, the repertoire of VDAC channels clustered at the ER/mitochondrial contact sites appears limiting for the Ca²⁺ uptake capacity of the organelle. Overexpression of VDAC channels causes a significant increase in the peak [Ca²⁺] values reached upon cell stimulation and, in high-speed single cell imaging experiments, reduces the delay between the cytosolic and mitochondrial uptake, thereby indicating that VDAC channels play a key role in the rapid transfer of the high Ca²⁺ microdomain from the surface of mitochondria to the intermembrane space to which the low-affinity mitochondrial

Ca^{2+} uniporter (MCU) is exposed. VDAC shuttles from an open and closed state (with reference to metabolite transport) and the closed state shows a higher permeability to Ca^{2+} . Interestingly, Ca^{2+} itself appears to control the conductance of VDAC, thus suggesting that the OMM barrier is dynamically reduced, and mitochondrial Ca^{2+} uptake facilitated, during the Ca^{2+} signal.

The Inner Membrane Transporters

The ion-impermeable inner membrane (IMM) is obviously the key site of mitochondrial Ca^{2+} transport, and the identification of the IMM Ca^{2+} transporters has been a long-pursued aim. In particular, major effort has been placed by numerous labs in giving a molecular identity to the MCU, the electrophoretic system that has the properties of a highly selective ion channel and allows Ca^{2+} to be accumulated down the electrical gradient established by the respiratory chain. The effort proved for a long time ineffective, because simpler model systems, such as yeast, do not have MCU activity, no specific inhibitor could be employed for protein purification, and homology search with known Ca^{2+} channels did not yield convincing candidates. Thus, although the general properties of mitochondrial Ca^{2+} handling have been known for nearly half a century, no mitochondrial Ca^{2+} transporters have been identified and validated for a long time.

In the past year, however, the molecular search of Ca^{2+} transporters has benefited from powerful genomic or proteomic approaches, yielding interesting candidates that require careful consideration. Clapham and co-workers carried out an impressive genome-wide RNA interference screening with a mitochondrially targeted pericam in *Drosophila* S2 cells, obtaining an interesting list of proteins affecting mitochondrial Ca^{2+} accumulation. They focused their attention on LETM1, previously reported to be a mitochondrial K^+/H^+ exchanger, and showed that its silencing depressed mitochondrial Ca^{2+} accumulation triggered by cell stimulation, while leaving the $\Delta\psi_m$ unaffected. Based on their data, the authors propose that LETM1 is a $\text{Na}^+/\text{Ca}^{2+}$ antiporter mediating Ca^{2+} accumulation into the matrix. These conclusions, however, remain controversial, as some key data appear in contrast with established evidence: (1) antiporter inhibition by CGP37157 augments, and not reduces, Ca^{2+} accumulation; (2) in contrast to the MCU, antiporters are known to be insensitive to Ruthenium Red; and (3) the accepted stoichiometry of the exchanger is $1\text{Ca}^{2+}/2\text{H}^+$ (or possibly $1\text{Ca}^{2+}/3\text{H}^+$), thus driving Ca^{2+} efflux, rather than influx, in energized mitochondria. Further work will allow us to confirm, or exclude, the direct role of LETM1 in mitochondria Ca^{2+} handling.

Another interesting approach has been pursued by Mootha and co-workers, who made the precious effort of generating a mitochondrial genome (MitoCarta), that is, a compendium of nuclear genes coding for proteins with mitochondrial localization validated by proteomic analysis and GFP tagging. By screening this list for expected properties of MCU (e.g., ubiquitous expression and absence in *Saccharomyces cerevisiae*), they initially identified a protein, which they named 'mitochondrial calcium uptake 1' (MICU1). MICU1 is a protein associated with the IMM, including two EF-hand Ca^{2+} -binding sites and a single putative transmembrane stretch. Its silencing abolishes mitochondrial Ca^{2+} uptake in intact and permeabilized cells, without

interfering with mitochondrial respiration or membrane potential. Moreover, MICU1 is required for the metabolic coupling between cytosolic Ca^{2+} transients and activation of matrix dehydrogenases. While it is unlikely that MICU1 is the bona fide channel, the data strongly argued for a key regulatory role and suggested that mitochondrial Ca^{2+} uptake depends on a macromolecular complex. Few months later, the judicious analysis of Mitocarta by two independent groups allowed to identify an integral inner membrane protein satisfying the criteria for being the MCU. The protein, denominated MCU, is ubiquitously expressed in mammals and has two predicted transmembrane regions connected through a loop enriched in acidic residues. Its small interfering RNA silencing inhibited mitochondrial Ca^{2+} accumulation in intact cells, whereas its overexpression dramatically increased it. The channel activity of purified MCU reconstituted in a planar lipid bilayer revealed the properties previously reported for the uniporter, thus conclusively demonstrating that MCU *per se* represents the bona fide calcium channel of the inner mitochondrial membrane. In addition, mutation of the putative pore region not only inhibited channel activity, but the mutant channel also acted as dominant negative in intact cells, thus suggesting that ion permeation requires the presence of multiple subunits in an oligomer.

Finally, molecular information has recently been obtained also on the efflux pathways, that is, the mitochondrial exchangers (the $\text{Na}^+/\text{Ca}^{2+}$ exchanger, mNCX, and $\text{H}^+/\text{Ca}^{2+}$ exchanger, mHCX) that re-extrude the cation out of the matrix, preventing the attainment of thermodynamic equilibrium and allowing the return of $[\text{Ca}^{2+}]_m$ to basal values after cell stimulation. Sekler and coworkers identified NCLX as the only mammalian member of a phylogenetically ancient branch of the $\text{Na}^+/\text{Ca}^{2+}$ exchanger superfamily, and proved it to be enriched in mitochondrial fractions and transport both Na^+ and Li^+ , as expected of the mNCX. Overexpression and silencing enhanced and reduced mitochondrial Ca^{2+} efflux in intact cells, and Ca^{2+} transport was inhibited by CGP37157. Overall, the data are fully consistent with the notion that the identified protein is the bona fide mitochondrial $\text{Na}^+/\text{Ca}^{2+}$ exchanger, opening this branch of the Ca^{2+} transport machinery to detailed molecular investigation.

Conclusions

Work in the past years has greatly re-evaluated mitochondrial Ca^{2+} homeostasis not only as a key event for the maintenance of ion balance within the organelle (avoiding deleterious short-circuiting of the chemiosmotic machinery), but also as a control mechanism for physiological and pathological processes occurring in mitochondria, or activated by mitochondrial factors. Thus, both in basic and applied research there is an expanding interest on this topic, that now benefits of the very recent identification of the key molecules involved (the NCX, MICU1, a uniporter regulator and MCU, the uniporter itself). The molecular machinery is thus essentially solved: now exciting times will come, in which proteomic analyses, transgenic models, and pharmacological studies will allow to unravel the function of mitochondrial Ca^{2+} homeostasis in normal and pathological conditions.

See also: **Bioenergetics: Mitochondrial Calcium Transport: Historical Aspects; Mitochondrial Membranes, Structural Organization; Mitochondrial Metabolite Carrier Protein Family; Mitochondrial Permeability Transition Pore.**

Further Reading

- Baughman JM, Perocchi F, Girgis HS, et al. (2011) Integrative genomics identifies MCU as an essential component of the mitochondrial calcium uniporter. *Nature* 476: 341–345.
- Carafoli E (2010) The fateful encounter of mitochondria with calcium: How did it happen? *Biochimica et Biophysica Acta* 1797: 595–606.
- De Stefani D, Raffaello A, Teardo E, Szabò I, and Rizzuto R (2011) A forty-kilodalton protein of the inner membrane is the mitochondrial calcium uniporter. *Nature* 476: 336–340.
- Duchen MR, Verkhratsky A, and Muallem S (2008) Mitochondria and calcium in health and disease. *Cell Calcium* 44: 1–5.
- Hajnóczky G and Csordás G (2010) Calcium signaling: Fishing out molecules of mitochondrial calcium transport. *Current Biology* 20: R888–891.
- Jiang D, Zhao L, and Clapham DE (2009) Genome-wide RNAi screen identifies Letm1 as a mitochondrial $\text{Ca}^{2+}/\text{H}^{+}$ antiporter. *Science* 326: 144–147.
- Kroemer G, Galluzzi L, and Brenner C (2007) Mitochondrial membrane permeabilization in cell death. *Physiological Reviews* 87: 99–163.
- MacCormack JG, Halestrap AP, and Denton RM (1990) Role of calcium ions in regulation of mammalian intramitochondrial metabolism. *Physiological Reviews* 70: 391–425.
- Palty R, Silverman WF, Hershfinkel M, et al. (2010) NCLX is an essential component of mitochondrial $\text{Na}^{+}/\text{Ca}^{2+}$ exchange. *Proceedings of the National Academy of Sciences of the United States of America* 107: 436–441.
- Perocchi F, Gohil VM, Girgis HS, et al. (2010) MICU1 encodes a mitochondrial EF hand protein required for Ca^{2+} uptake. *Nature* 467: 291–296.
- Ricchelli F, Sileikytė J, and Bernardi P (2011) Shedding light on the mitochondrial permeability transition. *Biochimica et Biophysica Acta* 1807: 482–490.
- Rizzuto R, Brini M, Murgia M, and Pozzan T (1993) Microdomains of high Ca^{2+} close to inositol-triphosphate sensitive channels are sensed by neighbouring mitochondria. *Science* 262: 744–747.
- Rizzuto R and Pozzan T (2006) Microdomains of intracellular Ca^{2+} : Molecular determinants and functional consequences. *Physiological Reviews* 86: 369–408.

Calpain

H Sorimachi, S Hata, and Y Ono, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Hiroyuki Sorimachi and Yasuko Ono, volume 1, pp. 300–306, © 2004, Elsevier Inc.

Glossary

C2 domain Common Ca^{2+} -binding structure composed of eight antiparallel β -strand structures (β sandwich structure).

Calpain homepage Calpain-related information (sequence, commercially available reagents) are summarized on the calpain homepages listed in the section Relevant Websites.

Cysteine protease A peptide bond hydrolyzing enzyme whose active site is composed of a catalytically active cysteine residue.

EF-hand motif Common Ca^{2+} -binding motif composed of two α -helices (E- and F-helices) and a Ca^{2+} -binding loop in between them.

Limb-girdle muscular dystrophy (LGMD) Heterogeneous sets of muscular dystrophies mainly affecting proximal muscles. Thus far, close to 20 responsible genes have been identified. Among them, LGMD type 2A (LGMD2A) is caused by mutations in muscle-specific calpain, p94/calpain 3, and also called ‘calpainopathy’.

Lissencephaly Lissencephaly literally means smooth brain. A rare brain formation disorder caused by defective neuronal migration during the 12th to 24th weeks of gestation, resulting in a lack of development of brain folds and grooves. Children with lissencephaly are severely neurologically impaired and often die within several months of birth.

Musculardystrophy Progressive deterioration of muscle tissue and resultant weakness caused by a defect of the number of muscle genes such as dystrophin, sarcoglycan, merosin, laminin, and calpain. The information on pathogenic mutations of muscular dystrophies is available in the Leiden Muscular Dystrophy pages.

Noninsulin-dependent diabetes mellitus (NIDDM) Also called diabetes mellitus type 2, type 2 diabetes, or adult-onset diabetes. A disorder that is characterized by high blood glucose in the context of insulin resistance and relative insulin deficiency.

Introduction

Calpains denote a unique branch of cysteine proteases, which have amino acid sequences significantly similar to that of the protease domain of human μ -calpain, the major Ca^{2+} -dependent protease, and constitute a superfamily (Clan CA, family C02, EC 3.4.22.17). Among them, most studied calpains are μ -calpain and m-calpain, which are called ‘conventional calpains’. Mammalian conventional calpains are mainly localized in the cytosol, show ubiquitous expression, and display limited proteolytic activity at neutral pH. The mode of its action is processing, rather than digesting; calpain proteolyzes substrates at one or very limited number of sites to transform and modulate their structures and activities. Thus, calpain is viewed as a representative intracellular modulator protease that governs various cellular functions such as signal transduction and cell morphogenesis.

Calpain belongs to the papain superfamily that is composed of the three distinct kingdoms, that is, calpain, papain, and bleomycin-hydrolase subfamilies. The human genome has 15 genes that encode a calpain-like protease domain (see [Figure 1](#) and [Table 1](#)). These generate diverse types of calpain homologs possessing combinations of several functional domains such as a Ca^{2+} -binding domain (C2-type and penta-EF-hand (PEF)-type) and a Zn-finger domain. Furthermore, the progress in genome projects revealed that calpain homologs exist in various living organisms including insect, nematode, trypanosome, plant, fungus, yeast, and even some bacteria, thus constituting a superfamily, even in an evolutionary aspect, possessing versatile functions ([Figure 1](#)). The importance of the physiological roles

of calpains is reflected by the fact that particular defects in calpain functionality cause a variety of deficiencies in many different organisms. These include muscular dystrophies, lissencephaly and tumorigenesis in humans, embryonic lethality in mice, impaired neurogenesis in flies, lack of definitive sex determination in nematodes, defects in aleurone cell development in maize, and alkaline/osmotic stress susceptibility in fungi and yeasts. In the following sections, with several disease-related topics of calpains, divergent physiological roles of calpains are described mainly from the viewpoint of structure–function relationship.

History and Nomenclature

An enzyme corresponding to calpain was first described in 1964 by Gordon Guroff in the *Journal of Biological Chemistry*. After several reidentifications, calpain, which was called ‘calcium-activated neutral protease (CANP)’ at that time, was finally molecularly cloned in 1984 by Koichi Suzuki’s group. The complementary DNA (cDNA) sequence for the catalytic subunit of calpain revealed that calpain is a chimeric molecule consisting of a cysteine protease and a calmodulin-like Ca^{2+} -binding module. Both names, calpain and CANP, were unified as calpain in 1991 in the annual meeting of the International Committee on Proteolysis held in Munich. In the two decades following this event, hundreds of calpain-related molecules, including its endogenous specific inhibitor protein, calpastatin, have been identified through the use of cDNA cloning and genome/expressed sequence tag (EST) projects ([Figures 1 and 2](#)).

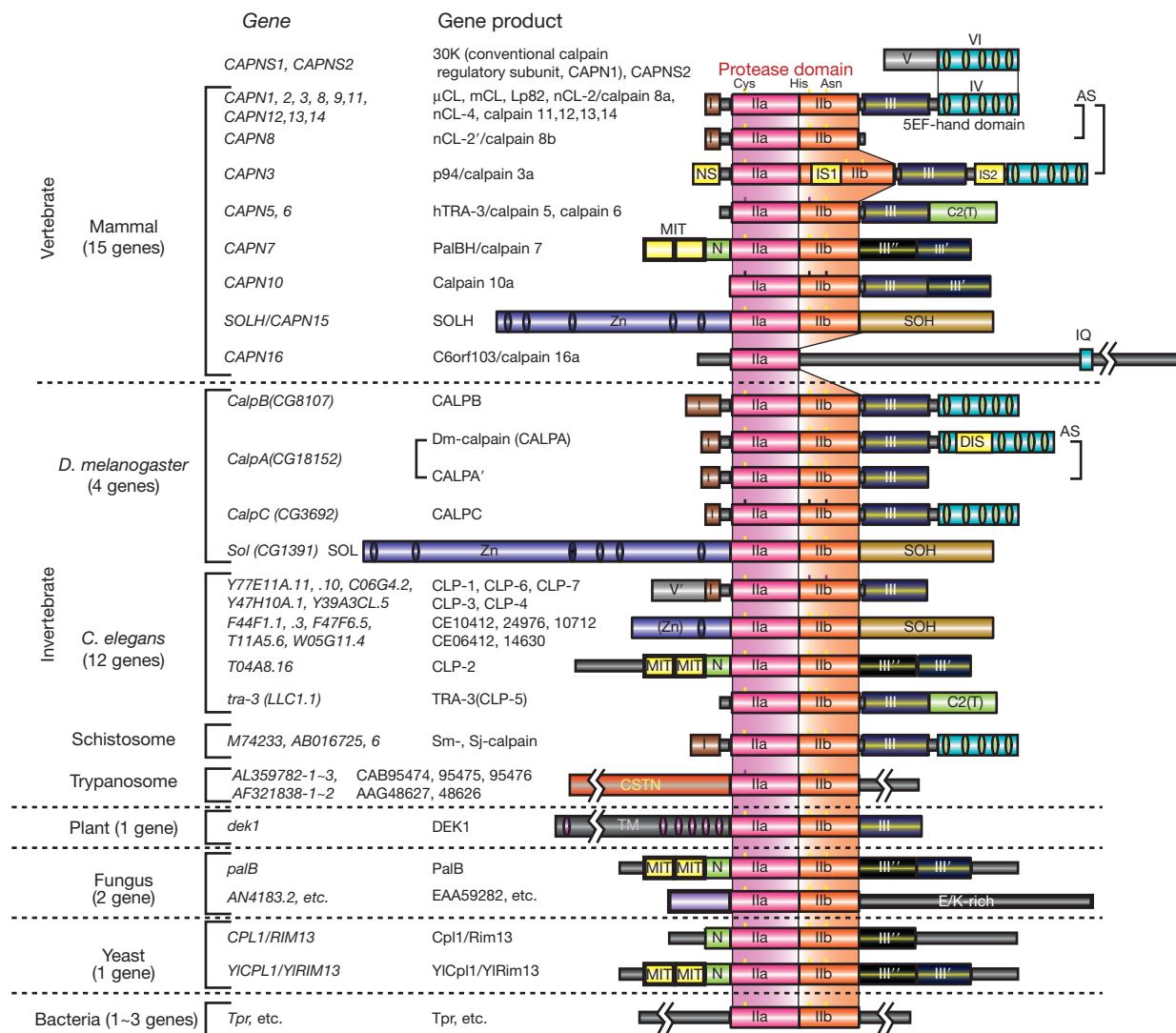


Figure 1 Schematic structures of calpain superfamily members. Calpain homologs have been identified in vertebrates, insects, nematodes, schistosomes, trypanosomes, plants, fungi, yeasts, some bacteria, etc. Symbols used are: I, the N-terminal domain with little homology; IIa and IIb, the protease sub-domains containing the active sites Cys and His, respectively; III, the C2-like Ca^{2+} -binding domain; III' and III'', domains moderately and very weakly similar to the domain III, respectively; IV and VI, the 5-EF-hand Ca^{2+} -binding domain; V, Gly-rich hydrophobic domain; NS, IS1, and IS2, p94-specific sequences; T, TRA-3 subfamily-specific C2 domain; MIT, Microtubule-interacting and trafficking domain; Zn, Zn-finger motif containing domain; SOH, SOL subfamily homology domain; DIS, CALPA-specific insertion sequence; TM, transmembrane domain; CSTN, the domain weakly similar to calpastatin.

Two major ubiquitous calpains are found in mammals, μ - and m-calpains. As most of the calpain studies have been actually the studies about μ - and m-calpains, they are referred to as the conventional calpains. As the names suggest, μ -calpain and m-calpain are activated by μM and mM levels of Ca^{2+} *in vitro*, respectively. They are both hetero-dimers and consist of a common calpain small regulatory subunit (CAPNS1, also called '30 K'; *c.* 28 kDa) and large distinct μ - and m-calpain catalytic subunit (μCL and mCL , respectively; *c.* 80 kDa), which have $\sim 60\%$ amino acid identity. Fifteen human calpain homologs have been numbered according to the gene name as shown in Table 1 and Figure 1. As an example, μCL is now called 'calpain 1' or CAPN1, that is, μ -calpain is a hetero-dimer consisting of a CAPNS1 regulatory subunit and calpain 1

catalytic subunit. To avoid unnecessary confusions, however, the original name and the gene product name are adopted in this article, for example, μCL /CAPN1, mCL /CAPN2, p94/CAPN3, 30 K/CAPNS1, hTRA-3/CAPN5, etc.

Structure and Function of Conventional Calpains

In general, substrates for calpains are limited and specific. Most short oligopeptides investigated are poor substrates for calpains. Although casein is not an *in vivo* substrate of calpain, it is a very good *in vitro* substrate and is used to assay calpain activity. Although the complete rules governing substrate specificity remain unclear, some preferences on cleavage site

Table 1 Human calpain-related genes

Gene	Chr.	Name in number	Originally described names(s) of the gene product	Other name(s) used	Protease activity ^a	Domains ^b			Expression	Phenotype(s) of gene deficiency ^c	Note
						C2-like	C2	PEF			
The catalytic subunits											
CAPN1	11q13	Calpain 1	μCANP ^d /calpain-I large subunit	μ-Calpain large subunit (μCL), μ80K	+	+	—	+	Ubiquitous	Platelet dysfunction ~ n.s.	+30 K/CAPNS1 = μ-calpain
CAPN2	1q41-q42	Calpain 2	mCANP/calpain-II large subunit	m-Calpain large subunit (mCL), m80K	+	+	—	+	Ubiquitous except for mammalian erythrocytes	Embryonic lethal	+30 K/CAPNS1 = m-calpain
CAPN3	15q15.1-q21.1	Calpain 3a (not in human)	p94 Lp82, Lp85, etc.	Calpain 3, nCL-1 —	+	+	—	+	Skeletal muscle Lens, retina	Muscular dystrophy	Termination codon in the human first exon
		Calpain 3b, 3c, . . . , 3n	p94:Δex15, Tp36, Mp18, Up84 etc.	—	n.d.	+	—	+	Ubiquitous		Ups, Tps, and Mps are driven from alternative promoter
CAPN5	11q14	Calpain 5	Calpain 5, hTRA-3	nCL-3	+	+	+	—	Testis, brain	Sudden death ~ n.s.	Nematode TRA-3 homolog
CAPN6	Xq23	Calpain 6	Calpain 6	Calpamodulin, CANPX	—	+	+	—	Placenta, embryonic muscles	n.d.	Nematode TRA-3 homolog, but no Cys at the active site
CAPN7	3p24	Calpain 7	Calpain 7, PalBH	—	n.d.	++	—	—	Ubiquitous	n.d.	Aspergillus PalB homolog
CAPN8	1q41	Calpain 8a	nCL-2	—	+	+	—	+	Stomach	n.d.	
		Calpain 8b	nCL-2'	—	+	—	—	—	Stomach		Ca ²⁺ -dependent
		Calpain 8c	nCL-2:Δ Dex10,17~fs	—	+	+	—	+/- —	Stomach		
CAPN9	1q42.11-q42.3	Calpain 9a	nCL-4	—	+	+	—	+	Digestive tracts	n.d.	30 K is required for activity
		Calpain 9b	nCL-4:Δex8	—	+	+	—	+	Digestive tracts		
CAPN10	2q37.3	Calpain 10a	Calpain 10a	—	n.d.	++	—	—	Ubiquitous	n.s.	SNP is related to NIDDM
		Calpain 10b, 10c, . . . , 10 h	Calpain 10b, 10c, . . . , 10 h	—	n.d.	+ or —	—	—			
CAPN11	6p12	Calpain 11	Calpain 11	—	n.d.	+	—	—	Testis	n.d.	

(Continued)

Table 1 (Continued)

Gene	Chr.	Name in number	Originally described names(s) of the gene product	Other name(s) used	Protease activity ^a	Domains ^b			Expression	Phenotype(s) of gene deficiency ^c	Note
						C2-like	C2	PEF			
CAPN12	19q13.2	Calpain 12a	Calpain 12a	—	n.d.	+	—	+	Hair follicle	n.d.	
		Calpain 12b, 12c	Calpain 12b, 12c	—	n.d.	+	—	+/-			
CAPN13	2p22-p21	Calpain 13	Calpain 13	—	n.d.	+	—	+	Ubiquitous	n.d.	mRNA not detected
CAPN14	2p23.1-p21	Calpain 14	Calpain 14	—	n.d.	+	—	+	n.d.	n.d.	
SOLH/ CAPN15	16p13.3	Calpain 15	SOLH	—	n.d.	—	—	—	Ubiquitous	n.d.	<i>Drosophila</i> SOL homolog
C6ORF103/ CAPN16	6q24.3	Calpain 16	Demi-calpain	—	—	—	—	—	Ubiquitous	n.d.	Only domain IIa
<i>The regulatory subunits</i>											
CAPNS1	19q13.1	CAPNS1	CANP/calpain small subunit	30 K, css1	—	—	—	+	Ubiquitous	Embryonic lethal	Regulatory subunit for μ CL and mCL
CAPNS2	16q12.2	CAPNS2	Calpain small subunit 2	30 K-2, css2	—	—	—	+	Ubiquitous	n.d.	Intron less gene, overlap with lysophosphatidylcholine acyltransferase 2 gene (LPCAT2)
<i>The calpain inhibitor</i>											
CAST	5q15	CAST	CANP inhibitor/calpastatin a~l	—	—	—	—	—	ubiquitous	Excitotoxicity in neuronal cells ~ n.s.	Specific inhibitor for μ - and m-calpains, nCL-2 and nCL-4

^a "+" indicates that it was experimentally shown to have protease activity, and "—" means that it does not have active site cysteine, asparagine, and/or histidine residues.

^b "+" or "—" indicates that it has or does not have a corresponding domain, respectively, and "+/-" indicates that it has less than five EF-hand motifs.

Human disease and results of knock-out mice were taken together.

^c n.d., not yet determined/done; n.s., no significant phenotype; CANP, Ca^{2+} -activated neutral protease; nCL, novel calpain large subunit; LGMD2A, limb-girdle muscular dystrophy type 2A; TRA-3, transform genotypic hermaphrodites; PalB, phosphatase mutants: loss in activity at alkaline pH but normal or increased activity at acidic pH; PalBH, PalB homolog; SNP, single nucleotide polymorphism; NIDDM, non-insulin-dependent diabetes mellitus; SOL, small optic lobes; SOLH, SOL homolog.

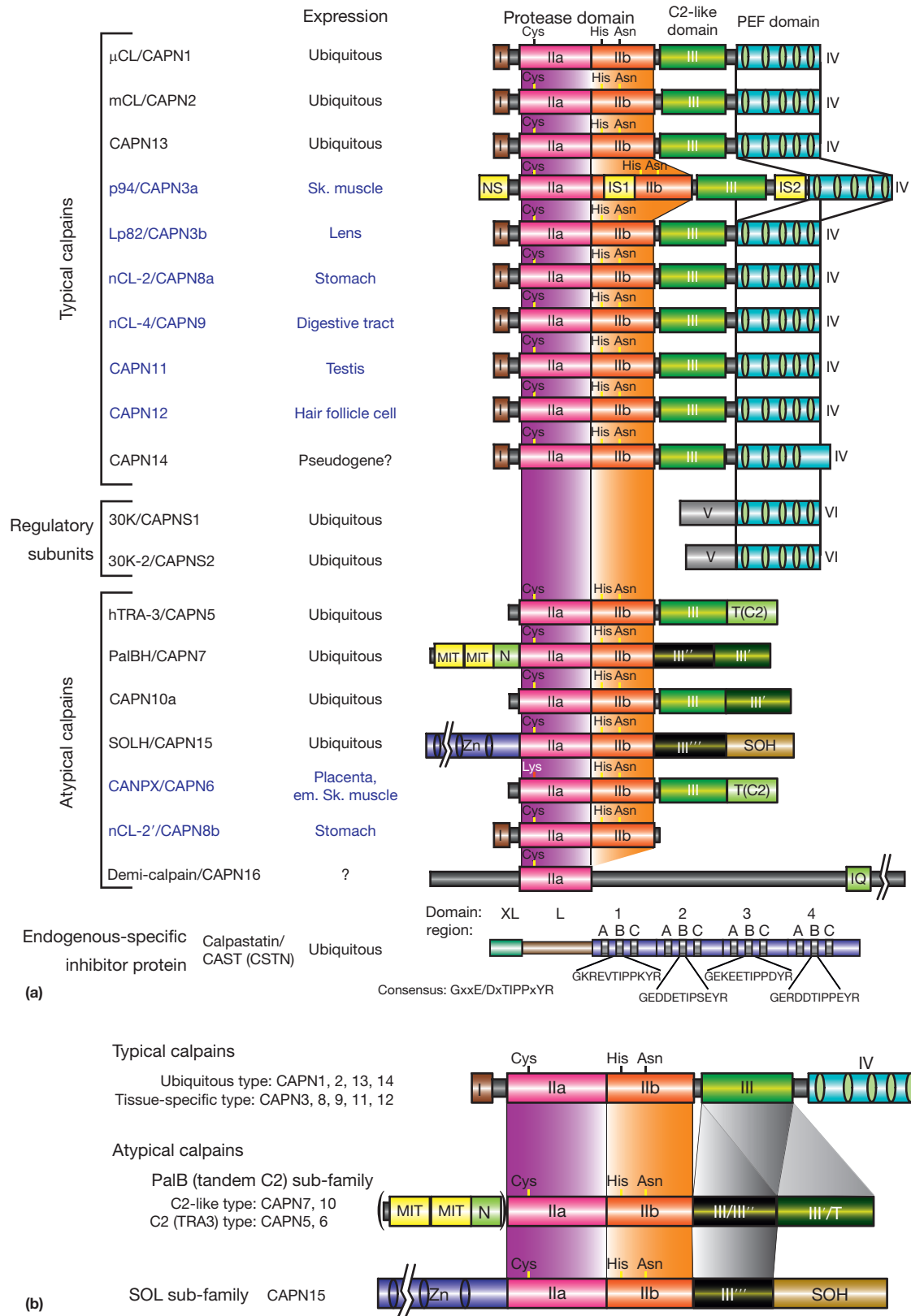


Figure 2 Schematic structures of human calpains. (a) Schematic structures of representative products of human genes that encode calpain homologs. Blue letters indicate that these calpains are expressed tissue-specifically, whereas the others in black letters are ubiquitously expressed (see also Table 1). (b) Classification of mammalian calpains by domain structures. Symbols used are: IQ, a region interactive with calmodulin; L and XL, N-terminal and extended N-terminal regions of calpastatin; MIT, a domain interactive with microtubule; see Figure 1 for other symbols.

sequences have been reported. On the other hand, calpain is thought to also recognize a large scope of three-dimensional (3D) substrate structures. For example, protein kinases, phosphatases, phospholipases, cytoskeletal proteins, membrane proteins, cytokines, transcription factors, lens proteins, and calmodulin-binding proteins are just some of the proteins suggested to be *in vivo* substrates. There has been no report suggesting differences in substrate specificity between μ - and m-calpains.

Calpain has a very specific proteinaceous inhibitor *in vivo* called calpastatin. Calpastatin has four repeats of an inhibitor unit, each of which inhibits one molecule of calpain (Figure 2 (a)). Both μ - and m-calpains possess similar susceptibility to calpastatin. Among other calpain homologs, novel calpain large subunit (nCL)-2/CAPN8 and nCL-4/CAPN9, but not p94/CAPN3, are also inhibited by calpastatin. The μ - and m-calpains are ubiquitously expressed in vertebrate cells. Thus, their function is thought to be fundamental and essential. Many functions including the regulation of signal transduction systems, cell motility, and apoptosis have been suggested, although their precise physiological roles are still elusive. *Capn1* knockout mice showed almost no phenotype, while *Capn2* and *Capns1* knockout mice resulted in embryonic lethality, indicating the indispensable roles of conventional calpains and, at the same time, differentiating between μ - and m-calpain functionality and/or amounts expressed.

As shown in Figure 2(a), the catalytic and regulatory subunits of conventional calpains can be divided into four and two domains, respectively. The N-terminus of domain I of the large subunit is autolyzed upon the initial activation by Ca^{2+} . This results in a lower requirement for Ca^{2+} , reveal of different substrate specificity, and the dissociation of both subunits. Therefore, autolysis is one of the important regulation mechanisms of calpain activity and specificity.

3D structural studies revealed that the protease domain (domain II) in the absence of Ca^{2+} is divided into two sub-domains – domains IIa and IIb (also called domains I and II), which are folded into one domain upon Ca^{2+} binding. This domain is highly conserved among calpain family members, suggesting the functional importance of this domain (Figure 1). Rather surprisingly, the protease domain without other domains of μ - and m-calpains showed Ca^{2+} -dependent protease activity. This was explained once the 3D structural studies of the protease domain in the presence of Ca^{2+} showed that Ca^{2+} bound to domains IIa and IIb. Thus, the whole calpain molecule mediates Ca^{2+} dependency because all of the domains IIa, IIb, III, IV, and VI bind at least one Ca^{2+} with varying affinities. The 3D structure of active (Ca^{2+} -bound form of) calpain in complex with calpastatin peptide is shown in Figure 3.

The 3D structure of domain III consists of eight antiparallel β -strands (β -sandwich structure), a structure very similar to tumor necrosis factor (TNF)- α and the C2-domains found in various Ca^{2+} -regulated proteins such as protein kinase C (PKCs) and synaptotagmins. Although the primary structures of domain III are highly conserved in calpain homologs, they have no similarity to any other proteins including TNF- α and C2-domains. This C2-like domain actually binds Ca^{2+} and may play an important role in the Ca^{2+} -dependent membrane translocation of calpains.

Domain IV is very similar to domain VI of the regulatory subunit and each contains five EF-hand motifs. Thus, these domains are referred to as PEF domains. *In vitro* experiments together with 3D structural studies showed that not all EF-hands bind Ca^{2+} . The fifth EF-hand motif is involved in the dimerization of the catalytic and regulatory subunits.

Domain V of the calpain regulatory subunit contains clusters of glycine residues making it hydrophobic. This domain is thought to interact with membrane and/or membrane proteins

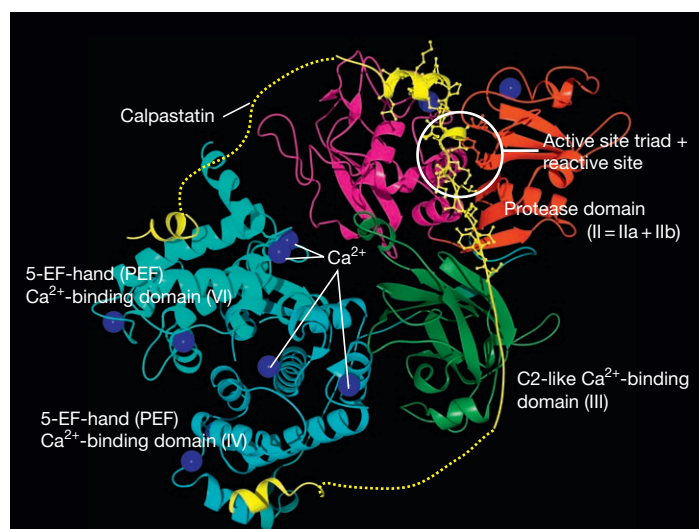


Figure 3 Schematic 3D structure of active rat m-calpain in complex with calpastatin peptide. Ribbon-type schematic 3D figures are drawn by FlatLux MolFeat Ver.4.5 using PDB data, 3DF0. Domain III and two PEF domains (domains IV and VI) are indicated by green and blue, respectively. The protease domain (domain II, i.e., sub-domains IIa+IIb) is given in pink and red color. Active site residues, Cys105, His262, and Asn286 (NP_058812), in the protease domain are indicated by ball-stick drawing and a white circle. Blue balls (visible in domains IIa, IIb, IV, and VI) represent Ca^{2+} . Calpastatin reactive site residues bound to domains IIa, IIb, and III of m-calpain are shown by ball-stick drawing. Dotted lines indicate invisible calpastatin polypeptide due to strong thermal vibration, which indicates the state unbound to m-calpain.

through hydrophobic interactions. Most of this domain is autolyzed during activation, indicating no involvement of this region in protease activity. In humans, the *CAPNS2* gene encodes a regulatory subunit homolog, whose physiological roles remain unclear.

Calpain Superfamily and Its Members

Classifications

As for calpain homologs other than conventional calpains, domains other than the protease domain are not necessarily conserved among homologs (Figures 1 and 2). Amino acid sequence identities of the protease domains vary, depending on the molecules, from less than 30% to more than 75%. Several kinds of domains, putatively originated from independent genes, exist in both N- and C-terminal sides of the protease domain. These include C2 and C2-like domains, a PEF domain, a transmembrane domain, Zn-finger domain, as well as conserved domains with unknown functions (e.g., SOL homology domain (SOH)). These features, together with the organization of mammalian calpain genes, strongly suggest that calpain molecules are the result of the combination of the gene for the ancestral calpain-type cysteine protease with genes encoding other functions.

Mammalian calpain homologs can be divided into two categories from the structural viewpoint. The first group consists of molecules having domains II, III, and IV (Figure 2). In other words, these group members have a typical structure highly similar to conventional calpain catalytic subunits. Beside mammalian μ CL/CAPN1 and mCL/CAPN2, typical calpains include p94/CAPN3, nCL-2/CAPN8, nCL-4/CAPN9, CAPN11, 12, 13, and 14. Chicken and *Xenopus laevis* are reported to additionally have μ /mCL, which may correspond to CAPN11. In invertebrates, only five typical calpains have been identified thus far. Three are found in *Drosophila melanogaster* as Dm-calpain/CALPA, CALPB, and CALPC. *Schistosoma mansoni* and *Schistosoma japonicum* also have at least one typical calpain (Sm-calpain and Sj-calpain). No typical calpain homologs have been found in *Caenorhabditis elegans*, plants, fungi, trypanosomes, and *Saccharomyces cerevisiae*.

The second group (atypical calpains) contains various molecules that have the protease domain but do not have overall similarity as to domains III and IV with extra domain(s). These atypical calpains are thought to have somewhat different mode of action compared with those of typical calpains. Atypical calpains include the small optic lobe (SOL) and PalB subfamilies, and several alternative splicing products of, for example, *Capn8* (nCL-2') and *CalpA*, and others. Although the structure is called atypical, PalB subfamily members are conserved in animals, fungi, and yeast, and SOL subfamily members are found in animals and green algae, but typical calpains only exist in vertebrates and *Drosophila*.

In addition to the structural features, independent classification of mammalian calpains is possible according to their tissue/organ distribution. μ CL/CAPN1, mCL/CAPN2, PalB homolog (PalBH)/CAPN7, CAPN10, CAPN13, and SOL homolog (SOLH)/CAPN15 are ubiquitously expressed, whereas p94/CAPN3, nCL-2/CAPN8, nCL-4/CAPN9, hTRA-3/CAPN5, CAPN6, 10, 11, and 12 are predominantly expressed in specific

organs (Table 1). Ubiquitous calpains play fundamental roles for all cells, whereas tissue-specific calpains are involved in functions evolved specifically in tissues they express. Therefore, defects in ubiquitous calpains often cause lethality like CAPN2, while those in tissue-specific calpains are responsible for tissue-specific diseases such as muscular dystrophy. In addition, under various disease conditions such as muscular dystrophies, cardiomyopathies, ischemic disorders, and lissencephaly, conventional calpains have trends to be over-activated probably due to the increase of intracellular Ca^{2+} concentration caused by diseases, which often aggravate the diseases. Thus, inhibitors for conventional calpains are used to prevent progression of such diseases.

Structure and Functions of Calpain Superfamily Members

Skeletal muscle-specific calpain, p94/CAPN3

p94/CAPN3, the first tissue-specific calpain found in 1989, is c. 60% identical to the large subunits of μ - and m-calpains and has a conserved domain structure (Figure 1). However, p94/CAPN3 contains three p94-specific regions, NS, IS1, and IS2, which are closely related to unique features of p94/CAPN3.

Messenger RNA (mRNA) for p94/CAPN3 is expressed predominantly in skeletal muscle and the amount expressed is approximately 10 times larger than those of conventional calpains. p94/CAPN3 possesses several unique properties. For example, p94/CAPN3 protein undergoes extremely rapid autolysis (half-life *in vitro* is less than 10 min) starting with a cut in NS and IS1, and this autolysis is obviated by deletion of IS1 or IS2. Specific inhibitors of μ - and m-calpains such as calpastatin, E-64, and leupeptin have little effect on autolysis. Furthermore, p94/CAPN3 possesses a nuclear localization signal-like sequence in IS2 and is localized in the nucleus in addition to the cytosol. p94/CAPN3 binds to gigantic muscle protein, connectin/titin, specifically through the region close to IS2. The protease activity of p94/CAPN3 is suppressed *in vivo*, at least partly by binding to connectin/titin N2A region, where other important subcellular structures of muscle, such as sarcoplasmic reticulum and T-tubules, are aligned.

Some alternative splicing products (Lp82, Up86, etc.) of *Capn3* are expressed in tissues other than muscles. Lp82 showed lens-specific expression and Ca^{2+} -dependent protease activity against β A3 and α B crystallins. The activity is inhibited by E-64, but not by calpastatin. Some other splicing variants are expressed ubiquitously, and in embryonic skeletal muscles, although the physiological functions of these variants remain clear.

In 1995, mutations in CAPN3 were shown to be responsible for limb-girdle muscular dystrophy type 2A (LGMD2A), also called 'calpainopathy'. Positions of the mutation found in the LGMD2A patients were distributed widely within CAPN3, with more than half of the mutations being missense mutations (see Figure 3). No hot point was found, making its diagnosis very difficult. A primary cause of LGMD2A is a defect in protease activity, and not structural property, of p94.

nCL-2 and nCL-2' (CAPN8)

nCL-2 and nCL-2' are alternative splicing products of *Capn8*, with and without, respectively, domains III+IV. They are predominantly expressed in the stomach with lesser amounts

in intestines. nCL-2/CAPN8 is highly similar to mCL/CAPN2 along the whole molecule (*c.* 62% identical). Moreover, CAPN8 and CAPN2 are closely located, and their transcripts have overlap, that is, complementary sequences. Recombinant nCL-2 and nCL-2' proteins were expressed in *Escherichia coli* and showed Ca^{2+} -dependent caseinolytic activities. *Xenopus laevis* possesses an nCL-2 ortholog, xCL-2, the disruption of which causes severe developmental defects. Unlike m-calpain, however, nCL-2 does not form a hetero-dimer with 30 K, but instead has activity as a monomer and also forms homo-oligomer via domain III *in vitro*.

nCL-4/CAPN9

nCL-4/CAPN9 is also a typical calpain homolog that is predominantly expressed in the gastrointestinal tract. It possesses overall similarity to μ CL/CAPN1 and mCL/CAPN2 and requires 30 K/CAPNS1 for its activity *in vitro*. Recombinant human nCL-4+30 K protein showed Ca^{2+} -dependent caseinolytic activity, which was inhibited by calpastatin and other cysteine protease inhibitors, as in the case of conventional calpains. nCL-4 is reported to be involved in tumor suppression.

Transform genotypic hermaphrodites (TRA-3) and its orthologs

TRA-3 is involved in the sex determination cascade of *C. elegans*. Although enzymatic characterization of the purified enzyme has not yet been reported, protease activity of TRA-3 is Ca^{2+} -dependent and necessary for female development in XX hermaphrodites through the processing of TRA-2A membrane protein. Interestingly, TRA-3 as well as another *C. elegans* calpain, CLP-1, is also involved in neuronal cell necrosis in cooperation with aspartic proteases ASP-3 and ASP-4. Mammals possess two orthologs of TRA-3, hTRA-3/CAPN5 and CAPN6, whose amino acid sequences are more than 30% identical to that of TRA-3. The T domain, which is conserved in all three molecules, has weak similarity to the C2-domain and, thus, they belong to the tandem C2 PalB subfamily described below. Surprisingly, CAPN6 apparently has no active site residues (active site Cys is substituted with Lys), strongly suggesting that CAPN6 has no proteolytic activity. CAPN6 was reported to be involved in the regulation of microtubule dynamics.

Cpl1, PalB, and its orthologs (PalB subfamily)

Cpl1 is the only calpain homolog found in *S. cerevisiae* and is considered both structurally and functionally to be the ortholog of *Aspergillus* PalB, which plays important roles in the adaptation of fungi to alkaline conditions. *CPL1*, also referred to as *RIM13*, is involved in both the alkaline adaptation and sporulation process of yeast through its processing activity. Rim101 and the *Aspergillus* ortholog PacC, are *in vivo* substrates for Cpl1 and PalB, respectively, and activated by their C-terminal processing. Several other yeasts such as *Candida albicans* and *Pichia pastoris* also have Cpl1 orthologs. In the C-terminus of domain III, Cpl1, PalB, and their orthologs share a somewhat conserved domain, which also has weak similarity to the domain III. Thus, they have two C2-like domains tandemly. Mammals have one ortholog, PalBH/CAPN7, whose physiological functions are elusive, but its involvement in membrane trafficking around endosomes is suggested according to the biochemical studies and analogy to yeasts and fungi.

Calpain 10/CAPN10

This homolog was identified by reverse genetics of non-insulin-dependent diabetes mellitus (NIDDM), type 2 diabetes. The single nucleotide polymorphism (SNP) in intron 3 of *CAPN10* is statistically related to a risk of NIDDM. This SNP may affect transcription levels of *CAPN10*, although direct relationship between CAPN10 and NIDDM has not yet been elucidated. *CAPN10* generates several alternative splicing products. The longest, CAPN10a, has two C2-like domains moderately and weakly similar to that of domain III and, thus, CAPN10a also belongs to PalB subfamily. The physiological roles of calpain 10/CAPN10 are unclear.

SOL and its orthologs (SOL subfamily)

The *Drosophila* gene responsible for a defect in neuronal cells (small optic lobes) corresponding to mammalian retina was positionally cloned and shown to encode a calpain homolog with several Zn-finger motifs located at the N-terminus. Mammals possess one ortholog, SOLH/CAPN15, while *C. elegans* has several. They share a conserved C-terminal structure called the SOL homology domain (SOH).

Defective kernel 1 (DEK1) and its orthologs (DEK1 subfamily)

The maize *defective kernel 1* gene required for aleurone cell development in the endosperm of maize grains encodes a plant calpain homolog. The primary structure of DEK1 is unique in that it has 21 transmembrane regions located at the N-terminus and a C2-like domain, significantly similar to domain III, located at the C-terminus. Rice, *Arabidopsis*, and loblolly pine have very conserved orthologs (above 70% identity), and the whole genome sequence of *Arabidopsis* has revealed that DEK1 is the only calpain homolog in the genome. DEK1 plays a key role in growth regulation in *Arabidopsis*, and although full-length DEK1 protein localizes to membranes, it undergoes intramolecular autolytic cleavage events that release the calpain domain into the cytoplasm, which is sufficient for full complementation of *dek1* mutants.

Other calpain homologs

As described with the TRA-3 subfamily, some of the calpain homologs possess substituted residues in one of the very conserved active site residues, Cys, His, and Asn. Besides CAPN6 and *Drosophila* CALPC, some of the *C. elegans* homologs and all of the trypanosome homologs, do not possess one or more of the active site residues. These molecules are thought not to possess Cys protease activity. Some of the *C. elegans* calpain homologs possess Gly-rich sequences, which is a characteristic of domain V of the calpain regulatory subunits (CAPNS1 and CAPNS2). Trypanosome calpains have N-terminal domains weakly similar to calpastatin. The physiological significance of these structural features has yet to be determined.

Conclusion

This article introduces unique characteristics of various members of the calpain superfamily, mainly from the viewpoint of structure. Various calpains are involved in a variety of biological events and, thus they, if not properly carried on, cause lethality or diseases. Precise molecular roles of most

calpains, however, are still elusive. Prospective approaches for a breakthrough in the field of calpain include studies using genetically modified living organisms such as yeasts, plants, nematodes, drosophila, and mice. These rapidly developing methodologies are revealing novel relationships between calpain and other molecules, which will lead to elucidation of physiological functions of the calpain superfamily at the molecular level.

See also: **Bioenergetics:** Calcium-Binding Proteins: Cytosolic (Annexins, Gelsolins, and C₂-Domain Proteins); Calcium-Modulated Proteins (EF-Hand); **Cell Architecture and Function:** Unfolded Protein Responses; **Metabolism Vitamins and Hormones:** Diabetes; **Protein/Enzyme Structure Function and Degradation:** Cysteine Proteases; Protein Degradation.

Further Reading

- Denison SH, Orejas M, and Arst HN, Jr. (1995) Signaling of ambient pH in *Aspergillus* involves a cysteine protease. *Journal of Biological Chemistry* 270: 28519–28522.
- Futai E, Maeda T, Sorimachi H, et al. (1999) The protease activity of a calpain-like cysteine protease in *Saccharomyces cerevisiae* is required for alkaline adaptation and sporulation. *Molecular and General Genetics* 260: 559–568.
- Goll DE, Thompson VF, Li H, Wei W, and Cong J (2003) The calpain system. *Physiological Review* 83: 731–801.
- Hanna RA, Campbell RL, and Davies PL (2008) Calcium-bound structure of calpain and its mechanism of inhibition by calpastatin. *Nature* 456: 409–412.
- Horikawa Y, Oda N, Cox NJ, et al. (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nature Genetics* 26: 163–175.
- Khorchid K and Ikura M (2002) How calpain is activated by calcium. *Nature Structural Biology* 9: 239–241.
- Kimura Y, Koga H, Araki N, et al. (1998) The involvement of calpain-dependent proteolysis of the tumor suppressor NF2 (merlin) in schwannomas and meningiomas. *Nature Medicine* 4: 915–922.
- Lid SE, Gruis D, Jung R, et al. (2002) The defective kernel 1 (dek1) gene required for aleurone cell development in the endosperm of maize grains encodes a membrane protein of the calpain gene superfamily. *Proceedings of the National Academy of Sciences of the United States of America* 99: 5460–5465.
- Margis R and Margis-Pinheiro M (2003) Phytocalpains: Orthologous calcium-dependent cysteine proteinases. *Trends in Plant Science* 8: 58–62.
- Moldoveanu T, Hosfield CM, Lim D, et al. (2002) A Ca²⁺ switch aligns the active site of calpain. *Cell* 108: 649–660.
- Moldoveanu T, Hosfield CM, Lim D, Jia Z, and Davies PL (2003) Calpain silencing by a reversible intrinsic mechanism. *Nature Structural Biology* 10: 371–378.
- Richard I, Broux O, Allamand V, et al. (1995) Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. *Cell* 81: 27–40.
- Sorimachi H, Ishiura S, and Suzuki K (1997) Structure and physiological function of calpains. *Biochemical Journal* 328: 721–732.
- Sorimachi H and Suzuki K (2001) The structure of calpain. *Journal of Biochemistry* 129: 653–664.
- Strobl S, Fernandez-Catalan C, Braun M, et al. (2000) The crystal structure of calcium-free human m-calpain suggests an electrostatic switch mechanism for activation by calcium. *Proceedings of the National Academy of Sciences of the United States of America* 97: 588–592.
- Suzuki K, Hata S, Kawabata Y, and Sorimachi H (2004) Structure activation, and biology of calpain. *Diabetes* 53: S12–S18.
- Syntichaki P, Xu K, Driscoll M, and Tavernarakis N (2002) Specific aspartyl and calpain proteases are required for neurodegeneration in *C. elegans*. *Nature* 419: 939–944.
- Yamada M, Yoshida Y, Mori D, et al. (2009) Inhibition of calpain increases LIS1 expression and partially rescues *in vivo* phenotypes in a mouse model of lissencephaly. *Nature Medicine* 15: 1202–1207.

Relevant Websites

- <http://cutdb.burnham.org> – A Proteolytic Event Database.
- <http://calpain.org> – Calpain for Modulatory Proteolysis.
- <http://www.dmd.nl> – Leiden Muscular Dystrophy pages.
- <http://merops.sanger.ac.uk> – Merops – the Peptidase Database.
- <http://www.rinshoken.or.jp> – The Tokyo Metropolitan Institute of Medical Science; Calpain Project Home Page; Calpain.
- <http://ag.arizona.edu> – The University of Arizona; The calpain family of proteases.
- <http://www.worldmusclesociety.org> – World Muscle Society.

Carbohydrate Chains: Enzymatic and Chemical Synthesis

T J Tolbert, Indiana University, Bloomington, IN, USA

C-H Wong, The Scripps Research Institute, La Jolla, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Anomeric carbon The carbon of a cyclic sugar which forms a hemiacetal or hemiketal. In the linear form of the sugar, the carbon that will become the anomeric carbon when the sugar cyclizes is the carbonyl carbon.

Anomeric effect The stabilization of axial orientation over equatorial orientation of electron withdrawing groups attached to the C1 of pyranose sugars.

Anomers The pair of diastereomers, termed the α - or β -anomer, that results when a linear sugar forms a cyclic hemiacetal or hemiketal.

Glycosidic linkage The linkage formed between the anomeric carbon of a sugar and an alcohol.

Difficulties in the Synthesis of Carbohydrate Chains

The synthesis of oligosaccharides can present several technical difficulties that do not occur in the synthesis of other biopolymers. Unlike nucleic acids and proteins, the linkage between each subunit (i.e., sugar) of a carbohydrate chain forms a chiral center at the sugar's anomeric carbon that has two possible configurations, termed the α - or β -anomer (Figure 1(b)). Each time a sugar is joined to a carbohydrate chain through a glycosidic linkage, the stereochemistry of the bond that is formed must be controlled or directed in some manner to ensure that the correct anomer is produced. In addition, carbohydrate chains are not just simple linear chains as DNA and proteins are, but can also be branched chains with increased complexity. Glycosidic linkages can be formed through each hydroxyl of a sugar, and when multiple hydroxyls on a single sugar form glycosidic linkages, branched oligosaccharide structures result. Distinguishing between the different hydroxyl moieties on sugars is another difficulty of carbohydrate chain synthesis. Each sugar of a carbohydrate chain can have several hydroxyl moieties that are nearly equivalent in chemical reactivity, making them difficult to differentiate from one another and adding several steps to most chemical syntheses. To overcome these difficulties, several enzymatic and chemical methods have been developed.

Enzymatic Synthesis of Oligosaccharides

Enzymes have frequently been employed to synthesize carbohydrate chains, because they offer a few technical advantages over traditional organic synthesis. Enzymes generally do not require protecting groups to select the correct chemical moiety on their substrates, and in carbohydrate chain synthesis this is a great advantage. The many nearly chemically equivalent hydroxyl moieties of sugars make selective chemical protection of sugar hydroxyls a laborious task. The use of enzymes in carbohydrate chain synthesis can eliminate the need for both selective hydroxyl protection before forming the glycosidic linkage, and removal of the protecting groups after forming

the glycosidic linkage, greatly reducing the number of steps required for synthesis of an oligosaccharide. Another advantage of the use of enzymes in carbohydrate chain synthesis is the stereoselectivity of glycosidic bond-forming enzymes, which often produce a single anomer during the formation of a glycosidic bond. By contrast, chemical methods frequently produce mixtures of anomers during glycosidic bond formation, which must be purified from one another after formation of the glycosidic bond. Though enzymatic synthesis of oligosaccharides has many advantages, it is often limited by the availability of specific enzymes needed to form certain types of carbohydrate structures, but with the increased number of enzymes being discovered by genomic research this should improve in the future. Two classes of enzymes that have been used to form carbohydrate chains will be discussed further below, glycosidases and glycosyltransferases.

Carbohydrate Chain Synthesis Utilizing Glycosidases

Glycosidases are enzymes that normally break glycosidic bonds during glycoprocessing or catabolism of oligosaccharides, but by placing glycosidases under certain controlled reaction conditions they can be utilized to form, rather than break, glycosidic bonds. Most often, glycosidases are used to form glycosidic bonds in transglycosylation reactions, where a glycosidic bond is broken in a glycosyl donor glycoside, and a new glycosidic bond is formed with a glycosyl acceptor (Figure 2(a)). Several approaches can be utilized to favor formation of the desired glycosidic bond, including the use of activated glycosyl donors such as *p*-nitrophenyl glycosides, elevated concentrations of glycosyl donors and acceptors, and organic cosolvents. The use of glycosidases to form glycosidic bonds generally results in low to medium yields ranging from 20% to 40%, although yields as high as 90% have been reported using mutated glycosidases. Glycosidase-catalyzed synthesis generally has good stereoselectivity, forming a single α - or β -anomer, but sometimes has low regioselectivity for the different hydroxyls on the acceptor sugar resulting in multiple products being formed.

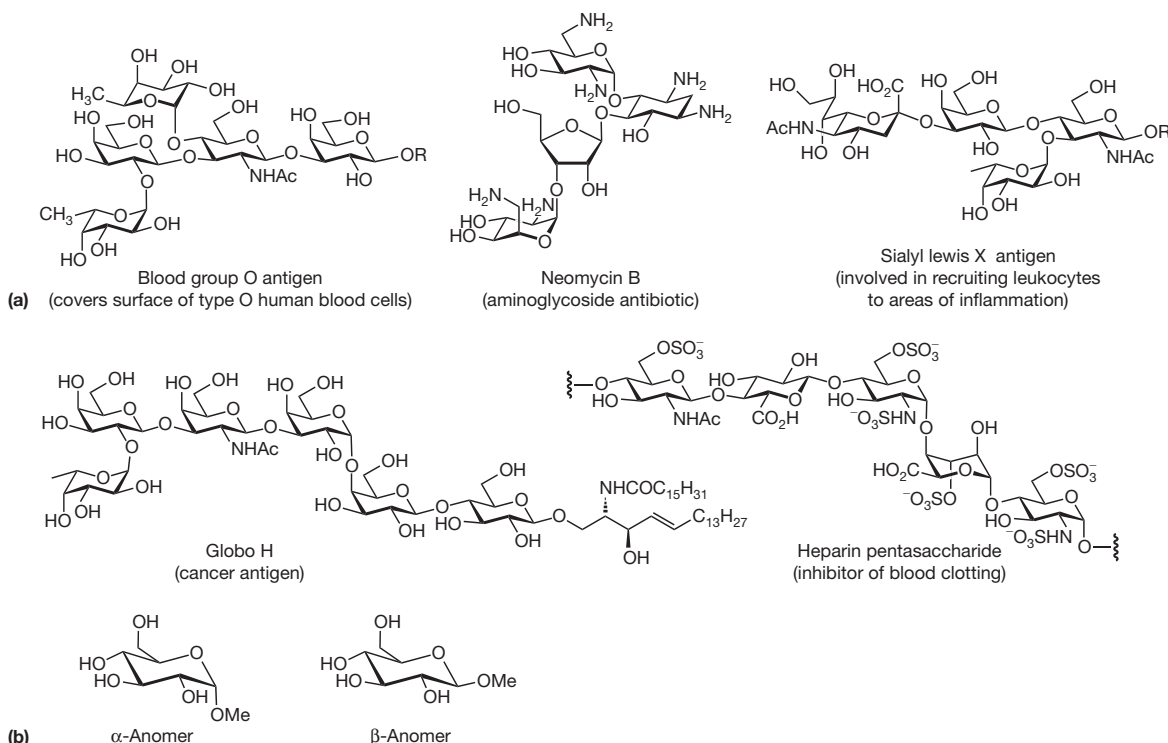


Figure 1 Carbohydrate chains: (a) some biologically important oligosaccharides and (b) 1-*O*-methyl-glucopyranosides, α - and β -anomers.

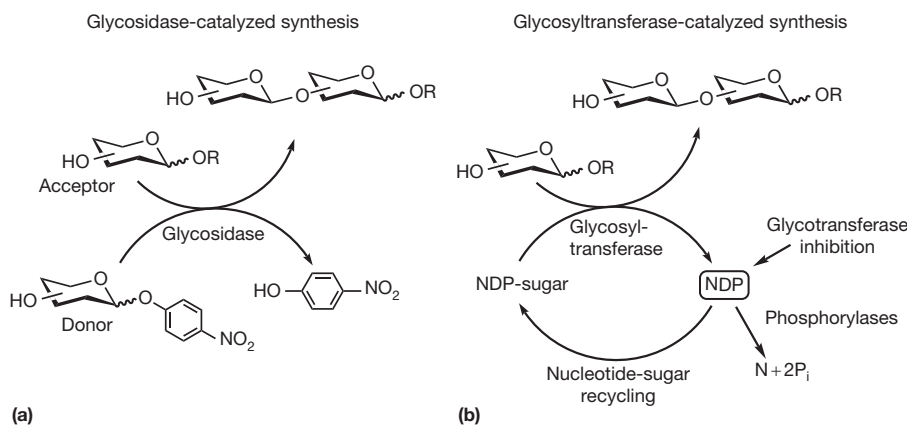


Figure 2 Enzymatic synthesis of oligosaccharides using: (a) glycosidases and (b) glycosyltransferases.

Though glycosidase reactions can suffer from low yield and low regioselectivity, there is a wide range of glycosidases available to catalyze the formation of many different types of glycosidic linkages.

Synthesis of carbohydrate chains using mutated glycosidases

Mutation of key amino acid residues in the active sites of many glycosidases can reduce or abolish glycosidic bond cleaving activity while retaining the ability to conduct transglycosylation reactions with natural oligosaccharides or activated glycosyl donors. Such mutated glycosidases are often called glycosynthases in reference to their greater transglycosylation

activities. Using glycosynthases with activated donors has resulted in yields as high as 90% for the formation of glycosidic bonds. Directed evolution is currently being applied to many recombinant glycosidases to expand the range of glycosynthases available for the synthesis of carbohydrate chains.

Carbohydrate Chain Synthesis Utilizing Glycosyltransferases

Glycosyltransferases are enzymes that catalyze the transfer of activated monosaccharide donors to carbohydrates during the biosynthesis of oligosaccharides. They are very useful in synthesizing oligosaccharides *in vitro* because they exhibit high

regioselectivity and stereoselectivity in the formation of glycosidic bonds. Glycosyltransferases that utilize nucleotide sugars as activated monosaccharide donors have often been used in the *in vitro* synthesis of carbohydrate chains (Figure 2(b)). A wide variety of oligosaccharide structures can be produced using the available glycosyltransferases and activated nucleotide sugars such as UDP-glucose (UDP-Glc), UDP-*N*-acetylglucosamine (UDP-GlcNAc), UDP-galactose (UDP-Gal), UDP-*N*-acetylgalactosamine (UDP-GalNAc), UDP-glucuronic acid (UDP-GlcUA), GDP-mannose (GDP-Man), GDP-fucose (GDP-Fuc), and CMP-sialic acid (CMP-NeuAc). Generally, each glycosyltransferase is selective for a specific nucleotide-sugar donor, and also has specificity for certain oligosaccharide structures in its substrates. Because of this it is often difficult to make unnatural oligosaccharides using glycosyltransferases, but in some cases relaxed substrate and nucleotide-sugar specificity has been utilized to make unnatural carbohydrate chains. Formation of oligosaccharides with glycosyltransferases requires the substrate oligosaccharide, a divalent cation, the nucleotide-sugar donor, and the glycosyltransferase itself in an appropriate buffer. Glycosyltransferase-catalyzed synthesis of carbohydrate chains usually results in high yields, approaching 100%, but the enzymatic reaction can suffer from product inhibition that has to be overcome to achieve those high yields. The use of glycosyltransferases in oligosaccharide synthesis is limited by the high cost of nucleotide sugars and the availability of glycosyltransferases that form certain oligosaccharide structures. Sugar-nucleotide recycling alleviates some of the former problem while genomic sequencing and research has started to alleviate some of the latter problem.

Product inhibition in glycosyltransferase reactions

Though glycosyltransferases can be used to form oligosaccharides in nearly 100% yield, they suffer from product inhibition from the nucleoside diphosphates and nucleoside monophosphates that are produced as the activated nucleotide-sugar donors are transferred to the growing carbohydrate chain. This product inhibition can drastically slow the enzymatic reaction and also reduce the yield of product, and so it is desirable to remove the nucleotide by-products of the glycosyltransferase reactions. Two approaches can be used to overcome product inhibition in glycosyltransferase reactions, use of phosphorylases and nucleotide-sugar recycling (Figure 2(b)).

Reduction of glycosyltransferase inhibition with phosphorylases. Nucleoside diphosphates and monophosphates can be converted into nucleosides, which do not inhibit glycosyltransferases, using phosphorylases. This is a very simple method, which requires only one or two additional enzymes to be added to the glycosyltransferase reaction. Unfortunately, this method requires a stoichiometric amount of nucleotide-sugar donor for the formation of the desired oligosaccharide, which can be quite costly on larger scales since nucleotide sugars are generally very expensive.

Reduction of glycosyltransferase inhibition with nucleotide-sugar recycling. Another method to remove nucleotide by-products of glycosyltransferase reactions is to regenerate the nucleotide-sugar donors from the nucleotide by-products using an enzymatic recycling reaction. This method of overcoming product inhibition is somewhat more complicated than using

phosphorylases, requiring several additional enzymes for the nucleotide-sugar recycling reaction, but it allows a catalytic amount of nucleotide-sugar donor to be used in the glycosyltransferase reaction. Since nucleotide-sugars are generally expensive, this method is desirable for larger-scale glycosyltransferase reactions.

Recombinant expression of glycosyltransferases

Genomic sequencing efforts have made the DNA sequences of many mammalian and bacterial glycosyltransferases freely available. This has enabled efforts to recombinantly overexpress glycosyltransferases for use in the synthesis of carbohydrate chains. The use of bacterial and yeast expression systems has allowed several mammalian glycosyltransferases to be produced on relatively large scale, and it has also been recognized that many bacterial glycosyltransferases, which are often easier to express, can be utilized to produce mammalian-type carbohydrate structures. These efforts are ongoing and are continually increasing the range of carbohydrate structures that can be formed using glycosyltransferases.

Chemical Synthesis of Oligosaccharides

Chemical synthesis of oligosaccharides is often the method of choice for constructing carbohydrate chains because chemical methods are flexible, can be used to produce both natural and unnatural oligosaccharides, and are also not limited by the availability of specific enzymes. Formation of glycosidic bonds by chemical methods usually relies upon activation of a leaving group on the anomeric carbon of a glycosyl donor with a Lewis acid, which once activated will react with a free hydroxyl upon a glycosyl acceptor (Figure 3(a)). A wide variety of glycosyl leaving groups have been utilized for carbohydrate chain synthesis some of which are shown in (Figure 3(b)).

Control of Stereochemistry in Chemical Glycosidic Bond Formation

Control of the stereochemistry of the anomeric linkage in the products of glycosylation reactions can be very complex, and many factors including types of protecting groups on the sugars, solvent, temperature, and leaving groups can be used to influence the α/β ratio of the products formed in glycosylation reactions. In general, the anomeric effect, stabilization of axial orientation over equatorial orientation of electron-withdrawing groups attached to the C1 of pyranose sugars, can be utilized to produce α -linked glucosides and galactosides. Participation of protecting groups of the 2-OH, such as acetate, and the use of polar solvents in glycosylation reactions can direct products toward β -linked glucosides and galactosides through dioxocarbenium or solvent intermediates (Figure 3(a)). Other methods for controlling the stereochemistry of glycosidic bond formation include tethering glycosyl donors to glycosyl acceptors and utilization of protecting groups to direct stereochemistry through steric and electronic effects. Sometimes it is difficult to obtain the desired anomer using chemical synthesis, as in the case of sialic acid glycosides,

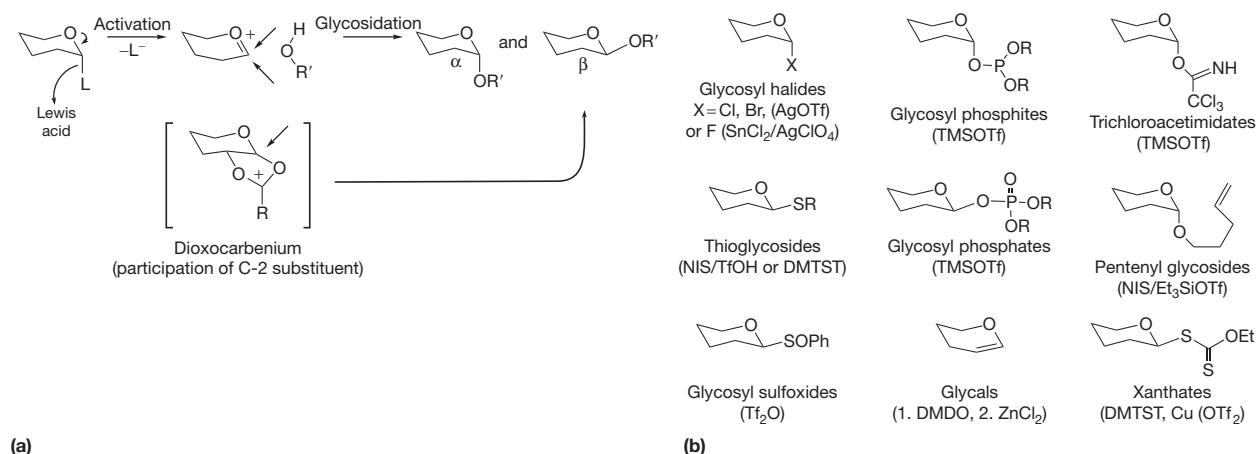


Figure 3 Chemical glycosylation reactions. (a) Mechanism of glycosylation. (b) Some commonly used glycosylation leaving groups and the reagents used to activate them (in parentheses). Abbreviations: L, leaving group; R, variable group; TMSOTf, trimethylsilyl triflate; NIS, *N*-iodosuccinimide; TfOH, triflic acid; Et₃SiOTf, triethylsilyl triflate; Tf₂O, triflic anhydride; DMDO, 3,3-dimethyldioxirane; DMTST, dimethylthiosulfonium triflate.

where the natural sialic acid linkage is exclusively the α -anomer, but the β -anomer is obtained from most chemical glycosylation reactions. Controlling the anomeric outcome of chemical glycosylation reactions can still be a difficult problem, and requires several approaches for different types of glycosidic linkages and sugars.

Protection of Sugar Hydroxyls in Chemical Carbohydrate Synthesis

Chemical oligosaccharide synthesis relies heavily upon sugar protecting groups both to distinguish between sugar hydroxyls with similar chemical reactivity and also to control the stereochemical outcome, either α or β , of glycosylation reactions. Because of this, chemical oligosaccharide synthesis typically requires a large amount of protecting group manipulations. Normally, protecting groups must be placed on hydroxyls that are not going to be used to form glycosidic linkages to prevent them from reacting in glycosylation reactions. In addition, the synthesis of carbohydrate chains longer than two sugars requires sugar subunits that can act both as glycosyl donors and acceptors, and this usually requires selective protection and deprotection of the hydroxyls that are to form glycosidic linkages. Selective protection of sugar subunits for glycosylation reactions often requires many chemical steps to produce each selectively protected subunit for the synthesis of an oligosaccharide, and this is one factor that can contribute to the large number of steps required for chemical oligosaccharide syntheses. To make the process of producing protected sugar subunits more efficient, one-pot regioselective approaches are being developed to install orthogonal protecting groups on sugars in a single reaction vessel. These one-pot procedures simplify protecting group manipulations by eliminating time-consuming purification and characterization of intermediates. Some protecting groups commonly used in chemical carbohydrate chain synthesis include benzyl ethers, *p*-methoxybenzyl ethers, *tert*-butyl-diphenylsilyl ethers, *tert*-butyl-dimethylsilyl ethers, allyl ethers, acetate esters, benzoate esters, levulinoyl esters, and dimethyl acetals.

Solution-Phase Chemical Oligosaccharide Synthesis

In conventional solution-phase oligosaccharide synthesis, a donor sugar is activated by a Lewis acid at the anomeric carbon, and then reacted with an acceptor sugar that contains a free hydroxyl. Once the glycosidic bond has been formed, the product disaccharide must be purified away from other glycosylation reaction reagents and side products, and then selectively deprotected to unmask a single hydroxyl so that it can then act as a glycosyl acceptor in the next glycosylation reaction (Figure 4(a)). Purification is often necessary after each glycosylation and deprotection step in the preparation of an oligosaccharide by solution-phase synthesis. When combined with the number of steps necessary to produce selectively protected sugar subunits, the number of chemical steps and purifications necessary to construct even small oligosaccharides can be very large.

Automated Oligosaccharide Synthesis

Because of the large amount of work necessary to construct even relatively small oligosaccharides by solution-phase synthesis, many approaches toward automating oligosaccharide synthesis have been developed in hopes of minimizing the number of chemical steps and purifications required to synthesize oligosaccharides. Two approaches to automated oligosaccharide synthesis will be discussed below: solid-phase oligosaccharide synthesis and programmable one-pot oligosaccharide synthesis.

Solid-phase chemical oligosaccharide synthesis

In solid-phase oligosaccharide synthesis, either the glycosyl donor or acceptor of a glycosylation reaction is attached to the solid phase. The advantage of this is that large excesses of the other components of the glycosylation reaction can be used to increase the rate of the reaction and ensure that it goes to completion, and then rinsed away by simple filtration. This has the potential of increasing the yield of glycosylation reactions and also eliminates many of the laborious

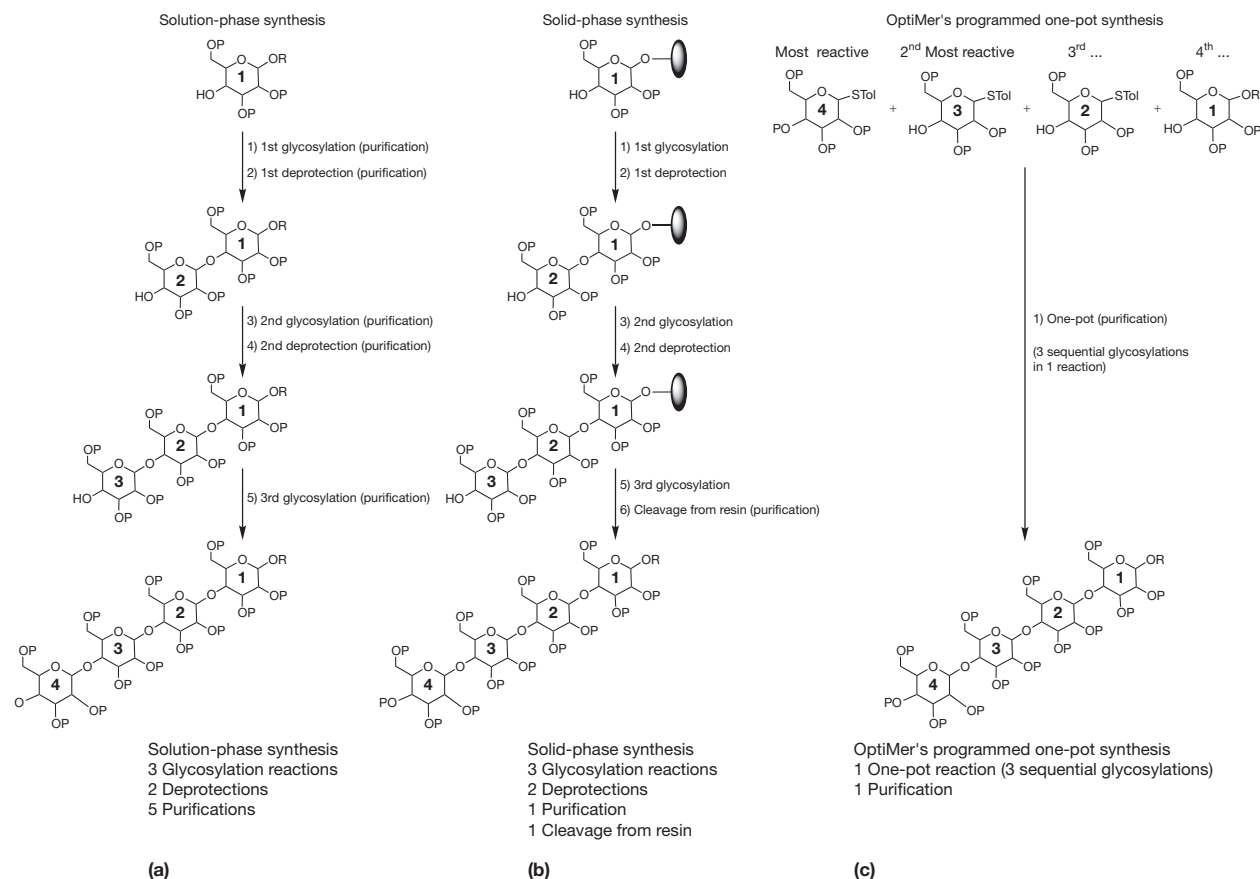


Figure 4 Chemical synthesis of oligosaccharides: (a) solution-phase synthesis, (b) solid-phase synthesis, and (c) OptiMer's programmed one-pot synthesis.

purification steps that are necessary in solution-phase oligosaccharide synthesis (Figure 4(b)). Unfortunately, there are many technical difficulties that must be overcome to successfully apply solid-phase synthesis to the production of oligosaccharides. Solid supports that function well with many solvents must be selected, because a wide variety of solvents are used to affect the outcome of glycosylation reactions. Linkers must be developed that are not sterically hindering for glycosylation reactions, stable to glycosylation conditions, and easily cleaved once the oligosaccharide has been synthesized. Strategies for protecting sugar monomers that allow highly flexible and selective deprotection of sugar hydroxyl groups in the presence of many other protected hydroxyls must be worked out. Since the reaction conditions of glycosidic bond formation can vary widely depending on oligosaccharide structure and the types of glycosidic linkages to be formed, there has been great difficulty in developing broadly applicable general methods for solid-phase oligosaccharide synthesis. Nevertheless, there have been many notable examples of successful solid-phase syntheses, including the synthesis of a dodecasaccharide by both the Nicolaou and Seeberger groups.

Programmable one-pot oligosaccharide synthesis

One-pot reaction approaches that involve conducting several sequential glycosylation reactions in one reaction flask have

been developed to facilitate automated oligosaccharide synthesis. The one-pot approach relies upon using a reactivity profile of protected sugars to determine what sequence of sugars to add to glycosylation reactions to obtain the desired products. Protected sugars are added to one-pot reactions in the order of most reactive to least reactive, thereby controlling the order of glycosidic bond formation during the synthesis of carbohydrate chains (Figure 4(c)). Since it has been shown that the types of protecting groups and type of anomeric activating groups used on sugars can greatly alter the speed at which a sugar will react in a glycosylation reaction, a wide range of reactivities can be obtained for a single type of sugar, allowing wide flexibility in oligosaccharide synthesis using this approach. The one-pot approach greatly reduces the number of steps necessary to construct oligosaccharides and eliminates several tedious purifications by combining multiple glycosylation reactions into a single one-pot reaction. The one-pot approach also minimizes protecting group manipulations by eliminating protecting group manipulations after construction of the building blocks. A large number of *p*-methylphenyl thioglycosides have been synthesized and had their reactivities measured to facilitate the use of this strategy. A computer program, OptiMer, has been developed that uses these *p*-methylphenyl thioglycosides as a set of building blocks to choose from to build oligosaccharides. When a carbohydrate chain structure is entered as input, the OptiMer program will calculate the best set of reactants for the

construction of that carbohydrate chain. Because the programmable one-pot strategy requires a minimum of protecting group manipulations and has a large library of building blocks, a wide range of oligosaccharide structures can be synthesized rapidly using this method.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Free Oligosaccharides: Formation and Degradation (Free Oligosaccharides Related to N-Linked Glycans); **Protein/Enzyme Structure Function and Degradation:** Enzyme Reaction Mechanisms: Stereochemistry.

Further Reading

- Boltje TJ, Buskas T, and Boons G-J (2009) Opportunities and challenges in synthetic oligosaccharide and glycoconjugate research. *Nature Chemistry* 1: 611–622.
- Chen R, Pawlicki MA, Hamilton BS, and Tolbert TJ (2008) Enzyme-catalyzed synthesis of a hybrid N-linked oligosaccharide using *N*-acetylglucosaminyltransferase I. *Advanced Synthesis and Catalysis* 350: 1689–1695.
- Hsu C-H, Chu K-C, Lin Y-S, et al. (2010) Highly alpha-selective sialyl phosphate donors for efficient preparation of natural sialosides. *Chemistry – a European Journal* 16: 1754–1760.
- Koeller KM and Wong C-H (2000) Synthesis of complex carbohydrates and glycoconjugates: Enzyme-based and programmable one-pot strategies. *Chemical Reviews* 100: 4465–4493.
- Nicolaou KC, Watanabe N, Li J, Pastor J, and Winssinger N (1998) Solid-phase synthesis of oligosaccharides: Construction of a dodecasaccharide. *Angewandte Chemie International Edition* 37: 1559.
- Plante OJ, Palmacci ER, and Seeberger PH (2001) Automated solid-phase synthesis of oligosaccharides. *Science* 291: 1523.
- Sears P and Wong C-H (2001) Toward automated synthesis of oligosaccharides and glycoproteins. *Science* 291: 2344–2350.
- Seeberger PH and Haase W-C (2000) Solid-phase oligosaccharide synthesis and combinatorial carbohydrate libraries. *Chemical Reviews* 100: 4349–4393.
- Shaikh FA and Withers SG (2008) Teaching old enzymes new tricks: Engineering and evolution of glycosidases and glycosyl transferases for improved glycoside synthesis. *Biochemistry and Cell Biology* 86: 169–177.
- Varki A, Cummings R, Esko J, et al. (eds.) (1999) *Essentials of Glycobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Wong C-H (2009) Enzymes for glycoprotein synthesis. *Chimia* 63: 318–326.
- Wong C-H, Halcomb RL, Ichikawa Y, and Kajimoto T (1995) Enzymes in organic synthesis: Application to the problems of carbohydrate recognition: Part 1. *Angewandte Chemie International Edition* 34: 412–432.
- Wong C-H, Halcomb RL, Ichikawa Y, and Kajimoto T (1995) Enzymes in organic synthesis: Application to the problems of carbohydrate recognition: Part 2. *Angewandte Chemie International Edition* 34: 521–546.
- Zhang Z, Ollman IR, Ye X-S, et al. (1999) Programmable one-pot digosaccharide synthesis. *Journal of the American Chemical Society* 121: 734.

Carbohydrate Metabolism in the Central Nervous System

I Simpson, Penn State College of Medicine, Hershey, PA, USA

S J Vannucci, Weill Cornell Medical College of Cornell University, New York, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

2-Deoxyglucose (2-DG) A glucose analog that can be transported and phosphorylated the same as glucose but essentially cannot be further metabolized. When labeled with ^{14}C or ^{18}F , 2-DG can be used as method for assessing glucose metabolism.

Aerobic metabolism This form of chemical breakdown involves oxygen and results in the more complete oxidative metabolism of a compound. In the case of glucose, oxidative metabolism couples the initial glycolytic pathway with the subsequent passage of pyruvate into the mitochondria where it is further metabolized to CO_2 and H_2O . This oxidation is coupled to the production of NADH and FADH_2 , which are further oxidized by the electron transport chain, which is coupled to the production of ATP. The aerobic metabolism of 1 mol of glucose generates approximately 30 mol of ATP.

Anaerobic metabolism This is strictly defined as the chemical breakdown of a compound in the absence of oxygen. In the case of glucose, glycolysis is an anaerobic sequence of reactions that is responsible for the conversion of glucose to lactate that results in the generation of two molecules of adenosine triphosphate (ATP)/mole of glucose.

Blood–brain barrier Morphologically, the barrier is comprised of the microvascular endothelial cells which are

joined by tight junction proteins, effectively preventing paracellular diffusion/leakage of molecules between the blood to the brain. Functionally, access from the blood to the brain is mediated by specific transporter proteins that transport compounds such as glucose or amino acids across both the luminal (blood side) and abluminal (brain side) membranes of the endothelial cells.

GLUTs A family of 14 facilitative glucose transporters. These transporters have differing affinities and capacities to transport glucose and other hexoses.

Glycogen A polymer of glucose comprised of linear chains of glucose molecules in a 1–4 linkage that are branched by the insertion of a 1–6 glucose linkages. It serves as an energy storage molecule and is found exclusively in astrocytes.

MCTs A family of 14 H^+ - or Na^+ -coupled monocarboxylate-related transport proteins. Only MCTs 1–4 are definitively monocarboxylate transporters.

Pentose-phosphate pathway This pathway, also known as the hexose monophosphate pathway, is an oxidative pathway responsible for approximately 5% of basal cerebral glucose metabolism. Its major functions are to generate NADPH for biosynthetic reactions and the maintenance of the levels of the antioxidant, reduced glutathione levels.

It is also the source of pentoses for RNA and DNA synthesis.

Glucose is central to the energy metabolism of most mammalian cells and it is the primary fuel for the mammalian brain. Unlike peripheral tissues such as heart and muscle, the brain is unable to access other energetic fuels, that is, circulating fatty acids and triglycerides, due to the restricted movement of these compounds across the blood–brain barrier (BBB). The importance of glucose as a metabolite derives from the fact that it is a rich source of energy (i.e., adenosine triphosphate (ATP) production), as the complete oxidation of glucose to CO_2 and H_2O proceeds with a standard free energy change of $-2840 \text{ kJ mol}^{-1}$. Glucose can be stored as multiple hexose units in the form of glycogen, providing a quick source of energy to the cell either aerobically or anaerobically. In addition to providing energetic fuel, glucose serves an anabolic role as a carbon source for numerous metabolic intermediates. In this regard, carbohydrate (glucose) metabolism in the brain is similar to that of peripheral tissues. The major pathways of glucose metabolism in brain are (1) glycolysis to provide three-carbon compounds, pyruvate to be used for biosynthetic purposes or to be oxidized in the tricarboxylic acid (TCA) cycle to generate energy as ATP through oxidative phosphorylation (90% of glucose utilization), or lactate for export; (2) glycogen synthesis for glucose storage; and (3) the pentose phosphate pathway for the formation of reducing equivalents in the form of reduced

nicotinamide adenine dinucleotide phosphate (NADPH) for lipid and antioxidant biosynthesis and nucleotide precursors (Figure 1). Each of these pathways proceeds in the brain as in the periphery. Lactate can be transported either from the periphery into the brain or between neural cells as an additional energy source. However, the brain differs from peripheral tissues in that it has quite unique characteristics that impact on various aspects of glucose metabolism. This article presents a review of cerebral glucose metabolism and the limitations and demands specific to the brain.

The human brain weighs approximately 1 kg which is equivalent to approximately 2% of the total body mass, but at rest it receives 15% of the cardiac output and is responsible for 20% of the total oxygen consumption and energy utilization, that is, a very expensive organ. Unlike most peripheral tissues, the brain requires a continual energy/glucose supply to maintain its level of activity and, unlike liver or muscle, has a limited capacity to store energy as glycogen, roughly equivalent to only 5 min worth in the event of a total cessation of supply. Moreover, the concentration of glucose that the neural cells are exposed to (i.e., the brain interstitial concentration) is approximately 20–30% of that observed by peripheral cells. Thus, given the normal levels of circulating glucose of 4.5–6 mM in humans and somewhat higher in rodents, the

levels of interstitial glucose concentrations range between 1 and 2 mM. During hypoglycemia, circulating levels can drop to 2 mM and the effective interstitial concentration approaches zero. Lactate is an intermediate in the metabolism of glucose and can also be an independent energy source for brain, providing approximately 8% of cerebral energy needs under normal conditions. In contrast to glucose, the circulating and interstitial lactate concentrations are approximately equal at 1 mM and remain in steady state. However, under circumstances where the circulating levels of lactate are increased, such as following intensive exercise, lactate uptake and utilization will increase with a concomitant decrease or sparing in cerebral glucose utilization.

The brain is not as homogeneous a tissue as peripheral organs, and demands for energy in the form of glucose demonstrate profound regional variation as illustrated in Figure 2. The figure depicts regional cerebral glucose utilization (rCMRglu) in rat (Figure 2(a)) and human (Figure 2(b)) brain as measured by the accumulation of isotopically labeled 2-deoxyglucose. This methodology, originally developed by Dr. Louis Sokoloff and colleagues, takes advantage of the inability of glycolytic enzymes to metabolize the glucose analog ^{14}C -2-deoxyglucose (2-DG) beyond the hexokinase product ^{14}C -2-deoxyglucose 6-P, which then accumulates in the cell in proportion to the cells' uptake and metabolism of glucose. The extent of accumulation at steady state

is measured quantitatively and expressed as $\mu\text{mol}/100\text{ g tissue}/\text{minute}$ as illustrated in Figure 2(a). An analogous approach is used to determine activity in human brain (Figure 2(b)), except that ^{18}F -2-DG is used as the glucose analog. ^{18}F is a positron emitter and the accumulation of ^{18}F -2-DG-6-P is then measured by positron emission tomography (PET scanning) in virtual axial sections that are taken through the brain. It is readily apparent from the numerical scales of Figure 2 that the rat brain metabolizes more glucose per gram of tissue/minute than does the human brain. This may well relate to the greater proportion of both white matter and glia in the human brain, which have lower energy demands than neurons. Rates of CMRglu are color coded as depicted in the scales adjacent to each figure, and the regional heterogeneity in glucose utilization and energy demand is apparent in the color variation apparent in both brains.

In addition to its regional heterogeneity, the brain is comprised of different populations of cells, each with profoundly different functions and consequent energy requirements. The major cell types are neurons and glia, including astrocytes, oligodendrocytes, and microglia, as illustrated in Figures 3 and 4. These neural cells are separated from the periphery by the microvascular endothelial cells that comprise the BBB. The BBB endothelial cells are the primary gatekeepers for the brain and all energy requirements must first be fulfilled by the passage of sufficient fuel, predominantly glucose, from the circulation across the two membranes of these cells. Thus, the two important components determining cellular glucose utilization are the energetic demand, that is, the cellular work and the major metabolic pathways, and the supply of the essential nutrients.

Figure 3 illustrates the integral membrane nutrient transporters (glucose transporters (GLUTs) for glucose; monocarboxylate transporters (MCTs) for monocarboxylic acids) that facilitate the uptake (supply) and export of the substrates into and out of the respective cells, while Figure 4 illustrates the metabolic pathways that predominate in each of the cells and constitute the demand.

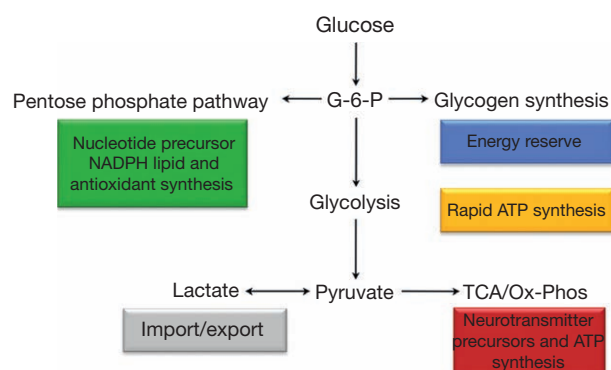


Figure 1 A diagrammatic representation of the various pathways that glucose and lactate are metabolized in the mammalian brain.

Glucose Transporters

The transport of glucose and other hexoses into most mammalian cells is mediated by the SLC2 family of 14 transport

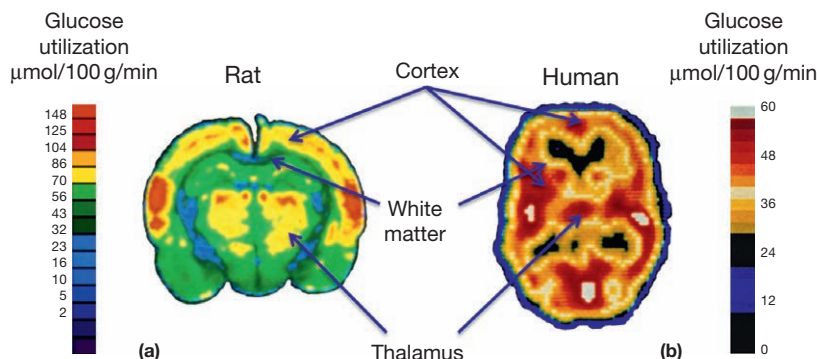


Figure 2 Cerebral glucose utilization in rat and human brain. The figure depicts a digitized autoradiograph of ^{14}C -2-deoxyglucose uptake/phosphorylation in conscious rat, coronal section (a), and a digitized positron emission derived from ^{18}F -2-deoxyglucose uptake in a virtual axial slice obtained from human brain (b). The color coding reflects the calculated cerebral glucose utilization in the respective slices using the algorithms originally formulated by Sokoloff and colleagues. Kindly provided by Dr. Carolyn Beebe Smith.

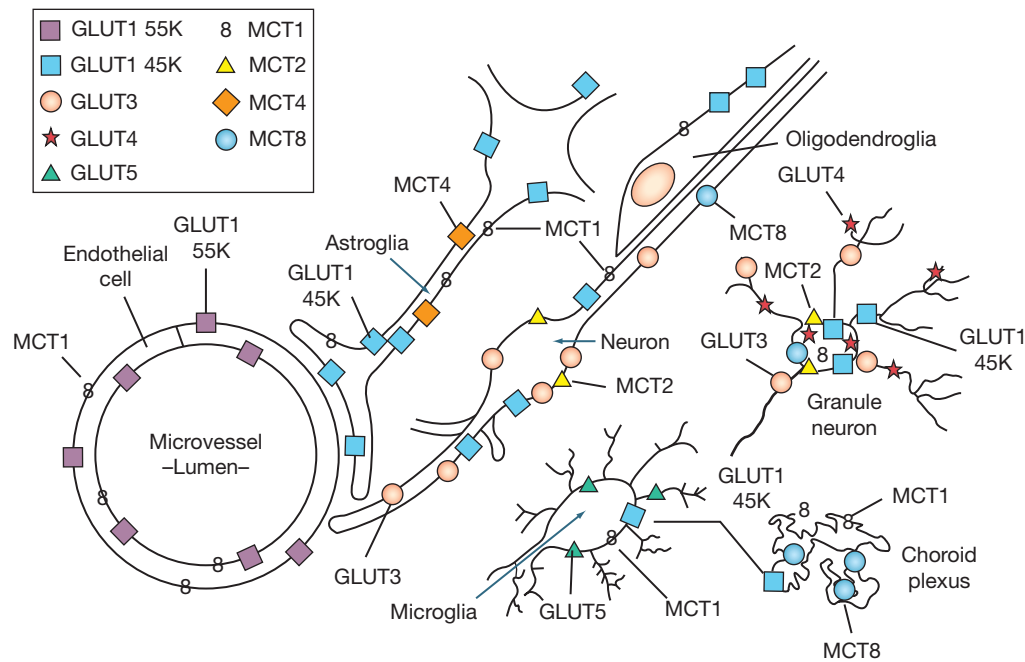


Figure 3 Neural localization of glucose and monocarboxylate transporters. The figure depicts the cellular localization of the various glucose (GLUTs) and monocarboxylate (MCTs) transporters.

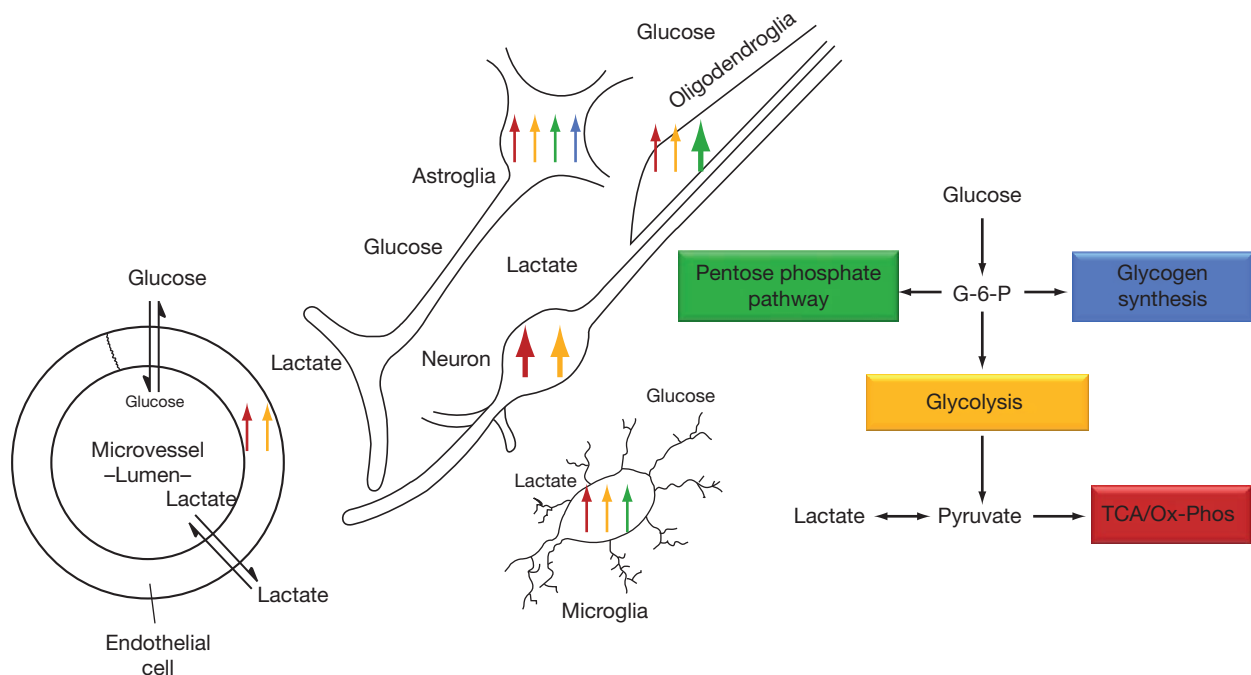


Figure 4 Neural localization of metabolic pathways for glucose. The figure illustrates the relevant metabolic pathways, originally described in [Figure 1](#), in the respective neural cells with corresponding color-coded arrows, and thickness related to the predominance of that pathway within a given cell.

proteins, GLUT1 to GLUT14, which includes HMIT, the proton-driven myo-inositol transporter. With the exception of GLUT1, which is ubiquitously expressed, the other family members have fairly distinct tissue- and cell-specific patterns of expression. As illustrated in [Figure 3](#), the brain with its

complex cellular heterogeneity expresses most, if not all, SLC2 family members. Much of the data supporting the indicated cellular distribution have been reviewed previously and will not be further discussed here. However, it is important to note that many of the transporters indicated in [Figure 3](#) either are unable

to transport glucose (GLUT5, -6, and -11, and HMIT have very low affinity for glucose) or have very limited localization and concentrations (GLUT2 and -4), whereas for others, their location, ability, or capacity to transport glucose have yet to be fully assessed (GLUT8, -9, and -10). Thus, with all of the above exclusions, the predominant transporters in mammalian brain involved in cerebral glucose utilization are GLUT1 and -3, which are expressed throughout the brain and whose localization and kinetic characteristics are well established. Two molecular forms of GLUT1 have been identified: a 55-kDa form in the endothelial cells and a 45-kDa form in the various glia (astrocytes, oligodendrocytes, and microglia) and choroid plexus. The difference in their relative molecular weight is accounted for by a differing extent of glycosylation, but this additional glycosylation does not appear to alter their protein structure or kinetic characteristics. GLUT3 is expressed predominantly in neurons and is distinct from GLUT1 in that its affinity for glucose is higher (K_m , 2–3 vs. 5–8 mM) and the ability to transport glucose is faster (K_{cat} , 6512 vs. 1166 s^{-1} @ 37 °C). An interesting feature of neuronal GLUT3 is that both messenger RNA (mRNA) and protein levels *in vivo* appear to be regulated in direct relation to changes in neuronal activity and rates of cerebral glucose utilization in several conditions underscoring the importance of glucose to neuronal energy demand.

Monocarboxylate Transporters

Also included in [Figure 3](#) is the cell-specific localization of the relevant MCTs, which are responsible for the transport of lactate, pyruvate, and the ketoacids, acetoacetate, and β -hydroxybutyrate.

The MCTs comprise the gene family SLC16, which has 14 members. However, of these 14 family members, only MCT1–4 have been definitively shown to transport metabolically relevant monocarboxylic acids such as lactate, pyruvate, and the ketone bodies, acetoacetate and β -hydroxybutyrate. Moreover, as illustrated in [Figure 3](#), all four are found in brain although MCT3 expression is limited to choroid plexus, and the pigment epithelial cells of the retina. The localization and kinetic characteristics of the MCTs in brain have been the subject of several reviews, and therefore only a brief synopsis is presented here. MCT1 has a very similar pattern of expression to GLUT1 in that it has ubiquitous tissue expression and is located in glia and endothelial cells in the adult rodent brain. The expression of MCT1 mRNA and protein are strongly developmentally regulated in the rodent brain as it serves as a ketone body transporter, which is an important cerebral substrate during development, as will be discussed later. The extent of MCT2 expression is species specific because, although widely expressed in rodents, expression in human tissue is very restricted. In rodent brain, MCT2 is expressed almost exclusively in neurons and subsequent studies have shown it to be extensively expressed in dendrites and associated with postsynaptic densities. MCT4 has been extensively characterized in highly glycolytic muscle, where it is responsible for lactate export. It is also expressed in chondrocytes, white blood cells, and in placenta, where it exports lactate from the fetal circulation. In rodent brain, MCT4 appears to be localized exclusively to astrocytes. In contrast to the glucose transporters, which transport glucose

by a facilitative mechanism that enables the bidirectional transport of glucose down its concentration gradient until equilibrium is reached, the MCTs transport the ketoacids by a proton-mediated co-transport mechanism. The binding constant (K_s) for the proton is 0.2 $\mu\text{mol l}^{-1}$, which corresponds to a pH of 6.7. As with the glucose transporters, the affinities of the MCTs for the ketoacids differ, with MCT2 having a greater affinity for lactate and pyruvate than MCT1 (K_m lactate 0.7 vs. 3.5 mM, pyruvate 0.1 vs. 0.7 mM) and both having much higher affinities for lactate and pyruvate than MCT4 (28 and 150 mM, respectively). Not included in [Figure 3](#) are the Na^+ -dependent glucose transporter(s) or the Na^+ -dependent MCTs whose presence has been shown in rodent brain but whose localization, concentration, and kinetic properties remain to be clarified.

Cellular Metabolic Pathways

The major energy demand for the brain is to fuel neurotransmission, which involves continual depolarization and repolarization of the neuronal membrane via the ionic pumps, Na^+/K^+ ATPase and Ca^{2+} -ATPase. It has been calculated that these action potentials and postsynaptic effects of neurotransmitters (predominantly glutamate) consume 47% and 34% of the total energy, respectively, with the maintenance of the resting membrane potential consuming a smaller amount (13%); glutamate recycling between astrocytes and neurons requires only 3%. To meet the high energetic demand of the neuron, the majority of the ATP is generated from glucose and lactate via the pathways of glycolysis, TCA, and final oxidative phosphorylation ([Figure 4](#)). The neurons are specifically equipped for this through their high mitochondrial content as well as expression of the neuron-specific GLUT3 glucose transporter ([Figure 3](#)) with the highest transport capacity (K_{cat}) determined to date. Thus, neurons appear to be the best equipped for rapid glucose uptake to meet their needs and it was generally considered that glucose was their major and obligate fuel. However, it has more recently been proposed that this function is actually served by lactate derived from astrocytic glycolysis, referred to as the astrocyte–neuron lactate shuttle (ANLS) which is analogous to the situation in skeletal muscle in which lactate is transferred from white-glycolytic to red-oxidative fibers. However, this situation in brain is quite different from skeletal muscle such that this shuttle may be less relevant under most normal circumstances. Indeed, the neuron has both the greatest energy demand and the capacity to utilize an array of potentially available substrates, including glucose, lactate, pyruvate, the ketone bodies (acetoacetate and β -hydroxybutyrate), and various amino acids, dependent on their relative availability. In support of the utilization of these monocarboxylic acids, the neuron also has the low- K_m and high-capacity MCT2 transporter. The actual ratio of glucose to lactate utilization by the neuron remains somewhat contentious, as is the source of the lactate (neurons or astrocytes) and the extent to which this ratio may be altered in activated versus basal state.

The glial component of the brain is comprised of astrocytes, oligodendrocytes, and microglia and the 45-kDa GLUT1 is the predominant glucose transporter isoform that supplies glucose to all of these cells ([Figure 3](#)). The primary energy-requiring

function of the astrocytes is the maintenance of the extracellular environment, including glutamate uptake for synthesis and recycling back to the neuron as glutamine, that is, the glutamate–glutamine shuttle, and the maintenance of K^+ , osmotic, and pH homeostasis. The astrocytes are the major, if not sole, neural cells that express the enzyme pyruvate carboxylase that is necessary for the *de novo* synthesis of glutamine, which is the precursor for the neurotransmitters, glutamate and γ -aminobutyric acid (GABA). In keeping with their role in maintaining cerebral homeostasis, the astrocyte is the primary site of glycogen synthesis and energy storage as glucosyl units. Thus, the predominant pathways of glucose utilization in the astrocyte are glycolysis to generate lactate for potential export, further oxidation or neurotransmitter biosynthesis via the TCA cycle, and glycogen synthesis (Figure 4). The provision of antioxidants, in the form of reduced glutathione, is also an important component of the homeostatic function of astrocytes and thus they also rely on the pentose phosphate pathway for the generation of reducing equivalents in the form of NADPH. The primary function of the oligodendrocyte is the elaboration and maintenance of the myelin sheath that encases and insulates the neuronal axon. Myelin, which is composed of 80% lipid and 20% protein, is synthesized from fatty acids and cholesterol and requires a supply of the reducing equivalents NADPH generated via the pentose phosphate pathway (Figure 4). The flux through this pathway is greatest during cerebral development and active myelination. It is also during development that the brain is most active in the uptake of the ketone bodies, β -hydroxybutyrate and acetoacetate, which serve as precursors in lipid synthesis, and are transported by MCT1. The third glial cell is the microglial cell, the resident immune cell of the brain. As such, microglia are the only antigen-presenting cells in the brain and initiate and orchestrate an inflammatory cascade in response to a variety of stimuli. Central to this inflammatory response is the generation of reactive oxygen species via a number of NADPH-requiring reactions. Thus, in addition to glycolysis and oxidative phosphorylation for the generation of ATP, the pentose–phosphate pathway is active in microglia as well. In addition to GLUT1, the microglia in human and rat brain express the GLUT5 transporter (Figure 3). GLUT5 is actually a fructose transporter and is active as such in peripheral cells, such as intestinal epithelium. However, there is no significant fructose in the brain; thus, the substrate and function of microglial GLUT5 remain to be determined.

All of the above-mentioned metabolic pathways depend on the delivery of glucose (or alternative substrate) from the circulation to the brain and this is the work of the microvascular endothelial cells that comprise the BBB. The primary energy-requiring functions of these cells are the fueling of the Na^+ - K^+ -ATPase components of the wide variety of active transport systems in the abluminal (brain-facing) membrane. The ATP required to support these transport systems is derived from the oxidative metabolism of glucose and potentially other substrates. The major nutrient transporters of the BBB endothelial cells are GLUT1, which exists at this site as a highly glycosylated 55-kDa form, and MCT1. Clearly, these transporters function both for the bidirectional transport of substrates across the two membranes of the cell, and for the provision of energetic substrate to the endothelial cell itself. Under normal circumstances in the adult brain, the supply of glucose is not limited by the transport capacity in the BBB. However, there are some

notable exceptions, which involve either increased demand and/or reduced supply, such as seizures, hypoglycemia, and cerebral ischemia or stroke. Genetic examples of transport limitation to cerebral glucose utilization are seen in the GLUT1 deficiency syndrome, in which affected individuals have only one allele for the GLUT1 glucose transporter and glucose delivery to the brain is significantly compromised. These individuals can be treated by the provision of an alternate energy substrate in the form of ketone bodies, generated by the ketogenic diet.

The above discussion of cerebral glucose utilization applies to the normal adult brain of a range of mammalian species from rodent to human. Although glucose is also the predominant energetic fuel for the immature brain, there are notable distinctions in both energy demand and substrate supply that characterize cerebral development. The energetic demands of the newborn brain are significantly lower than the adult, as are rates of cerebral glucose utilization, approximating only 10–20% of the adult value at birth in the rat to 50–70% for the human. Since the majority of cerebral glucose utilization is to provide the energy to maintain the resting membrane potential, the magnitude – and regional variation – in glucose utilization is directly related to the extent of connectivity of the brain in the resting state, or the number of active synapses. Thus, energy demand, cerebral blood flow, and substrate (glucose) utilization increase as neurons undergo synaptogenesis and form active connections, which is largely postnatal in the rat but begins prenatally in the human infant. During the early stages of development, energy demand increases coincident with the acquisition of new skills, all requiring the successful connection of neural networks. Measurements of cerebral glucose utilization during human development demonstrate a steep increase from birth to approximately 4 years of age, at which time the values are roughly twice those seen in healthy young adults, and this is roughly maintained over the next 4–5 years, before declining to the adult level. This is because during normal development there is an overproduction of neurons, synaptic connections, as well as neurotransmitters, which decline during the period of synaptic pruning and continues into adolescence. Comparable studies in developing rodents have not observed this pattern of cerebral glucose utilization, but instead demonstrate a steady increase from birth to adulthood. However, measurements of cerebral blood flow and energy utilization in developing rats do reach peak levels at P17 and P14, respectively, reflecting the period of increased activity/synaptic pruning at the end of the period of active brain growth. The discrepancy between the human and rodent data relates to the greater importance of circulating ketone bodies, β -hydroxybutyrate and acetoacetate, as energetic fuels to the rodent brain. Although both human and rodent milk are high in fat and generate a significant level of circulating ketone bodies in the suckling offspring to support biosynthetic and energetic pathways, the levels attained in the rodent are greater and the increased energy utilization is met by a combination of glucose and ketone bodies. The delivery of ketone bodies to the developing brain is facilitated by very high levels of MCT1 in the microvascular endothelial cells of the BBB, which also reach peak levels at P14–17 before declining to adult levels after suckling. The primary adaptation to the increased utilization of ketone bodies appears to be at the

level of the BBB, as the nonvascular levels of both MCT1 and MCT2 in the glial and neuronal cells, respectively (Figure 3), do not decrease after suckling, likely reflecting ongoing metabolic trafficking between cells in the mature brain. The other substrate that is of energetic importance to the neonatal brain is lactate, especially during the first hours to days of life in both humans and rodents, and the high levels of BBB MCT1 will facilitate delivery of this substrate as well. However, by P15 in the rodent, glucose has become the predominant energetic fuel accounting for 57% of total energy supply, compared with 23% and 20% for lactate and β -hydroxybutyrate, respectively.

Summary

The brain is a highly energy-demanding organ that consumes a seemingly disproportionate amount of energy of the whole body. It differs from other organs in the body because of the following:

1. it is highly oxidative and uses glucose as its primary energy source as it is unable to access circulating triglycerides and fatty acids;
2. it has a limited capacity to store energy and does so in the form of glycogen;
3. the basal rate of metabolism is high and thus a continual supply of glucose is necessary to sustain activity;
4. because basal cerebral energy utilization is high, activation produces only twofold increases in metabolic activity, in contrast to muscle where 50-fold increases are observed upon activation, even though the activated metabolic rates are comparable between the two tissues;
5. the brain is a heterogeneous tissue and the various cells that make up the brain have markedly different metabolic demands reflecting their differing cellular activities; and
6. restoration of neuronal membrane potentials represents the major energy demand and the neurons are preferentially equipped with the nutrient transporters GLUT3 and MCT2, which have the highest affinity for the available energy substrates – glucose, lactate, or ketoacids.

See also: **Metabolism Vitamins and Hormones: Carbohydrate-Responsive Element Binding Protein; Glucose/Sugar Transport in Mammals; Glycolysis Overview.**

Further Reading

- Attwell D and Laughlin SB (2001) An energy budget for signaling the grey matter of the brain. *Journal of Cerebral Blood Flow and Metabolism* 21: 1133–1145.
- Aubert A, Pellerin L, Magistretti PJ, and Costalat R (2007) A coherent neurobiological framework for functional neuroimaging provided by a model integrating compartmentalized energy metabolism. *Proceedings of the National Academy of Sciences of the United States of America* 104: 4188–4193.
- Chih CP and Roberts EL, Jr. (2003) Energy substrates for neurons during neural activity: A critical review of the astrocyte–neuron lactate shuttle hypothesis. *Journal of Cerebral Blood Flow and Metabolism* 23: 1263–1281.
- Chugani HT (1998) A critical period of brain development: Studies of cerebral glucose utilization with PET. *Preventive Medicine* 27: 184–188.
- Dienel GA and Hertz L (2001) Glucose and lactate metabolism during brain activation. *Journal of Neuroscience Research* 66: 824–838.
- DiNuzzo M, Mangia S, Maraviglia B, and Giove F (2010) Changes in glucose uptake rather than lactate shuttle take center stage in subserving neuroenergetics: Evidence from mathematical modeling. *Journal of Cerebral Blood Flow and Metabolism* 30: 586–602.
- Halestrap AP and Meredith D (2004) The SLC16 gene family – from monocarboxylate transporters (MCTs) to aromatic amino acid transporters and beyond. *Pflügers Archiv* 447: 619–628.
- Hertz L and Dienel GA (2002) Energy metabolism in the brain. *International Review of Neurobiology* 51: 1–102.
- Magistretti PJ, Pellerin L, Rothman DL, and Shulman RG (1999) Energy on demand. *Science* 283: 496–497.
- Mangia S, Simpson IA, Vannucci SJ, and Carruthers A (2009) The in vivo neuron-to-astrocyte lactate shuttle in human brain: Evidence from modeling of measured lactate levels during visual stimulation. *Journal of Neurochemistry* 109(supplement 1): 55–62.
- Manolescu AR, Witkowska K, Kinnaird A, Cessford T, and Cheeseman C (2007) Facilitated hexose transporters: New perspectives on form and function. *Physiology (Bethesda)* 22: 234–240.
- Pierre K and Pellerin L (2005) Monocarboxylate transporters in the central nervous system: Distribution, regulation and function. *Journal of Neurochemistry* 94: 1–14.
- Seidner G, Alvarez MG, Yeh JI, et al. (1998) GLUT-1 deficiency syndrome caused by haploinsufficiency of the blood–brain barrier hexose carrier: Imaging the metabolic footprint of Glut1 deficiency on the brain. *Nature Genetics* 18: 188–191.
- Simpson IA, Carruthers A, and Vannucci SJ (2007) Supply and demand in cerebral energy metabolism: The role of nutrient transporters. *Journal of Cerebral Blood Flow and Metabolism* 27: 1766–1791.
- Sokoloff L, Reivich M, Kennedy C, et al. (1977) The [^{14}C]deoxyglucose method for the measurement of local cerebral glucose utilization: Theory, procedure, and normal values in the conscious and anesthetized albino rat. *Journal of Neurochemistry* 28: 897–916.
- Vannucci SJ, Maher F, and Simpson IA (1997) Glucose transporter proteins in brain: Delivery of glucose to neurons and glia. *Glia* 21: 2–21.
- Vannucci SJ and Simpson IA (2003) Developmental switch in brain nutrient transporter expression in the rat. *American Journal of Physiology – Endocrinology and Metabolism* 285: E1127–E1134.
- Wood IS and Trayhurn P (2003) Glucose transporters (GLUT and SGLT): Expanded families of sugar transport proteins. *British Journal of Nutrition* 89: 3–9.

Carbohydrate-Responsive Element Binding Protein

K Uyeda, University of Texas Southwestern Medical Center & Veteran Affairs Medical Center, Dallas, TX, USA

H Yamashita, Okayama Prefectural University, Okayama, Japan

© 2013 Elsevier Inc. All rights reserved.

Glossary

Carbohydrate response element A DNA sequence within the promoter of a gene to which a specific transcription factor binds in response to excess carbohydrate.

Liver pyruvate kinase An isozyme of pyruvate kinase that occurs mainly in liver.

Transcriptional Activation of the Liver Pyruvate Kinase and the Lipogenic Enzyme Genes

Liver pyruvate kinase (LPK) catalyzes the final step of glycolysis, and generates pyruvate and adenosine triphosphate (ATP) from phosphoenolpyruvate (PEP) and adenosine diphosphate (ADP). Pyruvate is then converted into citrate via acetyl-CoA in the mitochondria and citrate is subsequently metabolized back into acetyl-CoA by ATP-citrate lyase in the cytoplasm. The cytoplasmic acetyl-CoA serves as a substrate for triglyceride synthesis.

The genes encoding LPK and the lipogenic enzymes are upregulated in hepatocytes by a high-carbohydrate diet. A short 88-bp segment (−183 to −97) upstream from the transcription initiation site of the LPK gene, referred to as the major late transcription factor (MLTF) site or glucose responsive element, is sufficient to allow glucose-stimulated induction of the promoter. The MLTF site consists of two CACGGG sequences (E-boxes) separated by five bases which are similar to the consensus-binding site (CACGTG) for the c-Myc and upstream stimulating factor (USF) of transcription factor families. Similar elements have been found in the promoter sites of acetyl-CoA carboxylase (ACC) and fatty acid synthetase (FAS) genes. Many other lipogenic genes likely contain the carbohydrate-responsive element (ChRE) within their promoter sites or other regulatory regions that mediate glucose response. The MLTF sequences have been used to identify and characterize ChRE-binding proteins (ChREBPs), but the identity of the carbohydrate-responsive transcription factor remained elusive for a number of years.

Discovery, Purification, and Identification of a New Transcription Factor that Binds to ChRE of the LPK Gene

The identification and purification of ChREBP, a goal of investigators that remained elusive for decades, was finally overcome using an innovative purification scheme. Conventional nuclear extract preparation methods, which had been widely used, were ineffective. When nuclear extracts were instead prepared using polyethylene glycol, which protected the transcription factor against denaturation and allowed for direct assays of DNA-binding activity, ChREBP was readily detected (Figure 1). Subsequent studies indicated that the expression pattern, regulation, and selectivity of the factor were all consistent with its

assignment as a carbohydrate-responsive transcription factor. ChREBP is activated by a high-carbohydrate diet, but not by high-fat diet or starvation. Its DNA-binding specificity under varying glucose concentrations correlates exactly with LPK promoter activity measured *in vivo*. ChREBP activity is highest in liver nuclei under high glucose conditions, coincident with the expression pattern of the LPK target gene under these conditions. These lines of evidence in summary are consistent with its assignment as cChREBP.

Species and Tissue Distribution of ChREBP and a Human Disease Candidate

ChREBP is highly conserved among human, rat, and mouse with over 80% amino acid sequence identity. Orthologs are also present in *Drosophila* and *Caenorhabditis elegans*. The tissue distribution of ChREBP messenger RNA (mRNA) in mammals in decreasing order is liver, adipose, islets, intestine, kidney, cardiac muscle, skeletal muscle, and brain. A candidate human disease involving ChREBP is the Williams–Beuren syndrome, a disorder due to a chromosomal deletion that removes 17 genes including ChREBP. This syndrome, also known as WBSR 14, is a neurodevelopmental disorder characterized by cardiac and vascular defects, dysmorphic facial features, and mental retardation.

Structure and Function of ChREBP

ChREBP is a large protein (Mr of mouse ChREBP = 94 600 Da) consisting of 864 amino acids. The transcription factor contains several domains including polyproline regions, a basic helix–loop–helix and leucine zipper motif (bHLH/Zip), and a zip domain. In addition, there is a nuclear localization signal motif located near the amino terminus as well as multiple phosphorylation sites (Figure 2). The ChREBP molecule can be divided into two functional domains: (1) the amino-terminal region (amino acids 1–250) containing nuclear/cytosol localization signal sites, two nuclear export signal sites (NES1 and NES2), and a nuclear localization signal (NLS) site, which are responsible for the subcellular localization of ChREBP, as well as glucose sensing capacity; and (2) the carboxy-terminal region of ChREBP containing a bHLH/Zip motif which is responsible for DNA-binding and transcription activity.

ChREBP senses circulating glucose level, and localizes either in the cytoplasm under low glucose conditions, as in the starved state, or in the nucleus under high glucose conditions, as in the fed state. However, the mechanism by which ChREBP senses the glucose level (glucose signaling) is not known, and is a focus of current investigations. A metabolite(s) of glucose, rather than glucose itself, serves as signaling compounds for ChREBP, since non-metabolizable carbohydrates cannot serve in the glucose signaling. For export out of the nucleus under low glucose conditions, ChREBP interacts with

a dimeric form of 14-3-3 protein and forms a heterodimer, which then binds to CRM1 (exportin) and translocates in the cytoplasm. In response to high glucose levels, ChREBP interacts with Max-like protein X (Mlx) to form a heterodimer of ChREBP-Mlx. ChREBP itself or a ChREBP-Mlx heterodimer then complexes with importin α and this complex binds to importin β , allowing for localization to the nucleus and transcriptional initiation of lipogenic genes.

Regulation of ChREBP Activity

Glucose and glucagon are opposing forces in maintaining glucose homeostasis by controlling glucose metabolism via glycolysis, gluconeogenesis, and glycogenolysis/glycogenesis pathways. The regulation of these pathways involves both short-term and long-term mechanisms. The short-term regulation involves controls of key regulatory enzymes by allosteric mechanisms involving metabolites and small ligands, and also includes control mediated by covalent modifications. The long-term regulation of these pathways involves altering the expression of genes encoding these same regulatory enzymes.

One mechanism for regulating ChREBP activity is by post-translational modifications. ChREBP contains multiple phosphorylation sites, whose status can be affected by various hormonal and nutrient stimuli. For example, in cell culture models, ChREBP can be phosphorylated at Ser-196, Ser-626, and Ser-666 by cyclic adenosine monophosphate (cAMP)-dependent protein kinase (PKA), which results in the inhibition of nuclear import and DNA-binding/transcriptional activities. In response to the rise in adenosine monophosphate (AMP), AMP-kinase-mediated modification of Ser-568 also inhibits DNA binding. However, the exact *in vivo* role of these phosphorylation sites, as well as the location of *bona fide* phosphorylation sites, in response to various stimuli is unknown. Nevertheless, phosphorylation of ChREBP renders it inactive.

In contrast to glucagon action, a rise in circulating glucose levels stimulates glucose metabolism and promotes the formation of xylulose-5-phosphate (Xu5P), an intermediate of the pentose phosphate shunt pathway. Xu5P specifically activates protein phosphatase 2A δ (PP2A δ) to dephosphorylate an inactive/phosphorylated-ChREBP, and thereby, activates nuclear localization as well as DNA binding of ChREBP (Figure 3). Xu5P-mediated glucose signaling was first discovered in the dephosphorylation/activation catalyzed by the same protein phosphatase (PP2A δ) of the phosphorylated/

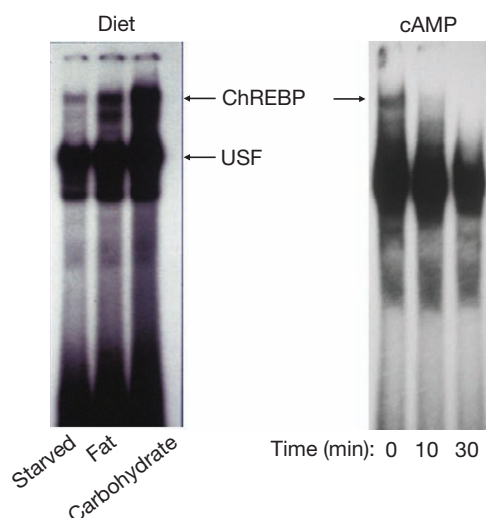


Figure 1 Response of ChREBP to dietary content or cyclic-AMP. (a) Gel-shift analysis of hepatic nuclear factors that recognize the LPK carbohydrate responsive element (ChRE). Rats were either starved (lane 1), or fed a high-fat diet (lane 2) or a high-carbohydrate diet (lane 3). Nuclear extracts were prepared from rat livers and incubated with a radiolabeled oligonucleotide corresponding to LPK-ChRE. Lanes 2 and 3 were equalized for protein loads. Arrows indicate the positions of the DNA-binding complexes containing USF or ChREBP. (b) Effect of cAMP on the DNA-binding activity of ChREBP *in vivo*. Rats were fed a high-carbohydrate diet after starvation, and db-cAMP was then injected via the intraperitoneal method. Rats were sacrificed at 10 and 30 min after the injection, and the DNA-binding activity of ChREBP was examined using a gel mobility shift assay. With permission from Yamashita, H., Takenoshita, M., Sakurai, M., Bruick, R.K., Henzel, W.J., Shillinglaw, W., Arnot, D., and Uyeda, K., (2001) A Glucose-Responsive Transcription Factor that regulates Carbohydrate Metabolism in the liver. *Proc Natl. Acad. Sci.* 98, 9116–9121. National Academy of Sciences, USA.

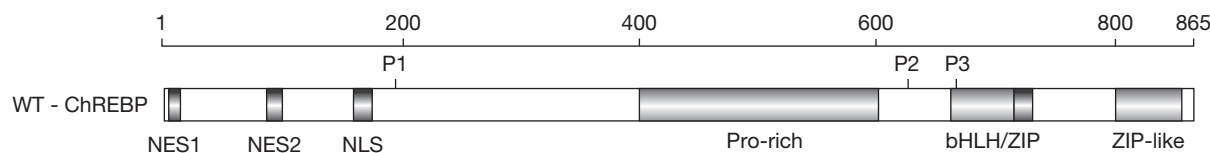


Figure 2 Schematic representation of ChREBP functional domains. Domains that are present include two nuclear export signal (NES) sites, a nuclear localization signal (NLS) site, a basic helix-loop-helix/Zipper DNA-binding domain, and a proline-rich domain implicated in protein-protein interactions. The amino acids designated P1 (Ser-196) and P2 (Thr-666), which are the residues phosphorylated by cAMP-protein kinase whereas P3 (Ser-568) is the amino acid phosphorylated by AMP-dependent protein kinase. The amino-terminal amino acids 1–251 are responsible for glucose-sensing and nuclear import/export. The carboxy-terminal bHLH/Zip region interacts with Mlx to form a heterodimer, which binds to DNA. With permission from Fukasawa, M., Ge, Q., Wynn, M.W., Ishii, S., and Uyeda, K. (2010) Coordinate regulation/localization of the carbohydrate responsive binding protein (ChREBP) by two nuclear export signal sites. *Discovery of a new leucine-rich nuclear export signal site. Biochem. Biophys. Res. Commun.* 391, 1166–1169.

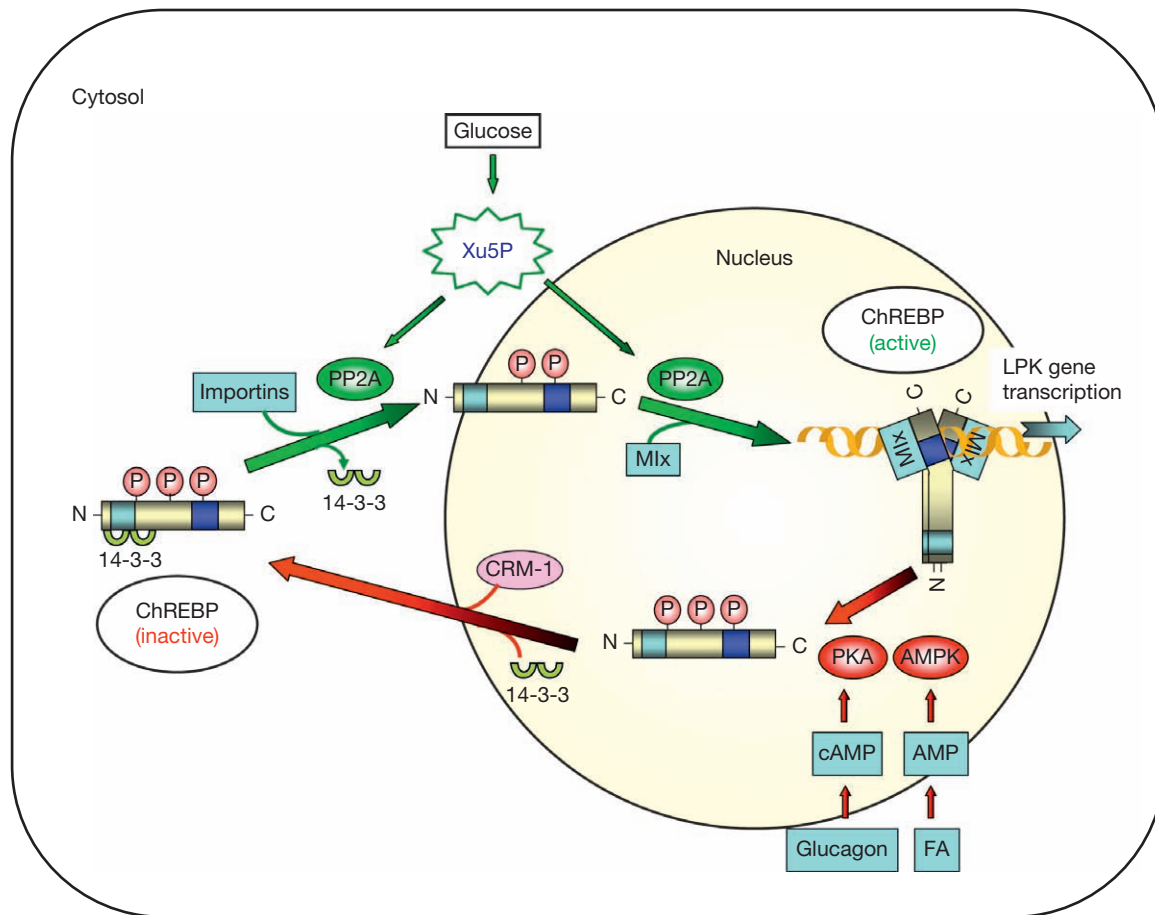


Figure 3 Stimulus-responsive regulation of ChREBP activity. Phosphorylated ChREBP (inactive form) binds to a 14-3-3 protein, complexes with CRM-1, and then is exported out of the nucleus to the cytosol (red arrows). Exposure of hepatocytes to high glucose conditions results in an increased xylulose 5-phosphate (Xu5P) concentration, the stimulus for glucose signaling, which activates protein phosphatase 2A (PP2A). PP2A dephosphorylates ChREBP (active form) and the dephosphorylated ChREBP subsequently binds to importin α and importin β (green arrows), thereby promoting nuclear translocation. Nuclear ChREBP is further dephosphorylated by PP2A, which promotes DNA binding of ChREBP to the ChREs in the promoters of the ChREBP target genes (LPK and lipogenic enzyme genes). In response to glucagon and high fatty acids, PKA and AMPK activities are increased, resulting in phosphorylation of ChREBP. Phosphorylated ChREBP exhibits decreased DNA-binding activity and is actively exported out of the nucleus back into the cytoplasm (red arrows).

inactive bifunctional enzyme, 6-phosphofructo-2-kinase/fructose-2,6-P₂-bisphosphatase (F6P2K/F26Pase; **Figure 3**).

Fatty acids, especially polyunsaturated fatty acids (PUFAs), are potent inhibitors of glycolysis and lipogenesis in the liver. The molecular mechanism for these observations became evident after the discovery of ChREBP, which stimulates the gene expression of LPK and nearly all lipogenesis enzymes. Elevated PUFA levels cause decrease in glycolysis as well as decreased flux through the pentose shunt pathway, and therefore results in decreased Xu5P concentrations. The reduction in Xu5P leads to increased ChREBP phosphorylation, due to the reduced action of PP2A δ , and consequently inhibition of ChREBP nuclear localization and low transcriptional activity. More recently, it was shown in cell lines that PUFA, but not saturated fatty acids, also accelerates ChREBP mRNA decay and thereby decreases ChREBP protein levels. However, the total amount of ChREBP protein usually does not change in primary hepatocytes or in liver upon feeding fatty acids, although the subcellular distribution of ChREBP may be affected.

An alternative mechanism to explain fatty acid inhibition of ChREBP activities involves AMP-activated protein kinase (AMPK). The first step in short-chain, medium-chain, and long-chain fatty acid metabolism is the activation of fatty acids to fatty acyl-CoA and AMP, a reaction that is catalyzed by acyl-CoA synthetase. During this activation reaction, a 30-fold increase in free AMP activates AMPK, which in turn phosphorylates S568 of ChREBP and thereby inactivates ChREBP DNA-binding activity. The AMP-mediated fatty acid inhibition of ChREBP may be distinct from that of PUFA and may play a functional role under different physiological conditions.

Genetic Ablation Studies of ChREBP-Gene in Mice Reveal the Biological Role of ChREBP

ChREBP^{-/-} mice are viable and have a normal life span. However, deletion of ChREBP decreases gene expression of LPK and all lipogenic enzymes including ATP-citrate lyase, acetyl-CoA

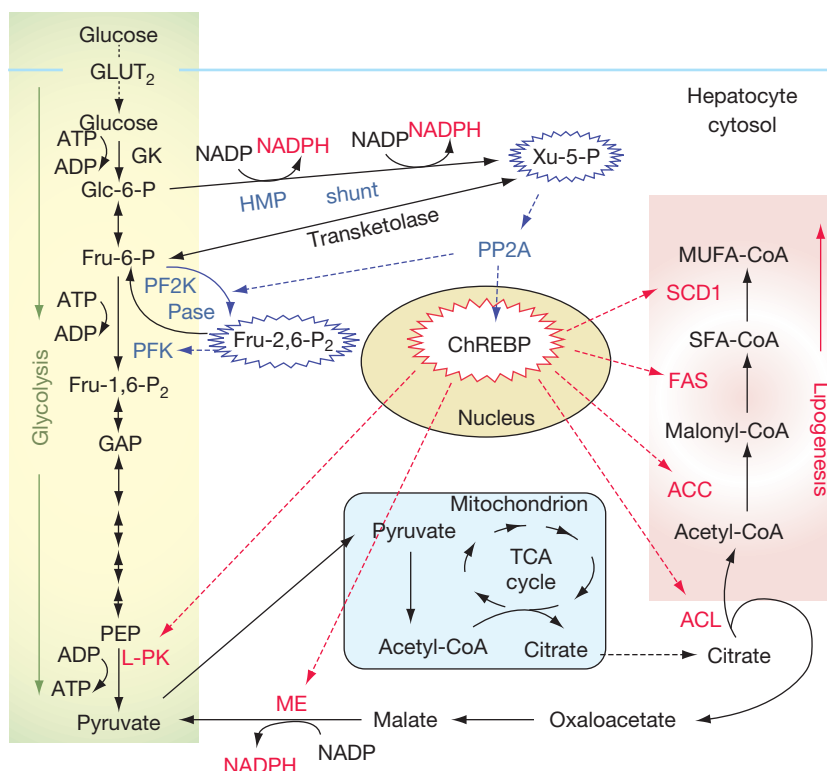


Figure 4 The central role of ChREBP in coupling glycolysis and lipogenesis in the liver. Glucose metabolism yields pyruvate in the final step of glycolysis. Pyruvate is then oxidized in the mitochondria to produce citrate. After citrate is transported to the cytoplasm, it is cleaved to regenerate acetyl-CoA, the substrate for fatty acid synthesis. The key regulatory enzyme in glycolysis, PFK, is activated by fructose-2-P₂ (Fru-2,6-P₂), which is synthesized (and also hydrolyzed) by 6-phosphofructo-2-kinase:fructose 2,6-phosphatase (PF2K-Pase). This bifunctional enzyme is activated by the same glucose signaling mechanism used by ChREBP, activation of PP2A by Xu5P. The final step of pyruvate production in glycolysis is controlled by pyruvate kinase (LPK). The long-term regulation of both glycolysis and fatty acid synthesis, mediated through regulation of LPK and of all lipogenesis enzyme genes respectively, is controlled by ChREBP (shown in red). Production of Xu5P is involved in both glucose signaling and the production of reducing equivalents (NADPH, produced by malic enzyme (ME)), as well as for fat synthesis. With permission from Uyeda, K., and Repa, J.J., (2006) Carbohydrate response element binding protein, ChREBP, a transcription factor coupling hepatic glucose utilization and lipid synthesis. *Cell Metabolism* 4, 107–110.

carboxylase 1, fatty acid synthase, stearyl-CoA desaturase, and LCE elongase (Figure 4). Interestingly, malic enzyme, which generates reduced nicotinamide adenine dinucleotide phosphate (NADPH) required for fat synthesis, is the most significantly reduced lipogenic enzyme. In addition, the expression of a glucose transporter gene (*Glut2*) is significantly reduced. These observations suggest that ChREBP coordinately regulates glucose utilization and fatty acid synthesis by tightly coupling the availability of the substrate, acetyl-CoA, for fat synthesis and the activities of lipogenic enzymes.

ChREBP deficiency results in less severe glucose intolerance, perhaps due to repressed glycolysis. Decreased expression of key gluconeogenic enzymes, glucose-6-phosphatase and PEP carboxykinase, in ChREBP^{-/-} mice suggests that increased glucose and glycogen in these animals does not result from increased gluconeogenesis. ChREBP^{-/-} mice are fructose intolerant because of reduced gene expression of two enzymes involved in the fructose metabolism pathway, namely fructose kinase and triose kinase. Due to these coordinated changes in carbohydrate metabolism and lipogenesis, it is not surprising that the rate of hepatic fatty acid synthesis in intact ChREBP^{-/-} mice is reduced by 65% compared to wild-type age- and gender-matched littermates.

The hepatic energy metabolism of ChREBP^{-/-} mice is also decreased compared to wild-type mice. The pyruvate dehydrogenase (PDH) complex is significantly more active in the ChREBP^{-/-} liver because of reduced PDH kinase (PDK3), which inhibits PDH activity. Greater PDH activity causes stimulation of pyruvate/lactate oxidation and therefore decreased fatty acid oxidation in perfused liver of ChREBP^{-/-} mice. This shift in the substrate in mitochondrial substrate for oxidation eventually leads to a threefold reduction in free cytosolic nicotinamide adenine dinucleotide/nicotinamide adenine dinucleotide phosphate (NAD/NADH) ratio, and a 1.7-fold increase in the free mitochondrial NAD/NADH ratio. The net result of the shift from fat to pyruvate/lactate utilization is a decreased energy state in the livers of ChREBP^{-/-} mice.

ChREBP^{-/-} Mice Crossed with Leptin-Deficient ob/ob Mice

As a result of the intercross of ChREBP^{-/-} mice with ob/ob mice, expression of all lipogenic enzymes was significantly lower compared to ob/ob animals. This results in decreased

fat synthesis and normalization of plasma free fatty acids and triglycerides. The overall weight gain was reduced in the doubly deficient ob/ob:ChREBP^{-/-} mice, which was attributed to decreased food intake and possibly from reduced expression of a hypothalamic appetite-stimulating neuropeptide. Activity of gluconeogenic enzymes also was lower in the doubly deficient mice contributing to lower blood glucose. These results suggest that inactivation of ChREBP reduces fat synthesis and obesity, and also improved glucose tolerance and appetite control, in mice genetically prone to development of diabetes.

Summary

ChREBP is required for both basal- and carbohydrate-induced expression of a number of liver enzymes essential for coordinated regulation of glucose metabolism, fatty acids, and the synthesis of fatty acids and triglycerides *in vivo*. Its regulation involves modifications induced after stimulation of a number of signal transduction pathways in response to dietary and other stimuli. The pivotal role that ChREBP plays in intermediary metabolism places it in a privileged position shared by only a handful of transcriptional regulators. Future advances in our understanding of the molecular and biochemical mechanisms responsible for activation and repression of

ChREBP will have important translational implications for the treatment of human disease states.

See also: **Metabolism Vitamins and Hormones:** Carbohydrate Metabolism in the Central Nervous System; Diabetes; Fatty Acid Structure and Synthesis; Fatty Acyl-CoA Synthetases; Glycolysis Overview; Pentose Phosphate (Hexose MonoPhosphate) Pathway; Pyruvate Kinase; Structure and Regulation of Pyruvate Dehydrogenases.

Further Reading

- Kabashima T, Kawaguchi T, Wadzinski BE, and Uyeda K (2003) Xylulose 5-phosphate mediates glucose-induced lipogenesis by xylulose 5-phosphate-activated protein phosphatase in rat liver. *Proceedings of the National Academy of Sciences of the United States of America* 100: 5107–5112.
- Postic C, Dentin R, Denechaud PD, and Girard J (2007) ChREBP, a transcriptional regulator of glucose and lipid metabolism. *Annual Review of Nutrition* 27: 179–192.
- Uyeda K and Repa JJ (2006) Carbohydrate response element binding protein, ChREBP, a transcription factor coupling hepatic glucose utilization and lipid synthesis. *Cell Metabolism* 4: 107–110.
- Yamashita H, Takenoshita M, Sakurai M, et al. (2001) A glucose-responsive transcription factor that regulates carbohydrate metabolism in the liver. *Proceedings of the National Academy of Sciences of the United States of America* 98: 9116–9121.

Cardiac Excitation–Contraction Coupling

D M Bers and S Despa, University of California at Davis, Davis, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Channel inactivation Reduction in current amplitude under sustained depolarization.

Electrochemical gradient The difference in ion concentration and electrical potential from one point to another, so that ions tend to move passively along it.

Fibrillation Incoordinated twitching of the heart muscle fibers.

Inward rectifying channel A channel that conducts inward current (positive charge flows into the cell) better than outward current.

Open probability of a channel The fraction of time the channel stays in the open conformation.

Tachycardia A faster than normal heart rate that occurs when an abnormality in the heart produces rapid electrical signals.

Cardiac Excitation–Contraction Coupling: Interplay between Electrical Activity of the Cell, Cytosolic Ca, and the Myofilaments

Cardiac excitation–contraction (E–C) coupling is the process from electrical excitation of the cardiac muscle to contraction, resulting in the pumping of the blood out of the heart. The normal heartbeat is initiated by an electrical impulse that originates in the pacemaker cells of the sino-atrial node. This electrical impulse then propagates as a wave of depolarization through the atrium and atrioventricular (AV) node and reaches the cardiac myocytes in the ventricle. Here, it triggers an action potential (AP), that is, membrane depolarization that follows a finely tuned waveform. The AP then initiates a complex cascade of events culminating in cell contraction. The ubiquitous second-messenger Ca plays a crucial role in these events as it is the direct activator of the myofilaments, the contractile apparatus of the cardiac myocytes. Cytosolic Ca also participates in shaping the AP. Thus, cardiac E–C coupling is a finely orchestrated interplay between electrical activity of the cell, cytosolic Ca, and the myofilaments.

Cardiac AP

When the wave of depolarization generated in the sino-atrial node reaches a ventricular myocyte, voltage-dependent Na channels open, resulting in the inward flux of Na ions down their electrochemical gradient. This Na current (I_{Na}) causes the upstroke (or phase 0) of the AP (Figure 1). The rapid upstroke of the AP is primarily due to the regenerative activation of I_{Na} ; that is, once the membrane potential reaches a threshold voltage in the -70 to -60 mV range, the activation of Na channels leads to further depolarization, which activates additional Na channels and depolarization and so on. This positive feedback creates a very rapid depolarization (200 V s^{-1}) toward the equilibrium potential for Na. In cardiac myocytes, the peak of the AP typically reaches $+30$ mV or more. Depolarization stops in part because Na channels inactivate rapidly (within a few milliseconds). This constitutes a built-in negative feedback on I_{Na} , which limits the gain in intracellular Na to $6\text{--}15 \mu\text{mol l}^{-1}$. Voltage-gated K channels also open during depolarization, but with some latency. Opening of a special type of K channel, which carries a

short-lived, voltage-gated outward K current called ‘transient outward current’ (I_{to}), is responsible for the early rapid repolarization (the notch or phase 1 of the AP; Figure 1).

During depolarization, Ca enters cardiac myocytes through voltage-dependent Ca channels (mainly long-lasting, or L-type channels) which generates an inward Ca current (I_{Ca}). I_{Ca} activation, combined with the transient nature of I_{to} , delays the repolarization and results in a plateau phase in the AP (phase 2; Figure 1). This plateau phase prolongs the AP duration and distinguishes cardiac APs from the much shorter APs found in neurons and skeletal muscle. The AP in rat and mouse ventricular myocytes has almost no plateau phase because I_{to} is much larger than in other species and consequently there is much greater early repolarization. During the plateau phase present in most mammals other than rat and mouse, the inward and outward currents are nearly balanced. The outward current is carried by delayed rectifier K channels (known as I_{Kur} , I_{Ktr} , and I_{Ks}). At some point, repolarization accelerates greatly and the membrane potential returns to the diastolic level (late rapid repolarization or phase 3 of the AP, Figure 1). Repolarization is favored by inactivation of L-type Ca channels, mainly through a mechanism dependent on Ca concentration at the cytosolic side. This Ca-dependent inactivation is a very local effect and is mediated by calmodulin bound to the carboxyl-terminal of the Ca channel protein. I_{Kr} is particularly important in the accelerating phase of late repolarization. As repolarization proceeds, I_{Kr} deactivates at more negative potentials. Below -30 mV, another K current, the inward rectifying current (I_{K1}), increases with repolarization, again accelerating repolarization. I_{K1} is active at resting membrane potentials and is responsible for stabilizing the potential at about -80 mV until the next wave of depolarization reaches the myocyte (phase 4 of the AP; Figure 1). This resting membrane conductance via I_{K1} channels is responsible for the very negative membrane potential seen in cardiac (and skeletal muscle) myocytes versus other cell types (including sino-atrial and AV node cells in the heart and smooth muscle myocytes and neurons).

Cytosolic Ca and Contraction during Cardiac E–C Coupling

L-Type Ca channels are located primarily in invaginations of the cellular membrane called ‘T-tubules’, where they are

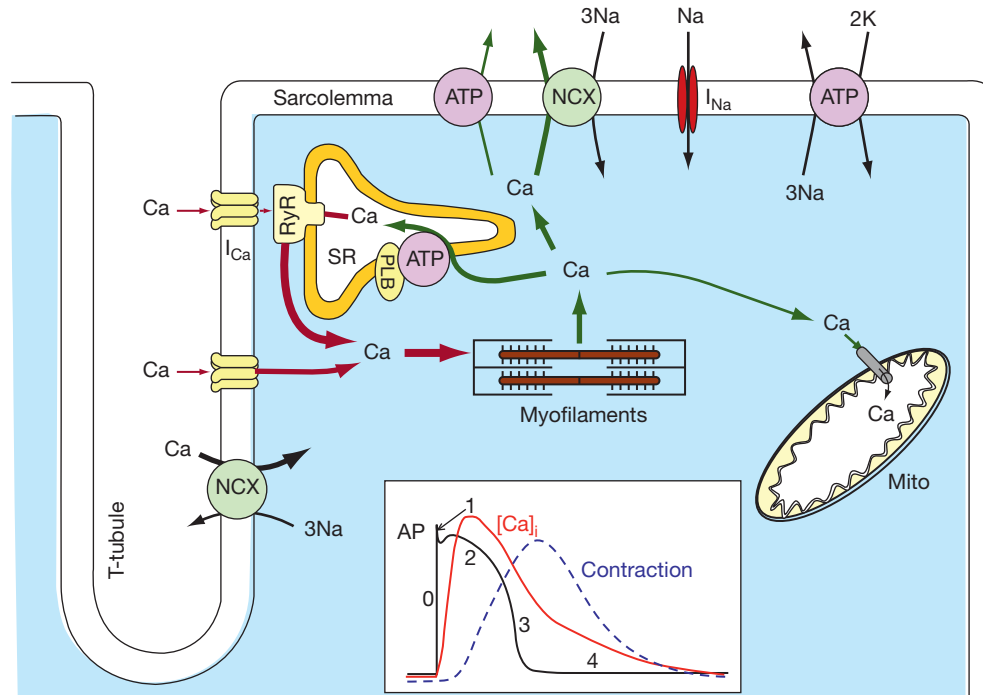


Figure 1 E–C coupling in cardiac ventricular myocytes. Inset shows the time course of an action potential (AP), Ca transient ($[Ca]_i$), and contraction measured in a rabbit ventricular myocyte at 37 °C. SR, sarcoplasmic reticulum; NCX, Na/Ca exchanger; RyR, ryanodine receptor; ATP, ATPase; PLB, phospholamban; I_{Ca} , L-type Ca channels current; I_{Na} , voltage-dependent Na current. Modified from Bers DM (2001) *Excitation–Contraction Coupling and Cardiac Contractile Force*, 2nd edn. Dordrecht: Kluwer.

physically close to Ca channels from the apposing sarcoplasmic reticulum (SR) membrane, the ryanodine receptors (RyRs; Figure 1). The open probability of RyRs depends on cytosolic $[Ca]$. Ca activation begins at submicromolar $[Ca]$, reaches a broad maximum near $100 \mu\text{mol l}^{-1}$ and decreases at very high $[Ca]$ ($5\text{--}10 \text{ mmol l}^{-1}$). Ca entering through I_{Ca} during the plateau phase of the AP locally increases $[Ca]$ in the restricted space between the T-tubules and junctional SR to levels as high as $10 \mu\text{mol l}^{-1}$. This is enough to activate the RyRs and trigger Ca release from the SR, a mechanism called ‘Ca-induced Ca release’. SR Ca release contributes to Ca-dependent inactivation of I_{Ca} . For example, total Ca influx via I_{Ca} is reduced by $\sim 50\%$ (from ~ 12 to $\sim 6 \mu\text{mol Ca/l cytosol}^{-1}$) when SR Ca release occurs. Thus, SR Ca release and I_{Ca} create local negative feedbacks on Ca influx. When there is high Ca influx or release, further Ca influx is turned off. The amount of Ca released from the SR also depends on the SR Ca content. Higher SR Ca load not only intrinsically increases Ca flux rate through the RyRs, but also greatly enhances the fraction of SR Ca that is released for a given $\sqrt{Ca}$ trigger. This is due, at least in part, to a stimulatory effect of high intra-SR free Ca concentration on the open probability of RyRs. At moderately low SR Ca content, I_{Ca} can fail to induce SR Ca release. This may help the SR reload if it becomes relatively depleted. Indeed, low SR Ca release allows more Ca influx via I_{Ca} (less inactivation) and reduces Ca extrusion through Na/Ca exchanger (see below). A decline in SR Ca load, even locally, may contribute dynamically to the turnoff of SR Ca release during E–C coupling. SR Ca content can be raised by increasing Ca influx, decreasing Ca efflux, or enhancing SR Ca uptake (e.g., by adrenergic stimulation or increase

in stimulation frequency, AP duration, I_{Ca} , or intracellular Na level).

Ca-induced Ca release is inherently a positive feedback mechanism, but its turnoff is essential for diastolic refilling of the heart. Two mechanisms contribute to the termination of the SR Ca release: SR Ca depletion (since the concentration of free Ca in the SR modulates RyR gating, see above) and RyR inactivation (or adaptation). There are two types of RyR inactivation, and both depend on intracellular Ca. One is an absorbing inactivation (as in the case of Na channels) where the RyR is unavailable for reopening until it recovers. The second type is called adaptation, in which the RyR relaxes after activation to a lower open probability, but can still be reactivated by a higher $[Ca]_i$. RyR inactivation results in refractoriness in cellular and local SR Ca release events and may be important in minimizing inappropriate SR Ca release events between heartbeats.

The combination of Ca influx and release raises free intracellular $[Ca]$ ($[Ca]_i$), allowing Ca binding to the myofilament protein troponin C, the process that switches on the contractile machinery of the myocyte. The dependence of the contraction force on $[Ca]_i$ is highly nonlinear due to high myofilament cooperativity with respect to $[Ca]_i$. The physiological contraction generates both isometric force (or ventricular pressure) and rapid shortening (to eject blood) when the intraventricular pressure exceeds the aortic pressure. There are two main ways to change the strength of cardiac contraction: (1) by altering the amplitude or duration of the Ca transient and (2) by altering the sensitivity of the myofilaments to Ca. Myofilament Ca sensitivity is enhanced dynamically by stretching the myofilaments (as the heart fills with blood), resulting in a stronger

contraction. This is due, in part, to transverse filament lattice compression upon stretch, which enhances the actin–myosin interaction. This is an important autoregulatory mechanism by which the heart adjusts to altered diastolic filling (the classic Frank–Starling response). Myofilament Ca sensitivity is reduced by acidosis and by elevated phosphate and Mg concentrations (all three of which occur during ischemia). Myofilament Ca sensitivity is also reduced by β -adrenergic activation, but enhanced by caffeine and certain inotropic drugs. The rapid Ca transient kinetics prevent the myofilaments from fully equilibrating with $[Ca]_i$ during a normal twitch (especially in the rising phase). Thus, while contractile strength is indicative of underlying Ca transients, there is a dynamic interplay between Ca and myofilaments during E–C coupling contraction and relaxation.

Relaxation of the Heart

For relaxation to occur, $[Ca]_i$ must decline, allowing Ca to dissociate from the myofilaments. There are four pathways that remove Ca from the cytosol: the SR Ca-adenosine triphosphatase (ATPase), the sarcolemmal Na/Ca exchanger (NCX), the sarcolemmal Ca-ATPase, and the mitochondrial Ca uniporter, which takes Ca up into mitochondria (Figure 1). The last two pathways are collectively known as ‘slow’ Ca extrusion pathways and they have a minor contribution in myocytes from healthy hearts. The SR Ca-ATPase (SERCA) takes Ca back into the SR and thus is essential for the refilling of the SR with Ca. SERCA is under reversible regulation by phospholamban (PLB). Dephosphorylated PLB inhibits SERCA activity by reducing its apparent affinity for Ca, with no effect on the maximum pumping rate. Phosphorylation of PLB during sympathetic activation by either cyclic adenosine monophosphate (cAMP)-dependent protein kinase (PKA) or Ca-calmodulin-dependent protein kinase II (CaMKII) relieves this inhibition, allowing faster twitch $[Ca]_i$ decline and relaxation. Since the SERCA competes better with the Na/Ca exchange for extruding Ca, this also enhances the SR Ca content. Thus, relief of the inhibitory effect of PLB on SERCA is a major contributor to the overall positive inotropic and lusitropic effects of sympathetic stimulation (enhanced contraction strength and faster relaxation).

The NCX transports one Ca ion in exchange for three Na ions and can operate in both Ca efflux (or direct) and Ca influx (or reverse) mode, depending on the internal and external concentration of both Na and Ca, as well as on the membrane potential. High $[Ca]_i$ favors Ca efflux, whereas positive membrane potential and high $[Na]_i$ favor Ca influx. In normal myocytes, under physiological conditions, the NCX works almost exclusively in the Ca extrusion mode, driven mostly by the Ca transient. However, the positive membrane potential during the AP plateau can limit Ca extrusion. When $[Na]_i$ is elevated (e.g., by digitalis glycosides, which block the Na/K-ATPase), the amount of Ca influx through NCX can increase greatly. This will raise the cellular and SR Ca content, resulting in larger Ca transients and therefore enhanced contractility. Indeed, the inotropic effect of increasing $[Na]_i$ is well known. Even a small rise in $[Na]_i$ (e.g., 1 mM) can dramatically increase contractile force.

The relative importance of SERCA and NCX in extruding Ca from the cytosol varies between species. In human, rabbit, dog, and guinea-pig ventricular myocytes, the SERCA pump

removes approximately 70% of the activator Ca, and Na/Ca exchange removes 28%, leaving only ~1% to be removed via each of the slow systems. The amount of Ca which leaves the cytosol by entry into mitochondria is inconsequential with respect to E–C coupling, but slow cumulative changes in intra-mitochondrial $[Ca]$ can stimulate key dehydrogenases that increase nicotinamide adenine dinucleotide (NADH) and ATP production to match increased energetic demands that typically parallel the elevated Ca levels. In rat and mouse ventricle, SERCA function is higher than in rabbit (due to a greater concentration of pump molecules) and Ca removal via Na/Ca exchange is lower, resulting in a balance of 92:7:1% for SERCA, Na/Ca exchange, and slow systems. Thus, mouse and rat ventricle (which also exhibit very spike-like APs) poorly mimic human with respect to quantitative balance of cellular Ca flux.

The amount of Ca extruded from the cell during twitch relaxation must be the same as the amount of Ca entry for each beat; otherwise, the cell would gain or lose Ca (i.e., not be in steady state). Indeed, complementary measurements of Ca influx and SR Ca release during a twitch confirm this expectation in rabbit and rat. L-type Ca channels bring ~10 $\mu\text{mol l}^{-1}$ Ca into the myocyte during the plateau phase of the AP. When NCX removes this Ca, it brings ~32 $\mu\text{mol l}^{-1}$ Na into the cell (due to the 3Na:1Ca stoichiometry). Voltage-gated Na channels also bring in about 8 $\mu\text{mol l}^{-1}$, which means that the total Na entering the myocyte during the cardiac cycle is ~40 $\mu\text{mol l}^{-1}$. This excess Na is then extruded by the Na/K-ATPase to keep the intracellular Na level constant.

Altered E–C Coupling in Failing Hearts

Heart failure (HF) is a syndrome that develops when the amount of blood pumped from the heart is inadequate to meet the metabolic demands of the body. In its early stages, there is a reduction in exercise capacity, but as the syndrome progresses, the heart is eventually unable to pump a sufficient quantity of blood to meet the normal metabolic needs of the tissues, even at rest. HF is associated with a poor prognosis (~50% 5-year mortality rate), and for patients with severe symptoms, there is a nearly 50% 2-year mortality rate. Contractile dysfunction of cardiac myocyte is a main cause of HF, but other factors, such as changes in cardiac structure (dilatation), cell death (apoptosis), altered vascular structure and reactivity, abnormal energy utilization, and neurohormonal disturbances, also contribute to the progression of HF, at least under certain conditions.

Defective cardiac myocyte contractility in HF is generally due to altered cellular Ca handling. Myocyte Ca transients are depressed in HF and this causes systolic dysfunction. A major cause for smaller Ca transients in HF is a lower SR Ca content, while Ca entry via Ca channels and the SR fractional release are generally not altered. Reduced SR Ca load not only directly decreases the amount of Ca available for release but also reduces the fraction of SR Ca that is released for a given I_{Ca} trigger, due to the effect of intra-SR free $[Ca]$ on the open probability of RyRs. There are three factors that result in a reduced SR Ca content in HF (see Figure 2): (1) reduced SERCA function, (2) increased expression and function of NCX, and (3) enhanced diastolic SR Ca leak. Typically, SERCA

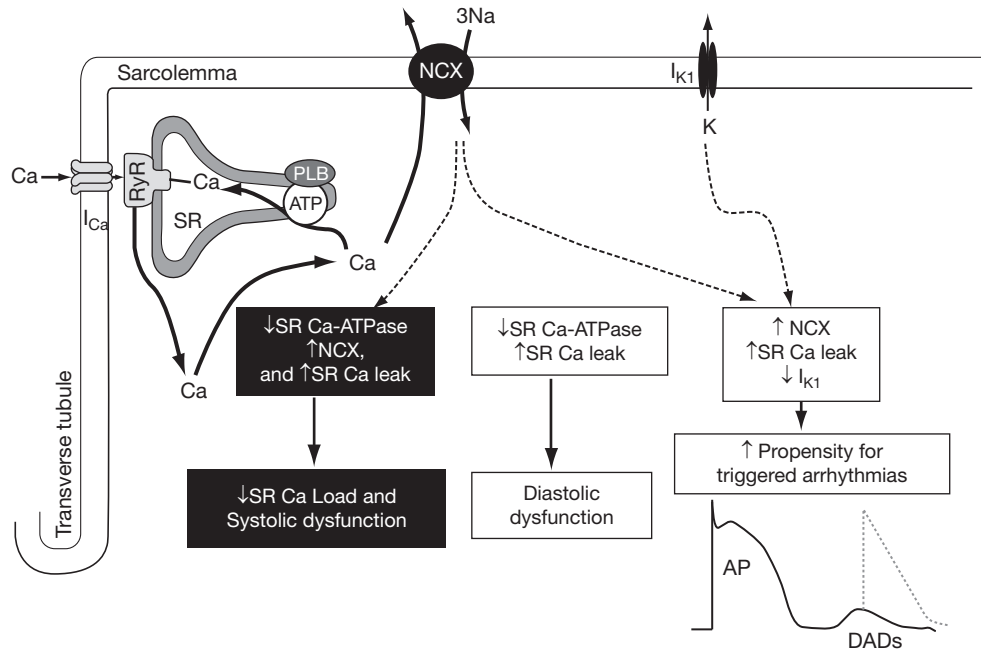


Figure 2 Cartoon showing the main alterations in E–C coupling that occur in heart failure myocytes. Modified from Bers DM (2001) *Excitation–Contraction Coupling and Cardiac Contractile Force*, 2nd edn. Dordrecht: Kluwer.

expression is reduced while the PLB level is unaltered in HF, although the extent of PLB phosphorylation is reduced. Therefore, the PLB:SERCA ratio is generally increased and the PLB that is there is more inhibitory. Both factors lower SERCA function in HF. The decreased SERCA activity in HF (typically by ~24–50%) coupled with an increased NCX activity (typically 50–100%) allows NCX to compete better with the SERCA in extruding Ca from the cytosol during relaxation. This brings the overall Ca transported by the SERCA and NCX closer to equal in HF and results in lower SR Ca content. There is a spectrum of increase in NCX and decrease in SERCA expression in individual HF patients, suggesting that there is individual variation as to how much the decrease in SR Ca load in HF is due to increased NCX and decreased SERCA function. Hearts with decreased SERCA and unchanged NCX had diastolic dysfunction, whereas hearts with increased NCX expression and relatively unaltered SERCA had preserved diastolic function. Thus, while downregulation of SERCA and upregulation of NCX both result in a lower SR Ca content and impaired contractility, reduced SERCA function also contributes to diastolic dysfunction (Figure 2). Increased expression and function of NCX not only preserve diastolic function but also lead to a higher propensity for triggered arrhythmias in HF (see below).

Elevated diastolic SR Ca leak also contributes to lower SR Ca content in HF (Figure 2). It has been reported that RyRs are hyperphosphorylated by PKA in HF and this leads to an increased SR Ca leak. However, not all studies agree about hyperphosphorylation. Moreover, PKA-dependent RyR phosphorylation does not enhance diastolic Ca leak via RyR (measured by direct block of leak and also via Ca spark measurements) at constant SR Ca load in control myocytes. RyR2 can also be phosphorylated by CaMKII (which is also bound to the RyR2) and this activates RyR opening and diastolic SR Ca release. CaMKII expression and activity and association with

RyR are increased in HF, and CaMKII blockade, but not PKA inhibition, reduced the diastolic SR Ca leak in ventricular myocytes from rabbits with pressure- and volume-overload-induced HF. Interestingly, CaMKII blockade restored SR Ca content, but only slightly enhanced systolic function. This may be because the increased RyR Ca-sensitivity that caused the leak also enhanced the fractional SR Ca release during E–C coupling. However, this sensitization of RyR to Ca in HF may play a very important role in diastolic dysfunction and triggered arrhythmias (see below). In summary, we think that the systolic dysfunction in HF is due mainly to lowered SR Ca content as a consequence of reduced SERCA function and enhanced NCX function while the enhanced SR Ca leak and NCX function may both contribute to an increased propensity for arrhythmias (Figure 2).

Increased Propensity for Arrhythmias in HF

Although many HF patients die due to pump failure (i.e., the heart cannot pump enough blood to maintain the metabolic processes of the organism), nearly half of all deaths occur suddenly, primarily from ventricular tachycardia degenerating to ventricular fibrillation. Myocytes from both patients with end-stage HF and animal models of ventricular hypertrophy and HF exhibit longer AP, which may be due to decreased outward K currents such as I_{to} or I_{K1} . Slow conduction and heterogeneous repolarization, which have been demonstrated in experimental models of nonischemic HF and in patients with idiopathic dilated cardiomyopathy, could contribute to reentrant arrhythmias in HF. Reentrant mechanisms are responsible for about half of the cases of sustained monomorphic ventricular tachycardia in patients with ischemic cardiomyopathy. In nonischemic HF (as well as in ~50% of ischemic HF cases), ventricular

tachycardias initiate mainly by a nonreentrant mechanism. This mechanism involves triggered activity from either delayed afterdepolarization (DAD) or early afterdepolarization (EAD) (triggered arrhythmias).

DADs are initiated by spontaneous SR Ca release at diastolic membrane potentials causing a Ca-activated transient inward current (I_{ti}), which depolarizes the membrane toward the threshold for a triggered AP. The depolarizing I_{ti} is mediated mainly by the NCX, although a nonspecific cationic current and a Ca-activated chloride current had once been thought to be involved in I_{ti} . DADs are more common at high heart rates, because the SR Ca content, and thus the propensity for spontaneous Ca release, is elevated. EADs typically occur in the setting of prolonged repolarization due to alterations in K or Na currents (e.g., long QT syndromes, hypokalemia, and class III antiarrhythmic agents), and also occur in failing myocytes. EADs generally result from reactivation of L-type Ca current, especially when the AP is prolonged. These early EADs occur early during repolarization (during the plateau phase of the AP) and are favored at very low heart rates. However, EADs have also been observed at rapid heart rates. These late EADs generally occur late during repolarization, when the membrane potential is below -40 mV, and are produced by the inward current generated by spontaneous SR Ca release. This mechanism is similar to that leading to DADs occurrence, except that the SR Ca release occurs during the AP.

Although both DADs and EADs have been reported in HF, the nonreentrant initiation of ventricular tachycardias in HF generally involves DADs, whereas EADs are less prominent. This is because EADs generally occur at low heart rates, whereas most ventricular tachycardias in HF occur at more rapid rates (although bradyarrhythmic deaths in HF have also been reported). Second, although hypertrophied and HF myocytes exhibit prolonged AP duration (vs. controls) at low pacing rates, these differences are minimal at more rapid (and physiologic) heart rates at which ventricular arrhythmias occur.

There are alterations in Ca handling and ion channels in HF that could contribute to the increased propensity for DADs despite reduced SR Ca content (Figure 2). When a spontane-

ous SR Ca release occurs, the upregulated NCX in HF causes more inward I_{ti} (NCX current) for a given SR Ca release. There is also a large reduction in the inward rectifier K current that is responsible for stabilizing the resting membrane potential (I_{K1}), which means that any given I_{ti} will cause more membrane depolarization. These changes in NCX and I_{K1} combine to increase the likelihood that an SR Ca release event will induce a DAD that is sufficiently large to trigger a full AP.

See also: **Bioenergetics:** Calcium/Calmodulin-Dependent Protein Kinase II; ER/SR Calcium Pump: Structure; Membrane Transporters: $\text{Na}^+/\text{Ca}^{2+}$ Exchangers; P-Type Pumps: H^+/K^+ Pump.

Further Reading

- Ai X, Curran JW, Shannon TR, Bers DM, and Pogwizd SM (2005) Ca^{2+} /calmodulin-dependent protein kinase modulates cardiac ryanodine receptor phosphorylation and sarcoplasmic reticulum Ca^{2+} leak in heart failure. *Circulation Research* 97: 1314–1322.
- Bers DM (2001) *Excitation–Contraction Coupling and Cardiac Contractile Force*, 2nd edn., p. 427. Dordrecht: Kluwer.
- Bers DM (2002) Cardiac excitation–contraction coupling. *Nature* 415: 198–205.
- Hasenfuss G, Schillinger W, Lehnart SE, et al. (1999) Relationship between $\text{Na}^+/\text{Ca}^{2+}$ -exchanger protein levels and diastolic function of failing human myocardium. *Circulation* 99: 641–648.
- Maier LS, Zhang T, Chen L, et al. (2003) Transgenic CaMKII δ C overexpression uniquely alters cardiac myocyte Ca^{2+} handling: Reduced SR Ca^{2+} load and activated SR Ca^{2+} release. *Circulation Research* 92: 904–911.
- Marx SO, Reiken S, Hisamatsu Y, et al. (2000) PKA phosphorylation dissociates FKBP12.6 from the calcium release channel (ryanodine receptor): Defective regulation in failing hearts. *Cell* 101: 365–376.
- Piacentino V, III, Weber CR, Chen X, et al. (2003) Cellular basis of abnormal calcium transients of failing human ventricular myocytes. *Circulation Research* 92: 651–658.
- Pogwizd SM, Schlotthauer K, Li L, Yuan W, and Bers DM (2001) Arrhythmogenesis and contractile dysfunction in heart failure: Roles of sodium–calcium exchange, inward rectifier potassium current, and residual beta-adrenergic responsiveness. *Circulation Research* 88: 1159–1167.
- Shannon TR, Pogwizd SM, and Bers DM (2003) Elevated sarcoplasmic reticulum Ca leak in intact ventricular myocytes from rabbits in heart failure. *Circulation Research* 93: 592–594.

Carnitine and β -Oxidation

J Kerner, Case Western Reserve University School of Medicine, Cleveland, OH, USA

C L Hoppel, Case Western Reserve University School of Medicine, Cleveland, OH, USA; Veterans Affairs Medical Center, Cleveland, OH, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Carnitine A water-soluble small molecular-weight compound derived from the essential amino acids, lysine and methionine, and is involved in reversible transesterification reactions with acyl-CoAs of different chain lengths. It serves as a general transport vehicle for activated fatty acids.

CoASH Coenzyme A is a cosubstrate of all acylcarnitine transferases catalyzing the reversible transesterification as

well as of numerous other enzymes central to energy metabolism.

Mitochondria Eukaryotic organelles that are surrounded by an inner and an outer membrane and are the sites of energy production.

β -Oxidation A series of enzyme reactions within the mitochondria, called mitochondrial matrix, in which fatty acids are progressively chain-shortened by the removal of two-carbon units as acetyl-CoA.

Carnitine (structure in [Scheme 1](#)) is derived from the essential amino acids, lysine and methionine, is ubiquitous in nature, and is found in especially high concentration in muscle tissue of higher organisms. It functions as a transport vehicle for activated fatty acids of different chain length through membranes within the cell. This transport function is best characterized in mitochondrial oxidation of long-chain fatty acids. This latter process, also known as mitochondrial β -oxidation, represents the repetitive oxidative cleavage of long-chain fatty acids into two carbon units, acetyl-CoA. Acetyl-CoA is then either further oxidized for energy production (all tissues) or used to synthesize ketone bodies (liver) as metabolic fuel for peripheral tissues.

Carnitine Biosynthesis and Homeostasis

The first clue about the physiological function of carnitine came from studies in the mealworm, *Tenebrio molitor*, which suggested that carnitine plays a vital function in fat catabolism. However, carnitine is not a vitamin for higher animals; the daily need is met by endogenous synthesis and dietary intake, mostly from meat products. For endogenous synthesis, the ultimate precursors are the essential amino acids, lysine and methionine. Protein-bound lysine is methylated to trimethyllysine using S-adenosylmethionine, and following liberation by proteolysis, the free trimethyllysine is converted (mostly in muscle) by a series of reactions to butyrobetaine, the ultimate carnitine precursor. Although most tissues are capable of synthesizing butyrobetaine, the hydroxylation of butyrobetaine to carnitine is restricted to the liver (and to a lesser extent to the kidney and brain). Following release of carnitine from the liver into the plasma, carnitine is taken up by other tissues by a carrier-mediated transport process ([Scheme 2](#)).

The body distribution of carnitine is determined by a series of systems that transport carnitine into cells against a concentration gradient, an independent efflux process, and an exchange mechanism in a tissue-specific fashion. Under physiological conditions, plasma carnitine concentration is maintained within a narrow range, predominantly by renal clearance, which is determined by the kinetic properties of carnitine transport system in the brush border.

Functions of Carnitine

In higher animals, the only reaction carnitine undergoes is a reversible transesterification with acyl-CoAs catalyzed by carnitine acyltransferases of differing chain length specificities (carnitine acetyl-, octanoyl-, and palmitoyltransferases) ([Scheme 3](#)).

All known functions of carnitine are based on the reversible transesterification with acyl-CoA esters. The physiologically most important and, therefore, the best-studied metabolic pathway in which carnitine plays a significant role is the mitochondrial oxidation of long-chain fatty acids.

Mitochondrial β -Oxidation

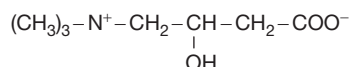
Cellular uptake and activation of long-chain fatty acids

Long-chain fatty acids represent a main energy source for many organs, especially for muscle (heart and skeletal) and liver. Since most tissues contain only small amounts of storage lipids, they depend, for energy production, on a continuous supply of fatty acids, mostly from adipose tissue. Following mobilization by lipolysis in adipose tissue and transport in the blood bound to albumin, fatty acids are taken up by the tissues in a process mediated by transport proteins present in the plasma membrane. Once within the cell, free fatty acids are bound to fatty acid binding proteins, which are present in the cytosol in large amounts. Depending on the tissue and its metabolic demand, fatty acids are converted to triglycerides and stored, secreted as very low density lipoproteins (liver), or oxidized in the mitochondria for energy production. Before being directed into storage or oxidation, fatty acids first are activated to acyl-CoA esters. This reaction is catalyzed by long-chain acyl-CoA synthetase, an enzyme present in abundance both in microsomes and in mitochondria. In mitochondria, long-chain acyl-CoA synthetase is associated with the mitochondrial outer membrane and appears to be a transmembrane protein with its active site exposed to the cytosol. This orientation of long-chain acyl-CoA synthetase gives rise to cytosolic production of long-chain acyl-CoAs, which, for their further β -oxidation, must cross both mitochondrial boundary membranes. Although the impermeability of the inner membrane to

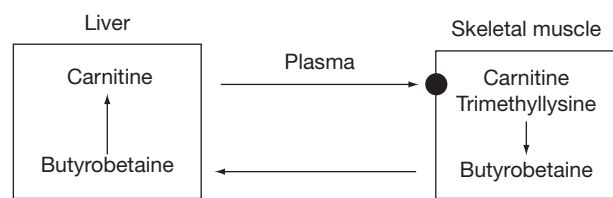
metabolites is well known, it is now recognized that the outer membrane is not freely permeable to solutes as was previously thought. A potential route for long-chain acyl-CoAs to cross the outer membrane could be the voltage-dependent, anion-selective channel (VDAC), also called mitochondrial porin, which occurs at high density in the mitochondrial outer membrane and regulates the permeability of this membrane to ions and metabolites (Figure 1).

Mitochondrial uptake of activated fatty acids – the mitochondrial carnitine system

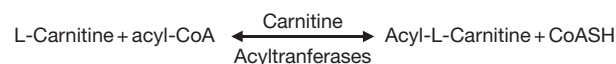
The mitochondrial carnitine system plays an obligatory role in β -oxidation of long-chain fatty acids by providing a means for their transport into the mitochondrial matrix. This transport



Scheme 1 Structure of L-carnitine (4-*N*-trimethylammonio-3-hydroxybutanoate).



Scheme 2 Interorgan relationship in carnitine biosynthesis.



Scheme 3 Reversible transesterification of carnitine catalyzed by carnitine acyltransferases with different chain-length specificity.

system consists of the malonyl-CoA-sensitive carnitine palmitoyltransferase-I (CPT-I) localized in the mitochondrial outer membrane, the carnitine:acylcarnitine translocase, an integral inner membrane protein, and carnitine palmitoyltransferase-II (CPT-II) localized on the matrix side of the inner membrane. In the schema depicted in Figure 1, the catalytic site of CPT-I faces the intermembrane space. In an alternative model, the catalytic site of CPT-I faces the cytosol. In this scenario, the carnitine esters must cross both mitochondrial membranes. In either case, transesterification of long-chain acyl-CoA esters to their carnitine esters by CPT-I commits activated fatty acids to oxidation in the mitochondrial matrix. CPT-I, by virtue of its inhibition by malonyl-CoA, represents a key regulatory site controlling the flux through β -oxidation. Consistent with its central role in mitochondrial fatty acid oxidation, the enzyme exists in at least two isoforms with significantly different kinetic and regulatory properties. The liver isoform displays a higher affinity for carnitine and a lower affinity for the physiological inhibitor, malonyl-CoA, as opposed to the muscle isoform. Long-term regulation of both isoforms is accomplished at the transcriptional level. Short-term or acute regulation is achieved via changes in tissue malonyl-CoA concentration, and for the liver isoform also through changes in the enzyme's sensitivity to malonyl-CoA inhibition.

The long-chain acylcarnitines formed by CPT-I in the mitochondrial outer membrane enter the mitochondrial matrix in exchange for free carnitine catalyzed by the carnitine:acylcarnitine translocase, an integral inner membrane protein. The carnitine:acylcarnitine translocase catalyzes a homologous/heterologous carnitine exchange (in addition to the slow unidirectional carnitine transport) and displays increasing affinity for carnitine esters with increasing acyl chain-length. Experimental data suggest the existence of a single transporter protein with a substrate specificity ranging from free carnitine to long-chain carnitine esters. This is in contrast to the known substrate specificity of the mitochondrial carnitine acyltransferases (carnitine palmitoyl- and acetyltransferase) and implies that, in addition to its role in mitochondrial fatty acid oxidation, the same transporter is also engaged in other functions involving membrane transport of acyl groups.

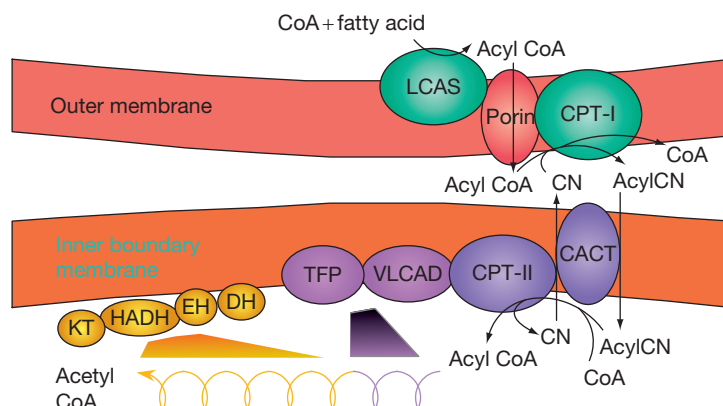


Figure 1 Schematic depiction of the carnitine-dependent (CN) transport of activated fatty acids into the mitochondrial matrix and of their oxidative chain shortening via the β -oxidation pathway. LCAS, long-chain acyl-CoA synthetase; CPT-I, malonyl-CoA-sensitive carnitine palmitoyltransferase; CN, carnitine; CACT, carnitine:acylcarnitine translocase; CPT-II, malonyl-CoA-insensitive carnitine palmitoyltransferase; VLCAD, very long-chain acyl-CoA dehydrogenase; TFP, trifunctional protein; DH, acyl-CoA dehydrogenase; EH, enoyl-CoA hydratase; HADH, 3-hydroxyacyl-CoA dehydrogenase; KT, 3-ketoacyl-CoA thiolase.

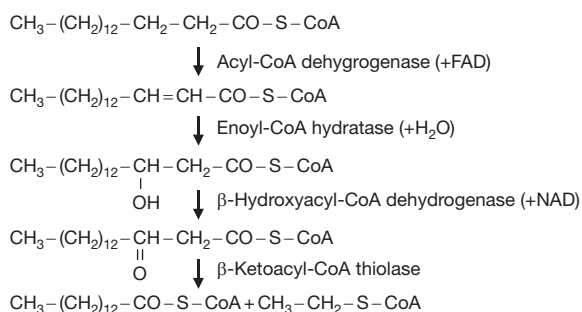
Following the translocation of long-chain acylcarnitines by carnitine:acylcarnitine translocase into the mitochondrial matrix, the carnitine esters are reconverted to their respective CoA esters by CPT-II, thus completing the carnitine-dependent uptake of activated fatty acids. Within the matrix, CPT-II is anchored to the inner membrane as a peripheral membrane protein. The enzyme has been isolated in catalytically active form from a variety of tissues and its physical, kinetic, and immunological properties determined. CPT-II is the product of a single gene and is expressed as the same protein in all tissues within the body. CPT-II is not inhibited by malonyl-CoA, either in its native environment (membrane-bound) or when isolated.

Pathway of mitochondrial β -oxidation

In the fatty-acid oxidation cycle, the activated long-chain fatty acid undergoes a consecutive oxidative chain shortening. Each cycle of this sequence of four reactions results in the removal of two carbon atoms from the fatty acyl residue in the form of acetyl-CoA (Figure 1 and Scheme 4) and two reducing equivalents (flavin adenine dinucleotide (FADH) and nicotinamide adenine dinucleotide (NADH)). Thus, the complete oxidation of palmitoyl-CoA requires seven cycles yielding eight acetyl-CoAs and seven FADH and NADH molecules. Acetyl-CoA is then further oxidized in the tricarboxylic acid cycle to CO_2 and reducing equivalents, each acetyl-CoA yielding three NADH and one FADH molecule(s). The reoxidation of these reducing equivalents in the electron transport chain then produces adenosine triphosphate (ATP). The complete oxidation of one mole palmitic acid thus results in the formation of 131 mol ATP. On a mole basis, this is 3.6-times more and on a weight basis 2.5-times more ATP generated compared with an equivalent amount of glucose.

The first of the four reactions in mitochondrial β -oxidation of activated long-chain fatty acids is catalyzed by the acyl-CoA dehydrogenases, a group of enzymes with differing but somewhat overlapping chain length specificity (very long-chain, long-chain, medium-chain, and short-chain) resulting in the formation of 2-*trans*-enoyl-CoA. They are composed of four identical subunits, each of which contains a non-covalently bound FAD. The enzyme is reoxidized via the electron-transferring flavoprotein and then in the electron transport chain by electron-transferring flavoprotein:ubiquinone oxidoreductase.

The second step of β -oxidation is catalyzed by 2-enoyl-CoA hydratases converting 2-*trans*-enoyl-CoA to 3-hydroxyacyl-CoA.



Scheme 4 Reactions of mitochondrial β -oxidation of saturated long-chain acyl-CoAs.

This reaction is catalyzed by the membrane-bound long-chain trifunctional protein, TFP, formerly known as the long-chain enoyl-CoA hydratase. The TFP comprises two non-identical subunits, the α and β subunits, present in a 1:1 molar ratio. The enzyme comprises the catalytic activities of 2-enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase (α subunit) and 3-ketoacyl-CoA thiolase (β subunit). The soluble short-chain specific form, also known as crotonase, catalyzes the hydration of short-chain 2-enoyl-CoAs.

The third reaction is catalyzed by 3-hydroxyacyl-CoA dehydrogenases. At least two 3-hydroxyacyl-CoA dehydrogenase activities are found in the mitochondria. In addition to the TFP, there is also a short-chain-specific form. The latter is composed of two identical subunits. Both activities require NAD as a co-factor. NADH formed is then reoxidized by complex I of the electron transport chain.

The last reaction of the β -oxidation spiral is the thiolytic cleavage of 3-ketoacyl-CoA catalyzed by the enzymes, 3-ketoacyl-CoA thiolases. The general acyl-CoA thiolase and the long-chain 3-ketoacyl thiolase of the TFP have overlapping substrate specificities. The reaction is greatly in favor of product formation. There is also a third thiolase enzyme present in mitochondria, called acetyl-CoA acetyltransferase or acetoacetyl-CoA thiolase. This latter enzyme is specific for acetoacetyl-CoA and 2-methylacetoacetyl-CoA and probably is involved in branched-chain amino acid metabolism, in ketogenesis, and in ketone body utilization.

The oxidation of (poly)unsaturated fatty acids requires auxiliary enzymes for their complete β -oxidation. Most unsaturated fatty acids have their double bonds in *cis* configuration with the first double bond between carbon atoms 9 and 10. Additional double bonds, if any, occur at three-carbon intervals. Several cycles of β -oxidation yields a chain-shortened acyl-CoA with a *cis* double bond in 3, 4 position that is not a substrate for enoyl-CoA hydratase. Isomerization catalyzed by 2,3-enoyl-CoA isomerase leads to 2-*trans*-enoyl-CoA, which can undergo further β -oxidation yielding, after two additional cycles, *trans*-2, *cis*-4-dienoyl-CoA, which is also a poor substrate for enoyl-CoA hydratase. The *cis*-4 double bond is reduced by the enzyme 2,4-dienoyl-CoA reductase producing *trans*-3-enoyl-CoA, which is isomerized to *trans*-2-enoyl-CoA by 2,3-enoyl-CoA isomerase.

The rate of β -oxidation is adjusted to physiological needs and is subject to control in which substrate supply is important. The reaction catalyzed by CPT-I committing activated fatty acids toward oxidation in the mitochondria often is considered rate limiting. The regulatory property that imparts such control to this otherwise near-equilibrium reaction is the inhibition of CPT-I by malonyl-CoA. The muscle isoform of CPT-I is extremely sensitive to inhibition by malonyl-CoA and it is thought that, in heart and skeletal muscle, CPT-I is regulated only by changes in tissue malonyl-CoA concentrations. The liver isoform of CPT-I is much less sensitive to malonyl-CoA inhibition and, in liver, the entry of activated fatty acids into the mitochondria is controlled by changes in the enzyme's sensitivity to malonyl-CoA as well as by changes in hepatic malonyl-CoA concentrations.

Although the mitochondrial uptake of long-chain activated fatty acids is absolutely carnitine dependent, the efficiency of the system is such that, under physiological conditions, tissue

carnitine concentrations seemingly have no significant impact on this process.

The control of β -oxidation flux by the redox state of NAD/NADH and FAD/FADH (electron-transferring flavoprotein (ETF/ETFH₂)) as well as by the acetyl-CoA/CoASH ratio is important under some conditions.

Regulation of Mitochondrial Acyl-CoA/CoA Ratio

The β -oxidation pathway is not the only one in which carnitine plays a role as a carrier of activated acyl groups. In conjunction with carnitine acetyltransferase and carnitine:acylcarnitine translocase, carnitine forms an effective shuttle system for acetyl groups out of the mitochondria, thus preventing sequestration of intramitochondrial free CoASH. This acetyl-CoA buffer function allows CoASH-dependent reactions such as pyruvate oxidation to proceed unimpaired. The CoASH buffer function is not restricted to acetyl-CoA. Under normal physiological conditions, the direction of long-chain fatty acid movement is into the mitochondrial matrix. However, the carnitine:acylcarnitine translocase and CPT-II/carnitine acetyltransferase system can under certain conditions operate in the reverse direction, for example, removing activated short-, medium- and long-chain acyl groups from the matrix. This buffering has been demonstrated with isolated mitochondria, as well as with cultured fibroblasts and lymphocytes from patients with defects in mitochondrial fatty acid oxidation. The accumulation of cellular acyl-CoAs as a result of an inborn error of fatty

acid metabolism leads to a corresponding increase in acylcarnitines in the cell. In contrast to the acyl-CoAs, the acylcarnitines can be transported across the plasma membrane into the blood. The identification of these acylcarnitines in blood is thus a signpost for the metabolic defect.

See also: **Metabolism Vitamins and Hormones:** Coenzyme A; Diabetes; Fatty Acyl-CoA Synthetases; Starvation.

Further Reading

- Eaton S (2002) Control of mitochondrial β -oxidation flux. *Progress in Lipid Research* 41: 197–239.
- Fraenkel G and Friedman S (1957) Carnitine. In: Harris RS, Marian GF, and Thiman KV (eds.) *Vitamins and Hormones*, vol. 15, pp. 73–118. New York, NY: Academic Press.
- Guzman M and Geelen MJH (1993) Regulation of fatty acid oxidation in mammalian liver. *Biochimica et Biophysica Acta* 1167: 227–241.
- Kerner J and Hoppel C (1998) Genetic disorders of carnitine metabolism and their nutritional management. *Annual Review of Nutrition* 18: 179–206.
- Kerner J and Hoppel C (2000) Fatty acid import into mitochondria. *Biochimica et Biophysica Acta* 1486: 1–17.
- McGarry JD and Brown NF (1997) The mitochondrial carnitine palmitoyltransferase system. *European Journal of Biochemistry* 244: 1–14.
- McGarry JD and Foster DW (1980) Regulation of hepatic fatty acid oxidation and ketone production. *Annual Review of Biochemistry* 49: 395–420.
- Vaz FM and Wanders RJA (2002) Carnitine biosynthesis in mammals. *Biochemical Journal* 361: 417–429.
- Zammit VA (1999) Carnitine acyltransferases: Functional significance of subcellular distribution and membrane topology. *Progress in Lipid Research* 38: 199–224.

Caspases and Cell Death

D W Nicholson, Merck Frosst Centre for Therapeutic Research, Pointe Claire – Dorval, QC, Canada

G Melino, Leicester University, Leicester, UK; University of Rome “Tor Vergata”, Biochemistry IDI-IRCCS laboratory, Rome, Italy

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Don W. Nicholson, Pierluigi Nicotera, and Gerry Melino, volume 1, pp. 319–327, © 2004, Elsevier Inc.

Glossary

Apoptosis Physiological mechanism of programmed cell death mediated by caspases (cysteine proteases), regulated by the Bcl-2 family of proteins, and triggered from the membrane (death receptors and adhesion), cytosolic stress, or nuclear signals (DNA damage). Excessive or defective apoptosis is involved in diseases such as ischemia, acquired immunodeficiency syndrome (AIDS), neurodegeneration, autoimmunity, and cancer.

Apoptosome Multiprotein complex of 700-kDa molecular weight formed by procaspase-9, apoptotic protease activating factor 1 (Apaf-1), deoxyadenosine triphosphate (dATP), and cytochrome-*c*. It proteolytically activates caspase-9, leading to the activation of downstream effector caspases, and subsequently death of the cell.

Baculovirus IAP repeat (BIR) A protein structural domain, conserved in both mammalian and viral proteins able to interfere with the apoptosome or with the active caspases. It characterizes the IAP proteins.

CysteinyI aspartate-specific protease (caspase) Cysteine proteases which cleave after an aspartate (D). The family

includes 13 enzymes involved either in programmed cell death (apoptosis) or in inflammation.

Death initiation signaling complex (DISC) Multiprotein complex beneath the cell membrane activated by trimeric death receptors (e.g., CD95, TRAIL-R, and tumor necrosis factor (TNF)). It includes the receptor, adaptor molecules, procaspases (caspases 8 and 10), and their related substrates (e.g., Bid) that signal to the mitochondria–apoptosome to activate the caspases, and therefore kill the cell by apoptosis.

Death receptors Cell membrane receptors (e.g., CD95, TRAIL-R, and TNF) belonging to the TNF superfamily, able to activate apoptosis via the formation of a DISC when bound by their specific ligands (e.g., CD95L, TRAIL, and TNF). These are essential in regulating the immune response.

Inhibitors of apoptosis (IAPs) Mammalian or viral proteins able to inhibit the activation of the apoptosome or the active caspases. They are characterized by structural domains called BIR; consequently, all IAP proteins are also called BIR proteins (BIRp).

Apoptosis, a major form of cell death used to remove unwanted or excess cells, is characterized by a plethora of morphological and biochemical changes resulting primarily from the activation of a family of cysteine proteases, which cleave their substrates at specific aspartate residues. Caspases are cysteine proteases involved in cell death. The name caspase is derived from cysteinyl aspartate-specific protease, that is to say proteases which cleave after an aspartate (D). The caspase family is comprised of 13 proteins in human (Figure 1), seven in *Drosophila melanogaster*, and a single protein in *Caenorhabditis elegans*. As depicted in Figure 2, all caspases consist of three structural domains: a prodomain, a large subunit, and a small subunit. The prodomain is important for the regulation of the enzymatic activation of caspases; the active enzyme consists of the small and large subunits together. To obtain an enzymatically active enzyme, the three subunits must be proteolytically cleaved, which is effected by other endoproteases, normally also a caspase. The site of cleavage of the substrate is indicated by the positions P4, P3, P1, and P1, and is always Asp (D) (Figure 3). The other positions are highly variable.

History and Classification

Caspases were first identified in 1989 by Sleath and Schmidt as a protease required for the activation of interleukin 1 β .

In 1993, the group of Bob Horvitz discovered that the cell death gene *ced-3* is a protease belonging to the same family. In 1999, the first human genetic disease caused by mutation of a caspase (type 10) was identified by Michael Lenardo.

Phylogenetic analyses, and also studies of substrate specificity, indicate that caspases can be subdivided into three different groups. Figure 4 reports the present classification of the human caspases.

Group 1 includes caspases 1, 4, 5, and 12, which are involved in inflammation. The prototype of this group is caspase 1, previously known as interleukin conversion enzyme 1 β , or ICE. Its enzymatic activity was identified in 1989, but the enzyme was only purified and cloned in 1992, the same year in which Lois Miller identified the first inhibitor, the viral protein CrmA. Caspase 1, or ICE, controls the maturation of interleukin-1 beta (IL-1 β) actively participating in the inflammatory response. Site P4 of the substrate is preferentially a hydrophobic residue.

Group 2 includes caspases 2, 3, and 7, shown by a star in Figure 4. The prototype of this group is caspase 3, the homolog of *ced-3* in *C. elegans*. Group 2 caspases are directly involved in apoptosis in the terminal effector phases. Group 2 caspases have a very small prodomain, indicating a simple regulation of their enzymatic activation. These caspases are in general activated directly by another caspase, in a cascade of enzymatic amplification. Site P4 of the substrate is preferentially an Asp

(D) residue. Site P3 of the substrate is preferentially a Glu (E) residue. Therefore, this subfamily preferentially cleaves substrates with the sequence DEXD.

Group 3 consists of caspases 6, 8, 9, and 10. These caspases have a very large prodomain, and therefore a complex mechanism of activation. In fact, these caspases need molecular adaptors to be enzymatically activated, in contrast to group 2 caspases. Caspases 8 and 10 are part of the signal transduction mechanism of apoptosis receptors such as CD95; in this case, the adaptor molecule is Fas-associated protein with death domain (FADD; **Figure 3(a)**). Caspase 9 is activated in the apoptosome, and the necessary adaptor molecule is apoptotic protease activating factor 1 (Apaf-1; see **Figure 3(b)**). Site P4 of the substrate is preferentially an aliphatic residue.

Structure and Active Site

As already mentioned, caspases consist of three structural domains: a prodomain, a large subunit (also named as p20), and a small subunit (called p10; **Figure 5**).

The enzyme is constituted of the small and large subunits (p20/p10, around 30 kDa). The active site is in fact in the p20 subunit: it can be further classified into four different subsites

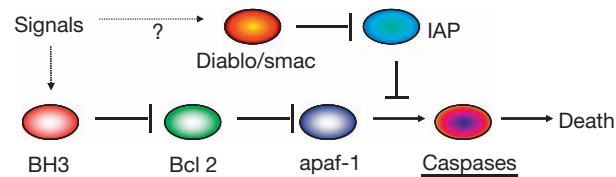


Figure 1 Caspases and cell death. Caspases are the proteolytic enzymes responsible for cell death. They are activated by an adaptor molecule (apaf-1), which in turn are controlled by the Bcl-2 family, and in particular by the BH3-only subfamily and the Bcl2-subfamily. The final stages of activation of the caspase can be regulated by a different pathway, the IAPs. This simple scheme is conserved in evolution from *C. elegans* and *D. melanogaster*.

S4–S1. The S1 subsite, which attacks the carboxyl group of the polypeptide side chain in P1 (the site of cleavage is Asp -D), is composed of p20 as well as p10 subunits. Similarly, the site for recognition and positioning of the substrate (S4–S1) is furnished by both p20 and p10, even though the residues mostly responsible for specificity (S4) are contained in p10 (see **Figure 6**).

Caspases enzymatically function as dimers of two identical molecules bound to each other. **Figure 5** shows a very schematic view of the dimer formation.

Each monomer (p20/p10) is formed by a compact cylinder dominated by six β -sheets and five α -helices distributed on the opposite site of the plane formed by the β -sheets. The dimer (p20/p10)₂ contains two (p20/p10) units aligned head to tail, which therefore position their respective active sites on opposite sides of the molecule. While there are two active sites, their cooperation has never been observed. The orientation of the molecule indicates a specific mechanism of activation, well regulated with security systems (a 'safety catch').

Substrate Recognition

The substrate recognition site of the caspase is indicated as S4–S1, at which the substrate is bound at residues P4–P1 (**Figure 6**). This recognition guarantees an absolute specificity. The S4–S1 substrate recognition site varies in each caspase, even if S1 is absolutely constant and highly tight. This is responsible for the absolute specificity to bind Asp (D) as the substrate residue P4. In fact, Asp is physically constrained and held by a hydrogen bond with Arg179, Gln283, Arg341 (the numbers refer to caspase 3). The dimensions of the S1 subsite are reduced to a minimum, and are therefore responsible for the lack of tolerance for P1 (Arg). Bond sites P2 and P3 (subsites S2 and S3) are more tolerant, less specific, and more heterogeneous in the various caspases. Therefore, the substrate (or the inhibitor) positions itself within the S1–S4 site, creating interactions as hydrogen bonds with Ser339 and Arg341 (conserved in practically all of the caspases). The P4 (S4) binding site is the major determinant of specificity (**Table 1**).

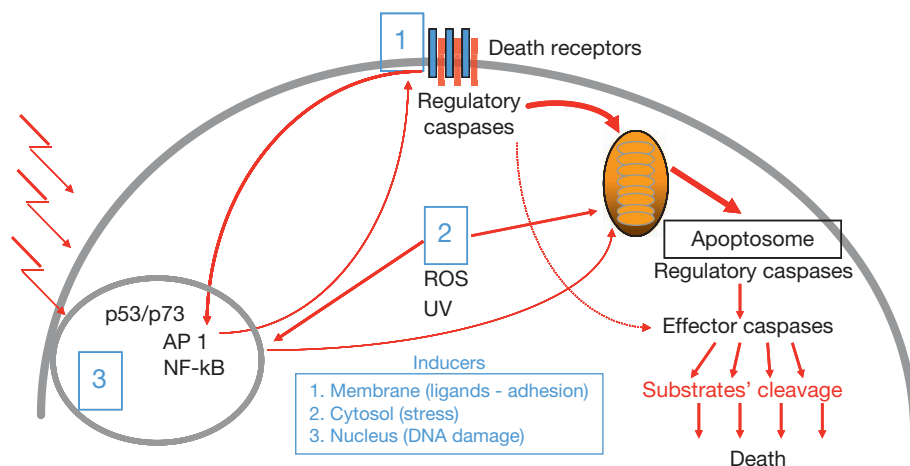


Figure 2 The apoptotic pathways in humans. In mammalian cells, apoptosis can be triggered at three different levels, as indicated. The signal converges to the apoptosome where the final effector phase occurs. Caspases are involved both in the upstream regulatory phase and in the final terminal phase. The regulation at the level of the death receptor, and of the apoptosome, is shown in greater detail in **Figure 3**.

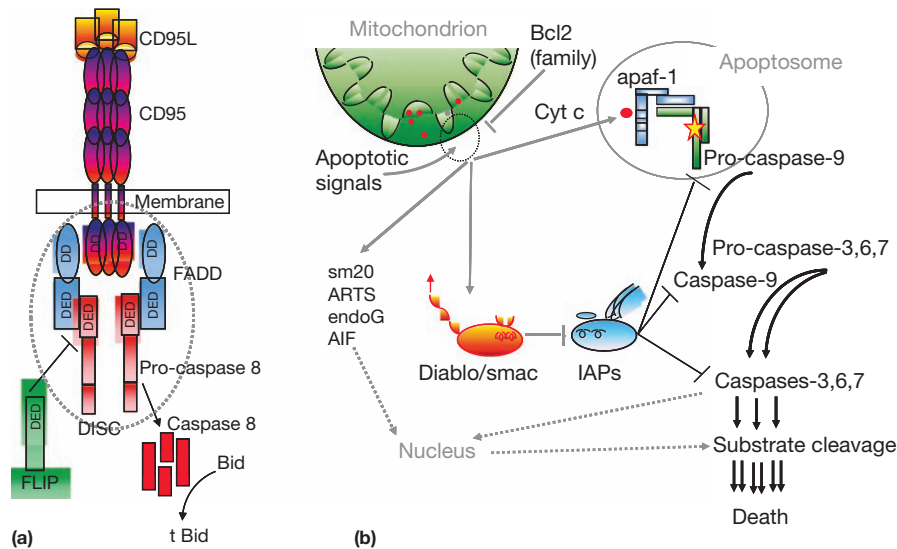


Figure 3 Regulation of the apical regulatory caspases. The general scheme is reported as in Figure 2. (a) Formation of the death-inducing signaling complex (DISC), following the activation of a death receptor such as CD95 (also called Fas or APO-1). Three molecules of ligand (CD95L) bind three molecules of receptor (CD95), allowing the recruitment of the adaptor molecule FADD via their death domain (DD). In turn, FADD, via its death effector domain (DED), recruits and activates caspase 8, which cleaves the specific substrate Bid. Truncated Bid (tBid) is in fact the molecular signal that propagates the death signal to the mitochondria and the apoptosome. This mechanism can be inhibited by the molecule FLIP, via its DED. (b) When activated by apoptotic signals (e.g., by tBid), mitochondria release several molecules, as indicated. In particular, cytochrome *c* goes to the apoptosome (formed by cytochrome *c*, apaf-1, pro-caspase 9, and dATP), and allows the activation of caspase 9. This, in turn, activates several molecules of downstream effector caspases (type 3, 6, and 7), and, consequently, the cleavage of many cellular substrates results in cell death. Still at its terminal stage, cell death can be inhibited by specific inhibitory proteins (IAPs), but another mitochondria-released protein, Diablo/Smac, can remove the protection by IAPs. The cascade of proteolytic amplification created by caspase 9 (apical regulatory caspase) and caspases 3, 6, and 7 (downstream effector caspases) is extremely powerful.

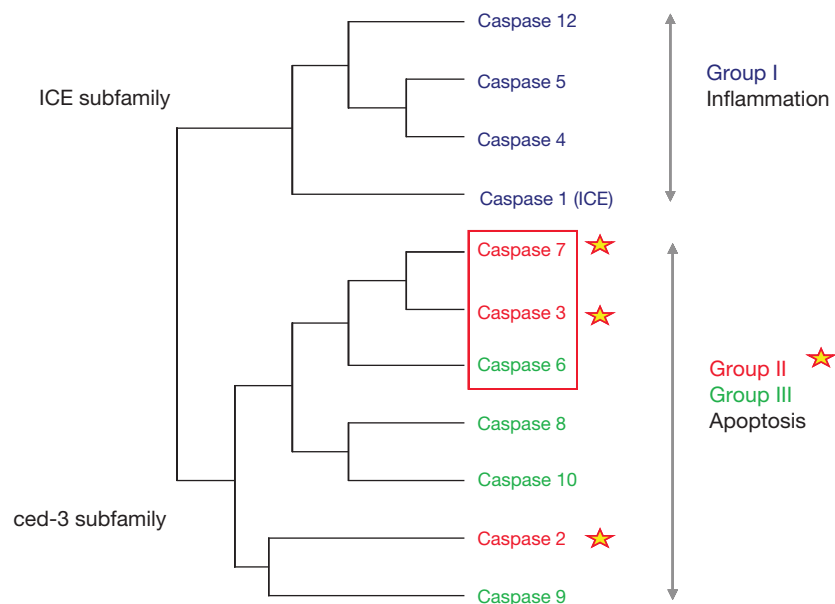


Figure 4 The human caspases family: Classification of the human caspases. According to the phylogenetic tree, caspases are subdivided into two major subfamilies, ICE and ced 3. The former are involved in inflammation, and the latter in apoptosis. The caspases with a very short prodomains (<30 aa) are boxed (type 3, 6, and 7); all the other enzymes have long prodomains (>10 kDa) subject to complex regulation. The enzymes can be divided according to their proteolytic specificity into group I, involved in cytokine maturation, group II, involved in the final effector phase of apoptosis (indicated by a star), and group III, involved in the upstream regulation of apoptosis. At least two gene clusters have been identified, consistent with some caspases arising from tandem gene duplication. Caspase 13 is an error of sequencing; caspase 12 is mutated into a nonfunctional enzyme in up to 90% of the human population. The scheme has been modified from Nicholson DW (1999) Caspase structure, proteolytic substrates, and function during apoptotic cell death. *Cell Death and Differentiation* 6: 1028–1042.

Table 1 Caspase substrates

<i>DxxD site</i>	<i>Non-DxxD site</i>
DNA-PKs	STAT1
PARP	Sp-1
Rad 51	SRP p72
Acinus	NF- κ B
CAD/DFP45	PITSLRE kinases
DNA-RTC140	PAK-2
Rb	p59 fyn
mdm2/HDM	CaMK-IV
p21-waf1/cip	p28 Bap31
NuMA	actin
ATM	Gas2
U1-70KsnRNP	Lamin A and B
hnRNP-C1/C2	Bcl-XI
SREB	Bid
I κ B- α	APP- β
D4-GDI	pro-IL-16
cPLA2	pro-caspases
PKC- δ , - ϵ , - ζ	
MEKK-1	
Mst1	
PRK2	<i>Unknown site</i>
PP2A	wee1
FAK	CDC27
Fodrine (all)	SAF-A/hnRNP-U
Gelsolin	hnRNP-A1
Keratin-18	Ras GAP
LAP2	Raf 1
Nup153	Akt 1
Rabaptin-5	Cbl e Cbl-b
APC	PKN
Hsp90	catenin - β , - γ
UbqCE NEDD4	kinectin
Bcl 2	caplastatin
Presenilin 2	ataxin-3
Huntingtin	AMPA receptors
SBMA-AR	p27 kip1
Atrofin-1	MCM3

This list of substrates is incomplete; a more comprehensive list of over 1000 substrates, with their role in cell death, is reported in Nicholson (1999) and Fischer et al. (2003). The substrates of caspases can have single cleavage site (e.g., DQTD in Gelsolin), nested multiple sites (e.g., DEVGDVD in PARP), redundant clustered sites (e.g., DSLD-(X13)-DEED-(X16)-DLND-(X32)-DGTD in Huntingtin), or distal multiple sites (e.g., DEPD and DAVD in ICAD). See website CASBAH.

Mechanism of Action

Caspases are cysteine proteases, possessing therefore a catalytic triad which includes a cysteine (active site) in the vicinity of an histidine. The third component of the classical catalytic triad typical of cysteine proteases is not constant in caspases, and its function can be differently substituted by different groups in different components of the family. For example, if we consider caspase 3, subsequent to substrate binding, the catalysis can take place with the catalytic diad Cys285 and His237. The third element of the cysteine protease triad is in fact substituted by an ossianionic bond formed with Gly238 and Cys285 (conserved in all caspases).

Table 2 Peptide inhibitors of caspases

	<i>zVAD-fmk</i>	<i>aDEVd-cho</i>	<i>aYVAD-cho</i>
Caspase 1	2	15	1
Caspase 2	2400	1700	>10 000
Caspase 3	40	1	>10 000
Caspase 4	130	130	360
Caspase 5	5	200	160
Caspase 6	100	30	>10 000
Caspase 7	40	2	>10 000
Caspase 8	2	1	350
Caspase 9	4	60	1000
Caspase 10	-	10	400

Inhibition constant (K_i , nM) shown is $t_{1/2}$ at 1 μ M in seconds.

The table has been modified from Ekert PG, Silke J, and Vaux DL (1999) Caspase inhibitors. *Cell Death and Differentiation* 6: 1081–1086; see also Nicholson (2000).

Sequence of Action

While the regulatory caspases of the death receptor (type 8 and 10) or of the apoptosome (type 9) participate in signal transmission, the effector caspases (type 3, 6, and 7) determine the death of the cell. Effector caspases are diverse, and are not activated in parallel, but in a sequential manner, resulting in an amplification cascade of proteolytic cleavage (vaguely similar to the cascade of proteolytic activation and amplification of the complement system, or blood coagulation). These cascades have the effect of augmenting the signal. This sequence of activation is somewhat reminiscent of the complement activation or of the coagulation cascade (Figure 7).

In the intrinsic pathway, mitochondrial perturbation results in the release of cytochrome *c* to the cytosol, which, in conjunction with deoxyadenosine triphosphate/adenosine triphosphate (dATP/ATP), causes Apaf-1 oligomerization and caspase-9 activation. Caspase-9 bound within the apoptosome complex to Apaf-1 through their CARD (caspase activation and recruitment domain) domains is two to three orders of magnitude more active than free caspase-9 within the cytosol. In the apoptosome therefore, the first caspase to be activated is caspase 9. Within the apoptosome, caspase-9 processes and activates both caspase-3 and -7 by cleaving between their large and small subunits. Therefore, caspase-9, in turn, activates caspases 3 and 7. Caspase 3 is also able to proteolytically activate caspase 9, thereby creating a reciprocating loop of potentiation. Caspase 3 proteolytically activates caspases 6 and 2. At this point, caspases 3, 7, 6, and 2 are highly activated and work to proteolytically cleave their intracellular substrates and determine cell death itself. Although the initial sequential processing and activation of caspase-9, -3, and -7 are generally well established, the precise hierarchical ordering of other caspases within the caspase cascade is less clear, being subject of current investigations, particularly in cells undergoing apoptosis. A hierarchy of caspase activation was established using a model system of dATP-activated lysates to mimic events following activation of the intrinsic pathway and this paradigm is still believed to hold. Although this model of caspase activation is currently widely accepted, there are some inconsistencies between the model and the phenotype from knockout mice for caspase-3-/-, caspase-7-/-, and

Table 3 Natural caspase inhibitors

	<i>Viral</i>			<i>Cellular</i>					
	<i>crmA</i>	<i>p35</i>	<i>OplAP</i>	<i>XIAP</i>	<i>cIAP1</i>	<i>cIAP2</i>	<i>Survivin</i>	<i>NAIP</i>	<i>PI9/GBI</i>
Caspase 1	0.01	9	Inib	NI	NI	NI			NI
Caspase 2	>10 000	Inib	Inib						
Caspase 3	1600	0.1	NI	0.7	100	35	Inib	NI	
Caspase 4	1								
Caspase 5	0.1								
Caspase 6	110	0.4		NI	NI	NI			
Caspase 7	>10 000	2		0.2	40	30	Inib	NI	
Caspase 8	0.8	0.5		NI	NI	NI	NI		NI
Caspase 9	2			Inib	Inib	Inib			
Caspase 10	17	7							
Granzyme B	NI	NI							Inib

Inhibition constant (K_i , nM) shown is $t_{1/2}$ at 1 μ M in seconds.

NI, noninhibitory; Inib, inhibitory.

The table has been modified from Ekert PG, Silke J, and Vaux DL (1999) Caspase inhibitors. *Cell Death and Differentiation* 6: 1081–1086.

Table 4 Caspase knockouts

	<i>Development</i>	<i>Apoptotic phenotype</i>
Caspase 1	Normal	Inflammation, (CD95?)
Caspase 2	Normal	Germinal cells (stem cells?)
Caspase 3	Lethal PN	Neural
Caspase 6	Normal	
Caspase 7	Lethal E	
Caspase 8	lethal E	Death receptors (CD95, TNF, and DR3)
Caspase 9	lethal E	Neural
Caspase 11	normal	thymocytes, neural (CD95?)

A more comprehensive discussion is reported in Zheng et al. (1999).

Often, the severity of the phenotype depends on the genetic background of the animals, as, for example, described in caspase 3.

E, embryonically lethal; PN, perinatally lethal.

caspase-3^{-/-}/caspase-7^{-/-} double-knockout mice. The double knockout mice died shortly after birth because of a defect in cardiac development and caspase-3 seemed to play a prominent role in DNA degradation, whereas caspase-3 and -7 were important in cell viability. Furthermore, caspase-7^{-/-} mice had a relatively mild phenotype most likely because of compensation from caspase-3 and both caspase-3 and -7 were key mediators of the loss of mitochondrial membrane potential. In agreement with this, we have recently shown that caspase-6 can be processed by caspase-7 in a caspase-3 independent manner in cells undergoing apoptosis in contrast to the *in vitro* model system using dATP-activated lysates (Inue 2009). A diagram of the currently accepted order of activation is depicted in Figure 7.

Caspase 6 can also proteolytically activate the apical regulatory caspases 8 and 10, creating therefore a further loop of potentiation, which further enhances the death receptor signals.

Schematically we can redefine the molecular phases of apoptosis in the following manner (see Figure 7): (1)Initiation, activation of death receptors (caspases 8 and 10); (2) determination, activation of the apoptosome (caspase 9); (3) amplification, increase in the type and number of active caspases (caspases 3, 7, 6, and 2) and apical reactivation loop (caspases

8 and 10); and (4)demolition, proteolytic cleavage of over 1000 different substrates which determine the death of the cell.

Substrates during Apoptosis

Experimental studies indicate that over 700 different polypeptides can be generated during apoptosis. Up to 2003, according to SchultzeOstoff, over 1000 proteins have been found to be cleaved by caspases during the execution of apoptosis. Currently, over 100 substrates of caspases have been identified *in vitro*, or *in vivo*. A new website, named CASBAH, has been created with all specific detailed substrates.

Table 1 indicates some examples of the principal substrates. The proteolytic cleavage of these substrates clearly renders the cell incapable of performing its proper functions.

Inhibitors

For practical reasons, we distinguish pharmacological (Table 2) and natural (Table 3) inhibitors of caspases.

The understanding of the molecular mechanism of catalysis and of the properties of the sites P4–P1 of the substrate was the basis for designating pharmacological inhibitors, both of polypeptide type and of low-molecular-weight compounds. The electrophilic groups which interact irreversibly with the catalytic Cys are aldehydes, nitriles, and ketones, of which the latter are much more stable than the others *in vivo* and therefore are most easily applied to pharmaceuticals. Caspase inhibitors are all constituted according to the following schema: [tetrapeptide]-C)-CH₂-X, where X signifies a –Cl or –F (fluoro- or chloro-methyl ketone, abbreviated fmk), –N₂ (diazomethylketone), or –oCOR (acyloxymethylketone). The peptide portion, stabilized, determines the specific selectivity of the various caspases. Thus, zVAD-fmk is a nonspecific inhibitor of almost all of the caspases. DEVD-fmk, on the other hand, is specifically selective for effector caspases such as caspases 3, 6, and 7. Table 2 gives an idea of the specific inhibitors for different caspases.

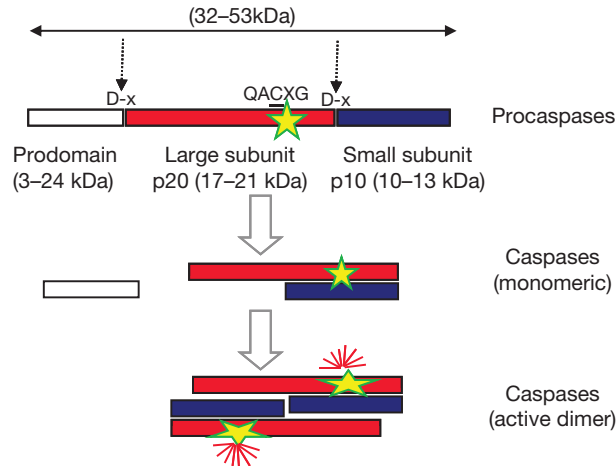


Figure 5 Structural organization and activation of caspases. The caspases are produced as pro-enzymes (32–53 kDa), including a prodomain (3–24 kDa), the large subunit (17–21 kDa) containing the active site, and the small subunit (10–13 kDa). To become enzymatically active, these three components must be proteolytically cleaved (D-x site), a phenomenon that is regulated by the prodomain itself; this allows the assembly of the large and small subunits, with the dimerization into a functional enzyme.

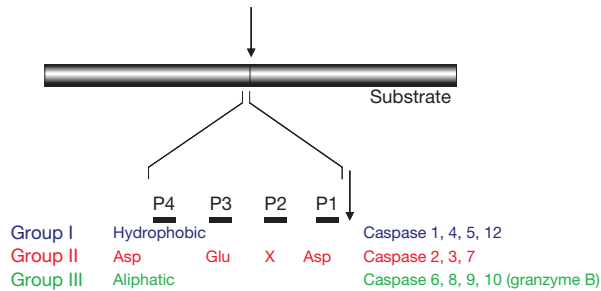


Figure 6 Caspase proteolytic specificity. The caspases recognize a tetrapeptide motif corresponding to the four residues P4–P3–P2–P1. While the position at P3 and P1 seems to be obligatory, the position P4 allows the classification of three subfamilies (see also Figure 1). This property has facilitated the identification of group-specific inhibitors.

In 1991, Lois Miller first identified a natural inhibitor of apoptosis, even before the discovery of caspases. The protein p35 produced by baculovirus (*Autographa californica*) is a potent and selective inhibitor of caspases. Subsequently, numerous other viral caspase inhibitors were identified, such as cytokine response modifier A (CrmA). Inhibition of caspases by these specific proteins is one of the important pathways among which viruses regulate apoptosis.

Numerous cellular proteins exist with similar actions to these viral caspase inhibitors. They are called inhibitors of apoptosis (IAPs). Table 3 shows the different viral and human IAPs with their relative inhibitory effects on each caspase.

All of the proteins which inhibit caspases, whether viral or cellular, are characterized by structural domains called Baculovirus IAP repeat (BIR); consequently, all IAP proteins are also called BIR proteins (BIRp). On the other hand, the contrary does not hold: there are BIR proteins which do not show IAP

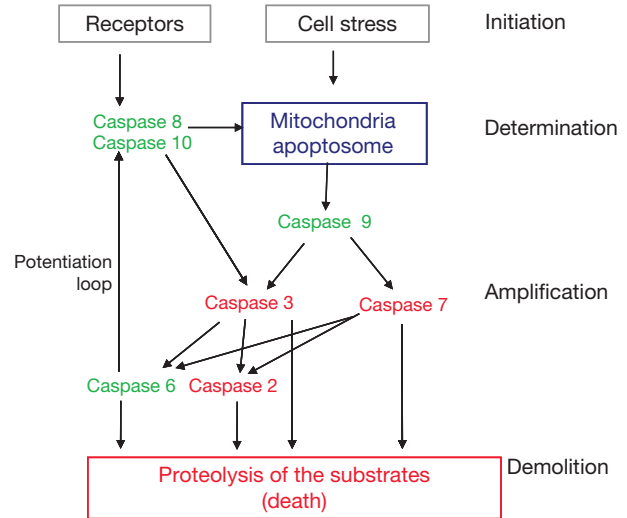


Figure 7 Sequential activation of caspases. In general, caspases with a long prodomain are involved in the upstream regulation and activation of the apoptotic pathway; they require a tight regulation and activation, and furthermore, they cleave very specific substrates. Caspases with a short prodomain are involved in the amplification of the effector cascade, thus they have a very simple, fast, and direct activation, and they also have many substrates, which inactivation finally kills the cell. Caspase 6 seems to be able to reactivate the upstream regulatory caspases, creating a feedback forward potentiation loop that strongly enhances the amplification of the death signal. Color code is in keeping with Figure 4. The scheme has been modified from Slee EA, Adrian C, and Martin SJ (1999) Serial killers: Ordering caspase activation events in apoptosis. *Cell Death and Differentiation* 6: 1067–1074.

activity. This fact is due to the prokaryotic origin of BIR proteins which are multifunctional. Many BIR proteins in humans (e.g., survivin) are passenger proteins in the mitotic spindle and are implicated in cell division.

Caspase Knockouts

The creation of transgenic mice lacking specific caspases has been accomplished for most caspases. Table 4 displays the phenotypes for these mice. Because caspases play a determining role in apoptosis, sometimes specifically, and sometimes redundantly, the phenotypes are extremely diverse, as indicated by the table.

The role of caspases in the receptor mechanism is indicated by the caspase-8 knockout.

The similarity of the phenotypes of the knockouts of caspases 3, 9, and Apaf-1 indicates the presence of a very important mechanism in the apoptosome by which the amplification of the caspase cascade is initiated and regulated.

Caspase 8 Plays a Crucial Role in the Regulation of Death Receptors

Caspase is the initiator caspase during the activation of death receptors like TNF and CD95. Here, it is a crucial component of the death initiation signaling complex, shown in some details in Figure 8(a). In addition to activate apoptosis, recent work has provided concluding evidence that caspase 8 inhibits

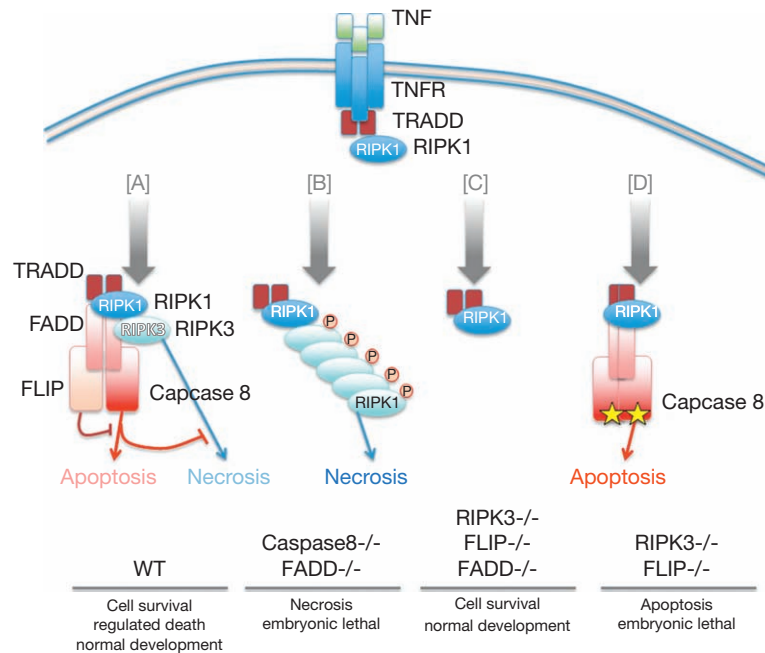


Figure 8 Caspase 8 plays a crucial role in the regulation of death receptors. Downstream of death receptors like TNF, caspase 8 is part of the death inducing signaling complex, recruited by its interaction with FADD (a). Here, it inhibits the activation of RIPK3 by RIPK1, thus regulating necrosis. At the same time, caspase 8 is bound and regulated in its pro-apoptotic function by its interaction with Flip. Consequently, both apoptosis and necrosis can result from activation of the death receptor, in a regulated manner. (a) In a normal receptor, the regulated interaction of these elements allows a controlled cell death. Wild-type mice, WT, therefore have a normal development with a regulated death pathway. (b) Knockout mice for caspase 8, caspase 8^{-/-}, or for FADD, FADD^{-/-}, or double knockouts for both show an embryonic lethal phenotype due to uncontrolled necrosis by active (phosphorylated) RIPK3. (c) Triple knockouts for FADD^{-/-}, FLIP^{-/-}, and RIPK3^{-/-} have a normal cell death and develop to a normal birth, due to the absence of necrosis (driven by RIPK3) and apoptosis (driven by FADD inability to recruit caspase 8). (d) Double knockout mice for RIPK3^{-/-}; FLIP^{-/-} die during embryonic development for uncontrolled apoptosis driven by active (starts) caspase 8. Modified from Dillon CP, Oberst A, Weinlich R, et al. (2012) Survival function of the FADD–CASPASE-8–FLIP complex. *Cell Reports* 1(5): 401–407.

necrosis by inhibiting the activation by phosphorylation of the kinase RIPK3 by the kinase RIPK1. The biological evidence for this pathway has been provided very recently by work based on multiple knockout mice. Indeed, caspase 8 knockout mice (caspase 8^{-/-}), FADD knockout mice (FADD^{-/-}), and FLIP knockout mice (FLIP^{-/-}) are embryonic lethal due to uncontrolled necrosis. Complementation with RIPK3^{-/-} reverts the lethal phenotype allowing normal development. **Figure 8** shows the effect of different double and triple knockout mice lacking different components of this crucial complex.

Pathological Implications

The execution of apoptosis is dependent on the activation of caspases, and, as a result, caspases play a determining role in all pathologies of excessive apoptosis. For example, caspases play a crucial role in the *in vivo* activation of death described in myocardial infarction and in cerebral ischemia, conditions in which cell death occurs both by apoptosis and by necrosis.

Caspases can also play an aberrant role in regulatory processes determining the major propensity or vulnerability of cells to lethal insults. Examples are the possible roles played by caspases in triplet diseases such as Huntington's disease and in neurodegeneration like in Alzheimer's disease.

In the Huntington disease, low level of intracellular activation of caspase 3 liberates an amino terminal fragment of the

protein Huntingtin (four DxxD sites in a cluster), containing the Glu expansion, facilitating a pathological aggregation. In turn, this aggregate sensitizes more caspase 3, forming a vicious cycle of activation and aggregation of poly-Glu. The cell, thus sensitized, facilitates the activation of caspase 8 (from the CD95 receptor, or directly by an effector caspase, or even stimulated indirectly by the poly-Glu aggregate), determining the death of the cell. This sensitization can also occur extremely slowly (years).

In Alzheimer's disease, caspase 3 alters the normal processing of the amyloid beta precursor protein (APP), removing the carboxy terminal. APP, deprived of the intramolecular reinternalization signal, proceeds down a route of degradation that results in the formation of the amyloid beta (Aβ) peptide, which can, in turn, favor the activation of apoptosis by means not yet clearly understood.

In either case, caspase 3 seems to augment the baseline levels (threshold?) of activation of cell death, rendering the cell more sensitive to death, and participating therefore in the pathogenetic mechanisms of these important illnesses.

Therapeutic Outlook

Selective peptide inhibitors of different caspases (see **Table 2**) have been successfully developed, even though they show

Pan-caspases	IDN-6556	Caspase inhibitor	Idun	Antiapoptotic, anti-inflammatory antifibrotic	Phase-2 for HCV-HBV infection and reperfusion
	IDN-6734	Caspase inhibitor	Idun	Reduces heart attack	Phase 1, myocard infarct
	VX-799	Caspase inhibitor	Vertex / Serono	Sepsis and neuronal death	Phase1, sepsis
	MX1013	Caspase inhibitor	Maxim	Myocardial infarct, stroke, and acute liver failure	Preclinical, myocardial infarct, stroke
	Mx-2060 derivatives	Caspase activator	Maxim	Growth inhibition	Preclinical
	M-920	Broad-spectrum caspase-inhibitor	Merck-Frosst	Effective against septic shock in a mouse model	Preclinical
	LOW MW compounds	Caspase activators	Gemin X	Caspase activators in cancer but not normal cells	Preclinical
	RGD peptides	Caspase activators	Merck, Maxim	Apoptosis induction in tumor cell lines	Preclinical
Caspase 1	IDN-11104	ICE inhibitor	Idun		Preclinical
	VX-740 (Pralnacasan)	ICE inhibitor	Vertex / Aventis	Anti-inflammatory	Phase 2, rheumatoid arthritis
	VX-756	ICE inhibitor	Vertex	Anti-inflammatory	Phase 2 for 2004
Caspase 3	M-826	Reversible caspase-3 inhibitor	Merck-Frosst	Protects hypoxia-ischemia, and Huntington's model	Preclinical
	M-791	Caspase-3 specific inhibitor	Merck-Frosst	Effective against septic shock in a mouse model	Preclinical
	Immunocasp-3	Cell-permeable HER2 mAb fused to caspase-3	Jia et al. (2003)	Growth inhibition of HER2-positivetumors in a mouse xenograft model	Preclinical
	Ad-G/iCasp3	Adenoviral induc. caspase-3	Shariat et al. (2001)	Reduces tumor growth in prostate cancer	Preclinical
	PEF-F8-CP3	Caspase-3 fusion with antibody	Tse and Rabbitts (2000)	Antigen-dependent induction of apoptosis	Preclinical
Caspase 6	Immunocasp-6	HER2 mAb fused to caspase-6	Xu et al. (2004)	Growth inhibition of tumors	Preclinical
Caspase 9	FKBP 12/casp-9 fusion prot	Dimerization of caspase-9	Nor et al. (2002)	Anti-angiogenic upon caspase-9 dimerization	Preclinical

Figure 9 Clinical trials based on Caspase regulators. The caspases' active site has been targeted to develop specific inhibitors. The figure lists some of the clinical trials that have been carried out with caspase inhibitors, just to provide a flavor of the current clinical perspectives.

serious problems penetrating the cell membrane, as well as having electrophilic promiscuity and stability in solution which make them able to be attacked by biological nucleophiles such as cathepsin. While this makes them not easily adapted to clinical use, they have allowed scientists to obtain many interesting *in vitro* results. More recently, Merck (like SmithKline Beecham, BASF, Idun/Novartis, and Vertex) has in advanced stages of development new classes of low-molecular-weight nonpeptide inhibitors. These show a much higher potency and specificity, opening new therapeutic perspectives. Nonpeptide inhibitors such as isatins attain potencies 100–1000 times greater than the class of zVAD-fmk, with substantial improvements *in vivo*. These studies are still in early stages (see [Figure 9](#)), and we hope that the near future will tell as the best clinical application for this new class of drugs.

See also: [Bioenergetics: Cell Death by Apoptosis and Necrosis](#); [Cell Architecture and Function: Bcl2 Family: Cell Death Enhancers and Inhibitors](#); [Cell Cycle: Control of Entry and Progression through S Phase](#); [Signaling: p53 Family](#).

Further Reading

- Black RA, Kronheim SR, and Sleath PR (1989) Activation of interleukin 1 β by a co-induced protease. *FEBS Letters* 247: 386–390.
- Davis BK, Wen H, and Ting JP (2011) The inflammasome NLRs in immunity, inflammation, and associated diseases. *Annual Review of Immunology* 29: 707–735.
- Dillon CP, Oberst A, Weinlich R, et al. (2012) Survival function of the FADD –CASPASE–8 –FLIP complex. *Cell Reports* 1(5): 401–407.

- Ekert PG, Silke J, and Vaux DL (1999) Caspase inhibitors. *Cell Death and Differentiation* 6: 1081–1086.
- Fischer U, Laénicke RU, and Schultze-Ostf K (2003) Many cuts to ruin: A comprehensive update of caspase substrates. *Cell Death and Differentiation* 10: 76–100.
- Galluzzi L, Vitale I, Abrams JM, et al. (2011) Molecular definitions of cell death subroutines: Recommendations of the Nomenclature Committee on Cell Death 2012. *Cell Death and Differentiation*. doi:10.1038/cdd.2011.96.
- Gunther C, Martini E, Wittkopf N, et al. (2011) Caspase-8 regulates TNF- α -induced epithelial necroptosis and terminal ileitis. *Nature* 477: 335–339.
- Inoue S, Browne G, Melino G, and Cohen GM (2009) Ordering of caspases in cells undergoing apoptosis by the intrinsic pathway. *Cell Death and Differentiation* 16(7): 1053–1061.
- Kaiser WJ, Upton JW, Long AB, et al. (2011) RIP3 mediates the embryonic lethality of caspase-8-deficient mice. *Nature* 471: 368–372.
- Kanneganti TD (2010) Central roles of NLRs and inflammasomes in viral infection. *Nature Reviews Immunology* 10(10): 688–698.
- Kostura MJ, Tocci MJ, Limjuco G, et al. (1989) Identification of a monocyte specific pre-interleukin 1 β convertase activity. *Proceedings of the National Academy of Sciences of the United States of America* 86: 5227–5231.
- Lüthi AU and Martin SJ (2007) The CASBAH: A searchable database of caspase substrates. *Cell Death and Differentiation* 14(4): 641–650.
- Melino G (2001) The Syren's song (concept: apoptosis). *Nature* 412: 23.
- Nicholson DW (1999) Caspase structure, proteolytic substrates, and function during apoptotic cell death. *Cell Death and Differentiation* 6: 1028–1042.
- Nicholson DW (2000) From bench to clinic with apoptosis-based therapeutic agents. *Nature* 407: 810–816.
- Oberst A, Dillon CP, Weinlich R, et al. (2011) Catalytic activity of the caspase-8-FLIP(L) complex inhibits RIPK3-dependent necrosis. *Nature* 471: 363–367.
- Oberst A and Green DR (2011) It cuts both ways: reconciling the dual roles of caspase 8 in cell death and survival. *Nature Reviews Molecular Cell Biology* 12: 757–763.
- Slee EA, Adrian C, and Martin SJ (1999) Serial killers: Ordering caspase activation events in apoptosis. *Cell Death and Differentiation* 6: 1067–1074.
- Wang J, Zheng L, Lobito A, et al. (1999) Inherited human caspase 10 mutations underlie defective lymphocyte and dendritic cell apoptosis in autoimmune lymphoproliferative syndrome type II. *Cell* 98: 47–58.
- Welz PS, Wullaert A, Vantis K, et al. (2011) FADD prevents RIP3-mediated epithelial cell necrosis and chronic intestinal inflammation. *Nature* 477: 330–334.
- Yuan J, Shaham S, Ledoux S, Ellis HM, and Horvitz HR (1993) The *C. elegans* cell death genes *ced-3* encodes a protein similar to mammalian interleukin 1 β -converting enzyme. *Cell* 75: 641–652.
- Zhang H, Zhou X, McQuade T, et al. (2011) Functional complementation between FADD and RIP1 in embryos and lymphocytes. *Nature* 471: 373–376.
- Zheng TS, Hunot S, Kuida K, and Flavell RA (1999) *Caspase knockouts: Matter of life and death*, 6: 1043–1053.

Relevant Websites

- <http://bioinf.gen.tcd.ie> – CASBAH provides a comprehensive list of caspase substrates as a searchable web resource containing information pertaining to all currently known caspase substrates.
- www.nature.com – The same editorial team have now launched a new scientific journal, *Cell Death and Disease*, a peer-reviewed online journal in the field of translational cell death.
- www.nature.com – The scientific journal *Cell Death and Differentiation*: scientific excellence in the cell biology, molecular biology, and biochemistry of cell death and disease. *Cell Death and Differentiation* provides a unified forum for scientists and clinical researchers, publishing high-quality original papers together with topical reviews and special issues.
- <http://deathbase.org> – This is a database of proteins involved in cell death processes. Updated continuously by a team of expert curators whose areas of expertise cover all proteins involved in cell death in five model species, DeathBase provides a useful tool for anyone with an interest in apoptosis and cell death.
- <http://p53.free.fr> – This is a reference for all scientists working on p53, including information on p53 structure and function, the description of all p53 monoclonal antibodies, and the phylogenetic analysis of p53. The p53 website also hosts the UMD p53 mutation database that can be downloaded for clinical or epidemiological analysis.
- <http://bcl2db.ibcp.fr> – This website holds a collection of sequence, structural and functional information about one of the most-studied protein families in cell biology, enabling researchers to readily retrieve Bcl-2 family sequences and related data.

Catalysis by RNA, Structural Themes: Group I Introns

B L Golden, Purdue University, West Lafayette, IN, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Exons Segments within a gene that encode a functional cellular protein or RNA.

Introns Noncoding DNA sequences that interrupt functional genes and are removed by

splicing once the gene has been transcribed into RNA.

RNA splicing Process by which intervening sequences are removed from a transcribed RNA and the flanking exons are joined to make a functional cellular RNA.

The Self-Splicing Reaction

The self-splicing reaction of group I introns is accomplished by two sequential phosphotransesterification reactions (**Figure 1(a)**). In the first step of splicing, a guanosine nucleoside substrate (guanosine, guanosine monophosphate, guanosine diphosphate, or guanosine triphosphate) is bound by the intron and positioned for nucleophilic attack at the 5' splice site. The 3'-OH of the guanosine attacks at the phosphorus of the 5' splice site and displaces the 3'-OH of the 5' exon. This reaction results in release of the 5' exon and covalent attachment of the guanosine molecule to the intron. In the second step of splicing (**Figure 1(a)**), a conserved guanosine at the 3' end of the intron (called ω G) is positioned by the intron in a manner similar to binding of the exogenous guanosine substrate in the first step of splicing. In a reaction chemically analogous to the reverse of the first step of splicing, the 3'-OH of the 5' exon displaces the 3'-OH of ω G from the phosphate at the 3' splice site. Ligation of the 5'- and 3'-exons and release of the intron results.

Secondary Structure

Structured ribonucleic acids (RNAs), such as transfer RNAs, ribosomal RNAs, and ribozymes, do not have an extended linear structures; they are folded. Although not completely double-stranded-like genomic DNA, they have short base-paired regions formed from self-complementary sequences within the primary structure of the RNA. Group I introns have a conserved secondary structure consisting of a series of double-stranded regions labeled P1–P9 (**Figure 2(a)**). These helical regions are linked by 'single stranded', or joining, stretches of RNA labeled for the two helices that they connect (e.g., J8/7) and are capped by loops designated with an 'L'. Although most joining and loop regions are not involved in forming Watson–Crick base pairs, they usually form well-defined structures, including base pairs. These helices, loops and joining nucleotides are organized into three domains designated P1–P2, P3–P9, and P4–P6 (**Figure 2(a)**).

The P1–P2 domain, or substrate domain, spans the 5' end of the molecule and contains the 5' splice site, which is located within the short helix P1. The final nucleotide in the 5' exon is a uridine, which is base paired to guanosine on the

complementary RNA strand of the intron, forming a G•U wobble base pair. The P3–P9 domain immediately follows the P1–P2 domain in the primary structure of the intron. This domain contains the helix P7 that binds to an exogenous guanosine substrate in the first step of splicing, and to ω G in the second step of splicing. The P3–P9 domain also contains a single-stranded stretch of 5–6 nucleotides called J8/7 that is involved in binding and positioning the P1 helix, and thereby the 5' splice site. The P3–P9 domain is interrupted by the P4–P6 domain. The P4–P6 domain, folds together with the P3–P9 domain, to comprise most, if not all, of the catalytic core of the intron. This domain also participates in splice-site recognition by binding the guanosine within the G•U wobble base pair at the 5' splice site.

Tertiary Structure

The secondary structures in the intron fold in a magnesium-dependent manner to form a compact tertiary structure much like a globular protein enzyme (**Figure 2(b)**). Formation of the folded structure is absolutely dependent on divalent metal ions, usually Mg (II), which serve to stabilize the structure by neutralizing the negative charges of the phosphate groups on the RNA backbone and by organizing the RNA into specific structures required for proper folding. In crystal structures of group I introns and their domains, magnesium ions are seen bound within the major groove, mediating the close approach of RNA helices, and stabilizing unusual conformations that allow the formation of intramolecular RNA–RNA contacts. Folding of the RNA creates binding sites for the P1 helix and the guanosine nucleotide cofactor described below.

The tertiary structure of group I introns has been extensively investigated by X-ray crystallography. These studies have provided high-resolution structural information for one of the domains and medium-resolution for three different group I introns. Complementary work has been performed using solution biochemistry to probe the rate of the chemical reactions of vast array of group I introns containing single atom modifications. These data can be put together to help elucidate the structural basis for group I intron catalysis.

The crystal structures of the group I introns reveal many of the strategies used by RNAs to form tertiary structures. Perfectly double-stranded helical RNAs do not tend to form stable tertiary structures. The structural information encoded

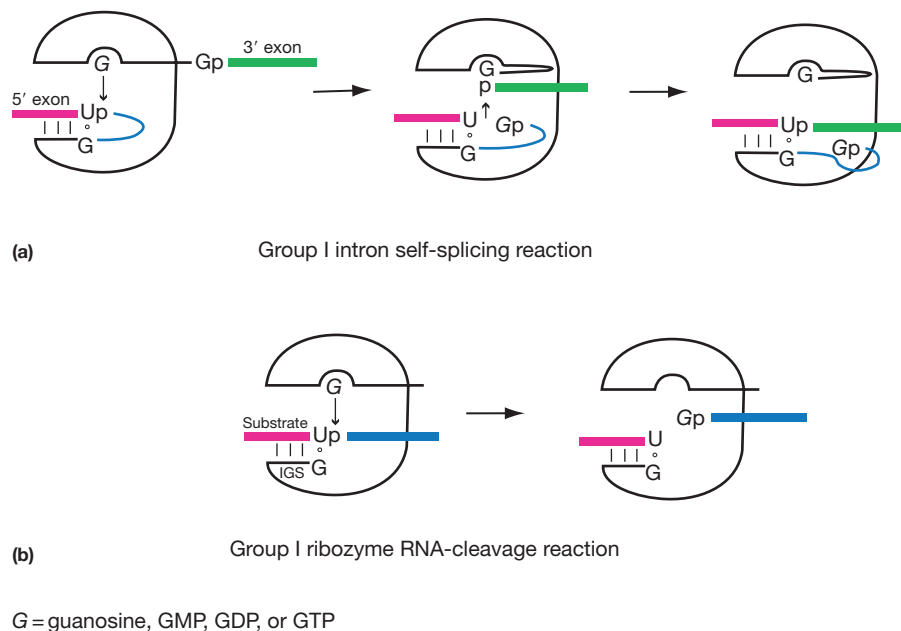


Figure 1 Reactions of a group I intron RNAs. (a) The self-splicing reaction is initiated by binding of a guanosine nucleoside. The 3'-OH of the nucleoside is activated for attack at the 5' exon–intron junction. In the first step of splicing, the guanosine becomes covalently attached to the intron, and the 5' exon (pink) is cleaved from the intron (blue). In the second step of splicing, the 3'-OH of the 5' exon attacks at the junction of the intron and the 3' exon (green). This reaction links the 5' and 3' exons and releases the intron. (b) The intron can be converted into a ribozyme capable of cleaving an RNA substrate using a guanosine nucleoside. The substrate RNA, like the 5' exon in the self-splicing molecule, can base pair to the 5' end of the ribozyme that is called the internal guide sequence (IGS).

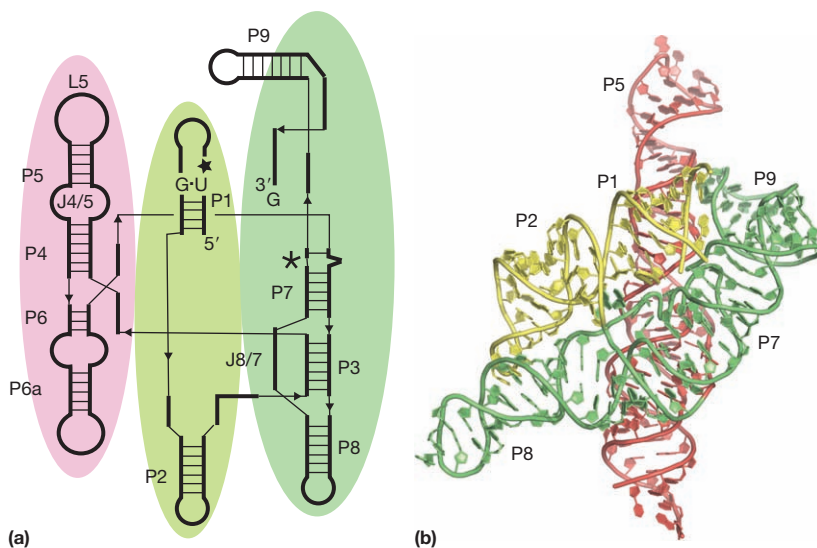


Figure 2 Secondary and tertiary structure of a group I intron. (a) The sequence of a group I intron folds into a secondary structure containing a series of base-paired regions called P1–P9. These short duplexes are linked by 'single-stranded' nucleotides designated 'J' and capped with loop sequences designated 'L'. These secondary structural elements are organized into three domains termed P1–P2, P3–P9, and P4–P6 that are shaded yellow, green, and pink, respectively. The P1 helix is formed by base pairing between the 5' end of the ribozyme and the 5' intron. Folding of the ribozyme brings the 5' splice site (*) and the guanosine substrate-binding site (*) into proximity. (b) The three-dimensional structure of a group I intron from *Azoarcus*. The P1–P2 domain (yellow) is bound and positioned by cooperative action of the P4–P6 (red) and P3–P9 (green) domains.

by the RNA sequence is sequestered within the Watson–Crick base pairs of a perfect helix, and the exposed, accessible surface is limited to the backbone and the shallow minor groove.

To facilitate tertiary structure formation, 'flaws' in the helices are often present, containing base pairs other than the standard Watson–Crick A–U and G–C base pairs. These can readily form in between helices and, when present, they

distort the major and minor grooves of the double-stranded helix to create distinct shapes and to expose hydrogen-bonding groups on the bases. This forms complementary surfaces that allow RNA domains to pack together. Loops, particularly four nucleotide sequences called tetraloops, which cap helices can also form distinct structures that mediate long-range tertiary interactions. These interactions are further stabilized by hydrogen bonds that can form between the 2'-OH groups of the ribose backbone. These tertiary structures are termed 'ribose zippers'.

Substrate Recognition

Group I introns have binding sites for two substrates: a guanosine nucleophile and a short duplex containing a G•U wobble base pair.

The Guanosine-Binding Site

The guanosine substrate (in the ribozyme reaction and in the first step of splicing) and ω G (in the second step of splicing) are recognized by a single guanosine-binding site located in the major groove of helix P7, a distorted double helix with a narrowed major groove. A G–C base pair near the 'top' of the P7 helix provides significant specificity for the guanosine base (Figures 3(a) and 3(b)). Recognition of the guanosine base can largely be described as a base triple. The recent crystal structures reveal that this base triple is stacked between two additional base triples, forming a snug-binding pocket for the guanosine substrate. The orientation of the ribose group of the guanosine comes from magnesium ions bound within the intron's active site (Figure 3(b)).

Helix P1 and 5' Splice-Site Recognition

The 5' splice site is recognized in the context of a short double-stranded region called P1. In self-splicing introns, helix P1 is formed by intramolecular base pairing between the 5' exon and a complementary sequence near the 5' end of the intron, creating a stem-loop structure (Figure 2(a)). The helix is 3–6 base pairs in length, depending on the specific intron, and the final base pair is a G•U wobble pair composed of the last nucleotide in the 5' exon, and a guanosine near the 5' end of the intron. G•U wobble pairs, like A–U base pairs, contain two hydrogen bonds between the two bases (Figures 3(a) and 3(b)), but the geometry of the pair is slightly altered compared to a Watson–Crick base pair: the guanosine base is pushed slightly into the minor groove and the uridine base is pushed slightly into the major groove. This base pair can be readily accommodated within an RNA double helix and is about as stable as an A–U base pair.

The intron may be converted into a ribozyme by splitting the P1 helix into two molecules (Figure 1(b)). The first strand is formed from the very 5' end of the ribozyme and is called an internal guide sequence (IGS; Figure 1(b)). The second strand serves as a substrate for the ribozyme. It is complementary to the first strand, maintains the G•U wobble base pair, and extends for at least one nucleotide beyond the cleavage or splice site.

P1, whether composed of a hairpin in the self-splicing intron or composed of an intermolecular duplex in a ribozyme form, docks into the active site of the ribozyme. Docking is stabilized by interactions between the single-stranded regions J2/3 and J8/7 and the minor groove of P1. These contacts are primarily hydrogen bonds to the ribose backbone and, therefore, are sequence independent. Except for the G•U wobble

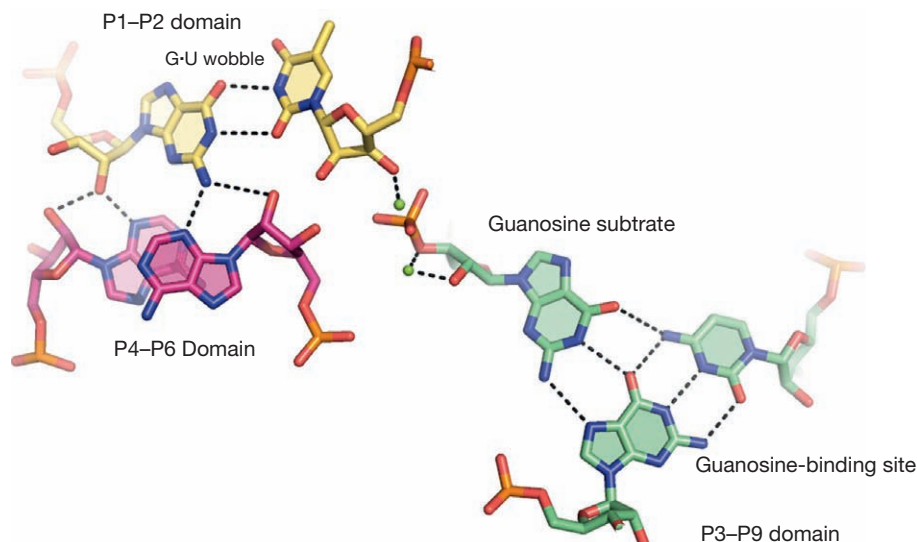


Figure 3 Substrate interactions with the group I intron core. (a) Interactions between the three domains in the transition state. Near the transition state for the ribozyme reaction, the bond between the 5' exon and the intron is partially broken and the bond between the guanosine substrate and the intron is partially formed. At least two magnesium ions help orient the substrates and stabilize the transition state, facilitating catalysis. The G•U wobble at the 5' splice site is shaded yellow. The guanosine bound in the guanosine-binding site is shaded green. Adenosines within the P4–P6 domain involved in positioning the G•U wobble are shaded pink. Magnesium ions that participate in catalysis are drawn as spheres. Hydrogen bonds and metal interactions are indicated by dashed lines and partial bonds are indicated by dotted lines. (b) The crystal structures of group I introns reveal the interactions that orient the G•U wobble that defines the 5'-cleavage site and the guanosine-nucleotide substrate.

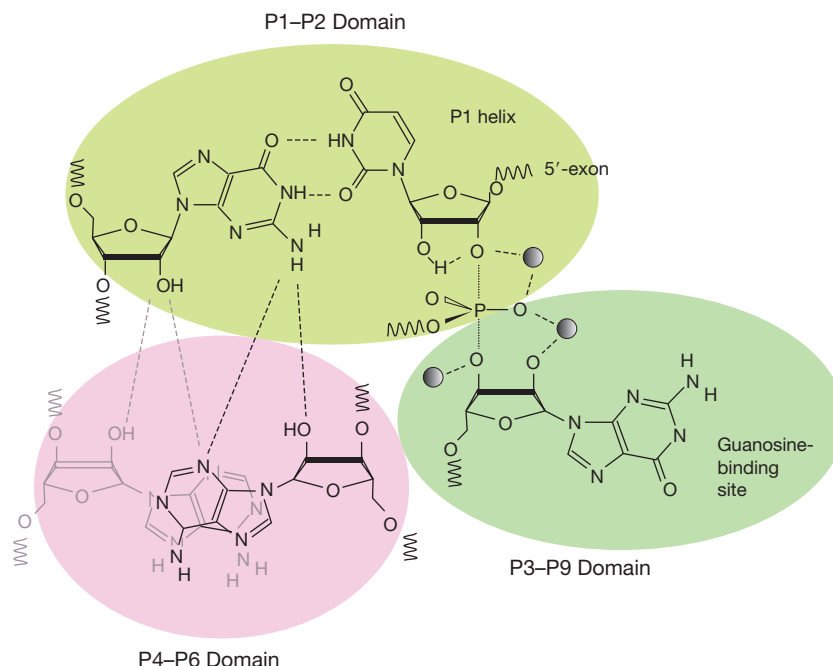


Figure 4 Interactions between the three domains in the transition state. Near the transition state for the ribozyme reaction, the bond between the 5' exon and the intron is partially broken and the bond between the guanosine substrate and the intron is partially formed. Three magnesium ions help orient the substrates and stabilize the transition state, facilitating catalysis. The G•U wobble at the 5' splice site is shaded yellow. The guanosine bound in the guanosine-binding site is shaded green. Adenosines within the P4–P6 domain involved in positioning the G•U wobble are shaded pink. Magnesium ions that participate in catalysis are drawn as spheres. Hydrogen bonds and metal interactions are indicated by dashed lines, and partial bonds are indicated by dotted lines.

base pair, identity of base pairs within the P1 helix is not critical for recognition. Thus, by appropriate variation of the nucleotides in the IGS, almost any RNA substrate containing a uridine can be recognized and cleaved by a group I ribozyme.

The guanosine of the G•U wobble base pair is recognized by the J4/5 region of the P4–P6 domain. The J4–5 region contains tandem A•A non-Watson–Crick base pairs that interact with the exocyclic amine and the 2'-OH of the guanosine in the wobble pair (Figures 3(a) and 3(b)). With the ribozyme 'pinched' down upon the guanosine in the wobble pair, the phosphate that follows the uridine is placed in an optimal position to interact with the guanosine nucleophile and the intron's active site. The requirement for a G•U instead of a G–C Watson–Crick base pair at this position can be attributed to two characteristics of this special base pair. First, the exocyclic amine of the guanosine is displaced into the minor groove, allowing it to hydrogen bond to the J4/5 region. The exocyclic amine of a guanosine within a G–C base pair is involved in hydrogen bonding to the cytosine and is, therefore, sterically occluded from this interaction. Second, if the guanosine base in a G•U base pair is superposed on the guanosine G–C base pair, the position of the pyrimidine (U or C), and, therefore, the phosphate at the 3' position, is quite different in the two structures. Displacement of the uridine toward the major groove as occurs in a wobble base pair is required to place the phosphate at the 5' splice site into the active site of the intron.

In summary, group I introns may be thought of as RNA restriction endonucleases, capable of recognizing double-

stranded RNA and specific for cleavage at G•U wobble base pairs. The catalytic activity of group I introns is dependent on proper folding of the RNA into a stable three-dimensional structure containing three domains. The three domains act cooperatively to position the 3'-OH of the guanosine nucleophile, the 5' splice-site phosphate, and at least two magnesium ions bound in the heart of the ribozyme (Figure 4), and thereby promote catalysis.

See also: **Molecular Biology:** DNA Secondary Structure; Riboswitches; Ribozyme Structural Elements: Group II Introns and the Spliceosome; Ribozymes and Evolution.

Further Reading

- Adams PL, Stahley MR, Kosek AB, Wang J, and Strobel SA (2004) Crystal structure of a group I intron splicing intermediate. *Nature* 430: 45–50.
- Golden BL (2007) Group I introns: Biochemical and crystallographic characterization of the active site structure. In: Lilley DJM and Eckstein F (eds.) *Ribozymes and RNA Catalysis*, pp. 178–201. Cambridge: Royal Society of Chemistry, Cambridge.
- Golden BL, Kim H, and Chase E (2005) Crystal structure of a phage Twort group I ribozyme-product complex. *Nature Structural and Molecular Biology* 12: 82–89.
- Guo F, Gooding AR, and Cech TR (2004) Structure of the *Tetrahymena* ribozyme: Base triple sandwich and metal ion at the active site. *Molecular Cell* 16: 351–362.

Cell Cycle: Control of Entry and Progression through S Phase

S L Forsburg, University of Southern California, Los Angeles, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Cyclin-dependent kinase (CDK) A family of kinases that requires binding to a regulatory subunit called a cyclin to be active against exogenous substrates.

DDK DBF4-dependent kinase Cdc7, required for replication initiation.

Go-Ichi-Ni-San (GINS) Initiation complex, named for the Japanese Go-Ichi-Ni-San, or 5-1-2-3 after the names of the genes.

Origin Site of replication initiation in the genome; usually intergenic, but often has no defining sequence.

preRC Pre-replication complex assembled before origin firing comprises ORC, Cdc6, Cdt1, and MCM proteins.

The process of DNA synthesis during S phase requires regulation at multiple levels to ensure that DNA replication is coordinated with overall cell-cycle progression. There are numerous mechanisms to ensure that S phase occurs in a timely and regulated fashion. These ensure that the onset of replication occurs in response to the appropriate signals, and they also prevent the re-replication of the genome in a single cell cycle. Much of this regulation occurs by controlling the activation of DNA synthesis at individual origins of replication.

The study of S phase and its many levels of control is continuing in many laboratories. This work has benefited from genetic systems such as the yeasts *Saccharomyces cerevisiae* and *Saccharomyces pombe*, and the fruit fly *Drosophila*. Biochemical studies, primarily in *Xenopus* extracts, complement early work that reconstituted the replication apparatus using viral systems and human cells. Given this long and distinguished history, there are inevitably different names for the conserved proteins involved. Table 1 provides a listing of the equivalents that are most relevant to this article.

Initiation and Elongation of DNA Replication Origins

The major control of DNA replication occurs at initiation, by regulating assembly and activation at the replication origins of the multiple proteins of the pre-replication complex (preRC). Following initiation, the enzymes of the DNA synthesis machinery take over and extend replication forks to duplicate the genome.

Origins

DNA synthesis begins at replication origins. Generally, there are many replication origins in each chromosome, to facilitate the timely duplication of the genome. In a few species, the origin is defined by a consensus sequence in the DNA. More commonly, origins are defined by general features of the sequence rather than a linear consensus, and are typically A/T rich and easily unwound. In some cases, origins appear to be distributed not by sequence, but by a spacing mechanism. Not all origins are activated at the same time: some regions of

the genome may undergo initiation early in S phase, others are activated late. Moreover, not all origins are used in every cell cycle, leading to a stochastic distribution of origin activity. Some of this may reflect differences in chromatin structure and access, possibly depending on histone acetylation. Despite these differences in origin function, the proteins responsible for replication initiation at individual origins are remarkably similar among eukaryotes, and many of these factors can also be found in the Archaea.

Assembly of the preRC

The preRC is assembled by sequential binding of conserved proteins (Figure 1). This occurs at the end of M phase in cycling cells. First, the six proteins of the origin recognition complex (ORC) bind DNA near what will become the initiation site in the replication origin. Next, Cdc6 (called Cdc18 in fission yeast) and Cdt1 bind. These two proteins are highly regulated so that they are only available for this function at the beginning of S phase. It is likely that Cdc6 acts as a clamp loader to assemble the six-membered ring of proteins called the 'mini-chromosome maintenance (MCM)' complex (Mcm2–7). Biochemical studies suggest that once the MCM complex is loaded, the other preRC components are no longer required for replication initiation. Curiously, MCMs are extremely abundant proteins, far outnumbering the replication origins, and it remains to be determined why the cell requires so many of them (the MCM paradox). One model suggests that the spreading of MCMs across the unreplicated chromatin serves to distribute sites of replication, but this remains unproven. The assembled preRC is now poised for initiation of the origin. It is likely that more origins are assembled than will be needed, probably as a backup mechanism in the case of DNA damage or replication fork defects.

Initiation

Initiation at each origin occurs when several activating proteins are recruited sequentially to the assembled preRC in response to specific kinase activity (see section Regulation of Origin Firing by Kinases). Although the details may differ slightly from system to system, there are broad similarities. In all

Table 1 Proteins required for S-phase initiation

<i>S. cerevisiae</i>	<i>S. pombe</i>	Metazoans	Product
<i>preRC assembly</i>			
Orc1-6	Orp1-6	Orc1-6	ORC
Cdc6	Cdc18	Cdc6	Activator
Tah1	Cdt1	Cdt1/Dup	preRC factor
Mcm2-7	Mcm2-7	Mcm2-7	MCM complex members
No homolog ^b	No homolog ^b	Geminin	Cdt1 inhibitor
<i>Initiation</i>			
Sld2	Drc1/Sld2	No homolog ^b	Loading factor
Dpb11	Rad4/Cut5	TopBP1/Mus101	Possible assembly factor?
Sld3	Sld3	No homolog ^b	Loading factor
GIN5 (Psf1, Psf2, Psf3, Sld5)	GIN5 (Psf1, Psf2, Psf3, Sld5)	GIN5 (Psf1, Psf2, Psf3, Sld5)	Initiation factor
Mcm10	Cdc23	Mcm10	Initiation factor
Cdc45	Sna41/Cdc45	Cdc45	Initiation factor
No homolog ^b	No homolog ^b	Treslin	Loading factor?
<i>Elongation</i>			
RFC	RFC	RFC	Replication factor C: PCNA clamp loader
Cdc17/Pol1	Swi7/Pol1	Pol α	DNA pol alpha
Cdc2/Pol3	Cdc6/Pol 3	Pol δ	DNA pol delta
Pol 2	Cdc20/Pol2	Pol ϵ	DNA pol epsilon
Cdc44	Pcn1	PCNA	Clamp
RPA	RPA	RPA	Replication Protein A: single strand DNA binding protein
Ctf4	Mcl1	AND-1	Replication fork linkage
Mrc1	Mrc1	Claspin	Replication fork linkage
<i>Regulatory kinases</i>			
Cdc7	Hsk1	Cdc7	DDK kinase
Cdc28	Cdc2	Cdk2 and Cdk4	S phase CDK
Sic1 ^a	Rum1 ^a	p27 ^{Kip1a}	Nonhomologous CKIs

The same protein may have synonyms even in one species (indicated by a slash).

^anot related by sequence.

^bsequence homolog not (yet) identified.

cases, the Cdc45 protein is recruited, probably by a loading factor sensitive to Cdc7 and cyclin-dependent kinase (CDK) kinase. Subsequently, a complex including the BRCA1 C-terminal (BRCT)-domain protein Dpb11/Rad4/TopBP1 is recruited and helps load DNA polymerase epsilon and the go-ichi-ni-san (GIN5) proteins. The loading factors and Dpb11/Rad4/TopBP1 then leave the complex, but Cdc45 and GINs activate and travel with the MCM helicase. This creates a core unwinding complex called variously the 'replication progression complex (RPC)' or 'Cdc45-MCM-GIN5 (CMG)'. This recruitment allows initial unwinding of the DNA and binding by the single-stranded-DNA-binding protein replication protein A (RPA). This is the essential first step that leads to the full replisome.

Although Cdc45 and Dpb11/Rad4/TopBP1 are conserved, the other loading factors are not. The yeasts use proteins called Sld2 and Sld3. Vertebrate cells appear to undergo a similar regulation, but the likely candidates are not strongly related to the yeast proteins (RecQ4 and treslin).

Elongation

Elongation occurs as the cell couples DNA polymerases to the unwinding apparatus. The initial DNA synthesis relies on DNA polymerase alpha and an RNA primase. As the leading strand is elongated, a polymerase switch occurs, and the highly processive DNA polymerase epsilon assumes the job of DNA

synthesis. It is locked into place by a clamp called proliferating cell nuclear antigen (PCNA), which is loaded onto the DNA by a clamp-loader called replication factor C (RFC). Replication proceeds bidirectionally from the origin, with each leading strand generating a replication fork structure. While DNA polymerase epsilon elongates the leading strand, the opposite lagging strand is synthesized in a series of short fragments. These are initiated by DNA polymerase alpha and extended by DNA polymerase delta. These Okazaki fragments are joined together by a group of processing enzymes including DNA ligase, to create a single continuous strand.

In order to maintain the attachment between the DNA polymerases and the unwinding activity of the helicase, several additional proteins are recruited to the replisome as linkers. These proteins, AND-1/Ctf4 and Claspin/Mrc1, contribute to the stability of the replication fork and prevent the helicase from unwinding too far ahead of the polymerases. Loss of these proteins makes the cells more sensitive to DNA-damaging agents that result in replication fork stalling or pausing.

Processive DNA replication may be interrupted if the polymerase encounters lesions in the template, or if it runs out of nucleotide precursors. Checkpoint response mechanisms and DNA repair factors can be recruited to repair the defect and allow replication to continue. These are discussed in other articles.

Little is known about the resolution of two colliding replication forks from adjacent origins, or the termination of

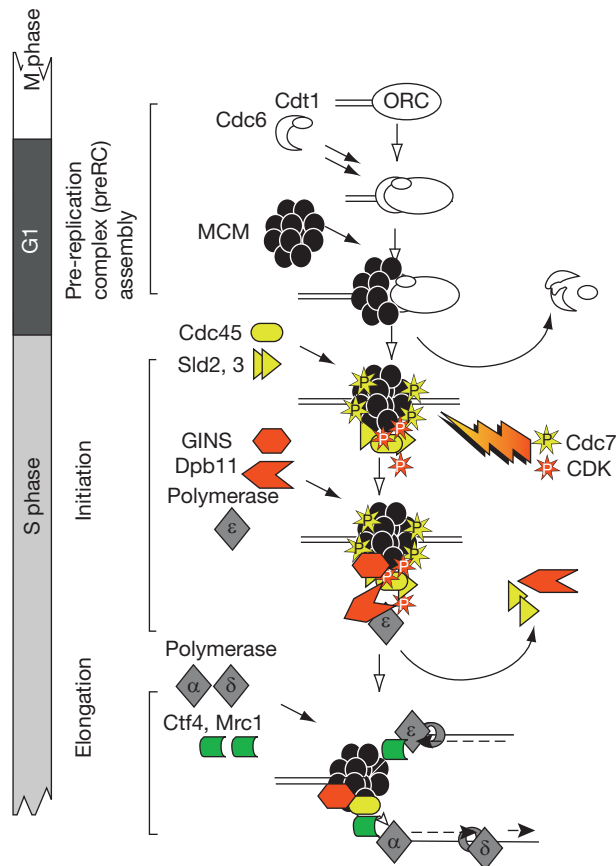


Figure 1 Ordered assembly of the budding yeast pre-replication complex (pre-RC) is a model for eukaryotes and occurs during late M and G1 phase. Initiation at individual replication origins occurs during S phase when the MCM helicase is activated. Phosphorylation is indicated by stars and results in recruitment of DNA synthesis factors. Loading factors Sld2 and 3 and Dpb11 assemble the helicase processivity factors Cdc45 and GINS. Elongation occurs with the recruitment of polymerases: leading strand (pol ϵ), lagging strand (pol δ), and primase (pol α), and coupling factors Ctf4 and Mrc1.

replication. However, studies show that the MCM proteins are dislodged from the chromatin as S phase progresses, and by the conclusion of DNA synthesis they are no longer DNA associated.

Regulation of Origin Firing by Kinases

Activation of DNA synthesis at the assembled preRC requires the activity of at least two kinases: a CDK and a related kinase called Cdc7 (DDK). Investigators are still examining the molecular effect of these kinases, but a general model for their function has been determined (Figure 2).

Cdc7 (DDK)

The Cdc7 kinase is essential for initiation of DNA replication. It acts at individual replication origins to initiate DNA synthesis and, thus, is rate limiting for initiation. Its activity requires

binding to a regulated subunit, originally identified as Dbf4, which is required for the catalytic Cdc7 subunit to recognize its substrates and target it to the preRC. Dbf4 is regulated both transcriptionally and posttranslationally and restricts kinase activity to the S phase of the cell cycle. The Cdc7 kinase is sometimes called a DDK, or Dbf4-dependent kinase, in recognition of its similarity to cyclin-dependent kinases. *In vitro* data indicate that Cdc7 kinase can phosphorylate components of the preRC and initiation proteins, including MCM proteins, Cdc45, and DNA polymerase alpha.

A change-of-function mutation in *S. cerevisiae* MCM5 was isolated that completely bypasses the requirement for Cdc7 kinase. This led to the suggestion that phosphorylation of the MCM complex by Cdc7 results in a conformational change in the complex required for activation. Recent studies in budding yeast suggest that Cdc7 phosphorylation of the N-terminus of Mcm4 is required for MCM activation, although this has not been demonstrated for other systems. However, together the data suggest that Cdc7 activity is required for assembly of Cdc45. As described above, Cdc45 binding is a prerequisite for assembly of RPA and primase, and is thought to be limiting for replication. Thus, Cdc7 can be thought of as activating the proteins at the individual origins.

Cdc7 also has a variety of other substrates in S phase. It is required for proper timing of centromere DNA replication, centromere cohesion, and response to DNA-damaging agents. It also functions in meiosis to regulate programmed recombination and cohesin. These different activities can be mapped onto different domains of the Dfp1 protein.

Cyclin-Dependent Kinases

CDKs are the engine of the cell cycle and their oscillation controls multiple events. In yeasts, a single kinase subunit switches cyclin partners during the cell cycle to provide substrate specificity. In metazoans, multiple kinases combine with multiple cyclins to achieve the same specificity. Studies *in vitro* and *in vivo* investigating replication proteins show that CDK phosphorylates ORC, MCMs, RPA, and DNA polymerase alpha, as well as the specific loading factors required to assemble Cdc45 and GINS. CDKs play both positive and negative roles in the regulation of DNA replication: although initiation clearly requires CDK activity, the kinase is also required to prevent re-initiation of origins within the same cell cycle (Figure 2).

Data from several systems suggest that the positive role of CDK in replication initiation operates in a pathway parallel to Cdc7 activity. In the two yeasts, the conserved Dpb11/Rad4/TopBP1 protein interacts with Drc1/Sld3, which is a known CDK substrate. If Drc1 is not phosphorylated, it cannot bind to Dpb11. Data from *Xenopus* suggest that the Xmus101 protein (related to Dpb11) is required for Cdc45 binding. Together, these studies suggest that two components are required for Cdc45 and GINS binding: Cdc7 activation of the preRC and CDK activation of the Dpb11 pathway. This model is speculative. However, it provides an attractive fail-safe mechanism for the cell to ensure that multiple signals are integrated before it fires a single replication origin, which would significantly reduce the chance of aberrant initiation.

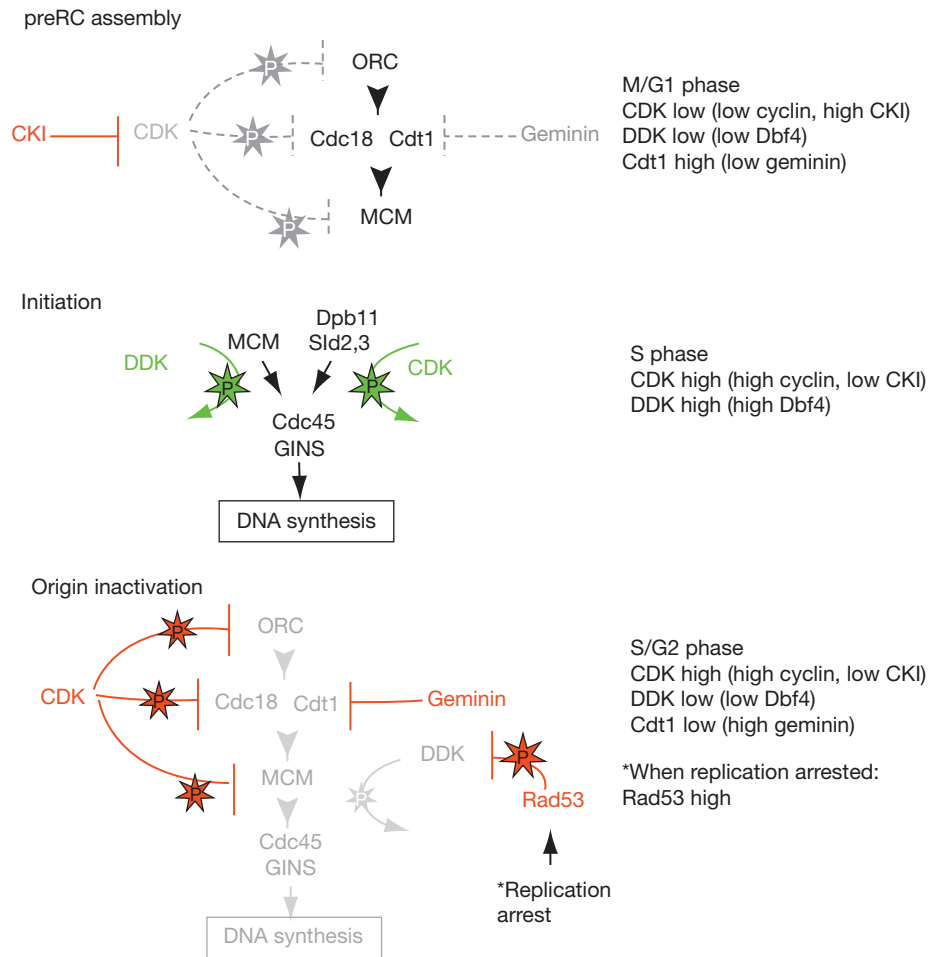


Figure 2 Positive and negative regulation of S-phase initiation prevents re-replication. The preRC is assembled on origins when CDK activity is low, and re-assembly is prevented when CDK activity is high. Origins only fire when both CDK and DDK are active on preRCs. Fluctuation in the levels or activity of cyclins, cyclin-kinase inhibitors (CKI), Dbf4 protein, and geminin contribute to this regulation. Under conditions when defects in replication block synthesis, Rad53 checkpoint kinase is activated to prevent further origin firing and to protect the structure of active replication forks. Gray text/arrows indicates inactivated pathway components.

Coupling Replication to the Cell-Cycle Engine

Activation and inhibition of replication from individual origins is controlled at multiple levels that ultimately depend upon the oscillation of CDK activity during the cell cycle. This oscillation reflects the control of the individual cyclins. The DDK subunit Dbf4 is also subject to this regulation (Figure 2).

CDK Inhibitors and Proteolysis

CDK activity is regulated at multiple levels, including phosphorylation. Most obviously, association of the kinase with transiently produced cyclin proteins provides temporal specificity (Figure 2). The cyclins are targeted for destruction by specific ubiquitination complexes at various points in the cell cycle. The Dbf4 subunit of the DDK in yeasts (and probably in larger cells) also oscillates in the cell cycle. It is regulated transcriptionally, and is also a target for ubiquitination during M phase, which restricts its activity to the S phase of the cell cycle. Thus, destruction of proteins in M phase helps maintain the timing of events in S phase.

Specific inhibitor molecules expressed during G1 also modulate CDK activity prior to S phase. These are usually specific to particular cyclin forms of the CDK. These inhibitors keep kinase activity repressed, which allows assembly of a preRC. As cells proceed toward S phase, the increasing levels of cyclin lead to increasing CDK activity which ultimately overcomes the inhibitor. The accumulating CDK phosphorylates its inhibitor, leading to its ubiquitination by an S-phase-specific complex that targets the inhibitor for destruction. Although this same general pattern describes most cell types, surprisingly the individual CDK inhibitor molecules in different species such as *S. cerevisiae* Sic1, *S. pombe* Rum1, and human p27^{Kip1} are highly diverged and show no primary sequence homology.

In mammals, cyclin D/CDK4 phosphorylates and inactivates the Rb protein. Active Rb inhibits E2F, a component of the S-phase-inducing transcription factor; thus, Rb needs to be inactivated before S phase can initiate. Because Rb inhibits entry into S phase and the cell cycle, its loss can lead to dysregulation of cell-cycle progression, so it is not surprising that Rb was originally identified as a tumor suppressor.

Activation of the preRC: Once and Only Once

The negative role of CDKs in preventing re-initiation of DNA replication is dramatically illustrated in fission yeast, in which overproduction of the CDK inhibitor Rum1 results in cells repeating S phase many times, without an intervening mitosis. The CDK is inhibited so it cannot promote mitosis, or prevent re-initiation of replication origins; however, the S-phase-promoting form of the CDK remains active.

The targets of the CDK for this inhibitory activity include several factors at the replication origin. Data suggest that the ORC complex is a CDK target that prevents re-replication. Similarly, Cdc6 protein is tightly regulated so that it is only present in the nucleus at the time of preRC assembly (between mitosis and S phase). It is a CDK substrate, and depending on the system, is either degraded or exported from the nucleus, in response to phosphorylation. In metazoans, Cdt1 associates with an inhibitor molecule called geminin, which itself is degraded during mitosis to free active Cdt1. Thus, active forms of Cdc6 and Cdt1 are only present at the time of preRC assembly to allow MCM loading. This provides one way to ensure that origins cannot be re-activated in a single cell cycle.

The MCM complex, although present throughout the cell cycle, can only bind at the origin in the presence of Cdc6 and Cdt1. The MCM complex is dislodged during replication, and some subunits are phosphorylated by CDK. In the yeast *S. cerevisiae*, this phosphorylation results in the active export of MCMs from the nucleus as S phase proceeds. The dislodged proteins remain in the nucleus in other species, but are unable to bind the chromatin. The MCMs cannot regain access to the origin (or in *Saccharomyces*, the nucleus) until CDK levels fall at the end of mitosis, and Cdc6 and Cdt1 are active again in the next G1 phase.

This sequential requirement for origin binding allows tight cell-cycle regulation of initiation. Low CDK activity allows pre-RC assembly. High CDK activity allows origin firing, but prevents re-assembly of initiation factors. The CDKs, thus, play a dual role by both promoting initiation and preventing re-replication at individual origins.

Origins are also inactivated during times of replication stress, which reduces the risk of catastrophic replication fork collapse and double-strand breaks. This relies on a specific replication checkpoint system. In budding yeast, the Rad53 kinase regulates the Cdc7 kinase to prevent origin activation under these conditions.

Regulating Chromosome Structure during S Phase

The events of S phase are intimately linked to chromosome structure. Cohesion between sister chromatids is established

during S phase, as the sister chromatids are generated. Genetic studies link cohesion to other proteins known to act at the replisome including Ctf4. Cohesion is essential for proper segregation of the chromosomes during mitosis. It prevents chromosome loss, contributes to centromere structure, and helps orient the chromatids to the opposite spindles. Whether the passing replication fork activates the cohesin proteins, or they are activated by the S-phase kinases, remains to be determined.

Chromatin assembly is also affected during S phase. The cell must assemble its newly duplicated DNA into nucleosomes and higher-order chromatin structures. Moreover, any epigenetic modifications on the histones or the DNA must be transferred to the new chromatin. Several conserved complexes are implicated in chromatin assembly and disassembly. The facilitates chromosome transactions (FACT) complex helps to dislodge the DNA from the chromatin and contributes to subsequent reassembly. Studies place FACT in front of the replisome. The chromatin assembly factor (CAF) complex helps recruit new histones into the new DNA. Chromatin structure may, in turn, regulate S-phase timing. Recent studies suggest that modification of histones by acetylation may help determine the timing of firing for individual origins.

Conclusions

Events during S phase have profound effects on chromosome biology throughout the cell cycle. The replication fork has an enormous capacity to remodel chromatin. In addition, activation of specific kinases only during S phase can provide a temporal link between DNA replication and modification of other proteins. Clearly, investigators are only beginning to trace these links. The study of S-phase regulation of multiple chromosome events will expand for years to come.

See also: **Cell Architecture and Function:** Cell Cycle: DNA Damage Checkpoints; Cell-Cycle Controls in G₁ and G₀; Chromosome Organization and Structure, Overview; **Molecular Biology:** DNA Helicases: Hexameric Enzyme Action; DNA Replication: Eukaryotic Origins and the Origin Recognition Complex; DNA Replication Fork, Eukaryotic.

Further Reading

- Labib K (2010) How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? *Genes and Development* 24: 1208–1219.
- Mechali M (2010) Eukaryotic DNA replication origins: Many choices for appropriate answers. *Nature Reviews. Molecular Cell Biology* 11: 728–738.

Cell-Cycle Controls in G₁ and G₀

A Narasimha, W Shi, and S F Dowdy, University of California, La Jolla, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Wenge Shi and Steven F. Dowdy, volume 1, pp. 328–331, © 2004, Elsevier Inc.

Glossary

Cdk inhibitors (CKI) There are two families of CKI: the Ink4 family proteins (inhibitors of Cdk4: p15INK4b, p16INK4a, p18INK4c, and p19INK4d) which specifically bind monomeric Cdk4 and Cdk6 to inhibit cyclin D-dependent kinase activity and the Cip/Kip family (p21WAF1, p27Kip1, and p57Kip2) that specifically binds and inhibits cyclin:Cdk complexes (cyclin D:Cdk4/6; cyclin E:Cdk2; and cyclin A:Cdk2).

Cyclin-dependent kinases (Cdk) Key activators of the cell cycle. A group of serine/threonine-specific protein kinase complexes composed of a cyclin regulatory subunit and a Cdk catalytic subunit. Activation of the kinase requires cyclin binding to Cdk.

Restriction point A key cell-cycle transition between the early G₁ mitogen-dependent phase into the late G₁ mitogen-independent phase prior to initiation of DNA synthesis in S phase.

The cell cycle is the process by which one cell becomes two. Somatic cell division involves cell growth (increase in cellular components, such as ribosomes, membranes, and organelles) throughout the cell cycle, faithful replication of its DNA during the S phase of the cell cycle, and precise distribution of DNA between daughter cells during the mitosis (M) phase. In addition, with the exception of early embryonic cell divisions, two gap phases, Gap 1 (G₁) and Gap 2 (G₂), are separated by S phase. As multicellular eukaryote organisms regulate cell division tightly to maintain tissue homeostasis, the most important regulatory decision is made during exiting of the resting state so-called the G₀ phase and G₁ phase of the cell cycle before cells become committed to initiate DNA synthesis and complete the cell cycle.

G₁ and G₀ Phase of the Cell Cycle

The majority of cells in adult metazoans are permanently withdrawn from the cell cycle in a terminally differentiated state. Only small numbers of cells, such as hematopoietic and epithelial early progenitor cells, are actively proliferating. Other cells are reversibly withdrawn from the cell cycle and remain in a quiescent stage or the G₀ phase of the cell cycle. Upon proper stimulation, these cells can re-enter the cell cycle. For example, highly differentiated hepatocytes in adult liver are present in the G₀ resting phase and rarely replicate normally. However, in response to acute liver injury or distress, these G₀ hepatocytes can be stimulated to re-enter the cell cycle and regenerate the liver. Similarly, primary peripheral blood lymphocytes (T and B cells), which are present in the G₀ phase, can re-enter the cell cycle and start clonal expansion when presented with the appropriate antigen. These examples serve to demonstrate that certain cell types may enter and exit the cell cycle pending the appropriate stimulus, whereas the vast majority of cells in a matured metazoan have permanently exited the cell cycle.

The most important decision for cell-cycle progression is made during the G₁ phase. In G₀ phase, cells respond to extracellular signals by entering the G₁ phase of the cell cycle.

However, prior to transiting into the late G₁ phase, they must traverse the growth-factor-dependent restriction point. This is a critical regulatory phase of the cell cycle where a cell becomes committed to enter S phase and finish the remaining cell cycle. In contrast to normal cells, tumor cells are often less growth factor dependent and fail to respond to growth inhibitory signals. Consequently, tumor cells select for genetic mutations that disrupt the important decisions performed at the restriction point and this is, in fact, one of the hallmarks of cancer.

Regulators of G₀ Exit and G₁–S Progression

One of the key negative regulators of the restriction point and early G₁ to G₀ cell-cycle exit is the retinoblastoma tumor suppressor protein (Rb) and two closely related family members, p107 and p130. Murine embryonic fibroblasts (MEF) deficient for all three pocket proteins fail to respond to G₁ phase arrest signals following contact inhibition, serum starvation, or DNA damage. The antiproliferation function of the pocket proteins depends, at least in part, on their interaction with E2F/DP transcriptional factors, which regulate the expression of key genes required for DNA synthesis, DNA repair, DNA-damage checkpoint, apoptosis, and mitosis. Direct interactions of Rb family members with E2F/DP complexes and the recruitment of chromatin-modifying enzyme complexes, such as histone deacetylases (HDAC), polycomb group proteins, the SWI/SNF complex, and histone methyl transferases, to E2F-responsive genes results in inhibition of target gene expression.

Distinct Rb family member–E2F repressor complexes exist in different cell-cycle phases. As an example, during the G₀ phase, p130 and p107 interact with E2F4 and E2F5, whereas Rb remains unbound. E2F4 and E2F5 are expressed constitutively and are involved in E2F-dependent gene repression during cell-cycle exit and terminal differentiation. By contrast, during early G₁ phase, Rb interacts with E2F1, E2F2, and E2F3, the so-called activator E2Fs. When the cell passes through the restriction point, Rb becomes inactivated by hyperphosphorylation, releasing E2Fs 1–3 and promoting the activation of E2F-responsive

genes for cellular proliferation. Recent evidence has shown that the Rb–E2F interaction is a bistable switch, and is necessary and sufficient for restriction point control. Not surprisingly, MEFs deficient for these E2F1, E2F2, and E2F3 exhibit impaired E2F-responsive gene induction in response to serum and significant proliferative defects.

Rb is physiologically regulated by a group of serine/threonine-specific protein kinases called cyclin-dependent kinases (Cdks). Specifically, cyclin D:Cdk4/6 and cyclin E:Cdk2 kinases are responsible for phosphorylating Rb during the G₁ phase of the cell cycle. However, cyclin D and cyclin E are regulated differently and play distinct roles in the cell-cycle regulation of Rb (Figure 1). D-type cyclins are not expressed in G₀ cells. However, cyclin D levels increase after stimulation of growth factors as a result of increased transcription mediated by Myc and Ras/MEK/MAPK signaling pathway and increased stability mediated by the PI(3)K/Akt signaling. In cycling cells, cyclin D levels and associated kinase activity remain relatively constant. Therefore, the D-type cyclins control cell-cycle entrance and exit by linking the mitogenic pathway to the core cell-cycle machinery.

Unlike cyclin D, the expression of cyclin E is mitogen independent and dependent on the action of E2Fs after the hyperphosphorylation and inactivation of Rb at the late G₁ phase of the cell cycle. However, because Rb represses cyclin E expression, questions remain as to how cyclin E gene expression and cyclin–Cdk2 activity is initially turned on. Once cyclin E levels reach a critical threshold, its levels can be further increased as a result of a positive feedback loop. Another potential mechanism of activating Cdk2–cyclin complexes is through the function of a Cdk-activating kinase (CAK). Studies are now being made on identifying this kinase, as it could be a potential target for therapeutics. Recently, a hypothesis has shown that the CAK could be Cdk7, which is known to be

part of the TFIIF transcription complex. However, more studies have to be made in order to substantiate this finding.

Rb undergoes cell-cycle-specific phosphorylation through the distinct phases of the cell cycle and subsequent dephosphorylation during mitosis. In G₀, Rb is unphosphorylated. Unphosphorylated Rb is considered inactive, as it is not able to bind E2F transcription factors. When a cell is stimulated to enter the cell cycle, cyclin D:Cdk4/6 complexes put one phosphate on Rb. This monophosphorylation can occur on at least nine different serines or threonines located throughout Rb. Monophosphorylated Rb is the active, transcriptional-repressing form of Rb and can simultaneously bind to both E2F and HDAC. Once the cell nears the restriction point, Rb is rapidly hyperphosphorylated by cyclin E:Cdk2 to give phosphorylation species of greater than 12 phosphates. This allows for the release of E2F and subsequent transcription of E2F-dependent targets and progression through the cell cycle.

The consequence of cyclin D-dependent and cyclin E-dependent kinase activity on Rb also results in differential regulation. It is apparent that cyclin D-dependent kinases are not sufficient for disrupting the Rb–E2F–HDAC complex. Indeed, this is most dramatically supported by the detection of functional Rb–HDAC–E2F complex in p16^{INK4a}-deficient cancer cells with deregulated and active cyclin D-dependent kinases. Furthermore, in T cells entering the cell cycle, E2F4 associates with Rb only after cyclin D kinase activity is turned on. By contrast, cyclin E-dependent kinase activity and hyperphosphorylation of Rb are required for disrupting E2F–Rb and HDAC–Rb interactions. Moreover, in addition to Rb family members, cyclin E-dependent kinase phosphorylates target proteins involved in chromatin remodeling, DNA synthesis, and centromere functions, further triggering the cell to traverse across the restriction point into late G₁ and S phase.

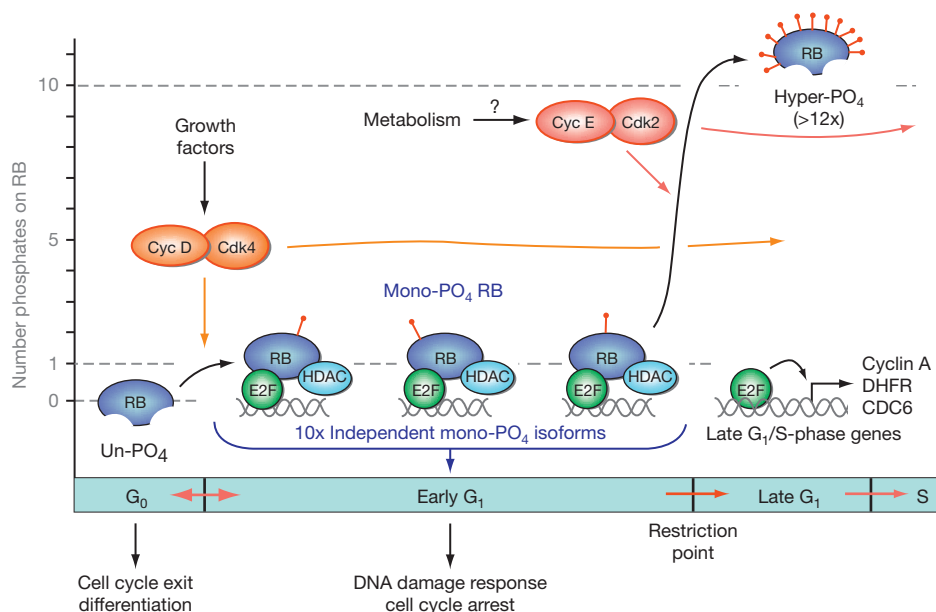


Figure 1 Model of G₀ and G₁ cell-cycle progression. Activation of cyclin D:Cdk4/6 kinases by growth factor stimulation of G₀ cells results in increased metabolism or growth and monophosphorylation of Rb driving assembly of pRb:E2F complexes. By an unknown mechanism, increased metabolism and cell mass leads to activation of cyclin E:Cdk2 complexes at the restriction point resulting in hyperphosphorylation (inactivation) of Rb, release of E2Fs, and activation of late-G₁-specific genes involved in DNA synthesis and late-G₁-cell-cycle regulation.

Stem Cells and Cell Proliferation

Recent studies have provided an insight into the differences in cell-cycle regulation between somatic and stem cells. Embryonic stem cells have an abbreviated cell cycle and, more importantly, an abbreviated G₁. Mouse embryonic stem cells have constitutive cyclin E:Cdk2 kinase activity, which causes rapid hyperphosphorylation of Rb and subsequent bypass of the restriction point. Thus, embryonic stem cells are not dependent on cyclin D levels and growth factor stimulation. In fact, growing embryonic cells in the presence of MAPK inhibitors will promote self-renewal and prevent differentiation into somatic cells. Less is known about human embryonic stem cells. It has been reported that Cdk4 is upregulated upon reentry into the cell cycle suggesting a dependence on D-type cyclin activity. However, like mouse embryonic stem cells, human embryonic stem cells have a shortened G₁ phase and cell-cycle-independent expression of cyclin E, which leads to hyperphosphorylation of Rb.

Adult stem cells do not proliferate as fast as embryonic stem cells. In fact, about 75% of adult stem cells are quiescent and will enter early G₁ only upon stimulation. Adult stem cells need to proliferate for either self-renewal or differentiation. Knockout of certain markers, such as the Cdk inhibitor p18, cause stem cells to self-renew. Knockout of other markers, such as the gene *HoxB4*, promotes differentiation.

Cell-Growth Regulation

The other aspect of cell-cycle regulation is the increase in cell mass or cell growth. In general, cells must grow to a minimum size to be able to divide and generate two daughter cells equal in size to the starting mother cell. Premature entry into S phase results in smaller daughter cells. Therefore, it has been realized, but not well understood, that generally with the exception of some developmental programs, such as early embryonic cell division and endo-reduplication, cell division and cell growth must be coordinated. Although cell growth is often reflected by an increase in cell size, cell growth needs to be defined as an increase in cell mass, or more precisely, the underlying global RNA and protein synthesis capacity, and additional macromolecule synthesis and storage.

How mitotic cells coordinate cell growth and division remains an open question. Some of the key regulators involved in cell-cycle regulation also control cell growth. For example, the Myc transcription factor has long been implicated in cell-growth regulation in both *Drosophila* and mammalian cells. Besides the gene products involved in cell-cycle regulation, such as cyclin D2, Cdk4, cdc25, cyclin E, and cyclin A, many Myc-target gene products are involved in ribosome biosynthesis, protein synthesis, and specific metabolic pathways. Another example of a key cell-cycle regulator is Rb. In addition to regulating E2F-dependent gene transcription, Rb can also repress RNA polymerase I- and III-dependent transcription that are responsible for the synthesis of ribosomal RNA (rRNA), various small RNAs such as 5S rRNA, and transfer RNA (tRNA). This regulation is mediated by direct interaction of pRb with the upstream-binding factor (UBF) and with the TF-IIIIB coactivator, respectively.

On the other hand, cell growth may influence cell division. Understanding of cell-growth regulation comes from studying

immune-suppressant rapamycin. Treatment of both yeast cells and mammalian cells with rapamycin delays cell growth and proliferation. In budding yeast, signaling of the target of rapamycin (TOR), a protein kinase, coordinates nutrient availability, such as carbon source and nitrogen, with cell growth and proliferation. It turns out that the yeast G₁/S cyclin Cln3 is both translationally and posttranslationally regulated in response to nutrient levels. Cell division occurs when cells reach a certain size. In metazoans, the situation is more complex because nutrient levels are maintained by tissue homeostasis. Not only does mammalian TOR (mTOR) signaling coordinate protein synthesis with glucose and amino acid availability (linked to nutrient sensing signaling pathways), but it also mediates hormonal and mitogenic factor responses. mTOR lies downstream of PI(3) kinase/Akt, although it is currently unclear that they belong to a linear signaling cascade.

PI(3)K/Akt may also activate further downstream targets through tumor-suppressor tuberous sclerosis (TSC) complexes. It turns out that the mTOR pathway in conjunction with signaling through the PI(3)K pathway regulates the translational initiation and elongation, ribosome biosynthesis, amino acid import, and metabolism. One of the important targets in ribosome synthesis regulated by the PI(3)K/Akt/mTOR pathway is S6 kinase, which phosphorylates ribosomal protein S6. In *Drosophila*, S6 is associated with body size and cell growth. S6 phosphorylation is required for the translation of a group of mRNAs with a 5' terminal oligopyrimidine tract (5'TOP), including ribosomal protein messenger RNAs (mRNAs) and mRNAs encoding for translation machinery. In addition, the Akt/mTOR pathway regulates protein translation by phosphorylating and inactivating the translation initiation factor 4E-binding protein (4E-BP1). As a result, PI(3)K/Akt and mTOR are linked to translation initiation by regulating the eukaryotic translation initiation factor 4E (eIF-4E), which enhances the translation of essential genes for cell-cycle regulation, including cyclin D and p27^{Kip1}. Future studies will ultimately resolve how cell growth signaling impinges on and is incorporated into the cell-cycle regulatory machinery.

Future Perspectives

There has been an increased understanding of signal transduction and cell division. A lot of focus has been on protein posttranslational modification and gene expression regulation. However, until recently, the importance of cell growth has not been fully appreciated. Clearly there is cross-talk between cell growth and cell-cycle regulation. Future studies should fill in the gaps in our understanding of cell growth, cell cycle progression, and their subsequent contributions to neoplastic transformation.

See also: **Cell Architecture and Function:** Cell Cycle: Control of Entry and Progression through S Phase; Cell Cycle: Mitotic Checkpoint; **Molecular Biology:** RNA Polymerase I and RNA Polymerase III in Eukaryotes; RNA Polymerase Reaction in Bacteria; RNA Polymerase Structure, Bacterial.

Further Reading

- Abraham RT (2002) Identification of TOR signaling complexes: More TORC for the cell growth engine. *Cell* 111: 9–12.
- Chen HZ, Chen SY, and Leone G (2009) Emerging roles of E2Fs in cancer: An exit from cell cycle control. *Nature Reviews. Cancer* 9: 785–797.
- Ho A and Dowdy SF (2002) Regulation of G1 cell-cycle progression by oncogenes and tumor suppressor genes. *Current Opinion in Genetics and Development* 12(1): 47–52.
- Malumbres M and Barbacid M (2001) To cycle or not to cycle: A critical decision in cancer. *Nature Reviews. Cancer* 1: 222–231.
- Orford KW and Scadden DT (2008) Deconstructing stem cell self-renewal: Genetic insights into cell-cycle regulation. *Nature Reviews. Genetics* 9: 115–128.
- Saucedo L and Edgar B (2002) Why size matters: Altering cell size. *Current Opinion in Genetics and Development* 12(5): 556–565.
- Stevaux O and Dyson NJ (2002) A revised picture of the E2F transcriptional network and RB function. *Current Opinion in Cell Biology* 14: 684–691.

Cell Cycle: DNA Damage Checkpoints

M D Weitzman, Children's Hospital of Philadelphia, Philadelphia, PA, USA

J Y J Wang, University of California, San Diego, La Jolla, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Jean Y. J. Wang, volume 1, pp. 338–344, © 2004, Elsevier Inc.

Glossary

Apoptosis A form of programmed cell death that kills a cell through condensation, shrinkage, fragmentation, and engulfment of the dead cell fragments by neighboring healthy cells.

ATM A large-molecular-weight protein encoded by the gene that is mutated in the human disease ataxia–telangiectasia (A–T). A–T patients suffer from cerebella degeneration, extreme sensitivity to ionizing radiation, increased cancer risk, and sterility. ATM-knockout mice exhibit most of the A–T phenotypes, except cerebella degeneration. ATM contains a PI3-kinase homology domain and its function is required for ionizing radiation (IR) to activate the cell-cycle checkpoints, apoptosis, and senescence.

ATM and Rad3-related (ATR) A protein with a PI3-kinase homology domain. ATR function is essential to mammalian cells as the knockout of ATR in mice causes embryonic lethality, and the knockout of ATR in cells causes passive death resulting from S-phase defects. ATR is required for the activation of cell-cycle checkpoints induced by UV irradiation.

BRCT A modular protein–protein interaction domain conserved through evolution and found in proteins involved in checkpoint activation and DNA repair.

Checkpoint kinase-1 (Chk1) First identified in the fission yeast and conserved through evolution. DNA damage activates Chk1 to maintain Cdk1/cyclin B (also known as mitosis promoting factor (MPF)) in the latent, inactive state. Chk1 is not an essential gene in yeasts, but Chk1-knockout mouse ES cells die from mitotic catastrophe.

Checkpoint kinase-2 (Chk2) First identified in the fission yeast as the Cds1 kinase and conserved through evolution. The budding yeast contains two Chk2-like kinases: RAD53 and DUN-1. Chk2 also inhibits MPF activation, and it phosphorylates and activates p53.

Cyclin-dependent protein kinase (Cdk) It forms complex with and is activated by cyclin. Cdk1 is also known as Cdc2 kinase.

Cyclins Activators of cyclin-dependent kinases. Cyclins are expressed and degraded periodically through the cell cycle.

DNA damage response (DDR) The cellular machinery that senses DNA damage and activates signaling cascades to promote effector functions comprising checkpoint activation and DNA repair.

FHA A modular protein–protein interaction domain conserved through evolution. FHA domain binds peptides with phosphorylated threonine. The FHA domain is found in a variety of proteins with diverse functions, including those involved in DNA damage signal transduction.

Overview

Error-free DNA replication and chromosome segregation are critical to the maintenance of genome integrity. A variety of DNA lesions are generated during normal metabolic processes or by exposure to external genotoxic agents. To maintain genome integrity, an evolutionarily conserved cellular surveillance network continuously monitors DNA lesions to coordinate DNA repair with cell-cycle progression and cell fate decisions. This surveillance network includes biochemical pathways that act as checkpoints to delay or arrest cell-cycle progression in response to DNA damage. This article discusses three distinct cell-cycle checkpoints that are activated by DNA damage: the G1/S checkpoint, the intra-S checkpoint, and the G2/M checkpoint (Figure 1).

Cell-Cycle Checkpoints Activated by DNA Damage

The G1/S checkpoint inhibits S-phase entry in G1 cells that have not yet committed to DNA replication. The intra-S checkpoint prevents initiation of DNA replication at origins that

have not yet been activated in S-phase cells. The G2/M checkpoint inhibits entry into mitosis. The cell-cycle checkpoints are reversible, allowing resumption of cell-cycle progression after the signals from DNA lesions are extinguished.

The key targets of inhibition by the cell-cycle checkpoint pathways are cyclin-dependent protein kinases (Cdks). Cdks are activated by binding to their cyclin partners, which are strictly controlled by temporally regulated expression, subcellular localization, and ubiquitination-dependent degradation. The kinase activity of cyclin-bound Cdk is further regulated by threonine and tyrosine phosphorylations in the Cdk kinase domain. The DNA damage checkpoint pathways inhibit the expression of cyclins, activate the degradation of cyclins, and promote the inhibitory phosphorylations of Cdks to halt cells in G1 or G2 phases.

The DNA damage checkpoint pathways consist of lesion sensors, adaptors/mediators that assemble signaling complexes at DNA lesions, protein kinases, and their substrate effectors (Table 1). The proteins identified as components of the DNA damage checkpoint pathways are part of the surveillance network that also regulates the orderly progression through an unperturbed cell cycle. Although the G1/S, intra-S,

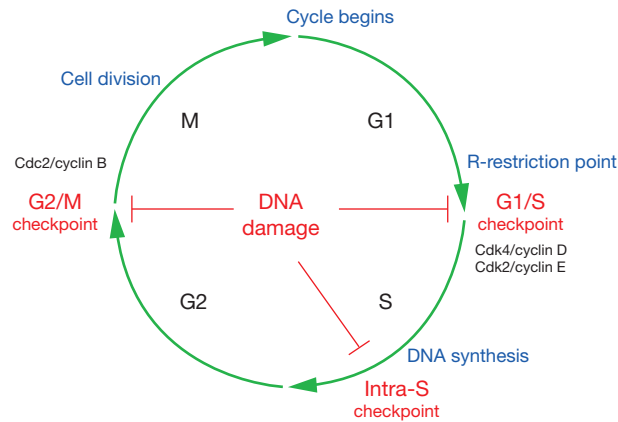


Figure 1 The cell cycle and DNA damage checkpoints. The cell cycle consists of four distinct phases: G1 phase, S-phase (synthesis), G2-phase, and M-phase (mitosis). Activation of each phase depends on the proper progression and completion of the previous one. The G1-phase is marked by synthesis of enzymes required for DNA replication in S-phase. During S-phase the genome is replicated and the amount of DNA in the cell effectively doubles. During G2 the proteins required for mitosis are produced. Mitosis is the process in which the cell separates the chromosomes into two identical sets in two daughter nuclei. Sequential progression through the cell cycle is regulated by cyclins and cyclin-dependent kinases (Cdks).

Table 1 DNA damage checkpoint proteins

Protein function	Mammals	<i>S. pombe</i>	<i>S. cerevisiae</i>
Sensors			
MRN/X	Mre11/Rad50/ Nbs1	Mre11/Rad50/ Nbs1	Mre11/Rad50/ Xrs2
PCNA-like	Rad9/Rad1/ Hus1	Rad9/Rad1/ Hus1	Ddc1/Rad17/ Mec3
RFC-like	Rad17	Rad17	Rad24
PI3-kinases	ATM ATR (ATRIP) DNA-PK	Tel1 Rad3 (Rad26)	Tel1 Mec1 (Ddc2)
Mediators	Mdc1 53BP1 Claspin BRCA1 TopBP1	Crb2 Mrc1	Rad9 Mrc1
Effector kinase	Chk1 Chk2	Chk1 Cds1	Chk1 Rad53

and G2/M checkpoints are distinct, many of the damage sensors and signaling molecules are shared by all three pathways or play a primary role in one pathway and a back-up role in the others. The checkpoint pathways contain self-reinforcing amplification loops with complex layers of spatial and temporal organization that are still under investigation.

G1/S Checkpoint

Cells in the G1-phase of the cell cycle become committed to enter the S-phase at a stage referred to as the Restriction

point (R) in mammalian cells and Start in budding yeast. The G1/S checkpoint serves to prevent cells from entering S-phase in the presence of DNA damage and functions to inhibit the initiation of replication (Figure 2). In the case of double-stranded breaks (DSBs), caused by ionizing radiation or radiomimetic agents, the ataxia telangiectasia mutated (ATM) kinase is activated and phosphorylates many downstream effectors, notably p53 and Chk2. Activation of these signal transduction pathways serves to initiate (Chk2) and maintain (p53) the G1/S arrest. The initiation of G1/S checkpoint requires Chk1- or Chk2-dependent phosphorylation of the Cdc25A phosphatase, leading to its nuclear exclusion and ubiquitin-mediated degradation. The absence of Cdc25A prevents the dephosphorylation of Cdk2 at the inhibitory phosphorylation sites and thus enforces the inactivation of Cdk2/cyclin complexes. Without active Cdk2, the Cdc45 protein, which is one of the key components of the pre-replication complex, is not phosphorylated and is unable to be loaded onto chromatin to initiate replication. This results in a block to the initiation of DNA synthesis. The phosphorylation-regulated pathway is implemented rapidly in response to DNA damage but is relatively transient and delays G1/S transition for just a few hours. Maintenance of the G1/S arrest requires DNA damage-induced activation of p53, which stimulates the expression of p21/Cip1/WAF1, an inhibitor of Cdk2/cyclin complexes. Since upregulation of p21Cip1 involves transcription and new protein synthesis, this p53-mediated G1 arrest is a slower response that complements the transient acute inhibition of Cdk2 through Cdc25A degradation.

Intra-S Checkpoint

In cells that have passed through the R point and have irreversibly committed to DNA replication, exposure to ionizing radiation activates an Intra-S checkpoint that causes a transient inhibition of DNA synthesis (Figure 2). In eukaryotic cells, the genome is divided into tens of thousands of replicons, each with an origin of replication. The replication origins are not activated synchronously at G1/S transition. Instead, some origins are activated early in S-phase, while others are activated later, with the timing of origin activation being dependent on the chromatin conformation. While DNA damage does not cause S-phase arrest, activation of the checkpoint pathways during S-phase, for example, by DSBs, does cause the inhibition of firing of late origins. The DSB-induced intra-S checkpoint critically depends on the ATM kinase, as illustrated by the radiation-resistant DNA synthesis (RDS) phenotype that characterizes cells lacking functional ATM. Although ATM is required for the activation of the intra-S checkpoint, some other components of the G1/S checkpoint pathway, that is, Chk1, Chk2, CDC25A, p53, and p21Cip1, are not involved in DSB-induced inhibition of late replication origins. Instead, the intra-S checkpoint activation appears to require the ATM-dependent phosphorylation of SMC1 and SMC3, which are components of the cohesin complex that regulates the pairing of sister chromatids. The precise biochemical mechanism underlying the inhibition of late origin firing by ATM and cohesin remains to be elucidated.

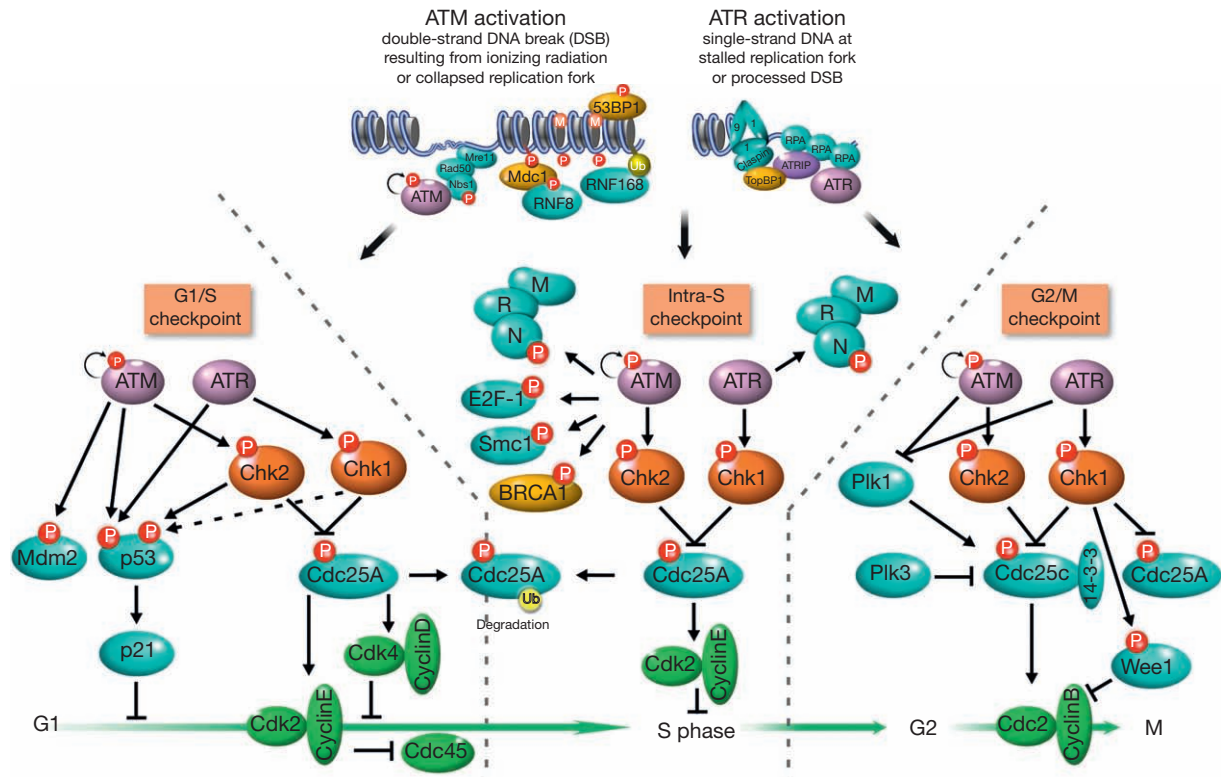


Figure 2 Activation and signaling of DNA damage checkpoints. DNA damage triggers a rapid cascade of phosphorylation events initiated by the ATM and ATR protein kinases. ATM is activated by DSBs through the MRN complex that acts the damage sensor. A series of posttranslational modifications, including phosphorylation (P), ubiquitination (Ub), and methylation (M), results in retention of damage proteins at the lesion and activation of checkpoint signaling. ATR is recruited to ssDNA through the ATRIP cofactor that binds RPA and is activated by the 9-1-1 complex, claspins, and TopBP1. The ATM and ATR signaling pathways culminate in the G1/S, intra-S, and G2/M checkpoints. See text for further details.

G2/M Checkpoint

The G2/M checkpoint prevents cells from undergoing mitosis in the presence of DNA damage (Figure 2). Depending on the type of DNA lesion, the ATM-Chk2 pathway and/or the ATR-Chk1 pathway are activated to arrest the cell cycle in G2. In response to UV radiation, the ATR-Chk1 pathway initiates G2 arrest and the ATM-Chk2 pathway maintains the G2 arrest. By contrast, G2 arrest in response to lesions such as ionizing radiation-induced DSBs is initiated by the ATM-Chk2 pathway. Mitotic entry requires Cdc2-mediated dephosphorylation of Cdc2 in complex with cyclin B. The G2/M checkpoint pathways inhibit Cdc2/cyclin B activity by stimulating the inhibitory phosphorylation of Cdc2 through the Wee1 kinase and blocking the dephosphorylation of Cdc2 by the Cdc25 phosphatase. Although it was initially suggested that the Cdc25C phosphatase was the key target of the G2/M checkpoint pathway, it now appears that all three members of Cdc25 family (Cdc25A, Cdc25B, and Cdc25C) cooperate to regulate Cdc2/cyclin B activity. Chk1- or Chk2-mediated phosphorylation of Cdc25 family of phosphatases can induce Cdc25 binding by the 14-3-3 proteins, leading to cytoplasmic sequestration and degradation of Cdc25 by the ubiquitin–proteasome pathway. Other signaling modules that contribute to G2/M arrest include the p38 family of stress-activated protein kinases, the mediators such as 53BP1 and BRCA1, and upstream regulators such as Polo-like kinases (Plk1 and Plk3).

DNA Damage Signal Transduction

DNA Lesion Sensors

The damage surveillance network begins with protein sensors that recognize specific lesions in the genomic DNA (Figure 3). Some of the sensors bind to modified bases, while others respond to nicks, regions of single-stranded DNA (ssDNA), or DSBs. The majority of the lesion sensors are part of the DNA repair pathways. For example, the Mre11/Rad50/Nbs1 (MRN) complex functions as a sensor of DSBs and plays essential roles in both the activation of ATM-dependent checkpoint pathways and DSB repair pathways. Likewise, the MSH2/MSH6 complex functions as a sensor of mismatched, alkylated, or oxidized base pairs and plays essential roles both in the activation of ATM- and ATR-dependent checkpoint pathways and in mismatch repair. The dual roles of lesion sensors in activating checkpoints and repair allow for the coordination of cell-cycle progress with the removal of DNA lesions.

PIKK-Family of Kinases

The two main signal transducers in the DNA damage response are ATM and ATR. They are members of the phosphoinositide 3-kinase-like kinase (PIKK) family and share common domain structure (Figure 4). Another member of the PIKK family involved in regulating cellular

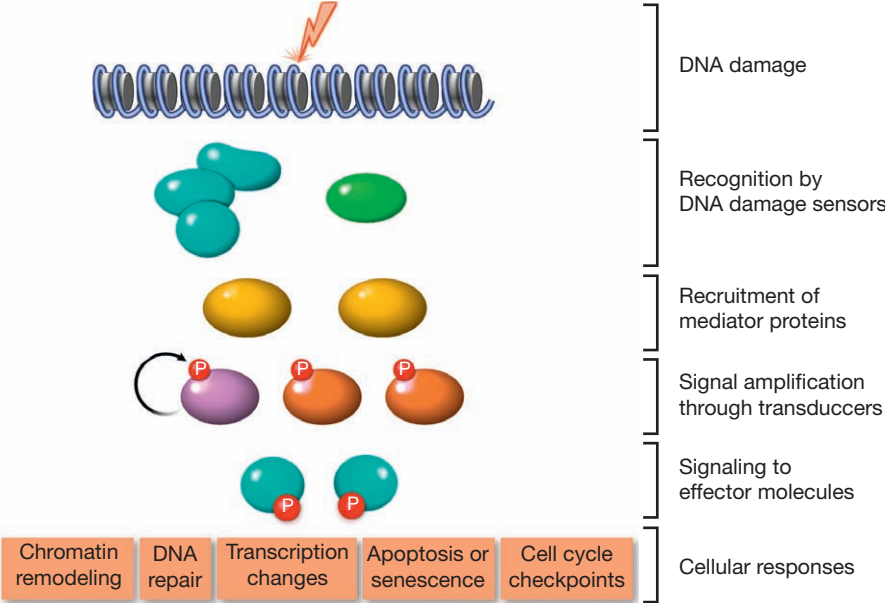


Figure 3 The DNA damage response. Damage lesions in DNA are recognized by sensor proteins that initiate recruit mediators and transducers of the damage response. Signal amplification through effector molecules produces the signaling networks that induce the cellular responses and facilitate repair.

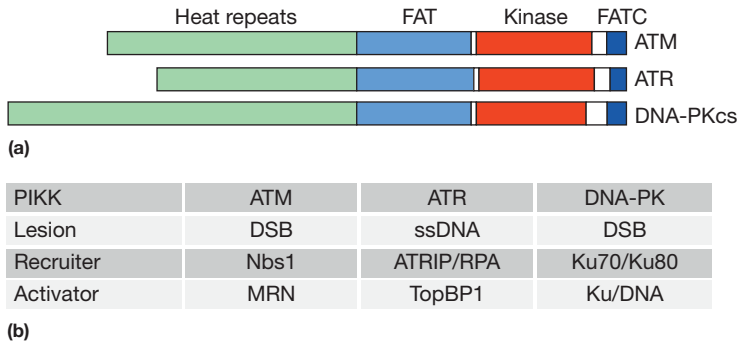


Figure 4 The PIKK family of protein kinases. (a) Schematic of the PIKK domain structure. FAT = FRAP, ATM, TRRAP domain; FATC = FAT C-terminus domain; HEAT = Huntingtin, elongation factor 3, A subunit of protein phosphatase 2A, TOR1. (b) Each PIKK responds to a specific DNA damage lesion. Binding to partner proteins recruits the PIKK to the lesion and interactions with activators induce their full activation and signaling to substrates.

responses to DNA damage is the DNA-dependent protein kinase DNA-PK, although its role in the cell-cycle checkpoint pathways is secondary to its major function in nonhomologous end-joining repair of DSBs. The kinase domain in the C-terminus of these proteins is flanked by the FAT domain and HEAT repeats on the one hand and the FATC domain (required for PIKK kinase activity) on the other (Figure 4). In response to DNA damage, ATM and ATR phosphorylate a large number of cellular proteins at serine and threonine sites in SQ or TQ sequence motifs. ATM is mainly activated by DSBs, while ATR is activated by the exposure of ssDNA. ATM is recruited to DSBs through binding to the MRN complex at DNA ends. ATR is recruited to ssDNA by binding to its partner ATRIP, which associates with the ssDNA-binding protein replication protein A (RPA). A shared sequence motif located at the C-terminus of Nbs1 and the C-terminus of ATRIP has been proposed to be required for recruitment of ATM and ATR, respectively, to damaged DNA.

ATM Activation

ATM is localized predominantly in the nucleus of most cell types, and its levels do not change in response to DNA damage. However, its protein kinase activity is stimulated *in vivo* by agents that induce DNA damage and *in vitro* by linear DNA. The ATM protein is thought to exist in an inactive dimer or higher-order oligomer, and auto-phosphorylation at Ser1981 converts it into monomers, which appear to be the form of the enzyme that activates checkpoint pathways. The MRN complex is important for recruitment of ATM to DSBs and also for its maximal activation. Some substrates of ATM accumulate at the DSBs, whereas others appear only transiently at the damage site, for example, the downstream checkpoint kinase Chk2. Compromising the MRN function (due to inherited hypomorphic mutations, shRNA-mediated knockdown, or viral-mediated inactivation) causes defective ATM auto-phosphorylation and signaling to some downstream substrates. There is evidence that ATM

can also be activated by changes in chromatin structure. ATM auto-phosphorylation can be observed in cells following exposure to hypotonic buffer or by inhibition of histone deacetylases. In addition to MRN, other proteins recruited through ATM interactions, such as Tip60, may also contribute to ATM activation and chromatin modification at DSBs. Stimuli that alter chromatin and induce ATM activation in the absence of DSBs may involve additional ATM-binding proteins, such as ATMIN, that could compete with Nbs1 to interact with ATM.

ATR Activation

ATR is activated during each S-phase in order to regulate replication origin firing and to repair damaged replication forks. While the RPA–ssDNA complex is required to recruit ATR-ATRIP to DNA lesions, it is not sufficient to activate ATR. Functioning at the top of the ATR-signaling network is a heterotrimeric DNA clamp that is conserved from yeast to man. It is structurally similar to the proliferating cell nuclear antigen (PCNA)-homotrimer, and based on the fission yeast gene names (Rad9, Rad1, and Hus1), this trimer is referred to as the 9-1-1 complex. While the PCNA clamp is loaded onto replicating DNA by replication factor C (RFC), the 9-1-1 complex is loaded onto DNA by a specialized clamp loading complex composed of the Rad17 protein (according to the fission yeast gene name) and RFC. Binding of RPA to ssDNA guides the loading of the 9-1-1 complex to the 5' primer end at the ssDNA/dsDNA junctions of resected DSBs or stalled replication forks. The ATR kinase activity is then stimulated by the topoisomerase-binding protein 1 (TopBP1), which is recruited to ATR by binding the C-terminus of Rad9. Therefore, TopBP1 is required for the phosphorylation of ATR substrates and for checkpoint signaling.

Downstream of the ATM and ATR kinases is a network of more than 700 different substrates with a wide range of biological activities and varying degrees of evolutionary conservation. Among the highly conserved substrates are the Chk1 and Chk2 protein kinases that play essential roles in regulating cellular responses to DNA damage. The second group of highly conserved substrates of the ATM and ATR kinases are the mediators, exemplified by Rad9 and Mrc1 proteins in yeast, and the Mdc1 and 53BP1 proteins in mammals. The third group of highly conserved substrates are histones, in particular, the H2AX subunit of the core histones. Phosphorylation by the ATM and ATR kinases generates in these substrates pSQ or pTQ motifs that function as phospho-epitopes in protein–protein interactions that are critical for signaling.

Mediators and Phospho-Dependent Protein-Binding Domains

The transduction of DNA damage signal from the lesion sensors to the protein kinases requires specific phospho-peptide-binding domains, in particular, the forkhead-associated (FHA) and the BRCA1 C-terminal (BRCT) domains. These domains are found in many proteins of the DNA damage network, and they recognize short amino acid sequences flanking phosphorylated serine or threonine residues. The FHA domain binds exclusively to phospho-threonine-containing peptide motifs. FHA folds into a beta-sandwich consisting of two beta sheets. The FHA domain is found in hundreds of prokaryotic and eukaryotic

proteins, including the damage-signaling proteins Mdc1 (a mediator), Nbs1 (a component of the MRN complex), and Chk2. The BRCT domain is often found in tandem repeats in proteins involved in DNA damage signaling and repair, including BRCA1, Mdc1, Nbs1, 53BP1, and DNA ligase IV. Binding to phospho-peptide motifs requires two tandem BRCT domains. The N-terminal BRCT domain stabilizes the phospho-serine, and the C-terminal tandem domain forms a hydrophobic pocket that interacts with the aromatic residue preferentially found at the +3 position in the consensus binding sequence.

Effector Kinases: Chk1 and Chk2

Downstream of ATR and ATM in the cell-cycle checkpoint pathways are two evolutionarily conserved protein kinases Chk1 and Chk2. While ATM and ATR belong to the PIKK family of kinases, Chk1 and Chk2 are members of the protein kinase super-family, but they segregate into different subfamilies of serine/threonine kinases. The Chk1 and Chk2 kinases are directly phosphorylated by ATR and ATM, with Chk1 being primarily activated by ATR and Chk2 by ATM. There is significant crosstalk between the ATR-Chk1 and ATM-Chk2 signaling cascades. The activation of Chk2 is mediated by ATM-phosphorylation-induced multimerization via its FHA domain, resulting in Chk2 auto-phosphorylation and enhanced kinase activity. Interactions with mediator proteins also serve to control the biological activities of the Chk1 and Chk2 kinases through regulation of their subcellular distribution and their stability. As discussed above, the Chk1 and Chk2 kinases negatively regulate the Cdc25 family of phosphatases through phosphorylations to activate the G1/S and the G2/M checkpoints. These effectors also regulate gene expression via phosphorylation of transcriptional activators and repressors, and DNA repair pathways through modification of repair proteins.

Chromatin Modifications

Chromatin modification is an important component of the checkpoint network. Posttranslational modifications on histones, together with chromatin remodeling proteins, facilitate access of repair proteins to damaged DNA and contribute to amplification as well as termination of the damage signal. Phosphorylation of H2AX occurs rapidly in response to damage, and the resulting γ H2AX acts as a platform to recruit and retain signaling and repair proteins. This chromatin landmark plays an essential role in checkpoint activation and chromatin remodeling that involves factors such as INO80, NuA4, and Swr1. Additional posttranslational modifications on H2AX and other histones include tyrosine phosphorylation of H2AX that appears to inhibit ATM/ATR-mediated serine phosphorylation, ubiquitination of H2A and H2AX, sumoylation of H2AZ, tyrosine phosphorylation on H3, acetylation on H2AX, and methylation of histone H4. The pathways leading from DNA lesions to these histone modifications are an active area of investigation.

Spatial and Temporal Organization of the DNA Damage Response

The dynamic redistribution of DNA repair and damage-signaling proteins has been observed in response to the acute

accumulation of DNA lesions. For example, ionizing radiation and radiomimetic drugs induce the formation of nuclear foci (aka IRIF). The techniques of laser micro-irradiation and live cell imaging have been combined to generate spatially defined lesions that facilitate monitoring of the accumulation and phosphorylation of specific proteins to damage sites. Although the current imaging technology may not detect proteins of low abundance and/or proteins that do not accumulate at lesions, they have revealed a series of temporally distinct steps. Immediately following the generation of DSBs, the acetylation and auto-phosphorylation of ATM are observed, concomitant with the phosphorylation of H2AX on Ser139 to generate γ H2AX. Formation of γ H2AX occurs initially near the break site and is then extended outward from the break site for several megabases. The resulting pSQ of γ H2AX is recognized by the tandem BRCT domains of the mediator Mdc1. A repeated pSTD motif in Mdc1 interacts with the FHA and BRCT domains of Nbs1, and facilitates retention of the MRN complex at the break site. The Mdc1 protein is also directly phosphorylated by ATM, and through this phosphorylation event, Mdc1 binds the FHA domain of the ubiquitin ligase RNF8. Recruitment of RNF8 leads to ubiquitination of H2A and H2AX, and the ubiquitinated marks are recognized by RNF168, a second ubiquitin ligase that amplifies the ubiquitination output at the break site. The resulting polyubiquitin chains that accumulate at the break sites further recruit additional factors, for example, the Rap80 protein which contains tandem ubiquitin interacting motifs (UIMs) that bind to polyubiquitin chains. Rap80, together with a coiled-coil scaffold protein Abraxis, mediates the recruitment of BRCA1, another ubiquitin ligase involved in the regulation of homologous recombination and DSB repair. Isopeptidases involved in the breakdown of ubiquitin chains are also part of the complexes assembled at the break sites to regulate DNA damage signaling. The current results suggest that phosphorylation- and ubiquitin-mediated interactions through γ H2AX, Mdc1, ATM, and Nbs1 create a positive feedback loop that amplifies the lesion signals to stimulate repair and cell-cycle checkpoints.

DNA Damage-Induced Apoptosis and Senescence

Under normal physiological conditions, the damage surveillance network coordinates DNA repair with cell-cycle progression to achieve the biological goal of maintaining genome integrity and cell proliferation. The importance of this surveillance network in normal cell growth is best illustrated by the findings that genetic ablation of either ATR or Chk1 cannot be tolerated and leads to cell lethality within one to two cell divisions. The death of ATR-deficient cells cannot be rescued by inactivating the cell death machinery, because the death is rooted in the inability of cells to replicate the genome properly. In contrast to this passive death caused by the loss of ATR, the damage surveillance network can actively stimulate programmed cell death, either in the form of apoptosis that shrinks and breaks a damaged cell or in the form of premature senescence that causes a damaged cell to withdraw irretrievably from the cell cycle.

Unlike the reversible mechanisms of cell-cycle checkpoints, apoptosis and senescence are both irreversible cell fate

decisions. Furthermore, while cell-cycle checkpoints are activated to protect the damaged cell, apoptosis and senescence are activated to destroy the damaged cell. Thus, the DNA damage signaling network has the capability of achieving opposing biological outcomes. Indeed, genetic studies have demonstrated that ATM and Chk2 are required for DSBs to activate cell-cycle checkpoints and apoptosis. The key determinant in the biological outcome downstream of ATM/Chk2 activation is time, because the G1/S, intra-S, and G2/M checkpoints are activated immediately following kinase activation, whereas apoptosis or senescence is invariably a delayed response occurring hours later. The delayed activation of apoptosis or senescence suggests that these biological outcomes may be regulated by the strength and/or the duration of the damage signals.

One of the key effectors in DSB-induced apoptosis or senescence is p53, which is a transcription factor activated by ATM and Chk2. As discussed above, p53 contributes to the maintenance of the G1/S checkpoint by stimulating the expression of p21Cip1 to inhibit Cdk2/cyclin. Activated p53 also stimulates the expression of Mdm2, which ubiquitinates p53 for proteasome-mediated degradation, and this negative feedback loop limits the duration of p53 activation. At doses that do not trigger apoptosis or senescence, a single peak of p53 activity is observed in cells following IR exposure. However, repeated rise and fall in p53 activity is observed in cells treated with IR doses that trigger apoptosis or senescence. These findings suggest that persistent signaling from unresolved DNA lesions causes repeated activation of p53 to trigger the delayed responses. It should be noted that activation of p53 can lead either to apoptosis or to senescence, depending on the cell type. It is therefore important to emphasize that the function of p53 *per se* does not determine the biological outcome following DNA damage. Rather, the strength and the duration of the lesion-induced signal, as well as the cell context, are the critical determinants that control the irreversible cell fate decisions activated by the DNA damage surveillance network.

The ATM-Chk2-p53 pathway is also required for the induction of apoptosis or senescence in response to telomere shortening in somatic mammalian cells. This finding illustrates yet another biological function of the DNA damage surveillance network and underscores the many diseases associated with defects in this signaling network.

Checkpoint Proteins and Human Diseases

Defects in components of the DNA damage response pathways are the underlying causes of a number of human disorders. Cells derived from patients with these disorders have provided valuable reagents to dissect the workings of these pathways. Notably, homozygous mutations in ATM cause ataxia-telangiectasia (A-T), a condition characterized by cerebellar degeneration, immunodeficiency, sterility, genome instability, clinical radiosensitivity, and cancer predisposition. Cells from A-T patients exhibit defects in G1/S, intra-S, and G2/M checkpoints in response to IR. Hypomorphic ATR mutation causing aberrant splicing and partial loss of ATR activity in humans have been associated with the autosomal recessive disorder Seckel syndrome, which shares several clinical features in common with A-T. The Seckel syndrome cells exhibit increased

sensitivity to UV-induced stalling of DNA replication forks and defective G2/M arrest. Hypomorphic mutations in the *NBS1* gene result in Nijmegen breakage syndrome (NBS), a rare autosomal recessive disorder characterized by microcephaly, immunodeficiency, and predisposition to cancer. The most common mutation in NBS patients is 657del5, which results in truncated proteins that maintain some functions of the full-length protein. Hypomorphic mutations in human *MRE11* lead to ataxia-telangiectasia-like disorder (A-TLD), in which patients display ataxia and neurodegeneration, resembling the phenotypes of ATM deficiency. A single patient has been reported with a hypomorphic mutation in *RAD50*, who exhibited phenotypes similar to NBS. Consistent with the involvement of MRN in cell-cycle checkpoint signaling and DNA repair, cells from patients with NBS and A-TLD display increased radio-sensitivity and are defective for checkpoint activation. The RNF168 ubiquitin ligase important for the recruitment of damage-signaling proteins to DSBs is mutated in a syndrome of radiosensitivity, immunodeficiency, and learning difficulty, referred to as RIDDLE syndrome.

With its critical roles in genome maintenance and the destruction of damaged cells, the DNA damage surveillance and response network function as a barrier to neoplastic transformation. Mutations in genes that encode proteins in this network contribute to hereditary and sporadic cancers. For example, germline mutations in p53 or Chk2 are found in Li-Fraumeni syndrome patients, which develop multiple tumors at various organ sites at young age. As discussed above, A-T patients are also prone to cancer development. Other notable tumor suppressors from this network include the mismatch repair proteins, MSH2 and MLH1, BRCA1 and its interacting proteins.

Concluding Remarks

This article describes in broad strokes the basic principles of DNA damage-activated cell-cycle checkpoints. Even at this level of discussion, it is clear that much remains to be deciphered regarding the mechanisms of damage recognition and checkpoint activation. This article focuses on the activation of the checkpoints and its biological consequences, but obviously a complementary complex system exists to terminate the checkpoints and allow cell-cycle progression following DNA repair. Inhibitors of the checkpoint kinases as well as protein phosphatases are likely to be involved in checkpoint termination upon the resolution of DNA lesions or stalled replication

forks. Although little is known about the checkpoint termination processes in mammalian systems, the PP2A phosphatase can regulate ATM phosphorylation in human cells, and phosphatases Wip1/PPM1D may antagonize the activation of Chk1 and Chk2. In addition, ATR-mediated phosphorylation of Chk1 not only activates this protein kinase but also marks it for ubiquitination and proteasomal degradation. Checkpoint termination may also require re-establishment of the original chromatin conformation around the break sites. Independent of signal termination, parts of the checkpoint pathways can be inactivated to restore cell-cycle progression despite the persistence of DNA lesions. The mechanisms underlying checkpoint adaptation to DNA lesions are not completely understood. It is obvious that adaptation would restore cell division at the expense of genome integrity; thus, checkpoint adaptation may promote malignant progression of cancer cells. Since genome integrity is the foundation of normal cellular functions, it will be important to achieve a comprehensive understanding of the DNA damage surveillance and response network.

See also: [Cell Architecture and Function: Cell Cycle: Control of Entry and Progression through S Phase](#); [Cell-Cycle Controls in G₁ and G₀](#).

Further Reading

- Borges HL, Linden R, and Wang JYJ (2008) DNA damage-induced cell death: Lessons from the central nervous system. *Cell Research* 18: 17–26.
- Chaurushiya MS and Weitzman MD (2009) Viral manipulation of DNA repair and cell cycle checkpoints. *DNA Repair* 8: 1166–1176.
- Cimprich KA and Cortez D (2008) ATR: An essential regulator of genome integrity. *Nature Reviews Molecular Cell Biology* 9: 616–627.
- Harper JW and Elledge SJ (2007) The DNA damage response: Ten years after. *Molecular Cell* 28: 739–745.
- Harrison JC and Haber JE (2006) Surviving the breakup: The DNA damage checkpoint. *Annual Review of Genetics* 40: 209–235.
- Huen MS and Chen J (2008) The DNA damage response pathways: At the crossroads of protein modifications. *Cell Research* 18: 8–16.
- Jackson SP and Bartek J (2009) The DNA-damage response in human biology and disease. *Nature* 461: 1071–1078.
- Lavin MF (2007) ATM and the Mre11 complex combine to recognize and signal DNA double-strand breaks. *Oncogene* 26: 7749–7758.
- Li L and Zou L (2005) Sensing, signaling, and responding to DNA damage: Organization of the checkpoint pathways in mammalian cells. *Journal of Cellular Biochemistry* 94: 298–306.
- Sancar A, Lindsey-Boltz LA, Unsal-Kagmaz K, and Linn S (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annual Review of Genetics* 73: 39–85.
- Stracker TH, Usui T, and Petrini JH (2009) Taking the time to make important decisions: The checkpoint effector kinases Chk1 and Chk2 and the DNA damage response. *DNA Repair* 8: 1047–1054.

Cell Cycle: Mitotic Checkpoint

T J Yen, Fox Chase Cancer Center, Philadelphia, PA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Anaphase-promoting complex (APC) A multisubunit E3 ubiquitin ligase that promotes the proteolytic degradation of substrates such as cyclin B and securin in order for cells to exit mitosis.

Kinetochores A macromolecular complex that is assembled at centromeres near the onset of mitosis that is essential for chromosomes to attach to the spindle.

Kinetochores tension The force that is developed between sister kinetochores because of opposing poleward forces that attempt to separate them.

Monopolar and bipolar States of attachment of chromosomes to either a single pole or to both poles of a spindle.

Wait anaphase signal An inhibitory signal generated from unattached kinetochores that diffuses throughout the cell to block the onset of anaphase.

The mitotic checkpoint is a fail-safe mechanism that evolved to ensure that cells with even a single unaligned chromosome do not exit mitosis to produce aneuploid cells. The mitotic checkpoint solves an inherent problem that arises because of the stochastic nature by which chromosomes attach to the spindle. At the onset of mitosis, chromosomes are randomly distributed throughout the cell so that not all chromosomes will achieve alignment at the spindle equator at the same time. Chromosomes located in the center of the spindle will rapidly establish bipolar attachments due to their higher frequency of encounters with microtubules. By contrast, chromosomes located near a pole will rapidly attach to that pole but attachment to the opposite pole will take more time given the lower frequency that it will find the rare microtubule that originates from the opposite pole. As the kinetochores are the structure on chromosomes that establishes connections with spindle microtubules, the checkpoint monitors this site to determine if chromosomes are properly attached to the spindle and whether it has achieved alignment.

The goal of this article is to discuss current models that explain how the checkpoint monitors kinetochores functions and how a localized defect that is restricted in space can alter the global biochemical status of a cell.

The Mitotic Checkpoint Monitors Kinetochores

The checkpoint is able to sense the status of chromosome alignment by monitoring microtubule occupancy and tension (or the lack thereof) at kinetochores. Once a defect is detected, the kinetochores generate an inhibitory signal to block the onset of anaphase. Early studies showed that a single unattached kinetochores is sufficient to arrest cells in mitosis. Thus, the mitotic checkpoint functions much like a signal transduction cascade where a defective kinetochores generates a signal that must be amplified throughout the cell to inhibit its targets. The discovery of the molecular components of the mitotic checkpoint has shed considerable light into how the mechanical activities at the kinetochores can regulate mitotic progression.

Molecular Components of the Mitotic Checkpoint

The mitotic checkpoint is specified by six evolutionarily conserved genes, BUB1, BUB3, MAD1, MAD2, MAD3, and MPS1. All of these genes are essential as inactivation of any single gene prevents cells from delaying mitosis in the presence of unaligned chromosomes. BUB1 and MPS1 are protein kinases, while the biochemical activities of the remaining proteins are unknown. It appears that MAD3 in metazoans has evolved into BUBR1, a third protein kinase of the mitotic checkpoint. It is interesting that BUBR1 may have co-evolved with centromere protein E (CENP-E), a kinetochores motor whose activity is thought to be monitored by BUBR1 and is also only present in metazoans. With the identification of checkpoint genes, the outstanding question is how their gene products interact at the molecular and biochemical levels to specify the signaling pathway that links a defective kinetochores to inhibition of the anaphase-promoting complex (APC).

Monitoring Microtubule Occupancy and Tension at Kinetochores

The discovery that mitotic checkpoint proteins preferentially localize to unattached kinetochores suggested their roles in monitoring kinetochores activities and generating the wait anaphase signal. In this regard, the role of the MAD2 checkpoint protein at kinetochores has been extensively characterized. Quantitative studies have shown that the staining intensity of MAD2 at unattached kinetochores can be nearly 100-fold higher than that detected at kinetochores that are fully saturated with microtubules. By contrast, the intensity of BUB1 and BUBR1 staining varies only three- to fivefold between unattached and attached kinetochores. It is now evident that MAD2 localization is sensitive to microtubule attachments rather than to kinetochores tension. When microtubule dynamics is suppressed by drugs (taxol or low concentrations of nocodazole or vinblastine) or low temperature, kinetochores are fully attached with microtubules, but no tension develops.

As MAD2 is not detected at these kinetochores, the conclusion was that MAD2 monitors microtubule attachment, but not the tension. These studies therefore reinforced an early notion that kinetochore tension is monitored by the checkpoint.

How tension is monitored by the checkpoint is not fully understood but it clearly involves aurora B/Ipl1, a protein kinase that is situated between sister kinetochores where tension is developed. Recent studies in yeast and mammalian cells showed that aurora B/Ipl1 is required for cells to arrest in mitosis when their kinetochores fail to establish tension despite being saturated with microtubules. Thus, aurora B was found to be essential for cells to arrest in the presence of taxol, a drug that suppresses microtubule dynamics and thus prevents kinetochores from developing tension. Despite this data, auroraB/Ipl1 does not behave like the other checkpoint proteins, as it is not required for cells to arrest in mitosis in response to the loss of microtubule attachments. Aurora B/Ipl1 may therefore provide tension-sensing checkpoint functions.

AuroraB/Ipl1 has been proposed to be part of a mechanism that ensures that improperly attached kinetochores are provided the opportunity to establish proper connections to the spindle. The need for a self-correcting mechanism is evident because of the error-prone nature by which kinetochores establish microtubule connections. Merotelic and syntelic attachments are conditions where both kinetochores are connected to the same pole or when one kinetochore is attached to opposite pole, respectively. As kinetochores with merotelic and syntelic attachments are fully saturated with microtubules, a checkpoint that is only sensitive to microtubule occupancy at kinetochores will fail to detect these aberrant connections. AuroraB/Ipl1 is thought to monitor merotelic and syntelic attachments as these abnormal connections accumulated when this kinase was inactivated. Aurora B/Ipl1 is thought to resolve merotelic and syntelic attachments by stimulating the release of microtubules from these kinetochores and thus promoting new rounds of interactions. Given this scenario, it is possible that kinetochores lacking tension do not directly activate the checkpoint but do so as a result of microtubule detachments that is stimulated by aurora B kinase. This could explain why, in taxol arrested cells, there is on average one kinetochore that exhibits MAD2 staining. It may be possible that this is sufficient to arrest these cells in mitosis.

The mechanism by which aurora B promotes microtubule release from kinetochores has not been clarified but is likely mediated through mitotic centromere-associated kinesin, an unconventional kinesin-like protein that colocalizes with aurora B and functions to depolymerize microtubules.

The Mitotic Checkpoint Monitors a Kinetochore Motor

CENP-E is a kinesin-like protein that binds to kinetochores at the onset of mitosis where it plays a critical role in establishing microtubule attachments. The link between CENP-E and the mitotic checkpoint was first established when cells whose kinetochores were depleted of CENP-E were found to arrest in mitosis. Cells defective for CENP-E functions characteristically accumulate a few chromosomes with monopolar attachments while the majority establish bipolar connections. The inability of monopolar chromosomes to establish bipolar

connections is thought to reflect the importance of CENP-E in allowing kinetochores to capture the rare microtubules that emanate from the opposite pole. On the other hand, chromosomes that are positioned near the middle of the cell can establish bipolar attachments because the higher-frequency encounters with microtubules from both poles can compensate for the loss of CENP-E. Quantitative electromagnetic analysis showed that the bipolar-attached kinetochores lack tension despite attaining near normal numbers of microtubule attachments. By contrast, the kinetochores of the monopolar chromosomes established very few microtubule connections. The defects of the monopolar versus the bipolar-attached chromosomes are viewed differently by the mitotic checkpoint. Consistent with its role in detecting microtubule attachments, MAD2 accumulated at kinetochores of the monopolar chromosomes, while it was undetectable at kinetochores of the bipolar chromosomes. These findings clearly show that the mitotic delay induced by the loss of CENP-E may be due to loss of microtubule attachments and tension.

The precise role of CENP-E in the mitotic checkpoint has been somewhat controversial. Studies in human cells suggested that it was not essential for the checkpoint as cells were able to arrest in the absence of CENP-E. By contrast, *Xenopus* egg extracts depleted of CENP-E and hepatocytes derived from CENP-E null mice failed to delay mitosis in the presence of unaligned chromosomes. The difference in response, however, is most likely a function of whether the assembly of checkpoint proteins to kinetochores is dependent on CENP-E. In egg extracts, the localization of a number of checkpoint proteins to kinetochores was absolutely dependent on CENP-E. Thus, the failure of egg extracts to delay mitosis in the absence of CENP-E can be ascribed to the absence of checkpoint proteins at unattached kinetochores. This contrasts with human cells where checkpoint proteins present at kinetochores that were depleted of CENP-E. However, quantitative analysis showed that in mouse and human cells, CENP-E did affect the assembly of checkpoint proteins at kinetochores. In both species, kinetochores depleted of CENP-E had approximately two- to fourfold lower levels of MAD1, MAD2, and BUBR1 than normal. However, this did not resolve why mouse cells failed to arrest in mitosis in the absence of CENP-E while human cells arrested. The discrepancy could be resolved if we take into account the fact that far fewer monopolar chromosomes accumulated in CENP-E depleted mouse cells than in human cells (two vs. eight). Given that the reduction in the amount of checkpoint proteins at these kinetochores likely compromised their ability to generate the wait anaphase signal, the combined output from all of the unattached kinetochores in mouse cells may not reach a threshold level that is required to sustain a prolonged arrest. In human cells, this threshold may be achieved as a result of the higher numbers of unattached kinetochores. Indeed, both mouse and human cells depleted of CENP-E were able to arrest in mitosis if the number of unattached kinetochores were increased by using drugs that inhibit spindle formation.

The mechanism by which CENP-E function is monitored by the checkpoint is believed to be mediated by the BUBR1 kinase. This connection was uncovered when hBUBR1 was identified in a yeast two-hybrid screen for proteins that interacted with CENP-E. This interaction was subsequently validated

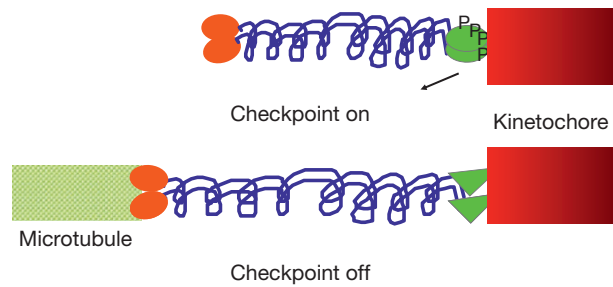


Figure 1 BUBR1 kinase acts as a mechanosensor that monitors centromere protein E (CENP-E) at kinetochores. BUBR1 kinase (green) interacts near the carboxy-terminus of CENP-E where it is postulated to monitor the microtubule-binding status of CENP-E. In the absence of microtubule attachments, BUBR1 is phosphorylated and activated (green oval) so that it generates the wait anaphase signal. When CENP-E is properly attached to microtubules, it undergoes a conformational change that inhibits BUBR1 kinase (green triangles) and thus extinguishes the wait anaphase signal.

when CENP-E and hBUBR1 were found to form a complex in cells. This finding coupled to the fact that hBUBR1 is an essential checkpoint protein suggested that hBUBR1 might act as a mechanosensor to monitor the activities of CENP-E at kinetochores (Figure 1). Consistent with this, hBUBR1 was found to be required for cells to arrest in mitosis when CENP-E functions were inhibited. The mechanism by which hBUBR1 monitors CENP-E activity remains to be clarified, but the working hypothesis is that hBUBR1 kinase activity is sensitive to interactions between CENP-E and microtubules. In the absence of microtubule interactions, CENP-E assumes a conformation that stimulates hBUBR1 to generate the wait anaphase signal. When CENP-E is interacting productively with microtubules, then hBUBR1 is silenced. Recent *in vitro* data showed that both *Xenopus* and human CENP-E can directly bind to BUBR1, and this binding stimulated its kinase activity. As microtubules were not present in these reactions, it is possible that the interaction reflected the checkpoint-on state. Addition of a CENP-E antibody to the CENP-E:BUBR1 complex inhibited kinase activity without disrupting the complex. The interpretation of this finding was that antibody binding altered the conformation of CENP-E so that it was unable to stimulate BUBR1 kinase activity. Thus, kinase activity of BUBR1 may be regulated allosterically by different conformational states of CENP-E. Although this model may indeed apply to the situation *in vivo*, the exact details of the mechanism remain to be clarified. For example, the *in vitro* dependence of BUBR1 kinase activity on CENP-E does not account for how cells that are depleted of CENP-E arrest in mitosis.

Inhibition of Mitotic Exit by the Checkpoint

Checkpoint proteins are functionally complex as they participate in multiple steps along the checkpoint-signaling pathway. A confluence of genetic and biochemical studies showed that the target of the checkpoint is the APC, an E3 ubiquitin ligase that promotes the degradation of substrates that inhibit the onset of anaphase.

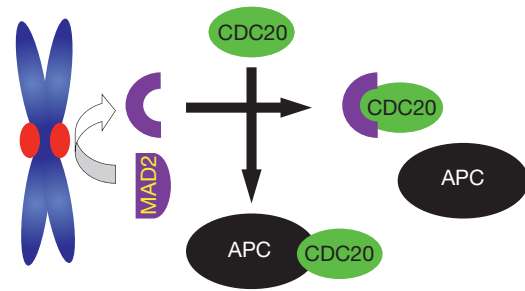


Figure 2 Sequestration model. The MAD2 checkpoint protein (purple semicircle) transiently binds to unattached kinetochores (red ovals) where they are converted to a form that upon dissociation will bind to CDC20. CDC20 when associated with MAD2 is incapable of activating the anaphase-promoting complex (APC). In this model, the pool of MAD2 that is released from kinetochores acts as the wait anaphase signal by indirectly preventing the activation of APC.

Sequestration Model

Studies in budding and fission yeast identified checkpoint-defective alleles of CDC20/Slp1, a protein that was shown biochemically to recruit substrates to the APC. These CDC20 and Slp1 mutants were unable to bind to MAD2 and thus supported *in vitro* data that MAD2 can inhibit the ubiquitin ligase activity of CDC20-dependent APC. These findings coupled with the *in vivo* observation that MAD2 exhibited a rapid rate of turnover at unattached kinetochores led to a model that described how MAD2 might act as the diffusible wait anaphase signal. In this sequestration model, unattached kinetochores are thought to catalytically convert MAD2 into an activated state, which upon release from kinetochores binds to CDC20 and sequesters it from the APC (Figure 2). The existence of an activated MAD2 initially gained support when it was discovered that recombinant MAD2 existed as monomers and oligomers (di- and tetramers), but the oligomeric form was more potent at inhibiting APC than the monomeric form *in vitro*. However, the existence of a MAD2 oligomer *in vivo* has been elusive. Furthermore, MAD2 mutants that failed to form tetramers *in vitro* were found to be well tolerated in yeast. These data suggest that if an activated MAD2 does exist *in vivo*, it is unlikely to be a tetramer.

Indeed, structural studies showed that MAD2 adopts two conformations that are either active or inactive. O-MAD2 is the open conformer that is inactive while the C-MAD2, closed conformer, exhibits checkpoint activity. The two conformers exhibit distinct structures mainly at the N- and C-termini of the molecules, specifically in the arrangement and position of two C-terminal β -strands and connecting loops. When MAD2 is bound to either MAD1 or CDC20, this region adopts a closed safety belt-like loop that defines the C-MAD2 conformer. Although the predominant conformer in cells is usually O-MAD2, during mitosis a MAD1:C-MAD2 complex that is present at unattached kinetochores is believed to catalytically convert the cytosolic pool of O-MAD2 into C-MAD2. In this template model, C-MAD2 is generated at kinetochores and, upon its release, it can form CDC20:C-MAD2 complexes that

may further amplify the production of C-MAD2 by converting additional molecules of O-MAD2.

A MAD2 dimerization domain, mainly involving residues at its α C helix (122–142 residues in human MAD2), is fundamental in mediating the O to C-MAD2 conversion. However, the rate at which O-MAD2 spontaneously converts into C-MAD2 is very slow. Thus, recombinant wild-type MAD2 consists largely of the O-MAD2 monomer, with small amounts of O:C heterodimer and C:C homodimer. Acceleration of the O→C conversion has been achieved *in vitro* with purified liganded C-MAD2, in the form of MAD1:C-MAD2 or CDC20:C-MAD2 complexes. This conversion depends on transient O:C heterodimerization. Liganded C-MAD2 seems incapable of forming C:C homodimers due to steric clashes at the MAD2 dimerization interface. Interestingly, p31^{comet}, a negative regulator of the mitotic checkpoint, was shown to exploit the dimerization interface to block O→C-MAD2 conversion during mitosis. Through structural mimicry, p31^{comet} binds to C-MAD2 at the α C helix thus preventing access of O-MAD2 for C-MAD2 amplification. Recent results also suggested that p31^{comet} may function downstream of the O→C conversion step during mitotic checkpoint signal transduction. The key question is to understand how p31^{comet} activity is regulated so that it knows when it is needed to silence the checkpoint to promote mitotic exit.

The CDC20:C-MAD2 interaction was thought to represent the terminal step of the mitotic checkpoint pathway, by sequestering CDC20 from APC/C activation. However, the long-known association of MAD2 with core APC/C subunits is inconsistent with the simple sequestration model. Although C-MAD2 is clearly essential for checkpoint activity, a mathematical model taking into consideration all known biochemical properties of C-MAD2 predicted that >5 h would be required to sequester 50% of intracellular CDC20. As an unperturbed mitosis is completed (onset of anaphase) within

30–40 min of mitotic entry, additional mechanisms must exist for rapid and efficient APC/C^{CDC20} inhibition. The checkpoint protein BUBR1 displays higher affinity toward CDC20 compared to MAD2 and has also been proposed to sequester CDC20 from APC/C. However, as with MAD2, BUBR1 alone is unable to inhibit pre-assembled mitotic APC/C^{CDC20} holoenzyme at physiologically relevant concentrations.

Direct Inhibitor Model

The direct inhibitor model differs from the sequestration model in that the APC is directly inhibited by checkpoint proteins and posits that the unattached kinetochore may sensitize the APC to its inhibitor (rather than generating a factor that sequesters an activator of the APC) (Figure 3). This model originated from the discovery of the mitotic checkpoint complex (MCC), a factor that was identified by testing fractionated cell extracts for inhibitory activity against mitotic APC. The MCC was originally identified in Hela cells, where it was found to consist of near stoichiometric amounts of hBUBR1, hBUB3, CDC20, and MAD2. The interesting finding was that the MAD2 that was present in the MCC represented a very small fraction (<5%) of the total pool of MAD2. Despite the presence of a large pool of monomeric MAD2, this pool was not found to inhibit the mitotic APC. Furthermore, MCC was more than 3000-fold more potent inhibitor of the APC than recombinant MAD2 (which was used in all the previous studies). The large difference in the inhibitory activities suggests that Hela cells do not express sufficient amounts of monomeric MAD2 to inhibit the APC. The mechanism by which MCC inhibited APC activity is not clear but is likely to depend on its ability to bind to the APC. When purified from cells, APC that was associated with MCC was found to be inactive, while

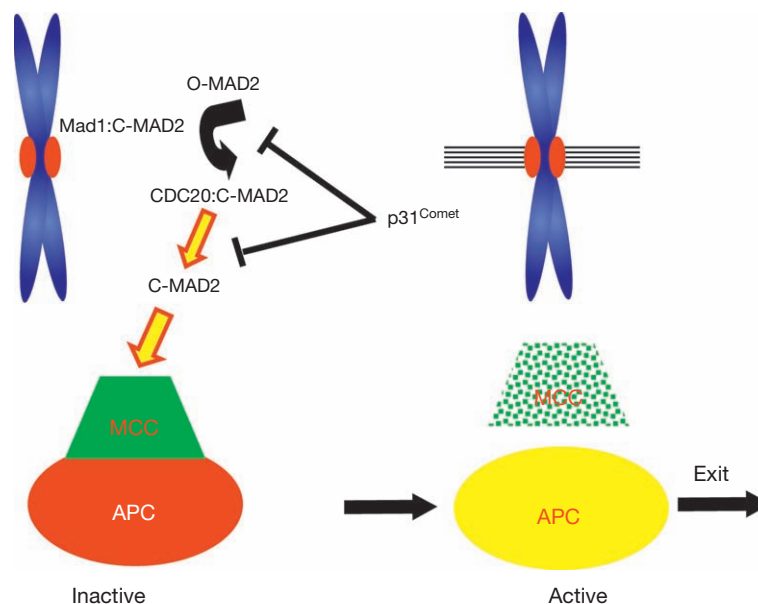


Figure 3 Direct inhibition model. BUBR1, Bub3, CDC20, and MAD2 form the mitotic checkpoint complex (MCC), which binds and directly inhibits the APC. Formation of MCC depends on C-MAD2, which is generated from O-MAD2 through its conversion at kinetochores (red ovals). Once kinetochores have established proper attachments to the spindle, p31^{Comet} is activated in an unknown way to block the O to C conversion. Loss of C-MAD2 production leads to disassembly of the MCC (stippled) and release of the APC from checkpoint inhibition.

APC that lacked MCC was highly active. Based on these findings, the MCC was proposed to be the primary checkpoint inhibitor of the APC in mitotically arrested cells.

Although MCC was identified in mitotic HeLa cells, it was also reported to be present and active throughout the cell cycle. As kinetochores are only present in mitosis, the existence of MCC during interphase suggested that MCC formation occurred independently of kinetochores. More recent studies, however, established that MCC is formed only in mitosis, but subcomplexes of BubR1/Bub3 and BubR1/Bub3/Cdc20 are detected as cells progress from early to late stages of interphase (G1 through G2), respectively. Why MCC formation occurs through intermediates consisting of BubR1/Bub3 to BubR1/Bub3/Cdc20 and, finally, BubR1/Bub3/Cdc20/Mad2 remains unclear. The latest findings show that MCC formation is dependent on kinetochores, and the previously reported APC inhibitory activity of interphase MCC was likely mediated by intermediates of MCC (BubR1/Bub3/Cdc20).

The existence of the MCC has been confirmed in other species including *Xenopus* and yeast. Nevertheless, the connection between MCC and the wait anaphase signal (C-MAD2 in particular) has only been postulated but not experimentally established. In fact, discrepancies in the description of the MCC has clouded its role in the checkpoint pathway. The discrepancies included whether MAD2 was part of the MCC, and whether MCC formation was dependent on unattached kinetochores. Differing reports claimed no or only substoichiometric amounts of MAD2 associated with BUBR1 or APC/C; and whether MCC formation was detected in interphase cells or its formation was independent of unattached kinetochores.

These issues were recently resolved with the discovery that MCC selectively recruits only C-MAD2. These studies were initiated because many of the reports that claimed MCC did not contain MAD2 relied on experiments that used wild-type MAD2, which exists predominantly in the inactive O-conformer. The availability of a collection of MAD2 mutants that were locked in either the O or C-conformers, in combination with dimerization defective mutants, prompted a systematic analysis of MCC formation. Both *in vivo* and *in vitro* data established that MCC selectively incorporated C-MAD2, more specifically, C-MAD2 homodimers (see below). As C-MAD2 is generated from kinetochores, it explains how MCC formation is limited to mitosis. The selectivity of C-MAD2 for the MCC also clarifies discrepancies among earlier reports regarding the presence and amounts of MAD2 in the MCC. Differences in MAD2 levels reported in MCC preparations may be ascribed to variable O:C ratios of MAD2^{WT} that exist in those preparations. MAD2^{WT} is composed of both O- and C-conformers at steady state, but as the O-conformer is the predominant topological form *in vitro* and *in vivo*, the majority of MAD2^{WT} would not be expected to be part of the MCC. Based on our current data, the MAD2^{WT} present in the MCC isolated from cells or assembled *in vitro* represents the small fraction of C-conformer.

Several lines of evidence from these studies showed that incorporation of C-MAD2 into the MCC stimulated both its binding and inhibition of APC/C. The results are in agreement with the original report that partially purified MCC displayed potent inhibitory activity toward mitotic APC/C than recombinant Mad2 alone. Recently, Herzog et al. demonstrated that in

HeLa cells that were arrested in mitosis, APC/C existed either as apo-APC/C, that lacked the CDC20 activator, or as a supra-complex with MCC containing MAD2. The apo-APC/C might be due to sequestration of CDC20 by C-MAD2 and BUBR1:BUB3 complex, but MCC provides an added layer of inhibition by targeting APC/C^{CDC20}.

The finding that C-MAD2 homodimers are incorporated into the MCC may shed light on the control mechanisms for silencing the mitotic checkpoint. Although homodimers of C-MAD2 have been reported, the functional role of this complex has remained enigmatic. This is due in part to structural studies that suggest that a C-MAD2 homodimer is incapable of catalyzing the conversion of O-MAD2. A role of the homodimer in MCC formation not only provides a functional role for this complex, but also fulfils a prediction based on mathematical modeling. Sear and Howard suggested the existence of two species of APC/C inhibitors to satisfy the time spans observed for both checkpoint activation and inactivation. Their model predicted that one species is dependent upon unattached kinetochores and is auto-amplifiable. C-MAD2 fulfils these criteria, but the calculations estimated that it would take >5 h for sufficient C-MAD2 to be made to sequester the cellular pool of CDC20. The model predicted the existence of a second species of inhibitor that would allow timely activation and inactivation of the checkpoint inhibitory signal. The model predicted that the second species is derived from the first species, but is incapable of auto-amplification. The C-MAD2 homodimer could be the second inhibitor given that it has been found to be part of the MCC. In accordance with the model, C-MAD2 homodimer sustains potent APC/C inhibitory activity but cannot generate more inhibitors (O to C-MAD2 conversion). If the affinity constant for the interaction between BUBR1 and C-MAD2 homodimer is relatively low, it ensures that MCC will disassemble when the production of C-MAD2 is halted after kinetochores have satisfied the checkpoint. On the other hand, it is more likely that separate mechanisms that disassemble MCC and its association with the APC/C may involve ubiquitylation of CDC20 or other ATP-dependent processes.

These latest findings resolve discrepancies regarding the existence of the MCC. These studies also confirmed that formation of the MCC occurs only in mitosis, and is therefore likely to depend on kinetochores. This eliminates an earlier model that suggested that MCC formation was independent of kinetochores.

Dual Roles for Mitotic Checkpoint Proteins

The discovery that the MCC contains proteins that are also localized at kinetochores suggested that they might play dual roles in the checkpoint. These proteins participate in monitoring kinetochore defects and then generating the wait anaphase signal. The same proteins also act downstream of the kinetochore by being part of the MCC that directly inhibits the APC.

Early studies of the human BUBR1 kinase showed that its kinase activity was essential for the mitotic checkpoint. Nevertheless, hBUBR1 purified from interphase cells as part of the MCC exhibits no detectable kinase activity, yet is fully capable of inhibiting APC. Likewise, *in vitro* studies showed that

inhibition of the CDC20-dependent APC activity by recombinant BUBR1 can also be independent of its kinase activity. The apparent contradictory findings could be resolved by postulating that kinase activity of BUBR1 was required for the kinetochore while kinase activity was not required for downstream components such as the MCC. This hypothesis has now been confirmed in *Xenopus* egg extracts by BUBR1 depletion and add-back experiments. When extracts depleted of BUBR1 are reprogrammed with a kinase-dead mutant, they are no longer able to activate their checkpoint. Interestingly, if the BUBR1 (wild type or mutant) was supplemented to 20% of the endogenous level, it is sufficient to saturate all the kinetochores to levels seen in normal kinetochores. However, the presence of BUBR1 at kinetochores is still insufficient to activate the checkpoint unless the cytosolic pool of BUBR1 is also restored. This suggests that BUBR1 might provide two separable functions in the checkpoint. Whether the cytosolic BUBR1 acts by sequestering CDC20 or as part of the MCC is not clear. The finding that either wild-type or kinase-dead BUBR1 can restore the checkpoint to kinetochores that had assembled BUBR1 wild-type kinase but not the mutant supports the idea that its kinase activity was essential at kinetochores but not in the cytosol.

Wait Anaphase Output Regulated by Checkpoint Proteins

Studies of hNuf2 and CENP-I led to the remarkable finding that MPS1, MAD1, and MAD2 do not appear to play an essential role at the kinetochore. How do kinetochores that lack MPS1, MAD1, and MAD2 maintain the checkpoint? This is most likely achieved in part by BUBR1 and BUB1 whose levels were not noticeably depleted. However, these proteins by themselves do not appear to be sufficient to generate a robust wait anaphase signal from unattached kinetochores. Cells defective for CENP-I accumulate a few monopolar chromosomes, while most chromosomes appear aligned (similar to loss of CENP-E). These cells are unable to sustain a prolonged mitotic arrest because the unattached kinetochores cannot generate sufficient amounts of wait anaphase signal to sustain a prolonged arrest. Consistent with this explanation, if the total number of unattached kinetochores was increased by disrupting the spindle with microtubule poisons, the collective output from a large number of unattached kinetochores must reach a critical threshold that is required for cells to arrest in mitosis. The caveat of this interpretation is that, after nocodazole treatment, kinetochores depleted of CENP-I exhibited detectable MAD2, although it was 20-fold lower than the level that is present at in-control cells. Nocodazole-treated

Hela cells easily contain more than 20 unattached kinetochores (Karyotype >60 chromosomes) and should therefore be able to generate a threshold level that can normally be achieved by a single unattached kinetochore.

See also: [Cell Architecture and Function: Cell Cycle: Control of Entry and Progression through S Phase](#); [Cell Cycle: DNA Damage Checkpoints](#); [Chromosome Organization and Structure, Overview](#); [Mitosis](#).

Further Reading

- Chan GK, Jablonski SA, Sudakin V, Hittle JC, and Yen TJ (1999) Human BUBR1 is a mitotic checkpoint kinase that monitors CENP-E functions at kinetochores and binds the cyclosome. *APC Journal of Cell Biology* 146: 941–954.
- Ciliberto A and Shah JV (2009) A quantitative systems view of the spindle assembly checkpoint. *EMBO Journal* 28: 2162–2173.
- Fang G, Yu H, and Kirschner MW (1998) Direct binding of CDC20 protein family members activates the anaphase-promoting complex in mitosis and G1. *Molecular Cell* 2: 163–171.
- Hauf S, Cole RW, La Terra S, et al. (2003) The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *Journal of Cell Biology* 161: 281–294.
- Huang H, Hittle J, Zappacosta F, et al. (2008) Phosphorylation sites in BubR1 that regulate kinetochore attachment, tension, and mitotic exit. *Journal of Cell Biology* 183: 667–680.
- Hunter AW, Caplour M, Coy DL, et al. (2003) The kinesin-related protein MCAK is a microtubule depolymerase that forms an ATP-hydrolyzing complex at microtubule ends. *Molecular Cell* 11: 445–457.
- Liu ST, Hittle JC, Jablonski SA, et al. (2003) Human CENP-I specifies localization of CENP-F, MAD1, and MAD2 to kinetochores and is essential for mitosis. *Nature Cell Biology* 5: 341–345.
- Mao Y, Abrieu A, and Cleveland DW (2003) Activating and silencing the mitotic checkpoint through CENP-E-dependent activation/inactivation of BubR1. *Cell* 114: 87–98.
- Martin-Lluesma S, Stucke VM, and Nigg EA (2002) Role of Hec1 in spindle checkpoint signaling and kinetochore recruitment of Mad1/Mad2. *Science* 297: 2267–2270.
- McEwen BF, Chan GKT, Zubrowski B, et al. (2001) CENP-E is essential for reliable bioriented spindle attachment, but chromosome alignment can be achieved via redundant mechanisms in mammalian cells. *Molecular Biology of the Cell* 12: 2776–2789.
- Sear RP and Howard M (2006) Modeling dual pathways for the metazoan spindle assembly checkpoint. *Proceedings of the National Academy of Sciences of the United States of America* 103: 16758–16763.
- Sudakin V, Chan GKT, and Yen TJ (2001) Checkpoint inhibition of the APC/C in HeLa cells is mediated by a complex of BUBR1, BUB3, CDC20, MAD2. *Journal of Cell Biology* 154: 925–936.
- Teichner A, Eytan E, Sitry-Shevah D, et al. (2011) p31comet promotes disassembly of the mitotic checkpoint complex in an ATP-dependent process. *Proceedings of the National Academy of Sciences of the United States of America* 108: 3187–3192.
- Tipton AR, Wang K, Link L, et al. (2011) BUBR1 and closed MAD2 (C-MAD2) interact directly to assemble a functional mitotic checkpoint complex. *Journal of Biological Chemistry* 286: 21173–21179.
- Yang M, Li B, Tomchick DR, et al. (2007) p31comet blocks Mad2 activation through structural mimicry. *Cell* 131: 744–755.

Cell Death by Apoptosis and Necrosis

V Ermer and P Nicotera, German Centre for Neurodegenerative Diseases, DZNE, Bonn, Germany

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Pierluigi Nicotera, volume 1, pp. 352–355, © 2004, Elsevier Inc.

Glossary

Calpains A family of Ca^{2+} -activated proteases that mediates some of the physiological effects of Ca^{2+} signals but that can also mediate cell death. Substrates include cytoskeletal proteins, other proteases, and Ca^{2+} transport proteins.

Caspases Cysteine aspartases that are activated during inflammatory processes and in apoptosis.

NCX The sodium–calcium exchanger that operates the transport of these two ions depending on their electrochemical gradients.

Plasma membrane Ca^{2+} pump ATPase (PMCA) Transmembrane Ca^{2+} pump that uses ATP to pump Ca^{2+} out of cells.

Various genetically encoded programs involved in the signaling, initiation, and execution of cell death decide cells' fate during development and adult life. These programs can execute physiological cell death during development or tissue turnover, and are also involved in the inappropriate elimination of cells under pathological conditions. Because balanced cell turnover is essential for life, defects in cell elimination can also result in disease, the foremost example being cancer. In many circumstances, both physiological cell death and cell death in pathological settings have similar morphological and biochemical characteristics. Perhaps the best-characterized biochemical and morphological changes during a cell death program are those defined as apoptosis. Apoptosis is characterized by condensation and fragmentation of the nucleus with shrinkage of the cytoplasm and exposure of surface molecules that facilitate recognition of the dying cells by phagocytes. However, other types of cell death are present and are strictly regulated *in vivo*, including cell lysis/necrosis or autophagy. Imbalance in cellular calcium regulation has been involved in both apoptotic and nonapoptotic cell death. Calcium can be a signal for cell death or simply a downstream consequence of the activation of the death machinery.

Ca^{2+} as a Signal for Cell Death

A sustained Ca^{2+} overload, such as that resulting from dysfunction of the main routes of Ca^{2+} entry or efflux or from the irreversible loss of intracellular buffering capacity, can be lethal. The idea that Ca^{2+} may be cytotoxic dates back to A Fleckenstein's suggestion in 1968 that excessive entry of Ca^{2+} into myocytes could be the underlying mechanism of cardiac pathology following ischemia. Subsequent studies showed that agonist stimulation or cytotoxic agents could cause lethal Ca^{2+} entry into cells. Cellular Ca^{2+} overload involves multiple intra- and extracellular routes, most of which are also used for physiological signaling, which implies that not only alterations of the normal Ca^{2+} homeostasis but also changes in Ca^{2+} signaling can have adverse effects. The following have been shown in a large number of experimental paradigms: (1) direct sustained elevation of $[\text{Ca}^{2+}]_i$ (e.g., by exposure of cells to ionophores or

to conditions that cause prolonged gating of inward-directed channels) causes cell death; (2) a $[\text{Ca}^{2+}]_i$ elevation precedes cell death induced by pathophysiological stimuli; (3) prevention of $[\text{Ca}^{2+}]_i$ elevation during such experiments can inhibit cell death; and (4) alterations of Ca^{2+} -signaling pathways (e.g., potentiation or inhibition of Ca^{2+} currents) can result in cytotoxicity.

Executors of Ca^{2+} Death Signals

Intracellular Ca^{2+} signals can set off cell demise via Ca^{2+} -dependent processes, change the mode of cell death from apoptosis to necrosis, or synergize with elements of the apoptotic death program. In particular, Ca^{2+} -activated proteases (calpains) can synergize with caspases and amplify apoptotic death routines, while modulators of apoptosis such as members of the Bcl-2 protein family can modulate Ca^{2+} compartmentalization. Hydrolytic enzymes, which include calpains, various DNases, and lipases, are the best-characterized effectors of cell death directly mediated by calcium overload. Calpains are Ca^{2+} -activated cysteine proteases that have been implicated in toxic cell death in the liver and in excitotoxic neuronal death. Calcium-dependent DNases can be responsible for DNA degradation, although the nature of the Ca^{2+} -dependent enzyme(s) responsible for the typical oligonucleosomal DNA cleavage has remained elusive. Among lipases, the Ca^{2+} -dependent phospholipase A_2 (PLA_2) has been implicated in neurotoxicity. Its activation results in the release of arachidonic acid and related polyunsaturated fatty acids, which are further metabolized by lipoxygenases or cyclooxygenases with concomitant generation of reactive oxygen species. In addition, PLA_2 activation generates lysophosphatids that alter the membrane structure, which may facilitate Ca^{2+} influx and Ca^{2+} release from internal stores.

Ca^{2+} , Excitotoxicity, and Death during Brain Ischemia

Excitotoxicity is a phenomenon typically encountered in neurons or myocytes following receptor stimulation by excitatory

amino acids that exceeds the physiological range with respect to duration or intensity. Typical excitotoxic stimulators are capsaicin, acetylcholine, or – most important in the central nervous system – glutamate. Direct injection of glutamate is selectively neurotoxic *in vivo*. Also, inhibition of excess synaptic activity by inhibitors of glutamate receptor subtypes (mainly *N*-methyl-D-aspartate, or NMDA) protects neurons from hypoxia. Generally, excitotoxicity is induced by conditions favoring glutamate accumulation in the extracellular space. Typical conditions leading to increased extracellular glutamate concentration are depolarization of neurons, energy depletion due to hypoglycemia or hypoxia, or defects in the glutamate reuptake systems.

Overall, three different lines of evidence suggest the key role of Ca^{2+} in excitotoxicity: (1) There is an obvious increase in intracellular Ca^{2+} concentration ($[\text{Ca}^{2+}]_i$) in both *in vivo* and *in vitro* models of excitotoxic cell death. This has been observed in ischemic brain and in brain slices exposed to NMDA agonists or anoxia. In addition, glutamate-stimulated Ca^{2+} influx in cultured neurons has been shown by the Ca^{45} technique, and increased $[\text{Ca}^{2+}]_i$ after NMDA stimulation has been observed repeatedly using fluorescent probes. (2) Prevention of Ca^{2+} entry into the cell by removal of extracellular Ca^{2+} depletion of NMDA or by pharmacological inhibition of glutamate receptors or voltage-dependent Ca^{2+} channels prevents neuronal death in many paradigms of excitotoxicity. (3) Prevention of neurotoxicity by inhibition of downstream effects of Ca^{2+} overload strongly suggests a causal role of Ca^{2+} in excitotoxicity. Intracellular Ca^{2+} chelators can prevent ischemic damage *in vivo* and excitotoxic neuronal damage *in vitro*. Also, inhibition of effectors of Ca^{2+} toxicity such as calmodulin, calcineurin, and bNOS protects neurons from the toxicity of excitatory amino acids. On the other hand, there could potentially be other routes for Ca^{2+} entry under excitotoxic conditions.

Using a model of hypoxia (oxygen/glucose deprivation, or OGD), a new lethal pathway that involves the activation of a cation conductance (I_{OGD}) has recently been unveiled. This leads to neuronal Ca^{2+} overload in the absence of excitotoxic stimulation. I_{OGD} requires the TRMP7 ion-channel protein, a member of the TRP (transient receptor potential) cation channel super family. Gating of TRMP7 occurs because of the generation of an excess of reactive oxygen/nitrogen species in anoxic neurons. Because gating TRMP7 allows Ca^{2+} entry into the neuron and Ca^{2+} can stimulate further radical production, a vicious loop leading to sustained Ca^{2+} overload is established in the absence of NMDA-R stimulation by excess glutamate.

In addition to excessive entry of Ca^{2+} through membrane channels, mitochondrial Ca^{2+} sequestration and subsequent release play a central role in ischemia- or glutamate-mediated cell death. Nevertheless, mitochondrial Ca^{2+} release and increased Ca^{2+} influx into neurons cannot fully account for the irreversible buildup of intracellular Ca^{2+} after excitotoxic stimulation. The bulk increase in cellular Ca^{2+} should be rectified over time, unless cellular Ca^{2+} extrusion is inhibited. Inhibition of cellular Ca^{2+} efflux from cells is sufficient to trigger cell death in non-neuronal cells, and in neurons may be brought about by oxidative damage downstream of mitochondrial dysfunction. Recent work shows that calpains can cleave the plasma membrane $\text{Na}^+/\text{Ca}^{2+}$ exchanger (NCX) in brain ischemia and in cerebellar granule neurons exposed to

glutamate. Calpain-mediated NCX proteolysis is necessary for the delayed excitotoxic Ca^{2+} deregulation leading to neuronal death (for a schematic summary of excitotoxicity, see Figure 1).

Apoptosis, Necrosis, and Other Ca^{2+} -Dependent Forms of Cell Death in Brain Ischemia

Neuronal demise and neurological dysfunction in brain ischemia are clearly due to several components. Apoptosis and necrosis, in their classical definition, are two fundamentally different modes of cell death. Whereas apoptosis is characterized by a preservation of membrane integrity until the cell is phagocytosed, this is not the case in necrosis/lysis of cells. The duration and extent of Ca^{2+} influx may determine if neurons survive, die by apoptosis, or undergo necrotic lysis. Very low $[\text{Ca}^{2+}]_i$ or the prolonged inhibition of Ca^{2+} influx may be neurotoxic. A continuous moderate increase in $[\text{Ca}^{2+}]_i$, such as that produced by a sustained slow influx, may cause apoptosis, whereas an exceedingly high influx causes rapid cell lysis. For example, stimulation of cortical neurons with high concentrations of NMDA results in necrosis, whereas exposure to low concentrations causes apoptosis. Accordingly, neuronal death in experimental stroke models is predominantly necrotic in the ischemic core, but apoptosis occurs in the less severely compromised penumbra or border regions. The same applies to several other neuropathological conditions in which apoptosis and necrosis have been observed to occur simultaneously. One sensor that switches neurons toward one of the two fates is the ability of cells to generate adenosine triphosphate (ATP). A complete de-energization of the cell (e.g., failure of all mitochondria and of glycolysis) does not allow the ordered sequence of changes required for the apoptotic demise. In such a case, other processes result in rapid, uncontrolled cell lysis/necrosis. Therefore, under conditions of Ca^{2+} overload, apoptosis ensues when sufficient energy production (ATP) is available to execute the death program. The main ATP-requiring step for the execution of apoptosis is the formation of the apoptosome protein complex between cytochrome *c* released from damaged mitochondria, the cytosolic protein APAF-1, and procaspases. Either ATP or dATP is required to promote the functional activation and assembly of this complex, which then leads to caspase activation and cell breakdown.

Ca^{2+} Signals in Neurodegenerative Disorders

Neurons rely on complex synaptic activity to function and substantial loss of synapses can lead to neurodegeneration. The initial proposal that excitotoxic mechanism may contribute to neurodegenerative disorders including Alzheimer's and Parkinson's diseases has fostered intense research on the possible role of Ca^{2+} signaling dysfunction during early stages of neurodegeneration. Recent work has highlighted defects in calcium transport and release from the endoplasmic reticulum (ER) in mouse models of Alzheimer's diseases as well as relevant Ca^{2+} signal alterations at synapses. For example, deficits in NMDARs function at synapses of young Alzheimer's disease mice have been associated with increased ryanodine receptor activity. Abeta oligomers can cause localized calcium elevation and tau missorting in dendrites followed by disruption of

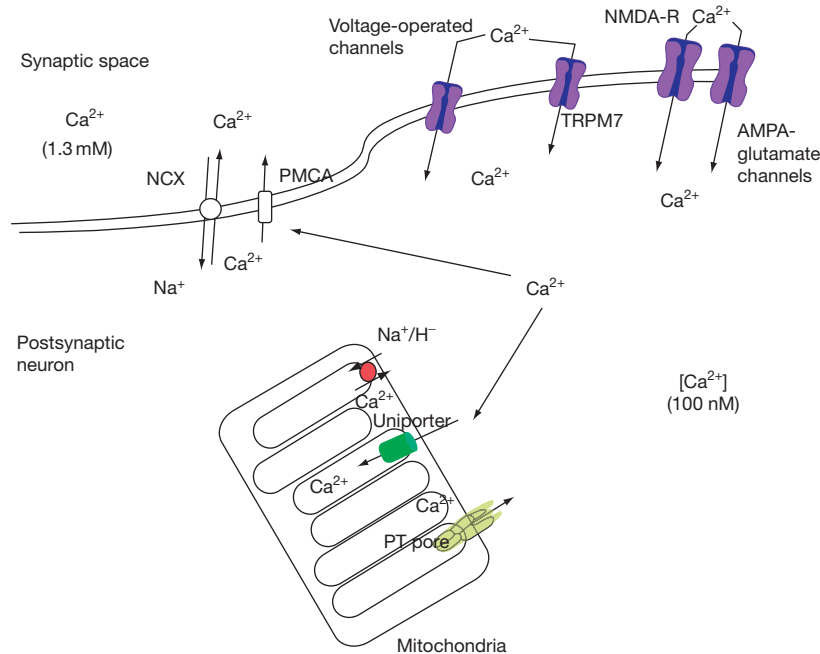


Figure 1 The term neuronal excitotoxicity defines as consequences of excessive stimulation of postsynaptic glutamate receptors. Excess glutamate in the synapse causes gating of AMPA and NMDA receptors. The former allows entry of Na^+ and Ca^{2+} into the neuron. Na^+ hyperpolarizes the plasma membrane and removes the Mg^{2+} block from the NMDA receptor (the NMDA receptor is normally 'plugged' by Mg^{2+}). The NMDA receptor is then gated to Ca^{2+} that flows into the neuron. Additional routes for Ca^{2+} entry involve the TRPM7 cation channel and voltage-operated Ca^{2+} channels. Ca^{2+} accumulates in mitochondria as amorphous calcium phosphate. When the buffering capacity of mitochondria is overwhelmed, Ca^{2+} is released in a process known as permeability transition (PT). Efflux of Ca^{2+} from the neuron is operated by the NCX and the PMCA. When mitochondria release their Ca^{2+} and the efflux systems are unable to extrude the excess intracellular Ca^{2+} , Ca^{2+} overload becomes irreversible and causes neuronal demise.

microtubules and spines. It is apparent that synaptic dysfunction either caused by alterations in Ca^{2+} signaling or leading to alterations in Ca^{2+} transport can enhance synaptic pruning and lead to a loss of plasticity. Both these events feature prominently in age-related neurodegenerative disorders including dementia and motor disorders.

Ca^{2+} Signals as Subroutines of the Apoptotic Program

The role of Ca^{2+} in apoptosis signaling was initially suggested by studies in thymocytes and lymphocytes showing that sustained Ca^{2+} increases could trigger the DNA fragmentation typical of apoptosis and all the other features of the apoptotic demise. Ca^{2+} may be central in the unfolding of the apoptotic program at different stages. Ca^{2+} -mediated mitochondrial permeability transition is recognized as one important mechanism for cytochrome *c* release and caspase activation. In addition, ER-mitochondrial Ca^{2+} fluxes are modulated by the Bcl-2 protein family, and caspase-mediated cleavage of Ca^{2+} transporters then results in further disruption of ER Ca^{2+} handling, with subsequent mitochondrial Ca^{2+} overload and permeability transition. Ample evidence has documented cross-talk among calpains, caspases, and other protease families, which alters the downstream effects of these families on cellular Ca^{2+} fluxes. Finally, Ca^{2+} -regulated processes are also involved in the ultimate fate of dying cells, their clearing by phagocytes due to Ca^{2+} -dependent exposure of surface recognition molecules and secondary lysis. The latter can be brought about by caspase-dependent cleavage of the plasma

membrane Ca^{2+} ATPase (PMCA), which results in a secondary Ca^{2+} overload and the activation of Ca^{2+} -dependent mechanisms causing cell lysis.

See also: **Bioenergetics:** Plasma-Membrane Calcium Pump: Structure and Function; **Cell Architecture and Function:** Autophagy in Fungi and Mammals; Bcl2 Family: Cell Death Enhancers and Inhibitors; Caspases and Cell Death; **Protein/Enzyme Structure Function and Degradation:** Calpain.

Further Reading

- Aarts M, Iihara K, Wei W-L, et al. (2003) A key role for TRPM7 channels in anoxic neuronal death. *Cell* 115: 863–877.
- Ankarcrona MDJ, Bonfoco E, Zhivotovsky B, et al. (1995) Glutamate-induced neuronal death: A succession of necrosis or apoptosis depending on mitochondrial function. *Neuron* 14: 961–973.
- Bano D, Young KW, Guerin CJ, et al. (2005) Calpain cleavage of the plasma membrane $\text{Na}^+/\text{Ca}^{2+}$ exchanger in excitotoxicity. *Cell* 120: 275–285.
- Han BH, Xu D, Choi J, et al. (2002) Selective, reversible caspase-3 inhibitor is neuroprotective and reveals distinct pathways of cell death after neonatal hypoxic-ischemic brain injury. *Journal of Biological Chemistry* 277: 30128–30136.
- Lee JM, Zipfel GJ, and Choi DW (1999) The changing landscape of ischaemic brain injury mechanisms. *Nature* 399: A7–14.
- Lipton SA (1996) Similarity of neuronal cell injury and death in AIDS dementia and focal cerebral ischemia: Potential treatment with NMDA open-channel blockers and nitric oxide-related species. *Brain Pathology* 6: 507–517.
- Orrenius S, Zhivotovsky B, and Nicotera P (2003) Regulation of cell death: The calcium apoptosis link. *Nature Reviews Molecular Cell Biology* 4: 552–565.
- Scorrano L, Oakes SA, Opferman JT, et al. (2003) BAX and BAK regulation of endoplasmic reticulum Ca^{2+} : A control point for apoptosis. *Science* 300: 135–139.

Cell–Matrix Interactions

J A Askari and M J Humphries, University of Manchester, Manchester, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Extracellular matrix (ECM) A network of proteins and polysaccharides underlying and surrounding cells.

Focal adhesions Points of anchorage of the cell to the underlying matrix allowing communication between the inside and outside of the cell via integrins.

Inside-out signaling The modulation of integrin–ligand binding activity by intracellular events, leading to integrin clustering and conformational changes.

Integrins A family of $\alpha\beta$ heterodimeric cell surface receptors for extracellular matrix proteins.

Outside-in signaling The modulation of cell phenotype or behavior by extracellular events such as ligand binding and application of force.

Syndecans A family of transmembrane glycoprotein cell surface receptors for extracellular matrix proteins that can synergize with integrins.

The extracellular matrix (ECM) is a network of macromolecules that underlies all epithelia and endothelia and that surrounds all connective tissue cells. The ECM provides mechanical support and also profoundly influences the behavior and the differentiation state of cells in contact with it. The main receptors mediating the interaction of cells with ECM proteins are known as integrins, referring to their function of integrating the cell's exterior with its interior. Following ligand binding, the cytoplasmic domains of integrins connect to the cytoskeleton and trigger the assembly of signaling complexes. Conversely, the binding of intracellular cytoplasmic components influences cell adhesiveness by altering integrin conformation. Thus, a large variety of complex signaling events can be transduced by integrins in a bidirectional manner across the cell membrane. These events serve to modulate and coordinate many aspects of cell behavior, such as proliferation, survival, shape, polarity, motility, gene expression, and differentiation that are required for fundamental processes such as development, tissue morphogenesis, and wound healing within multicellular organisms. Integrins are also implicated in several disease processes such as inflammation, thrombosis, and cancer metastasis, whereas mutations in integrin genes lead to deficiencies in leukocyte adhesion, myopathy, and blistering skin diseases.

The Extracellular Matrix

Main Components

The ECM components are diverse in composition, but they generally comprise a mixture of fibrillar proteins, polysaccharides, and glycoproteins synthesized, secreted, and organized by neighboring cells. Collagens, fibronectin (FN), and laminins are the principal components involved in cell–matrix interactions; other components, such as vitronectin, thrombospondin, and osteopontin, although less abundant, are also important adhesive molecules.

Over 20 different collagens exist in mammals, but the most abundant exist as helical molecules that assemble into fibrils

and provide tensile strength for many tissues. The ECM also contains a large number of glycoproteins, the best studied of which is FN, a dimeric molecule that also forms insoluble fibrils. The laminins are heterotrimeric glycoproteins that fold into cruciform-shaped molecules and are important components of basement membranes. Fibrinogen and von Willebrand factor are also considered to be matrix proteins because they function as major adhesive molecules in blood, but they can also pass into extravascular fluid and influence cell function.

Adhesive Motifs

Although matrix proteins are large molecules, their major integrin recognition sites are very short peptide motifs of only three to six amino acids. The best known and most widespread of these adhesive motifs is arginine–glycine–aspartic acid (RGD, using the single-letter amino acid nomenclature). This motif was first discovered in 1984 by Pierschbacher and Ruoslahti in the center of FN, but it is also present and functional in vitronectin, von Willebrand factor, and thrombospondin. The crystal structure of the area of FN containing this motif has been solved and shows the tripeptide extending out as a loop from the surface of the protein. Similar short peptide motifs are present in other matrix proteins (see [Table 1](#)). A constant feature of the motifs is the presence of an acidic residue, either aspartic or glutamic acid, which is an absolute requirement for the adhesive activity of the proteins in which they reside. As yet, no corresponding motif has been determined in laminin; however, integrin binding has been localized to distinct regions within the molecule.

In FN, a second region of the molecule, termed the synergy site, has been shown to work in concert with the RGD motif and to enhance integrin-binding affinity. Part of the synergy site is a pentapeptide proline–histidine–serine–arginine–asparagine (PHSRN). Although they probably exist, no synergy sequences have as yet been defined for other matrix proteins.

Table 1 Adhesive sequences in matrix proteins and their integrin receptors^a

Matrix protein	Adhesive sequence	Integrin receptor
Collagens	GFOGER	$\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 10\beta 1$, $\alpha 11\beta 1$
Fibronectin	RGD	$\alpha 5\beta 1$, $\alpha V\beta 3$, $\alpha 8\beta 1$, $\alpha V\beta 1$, $\alpha V\beta 6$, $\alpha 11b\beta 3$
	LDV	$\alpha 4\beta 1$, $\alpha 4\beta 7$
	REDV	$\alpha 4\beta 1$
Laminins	E1' fragment	$\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 10\beta 1$
	E8 fragment	$\alpha 3\beta 1$, $\alpha 6\beta 1$, $\alpha 7\beta 1$, $\alpha 6\beta 4$
Vitronectin	RGD	$\alpha V\beta 3$, $\alpha 11b\beta 3$, $\alpha V\beta 5$, $\alpha V\beta 1$, $\alpha V\beta 8$
Fibrinogen	RGD	$\alpha V\beta 3$
	KQAGDV	$\alpha 11b\beta 3$
von Willebrand factor	RGD	$\alpha 11b\beta 3$, $\alpha V\beta 3$

^aGFOGER, glycine–phenylalanine–hydroxyproline–glycine–glutamic acid–arginine; KQAGDV, lysine–glutamine–alanine–glycine–aspartic acid–valine; LDV, leucine–aspartic acid–valine; REDV, arginine–glutamic acid–aspartic acid–valine; RGD, arginine–glycine–aspartic acid.

Integrins

Structure

Integrins are noncovalently linked dimers consisting of an α - and a β -subunit, and the 18 α - and 8 β -subunits that have been identified in humans combine to form 24 different receptors. Integrin homologs are present in organisms ranging from sponges to humans, indicating their central role in metazoan evolution. Integrins possess large extracellular domains, approximately 1200 amino acids in α -subunits and 800 in β -subunits, a transmembrane domain, and short cytoplasmic regions of 50 residues or less. The exception to this rule is the $\beta 4$ -subunit, which has a cytoplasmic domain of over 1000 amino acid residues. In addition, half of all α -subunits contain an extra 200 residue module toward their N-terminus (see Table 2), which has homology with the von Willebrand factor A-domain and is referred to as the α A- or I- (for inserted) domain. Several α A-domains have been crystallized and all adopt a Rossmann fold characterized by central β -sheets surrounded by α -helices. A conserved motif, the metal-ion-dependent adhesion site (MIDAS), which includes the sequence aspartic acid–any residue–serine–any residue–serine (DxSxS), coordinates a divalent metal cation at the top of the domain. The N-terminus of the β -subunit also contains an A-domain with a MIDAS motif. Electron micrographs show the integrin dimer to have a globular head 8–12 nm in diameter that constitutes the ligand-binding domain from which two stalks 2-nm thick and 14–20 nm in length project; the distal regions of these contain hydrophobic sequences representing the transmembrane domains.

Those α -subunits that do not contain A-domains undergo posttranslational cleavage into a light and heavy chain held together by a disulfide bond, the exception being $\alpha 4$, which is cleaved at a more central position to yield fragments of 70 and 80 kDa. In addition, several integrin subunits, $\alpha 3$, $\alpha 6$, $\alpha 7$, $\beta 1$, $\beta 3$, $\beta 4$, and $\beta 5$, are subject to alternative splicing mainly in their cytoplasmic domains, yielding isoforms that are differentially expressed in specific patterns. This splicing serves to

Table 2 Properties of integrins^a

Integrin	Alternative names	Matrix ligands
<i>With an A-domain</i>		
$\alpha 1\beta 1$	VLA 1, CD49a/CD29	CO, LM
$\alpha 2\beta 1$	VLA 2, GPIIb/IIIa, CD49b/CD29	CO, LM
$\alpha 10\beta 1$		CO, LM
$\alpha 11\beta 1$		CO
$\alpha D\beta 2$	CD11d/CD18	b
$\alpha L\beta 2$	LFA-1, CD11a/CD18	b
$\alpha M\beta 2$	Mac1, CD11b/CD18	FG
$\alpha X\beta 2$	p150,95, CD11c/CD18	FG
$\alpha E\beta 7$	$\alpha_{IEL}\beta 7$, HML-1 antigen, CD103(αE)	b
<i>Without an A-domain</i>		
$\alpha 3\beta 1$	VLA 3, CD49c/CD29	LM
$\alpha 4\beta 1$	VLA 4, CD49d/CD29	FN, OP
$\alpha 5\beta 1$	VLA 5, FNR, CD49e/CD29	FN, OP
$\alpha 6\beta 1$	VLA 6, CD49f/CD29	LM
$\alpha 7\beta 1$		LM
$\alpha 8\beta 1$		FN, TN, NN
$\alpha 9\beta 1$		TN, OP
$\alpha V\beta 1$	CD51/CD29	FN, VN
$\alpha 11b\beta 3$	GPIIb/IIIa, CD41/CD61	FN, FG, VN, vWF, Tsp
$\alpha V\beta 3$	VNR, CD51/CD61	VN, FG, FN, vWF, Tsp
$\alpha V\beta 5$		VN
$\alpha V\beta 6$		FN
$\alpha V\beta 8$		VN
$\alpha 6\beta 4$	CD49f/CD104	LM
$\alpha 4\beta 7$	LPAM-1	FN

^aCO, collagen; FG, fibrinogen; FN, fibronectin; LM, laminin; NN, nephronectin; OP, osteopontin; TN, tenascin; Tsp, thrombospondin; VN, vitronectin; vWF, von Willebrand factor.

increase the cell–matrix interaction repertoire and hone the temporal and spatial adhesive responses of cells.

The field of integrin research was advanced enormously by the publication of the crystal structure of the extracellular domain of $\alpha V\beta 3$ in 2001 by a group led by Arnaout. This structure shows that the globular head of the integrin is made up of a seven-bladed β -propeller contributed by the α -subunit and the A-domain from the β -subunit, together with an immunoglobulin fold, termed the hybrid domain, made up of polypeptide sequences from either side of the β A-domain. The integrin stalks are folded into three β -sandwich domains in the α -subunit and four epidermal growth factor like repeats in the β -subunit. The β -subunit also has an N-terminal plexin–semaphorin–integrin domain and a novel cystatin-like fold just before the transmembrane region. An arginine residue, R261, in the $\beta 3$ -subunit A-domain extends into the core of the α -subunit propeller, where it is held in place by aromatic residues, and is the main area of subunit association.

Ligand Binding

All interactions between integrins and matrix proteins are dependent on divalent cations, whereas the ligand-binding specificity of integrin molecules is determined by the particular α – β combination. Some matrix proteins (e.g., FN) can bind to several integrins, or, conversely, one integrin can recognize ligands of diverse structure; for example, $\alpha 2\beta 1$ binds to both

laminin and collagen. In general, each integrin has a specific nonredundant function, which is emphasized by the distinct phenotypes of knockout mice.

The integrin dimers can be broadly divided into three families consisting of the $\beta 1$, $\beta 2/\beta 7$, and $\beta 3/\alpha V$ integrins. $\beta 1$ associates with 12 α -subunits and can be further divided into RGD-, collagen-, or laminin-binding and the related $\alpha 4/\alpha 9$ integrins that recognize both matrix and vascular ligands. $\beta 2/\beta 7$ integrins are restricted to leukocytes and mediate cell–cell rather than cell–matrix interactions, although some recognize fibrinogen. The $\beta 3/\alpha V$ family members are all RGD receptors and comprise $\alpha IIb\beta 3$, an important receptor on platelets, and the remaining β -subunits, all of which associate with αV . It is the collagen receptors and the leukocyte-specific integrins that contain α A-domains.

In non- α A-domain-containing integrins, both the α - and β -subunits are required for ligand binding. An integrin–ligand complex crystal structure showed that RGD peptide binds at the α – β interface, with its arginine residue contacting the α -subunit propeller and its aspartate residue helping to coordinate the divalent cation at the MIDAS site via a carboxyl linkage, which, in the presence of ligand, is occupied. Two further cations are present in the ligand-bound structure, one at a site adjacent to the MIDAS (termed ADMIDAS) and the other at a site in close proximity, termed the ligand-induced metal-binding site. Isolated recombinant α A-domains are able to bind peptide and macromolecular ligands with the same affinity as intact dimer. A crystal structure of the A-domain of $\alpha 2$ in complex with a triple-helical collagenous peptide containing the glycine–phenylalanine–hydroxyproline–glycine–glutamic acid–arginine (GFOGER) motif also shows the carboxyl from an acidic residue, in this case glutamate, from the ligand directly completing the coordination sphere of the metal ion in the MIDAS. This explains the absolute requirement for either aspartic or glutamic acid in matrix adhesive proteins and the dependence on divalent cations for integrin–ligand interactions.

Cation Modulation

The binding of ligand to integrin cannot take place without a divalent cation, which directly contacts a carboxyl group from the ligand. However, cations also play an important role in the regulation of integrin affinity. Integrin $\alpha V\beta 3$ contains six cation-binding sites in the unliganded crystal structure and eight in the presence of cyclic RGD peptide, at least three of which are affected by ligand binding. In general, integrin–ligand binding is stimulated by magnesium and manganese ions and inhibited by calcium ions. However, this situation is complicated by the fact that the binding of one cation can affect the binding of another at a different site. In addition, several cation-binding sites within an individual integrin can influence ligand binding. The physiological relevance of cation modulation of integrin affinity has yet to be resolved.

Activation and Signaling

Focal Adhesions

When cells attach and spread on a matrix ligand, there is an initial clustering of integrins in the membrane, followed by an

accumulation of cytoskeletal and signaling molecules into dynamic structures known as focal adhesions or focal contacts. These structures represent the anchor points of the cell where integrins link the cell to both the underlying matrix and the intracellular actin filaments of the cytoskeleton. The contacts appear as dense plaques, often located near the cell periphery, and provide both a scaffold and a means whereby cells can generate traction during migration. More important, focal adhesions also serve as nucleation points for the recruitment of not only structural proteins such as vinculin, talin, and paxillin, but also many signaling and adaptor molecules able to trigger a cascade of phosphorylation events that can activate numerous downstream targets.

Conformational Changes

Many integrins are not constitutively active, and adhesion of cells to matrix proteins needs to be strictly controlled and regulated in response to environmental changes. Integrins exist in at least three states: inactive, active, and ligand bound; the switching of integrins from an inactive to an active state involves conformational changes not only in the ligand-binding pocket, but also across the whole of the extracellular domain and in the cytoplasmic face. Thus, the activation of the ligand-binding domain in the head and the binding of ligand are coupled via long-range conformational changes to signaling events in the cytoplasm. Conversely, intracellular events effect changes in the cytoplasmic domains that are translated in the opposite direction to the integrin head, allowing activation and ligand binding. This bidirectional communication is termed outside-in and inside-out signaling. There is increasing evidence that these signaling responses are additionally modulated by the application of forces on integrins in adherent cells generated by linkages with the ECM and the cytoskeleton.

Cytoplasmic Domains

The short cytoplasmic domains of integrins play a vital role in integrin function because they control the activation states of integrins and are required to maintain integrins in an inactive state. They are also the sites of interaction with, and linkage to, both cytoskeletal and signaling molecules when clustered in focal adhesions. Interactions between the α and β tails occur by a salt bridge and hydrophobic and electrostatic contacts in the α -helical, membrane-proximal region of both domains, and it is likely that this association is lost on integrin activation, allowing the tails to separate and effector molecules to bind. The β -subunit tail is the principal site for binding of cytoplasmic molecules, whereas the α -subunit plays a more regulatory role in controlling activation.

Many proteins are intimately associated with integrins in focal adhesions, but direct interaction with integrin $\beta 1$ and $\beta 3$ cytoplasmic domains has only been proven for the binding of the structural protein talin and the kindlin family of proteins. The sites of interaction are conserved motifs (a membrane-proximal asparagine–proline–any residue–tyrosine (NPXY) for talin and a membrane-distal NxXY for kindlin) in the integrin β tail with a phosphotyrosine-binding-like (PTB) subdomain present in the other molecule. This interaction may be the prototype for other protein interactions with integrin β tails through

PTB domains. While the binding of talin is established as the essential, final common step in integrin activation, it is clear that the kindlins play a major role in regulating this process.

Signaling by Integrins

Integrin-mediated signaling can be broadly divided into two categories, direct and collaborative. In the first, ligation and clustering of integrins are the only stimuli and adhesion to ECM proteins activates cytoplasmic tyrosine kinases, for example, focal adhesion kinase and serine/threonine kinases such as those in the mitogen-activated protein kinase cascade. Direct signaling also induces ionic transients (e.g., Ca^{2+} and Na^+/H^+) and stimulates lipid metabolism. In collaborative signaling, integrin–ECM adhesion modulates signaling initiated by other types of receptors, such as receptor tyrosine kinases and G-protein-coupled receptors, allowing cells to integrate positional information concerning matrix contacts with information about the availability of soluble growth or differentiation factors that are also located in the ECM. An important example of collaborative signaling has been demonstrated between integrins and syndecans, heavily glycosylated proteins that are receptors for ECM proteins as well as soluble growth factors. Most ECM molecules contain binding sites for both integrins and syndecans and there is substantial evidence that a full cell-adhesion response requires engagement of both types of receptor.

Cells that are deprived of anchorage to the ECM eventually die by a specialized form of programmed cell death (apoptosis) known as anoikis. Cell–matrix interactions via

integrins are thus essential for cell survival. Integrins also directly affect the organization of the cytoskeleton, and consequently cell motility, by activating the Rho guanosine triphosphatases (GTPases), a branch of the Ras GTPase superfamily, particularly CDC42, Rac1, and RhoA. Rho promotes the formation and maintenance of actin stress fibers, whereas Rac and CDC42 regulate structures such as lamellipodia and filopodia, respectively.

It is worth remembering that each signaling pathway mentioned influences others at some point, so signaling within a cell is better considered as a series of networks rather than a direct path. Thus, integrins are able to modulate, either directly or indirectly, every signaling pathway within the cell, a fact that emphasizes the importance of cell–matrix interactions on all aspects of cell behavior.

See also: **Cell Architecture and Function:** Focal Adhesions and Related Integrin Contacts; Rho GTPases and Actin Cytoskeleton Dynamics; **Signaling:** Integrin Signaling.

Further Reading

- Askari JA, Buckley PA, Mould AP, and Humphries MJ (2009) Linking integrin conformation to function. *Journal of Cell Science* 122: 165–170.
- Humphries JD, Byron A, and Humphries MJ (2006) Integrin ligands at a glance. *Journal of Cell Science* 119: 3901–3903.
- Hynes RO (2002) Integrins: Bidirectional, allosteric signaling machines. *Cell* 110: 673–687.
- Xiong J-P, Stehle T, Diefenbach B, et al. (2001) Crystal structure of the extracellular segment of integrin $\alpha V\beta 3$. *Science* 294: 339–345.

Cell Migration

J V Small, IMBA-Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Vienna, Austria

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by J. Victor Small and Emmanuel Vignal, volume 1, pp. 356–361, © 2004, Elsevier Inc.

Glossary

Actin filament A filament of ~8 nm in diameter, made up from two helical strands of actin monomers.

Filopodia Bundles of unipolar actin filaments, ~0.2 µm in diameter, that protrude from the cell edge, normally in association with lamellipodia and ruffles. Filopodia and lamellipodia are interconvertible assemblies of actin filaments. Bundles resembling filopodia that do not extend beyond the cell edge are sometimes referred to as 'microspikes'.

Focal adhesions Adhesion foci, longer-lived than focal complexes that link to contractile actin bundles (stress fibers) in the actin cytoskeleton. Both focal complexes and focal adhesions harbor more than 50 proteins, involved in structural and signaling activities.

Focal complex A site of early adhesion to the extracellular matrix, formed beneath a lamellipodium or a filopodium.

Focal complexes either can form and dissolve, within 1–2 min, or can mature into focal adhesions.

Integrins A family of transmembrane receptor molecules that cluster at adhesion foci to link the matrix on the outside to the actin cytoskeleton on the inside.

Lamellipodia Thin, membrane-bound leaflets of cytoplasm, 0.2–0.3 µm thick and up to several microns wide that are protruded at the cell edge, close to and parallel to the substrate. They are composed of unipolar networks of actin filaments.

Rho (Ras homology) proteins A subfamily of small guanosine triphosphatases whose roles include signaling to the actin cytoskeleton.

Ruffles Manifestations of lamellipodia, often including filopodia, protruding upward from the dorsal cell surface, and migrating generally rearwards over it. Ruffles can also occur in dorsal circular arrays, especially in malignant cells.

The morphogenesis of multicellular organisms involves the extensive migration of cells from primordial sites to locations destined for specific tissue development. Among the most dramatic are the movements of cells of the neural crest, which travel from the neural tube to distant sites where they differentiate into diverse cell lineages. The fusion of epithelial layers, such as those which occur during closure of the neural tube, entails coordinated cell migration. In the adult, wound closure and tissue repair depend on the migration of surrounding cells to effect the regeneration process. In defense against foreign organisms, cells of the immune system are mobilized and migrate to the sites of inflammation to engage with the enemy. In another context, cell migration contributes to the dissemination of malignant cells in the spread of cancer. Understanding cell migration has therefore much to do with life and death, which concerns all.

Metazoan cells migrate by a crawling mechanism that entails continuous changes in shape. Crawling, in turn, requires traction and this is provided by the development of transient points of anchorage with the connective tissue scaffold and with other cells. In both processes, shape change and traction, the intracellular polymer framework of the cell, the so-called cytoskeleton, plays a central role. Changes in the cytoskeleton framework are influenced, in turn, by the chemical and mechanical properties of the surrounding matrix, so there is an active crosstalk between the two. Most of what is known about the mechanisms of cell migration comes from studies of cells moving on planar surfaces *in vitro* and from investigations of the test tube properties of cytoskeleton polymers and their associated proteins. However, parallel studies indicate that the basic principles derived from these approaches apply also to cell movement in a tissue environment.

The Cytoskeleton

The cytoskeleton is composed of three distinct, but interlinked networks of polymers, composed of actin, tubulin, and proteins of the intermediate filament family, together with many associated proteins. The term 'cytoskeleton' is a misnomer, because the cytoplasmic filament networks that make it up are in a state of continuous turnover and rearrangement. Cell migration is driven primarily by the programmed turnover of the actin cytoskeleton, but microtubules exert an important influence on polarization and guidance, which is required for directional motility. Intermediate filaments do not appear to play an important role in cell motility, but a subtle modulatory role cannot yet be excluded.

The Actin Cytoskeleton

The actin cytoskeleton is composed of actin filaments organized in networks and bundles: the salient features of the actin cytoskeleton in a fibroblast are illustrated in [Figures 1](#) and [2](#). The cell periphery is delimited either by bundles of actin filaments, running parallel to the cell edge or by dense actin meshworks that are commonly punctuated by small, radial actin bundles, termed filopodia or microspikes (Fp, Ms; [Figures 1](#) and [2](#)). The actin meshworks (Lp, [Figure 3](#)) are the structural component of the sheet-like protrusions at the cell periphery called 'lamellipodia' (or ruffles, when they fold upwards). The body of the cell is pervaded by a loose network of actin filaments, a proportion of which are organized into prominent bundles, called stress fibers (Sf, [Figures 1](#) and [2](#)).

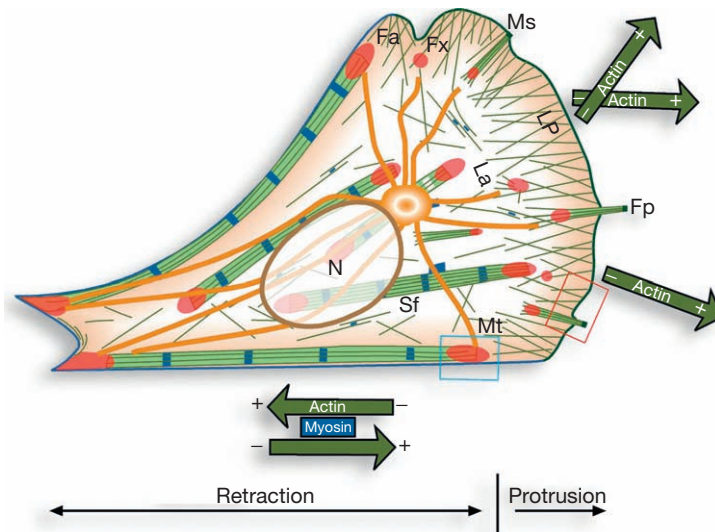


Figure 1 General features of the actin and microtubule cytoskeletons of a migrating cell. Actin filaments are in green, microtubules (Mt) in orange. Protrusion of the cell is driven by the polymerization of actin filaments organized in sheet-like meshworks (lamellipodia, Lp) and bundles (filopodia, Fp; microspikes, Ms). Early adhesions formed beneath protrusions as focal complexes (Fx) can mature into larger, focal adhesions (Fa) connecting the ends of actin bundles, called stress fibers (Sf) via transmembrane receptors to the matrix. Actin filaments in protruding regions are similarly polarized, with their fast-growing +ends directed forwards. Immediately behind the protruding front, in the lamella (La) and the rest of the cell, actin filaments interact in anti-parallel arrays with myosin filaments (blue) to form contractile assemblies, which function in retraction. The polarization of the cell is reflected in a polarization of the adhesion pattern, with stationary focal complexes and focal adhesions at the front and sliding focal adhesions at the rear. Microtubules emanating from the centrosome interact with adhesions and influence their turnover. N, nucleus. Details of the boxed areas are provided in [Figure 4](#). Modified from Small JV and Rottner K (2010) Elementary cellular processes driven by actin assembly: Lamellipodia and filopodia. In: Carlier M-F (ed.) *Actin Based Motility: Cellular, Molecular and Physical Aspects*. Springer Science.

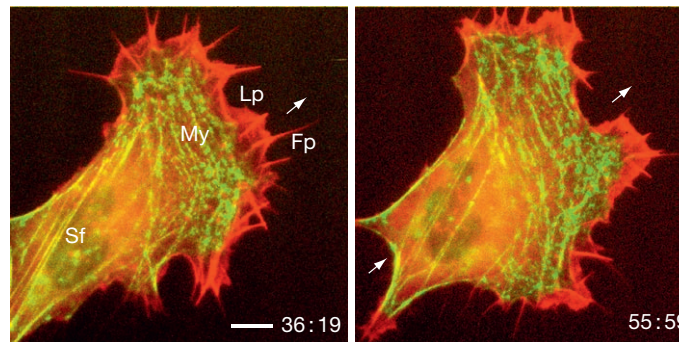


Figure 2 Sequential video frames of a migrating fibroblast, taken at the times indicated (min/s), illustrating protrusion at the front followed by retraction at the rear. The cell was transfected with mCherry-actin (red) and GFP-myosin light chain (My, green). The protruding cell front comprises lamellipodia (Lp) and filopodia (Fp). Myosin forms contractile bundles with actin (stress fibers, Sf) in the cell body. Figure supplied by Maria Nemethova. Bar=10 μ m. Modified from Small JV and Rottner K (2010) Elementary cellular processes driven by actin assembly: Lamellipodia and filopodia. In: Carlier M-F (ed.) *Actin Based Motility: Cellular, Molecular and Physical Aspects*. Springer Science.

that contain also myosin (My, [Figure 2](#)) to confer contractility. According to cell type, the ratio of peripheral actin meshworks to stress fibers differs and, as a general rule, cells that migrate faster have fewer stress fiber bundles in the body of the cell. The different organizations of actin filaments that make up the subcompartments of the actin cytoskeleton are signaled via pathways involving different members of the Rho family of small guanosine triphosphatases (GTPases): Rho for stress fibers, Rac for lamellipodia, and Cdc42 for filopodia.

One important feature of the actin cytoskeleton is its linkage to the cell membrane, which explains its pivotal influence

on cell form. Major sites of linkage occur at the termini of the stress fiber bundles, which in turn correspond to sites of adhesion of the cell to the extracellular matrix. Because of their focal nature, these sites are called focal adhesions (Fa, [Figure 1](#)). They comprise more than 50 structural, adaptor, and signaling proteins that form and regulate the linkage of the actin cytoskeleton, via transmembrane receptors of the integrin family, to the extracellular matrix (see also [Figure 4\(c\)](#)). The precursors of these anchorage sites are assembled in association with lamellipodia and filopodia at the cell periphery and are called focal complexes (Fx, [Figure 1](#)).

Phases of Movement

In general terms, the mode of translocation of a cell can be divided into three phases: (1) protrusion of a cell front; (2) the development of adhesion to the substrate for the purpose of traction; and (3) retraction of the cell rear.

Protrusion

In the first step of movement, the cell protrudes lamellipodia and filopodia: sheet-like and rod-like processes $\sim 0.2 \mu\text{m}$ thick and several microns in length (Figures 3, 4(a) and 4(b)). The rate of protrusion varies in the range of $\sim 2\text{--}15 \mu\text{m min}^{-1}$. Protrusion is

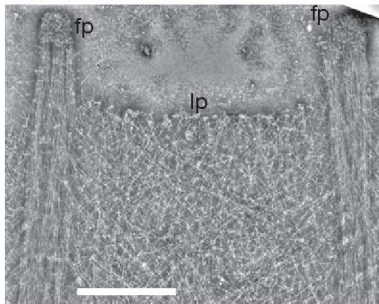


Figure 3 Electron micrograph of the protruding front of a fibroblast showing the actin filament network in a lamellipodium (Lp) and bundled filaments in filopodia (Fp). Image supplied by Stefan Koestler. Bar = $0.5 \mu\text{m}$. From Small JV and Rottner K (2010) Elementary cellular processes driven by actin assembly: Lamellipodia and filopodia. In: Carrier M-F (ed.) *Actin Based Motility: Cellular, Molecular and Physical Aspects*. Springer Science.

effected by the local nucleation and unidirectional polymerization of actin filaments on the cell membrane. Complexes of proteins including actin nucleation promoting factors and elongation factors are recruited at these membrane sites to signal and drive actin polymerization, downstream from Rac and Cdc42. A subset of these protein complexes collaborates in the formation of actin networks to form lamellipodia and another induces actin bundles to form filopodia (Figures 4(a) and 4(b)). Filaments from lamellipodia can be recruited to form filopodia so there is a mutual interplay, not yet understood, between the different nucleation and elongation factors.

Additional proteins are responsible for cross-linking the formed actin filaments into meshworks (lamellipodia) and bundles (filopodia) and for disassembling the majority of actin filaments at the base of lamellipodia and filopodia to replenish the actin monomer pool (Figure 4). Protrusion therefore involves a regulated treadmilling of actin monomers from the front to the rear of lamellipodia and filopodia. This treadmilling can be experimentally visualized as a retrograde flow by following the recovery of fluorescence after photobleaching of cells expressing actin-green fluorescent protein (GFP) (Figure 5).

Adhesion

In addition to their role in protrusion, lamellipodia and filopodia initiate adhesion to the extracellular matrix. This involves the recognition of matrix ligands on the outside and the accumulation of integrins and proteins of the adhesion machinery to form specific focal points, the focal complexes (Figure 1). The accumulation of proteins in focal complexes is likely linked to the cycling of proteins through lamellipodia and filopodia by retrograde flow.

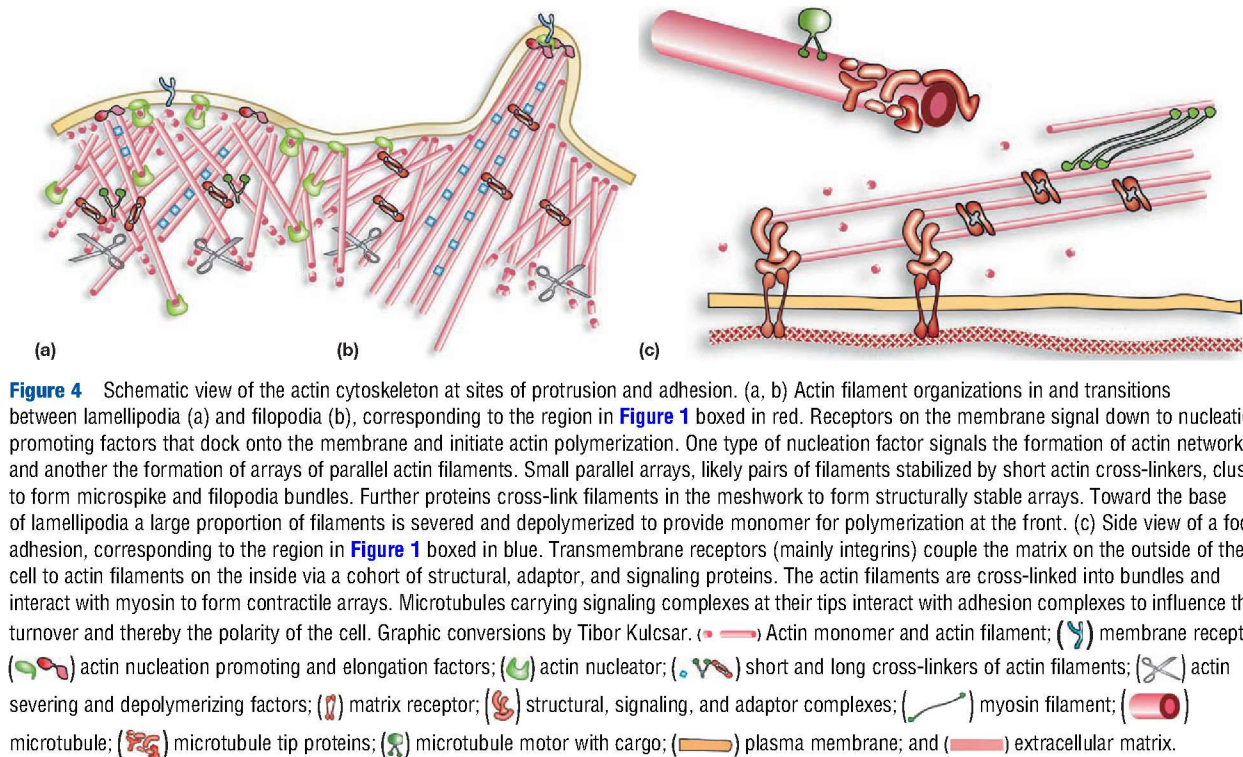


Figure 4 Schematic view of the actin cytoskeleton at sites of protrusion and adhesion. (a, b) Actin filament organizations in and transitions between lamellipodia (a) and filopodia (b), corresponding to the region in Figure 1 boxed in red. Receptors on the membrane signal down to nucleation promoting factors that dock onto the membrane and initiate actin polymerization. One type of nucleation factor signals the formation of actin networks and another the formation of arrays of parallel actin filaments. Small parallel arrays, likely pairs of filaments stabilized by short actin cross-linkers, cluster to form microspike and filopodia bundles. Further proteins cross-link filaments in the meshwork to form structurally stable arrays. Toward the base of lamellipodia a large proportion of filaments is severed and depolymerized to provide monomer for polymerization at the front. (c) Side view of a focal adhesion, corresponding to the region in Figure 1 boxed in blue. Transmembrane receptors (mainly integrins) couple the matrix on the outside of the cell to actin filaments on the inside via a cohort of structural, adaptor, and signaling proteins. The actin filaments are cross-linked into bundles and interact with myosin to form contractile arrays. Microtubules carrying signaling complexes at their tips interact with adhesion complexes to influence their turnover and thereby the polarity of the cell. Graphic conversions by Tibor Kulcsar. (●) Actin monomer and actin filament; (Y) membrane receptor; (●●) actin nucleation promoting and elongation factors; (●) actin nucleator; (●●) short and long cross-linkers of actin filaments; (✂) actin severing and depolymerizing factors; (■) matrix receptor; (■) structural, signaling, and adaptor complexes; (—) myosin filament; (■) microtubule; (■) microtubule tip proteins; (■) microtubule motor with cargo; (■) plasma membrane; and (■) extracellular matrix.

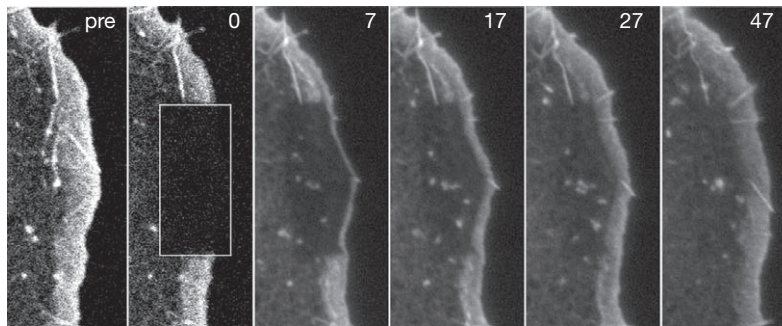


Figure 5 Actin filaments push by incorporating actin monomers between filament ends and the leading membrane of lamellipodia giving rise to a retrograde flow. Images show the lamellipodium of a B16 melanoma cell that was expressing GFP-actin. At time '0' the fluorescence in the boxed region was bleached by a laser beam. Fluorescence recovered from the front by the incorporation into filaments of fluorescent actin monomers that diffused from unbleached regions of the cell. Pre, video frame before photobleach. Numbers at top right indicate time in sec. Bar=3 μ m. From Lai FPL, Szczodrak M, Block J, et al. (2008) Molecular dissection of lamellipodium protrusion by FRAP. Dissection of actin network turnover in lamellipodia. *EMBO Journal* 27: 982–992.

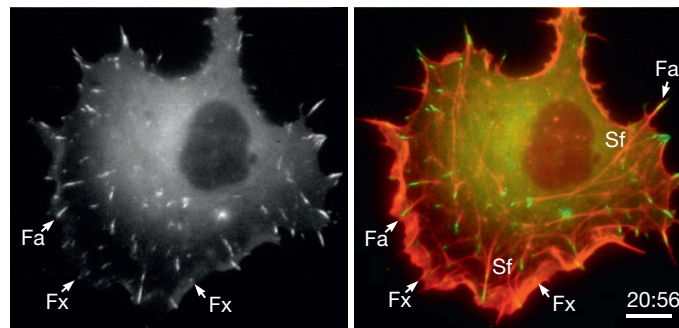


Figure 6 Adhesion sites in a migrating fibroblast. Images correspond to video frames of a living fibroblast expressing mCherry-actin (red) and GFP-paxillin (a component of adhesion sites; green) moving on fibronectin-coated glass. The left panel shows the image for GFP-paxillin alone. Fx, focal complexes; early adhesion sites created under filopodia and lamellipodia. Fa, focal adhesions; adhesions which developed from focal complexes to form the ends of stress fibre bundles (Sf). Focal adhesions at the rear slide during retraction (Fa*). Bar=10 μ m. Modified from Nemethova M, Auinger S, and Small JV (2008) Building the actin cytoskeleton: Filopodia contribute to the construction of contractile bundles in the lamella. *Journal of Cell Biology* 180: 1234–1244.

Focal complexes do not move relative to the substrate and can experience one of two fates. They either exist transiently for a few minutes and then disperse or enlarge and differentiate into larger anchorage sites at the ends of stress fiber bundles, the focal adhesions (Figures 1, 4(c), and 6). The transition from focal complexes to focal adhesions is linked to a switch in signaling from Rac/Cdc42 to Rho and to the engagement of muscle-type myosin with actin to form contractile bundles. Filopodia are often involved in focal complex formation and in initiating the assembly of contractile bundles (Figure 6).

The regulation of adhesion formation and turnover is a complex process that involves both enzymatic and mechanosensory pathways. Both focal complexes and focal adhesions are enriched in tyrosine kinases as well as their substrates, and changes in their activities modulate adhesion complex turnover. Focal adhesion formation and maintenance depend also on mechanical stress in the actin cytoskeleton. This is illustrated dramatically by their disappearance when cells are treated with drugs that inhibit the interaction between muscle-type myosin and actin. Mechano-sensory mechanisms therefore play a role in regulating adhesion dynamics, most likely through mechanically induced conformational changes of specific proteins in adhesion sites.

Adhesion Asymmetry and Polarization

For a cell to move, it must develop an advancing front and a retracting tail (Figure 2). This polarization process is reflected in an asymmetry of the pattern of adhesions developed with the extracellular matrix (Figure 6). It is interesting to consider how this asymmetry is established as it is relevant to the question of how polarization is determined.

As long as a cell edge protrudes, focal complexes are created in newly won territory under lamellipodia and filopodia, and a proportion differentiates into focal adhesions. In general, however, lamellipodia and filopodia do not protrude in a persistent manner. Instead, forward advancement is the net result of repetitive protrusion and retraction events. Retraction involves the withdrawal of lamellipodia or filopodia, or their backfolding, as ruffles. When this occurs, focal complexes dissolve, or convert into focal adhesions that temporarily tether the retracted cell edge.

In the above context, polarization can be considered the result of regional changes in the ratio of protrusion and retraction events at the cell edge. Thus, at the advancing front of a migrating cell the duration of protrusion exceeds that of

retraction. Elsewhere, at the rear and flanks, protrusion can occur, but retraction dominates. In the simplest terms, the front edge of a migrating cell is accordingly populated by focal complexes and the rear edge by focal adhesions. In many cases, focal adhesions are also formed behind the front edge and contribute to traction (**Figure 6**). This asymmetry of adhesion site development is the hallmark of a polarized cell. An important difference between focal adhesions at the front and rear of migrating cells is that those at the front remain fixed relative to the substrate, whereas those at the retracting rear and flanks can slide. The anterior adhesions therefore provide anchorage points for the actin cytoskeleton that support the retraction of the trailing cell body. Retraction itself is driven by the interaction of myosin with the actin filaments of the cytoplasmic network (**Figures 1 and 2**).

Microtubules and Cell Guidance

When fibroblasts are treated with drugs that disassemble microtubules, they become depolarized and extend protrusions in all directions. At the same time, tension in the actin cytoskeleton increases and focal adhesions grow in size. These changes are linked to an increase in the activity of Rho. Microtubules are therefore required for polarization and they exert their influence on polarity via a cross-talk with the actin cytoskeleton. The cross-talk apparently takes place between the tips of microtubules and the focal adhesion sites at the ends of actin bundles (**Figures 1 and 4(c)**).

Microtubules grow from the centrosome toward the cell periphery, but not in a continuous manner. Instead, they exhibit alternating periods of polymerization and depolymerization regulated by some of a growing number of proteins that accumulate at microtubule tips. When they reach the cell periphery, microtubules target focal adhesions and this targeting can occur in a repetitive manner. Targeting is more frequent at retracting cell edges and is correlated with the dispersal of focal adhesions or with their release from the substrate.

The global effect of microtubules is to reduce cell contractility. At the level of focal adhesions, microtubules appear to mediate the localized relaxation of actomyosin interactions

at the ends of stress fibers, to promote focal adhesion disassembly. This occurs in a mechano-sensory feedback mode, whereby an increase in mechanical stress at adhesion sites signals the polymerization of microtubules into them. The dependence of a cell on microtubules to maintain polarity is linked to the extent of formation of focal adhesions. Very motile cells, such as neutrophils, do not form stress fibers and typical focal adhesions and are less dependent on microtubules for directional locomotion. Likewise, fish epidermal keratocytes exhibit focal complexes, but lack focal adhesions and can also migrate without microtubules. Some cells appear to have the inherent ability to segregate protruding and contractile domains. However, most cells are unable to maintain this segregation without the modulatory input of microtubules. The nature of the putative molecular cross-talk between microtubules and adhesion sites, likely involving microtubule tip proteins, remains to be clarified.

Actin Filament Recycling: From Protrusion to Retraction

In their role as protruding organelles, lamellipodia and filopodia are major sites of actin filament generation in a motile cell. Whereas a proportion of the filaments generated for protrusion is depolymerized at the base of lamellipodia and filopodia, some filaments survive to seed formation of the rest of the actin cytoskeleton behind, which later serves in retraction. An important feature of this process is the conversion of filaments organized in polarized arrays for protrusion, into arrays of filaments organized in antiparallel arrays, for collaboration with myosin in contraction (**Figures 1 and 2**). The conversion from parallel to antiparallel arrays results from the active contralateral and rearward movements of filopodia and from the folding backward of some lamellipodia and filopodia into the cell as ruffles (**Figure 7**). In addition, those filopodia active in adhesion formation relinquish their basal parts to initiate the formation of contractile, stress fiber bundles. This recycling activity occurs around the entire cell periphery, to a more or lesser extent. The direction of movement is then determined by the region of the cell with the upper hand in protrusion, namely where lamellipodia and filopodia dominate (**Figure 7**).

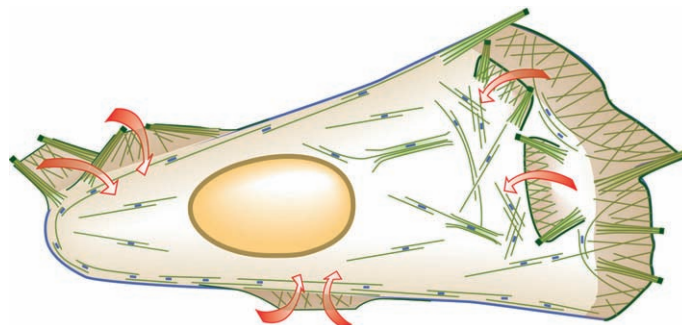


Figure 7 Actin filament re-cycling, from protrusion to retraction. Schematic illustration of the generation of contractile arrays in a hypothetical cell lacking stress fiber bundles (e.g., a neutrophil). Ruffles, including lamellipodia and filopodia, fold back into the cell and contribute actin filament seeds for the formation of antiparallel assemblies with myosin that interconnect throughout the cell body. The bilateral flow of filaments in lamellipodia also contributes antiparallel filament arrays via the base of the lamellipodium. The contractile bundle around the periphery of the trailing part of the cell is formed via prior ruffling activity at those locations. Actin filaments are in green, myosin filaments in blue. Modified from Small JV and Rottner K (2010) *Elementary cellular processes driven by actin assembly: Lamellipodia and filopodia*. In: Carrier M-F (ed.) *Actin Based Motility: Cellular, Molecular and Physical Aspects*. Springer Science.

Migration *In Vivo*

Cells migrating in tissues exhibit both filopodia and lamellipodia. These thin protrusive extensions are ideal for exploring and penetrating tissue spaces. They are also well suited for intercalating between cells, such as during the migration of leukocytes across endothelial layers. Stress fibers, as seen in cultured cells, are not obvious features of cells moving in a three-dimensional environment. However, inhibition of the Rho pathway – which signals contractility – inhibits the migration of primordial cells during embryogenesis and prevents tail retraction during the transendothelial migration of monocytes. Migration *in vivo* therefore involves protrusion of lamellipodia and filopodia signaled by Rac and Cdc42 as well as retraction signaled by Rho. The closure of epithelial layers during embryogenesis is likewise dependent on protrusion and retraction events, whereby the contractility of actin bundles parallel to the epithelial boundary contributes to closure in a zipper-like mode.

Further factors important for cell migration have not been discussed here. These include the cues initiating migration at an appropriate timepoint and the gradients of chemotactic factors that define the destinations of migrating cells. Needless to say, these must impinge on the network of pathways that signal to the locomotion machinery.

See also: **Cell Architecture and Function:** Actin Assembly/Disassembly; Centromeres; Focal Adhesions and Related Integrin Contacts; **Signaling:** Integrin Signaling.

Further Reading

Bray D (2000) *Cell Movements, from Molecules to Motility*. New York: Garland.
Chesarone MA and Goode BL (2009) Actin nucleation and elongation factors: Mechanisms and interplay. *Current Opinion in Cell Biology* 21: 28–37.

- Faix J, Breitsprecher D, Stradal TE, and Rottner K (2009) Filopodia: Complex models for simple rods. *International Journal of Biochemistry and Cell Biology* 41: 1656–1664.
- Friedl P and Gilmour D (2009) Collective cell migration in morphogenesis, regeneration and cancer. *Nature Reviews Molecular Cell Biology* 10: 445–457.
- Geiger B, Spatz JP, and Bershadsky AD (2009) Environmental sensing through focal adhesions. *Nature Reviews Molecular Cell Biology* 10: 121–133.
- Heasman SJ and Ridley AJ (2008) Mammalian Rho GTPases: New insights into their functions from *in vivo* studies. *Nature Reviews Molecular Cell Biology* 9: 690–701.
- Koestler SA, Auinger S, Vinzenz M, Rottner K, and Small JV (2008) Differentially oriented populations of actin filaments generated in lamellipodia collaborate in pushing and pausing at the cell front. *Nature Cell Biology* 10: 306–313.
- Ladwein M and Rottner K (2008) On the Rho'd: The regulation of membrane protrusions by Rho-GTPases. *FEBS Letters* 582: 2066–2074.
- Laemmernann T and Sixt M (2009) Mechanical modes of “amoeboid” cell migration. *Current Opinion in Cell Biology* 21: 1–9.
- Lai FPL, Szczodrak M, Block J, et al. (2008) Molecular dissection of lamellipodium protrusion by FRAP. Dissection of actin network turnover in lamellipodia. *EMBO Journal* 27: 982–992.
- Le Clainche C and Carlier M-F (2008) Regulation of actin assembly associated with protrusion and adhesion in cell migration. *Physiological Review* 88: 489–513.
- Nemethova M, Auinger S, and Small JV (2008) Building the actin cytoskeleton: Filopodia contribute to the construction of contractile bundles in the lamella. *Journal of Cell Biology* 180: 1234–1244.
- Pollard TD (2007) Regulation of actin filament assembly by arp2/3 complex and formins. *Annual Review of Biophysics and Biomolecular Structure* 36: 451–477.
- Small JV and Rottner K (2010) Elementary cellular processes driven by actin assembly: Lamellipodia and filopodia. In: Carlier M-F (ed.) *Actin Based Motility: Cellular, Molecular and Physical Aspects*. Springer Science.
- Steffen A, Rottner K, Ehinger J, et al. (2004) Sra-1 and Nap1 link Rac to actin assembly driving lamellipodia formation. *EMBO Journal* 23: 749–759.
- Takenawa T and Suetsugu S (2007) The WASP-WAVE protein network: Connecting the membrane to the cytoskeleton. *Nature Reviews Molecular Cell Biology* 8: 37–48.

Relevant Websites

<http://cellix.imba.oeaw.ac.at> – A Video Tour of Cell Motility.
<http://www.cellmigration.org> – Nature, Cell Migration Research and Resources.
<http://mitchison.med.harvard.edu> – Welcome to the Mitchison Lab, Department of Systems Biology, Harvard Medical School.

Cell Migration within Three-Dimensional Matrices

P J Keely, University of Wisconsin, Madison, WI, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Adaptor protein A signaling protein that has no intrinsic catalytic activity but serves as a scaffold to activate other proteins or bring signaling proteins together.

Basement membrane A specialized, thin (~0.2- μ m-thick) extracellular matrix (ECM) that is found at the basal surface of epithelial cell sheets, composed of collagen IV, laminins, proteoglycans, entactin, and other glycoproteins.

Cell adhesion The process by which a cell attaches to the ECM using specific cell-surface receptors, including integrins and proteoglycans.

Cell-cell adhesion The process of cell attachment to other cells, usually through cadherins, cell-adhesion molecules (CAMs), and integrins.

Cell migration The process by which a cell moves from one location to another, involving membrane protrusion, cell adhesion, and forward translocation of the cell body.

Collagen The major structural protein of vertebrates and comprising a family of over 18 types of collagen that are fibrillar and nonfibrillar.

Extracellular matrix The proteins, proteoglycans, and carbohydrates that comprise the environment and connective tissue surrounding cells.

Fibronectin A specific ECM glycoprotein deposited largely by fibroblasts and containing several binding sites for cellular receptors, collagens, and proteoglycans.

Focal adhesions The site where integrin and proteoglycan receptors are clustered and connect to the ECM on the outside of the cell and the actin cytoskeleton on the inside of the cell; focal adhesions are additionally the site where over 100 signaling molecules are scaffolded in response to cell adhesion.

Glycoproteins Extracellular proteins that contain sugar moieties covalently attached.

Integrins The major class of cell-adhesion receptors for the ECM, integrins, are obligate heterodimers comprised of an α - and a β -subunit that can combine to result in ligand specificity for the ECM.

Invasion The process by which cells migrate into tissues and three-dimensional (3D) matrices, which involves the steps of cell migration, and may include proteolysis of the ECM.

Microenvironment The conditions, including ECM, growth factors, pH, glucose, oxygen, and other components that are proximal to cells.

Proteases Proteins that have enzymatic function to cleave other proteins.

Proteoglycans A family of molecules comprised of a protein core and covalently linked glycosaminoglycan side chains.

Rho GTP-binding proteins A subfamily of small guanosine triphosphate (GTP)-binding proteins within the Ras superfamily comprised of Rac isoforms, Rho isoforms, and Cdc42; small GTP-binding proteins cycle between an inactive, GDP-bound form, and an active, GTP-bound form; activation is accompanied by a conformational change that allows the protein to bind to and activate down-stream effector molecules.

SH2 domains A specialized protein domain that can bind to phosphorylated tyrosines when they are in a Y-x-x-P motif (tyrosine-any peptide-any peptide-proline).

Tyrosine kinase A signaling protein that has catalytic activity allowing it to add a phosphate group from the γ phosphate of adenosine triphosphate (ATP) to a tyrosine residue of a protein; usually the tyrosine residue occurs in a particular context, and not all tyrosines are substrates for a particular tyrosine kinase.

Introduction

For a few decades, biologists have studied cell migration in the hopes of better understanding this complex event. Much of the work has been performed in cell culture environments, often on a two-dimensional (2D) plastic or glass surface coated with proteins. The study of cell migration in 2D has allowed the simplification of conditions to better understand the contributions of the extracellular matrix (ECM), cell-surface receptors, and signaling pathways that are involved. The relative ease of imaging a mostly flat cell crawling along a 2D surface has allowed detailed dynamic analysis of adhesion structures and the associated cytoskeleton at various stages of cell migration. From this seminal body of work, much understanding has emerged regarding the dynamics of cell migration, and we now appreciate that cell migration in 2D involves several

steps: forward protrusion of the membrane driven in part by dynamic polymerization and de-polymerization of both actin and tubulin-based cytoskeletal structures, formation of early or nascent adhesions to the underlying substratum, forward translocation of the cell body, strengthening of adhesions at the front, and weakening or tearing of adhesions at the rear to pull forward the rear of the cell (Figure 1).

Using this knowledge as a base, the field of cell migration is now ready to tackle the more complex questions of how cells migrate within 3D matrices, which is necessary in order to understand cell migration *in vivo*. There are several examples of cell migrations during both normal development and pathological conditions to which this knowledge is relevant. This article focuses specifically on our current understanding of cell migration and invasion within 3D matrices, both in culture and *in vivo*.

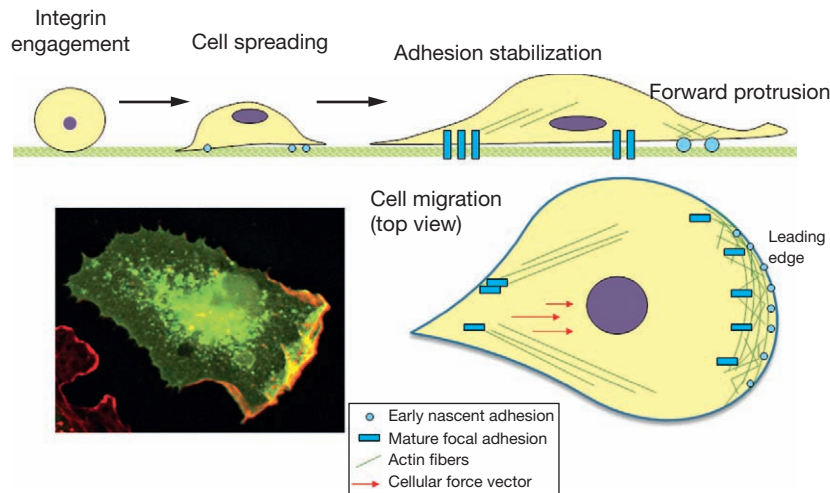


Figure 1 Stages in cell migration. Cells adhere to the ECM using integrin receptors and quickly reorganize their actin to initiate cell spreading. Cells migrate with a forward thrust of the leading edge, which involves the polymerization of actin into a branching network. This allows the cell to contact the ECM at this protrusion. Cell adhesions start out small and dynamic at the leading edge, and over time mature as more proteins are clustered into the adhesions. This stabilizes the front of the cells. At the rear of the cell, contractile forces move the cell forward and result in release or tearing of the rear adhesions, and forward translocation of the cell body. Inset: A Cos-7 cell stained to reveal the dense actin network at the leading edge (actin is stained red and appears yellow in this figure where it co-localizes with the signaling molecule, R-Ras).

Structure and Function of the ECM

Cell Migration and Invasion Occurs within the ECM

ECM is a term applied to the collection of glycoproteins and proteoglycans (PGs) that are secreted by cells and deposited into the extracellular spaces surrounding cells. The ECM can be a few proteins surrounding a cell, as part of the cells' microenvironment, or it can encompass several hundred microns or even centimeters of structure that we term connective tissue. Although specialized ECMs exist in specific tissues such as brain, heart, nerves, blood vessels, bone, and cartilage, we can generalize our consideration of the organization and structure of ECM by discussing epithelial organs, as the concepts are similar and relevant to other organ sites.

Epithelial cells sit on a specialized ECM that is a mere 0.2 μm thick, termed the basement membrane (Figure 2). Collagen IV, laminins, PGs, and entactin are major components of this structure and form a densely woven mesh that serves both a biochemical (as a ligand to adhesion receptors) relationship and a structural (restraining) function to the gland. Epithelial cells attach directly to the basement membrane via adhesion receptors found at hemidesmosomes. A similar structure that is laminin rich surrounds individual nerve cells and bundles, and underlies endothelial cells in blood vessels. Thus, basement membranes serve to delineate specialized cells from the surrounding loose stromal connective tissue in multicellular organisms. Such a structure is found in invertebrates such as worms and insects all the way through insects and all vertebrate organisms, demonstrating its important role in the development and maintenance of tissue architecture. Indeed, disruption of the basement membrane is a hallmark of invasive cancer, a process that involves a loss of differentiation.

Surrounding the basement membrane is the stromal matrix, composed largely of collagen I and, to a lesser extent, other fibrillar collagens such as collagen III, as well as fibronectin and PGs. The stromal matrix is loosely organized around epithelial cells, yet also serves both biochemical and structural roles (Figure 2). A similar stromal matrix surrounds nerves, muscle, and blood vessels, although it should be noted that as our understanding of ECM increases, so does our realization that subtle tissue-specific differences exist. Fibroblasts found in the stroma are the predominant source of ECM proteins, although other cell types including macrophages can also deposit ECM. Moreover, in culture, epithelial cells deposit ECM proteins, especially those found in the basement membrane, suggesting they may contribute to this structure.

The ECM Has Both Biochemical and Mechanical Properties

The components of the ECM have long been studied as true biochemical ligands for cell-surface receptors, including integrins and PGs. The ECM exhibits multiple and discrete binding sites within the proteins that compose it. For example, fibronectin has several such sites, including an arginyl-glycine-aspartate-serine (RGDS) site that binds to several integrins, heparin-binding sites, and binding sites within the N-terminus. The RGD sequence has been found in several other ECM proteins, including vitronectin, fibrinogen, and denatured collagen. In addition to sites important for cell adhesion, fibronectin has sites that help cells assemble it into a fibrillar matrix at its N-terminus, as well as sites that bind other ECM molecules such as collagen. A similar story is found for laminins, which have both integrin-binding sites, such as the YIGSR peptide, and heparin-binding sites.

It has recently been suggested that the mechanical and structural properties of the ECM contribute to its tissue

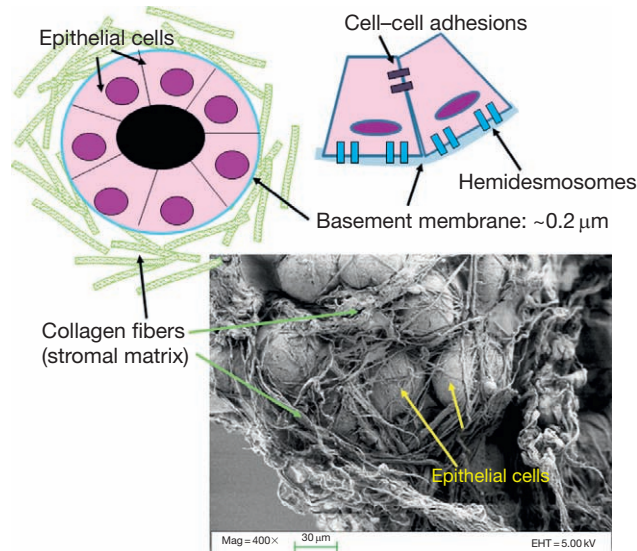


Figure 2 Cells interact with a basement membrane. Epithelial cells organized into the terminal functional unit of a duct, the ascinus, are diagrammed in cross section. Cells attach to the 0.2- μ m-thick basement membrane through integrins at hemidesmosomes and to one another at cell–cell adhesions. Surrounding the basement membrane is the stromal matrix, comprised of collagens, fibronectins, and PGs. Image: a scanning electron micrograph (SEM) image of a mouse mammary ascinus demonstrating collagen fibers wrapped around the ascinus. Collagen fibers appear as ropes of various sizes, depending on the number of fibrils in the bundle. SEM courtesy of Dr. Paolo Provenzano.

specificity and the resulting cellular response. Matrices that have identical composition can vary in stiffness, or elastic modulus. Matrix stiffness can result due to cross-linking of proteins within the matrix, deposition of more proteins, or rearrangement of components. Collagen fibers in particular contribute to matrix stiffness, as an increase in the number of fibers or their cross-linking or bundling results in matrices that are stiffer. The idea that collagen fibers contribute a mechanical constraint on cells becomes intuitive when looking at the scanning electron micrograph in **Figure 2**.

Matrix stiffness can also determine the fate of developing cells. Mesenchymal stem cells that are cultured on stiff matrices that mimic the elastic modulus found in bone differentiate into osteocytes. When the same cells are cultured on an identical matrix differing only in stiffness, softer matrices that mimic muscle stiffness induced cells to become myocytes.

Matrix stiffness is also important in the progression of lung fibrosis in asthma and chronic pulmonary disease, as deposition of collagen and other ECM components causes the lung to become stiffer, aggravating inflammation and diminishing the capacity of the alveoli to expand during breathing. A similar situation is observed in atherosclerosis, where increased deposition of ECM around blood vessels makes the vessels stiffer, resulting in ‘hardening of the arteries’ and an increase in blood pressure as the vasculature loses its elasticity. In atherosclerotic plaques, smooth muscle cells that have infiltrated into the inside of the blood vessel deposit collagen, which forms a stiff nodule. This collagen deposition works in a feed-forward mechanism to promote the proliferation of smooth muscle cells, resulting in an ever-growing plaque that will ultimately occlude a blood vessel.

Interestingly, matrix stiffness accompanies cancer progression as well. A well-known pathological finding in many

cancers is the deposition of ECM, predominantly collagens, around tumors, a process termed desmoplasia. Desmoplasia is associated with poor outcome, and recently several groups have demonstrated that stiff matrices can promote cell proliferation and cancer progression. This finding is consistent with the observation that many tumors can be found by palpation because they are stiffer than the surrounding tissue.

Receptors for the ECM: Integrins

Cell migration will only occur if the cells can attach to the ECM, and thus it is important to consider the adhesive receptors employed by cells for this purpose. The integrin family of cell-adhesion receptors is crucial for cell migration, as blocking these receptors blocks cell adhesion and migration. Integrins are composed of an α - and β -subunit, which can associate in a combinatorial manner to create several receptors (**Figure 3**). The subfamilies are based on the β -subunit, with β 1-integrins being receptors for ECM proteins, including collagen, fibronectin, laminin, vitronectin, tenascin, and other extracellular glycoproteins.

The β 2-integrins are found on hematopoietic cells and mediate leukocyte trafficking out of blood vessels and cell–cell interactions. The β 3-integrins are most important for the unique member, α IIb β 3, which is found only on platelets and mediates platelet aggregation by binding to fibrinogen/fibrin. The α v-subunit is the most promiscuous of the α -subunits, combining with β 1, β 3, β 7, β 8 and predominantly mediating adhesion to vitronectin. The β 4-integrin subunit is unique as it harbors an unusually long cytoplasmic domain that is thought to have additional signaling function. When combined with the α 6-subunit, the α 6 β 4-integrin binds to laminin in the

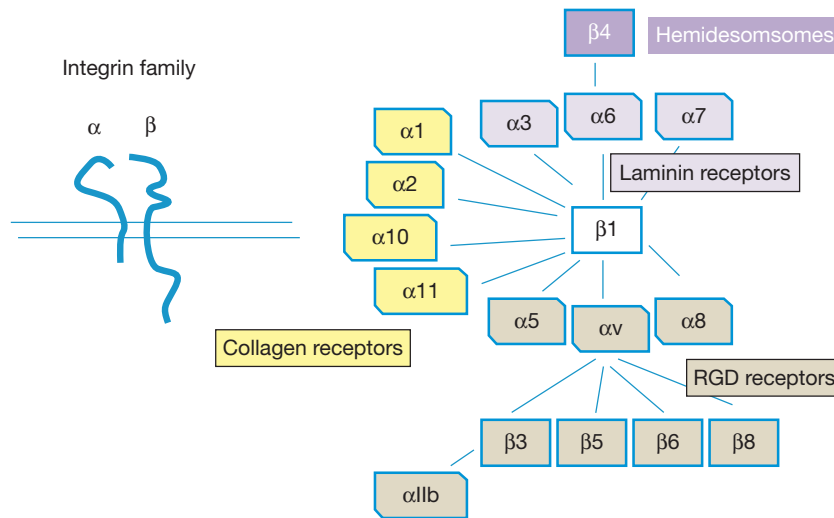


Figure 3 The integrin family of receptors is composed of an $\alpha\beta$ -heterodimer. These heterodimers are assembled in a combinatorial fashion that determines their ligand specificity and intracellular signaling events. The RGD-binding integrins will bind to exposed arg–glu–asp (RGD) peptide sequences in several molecules, including fibronectin, vitronectin, denatured collagen, and fibrinogen. Adapted from Hynes RO (2002) Integrins: Bidirectional, allosteric signaling machines. *Cell* 110: 673–687.

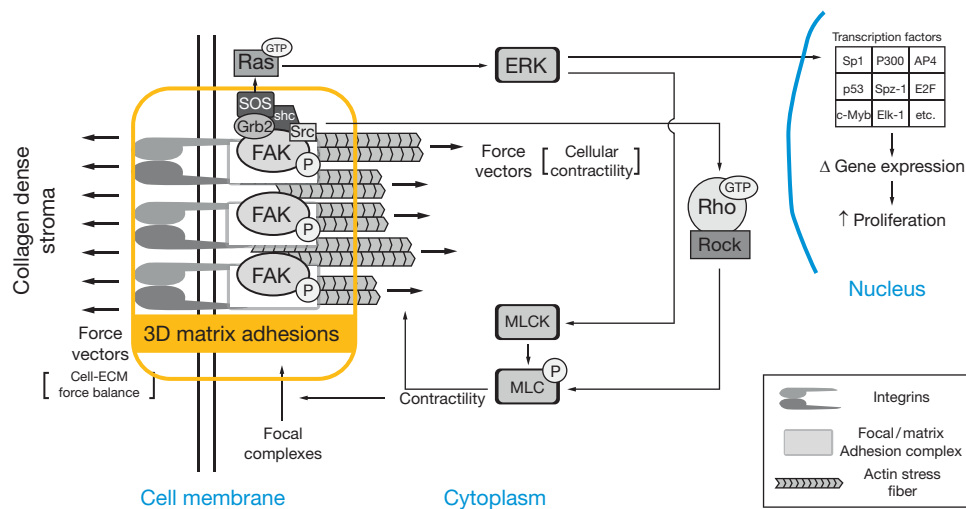


Figure 4 Signaling pathways associated with matrix adhesions. Integrins are clustered as a result of opposing forces due to contraction of the actin cytoskeleton and the opposing stiffness of the ECM. Contractility is driven by the Rho GTP-binding protein, and its effector, Rho-kinase (ROCK), which in turn regulates MLC phosphorylation, leading to myosin contraction of actin. Clustering assembles numerous signaling molecules, including focal adhesion kinase (FAK) and Src. In turn, the Ras pathway is activated, activating extracellular regulated kinase (ERK), which will translocate into the nucleus and regulate gene expression. ERK can also regulate MLC-kinase (MLCK) and thus impinges upon contractility. Adapted from Provenzano PP, Inman DR, Eliceiri KW, and Keely PJ (2009) Matrix density-induced mechanoregulation of breast cell phenotype, signaling and gene expression through a FAK-ERK linkage. *Oncogene* 28: 4326–4343.

basement membrane and is found in unique adhesion structures, termed hemidesmosomes, at the basal surface of epithelial cells. Upon loss of adhesion to the basement membrane in cancer progression, $\alpha 6\beta 4$ -integrin is recommissioned to mediate cell migration along laminin fibers.

Integrins mediate signal transduction via binding proteins to their cytoplasmic tails, which are generally rather short. Nonetheless, several proteins have been demonstrated to bind directly to integrins, and include actin-binding and signal-scaffolding proteins such as filamin, talin, and paxillin, signaling proteins such as focal adhesion kinase (FAK),

Src-family kinases and phosphatidylinositol kinases, and signaling adaptors such as p130Cas and Shc. Thus, integrins are well poised to transmit signals from the ECM to the inside of the cell (Figure 4).

Proteoglycans

PGs are generally large molecules composed of a protein core to which are attached several sulfated glycosaminoglycans, such as chondroitin sulfate or heparan sulfate. Several PGs are secreted into the ECM, where they bind to ECM

glycoproteins and modify the organization of the matrix. PGs can also exist on the cell surface as trans-membrane proteins. Cell-surface PGs, including the syndecans (1, 2, 3, 4), and melanoma chondroitin sulfate proteoglycan (MCSP) also bind to the ECM ligands at heparin-binding sites within the ECM. Cell-surface PGs are necessary for proper adhesion structures to be formed and collaborate with integrins to mediate cell migration. It has recently been demonstrated that cell-surface PGs regulate integrin function; for example, syndecan-1 can help to activate $\beta 1$ integrins. Moreover, cell-surface PGs can transduce signals: MCSP activates the small guanosine triphosphate (GTPase), Cdc42, while syndecan-4 binds, through its cytoplasmic tail, to protein kinase C.

Clustering of Receptors into Focal Adhesions

Cell-adhesion receptors cluster into specialized adhesion structures, termed focal adhesions, which connect within the cell to actin stress fibers. Indeed, integrins were named decades ago because they appear to integrate aligned ECM fibers outside the cell with aligned actin fibers inside the cell. Early adhesions that are not yet associated with stress fibers, but rather a network of actin filaments, are termed focal contacts. Evidence suggests that these early adhesions are more dynamic than focal adhesions and that focal contacts often mature into focal adhesions as actin stress fibers are formed. The concept of focal contacts and focal adhesions is one that has been well described for cell adhesion and migration on 2D surfaces *in vitro*, but has been challenged for cells within 3D matrices and *in vivo*. However, solid evidence exists for the formation of an adhesive structure, termed matrix adhesion, that is found in cells resident within 3D matrices. When live cells expressing a fluorescent version of the $\beta 1$ -integrin subunit (GFP- $\beta 1$) are imaged in a 3D collagen matrix, we find that the $\beta 1$ -integrin clusters along the collagen fibers (Figure 5). Thus, in this article we use the terms focal contact and focal adhesion to indicate data obtained in 2D studies, and matrix adhesion to indicate data obtained specifically in 3D.

Signaling Molecules that Regulate Cell Adhesion and Migration

Rho GTP-Binding Proteins

The dynamics and maturation of focal adhesions are regulated in part by small guanosine triphosphate (GTP)-binding proteins of the Rho family, Rho, Rac, and Cdc42. These small GTP-binding proteins are members of the Ras superfamily of proteins. As such, they cycle between a guanosine diphosphate (GDP)-bound inactive state and a GTP-bound active state. The active state is accompanied by a conformational change in the protein that allows the GTP-binding protein to bind to other proteins and activate them. These downstream proteins are termed effectors, as they mediate, or effect, the outcome of GTP-binding protein activation. Each small GTP-binding protein can activate several effectors, although exactly which effector is bound under specific conditions is still under investigation.

Both Rac and Cdc42 are found at the leading edge of migrating cells, where they promote membrane protrusion and control the formation of lamellipodia and filopodia by

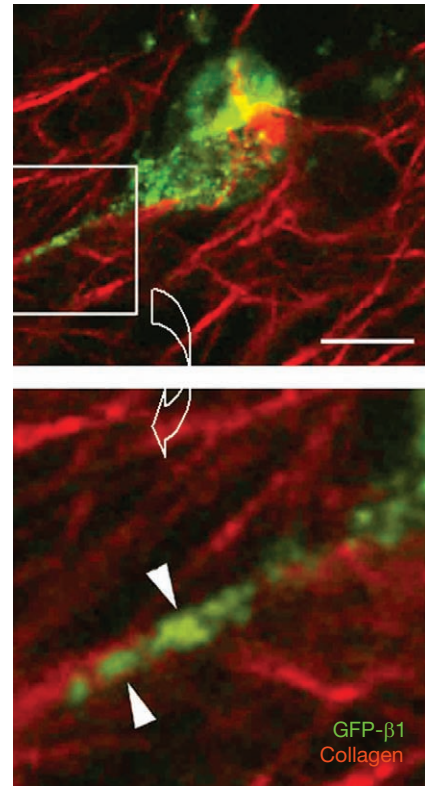


Figure 5 Integrins cluster at sites of cell adhesion along a collagen fiber. Breast carcinoma cells were engineered to express GFP- $\beta 1$ integrin and cultured within a 3D collagen matrix. Live cells were imaged using multiphoton microscopy to visualize collagen (red, second harmonic generation) and $\beta 1$ integrins (green). Scale = 10 μm . Micrograph courtesy of Kristin Riching.

their regulation of the actin network. This occurs through several effectors that include N-WASP, WAVE, and p21-activated kinase (PAK). Rac and Cdc42 both regulate early focal contact formation and are further activated by cell adhesion. Rho, on the other hand, drives stress fiber formation and maturation of focal adhesions through its effectors, including Rho-kinase (ROCK) and mDia (Figure 6).

Rho is the central player in generating cellular contractility, which is important for cell migration in several ways. Rho signals to ROCK to increase the phosphorylation of myosin light chain (MLC). When phosphorylated, MLC exposes the active site on myosin, engaging its adenosine triphosphatase activity and the subsequent contraction of actin. The bundling of actin into stress fibers and the clustering of integrins into focal adhesions are driven by this actomyosin contractility, initiated by Rho activation. As Rho-generated contractility occurs, the actin cytoskeleton clusters integrins through inside-out signaling, resulting in focal adhesion or focal contact formation.

In 3D matrices, the outcome becomes a bit more complicated. Focal adhesion-like structures form in 3D matrices, but are smaller and differ in the proteins involved compared to those that form on a 2D glass surface. These adhesions, termed matrix adhesions, are formed only when cells are within a 3D matrix that is stiff enough to resist the contraction of the cell

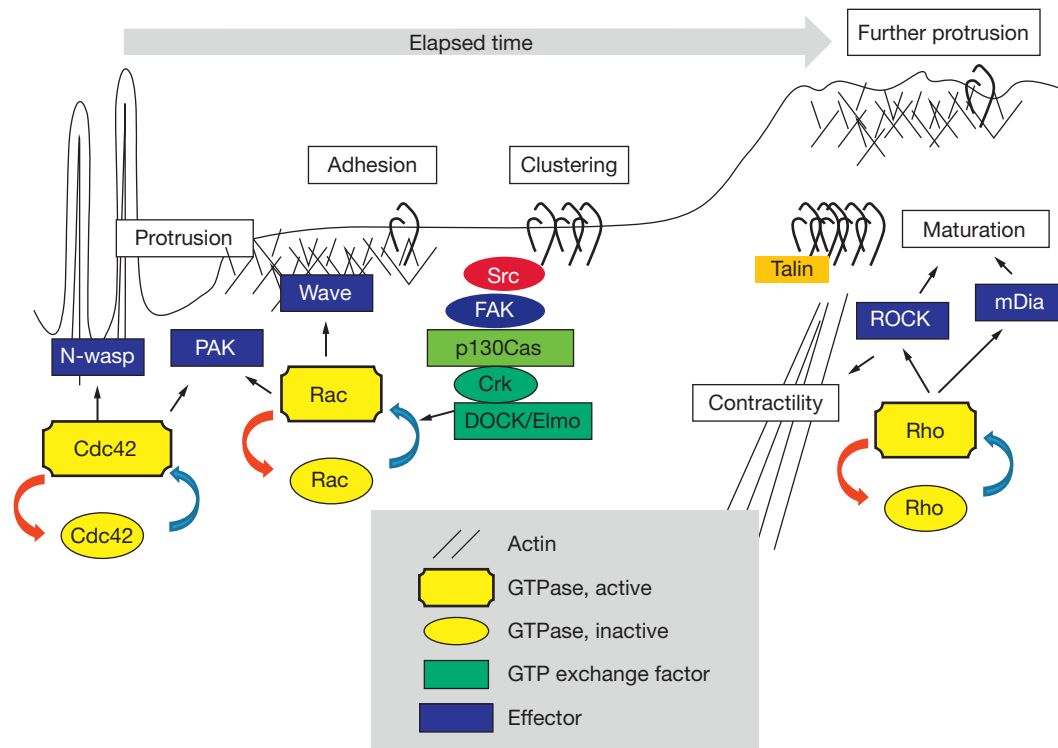


Figure 6 Signaling pathways linked to small GTP binding proteins involved in cell migration. Small GTP binding proteins cycle from an inactive, GDP-bound, form to an active, GTP-bound form. This is regulated by GDP exchange factors. Cdc42 and Rac activate their effectors, including N-WASP and WAVE, to regulate early events of actin polymerization resulting in filopodial and lamellipodial protrusions. These protrusions allow early adhesions to form at sites of ECM contact. Rho activates several effectors including ROCK and mDia to enhance actin bundling, integrin clustering, and adhesion maturation. This stabilizes the cell, allowing further forward protrusion of the leading edge.

against the matrix. If the matrix is compliant, rather than generating matrix adhesions, the cell will contract the ECM, condensing it around the cell. Increasing the signaling through the Rho pathway therefore allows cells to tune their response and contract a stiffer matrix.

The result of this in terms of cell migration is twofold. For cells that find themselves in a stiff matrix, cellular contraction allows the cell to pull against the matrix and move within it. In contrast, within a compliant matrix, such as one that is not organized, cellular contractility allows cells to exert forces on collagen and fibronectin fibers.

Focal Adhesion Proteins

Several proteins are recruited to focal complexes and focal adhesions upon clustering, and serve to transduce downstream signals that regulate the behavior of the cell. A central player is FAK, a tyrosine kinase that auto-phosphorylates itself at Y379 upon clustering (Figure 4). The resulting addition of a phosphate group to Y397 results in a docking site for other signaling molecules that contain SH2 domains, including Src, Grb7, and phosphoinositide-3-kinase. Importantly, all of these molecules that bind to FAK promote cell migration.

The Src tyrosine kinase in turn phosphorylates FAK on Y925 to provide an additional docking site for the adaptor, Grb2. The SHC and Grb2 adaptors are important scaffolds that lead to activation of the Ras GTP-binding protein through its exchange factor, SOS (Figure 4). Ras activation is associated

with cell proliferation, cell survival, and increased cell migration, all properties of tumor cells.

FAK and Src also serve to scaffold the adaptor protein, p130Cas. This molecule, like other scaffolding molecules, has no intrinsic enzymatic function but serves to activate other molecules through binding. In the case of p130Cas, it subsequently binds Crk and the Rac exchange factor DOCK 180/ELMO. Activation of this pathway will thus lead to activation of the Rac GTP-binding protein (Figure 6).

Cell Migration *In Vivo* Has Both Normal and Pathological Outcomes

Cell Migration Is Key to Proper Development

One of the first major events dependent on cell migration occurs early in development during gastrulation, in which invagination and migration of the presumptive ectoderm toward the center of the embryo forms the three layers of the embryo: ectoderm, mesoderm, and endoderm. This process does not occur in embryos lacking fibronectin, indicating the importance of the ECM (and cell migration) during development. As development proceeds, more precise migration events are important, as movement of various cell groups sets up the architecture and specialization of the embryo. Invagination of epithelial sheets via collective cell migration forms glandular organs such as salivary glands and breast from ectoderm, and liver and pancreas from endoderm.

During development of branching glandular organs, such as the lung, breast, and salivary gland, fibronectin becomes deposited along the developing gland at locations where branches are to occur. As the gland develops, it must invade into the surrounding fat pad, and it is therefore of interest to note that collagen fibers are found at these tips, radiating outward in the direction the gland will grow. It is thus postulated that collagen fibers represent a path for the growing end buds to follow as the gland is growing.

The ECM guides cell migration in the peripheral nervous system as well. The peripheral nervous system is derived by migration of neural crest cells from their location ventral and proximal to the neural tube – at the crest of the neural tube – into various locations throughout the body. Once resident in their intended location, peripheral neurons then undergo a specialized form of cell migration in which the neuronal cell body does not move, and long cell extensions, termed neurites, migrate into the target organ(s) and into the central nervous system. Laminin is located along paths on which neurites migrate, and neural crest cells find their way along collagen and fibronectin matrices.

Cell Migration in Immune Trafficking and Wound Healing

Immune cells rely on cell migration throughout adulthood in order to traffic between the blood stream and into tissues to maintain immune surveillance. For example, neutrophils and macrophages are involved in inflammation and wound healing, and readily move out of the blood stream to migrate through tissues toward inflamed or wound sites. These cells are able to recognize chemical mediators, including chemokines released at the wound site to migrate directionally toward the wound, a process termed chemotaxis.

The presence of a wound and its resultant blood clot activates local fibroblasts, which migrate into the wound and deposit a provisional ECM. The provisional matrix in turn allows epithelial cells at the wound site to migrate and close the wound, restoring tissue integrity.

Cell Migration during Pathological Conditions

While wound healing is a necessary process for complex organisms, sometimes this process becomes mis-regulated and pathologic inflammation results. Excess activation and infiltration of immune cells into the lungs of asthma patients is one such example. Many diseases are now thought to either involve directly or are made worse by inflammation, including Alzheimer's dementia, coronary artery diseases, diabetes, and cancer.

Atherosclerotic plaques are characterized by the infiltration of smooth muscle cells, which migrate from their location surrounding arterioles into the lumen of the blood vessel, a location where they are not normally found. These smooth muscle cells exacerbate the plaque by depositing the ECM protein collagen, which recruits yet more smooth muscle cells and possibly immune cells. Rupture of the plaque exposes this deposited collagen, causing platelets to attach to the newly exposed collagen and form a clot. This process activates the platelets to secrete many chemical signals that attract additional cells, which then migrate into the area.

Cell migration in 3D matrices is particularly important in cancer progression. Local invasion into the surrounding ECM is one of the hallmarks of a malignant tumor compared to one that is considered benign. In addition to local spread, as tumors progress they spread to more distant organs, a process termed metastasis. Metastasis is a multistep process that includes local invasion away from the tumor through the surrounding stromal ECM (Figure 7), migration into either lymph or blood vessels (intravasation), flow and trafficking through the blood stream, exit from the lymph or blood vessel (extravasation), and survival and growth in the target organ.

Cancer Invasion Is Facilitated by the Structure of Stromal Matrices

Once formed, epithelial cells normally reside within the confines of the basement membrane. However, transformation of cells to a cancerous state fundamentally alters their relationship to the ECM, such that cancer cells both cross the basement membrane and transit into and through the stromal matrix. It is generally believed that protease activity is necessary for cells to cross the dense meshwork of the basement membrane.

A dynamic feedback relationship exists between a developing tumor and the stromal matrix. As tumors arise and progress, they secrete several cytokines and growth factors that recruit both stromal fibroblasts as well as bone marrow-derived inflammatory cells including macrophages. Intriguingly, the fibroblasts that surround a developing tumor are fundamentally different in their gene expression and behavior than normal resident stromal fibroblasts, and are termed carcinoma-associated fibroblasts (CAFs). When co-cultured with normal cells, these CAFs can make normal cells behave like cancer cells such that the normal cells become invasive and proliferative.

The infiltration and proliferation of cells in the stroma, predominantly CAFs and macrophages, result in an increase in matrix deposition. What is particularly intriguing is that not only do fibroblasts and macrophages step up their production of collagen and fibronectin, but it has recently been shown that they can deposit these fibers as an aligned matrix. During breast carcinoma progression in both animal models and humans, changes in the organization of collagen from a loose ECM to straight and ultimately aligned collagen is a marker of tumor progression (Figure 7).

The production of an aligned matrix is significant, as cell migration *in vitro* is enhanced along 3D matrices that have aligned collagen fibers compared to matrices that are randomly organized. Moreover, the orientation of the collagen is important, such that cells prefer to migrate along fibers that are oriented perpendicular to the tumor boundary, providing a surface akin to a highway on which to migrate. In contrast, when fibers are oriented tangential to the tumor boundary, cell migration is minimal. *In vivo*, cells have been shown to migrate from tumors along collagen fibers, appearing almost tight-rope walkers across the stroma. The finding that perpendicular, aligned collagen forms a preferred substratum for 3D cell migration, and the observation of tumor cells *in vivo* using collagen fibers for migration, suggests that the alteration in collagen alignment that surrounds tumors is a key contributor to local invasion.

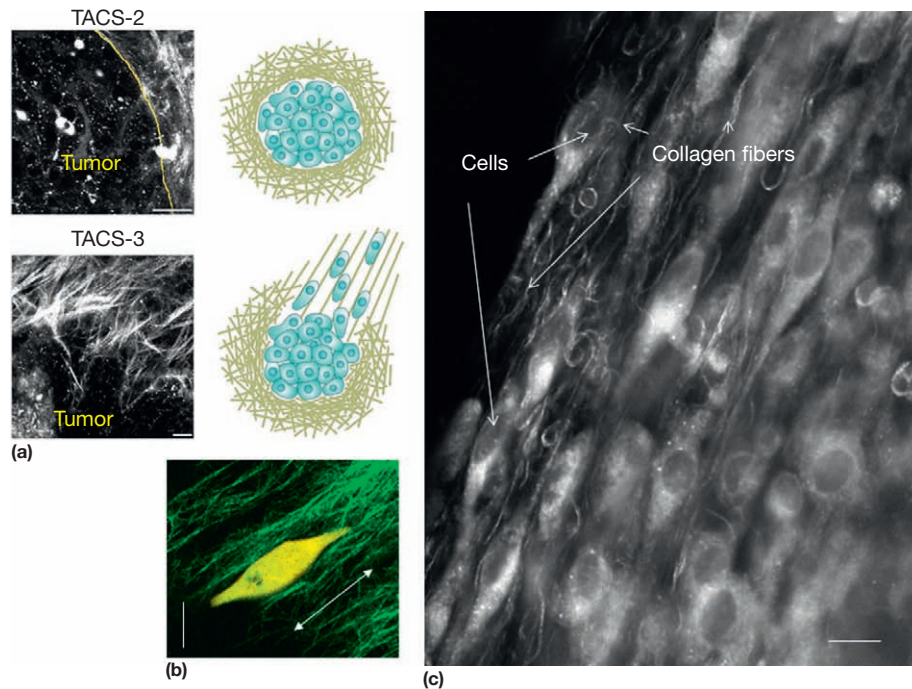


Figure 7 Carcinoma cells invade along collagen fibers. (a) Collagen alignment is changed during tumor progression. We have defined tumor-associated collagen signatures (TACSs) that describe these changes. TACS-2: Alignment of collagen fibers tangential to the tumor boundary; TACS-3: Alignment of collagen fibers perpendicular to the tumor boundary. Images are of collagen in mouse mammary tumors (collagen appears white, imaged by SHG). The tumor cells are not visible, but reside in the dark space. The tumor/stromal boundary is drawn in yellow. The cartoon depicts the relationship of the tumor cells to the collagen fibers. (b) Breast carcinoma cells cultured in an aligned collagen matrix align themselves relative to this matrix. Arrow indicates axis of collagen alignment. Cell (yellow) and collagen (green, imaged by SHG). Image courtesy of Dr. Steve Trier. (c) Breast carcinoma cells invade through the collagen-rich stroma in a mouse mammary tumor. Cells were unstained, imaged in a tumor explants. Note the numerous cells migrating in parallel along collagen fibers, which are aligned and guide the cells. Scale = 10 μm . Image courtesy of David Inman.

Migration along aligned collagen fibers may reflect a local density in the presentation of the cell-binding sites along the fibers. It has been observed that cells crawl toward increased density of immobilized ligands, a process termed haptotaxis. Alternatively, or in addition, migration along aligned fibers may reflect an advantage in the mechanics of these fibers, which are likely stiffer along the axis of alignment. Indeed, cells have been demonstrated to choose stiffer substrata for migration in 2D studies, a process termed durotaxis. As an analogy, picture a rope strung between two points, attached at both ends and a bit taut. If you stand at one attachment point and pull the rope along the axis, you will not readily be able to deflect it, and instead the rope will feel stiff. In contrast, if you stand to the side and pull the rope cross-wise from the axis, less force is required to deflect the rope, and it will feel more compliant. Thus, one can envision that a cell crawling along aligned fibers has the advantage of a stiff surface against which to generate contraction forces as well as the presentation of ligands laid out before it in a locally dense and near linear manner.

The Role of Proteases in 3D Cell Migration and Invasion

Proteases Degrade the ECM

The ECM is frequently remodeled by the action of cellular proteases. Proteases that act on the ECM include cysteine proteases (cathepsins), serine proteases such as plasmin

and thrombin, and matrix metalloproteinases (MMPs). The MMPs require a conjugated Zn^{2+} ion and include collagenase (MT1-MMP), gelatinases (MMP2 and MMP9), and stromelysins (MMP-3, MMP-10, and MMP-11). MMPs are secreted as inactive zymogens and are activated upon proteolytic cleavage and inactivated by endogenous inhibitors, the most prominent being the tissue inhibitor of metalloproteinases (TIMPs) family (TIMP-1, -2, -3, and -4). Thus, the MMPs are controlled in part by the balance of activators and inhibitors.

While many of these proteases are secreted into the extracellular space, MT1-MMP is particularly interesting because it is a trans-membrane protein and thus is localized to the cell surface. MT1-MMP is found in specialized protrusions that are involved in matrix degradation, called invadopodia. While invadopodia are similar to lamellipodia and filopodia in their actin-based structure, they differ due to the presence of MT1-MMP. In addition, invadopodia can be identified by the presence of cortactin, an actin binding protein that accumulates in invadopodia and may help to localize signaling molecules or stabilize this structure.

There is currently much debate regarding the role of proteases in 3D migration and invasion. *In vivo*, it is thought that proteases are necessary for carcinoma cells to degrade the basement membrane, which is very dense despite being only 0.2 μm thick. Thus, protease activity may represent an obligate first step before carcinoma invasion can occur. Evidence from several labs suggests that once past the basement membrane,

cells can migrate through the collagen and fibronectin-rich stromal matrix in the absence of protease activity, especially if they have aligned collagen on which to migrate.

Proteases Are Required to Form the ECM

Proteases not only have a degradative function, they are also essential for the formation of the ECM. Collagens are secreted as pro-peptides that must have their N-terminal and C-terminal pro-peptides removed by protease activity. Bone morphogenetic protein-1 (BMP-1) is a protease essential for this process. Fibrin-based provisional matrices at the sites of wound healing are a result of thrombin cleavage of fibrinogen, which is soluble in the serum, into fibrin, which is insoluble and forms a clot to achieve hemostasis.

Proteolytic Fragments Are Functional

One of the other aspects to consider is the role of the proteolytic fragments of the ECM. Several ECM molecules, when cleaved, become active peptides that have function. In the lung, which is rich in elastin, elastase cleaves elastin into peptides that are chemotactic for infiltrating neutrophils. In tumors, cleavage of collagen XVIII forms a molecule called endostatin. Endostatin is released into the blood stream and inhibits angiogenesis in distant sites. There is an observation in cancer biology that removal of a primary tumor sometimes results in growth of the metastases. It is postulated that removal of the primary tumor removes the activity of proteases and the release of endostatin, which was controlling angiogenesis to the metastases.

Proteolysis Can Activate Signaling Molecules

In addition to cleaving the ECM outright, proteases can activate growth factors by releasing them from the ECM. For example, transforming growth factor- β (TGF β) is bound into the ECM in a latent form, and becomes active due to cleavage by MMP-9 or other serum proteases such as plasmin. Like the MMPs, plasmin is secreted as an inactive zymogen, plasminogen, and is only able to activate TGF β in a localized manner where it has been activated.

Finally, cell-surface receptors themselves can be activated by proteolytic cleavage. The protease activated receptor (PAR) family (PAR-1, -2, -3, and -4) is activated upon cleavage by proteases: thrombin cleavage of a small extracellular peptide from PAR-1, -3, and -4 will activate downstream intracellular signaling. MMPs are able to activate cell-surface receptors as well, including the epidermal growth factor (EGF) family of receptors. Thus, in considering cell migration within complex extracellular matrices, many levels of regulation and many sources of activation are at play.

Cell-Cell Adhesion

In tissues, cells are in close contact with their neighbors. In epithelial cells, these cell-cell contacts are maintained by E-cadherins: transmembrane receptors that bind homotypically to E-cadherins on neighboring cells (Figure 2). Similar

to focal contacts, E-cadherin receptors link cytoplasmic signaling molecules to the actin cytoskeleton in structures called adherens junctions. Signaling molecules that are involved in adherens junctions include β -catenin, α -catenin, p120 catenin, and Rho. E-Cadherins are important to signal appropriate cell proliferation, and to provide integrity to epithelial tissues. Similar receptors exist on most cells: N-cadherin and N-CAMs (cell-adhesion molecules) mediate neuronal cell-cell attachments, while PE-CAM mediates attachment of endothelial cells. Cadherins are important to maintain the differentiated state and tissue integrity and regulate processes such as cell proliferation, tissue-specific gene expression, and tissue architecture.

An important aspect in carcinoma migration is the loss of E-cadherin adhesion, which allows cells to escape some of the controls on their growth and movement. In fact, inhibition of E-cadherin is often enough to stimulate cell migration. Many growth factors that promote cell migration do so in part by their negative regulation of cell-cell adhesions. Some studies have suggested that the loss of cadherin-mediated adhesions tips the balance in favor of increased cell-matrix adhesion.

Cells also make use of cell-cell attachments to migrate. For example, during development of the nervous system, N-cadherin and N-CAMs facilitate a process called pathfinding, and allow growing neurites to extend along the neurite processes of their predecessors. This process may be co-opted by migrating tumor cells, which frequently start expressing N-cadherin when they lose E-cadherin. The increase in N-cadherin expression is often associated with an increased ability of carcinoma cells to migrate, and may allow carcinoma cells to track along nerves as part of their dissemination.

Collective and Individual Cell Migration

Cells can migrate either as individual cells, or collectively. When cells migrate collectively, many cells follow a lead cell and migrate as single-files, or as sheets of cells that maintain their cell-cell contact. Several aspects of cell migration that have been mentioned in this article regulate whether cells will migrate individually or collectively (Figure 8). Whether cells migrate individually or collectively is thought to be important in how effective they are at invading into tissues and in disseminating to other organs, and may have important implications for cancer metastasis in particular.

Protease activity is one such regulator. Experiments in which protease activity has been either inhibited or augmented demonstrate that cells without protease activity will migrate individually, while the ability to degrade the stromal matrix is needed for collective migration. This is thought to be a simple restraining function of the ECM: if it is not degraded, the cells must squeeze through the openings available, forcing the cells to behave in an amoeboid manner and migrate more individually. An aligned matrix can overcome the need for proteases. When cells track along aligned fibers, they are able to move in single files that do not require protease activity.

Integrin function and cell-cell adhesion regulate individual migration. Rather surprisingly, when β 1-integrin function is blocked, cells do not migrate as a collective group, but rather as individual cells. Regulation of classic cell-cell attachments is also involved. In carcinoma cells, loss of E-cadherin can

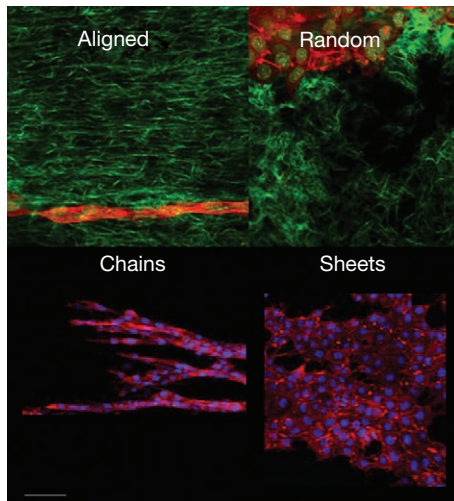


Figure 8 Collective migration in chains and sheets. Upper panels: Visualization of collagen fibers (green) by SHG and fluorescently labeled mammary epithelial cells (red). Collagen matrices are aligned (left) or random (right). Bottom panels: View of mammary epithelial cells migrating into an aligned (left) or random (right) collagen gel. Cells are stained to view f-actin (red) and nuclei (blue). Note that cells invade as chains along aligned collagen fibers. If such topographical cue is missing, as in a random collagen matrix, then cells invade as a sheet. Scale = 100 μ m. Image courtesy of Kristin Riching.

stimulate cell migration and releases the cells from the control of a polarized epithelium. However, often carcinoma cells will switch to other cell–cell adhesions such as N-cadherin, which may also regulate collective cell migration.

Summary

Cell migration through complex ECMs is regulated at several levels: intracellular signaling, cell-surface receptors, the composition and organization of the ECM, protease activity, cell–cell adhesions, and the presence of extracellular stimuli such as growth factors. While cell migration is necessary to set up complex tissues during development, and to maintain immune surveillance and repair wounds in adults, this process can go

awry during inflammation and various diseases. In particular, cancer metastasis is an area where cell invasion and migration through tissues is a key component. As our understanding of the complex events that occur during 3D cell migration advances, so too will our ability to regulate these processes.

See also: **Cell Architecture and Function:** Actin Organization; Actin-Related Proteins; Cadherin-Mediated Cell–Cell Adhesion; Cell Migration; Eukaryotic Chemotaxis; Focal Adhesions and Related Integrin Contacts; Metalloproteinases, Matrix; **Lipids Carbohydrates Membranes and Membrane Proteins:** Proteoglycans; **Signaling:** FAK Family; Integrin Signaling; Mitogen-Activated Protein Kinase Family; Ras Family; Small GTPases; Src Family of Protein Tyrosine Kinases.

Further Reading

- Condeelis JS, Wyckoff JB, Bailly M, et al. (2001) Lamellipodia in invasion. *Seminars in Cancer Biology* 11: 119–128.
- Cukierman E, Pankov R, and Yamada KM (2002) Cell interactions with three-dimensional matrices [Review]. *Current Opinion in Cell Biology* 14: 633–639.
- Friedl P and Wolf K (2003) Tumour-cell invasion and migration: Diversity and escape mechanisms. *Nature Reviews Cancer* 3: 362–374.
- Huttenlocher A, Sandborg RR, and Horwitz AF (1995) Adhesion in cell migration. *Current Opinion in Cell Biology* 7: 697–706.
- Hynes RO (2002) Integrins: Bidirectional, allosteric signaling machines. *Cell* 110: 673–687.
- Nelson CM, Vanduijn MM, Inman JL, Fletcher DA, and Bissell MJ (2006) Tissue geometry determines sites of mammary branching morphogenesis in organotypic cultures. *Science* 314: 298–300.
- Page-McCaw A, Ewald AJ, and Werb Z (2007) Matrix metalloproteinases and the regulation of tissue remodelling. *Nature Reviews Molecular Cell Biology* 8: 221–233.
- Paszek MJ and Weaver VM (2004) The tension mounts: Mechanics meets morphogenesis and malignancy. *Journal of Mammary Gland Biology and Neoplasia* 9: 325–342.
- Provenzano PP, Inman DR, Eliceiri KW, Trier SM, and Keely PJ (2008) Contact guidance mediated three-dimensional cell migration is regulated by Rho/ROCK-dependent matrix reorganization. *Biophys Journal* 95: 5374–5384.
- Wolf K, Wu YI, Liu Y, et al. (2007) Multi-step pericellular proteolysis controls the transition from individual to collective cancer cell invasion. *Nature Cell Biology* 9: 893–904.
- Wozniak MA, Modzelewska K, Kwong L, and Keely PJ (2004) Focal adhesion regulation of cell behavior. *Biochimica et Biophysica Acta* 1692: 103–119.

Centromeres

B A Sullivan, Duke University, Durham, NC, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Epigenetic Referring to or describing any heritable influence on chromosome or gene function that is not correlated with, or dependent on, a change in DNA sequence.

Heterochromatin Cytologically defined regions of the genome that contain repetitive DNA (satellite DNA, transposable elements). A defining feature of heterochromatin is the ability to silence gene expression.

Kinetochores The proteinaceous structure on each chromosome that is responsible for chromosomal attachment to and movement along spindle microtubules.

RNA interference A cellular defense mechanism for specific gene silencing that is induced by double-stranded RNAs that are processed into 21–23-nucleotide small interfering RNAs, causing degradation of homologous endogenous messenger RNA.

The centromere is a specialized mixture of DNA and proteins, and ensures chromosome inheritance and genome stability. As a chromosomal locus, the centromere is the minimal DNA or chromatin element that promotes formation of the proteinaceous kinetochore complex and coordinates chromosome movement in mitosis and meiosis. The centromere also synchronizes aspects of chromosome structure, such as heterochromatin formation, sister chromatid cohesion, and chromosome condensation. It is a multidomain locus, recruiting a variety of proteins with distinct functions.

Organization of the Centromere Region

Cytologically, centromeres have been defined by the visible primary constriction on metaphase chromosomes. This chromosomal locus is structurally complex, and contributes to various processes that ensure chromosome stability. The centromere region can be broadly classified into two major domains that encode kinetochore and heterochromatin functions. The kinetochore and heterochromatin are assembled independently, but each is equally important for complete centromere function and for ensuring chromosome and genome stability. The centromere/kinetochore domain comprises both DNA and proteins involved in chromatin assembly and structural aspects of the kinetochore. The heterochromatin domain is located adjacent to, or flanks, centromeric chromatin. Studies in various organisms have shown that heterochromatin is equally important for centromere function and chromosome inheritance as the kinetochore domain.

Kinetochore Domain

At metaphase, the kinetochore, a proteinaceous multidomain structure, is assembled on the outer surface of the centromere, promoting attachment of the chromosome to spindle microtubules and movement during anaphase.

The inner kinetochore

The inner kinetochore is the region most intimately associated with centromeric DNA and/or chromatin. Many structural

proteins that bind centromeric DNA or contribute to specialized chromatin structure are located here (Table 1). The inner kinetochore contains constitutive proteins that serve as the foundation for the kinetochore, often termed the prekinetochore. CENP-A, a centromere-specific histone H3 variant that replaces H3 in centromeric nucleosomes, is located here and serves as an initiatory signal for kinetochore assembly by recruiting other inner and outer kinetochore proteins.

The outer kinetochore

The outer kinetochore region contains primarily microtubule-associated and chromosomal motor proteins that are involved in chromosome congression to the midzone at metaphase or engage spindle microtubules and move chromosomes to spindle poles in anaphase. In addition, the outer kinetochore contains surveillance or checkpoint proteins that monitor kinetochore attachments to the spindle and regulate the transition from metaphase to anaphase.

The central domain

The region spanning the interior of the centromere and connects sister kinetochores comprises the interior central domain. Centromeric DNA-binding proteins are concentrated here, as well as cohesion and condensation proteins and transiently associated proteins (chromosomal passengers) that coordinate chromosome segregation and cytokinesis.

Heterochromatin Domain

Heterochromatin is a cytologically dense material that is typically found at centromeres and telomeres. It mostly consists of repetitive DNA sequences and is relatively gene poor. Its most notable property is its ability to silence euchromatic gene expression. Centromeres in yeast, fruit flies, and mammals are flanked by heterochromatin, indicating that its repetitive composition or dense chromatin structure may represent an important, conserved function in centromere structure and function. Heterochromatin assembly is linked to chromatin regulation, occurring in a pathway that initiates with methylation of histone H3 at amino acid residue lysine 9 in order to

Table 1 Homologous centromere region proteins in different species

Location	Function	<i>S. cerevisiae</i>	<i>S. pombe</i>	<i>C. elegans</i>	<i>D. melanogaster</i>	<i>H. sapiens</i>
Kinetochores	Centromere-specific histone	Cse4p	Cnp1	HCP-3	CID	CENP-A
	Centromeric chromatin architecture		Mis6 Mis12 Mal2 Sim4			CENP-I
	Inner plate: structure and DNA binding	Mif2p	Cnp3	HCP-4		CENP-C, CENP-G, CENP-H
	Outer plate: chromosome congression and movement	CBF1, CBF3		HCP-1,2	Cenpmeta, Cenpana, Zw10, Rod	CENP-E, CENP-F, ZW10, ROD
Heterochromatin	Histone H3 methyltransferase		Clr4		Su(var)3-9	SUVAR39H1
	Heterochromatin formation		Swi6Chp1Rik1		Su(var)2-5/HP1 HP2	HP1
	Sister chromatid cohesion	Scc1/Mcd1, Scc3Pds5	Rad21/Scc1 Mis4,6,12	SCC-1/COH-2 SCC-3 EVL-14/PDS-5	dRad21/Scc1	RAD21/SCC1

recruit heterochromatin proteins (HPs), such as HP1 (heterochromatin protein 1). Once heterochromatin is established, cohesion and condensation proteins accumulate between sister kinetochores and chromatids. Mutations in HPs or certain histone-modifying enzymes lead to chromosome mis-segregation and mitotic defects, indicating that heterochromatin contributes significantly to chromosome stability and segregation.

Specification of Centromere Identity and Function

The Centromere as DNA

How are the distinct kinetochore and heterochromatin domains assembled at centromeric regions? A well-debated question in centromere biology has been the role of primary DNA sequence in centromere identity and assembly. In a few organisms, specific DNA sequences are required for centromeric protein binding, while in other organisms, centromere-specific DNA sequences have not been found. In addition, epigenetic mechanisms often determine centromere identity. In short, a centromere is established, and it is consistently maintained at a genomic region, but the site of formation is not determined by the underlying sequence itself. Recent studies indicate that RNA also participates in centromere assembly.

Point centromeres

Centromeres in the budding yeast *Saccharomyces cerevisiae* are the best studied and understood. *S. cerevisiae* centromeres are encoded by three distinct DNA elements (CDE I, II, and III) within a 125-bp region (Figure 1). Two elements (CDE I and III) are absolutely conserved and required to recruit centromere and kinetochore proteins. The centromeric histone CENP-A (Cse4p) is recruited to CDE II, the centromeric element that varies in sequence, but not size, from chromosome to chromosome. The small size of budding yeast centromeres as well as the strict reliance on DNA-protein interactions for

assembly have resulted in detailed molecular maps and models of centromere assembly that are lacking in larger eukaryotes.

Regional centromeres

Centromeres in fission yeast, *Schizosaccharomyces pombe*, and in flies, worms, mammals, and plants are monocentric, forming on a limited, specific region of the chromosome (Figure 1). These centromeres are comprised of multiple subunits encoding kilobase- or megabase-sized genomic regions. Unlike in budding yeast, centromeric DNA sequences are not obviously conserved, and centromere identity and assembly depend largely on epigenetic factors.

Schizosaccharomyces pombe

The three centromeres in fission yeast consist of unconserved central core sequences flanked by inverted arrays of inner and outer repeats. Both the central core and portions of the inner repeats are required for establishment and maintenance of centromere function and for recruitment of CENP-A (Cnp1) and other centromeric proteins. The outer repeats are essential for assembly of centric heterochromatin through histone modifications and noncoding RNAs.

Multisubunit centromeres: flies, humans, and plants

Metazoan and plant centromeres are large, often ranging in size from a few hundred kb to 5 Mb. They primarily consist of thousands of copies of satellite DNA, often arranged in tandem (mammals and plants). In flies and plants, these repeats are often interspersed with transposable elements. In humans, *de novo* centromere assembly via artificial chromosomes has been most efficiently achieved when centromeric satellites (alpha-satellite DNA) are introduced into cells. However, not all human alpha-satellite DNAs form centromere *de novo*, indicating that other sequences or factors are required to assemble and maintain human centromeres.

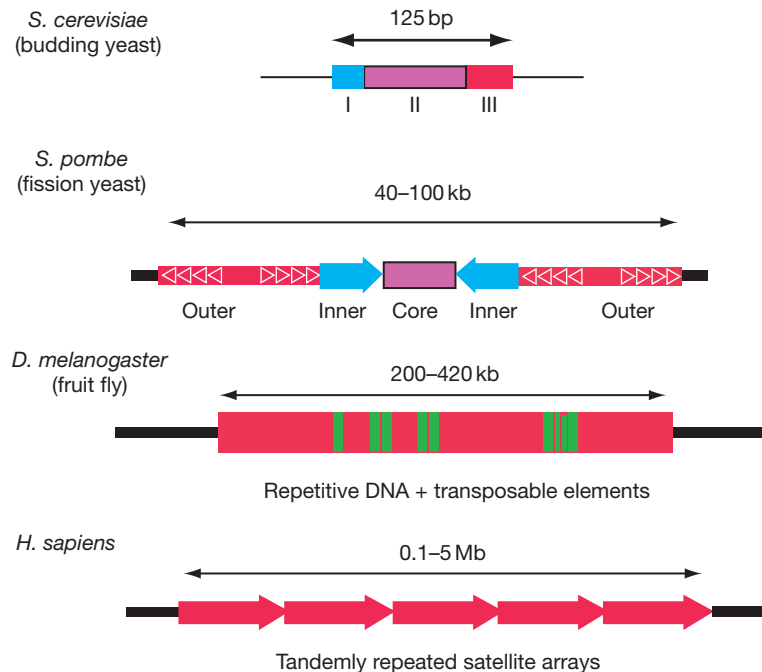


Figure 1 Centromeric structure in various eukaryotes. Schematic diagram of eukaryotic centromeres. The DNA sequence of centromeres differs between species, but the organization of the centromere and the presence of centromere proteins are conserved. Budding yeast (*S. cerevisiae*) centromeres are 125 bp and are composed of three distinct elements, two of which are conserved (I and III). *S. pombe* (fission yeast) centromeres contain a unique central core flanked by inverted inner and outer repeats. *Drosophila* centromeres extend for 200–420 kb and contain repetitive DNA (red boxes) that is interspersed with transposable elements (green lines). Human centromeres consist of tandemly repeated alpha-satellite DNA arranged into higher-order repeats that extend over several megabases.

Holocentric chromosomes

In the nematode *Caenorhabditis elegans*, crayfish, and some insects, the holocentric chromosomes assemble centromeres along their lengths. *C. elegans* centromeres are the best-studied holocentrics. Most of the known eukaryotic centromere and kinetochore proteins have been identified in worms and the order of assembly, with CENP-A (CeCenp-A or HCP-3) at the top of the assembly pathway, are conserved. Proteins are recruited into distinct foci at prophase, but by metaphase, these foci spread evenly into ribbons along the poleward face of the chromosome arms. The distribution of centromere and kinetochore proteins into inner and outer kinetochore regions is conserved in *C. elegans*.

The Centromere as Chromatin

Chromatin is an important regulator of gene expression and compartmentalization of the genome in interphase. Establishment of distinct chromatin domains (kinetochore and heterochromatin) is necessary for complete centromere function; thus, chromatin may be a more important determinant of centromere function than DNA sequence.

Epigenetic centromeres

The simple presence of centromeric DNA in a cell or a chromosome does not necessarily correlate with centromere function. Besides the obvious lack of sequence conservation at eukaryotic centromeres, additional evidence exists to argue the role of underlying DNA sequence in centromere assembly. First, at

normal mammalian centromeres, only a fraction of the arrays actually participates in centromere assembly. In humans, the kinetochore domain typically comprises one-half to two-thirds of the satellite DNA found at eukaryotic centromeres. The remainder of the satellite DNA is involved in forming heterochromatin and cohesion. Second, naturally occurring or engineered dicentric chromosomes that contain two regions of centromeric DNA are stably transmitted in flies and humans only after inactivation of one centromere. Finally, neocentromeres are formed in flies and humans on stable marker chromosomes that completely lack centromeric sequences. None of the neocentromeres share sequence homology and even those that are derived from similar genomic regions assemble centromere, kinetochore, and HPs on different sequences.

Specification by CENP-A

All centromeres are associated with the centromere-specific histone CENP-A. CENP-A is required to recruit all other centromere and kinetochore proteins, with the exception of HP1, making it a strong candidate for a protein that specifies and maintains the site of kinetochore assembly. Unlike replication-dependent histones, CENP-A is dispersed in a semiconservative manner, thus providing the scaffold on which new CENP-A can be loaded and the centromere can be maintained irrespective of the underlying DNA sequence. The organization of centromeric chromatin may further facilitate its maintenance and/or function. In two dimensions, centromeric chromatin in flies and humans contains interspersed subunits of CENP-A and H3 nucleosomes (Figure 2). CENP-A subunits coalesce

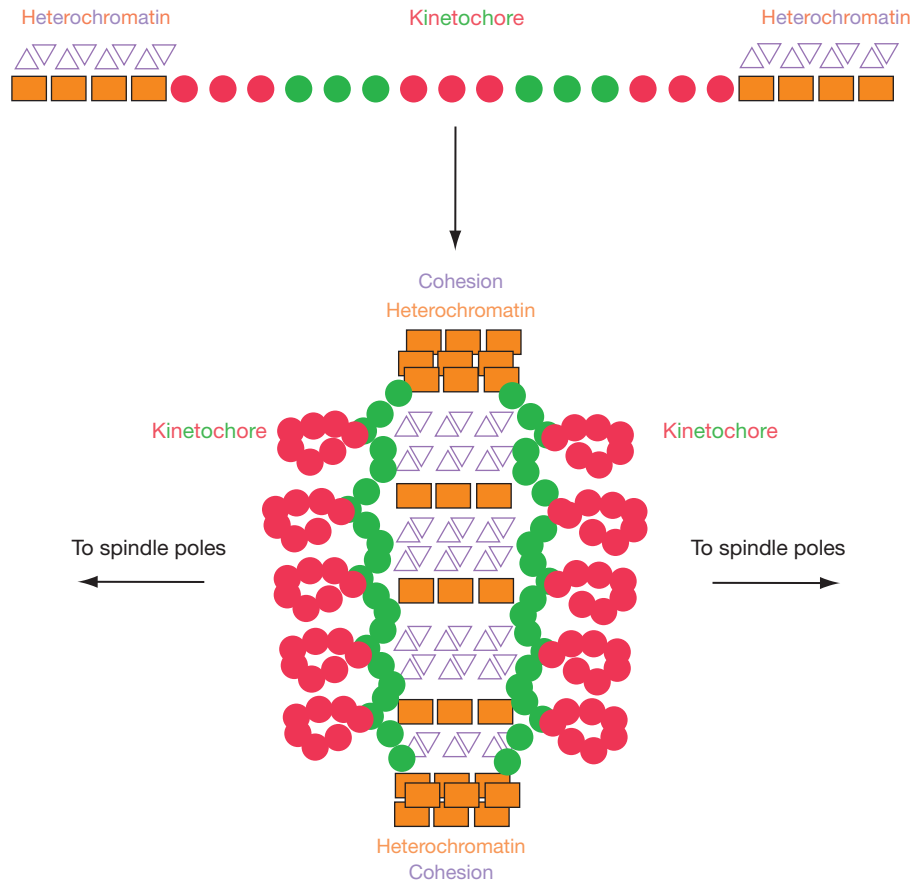


Figure 2 Unique organization of centromeric chromatin creates proper structure of the centromere region. In 2D, CENP-A (red circles) and H3 (green circles) nucleosomes within centromeric chromatin are interspersed, forming the foundation of the kinetochore domain, which is flanked by heterochromatic proteins (squares) and cohesion proteins (triangles). In metaphase (3D), the blocks of CENP-A and H3 nucleosomes may be arranged in a spiral, orienting CENP-A nucleosomes to the poleward face of the chromosome, presumably to bi-orient the chromosome and facilitate recruitment and interactions with other kinetochore proteins. H3 nucleosomes are sequestered to the region between sister kinetochores, where heterochromatin and cohesion proteins are recruited. Flanking heterochromatin that is assembled between sister kinetochores may be as important as CENP-A:H3 interspersions for ensuring that CENP-A opposes spindle poles.

to form a three-dimensional (3D) interface that promotes recruitment of additional kinetochore proteins, biorientation of chromatids to opposite spindle poles, and attachments to microtubules. Conversely, the H3 subdomains are oriented toward the interior of the chromosome to establish a platform for recruiting cohesion and condensation proteins. The unique interspersions of CENP-A and H3 nucleosomes is important for centromere function, because depletion of CENP-A by RNA interference (RNAi) or in CENP-A mutants alters the ratio of CENP-A:H3 and leads to chromosome segregation defects and mitotic arrest.

Incorporation of CENP-A into chromatin is temporally regulated by a unique chaperone. It is clear that CENP-A is a crucial component of centromeres and is required to recruit other proteins that are the structural and functional basis of the kinetochore. As such, it must be replaced each cell cycle, as it is semi-conservatively depleted with each round of DNA replication. When and how CENP-A is loaded specifically at centromere regions and how it is replenished or maintained at these sites have been long-standing questions. Recent evidence has

emerged indicating that CENP-A is loaded at a specific time during the cell cycle and by a specific protein. The timing of CENP-A incorporation occurs in late telophase/early G1 in humans and fission yeast and during metaphase/anaphase in *Drosophila*. In humans, the CENP-A chaperone, Holliday junction recognition protein (HJURP) associates with CENP-A and is recruited to centromeres at the same time that CENP-A is deposited in early G1. HJURP directly recognizes the CENP-A targeting domain (CATD), a motif that is required for both proper CENP-A targeting to the centromere and unique structural and functional properties of the CENP-A nucleosome. In budding yeast, Scm3 may actually chaperone CENP-A^{Cse4p} to the location of nucleosome assembly. It is functionally similar to HJURP, although there is less sequence similarity between the two protein families. HJURP or Scm3 proteins have not been identified in *Drosophila*. Instead, CAL1, a newly identified protein, and CENP-C, a constitutive component of the kinetochore that is present in all organisms, have been shown to be essential for loading newly synthesized CENP-A at centromeres during metaphase–anaphase. Although the protein players vary slightly

among organisms, these findings indicate that CENP-A incorporation at eukaryotic centromeres involves coordination of novel and common centromere proteins.

Current model proposes that existing, or old, CENP-A acts as a centromeric flag or beacon to help target new CENP-A. However, other features of centromeric chromatin, such as the presence of H3 nucleosomes interspersed with CENP-A, post-translational modifications of the N-terminal tails of CENP-A or H3 or chromatin environment, may be part of the signal. In human cells, CENP-N specifically recognizes CENP-A nucleosomes and not H3 nucleosome located nearby. It is thought to read and interpret the chromatin landscape at the centromere and communicate to assembly factors that more CENP-A is needed. At the same time, CENP-T and CENP-W, components of the DNA-proximal constitutive centromere-associated network (CCAN), associate with H3-containing nucleosomes within the centromere region. Studies in the fission yeast argue for a role of chromatin environment in centromere formation. In *S. pombe*, heterochromatin marked with H3K9 methylation by Clr4/Suv39 is crucial for recruiting CENP-A^{Cnp1} to neighboring central core sequences and establishing a functional centromere. Neocentromeres, or ectopic centromeres that form *de novo* at noncentromeric sites, are established at telomere and centromere-proximal regions in *S. pombe* and *Candida albicans*. Both of these regions are normally heterochromatic, suggesting that heterochromatin promotes centromere assembly, at least in fungi. It is unclear if heterochromatin has a similar positive impact on CENP-A deposition in larger eukaryotes, or if independent H3 and CENP-A assembly systems coordinate their actions to establish centromeric chromatin.

The Centromere as RNA

One of the most exciting discoveries in the past few years has been the involvement of double-stranded RNA (dsRNA) in specifying chromosomal, and particularly centromere, function. Although originally considered junk DNA, the arrays of repeats that surround active centromeres have been shown to be functionally significant. In particular, noncoding, antiparallel RNAs are transcribed from the outer centromeric repeats in *S. pombe*. These transcripts are typically unstable and are normally processed by RNAi machinery that mediates the formation of heterochromatin. Full-length double-stranded transcripts from the repeats are reduced into small interfering RNAs (siRNAs) by the protein Dicer, a member of the RNA-induced silencing complex (RISC). The siRNAs then initiate methylation of histone H3 at lysine 9 that subsequently recruits HPs. The siRNAs are only required for establishment, but not

maintenance of heterochromatin. However, RNAs appear to be important for centromere assembly as well. In fission yeast, pericentric heterochromatin and the RNAi components necessary to form it are required to establish CENP-A chromatin on neighboring sequences. RNAs produced from CENP-A regions have been identified in plants and are important for structural and functional aspects of CENP-C binding at centromeres. Transcription from an L1 retrotransposon within the CENP-A region of a human neocentromere has been shown to be important for CENP-A assembly and maintenance. Although transcripts originating from alpha-satellite DNA have been identified at endogenous human centromeres, it is unclear if they are required for CENP-A assembly. With recent findings in other organisms supporting a role for RNA in centromere assembly and function, it is tempting to speculate that kinetochore-specific transcripts from human centromeres might initiate *de novo* centromere assembly that is then maintained by retention of CENP-A and/or other proteins through many rounds of replication and division.

See also: [Cell Architecture and Function: Chromosome Organization and Structure, Overview](#); [Molecular Biology: Chromatin: Physical Organization; Regulation of Chromatin Dynamics](#).

Further Reading

- Buscaino A, Allshire R, and Pidoux A (2010) Building centromeres: Home sweet home or a nomadic existence? *Current Opinion in Genetics and Development* 20: 1–9.
- Choo KHA (1997) *The Centromere*. New York: Oxford University Press.
- Cleveland DW, Mao Y, and Sullivan KF (2003) Centromeres and kinetochores: From epigenetics to mitotic checkpoint signaling. *Cell* 112: 407–421.
- Dalal Y and Bui M (2010) Down the rabbit hole of centromere assembly and dynamics. *Current Opinion in Cell Biology* 22: 1–10.
- Dunleavy EM, Roche D, Tagami H, et al. (2009) HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell* 137: 485–497.
- Foltz DR, Jansen LET, Bailey AO, et al. (2009) Centromere-specific assembly of CENP-A nucleosomes is mediated by HJURP. *Cell* 137: 472–484.
- Karpen GH and Allshire RC (1997) The case for epigenetic effects of centromere identity and function. *Trends in Genetics* 13: 489–496.
- Martienssen RA (2003) Maintenance of heterochromatin by RNA interference of tandem repeats. *Nature Genetics* 35: 213–214.
- Mellone BG and Allshire RC (2003) Stretching it: Putting the CEN(P-A) in centromere. *Current Opinion in Genetics and Development* 13: 191–198.
- Sharp JA and Kaufman PD (2003) Chromatin proteins are determinants of centromere function. *Current Topics in Microbiology and Immunology* 274: 23–52.
- Sullivan BA, Blower MD, and Karpen GH (2001) Determining centromere identity: Cyclical stories and forking paths. *Nature Reviews Genetics* 2: 584–596.
- Sullivan KF (2001) A solid foundation: Functional specialization of centromeric chromatin. *Current Opinion in Genetics and Development* 11: 182–188.

c-fes Proto-Oncogene

T E Smithgall, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Thomas E. Smithgall and Robert I. Glazer, volume 1, pp. 377–382, © 2004, Elsevier Inc.

Glossary

Promoter The DNA sequence upstream to the transcription initiation site that directs expression of the gene.

STAT (signal transducers and activators of transcription) Transcription factors originally identified as mediating the transcriptional effects of interferon.

Structure and Regulation

The human *c-fes* gene is located on chromosome 15 at position 15q26.1 and is comprised of 19 exons, the first of which is noncoding. The *c-fes* locus encodes a 93-kDa protein-tyrosine kinase with several distinct structural regions (Figure 1). These include an N-terminal Fes-CIP4 homology-bin-amphiphysin-rvs domain (F-BAR) region made up of Fes-Cdc42-interacting protein 4 (CIP4) homology (FCH) and coiled-coil (CC) domains, an Src homology 2 (SH2) domain, and a C-terminal kinase domain. Absent are an Src homology 3 (SH3) domain and signal sequences for lipid modification, such as the N-terminal myristoylation site associated with c-Src, c-Abl, and other nonreceptor tyrosine kinases. Fes tyrosine kinase activity can be readily demonstrated *in vitro*. However, unlike its transforming viral counterparts, Fes kinase activity is tightly regulated in cells, even upon high-level overexpression. Both the CC and SH2 domains play important roles in kinase regulation.

F-BAR Region

The N-terminal region of Fes is defined by an N-terminal Fes/Fer/CIP4 homology (FCH) domain followed by at least two CC homology motifs. Together, these domains define an F-BAR region, a structural arrangement encoded by at least 21 other human genes. Proteins with F-BAR domains have been shown to regulate membrane curvature and contribute to cellular processes such as endocytosis, exocytosis, and motility. The presence of an F-BAR region in Fes (and its close homolog Fer) is unique among protein-tyrosine kinases.

The FCH domain was originally defined as a region of homology between Fes, Fer, and a binding partner for the Rho guanosine triphosphatase (GTPase), Cdc42 (CIP4). The *c-Fes* FCH domain has been linked to the regulation of microtubule dynamics, as it associates with soluble tubulin dimers and contributes to Fes-induced nucleation of microtubules both *in vitro* and in cells.

One unusual feature of Fes is its ability to form large homotypic oligomers, both *in vitro* and in cells. Computational analysis of the N-terminal region revealed the presence of at least two CC domains; subsequent biological studies demonstrated that the CC regions are essential for Fes oligomerization. In addition, deletion or mutagenesis of the more N-terminal CC domain strongly upregulates the tyrosine kinase and biological activities of Fes in cells, consistent with a crucial role for this domain in negative regulation. Recent

analysis of Fes oligomerization using a live-cell bimolecular fluorescence complementation approach showed that oligomerization is indeed dependent upon the CC domains, but is independent of kinase activity. Thus, the impact of the CCs on the kinase domain is more likely to involve conformational rearrangements within a preformed oligomeric Fes structure.

The N-terminal region of Fes may also contribute to the recruitment of signaling partners. One example is the breakpoint-cluster region protein (Bcr), better known in the context of the Bcr-Abl oncoprotein associated with chronic myelogenous leukemia (CML). Bcr is a multidomain signaling protein with an N-terminal Ser/Thr kinase domain, followed by a central guanine nucleotide exchange region and C-terminal GTPase-activating domain for small GTPases of the Rho family. Bcr is phosphorylated by Fes on a cluster of sites localized to its N-terminal kinase domain. Expression of this Bcr region suppresses Fes-induced neurite outgrowth in rat PC12 cells, suggesting that Bcr may couple Fes to Rho family small GTPases in neuronal cells. Fes also interacts with Bcr-Abl through the Bcr-derived portion of the protein, an interaction that may contribute to the observed suppression of Bcr-Abl oncogenic signaling by Fes.

Yeast two-hybrid screens using the Fes N-terminal CC domain as bait identified Kruppel-associated box-associated protein-1 (KAP-1) as a unique Fes substrate and binding partner. Like Fes, KAP-1 also has a CC domain, and heterologous CC interactions appear to drive the association of Fes with this transcriptional regulator. Fes:KAP-1 complexes and Fes-mediated KAP-1 tyrosine phosphorylation have been observed in myeloid cells, suggesting a possible link between Fes and the regulation of chromatin structure via KAP-1.

SH2 Domain

SH2 domains are signaling modules that mediate protein-protein interactions in tyrosine kinase signal transduction pathways. By binding with high affinity and specificity to short, tyrosine-phosphorylated target sequences, SH2 domains mediate the assembly of multiprotein complexes in response to tyrosine kinase activation. Although the role of SH2 domains in the recruitment of effector proteins to autophosphorylated growth factor receptors is well known, SH2 domains were first identified within the sequences of nonreceptor tyrosine kinases. In fact, the SH2 domain of the Fes-related oncoprotein v-Fps was among the first to be described, where it was shown to be required for full catalytic activity. Later work established that

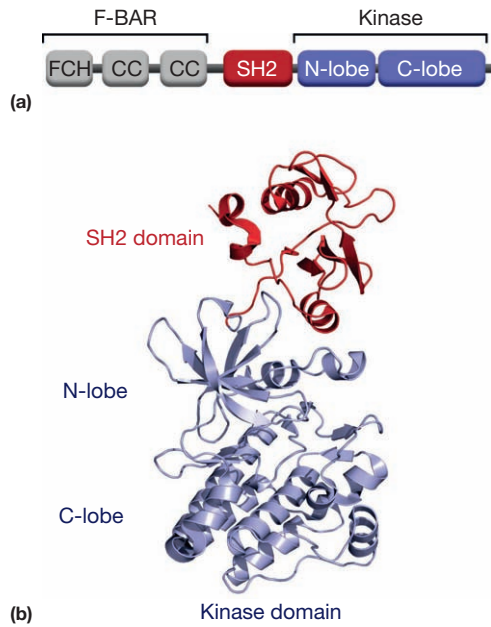


Figure 1 (a) Primary structure of the c-Fes protein-tyrosine kinase. Fes consists of an N-terminal F-BAR region, a central SH2 domain, and a C-terminal kinase domain. The F-BAR region includes the N-terminal FCH domain involved in interactions with the tubulin cytoskeleton as well as two coiled-coil (CC) oligomerization domains. (b) X-ray crystal structure of the c-Fes SH2-kinase unit. Note the juxtaposition of the SH2 domain (red) and the N-lobe of the kinase domain (blue), which serves to stabilize the active conformation of the kinase (see text). This model was produced using PyMol and PDB coordinates 3BKB.

v-Fps-induced phosphorylation of transformation-related substrates was also dependent upon this noncatalytic domain. More recent studies revealed a similar dual role for the human Fes SH2 domain in the regulation of kinase activity and biological function. Deletion of the SH2 domain greatly reduces Fes kinase activity in terms of both substrate phosphorylation and autophosphorylation *in vitro*. Recent structural analysis of the Fes SH2:kinase unit provides a mechanistic basis for the positive influence of the SH2 domain on kinase activity. In the crystal structure, the SH2 domain makes direct contacts with the small lobe of the kinase domain, holding the critical α C regulatory helix in an active conformation (Figure 1). Interestingly, ligand binding to the SH2 domain stabilizes the SH2:kinase domain interaction, suggesting that SH2-based substrate recruitment and kinase activation are coupled. Given that the N-terminal CC domain is involved in the repression of kinase activity, it is reasonable to speculate that the N-terminal region may allosterically repress SH2:kinase domain interaction as part of the regulatory mechanism. A structure of the full-length, inactive form of Fes is required to fully appreciate the unique mechanism responsible for negative regulation of its kinase activity.

Kinase Domain

The Fes catalytic domain is localized to the C-terminal region of the protein and exhibits structural features typical of other protein-tyrosine kinases. These include a conserved lysine in the adenosine triphosphate (ATP)-binding pocket (Lys590)

and a major site of tyrosine autophosphorylation (Tyr713) which localizes to the activation loop. Mutagenesis of Tyr713 to Phe greatly reduces catalytic activity both *in vitro* and *in vivo*, suggesting that autophosphorylation is required for kinase activation. Substitution of Lys590 with Glu or Arg completely abolishes kinase activity.

Biological Functions

Myeloid Differentiation

Early studies describing Fes expression patterns noted a striking correlation with terminal differentiation of myeloid hematopoietic cells. Fes expression was absent in myeloid leukemia cell lines resistant to differentiation inducers or selected for differentiation resistance, while expression increased following induction of differentiation with a variety of agents. Chicken bone marrow cells infected with Fujinami sarcoma virus, which carries the *fes*-related oncogene, *v-fps*, undergo macrophage differentiation in the absence of macrophage colony-stimulating factor. Similarly, expression of active c-Fes mutants is sufficient to induce differentiation of myeloid leukemia cells along the macrophage lineage. Consistent with these observations, Fes activation is linked to a number of hematopoietic cytokines and growth factors, including interleukin-3 (IL-3), IL-4, granulocyte-macrophage colony-stimulating factor (GM-CSF), erythropoietin, and Kit ligand (stem cell factor). Upstream regulators of Fes activity also include Fc-linked immune receptors found on macrophages and other cells of innate immunity.

Evidence supporting a requirement for Fes in myeloid differentiation comes from protein knockdown experiments. Suppression of Fes in HL-60 promyelocytic leukemia cells using antisense oligonucleotides blocked myeloid differentiation in response to phorbol esters and other chemical inducers. However, targeted disruption of Fes in mice has yielded contradictory findings with respect to its role in hematopoiesis. Homozygous replacement of the murine *fes* alleles with kinase-defective mutants did not show any remarkable effect on hematopoiesis, although signal transducers and activators of transcription (STAT)3 and STAT5A phosphorylation in response to GM-CSF was markedly reduced. By contrast, homozygous deletion of Fes led to a runted appearance, abnormal myeloid proliferation, but increased STAT3 activation in response to GM-CSF and IL-6. One possible explanation for these apparently contradictory results may involve competition of kinase-dead Fes with Jak2 for STAT3, resulting in a lower degree of STAT3 activation. By contrast, the complete absence of Fes protein may result in less competition with Jak2 for STAT3, resulting in enhanced STAT3 phosphorylation.

Angiogenesis

Other evidence supports a role for Fes in the development of the vascular endothelium. Expression of Fes in transgenic mice under the control of a nonspecific SV40 promoter-induced hypervascularity which progressed to multifocal hemangiomas. The Fes transgene used in these experiments was modified to encode an N-terminal myristoylation sequence on its N-terminus; this modification leads to addition of myristate, membrane targeting, and mild activation of kinase activity. Expression of endogenous Fes was also demonstrated in primary human vascular endothelial cells at levels comparable

to those detected in myeloid cells. Consistent with these findings are the observations that Fes is activated by the angiogenic growth factors fibroblast growth factor-2 and angio-pietin-2 in cultured capillary endothelial cells, and that over-expression of Fes is sufficient to induce formation of tube-like structures and chemotaxis in this cell type.

Is Fes an Oncogene or a Tumor Suppressor?

The observations that Fes may contribute to angiogenesis raises the possibility that it may contribute to neo-vascularization, which is essential for solid tumor growth. In this context, Fes may indirectly support tumor formation. Fes may also contribute to oncogenesis by coupling to c-Kit, the receptor tyrosine kinase required for stem cell factor signaling. Activating mutations have been reported within the c-Kit juxtamembrane region, which lead to ligand-independent kinase activation and drug resistance in gastrointestinal stromal tumors and other forms of cancer. Fes has been implicated as a key downstream signaling partner for mutationally activated Kit, suggesting that selective inhibitors of Fes may have antitumor effects in certain situations. Along these lines, small interfering RNA knockdown of Fes expression has been found to suppress the growth of renal carcinoma cell lines, implicating Fes as an oncogenic factor in this cell type as well.

By contrast, Fes-induced differentiation may overcome signals for transformation in certain forms of leukemia. An early demonstration of this phenomenon involved transfer of the *c-fes* gene into the human CML cell line K-562, resulting in growth suppression and differentiation to macrophage-like cells. K-562 cells were established from a CML patient in blast crisis and exhibit no detectable Fes expression; later work demonstrated hypermethylation of the *c-fes* promoter in this cell line was responsible for gene silencing. Because they are CML-derived, K-562 cells exhibit the Philadelphia chromosome translocation and express p210 Bcr-Abl, the oncogenic tyrosine kinase responsible for CML pathogenesis. Thus, the Fes signal for differentiation is dominant to the Bcr-Abl signal for growth and survival. Fes can also suppress Bcr-Abl-induced transformation of murine myeloid leukemia cells to cytokine independence, providing further support for the idea that Fes may delay the progression of CML.

More recent studies suggest an unexpected role for c-Fes as a tumor suppressor in cancers of epithelial origin. Nucleotide sequence analysis of the tyrosine kinome of 182 colorectal cancer cell lines revealed that *c-fes* was one of only seven genes exhibiting consistent cancer-associated kinase domain mutations. Although these mutations were initially predicted to be activating and contribute to tumorigenesis, subsequent studies established that these mutations rendered Fes either catalytically inactive or had no effect on kinase activity. Subsequent work showed that Fes is strongly expressed in normal colonic epithelial cells from both mice and humans. By contrast, Fes expression was reduced or absent in more than two-thirds of primary colon tumor specimens. Similarly, Fes protein expression was significantly reduced or absent in colorectal cancer cell lines. Re-expression of Fes from a retroviral vector suppressed soft-agar colony formation by two colorectal cancer cell lines where endogenous Fes expression is silenced (HT-29 and HCT 116). Taken together, these results suggest

that loss of c-Fes expression is a common finding in colorectal cancer, an observation that fits with a tumor-suppressor function for c-Fes in this tumor site. Indeed, promoter methylation has been demonstrated to play an important role in suppression of Fes expression in both colorectal cancer cell lines and primary tumors. Along similar lines, breast tumor onset has been observed to occur more rapidly in mice targeted with either null or kinase-inactivating *c-fes* mutations, and introduction of a wild-type *c-fes* transgene rescued the kinetics of tumor onset in the *c-fes* null mice.

Downstream Signaling Pathways

Small G Proteins

Ras and other small GTPases are critical signaling intermediates for most protein-tyrosine kinases, and Fes is no exception. Activation of Ras and the related small GTPases, Rac and Cdc42, are required for fibroblast transformation by Fes oncogenes. Dominant-negative mutants of all three small G proteins inhibit fibroblast colony-forming activity by Myr-Fes and v-Fps, and transformation correlates with constitutive activation of both the Erk and Jnk pathways downstream. Interestingly, sustained Erk activation has also been linked to myeloid differentiation of K-562 cells, suggesting that Fes-induced differentiation in these cells may be due in part to activation of the Ras/Erk pathway. Thus, the ultimate biological response (e.g., differentiation vs. transformation) may be dependent upon the cellular context in which the kinase is active.

The activation and termination of small GTPase signaling are regulated by protein factors (guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs), respectively), which link small GTPases with upstream tyrosine kinases, including Fes and its transforming viral homologs. Transformation of fibroblasts by v-Fps has been shown to induce tyrosine phosphorylation of p120 Ras GAP, and similar results have been observed *in vitro* with purified Fes and GAP proteins. Fibroblast transformation by v-Fps also correlates with tyrosine phosphorylation of Shc, an adaptor protein that links tyrosine kinases to the Ras GEF complex, Grb-2/Sos. As mentioned above, endogenous Bcr is also tyrosine-phosphorylated in fibroblasts transformed by v-Fps. Tyrosine phosphorylation of Bcr in v-Fps-transformed cells induces its association with Grb-2/Sos via the Grb-2 SH2 domain. Like Shc, Bcr may serve as an intermediate between Fes and the activation of Ras downstream. In addition, tyrosine phosphorylation of Bcr may modulate either its GEF or GAP activities toward Rac or other Rho-family GTPases.

STAT Signaling

STATs are transcription factors with SH2 domains that are activated by tyrosine phosphorylation, often in response to cytokine and growth factor stimulation. Although kinases of the Jak family are often associated with STAT activation, a growing body of evidence indicates that Fes and other tyrosine kinases also induce STAT activation. For example, STAT3 DNA-binding activity is strongly activated following co-expression with Fes in Sf-9 insect cells, which lack endogenous Jak kinases. In addition, macrophages from mice with targeted inactivating

mutations in both *fes* alleles show blunted STAT3 and STAT5 activation in response to GM-CSF. STAT3 has been implicated in myeloid differentiation, suggesting that Fes may influence differentiation-related gene expression via a STAT3-dependent pathway.

PI3K Pathway

PI3K phosphorylates phosphoinositol lipids at the 3'-OH position, generating second messengers such as phosphatidylinositol-3,4,5-trisphosphate (PIP3). PI3K is a heterodimer consisting of 85-kDa regulatory (p85) and 110-kDa catalytic (p110) subunits, and interacts with tyrosine phosphorylated target proteins through SH2 domains found in the p85 subunit. Several reports have linked Fes to the activation of PI3K signaling. Early work demonstrated elevated PI3K activity in chick embryo fibroblasts transformed by v-Fps as well as other transforming tyrosine kinases. More recently, IL-4 was shown to induce the association of endogenous Fes with PI3K in IL-2-dependent T-cell lines. The SH2 domain of p85 forms a stable complex with Fes, suggesting a possible mechanism for factor-induced complex formation. The PI3K pathway is often associated with survival signaling, and activation of this pathway by Fes may contribute to macrophage survival. Activation of PI3K is also essential to Fes-induced neurite extension in PC12 cells, where it may couple Fes to Rac activation.

Other Substrates

Many of the substrate proteins reported for Fes have been identified using fibroblast transformation as a model system. However, overexpression of Fes in a macrophage cell line, which represents a physiological cellular context, led to the phosphorylation of proteins related to integrin-mediated responses including adhesion and cell-cell contact. Interestingly, phosphorylation of Shc and other proteins associated with mitogenic responses were not observed. A role for Fes in cell adhesion is further suggested by the interaction of the Fes SH2 domain with cortactin, which is also phosphorylated by Fes. These results suggest a mechanism for the localization of Fes with focal adhesions, which has been observed in fibroblasts and in vascular endothelial cells.

Transcriptional Regulation of Fes

Transgenic mice expressing a 13-kb human genomic *c-fes* clone faithfully reproduced the myeloid expression pattern of the endogenous gene, suggesting that the relatively compact 5' untranslated region (UTR) contains a tissue-specific promoter. Myeloid-specific expression has been attributed to a tissue-specific repressor acting at promoter region +441 to +454 (relative to the transcription start site), a Fes-specific transcription factor (*c-Fes* expression factor, FEF) binding to a *cis*-acting element at -9 to -4 and to transcription factors Sp1, PU.1, and Spi-1 binding to elements within the -425 to +75 promoter sequence. Although these studies account for the myeloid-selective expression of Fes, they do not address Fes expression in other tissues, such as vascular endothelial and gut

epithelial cells. In addition to myeloid cells, Fes gene upregulation occurs during neuronal differentiation and accelerates neurite outgrowth in PC12 cells. Whether alternative *cis*-acting elements control Fes expression in different cell types will require further investigation.

Conclusions

The *c-fes* proto-oncogene encodes a unique cytoplasmic protein-tyrosine kinase involved in myelopoiesis, angiogenesis, and CML progression. Fes is also expressed in mammary and gut epithelium, where it may serve as a tumor suppressor. Although the sequence, structure, and position of the SH2 and kinase domains found in Fes share close homology with other nonreceptor tyrosine kinases, the N-terminal F-BAR region is structurally distinct. The presence of CC oligomerization domains within this region suggests a mode of kinase regulation and substrate recruitment unique to this tyrosine kinase family. Fes is associated with terminal differentiation of myeloid hematopoietic cells, as opposed to other tyrosine kinases that drive proliferation and survival. Work implicating Fes in the regulation of vascular endothelial cell migration suggests a role in tumor angiogenesis. Fes may also regulate cell survival and apoptosis during mammary gland differentiation, as well as neurite outgrowth. Loss of Fes function and expression in colorectal cancer suggests a key role for this unique kinase in the maintenance of epithelial tissue architecture as well.

See also: [Signaling: Fibroblast Growth Factor Receptors and Cancer-Associated Perturbations](#); [Hematopoietin Receptors](#); [Inositol Phosphate Kinases and Phosphatases](#).

Further Reading

- Bardelli A, Parsons DW, Silliman N, et al. (2003) Mutational analysis of the tyrosine kinome in colorectal cancers. *Science* 300: 949.
- Cheng HY, Rogers JA, Dunham NA, and Smithgall TE (1999) Regulation of c-Fes tyrosine kinase and biological activities by N-terminal coiled-coil oligomerization domains. *Molecular and Cellular Biology* 19: 8335–8343.
- Delfino FJ, Stevenson HM, and Smithgall TE (2006) A growth-suppressive function for the c-Fes protein-tyrosine kinase in colorectal cancer. *Journal of Biological Chemistry* 281: 8829–8835.
- Filippakopoulos P, Kofler M, Hantschel O, et al. (2008) Structural coupling of SH2-kinase domains links Fes and Abl substrate recognition and kinase activation. *Cell* 134: 793–803.
- Filippakopoulos P, Muller S, and Knapp S (2009) SH2 domains: Modulators of nonreceptor tyrosine kinase activity. *Current Opinion in Structural Biology* 19: 643–649.
- Greer P, Maltby V, Rossant J, Bernstein A, and Pawson T (1990) Myeloid expression of the human *c-fps/fes* proto-oncogene in transgenic mice. *Molecular and Cellular Biology* 10: 2521–2527.
- Greer PA (2002) Closing in on the biological functions of Fps/Fes and Fer. *Nature Reviews. Molecular Cell Biology* 3: 278–289.
- Kanda S, Miyata Y, Kanetake H, and Smithgall TE (2007) Non-receptor protein-tyrosine kinases as molecular targets for antiangiogenic therapy. *International Journal of Molecular Medicine* 20: 113–121.
- Laurent CE, Delfino FJ, Cheng HY, and Smithgall TE (2004) The human c-Fes tyrosine kinase binds tubulin and microtubules through separate domains and promotes microtubule assembly. *Molecular and Cellular Biology* 24: 9351–9358.

- Pawson T (1988) Non-catalytic domains of cytoplasmic protein-tyrosine kinases: Regulatory elements in signal transduction. *Oncogene* 3: 491–495.
- Sangrar W, Zirgibl RA, Gao Y, et al. (2005) An identity crisis for fps/fes: Oncogene or tumor suppressor? *Cancer Research* 65: 3518–3522.
- Shaffer JM, Hellwig S, and Smithgall TE (2009) Bimolecular fluorescence complementation demonstrates that the c-Fes protein-tyrosine kinase forms constitutive oligomers in living cells. *Biochemistry* 48: 4780–4788.
- Shaffer JM and Smithgall TE (2009) Promoter methylation blocks FES protein-tyrosine kinase gene expression in colorectal cancer. *Genes Chromosomes Cancer* 48: 272–284.
- Smithgall TE, Rogers JA, and Peters KL (1998) The c-Fes family of protein-tyrosine kinases. *Critical Reviews in Oncogenesis* 9: 43–62.
- Yu G, Smithgall TE, and Glazer RI (1989) K562 leukemia cells transfected with the human *c-fes* gene acquire the ability to undergo myeloid differentiation. *Journal of Biological Chemistry* 264: 10276–10281.

Chaperonins

A Bracher and F U Hartl, Max Planck Institute of Biochemistry, Martinsried, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

CCT Chaperonin containing Tcp-1, eukaryotic group II chaperonin.

Chaperonin Nanocage-forming molecular chaperone.

Cpn10 Chaperonin of 10 kDa, chloroplast GroES homolog.

Cpn60 Chaperonin of 60 kDa, chloroplast group I chaperonin.

GroEL Growth E locus, large gene, bacterial group I chaperonin.

GroES Growth E locus, small gene, bacterial group I chaperonin cofactor.

Hsp10 Heat shock protein of 10 kDa, mitochondrial GroES homolog.

Hsp60 Heat shock protein of 60 kDa, mitochondrial group I chaperonin.

Hsp70 Heat shock protein of 70 kDa, a major heat shock protein.

Htt Huntingtin, a human neuronal poly-glutamine protein.

Molecular chaperone Cellular protein folding catalyst/mediator or assembly factor.

PhLP Phosducin-like protein, a co-chaperone of group II chaperonins.

Prefoldin/GimC Co-chaperone of group II chaperonins.

RuBisCO Ribulose biphosphate carboxylase/oxygenase.

TIM barrel Triose phosphate isomerase fold, an eight-stranded β -barrel.

Thermosome Archaeal group II chaperonin.

TRiC Tcp-1-containing ring complex, eukaryotic group II chaperonin.

VHL protein Van Hippel Lindau protein, a human Trp/CCT substrate.

WD-40 repeat Forty-Residue repeat with a characteristic Trp-Asp motif; forms one blade of a β -propeller structure.

Introduction

The folding and assembly of newly synthesized proteins is a process of fundamental biological significance. *In vitro*, many proteins can fold spontaneously, solely based on the steric information contained in their amino acid sequences. However, under the highly crowded conditions prevailing in the cell, the efficiency of protein folding is often limited by aggregation, that is, the association of non-native states that expose hydrophobic amino acid residues and regions of unstructured polypeptide backbone to the solvent. Molecular chaperones, themselves proteins, have evolved to prevent aggregation and promote proper folding. They act by recognizing and shielding the non-native states of newly synthesized proteins and of pre-existent proteins that have unfolded under stress conditions, such as heat stress. Several classes of chaperones act as adenosine triphosphate (ATP)-driven molecular machines, including the members of the Hsp70 and Hsp90 families, the AAA+ proteins, and the chaperonins. In the following, the structure and function of the chaperonins will be discussed in greater detail.

Architecture of Group I and Group II Chaperonins

Chaperonins are found in all domains of life, in archaea, eubacteria, and eukaryotes, forming a group of evolutionary conserved proteins consisting of subunits of ~55-kDa molecular weight. These proteins form 800–1000-kDa double-ring complexes with seven to nine subunits per ring. Each subunit consists of an equatorial, an intermediate, and an apical domain. Both N- and C-termini of the sequence are located in the equatorial domain and thus the intermediate domain is inserted into a loop segment of the equatorial domain, while the apical

domain represents an insertion into the intermediate domain. The ATP-binding pocket is located in the equatorial domain, which has a unique ATPase fold only found in chaperonins. The equatorial domains also form the interface between the two rings. The intermediate domains are positioned on top of the equatorial domains, connecting them with the apical domains which form the pore-like opening of the complex (Figure 1).

Two groups of chaperonin can be distinguished: group I chaperonins and group II chaperonins. Group I chaperonins are found in eubacteria and the organelles of eubacterial origin, mitochondria and chloroplasts. These chaperonins are named growth E locus, large gene, bacterial group I chaperonin (GroEL), heat shock protein of 60 kDa (Hsp60), and chaperonin of 60 kDa (Cpn60), respectively. All group I chaperonins consist of seven-membered rings. GroEL and Hsp60 rings are homomeric, whereas Cpn60 contains two types of homologous subunits, designated α and β . Apparently, the eukaryotic group I chaperonins are part of the prokaryotic protein folding machinery that has been maintained during evolution from the endosymbiont precursors of mitochondria and chloroplasts, proteobacteria, and cyanobacteria, respectively. Catalysis of folding by group I chaperonins strongly depends on a heptameric cofactor complex named growth E locus, small gene, bacterial group I chaperonin cofactor (GroES) in bacteria (Hsp10 and Cpn10 in mitochondria and chloroplasts, respectively). Structural analysis revealed that binding of GroES to a GroEL ring creates a large internal cavity that provides a secluded folding chamber for substrate proteins (Figure 1).

By contrast, the group II chaperonins in the cytosol of archaea and eukaryotes do not require such a cofactor but instead carry α -helical extensions on their apical domains. The crystal structure of the archaeal chaperonin, the so-called thermosome, showed that these extensions can form an in-built,

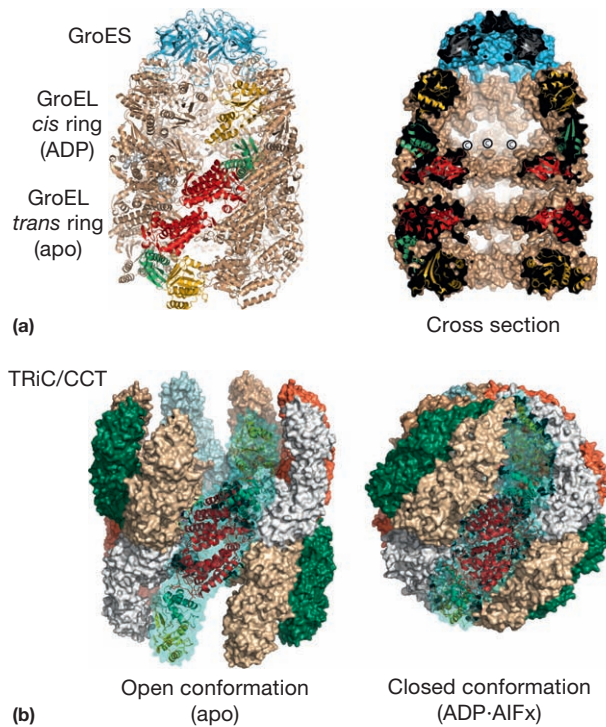


Figure 1 Structures of (a) group I and (b) group II chaperonins. On top, the structure of the GroEL₁₄ ADP₇-GroES₇ complex is shown (Protein Data Bank (PDB): 1PF9). The lower, so-called *trans*-GroEL₇ ring has no nucleotide bound and the subunits expose the substrate binding cleft toward the ring opening. The upper GroEL₇ ring has ADP bound in the crystal structure. The GroES₇ heptamer sits on top (the *cis* ring), indicated in blue. In two GroES subunits, the equatorial, intermediate, and apical domains are indicated in red, green, and blue, respectively. On the right, a vertical cross section through the complex is shown, highlighting the protein-folding chamber in the *cis* ring. The Gly-Gly-Met sequence repeats at the C-termini (indicated) are disordered in the crystal structure and thus not shown. These flexible tails presumably point into the cavity. Below, electron microscopy reconstructions of the thermosome from *Methanococcus maripaludis* are shown (PDB files 3IYE and 3IYF). The corresponding data for TRiC/CCT are very similar, but at lower resolution. The subunits are color-coded to show the subunit arrangement in TRiC/CCT. Two pairs of the subunits make homomeric contacts across the ring-ring interface. For one of these subunits the domain structure is shown using the same color code as for GroEL. Note that the chaperonin subunits have a staggered arrangement in group I chaperonins but are directly opposed in group II chaperonins. The symmetrical ring arrangement in the group II chaperonin structures is presumably an experimental artifact caused by either the complete absence or high abundance of nucleotide. Kinetic data suggest that these chaperonins also assume asymmetrical conformations with one ring open and the other ring closed at physiological nucleotide levels.

iris-like lid above the cavity (Figure 1). Most group II chaperonins form eight-membered rings but some archaeal chaperonins consist of nine-membered rings. Most group II chaperonins are heterooligomeric: thermosomes were reported to contain up to three different subunit types; the eukaryotic cytosolic chaperonin TRiC/CCT comprises eight evolutionary conserved, paralogous subunits.

Besides the group I and group II chaperonins, several chaperonin homologs appear to adopt alternative assemblies

other than typical chaperonin cages. For example, one of the two chaperonins of *Mycobacterium tuberculosis* is dimeric in solution. Metazoans contain three chaperonin homologs, Bbs6, Bbs10, and Bbs12, which are implicated in humans in the inherited disease, Bardet-Biedl syndrome. Such unusual chaperonin subunits might become incorporated into the ring complexes of canonical chaperonins, adapting them to specific substrate proteins.

Bacterial GroEL – Structure and Mechanism

GroEL of *Escherichia coli* is by far the best-studied chaperonin. A plethora of structural and functional data is available for this complex. The function of GroEL is based on an intricate ATP- and GroES-dependent conformational cycle, which involves both chaperonin rings (Figure 2). In the nucleotide-free state, the apical domains of GroEL expose a hydrophobic cleft suited for the binding of short hydrophobic peptides in an extended or amphiphilic α -helical conformation. This state is characterized by a collapsed ring conformation with only small internal cavities. ATP binding primes the rings for a large conformational change in the subunits, specifically an upward and clockwise turning motion of the apical domains, which is completed and stabilized upon binding of the GroES heptamer (Figure 1). The ATP hydrolysis reaction exhibits positive cooperativity within individual rings, that is, hydrolysis of ATP bound to one subunit triggers ATP hydrolysis in neighboring subunits, and negative cooperativity between rings, such that only one ring is active at a time. In the absence of bound substrate, GroES is a potent inhibitor of ATP hydrolysis by GroEL, preventing nonproductive cycling. The GroES-binding site partially overlaps with the substrate-binding site; thus, the binding of GroES promotes the displacement of the substrate into the newly formed interior cavity. This cavity is large enough to accommodate proteins of ~60 kDa and thus appears to be sufficient for the great majority of the identified GroEL substrates. Some larger proteins may utilize GroEL for aggregation prevention but do not undergo global encapsulation for folding. Following GroES binding, the ATP in the seven GroES-interacting subunits of GroEL (the so-called *cis* ring) is hydrolyzed (Figure 2). During this period of ~10 s, the enclosed substrate protein is free to fold in the *cis* cavity. Upon completion of ATP hydrolysis in the *cis* ring, the opposite (*trans*) ring becomes primed for a new cycle of substrate and nucleotide binding. Specifically, ATP binding to the *trans* ring elicits a negative allosteric signal that causes the dissociation of the ligands from the *cis* ring, 7ADP and GroES (ADP, adenosine diphosphate), thereby allowing substrate to exit the folding cage. Incompletely folded protein molecules will rebinding rapidly, either to the same or to another GroEL ring, to undergo a new attempt at folding in the chaperonin cage (Figure 2). Some proteins have been observed to require multiple interaction cycles with GroEL for completion of folding.

The crystal structure of a GroEL₁₄ ADP₇ GroES₇ complex (Figure 1) showed that the interior wall of the folding cage has mainly polar character and contains numerous charged side chains; the hydrophobic patches on the apical domains are buried at the interface to GroES. Thus, the folding substrate may experience conditions of infinite dilution, provided that

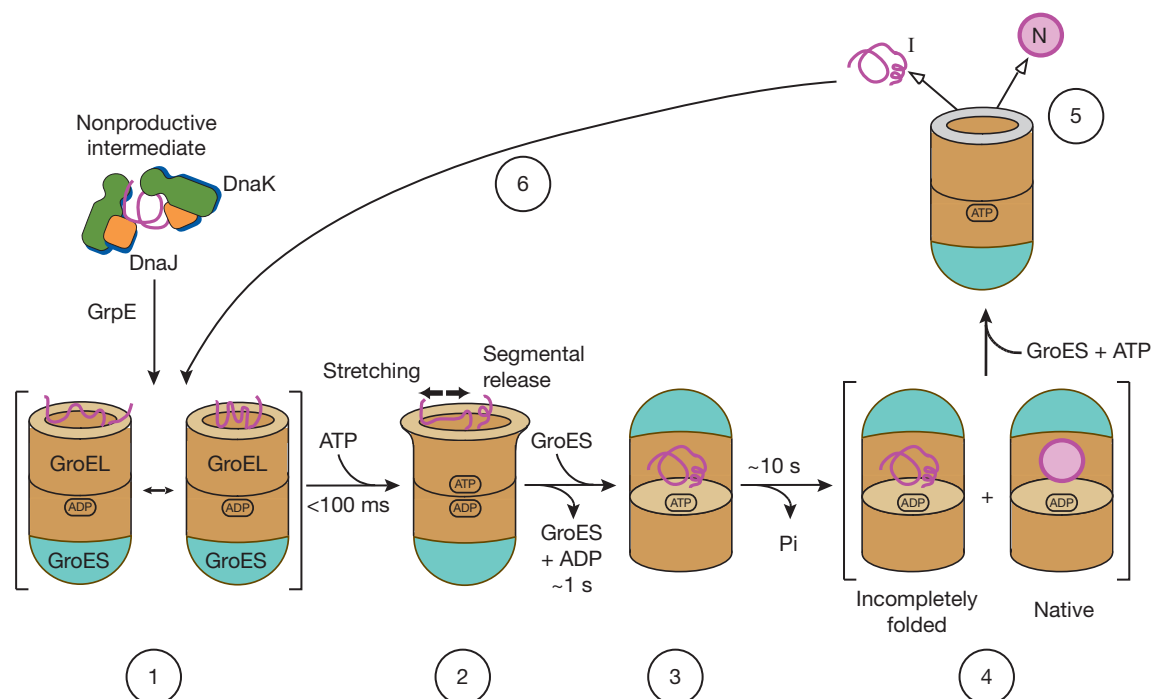


Figure 2 Protein-folding cycle of GroEL–GroES. Kinetically trapped or misfolded substrate protein may be delivered to GroEL by upstream chaperones such as DnaK–DnaJ (the Hsp70 chaperone system) (1). Upon binding to GroEL, substrate may undergo local unfolding, occupying multiple apical domain-binding sites of GroEL. ATP-dependent domain movement of the apical GroEL domains can result in a stretching of tightly bound regions of substrate and in release and partial compaction of less stably bound regions (2). Compaction is completed upon global substrate encapsulation by GroES (3). Subsequently, folding proceeds in the chaperonin cage (4). ATP binding to the opposite ring then initiates a new folding cycle, releasing GroES and allowing substrate exit (5). Incompletely folded substrates might rebound to GroEL. Note that binding of a second substrate molecule to the open ring of GroEL in steps (4) and (5) as well as the transient formation of a symmetrical GroEL–GroES₂ complex are omitted for simplicity. N, native state; I, folding intermediate. Modified from F Ulrich Hartl & Manajit Hayer-Hartl (2009) Converging concepts of protein folding *in vitro* and *in vivo*. *Nature Structural and Molecular Biology* 16: 574–581.

sufficient water was enclosed (water can freely access the cage). In this model, known as the Anfinsen cage model, the protein is thought to fold inside the cage at the same speed and efficiency as it would in bulk solution in the absence of aggregation, as confirmed for the 30-kDa protein rhodanese. However, this model cannot explain how GroEL–GroES accelerate the folding of certain proteins above their rate of spontaneous folding (measured in the absence of rate-limiting aggregation). GroEL-dependent substrates typically are ~35–60 kDa in size and have complex domain topologies. Such proteins tend to accumulate kinetically trapped folding intermediates that need activation to proceed rapidly toward the native state. Assuming that kinetic trapping is caused by the presence of non-native intra-chain contacts, local unfolding upon substrate binding to multiple GroEL subunits or mediated by the ATP-dependent expansion of the GroEL ring may destabilize these misfolded states (Figure 2). Accordingly, iterative cycles of unfolding followed by folding trials in the chaperonin cage (iterative annealing) may increase the rate and/or the yield of the folding reaction. However, accelerated folding of proteins such as RuBisCo (~50 kDa) is also observed upon single-round encapsulation in a noncycling GroEL variant, arguing against iterative annealing as the mechanism of rate acceleration and suggesting that the physical environment of the cage is critical for this effect. Indeed, geometric confinement in the cage may significantly reduce the number of accessible

conformational states, thereby entropically destabilizing flexible folding intermediates and promoting their conversion to the native state. Furthermore, the cavity walls are lined with charged side chains whose electric potential may be difficult to dissipate through solvation by water in the presence of large enclosed proteins. As charged residues are normally located at the surface but not in the hydrophobic interior of natively folded proteins, the charged groups on the cavity walls might thus help to drive folding of the substrate into a near-native state.

Another feature frequently found in bacterial group I chaperonins (and also in group II chaperonins) are flexible C-terminal sequences extending into the cavity, often containing multiple Gly–Gly–Met repeats. *In vitro*, the length of these extensions modulates the folding efficiency, apparently by adapting the effective volume of the cage to substrate size.

The Eukaryotic Chaperonin TRiC/CCT

The paradigm for group II chaperonins is the eukaryotic cytosolic chaperonin TRiC (alternatively named CCT). X-ray crystallographic data are available for the less complex thermosome but only little is known about the folding substrates and the function of these chaperonins. The subunit sequence within each ring of TRiC/CCT and also the relative orientation

of the two rings appears to be invariant, but the exact arrangement of the eight subunits around the ring has not yet been unequivocally determined.

Despite the close structural homology between group I and II chaperonin subunits, the architecture of their complexes deviates markedly. In group I chaperonins, the equatorial domains are interdigitated; each subunit contacts two subunits in the opposite ring. In group II chaperonins the subunits are directly opposed, limiting the contacts across the equator to one subunit (Figure 1). Furthermore, the closure of the in-built lid in group II chaperonins requires a regulatory mechanism distinct from the association and dissociation of the GroES cofactor in group I chaperonins. Similar to GroEL, ATP hydrolysis within one ring exhibits positive cooperativity, whereas negative coupling occurs between the rings in TRiC/CCT. However, it is rather the ATP hydrolysis transition state than ATP binding that converts the open to the closed conformation, which is characterized by the association of the apical domain extensions and formation of the internal cavity. In this conformation, the putative substrate-binding sites in the apical domains become buried at the subunit-subunit interfaces. Electron microscopy suggests that the conformational transition proceeds sequentially around the ring, perhaps reflecting different ATP affinities and hydrolysis activities of the distinct TRiC/CCT subunits. In the open state, the subunits assume an extended, relaxed conformation, in which the apical domains are readily accessible for substrate.

Electron microscopic structures of TRiC/CCT in complex with substrates suggest that multiple subunits are involved in substrate binding and that these contacts occur in a specific binding pattern, providing a possible explanation for the non-redundancy of the TRiC/CCT subunits. Based on the conservation of signature residues in individual subunits, it has been suggested that the substrate interfaces in the apical domains of TRiC/CCT are rather flat and polar in character, more suited for interactions with late or near-native folding intermediates than with the molten globule-like states recognized by GroEL. However, similar to GroEL, TRiC/CCT also recognizes short hydrophobic sequence motifs in model substrates such as Van Hippel Lindau (VHL). Probably, the TRiC/CCT complex employs a mosaic of binding sites for folding intermediates with distinct physico-chemical properties.

Cellular Functions of Chaperonins

In bacteria and eukaryotes, the chaperonins participate in the folding of 10–15% of the cellular proteins, making them one of the central components of the folding machinery besides Hsp70 and Hsp90. Both GroEL/GroES and all TRiC/CCT subunits are essential in *E. coli* and *Saccharomyces cerevisiae*, respectively.

Proteomic analysis indicated that ~250 different cytosolic proteins interact with *E. coli* GroEL. The molecular weights of most of the GroEL interactors fall in the range of 20–50 kDa, matching the cavity size of GroEL. Majority of GroEL-dependent proteins have complex α/β or $\alpha + \beta$ topologies, with a predominance of the triose phosphate isomerase (TIM) barrel fold. For example, the large subunit of RuBisCO, the obligate bulk substrate of the chloroplast group I chaperonin, has such a TIM barrel fold. Similar to the large RuBisCO subunit, many GroEL substrates form homo- or hetero-

oligomeric assemblies. As in the case of RuBisCO, the monomeric precursors of some of these proteins may be structurally labile, resulting in a need for specific chaperones acting downstream of chaperonin in mediating assembly.

TRiC/CCT was first recognized as the essential folding catalyst of the cytoskeletal proteins actin and tubulin, both of which are unable to fold spontaneously. Today, it is clear that many other eukaryotic proteins are also substrates of TRiC/CCT. In recent proteomics and genomics studies, more than 100 potential TRiC/CCT substrates were identified. Bioinformatic analysis revealed that multidomain proteins ranging in size between 40 and 75 kDa and proteins rich in regions of β -strand propensity are over-represented in the TRiC/CCT interactome. These proteins are predicted to be slow folding and aggregation prone. Many TRiC/CCT substrates contain β -propeller domains with a characteristic WD-40 repeat pattern, a structural building block frequently found in eukaryotic proteins. Many of the TRiC/CCT substrates are subunits of important eukaryotic complexes, such as the anaphase-promoting complex, the nuclear pore complex, and the septin complex. TRiC/CCT is also implicated in the suppression of toxicity arising from the amyloid-forming protein huntingtin, which causes neuronal cell death in Huntington's disease.

Although TRiC/CCT was also found associated to translating ribosomes, it is assumed that other chaperones generally act upstream in stabilizing nascent chains and mediating their transfer to TRiC/CCT. A molecular chaperone specialized for the cooperation with group II chaperonins is prefoldin/GimC, a jelly fish-shaped complex, which can stabilize folding substrates between six tentacle-like coiled-coil extensions. The prefoldin complex is conserved in both archaea and the eukaryotic cytosol. The archaeal version contains two types of subunits; the eukaryotic has six subunits, mirroring the increase in complexity in subunit composition that occurred during evolution from the thermosome to TRiC/CCT. Electron microscopy of the prefoldin-TRiC/CCT complex suggests that the substrate might be handed over directly into the folding cavity. Deletion of prefoldin subunits impairs actin and tubulin function to varying degrees, implicating prefoldin in the folding of these TRiC/CCT substrates. Besides prefoldin, the Hsp70 chaperone acts upstream of TRiC/CCT in eukaryotes. An electron microscopic structure of the TRiC/CCT-Hsp70 complex indicated that Hsp70 specifically recognizes the apical domain of one chaperonin subunit for the handover of substrates.

Another class of co-chaperones of TRiC/CCT are the phosducin-like proteins (PhLPs), which are present in several isoforms. In contrast to prefoldin, the PhLPs form stable complexes with the chaperonin and could potentially influence chaperonin function directly, for example, by enlarging the cavity size or regulating the ATPase activity of TRiC/CCT. Genetic studies in *S. cerevisiae* point to an involvement of PhLPs in actin and tubulin folding. A role of PhLPs in folding of the α -subunit of trimeric G-proteins, a β -propeller protein, has also been demonstrated.

Conclusions

The chaperonins constitute an essential part of the cellular folding machinery in every organism, with the GroEL-GroES

complex currently being the best-characterized molecular chaperone system. In eukaryotes, the TRiC/CCT complex is a central element of the cellular protein quality control network, intricately involved in processes that contribute to aging and human disease. Of particular interest for future studies is a detailed comparison of the folding mechanisms of spontaneous and chaperonin-assisted folding.

See also: [Cell Architecture and Function: Heat/Stress Responses](#); [Protein/Enzyme Structure Function and Degradation: Protein Folding and Assembly](#).

Further Reading

- Booth CR, Meyer AS, Cong Y, et al. (2008) Mechanism of lid closure in the eukaryotic chaperonin TRiC/CCT. *Nature Structural and Molecular Biology* 15: 746–753.
- Brinker A, Pfeifer G, Kerner MJ, et al. (2001) Dual function of protein confinement in chaperonin-assisted protein folding. *Cell* 107: 223–233.
- Bukau B and Horwich AL (1998) The Hsp70 and Hsp60 chaperone machines. *Cell* 92: 351–366.
- Cuellar J, Martín-Benito J, Scheres SH, et al. (2008) The structure of CCT–Hsc70 NBD suggests a mechanism for Hsp70 delivery of substrates to the chaperonin. *Nature Structural and Molecular Biology* 15: 858–864.
- Dekker C, Stirling PC, McCormack EA, et al. (2008) The interaction network of the chaperonin CCT. *EMBO Journal* 27: 1827–1839.
- Ditzel L, Löwe J, Stock D, et al. (1998) Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell* 93: 125–138.
- Hartl FU and Hayer-Hartl M (2009) Converging concepts of protein folding *in vitro* and *in vivo*. *Nature Structural Molecular Biology* 16: 574–581.
- Kerner MJ, Naylor DJ, Ishihama Y, et al. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* 122: 209–220.
- Martin J, Langer T, Boteva R, et al. (1991) Chaperonin-mediated protein folding at the surface of GroEL through a 'molten globule'-like intermediate. *Nature* 352: 36–42.
- Martín-Benito J, Bertrand S, Hu T, et al. (2004) Structure of the complex between the cytosolic chaperonin CCT and phosducin-like protein. *Proceedings of the National Academy of Sciences of the United States of America* 101: 17410–17415.
- Martín-Benito J, Boskovic J, Gómez-Puertas P, et al. (2002) Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT. *EMBO Journal* 21: 6377–6386.
- Mayhew M, Da Silva ACR, Martin J, et al. (1996) Protein folding in the central cavity of the GroEL–GroES chaperonin complex. *Nature* 379: 420–426.
- Meyer AS, Gillespie JR, Walther D, et al. (2003) Closing the folding chamber of the eukaryotic chaperonin requires the transition state of ATP hydrolysis. *Cell* 113: 369–381.
- Rye HS, Burston SG, Fenton WA, et al. (1997) Distinct actions of *cis* and *trans* ATP within the double ring of the chaperonin GroEL. *Nature* 388: 792–798.
- Sharma S, Chakraborty K, Müller BK, et al. (2008) Monitoring protein conformation along the pathway of chaperonin-assisted folding. *Cell* 133: 142–153.
- Spiess C, Meyer AS, Reissmann S, and Frydman J (2004) Mechanism of the eukaryotic chaperonin: Protein folding in the chamber of secrets. *Trends in Cell Biology* 14: 598–604.
- Tang YC, Chang HC, Roeben A, et al. (2006) Structural features of the GroEL–GroES nano-cage required for rapid folding of encapsulated protein. *Cell* 125: 903–914.
- Weissman JS, Rye HS, Fenton WA, Beechem JM, and Horwich AL (1996) Characterization of the active intermediate of a GroEL–GroES-mediated protein folding reaction. *Cell* 84: 481–490.
- Xu Z, Horwich AL, and Sigler PB (1997) The crystal structure of the asymmetric GroEL–GroES–(ADP)₇ chaperonin complex. *Nature* 388: 741–750.
- Yam AY, Xia Y, Lin HT, et al. (2008) Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. *Nature Structural and Molecular Biology* 15: 1255–1262.

Chemiluminescence and Bioluminescence

T O Baldwin, University of Arizona, Tucson, AZ, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 399–404, © 2004, Elsevier Inc.

Glossary

Bioluminescence The light that results from a chemiluminescent reaction in a living organism or from components of a living system, such as firefly tails.

Chemiluminescence The light that is emitted from certain chemical reactions without the production of significant heat.

Flavin mononucleotide Riboflavin-5'-phosphate, the 5'-phosphate derivative of vitamin B₂.

Flavin monooxygenase An enzyme that employs a flavin coenzyme to split the two atoms of molecular

oxygen apart, depositing one in an organic substrate as a hydroxyl group and the other in water.

Fluorescence The phenomenon of light emission that occurs from certain molecules as they return to the ground state from the lowest excited singlet excited state. Population of the singlet excited state is the result of absorption of energy from incident light.

Luciferase An enzyme that catalyzes a reaction that emits visible light with high efficiency. Luciferases are exceptionally diverse, many with no evolutionary relationship to the others.

Chemiluminescence and bioluminescence are terms that refer to the same physical process of light emission without heat. Bioluminescence is the process by which a living system or components isolated from a living system, such as a firefly tail, carry out a series of reactions that result in emission of light. Chemiluminescence is the same kind of process, but involving molecules that are not of biological origin. Chemiluminescence should be distinguished from incandescence, which occurs at high temperatures. Both bioluminescence and chemiluminescence are well known to the public, and both hold great fascination for all observers. Light without heat has intrigued people since the beginning of recorded history, and surely before.

Fluorescence: The Underlying Property

Fluorescence is a characteristic common to molecules formed as a result of chemiluminescent and bioluminescent reactions. At normal ambient temperatures, the vast majority of molecules exist in what is known as the ground state. However, under certain conditions, such as absorption of light energy, a molecule may be converted into an excited state. For fluorescent molecules, the excited state can return to the ground state by emitting energy as light. Virtually, all molecules are fluorescent to some extent, but most emit such a small amount of light that sensitive instruments are required to detect it. A graphic description of fluorescence is shown in Figure 1.

In the case of fluorescence, the energy input is the absorption of light. The chemical characteristics of a molecule will determine which wavelengths (color) of light it will absorb. As in the ground state, the excited state will have multiple vibrational modes at discrete energy levels. At normal ambient temperatures, most of the molecules will be in the lowest vibrational energy level, so following absorption of the input energy, the molecules will relax to the lowest vibrational level of that excited state. Excited states tend to be very unstable,

with lifetimes from the nanosecond to picosecond range. There are many ways for a molecule to return to the ground state. In solution, the excited molecules may collide with other molecules in solution and impart energy through the collisions, thereby increasing the vibrational and translational energies of the other molecules while returning to the ground state. Alternatively, some molecules, referred to as fluorescent molecules, have the ability to return to the ground state by emission of light. Note that the emitted light will be of lower energy (red-shifted in the spectrum) relative to that of the input energy due to the relaxation of the excited state. This effect is known as the Stokes shift.

Chemiluminescence: Fluorescence from a Chemical Reaction

In the case of chemiluminescence, the input energy is derived from the making and breaking of bonds that occur during a chemical reaction. Other than this important detail, fluorescence and chemiluminescence result from the same fundamental characteristics of the molecules involved. For a reaction to be chemiluminescent, one of the products must be fluorescent, and the chemical step that results in generation of the fluorescent product must be of sufficiently high energy to result in formation of the excited state. Therefore, the efficiency of light emission from a chemiluminescent reaction is the product of the efficiencies of each step. That is, the chemiluminescent quantum yield, Φ_{CL} , or the yield of photons per molecule of reactant consumed in the reaction:

$$\Phi_{CL} = \Phi_R \times \Phi_{CE} \times \Phi_F$$

All of these efficiencies have values between 0 and 1. The reaction yield (Φ_R) is the chemical yield of the correct products, rather than the products of side reactions, which may not be fluorescent, and is typically near 1. The chemical excitation yield (Φ_{CE}) is the fraction of the fluorescent product molecules that are produced in the excited state, rather than the ground

state. To be considered a chemiluminescent reaction, this value should be 10^{-3} or greater. Finally, the fluorescence quantum yield (Φ_F) is the proportion of those molecules that are produced in the excited state that return to the ground state by emission of light rather than by collisional quenching or some other process, as depicted in **Figure 1**. The fluorescence quantum yield is typically 0.1 or greater. Conditions that impact any of these three factors will impact the overall yield of light from the reaction.

In some cases, product molecules are produced with exceptionally high chemical yield in the excited state, but the product is itself nonfluorescent (very low Φ_F). In these cases, the yield of luminescence can be increased dramatically by addition of a highly fluorescent molecule to the reaction, which can, by a process called energy transfer, accept the energy from the excited product to become excited itself. The overall chemiluminescence quantum yield in such cases becomes

$$\Phi_{CL} = \Phi_R \times \Phi_{CE} \times \Phi_F \times \Phi_{ET}$$

where Φ_{ET} is the efficiency of energy transfer from the primary excited state produced in the chemical reaction to the acceptor, which then emits with the efficiency given by Φ_F . The chemiluminescence quantum yields in some sensitized chemiluminescent reactions can be very high indeed, approaching 1.

Much research has been reported on the chemiluminescence properties of a wide array of peroxides, especially the cyclic peroxides such as the dioxetanes. Tetramethyldioxetane will decompose to yield two molecules of acetone, one in the singlet excited state and the other in the triplet (Figure 2). Of

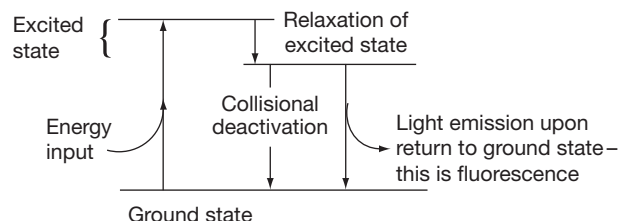


Figure 1 Diagram showing the processes of generation of the excited state and return of the excited state to the ground state.

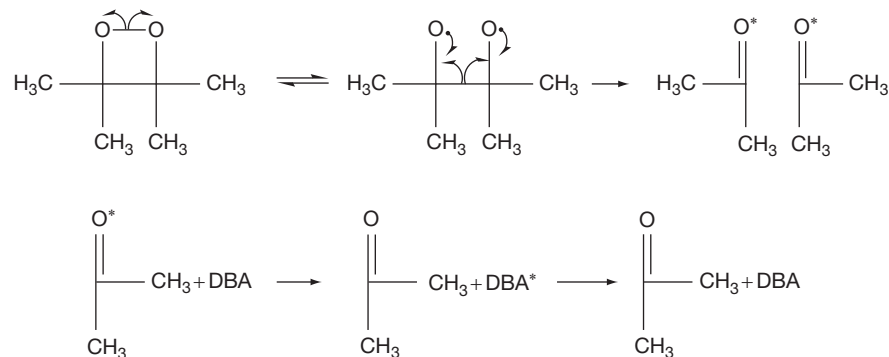


Figure 2 Chemiluminescent decomposition of tetramethyldioxetane. The first step of the reaction is the hemolytic cleavage of the oxygen–oxygen bond, followed by hemolytic cleavage of the carbon–carbon bond to yield the diradical, which recombines to form acetone in the excited state. In the absence of the dye 9,10-dibromoanthracene, the reaction is dark. However, the dye can accept the energy from the excited state carbonyl to yield the excited state of the dye, which returns to ground state with the emission of light.

course, acetone is nonfluorescent, so the decomposition reaction is essentially dark. However, if 9,10-dibromoanthracene or 9,10-diphenylanthracene is added to the reaction, intense light emission is detected as a result of the energy transfer from the excited-state acetone product to the fluorescent dye, dibromoanthracene (DBA) or diphenylanthracene (DPA).

These and related technologies have found wide use in the form of chemiluminescence-based diagnostics and other applications. Perhaps one of the best-known applications is the light-stick technology used by campers and hikers, and for fun and safety during Halloween. The light stick consists of a sealed, flexible plastic tube containing a solution of oxalic phthalate ester and a sealed glass vial containing a solution of hydrogen peroxide. Bending of the plastic tube will cause the glass vial to break, mixing the two solutions. The hydrogen peroxide will oxidize the ester, yielding two molecules of phenol and one molecule of 1,2-dioxetane-3,4-dione. The dioxetane will decompose to yield two molecules of carbon dioxide in the excited state. By including a dye sensitizer in the solution, the energy of the excited state carbon dioxide will result in light emission (see **Figure 3**). The color of the light emitted is determined by the fluorescence properties of the dye used in the light stick and can range from blue to red. Examples of dyes used are shown in **Figure 3**. One characteristic common to most fluorescent compounds and demonstrated in these examples is a system of conjugated double bonds.

Bioluminescence: Chemiluminescence from a Biological Source

Bioluminescence reactions, unlike chemiluminescence, require an enzymatic catalyst for the reaction to occur. These enzymes are generically referred to as luciferases. Even though they have the same name, they catalyze vastly different reactions. For comparison, other groups of enzymes that have a common name, such as proteases, all catalyze the same kind of reactions. Proteases all hydrolyze peptide bonds, but luciferases catalyze reactions on completely different substrates, and they have no evolutionary relationship. Firefly luciferase, bacterial luciferase, and Renilla luciferase catalyze different reactions, having in common only the fact that light is emitted from a product of

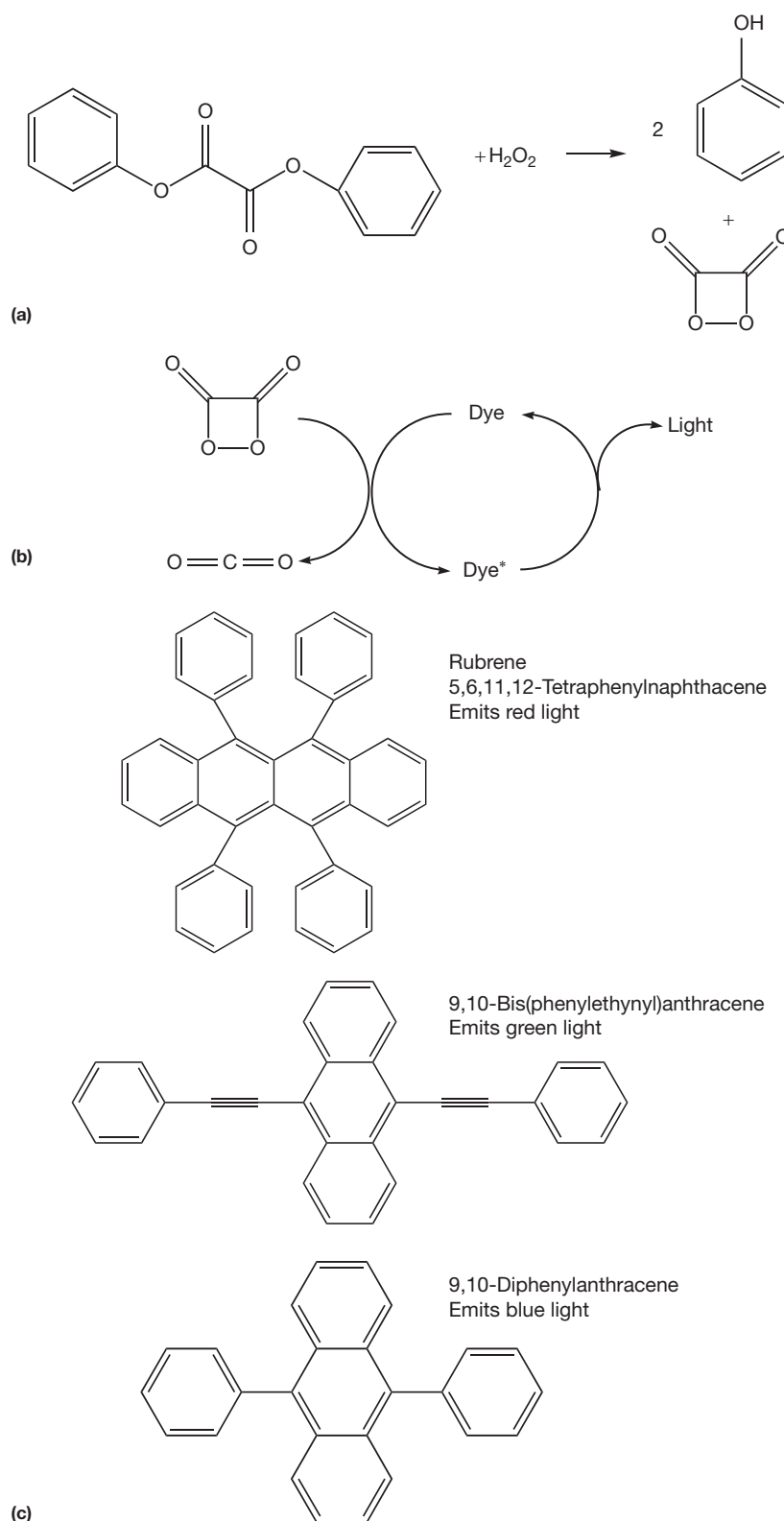


Figure 3 The chemistry of the light stick. The reaction that occurs in a light stick upon breaking the glass ampule and allowing the two solutions to mix is shown in panel (a). The oxalic phthalate ester is oxidized to yield two molecules of phenol and one molecule of the dioxetane. In panel (b), the dioxetane decomposes by the same mechanism shown in [Figure 2](#) to yield two molecules of carbon dioxide. The excited-state carbon dioxide transfers energy to the dye to yield the excited state of the dye, which in turn emits light. The color of the light is determined by the chemistry of the dye. Three examples are shown in panel (c), spanning the spectrum from red to blue.

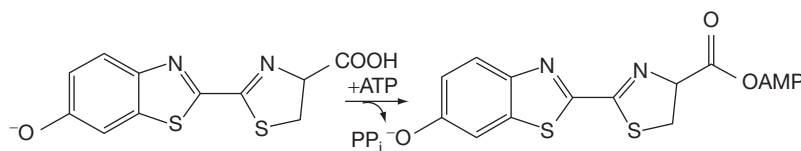


Figure 4 Proposed mechanism of the firefly luciferase reaction. The luciferin reacts with ATP to form the luciferyl adenylate, thus preparing the molecule for reaction with oxygen.

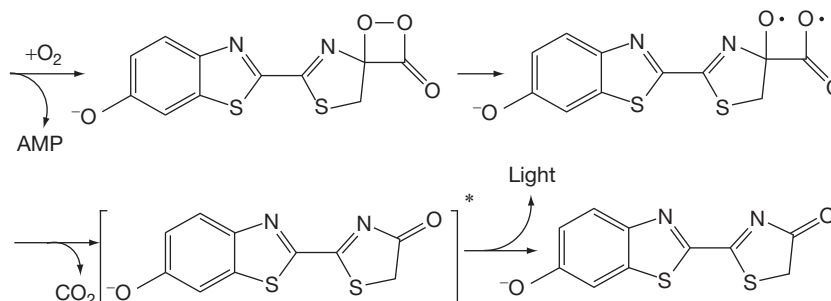


Figure 5 Proposed mechanism of the firefly luciferase-catalyzed reaction. The luciferyl adenylate (Figure 4) reacts with molecular oxygen, eliminating the AMP and forming the cyclic peroxide dioxetanone structure. The dioxetanone then decomposes by homolytic cleavage of the O–O and C–C bonds, liberating CO₂ and yielding the excited state of the oxyluciferin. Light emission occurs as the excited state returns to ground state.

the reaction. It is therefore very important when discussing bioluminescence to stipulate the biological source of the enzyme.

There are many examples of bioluminescence that resemble the nonbiological chemiluminescence reactions described previously. For example, the reaction that occurs in the tail of the firefly is very similar to that of the light stick, but unlike the light stick, the firefly reaction requires the participation of an enzyme, firefly luciferase. As with chemiluminescent reactions, the firefly luciferase-catalyzed reaction has found numerous practical applications, primarily due to the involvement of adenosine triphosphate (ATP) in the reaction (Figure 4). Many of these applications involve accurate and sensitive analysis of ATP levels within samples. The role of the ATP is to activate the carboxyl group of the firefly luciferin, the substrate for the luciferase-catalyzed reaction. As a result of this adenylation reaction, the luciferyl adenylate is now poised to react with molecular oxygen, eliminating adenosine monophosphate (AMP) and forming the dioxetanone intermediate (Figure 5). The dioxetanone ring then decomposes by a mechanism similar to that shown in Figure 2, ultimately yielding CO₂ and the excited state of the product oxyluciferin, which emits light as it returns to ground state.

Different species of firefly emit light of different colors, but they all appear to use the same substrate luciferin, and there appear to be no energy transfer systems involved. At present, the detailed mechanism by which the insects emit light of different color is unknown, but it surely has to do with the details of the interactions of the excited state with the luciferase enzyme, since the enzymes are slightly different between the species. The color of the light emitted ranges from green to red, and the quantum yield of these reactions approaches 1.

Unlike firefly luciferase, which is a single polypeptide and employs ATP and a special substrate firefly luciferase, bacterial

luciferase is a heterodimer ($\alpha\beta$) consisting of two similar but nonidentical polypeptides. In bacterial luciferase, there is a single active center on the α -subunit, but both subunits are required for the high quantum yield reaction. The substrates for the bacterial luciferase reaction are reduced flavin mononucleotide, molecular oxygen, and a long-chain saturated aldehyde. The enzyme formally is a flavin monooxygenase, catalyzing the cleavage of molecular oxygen and inserting one atom of oxygen into a substrate and the other into water. The mechanism of light emission in the bacterial luciferase-catalyzed reaction is still under debate, but there is general agreement regarding most of the reaction. In the first step of the reaction, reduced flavin mononucleotide (FMNH₂) reacts with O₂ to yield a reduced flavin peroxide (step 1, Figure 6). In the second step, the peroxide reacts with the aldehyde substrate to form the tetrahedral intermediate shown in Figure 6, step 2. There is general agreement in the field that this tetrahedral intermediate forms in the reaction, but how the intermediate decomposes to yield the excited state is under debate. One possible mechanism is shown in step 3 of Figure 6.

The identity of the emitter in the bacterial bioluminescence has been difficult to identify unambiguously because the product of the reaction, oxidized flavin mononucleotide, although fluorescent in solution, is nonfluorescent when bound to the luciferase. Various lines of evidence suggest that the emitter in the reaction is the flavin pseudobase shown in step 3 of Figure 6. The primary excited state, however, must be some other molecule, possibly the excited carbonyl shown in step 4 of Figure 6. In some species of bioluminescent bacteria, light emission from the living bacteria is significantly blue-shifted compared to that from the purified luciferase enzyme. It has been shown that the blue light comes from another protein, the lumazine protein, which becomes excited through some

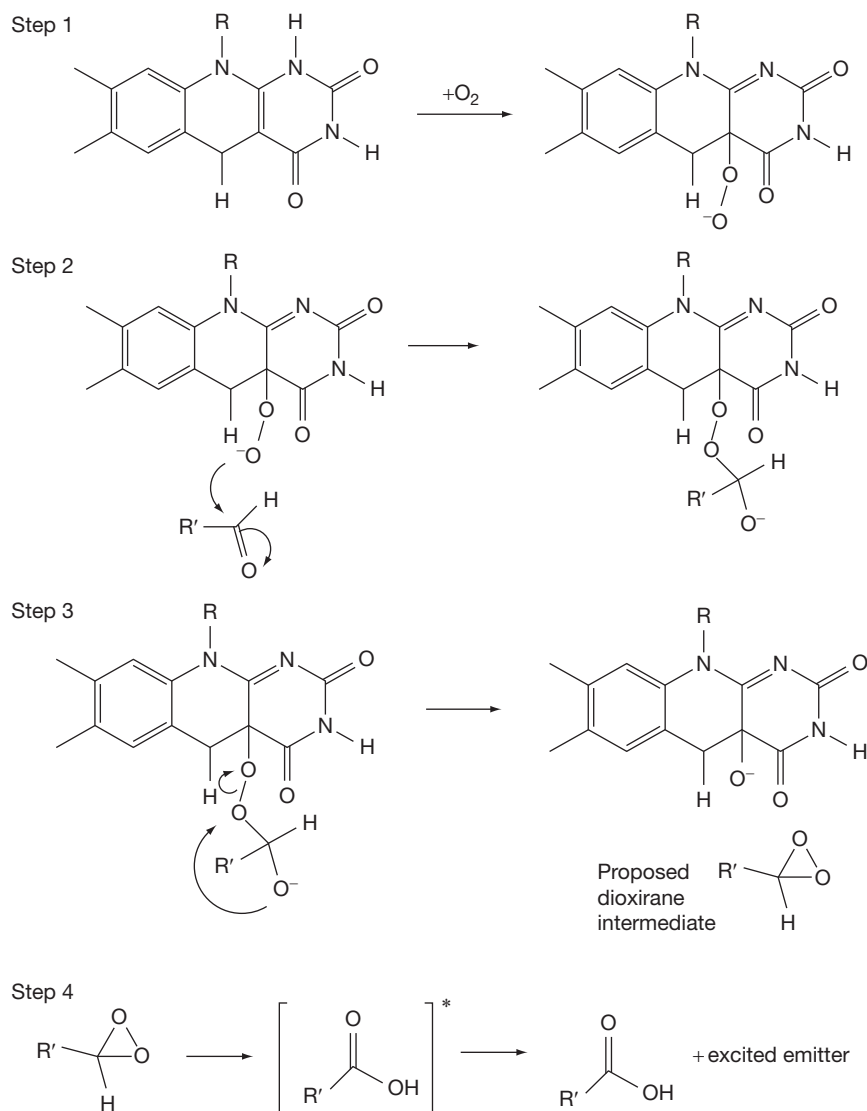


Figure 6 Proposed mechanism of the bacterial luciferase-catalyzed reaction, as described in the text.

form of interaction with the luciferase during the reaction. This is an example of an energy transfer process in a bioluminescence system, similar to the chemiluminescent examples described in Figure 3. But for the energy transfer to be efficient, it must occur in an energetically downhill direction. That is, the primary excited state must have an energy level higher than the excited flavin, which emits at a longer wavelength (lower energy) than does the lumazine protein. It is this fact, among others, that leads to the belief that the primary excited state in this reaction is likely to be an excited carbonyl, rather than the excited flavin. In the absence of the lumazine protein, energy transfer can occur to the flavin with emission from the excited flavin.

Another example of energy transfer in the bacterial bioluminescence system is from a yellow-emitting strain of *Vibrio fischeri*. These bacteria have a luciferase that emits blue-green light, similar to other bacterial luciferases, but *in vivo* and at temperatures of 20 °C or lower, the color of the light emitted is bright yellow. This color shift is the result of a yellow

fluorescence protein that has a highly fluorescent-oxidized flavin mononucleotide chromophore. It is interesting that the lumazine protein, which causes a blue shift, and the yellow fluorescence protein appear to be homologous and carry out similar functions of energy transfer but do so with different chromophores.

There are many other examples of bioluminescent and chemiluminescent reactions, many of which have found important and valuable uses. The green fluorescent protein that was isolated from jellyfish is a very important tool in the study of protein trafficking and other molecular imaging applications. In immunoassays, chemiluminescent and bioluminescent reactions offer sensitivity, but without the hazards of radioactive labels. In recent years, work in the field has become primarily focused on applications of the technology; by comparison, very little effort is being directed at a basic understanding of the mechanisms of the processes involved. As should be apparent from this brief overview, much remains to be learned.

See also: **Protein/Enzyme Structure Function and Degradation:** Flavins.

Further Reading

Baldwin TO (1996) Firefly luciferase: The structure is known, but the mystery remains. *Structure* 4: 223–228.

Baldwin TO and Ziegler MM (1992) The biochemistry and molecular biology of bacterial bioluminescence. In: Müller F (ed.) *Chemistry and Biochemistry of Flavoenzymes*, vol. III, pp. 467–530. Boca Raton, FL: CRC Press.

Harvey EN (1952) *Bioluminescence*. New York, NY: Academic Press.

Harvey EN (1957) *A History of Luminescence from the Earliest Times to 1900*. New York, NY: American Philosophical Society.

Lakowicz JR (1999) *Principles of Fluorescence Spectroscopy*. New York, NY: Kluwer Academic/Plenum.

Ziegler MM and Baldwin TO (eds.) (2000) *Bioluminescence and Chemiluminescence, Part C: Methods in Enzymology*, vol. 305. New York, NY: Academic Press.

Chemiosmotic Theory

P R Rich, University College London, London, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Electrogenic A process that produces an electric field due to net movement of charged species across a region of low dielectric strength such as a lipid bilayer membrane.

Electroneutral A process that results in no net charge change and does not produce an electric field.

Inner mitochondrial membrane The innermost membrane of mitochondria that houses the protonmotive electron transfer chain and adenosine triphosphate (ATP) synthase and separates the intermembrane space and matrix.

NAD(P)⁺ and NAD(P)H The oxidized and reduced form of nicotinamide adenine dinucleotide (phosphate).

NADH is produced by metabolic dehydrogenation reactions and is reoxidized by complex I of the respiratory chain.

NADPH is generated by reductants produced by photosystem I. They are donors/acceptors of two electrons and only one proton, but are sometimes shown in textbooks as NAD(P) and NAD(P)H₂.

Protonmotive force A protonmotive force is generated when protons move across a proton-impermeable barrier. It is a free energy term arising from an electrical charge difference and a pH difference across the barrier.

Thylakoid membrane The membrane within chloroplasts that house the photosynthetic electron transfer chain and ATP synthase.

Vectorial reaction A reaction in which the reactants are arranged spatially such that the reaction products are produced in a spatial manner.

Introduction

All living cells require energy to survive. The energy can be provided by food, by light, or, in diverse bacteria, by a myriad of possible environmental materials. These are used in large part to produce adenosine triphosphate (ATP) from adenosine diphosphate (ADP) and inorganic phosphate. ATP is a stable and transportable intracellular energy source. Its hydrolysis back to ADP and inorganic phosphate releases energy that can be used to drive energy-requiring cellular processes. Some ATP synthesis occurs by classical enzymological mechanisms in which an energy-releasing chemical transformation is directly coupled to ADP phosphorylation within a single enzyme complex. Such substrate level phosphorylations occur in the glycolytic pathway. The majority of ATP synthesis, however, is usually associated with respiratory and photosynthetic processes that are catalyzed by complex redox enzymes embedded in ion-impermeable lipid membranes. The way in which these redox reactions are coupled to ATP synthesis is quite different to that of the substrate level phosphorylations because the energy-releasing redox reactions and energy-requiring ATP-synthesizing reactions do not occur within the same enzyme structures. Instead, they are coupled indirectly through an intermediary energy-storing proton gradient that is formed across the membrane in which they are embedded. This revolutionary chemiosmotic hypothesis of the mechanism of such coupling was first postulated by Peter Mitchell in 1961. It took many years before it, and its ramifications, gained widespread acceptance as the chemiosmotic theory, though even today some aspects of development of the ideas have remained controversial.

Historical Background

When Mitchell's first chemiosmotic paper appeared in 1961, it was already known that mitochondria and chloroplasts were primary producers of ATP and that the enzymes that harness the energy source and the ATP synthases are embedded in the same

lipid bilayer membranes. It was also known that the redox processes of respiration and photosynthesis could be uncoupled from each other, a phenomenon that was difficult to explain in terms of conventional enzymology. The prevailing ideas on how they are coupled involved mechanisms akin to those of substrate level phosphorylations or to the existence of a high-energy chemical intermediate (X~I) that could provide the linkage between the sources and sinks of energy. Despite considerable efforts to identify the nature of X~I, it remained elusive. Nevertheless, Mitchell's proposal that the coupling agent could instead be a proton gradient across the lipid membrane remained highly controversial and largely ignored.

It took more than a decade of experimental support, both from Mitchell himself and from other laboratories around the world, for his ideas to become accepted widely. Mitchell, together with his colleague Jennifer Moyle, left University in 1964 and set up the independent Glynn Research Institute in relatively remote Cornwall. There, together with a small research team, they set about testing the central tenets of the hypothesis by experiment. At the same time, Mitchell further developed his chemiosmotic hypothesis and placed it firmly within the context of the rapid developments that were occurring in the field of experimental bioenergetics. This work was published in two monographs, the 'Grey Books', in 1966 and 1968. These provided an expanded account of the chemiosmotic hypothesis and its application both to the coupling between respiratory and photosynthetic electron transfer and ATP synthesis and to energy-linked solute transport across membranes. Over the next decade it slowly gained support, often after acrimonious exchanges, and, in 1978, Mitchell was awarded the Nobel prize in chemistry for his work. The background and details of this remarkable case of difficulties of introduction of a radical idea into mainstream scientific thinking has been the subject of study and analysis by scientific sociologists and historians. The most important pieces of documentation have been archived and are held in Cambridge University Library.

The Central Features of the Chemiosmotic Theory

The chemiosmotic hypothesis was developed in the context of respiratory and photosynthetic electron flow, and the energetic coupling of these reactions to ATP generation and to active solute transport. It had four essential postulates in the form presented in the 1966 Grey Book, which can be paraphrased as:

1. The ATP synthase systems are membrane located and their normal function is to couple reversibly the translocation of protons across the membrane to the condensation reaction between ADP and inorganic phosphate to form ATP.
2. The electron transfer chains are also membrane located and their function is to couple reversibly the flow of reducing equivalents to the translocation of protons across the membrane.
3. The membranes also contain carrier systems that permit reversible transmembrane exchange of ions with protons or hydroxide ions. Their function is for osmotic and pH stabilization and for essential metabolite transport.
4. Systems of postulates 1–3 above are located in a coupling membrane that has a low permeability to protons and to anions and cations generally.

These features are discussed in more detail in a modern context below.

The Ion-Impermeable Membrane and the Protonmotive Force

A central component of chemiosmotic systems is a lipid bilayer membrane that separates two aqueous phases. In mitochondria, this is the inner mitochondrial membrane that separates the internal matrix from the intermembrane space and cytoplasm. In chloroplasts, it is the thylakoid membrane that separates the chloroplast lumen and stromal aqueous phases. In bacteria, it is the plasma membrane separating the cytoplasm from the external medium.

These membranes are bilayers of hydrophobic lipids that are impermeable to charged ions including protons. Hence, when protons are moved from one side to the other across the membrane barrier through the actions of the enzymes involved, a stable gradient of protons is established. This is akin to the charging of a capacitor in electronics and represents a form of stored energy. In this case, the stored energy is an electrochemical proton gradient and it has two energetic components, a pH gradient (ΔpH) and, because protons are charged, an electrical potential difference ($\Delta\psi$). Mitchell called this the protonmotive force (PMF or Δp). The PMF is the sum of the pH and electrical difference terms. It is usually expressed in units of electrical potential by conversion of the ΔpH term into the same units as $\Delta\psi$:

$$\Delta p = \Delta\psi - 2.303RT/F \Delta\text{pH}$$

where F = the Faraday constant ($96,487 \text{ J}\cdot\text{V}^{-1}\cdot\text{mol}^{-1}$), R is the gas constant ($8.3145 \text{ J}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$), and T is temperature in K. At 25°C , $2.303RT/F \approx 59 \text{ mV}$ so that

$$\Delta p \text{ (in mV)} = \Delta\psi - 59\Delta\text{pH}$$

Δp is the free energy required to move protons across the membrane against the prevailing electrochemical proton

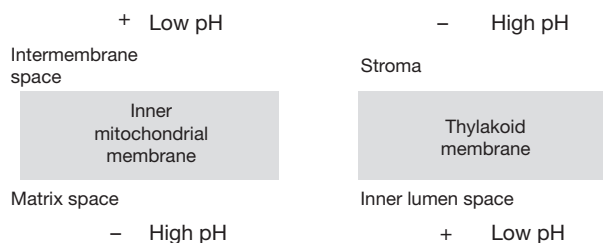


Figure 1 Components of PMF and their orientations in the coupling membranes of mitochondria and chloroplasts.

gradient, or the free energy available when moving protons back across the membrane to restore equilibrium.

The size of the electrical component is determined by the net charge difference and the capacitance of the membrane. The balance between the ΔpH and $\Delta\psi$ terms of Δp can be very different in different membrane systems. For example (Figure 1), in mitochondria under conditions when the Δp is maximal, the $\Delta\psi$ is typically close to 190 mV (positive on side facing the cytoplasm, Figure 1) and ΔpH is around 0.3 pH units (with highest pH in the matrix), resulting in a Δp of approx. 210 mV. In thylakoids in contrast, the Δp can be of a similar magnitude, but the $\Delta\psi$ -driven transfer of other ions across the membrane decreases the $\Delta\psi$ and allows the ΔpH component to rise substantially. The ΔpH can be up to 4 pH units (acidic in the lumen) but the $\Delta\psi$ is 30 mV or less (positive on the lumen side).

Membrane Location of Electron Transfer Chains and ATP Synthase

In order to be able to generate a PMF across the lipid bilayer membrane and to utilize it to generate ATP, it is essential that components of the enzymes involved must be arranged spatially in a specific manner within the membrane. This vectorial arrangement results in the redox reactions being coupled to the net movement of protons across the membrane. This in turn means that the free energy available from the redox reactions is not lost as heat but instead becomes stored in the form of the electrochemical proton gradient that they generate. The term 'chemiosmotic' was introduced to describe this type of process. Protein components of the ATP synthases embedded in the same membrane system provide a pathway for the return of the protons back across the membrane in a manner that drives ATP synthesis. The result is that the free energy available from the PMF-driven passage of protons across the membrane drives the otherwise endergonic synthesis of ATP.

The Vectorial Redox Loop

Importantly, Mitchell also proposed a model of the spatial arrangement and chemistry that would be required in a series of membrane-bound electron transfer components so that electron transfer would necessarily result in coupled proton translocation across the membrane. This elegant and simple proposal was termed the 'vectorial redox loop' (Figure 2(a)). It was postulated that the components of the redox chain alternated between electron carriers and hydrogen atom carriers and that

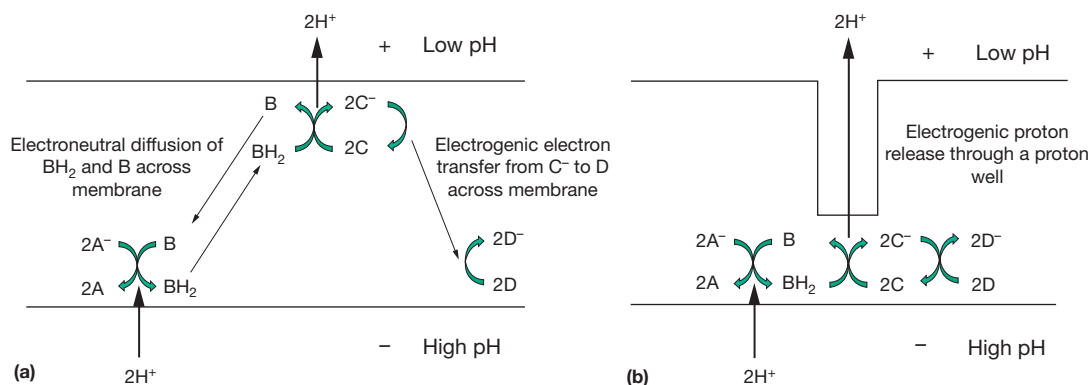


Figure 2 The redox loop proposal for coupling of electron transfer to transmembrane proton transfer. In its simplest form (a), an electron carrier (A) reduces an H-atom carrier (B). The site of reduction is close to the negative aqueous phase, so that reduction of B (shown here as a two-electron acceptor) results in proton uptake from that phase. The reduced form BH_2 then diffuses to the next acceptor C. Since C can only accept electrons and the reaction site is close to the positive aqueous phase, the two protons on BH_2 are released into this phase. The redox loop is completed by electrogenic transfer of the electrons to the next electron acceptor D that is located close to the negative aqueous phase. (b) This illustrates the same reactions, but in this case all reactions take place close to the negative aqueous phase. On oxidation of BH_2 , protons are released into a proton well that leads to the positive aqueous phase. The electrogenic movement of protons along this well to the positive aqueous phase generates the electric field. Most PMF-generating proteins tend to fall between these extremes of structure with both electron and proton transfer playing a part in electric field generation.

these were spatially arranged in a definite way in the membrane. The hydrogen carriers have redox-linked pK values such that reduction causes protonation and oxidation causes deprotonation. The hydrogen atom carriers are effectively mobile in the membrane, and diffuse from a site of reduction (and proton uptake) on one side of the membrane to a site of oxidation (and proton release) close to the opposite membrane surface. Since protons and electrons are transferred across the membrane together, the process is electroneutral. The loop is completed, however, by the transfer of the electrons back across the membrane *via* appropriately arranged electron carriers. In this case, uncompensated charge is moved across the membrane, and so an electric field is generated across the insulating membrane, i.e. the reaction is electrogenic. The net result is the appearance of one proton (with its positive charge) across the membrane for each electron passing through the redox loop, and the conversion of redox energy into an electrochemical potential gradient of protons, the PMF, across the membrane.

The classic example of a mobile hydrogen atom carrier behavior can be seen in the role of ubiquinone in mitochondria, and that of the related menaquinone and plastoquinone in bacterial and photosynthetic systems. A range of dehydrogenases and photosystems have quinone reduction sites that ensure that protons are taken up on one side of the membrane. As in **Figure 2(a)**, the quinone is reduced with two electrons and binds two protons to produce the reduced quinol form. The quinol then dissociates from its reduction site and physically diffuses to proteins with oxidation sites linked to the other membrane surface where it is reoxidized and the protons are released. Hydrophobic quinones play exactly this role as one part of the protonmotive reactions in mitochondria, chloroplasts, and many bacterial photosynthetic and respiratory chains.

In protonmotive enzymes, this simple redox loop notion is complicated by the role of the protein structures in the generation of the electrical and chemical potential components of the PMF. Usually, the catalytic sites where the key chemical and

redox transformations take place tend to be buried within the protein and can be located at any distance in terms of position across the lipid bilayer barrier. In this case, if protons are involved, channels in the protein structure must allow the protons into or out of the reaction site to the membrane surface. If the movement of the charged protons is not charge compensated by movement of counter ions then the proton movement, rather than the electron transfer, can be the primary electrogenic process. In the extreme example shown in **Figure 2(b)**, the highly localized redox loop takes place in reaction sites toward the negatively charged side of the membrane. As a result, the electron transfer steps do not generate an electric field. Instead, protons are released into a proton channel leading to the aqueous positive phase. The protons diffuse out electrogenically along the channel into the positive aqueous phase and in doing so generate an electric field. Hence, in this case it is proton movement, rather than electron transfer, which is electrogenic. Real protein systems tend to be somewhere between the extremes shown in **Figure 2**, with both electron and proton transfers playing a part in the generation of the electric field.

The Protonmotive Enzymes of Respiration and Photosynthesis

Tremendous progress has been made on understanding of the structures and mechanisms of the proteins and cofactors that are involved in generation and utilization of the PMF. In many systems, PMF generation is achieved by membrane-spanning enzymes that catalyze electron transfer and/or proton transfer toward buried catalytic centers as outlined in **Figure 2**. In many cases, the enzymes that oxidize the source of reductant are connected by a hydrophobic quinone to enzymes that pass the electrons to a terminal electron sink. The quinone often acts as a mobile hydrogen atom carrier as envisaged in the vectorial loop proposal.

However, several central enzymes have evolved additional protonmotive mechanisms that do not conform to the simple

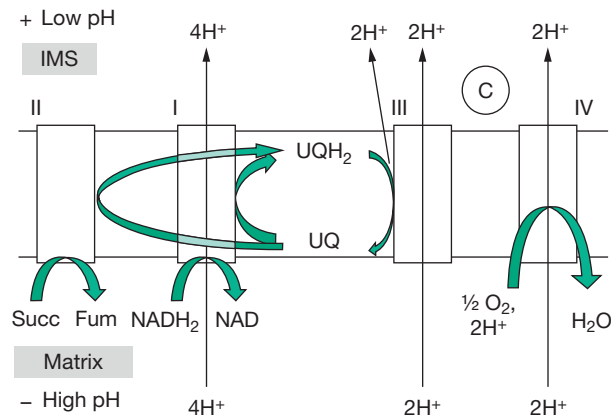


Figure 3 The protonmotive electron transport chain of mitochondria. The mitochondrial inner membrane is shown separating the internal matrix space and the intermembrane space (IMS) and the numbers refer to the passage of two reducing equivalents along the chain of redox enzymes. The four major enzymes are complex I (NADH:ubiquinone oxidoreductase), complex II (succinate:ubiquinone oxidoreductase), complex III (ubiquinol:cytochrome *c* oxidoreductase) and, connected by cytochrome *c*, complex IV (cytochrome *c* oxidase). Reduction of ubiquinone by NADH in complex I is linked to translocation of four protons across the membrane. When ubiquinol is oxidized by complex II, two protons are released and a further two are translocated. At complex IV, molecular oxygen is reduced water with uptake of two protons from the matrix and a further two are translocated. Complex II has no protonmotive activity. Overall, for succinate and NADH oxidation by oxygen, 10 and six protons, respectively, are lost from the matrix and appear on the cytoplasmic side.

vectorial loop mechanism. In mitochondria, these enzymes are complex I (an NADH:ubiquinone oxidoreductase), complex III (a quinol:cytochrome *c* oxidoreductase), and complex IV (cytochrome *c* oxidase). Diverse homologs of these enzymes, which retain the additional protonmotive functions, are found in archaea, bacteria, plants, and fungi. Hence, it seems that these complex functions have evolved very early and have been retained throughout the diversity of evolution. The mechanism operative in complex III was itself explained by Mitchell in his ingenious extrapolation of the vectorial loop ideas to a Q cycle mechanism of action. For complexes I and IV, their protonmotive mechanisms remain controversial. The current understanding of the PMF-generating steps of the classical respiratory and thylakoids chains are summarized in Figures 3 and 4.

ATP Synthase and Coupling to the Proton Gradient

The enzymes that harness the energy of the PMF to produce ATP are the F_0F_1 -type ATP synthases. As their structure emerged, it became clear that the ATP-synthesizing catalytic centers, located in the F_1 domain, are at a considerable distance from the membrane-embedded protein domains, linked only by relatively narrow connecting structures. It is now clear that the PMF drives a membrane-embedded rotor of identical *c* subunits. This in turn drives the mechanical rotation of a central stalk within the catalytic F_1 subunits, which are themselves prevented from rotating by a peripheral protein stator that links the membrane-located and catalytic F_1 domains. The rotating stalk forces a cycle of conformational changes on the

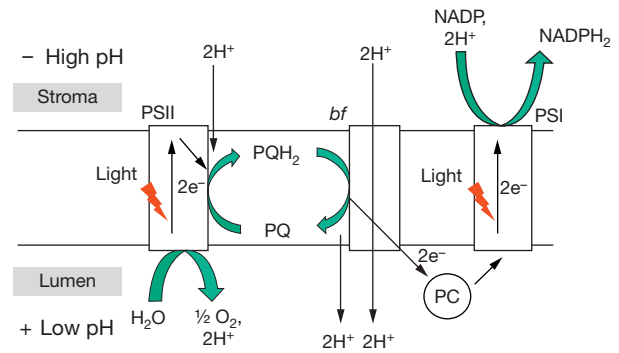


Figure 4 The protonmotive electron transport chain of chloroplasts. The thylakoids membrane separates the stromal and lumen aqueous phases of the chloroplast. The numbers refer to the passage of two reducing equivalents along the chains. Photosystem II is activated by two separate photons that cause electrogenic charge separation by passage of two electrons across the membrane. This results in water oxidation to $\frac{1}{2} O_2$ with release of two protons into the lumen, together with plastoquinone reduction with uptake of two protons from the stroma. The plastoquinol diffuses along and across the membrane where it is oxidized by the cytochrome *bf* complex with release of two protons. The cytochrome *bf* complex is homologous to mitochondrial complex III and is capable of translocating a further two protons from stroma to lumen by a Q-cycle mechanism as indicated in the diagram. However, there is no consensus on whether this occurs under all conditions. The electrons pass via plastocyanin (PC) to photosystem I where two separate photons again cause them to traverse the membrane electrogenically, finally reducing NADP to NADPH with uptake of two protons from the stroma. Hence, with operation of a Q cycle, the passage of two electrons can result maximally in six protons (with their six positive charges) being lost from the stroma and appearing in the lumen.

active site polypeptides that power the endergonic process of condensation of ADP and inorganic phosphate to form and release free ATP. F_0F_1 -type ATP synthases are fully reversible and their overall effect is to bring into equilibrium the free energy of the PMF and that of the ATP/ADP reaction. In some systems, they act primarily in reverse to generate a PMF by hydrolyzing ATP synthesized elsewhere.

The Proton Circuit and Stoichiometry of ATP Synthesis

The operation of the respiratory or photosynthetic electron transfer chains with the ATP synthases, linked through the PMF, creates a proton circuit that allows the energy released from the exergonic electron transfer to drive endergonic ATP synthesis. The yield of ATP synthesis is then governed by the numbers of protons translocated across the membrane by the components involved.

For the mammalian respiratory chain (Figure 3), it is generally agreed that oxidation of NADH through complexes I, III, and IV is coupled overall to the net transfer of 10 H^+ across the mitochondrial inner membrane. Succinate oxidation through complexes II, III, and IV is coupled to the net transfer of 6 H^+ . For the ATP synthase, one complete rotation of the rotor results in synthesis of three ATP inside the mitochondrial matrix. The number of protons required for a full rotation is governed by the number of subunits that comprise the membrane-embedded ring of identical subunits (the *c* subunits), since one proton is used to rotate this ring by one subunit. It has

been found that the c ring is composed of different numbers of subunits in different organisms. For example, yeast mitochondrial ATP synthase has 10 c subunits but the mammalian mitochondrial form has only eight.

A further consideration in eukaryotic systems is that transfer of each ATP out of the mitochondrial matrix into the cytoplasm occurs by the action of transporters (see below) that exchange one ATP for an ADP and a phosphate from the cytoplasm. In doing this, one H^+ of the proton gradient is also transferred back into the mitochondrial matrix. Hence, the total number of protons required to produce three molecules of ATP in the cytoplasm is given by the number of c subunits plus the three protons required to export them into the cytoplasm.

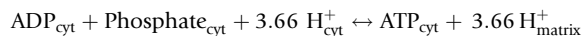
For example, the maximum extent of synthesis of ATP by oxidation of glucose in mammalian cells can be calculated as follows. Full oxidation to CO_2 and water, catalyzed by the enzymes of glycolysis and the citric acid cycle, results overall in the production in the cytoplasm of two molecules of ATP by substrate-level phosphorylation and two molecules of NADH. The NADH are transported in most cells into the mitochondrial matrix where another eight NADH are produced by the citric acid cycle, together with two guanosine triphosphates (GTPs).

Hence, overall, full oxidation of one glucose can provide the respiratory chain with 10 NADH and also, at the step of succinate oxidation in the citric acid cycle, two succinates that are oxidized to fumarate by complex II. The oxidation of the 10 NADH and two succinates by the protonmotive respiratory chain will result in 112 H^+ translocated across the inner mitochondrial membrane. Since the mammalian ATP synthase has eight subunits in its c ring, this means that synthesis of three ATP will require $8 + 3 = 11$ protons. Hence, the respiratory chain could maximally produce slightly in excess of 30 ATP in mammalian cells.

Under conditions close to what might be expected physiologically in a mammalian cell (pH 7; 10 mM Mg^{2+} ; 10 mM phosphate), the equilibrium constant for the reaction



is approximately $10^{-5} (M^{-1})$, that is, strongly in favor of ADP with a positive ΔG_o of $\approx +29 \text{ kJ} \cdot \text{mol}^{-1}$. As stated above, the maximum PMF is typically around +210 mV, which in the same units is $\approx 20 \text{ kJ} \cdot \text{mol}^{-1}$. Since 3.66 protons are used to produce one ATP in the cytoplasm (ATP_{cyt}), the fully coupled overall reaction is



This would have a net ΔG of around $-45 \text{ kJ} \cdot \text{mol}^{-1}$, so that the ATP synthase will drive the reaction very much in favor of ATP. In practice, the PMF in active, living cells is not held at this maximum value and, in addition, there are processes that cause some of the protons to leak back across the membrane without passing through the ATP synthases. Hence, the yield of ATP, and the ratio of ATP/ADP achieved, is somewhat lower. Nevertheless, mitochondria have been shown to be able to drive the cytoplasmic ATP/(ADP.phosphate) ratio under similar conditions to at least 10^5 , with ATP/ADP around 10^3 .

Similar calculations of ATP yield and maximum phosphate potential can be made for other systems provided the proton

stoichiometries of the components concerned, and the magnitude of the PMF, are known. Figure 4 summarizes the current view of the PMF-generating reactions during electron transfer from oxygen to NADP in higher plant chloroplasts. For the plant thylakoid F_0F_1 ATP synthase, the most recent structural studies suggest that there are 14 c subunits in its c ring rotor. Since the catalytic F_1 domain is located in the stromal phase where ATP is required, this means that 14 H^+ are required for generation of three ATP at their site of usage. Bacterial F_0F_1 ATP synthases can have c rings with up to 15 subunits. Presumably, these different stoichiometries are tailored to balance an optimal ATP/ADP ratio (which can be driven higher with higher H^+ /ATP coupling ratios) with the net yield of ATP (which decreases as the H^+ /ATP coupling increases).

Transport Systems

A further major insight provided by the chemiosmotic theory is the role played by the PMF in active transport across the chemiosmotic membrane. The lipid membranes themselves are relatively impermeable not only to protons, but also to other charged or hydrophilic ions and metabolites. Transport proteins present in the same membrane are required to catalyze transport across the membrane. Mitchell recognized that the pH and electrical components of the PMF could be used to drive such transport processes provided that the transport proteins have the required properties.

If such transporters result in a net charge change, then the exchange can be driven by the electric field component of the PMF. An example is the adenine nucleotide translocator that catalyzes exchange of ATP (with four negative charges) with ADP (with three negative charges) across the inner mitochondrial membrane (Figure 5). Because the $\Delta\psi$ component of the PMF is negative in the matrix phase, it drives the export of the more negatively charged ATP into the cytoplasm and the ADP into the matrix. This results in a higher ATP/ADP ratio in the cytoplasm than in the mitochondrial matrix.

If the transporter is coupled to a net proton exchange but is overall electroneutral, then the metabolite can be actively

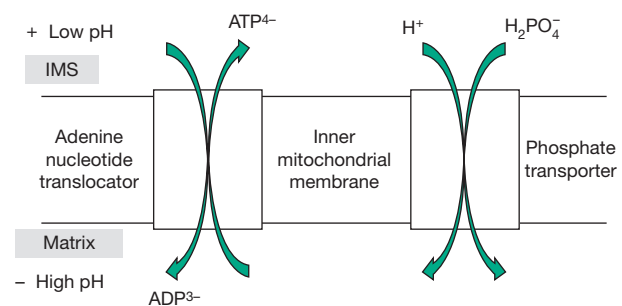


Figure 5 Mitochondrial transport of ADP, ATP, and inorganic phosphate. The mitochondrial inner membrane separates the internal matrix space and the intermembrane space (IMS). The adenine nucleotide translocator catalyzes exchange of one ATP for one ADP. Because ATP^{4-} has one more negative charge than ADP^{3-} , the exchange is electrogenic and will be driven by the electric field in the direction shown. The phosphate transporter catalyzes the transport of $H_2PO_4^-$ with one proton and is electroneutral. It will be driven in the direction shown by the ΔpH component of the PMF.

transported against its concentration gradient, driven by the pH difference across the membrane. An example of this is the phosphate transporter of the inner mitochondrial membrane. This couples transport into the mitochondrial matrix of inorganic phosphate (as H_2PO_4^-) with the co-transport of a proton (or, possibly, with the transport in the opposite direction of a hydroxide). The overall effect is that phosphate transport is electroneutral and is driven by the ΔpH component of the PMF. Bacteria display a wide range of proton-coupled transport systems for active uptake of nutrients. A particularly well-studied example is the lactose transporter of *Escherichia coli*. This co-transporters one H^+ with each lactose so that the lactose can be actively accumulated using both components of the PMF.

Further Aspects

The chemiosmotic theory has provided a sound theoretical basis for understanding of the mechanisms and energetics of energy coupling in respiratory and photosynthetic systems and in membrane-transport processes. New enzymes whose actions contribute to PMF generation, particularly in bacteria, are still being discovered. It is also now known that some bacteria can use a sodium ion gradient in place of a proton gradient to synthesize ATP. The sodium ion gradient can be formed from a proton gradient by the action of transporter that catalyzes electroneutral H^+/Na^+ exchange. However, some bacterial membrane enzymes couple electron transfer or other metabolic reactions to the direct transmembrane movement of sodium ions. Some bacteria, the alkaliphiles, can survive under conditions of extremely high pH; how they manage to do so while maintaining a PMF across their plasma membrane is still not fully understood. Another area of active interest is the finding that animal tissues tend to lose a significant part – up to 50% in some cases – of their proton current as heat. This can occur by slippage within the protonmotive enzymes

themselves, by possible direct proton leak across the lipid membrane at high PMF values and in some systems by the presence of uncoupler proteins that catalyze proton leaks. Besides having a role in thermogenesis in specialized tissues, their other functions are still not fully understood but may include regulation of levels of damaging reactive oxygen species that can form as side products if the PMF rises too high and roles in cellular signaling pathways.

See also: [Bioenergetics: F-ATPases](#); [Mitochondrial Channels](#); [Uncoupling Proteins](#); [Metabolism Vitamins and Hormones: Photosynthesis](#); [Tricarboxylic Acid Cycle](#).

Further Reading

- Brand MD (2005) The efficiency and plasticity of mitochondrial energy transduction. *Biochemical Society Transactions* 33: 897–904.
- Mitchell P (1961) Coupling of phosphorylation to electron and proton transfer by a chemi-osmotic type of mechanism. *Nature* 191: 144–148.
- Mitchell P (1966) *Chemiosmotic Coupling in Oxidative and Photosynthetic Phosphorylation*. Bodmin, UK: Glynn Research Ltd. (Republished in full in *Biochimica et Biophysica Acta* (2011) 1807: 1507–1538.)
- Mitchell P (1968) *Chemiosmotic Coupling and Energy Transduction*. Bodmin, UK: Glynn Research Ltd.
- Nicholls DG (2010) Mitochondrial ion circuits. *Essays in Biochemistry* 27: 25–35.
- Nicholls DG and Ferguson SJ (2001) *Bioenergetics* 3. London: Academic Press.
- Palmieri F and Pierri CL (2010) Mitochondrial metabolite transport. *Essays in Biochemistry* 27: 37–52.
- Prebble J and Weber B (2003) *Wandering in the Gardens of the Mind. Peter Mitchell and the Making of Glynn*. Oxford: Oxford University Press.
- Rich PR (2008) A perspective on Peter Mitchell and the chemiosmotic theory. *Journal of Bioenergetics and Biomembranes* 40: 407–410.
- Rich PR and Maréchal A (2010) The mitochondrial respiratory chain. *Essays in Biochemistry* 27: 1–23.
- Rich PR and Maréchal A (2011) Respiratory chains: Structures, mechanisms and energy coupling. In: *Comprehensive Biophysics*, vol. 8, ch. 6. The Netherlands: Elsevier.

Chemistry of Alzheimer Disease

G H Sack Jr., Johns Hopkins University, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Amyloid Fibrillar protein arrays that characteristically and diagnostically bind Congo red or Thioflavin T dyes and show birefringence in polarized light.

Amyloid precursor protein (APP) A large polypeptide with three isoforms. A small, partially intramembrane domain must be proteolytically cleaved to release the 'beta protein'.

APOE-ε4 An allele of the human APOE protein. The presence of two copies of this allele (homozygosity) is associated with an approximately eightfold increased risk of Alzheimer disease.

Beta protein (Aβ) A 40–42 amino acid polypeptide derived by proteolytic cleavage of APP. This polymerizes to form amyloid fibrils in Alzheimer disease.

Dementia Loss of skills related to thinking (cognition) and activities of daily living.

Down syndrome A disorder characterized by three copies of the genetic material from human chromosome 21 (normal

individuals have two copies), usually seen with three (rather than two) copies of the entire chromosome. Characteristic physical changes and a varied spectrum of mental retardation are prominent features.

Fibril A thread-like entity comprised of multiple monomers. Fibrils in Alzheimer disease are highly ordered and have a diameter of ~10 nm.

Plaque A deposit in nerve tissue largely comprised of amyloid (Aβ) fibrils and found in association with dementia.

Presenilin Protease responsible for cleavage of APP to yield Aβ.

Tangle A deposit in nerve cells associated with dementia with Tau protein as a major constituent.

Tau A family of protein isoforms whose normal function(s) involve assembly and stabilization of neurofilaments. Tau proteins are the main constituent of neurofibrillary tangles.

Introduction

Alzheimer disease (AD) is the most commonly recognized form of dementia in the US. Establishing the diagnosis can be difficult because the accepted definition relies on examination of brain tissue and this is usually available only after death. Characteristic features include slowly progressive dementia and diffuse atrophy of the cerebral cortex. It is important, however, to appreciate that the clinical endpoint of dementia may be a final pathway for at least three distinct problems – in addition to AD itself, vascular disease (particularly related to small vessel changes), and Lewy body disease. Thus, dissecting the mechanism(s) of AD alone may be impossible or at least very difficult when based on clinically defined individuals.

Pathologic Changes

Brain tissue from affected individuals shows distinctive changes. Cholinergic neurons are the site(s) of most visible changes and dysfunction. Two types of changes are common – amyloid plaques and neurofibrillary tangles (see Figure 1). Chemical features of these findings have been studied extensively.

Amyloid plaques comprise fibrils ~10 nm in diameter whose main protein constituent is the small protein 40–42 aa, the so-called β protein (see below). They show characteristic apple green birefringence when stained with Congo red or thioflavin T and viewed under polarized light (this feature is common to all types of amyloid fibrils). These fibrils are found extracellularly as well as in microvessels of cerebral tissues and meninges. Within neuronal regions, plaques are generally associated with dystrophic neurons (often containing tangles) as well as reactive glial cells.

Interestingly, nonfibrillar deposits of the β protein may be found prior to the appearance of plaques with their aforementioned fibrillar and tinctorial features. These deposits are generally amorphous and seen in areas (e.g., the cerebellar molecular cortex) not known to be affected in AD. Such diffuse deposits are common in older humans and other primates but are generally not associated with cellular or physiologic evidence of neuronal damage.

Neurofibrillary tangles are intracellular collections of paired filaments derived from tau proteins. These are either straight or paired in helices (so-called paired helical filaments or PHFs). They are not unique to AD but are also found in other disorders of cortical neurons including Pick disease and the parkinsonism-dementia complex of Guam. In other neurodegenerative disorders including progressive supranuclear palsy, corticobasal degeneration and argyrophilic brain disease tangles occur in both neuronal and glial cells.

Understanding biochemical changes responsible for these changes is not yet complete but much important information is known about the proteins involved. The final picture likely results from interaction(s) of multiple factors.

Beta Protein (Aβ)

The most prominent component of amyloid plaques isolated from the brain is a 40–42-amino-acid hydrophobic polypeptide that has been called the beta protein (Aβ). Aβ₄₂ is more likely to form amyloid fibrils than Aβ₄₀. Aβ is a normal secreted polypeptide of unknown function. It can be detected in cerebrospinal fluid and is cleared through vascular pathways. Aβ is derived by specific endoproteolytic cleavage of a much larger

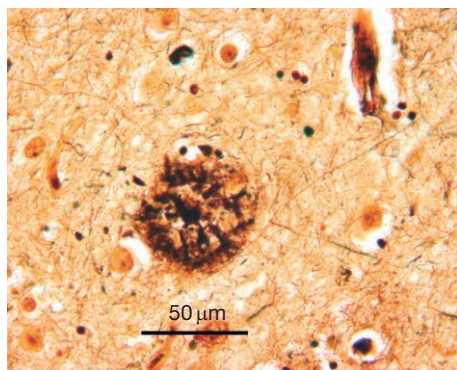


Figure 1 Brain tissue from an individual with Alzheimer disease. The central object is a neuritic plaque. A prominent group of neurofibrillary tangles is at the upper right. Bielschowsky silver stain. Photo courtesy of Dr. Juan Troncoso.

precursor, referred to as the amyloid precursor protein (APP, see **Figure 2**). APP contains a large extracellular domain, a membrane-spanning domain, and a relatively short intracellular domain. There are three APP isoforms of varying length, derived from the same gene by differential splicing. Recent data indicate that APP is important in presynaptic regions for neurotransmission although precise definition of its function(s) has been elusive. $A\beta$ levels in the interstitial fluid of the nervous system are increased with sleep deprivation.

Genetic Relationships with APP

An important correlation is that the gene for APP resides on human chromosome 21 (locus at 21q21) and essentially all individuals with Down syndrome (who have three copies (trisomy) of chromosome 21) develop neuropathic changes of AD by age 40 (considerably earlier than seen otherwise). Individuals with APP gene duplications (independent of trisomy 21) have also been identified and they have pathologic brain changes similar to those seen in AD (these are occasionally limited to association with brain hemorrhage). Over 20 missense mutations have been identified in the APP protein and are associated with a subset of AD known to be familial with a relatively early onset. Most of these mutations flank the $A\beta$ region (**Figure 2**) and may affect proteolysis. Some, however, are within the $A\beta$ sequence itself and likely affect fibril formation. It is important to note that most mutations identified in familial forms of AD are not related to the APP gene.

Proteolysis

The small size of the $A\beta$ protein itself compared with its long APP precursor(s) led to efforts to characterize the responsible proteases. These are known as the β - and γ -secretases; their endoproteolytic cleavage sites are different. The β -secretase cleaves at the N-terminus of what will become the $A\beta$ protein while the γ -secretase cleaves within the transmembrane region and defines the C-terminus of the $A\beta$ molecule (see **Figure 2**). In addition, an α -secretase cleaves within the $A\beta$ protein but outside the transmembrane domain and thus cleavage by the α -secretase prevents formation of the full-length $A\beta$ protein.

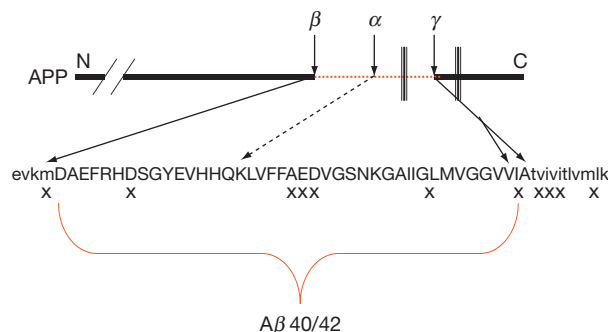


Figure 2 Formation of the $A\beta$ protein from the APP precursor involves specific endoprotease cleavages (β - and γ -secretases) as shown. Vertical lines denote membrane boundaries. Note that α -secretase cleavage would interrupt the $A\beta$ sequence. Site(s) of recognized mutations are indicated (x).

Presenilins

Identifying proteins related to AD has been aided by studying families segregating early onset AD but who show no genetic relationship to APP. Presenilin-1 (on chromosome 14) forms part of the aspartyl protease complex serving as the γ -secretase. Presenilin-2 (on chromosome 1) is closely related to presenilin-1. Mutations in presenilin-2 are also associated with familial transmission but are less common. The β -secretase is also an aspartyl protease but no pathologic mutations have yet been associated with it.

Neurophysiologic study indicates that presenilins affect presynaptic nerve termini. Inactivating presenilins alters both neurotransmitter release and induction of long-term potentiation. These effects were not seen when presenilins were inactivated in postsynaptic termini. This is consistent with finding (noted above) that APP is involved in presynaptic physiology.

A recent report (Green et al.) showed that using a γ -secretase modulator (γ -secretase cannot be totally inhibited; doing so suppresses Notch signaling) did not alter the progression of AD dementia.

Amyloid Fibril Formation

The characteristic fibrillar arrays of amyloid are formed from $A\beta$ monomers. $A\beta_{42}$ more readily forms fibrils than $A\beta_{40}$. $A\beta$ fibrils share features common to other types of amyloid including staining with dyes such as Congo red and thioflavin T. The fibrils have a cross- β structure and a diameter of ~ 10 nm and are easily seen by electron microscopy (**Figure 3**). The chemical and biophysical properties of amyloid fibrils are likely central to the pathologic states in which they are found. Fibril assembly, fibril structure, and physiologic consequences of fibrils within and around cells have been studied intensively.

Fibril assembly and growth occur through gradual addition of monomers. Intracerebral injection of $A\beta$ -containing brain extracts from humans with AD or mice transgenic for β -APP led to cerebral β -amyloidosis in APP transgenic mice in a time- and concentration-dependent manner. The findings depended on integrity of the introduced $A\beta$ as well as the status of the recipient mice (i.e., they had to be transgenic for APP). This is consistent with a seed model such as has been proposed for



Figure 3 Electron microscopy of amyloid fibrils from Alzheimer brain. Fibril diameter is ~ 10 nm.

prion-induced neuropathy where a (likely) specific three-dimensional (3-D) structure underlies the inducibility. It also emphasizes the cooperative nature of the process, since non-transgenic mice do not develop the changes. Intercellular transmission of protein aggregates is likely important in pathologic deposits. It is also important to note that injecting brain extracts of misfolded tau proteins (see below) into brains of mice was followed by a spreading pattern of misfolded and aggregated endogenous tau.

Obtaining molecular details of amyloid fibril structure has been difficult. While it has been clear that a cross- β structure is central to fibril formation, stability, and characteristic interaction(s) with planar dyes, all of the atomic interactions are not yet defined. Fibrils are generally too large for crystallographic resolution and their hydration states influence many details. Nevertheless, several features are consistent between studies: (1) N-terminal residues 1–17 of $A\beta_{42}$ are not structured in studies to date, (2) two amino acid stretches (i.e., 18–26 and 31–42) can form β -sheets, (3) two molecules of $A\beta_{42}$ interact to form a protofilament, and (4) the strands in the protofilament are oriented perpendicular to the long axis of the visible fibril.

Lührs et al. proposed a fibril model based on solid-state and quenched hydrogen/deuterium exchange nuclear magnetic resonance analysis. As shown in **Figure 4(a)**, their proposed repeat unit is a bimolecular protofilament. The critical associating region relies on a β -strand-turn- β -strand motif with two intermolecular parallel β -sheets ($\beta 1$ = residues 18–26; $\beta 2$ = residues 31–42). Residues 1–17 are disordered in the structure. Intermolecular side-chain contacts are formed between the odd-numbered residues of $\beta 1$ of the n th molecule and the even-numbered residues of $\beta 2$ of the $(n-1)$ th molecule. As shown, the β strands are oriented perpendicular to the long axis of the fibril. Intermolecular interactions are stabilized by hydrogen bonding and salt bridges. The fibril is polarized with even and odd ends, and growth is toward the odd end. A newly added monomer is less stably connected and, thus, adding yet another monomer provides additional bond energies. This structure implies a sequence-specific, unidirectional, cooperative model for fibril extension consistent with first-order kinetics.

More recently, Schmidt et al. evaluated mass-per-length (MPL) data and electron cryomicroscopy with 3-D reconstruction to assess details of homogeneous fibril populations. They confirmed earlier data suggesting ~ 2.5 $A\beta_{42}$ monomers per protofilament repeat, a number not compatible with the model of Lührs et al. Their model for the protofilament

(**Figure 4(b)**) does not rely on the β -strand-turn- β -strand nucleus but, rather, aligns the C-terminal regions across from each other to form a cross- β structure and leaves the N-terminal domains unstructured (within the resolution of their technology). To address the apparently anomalous MPL figure of 2.5, they propose that additional monomers might periodically decorate the core dimer and/or that peripheral monomers could swap domains with the core dimer at a frequency of ~ 0.5 exogenous monomers per protofilament dimer. Interestingly, their data for $A\beta_{40}$ fibrils indicate that the repeat unit could comprise two of the structures in the $A\beta_{42}$ model as shown.

Although current technology cannot define the precise details, these two models emphasize the dimer as the basic structural unit. The dimers in each model are stabilized by cross- β interactions, as has been noted in all amyloid fibril models. Both models address the possibilities for self-assembly and impressive stability of the fibrils. In most physiologic situations, amyloid fibrils can function as molecular sinks, continually adding monomers, and are very stable in biologic time and environments.

Interestingly, in a mouse model, neurotoxicity appears related to smaller $A\beta_{42}$ oligomers that interfere with synaptic transmission and memory. Structural data on these oligomers are not available. Fibrils (and visible amyloid formation) may accumulate later in the course for these mice. It is important, however, to recall the short life span of the mouse versus the later onset of changes in humans. Mice may not live long enough to develop the full spectrum of pathologic changes, including fibrils.

Tau Proteins

Tau proteins are the primary constituent of the intracellular neurofibrillary tangles. There are six tau isoforms in the brain, all derived from a single gene (see **Figure 5**). Tau proteins are important in assembling and stabilizing microtubules and they form two groups as determined by the number of microtubule-binding repeats (either three or four). The population of isoforms in AD brains appears to be the same as in normal brains.

Formation of filaments from tau proteins is related to serine/threonine phosphorylation of individual tau monomers. This can be extensive; 79 modification sites have been identified in the longest isoform. Phosphorylation can regulate binding of tau to microtubules. Thirty phosphorylated residues have been identified in AD and the proline-directed kinase GSK3 is an important mediator of this modification. Tau filaments are not unique to AD; they are recognized in other neurodegenerative disorders as well (see **Table 1**). Some of these disorders have hyperphosphorylation of distinct tau isoforms; others do not. Pick disease, another neuropathy, is related to a mutation in presenilin-1 that is associated with tau hyperphosphorylation (probably through reduced cadherin binding, reduced PI3K function and increased GSK3 activity) but not AD.

Tau Mutations

Tau mutations (39 known) have been identified in association with specific forms of neurodegenerative disease, classified as tauopathies (see **Table 1**). Two different categories of tau mutations are recognized: (1) mutations affecting splicing and

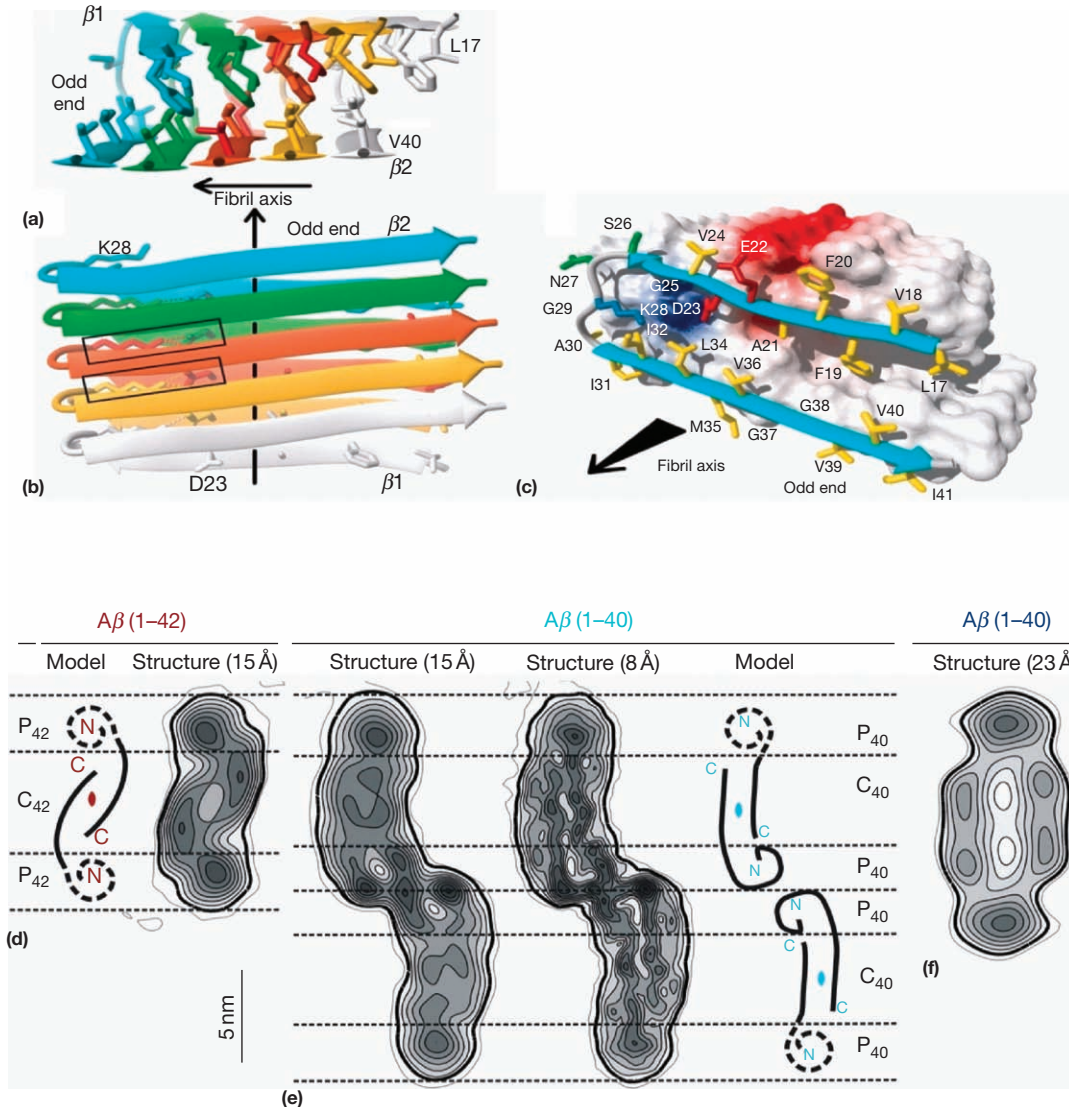


Figure 4 Proposed Aβ₄₂ protofibril structures. (a, b) Ribbon diagrams of proposed core (aa 17–42) showing intermolecular nature of inter-β-strand interactions. Arrows indicate β strands. Rectangles indicate salt bridges formed by central part of molecules. (c) van der Waals surface and ribbon diagram at the odd end of the Aβ protofibril aa 17–42. Arrow indicates direction of fibril axis (even to odd). (d) Density map of Aβ₄₂ (right) and possible peptide placement (left), based on interpretations shown in (e). (e) Cross sections of 20 nm Aβ₄₀ fibril density maps at two densities. F: 13 nm Aβ₄₀ fibril density map. Note unstructured N-terminal region and β-strand interactions for two Aβ monomers. (a–c) Reproduced from Lührs T, Ritter C, Adrian M, et al. (2005) 3D structure of Alzheimer's amyloid-β(1–42) fibrils. *Proceedings of the National Academy of Sciences of the United States of America* 102: 17342–17347, copyright (2005) National Academy of Sciences, with permission from PNAS. (d, e) Reproduced from Schmidt M, Sachse C, Richter W, et al. (2009) Comparison of Alzheimer Aβ(1–40) and Aβ(1–42) amyloid fibrils reveals similar protofibril structures. *Proceedings of the National Academy of Sciences of the United States of America* 106: 19813–19818, with permission from PNAS.

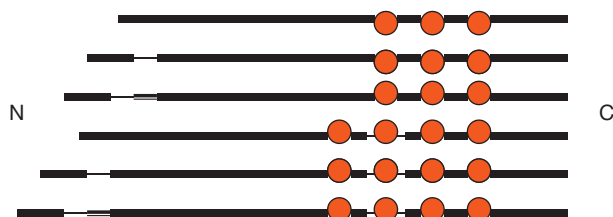


Figure 5 Tau protein family showing repeat units and phosphorylation patterns. Red circles indicate microtubule-binding repeats. Heavy lines are constant regions, thin/alternative lines represent alternative splicing.

(2) mutations causing protein changes. Most of the latter affect the microtubule-binding domain. Some mutations enhance filament formation. Mutations within Exon 10 or adjacent introns are associated with higher proportions of tau with four microtubule-binding repeats. This distortion from the normal ratio of tau isoforms apparently enhances the likelihood of neuropathic changes. Different mutations are associated with different filament structures and involve different cell types. Tau proteins may be related to brain development. Learning disabilities have been found in individuals with deletions of 17q21, a region including tau and other genes.

Table 1 Neurodegenerative disorders

Alzheimer disease
ALS/Parkinson–dementia
Argyrophilic grain dementia
Corticobasal degeneration
Creutzfeldt–Jakob disease
Dementia pugilistica
Diffuse neurofibrillary tangles with calcification
Down syndrome
Frontotemporal dementia with parkinsonism (also Pick disease)
Gerstmann–Sträussler disease
Motor neuron disease with neurofibrillary tangles (non-Guamanian)
Myotonic dystrophy
Niemann–Pick disease, type C1
Pantothenate kinase-associated neurodegeneration
Parkinsonism (postencephalitic)
Prion protein cerebral amyloid angiopathy
Progressive subcortical gliosis
Progressive supranuclear palsy
Subacute sclerosing panencephalitis
Tangle-only dementia

Enhanced microscopic tau and filament changes have been reported in cell lines expressing tau mutations that have been crossed with lines expressing APP mutants. By contrast, mice without tau proteins ($\tau^{-/-}$) resisted excitotoxicity associated with $A\beta$ in mice transgenic for the latter.

Risk Factors for AD

Individuals whose AD can be attributed to APP or presenilin mutations account for $\sim 2\%$ of those affected and thus other associations have been sought. The most prominent risk factor for what appears to be sporadic AD is the presence of the $\epsilon 4$ allele of apolipoprotein E (*APOE- $\epsilon 4$*). Homozygosity for this allele imparts an approximately eightfold increase in the likelihood of having AD and of developing changes earlier in life. β -Amyloid fibrils are found more frequently in brains of $\epsilon 4$ homozygotes.

About 25% of individuals with late-onset AD have had a close relative with dementia. A genome-wide association study based on single-nucleotide polymorphism analysis of individuals from 410 AD families confirmed the *APOE- $\epsilon 4$* association and also implicated a locus on 14q31. Although the molecular identity of the locus is not yet known, it appears to affect age of onset of clinical AD changes, thus providing evidence for additional genetic contributions to the neuropathology. The strength of this association is lower than that for *APOE* but it is likely to be a valid risk factor with, probably, additive effect. At least two other loci were identified but must be considered tentatively associated. These studies require very large databases to detect low-level effects although the results support the notion that multiple variation(s) can contribute to AD pathology.

Increasing age is a known risk factor, but it is more difficult to interpret because visible microscopic brain changes may precede clinically evident dementia by several decades. The first visible changes are generally tau filaments but in the neocortex $A\beta$ deposits (not fibrils) are often seen before tau filaments. This is especially prominent in the (rare) individuals

who have underlying APP mutations leading to an excess of $A\beta_{42}$. By contrast, tau mutations are not associated with β -amyloid formation.

Biomarkers and Detection Approaches

Although the histologic changes of plaques and tangles are diagnostic of AD, they are not generally detected during life because tissue is not available. Autopsy studies have shown, however, that such changes can be detected prior to cognitive decline – sometimes 10 or more years earlier. Thus, for both patient care and improved understanding of early events in AD other, noninvasive, approaches to presymptomatic detection have been sought. Serial analysis of high-resolution quantitative magnetic resonance imaging (MRI) images can be used to distinguish AD changes from healthy age-related losses. Positron emission (PET) scanning using the compound ^{11}C -PIB has been used to assess $A\beta$ in the nervous system. When combined with measuring $A\beta_{42}$ levels in the spinal fluid, this has shown reasonable discrimination of AD in early studies. In addition, atrophic brain regions show reduced glucose metabolism by FDG PET.

Measuring $A\beta_{42}$ levels in spinal fluid shows some promise for correlation with amyloid plaques (the presumption is that $A\beta_{42}$ monomers are readily partitioned into fibrils and, hence, monomer levels may be lower elsewhere). Measuring plasma $A\beta_{42}$ levels has not proved useful. Higher levels of both tau and phosphorylated tau proteins in the spinal fluid are seen in individuals with AD than in controls. Sampling spinal fluid is not a simple procedure, however, and doing so is not easily adaptable to large studies. For example, in a large study at 12 centers, measuring spinal fluid levels of $A\beta_{42}$, total tau, and tau phosphorylated at thr 181 could identify incipient AD in individuals with mild cognitive impairment with a sensitivity of 83% but standards were difficult to establish between centers and the approach was not endorsed for routine clinical use. Recently, high blood leptin levels were associated with a reduced incidence of dementia and improved cerebral volume. This finding, if validated, could provide a serum marker for prognosis as well as implicate other pathways in pathogenesis. Other possible biomarkers including proteomics assays have not yet been validated.

Pathophysiology of Alzheimer Changes

It is important to emphasize that all details of Alzheimer neuropathic development are not understood. The histologic changes found in the brains of affected individuals are generally observed late in the course of the progressive dementia, considerably after the cognitive neurologic features have appeared (recall that most studies have been based on autopsy findings). While the visible changes are recognizable, the earlier biochemical events leading to their development are not established. Although specific mutations are recognized in a subset of individuals, they are infrequent and polygenic aspects of AD risk are apparent (see above). Most of the underlying variation(s) likely contribute relatively small amounts to the total risk. In addition, the changes occur over many years (50 or more). Thus, measuring late molecular changes may not identify the primary events and could be misleading. In

addition, animal models such as the mouse cannot be assumed to reflect a simple acceleration of human pathophysiology.

It remains likely that at least some of the brain changes are not unique to what has been called AD. Rather, the classic plaques and tangles may develop relatively late and thus mark common endpoints of multiple pathways of dysfunction (this possibility has been proposed as the so-called tombstone hypothesis).

Several categories of dysfunctional changes have been studied extensively. These categories are not necessarily exclusive.

A β toxicity and mitochondrial integrity

Mitochondria in AD brains are small with fragmented cristae, consistent with excessive mitochondrial fission. Modifying the fission-inducing protein, dynamin-related protein 1 (Drp1) is associated with damage to synapses and neurons. S-Nitrosylation of Drp1 (apparently at Cys644) leads to dimerization and increased guanosine triphosphatase activity, accelerating mitochondrial fragmentation. Physiologically relevant levels of S-NO-Drp1 are found in AD brains. Copper (Cu²⁺) may mediate this reaction and copper levels have been found to be elevated in AD brains (see below). The S-NO-Drp1 appears necessary for nitric-oxide-mediated neuronal damage, and oligomers of A β induce mitochondrial fission and neuronal damage in an NO-mediated manner. Synaptic damage occurs early with loss of dendritic spines. In Drp1 mutants lacking the presumed nitrosylation site (C644A), the A β effect on the spines is blunted. This pathway is consistent with the measured levels of constituents and reactants but does not explain why A β levels are elevated to initiate the process.

Copper

The N-termini of both APP and A β contain high-affinity Cu²⁺-binding domains. The copper-binding site is structurally similar to the site on superoxide dismutase (which has femtomolar copper affinity). A β becomes redox active in excess metal environments. Excess APP or its C-terminal 100 residues can bind copper and thus decrease free copper levels in transgenic mice. Could A β be protective and lessen the effect of copper excess?

Copper levels rise in the aging brain and plaques and Tanzi et al. have proposed that the combination of A β and copper can lead to both H₂O₂ production via a redox mechanism as well as aggregation. Copper and zinc can coordinate A β and precipitate A β at submicromolar concentrations.

Cholesterol

There are epidemiologic associations between high cholesterol levels and AD. Cholesterol itself is a reducing agent for A β -Cu²⁺ and this combination may potentiate formation of reactive oxygen species. This suggests possible synergy between copper and cholesterol. Lowering cholesterol levels with statins can modify vascular risk factors which, as noted above, may contribute to the phenotype of dementia but such treatment is generally begun long before the emergence of neurologic changes late in life. The established relationship to the APOE gene was noted above. A recent study showed that the homozygosity for the isoleucine $\rightarrow$ valine (I405V) change in the cholesterol ester transfer protein was associated with both slower

memory decline and a lower incidence of dementia and AD risk in a prospective cohort study of adults 70 or older without dementia.

Calcium

Regulation of intracellular calcium concentrations has been affected in several model systems. A β has been suggested to form ion channels transporting calcium into neurons. Alternatively, it may distort membranes to produce the same effect. It has also been noted that A β increases calcium influx upon N-methyl-D-aspartate (NMDA) receptor stimulation by glutamate.

Presenilin mutations lead to increased calcium release from endoplasmic reticulum (ER) stores after glutamate stimulation. Another explanation may be that the mutants increase ER calcium levels which subsequently can be released upon cell stimulation. A third possibility is that presenilin mutations interact with other channels – IP₃ and/or ryanodine – to increase calcium release.

Integration

AD pathologic change is a complex process with multiple, independent as well as interacting factors. Lack of neuronal integrity likely leads to apoptosis with mitochondrial fragmentation. In addition, early events can cause an inflammatory response, likely mediated by glial cells, with high local levels of acute-phase proteins and cytokines. Both the aggregates of tau proteins and fibril formation by A β are likely to be relatively late changes in the cascade of neurologic dysfunction. Once begun, however, both relatively disordered aggregates (tau) and highly ordered fibrils (A β) are likely permanent in biologic time. Clarifying the chemistry of the early phases of neuronal dysfunction, which are likely multifactorial, should suggest approaches for minimizing the end-stage findings of advanced AD.

See also: [Cell Architecture and Function: Neuronal Intermediate Filaments](#); [Protein/Enzyme Structure Function and Degradation: Amyloid](#).

Further Reading

- Avila J (2006) Tau phosphorylation and aggregation in Alzheimer's disease pathology. *FEBS Letters* 580: 2922–2927.
- Bertram L, Lange C, Mullin K, et al. (2008) Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. *American Journal of Human Genetics* 83: 623–632.
- Bird TD (2008) Genetic aspects of Alzheimer disease. *Genetics in Medicine* 10: 231–239.
- Bitan G, Kirkitadze MD, Lomakin A, et al. (2003) Amyloid β -protein (A β) assembly: A β 40 and A β 42 oligomerize through distinct pathways. *Proceedings of the National Academy of Sciences of the United States of America* 100: 330–335.
- Bush AI, Masters CL, and Tanzi RE (2003) Copper, β -amyloid, and Alzheimer's disease: Tapping a sensitive connection. *Proceedings of the National Academy of Sciences of the United States of America* 100: 11193–11194.
- Cho D-H, Nakamura T, Fang J, et al. (2009) S-Nitrosylation of Drp1 mediates β -amyloid-related mitochondrial fission and neuronal injury. *Science* 324: 102–105.
- Goedert M and Spillantini MG (2006) A century of Alzheimer's disease. *Science* 314: 777–778.
- Green RC, Schneider LS, Amato DA, et al. (2009) Effect of tarenflurbil on cognitive decline and activities of daily living in patients with mild Alzheimer disease – a randomized controlled trial. *Journal of the American Medical Association* 302: 2557–2564.

- Lee VM-Y, Goedert M, and Trojanowski JQ (2001) Neurodegenerative tauopathies. *Annual Review of Neuroscience* 24: 1121–1159.
- Lieb W, Beiser AS, Vasan RS, et al. (2009) Association of plasma leptin levels with incident Alzheimer disease and MRI measures of brain aging. *Journal of the American Medical Association* 302: 2565–2572.
- Lührs T, Ritter C, Adrian M, et al. (2005) 3D structure of Alzheimer's amyloid- β (1–42) fibrils. *Proceedings of the National Academy of Sciences of the United States of America* 102: 17342–17347.
- Mattson N, Zetterberg H, Hansson O, et al. (2009) CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *Journal of the American Medical Association* 302: 385–393.
- Meyer-Luehmann M, Coomaraswamy J, Bolmont T, et al. (2006) Exogenous induction of cerebral β -amyloidogenesis is governed by agent and host. *Science* 313: 1781–1784.
- Montine T and Sonnen J (eds.) (2009) Mini-forum on clinical-pathologic correlations in population- and community-based studies of brain aging. *Journal of Alzheimer's Disease* 18: 643–745.
- Perrin RJ, Fagan AM, and Holtzman DM (2009) Multimodal techniques for diagnosis and prognosis of Alzheimer's disease. *Nature* 461: 916–922.
- Sanders AE, Wang C, Katz M, et al. (2010) Association of a functional polymorphism in the cholesterol ester transfer protein (*CETP*) gene with memory decline and incidence of dementia. *Journal of the American Medical Association* 303: 150–158.
- Schmidt M, Sachse C, Richter W, et al. (2009) Comparison of Alzheimer A β (1–40) and A β (1–42) amyloid fibrils reveals similar protofilament structures. *Proceedings of the National Academy of Sciences of the United States of America* 106: 19813–19818.
- Selkoe DJ (1989) Aging, amyloid, and Alzheimer's disease. *New England Journal of Medicine* 320: 1484–1487.
- Sparks DL and Schreurs BG (2003) Trace amounts of copper in water induce β -amyloid plaques and learning deficits in a rabbit model of Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America* 100: 11065–11069.
- Tanzi RE (2005) Tangles and neurodegenerative disease – a surprising twist. *New England Journal of Medicine* 353: 1853–1855.
- Tycko R (2004) Progress towards a molecular-level structural understanding of amyloid fibrils. *Current Opinion in Structural Biology* 14: 96–103.
- Zhang C, Wu B, Beglopoulos V, et al. (2009) Presenilins are essential for regulating neurotransmitter release. *Nature* 460: 632–636.

Chemokine Receptors

T Sobolik-Delmaire, D Raman, and J Sai, Vanderbilt University School of Medicine, Nashville, TN, USA

G-H Fan and A Richmond, VA Medical Center, Nashville, TN, USA; Vanderbilt University School of Medicine, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Ann Richmond and Guo-Huang Fan, volume 1, pp. 413–418, © 2004, Elsevier Inc.

Glossary

Angiogenesis The growth of new vasculature from pre-existing blood vessels.

Chemotaxis Directed migration of leukocytes toward an inflammatory site.

Coreceptor Any of certain chemokine receptors that are used together with CD4 by human immunodeficient virus to enter the target cells.

Desensitization Attenuation of the functional response of chemokine receptors to their ligands following repeated ligand stimulation.

Internalization Translocation of chemokine receptors from the cell membrane to the cytoplasmic compartments.

Metastasis The migration of cancer cells to sites distant from the primary tumor.

Ligands

The ligands that bind and activate chemokine receptors are chemokines or chemoattractant cytokines. Approximately 50 chemokines have been identified, and are classified into four (CXC, CC, C, and CX3C) subfamilies based on their primary amino acid sequences (Table 1). The CXC subfamily has seven members, which have one amino acid (X) interrupting the first two of their four conserved cysteine residues. The CC subfamily has 10 members with no intervening amino acid between the first two of their four cysteine residues. In both the CXC and CC subfamilies, disulfide bonds are formed between the first and third cysteines and between the second and fourth cysteines to establish a stable tertiary structure with the molecular mass of 7–9 kDa. The C subfamily has two members that are 16 kDa in molecular size. The CX3C subfamily has only one member (CX3CL1), which has three intervening amino acids between the N-terminal cysteines, with a molecular mass of 38 kDa, larger than any other known chemokine. Chemokines have two main sites of interaction with their receptors, the flexible N-terminal region and the conformationally rigid loop that follows the second cysteine. Chemokines dock onto receptors by means of the loop region to facilitate the binding of the N-terminal region to the receptor resulting in receptor activation. Chemokines possess heparin-binding capacity at their C-terminal end enabling them to bind to glycosaminoglycans and other negatively charged sugar moieties on cell surfaces and matrix glycoproteins. This property may result in the adsorption of chemokines onto the endothelial cell lining of the blood vessels, connective tissues, and cell matrices. Thus, chemokines immobilized on tissue or matrix surfaces may induce haptotactic migration of target cells.

Structure and Cell Signaling

Chemokine receptors are seven-transmembrane G-protein-coupled receptors composed of approximately 340–370 amino acid residues with 25–80% identity. These receptors

share a common putative structural topology composed of seven hydrophobic transmembrane domains, an N terminus outside the cell surface, three extracellular and three intracellular loops, and a C terminus in the cytoplasmic compartment. The poorly conserved N-terminal domains together with a second binding site in the extracellular loops determine the specificity for ligand binding. The sequence DRYLAIVHA, or a variation of it, in the second intracellular loop and the third intracellular loop is required for G-protein coupling. A cysteine residue in each of the four extracellular domains is required for the disulfide-bond formation that is critical for cell-surface expression. The C terminus of the receptors contains a number of serine and threonine residues that, upon phosphorylation, are involved in signaling and receptor desensitization. A leucine–leucine or isoleucine–leucine motif in the C terminus is required for the receptor internalization (Figure 1). Some chemokine receptors such as CCR2, CCR5, and CXCR4 form homodimers, which may be needed for signal transduction.

Ligand binding to chemokine receptors initiates a cascade of intracellular events (Figure 2). Chemokines bind to the chemokine receptors in the amino terminus resulting in phosphorylation of serine/threonine residues in the carboxyl terminus, conformational changes of the receptor, and activation of the heterotrimeric G-protein complex bound in the intracellular domain of the receptor. Activation leads to the exchange in the α -subunit of the G proteins from the guanosine diphosphate to guanosine triphosphate (GTP)-bound state dissociating the α from the β and γ G-protein subunits. The $G\beta\gamma$ subunits activate phospholipase C (PLC) that hydrolyzes phosphatidylinositol 4,5-bisphosphate [(PIP₂] to produce inositol trisphosphate (IP₃) and diacylglycerol (DAG). IP₃ mediates Ca²⁺ release from the intracellular stores. DAG activates protein kinase C (PKC) resulting in phosphorylation of a number of effector molecules. The activation of isotopes of PLC and hydrolysis of PIP₂ vary among chemokine receptors and from cell type to cell type, accounting for the divergent cellular responses to a given chemokine. Another important signaling pathway of chemokine receptors is the activation of phosphoinositide 3-kinases (PI3Ks), which convert the plasma-membrane lipid PI (4,5)P₂ to

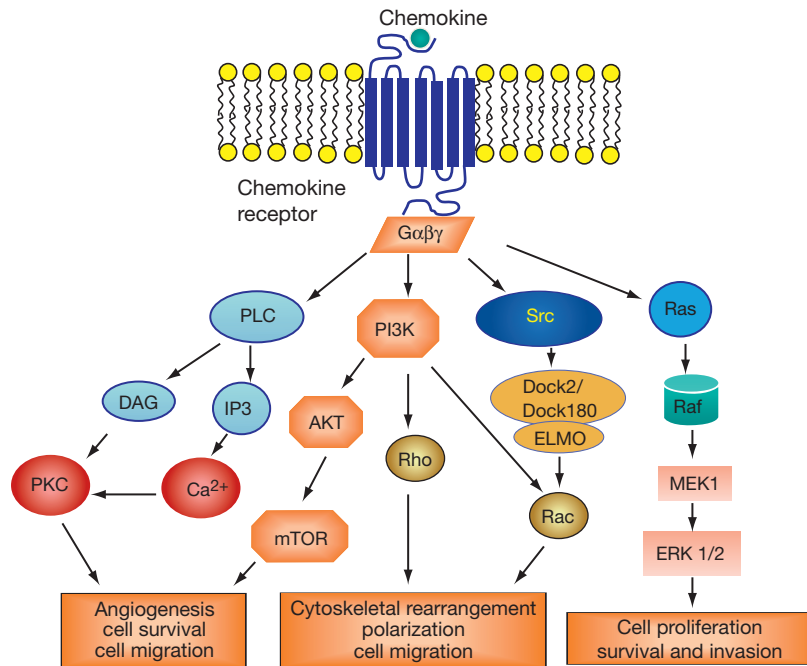


Figure 2 The major intracellular signaling events that are activated by stimulation of a chemokine receptor CXCR2 with the chemokine CXCL8. Stimulation of a chemokine receptor by its ligand results in a conformational change of the receptor and activation of the coupled G proteins. The $G\alpha$ proteins coupled with chemokine receptors are generally the inhibitory $G\alpha$ proteins ($G\alpha_i$). The β - and γ -subunits of the G-protein activate phospholipase C (PLC) which cleaves PIP₂ to produce diacylglycerol (DAG) and inositol triphosphate (IP₃). IP₃ initiates mobilization of the intracellular free Ca^{2+} and DAG activates protein kinase C (PKC). Stimulation also results in activation of nonreceptor tyrosine kinases such as Src, which leads to activation of Dock2/Dock180/Elmo complex and activation of Rac. CXCL8 signaling also regulates several pathways, including mitogen-activated protein kinase (MAPK) pathway, phosphatidylinositol 3-kinase (PI3K)/protein kinase B (Akt) pathway, as well as Rho-signaling pathway. These pathways are required for cytoskeleton dynamics and cell movement. ERK1/2, extracellular signal-regulated kinases 1 and 2; MEK, mitogen-activated protein kinase; Ras and Rho, small G proteins.

phosphatidylinositol-3,4,5-trisphosphate [PI(3,4,5)P₃], resulting in the phosphorylation and activation of the serine-threonine kinase Akt (PKB) by PDK-1 phosphorylation of Ser 308. An additional part of the signals that evoke the motility response involves activation of the Rho GTPase family including Rho A, Rac, and Cdc42. Rac and Cdc42 activate P21-activated kinase, Wiskott-Aldrich syndrome protein or Scar/WASP family Verprolin-homologous protein (WAVE), which, in turn, interact with cofilin or Arp2/3. Rho A mediates cell retraction by activating Rho kinase (ROCK) and Rhotekin, which then activate myosin light chain to form myosin-centered retraction of the uropod. $G\beta\gamma$ subunits also activate nonreceptor tyrosine kinases, Src-family kinases (SFK). In addition to functioning in the cell focal adhesion, SFK was also found to increase Rac activation through Dock180/ELMO1 complex. Chemokines also induce tyrosine phosphorylation of p130 Cas, focal adhesion kinase, and members of the src family in several cell types. Through activation of the Ras, Raf, mitogen-activated protein kinase (MAPK) cascade and through activation of the NF- κ B cascade, transcription factors are activated and gene expression is affected. The signaling pathways of chemokine receptors play a role in cell growth, cell migration, cell survival, and superoxide production. Chemokine receptors also activate signaling pathways independent of G proteins, including JAK/STAT and p38MAPK to regulate gene transcription and cellular migration.

In addition, another class of chemokine receptors that are structurally unable to elicit migration or signaling responses

acts as decoy receptors and scavengers to regulate inflammatory and immune reactions. This subfamily of nonsignaling chemokine receptors includes DARC (Duffy antigen receptor for chemokines), D6, and CCX-CKR (also known as CCRL1).

Desensitization

The ligand-induced activation of chemokine receptors is transient. The functional response of chemokine receptors is rapidly reduced following repeated stimulation by chemokines. This process is called desensitization. Agonist-dependent (homologous) desensitization occurs when a receptor binds chemokine and is phosphorylated by a G-protein receptor-coupled receptor kinase. The phosphorylated receptor recruits β -arrestin, an adaptor protein that uncouples the receptor with G proteins, thus retaining the receptor in a desensitized state, resulting in attenuation in cell excitability (Figure 3). This mechanism plays a major role in determining the duration of leukocyte trafficking, migration, or sequestration in certain situations (see later discussion). Chemokine receptors can also be phosphorylated by a kinase, such as PKC, activated by a different signaling cascade. This is called heterologous desensitization. For example, the chemokine receptor CXCR1 or CXCR2 can be phosphorylated and desensitized through the activation of formyl peptide receptors by the tripeptide formylmethionyl-leucyl-phenylalanine (fMLP). The importance of this phenomenon in inflammation

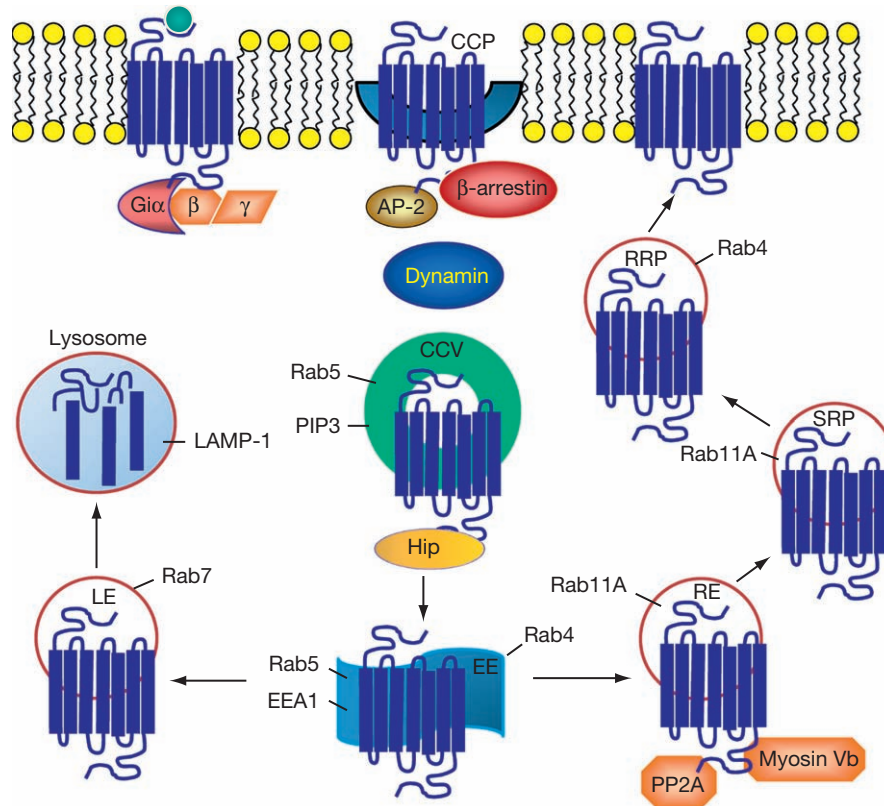


Figure 3 Homologous desensitization and intracellular trafficking of chemokine receptors. In response to chemokine stimulation, the receptor undergoes desensitization, in which receptor phosphorylation by a G-protein-coupled receptor kinase or another kinase leads to uncoupling of G proteins, binding of β -arrestins and adaptin-2 (AP-2), and subsequent internalization of the receptor through clathrin-coated pits. After pinching off from the membrane, the clathrin-coated vesicles fuse to early endosomes, where the chemokine receptors are dephosphorylated by protein phosphatase 2A and transported to late endosomes and lysosomes for degradation. After removal of the extracellular ligands, the receptors are transported from early endosomes to recycling endosomes to allow a reexpression of receptor at the cell surface. AP2, adaptin 2; CCP, clathrin-coated pit; CCV, clathrin-coated vesicle; EE, early endosome; LE, late endosome; PP2A, protein phosphatase 2A; Rab 4, Rab5, Rab11a, Rab7, small GTPases that regulate receptor trafficking; RE, recycling endosome; SRP, slow recycling pathway; RRP, rapid recycling pathway; LAMP-1, lysosomal-associated membrane protein-1.

and leukocyte trafficking is not completely clear, but the capacity for heterologous desensitization appears to be different among chemokine receptors, suggesting a hierarchy of chemokine/chemokine receptor interactions.

Internalization

After agonist occupancy, chemokine receptors are rapidly translocated from the cell membrane to the cytoplasm, in a process known as internalization. Chemokine receptors generally undergo internalization through clathrin-coated pits, a complex process that involves clathrin and adaptor proteins such as β -arrestins and adaptin-2 (AP-2). β -Arrestin binds to both the carboxyl terminus and third intracellular loop of the chemokine receptor. Studies have shown that AP-2 binds to Leu-Leu, Ile-Leu, and Leu-Ile motifs in the carboxyl terminus of CXCR2 and CXCR4. The association of the adaptor proteins with the chemokine receptor results in recruitment of clathrin and formation of clathrin-coated pits which pinch off from the cell membrane through a dynamin-dependent process to become clathrin-coated vesicles. The clathrin-coated vesicles containing the internalized receptors fuse with early endosomal compartments.

When there is a low pH value in the early endosomes, the receptors become dephosphorylated by a specific serine/threonine protein phosphatase such as protein phosphatase 2A. The dephosphorylated receptors are either transported to recycling endosomes and subsequently re-expressed on the cell surface after removal of the extracellular ligand or delivered to late endosomes and lysosomes for degradation in the continued presence of a high concentration of ligands. The intracellular trafficking processes are regulated by a number of molecules, including the Ras-like GTPases (Rabs) (Figure 3). Rab4 and Rab5 associate with the early endocytic compartment, while Rab11 associates with the perinuclear recycling compartment. Such trafficking may be important for both transmission and termination of the receptor signals and may play an important role in mediating cell chemotaxis (see later discussion).

Multiple Roles

Chemotaxis

Neutrophils, lymphocytes, and other immune cells migrate to the sites of injury and infection. This process, which is called chemotaxis, occurs through the dynamic response of

chemokine receptors expressed on these cells to the gradient concentrations of chemokines produced in the inflammatory sites. Chemokine receptor signaling leading to the establishment of cell polarity, cytoskeletal rearrangement, and interaction with the extracellular matrix plays an essential role in this complex process. In addition, the desensitization and internalization of chemokine receptors play a regulatory role. For example, human immature dendritic cells derived from monocytes express the inflammatory chemokine receptors CXCR1, CCR1, CCR2, CCR5, and CCR7 allowing these cells to follow chemotactic gradients to inflammatory sites. Once there, dendritic cells process antigen and become exposed to the maturation-stimulating cytokines. Maturing dendritic cells express large amounts of chemokines such as CCL3, CCL4, CCL5, CCL8, and CCL10. One consequence of this increased chemokine expression is the downregulation of chemokine receptors on maturing dendritic cells, particularly CCR1 and CCR5, by receptor desensitization and internalization. The second consequence of this upregulation of chemokines is that it strengthens the original chemotactic gradient, further boosting recruitment of immature dendritic cells, monocytes, and lymphocytes.

Human Immunodeficiency Virus Entry

Human immunodeficiency virus (HIV) is the causative agent of acquired immune deficiency syndrome (AIDS). Type 1 HIV (HIV-1) uses chemokine receptor as a coreceptor together with CD4 to enter the CD4-positive (CD4+) target cells. The main cells targeted by HIV-1 are T lymphocytes (T cells), macrophages, and dendritic cells. The chemokine receptors CXCR4 and CCR5 are the major coreceptors for T-cell line-tropic and macrophage-tropic HIV-1 strains, respectively, although many other chemokine and orphan receptors have been identified as potential coreceptors for HIV-1. The fusion of the viral envelope and the lymphocyte occurs when the envelope glycoprotein on the surface of virus particles (comprising a trimer of three gp120 and three transmembrane gp41 molecules) binds to CD4 triggering a structural change exposing a binding site for a coreceptor. Further structural rearrangements are initiated when the coreceptor is bound. These changes occur predominantly in gp41 and are sufficient to trigger fusion of viral and cellular membranes and entry of the virion core into the cytoplasm.

Angiogenesis

Angiogenesis is the formation of new blood vessels from pre-existing microvasculature. This biological process is critical to both physiological and pathological processes such as wound healing and tumor growth. Angiogenesis is regulated by an opposing balance of angiogenic and angiostatic factors. The CXC chemokines are a family of cytokines unique in their ability to behave in a disparate manner in the regulation of angiogenesis. The N terminus of the majority of the CXC chemokines contains three amino acid residues (Glu-Leu-Arg; the ELR motif). This motif, in general, determines whether these chemokines promote angiogenesis. Members that

contain the ELR motif (ELR+), such as CXCL1, CXCL2, and CXCL8, are potent promoters of angiogenesis. A novel link between CXCL1 and prostaglandin E2 has been shown to promote angiogenesis in a mouse model for colon cancer. By contrast, members that are inducible by interferons and lack the ELR motif (ELR-), such as CXCL10, are potent inhibitors of angiogenesis (angiostatic chemokines). ELR+ CXC chemokines bind to CXCR2 and CXCR1, whereas non-ELR- CXC chemokines bind to CXCR3, CXCR4, CXCR5, and CXCR6. CXCL12/CXCR4 axis has been shown to recruit and retain 'recruited bone marrow-derived circulating cells' and support neovascularization of the tumor.

Tumor Growth and Metastasis

Many cancer cells such as melanomas express a number of chemokines, including CXCL8, CXCL1-3, CCL5, and CCL2, which have been implicated in tumor growth and progression. Recent studies have demonstrated organ-specific patterns of melanoma metastasis that correlate with their expression of specific chemokine receptors, including CXCR4, CCR7, CCR9, and CCR10. The chemokine receptors CXCR4 and CXCR7 are also found on breast cancer cells and their ligands are highly expressed at sites associated with breast cancer metastases. Other models in which CXCR4 has been suggested to play a role in metastasis are ovarian, prostate, and lung cancers. In addition, studies have shown that decreased expression of Duffy and D6 decoy receptors in breast cancer inversely correlate with lymph node metastases and increased survival rates. Increased levels of CCL2 expression in breast cancer have been associated with increased levels of tumor-associated macrophages, correlating with invasive phenotype and poor prognosis. It is postulated that chemokine receptors and their ligand pairs play a role in the migration of tumor cells from their primary site via the circulation to the preferential sites of metastases.

Central Nervous System

In addition to their well-established role in the immune system, chemokines and chemokine receptors are involved in the development, homeostasis, synaptic transmission, and neuroinflammation of the central nervous system (CNS). In the CNS, chemokines function as neuromodulators by coordinating the paracrine communication between the neurons, glial and immune cells acting through their chemokine receptors present on these cells.

Chemokine receptors such as CXCR2, CXCR4, CXCR7, CCR1, and CCR5 and chemokine CXCL12 are constitutively expressed in both developing and adult brain. Other chemokines and chemokine receptors are induced depending on the pathological context. Expression of the chemokine CCL21 is induced in neurons around the ischemic brain lesions and spinal cord injury model. Interestingly, CCL21 activates microglia through CXCR3 and not through CCR7. Chemokines and their receptors target the neurons to their destination by modulating migration, differentiation, and proliferation of neuronal and glial cells. The chemokine receptor CXCR4 and its

chemokine ligand CXCL12 play a role in neuronal migration to hippocampus, cerebellum, axonal path finding. CXCL12 antagonizes the axon repellent properties of slit/robo in a retinal neuron model. Perturbation of CXCL12/CXCR4 signaling causes premature cortical plate invasion by GABAergic cortical interneurons leading to positional defects in postnatal cortex. The chemokine receptor CXCR7 coordinates the CXCR4-driven neuronal migration by shaping the extraneuronal CXCL12 gradient. Apart from neuronal homeostasis, chemokines are involved in neuroinflammation as mediators of leukocyte infiltration and glial-cell activation. Their overexpression has been implicated in several neurological disorders, such as multiple sclerosis, trauma, stroke, late stages of Alzheimer's disease, chronic pain, tumor progression, and AIDS-associated dementia. Chemokine receptor antagonists seem to be beneficial in CNS disorders where there is an overexpression of the chemokines or overactivation of the chemokine receptors. In addition, several chemokines and their receptors are implicated in neuroprotection. The employment of chemokines CXCL1, CXCL8, CXCL12, and CX3CL1 are implicated in neuroprotection in different models of neurotoxicity such as gp120, N-methyl-D-aspartate-induced excitotoxicity, and early stages of Alzheimer's disease.

See also: ; **Signaling:** G-Protein-Coupled Receptor Kinases and Arrestins; Phosphatidylinositol Bisphosphate and Trisphosphate; Phospholipase C; Protein Kinase C Family; Ras Family.

Further Reading

- Bajetto A, Bonavia R, Barbero S, and Schettini G (2002) Characterization of chemokines and their receptors in the central nervous system: Physiopathological implications. *Journal of Neurochemistry* 82: 1311–1329.
- Belperio JA, Keane MP, Arenberg DA, et al. (2000) CXC chemokines in angiogenesis. *Journal of Leukocyte Biology* 68: 1–8.
- Lazennec G and Richmond A (2010) Chemokines and chemokine receptors: New insights into cancer-related inflammation. *Trends in Molecular Medicine* 16: 133–144.
- Murphy PM, Baggiolini M, Charo IF, et al. (2000) International Union of Pharmacology, XXII. Nomenclature for chemokine receptors. *Pharmacological Reviews* 52: 145–176.
- Neel N, Schutyser E, Sai J, Fan G, and Richmond A (2005) Chemokine receptor internalization and intracellular trafficking. *Cytokine and Growth Factor Reviews* 16: 637–658.
- Raman D, Baugher P, Thu Y, and Richmond A (2007) Role of chemokines in tumor growth. *Cancer Letters* 256: 137–165.
- Richmond A (2002) NF-kappa B, chemokine gene transcription and tumour growth. *Nature Reviews Immunology* 2: 664–674.
- Thelen M and Stein J (2008) How chemokines invite leukocytes to dance. *Nature Immunology* 9: 953–959.

Chemolithotrophy

A B Hooper, University of Minnesota, St. Paul, MN, USA

A A DiSpirito, Iowa State University, Ames, IA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Alan B. Hooper, volume 1, pp. 419–424, © 2004, Elsevier Inc.

Glossary

Aerobic respiration Movement of electrons from a growth-supporting substrate, through an electron transport chain to dioxygen as the terminal electron acceptor.

Anaerobic respiration Respiration utilizing a terminal electron acceptor other than dioxygen.

Autotrophy A mode of bacterial growth in which carbon dioxide is the source of all or most of the carbon for biosynthesis. These bacteria are autotrophic. Autotrophy contrasts with heterotrophy, a mode of growth where reduced carbon compounds can provide all carbon for biosynthesis.

Chemolithotrophy A mode of bacterial growth in which the reductive substrate for energy generation and for biosynthesis is an inorganic or a single-carbon compound. These bacteria are chemolithotrophic. Chemolithotrophy contrasts with organotrophy or phototrophy in which the energy generation occurs by the oxidation of complex organic molecules or absorption of light, respectively.

Energy substrate A compound that supplies the energy necessary to: (1) drive the phosphorylation of adenosine

diphosphate (ADP) to adenosine triphosphate (ATP) and (2) generate a membrane potential.

Growth-supporting reductant A compound whose oxidation can be coupled biologically to the conservation of energy in forms of general use to cellular activities: as an electrochemical gradient, as a phosphate anhydride bond in ATP, or as the low-potential reductant, reduced nicotinamide adenine dinucleotide.

Methylotrophy A mode of bacterial growth in which the reductive substrate for energy generation and for biosynthesis as well as the carbon source is a single-carbon compound.

Oxic environment An environment containing dioxygen (O_2). The word aerobic is also used. Oxic contrasts with anoxic (or anaerobic) which applies to environments in which oxygen is absent.

Redox Oxidation–reduction reaction.

Terminal electron acceptor (TEA) A compound whose reduction can be coupled to the stoichiometric oxidation of the growth-supporting reductant and conservation of energy in cells.

The chemolithotrophic bacteria and archaea are defined by their ability to oxidize inorganic atoms or molecules as a growth-supporting reductant (Table 1). The enzymes catalyzing these oxidations are unique in nature. A metabolically related group, the methylotrophic microorganisms, utilizes one-carbon compounds (C_1) as a growth-supporting reductant and sole carbon source. The enzymes catalyzing the C_1 oxidations are similar to those of the chemolithotrophs that are described here. The various overall growth-supporting reactions are tabulated here with the reductive and oxidative half-reactions of intermediate enzymatic steps. These half-reactions are steps in the cycling of carbon, hydrogen, nitrogen, sulfur, and metals in nature and have played significant roles in the geochemical history of earth.

Oxidation of the growth-supporting reductant initiates a chain of oxidation–reduction electron-transfer reactions ending with reduction of a terminal electron acceptor (TEA) such as dioxygen. This results in the generation of a membrane potential that can provide the driving force for synthesis of high-energy compounds such as adenosine triphosphate (ATP). Alternatively, the electrons from the growth-supporting substrate reduce compounds such as nicotinamide adenine dinucleotide ($NADH + H^+$), which are employed as universal electron donors in reductive biosynthetic reactions.

Universal Bioenergetic Reactions; Oxidation of the Growth-Supporting Substrate Drives Synthesis of ATP

All organisms require a source of energy to drive biosynthesis, mechanical work, and translocation of molecules across a membrane against a gradient. In cells, potential energy is commonly stored in three basic forms: (1) transmembrane electrochemical potential, (2) the potential-free energy of hydrolysis of a phosphate anhydride bond of ATP, and (3) the low redox potential of the reaction $NAD^+ + 2e^- + 2H \leftrightarrow NADH + H^+$.

The hydrolysis of the terminal phosphoric anhydride bond of ATP can drive an otherwise thermodynamically unfavorable reaction when an enzyme functionally couples the two reactions. The re-phosphorylation of the resulting adenosine diphosphate (ADP) is catalyzed by ATP synthase, an enzyme that is able to use the potential energy of a steady-state transmembrane electrochemical gradient; as protons are relayed across the membrane through the ATP synthase, the enzyme couples ADP and phosphate (P_i) to ATP. In turn, maintenance of the proton electrochemical gradient requires the continuous turnover of the low potential half-reaction (the oxidation of the growth-supporting reducing substrate). The continuation of the latter oxidative reaction is dependent on a high-potential

Table 1 Reductive reactions employed by bacteria

Half-reaction	Sample enzyme/process	Sample species
Aerobic conditions		
a. $O_2 + 4e^- + 4H^+ \rightarrow 2H_2O$	Cytochrome <i>c</i> oxidase	<i>Azotobacter vinlandii</i>
Anaerobic conditions		
b. $NO_3^- + 2e^- + 2H^+ \rightarrow NO_2^- + H_2O$	Periplasmic nitrate reductase	<i>Paracoccus denitrificans</i>
c. $NO_2^- + e^- + 2H^+ \rightarrow NO + H_2O$	Cytochrome <i>cd</i> nitrite reductase	<i>Paracoccus denitrificans</i>
d. $NO_2^- + 6e^- + 8H^+ \rightarrow NH_4^+ + 2H_2O$	Pentaheme nitrite reductase	<i>Campylobacter jejuni</i>
e. $2NO + 2e^- + 2H^+ \rightarrow N_2O + H_2O$	Cytochrome <i>c</i> -dependent nitric oxide reductase	<i>Paracoccus denitrificans</i>
f. $N_2O + 2e^- + 2H^+ \rightarrow N_2 + H_2O$	Nitrous oxide reductase	<i>Paracoccus denitrificans</i>
g. $2H^+ + 2e^- \rightarrow H_2$	Hydrogenase	<i>Rhodospirillum rubrum</i>
h. $C_4H_4O_4 + 2e^- + 2H^+ \rightarrow C_4H_6O_4$	Fumarate reductase	<i>Escherichia coli</i>
i. $CO_2 + 8e^- + 8H^+ \rightarrow CH_4 + 2H_2O$	Methanogenesis	<i>Methanosarcina barkeri</i>
j. $2CO_2 + 8e^- + 8H^+ \rightarrow CH_3COOH + 2H_2O$	Acetogenesis	<i>Acetobacterium woodii</i>
k. $SO_4^{2-} \rightarrow SO_3^{2-} \rightarrow S_4O_6^{2-} \rightarrow S_2O_3^{2-} \rightarrow S^0 \rightarrow S^2$	Sulfate reduction	<i>Desulfovibrio desulfuricans</i>
l. $Fe^{3+} + e^- \rightarrow Fe^{2+}$	Iron reductase	<i>Shewanella putrefaciens</i>
m. $Hg^{2+} + 2e^- \rightarrow Hg^0$	Mercury reductase	<i>Shewanella oneidensis</i>
n. $Mn^{4+} + e^- \rightarrow Mn^{2+}$	Manganese reductase	<i>Shewanella putrefaciens</i>
o. $U^{6+} + 2e^- \rightarrow U^{4+}$	Uranium reduction	<i>Shewanella putrefaciens</i>
p. $Cr(VI) + 2e^- \rightarrow Cr(IV)$	Chromium/uranium oxidoreductase	<i>Desulfovibrio desulfuricans</i>
q. $Cs(VI) + 3e^- \rightarrow Cs(III)$	Cesium reductase	<i>Shewanella oneidensis</i>
r. $Se(VI) + 4e^- \rightarrow Se(0)$	Selenium reductase	<i>Shewanella oneidensis</i>
s. $Tc(VI) + e^- \rightarrow Tc(V)$	Technetium reductase	<i>Shewanella oneidensis</i>
t. $As(V) + 2e^- \rightarrow As(III)$	Arsenate reductase	<i>Desulfovibrio</i>
u. $(CH_3)_3NO + 2e^- + 2H^+ \rightarrow (CH_3)_3N + H_2O$	Trimethylamine reductase	<i>Shewanella putrefaciens</i>
v. $(CH_3)_2SO + 2e^- + 2H^+ \rightarrow (CH_3)_2S + H_2O$	Dimethylsulfoxide reductase	<i>Shewanella putrefaciens</i>

The substrates of these reactions are terminal electron acceptors.

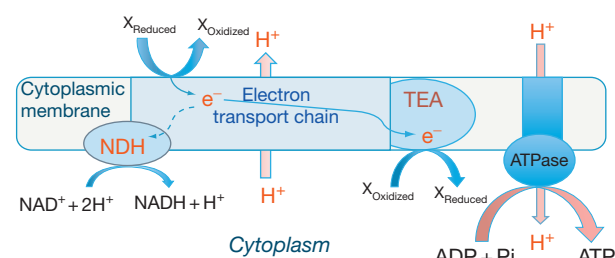


Figure 1 General scheme for oxidation of growth supporting substrate X, listed in **Tables 3–6**, generation of a proton gradient, and phosphorylation of ADP + Pi to ATP in chemolithotrophic and methylotrophic microorganisms. Abbreviations: ATPase, ATP synthase; NDH, NADH dehydrogenase; and TEA, terminal electron acceptor.

half-reaction by which a growth-supporting oxidant substrate, TEA, is reduced by electrons originating from the growth-supporting reductant (**Figure 1**). A series of metallo-enzymes, electron transfer proteins, and quinones are oriented topologically in the cytoplasmic membrane so as to force the net vectorial transfer of protons from the cytosol to the external or extracytoplasmic side of the membrane, resulting in an electrochemical gradient.

To summarize, the potential energy of oxidation of growth substrate by the growth TEA is recovered by coupling these redox reactions to the vectorial transfer of protons from the cytosol to the extracytoplasmic side of the cell membrane and the energy stored in the form of a membrane electrochemical potential. The energy stored in the electrochemical gradient

can then be used to drive enzyme-catalyzed reactions such as the phosphorylation of ADP to ATP.

The Variety of TEAs Used by Bacteria

Many bacteria are able to use one or more of the half-reactions shown in **Table 1** as an electron sink supporting the continuity of the energy-generating reactions; these substrate molecules are TEAs. Because of its high redox potential and resulting high potential difference with the growth-supporting half-reaction, half-reaction 2a is favored by bacteria if oxygen is available. The other TEAs are used when oxygen is absent (i.e., in anoxic or anaerobic rather than oxic or aerobic environments). Note that the reduction of protons in half-reaction g requires an especially low-potential reductant. Line i is a multi-step process carried out by the methanogenic bacteria. In line k, each step involves two electrons. The sulfur-reducing bacteria vary in the steps that they are able to use.

Reductant for Biosynthesis

Biosynthetic reactions often require an electron- or hydride-donor reaction with a low enough potential to drive reductive reactions. The electron donor molecule is often the reduced form of NADH. Hence, a second role of the growth-supporting reductant is the reduction of NAD^+ ($NAD^+ + 2e^- + H^+ \rightarrow NADH$) (see **Figures 1** and **2**). In the chemoautotrophs, $NADPH + H^+$ is used as reductant for biosynthetic reactions.

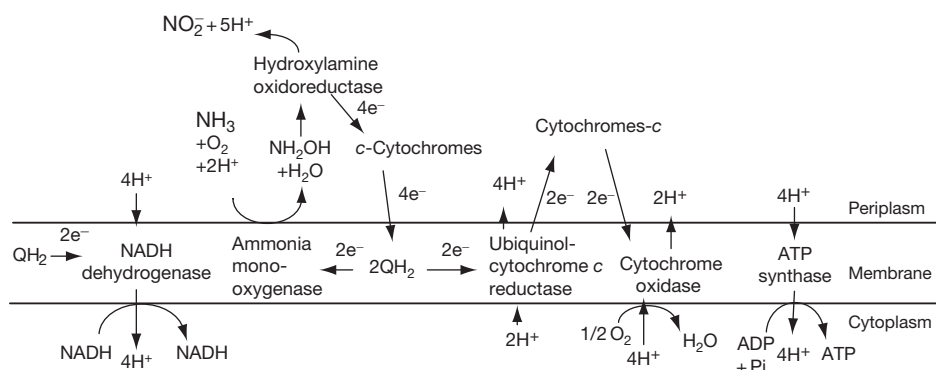


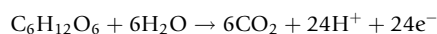
Figure 2 Catalysis of the oxidation of ammonia to nitrite coupled to generation of ATP and NADH in the chemolithotrophic bacterium, *Nitrosomonas europaea*. The four enzymes in the center of the figure create a proton gradient, which is employed by the enzymes on the left and right to drive production of NADH and ATP, respectively. Electrons from hydroxylamine oxidoreductase are relayed by heme-containing electron-transferring proteins, cytochromes in which Fe cycles between Fe^{2+} and Fe^{3+} , to reduce a quinone to a quinol (QH_2). The quinol is reoxidized either by ammonia monooxygenase or by an electron-transfer chain including a proton-pumping ubiquinol–cytochrome *c* reductase, a cytochrome *c*, and a proton-pumping cytochrome oxidase. The proton-pumping NADH-dehydrogenase functions in reverse to form NADH as shown in the figure. ATP synthase catalyzes the proton gradient-driven phosphorylation of ADP. The periplasm is the region outside the cell membrane; the cytoplasm is inside the cell membrane. The membrane itself is impermeable to protons.

Table 2 Classification of microbes based on reductant for biosynthesis and energy-driving ATP-forming reactions and by carbon source

Physiological type	Common name	Carbon source	Electron source	Energy source
Term				
Photo-				Light
Chemo-				Chemical
Autotroph		CO_2		
Organotroph		Organic	Organic	
Lithotroph			Inorganic	
Types				
Photolithoautotroph	Photosynthetic	CO_2	Inorganic	Light
Photoorganotroph	Photoheterotroph	Organic	Organic	Light
Chemolithoautotroph	Chemolithotroph/chemoautotroph	CO_2	Inorganic	Inorganic
Chemoorganotroph	Heterotroph	Organic	Organic	Organic

Categories of Growth of Bacteria and Archaea Based on the Nature of the Growth-Supporting Reductant

Organisms for which the growth-supporting reductant is a reduced multi-carbon molecule such as glucose are called chemoorganotrophs (Table 2). The sum of many reactions, including dehydrogenations, in the pathway of aerobic glucose oxidation is

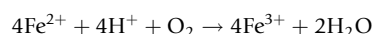
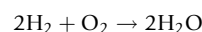
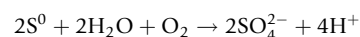
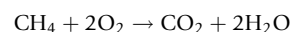
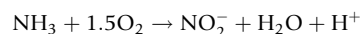


Organisms defined as phototrophs are able to photo-chemically activate an internal chlorophyll to a lower potential reductant. Re-oxidation of the latter drives the creation of the proton electrochemical gradient and then synthesis of ATP. Electrons can either be recycled back to the center of

photochemical reduction resulting in a continuous steady-state gradient and ATP production or can be used to reduce NAD^+ and end up in fixed carbon molecules. Therefore, phototrophs require an external electron source, such as H_2O , to support biosynthesis. Phototrophic bacteria can use either organic or inorganic (e.g., H_2 or H_2S) reductants and are called photo-organotrophs or photolithotrophs, respectively.

Chemolithotrophy

Microorganisms that are specialized to use inorganic compounds as their growth-supporting reductant and energy source are called chemolithotrophs. Bacteria that are able to grow only in this manner are obligate chemolithotrophs; facultative chemolithotrophs have the enzymes necessary for chemoorganotrophic, photoorganotrophic, photolithotrophic, or chemolithotrophic growth. The reducing half-reactions of the major chemolithotrophic growth-supporting substrates are listed in Tables 3–7. In many chemolithotrophic bacteria, the oxidation of the growth-supporting reductant is coupled to reduction of oxygen (Table 1(a)). The overall reactions supporting ATP synthesis and NAD^+ -reduction in aerobic bacteria, which oxidize ammonia, methane, sulfur, hydrogen, or ferric ion with dioxygen as TEA, are as follows:



Bacteria which use a growth-substrate half-reaction with a sufficiently low redox potential can also grow anaerobically

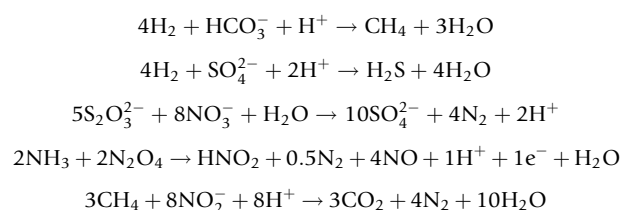
Table 3 Oxidation of the growth-supporting reductant in the nitrifying bacteria

Half-reaction	Sample enzyme/process	Sample species
Aerobic ammonia oxidizing bacteria		
<i>Overall growth-supporting half reaction</i>		
a. $\text{NH}_3 + \text{O}_2 \rightarrow \text{HNO}_2 + \text{H}_2\text{O} + 2\text{H}^+ + 2\text{e}^-$		<i>Nitrosomonas europaea</i>
<i>Enzymatic steps</i>		
i. $\text{NH}_3 + \text{O}_2 + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{NH}_2\text{OH}$	Ammonia monooxygenase	
ii. $\text{NH}_2\text{OH} + \text{H}_2\text{O} \rightarrow \text{HNO}_2 + 4\text{e}^- + 4\text{H}^+$	Hydroxylamine oxidoreductase	
Anaerobic ammonia oxidizing archaea		
<i>Overall growth-supporting half-reaction</i>		
b. $\text{HNO}_2 + \text{NH}_4^+ \rightarrow \text{N}_2 + 2\text{H}_2\text{O}$		<i>Brocadia anammoxidans</i>
<i>Enzymatic steps</i>		
i. $\text{HNO}_2 + \text{e}^- + \text{H}^+ \rightarrow \text{NO} + \text{H}_2\text{O}$	Nitrite reductase	
ii. $\text{NH}_3 + \text{NO} + 3\text{e}^- + 3\text{H}^+ \rightarrow \text{N}_2\text{H}_4 + \text{H}_2\text{O}$	Hydroxylamine oxidoreductase	
iii. $\text{N}_2\text{H}_4 \rightarrow \text{N}_2 + 4\text{e}^- + 4\text{H}^+$	Hydrazine oxidoreductase	
Nitrite oxidizing bacteria		
c. $\text{HNO}_2 + \text{H}_2\text{O} \rightarrow \text{HNO}_3 + 2\text{H}^+ + 2\text{e}^-$	Nitrite oxidoreductase	<i>Nitrobacter winogradskyi</i>

Table 4 Oxidation of the growth-supporting reductant in the methanotrophic bacteria

Half-reaction	Sample enzyme/process	Sample species
Aerobic methane-oxidizing bacteria (methanotrophic bacteria)		
<i>Overall-growth supporting half-reaction</i>		
a. $\text{CH}_4 + \text{O}_2 \rightarrow \text{CO}_2 + 4\text{H}^+ + 4\text{e}^-$		<i>Methylococcus capsulatus</i> Bath
<i>Enzyme steps</i>		
i. $\text{CH}_4 + \text{O}_2 + 2\text{H}^+ + 2\text{e}^- \rightarrow \text{CH}_3\text{OH} + \text{H}_2\text{O}$	Methane monooxygenase	
ii. $\text{CH}_3\text{OH} \rightarrow \text{CHOH} + 2\text{H}^+ + 2\text{e}^-$	Methanol dehydrogenase	
iii. $\text{CHOH} + \text{H}_2\text{O} \rightarrow \text{CHOOH} + 2\text{H}^+ + 2\text{e}^-$	Formaldehyde dehydrogenase	
iv. $\text{CHOOH} + \text{H}_2\text{O} \rightarrow \text{H}_2\text{CO}_3 + 2\text{H}^+ + 2\text{e}^-$	Formate dehydrogenase	
v. $\text{H}_2\text{CO}_3 \rightarrow \text{CO}_2 + \text{H}_2\text{O}$	Carbonic anhydrase	
Anaerobic methane-oxidizing bacteria		
<i>Overall growth-supporting half-reaction</i>		
b. $3\text{CH}_4 + 8\text{NO}_2^- + 8\text{H}^+ \rightarrow 3\text{CO}_2 + 4\text{N}_2 + 10\text{H}_2\text{O}$		<i>Methyloirabillia oxyfera</i>
<i>Enzyme steps</i>		
i. $2\text{NO}_3^- + 2\text{e}^- \rightarrow \text{NO}_2^-$	Periplasmic nitrate reductase	
ii. $2\text{NO}_2^- + 2\text{e}^- + \text{H}_2\text{O} \rightarrow 2\text{NO}$	Nitrite reductase	
iii. $2\text{NO} \rightarrow \text{N}_2 + \text{O}_2$	Hypothetical	(O_2 used in rx. iv)
iv. $\text{CH}_4 + \text{O}_2 + 2\text{e}^- \rightarrow \text{CH}_3\text{OH} + \text{H}_2\text{O}$	Methane monooxygenase	
v. $\text{CH}_3\text{OH} \rightarrow \text{HCOH} + 2\text{e}^-$	Methanol dehydrogenase	
vi. $\text{HCOH} \rightarrow \text{HCOOH} + 2\text{e}^-$	Tetrahydrofolate-dependent C_1 transfer reactions	
vii. $\text{HCOOH} \rightarrow \text{CO}_2 + 2\text{e}^-$	Formate dehydrogenase	

using TEAs other than oxygen as shown in the following examples:



Examples of Half-Reactions of the Growth-Supporting Reductant in the Chemolithotrophic Bacteria

As illustrated below, the origin of the O atom in the oxide product in some cases is O_2 and in some cases is H_2O .

Nitrifying Bacteria (Oxidation of Ammonia to Nitrite and Then Nitrate)

The half-reactions of the nitrifying bacteria make up the oxidative arm of the nitrogen cycle (Table 3). The recently discovered chemolithotrophic anaerobic ammonia oxidizing (ANAMOX) bacteria generate energy in a series of reactions involving ammonia and nitrite with hydrazine as an intermediate and dinitrogen as the end product.

Methane-Oxidizing Bacteria

Both aerobic and anaerobic oxidation of methane by the methanotrophic bacteria (Table 4(a, b)) involves an electron-consuming methane monooxygenase (Table 4(a)(i)) and successive methanol-, formaldehyde-, and formate-dehydrogenases

(Table 4(a)(ii)–(iv)). The conversion of carbonic acid to CO₂ can be spontaneous or enzyme-catalyzed (Table 4(a)(v)). Methylotrophs are bacteria which grow on C₁-compounds, including methane, methanol, formaldehyde, formic acid, methyl amines, dichloromethane, and methane sulfonic acid.

Hydrogen and Carbon Monoxide-Oxidizing Bacteria

Hydrogenase of hydrogen-oxidizing bacteria oxidizes hydrogen to protons and electrons (Table 5(a)). In the carboxydo-bacteria, carbon monoxide dehydrogenase catalyzes the net extraction of two electrons and two protons to produce carbon dioxide from carbon monoxide and water (Table 5(b)).

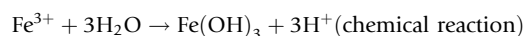
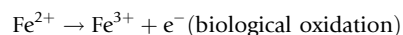
Sulfur-Oxidizing Bacteria

Depending on the specific bacterium, oxidation of a variety of reduced sulfur compounds is used to support growth. Some examples of the half-reactions are shown in Table 6(a–d). Some bacteria utilize the adenosine phosphosulfate (APS)

pathway for synthesis of ATP coupled to the oxidation of sulfite and represent a rare example of substrate level phosphorylation couple in chemolithotrophic bacteria (Table 6(c)).

Metal-Oxidizing Bacteria

The half-reactions shown in Table 7(a–g) are carried out by a variety of bacteria, chemolithotrophically or photolithotrophically, under both aerobic and anaerobic conditions. For example, a metallo-enzyme catalyzes the oxidation of ferrous iron to ferric ion, which hydrolyzes to form an orange precipitate of ferric hydroxides:



Mechanisms of Energy Transduction and ATP Synthesis Somewhat Specific to Chemolithotrophs

Generation of the Electrochemical Gradient and ATP Synthesis

Some generalizations regarding location of the enzymes catalyzing the oxidation of substrate, generation of a proton gradient, and synthesis of ATP are illustrated in the oxidation of ammonia to nitrite by the nitrifying bacterium *Nitrosomonas*, an obligate chemolithoautotroph (Figure 2). Although ammonia is the substrate required for growth, it is the oxidation of hydroxylamine that generates low potential electrons for energy conversion and reductive biosynthetic reactions. Four electrons are transferred from hydroxylamine oxidoreductase (Table 3(a) (ii)) to an electron-transfer chain made up of

Table 5 Oxidation of the growth-supporting reductant in the hydrogen- and carbon monoxide-oxidizing bacteria

Half-reaction	Sample enzyme/process	Sample species
Hydrogen-oxidizing bacteria		
a. $\text{H}_2 \rightarrow 2\text{H}^+ + 2\text{e}^-$	Hydrogenase	<i>Alcaligenes eutrophus</i>
b. $\text{CO} + \text{H}_2\text{O} \rightarrow \text{CO}_2 + \text{H}_2$	Carbon monoxide dehydrogenase	<i>Carboxydotherrmus hydrogenoformans</i>

Table 6 Oxidation of the growth-supporting reductant in the sulfur-oxidizing bacteria

Half-reaction	Sample enzyme/process	Sample species
a. $\text{H}_2\text{S} + 4\text{H}_2\text{O} \rightarrow \text{SO}_4^{2-} + 10\text{H}^+ + 8\text{e}^-$	Sulfide oxidase	<i>Thiobacillus thiooxidans</i>
b. $\text{S}^0 + 4\text{H}_2\text{O} \rightarrow \text{SO}_4^{2-} + 8\text{H}^+ + 6\text{e}^-$	Sulfur oxidase	<i>Thiobacillus thiooxidans</i>
c.1. $\text{S}_2\text{O}_3^{2-} + 5\text{H}_2\text{O} \rightarrow 2\text{SO}_4^{2-} + 10\text{H}^+ + 8\text{e}^-$	Periplasmic thiosulfate-oxidizing multienzyme complex	<i>Thiobacillus versutus</i>
c.2. <i>Tetrathionate path for the oxidation of thiosulfate</i>		
i. $2\text{S}-\text{SO}_3^{2-} \rightarrow \text{O}_3\text{S}-\text{S}-\text{S}-\text{SO}_3^- + 2\text{e}^-$	Thiosulfate dehydrogenase	<i>Paracoccus pantotrophus</i>
ii. $\text{O}_3\text{S}-\text{S}-\text{S}-\text{SO}_3^- + \text{H}_2\text{O} \rightarrow \text{O}_3\text{S}-\text{S}-\text{S}^- + \text{SO}_4^{2-} + 2\text{H}^+$	Tetrathionate hydrolase	
iii. $\text{O}_3\text{S}-\text{S}-\text{S}^- + \text{H}_2\text{O} \rightarrow \text{S}-\text{SO}_3^- + \text{SO}_4^{2-} + 2\text{H}^+$	Trithionate hydrolase	
<i>Protein cysteine S-thiosulfonate-intermediate pathway for the oxidation of thiosulfate</i>		
$\text{prot-SH} + \text{S}-\text{SO}_3^- \rightarrow \text{prot-S-S-SO}_3^- + 2\text{e}^- + \text{H}^+$	Thiosulfate dehydrogenase	<i>Thiobacillus denitrificans</i>
$\text{prot-S-S-SO}_3^- + \text{H}_2\text{O} \rightarrow \text{prot-S-S}^- + \text{SO}_4^{2-} + 2\text{H}^+$	Protein cysteine S-thiosulfonate hydrolase	
$\text{prot-S-S}^- + 3\text{H}_2\text{O} \rightarrow \text{prot-S-SO}_3^- + 6\text{e}^- + 6\text{H}^+$	Protein cysteine S-thiol dehydrogenase	
$\text{prot-S-SO}_3^- \rightarrow \text{prot-SH} + \text{SO}_4^{2-} + \text{H}^+$	Protein cysteine S-thiol dehydrogenase	
Adenosine phosphosulfate pathway for thiosulfate oxidation		
c.3. <i>Overall-growth-supporting half-reaction</i>		
$\text{S}_2\text{O}_3^{2-} + 5\text{H}_2\text{O} + \text{AMP} \rightarrow 2\text{SO}_4^{2-} + 8\text{H}^+ + 6\text{e}^- + \text{ADP}$		
<i>Enzyme steps</i>		
i. $\text{S}_2\text{O}_3^{2-} \rightarrow \text{S}^0 + \text{SO}_3^-$	Rhodanese	
ii. $\text{SO}_3^- + \text{AMP} \rightarrow \text{adenosine phosphosulfate (APS)}$	Adenosine phosphosulfate reductase	
iii. $\text{APS} + \text{Pi} \rightarrow \text{ADP} + \text{SO}_4^{2-}$	ADP sulfurylase	
iv. $\text{S}^0 + 4\text{H}_2\text{O} \rightarrow \text{SO}_4^{2-} + 8\text{H}^+ + 6\text{e}^-$	Sulfur oxidase	
d. $\text{SO}_3^{2-} + \text{H}_2\text{O} \rightarrow \text{SO}_4^{2-} + 2\text{H}^+ + 2\text{e}^-$	Sulfite cytochrome <i>c</i> oxidoreductase	<i>Thiobacillus thiooxidans</i>

Note: Sulfur-oxidizing bacteria will oxidize a number of different compounds; some of the more common substrates are given in this table.

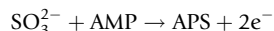
Table 7 Oxidation of the growth-supporting reductant in the metal-oxidizing bacteria

Half-reaction	Sample enzyme/process	Sample species
a. $\text{As}^{3+} \rightarrow \text{As}^{5+} + 2\text{e}^-$		<i>Agrobacterium tumefaciens</i>
b. $\text{Cu}^+ \rightarrow \text{Cu}^{2+} + \text{e}^-$	Iron oxidase	<i>Thiobacillus ferrooxidans</i>
c. $\text{Fe}^{2+} \rightarrow \text{Fe}^{3+} + \text{e}^-$	Iron oxidase	<i>Thiobacillus ferrooxidans</i>
d. $\text{Mn}^{2+} \rightarrow \text{Mn}^{4+} + 2\text{e}^-$	Manganese(II) oxidase	<i>Pseudomonas putida</i> strain GB-1
e. $\text{Sb}^{3+} \rightarrow \text{Sb}^{5+} + 2\text{e}^-$	Unidentified	<i>Agrobacterium tumefaciens</i>
f. $\text{Se}^0 \rightarrow \text{Se}^{6+} + 6\text{e}^-$	Unidentified	<i>Leptothrix</i> sp. Strain MNB-1
g. $\text{U}^{4+} \rightarrow \text{U}^{6+} + 2\text{e}^-$	Iron oxidase	<i>Thiobacillus ferrooxidans</i>

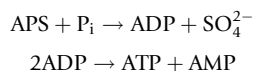
two heme-containing proteins (cytochromes) to two quinone molecules in the membrane. Two electrons from the reoxidation of one quinol are utilized by ammonia monooxygenase to regenerate a molecule of hydroxylamine, that is, to re-prime the pump. The two electrons (per hydroxylamine oxidized) from the second quinol may reduce dioxygen in a pathway which involves the two membrane enzymes, ubiquinol cytochrome *c* reductase and cytochrome *c* oxidase, each of which couples the reaction to the transport of protons against a gradient to the outside (the periplasm of the bacterial cell). In this way, a steady-state electrochemical potential (outside positive) is maintained across the membrane and drives ATP synthesis.

Substrate-Level Phosphorylation of ADP

In some chemolithotrophic sulfite-oxidizing bacteria, the net phosphorylation of ADP occurs by a substrate-level process, which is independent of the membrane electrochemical gradient (Table 5(c) (iii)). The key reaction involves the simultaneous dehydrogenation of sulfite and reaction with adenosine monophosphate (AMP) to form the anhydride of phosphoric and sulfuric groups in the product, APS:



Phosphate can then displace sulfate to form ADP and the subsequent dismutation of two molecules of ADP provides an ATP:



Periplasmic Release and Cytoplasmic Utilization of Protons in Chemolithotrophic Reactions

In the reducing half-reactions of the major chemolithotrophic growth-supporting reductants (Tables 3–7), the energetically fruitful reactions are, in effect, dehydrogenases where the products are free protons and electrons are transferred to an enzyme redox center which is often a metal such as Fe or Cu.

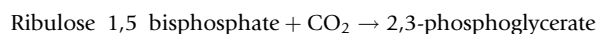
The protons originate from the substrate (in the oxidation of ethanol, formaldehyde, H_2 , N_2H_4 , H_2S), H_2O (in the oxidation of CO, nitrite, S^0 , SO_3^{2-}), or both (in the oxidation of NH_2OH). In all cases where location of the enzyme has been determined, the proton-yielding reactions are found in the periplasm or periplasmic face of the membrane and thus contribute to maintenance of the positive charge on the outer side of the membrane. Correspondingly, almost all TEA reactions (Table 1) that consume protons are located on the cytoplasmic side of the membrane, thus contributing to the maintenance of the negative charge on that side. This is illustrated in the location of hydroxylamine dehydrogenation and oxygen reduction in *Nitrosomonas* (Figure 2).

Production of NADH by Chemolithotrophic Bacteria

With many chemolithotrophs, the potential of the half-reaction that drives metabolism (in the case of *Nitrosomonas*, the dehydrogenation of hydroxylamine) is apparently not low enough to reduce NAD^+ directly. Thus, chemoautotrophs must drive electrons in the reverse direction to reduce NAD^+ to $\text{NADH} + \text{H}^+$ in a process called reverse electron transport. The final reaction in this process is catalyzed by the NADH dehydrogenase. In chemoorganotrophs, the NADH dehydrogenase oxidizes $\text{NADH} + \text{H}^+$ to NAD^+ , reduces quinone, and pumps protons to the outside of the cell. However, in the case of *Nitrosomonas* (Figure 2) and other chemolithotrophs, the reaction runs in the reverse direction to form NADH. This process is not needed in bacteria, such as the hydrogen oxidizers, where the growth-supporting potential of the reductive half reaction ($\text{H}_2 \rightarrow 2\text{H}^+ + 2\text{e}^-$) has a low enough value that NAD^+ is reduced directly.

Carbon Source for Chemolithotrophic Bacteria

The carbon source for most chemolithotrophic bacteria is CO_2 . This mode of growth is called autotrophic, that is, the organisms are chemolithoautotrophs (Table 2). Most chemolithotrophs are obligate autotrophs, meaning that they can only grow having synthesized all or most of their reduced carbon compounds from CO_2 . This contrasts with heterotrophic organisms, which can fulfill all their need for carbon compounds with previously fixed carbons such as sugars. Facultative autotrophs are able to alter their content of biosynthetic enzymes to catalyze either the autotrophic or heterotrophic mode of life. The most common pathway of carbon dioxide fixation in the chemolithoautotrophs is the Calvin cycle, starting with the reaction of ribulose bis-phosphate carboxylase, which catalyzes the following reaction:



Considering the reduction of 3-phosphoglycerate and the regeneration of ribulose 1,5-bis-phosphate in the Calvin cycle, approximately 12 molecules of NADH and 18 of ATP are required per hexose synthesized from CO_2 . Remarkably, the obligate chemolithoautotrophs contain the enzymes for biosynthesis of all cellular constituents. Further, they have genes for enzymes for the uptake, metabolism, and incorporation of carbon from extremely few carbon compounds other than

CO₂. This was illustrated by analysis of the genes present and absent in the genome of *Nitrosomonas*.

Growth Yields of Chemolithoautotrophs

Since the reduction potential of the half-reactions utilized by the chemolithotrophs often has a low value, the energy yield from the overall growth reaction is lower than commonly observed with organotrophs. Given the required expenditure of energy required to drive the synthesis of ATP and the reduction of NAD⁺, which are then used in biosynthetic reactions, the yield of biomass per mole of substrate oxidized is usually much lower in the chemolithoautotrophs than with heterotrophic organotrophs. Profiting from their unique catalytic capabilities, chemolithotrophs grow slowly and live frugally with a highly selective diet of a thin low-calorie soup of inorganic compounds or small carbon molecules.

Ecology

The chemolithotrophs and methylotrophs represent the oxidative segment of the biological cycles of inorganic compounds such as hydrogen, nitrogen, sulfur, metal ions, and carbon. The vast majority of these steps occurs aerobically. These reactions and the bacterial reduction of inorganic TEAs have played significant parts in the geochemical history of carbon, nitrogen, oxygen, sulfur, and metals on earth.

Most of the reduced substrates for chemolithotrophs are generated by the terminal electron accepting reactions of anaerobic organisms (Table 1). For example, hydrogen from reduction of protons (Table 1(g)) by organotrophic bacteria in cattle rumen is oxidized (Table 5(a)) by chemolithotrophic hydrogen-oxidizing bacteria using the conversion of carbon dioxide to methane as the TEA half-reaction (Table 1(j)). Methane released from ruminants finds its way to an oxic (aerobic) environment where it is oxidized by the methanotrophic bacteria (Table 4(a)). The chemolithotrophs and the anaerobes are sometimes found at oxic/anoxic (aerobic/anaerobic) interfaces where the respective oxidative and reductive reactions provide a cycle to the advantage of both types of organism. There are many examples of economically significant effects of chemolithotrophic activity. For example, reduction of sulfuric acid by the sulfur-oxidizing bacteria facilitates corrosion of metals and has been used to extract metals in commercial mining processes.

See also: **Bioenergetics:** Algal Hydrogen Production; Chlorophylls and Carotenoids; Cytochrome *c*; Cytochrome Oxidases, Bacterial; Dynamic Behavior of Photosystem II Light Harvesting; F-ATPases;

Ferredoxin-NADP⁺ Reductase; Hydrogenases, Structure and Function; Membrane Transport, General Concepts; Membrane-Associated Energy Transduction in Bacteria and Archaea; Nicotinamide Nucleotide Transhydrogenase; Periplasmic Electron-Transport Systems in Bacteria; Photosystem I; Photosystem II: Redox and Protein Components; Photosystem II: Water Oxidation, Overview; Proton Pumping in the Respiratory Chain; Purple Bacteria: Electron Acceptors and Donors; Quinones; Renewable Hydrogen from Biomass; Respiration in Phototrophic Microorganisms; Respiratory Chain and Adenosine Triphosphate Synthase; **Metabolism Vitamins and Hormones:** Photosynthesis; Photosynthetic Carbon Dioxide Fixation; **Protein/Enzyme Structure Function and Degradation:** Heme Proteins.

Further Reading

- Chain P, Lamerdin J, Larimer F, et al. (2003) Complete genome sequence of the ammonia oxidizing bacterium and obligate chemolithoautotroph *Nitrosomonas europaea*. *Journal of Bacteriology* 105: 2759–2773.
- Ettwig KF, Butler MK, LePaslier D, et al. (2010) Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* 464: 543–548.
- Friedrich CG, Bardischewsky F, Rother D, Quentmeier A, and Fischer J (2005) Prokaryotic sulfur Oxidation. *Current Opinion in Microbiology* 8: 253–295.
- Gadd GM (2010) Metals, minerals and microbes: Geomicrobiology and bioremediation. *Microbiology* 156: 609–643.
- Hooper AB, Arciero DM, Bergmann D, and Hendrich MP (2003) The oxidation of ammonia as an energy source in bacteria. In: Zannoni D (ed.) *Respiration in Archaea and Bacteria, Vol. 16, Diversity of Prokaryotic Respiratory Systems*, pp. 122–147. The Netherlands: Kluwer Scientific.
- Hooper AB and DiSpirito AA (1985) In bacteria which grow on simple reductants, generation of a proton gradient involves extracytoplasmic oxidation of substrate. *Microbiology Reviews* 49: 140–157.
- Jetten MSM, Niftrik L, Strous M, et al. (2009) Biochemistry and molecular biology of anammox bacteria. *Critical Reviews in Biochemistry and Molecular Biology* 44(2): 65–84.
- King GM and Weber CF (2007) Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nature Reviews Microbiology* 5: 107–118.
- Lengeler JW, Drews G, and Schlegel HG (1999) *Biology of the Prokaryotes* Blackwell. New York, Stuttgart: Thieme.
- Oelgeschläger E and Rother M (2008) Carbon Monoxide-dependent energy metabolism in anaerobic bacteria and archaea. *Archives of Microbiology* 190: 257–269.
- Sauve V, Roversi P, Leath KJ, et al. (2009) Mechanism for the hydrolysis of a sulfur-sulfur bond based on the crystal structure of the thiosulfohydrolase SoxB. *The Journal of Biological Chemistry* 284: 21707–21718.
- Schlegel HG and Bowein B (1987) *Autotrophic Bacteria*. Madison, WI: Science Tech. Berlin: Springer.
- Schmidt I, Hermelink C, van de Pas-Schoonen K, et al. (2002) Anaerobic ammonia oxidation in the presence of nitrogen oxides (NO_x) by two different lithotrophs. *Applied and Environmental Microbiology* 68: 5351–5357.
- Semrau JD, DiSpirito AA, and Yoon S (2010) Methanotrophs and copper. *FEMS Microbiology Reviews* 24: 496–531.
- Stolz JF, Basu P, Santini JM, and Oremland RS (2006) Arsenic and selenium in microbial metabolism. *Annual Review of Microbiology* 60: 107–130.
- White D (2000) *The Physiology and Biochemistry of Prokaryotes*. New York: Oxford University Press.

Chemotactic Peptide/Complement Receptors

E R Prossnitz and L A Sklar, University of New Mexico, Albuquerque, NM, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Chemokine A family of structurally related 70–90-amino-acid-long glycoproteins that exhibit potent chemotactic and stimulatory properties on leukocytes.

Complement system A series of 30 components that include proteolytic pro-enzymes, nonenzymatic components that form functional complexes, regulators, and receptors. The proteolytic pro-enzymes become sequentially activated. C5a and C3a are terminal nonenzymatic cleavage products.

G protein A protein that binds and is regulated by guanine nucleotides. Heterotrimeric members of this family couple to GPCRs. Low-molecular-weight monomeric members of the family are activated downstream of heterotrimeric G proteins.

G-protein-coupled receptor (GPCR) Transmembrane receptor containing seven membrane-spanning helices that activates cells through coupling to heterotrimeric G proteins.

A critical aspect in the normal function of many cell types is their ability to migrate in response to soluble ligands or chemoattractants. Directional migration, where the cell seeks out the highest concentration of chemoattractant, is referred to as chemotaxis. This is of paramount importance in the immune system where a large collection of G-protein-coupled receptors (GPCRs) exists to coordinate the complex interactions of immune cells. These receptors can be divided into two families, the chemotactic or chemoattractant GPCRs that respond to a wide variety of endogenous as well as exogenous ligands and the chemokine GPCRs that respond to the structurally conserved family of endogenous peptide chemokines. The chemotactic GPCRs respond to ligands as varied as oligopeptides, such as bacterially derived N-formylated peptides; proteins, such as complement components; and lipids, such as leukotrienes and platelet-activating factor.

Leukocytes and the Immune System

The immune system is responsible for numerous activities in the body including defense against pathogens (bacterial, viral, and parasitic), removal of dead and cancerous cells as well as the rejection of foreign cells. Leukocytes, or white blood cells, which represent the principal cells of the immune system, are produced as progenitor cells in the bone marrow and mature in various sites throughout the body, including the spleen, thymus, and lymph nodes. Leukocytes can be divided into five broad categories: neutrophils (highly mobile phagocytic cells), basophils (allergy mediators through histamine release), eosinophils (parasitic killing and allergy), lymphocytes (B cells that produce antibodies and T cells that affect cell-mediated immunity), and monocytes/macrophages (tissue-specific phagocytic cells). These cells differentially participate in two types of immune responses: innate immunity, an inherent and rapid response of neutrophils, macrophages, and natural killer lymphocytes to foreign materials; and acquired immunity, a selective trained response involving lymphocytes requiring prior exposure to an agent. An additional noncellular defensive mechanism, the complement system consisting of at least 20–30 distinct serum proteins,

destroys foreign cells through membrane lysis of antibody-targeted cells. Two of the smaller bioactive fragments released by proteolysis from the C3 and C5 components, C3a and C5a, are known as anaphylatoxins and play a major role in inflammation. Immune cells for the most part circulate in the vascular and lymphatic systems, although tissue resident cells such as macrophages are essential to proper immune function. In almost any immune response, however, leukocytes must migrate to the site of damage/infection from the blood. This is a complex multistep process involving inflammatory mediators (such as histamine) that induce vasodilation and increased vascular permeability, activation of adhesion molecules on capillary lining endothelial cells to slow leukocytes, and release of chemoattractants to activate leukocytes resulting in firm adhesion to endothelial cells. The concentration gradients of chemoattractants induce the leukocyte to transverse the endothelium and migrate through the tissue to the peak concentration of chemoattractant at the site of inflammation or infection. Having arrived at this site, neutrophils, for example, phagocytose bacteria, generate superoxide radicals and release granules containing degradative proteases. Most of these responses are initiated and regulated by a family of cell surface receptors that bind inflammatory mediators.

Receptors

The largest family of proteins in the human genome is the GPCR superfamily consisting of at least 600 members. GPCRs possess a highly conserved structure, consisting of seven transmembrane segments with the amino terminus on the cell exterior and the carboxy terminus in the cytoplasm. The chemoattractant family of receptors consists of proteins with 350 amino acids. The transmembrane regions consist primarily of α -helices arranged in a bundle. Despite this conserved structure, there is substantial sequence diversity between members of this receptor family. Binding of small ligands, such as chemotactic peptides, occurs deep within the receptor in the plane of the membrane. Ligand binding and receptor activation result in conformational changes that activate heterotrimeric

G proteins, which further stimulate a wide variety of effector proteins. Following activation, GPCRs undergo phosphorylation, resulting in desensitization and internalization, which in turn can lead to receptor degradation or recycling back to the cell surface.

Chemotactic Peptide Receptors

Though identified as a neutrophil receptor for N-formylated peptides in the 1970s, it was not until 1990 that the N-formyl peptide receptor (FPR) gene was isolated, representing the first chemoattractant or chemokine receptor to be cloned. Two human homologs were subsequently identified and designated FPRL1 and FPRL2. Of these, only the former responds, though poorly, to formyl peptides. Although the FPR is expressed predominantly in neutrophils, it also appears to be expressed in monocytes, dendritic cells, hepatocytes, endothelial cells, and the nervous system. FPRL1 and FPRL2 are expressed in some of the same cell types as well as unique sites. Although the human FPR gene family consists of only three members, the number varies considerably in other species, with six members in the murine gene family for example. Thus, distinct evolutionary pressures seem to dictate the repertoire of formyl peptide receptors in a given organism.

The role of the FPR in antimicrobial defense has recently been directly demonstrated in knockout mice lacking mFPR1, the homolog of the human FPR. Although such mice appear normal under unstressed conditions, they demonstrate a lack of neutrophil responses to N-formyl-methionyl-leucyl-phenylalanine (fMLP or fMLF) challenge as well as an increased susceptibility to infection with *Listeria monocytogenes*. Furthermore, in humans, two variant alleles of the FPR representing single point mutations that disrupt normal function are associated with localized juvenile periodontitis, a result of periodontal bacterial infection. It is perhaps surprising that such defects do not lead to more serious conditions.

Complement Receptors

C5a Receptor

At the same time that formylated peptides were being characterized as chemoattractants, proinflammatory peptides derived from the complement system were similarly being studied. Two receptors have been identified that respond to the cleavage products of the complement components C3 and C5. The C5a receptor is similar in size to the FPR and closely related to it phylogenetically. Interestingly, there exists little interspecies homology in the extracellular regions of the C5a receptor, despite apparent contact sites with C5a and interspecies activation (C5a from one species activating the receptor from another species). Antibodies directed against peptides representing the amino terminus of the receptor and amino-terminal truncations of the receptor prevent C5a binding. As the antibodies do not block activation of the receptor by synthetic C5a carboxy-terminal peptides, the amino terminus of the receptor may represent a secondary nonactivating binding site and not the primary effector site. Mutagenesis studies suggest a critical role for a number of aspartic acid residues in the

amino terminus. Recent evidence is consistent with C5a receptor dimerization, a phenomenon now demonstrated for many GPCRs with unclear functional consequences. Studies of genetically engineered mice lacking the C5a receptor demonstrated that this receptor plays a critical role in clearing pulmonary pseudomonas infections, with mice succumbing to pneumonia despite a marked neutrophil influx.

C3a Receptor

The C3a receptor is unusual in that it possesses a second extracellular loop of nearly 175 amino acids compared to about 30 amino acids for most GPCRs including the C5a receptor and FPR family. Deletion analysis of this loop suggests that it plays a role in C3a binding and receptor activation, although as much as two-thirds of the loop can be deleted without affecting activity. It is thought that the second extracellular loop may replace the function of the amino terminus of the C5a receptor in providing a secondary nonactivating binding site. Recent studies with C3a receptor knockout mice suggest a role for the receptor in allergic airway diseases such as asthma.

Ligands

Chemotactic Peptide Receptors

The first leukocyte chemotactic factors to be defined structurally were the N-formyl peptides. These consist of short oligopeptides consisting of three to six hydrophobic amino acids, with the amino terminal residue being formylated. Such peptides are not synthesized by eukaryotic cells *per se*. There are only two sources of N-formylated peptides in nature: one endogenous to eukaryotic cells, mitochondria, and the other exogenous, bacteria. Each of these sources initiates protein synthesis with an N-formyl methionine tRNA. This suggests an ideal targeting mechanism for leukocytes to sense bacterial infection or cell damage through the release of peptides containing N-formyl-methionyl peptides. The first chemoattractant peptide was isolated from bacterial lysates and was identified in 1976 as fMLP. This agonist is capable of eliciting the full complement of leukocyte functions from chemotaxis to degranulation to superoxide production at concentrations ranging from pM to nM. The FPR is the primary receptor responsible for cellular stimulation in these concentration ranges, whereas FPRL1 responds to fMLF at μ M concentrations. Both the FPR and FPRL1 have recently been shown to be activated by a diverse collection of endogenous as well as exogenous peptides. These include nonformylated peptides from exogenous sources such as *Helicobacter pylori* and HIV-1 envelope proteins as well as endogenous host-derived proteins such as serum amyloid A and peptides from annexin I, amyloid precursor protein, and prion protein among others. Lipoxin A4, a lipid metabolite, has also been reported to activate FPRL1. The biological relevance of these agonists remains to be determined. Finally, a number of nonformylated high-affinity peptide agonists for the FPR and FPRL1 have been isolated from peptide libraries. A number of antagonists have also been identified, including Boc-FLFLF, cyclosporine H (a fungal metabolite), and deoxycholic acid (a bile acid).

Complement Receptors

C5a

The anaphylatoxin C5a, the most potent plasma-derived chemotactic factor, is a 74-amino-acid peptide/protein released from the α -chain of C5 by either the classical or alternative pathway C5 convertases. Structurally, the protein consists of four antiparallel α -helices held together by disulfide bonds and an 11-amino-acid carboxy terminal tail. Unlike the small formyl peptides, C5a appears to interact with its receptor at two sites. The α -helical bundle interacts with the receptor's amino-terminal domain, whereas the tail appears to insert itself into the transmembrane bundle of the receptor, thereby activating the receptor. Synthetic peptides representing the carboxy-terminal ten amino acids are full agonists, though with potencies 1000–10 000 lower than C5a itself. The sequence Gln–Leu–Gly–Arg is a highly conserved effector sequence in C5a from numerous species. Substitution of residues within this region can greatly augment potency. Removal of the final arginine residue in C5a results in greatly decreased potency, ranging from 1000-fold decrease for spasmogenic activity to 10-fold for chemotactic activity.

C3a

Less is known regarding the interactions between C3a and its receptor. C3a is a 77-amino-acid protein released from the α -chain of C3 by either the classical or alternative pathway C3 convertases. C3a and C5a, despite having common genetic ancestry, have markedly different sequences. Only 13-amino-acid positions have been totally conserved between C3a, C5a, and the related C4a molecules from numerous species analyzed. Six of these residues represent immutable cysteine residues that direct folding of the protein. Again, the carboxy-terminal sequence seems to define molecular action as the terminal sequence Leu–Gly–Leu–Ala–Arg is conserved in all known C3a molecules. C3a appears to interact with both portions of the large second extracellular loop of the receptor as well as charged residues at the transmembrane boundaries of this loop. For both the C5a and C3a receptors, insertion of the carboxy terminus of the anaphylatoxin into a binding pocket formed by the transmembrane helices of the receptor appears essential to receptor activation.

G Proteins

Following ligand binding and the commensurate activation of the receptor, GPCRs interact with G proteins initiating a vast array of intracellular signaling (Figure 1). Of the four major G protein subfamilies ($G_{s/r}$, G_i , $G_{q/11}$, and $G_{12/13}$), chemoattractant/complement receptors activate primarily the G_i family of G proteins. This is demonstrated by the inhibition of virtually all chemoattractant receptor-mediated cell activation by pertussis toxin, a bacterial toxin that inactivates G_i (and closely related G_o) α -subunits of G proteins. G proteins consist of three subunits: the guanine nucleotide-binding α -subunit, and the nondissociable membrane-associated $\beta\gamma$ -subunit dimer. Receptor activation catalyzes the exchange of guanosine diphosphate (GDP) for guanosine triphosphate (GTP) on the α -subunit followed by activation of the α - and $\beta\gamma$ -subunits. Hydrolysis of the bound GTP by the intrinsic activity of the G_{α}

protein, accelerated by a family of regulators of G protein signaling (RGS) proteins, inactivates the protein, and allows reassembly of the inactive $\alpha\beta\gamma$ -heterotrimer.

Effector Proteins

Chemoattractant-mediated G protein activation leads to a plethora of downstream signaling events. Although G_i proteins are generally considered to inhibit adenylyl cyclase activity, in the case of most chemoattractant and chemokine receptors, G_i activation leads to $G_{\beta\gamma}$ -mediated activation of phospholipase $C\beta_2$, resulting in the production of inositol trisphosphate (IP_3) and diacylglycerol (DAG). IP_3 generation results in the release of calcium from intracellular endoplasmic reticulum stores, often followed by the influx of calcium from the extracellular environment through ion channels. The calcium and DAG in turn activate protein kinase C (PKC). Receptor-mediated phospholipase A_2 activation results in the generation of arachidonic acid and subsequently prostaglandins and leukotrienes. Phosphatidylinositol 3-kinase (PI3K) also plays a critical role in leukocyte activation and chemotaxis in particular. MAP kinase family members ERK, p38, and Jnk have all been shown to be essential to cellular activation events. The low-molecular-weight monomeric G proteins Rac and Rap, as well as their associated regulatory proteins and downstream effectors, are important mediators of actin assembly/chemotaxis and superoxide production.

Receptor Regulation

The dysregulated activation of leukocytes is a serious contributing factor to numerous chronic inflammatory diseases such as arthritis and reperfusion injury in ischemic tissues following myocardial thrombosis and stroke. Activated chemoattractant/complement receptors are rapidly phosphorylated following ligand binding and receptor activation. This can be mediated either by G-protein-coupled receptor kinases (GRKs) that are specific for the active state of GPCRs or by second messenger kinases such as protein kinase A and PKC that phosphorylate active and inactive receptors. The latter provides a mechanism (termed heterologous desensitization) by which GPCRs such as the FPR, which has no PKC phosphorylation sites, can desensitize the C5a receptor, which is phosphorylated by PKC, leading to decreased G protein coupling (see Figure 1). Since any inflammatory process is likely to lead to the generation of multiple simultaneous chemoattractant signals, these regulatory mechanisms create a hierarchy of chemotactic responsiveness to different ligands. GRK phosphorylation of chemoattractant receptors, which occurs within the carboxy terminus of the receptor at sites distinct from PKC sites, leads to binding of arrestins and receptor internalization. Arrestin binding further uncouples receptors from G proteins by physically preventing access of G proteins to the receptor (termed homologous desensitization). For many GPCRs, arrestin binding also mediates interactions with clathrin and adapter proteins such as AP-2, initiating receptor translocation to clathrin-coated pits and subsequent internalization. Contrary to this paradigm, the FPR internalizes in an arrestin-independent manner. However, arrestin is required for recycling of the FPR back to the cell surface. Arrestins have also been shown to play

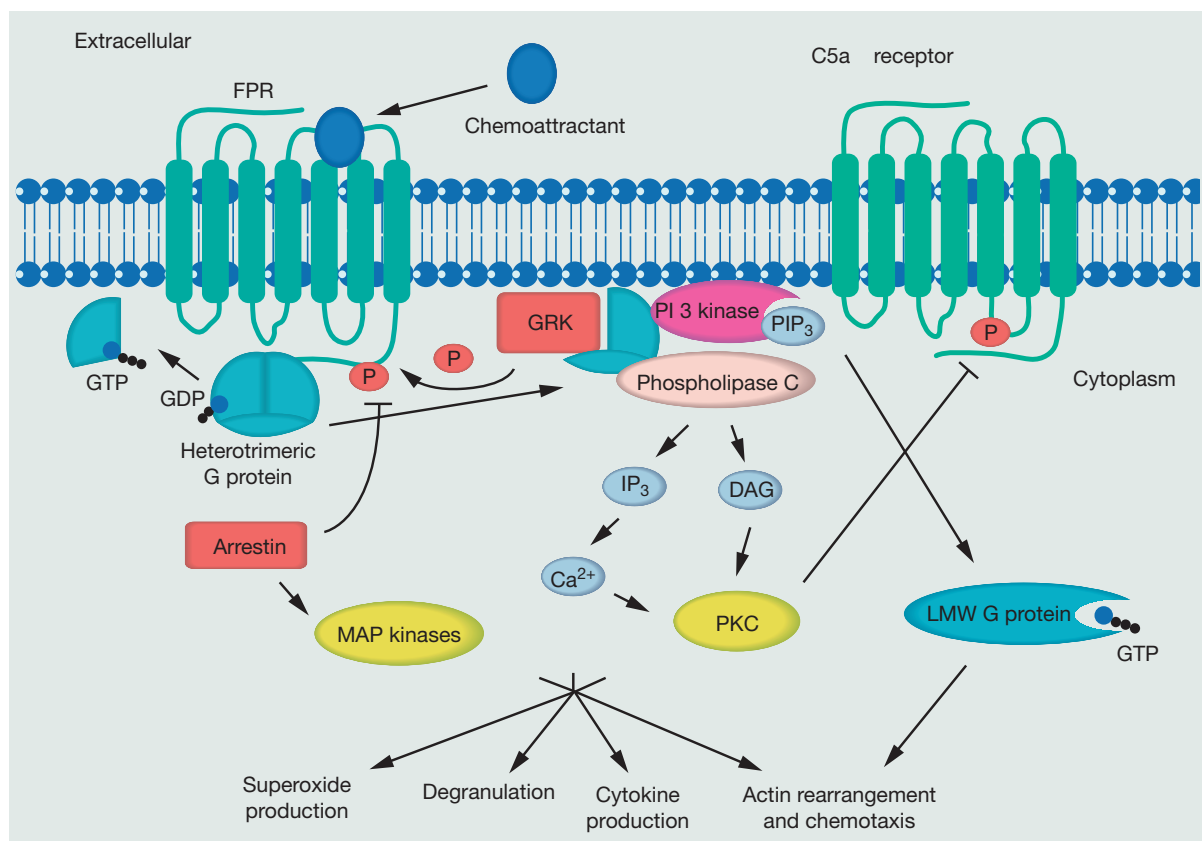


Figure 1 Signaling and regulation of chemoattractant/complement receptors in leukocytes. Binding of a chemoattractant (agonist) results in a conformational change in the FPR, which activates G proteins. Dissociation of GDP from the α -subunit is followed by GTP binding and dissociation of the α -subunit from the $\beta\gamma$ -subunits, both of which can stimulate downstream effectors. The $\beta\gamma$ -subunits provide a docking site for G-protein-coupled receptor kinases (GRKs), which phosphorylate (P) the ligand-bound, activated receptor. $\beta\gamma$ -Subunits also activate phospholipases, which generate diacyl glycerol (DAG) and inositol trisphosphate (IP_3) and phosphatidylinositol-3 kinase (PI 3 kinase), which generates phosphatidylinositol trisphosphate (PIP_3). PIP_3 is an activator of numerous pathways, including those culminating in low-molecular-weight (LMW) monomeric G proteins of the Ras superfamily that regulate actin assembly and ultimately chemotaxis. IP_3 production results in the release of Ca^{2+} from intracellular stores and, together with DAG, activates protein kinase C (PKC). PKC is capable of phosphorylating many proteins, including complement receptors, such as the C5a receptor, resulting in their heterologous desensitization. Desensitization of the FPR occurs through the binding of arrestin to GRK-phosphorylated receptors. In addition to preventing further signaling through heterotrimeric G proteins, arrestins serve as scaffolds for multiple MAP kinases, resulting in their localized activation. Together, these and many additional pathways produce cellular responses, including superoxide generation, degranulation, chemotactic cell movement, and transcriptional activation resulting in cytokine production.

an important role in the scaffolding of numerous intracellular signaling cascades, including Src and MAP kinases. This mechanism is believed to provide both spatial and temporal regulation of the downstream effectors.

Drug Discovery and Receptor Technology

GPCRs represent 50% of the targets of drugs on the market today. In particular, the family of chemokine/chemoattractant receptors is highly sought after target as the receptors are involved in immune disorders such as chronic inflammation and asthma, cancer, and many more disease states. The formyl peptide receptor, in particular, has served as an important model for the real-time analysis of ligand-receptor interactions based on the detection of fluorescent peptide ligands by solution fluorescence, flow cytometry, and microscopy. These

ligands and receptors have provided the basis for developing subsecond kinetic resolution of molecular assemblies and disassemblies by flow cytometry as well as small volume, high-throughput flow cytometric platforms for screening. Recently, small molecule (nonpeptide) agonists and antagonists have been identified that exhibit selectivity for the FPR and FPRL1.

See also: [Signaling: Chemokine Receptors; G Protein Signaling Regulators; Gi Family of Heterotrimeric G Proteins; G-Protein-Coupled Receptor Kinases and Arrestins; Protein Kinase C Family.](#)

Further Reading

Cicchetti G, Allen PG, and Glogauer M (2002) Chemotactic signaling pathways in neutrophils: From receptor to actin assembly. *Critical Reviews in Oral Biology and Medicine* 13: 220–228.

- Ember JA and Hugli TE (1997) Complement factors and their receptors. *Immunopharmacology* 38: 3–15.
- Haribabu B, Richardson RM, Verghese MW, et al. (2000) Function and regulation of chemoattractant receptors. *Immunologic Research* 22: 271–279.
- Le Y, Murphy PM, and Wang JM (2002) Formyl-peptide receptors revisited. *Trends in Immunology* 23: 541–548.
- Prossnitz ER and Ye RD (1997) The *N*-formyl peptide receptor: A model for the study of chemoattractant receptor structure and function. *Pharmacology and Therapeutics* 74: 73–102.
- Rabiet MJ, Huet E, and Boulay F (2007) The *N*-formyl peptide receptors and the anaphylatoxin C5a receptors: An overview. *Biochimie* 89: 1089–1106.
- Sklar LA, Edwards BS, Graves SW, Nolan JP, and Prossnitz ER (2002) Flow cytometric analysis of ligand–receptor interactions and molecular assemblies. *Annual Review of Biophysics and Biomolecular Structure* 31: 97–119.
- Ye RD, Boulay F, Wang JM, et al. (2009) International Union of Basic and Clinical Pharmacology. LXXIII. Nomenclature for the formyl peptide receptor (FPR) family. *Pharmacological Reviews* 61: 119–161.

Chlorophylls and Carotenoids

H Scheer, Universität München, Munich, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

Absorption The process of interaction of a molecule with a photon to produce an excited state.

Excited states More precisely, electronically excited states; they are usually generated by light absorption in molecules containing a large number of conjugated double bonds. Singlet and triplet states are distinguished by their spin states, leading to large differences in their lifetimes. If the ground state is a singlet, as in practically all dyes, excited singlet states are short lived (10^{-12} – 10^{-8} s), and the lowest triplet states are long lived (up to milliseconds). If the ground state is a triplet, one of the rare cases being molecular oxygen, the lowest excited singlet state is long lived.

Fluorescence It is the spontaneous decay of an excited state to the ground state with the concomitant emission of the energy in the form of a photon. Together with phosphorescence, which is distinguished from fluorescence mainly by its longer timescale ($>10^{-6}$ vs. $<10^{-8}$ s), they are often summarized as luminescence.

Internal conversion (IC) The spontaneous decay of an excited state to the ground state with the concomitant release of the energy in small packages as heat. It is a so-called loss channel, because the heat is (generally) of no use in biology, but is, by the same token, important in light protection (e.g., by preventing photosensitized generation of reactive oxygen species).

Intersystem crossing (ISC) Conversion of a triplet to a singlet state, or vice versa. ISC is generally slow (μ s), but can be sped up by heavy metals or in radical pairs.

The latter process is responsible for the generation of triplets in reaction centers.

Isoprenoids These compounds are formally derived from the branched C_5 -unit, isoprene. Besides carotenoids, important natural isoprenoids are the steroids, gibberellins (plant hormones), mono-, sesqui-, and diterpenes (fragrances and flavors), and polymers such as rubber.

Photodynamic therapy (PDT) This is the diagnostic and therapeutic application, for example, in treatment of cancer, which relies on photosensitizing dyes, light, and oxygen to selectively attack cells, virus, tissue, or organs.

Photosensitization The process of indirectly generating reactive oxygen species by triplet energy transfer. A sensitizing dye is excited by light to the singlet state. This converts spontaneously by ISC to a long-lived triplet state, capable of generating reactive oxygen species by several mechanisms.

Photosynthesis The natural process of plants, algae, and certain bacteria by which (sun)light is converted to biochemical energy (carbohydrates), which is eventually the primary source for life on Earth.

Reactive oxygen species (ROS) These highly aggressive and cytotoxic chemicals, including singlet oxygen, superoxide, peroxide, and hydroxyl radicals, are derived from oxygen. They are often produced, directly or indirectly, by action of light.

Tetrapyrroles These are compounds comprised of four pyrrole rings. In natural tetrapyrroles, they are linked either directly or by single carbon bridges, and they can be linked in a linear (bile pigments) or a cyclic fashion (hemes, chlorophylls, vitamin B_{12} , and F430).

The green (Greek 'chloros') of leaves (Greek 'phyllos') is a mixture of two water-insoluble pigment classes: chlorophylls, mainly the blue chlorophyll (Chl) *a*, and carotenoids, mainly β -carotene. Both pigment classes cooperate in photosynthesis to safely capture light provided by the Sun, as the primary energy source for life on Earth. Chlorophylls are responsible for light harvesting and its transduction to an electrochemical potential across the photosynthetic membrane which ultimately serves to reduce atmospheric carbon dioxide to carbohydrates. Thus, the energy of the fleeting photons is stored in increasingly longer-lived products: first, Chl excited states are generated with lifetimes in the order of several nanoseconds (10^{-9} s), then a membrane potential which is stable for seconds, then high-energy products such as adenosine triphosphate and reduced nicotinamide adenine dinucleotide phosphate that live for minutes or even hours, and eventually storage products such as starch and sugars.

The carotenoids can act, too, as light-harvesting pigments, in the blue spectral regions (450–530 nm) where Chls have

only little absorption. Their major and indispensable function is, however, to safely drain excess excitation energy from the chlorophylls and convert it into heat. By biological standards, photons are a high-energy source which is potentially highly toxic (sunburn). Mutants lacking carotenoids, or enriched in chlorophylls, are quickly bleached and killed in sunlight. The combination of chlorophylls with carotenoids allows photosynthetic organisms a careful balance between successfully competing with other organisms for capturing light, and being killed by an overdose of it. Photosynthesis based on these two pigments has, therefore, conquered nearly all habitats on earth where light is available. It has produced the atmospheric oxygen we breathe, and fixes $\sim 10^{11}$ tons of carbon annually. The greening and degreening of the vegetation is probably the most obvious life process on Earth, visible clearly from outer space. The widespread occurrence of photosynthesis in nearly all ecotopes is paralleled by considerable variations of the photosynthetic apparatus, including the pigments. It is this variety of chlorophylls and carotenoids,

and also basic similarities among the different species, which are dealt within this article.

Chlorophylls

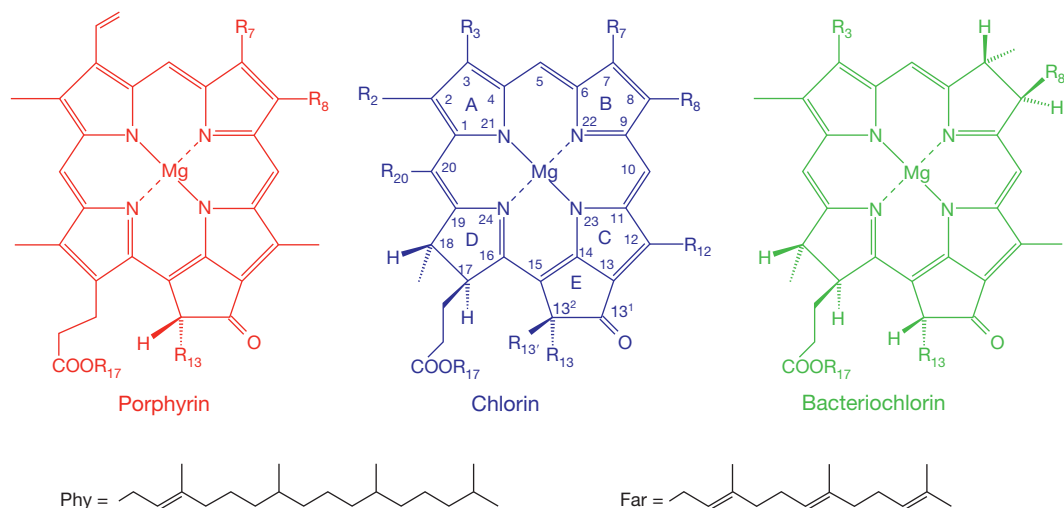
Structures

The basic structure of chlorophylls is the porphyrin macrocycle. It is comprised of four pyrrole rings containing each four carbons and one nitrogen, which are joined via one-carbon bridges to an aromatic macrocycle (Figure 1). The four nitrogens face inward, creating a hole which is ideal for binding metal ions. In the chlorophylls, this central metal is almost always magnesium (Mg). Although many of the peripheral substituents vary among the different structures, all chlorophylls contain an additional five-membered ring E and, with the exception of some Chl *c*, the C-17 propionic acid side chain is esterified by a long-chain alcohol, most commonly the C₂₀

terpenoid alcohol, phytol. The chlorophylls can further be classified by the degree of unsaturation of the macrocycle, into the fully unsaturated true porphyrins, the chlorins with an unsaturated ring D and the bacteriochlorins in which ring B is unsaturated, too.

Spectroscopy

The spectral properties of chlorophylls are described by a modified version of the general four-orbital model of porphyrins resulting in four major absorptions termed Q_Y, Q_X, B_Y, and B_X, in the near-ultraviolet (NUV), visible (Vis), and near-infrared (NIR) spectral regions. The type of spectrum is mainly determined by the degree of unsaturation of the tetrapyrrole macrocycle (Figure 2): in the fully unsaturated Chl *c* containing a porphyrin macrocyclic system, the B-bands around 400 nm overlap, the Q_Y-band around 620 nm is weak, and Q_X-band is visible only with special techniques. In chlorin-type



Pigment	R ₂	R ₃	R ₇	R ₈	R ₁₂	R _{13'}	R ₁₃	R ₁₇	R ₂₀	Macrocycle
Chl <i>a</i>	CH ₃	C ₂ H ₃	CH ₃	C ₂ H ₅	CH ₃	H	COOCH ₃	Phy ^c	H	Chlorin
Chl <i>b</i>	CH ₃	C ₂ H ₃	CHO	C ₂ H ₅	CH ₃	H	COOCH ₃	Phy	H	Chlorin
Chl <i>d</i>	CH ₃	CHO	CH ₃	C ₂ H ₅	CH ₃	H	COOCH ₃	Phy	H	Chlorin
Chl <i>f</i>	CHO	C ₂ H ₃	CH ₃	C ₂ H ₅	CH ₃	H	COOCH ₃	Phy	H	Chlorin
Phe <i>a</i> ^d	CH ₃	C ₂ H ₃	CH ₃	C ₂ H ₅	CH ₃	H	COOCH ₃	Phy	H	Chlorin
BChl <i>c</i>	CH ₃	CHOH-CH ₃	CH ₃	C ₂ H _{5-n} (CH ₃) _n	CH ₃ /C ₂ H ₅	H	H	Far + others	CH ₃	Chlorin
BChl <i>d</i>	CH ₃	CHOH-CH ₃	CH ₃	C ₂ H _{5-n} (CH ₃) _n	CH ₃ /C ₂ H ₅	H	H	Far + others	H	Chlorin
BChl <i>e</i>	CH ₃	CHOH-CH ₃	CHO	C ₂ H _{5-n} (CH ₃) _n	CH ₃ /C ₂ H ₅	H	H	Far + others	CH ₃	Chlorin
BChl <i>a</i>	CH ₃	CO-CH ₃	CH ₃	C ₂ H ₅	CH ₃	H	COOCH ₃	Phy + others	H	Bacteriochlorin
BChl <i>b</i>	CH ₃	CO-CH ₃	CH ₃	=CH-CH ₂	CH ₃	H	COOCH ₃	Phy + others	H	Bacteriochlorin
BChl <i>g</i>	CH ₃	C ₂ H ₃	CH ₃	=CH-CH ₂	CH ₃	H	COOCH ₃	Far	H	Bacteriochlorin
BPhe <i>a</i> ^d	CH ₃	CO-CH ₃	CH ₃	C ₂ H ₅	CH ₃	H	COOCH ₃	Phy	H	Bacteriochlorin
Chl(ide) <i>c</i> ₁	CH ₃	C ₂ H ₃	CH ₃	C ₂ H ₅	CH ₃	H	COOCH ₃	H ^{a,b}	H	Porphyrin
Chl(ide) <i>c</i> ₂	CH ₃	C ₂ H ₃	CH ₃	C ₂ H ₃	CH ₃	H	COOCH ₃	H ^{a,b}	H	Porphyrin
Chl(ide) <i>c</i> ₃	CH ₃	C ₂ H ₃	COOCH ₃	C ₂ H ₅	CH ₃	H	COOCH ₃	H ^{a,b}	H	Porphyrin
[8-Vinyl]-PChlide <i>a</i>	CH ₃	C ₂ H ₃	CH ₃	C ₂ H ₅	CH ₃	H	COOCH ₃	H	H	Porphyrin

(a) Sometimes esterified, (b) Acrylic side chain at C-17 (17¹-17² double bond).

(c) 2,6-Phytadienol in green bacterial reaction centers, (d) no central metal.

Figure 1 Structures of common chlorophylls. In particular, reaction centers contain a series of specialized chlorophylls that are being dealt with in a survey by M Kobayashi and coworkers in 2006. The structure of chlorophyll *f* has been described by Chen et al. in 2010.

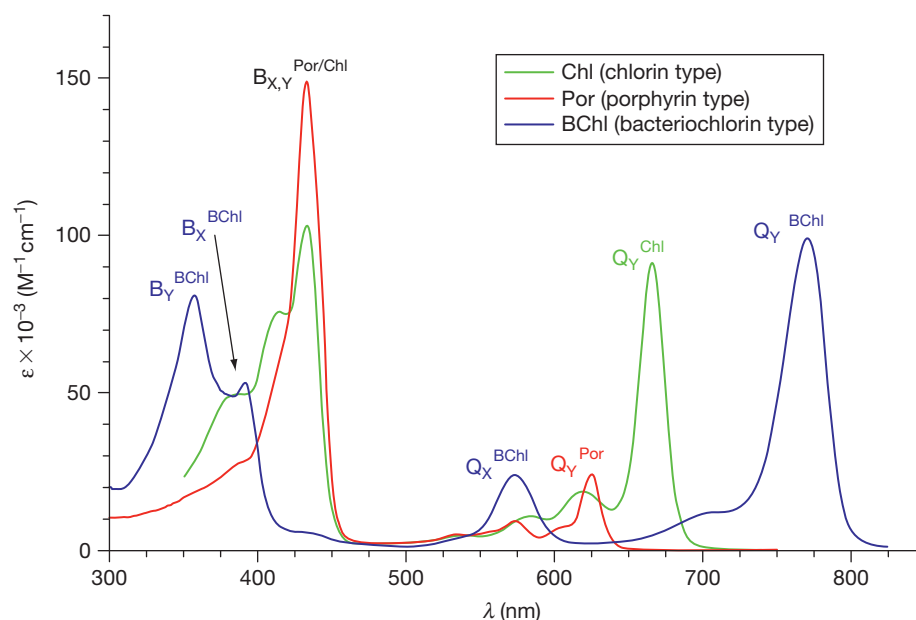


Figure 2 Type spectra of the three chlorophyll classes: porphyrins (Por), chlorins (Chl), and bacteriochlorins (BChl). Note that the bacteriochlorophylls *c*, *d*, and *e* are not bacteriochlorins, but rather chlorins (see [Figure 1](#)).

chlorophylls (Chl *a*, *b*, *d*, BChl *c*, *d*, *e*), the B-bands are reduced in intensity, the Q_x-band increased to near equal intensity and redshifted to ~660 nm. In the bacteriochlorin-type chlorophylls (BChl *a*, *b*, *g*), the B-bands are blueshifted to <400 nm and well separated, the Q_x-band is even more increased in intensity and redshifted to ~780 nm, and in these pigments the Q_y-band also has gained in intensity and becomes visible around 570 nm. By these characteristics, the chlorophyll type can be readily determined from the spectra, and often also other details like substitution pattern, ligation to the central metal (e.g., by the protein), or aggregation.

Chlorophylls dissolved in organic solvents have long-lived-excited states (~10⁻⁸ s). They are, therefore, highly fluorescent. Consequently, there is significant intersystem crossing possible to the triplet state. This gives rise to phosphorescence and, more significantly, under aerobic conditions to the generation of highly reactive oxygen species (ROS) like singlet oxygen. Chlorophyll solutions, therefore, bleach rapidly by attack of the pigment by these ROS, and chlorophylls injected into animals lead to severe 'sunburn' and destruction of tissue. This effect is used in photodynamic therapy of cancer.

In chlorophyll aggregates, there is generally a pronounced redshift of the Q_y-band, and the excited-state lifetimes (and thereby fluorescence and phosphorescence) are drastically reduced due to rapid conversion of the excitation energy into heat by internal conversion (IC). These effects are summarized under the term 'concentration quenching' and probably due to the presence of 'traps' in the large aggregates in which the excitation is highly delocalized over many pigment molecules. An exception are the chlorosomes of green bacteria: they contain aggregates of bacteriochlorophylls *c*, *d*, or *e* which show little concentration quenching, or only under conditions of excess light.

Aggregation is also observed in photosynthetic pigment-protein complexes, where it is considered a major organizing

force. In addition, there is a general redshift of all bands brought about by the protein environment. There is, however, a very distinct difference to aggregates formed *in vitro*: although excitation energy is mainly transformed to heat in the latter, this loss is avoided in the pigment protein complexes. Isolated light-harvesting complexes (LHCs), which contain the vast majority (≥99%) of chlorophylls, therefore, show, under moderate light intensities, high fluorescence. The avoidance of concentration quenching in these aggregates is in fact one of the major, mechanistically still unresolved tricks of efficient photosynthesis. In the reaction centers (RCs), the lifetime is short, but here the excitation energy is used to create, with quantum efficiencies near 100%, a charge separation across the photosynthetic membrane which is otherwise an insulator, thus creating a long-lived membrane potential. This energy transduction process is at the heart of photosynthesis, and the high-resolution crystal structures of RC together with extensive spectroscopy at high time resolution have provided a general picture on how this process is working. In the photosynthetic apparatus, LHCs containing >99% of the chlorophylls are coupled to RC, such that now the excitation energy of the former is efficiently transferred to the latter. By this means, the absorption of the RC is increased by orders of magnitude, and different numbers and types of LHC can be combined to adapt to the prevailing light conditions. As the amount of protein is much less in LHC than in RC, the biosynthetic expenses for such a coupled system are at the same time strongly reduced. There is nonetheless an often overlooked loss of energy that is inherent to the antenna/RC concept: antennas absorb over most of the Vis and NIR spectral region, and the energy is transported to the RCs absorbing, and working with low-energy red or NIR excitation. By this process, any excess energy of the absorbed photons (compared to the RC energy) is lost as heat. Depending on the system used, these losses can be substantial, but must be ecologically still less than with other solutions.

Occurrence and Functions

The majority of chlorophylls, both in terms of structure diversity and quantity, serve as light-harvesting pigments. Taken together, they absorb light across most of the spectral range from 350 to 1020 nm, with a gap only in the region from 500 to 600 nm where carotenoids and biliproteins add to light harvesting. No organism contains all types of chlorophylls; however, so different organisms occupy different spectral niches, and the pigmentation is an important taxonomic criterion. Due to the more demanding energetics, oxygenic organisms (plants, algae, and cyanobacteria) can only use light <750 nm, while the anoxygenic photosynthetic bacteria have specialized in the NIR with extremes such as the purple bacterium, *Blastochloris viridis*; reaching 1020 nm. RCs contain chlorophylls, too, including besides the bulk pigments some specialized derivatives. They form a chain of pigments across the membrane, over which in a stepwise fashion electrons are transferred from the primary donor, a pair of chlorophylls located on the periplasmic side of the membrane, to the acceptors on the cytoplasmic side. It should be mentioned that some chlorophyll-like pigments perform other functions in nature: examples are certain deep-sea fish using chlorophylls as visual pigments, and the marine worm, *Bonella viridis*, where chlorophyll derivative acts as a sex determinant.

Chemical Properties

Chlorophylls are large, aromatic molecules with a central hole fit to complex many metals. They are basically planar and fairly rigid, but small deviations from planarity are observed in most structures of chlorophyll proteins. Although the macrocyclic system is quite stable, the functional groups and the central Mg render the chlorophylls quite labile. Under acidic conditions, the central Mg is lost rapidly, and the long-chain alcohol residue hydrolyzed. The central Mg is also readily replaced by other, more stable metals such as Zn, Ni, Ag, Cu. As most of these replacements lead to pigments that are less efficient for photosynthesis, part of the metals' toxicities relate to impairment of photosynthesis. There is one exception, however: a photosynthetic bacterium from a strongly acidic biotope contains a bacteriochlorophyll *a* in which the central Mg is replaced by Zn, thereby probably stabilizing it to acid-induced demetalation. Under alkaline conditions, the isocyclic ring is modified extensively. The macrocycle can be opened at the methine bridges, both (photo) chemically and enzymatically during degradation.

After binding to the tetrapyrrole, the central Mg still has free valences, which are important for aggregation and, in particular, for interactions with the natural protein environment. By interactions with suitable amino acids (e.g., histidin, methionin, and glutamate) they can be positioned properly for efficient energy or electron transfer. In the LHC of green bacteria, the chlorosomes, the coordinating properties of the central Mg and the ligation of special peripheral substituents combine to form large aggregates of bacteriochlorophylls *c*, *d*, and *e* that are nearly devoid of protein.

Metabolism

The formation of chlorophylls can be separated in stages. The first is the synthesis of 5-aminolevulinic acid (ALA), the basic

building block of all tetrapyrroles. In most photosynthetic organisms, it is formed from glutamic acid, involving an intermediate (Glu-tRNA^{Glu}) which is otherwise only involved in protein synthesis. Some photosynthetic bacteria (purple bacteria) use, like humans, the alternative C₄+1-pathway in which succinyl coenzyme A is condensed with glycine. The second stage is ubiquitous in all organisms and for all natural tetrapyrroles. The two ALA first react to yield a pyrrole (porphobilinogen), a five-membered ring containing one nitrogen atom. Four such pyrroles condense to a linear tetrapyrrole, and then cyclize in a remarkable reaction in which one of the pyrroles is flipped. At this stage, the pathways to vitamin B₁₂, certain hemes, and certain Ni-tetrapyrroles branch off. In chlorophyll biosynthesis, the resulting cyclic tetrapyrrole is transformed by a series of decarboxylation and oxidation reactions to protoporphyrin. It is noteworthy that the latter is the first colored product in the synthesis, and highly phototoxic like most porphyrins. The organisms, therefore, avoid, over large stretches of biosynthesis, the formation of such potentially harmful precursors. They also regulate very tightly the amount of ALA in order to avoid any overproduction of protoporphyrin. Any deregulation of this process, as any other alteration of the biosynthesis can result in severe damage and often death. The next stage of the pathway is the insertion of Mg, which is thermodynamically unfavorable and energy requiring, and the formation of the isocyclic ring from the propionic acid side chain at C-13. Although the synthesis of Chl *c* is practically complete at this stage, all other chlorophylls require further modifications, viz., the reduction of ring D (chlorins) and B (bacteriochlorins), modifications at the periphery, and the condensation of the C-17 propionic acid with the long-chain alcohol derived from the isoprenoid pathway.

Chlorophyll degradation has been studied in some detail for Chl *a* and *b*, it involves the (light independent) ring opening to the much less-phototoxic open-chain tetrapyrroles (bilins). Very little is known on the degradation of the other pigments, and on potential functions of the degradation products.

Applications

Unmodified chlorophylls are too labile for most practical use, but some derivatives are used as dyes for cosmetics and food (Cu-chlorophyllin), and in photodynamic tumor therapy (chlorins and bacteriochlorins). The chlorophyll used, for example, in certain health care products is a complex mixture of degradation products. The best source for Chl *a* is the cyanobacterium *Spirulina platensis*, which is available commercially. Chl *a/b* mixtures can be obtained from all green plants. Bacteriochlorophyll *a* can be obtained from certain photosynthetic bacteria that can be grown in the dark. All other chlorophylls are less readily accessible, thereby limiting their applications.

Carotenoids

Structures

Carotenoids are much more widespread and not confined to photosynthetic organisms, and structurally and functionally more diverse than chlorophylls. There are currently more than

800 natural carotenoids known, each can, furthermore, form several *cis-trans* isomers and be additionally modified. Chemically, most carotenoids are tetraterpenes: two C_{20} -units (originally geranyl-geraniol) are joined tail-to-tail to a chain of 32 carbon atoms bearing eight methyl side chains. Often, this basic carbon C_{40} -skeleton is either retained, or only slightly modified, for example, by cyclization at one or both ends (Figure 3). However, much more extensive modifications are possible, including isomerization and rearrangement of the double bonds, removal of carbon atoms, the introduction of oxygen-containing functional groups, and their glycosylation or acylation. These modification seems to be particularly

far-reaching in carotenoids dedicated to light harvesting, as exemplified by peridinin; a highly modified C_{37} pigment from algae. Some carotenoids are also derived from C_{30} or C_{50} skeletons.

Spectroscopy

The rod-shaped carotenoids show the typical absorption spectra of linear polyenes, which are characterized by some unusual features. (1) The lowest energy $S_0 > S_1$ transition located in the red to NIR spectral range, is extremely weak (optically forbidden) in most carotenoids, and has been accurately determined

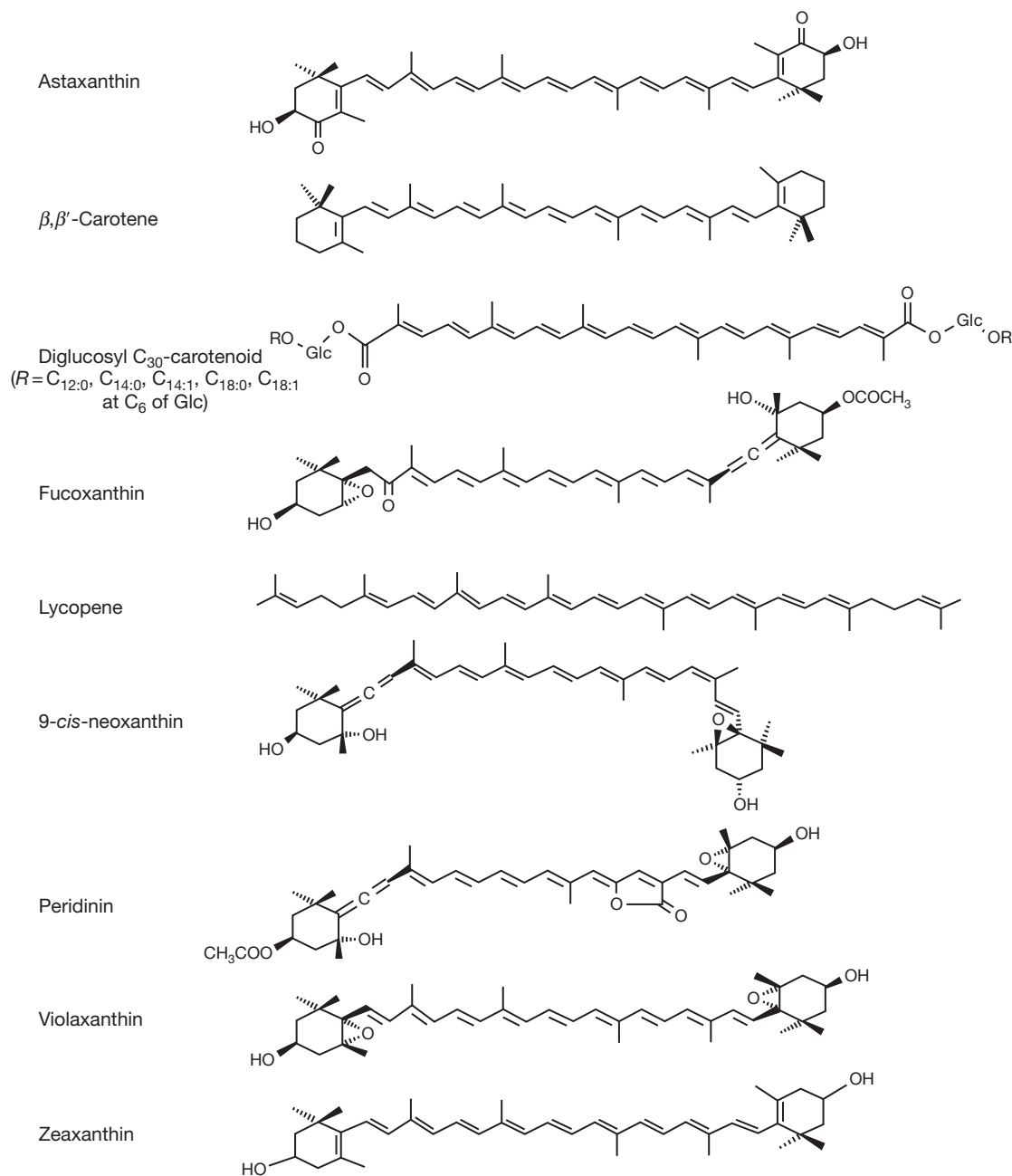


Figure 3 Examples of carotenoid structures.

only in few pigments by special techniques. (2) The most intense absorption, responsible for the yellow–orange color of most carotenoids, is a series of closely spaced, sometimes overlapping bands in the 400–550 nm range. With an increasing number of conjugated double bonds, this absorption is shifted, in an asymptotic fashion, to the red. The sub-bands are generally well resolved, but can be broadened to a degree that they appear only as a single, somewhat structured band. (3) There several other transitions between these two bands, which are also optically forbidden and, therefore, very weak. (4) *cis*-Carotenoids show an additional, typical band 100–150 nm to the blue of the main absorption (see Figure 4 for type spectra). Although the ‘forbidden’ bands do not contribute to absorption, they are fundamentally important for the biological functions of carotenoids. The states responsible for them can be reached indirectly, for example, after absorption into the energy-rich major absorption band and IC, or by energy transfer from neighboring pigments like chlorophylls by electron exchange and they thereby contribute to energy transfer and dissipation.

Most carotenoids show extremely rapid IC from all excited states and, therefore, negligible fluorescence. Their excited state have generally lifetimes of only a few picoseconds, and reach even in extreme cases like peridinin only 100 ps. Carotenoids have unusually low-lying triplet states, many of them even below the energy of singlet oxygen (1250 nm). Again, they can be reached only indirectly, mainly via short range (≤ 3.5 Å) energy transfer from triplets (e.g., of chlorophylls) or from singlet oxygen. As the latter are highly cytotoxic, this property is functionally very important, too, because it is the basis for the protective effects of carotenoids.

The spectral properties of carotenoids can be modified considerably by interactions with the protein. A spectacular case is the color change of astaxanthin from orange to green when it is bound in the crustacean protein, α -crustaxanthin, and its reversion when the protein is denatured by boiling.

Occurrence and Functions

Carotenoids are almost ubiquitous in living organisms, even though animals including man cannot synthesize them but have to rely on dietary supplies.

Their functions are as diverse as their structures. In photosynthesis, their essential function is the protection of the photosynthetic apparatus under conditions of light overload. Whenever the RC cannot follow the energy input from the LHC, the long-lived-excited states of the chlorophylls pose a deadly danger to these organisms due their potential of generating ROS. This is prevented by carotenoids at several levels. They can accept excess energy via their low-lying forbidden states, they can quench chlorophyll triplets by triplet energy transfer to the low-lying carotenoid triplets, and they can quench ROS by energy transfer, for example, from singlet oxygen, or addition of ROS to the double-bond system. All these processes require very close distances between donor and acceptor, which are nicely corroborated by X-ray structures of photosynthetic complexes. All chlorophyll-containing LHC have carotenoids in contact with certain chlorophylls, and specialized 15-*cis*-carotenoids are found in the RC. In the LHC, protective carotenoids seem to be positioned strategically at critical sites where the energy is funneled to the RC. As energy will flow preferentially downhill, the accepting states of protective carotenoids need to lie very close to those of the donors.

As ROS are also important in the attack of infectious agents by the immune system, bacteria synthesize carotenoids for their protection, and many of the more virulent forms are deeply colored by carotenoids.

Carotenoids also protect from light, simply by their function as light filters, removing NUV and blue light by virtue of their high absorption in these spectral regions, and their capacity to rapidly degrade the excitation energy to heat by IC. Certain algae tolerant to extreme light stress contain droplets of pure carotenoids, and many nonphotosynthetic organisms

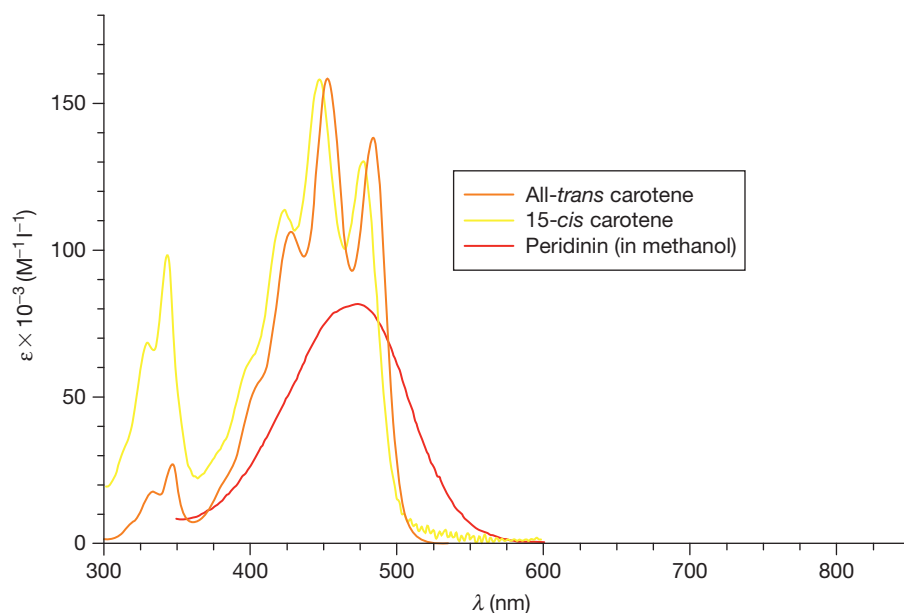


Figure 4 Type spectra of carotenoids.

synthesize more carotenoids when subjected to increasing light intensities. As nonphototoxic pigments, carotenoids often function as 'safe' dyes in nature. Examples are the colors of flowers used for attracting insects, or of crustaceans to blend into the marine background.

In spite of their general protective function, carotenoids have also been adapted by photosynthetic organisms for light harvesting in the 'green gap' (470–600 nm) where chlorophylls absorb only poorly, in particular in microalgae and phototrophic bacteria which, unlike higher plants, cannot outgrow their competitors. Two factors are important to this function. One is the development of carotenoids in which the excited-state lifetime is somewhat increased and, therefore, IC reduced. The two most-abundant carotenoids, fucoxanthin and peridinin, show this very nicely with lifetimes reaching up to 100 ps. The second factor is positioning of carotenoids in contact with chlorophylls, thereby speeding up energy transfer sufficiently to be effective within the short excited-state lifetimes (electron-exchange mechanism). As the direction of energy transfer depends on the relative energies of the excited states, they have to be properly ordered: for light-protection the carotenoid excited state needs to be below that of a neighboring chlorophyll; for light harvesting it needs to be above. There is evidence that plants and some algae use, in the so-called violaxanthin cycle, subtle structural modifications in order to manipulate the excited state energies of carotenoids, thereby converting a light harvesting into a protecting pigment, and vice versa, in response to the light supply and energetic demands.

Last but not least are carotenoids precursors for certain important metabolites. Only two examples are stated. (1) Retinal, the chromophore of the visual pigment rhodopsin, is derived from carotene. As the latter cannot be synthesized by animals, they need it supplied as provitamin A. Retinal derivatives are also required for other regulatory functions. (2) Certain fragrances of roses have also been shown not to be synthesized directly, but rather to be breakdown products of the flowers' carotenoids.

Chemical Properties

Due to their hydrocarbon skeleton, almost all carotenoids are insoluble in water and well soluble in nonpolar solvents. The conjugated double-bond system is chemically moderately stable. It is subject to rearrangements, in particular in the light, and to additions, for example, of oxygen. The chemical properties of the more highly modified carotenoids are determined by their particular functional groups. Only few examples shall be given: the epoxide function present in violaxanthin is very acid sensitive, *cis*-carotenoids are readily (photo)converted to the more stable *trans*-carotenoids, and lactones or esters can be hydrolyzed by acid or base.

Metabolism

Carotenoids, like all terpenoids, are products of the isopentenyl-pyrophosphate metabolism. The basic C₅ building block can be synthesized either by the mevalonate pathway from acetyl-coenzyme A, or via the more recently discovered deoxyxylulose pathway. The latter seems to be the pathway leading to carotenoids in most photosynthetic organisms.

One of the resulting C₅-units (dimethylallyl-pyrophosphate) is condensed sequentially by prenyl transferases in a head-to-tail fashion with three isomeric C₅-units (isopentenyl-pyrophosphate) to yield geranylgeranyl-pyrophosphate (GGPP), an important branching point in terpenoid metabolism. GGPP is, for example, directly or indirectly, the substrate for esterifying the C-17 propionic acid side chain of most chlorophylls, thereby linking the tetrapyrrole- and terpenoid pathways. The biosynthesis directed to most carotenoids begins with a tail-to-tail condensation of two molecules of GGPP by the dedicated enzyme, phytoene synthase. Having only three conjugated double bonds, the resulting phytoene is still uncolored. Dehydrogenation by desaturases via phytofluene to lycopene is common to most carotenoids currently known. C₃₀ carotenoids are derived from condensation of two C₁₅-units (farnesol-pyrophosphate) in an otherwise very similar reaction sequence. A wide variety of enzymes, many of them still only partly characterized, is responsible for further structural modifications. They include cyclases to form the end rings characteristic of many carotenoids, oxygenases to introduce OH groups, isomerases and (de)hydrogenases to modify the double-bond system, and many more follow-up enzymes allowing for the structural variety of carotenoids. Although *cis-trans* isomerases related to the visual systems are well studied, the formation of 15-*cis*-carotenoids in the RC is unclear and may be a spontaneous reaction during incorporation of the respective 15-*trans* isomer.

Applications

The most frequent applications of carotenoids, which are of considerable economic importance, are as food colorants, as dietary supplies, and as sun screens in cosmetics. Food is also frequently colored indirectly, for example, by supplying carotenoids or carotenoid-rich algae in the feedings of fish (salmon) or poultry (chicken eggs). Provitamin A (β-carotene) is supplied not only directly but also indirectly, for example, as milk supplement. Industry-scale synthetic methods have been developed, natural sources are algae such as *Hematococcus*, plants such as carrots, and genetically manipulated bacteria.

See also: **Bioenergetics:** Chloroplasts; Chloroplast Signaling and Redox Control; Light-Harvesting Complex I and II: Pigments and Proteins; Structure–Function of the Cytochrome *b₆f* Complex of Oxygenic Photosynthesis; **Metabolism Vitamins and Hormones:** Photosynthesis; Photosynthetic Carbon Dioxide Fixation.

Further Reading

- Britton G, Liaaen-Jensen S, and Pfander H (eds.) (1995) *Carotenoids*. Basel: Birkhäuser.
- Chen M, Schliep M, Willows RD, et al. (2010) A red-shifted chlorophyll. *Science* 329: 1318–1319.
- Deisenhofer J and Norris JR (eds.) (1993) *The Photosynthetic Reaction Center*. New York, NY: Academic Press.
- Eisenreich W, Rohdich F, and Bacher A (2001) Deoxyxylulose phosphate pathway to terpenoids. *Trends in Plant Science* 6: 78–84.
- Frank HA, Young AJ, Britton G, and Cogdell RJ (eds.) (1999) *The Photochemistry of Carotenoids*. Dordrecht: Kluwer.

- Green B and Parson WW (eds.) (2003) *Light-Harvesting Antennas in Photosynthesis*. Dordrecht: Kluwer.
- Grimm B, Porra R, Rüdiger W, and Scheer H (eds.) (2006) *Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications*. Dordrecht: Springer.
- Kobayashi M, Akiyama M, Kise H, and Watanabe T (2006) Unusual terapyrrole pigments of photosynthetic antennas and reaction centers: Specially tailored chlorophylls. In: Grimm B, Porra R, Rüdiger W, and Scheer H (eds.) *Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications*. Dordrecht: Springer.
- Moser JG (ed.) (1998) *Photodynamic Tumor Therapy: 2nd and 3rd Generation Photosensitizers*. Amsterdam: OPA.
- Moser S, Müller T, Oberhuber M, and Kräutler BM (2009) Chlorophyll catabolites – chemical and structural footprints of a fascinating biological phenomenon.

- European Journal of Organic Chemistry* 2009: 31–35. <http://onlinelibrary.wiley.com/doi/10.1002/ejoc.200800804/pdf> (accessed April 2011).
- Scheer H (ed.) (1991) *Chlorophylls*. Boca Raton, FL: CRC Press.
- Vernon LP and Seely GR (eds.) (1966) *The Chlorophylls*. New York, NY: Academic Press.

Relevant Websites

- <http://www.carotenoidsociety.org> – International Carotenoids Society.
- <http://www.photosynthesisresearch.org> – International Society of Photosynthetic Research.
- <http://terra.nasa.gov> – National Aeronautics and Space Administration; Images and Data.

Chloroplasts

N Rascio, University of Padua, Padua, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Binary fission Mechanism, similar to that occurring in bacteria, by which the chloroplast divides independently from the cell division.

Envelope Pair of concentric membranes enclosing the chloroplast.

Etioplast Organelle which is the counterpart of chloroplast and develops in dark-grown seedlings of flowering plants.

Prolamellar body Tridimensional semi-crystalline aggregate of the etioplast tubular membranes.

Thylakoids Flattened cisternae or saccules intercommunicating, which form the closed inner membrane system of chloroplast.

Chloroplast Origin

Some 3.5 billion years ago, an ancestral form of cyanobacterium with the contemporary presence in the cell of a green and a purple bacterium anoxygenic photosystem appeared on the earth. A deep re-elaboration of the new photosynthetic apparatus, which required about 1 billion years, led, at the end, this kind of microorganism to gain the ability to use water as source for electrons, thus giving rise to the oxygen-evolving photosynthesis. This occurred through: the link in series of the two photosystems; the change of bacteriochlorophyll with chlorophyll absorbing at shorter and more energetic wavelength; the rise of the P680 (the photoactive chlorophyll of photosystem II, derived from the purple bacterium one) midpoint reduction potential over that of $\text{H}_2\text{O}/\text{O}_2$ (+0.82 V); the evolution of a Mn-protein complex able to split water-releasing oxygen, electrons, and protons.

The start of the oxygenic photosynthesis was a central event in the life development on our planet, even deemed the Big Bang of life. The oxygen release in the environment changed the face of the Earth by transforming the anaerobic atmosphere to one suitable for the evolution of aerobic forms of life.

A free-living aerobic prokaryote (probably a *rickettsia*-like α -proteobacterium), endosymbiotically enslaved by an anaerobic protoeukaryote, gave rise to the mitochondrion and this allowed eukaryotic heterotroph organisms to evolve.

Successively, a cyanobacterium-like prokaryote with oxygenic photosynthesis, engulfed by an aerobic heterotroph protoeukaryote, originated the chloroplast.

This latter endosymbiotic event, which took place about 1 billion years ago, was of enormous importance, because it triggered the evolution of photosynthetic eukaryotes and even led, as a consequence, to the expansion of the heterotroph forms of life on the Earth. The photoautotroph organisms, in fact, are the primary producers which supply with energy and organic matters to most of ecosystems.

During the evolutionary route which changed it into the chloroplast, the cyanobacterium-like ancestor underwent substantial modifications. A significant event was the transfer of most prokaryotic DNA to the host cell nucleus. The genome of a present chloroplast, actually, codes for ~100 proteins, while

the DNA of a free-living cyanobacterium, like *Synechocystis*, whose entire sequence is known, encodes more than 3000 proteins. The massive 'endosymbiotic gene transfer' possibly occurred to avoid the mutagenic load for organelle genes due to free radicals produced by photosynthetic light reactions, and to escape the accumulation of deleterious mutations caused by the genetic isolation of the organelle DNA in the host cells (the so-called Muller's ratchet).

This transfer of genetic information made it necessary for the chloroplast to recover the proteins encoded in the nucleus and this was achieved by inserting in the nuclear genes a presequence coding for a transit peptide which targeted the chloroplast proteins and redirected them to the organelle.

Another essential change undergone by the evolving chloroplast concerned the photosynthetic light-harvesting pigments. With the exception of some algal groups (Rhodophytes and Cryptophytes), the original phycobiliproteins were replaced by chlorophyll forms, in particular by chlorophyll *b* in green organisms.

Chloroplast Organization

Algae and lower plants have chloroplasts which can vary considerably in shape and size, whereas in higher plants these organelles are commonly lens-shaped and measure ~5–10 μm .

Envelope

The chloroplast (Figure 1) is enclosed by a pair of concentric membranes forming the envelope. The inner envelope membrane is highly selective and contains numerous carriers which regulate the flows of metabolites and ions between chloroplast and the cytosol. The most abundant is the phosphate translocator, which plays a major role in supplying the cell with the products of photosynthesis. The inner membrane also supports some enzymes of biosynthetic pathways, like those of carotenoids, tetrapyrroles, and fatty acids. The outer envelope membrane contains a nonspecific porine-like protein,

which allows ions and metabolites of up to ~10 kDa to pass. It is essentially involved in recognizing the cytosol-synthesized chloroplastic proteins which must be transferred into the organelle. This translocation occurs through two multiproteic complexes, one of them inserted in the outer membrane (translocation outer complex) and the other in the inner membrane (translocation inner complex) of the envelope.

Stroma

The chloroplast internal milieu, called stroma, looks finely granular due to the presence of 70S ribosomes. It contains prokaryotic DNA (Figure 1), RNA, and the entire protein synthesis machinery allowing the organelle to translate its own genetic information. In the stroma are also inserted the soluble enzymes engaged in the numerous biosynthetic pathways. The ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), which catalyzes CO₂ fixation in the first reaction of the Calvin-Benson cycle, is the most abundant enzyme, and is

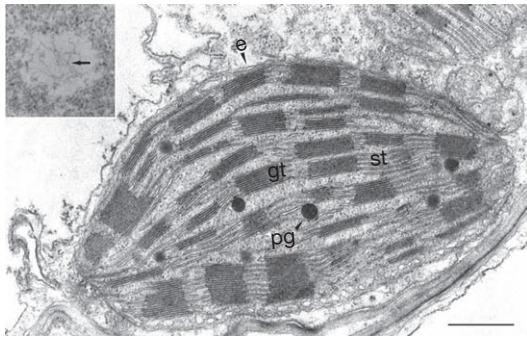


Figure 1 Transmission electron microscope (TEM) micrograph of a chloroplast from a mesophyll cell in a maize leaf. Note the envelope (e) the inner membrane system formed by stacked granal thylakoids (gt) and single stromal thylakoids (st). Some plastoglobules (pg) are present in the stroma, which look finely granular due to the presence of 70S ribosomes. Scale = 1 μ m. The inset shows the region of a chloroplast stroma with DNA microfibrils (arrow).

even the most abundant protein in the world. It is, actually, a large protein complex consisting of eight small subunits (SSUs) and eight large subunits (LSUs) encoded by nuclear genes and plastidic genes, respectively.

Small roundish inclusions, named plastoglobules, made up of plastoquinones, carotenoids, and proteins, are common in the stroma, where also starch grains of photosynthetic origin can occasionally accumulate.

Thylakoid System

Within the chloroplast a membrane system, the thylakoid system, is present, which contains the multiproteic complexes involved in the light reactions of photosynthesis. The inner membrane system of the chloroplast (Figure 1) consists of coupled lamellae (thylakoid membranes) separated by a narrow intermembrane space. They form flattened cisternae or saccules (thylakoids), intercommunicating to constitute a continuous and closed membrane system, whose internal space (lumen) is totally isolated from the stromal environment.

A feature common to the green plant chloroplasts (Figure 2) is the stacking of thylakoids (granal thylakoids) one upon the other to make piles of saccules named grana (singular: granum), which are interconnected by single unstacked thylakoids running in the stroma (stromal thylakoids). This kind of organization, which gives rise to a huge surface area of thylakoid membranes, acquires functional significance, taking into account that, as stated above, these membranes support the complexes carrying out the photosynthetic light reactions and the correlated electron transport and adenosine triphosphate (ATP) synthesis (precisely: photosystem I (PSI), photosystem II (PSII), the cytochrome *b₆f* complex (Cyt *b₆f*), and the CF₀-CF₁ ATP synthase). These components are diversely and specifically located in the thylakoid membranes (Figure 3). Most of PSII resides in the approached regions of stacked thylakoids (partitions), while PSI and the ATP synthase only occur in the stromal thylakoids and in all the unstacked regions of granal thylakoids (end granal membranes and margins). Finally, the Cyt *b₆f* is rather homogeneously distributed in the thylakoid system.

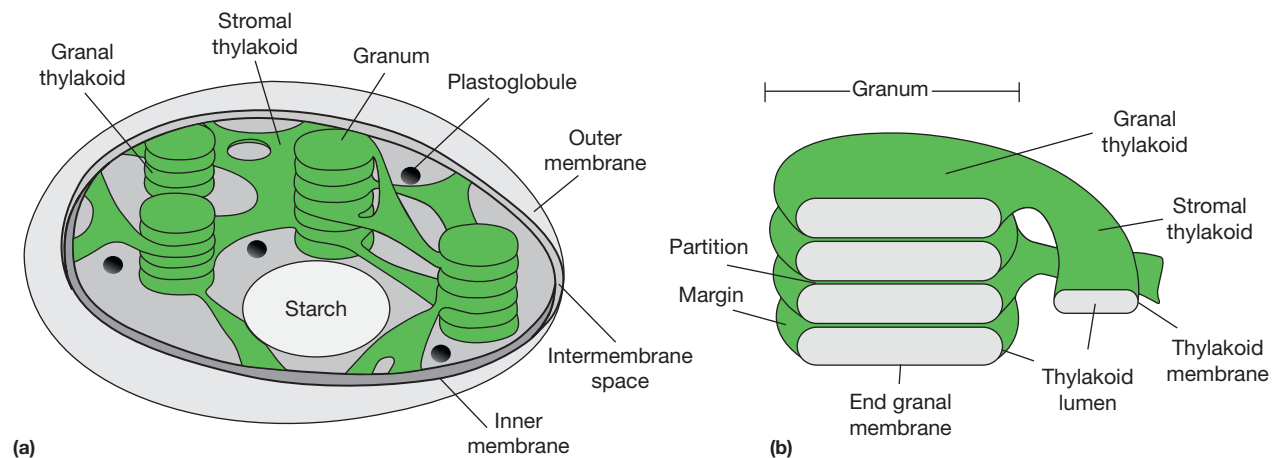


Figure 2 (a) Three-dimensional scheme of a chloroplast. (b) Particulars of the thylakoid system organization.

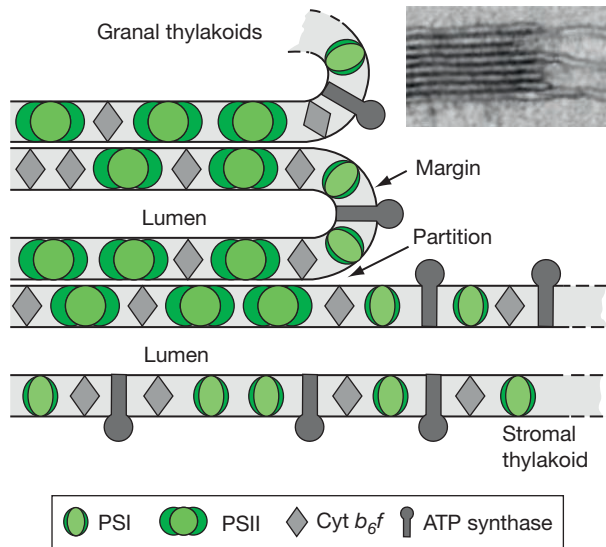


Figure 3 Distribution of the multiprotein complexes in the thylakoid membranes.

The thylakoid membranes contain several other proteins, among which of particular interest are a plastid-encoded reduced nicotinamide adenine dinucleotide phosphate (NADPH): plastoquinone oxidoreductase homologous to the complex I of the bacterial respiratory chain and a nucleus-encoded terminal plastoquinol oxidase homologous to the plant mitochondrial cyanide-resistant alternative oxidase. Both these protein components are involved in a physiological process named chlororespiration.

A peculiarity of chloroplast membranes, which distinguishes them from the other cell membranes, is the great abundance of glycolipids, represented by two neutral galactolipids, the monogalactosyldiacylglycerol (MGDG) and the digalactosyldiacylglycerol, which account for 70% of the entire acyl-lipid moiety, and by the anionic sulfolipid sulfoquinovosyldiacylglycerol. Phosphatidylglycerol is the only phospholipid present in the membranes. A feature of all thylakoid lipids is the high concentration of polyunsaturated fatty acids, which maintain membrane fluidity in spite of the high protein/lipid ratio. Nonpolar lipids, such as plastoquinones, phylloquinones, and tocopherols, are inserted in the hydrophobic region of the membrane matrix. Plastoquinone-9, a component of the photosynthetic electron transport chain, is particularly abundant.

Chloroplast Biogenesis

In higher plants, chloroplasts develop from proplastids, which are small undifferentiated organelles present in the meristematic cells. Chloroplast differentiation involves the synthesis of pigments, proteins, and lipids and the structural and functional organization of the whole thylakoid system, as well as the acquisition of the enzymatic components carrying out the numerous metabolic pathways of the mature organelle. These events are accompanied by the increase in the developing chloroplast size and the rise of the organelle number in the cells. Proplastids and young chloroplasts, indeed, can divide through a process of 'binary fission' similar to that occurring in bacteria.

The chloroplast division is independent of cell division, being part of the developmental program that defines the final number of photosynthetic organelles in the mature green cell.

The dependence on both nuclear and own genetic information makes chloroplast biogenesis a quite complex process requiring the coordinated expression of genes located in the two cellular compartments. This takes place through environmental cues, the most important of which is light, and through endogenous nucleus–chloroplast signaling. The nucleus exerts a preeminent role in controlling the chloroplast gene expression. However, also chloroplast-to-nucleus signaling pathways exist that greatly affect the transcription of nuclear genes encoding photosynthesis-related proteins. This kind of 'retrograde' informational flow is triggered by different chloroplast-generated signals, involving, for instance, chlorophyll precursors (porphyrins) or the plastoquinone redox state, and serves not only to correlate the nuclear gene expression with the chloroplast-developmental stage and the organelle functionality but also to achieve a swift photosystem adjustment in response to qualitative and quantitative light changes in the plant growth environment. It is interesting to point out that during the massive transfer of genetic information to the host cell nucleus, the cyanobacterium-like ancestor, which was changing into chloroplast, kept the genes coding for the reaction center proteins in own DNA. The import of this strategy was to make the chloroplast able to adapt more quickly the synthesis of these photosystem key components to the environmental conditions.

Light and Chloroplast Differentiation

In lower vascular plants, as well as in most gymnosperms, the events leading to chloroplast differentiation can occur both in light and in darkness. On the contrary, in the angiosperms, light is a key factor for chloroplast biogenesis and the buildup of the thylakoid system. In these more evolved plants, light, by mediation of phytochromes and cryptochromes (sensing red and blue light, respectively), induces the expression of different nuclear genes (photo-genes) coding for essential chloroplast components. Light, for instance, is required for the transcription of the *RbcS* genes coding for the Rubisco SSUs and for that of the *Cab* genes encoding numerous chlorophyll *a/b*-binding proteins. Moreover, chlorophyll is not produced in the dark because one of the final steps of its biosynthetic pathway, precisely the protochlorophyllide (Pchl_{id}) to chlorophyllide (Chl_{id}) reduction, is carried out by the enzyme Pchl_{id} oxidoreductase (POR) which, in flowering plants, has become light dependent. As a consequence, in dark-grown (etiolated) seedlings of angiosperms, peculiar organelles, named 'etioplasts', originate from proplastids in cells that would become green in the presence of light.

Etioplasts

Etioplasts, which are the chloroplast counterparts in darkness, can be formed in nature during the first phase of plantlet growth, before the emergence from soil. An inner membrane system very different from the thylakoid develops in etioplasts

(Figure 4). Most membranes have a tubular arrangement and aggregate in a tridimensional semi-crystalline network of interconnected tubules, defined 'prolamellar body', from which some lamellar membranes (prothylakoids) extend. The prolamellar body membranes contain the POR associated with Pchl_a and NADPH to form a stable ternary complex, and are particularly rich in MGDG, which favors the tubular arrangement due to its cone-shaped molecular configuration.

Etioplasts exposed to light convert to chloroplasts. The first event of this greening process is the Pchl_a reduction that the photo-activated POR carries out by using the NADPH of the ternary complex. Successively, the etioplast membranes

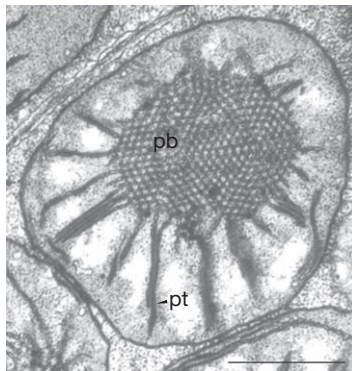


Figure 4 TEM micrograph of an etioplast from a dark-grown bean cotyledon. The inner tubular membranes are arranged in a semi-crystalline prolamellar body (pb) from which lamellar prothylakoids (pt) extend. Scale = 1 μ m.

rearrange to form primary thylakoids and then a massive synthesis *de novo* of chlorophylls and membranes leads to the formation of the well-organized and photosynthetically competent chloroplast.

Plastids

The chloroplast belongs to the organellar class of plastids. In lower photoautotroph organisms, such as algae, which do not possess cell types with very distinct functions, the photosynthetic chloroplast has remained the sole kind of plastid. On the contrary, in terrestrial plants, which evolved highly specialized tissues and organs, different kinds of plastids with specific structures and functions arose from the original chloroplast. In these plants, the class of plastids, besides the green chloroplasts, includes uncolored amyloplasts, whose function is to store large amounts of starch (Figure 5) and yellow-red-brown chromoplasts which synthesize a lot of carotenoids (Figure 5) and have attractive functions. Apart from the lack of thylakoids, these nonphotosynthesizing plastids share common characteristics with the chloroplasts, like the envelope membranes, the prokaryotic DNA and the 70S ribosomes. Furthermore, they maintain several biosynthetic pathways carried out by enzymes inserted in the stroma or bound to the envelope, such as those of fatty acids and terpenoids.

All the plastids can differentiate from proplastids. However, the possibility also exists that a plastid derives from the conversion of another kind of plastid, according to the

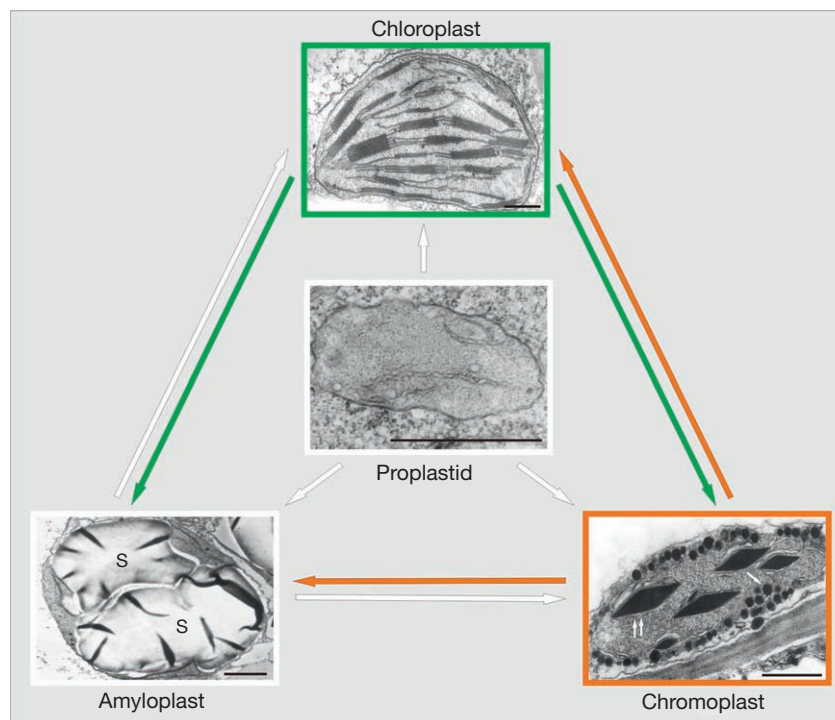


Figure 5 The cyclic model of plastid interconversions. The amyloplast from a cell of *Raphanus* hypocotyl contains large starch grains (S). Globular (arrow) and crystalline (double arrow) carotenoid masses are inserted in the chromoplast from a cell of *Ranunculus* petal. Chloroplast and proplastid are from a mesophyll cell and a basal meristem cell of a maize leaf, respectively. Scale = 1 μ m.

'cyclic model' (Figure 5) proposed already by Schimper at the end of 1800. These reversible plastid interconversions depend on cell-developmental programs as well as on endogenous (hormones and nutrients) or environmental (light and temperature) signals. Only the chromoplasts (named gerontoplasts) derived from chloroplast degeneration during a green tissue senescence cannot undergo further conversion.

See also: [Bioenergetics: Chlorophylls and Carotenoids](#); [Chloroplast Signaling and Redox Control](#); [Structure–Function of the Cytochrome *b₆f* Complex of Oxygenic Photosynthesis](#).

Further Reading

- Cornah J, Terry MJ, and Smith AG (2003) Green or red: What stops the traffic in the tetrapyrrole pathway. *Trends in Plant Science* 5: 224–230.
- Goldschmidt-Clermont M (1998) Coordination of nuclear and chloroplast gene expression in plant cells. *International Review of Cytology* 177: 115–179.
- Jarvis P (2003) Intracellular signalling: The language of the chloroplast. *Current Biology* 13: R314–R316.
- Kleinig H (1989) The role of plastids in isoprenoid biosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology* 40: 39–59.
- Larkin RM and Ruckle ME (2008) Integration of light and plastid signals. *Current Opinion in Plant Biology* 11: 593–599.
- Martin W and Herrmann RG (1998) Gene transfer from organelles to the nucleus: How much, what happens, and why? *Plant Physiology* 118: 9–17.
- McFadden GI (1999) Endosymbiosis and evolution of the plant cell. *Current Opinion in Plant Biology* 2: 513–519.
- Ohlrogge J and Browse J (1995) Lipid biosynthesis. *Plant Cell* 7: 957–970.
- Peltier G and Cournac L (2002) Chlororespiration. *Annual Review of Plant Biology* 53: 523–550.
- Sundqvist C and Dahlin C (1997) With chlorophyll pigments from prolamellar bodies to light-harvesting complexes. *Physiologia Plantarum* 199: 748–759.
- Willows RD (2003) Biosynthesis of chlorophylls from protoporphyrin IX. *Natural Product Reports* 20: 327–341.

Chloroplast Signaling and Redox Control

J-D Rochaix, University of Geneva, Geneva, Switzerland

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by John F. Allen, volume 1, pp. 438–445, © 2004, Elsevier Inc.

Glossary

Calvin–Benson cycle A series of enzymatic reactions in the stromal phase of chloroplasts in which reduced nicotinamide adenine dinucleotide phosphate (NADPH) and adenosine triphosphate (ATP) produced by the light reactions of photosynthesis are consumed to fix CO₂ and form precursors of sugars. Each cycle consumes three ATP and two NADPH molecules and fixes one molecule of CO₂, and the intermediates of the cycle are regenerated.

Glutathione–ascorbate cycle Metabolic pathway that detoxifies hydrogen peroxide (H₂O₂), which is a reactive-oxygen species. In this cycle, ascorbate and glutathione are regenerated and net electron flow is from NADPH to H₂O₂.

Light-harvesting system Antenna system of PSII or PSI consisting of light-harvesting proteins with bound chlorophyll, carotenoid, and lipid molecules, which captures light energy and transfers it to the reaction centers of the two photosystems for the photochemical reactions.

Pasteur effect Shortage of ATP leads to the stimulation of the breakdown of sugars (glycolysis) with

concomitant production of reducing equivalents such as NADH.

Q-cycle Upon oxidation of PQH₂ by the Cytb₆f complex at the Q_o site, two electrons are released. One electron is transferred through a high-potential chain to Cyt_f, whereas the other is transferred through a low-potential chain to two cytochromes within Cytb₆f, and finally to a quinone or semiquinone molecule at the Q_i site. In this way, four protons are pumped into the lumen per two electrons transferred through the Cytb₆f complex.

Quinone Small molecule acting as an electron carrier which can exist in oxidized (PQ) or reduced form (PQH₂).

Retrograde signaling Upon chloroplast dysfunction or plastid-developmental arrest, signals are generated within the chloroplast that are sensed and lead to the activation of a signaling chain from the chloroplast to the nucleus where expression of specific genes is altered in response to the chloroplast signals.

Singlet oxygen A highly reactive form of oxygen formed upon charge recombination within the PSII reaction center or through photosensitization of porphyrins.

Introduction

A remarkable feature of plant and algal cells is the presence of three genetic systems located in the nucleus, chloroplast, and mitochondria, which are autonomous but need to be coordinated for proper functioning and regulation of cellular metabolism. Although the majority of chloroplast proteins is encoded by nuclear genes and the plastid genome has a modest size, the chloroplast can also influence nuclear gene activity under various stress conditions that affect the photosynthetic apparatus. In land plants and algae, the photosynthetic system uses solar energy to convert water and carbon dioxide into sugars with concomitant oxygen evolution. The primary events of photosynthesis occur in the thylakoids, an internal membrane system within chloroplasts consisting of appressed membranes called ‘grana’ and of ‘stromal lamellae’ (Figure 1). The photosynthetic electron transport chain consists of several multicomponent complexes, including proteins and pigments, which are embedded in the thylakoid membranes. At one end of the chain, photosystem II (PSII) uses the light excitation energy absorbed by its light-harvesting system (LHCII) to photo-oxidize a chlorophyll dimer within the PSII reaction center. This event creates a very powerful oxidant capable of splitting water with the concomitant release of protons in the thylakoid lumen and the transfer of electrons to the acceptors of PSII, the quinones Q_A and Q_B, and subsequently to the plastoquinone pool. The electrons are then transferred to

the cytochrome *b₆f* complex (Cytb₆f), which pumps protons into the thylakoid lumen through the Q-cycle and transfers the electrons further to plastocyanin. This soluble protein in the thylakoid lumen acts as an electron carrier between Cytb₆f and PSI. This complex is capable of using light energy captured by its light-harvesting system (LHCI) to induce a charge separation across the membrane, resulting in the oxidation of plastocyanin and the reduction of ferredoxin on the stromal side of PSI and the subsequent reduction of nicotinamide adenine dinucleotide phosphate (NADP⁺) to reduced nicotinamide adenine dinucleotide phosphate (NADPH). This linear photosynthetic electron flow is coupled with proton translocation across the thylakoid membrane and the resulting proton gradient is used by the adenosine triphosphate (ATP) synthase to produce ATP. Both ATP and NADPH are used for driving the Calvin–Benson cycle in the soluble phase of the chloroplast, called stroma, which results in CO₂ fixation and in the formation of sugars.

Although PSII and LHCII are located in the grana, PSI and the ATP synthase are located mainly in the stromal regions and also in the end membranes and margins of the grana. This is due to the fact that both PSI and the ATP synthase contain bulky stromal domains that do not fit between the tightly packed grana stacks.

The different complexes of the photosynthetic electron transport chain have been studied intensively and spectacular advances have been achieved in the structure determination of

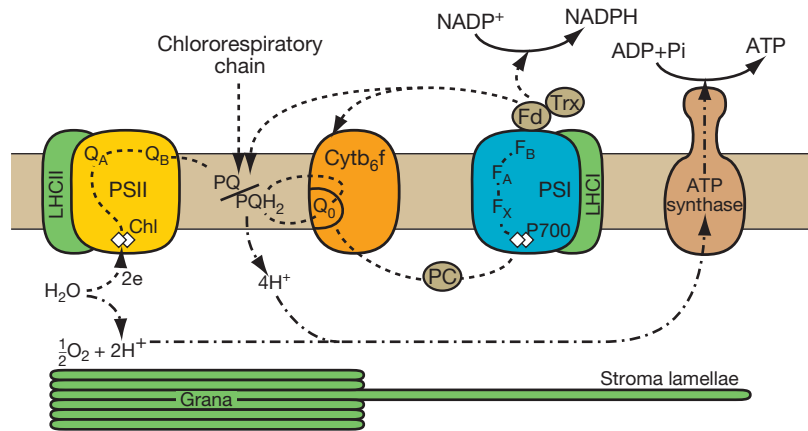


Figure 1 Photosynthetic electron transport chain. Light energy absorbed by the light-harvesting system of photosystem II (LHCII) is directed to the reaction center of PSII and used to induce charge separations across the thylakoid membrane followed by extraction of electrons from water, which are subsequently transferred to the plastoquinone pool and the *Cytb₆f* complex. The electrons are further transferred to plastocyanin and to PSI. The latter functions as a light-driven plastocyanin–ferredoxin oxidoreductase. The photochemical water oxidation releases two protons in the lumen and injects two electrons in the electron transport chain, which are used to reduce one plastoquinone molecule. Oxidation of PQH₂ by the *Cytb₆f* complex and operation of its proton pump leads to the further release of four protons in the lumen and, thus, to a total of six protons per two electrons transferred. As 2 electrons are required for the reduction of NADP⁺ and 14 protons are consumed by the ATP synthase to produce three ATP, this leads to an ATP/NADPH ratio of 9/7. Thus, linear electron flow is insufficient for reaching an ATP/NADPH ratio of 1.5 required for the operation of the Calvin cycle. The remaining ATP can be generated through cyclic electron flow in which reduced ferredoxin transfers electrons back to the plastoquinone pool or directly to the *Cytb₆f* complex. In this mode, only ATP is produced. Reducing equivalents can also be transferred to the plastoquinone pool through the chlororespiratory chain. PSII and LHCII are localized in the grana regions, while PSI and ATP synthase are found mostly in the stroma lamellae.

these complexes. PSII, LHCII, *Cytb₆f*, PSI, and ATP synthase have been crystallized and their atomic structure determined. Although we know a great deal about the individual complexes and their structure–function relationship, less is known on how electron flow is regulated and how the redox state of individual components of the photosynthetic electron transport chain influences signaling events within the chloroplast. Two key components in chloroplast redox signaling are the plastoquinone pool and the ferredoxin/ thioredoxin system.

Electron Flow Pathways in the Chloroplast

According to recent estimates, linear electron flow in plants leads to an ATP:NADPH ratio of 9/7, which is below the value of 1.5 required for the operation of the Calvin–Benson cycle (Figure 1). The additional ATP must be provided by another pathway called cyclic electron flow in which reduced ferredoxin transfers its electrons back to the plastoquinone pool or to *Cytb₆f*. In this pathway, two protons are pumped across the thylakoid membrane for each cycling electron without generating reducing power. Hence, the ATP/NADPH ratio can be finely tuned through adjustment of the relative levels of linear and cyclic electron flow to match changing metabolic demands. Alternatively, in a pathway called ‘pseudocyclic electron flow’, molecular oxygen can act as electron acceptor instead of NADP⁺. In addition to these three pathways, the chlororespiratory chain feeds reducing power from the stromal phase into the plastoquinone pool and can, therefore, also modify the redox state of the plastoquinone pool.

The redox poise of the photosynthetic electron transport chain is important for its proper functioning. Both complete reduction and complete oxidation of this chain would lead to

an arrest of electron flow, because in the first case, there is no place for the electron to move to, except for reacting with oxygen, and in the second case, no electrons are available. Hence, the system attempts to maintain a redox poise which, however, can be profoundly affected by environmental factors such as temperature and the limited availability of CO₂, water, and nutrients. Many of these changes lead to a decrease in the activity of the Calvin–Benson cycle which, in turn, leads to an over-reduction of the electron transport chain and to the generation of reactive-oxygen species (ROSs). These ROSs may cause oxidative damage under extreme conditions but are usually detoxified by an efficient plastid-scavenging system, including superoxide dismutase, ascorbate peroxidase, and the glutathione–ascorbate cycle. Signaling mechanisms that connect photosynthetic electron flow with gene expression include reversible changes in redox poise that occur either at the level of the plastoquinone pool or of the ferredoxin/ thioredoxin system at the acceptor side of PSI.

Redox State of Plastoquinone Pool and Chloroplast Signaling

The plastoquinone pool plays a key role in photosynthetic electron transfer. Its redox state depends on the relative activities of PSII and PSI. Because the light-harvesting systems of these two photosystems have different light-absorption properties, the excitation energy transferred to the PSII and PSI reaction centers varies when the quality of the incident light changes. In turn, these changes alter the redox state of the plastoquinone pool. Thus, excess stimulation of PSII relative to PSI leads to a reduction of the plastoquinone pool, whereas excess stimulation of PSI relative to PSII leads to its oxidation.

Changes in the redox state of the plastoquinone pool act as a read-out of any imbalance in the activities of the two photosystems and are sensed by a signaling system that corrects the imbalance both in the short term and in the long term, and thereby restores an optimal photosynthetic yield especially under low-light conditions.

Short-Term Response: State Transitions

A typical short-term response that occurs within minutes is state transition, a process whereby any imbalance in the activities of PSII and PSI resulting from changes in light or metabolic conditions is compensated by a change in the size of the antennae of PSII and PSI. In this way, the two serially connected photosystems function at the same rate and, thus, ensure an optimal photosynthetic yield, especially under low irradiance (Figure 2). When the activity of PSII exceeds that of PSI, the redox state of the plastoquinone pool is more reduced and binding of plastoquinol, the reduced form of plastoquinone, to the Q_o site of $Cytb_6f$ leads to the activation of a protein kinase that phosphorylates a portion of LHCII. The $Cytb_6f$ complex is essential for this activation as mutants deficient in $Cytb_6f$ are unable to phosphorylate LHCII. This phosphorylation causes the release of the mobile portion of LHCII from PSII and favors its binding to PSI, thereby increasing the cross section of the PSI antenna at the expense of the light-harvesting system of PSII. This state is called state 2. The process is reversible as excess stimulation of PSI relative to PSII leads to the oxidation of the plastoquinone

pool and to the inactivation of the kinase which is followed by dephosphorylation of LHCII and its return to PSII. This state is called state 1 (Figure 2).

In the green unicellular alga *Chlamydomonas*, as much as 80% of LHCII is mobile, whereas in land plants only 15–20% of LHCII is mobile during state transitions. This difference can be correlated with the fact that the major function of state transitions in *Chlamydomonas* is to adjust the ATP level to cellular demands. In this alga, a decrease in the amount of ATP induces a transition from state 1 to state 2. Lower ATP levels induce glycolysis through the Pasteur effect and the reducing power generated is transferred through the chlororespiratory chain to the plastoquinone pool, a process that leads to the reduction of this pool and to the activation of the LHCII kinase. In *Chlamydomonas*, transition from state 1 to state 2 is accompanied by a switch from linear to cyclic electron transport in which most of the PSII is disconnected from the other components of the electron transport chain (Figure 2). Another remarkable feature is that state transitions are associated with a remodeling of thylakoid membranes. Major rearrangements occur in these membranes with breakage of bridges between the grana stacks and fusion events.

Both the kinase and phosphatase involved in state transitions have been identified. The kinase Stt7/STN7 is a thylakoid membrane protein associated with $Cytb_6f$ and specifically required for the phosphorylation of LHCII under state 2 conditions. It is well conserved in all eukaryotic organisms performing oxygenic photosynthesis. It is present in vastly substoichiometric amounts relative to $Cytb_6f$, indicating that it must act in catalytic amounts. In its absence, the photosynthetic apparatus is locked in state 1. The TAP38/PPH1 phosphatase is specifically required for the dephosphorylation of LHCII in land plants and mainly associated with the stromal lamellae of thylakoid membranes. Loss of TAP38/PPH1 leads to a persistent state 2 with phosphorylated LHCII and an increased size of the PSI antenna. Thus, the phosphorylation status of LHCII and the control of state transitions are determined specifically by the kinase phosphatase pair STN7 and TAP38/PPH1.

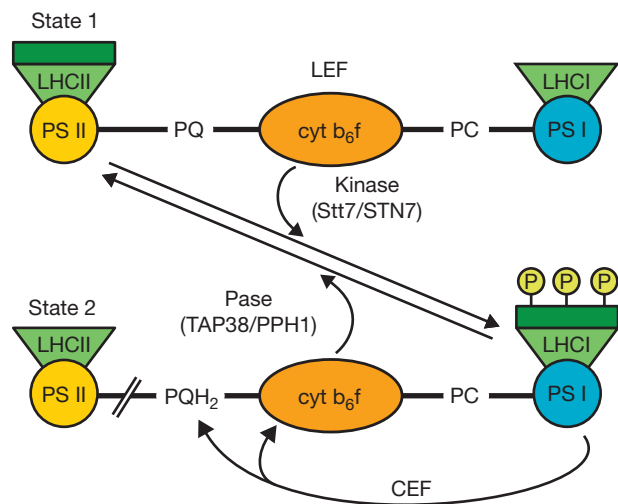


Figure 2 Scheme of state transitions. This process balances the light excitation energy between PSII and PSI upon changes in the spectral composition of light, and thereby ensures an optimal photosynthetic rate. During a state 1 to state 2 transition, excess stimulation of PSII leads to the reduction of the plastoquinone pool that favors docking of plastoquinol (PQH_2) to the Q_o site of $Cytb_6f$. This event activates the LHCII kinase (probably Stt7/STN7), which phosphorylates the mobile portion of LHCII leading to its detachment from PSII and to its binding to PSI (state 2). Preferential stimulation of PSI oxidizes the plastoquinone pool and inactivates the LHCII kinase. Dephosphorylation of LHCII by the phosphatase PPH1/TAP38 is followed by its return to PSII (state 1). In *Chlamydomonas* transition from state 1 to state 2 is accompanied by a switch from linear electron flow (LEF) to cyclic electron flow (CEF), which favors mainly ATP synthesis.

Long-Term Response

Although state transitions represent a short-term response, changes in the redox state of the plastoquinone pool also trigger a long-term response, which occurs in the range of hours or days. Continued stimulation of PSII relative to PSI will in the long term lead to a compensatory response with an increase in the level of PSI to restore the balance between the two photosystems. This long-term response can also be followed by an increase in the ratio between chlorophyll *a* (Chl *a*) and Chl *b* due to the fact that PSI contains less Chl *b* than PSII. Because the long-term response also involves nuclear genes besides chloroplast genes, this implies that the changes in the redox state of the plastoquinone pool are sensed by the nucleus through a retrograde signaling chain. The components of this chain are still largely unknown except for the Stt7/STN7 kinase which is involved both in the long- and short-term responses. Mutants lacking STN7 are unable to adjust the photosystem stoichiometry upon changes in light conditions. The long- and short-term signaling pathways appear to diverge at, or immediately downstream of STN7. Comparison between different

organisms reveals that the redox state of the plastoquinone pool regulates expression of the nuclear LHCII genes either at the transcriptional or posttranscriptional level.

Chloroplast Kinase Network

Besides the Stt7/STN7 kinase, other chloroplast protein kinases have been identified. The Stt7/STN7 kinase is structurally related to Stt1/STN8, another thylakoid protein kinase that is specifically required for the phosphorylation of the PSII core proteins D1, D2, and CP43. Both kinases are phosphorylated under state two conditions and the phosphorylation of Stt1 is dependent on Stt7, suggesting the existence of a chloroplast protein kinase cascade. Mutants of *Arabidopsis* deficient in both STN7 and STN8 are nearly fully deficient in thylakoid protein phosphorylation and display a significant decrease in fitness when grown under changing light conditions or in field tests (Figure 3).

Phosphorylation of PSII and LHCII proteins has been proposed to play a role in the folding of thylakoid membranes through charge repulsion. The availability of mutants of *Arabidopsis* lacking either STN7 or STN8 has provided new tools for examining this question. It appears that loss of PSII core protein phosphorylation leads to marked changes in membrane folding with a significant increase in grana size. By contrast, loss of LHCII phosphorylation does not cause any detectable changes in thylakoid membrane organization. Thus, PSII core protein phosphorylation appears to be more important for shaping the grana than LHCII phosphorylation.

The TAK1 protein kinase of *Arabidopsis*, which forms a small family together with two other similar kinases, has also been proposed to be involved in state transitions based on the observation that lines deficient in TAK1 show diminished LHCII phosphorylation and a partial deficiency in state transitions. This kinase appears to be associated with *Cytb₆f* and PSII but its precise role is still unknown.

In view of the endosymbiotic origin of chloroplasts, one might expect to find conserved bacterial signaling components in plastids. Two-component signal transduction is widespread in prokaryotes and includes a sensor kinase and a response regulator that sense specific environmental cues and trigger appropriate responses mostly at the transcriptional level. A sensor-like kinase-dubbed chloroplast kinase network (CSK) was identified in *Arabidopsis* based on its sequence similarity to corresponding bacterial proteins. Chloroplast transcript accumulation was changed in CSK knockout lines indicating that CSK affects plastid gene expression either directly or indirectly. Moreover, the typical increase in the chl *a*/chl *b* ratio, which occurs under prolonged stimulation of PSII, was abolished in the *csk* mutants. Whether CSK acts as a redox regulatory coupling system between photosynthetic electron transport and chloroplast gene expression remains to be proven.

Another protein kinase implicated in redox-mediated regulation of plastid gene expression is the casein kinase CK2 α that is associated with the plastid RNA polymerase and known to phosphorylate proteins involved in transcriptional and post-transcriptional regulation. Interestingly, the sequence of one phosphorylation site of STN7 conforms to the CK2 α substrate consensus raising the possibility that CK2 α could be involved in the phosphorylation of LHCII through the STN7 kinase. The picture that emerges from these studies is a complex chloroplast kinase and phosphatase signaling network which is involved in both short- and long-term acclimation in response to changing environmental conditions (Figure 3).

Ferredoxin/Thioredoxin and Chloroplast Signaling

Besides the redox state of the plastoquinone pool, that of the ferredoxin/thioredoxin system comprised of ferredoxin, thioredoxin and ferredoxin–thioredoxin reductase also plays a central role in chloroplast signaling. This system is involved in an

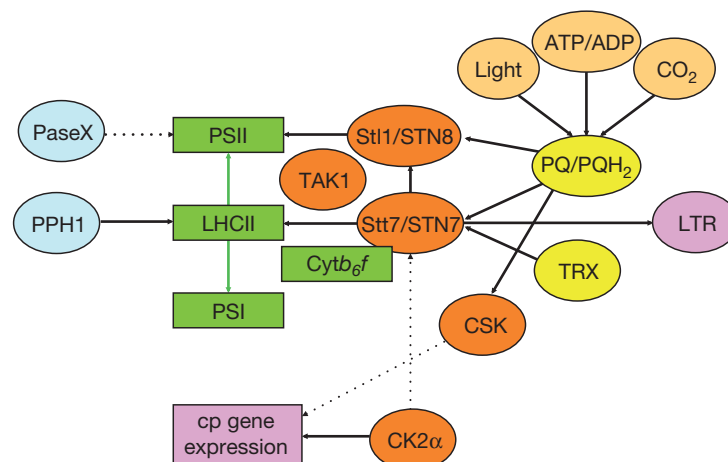


Figure 3 Thylakoid protein kinase network. The plastoquinone redox state depends on light quality and quantity, ATP, and CO₂ levels, and modulates the activity of the Stt7/STN7 and Stt1/STN8 kinases required for the phosphorylation of LHCII and PSII, respectively. It may also act on CSK. Activation of the Stt7/STN7 kinase requires a close interaction with the Cytb₆f complex. Phosphorylated LHCII moves from PSII to PSI. Dephosphorylation is mediated by the phosphatase TAP38/ PPH1. The redox state of Trx also influences the activity of Stt7/STN7. The proposed action of CK2 α on Stt7/STN7 and of CSK on chloroplast gene expression is still tentative and will need further experimental support. Besides its role in short-term acclimation, STN7 is also involved in the long-term response (LTR). TAK1 may also contribute to state transitions but its exact role is not known. Protein kinases are highlighted in orange.

extensive SH-group redox regulation and is known to link light to the activity of enzymes in the chloroplast stroma. In their light-induced reduced state, thioredoxins are implicated through disulfide bond reduction in the activation of numerous enzymes functional in biosynthesis. In other cases, the reduction of disulfide bonds leads to the inactivation of enzymes involved in carbohydrate degradation and their activation in the dark occurs through oxidation of the thiol groups. This also appears to hold for lumen proteins that are oxidatively activated in the light. In this respect, the case of Stt7/STN7 is particularly interesting. This kinase, known to be inactivated under high light, is a transmembrane protein with its catalytic domain on the stromal side and the N-terminal end on the lumen side, with two conserved Cys that are critical for the activity of the kinase and likely candidates for targets of thioredoxin. However, thioredoxins are localized on the stromal side of the thylakoid membrane, whereas the target Cys residues are on the lumen side. One possibility is that the redox signal from thioredoxin is shuttled through the membrane by the CcdA and Hcf164 proteins known to transduce thiol-reducing equivalents across the thylakoid membrane.

The reversible changes in redox state of the ferredoxin/thioredoxin system operate at several levels of gene expression, including both transcriptional and posttranscriptional steps. In *Chlamydomonas*, a redox-responsive multiprotein complex binding to the 5'-untranslated region of chloroplast messenger RNA plays a role in the light-induced activation of translation initiation. Other steps in chloroplast gene expression such as RNA degradation, RNA splicing, and translation elongation also appear to be redox modulated.

Retrograde Signaling Pathways

Gun Mutants

Plastid-to-nucleus retrograde signaling is triggered by plastid dysfunction or plastid-developmental arrest and refers to signaling initiated in the chloroplast, which is sensed by the nucleus and leads to altered gene expression. Thousands of nuclear genes in *Arabidopsis* can be affected through these pathways. Several different retrograde signaling chains have been identified which are triggered by inhibition of chloroplast gene expression, perturbations in the tetrapyrrole pathway, and changes in the redox state of the photosynthetic electron transport chain or stress conditions that lead to the formation of ROS. In all these cases, expression of the LHCII gene and of other nuclear genes involved in photosynthesis is decreased during these perturbations. A genetic screen for mutants in which LHCII expression is partially restored has yielded *gun* (genome uncoupled) mutants. Most of these mutants are deficient in components involved in tetrapyrrole biosynthesis, suggesting that tetrapyrroles or their regulators are involved in retrograde signaling. These different retrograde signaling chains appear to be integrated upstream of GUN1, a chloroplast protein associated with the chloroplast nucleoid which comprises the plastid DNA and its associated proteins. The components downstream of GUN1 are still unknown except for the ABI4

transcription factor which has been proposed to mediate repression of nuclear genes upon activation of the retrograde signaling chain.

ROS Signaling: The Case of Singlet Oxygen

ROSs are generated as an inevitable consequence of the multiple photosynthetic electron transfer reactions. They can cause oxidative damage and elicit stress responses. Several chemically distinct ROSs are produced, including hydrogen peroxide and superoxide anion radical generated as a result of the reduction of oxygen at PSI and singlet oxygen produced in PSII reaction centers. Because these ROSs are usually produced simultaneously, it has been difficult to determine which ROS is involved in a specific stress response. This problem could be circumvented through the use of the *flu* mutant of *Arabidopsis*. This mutant accumulates protochlorophyllide, a precursor of chlorophyll in the dark, which is a potent photosensitizer. Upon illumination, this compound generates singlet oxygen. After a transfer from the dark to the light, growth of mature *flu* mutant plants is arrested while *flu* seedlings bleach and die. At first sight, this suggests that plants are damaged due to the toxic levels of singlet oxygen. However, further genetic studies revealed that inactivation of the *Executer 1* gene is sufficient for abolishing the stress response caused by the release of singlet oxygen. Thus, it is not the ROS generated during the stress response which causes damage. Instead, a genetically determined stress response program is activated through the *Executer 1* protein after release and perception of singlet oxygen by the plant that subsequently leads to seedling lethality and growth inhibition of mature plants.

See also: **Bioenergetics:** Chemiosmotic Theory; Chloroplasts; Dynamic Behavior of Photosystem II Light Harvesting; Light-Harvesting Complex I and II: Pigments and Proteins; Photosystem I; Photosystem II: Redox and Protein Components; Structure-Function of the Cytochrome *b₆f* Complex of Oxygenic Photosynthesis; **Metabolism** **Vitamins and Hormones:** Photosynthesis; Photosynthetic Carbon Dioxide Fixation; **Signaling:** Serine/Threonine Phosphatases.

Further Reading

- Allen JF (2003) Cyclic, pseudocyclic and noncyclic photophosphorylation: New links in the chain. *Trends in Plant Sciences* 8: 15–19.
- Buchanan BB and Luan S (2005) Redox regulation in the chloroplast thylakoid lumen: A new frontier in photosynthesis research. *Journal of Experimental Botany* 56: 1439–1447.
- Eberhard S, Finazzi G, and Wollman FA (2008) The dynamics of photosynthesis. *Annual Review of Genetics* 42: 463–515.
- Laloi C, Przybyla D, and Apel K (2006) A genetic approach towards elucidating the biological activity of different reactive oxygen species in *Arabidopsis thaliana*. *Journal of Experimental Botany* 57: 1719–1724.
- Lemeille S and Rochaix JD (2010) State transitions at the crossroad of thylakoid signalling pathways. *Photosynthesis Research* 106: 33–46.
- Pfannschmidt T (2003) Chloroplast redox signals: How photosynthesis controls its own genes. *Trends in Plant Sciences* 8: 33–41.
- Pogson BJ, Woo NS, Förster B, and Small ID (2008) Plastid signalling to the nucleus and beyond. *Trends in Plant Science* 13: 602–609.

Cholesterol Synthesis and Regulation

P J Espenshade, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Peter A. Edwards, volume 1, pp. 451–455, © 2004, Elsevier Inc.

Glossary

HMG-CoA reductase 3-Hydroxy-3-methylglutaryl coenzyme A reductase, a rate-limiting enzyme of cholesterol biosynthesis.

Scap SREBP cleavage activating protein. Scap is required for the transport of SREBP from the endoplasmic reticulum to the Golgi.

SREBP Sterol regulatory element binding protein, a transcription factor that is proteolytically processed before entering the nucleus and binding to a *cis* element on the DNA, termed sterol regulatory element.

Statins A class of drugs that inhibit the activity of HMG-CoA reductase.

Properties of Cholesterol

Cholesterol is a cyclic hydrocarbon that can be esterified with a fatty acid to form a cholesteryl ester. Both cholesterol and cholesteryl esters are lipids and are essentially insoluble in aqueous solution but soluble in organic solvents. Excess cholesterol esters are stored as lipid droplets within the cytosol. Such droplets are prevalent in steroidogenic tissues where they serve as precursors of the steroid hormones. In eukaryotic cells, cholesterol is solubilized as a result of its interaction either with membrane phospholipids or with phospholipids and bile acid micelles in the gall bladder. In blood, cholesterol and cholesterol esters are solubilized within lipoprotein complexes. These lipoproteins, which include low-density lipoprotein (LDL), very low density lipoprotein (VLDL), chylomicrons, and high-density lipoprotein (HDL), function to transport insoluble lipids around the body.

Functions of Cholesterol

Membrane Structure

Cholesterol is absent from prokaryotic cells. However, cholesterol plays an essential structural role in maintaining the fluidity of eukaryotic cell membranes. Cholesterol is not equally distributed in all membranes; the membranes of mitochondria, peroxisomes, and endoplasmic reticulum are cholesterol poor, whereas the plasma membrane is enriched in the sterol. However, the concentration of cholesterol varies significantly even within the plasma membrane; it is highly enriched in two specialized areas termed lipid rafts and caveolae. Since many receptors are localized to these cholesterol- and sphingomyelin-rich domains, it has been suggested that lipid rafts and caveolae function as signaling gateways into the cell.

Precursor of Signaling Molecules

As shown in [Figure 1](#), cholesterol is a precursor of bile acids, steroid hormones, and oxysterols (oxidized cholesterol). These cholesterol metabolites function to activate specific nuclear receptors that control many metabolic and developmental processes. In addition, intermediates in the cholesterol

biosynthetic pathway are themselves precursors for other critical pathways. For example, 7-dehydrocholesterol is a precursor of vitamin D, and farnesyl diphosphate is a precursor of geranylgeranyl diphosphate, ubiquinone, Heme a, and dolichols ([Figure 1](#)). Although beyond the scope of this article, it is now clear that modification (prenylation) of many proteins by either farnesyl diphosphate or geranylgeranyl diphosphate is critical for cell survival. Thus, hypolipidemic drugs (e.g., statins) that inhibit HMG-CoA reductase activity must be used at doses that result in partial, not total, inhibition of the pathway.

Sources of Cellular Cholesterol

Diet and Lipoproteins

Insects are cholesterol auxotrophs and must obtain cholesterol from the diet. In contrast, other eukaryotes obtain cholesterol from two sources: the diet and endogenous synthesis. Mammals absorb 40–50% of the cholesterol in their diet and transport it to the liver and other tissues in lipoproteins. The plasma lipoproteins bind to cell-surface receptors prior to the delivery of the lipid cargo to the cell. Such receptors include the LDL receptor, SR-B1, SR-A, and CD36. The LDL receptor is expressed on the surface of most cells and plays a particularly important role in cholesterol homeostasis. This transmembrane protein binds cholesterol-rich LDL and facilitates the delivery of cholesterol into the cell via specialized areas on the cell surface called clathrin-coated pits. However, virtually all mammalian cells also synthesize cholesterol ([Figure 1](#)).

Cholesterol Synthesis

There are many reviews on the biosynthesis of cholesterol, possibly in part because work in this area has been extensive and has led to numerous Nobel prizes. All mammalian cells, with the exception of mature red blood cells, utilize more than 30 enzyme reactions and 11 molecules of dioxygen to convert the 2 carbon acetate to the 27 carbon cholesterol via isoprenoid intermediates ([Figure 1](#)). The rate-limiting enzyme in this pathway, HMG-CoA reductase, is the target for many hypolipidemic drugs ([Figure 1](#)). Together, the cholesterol biosynthetic pathway and the LDL receptor pathway provide sufficient

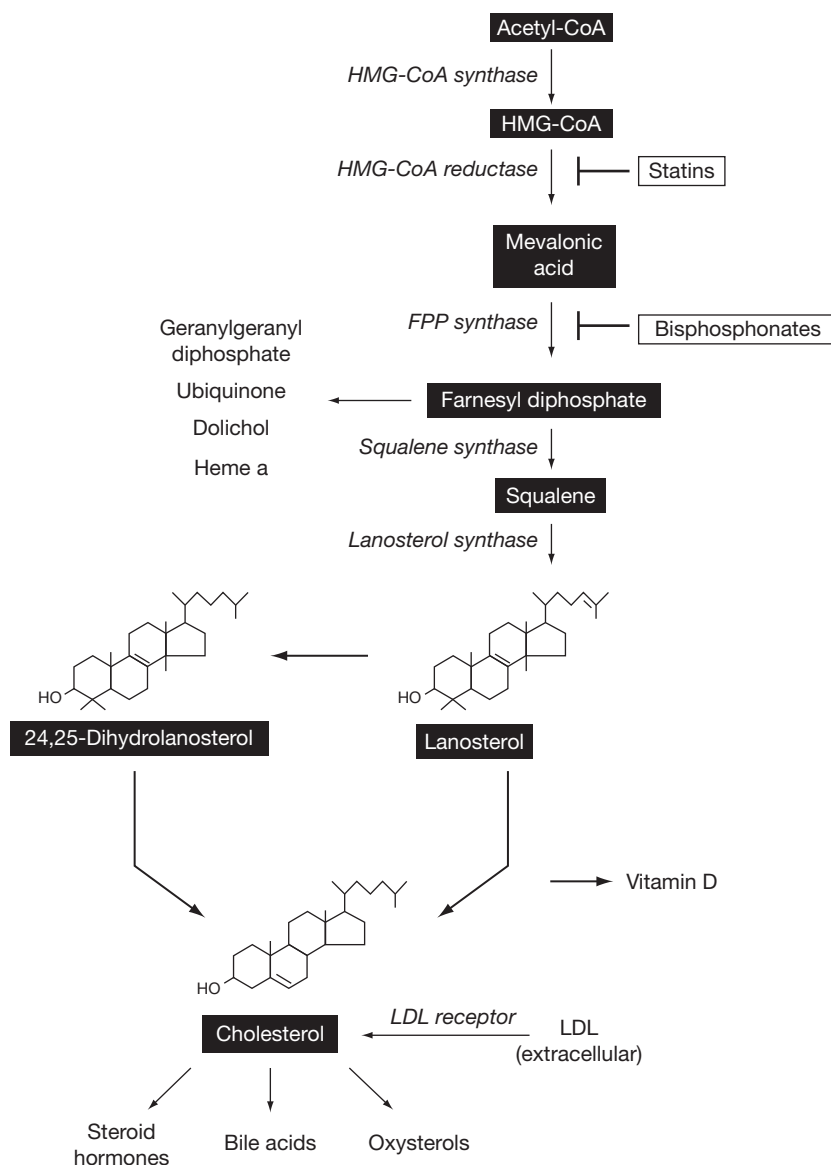


Figure 1 The cholesterol biosynthetic pathway. Selected intermediates (black boxes) and enzymes of the cholesterol synthesis pathway are shown for simplicity. SREBP-2 regulates transcription of genes throughout the pathway, and a few are shown (*italics*). Statins and bisphosphonates inhibit HMG-CoA reductase and farnesyl diphosphate synthase, respectively. Isoprenoids and sterols are precursors for important products, such as ubiquinone, vitamin D, steroids, and bile acids. Modified from Edwards P (2004) In: Lennarz WJ and Lane MD (eds.) *Encyclopedia of Biological Chemistry*, 1st edn., pp. 451–455. Amsterdam: Elsevier.

cholesterol necessary for cell growth and division. These two pathways are coordinately controlled by an exquisite mechanism that senses cellular cholesterol levels. The mechanism involves the generation of three functionally active transcription factors termed sterol regulatory element binding proteins (SREBPs).

Transcriptional control of cholesterol synthesis

There are two SREBP genes, SREBP-1 and SREBP-2. The SREBP-2 gene produces one protein that preferentially activates genes involved in cholesterol homeostasis. In contrast, two proteins, SREBP-1a and SREBP-1c, are produced from the SREBP-1 gene. SREBP-1c preferentially activates genes involved in fatty acid

synthesis. SREBP-1a activates both pathways but is expressed at relatively low levels. Newly synthesized SREBP proteins contain two transmembrane domains that form a hairpin-like structure and anchor the proteins in the endoplasmic reticulum, outside the nucleus (**Figure 2**). If these proteins are anchored to membranes outside the nucleus, how can they function as transcription factors in the nucleus?

The answer came from a series of elegant studies initiated in the laboratory of Goldstein and Brown in Dallas. They showed that generation of nuclear-localized SREBPs requires additional proteins named Insig, Scap, the site 1 protease (S1P), and the site 2 protease (S2P) (**Figure 2**). The roles of these proteins vary; Scap functions both as a cholesterol

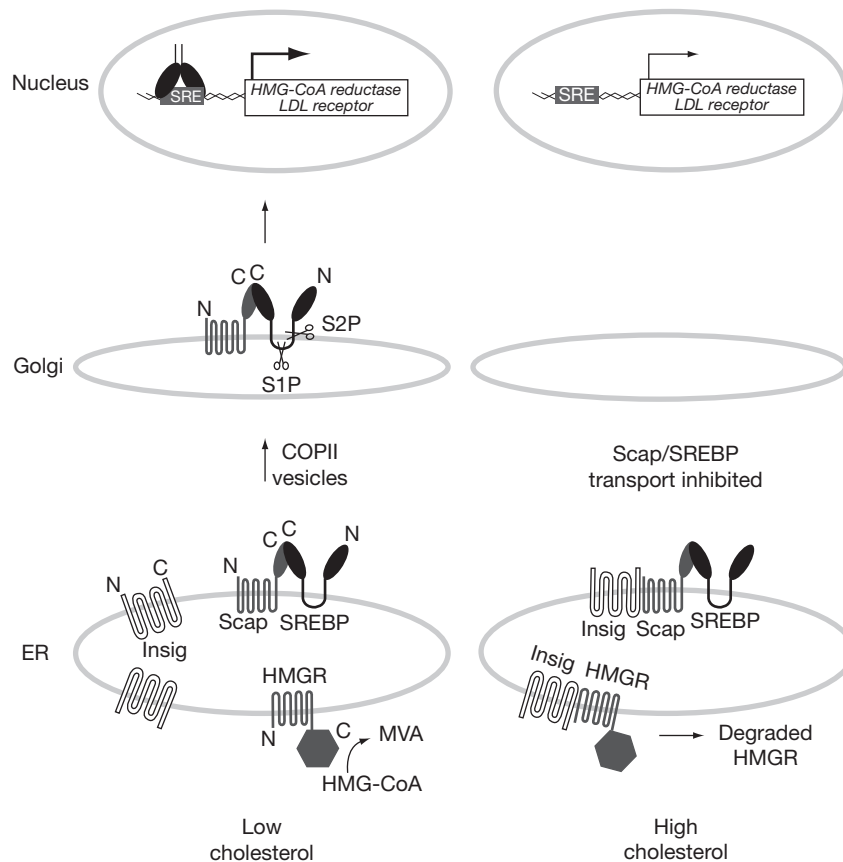


Figure 2 Mechanisms involved in the nuclear targeting of SREBP and the altered stability of HMG-CoA reductase. The left side represents conditions where the cholesterol concentration in the cell is low. Under these conditions, SREBP is transported from the endoplasmic reticulum (ER) to the Golgi, where SREBP is cleaved by S1P and S2P. Cleaved SREBP then travels to the nucleus where it binds DNA as a dimer and activates many target genes. For simplicity, only the LDL receptor and HMG-CoA reductase genes are shown. HMG-CoA reductase (HMGCR) protein also accumulates in the ER where it catalyzes the reduction of HMG-CoA to mevalonic acid (MVA). The right side represents conditions where cellular cholesterol levels are higher than normal. Under these cholesterol-rich conditions, Insig binds to Scap/SREBP and inhibits transport to the Golgi and nuclear localization of SREBP. In addition, the rate of degradation of HMG-CoA reductase is increased. Degradation is preceded by the formation of an Insig/HMG-CoA reductase complex. Modified from Edwards P (2004) In: Lennarz WJ and Lane MD (eds.) *Encyclopedia of Biological Chemistry*, 1st edn. Amsterdam: Elsevier.

sensor and as an escort protein to transport SREBP to the Golgi where two proteases cleave SREBPs to release the soluble amino-terminal fragment. The process is also regulated by Insig, which functions to control the exit of Scap/SREBP from the endoplasmic reticulum (Figure 2). In brief, when cells become cholesterol-poor, for example, after addition of a drug that inhibits cholesterol synthesis, the Scap/SREBP-2 heterodimeric complex is transported in coat protein complex II (COPII) vesicles that bud from the endoplasmic reticulum and then fuse with the Golgi (Figure 2). The Golgi contains two resident proteases, S1P and S2P, that sequentially cleave the newly arrived SREBP-2 protein; first S1P cleaves SREBP into two parts by cleaving the hairpin loop in the lumen of the Golgi, then S2P cleaves the protein at a second site within the transmembrane segment to release a soluble amino-terminal domain (Figure 2). This soluble amino-terminal fragment, containing a transcriptional activation domain and a basic helix-loop-helix-leucine zipper DNA-binding domain, migrates to the nucleus where it binds to specific DNA sequences termed sterol response elements (SREs) and activates transcription. SREs have been

identified in the promoters of many genes involved in cholesterol and fatty acid and lipid synthesis. Such genes include HMG-CoA synthase, HMG-CoA reductase, farnesyl diphosphate synthase, squalene synthase, lanosterol synthase, and the LDL receptor (Figure 1). In summary, a transient decrease in cellular cholesterol levels leads to increased levels of transcriptionally active SREBP-2 in the nucleus and activation of genes involved in cholesterol synthesis and uptake of LDL. The net result is a rapid return of the cellular sterol levels to normal.

In contrast, when cells accumulate cholesterol, Scap binds cholesterol and undergoes a conformational change that induces binding of the Scap/SREBP heterodimer complexes to Insig. Since Insig is a resident protein of the endoplasmic reticulum, the Insig/Scap/SREBP-2 complex remains localized to this membrane (Figure 2). Consequently, transport of SREBP-2 to the Golgi and subsequent cleavage and nuclear localization of the protein is blocked. As a result, transcription of genes coding for the LDL receptor and cholesterol synthesis enzymes is low. Such changes in gene expression will prevent or reduce subsequent cholesterol accumulation in the cell.

SREBP-2 proteolytic activation is also suppressed in response to elevated levels of oxysterols, such as 25-hydroxycholesterol. Oxysterols bind to Insig and stimulate Insig to bind Scap, thus retaining the Scap/SREBP-2 complex in the endoplasmic reticulum.

Regulated degradation of HMG-CoA reductase

HMG-CoA reductase catalyzes the nicotinamide adenine dinucleotide phosphate (NADPH)-dependent reduction of HMG-CoA to mevalonic acid (MVA in [Figure 2](#)). It is considered to be the rate-limiting enzyme of the cholesterol biosynthetic pathway. Thus, changes in the activity of the enzyme are paralleled by changes in cholesterol synthesis. The activity of the enzyme is regulated by changes in transcription, translation (mechanism unknown), and protein stability.

HMG-CoA reductase contains two domains: an eight-transmembrane spanning region that localizes the protein to the endoplasmic reticulum and a carboxy-terminal domain that projects into the cytosol and contains the catalytic activity ([Figure 2](#)). Degradation of HMG-CoA reductase is regulated, and the half-life (a measure of protein stability) varies at least 10-fold. When cellular cholesterol levels are low, the enzyme is relatively stable (half-life ~ 10 h). Thus, active HMG-CoA reductase enzyme accumulates in the endoplasmic reticulum in response to (1) a slow rate of degradation of the protein and (2) increased transcription and enzyme synthesis ([Figure 2](#)). Since HMG-CoA reductase is the rate-limiting enzyme, the net result is increased cholesterol synthesis.

However, when cellular cholesterol levels increase, the rate of degradation of HMG-CoA reductase protein is enhanced ≥ 10 -fold (half-life < 45 min). HMG-CoA reductase degradation is accelerated when cells accumulate the sterol intermediate 24,25-dihydrolanosterol ([Figure 1](#)) and protein turnover is further stimulated by a nonsterol isoprenoid derived from farnesyl diphosphate. Elevated 24,25-dihydrolanosterol induces binding of HMG-CoA reductase to Insig, which forms a complex with the ubiquitin E3 ligase gp78. Subsequent ubiquitinylation of HMG-CoA reductase by gp78 results in its degradation by the proteasome through a mechanism known as endoplasmic reticulum-associated degradation (ERAD; [Figure 2](#)).

Thus, Insig has a dual role in cholesterol-enriched cells; it interacts with HMG-CoA reductase to promote degradation of the enzyme, and with Scap to retain SREBP-2 in the endoplasmic reticulum ([Figure 2](#)). These interactions depend on the partially conserved sterol-sensing domains identified in both HMG-CoA reductase and Scap. In summary, cholesterol-loaded cells have low levels of HMG-CoA reductase protein (and hence low rates of cholesterol synthesis) because (1) the enzyme is rapidly degraded and (2) transcription of the gene is low as a result of the retention of SREBP-2 in the endoplasmic reticulum ([Figure 2](#), right side). Such changes will minimize subsequent cholesterol accumulation in the cell.

Effect of Mutations in the Cholesterol Synthetic Pathway

Mutations in the cholesterol synthetic pathway are rare because they are usually incompatible with life. Mutations in genes encoding mevalonic acid kinase (an early enzyme in the

pathway) and 7-dehydrocholesterol (a late enzyme in the pathway) result in mevalonic acid urea and Smith-Lemli-Opitz syndrome, respectively. It should be emphasized that patients with mevalonic acid urea have residual mevalonic acid kinase activity, sufficient to provide isoprenoid intermediates that are necessary for cell survival. Indeed, a complete blockage of early steps in cholesterol biosynthesis leads to embryonic death. Patients with mutations in genes that result in impaired cholesterol synthesis exhibit various neurological dysfunctions. Such observations emphasize the importance of cholesterol to the neurological system.

Inhibitors of Cholesterol Synthesis: Treatment for Hyperlipidemia and Osteoporosis

Statins

Elevated LDL cholesterol levels (hypercholesterolemia) are relatively common in the West. Hypercholesterolemia is associated with the development of atherosclerosis, coronary heart disease, myocardial infarction, and increased mortality. One class of drugs, referred to collectively as statins, has proved to be extremely effective in lowering plasma LDL levels and in reducing the progression of atherosclerosis and clinically associated problems.

Mevinolin is a fungal metabolite that is structurally related to HMG-CoA. It was originally discovered in 1976 by Akira Endo in Japan and shown to be a potent competitive inhibitor of HMG-CoA reductase; it binds to the enzyme with a $K_i \sim 10^{-9}$ M, compared to a $K_m \sim 10^{-6}$ M for the natural substrate HMG-CoA. Subsequent studies by many investigators led to the identification of a number of natural and synthetic compounds that also inhibit the enzymatic activity of HMG-CoA reductase. These drugs are collectively known as statins. When taken orally, these drugs are absorbed and targeted to the liver where they bind and partially inhibit HMG-CoA reductase and cholesterol synthesis. The liver cells sense this decrease in newly synthesized cholesterol and respond by enhancing the cleavage and nuclear localization of SREBP-2 ([Figure 2](#)). As a result, expression of a number of genes, including the LDL receptor and HMG-CoA reductase, is increased. The increased expression of the LDL receptor protein on the hepatocyte cell surface results in increased clearance of cholesterol-rich LDL from the plasma into the liver. This is the molecular basis for the clinical effects of the drug. Fortunately, the liver does not become loaded with lipid because it converts the LDL-derived cholesterol to bile acids and excretes the latter in the bile. In summary, oral administration of statins results in an increased rate of removal of LDL from the blood and a decline in blood LDL cholesterol levels. Such a change has been shown to reduce both the progression of coronary atherosclerosis and the incidence of myocardial infarction and death.

Bisphosphonates

Bisphosphonates are potent inhibitors of bone resorption and are used widely to treat osteoporosis. After absorption, these drugs are targeted to bone, where they are taken up by osteoclasts, the bone-resorbing cells. The nitrogen-containing class of bisphosphonates inhibits farnesyl diphosphate synthase and

isoprenoid synthesis (Figure 1). As a result, the osteoclasts undergo apoptosis and bone resorption is attenuated.

See also: **Cell Architecture and Function:** Endoplasmic Reticulum-Associated Protein Degradation; **Metabolism Vitamins and Hormones:** Bile Salts; Fatty Acid Oxidation; Fatty Acid Structure and Synthesis; **Protein/Enzyme Structure Function and Degradation:** Regulated Intramembrane Proteolysis.

Further Reading

- Goldstein JL, DeBose-Boyd RA, and Brown MS (2006) Protein sensors for membrane sterols. *Cell* 124: 35–46.
- Horton JD, Goldstein JL, and Brown MS (2002) SREBPs: Activators of the complete program of cholesterol and fatty acid synthesis in the liver. *Journal of Clinical Investigation* 109: 1125–1131.
- Jo Y and DeBose-Boyd RA (2010) Control of cholesterol synthesis through regulated ER-associated degradation of HMG-CoA reductase. *Critical Reviews in Biochemistry and Molecular Biology* 45: 185–198.
- Osborne TF and Espenshade PJ (2009) Evolutionary conservation and adaptation in the mechanism that regulates SREBP action: What a long, strange tRIP it's been. *Genes and Development* 23: 2578–2591.
- Porter FD (2002) Malformation syndromes due to inborn errors of cholesterol synthesis. *Journal of Clinical Investigation* 110: 715–724.
- Tobert JA (2003) Lovastatin and beyond: The history of the HMG-CoA reductase inhibitors. *Nature Reviews Drug Discovery* 2: 517–526.

Chromatin: Methyl-CpG-Binding Proteins

D G Skalnik, Indiana University, Indianapolis, IN, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

CpG island Region of DNA containing a cluster of unmethylated CpG dinucleotides, often found near promoters of actively expressed genes.

Epigenetics Heritable patterns of gene expression that occur in the absence of altered DNA sequence.

Euchromatin Open form of chromatin associated with transcriptionally active genes.

Genomic imprinting Phenomenon whereby maternally derived and paternally derived alleles of a gene exhibit distinct expression patterns as a result of distinct epigenetic modifications established during gametogenesis.

Heterochromatin Highly compact form of transcriptionally inactive chromatin.

The development of multicellular organisms requires the generation of a variety of terminally differentiated cell types from a common set of totipotent stem cells. Integral to this process is the phenomenon of cell memory (e.g., a liver cell gives rise to progeny liver cells upon cell division, rather than kidney or spleen cells). At the molecular level, cell memory derives from distinct patterns of gene expression that are correlated with each type of cell lineage. However, nearly all somatic cells in a multicellular organism contain the same genome. The term epigenetics refers to heritable patterns of gene expression that occur in the absence of altered DNA sequence, and represents the underlying molecular mechanism for cell memory. The CpG dinucleotide represents an important regulatory component of mammalian genomes. The cytosine of this dinucleotide serves as the target of DNA methyltransferases, which catalyze the formation of 5-methylcytosine, a critical epigenetic modification of DNA. Methylated DNA is associated with transcriptionally inactive genes and is generally tightly packaged as heterochromatin, while actively expressed genes are generally hypomethylated and present in a more open euchromatin configuration. Methyl-CpG-binding proteins contribute to both the formation of heterochromatin and the repression of transcription.

Epigenetic Regulation of Gene Expression

The DNA of higher eukaryotes is found tightly associated with proteins *in vivo*. In fact, at least 50% of the mass of a chromosome is contributed by proteins. Although originally thought to provide primarily a packaging function, it is now appreciated that chromatin structure is highly dynamic and plays a critical function in the epigenetic regulation of gene expression. It is necessary to understand the nature of chromatin structure to appreciate the function of methyl-CpG-binding proteins.

Nucleosomes

The primary level of chromatin structure is the nucleosome, in which a DNA double helix is wrapped twice around a core

octamer of eight histone protein molecules (two copies each of histone H2A, H2B, H3, and H4). Nucleosomes are separated by short linkers of DNA, and 200 base pairs of DNA are present in each nucleosome repeat. Chromatin in this configuration is referred to as a 'bead-on-a-string' structure, or euchromatin, and can be transcriptionally active. Additional packaging of nucleosomes into a 30-nm helical array is found in heterochromatin, which is transcriptionally inactive, and is also tightly associated with another species of histone (H1).

Histone Code

Determination of the structure of the nucleosome revealed that the amino terminus of each histone molecule extends away from the octamer core. These histone tails are sites of covalent modification, such as phosphorylation, acetylation, and methylation. At least three-dozen distinct sites of covalent modifications within histone tails have been identified, and many of these are associated specifically with either euchromatin or heterochromatin (Figure 1). The sum of histone modifications found in chromatin is referred to as the histone code, and provides important regulatory information with respect to chromatin structure and ultimately gene expression. For example, acetylated histones are found preferentially in euchromatin. The net acetylation state of a specific region of chromatin is the result of balance between the action of histone acetyltransferases (HATs) and histone deacetylases (HDACs). Similarly, methylated lysine 4 of histone H3 is found in euchromatin, while heterochromatin preferentially contains histone H3 methylated at the lysine 9 position. Lysine residues in histone proteins can be mono-, di-, or trimethylated. Posttranslationally modified histone molecules serve as binding sites for effector proteins implicated in the control of chromatin structure. For example, methylated lysine 9 of histone H3 serves as a binding site for the heterochromatin-associated protein HP1, which recruits a histone methyltransferase to the chromatin. This leads to the methylation of adjacent histone H3-lysine 9 residues, formation of additional HP1-binding sites, and spreading of the heterochromatin configuration.

Histone H2A	Histone H2B	Histone H3	Histone H4
Ser 1, P	Lys 5, Ac	Arg 2, Me	Ser 1, P
Lys 5, Ac	Lys 12, Ac	Lys 4, Me	Arg 3, Me
Lys 9, Ac	Lys 15, Ac	Lys 9, Me	Lys 5, Ac
Lys 119, Ub	Lys 20, Ac	Ser 10, P	Lys 8, Ac
	Lys 24, Ac	Lys 14, Ac	Lys 12, Ac
	Lys 123, Ub	Arg 17, Me	Lys 16, Ac
		Lys 18, Ac	Lys 20, Me
		Lys 23, Ac	
		Arg 26, Me	
		Lys 27, Me	
		Ser 28, P	
		Lys 36, Me	
		Lys 79, Me	

Figure 1 Covalent modifications of histones. Several-dozen covalent histone tail modifications that contribute to the histone code are summarized. For each species of histone protein, the amino acid residue, position, and type of modification are presented. Me, mono-, di-, or trimethylation; P, phosphorylation; Ac, acetylation.

Cytosine Methylation

Another covalent modification associated with heterochromatin and repression of gene expression is cytosine methylation, which occurs in the DNA double helix predominantly in the context of the 5'-CpG-3' dinucleotide (where p denotes the phosphodiester linkage between the cytosine and guanine nucleotides). In mammalian genomes, 75% of CpG motifs are methylated, and these are concentrated in areas of repetitive DNA sequences. This dinucleotide is the target of DNA methyltransferase proteins, which catalyze the addition of a methyl group to the C5 position of the cytosine ring. DNA methyltransferase 1 is the major maintenance methyltransferase. It exhibits high affinity for hemi-methylated DNA (the immediate product of semi-conservative DNA replication), and catalyzes the addition of a methyl group to the CpG cytosines within the newly synthesized DNA strand. DNA methyltransferases 3a and 3b are *de novo* DNA methyltransferases that are responsible for establishing specific patterns of cytosine methylation during gametogenesis and early development. The mechanisms by which appropriate patterns of cytosine methylation are initially established are not well understood.

Functions of Cytosine Methylation

Appropriate patterns of cytosine methylation within DNA are essential for numerous biological processes. This is clearly reflected in the finding that mice lacking DNA methyltransferase gene(s) die prior to or shortly after birth.

Regulation of Gene Expression

As mentioned above, cytosine methylation is associated with heterochromatin and repressed gene expression. As the majority of 5-methylcytosine is found in repetitive DNA elements, this may provide a defense mechanism against expression of parasitic

DNA. Indeed, loss of appropriate cytosine methylation results in the induction of repetitive DNA expression. Maintenance of repetitive elements in heterochromatin configuration may also be important for chromosome stability by preventing homologous recombination between repetitive element sequences.

X-Chromosome Inactivation

X-chromosome inactivation refers to the irreversible inactivation during early embryonic development of one of the X-chromosomes present in each cell of a female. The inactive X-chromosome becomes heavily methylated at CpG motifs and takes on a heterochromatin configuration.

Genomic Imprinting

Distinct patterns of cytosine methylation are established throughout the genome during spermatogenesis and oogenesis, leading to parent-of-origin epigenetic marks that distinguish maternally derived alleles from paternally derived ones. Such genomic imprinting can result in a uni-allelic pattern of gene expression. Genomic imprinting is essential for normal development, as zygotes carrying two sets of paternal or two sets of maternal alleles fail to develop normally, even though they carry an appropriate diploid complement of the DNA genome. To date, 150 mammalian genes have been identified that exhibit genomic imprinting.

Cancer

A large number of DNA mutations in oncogenes and tumor suppressor genes have been identified in cancer cells. More recently, it has been recognized that cancer cells also exhibit epigenetic aberrations, including global cytosine hypomethylation, as well as hypermethylation of CpG motifs, specifically within the promoters of tumor suppressor genes. Pharmacologic modulation of the epigenome represents a novel chemotherapy strategy. For example, inhibition of HDAC enzymes is being utilized to induce expression of inappropriately silenced tumor suppressor genes in cancer patients.

Methyl-CpG-Binding Proteins

The function of methylated CpG motifs is mediated in part by methyl-CpG-binding proteins, which specifically bind to DNA at sites of methyl-CpG and affect transcription rates and local chromatin structure. A small family of proteins share a common DNA-binding domain, the methyl-CpG-binding domain (MBD) (Figure 2).

MeCP2

MeCP2 is the founding member of the MBD protein family. It localizes to heterochromatin *in vivo* and represses transcription via a carboxyl terminus transcriptional repression domain (TRD). MeCP2 also contains an amino terminus MBD DNA-binding domain. One mechanism for MeCP2-mediated

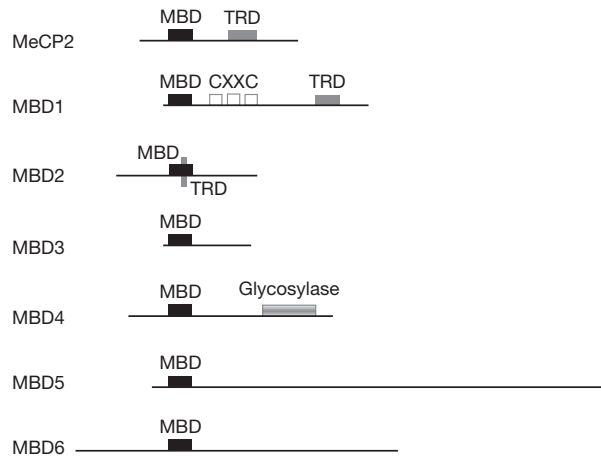


Figure 2 Schematic comparison of the MBD family of proteins. The conserved methyl-CpG-binding domains (MBD) present in each protein are aligned. TRD, transcriptional repression domain.

transcriptional repression appears to be regulation of chromatin structure, as the TRD of MeCP2 interacts with a co-repressor that contains an HDAC. Disruption of the X-chromosome-linked MeCP2 gene in mice leads to premature death, and mutations in the human MeCP2 gene causes Rett syndrome, a progressive neurodegenerative disease found nearly exclusively in girls. Deficiency of MeCP2 in neural cells leads to increased histone acetylation and inappropriate transcription of repetitive DNA elements. The severity of symptoms in Rett syndrome depends on the exact nature of mutation present in the MeCP2 gene (most of which alter the sequence of the MBD or TRD domains), and the mosaic pattern of mutant gene expression resulting from random X-chromosome inactivation.

MBD1

Similar to MeCP2, MBD1 contains an amino terminus MBD DNA-binding domain and a carboxyl terminus TRD, although the TRD exhibits no sequence similarity to the MeCP2 TRD. MBD1 binds methylated DNA targets and represses gene expression. Various isoforms of MBD1 produced by differential RNA splicing additionally contain two or three CXXC domains, a cysteine-rich motif also found in DNA methyltransferase 1, human trithorax (HRX), and CXXC finger protein 1 (CFP1). The CXXC domain constitutes the DNA-binding domain of CFP1, and is responsible for that factor's binding specificity for DNA containing unmethylated CpG motifs. Similarly, HRX also binds unmethylated-CpG motifs, and the third CXXC domain of MBD1 is responsible for this factor's ability to bind unmethylated target sequences and repress transcription. The significance of the CXXC domain for MBD1 function is unclear, as it is not required for binding to sites of methyl-CpG motifs. Mice lacking MBD1 are viable, but suffer from reduced neural differentiation and increased genomic instability.

MBD2 and MBD3

MBD2 and MBD3 are the most structurally related members of the MBD family, sharing 75% similarity over the entire length

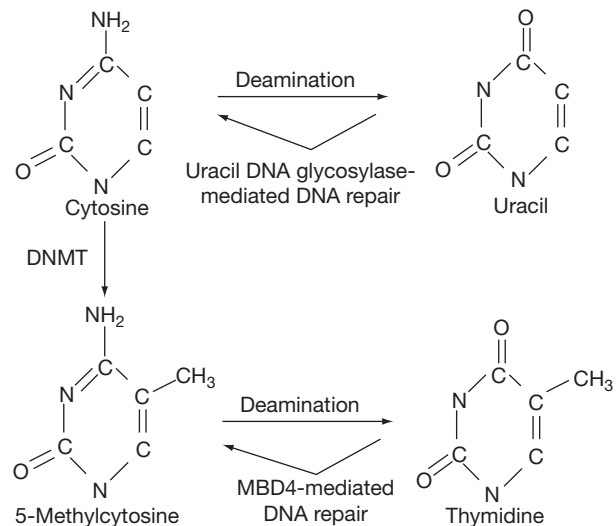


Figure 3 Summary of chemical relationships between cytosine, 5-methylcytosine, uracil, and thymidine, including how these species inter-convert *in vivo*. DNMT, DNA methyltransferases.

of MBD3, the shortest member of the family. MBD2 contains an additional 140-amino-acid amino terminus extension. Despite this structural similarity, these MBD proteins exhibit quite dissimilar properties. MBD2 binds to methylated DNA and represses transcription via overlapping MBD and TRD domains. MBD2 interacts with HDAC1, and is a component of the methyl-DNA-binding complex MeCP1. The ability of MBD2 to repress transcription is sensitive to inhibitors of HDAC activity, suggesting a mechanism for cross-talk between cytosine methylation and chromatin structure. Surprisingly, mice lacking the MBD2 gene are viable despite lacking the MeCP1 complex. By contrast, MBD3 contains a divergent MBD domain and fails to interact directly with methylated DNA. However, similar to MeCP2, MBD3 is a component of a co-repressor complex. The importance of MBD3 is illustrated by the finding that mice lacking an MBD3 gene fail to develop to birth.

MBD4

MBD4 is an unusual member of the MBD protein family, as it apparently plays no role in transcriptional repression; rather, it functions as a DNA-repair protein. Cytosine exhibits a propensity to deaminate spontaneously to form uracil (Figure 3). Cells contain DNA repair enzymes (e.g., uracil DNA glycosylase) that recognize a uracil base in DNA as a site of damage, and excise this nucleotide to permit replacement with cytosine. This may explain the presence of thymidine in DNA rather than uracil, as it would be more difficult for a cell to recognize a site of cytosine deamination if uracil were a normal nucleotide used for encoding genetic information in DNA. However, deamination of 5-methylcytosine leads to the formation of thymidine. It is this rate of spontaneous deamination that leads to the relative rarity of the CpG dinucleotide in mammalian genomes (only ~10% of the expected frequency). Conversely, CpG motifs near widely expressed genes are usually unmethylated, and hence avoid this

mutagenic pressure. This is hypothesized to explain the clustering of unmethylated CpG motifs in CpG islands near the promoters of 50% of mammalian genes. The MBD domain of MBD4 binds with highest affinity to methylated CG/TG mismatched sequences, and the protein also contains a glycosylase activity in its carboxyl terminus, which removes the mismatched thymidine base from DNA to permit replacement with cytosine. The biological significance of this enzymatic activity is illustrated by an elevated rate of mutation at CpG motifs and increased tumorigenesis in mice lacking MBD4.

MBD5 and MBD6

Relatively little is known about these widely expressed MBD-containing proteins. However, haplo-insufficiency of MBD5 leads to severe mental retardation in humans.

Cross-Talk between DNA Methylation and Chromatin Structure

The nature of epigenetic modifications associated with transcriptionally active or inactive genomic loci is now well established. Current investigations are focused on the regulation of these modifications. How does a cell choose which regions of DNA to methylate and/or assemble into heterochromatin? Recently, it has become clear that DNA methylation and chromatin structure are highly interdependent epigenetic processes. For example, perturbations of the cellular machinery responsible for nucleosome remodeling or histone modifications also result in aberrations in the pattern of cytosine methylation. Conversely, the association of MBD family members

with chromatin-modifying enzymes (e.g., HDACs) suggests a mechanism whereby cytosine methylation may provide an epigenetic mark that permits targeted recruitment of chromatin-remodeling enzymes to specific genomic loci. Hence, cytosine methylation patterns and the histone code may represent mutually reinforcing processes that lead to the stable maintenance of chromatin packaging and gene expression patterns characteristic of cellular memory.

See also: [Cell Architecture and Function: Chromosome Organization and Structure, Overview](#); [Molecular Biology: Chromatin: Nucleosome Positioning – the GAL Promoter](#); [Chromatin: Physical Organization; Regulation of Chromatin Dynamics](#).

Further Reading

- Ballestar E and Wolffe AP (2001) Methyl-CpG-binding proteins: Targeting specific gene repression. *European Journal of Biochemistry* 268: 1–6.
- Bienvenu T and Chelly J (2006) Molecular genetics of Rett syndrome: When DNA methylation goes unrecognized. *Nature Reviews Genetics* 7: 415–426.
- Esteller M (2008) Epigenetics in cancer. *New England Journal of Medicine* 358: 1148–1159.
- Felsenfeld G and Groudine M (2003) Controlling the double helix. *Nature* 421: 448–453.
- Jaenisch R and Bird A (2003) Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature Genetics* 33: 245–254.
- Jones PA and Baylin SB (2007) The epigenomics of cancer. *Cell* 128: 683–692.
- Jorgensen HF and Bird A (2002) MeCP2 and other methyl-CpG-binding proteins. *Mental Retardation and Developmental Disabilities Research Reviews* 8: 87–93.
- Nakao M (2001) Epigenetics: Interaction of DNA methylation and chromatin. *Gene* 278: 25–31.
- Wade PA (2001) Methyl CpG-binding proteins and transcriptional repression. *BioEssays* 23: 1131–1137.
- Wade PA (2001) Methyl CpG-binding proteins coupling chromatin architecture to gene regulation. *Oncogene* 20: 3166–3173.

Chromatin: Nucleosome Positioning – the GAL Promoter

D Lohr, Arizona State University, Tempe, AZ, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Activator Proteins that enable or enhance the transcription of genes.

Gene-specific factors Proteins whose function is associated with expression of a specific gene or gene family.

General factors Proteins whose function is associated with expression of many and unrelated genes.

Nucleosome The octameric complex of core histone proteins wrapping up 147 bp of DNA; the basic structural unit of eukaryotic chromosomes.

Promoter The DNA sequences through which control of transcription is implemented.

The impacts of nucleosomes and, in particular, nucleosome locations (positioning) on biological functions such as transcription have been subjects of intense interest from the earliest days of nucleosome study. On some eukaryotic promoters, nucleosomes are positioned on regulatory elements when genes are in the inactive state of expression but undergo alteration (remodeling and removal) during transcription activation. Yeast PHO5 and mammalian mouse mammary tumor virus (MMTV) promoters provide well-studied examples of this organizational mode. On genes, such as PET56 and DED1, promoter elements remain largely nucleosome-free regardless of the state of expression of the associated gene. The GAL1–10 genes in the budding yeast *Saccharomyces cerevisiae* demonstrate both types of promoter nucleosome organization, affecting different classes of promoter elements. Furthermore, the promoter chromatin organization on GAL1–10 has been shown to play a key role in transcriptional regulation of these genes.

GAL1–10: Gene Expression and Chromatin Structural Features

GAL1 and GAL10 encode enzymes that function in the conversion of galactose to a glycolytic intermediate, thus allowing galactose to be used as a carbon source for cell growth. Genetic characterization of the GAL genes and dissection of their regulatory mechanisms by Douglas and Hawthorne in the 1960s provided a firm and crucial foundation for the molecular level characterization of GAL1–10 gene expression and its regulation that began in the late 1970s and the analysis of GAL1–10 chromatin structure that began in the mid-1980s.

Expression of GAL1 and GAL10 has a striking carbon source dependence; transcription occurs at extremely high levels in galactose-grown cells but is completely undetectable in cells grown in other carbon sources, such as glucose or glycerol–lactate (gly–lac). Their clear-cut and highly efficient control (the genes transcribed only when the gene products are needed), plus their genetic and biochemical tractability have made GAL1–10 a very useful model system for studies of gene regulation and its relation to nucleosome organization. GAL1–10 regulation occurs mainly at the transcriptional level, involving inputs from DNA sequence (promoter) elements, protein

factors (gene-specific and general), and chromatin structure. In the following discussion, genes are referred to as GAL1, protein factors as Gal1p.

Promoter DNA Elements

GAL1 and GAL10 share a common, ~600-bp intergenic promoter DNA region from which they are co-regulated and divergently transcribed (Figure 1). Within this region lie the most important promoter elements for GAL1–10 expression, the gene-specific upstream activation sequence (UAS_G) elements. As shown by the Ptashne laboratory, these ~17-bp motifs are necessary and sufficient for galactose-dependent transcription; they provide the binding sites for Gal4p, the specific activator of GAL gene transcription. GAL1–10 share four UAS_G elements asymmetrically located between the genes (Figure 1). GAL1 and GAL10 each have their own TATA sequence for binding the general transcription factor, TATA-binding protein (TBP).

Regulatory and Transcription Factors

Major gene-specific factors

Gal4p

The major and required activator for GAL gene expression is Gal4p. In cells growing in galactose, Gal4p activates transcription through a domain located near its carboxy terminus, residues 767–881, while bound to (UAS_G) DNA via a domain located near its amino terminus. DNA binding and transcription activation are separable functions on Gal4p (established by the Ptashne laboratory). Gal4p binds as a homodimer to individual UAS_G elements.

Gal80p

In carbon sources other than galactose, the negative regulator Gal80p binds to the C-terminal activation domain of Gal4p and masks Gal4p activation activity, thus preventing GAL1–10 transcription. Gal80p binds Gal4p quite strongly ($K_d \sim 5$ nM).

Gal3p

In the presence of galactose, the Gal80p-mediated inhibition of Gal4p is relaxed via a Gal3p-dependent process, thus freeing the Gal4p activation domain to activate GAL1–10 transcription.

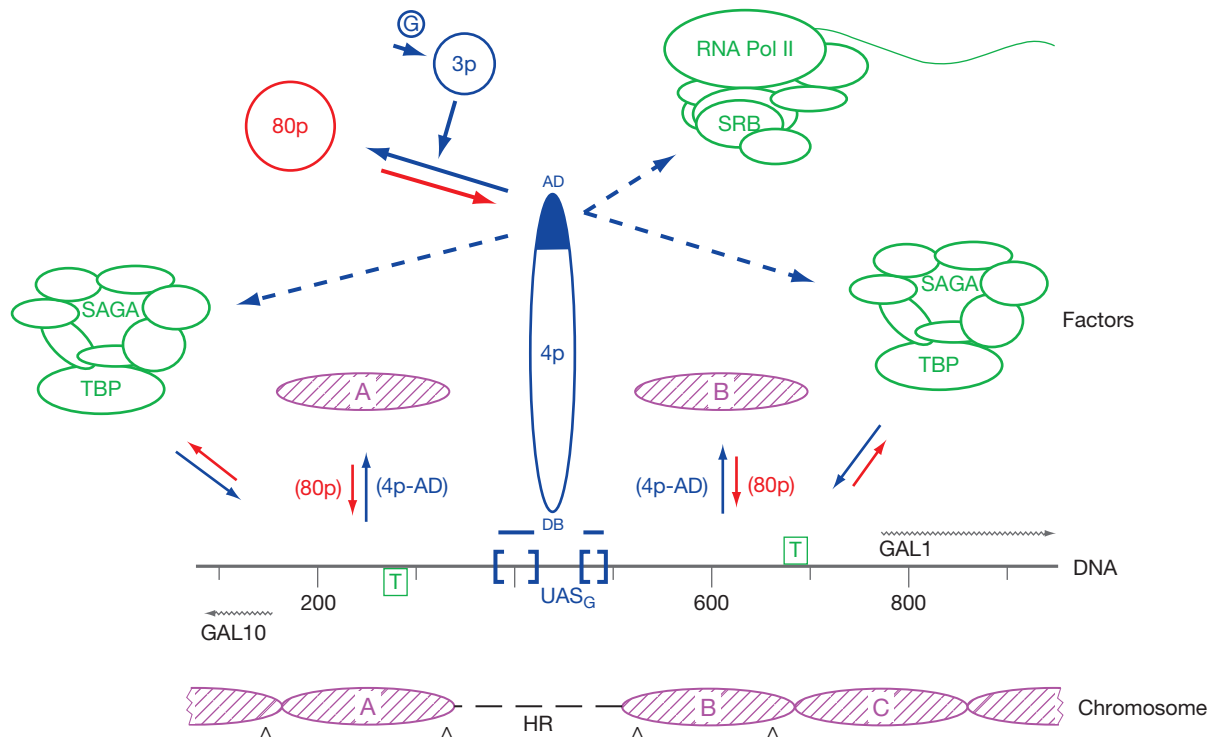


Figure 1 An outline of GAL1–10 promoter organization and gene regulation. The DNA sequence organization of the promoter region is shown at the center of the figure (DNA), with UAS_G (blue brackets) and TATA boxes (green, boxed T" located to scale along the sequence (gray line). The numbers refer to base pairs from an EcoRI site in the GAL10 gene. The transcription start sites lie at the origins of the two wavy arrows below (GAL10) or above (GAL1) the line. The chromosomal structure of this region in the inactive state of gene expression is shown at the bottom of the figure (Chromosome) with nucleosomes A, B, and C (pink) located at their correct sites on the sequence and the nucleosome-free region in the chromosome indicated by a dashed line (HR). The locations of intrinsic, sequence-specific DNA bends are identified by the symbol $\wedge$. In the topmost portion of the figure (Factors), various factors and complexes known or thought to be involved in regulating and implementing GAL1–10 expression are shown. The GAL-specific expression-enabling factors (3p, Gal3p; 4p, Gal4p) are shown in blue and solid blue arrows designate the reactions that occur in, or are characteristic of, the active (galactose-induced, G) state. Dotted blue arrows indicate the suggested recruitment of complexes by Gal4p during gene activation, such as RNA polymerase II (RNA pol II), including the suppressor of RNA polymerase B subunit (SRB) proteins and the Spt-Ada-Gen5-acetyltransferase (SAGA) complex. The activation domain (AD) of Gal4p is solid blue and the DNA-binding domain is identified (DB). The GAL-specific negative regulator, Gal80p (80p), is shown in red and solid red arrows designate the reactions that occur in, or are characteristic of, the inactive state of gene expression. The association/dissociation reactions between TBP and the GAL1 and GAL10 TATA and the disruption/redeposition reactions of promoter nucleosomes A and B are also designated by solid blue/red arrows, again color-coded to reflect the situation in the active/inactive state of expression. The involvement of Gal4p and Gal80p in the promoter nucleosome disruption/redeposition cycle is indicated by their appearance in parentheses next to the appropriate arrow. Nucleosome C, which also undergoes similar disruption/redeposition behavior to A and B, is not shown in the upper portion of this figure, for simplicity.

Interestingly, the GAL1 gene product, Gal1p, contains a Gal3p-like activity that should make it capable of fulfilling a similar function to Gal3p.

General factors

Factors associated with global processes, notably the general glucose repression system, also participate in GAL1–10 regulation and elements of the RNA polymerase II machinery (cf. SAGA, Mediator, and Swi-Snf) function in GAL1–10 transcription. Genomic studies have identified other galactose-sensitive and Gal4p-regulated genes, but whether they are involved in GAL1–10 regulation remains unknown.

Promoter Chromatin Organization

The promoter of GAL1–10 has a distinctive and functionally suggestive chromatin organization (as first shown in the

Lohr laboratory). Within the intergenic domain, there is a sizable (at least 150 bp), highly accessible stretch of DNA. This corresponds to a nucleosome-free, hypersensitive region (HR, Figure 1). This HR region is present regardless of the state of gene expression, that is, in all carbon sources, and all four UAS_G elements reside well within the region. In the inactive states of gene expression (nongalactose carbon sources), positioned nucleosomes (A–C, Figure 1) bracket the HR region. These promoter nucleosomal regions include the TATA and transcription start sites for each gene. Note that nucleosomes cover the GAL10 TATA and GAL1 start site but bracket the GAL10 start site and GAL1 TATA rather than uniformly covering all of these elements. Whether this is significant is unclear. The coding regions of the genes are also nucleosome covered (not shown).

Just how precisely promoter nucleosomes A–C are positioned remains uncertain but both the HR–nucleosomal

region boundaries and the internucleosomal boundaries are sharp and constant in location. Thus, the UAS_C elements are always nucleosome-free and the adjacent TATA and transcription start sites nucleosome involved. The precision of nucleosome positioning does clearly diminish away from the promoter region (in the coding sequences). As shown by the Lohr laboratory, the DNA regions near the termini of nucleosomes A and B contain strong intrinsic, sequence-dependent DNA bends (Figure 1), which could contribute to the preferential location of these nucleosomes *in vivo*. However, *in vivo* positioning must also involve other features because *in vitro* nucleosome reconstitution of GAL1–10 intergenic region DNA does not produce detectable positioning (shown by the Lohr laboratory). It is likely that the nucleosome-free HR region, which is a constant and unchanging feature of GAL1–10 promoter chromatin, helps position nucleosomes A–C *in vivo*, perhaps by acting as a boundary. HR region DNA readily forms nucleosomes *in vitro*, so its nucleosome-free nature *in vivo* must reflect specific cellular nucleosome-exclusion mechanisms, perhaps involving higher-order chromosomal features. What these might be remains unknown.

This promoter chromatin organization – accessible HR, flanking nucleosomes – has been established by micrococcal nuclease and chemical cleavage studies, as well as a novel DNase I analysis that tested small stretches across the intergenic chromatin region for their ability to produce the nucleosome-diagnostic, 10-bp ladder pattern. This organization has been observed in nuclei and *in vivo*, in studies from several laboratories. The promoter regions of other GAL genes, including the regulatory gene GAL80, show a similar chromatin organization. Thus, it may be a general feature of GAL genes.

In the carbon source galactose (the active state of expression), the three promoter nucleosomes A–C appear to be completely removed (disrupted), thus exposing the underlying DNA sequences, including the TATA and start sites, to transcription factors. The HR region remains highly accessible. Recent *in vitro* fluorescence studies (involving the Lohr laboratory) have shown that nucleosome A, which covers the GAL10 TATA, has a lower intrinsic stability and higher levels of inherent DNA dynamics than nucleosomes typically studied. Thus, this nucleosome may be especially susceptible to disruption.

In contrast to the promoter nucleosomes, the GAL1 coding sequence nucleosomes remain present under actively induced conditions. However, they do appear to have undergone alteration compared to their structure in the inactive state. This probably involves conformational change and perhaps a reshuffling of locations, although the precise nature of the alteration remains unclear. Coding sequence and promoter nucleosomes on GAL1–10 show other differences (see below).

The Role of Promoter Chromatin Structure in GAL1–10 Regulation

Factors and Chromosomal Organization

The active (galactose-induced) state

In cells growing in galactose, the transcription activation function of Gal4p is free of Gal80p inhibition and the GAL1–10 genes are transcribed at very high levels. Under these conditions, Gal4p is bound, via its N-terminus, to the UAS_C elements in the

HR region between GAL1 and GAL10 (Figure 1) while activating transcription through its C-terminal activation domain. The UAS_C elements are more than 200 bp (>680 Å in linear DNA distance) from the GAL1 to GAL10 transcription start sites, suggesting that the three-dimensional structure of Gal4p and promoter chromatin architecture, both unknown, may impact the transcription process.

In this active state, the DNA in the TATA and transcription start site regions is freed from nucleosome protection, due to the complete disruption of nucleosomes A–C. This DNA exposure should greatly facilitate the binding of factors to these regions, such as the binding of TBP to the GAL1 and GAL10 TATA elements. TBP cannot bind to DNA that is nucleosome covered.

Inactive states (poised or repressed)

There are two distinct types of inactive states, repressed (in glucose) or poised for expression (in nonrepressing, noninducing carbon sources such as gly–lac). In both types of inactive states, there is no detectable transcription of the GAL1–10 genes. In both, Gal80p is bound to the Gal4p activation domain, inhibiting its activity, and nucleosomes A–C are present on the TATA and transcription start site regions, inhibiting factor access to those DNA sequences.

In the poised state (cells growing in gly–lac), the inactive GAL1–10 genes can be rapidly (within minutes) brought to full, high-level expression by the addition of galactose to the growth medium. In gly/lac, Gal4p is bound to the GAL1–10 UAS_C elements, even though there is no transcription. Moreover, the level of UAS_C protection by Gal4p, and thus the strength of Gal4p–UAS_C binding, appears to be similar to that in galactose-grown cells. Elevated Gal4p levels (GAL4 is most highly expressed in carbon sources such as gly–lac) plus constitutive accessibility of the UAS_C elements in the chromosome structure probably account for the UAS_C occupation by Gal4p, and the rapid inducibility, in these carbon sources. Thus, although inactive, the GAL1–10 genes are poised for the rapid induction of expression.

The gly–lac results show that the DNA binding and transcription activation functions of Gal4p act independently in the cell and respond to different controls: DNA binding to Gal4p levels and transcription activation to Gal80p. In addition, location of the UAS_C elements within a large, nucleosome-free region means that Gal4p does not have to compete with nucleosomes nor depend on the action of nucleosome modification factors to gain access to its target elements. The constitutive accessibility of the UAS_C and the independence of Gal4p DNA binding and activation functions are key features (along with elevated Gal4p levels) that permit the existence of a poised state. Such a state would have been a major advantage to wild yeast, allowing cells growing on poorer carbon sources to efficiently utilize galactose, a relatively good carbon source, even if it is only transiently available.

The presence of glucose imposes an additional negative regulation on the GAL1–10 genes, the absence of Gal4p from the UAS_C. This feature prevents GAL1–10 from being (rapidly) inducible in glucose. The UAS_C elements reside in the same size, highly accessible, nucleosome-free region in glucose- or gly/lac-grown cells and thus should be equally available for Gal4p binding in both carbon sources. The absence of Gal4p from the UAS_C is thought to result, at least in large part, from low Gal4p levels due to decreased expression of the GAL4 gene

in glucose. This regulation is imposed by the global, glucose repression apparatus, a system that helps insure the preferential utilization of glucose. The decrease in GAL4 expression is modest but Gal4p–UAS_G binding should be very sensitive to Gal4p levels due to its highly cooperative nature (multiple UAS_G/two Gal4p per UAS_G).

Regulating GAL1–10 Expression: A Dynamic Interplay of Promoter Nucleosomes and Regulatory Factors

Chromatin structure is now seen as a major facet of eukaryotic gene regulation; studies of GAL1–10 promoter chromatin, beginning in the mid-1980s (in the Lohr laboratory), provided early indications of this important association. Studies since that time, in several laboratories, have provided insights into various aspects of this relationship on GAL1–10. These studies have shown that regulation of GAL1–10 expression involves complex interactions among the gene-specific regulatory factors, general factors, and the promoter nucleosomes A–C. The rigor of these conclusions reflects the ability to couple biochemical and genetic analyses in the study of genes in single-copy number.

In addition to their other functions, the specific regulators Gal4p and Gal80p also play crucial roles in determining nucleosome occupation on the GAL1–10 promoter region. Gal4p mediates the characteristic disruption of the promoter nucleosomes A–C observed in the induced state. As noted above, this provides DNA access for factors such as TBP. This nucleosome disruption has been shown to be a specific and dedicated activity of the Gal4p activation domain and not simply a secondary consequence of the transcription process (by the Majors laboratory). Whether Gal4p acts directly or through other factors is unknown. Other activation-associated activities of Gal4p include the recruitment of cellular complexes such as SAGA, Mediator, probably Swi-Snf, and perhaps RNA polymerase II itself (Figure 1).

In addition to inhibiting Gal4p activation activities in nongalactose carbon sources, the negative regulator Gal80p also mediates the replacement (reposition) of disrupted promoter nucleosomes A–C when conditions are unfavorable (loss of activation signals, low cellular energy levels) for transcription (shown by the Lohr laboratory). This can even occur in cells initially grown in galactose, when they are starved for energy. Energy sensitivity is not unexpected because it links the disrupted state, and thus transcription capability, to cellular energy levels. Whether Gal80p carries out redeposition directly or through other factors is unclear.

Gal80p modulates the levels of GAL1–10 induced expression, keeping them lower than inherently possible. This may explain the surprising observation that expression of this negative regulator is actually enhanced in galactose (5–10-fold); enhancement is Gal4p-dependent and mediated through a UAS_G on the GAL80 promoter. Gal80p is also involved in maintaining promoter nucleosome coverage under inactive (noninduced) conditions; Gal80p removal results in a partial loss of protection in the nucleosome A–C regions, and low levels of GAL1 expression, under these conditions. Both promoter modulation under induced conditions and nucleosome maintenance under noninduced conditions may involve the nucleosome redeposition activity of Gal80p.

The reciprocal processes of nucleosome disruption, mediated by Gal4p, and redeposition, mediated by Gal80p, are dynamic and reversible. Galactose addition can trigger promoter nucleosome disruption in minutes; nucleosome redeposition, during starvation for example, occurs in well under an hour (and perhaps minutes as well). The three promoter nucleosomes, A–C, are always found in the same state, disrupted or intact, and all three show a similar disruption/redeposition response, suggesting that they are affected in unison and by the actions of common, concerted processes.

Gal4p/Gal80p-dependent disruption/redeposition is a unique response of the promoter nucleosomes A–C. The GAL1 coding region nucleosomes do not show such a response (nor do they undergo complete disruption under induced conditions (see above)). That the promoter nucleosomes are specific targets of Gal4p/Gal80p action suggests they have a unique, regulatory type of role as opposed to the coding sequence nucleosomes whose responses are consistent with secondary consequences of the expression process. Gal4p–UAS_G binding in the HR region has no detectable effect on nucleosomes A–C (cf. in gly/lac) and disruption of nucleosomes A–C (in galactose) occurs with no apparent effect on the accessibility of the HR region. Thus, the nucleosome-free HR, the promoter nucleosomes A–C, and the coding sequence nucleosome region seem to behave independently, like small mini-domains.

The Grunstein laboratory has studied, in some detail, the role of nucleosomes and histones in GAL1–10 expression. Nucleosomes have a general repressive effect; their (partial) depletion allows some TATA-driven, Gal4p-independent GAL1 transcription, even in glucose. Repression requires the N-terminal tails of all four histones. The promoter nucleosomes and particular histone components also play specific roles in GAL1–10 expression. The N-terminal tails of histone H4 are needed for full levels of galactose-induced GAL1 expression; this positive role involves the nucleosome B region on the promoter. By contrast, the H3 tails function in a process that modulates induced expression; their removal results in an elevated level of GAL1 transcription in galactose. This effect depends on the UAS_G elements. The removal of H4 N-terminal tails decreases DNA accessibility just upstream of the GAL1 TATA, in the nucleosome B region. Deletion of H3 tails has no effect on the promoter nucleosome regions, consistent with the UAS_G association of the H3 effect.

These H3/H4 effects probably involve acetylation because mutation of the tail acetylation sites produces the same effects as tail removal. One possibility is that these tails are contact sites for nucleosome transactions such as disruption/redeposition. For example, removal of Gal80p and H3 tail removal produce similar elevations in induced expression. Notably, the effects of H3 and H4 tail deletions on coding sequence nucleosomes are similar to one another, so the above specific effects are further examples of unique promoter nucleosome features.

Binding of TBP to the TATA regions on the GAL1–10 promoter is undoubtedly a crucial feature and a critical early step in the transcription process. TATA occupancy on the GAL1–10 promoter can be viewed as a competition between TBP and the promoter region nucleosomes A and B/C (Figure 1); the presence of the latter excludes the former. Active conditions (growth in galactose) produce the disruption of nucleosomes A–C, via a Gal4p-dependent process. The attendant exposure

of the TATA-containing DNA regions favors TBP binding and transcription. Under gene-inactive conditions (growth in non-galactose carbon sources or low cellular energy), nucleosome presence on the A–C regions is favored, mediated via a Gal80p-dependent process, and TBP binding and transcription are disfavored (**Figure 1**). Therefore, TATA availability, and thus transcriptional activity, reflects a nucleosome disruption/redeposition competition mediated by the gene-specific regulators Gal4p and Gal80p. Such a competition can provide a dynamic switch for the control of Gal1–10 transcription. Gal4p and Gal80p may be well suited for such a control function because they probably associate with each other in all carbon sources, albeit in differing manners.

In summary, the major GAL1–10 promoter elements (UAS_G) reside in a constitutively accessible, nucleosome-free region, always available for binding by the major activator Gal4p; Gal4p does not have to compete with nucleosomes to access the UAS_G. Positioned, promoter nucleosomes surround this accessible region; their presence or absence depends on dynamic, gene-specific regulatory factor-mediated processes that control DNA access in these regions and thus TATA and transcription start site availability.

Acknowledgment

This article is dedicated to the memory of Ralph Bash and his skill and dedication to the study of chromatin as well as to composing illustration art such as **Figure 1**.

See also: **Molecular Biology:** Chromatin: Physical Organization; Gene Expression in Eukaryotes: RNA Polymerase II Structure; Nuclear Organization, Chromatin Structure, and Gene Silencing; Regulation of Chromatin Dynamics.

Further Reading

- Bash R and Lohr D (2001) Yeast chromatin structure and regulation of GAL gene expression. *Nucleic Acid Research and Molecular Biology* 65: 197–259.
- Johnston M (1987) A model fungal gene regulatory mechanism: The GAL genes of *Saccharomyces cerevisiae*. *Microbiological Reviews* 51: 458–476.
- Johnston M and Carlson M (1992) Regulation of carbon and phosphate utilization. In: Jones EW, Pringle JR, and Broach JR (eds.) *The Molecular and Cellular Biology of the Yeast Saccharomyces: Gene Expression*, vol. 2, pp. 193–281. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Reece R and Platt A (1997) Signalling activation and repression of RNA polymerase II transcription in yeast. *BioEssays* 19: 1001–1010.
- Lohr D (1997) Nucleosome transactions on the promoters of the yeast GAL and PHO genes. *Journal of Biological Chemistry* 272: 26795–26798.
- Lohr D, Venkov P, and Zlatanova J (1995) Transcriptional regulation in the yeast GAL gene family: A complex genetic network. *FASEB Journal* 9: 777–787.

Chromatin: Physical Organization

C L Woodcock, University of Massachusetts, Amherst, MA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Chromatin Generic term used for (1) material in the nucleus that stains strongly with basophilic dyes and (2) the complex of DNA and chromatin that may be extracted from nuclei for *in vitro* study.

Histone Family of highly conserved, small basic proteins that constitute the principal protein components of chromatin.

Histone modification Covalent addition of one or more chemical groups to a specific amino acid in a histone. Modifications include acetylation, methylation, phosphorylation, poly-ADP ribosylation, and ubiquitination (ADP, adenosine diphosphate).

Histone variant (subtype) A histone molecule derived from a different gene and having specific amino acid differences from the canonical histone type.

Nonhistone chromatin proteins Generic term for proteins other than histones that may be bound to chromatin.

Nucleosome The complete repeat unit consisting of a nucleosome core particle and the linker DNA that continues to the next nucleosome core particle, often with histone H1.

Nucleosome core particle The disk-shaped (11 nm diameter, 6 nm high) complex of 146 bp of DNA and octamer of histones H2A, H2B, H3, and H4 that is the fundamental structural unit of chromatin. The X-ray structure of the core particle has been solved.

The term 'chromatin' was coined by cytologists to designate the brightly staining portions of nuclei exposed to basic dyes. Although this terminology is still in effect, in the context of biological chemistry, chromatin usually refers to the complex of DNA and protein that is present in the interphase nucleus, and can be isolated for study *in vitro*. As with all biological systems, the physical organization of chromatin results from the intrinsic biochemical properties of the components as influenced by the composition of the aqueous milieu. As chromatin organization plays a crucial role in many aspects of gene regulation, there is much interest in understanding the basic principles involved. Here, the current status of this challenging task is discussed in terms of the major components of chromatin, their interactions, and the complex, dynamic biopolymer that results.

A crude analysis of isolated chromatin reveals a 1:1 ratio of DNA and histones, small, highly conserved basic proteins. Four species of histone (H2A, H2B, H3, and H4) contribute two copies each to the histone core of the nucleosome, a disk-shaped structure 11 nm in diameter and 6 nm in height, while histone H1 binds to the exterior of the nucleosome. Other proteins bound to DNA in chromatin are collectively known as nonhistone chromatin proteins (NHCPs) and include both structural and enzymatic components. Thus, at the simplest level, the physical organization of chromatin consists of very long linear arrays (one for each chromosome) of nucleosomes arranged in a beads-on-a-string conformation, together with bound NHCPs. This primary structure is variously folded and compacted, and (in the interphase nucleus) occupies a distinct volume or territory with little or no intermixing with other chromosome territories. Prior to cell division, a complex series of events transforms interphase chromatin into the familiar linear metaphase chromosomes. It is clear that the arrays of nucleosomes are exposed to a wide range of local and global factors that influence the degree of chromatin compaction, and the challenge has been to understand the impact of these

factors on the physical organization. There is strong evidence that features of chromatin organization such as compaction level are related to mechanisms of differential gene expression. Further, some aspects of chromatin organization are important in epigenetic inheritance, whereby properties such as differential gene silencing or activation are inherited not as differences in DNA sequence but as specific chromatin structures and their components.

Global factors that influence the physical organization of chromatin include the ionic milieu of the nucleus, especially the concentrations of monovalent and divalent cations, and polycations such as polyamines. The combination of charge shielding and charge neutralization that results from exposure of chromatin to cations *in vitro* results in stronger intra-chromatin interactions (as the self-repulsive properties of DNA are negated) and greater compaction. Indeed, at the concentration of cations thought to be present in the living nucleus, chromatin *in vitro* tends to aggregate and essentially precipitate. Modulating the effects of cations are local variations in nucleosomes (histone variants and histone modifications), DNA modifications (chiefly cytosine methylation), and bound NHCPs (often related to histone modifications). Thus, the chromatin template has the potential to assume a continuum of compaction levels, and it is in this context that its physical organization has to be viewed. In this article, the principal variables that can influence the structure and function of a given portion of the genome are outlined, after which the question of physical organization is discussed.

Variables That May Influence Chromatin Physical Organization

Posttranslational Modifications of Core Histones

The presence of a large variety of core histone modifications has been known for many years, but its role in chromatin

was very poorly understood. Recently, however, evidence has grown that the pattern of histone modifications on an individual nucleosome or array of nucleosomes may encode information that is essential in determining a variety of functional states. Many modifications occur on the N-termini of the core histones which project from the nucleosome and are unstructured in the sense that they are not detected in nucleosome crystals by X-ray analysis. Some, such as acetylation and methylation of lysine residues, result in the net loss of a positive charge and may result in a change in chromatin conformation *per se*. However, it is clear that the functional implications of histone modification result from far more complex effects than modulations in compaction resulting from alterations in charge balance.

For example, H4 acetylation can occur at one or more of four lysines in the N-terminus, and these modifications alone, and in combination, carry important functional signals. The net acetylation pattern is the result of the balance between histone acetylase (HAT) enzymes and histone deacetylase (HDAC) enzymes. Both HAT and HDAC enzymes tend to occur in large complexes that may contain proteins that serve to target the enzymatic activity to the appropriate gene(s). Specific acetylation patterns also occur on newly synthesized chromatin and are erased after deposition as nucleosomes move close to the replication fork.

Another key set of histone modifications involves the methylation of lysines on histone H3. Here, another level of complexity arises because the methyl transferases may deposit one, two, or three methyl groups, and the number of methyl groups provides another signal. Certain methylation patterns are important in providing binding sites for a group of NHCPs known as 'heterochromatin protein 1 (HP1)'. HP1 tends to accumulate on transcriptionally silent chromatin and appears to be involved in the downregulation of genes.

Variants of Core Histones

Typically, higher organisms have several genes for each of the core histones, some of which may be identical in sequence, while others vary in amino acid sequence. The majority of core histone genes is transcribed only during DNA replication and is incorporated into nucleosomes in the wake of the replication fork. Others are synthesized constitutively and incorporated into nucleosomes only after the relatively rare event of displacement of a complete histone octamer. An interesting example is provided by the replication coupled 'bulk' H3 and the H3.3 variant which can be deposited in both replication-coupled and replication-independent pathways, and which differ at only four amino acid positions. Single amino acid changes of H3 toward H3.3 allow replication-independent deposition. H3.3 accumulates modifications that favor transcription, while H3 accumulates lysine methylations, a mark of inactive chromatin. The ratio of H3.3 to H3 increases steadily in nondividing long-lived cells (such as neurons) and, through this mechanism, may become the major H3 variant. Another H3 variant which is highly conserved evolutionarily occurs only in centromeric chromatin and plays a crucial role in cell division.

Histone H1

Histone H1, also known as 'linker histone', plays a dominant role in establishing the compaction state of an array of nucleosomes as well as influencing the conformation. H1 has three domains: a central globular domain that binds near the entry/exit site of linker DNA on the nucleosome, and extended N- and C-terminus. The C-terminus is particularly rich in the basic amino acids lysine and arginine and, thus, has a strong propensity to bind DNA. H1 binding protects an extra 20 bp of nucleosomal DNA from nuclease digestion, a protection conferred by the globular region alone. There is a general correlation between the amount of H1 in chromatin and its ability to be transcribed, and most models of transcriptional regulation call for the temporary displacement of H1. The silencing effect of H1 could be due to the sealing of two turns of nucleosomal DNA, thus preventing the unwrapping needed for RNA polymerase, or due to the H1-induced physical compaction of arrays of nucleosomes, or most likely to a combination of both.

Like the core histones, there are several variants of H1 in most higher organisms, but little is known about their individual properties. Genetic manipulations of mice that knock out one or two variants produce no observable phenotype, and the remaining variants are upregulated to compensate for the loss. The loss of additional variants leads to a depletion in total H1 and eventually to embryonic lethality, although embryonic stem cells may be rescued from H1-depleted nonviable embryos and grown in culture. H1-depleted nuclei do not exhibit a general depression of transcription, as predicted by the simple model that H1 loss should create open and permissive chromatin.

Nonhistone Chromatin Proteins

A two-dimensional (2D) gel of the proteins bound to the chromatin of an active nucleus reveals hundreds of spots, and it is clear that there are many other proteins present in low abundance that are revealed only after staining with specific antibodies. To date, relatively few have been studied in detail, and most of these are either abundant in the nucleus or came to researchers' attention through functional studies of specific genes or classes of genes. Most NHCPs can be eluted from chromatin by treatment with 0.3–0.4 M monovalent ions (usually NaCl), in contrast to H1 which requires 0.5–0.6 M NaCl and the core histones which elute between 0.8 and 2.0 M salt. There is some evidence that the few proteins that remain bound in 2.0 M salt contribute to a nuclear skeleton (karyoskeleton and nuclear matrix), which provides structural integrity to the nucleus and has an important dynamic role in its spatial organization as well as replication and transcription. However, despite much effort, it has not been possible to reveal an *in vivo* karyoskeleton analogous to the cytoskeleton, and the possibility that the residual proteins in salt-extracted (and usually DNase-treated) nuclei represent an aggregation phenomenon has not been excluded. The finding that the most abundant nuclear matrix proteins are associated with the peripheral nuclear lamina or are part of the RNA transcript processing machinery has reinforced these concerns.

Table 1 Hierarchical classification scheme for chromatin physical organization

<i>Level of chromatin physical organization</i>	<i>Examples of global features</i>	<i>Examples of local features</i>
Primary – the linear arrangement of nucleosomes on DNA	The nucleosome repeat length	Preferred locations of nucleosomes and features such as DNase hypersensitive sites on a specific DNA sequence.
Secondary – structures formed by interactions of nucleosomes	The 30-nm chromatin fiber	3D architecture of nucleosomes and regulatory proteins on a specific DNA sequence.
Tertiary – structures formed by interactions between secondary structures	Thicker fibers seen in nuclei and postulated to be composed of 30-nm fibers	Long-distance contacts possibly involving locus control regions, enhancers and promoters, or looped chromatin domains.
Quaternary and above – structures formed by interactions between tertiary structures, etc.	No unambiguous examples – see text	Interphase chromosome territories; metaphase chromosomes

The ubiquitous and abundant (typically 1 per 10–15 nucleosomes) high-mobility group (HMG) proteins elute from nuclei in 0.3–0.4 M salt and were originally grouped together on the basis of their solubility properties and rapid mobility in gels. Members of the HMG-B group share the HMG-box functional domain, which contains three α -helices and, when bound to double-stranded DNA, induces a bend of 90°. By contrast, the HMG-N family of proteins bind to nucleosomes via a nucleosome-binding domain (NBD), resulting in a change of conformation and the stimulation of transcription and replication. Finally, the HMG-A family members contain three A-T hook motifs that bind preferentially to the minor groove of DNA rich in AT sequences. HMG-A proteins appear to participate in a wide range of chromatin-based processes and are often upregulated in cancerous cells. The lack of sequence specificity of all the HMG proteins suggests that they are targeted to the appropriate sites by interactions with other nuclear proteins, and evidence is accumulating for many such interactions.

The HP1 family of NHCPs, known to play a role in epigenetic inheritance, is also conserved across a wide range of eukaryotes, with three isoforms, α , β , and γ , found in mammalian nuclei. HP1 (as well as several other important NHCPs) molecules share a chromo domain which binds not to DNA but to the methylated form of lysine 9 on histone H3. HP1 also interacts with numerous other proteins, including an enzyme that methylates H3. This illustrates the complex interactions and feedback loops that are probably widespread in the nucleus.

The protein families discussed above are certainly only a small subset of the repertoire of NHCPs, but perhaps representative of the different modes of action, which range from modifying chromatin architecture to acting as targets for other proteins. Most NHCPs are not stably bound to chromatin, but are remarkably mobile within the nucleus.

Levels of Physical Organization

The large number of variables in chromatin composition has the potential to influence its global and local physical organization, especially the degree of compaction, and the related level of transcriptional competence (higher compaction correlates with reduced transcription). A useful approach is to consider chromatin compaction in terms of hierarchical levels of folding, using a nomenclature based on that devised for

proteins (Table 1). For chromatin, it is necessary to distinguish between local structures that may be confined to specific genes, portions of genes, or groups of genes and global structures that are nucleus wide.

Global Primary Structures – the Nucleosome Repeat Length

While the length of DNA incorporated into the nucleosome is fixed at 146 bp, the length of linker DNA between nucleosomes is variable. At the level of the whole nucleus, each cell type in an organism has a characteristic average nucleosome repeat length (NRL), most mammalian cells having NRLs between 175 and 195 bp. The longest NRLs, measured to date, ~240 bp, are in echinoderm sperm (which have unusual core and H1 histones) and the shortest, ~165 bp, are derived from yeast (*Saccharomyces cerevisiae*), which lack a typical histone H1. The number of H1 molecules per nucleosome varies among cell types; there is a straight-line relationship between H1 abundance and NRL, cells with higher H1/nucleosome ratios having longer NRL values. Such a relationship tends to maintain a fixed net charge within nuclei.

Local Primary Structures

For a given cell type, local NRLs vary about the global mean, with spacings between individual nucleosomes somewhat influenced by the underlying DNA sequence. In addition, there are many instances where nucleosomes on particular genes and gene-regulatory sequences may be positioned with base-pair precision, and thus the local NRL may differ from the global NRL. This local positioning often contributes to transcriptional regulation. For example, a nucleosome positioned on a promoter sequence may block the binding of proteins and protein complexes that are required for transcription initiation. Relieving such a transcriptional block may be achieved by shifting nucleosomes, a process catalyzed in an adenosine triphosphate (ATP)-dependent reaction by several families of large, multi-component nucleosome remodeling complexes. Other cases are known where a specific DNA sequence is normally nucleosome free. These sites, which are most common in gene-regulatory sequences and are probably occupied by NHCPs *in vivo*, may be detected by their hypersensitivity to endonucleases.

Specialized DNA sequences may have a characteristic NRL that differs from the global value. For example, the long stretches of gene-free chromatin at the ends (telomeres) of chromosomes, which consist of tandem repeats of TTAGGG, typically have a shorter NRL than bulk chromatin.

Local primary structures may also be influenced by cytosine methylation of CpG dinucleotides, followed by binding of methylated DNA-binding proteins (MBDs). Some MBDs have binding sites for HDAC complexes, leading to local histone deacetylation and hence transcriptional silencing, and there is evidence that MBDs may promote chromatin compaction.

Global Secondary Structures – the 30-nm Chromatin Fiber

If chromatin is isolated from nuclei and observed by electron or atomic force microscopy, it is seen to have a conformation highly dependent on the ionic strength of the medium – at low ionic strength (5–25 mM Na^+), the linear array of nucleosomes is clearly seen, while at higher ionic strengths, increasing compaction occurs, leading to the formation of an irregular fiber ~ 30 nm in diameter. This reversible transformation, which is dependent on H1, can also be monitored by changes in physical properties such as sedimentation velocity. Much research effort has been expended on determining the physical organization of the 30-nm fiber, and several different models have been proposed. The most widely discussed are solenoidal structures in which the linker DNA continues the superhelix established in the nucleosome, creating a stack of coins which subsequently coils into a 30-nm diameter fiber, and zigzag structures in which the linker DNA remains more or less straight and compaction results from the accordion-like compression of the zigzag. Among recent results supporting a zigzag arrangement are computer models derived from force-extension curves of individual chromatin fibers (Figure 1), which bear a striking similarity to native chromatin fibers imaged in the hydrated form (Figure 2). The physical irregularity of native chromatin fibers, which probably results largely from the local variations in NRL, has contributed to the ongoing controversy. The ability to reconstitute artificial chromatin with regularly spaced nucleosomes has provided important insights into the structure of the 30-nm fiber. Using this

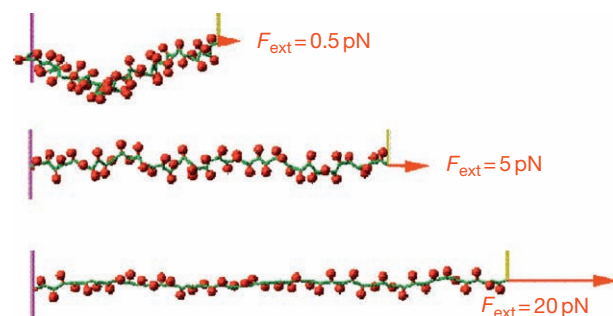


Figure 1 Computer models of chromatin secondary structure at different levels of extension (F_{ext}) based on force-extension data derived from pulling single chromatin fibers. Nucleosome core particles are represented by red spheres and linker DNA by green rods. Modified from Katritch V, Bustamante C, and Olson WK (2000) Pulling a single chromatin fiber: Computer simulations of direct physical manipulations. *Journal of Molecular Biology* 295: 29–40.

approach, it has been possible to crystallize and solve the X-ray structure of defined tetra-nucleosomes that lacked H1 but were compacted using high levels of divalent cations. These clearly adopt a zigzag conformation with straight linker DNA segments, such that nucleosome 1 associates with nucleosome 3 and nucleosome 2 with nucleosome 4. Other approaches using sophisticated computer modeling that incorporate thermodynamic and charge effects also suggest that long polynucleosomes adopt a primarily zigzag conformation. However, some models call for a subset of nucleosomes with bent linker as would be predicted for a solenoidal architecture. This type of heteromorphic organization has been observed by direct examination of defined polynucleosomes. *In vivo*, local variations in NRL, histone modifications, histone variants, and occupancy by NHCPs are likely to create a highly irregular chromatin fiber organization.

Local Secondary Structures

There is much evidence that local chromatin secondary structures play a large role in chromatin function by promoting or repressing transcription, but little is known of the precise physical organization of these local structures. In some instances, it has been demonstrated that a stretch of chromatin may accumulate many copies of NHCP that acts as a transcriptional silencer (e.g., HP1), but the difficulty of isolating single-gene chromatin for observation precludes rapid progress.

Tertiary and Higher-Level Features

Although it is often assumed that native chromatin exists as a series of hierarchical structures above the 30-nm chromatin fiber (and such illustrations are common in textbooks), there is very little solid supporting data. Methods used to examine isolated 30-nm fibers fail to reveal tertiary or higher levels of folding as the ionic strength is increased toward the physiological level. Rather, chromatin tends to aggregate. Examination of thin sections of nuclei in the electron microscope reveals many different fiber diameters ranging up to 240 nm, with no clear

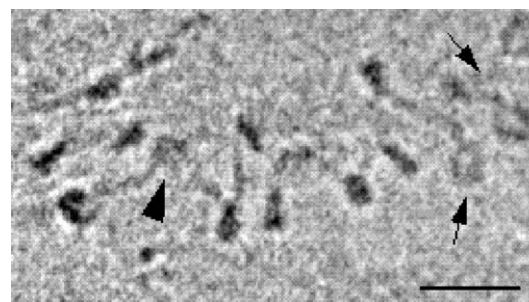


Figure 2 Chromatin fiber observed in the frozen hydrated state in 20-mM NaCl. Many of the nucleosomes are viewed edge-on and appear as dark ovals. Arrowheads point to two nucleosomes showing the circular enface profile. The zigzag secondary structure bears a strong resemblance to the model in Figure 1. Scale = 30 nm. From Bednar J, Horowitz RA, Grigoryev SA, et al. (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proceedings of the National Academy of Sciences of the United States of America* 95: 14173–14178.

hierarchical arrangement. Indeed, only under rather specialized conditions and with inactive nuclei can 30-nm fibers be routinely observed in sections. Whether this lack of order *in situ* is due to problems of adequate structural preservation, or simply due to a very unstructured and dynamic native organization remains to be seen.

Chromatin Organization *In Vivo*

Most of the work on chromatin physical organization discussed above has been conducted *in vitro* using chromatin isolated from nuclei or reconstituted from the individual components. However, it is important to verify that structures studied *in vitro* actually exist *in vivo*. For chromatin, such verification has been difficult. For example, sections of nuclei observed by EM mostly show dense masses of chromatin with no signs of individual 30-nm chromatin fibers. The few types of cells where 30-nm fibers can be seen in thin sections tend to be transcriptionally inert and have specialized H1-type histones and long NRLs (e.g., echinoderm sperm and avian erythrocytes). It has been suggested that the absence of a fiber structure even in frozen hydrated sections of unfixed nuclei and chromosomes is due to a tendency for nucleosomes in adjacent fibers to interdigitate, with the result that the chromatin behaves like a molten polymer. The high density of chromatin *in vivo* will favor this type of behavior. When removed from the nucleus, segments of chromatin are greatly diluted from their *in vivo* state, revealing the underlying 30-nm fiber organization.

Although *in vivo* studies have failed to shed much light on chromatin secondary structures, they have provided important insights into higher levels of organization through a number of related techniques collectively known as ‘chromatin conformation capture’ or 3C. The underlying strategy is to cross-link living nuclei with formaldehyde so as to capture segments of chromatin in close contact. Chromatin is then isolated and cleaved into small fragments, after which a desired subset containing the feature of interest, such as a histone variant, histone modification, and specific NHCP, is purified typically via immunoprecipitation. Finally, the cross-links are reversed, and the associated DNA sequences identified. To date, 3C and variants of the technique have revealed important features of chromatin architecture, including the suggestion that chromatin adopts a fractal organization *in vivo*. 3C methods are rapidly evolving, and as resolution increases through the use of massively parallel high-throughput screening, rapid progress can be anticipated.

Chromatin Dynamics

Recent work has revealed that living nuclei are remarkably dynamic. While the nucleosome core histones are long-lived

and stably incorporated, histone H1 binds to chromatin with a half-life of a few minutes, and most NHCPs have half-lives of a few seconds. Modeling of chromatin fibers based on thermodynamic principles indicates a constant fluctuation in local compaction. Also, individual clusters of genes may change dramatically in compaction state and even location in the nucleus when transcription is stimulated. These observations reinforce the concept of the living nucleus as a highly dynamic entity, both in terms of genetic activity and physical organization.

See also: [Molecular Biology: Nuclear Organization, Chromatin Structure, and Gene Silencing; Regulation of Chromatin Dynamics.](#)

Further Reading

- Allis CD and Jenuwein T (2001) Translating the histone code. *Science* 293: 1074–1080.
- Bednar J, Horowitz RA, Grigoryev SA, et al. (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proceedings of the National Academy of Sciences of the United States of America* 95: 14173–14178.
- Bustin M (2001) Chromatin folding and activation by HMGN chromosomal proteins. *Trends in Biochemical Sciences* 26: 431–438.
- Duan Z, Andronescu M, Schutz K, et al. (2010) A three-dimensional model of the yeast genome. *Nature* 465: 363–367.
- Eltsov M, MacLennan KL, Maeshima K, Frangakis AS, and Dubochet J (2008) Analysis of cryo-electron microscopy images does not support the existence of 30 nm fibers in mitotic chromosomes *in situ*. *Proceedings of the National Academy of Sciences of the United States of America* 105: 19732–19737.
- Grigoryev SA, Arya G, Correll S, Woodcock CL, and Schlick T (2009) Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proceedings of the National Academy of Sciences of the United States of America* 106: 13317–13322.
- Katritch V, Bustamante C, and Olson WK (2000) Pulling a single chromatin fiber: Computer simulations of direct physical manipulations. *Journal of Molecular Biology* 295: 29–40.
- Lieberman-Aiden E, van Berkum NL, Williams L, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326: 289–293.
- Misteli T (2001) Protein dynamics: Implications for nuclear structure and function. *Science* 291: 843–847.
- Schlach T, Duda S, Sargent DF, and Richmond TJ (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* 436: 138–141.
- Singh PB and Georgatos SD (2003) HP1: Facts, open questions, and speculation. *Journal of Structural Biology* 140: 10–16.
- Woodcock CL and Dimitrov S (2001) Higher-order structure of chromatin and chromosomes. *Current Opinion in Genetics and Development* 11: 130–135.
- Woodcock CL and Ghosh RP (2010) Chromatin higher-order structure and dynamics. *Cold Spring Harbor Perspectives in Biology* 2: a000596.

Chromosome Organization and Structure, Overview

E Gracheva, Medros Inc., St. Louis, MO, USA

S C R Elgin, Washington University, St. Louis, MO, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Chromatin The highly dynamic complex of DNA, histones, and nonhistone proteins from which the eukaryotic chromosome is built; it is generally used in reference to the interphase form and can refer to the sum of all chromosomes.

Chromosome Structural unit of the genetic material in eukaryotes. It consists of one double-stranded DNA molecule, packaged as interphase chromatin or a metaphase chromosome. Eukaryotes have

from 1 to over 200 chromosomes, depending on the species.

Nucleosome Fundamental repeating subunit of the chromatin fiber, consisting of a core histone octamer (made up of two molecules each of H2A, H2B, H3, and H4) with DNA wrapped around the surface in nearly two left-handed turns, connected by linker DNA associated with one molecule of linker histone (usually H1) to the next subunit. Each repeating unit (core plus linker) packages about 200 bp of DNA.

The Structure of Chromatin

The nucleotide sequence of DNA represents the primary level of chromosome organization. When a DNA molecule is packaged into chromatin, the latter does not behave as a uniform fiber. The chromosome conformation is affected by regulatory and structural features. These sequence elements include promoter proximal elements that control the activity of individual genes, locus control regions that impact a cluster of genes, and boundary elements that potentially regulate the activity of large (10–100 kb) chromosomal domains. In addition to the unique sequences that code for genes, eukaryotic genomes include a high percentage of repetitious DNA, both tandem arrays of short-repeating motifs (often identified as satellite DNA because of its different density) and dispersed relics of transposable elements, retroviruses, and the like. Some of the repetitious DNA is utilized in the generation of special chromosome structures such as the centromere, the site of microtubule attachment during mitosis, and the telomeres, structures that protect the ends of the DNA molecules.

The discovery of nucleosomes, repeating subunits based on DNA folding around an octamer of histones, provided a molecular description of the fundamental unit of chromatin folding and accounted for the first six- to sevenfold linear compaction of the DNA. The octamer, containing two molecules each of the core histones H2A, H2B, H3, and H4 (a tetramer of $[H3+H4]_2$ plus two dimers of $[H2A+H2B]$) with 147 bp of DNA wrapped around it in 1 2/3 left-handed turns, makes up the core particle. The high-resolution 2.8-Å structure reveals the details of histone organization within the nucleosome (Figure 1). The C-terminal two-thirds of the histone molecules, folded in a hand-shake motif, form the core of the nucleosome, while their amino-terminal tails extend outward and are potentially involved in further interactions with DNA, with other proteins, and between adjacent nucleosomes. Typically, 170 bp of DNA follows a bent path round the histone core, held strongly by electrostatic bonds between the basic amino acids of the histones and the phosphates of the DNA. The segments of DNA that link adjacent nucleosome core particles (linker DNA) are bound by linker

histones (H1 and its variants) or HMG proteins. These proteins contribute to the further folding of the chromatin fiber, playing a role in the condensation of the 10-nm fiber to form a 30-nm fiber (Figure 2).

The arrangement of nucleosomes within the 30-nm fiber is still very controversial. A widely utilized model, derived primarily from physical measurements and diffraction studies, is based on a solenoid, in which the nucleosomal fiber is wound in a regular helix, with about six nucleosomes per turn. The linker histones face inward and help to stabilize the solenoid. However, advanced single-molecule methods, using microscopy and spectroscopy to study single chromatin fibres, suggest that the chromatin fiber is an irregular helix, with a fluctuating zigzag secondary structure. Comparative studies suggest that the linker DNA with the associated linker histone affects the DNA topology and the higher-order folding and compaction of chromatin along the 30-nm fiber axis (Figure 2(b)).

The 30-nm fiber is thought to be the basic form of much of the chromatin during interphase. However, a subfraction of the chromatin remains condensed, and darkly staining (heteropycnotic), as the cell returns from metaphase to interphase. This condensed fraction is referred to as 'heterochromatin', while the more diffuse material is labeled 'euchromatin'.

Heterochromatic domains contain a high proportion of repetitious DNA sequences (satellite DNA, retrotransposons, and other transposable element relics), have a low gene density, are replicated late in S phase, and show little or no meiotic recombination. These domains are typically associated with methylation of histone H3 at lysine 9, with concomitant association of heterochromatin-specific proteins such as HP1. Genes normally resident in euchromatin become inactive when placed in or near a heterochromatic environment by rearrangement or transposition. The factors that define specific chromosomal domains as preferred sites of heterochromatin assembly are not well understood, but it appears likely that heterochromatin formation is linked to the presence of repeated DNA sequences. Heterochromatic regions show a more regular nucleosome array (more even spacing) than

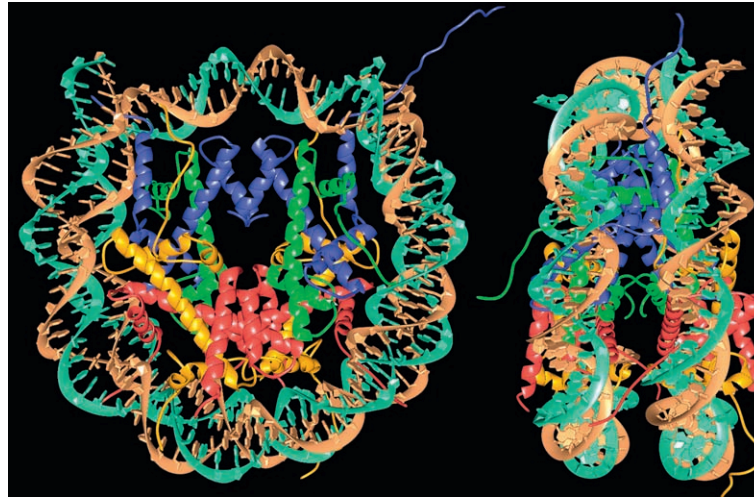


Figure 1 The structure of the nucleosome core particle. Brown and green ribbon traces represent the double-stranded 146-bp fragment of DNA. H2A molecules are shown in yellow, H2B in red, H3 in blue, and H4 in green. Reprinted from Luger K, Mader AW, Richmond RK, Sargent DF, and Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389: 251–260.

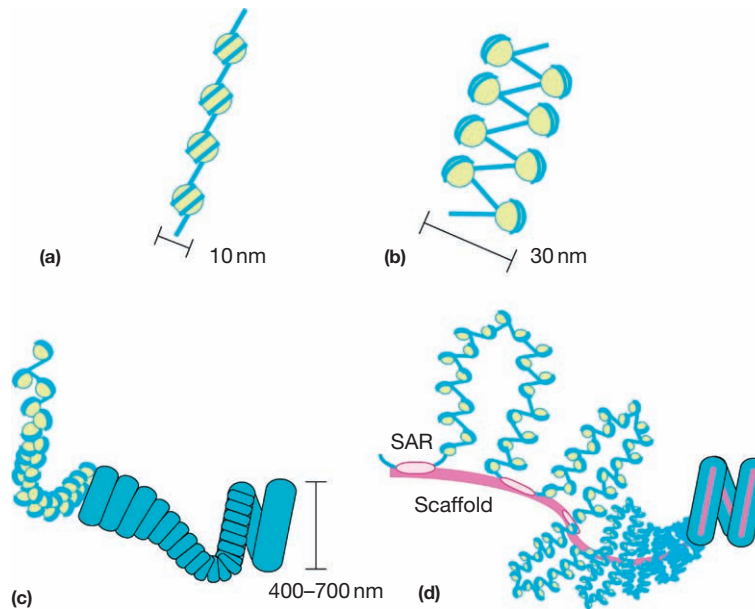


Figure 2 Hierarchical levels of chromatin organization. (a) Structure of the 10-nm nucleosome fiber. Histone octamers are shown as yellow circles, the blue line designates double-stranded DNA wrapped around an octamer. (b) Possible structure of the 30-nm chromatin fiber based on a three-dimensional zigzag arrangement. (c) Formation of higher-order chromatin structure by coiling into thicker (400–700 nm) chromatin fibers. (d) Organization of the chromatin into loops radiating out from a scaffold (magenta). Pink ovals represent SAR elements. Reprinted from Swedlow JR and Hirano T (2003) The making of the mitotic chromosome: Modern insights into classical questions. *Molecular Cell* 11(3): 557–569.

euchromatic regions, as well as a loss of nucleosome-free regulatory regions; however, the nature of the difference in higher-order packaging between heterochromatin and euchromatin remains unknown.

Chromatin structure is intimately involved in the regulation of chromosome activities: replication, transcription, and repair. The packaging of a DNA protein-binding motif with histones in a nucleosome often renders that signal invisible to the protein. For example, packaging of a TATA box sequence into a nucleosome blocks initiation of transcription; assembly of a restriction site

into a nucleosome blocks cleavage by that restriction enzyme. Thus, it is no surprise that chromatin needs to undergo significant structural change, or remodeling, to enable the cellular machines that perform various functions to gain access to their template, the double-helical DNA. The challenge to understand chromatin structure and dynamics, and to link chromatin structure to function, is enormous. Current research has highlighted the role of posttranslational modifications of both the histones and the DNA in marking different structural states and providing appropriate signals to the cellular machines.

The core histone proteins are highly conserved. In the most extreme example, there are only two amino acid changes between the sequences of histone H4 from plants and animals. While the globular carboxy-terminal domains make up the nucleosome core, the flexible amino-terminal tails protrude outward from the core structure. These basic tails function as acceptors for a variety of posttranslational modifications, including acetylation, methylation, and ubiquitination of lysine (K) residues, phosphorylation of serine (S) and threonine (T) residues, and methylation of arginine (R) residues. Several studies have shown that histone modifications can be cooperative or conflicting, one modification facilitating or blocking another. The combination of histone modifications present contributes to the chromatin status of small and large regions of the genome and mediates its functional activity. In particular, histone acetylation is associated with gene expression, while deacetylation, with methylation of histone H3 on lysines 9 and/or 27, is associated with silencing. These patterns can be altered by multiple extracellular and intracellular stimuli that signal to the nucleus. In fact, recent studies have shown that many proteins characterized as activators or silencers of transcription work by causing or facilitating a change in the modification state of the histones at a given locus. Thus, chromatin itself has been proposed to serve as a signalling platform and to function as a genomic integrator of various signaling pathways. Covalent modifications of histone tails are believed to act in concert to establish the 'histone code' essential for assembly of the desired chromatin state.

In addition to the proteins that package DNA, there are also sets of nonstructural proteins that interact with specific DNA sequences to control gene expression, interacting with promoters, enhancers, locus control regions, etc. While some such proteins can recognize DNA sequences in a nucleosome, many cannot. Both targeted alterations in histone modification and the remodeling of nucleosomes are often required to allow these latter proteins to function. The order of the required operations for activation of transcription varies at different loci.

Other nonhistone proteins contribute to structural elements such as insulators and/or boundary complexes. Insulators define a region of regulatory activity by managing promoter interactions with proximal and distant enhancers; while boundaries (or barriers) limit the spread of heterochromatin into a euchromatic domain. The arrangement of these genomic elements contributes to the functional and structural state of a locus and region. Unfortunately, it has not yet been possible to relate boundary elements, defined by their functional role described above, with the structural elements that anchor chromosomal loops, as seen by electron microscopy, or with the protein scaffold that constrains DNA in supercoiled domains of 10–100 kb. A better understanding of this level of organization will be essential to develop a detailed model of how chromosomes function.

The Role of Spatial Organization of the Genome in the Nucleus

The next level of chromosome organization must take into account the three-dimensional (3D) arrangement of the chromosomes within the cell nucleus. Approaches to microscopy

that allow us to examine the nucleus in 3D have shown that individual chromosomes occupy defined subdomains within the interphase nucleus. This organization involves the interaction of different factors bound to specific sites in the chromosomes with regions of the nuclear envelope and the protein-facilitated juxtaposition of chromosome sites to form stable connections between distant sequences, with intervening loops of the chromatin fiber. There is increasing evidence that both actin and myosin, found inside the nucleus, might have an impact on its internal structure. Noncoding RNAs have also been shown to affect higher-order chromatin organization. Analysis of the spatial distribution of chromatin within the nuclei of individual mammalian cells has revealed distinct variations between cells; these findings may help to explain the epigenetic diversity among genetically identical cells. Unfortunately, our understanding of this higher-level packaging is very meagre compared to our knowledge of the primary levels of organization. Nonetheless, such packaging can be thought of as the culmination of the hierarchy of chromatin structures that functionally connect the linear genomic sequence to the mechanisms for information processing inside the cell nucleus.

The Mitotic Chromosome

Substantial differences in the organization of chromosomes occur during mitosis. The most obvious difference is that the chromosome structure is more compact and the shape much better defined than in interphase chromosomes. This more stable structure has offered some advantages for examining the organization and packaging of the single DNA molecule of a chromatid. Several biochemical and structural studies have argued for the partitioning of the genome into a series of defined units (domains) containing 10–100 kb of DNA. Early EM studies suggested the presence of a chromosome scaffold from which loops of chromatin radiate. The scaffold might be considered as either an axial core, most likely made up of DNA and/or protein, or a loose network of DNA/protein complexes whose regulated assembly controls the higher-order structure of the chromosome. In general, scaffold models presume interactions between specific protein factors and *cis*-acting DNA sequences (SARs, or scaffold attachment regions) that determine the limits of the domains. In most models, two neighboring elements, separated by up to 100 kb of DNA, bind to the scaffold (or to each other) forming a loop. Experimentally, a group of very AT-rich sequences has been identified as potential SARs, based on the association of these sequences with the protein scaffold remaining after histones have been extracted from chromosomes. Using fluorescence *in situ* hybridization (FISH), SARs were visualized as a long queue running the length of the chromosome that alternately followed a fairly linear or tightly coiled path. These images were highly variable. Helically folded mitotic chromosomes have been visualized under certain conditions, suggesting that the highest order of folding is a helical winding of the chromosome arm.

However, *cis*-acting elements affecting chromosome architecture have not been identified by genetic criteria, nor have the proposed SARs (identified by biochemical assay) been mapped successfully in relation to cytological preparations. Recently, biophysical approaches for quantifying the elastic

properties of mitotic chromosomes have provided important insights into their structure. The mechanical integrity of mitotic chromosomes appears to be maintained by links between chromatin fibers, not by a proteinaceous axial core. Isolation of whole chromosomes and reconstitution of mitotic chromosomes have identified topoisomerase IIa, a five-subunit complex termed condensin, chromokinesin, and the chromatin remodeling ATPase ISWI as major constituents. All presumably contribute to the change from the interphase form to the condensed mitotic chromosome. ATPases are very abundant in mitotic chromosome fractions. Probably, these enzymes provide the energy for the aforementioned complexes to induce the local and global conformational changes in chromosomes required to support condensation, and microtubule-dependent movement of assembled chromosomes to opposite poles prior to cell division.

The centromere, observed in all higher eukaryote chromosomes, is a well-known landmark. It takes the form of a distinct primary constriction where the sister chromatids are held together until separated. The constricted region comprises a differentiated chromatin structure onto which microtubules bind to provide proper chromosome movements during the processes of mitosis and meiosis. The centromere in higher eukaryotes is generally made up of massive regions of repetitive sequences. Many observations suggest that there is no specific DNA sequence that is required for centromere formation, strongly implicating an epigenetic mechanism. Five mammalian CENP proteins (CENTromere Proteins) – CENP-A, CENP-B, CENP-C, CENP-G, and CENP-H – are unique and essential to the centromere. Three of these proteins – CENP-A, CENP-B, and CENP-C – have demonstrated DNA-binding activity and are potential candidates for marking the site. CENP-A is a histone H3-like protein that is conserved in mammals and other organisms; it is present only in the nucleosomes of active centromeric regions, where its presence triggers kinetochore formation. A mechanism for initiating centromere formation is not identified at present. Most of the chromatin properties of centromeres are not unique. For example, hypoacetylation of histones, a high level of DNA methylation and late replication, are also characteristic features of noncentromeric heterochromatin. The most attractive candidate for the centromeric mark is CENP-A, but the mechanism for incorporation of CENP-A into only one site per chromosome remains a puzzle.

The ends of eukaryotic chromosomes are organized into telomeres, nucleoprotein structures essential for maintaining the integrity of the genome, the so-called ‘guardians of the chromosome’. Telomeric DNA in the majority of eukaryotes is made up of tandem repeats of short guanine-rich sequences. In some cases (e.g., in *Drosophila*), a subset of transposable elements is maintained at the chromosome ends and guanine-rich repeats are not observed. The telomeres show a very special pattern of chromatin fiber organization, distinct from the rest of the genome. The G-rich sequences can form a quadruple helix. *In vivo*, telomeres are not necessarily linear, but can form looped foldback structures, stabilized by protein components. Many of the key proteins in the telomere have

been identified, although their functions are still incompletely understood. Interestingly, telomeres are not transcriptionally silent. Noncoding RNAs are being produced; it is very likely they are involved in functional and structural regulation of telomeres.

Summary

A chromosome, then, can be considered to be a single double helical DNA molecule, running from one telomere through the centromere to the other telomere. Multiple origins of replication, and many genes, are typically found distributed along the length of the DNA molecule. The DNA is packaged first in a nucleosome array, subsequently in a 30-nm fiber, and finally in higher-order structures. This packaging contributes to gene regulation both at the level of individual genes and at the level of large domains.

See also: **Molecular Biology:** Chromatin: Physical Organization; Nuclear Organization, Chromatin Structure, and Gene Silencing; Regulation of Chromatin Dynamics.

Further Reading

- Bloom K and Joglekar A (2010) Towards building a chromosome segregation machine. *Nature* 28: 446–456.
- Fischle W, Wang Y, and Allis CD (2003) Histone and chromatin cross-talk. *Current Opinion in Cell Biology* 15(2): 172–183.
- Goetze S, Mateos-Langerak J, and van Driel R (2007) Three-dimensional genome organization in interphase and its relation to genome function. *Seminars in Cell and Developmental Biology* 18(5): 707–714.
- Grewal SI and Elgin SC (2002) Heterochromatin: New possibilities for the inheritance of structure. *Current Opinion in Genetics and Development* 12(2): 178–187.
- Kruihof M, Chien F, Routh A, et al. (2009) Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nature Structural and Molecular Biology* 16: 534–540.
- Luger K, Mader AW, Richmond RK, Sargent DF, and Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389: 251–260.
- Maeda RK and Karch F (2007) Making connections: boundaries and insulators in *Drosophila*. *Current Opinion in Genetics and Development* 17(5): 394–399.
- Mellone BG and Allshire RC (2003) Stretching it: Putting the CEN(P-A) in centromere. *Current Opinion in Genetics and Development* 13(2): 191–198.
- Neidle S and Parkinson GN (2003) The structure of telomeric DNA. *Current Opinion in Structural Biology* 13(3): 275–283.
- O'Brien TP, Bult CJ, Cremer C, et al. Genome function and nuclear architecture: From gene expression to nanoscience. *Genome Research* 13(6A): 1029–1041.
- Schoeffner S and Blasco MA (2009) A ‘higher order’ of telomere regulation: Telomere heterochromatin and telomeric RNAs. *EMBO Journal* 19: 2323–2336.
- Sims RJ III and Reinberg D (2008) Is there a code embedded in proteins that is based on post-translational modifications? *Nature Reviews Molecular Cell Biology* 9(10): 815–820.
- Summer AT (2003) *Chromosomes: Organization and Function*. Malden, MA: Blackwell Publishing.
- Swedlow JR and Hirano T (2003) The making of the mitotic chromosome: Modern insights into classical questions. *Molecular Cell* 11(3): 557–569.
- Trojer P and Reinberg D (2007) Facultative heterochromatin: Is there a distinctive molecular signature? *Molecular Cell* 28(1): 1–13.
- West AG, Gaszner M, and Felsenfeld G (2002) Insulators: Many functions, many mechanisms. *Genes Development* 16(3): 271–288.
- Wu C, Bassett A, and Travers A (2007) A variable topology for the 30-nm chromatin fibre. *EMBO Report* 8(12): 1129–1134.

Coenzyme A

M D Lane, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Coenzyme A small molecule (molecular weight usually ≤ 1000) that participates chemically in a reaction catalyzed by an enzyme and upon which the reaction depends.

Electrophile A compound with an electron-deficient atom that can accept an electron pair from a nucleophilic compound with an electron-rich atom.

Nucleophile A compound with an electron-rich atom having an unshared electron pair that can be contributed to an electron-poor atom of an electrophilic compound.

Thioester The condensation product of a carboxylic acid and a thiol ($-\text{SH}$) group formed by removal of a water molecule, for example, $\text{R}-\text{CO}_2\text{H} + \text{Co}-\text{SH} \rightarrow \text{H}_2\text{O} + \text{R}-\text{CO}-\text{S}-\text{CoA}$.

Background and Discovery

Coenzyme A (CoA) was discovered by Fritz Lipmann and his colleagues in the early 1950s. The coenzyme was first isolated from large quantities of pig liver extract as the factor required for the acetylation of sulfanilamide, the assay system used to track CoA during its purification. The discovery of CoA and the characterization and determination of its structure (Figure 1) led Lipmann being awarded the 1953 Nobel prize in physiology or medicine. Lipmann's findings opened the door for the discovery of innumerable roles of CoA, most notably the discovery by Feodor Lynen that active acetate was acetyl-CoA, a key intermediate in the metabolism of carbon compounds by all organisms. In 1964, Lynen was awarded the Nobel prize in physiology or medicine for his discovery of acetyl-CoA and many of the metabolic systems that CoA functions. We now know that CoA plays a key role in carbohydrate, lipid, and amino acid metabolism.

Enzymatic Functions of CoA and 4'-Phosphopantetheine

Chemical Mechanisms of Action

CoA occurs in cells either as free CoA or in the form of its acyl-S-CoA thioester or as 4'-phosphopantetheine (4'-PP), which is part of a multifunctional enzyme. The sulfhydryl ($-\text{SH}$) group of the coenzyme is the functional site that endows it with its unique chemical/enzymatic properties. In all of the reactions in which it participates, CoA serves to activate carboxyl groups ($\text{R}-\text{CO}-\text{OH}$) in the form of an acyl-CoA thioester derivative ($\text{R}-\text{CO}-\text{S}-\text{CoA}$). In this form the acyl group has favorable energetic and mechanistic properties. Energetically, CoA thioesters have high transfer potential (i.e., a high negative free energy change associated with breaking the C-S bond of the thioester) that facilitates driving the acyl transfer reaction in which it participates toward completion. Mechanistically, the chemistry of CoA thioesters facilitates reactions either at the carbonyl carbon ($-\text{CH}_2-\text{*CO}-\text{S}-\text{CoA}$) or the adjacent α -carbon atom ($-\text{*CH}_2-\text{CO}-\text{S}-\text{CoA}$).

Activation of the carbonyl- and α -carbon atoms in acyl-CoA thioesters

All enzymatic reactions of CoA thioesters in metabolism involve activation either of the carbonyl-carbon or the adjacent-carbon atom for facile chemical reaction. Mechanistically, the $-\text{CO}$ -carbon (i.e., $\text{R}-\text{CH}_2-\text{CO}-\text{S}-\text{CoA}$) of the acyl group reacts with a nucleophile (R^- ; see Figure 2, reaction 1) or the adjacent alpha carbon (i.e., $\text{R}-\text{*CH}_2-\text{CO}-\text{S}-\text{CoA}$) reacts with an electrophile (R^+ ; see Figure 2, reaction 2) to produce a new carbon-carbon bond. These reaction types are illustrated in the figure.

CoA as Donor of 4'-PP for Fatty Acid Synthase

Fatty acid synthase (FAS) is a large multifunctional enzyme that catalyzes all steps in fatty acid synthesis, catalysis being facilitated by a covalently linked 4'-PP prosthetic group. The 4'-PP group, which is derived from CoA (Figure 1), is enzymatically transferred to the FAS apoenzyme where it becomes covalently linked to the enzyme. The 4'-PP prosthetic group acts as a long sidearm to which intermediates of the pathway are covalently linked. The 4'-PP sidearm allows translocation of the growing fatty acyl chain intermediate from one catalytic site on FAS complex to the next in the cyclic sequence, which leads to formation of a long-chain fatty acid.

Enzymatic Synthesis of Acyl-S-CoA Derivatives from Free Carboxylic Acids and CoA-SH

The enzymatic reactions by which acyl-S-CoA derivative is synthesized from carboxylic acids, such as free fatty acids, occur through a common mechanism (Figure 3).

The reaction illustrated uses fatty acid as example; however, it should be noted that virtually all acyl-CoAs are formed from their cognate carboxylic acids by the same mechanism. The enzymes that catalyze these reactions are referred to as 'acyl-CoA synthetases'. In the initial step (Figure 3, reaction 1), the fatty acid attacks the (innermost) phosphate of ATP-releasing pyrophosphate (PP_i) forming an enzyme-bound acyl-AMP intermediate. In the second step (Figure 3, reaction 2), CoA-SH

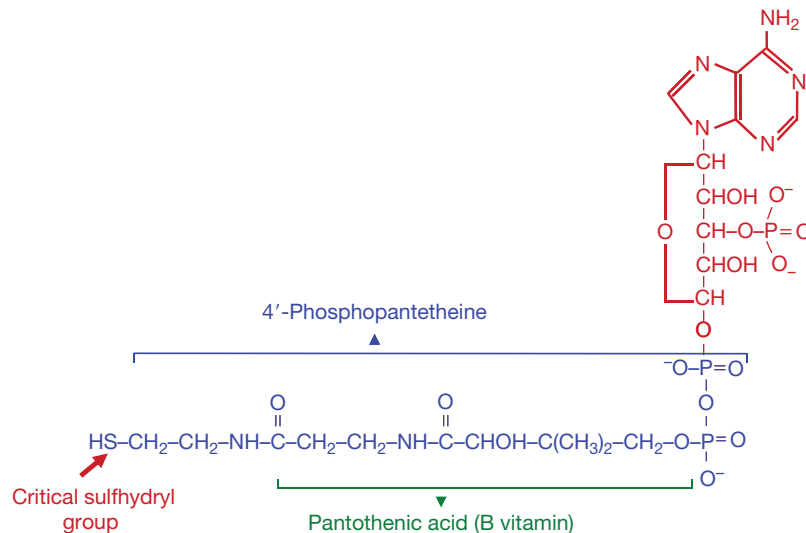


Figure 1 Structure of CoA.

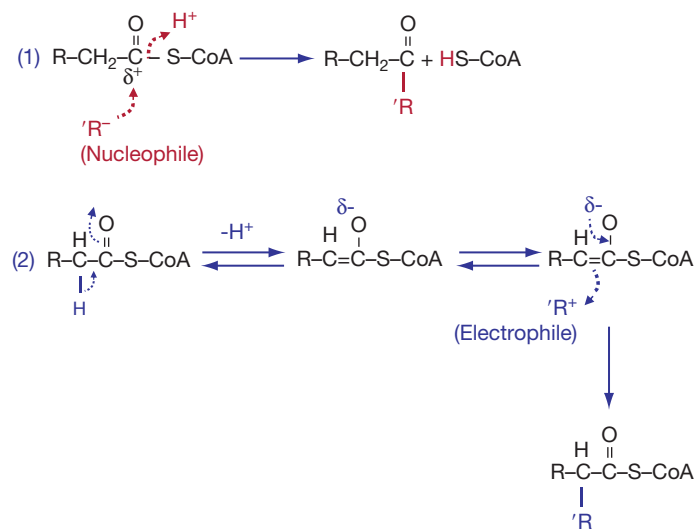


Figure 2 Mechanisms of activation of the acyl carbonyl- and α -carbon atoms in CoA thioesters.

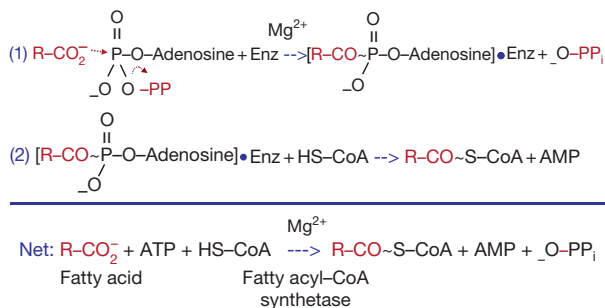


Figure 3 Enzymatic synthesis of acyl-S-CoAs.

attacks the acyl-CO-carbon of the intermediate producing an acyl-S-CoA product and AMP.

Metabolic Roles/Functions

CoA and its derivatives function in a wide variety of metabolic pathways, including the tricarboxylic cycle, fatty acid oxidation, ketogenesis, fatty acid biosynthesis, sterol synthesis, complex lipid synthesis, amino acid metabolism, and porphyrin synthesis. (Since most of these are covered elsewhere in this encyclopedia, they are not dealt with in this article.) As FAS

system is unique, in that it makes use of only a fragment of the CoA molecule, 4'-PP, a brief discussion of FAS is warranted.

Fatty Acid Synthase

Fatty acid synthesis in animals is catalyzed by a single large (molecular weight, 5×10^5 , a dimer consisting of two identical $\sim 2.5 \times 10^5$ subunits) multifunctional enzyme. All eight steps and thus all eight catalytic centers that carry out fatty acid synthesis occur with the intermediates tethered to the FAS. The intermediates are covalently linked by thioester bonds to the -SH group of the long 4'-PP sidearm which facilitates translocation of intermediates from one catalytic center to the next in sequence until the multiple steps of long-chain fatty acid synthesis are completed. Each round of elongation lengthens the chain by two carbons, a process that is repeated 7 or 8 times for the synthesis of a 16- or 18-carbon-containing fatty acids.

The process is initiated by the transfer of an acetyl group to 4'-PP from acetyl-S-CoA. The acetyl group linked to 4'-PP serves as the primer onto which the long-chain fatty acid is built. Malonyl units from malonyl-CoA, which serve as the chain-elongating group, condenses with the acetyl-primer concomitant with decarboxylation to produce a four-carbon

intermediate that then undergoes two reductive and one dehydration steps. Successive malonyl groups are transferred to FAS from malonyl-CoA to provide the basic units for successive steps in the elongation process. The terminal step is catalyzed by a thioesterase releasing the long-chain fatty acid product.

See also: [Metabolism Vitamins and Hormones: Fatty Acid Structure and Synthesis.](#)

Further Reading

- Berg J, Tymoczko JL, and Stryer L (2006) *Biochemistry*, 6th edn. New York, NY: WH Freeman.
- Lipmann F (1953) Development of the acetylation problem: A personal account, Nobel Prize Lecture (in physiology or medicine). <http://www.nobel.se> (accessed April 2010).
- Lynen F (1964) The pathway from "activated acetic acid" to terpenes and fatty acids. Nobel Prize Lecture (in physiology or medicine). <http://www.nobel.se> (accessed April 2010).
- Nelson DL and Cox MM (2005) *Lehninger: Principles of Biochemistry*, 4th edn. New York, NY: WH Freeman.
- Walsh C (1979) *Enzymatic Reaction Mechanisms*. San Francisco, CA: WH Freeman.

Collagenases

K Holmbeck and H Birkedal-Hansen, National Institutes of Health, Bethesda, MD, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 478–481, © 2004, Elsevier Inc.

Glossary

Cathepsins Lysosomal proteinases, mostly thiol-dependent proteinases of the papain family. They are present in multiple forms of which cathepsin K seems to be the most potent collagen-cleaving enzyme.

Collagen Major structural proteins of the body with type I being the most abundant protein in soft connective tissue and bone. The major fibrillar collagens (types I, II, and III) provide much of the structural framework for soft connective tissues, cartilage, and bone. They are generally highly resistant to proteolysis.

Collagenase Enzyme capable of cleaving native collagen under specified conditions (temperature and pH).

Knockout mouse Mouse rendered deficient for a selected gene by targeted inactivation of the encoding locus in embryonic stem cells, subsequent transmission of the trait to the germ line and breeding of the trait to a homozygous state.

Matrix metalloproteinases (MMPs) A group of zinc endopeptidases (metzincins) characterized by zinc-binding VAAHEXGHXXGXXH amino acid consensus sequence in the catalytic domain and activation by a so-called cysteine switch. This residue is harbored in the latency-conferring prodomain that is proteolytically cleaved upon activation. This group of enzymes does not include a disintegrin and metalloproteinases (ADAMs), which contains a characteristic disintegrin domain.

What Is a Collagenase?

The term collagenase was first used by Gross and Lapiere in their 1962 description of an activity released from cultured tissue explants of involuting tadpoles, which dissolved otherwise highly proteinase-resistant, reconstituted collagen fibrils. This was a groundbreaking finding. It was the first demonstration of the existence of an activity operating at neutral pH that was fully capable of dissolving and cleaving native collagen fibrils. It marked the beginning of an exciting period of discovery that led to the identification of a large number of distinct matrix metalloproteinases (MMPs), almost six to seven in number with collagen-cleaving capabilities under various *in vitro* conditions (Table 1).

Original Definition of Collagenases

The unique resistance of native collagen molecules to proteolysis is a result of the tightly wound, semirigid triple helical structure. However, the resistance is only relative and not absolute. Generally, one of the two tests was applied to ascertain whether an enzyme or an activity represented a true collagenase: (1) ability to dissolve reconstituted type I collagen fibrils in the temperature range 35–37 °C and/or (2) ability to cleave type I collagen in solution at neutral pH in the temperature range 15–25 °C into two distinct and recognizable 1/4–3/4 fragments (classical collagenase signature cleavage pattern). That the 3/4–1/4 cleavage site is actually utilized *in vivo* was elegantly shown by Robin Poole and co-workers by neo-epitope immunohistochemistry which convincingly identified collagen molecules in several tissues, which were cleaved at this particular site. Since a number of collagenases cleave this site, the mere detection of this neo-epitope does not necessarily identify any single enzyme responsible for the cleavage. Despite this, the 3/4–1/4 cleavage

products are highly useful hallmarks in the analysis and characterization of collagenase activity.

Several lines of evidence support the notion that the collagenase-sensitive cleavage site is subject to local unwinding and therefore constitutes a site of minor resistance. For example, type III collagen has an arginine substitution in the susceptible region, and is readily cleaved by trypsin producing a 3/4–1/4 pattern. However, the collagenase-sensitive site in the $\alpha 1$ chain of type I collagen can be mutated in mice without yielding an acute collagen indigestion phenotype. The modification leads to a progressive fibrosis, but at a rate indicating that the collagen molecule is cleaved with some efficiency elsewhere. The term collagenase based on site specificity may thus be only partially applicable and may in effect obscure the fact that collagen is cleaved at other sites by collagenases, which do not qualify for this designation based on the criteria of generating 3/4–1/4-specific cleavage products.

Revised Definition of Collagenases

Based on the observations outlined above, the original definition of a collagenase will therefore have to be revised and broadened to also accommodate other proteinases such as cathepsin K, which clearly is a potent collagen-cleaving enzyme and yet utilizes other cleavage sites in the triple helical region and functions at acidic pH. The ability to cleave native type I collagen in solution within the temperature range 15–25 °C at either neutral or acidic pH, or completely dissolve native or reconstituted collagen fibril matrices at 35–37 °C, still appears to be a valid requirement. It is now increasingly being recognized that type I collagen-cleaving ability *in vitro* does not necessarily mean that the enzyme in reality functions as a collagenase *in vivo*. In fact, the assay conditions ordinarily used to assess the collagenolytic activity of putative

Table 1 Fibrillar collagenases (proteinases with activity against fibrillar collagens)

Enzyme	Other name	Classification	Species
MMP-1	Collagenase-1, interstitial collagenase	Matrix metalloproteinase	Widely expressed except in rodents
McolA		Matrix metalloproteinase	Mice
MMP-2	Gelatinase, type IV collagenase	Matrix metalloproteinase	Widely expressed
MMP-8	Collagenase-2, PMN collagenase	Matrix metalloproteinase	Widely expressed
MMP-13	Collagenase-3, interstitial collagenase	Matrix metalloproteinase	Widely expressed
MMP-14	MT1-MMP	Matrix metalloproteinase	Widely expressed
MMP-18	Collagenase-4	Matrix metalloproteinase	Frog
Cathepsin K		Acidic thiol proteinase	Widely expressed

collagenases may rarely, if ever, entirely mimic the conformation and configuration of the substrate when it is associated with multiple other components of the extracellular matrix (ECM) and packed to densities rarely achieved in a cell-free system. Conversely, an argument may also be made that enzymes displaying little collagenolytic activity *in vitro* may conceivably be potent collagenases *in vivo*, if by association with other ECM macromolecules, the collagen molecules are made more readily available for cleavage. Standard enzyme kinetic assays measuring substrate conversion are often ill-suited to account for reactions, which take place in a two-dimensional, diffusion-restricted environment such as a cell-membrane or a solid-substrate surface. For collagenolytic enzymes with trans-membrane spanning domains, a solution-based assay can obviously change the basic environmental properties under which a molecule ordinarily works, and molecules commonly recognized as secreted exert their function in close proximity to the peri-cellular (solid phase) environment. In addition, the actual enzyme-to-substrate ratio *in vitro* may not adequately mirror that which exists *in vivo*. It may therefore be necessary to revise the definition collagenase and assign the term only to enzymes for which there is some evidence, for instance, from naturally occurring mutations or gene ablation studies, that they truly function to cleave collagen *in vivo*.

The Collagenases

Within a few years of the discovery of tadpole enzyme, similar activities were identified in many vertebrate species including humans. Among these were an enzyme from polymorphonuclear neutrophil (PMN) leukocytes that later turned out to be a genetically distinct collagenase (MMP-8). Gelatinase A/type IV collagenase (MMP-2) was also identified early on and recognized as a probably distinct and separate activity because of its ability to cleave type IV collagen and gelatin. However, it was not recognized until Aimes and Quigley's important study in 1995 that MMP-2 can also be a genuine type I collagen-specific collagenase once the inhibitor TIMP-2, with which it is often complexed, is removed. This finding raises the question whether, and under what conditions, TIMP-2-free proMMP-2 exists *in vivo*? MMP-13 or collagenase-3 is found in both humans and rodents and displays partial homology to the human interstitial collagenase, MMP-1. It was thus originally thought to be the rodent ortholog of MMP-1, but turned out to be a genetically distinct collagenase capable of also dissolving the

otherwise highly resistant and slowly cleaved type II collagen. Surprisingly, rodents appear to express MMP-13 in most places where MMP-1 is expressed in humans, thus suggesting that it is a functional equivalent. Yet a closer rodent ortholog of human MMP-1 called Mcol-A has been identified, but so far has been characterized only to some extent. Based on the initial analysis, Mcol-A appears to have some collagen-cleaving ability against type I collagen although the specific activity seems fairly low. A collagenase (MMP-18), which was distinct from MMP-1, MMP-13, and MMP-8, was discovered in the frog but no vertebrate ortholog has been identified. The first membrane-type MMP (MT1-MMP or MMP-14) was discovered in 1994 and, shortly after, it was shown to have collagen-cleaving activity at least *in vitro*. Later five additional membrane-bound MMPs were identified but it is not yet known whether they possess collagen-cleaving activity. Although most of the collagenases mentioned so far have been MMPs, it is, as previously mentioned, necessary today to include also the thiol cathepsin, cathepsin K. This potent collagen-cleaving enzyme is utilized in the phagocytic pathway of collagen degradation and in osteoclastic bone resorption, when mineralized type I or type II matrices are dissolved. Various collagenases display somewhat varying substrate preference against type I, type II, and type III collagens. Type I collagen is most readily cleaved by MMP-8 and MMP-14, while MMP-1 shows preference for type III collagen and MMP-13 for type II collagen.

Recent New Insights

The past decade has dramatically altered our view of the role of classical collagenases in normal growth and development. The advent of gene ablation techniques in combination with various mouse models has enabled detailed exploration of the biologic function of the individual collagenases *in vivo*. Unfortunately, MMP-1 cannot be subjected to this approach because it is not expressed in mice and no human mutations of MMP-1 have been characterized. Although a potent collagenolytic enzyme *in vitro*, the true role of MMP-1 *in vivo* may therefore not be known until a human null-mutation is discovered. No reports have emerged yet with regard to ablation of the weakly collagenolytic, putative murine MMP-1 ortholog, Mcol-A. MMP-2 has been knocked out, but the phenotype is very mild when considering that this enzyme is perhaps the most widely expressed MMP. It was subsequently shown that primary tumor mass and metastatic spread were significantly reduced

in MMP-2-deficient animals in two independent tumor models, but it is uncertain whether and/or how these observations relate to collagen degradation. So, at this point there is only limited evidence that MMP-2 indeed serves as a collagenase in mice. Adding confusion to this scenario, however, is the finding by Martignetti and co-workers of the first human MMP-mutation (in MMP-2), which yields a rather spectacular generalized osteolytic bone phenotype. It is dramatically different from that of the mouse but it does have features of a collagen indigestion phenotype. Recently, ablation of the PMN collagenase MMP-8 was reported. Interestingly, MMP-8 deficiency, particularly in male mice, resulted in a dramatic rise in 7,12-dimethylbenz(a) anthracene/12-O-tetradecanoylphorbol-13-acetate-induced skin tumors, diminished inflammatory response, and chemokine processing, suggesting that an important protective role, conferred by MMP-8, had been lost. It is so far unresolved whether this effect is a function of the loss of collagenolytic activity by PMNs. An MMP-13-deficient mouse has been generated which shows widening of the growth plate presumably as a result of impairment of cleavage of collagen type II in this area. Moreover, osteoclast differentiation and activity are impaired (SM Krane, personal communication) and bone formation is increased. The phenotype is thus consistent with MMP-13 serving as a collagen-cleaving enzyme *in vivo*. MT1-MMP (MMP-14)-deficient mice display a profound and complex phenotype in which growth, development, and maintenance of multiple collagen-containing matrices (bone, skin, tendon, joints, and ligaments) are severely impaired due to the loss of an indispensable and uncompensated collagenolytic activity. Since the MT1-MMP knockout phenotype does not resemble those of either MMP-2 or MMP-13, it is highly unlikely that MT1-MMP functions primarily through the activation of these enzymes as has been suggested. Rather, MT1-MMP appears to serve as a highly effective collagenase in its own right *in vivo*. However, it is interesting to note that MT1-MMP displays somewhat weaker collagenolytic activity *in vitro* than, for instance, MMP-1. This finding again points to some level of disconnect between collagenolytic capacity *in vivo* and *in vitro*. In the case of cathepsin K deficiency, the predictions of collagenolytic impairment, however, do hold up *in vivo*. Mouse and human cathepsin K gene mutations give rise to pycnodysostosis in humans and osteopetrosis in the mouse consistent with impairment of osteoclast-mediated type I collagen cleavage. These findings help identify cathepsin K as the acidic collagenase required for the removal of type I and type II collagens from mineralized matrices, which can only be dissolved in an acidic environment. It is highly likely that cathepsin K also is responsible for the cleavage of soft tissue collagen fibrils in phago-lysosomes during the often-overlooked phagocytic pathway.

Conclusion

In the aggregate, recent findings have demonstrated that true collagenases as defined by their substrate specificity *in vivo* can be found in at least two different classes of proteolytic enzymes (MMPs and cathepsins). Moreover, it is now evident that not all predictions based on enzyme kinetics obtained *in vitro* hold up *in vivo*. Collagenases work in the complex environment of an ECM. Fortunately, the tools for characterization of enzyme substrate interactions have now entered an exciting new era, which offer multiple genetic approaches to elucidation of these questions. Based on such approaches we have already come to realize that seemingly identical enzymes can perform different roles depending on the species. With this knowledge in mind, the understanding of and search for the collagen-metabolizing enzymes should continue.

See also: [Cell Architecture and Function: Metalloproteinases, Matrix; Protein/Enzyme Structure Function and Degradation: Collagens.](#)

Further Reading

- Balbin M, Fueyo A, Tester AM, et al. (2003) Loss of collagenase-2 confers increased skin tumor susceptibility to male mice. *Nature Genetics* 35(3): 252–257.
- Billinghurst RC, Dahlberg L, Ionescu M, et al. (1997) Enhanced cleavage of type II collagen by collagenases in osteoarthritic articular cartilage. *Journal of Clinical Investigation* 99: 1534–1545.
- Brinckerhoff CE and Matrisian LM (2002) Matrix metalloproteinases: A tail of a frog that became a prince. *Nature Reviews Molecular Cell Biology* 3: 207–214.
- Egeblad M and Werb Z (2002) New functions for the matrix metalloproteinases in cancer progression. *Nature Reviews Cancer* 2: 161–174.
- Everts V, Hou WS, Rialland X, et al. (2003) Cathepsin K deficiency in pycnodysostosis results in accumulation of non-digested phagocytosed collagen in fibroblasts. *Calcified Tissue International* 73: 380–386.
- Garnero P, Borel O, Byrjalsen I, et al. (1998) The collagenolytic activity of cathepsin K is unique among mammalian proteinases. *Journal of Biological Chemistry* 273(48): 32347–32352.
- Hotary KB, Allen ED, Brooks PC, et al. (2003) Membrane type I matrix metalloproteinase usurps tumor growth control imposed by the three-dimensional extracellular matrix. *Cell* 114: 33–45.
- Krane SM, Byrne MH, Lemaitre V, et al. (1996) Different collagenase gene products have different roles in degradation of type I collagen. *Journal of Biological Chemistry* 271(45): 28509–28515.
- Liu X, Wu H, Byrne M, et al. (1995) A targeted mutation at the known collagenase cleavage site in mouse type I collagen impairs tissue remodeling. *Journal of Cell Biology* 130(1): 227–237.
- Saftig P, Hunziker E, Wehmeyer O, et al. (1998) Impaired osteoclastic bone resorption leads to osteopetrosis in cathepsin-K-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America* 95: 13453–13458.
- Woessner J and Nagase FH (eds.) (2000) *Matrix Metalloproteinases and TIMPs*. Oxford: Oxford University Press.
- Zucker S and Chen W (eds.) (2003) *Cell Surface Proteases*. San Diego, CA: Academic Press.

Collagens

D J Prockop, Tulane University Health Sciences Center, New Orleans, LA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 482–487, © 2004, Elsevier Inc.

Glossary

Collagen fibrils Thin structures, 40–500 nm in diameter. Fibrils are assembled laterally into much larger fibers in tissues such as tendons.

Collagen monomer A protein containing a collagen triple helix.

Collagen triple helix A unique protein structure in which each of three chains with a (Gly–XY–)_n sequence is

coiled in a left-handed helix and the three chains are wrapped around each other in a more extended right-handed helix.

Procollagens The soluble precursors of fibrillar collagens that contain large, noncollagenous extensions of both ends of the long triple-helical domains.

The Signature (Gly–XY–)_n Sequence of Collagens

In the signature amino acid sequence of collagens, every third amino acid is glycine, the amino acid following glycine is usually the five-member ring amino acid proline, and the next amino acid is hydroxyproline. Therefore, the signature sequence is usually depicted as (Gly–XY)_n in which the X position is frequently proline and the Y position is frequently hydroxyproline. Under physiological conditions, three chains with the signature (Gly–XY–)_n sequence spontaneously associate into a three-stranded helix that is unique to collagens. The subclass of collagens that form large fibrils consists almost entirely of long (Gly–XY–)_n sequences tightly coiled into triple-helical structures referred to as collagen monomers. The long triple-helical monomers, in turn, spontaneously associate into long, highly ordered fibrils. In some tissues, the long fibrils associate laterally into thick fibers. The two steps in which (Gly–XY–)_n sequences assemble into triple-helical monomers and the triple-helical monomers assemble into fibers resemble the crystallization of small molecules (Figures 1 and 2). Each step is entropy driven by loss of surface water as the larger structure is formed, much as is seen in many crystallization processes. Also, as with crystallizations, each step involves first the slow formation of a small nucleus of a defined structure, and then rapid propagation of the nucleus to form a large structure in which the molecular architecture of the nucleus is extensively replicated. Glycine, the smallest amino acid, must be in every third position in the collagen triple helix because it occupies a very small space where the three chains come together. In addition, the hydrogens on the α -carbon of glycine allow it to form hydrogen bonds across the three chains. The closed-ring structures of the proline and hydroxyproline residues limit rotation of the polypeptide backbone, thereby providing rigidity to the structure.

Hydroxyproline residues, which are rarely found in other proteins, play a special role in the triple helix. The hydroxyl groups lock the slightly flexible ring of the amino acid into a conformation that further stabilizes the polypeptide backbone. Charged amino acids that occupy the X and Y positions in the repetitive (Gly–XY–)_n sequences point out from the center of

the triple helix. The orientation of the charged groups allows them to form salt bridges that bind each triple-helical monomer to surrounding monomers. Most of the monomers in fibrils are staggered one-quarter of their length relative to their nearest neighbors with short gaps between their ends. The gaps are readily seen by staining the surface of the fibrils and examining them by electron microscopy: each fibril has a cross-striated pattern in which the distances between the cross-striations correspond to about one-quarter of the lengths of the monomers. After assembly of the fibrils, the structures are further stabilized by a cross-linking enzyme that generates highly reactive aldehydes by deaminating some of the terminal amino groups on lysine residues that project out from the surface of the monomers. The aldehydes then spontaneously form covalent bonds that link most of the monomers throughout the width and length of the fibril. The result is a highly ordered structure with about the same strength as a steel wire. In essence, the signature (Gly–XY–)_n sequences contain essentially all of the information and driving forces needed to spontaneously generate the highly ordered fibrils and fibers that in some tissues can be several feet long.

The Two Subtypes: The Fibrillar and Nonfibrillar Collagens

The fibril-forming or fibrillar collagens contain long and uninterrupted (Gly–XY–)_n sequences (Figure 1(a)). They are among the most abundant proteins in nature, and the most abundant of the fibrillar collagens are referred to as types I, II, and III. They are similar in structure in that each contains three polypeptide chains (α -chains) with approximately 330 –Gly–XY– triplets. They are also similar in that the α -chains spontaneously self-assemble into similar triple-helical monomers and the monomers spontaneously assemble into fibrils. Type I collagen accounts for well over 50% of the dry weight of ligaments, tendons, and demineralized bone. However, the biological properties of the fibrils depend on the still-undefined forces that orient the fibers in tissues. For example, the type I collagen fibrils in ligaments and tendons are present in parallel

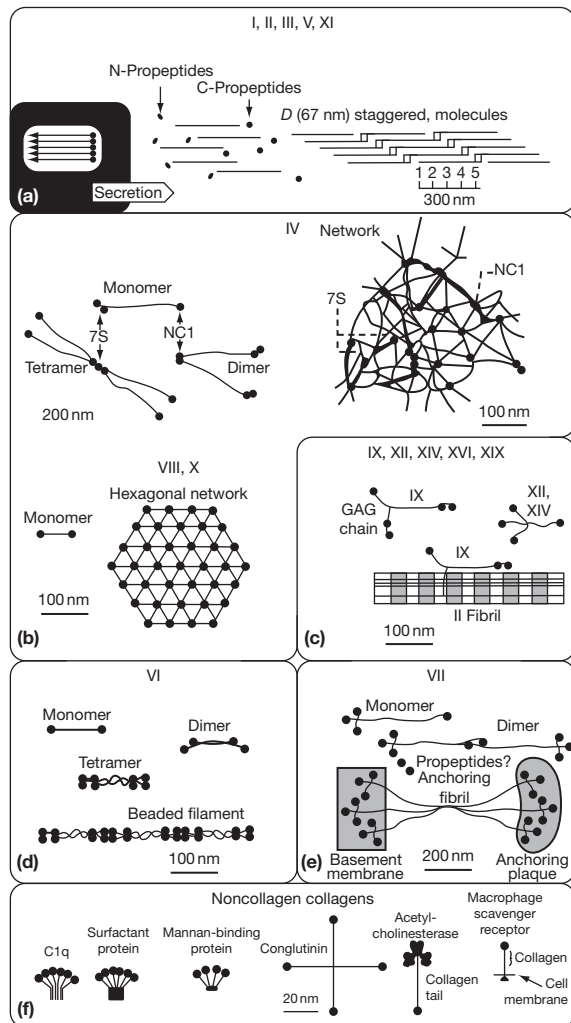


Figure 1 (a–f) Schematic for the structure of various collagens. Reproduced from Prockop DJ and Kivirikko KI (1995) Collagens: Molecular biology, diseases, and potentials for therapy. *Annual Review of Biochemistry* 64: 403–434.

bundles of thick fibers. In skin, they are randomly oriented in the plane of the skin. In many regions of bone, type I collagen fibrils form highly symmetrical concentric helical arrays around the Haversian canal through which capillaries pass. Type II collagen is very similar to type I collagen in structure, but it is found almost exclusively in cartilage, where it forms very thin fibrils that trap highly charged proteoglycans and water. The result is an arcade-like structure that makes cartilage highly resilient to compression. Type III collagen is also similar to type I collagen in structure and its spontaneous assembly into fibrils. Type III collagen is associated in relatively small amounts with type I collagen in many tissues, but, for reasons that are not apparent, it is particularly abundant in the aorta.

The family of collagen proteins also includes a large number of proteins referred to as nonfibrillar collagens because they contain either only short (Gly-XY)_n sequences or (Gly-XY)_n sequences that are interrupted by noncollagen sequences. As a result, the proteins do not assemble into highly ordered fibrils

(Figures 1(b)–1(f)). One important example of a nonfibrillar collagen is the type IV collagen, which is an abundant constituent of basement membranes (Figure 1(b)). The protein is a large trimeric monomer with multiple (Gly-XY)_n sequences that are interrupted by noncollagen sequences; it also contains large globular domains at both ends. The interruptions make the protein flexible, and the globular ends enable the protein to form flexible networks that bind multiple other proteins. The flexible networks bind epithelial cells and form the barriers that protect the organism from the external environment. Other nonfibrillar collagens are much less abundant and fulfill more specialized functions. Some bind to the surfaces of fibers formed by fibrillar collagens and help either to regulate the assembly process or to provide binding sites for cells or other components of the extracellular matrix. For example, type IX is a short collagen that is flexible because its (Gly-XY)_n sequences are interrupted by two noncollagenous sequences (Figure 1(c)). Type IX collagen binds to both fibrils of type II collagen and proteoglycans to provide one of the connecting links in the arcade-like structure of cartilage. Some nonfibrillar collagens contain very short (Gly-XY)_n sequences and triple-helical domains that serve primarily as extension rods for more interactive domains on the ends of the same proteins (Figure 1(f)).

Collagens and collagen-like proteins are abundant in many multicellular organisms. For example, they are very abundant in diverse invertebrates such as sponges and worms, including the worms found in deep-sea hydrothermal vents. The smallest collagens appear to be minicollagens that contain only 12–16 – Gly-XY– repeats, and that are found in the nematocysts with which hydra sting their prey.

The Need to Control the Self-Assembly of (Gly-XY)_n Sequences

Viewed with evolution in mind, the special properties of (Gly-XY)_n sequences can be seen to have a variety of different purposes. One of the most important of these purposes was to produce tough fibrils to hold together cells and thereby provide a means of developing multicellular organisms. However, it is clear that there had to be controls on the tendency of long polypeptides with repetitive (Gly-XY)_n sequences to spontaneously assemble into large fibers. In living systems, the assembly of fibrils cannot be allowed to occur prematurely. If they did, the fibrils would either destroy the cell synthesizing the protein or encase the cell in an impenetrable matrix. Also, the fibrils must be strong but not as rigid as crystals. Their flexibility is critical for most tissues. It is apparent, for these reasons, that the synthesis of collagens is an elaborate process in which a series of specialized enzymes modifies the polypeptide chains as they are assembled on ribosomes and then pass into the cisternae of the endoplasmic reticulum en route to the Golgi apparatus and secretion from the cell.

Why the Hydroxyproline?

One of the initial intriguing questions about collagens was why they contain large amounts of hydroxyproline, an amino acid

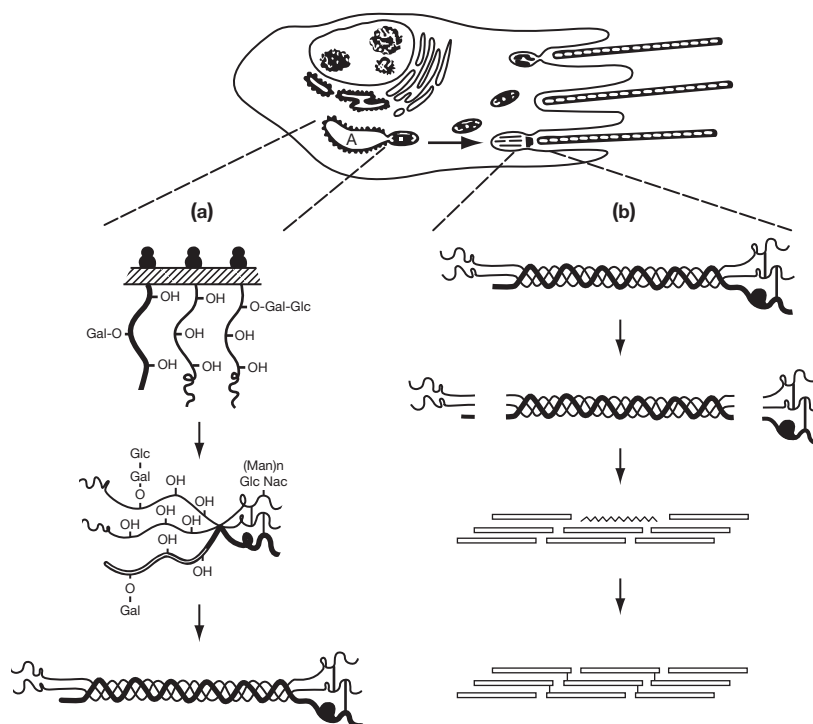


Figure 2 Schematic for the biosynthesis of a fibril-forming collagen. (a) Intracellular events that involve posttranslational hydroxylations and glycosylations, association of polypeptide chains, and folding of the triple helix. (b) Extracellular events that involve cleavage of the N- and C-propeptides, self-assembly of collagen into fibrils, and cross-linking of the fibrils. Reproduced from Prockop DJ and Kivirikko KI (1995) Collagens: Molecular biology, diseases, and potentials for therapy. *Annual Review of Biochemistry* 64: 403–434.

very rarely found in other proteins. Early isotope studies demonstrated that after administration of labeled proline to rodents, both labeled proline and hydroxyproline were present in proteins isolated from the animals. After administration of labeled hydroxyproline, however, essentially no labeled hydroxyproline was recovered. Therefore, long before the genetic code was solved, it was apparent that the hydroxyproline in collagen was introduced by an event that did not occur during the initial assembly of the protein but that occurred as a posttranslational event. These observations led to the discovery that the hydroxyproline in collagens was synthesized by prolyl hydroxylase, an enzyme that specifically hydroxylates proline in the Y position in peptides with $(\text{Gly-XY-})_n$ sequences. The question of why hydroxyproline is introduced in this manner was resolved by demonstrating that if prolyl hydroxylase is inhibited, cells synthesize complete collagen polypeptide chains that cannot fold into a collagen triple helix unless the protein is cooled to about 15°C below body temperature (Figure 2). The protein folds into a triple helix at body temperature only if about 100 proline residues in each of the 1000-amino-acid-long α -chains of fibrillar collagen are hydroxylated to hydroxyproline. After the protein folds into the triple-helical conformation, prolyl hydroxylase can no longer modify it. In effect, prolyl hydroxylase titers the hydroxyproline content of the proteins so that it folds at body temperature. Remarkably, the protein is not secreted from cells in significant amounts until it acquires the requisite content of hydroxyproline and folds into a triple helix. This elaborate system is one means of preventing the protein from folding too early in the

biosynthetic process. The process by which prolyl hydroxylase gradually titers the hydroxyproline content of collagens serves a second function of ensuring that collagen fibrils are flexible. The monomers are secreted by cells before all the Y position proline residues are fully hydroxylated and before the molecule is a completely rigid rod. Instead, regions of the protein that have low contents of proline and hydroxyproline tend to breathe in the sense that they rapidly fold and unfold. Because the monomers are not rigid triple helices, the fibrils they assemble into are not fully crystalline and remain flexible. As might be expected, monomers for fibrillar collagens from organisms with body temperatures ranging from 18 to 40°C completely unfold at about 4°C above the body temperatures of the organisms. The fibrillar collagens from organisms with varying body temperatures have different amino acids found in the X and Y positions that in part explain their different stabilities, but each begins as an unfolded protein that folds as its hydroxyproline content is titrated to a critical level by prolyl hydroxylase. After the monomers assemble into fibrils, the lateral interactions of protein make the triple helix stable to about 16°C above body temperature.

Other Levels of Control

The process of fibril assembly by fibrillar collagens is also controlled at another step. The monomers are first synthesized as larger precursors that are known as procollagens. The procollagens have large noncollagenous domains at both ends that keep

the monomers soluble and prevent them from self-assembling until the domains are cleaved by proteinases found outside the cells (Figure 2). Therefore, the proteinases help to regulate the formation of fibrils. In addition, the first fibrils formed begin to accumulate other proteins and components of the extracellular matrix on their surfaces that prevent fusion of the fibrils into large fibers. The rate at which these components are removed provides still another control on the formation of fibrils and fibers.

One Surprise: The Enzymes Involved in Collagen Synthesis Have Other Functions

The biosynthesis of collagen involves processing by a series of separate enzymes that have a number of unusual properties (Figure 2). Because the enzymes carried out processing steps not seen in the synthesis of other proteins, it was initially felt that the enzymes were uniquely dedicated to production of collagen. Subsequently, it became apparent that several of the enzymes had important additional functions. For example, prolyl hydroxylase is unusual among the small number of enzymes that modify proteins in that it requires unfolded peptide substrates. It is also unusual in that it specifically hydroxylates prolyl residues in the Y position of $-\text{Gly-X-Pro}-$ sequences. In addition, it is unusual in that the reaction uses iron, molecular oxygen, and ascorbate as well as α -ketoglutarate that is oxidatively decarboxylated to succinate in the reaction. It was surprising, therefore, that cloning of the genes for the two subunits of the enzyme revealed that the β -subunit is identical to protein disulfide isomerase, an enzyme that is required to ensure the correct formation of disulfide bonds during the synthesis of a large number of noncollagen proteins. Perhaps even more surprising was the discovery of three similar but different prolyl hydroxylases that hydroxylate a single proline residue in hypoxia induction factor, a protein that triggers the response of tissues to ischemia. The proline residue in hypoxia induction factor must be hydroxylated for the protein to be degraded by proteasomes. One consequence of this discovery is that inhibitors of prolyl hydroxylase initially designed to limit the excessive deposits of collagen found in scars also increase tissue levels of hypoxia induction factor, thereby enhancing their response to hypoxia. Inhibitors of prolyl hydroxylase therefore appear to provide a new strategy for the treatment of diseases such as anemias that are now treated with erythropoietin.

Similar surprises have been encountered with the proteinases that cleave the amino acid extensions on the ends of the procollagen precursors of fibrillar collagens and that keep the proteins from prematurely assembling into collagen fibers. The proteinase that cleaved the large C-terminal globular domains was a large zinc metalloproteinase with an unusual specificity for the similar but different sequences in the cleavage sites for the precursors of types I, II, and III collagens. Therefore, it seemed reasonable to assume that it was a protein with a unique function. Subsequent work demonstrated that the same enzyme was essential for processing a component of basement membranes (laminin-5) and for processing the enzyme (lysyl oxidase) that generates the aldehydes that form

the covalent cross-links in fibrils of both collagen and elastin. Even more surprising was the discovery that isoforms of the same enzyme are essential in completely unrelated processes in early embryonic life: they establish a gradient for dorsolateral development by cleaving inhibitors of morphogenic proteins (decapentaplegic/BMP-2/BMP-4).

Evolutionary Origins and Selective Pressures on $(\text{Gly-X-Y})_n$ Sequences

The manner in which $(\text{Gly-X-Y})_n$ sequences arose and were distributed across biology during evolution cannot be defined any more accurately than many other aspects of evolution. One clue seemed to come from the observation that the genes for the major fibrillar collagens have an unusual structure: most of the codons for the triple helix were found in short exons that were 54 bases long and coded for 18 triplets of $(\text{Gly-X-Y})_n$ sequences. Therefore, it appeared that evolutionary pressures had trapped the 54-base exon and replicated it to produce genes for many different collagens. Unfortunately, this attractive hypothesis was not substantiated by data on other collagen genes in which the exons are also relatively short, but the 54-base motif is not a common theme. Independent evidence on the selective pressures on the structure of collagens and the biosynthetic pathway is provided by data on mutations found in patients and a few animals with genetic diseases. Studies on mutations in fibrillar collagens confirmed that the presence of glycine in $(\text{Gly-X-Y})_n$ sequences is critical. Hundreds of mutations that substitute an amino acid with a bulkier side chain for a single glycine distort the triple helix or prevent it from folding. As a result, the mutations produce a large series of diseases of bone, cartilage, blood vessels, and other tissues. Most of the mutations produce severe and rare diseases in children; a few produce variants of common diseases such as osteoporosis or osteoarthritis. The selective pressure for maintaining the signature acid sequences explains why the mutations in collagens differ only slightly from those in organisms that have evolved separately for millions of years. Mutations in the proteinase that cleaves the N-terminal extension on two fibrillar procollagens (types I and II) also produce severe diseases of the skeleton, skin, and other tissues. However, no mutations have been found in either prolyl hydroxylase or the proteinase that cleaves the C-terminal extensions on the procollagen precursors of fibrillar collagens. Such mutations are presumably lethal early in embryonic life.

Summary

The impressive ability of $(\text{Gly-X-Y})_n$ sequences to spontaneously undergo self-assembly has been used time and again in biology to provide structures with a variety of important features. However, biological systems need safeguards to control the spontaneous self-assembly of the $(\text{Gly-X-Y})_n$ sequences. The principal safeguard is a biosynthetic pathway in which the protein is assembled through a series of complex steps that prevent premature self-assembly and modulate how perfectly the protein crystallizes.

See also: **Protein/Enzyme Structure Function and Degradation:** Collagenases; **Metabolism Vitamins and Hormones:** Amino Acid Metabolism.

Further Reading

Boudko S, Frank S, Kammerer RA, et al. (2002) Nucleation and propagation of the collagen triple helix in single-chain and trimerized peptides: Transition from third to first order kinetics. *Journal of Molecular Biology* 317: 459–470.

Canty EG and Kadler KE (2002) Collagen fibril biosynthesis in tendon: A review and recent insights. *Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology* 133: 979–985.

Engel J (1997) Versatile collagens in invertebrates. *Science* 277: 1785–1786.

Myllyharju J and Kivirikko KI (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends in Genetics* 20: 43–45.

Prockop DJ and Kivirikko KI (1995) Collagens: Molecular biology, diseases, and potentials for therapy. *Annual Review of Biochemistry* 64: 403–434.

Pugh CW and Ratcliffe PJ (2003) Regulation of angiogenesis by hypoxia: Role of the HIF system. *Nature Medicine* 9: 677–684.

Color Vision

G H Jacobs, University of California, Santa Barbara, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Color vision The ability of an animal to reliably discriminate between lights that differ only in their spectral energy distributions.

Opsins G-protein-coupled receptors embedded in the membranes of photoreceptor cells. They are light sensitive and mediate the conversion of photons to neural signals.

Polymorphism The presence of two or more different phenotypes within the population of a species.

Spectrally opponent cells Cells in the visual system that signal the difference in absorption rates of two or more types of cone photopigments, thus providing the neural substrate for color vision.

Univariance Describes a response property of photopigments wherein they generate an equal-sized signal for every photon that is absorbed irrespective of its wavelength.

Introduction

Light reaching the eye from various locations in the environment can vary both in its distribution of spectral energies and in its overall intensity (total photon flux). Color vision is defined formally as the ability of an animal to make reliable discriminations based solely on former, irrespective of values of the latter. In perceptual terms this means that in tests to establish and characterize color vision potential cues to discrimination based on variations in object brightness or lightness are rendered irrelevant. A familiar example of the application of this strategy is exemplified by the various plate tests that are often used to test human color vision. These plates are composed of an array of small, nonoverlapping circles that vary in size, lightness, and spectral reflectivity. In the test the viewer is asked to identify an object appearing in the plate, typically a numeral or a geometric form, when the presence of that object is consistently defined only by its shared property of color. A strong implication of this definition of color vision is that any visual system capable of supporting that capacity must be so configured that it can disambiguate joint variations in the intensity and wavelength of light. Here we consider the biological elements making it possible.

Photoreceptors

In vertebrate retinas two types of photoreceptors, the familiar rods and cones, serve to provide the initial active stage in the visual process. Both are delicate and densely packed, elongate cells positioned at the outermost portion of the retina with their long axes aligned with the incoming light (**Figure 1(a)**). With few exceptions all vertebrate retinas contain both rods and cones, but the two types show significant variation across species in their overall numbers, their relative representation (cone/rod ratios), and their spatial distributions across the extent of the retina. It has long been appreciated that, in general, rods and cones subserve different domains of vision, cones being linked to daylight (photopic) vision providing the information required for high spatial and temporal resolution, as well as color vision, while rods, with their higher sensitivity and lower inherent noise properties, support vision when ambient light levels are generally low

(scotopic). Although rods are usually associated with achromatic vision, both types of photoreceptors can contribute to vision during those transitional twilight periods between these two light regimes (mesopic); under these conditions rods have been shown capable of making some contributions to color vision. Such instances are not considered here further.

Cone Photopigments

Cone photopigments are densely packed in a stack of membranous disks that make up the outer segments of the photoreceptors (**Figure 1(a)**). The total pigment density, and thus its capacity for trapping incoming photons, depends on the packing density of the photoreceptors and the lengths of the outer segments and these may vary considerably both within and across retinas. At their best, cones operate with high efficiency; for example, in the central part of the human retina cone pigment can absorb as many as 75–80% of all incident photons. These pigments are composed of a protein (opsin) linked to a chromophore that is an aldehyde of vitamin A. Opsins are members of a large family of G-protein-linked receptor molecules in the case of opsins being constituted as a chain of ~350 amino acids embedded in a lipid membrane in the form of seven transmembrane α -helical palisades with retinal bound via a Schiff base linkage to a lysine residue in the seventh helix (**Figure 1(b)**). Vertebrate photopigments are constructed from either of two chromophores – retinal or 3-dehydroretinal. Although there are some notable exceptions, the latter are generally found in freshwater fish, amphibians, and reptiles, whereas retinal is common in animals living in marine and terrestrial environments and is the only chromophore used to produce photopigments in all birds and mammals. From a functional point of view, replacing retinal with 3-dehydroretinal has the effect of displacing the wavelength of maximum absorbance (λ_{max}) of the photopigment toward the longer wavelengths, the magnitude of the displacement being greater for those pigments that have maximal absorbance in the longer wavelengths.

Absorption of a photon of light causes the chromophore to isomerize from an 11-*cis* to an all-*trans* configuration initiating a cascade of molecular events that result in a change in the flow

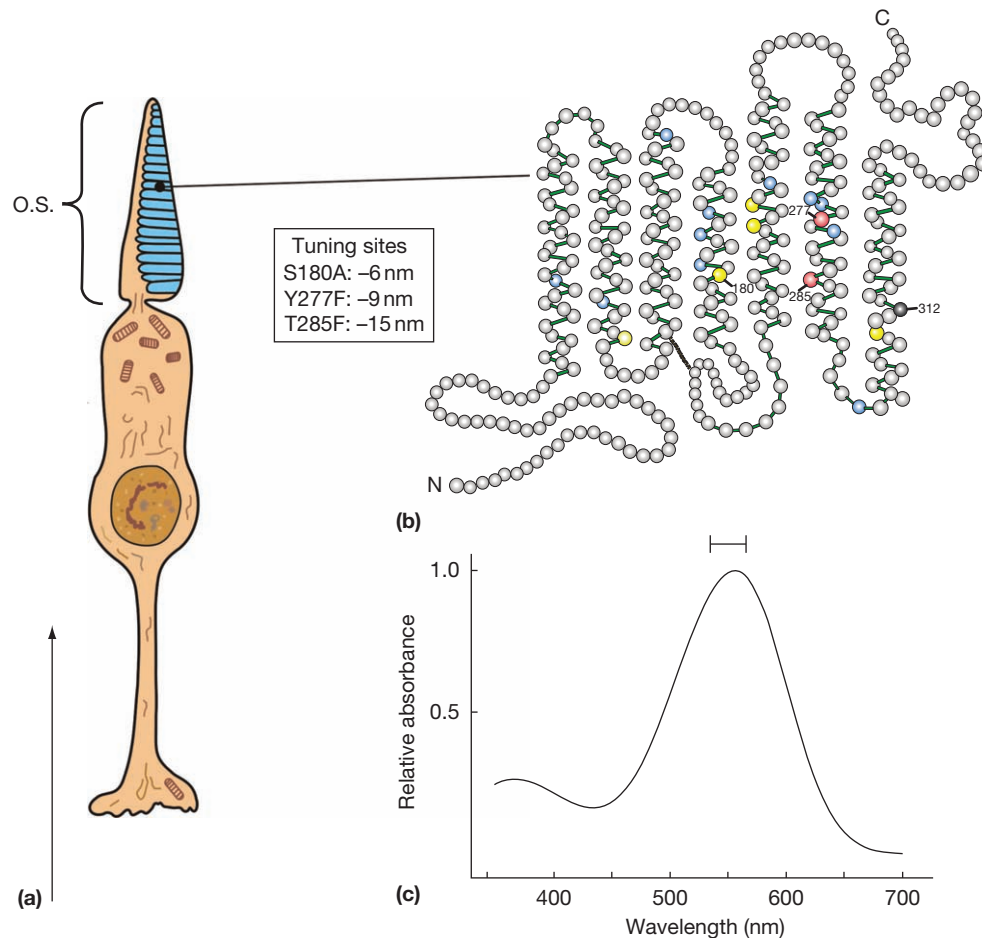


Figure 1 (a) Sketched version of the vertebrate cone photoreceptor with the normal pathway of incoming light indicated by the arrow. The outer segment (OS) consists of a stack of membrane-limited disks which contain the photopigment. The synaptic terminal (at the base of the cell) is the location where photoreceptors deliver signals to other retinal cells. (b) Schematic representation of the M/L primate cone opsin. The 364 amino acids are drawn as small circles. The three residues responsible for the spectral tuning of these pigments in New World monkeys are numbered (180, 277, and 285). Dimorphic amino acid substitutions at these sites shift the absorption spectrum of the pigment by the amounts and direction listed inside the box. The chromophore attachment site is at residue 312. (c) Absorption curve for an M/L cone photopigment. Various combinations of residue substitutions at the three indicated sites shift the absorption properties of the M/L pigments in New World monkeys so as to allow them to occupy five discrete spectral locations (all within the range indicated by the bar). Reproduced from Jacobs GH (2009) Cone photopigment evolution. In: Squire LR (ed.) *Encyclopedia of Neuroscience*, vol. 3, pp. 59–65. Oxford: Academic Press, with permission from Elsevier.

of ionic current across the cell membrane. This cascade occurs rapidly and entails a considerable amplification of signal strength. This signal spreads along the extent of the photoreceptor membrane and modulates the release of the neurotransmitter glutamate from cone terminals, thus providing an input signal to second-order cells in the retina. Retinal absorbs maximally in the short wavelengths (~ 440 nm) but, when bound to an opsin, the absorbance spectrum is shifted. The magnitude and direction of this shift, and thus the $\lambda_{\max}$ of photopigments, depend on the structure of the particular opsin. When properly scaled, the shapes of the absorbance spectra of all photopigments are very similar and so are conventionally characterized according to their $\lambda_{\max}$ values (Figure 1(c)). Photoreceptors behave univariantly meaning that once absorbed every photon yields an identical sized signal. Note also that photopigments display a secondary sensitivity peak (β -band) in the short wavelengths (Figure 1(c)).

Cone Opsin Genes

The cones of the human retina contain any of three separate classes of cone photopigment. The genes specifying these three opsins were first isolated and sequenced about 25 years ago and, as had long been apparent from studies of the inheritance of human color vision defects, genes specifying the opsins for two of these cone pigments (having $\lambda_{\max}$ at ~ 530 and 560 nm and conventionally called M and L, respectively) were localized to the X-chromosome (Xq28) where they are positioned in a head-to-tail tandem array. Each of these genes specifies an opsin made up of 364 amino acids. Having a recent common origin, the amino acid sequences for the M and L opsins are 96% identical. The gene for the opsin of the third type of human cone pigment (with $\lambda_{\max} \sim 420$ nm and termed S) was mapped to chromosome 7 (7q31–32). The S-opsin consists of ~ 350 amino acids and, reflective of the fact that the genes are

more distantly related, has a sequence identity to those for M and L of only about 40–45%.

In recent years opsin genes have been isolated and characterized for many other species (200 or more vertebrate species along with many invertebrates thus far) and that number continues to grow rapidly. This sequence information can be accessed at a website maintained by the National Center for Biotechnology Information (NCBI). These data have fostered the derivation of direct correlations between opsin sequences and photopigment absorption spectra, thus providing critical links between opsin genes and visual function. One major conclusion is that the spectral positioning of the absorption spectra of photopigments is controlled at a small number of critical locations in the opsin structure. A recent estimate concludes that the spectral positioning of all vertebrate photopigments involves changes at no more than a total of 30 different residues and within any subgroup of vertebrate photopigments that number is considerably more restricted. For example, the M/L cone photopigments of New World (platyrrhine) monkeys are polymorphic being positioned at any of five different $\lambda_{\max}$ values between ~530 and 560 nm. Those variant spectral positions reflect dimorphic amino acid substitutions at only three opsin locations (Figure 1(b)). These same three sites are also implicated in controlling the spectral positioning of the M/L photopigments in all the primates and, indeed, across all mammals only five amino acid sites are linked to shifts in absorption spectra over this region of the spectrum.

Large-scale comparisons of opsin gene sequences obtained from many species show that all vertebrate photopigments can be classified as belonging to five paralogous groups (i.e., sharing a common ancestry). An opsin gene family tree illustrating this fact is shown in Figure 2, which incorporates results obtained from a least one representative of each of the five major vertebrate groups. In total, these opsin genes code for photopigments whose $\lambda_{\max}$ values span a range from about 360 to 620 nm. Genes from one of the five groups (Rh1) are consistently linked to pigments expressed in rod photoreceptors; the other four families yield cone pigments. Note that genes from all five families (termed LWS, Rh2, Rh1, SWS2, and SWS1) yield pigments encompassing different spectral ranges and that there is some overlap in the achieved $\lambda_{\max}$ values across the various families. An additional important fact (Figure 2) is that representative birds, reptiles, and fishes may have pigments drawn from each of the four cone opsin gene families that is not true for either the amphibian (newt) or the two mammals (mouse and human) represented in the figure. Amphibians lack Rh2 genes, while humans, like all other eutherian (placental) mammals, derive cone pigments from only two of the families (SWS1 and LWS).

Linking Photopigments, Nerve Signals, and the Dimensionality of Color Vision

The univariant property of photopigments noted previously implies that a retina containing only a single type of photopigment is incapable of providing the information necessary to support color vision and, indeed, tests run on such animals show that they lack color vision. On the other hand, light reaching the retina that contains some spectral variation will inevitably

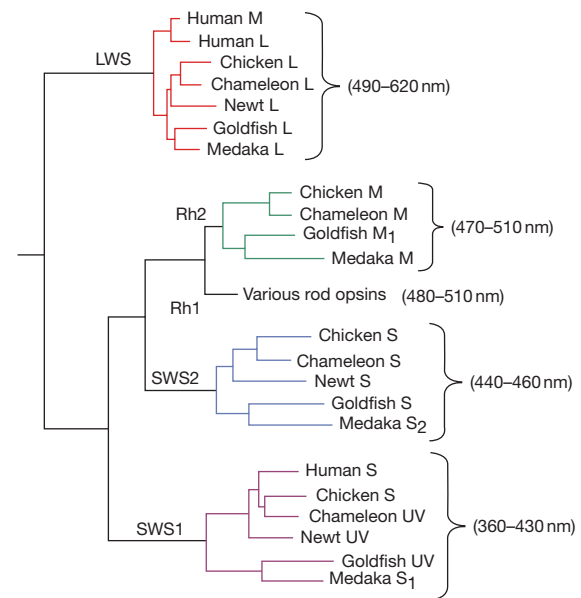


Figure 2 Opsin gene family tree. Indicated are representative contemporary species that draw their pigments from each of the five gene families (LWS, Rh2, Rh1, SWS2, and SWS1). The total range of $\lambda_{\max}$ values for pigments derived from genes of each of the families are listed to the right. These values are appropriate for photopigments that utilize the chromophore retinal. Reproduced from Jacobs GH (2009) Cone photopigment evolution. In: Squire LR (ed.) *Encyclopedia of Neuroscience*, vol. 3, pp. 59–65. Oxford: Academic Press, with permission from Elsevier.

be differentially absorbed by two types of photopigment, and that difference will be maintained irrespective of the overall intensity of the light. Thus, a nervous system that allows a comparison of signals emerging from two types of photoreceptors containing different photopigments can provide the substrate for color vision. Such an arrangement exists in the retinas of all vertebrate animals that are known to have color vision. Although the exact details of the neural circuitry to allow this show variations from species to species, in every case signals derived from one class of cone are merged with signals derived from the second class of cone onto recipient cells in a push/pull, excitatory/inhibitory fashion. Cells of this type, typically called spectrally opponent, effectively code the magnitude of the difference in absorption of the two cone pigments for any particular spectral stimulus.

As an example of this arrangement, consider the case of the human retina with its three classes of cone photopigment. Outputs from cones containing M and L pigment are combined in spectrally opponent cells, thereby generating a signal that is, effectively, L–M. Such signals are known to support color discriminations over the middle to long wavelengths and the emergent dimension of color vision is thus termed (for our species) red–green. If there are only two types of cone pigments and a single type of spectrally opponent cell, the system supports a single dimension of color vision (dichromatic). The human retina contains a third type of photopigment (S) and a separate set of neural connections in the retina allows the emergence of a second class of spectrally opponent cells – in this case involving signals drawn from all three cone classes in a combination that is effectively S – (M + L). This

allows an additional dimension of color vision, called yellow–blue in humans, and in conjunction with the red–green system renders humans trichromatic. Additional dimensions of color vision may emerge in animals whose retinas contain additional types of cone pigment and spectrally opponent cells.

Of course, color vision encompasses much more than just being able to make simple discriminations between spectrally variant stimuli (e.g., we link color percepts to surfaces and objects, use color as a source of information about the quality of objects and locations in the environment, etc.). All these operations are supported by additional neural processing that occurs along the projection pathways into the visual brain and beyond.

Evolution of Color Vision

Comparisons of gene structures have yielded reasonable agreement on a likely phylogenetic scheme for the evolution of the five vertebrate opsin gene families (Figure 2), but it is still uncertain as to when these events occurred and what functional consequences supported their evolution. Estimates based on the rate of gene divergence indicate that the four opsin cone classes must have emerged very early in vertebrate history, perhaps more than 500 million years ago (mya). This estimate finds support in the linked facts that is also about the time when jawless and jawed vertebrates diverged and that modern descendants of both lineages are known to have representation from all four cone gene classes. From similar kinds of observations, it appears that the Rh1 gene family (specifying rod photopigments) appeared later, probably as a result of a duplication of ancestral Rh2 cone opsin genes. The strong functional implication from this is that rods are a secondary major adaptation for vertebrate sight that have emerged as a specialization for supporting vision at low light levels. Although all four cone opsin gene families have been maintained in modern fishes, birds, and reptiles, both amphibians and mammals have each lost representation from these gene families. In the case of the amphibians, Rh2 genes are absent probably having been lost early in the evolution of this group.

A separate scenario for the evolution of photopigment genes in mammals is sketched in Figure 3. Note that no mammals have representation from the Rh2 gene family indicating that family was lost very early in mammalian history. An SWS2 cone

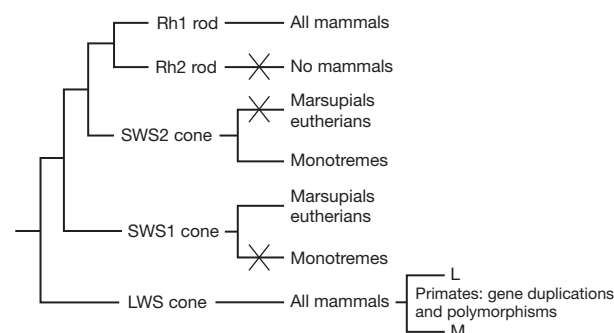


Figure 3 Schematic illustration of the pattern of photopigment loss and retention believed to have occurred during the evolution of mammals. The X's indicated the loss of photopigment types due to a loss of a gene family. Further details of this pattern are described in the text.

pigment gene has been detected in contemporary monotremes (platypus, echidna), but not in any of the marsupials or eutherian mammals, implying that the SWS2 gene family was lost prior to the time at which the latter two groups diverged from the monotremes (~ 160 – 170 mya). SWS1 cones survive in marsupials and eutherians but are absent in monotremes so that loss must be specific to the monotremes. All mammals retain representation from the LWS cone opsin gene family. Many species from one family of mammal, the primates, have acquired two or more representatives from the LWS gene family. Features of the primate arrangement are considered separately below. The circumstances that surround these various losses of opsin gene families have not been established, but there are some speculations: for example, one idea is that these all occurred during that early period when mammals were predominantly nocturnal and were thus perhaps subject to diminished adaptive pressures to maintain cone pigments given that cone operation provides predominant support for daylight vision.

The most common change to occur during the evolution of cone pigments involved not the addition or loss of gene families but rather shifts in the absorption spectra of the cone pigments. Such shifts have mostly come about as the result of residue changes at critical tuning sites in the opsin. For example, the ancestral SWS1 cone pigment in mammals is believed to have had $\lambda_{\max}$ in the ultraviolet, at ~ 360 nm. Although some contemporary mammals have retained this ultraviolet (UV) pigment (e.g., mouse and rat), in most mammalian lineages the SWS1 pigment was shifted into the visible wavelengths where it peaks, depending on the species, somewhere in the range ~ 420 – 450 nm. Pigment peak shifts of this type are functionally important as they can equip an animal to more efficiently harvest photons in some particular spectral band, thus supporting new behavioral repertoires or allowing the exploitation of new photic environments. In their interactions with signals derived from other pigment types, shifts in the absorption spectra will also alter the nature of derived color vision. It seems clear that across the full range of vertebrates such cone pigment peak shifts have occurred very frequently.

Given the known linkages between cone opsin genes, cone photopigments, and the dimensionality of color vision in some contemporary species, it is tempting to conclude that the evolutionary history of the cone opsin genes is also that for color vision. However, one should be cautious about such jumps because more is required to support color vision than just the presence of multiple types of cone opsin genes; for example, these genes have to be expressed in sufficient numbers of photoreceptors and the outputs from them have to be actively contrasted in order to yield color vision. Because of these additional organizational requirements, so far mostly unevaluated, we cannot be absolutely certain that the early emergence of multiple types of cone pigments necessarily implied a similar early emergence of elaborate color vision. Despite these reservations, it is nevertheless true that multiple types of cone pigments appeared early in vertebrate history and that arrangement has been maintained in various guises throughout the evolution of vertebrates. Those facts make it seem likely that a capacity for color vision has also been present throughout the long history of vertebrates and, given that, there is also strong support for the idea that color vision contributes significantly to survival in many different types of animal.

Primate Color Vision

Primate color vision is unique. Like other mammals, the earliest primates (dating to ~80–90 mya) were nocturnal and believed to have two types of cone pigment and dichromatic color vision. However, as primates subsequently diverged many acquired diurnal lifestyles supported by a variety of visual system alterations to enhance cone-based vision. Among the changes were significant improvements in color vision.

One signal event in the evolution of primate color vision occurred uniquely in the catarrhine primates (humans, apes, and Old World monkeys). About 35 mya the X-chromosome opsin gene duplicated. This added a second X-chromosome cone opsin gene which was subsequently altered in structure to encode a second LWS-based pigment (Figure 2) and that, in turn, set the stage for an added dimension of color vision making trichromatic color vision a routine capacity in all contemporary catarrhines. Most New World monkeys have followed a different path to enhance color vision. As noted earlier, many of these monkeys have highly polymorphic color vision. The polymorphism is based on the presence of multiple (usually three) alternative versions (alleles) of the X-chromosome LWS opsin genes each allele coding for a pigment having different spectral absorption properties (Figure 1). Since male monkeys have only a single X-chromosome, they get any one of the three versions of the LWS opsin gene and in conjunction with the SWS1 autosomal opsin genes have two cone pigments and dichromatic color vision. Females have two X-chromosomes so individuals that are homozygous at this gene locus also have any of the three M/L pigments and dichromatic color vision; however, heterozygous females have two different LWS opsin genes yielding a total of three cone pigments endowing them with trichromatic color vision. The polymorphic gene arrangement of the New World monkeys thus produces a rich array of individual differences in color vision. The third large group of contemporary primates, the more primitive strepsirrhines (lemurs, bushbabies and the like), shows considerable species variations in their opsin gene and color vision arrangements such that although some gain a polymorphic trichromacy of the sort just described, most have dichromatic color vision or they lack color vision entirely.

Humans normally have two different X-chromosome opsin genes supporting trichromatic color vision, but there are well-studied polymorphisms that result in dramatic individual variations in human color vision. Of these, the most familiar is red/green color blindness. About 2% of all Caucasian males have only a single photopigment absorbing in the middle to long wavelengths and so have dichromatic color vision. This defect apparently arose from an unequal meiotic recombination event. Since the two cone opsin genes are adjacent

to one another on the X-chromosome, such an unequal recombination can result in the production of an X-chromosome that has only a single opsin gene; since males have only a single X-chromosome such individuals will produce only a single type of M/L photopigment and become dichromatic. This same condition is much rarer in females simply because they have two X-chromosomes and thus have to be homozygous for this same alteration in order to have only a single M/L pigment. There are even more frequent red–green color vision variations where unequal recombination led not to a loss of a gene but rather to alterations in gene structures that shift the absorption spectrum of the second pigment. This too will produce clear changes in color vision that are also principally male afflictions. It is curious that color blindness, a condition reasonably common among humans, is effectively absent in all other catarrhine primates. To date there are no very satisfactory explanations for this difference.

See also: **Signaling: Photoreceptors.**

Further Reading

- Bowmaker JK (2008) Evolution of vertebrate visual pigments. *Vision Research* 48: 2022–2041.
- Carroll J and Jacobs GH (2008) Mammalian photopigments. In: Masland RH and Albright TD (eds.) *The Senses: A Comprehensive Reference*, pp. 247–268. Amsterdam: Elsevier.
- Hunt DM, Carvalho LS, Cowing JA, and Davies WL (2009) Evolution and spectral tuning of visual pigments in birds and mammals. *Philosophical Transactions of the Royal Society of London B* 364: 2941–2956.
- Jacobs GH (1993) The distribution and nature of colour vision among the mammals. *Biological Reviews* 68: 413–471.
- Jacobs GH (2007) Primate color vision: A comparative perspective. *Visual Neuroscience* 25: 619–633.
- Jacobs GH (2009) Cone photopigment evolution. In: Squire LR (ed.) *Encyclopedia of Neuroscience*, vol. 3, pp. 59–65. Oxford: Academic Press.
- Jacobs GH (2009) Evolution of colour vision in mammals. *Philosophical Transactions of the Royal Society of London B* 364: 2957–2967.
- Jacobs GH (2010) The Verriest Lecture 2009: Recent progress in understanding mammalian color vision. *Ophthalmic and Physiological Optics* 30(5): 422–434.
- Jacobs GH and Nathans J (2009) The evolution of primate color vision. *Scientific American* 300: 40–47.
- Solomon SG and Lennie P (2007) The machinery of colour vision. *Nature Neuroscience Reviews* 8: 276–286.
- Yokoyama S (2008) Evolution of dim-light and color vision pigments. *Annual Review of Genomics and Human Genetics* 9: 259–282.

Relevant Website

<http://www.ncbi.nlm.nih.gov> – NCBI; Opsin Sequence Information is Catalogued by the NCBI

Conservative Site-Specific Recombination

M C M Smith, University of Aberdeen, Aberdeen, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Site-specific recombinase The proteins that catalyse site-specific recombination reactions. These are named by the type of DNA rearrangement they mediate, that is integrase, resolvase, invertase, or transposase.

Site-specific recombination DNA breakage and reunion reactions that occur at specific DNA sequences.

Overview

Site-specific recombination systems mediate certain types of genetic exchange in bacteria that are important for growth, adaptation, and evolution. Mobile genetic elements such as phages, plasmids, and transposons, frequently encoding antibiotic resistance determinants or virulence genes, require site-specific recombination systems for their maintenance and/or cell-to-cell movement. Most bacteria rely on site-specific recombinases for the stable maintenance of chromosomes, ensuring that each of the daughter cells receives one copy of the chromosome. Site-specific recombination systems are also used to control gene expression. The recombinases fall into one of two protein families, the serine and the tyrosine recombinases, named after the amino acid residues that transiently become covalently linked to the DNA substrates during recombination. The two families are mechanistically and evolutionarily distinct and mediate precise and highly controlled DNA rearrangements. Despite there being only three outcomes of site-specific recombination, DNA integration, resolution (excision), and inversion, the control mechanisms used by different recombinases vary enormously.

The Biology of Conservative Site-Specific Recombination

The first description of a site-specific recombination reaction was by Allen Campbell, who proposed that the phage lambda prophage attaches to the *Escherichia coli* genome during the lysogenic growth cycle. Campbell's idea, which was later proved to be an important principle for most temperate phages, was that the host and the circularized phage genomes each have an attachment site, *attB* and *attP*, respectively, where a single crossover integrates the phage genome so that it becomes a part of the host chromosome (Figure 1(a)). The endpoints of the integrated prophage are marked by two new attachment sites, *attL* and *attR* (for left and right). During excision, *attL* and *attR* undergo a single crossover to regenerate *attB* and *attP*. The proteins that mediate these integration and excision reactions are integrases. Integrases, transposases, resolvases, and invertases are now understood to mediate conservative site-specific recombination, so called because the energy of the phosphodiester bond from the DNA backbone is maintained by a covalent protein-DNA recombination intermediate. In addition, site-specific recombination is not accompanied by overall loss or gain of

DNA. Although being ubiquitous in bacteria, conservative site-specific recombination systems are rare in eukaryotes.

Bacterial mobile genetic elements, such as plasmids, transposons, integrons, and phages, frequently rely on a site-specific recombination system for some aspect of their maintenance or movement. Some elements such as the integrative conjugative elements (ICEs) resemble temperate phage genomes, in that they need to excise from and integrate into the bacterial chromosome, frequently using specific attachment sites (e.g., *attB* and *attP* or *attTn*). Cell-to-cell transfer of ICEs occurs by conjugation of an excised element to a recipient cell, in whose genome the ICE integrates. Integrons are specialized genetic loci for the expression and exchange of gene cassettes, again requiring DNA integration and excision. Superintegrons such as those in *Vibrio* species can contain hundreds of gene cassettes, some of which confer antibiotic resistance and others are implicated in virulence, but majority are of unknown function. The integron platform comprises an integrase gene, *intI*, and a recombination site, *attI*, located just downstream of a promoter. Integrase mediates recombination between the *attI* and an *attC* site located on the incoming circular gene cassette placing the gene cassette so that it is expressed by the promoter. Recombination regenerates the *attI* site, enabling integration of further gene cassettes, and recombination between *attC* sites permits excision.

Resolution systems generally act to split fused genetic elements (co-integrates or tandem repeats) into separate molecules (Figure 1(b)). Typically, the Tn3 family of replicative transposons employs a resolvase to mediate resolution of a co-integrate containing a fusion of the donor and recipient DNA molecules and two copies of the transposon. Resolvase acts on a site within each copy of the transposon to regenerate the donor molecule and to release the recipient, having acquired a complete copy of the transposon. The most general site-specific recombination system in bacterial cells is the XerC/XerD/*dif* system that resolves chromosome dimers formed by homologous recombination during genome duplication, enabling stable chromosome segregation during cell division. Remarkably, the same recombinases, XerC and XerD, are used in bacteria for at least two other purposes: resolution of plasmid multimers to maintain plasmid stability and phage genome integration (e.g., *Vibrio cholera* phage CTX ϕ ; see below).

Inversion systems have evolved in bacteria and their phages to mediate on/off switches of gene expression (phase variation) or to switch between expression of gene coding for functionally

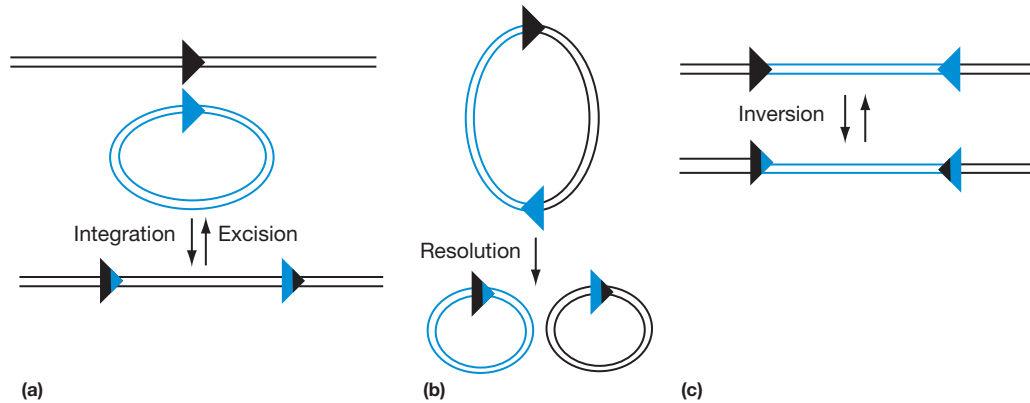


Figure 1 Recombination outcomes. (a) Integration and excision reactions, typically phage or plasmid DNA into and out of the host chromosome. The directionality (i.e., integration vs. excision) of the reaction often depends on the presence of an accessory protein, the recombination directionality factor (RDF). (b) A resolution reaction is usually unidirectional, typically chromosome or plasmid resolution or transposition intermediates. Note that a resolution reaction is essentially the same as excision in panel (a). (c) Inversion switches are usually reversible. In all outcomes the recombination sites are shown as blue or black or blue/black arrowheads. Arrowheads are used to show that the polarity of each site is always maintained after recombination. The different colors do not necessarily mean that the DNA sequences of pairs of recombination sites are different; many reactions use identical recombination sites.

alternative proteins (antigenic variation) (Figure 1(c)). The best-understood inversion system mediates the switching between two antigenic forms of flagella in *Salmonella typhimurium* by Hin invertase. An invertible element contains a promoter that, in one orientation, reads into a gene encoding the H2 flagellin and, in the other, relieves the inhibition on the expression of the gene encoding the H1 flagellin. Although many invertible loci flip the orientation of a promoter, others change the coding sequence itself. In plasmid R64 in *E. coli*, there is a cluster of invertible DNA segments that mediate diversity of the mating pilus, encoded by the *pilV* locus, required for conjugation in liquid cultures. The *pilV* locus consists of a constant 5'-end coding sequence that becomes fused to one of seven possible 3'-end coding sequences. The term coined for this type of complex, and genetically plastic locus is a shufflon. In the opportunistic gut pathogen, *Bacteroides fragilis*, the incidence of invertible systems is taken to a new extreme. There are 23 invertible promoter-containing segments, 4 shufflons, 3 invertible elements containing complete coding sequences, and 30 potential invertases. Many of these switches turn on or off the expression of enzymes for surface polysaccharides or outer membrane proteins, thus changing the antigenicity of the bacterial surface. These switches are stochastic and combinatorial, yielding a highly diverse population of cells on which selection acts. Switching the surface structures of bacteria is a common feature of pathogenic bacteria.

Recombinases and Mechanisms

The minimum requirement for site-specific recombination is a recombinase that recognizes and acts on two recombination sites, which can be identical or of different sequences. The recombinases fall into one of two evolutionary and mechanistically distinct families, the tyrosine and the serine recombinases. In both families, the overall reaction pathway is similar; the recombinase binds to the recombination substrates and

brings them together to form a synapse within which DNA cleavage occurs. Strand exchange then repositions the cleaved DNA ends such that joining of the DNA backbone forms the products.

Tyrosine Recombinases

The tyrosine recombinases include phage, plasmid and integron integrases, resolvases, invertases, and transposases. The eukaryotic type IB topoisomerases also share the chemistry and enzymology of the DNA cleavage–ligation reactions. The tyrosine recombinases have limited primary sequence similarity with most identity in the C-terminal catalytic domain. Sequence comparisons and structural data showed that the proteins contain four strictly conserved amino acids required for catalysis: an arginine–histidine–arginine catalytic triad and the tyrosine nucleophile. Extensive biochemical experiments with phage λ integrase (λ Int) showed that recombination occurs in two stages with one pair of strands exchanged before the second pair. The intermediate in which only one pair of strands has exchanged resembles a Holliday junction (Figure 2).

Biochemical experiments and structural data have yielded a good understanding of the mechanism of the tyrosine recombinases (Figure 2). The simplest examples include the Cre protein from phage P1 and the FLP protein from *Saccharomyces cerevisiae*. Both of these proteins recognize simple recombination sites, *loxP* by Cre and *FRT* by FLP. Each site consists of two binding elements 11–13 bp in length in inverted repeat orientation separated by a 6–8 bp unique sequence where the crossovers occur. Both the N- and C-terminal domains of Cre or FLP interact extensively with the binding element on both faces of the DNA. Binding by Cre monomers to two *loxP* sites is cooperative, and synapse formation is brought about by the formation of a tetramer of subunits with pseudo-fourfold symmetry. The orientation of the DNA in the synapse is antiparallel with respect to the directional sequence

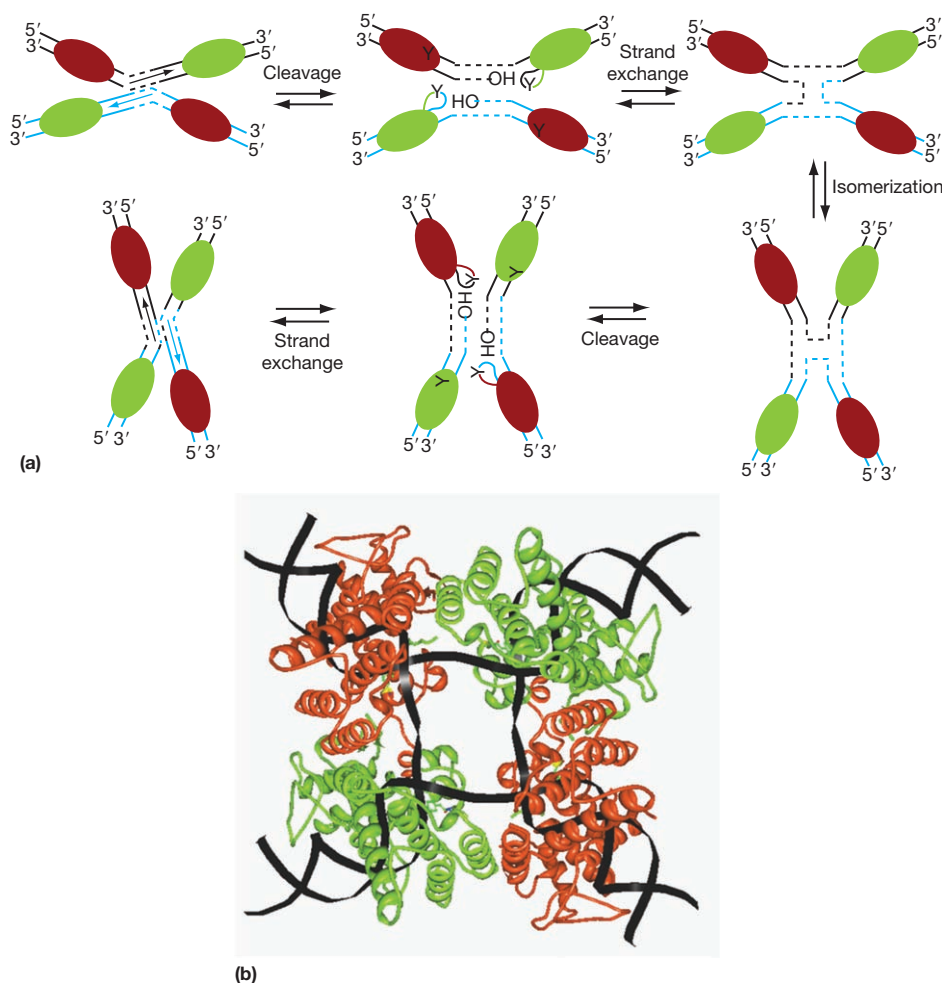


Figure 2 The mechanism of the tyrosine recombinases. (a) Reaction mechanism in Cre-loxP recombination based on insights from Cre/DNA structures. Two loxP sites are blue and black parallel lines, each site containing two binding elements for a Cre subunit. A tetramer of Cre bound to the four binding elements brings the loxP sites together in synapse where the central crossover regions (dotted lines) are reading in opposite directions. The active site tyrosines on two opposing Cre monomers are activated (green) to initiate exchange of two of the four DNA strands to form a Holliday Junction intermediate. Isomerization is a subtle change in DNA conformation in the Holliday junction intermediate that then activates the previously inactive Cre subunits (red) which then initiate exchange of the remaining strands. (b) Structure of the Holliday junction intermediate. The opposing pairs of Cre subunits that are simultaneously active in the tetramer are red and green; the DNA strands are black. The DNA lies almost entirely in one plane. The complex has the N-terminal domains closest to the viewer (PDB entry, 2CRX viewed with Protein Workshop). Reproduced from Van Duyne GD (2001) Crystal structure of the Holliday junction intermediate bound by four Cre subunits. *Annual Review of Biophysics and Biomolecular Structure* 30: 87–104.

at the crossover sequences. Activation of two opposing subunits bound to different substrates cleaves two (so-called ‘top’) strands using the tyrosine nucleophiles, activated by the arginine–histidine–arginine triad that coordinates the scissile phosphate. Acid–base chemistry generates a transient covalent phosphotyrosine linkage to the 3′-end of the cleaved DNA. The free 5′-ends peel off their partner strands and anneal to the unpaired base pairs from the other DNA substrate in a strand swap. The 5′-OH then acts as the nucleophile to attack the phosphotyrosine bond resulting in ligation of the phosphodiester backbone and formation of the Holliday junction. An isomerization reaction occurs that activates the previously inactive pair of opposing recombinase subunits; these cleave the bottom strands and essentially repeat the reactions for the first strand swap, thus completing recombination and

releasing both products and the recombinase. Mechanisms to control the outcome of recombination (i.e., integration versus excision, resolution, and inversion) by the tyrosine recombinases are superimposed on this reaction mechanism (see below).

There are significant variations on the reaction mechanism itself. For most tyrosine recombinases, the identity of the crossover region is essential for efficient recombination, and past and present proposals for the reaction mechanism always ensure that there is a way of reading the identity of this sequence. However, there are some recombinases, notably Tn916 and CTnDOT integrases that do not require extensive DNA identity at the crossover site. These recombinases only need about 2 bp of sequence identity after the position of the first strand cleavage; thereafter, the sequences

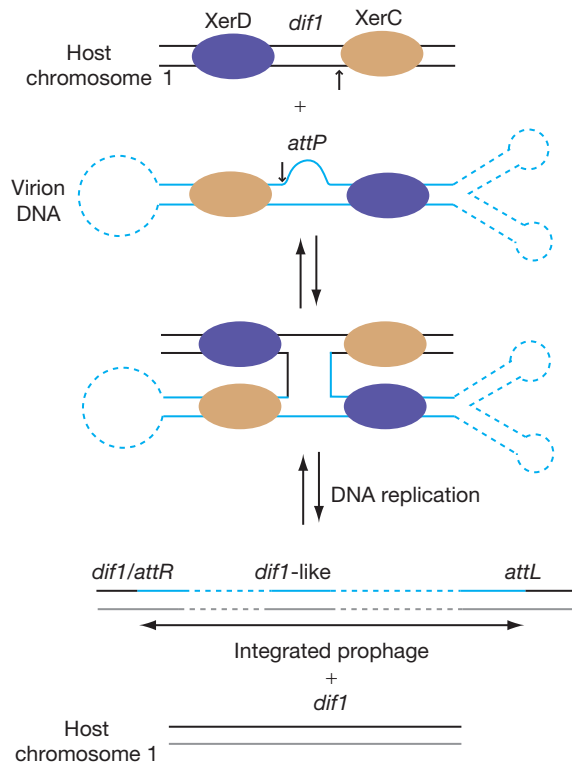


Figure 3 XerC and XerD bind to their cognate-binding elements at the *dif* site on chromosome 1 of *V. cholera*, *dif1*, and to the folded *attP* site formed from the CTX ϕ virion DNA. First strand cleavage and rejoining by XerC covalently inserts the ssDNA CTX ϕ genome, and either replication or gap repair then converts this to a dsDNA prophage integrated at *dif1*. The *dif1* site is regenerated in this process, but, at the other end of the prophage, is a *dif*-like site that cannot take part in a reverse reaction and this single integrated phage genome cannot therefore excise. Further integration of CTX ϕ genomes into *dif1* can occur by Xer recombination or into the prophage by homologous recombination; excision occurs between two tandem CTX ϕ genomes by homologous recombination.

at the crossover and second-strand cleavage site can be different. This explains how these transposons can target relatively random sequence.

The integron integrase, IntI, and XerC/XerD can mediate recombination where one of the substrates might be circular single-stranded DNA (ssDNA). In the integron platform, the *attI* site in the chromosome resembles a simple recombination substrate comprised of two recombinase-binding elements flanking the crossover sequence. However, the other recombination substrate, the *attC* site, is generated from a folded circular ssDNA molecule. In *V. cholera*, the host XerC/XerD enzymes and the *dif1* site are exploited for the integration of a single-stranded phage genome, CTX ϕ . The self-pairing in the *attC* sites and the CTX ϕ *attP* sites creates the recombinase-binding elements separated by the crossover sequences (Figure 3). In both cases, the reaction pathway is halted after exchange of only one pair of strands, that is, at the Holliday junction intermediate. DNA replication of the recombination product then yields two different daughter strands, only one of which contains an insertion.

Serine Recombinases

The serine recombinases include phage integrases, resolvases, invertases, and transposases. The resolvase/invertase recombinases are the best understood mechanistically (e.g., $\gamma\delta$ and Tn3 resolvases required for co-integrate resolution, the plasmid resolvase, Sin, and the invertases, e.g., Hin and Gin). The resolvase/invertases are small proteins that contain an approximately 140-amino-acid (aa) N-terminal domain and a small C-terminal domain of 40–50 aa, required for recognition and binding to the recombination substrates. The substrates for the resolvases, the *res* sites, are generally quite large (Tn3 *res* is 114 bp in length) containing three binding sites, called subsites, for three resolvase dimers. Only one of these subsites (i.e., subsite I) is where recombination occurs with the other two being required solely for the control of directionality (see below). Subsite I is an inverted repeat about 28 bp in length. The substrates for the invertases are also inverted repeats; the *hixL* and *hixR* sites recognized by Hin invertase are 26 bp in length. Both Hin and Gin require an accessory protein, Fis, bound to an enhancer sequence to control recombination.

The mechanism of the resolvase/invertases is described here by considering only the events at the recombination sites (Figure 4). The serine recombinase N-terminal domain is required for catalysis, strand exchange, and subunit–subunit interactions. In solution serine recombinases are dimers, mediated through a dimer interface in the catalytic domain. Recombinase dimers bind to their recombination substrates, and further protein–protein interactions, also mediated through the N-terminal domains, generate a tetramer of recombinase subunits and bring the two DNA substrates together in a synaptic complex. Large conformation changes occur within the synaptic tetramers that are presumed to activate catalysis. The active site serine residues on each of the four recombinase subunits attack the scissile phosphates on all four DNA strands, transferring the 5'-phosphates to the serine hydroxyl. The DNA cleavage leaves a 2-bp staggered break with a free 3'-OH and the covalent phosphoserine linkage to the recessed 5'-end. A pair of recombinase subunits bound to different DNA substrates then rotates 180° relative to the other pair to align the DNA ends, such that ligation of products occurs through the attack by the free 3'-OH groups on the phosphoserine bonds. As cleavage occurs in all four DNA strands at the same time, only the subunit–subunit interactions hold the DNA together. A crystal structure of a tetrameric covalent intermediate of $\gamma\delta$ resolvase reveals a large flat hydrophobic interface within the tetramer that would enable subunit rotation without dissociation.

Other members of the serine recombinase family have different domain structures. The large serine recombinases have the N-terminal catalytic domain but have a large 300–500-aa C-terminal domain. The large serine recombinases include phage-encoded serine integrases that, like λ Int, use an *attP* and *attB* site to integrate phage DNA. The integrases from the *Streptomyces* phage ϕ C31 and from a mycobacteriophage Bxb1 are the most studied of this group. The ϕ C31 and Bxb1 integrases have a similar mechanism of strand cleavage and strand exchange to the resolvase/invertases. The large C-terminal domains are required for substrate recognition

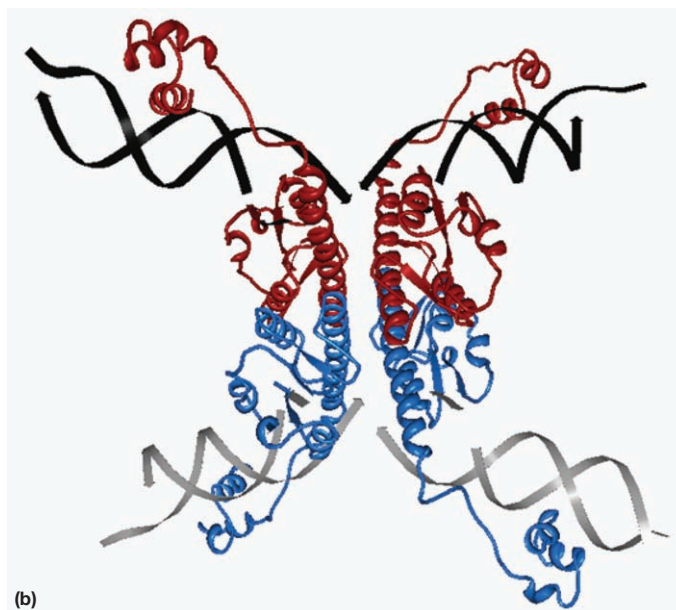
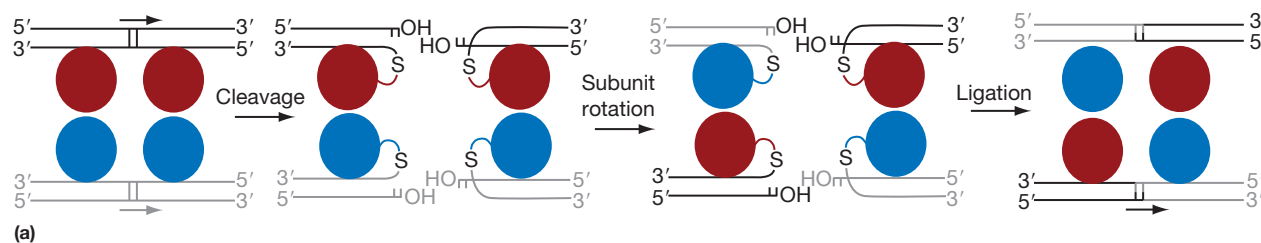


Figure 4 Mechanism of a serine recombinase. (a) The reaction mechanism at the recombination sites for the resolvase/invertases and for serine integrases. Serine recombinases bind to a pair of recombination sites (shown here as two parallel black or gray lines with the 2-bp crossover sequences shown as two short vertical lines) as dimers (two red and two blue subunits) and then form a tetramer, bringing the DNA sites together in a synapse. All four subunits are simultaneously activated to cleave the DNA, forming in each subunit a phosphoserine link to a recessed 5'-end and leaving a free 2-bp 3'-overhang. Subunit rotation exchanges two DNA half sites, repositioning the half sites into a recombinant format. Ligation of the DNA backbone generates the recombinant molecules. Note that base pairing must occur between the 3'-overhangs and the recessed 5'-ends in the recombinants; mismatches in the recombinants lead to reformation of the substrates trapping knots or catenanes in the process. If the 2-bp sequence at the crossover is nonpalindromic, then only one of two synapse structures that differ by the relative orientations of the recombination sites will avoid mismatches in the recombinants. (b) The tetrameric covalent intermediate of an activated mutant of $\gamma\delta$ resolvase (PDB 1ZR4, viewed with Protein Workshop). The $\gamma\delta$ subunits colored red and blue represent the dimers that initially bound either a black or gray recombination site, respectively. The structure is arbitrarily shown just after cleavage and prior to strand exchange (two black half sites and two gray half sites are still adjacent to each other in a parental format). The large flat interface that mediates subunit rotation is easily visible down the center of the tetramer. (a) Reproduced from Stark WM, Boocock MR, and Sherratt DJ (1992) Catalysis by site-specific recombinases. *Trends in Genetics* 8(12): 432–439.

and for directional control (see below). There is also a group of uncharacterized transposases (e.g., IS607) that have the serine recombinase catalytic domain at the C-terminus and a small DNA-binding domain at the N-terminus.

Control of Recombination

Maintaining Site Polarity

An important feature of site-specific recombination systems is the need to maintain the integrity of the recombination sites through repeated rounds of recombination. To this end, all recombination systems have a mechanism to distinguish the polarity of the substrates such that the left side of one recombination site always joins to the right side of the other and vice

versa. This strict requirement for left–right joining also ensures a predictable outcome for recombination. For example, in the simple systems such as Cre-*loxP* and FLP-FRT, the recombination sites have the directional crossover sequence that defines polarity. Cre recombines two *loxP* sites in head-to-tail configuration on a plasmid to produce two separate plasmid circles (resolution products), but if the *loxP* sites are fully symmetrical (*loxS*) both resolution and inversion occur. Many serine recombinases also rely on the crossover site to define polarity as base pairing at the 2-bp staggered break is important for ligation of the recombinants. As for the *loxS* site, if the 2-bp crossover sequences are made symmetrical in both recombination substrates, normal left side to right side joining is lost. Thus far, all serine integrases and invertases have nonsymmetrical 2-bp crossover sites to avoid loss of substrate polarity.

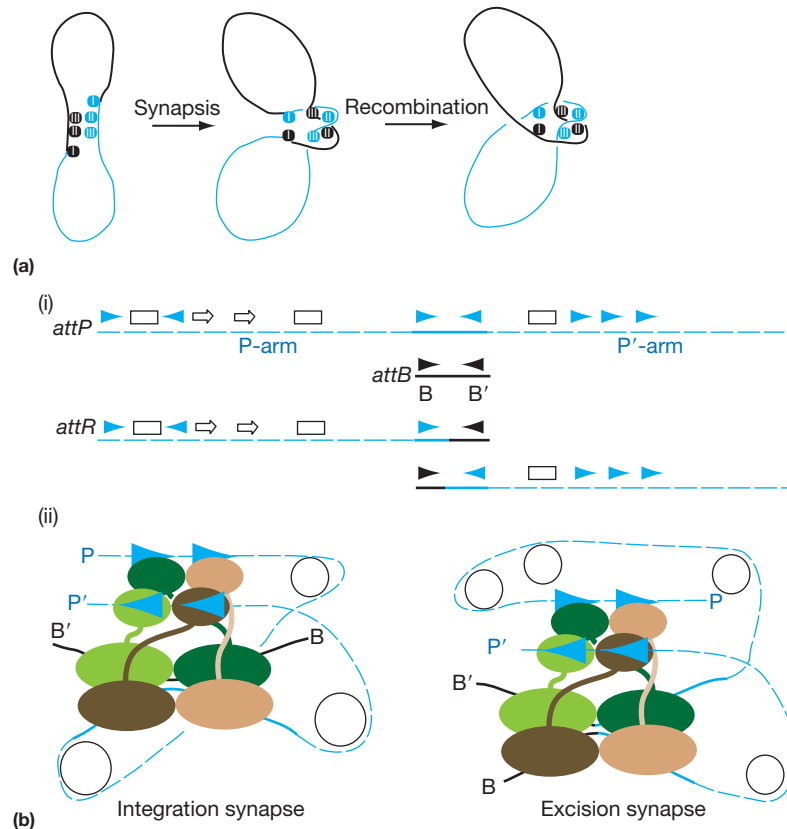


Figure 5 Two common mechanisms of control of recombinases. (a) The topological filter as displayed by $\gamma\delta$ and Tn3 resolvases. Dimers of resolvase bind to the three subsites in the transposon *res* sites that lie at the boundaries of the two elements to be resolved (illustrated by thick and a thin lines within the co-integrate). Subunit interactions between resolvases bound at sites II and III trap three negative nodes and bring the resolvases bound at site I together to form an active tetramer. Subunit rotation is right handed, driven by stored energy from supercoiling. The products are almost entirely singly linked catenanes that are resolved by the bacterial topoisomerase IV. (b) Control of integration vs. excision by λ Int: (i) arrangement of Int and accessory protein-binding sites on the attachment sites. *attP* is the blue line comprising the arms (dashed) and the core (solid) and *attB* is black. *attL* and *attR* comprise a half site from *attP* (blue) and a half site from *attB* (black). λ Int binds to arm sites (small blue arrowheads) and core sites (longer blue and black arrowheads). The *attP* site also has binding sites for IHF (white rectangles) and for Xis (white arrows). (ii) λ Int integration and excision is controlled by the formation of two different synapse structures depending on the presence of different DNA bending proteins (shown as white circles) and the nature of the substrates. At the center of both synaptic complexes is a tetramer of Int subunits contacting arm and core sites simultaneously. Each int subunit is shown in a different color, with the large oval being the core-binding and catalytic (CB) domains and the smaller oval is the N-terminal, arm-binding domain. Only once the synapse is formed are the CB domains activated to initiate strand exchange using the mechanism described for the tyrosine recombinases (Figure 2). (a) Adapted from Grindley ND, Whiteson KL, and Rice PA (2006) Mechanisms of site-specific recombination. *Annual Review of Biochemistry* 75: 567–605. (b) Reproduced from Van Duyne GD (2005) Lambda integrase: Armed for recombination. *Current Biology* 15(17): R658–660, with permission from Elsevier.

Controlling the Outcome

Site-specific recombinases have the capacity to cleave DNA, an activity that, if it goes uncontrolled, could lead to many unwanted DNA rearrangements or potentially lethal double-strand breaks. All site-specific recombinases are therefore inactive until the correct protein–protein and protein–DNA interactions are made in a synaptic complex containing both substrates and the recombinases. Cre and FLP are unusual systems as they are reversible (although with Cre-*loxP* resolution is greatly favored over integration solely due to kinetic factors). Most recombination systems have accessory proteins and/or accessory binding sites to control the directionality or outcome of a recombination reaction.

Many resolvases of the serine recombinase family such as Tn3 and $\gamma\delta$ resolvases have accessory resolvase-binding sites close to

the recombination site that wrap the DNA to form a nucleoprotein complex with a specific topology (Figure 5(a)). Only when the two *res* sites are in the same DNA molecule and in a head-to-tail orientation is the synapse formed that activates catalysis and recombination. This topological filter ensures that recombination cannot occur between two recombination sites in head-to-head orientation on the same molecule or between sites on different molecules. The substrates for recombination by Sin recombinase have an accessory-binding site for a Sin dimer that directly contacts Sin bound at the recombination site and one binding site for a histone-like protein whose function is to help bend the DNA into the correct synapse. This arrangement of binding sites also acts as a topological filter.

Invertases such as Hin and Gin also use topological filters to ensure inversion rather than resolution and to prevent integration. With the Hin system a recombination enhancer segment

binds two Fis dimers that contact Hin dimers bound to *hixL* and *hixR*. The enhancer must be located on the same molecule as the invertible segment and up to about 4 kbp from the nearest *hix* site. In addition, the DNA in this synaptic complex (or invertosome) follows a specific topological path ensuring that only sites on the same molecule and in inverted orientation can recombine.

The bacterial-encoded tyrosine recombinases, XerC and XerD, are the most versatile of all the recombinases being required for chromosome dimer resolution, plasmid resolution, and integration of some phage genomes. Each different function requires a different set of accessory proteins and pairs of recombination sites. XerC and XerD bind to their cognate-binding elements on either side of the crossover sequence (e.g., [Figure 3](#)). In the resolution of chromosome, dimers XerC and XerD bind to two identical *dif* sites, which are the recombination substrates located in the *ter* region of the chromosome. The recombinases are activated by FtsK, a protein located in the closing cell division septum and that pulls the DNA toward the XerC/XerD/*dif* complex. FtsK activates the XerD pair to make the first crossover followed by XerC activation and the second crossover. In plasmid resolution, the recombination sites, *cer*, have slightly different sequences from *dif* such that any strand cleavage is inhibited until two accessory proteins, PepA and ArgR, bind generating a topologically defined complex similar to that made by the $\gamma\delta$, Tn3, and Sin resolvases. In phage integration, XerC and XerD are again required, this time without needing an accessory protein but using a folded ssDNA *attP* site (see above and [Figure 3](#)).

Phage integrases such as λ Int have an elaborate mechanism for controlling integration and excision of the phage genome involving different juxtapositions of DNA-binding sites for both integrase and accessory proteins. λ Int is a tyrosine recombinase and has the core-binding and catalytic domain analogous to Cre and Xer recombinases but has, in addition, a small N-terminal domain, the arm-binding domain. This domain binds to specific arm sequences in the *attP* site, which is about 240 bp in length ([Figure 5\(b\)\(i\)](#)). Thus, the heterobivalent DNA-binding activity of λ Int contacts both the core site where the crossover occurs, and the arm-binding sites simultaneously. These contacts are facilitated by a host protein, integration host factor (IHF), a bending protein that binds to its cognate-binding sites in *attP*. In integration, binding to *attP* by Int and IHF forms a nucleoprotein structure (the intasome) that then captures the small *attB* site consisting only of a core site to form the integration synapse ([Figure 5\(b\)\(ii\)](#)). Recombination occurs within this tetramer to yield *attL* and *attR*. Recombination between *attL* and *attR* (excision) requires another protein, Xis, encoded by the phage genome and which binds to *xis*-binding sites in the *attP* derived half of *attR*. In the presence of λ Int, Xis, IHF, and, under some circumstances, Fis, an excision synapse is assembled that activates *attL* $\times$ *attR* recombination ([Figure 5\(b\)\(ii\)](#)).

The phage-encoded serine integrases such as Bxb1 and ϕ C31 integrases have evolved a much simpler system for the regulation of directionality. Only *attP* $\times$ *attB* recombination occurs with integrase alone, and like λ Int an accessory protein (Xis or recombination directionality factor (RDF)) is required for *attL* $\times$ *attR* recombination. Unlike λ Int system, the recombination sites are both relatively short (<50 bp) and are

different sequences. No other proteins are required other than integrase and the RDF. The mechanism of directionality relies on different DNA-induced recombinase conformations that enable synapsis and activation of catalysis. The default reaction for the serine integrases is integration, and the recombination substrates *attP* and *attB* are both imperfect inverted repeats. The products *attL* and *attR* contain half sites from *attP* and *attB*, and the integrase subunits bound to these half sites have different conformations that prevent the excision reaction. The role of the RDF identified from the mycobacteriophage Bxb1 is to bind directly to integrase, preventing the *attP* $\times$ *attB* reaction but stimulating the *attL* $\times$ *attR* reaction.

Conclusions and Applications

Bacteria use conservative site-specific recombination systems for a wide variety of processes that are profoundly linked to both the stability of genomes and their evolution. Mechanistic studies on the two distinct protein families that mediate site-specific recombination reactions have revealed precedents in protein function and DNA recognition. Because some recombination systems are simple requiring only the recombinase and the two recombination sites, many are being exploited as tools for manipulation of very complex genomes and potentially as vectors for gene therapy in humans and animals. Optimization of the recombinases in synthetic biology applications depends on a detailed understanding of their mechanism and regulation.

See also: **Molecular Biology:** DNA Sequence Recognition by Proteins; DNA Topoisomerases: Type I; DNA Topoisomerases: Type II; Nonhomologous Recombination: Bacterial Transposons.

Further Reading

- Branda CS and Dymek SM (2004) Talking about a revolution: The impact of site-specific recombinases on genetic analysis in mice. *Developmental Cell* 6: 7–28.
- Grindley NDF, Whiteson KL, and Rice PA (2006) Mechanisms of site-specific recombination. *Annual Review of Biochemistry* 75: 567–605.
- Lesterlin C, Barre F-X, and Cornet F (2004) Genetic recombination and the cell cycle: What we have learned from chromosome dimers. *Molecular Microbiology* 54: 1151–1160.
- Mazel D (2006) Integrons: Agents of bacterial evolution. *Nature Reviews Microbiology* 4: 608–620.
- Radman-Livaja M, Biswas T, Ellenberger T, Landy A, and Aihara H (2006) DNA arms do the legwork to ensure the directionality of lambda site-specific recombination. *Current Opinion in Structural Biology* 16: 42–50.
- Rice PA, Mouw KA, Montano SP, et al. (2010) Orchestrating serine resolvases. *Biochemical Society Transactions* 38: 384–387.
- Smith MCM, Brown WRA, McEwan AR, and Rowley PA (2010) Site-specific recombination by ϕ C31 integrase and other large serine recombinases. *Biochemical Society Transactions* 38: 388–394.
- Val M-E, Bouvier M, Campos C, et al. (2005) The single-stranded genome of phage CTX is the form used for integration into the genome of *Vibrio cholera*. *Molecular Cell* 19: 559–566.
- Van Duyn GD (2001) A structural view of Cre-loxP site-specific recombination. *Annual Review of Biophysics and Biomolecular Structure* 30: 87–104.
- Venken KJT and Bellen HJ (2005) Emerging technologies for gene manipulation in *Drosophila melanogaster*. *Nature Reviews Genetics* 6: 167–178.

Cyclic AMP Receptors of *Dictyostelium*

D Hereld, National Institute on Alcohol Abuse and Alcoholism, Rockville, MD, USA

Y Kamimura and P N Devreotes, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Chemotaxis Directed movement of cells toward (or away from) the source of diffusible chemoattractant (or repellent) molecules.

G protein Heterotrimeric proteins with inactive (guanosine diphosphate (GDP)-bound) and active (guanosine triphosphate (GTP)-bound) states which are activated by ligand-occupied receptors and, in turn, activate downstream targets within the cell.

G-protein-coupled receptors Cell surface, integral membrane proteins possessing seven membrane-spanning α -helices and usually capable of activating cytosolic G proteins upon binding specific extracellular signaling molecules (also known as seven-transmembrane or serpentine receptors).

Signal transduction Molecular events by which the perception of an extracellular signal by a cell is translated into an appropriate cellular response.

Dictyostelium Development and Cyclic Adenosine Monophosphate Signaling

Dictyostelium discoideum is an amoeba found in soil where it feeds on bacteria. In order to survive periods of starvation, 10^4 – 10^5 amoebae aggregate and execute a 24-h developmental program that yields a fruiting body comprised of a round mass of spores held aloft by a slender stalk (Figure 1(a)). When nutrients return to the environment, spores germinate to yield amoebae which resume cell division.

Not long after Sutherland and his colleagues discovered cyclic adenosine monophosphate (cAMP) to be an important intracellular second messenger in hormonal signaling, Konijn and his associates demonstrated that cAMP is a potent chemoattractant for *Dictyostelium* and correctly speculated that it was the so-called acrasin secreted by starving amoebae which mediates their aggregation. Shaffer had proposed that the acrasin would be emitted periodically by cells at aggregation centers and relayed outwardly as waves by surrounding cells. Indeed, exogenous cAMP was shown to elicit the transient activation of adenyl cyclase and secretion of cAMP. Tomchik and Devreotes later demonstrated the concentric waves of extracellular cAMP waves which arise every ~6 min at aggregation centers and travel radially outward through aggregating populations. Subsequent pharmacologic characterization of these responses established the framework for the identification of the cAMP receptors.

Identification and Properties of cAMP Receptors

The molecular identification of the first cAMP receptor, cAR1, began with its photoaffinity labeling with the cAMP analog, 8-azido-[32 P]cAMP. This approach identified a protein of either 40 or 43 kDa depending on whether or not the cells had been exposed to cAMP, suggesting the existence of a reversible, ligand-induced modification which proved to be phosphorylation. Radiolabeling of cells with [32 P]phosphate permitted purification of the 43-kDa phosphorylated form of the receptor.

Antibodies directed against purified cAR1 led to the isolation of a complementary DNA (cDNA) that encoded a protein of the expected size and hybridized to a messenger RNA (mRNA) expressed transiently in early development consistent with cAMP binding. Formal proof that the isolated cDNA did indeed encode cAR1 came from expression of the cDNA, which resulted in increased cAMP binding, and disruption of the corresponding gene, which obviated cAMP binding and cAR1-mediated responses. The deduced sequence of cAR1 possessed seven putative transmembrane domains (Figure 2) and exhibited weak homology to mammalian G-protein-coupled receptors (GPCRs). cAMP stimulation of guanosine triphosphate (GTP) binding to isolated membranes and GTP hydrolysis was further evidence that cAR1 is indeed coupled to a G protein. This together with studies of yeast pheromone signaling provided the earliest indications that virtually all eukaryotes have inherited these ancient sensory mechanisms.

Three other highly homologous cAMP receptors, designated cAR2–4, were subsequently identified by hybridization with a cAR1 probe. The cARs are expressed successively during development, peaking in expression at roughly 5-h intervals in the order: cAR1, cAR3, cAR2, and cAR4 (Figure 1(b)). Disruption of the genes encoding each cAR results in developmental defects consistent with the timing of their expression. cAR1[−] cells fail to aggregate. cAR2[−] cells arrest shortly after aggregating at the mound stage. cAR3 gene disruption has been variously reported either to have no apparent effect or to interfere with spore cell differentiation late in the mound stage and consequently yield fruiting bodies predominantly comprised of stalk cells. cAR4[−] cells develop normally beyond the mound stage but exhibit defects in culmination, resulting in misshapen fruiting bodies.

The cARs differ markedly in their affinities for cAMP. The early cARs, cAR1 and cAR3, have high affinities (i.e., low- to mid-nM K_d 's), whereas those expressed later in development, cAR2 and cAR4, have low affinities (K_d 's > 1 μ M). These affinities are appropriate for the extracellular cAMP concentrations that exist at these stages of development (Figure 1(c)). During aggregation, the cAMP signal oscillates from sub-nM to near μ M

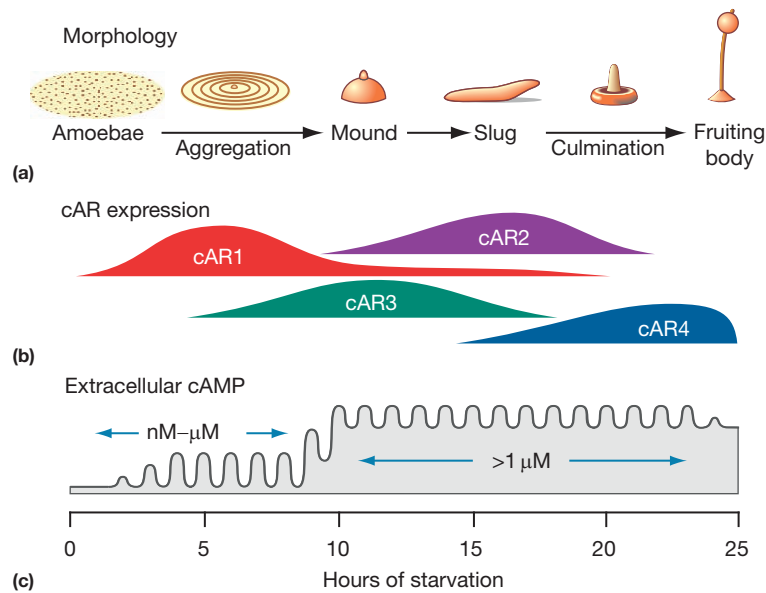


Figure 1 Correlation of developmental morphology, cAR expression, and extracellular cAMP levels. See text for additional details.

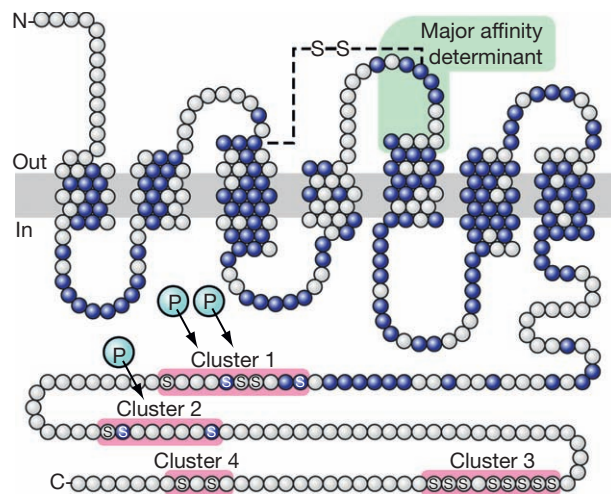


Figure 2 Model of cAR1. Dark blue spheres represent amino acids conserved in all four cARs. Densely packed amino acids in the plane of the membrane (gray bar) signify the seven hydrophobic transmembrane helices. Nonconserved amino acids (light gray spheres) in the region labeled major affinity determinant are largely responsible for the widely differing cAMP affinities of cAR1 and cAR2 (and possibly other cARs). Clusters of serines in the C-terminal cytoplasmic domain (S), the distribution of ligand-induced phosphorylation (P), and the putative disulfide bridge (—S—S—) are also indicated. cAR2–4 differ in the lengths and sequences of their C-terminal cytoplasmic domains.

concentrations. By contrast, external cAMP oscillates at elevated concentrations exceeding $1\text{ }\mu\text{M}$ in the multicellular stages. Nonconserved residues in the second extracellular loop largely determine whether a cAR has a high or low affinity (Figure 2). By analogy with rhodopsin, this extracellular loop, positioned by disulfide linkage to the extracellular end of the third transmembrane helix, is likely to lie at the entrance to the cAMP-binding cleft of cARs and, thereby, influence binding.

Signaling Pathways

cAR1 is perhaps the most versatile GPCR yet to be characterized as it regulates a wide range of downstream effectors and biological responses. Consequently, it serves as a valuable model for understanding diverse modes of GPCR signaling. cAR1 mediates three principal cellular responses during aggregation: (1) propagation of cAMP waves, (2) chemotaxis up the cAMP gradient of each oncoming wave, and (3) regulation of genes required for development. The pathways underlying these responses have been determined to a large extent (Figure 3(a)). Most striking is the dichotomy between signaling pathways that involve G proteins and those that do not. Comparatively less is known about the pathways governed by cAR2–4 in large measure due to technical challenges posed by multicellularity, although significant progress has been made toward elucidating mechanisms by which these cARs promote cell differentiation in multicellular stages.

G-Protein-Dependent Pathways

Dictyostelium possesses at least nine heterotrimeric G proteins comprised of distinct α subunits (designated $G\alpha 1$ – 9) and common β and γ subunits. Genetic and biochemical evidence indicate that $G\alpha 2\beta\gamma$ is the principal G protein to which cAR1 couples. Activation of $G\alpha 2\beta\gamma$ by cAR1 liberates the $G\beta\gamma$ dimer and GTP-bound $G\alpha 2$, which, in turn, activate various effectors.

The activated G protein triggers a network of interconnected pathways that ultimately control cell–cell signaling and chemotaxis. The $G\beta\gamma$ dimer-mediated activation of phosphoinositide-3-kinase (PI3K) converts the membrane phospholipid PIP_2 (phosphatidylinositol-4,5-bisphosphate) into PIP_3 (phosphatidylinositol-3,4,5-trisphosphate). PIP_3 , in turn, binds to the PH domain of cytosolic regulator of adenylyl cyclase (CRAC), thus recruiting it to the plasma membrane. CRAC participates in the activation of adenylyl cyclase A (ACA) by an unknown mechanism. The resulting cAMP functions intracellularly via

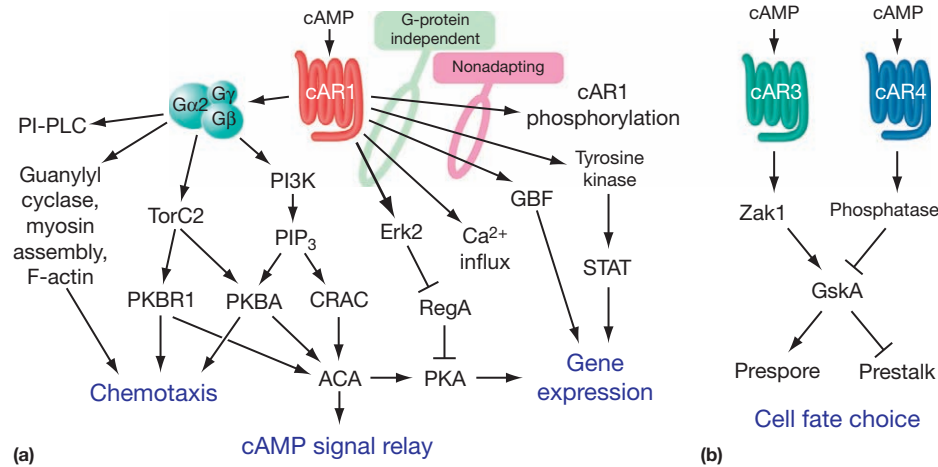


Figure 3 cAMP receptor signaling pathways. Pointed (↓) and flatheaded (⊥) arrows indicate activation and inhibition, respectively. (a) cAR1 pathways leading to chemotaxis, cAMP relay, and gene regulation. G-protein-independent and nonadapting pathways are indicated. In multicellular stages, cAR2–4 might activate some of the pathways shown for cAR1. (b) Role of cARs in the GskA-mediated determination of cell fate. cAR2 (not shown) might promote prestalk cell differentiation at the mound stage in the manner shown for cAR4.

protein kinase A (PKA) to modulate gene expression and, in addition, is secreted from the cell in order to relay the cAMP signal to neighboring cells.

The receptor–G protein system leads to the activation of two protein kinase B (PKB) isoforms that are critical for chemotaxis. Protein kinase B type A (PKBA) is a PH domain-containing protein that depends on PIP₃ for its activation. In addition, protein kinase B-related 1 (PKBR1) is myristoylated and activated independently of PIP₃. Both PKBs are regulated by target of rapamycin complex 2 (TorC2) via phosphorylation of their hydrophobic motifs. In cells undergoing chemotaxis in a shallow cAMP gradient, these PKBs are exclusively activated at the cells' leading edge where they presumably promote actin assembly and pseudopod extension. These mechanistic insights into cAR1-mediated chemotaxis, namely the involvement of PI3K activation and recruitment of PKB to the leading edge, have since been found to also pertain to the chemotaxis of a variety of mammalian cells including neutrophils.

The Gα2 subunit is implicated in the cAR1-dependent activation of guanylyl cyclase which provides another important input to the chemotaxis machinery. The product of guanylyl cyclase, cGMP, regulates myosin heavy chain kinases and, thereby, promotes the assembly of conventional myosin II in posterior and lateral regions of chemotaxing cells where it can propel the rear of the cell forward upon contraction and suppress lateral pseudopod formation.

G-Protein-Independent Pathways

Several cAR-mediated responses appear to be G-protein-independent based on their preservation in cells lacking what is believed to be the sole G-protein β-subunit gene. These include phosphorylation of cAR1, uptake of Ca²⁺, and activation of the mitogen-activated protein kinase Erk2, the transcriptional regulators G-box-binding factor (GBF) and signal transducer and activator of transcription (STAT), and the glycogen synthase kinase-3 (GSK3) homolog GskA.

cAR1 is phosphorylated on serine residues within its C-terminal cytoplasmic domain. The 18 serines in this domain exist in four clusters (Figure 2). In unstimulated cells, cAR1 is basally phosphorylated within clusters 2 and 3. Upon stimulation with cAMP, cAR1 becomes reversibly hyperphosphorylated due to the addition of approximately two phosphates within cluster 1 and a third phosphate within cluster 2. Cluster 1 phosphorylation causes the 40- to 43-kDa electrophoretic shift. Other cARs also undergo cAMP-induced phosphorylation commensurate with their affinities.

Erk2 promotes the intracellular accumulation of cAMP by negatively regulating RegA, a cAMP-specific phosphodiesterase. Erk2 is presumably the third of three kinases in a typical mitogen-activated protein (MAP) kinase cascade. The identity of the upstream kinases as well as the G-protein-independent mechanism by which cAR1 activates the cascade remains to be determined.

cAMP triggers the rapid and transient influx of Ca²⁺ ions in aggregation-competent cells. This is mediated largely by cAR1 due to its natural abundance at this stage in development. However, expression of cARs in vegetative cells indicates that the components required for uptake are expressed at this stage and that other cARs can also mediate Ca²⁺ uptake. The magnitude of Ca²⁺ uptake is roughly proportional to receptor level, indicating that cARs are the limiting component for this response.

GBF binds G-rich elements in early post-aggregative genes and is required for their induction by cAMP. Gene disruptions suggest that either cAR1 or cAR3 mediate the G-protein-independent activation of GBF. In contrast to many aggregation-stage genes whose expression requires periodic cAMP pulses that mimic natural cAMP waves, GBF-mediated gene expression is induced by constant cAMP, indicating that it is not subject to adaptation. Thus, the GBF pathway is appropriately activated upon aggregation when extracellular cAMP rises to levels that persistently occupy cAR1.

STAT proteins have been extensively studied in the context of cytokine signaling in mammalian immune cells. In *Dictyostelium*,

exogenous cAMP triggers STaT's phosphorylation on tyrosine, SH2 domain-mediated dimerization, and translocation to the nucleus where it governs prestalk gene expression. cAR1 is required for this response but it can be substituted in this capacity with cAR2, suggesting that multiple cARs might activate STaT during development. The cAR-activated tyrosine kinase involved remains to be identified.

GskA, a homolog of GSK3, is an important regulator of cell fate in *Dictyostelium* and cARs play key roles in regulating its activity. GskA activity promotes spore cell differentiation, whereas inactivity results in stalk cell differentiation (Figure 3 (b)). Independent of G proteins, cAR3 activates the nonreceptor tyrosine kinase Zak1 which, in turn, phosphorylates and activates GskA. On the other hand, cAR4 (and perhaps cAR2) activates a phosphotyrosine phosphatase, resulting in the dephosphorylation and inactivation of GskA. The role of G proteins in the latter process, if any, is not known. Therefore, cAR3 signaling promotes spore differentiation and cAR2 and cAR4 favor stalk cell differentiation. Because cAR2 and cAR4 are expressed predominantly in prestalk cells, it is unclear whether these mechanisms determine cell fate or act thereafter in cell-type maintenance.

One known target of cAMP-activated GskA is STaT. GskA phosphorylates multiple serines of STaT, causing it to be exported from the nucleus. This is one mechanism by which cAR3 signaling might oppose prestalk differentiation. As has been described for other developmental systems, GskA may also determine cell fate by phosphorylating β -catenin and thus targeting this transcriptional co-activator for destruction.

Mechanisms of G-Protein-Independent Signaling

It remains to be determined how cAMP-occupied cAR1 communicates with and activates these effectors. By analogy with G proteins and also mammalian G-protein-coupled receptor kinases (GRKs), the yet to be identified cAR kinase might interact selectively with the ligand-occupied conformation of the receptor's cytoplasmic loops. Other G-protein-independent processes such as Ca^{2+} uptake could involve lateral signal transduction within the plane of the membrane by direct interaction of cAR1 with another integral membrane protein, analogous to the interaction of *Halobacterium* sensory rhodopsins (SI and SRII) with their associated histidine kinases (HtrI and HtrII).

Desensitization Mechanisms

In general, receptor-mediated responses are governed by various desensitization mechanisms which attenuate the cell's responsiveness. cAMP triggers the sequential phosphorylation ($t_{1/2} \sim 2$ min) and internalization ($t_{1/2} \sim 15$ min) of cAR1. cAR1 phosphorylation causes a several-fold reduction in the receptor's intrinsic affinity for cAMP, which may extend the range of cAMP concentrations to which the receptor can respond during aggregation. In addition, preliminary results indicate that phosphorylation of cAR1 is a prerequisite for its internalization as in other systems. Prolonged cAMP exposure results in downregulation of cAR1 levels, the combined effect of diminished cAR1 gene transcription and cAR1 degradation. Degradation presumably results from delivery of internalized receptors to lysosomes.

On a more rapid timescale, nearly all of the cAR1-mediated responses to abrupt cAMP increases are transient, returning to prestimulus levels in 30 s to several minutes despite constant stimulation. This rapid and reversible attenuation of responses is referred to as 'adaptation'. The few responses that do not adapt include cAR1 phosphorylation and GBF activation. The mechanisms of adaptation are poorly understood and might be distinct for each pathway. cAR1 phosphorylation appears not to be involved as elimination of phosphorylated serine residues in cAR1 by site-directed mutagenesis has little impact on the kinetics of these responses. Cells adapted to one cAMP concentration can respond to yet higher concentrations (provided the receptor is not saturated). This observation indicates that adaptation is a graded signal that is just sufficient to offset the excitatory signal, the strength of which also reflects receptor occupancy.

Studies suggest that adaptation occurs downstream of G protein activation. The activation of the G protein can be monitored by observing the loss of fluorescence resonance energy transfer (FRET) occurring between coexpressed G α -CFP and G β -YFP subunits. These FRET experiments have shown that the G protein is persistently dissociated in adapted cells. Yet, most downstream events including PIP₃ production and PKB, TorC2, and ACA activation adapt. Taken together, these findings suggest that an adaptation pathway emanates from cAR1 and acts upon the excitatory pathway somewhere beyond the G protein.

Gradient Sensing in Chemotaxis

Temporal challenges with fixed cAMP concentrations as described above have been invaluable for deciphering cAR-mediated pathways and revealing the existence of adaptation mechanisms. However, natural cAMP waves also contain spatial information which the cells must rapidly and accurately sense for efficient chemotaxis. Although cAR1 is uniformly distributed in the plasma membrane, shallow cAMP gradients differing by as little as 2% across the length of the cell prompt highly asymmetric localization of various proteins, indicating that the cell senses and amplifies small differences in cAR1 occupancy on this surface. Proteins with PH domains including PKB and CRAC are highly localized to the plasma membrane of the cell's leading (or anterior) edge, indicative of elevated PIP₃ levels. This reflects the recruitment of PI3K to the anterior plasma membrane and its subsequent activation. The membrane translocation of PI3K is mediated by its N-terminal domain which presumably binds an entity in the plasma membrane that is generated in response to cAR1 activation. Conversely, phosphatase and tensin homolog (PTEN) is associated with the posterior plasma membrane via specific interactions with PIP₂. Thus, PTEN is excluded from the anterior membrane where PI3K actively converts PIP₂ to PIP₃ and instead localizes to posterior regions where PIP₂ should be more abundant. Localization of the antagonistic activities of PI3K and PTEN to opposing poles of the cell should result in a steep gradient of PIP₃ and associated PH domain-containing proteins. Other pathways regulated by cAR1 are also polarized in chemotaxing cells. For example, phosphorylated PKBR1 is found selectively at the cell's leading edge. Precisely how the

cell amplifies the directional information of a shallow chemoattractant gradient to achieve extreme gradients of activities within the cell is likely to be critically important for efficient chemotaxis and is under intensive investigation.

See also: **Cell Architecture and Function:** Cell Migration; Eukaryotic Chemotaxis; **Molecular Biology:** DNA Ligases: Structures; **Signaling:** Adrenergic Receptors; Eicosanoid Receptors; G_q Family; Gs Family of Heterotrimeric G Proteins; Opioid Receptors; P2Y Receptors.

Further Reading

- Brzostowski J and Kimmel A (2001) Signaling at zero G: G-protein-independent functions for 7-TM receptors. *Trends in Biochemical Sciences* 26: 291–297.
- Chung CY, Funamoto S, and Firtel RA (2001) Signaling pathways controlling cell polarity and chemotaxis. *Trends in Biochemical Sciences* 26: 557–566.
- Iijima M, Huang YE, and Devreotes P (2002) Temporal and spatial regulation of chemotaxis. *Developmental Cell* 3: 469–478.
- Kessin R (2001) *Dictyostelium: Evolution, Cell Biology, and the Development of Multicellularity*. Cambridge: Cambridge University Press.
- Kim L and Kimmel AR (2000) GSK3, a master switch regulating cell-fate specification and tumorigenesis. *Current Opinion in Genetics and Development* 10: 508–514.
- King JS and Insall RH (2009) Chemotaxis: Finding the way forward with *Dictyostelium*. *Trends in Cell Biology* 19: 523–530.

Cyclic GMP Phosphodiesterases

S H Francis and J D Corbin, Vanderbilt University School of Medicine, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Allosteric binding Binding of ligands to regulatory sites within a protein.

Autoinhibition Effect of one region of an enzyme to block the action of the catalytic site of another region of that same enzyme.

CAP-related domains Cyclic nucleotide (cN)-binding domains containing ~110 amino acids that provide for cN binding in PKA, PKG, guanine-nucleotide exchange factor (GEFs), and cN-gated cation channels and are evolutionarily related to the bacterial catabolite-gene activator protein.

Chimeric proteins Proteins that are composed of multiple subdomains each of which provides for a specific function of that protein.

Dual specificity Proteins that are not entirely selective among closely related ligands, for example, cAMP and cGMP.

GAF domains Domains of ~120 amino acids that provide for binding of a variety of ligands and protein–protein interactions in diverse proteins.

Phosphohydrolase Enzyme that breaks a phospho-ester bond.

Introduction

Cyclic guanosine monophosphate (cGMP) and cyclic adenosine monophosphate (cAMP) are prominent second-messenger molecules (Figure 1) that are synthesized by guanylyl cyclases (GCs) and adenylyl cyclases (ACs), respectively. The balance between the activities of the cyclases and phosphodiesterases (PDEs) determines cellular cyclic nucleotide (cN) levels (see cGMP-signaling pathway; Figure 2). Although cNs can be exported from cells by certain members of the multidrug transport proteins or exit via gap junctions, rapid lowering of cNs by action of PDEs is quantitatively the most important process. The cyclic phosphate ring of cGMP and cAMP is a key feature by which PDEs recognize their substrates, and it is highly resistant to most other phosphohydrolases. Nucleotides such as guanosine diphosphate (GDP), 5'-GMP, adenosine diphosphate (ADP), or 5'-AMP that lack this ring do not significantly interact with the catalytic sites of PDEs even at high concentrations. Structural features of the purine (Figure 1) also provide critical features for interaction of the cN substrates with PDEs. In the catalytic sites of cGMP-specific PDEs, a hydrogen bond between an invariant glutamine and the oxygen at the C6 position of the guanine is extremely important as is hydrophobic stacking between a conserved phenylalanine and the heterocyclic ring of guanine. Potent and specific inhibitors also utilize these contacts as well as novel contacts within and near the catalytic pocket.

Distribution of cAMP and cGMP

cAMP is apparently present in all eukaryotes, but cGMP is less widespread (Figure 1). The yeast genome lacks identifiable sequences for GCs or known cGMP-binding domains. However, *Dictyostelium*, *Paramecium*, *Tetrahymena*, *Trypanosomes*, and *Plasmodium* contain GCs. The role of cGMP in most of these organisms is not understood, but in *Dictyostelium* cGMP participates importantly in chemotaxis and osmoregulation. In mammals, cGMP-signaling pathways are prominent in many

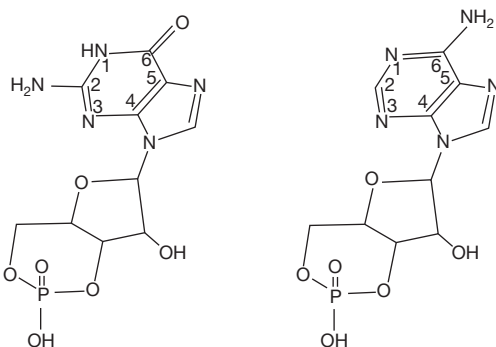
processes including smooth muscle proliferation and contraction/relaxation, visual transduction, platelet disaggregation, neuronal viability, bone growth, apoptosis, water and electrolyte homeostasis, memory and learning, cardiac functions, and aldosterone synthesis.

Classification of PDEs

Two evolutionarily distinct classes of PDEs, class I and class II, have been identified; to date, both class I and class II PDEs occur in some genera, for example, *Dictyostelium* and *Saccharomyces*. Class I PDEs are found in both protozoans and metazoans, but all known mammalian PDEs are class I PDEs. In eukaryotes, the catalytic sites and allosteric cN-binding sites in class I PDEs are functionally and evolutionarily distinguished from each other and from those of other known intracellular cN receptors; these include cN-binding sites in proteins such as the cAMP- and cGMP-dependent protein kinases (PKA and PKG, respectively) that belong to the bacterial catabolite gene-activator protein (CAP-related family), the cGMP-binding domains in some PDEs, and anion transporters. However, class II PDEs contain an evolutionarily distinct catalytic domain, and allosteric cN-binding sites in these proteins belong to the CAP-related family of proteins.

Class I PDEs

The class I PDE superfamily in mammals contains 11 PDE families (PDEs 1–11) that are products of 21 separate genes. Classification is based on the order of discovery, extent of similarity in the amino acid sequences, substrate specificities, regulatory features, and DNA sequences. Based on *in vitro* determinations of substrate preferences, three are cAMP-specific PDEs (PDEs 4, 7, and 8), three are cGMP-specific PDEs (PDEs 5, 6, and 9), and five are dual-specificity PDEs (PDEs 1, 2, 3, 10, and 11) (Figure 3). In the latter group, the relative affinities and catalytic rates for cGMP and cAMP breakdown and the affinities



Guanosine 3', 5'-monophosphate (cyclic GMP) Adenosine 3', 5'-monophosphate (cyclic AMP)

Figure 1 Molecular structures of cGMP and cAMP. The structures of cGMP and cAMP differ only at the N-1, C-6, and C-2 positions, but these differences provide for specific and selective interaction with phosphodiesterase catalytic sites. The *syn* conformation of both nucleotides is shown.

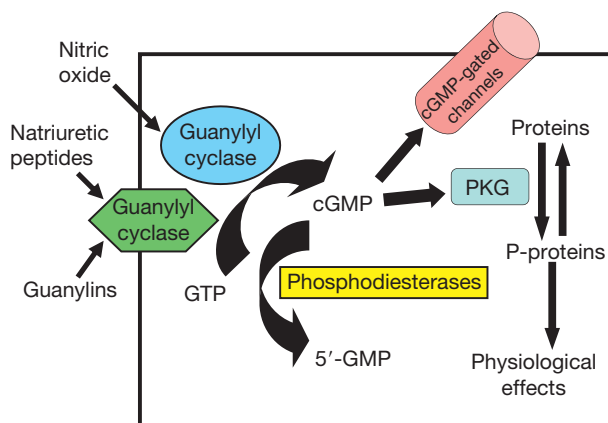


Figure 2 Model of cGMP signaling in cells. Intracellular levels of cGMP are determined by the relative activities of guanylyl cyclases and phosphodiesterases. There are two families of guanylyl cyclases; form that is largely cytosolic is activated by nitric oxide and a membrane-bound form is activated by natriuretic peptides and guanylylins. cGMP breakdown is catalyzed by the catalytic sites of phosphodiesterases that are either cGMP-specific or PDEs that hydrolyze both cGMP and cAMP, that is, dual-specificity PDEs. cGMP also interacts with allosteric cGMP-binding sites on cGMP-dependent protein kinase (PKG), cGMP-gated cation channels, and cGMP-binding phosphodiesterases to bring about its physiological effects.

for certain inhibitors can vary significantly among members of a single family as exemplified by PDE1 isoenzymes.

Domain Structure

Each class I PDE monomer is a chimeric protein that includes a conserved catalytic domain (C domain) of ~270 amino acids that is located more C-terminal to its regulatory domain (R domain). Among all known class I PDEs, only 14 amino acids are invariant, and most of these are located in the catalytic pocket. These most likely provide for critical structural and

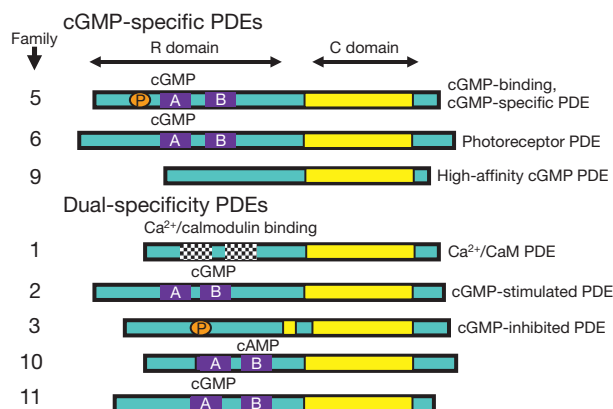


Figure 3 Schematic depiction of structural features of mammalian class I cGMP-hydrolyzing phosphodiesterases (PDEs). Arrangements of the regulatory (R) domains and catalytic (C) domains are shown for a monomer, although most known mammalian phosphodiesterases exist as dimers or oligomers. Known sites of phosphorylation (P) are indicated. The highly selective allosteric cGMP-binding is associated with GAF-A domains in PDEs 5 and 11 and with GAF-B in PDE2. All isoenzymes of the PDE2, PDE5, and PDE6 families contain two GAF domains. GAF-B in PDE10 binds cAMP with high affinity and selectivity. The role(s) of GAF domains in PDEs 10 and 11 are not known; splice variants within the PDE11 family contain 1, 2, or no complete GAF domains. The open block in PDE3 is a novel 44 amino acid insert in the catalytic domain of this family. Cross-hatched boxes denote calmodulin-binding domains (1 or 2) in the R domain of PDE1.

catalytic features. The R domains of PDEs, which are structurally and functionally more diverse, confer distinct regulatory features for each PDE family (Figure 3). The R domains contain varied subdomains including phosphorylation sites, autoinhibitory sequences, subcellular localization signals, dimerization motifs, and GAFs; GAFs are ancient domains of ~120 amino acids that occur in five PDE families. The name is derived from recognition of this motif in diverse proteins (cGMP-binding PDEs, *Anabaena* AC, and *Escherichia coli* FhlA transcription factor). Depending on the PDE in which they are located, GAFs can contribute to dimerization, interaction with other proteins, or interaction with ligands such as cGMP or cAMP. Cyclic GMP has high affinity and strong selectivity for GAFs in PDEs 5, 6, and 11, but cAMP interacts weakly with these sites in PDE2. GAFs in PDE10 have high affinity and selectivity for cAMP.

Catalytic Domains

X-ray crystallographic structures of C domains of class I PDEs reveal that they are primarily comprised of alpha-helices; three regions come together to form the catalytic pocket. Within the catalytic site, there are two prominent functional subdomains: one contains a novel binuclear metal-binding site that binds two divalent cation(s) and provides the catalytic machinery; the other subdomain provides interactions with the cN purine. The metals and cN apparently bind independently. The subdomains are closely juxtaposed in the tertiary structure in order to bind ligands and effect catalysis. Results of site-directed mutagenesis studies in PDE holoenzymes and X-ray

crystallographic structures of the C domains have identified important amino acids in each subdomain.

Metal Requirements

PDEs require divalent cation(s) to support catalysis. Class I PDEs contain two metals that are bound in close proximity and are critical to the catalytic process. Evidence indicates that an activated hydroxyl derived from solvent water and bridging the metals is inserted at the phosphate, thereby cleaving the phosphodiester bond. In two cGMP-specific PDEs (PDEs 5 and 6), Zn^{2+} is required for catalysis. Results of site-directed mutagenesis and X-ray crystallographic structures of most of the catalytic domains of class I PDEs indicate that these PDEs contain a tightly bound Zn^{2+} that is coordinated by histidines, aspartic acids, and water molecules. The second metal, which is more weakly bound, is coordinated through interactions with an aspartic acid and several water molecules that are, in turn, linked to histidines and aspartic acid. The metal occupying the second metal site in PDEs is still unclear and may vary among PDEs. Both Mg^{2+} and Mn^{2+} have been suggested to be the physiologically relevant cations at this second site, but this remains to be definitively demonstrated.

Features That Contribute to cN Specificity

The molecular basis that provides for the affinity and selectivity of PDEs for either cN is poorly understood. cAMP and cGMP interactions with the PDE catalytic sites share some common features but different topographies in the various sites dictate positioning of the respective cNs. There are two major types of interactions between cN substrates and PDE catalytic sites that contribute to the affinity of interaction, that is, hydrogen bonding with an invariant glutamine and hydrophobic stacking with residues that sandwich the purine. Based on X-ray crystallographic structures of isolated C domains of PDEs, it was suggested that the invariant glutamine forms two hydrogen bonds with either cGMP or cAMP and that constraint in the orientation of that glutamine side chain determines specificity for interaction with the guanine or adenine in cGMP or cAMP, respectively. Results of site-directed mutagenesis studies in PDE5 and PDE11 support a single hydrogen bond with substrates, and a mutation in PDE5 allowing for free rotation of the invariant glutamine side chain does not improve affinity for cAMP as would be predicted from the model. Results of X-ray crystallographic studies of PDEs 2, 9, and 10 also support a single hydrogen bond between the glutamine and the purine. The combined results indicate that multiple factors are likely to dictate specificity for either cGMP or cAMP.

In solution, cNs are in equilibrium between two conformations, *syn* and *anti*, based on orientation between the purine and ribose moieties. The preference of the catalytic site for binding cAMP or cGMP in either the *syn* or *anti* configuration varies among PDE families, but the *anti* conformer is most commonly preferred. Large substitutions on the ribose 2'-OH or the N-1 and C-2 positions of the purine are well tolerated by the catalytic sites of both cGMP-specific PDEs 5 and 6 and the dual-specificity PDEs 1, 2, and 11. In addition to the preference of

particular PDEs for the *syn* or *anti* conformation, it has now been shown that, even in dual-specificity PDEs, such as PDE10, which binds cAMP and cGMP in the *syn* conformation, the orientation of the purine ring of cAMP or cGMP can differ in the catalytic pocket presumably in order to optimize contacts in the site.

Interaction with Inhibitors

Most of the known PDE inhibitors compete with cN for access to the catalytic site. Specific structural features in the catalytic sites of PDEs provide for interaction with different inhibitors. Many inhibitors that compete for access to the catalytic site of cGMP-hydrolyzing PDEs mimic the structure of cGMP. This is true for three potent inhibitors of PDE5, that is, sildenafil (Viagra), vardenafil (Levitra) and udenafil (Zydena); tadalafil (Cialis) also competes with cGMP for the catalytic site, but its structure is less closely patterned on that of cGMP, and it forms different contacts with PDE5 than occur with sildenafil or vardenafil. Results from site-directed mutagenesis studies indicate that sildenafil and vardenafil utilize many amino acid contacts that provide for interaction of cGMP with PDE5. Affinity of PDE5 for these inhibitors is ~1000–25 000 times greater than for cGMP mainly due to the formation of many additional contacts between the respective inhibitors and the protein, but contacts common to cGMP and the inhibitors may be strengthened as well.

PDEs 5, 6, and 11 are closely related and share ~50% amino acid sequence homology. Sildenafil and vardenafil inhibit both PDEs 5 and 6 but have ~4–10-fold lower affinity for PDE6 isoenzymes than for PDE5; both are very weak inhibitors of PDE11. By contrast, tadalafil (Cialis) potently inhibits PDE5, is ~40-fold less effective against PDE11, and has little effect on PDE6. PDE1, a dual-specificity PDE, is inhibited by sildenafil with ~100-fold lower potency than that for PDE5. The catalytic sites of PDEs 1, 5, and 6 have similar preferences for cN analogs, so it is likely that sildenafil and vardenafil exploit shared structural features among these PDEs. By contrast, some other inhibitors selectively inhibit the dual-specificity PDEs 1, 2, and 3 and are ineffective against PDEs 5 and 6. The approach for development of new selective PDE inhibitors will continue to use X-ray crystallographic structures of catalytic sites as well as the traditional medicinal chemistry approach to map patterns of molecular selectivity for compounds that preferentially inhibit particular cGMP PDEs. PDE6 isoenzymes are novel in that they are inhibited by members of a specific protein family, $\text{P}\gamma$. $\text{P}\gamma$ inhibitor proteins appear to be entirely specific for the PDE6 isoenzymes, and like the PDE6 isoenzymes, expression of these proteins is restricted to photoreceptor cells and the pineal gland.

Regulation of cGMP-Hydrolyzing PDEs

cGMP-hydrolyzing PDEs are regulated by both short- and long-term regulatory mechanisms. These include ligand activation, for example, Ca^{2+} /calmodulin (PDE1) and cGMP binding to allosteric sites provided by GAFs (PDEs 2, 5, and 6); phosphorylation (PDEs 1, 3, 5, and 11) (Figure 3); selective subcellular localization; and modulation of PDE gene expression and protein levels. In a physiological setting, these frequently

work in concert to provide sensitive control of PDE activities in response to many stimuli.

Each monomer of PDEs 2, 5, 6, 10, and 11A4 contains two homologous GAFs (GAF-A and GAF-B) that are evolutionarily distinct from the catalytic sites. GAF-B in PDE2 and GAF-A in PDE5, PDE6, and PDE11 provide for allosteric cGMP binding whereas GAF-B in PDE10 provides for allosteric cAMP binding. Molecular determinants for cGMP interaction with the catalytic site differ markedly from those for cGMP binding to GAFs. In PDE2, cAMP binds to GAF-B but with >10-fold lower affinity than that for cGMP binding. However, cAMP is a very poor competitor for cGMP binding in PDEs 5, 6, and 11. In the X-ray crystallographic structure of the PDE2 R domain, cGMP is almost entirely buried within the GAF-B site. The tight fit of cGMP agrees well with the fact that few cGMP analogs compete with cGMP binding at this site.

Characteristics of cGMP-Specific Class I PDEs

PDE5

PDE5 is known as the cGMP-binding, cGMP-specific PDE. There are three PDE5 isoenzymes derived from a single gene. PDE5 has >100-fold selectivity for cGMP over cAMP at both its catalytic and allosteric sites; it is a homodimer of ~100 kDa subunits; each R domain contains two GAFs (A and B). GAF-A provides for allosteric cGMP binding, which activates catalysis. Ligand binding to GAF-B has not been shown, but this GAF provides important functional and regulatory features to the enzyme, and both GAFs contribute to dimerization. The K_m for cGMP hydrolysis is ~3 μM and the $V_{\max}$ is ~5 $\mu\text{mol min}^{-1} \text{mg}^{-1}$. PDE5 is abundant in smooth muscle cells, gastrointestinal epithelial cells, platelets, and certain neuronal cells. It is typically coexpressed with other components of cGMP signaling, including PKGs and GCs. Catalytic activity is stimulated by cGMP binding to the allosteric cGMP-binding site, phosphorylation of Ser-102 near the N terminus, and chemical reduction. Phosphorylation occurs in intact tissues in response to elevation of cGMP and PKG action and is enhanced by cGMP binding to GAF-A. PKA can phosphorylate Ser-102 but with >10-fold lower efficacy.

PDE5 is the major cGMP-hydrolyzing PDE in platelets, lung, and the vascular smooth muscle (VSM) of the penile corpus cavernosum (PCC). Sildenafil (Viagra and Revatio), tadalafil (Cialis), vardenafil (Levitra) or udenafil (Zydena) blocks cGMP breakdown by PDE5. In the presence of ongoing cGMP synthesis, these inhibitors promote accumulation of intracellular cGMP and relaxation of the VSM of blood vessels in the lung and PCC. The relaxation of VSM increases blood flow and capacitance of the vascular structures in the PCC, thereby facilitating accumulation of blood and improved erectile function; relaxation of VSM in pulmonary blood vessels increases perfusion of the lung, thereby increasing oxygen exchange and decreasing back pressure on the right heart. Sildenafil has recently shown promise in treatment of cardiac hypertrophy resulting from pressure overload diseases involving the vasculature and other maladies.

PDE6

PDE6 occurs only in the pineal gland and outer segments of mammalian retinal photoreceptor cells. PDE6 in cone and

rod photoreceptor cells provides for the visual response to high- and low-intensity light, respectively. In both cell types, the PDE6 isoenzymes are highly abundant. Rods contain PDE6 $\alpha\beta$, a heterodimer of ~100-kDa subunits that is bound to two inhibitory $P\gamma$ subunits specific for the rod enzyme. Cones contain PDE6 $\alpha'\alpha'$, a homodimer of ~100-kDa subunits that is complexed with two inhibitory $P\gamma$ subunits specific for the cone enzyme. In the absence of light, the catalytic activities of PDE6 $\alpha\beta$ and PDE6 $\alpha'\alpha'$ are inhibited by the respective $P\gamma$ s. Activated PDE6 isoenzymes have high catalytic activity ($V_{\max}$ ~300 $\mu\text{mol min}^{-1} \text{mg}^{-1}$) and K_m ~35 μM . GAF-A in each PDE6 subunit provides for allosteric cGMP binding and a major portion of cGMP in the photoreceptors is bound to PDE6 isoenzymes.

In the dark, cGMP in photoreceptors is high and binds to a cGMP-gated cation channel, which then allows for increased conductance of a 'dark current'. When light strikes the retina, transducin, a G protein, is activated, and the α -subunit of transducin interacts with the $P\gamma$ in each PDE6 isoenzyme to relieve $P\gamma$ -mediated inhibition. The resulting burst of PDE6 activity causes a sudden drop in cGMP. At low cGMP, cGMP dissociates from the channel, which then closes and the 'dark current' is terminated. The resulting change in the membrane potential produces the sensation of light. PDE6 isoenzymes are the only PDEs that have been found thus far to be regulated by a G protein. Many other complex regulatory processes also contribute to PDE6 functions and provide for optimal visual response to a broad range of light intensities.

PDE9

Little is known about the tissue distribution or physiological role of PDE9. It is highly selective for cGMP and has a K_m of 0.1–2 μM for cGMP versus 230 μM for cAMP. It is inhibited by zaprinast (IC_{50} ~35 μM) but is insensitive to a wide range of other PDE inhibitors including sildenafil, vinpocetine, IBMX, and dipyridamole. Inhibitors that target PDE9 may prove to be useful in treatment of Alzheimer's disease.

Characteristics of Class I Dual-Specificity PDEs

A number of mammalian PDEs (PDEs 1–3, 10, and 11) hydrolyze both cGMP and cAMP. The physiological role of each PDE in hydrolyzing the two nucleotides is likely to be complex. Activities of these PDEs toward either cGMP or cAMP in cells will reflect the respective K_m and $V_{\max}$ values for each. Selective action toward one nucleotide can occur simply by mass action as the level of that cN increases in the cell. For example, if the cGMP/cAMP content of a tissue selectively changes, increased hydrolysis of one nucleotide (e.g., in response to elevated amounts of that nucleotide) would tend to decrease hydrolysis of the other nucleotide through simple competition for the catalytic site. As a result, the concentration of the second cN would increase and activate its signaling pathway.

PDE1

The PDE1 family is regulated in the short term by Ca^{2+} /calmodulin binding following elevation of Ca^{2+} derived primarily from extracellular sources, and in the long term by

changes in PDE protein level (Figure 3). PDE1 isoenzymes are widely expressed and are abundant in brain, heart, VSM, testis, macrophages, lymphocytes, and liver. PDE1 isoenzymes catalytic subunits occur as dimers, and each monomer is comprised of two calmodulin-binding sites, an autoinhibitory subdomain, and a catalytic domain. Ca^{2+} /calmodulin binding relieves autoinhibition and increases the $V_{\max}$ with no change in K_m . Activities of the PDE1 family are implicated in diverse processes that include regulation of smooth muscle proliferation, learning and memory, cardiac hypertrophy, and olfaction.

The PDE1 isoenzymes (PDE1A, 1B, and 1C) are derived from three genes and are represented in tissues by many splice variants that substantially differ in size. Among these isoenzymes, the relative affinities and catalytic rates for cGMP and cAMP vary considerably. Affinities for certain inhibitors and sensitivities to stimulation by Ca^{2+} /calmodulin also vary. In tissues containing primarily one isoenzyme, the relative hydrolytic contribution of the PDE1 family to breakdown of cAMP or cGMP will reflect the kinetic features of this dominant PDE1. PDE1 isoenzymes are primarily cytosolic, but they are also found in association with the plasma membrane and other particulate cellular components. Depending on location in the cell, members of the PDE1 family can differentially respond to selective changes in Ca^{2+} signaling and thereby provide for variations in the spatial and temporal changes in cGMP and/or cAMP levels. Inhibitors that show selectivity for PDE1 isoenzymes include vinpocetine, IC224, and SCH51866.

PDE2

PDE2 is known as the cGMP-stimulated PDE, and its three variants are derived from a single gene. It is a homodimer comprised of ~105-kDa monomers and is expressed in many tissues including brain, heart, liver, adrenal gland, macrophages, vascular endothelial cells, and platelets. $V_{\max}$ values for hydrolysis of cAMP and cGMP are similar ($150 \mu\text{mol min}^{-1} \text{mg}^{-1}$) and K_m values differ nearly twofold ($15\text{--}30 \mu\text{M}$ for cGMP vs. $30\text{--}50 \mu\text{M}$ for cAMP). cGMP binds to GAF-B in the R domain and stimulates breakdown of either cAMP or cGMP at the catalytic site. PDE2 can, therefore, potentially lower either or both cNs. PDE2 is abundant in adrenal cortex cells, which produce aldosterone in response to cAMP elevation. In these cells, elevation of cGMP results in cGMP binding to the PDE2 GAF-B, causing increased cAMP hydrolysis and decreased aldosterone production. A similar physiological effect to increase cGMP hydrolysis by cGMP binding has not been experimentally demonstrated, but it is likely to occur and would provide a negative feedback pathway for regulation of cGMP level. PDE2 has also been shown to be important in regulation of cGMP pools near the plasma membrane in cardiomyocytes and is implicated along with PDE3 in control of permeability of the vascular endothelium.

PDE3

PDE3 is abundant in many tissues and occurs as two groups of isoenzymes (PDE3A and 3B) that are largely membrane bound and exhibit ~80% amino acid sequence identity in their C domains. Both hydrolyze cAMP with a greater $V_{\max}$ than that for cGMP, but there is little difference in K_m values for the cNs. PDE3B is activated by phosphorylation in its R domain. Historically, the moniker for PDE3 was the

'cGMP-inhibited PDE', but this is a misleading descriptor because cGMP 'inhibits' cAMP hydrolysis by competing for access to the catalytic site. The action of PDE3 isoenzymes is implicated in myriad physiological processes including cardiovascular functions, lipolysis, carbohydrate metabolism, insulin secretion, and platelet disaggregation. In some tissues, evidence suggests that some effects of cGMP may be mediated, at least in part, through increased cAMP signaling due to cGMP competition with cAMP hydrolysis by PDE3. Elevation of cGMP would increase its effectiveness in competing with cAMP for the PDE3 catalytic site, thereby 'protecting' cAMP from hydrolysis.

The catalytic activities of PDE3B isoenzymes are activated by phosphorylation by PKA as well as by the $\text{PI}_3\text{K}/\text{PKB}$ pathways in response to a number of hormone signals. This activation produces a negative feedback loop to blunt or terminate elevations in cAMP and cAMP-signaling. In addition to this negative feedback role, PDE3 is activated by insulin and other signaling molecules by stimulation of the $\text{PI}_3\text{K}/\text{PKB}$ pathway in adipose tissue and liver, thereby lowering cAMP level. This results in inhibition of adipose tissue lipolysis or liver glycogenolysis and gluconeogenesis. PDE3 is physiologically important in many tissues including heart, platelets, oocytes, smooth muscle, liver, and adipocytes. There are a number of inhibitors (cilostazol, milrinone, and amrinone) that are selective for PDE3 isoenzymes; cilostazol (Pletal) is used clinically for treatment of intermittent claudication and milrinone (Primacor) is used for treatment of heart failure.

PDE10

Little is known about the physiological functions or regulation of PDE10. PDE10 isoenzymes (PDE10A1-4) are derived from one gene and are most abundant in neural tissues including the striatal area, hippocampus, cerebellum, and thalamus; it may play a particularly prominent role in striatal medium spiny neurons. Transcripts are also found in non-neural tissues including testis, thyroid, pituitary, heart, prostate, liver, and skeletal muscles. It has higher affinity for cAMP versus cGMP (K_m values ~0.05 and $3 \mu\text{M}$, respectively), but the $V_{\max}$ for cGMP is ~4 times that for cAMP. PDE10 R domain contains GAFs, and cAMP binds tightly to GAF-B. However, a regulatory function for this association in PDE10 holoenzyme has not been established. PDE10A2 is phosphorylated by PKA at Thr-16; this promotes relocation of PDE10 from the Golgi to the cytosol. Based on its abundance and location in the brain as well as changes in various disease models, PDE10 has been considered as a target for treatment of dysfunctions such as schizophrenia, obsessive-compulsive disorders, addictions, Parkinson's disease, and Huntington's disease. It is potently inhibited by papaverine ($\text{IC}_{50} \sim 30 \text{ nM}$) as well as by the nonspecific inhibitors 3-isobutyl-1-methylxanthine (IBMX), dipyridamole, and zaprinast.

PDE11

PDE11 is closely related to PDE5 (50% homology). Very little is known about its physiological function or biochemical regulation. There are four isoenzymes (PDE11A1-4) derived from one gene; PDE11A4 is the only form detected thus far in

human tissues and is most abundant in prostate, skeletal muscle, pituitary, heart, and liver. The R domains of PDE11 isoenzymes contain one, two, or partial GAFs, and cGMP binds with high affinity to GAF-A. However, no regulatory role has been established for this interaction. PDE11 hydrolyzes cGMP and cAMP at similar rates, but K_m for cAMP is slightly lower than that for cGMP (0.5 vs. 1 μ M, respectively). The most potent inhibitor of PDE11 is tadalafil (Cialis) ($IC_{50} \sim 73$ nM); it is also inhibited by dipyrindamole ($IC_{50} \sim 400$ –900 nM), vardenafil ($IC_{50} \sim 840$ nM) and sildenafil ($IC_{50} \sim 3800$ nM).

Characteristics of Class II PDEs

Class II PDEs have now been identified in various lower eukaryotes, fungi, and bacteria and based on analysis of the sequences of their catalytic domains have been subdivided into two related groups of PDEs. They share no apparent homology with class I PDEs, and the enzymatic characteristics of most of these enzymes have not been studied. The catalytic domains of the two subgroups vary in length, and the amino acid sequence of one of the groups resembles that of the metallo- β -lactamases. No X-ray crystallographic structures are available, and studies on the effects of site-directed mutagenesis on enzyme functions are limited. The catalytic domains of class II PDEs contain a Zn^{2+} -binding motif that is different from the Zn^{2+} -binding motif in class I PDEs, and this motif is thought to provide critical features for hydrolysis of cN. This class of PDEs can hydrolyze cGMP and cAMP and, in some instances, have catalytic action on polynucleotides.

Catalytic Activity

Like class I PDEs, class II PDEs require divalent cation(s) for catalytic activity and exhibit varying selectivities for cGMP and cAMP as substrates. Catalytic activity in a class II PDE from *Saccharomyces cerevisiae* has been shown to require Zn^{2+} . Class II PDEs that are known to hydrolyze cGMP include a dual-specificity PDE from *Vibrio fischeri* that is expressed in the periplasmic space and hydrolyzes cAMP and cGMP with very high turnover rates; dual-specificity PDEs from *Dictyostelium discoideum* (DdPDE1, DdPDE6, and DdPDE7); and a cGMP-specific PDE from *Dictyostelium discoideum* (DdPDE5), which is the major cGMP-degrading enzyme in the intact cell (Figure 4). Many of these PDEs are insensitive to IBMX, a nonspecific inhibitor of most class I PDEs. DdPDE1 in *Dictyostelium* can be exposed on the cell surface or secreted into the medium; although it is a dual-specificity PDE, its main function is apparently to hydrolyze cAMP in the extracellular milieu.

Regulation

Class II PDEs can be regulated by short- and long-term processes. The catalytic activities of two *Dictyostelium* PDEs (DdPDE5 and DdPDE6) are increased by cN binding to site (s) in their R domains (Figure 4). In DdPDE5, catalytic activity is stimulated by allosteric cGMP binding, whereas allosteric binding of either cGMP or cAMP stimulates catalysis by

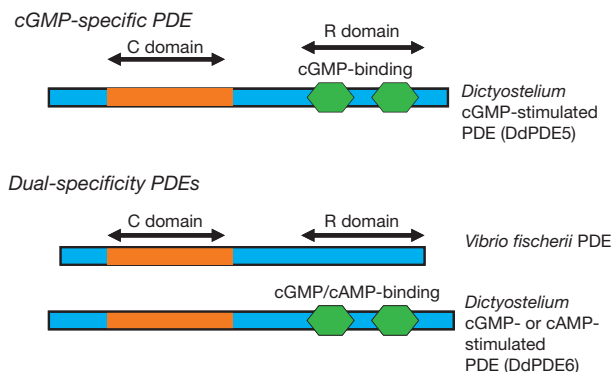


Figure 4 Schematic depiction of select class II cGMP-hydrolyzing PDEs. Arrangements of functional domains are shown for a monomer. The allosteric cN-binding sites belong to the CAP-related family of cN-binding proteins and are distinct from the allosteric cGMP-binding sites in class I PDEs.

DdPDE6. This resembles the stimulatory effect of cGMP on catalysis in class I PDEs 2 and 5. However, GAFs are absent in class II PDEs. Instead, both contain two CAP-related cN-binding domains that are located C-terminal to the catalytic domain (Figure 4). Thus, despite no evolutionary relatedness between class I and II PDEs, similar regulatory mechanisms are used, and, remarkably, evolutionarily distinct motifs and arrangements of the motifs in the structures have been employed. These developments are striking examples of convergent evolution. Moreover, regulation of the expression both classes of PDEs occurs throughout the growth cycle and aggregate phases of *Dictyostelium*.

Concluding Remarks

cGMP levels are physiologically modulated through the action of cGMP-specific PDEs and dual-specificity (cGMP and cAMP) PDEs. Both must be considered when studying physiological changes in intracellular cGMP. Two evolutionarily unrelated classes of these enzymes exist, but only class I is present in mammals. Several mammalian PDEs have been successfully targeted by new medications for treatment of various diseases.

See also: [Signaling: Cyclic AMP Receptors of *Dictyostelium*](#); [Cyclic Nucleotide Phosphodiesterases](#); [Cyclic Nucleotide-Dependent Protein Kinases](#); [Cyclic Nucleotide-Regulated Cation Channels](#).

Further Reading

- Bader S, Kortholt A, and Van Haastert JM (2007) Seven *Dictyostelium discoideum* phosphodiesterases degrade three pools of cAMP and cGMP. *Biochemical Journal* 402: 153–161.
- Beavo JA, Francis SH, and Houslay MD (eds.) (2007) *Cyclic Nucleotide Phosphodiesterases in Health and Disease*. Boca Raton, FL: CRC Press.
- Bender AT and Beavo JA (2006) Cyclic nucleotide phosphodiesterases: Molecular regulation to clinical use. *Pharmacological Reviews* 58: 488–520.
- Bosgraaf LH, Russcher H, Snippe H, et al. (2002) Identification and characterization of two unusual cGMP-stimulated phosphodiesterases in *Dictyostelium*. *Molecular Biology Cell* 13: 3878–3889.

- Conti M and Beavo J (2006) Biochemistry and physiology of cyclic nucleotide phosphodiesterases: Essential components in cyclic nucleotide signalling. *Annual Reviews of Biochemistry* 76: 481–511.
- Corbin JD and Francis SH (1999) Cyclic GMP phosphodiesterase-5: Target of sildenafil. *Journal of Biological Chemistry* 274: 13729–13732.
- Francis SH, Corbin JD, and Bischoff E (2009) Cyclic GMP-hydrolyzing phosphodiesterases. *Handbook of Experimental Pharmacology* 191: 409–421.
- Goraya TA and Cooper DMF (2005) Ca^{2+} -calmodulin-dependent phosphodiesterase (PDE1): Current perspectives. *Cellular Signaling* 17: 789–797.
- Kawamura S and Tachibanaki S (2008) Rod and cone photoreceptors: Molecular basis of the difference in their physiology. *Comparative Biochemistry and Physiology. Part A. Molecular and Integrative Physiology* 150: 369–377.
- Ke H and Wang H (2007) Crystal structures of phosphodiesterases and implications on substrate specificity and inhibitor selectivity. *Current Topics in Medicinal Chemistry* 7: 391–403.
- Martinez SE, Wu AY, Glavas NA, et al. (2002) The two GAF domains in phosphodiesterase 2A have distinct roles in dimerization and in cGMP binding. *Proceedings of the National Academy of Sciences of the United States of America* 99: 13260–13265.
- Murata T, Shimizu K, Hiramoto K, and Tagawa T (2009) Phosphodiesterase 3 (PDE3): structure, localization and function. *Cardiovascular and Hematological Agents in Medicinal Chemistry* 7: 206–211.
- Xu RX, Hassell AM, Vanderwall D, et al. (2000) Atomic structure of PDE4: Insights into phosphodiesterase mechanism and specificity. *Science* 288: 1822–1825.
- Yarfitz S and Hurley JB (1994) Transduction mechanisms of vertebrate and invertebrate photoreceptors. *Journal of Biological Chemistry* 269: 14329–14332.
- Zoraghi R, Francis SH, and Corbin JD (2007) Critical amino acids in phosphodiesterase-5 catalytic site that provide for high-affinity interaction with cyclic guanosine monophosphate and inhibitors. *Biochemistry* 46: 13554–13563.

Cyclic Nucleotide-Dependent Protein Kinases

S H Francis and J D Corbin, Vanderbilt University School of Medicine, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Autoinhibition The effect of one region of an enzyme blocking the action of the catalytic site of another region of that same enzyme.

Autophosphorylation Protein kinase-catalyzed transfer of the γ -phosphate of adenosine triphosphate (ATP) to a residue within the protein kinase itself.

Consensus phosphorylation sequence The amino acid sequence that contains the determinants that provide for the

specific phosphorylation of that sequence by a particular protein kinase.

Heterophosphorylation Protein kinase-catalyzed transfer of the γ -phosphate of ATP to another protein substrate.

Pseudosubstrate sequence An amino acid sequence that closely resembles the consensus phosphorylation sequence for a protein kinase substrate but lacks the residue that can be phosphorylated.

Introduction

Upon elevation of 3',5'-cyclic monophosphate (cAMP) or guanosine 3',5'-cyclic monophosphate (cGMP) in cells, these nucleotides bind to numerous proteins including the cyclic nucleotide-dependent protein kinases, cyclic nucleotide-gated channels, and cyclic nucleotide phosphodiesterases (PDEs); cAMP also binds to exchange proteins that are activated by cAMP (Epacs). The actions of these target proteins elicit a panoply of biological changes in cells, but the cAMP- and cGMP-dependent protein kinases (PKAs and PKGs, respectively) play a central role in mediating the effects of the cyclic nucleotides. Intracellular levels of cAMP and cGMP are primarily determined by the balance between their synthesis and breakdown. They are synthesized from either adenosine triphosphate (ATP) or guanosine triphosphate (GTP) by adenylyl cyclases or guanylyl cyclases, respectively, in response to first-messenger signals in the body (e.g., hormones, neurotransmitters, and environmental stimuli), and they are broken down by cyclic nucleotide PDEs. The important concept of second-messenger signaling as occurs with cAMP and cGMP is that many hormones, neurotransmitters, or other stimuli can elicit their physiological effects without entering the cell. Second-messenger signaling was first demonstrated for cAMP.

even slight variations in the cellular concentration of cAMP or cGMP can alter the activation state of the respective kinase significantly.

In addition to their breakdown by PDEs, both nucleotides can be extruded from cells. The action of PDEs is the major mechanism for rapid lowering of cyclic nucleotides in the cell (Figure 2). Following activation by cyclic nucleotide binding, PKA or PKG then catalyzes transfer of the γ -phosphate of ATP to selected serines or threonines that are located in a consensus sequence in many cellular proteins; this process is known as phosphorylation. The covalent attachment of phosphate to these proteins converts them into phosphoproteins, a process known as heterophosphorylation, because in this case PKA or PKG phosphorylates proteins other than itself. When PKA or PKG catalyzes the covalent attachment of a phosphate to sites within its own structure, this is known as autophosphorylation. The phosphoproteins produced by these modifications frequently exhibit changes in activities, subcellular localizations, and interactions with other cellular components, including other proteins, DNA, RNA, and lipids. The altered functions of these proteins that have been phosphorylated by PKA or PKG mediate the changes induced by many hormones, neurotransmitters, and other stimuli, including those in metabolism, gene transcription, neurotransmission, blood pressure, and fluid homeostasis.

Cyclic Nucleotide-Dependent Protein Kinases as Intracellular Receptors for cAMP and cGMP

The molecular structures of cAMP and cGMP are very similar (Figure 1), but the structural distinctions between these compounds provide for selective binding to and activation of PKA and PKG, respectively, which then mediate many of the biological actions of these two second messengers (Figure 1). PKA and PKG belong to the family of ABC kinases, are closely related, and share many functional and regulatory characteristics. In the presence of small amounts of cAMP or cGMP (e.g., concentrations ranging from 10^{-8} to 10^{-7} M), these kinases are activated by binding the cyclic nucleotide so that

Overall Structures of Cyclic Nucleotide-Dependent Protein Kinases

There are several isoforms of PKA and PKG, but all have a similar organization of their functional domains (Figure 3). Both are chimeric proteins comprising a regulatory domain and a catalytic domain, and within each of these there are several subdomains that provide for specific functions. PKA regulatory and catalytic domains are located on separate subunits, the regulatory (R) and catalytic (C) subunits, respectively. In PKG, the regulatory and catalytic domains are on a single polypeptide chain. Phosphorylation at a threonine in the active sites of PKG and PKA C subunit is required for

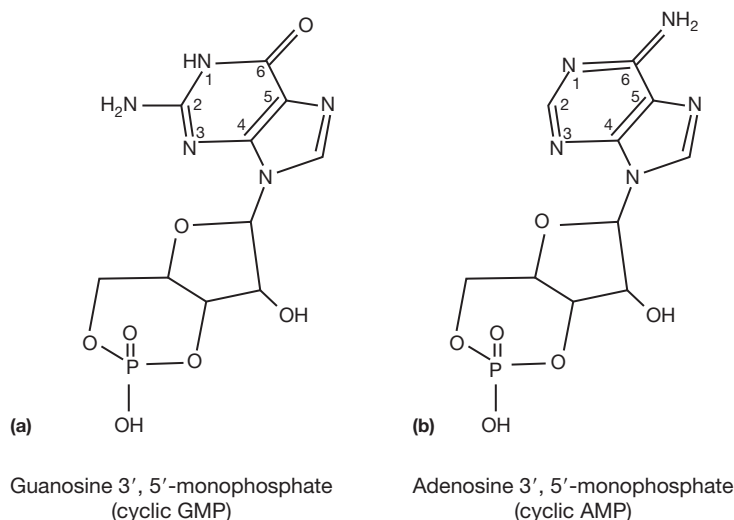


Figure 1 Structural models of (a) cGMP and (b) cAMP. cGMP and cAMP differ only in their purine rings. These differences in the purine ring provide for the specificity with which each kinase interacts preferentially with that nucleotide. The cyclic phosphate moiety is required for binding to the cyclic nucleotide-binding sites on PKG and PKA, and the 2'-OH of the ribose provides an important contact in the respective sites. There are two conformations of these nucleotides, *syn* (shown) and *anti*, due to rotation around the bond linking the purine and the ribose. Both PKA and PKG bind the *syn* conformer. Important features of cGMP that contribute to its interaction with PKG include the amino group at C-2 (which interacts with a conserved threonine in sites preferring cGMP), the oxygen at C-6, and a protonated N-1. As compared with cGMP, cAMP has a single substitution on the purine (i.e., an amino group at C-6 and no protonation at N-1).

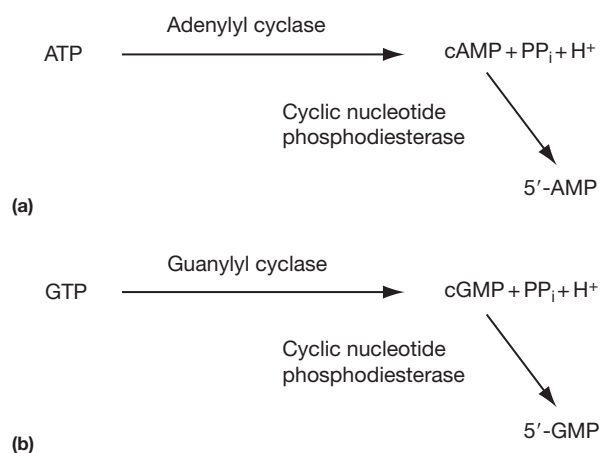


Figure 2 The balance between cyclase and phosphodiesterase activities determines cyclic nucleotide levels in the cell. cAMP and cGMP are synthesized from either (a) ATP by adenylyl cyclase or (b) GTP by guanylyl cyclase; when the activities of these cyclases are increased, more cAMP or cGMP is produced. The breakdown of the cyclic nucleotides is catalyzed by cyclic nucleotide phosphodiesterases (PDEs), and the activities of these enzymes lower cAMP and cGMP.

activity. Both families of kinases form dimers through interactions among residues near the N terminus within their regulatory domains. However, the salient features of regulation and catalysis are largely retained in monomeric forms of PKG and in PKA formed by the combination of a single R subunit and C subunit.

Although dimerization is not necessary for many PKA functions, it is required for anchoring a portion of the PKA to

specific subcellular compartments through a family of proteins known as A-kinase anchoring proteins (AKAPs). In many instances, anchoring of PKA brings it into close proximity with substrates, thereby facilitating rapid and efficient cAMP signaling. Likewise, dimerization of PKGs is not necessary for many of PKG-mediated functions; however, in some instances, the N-terminal dimerization domain of PKGs provides for specific interaction with certain substrates allowing for PKG-mediated phosphorylation. PKA R subunits are aligned in an antiparallel arrangement within the dimer and use hydrophobic interactions, whereas PKG monomers are aligned in a parallel arrangement within the dimer and use interactions formed by an extended leucine zipper motif.

cAMP-Dependent Protein Kinase

PKA is present in the cytosolic and particulate fraction of all mammalian tissues and occurs at concentrations ranging from 0.2 to 2 μM . It is a tetramer comprising two C subunits and two R subunits (R₂C₂) (Figure 3). There are three genes for mammalian C subunit (C α , C β , and C γ), each is $\sim 36\,000$ Da. There are four major genes for R subunits (RI α , RI β , RII α , and RII β), each is $\sim 45\,000$ Da. Nomenclature for PKAs is based on the R subunits; PKAs containing RI subunits are called type I PKA, and those containing RII subunits are called type II PKA. Both types may contain either form of C subunit. Most tissues contain both types of PKA, and while either type can frequently mediate cAMP action, selective colocalization of one form with a particular substrate may be required. Type II PKA is abundant in the cytosol and is also commonly anchored in membrane compartments. Although type I PKA is also anchored in some instances, it is

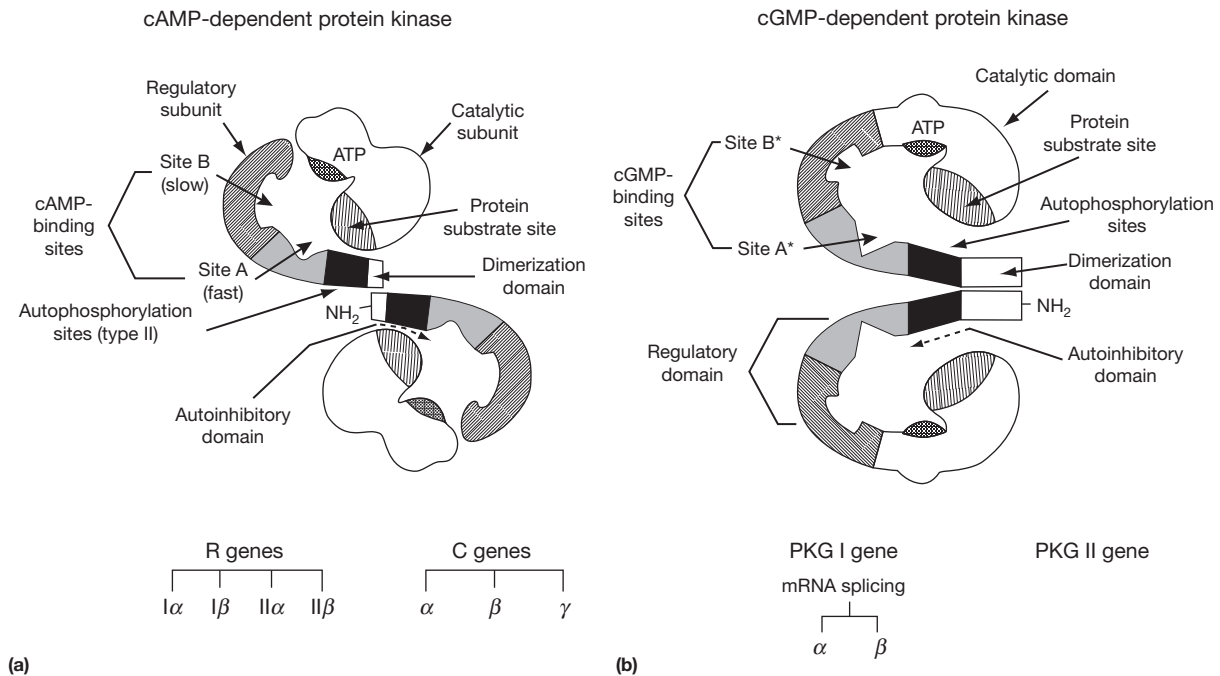


Figure 3 Working models of (a) PKAs and (b) PKGs, depicting the overall organization of the functional domains within these enzymes. The major isozymic forms that are products of different genes for each of the kinases are listed below each model. C_γ is the product of a pseudogene that is not present in all species. The alignment of the low-affinity and high-affinity cyclic nucleotide sites within PKA R subunit is the same for all known PKAs. The linear arrangement of the cGMP-binding sites in PKG-II is the same as that found in PKAs. However, in PKG-I, the sites are transposed so that the more N-terminal site is the high-affinity cGMP-binding site and the more C-terminal position is the low-affinity site.

predominantly cytosolic. The RI and RII content of a particular tissue varies among species.

In the absence of cAMP, an R subunit interacts with a C subunit with high affinity (~ 0.2 nM) and suppresses its catalytic activity through a process known as autoinhibition. Cyclic AMP binding to two cAMP-binding sites on an R subunit causes a large conformational change that decreases the R subunit affinity for C subunit by 10 000- to 100 000-fold. An active monomeric C subunit dissociates from the complex and can diffuse throughout the cell, including the nucleus, which excludes R subunits and PKA holoenzymes. The diffusibility of the C subunit provides for the mechanism by which cAMP elicits its effect in a number of systems. The R subunits remain dimerized and, if anchored to AKAPs, they remain anchored.

X-ray crystallographic structures of the isolated C subunit, R subunits, and complexes between R and C subunits have provided myriad molecular insights into structure/function relationships of PKAs. Moreover, these results have advanced understanding of all kinases that are related to PKA. The catalytic site of the C subunit is formed at the interface of two lobes where the γ -phosphate of ATP is transferred to the protein substrate. The smaller, more N-terminal lobe makes most of the contacts with Mg^{2+} /ATP. The larger, more C-terminal lobe makes most of the contacts that precisely position the protein substrate in the catalytic cleft and also forms extensive contacts with the R subunit. The binding of Mg^{2+} /ATP and a protein substrate into the catalytic cleft allows a complex set of enzymatic actions that act in concert to rapidly transfer the phosphate from ATP to substrate. Release of the product, adenosine diphosphate (ADP), is rate limiting in most cases.

In order to efficiently transfer the phosphate to a peptide or protein, PKA strongly prefers an amino acid sequence of -ArgArgXSer/ThrX- (called a PKA consensus phosphorylation site) in its substrates. The sequence at the phosphorylation site provides in large part for the efficiency with which the C subunit phosphorylates a protein, but co-localization of PKA and some substrates via interactions with AKAPs also contribute importantly. Although arginines (Arg) are highly preferred in the two indicated positions, some variation is tolerated. The amino acids in the positions marked as X can affect the affinity of the C-subunit interaction with a substrate.

Each R subunit contains several subdomains. These include an N-terminal dimerization/docking subdomain, which includes the autoinhibitory and autophosphorylation regions, and a cyclic nucleotide-binding subdomain that contains two cAMP-binding sites (A and B) arranged in tandem (Figure 3). In the absence of cAMP, a substrate-like sequence in the R subunit binds tightly to the catalytic site in the C subunit. Although other contacts also contribute to stabilizing the R-C interaction, this substrate-like sequence in the autoinhibitory domain competes with protein substrate binding and provides a major portion of the interactions that hold PKA in a catalytically inactive form in the absence of cAMP. This is known as autoinhibition. The binding of cAMP to both cAMP-binding sites in each R subunit disrupts the R-C interaction.

Most commonly, the PKA tetramer contains a single type of R subunit (either RI or RII). RII subunits contain a sequence in the autoinhibitory domain that mimics a PKA substrate motif, and the serine in this sequence can be autophosphorylated by the C subunit when the subunits are complexed (i.e., in the

absence of cAMP). Autophosphorylation of this site in RII increases its affinity for cAMP so that PKA can be activated at lower cAMP concentrations. Type I R subunits contain a pseudosubstrate sequence (-ArgArgXGlyX-) that is a substrate-like sequence lacking a phosphorylatable residue. Therefore, type I PKA R subunits cannot undergo autophosphorylation.

PKA plays a critical role in functions of all mammalian cells. As a result, it is likely that survival requires the presence of the PKA C subunit. A null mutation for all three C subunits in yeast is lethal. Mice carrying null mutations for either *C α* or *C β* are viable, but differ phenotypically; mice lacking either RI or RII subunits are also typically viable and differ phenotypically. These results suggest that these C subunits and R subunits have selectivity in some functions.

cGMP-Dependent Protein Kinase

PKGs play a central role in regulating smooth muscle tone, platelet aggregation, bone growth, neuronal plasticity, and water-salt homeostasis. Cellular concentration of PKG ranges from infinitesimally low to 4 μ M. There are two families of PKG (PKG-I and PKG-II) that are the products of separate genes. These families are rarely co-expressed and have selective functions because the phenotype of PKG-I-null mice differs substantially from that of PKG-II-null animals. PKG-I has two alternative splice variants (PKG-I α and PKG-I β) that vary in the N terminal ~100 amino acids and are frequently co-expressed. PKG-I is largely cytosolic, whereas PKG-II is membrane-bound through an N-terminal myristoyl group. The membrane association of PKG-II co-localizes PKG-II with intended substrates, thereby providing for specific phosphorylations that do not occur when PKG-II is in its soluble form. The PKG-I isoenzymes, PKG-I α and PKG-I β , are largely cytosolic, but they too are specifically localized in some cases. Where this has been studied, localization for PKG-I isoenzymes is selectively mediated through interactions involving their respective unique leucine zipper motifs and specific sites in other proteins, for example, the myosin-binding protein of myosin light chain phosphoprotein phosphatase. In many cases, the interaction is selective for either PKG-I α or PKG-I β , indicating that the specificity of the interaction involves sequences that are novel to that isoenzyme.

PKG regulatory domains contain several subdomains including an N-terminal dimerization subdomain involving a leucine zipper motif, a subdomain containing the overlapping autoinhibitory and autophosphorylation regions, and a cyclic nucleotide-binding subdomain that contains two cGMP-binding sites arranged in tandem (Figure 3). The role of dimerization is unclear because, at least for PKGI, the monomeric enzyme retains the salient features of the dimer. However, interactions within the PKGI leucine zipper that provide for dimerization also stabilize structural features in this region that confer PKGI localization with some substrates and may afford catalytic advantages for dimeric PKGI.

The autoinhibitory-autophosphorylation subdomain is located just C-terminal to the dimerization subdomain. This region accounts for autoinhibition of the catalytic site and contains multiple autophosphorylation sites, most of which do not resemble a consensus substrate sequence and are not

conserved among PKGs. Autophosphorylation of these sites is catalyzed by the catalytic domain within the same monomer and increases affinity for cGMP and basal activity. Autophosphorylation increases in the presence of cyclic nucleotide and promotes PKG ubiquitination and degradation through the proteosomal degradation system.

The mechanism of autoinhibition of PKGs also differs somewhat from that of PKA. The autoinhibitory subdomain of all PKGs contains either pseudosubstrate sequences, for example, -LysArgGlnAlaIle- in PKG-I β , or sequences that only weakly mimic a pseudosubstrate site, for example, -ArgAlaGlnGlyIle- in PKG-I α , where the italicized residue indicates the phosphorylation position in the autoinhibitory sequence. Furthermore, sequences outside the pseudosubstrate sequence contribute more prominently to the autoinhibition of both PKG-I and PKG-II compared to PKA.

Each of the two cGMP-binding sites of PKG are comprised of ~110 amino acids arranged in tandem. A small segment of protein connects the regulatory and catalytic domains of the enzyme. The structure of the catalytic domain is thought to closely resemble that of PKA because the enzymes are homologous. However, the two enzymes have numerous differences in regulatory mechanisms that are most likely due to differences in their overall structures. PKG catalytic activity, which is latent in the absence of cGMP, is increased upon cGMP binding, which causes a marked elongation of the PKG monomer and relief of autoinhibition.

Cyclic Nucleotide-Binding Specificity and Affinity

The two homologous cyclic nucleotide-binding sites (~110 amino acids each) in PKA and PKG are products of an ancient gene duplication that occurred prior to the divergence of these proteins. The sites are evolutionarily related to a bacterial cAMP-binding protein, the catabolite-gene-activating protein (CAP) family of cyclic nucleotide-binding proteins. The cyclic nucleotide-binding sites of cyclic nucleotide-gated channels and exchange proteins directly activated by cAMP (Epacs) are also members of this family. In both PKA and PKG, the two intra-subunit sites differ approximately 10-fold in their affinities for cyclic nucleotide binding and also differ in specificities for cyclic nucleotide analogs. In PKA RI and RII subunits, the more N-terminal site has a lower affinity for cAMP (fast cAMP dissociation) compared to the more C-terminal site (slow cAMP dissociation) (Figure 3). In PKG-II, the kinetically different cGMP-binding sites are arranged as in PKA, but in PKG-I the high-affinity site is the more N-terminal site. Full activation of either PKA or PKG requires saturation of all the cyclic nucleotide-binding sites.

PKA and PKG have a 50- to 200-fold selectivity for cAMP and cGMP, respectively, and most commonly the cellular effect of cAMP or cGMP is mediated by the respective kinase. However, in several systems, cAMP or cGMP has been shown to activate the other kinase through a process known as cross-activation. In the X-ray crystallographic structure of the PKA R subunit, cAMP is bound in a deep pocket within each site, and several conserved amino acids (glutamate, arginine, and several glycines) provide direct contact with the cyclic nucleotide or contribute to critical structural features of the site. In sites

preferring cGMP, an invariant threonine provides a major portion of the selectivity for cGMP by interacting with the 2'-amino group in the purine (Figure 1), but an X-ray crystal structure has not been determined for PKGs. Hydrophobic amino acids in each of the sites contribute importantly to the affinity with which cyclic nucleotide is bound.

Substrate Specificities of PKA and PKG

PKA and PKG-I share many similarities in substrate specificities, and the consensus sequence for phosphorylation by either enzyme is defined as -ArgArgXSer/ThrX-. These kinases can frequently phosphorylate the same proteins *in vitro*, although there are differences. PKG-I phosphorylates numerous proteins in sites that do not conform to this ideal consensus phosphorylation sequence; these sites are not phosphorylated well by PKA. A phenylalanine located C-terminal to the phosphorylation site strongly discriminates against PKA phosphorylation, but not that of PKG, and a basic amino acid C-terminally adjacent to the phosphorylation site favors phosphorylation by PKG-I. The substrate specificity of PKG-II is quite different from those of PKG-I and PKA.

Negative Feedback and Feed-Forward Control of Cyclic Nucleotide Pathways

A two- to fourfold increase in either cAMP or cGMP typically accounts for the physiological responses to these nucleotides, although larger increases can occur in some instances. Negative feedback control mechanisms restrict the magnitude or persistence of cyclic nucleotide signaling, whereas feed-forward mechanisms potentiate the response to cyclic nucleotide elevation. Both types of mechanisms have been established for the cAMP- and cGMP-signaling cascades (Figure 4). First, PKA and PKG can phosphorylate and activate PDEs, which accelerates the breakdown of cyclic nucleotides, thereby favoring inactivation of the kinases. Second, when the beta 2-adrenergic receptor is bound to an agonist such as epinephrine, it is rapidly phosphorylated by PKA, as well as by the beta-adrenergic receptor kinase (BARK), resulting in desensitization to further stimulation by agonists. For cAMP signaling, both processes provide for a powerful negative feedback response to blunt the effects of elevated cAMP. Elevation of cGMP promotes autophosphorylation of PKGI α , which targets the enzyme for ubiquitination and degradation, thereby blunting the cGMP-signaling in those cells.

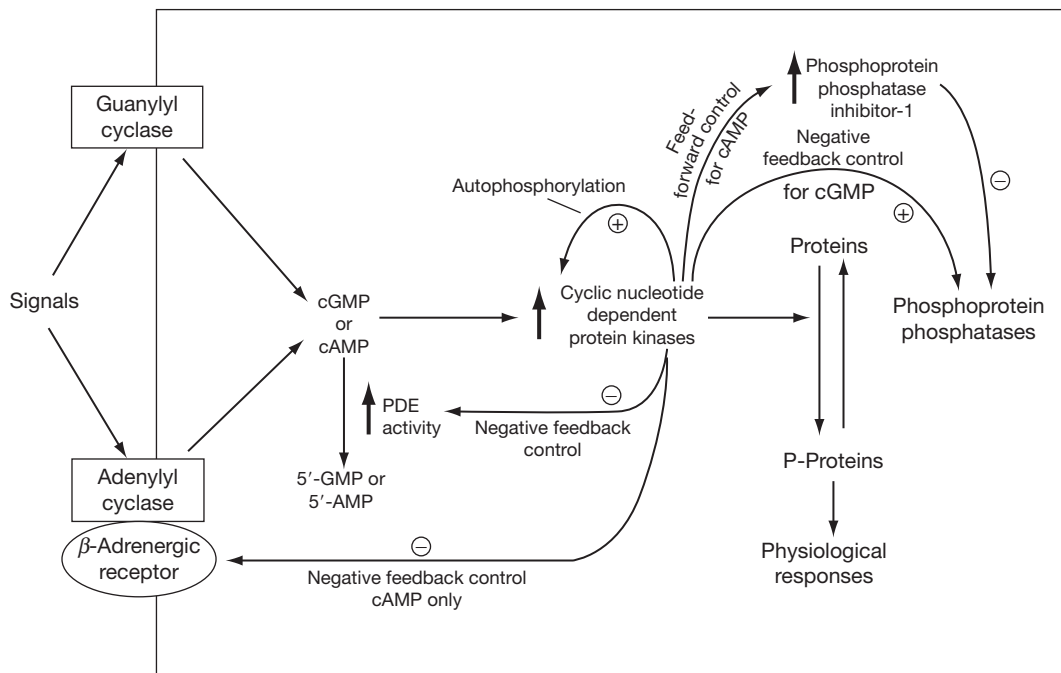


Figure 4 Mechanisms involving phosphorylation that contribute to modulation of the cAMP- and cGMP-signaling pathways. Feedback control refers to processes that tend to alter the overall effectiveness of signaling through cyclic nucleotide pathways. Negative feedback control mechanisms decrease the effectiveness of the cyclic nucleotide signal; these include increased activities of cyclic nucleotide phosphodiesterases (PDE activity) that break down the cAMP and cGMP, decreased cAMP synthesis by adenylyl cyclases that are coupled to G proteins, and increased activities of phosphoprotein phosphatases that remove phosphates from phosphoproteins that have been phosphorylated by PKG. Feed-forward control (i.e., positive feedback control) mechanisms enhance the effectiveness of the signal. These include autophosphorylation by PKA and PKGI, which increases the affinity of these enzymes for the respective cyclic nucleotides as well as increasing catalytic activity. In addition, PKA increases phosphorylation of the phosphoprotein phosphatase inhibitor-1, which inhibits a phosphoprotein phosphatase and blocks its effect of removing phosphate from phosphoproteins involved in the signaling pathway. Circled + and – denote effects that either increase or decrease as the levels of cyclic nucleotides increase or decrease; arrows, ↑ and ↓, denote a change in the catalytic activity of the indicated enzymes that is induced by PKA- or PKG-mediated phosphorylation or the state of phosphorylation of the phosphoprotein phosphatase inhibitor-1 in response to increased PKA activity.

Both the cAMP- and cGMP-signaling cascades also have feed-forward control because the autophosphorylation of type II PKA or PKG-I increases kinase activity and cyclic nucleotide-binding affinity. The cAMP system has an additional feed-forward control because PKA phosphorylates phosphoprotein phosphatase inhibitor-1 to produce a potent inhibition of phosphoprotein phosphatase action, thereby potentiating the effects of PKA phosphorylation.

See also: [Signaling: Adenylyl Cyclases; Cyclic Nucleotide Phosphodiesterases; Nitric Oxide Signaling.](#)

Further Reading

- Francis SH and Corbin JD (1999) Cyclic nucleotide-dependent protein kinases: Intracellular receptors for cAMP and cGMP action. *Critical Reviews of Clinical Laboratory Science* 52: 275–328.
- Hofmann F, Bernhard D, Lukowski R, and Weinmeister P (2009) cGMP regulated protein kinases (cGK). *Handbook of Experimental Pharmacology* 191: 137–162.
- Hofmann F, Feil R, Kleppisch T, and Schlossmann J (2006) Function of cGMP-dependent protein kinases as revealed by gene deletion. *Physiological Reviews* 86: 1–23.
- Kim C, Cheng CY, Saldanha SA, and Taylor SS (2007) PKA-I holoenzyme structure reveals a mechanism for cAMP-dependent activation. *Cell* 130: 1032–1043.
- Lincoln TM, Wu X, Sellak H, Dey N, and Choi CS (2006) Regulation of smooth muscle cell phenotype by cyclic GMP and cyclic GMP-dependent protein kinase. *Frontiers in Bioscience* 11: 356–367.
- Lohmann SM, Vaandrager AB, Smolenski A, Walter U, and De Jonge HR (1997) Distinct and specific functions of cGMP-dependent protein kinases. *Trends in Biochemical Science* 22: 307–312.
- Masterson LR, Mascioni A, Traaseth NJ, Taylor SS, and Veglia G (2008) Allosteric cooperativity in protein kinase A. *Proceedings of the National Academy of Sciences of the United States of America* 105: 506–511.
- Newhall KJ, Cummings DE, Nolan MA, and McKnight GS (2005) Deletion of the RII beta-subunit of protein kinase A decreases body weight and increases energy expenditure in the obese, leptin-deficient ob/ob mouse. *Molecular Endocrinology* 19: 982–991.
- Pilz RB and Broderick KE (2005) Role of cyclic GMP in gene regulation. *Frontiers in Bioscience* 10: 1239–1268.
- Schlossmann J and Desch M (2009) cGK substrates. *Handbook of Experimental Pharmacology* 191: 163–193.
- Schlossmann J, Feil R, and Hofmann F (2003) Signaling through NO and cGMP-dependent protein kinases. *Annals of Medicine* 35: 21–27.
- Scott JD and Pawson T (2009) Cell signaling in space and time: Where proteins come together and when they're apart. *Science* 326: 1220–1224.
- Skalhegg BS, Huang Y, Su T, et al. (2002) Mutation of the C alpha subunit of PKA leads to growth retardation and sperm dysfunction. *Molecular Endocrinology* 16: 630–639.
- Taylor SS, Kim C, Cheng CY, et al. (2008) Signaling through cAMP and cAMP-dependent protein kinase: diverse strategies for drug design. *Biochimica et Biophysica Acta* 1784: 16–26.
- Walter U and Gambaryan S (2009) cGMP and cGMP-dependent protein kinase in platelets and blood cells. *Handbook of Experimental Pharmacology* 191: 533–548.

Cyclic Nucleotide Phosphodiesterases

V C Manganiello, National Institutes of Health, Bethesda, MD, USA

E Degerman, Lund University, Lund, Sweden

© 2013 Elsevier Inc. All rights reserved.

Glossary

Cyclic nucleotide phosphodiesterases (PDEs) Enzymes that catalyze hydrolysis of the 3'-5'-phosphodiester bond of cyclic adenosine monophosphate (cAMP) and/or cyclic guanosine monophosphate (cGMP).

Desensitization Prolonged or repeated exposure of receptors to their ligands leads to reduction or loss of responsiveness,

either homologous desensitization with loss of responsiveness only to a specific ligand or heterologous desensitization with loss of responsiveness to multiple ligands.

Family-specific inhibitors Drugs that inhibit the activity of one phosphodiesterase (PDE) gene family with 10–100-fold greater potency than the activities of other PDE gene families.

Cyclic nucleotide phosphodiesterases (PDEs) constitute a large, diverse, and complex superfamily of metal hydrolases which cleave the 3',5'-cyclic phosphate bond of cyclic adenosine monophosphate (cAMP) and cyclic guanosine monophosphate (cGMP), resulting in production of 5'-AMP and 5'-GMP, respectively. Based on differences in primary structures, PDEs have been divided into three major classes, I, II, and III. Class I PDEs, which contain a conserved catalytic domain (~250–300 amino acids), comprise the 11 structurally related, highly regulated, and functionally distinct mammalian gene families (PDEs 1–11). The class I catalytic domain contains a signature motif H(X)₃H (X)_{25–35}(D/E), which is also found in metal hydrolases. Class I PDEs are found in the parasites *Trypanosoma brucei* and *cruzi*, and in *Drosophila*, nematodes, yeast, and sponges. Class II PDEs, which are structurally unrelated to class I PDEs, are found in fungi, slime mold, and amoebae. Class III PDEs, with a β -lactamase-type catalytic domain, are found in *Dictyostelium discoideum*. The evolutionary relationship between the three PDE classes is not certain; this brief article focuses on mammalian class I PDEs.

By catalyzing hydrolysis of cyclic nucleotides, mammalian PDEs regulate their intracellular concentrations, and, consequently, their signaling pathways and myriad physiological effects, including myocardial contractility, visual transduction, vascular and airway smooth muscle relaxation, immune/inflammatory responses, cell proliferation and apoptosis, memory, and many others. There is a growing interest in identifying mutations in PDE genes that are associated with inherited diseases. The mammalian PDE superfamily is a major target for drug discovery in the treatment of clinically important diseases, with several PDE inhibitors being approved for use by the Food and Drug Administration (FDA), for example, PDE5 inhibitors for erectile dysfunction and pulmonary hypertension, and certain PDE3 inhibitors for acute heart failure and intermittent claudication.

Molecular Diversity

cAMP and cGMP are important intracellular second messengers that modulate many biological processes. The classical mechanism for transduction of cyclic nucleotide signals involves cyclic-nucleotide-induced activation of cAMP- and

cGMP-dependent protein kinases (PKA and PKG, respectively), with subsequent phosphorylation of critical downstream regulatory effectors. A second mechanism, however, assumed ever-increasing importance. It is clear that cyclic nucleotide-binding proteins can serve as direct mediators of many cyclic nucleotide actions, for example, cyclic-nucleotide-gated ion channels, cAMP-activated guanine nucleotide exchange factors (exchange protein activated by cAMP (EPACs)) which regulate Rap1 guanosine triphosphatases (GTPases), and several PDEs, especially PDEs 2, 5, 6, which contain allosteric, noncatalytic cyclic-nucleotide-binding sites.

Molecular genetics has revealed the diversity and complexity of the 11 mammalian PDE gene families (PDEs 1–11). PDE families differ in their primary amino sequences, substrate specificities, sensitivities to endogenous effectors and pharmacological agents, cellular functions, and mechanisms whereby they are regulated. Most PDE families comprise more than one gene; within these multiple families, closely related isoforms are generated from the same gene or different genes via alternative messenger RNA (mRNA) splicing or utilization of different promoters/transcription initiation sites. More than 20 PDE genes probably encode more than 50 PDE proteins. Some cells are relatively enriched in specific PDEs, for example, photoreceptor PDE6 is virtually exclusively expressed in the retina. Most cells, however, contain representatives of multiple PDE gene families, and different members of the same family, but in different amounts, proportions, and subcellular locations. Redundancy in PDEs, that is, the presence in the same cell of multiple enzymes which essentially perform the same reaction of hydrolyzing cyclic nucleotides, does not merely serve a survival or protective function, but rather allows cells to use distinct subsets of differentially regulated and localized PDEs to specifically and selectively regulate and segregate the generation, amplitude, duration, and compartmentation of cyclic nucleotide signals and actions. It is generally accepted that the cellular capacity to degrade cyclic nucleotides far exceeds their synthesis, and is a major factor in the very rapid turnover of intracellular cAMP and cGMP. The physiological significance and functional consequences of this rapid turnover, and concomitant heat generation (cyclic nucleotide hydrolysis is accompanied by release of $\sim 7 \text{ cal mol}^{-1}$), are not completely understood. This brief article discusses some

general characteristics of PDEs and then focuses on the cellular biology and diverse functions of different PDE isoforms and their potential as therapeutic targets.

Structure/Function Analyses

Catalytic Domain

Mammalian PDEs exhibit a common structural organization, with a conserved catalytic domain (~250–300 amino acids) in the C-terminal portion of the molecules and divergent regulatory domains and modules in N-terminal portions (Figure 1). Many PDEs exist as asymmetric homo- or heterodimers. The catalytic core, more highly conserved among members of the same gene family than different gene families, contains a signature motif [HD (X₂) H (X₄) N], common to all PDEs. Some PDE families are relatively specific for hydrolysis of cAMP (PDEs 4, 7, and 8); others for cGMP (PDEs 5, 6, and 9); and some exhibit mixed specificity for both cAMP and cGMP (PDEs 1, 2, 3, 10, and 11). Although methylxanthines inhibit almost all PDEs, relatively specific and selective inhibitors, that is, drugs that target individual PDE families with 10–100-fold greater potency than other PDE families, are available for several families, for example, PDEs 1, 2, 3, 4, 5, and 6. These family-specific inhibitors are important for both basic research and clinical applications.

Analyses of the crystal structures of the catalytic domains of seven PDE families (PDEs 1–5, 7, and 9) indicate that these core catalytic domains assume a compact structure consisting of 14 α -helices consisting of three subdomains. At the interface of these domains, the active site, which contains two metal-binding sites, forms a deep hydrophobic pocket lined with highly conserved residues. All families have a uniform conformation, except for PDE5. Cyclic nucleotide substrate specificity may be determined by the orientation (stabilized by a hydrogen-binding network) of an amide group of an invariant glutamine at the active site. Available data,

however, do not completely support this mechanism for all dual-specificity PDEs which hydrolyze both cAMP and cGMP. Co-crystallization studies of PDE-inhibitor complexes suggest that inhibitors may interact with the invariant glutamine at the active site. Inhibitors bind in a common subpocket which contains structural elements responsible for nonselective binding of inhibitors and for the hydrolysis of cAMP and cGMP, as well as family-specific sequences responsible for differences in substrate affinities and sensitivities to family-specific inhibitors, such as IC224 and IC295 (PDE1); EHNA (erythro-9-(2 hydroxy-3-nonyl)-adenine) and BAY60-7750 (PDE2); milrinone, cilostazol, and cilostamide (PDE3); rolipram and roflumilast (PDE4); sildenafil, vardenafil, and tadalafil (PDE5).

Regulatory Domain

N-terminal portions of PDE molecules are highly divergent, containing structural determinants and specific amino acid sequences that allow different PDEs to respond selectively to specific regulatory signals, and to interact and communicate with the surrounding cellular milieu (Figure 1). These regulatory regions include dimerization domains and autoinhibitory modules (e.g., in PDEs 1, 4, and 5), as well as sites and domains that are subject to different types of covalent modification (e.g., sites in PDEs 1, 3, 4, 5, 10, and 11 for phosphorylation by various protein kinases), or sites that interact with allosteric ligands (e.g., cGMP-binding sites in GAF domains), specific effectors (e.g., Ca²⁺/calmodulin), protein partners or molecular scaffolds (e.g., A kinase anchoring protein scaffolding proteins (AKAPs), β -arrestin, and caveolin), and thereby regulate catalytic activity, protein–protein interactions, and/or subcellular compartmentation and localization (Figure 1). Five PDE families (PDEs 2, 5, 6, 10, and 11) contain homologous so-called GAF domains, an acronym for proteins (cGMP-binding PDEs, Anabena adenyl cyclase and *hflA*, an *Escherichia coli* transcriptional regulator) that contain these sequences. Although cGMP binding is not the primary function of GAF domains, in three PDE families

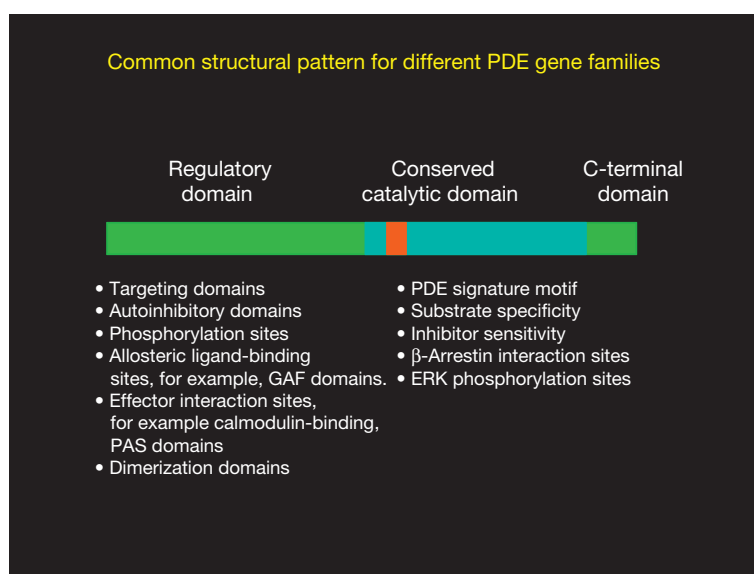


Figure 1 Common structural pattern for different phosphodiesterase (PDE) gene families.

(PDEs 2, 5, and 6), GAF domains bind cGMP with high affinity, but with different functional consequences. PDE8 isoforms contain PAS domains, an acronym for proteins (period clock protein, aryl hydrocarbon receptor nuclear translocator, and single-minded protein) that contain these sequences. In PDE4 isoforms, ligand binding to upstream conserved regions (UCRs) may induce profound conformational changes that allow interactions between the ligand-bound UCR and the catalytic site, resulting in allosteric modulation of catalytic activity.

PDE Functions

PDEs: Components of Spatially Organized Signaling Networks and Microdomains

PDEs are critical regulators of the generation, amplitude, duration, and termination of intracellular cyclic nucleotide signals. In addition to tight regulation of their concentrations and turnover, intracellular pools of cAMP and cGMP are also temporally, spatially, and functionally compartmentalized. In cardiac myocytes, for example, PGE₁ and catecholamines increase cAMP and activate PKA in different compartments, and cAMP diffusion is spatially restricted and regulated, at least in part, by PDEs. Newer techniques, for example, use of fluorescence-resonance energy transfer (FRET) and cyclic nucleotide biosensors, also suggest that PDEs play an important role in these spatially constrained microdomains, that is, discrete subcellular regions in which cyclic nucleotide gradients are generated, monitored, and regulated, and their signals and effects channeled, transduced, and modulated.

By virtue of their unique N-terminal regions, the multiple PDE4 N-terminal splice variants are ideally designed for intracellular targeting to different subcellular compartments, molecular scaffolds (including AKAPs, β -arrestin, RACK1 (receptor for activated protein kinase C 1), myomegalin, spectrin, disrupted in schizophrenia (DISC1)), and other interacting protein partners. In these complexes and signaling microdomains, PDE4 isoforms determine local cAMP gradients, and, together with sequestered PKA and EPAC partners, gate activation of spatially segregated, localized signaling pathways. The 'classical' molecular basis for compartmentalization of cAMP signaling involves anchoring of PKA isoforms at specific intracellular sites via A Kinase Anchoring Protein scaffolding proteins (AKAPs). These proteins organize formation of localized signaling modules/microdomains consisting of kinases, kinase substrates, phosphatases, PDE (especially PDE4) isoforms, and other proteins, including EPACs. AKAPs anchor PKA in close proximity to substrate targets, leading to phosphorylation of compartmentalized PKA substrates. These modules sense intracellular cAMP gradients and effectively compartmentalize activation of substrates and biological responses. In these microdomains, cAMP-induced activation of PKA also results in phosphorylation/activation of PDE4 isoforms associated with AKAPs, resulting in modulation and termination of cAMP signals, and return of PKA to its basal state. For example, in cardiac myocytes, PDE4 isoforms are associated with AKAP complexes thought to modulate effects of cAMP on L-type and ryanodine-sensitive Ca²⁺ channels (RyR2). By limiting β -adrenoreceptor (BAR) stimulation/phosphorylation of RyR2, PDE4 may protect against proposed pro-arrhythmogenic effects of hyper-phosphorylated RyR2.

This signaling microdomain or module may have clinical relevance to human heart failure, in that it has been suggested that, in some patients, decreased PDE4 in this AKAP-centered macromolecular complex may contribute to disease progression via chronic dysregulation of PKA activation of Ca²⁺ transport and development of arrhythmias. This information has raised theoretical concerns regarding chronic administration of PDE4 inhibitors to older patients with heart failure.

PDE4 isoforms, PKA, and other molecules also interact/associate with β -arrestins, which are scaffolding proteins located in the vicinity of plasma membranes (PMs) that interact with activated β receptors and are involved in coordination of β receptor signaling and trafficking. This targeting and effective concentrating of PKA and PDE4 in proximity to the activated β receptors not only reduces cAMP generation (via phosphorylation, uncoupling, and internalization of β receptors) but also enhances localized degradation of cAMP and desensitizes the complex with respect to further signaling via the β receptor and cAMP.

Experiments in isolated cardiomyocytes strongly suggest that specific PDEs are functionally coupled with discrete pools of adenylyl cyclase/G-protein-coupled receptors that respond to specific hormones. For example, β_1 AR-induced cAMP is regulated by PDE3 and PDE4, whereas glucagon R-induced cAMP is regulated by PDE4 alone. Although PDE2 accounts for a minor fraction of PDE activity in neonatal rat cardiomyocytes, it seems to be specifically involved in the hydrolysis of cAMP induced by β AR stimulation. Since PDE2 is stimulated by cGMP, it mediates inhibitory effects of cGMP on β AR-induced increases in cAMP and contractility.

The subcellular localization of different PDEs, their interactions with molecular scaffolds and interacting partners, and their inclusion in, and contribution to the function of, macromolecular signaling microdomains is becoming a common theme in PDE biology. In human platelets, PDE5, PKG1 β , and inositol-1,4, 5-triphosphate receptor type 1 (IP(3)R1) are integral components of a complex that regulate effects of cGMP on IP(3)R1-mediated Ca²⁺ release. In 3T3-L1 adipocytes, PDE3B is found in caveolin-enriched PM fractions, and endoplasmic reticulum (ER) and Golgi membrane fractions. Treatment of adipocytes with insulin resulted in phosphorylation/activation of PDE3B associated with internal membranes (ER/Golgi), whereas treatment with the β_3 -receptor agonist, CL316,243 (CL), resulted in phosphorylation/activation of PDE3B associated with PM/caveolae fractions. Activation of PDE3B by insulin and CL involves the reversible assembly of distinct macromolecular complexes that contain phosphorylated PDE3B and different sets of signaling effectors thought to be involved in activation of PDE3B by insulin or CL. Insulin- or CL-mediated macromolecular complex formation was significantly inhibited by caveolin-1 knockdown, suggesting that cav-1 acts as a molecular chaperone or scaffolding molecule in cholesterol-rich lipid rafts that may be necessary for the proper stabilization and activation of PDE3B in response to CL and insulin.

PDEs: Homeostatic Regulators

Signaling pathways, in general, also include mechanisms for negative feedback control. PDE3 and PDE4 activities, for

example, are acutely upregulated by cAMP-induced activation of PKA, which results in phosphorylation/activation of PDE3 and PDE4, and enhanced destruction of cAMP. This rapid destruction/turnover is crucial in the transient accumulation of intracellular cAMP that usually occurs in response to acute stimuli, and important in resetting cAMP homeostasis. Feedback regulation of cAMP destruction in cAMP-signaling microdomains described above maintains compartmentalized cAMP, and controls diffusion of cAMP signals. Chronic elevation of cAMP also provides negative feedback, by increasing transcription of PDE3 and PDE4 genes and protein synthesis, resulting in increased enzymatic activities. In some cells, this latter phenomenon is part of mechanisms involved in ligand-induced downregulation of cellular responses, such as tachyphylaxis or desensitization. Negative feedback control of cGMP hydrolysis by PDE5 is somewhat different, due to the presence of high-affinity, noncatalytic, allosteric-binding sites for cGMP in the GAF domains of PDE5. Binding of cGMP to GAF domains results in allosterically induced conformational changes that increase affinity of the catalytic site for cGMP and also facilitates phosphorylation and activation of PDE5 by PKG. Thus, elevation in intracellular cGMP provides negative feedback control and enhances its own destruction, via both direct cGMP-induced allosteric activation as well as indirect activation due to phosphorylation of PDE5 by PKG.

PDEs: Intracellular Effectors

PDEs not only are important determinants in the stringent regulation of intracellular cAMP and cGMP concentrations, but also serve as effectors of cyclic nucleotide actions. Historically, in cells containing multiple PDEs, family-specific PDE inhibitors were initially utilized to pharmacologically define roles of specific PDEs in regulating specific signaling pathways and discrete cellular functions. In cultured mammalian oocytes, PDE3 inhibitors, not PDE4 or 5 inhibitors, inhibit meiotic progression and oocyte maturation, implying that PDE3 regulates a cAMP pool that controls maturation. (In fact, female PDE3A knockout (KO) mice are sterile.) In cultured renal mesangial cells, experiments with PDE3 and PDE4 inhibitors indicated that PDE3 and PDE4 selectively regulated functionally distinct cAMP pools that controlled cell growth and generation of reactive oxygen species, respectively. These types of studies first suggested a role for PDEs in spatial and/or functional compartmentalization of cyclic nucleotide signals and effects.

In some instances, a specific PDE serves as a crucial effector system and regulates a unique cellular function, for example, in retinal rods and cones, light-induced activation of photoreceptor PDE6 results in hydrolysis of cGMP and initiation of visual signal transduction. Activation of PDE3s by insulin, IGF-1, and leptin is apparently important in their ability to reduce cAMP and thereby regulate lipolysis in adipocytes, glycogenolysis in hepatocytes, insulin secretion from pancreatic β cells, and oocyte maturation. In these instances, PDE3 inhibitors can attenuate or block the effects of insulin, IGF-1, and leptin. In monocytes/macrophages, upregulation of a specific PDE4B splice variant is required to relieve cAMP inhibition of tumor necrosis factor (TNF) production and toll-like receptor signaling.

Studies of PDE KO mice do indicate that the ability of specific PDEs to regulate discrete cAMP/cGMP signaling pathways and actions is genetically determined. Female PDE3A KO mice are sterile, most likely due to inhibitory effects of cAMP on meiotic progression and maturation of oocytes and, consequently, their competency for fertilization. PDE3B KO mice, not frankly diabetic, demonstrate signs of disruption in insulin secretion and insulin resistance. Compared with wild-type (WT) mice, PDE3B KO mice and PDE4B KO mice were leaner, with lower fat pad weights and smaller adipocytes, and with reduced inflammation and macrophage markers in white adipose tissue (WAT). PDE3B KO WAT exhibited phenotypic characteristics of brown adipose tissue. These data suggest a role for both PDE3 and PDE4 in regulation of energy homeostasis, and are of particular interest, because some of the beneficial metabolic effects of the polyphenol resveratrol, at least in rodents, may be mediated by inhibition of PDE3 and PDE4. Production of TNF α in response to administration of lipopolysaccharide (LPS) in PDE4B KO mice is profoundly (~90%) inhibited. On the other hand, PDE4D KO mice do not exhibit acetylcholine-induced contraction and hyper-reactivity of airway smooth muscle. In PDE4D KO mice, the duration of β -adrenergic-induced anesthesia is reduced by more than 50%, and fertility is decreased due to reduced ovulation. Targeted inactivation of PDE1B alters locomotion activity, whereas sperm function is impaired in PDE11 KO mice.

PDEs: Signal Integrators

With their different intrinsic properties and different responses to regulatory signals, PDEs also integrate multiple inputs and are a locus for cross-talk between different signaling pathways. For example, PDE2, which is allosterically activated by cGMP, is highly concentrated in adrenal zona glomerulosa cells. In these cells, atrial natriuretic factor (ANF) increases cGMP synthesis; cGMP, in turn, activates PDE2, leading to a decrease in cAMP and PKA activity and inhibition of cAMP-stimulated aldosterone production. On the other hand, in other cells, NO activates guanylyl cyclase and increases cGMP, which inhibits PDE3 and increases intracellular cAMP content, resulting, for example, in stimulation of renin secretion from the juxtaglomerular apparatus in the kidney and inhibition of platelet aggregation.

Human platelets offer an excellent example of signal integration by PDE2, PDE3, and PDE5. At low cAMP, PDE3 is important and is inhibited by endogenous cGMP, produced by nitrovasodilator-activation of guanylyl cyclase or by inhibition of the cGMP-specific PDE5. This inhibition of PDE3 results in increased cAMP and inhibition of platelet aggregation. At higher cGMP (or cAMP) concentrations, PDE2 is allosterically activated to hydrolyze cAMP. Thus, activation of PDE2 limits accumulation of cAMP in platelets and aids in maintenance of cAMP homeostasis.

Newer approaches have suggested that different PDE isoforms may not only coordinate regulation of signals and responses in individual cells, but that differential cellular distribution within tissues may be important in intercellular communication and regulation of tissue function. Immunohistochemical studies indicate, for example, that PDE1C and PDE2A are expressed in different sets of neurons in the olfactory epithelium, whereas PDE1C

and PDE4A are present in different subcellular regions of the same neurons. In certain cells, for example, pancreatic beta-cells and olfactory epithelium, Ca^{2+} /calmodulin regulation of PDE1, together with adenylyl cyclases 1 and 8, allow coupling of cAMP and Ca^{2+} oscillations, crucial for rapid functional responses in these cells.

Inherited Diseases Associated with PDEs

At present, mutations in the PDE6 and PDE11 genes represent the clearest examples of association of PDE mutations with human disease. Recessive mutations in rod PDE6B have been found in 5–6% of patients with retinal degeneration; mutations in PDE6A are associated with autosomal recessive retinitis pigmentosa. Mutations in PDE11A gene have been found in a subgroup of patients presenting with bilateral adrenal hyperplasia and Cushing's syndrome. Although controversial and not replicated in all studies, there have been reports correlating PDE4D haplotypes and single-nucleotide polymorphisms and stroke, and perhaps, PDE4B and schizophrenia.

Family-Specific Inhibitors: Clinical Applications

Molecular diversity of PDEs has occasioned the development of family-specific PDE inhibitors to replace nonselective PDE inhibitors (theophylline and caffeine) as therapeutic agents for a wide variety of major diseases. PDE5 inhibitors, especially Viagra®, Levitra®, and Cialis®, have been very successful in the treatment of erectile dysfunction, and are approved by the FDA for pulmonary hypertension. The PDE3 inhibitor Pletal® is used to alleviate symptoms in a type of peripheral vascular disease, that is, intermittent claudication. PDE3 inhibitors, which enhance myocardial contractility and relaxation of smooth muscle and inhibit platelet aggregation, failed in long-term clinical trials for the treatment of chronic heart failure, but one PDE3 inhibitor, milrinone, is approved for acute and short-term treatment of adult patients with decompensated and refractory cardiac failure. PDE4 enzymes are relatively highly concentrated in immune/inflammatory cells, and specific PDE4 inhibitors exhibit potent anti-inflammatory actions. For these reasons, selective PDE4 inhibitors, targeted toward the active catalytic site, have been extensively studied as anti-inflammatory drugs for treatment of asthma and chronic obstructive pulmonary disease (COPD), and also (on a smaller scale) as possible therapeutic agents for rheumatoid arthritis, multiple sclerosis, type II diabetes, septic shock, atopic dermatitis and other autoimmune diseases, and depression. Virtually, all clinical trials with PDE4 inhibitors, however, have not been successful, because tolerability problems, especially emesis and diarrhea, have precluded achievement and maintenance of clinically effective concentrations. A novel strategy has recently been initiated in the development of PDE4D inhibitors that may allow these drugs to achieve their rich and highly anticipated therapeutic promise. This new class, termed 'allosteric modulators', which limit access of substrate to the PDE4D active site, are apparently potent in pre-clinical assays, but have much less potential for emesis than existing PDE4 inhibitors.

Conclusions

By virtue of their distinct intrinsic characteristics and differential regulation, their intracellular targeting to different subcellular locations and microdomains, and their interactions with cellular structural elements, regulatory partners, and molecular scaffolds such as AKAPs, β arrestin, and caveolin, different PDEs can integrate multiple cellular inputs and modulate the intracellular diffusion and functional compartmentalization of cyclic nucleotide signals. The combined enormous molecular diversity of receptors and their ligands, adenylyl and guanylyl cyclase systems, PDEs, and cyclic-nucleotide-regulated effector systems, coupled with their physical and functional compartmentalization, provides for the complex integration, specificity, and variety of networks and pathways involved in generation, transduction, modulation, and termination of cyclic-nucleotide-gated signals and actions. The ability of scaffolding proteins to spatially organize signaling molecules, including protein kinases, phosphatases, kinase substrates, PDEs, and other signaling and effector molecules, effectively allows cells to generate signaling specificity by using small, discrete, and spatially segregated subsets of proteins which exhibit overlapping or redundant functions. Integration of these myriad combinations establishes the unique cyclic nucleotide networks and phenotypes that characterize individual cells.

Acknowledgments

VM is supported by the Intramural Research Program, NHLBI, NIH; ED, in part, by Swedish Medical Research Council grant 3362.

See also: **Signaling:** Adenylyl Cyclases; Cyclic GMP Phosphodiesterases; G-Protein-Coupled Receptor Kinases and Arrestins.

Further Reading

- Ahmad F, Lindh R, Tang Y, et al. (2009) Differential regulation of adipocyte PDE3B in distinct membrane compartments by insulin and the beta 3-adrenergic receptor agonist CL316243: Effects of caveolin-1 knockdown on formation/maintenance of macromolecular signaling complexes. *Biochemical Journal* 424: 399–410.
- Borland G, Smith BO, and Yarwood SJ (2009) EPAC proteins transduce diverse cellular actions of cAMP. *British Journal of Pharmacology* 158: 70–86.
- Burgin AB, Magnusson OT, Singh J, et al. (2009) Design of phosphodiesterase 4D (PDE4D) allosteric modulators for enhancing cognition with improved safety. *Nature Biotechnology* 28: 63–72.
- Conti M (2004) A view into the catalytic pocket of cyclic nucleotide phosphodiesterases. *Nature Structural and Molecular Biology* 11: 809–810.
- Conti M and Beavo J (2007) Biochemistry and physiology of cyclic nucleotide phosphodiesterases: Essential components in cyclic nucleotide signaling. *Annual Review of Biochemistry* 76: 481–511.
- Degerman E and Manganiello VC (2007) PDE3B: An important regulator of energy homeostasis. In: Beavo JA, Francis SH, and Houslay MD (eds.) *Cyclic Nucleotide Phosphodiesterases in Health and Disease*, pp. 79–98. Boca Raton, FL: CRC Press.
- Dodge-Kafka KL, Langeberg L, and Scott JD (2006) Compartmentation of cyclic nucleotide signaling in the heart: The role of A-kinase anchoring proteins. *Circulation Research* 98: 993–1001.

- Houslay MD (2009) Underpinning compartmentalized cAMP signaling through targeted cAMP breakdown. *Trends in Biochemical Sciences* 35: 91–100.
- Houslay MD, Schafer P, and Zhang YJ (2005) Phosphodiesterase-4 as a therapeutic target. *Drug Discovery Today* 10: 1503–1519.
- Ke H and Wang H (2007) Crystal structures of phosphodiesterases and implications on substrate specificity and inhibitor selectivity. *Current Topics in Medicinal Chemistry* 7: 391–403.
- Lehnart SE and Marks AR (2006) Phosphodiesterase4D and heart failure: A cautionary tale. *Expert Opinion on Therapeutic Targets* 10: 677–688.
- Mongillo M and Zaccolo M (2006) A complex phosphodiesterase system controls β -adrenoreceptor signaling in cardiomyocytes. *Biochemical Society Transactions* 34: 509–511.
- Thompson PE, Manganiello V, and Degerman E (2007) Rediscovering PDE3 inhibitors – new opportunities for a long-neglected target. *Current Topics in Medicinal Chemistry* 7: 421–436.
- Wilkins MR, Wharton J, Grimminger F, and Ghofrani HA (2008) Phosphodiesterase inhibitors for the treatment of pulmonary hypertension. *European Respiratory Journal* 32: 198–209.

Cyclic Nucleotide-Regulated Cation Channels

M Biel, Ludwig-Maximilians-Universität, Munich, Germany

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Martin Biel and Franz Hofmann, volume 1, pp. 512–515, © 2004, Elsevier Inc.

Glossary

Channel states Voltage-gated channels have at least three states: a closed, an open, and an inactivated state. Ions pass the channel at the open state.

Ligand-gated channel In contrast to voltage-gated channels, ligand-gated channels are opened by the occupation of the ligand-binding site. Hyperpolarization-activated cyclic nucleotide-gated (HCN) channels are gated by the membrane potential and are modulated

by the binding of cyclic adenosine monophosphate (cAMP).

Olfaction The processing of smell in the nose and olfactory bulb.

Pacemaker cells Specialized cells that depolarize at a known frequency. The best-known example is the sinoatrial node cells of the heart.

Photoreceptors Light-sensitive neurons (rods and cones) of the retina that confer vision.

Introduction

Cyclic nucleotides exert their physiological effects by binding to four major classes of cellular receptors: cyclic adenosine monophosphate (cAMP)- and cyclic guanosine monophosphate (cGMP)-dependent protein kinases, cyclic GMP-regulated phosphodiesterases, cAMP-binding guanine nucleotide exchange factors, and cyclic nucleotide-regulated cation channels. Cyclic nucleotide-regulated cation channels are unique among these receptors because their activation is directly coupled to the influx of extracellular cations into the cytoplasm and to the depolarization of the plasma membrane. Two families of channels regulated by cyclic nucleotides have been identified, the cyclic nucleotide-gated (CNG) channels and the hyperpolarization-activated cyclic nucleotide-gated (HCN) channels. The two channel classes differ from each other with regard to their mode of activation. CNG channels are opened by direct binding of cAMP or cGMP. By contrast, HCN channels are principally operated by voltage. These channels open at hyperpolarized membrane potentials and close on depolarization. Apart from their voltage sensitivity, HCN channels are also activated directly by cyclic nucleotides, which act by increasing the channel open probability.

General Features of Cyclic Nucleotide-Regulated Cation Channels

Structurally, both CNG and HCN channels are members of the superfamily of pore loop cation channels. Like other subunits encoded by this large gene family, CNG and HCN channel subunits assemble to tetrameric complexes. The proposed structure and the phylogenetic relationship of mammalian CNG and HCN channel subunits are shown in [Figure 1](#). The transmembrane channel core consists of six α -helical segments (S1–S6) and an ion-conducting pore loop between the S5 and S6. The amino- and carboxy-termini are localized in the cytosol. CNG and HCN channels contain a positively charged S4 helix carrying three to nine regularly spaced

arginine or lysine residues at every third position. In HCN channels, as in other voltage-operated members of the channel superfamily, the S4 helix functions as a voltage-sensor conferring voltage-dependent gating. In CNG channels, which are not gated by voltage, the specific role of S4 is not known.

CNG and HCN channels reveal different ion selectivities. CNG channels conduct both Ca^{2+} and monovalent cations with permeability ratios $P_{\text{Ca}}/P_{\text{Na}}$ ranging from about 2 to 25 depending on the respective channel type and the cyclic nucleotide concentration. By providing an entry pathway for Ca^{2+} , CNG channels control a variety of cellular processes that are triggered by this cation. Under physiological conditions, HCN channels pass Na^+ and K^+ ion with a relative permeability ratio $P_{\text{Na}}/P_{\text{K}}$ of about 0.15–0.25. In addition, there is recent evidence for a small but significant Ca^{2+} permeability of these channels. The functional relevance of Ca^{2+} entry through HCN channels is not clear at the moment.

In the carboxy-terminus, CNG and HCN channels contain a cyclic nucleotide-binding domain (CNBD) that has significant sequence similarity to the CNBDs of most other types of cyclic nucleotide receptors. In CNG channels, the binding of cyclic nucleotides to the CNBD initiates a sequence of allosteric transitions that lead to the opening of the ion-conducting pore. In HCN channels, the binding of cyclic nucleotides is not required for activation. However, cyclic nucleotides shift the voltage dependence of channel activation to more positive membrane potentials and thereby facilitate voltage-dependent channel activation. Despite the fact that the CNBDs of HCN and CNG channels show significant sequence homology, the two channel classes reveal different selectivities for cyclic nucleotides. HCN channels display an approximately 10-fold higher affinity for cAMP than for cGMP, whereas CNG channels select cGMP over cAMP.

CNG Channels

CNG channels are expressed in retinal photoreceptors and olfactory neurons and they play a key role in visual and

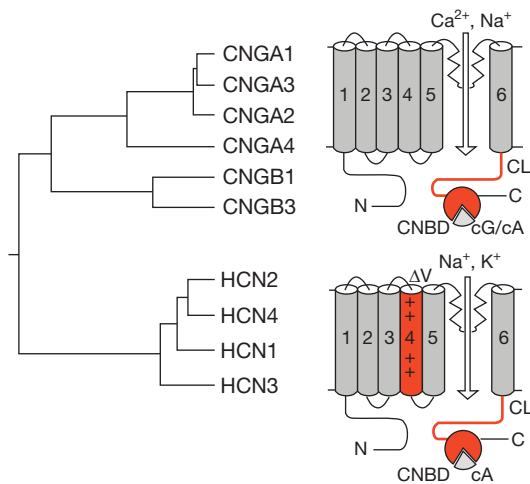


Figure 1 Phylogenetic tree and structural model of cyclic nucleotide-regulated cation channels. The cyclic nucleotide-gated (CNG) channel family comprises six members, which are classified into A subunits (CNGA1–4) and B subunits (CNGB1 and CNGB3). The hyperpolarization-activated cyclic nucleotide-gated (HCN) channel family comprises four members (HCN1–4). CNG and HCN channels share a common transmembrane topology, consisting of six transmembrane segments (1–6), a pore loop, and a cyclic nucleotide-binding domain (CNBD). CNG channels are activated *in vivo* by binding of either cAMP (cA) or cGMP (cG), depending on the channel type. HCN channels activate on membrane hyperpolarization (ΔV) and are enhanced by binding of cAMP. Structures involved in channel gating are shown in red. The positively charged amino acid residues in the S4 segment of HCN channels are indicated by (+). CL denotes the C-linker involved in activation gating of CNG and HCN channels.

olfactory signal transduction. CNG channels are also found at low density in some other cell types and tissues such as brain, testis, and kidney. Although the function of CNG channels in sensory neurons has been unequivocally demonstrated, the role of these channels in other cell types remains to be established. Based on phylogenetic relationship, the six CNG channels identified in mammals are divided into two subfamilies, the A subunits (CNGA1–4) and the B subunits (CNGB1 and CNGB3) (Figure 1). CNG channel A subunits (with the exception of CNGA4) can form functional homomeric channels in various heterologous expression systems. By contrast, B subunits do not give rise to functional channels when expressed alone. However, together with CNGA1–3, they confer novel properties (e.g., single-channel flickering, increased sensitivity for cAMP, and *L-cis*-diltiazem) that are characteristic of native CNG channels. The physiological role and subunit composition is known for three native CNG channels: the rod photoreceptor channel, the cone photoreceptor channel, and the olfactory channel. The CNG channel of rod photoreceptors is a tetramer composed of three CNGA1 and one long isoform of the CNGB1 subunit (CNGB1a). The cone photoreceptor channel consists of two copies of the CNGA3 and the CNGB3 subunit, respectively. CNG channels control the membrane potential and the calcium concentration of photoreceptors. In the dark, both channels are maintained in the open state by a high concentration of cGMP. The resulting influx of Na^+ and Ca^{2+} (dark current) depolarizes the photoreceptor and

promotes synaptic transmission. Light-induced hydrolysis of cGMP leads to the closure of the CNG channel. As a result, the photoreceptor hyperpolarizes and shuts off synaptic glutamate release. Mutations in CNG channel genes have been linked to retinal diseases. Mutations in the CNGA and CNGB1 subunits have been identified in the genome of patients suffering from retinitis pigmentosa. The functional loss of either the CNGA3 or the CNGB3 subunit causes total color blindness (achromatopsia) and degeneration of cone photoreceptors.

The CNG channel expressed in cilia of olfactory neurons consists of three different subunits: CNGA2, CNGA4, and a short isoform of the CNGB1 subunit (CNGB1b). The channel is activated *in vivo* by cAMP that is synthesized in response to the binding of odorants to their cognate receptors. The olfactory CNG channel is thought to conduct almost exclusively Ca^{2+} under physiological ionic conditions. The resulting increase in cellular Ca^{2+} activates a Ca^{2+} -activated Cl^- channel that further depolarizes the cell membrane. Ca^{2+} not only is a permeating ion of the olfactory CNG channel, but also represents an important modulator of this channel. CNG channels are desensitized by Ca^{2+} -calmodulin (CaM)-mediated feedback inhibition, which lowers the cAMP sensitivity. Although all three olfactory CNG channel subunits contain CaM-binding sites, only the binding site present in the CNGB1b subunit renders the channel sensitive to CaM. Unlike originally proposed, CaM-mediated CNG channel desensitization is not necessary for adaptation to repeated odorant exposure but rather acts as a termination mechanism for shaping responses of olfactory neurons during and after odorant stimulation.

HCN Channels

A cation current that is slowly activated by membrane hyperpolarization (termed I_h , I_f , or I_q) is found in a variety of excitable cells including neurons, cardiac pacemaker cells, and photoreceptors. The best understood function of I_h is the control of heart rate and rhythm by acting as pacemaker current in the sinoatrial (SA) node. I_h is activated during the membrane hyperpolarization following the termination of an action potential and provides an inward Na^+ current that slowly depolarizes the plasma membrane. Sympathetic stimulation of SA node cells raises cAMP levels and increases I_h by a positive shift of the current activation curve, thus accelerating diastolic depolarization and heart rate. Stimulation of muscarinic receptors slows down heart rate by the opposite action. In the brain, I_h fulfills diverse basic functions including control of neuronal rhythmicity, determination of resting membrane potential, dendritic integration, and synaptic plasticity.

HCN channels represent the molecular correlate of I_h . In mammals, the HCN channel family comprises four members (HCN1–4) that share about 60% sequence identity to each other and about 25% sequence identity to CNG channels. The highest degree of sequence homology between HCN and CNG channels is found in the CNBD. When expressed in heterologous systems, HCN channels generate currents displaying the typical features of native I_h : (1) activation by membrane hyperpolarization, (2) permeation of Na^+ and K^+ , (3) positive shift of the voltage dependence of channel activation by direct binding of cAMP, and (4) channel blockade by

extracellular Cs^+ . HCN1–4 mainly differ from each other with regard to their speed of activation and the extent by which they are modulated by cAMP. HCN1 is the fastest channel, followed by HCN2, HCN3, and HCN4. Unlike HCN2 and HCN4, whose activation curves are shifted by about +15 mV by cAMP, HCN1 and HCN3 are only weakly, if at all, affected by cAMP. Determination of the crystal structure of the CNBD combined with site-directed mutagenesis has provided insight into the complex mechanism underlying dual HCN channel activation by voltage and cAMP. Surprisingly, the voltage-dependent movement of the positively charged S4 helix is fully conserved between HCN channels and depolarization-activated channels such as Shaker K-channels. However, the allosteric coupling between S4 movement and the activation gate is different in the two channel types. In HCN channels, inward movement of S4 leads to the opening of the channel gate, whereas it closes Shaker K-channels. Major determinants affecting channel activation are the intracellular S4–S5 loop, the S5–P linker, the S1 segment, and the extracellular S1–S2 loop. The CNBD fulfills the role of an autoinhibitory channel domain. In the absence of cAMP, the cytoplasmic carboxy-terminus inhibits HCN channel gating by interacting with the channel core and thereby shifts the activation curve to more hyperpolarizing voltages. Binding of cAMP to the CNBD relieves this inhibition. Differences in the magnitude of the response to cAMP among the four HCN channel isoforms are largely due to differences in the extent to which the CNBD inhibits basal gating. It remains to be determined if the inhibitory effect of the CNBD is conferred by a direct physical interaction with the channel core domain or by some indirect pathway. There is evidence that the so-called C-linker, a peptide of about 80 amino acids that connects the last transmembrane helix (S6) to the CNBD, plays an important role in this process (Figure 1). The C-linker was also shown to play a key role in the gating of CNG channels, suggesting that the functional role of this domain has been conserved during channel evolution. The functional properties of HCN channels may also be influenced by auxiliary subunits such as MiRP1 or Trip8b.

HCN channels are found in neurons and heart cells. In brain, all four HCN channel isoforms have been detected with HCN2 being the most abundant isoform. By contrast, HCN1, HCN3, and HCN4 are enriched in specific regions of the brain such as thalamus (HCN4), hippocampus (HCN1), or olfactory bulb and hypothalamus (HCN3). HCN channels have also been detected in the retina and some peripheral neurons such as dorsal root ganglion neurons. In SA node cells, HCN4 represents the predominantly expressed HCN channel isoform. In addition, low amounts of HCN1 and HCN2 are also present in these cells. Insights into the (patho)physiological relevance of HCN channels have been gained from the analysis of mouse lines lacking individual HCN channel isoforms. Disruption of HCN1 impairs motor learning but enhances spatial learning and memory. Deletion of HCN2 results in absence epilepsy, ataxia, and sinus node dysfunction. Mice lacking HCN4 die *in utero* because of the failure to generate mature SA pacemaker cells. By contrast, mice in which HCN4 is

deleted in adult stage are viable but display cardiac arrhythmia characterized by recurrent sinus pauses. The key role of HCN4 in controlling heart rhythmicity is corroborated by genetic data from human patients. Mutations in the human HCN4 gene, leading to mutated or truncated channel proteins, have been found to be associated with sinus bradycardia and complex cardiac arrhythmia.

See also: **Bioenergetics:** Nuclear Calcium in Synapse-to-Nucleus Communication: Common Regulator of Persistent Adaptations in the Nervous System; Transient Receptor Potential Channels; Voltage-Dependent K^+ Channels; Voltage-Gated Ca^{2+} Channels; Voltage-Gated Sodium Channels: Structure, Function, and Pathophysiology; **Lipids Carbohydrates Membranes and Membrane Proteins:** Ion Channel Protein Superfamily; **Metabolism Vitamins and Hormones:** Color Vision; **Signaling:** Cyclic Nucleotide-Dependent Protein Kinases; Photoreceptors.

Further Reading

- Baruscotti M, Bucchi A, and DiFrancesco D (2005) Pacemaker channels. *Annals of the New York Academy of Sciences* 1015: 111–121.
- Biel M and Michalakakis S (2007) Function and dysfunction of CNG channels: Insights from channelopathies and mouse models. *Molecular Neurobiology* 35: 266–277.
- Biel M, Wahl-Schott C, Michalakakis S, and Zong X (2009) Hyperpolarization-activated cation channels: From genes to function. *Physiological Reviews* 89: 847–885.
- Biskup C, Kusch J, Schulz E, et al. (2007) Relating ligand binding to activation gating in CNGA2 channels. *Nature* 446: 440–443.
- Gauss R, Seifert R, and Kaupp UB (1998) Molecular identification of a hyperpolarization-activated channel in sea urchin sperm. *Nature* 393: 583–587.
- Kaupp UB (2010) Olfactory signalling in vertebrates and insects: Differences and commonalities. *Nature Reviews: Neuroscience* 11: 188–200.
- Kaupp UB and Seifert R (2002) Cyclic nucleotide-gated ion channels. *Physiological Reviews* 82: 769–824.
- Ludwig A, Budde T, Stieber J, et al. (2003) Absence epilepsy and sinus dysrhythmia in mice lacking the pacemaker channel HCN2. *EMBO Journal* 15: 216–224.
- Ludwig A, Zong X, Jeglitsch M, Hofmann F, and Biel M (1998) A family of hyperpolarization-activated mammalian cation channels. *Nature* 393: 587–591.
- Luo DG, Xue T, and Yau KW (2008) How vision begins: An odyssey. *Proceedings of the National Academy of Sciences of the United States of America* 105: 9855–9862.
- Männikkö R, Elinder F, and Larsson HP (2002) Voltage-sensing mechanism is conserved among ion channels gated by opposite voltages. *Nature* 419: 837–841.
- Nolan MF, Malleret G, Lee KH, et al. (2003) The hyperpolarization-activated HCN1 channel is important for motor learning and neuronal integration by cerebellar Purkinje cells. *Cell* 115: 551–564.
- Robinson RB and Siegelbaum SA (2003) Hyperpolarization-activated cation currents: From molecules to physiological function. *Annual Review of Physiology* 63: 453–480.
- Sanotoro B, Liu DT, Yao H, et al. (1998) Identification of a gene encoding a hyperpolarization-activated pacemaker channel in brain. *Cell* 93: 717–729.
- Song Y, Cygнар KD, Sagdullaev B, et al. (2008) Olfactory CNG channel desensitization by Ca^{2+} /CaM via the B1b subunit affects response termination but not sensitivity to recurring stimulation. *Neuron* 58: 374–386.
- Wainner BJ, DeGennaro M, Sanotoro B, Siegelbaum SA, and Tibbs GR (2001) Molecular mechanism of cAMP modulation of HCN pacemaker channels. *Nature* 411: 805–810.
- Zagotta WN, Olivier NB, Black KD, et al. (2003) Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature* 425: 200–205.
- Zhong H, Molday LL, Molday RS, and Yau KW (2002) The heteromeric cyclic nucleotide-gated channel adopts a 3A:1B stoichiometry. *Nature* 420: 193–198.

Cysteine Proteases

D J Buttle, University of Sheffield, Sheffield, UK

J S Mort, Shriners Hospital for Children, Montréal, QC, Canada

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 516–520, © 2004, Elsevier Inc.

Glossary

Endocytosis The taking in, by a cell, of soluble or membrane-bound material from outside the cell.

In the case of eukaryotic cells, endocytosed material passes into the endosome, a membrane-limited organelle that can fuse with lysosomes to allow digestion of the contents.

Homeostasis The normal balance of a healthy living organism, in which the status quo is maintained by regulatory mechanisms such as feedback control. In protein turnover, it is the balance between protein synthesis and protein breakdown.

Hydrolase An enzyme that breaks peptide, ester, or glycoside bonds by the addition of water. In the case of proteases, water is split, and the resulting oxygen is added to the newly formed carboxyl terminus while the two

hydrogens are added to the new amino terminus formed by the breaking of the peptide bond.

Protease An enzyme that cleaves proteins or peptides (also called proteinase, peptidase, or proteolytic enzyme).

Specificity Susceptibility of an amino acid sequence of a peptide or protein to cleavage by a particular protease. Most proteases possess specificity for the cleavage of particular sequences. For instance, caspases cleave immediately after aspartic acid residues (they cleave aspartyl bonds); most lysosomal cysteine proteases cleave the second peptide bond on the carboxyl side of a hydrophobic residue, the exception being legumain, which cleaves asparaginyl bonds. Specificity is conferred by the structure of the substrate-binding groove on the protease.

Phylogenetic Relationships

The use of the cysteine residue thiolate ion as the nucleophile for peptide-bond cleavage appears to have been derived at least 5 times during evolution using different structural frameworks, and members of each of these protease families are widely distributed. However, the bulk of these enzymes are members of clans CA and CD as classified in the MEROPS database (a regularly updated and invaluable source of information on proteases) (Figure 1). Clan CA contains the earliest identified cysteine proteases, such as the papaya protease papain, and the mammalian proteases, for example, cathepsins B, L, S, and K. These enzymes are generally present in lysosomes or in other membrane-bound compartments where they mediate antigen processing and presentation, or they are secreted from the cell. As a group, clan CA proteases show cleavage selectivity for the residue 2 positions N-terminal to the cleavage site. A related subfamily of papain-like enzymes is the calcium-dependent calpains, which are found in the cytoplasm and mediate various intracellular processes.

The second major group of cysteine proteases, clan CD, has been described much more recently and includes the caspases, legumains, several viral proteases, and separase, which is responsible for the separation of sister chromatids during mitosis. Unlike the papain family, these proteases generally show exquisite specificity for cleavage C-terminal to particular amino acid residues – aspartic acid in the case of caspases, asparagine for legumains, and arginine for separase. The caspases are the mediators of apoptosis.

Mechanism

Proteolytic enzymes are very efficient catalysts that can increase the rate of peptide-bond hydrolysis a billionfold. In the case of the cysteine proteases, the catalytic mechanism has been most extensively investigated for the clan CA enzymes. Despite their distinct evolutionary origins, it appears that the mechanism of other cysteine proteases is quite similar (an example of convergent evolution). Cleavage of the peptide bond depends on a thiolate–imidazolium ion pair provided by the cysteine and histidine residues in the active site (Figure 2). The environment of the active site results in an unusually low acid dissociation constant (pK_a) of around 4 for the cysteine thiol group. Nucleophilic attack on the carbonyl group results in the formation of a tetrahedral oxyanion intermediate. This then accepts a proton from the imidazolium group, resulting in the formation of an acyl enzyme intermediate and the release of the C-terminal portion of the substrate. In a second reaction, the acyl enzyme intermediate is deacylated by a water molecule and the remaining portion of the substrate is released.

Control

Activation

As is the case with most proteolytic enzymes, cysteine proteases are synthesized as inactive precursors, allowing them to be delivered to their intended site of action without harming the biosynthetic machinery of the cell. Often these proenzymes are

retained in an inactive form until their services are required. Very different strategies have evolved for the activation of the precursors of the different cysteine protease groups. In the case of the papain family, an N-terminal propeptide blocks access to the fully functional active site by binding in the reverse orientation to that required for peptide-bond cleavage (Figure 1). Its removal via inter- or intramolecular proteolysis liberates the functional protease. In contrast, the caspases undergo dimerization, which signals a rearrangement of the protein to form the previously incomplete active site.

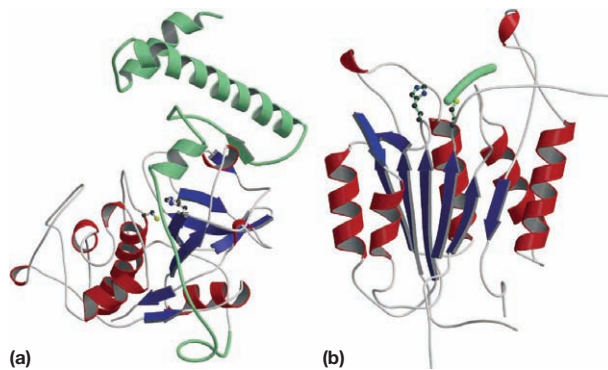


Figure 1 Cysteine protease structure. Ribbon diagrams of representative members of the two principal cysteine peptidase clans: (a) cathepsin K (pdb 1by8), a member of the papain family in clan CA shown as the proenzyme, and (b) caspase-3 (pdb 1cp3), a member of clan CD. The active site cysteine and histidine residues are shown in ball and stick representation. For caspase-3, the position occupied by the substrate is indicated by the green rod. For procathepsin K, the proregion is depicted in green. The extended coil linking the helical region of the propeptide passes through the active site in the region normally occupied by the substrate. Processing of the proforms of papain family members requires the proteolytic removal of the proregion without any change to the structure of the protease component. In contrast, major restructuring of the active site occurs during caspase activation and the active form of the enzyme is a dimer of the structure shown.

Inhibitors

A further level of control for unwanted proteolytic activity is provided by the action of a series of cysteine protease inhibitors. Members of the cystatin family are highly effective inhibitors of papain-like enzymes. The N-terminus of the cystatin molecule binds in the active site cleft in a substrate-like manner, but the inhibitor evades cleavage by the target protease due to the presence of a glycine residue at what would normally be the site of cleavage. Additional elements in the inhibitor bind to the protease to maneuver the N-terminal substrate region just out of the range of the catalytic residues. A separate family of inhibitors, termed inhibitors of apoptosis proteins (IAPs), is available for the caspases. These proteins bind across the active caspase, obstructing normal substrate access but making limited interactions with the protease. Analogous to the propeptides of the papain family members, IAPs ensure their resistance to proteolysis by binding in the reverse direction to that required for substrate cleavage.

Physiology

Lysosomal Proteolysis

Virtually, all eukaryotic cells possess membrane-enclosed organelles known as lysosomes. This structure is the site of breakdown of endocytosed or phagocytosed material, and it is also the location for the turnover of certain cytosolic proteins and organelles. In most eukaryotic cells, the lysosome has the highest concentration of cysteine proteases and other hydrolases, which implies that these enzymes are involved in protein degradation for the recycling of amino acids for protein synthesis. This is indeed likely to be the major physiological function of ubiquitously expressed cysteine proteases such as cathepsin L, legumain, cathepsin B, and cathepsin H. Indeed, the last two enzymes, which fragment polypeptides (endopeptidase action) and then attack the fragments either at the C-terminal end, in the case of cathepsin B (carboxypeptidase

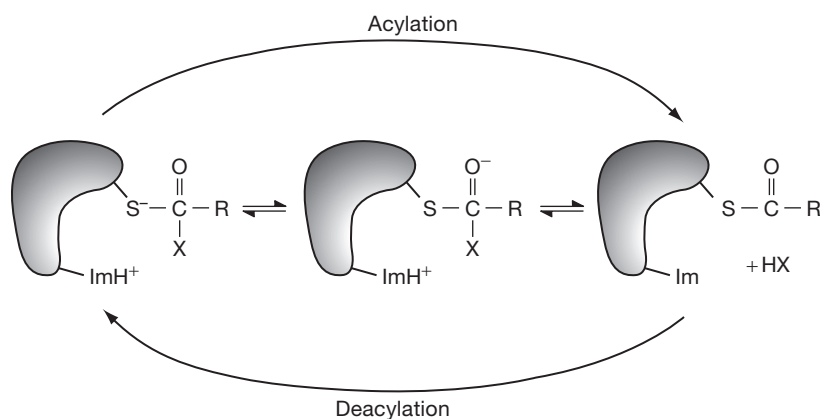


Figure 2 Cysteine protease mechanism. Scheme for the reaction mechanism of cysteine protease action. The active-site cysteine and histidine residues are present as a thiolate (S^-)-imidazolium (ImH^+) ion pair. The substrate is represented by the carbonyl group under attack connected to the N-terminal peptide component R and the C-terminal component X. Upon substrate binding, nucleophilic attack by the thiolate ion results in the formation of a tetrahedral oxyanion intermediate. This, following proton donation from ImH^+ , breaks down to form an acyl enzyme (thiol ester) intermediate with the release of the C-terminal fragment HX. Adapted from Polgár L and Asbóth B (1986) The basic difference in catalysis by serine and cysteine proteinases resides in charge stabilization in the transition state. *Journal of Theoretical Biology* 121: 323–326.

action), or at the amino terminus, in the case of cathepsin H (aminopeptidase action), are functionally well equipped to fulfill a digestive role.

Antigen-Ii Processing

In the major histocompatibility complex (MHC), class II antigen-presenting cells, such as dendritic cells, macrophages, and B lymphocytes, the lysosomal or endocytic localization of some cysteine proteases has allowed them to participate in antigen presentation to T cells. In this process, the antigen is endocytosed and fragmented into peptides of between 13 and 26 amino acids in length. In order to allow the peptides thus generated to bind to the MHC complexes for presentation at the cell surface, the invariant chain Ii must be cleaved in order to remove it from the peptide-binding groove. Cathepsin S has been known for some time to be highly expressed in antigen-presenting cells, and recently the use of selective cathepsin S inhibitors and the generation of cathepsin S-null mice have demonstrated an important, but not solitary, role for this enzyme in Ii processing.

It is likely that a number of endosomal proteases are involved in antigen processing and also that the protease involved depends on the nature of the antigen. In some cases at least, legumain appears to play an important role.

Bone Turnover

Bone is continually being broken down and replenished. This is important in calcium homeostasis, in which soluble calcium is provided by the hydrolysis of the mineral fraction of bone. At the same time, the organic component of bone, consisting largely of type I collagen, is also removed. Although a normal physiological process, under some circumstances bone resorption can outstrip synthesis. This can lead to the condition known as osteoporosis or brittle bone disease, which largely affects postmenopausal women. Other situations in which bone resorption is outside normal homeostatic control include bone metastases and some rare genetic diseases. The process of bone resorption is mediated largely by the multinucleated cell type known as an osteoclast. During periods of active resorption, osteoclasts sitting on bone surfaces generate a ruffled membrane, effectively sealing off the area immediately beneath the cell into which proteolytic enzymes and acid are secreted to remove the organic and mineral components, respectively. Both cysteine proteases and matrix metalloproteinases have been implicated in the hydrolysis of the organic component of bone, and it is likely that the relative contributions played by these different enzymes depend on the type of bone, with cysteine proteases playing a dominant role in the resorption of long bone.

The discovery that pycnodysostosis, an autosomal recessive disease leading to short stature and bone fragility, was due to the deficiency in cathepsin K activity, provided the first indication that this cysteine protease played a major role in bone resorption. Laboratory experiments have demonstrated this enzyme's efficiency at cleaving type I collagen, and consequently it has become a major drug target for the treatment of bone-resorbing diseases.

Other cysteine proteases are likely to be involved in bone resorption. For instance, legumain acts as an inhibitor of bone resorption by inhibiting the formation of multinucleated osteoclasts from their mononuclear precursors. The precise mechanism is not known, but it is not linked to the catalytic activity of the enzyme, and in fact the active component is the C-terminal peptide that is removed upon generation of full catalytic activity.

Thyroglobulin Processing

The distribution of cathepsin K is quite restricted, being found predominantly in the ovary and osteoclasts, the bone-degrading cells. The other site where cathepsin K is highly expressed, alongside the more ubiquitously distributed cathepsins B and L, is in the thyroid follicles, which are the site of storage of thyroglobulin, the precursor of the thyroid hormones. It appears from experiments using single- and double-knockout mice that all three of these cysteine proteases are involved in thyroid hormone generation from thyroglobulin. All are found in the lumen of thyroid follicles, where cathepsin L appears to be involved in both the solubilization of cross-linked thyroglobulin and the subsequent generation of thyroid hormones. In contrast, cathepsin B may be primarily involved in thyroglobulin solubilization and cathepsin K in thyroid hormone generation.

Apoptosis

The importance of programmed cell death to developmental processes has been recognized for many years. The discovery that the proapoptotic *ced-3* gene of *Caenorhabditis elegans* encodes a protease homologous to the mammalian cysteine protease interleukin-1 β -converting enzyme (ICE), together with the demonstration that ICE induces apoptosis in mammals, ignited a great deal of renewed interest in the field, due, in large part, to the desire of the pharmaceutical industry to control the process. This article is too short to do justice to such a large area of research, which has been the subject of a number of good reviews. Briefly, the proteases most intricately involved in the process of apoptosis are the caspases (including ICE), all of which have acute specificity for the cleavage of aspartyl bonds and which function in a cascade reaction. Upstream caspases activate downstream caspases, which cleave proteins whose functions are essential to cell survival. One of the initiators of this cascade event is the occupation of cell surface death receptors of the tumor necrosis factor receptor 1 family, which triggers the activation of caspase 8. Alternatively, the formation of a complex with cytochrome c and Apaf-1 in response to DNA damage leads to the activation of caspase 9. Both of these initiating caspases then activate the downstream executioner caspases, such as caspase 3. In cell killing by cytotoxic T cells or natural killer (NK) cells, this pathway is hijacked in the target cell through the ability of the killing cell's serine proteinase, granzyme B, to activate caspases.

It has become increasingly clear that, under some circumstances, cysteine proteases other than caspases are involved in apoptosis. For instance, cathepsin B may be involved in death receptor-mediated apoptosis of tumor cells if downstream

caspases are inhibited. This process is expedited by tumor necrosis factor- and caspase 8-induced increases in cytosolic cathepsin B and subsequent release from mitochondria of cytochrome c.

Pathology

Cysteine proteases have been implicated in a wide range of pathologies, including cancer, osteoporosis, and arthritis. Here, we focus on their involvement in two other pathological situations that may be less fully appreciated.

Allergenic Proteinases

Proteases represent a disproportionately high number of allergens. Examples include the major allergen from the house dust mite (a cysteine protease that is the product of the *Der p1* gene) and many plant cysteine proteases such as actinidain, bromelain, papain, and glycyI endopeptidase. The normal route for the generation of atopy involves the initiation of a Th2 lymphocyte phenotype and enhanced IgE responses by B cells. Although the mechanisms by which cysteine (and other) proteases favor a switch to a Th2-like response are not completely elucidated, a link with the catalytic activity of these enzymes has been made. The mechanisms may be many and varied and include the cleavage (and activation?) of IgE and IL-2 receptors and disruption of epithelial tight junctions, thus enhancing access by allergens to immune cells. However, a systematic analysis of potential molecular targets for these enzymes has not been undertaken.

Parasitic Proteinases

Many parasites contain cysteine proteases in their phagocytic vacuoles, where they are involved in the digestion of engulfed material to be used in satisfying the nutritional requirements of the organism. As such, these enzymes are potential therapeutic targets. Examples include *Entamoeba histolytica*, a causative agent of amebic dysentery; trypanosomatids that cause

African sleeping sickness and Chagas' disease; schistosomal organisms and the plasmodial parasites that are the causative agents of malaria; and a number of parasitic helminths. In the case of plasmodial parasites, the cysteine proteases are known to be instrumental in the degradation of host hemoglobin, the principal source of amino acids for the parasite, because enzyme inhibition blocks hemoglobin degradation and parasite development. In a mouse model of the human disease, an orally administered inhibitor of the enzyme produced a 40% cure rate, and a delayed progression of the disease in the remaining animals. New treatments for this and other parasitic diseases that kill millions every year are badly needed, due to the increasing resistance to currently available drugs.

See also: **Bioenergetics:** Cell Death by Apoptosis and Necrosis; **Protein/Enzyme Structure Function and Degradation:** Calpain; Metalloproteases.

Further Reading

- Barrett AJ, Rawlings ND, and Woessner JF (eds.) (1998) *Handbook of Proteolytic Enzymes*. London: Academic Press.
- Chapman HA, Riese RJ, and Shi G-P (1997) Emerging roles for cysteine proteases in human biology. *Annual Review of Physiology* 59: 63–88.
- Cygler M and Mort JS (1997) Proregion structure of members of the papain superfamily: Mode of inhibition of enzymatic activity. *Biochimie* 79: 645–652.
- Delaissé J-M, Engsig MT, Everts V, et al. (2000) Proteinases in bone resorption: Obvious and less obvious roles. *Clinica Chimica Acta* 291: 223–234.
- Elliott E and Sloane BF (1996) The cysteine protease cathepsin B in cancer. *Perspectives in Drug Discovery and Design* 6: 12–32.
- Friedrichs B, Tepel C, Reinheckel T, et al. (2003) Thyroid functions of mouse cathepsins B, K, and L. *Journal of Clinical Investigation* 111: 1733–1745.
- Lang A, Hörler D, and Baici A (2000) The relative importance of cysteine peptidases in osteoarthritis. *Journal of Rheumatology* 27: 1970–1979.
- Polgár L and Asbóth B (1986) The basic difference in catalysis by serine and cysteine proteinases resides in charge stabilization in the transition state. *Journal of Theoretical Biology* 121: 323–326.
- Shakib F, Schulz O, and Sewell H (1998) A mite subversive: Cleavage of CD23 and CD25 by Der p1 enhances allergenicity. *Immunology Today* 19: 313–316.
- Thornberry NA and Lazebnik Y (1998) Caspases: Enemies within. *Science* 281: 1312–1316.

Cytochrome *bc*₁ Complex (Respiratory Chain Complex III)

B L Trumpower, Dartmouth Medical School, Hanover, NH, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 528–534, © 2004, Elsevier Inc.

Glossary

Protonmotive Q cycle The mechanism of electron transfer by which the cytochrome *bc*₁ complex transfers electrons from ubiquinol to cytochrome *c* and links the electron transfer to proton translocation. The mechanism is so named because the enzyme oxidizes ubiquinol on one side of the membrane and re-reduces ubiquinone on the other side in a cyclic manner, thus bringing about transmembrane

movement of protons, carried through the membrane as hydrogen atoms on the quinol hydroxyl groups.

Ubiquinone 2,3-Dimethoxyl-5-multiprenyl-6-methyl-1,4-benzoquinone, where the multiprenyl group is an isoprenoid side chain consisting of 6–10 isoprenyl groups, is a lipid-soluble benzoquinone that is reduced to quinol by various dehydrogenases and reoxidized to quinone by the cytochrome *bc*₁ complex.

The cytochrome *bc*₁ complex (ubiquinol: cytochrome *c* oxidoreductase complex, E.C. 1.10.2.2) is an energy-transducing, electron-transfer enzyme located in the inner mitochondrial membrane of oxygen-utilizing eukaryotic cells, where it participates in cell respiration. A functionally similar but structurally simpler version of the *bc*₁ complex is located in the plasma membrane of many, but not all, bacteria, where it takes part in respiration, denitrification, nitrogen fixation, and cyclic photosynthetic electron transfer, depending on the species. In all of these organisms, the *bc*₁ complex oxidizes a membrane-localized quinol and reduces a water-soluble, *c*-type cytochrome and links this redox reaction to translocation of protons across the membrane in which the *bc*₁ complex resides. The *bc*₁ complexes from mitochondria of several species have been crystallized and the mechanism of the enzyme is generally well understood, although some questions remain outstanding.

Function of the *bc*₁ Complex

The cytochrome *bc*₁ complex participates in respiration in oxygen-utilizing cells and also participates in electron transfer in numerous bacteria that utilize alternative terminal-electron acceptors in addition to oxygen. The functional relationship of the *bc*₁ complex to other redox enzymes in these linked electron-transfer systems is illustrated in [Figure 1](#). In mitochondria the *bc*₁ complex is a confluence point for reducing equivalents from the various dehydrogenases and it is essential for mitochondrial respiration, as there is no alternative route to oxidize ubiquinol by molecular oxygen. Bacteria such as *Paracoccus denitrificans* that have a *bc*₁ complex usually have alternative mechanisms to oxidize ubiquinol or otherwise bypass the *bc*₁ complex. Consequently, the enzyme is not essential in these organisms. Hydroxyquinone analogs of ubiquinone, such as atovaquone, inhibit the *bc*₁ complex and are used therapeutically against fungal (e.g., pneumocystis) and parasitic (e.g., malaria) infections, but are ineffective against bacteria in which the enzyme is not essential. In some photosynthetic bacteria, such as *Rhodobacter*, the *bc*₁ complex is essential for cyclic photosynthetic electron transfer in the absence of a terminal electron acceptor, but is not essential when the cells are

growing heterotrophically. In all of these organisms, the *bc*₁ complex is located in the organelle or plasma membrane and converts the energy associated with electron transfer from ubiquinol to cytochrome *c* into an electrochemical proton gradient across the membrane in which the enzyme resides. The resulting protonmotive force is used by the cell for energy-requiring reactions, including adenosine triphosphate (ATP) synthesis, ion and metabolite transport, and flagellar motion.

Structure, Composition, and Properties of the *bc*₁ Complex

Protein Subunits

All cytochrome *bc*₁ complexes contain three protein subunits with redox prosthetic groups, a diheme cytochrome *b* containing a relatively high-potential *b*_H heme and a lower potential *b*_L heme, cytochrome *c*₁, and a Rieske iron–sulfur protein with a 2Fe–2S cluster. In some bacteria, such as *P. denitrificans* and *Rhodospirillum rubrum*, the *bc*₁ complex contains only these three subunits. Other bacteria, including the *Rhodobacter*, contain a fourth subunit of unknown function that lacks prosthetic groups. *Rhodobacter* can adapt to grow photosynthetically when this subunit is absent, but the stability of the enzyme is compromised, suggesting at least a structural role.

The cytochrome *bc*₁ complexes of mitochondria contain as many as seven or eight supernumerary subunits. The enzyme from *Saccharomyces cerevisiae* contains 10 subunits, while those from bovine heart and *Schizosaccharomyces pombe* contain 11. The difference between the 10- and 11-subunit enzymes is due to the fact that the 11-subunit enzymes contain one small subunit that is formed when the presequence that targets the Rieske protein to the mitochondria is cleaved from the protein. The cleaved presequence is retained in the *bc*₁ complex as an additional subunit. In *S. cerevisiae*, the presequence is cleaved from the Rieske protein in two steps and degraded, thus that *bc*₁ complex contains only 10 subunits.

The functions of the supernumerary subunits in the mitochondrial enzymes are not known. They are not required for the electron transfer and proton translocation activities of the enzyme, since the three-subunit enzyme from *Paracoccus*

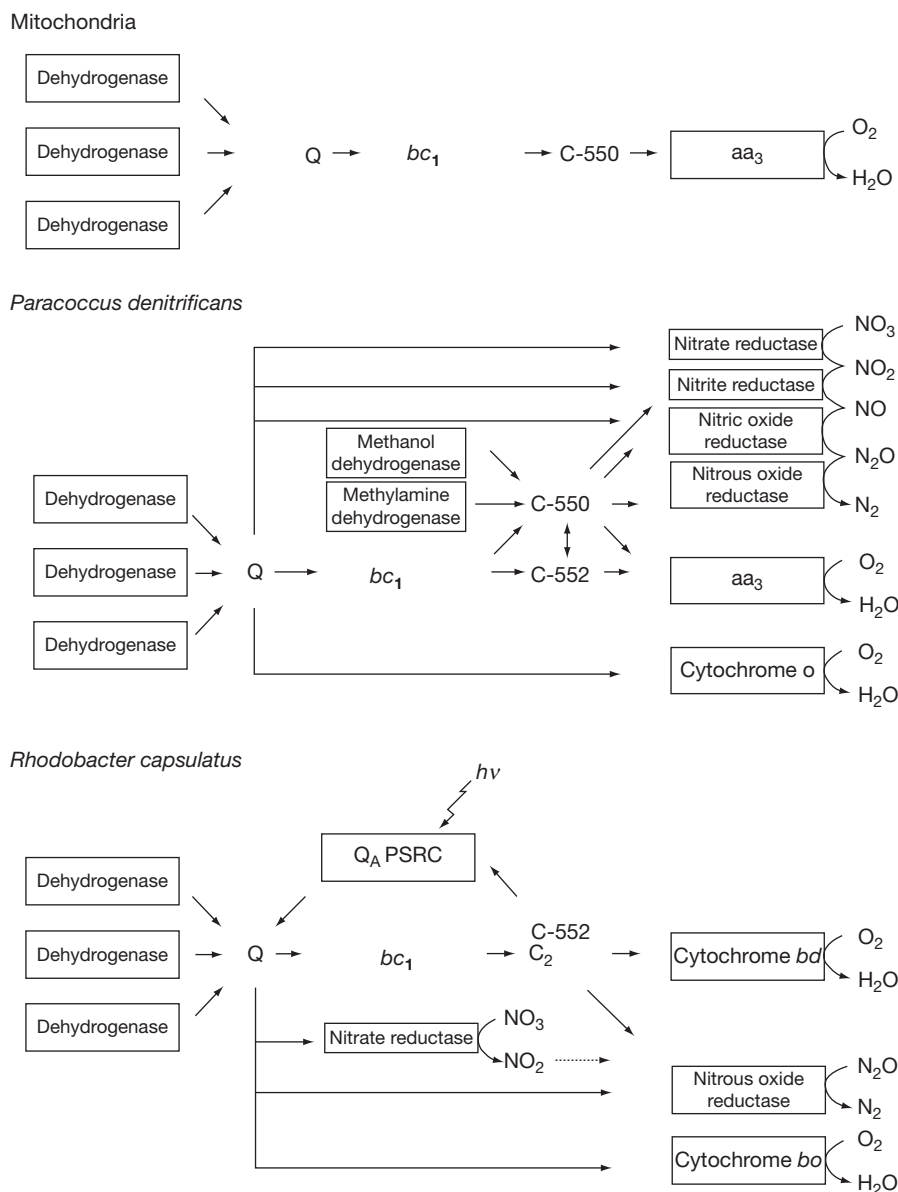


Figure 1 Function of the cytochrome *bc*₁ complex in mitochondria and bacteria. In mitochondria, the *bc*₁ complex oxidizes ubiquinol that is formed by reduction of ubiquinone by various dehydrogenases including glycerol phosphate dehydrogenase, succinate dehydrogenase, NADH dehydrogenase, and ETF dehydrogenase. In bacteria, such as *Paracoccus denitrificans*, ubiquinol can be oxidized by various routes that bypass the *bc*₁ complex, including a terminal oxidase, cytochrome *o*, which oxidizes ubiquinol directly, without the intervention of cytochrome *c* and cytochrome *c* oxidase. These bacteria can also use alternative terminal electron acceptors, such as nitrate, in which case the *bc*₁ complex is also not essential for electron transfer from the quinol. In photosynthetic bacteria, such as the *Rhodobacter*, the *bc*₁ complex participates in cyclic electron transfer when the bacteria are grown photosynthetically, and is essential under those growth conditions. However, *Rhodobacter* can also grow heterotrophically on a variety of nutrients and can use either oxygen or nitrogen in the form of nitrate and nitrous oxide as terminal electron acceptor. Under these conditions, the ubiquinol that is formed by the various dehydrogenases can be oxidized by the *bc*₁ complex and electrons transferred via the *c* cytochromes to oxygen or nitrous oxide, or electrons from the quinol can bypass the *c*-cytochrome pool and be transferred directly to nitrate, nitrous oxide, or oxygen.

has the same electron transfer and proton translocation functionality as the mitochondrial enzymes. Possible functions for these nonredox subunits include docking sites for ternary complex formation with the dehydrogenase and oxidase complexes, regulation of half-of-the-sites activity of the dimeric enzyme, structural stability in the relatively long-lived eukaryotes, and protection against oxygen radicals.

Spectroscopic and Thermodynamic Properties of the *bc*₁ Complex

The *bc*₁ complex is visibly red due to the cytochromes. Optical spectra of the yeast enzyme in the visible region of the spectrum are shown in Figure 2. When the enzyme is reduced by dithionite, a difference spectrum of the dithionite-reduced

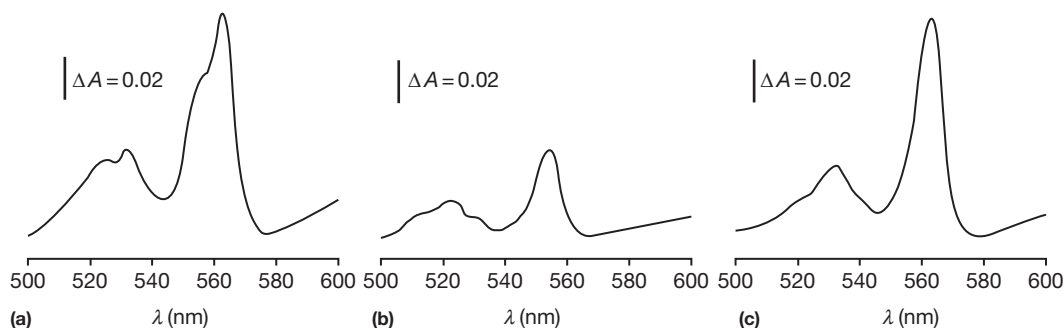


Figure 2 Optical spectra of the yeast cytochrome bc_1 complex. Difference spectra of (a) the dithionite-reduced vs. ferricyanide-oxidized enzyme and (b) the ascorbate-reduced versus ferrocyanide-oxidized enzyme. (c) Calculated difference spectrum, obtained by subtracting the spectrum of the ascorbate-reduced enzyme from the spectrum of the dithionite-reduced enzyme.

versus ferricyanide-oxidized enzyme exhibits an absorption maximum at 562 nm due to ferro-cytochrome b and a shoulder at ~ 553 nm due to ferro-cytochrome c_1 . When the enzyme is reduced with ascorbic acid, the reduced versus ferricyanide-oxidized difference spectrum consists of only cytochrome c_1 , with a maximum at 553 nm (Figure 2(b)). The extinction coefficient of reduced versus oxidized cytochrome c_1 at 553 nm versus 539 nm is $17.5 \text{ mM}^{-1} \text{ cm}^{-1}$, which was obtained from the purified protein. By subtracting the spectrum of c_1 from that of $b + c_1$, it is possible to derive a calculated difference spectrum of the cytochrome b as shown in Figure 2(c). The extinction coefficient of reduced versus oxidized cytochrome b at 562 versus 575 nm is $25.6 \text{ mM}^{-1} \text{ cm}^{-1}$, which was determined by measuring the fluorescence quenching upon stoichiometric binding of antimycin to the bovine enzyme. This value is an average of the two b hemes. The individual extinction coefficients of the b_H and b_L hemes have not been determined for most bc_1 complexes, but the contribution from the b_H heme at 562 is significantly greater than that from the b_L heme, which has a split absorption maximum with peaks at 564–566 and ~ 558 nm.

The Rieske iron–sulfur protein is a pale yellow-gold color when isolated from the bc_1 complex in the oxidized form but bleached to colorless when reduced and thus makes no contribution to the optical spectrum of the reduced enzyme. The redox status of the $2\text{Fe}-2\text{S}$ cluster is monitored by electron paramagnetic resonance (EPR) spectroscopy as the reduced cluster is paramagnetic. The Rieske cluster exhibits a unique EPR spectrum with peaks at approximately $g_x = 1.76$, $g_y = 1.90$, and $g_z = 2.03$, resulting in $g_{av} = 1.90\text{--}1.91$, which is atypical of binuclear iron–sulfur clusters. The unusual EPR spectrum of the Rieske protein is due to bis-histidine coordination of one of the two iron atoms, which was first suggested by spectroscopic studies and confirmed by crystal structures of the enzyme.

The midpoint potentials of the redox centers in the bc_1 complex determine the thermodynamic parameters and in some cases control electron transfer reactions within the enzyme. In *S. cerevisiae* the midpoint potentials at pH 7 are as follows: cytochrome c_1 , +240 mV; Rieske iron–sulfur cluster, +280 mV; cytochrome b_H , +120 mV; and cytochrome b_L , –30 mV. Although there are species differences in these midpoint potentials, it is generally true that the midpoint potential of the Rieske center is $\sim 20\text{--}40$ mV more positive than that of c_1 , and that of the b_L heme is 60–120 mV lower

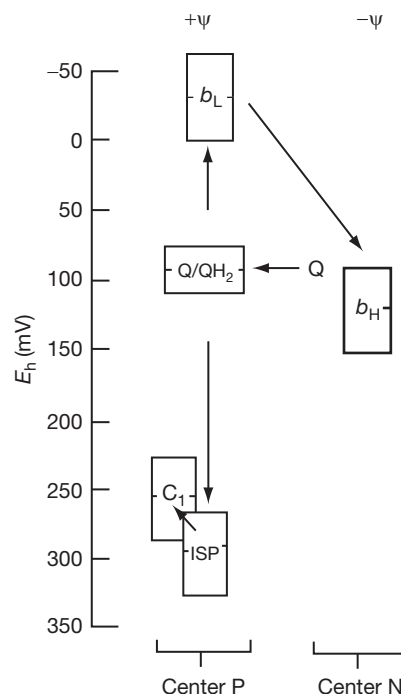


Figure 3 Thermodynamic profile of the *Saccharomyces cerevisiae* cytochrome bc_1 complex. The figure illustrates the thermodynamic relationship between the redox components of the cytochrome bc_1 complex. Arrows indicate electron transfer reactions in the enzyme. The redox groups are arranged vertically according to their oxidation–reduction potentials, with the center of the boxes positioned vertically at the midpoint potentials of the redox groups. The open boxes depict the approximate range of potentials spanned by the redox components as their oxidation–reduction status varies in response to changes in rates of electron transfer. The midpoint potentials of the redox components are cytochrome b_L , –30 mV, cytochrome b_H , +120 mV, cytochrome c_1 , +240 mV, and Rieske iron–sulfur protein, +280 mV.

than that of the b_H heme. The midpoint potentials of the b hemes are influenced by the detergents used to purify the enzyme, and may be similarly influenced by bound phospholipids. The midpoint potential at pH 7 of ubiquinone is +90 mV in the membrane, but this value may be altered by preferential binding of either quinone or quinol to reaction

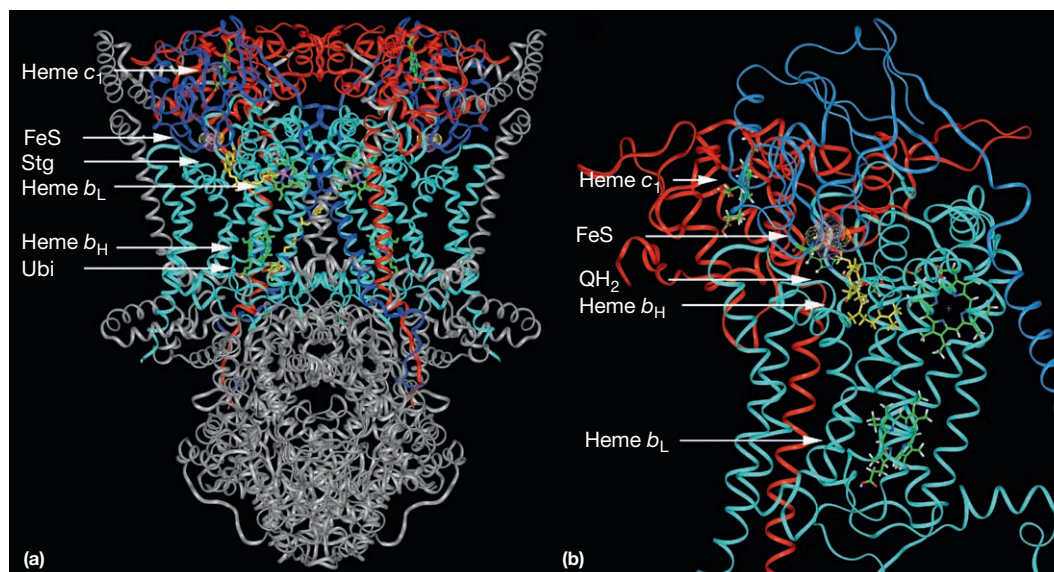


Figure 4 Crystal structure of the cytochrome *bc*₁ complex. The structure in (a) is of the yeast *bc*₁ complex with stigmatellin bound, resolved to 2.3 Å by Hunte and co-workers, and shows all of the subunits of the enzyme dimer, except subunit 10, which is lost during purification. The structure is oriented so that the mitochondrial matrix would be at the bottom and the intermembrane space at the top of the figure. Cytochrome *b* is colored cyan, cytochrome *c*₁ red, and the Rieske protein blue. The supernumerary subunits are colored gray. The heme groups (heme *b*_L and heme *b*_H) are shown as stick structures, colored green. Stigmatellin (Stg) bound at center P in only one monomer is shown as a stick structure, with carbon atoms colored yellow and oxygen atoms red. Ubiquinone (Ubi) bound at center N in only one monomer is also shown as a stick structure, with carbon atoms colored yellow and oxygen atoms red. The iron-sulfur cluster (FeS) is shown with its van der Waals radius, with iron atoms colored purple and sulfur atoms colored yellow. The structure in (b) is of a functional monomer, consisting of cytochrome *b* and cytochrome *c*₁ from one monomer and the Rieske protein from the other. Ubiquinol (QH₂) containing two isoprenyl groups is hydrogen-bonded between His181 of the Rieske protein and Glu272 of cytochrome *b* to form a putative electron-donor complex. An energy-minimized structure of the docked quinol was obtained by replacing stigmatellin in the crystal structure, followed by molecular dynamics and energy minimization calculations. The protein subunits, heme groups, and ubiquinol are colored as in (a).

sites in the enzyme. These midpoint potentials result in a thermodynamic profile of the *bc*₁ complex as shown in Figure 3.

Three-Dimensional Structure of the *bc*₁ Complex

Cytochrome *bc*₁ complexes from bovine, chicken, and yeast mitochondria have been crystallized and their structures solved to atomic resolution. The yeast enzyme, shown in Figure 4, is the highest-resolution structure available at this time. The mitochondrial *bc*₁ complex is a symmetrical, oligomeric dimer, in which the three redox subunits are surrounded by a periphery of nonredox, supernumerary subunits. The dimeric structure of the enzyme was suspected from hydrodynamic measurements and confirmed by the first crystal structure of the bovine enzyme. Although the bacterial enzyme has not yet been crystallized, it will presumably turn out to be a dimeric structure too, but lacking the large number of supernumerary subunits.

One unusual aspect of the dimeric structure is that the iron-sulfur protein spans the dimer. The Rieske protein is anchored in one monomer by a single transmembrane helix, as shown in Figure 4(a), while the peripheral domain that contains the 2Fe-2S cluster is located in the other monomer, where it forms part of the ubiquinol oxidation site. Crystal structures of the chicken *bc*₁ complex in the absence and presence of stigmatellin, an inhibitory ligand, also provided

striking evidence that the peripheral domain of the Rieske protein moves back and forth between positions proximal to cytochrome *b* and cytochrome *c*₁. This suggested that movement of the Rieske protein facilitates electron transfer, which was confirmed by site-directed mutagenesis studies that demonstrated that such movement of the Rieske protein was essential for enzyme activity.

Functionally, the enzyme consists of the cytochrome *b* and *c*₁ subunits from one monomer and the Rieske protein from the other, as shown in Figure 4(b). The crystal structures established the location of the sites of ubiquinol oxidation and ubiquinone reduction at topographically separated sites within the enzyme. The structures also confirmed the transmembrane disposition of the *b* hemes, which form a conduit through which electrons are cycled across the membrane in which the enzyme resides. These structural details provided the final confirmatory evidence of the protonmotive Q cycle mechanism of the enzyme.

Mechanism of the Enzyme

The Protonmotive Q Cycle

The *bc*₁ complex oxidizes ubiquinol and transfers two electrons to two molecules of cytochrome *c*. During this electron transfer reaction, two protons are taken up from the mitochondrial

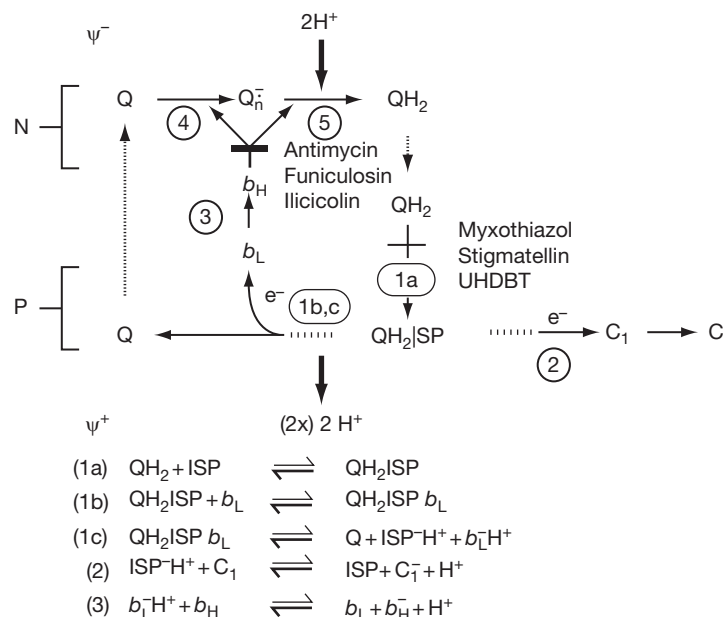


Figure 5 Mechanism of electron transfer through the cytochrome *bc*₁ complex. The scheme shows the pathway of electron transfer from ubiquinol to cytochrome *c* through the redox centers of the cytochrome *bc*₁ complex. The circled numbers designate electron transfer reactions. Dashed arrows designate movement of ubiquinol (QH₂) or ubiquinone (Q) between center P on the positive side of the membrane and center N on the negative side of membrane, and movement of the iron-sulfur protein between cytochrome *b* and cytochrome *c*₁. Solid black bars indicate sites of inhibition by antimycin, funiculosin, illicicoline, myxothiazol, UHDBT, and stigmatellin. In reaction 1, ubiquinol oxidation delivers two electrons divergently to the Rieske iron-sulfur cluster and the *b*_L heme, two protons are released, and the resulting ubiquinone leaves center P. In reaction 2, the reduced Rieske cluster oscillates to within electron transfer distance of cytochrome *c*₁, resulting in electron transfer from the iron-sulfur cluster to the *c*₁ heme. In reaction 3, an electron is transferred from the *b*_L to *b*_H heme, which in turn reduces ubiquinone to ubiquinol (reaction 4). Following oxidation of a second ubiquinol at center P and reduction of the *b* cytochromes the *b*_H heme reduces ubiquinol to ubiquinol (reaction 5), accompanied by uptake of two protons at center N.

matrix or bacterial cytoplasm and four protons are deposited on the other side of the membrane. The mechanism by which the *bc*₁ complex links the electron transfer and proton translocation reactions is the protonmotive Q cycle, as shown in Figure 5. In the Q cycle mechanism proton translocation is the net result of topographically segregated reduction of quinone and reoxidation of quinol on opposite sides of the membrane, with protons being carried across the membrane as hydroxyl hydrogen atoms on the quinol. The sites where quinone is reduced and quinol is oxidized are referred to as center N and center P, respectively, as they are located toward the electronegative and electropositive sides of the membrane. Protons are taken up at center N, carried across the membrane by the quinol, and released at center P.

In the first step of the Q cycle, ubiquinol is oxidized at center P in a reaction that divergently transfers the two electrons from ubiquinol to the Rieske iron-sulfur cluster and the cytochrome *b*_L heme (reactions 1a–1c in Figure 5). In reaction 2, the reduced Rieske cluster oscillates to within electron transfer distance of cytochrome *c*₁ and an electron is transferred from the iron-sulfur cluster to the *c*₁ heme. Two protons are released from center P coincident with ubiquinol oxidation. In reaction 3, an electron is transferred from the *b*_L to *b*_H heme, which in turn reduces ubiquinone to ubiquinol (reaction 4). Following oxidation of a second ubiquinol at center P and reduction of the *b* cytochromes the *b*_H heme reduces ubiquinol to ubiquinol (reaction 5), accompanied by uptake of two protons at center N.

There is evidence that the mitochondrial *bc*₁ complex functions by an alternating-site mechanism, in which binding of ubiquinol in one monomer exerts negative cooperativity on binding of ubiquinol in the other monomer. If the enzyme can be switched from a form in which ubiquinol oxidation alternates between the two monomers to a form where both monomers are simultaneously active, it would provide a mechanism to regulate the activity of the enzyme, perhaps in response to cellular energy needs. Whether such a regulatory mechanism exists and, if so, how it operates, is not known.

Proton Conduction Pathways

In the Q cycle mechanism, protons are carried across the membrane as hydrogen atoms on the hydroxyl groups of ubiquinol. However, the sites where ubiquinol is oxidized at center P and ubiquinol is reduced at center N are not freely accessible to the bulk aqueous phase at the membrane surface. Consequently, the linkage of proton chemistry to electron transfer requires mechanisms for moving protons to and from the aqueous phase and the hydrophobic sites of quinol and quinone redox reactions.

The crystal structures of the *bc*₁ complex with bound stigmatellin (Figure 4(a)) show that the inhibitor is hydrogen-bonded between His181 (in the yeast numbering system) of the Rieske protein, which is a ligand to the iron-sulfur cluster, and Glu272 of cytochrome *b*. As stigmatellin presumably

mimics an intermediate in ubiquinol oxidation, this suggests that ubiquinol forms an electron-donor complex involving hydrogen bonds between the quinol hydroxyl groups and these two amino acids, as shown in [Figure 4\(b\)](#).

When ubiquinol reduces the Rieske protein, an electron is transferred to the iron–sulfur cluster, while a proton from the quinol hydroxyl group protonates the imidazole nitrogen on His181 of the Rieske protein. The Rieske protein acts as a hydrogen carrier and releases the proton to the aqueous phase when it moves proximal to cytochrome *c*₁ and is oxidized. The second proton from ubiquinol is transferred to Glu272 and then to a propionate of the cytochrome *b* heme as an electron is transferred to the *b*_L heme. From the crystal structure of the yeast enzyme it appears that this proton can then access the aqueous surface from the *b*_L heme via a conserved arginine (Arg79). In this manner, two protons are released from center P with minimal charge separation as the two electrons are divergently transferred from the quinol to the high-potential and low-potential electron acceptors ([Figure 3](#)).

The crystal structure of the yeast enzyme has also revealed pathways for conduction of two protons into center N. One of these pathways, the (CL)/K pathway, involves a conserved lysine and a bound cardiolipin, which may function as a buffer to concentrate protons at the entrance to the pathway. The second proposed proton conduction pathway at center N is referred to as the E/R pathway and involves a conserved glutamate and a conserved arginine. The high resolution of the yeast crystal structure also reveals that one or two bound water molecules may serve as direct proton donors during quinone reduction. Notably, whereas ubiquinol oxidation at center P proceeds with minimal charge separation, ubiquinone reduction at center N begins with substantial charge separation as electrons and protons enter the reaction site from opposite directions, and the redox reaction is accompanied by charge compensation, which enhances the reaction rate.

Inhibitors

There are numerous inhibitors that act on the cytochrome *bc*₁ complex and that bind specifically to either the ubiquinol oxidation site at center P or the ubiquinone reduction site at center N. These have proved to be especially useful in establishing the mechanism of the enzyme and for selectively blocking electron transfer reactions at either of the two reaction sites in the enzyme to facilitate kinetic analysis. In addition, some of the inhibitors have antifungal and antiparasitic activity.

Antimycin, funiculosin, and ilicicolin H are produced by microorganisms as defensive toxins and inhibit the *bc*₁ complex by binding to center N with near stoichiometric affinity. Under conditions of catalytic turnover, they block reduction of ubiquinone by cytochrome *b* by binding to the ubiquinone

reduction pocket at a site proximal to the *b*_H heme, causing electrons to accumulate in the *b* hemes. Under presteady-state conditions where the enzyme is reduced by ubiquinol, they block reduction of the *b*_H heme by quinol.

Stigmatellin, also produced by a microorganism, inhibits the *bc*₁ complex by binding to the ubiquinol oxidation pocket, simultaneously forming hydrogen bonds to the imidazole nitrogen of His181 on the Rieske protein and a carboxyl oxygen of Glu272 on cytochrome *b*. In this manner, the inhibitor locks the Rieske protein in a conformation proximal to cytochrome *b*. Stigmatellin raises the midpoint potential and shifts the EPR spectrum of the iron–sulfur cluster, indicative of its interaction with the Rieske protein.

Hydroxyquinones, such as UHDBT, a benzoxothiazole (3-undecyl-2-hydroxy-1,4-benzoxothiazole) and atovaquone, a hydroxynaphthoquinone (2-[*trans*-4-(4'-chlorophenyl) cyclohexyl]-3-hydroxy-1,4 hydroxy-naphthoquinone), also bind to the ubiquinol oxidation pocket and form a hydrogen bond to His181 of the Rieske protein. Like stigmatellin, these ligands raise the midpoint potential and change the EPR spectrum of the iron–sulfur cluster. However, the hydroxyquinone inhibitors do not hydrogen-bond to Glu272 of cytochrome *b*; instead, the crystal structure of the yeast *bc*₁ complex with an analog of UHDBT bound shows Glu272 rotated toward the heme propionate, which suggested the proton conduction pathway described above.

Methoxyacrylates, such as myxothiazol, also bind to the ubiquinol oxidation pocket but do not interact directly with the Rieske protein. These inhibitors bind proximal to the *b*_L heme and affect its spectral and thermodynamic properties.

See also: **Bioenergetics:** Heme Synthesis; Membrane-Associated Energy Transduction in Bacteria and Archaea; Mitochondrial DNA; Respiratory Chain and Adenosine Triphosphate Synthase; **Protein/Enzyme Structure Function and Degradation:** Heme Proteins; Ubiquitin-Like Proteins.

Further Reading

- Hunte C, Koepke J, Lange C, Roßmann T, and Michel H (2000) Structure at 2.3 angstrom resolution of the cytochrome *bc*₁ complex from the yeast *Saccharomyces cerevisiae* co-crystallized with an antibody Fv fragment. *Structure* 8: 669–684.
- Link TA (1999) The structures of Rieske and Rieske-type proteins. *Advances in Inorganic Chemistry* 47: 83–157.
- Trumpower BL (1990) The protonmotive Q cycle: Coupling of proton translocation to electron transfer by the cytochrome *bc*₁ complex. *Journal of Biological Chemistry* 265: 11409–11412.
- Von Jagow G and Link TA (1986) Use of specific inhibitors of the mitochondrial *bc*₁ complex. *Methods in Enzymology* 126: 253–271.
- Zhang ZL, Huang LS, Shulmeister VM, et al. (1998) Electron transfer by domain movement in cytochrome *bc*₁. *Nature* 392: 677–684.

Cytochrome *c*

H Tuppy and G Kreil, University of Vienna, Vienna, Austria

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 535–538, © 2004, Elsevier Inc.

Glossary

Apoptosis A means for the multicellular organism to eliminate cells that are no longer needed or are damaged, also called 'programmed cell death'.

Cytochromes Literally cell pigments, heme proteins chiefly involved in cell respiration and energy supply.

Heme An iron porphyrin complex that is attached (as a prosthetic group) to protein moieties in the red

blood pigment hemoglobin, in the cytochromes and in many other heme proteins.

Mitochondria Intracellular organelles, the main site of oxidative metabolism that supplies the cell with utilizable energy in the form of adenosine triphosphate. In the course of cellular respiration, elementary oxygen is taken up by mitochondria and reduced to water. The electrons required for reduction are transmitted to oxygen via the respiratory chain, one of the electron transporters being cytochrome *c*.

Cytochrome *c* is a heme protein that is present in and can easily be isolated from mitochondria of all eukaryotic organisms. The amino acid sequence of the protein moiety was among the first sequences that could be elucidated. This was the starting point for comparative studies about sequence variations found in cytochrome *c* from a wide range of species. A phylogenetic tree constructed on the basis of this information was found to be biologically significant and became exemplary for subsequent studies on molecular evolution. The function of cytochrome *c* in the respiratory chain as an electron carrier is well established. More recently, an additional role of cytochrome *c* was discovered: its release from mitochondria into the cytosol triggers apoptosis – the programmed cell death.

Introduction

Cytochromes are proteins that contain heme as their prosthetic group and whose principal biological function, in the cells of animals, plants, and microorganisms, is electron transport. The foundations of the knowledge of heme proteins and their roles as electron carriers in cell respiration were laid by David Keilin (1887–1963), ~80 years ago. In the cytochromes, the iron which is coordinately linked with four nitrogen atoms within the prosthetic group and with two additional ligands provided by the protein moiety, can alternate between a reduced Fe^{2+} and an oxidized Fe^{3+} state. Different classes of cytochromes (*a* type, *b* type, and *c* type) differ in the nature of their heme prosthetic groups. They can be observed and differentiated spectroscopically, on the basis of characteristic absorption bands. The absorption peaks of reduced cytochrome *c* and other *c*-type cytochromes (such as *c*₁) are at ~550, 520, 416, and 270 nm. The *b*-type and *a*-type cytochromes absorb visible light at higher wavelengths. In eukaryotic cells, cytochromes of types *a*, *b*, and *c* are found predominantly within mitochondria. In the mitochondrial respiratory chain, cytochrome *c* accepts electrons from complex III, which contains cytochromes *b* and *c*₁ (*bc*₁), and transmits them to complex IV (cytochrome oxidase), which has two heme prosthetic groups (*aa*₃). These respiratory complexes, with their respective cytochromes, are firmly integrated in the inner membrane of the

mitochondrion. Cytochrome *c*, by contrast, is only loosely bound in the space between the inner and outer mitochondrial membranes and shuttles between *bc*₁ and *aa*₃. As shown by Keilin and others, cytochrome *c* can be easily removed experimentally from and reincorporated into isolated mitochondria, their respiratory capacity thereby being impaired and restored, respectively. It has recently been discovered that the release of cytochrome *c* from mitochondria into the cytoplasm is an important step in programmed cell death. This participation in apoptosis of eukaryotic cells is now considered to be another significant function of cytochrome *c*.

Cytochrome *c* and *c*-type cytochromes occur in all eukaryotic organisms. The comparative analysis of the protein structures of cytochromes isolated from a multitude of different species of organisms has made it possible to establish evolutionary relationships between them. The phylogenetic tree based on the slow changes of the structure of cytochrome *c* in the course of evolution is most impressive and biologically relevant.

The *c*-type cytochromes are also present in prokaryotes, where they may function in a respiratory chain or in photosynthesis; examples are cytochrome *c*₂ from *Rhodospirillum rubrum* (purple bacteria) and cytochrome *c*₅₅₁ from *Pseudomonas aeruginosa* (Gram-negative bacteria). These are distantly related to mitochondrial cytochromes *c*; however, a more detailed discussion of these diverse hemoproteins is outside the scope of this article.

Attachment of the Heme Prosthetic Group to Cytochrome *c* Apoprotein

The prosthetic group of *c*-type cytochromes, such as cytochrome *c* and *c*₁ in mitochondria and cytochrome *f* in chloroplasts, unlike that of *a*- and of *b*-type cytochromes, is covalently linked to the polypeptide moiety. The thiol groups of two cysteinyl residues in a Cys–X–Y–Cys–His peptide motif are attached to two vinyl groups of heme through thioether bonds. This linkage is resistant to heat and hydrolysis, but can be broken with the help of silver or mercury salts. Thus, the polypeptide moiety of cytochrome *c*, or of heme peptides derived from it by proteolytic cleavage, could be obtained. In

eukaryotic cells, the apoprotein of cytochrome *c* is encoded by a nuclear gene, translated on cytoplasmic ribosomes, and translocated into the intermembrane space of mitochondria, where an enzyme, cytochrome *c* heme lyase, combines it with heme. An *in vitro* synthesis of a microbial *c*-type cytochrome from its apoprotein and reduced heme has recently been achieved.

Amino Acid Sequence Studies

Cytochrome *c* is a small, water-soluble protein that can easily be isolated and purified from cells and tissues of many different eukaryotic organisms. In proteolytic digests of cytochrome *c*, one fragment containing the covalently linked heme group can be separated from all the other ones. Studies on the amino acid sequence of these heme peptides isolated from cytochromes *c* of diverse origin thus became feasible. It turned out that certain sequence elements, as the one mentioned in the Introduction, were highly conserved from yeast to mammalian cytochromes *c*.

Expanding on this early work, cytochrome *c* from horse heart was one of the first proteins whose complete amino acid sequence could be determined. It consists of a single polypeptide chain of only 104 residues that could be analyzed even with, by modern standards, rather primitive tools available around 1960. The protein contains many (up to 19) lysine residues, whose positive charges can form ionic bonds with constituents of the inner mitochondrial membrane.

With the horse sequence as a reference, cytochrome *c* isolated from diverse species was subsequently analyzed in quick succession. These included other mammals (human, rhesus monkey, pig, dog, rabbit, whale, and kangaroo), birds (chicken, pigeon, king penguin, and turtle), reptiles (rattlesnake and alligator), amphibia (bullfrog), fish (tuna, carp, and dogfish), and the chordate pacific lamprey. Further, with unabated pace, additional sequences were determined for cytochromes *c* from flies, moths, yeasts, *Neurospora*, and a large variety of plants. By now, we know the sequence of this protein from more than 100 eukaryotic species. As a result of these efforts, cytochrome *c* became the first and still is a standard example for studies on the molecular evolution of proteins. Nuclear genes encoding cytochrome *c* have been analyzed from various species. The mammalian gene contains a small intron in the coding sequence and a larger one upstream of the initiation codon.

The recent progress in the sequencing of whole eukaryotic genomes has also yielded information about the number of cytochrome *c* genes in various species. For example, in the human genome, a single gene for cytochrome *c* is present on chromosome 7, along with 49 pseudogenes located on many different chromosomes. These are mostly processed pseudogenes, which were formed by retro-transcription of a messenger RNA. In rodents, two cytochrome *c* genes are expressed, one of them exclusively in the testis. This latter gene is still present as a pseudogene in the human genome.

Evolutionary Aspects

Cytochrome *c* is a highly conserved protein, which retained many structural characteristics over the eons of evolution of

animals and plants. This makes it an ideal tool for amino acid sequence comparisons over a wide range of species (Figure 1). From the sequence differences, a molecular pedigree can be constructed based on the original assumption, now corroborated by many examples, that the longer ago the common ancestor of two species existed, the more amino acid changes will have accumulated since. As it turned out, the phylogenetic trees constructed on the basis of the sequence of a single protein were surprisingly similar to those derived from comparative anatomy, fossils, etc. Moreover, as cytochrome *c* could be extracted from species for which no reliable data were available from other sources, these could now also be included in this analysis.

In this comparison of cytochromes *c* from diverse eukaryotic species, it was found that more than a quarter of the amino acids have been conserved during more than a billion years of evolution. Among these immutable residues are the two cysteines to which the heme is bound via thioether linkages. The central iron of the planar heme is bound not only to four nitrogen atoms of the porphyrin but also to the side chains of two invariant amino acids of the protein, one being a histidine adjacent to the second cysteine (residue #18 in the vertebrate sequences), the other the sulfur atom of methionine 80, present on opposite sides. Moreover, certain amino acids with hydrophobic side chains, some of which line the crevice where the heme is buried, are also conserved. Surprisingly, it was found that several glycine residues in the sequence were also invariant. Glycine is the smallest amino acid containing only a hydrogen atom where the 19 others have more or less

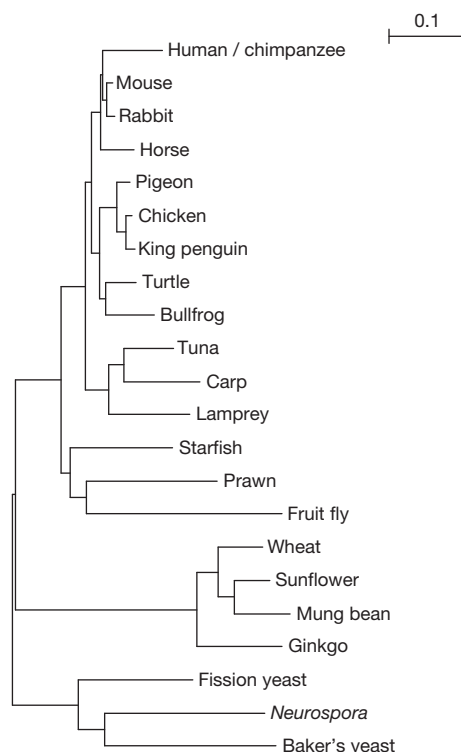


Figure 1 Phylogenetic tree constructed from the amino acid sequences of cytochrome *c* using the ClustalW program. The length of the branches corresponds to evolutionary distances (0.1 = 10% amino acid substitutions).

bulky side chains. From the three-dimensional structure, this first puzzling observation could be explained. At these sites, there is simply no space in this tightly folded protein and every other amino acid would presumably prevent proper folding of the polypeptide chain and thus be deleterious. In addition, some charged amino acids, mostly lysines, have been conserved. These may be essential for the ionic interaction with two of its reaction partners, as well as with acidic phospholipids present in the inner mitochondrial membrane.

Three-Dimensional Structure

Using X-ray crystallographic analysis, the three-dimensional structure of tuna cytochrome *c* in its oxidized and reduced forms could be determined. As shown by these studies, the polypeptide chain is tightly wrapped around the heme, which sits in a deep pocket, exposed only on one edge to the solvent. In accordance with the oil-droplet model of folded proteins, the hydrophobic amino acids are mostly buried inside the protein, some of them surrounding the heme. In the respiratory chain located in the inner membrane of mitochondria, cytochrome *c* shuttles between a reduced ferro- (Fe^{2+}) and ferri- (Fe^{3+}) form. In this redox cycle, the protein changes its conformation, with the reduced form having a more compact structure.

Interactions of Cytochrome *c* in the Respiratory Chain

Cytochrome *c* is loosely bound to the outer surface of the inner mitochondrial membrane. In the respiratory chain, an electron is transferred to cytochrome *c* from cytochrome *bc*₁. These cytochromes are part of a large, membrane-bound complex, also termed complex III composed of 11 proteins. Subsequently, the electron is delivered to the cytochrome *aa*₃ of the complex IV, the cytochrome oxidase, which contains 13 proteins. Some of the molecular details of this electron transport have recently also been clarified. In the crystal structure of the yeast cytochrome *bc*₁ complex with its bound substrate cytochrome *c*, the heme groups of cytochrome *c*₁ and *c* are in close proximity (see Figure 2). This suggests that the redox process takes place by direct heme-to-heme electron transfer. The same may also be true for the oxidation of cytochrome *c* by cytochrome *aa*₃.

In both redox reactions, a proton is pumped across the inner membrane. The proton gradient thus formed drives the adenosine triphosphate (ATP) synthetase motor, which results in the synthesis of ATP.

Cytochrome *c* and Apoptosis

Cells that are damaged or are no longer needed may undergo a form of cell death that is controlled and executed by intracellular programs. One major apoptotic program involves the release of cytochrome *c* from the intermembrane space of mitochondria into the cytosol, together with some other mitochondrial proteins. Once in the cytoplasm, cytochrome *c* combines with an apoptosis activating factor-1 (Apaf-1) and thus triggers the assembly of a multimeric protein complex, the so-called apoptosome. An inactive proform of a proteolytic enzyme, procaspase-9, is then recruited to the apoptosome and activated. Active caspase-9, in turn, activates other

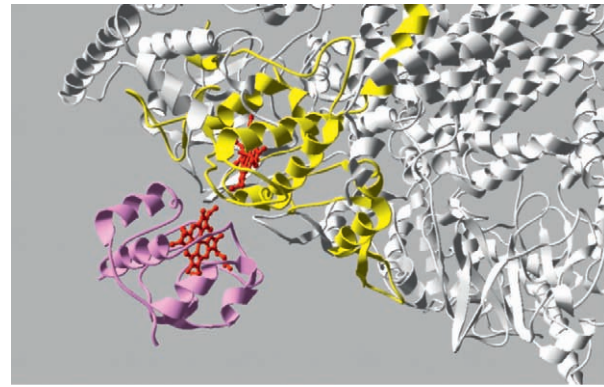


Figure 2 Part of the yeast cytochrome *bc*₁ complex and the bound substrate cytochrome *c* (Protein Data Bank, accession number 1KY0). Cytochrome *c*₁, yellow; cytochrome *c*, magenta; hemes, red.

caspases, whose proteolytic actions finally lead to cell death. The crucial role of cytochrome *c* in this process is shown by the fact that microinjection of cytochrome *c* into the cytoplasm suffices to induce apoptosis in several cell lines. In the interaction with Apaf-1, as in the interaction with cytochrome *c*₁, the exposed heme edge of cytochrome *c* is involved.

When cytochrome *c* molecules are released from the mitochondrion, they have to pass through pores in the outer membrane of the organelle. A group of related proteins, the Bcl-2 family, regulates apoptosis by controlling the permeability of this membrane. Bcl-2 itself and other antiapoptotic proteins are located in the outer mitochondrial membrane and inhibit cytochrome *c* release. Conversely, release of cytochrome *c* is triggered by proapoptotic members of the Bcl-2 family, which are induced by a number of stimuli to translocate to the mitochondria. Since one of the functions of apoptosis is to help the multicellular organism to eliminate cancer cells, the pro- and antiapoptotic Bcl-2 family members are considered to be oncogenic and antioncogenic proteins, respectively. They control whether cytochrome *c*, besides being a vital mediator of electron transfer in respiration, also gets involved in cell death.

See also: [Bioenergetics: Cell Death by Apoptosis and Necrosis; Cytochrome Oxidases, Bacterial; Heme Synthesis; Purple Bacteria: Electron Acceptors and Donors; Purple Bacteria: Photosynthetic Reaction Centers; Protein/Enzyme Structure Function and Degradation: Heme Proteins.](#)

Further Reading

- Dickerson RE (1972) The structure and history of an ancient protein. *Scientific American* 226: 58–70.
- Dickerson RE (1980) Cytochrome *c* and the evolution of energy metabolism. *Scientific American* 242: 98–109.
- Hengartner MO (2000) The biochemistry of apoptosis. *Nature* 407: 770–776.
- Lange C and Hunte C (2002) Crystal structure of the yeast cytochrome *bc*₁ complex with its bound substrate cytochrome *c*. *Proceedings of the National Academy of Sciences of the United States of America* 99: 2800–2805.
- Saraste M (1999) Oxidative phosphorylation at the *fin de siècle*. *Science* 283: 1488–1493.
- Slater EC (2003) Keilin, cytochromes, and the respiratory chain. *Journal of Biological Chemistry* 278: 16455–16461.
- Zhang Z and Gerstein M (2003) The human genome has 49 cytochrome *c* pseudogenes, including a relic of a primordial gene that still functions in mice. *Gene* 312: 61–72.

Cytochrome Oxidases, Bacterial

P Brzezinski and P Ädelroth, Stockholm University, Stockholm, Sweden

© 2013 Elsevier Inc. All rights reserved.

Glossary

Aerobe Organism that uses oxygen (O₂) as the terminal electron acceptor in respiration.

Cytochrome A heme protein serving as an electron carrier.

Proton electrochemical gradient Sum of the contributions of the proton concentration difference (pH) and voltage across a membrane (in units of energy, e.g., kJ mol⁻¹ or eV).

Proton pump Integral membrane protein that translocates protons across the membrane without the use of mobile

proton carriers (which would bodily carry protons across the membrane, e.g., a quinone).

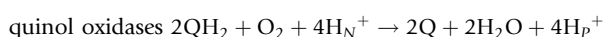
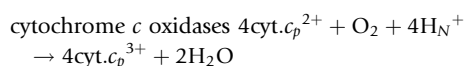
Proton-transfer pathway Typically, an arrangement of protein-bound water molecules and hydrophilic or protonatable amino acid residues used for the transfer of a proton from a donor to an acceptor (where one of these may be the bulk solution).

Redox-active cofactor or redox site Group within a protein (e.g., a metal ion) that can accept and give electrons in its oxidized and reduced states, respectively.

Bacterial cytochrome *c* oxidases are integral membrane proteins that catalyze the four-electron reduction of dioxygen (O₂) to water and oxidation of different types of water-soluble or membrane-anchored cytochromes *c* (cyt.*c*). The oxygen-reducing site of the enzymes consists of a heme–copper center, which is buried within the protein. Hence, cytochrome *c* oxidases belong to the superfamily of heme–copper oxidases, a class of enzymes that also includes the quinol oxidases, which use different types of lipid-soluble quinols (QH₂) instead of cytochrome *c* as their electron donor. The heme–copper oxidases are also often termed ‘respiratory oxidases’ or ‘terminal oxidases’, terms that refer to the enzymes being the last components of the respiratory chains of aerobic organisms. However, not all terminal oxidases belong to the heme–copper oxidase superfamily. One example is the group of so-called alternative oxidases (e.g., in plants), which do not contain any heme groups.

Chemical Process

In the cytochrome *c* oxidases the protons used for O₂ reduction to water (substrate protons) are taken up from the inner (cytosol) side of the membrane, while cytochrome *c* reacts on the opposite, outer (periplasm) side of the membrane. In the quinol oxidases, the quinol binds from the core of the membrane and, upon oxidation, the protons are released to the periplasm. Thus, the reaction catalyzed by the heme–copper oxidases has a direction not only in time (cf. any chemical reaction), but also in space, which is referred to as ‘vectorial chemistry’. The charge distribution across the membrane is such that there are more positive charges on the outer side than on the inner side. Hence, the outer and inner sides are referred to as the positive (*P*) and negative (*N*) sides, respectively:



where the subscripts *N* and *P* refer to the negative and positive sides, respectively. In both cases outlined above, a net of four positive charges are generated on the *P*-side, while four positive charges are consumed on the *N*-side. Consequently, the chemical reactions catalyzed by heme–copper oxidases are arranged topographically in such a way that they result in a charge separation corresponding to the net transfer of one positive charge from the *N*-side to the *P*-side of the membrane per electron transferred to O₂. In addition, for most heme–copper oxidases characterized to date, part of the free energy released in the catalytic reaction is also used to pump (translocate) protons from the *N*-side to the *P*-side of the membrane, with an average stoichiometry of one proton per electron (see Figure 1 and more detailed discussion below). Hence, on average, two charges are transferred across the inner membrane per electron transferred to oxygen. As suggested by Peter Mitchell and formulated in the chemiosmotic theory, this transmembrane proton and voltage gradient generated in part by the heme–copper oxidases is used, for example, for synthesis of adenosine triphosphate (ATP) by the ATP synthase.

Structures

The bacterial heme–copper oxidases consist of several (in many cases four) subunits, located in the (inner) cell membrane of the bacterium. The minimal functional unit of the heme–copper oxidases is the subunit (SU) I–II complex (see also below). The heme–copper oxidase family is defined by the presence in SU I of six conserved histidine residues that coordinate three redox-active metal sites: (1) a six-coordinated heme group which has two axial His ligands and in which the iron ion is in a low-spin state; (2) a five-coordinated heme group with one axial His ligand and in which the iron ion generally is found in the high-spin state; and (3) a copper ion (Cu_B) which is bound by three His ligands. The SU I scaffold, which holds the redox centers, comprises at least 12 transmembrane helices, which span the membrane. The five-coordinated heme is usually denoted with a subscript 3, for example, heme *a*₃. In the examples and discussion below,

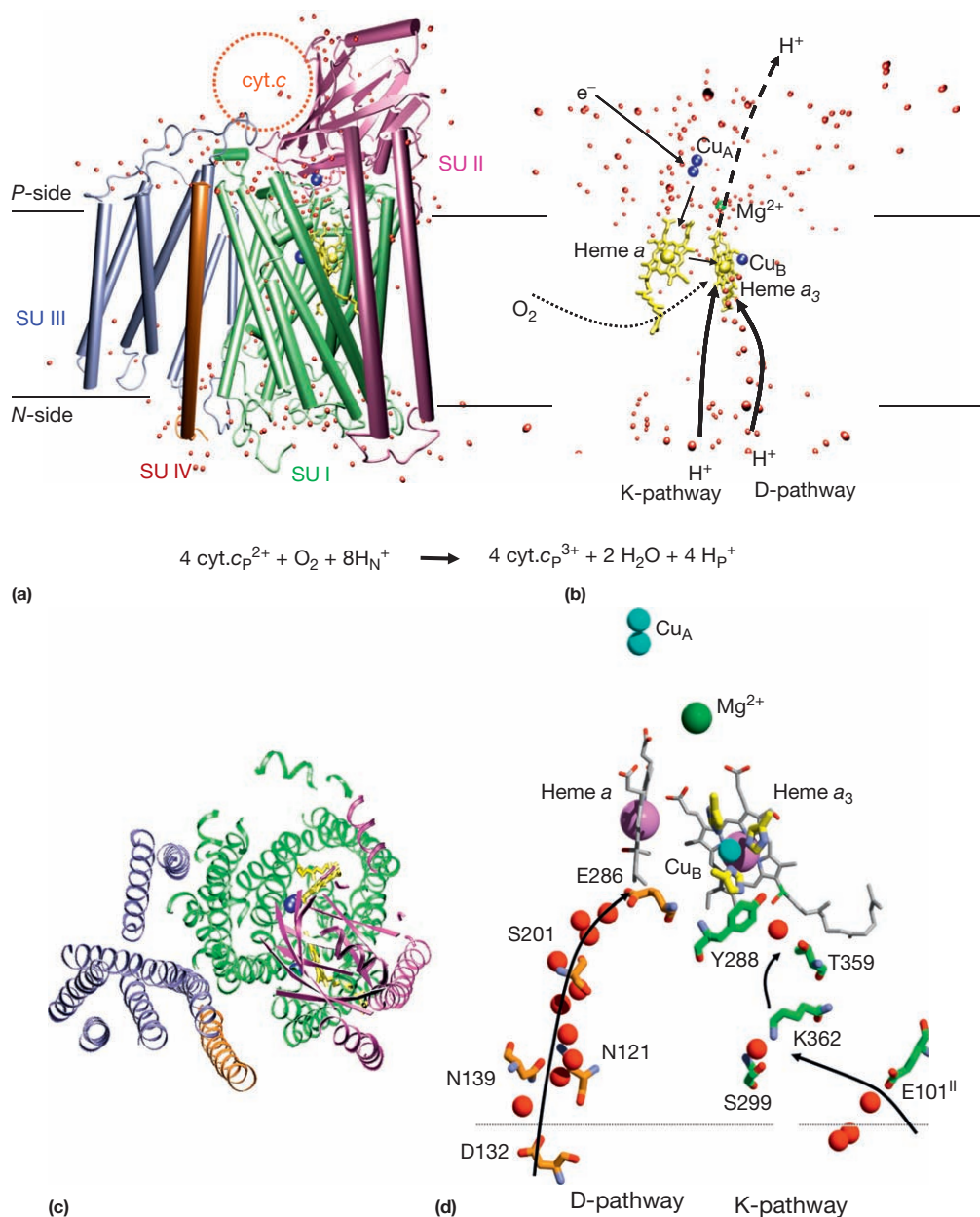


Figure 1 (a) The overall structure of the *Rhodobacter sphaeroides* cytochrome *aa*₃ (cytochrome *c* oxidase). Subunits I, II, III, and IV are shown in different colors as indicated. The heme groups are shown in yellow, the copper ions in blue, and a redox inactive Mg^{2+} ion in green. Three redox active sites, heme *a*, and the catalytic site consisting of heme *a*₃ and Cu_B are found in SU I. The Cu_A center is found in SU II. A likely location of the cytochrome *c* binding site is indicated. Water molecules resolved in the structure are shown as red spheres. (b) Only the cofactors are shown together with the water molecules resolved in the structure. Substrate and pumped protons (thick arrows) are taken up from the inner, cytosol side (N-side) and the pumped protons (dashed arrow) are released on the outer, periplasm side (P-side), as shown. e^- is the electron from cyto *c* (thin arrows). (c) Top view of the enzyme. (d) The redox-active cofactors (and the Mg^{2+} ion) and proton-transfer pathways of cytochrome *c* oxidase. The overall reaction catalyzed by cytochrome *c* oxidase is described by the reaction formula, where the subscripts *P* and *N* refer to the two sides of the membrane. Subunits I–III of the *R. sphaeroides*, *Paracoccus denitrificans*, and the mitochondrial enzymes are highly homologous and their overall structures are very similar.

we call the two hemes of the enzyme hemes *a* and *a*₃, respectively.

Unlike the six-coordinated heme, heme *a*₃ binds ligands such as O_2 , CO, and CN^- . Heme *a*₃ and Cu_B are collectively called the binuclear center, which is the catalytic site at which O_2 is reduced to water. The heme groups may be of different types, such as heme *a*, *b*, or *o*, where the two hemes in SU I may

be of the same or of different types. Heme–copper oxidases are also often given a name that indicates the types of hemes present, for example, cytochrome *aa*₃. Many bacteria have several types of oxidases, which are expressed to various degrees depending on the environmental conditions (e.g., the O_2 concentration). The X-ray structures of heme–copper oxidases that have been determined to date are summarized in Table 1.

Table 1 Structures of heme–copper oxidases from the protein data bank (PDB)^a (the bovine heart enzyme is also included)

Heme–copper oxidase	Source	PDB code	Resolution (Å)	Comment
cyt <i>aa</i> ₃	<i>P. denitrificans</i>	1AR1	2.70	Two-subunit enzyme
		1QLE	3.00	Four-subunit enzyme
		3HB3	2.25	Two-subunit enzyme
		3EHB	2.32	Asn131Asp mutant
cyt <i>aa</i> ₃	<i>R. sphaeroides</i>	1M56	2.30	Wild-type enzyme
		1M57	3.00	Glu286Gln mutant
		3FYE	2.15	Two-subunit enzyme, reduced state
		3FYI	2.20	Two-subunit enzyme, reduced state with CN [−]
		3DTU	2.15	Complexed with deoxycholic acid
		2GSM	2.00	
		2OCC	2.30	Oxidized state
cyt <i>aa</i> ₃	Bovine heart mitochondria	2ZXW	2.50	Fully oxidized state
		2DYR	1.80	Fully oxidized state
		1V54	1.80	Fully oxidized state
		1OCR	2.35	Reduced state
		1V55	1.90	Fully reduced state
		2EIJ	1.90	Fully reduced state
		2EIK	2.10	Fully reduced state, Cd ²⁺ ion binding
		2EIL	2.10	Fully oxidized state, Cd ²⁺ ion binding
		2EIM	2.60	Fully reduced state, Zn ²⁺ ion binding
		2EIN	2.70	Fully oxidized state, Zn ²⁺ ion binding
		2DYS	2.20	Modified by DCCD
		1OCO	2.80	With CO bound
		1OCZ	2.90	With azide bound
cyt <i>ba</i> ₃	<i>T. thermophilus</i>	1EHK	2.40	
		3EH5	2.80	Reduced state
		3S8F	1.8	Lipidic cubic phase
cyt <i>bo</i> ₃	<i>E. coli</i>	1FFT	3.5	
cyt <i>cbb</i> ₃	<i>P. stutzeri</i>	3MK7	3.2	

^aNote. The protein data bank can be found at www.rcsb.org.

A common feature of most (but not all) cytochrome *c* oxidases is a copper center (Cu_A) bound in SU II, which is the primary electron acceptor from cytochrome *c*. Hence, the binding site for the water-soluble cytochrome *c* is located near Cu_A (see **Figure 1**). The Cu_A center is composed of two copper ions coordinated by two –SH groups of two Cys residues (among other ligands), which resembles the [2Fe–2S] centers of iron–sulfur proteins. The Cu_A center can accept and donate one electron switching between the formal charges [Cu¹⁺–Cu¹⁺] and [Cu^{1.5+}–Cu^{1.5+}] in the reduced and oxidized form, respectively. Many cytochrome *c* oxidases that do not have the Cu_A site instead have a cytochrome *c* subunit that protrudes into the inter-membrane space and acts as the primary electron acceptor. In the ubiquinol oxidase from *Escherichia coli*, the ubiquinol binding site was suggested from an analysis of the three-dimensional (3D) structure and from experiments using site-directed mutagenesis to be found in the membrane-spanning part of SU I at a cluster of polar residues exposed to the interior of the lipid bilayer.

Subunit III does not contain any redox-active cofactors. Even though the enzyme with SU III removed displays O₂-reduction activity and pumps protons (although with a stoichiometry of less than one proton per electron), this form of the enzyme becomes inactivated after a limited number of reaction cycles. Consequently, SU III is important for the stability of the protein.

Proton-Transfer Pathways

As seen in **Figure 1**, the O₂-reducing site is located in the membrane-spanning part of the enzyme. Therefore, proton-transfer pathways, leading from the protein surface on the *N*-side to the binuclear center, are needed. Such pathways are typically composed of polar and protonatable amino-acid residues as well as of water molecules forming a hydrogen-bonded chain. The detailed composition of these pathways varies in the bacterial heme–copper oxidases. Two pathways have been identified from an analysis of the 3D structures of heme–copper oxidases determined to date, from a comparison of the amino acid residue sequences of a large number of other heme–copper oxidases, and from the combined use of site-directed mutagenesis and functional studies. These pathways are found in SU I and are named the D- and K-pathways, after residues Asp(D)132 and Lys(K)-362, respectively. We use the *Rhodobacter sphaeroides* cytochrome *aa*₃ amino-acid residue numbering because the structure shown in **Figure 1** is from that enzyme. Both pathways start at the inner-side surface and lead to the catalytic site (**Figures 1** and **2**). A common feature of a large number of oxidases is a Glu residue in the D-pathway, E(I-286) in the *R. sphaeroides* cytochrome *aa*₃, located about 25 Å from D(I-132) and about 10 Å from the binuclear center. In heme–copper oxidases in which the Glu is not

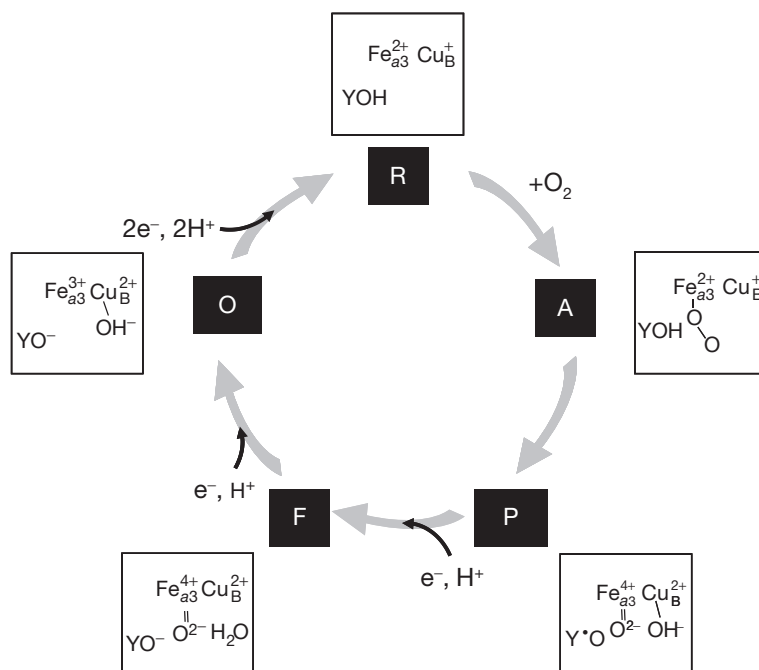


Figure 2 The reaction cycle of cytochrome *c* oxidase. State **O** is the oxidized enzyme. The two-electron-reduced catalytic site (state **R**) binds O_2 (forming state **A**). In the first step of O_2 reduction, a ferryl state is formed (**P**), where one electron and one proton are transferred from a Tyr residue at the catalytic site. The transfer of the third electron to the catalytic site is accompanied by proton uptake from the bulk *N*-side solution and the formation of state **F**. In the next step, the fourth electron is transferred to the binuclear center associated with proton uptake from the bulk *N*-side solution.

conserved, other protonatable residues are found at about the same location in space.

Oxygen Binding and Reduction

Even though the reduction of O_2 to water is a highly exergonic reaction, the O_2 molecule is kinetically stabilized against reduction due to two factors. First, the one-electron reduction of O_2 to form the superoxide radical $O_2^{\bullet -}$ is associated with a positive free energy change, which imposes a thermodynamic barrier in the initial step of the reduction process. Second, in the ground electronic state, O_2 has two unpaired electrons (triplet state), which imposes spin restrictions on many of its reactions. The heme-copper oxidases overcome these problems by binding O_2 to the heme a_3 iron in the high-spin state and by providing a second electron donor, Cu_B , in the immediate vicinity. The rates at which bacterial oxidases catalyze the O_2 reduction are remarkably rapid with turnover numbers of up to several hundred O_2 molecules per second. The rapid uptake of O_2 may be facilitated by hydrophobic channels leading from the membrane core to the catalytic site, as suggested from analyses of the crystal structures of heme-copper oxidases.

The mechanism of the O_2 -reduction has been primarily investigated in detail in cytochromes aa_3 from *R. sphaeroides*, *Paracoccus denitrificans*, the closely related mitochondrial enzyme, as well as in cytochrome bo_3 from *E. coli*. The reaction sequence is most likely the same for all oxidases, but the mechanism described below has been identified primarily from studies of these enzymes.

Initially, the electrons are transferred one at a time to the binuclear center. When both heme a_3 and Cu_B are reduced

(i.e., $Fe_{a_3}^{2+}$ and Cu_B^+ , state **R** in Figure 2), O_2 binds to heme a_3 (the state is called **A**). After binding of O_2 the O–O bond is cleaved, which results in formation of a ferryl group ($Fe_{a_3}^{4+} = O$, see Figure 2), where one electron and a proton are donated by an internal group, presumably a highly conserved Tyr at the catalytic site, which would form a radical. Even though the binuclear center is in a two-electron reduced state, formally four electrons are donated to O_2 in this reaction because two electrons are donated by Fe_{a_3} ($Fe_{a_3}^{2+} \rightarrow Fe_{a_3}^{4+}$), one by Cu_B ($Cu_B^+ \rightarrow Cu_B^{2+}$) and one by the Tyr. For historical reasons, this state is called ‘peroxy’ and denoted **P**.

The transfer of the third electron to the binuclear center is associated with proton uptake from the inside bulk solution. The intermediate that is formed is called ‘ferryl’ and is denoted **F** (Figure 2). It is assumed to have the same chemical structure as **P**, except for the additional electron and proton at the binuclear center, which are presumably transferred to the Tyr radical. The transfer of the next, fourth electron to the binuclear center is also associated with proton uptake from the inside bulk solution and results in formation of the oxidized binuclear center, denoted **O**.

Proton Pumping

In 1977, Mårten Wikström showed that in addition to the charge separation associated with the O_2 reduction reaction, cytochrome *c* oxidase also releases protons to the *P*-side of the membrane. In other words, the enzyme is a proton pump, which translocates one proton from the *N*-side to the *P*-side of the membrane per electron transferred to O_2 . Consequently, since there is no proton carrier in the enzyme that can bodily

move protons across the membrane, there must be a specific mechanism by which the enzyme uses the free energy from reduction of O_2 to translocate protons across the membrane. This function must be reflected in specific structural elements of the molecular machine. The molecular mechanism by which heme–copper oxidases pump protons is not known. The general view is that during the process the enzyme must provide an alternating access of protons to the two sides of the membrane and have this alteration strictly coupled to specific transitions of the catalytic cycle. Such changes in the accessibility for protons to the two sides of the membrane are likely to be achieved by breakage and formation of hydrogen bonds in the proton-transfer pathways, for example, through local rearrangements of amino acid residue side chains. It has been shown that the $R \rightarrow P$ transition is not associated with proton pumping, while the $P \rightarrow F$ and $F \rightarrow O$ transitions are associated with the translocation of one proton in each step. The definite assignment of other reaction steps in which proton translocation occurs and conditions under which this takes place is at present not definitely settled.

As indicated above, in the initial step of the reaction, the O_2 molecule is formally reduced by four electrons (to form state P, [Figure 2](#)). An advantage of such a four-electron reduction of O_2 in a single step is that it prevents accumulation of potentially harmful, partly reduced O_2 intermediates. However, since the formation of P is not associated with proton translocation, this mechanism requires that the free energy available from the O_2 -reduction reaction is conserved within the enzyme, for example, by the formation of high- pK_a internal proton acceptors; and the free energy can be used for proton translocation in the following steps of the reaction.

Some Experimental Techniques Used to Study Heme–Copper Oxidases

Present knowledge about the O_2 -reduction mechanism is derived from the use of a large number of experimental techniques. One of these is the flow-flash technique, pioneered by Quentin Gibson and Colin Greenwood. The enzyme reduced with 2–4 reduction equivalents and with CO bound to heme a_3 is mixed with an O_2 solution after which the CO ligand is flashed off by means of pulsed illumination, which initiates the reaction of the enzyme with O_2 . The application of this technique has made it possible to observe, with microsecond time resolution, the transitions between different states. The progress of the reaction is followed, for example, by observing absorbance changes of the redox sites, by measuring voltage

changes that originate from the movement of charges in enzyme reconstituted in lipid vesicles, or by using time-resolved resonance-Raman spectroscopy to obtain information about the structure of the oxygen intermediates. In addition, Fourier transform infrared difference spectroscopy is used to identify specific residues that undergo changes in their protonation state and/or hydrogen-bonding patterns, and electron paramagnetic resonance spectroscopy is used to explore the electronic configuration and the chemical environment of the redox centers. The use of the techniques discussed above, combined with the use of site-directed mutagenesis, has provided important insights into the catalytic mechanism.

See also: [Bioenergetics: Cytochrome *c*; P-Type Pumps: Copper Pump; Protein/Enzyme Structure Function and Degradation: Heme Proteins.](#)

Further Reading

- Abramson J, Riistama S, Larsson G, et al. (2000) The structure of the ubiquinol oxidase from *Escherichia coli* and its ubiquinone binding site. *Nature Structural Biology* 7: 910–917.
- Brzezinski P and Ådelroth P (2006) Design principles of proton-pumping heme–copper oxidases. *Current Opinion in Structural Biology* 16: 465–472.
- Brzezinski P and Gennis RB (2008) Cytochrome *c* oxidase: Exciting progress and remaining mysteries. *Journal of Bioenergetics and Biomembranes* 40: 521–531.
- Faxén K, Gilderson G, Ådelroth P, and Brzezinski P (2005) A mechanistic principle for proton pumping by cytochrome *c* oxidase. *Nature* 437: 286–289.
- Ferguson-Miller S and Babcock GT (1996) Heme/copper terminal oxidases. *Chemical Reviews* 96: 2889–2907.
- Iwata S, Ostermeier C, Ludwig B, and Michel H (1995) Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Nature* 376: 660–669.
- Qin L, Hiser C, Mulichak A, Garavito RM, and Ferguson-Miller S (2006) Identification of conserved lipid/detergent-binding sites in a high-resolution structure of the membrane protein cytochrome *c* oxidase. *Proceedings of the National Academy of Sciences of the United States of America* 103: 16117–16122.
- Shimokata K, Katayama Y, Murayama H, et al. (2007) The proton pumping pathway of bovine heart cytochrome *c* oxidase. *Proceedings of the National Academy of Sciences of the United States of America* 104: 4200–4205.
- Soulimane T, Buse G, Bourenkov GP, et al. (2000) Structure and mechanism of the aberrant ba_3 -cytochrome *c* oxidase from *Thermus thermophilus*. *EMBO Journal* 19: 1766–1776.
- Svensson-Ek M, Abramson J, Larsson G, et al. (2002) The X-ray crystal structures of wild-type and EQ(I-286) mutant cytochrome *c* oxidases from *Rhodobacter sphaeroides*. *Journal of Molecular Biology* 321: 329–339.
- Wikström MKF (1977) Proton pump coupled to cytochrome *c* oxidase in mitochondria. *Nature* 266: 271–273.
- Wikström M (2004) Cytochrome *c* oxidase: 25 years of the elusive proton pump. *Biochimica et Biophysica Acta* 1655: 241–247.
- Zaslavsky D and Gennis RB (2000) Proton pumping by cytochrome *c* oxidase: Progress, problems, and postulates. *Biochimica et Biophysica Acta* 1458: 164–179.

Cytochrome P-450

R Bernhardt, Saarland University, Saarbrücken, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

CYP Nomenclature for the enzymes of the cytochrome P-450 superfamily.

Cytochrome P-450 Hemoprotein showing an unusual absorption of the reduced CO complex at 450 nm.

Genetic polymorphism Inherited deficiencies of single enzymes such as CYP2D6 and CYP2C19.

Glucocorticoids Steroid hormones regulating glucose metabolism, regulating the stress response of the body, and suppressing inflammation.

Hormone Chemical messenger secreted into the circulating blood.

Mineralocorticoids Steroid hormones regulating the salt and water levels of the body and in this way the blood pressure.

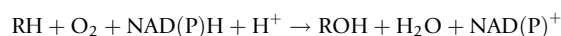
Cytochromes P-450 are ubiquitously distributed multicatalysts, which were discovered approximately 50 years ago. They possess a high degree of complexity and display a broad field of activity. Although already more than 18 000 different P450 genes have been described, new forms are constantly being found, opening up new research fields. They are the key enzymes responsible for the metabolism of many drugs, carcinogens, alkaloids, pesticides, and other important xenobiotics. Moreover, they are involved in a variety of physiological processes such as steroid hormone, eicosanoid, vitamin D, and bile acid biosynthesis. Cytochromes P-450 are hemoproteins that use molecular oxygen to catalyze various reactions. In most cases, they get the electrons necessary for oxygen activation and substrate hydroxylation from NADH (the reduced form of nicotinamide adenine dinucleotide (NAD)) or NADPH (the reduced form of nicotinamide adenine dinucleotide phosphate (NADP⁺)), in general via the action of electron-transferring proteins. Most eukaryotic cytochromes P-450 are membrane bound. Many of them are inducible, but some are constitutively expressed. Defects in some of the mammalian cytochromes P-450 lead to pathological effects, such as congenital adrenal hyperplasia and hypertension, and adverse drug effects.

Cytochrome P-450 – General Aspects

There are two general classes of enzymes involved in oxygen metabolism: oxidases, transferring electrons from a substrate to oxygen, and oxygenases, transferring oxygen to a substrate after reductive splitting of molecular oxygen. Oxygenases can be divided into dioxygenases and monooxygenases. Monooxygenases (mixed-function oxidases) catalyze the incorporation of a single atom of molecular oxygen into a substrate with the concomitant reduction of the other atom to water. The monooxygenases are divided into two classes, the internal and the external monooxygenases. Internal monooxygenases extract two reducing equivalents from the substrate to reduce one atom of dioxygen to water, whereas external monooxygenases use an external reductant. Cytochromes P-450 are external monooxygenases. Initially the microsomal drug and xenobiotic-metabolizing enzymes were referred to as mixed-function oxidases, but in more recent years the term monooxygenase has become the more accepted one.

Cytochromes P-450 got their name from their character as a hemoprotein and from their unusual spectral properties, displaying a typical absorption maximum of the reduced CO-bound complex at 450 nm (**Figure 1**) – cytochrome stands for a hemoprotein, P for pigment, and 450 reflects the absorption peak of the reduced CO-bound complex at 450 nm. The ability of reduced P450 to produce an absorption peak at 450 nm upon CO binding is still used for the estimation of P450 content. The red shift of approximately 30 nm observed in cytochromes P-450 means that the distribution of electron density at the heme is significantly perturbed compared to other cytochromes. It has been documented that it is the thiolate sulfur that causes this effect by means of a direct bond to the iron. The Soret band (named after its discoverer) describes the absorption band of hemoproteins at approximately 380–420 nm.

In general, cytochrome P-450 systems catalyze the following reaction:



They catalyze reactions as diverse as hydroxylation; N-, O-, and S-dealkylation; sulfoxidation; epoxidation; deamination; desulfuration; dehalogenation; peroxidation; and N-oxide reduction. Their substrates include fatty acids, steroids, and prostaglandins, as well as a multitude of foreign compounds such as drugs, anesthetics, organic solvents, ethanol, alkylaryl hydrocarbon products, pesticides, and carcinogens. This diversity of catalyzed reactions and acceptable substrates has attracted researchers from diverse fields to the study of cytochrome P-450 systems (**Figure 2**). In addition to pharmacologists and toxicologists, endocrinologists, physiologists, microbiologists, organic chemists, plant biologists, and environmental scientists also are working on diverse aspects of P450 function and regulation.

It is obvious that this diversity of substrates and catalyzed reactions cannot be managed by only a few different isoforms. In the human genome alone 57 different P450 genes have been found, and in total more than 18 000 different P450 genes have been described (D. Nelson lists the available P450 sequences identified so far on his website). This superfamily of proteins (and their respective genes) is divided into families, subfamilies, and finally into individual members according to

similarities in primary structure. A new nomenclature has been introduced in which CYP and a series of numbers and letters are used to characterize each P450 as a hemoprotein. For example, in CYP1A1, for cytochrome P4501A1 (previously called P450c), the first Arabic number defines the gene family, the following letter defines the subfamily, and the second number defines the individual enzyme. Members of the same gene family are defined as usually having >40% sequence identity to a P450 protein from any other family. Mammalian sequences within the same subfamily are always >55% identical. The numbers of individual P450 enzymes in different species differ, the highest numbers observed so far being in plants.

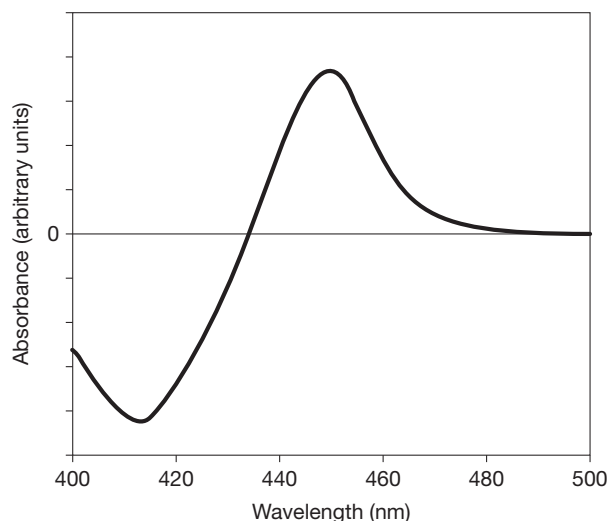


Figure 1 CO-difference spectrum of reduced vs. oxidized cytochrome P-450 (CYP106A2).

Structural Organization of Cytochrome P-450 Systems

As already mentioned, cytochromes P-450 belong to the external monooxygenases. This means that they need an external electron donor to transfer the electrons necessary for oxygen activation and the following substrate hydroxylation. Although a very high variety of the electron supporting systems has been found, two main classes of cytochromes P-450 can be defined with respect to their electron-supporting system: the microsomal type and the mitochondrial/bacterial type (Figure 3).

Microsomal cytochromes P-450 are membrane bound and accept electrons from a microsomal NADPH-cytochrome P-450 reductase, containing flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN). All drug- and xenobiotic-metabolizing cytochromes P-450 isolated so far belong to this class. In addition, CYP102 (P450BM-3) isolated from *Bacillus megaterium* was shown to belong to this class. This P450 system consists of a polypeptide chain with two different domains, one containing the hemoprotein and the other one containing a flavoprotein with FAD and FMN. Many of the other bacterial cytochromes P-450 belong to the second class. They are soluble and obtain the electrons necessary for the reaction mechanism from an NADH-dependent FAD-containing reductase via an iron-sulfur protein of the [2Fe-2S] type. Other bacterial P450s accept electrons from a variety of other partners.

Mitochondrial cytochromes P450, which are involved, for example, in the side-chain cleavage of cholesterol, the 11 β -hydroxylation of deoxycortisol, and the production of aldosterone, also belong to the second class. These cytochromes P-450 are localized in the inner mitochondrial membrane, where the [2Fe-2S] protein, called adrenodoxin in the case of adrenal steroid hydroxylase systems, is a soluble protein of the matrix. The FAD-containing reductase, adrenodoxin reductase, is associated with the inner mitochondrial membrane.

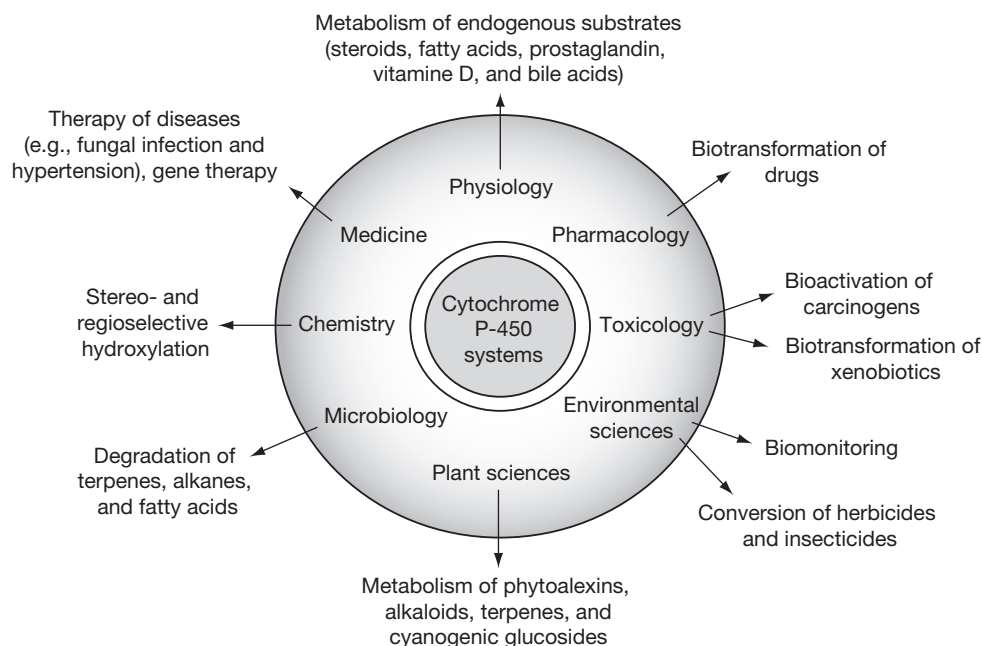


Figure 2 Cytochrome P-450 research and application fields.

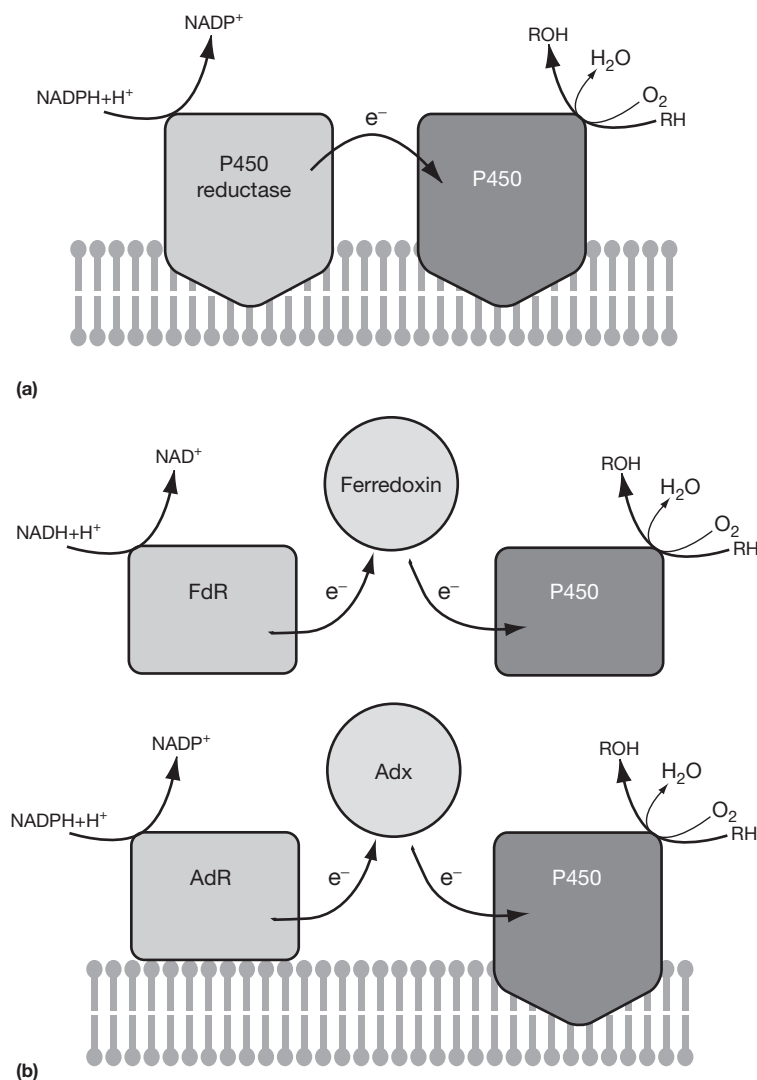


Figure 3 Model for the organization of the two main classes of cytochrome P-450 systems. (a) Microsomal type. (b) Mitochondrial/bacterial type.

The interaction of the cytochromes P-450 with their corresponding electron donors is a necessary prerequisite of the catalytic cycle. Its specificity guarantees a sufficient reaction rate of catalysis and likewise a discrimination between the different potential donors and acceptors of electrons to protect the system from shunt reactions.

Because many different isoenzymes in liver microsomes have to interact with only one type of reductase, the binding site for reductase is very similar or identical on various cytochromes P-450. Salt bridges are responsible for the recognition of reductase by the cytochromes P-450 and for the correct orientation of the proteins to one another. In addition to microsomal reductase, some microsomal cytochromes P-450 are able to accept the second electron from cytochrome b5. Cytochrome b5 has also been shown to exert a differential stimulatory action, depending on the form of cytochrome P-450 and the substrate metabolized.

In mitochondrial steroid hydroxylases and in the camphor hydroxylating bacterial cytochrome P-450 (CYP101) system, a charge-pair interaction mechanism has been demonstrated by

chemical modification, site-directed mutagenesis studies, and structural data of electron-transfer complexes. Like microsomal reductase, the mitochondrial ferredoxin has been shown to deliver electrons to different cytochromes P-450. From the available data, a shuttle model is favored, in which the oxidized ferredoxin interacts first with the ferredoxin reductase to undergo reduction, with the formation of an $\text{Fe}^{3+}-\text{Fe}^{2+}$ iron-sulfur cluster. It dissociates from the reductase and then interacts with the respective cytochrome P-450, to which it delivers this electron before going back to the reductase, and transferring the second electron to the cytochrome P-450. The mechanism of electron transfer between the components of the different cytochrome P-450 systems, one of the fundamental problems in life sciences, is not well understood.

Reaction Cycle

The generally accepted mechanism of cytochrome P-450-dependent substrate conversion is depicted in the overall

scheme presented in Figure 4. The first step of the reaction cycle is the formation of the substrate–enzyme complex. Substrate binding induces structural changes in the cytochrome P-450 that may result not only in a spin shift, but also in changes of the redox potential and in changed binding affinities between interacting components of the cytochrome P-450 system. The second step of the reaction cycle is the introduction of the first electron, either by NADPH-dependent reductase or via a ferredoxin (see Figure 3). Iron ($3+$) is reduced to iron ($2+$). Although in some cytochromes P450 the substrate binding is a prerequisite for electron transfer, this is not a universal requirement. In the third step of the reaction cycle, the hemo-protein reduced by one electron binds the dioxygen molecule. As shown in Figure 4, from this complex a superoxide anion radical can be released. Step 4 of the reaction cycle is the introduction of the second electron. In some instances, another microsomal hemoprotein, cytochrome b_5 , can facilitate catalysis by providing the second electron. From the $(\text{RH})\text{Fe}^{3+}(\text{O}_2^{2-})$ complex, hydrogen peroxide can be split off. The structure of the activated oxygen complex and the precise mechanism of oxygen cleavage are not fully understood yet. Step 5 in the reaction cycle is the removal of the terminal oxygen atom from the dioxygen ligand, that is, the cleavage of the dioxygen bond. Heterolysis of the oxygen–oxygen bond also results in the formation of a putative oxyferryl species. Heterolytic cleavage in P450cam (CYP101), where most of the mechanistic studies have been performed, was shown to be facilitated by hydrogen bonding of the ferric hydroperoxide via a water molecule to a highly conserved threonine residue (Thr252 in CYP101), which is involved in a proton transfer network that promotes oxygen cleavage. The introduction of two electrons and two protons to the putative oxyferryl species may lead to the formation of water. Finally, in step 6 the hydroxylated product dissociates and the cycle can start again. Interestingly, in many but not all cytochromes P450 a shunt reaction can proceed (step 7), in which the substrate can be hydroxylated directly by peroxides such as hydrogen peroxide, cumene

hydroperoxide, and *tert*-butyl hydroperoxide without the necessity of an interaction with an electron-donating system. Taken together, cytochromes P-450 do not catalyze just monooxygenase, but also oxidase and peroxidase reactions. Variations of this scheme for the reaction mechanism of cytochromes P-450 occur with different cytochrome P-450 systems such as thromboxane and prostacyclin synthase, nitric oxide reductase (CYP55A1), and others.

Regulation of Cytochrome P-450 Systems

The regulation of enzyme systems is possible at different levels. When considering cytochrome P-450 systems, the complexity of the reaction cycle, the organization within membrane systems for most of the eukaryotic cytochromes P450, and their organ- and tissue-specific expression imply various possible ways for regulatory mechanisms to work. Interaction between these mechanisms then leads to the tuned response of these enzyme systems to endogenous and exogenous signals in terms of acute and long-term reactivities of cytochromes P450.

First, it is possible to regulate cytochromes P-450 on the level of the organism. Some of the cytochromes P-450 are expressed in age-, tissue-, and sex-dependent manners. These differences in the expression pattern are governed by sex or other hormones. There are also clearly developmental influences on the expression and function of various cytochrome P-450 systems.

Cellular regulation of cytochromes P-450 is extremely complex and can occur on at least two different levels, enzyme induction and posttranslational modification of enzymes. The induction of drug-metabolizing enzymes was noted in the early 1950s. Since then specific inducers for various cytochrome P-450 families and subfamilies have been identified. It can be demonstrated that the inducer functions by enhancing the rate of messenger RNA (mRNA) synthesis of a special cytochrome P450. In some cases, receptors (the best-studied

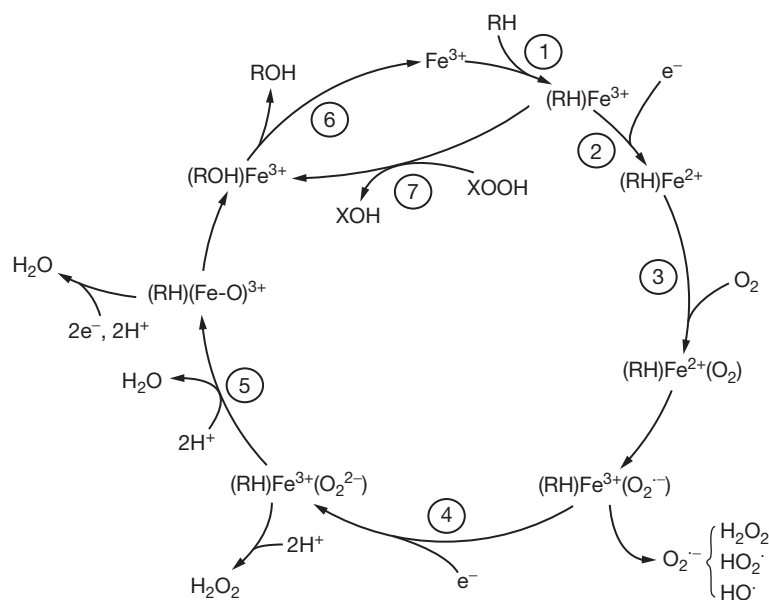


Figure 4 Reaction cycle of cytochromes P450.

system currently is the aryl hydrocarbon, Ah, receptor) are involved in mediating the effect of the inducer. Not all cytochromes P450 are, however, inducible; some of them are constitutively expressed. Moreover, regulation on the epigenetic level and by microRNA has been described for several P450s.

After the biosynthesis of cytochromes P450, the enzymes might be subject to posttranslational modifications to exert a short-term control of activity. One of the most common types of posttranslational modifications in eukaryotic organisms is protein phosphorylation, which has been demonstrated for several steroid hydroxylases such as the microsomal CYP7 and the mitochondrial CYP11A1 and also for the drug-metabolizing microsomal CYP2B4.

Because nearly all the eukaryotic cytochromes P-450 are bound to membranes, they can be regulated also at that level. By contrast, most of the bacterial cytochromes P450, such as CYP101, CYP102, and CYP108 are soluble. The composition of the membrane (e.g., protein–lipid as well as lipid-mediated protein–protein interactions) exerts a functional control on cytochromes P450. Lipids appear to function in at least three ways – they (1) stabilize and induce a functionally active conformation of cytochromes P-450 and the corresponding electron-transfer systems, (2) modulate the electron transfer, and (3) mediate interactions between cytochromes P-450 and the electron-donor systems.

Finally, cytochromes P-450 can be regulated at the molecular level. This includes regulation of the activities by changes of the primary sequence (e.g., by polymorphisms), the spin equilibrium, protein–electron donor interactions, and protein–protein interactions.

Important Functions of Cytochrome P-450 Systems

Cytochromes P-450 and Drug Metabolism

Cytochromes P-450 are able to perform the biotransformation of an enormous variety of substrates. Taking into account all drugs metabolized by enzymes, cytochromes P-450 are responsible for more than two-thirds of these reactions. The most important cytochrome P-450 form for biotransformation in humans is CYP3A4; it is involved in the metabolism of more than one-half of the known drugs and xenobiotics, such as nifedipine, cyclosporin, erythromycin, gestodene, and aflatoxins. This variety of substrates makes CYP3A4 one of the most important enzymes for drug metabolism. The fact that a single cytochrome P-450 is responsible for the metabolism of many different drugs may lead to competition among various drugs for the same enzyme and thus to drug–drug interactions, resulting in higher plasma levels of the less successful (i.e., displaying a lower affinity for the particular cytochrome P450) substrate and even to fatal side effects of the drug treatment.

Another fact complicating drug treatment is the presence of genetic polymorphisms in some of the drug-metabolizing enzymes. It has been shown that 5–10% of Caucasians suffer from a decreased ability to metabolize certain drugs such as debrisoquine and bufuralol (i.e., they are poor metabolizers), due to the presence of nonfunctional alleles of CYP2D6 in these individuals. A smaller portion of patients metabolize the corresponding drugs extremely rapidly (i.e., they are extensive metabolizers) due to having multiple copies of this

gene. These pharmacogenetic effects are being studied in many laboratories and will certainly lead to personalized pharmacotherapy.

Physiological Role of Cytochromes P450

In addition to being involved in drug metabolism and xenobiotica degradation, cytochromes P-450 also play a pivotal role in various physiological processes in humans. They are central in the biosynthesis of steroid hormones (sex hormones, glucocorticoids, and mineralocorticoids), vitamin D, retinoic acid, fatty acids, bile acids, and eicosanoids. Some severe defects are connected to mutations in cytochrome P-450 genes, such as congenital adrenal hyperplasia, which is mostly due to a steroid 21-hydroxylase (CYP21) deficiency. In addition, human essential hypertension can be caused by the overproduction of aldosterone produced by CYP11B2. More recently, products of cytochrome P-450 reactions, such as oxysterols and the endothelium-derived hyperpolarizing factor (EDHF), have been found to play important roles in cellular cholesterol homeostasis, inhibition of cellular proliferation, gene regulation, endothelium-dependent dilation, and enhanced endothelial cell proliferation.

Cytochromes P450 as Drug Targets

While fungal P450s are long established, successful targets in antimycotic therapy, there is increasing evidence that a variety of human P450s can also serve as valuable therapeutic targets. Only ~300 human-genome-derived proteins are used as targets so far, including several P450 forms. Well-known drug targets are the steroidogenic P450s. CYP19A1 catalyzing the aromatization of the steroid A ring represents a target of blockbuster drugs, especially against breast cancer. More recently, CYP11B2 catalyzing the last steps in aldosterone formation became a target for the development of drugs against hypertension and congestive heart failure. Moreover, the differential expression of various P450s in tumors opens up the possibility to use prodrugs leading after their biotransformation to selective toxicity exclusively in these cells.

Cytochrome P-450 Diversity

Cytochromes P-450 are found in all kingdoms – eubacteria, archaeobacteria, fungi, plants, fish, insects, and vertebrates. About 250 different forms have been found in the plant *Arabidopsis thaliana*. It was shown that plant cytochromes P450 play a role in the metabolism of a variety of secondary metabolites, in plant–insect interaction, in herbicide metabolism, and in other vital functions. Due to the diversity of catalyzed reactions, it can be anticipated that plant cytochromes P450 will offer a broad field for future applications.

Cytochromes P450 and Their Biotechnological Application

The interesting chemistry and the high regio- and stereoselectivity of hydroxylation reactions catalyzed by cytochromes

P450 opens up new potentials for practical application of these systems. So far, industrial application is, however, focused on only very few cases due to the complexity of these protein systems. The most prominent example is the production of transgenic plants displaying changed blossom colors (purple carnations and blue roses). The other example is the production of hydrocortisone using a designed strain of Baker's yeast expressing a total of four P450s and engineering 13 proteins.

More recently, site-directed mutagenesis and directed evolution of different P450s has been used to produce valuable drug metabolites and fine chemicals.

Furthermore, the diversity of P450s becoming available through the sequencing of genomes of a variety of organisms opens up new sources for application.

See also: **Bioenergetics: Oxygenases; Metabolism Vitamins and Hormones: Vitamin D.**

Further Reading

- Bernhardt R (1996) Cytochrome P450: Structure, function, and generation of reactive oxygen species. *Reviews of Physiology, Biochemistry, and Pharmacology* 127: 137–221.
- Bernhardt R (2006) Cytochromes P450 as versatile biocatalysts. *Journal of Biotechnology* 124: 128–145.
- Bernhardt R and Waterman MR (2007) Cytochrome P450 and steroid hormone biosynthesis. In: Sigel A, Sigel H, and Sigel RKO (eds.) *Metal Ions in Life Sciences* 3, pp. 361–396. Chichester: Wiley.
- Gomez A and Sundberg MJ (2009) Epigenetic, and microRNA-dependent control of cytochrome P450 expression: A gap between DNA and protein. *Pharmacogenomics* 10: 1067–1076.
- Guengerich FP (2007) Mechanisms of cytochrome P450 substrate oxidation. *Journal of Biochemical and Molecular Toxicology* 21: 163–168. (MiniReview).
- Hakki T and Bernhardt R (2006) CYP17- and CYP11B-dependent steroid hydroxylases as drug development targets. *Pharmacology and Therapeutics* 111: 27–52.
- Hannemann F, Bichet A, Ewen KM, and Bernhardt R (2007) Cytochrome P450 systems – biological variations of electron transport chains. *Biochimica et Biophysica Acta* 1770: 330–344.
- Ingelman-Sundberg M, Sim SC, Gomez A, and Rodriguez-Antona C (2007) Influence of cytochrome P450 polymorphisms on drug therapies: Pharmacogenetic, pharmacoeconomic, and clinical aspects. *Pharmacology and Therapeutics* 116: 496–526.
- Meyer UA (2000) Pharmacogenetics and adverse drug reactions. *Lancet* 356: 1667–1671.
- Meyer UA (2004) Pharmacogenetics – five decades of therapeutic lessons from genetic diversity. *Nature Reviews Genetics* 5: 669–676.
- Ogata J, Kanno Y, Itoh Y, Tsugawa H, and Suzuki M (2005) Plant biochemistry: Anthocyanin biosynthesis in roses. *Nature* 435: 757–758.
- Ruckpaul K and Rein H (eds.) (1989–1994) *Frontiers in Biotransformation*, vol. 1–9. Berlin: Akademie-Verlag.
- Schuster I and Bernhardt R (2007) Inhibition of cytochromes P450. Existing and new promising therapeutic targets. *Drug Metabolism Reviews* 39: 481–499.
- Szcebara FM, Chandelier C, Villeret C, et al. (2003) Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nature Biotechnology* 21: 143–149.

Cytokines

S J Levine, National Institutes of Health, Bethesda, MD, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Andrea L. Wurster and Michael J. Grusby, volume 1, pp. 550–555, © 2004, Elsevier Inc.

Glossary

Apoptosis Programmed cell death.

Chemokine Chemotactic cytokines that mediate the tissue localization of immune, inflammatory, vascular and cancer cells.

Cytokine Small protein intercellular messengers that mediate key biological functions.

Interleukin Cytokines that were initially thought to solely mediate signaling between leukocytes, but have since been recognized to be produced and act on a variety of cell types.

Trans-signaling The ability of soluble cytokine receptor to function as an agonist and mediate signaling by cells that do not express the receptor.

Cytokines are small protein intercellular messengers that mediate key biologic functions related to inflammation, host defense, tumor surveillance, and innate and adoptive immune responses. Efforts to identify a soluble factor produced by leukocytes that functions as an endogenous pyrogen led to the discovery of the first cytokine, named ‘interleukin-1 (IL-1)’, which was purified in 1977 and cloned in 1984. Although interleukins were initially thought to solely mediate signaling between leukocytes, it has since been recognized that they can be produced and act upon a variety of cell types. Multiple additional classes of cytokines have since been identified, which include the tumor necrosis factor superfamily, chemokines, and the transforming growth factor family.

Interleukins

Significant progress has been made in interleukin biology since the discovery of IL-1. There are currently 37 named interleukins to which functions have been identified, with one additional interleukin yet to be named. Interleukins can be grouped into several families that are discussed below.

IL-1 Family

The IL-1 family is comprised of 11 members, which include IL-1 α (IL-1 F1), IL-1 β (IL-1 F2), IL-1 receptor antagonist (IL-1Ra, IL-1 F3), IL-18 (IL-1 F4), IL-33 (IL-1 F11), IL-36 α (IL-1 F6), IL-36 β (IL-1 F8), IL-36 γ (IL-1 F9), IL-36 receptor antagonist (IL-36Ra, IL-1 F5), IL-37 (IL-1 F7), and IL-38 (IL-1 F10). Activation of IL-1 family members is highly regulated and coordinated to prevent excessive inflammatory responses. IL-1 β and IL-18 are synthesized as inactive proforms that are activated following cleavage of their pro-domains by the cysteine protease, caspase-1. Caspase-1 functions within the context of molecular platforms, called ‘inflammasomes’, which are pattern-recognition receptors that sense exogenous and endogenous danger signals, such as pathogen-associated molecular patterns (PAMPs) and danger-associated molecular patterns (DAMPs). Activation of the NLRP3 inflammasome results in the clustering and activation of caspase-1. Activated

caspase-1 cleaves the proforms of IL-1 β and IL-18 to generate mature, secreted pro-inflammatory interleukins that play key roles in cellular responses to infection and injury.

The IL-1 receptor family is similarly comprised of 11 members that are characterized by extracellular immunoglobulin domains and an intracellular toll-IL-1 receptor (TIR) domain that interacts with adaptor proteins to mediate signaling. For example, both IL-1 α and IL-1 β can bind the type I IL-1 receptor (IL-1R1), which then recruits the IL-1 receptor accessory protein, IL-1RAcP, to function as a co-receptor chain (Figure 1). This facilitates the recruitment of myeloid differentiation primary response protein 88 (MyD88) via a TIR domain interaction, as well as IL-1R-associated kinase 4 (IRAK4), and IRAK1 or IRAK2, which form the Myddosome complex. Activated IRAK1 then dissociates from the Myddosome complex to activate tumor necrosis factor receptor (TNFR)-associated factor 6 (TRAF6), which signals via downstream pathways, such as nuclear factor- κ B (NF- κ B) and mitogen-activated protein (MAP) kinases.

IL-18 signals via binding to IL-18R α , which recruits the co-receptor, IL-18R β , to induce signaling. IL-18 binding protein (IL-18BP), which is distantly related to the IL-1 receptor family, is a soluble protein that binds and sequesters IL-18, thereby preventing activation of cellular IL-18 receptors. ST2 (IL-33R α and IL-1RL1), the receptor for IL-33, and IL-1Rrp2 (IL-1RL2), which is the receptor for IL-36 α (IL-1 F6), IL-36 β (IL-1 F8), and IL-36 γ (IL-1 F9), also recruit the shared receptor subunit IL-1RAcP to activate downstream signaling pathways. The receptors for IL-37 (IL-1 F7) and IL-38 (IL-1 F10) have yet to be identified. In addition to binding and signaling via cell-surface receptors, the pro-forms of IL-1 α and IL-33 can translocate to the nucleus and thereby modulate gene transcription.

Multiple layers of regulation exist to prevent excessive or undesired IL-1 signaling. IL-1 receptor antagonist (IL-1Ra) avidly binds IL-1R1 without recruiting IL-1RAcP and thereby antagonizes the activity of IL-1 β and IL-1 α . IL-1RII functions as a decoy receptor that binds IL-1 β with greater affinity than IL-1R1, but has a short cytoplasmic domain that does not bind Myd88 and therefore does not signal. IL-1RII can also recruit and sequester IL-1RAcP to attenuate IL-1 signaling. IL-36 receptor antagonist (IL-36Ra) binds IL-1Rrp2, but does

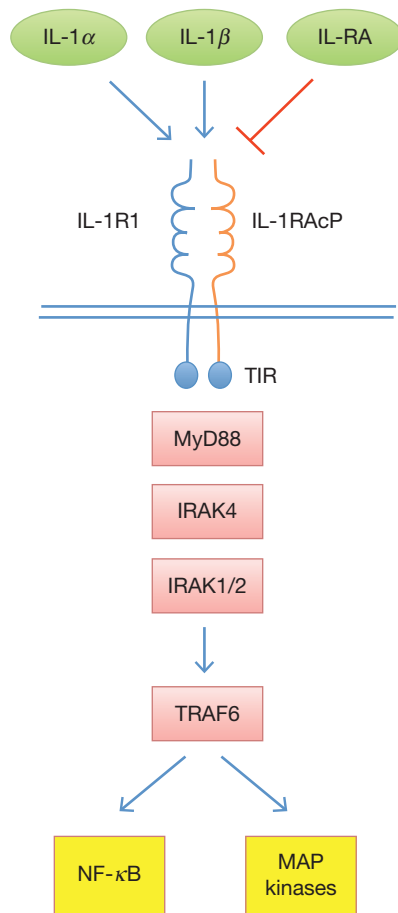


Figure 1 Interleukin 1 signaling.

not mediate the recruitment of IL-1RAcP and thereby functions as a receptor antagonist that attenuates the biological activity of IL-36 α , IL-36 β , and IL-36 γ . A soluble form of ST2 that binds and inhibits IL-33 activity can be generated by alternative messenger RNA (mRNA) splicing. Lastly, single immunoglobulin IL-1R-related (SIGIRR) molecule or TIR8 is an IL-1 receptor family member expressed by epithelial cells that does not signal due to the absence of a functional TIR domain and thereby attenuates signaling by IL-1, IL-18, and IL-33.

IL-1 family members play key roles in inflammation. Administration of a recombinant IL-1RA has been used clinically to successfully attenuate the manifestations of auto-inflammatory syndromes characterized by excessive IL-1 signaling, such as familial cold auto-inflammatory syndrome (FCAS), neonatal-onset multisystem inflammatory disease (NOMID), and Muckle-Wells syndrome (MWS). These auto-inflammatory syndromes are caused by mutations in the cryopyrin or NALP3 genes that result in spontaneous inflammasome formation and increased IL-1 β secretion. IL-1 may also play a role in the pathogenesis of gout due to inflammasome activation by uric acid crystals, as well as in type II diabetes mellitus and ischemia-induced inflammation. IL-18, which was initially cloned as an interferon-gamma-inducing factor that augments natural killer cell activity, promotes Th1 responses and is synthesized by macrophages, dendritic cells, epithelial cells, and keratinocytes.

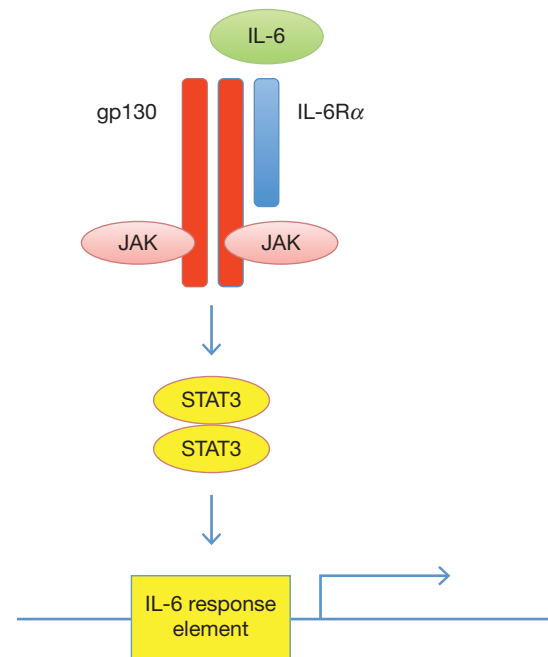


Figure 2 Interleukin 6 signaling.

By contrast, IL-33 promotes Th2 responses and has been implicated in the pathogenesis of asthma and atopic dermatitis.

IL-6 Family

The IL-6 family, which are class I cytokines (see below), is comprised of IL-6, IL-11, IL-27, IL-31, ciliary neurotrophic factor (CNTF), cardiotrophin-1 (CT-1), cardiotrophin-like cytokine (CLC), leukemia inhibitory factor (LIF), neuropoietin (NPN), and oncostatin M (OSM). IL-6 was initially identified in 1986 as a B-cell stimulatory factor that promotes immunoglobulin G (IgG) production. IL-6 is a pro-inflammatory cytokine with pleiotropic biologic functions that include the regulation of immune and inflammatory responses, as well as the modulation of neoplasia, metabolism, regeneration, and neuronal signaling. Furthermore, IL-6 has been implicated in the pathogenesis of rheumatoid arthritis, multiple sclerosis, asthma, diabetes, inflammatory bowel disease, and colitis-associated colon cancer. IL-6 is produced by a diverse variety of cells, with macrophages, fibroblasts, and endothelial cells representing the main cellular sources.

IL-6 and IL-11 bind to IL-6R α and IL-11R α , respectively, which then associate with gp130 homodimers (**Figure 2**). All IL-6 family members, except for IL-31, signal via receptors that contain glycoprotein 130 (gp130) as a common subunit. The gp130 subunit does not possess intrinsic kinase activity, but instead noncovalently associates with janus kinases (JAK1, JAK2, and TYK2) which undergo autophosphorylation and activation upon ligand binding, with subsequent activation of signal transducer and activator of transcription (STAT) transcription factors, STAT3 or STAT1. IL-6 can also activate the transcription factor CAAT-enhancer-binding protein (C/EBP), as well as signal via mitogen-activated protein kinases (MAPKs). LIF, CNTF, OSM, CT-1, NPN, and NNT-1 signal via

complexes comprised of the LIF receptor (LIFR) and gp130. IL-27 signals via the IL-27 receptor alpha chain (IL-27RA and WSX1) and gp130, while OSM signals via a complex comprised of the OSM receptor (OSMR) and gp130. By contrast, gp130 is not necessary for IL-31 signaling which occurs via the OSMR and the gp130-like receptor.

IL-6 can also activate cells that do not express IL-6R via a process called 'trans-signaling'. This occurs via an interaction between membrane-bound gp130, which is expressed on most cells, and a soluble form of IL-6R that binds soluble IL-6. Thus, in contrast to soluble forms of most receptors that are antagonists, soluble IL-6 receptors serve as agonists.

IL-10 Family

The IL-10 family is comprised of six members, which include IL-10, IL-19, IL-20, IL-22, IL-24, and IL-26. There are also four viral homologs produced by the Epstein-Barr virus, the equine herpesvirus type 2, Orf parapoxvirus, and simian cytomegalovirus. In addition, there are three IL-10 family members that are classified as type III interferons, IL-28 α (IFN- λ 1), IL-28 β (IFN- λ 2), and IL-29 (IFN- λ 3). Despite weak sequence identity, IL-10 family members share significant structural similarity and are characterized by a compact, bundle-like shape that is comprised of six or seven α -helices. The IL-10 protein exists as a homodimer, whereas IL-19 and IL-22 are monomers. IL-10 is primarily produced by activated monocytes and T cells to suppress inflammatory and immune responses.

IL-10 members belong to the class II cytokine family and bind class II cytokine receptors that lack a WSXWS motif in the extracellular domain. IL-10 receptors exist as heterodimers, comprised of a R1 and an R2 chain, which are shared by IL-10 family members. The receptor for IL-10 is comprised of IL-10R1, which is specific for IL-10, and IL-10R2, which is shared by IL-22, IL-26, IL-28 α , IL-28 β , and IL-29 (Figure 3). The type I IL-20 receptor, which is comprised of IL-20R1 and IL-20R2, binds IL-19, IL-20, and IL-24, whereas the type II IL-20 receptor, which is comprised of IL-22R1 and IL-20R2, binds IL-20 and IL-24. The IL-22 receptor is comprised of IL-22R1 and IL-10R2, while IL-26 binds to a heterodimeric receptor comprised of IL-20R1 and IL-10R2. In addition, there is a naturally occurring soluble receptor, IL-22 binding protein (IL-22BP), which binds and neutralizes the biological activity of IL-22.

IL-10 receptor family members do not have intrinsic receptor tyrosine kinase activity, but instead signal via the JAK-STAT pathway, with Jak1 binding to IL-10R1 and IL-22R1, while Tyk2 binds IL-10R2. This leads to the activation of STAT3, whereas STAT1 and STAT5 may also be activated in specific cell types. IL-10 family members can also activate MAPK signaling pathways.

IL-17 Family

The IL-17 family is a recently identified family of cytokines that plays important roles in inflammation and immune responses. IL-17A, the founding member of the IL-17 family, is produced primarily by T helper 17 (Th17) cells, which are a subset of CD4⁺ T cells that regulate allergic disorders, such as asthma, and autoimmune disorders, such as rheumatoid arthritis, multiple sclerosis, psoriasis, and inflammatory bowel disease.

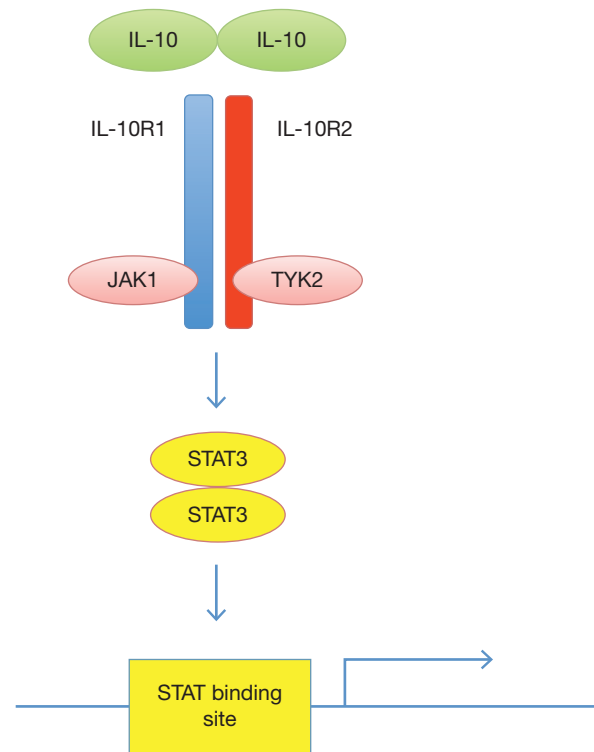


Figure 3 Interleukin 10 signaling.

IL-17A also participates in host defense against bacterial and fungal infections, as well as in granulopoiesis. $\gamma\delta$ T cells also express IL-17A. IL-17A and IL-17 F exist either as homodimers or as heterodimers. Four additional IL-17 family members have been identified, IL-17B, IL-17 C, IL-17D, and IL-17E (also known as IL-25). IL-17 F is primarily involved in mucosal host defense, whereas IL-17E (IL-25) amplifies immune responses by Th2 cells.

The five members of the IL-17 receptor family include IL-17RA, IL-17RB, IL-17RC, IL-17RD, and IL-17RE. The IL-17 receptor is a heterodimer comprised of IL-17RA and IL-17RC that bind homodimers of IL-17A or IL-17 F, as well as IL-17A/IL-17 F heterodimers (Figure 4). Biological effects of IL-17A and IL-17 F may be regulated at the level of receptor binding, as IL-17A has higher affinity for IL-17RA, while IL-17 F has a higher affinity for IL-17RC. Furthermore, cellular patterns of expression differ for these receptors, as IL-17RA is primarily expressed by immune cells, whereas IL-17RC is primarily expressed by nonimmune cells. Homodimeric IL-17E (IL-25) signals via the IL-25 receptor, which is comprised of IL-17RA and IL-17RB. IL-17RB and IL-17RE are the receptors for IL-17B and IL-17C, respectively. The receptor for IL-17D has not been identified nor has the ligand for IL-17RD.

The IL-17 receptor signals via the recruitment of the adaptor protein Act1 (TRAF3IP3 and CIKS), which directly associates with IL-17RA and IL-17RC via a SEFIR (SEF/IL-17R) domain interaction. This recruits TRAF6 and TAK1, leading to the activation of NF- κ B, as well as MAPKs and transcription factors, AP-1 and C/EBP. TRAF3 can also associate with Act1 and negatively regulate IL-17A-mediated activation of NF- κ B and MAPK. Binding of IL-17E (IL-25) to the IL-25R, which is a

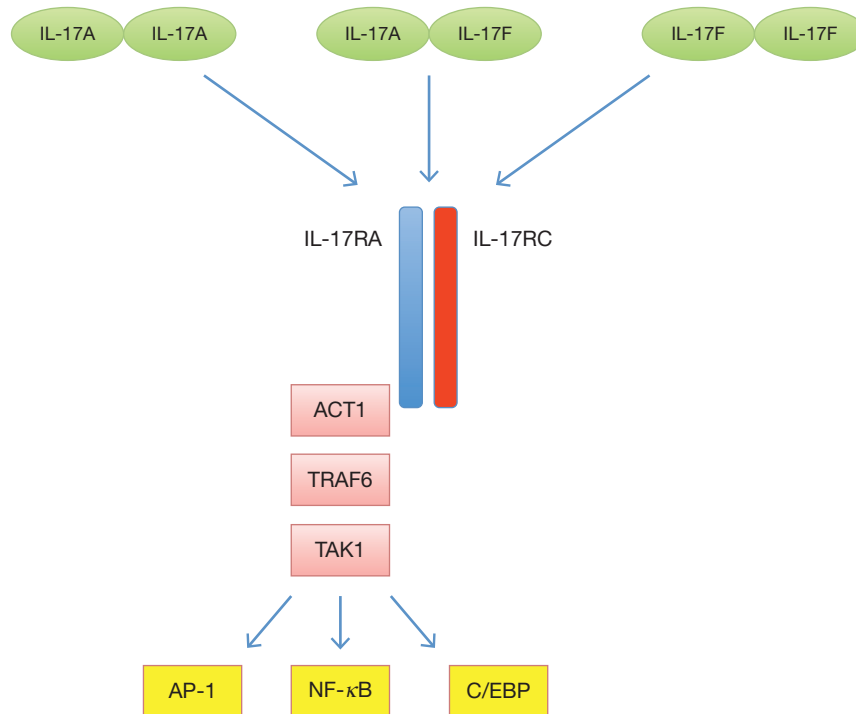


Figure 4 Interleukin 17 signaling.

heterodimer of IL-17RA and IL-17RB, also signals via the recruitment of Act1, TRAF6, and TAK1.

Type I Cytokines

Type I cytokines, which have a three-dimensional structure comprised of four α -helical bundles, can be divided into subgroups characterized by short chains of approximately 15 amino acids, or long chains of 25 amino acids. Short-chain type I cytokines can be further subdivided based upon whether they share common receptor subunits of multimeric type I cytokine receptors. For example, type I cytokines whose receptors share the common cytokine receptor γ -chain (γ_c) include, IL-2, IL-4, IL-7, IL-9, IL-13, IL-15, IL-21, and thymic stromal lymphopoietin (TSLP), whereas those that share the common cytokine receptor β -chain (β_c) include IL-3, IL-5, and granulocyte macrophage colony-stimulating factor (GM-CSF). Receptors for type I cytokine family members, as well as type II cytokines, do not possess intrinsic tyrosine kinase activity, but instead signal via the JAK/STAT pathway. Additional short-chain type I cytokines include macrophage-colony stimulating factor (M-CSF) and stem cell factor (SCF). Cytokines that share the common γ_c receptor activate JAK1, JAK2, JAK3, and TYK2, which leads to the downstream activation of STAT3, STAT5A, STAT5B, or STAT6 (Figure 5). By contrast, cytokines that share the common β_c receptor signal via JAK2 and STAT5A and STAT5B. Long-chain type I cytokines are typified by the IL-6 family that signal via receptors that share gp130. Additional long-chain type I cytokines include IL-12, growth hormone, prolactin, erythropoietin, thrombopoietin, leptin, and granulocyte colony-stimulating factor (G-CSF).

The importance of class I cytokines is evidenced by the role of the common cytokine receptor γ -chain (γ_c) in T-cell development. X-linked severe combined immunodeficiency (SCID) occurs in individuals with mutations in the γ_c chain that results in the absence of T cells and natural killer (NKT) cells, as well as nonfunctional B cells. Patients with SCID typically die in their first year if they do not undergo bone marrow transplantation. An autosomal mutation in JAK3 can also cause SCID as a result of defective signaling via the γ_c receptor. JAK1- and JAK2-deficient mice display perinatal or fetal lethality, which may reflect their effects in humans.

Type II Cytokines

Type II cytokines are comprised of interferons (IFN- α/β and IFN- γ) and IL-10 family members (see above). IFNs play key roles in host defense against DNA and RNA viruses, as well as tumor cells. IFNs also modulate innate and adaptive immune responses, as well as autoimmunity. Type I IFNs are comprised of 13 IFN- α genes, as well as single genes for IFN- β , IFN- ω , IFN- ϵ , and IFN- κ . Type I IFNs are nonglycosylated proteins of 165–200 amino acids which contain five α -helical bundles that are positioned by two disulphide bonds. IFN gene transcription is initiated via toll-like receptor signaling. For example, binding of double-stranded RNA to TLR3 or cytoplasmic RNA helicases (retinoic acid-inducible gene I (RIG-I) and melanoma differentiation-associated protein 5 (MDA5)) activates IFN regulatory factors, IRF3 and IRF7, as well as NF- κ B, with resultant transcription of type I IFNs. The type I IFNs then bind a heterodimeric receptor comprised of IFN receptor 1 (IFNAR1) and IFN receptor 2 (IFNAR2), which are associated with TYK2 and JAK1 (Figure 6). This induces the

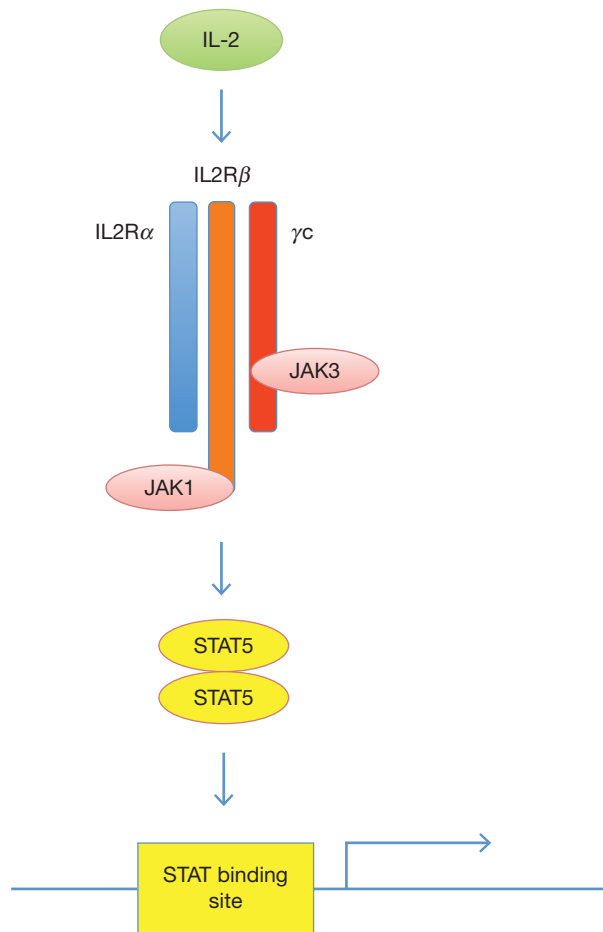


Figure 5 Interleukin 2 signaling.

phosphorylation of STAT1 and STAT2 to form a heterotrimeric complex with IRF9 that binds IFN-stimulated response elements (ISREs) in the promoters of IFN- α -activated genes.

IFN- γ is a type II IFN that is expressed by NKT cells and activated T cells. The IFN- γ protein is a single glycosylated protein of 140 amino acids that exists as an antiparallel homodimer that binds to a receptor complex comprised of the IFN- γ receptor 1 (IFNGR1) and IFN- γ receptor 2 (IFNGR2). IFN- γ binds to two IFNGR1 receptor chains, which are stabilized by two IFNGR2 chains (Figure 7). IFNGR1 binds JAK1, whereas IFNGR2 binds JAK2, which leads to the activation of a STAT1 homodimer that binds to IFN- γ -activated sites (GAS) to induce the transcriptional activation of IFN- γ -activated genes.

Type III IFNs (IFN- λ) are members of the IL-10 cytokine family (see above) and signal via a heterodimeric receptor comprised of the IL10R2 (IL-10RB) and IFN- λ receptor 1 (IFNLR1 or IL28RA), which activates the same pathway as type I IFNs.

Tumor Necrosis Factor Family

The tumor necrosis factor (TNF) family is a large family of at least 17 members that play key roles in inflammation and immunity, as well as a broad range of diverse biological functions, including host defense, neoplasia, tissue modeling and

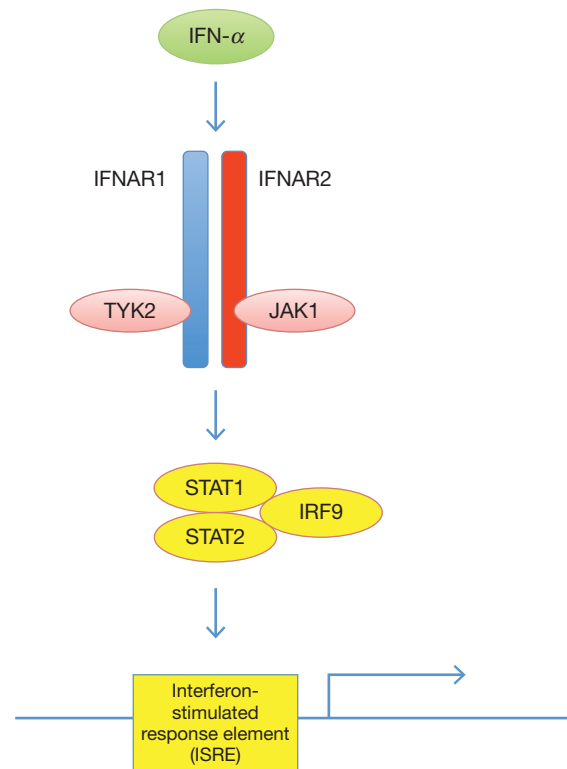


Figure 6 Interferon- α signaling.

remodeling, and neuronal development. TNF ligands exist as type 2 membrane proteins that have both a membrane-associated pro-form, as well as a cleaved, soluble mature form. The TNF receptor family is similarly large, with at least 27 members that exist as type I membrane proteins.

TNF, the prototypical member of the TNF superfamily, was originally recognized as the immune cell product that induces tumor cell cytotoxicity and necrosis. TNF was cloned in 1984 and found to mediate pleiotropic biologic functions that include tumor cell apoptosis, inflammation, immune cell function, host defense, autoimmunity, and organ development. TNF exists as a stable homotrimer of type II transmembrane proteins that undergoes proteolytic cleavage by ADAM17 (TNF- α converting enzyme (TACE)) to release a soluble form comprised of three 17-kDa protomers. TNF binds to two receptors, the type 1 TNF receptor (TNFR1 and CD120a) and the type 2 TNF receptor (TNFR2 and CD120b). TNFR1 and TNFR2 exist as homotrimers that are pre-assembled via preligand binding assembly domains.

TNFR1 can activate signaling pathways that have opposite outcomes and promote either inflammation or apoptosis. The pro-inflammatory/anti-apoptotic pathway is mediated via recruitment of signaling complex I, which consists of TNF receptor-associated protein with a death domain (TRADD), receptor interacting protein (RIP), cellular inhibitor of apoptosis proteins 1 and 2 (cIAP1/2), and TNF receptor-associated factor 2 (TRAF2) (Figure 8). This occurs following TNF-induced translocation of TNFR1 to lipid rafts and is associated with the ubiquitination of TNFR1, TRAF2, and RIP1. This leads to the recruitment of the I κ B kinase

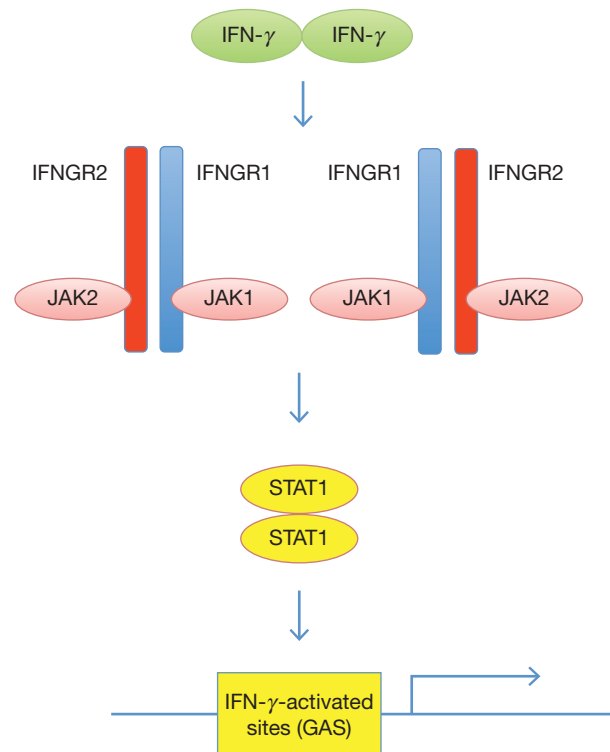


Figure 7 Interferon- γ signaling.

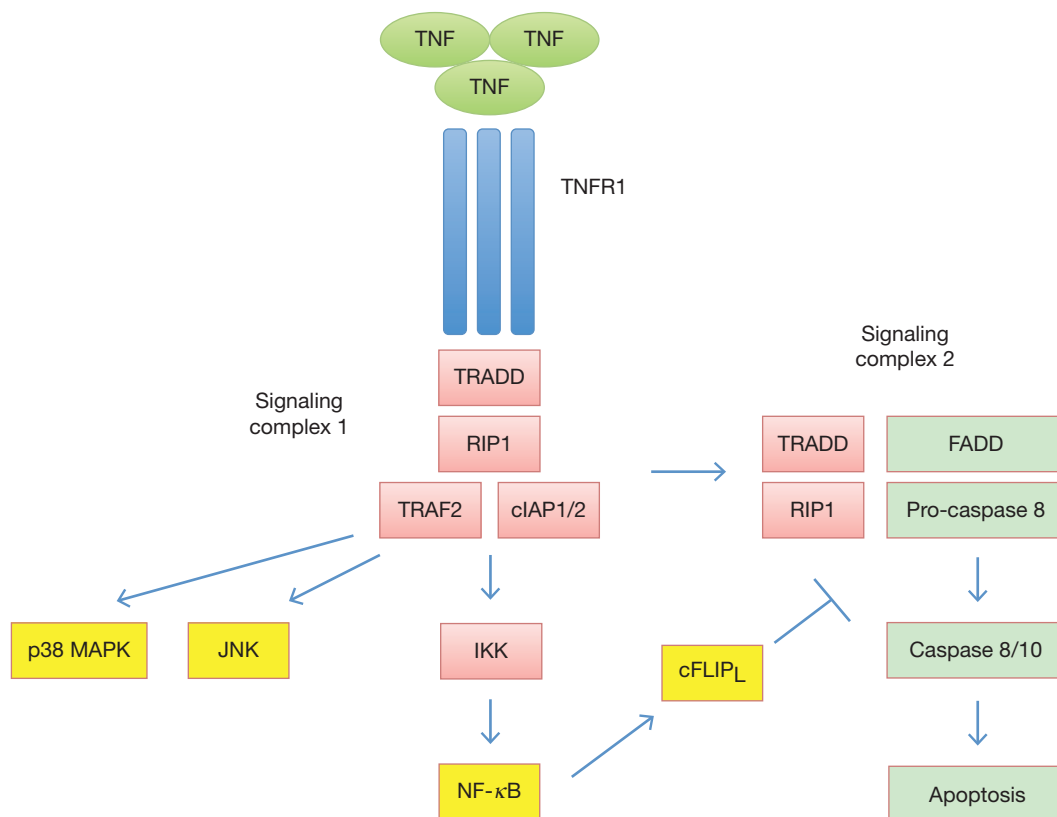


Figure 8 Tumor necrosis factor (TNF) signaling.

(IKK) complex and the IKK-activating complex, comprised of TAK1, TAB2, and TAB3, with the resultant activation of NF- κ B, which has pro-inflammatory and anti-apoptotic effects. Signaling complex 1 can also activate p38 MAPK and JNK kinases. Activation of pro-apoptotic pathways occurs when RIP and TRADD dissociate from the TNFR1 complex at the plasma membrane and recruit Fas-activated via death domain (FADD), TNFRSF6) and pro-caspase 8 to form signaling complex 2. When NF- κ B is activated by signaling complex 1, the caspase-8 inhibitor, cellular FLICE inhibitory protein long (cFLIP_L) is recruited to signaling complex 2, leading to cell survival. In the absence of NF- κ B activation, pro-caspase 10 is recruited to signaling complex 2, which results in programmed cell death.

Several lines of evidence support the key role of TNF in the regulation of inflammation. First, TNF receptor-associated periodic syndrome (TRAPS) is an auto-inflammatory disorder characterized by recurrent episodes of fever, myalgia, rash, abdominal pain, and conjunctivitis that is caused by missense mutations in TNFR1 (TNFRSF1A). The TRAP-associated TNFR1 mutations lead to the intracellular accumulation of mutant TNFR1 molecules that do not signal, but instead display enhanced secretion of pro-inflammatory cytokines in response to lipopolysaccharide. Second, TNF inhibitors have entered clinical use for the treatment of inflammatory arthritides, such as rheumatoid arthritis. Third, TNF blockade has been associated with opportunistic infections, such as tuberculosis, which illustrates the important role of TNF in host defense against bacterial, fungal, parasitic, and mycobacterial infections.

Chemokines

Chemokines are chemotactic cytokines that mediate the tissue localization of immune, inflammatory, vascular, and cancer cells. Chemokines constitute the largest cytokine family in humans, with 44 different chemokine genes. Of note, the largest number of chemokine genes is present in the zebrafish genome,

which contains more than 100. Chemokine ligands, which have four conserved cysteines, are grouped into five subfamilies based upon the location of the two amino-terminal cysteines. The conserved cysteines form two intra-chain disulfide bonds that link cysteine 1 to 3 and 2 to 4. The CC subfamily has adjacent cysteines, whereas the CXC and CX3C subfamilies have one or three amino acids separating the first and second cysteines. The XC or C subfamily has a single cysteine at the amino terminus, whereas the first two cysteines are omitted in the XC subfamily. Chemokines can act on a variety of cell types via binding to multiple chemokine receptors.

Chemokine receptors are members of the seven transmembrane domain receptor families that signal via heterotrimeric G proteins (Figure 9). Humans have 22 chemokine receptors that can be divided into subfamilies that reflect the structural classification of the chemokine ligands. These include six CXC chemokine receptors (CXCR1 through CXCR6), 10 CC chemokine receptors (CCR1 through CCR10), one XC chemokine receptor (XCR1), and one CX3C chemokine receptor (CX3CR1). Binding of chemokine ligands are restricted to chemokine receptors within their subfamily. However, chemokines may bind multiple receptors within their chemokine receptor subfamily, which facilitates the recruitment of distinct populations of functionally related cells. For example, CXC chemokines primarily mediate the chemotaxis of polymorphonuclear neutrophils, whereas CC receptors primarily recruit subsets of T cells, eosinophils, basophils, dendritic cells, and monocytes. Binding of CX3CL1 to CX3CR1 recruits monocytes, immature dendritic cells and T-cell subsets, whereas binding of XCL1 or XCL2 to XCR1 mediates the recruitment of T cells and NKT cells.

TGF- β Family

The transforming growth factor- β (TGF- β) family of cytokines is comprised of 33 structurally related proteins in humans that

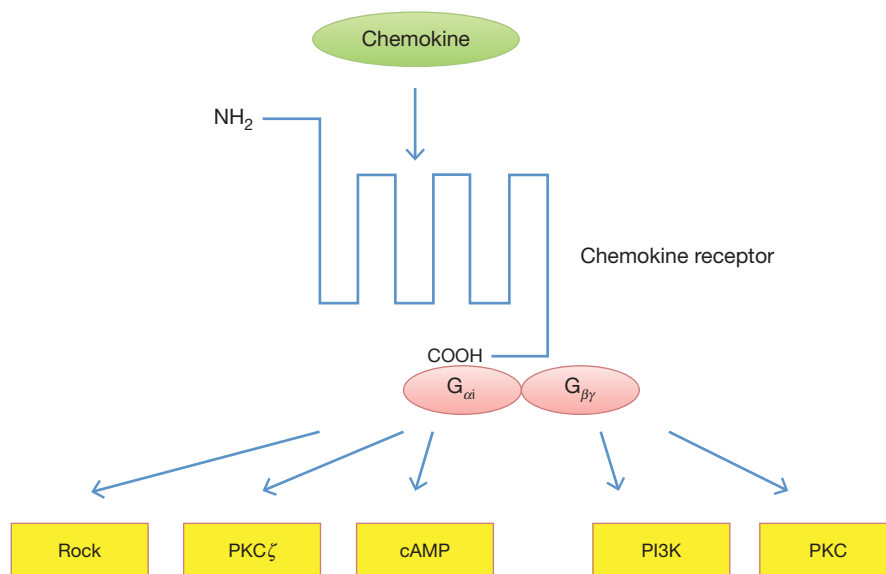


Figure 9 Chemokine signaling.

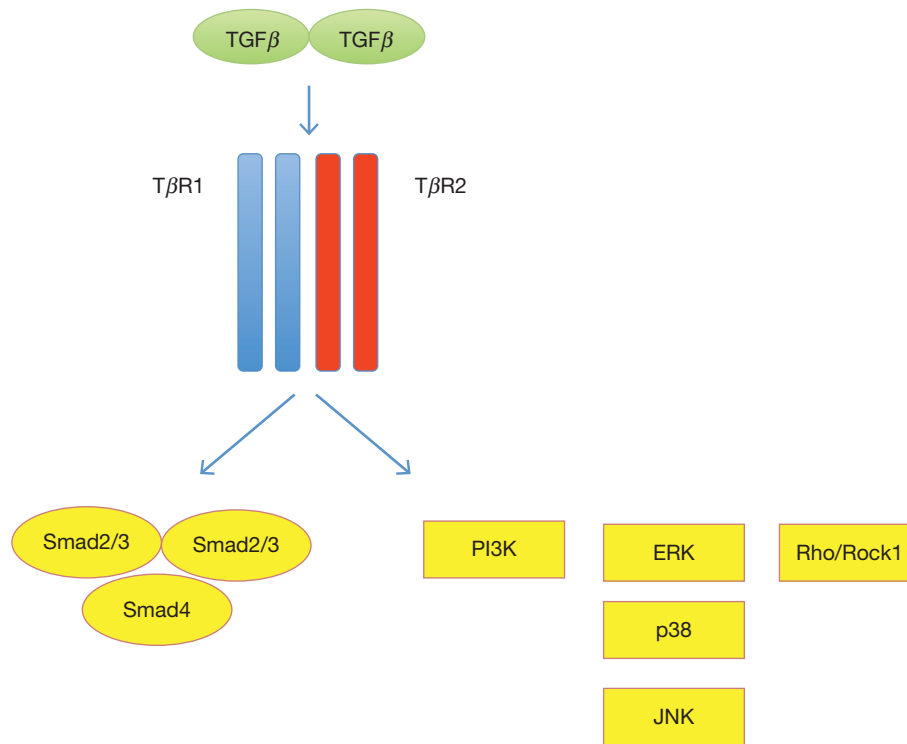


Figure 10 Transforming growth factor- β signaling.

regulate development, homeostasis, and disease. TGF- β is the prototypical member of this family and plays important roles in regulating cellular proliferation, differentiation, apoptosis, and adhesion. TGF- β also regulates cellular microenvironment and functions as a tumor suppressor. The TGF- β family can be divided into two branches. One branch is comprised of TGF- β isoforms (TGF- β 1, - β 2, and - β 3), activins (activin-A, activin-B, and activin-C), inhibin, lefty proteins, nodal, and myostatin. The other branch is comprised of bone morphogenetic proteins (BMPs), growth and differentiation factors (GDFs), and anti-muellerian hormone (AMH or MIS).

TGF- β is expressed as a precursor protein that contains a pro-peptide which is cleaved by furin-like proteases in the secretory pathway. The pro-peptide or latency-associated peptide (LAP) remains complexed with the mature peptide and functions as a chaperone during exocytosis, as well as facilitates the deposition of TGF- β in the extracellular matrix via an association with extracellular matrix proteins and TGF- β -binding proteins, which are large secreted proteins. The active TGF- β dimer signals via heterotetrameric receptors comprised of type I (also known as 'activin receptor-like kinases' or 'ALKs') and type II receptors that possess intrinsic serine/threonine kinase activity, as well as weaker tyrosine kinase activity (Figure 10). Binding of TGF- β stimulates the phosphorylation of the T β RI type I receptor by the T β RII type II receptor, which then signals via Smad transcription factors. TGF- β branch receptors signal via phosphorylation of Smads 2 and 3, while BMP branch receptors signal via phosphorylation of Smads 1, 5, and 8. Phosphorylation of two receptor-activated Smads (RSmads), such as Smad2 or Smad3, recruits a single common-mediator Smad, such as Smad4, to form a trimeric

complex that translocates to the nucleus and binds chromatin, as well as interacts with additional transcription factors, to regulate the transcription of target genes. TGF- β receptors can also activate non-Smad signaling pathways, such as MAP kinases, PI3K, and Rho-ROCK1.

BMPs play important roles in bone development, as well as bone remodeling and repair. This has led to the development of BMPs for the treatment of orthopedic surgery applications. Both recombinant human BMP-2 and BMP-7 have been approved by the US Food and Drug Administration for clinical use to repair bone injuries by promoting fracture healing or spinal fusion.

Acknowledgment

This work was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute.

See also: [Cell Architecture and Function: Septins and Cytokinesis](#); [Metabolism Vitamins and Hormones: Insulin- and Glucagon-Secreting Cells of the Pancreas](#).

Further Reading

- Chang SH and Dong C (2010) Signaling of interleukin-17 family cytokines in immunity and inflammation. *Cell Signal*, doi:10.1016/j.cellsig.2010.11.022.
- Chen G and Goeddel DV (2002) TNFR1 signaling: A beautiful pathway. *Science* 296: 1634–1635.
- Dinarello CA (2010) IL-1: Discoveries, controversies and future directions. *European Journal of Immunology* 40: 599–606.

- Dinareello CA (2011) Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. *Blood* [online].
- Gabay C, Lamacchia C, and Palmer G (2010) IL-1 pathways in inflammation and human diseases. *Nature Reviews Rheumatology* 6: 232–241.
- George J, Mothswene PG, Wang H, et al. (2011) Two MYD88 variants, S34Y and R98C, interfere with MyD88-IRAK4-Myddosome Assembly. *Journal of Biological Chemistry* 286: 1341–1353.
- Hull KM, Drewe E, Aksentijevich I, et al. (2002) The TNF receptor-associated periodic syndrome (TRAPS): Emerging concepts of an autoinflammatory disorder. *Medicine* 81: 349–368.
- Iwakura Y, Ishigame H, Saijo S, and Nakae S (2011) Functional specialization of interleukin-17 family members. *Immunity* 34: 149–162.
- Kehrl JH (2006) Chemoattractant receptor signaling and the control of lymphocyte migration. *Immunologic Research* 34: 211–227.
- Leonard WJ (2001) Cytokines and immunodeficiency disease. *Nature Reviews Immunology* 1: 200–208.
- Leonard WJ (2008) Type I cytokines and interferons and their receptors. In Paul W (ed.) *Fundamental Immunology*, 6th edn., pp. 706–749. PA: Lippincott Williams and Wilkins.
- Locksley RM, Killeen N, and Lenardo MJ (2001) The TNF and TNF receptor superfamilies: Integrating mammalian biology. *Cell* 104: 487–501.
- Mantovani A, Savino B, Locati M, et al. (2010) The chemokine system in cancer biology and therapy. *Cytokine and Growth Factor Reviews* 21: 27–39.
- Massagué J (2008) TGF β in cancer. *Cell* 134: 215–230.
- Micheau O and Tschopp J (2003) Induction of TNF receptor I-mediated apoptosis via two sequential signaling complexes. *Cell* 114: 181–190.
- Moustakas A and Heldin CH (2009) The regulation of TGF β signal transduction. *Development* 136: 3699–3714.
- Neurath MF and Finotto S (2011) IL-6 signaling in autoimmunity, chronic inflammation, and inflammation-based cancer. *Cytokine and Growth Factor Reviews* 22: 83–89.
- Nomiyama H, Osada N, and Yoshie O (2010) The evolution of mammalian chemokine genes. *Cytokine and Growth Factor Reviews* 21: 253–262.
- Ouyang W, Kolls JA, and Zheng Y (2008) The biological functions of T helper 17 cell effector cytokines in inflammation. *Immunity* 28: 454–467.
- Rose-John S, Scheller J, Elson G, and Jones SA (2006) Interleukin-6 biology is coordinated by membrane-bound and soluble receptors: Role in inflammation and cancer. *Journal of Leukocyte Biology* 80: 227–236.
- Sabat R (2010) IL-10 family of cytokines. *Cytokine and Growth Factor Reviews* 21: 315–324.
- Scheller J, Chalaris A, Schmidt-Arras D, and Rose-John S (2011) The pro- and anti-inflammatory properties of the cytokine interleukin-6. *Biochimica et Biophysica Acta*, doi:10.1016/j.bbamcr.2011.01.034.
- Schroder K and Tschopp J (2010) The Inflammasomes. *Cell* 140: 821–832.
- Senta H, Park H, Bergeron E, et al. (2009) Cell responses to bone morphogenetic proteins and peptides derived from them: Biomedical applications and limitations. *Cytokine and Growth Factor Reviews* 20: 213–222.
- Simon A, Park H, Maddipati R, et al. (2010) Concerted action of wild-type and mutant TNF receptors enhances inflammation in TNF receptor 1-associated periodic fever syndrome. *Proceedings of the National Academy of Sciences of the United States of America* 107: 9801–9806.
- Sims JE and Smith DE (2010) The IL-1 family: Regulators of immunity. *Nature Reviews Immunology* 10: 89–102.
- Theofilopoulos AN, Baccala R, Beutler B, and Kono DH (2005) Type I interferons (α/β) in immunity and autoimmunity. *Annual Reviews in Immunology* 23: 307–336.
- Vilcek J and Lee TH (1991) Tumor necrosis factor: New insights into the molecular mechanisms of its multiple actions. *Journal of Biological Chemistry* 266: 7313–7316.
- Wajant H, Pfizenmaier K, and Scheurich P (2003) Tumor necrosis factor signaling. *Cell Death and Differentiation* 10: 45–65.
- Wajant H and Scheurich (2011) TNFR1-induced activation of the classical NF- κ B pathway. *FEBS Journal* 278: 862–876.
- Wallach D and Kovalenko A (2009) 12th international TNF conference: The good, the bad, and the scientists. *Cytokine and Growth Factor Reviews* 20: 259–269.
- Wertz IE and Dixit VM (2008) Ubiquitin-mediated regulation of TNFR1 signaling. *Cytokine and Growth Factor Reviews* 19: 313–324.
- Winthrop KL and Chiller T (2009) Preventing and treating biologic-associated opportunistic infections. *Nature Reviews Rheumatology* 5: 405–410.
- Zdanov A (2010) Structural analysis of cytokines comprising the IL-10 family. *Cytokine and Growth Factor Reviews* 21: 325–330.

Cytokinesis

V Srivastava and D N Robinson, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Abscission The final severing of the intercellular bridge, which is formed through ingression of the cleavage furrow.

Asters The radial array of microtubules originating from each spindle pole that extends toward the cell cortex.

Cell cortex The network of actin filaments and other proteins in close proximity to the cell membrane.

Central spindle The microtubule network extending between the two microtubule organizing centers of a dividing cell.

Cleavage furrow The region of cortex of a dividing cell where the cell constricts to form the intercellular bridge.

Contractile ring A dense array of actin and myosin II arranged in a ring-like structure in the cleavage furrow of many dividing cells.

Cortical tension The energy cost per unit increase in the cell surface area.

Cytokinesis The process of physical separation of a dividing cell into daughter cells.

Intercellular bridge The narrow, cylindrical connection between two daughter cells.

Laplace pressure The pressure generated at a curved fluid surface due to the surface tension of the fluid.

Midbody A compact protein-rich structure that contains a dense anti-parallel microtubule array, which is found in the intercellular bridge during late cytokinesis.

Cytokinesis

Subsequent to chromosome segregation during mitosis, the cell cytoplasm and other organelles are partitioned into two daughter cells through the process of cytokinesis. Correct cytokinesis is relevant for both normal development and disease. The uniform partitioning of cellular material is critical for normal cell proliferation, while asymmetric cell division is important in processes such as stem cell maintenance. In addition, cytokinesis defects have been implicated in many diseases, including cancer. Thus, detailed studies aimed at identifying the mechanisms that control cytokinesis are necessary to understand cellular behavior and for therapeutic applications.

In order to understand the mechanisms that regulate cytokinesis, it is essential to answer the following questions:

1. When and where does the division occur?
2. What are the factors responsible for this division?
3. How do those factors interact with each other to help the cell divide?

Like many biological processes, cytokinesis is a complex phenomenon involving a large number of players that interact with each other through multiple overlapping biochemical and mechanical pathways. Thus, cytokinesis is a great example of a biological control system, where the various feedback loops act in tandem to ensure the fidelity and robustness of cell division. Although the exact biochemical interactions are still unresolved to a large extent, a great deal is now known about the underlying mechanisms.

In spite of large cell-specific variations in the details of cytokinesis, there are certain universal characteristics of the process. Typically, cytokinesis proceeds through a series of stereotypical cell shape changes where the cell first rounds up, then elongates forming a cleavage furrow near the middle, whose constriction finally results in the separation of the daughter cells (Figure 1). These cell shape changes imply that cytokinesis is fundamentally

a mechanical process that requires major reorganization of the cytoskeleton and associated proteins to promote cellular contractility at the cleavage site. Hence, it is essential to supplement biochemical and genetic information with biophysical and mechanical studies to understand force generation, sensing, and transduction in a dividing cell.

Proteins That Regulate Cytokinesis

Methods

A variety of genetic, biochemical, and biophysical techniques have been used to identify and characterize cytokinesis genes. A traditional genetic approach involves phenotypic screening of mutant cell lines for cytokinesis defects, such as an increase in the number of bi- or multi-nucleated cells. Partial knockdown of protein expression using RNA interference (RNAi) has allowed functional analysis of essential genes, whose deletion can be lethal. By contrast, overexpression of genes has helped identify proteins that possess dominant-negative activities. Finally, live cell imaging of fluorescently labeled proteins gives information about their localization and dynamics *in vivo*. These techniques provide a powerful toolset for the functional analysis of protein dynamics, and have been combined with high-throughput methods for genome-wide screening to identify genes that regulate cytokinesis. In addition, directed screens and genetic suppression have helped map out biochemical pathways. However, to uncover entire pathways, these studies need to be supplemented with detailed characterization of protein interactions and *in vitro* biochemical reconstitution. Additionally, the use of pharmacological agents that modify the activity of specific proteins has allowed real-time phenotypic manipulation. This approach is especially useful for studying elements that are essential for cell survival or those that are required at multiple stages during the cell cycle. For example, cytokinesis defects were observed in post-mitotic cells treated with the microtubule-destabilizing compound, nocodazole. This established the involvement of

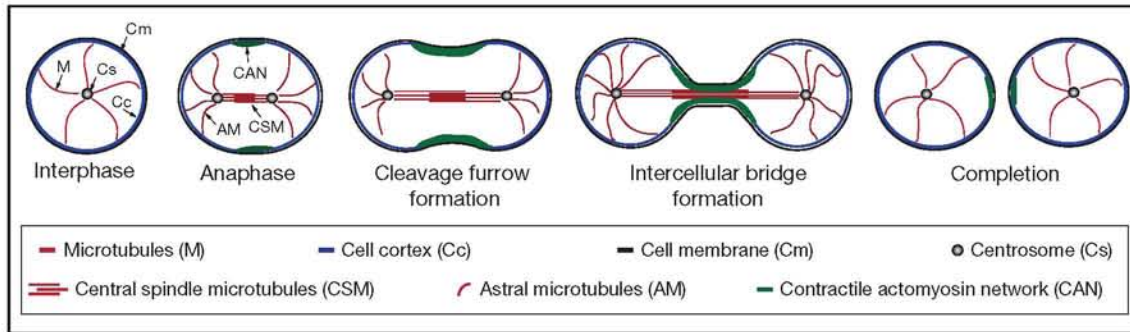


Figure 1 Stages of cytokinesis: a schematic diagram depicting the progression of cytokinesis in an amoeboid cell. Note that the contractile actomyosin network (CAN) is found throughout the cell cortex, but is enriched in the cleavage furrow.

microtubules in cytokinesis completion, apart from their role in chromosome segregation and cleavage site selection. Using a combination of various genetic and biochemical studies, a partial parts list of cytokinesis proteins has been compiled for many model organisms. Many of these are cytoskeleton-related proteins and kinases, while a large number of genes still remain uncharacterized.

Proteins

In most eukaryotic cell types, actin and the motor protein myosin II are known to form a contractile structure in the equatorial region of a dividing cell, whose ingression drives cytokinesis. Actin controls the cell mechanics by forming a highly dynamic network of semi-flexible filaments (Figure 2). During the myosin II power stroke, myosin pulls on the actin filaments as it releases the adenosine triphosphate (ATP) hydrolysis products, thereby generating mechanical force. Two myosin heavy chains combine with two essential and two regulatory light chains (ELC and RLC, respectively) to form a myosin II hexameric monomer, which then assembles into functional bipolar thick filaments (BTFs). Rho-kinase (ROCK) directs BTF assembly by activating myosin II through the phosphorylation of RLC. Additionally, heavy chain phosphorylation also controls BTF assembly and disassembly, both of which are required for normal cytokinesis. Although myosin II is the major mechanoenzyme during cytokinesis, it is not necessary for cytokinesis. Adherent cells can divide fairly normally in the absence of myosin II using traction forces to help with the initial cell elongation followed by cortical tension-driven furrow thinning. The effects of cortical tension, which serves to minimize the surface area-to-volume ratio, are highly reminiscent of the surface tension of a liquid droplet, which also helps drive droplet breakup.

In addition, several other actin-binding proteins, such as anillin and α -actinin in animals and cortexillin in *Dictyostelium*, help form a cross-linked actin network in the cell cortex (Figure 2), thereby regulating mechanical properties of the cortex. These proteins differ in their structure, actin-binding kinetics, force sensitivity, and cellular location. Collectively, these promote load-dependent force generation by myosin II and thereby cellular contractility. Some cross-linkers also contain lipid-binding domains (such as the pleckstrin-homology (PH) domain) that facilitate membrane attachment of the actin network. In animal cells, anillin is a potential scaffolding

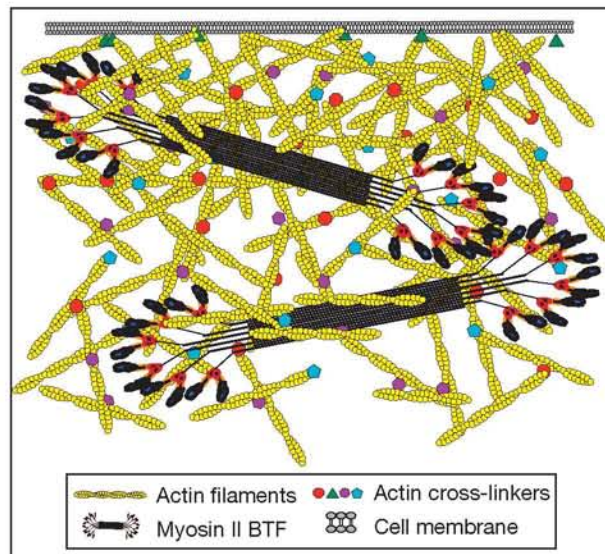


Figure 2 A stylized diagram showing major components of the actin cytoskeleton in the cell cortex. Actin filaments form a dynamic highly cross-linked network that is connected through various actin cross-linking proteins. These proteins can regulate the tension in the network, thereby controlling the contractility of the cortex. The actin network is also attached to the cell membrane through some actin-binding proteins. The mechanoenzyme myosin II forms bipolar thick filaments and generates mechanical stress within the network by pulling on actin filaments.

protein that may provide membrane anchoring and link Rho, actin, and myosin II in the furrow. Rho is a major regulator of animal cytokinesis, as it controls both actin polymerization through formins and myosin II activation through ROCK. The levels of guanosine triphosphate (GTP)-bound active Rho in the cleavage furrow are controlled by the guanine-exchange factor (GEF) ECT2 and the GTPase-activating protein (GAP) MgcRacGAP.

Another major player in cytokinesis is the mitotic spindle, which helps define the axis of cell division and is involved in many signaling pathways. The spindle can deliver signals to the cell cortex that modulate cortical mechanics and direct cleavage-site selection and actomyosin contractile structure formation. This signal can either be in the form of a biochemical factor or be in the form of a purely mechanical cue like a change in

cortical tension or membrane potential. Many microtubule-based proteins such as the kinesin-6 family of proteins (mitotic kinesin-like protein-1 (MKLP-1)) are known to be important in cytokinesis and may localize differentially in a dividing cell. These proteins are believed to promote communication between the central spindle and the cell cortex.

Even though the membrane is thought to have a relatively limited contribution to the cellular mechanical properties, proteins involved in membrane dynamics, membrane fission and fusion, and vesicle transport are important in cytokinesis. The surface area of a dividing cell increases significantly as the furrow constricts. This requires the deposition of new membrane in the furrow region. In addition, constant membrane remodeling is required to relieve mechanical stress.

As the search for new genes that regulate cytokinesis continues, many nonprotein factors, including lipids and small metabolites, are also being examined for their role in cytokinesis. For example, the phosphoinositol-4,5-bisphosphate (PIP₂) is enriched in the cleavage furrow and can control the accumulation and retention of PIP₂-binding proteins during cytokinesis. In addition, many mechanical parameters, such as those described in the section Mechanics of cytokinesis, can affect the kinetics of cell division. To ensure robustness of cytokinesis, current models support the existence of multiple interacting, as well as parallel, mechanisms, thereby making the compilation of a comprehensive cytokinesis parts list challenging.

Spatiotemporal Events during Cytokinesis

Cytokinesis is tightly coupled to the cell cycle to ensure the proper segregation of genetic material into daughter cells. The onset of anaphase triggers major restructuring of the cytoskeleton, which initiates cytokinesis. No obvious factors for controlling the timing of cytokinesis initiation have been identified yet, but it is thought that some of the factors involved in the mitotic phase also regulate cytokinesis initiation. For example, Cdc2/Cdk1 inactivation during early anaphase or anaphase-promoting complex (APC)-dependent proteolysis could help initiate cytokinesis. The mitotic-exit network (MEN) and septation-initiation network (SIN) are other pathways that are known to regulate cytokinesis in some systems, and are relatively conserved across yeast, fungi, and plants.

Along with the temporal control of cytokinesis, the process is also regulated spatially. In most cells, cytokinesis occurs in the plane perpendicular to the cellular long axis generally close to the center of the cell. Thus, even from a purely geometrical perspective, the mitotic spindle is likely to be important in symmetry breaking and cleavage-site selection. Both the central spindle and astral microtubules are required for the correct positioning of the furrow, and likely deliver chemical signals to the cell cortex. The microtubule-dependent symmetry breaking can be achieved by the following mechanisms: motor-based transport of molecules, tracking of differential microtubule density in the cortex, or signals from the plus ends of microtubules themselves. Two classical models have been proposed to explain the role of spindle and the nature of these signals. The polar relaxation model proposes that signals inhibiting contractility are delivered to the polar cortex by the asters, while the equatorial simulation model argues for positive cues being delivered to the equatorial cortex by either the

central spindle or astral microtubules. Recent evidence suggests that both mechanisms most likely act synergistically to ensure fidelity of furrow initiation, though the exact signaling molecules involved remain unknown.

Once the cleavage furrow position is established, the actin cytoskeleton must reorganize to form the contractile actomyosin network. In animals, the activation of Rho GTP-binding proteins in the furrow region stimulates actin polymerization and myosin activation. For the enrichment of active Rho in the furrow, its diffusion must be restricted. One hypothesis argues for the role of anillin as a scaffold protein, as it can specifically bind to PIP₂ as well as activated Rho. Another theory supports microtubule-dependent accumulation of equatorial proteins.

The organization of actin and myosin II filaments varies with cell type. Actin and myosin II form a well-defined contractile ring-like structure (Figure 3(a)) in a number of cell types including the fission yeast *Schizosaccharomyces pombe*, while in many cells such as fibroblasts and *Dictyostelium*, actin forms a contractile meshwork (Figure 3(b)). The length and orientation of individual filaments are also cell-type dependent, ranging from ~1 µm long, roughly concentric filaments in *S. pombe*, to a more disorganized network of short ~100 nm actin filaments in *Dictyostelium*. Myosin II is a force-sensitive mechanoenzyme whose actin-binding lifetime and force generation depend on the tension in the actin filament. Thus, by forming junctions in the actin network and tethering it to the plasma membrane, actin cross-linkers help generate tension in the filaments, thereby increasing the overall myosin II-dependent contractility of the network, which drives constriction of the furrow cortex. The dynamic properties of these cross-linkers regulate the kinetics and mechanics of furrow ingression. Actin cross-linkers localizing in the cleavage furrow generally promote ingression, while other global cross-linkers inhibit contractility. After the initial accumulation, the furrow concentration of actin and myosin II remains largely unchanged as the furrow volume decreases, highlighting the importance of cytoskeletal disassembly during contraction. In summary, while actin and myosin II together form the active contractile force generation machinery responsible for furrow ingression, other actin-binding proteins act as regulators of cellular contractility by assisting or inhibiting cleavage furrow constriction.

As furrow ingression proceeds, a thin intercellular bridge is formed connecting the daughter cells. During later stages, the midzone microtubules, which are the microtubules between segregated chromosomes, condense to form a dense protein-rich structure called the 'midbody' within the bridge. The midbody is comprised of condensed anti-parallel microtubules and several other spindle-associated proteins such as the components of the chromosome passenger complex (CPC; including proteins such as Aurora B kinase and inner centromere protein (INCENP)) and the centralspindlin (MKLP1 and MgcRacGAP). A dividing cell can continue to exist at the bridge stage for long periods of time (from minutes to hours, depending on cell type) before the final separation, which may enable the cell to ensure proper segregation of cellular content.

Multiple models have been proposed to explain abscission, which is the final severing of the intercellular bridge. The final push of cytoplasm from the bridge may be a largely physical process driven by Laplace pressure that favors minimization of surface area to volume ratio. A second model

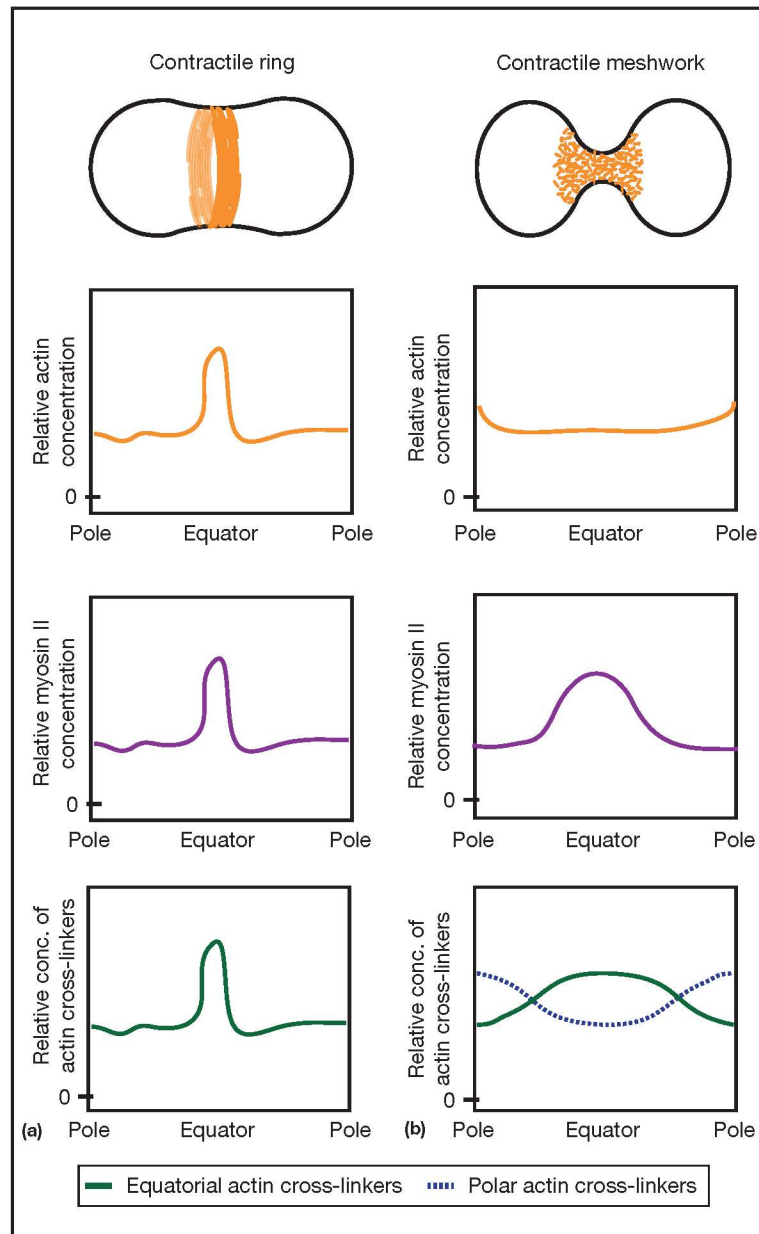


Figure 3 A comparison of actin cytoskeleton structure in cells forming a contractile ring (a) or a contractile meshwork (b). Note that the actin structures are shown to emphasize the cleavage furrow contractile organization, and the global actin network is not shown. The relative concentrations of actin, myosin II, and actin cross-linkers in the polar or equatorial cortex are also shown for both structures.

suggests that cell separation is promoted by the accumulation of secretory and endocytic vesicles adjacent to the midbody. Vesicle transport is also necessary to prevent membrane tearing as the furrow ingresses, increasing the total cellular surface area. A third model proposes that the invagination of the plasma membrane results first in hemifusion and finally cell cleavage. There is some evidence of differential membrane compositions in the equatorial versus polar cortices and in the inner versus outer membrane leaflet, which can activate signaling cascades and/or help in membrane fusion. Recently, the direct involvement of the endosomal sorting complex required for transport (ESCRT) in abscission has been shown using high-resolution imaging. However, similar to other steps

in cytokinesis, multiple mechanisms are likely to be working in parallel during abscission to ensure robustness.

Mechanics of Cytokinesis

Cytokinesis is a mechanical process during which a cell undergoes major mechanical deformation. Thus, an in-depth understanding of the effects of mechanical signals on cell mechanics and biochemistry is required. A diverse tool set is available to study cell mechanics during cytokinesis, allowing characterization of various mechanical parameters. While micropipette aspiration (MPA) studies are used to determine the elastic modulus and effective cortical tension, atomic force microscopy (AFM)

measures the bending modulus. The elastic modulus quantifies the deformability of the cell surface, and the cortical tension is a complex parameter that measures the energy cost per unit increase in cell surface area. The bending modulus reflects the stress required for bending a material. During cytokinesis, the initial deformation of a roughly spherical cell requires deviation from its quasi-steady state. This is resisted by the cortical surface tension, which favors a spherical cell. However, as the furrow continues to ingress, the curvature in the cleavage furrow changes so that the Laplace pressure eventually favors bridge thinning and abscission. Laser tracking microrheology (LTM) can be used to measure cortical viscoelasticity noninvasively. Viscoelasticity represents the time-dependent cellular response to stresses and affects the kinetics of furrow ingression by dampening the mechanical deformation, thereby allowing sufficient time for activation and stabilization of biochemical factors.

Treatment with actin depolymerizing drugs such as Latrunculin-A has established that the actin cytoskeleton is the main contributor of cell mechanics, though the cell membrane and microtubules also make some contribution. The actin cytoskeleton also undergoes remodeling with the application of internal or external mechanical stresses. The impact of mechanical stresses has been uncovered using micropipette aspiration, which allows the application of an external stress on the cell, similar in magnitude to stresses generated internally during cytokinesis. Many mechanosensitive proteins such as myosin II, which localize to the cleavage furrow cortex, also accumulate at sites where mechanical stress has been applied.

In contrast to the mechanical activation of biochemical reactions, the mechanical properties of the cell can be controlled biochemically. Knockdown of some actin cross-linkers softens the cell cortex significantly, leading to altered furrow ingression kinetics and a reduced ability to perform cytokinesis in suspension culture (where cell–substrate adhesion is absent). Interestingly, the overall deformability of the furrow is lower than the polar cortex, even though furrow undergoes major deformation during cytokinesis, which is attributed to a differential cortical distribution of mechanosensitive proteins during cytokinesis. This further illustrates the intricate interplay between biochemical and mechanical pathways during cytokinesis.

Species-Specific Cell Division Mechanisms

Though the essential sequence of events during cytokinesis is more or less conserved across organisms, there are significant differences in the mechanisms in various cell types. These become especially important in furrow positioning, where various organisms use diverse mechanisms to induce the initial symmetry breaking. Although actin and myosin II form the core contractile machinery in yeasts, protozoans, and animals, a similar system has not been found in plants. Cytokinesis in plants proceeds through a microtubule-dependent mechanism involving two additional microtubule structures, apart from the mitotic spindle, known as the ‘preprophase band (PPB)’ and ‘phragmoplast’. PPB is a ring-like microtubule and actin structure formed around the premitotic nucleus. During mitosis, actin dissociates from the PPB and the accumulation of a yet unknown factor marks the division site. The spindle then gives rise to the phragmoplast that expands toward the

cell cortex. Subsequent membrane trafficking and fusion at its edge leads to cell plate formation.

Similar to plants, the budding yeast cleavage site is determined early during mitosis by the accumulation of the small Rho GTP-binding protein Cdc42. Thus, the mitotic spindle is not crucial for budding site determination; rather the spindle aligns itself according to the polarization by Cdc42. By contrast, the cleavage furrow in fission yeast is determined by the position of the nucleus and the cylindrical geometry of the cell, as repositioning of the nucleus before early mitosis causes a shift in furrow location. Precursors of the contractile ring (or nodes) first appear near the center of the cell, which are attached to the membrane. These nodes then condense to form the contractile ring.

Cytokinesis in bacteria differs greatly from most eukaryotes as they do not contain organelles and the eukaryotic cytokinetic components. Rather, bacteria use an ancestral tubulin-related FtsZ cytoskeleton and have developed diverse intricate mechanisms for cytokinesis.

Conclusion

Cytokinesis is an essential biological process. Understanding the molecular mechanisms that regulate cytokinesis has great potential in the development of molecular targets against both cancer and infectious pathogens. It is an elegant biological control system comprised of multiple overlapping and parallel biochemical and mechanical pathways, promoting fidelity and robustness of cell division. Thus, along with the obvious therapeutic applications, cytokinesis studies are important for appreciating the complexity of biological shape morphogenesis.

See also: [Cell Architecture and Function: Actin-Related Proteins; Microtubule-Associated Proteins; Mitosis; Myosin Motors; Rho GTPases and Actin Cytoskeleton Dynamics.](#)

Further Reading

- Barr FA and Gruneberg U (2007) Cytokinesis: Placing and making the final cut. *Cell* 131: 847–860.
- Eggert US, Mitchison TJ, and Field CM (2006) Animal cytokinesis: From parts list to mechanisms. *Annual Review of Biochemistry* 75: 543–566.
- Erickson HP, Anderson DE, and Osawa M (2010) FtsZ in bacterial cytokinesis: Cytoskeleton and force generator all in one. *Microbiology and Molecular Biology Reviews* 74: 504–528.
- Glotzer M (2005) The molecular requirements for cytokinesis. *Science* 307: 1735–1739.
- Oliferenko S, Chew TG, and Balasubramanian MK (2009) Positioning cytokinesis. *Genes and Development* 23: 660–674.
- Pollard TD (2010) Mechanics of cytokinesis in eukaryotes. *Current Opinion in Cell Biology* 22: 50–56.
- Raiborg C and Stenmark H (2011) A helix for the final cut. *Science* 331: 1533–1534.
- Reichl EM, Effler JC, and Robinson DN (2005) The stress and strain of cytokinesis. *Trends in Cell Biology* 15: 200–206.
- Steigemann P and Gerlich DW (2009) Cytokinetic abscission: Cellular dynamics at the midbody. *Trends in Cell Biology* 19: 606–616.
- Surcel A, Kee YS, Luo T, and Robinson DN (2010) Cytokinesis through biochemical-mechanical feedback loops. *Seminars in Cell and Developmental Biology* 21: 866–873.

Cytokinin

T Schmülling, Free University of Berlin, Berlin, Germany

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 562–567, © 2004, Elsevier Inc.

Glossary

Cell cycle Sequence of events between mitotic divisions, divided into G1 (G standing for gap), S (synthesis phase), G2, and M (mitosis).

Cytokinin conjugate Compound formed by the union of a cytokinin and a sugar moiety.

Meristem Growing tip of roots and shoots.

Senescence Programmed aging leading to organ or plant death.

Two-component system Signal transduction system of bacteria, lower eukaryotes, and plants; involves autophosphorylation of a histidine kinase that transmits the signal via phosphorelay to response regulator proteins.

Cytokinins are plant-specific chemical messengers (hormones) that play a central role in the regulation of the plant cell cycle and numerous developmental processes. Cytokinins were discovered by F Skoog, C Miller, and co-workers during the 1950s as factors that promote cell division (cytokinesis). The first cytokinin discovered was an adenine (aminopurine) derivative named 'kinetin' (6-furfuryl-aminopurine), which was isolated as a DNA degradation product. The first common natural cytokinin identified was purified from immature maize kernels and named 'zeatin'. Several other cytokinins with related structures are known today. Cytokinins are present in all plant tissues. They are abundant in the root tip, shoot apex, and immature seeds. Their endogenous concentration is in the low nanomolar range. Typically, several types of cytokinins and their modified forms are present in a given tissue. Cytokinins can act over long distances or in the direct vicinity of the cytokinin-producing cells (paracrine signaling). Cytokinins may act also on the cell that produced them (autocrine signaling). Cytokinins are also produced by cyanobacteria, some plant pathogenic bacteria (e.g., *Agrobacterium tumefaciens*, *Pseudomonas savastanoi*, and *Rhodococcus fascians*), and the slime-mold *Dictyostelium discoideum*.

Cytokinin Structures

Naturally occurring cytokinins are adenine derivatives with a side chain at the N^6 -position (Figure 1). The structure and conformation of the N^6 -attached side chain can markedly influence the biological activity of the cytokinin. Depending on the structure of the N^6 -substituent, cytokinins are classified as isoprenoid or aromatic cytokinins. The biological activities of both classes are qualitatively similar but they may differ quantitatively in different processes. Isoprenoid cytokinins are the most abundant class. They are either isopentenyl (iP)-type cytokinins, having an isopentenyl N^6 -side chain, or zeatin-type cytokinins, having a hydroxylated isopentenyl N^6 -side chain. The side chain of a zeatin-type cytokinin occurs in either *cis* or *trans* configuration, depending on which of the two methyl groups is hydroxylated. The *cis* form is usually much less active. Reduction of the double bond in the side

chain leads to dihydrozeatin. Aromatic cytokinins have an aromatic benzyl group at N^6 . They occur more rarely and much less is known about them. Because of their greater stability, aromatic cytokinins are often used in tissue culture; an example is benzyladenine. In addition, there are the structurally unrelated phenylurea-type cytokinins (e.g., diphenylurea and thidiazuron), a class of synthetic cytokinins. These cytokinins are highly active but do not occur naturally.

Cytokinin Biosynthesis and Metabolism

The rate of *de novo* synthesis, metabolic interconversion, and breakdown are, together with transport processes, relevant to the regulation of cytokinin homeostasis in cells. Cytokinin metabolism includes mainly conversions among cytokinin bases, ribosides, ribotides, side-chain modification, conjugation and conjugate-hydrolyzing reactions, and cytokinin degradation.

Biosynthesis

The initial and rate-limiting step of biosynthesis of isoprenoid-type cytokinins is the transfer of the isopentenyl moiety from dimethylallyl pyrophosphate (DMAPP) to adenosine monophosphate (AMP), adenosine diphosphate (ADP), or adenosine triphosphate (ATP). The reaction is catalyzed by DMAPP::AMP/ADP/ATP isopentenyltransferases (IPT). ADP and ATP are the preferred substrates of most of the known plant IPT enzymes, while bacterial enzymes prefer AMP. The reaction leads to the formation of isopentenyl-AMP, -ADP, and -ATP, which are the precursor molecules of biologically active cytokinins. The isopentenyl side chain is subsequently hydroxylated to form zeatin-type cytokinins (Figure 1). An alternative pathway, in which an already hydroxylated side chain is directly added to the N^6 -position of the adenine moiety, may exist. IPT enzymes are encoded in *Arabidopsis* by a small gene family with seven members (*AtIPT1*, *AtIPT3*–*AtIPT8*). *AtIPT* genes are expressed in specific tissues of the root and shoot (e.g., vasculature), indicating that cytokinin synthesis occurs in all major organs. Another possible source of cytokinins is

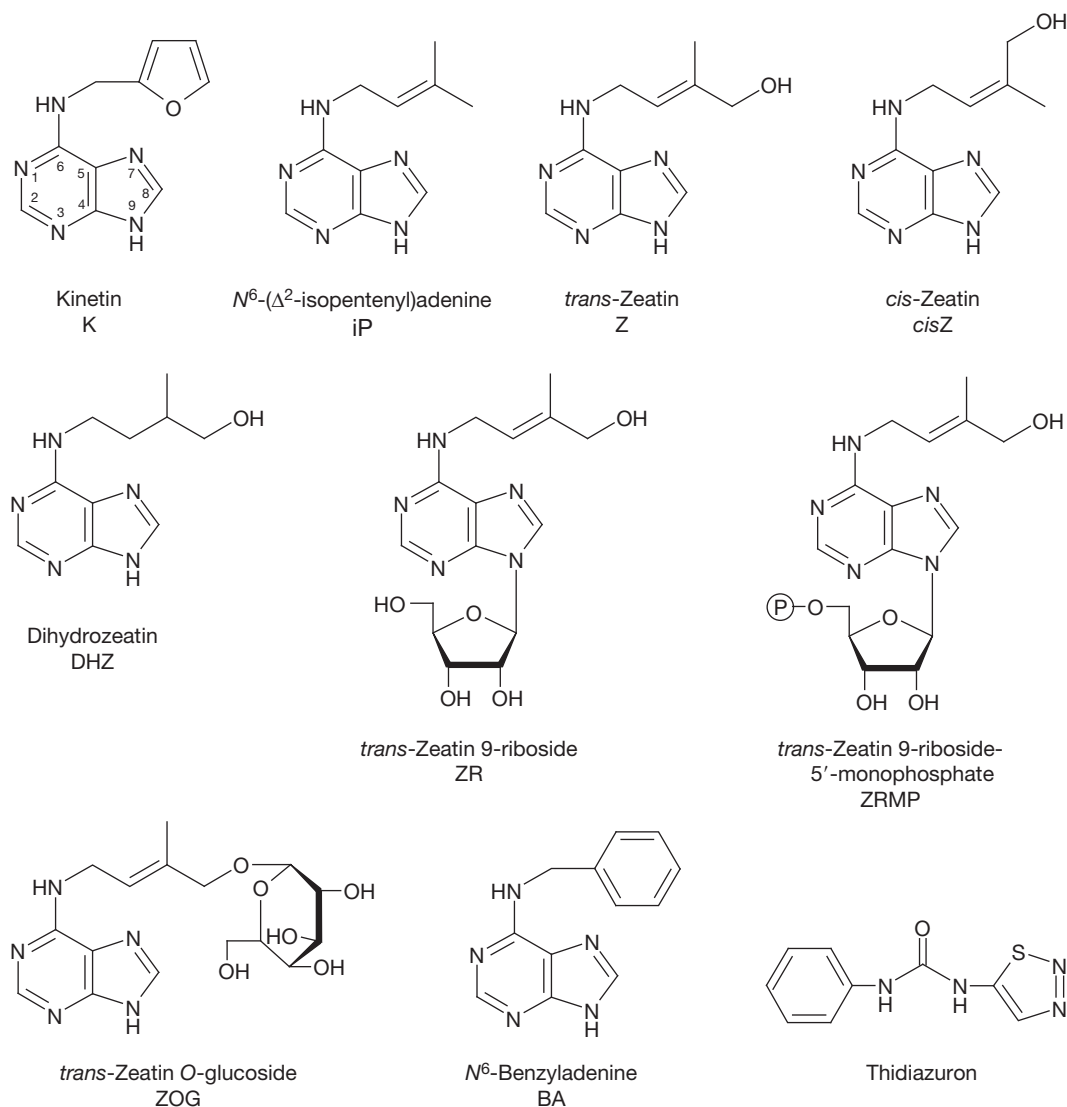


Figure 1 Chemical structures of some naturally occurring and synthetic cytokinins. Common names and abbreviations are indicated below the structures. The numbering of purine ring atoms is shown for kinetin.

transfer RNA (tRNA), since tRNAs of most organisms contain isopentenylated adenine and other structural derivatives with cytokinin activity. However, it is generally assumed that tRNAs play only a minor role, if any, as a cytokinin source.

Interconversion

A characteristic feature of cytokinin metabolism is the rapid metabolic interconversion of base, ribosides, and ribotides. The biologically most active form of cytokinins is the base. Attachment of ribose or ribose-5'-phosphate to the N⁹ atom of the adenine ring leads to the formation of ribosides and ribotides, respectively, which do have lower activities (**Figure 1**). Interconversion of cytokinins is presumably an important mechanism to regulate the concentration of active compounds. Cytokinin ribosides are probably relevant as a transport form. The interconversions may be catalyzed by the same enzymes that metabolize adenine, adenosine, and AMP. The conversion

between the *cis*- and *trans*-isomers of zeatin is catalyzed by the enzyme *cis-trans* zeatin isomerase. Zeatin is converted to dihydrozeatin by a nicotinamide adenine dinucleotide phosphate (reduced)-dependent zeatin reductase.

Conjugation

Cytokinins can be stably or transiently inactivated by glycosylation of the purine ring or of the side chain. The purine ring can be glycosylated at the N³-, N⁷-, and N⁹-positions. In addition, the N⁶-side chain group can form O-glycosyl conjugates if it bears a hydroxyl-group. Most often, glucose is the conjugated sugar molecule, more rarely xylose is attached. N⁷- and N⁹-conjugates are biologically inactive and extremely stable. Thus, they are irreversibly inactivated cytokinins. N³- and O-conjugates are biologically inactive but can be readily hydrolyzed. They are believed to be transient storage forms of cytokinins. Glycosyl conjugation is considered to be important in

the regulation of cytokinin activity levels, at least in some tissues and species. Several genes coding for cytokinin glycosyltransferases and glycosidases have been identified. Some conjugates of cytokinins and amino acids (alanine) have been described as well.

Catabolism

Cytokinins are irreversibly degraded in a single enzymatic step by oxidative cleavage of the N^6 -side chain. The reaction is catalyzed by cytokinin oxidases/dehydrogenases (CKX), which contain flavin adenine dinucleotide as a cofactor. The reaction products are adenine and an aldehyde. The preferred substrates of CKX are isopentenyladenine, zeatin, and their corresponding ribosides. Ribotides, *O*-glucosides, dihydrozeatin, and aromatic cytokinins are not degraded by CKX. The *Arabidopsis* genome contains seven *AtCKX* genes, which are preferentially expressed in zones of active cell division and growth. The corresponding enzymes are located in the endoplasmatic reticulum, in the apoplast, and in the vacuole.

Cytokinin Transport

Cytokinins are transported from roots to shoots in the xylem, and in the opposite direction in the phloem. Transported cytokinins may have a role in coordinating root and shoot development, for example, by carrying information about nutrient availability. Multiple cellular importers and exporters are required to allow efficient mobilization and targeted translocation of cytokinins, but very little is known about cytokinin transporters. Transport studies indicate that a common H^+ -coupled high-affinity purine transport system transports cytokinins.

Cytokinin Signaling

The mechanism of cytokinin signaling is just beginning to emerge. The cytokinin signal is perceived and transduced by a multistep phosphorelay system through a complex form of the two-component system (TCS) pathway. The TCS is common among prokaryotes and lower eukaryotes; among the higher eukaryotes, it is unique to plants. In this signaling system, a membrane-located receptor kinase with an extracellular ligand-recognition domain (sensor) dimerizes upon binding a ligand and autophosphorylates a histidine within its cytoplasmic transmitter domain. The phosphoryl group is first transferred to an aspartate residue within the receiver domain at the C terminus of the receptor and from there to a His-containing phosphotransmitter (Hpt), which ultimately phosphorylates and thus activates a response regulator (RR) at a central Asp residue (see [Figure 2](#)).

Signal Perception

Cytokinin receptors are histidine kinases consisting of an extracellular sensing domain, a cytoplasmic histidine kinase transmitter, and receiver domains. Three cytokinin receptors (CRE1/WOL/AHK4, AHK2, and AHK3) have been identified in *Arabidopsis*. They all share a ~270-amino-acid long extracellular

cyclases/histidine kinases associated sensing extracellular domain, which presumably recognizes cytokinin. This domain might have been acquired by plants through lateral gene transfer from cyanobacteria. Loss-of-function mutants of CRE1/WOL/AHK4 lack the phloem in their primary roots, indicating a role for cytokinins in embryo development.

Signal Transduction

Current knowledge suggests that downstream signaling components of the cytokinin signal-transduction pathway in *Arabidopsis* consist of five Hpt and 22 RRs of the A- or B-type. Hpts transmit the signal from the receptor, which is presumably localized in the plasma membrane, to B-type RRs, which are in the nucleus. B-type RRs consist of an N-terminal receiver domain and a C-terminal output domain, containing a DNA-recognition motif called genomic and RNA profiling (GARP), which is distantly related to the Myb repeat. The DNA motif optimal for binding is 5'-(A/G)GAT(T/C)-3'. Activated B-type ARRs transcribe primary response genes of cytokinins. Some of the known primary response genes, which are rapidly and specifically upregulated by cytokinins, code for type A RRs. Type A RRs resemble type B RRs but lack the C-terminal DNA binding and activation domain. Type A RRs fulfill at least two different functions. On the one hand, they exert a negative feedback regulation of the cytokinin signaling pathway through protein-protein interaction. On the other hand, they mediate the cytokinin-dependent modulation of other pathways, for example, light signaling. A-type RRs can be positive or negative regulators, depending on the individual RR and the output reaction analyzed.

Cytokinin Functions

Cell Cycle

Cytokinins are required for cell division during embryogenesis in the shoot apical meristem, young leaves, the cambium, and cultured plant cells. By contrast, they have a negative regulatory role in the root meristem. Cytokinin controls the exit of dividing cells from this meristem. Changes in cytokinin levels occur during the cell cycle of cultured cells, the level being highest during the late S and before the M phase. Cytokinins have been functionally linked to all stages of the cell cycle but their mechanism of action has only been partially elucidated. Cytokinin upregulates expression of the D-type cyclin gene *CycD3*, which is important in regulating the G1/S-transition of the cell cycle. Cytokinin increases the number of replication origins during S-phase and it may also play a role in regulating G2/M transition.

Plant Development and Growth

Cytokinin participates in regulating numerous aspects of plant development throughout the life cycle. These include seed germination, cotyledon expansion, chloroplast differentiation, de-etiolation, differentiation of vascular tissue, apical dominance (shoot branching), root elongation and branching, nutritional signaling, regulation of sink strength, the transition from the vegetative to the reproductive growth phase, flower

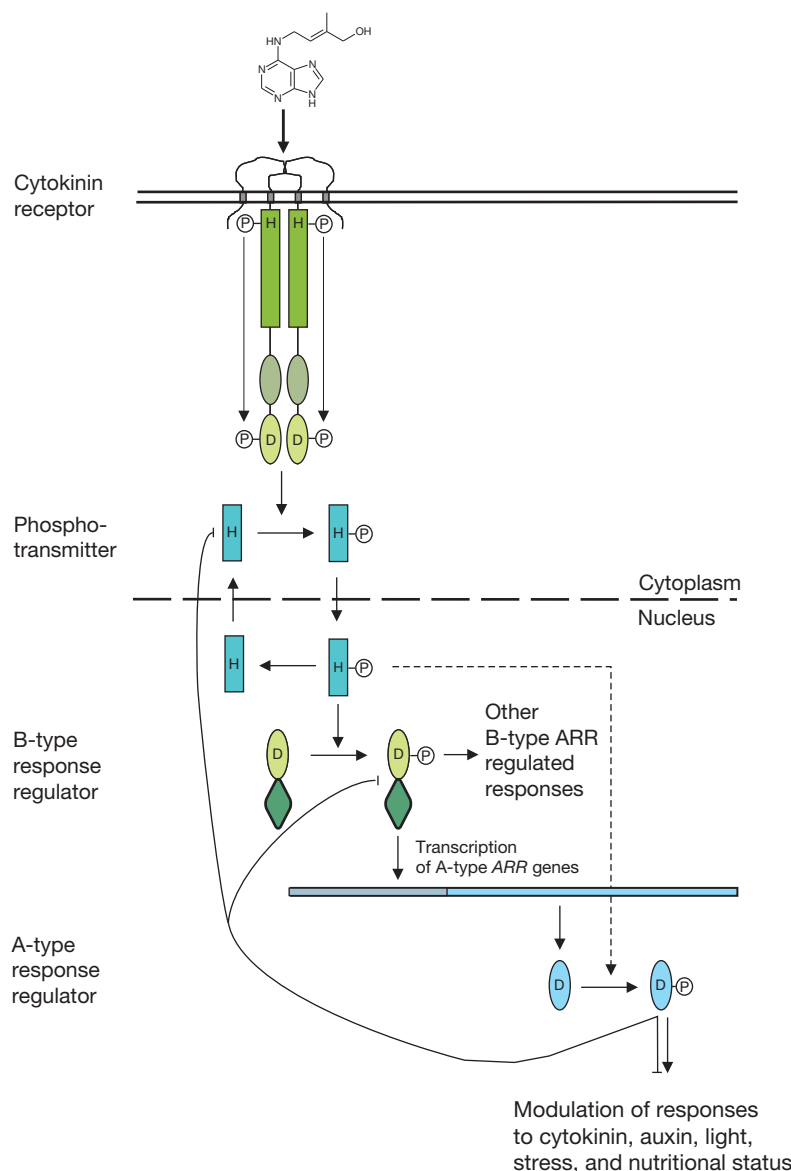


Figure 2 A model for cytokinin signal transduction via a His-to-Asp phosphorelay. The structure of CRE1/WOL/AHK4 is shown as an example. Ligand binding induces receptor dimerization and autophosphorylation. Transfer of the phosphoryl group by activated receptors activates histidine phosphotransmitter proteins (Hpts), which transport the signal from the cytoplasm to B-type RRs in the nucleus. B-Type RRs transcribe target genes, among them A-type *ARR* genes. A-Type RRs may downregulate the primary cytokinin signal response via a negative feedback loop, modulate downstream activities of cytokinins in a positive or negative fashion, or modulate other signaling pathways through protein–protein interaction. A more complex regulation than shown in the model may exist. D, aspartate residue; H, histidine residue; P, phosphoryl group.

and fruit development, leaf senescence, and plant–pathogen interactions. A role for the hormone in vascular morphogenesis during embryonic development is firmly established. During postembryonic development, cytokinins are required to maintain meristem activity and leaf development in the plant shoot. Local exogenous cytokinin application to the shoot leads to premature growth of lateral buds, retarded leaf senescence, partial photomorphogenesis in the dark, increased sink strength, and an altered vasculature. In contrast to their stimulatory activities in the shoot, cytokinins have a negative regulatory role in the control of root elongation and branching. Additionally, cytokinin regulates important physiological

parameters that determine biomass formation and distribution via central genes of primary metabolite pathways, including invertases, hexose transporters, and key genes of phosphate and nitrogen metabolism and signaling (e.g., nitrate reductase). Changes in cytokinin levels are generally positively correlated with levels of mineral nutrients, especially nitrogenous nutrients. Cytokinin levels are decreased by water stress. *In vitro*, the ratio of cytokinin to auxin determines the differentiation of cultured plant tissues to either shoots or roots. A high cytokinin to auxin ratio promotes shoot formation; a low ratio promotes root formation. Owing to their stimulatory effect on plant regeneration, cytokinins are widely used in plant tissue culture.

Pathogenicity

Cytokinins are produced by several plant pathogenic bacteria and play a role in pathogenicity. One such pathogen is *A. tumefaciens*, the causative agent of the crown gall disease. During the infection process, *A. tumefaciens* transfers a small stretch of DNA, the T-DNA, to the host plant, where it becomes integrated in the nuclear genome. The T-DNA harbors an *IPT* gene, which is expressed in the host cell and causes cytokinin overproduction. This leads, together with an enhanced auxin content, to tumorous cell proliferation. Other cytokinin-synthesizing pathogens are *Pseudomonas syringae*, which induces gall formation and *R. fascians*, which causes fascinations and a growth abnormality called 'witch's broom disease'. The root-nodule forming and nitrogen-fixing plant symbiont *Rhizobium* sp. is also known to produce cytokinin.

Biotechnology

Practical use of cytokinin in agriculture is currently limited. Modulation of the endogenous cytokinin content of plants or interfering with cytokinin signaling has a high potential for biotechnological applications in agriculture. Plants with increased cytokinin content are more branched and senesce later. Moreover, cytokinins alter sink-source relations, a promising approach to improve yield attributes. Plants with reduced cytokinin content develop a larger root system. An improved root system means improved acquisition of

minerals and water, factors which are often limiting for plant growth.

See also: **Cell Architecture and Function:** Cytokinesis; Septins and Cytokinesis.

Further Reading

- Davies PJ (1995) *Plant Hormones. Physiology, Biochemistry and Molecular Biology*. Dordrecht: Kluwer Academic.
- Heyl A and Schmölling T (2003) Cytokinin signal perception and transduction. *Current Opinion in Plant Biology* 6: 480–488.
- Hoykaas PJJ, Hall MA, and Libbenga KR (1999) *Biochemistry and Molecular Biology of Plant Hormones*. Amsterdam: Elsevier.
- Kakimoto T (2003) Biosynthesis of cytokinins. *Journal of Plant Research* 116: 233–239.
- Kakimoto T (2003) Perception and signal transduction of cytokinins. *Annual Review of Plant Biology* 54: 605–627.
- Kieber JJ (2002) Cytokinins. In: Somerville C, Meyerowitz E, and Rockville MD (eds.) *The Arabidopsis Book*, American Society of Plant Biologists. ASPB Publications. <http://www.aspb.org/publications/arabidopsis> (accessed July 2012).
- Mok DWS and Mok MC (1994) *Cytokinins. Chemistry, Activity, and Function*. Boca Raton, FL: CRC Press.
- Mok DWS and Mok MC (2001) Cytokinin metabolism and action. *Annual Review of Plant Physiology and Plant Molecular Biology* 52: 89–118.
- Schmölling T, Werner T, Rietler M, Krupková E, and Manns IB (2003) Structure and function of cytokinin oxidase/dehydrogenase genes of maize, rice, *Arabidopsis*, and other species. *Journal of Plant Research* 116: 241–252.
- Taiz L and Zeiger E (2002) *Plant Physiology*. Sunderland, MA: Sinauer.

Cytoskeletal Motors: General Principles

R S Rock, The University of Chicago, Chicago, IL, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Actin-binding protein (ABP) A protein that binds to actin and modulates its function. Functions of ABPs include nucleating new filaments, decorating the sides of filaments, bundling filaments, and severing filaments.

Gating A mode of regulation where biochemical or mechanical events on one catalytic domain of a motor cannot proceed until separate events that serve as checkpoints are completed on the other

catalytic domain; this is related to half of the sites reactivity.

Half of the sites reactivity A form of allostery where one out of two catalytic sites in a protein is inactive at any given time. Thus, the protein has half as many active sites as expected.

Microtubule-associated protein (MAP) A protein that binds to microtubules and modulates their function. Functionally similar to the ABPs.

A Walking Model for Cytoskeletal Motors

In the following narrative, we describe the structure of the cytoskeletal motors and the arrangement of their filament tracks. For simplicity, we consider these motors to be bipedal molecules, walking along their filament tracks much like we walk on the ground (Figure 1). In this analogy, the two feet are the two motor domains that bind to the track and turnover adenosine triphosphate (ATP) as fuel. The feet bind stereospecifically to the track, always facing the same end of the cytoskeletal filament. This feature allows motors to travel in one direction, set by the polarity of the track. The two legs are elongated structural elements that swing from one orientation to another. These legs vary considerably in length, flexibility, and structure between and within motor families. Finally, the hips define the point where the two legs join and where cargo or scaffolding proteins are attached. In describing the parts of cytoskeletal motors in this way, we are dispensing with the historical but confusing description of heads, necks, and tails. These motors take forward steps by picking up their trailing foot and placing it in the leading position, yielding a rapid and discrete step. The fuel for this process is ATP. Many motors are processive, meaning they move continuously and take many steps before detaching from their tracks. Those that are non-processive are specialized and usually operate in large arrays to support continuous movement.

Following the motions of these intricate machines required the development of fundamentally new biochemical and biophysical approaches. The key challenge was the difficulty in monitoring the main product of motor proteins, namely motion. One cannot simply mix motors, tracks, and ATP in a tube and measure total motion; one must track objects as they move. Furthermore, multiple motors would compete with each other unless the system was properly engineered. The Spudich group developed the first successful *in vitro* motility assays in the 1980s. Their gliding filament assay monitored the movement of labeled actin filaments over a surface coated with many myosin motors. This system was later adapted to track the motion of microtubules driven by kinesin or dynein. Further refinements led to single-molecule motility assays that allowed us to follow the action of individual motors, and to

avoid complications associated with tracking many unsynchronized motors at once.

These single-molecule assays now allow us to track the individual parts within the motor protein. For example, fluorescent probes placed on the hips will report the location of the center of mass of the motor protein as it steps along its track. Modern techniques allow us to track the location of this fluorescent probe to within a few nanometers. If a probe is instead placed on the leg near the foot, a specific long-short-long-short stepping pattern is seen. To see why this pattern emerges, imagine tracking the location of your right knee as you walk. With your right foot in front as you take a step with your left foot, your right ankle will rotate, with a small translation of your right knee forward. With your next step, you pick up your trailing right foot and swing it forward, resulting in a large translation of your right knee. Thus, by careful placement of probes, it is possible to map out the moving parts of the cytoskeletal motors. These observations largely confirm many aspects of the walking model for each of the three cytoskeletal motors. In a stunning recent development, the Ando group imaged myosin motors walking along actin filaments in solution with a video-rate atomic force microscope. In this experiment, the legs of a myosin motor are clearly resolved, allowing us to see the walking pattern of a motor protein in real time, and without the need to attach molecular probes to the motor.

Caveats of the Walking Model

Of course, the walking model described above is an oversimplification, and it is easy to stretch the analogy too far. The analogies extracted from walking around the macroscopic world is of little help when considering the motion of cytoskeletal motors as they are continuously buffeted by thermal forces from the solution. Here, we describe some of the features of cytoskeletal motors that diverge from our expectations derived from the motion of our legs as we walk.

The first major difference is in where the driving force is generated. When we walk, we are primarily using the large muscle groups found in our hips and thighs. For kinesins and myosins, however, the motor domain is found in the feet, where it drives the rotation of the ankle. The notable exception

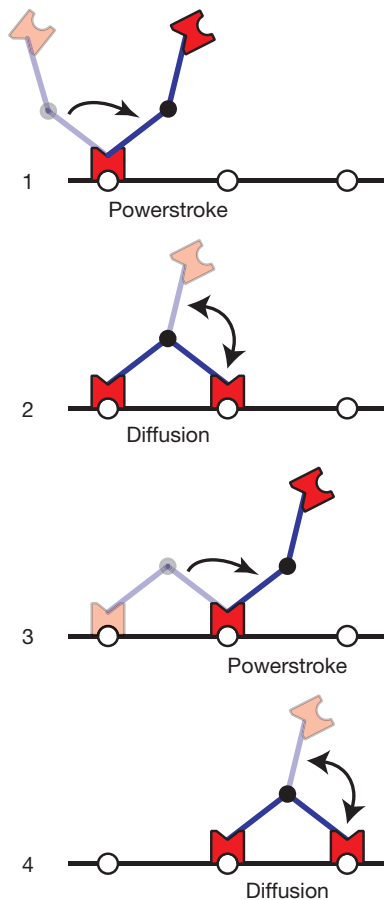


Figure 1 A general stepping scheme for cytoskeletal motors. A generic motor is shown, with motor domain feet (red), elongated legs that transmit motions (blue lines), and hips where the two legs join and cargo is attached (black circle). These motors walk along lattice sites on a cytoskeletal filament (black line with open black circles). The legs either may be rigid, as shown or may be flexible and buttressed against the feet, as is the case for kinesin. Dynein places the motor domain in the hips instead of the feet. The two phases of a step are shown. In phase 1, the leg swings, placing the free foot closer to its next lattice site on the filament. In phase 2, the free foot diffuses around the location of the hips, until it rebinds the filament. Phases 3 and 4 repeat the processes in 1 and 2, illustrating the transition to the next step.

is dynein, which, like us, seems to drive motility from the hips where its ATPase with associated activity (AAA) domains do the work. But without movement at the hips, how do these two cytoskeletal motors walk? We would certainly fall over with motion only at the ankles.

Here is where the massive thermal forces in the microscopic environment take over. At the instant the back foot releases from the track, the ankle of the front foot swings (an event sometimes called the ‘powerstroke’ or ‘working stroke’ of the motor). The free foot now undergoes a diffusive search for a new binding site on its track (Figure 1). Thus, the pivoting at the hips is entirely driven by Brownian motion of the free leg. This search is centered upon the location of the hips, and because the hips move forward when the bound leg swings forward, the search is biased toward forward sites on the track.

A second major difference has to do with the structure of the track. Actin filaments and microtubules both present a fixed lattice of binding sites where the feet may step. This feature places significant constraints on the stepping pattern of motor proteins. Whereas we can change our stride when walking along the ground, motor proteins are stepping on discrete sites, much like stepping stones in a river. Depending on the spacing of these stones, they have either many options (like myosin VI stepping along actin or dynein along microtubules) or few (like kinesin along microtubules, where only a single tubulin binding-site is available and stepsizes are therefore highly regular).

We might expect that the motor proteins have evolved to precisely match their tracks, much like the gears meshing in a Swiss watch. However, this level of precision does not seem to be the case for cytoskeletal motors. For example, a single leg of myosin V will naturally swing 25 nm with each step. However, the next binding site along actin is 36 nm away. The myosin V must rely on thermal forces to carry it the rest of the way. When the free foot lands on actin, the whole myosin must distort to make up the 11-nm distance. It is as if it must stretch to reach a stepping stone that is just a bit farther than its natural stride. This stretching is one way for a cytoskeletal motor to keep track of which foot is in front, and which foot is in back. A foot that senses a rearward tug is in front, and is in back if it senses a forward tug.

A third difference with our walking experience has to do with the timing of events in stepping. We walk in a rather rhythmic fashion, with our feet touching the ground after a nearly constant time in the air. However, cytoskeletal motor steps are coordinated through discrete biochemical events, and the timing of these events is stochastic.

Finally, our discussion of walking motors leaves out an entire class of motors that do not walk with two feet at all. There are certain cytoskeletal motors that may act more as tethers or crosslinkers, and can operate with only a single foot. There are others that have feet that spend most of their time detached from the track, up in the air, as it were. These motors will typically operate as part of a large ensemble, which virtually guarantees that at least one foot will be on the track at all times. Rather than being bipedal motor proteins, these motors (like skeletal muscle myosin) bear closer resemblance to centipedes with many legs, but with only a few feet on the ground at any given time.

Regulation of Stepping

Our legs are under conscious control; we decide when and where we want to move them. Clearly, molecular motors are not able to make the same decisions. Nevertheless, motors must be regulated in some manner; otherwise, they would make serious errors such as moving to their final destination before their cargo is attached. We have now uncovered a number of these forms of regulation.

The most direct forms of regulation are modifications of the motor or the track. A number of motors switch activity based on posttranslational modifications. For example, certain myosins are phosphorylated, either directly or on their associated light chains, which controls their ability to engage their actin

track or to assemble into larger structures. Certain kinesins are likewise regulated by phosphorylation. Track-based modifications may take a number of forms. Accessory proteins (ABPs for actin, MAPs for microtubules) or posttranslational modifications (acetylation of microtubules) can recruit or inhibit certain motors. Moreover, certain architectures of cytoskeletal filaments may be required for motor activity. Some kinesins are found at overlapping, antiparallel microtubules found at the central spindle, while some myosins select bundled actin filaments in preference to single actin filaments.

An emerging regulatory theme is the idea of cargo activation of cytoskeletal motors, which ensures that motors do not engage their tracks and start moving prematurely. Cargo binding activates the motor in a concerted manner, removing the requirement for separate modes of regulation for cargo binding and motor activation. There are a couple of common forms of cargo activation. In one, the motor protein will fold back on itself in the absence of cargo. This folded configuration, where the motor is grabbing its own feet, is autoinhibited and incapable of walking along the track. Cargo binding releases this folded configuration, and simultaneously activates the motor and allows it to commence walking. A related form of regulation occurs when cargo binding triggers motor dimerization. In the absence of cargo, certain motors (e.g., the kinesin KIF1a or myosin VI) are monomeric in solution, as if they have only one foot and leg. Individual monomeric motors can bind to cargoes to trigger motor dimerization and activation. Thus, it is not easy to predict the arrangement of feet and legs for a motor protein without considerable structural work on the cargo-binding domains and associated scaffold proteins.

A cautionary tale is warranted here. In recent years, a number of groups have claimed that certain molecular motors must operate as monomers within the cell, since full-length versions of the motors are monomeric in biochemical assays. These groups have bolstered these claims by expressing isolated dimerization domains from these motors and showing that they do not dimerize in their hands. However, these claims neglect the influence of the cargo-binding domains of the motors. If cargo binding is tight and leads to motor dimerization at the point of the cargo-binding domains, then the local concentration of the two motor domains can exceed one millimolar based on the volume defined by the dimensions of typical motor proteins. Therefore, extremely weak dimerization that is difficult to detect may still be critical for motor function. Indeed, motors that form regulated dimers likely require weak dimerization domains, so that dimers form only after the assistance of cargo and not spontaneously.

Coordination and Gating

We walk in a highly coordinated manner. A set of sophisticated neural circuits triggers muscle contraction in the correct sequence, so that our back leg lifts and our front leg swings. Proper coordination involves proprioceptive neurons that sense the orientation of our limbs. Likewise, cytoskeletal motors often employ forms of coordination to ensure that the back leg lifts before the forward leg, and also to ensure that both legs do not lift at the same time.

For cytoskeletal motors, such coordination is achieved through carefully tuned ATPase kinetics. In particular, these motors switch off the ATPase activity of one of the two feet.

This alternating inactivation is commonly called 'half of the sites reactivity' or 'gating'. It is thought that an allosteric, strain-sensitive mechanism is responsible for identifying front foot, rear foot, and switching of ATPase activity from one foot to the other, but the molecular details remain obscure.

There are two general ways for a motor to be gated. In one mechanism, the progress of the front foot is halted. The front foot knows that it is the front foot when its leg is leaning backward. In this mechanism, this front foot waits until the back foot releases. At this moment, the motor is vulnerable to detach from the track. To avoid this premature detachment, the kinetic pathway requires a slow event in the bound front foot, which occurs sometime after its leg swings forward. This slow event allows plenty of time for the free foot to rebind to the track before the bound foot releases. Thus, a kinetic competition exists between taking another forward step and detaching from the track, and the kinetics favor the landing of the free foot.

In the second mechanism, the action occurs on the back foot. A slow back foot waits for a tug from the front foot before continuing through its ATPase cycle and detaching. Although the essential features are similar to front-foot gating, a curious feature of this mechanism is that the allosteric communication mechanism does not actually increase the processivity of the motor. The strain that accelerates stepping is only felt once both feet are on the track and motor is no longer vulnerable to detach. Instead, strain is used here to increase the stepping speed. In practice, mixed models are possible, where gating occurs on both the front foot and the back foot at the same time.

Finally, some motors manage to walk processively even without gating or half of the sites reactivity. They do so with sticky feet that remain bound to the track nearly all of the time. Since the two feet step independently of each other, these motors make many mistakes, lifting the front foot before the back. The stepping pattern of these ungated motors resembles a biased diffusive walk rather than the regular forward stepping that we would expect. The common structural feature of these ungated motors is that they have long and flexible legs that do a poor job of transmitting strain. However, even these flexible motors must be gated some fraction of the time. If the two feet have taken diffusive steps and end up far apart on the track, all of the slack in the flexible legs may be absorbed. In this case, the feet are no longer independent: the motor would have no choice but to lift its back foot with the next step.

Mechanical Performance

All cytoskeletal motor proteins pull on filaments, generating force. Moreover, all motor proteins have fundamental limits on their ability to perform mechanical work. These limits are important to consider in the cellular context, where drag forces experienced by the cargoes or resisting tension in cytoskeletal filaments may become significant. In general, if an external force that opposes forward motion is applied to a motor protein, its forward stepping rate will decrease. Eventually, the forward stepping rate decreases to the point where it equals the backward stepping rate. At this force, the motor is stalled, meaning that it can move no farther, on average.

Rate-constants for biochemical events coupled to displacements, such as working strokes in cytoskeletal motors, depend

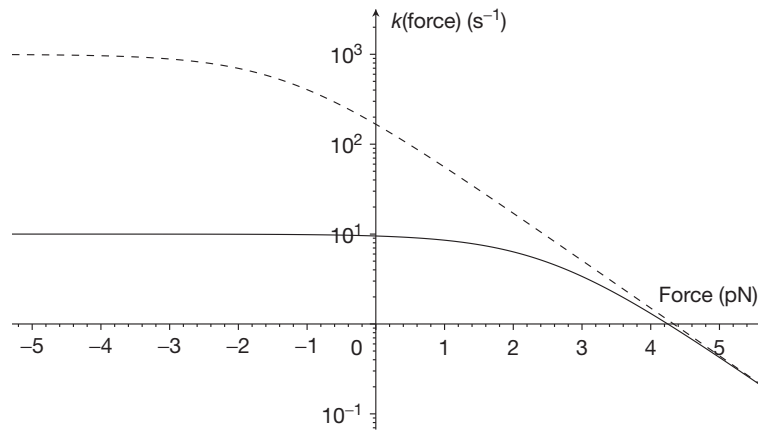


Figure 2 Overall stepping rates as a function of load. The general case is shown, with a load-dependent event and a load-independent event occurring in series with each mechanical step. The solid line is for $k_i = 10 \text{ s}^{-1}$, $k_0 = 200 \text{ s}^{-1}$, and $d = 5$. Note that the stepping rate does not vary with force around zero force, as k_i dominates. The dashed line uses the same parameters, except $k_i = 1000 \text{ s}^{-1}$. Here, the stepping rate does vary with force near zero force, as k_0 dominates in this region.

on the external force with a Boltzmann-like dependence, first described by Bell in 1978:

$$k(F) = k_0 e^{-Fd/k_b T}$$

Here, k_0 is the rate constant at zero force, F is the applied force, d is the distance to the transition state, and $k_b T$ is thermal energy ($\sim 4 \text{ pN nm}$ at room temperature). The sign convention here has positive forces (load) opposing the transition, so that the exponential factor is less than one and the rate constant is reduced with increasing force. The remarkable feature of Bell's equation is that it allows us to measure distances associated with a structural change solely through the effect on rate constants. Note that the distance, d , is the distance to the transition state, not the step size of the motor. Large values of d correspond to highly force-sensitive transitions, while low to zero values of d are transitions insensitive to force. For a unidirectional swing of the leg, we expect that d will be less than the motor step size.

Practical measurement of d requires instrumentation that can measure stepping rates as a function of applied force. Optical traps are ideal instruments to apply precise forces to cytoskeletal motors as they step. Because each step may require force-sensitive and force-insensitive events that occur in series, the general expression for the observed stepping rate is

$$k(F) = \frac{k_i k_0 e^{-Fd/k_b T}}{k_i + k_0 e^{-Fd/k_b T}}$$

where k_i is a composite rate for all force-insensitive transitions. The composite force-dependent rate, k_0 , could be separated into multiple individual force-dependent transitions, each with their own d , but it is rarely the case that the additional parameters are justified by the data. The most common scenario for cytoskeletal motors is a rate-limiting and load-independent transition described by k_i . The signature of this mechanism is load-independent stepping at low load, followed by a transition to load-dependent stepping over a narrow range of forces as k_0 eventually dominates (Figure 2, solid line). Alternatively, k_0 may itself be rate limiting, which requires load-dependent stepping near zero load (Figure 2, dashed line).

The manner in which a cytoskeletal motor approaches stall can significantly constrain its role in the cell. There are three general ways that a cytoskeletal motor can approach stall: (1) it can detach before reaching stall, (2) it can stop stepping altogether as it approaches the stall force (static stall), or (3) it can step forward and backward around the stall force (dynamic stall). To transport cargo, any of the three options are acceptable since the motor has an opportunity to reengage a track after a failed run. However, to regulate the tension of cytoskeletal filaments, the best option is (3), dynamic stall. As the motor experiences superstall loads, it can relax to the original stall force without detaching and suddenly dissipating tension. In contrast, a static stall motor acts more like a ratchet, reeling in slack filaments but stubbornly holding on as load increases. Likewise, detaching before the stall force is reached is a poor option for a cytoskeletal motor that regulates tension, since the detachment force will vary considerably with each run.

Conclusion

The biped walking picture of cytoskeletal motors is a useful starting framework for understanding how they operate within the cell, especially in understanding the order of the mechanical events. However, we must remain aware that there are important differences in the mechanical environment and the coordination mechanisms that affect how motors walk relative to how we walk.

See also: **Cell Architecture and Function:** Dynein; Kinesin Superfamily Proteins; Myosin Motors.

Further Reading

- Howard J (2001) *Mechanics of Motor Proteins and the Cytoskeleton*. Sunderland, MA: Sinauer Associates.
- Vale RD (2003) The molecular motor toolbox for intracellular transport. *Cell* 112: 467–480.
- Vale RD and Milligan RA (2000) The way things move: Looking under the hood of molecular motor proteins. *Science* 288: 88–95.

Desmosomes and Hemidesmosomes

V Todorović, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

K R Kligys, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

R L Dusek, Stanford University School of Medicine, Stanford, CA, USA

J C R Jones, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

K J Green, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Rachel L Dusek, Jonathan C R Jones and Kathleen J Green, volume 1, pp. 569–576, © 2004, Elsevier Inc.

Glossary

Cadherins Single-pass transmembrane proteins that mediate calcium-dependent cell–cell adhesion.

Epithelia Sheets of tightly packed cells that cover a body surface or line organs and body cavities.

Extracellular matrix An organized network of proteins and polysaccharides that fills the extracellular space of tissues and is produced by the surrounding cells.

Integrins Transmembrane linker proteins that are receptors for extracellular matrix proteins and link the extracellular matrix to components of the cytoskeleton.

Intermediate filaments Cytoskeletal filaments, typically 10 nm in diameter, found in higher eukaryotic cells.

Signal transduction The propagation of a chemical or mechanical stimulus to affect a cellular response.

The Desmosome

Desmosomes are prominent in tissues that are subjected to mechanical stress. Found most abundantly in stratified squamous and simple epithelial tissues, desmosomes have also been noted in the myocardium, brain meninges, and follicular dendritic cells of the lymph nodes. These specialized anchoring junctions function to mediate intercellular adhesion and maintain tissue integrity by connecting sites of cell–cell contact to the highly tensile intermediate filament (IF) cytoskeleton. In the myocardium, desmosomes have classically been described as part of a junctional complex called the intercalated disk, also containing adherens junctions and gap junctions, which chemically and electrically couples beating cardiac myocytes. More recently, it has been appreciated that in mammals a mixed type of cell–cell junction termed the ‘area composita’, consisting of desmosomal and adherens junction components, is formed postnatally. The functional significance of this junction or its species specificity is not understood.

Desmosome Ultrastructure

Ultrastructurally, the desmosome appears as a mirror image of a highly organized, electron-dense region adjacent to the cell membrane (Figure 1, left). The plasma membranes of adjacent cells are separated by a ~30-nm space containing an electron-dense

midline, likely composed of the extracellular domains of the transmembrane desmosomal components. Just internal to the plasma membrane is an electron-dense outer plaque containing the intracellular desmosomal components. Interior to this region is a less-dense, IF-rich fibrillar inner plaque.

Structure and Molecular Components of the Desmosome

The basic blueprint of a desmosome comprises molecular components that fall into three main gene families: desmosomal cadherins, armadillo family members (plakoglobin (Pg) and plakophilins (PKPs)), and plakins. Several studies suggest that these molecules are arranged linearly in this order from the cell-surface inward and function together to indirectly tether the IF cytoskeleton to the cell membrane. This structural feature is shared by the closely related actin cytoskeleton-tethering adherens junctions, which have a similar molecular blueprint (with classical cadherins, armadillo proteins (β -catenin and p120), and α -catenin in place of their desmosomal counterparts). Lateral clustering of these proteins enhances the strength of the junction. Nonadhesive functions have also been proposed for several of these molecules. The best-characterized desmosomal components (Figure 1, right) are discussed in detail below. However, minor components, including pinin, perp, corneodesmosin, and desmocalmin/keratocalmin, have also been described and may have tissue- or differentiation-specific functions.

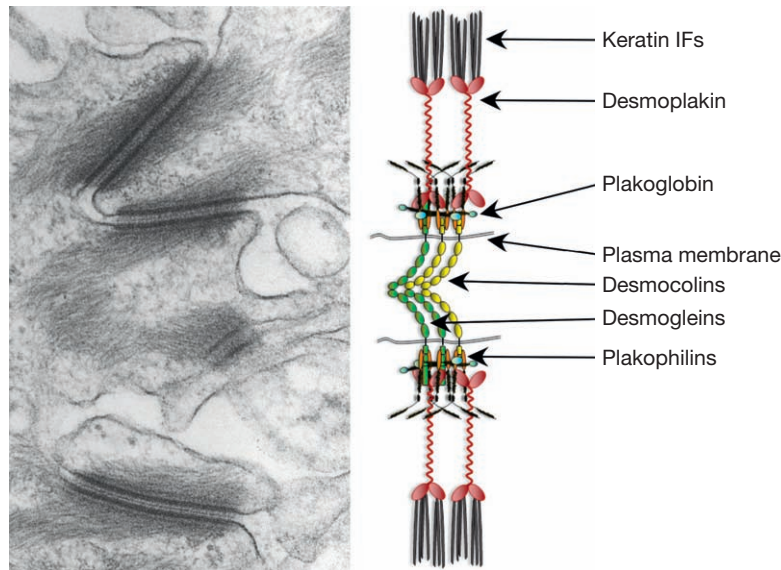


Figure 1 Molecular composition and structure of the desmosome. (Left) Desmosome ultrastructure is shown in this electron micrograph of bovine tongue epidermis. (Right) Desmosomal proteins, their interactions, and their approximate position within the organelle are depicted in this schematic diagram of the desmosome. Reproduced from Kowalczyk AP, Bornslaeger EA, Borgwardt JE, et al. (1994) Structure and function of desmosomal transmembrane core and plaque molecules. *Biophysical Chemistry* 50: 97–112.

Desmosomal cadherins

Two types of glycosylated, type I transmembrane adhesive cadherin proteins are found in the desmosome, desmogleins (Dsgs) and desmocollins (Dscs). Different genes encode four isoforms (1–4) of desmogleins and three isoforms (1–3) of desmocollins, which are expressed differentially in various cell types and in a differentiation-specific manner in complex stratified epithelia. The Dsc isoforms are further subdivided into two types, a longer a form and a shorter b form. Like classical cadherins, both Dsgs and Dscs have a highly conserved calcium-binding extracellular domain, membrane spanning region, and catenin-binding intracellular cadherin segment (ICS) (not present in Dsc b), while Dsgs contain additional, unique cytoplasmic subdomains. Emerging evidence implicates these unique subdomains of Dsg1 in epidermal differentiation through regulation of epidermal growth factor receptor (EGFR)–extracellular-signal-regulated kinase (Erk) signaling axis. Since most of the structural variability between Dsg family members rests in this unique region, it is possible that individual family members regulate distinct signaling pathways. Dsgs and Dscs are thought to function primarily in mediating homo- and/or heterophilic calcium-dependent adhesion across the membranes of adjacent cells. Severe blistering of the skin or lesions on the hands and feet are consequences of genetic, autoimmune, and bacterial diseases that compromise the adhesive function of these molecules (Table 1).

Armadillo family members

Proteins with homology to the *Drosophila* segment polarity protein, armadillo, are important structural components of the desmosome, linking the desmosomal cadherins at the plasma membrane with intracellular cytoskeleton components. These proteins contain a 42-amino-acid repeated motif called the arm repeat, which is thought to mediate protein–

protein interactions. Pg is one arm protein whose presence in both adherens junctions and desmosomes is selectively regulated. In the desmosome, it binds directly to the cytoplasmic tail of the desmosomal cadherins and to the obligate desmosomal protein, desmoplakin (DP). Pg's importance for tissue integrity is evident from the mouse model system where its deficiency leads to late-stage embryonic lethality due to bursting of the ventricles and massive blood leakage from the heart. Disruption in skin cell–cell adhesion also occurs in certain mouse backgrounds. Moreover, Pg mutation in humans results in cardiomyopathy, keratoderma, and wooly hair (Table 1). In addition to its adhesive role, Pg is also thought to function in a variety of signal transduction pathways, including interactions with its family member β -catenin and direct as well as indirect involvement in Wnt signaling pathway. Recently, regulation of Src and Rho pathways as well as regulation of cell–substrate interactions emerged as prominent Pg targets.

The PKPs (PKP1–3) are another group of armadillo family members that contribute to desmosomal adhesion. These proteins have a specific tissue distribution and are differentially expressed in complex stratified epithelia. PKPs are found in the desmosome as well as in the nucleus and messenger RNA (mRNA)-containing stress granules in the cytoplasm. In the desmosome, PKPs interact with the cytoplasmic tails of desmosomal cadherins, Pg, DP, and IFs. The role of individual PKP family members in desmosome assembly and IF attachment has not been completely unraveled; however, PKP1 is thought to enhance the recruitment of DP to the plasma membrane and to facilitate lateral clustering of plaque components. Genetic mutations in both PKP1 alleles negate the protein's function and result in excessive skin fragility (Table 1). Recent findings suggest that PKP2 acts as a scaffold, bringing together molecular components necessary for transport of DP to the membrane. One of the molecules recruited by PKP2 is protein kinase C

Table 1 Human disease associated with desmosomal components

<i>Genetic diseases</i>	<i>Desmosomal molecules mutated</i>	<i>Clinical manifestation</i>
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	Pkp2, DP, Dsg2, Pg, Dsc2	ARVC results in the replacement of right ventricular myocardium with fibro fatty tissue resulting in thinning of the right ventricular wall. Patients die due to sudden cardiac arrest arising from arrhythmia. Some ARVC patients (most notably Pg 188GCA insertion mutation) do not experience any skin symptoms.
Dilated cardiomyopathy (DC)	Dsg2, DP	DC results in ventricular dilation impairing myocardial systolic function resulting in heart failure.
Woolly hair with and without cardiomyopathy	DP, Dsc2, Pg	Woolly hair is a syndrome of the scalp hair and is characterized by frizzy and wiry hair giving it a wool-like appearance. Some patients with Pg mutations develop woolly hair, skin fragility, and diffuse palmoplantar keratoderma without ever developing cardiomyopathy.
Striate palmoplantar keratoderma (SPPK)	DP, Dsg1, Dsc2	SPPK results in the epidermal thickening of the palms and soles and, depending on the genetic background, is often combined with the appearance of woolly hair and cardiomyopathy.
Skin fragility ectodermal dysplasia syndrome	Pkp1	The syndrome results in trauma-induced blistering and subsequent thickening of the skin on palms and soles, abnormal hair, nails and fragile skin which blisters and peels.
Naxos disease	Pg	Naxos disease is a result of combination of ARVC, palmoplantar keratosis and woolly hair.
Localized autosomal hypotrichosis (LAH1)	Dsg4	Sparse, fragile hair with abnormal hair follicles, impaired hair keratinization and epidermal hyperproliferation.
Hypotrichosis and recurrent skin vesicles	Dsc3	Sparse, fragile hair with normal follicles. Recurring skin vesicles prone to bursting and leaving slow healing scars.
Hypotrichosis simplex of the scalp (HTSS)	CDSN (corneodesmosin)	HTSS results in the early hair loss (within the first decade of life) in the patients, with no other skin, hair, or nail symptoms.
Early-onset (type 1a) psoriasis	CDSN	Specific CDSN polymorphisms are strongly correlated with impaired desquamation in the early-onset psoriasis.
<i>Autoimmune diseases</i>	<i>Desmosomal molecules targeted</i>	<i>Clinical manifestation</i>
Pemphigus foliaceus (PF)	Dsg1	PF results in disruption of intercellular adhesion in a cell sheet leading to its acantholysis due to autoantibody against Dsg1. Endemic form of the disease is caused by an as of yet unknown environmental factor.
Pemphigus vulgaris (PV)	Dsg3	PV is a rare blistering autoimmune disease, with anti-Dsg3 antibody and the imbalance between Dsg3 specific helper and regulatory T cells being the cause of the disease.
Paraneoplastic pemphigus	DP and other plakin family members	Severe blistering of the skin and mucous membranes caused by circulating autoantibodies against plakin family proteins. Occurs mainly in the patients suffering from lymphoid malignancies, thymomas and poorly differentiated sarcomas.
IgA pemphigus	Dsg1, Dsg3, Dsc1	Milder form of pemphigus than one caused by IgGs. Neutrophil recruitment to skin by IgA against Dsgs and Dscs causes blistering, with mucosa, palms, and soles usually spared.
<i>Infectious diseases</i>	<i>Desmosomal molecules targeted</i>	<i>Clinical manifestation</i>
Bullous impetigo	Dsg1	Skin infection by group A streptococcus or <i>S. aureus</i> results in the appearance of fluid-filled blisters, usually on the trunk, arms, or the legs. Present mostly in children, spreads by touching.
Staphylococcal scalded-skin syndrome (SSSS)	Dsg1	SSSS results in loss of intercellular adhesion leading to detachment within the epidermal layer. It is caused by the exfoliative toxin produced by <i>S. aureus</i> , but unlike bullous impetigo it has systemic rather than localized effects.

(PKC) α , whose activity is necessary for the modulation of DP interaction with IF and assembly into cell–cell junctions. The importance of PKP2 for maintenance of tissue integrity is emphasized by numerous mutations correlated with congenital arrhythmias such as arrhythmogenic right ventricular cardiomyopathy (ARVC) (Table 1). PKP3 is the most recently discovered and the least studied of the three molecules. Whereas mouse knock-out model system suggests a role in hair follicle morphogenesis and regulation of inflammatory response in skin, some clinical data suggest a larger role for PKP3 in tumorigenesis of several epithelial tumors. PKPs are also engaged in a host of nonadhesive functions which are discussed below.

Plakins

Although desmosomes consist of various components in different tissues, DP, the most abundant protein in the desmosomal plaque, is an obligate desmosomal constituent. DP is the primary plakin family member found in the desmosome, although other less abundant components include envoplakin, periplakin, and plectin. Structurally, the DP molecule consists of two globular ends connected by a coiled-coil rod whose size is modulated by alternative splicing of the transcript to produce two major variants (I and II). The rod domain mediates DP self-dimerization. DP also directly associates with Pg and PKPs through its amino (N)-terminus and with IFs

through its carboxyl (C)-terminus. In addition, direct association of DP with desmosomal cadherin tails has been noted, as has an interaction with PKPs. DP is required for desmosome assembly, linking the IF cytoskeleton to the plasma membrane and anchoring IFs in the plaque. The importance of DP in maintaining tissue integrity is emphasized by the existence of inherited skin diseases of varying severity and cardiomyopathies leading to lethal arrhythmias, caused by mutations in the DP gene (Table 1). In addition to its essential role in desmosomal assembly and IF anchoring, DP is emerging as an important regulator of microtubule organization during epidermal differentiation. DP relocates ninein from centrioles to the sites of cell–cell adhesion, thereby shifting microtubules from radial to cortical organization in the differentiating skin.

Desmosome Assembly and Maintenance

Regulation of the assembly and disassembly of cell–cell junctions is important for normal tissue homeostasis and during junction remodeling processes such as development or wound healing. One crucial factor controlling desmosome assembly appears to be the formation of adherens junctions, which may be required to bring adjacent membranes close together and permit the interaction and clustering of desmosomal molecules. In cultured cells, desmosome assembly occurs in response to cell–cell contact in a calcium-dependent manner. In low calcium conditions, even though they are constantly synthesized and delivered to the plasma membrane, newly synthesized desmosomal components are unstable. Extracellular contact triggers a cascade of calcium-dependent events beginning with the accumulation of E-cadherin and the desmosomal plaque protein, PKP2. PKP2 is required for the accumulation of active RhoA at nascent junctions and associated cortical actin remodeling that drives later stages of desmosome plaque assembly. This process involves the myosin II-dependent translocation of desmosomal precursors containing DP and PKPs to reinforce the plaque. Plaque assembly is spatially and temporally coordinated with clustering of desmosomal cadherins and associated Pg, which are delivered together in vesicles to the plasma membrane. Mature desmosomes are highly insoluble and over time become independent of variations in extracellular calcium concentration. The assembly and maintenance of desmosomes are regulated in part by post-translational modification of protein components. Specifically, serine/threonine and tyrosine phosphorylation of several desmosomal components (including DP) regulates desmosomal assembly and maintenance by modulating the ability of desmosomal proteins to interact with one another.

Desmosome Function

Far from static spot welds, desmosomes are malleable structures which respond to extra- and intracellular signals to adapt their functions in cell–cell adhesion, morphogenesis, and signal transduction.

Cell–cell adhesive function

Although desmosomal cadherins were long assumed to function in calcium-dependent cell–cell adhesion, until recently there has

been a lack of direct evidence demonstrating that ectopically expressed desmosomal proteins promote adhesion in normally nonadherent cells. Evidence now suggests that both Dscs and Dsgs, along with the associated protein Pg, are required for adhesion. Autoimmune, genetic, and bacterial diseases, which affect the desmosomal cadherins and result in compromised epidermal adhesion (Table 1), support the idea that desmosomes function in intercellular adhesion. Intercellular adhesion is also disrupted by human diseases that perturb the structure of the desmosomal plaque (Table 1). Cell-culture models have demonstrated that IF attachment to the desmosome core is a critical aspect in the regulation of desmosome-mediated intercellular adhesion. This is especially important in tissues experiencing mechanical stress such as the skin and the heart, where intercellular adhesive defects can have grave consequences. In particular, several mutations in desmosomal components have been established as a major causative factor in various skin, hair, and nail abnormalities, as well as in congenital heart arrhythmias. Epidermal and heart symptoms may occur independently or in combination and vary in severity from mild to lethal (Table 1). Most notably, according to well-documented studies, ~20% of arrhythmogenic right ventricular cardiomyopathy subjects in Western Europe had a PKP2 mutation.

Morphogenetic function

Analysis of mutations of desmosomal components in animal models highlights the critical role of desmosomes in maintaining tissue integrity during embryonic development beyond the early post-implantation stage. Whereas engineered or spontaneous ablation of desmosomal cadherins, whose expression is primarily restricted to stratified epithelia, results in defects in the structure and function of skin and its appendages, elimination of more generally expressed Dsg2 leads to an early embryonic lethality immediately upon implantation, arguably due to defects in embryonic stem cell proliferation before cell–cell adhesion formation. This observation, combined with the reported pre-implantation mortality of Dsc3-deficient mice, suggests a desmosome-independent role in development for these desmosomal cadherins. DP-deficient mice (growing in the background of DP-expressing extra-embryonic tissues to prevent early adhesive defects) die early during embryogenesis due to numerous problems with heart, skin, and neuroepithelium. Conditional ablation of DP in heart and skin leads to organ-specific abnormalities and may serve as a model system for human DP-related diseases. Null mutations of the armadillo proteins Pg, PKP2, and PKP3 yield animals with lethal cardiac abnormalities (Pg and PKP2), skin fragility (Pg), and hair follicle and skin inflammation impairment (PKP3). Peptide inhibition of desmosomal adhesion has also demonstrated its necessity in normal alveolar morphogenesis by epithelial cells from mammary lumen. Normal desmosomal adhesion is also essential for polarized sorting of aggregated luminal and myoepithelial cells. Desmosome-mediated intercellular adhesion may be intimately involved in many morphogenetic events.

Signaling function

Desmosomes may also have a signaling function. A role in outside-in signaling is evidenced by the ligation of the extracellular domain of Dsg3 and Dsg1 by pemphigus vulgaris

(PV) and pemphigus foliaceus (PF) antibodies, respectively. Whereas PV antibodies sterically hinder cadherin adhesive interactions and PF antibodies do not, both can modulate the intracellular signaling including calcium concentration, phospholipid metabolism, and p38-dependent RhoA activity in keratinocytes. In addition, the antibody-mediated inactivation of RhoA was found to be a significant contributing factor in blistering caused by both autoimmune diseases, presenting itself as a potential therapeutic target. In cultured cells, ligation by PV antibody results in serine phosphorylation of Dsg3 and its reported dissociation from Pg. Pg-null keratinocytes were utilized to demonstrate that PV antibody ligation of Dsg3 leads to a Pg-dependent keratin retraction which may contribute to the epidermal blistering characteristic of this disease.

Intracellular signals can also be propagated by desmosomal components; for example, Pg has been shown to negatively regulate the activity of Src kinase and RhoGTPase family. Although Pg can be found in different junction types, the metabolic stability and cellular localization of Pg are tightly regulated, allowing it to function in cell–cell adhesion as well as in the regulation of cell–substrate interactions. Pg performs these functions through regulation of RNA and protein levels of a variety of extracellular matrix proteins, their integrin receptors, and other cell–substrate adhesion-related molecules. The growing list of regulatory roles of Pg includes transcriptional activation, proliferation, and programmed cell death.

Emerging data suggest the role for other desmosomal molecules in the regulation of intracellular signaling, including functional interactions between PKP2 and PKC α as well as the role of Dsg1 in the inhibition of EGFR/Erk signaling. Intriguingly, PKP1 and PKP3 are found in complexes with RNA-binding proteins as well as in stalled translation initiation complexes (stress granules) and, in the case of PKP1, are shown to directly stimulate

translation through activation of eIF4A1 translation initiation factor. These exciting findings point toward a broad regulatory role of desmosomal component in many critical processes in the cell beyond their function as adhesive molecules.

The Hemidesmosome

Hemidesmosomes are specialized junctions that connect basal epithelial cells to the basement membrane and underlying extracellular matrix. Hemidesmosomes promote the integrity and strength of epithelial tissues by indirectly connecting the intracellular keratin IF cytoskeleton to extracellular matrix proteins.

Hemidesmosome Ultrastructure

The ultrastructure of the hemidesmosome (Figure 2, left) includes an electron-dense cytoplasmic triangular plaque, which comprises an inner IF-rich region closest to the cell cytoplasm and a perimembrane plaque containing the cytoplasmic tails of the transmembrane hemidesmosome components. Just adjacent and external to the basal cell membrane is a sub-basal dense plate and thin, extracellular anchoring filaments that extend from the plate into the specialized extracellular matrix of the basement membrane.

Structure and Molecular Components of the Hemidesmosome

The major hemidesmosomal components together connect the basal epithelial cell cytoskeleton to the extracellular matrix (Figure 2, right). Hemidesmosomal proteins are arranged such that IFs are anchored to the plaque at multiple points, re-enforcing the connection to the substratum.

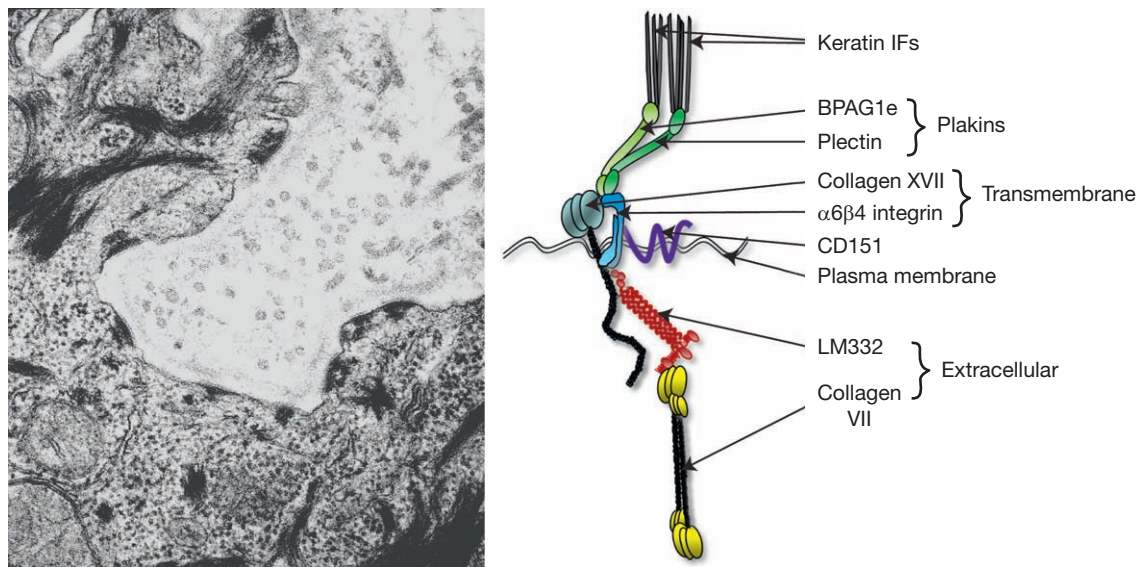


Figure 2 Molecular composition and structure of the hemidesmosome. (Left) Hemidesmosome ultrastructure is shown in this electron micrograph of bovine tongue epidermis. (Right) Hemidesmosomal proteins, their interactions, and their approximate position within the organelle are depicted in this schematic diagram of the hemidesmosome. Reproduced from Jones JCR, Yokoo KM, and Goldman RD (1986) Is the hemidesmosome a half desmosome? An immunological comparison of mammalian desmosomes and hemidesmosomes. *Cell Motility and the Cytoskeleton* 6: 560–569.

$\alpha 6 \beta 4$ integrin

The $\alpha 6 \beta 4$ integrin heterodimer is essential for hemidesmosome formation and maintenance of epithelial cell attachment to the basement membrane. The $\alpha 6 \beta 4$ integrin is a receptor for the matrix protein laminin-332, and its interaction with laminin-332 is thought to be the structural center of the hemidesmosome. The $\alpha 6 \beta 4$ integrin heterodimer is expressed predominantly in epithelial tissues, Schwann cells, and endothelial cells. The $\beta 4$ integrin subunit has an unusually long cytoplasmic tail, which is important for the specialized functions of this molecule and for interaction with the other hemidesmosomal proteins plectin, type XVII collagen, and BPAG1e. The interaction with plectin is thought to stabilize the association of $\alpha 6 \beta 4$ integrin with laminin-332 to enhance adhesion strength.

Type XVII collagen (BPAG2, BP180)

Type XVII collagen is a 180-kDa type II transmembrane protein component of the hemidesmosome. It was originally named bullous pemphigoid antigen II (BPAG2) or BP180 because autoantibodies from the serum of patients with the severe blistering disease bullous pemphigoid recognize and bind to this protein (Table 2). Type XVII collagen is a trimer that contains collagen-like repeats in its C-terminal extracellular domain which organize into a triple helical coiled coil. Type XVII collagen interacts with the $\beta 4$ integrin subunit and the hemidesmosomal protein BPAG1e through its N-terminal domain, and with the $\alpha 6$ integrin subunit via its C-terminal extracellular domain. The latter interaction seems to be

essential for hemidesmosome formation and stabilization. Although type XVII collagen can bind almost all the proteins within the hemidesmosome, its major function appears to be recruitment of BPAG1e to the hemidesmosome. Type XVII collagen is also known to interact with actinin family members and to co-distribute at sites of cell–cell contact with adherens junction proteins, suggesting a potential role for this molecule in mediating cross-talk between cell–substrate and cell–cell junctions.

BPAG1 (BP230)

Bullous pemphigoid antigen I (BPAG1 and BP230) is a 230-kDa plakin family member that, like type XVII collagen, was first identified as a protein recognized by autoantibodies of patients with bullous pemphigoid. Four main splice isoforms of BPAG1 have been identified. The epithelial specific form (BPAG1e) is a cytoplasmic plaque component of the hemidesmosome. BPAG1e is known to interact with IFs through its C-terminus, whereas the N-terminal domains interact with the $\beta 4$ integrin subunit and type XVII collagen. BPAG1e is thought to link the keratin IFs to type XVII collagen and the hemidesmosomal plaque. Skin isolated from BPAG1-null mice has poorly formed hemidesmosomes with few filament attachments.

Plectin

Plectin, also referred to as IFAP 300 or HD1, is a 500-kDa plakin family member. It is particularly important for hemidesmosome function in epithelial tissues, and as a cytoplasmic

Table 2 Human disease associated with hemidesmosomal components

<i>Genetic diseases</i>	<i>Hemidesmosomal molecules mutated</i>	<i>Clinical manifestation</i>
Herlitz or non-Herlitz junctional epidermolysis bullosa (JEB)	Laminin-332, $\beta 4$ integrin	Herlitz: Autosomal recessive: severe blistering at the dermal–epidermal junction often coincident with early mortality. non-Herlitz: Less severe form, often includes lifelong blistering, but improves with age, loss of hair and nails, and dental abnormalities.
Junctional epidermolysis bullosa with pyloric atresia (JEB-PA)	$\alpha 6$ or $\beta 4$ integrin	Mucocutaneous blistering associated with congenital intestinal abnormalities.
Dystrophic epidermolysis bullosa (DEB)	Collagen VII	The most severe form of epidermolysis bullosa characterized by deep blistering in dermis due to disruption of anchoring fibril stability.
Non-Herlitz JEB/generalized atrophic benign epidermolysis bullosa (GABEB)	BPAG2 (BP180)	Autosomal recessive generalized skin blistering with mild scarring, nail dystrophy, enamel hypoplasia, pigmentary changes, and hair loss.
Epidermolysis bullosa with muscular dystrophy (EB-MD)	Plectin	Skin blistering noted from birth with late-onset, progressive muscle weakness.
<i>Autoimmune diseases</i>		
	<i>Hemidesmosomal molecules targeted</i>	<i>Clinical manifestation</i>
Bullous pemphigoid	BPAG1, BPAG2	Generalized large subepidermal blisters caused by circulating autoantibodies against Type XVII Collagen (BPAG2, BP180) and BPAG1e (BP230).
Cicatricial pemphigoid	Laminin-332, BPAG2	Chronic subepithelial blistering of mucous membranes and, occasionally, of the skin. Caused by circulating autoantibodies against laminin-332 or type XVII collagen.
Ocular cicatricial pemphigoid	$\beta 4$ integrin	Subset of mucous membrane pemphigoid. Subepithelial blistering that mainly affects conjunctiva more than other mucous membranes. Leads to ocular keratinization, corneal scarring and blindness. Caused by circulating autoantibodies against $\beta 4$ integrin.
Oral Pemphigoid	$\alpha 6$ integrin	Subset of cicatricial pemphigoid. Blistering limited to oral cavity. Caused by circulating autoantibodies against $\alpha 6$ integrin.
Epidermolysis bullosa acquisita	Collagen VII	Minor skin trauma induces severe blistering below the dermal–epidermal junction. Caused by circulating autoantibodies against collagen VII.

cytoskeletal linker in other tissues, but has also been reported in less abundance in desmosomes. Plectin's importance for hemidesmosome stability and tissue integrity is emphasized in patients with genetic mutations in the plectin gene who suffer from severe skin blistering and muscular dystrophy (Table 2). Plectin can interact with itself, with IFs, and with multiple domains of the $\beta 4$ integrin tail. Plectin has been proposed to cluster $\alpha 6\beta 4$ molecules at the basal cell surface, a potentially critical step in the formation of hemidesmosomes. However, the primary role for plectin (like BPAG1) is to link the IF system to the hemidesmosome plaque, stabilizing the junction by connecting the IF cytoskeleton to the hemidesmosome core at multiple points.

CD151

CD151, a member of the tetraspanin family, has been recently identified as a component of the hemidesmosome. The $\alpha 6$ integrin subunit and CD151 interact directly, and, thus, CD151 appears concentrated at hemidesmosomes. The exact function of CD151 as a hemidesmosomal component requires further investigation, since the skin of mice lacking CD151 contains normal hemidesmosomes.

Laminin-332

Laminin-332 (previously laminin-5) is a member of the laminin family of large, heterotrimeric glycoproteins, which are major constituents of extracellular matrices. Laminin-332 is necessary for the firm attachment of basal epithelial cells to the subepithelial basement membrane by linking the hemidesmosome to the underlying tissue. In the hemidesmosome, laminin-332 binds to $\alpha 6\beta 4$ integrin to mediate cell–substrate attachment. Laminin-332 is highly expressed in many types of epithelia and is specifically composed of subunits $\alpha 3$, $\beta 3$, and $\gamma 2$, which interact in a noncovalent fashion to form a cross-shaped structure. The subunits of laminin-332 undergo proteolytic processing after the molecule is secreted into the matrix. Yet, laminin-332 is initially secreted in an unprocessed form. This precursor laminin-332 is thought to promote cell migration, whereas cleavage of the $\alpha 3$ subunit promotes hemidesmosome nucleation and stable adhesion in an $\alpha 6\beta 4$ integrin-dependent manner. Laminin-332 is also known to mediate epithelial cell attachment through the nonhemidesmosomal integrin $\alpha 3\beta 1$. Laminin-332 strongly attaches epithelial cells to the extracellular matrix by interacting with other basement membrane proteins, including type VII collagen. Genetic or autoimmune diseases involving laminin-332 are characterized by complete separation of the epidermis from the dermis (Table 2) and further support an important role for laminin-332 in the attachment of basal epithelial cells to the underlying basement membrane.

Hemidesmosome Assembly and Maintenance

The nucleation and assembly of hemidesmosomes depend upon the phosphorylation state of $\alpha 6\beta 4$ integrin and laminin-332. It has been suggested that laminin-332 induces clustering of $\alpha 6\beta 4$ integrin, followed by recruitment of plectin, BPAG1e, and type XVII collagen to the nucleation site. In addition, clustering of the $\alpha 6\beta 4$ integrin has been shown to promote phosphorylation of the $\beta 4$ integrin subunit

cytoplasmic tail and hemidesmosome assembly. The $\alpha 3\beta 1$ integrin is also suggested to be an initiator of hemidesmosome assembly, by forming pre-hemidesmosome clusters of CD151, $\alpha 3\beta 1$ integrin, and laminin-332. These clusters have been proposed to serve as docking sites for $\alpha 6\beta 4$ integrin.

Both bullous pemphigoid antigens are also thought to be important players in regulating hemidesmosome assembly. Patients with mutations in the type XVII collagen gene have normal $\alpha 6\beta 4$ integrin and laminin-332 localization but absent or underdeveloped hemidesmosomes. In addition, overexpression of BPAG1e in cells alters the localization of endogenous type XVII collagen, suggesting a role for BPAG1e in initiating early interaction of some hemidesmosome proteins prior to junctional incorporation.

Laminin-332 also contributes to the regulation and maintenance of hemidesmosomes. Processing of laminin-332 promotes hemidesmosome formation, as mentioned above. For adhesion, laminin-332 appears to initially interact with $\alpha 3\beta 1$ integrin to mediate attachment of the epithelial cells to the extracellular matrix. Ligation is then transferred to $\alpha 6\beta 4$ integrin for long-term, stable adhesion.

Hemidesmosome Function

Hemidesmosomes were long thought to be inert adhesive structures with primary importance in maintaining and stabilizing epithelial cell attachment to the underlying basement membrane. However, recent evidence suggests that the proteins that comprise these structures exhibit dynamic characteristics that may facilitate tissue remodeling processes during development and wound healing as well as permit rapid responses to extracellular signals.

Cell–substrate adhesive function

A structural function for hemidesmosomes in initiating and maintaining adhesion between basal epithelial cells and the extracellular matrix is well accepted today, in part because of evidence from autoimmune and genetic human skin diseases which result in epithelial cell–basement membrane separation and severe blister formation (Table 2). Other evidence supporting this structural role comes from the noted absence of hemidesmosomes in cells closing wounds and in invasive epithelial tumor cells. Thus, hemidesmosomes appear necessary for stable anchorage of cells to the basement membrane but may not be required in motile cells.

Morphogenetic function

The role that hemidesmosomes play in morphogenetic events such as breast epithelial tubule formation indicates the relative dynamic nature of these structures *in vivo*. Indeed, perturbation of the microfilament cytoskeletal architecture in cultured cells results in a redistribution of IF-tethered hemidesmosomes to the cell periphery, highlighting the dynamic nature of the hemidesmosome plaque and its ability to move laterally within the cell membrane.

Transmembrane signaling function

The proteins that comprise the hemidesmosome are important mediators of signal transduction; however, they activate signaling cascades when not assembled into a complete

hemidesmosome adhesion device. Furthermore, $\alpha 6 \beta 4$ integrin, the hemidesmosomal protein mainly involved in mediating signaling, can act in a ligand-dependent and -independent manner. The $\alpha 6 \beta 4$ integrin appears to be able to activate a number of signaling cascades, including the phosphoinositide 3-OH kinase (PI3K), mitogen-activated protein kinase (MAPK), c-jun N-terminal kinase (JNK), and small guanosine phosphatase (GTPase) Rac pathways. Therefore, in addition to promoting cell adhesion, the $\alpha 6 \beta 4$ integrin functions to promote cell migration, invasion, proliferation, and survival. Phosphorylation of specific tyrosine residues within the large cytoplasmic tail of the $\beta 4$ integrin subunit leads to activation of the distinct signaling pathways and subsequent cellular processes.

Although it is well accepted that $\alpha 6 \beta 4$ integrin and its ligand laminin-332 are essential for stable anchorage of cells to the extracellular matrix, there is emerging evidence that they also mediate the migration of cells – this is controversial. Nonetheless, studies investigating wound healing or tumor cell migration indicate that $\alpha 6 \beta 4$ integrin is localized at the leading edge of the lamellipodium, where it signals to activate Rac1 and the actin-severing protein cofilin. In addition, laminin-332 has been shown to be deposited at the leading front of migrating cells, suggesting that it provides a track upon which the cell can move. This mechanism is proposed to be $\alpha 6 \beta 4$ integrin dependent. Although direct evidence for $\alpha 6 \beta 4$ integrin as the main mediator of cell migration is lacking, several models have been proposed in which $\alpha 6 \beta 4$ integrin serves a dual function, that is, one of stable attachment in the confines of the hemidesmosome and the other as a mediator of cell migration when the hemidesmosome is disassembled.

Despite their ultrastructural and functional similarity as adhesive junctions, desmosomes and hemidesmosomes are unique structures. Significant progress has been made thus far in the fields of desmosome and hemidesmosome research. The structural organization of these junctions has been examined. Molecular components that comprise these junctions have been identified. Protein–protein interactions that contribute to the normal function of these organelles have been characterized. Future studies in these fields will likely focus on a growing array of nonadhesive functions for protein components of these junctional complexes, and their underlying signaling cascades.

Acknowledgments

Work in the Jones and Green labs is supported by grants from the National Institutes of Health, R01 AR043380 (K.J.G.), R01 AR041836 (K.J.G.), R01 CA122151 (K.J.G.), R01 HL092963 (J.C.R.J.), R01 AR054184 (J.C.R.J.), and the LeDucq Foundation (K.J.G.).

See also: [Cell Architecture and Function: Actin Organization; Cadherin-Mediated Cell–Cell Adhesion; Intermediate Filament Linker Proteins: Plectin and BPAG1; Intermediate Filaments; Keratins and the Skin; Rho GTPases and Actin Cytoskeleton Dynamics; Signaling: Cadherin Signaling; Integrin Signaling.](#)

Further Reading

- Bannon LJ, Goldfinger LE, Jones JCR, and Green KJ (2001) Desmosomes and hemidesmosomes. In: Beckerle MC (ed.) *Cell Adhesion*, 1st edn., vol. 39, pp. 324–368. New York, NY: Oxford University Press.
- Bass-Zubek AE, Godsel LM, Delmar M, and Green KJ (2009) Plakophilins: Multifunctional scaffolds for adhesion and signaling. *Current Opinion in Cell Biology* 21: 708–716.
- Garrod DR, Merritt AJ, and Nie Z (2002) Desmosomal cadherins. *Current Opinion in Cell Biology* 14: 537–545.
- Green KJ and Gaudry CA (2000) Are desmosomes more than tethers for intermediate filaments? *Nature Reviews Molecular and Cellular Biology* 1: 208–216.
- Green KJ and Simpson CL (2007) Desmosomes: New perspectives on a classic. *Journal of Investigative Dermatology* 127: 2499–2515.
- Jamora C and Fuchs E (2002) Intercellular adhesion, signaling, and the cytoskeleton. *Nature Cell Biology* 4: E101–E108.
- Jones JCR, Hopkinson SB, and Goldfinger LE (1998) Structure and assembly of hemidesmosomes. *BioEssays* 20: 488–494.
- Kligys K, Hamill K, and Jones JCR (2008) Hemidesmosomes and their components: Adhesion versus signaling in health and disease. In: LaFlamme SE and Kowalczyk AP (eds.) *Cell Junctions: Adhesion, Development and Disease*, pp. 109–133. Weinheim: Wiley-VCH.
- Leung CL, Green KJ, and Liem RK (2002) Plakins: A family of versatile cytolinker proteins. *Trends in Cell Biology* 12: 37–45.
- Nievers MG, Schaapveld RO, and Sonnenberg A (1999) Biology and function of hemidesmosomes. *Matrix Biology* 18: 5–17.
- Nishiyama T, Amano S, Tsunenaga M, et al. (2000) The importance of laminin-5 in the dermal–epidermal basement membrane. *Journal of Dermatological Science* 24(supplement 1): S51–S59.
- Uitto J and Pulkkinen L (2001) Molecular genetics of heritable blistering disorders. *Archives of Dermatology* 137: 1458–1461.

Detergent Properties

D R McCaslin, University of Wisconsin, Madison, WI, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Aggregation number Number of detergent molecules in the average micelle.

Critical micelle concentration (CMC) Concentration of detergent in aqueous solution above which micelles begin to form, essentially the maximum concentration of detergent in monomeric form.

Detergent A compound having spatially segregated hydrophilic and hydrophobic regions and which, when dissolved in water above the CMC, self-associates to form micelles.

Hydrophilic group Structure having an affinity for water with strong favorable noncovalent bonds with water molecules.

Hydrophobic group Structure exhibiting an aversion to water and preference for hydrocarbon-type liquids.

Micelle Structure formed by the noncovalent and highly cooperative self-association of a detergent molecule so as to form a hydrophobic core from which the hydrophilic groups project into the aqueous surroundings.

A molecule can be classified as a detergent when within its chemical skeleton one can identify spatially segregated regions, one with an affinity for water (a hydrophilic group) and one which, on its own, would have limited water solubility (a hydrophobic group). For well-behaved detergents in aqueous solution, this molecular polarity drives the formation of a mixture consisting of monomeric detergent molecules in equilibrium with a fairly monodisperse population of detergent aggregates known as 'micelles'. In the micelle, the hydrophobic groups are packed together to create a hydrophobic core with the attached hydrophilic groups projecting out from the surface of this core and protecting it from contact with water. To date, only limited biological functions for detergents have been identified such as the well-known role in digestion and possibly, to some extent, in membrane fusion events; however, they play necessary if not essential roles in the isolation, manipulation, and characterization of the constituents of biological membranes.

Chemical Structure of Detergents

The chemical structures of detergents which have been employed in biological studies are quite varied and continue to grow as investigators attempt to develop new entities that will permit the facile isolation of membrane constituents while enhancing their stability once isolated. To reach this goal, various combinations of hydrophobic and hydrophilic groups have been utilized which yield a variety of physical and chemical properties.

Hydrophobic Groups (Tails)

Simple hydrocarbon chains

The hydrocarbon chain is the most easily recognized hydrophobic group and when present is often referred to as the molecule's 'hydrophobic tail' (Figure 1). These chains are most often saturated hydrocarbons (Figure 1) and are

available in many lengths. The shorter the chain the less well defined the detergent properties tend to be, whereas longer tails become essentially insoluble. Unsaturation and more complex branching structures have been explored, as well as chains incorporating phenyl rings. The presence of the ultraviolet light-absorbing phenyl ring should generally be avoided as it can complicate various spectroscopic techniques.

Ring-based hydrocarbons

The naturally occurring bile salts synthesized from cholesterol and utilized in digestive processes are the prototypes for detergents based on ring systems. Figure 1 shows cholic acid, one of the abundant bile salts which is converted to deoxycholate (DOC) by removing the hydroxyl at the arrow. Identifying the spatial segregation of the hydrophilic and hydrophobic groups in these detergents may require careful inspection of their three-dimensional chemical structure. In cholic acid, the carboxyl group is an obvious hydrophilic group, but the effects of the hydroxyl groups located on the rings must also be considered. Viewing the steroid ring nucleus in the molecule as defining a plane, the spatial arrangement of the hydroxyls creates a hydrophilic face on one side and a hydrophobic face on the other side of this plane. The structure of DOC suggests the presence of only a hydrophilic edge along the ring plane, making the whole structure more hydrophobic than cholic acid, which may explain why DOC is generally a more aggressive solubilizer than cholic acid.

Hydrophilic Groups (Headgroups)

Ionic

Positively (e.g., amino) and negatively (e.g., carboxyl) charged groups have been utilized as hydrophilic groups. Single-tailed ionic detergents tend to denature all proteins, with the anionic ones being more aggressive denaturants than cationic molecules. One of the most familiar examples is the anionic detergent, sodium dodecyl sulfate, NaDodSO₄ (Figure 1). It is the denaturing effect of NaDodSO₄ on proteins that led to the

used when relatively freshly prepared; moreover, when using commercial sources one should be aware that antioxidants are sometimes included.

Connecting Head and Tail

Linkages between the hydrophobic and hydrophilic groups are most often either an ether or ester linkage. The latter can be susceptible to hydrolysis, especially in biological preparations, which in turn could generate an ionic group with denaturing properties.

Detergent Properties

Critical Micelle Concentration

The most common concentration-dependent behavior for detergents is illustrated in **Figure 2**. The monomer concentration increases until the critical micelle concentration (CMC) above which an equilibrium is established between the monomer and an increasing concentration of micelles. For well-behaved detergents, the monomer concentration increases very little after reaching the CMC and thus is in effect the highest possible concentration of monomeric detergent. All detergents in excess of the CMC are incorporated into micelles (dashed line in **Figure 2**). The concentration of micelles increases linearly with total detergent concentration above the CMC, but at a slower rate since a large number of molecules are incorporated into each micelle. Some detergents exhibit more complex behavior, such as secondary association of micelles or even new phases as the total concentration continues to increase.

As illustrated in **Figure 2**, the CMC is actually a narrow range of concentration over which the formation of micelles becomes dominant. Nonetheless, a single number is usually defined as the CMC and determined by plotting some physical observable whose response differs markedly above and below the CMC against the total concentration of detergent. The CMC is then taken as the intersection of lines extrapolated from the nearly linear regions above and below the transition as shown

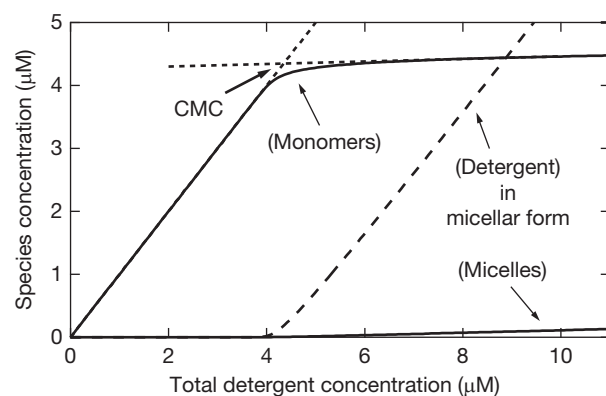


Figure 2 Concentration-dependent behavior of a detergent. The calculations are based on the theory of micelle formation developed by Tanford. An aggregation number of 50 was used and the equilibrium constant for micellization was chosen to yield the critical micelle concentration (CMC) indicated.

by the dotted lines in **Figure 2**. Using different observables or concentration ranges for the extrapolation can lead to somewhat different values for the CMC. One of the simplest methods to measure a CMC is the spectrophotometric monitoring of the solubilization of heme, which is essentially insoluble below the CMC and increases as the concentration of micelles increases; the increase in absorbance with detergent concentration would be similar to the dashed line in **Figure 2**.

CMCs range from nanomolar to several millimolar. For single-tailed detergents, the CMC decreases with increasing length of tail and with smaller headgroups. The CMC of nonionics is generally lower than that of ionics with the same length of tail. Ionic headgroups are more susceptible to solution variables than are nonionics. Ionic headgroups that can be titrated are sensitive to pH and may precipitate when the headgroup is in neutral form (both cholic acid and DOC are examples). The CMCs of ionics may also be influenced by the type and concentration of counterions present, for example, using potassium instead of sodium for dodecyl sulfate results in an insoluble salt. In general, higher salt concentrations will decrease the CMC of an ionic detergent which is a consequence of reducing the headgroup repulsion in the micelle. Even the POE-based nonionics may be influenced by presence of salts as the ether groups are capable of at least weakly complexing ions. The solubilization of other components in the micelle (e.g., mixed micelles with lipids) will generally decrease the CMC (i.e., the maximum monomer detergent concentration is decreased). Published values of CMCs are useful guides in experimental design, but should generally be confirmed under the actual experimental conditions.

Aggregation Number

Micelles are formed as a result of a highly cooperative self-association process in which many detergent molecules form a single noncovalent structure. While the number of molecules in each micelle is somewhat variable, the average number of molecules per micelle (aggregation number) can be determined by standard hydrodynamic methods (e.g., sedimentation equilibrium). The aggregation number is needed to calculate the micelle concentration, which is the total detergent concentration in excess of the CMC divided by the aggregation number. This allows one to be certain that the number of micelles to protein chains is high enough so as to minimize the possibility of adventitious association between protein chains.

The CMC transition becomes sharper as the aggregation number increases. The CMC is generally strongly dependent on the size of the hydrophobic moiety, but repulsion between headgroups is the dominant factor for the aggregation number. Thus, nonionics tend to have larger aggregation numbers than ionics of the same tail length. The aggregation number for ionic detergents can be strongly influenced by both the type and concentration of counterions present as well as pH since these factors can dramatically change the electrostatic repulsion between the charged headgroups. Longer hydrophobic tails tend to have larger aggregation numbers. Detergents with high CMCs tend to have more variability in their micelle size distribution. At high concentrations and under certain solution conditions (e.g., elevated temperature), larger

structures may be formed which in some cases are new micellar phases and in others simply secondary aggregation of smaller micelles.

Temperature Effects

As a thermodynamic equilibrium, the formation of micelles can be influenced by temperature. At low enough temperatures, solid detergent will exist in equilibrium with monomeric detergent; as the temperature increases, the monomer concentration increases until it reaches the CMC at the critical micelle temperature. Above this temperature, solid detergent will begin to go into solution as micelles. The temperature at which solid, micelles, and monomeric detergent at the CMC coexist is called the 'Kraft point' and for most detergents is the same as the critical micelle temperature. An often-observed Kraft point, which is near room temperature, is that of sodium dodecyl sulfate where upon cooling one sees precipitation of detergent. A second temperature effect observed at higher temperature, especially with nonionics containing POE, is the cloud point. Above the CMC at this temperature, the solution will turn turbid due to the formation of much larger aggregates. Both the critical micelle temperature and cloud point have been exploited in purification of membrane components.

Micelle Structure

Single-Tailed Detergents

In a typical micelle, the hydrophobic tails are sequestered into a central core structure with the hydrophilic groups projecting out from the surface of this core into the aqueous surrounding. The tails in the core are not fully extended but are quite flexible, some even lying along the surface of the core rather than within it. The surface of the core should be regarded as having a rippled texture with some tail methylene groups protruding above it. Experimental and theoretical arguments suggest that for most detergent micelles the overall shape is best described as an oblate ellipsoid, although for small aggregation numbers the shape cannot be readily distinguished from spherical. For $C_{12}E_8$ micelles (Figure 1), theoretical calculations constrained by hydrodynamic measurements yield an oblate ellipsoid with dimensions shown in Figure 3. Since the core cannot contain a void, one dimension must be less than the 3.4 nm length of two fully extended 12C chains. For the same length of tail, the overall dimensions of the micelle will depend on the nature of the headgroup. For POE-based detergents, the headgroup is in a random-coil configuration and consequently occupies a very large region of the space surrounding the core, as suggested in Figure 3.

Figure 3, while a convenient visualization of micelle structure, cannot convey the highly dynamic processes occurring. Detergent molecules are rapidly exchanged among micelles; whole micelles disappear and reform with slightly different aggregation numbers. The hydrophobic core acts for the most part like a simple liquid hydrocarbon with the tails constantly flexing. POE headgroups occupy a great deal of space around the core and are constantly changing their conformation as well. Finally, it should be obvious that the type and size of headgroup can generate dramatically different chemical

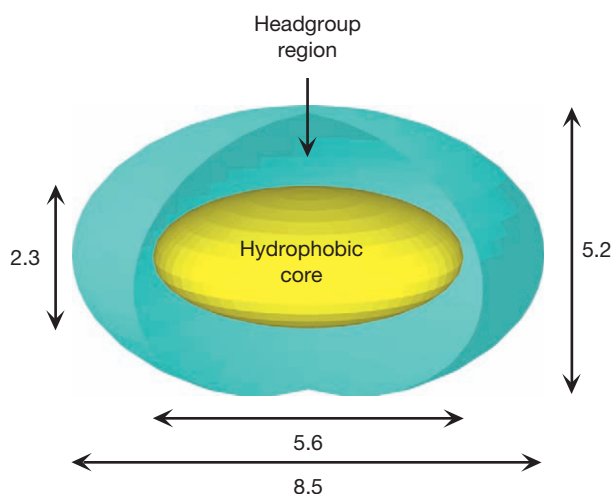


Figure 3 Shape of micelles formed by typical single-tailed detergents. The model shown is based on data and calculations for $C_{12}E_8$. The dimensions shown are in nm.

environments to which molecules solubilized by detergents are exposed; moreover, the solution composition near a micelle may be quite different from that of the bulk aqueous solution.

Ring-Based Detergents

Bile salts usually have relatively high CMCs and quite small aggregation numbers (4–10). The small aggregation number does not permit micelles as just described. The micelles are more heterogeneous and are probably best viewed as small assemblies with the hydrophobic surfaces facing each other, presenting their hydrophilic faces to the surrounding solution. The size and shape of these micelles are quite sensitive to the concentration and types of ions, and potentially pH.

Detergents as Tools in Membrane Studies

Membrane Solubilization

Initially, when a membrane is exposed to detergent, monomers partition into the bilayer, so one can imagine that a high CMC might be advantageous (e.g., OG has a very high CMC (>15 mM) and is an efficient solubilizer). As more detergent intercalates into the membrane bilayer, one reaches a stage where fragmentation occurs, resulting in large mixed micelles containing detergent, lipids, and proteins in various ratios. Continuing to add detergent will eventually disperse the membrane components to the point that any given micelle contains no more than a few lipids, a protein, or the detergent alone. One can apply isolation and enrichment techniques for specific targets to this solubilized mix. The choice of detergent can hinder or help in isolation, for example, a small aggregation number can facilitate separation of a large protein-containing micelle from those containing only lipid components. It is possible, and perhaps even desirable, to exchange detergents, so that one detergent might be used to speed initial membrane

solubilization and another for final isolation of a stable target. It is important to recognize that it is the concentration of micelles that needs to be controlled when one is studying membrane components solubilized by detergents, and generally one should maintain a ratio of several micelles to each solubilized component to avoid the possibility of spurious associations.

Membrane Proteins

A detergent can be found that can reasonably mimic the properties of the hydrophobic core of a native membrane environment. However, the membrane itself provides a physical constraint to a protein embedded within it, in that the protein cannot expand laterally in the plane of the membrane nor can it pull itself through the bilayer; in a detergent solution, such a physical constraint is greatly relaxed and the protein in effect has a much expanded conformational space including conformations where activity will be lost possibly irreversibly. Low CMCs and large aggregation numbers may partially compensate for the loss of the membrane's physical constraints. In addition, the native membrane environment provides a multitude of diverse interactions with a variety of lipid headgroups, tails, and with other proteins and, moreover, these interactions may differ from one side of the membrane to the other. The roles of these types of interactions in maintaining an active protein are not well understood, and they are radically changed or even eliminated upon solubilization. Studies as a function of micelle concentration should be performed, as there is always a possibility of adventitious interactions, due merely to the fact that too few micelles are present.

Reconstitution of Protein into a Bilayer

Many of the functions of membrane proteins are vectorial in nature, and one must for full characterization eventually reconstitute the protein into a lipid bilayer where these functions can be probed. Starting with mixed micelles of protein in detergent and lipids in detergent, one must effect a controlled removal of the detergent, permitting the formation of an artificial membrane bearing the protein. This is accomplished largely by trial and error. A high CMC permits the use of dialysis since the monomer should pass through the dialysis membrane. Hydrophobic beads to which the detergent absorbs have also proven effective in reconstitution studies.

Summary

Obviously single-tailed ionic detergents should generally be avoided because of their potential to denature any protein. While zwitterionic headgroups can be identical or at least similar to those of native lipids, to date they have not proven any more useful than others. High CMC detergents may be better choices for initial solubilization and eventual reconstitution. But long tails and large headgroups (usually accompanied by a low CMC) may enhance stability by restricting the conformational space accessible to the protein while in the detergent solution. The requirements for efficient solubilization and for the maintenance of activity may to a large extent be antagonistic. The possibility of using one detergent to solubilize and subsequently exchanging into another to enhance stability for isolation and study is always worth consideration. The thermodynamic interactions resulting from the dissolution of a protein in a detergent solution are complex and require great care when evaluating stability and functional data. One must recognize that the thermodynamic properties of all the components of the solution (protein, detergent, buffer, salts, etc.) are inextricably linked together (via the Gibbs–Duhem relationship) so that a simple change in salt concentration may appear to have altered a protein's stability, but in reality did so only indirectly via a change in some property of the detergent. Unfortunately, there is no perfect detergent for all situations, nor are there hard-and-fast rules for choosing a detergent for a specific task. In the end, the choice is still largely a matter of trial and error.

See also: **Bioenergetics:** Membrane Transport, General Concepts; **Lipids Carbohydrates Membranes and Membrane Proteins:** Multidrug Resistance Membrane Proteins.

Further Reading

- Gravito RM and Ferguson-Miller S (2001) Detergents as tools in membrane biochemistry. *Journal of Biological Chemistry* 276: 32403–32406.
- Helenius A, McCaslin DR, Fries E, and Tanford C (1979) Properties of detergents. *Methods in Enzymology* 56: 734–749.
- Helenius A and Simons K (1975) Solubilization of membranes by detergents. *Biochimica et Biophysica Acta* 415: 29–79.
- Tanford C (1980) *The Hydrophobic Effect: Formation of Micelles and Biological Membrane*. New York: Wiley.
- Tanford C and Reynolds JA (1976) Characterization of membrane proteins in detergent solutions. *Biochimica et Biophysica Acta* 457: 133–170.
- White SH, Ladokhin AS, Jayasinghe S, and Hristova K (2001) How membranes shape protein structure. *Journal of Biological Chemistry* 276: 32395–32398.

Diabetes

S M Blackman and D W Cooke, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by David W. Cooke, volume 1, pp. 582–592, © 2004, Elsevier Inc.

Glossary

β -Cell The insulin-producing cells in the Islets of Langerhans.

Glucose disposal The removal of glucose from the blood, usually in the context of its removal after having been added, as after absorption of a carbohydrate-containing meal.

Glucose intolerance The condition where the concentration of glucose in the blood after ingestion of carbohydrate is elevated above normal; it may also refer more generally to a condition associated with hyperglycemia.

Hyperglycemia An abnormally elevated concentration of glucose in the blood.

Kinase An enzyme that modifies a substrate through the addition of a phosphate group.

Osmotic diuresis An increase in urine output stimulated by an increase in osmotically active solute in the urine; in diabetes, the solute is glucose.

Phosphatase An enzyme that modifies a substrate through the removal of a phosphate group.

Prandial Meal associated.

Introduction

The defining feature of diabetes mellitus is an abnormal elevation of the blood glucose level. When the degree of hyperglycemia is mild, there may be no symptoms of diabetes. With greater degrees of hyperglycemia, the concentration of glucose in the blood exceeds the ability of the kidneys to completely reabsorb the glucose delivered to it. The resultant loss of glucose in the urine provides an osmotic force that draws increased water into the excreted urine, causing the patient to experience polyuria (increased urination; this glucosuria gives diabetes mellitus its name, which can be translated from the Greek as 'sweet urine' or more directly as 'honey siphon'). The osmotic diuresis from glucosuria induces polydipsia (increased drinking) in order to prevent dehydration. When uncontrolled, diabetes mellitus puts patients with these diseases at risk of acute metabolic decompensation, particularly when the insulin deficiency is complete, as in type 1 diabetes mellitus. While this decompensation is life threatening, since the introduction of insulin treatment over 80 years ago these events are generally preventable. It is now the long-term complications that all patients with diabetes mellitus are at risk that has become the most significant burden. Due to these long-term complications, diabetes is a leading cause of blindness and end-stage renal disease. In addition, diabetes is associated with accelerated atherosclerotic vascular disease, leading to a risk of myocardial infarction, stroke, and limb amputation that is increased and occurs at a younger age than in those without diabetes.

Etiology

For each type of diabetes mellitus, the relative contribution of a defect in insulin secretion versus a defect in insulin action will vary. For example, the two most common forms of diabetes are type 1, which is due to an isolated deficiency in insulin, and

type 2, which is due to varying degrees of insulin resistance and deficient insulin secretion. In addition, some forms of diabetes have a predominantly genetic etiology, such as the group of autosomal dominantly inherited forms of diabetes known collectively as maturity-onset diabetes of the young (MODY), whereas in other types both genetic and environmental factors are important, exemplified by type 2 diabetes, where the development of obesity is a significant contributor to the cause of insulin resistance that is central to the pathophysiology of this disease.

Type 1 Diabetes Mellitus

Type 1 diabetes accounts for ~5–10% of all cases of diabetes. It typically has its onset in childhood, although it can occur at any age. Type 1 diabetes is due to the destruction of the insulin-producing β -cells of the pancreas, resulting in an absolute deficiency of insulin. In the vast majority of cases, the destruction is due to a cell-mediated autoimmune attack of the β -cells. This process appears to be precipitated by environmental factors in genetically susceptible individuals. While viral infections, dietary components, and specific toxins have been suspected, the true identity of the environmental triggers remains unknown. About half of the genetic portion of the risk lies in the immune-recognition genes of the HLA locus. Variants in at least 16 other genetic loci also affect risk for type 1 diabetes. Several variants are also thought to act through the immune system; some of these diabetes risk variants operate in other autoimmune diseases including multiple sclerosis and autoimmune thyroid disease.

In almost all cases of type 1 diabetes, there is ultimately near complete destruction of the pancreatic β -cells. Because of this, these patients are dependent on treatment with exogenous insulin for their survival, as the absence of such treatment leads to rapid metabolic deterioration into diabetic ketoacidosis (DKA) and death.

Type 2 Diabetes Mellitus

Over 90% of diabetes is due to type 2 diabetes. The prevalence of this disease increases with age; previously, it had been uncommon for children or young adults to be diagnosed with this disease. However, there has been a recent dramatic increase in the prevalence of this disease, and with this, significant numbers of children are now being diagnosed with type 2 diabetes. Although much has been learned about the pathophysiology of type 2 diabetes in the past few decades, there remains much to be learned about the specific details of its cause. What is known is that two defects are present in patients with type 2 diabetes: insulin resistance and defective insulin secretion. Insulin resistance refers to a decreased effectiveness of insulin in activating signals distal to binding of insulin to the insulin receptor. Insulin resistance itself, except in the most extreme situation, will not lead to diabetes, as normal metabolic control can be maintained by a compensatory increase in insulin secretion. When a second defect results in an inability to respond to the requirement for increased insulin secretion imposed by insulin resistance, type 2 diabetes mellitus occurs.

There is strong familial clustering of type 2 diabetes, and the risk for this disease varies considerably across different ethnic populations. These and other findings indicate that there is a strong genetic component to the development of type 2 diabetes. However, multiple genes appear to be involved, with each individual with diabetes having altered function of multiple gene products. In addition, the relative contribution of insulin resistance versus defective insulin secretion can vary from patient to patient, so that there is likely a subset of altered genes contributing to the development of type 2 diabetes that differs from one person to the next. The genes involved include those encoding proteins affecting insulin signaling and insulin secretion, those controlling metabolic pathways, and those contributing to the development of obesity.

Environmental factors play a significant role in determining which genetically predisposed individuals will develop type 2 diabetes. Obesity and low levels of physical activity are strongly associated with insulin resistance, and are the factors to which the rising prevalence of type 2 diabetes in industrialized societies is ascribed. In addition, insulin sensitivity (the inverse of insulin resistance) decreases with age, accounting for much of the increased prevalence of this disease with increasing age.

Insulin resistance

Both central (hepatic) and peripheral (muscle and adipose) defects contribute to the insulin resistance of type 2 and obesity. (Interestingly, experiments in rodents have indicated that resistance in cells not classically characterized as being insulin responsive, including the brain and pancreatic β -cells, may contribute to the pathophysiology of obesity and type 2 diabetes.) Hepatic insulin resistance results in increased glucose output from the liver, predominantly due to increased gluconeogenesis. Insulin-resistant muscle and adipose cells have diminished insulin-stimulated glucose uptake, and the muscle cells have a marked decrease in glycogen synthesis. These defects, particularly in the muscle, result in decreased disposal of glucose after a meal. In addition, the decreased metabolism

of glucose to glycogen leads to increased glycolysis and production of lactate, which circulates to the liver and provides substrate for increased gluconeogenesis. The increased hepatic glucose production contributes most to fasting hyperglycemia, while the decreased glucose disposal into muscle contributes most to postprandial hyperglycemia.

Although the mechanisms remain incompletely understood, a number of factors have been demonstrated to contribute to the insulin resistance of type 2 diabetes, including free fatty acids and other factors secreted by adipocytes, as well as hyperglycemia itself (Figure 1). Free fatty acids impair insulin sensitivity, both in the liver and in muscle, and increased circulating levels of free fatty acids are found in obese and insulin-resistant subjects, due to increased lipolysis in adipocytes. Importantly, intra-abdominal fat tissue has a higher lipolytic rate compared to subcutaneous adipose tissue, and excess intra-abdominal adiposity has a much higher association with insulin resistance than obesity with a more peripheral (subcutaneous) distribution. Abnormal secretion by the adipocyte of a number of other factors (adipokines) has also been implicated in the pathophysiology of insulin resistance, including tumor necrosis factor α (TNF- α), resistin, adiponectin, and leptin. Hyperglycemia itself causes impaired insulin sensitivity. This may be due to activation of protein kinase C (PKC) by hyperglycemia. Recently, it has also been suggested that hyperglycemia-induced insulin resistance is in some way caused by increased flux of glucose through the glucosamine pathway.

Molecular mechanisms of insulin resistance

Insulin is a peptide hormone that binds to a transmembrane-spanning cell surface receptor (Figure 2). The insulin receptor is heterotetrameric in structure, containing two α - and two β -subunits. Insulin binds to the extracellular α -subunit, inducing a conformational change in the transmembrane β -subunit that derepresses intrinsic tyrosine kinase activity contained in the intracellular domain. The activated β -subunit tyrosine kinase activity first *trans*-phosphorylates the opposite β -subunit of the $\alpha_2\beta_2$ receptor complex. Once autophosphorylated, the insulin receptor is capable of phosphorylating intracellular proteins that propagate the insulin signal within the cell. Proteins of the insulin receptor substrate (IRS) family, the most important of these being IRS-1 and IRS-2, recognize and bind to phosphotyrosine motifs on the activated insulin receptor.

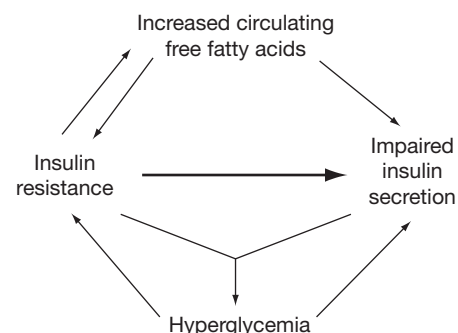


Figure 1 Factors contributing to insulin resistance and impaired insulin secretion include the ability for each of them to affect both themselves and each other through direct and indirect mechanisms.

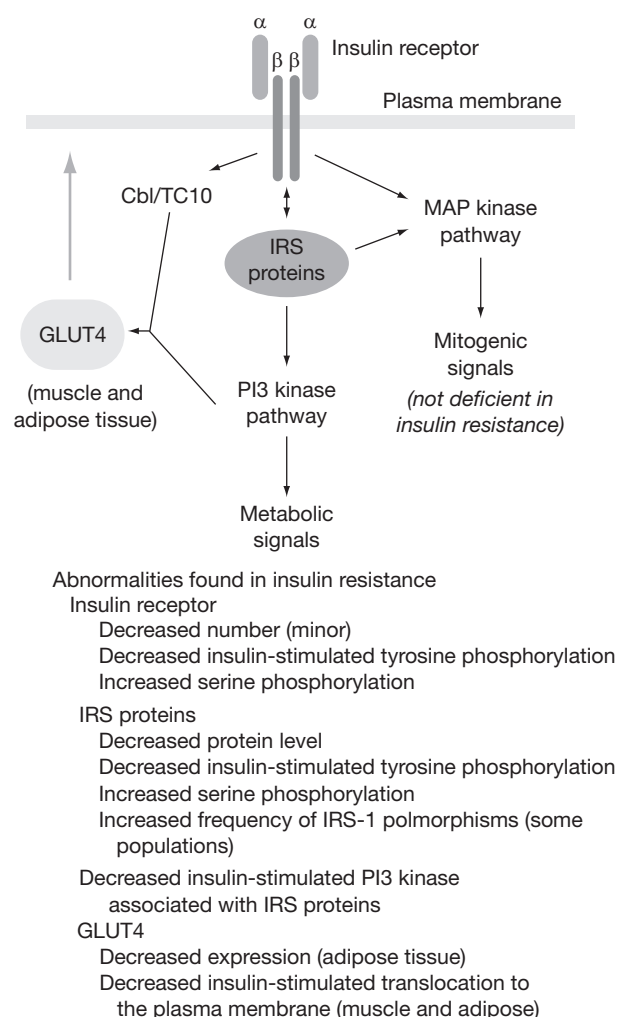


Figure 2 Major signaling pathways of insulin action and abnormalities found in insulin resistance.

Once bound, the IRS proteins are themselves phosphorylated on multiple tyrosine residues by the activated insulin receptor. The IRS proteins do not have intrinsic enzymatic activity, but serve as docking sites to bring other proteins into proximity to the insulin receptor. A signaling cascade is initiated, as proteins recognize and bind to the phosphotyrosine motifs on either the insulin receptor or the IRS proteins. This results in their activation, often through their phosphorylation by the insulin receptor.

The downstream components of the insulin-signaling cascade include protein kinases and phosphatases, and lipid kinases. The two pathways that have been investigated in detail are the mitogen-activated protein (MAP) kinase and the phosphatidylinositol-3-OH (PI3) kinase pathways. Although the distinction is not absolute, activation of the MAP kinase pathway is most responsible for the actions of insulin related to cell growth, while activation of the PI3 kinase pathway is predominantly involved in the metabolic actions of insulin such as glucose uptake and glycogen synthesis. This is significant, because activation of the PI3 kinase pathway is markedly impaired in the insulin resistance of type 2 diabetes, while signaling through the MAP kinase pathway is much less impaired. This difference may contribute to the pathogenesis

of the long-term complications of insulin resistance and diabetes. A third pathway involves phosphorylation and activation of the Cbl proto-oncogene by the insulin receptor, which results in activation of the G protein TC10. Many of the pathways that are activated by the insulin receptor are also activated by other means, particularly activation of other growth factor receptors. Specificity of the insulin signal is felt to be due to the integration of the multiple pathways activated, as well as to compartmentalization of the components of the cascade. For example, both the PI3 kinase and Cbl/TC10 pathways may be necessary for insulin-stimulated glucose uptake into muscle and adipose cells, although the importance of the Cbl/TC10 pathway in insulin signaling remains uncertain.

Abnormalities of each of the components of the insulin-signaling cascade, from the insulin receptor at the beginning to the enzymes, transcription factors, and other effector proteins at the end, are potential candidates for contributing to the insulin resistance of type 2 diabetes. Indeed, a polymorphism near the IRS-1 gene was found to associate with insulin resistance and moderately increased risk for type 2 diabetes in Europeans. Otherwise, however, mutations or significant alterations in expression in these signaling components have been found only in isolated instances and have not been found to play a major role in the insulin resistance of type 2 diabetes. There have, however, been clearly identified abnormalities in insulin signaling that occur in the insulin resistance of type 2 diabetes (Figure 2). Notably, insulin-stimulated tyrosine phosphorylation of the insulin receptor and IRS proteins is decreased in tissues in insulin-resistant subjects. This is often associated with an increase in phosphorylation on specific serine and/or threonine residues in these proteins, which may inhibit the protein-protein interactions necessary for proper signaling. Potential mechanisms for this include serine-phosphorylation of the insulin receptor by PKC, and serine/threonine-phosphorylation of IRS-1 and IRS-2 by downstream mediators of insulin signaling (mammalian target of rapamycin (mTOR), mammalian target of rapamycin complex 1/p70 S6 kinase (mTORC1/S6K), and protein kinase C ζ (PKC ζ)) and other kinases. In particular, increases in, for example, free fatty acids, inflammatory cytokines, and reactive oxygen species (ROS) are found in obesity-activated serine phosphorylation of IRS-1 and IRS-2 proteins via glycogen synthase kinase (GSK), inhibitor of nuclear factor kappaB kinase (IKK β), and c-Jun N-terminal kinase (JNK).

In order to enter a cell, glucose requires specialized transporters to be present on the cell surface. GLUT4, the insulin-responsive glucose transporter, is sequestered in intracellular vesicles in the basal state. When muscle or adipose cells are stimulated by insulin, these vesicles translocate and fuse with the plasma membrane, and the increase in transporters at the cell surface results in increased glucose uptake. In the insulin resistance of obesity and type 2 diabetes, there is a decrease in the insulin-stimulated translocation of GLUT4. There may also be a decrease in the intrinsic activity of GLUT4 in insulin-resistant states. In adipocytes (but not muscle cells) there is significantly decreased expression of GLUT4 that contributes to the diminished insulin-stimulated glucose uptake. Although quantitatively adipose tissue contributes only a small fraction to glucose disposal compared to muscle, evidence from animal models with tissue-specific insulin resistance demonstrates that insulin resistance in one tissue can affect insulin sensitivity

in other tissues. Thus, insulin resistance in the adipocyte can cause insulin resistance in liver and muscle, most likely through secreted factors including free fatty acids, retinol-binding protein 4, and adipokines.

Insulin secretory defect

Insulin secretion and β -cell mass increase to compensate for states of insulin resistance such as obesity, pregnancy, or cortisol excess, so that fasting and meal-stimulated insulin levels are elevated even as glucose levels remain normal. When insulin secretion fails to fully compensate for the degree of insulin resistance, glucose levels rise, first to those of mild glucose intolerance (glucose levels above normal but below criteria for a diagnosis of diabetes), and then to those of diabetes. Insulin levels early in this progression are still elevated compared to individuals with normal insulin sensitivity, but are not elevated sufficient to maintain normal glucose control. With time, in most patients, the insulin secretory defect continues to worsen, including loss of β -cells, ultimately requiring exogenous insulin treatment to maintain control of blood glucose levels.

The development of impaired insulin secretion is likely due to both genetic and environmental factors. Indeed, variants at more than 20 genetic loci have recently been identified as playing a role in type 2 diabetes risk; of these, most appear to affect insulin secretion, but detailed mechanisms of action remain to be elucidated. Gene variants may also affect an individual's response to nongenetic factors. One factor contributing to the loss of β -cell function may be a persistent stimulus to oversecrete insulin caused by insulin resistance (Figure 1). This may be due in part to the effect of excessive secretion of islet amyloid polypeptide (IAPP), which is cosecreted with insulin; in the islets in the majority of patients with type 2 diabetes there is the accumulation of amyloid, containing IAPP. Another factor is the elevated free fatty acid level present in insulin resistance, which can contribute to the defective insulin secretion, just as it can contribute to insulin resistance. While short-term exposure to lipids increases insulin secretion, long-term exposure impairs β -cell function, and may be responsible for β -cell death by apoptosis. Similarly, prolonged exposure to increased glucose levels imposes a glucotoxicity on the β -cell; if glucose levels are normalized in a patient with diabetes, endogenous insulin secretion will improve. Over time, however, this glucotoxicity may lead to irreversible β -cell damage, perhaps through oxidative injury. Inflammatory cytokines such as IL-1 β also appear to induce β -cell apoptosis, which may be important in proinflammatory states such as obesity. Finally, insulin resistance of the β -cell may itself lead to impaired β -cell function, as demonstrated in mice carrying genetic defects of insulin signaling in β -cells.

Recent investigations have implicated endoplasmic reticulum (ER) stress and the unfolded protein response (UPR) in the pathogenesis of type 2 diabetes. The strongest evidence supports a role for the UPR in contributing to β -cell loss, although there are also data indicating that it can contribute to insulin resistance. The UPR is activated in cells when protein synthesis exceeds the capacity of the ER to properly fold them. The response includes a global decrease in protein translation, an increase in the degradation of misfolded proteins, and a specific increase in production of proteins that assist in the folding of proteins within the ER, such as chaperones. When

these steps fail to prevent the accumulation of misfolded proteins, apoptosis is triggered. The rate of proinsulin production can approach 1 million molecules per minute per cell, making it clear why β -cells are at risk of activation of the UPR. A clear demonstration that the UPR can contribute to β -cell loss is the monogenic disease Wolcott-Rallison syndrome, a rare disorder that includes insulin-deficient diabetes with an onset in infancy. It is caused by a mutation in the *EIF2AK3* gene, which encodes a protein (eukaryotic initiation factor 2 α (eIF2 α)) kinase 3, equivalent to rodent protein kinase RNA-like endoplasmic reticulum kinase (PERK)) that is central to the UPR.

Other Specific Types

Diabetes of types other than type 1 or type 2 make up a small percentage of the cases of diabetes. For a number of these, however, a single gene defect results in the disease, providing potential insight into what may be the underlying cause of the insulin resistance or insulin secretory defect in type 2 diabetes.

Genetic Defects of β -Cell Function

Maturity-Onset Diabetes of the Young

MODY (also known as monogenic diabetes of the young) refers to a group of diabetes types caused by mutations in single genes and having autosomal dominant inheritance. MODY patients have early-onset diabetes presenting in young adulthood (usually before the age of 25 years). In contrast to type 2 diabetes, obesity and insulin resistance are not a major feature of MODY; the diabetes in these cases is due to the defective insulin secretion. Clinical characteristics of the diabetes depend on the underlying mutation, and the severity of the untreated diabetes can vary from mild, sometimes unrecognized hyperglycemia with little risk of long-term microvascular diabetic complications, to severe hyperglycemia with a very high risk of microvascular complications.

MODY-causing mutations in six genes have been established from studies in multiple populations, and three additional MODY-causing mutations have been more recently identified, so far in specific populations. One form of MODY (MODY2) is caused by heterozygous mutations in the gene for glucokinase, which is expressed in highest levels in the pancreatic β -cell and in the liver. In the β -cell, glucokinase catalyzes the phosphorylation of glucose to generate glucose-6-phosphate, the first step in glucose metabolism. Because glucose metabolism is the stimulus for insulin release, glucokinase functions as the glucose sensor in the β -cell as it is rate limiting for glucose entry into the glycolytic pathway. Thus, glucokinase mutations lead to decreased glucose-stimulated insulin secretion. In the liver, glucokinase affects the storage of glucose as glycogen; this defect may contribute to the hyperglycemia in patients with glucokinase mutations.

Other genes that have been identified as having mutations that cause MODY are each transcription factors expressed in the pancreatic β -cell: hepatocyte nuclear factor (HNF)-4 α (MODY1); TCF1, which encodes HNF-1 α (MODY3); insulin promoter factor-1 (IPF1, also known as PDX-1; MODY4); TCF2, which encodes HNF-1 β (MODY5); neurogenic differentiation 1/ β -cell E-box transactivator 2 (NeuroD1/BETA2;

MODY6); and Kruppel-like factor 11 (KLF11; MODY7). Each of these genes is expressed in the β -cell and regulates expression of the insulin gene and genes for other factors involved in the regulation of insulin signaling; heterozygous mutations lead to β -cell dysfunction sufficient to cause diabetes. IPF-1, NeuroD1/BETA2, and paired box 4 (PAX-4; MODY9) are also involved in pancreatic development.

A syndrome of pancreatic exocrine insufficiency and β -cell dysfunction causing diabetes (MODY8) is caused by mutations in the gene for carboxy-ester lipase, expressed in the exocrine pancreas. The mechanism of action is unknown, and to date this appears to be a rare cause of MODY.

Neonatal Diabetes

Rarely, children may inherit a defect in insulin production which causes diabetes in infancy. Neonatal diabetes, usually present before 6–12 months of age, is non-immune-mediated and may be transient or permanent. Individuals who had transient neonatal diabetes are at increased risk of developing nonimmune type 1 diabetes at a later age. Mutations causing neonatal diabetes have been identified in genes involved in insulin production and secretion in β -cells (insulin, glucokinase, and components of a key potassium channel involved in insulin secretion: Kir6.2 and sulfonylurea receptor (SUR)), transcription factors expressed in the pancreas whose dysfunction causes pancreatic agenesis (IPF-1 and PTF1A) and genes operating through other mechanisms (*EIF2AK3*, *FOXP3*, and *GLIS-3*).

Some of the genes involved in MODY or neonatal diabetes can be associated with different forms of diabetes based on the severity of the genetic abnormality. In the case of the glucokinase and IPF-1 genes, a specific mutation causes MODY (when one copy of the gene is mutated) or neonatal diabetes (when both copies are mutated). In other cases, different mutations in the same gene are associated with neonatal diabetes, MODY, and/or type 2 diabetes (e.g., *HNF-1 β* , *HNF-4 α* , Kir6.2, insulin gene, and KLF11). For the case of the glucokinase gene, other mutations associate with fasting hyperglycemia but not diabetes. Some type 2 diabetes risk variants also impact gestational diabetes (see below). In addition to the above cases predisposing to forms of diabetes, other mutations in several of the above genes can cause syndromes of excessive insulin secretion and hypoglycemia.

Mitochondrial DNA Mutations

Mutations in the mitochondrial genome can cause diabetes mellitus, the most common involving the mitochondrial tRNA^{LEU(UUR)}. Due to the maternal origin of mitochondria, this type of diabetes is maternally inherited. The pancreatic β -cell is rich in mitochondria and has a high rate of oxidative metabolism; the mitochondrial mutations result in a β -cell defect that causes diabetes through deficient insulin secretion. This form of diabetes is frequently associated with neurosensory deafness.

While inherited mitochondrial genome mutations are rare as a cause of diabetes mellitus, they may provide insight into the worsening insulin secretory defect that occurs in type 2 diabetes. The mitochondrial genome consists only of coding sequences, and its repair mechanisms are poor so that it is

highly susceptible to mutation. With age, there is an increase in ROS generated in mitochondria, and this may be exacerbated in β -cells with a higher rate of oxidative metabolism due to exposure to elevated glucose levels. With a low expression of enzymes that defend against oxidative damage, β -cells are particularly sensitive to oxidative injury. The inherited mitochondrial genome mutations that cause diabetes demonstrate that these mutations can impair insulin secretion (though the exact mechanism is unknown), and it is possible that an accumulation of mitochondrial DNA mutations due to oxidative damage contributes to the β -cell dysfunction of type 2 diabetes.

Genetic Defects in Insulin Action

Insulin Receptor Mutations

Insulin resistance can be caused by mutations of the insulin receptor, although these are rare as a cause of diabetes. Most of the subjects identified with insulin resistance and insulin receptor mutations have mutations of both alleles. Some cases of mild insulin resistance have been described in which only one allele carries a mutation, acting as a dominant negative inhibitor of the wild-type allele, due to their combination within the heterotetrameric structure of the receptor. The severity of the phenotype with two mutant insulin receptor alleles can vary from insulin resistance with relatively normal glucose control (type A insulin resistance) to a more severe phenotype evident in infancy (Rabson–Mendenhall syndrome and Leprechaunism, which is usually fatal in infancy).

Secondary Diabetes Mellitus

A number of systemic disorders can cause marked insulin resistance, which can lead to secondary diabetes mellitus (although as for type 2 diabetes and as discussed above, there must be some defect in insulin secretion to prevent complete compensation for the insulin resistance by the β -cell). Excess growth hormone, as in acromegaly, excess glucocorticoid, as in Cushing's disease or exogenous glucocorticoid treatment, or excess catecholamines, as in pheochromocytoma, as well as other diseases such as uremia and hepatic cirrhosis can lead to secondary diabetes primarily through their effect on insulin sensitivity. Similarly, diabetes mellitus will occur if a disorder produces sufficient β -cell damage. Examples of this include pancreatitis, cystic fibrosis, pancreatectomy, and hemochromatosis; in addition, somatostatin-secreting tumors can inhibit insulin secretion sufficient to cause diabetes mellitus.

Rarely, patients with autoimmune disease have been identified where antibodies against the insulin receptor block signaling, leading to an insulin resistant form of diabetes referred to as type B insulin resistance.

Strikingly, just as the excess adipose tissue in obesity leads to insulin resistance, the deficiency of adipose tissue that is found in lipoatrophy/lipodystrophy also causes insulin resistance, often severe. Mice that are genetically engineered to have a paucity of adipose tissue are also markedly insulin resistant, confirming this association. The cause of insulin resistance in syndromes of adipose deficiency is likely due to intrahepatic and intramuscular accumulation of triglyceride, as well as a loss of normal secreted signals from adipocytes.

Gestational Diabetes

Glucose intolerance that is first identified during pregnancy is referred to as gestational diabetes. The importance of this classification is the significant maternal and fetal morbidity related to diabetes during pregnancy, warranting screening for diabetes mellitus in all pregnant women except those at low risk. In some cases gestational diabetes is merely the identification of glucose intolerance that was previously unrecognized until screening is performed during a pregnancy, or to the temporal development of diabetes (type 1 or type 2) during a pregnancy, perhaps accelerated by the pregnant state, but otherwise predestined to develop. In the majority of cases, however, glucose regulation returns to normal after the pregnancy. Pregnancy is characterized by progressively decreasing insulin sensitivity, due to the insulin antagonizing effects of increases in prolactin, human placental lactogen, estrogens, progesterone, and unbound cortisol levels. In the majority of pregnancies, this insulin resistance is compensated for by increased insulin secretion. However, in the 2–5% of pregnancies that develop gestational diabetes mellitus, there is an inadequate compensation. Thus, gestational diabetes represents the same or similar pathophysiologic abnormality of insulin sensitivity and insulin secretion that occurs in type 2 diabetes, perhaps revealed by the insulin resistance of pregnancy. Indeed, women who have had gestational diabetes are at a substantially increased risk for the future development of type 2 diabetes. A number of common gene variants have been found to increase the risk for both type 2 diabetes and gestational diabetes, supporting the hypothesis that some of the underlying pathophysiology is shared.

Metabolic Syndrome and Prediabetes

Insulin resistance often precedes the development of diabetes mellitus by many years. In addition, in some people, β -cell dysfunction does not develop, and insulin resistance exists indefinitely without progressing to diabetes. However, insulin resistance is not without significance of its own. The metabolic syndrome (also referred to as syndrome X) is an association of findings that frequently coexist, and includes insulin resistance, obesity (particularly abdominal/visceral rather than peripheral/subcutaneous), hypertension, and a specific form of dyslipidemia: hypertriglyceridemia and decreased high-density lipoprotein (HDL) cholesterol. These subjects are at risk of developing type 2 diabetes if they do not already have it, but more importantly, even in the absence of progression to diabetes, these individuals are at a markedly elevated risk for atherosclerotic vascular disease. The causality of obesity and insulin resistance in this syndrome is supported by the fact that lipid levels and hypertension may both be improved by decreasing insulin resistance through exercise and weight loss.

The diagnosis of diabetes mellitus is based on blood glucose levels that exceed defined levels, either in the fasting state or after a challenge with an ingested glucose load. The cutoffs are based mostly on levels that predict an increased risk for diabetic microvascular complications. Some individuals with insulin resistance will have glucose levels that do not meet criteria for the diagnosis of diabetes, but exceed levels considered normal. Some of these subjects with impaired fasting

glucose or impaired glucose tolerance will have glucose levels that revert to normal, or remain impaired over time. However, a large percentage will have progressive β -cell dysfunction (or worsening insulin resistance, or both) that results in progression to type 2 diabetes — ~40% in 5–10 years.

Complications

Acute

Ketoacidosis

DKA is the life-threatening acute metabolic decompensation in patients with diabetes mellitus. Because the development of DKA is suppressed by even modest insulin action, DKA is generally only a risk for patients with type 1 diabetes. The pathophysiology of DKA is one of acidosis, hyperosmolarity, and dehydration.

Decompensation toward DKA begins with a fall in insulin action, either due to an absolute decline in the insulin level, or due to antagonism of insulin action by cytokines and stress hormones, including catecholamines, cortisol, glucagon, and growth hormone. As insulin action becomes inadequate, serum glucose levels rise; the loss of insulin suppression of gluconeogenesis and glycogenolysis in the liver results in an increase in hepatic glucose output, and the loss of insulin stimulation of glucose uptake in muscle and adipose tissue results in accumulation of the excess glucose in the blood. The loss of insulin suppression results in unrestrained lipolysis in the adipocyte, releasing free fatty acids and glycerol. Increased protein breakdown in muscle releases amino acids. These metabolites are taken up by the hepatocytes, where the free fatty acids are metabolized to ketoacids (β -hydroxybutyrate and acetoacetic acid), and the glycerol and amino acids serve as substrates for accelerated gluconeogenesis.

The hyperglycemia and ketoacidosis are initially moderated by the loss of glucose and ketoacids in the urine. However, the glucosuria produces an osmotic diuresis, and the ketoacidosis produces an ileus leading to nausea and vomiting that impairs the ability of the patient to drink to maintain their hydration. As dehydration develops, then worsens, the glomerular filtration rate of the kidney falls, and less glucose and ketoacids can be excreted in the urine, resulting in dramatic accumulation of glucose and ketoacids in the blood. The evolving disease stimulates (further) release of stress hormones that antagonize any remaining insulin action, and further exacerbate the hyperglycemia. The acidosis and the hyperosmolarity caused by the extreme hyperglycemia decrease cerebral function and impair the patient's consciousness, further compromising the patient's ability to drink. Without intervention, death is inevitable.

Nonketotic hyperosmolar coma

Patients with type 2 diabetes are not at much risk of developing DKA. They will generally have sufficient insulin action to restrain lipolysis, limiting free fatty acid delivery to the liver and subsequent ketoacid production. However, decompensation similar to that of DKA, but without the excess ketoacid production, can occur in patients with type 2 diabetes. Nonketotic hyperosmolar coma in patients with type 2 diabetes generally occurs when another illness induces a stress response, and there is an impaired ability to maintain hydration with

oral fluids. As in DKA, the increase in stress hormones antagonizes insulin action, resulting in increased hepatic glucose production and decreased peripheral glucose utilization. The rising hyperglycemia produces a glucosuria-driven osmotic diuresis. Dehydration is a result of an inability of oral fluid intake to compensate for the osmotic diuresis. As in DKA, the dehydration impairs the renal excretion of glucose, allowing the glucose level to continue to rise. The dehydration and hyperosmolarity impair the patient's consciousness, and oral intake falls further. The final situation is one of marked hyperglycemia, hyperosmolarity, and dehydration. As in DKA, the mortality of nonketotic hyperosmolar coma is high.

Chronic

Patients with diabetes will occasionally have bothersome symptoms and are at risk of severe acute metabolic decompensation. Even if treatment is successful in avoiding these issues, diabetes can lead to long-term complications that develop over years. These complications can be categorized as microvascular, which are unique to patients with diabetes, and macrovascular, which are complications of atherosclerotic vascular disease that also occur in people without diabetes, but occur with higher frequency in those with diabetes. These complications are felt to be due to a failure of treatment to adequately normalize insulin and glucose levels, as well as to changes produced by insulin resistance in type 2 diabetes.

Macrovascular disease

As in individuals without diabetes, atherosclerotic changes in medium- and large-sized arteries that either partly or completely occlude the vessel lumen lead to morbidity and mortality due to impaired blood flow to the heart (angina and myocardial infarction), the brain (stroke), or the periphery (claudication and ischemic damage requiring amputation). These complications occur at 2–15 times greater frequency and develop at younger ages in patients with diabetes; over half of patients with type 2 diabetes die from cardiovascular disease. It is not fully clear how diabetes contributes to the risk of atherosclerotic vascular disease. The factors that have been implicated include lipid abnormalities, hyperglycemia, and increases in prothrombotic and proinflammatory factors. In addition, the coexistence of other atherosclerotic risk factors, especially hypertension, smoking, and obesity, also plays a role.

In patients with type 2 diabetes, the insulin resistance that is a part of the pathophysiology is associated with a specific pattern of lipid abnormality that increases the atherosclerotic risk: decreased HDL cholesterol concentration, and elevated triglyceride concentration in the serum. In addition, the inadequate insulin action in poorly controlled diabetes of any type leads to inadequate stimulation of lipoprotein lipase activity and to increased lipolysis and elevated free fatty acids. The free fatty acids that are not oxidized to ketones in the liver are re-esterified to triglycerides and secreted as very low density lipoproteins (VLDLs). Thus, poorly controlled diabetes can lead to atherosclerotic risk by increasing triglyceride and VLDL levels; hypertriglyceridemia can exacerbate the risk by depressing HDL levels. Hyperglycemia may contribute to disordered

serum lipids by reducing expression of the heparin sulfate proteoglycan perlecan on hepatocytes, leading to increased levels of the atherogenic cholesterol-enriched apolipoprotein B-containing remnant particles. Finally, glycated LDL is more susceptible to oxidation and is more atherogenic.

Endothelial dysfunction is central to the development of atherosclerosis, with decreased production of the vasodilating molecule nitric oxide and increased production of the pro-thrombotic factor plasminogen activator inhibitor-1 (PAI-1) – two factors implicated in this dysfunction. In diabetes, hyperglycemia inhibits nitric oxide production and increases production of PAI-1. Hyperglycemia may also contribute to macrovascular disease indirectly through microvascular damage to the nutrient blood vessels supplying the walls of the medium- and large-sized arteries.

Another aspect that is likely to play a role in the development of atherosclerosis in type 2 diabetes relates to the pathway-specific insulin resistance that is present. As noted previously, the insulin resistance of type 2 diabetes has less of an effect on the activation of the intracellular MAP kinase pathway than the pathway activating phosphatidylinositol-3-kinase. This may lead to increased activation of the MAP kinase pathway by insulin signaling, which also decreases nitric oxide production and increases PAI-1 production. In addition, as the MAP kinase pathway is a mitogenic pathway, the increased MAP kinase signaling can increase proliferation of the vascular smooth muscle cells, contributing to the atherosclerotic process.

Microvascular disease

The hyperglycemia of diabetes mellitus leads to damage to the microvascular circulation that results in tissue and organ damage, most notably in the retina, kidneys, and nerves. Due to these microvascular complications, diabetes mellitus is a leading cause of blindness, end-stage renal disease, and neuropathy. Increasing degrees of hyperglycemia are the main risk factor for microvascular complications, with the Diabetes Control and Complications Trial (for type 1 diabetes) and the United Kingdom Prospective Diabetes Study (for type 2) definitively demonstrating a decreased incidence of microvascular complications with treatments that lower the degree of hyperglycemia. Unfortunately, even the best of current therapies do not completely normalize blood glucose levels. Also, elevated triglycerides, decreased HDL, and elevated apolipoprotein C-III independently correlate with progression of microvascular disease, and treatment of diabetic dyslipidemia delays progression of diabetic nephropathy. The risk of microvascular complications increases with time, with the earliest evidence uncommon prior to a duration of hyperglycemia of 5 years; note, however, that except for typical type 1 diabetes, hyperglycemia may be present for years before the diagnosis of diabetes is made. Finally, as yet undefined genetic factors also play a role in the risk of microvascular disease.

Microvascular disease begins with abnormal blood flow and increased vascular permeability due to decreased activity of vasodilators, increased activity of vasoconstrictors, and overexpression of permeability factors. Irreversible increases in vascular permeability occur with alterations in the extracellular matrix. Ultimately, there is cell loss, capillary occlusion from deposition of extravasated plasma proteins, and the overproduction of extracellular matrix. Microvascular disease affects the glomerulus

in the kidney, with the earliest evidence being leakage of albumin into the urine and progressing to loss of glomerular function. Damage to retinal capillaries causes hypoxic-ischemia, edema, new vessel formation, and hemorrhage. It is the development of new vessels, referred to as proliferative retinopathy, that can lead to retinal detachment, vitreous hemorrhage, and loss of sight. Diabetic neuropathy includes degeneration of motor, sensory, and autonomic nerves.

While it is clear that hyperglycemia leads to microvascular disease, the precise mechanism for this has not been completely demonstrated. However, five pathologic processes triggered by hyperglycemia have been identified and proposed as causative mechanisms: increases in the polyol and hexosamine pathways, production of advanced glycation end-products (AGEs), oxidative stress, and activation of PKC. *In vivo* studies with inhibitors, either in animal models or in human subjects, have supported a role for each of these pathways (except the recently implicated hexosamine pathway) in microvascular pathology. The relative contribution of each of these proposed mechanisms likely varies across cell types.

Elevated triglycerides, decreased HDL, and elevated apolipoprotein C-III also independently correlate with progression of microvascular disease, and treatment of diabetic dyslipidemia delays progression of diabetic nephropathy.

Polyol pathway

Hyperglycemia drives an increased intracellular generation of sorbitol through the nicotinamide adenine dinucleotide phosphate (NADPH)-dependent reduction of glucose by aldose reductase. Increased production of sorbitol could alter cellular functions in a number of ways, including increasing intracellular osmotic stress, decreasing myoinositol and taurine (as compensation for the increase in sorbitol in order to prevent or mitigate an increase in osmolarity), or through an increased NADH/NAD⁺ ratio (due to NAD⁺-dependent oxidation of sorbitol to fructose). However, the most likely mechanism for cellular dysfunction caused by increased flux through the polyol pathway is due to the consumption of NADPH. This leads to a lack of NADPH available to regenerate reduced glutathione, limiting the ability of the cell to protect itself against oxidative stress.

Hexosamine pathway

The hexosamine pathway generates uridine diphosphate (UDP)-*N*-acetylglucosamine (GlcNAc) from glucose to provide substrate for proteoglycan synthesis and for O-linked glycosylation of certain proteins. The activity or function of many proteins, including transcription factors, can be modified by O-linked glycosylation, often in concert with reciprocal modification by phosphorylation. With hyperglycemia, some of the excess glucose is shunted into the hexosamine pathway. As noted previously, increased flux through this pathway may play a role in insulin resistance. In addition, increased generation of UDP-GlcNAc could alter gene expression and cellular function through excess protein glycosylation, contributing to the pathogenesis of microvascular disease.

Advanced glycation end-products

Glucose will react with free amino groups on proteins in a nonenzymatic reaction to form glycated proteins. This reaction

starts with the formation of a Schiff base (aldamine), which then undergoes an internal Amadori rearrangement. The formation of the Schiff base is rapid and reversible, while the Amadori rearrangement is much slower and results in a more stable ketoamine. This process occurs at a rate proportional to the prevailing glucose level that a given protein is exposed to. The glycation of hemoglobin is used clinically to estimate the average blood sugar in patients with diabetes, as the percent of hemoglobin in the glycated A_{1c} form (HbA_{1c}) is proportional to the average glucose level present over ~3 months prior to testing (roughly the red blood cell life span).

Proteins modified irreversibly through nonenzymatic reaction with glucose or glucose-derived compounds are referred to as AGEs. Most AGEs are not formed through the reaction of proteins directly with glucose but rather with three reactive dicarbonyl compounds that are produced intracellularly from glucose: glyoxal (an auto-oxidation product of glucose), 3-deoxyglucosone (formed from decomposition of the Amadori product of glucose), and methylglyoxal (from fragmentation of glyceraldehyde-3-phosphate and dihydroxyacetone phosphate generated from glycolysis). The reactive dicarbonyls interact with the amino groups of proteins, and can introduce both intra- and intermolecular cross-links. AGE formation may lead to microvascular damage through alteration of the function of both intracellular and extracellular proteins modified in this way. In particular, the modification of extracellular matrix proteins by this process is likely to contribute to abnormal vascular permeability and elasticity. In addition, however, AGEs formed from plasma proteins bind to and activate specific AGE receptors on endothelial cells, mesangial cells, and macrophages (one such receptor is RAGE, a member of the immunoglobulin superfamily). Activation of AGE receptors induces production of cytokines, growth factors, procoagulatory factors, and ROS. Induction of vascular endothelial growth factor (VEGF) by AGE receptor activation most likely contributes to the increased permeability of capillary walls present in diabetes.

Activation of PKC

Hyperglycemia increases intracellular diacylglycerol (DAG) levels, which is an activator of many of the PKC isoforms. Hyperglycemia may also activate PKC isoforms indirectly by increasing ROS from activation of AGE receptors and from increased activity of the polyol pathway. As discussed above, activation of PKC may play a role in hyperglycemia-induced insulin resistance. Mechanisms through which activation of PKC may be involved in the development of microvascular disease include: inhibiting production of the vasodilator nitric oxide; increasing activity of the vasoconstrictor endothelin-1; increasing expression of VEGF, leading to increased cell permeability; and contributing to the accumulation of extracellular matrix protein by inducing expression of transforming growth factor (TGF)- β 1, fibronectin, and type IV collagen.

Oxidative stress

The increased metabolism of glucose in the presence of hyperglycemia increases the delivery of electron donors from the tricarboxylic acid cycle to mitochondria. This can lead to an increase of the proton gradient across the inner mitochondrial

membrane, inhibiting the electron transport chain and increasing production of superoxide and other ROS. Oxidative stress can directly contribute to the development of microvascular complications by stimulating signal transduction pathways resulting in altered cellular functions and gene expressions. In addition, ROS can react with and damage DNA, proteins, and lipids. ROS can trigger mitochondria to release cytochrome *c*, triggering apoptosis. In addition to these direct actions of ROS, oxidative stress can activate the other potential microvascular pathogenic pathways. Because superoxide inhibits the glycolytic enzyme glyceraldehyde 3-phosphate dehydrogenase, precursors can accumulate and be diverted into the polyol pathway (glucose), the hexosamine pathway (fructose-6-phosphate), production of DAG (glyceraldehyde-3-phosphate), which activates PKC isoforms, and production of the reactive dicarbonyl methylglyoxal (glyceraldehyde-3-phosphate), which produces AGEs.

Diabetes Treatment Medications – Mechanisms of Action

Patients with diabetes of any type can be treated with insulin to overcome the relative or absolute insulin deficiency. For patients with type 1 diabetes, the loss of the ability to produce endogenous insulin because of the loss of pancreatic β -cells makes insulin a required treatment. Similarly, in many patients with type 2 diabetes, the β -cell dysfunction will progress to a degree that insulin treatment is necessary to maintain glycemic control. However, for many patients with type 2 diabetes, a number of treatments may be effective in lowering the degree of hyperglycemia instead of or in addition to insulin.

Weight Loss and Exercise

For the majority of patients with type 2 diabetes, obesity is a significant contributor to the insulin resistance that underlies their disease. Therefore, in many cases, if the patient is able to lose weight, insulin sensitivity will improve sufficiently to return metabolic control to normal or near normal. In many cases, only modest weight loss can be effective. A sedentary lifestyle also leads to decreased insulin sensitivity. Therefore, exercise to improve fitness can also improve insulin sensitivity sufficient to ameliorate the diabetic state, an effect that can occur in the absence of weight loss. Key mechanisms for this improved insulin sensitivity are the exercise-induced stimulation of AMP-activated kinase (AMP kinase) and PI3 kinase activities. One of the effects of the activation of AMP kinase is to directly increase GLUT4 translocation to the plasma membrane, increasing cellular glucose uptake. Prolonged exercise training also increases expression of both AMP kinase and GLUT4.

Exercise also induces the upregulation of factors such as peroxisome proliferator-activated receptor γ coactivator 1 α (PGC-1 α), peroxisome proliferator-activated receptor- γ (PPAR- γ), and nuclear respiratory factor 1 (NRF1), which induce an increased synthesis of new mitochondria. This, along with the effect of AMP kinase activation on lipid metabolism, results in increased fatty acid oxidation which contributes to the improved insulin sensitivity.

Insulin Secretagogues

Sulfonylureas and meglitinide analogs lower blood sugar levels by stimulating endogenous insulin secretion. They act by binding to the SUR – one component of the β -cell potassium channel, the other being Kir6.2. Binding of these agents to SUR results in closure of the potassium channel, mimicking the effect of a glucose-driven increase in the cellular ATP/ADP ratio (ATP, adenosine triphosphate; ADP, adenosine diphosphate), and setting in motion the intracellular signal for exocytosis of insulin. Unfortunately, due to the worsening β -cell dysfunction and β -cell loss that occurs in most patients with type 2 diabetes, insulin secretagogues often become ineffective over time.

Metformin

Metformin is a biguanide compound that works by increasing insulin sensitivity in diabetic patients. Its main effect is on the liver, to decrease excess hepatic glucose production. The underlying mechanism for metformin's action has not been fully determined. However, it appears to function by activating AMP-activated protein kinase (AMPK). By activating AMPK, the hepatocyte shifts from lipogenesis to fatty acid oxidation. This reduces hepatic triglyceride content, which has been implicated in the cause of insulin resistance. Activation of AMPK may increase insulin sensitivity by interrupting the S6K-activated serine/threonine-phosphorylation of IRS protein, increasing their efficiency in insulin signaling. How metformin activates AMP kinase is uncertain, but may involve inhibition of the mitochondrial electron transport chain.

Thiazolidinediones

Thiazolidinediones (TZDs) are a second class of insulin-sensitizing drugs. In contrast to metformin, however, their primary action is to increase peripheral insulin sensitivity, thereby increasing glucose uptake into fat and muscle. TZDs are ligands for PPAR- γ , which is expressed in high levels only in adipose tissue, although low levels of expression are seen in other tissues, including in muscle, liver, and pancreatic β -cells. PPAR- γ is a transcription factor that alters gene expression in a ligand-dependent manner. The importance of PPAR- γ in the pathophysiology of diabetes is highlighted by the fact that a common amino acid polymorphism in the PPAR γ gene (proline instead of alanine at codon 12, occurring in both chromosomes in about 75% of Caucasians) confers decreased insulin sensitivity and insulin clearance and greater risk of type 2 diabetes. TZDs most likely exert their effect by altering PPAR- γ -dependent gene expression in adipocytes, which leads to improved insulin action in muscle and liver. However, TZDs have also been shown to stimulate AMPK in liver, muscle, and adipocytes cell culture systems. In addition, other mouse models suggest that PPAR- γ signaling in the pancreas may influence insulin secretion.

α -Glucosidase Inhibitors

Acarbose and miglitol inhibit the intestinal enzyme α -glucosidase. Given with meals, this delays carbohydrate absorption, leading to a requirement for a lower maximal insulin level to dispose of the

absorbed carbohydrate. These drugs do not significantly affect fasting glucose levels, and so are generally ineffective as single agent therapy. However, there is accumulating information demonstrating the importance of controlling postprandial hyperglycemia to minimize long-term diabetic complications.

Incretin Hormones

Glucagon-like peptide-1 (GLP-1) agonists and dipeptidyl-peptidase-4 (DPP-4) inhibitors act in different ways to activate insulin secretion via the incretin pathway. The incretin pathway is rapidly activated when ingestion of oral carbohydrate stimulates production of gastric inhibitory peptide (GIP) and GLP-1, mainly by L-cells in the small intestine. GIP and GLP-1 are rapidly degraded by peptidases, primarily DPP-4, and cleared by the kidneys. These hormones, termed incretins, bind receptors on pancreatic α - and β -cells to stimulate insulin and decrease glucagon secretion, and also inhibit gastric emptying and decrease appetite. Incretins have also been shown to decrease β -cell apoptosis *in vitro* and promote β -cell proliferation in rodent models. Medications which mimic the action of GLP-1 (e.g., exenatide) or prolong the action of endogenous GLP-1 by inhibiting DPP-4 (e.g., sitagliptin) are now in clinical use and have been shown to be effective at improving blood glucose as first- or second-line agents in the treatment of type 2 diabetes. Whether the antiapoptotic effects seen in model systems are clinically significant in humans is under investigation.

Bromocriptine

Bromocriptine is a sympatholytic dopamine D2 receptor agonist. It has been demonstrated to produce improved glycemic control in patients with type 2 diabetes. The mechanism of this action is unknown, but it is presumed to act centrally to modulate neuronal signals from the central nervous system that regulate metabolism.

Treatments Directed at Diabetic Complications

The above medications are primarily used to improve hyperglycemia and insulin sensitivity in diabetes patients. It is also important to treat other risk factors that contribute to diabetes complications, particularly the macrovascular complications. Thus, treatment is directed at dyslipidemia and hypertension, important risk factors for atherosclerotic vascular disease. Inhibitors of hydroxymethylglutaryl-coenzyme A reductase, statins, decrease cholesterol biosynthesis and reduce the risk of cardiovascular disease. These agents may also act by reducing inflammation and oxidative stress in the vasculature, possibly partly via activation of AMPK. A number of other lipid lowering agents are effective. Of note, fibric acid derivatives

may be more effective in hyperlipidemic diabetic patients, and colesvelam (which acts by increasing cholesterol excretion in the form of bile acids) may also reduce hyperglycemia in diabetic patients.

An increase in the urinary loss of albumin is an early indicator of diabetic renal disease. Treatment of patients who have such microalbuminuria with angiotensin converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARBs) decreases the rate of progression to more significant renal impairment in such patients. Treatment of hypertension is undertaken to reduce the risk of macrovascular complications; because of their role in mitigating progression of diabetic nephropathy, ACE inhibitors and ARBs are often selected as the initial treatment for hypertension.

See also: **Bioenergetics:** Phosphatidylinositol-3-Phosphate; **Metabolism Vitamins and Hormones:** Glycogen Metabolism; Insulin- and Glucagon-Secreting Cells of the Pancreas; **Signaling:** Protein Kinase C Family.

Further Reading

- American Diabetes Association (2010) Diagnosis and classification of diabetes mellitus. *Diabetes Care* 33(supplement 1): S62–S69.
- Bell GI and Polonsky KS (2001) Diabetes mellitus and genetically programmed defects in β -cell function. *Nature* 414: 788–791.
- Brownlee M (2001) Biochemistry and molecular cell biology of diabetic complications. *Nature* 414: 813–820.
- Cefalu WT (2001) Insulin resistance: Cellular and clinical concepts. *Experimental Biology and Medicine* 226: 13–26.
- De Silva NM and Frayling TM (2010) Novel biological insights emerging from genetic studies of type 2 diabetes and related metabolic traits. *Current Opinion in Lipidology* 21(1): 44–50.
- Eizirik DL, Cardozo AK, and Cnop M (2008) The role for endoplasmic reticulum stress in diabetes mellitus. *Endocrine Reviews* 29(1): 42–61.
- Evans JL, Goldfine ID, Maddux BA, and Grodsky GM (2002) Oxidative stress and stress-activated signaling pathways: A unifying hypothesis of type 2 diabetes. *Endocrine Reviews* 23: 599–622.
- Gallagher EJ, LeRoith D, and Karnieli E (2008) The metabolic syndrome – from insulin resistance to obesity and diabetes. *Endocrinology Metabolism Clinics of North America* 37(3): 559–579.
- Hawley JA and Lessard SJ (2008) Exercise training-induced improvements in insulin action (review). *Acta Physiologica (Oxford)* 192(1): 127–135.
- Kahn SE (2003) The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of type 2 diabetes. *Diabetologia* 46: 3–19.
- Lin Y and Sun Z (2009) Current views on type 2 diabetes. *Journal of Endocrinology* 204(1): 1–11.
- Maechler P and Wollheim CB (2001) Mitochondrial function in normal and diabetic β -cells. *Nature* 414: 807–812.
- Saltiel AL and Kahn CR (2001) Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* 414: 799–806.
- Schimmack G, Defronzo RA, and Musi N (2006) AMP-activated protein kinase: Role in metabolism and therapeutic implications. *Diabetes, Obesity and Metabolism* 8(6): 591–602.
- Staiger H, Machicao F, Fritsche A, and Häring HU (2009) Pathomechanisms of type 2 diabetes genes. *Endocrine Reviews* 30(6): 557–585.

Diacylglycerol Kinases and Phosphatidic Acid Phosphatases

M K Topham, University of Utah, Salt Lake City, UT, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Stephen M. Prescott and Matthew K. Topham, volume 1, pp. 593–597, © 2004, Elsevier Inc.

Glossary

Diacylglycerol (DAG) A lipid composed of a glycerol backbone with fatty acids attached by ester bonds to its first (*sn*-1) and second (*sn*-2) carbons.

Fatty acid Monobasic acids containing long hydrocarbon chains that can be either saturated (no double bonds) or unsaturated (one or more double bonds).

Phosphatidic acid (PA) DAG with a phosphate group attached to the third (*sn*-3) carbon of the glycerol backbone.

Phosphatidylinositol PA with an inositol group attached to the phosphate.

Phospholipid The major structural lipids of most cellular membranes. They contain one or more phosphate groups and, if derived from glycerol, are known as 'phosphoglycerides'.

Diacylglycerol (DAG) is an important lipid that is both an intermediate in lipid biosynthetic pathways and can also initiate specific intracellular signaling events. The majority of signaling DAG is generated by hydrolysis of phosphatidylinositol-4,5-bisphosphate (PIP₂) by the enzyme phospholipase C (PLC), but DAG can also be generated when phosphatidic acid phosphatases (PAPs) remove the phosphate head group from phosphatidic acid (PA) (Figure 1). DAG activates protein kinase C (PKC) isoforms, binds to and activates the RasGRP nucleotide exchange factors, and recruits the chimaerins, which are Rac guanosine triphosphatase (GTPase)-activating proteins (GAPs), to membranes. Because DAG can modulate a plethora of signaling events, it is crucial that intracellular DAG levels be tightly regulated. Though DAG can be metabolized in various ways, under most circumstances its major route for metabolism is by its phosphorylation, a reaction that is catalyzed by the diacylglycerol kinases (DGKs).

The DAG Kinases

Ten mammalian DGK isoforms have been identified (Figure 2). The heterogeneity of the gene family is similar to the PKC and PLC families, suggesting that the DGKs are not simply lipid biosynthetic enzymes, but that they also have signaling roles, because enzymes involved in biosynthetic pathways usually do not have extended families. One or only a few DGK isoforms have been identified in organisms such as *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Arabidopsis thaliana*. Single-celled organisms such as yeast and bacteria also have DGKs, and a multi-substrate lipid kinase that can phosphorylate DAG also exists in vertebrates, but these enzymes are not similar to mammalian DGKs and they will not be discussed here.

DGK isoforms are categorized based on common structural motifs. All DGK isoforms have a catalytic domain composed of catalytic and accessory motifs that is necessary for kinase activity. In most cases, the catalytic domain is a single motif, but DGKs δ , η , and κ have bipartite catalytic domains. All DGKs have at least two cysteine-rich regions homologous to the C1A and C1B motifs of PKCs. In theory, these domains may bind

DAG, perhaps localizing DGKs to where DAG accumulates. However, no DGK C1 domain has so far been conclusively shown to bind DAG. Structural predictions suggest that most DGK C1 domains may not bind DAG, and most DGKs tested are unable to bind long-lived DAG-like analogs. Houssa and van Blitterswijk noted that in DGKs, the C1 domain closest to the catalytic domain is highly conserved, including an extended motif of 15 amino acids not present in other C1 domains. Conserved residues in this extended motif have been shown to be crucial for DAG kinase activity. Together, these data suggest that DGK C1 domains are structurally different from C1 domains in other proteins and that the different C1 domains of a single DGK isoform may have unique functions. Distinct functions of individual C1 domains remain to be definitively demonstrated.

In addition to the C1 and catalytic domains, other regulatory domains are used to group the DGKs into five subfamilies. Type I DGKs have calcium-binding EF hand motifs and have been shown to be more active in the presence of calcium. Type II DGKs have a pleckstrin homology (PH) domain at their amino termini. This domain in DGK δ has been shown to bind weakly and nonselectively to phosphatidylinositols (PIs). DGKs δ and η also have sterile alpha motifs (SAM domains) at their C-termini that can modulate their cellular localization or act as protein–protein interaction sites. DGK ϵ , a type III enzyme, has an unusual specificity toward acyl chains of DAG, strongly preferring a specific fatty acid – arachidonate – at the *sn*-2 position. This preference suggests that DGK ϵ may be a component of the biochemical pathway that accounts for the enrichment of PIs with arachidonate. Type IV DGKs have a MARCKS phosphorylation domain and, at their C-termini, four ankyrin repeats. The type V enzyme, DGK θ , has three C1 domains and a PH domain.

Regulation of DGK Activity

Activation of the DGKs is complex and unique for each DGK isotype. In most cases, DGKs must translocate to a membrane compartment to access DAG. In addition, their activity can be

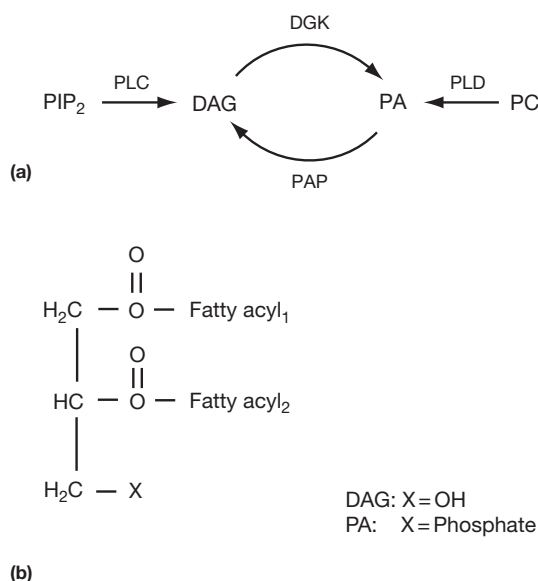


Figure 1 (a) Different enzymatic pathways can produce diacylglycerol (DAG) and phosphatidic acid (PA). Phospholipase C (PLC) enzymes generate DAG that can be phosphorylated by DGKs to produce PA. In another pathway, phospholipase D (PLD) hydrolyzes phosphatidylcholine (PC) – or phosphatidylethanolamine – making PA that can be further hydrolyzed by phosphatidic acid phosphatases (PAPs) to generate DAG. To date, there is little definitive evidence to suggest that the diacylglycerol kinase (DGK) and PAP reactions are coupled. Both DAG and PA can act as signaling lipids and are also intermediates in lipid biosynthetic pathways. (b) DAG and PA have the same general structure but contain different functional groups attached to the third (*sn*-3) carbon. Depending on their molecular structure, the fatty acyl groups confer different signaling properties to the DAG or PA.

modified by appropriate cofactors, and several DAG kinases are also regulated by posttranslational modifications. Finally, tissue-specific alternative splicing of DGKs β , γ , δ , ζ , ι , η , and possibly other isoforms allows additional opportunities for regulation. This complexity permits tissue- or cell-specific regulation of each DGK isotype, depending on the availability of cofactors and the type of stimulus that the cell receives.

DGK ζ is an example of the contextually dependent differential regulation of DGKs. DGK ζ translocates to at least two different membrane compartments in T lymphocytes depending upon the agonist used to activate the cells: from the cytosol to a perinuclear region in T cells stimulated with IL-2, and to the plasma membrane upon activation of the T-cell antigen receptor. Once at a membrane compartment, the DAG kinase activity of DGK ζ can be modified by the availability of several cofactors. Calcium is known to bind to EF hand structures and stimulates DGK ζ activity *in vitro*. Lipids also modify its activity: phosphatidylserine and sphingosine activate DGK ζ *in vitro* and likely *in vivo* as well. Finally, DGK ζ can be phosphorylated by several protein kinases, including some PKC isoforms and Src kinase, which may enhance its DAG kinase activity.

Like DGK ζ , other DGK isoforms appear to be sensitively regulated. For example, type II DGKs have a PH domain that may help localize these DGKs, modify their activity, or allow binding to other proteins. The PH domain of the type II

DGK δ binds PIs but its DAG kinase activity is not affected by these lipids. By contrast, the activity of DGK types III and IV can be modified by PIs and by phosphatidylserine, sometimes in opposing ways. For example, the type III enzyme, DGK ϵ , is inhibited by, whereas type IV DGKs are activated by, phosphatidylserine. Type IV DGKs are strongly regulated by subcellular translocation. These enzymes have a nuclear localization signal that is regulated by PKC phosphorylation, and there is evidence that the syntrophin family of scaffolding proteins further regulates their subcellular location by anchoring them in the cytoplasm. Finally, DGK θ , a type V DGK, can be regulated through its association with active RhoA, which abolishes its DAG kinase activity, while DGK ζ binds retinoblastoma family proteins, which stimulate its DAG kinase activity. Thus, depending on the context of activation, the availability of cofactors and binding partners, and the activation state of protein kinases, DGKs can be differentially regulated.

Activity of DGKs is Compartmentalized

Evidence suggests that DGK activity is restricted to the area of localized DAG pools generated after activation of receptors. This concept was initially tested in 1994 by Van der Bend et al., who detected DGK activity in cells following receptor activation but could not detect significant DAG kinase activity after treating the cells with exogenous PLC, which caused global, nonspecific DAG generation. Their data suggest that DGKs are active only in spatially restricted compartments following physiologic generation of DAG.

We are only beginning to understand the complexities of spatially restricted DGK function. Several DGKs localize to the cytoskeleton, where they may participate in regulating cytoskeleton dynamics. Houssa et al. showed that RhoA associates with DGK θ – but only when RhoA is in its active form – and that this interaction abolished DGK activity. RhoA appears to bind the catalytic domain of DGK θ , which probably physically impairs DAG kinase activity. Investigators have also noted that DGK activity associates with a complex of proteins including PI5K, Rac, Rho, Cdc42, and Rho-GDI, all of which regulate cytoskeleton dynamics. Evidence suggests that this might be DGK ζ , which helps promote Rac activation. Consistent with this possibility, DGK ζ localizes at the leading edge and cell extensions of glioblastoma cells and can disrupt cell spreading when overexpressed. We have also found in unpublished studies that other DGK isoforms co-immunoprecipitate with Rho family proteins when overexpressed in cells. The physiologic significance of these interactions is not clear, though, together, these data suggest that DGKs likely have a role in regulating the cytoskeleton.

In the nucleus, a distinct nuclear PI cycle is regulated separately from its plasma membrane/cytosolic counterpart. The function of DAG produced in the nucleus has received little attention, but data clearly indicate that, in most cases, it acts to promote cell growth independently of plasma membrane DAG. For example, some growth factors (e.g., IGF-1) can stimulate a pulse of nuclear DAG without causing a measurable change in extranuclear DAG, and several groups have demonstrated that nuclear DAG fluctuates independently of

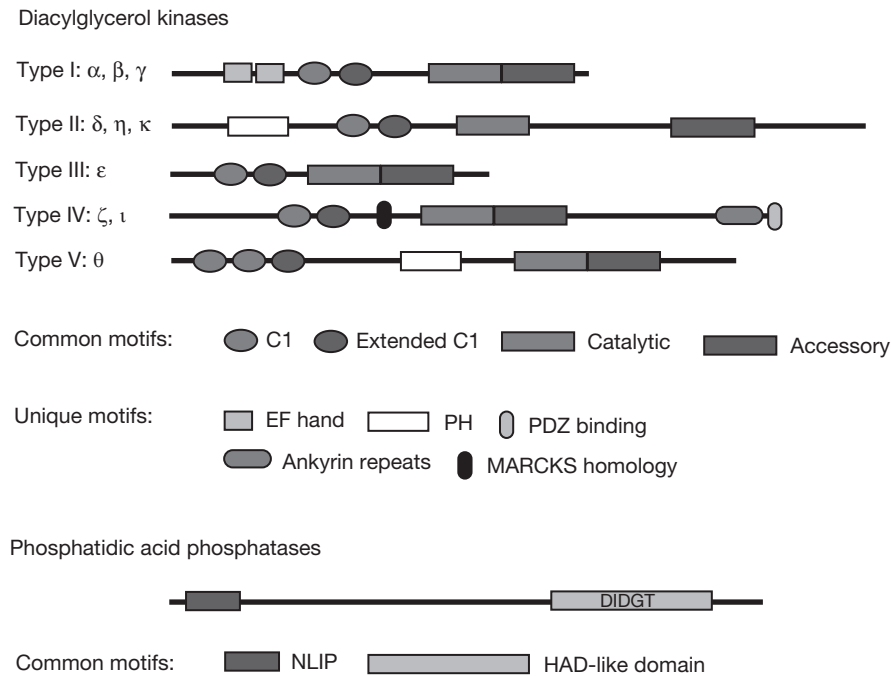


Figure 2 The mammalian DGK and PAP families. Based on structural motifs, the 10 mammalian DGKs are divided into five subtypes. Alternative splicing of some DGK isoforms generates even more structural diversity. Many of the DGKs contain other unique structural domains that are not shown. The PAP family is a smaller family of enzymes that have two motifs in common: the N-terminal NLIP domain is of unclear significance while the C-terminal haloacid dehalogenase (HAD)-like domain with a DIDGT motif is required for their activity.

extranuclear DAG during the cell cycle. Nuclear DAG was shown to peak shortly before S phase, suggesting that it may participate in the G1/S transition. Supporting this conclusion and emphasizing the importance of DAG signaling in the cell cycle, we demonstrated that when a nuclear DGK was over-expressed, cells accumulated at the G0/G1 transition, presumably because the kinase attenuated the nuclear DAG concentration. Within the nucleus, DAG signaling, like its extranuclear counterpart, appears to be compartmentalized. For example, in 1999 D'Santos et al. demonstrated independently fluctuating pools of nuclear DAG with distinct fatty acid compositions, strongly suggesting the existence of multiple, differentially regulated pools of nuclear DAG.

DGKs are present in nuclei and appear to have prominent and specific roles there. DGKs α , ζ , and ι translocate to the nucleus, while a fraction of DGK θ is localized there constitutively. Different DGK subtypes are confined to specific, separate compartments within the nucleus, suggesting specificity in their nuclear roles. For example, DGK α associates with the nuclear envelope, while DGK θ and DGK ζ are found in discrete regions within the body of the nucleus. Evidence suggests that DGKs likely affect nuclear signaling either by terminating DAG signals or by generating PA. For example, nuclear DGK ζ inhibits progression from G1 to S phase of the cell cycle, likely by metabolizing DAG. Conversely, in T lymphocytes, PA generated by nuclear DGK α appears to be necessary for IL-2-mediated progression to S phase of the cell cycle. Thus, DGK α and DGK ζ appear to have opposing roles in the nucleus, reflecting the complexity of lipid signaling and DGK activity there.

Coupling of DGKs with Other Signaling Proteins

Recent evidence suggests that in addition to acting locally at sites of DAG pools, DGKs can specifically associate with and regulate signaling proteins that are activated by either DAG or PA. For example, DGK ζ associates with and inhibits RasGRP1, a guanine nucleotide-exchange factor for Ras that requires DAG to function. This regulation may be selective: five other DGK isoforms did not significantly inhibit RasGRP1 activity. In *C. elegans*, Nurrish et al. found that *dgc-1*, an ortholog of human DGK θ , regulates DAG signaling that is necessary for acetylcholine release. Combined, these data demonstrate that DGK activity is not only targeted at sites of DAG production, but also specifically directed toward a subset of DAG-activated proteins.

DGK Gene Deletion Studies in Mice

In an effort to understand their function, the genes encoding several DGK isoforms have been deleted in mice. These studies have revealed that several DGKs modulate signaling events downstream of plasma membrane receptors. For example, DGKs α and ζ suppress immune cell activation through the T-cell receptor while DGK δ is required for proper EGFR signaling. DGKs ι and ζ modulate Ras signaling by virtue of their ability to bind and regulate RasGRP exchange factors. Finally, DGK ϵ was found to be necessary for the maintenance of several signaling lipids in neuronal cells, and deletion of DGK ϵ protected mice from induced seizures. Collectively, these studies

underscore the importance of DGKs in lipid signaling events and demonstrate the diversity of their functions.

PA as a Signal

The DGK reaction is of special interest because it removes one signaling molecule, DAG, while generating another, PA. PA can stimulate DNA synthesis and is potentially mitogenic. It modulates the activity of several enzymes, including PI 5-kinases (PI5Ks), PAK1, Ras-GAP, and PKC ζ , and it has a prominent role in vesicle trafficking. At the plasma membrane, PA also helps recruit Raf and Son of sevenless (Sos). Although phospholipase D (PLD) generates the bulk of signaling PA (Figure 1), DGKs likely also contribute to its intracellular concentration.

The PA species generated by DGK and PLD reactions are distinct from each other by virtue of their initial substrates. Phosphatidylcholine, which is largely composed of saturated and mono-unsaturated fatty acids is the predominant substrate of PLD, while PA produced by DGKs is derived from DAG that is enriched in polyunsaturated fatty acids, particularly arachidonate. There is good evidence that each PA species – saturated and unsaturated – can differentially activate targets. For example, saturated PA species induce mitogen-activated protein kinase (MAPK) activation to a greater extent than unsaturated PAs; Flores and colleagues presented evidence that the PA produced by DGK α was necessary for stimulated T lymphocytes to progress to S phase of the cell cycle. Abramovici et al. recently demonstrated that DGK ζ generated PA to activate PAK1 that then initiated events to alter actin dynamics. Thus, DGKs can influence signaling events either by metabolizing DAG or by generating PA.

PA Conversion to DAG by PA Phosphatases

Just as PA can be generated by more than one enzymatic pathway, DAG is also the product of several enzymatic reactions. The PLC reaction is generally thought to be the major route to generate signaling DAG, but some lipid phosphatases can also produce DAG by dephosphorylating PA (Figure 1). These phosphatases fall into two groups. One group is composed of integral membrane proteins called 'lipid phosphate phosphatases (LPPs)', which are promiscuous and hydrolyze a diverse array of lipids that include PA, lysophosphatidic acid, sphingosine-1-phosphate, and ceramide-1-phosphate. The other group of enzymes is called 'PAPs'. These enzymes have specificity for PA and will be the focus of this section.

PA Phosphatase Enzymes

PAP activity in cellular lysates was originally identified in the 1950s, but the enzymes were not cloned until 2006, when the yeast protein encoded by the *PAH1* gene was shown to have PAP activity. Pah1p, the protein product of the *PAH1* gene, has a predicted molecular weight of 95 kDa and contains a haloacid dehalogenase (HAD)-like domain with a DIDGT motif that is crucial for phosphatase activity. There is also an

N-terminal NLIP domain of unknown significance (Figure 2). Based on sequence homology to Pah1p, a group of mammalian enzymes called 'lipins' were subsequently recognized to have PAP activity, which is now believed to be their primary enzymatic function. There are three known mammalian lipins – lipin-1, -2, and -3 – and alternative splicing yields two lipin-1 products called lipin-1A and lipin-1B. Lipin and Pah1p orthologs are now recognized in most vertebrates and invertebrates, and all of these enzymes strictly require Mg²⁺ for activity. In humans, the lipin enzymes are expressed in numerous tissues, with lipins-1 and -2 being predominantly expressed in adipose tissue.

Physiologic Roles of PA Phosphatases

Given the central role of the PAP reaction in the biosynthesis of many cellular lipid components, it is not surprising that mutant yeast lacking PAP activity accumulate PA and have altered levels of numerous lipids including DAG and triacylglycerol. These lipid abnormalities verified that PA phosphatases are crucial for proper lipid biosynthesis. Prior to the identification of Pah1p as a PA phosphatase, *Lpin1* (the gene encoding lipin-1) was identified as the mutated gene in fatty liver dystrophy (*fld*) mice. The *fld* mice develop numerous abnormalities including fatty liver, insulin resistance, lipodystrophy, and susceptibility to atherosclerosis. These abnormalities indicate that lipin-1, like its ortholog in yeast, has a central role in lipid biosynthesis. Additional studies have demonstrated that lipin-2 also has an important function in hepatic lipid homeostasis, while lipin-3 has not been studied as extensively.

In addition to their roles in lipid biosynthetic pathways, it is possible that PA phosphatases might also function in lipid signaling pathways. Supporting this possibility, there are several reports demonstrating agonist-induced phosphorylation of PA phosphatases or translocation of Mg²⁺-dependent PA phosphatase activity. Specifically, it has been postulated that PA phosphatases might generate DAG from PA that could then activate PKCs and other proteins. Indeed, DAG levels fall in the setting of PA phosphatase deficiency. But there is a distinction between the DAG that is an intermediate in lipid biosynthetic pathways and the DAG that acts as a signaling lipid. The lipid abnormalities in PA-phosphatase-deficient yeast and mice suggest that PA phosphatases are predominantly active in lipid biosynthetic pathways. So far, there is little published work supporting that DAG derived from the PA phosphatase reaction can act as a signaling lipid. Most data suggest that PA phosphatase activity is coupled to PLD activity, which by virtue of its substrates, predominantly generates PA with saturated fatty acids. As such, DAG that is generated by sequential PLD and PAP activity would be composed of predominantly saturated fatty acids. But evidence suggests that DAG composed of saturated fatty acids has weak signaling properties compared to unsaturated DAG. So based on present data, PAPs might generate DAG, but because of its fatty acid components, this DAG is not thought to initiate signaling. Instead, the PAP reaction probably functions to terminate PA signals that were initiated by PLD. It is also possible that PAP activity could couple with DGK activity to regenerate unsaturated DAG that

might initiate signaling, but no one has demonstrated that this cycle exists. Thus, PAPs theoretically could contribute to DAG signaling but evidence is lacking that they do so. However, this issue needs to be explored more definitively. Instead, it is more likely that PAPs predominantly function in lipid biosynthetic pathways or serve to terminate PA signaling when coupled with PLDs.

Conclusions

As evidence accumulates, it is apparent that the structural diversity of DGK isoforms mirrors the range of functions and sites of action of these enzymes. Further studies of the regulation of lipid signaling should reveal additional roles for DGKs and more precise information about their regulation. PAPs are a family of enzymes that have been known to exist for decades but were only recently cloned. They are a less diverse family of enzymes but have crucial functions in lipid biosynthetic pathways. Their role in lipid signaling is less clear, but PAPs probably terminate PA signals generated by the PLD reaction. Although it is possible that PAPs can contribute to DAG signaling, there is little definitive evidence at this point to demonstrate that they do.

See also: **Signaling:** Phosphatidylinositol Bisphosphate and Trisphosphate; Phospholipase C; Phospholipase D.

Further Reading

- Brindley DN, Palquil C, Sariahmetoglu M, and Reue K (2009) Phosphatidate degradation: Phosphatidate phosphatases (lipins) and lipid phosphate phosphatases. *Biochimica et Biophysica Acta* 1791: 956–961.
- Carman GM and Han G-S (2009) Phosphatidic acid phosphatase, a key enzyme in the regulation of lipid synthesis. *Journal of Biological Chemistry* 284: 2593–2597.
- Hodgkin MN, Pettitt TR, Martin A, et al. (1998) Diacylglycerols and phosphatidates: Which molecular species are intracellular messengers? *Trends in Biochemical Sciences* 23: 200–204.
- Merida I, Avila-Flores A, and Merino E (2008) Diacylglycerol kinases: At the hub of cell signaling. *Biochemical Journal* 409: 1–18.
- Nurrish S, Segalat L, and Kaplan JM (1999) Serotonin inhibition of synaptic transmission: G_{α_o} decreases the abundance of UNC-13 at release sites. *Neuron* 24: 231–242.
- Sakane F, Imai S, Kai M, Yasuda S, and Kanoh H (2007) Diacylglycerol kinases: Why so many of them? *Biochimica et Biophysica Acta* 1771: 793–806.
- Sanjuan MA, Jones DR, Izquierdo M, and Merida I (2001) Role of diacylglycerol kinase α in the attenuation of receptor signaling. *Journal of Cell Biology* 153: 207–219.
- Sciorra VA and Morris AJ (2002) Roles for lipid phosphate phosphatases in regulation of cellular signaling. *Biochimica et Biophysica Acta* 1582: 45–51.
- Topham MK and Prescott SM (2001) Diacylglycerol kinase ζ regulates Ras activation by a novel mechanism. *Journal of Cell Biology* 152: 1135–1143.
- Topham MK and Prescott (2002) Diacylglycerol kinases: Regulation and signaling roles. *Thrombosis and Haemostasis* 88: 912–918.
- van Blitterswijk WJ and Houssa B (2000) Properties and functions of diacylglycerol kinases. *Cell Signalling* 12: 595–605.
- Van der Bend RL, de Widt J, Hilkmann H, and van Blitterswijk WJ (1994) Diacylglycerol kinase in receptor-stimulated cells converts its substrate in a topologically restricted manner. *Journal of Biological Chemistry* 269: 4098–4102.

Disulfide Bond Formation

H F Gilbert, Baylor College of Medicine, Houston, TX, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 1, pp. 598–602, © 2004, Elsevier Inc.

Glossary

Bacterial periplasm A specialized compartment in between the inner and outer membrane of bacteria. Disulfide formation occurs in this compartment.

Disulfide bond A covalent bond between two sulfhydryl groups. In proteins, this bond covalently connects the sulfur atoms of two cysteine residues.

Endoplasmic reticulum A specialized compartment of eukaryotic cells in which extracellular proteins are processed for secretion or inclusion in other cellular compartments.

Glutathione (GSH) A tripeptide, γ -glutamylcysteinylglycine found at high concentrations in cells. It is the major thiol compound in most cells involved in redox reactions and the

protection of cells against reactive oxygen species and toxic compounds.

Oxidation Disulfide bond formation (formally, the loss of electrons) oxidation.

Oxidative folding The coupling of disulfide formation to the folding of a protein into its correct three-dimensional (3-D) structure. Disulfides stabilize protein structure and vice versa.

Protein folding The process by which an unfolded protein gains its 3-D structure.

Reduction Formally, the gain of electrons. Disulfide bonds are reduced to two thiols.

Thiol A sulfhydryl (–SH) group.

Chemistry of Disulfide Formation as a Posttranslational Modification

Disulfide Formation

Extracellular and secreted proteins often contain many disulfides, with pairs of cysteines linked to each other in a specific configuration in the folded protein (Figure 1). Biochemically, disulfide formation is an oxidation (the loss of electrons). Not all of the cysteines in a protein form disulfides. Disulfides form only when the structure of the protein places two cysteines in the proper spatial location and when an oxidizing agent is available. The chemical reaction is reversible, and high concentrations of a reductant can reverse or prevent disulfide formation.

Disulfide Rearrangements

Through a chemical process called thiol–disulfide exchange, one disulfide may serve as an oxidizing agent to form a disulfide between two different cysteines (Figure 2). If this leads to connecting the cysteines in a different configuration, it is termed disulfide isomerization.

Oxidative Protein Folding

Since many disulfides will be buried in the core of the protein and link cysteines that are not adjacent in sequence, disulfide formation and rearrangements often occur as the protein attains a three-dimensional (3-D) structure (protein folding). The coupling of disulfide formation and protein folding is termed oxidative protein folding.

Structure Is Linked to Disulfide Formation

In the 1960s, Anfinsen and his colleagues used the reversible nature of disulfide formation to demonstrate that the primary

sequence of a protein contains sufficient information to define the final, biologically active structure of a protein. Disulfide bonds stabilize protein structure by organizing and destabilizing the denatured protein relative to the native structure. Usually, both the disulfides and noncovalent interactions (hydrophobic interaction, hydrogen bonds, etc.) are needed to specify the structure of ribonuclease A (RNase). Reducing the disulfides or using urea to disrupt the noncovalent structure causes a loss in biological activity. After removing the urea and including an oxidizing agent to reform the disulfides, the protein spontaneously refolds, forming the correct disulfides and a biologically active structure. However, if disulfide formation is permitted without removing the urea, random disulfides form (scrambled RNase) and there is no regain of biological activity. If disulfide formation is inhibited (by including a reducing agent), the native structure does not form when the urea is removed. Disulfide bonds are essential for forming the native structure but when the noncovalent structure is disrupted by denaturants disulfides are usually not sufficient to specify the tertiary structure.

Disulfide Bond Formation during Protein Folding

Even with two disulfides, there are numerous ways to connect the cysteines (Figure 3), and the problem increases greatly as the number of disulfides to be formed increases. Early in protein folding, disulfide formation tends to be error prone. Two types of errors have been found: (1) the disulfides may be formed between the wrong pairs of cysteines or (2) the correct cysteines may be paired but in a temporal order that interferes with the assembly of the rest of the structure. In both cases, the incorrect disulfides are replaced with the correct ones through disulfide isomerization. In the cases examined so far, many incorrect disulfides accumulate early in oxidative folding. These are slowly replaced by correct disulfides through isomerization which requires the breaking and reforming of disulfide bonds.

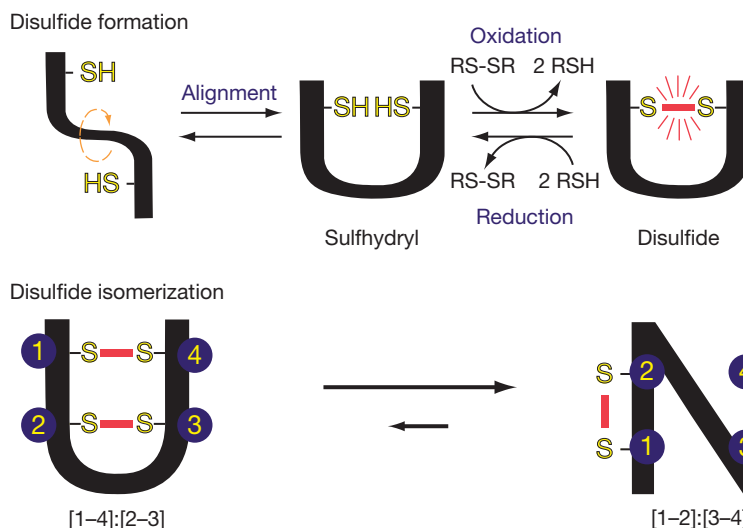


Figure 1 Disulfide formation and isomerization. An alignment step is needed to bring two sulfhydryl groups from cysteine residues into close proximity. Disulfide formation represents an oxidation where the electrons from disulfide formation must be transferred to an oxidizing agent. Disulfide isomerization does not formally require a change in the oxidation state of the protein; however, a thiol group is needed to initiate and propagate the change in disulfide connectivity.

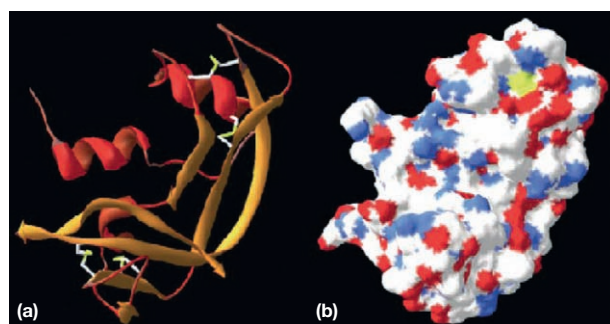


Figure 2 Structure of disulfide bonds: (a) disulfide bonds cross-link cysteines in various parts of the protein structure, (b) some disulfides (yellow) are exposed to solution in the tertiary structure of the protein but others are buried in the core of the protein and not exposed to solution.

Catalysis of Disulfide Formation

Soon after discovering that native proteins can be regenerated from reduced, denatured ones, Anfinsen realized that *in vitro* oxidative folding was often too slow (hours to days) to be biologically relevant. The resulting search for a biological catalyst for oxidative folding led to the discovery of the first catalyst of protein folding, protein disulfide isomerase (PDI).

Protein Disulfide Isomerase

PDI accelerates the regain of native structure by catalyzing the formation of disulfide bonds and by providing a way to correct mistakes in connecting cysteines. PDI is an enzyme that has both oxidase (disulfide formation) and isomerase (disulfide rearrangement) activities. PDI accelerates thiol/disulfide exchange reactions (Figure 1) with proteins and small-molecule substrates. The final outcome of the reaction will depend on the redox conditions. The catalytic sites are contained

in two thioredoxin homology domains located at each end of the PDI molecule. These active sites, both with the sequence CGHC, have the correct redox potential to effectively catalyze oxidation, reduction, and exchange reactions under the redox conditions of the endoplasmic reticulum. PDI interacts with its protein substrates but in a way that does not direct which cysteines are connected in the substrate. The mechanism involves several rounds of making and breaking disulfide bonds until the final structure becomes resistant to further isomerization of its disulfides.

The Thioredoxin Family

Many of the thiol–disulfide exchange reactions of the cell are mediated by members of the thioredoxin family. This large family of proteins is characterized by an active site sequence of CXXC and an α - β secondary structure. The structure greatly affects the redox potential for reducing the active site disulfide. Thioredoxin, the patriarch of the family, is an excellent reducing agent involved in the biosynthesis of deoxyribonucleotides for DNA synthesis and in helping maintain reduced proteins in the cellular cytoplasm. DsbA, a periplasmic oxidase in bacteria, is an excellent oxidizing agent that inserts disulfides into secreted proteins. The ease of forming the CXXC disulfide varies by a factor of over 10^5 , corresponding to a redox potential difference of more than 0.15 V.

Disulfide Formation in the Cell

Disulfide bond formation state is influenced by the thiol/disulfide redox environment of the cellular compartment where the protein is located. Most disulfides are found in secreted or extracellular proteins where the environment is oxidizing. Generally, the cytoplasm of most cells, bacterial and eukaryotic, is very reducing, due to the high concentration

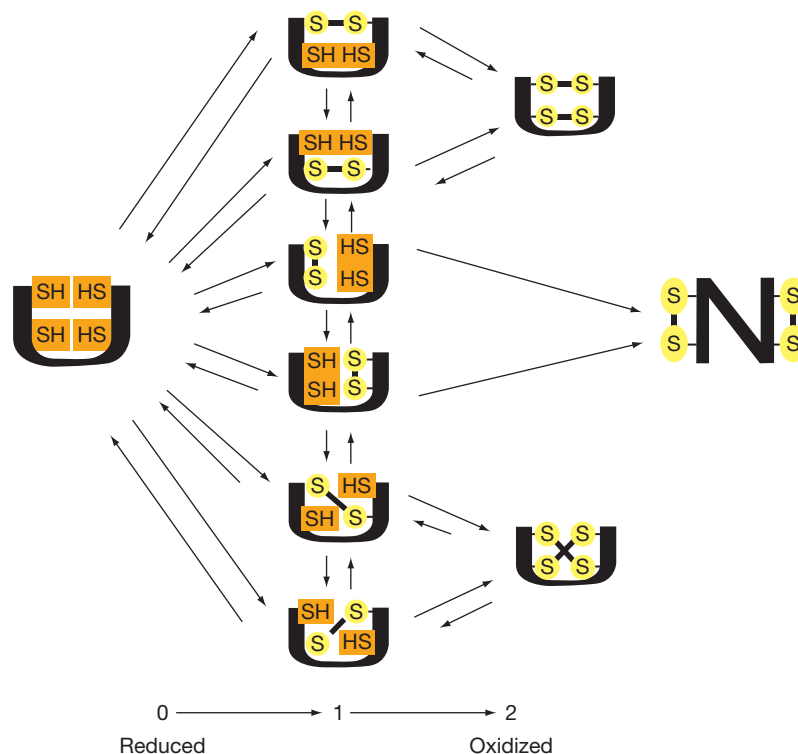


Figure 3 Generalized mechanism of oxidative protein folding. The initial stages of oxidative folding are error prone and often lead to the connection of cysteines that are not connected in the native structure. Different proteins will generate a different population of misoxidized species which are converted to the native structure (N) by disulfide bond isomerization.

of glutathione (GSH, a tripeptide, γ -glutamylcysteinylglycine) and the low concentration of glutathione disulfide.

Extracellular Proteins

Proteins destined for the extracellular environment acquire disulfides after they are secreted from the cytoplasm during translation. In bacteria, an elaborate system of periplasmic and membrane-bound proteins, comprising the Dsb system, provides the periplasm with oxidizing equivalents to enable disulfide formation and with reducing equivalents and catalytic isomerases to correct errors during protein folding (Figure 4). DsbB couples the bacterial electron transport system to disulfide formation, providing a source of oxidizing agent which are delivered directly to folding proteins by DsbA. Reducing equivalents to enable disulfide isomerization are provided by thioredoxin in the cytosol and transferred across the inner bacterial membrane by DsbD to DsbC, the periplasmic isomerase.

A similar system operates in eukaryotic cells, although all the elements of the system have not been found. Extracellular proteins are synthesized on the rough endoplasmic reticulum and translocated into the lumen (inside) during translation. The lumen of the endoplasmic reticulum is a more oxidizing intracellular compartment than the cytoplasm. Oxidizing equivalents are transferred from Ero1 to PDI and then to the folding proteins. The ultimate source of the oxidizing agent is not yet known. Reducing agents of the endoplasmic reticulum enable PDI and other catalysts to correct any errors in pairing

cysteines by reducing the incorrect disulfide bond. Reducing agents are provided by glutathione, presumably by importing GSH from the cytoplasm. The endoplasmic reticulum has an elaborate quality-control mechanism, which prevents misfolded proteins from exiting the endoplasmic reticulum or directs them to degradation pathways. The error-correction activity of PDI plays a critical role in this system. Depending on the details of the folding mechanism dictated by the primary sequence, different extracellular proteins may use different components of the complex folding and quality-control apparatus of the endoplasmic reticulum to gain their correct 3-D structures, including the disulfide bonds that stabilize the structures.

Disulfide Formation as a Regulatory Mechanism

Because disulfide bond formation is reversible, disulfide bonds can also regulate biological activity through their ability to stabilize specific protein structures. For example, the bacterial transcription factor, OxyR, senses the redox environment of the aerobic bacterial cell through reversible disulfide bond formation linked to the cellular redox state and oxygen metabolism. Reduced OxyR is inactive as a transcription factor and upon oxidation and disulfide formation, the transcription factor can regulate the transcription of genes involved in the protection of the cell against oxidative stress and reactive oxygen species. Similar mechanisms operate in eukaryotic cells.

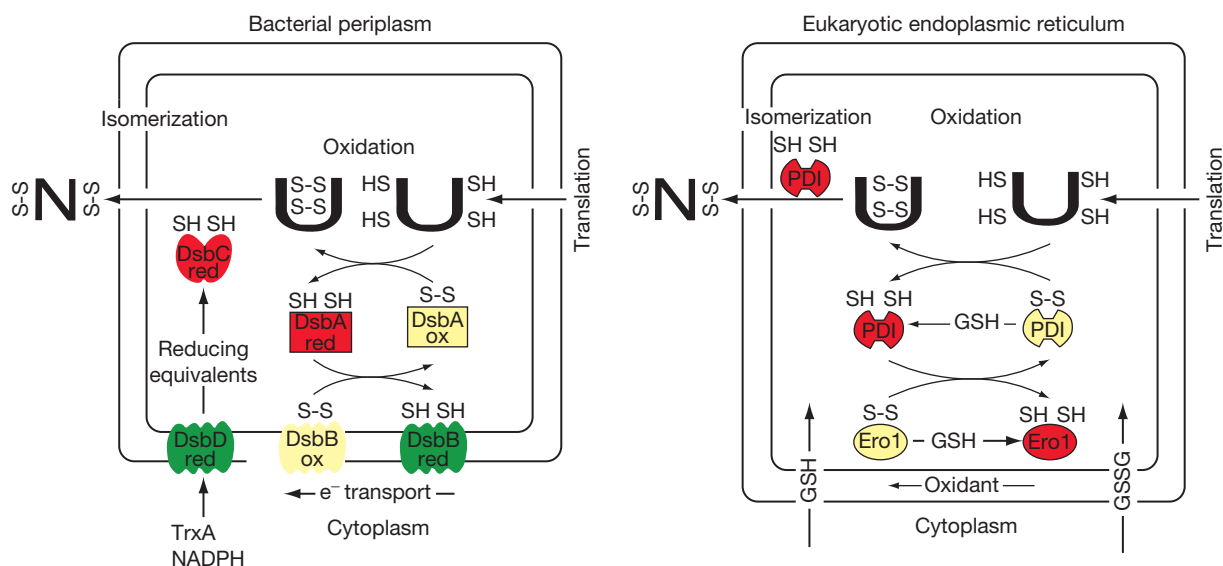


Figure 4 Disulfide bond formation in bacterial and eukaryotic cells. In both cells, the major pathways for disulfide formation are located in specialized cellular compartments where the redox state of the environment is maintained in a more oxidized state than the cytosol.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Glycoprotein Folding and Processing Reactions; **Protein/Enzyme Structure Function and Degradation:** Protein Folding and Assembly.

Further Reading

- Collet J-F and Bardwell JCA (2002) Oxidative protein folding in bacteria. *Molecular Microbiology* 44: 1–8.
- Creighton TE (1984) Disulfide bond formation in proteins. *Methods in Enzymology* 107: 305–329.
- Frand AR, Cuzzo JW, and Kaiser CA (2000) Pathways for protein disulphide bond formation. *Trends in Cell Biology* 10: 203–210.

- Gilbert HF (1998) Protein disulfide isomerase. *Methods in Enzymology* 290: 26–50.
- Gitler C and Danon A (eds.) (2003) *Cellular Implications of Redox Signaling*. London: Imperial College Press.
- Goldberger RF, Epstein CJ, and Anfinsen CB (1963) Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. *Journal of Biological Chemistry* 238: 628–635.
- Helenius A (2001) Quality control in the secretory assembly line. *Philosophical Transactions of the Royal Society B: Biological Sciences* 356: 147–150.
- Ritz D and Beckwith J (2001) Roles of thiol-redox pathways in bacteria. *Annual Review of Microbiology* 55: 21–48.
- Scheraga HA, Konishi Y, and Ooi T (1984) Multiple pathways for regenerating ribonuclease A. *Advances in Biophysics* 18: 21–41.
- Weissman JS and Kim PS (1991) Reexamination of the folding of BPTI: Predominance of native intermediates. *Science* 253: 1386–1390.

ENCYCLOPEDIA OF **BIOLOGICAL CHEMISTRY**

SECOND EDITION

Volume 2

DNA–L

DNA Base Excision Repair Pathways

R A O Bennett and B Dемple, Stony Brook University School of Medicine, Stony Brook, NY, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Hilde Nilsen and Tomas Lindahl, volume 1, pp. 603–608, © 2004, Elsevier Inc.

Glossary

Deamination The hydrolytic elimination of an amino group from a base or other compound.

Free radical A molecule, ion, or an atom having an unpaired valence electron that is typically highly reactive and short lived.

Hydrolysis Chemical decomposition in which a compound is split into other compounds by reaction with water.

Schiff base The metastable product of the chemical association of an aldehyde with a primary amine.

Introduction and Overview

Base excision DNA repair (BER) is one of the four major DNA repair pathways of mammalian cells, and in biology generally. BER systems process a majority of smaller lesions that do not greatly distort the helical structure of DNA. By contrast, nucleotide excision repair seems to be directed at lesions that cause significant helix distortion (e.g., pyrimidine photodimers), while mismatch repair detects primarily replication errors and provides ancillary functions in some other pathways, and recombinational repair and end-joining pathways handle DNA double-strand breaks. Additional repair pathways correct very specific types of DNA damage, such as the demethylation proteins that directly reverse mutagenic products such as O6-methylguanine or 1-methylthymine.

The DNA lesions targeted by BER include most of those due to oxidative damage, a major threat in aerobic organisms; N-alkylated bases that result from reactions with metabolites and environmental agents; and the DNA lesions that result from hydrolytic reactions that occur relentlessly in all organisms (e.g., depurination resulting in base loss or deaminations that alter the composition of heterocyclic bases). Specific examples of DNA damage efficiently processed by BER include the deamination product uracil in DNA, derived from cytosine; the oxidative lesions 8-oxo-guanine (8-oxoG; structure 5 in Figure 1), thymine glycol (structure 1 in Figure 1), and the ring-fragmentation product *N*-(2,4-diamino-6-oxo-5H-pyrimidin-5-yl)formamide or formamidopyrimidine (structure 2 in Figure 1); and the alkylation product N3-methyladenine (structure 7 in Figure 1). See Figure 1 for the structures of selected base alterations and abasic sites.

The steps of the simplest BER pathway were worked out for the repair of uracil residues in DNA, and they apply to many other lesions. Uracil in DNA is recognized by uracil–DNA glycosylase, the first such enzyme discovered, which hydrolyzes the base–sugar glycosylic bond. Several forms of uracil–DNA *N*-glycosylase are now known and, in total, 10 DNA glycosylases act on a range of oxidized, alkylated, or hydrolytic base derivatives. In addition, the MUTHY glycosylase (homologous to *Escherichia coli* MutY) specifically excises undamaged adenine opposite 8-oxoguanine residues, which offsets possible mutation when an oxidized DNA template is copied by DNA polymerase.

Removal of damaged or inappropriate bases by DNA glycosylases yields abasic sites (structure 3 in Figure 1) (specifically, apurinic or apyrimidinic sites, which we will term collectively AP sites). These are recognized by AP endonuclease enzymes, the most common forms of which cleave the 5′ phosphodiester of the abasic site to generate a single-strand break bracketed by a nucleotide-3′-hydroxyl and 5′-deoxyribose-5-phosphate ester (the AP site). Other enzymes may also incise abasic sites, which might lead to alternative processing (see section [The Core Enzymes of BER](#)).

The 3′-terminus at the AP endonuclease-generated strand break is then extended by DNA polymerase to replace the missing nucleotide; in mammalian cells, this reaction is carried out primarily by DNA polymerase β (Polβ) during BER in the nucleus. Before DNA ligation can complete the repair, the 5′-abasic residue or a displaced single strand containing it must be removed, and there are alternative subpathways that achieve this removal. Finally, DNA ligase joins the ends to restore the original DNA structure.

Sources and Occurrence of Lesions Handled by BER

As noted in the preceding section, there are important endogenous sources of DNA lesions that are processed in BER pathways. Some sources can be modulated if not completely eliminated (e.g., oxygen radicals in aerobic organisms), but at least one is simply unavoidable: the water that drives hydrolytic depurination and base deamination reactions. Beyond this endogenous burden of lesions, DNA is subject to damage by many environmental agents, including ultraviolet and ionizing radiation (e.g., X-rays) and environmental compounds such as methyl chloride. Still other environmental agents exacerbate cellular processes to inflict DNA damage. Important examples are the redox-cycling quinones and the herbicide paraquat, which set up a ‘redox cycle’ dependent on cellular reductases to generate superoxide radicals, thus exerting powerful oxidative stress in the cell.

Many of the same lesions result from both endogenous and environmental sources. For oxidative DNA damage, dozens of different base lesions have been identified, largely through the work of radiation chemists. Most well known among these are the highly mutagenic 8-oxoG, thymine glycols,

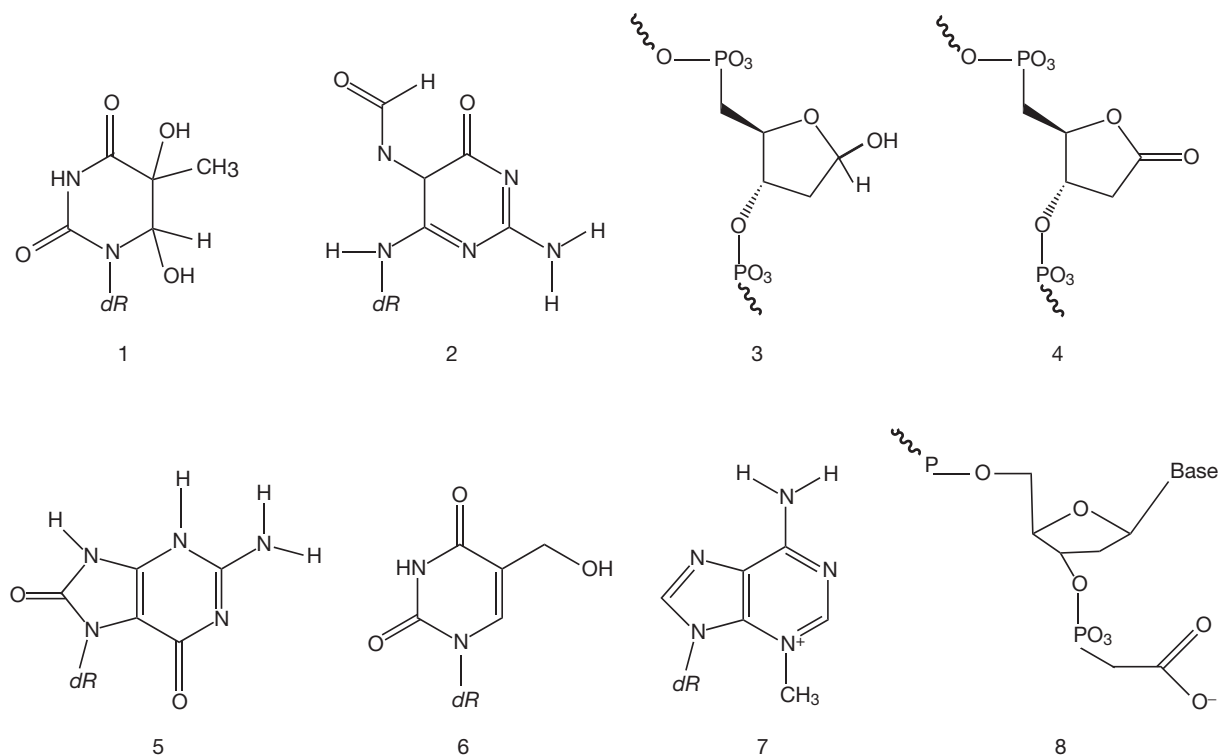


Figure 1 Structures of some selected DNA lesions processed by BER. Note that many more have been described: the examples shown here were chosen because their repair has been extensively characterized, or they have notable biological importance. 1, thymine glycol (representing multiple stereoisomers); 2, the formamidopyrimidine *N*-(2,4-diamino-6-oxo-5H-pyrimidin-5-yl)formamide; 3, an AP site (actually a mixture of the *a* and *b* anomers, which are in equilibrium); 4, 2-deoxyribonolactone (dL); 5, 8-oxoguanine (8-oxoG); 6, 5-hydroxymethyluracil; 7, 3-methyladenine; 8, a 3'-phosphoglycolate ester.

formamidopyrimidine derivatives of both guanine and adenine, and 5-hydroxymethylcytosine (structure 6 in [Figure 1](#)), but none constitutes more than a few percent of the total oxidative damage. Chemical oxidants such as H_2O_2 produce most of the same base lesions as ionizing radiation. Similarly, DNA alkylation produces an array of lesions, with 3-methyladenine having a substantial cytotoxic effect, while the more abundant 7-methylguanine accounts for a much greater proportion. It should be noted, though, that 7-methylguanine can undergo alkali-catalyzed opening of the imidazole ring to generate a methylated formamidopyrimidine, analogous to the fragmented purines formed by oxygen radicals.

Abasic sites are generated not only by DNA glycosylases, but also by acid-catalyzed hydrolysis liberating undamaged purines from DNA and, to a much lesser extent, pyrimidines. Modified bases such as 3-methyladenine have destabilized glycosylic bonds that promote their release. *In vitro* DNA depurination measurements gave an estimated rate of about 10 000 hydrolytic depurination events per day in a mammalian genome. Other endogenous and environmental sources of DNA damage would add to this burden. Free radicals also directly generate oxidized forms of abasic sites, including fragmentary products. BER enzymes deal with these lesions, with some oxidized abasic residues requiring processing by a special subpathway of BER.

Some radiomimetic drugs used in chemotherapy, such as bleomycin and calicheamicin, operate by oxidizing the deoxyribose sugar in DNA, leading to loss of the base. The cytotoxic

effectiveness of these agents depends on their ability to generate such lesions in clusters, which is also true of X-rays, with the hallmark production of double-strand breaks. Double-strand breaks and probably other types of lesion clusters need the intervention of recombinational repair pathways for their correction.

Finally, there are biological processes in the cell that actively generate BER lesions. Although the enzyme dUTPase (dUTP pyrophosphatase, the human *DUT* gene product) efficiently excludes dUTP from the pool of DNA synthesis precursors, small amounts of this nucleotide are inserted by DNA polymerases, which use them relatively efficiently. The nuclear form of mammalian uracil-DNA *N*-glycosylase is associated with replication centers, where it is poised to remove the occasional uracil residue incorporated during DNA synthesis. It is worth noting that uracil incorporated in this way, opposite template A residues, is not *per se* mutagenic. This situation contrasts with uracil formation by the deamination of cytosine in duplex DNA to generate a G:U base pair which, if processed by replication, would produce an A:U base pair in one product, and eventually an A:T transition mutation. Active mutagenic deamination occurs in the vertebrate immune system during antibody diversification in germinal centers (somatic hypermutation) and immunoglobulin class switching, where the mutagenic effect is desirable. It remains uncertain why BER does not counteract this process effectively.

The nucleotide pool can also react with the same agents that damage DNA itself. In this way, some 8-oxo-dGTP is generated,

which can be incorporated by DNA polymerases, occasionally opposite A in the template in a mutagenic reaction. This threat is largely offset by MutT-type (mammalian MTH1) enzymes that hydrolyze 8-oxo-dGTP to 8-oxo-dGMP, which is, thus, analogous to the cleansing of the nucleotide pool carried out by dUTPase. Localization of mammalian MUTYH to replication forks provides a second level of defense against this type of mutagenesis.

The-Core Enzymes of BER

DNA Glycosylases

To date, 11 distinct DNA glycosylase enzymes have been identified, which can be put into two broad functional categories (Table 1). All of the enzymes excise the target base by cleaving the N-glycosylic bond that connects the base to C1' of the deoxyribose sugar, resulting in an AP site in DNA. Most of the 'pure' glycosylases activate water for the hydrolysis reaction. However, five enzymes have both DNA glycosylase activity and AP lyase activity, enabling them to cleave the AP sites they produce, as well as AP sites generated by other routes. These enzymes act through a covalent Schiff base intermediate, linked to an amine (usually lysine) in the enzyme. The Schiff base intermediate is resolved by hydrolysis, which leaves a 2–3 unsaturated abasic residue attached to a 3' end at the nicked site.

This glycosylase/lyase group corresponds to the enzymes that act on oxidative lesions, oxidized or fragmented forms of purines and pyrimidines, including 8-oxoG, thymine glycols, and formamidopyrimidines. The prompt introduction of a strand break by these enzymes might provide a generalized mark of the damage, although that idea does not address why the glycosylases for uracil and methylated purines do not share

such an activity. Moreover, although the AP lyase activities can be detected *in vitro*, their contribution to AP site cleavage *in vivo* is unclear.

The MUTYH glycosylases are unique, in that they share structural features with the glycosylase/lyase enzymes, but act only as glycosylases. It seems likely that the lyase activity was lost as this enzyme evolved, which again puts the question of the role of the lyase functions in general.

Collectively, the DNA glycosylases remove a wide array of lesions, and some are present both in the nucleus and in mitochondria (see section BER in Mitochondria). The degree of specificity varies among the enzymes. Uracil–DNA glycosylase acts on a fairly narrow range of structurally related substrates, and maximally on uracil itself. By contrast, Nth-type enzymes remove a broad spectrum of lesions derived from pyrimidines, from thymine glycol and pyrimidine hydrates, to the fragmentation/contraction product 5-hydroxy-5-methyl hydantoin, and even urea, a minimal oxidation fragment. Evolution has further refined specificity in some cases. Methylpurine glycosylases are most active on 3-methyladenine, but they can also excise hypoxanthine (deaminated adenine). However, the mammalian MPG enzymes (Table 1) also have significant activity against longer alkyl chains (ethyl, propyl) and some exocyclic adducts (1, N6-etheno-adenine), and remove these more effectively than the *E. coli* enzyme. The bacteria instead contain a second protein AlkA that expands the repertoire of alkylated substrates.

AP Endonucleases

We distinguish AP endonucleases as enzymes that activate water to mediate cleavage of the phosphodiester bonds, and produce a nick with 5'-deoxyribose-5-phosphate. This contrasts with the glycosylase/lyases that produce nicks with the unsaturated abasic derivative attached to a 3' end (Figure 2). Cleavage of the latter type can even occur nonenzymatically in reactions catalyzed by small primary amine compounds.

Enzymatic and molecular genetic studies have revealed that AP endonucleases comprise two phylogenetic groups, each including both prokaryotic and eukaryotic members (Table 2). The prototypic proteins of the two groups are the two main AP endonucleases of *E. coli*: exonuclease III, encoded by the *xth* gene, and endonuclease IV, encoded by the *nfo* gene. The former enzyme was first discovered as an exonuclease and 3'-phosphatase; hence, the name exonuclease III. Later, the enzyme was shown to have an AP endonuclease activity about equal to its exonuclease activity. Endonuclease IV represents about 5% of the total AP endonuclease activity in *E. coli* under normal growth conditions, but the enzyme is strongly inducible in response to oxidative stress. Exonuclease III and endonuclease IV are structurally unrelated and engage the AP site in DNA differently, but they yield the same cleavage product.

The major AP endonuclease of the budding yeast *Saccharomyces cerevisiae* is termed Apn1 and is a relatively close homolog of endonuclease IV, but with a C-terminal extension that contains the nuclear localization sequence. Apn1 functions in yeast BER, and enzyme-deficient mutants are hypersensitive to agents that generate AP sites and base damage processed by BER, such as H₂O₂ and methylating chemicals. A second AP endonuclease is present in *S. cerevisiae* at low levels: the

Table 1 Human DNA glycosylases^a

Enzyme name	Associated AP lyase	Subcellular location ^b	Major substrates ^c
UNG	No	N/M	Uracil (U)
SMUG1	No	N	U
MBD4	No	N	U or thymine (T) opposite the guanine (G) at CpG sequences
TDG	No	N	U, T, etheno-C opposite G
OGG1	Yes	N/M	8-oxoG opposite cytosine (C)
MYH	No	N/M	Adenine (A) opposite 8-oxoG
NTH1	Yes	N/M	Ring-saturated or fragmented pyrimidines
MPG	No	N	3-methyl-A, etheno-A, hypoxanthine
NEIL1	Yes	N	Thymine glycols
NEIL2	Yes	N	Oxidative products of pyrimidines
NEIL3	Yes	N	Oxidative products of pyrimidines

^aAdapted from Wood RD, Mitchell M, and Lindahl T (2005) Human DNA repair genes. *Mutation Research* 577: 275–283.

^bN, nuclei; M, mitochondria.

^cA more comprehensive and regularly updated list can be found online at http://sciencepark.mdanderson.org/labs/wood/DNA_Repair_Genes.html

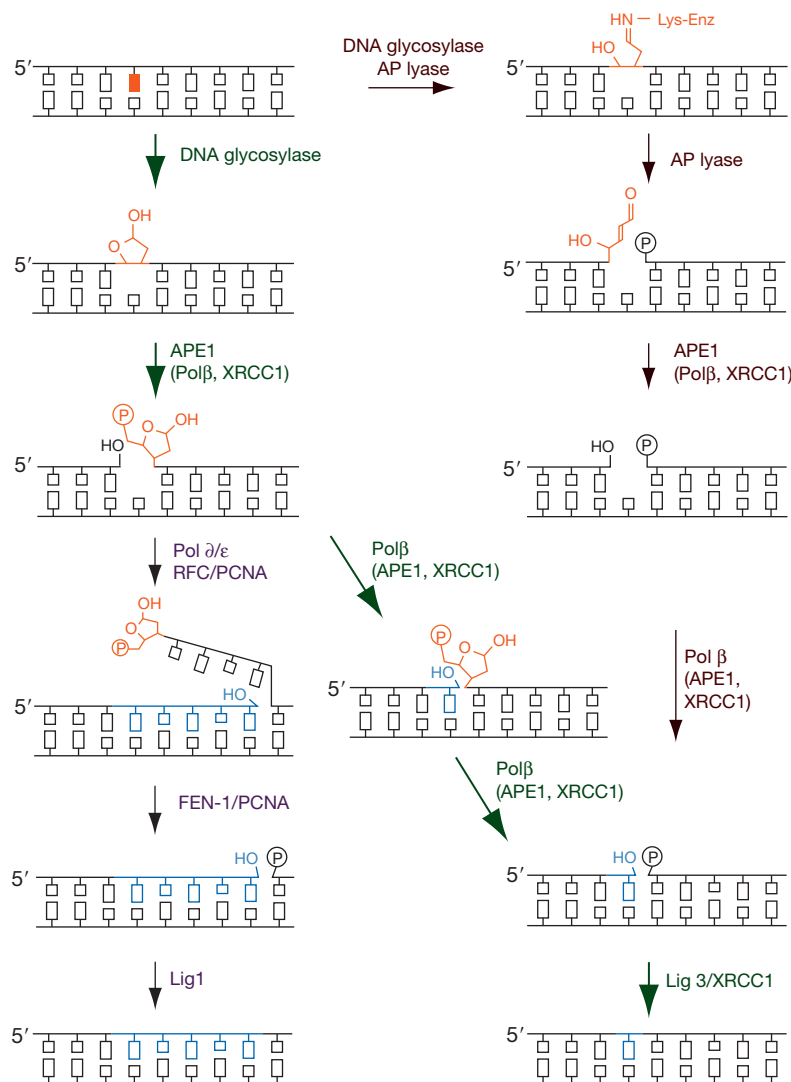


Figure 2 Branching pathways of BER. The topmost branch depends on the type of AP site incision: AP endonuclease (left) vs. AP lyase (right; the lyase Schiff base intermediate is shown). LP-BER flows down the left, and single-nucleotide BER on the right. See text for details. RFC, the loading activity for PCNA protein. Reproduced from Schärer OD (2003) Chemistry and biology of DNA repair. *Angewandte Chemie International Edition in English* 42: 2946–2974.

Apn2/Eth1 protein. This enzyme is much less active than Apn1 and may function as a backup activity, or perhaps it acts in an as-yet-unknown pathway. An Apn1-type AP endonuclease also seems to be the main such enzyme in the nematode *Caenorhabditis elegans*, where it is encoded by the *apn-1* gene. Suppression of this enzyme by RNA interference (RNAi) reproduces in *C. elegans* similar phenotypes of BER deficiency as those observed in yeast.

AP endonucleases of the exonuclease III family seem to be more common in metazoan organisms than endonuclease IV homologs appear to be. The predominant enzyme in mammalian cells belongs to this family, and is usually termed APEX or Ape1 protein. Ape1 proteins contain a ~280-residue segment homologous to exonuclease III, with a ~60-residue extension at the N-terminus that contains the nuclear localization signal. Ape1 is a robust AP endonuclease, even more active than its bacterial counterpart, and it acts on a variety of AP site

derivatives including oxidized products such as 2-deoxyribonolactone (dL; structure 4, Figure 1). The human protein was able to supply the AP endonuclease function in enzyme-deficient *E. coli* or *S. cerevisiae*, but establishing its role in mammalian cells was initially hampered by an inability to obtain stable *APE1*-null cell lines. *APE1*-knockout mutations cause embryonic lethality in mice, underscoring the important role played by this protein. Eventually, RNAi and conditional knockout techniques were used to show that the DNA repair activity of Ape1 is essential in mammalian cells. This contrasts with microbes, where AP endonuclease-deficient mutants show full viability; this difference may relate to the much greater amount of cellular DNA that has to be maintained in mammalian cells. Evidently, enough DNA damage accumulates under normal conditions in mammalian cells to cause proliferation arrest and apoptosis when Ape1-dependent BER is severely limited.

Table 2 Core BER enzymes in DNA base excision repair^a

Organism	Protein			
	ExoIII-like	EndoIV-like	DNA polymerase	DNA ligase
<i>Escherichia coli</i>	Xth	Nfo	Pol I	Lig
<i>Caenorhabditis elegans</i>		apn-1		
<i>Drosophila melanogaster</i>	Rrp1			
<i>Arabidopsis thaliana</i>	ARP			
<i>Danio rerio</i>	Apex1			
<i>Xenopus tropicalis</i>	APEX1			
<i>Mycobacterium genitalium</i>		nfo		
<i>Saccharomyces cerevisiae</i>	Apn2/Eth1	Apn1	Pol2 or Pol3	Cdc9
<i>Schizosaccharomyces pombe</i>	apn2/eth1	apn1	cdc6 or cdc20	cdc17
<i>Mus musculus</i>	Apex1			
<i>Rattus rattus</i>	Apex1			
<i>Cricetus griseus</i>	Apex1			
<i>Bos taurus</i>	APEX1			
<i>Homo sapiens</i>	APEX1 or Ape1		DNA polymerase β or δ/ϵ	Ligase I or III

^aWe have included only enzymes with demonstrated biochemical activity.

End-Processing Enzymes

Oxidative agents including ionizing radiation produce strand breaks with 3' termini that are unusable by DNA polymerases. A notable example is 3'-phosphoglycolate esters, which contain only the former 4' and 5' carbons of deoxyribose (structure 8 in Figure 1). The strand breaks produced by the antitumor drug bleomycin produce the same lesion. In both prokaryotic and eukaryotic cells, these esters are excised by AP endonucleases to provide active 3'-OH primers for DNA repair synthesis. Bacteria deficient in exonuclease III accumulate unrepaired 3'-phosphoglycolate esters when oxidative DNA damage is inflicted by H₂O₂. Although Ape1 is the predominant 3'-phosphoglycolate diesterase activity in mammalian cell extracts, other enzymes have recently been implicated in repair of this lesion, including Aprataxin, Artemis, and the tyrosyl-phosphodiesterase TDP1 that also corrects some abortive topoisomerase intermediates.

Oxidative agents also generate strand breaks with 3'-phosphate and 5'-hydroxyl residues. Such termini block DNA polymerases and can even bind them unproductively. The AP endonucleases of *E. coli* and *S. cerevisiae* readily remove these blocking phosphates, but the mammalian enzymes have far less activity. Instead, the major 3'-phosphatase in mammalian cells is the polynucleotide kinase-phosphatase PNKP, which also phosphorylates the 5' hydroxyls via a separate kinase domain. Note that this process at an oxidative strand break still leaves a one-nucleotide gap requiring DNA repair synthesis.

The glycosylase/lyase enzymes generate products that require 3' processing before DNA repair synthesis can commence. Many of these enzymes leave unsaturated deoxyribose (2,3-didehydro-2,3-dideoxyribose) at the 3' terminus of the nick, which is also the product of nonenzymatic β -elimination at AP sites. Their removal is accomplished by Ape1. Some glycosylase/lyase enzymes (*E. coli* Fpg, and the mammalian NEIL proteins) can also follow β -elimination with a second reaction, δ -elimination, that leaves only a 3'-phosphate that must be excised by PNKP.

DNA Polymerases

Multiple DNA polymerases participate in single-nucleotide and long-patch (LP)-BER (Figure 2). In the mammalian nucleus, the former pathway depends largely on DNA polymerase β , but DNA polymerase λ acts as a backup activity. The insertion of a single nucleotide displaces only the 5' abasic residue at the break. For the LP-BER pathway, DNA polymerase β may also participate, particularly in the early stage, but DNA polymerases δ and ϵ mediate the repair in concert with proliferating cell nuclear antigen (PCNA). LP-BER synthesis results in displacement of the downstream 5' end to form a flap structure (Figure 2).

5'-Deoxyribose-5-Phosphate Lyase

The abasic residue displaced by DNA repair synthesis must be removed before ligation can complete the BER reaction. For the single-nucleotide pathway, this is accomplished in bacteria by Fpg protein, in a reaction analogous to its incision of an intact AP site. In mammalian cells, the excision is mostly dependent on DNA polymerase β , which uses a lysine in its N-terminal domain to carry out the lyase reaction (5'-dRp lyase; Figure 3). This seems to be the more important function of polymerase β in single-nucleotide BER, in that resistance to monofunctional alkylating agents in *POLB*-knockout cells is provided by expression of the separated N-terminal lyase domain, but not when only the polymerase domain is expressed. Polymerases ι and λ also have 5'-dRp lyase activity. It should be noted that this activity is problematic for processing certain oxidized abasic sites (see section Special Problems in BER).

Flap Endonuclease

The flap structure generated by DNA synthesis in LP-BER renders the 5'-dRp residue unavailable for removal by the 5'-dRp lyase of polymerase β . Instead, this structure must be removed by Flap endonuclease (FEN1 protein), which incises the displaced single strand at its junction with double-stranded DNA (Figure 2).

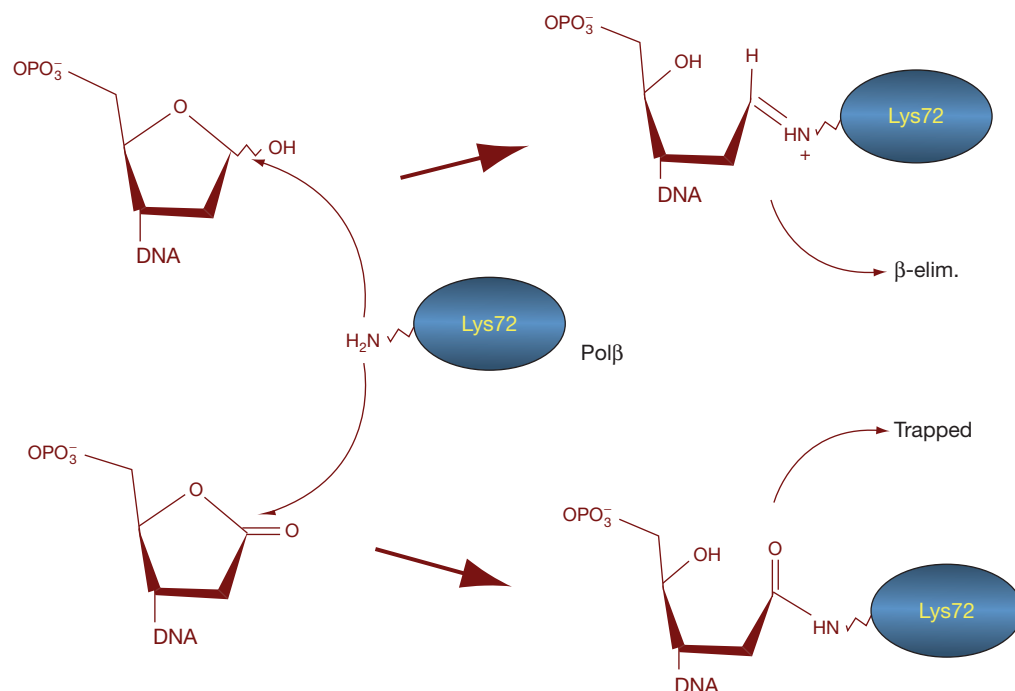


Figure 3 The 5'-dRp lyase activity of DNA polymerase β and cross-link formation with dL. Lysine-72 in the N-terminal domain of the enzyme attacks C1 of an AP site to produce a Schiff base intermediate, which is then hydrolyzed to release the enzyme, the abasic residue and the DNA in overall β-elimination reaction (upper pathway). Attack of the C1 of dL leads to an amide bond, which traps the protein on the DNA (lower pathway). Reproduced from Demple B and DeMott MS (2002) Dynamics and diversions in base excision DNA repair of oxidized abasic lesions. *Oncogene Reviews* 21: 8926–8934.

Like polymerases δ and ε, this enzyme is also a main player in DNA replication, where it participates in the removal of Okazaki fragments formed during lagging-strand synthesis.

DNA Ligases

Two vertebrate DNA ligases are involved in BER, ligase I and ligase III. Each interacts with a partner protein during its reaction in BER. DNA ligase I interacts with PCNA and DNA ligase IIIα forms a stable complex with XRCC1 protein. Although XRCC1 does not itself have enzymatic activity, without this protein ligase III is destabilized; XRCC1 also mediates the interaction of the ligase with DNA polymerase β during single-nucleotide BER. Ligase I participates in LP-BER; it is still another protein (such as FEN1 and PCNA) recruited from replication.

Coordination among BER Proteins

Numerous pairwise protein interactions can coordinate the reactions of BER. Many of the proteins in this pathway exhibit relatively strong product binding, such that they need to be displaced in order for the process to proceed. Several mammalian DNA glycosylases have been reported to interact with Ape1 protein, which accelerates the glycosylase reactions. In many cases, this effect may occur by the Ape1-mediated incision of their product AP sites. Similarly, Ape1 binds its 5'-incised abasic product more tightly than it does an intact AP site. This bound Ape1 recruits DNA polymerase β, and in the process

strongly activates the 5'-dRp lyase activity. DNA synthesis and dRp excision facilitate the departure of Ape1, and bound DNA polymerase β recruits the XRCC1-ligase IIIα complex to promote ligation. In LP-BER, PCNA interacts with both the DNA polymerases and with FEN1 protein to coordinate the overall repair reaction. Collectively, these interactions may prevent the accidental release of repair intermediates that could undergo aberrant reactions or trigger checkpoint pathways.

Special Problems in BER: A Rationale for the LP Pathway

When LP-BER was first discovered, the utility of this pathway was not obvious. When examined in cell-free extracts, lesions such as uracil, which in principle should be well handled by single-nucleotide BER, nevertheless undergo significant amounts of LP-BER (although the single-nucleotide pathway predominates). Some evidence suggested that modified forms of deoxyribose, perhaps oxidized derivatives, might require LP-BER. One such form was identified as dL (Figure 1), a major abasic lesion formed by radiation or chemical oxidation. Synthetic DNA substrates containing a single dL residue were readily cleaved by Ape1 protein, about as efficiently as an ordinary AP site. However, subsequent reaction with DNA polymerase β revealed that the enzyme forms a stable DNA-protein cross-link during attempted excision of the abasic residue. Instead of the hydrolysis-sensitive Schiff base formed during the normal 5'-dRp lyase reaction, the polymerase forms

a stable amide bond with its catalytic lysine-72 (Figure 3). This DNA–protein cross-link is resistant to removal by various enzymes, including FEN1 protein. However, substrates with a dL residue trigger the activation of LP-BER, such that this lesion is processed virtually completely by this pathway. As dL is produced in amounts similar to the other major oxidative lesions, LP-BER would be biologically essential for correction of this lesion. Other naturally occurring lesions that require LP-BER remain to be identified.

Other polymerases with 5'-dRp lyase activity also form cross-links with the incised dL lesion (see section [BER Pathways in Mitochondria](#)). In addition, attack on an unincised dL residue by some AP lyase enzymes might also produce stable DNA–protein cross-links. The mammalian enzymes do not seem to be prone to this reaction, however, and only endonuclease III of *E. coli* (the NTH1 counterpart) was reported to form stable cross-links with dL. None of the mammalian AP lyases was able to incise a dL residue.

BER Pathways in Mitochondria

Mitochondrial DNA (mtDNA) is as essential for eukaryotic function as is the nuclear genome, so it might seem obvious that systems to maintain mtDNA would be present. Moreover, mtDNA replication proceeds even in nonproliferating cells, which implies that maintenance of the mtDNA template is required in all tissues at all times. A particular threat to mtDNA may be oxidative damage, as the organelle is the site of oxidative phosphorylation. Although mitochondria seem to lack pathways available in the nucleus, notably nucleotide excision repair activity or proteins, several DNA glycosylases (Table 1) and Ape1 protein are present. In addition, the mtDNA polymerase γ has its own 5'-dRp lyase activity, and a form of DNA ligase III is present in the mammalian mitochondria. Thus, the basic single-nucleotide BER machinery is present to handle some deamination and oxidative lesions.

However, oxidation produces dL residues in amounts similar to 8-oxoG, such that LP-BER may also be needed. Indeed, DNA polymerase γ also forms stable cross-links with Ape1-incised dL in a manner analogous to polymerase β (Figure 3). As in the nucleus, this problem is averted by the presence of LP-BER activities, which led to the discovery that FEN1 protein is present in mitochondria. Mammalian mitochondria also contain a second activity that can process flap-containing intermediates formed during LP-BER, an isoform of the DNA2 protein. This distribution contrasts with *S. cerevisiae*, in which DNA2 seems to be exclusively nuclear.

The localization of the various BER proteins to mitochondria occurs by both classical and nonclassical pathways. In the case of uracil glycosylase, a specific isoform of the protein generated via alternative messenger RNA splicing is transported in the signal-peptide pathway. In the case of Ape1 protein, removal of its N-terminal nuclear-targeting sequence may allow the C-terminal mitochondrial transport sequence to function. The mitochondrial targeting sequence of FEN1 has not been identified, but a potential mitochondrial import determinant was identified in mammalian DNA2. It is unknown whether the distribution of these enzymes is altered under different conditions.

Other Functions of BER Proteins

The Ape1 AP endonuclease was independently identified as a factor termed Ref-1, which activates *in vitro* DNA binding by various transcription factors. The detailed mechanism of this reaction has not been determined, but it seems dependent on one or more cysteines in the N-terminal ~100 residues of the protein. As most of these residues are folded inside the native Ape1 structure, some partial unfolding may be required for Ref-1 activity. Some *in vivo* experiments in cell culture suggest that Ref-1 depends on reducing power from thioredoxin.

Ape1 may also directly regulate several promoters including that of the parathyroid hormone gene by binding upstream control elements. This binding appears to depend on two N-terminal lysine residues, which are partially acetylated in cells. Additional function of these acetylation sites has been proposed for an Ape1-specific function in the nucleolus, the site of ribosomal RNA synthesis.

BER Genes and Disease Risk

In view of the apparently fundamental role of BER pathways, it is surprising that deletion of individual DNA glycosylase genes in mice does not produce dramatic increases in rates of spontaneous cancer. For the uracil glycosylase gene, a modest elevation in certain cancers is observed in older animals, and for other glycosylase genes, their ablation increases the rate of tumor formation in response to carcinogens. One possibility is that there is redundancy for the most important lesions among the glycosylase enzymes. However, for OGG1 and MUTYH, human polymorphisms have been linked to increased rates of some human cancers. Additional studies point to possible contributions to cancer by polymorphisms affecting other BER proteins, such as XRCC1. It is unclear to what extent these effects may be due to alterations of nuclear as opposed to mitochondrial function.

In contrast to the glycosylases, all the genes of the central BER pathway are essential, complicating efforts to develop mouse models of defects in their activity. It may also be that carcinogenesis would not be the most prominent effect of such defects; for example, the demonstrable roles of some BER proteins in mtDNA maintenance may give effects in longevity and the aging process.

Some of the DNA end-processing enzymes have connections to human disease. Thus, PNKP mutations cause various neurological defects, which underscores the general importance of pathways to correct single-strand breaks in DNA. Repair of mtDNA would also be especially important in this context.

Conclusion

The incessant production of DNA lesions by spontaneous and endogenous reactions can be lethal in the absence of active BER functions. In addition, unrepaired lesions trigger mutations due to misreading during replication or the activity of translesion polymerases. BER pathways handle a broad variety of hydrolytic, oxidative, and methylation lesions through branch-point pathways that either replace just one nucleotide or generate

a multinucleotide repair patch (LP-BER). Many lesions can be handled by the single-nucleotide branch, but the LP-BER branch is necessary to process at least one type of oxidized abasic site. DNA glycosylases remove damaged bases or, in some specific cases, conventional DNA bases that are inappropriately positioned. Glycosylases produce AP sites, thus funneling a multitude of different lesions into a central pathway. The central pathway BER enzymes occur universally in nature, and in mammalian cells, they are essential even at the cellular level.

See also: **Bioenergetics:** Mitochondria: Source and Target of Free Radicals; Mitochondrial DNA; **Molecular Biology:** DNA Damage: Alkylation; DNA Glycosylases: Mechanisms; DNA Ligases: Mechanism and Functions; DNA Mismatch Repair and the DNA Damage Response; DNA Mismatch Repair in Mammals; DNA Oxidation; DNA Polymerase β , Eukaryotic; DNA Polymerases: Kinetics and Mechanism; Eukaryotic DNA Polymerase δ ; Nonhomologous End Joining in Eukaryotes; Nucleotide Excision Repair, Bacterial: The UvrABCD System;

Nucleotide Excision Repair: Biology; Nucleotide Excision Repair in Eukaryotes; UmuC D Lesion Bypass DNA Polymerase V.

Further Reading

- Barnes DE and Lindahl T (2004) Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annual Review of Genetics* 38: 445–476.
- Caldecott KW (2008) Single-strand break repair and genetic disease. *Nature Reviews Genetics* 9: 619–631.
- de Souza-Pinto NC, Wilson DM 3rd, Stevnsner TV, and Bohr VA (2008) Mitochondrial DNA, base excision repair and neurodegeneration. *DNA Repair* 7: 1098–1109.
- Friedberg EC, Walker GC, Siede W, et al. (2006) *DNA Repair and Mutagenesis*, 2nd edn., p. 1164. Washington, DC: ASM Press.
- Larsen E, Meza TJ, Kleppa L, and Klungland A (2007) Organ and cell specificity of base excision repair mutants in mice. *Mutation Research* 614: 56–68.
- Liu P and Demple B (2010) DNA repair in mammalian mitochondria: Much more than we thought? *Environmental and Molecular Mutagenesis* 51: 417–426.
- Tell G, Quadrioglio F, Tiribelli C, and Kelley MR (2009) The many functions of APE1/Ref-1: Not only a DNA repair enzyme. *Antioxidants and Redox Signaling* 11: 601–620.

DNA Damage: Alkylation

J L Tubbs and J A Tainer, The Scripps Research Institute, La Jolla, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Anton B. Guliaev and B. Singer, volume 1, pp. 609–613, © 2004 Elsevier Inc.

Glossary

Adduct Chemical group covalently attached to a DNA nucleotide; same as a lesion.

Alkylating agent Chemical compound that transfers an alkyl group to another compound or molecule.

DNA base The purine or pyrimidine group in a nucleotide.

DNA base-flipping A common strategy used by enzymes to gain access to nucleotide substrates in double-stranded DNA, whereby a base is rotated out of the DNA helix and into the enzyme active site pocket.

Lesion Chemical group covalently attached to a DNA nucleotide; same as an adduct.

Nucleotide Compound that consists of a nitrogenous base, a sugar, and one or more phosphate groups to form the basic DNA building blocks of guanine, adenine, thymine, or cytosine.

Purine A heteroaromatic compound, present in adenine and guanine, composed of carbon, hydrogen, nitrogen, and oxygen atoms to form fused five- and six-membered rings.

Pyrimidine A heteroaromatic compound, present in thymine and cytosine, composed of carbon, hydrogen, nitrogen, and oxygen atoms to form six-membered rings.

Introduction

DNA-alkylation lesions are among the most common types of DNA damage within cells. If unrepaired, some have serious mutagenic, and even cytotoxic, consequences. Fortunately, cells are protected by several DNA repair mechanisms. However, this repair can be adverse to certain cancer-alkylating therapies, underscoring the importance of understanding DNA alkylation damage and its repair. Here we discuss common types, causes, biological effects, and repair mechanisms for cellular DNA alkylation damage.

Types, Causes, and Deleterious Effects of DNA-Alkylation Damage

DNA base exocyclic oxygens and most of the ring nitrogens are susceptible to alkylation to varying degrees. DNA phosphate backbone oxygens are also susceptible to methylation, leading to methylphosphotriesters. In cells, methylation is most common, but larger adducts can also form, depending on the alkylating agent. Alkylating agents work by either unimolecular nucleophilic substitution (S_N1) or bimolecular nucleophilic substitution (S_N2), with the reaction mode determining the alkylation proportion at different sites. The predominant lesions formed by most alkylating agents are the innocuous 7-methylguanine (7mG) and the highly cytotoxic 3-methyladenine (3mA) lesions. In addition, S_N2 agents, such as methyl methanesulfonate (MMS) (Figure 1) and methyl halides, produce 1-methyladenine (1mA) and 3-methylcytosine (3mC), largely in single-stranded (ss) DNA and RNA. S_N1 agents, such as *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG) and *N*-methyl-*N*-nitrosourea (MNU) (Figure 1), react more readily

with DNA oxygens to produce larger quantities of the highly mutagenic and cytotoxic O^6 -methylguanine (O^6 -mG) and the less prevalent but more mutagenic O^4 -methylthymine (O^4 -mT) lesions. The biological effects resulting from DNA alkylation include the alteration of DNA structure and function, DNA strand breaks, and apoptosis. The most common types of DNA-alkylation damage, along with their biological effects and known repair mechanisms, are listed in Table 1.

DNA-alkylation damage arises from several sources. One source is the endogenous methyl donor *S*-adenosylmethionine (SAM), an S_N2 agent that generates the same DNA-alkylation products as MMS, but with a 2000-fold weaker reactivity. SAM is the methyl donor for the majority of *in vivo* enzymatic methylation reactions, and it also spontaneously methylates cellular nucleic acids and proteins at a low but significant rate due to its high transfer potential. Another source is environmental toxins. The most abundant are S_N2 methyl halides from decaying vegetation and biomass burning. Other toxins are potentially carcinogenic S_N1 nitrosamines generated by reacting nitrite with amines and amides in food or decaying matter and found in drinking water, diet, nicotine metabolites, or occupational hazards from industrial processes. A third source is chemotherapeutic alkylating drugs that cause cancer cell apoptosis. The first class of these drugs are methylating agents that produce various DNA monoadducts, including lethal O^6 -mG, which causes apoptosis via the DNA mismatch repair (MMR) pathway (discussed later). The second class are chloroethylating agents that form several DNA and protein adducts. The most toxic one results from guanine O^6 chloroethylation, followed by spontaneous cyclization to yield an unstable N^1, O^6 -ethylguanine intermediate, which readily reacts with the opposing cytosine N3 to form a DNA G–C crosslink that inhibits DNA replication and transcription and causes DNA double-strand breaks and apoptosis.

Repair of DNA-Alkylation Damage by Base Excision Repair

The DNA base excision repair (BER) pathway is the major repair mechanism for most alkylation damage. It repairs alkylated bases with destabilized N-glycosylic bonds, such as N³- and N⁷-substituted alkylpurines and O²-substituted alkylpyrimidines. BER is initiated through the recognition and excision of modified bases by hydrolysis of the N-glycosylic bond by lesion-specific DNA glycosylases to create DNA abasic (AP) sites. Next, the AP site is incised at the 5'-side by AP endonuclease or 3'-side by AP lyase, followed by 5'- or 3'-phosphodiesterases to remove the remaining sugar phosphate at the DNA termini. Finally, DNA polymerase fills the gap of one to several nucleotides and DNA ligase seals the nick.

DNA glycosylases that recognize alkylated bases are found in eukaryotic and prokaryotic organisms. The five known classes of these glycosylases are represented by: (1) 3mA DNA glycosylase I (*Escherichia coli* and *Salmonella*

typhi TAG), (2) 3mA DNA glycosylase II (*E. coli* AlkA), and (3) *Helicobacter pylori* 3mA DNA glycosylase III (MagIII), all belonging to the helix-hairpin-helix (HhH) superfamily of DNA glycosylases, (4) human AAG, and (5) prokaryotic HLR 3mA DNA glycosylases (*Bacillus cereus* AlkC and AlkD) (Figure 2). Also, methylated purines are excised by *Thermotoga maritima* methylpurine DNA glycosylase II (MpgII) and some enzymes of the endonuclease III (Nth) DNA glycosylase family. All known alkylbase glycosylases remove 3mA, but with varying efficiency depending on the surrounding DNA sequence. TAG, AlkA, MagIII, and AAG use similar substrate recognition mechanisms of DNA distortion, nucleotide flipping (usually ~180°), and DNA intercalation by amino acid side chains for stabilization, whereas a different mechanism is likely employed by AlkC and AlkD.

3mA DNA Glycosylase I (*E. coli* and *S. typhi* TAG)

The constitutively expressed TAG protein of the HhH family of DNA glycosylases removes 3mA, and also 3mG with lower affinity, but not 7mG. TAG is important for initial cellular protection against sudden alkylation challenge. TAG is a globular protein, with an α -helical domain containing the HhH motif, and a Zn²⁺-binding domain that connects the protein N- and C-termini (Figure 2(a)). The lesion-binding pocket is at the domain interface and contains specific substrate interactions. Despite structural similarity to other HhH-family glycosylases, TAG has notable features that distinguish it as an

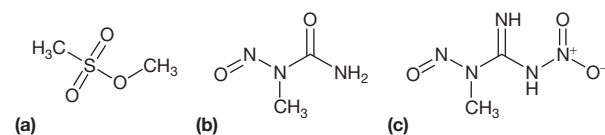


Figure 1 Chemical structures of the alkylating agents MMS (a), MNU (b), and MNNG (c).

Table 1 DNA alkylation sites and their repair

Lesion (abbreviation)	Alkylation pattern ^{a,b}	Biological effect ^{c,d}	DNA repair mechanism	Recognition proteins
Adenine				
N ¹ -Methyladenine (1mA)	S _N 2: 18%	Cytotoxic	Direct damage reversal	AlkB, AAG, ABH2, ABH3
N ³ -Methyladenine (3mA)	Most agents: 10%	Cytotoxic	BER	AlkA, TAG, MagIII, AAG, AlkC, AlkD
N ⁷ -Methyladenine (7mA)	Most agents: 2%	Rapidly depurinates	BER	AlkA
1,N ⁶ -Ethenoadenine (εA)		Cytotoxic	BER, Direct damage reversal	AlkA, AAG, AlkB, ABH3
Guanine				
N ¹ -Methylguanine (1mG)		Mutagenic, cytotoxic	BER, Direct damage reversal	AlkB, AAG
N ³ -Methylguanine (3mG)	Most agents: 1%	Cytotoxic	BER	AlkA, TAG, AlkC, AlkD
O ⁶ -Methylguanine (O ⁶ -mG)	S _N 1: 6% S _N 2: 0.3%	Mutagenic, cytotoxic	Direct damage reversal, NER, MMR	C-Ada, AGT, Ogt, ATL
N ⁷ -Methylguanine (7mG)	Most agents: 70%	Innocuous but depurinates	BER	AlkA, AAG, AlkD
8-Methylguanine	Methyl radicals		BER	AlkA
Thymine				
O ² -Methylthymine (2mT)	S _N 1: 0.1%	Cytotoxic	BER	AlkA
N ³ -Methylthymine (3mT)	S _N 2	Cytotoxic	Direct damage reversal	FTO, AlkB
O ⁴ -Methylthymine (O ⁴ -mT)	S _N 1: 0.4%	Mutagenic, cytotoxic	Direct damage reversal	Ogt, C-Ada, AGT
Cytosine				
O ² -Methylcytosine (2mC)	S _N 1: 0.1%	Cytotoxic	BER	AlkA
N ³ -Methylcytosine (3mC)	S _N 2: 10%	Cytotoxic	BER, Direct damage reversal	AlkB, AAG, ABH1, ABH2, ABH3
3,N ⁴ -Ethenocytosine (εC)		Mutagenic, cytotoxic	Direct damage reversal	AlkA, AAG, AlkB
DNA backbone				
Sp-Methylphosphotriester	S _N 1: 17%		Direct damage reversal	N-Ada

^aReviewed in Mishina Y, Duguid EM, and He C (2006) Direct reversal of DNA alkylation damage. *Chemical Reviews* 106: 215–232.

^bReviewed in Sedgwick B (2004) Repairing DNA-methylation damage. *Nature Reviews. Molecular Cell Biology* 5: 148–57.

^cReviewed in Shrivastav N, Li D, and Essigmann JM (2010) Chemical biology of mutagenesis and DNA repair: Cellular responses to DNA alkylation. *Carcinogenesis* 31: 59–70.

^dReviewed in Wyatt MD and Pittman DL (2006) Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chemical Research in Toxicology* 19: 1580–1594.

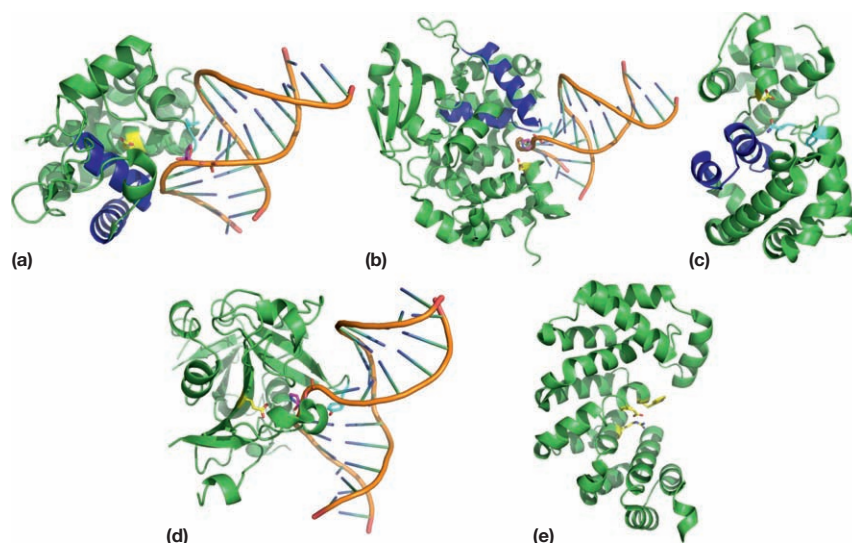


Figure 2 Crystal structures for representatives of each class of BER glycosylases that recognize alkylated bases. (a–c) Members of the HhH superfamily. HhH motif, blue. (a) *S. typhi* TAG bound to DNA. (b) *E. coli* AlkA bound to DNA. (c) *H. pylori* MagIII. (d) Human AAG bound to DNA. (e) *B. cereus* AlkD. Damage base (magenta), catalytic (yellow), and base flipping (cyan) residues are shown.

outlier, including a different strategy to access the damage site: intercalation of both DNA strands by a single hairpin loop and impartial rotation of the target nucleotide into its active site, and catalysis by a glutamic acid rather than the conserved aspartate of other known HhH glycosylases. TAG likely detects 3mA lesions through its extensive interactions with both DNA strands and uses a ground-state destabilization mechanism for the catalysis of base excision.

3mA DNA Glycosylase II (*E. coli* AlkA)

AlkA is an inducible HhH-family protein that recognizes N^3 - and N^7 -alkyl purines, O^2 -alkyl pyrimidines, deaminated adenine (inosine), and the cyclic ethenoadducts 1, N^6 -ethenoadenine (ϵ A) and 3, N^4 -ethenocytosine (ϵ C). AlkA family members include *E. coli* AlkA, *Saccharomyces cerevisiae* Mag, *Schizosaccharomyces pombe* Mag1 and Mag2, and *B. cereus* AlkE. *E. coli* AlkA is a globular three-domain protein (Figure 2(b)). The active site with conserved catalytic Asp238 is inside a large hydrophobic cleft between two domains. This active site cleft is unusually rich in aromatic residues and AlkA likely uses this abundance of electrons, in combination with shape and H-bonding, to recognize electron-deficient substrates. Leu125 wedges out alkylated nucleotides from DNA, accompanied by an induced 66° DNA bend and minor groove widening. AlkA primarily contacts the DNA damaged strand, anchoring it through a sodium ion bridge with the HhH motif. The extrahelical alkyl base is stabilized by a cation- π interaction with Trp272. Asp238 likely stabilizes a carbocation intermediate for catalysis to proceed through an S_N1 -type mechanism. Excision selectivity may result from the weakened glycosylic bond of the positively charged base.

H. pylori 3mA DNA Glycosylase III (MagIII)

H. pylori MagIII is highly specific for 3mA removal and shares 32% sequence identity with *T. maritima* MpgII, but its inability

to remove 7mG or hypoxanthine from DNA makes it more functionally similar to TAG. MagIII is similar in structure to HhH-family glycosylases (Figure 2(c)), with the notable differences that it lacks an iron sulfur cluster or other additional domains, and its C-terminal domain contains a longer helix that is part of the substrate-binding site and a lysine carbamylation site. The lesion-binding pocket is lined by hydrophobic residues and is specific for 3mA, but sterically excludes 7mG. Favorable π - π stacking interactions between the 3mA base and two aromatic side chains, rather than direct hydrogen bond interactions with the nucleobase, account for substrate specificity.

Human AAG

AAG, also known as MPG, is the human functional analog of AlkA and only known human BER DNA glycosylase that removes alkylation damage. It has a wide specificity range and removes other damage, similar to AlkA. AAG belongs to a distinct family of mammalian glycosylases characterized by a mixed α/β fold and no HhH motif. AAG is composed of a single α/β domain, with α -helices enclosing an antiparallel β -sheet (Figure 2(d)). AAG flips the alkylated nucleotide into a deep pocket, accompanied by side chain intercalation of β -hairpin residue Tyr162 into the DNA minor groove. Planar stacking and cation- π interactions by Tyr127, His136, and Tyr159 with the alkylbase account for substrate specificity. AAG excision specificity, like AlkA, likely depends on the chemical instability of positively charged, alkylated nucleobase glycosylic bonds. AAG bends substrate DNA by 20°, and widens the DNA minor groove by more than 2 Å. Although the active site can accommodate a variety of substrates, AAG uses a combination of nucleophilic activation/positioning and leaving group protonation for efficient alkylpurine catalysis. AAG simultaneously searches both DNA strands and bypasses DNA-bound protein roadblocks through hopping.

Prokaryotic HLR 3mA DNA Glycosylases (*B. cereus* AlkC and AlkD)

B. cereus AlkC and AlkD belong to a prokaryotic HLR 3mA DNA glycosylase superfamily. These enzymes contain no sequence homology to other known DNA glycosylases and likely catalyze nucleobase excision by a different mechanism. Both proteins efficiently excise 3mA. AlkC removes 3mG more efficiently than AlkD, whereas AlkD, but not AlkC, is proficient in excising 7mG. AlkD is a single α -helical domain protein, with 12 of its 14 helices arranged in antiparallel pairs to form six tandemly repeated Huntington/elongation/A subunit/target-of-rapamycin (HEAT)-like motifs that stack into a super-helical solenoid with a highly electropositive concave surface (Figure 2(e)). The concave surface contains a central shallow putative active site cleft of aromatic and charged residues. Stacking interactions between Trp109 and alkylated bases may be important for substrate recognition. An Asp113–Arg148 salt bridge is likely essential for alkylbase excision and may contribute to AlkD's specificity for positively charged alkylpurines. DNA was proposed to bind the concave surface in a novel mode of protein HEAT-repeat–DNA interaction, but no obvious DNA-intercalating side chains were identified. Surprisingly, AlkD DNA-binding models suggest that it may access the alkylbase for catalysis without flipping it into the aromatic cleft, raising the possibility of a different purpose for this cleft.

Repair of DNA-Alkylation Damage by Direct Damage Reversal Proteins

Repair of more stable DNA alkyl lesions, such as N1-substituted purines and O^6 -methylguanine, is accomplished by direct damage reversal proteins. These proteins reverse damage directly, without the need to break the DNA backbone, remove the damaged base, or further modify the nucleotide or DNA. DNA-alkylation direct damage reversal proteins fall into two categories: alkyltransferases and Fe(II)/2-oxoglutarate(2OG)-dependent dioxygenases.

Alkyltransferases

Alkyltransferases are widespread among prokaryotes and eukaryotes and repair O^6 -mG and O^4 -mT lesions. The N- and C-terminal domains of many prokaryotic alkyltransferases have distinct repair activities, as described later. For most eukaryotic alkyltransferases, the repair activity is completely

carried out by the C-terminal domain, whereas the N-terminal domain function is unknown. Alkyltransferase fusion proteins, such as *Caenorhabditis elegans* alkyltransferase-histone 1C and *Ferroplasma acidarmanus* alkyltransferase–endonuclease V, have also been identified.

N-Ada

E. coli contains two alkyltransferase proteins, the inducible Ada and the constitutively expressed Ogt. Ada is a multifunctional, two-domain protein, and each domain has a separate repair activity. The N-terminal domain of Ada (N-Ada) repairs Sp-methylphosphotriesters, but not Rp-methylphosphotriesters, via a zinc-activated Cys residue, resulting in covalent methylation of this domain. Methylated N-Ada has a significantly increased affinity for DNA and is thereby converted into an efficient transcriptional activator of several *E. coli* alkylation resistance genes within the *ada* regulon, including *alkA*, *alkB*, *aidB*, and *ada* itself. Thus, Ada is key to the *E. coli* adaptive response to alkylation damage.

N-Ada contains two subdomains, N-Ada_N and N-Ada_C, arranged along one face of the DNA helix and connected by a flexible linker (Figure 3(a)). N-Ada_C is an α -helical domain that binds the DNA major groove via a canonical helix–turn–helix (HTH) DNA-binding motif. N-Ada_N, a globular domain containing two α -helices enclosing a central β -sheet, contains zinc(II)-activated nucleophilic Cys38 and binds the DNA minor groove. To selectively recognize methylphosphotriester lesions and control its DNA-binding affinity and transcriptional regulation, N-Ada uses a methylation-dependent electrostatic switching mechanism whereby a negatively charged patch near the zinc–sulfur cluster at the protein–DNA interface is neutralized by a methylphosphotriester lesion or methylation of Cys38, thereby reducing charge–charge repulsion between the protein and the negatively charged DNA backbone and resulting in selective recognition of the neutral methylphosphotriester or enhanced DNA-binding affinity, respectively.

C-Ada, Ogt, and AGT

Two *E. coli* proteins, Ogt and the C-terminal domain of Ada (C-Ada), repair O^6 -mG and O^4 -mT. In higher eukaryotes and mammals, including humans, this repair is performed by O^6 -alkylguanine-DNA alkyltransferase (AGT), also known as hAGT or O^6 -methylguanine-DNA methyltransferase (MGMT) in humans. These proteins directly and covalently transfer the alkyl lesion from the damaged guanine to a conserved nucleophilic cysteine within a $-(I/V)PCHR(V/I)-$ active site motif in a stoichiometric, irreversible reaction. Alkylation damage can be

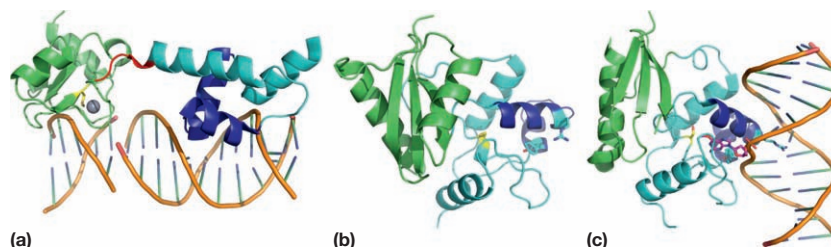


Figure 3 Crystal structures for direct damage reversal alkyltransferases. (a) *E. coli* N-Ada bound to DNA. Zn^{2+} is shown as a gray sphere. (b) *E. coli* C-Ada. (c) Human AGT bound to DNA. N-terminal domain (green), C-terminal domain (cyan), HTH motif (blue), base flipping residues (cyan sticks), and reactive Cys residue (yellow sticks) are shown.

repaired in ssDNA, but is preferentially repaired in dsDNA. C-Ada and other AGTs, including Ogt, can repair O^6 -alkylG lesions larger than methyl, but repair efficiency is affected by the alkyl group size. hAGT has a larger active site pocket than C-Ada and can more rapidly repair a broader range of bulky O^6 -alkylG lesions. The repair of O^4 -mT is most efficient in Ogt, followed by C-Ada, and finally hAGT, *in vitro*. However, *in vivo* reports of this repair are contradictory, and further studies are needed to clarify these inconsistencies. Interestingly, hAGT repair specificity and efficiency are altered by the modification of residues within and near the active site. In addition to preventing mutations, hAGT promotes tumor resistance to alkylating chemotherapies, making it a prime drug target.

The first of two C-Ada subdomains consists of two α -helices surrounding a β -sheet, whereas the second is mostly composed of α -helices and coils and contains conserved HTH and active site motifs (Figure 3(b)). hAGT is also a two-domain protein that shares C-Ada's α/β fold (Figure 3(c)). Like C-Ada, the hAGT C-terminal domain contains all known DNA-binding and alkyltransfer residues. For both C-Ada and hAGT, the reactive cysteine is in an interior cavity and activated for nucleophilic attack by a hydrogen bond network involving a histidine, glutamic acid, and structured water. hAGT binds the DNA minor groove via the HTH motif, flipping the alkylated guanine into its active site for repair. This nucleotide flipping is achieved by conserved arginine and tyrosine residues, with the arginine also stabilizing the extrahelical DNA conformation. AGT is a suicide protein because the alkyl adduct remains covalently attached and signals proteolytic degradation. AGTs likely detect DNA lesions by searching for weakened and/or distorted base pairs rather than the adduct. hAGT binds ss- and double-stranded (ds) DNA in a highly cooperative manner, which leads to a 3'-5' kinetic scanning bias aiding the search for damaged DNA through localization of multiple AGT molecules.

Fe(II)/2OG-Dependent Dioxygenases

Fe(II)/2OG-dependent dioxygenases are a large superfamily of proteins that catalyze a variety of biosynthetic and degradation reactions. DNA-alkylation damage is repaired by only certain family members, including *E. coli* AlkB, its human homologs ABH1, ABH2, ABH3, and the fat mass and obesity associated (FTO) dioxygenase. AlkB and its homologs repair 1mA and 3mC by coupling their oxidative demethylation to conversion of 2OG into succinate and CO_2 . The process requires iron(II), 2OG, and O_2 and the oxidized methyl group is spontaneously released as formaldehyde. Both AlkB and ABH3 prefer lesions in ssDNA and -RNA, whereas ABH2 preferentially repairs dsDNA lesions. AlkB also repairs 1mC, 3mT, 1, N^6 -ethenoadenine, 3, N^4 -ethenocytosine and analogs. ABH2 and ABH3 have a more narrow specificity, and ABH3 shows lower activity for 1, N^6 -ethenoadenine. FTO prefers 3mT in ssDNA or 3-methyluracil (3mU) in ssRNA. ABH1 has a very weak repair activity for 3mC in ssDNA and -RNA, but has stronger lyase activity for cleaving DNA at abasic sites.

AlkB, ABH2, and ABH3

AlkB, ABH2, and ABH3 have a core consisting of double-layered β -sheets in a jellyroll fold, matching other iron(II)/2OG-dependent dioxygenase superfamily members (Figures 4(a)–4(c)). The 2OG and iron cofactors are located near the conserved core H-X-D/E-Xn-H motif. AlkB has three extra β -strands that form a likely flexible nucleotide-recognition lid. Significant differences between AlkB and ABH3 include a more polar hABH3 substrate-binding cavity with key AlkB tryptophan replaced by tyrosine, structurally different DNA-binding interfaces, and different substrate-binding lid secondary structure suggesting distinct binding modes among AlkB homologs. Interestingly, ABH3 and ABH2 share some DNA-binding motifs, although substrate preferences differ. Notable AlkB and ABH2

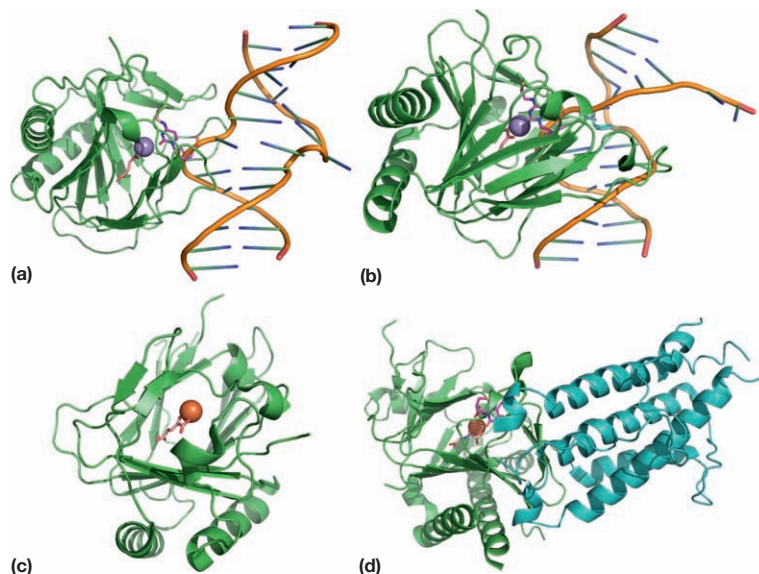


Figure 4 Crystal structures for direct damage reversal dioxygenases. (a) *E. coli* AlkB bound to DNA and Mn^{2+} . (b) Human ABH2 bound to DNA and Mn^{2+} . DNA-intercalating Phe shown in cyan. (c) Human ABH3 bound to Fe^{2+} . (d) Human FTO bound to Fe^{2+} and 3mT (magenta sticks) showing N-terminal dioxygenase domain (green) and C-terminal domain (cyan). Mn^{2+} (gray sphere) and Fe^{2+} (orange sphere), 2OG (pink sticks), and flipped damaged base (magenta sticks) are shown.

DNA interaction differences include AlkB predominant contacts with the DNA damaged strand versus ABH2 significant interactions with both strands. Also, to achieve nucleotide rotation, AlkB pinches together the two bases on either side of the damaged base so that they stack with one another, whereas ABH2 intercalates an aromatic phenylalanine finger residue into the DNA.

FTO

FTO contains an AlkB-like domain and a second domain with a novel fold, both of which are necessary for catalytic activity (Figure 4(d)). One side of FTO's conserved jelly-roll motif is covered by an extra loop, not found in other AlkB family members. This loop competes with the nondamaged DNA strand for binding FTO, and likely plays an important role in maintaining ssDNA selectivity. Hydrogen-bond interactions between active site residues and the carbonyl oxygen atoms of 3mT or 3mU likely account for the substrate specificity of FTO.

Other Responses against DNA-Alkylation Damage

AidB

E. coli AidB is a part of the *ada* regulon of the *E. coli* adaptive response. Although AidB provides some cellular protection against S_N1 methylating agents, its exact function is unknown. AidB, related to the acyl-CoA dehydrogenase (ACAD) family of flavoproteins in sequence and fold, is a tetramer composed of a dimer of dimers, with the two FAD cofactors located at each of the dimer interfaces. Although the active site conserves many of the ACAD active site general features, notable differences include blocking and alteration of the putative substrate access channel environment by a conserved tryptophan and differences in the FAD cavity, suggesting it is not an active site for fatty or amino acid acyl-CoA thioesters. DNA is thought to bind in one of two places: an approximately 20-Å-wide, highly electropositive concave groove or near the positively charged substrate access channel mouth. The AidB tetramer ends bind DNA nonspecifically. Together, these observations suggest that AidB protects DNA by deactivating alkylating agents, rather than repairing it.

Lesion Bypass Polymerases

O^6 -mG lesions block the replicative DNA polymerases Pol α and Pol δ . However, bypass of these lesions by translesion synthesis (TLS) DNA polymerases may provide a means of survival in mammalian cells. DNA polymerase η can insert nucleotides opposite O^6 -mG *in vitro*, and other DNA polymerases, such as polymerase κ or polymerase ζ , may extend from an O^6 -mG:T mismatch, indicating that several TLS polymerases may be needed for O^6 -mG bypass. *In vivo* sensitivity to methylating agents is seen for cells defective in the polymerase ζ catalytic subunit, REV3, and further increased through inactivation of polymerase κ .

Mismatch Repair

The DNA MMR pathway repairs base substitution mismatches and small insertion or deletion mismatch loops formed during DNA replication. MMR also recognizes and attempts to repair O^6 -mG:T mispairs, but this repair leads to apoptosis. If MMR is

inactivated, O^6 -mG is highly mutagenic but not cytotoxic. The mechanism of how MMR leads to apoptosis is controversial. One long-held explanation is that MMR erroneously removes the thymine, and subsequent resynthesis by DNA polymerase restores the mispair. This leads to a futile cycle of nucleotide removal and synthesis, resulting in DNA single- and double-strand breaks and eventually apoptosis. Another proposal is that MMR complex binding to O^6 -mG:T mispairs directly signals apoptosis. Regardless of mechanism, it is likely that MMR proteins induce apoptosis by binding O^6 -mG:T mispairs and signaling their presence.

Alkyltransferase-Like Proteins and Connection to Nucleotide Excision Repair

The DNA nucleotide excision repair (NER) pathway repairs bulky, unrelated, helix-distorting lesions by excising a long DNA patch containing the lesion. Although NER can also repair O^6 -alkylG and O^4 -alkylT *in vitro*, its recognition of these nonbulky adducts is relatively poor. Alkyltransferase-like (ATL) proteins are AGT homologs that provide a connection for efficient O^6 -alkylG repair by NER (Figure 5). ATLs lack AGT's nucleophilic cysteine and are catalytically inactive, yet they protect cells against alkylation damage. ATLs tightly bind and promote repair of various O^6 -alkylG lesions, including many bulky lesions not repaired by AGTs. *S. pombe* and *Vibrio parahaemolyticus* ATLs retain the AGT C-terminal domain structural fold and conserve identities and roles of key DNA-binding and nucleotide flipping residues. ATL flips O^6 -alkylG into its binding pocket, analogous to hAGT, with the exception that here, nucleotide flipping is not connected to enzymatic activity. The ATL-induced 45° DNA bend, significantly more than the 30° DNA bend achieved by hAGT, is accompanied by an open-to-closed binding site loop conformational change, not seen for hAGT. Together, these observations suggest for ATL, nucleotide flipping is a switch for pathway activation rather than catalysis.

ATLs provide cellular protection by promoting a novel repair mechanism for O^6 -alkylG damage through cross-talk between DNA-alkylation base damage responses and NER, thereby linking two classically distinct DNA repair pathways. *In vitro*, *E. coli*, *S. pombe*, and *Thermus thermophilus* ATLs interact with NER proteins. An *in vivo* epistatic relationship has been shown for *S. pombe* and *E. coli* ATLs with *S. pombe* NER proteins Rad13 and Swi10 (homologs of human XPG and ERCC1) and *E. coli* NER protein UvrA, respectively, establishing ATL's connection to NER. In organisms where AGT and ATL coexist, such as *E. coli*, O^6 -alkylG adducts are repaired by the AGT or NER pathways according to their size, leading to a proposed enhanced-NER pathway, where ATL mediates bulky O^6 -alkylG repair by NER, whereas AGT repairs O^6 -mG (Figure 5). Although most known ATLs are bacterial or yeast, ATLs have recently been discovered in archaea and the multicellular starlet sea anemone *Nematostella vectensis*, showing that ATLs exist in all three domains of life.

Conclusion

DNA-alkylation damage, prevalent within cells, can have detrimental cellular effects if left unrepaired. Cells use a vast

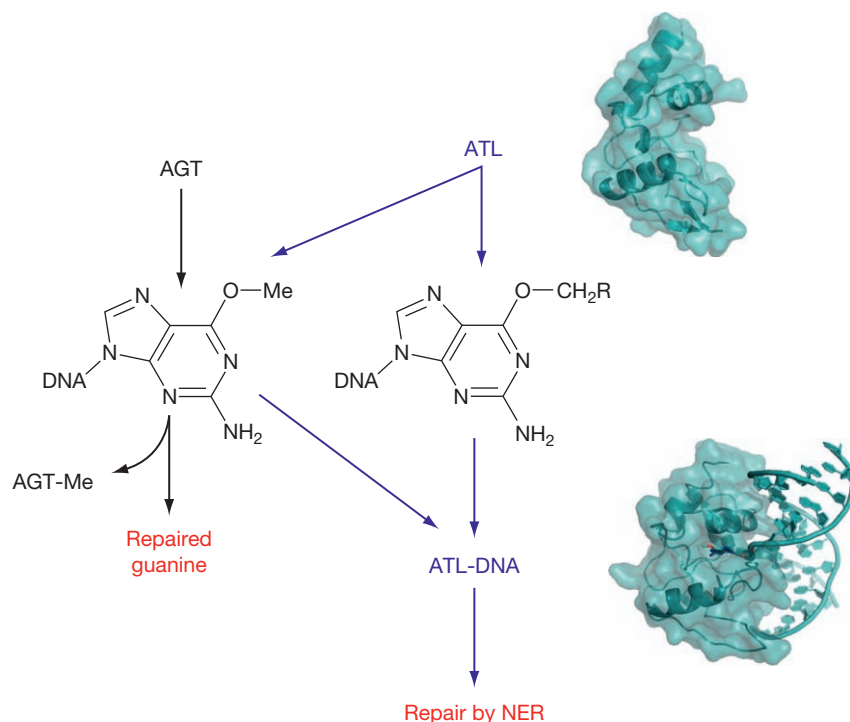


Figure 5 ATL binds O^6 -alkylG to direct its repair into the NER pathway. If AGT and ATL coexist, ATL and AGT compete for O^6 -alkylG repair. ATL apo (open) and DNA-bound (closed) structures show the nonenzymatic nucleotide flip switch that signals repair by NER.

number of DNA repair mechanisms to protect them from these potentially lethal effects. These include the main dealkylation mechanisms of BER and direct damage reversal, and the recently discovered ATL protein-mediated channeling of alkylation damage away from classical base responses and into NER. The significant threat to genomic stability posed by DNA alkyl lesions is reflected in their extensive and creative repair mechanisms within the cell.

Acknowledgments

Research on DNA base repair by the authors is supported by NIH grants CA097209 and GM046312.

See also: **Molecular Biology:** DNA Base Excision Repair Pathways; DNA Glycosylases: Mechanisms; DNA Mismatch Repair and the DNA Damage Response; DNA Mismatch Repair in Bacteria; DNA Mismatch Repair in Disease and Aging; DNA Mismatch Repair in Mammals; Nucleotide Excision Repair, Bacterial: The UvrABCD System; Nucleotide Excision Repair: Biology; Nucleotide Excision Repair in Eukaryotes.

Further Reading

Dalhus B, Laerdahl JK, Backe PH, and Bjørås M (2009) DNA base repair – recognition and initiation of catalysis. *FEMS Microbiology Reviews* 33: 1044–1078.
Hitomi K, Iwai S, and Tainer JA (2007) The intricate structural chemistry of base excision repair machinery: Implications for DNA damage recognition, removal, and repair. *DNA Repair* 6: 410–428.

Huffman JL, Sundheim O, and Tainer JA (2005) DNA base damage recognition and removal: New twists and grooves. *Mutation Research* 577: 55–76.
Lee CY, Delaney JC, Kartalou M, et al. (2009) Recognition and processing of a new repertoire of DNA substrates by human 3-methyladenine DNA glycosylase (AAG). *Biochemistry* 48: 1850–1861.
Margison GP, Butt A, Pearson SJ, et al. (2007) Alkyltransferase-like proteins. *DNA Repair* 6: 1222–1228.
Mishina Y, Duguid EM, and He C (2006) Direct reversal of DNA alkylation damage. *Chemical Reviews* 106: 215–232.
Mishina Y and He C (2006) Oxidative dealkylation DNA repair mediated by the mononuclear non-heme iron AlkB proteins. *Journal of Inorganic Biochemistry* 100: 670–678.
Pegg AE (2000) Repair of O^6 -alkylguanine by alkyltransferases. *Mutation Research* 462: 83–100.
Rubinson EH, Metz AH, O'Quin J, and Eichman BF (2008) A new protein architecture for processing alkylation damaged DNA: The crystal structure of DNA glycosylase AlkB. *Journal of Molecular Biology* 381: 13–23.
Sedgwick B (2004) Repairing DNA-methylation damage. *Nature Reviews: Molecular Cell Biology* 5: 148–157.
Seeberg E, Eide L, and Bjørås M (1995) The base excision repair pathway. *Trends in Biochemical Sciences* 20: 391–397.
Shrivastav N, Li D, and Essigmann JM (2010) Chemical biology of mutagenesis and DNA repair: Cellular responses to DNA alkylation. *Carcinogenesis* 31: 59–70.
Sundheim O, Talstad VA, Vågbø CB, Slupphaug G, and Krokan HE (2008) AlkB demethylases flip out in different ways. *DNA Repair* 7: 1916–1923.
Sundheim O, Vågbø CB, Bjørås M, et al. (2006) Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage. *EMBO Journal* 25: 3389–3397.
Tubbs JL, Latypov V, Kanugula S, et al. (2009) Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature* 459: 808–813.
Tubbs JL, Pegg AE, and Tainer JA (2007) DNA binding, nucleotide flipping, and the helix–turn–helix motif in base repair by O^6 -alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. *DNA Repair* 6: 1100–1115.
Wyatt MD and Pittman DL (2006) Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chemical Research in Toxicology* 19: 1580–1594.
Yang CG, Yi C, Duguid EM, et al. (2008) Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* 452: 961–965.

DNA Glycosylases: Mechanisms

A C Drohat and A Maiti, University of Maryland School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Daniel J. Krosky and James T. Stivers, volume 1, pp. 614–617, © 2004, Elsevier Inc.

Glossary

Abasic or apurinic/aprimidinic site A deoxyribose residue in DNA that lacks a nucleobase.

DNA glycosylase Enzymes that recognize and remove damaged (chemically altered) bases from DNA.

N-Glycosylic bond The bond connecting the anomeric carbon of the deoxyribose to the nitrogen of the purine or pyrimidine base.

Nucleotide flipping The process by which an enzyme extrudes a nucleotide from the DNA helix into the active site (also known as base flipping).

Transition state The state of most positive Gibbs free energy through which reactants must pass during the transformation to products. The difference in free energy between the reactants (ground state) and the transition state, termed the activation barrier, determines the rate of a reaction.

Introduction

DNA glycosylases recognize and remove damaged bases in DNA and are essential for maintaining genomic integrity, and some glycosylases play an important role in transcriptional regulation and the immune system. The four nucleobases of DNA are reactive and subject to continuous chemical modification, giving a large variety of lesions resulting from alkylation (i.e., 3-methyladenine), oxidation (8-oxoguanine), deamination (cytosine to uracil), and ultraviolet (UV) radiation (pyrimidine dimers). These lesions can be mutagenic, due to miscoding properties during DNA replication, and/or cytotoxic, because they can block DNA or RNA polymerases. Damage to individual bases in DNA is handled by the base excision repair (BER) pathway, one of the several DNA repair systems found in all forms of life.

We can think of BER in terms of two basic phases – recognition and excision of the altered base by a damage-specific DNA glycosylase, and subsequent steps that replace the damaged nucleotide using the undamaged strand as a template. DNA glycosylases use a remarkable nucleotide-flipping mechanism to extrude damaged nucleotides from the DNA helix into their active site, where they cleave the base-sugar (*N*-glycosylic) bond to release the damaged base. All organisms are armed with many different DNA glycosylases (11 are found in human cells), some of which are highly specific for a single lesion, while others can remove a broad range of lesions.

In addition to handling spontaneous DNA damage, some DNA glycosylases are required for other important cellular processes. Uracil DNA glycosylase (UNG) contributes to the acquired immune response by excising uracils created in antibody genes (by active cytidine deaminase), leading to antibody diversity through somatic hypermutation or class-switch recombination. In addition, replication of retroviruses and retrotransposons is inhibited by degradation of reverse transcripts via cytidine deamination and UNG-mediated uracil excision. In plants, DNA glycosylases mediate gene imprinting by excising 5-methylcytosine (m^5C), a modified base that signals for transcriptional silencing. Two DNA glycosylases appear to

participate in transcriptional regulation in mammals, thymine DNA glycosylase (TDG) and methyl binding domain IV (MBD4), although their role likely involves excision of T from G-T mismatches created by active m^5C deamination rather than excision of m^5C directly.

Types of DNA Glycosylases

DNA glycosylases are classified as either monofunctional, which have base-excision (glycosylase) activity only, or bifunctional, which also have DNA lyase activity (Figure 1). The mechanism of the glycosylase reaction and subsequent steps of the BER pathway differs substantially for these two types of enzymes. Monofunctional enzymes cleave the glycosylic bond by catalyzing a reaction in which water is the nucleophile, producing a common product, an apurinic/aprimidinic (AP) or abasic site in the DNA. This mutagenic and toxic BER intermediate is recognized and processed by an AP endonuclease, which hydrolytically cleaves the 5'-phosphodiester bond at the AP site. AP endonucleases yield a 3'-OH primer for repair synthesis, and other BER enzymes complete the repair process. Bifunctional enzymes excise damaged bases by catalyzing a glycosyl transfer reaction, where the nucleophile is an amine provided by the enzyme (Lys side chain or the amino terminus of the protein). The resulting Schiff base intermediate undergoes β -elimination, promoted by the lyase activity, cleaving the C3'-O bond of the damaged nucleotide. The remaining portion of the damaged nucleotide is recognized and removed by an AP endonuclease, leaving a 3'-OH primer for repair synthesis. Some bifunctional enzymes also catalyze a δ -elimination reaction that cleaves the C5'-O bond, which completely excises the damaged nucleotide. Following β - or β/δ -elimination, downstream BER proteins complete the repair process.

DNA glycosylases can be categorized according to overall fold into one of six superfamilies, some of which contain many enzymes that together recognize a very broad range of lesions. The helix-hairpin-helix (HhH, or Endo III) superfamily is a large and diverse collection of mono- and bifunctional enzymes that is represented by *Escherichia coli* enzymes MutY

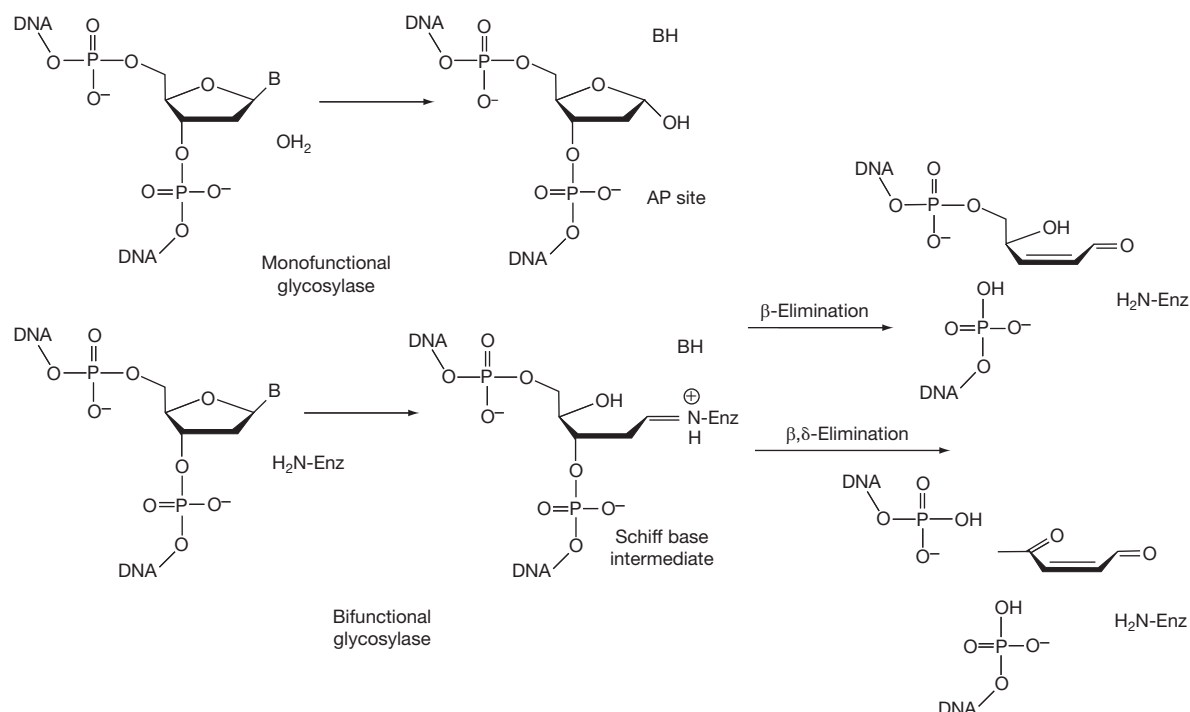


Figure 1 DNA glycosylases recognize and remove damaged bases by cleaving the base-sugar (*N*-glycosylic) bond. Monofunctional DNA glycosylases catalyze a hydrolytic reaction (water is the nucleophile) and produce an abasic (or AP) site, and downstream enzymes complete the repair process. Bifunctional enzymes cleave the glycosidic bond using an amine as the nucleophile (Lys side chain or amino terminus of protein), giving a Schiff base intermediate. The bifunctional enzymes also have lyase activity, and can cleave the DNA backbone 3' to the lesion (β -elimination) or 3' and 5' to the lesion (β,δ -elimination). Downstream BER enzymes restore the correct nucleotide.

(this removes adenine from A-8-oxoG pairs) and the alkylation damage-repair enzymes AlkA and TAG, and eukaryotic 8-oxoguanine glycosylase (OGG1). The helix-two-turn-helix (H2TH) superfamily includes bifunctional enzymes that handle oxidized bases, the MutM and Nei glycosylases. The UNG superfamily includes many monofunctional enzymes, some of which can remove a variety of uracil analogs and other modified pyrimidines. The alkyladenine DNA glycosylase (AAG) family of monofunctional enzymes, represented by human AAG, handles a broad range of lesions. The HEAT-like repeat (HLR) superfamily, including AlkC and AlkD, removes positively charged methylpurines and is found in prokaryotes and lower eukaryotes. The sixth superfamily, pyrimidine dimer glycosylases (PDGs), are bifunctional enzymes that mediate excision of cyclobutane pyrimidine dimers, and is best represented by the Endo V enzyme from bacteriophage T4.

Nucleotide Flipping: A Universal Trait

DNA glycosylases are charged with the daunting task of finding and removing damaged bases within the huge (million-fold) excess of undamaged DNA. Often, the lesion differs only slightly from the undamaged base and does not substantially perturb the overall DNA structure, complicating the search process. To overcome this problem, DNA glycosylases extrude the damaged nucleotide from the helix into an active site, a reversible step called nucleotide flipping (or base flipping) that precedes cleavage of the glycosylic bond (Figure 2). Nucleotide

flipping allows glycosylases to form interactions required for lesion recognition, provides access to the scissile C–N bond, and places the lesion in an active site with a chemical environment that promotes catalysis. The conformational change is quite large, as the target nucleotide is rotated by up to 180° out of the DNA helix. The finding that all glycosylases employ nucleotide flipping reflects the importance of this step in the enzymatic reaction, which couples lesion recognition and base excision. Nucleotide flipping is also used by other enzymes such as alkylguanine transferases that modify DNA and RNA, and it was first observed for a cytosine methyltransferase.

DNA glycosylases use several approaches to promote nucleotide flipping, which is energetically costly due to disruption of base pairing and stacking interactions and unfavorable DNA backbone conformational changes. Glycosylases typically interact with two or three phosphodiester groups on each side of the target nucleotide, and some enzymes contact phosphodiester groups in the complementary DNA strand. These contacts enable sequence-independent DNA binding, allow for rapid DNA scanning, and provide a scaffold for enzymatic distortion of the DNA to promote and/or stabilize nucleotide flipping. Structures of glycosylases bound to DNA, with a target lesion flipped into the active site, reveal that nucleotide flipping involves bending of the DNA helix at the lesion site. The degree of bending, and the point along the nucleotide-flipping pathway at which it occurs, depends on the glycosylase. Although our understanding of exactly how DNA bending facilitates nucleotide flipping remains incomplete, computational studies indicate that DNA bending lowers the energetic barrier to

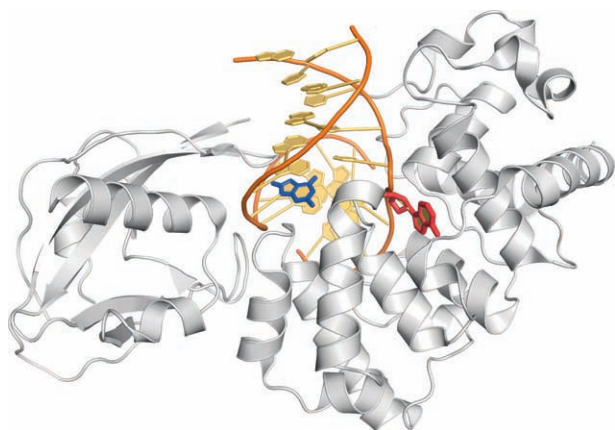


Figure 2 DNA glycosylases use a remarkable process known as nucleotide flipping (also called base flipping) to extrude nucleotides from the helix into their active site, where the glycosidic bond is cleaved. Shown is the crystal structure of MutY (*Bacillus stearothermophilus*) with the target nucleotide (dA, red) flipped into its active site (PDB ID: 3G0Q). MutY also recognizes the 8-oxoguanine (blue) that has been base-paired with the target dA in the DNA helix. Recognition of 8-oxoG, which remains stacked in the helix, is important for avoiding enzymatic excision of adenine from the million-fold excess of A-T base pairs. The figure was generated by Pymol, and the structure was solved by Lee and Verdine. From Lee S and Verdine GL (2009) Atomic substitution reveals the structural basis for substrate adenine recognition and removal by adenine DNA glycosylase. *Proceedings of the National Academy of Sciences of the United States of America* 106: 18497–18502.

nucleotide flipping. Glycosylases also promote and/or stabilize nucleotide flipping by inserting one or more amino acid side chains into the DNA helix. For example, most glycosylases fill the helical space vacated by the flipped nucleotide using a bulky side chain (Asn, Arg, Gln, Leu, or Tyr). Some insert a wedge between bases in the complementary DNA strand, typically adjacent to the base that is paired with the target nucleotide. Finally, many glycosylases form selective steric and/or electrostatic interactions with the target lesion in the active site, which stabilize nucleotide flipping and can trigger a conformational change that positions active-site residues for catalysis. Thus, most glycosylases promote nucleotide flipping by bending the DNA, plugging the helical space vacated by the flipped nucleotide, and/or retaining the lesion in the active site.

While some glycosylases appear to initiate nucleotide flipping by selectively disrupting base-pairing interactions to push target nucleotides out of the helix, other glycosylases simply trap nucleotides that spontaneously flip due to normal breathing of the DNA base pairs. Nuclear magnetic resonance and crystallographic studies show that UNG uses a passive trapping mechanism to find dU lesions in DNA. UNG traps dU and dT nucleotides that flip spontaneously from the DNA helix into an 'exo-site' that is distinct from the active site. The exo-site lies along the nucleotide-flipping pathway and interacts transiently with dU and dT nucleotides that have flipped $\sim 40^\circ$ out of the helix. UNG does not alter the rate at which dT or dU nucleotides flip out of the helix; rather, it slows the rate at which they flip back into the helix. However, dT does not flip stably into the UNG active site due to steric clashes with its C5-methyl group (dU has a hydrogen at C5). The exceedingly short

lifetime for the open state of an A·U base pair ($<0.1 \mu\text{s}$) indicates that UNG cannot simply bind dU nucleotides that have already flipped. Kinetics experiments show that the residence time of UNG at any given site in DNA ($\sim 0.5 \text{ ms}$) is long enough to encounter many spontaneous flipping events.

OGG1 has an exo-site that interacts transiently with extrahelical dG and oxo-dG nucleotides, but only oxo-dG flips fully into the active site. Single-molecule studies of OGG1 indicate that it slides along the DNA at a very rapid rate, sampling hundreds of base pairs before dissociating, and its residence time at any given site appears too short to trap oxo-dG nucleotides that flip spontaneously from DNA. This suggests that OGG1 selectively promotes the flipping of oxo-dG nucleotides, because a mechanism whereby all dG residues were flipped into the exo-site would be too time consuming to account for the observed rapid sliding on DNA. The process of searching for oxo-dG lesions is more fully understood for the MutM glycosylase, which also slides rapidly along the DNA. Structural studies indicate that MutM interrogates DNA using a process that involves DNA bending and selective distortion of oxo-G·C base pairs to promote flipping of oxo-dG lesions.

The Catalytic Mechanism of DNA Glycosylases

The *N*-glycosidic bond is quite stable for the four normal nucleotides in DNA, and most damaged nucleotides, and glycosylases employ a number of strategies to meet the significant catalytic challenge of cleaving this bond. The spontaneous hydrolysis of 2'-deoxynucleosides is exceedingly slow, and occurs with a half-life ranging from 12 years for dG to 43 years for dC, at pH 7 and 37°C . The nucleotides are even more stable in duplex DNA, where spontaneous hydrolysis occurs with a half-life of 400 and 11 000 years for purines and pyrimidines, respectively, suggesting that glycosylases attain some degree of catalysis by simply flipping the target nucleotide into their active site. DNA glycosylases increase the rate of glycosidic bond cleavage by 10^{12} -fold compared to the corresponding nonenzymatic reaction. We first consider the mechanism of enzymatic and noncatalyzed reactions, and then discuss how glycosylases catalyze these reactions.

Noncatalyzed *N*-glycosyl hydrolyses and *N*-glycosyl transfer reactions follow a stepwise or a highly dissociative mechanism, where rupture of the glycosidic bond fully precedes or is greatly advanced over the nucleophile approach (Figure 3). The transition state (TS) involves positive charge on the sugar and migration of negative charge into the departing base. The detailed reaction mechanism has been determined for two DNA glycosylases, UNG and MutY, both monofunctional enzymes (water is the nucleophile). The UNG reaction is stepwise, where cleavage of the glycosidic bond in the first TS gives a short-lived oxacarbenium ion–uracil anion intermediate, and water attacks the cationic sugar in the second TS (Figure 1). The reaction catalyzed by MutY is also stepwise, but the glycosidic bond breaks and re-forms many times before nucleophilic attack, which is the first irreversible step. Remarkably, this indicates that the most probable event following glycosidic bond cleavage is re-formation of the bond and dissociation of MutY from DNA substrate.

DNA glycosylases use one or more of three basic strategies to catalyze the excision of damaged bases from DNA in a timely

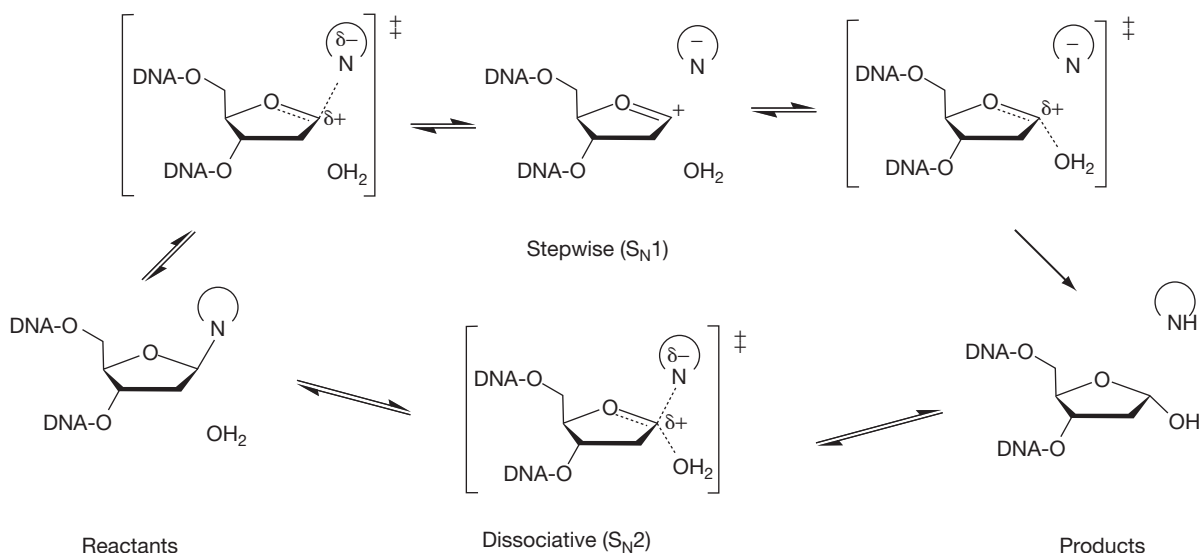


Figure 3 Two potential mechanisms for 2'-deoxynucleoside hydrolysis. A stepwise reaction (S_N1, upper pathway) has two transition states, one for C–N bond cleavage and one for nucleophile (water) addition, separated by a short-lived oxacarbenium ion intermediate. Hydrolysis of 2'-deoxynucleosides can also follow a highly dissociative mechanism (S_N2, lower pathway), where glycosidic bond cleavage is greatly advanced over nucleophile approach, and significant positive charge develops on the sugar in the transition state (TS). A stepwise mechanism is observed for nonenzymatic dAMP hydrolysis, and for enzymatic hydrolysis of dU in DNA by UNG and of dA in DNA by MutY and ricin. For the MutY reaction, the glycosidic bond is cleaved and reformed many times before water is added in the first irreversible step. In contrast, for UNG, the first irreversible step is likely C–N bond cleavage. While MutY protonates the adenine base prior to glycosidic bond cleavage, which activates it for departure, UNG does not protonate the uracil base, which departs as an anion (as indicated above).

manner to preserve genomic integrity. These include (1) stabilization of unfavorable positive charge that develops on the sugar and negative charge that develops on the departing base, (2) protonation of the base to activate it for departure (general acid catalysis, particularly for purines), and (3) orientation and/or activation (via deprotonation) of the nucleophile to promote attack of the anomeric carbon.

Cleavage of the glycosidic bond involves transfer of electrons into the nucleobase, which is unfavorable because the anionic base is a poor leaving group and a good nucleophile. DNA glycosylases can enhance the leaving ability of the base by protonation (general acid catalysis), which activates it for departure because the protonated base is electron poor. Alternatively, glycosylases can stabilize the negative charge that develops on the base in the TS (electrostatic catalysis). Detailed studies of the UNG reaction show that the uracil base departs as an anion, which is stabilized by a strong hydrogen bond provided by a conserved His side chain (to uracil O2). It is likely that excision of uracil or thymine by other glycosylases involves departure of the pyrimidine anion. In contrast, MutY catalyzes cleavage of dA nucleotides by transferring a proton to adenine N7 prior to bond cleavage, activating the adenine base for departure. Some alkylated bases carry a positive charge (are electron deficient) and are good leaving groups in the absence of enzyme activation, a feature that can be used by glycosylases to distinguish the alkylated base from its undamaged counterpart.

DNA glycosylases also attain substantial rate enhancement by stabilizing the glycosyl cation that develops in the TSs of the reaction. UNG stabilizes the oxacarbenium ion using a constellation of proximal anionic groups, including a conserved Asp side chain, phosphates of the DNA backbone, and the departing anionic uracil base. The observation that many glycosylases

bind more tightly to DNA containing a positively charged mimic of the oxacarbenium ion TS than to DNA containing a neutral mimic of the abasic site product suggests that they stabilize the glycosyl cation in a manner similar to UNG. However, unlike UNG, many glycosylases do not have a carboxylate side chain to stabilize the glycosyl cation, and for many glycosylases, the base departs in its neutral rather than anionic form. Thus, some glycosylases may stabilize the cationic sugar using primarily backbone phosphates, and some may stabilize TS by other means, including activation of the leaving group or the nucleophile.

Another strategy for TS stabilization for many DNA glycosylases is activation and/or positioning of the nucleophile for attack at the anomeric carbon (C1'). Many glycosylases have a conserved Asp or Glu side chain that can abstract a proton to activate the nucleophile (general base catalysis) or at least orient the nucleophile for attack. Some glycosylases have an Asn in the corresponding position, which can orient but not activate the nucleophile. The observation that many glycosylases have a nonpolar side chain at this position (i.e., Ala or Gly) indicates that the nucleophile can be activated by other means or that nucleophile activation is not needed for these glycosylases. Consistent with this conclusion are findings that for some glycosylases, substitution of the conserved Asp or Glu residue does not dramatically alter catalytic activity.

How Do DNA Glycosylases Attain Specificity?

DNA glycosylases use a number of approaches to meet the formidable task of finding and removing lesions while avoiding the million-fold excess of normal DNA. While some DNA

glycosylases are specific for one type of base damage, others are permissive and can remove a broad range of lesions. For many glycosylases, the active site is sterically restrictive and provides electrostatic interactions that select for the cognate lesion, excluding other lesions and undamaged DNA. UNG is highly specific for removing uracil from DNA; its active site sterically rejects adenine, guanine, and thymine, and it forms hydrogen bonds with uracil that are not compatible with cytosine. UNG rapidly excises uracil but cannot remove thymine due to steric hindrance of the CH₃ at C5 of thymine (uracil has a hydrogen at C5). In addition, some glycosylases exhibit strong specificity not only for a particular target nucleotide, but also for its base-pairing partner in the complementary strand. MutY is highly specific for removing adenine when it is base-paired with 8-oxoguanine (A-oxoG), but does not remove adenine from normal A-T pairs. The active site of prokaryotic TAG enzymes is highly specific for the excision of 3-methyladenine.

Given the limiting number of DNA glycosylases and the much larger number of potential lesions, some glycosylases must handle multiple types of damage. Not surprisingly, these enzymes have a relatively permissive active site and typically provide few selective interactions with their various target bases. For example, *E. coli* AlkA removes a broad range of purines and pyrimidine lesions arising predominantly not only from alkylation but also from deamination and oxidation, and AlkA can remove undamaged bases, albeit at a greatly reduced rate. The specificity of AlkA stems in part from certain properties of the damaged base. For example, alkylated bases carry a positive charge and are pre-activated for departure, such that the glycosylic bond is more readily cleaved. In addition, some damaged bases exhibit impaired hydrogen bonding and/or stacking interactions and are prone to be extrahelical, which promotes tighter binding to AlkA and faster base excision. Similar observations were made for human AAG, which removes a range of lesions, mostly alkylpurines and some deaminated purines such as hypoxanthine.

Some glycosylases with a permissive active site exhibit strong specificity for the base-pairing partner of the target nucleotide. Members of the TDG–mismatch uracil glycosylase (MUG) family, represented by human TDG and *E. coli* MUG, remove a variety of lesions and exhibit strong specificity for excising bases paired with guanine rather than adenine, which is consistent with findings that G-U or G-T mispairs are their predominant biological substrate. Given the strong specificity of TDG–MUG enzymes for removing bases paired with G, they

must employ a stringent mechanism to avoid excision of cytosine from normal G-C pairs. The broad substrate range of TDG, which includes modified cytosines, indicates that the specificity does not involve selective substrate interactions that exclude cytosine from the active site. Instead, specificity of TDG for G-T over G-C pairs arises largely at the chemical step of the reaction, where the scissile *N*-glycosylic bond is much more readily cleaved for dT relative to dC. Other glycosylases with an indiscriminate active site attain specificity at the chemical step. AlkA removes many lesions with a rate that depends on *N*-glycosylic bond stability.

See also: **Molecular Biology: DNA Damage: Alkylation; DNA Oxidation; Mammalian DNA Methyltransferase Structural Themes.**

Further Reading

- Bennett MT, Rodgers MT, Hebert AS, et al. (2006) Specificity of human thymine DNA glycosylase depends on *N*-glycosidic bond stability. *Journal of the American Chemical Society* 128: 12510–12519.
- Berti PJ and McCann JA (2006) Toward a detailed understanding of base excision repair enzymes: Transition state and mechanistic analyses of *N*-glycoside hydrolysis and *N*-glycoside transfer. *Chemical Reviews* 106: 506–555.
- Dalhus B, Laerdahl JK, Backe PH, and Bjoras M (2009) DNA base repair – recognition and initiation of catalysis. *FEMS Microbiology Reviews* 33: 1044–1078.
- David SS, O'Shea VL, and Kundu S (2007) Base-excision repair of oxidative DNA damage. *Nature* 447: 941–950.
- Fromme JC, Banerjee A, and Verdine GL (2004) DNA glycosylase recognition and catalysis. *Current Opinion in Structural Biology* 14: 43–49.
- Hitomi K, Iwai S, and Tainer JA (2007) The intricate structural chemistry of base excision repair machinery: Implications for DNA damage recognition, removal, and repair. *DNA Repair (Amst)* 6: 410–428.
- Lee S and Verdine GL (2009) Atomic substitution reveals the structural basis for substrate adenine recognition and removal by adenine DNA glycosylase. *Proceedings of the National Academy of Sciences of the United States of America* 106: 18497–18502.
- O'Brien PJ (2006) Catalytic promiscuity and the divergent evolution of DNA repair enzymes. *Chemical Reviews* 106: 720–752.
- O'Brien PJ and Ellenberger T (2004) The *Escherichia coli* 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site. *Journal of Biological Chemistry* 279: 26876–26884.
- Parker JB, Bianchet MA, Krosky DJ, et al. (2007) Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature* 449: 433–437.
- Priyakumar UD and MacKerell AD (2005) Computational approaches for investigating base flipping in oligonucleotides. *Chemical Reviews* 106: 489–505.
- Qi Y, Spong MC, Nam K, Banerjee A, et al. (2009) Encounter and extrusion of an intrahelical lesion by a DNA repair enzyme. *Nature* 462: 762–766.
- Stivers JT (2008) Extrahelical damaged base recognition by DNA glycosylase enzymes. *Chemistry* 14: 786–793.

DNA Helicases: Hexameric Enzyme Action

D Sen and S S Patel, UMDNJ-Robert Wood Johnson Medical School, Piscataway, NJ, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Smita S. Patel, volume 1, pp. 624–631, © 2004, Elsevier Inc.

Glossary

Dissociation constant (K_d) Measures the affinity between enzyme and its substrate. It has molar (M) unit, and smaller the number, tighter the binding/affinity.

Holliday junction A four-stranded DNA structure that is an intermediate in DNA recombination.

Introduction

Since its discovery in 1976 from *Escherichia coli*, we have come a long way in understanding helicases. It is now clearer than ever that these nucleic acid (NA) motor proteins are found ubiquitously and helicases alone constitute approximately 2% of all encoded proteins in eukaryotes. Helicases use the energy of nucleoside triphosphate (NTP) hydrolysis to perform an array of NA (RNA or DNA)-related functions including translocating on single-stranded (ss) and double-stranded (ds) NA, separation of the dsDNA strands, removal of the secondary structures in RNA, displacing proteins bound to NAs, remodeling chromatin, and moving the Holliday junction. Given the pivotal role played by helicases in NA metabolism, it is not surprising that, mutations in human genes coding for helicases are associated with premature aging and various types of cancers. In viruses, helicases play a key role in genome replication and its packaging/unpackaging. Hence, viral helicases are prospective targets for developing antiviral therapy. The topic of helicase structure, function, and mechanisms has been widely discussed in the existing literature. This review focuses primarily on the structure and mechanisms of ring-shaped hexameric helicases.

Ring-shaped helicases are found across all species, from bacteria to humans (Table 1). The most commonly studied ring-shaped helicases are DnaB, RepA, and Rho of the *E. coli*, T7 gp4, G40P, and T4 gp41 of bacteriophages, SV40 large T antigen and E1 of eukaryotic viruses, minichromosome maintenance (MCM) 2–7 in eukaryotes and mitochondrial helicase TWINKLE in humans. Hexameric helicases can broadly be divided into two types based on their conserved adenosine triphosphatase (ATPase) core, namely, RecA and AAA⁺. Most of the RecA-type helicases move in the 5′–3′ direction on ssNA, while the AAA⁺-type helicases move in the 3′–5′ direction.

Hexameric Helicase Structure with ssNA

The uniqueness of hexameric helicases lies in their structure. Structural studies show that the six subunits are arranged like a toroid with an outer diameter of ~120 Å, and a ~80-Å-long central channel with a diameter of ~20 Å (Figure 1). This three-dimensional structure of the hexameric helicase is conserved across bacteriophage, viruses, prokaryotes, and eukaryotes.

Most ring-shaped helicases are homo-hexamers, except for the eukaryotic MCM helicases, that assemble into a hetero-hexamer of MCM2–7 proteins. Biochemical and electron microscopy data have indicated that the NA binds in the central channel. The crystal structures of *E. coli* Rho and papilloma virus E1 (with 379 amino acids deleted from its N-terminal domain) helicases provided the first glimpse of the interactions of the ssNA within the central channel (Figure 1). The structures of Rho (with adenosine diphosphate.beryllium fluoride (ADP.BeF₃)) and E1 (with ADP) show five subunits of the six interacting with five nucleotides of ssNA – each subunit interacting with one nucleotide of ssNA. Two loops from each subunit protrude into the central channel forming a ‘spiral staircase’-like structure to interact with the sugar–phosphate backbone of the NA explaining the sequence-independent binding of helicases to ssNA. Despite their structural differences and opposite directionality of translocation – E1 being an AAA⁺ family of 3′–5′ helicase and the Rho being a Rec A family of 5′–3′ helicase – both bind ssNA with the same polarity. The 5′ end of the ssNA (top of the staircase) is proximal to the N-terminal domain of both the helicases. Structures suggest that Rho and E1 fire adenosine triphosphate (ATP) in opposite sequence of directions around the ring to drive NA translocation in opposite directions.

Loading of the Ring-Shaped Helicases on the NA

The annular conformation of hexameric helicases raise an obvious question: How does the NA enter the central channel? Odds are probably low under *in vivo* conditions (although possible *in vitro*) for the NA threading itself into the central channel like a thread through a needle. Most ring-shaped helicases rely on proteins called the loaders to help it encircle NA (Table 1). A loader is either a separate protein that interacts with its corresponding helicase and with the NA (e.g., DnaC for DnaB helicase in *E. coli*) or helicase’s own specific domain endowed with the loading task (e.g., primase domain in T7 gp4 protein and N-terminal RNA-binding domain in Rho).

These proteins help to load the helicase on the NA by acting either as ring breakers or as ring makers. Helicases with low K_d for hexamerization are stable rings without the NA. In this case, the loader needs to break the ring either by completely disassembling the ring or, more likely, by creating openings between the subunits to get the NA into the central hole.

Table 1 RecA and AAA⁺ ring-shaped helicases

Protein fold	Name	Polarity	Function	Loader
RecA	Bacterial DNAB	5'–3'	DNA unwinding	DNAC
	Phage T7 gp4	5'–3'	DNA replication	N-terminal primase domain
	Phage T4 gp41	5'–3'	DNA unwinding	gp59
	G40P	5'–3'	DNA unwinding	G39P
	Bacterial Rho	5'–3'	RNA–DNA heteroduplex unwinding and transcription termination	N-terminal domain
	Human mitochondrial TWINKLE	5'–3'	DNA unwinding	?
	Φ 12 P4	5'–3'	ssRNA packaging	?
AAA ⁺	pRSF1010RepA	5'–3'	DNA unwinding	?
	Papilloma virus E1	3'–5'	DNA unwinding	E2
	Simian virus 40 Large T-antigen	3'–5'	DNA unwinding	Loads on its own. Loader domain?
	Eukaryotic or archaeal MCMs	3'–5'	DNA unwinding	Cdc6–Cdt1 in <i>S. cerevisiae</i>
	Bacterial RuvB	dsDNA	Resolving Holliday junctions	RuvA

For helicases that are not stable hexamers, the loader actually builds the ring by putting the monomeric subunits together around the NA.

ATP binding/hydrolysis by the loaders is believed to play a crucial part in loading of hexameric helicases. When a separate protein acts as a loader, the loaders are ATPases, for example, Cdc18 of *Saccharomyces pombe* Mcm2-7 is an AAA⁺ ATPase, Cdc6 of *S. cerevisiae* Mcm2-7 requires ATP hydrolysis for loading, and DnaC for *E. coli* DnaB is an ATPase. Where two separate loaders are involved, at least one performs ATP hydrolysis; for example, DnaI for *Bacillus subtilis* DnaC is an ATPase only in the presence of another loader DnaB and this ATPase activity is further stimulated in the presence of ssDNA, but DnaI does not hydrolyze ATP in the absence of DnaB. However, for helicases with a domain of its own acting as a loader, ATP hydrolysis is not required for loading. For example, the N-terminal domain of Rho does not show any ATP-binding motif. For phage T7 gp4, the presence of deoxythymidine triphosphate (dTTP), but not its hydrolysis, is required for hexamer assembly around the DNA. Hence, the role of ATPase might be as a gatekeeper at the transition of loading and unwinding.

The protein loaders interact directly with NA in addition to their substrate helicase. This interaction has been reported in loaders such as DnaB, the co-loader for *B. subtilis* DnaC, DnaC the loader for DnaB in *E. coli*, gp59 loader for phage T4 gp41 helicase, Cdc6 loader for *S. cerevisiae* MCM 2–7, primase domain of phage T7 gp4, and RNA-binding domain of Rho, to mention a few. There is no information available for heterologous loader loading a helicase; hence, it is reasonable to assume that the interaction between loader and its corresponding helicase is extremely specific. For example, DnaC in *E. coli* binds the N-terminal domain of DnaB helicase.

After loading is completed, the helicase does not lug the loader along during unwinding. For example, DnaC loader for DnaB helicase in *E. coli* simply acts as a loader and gatekeeper between loading and unwinding steps and dissociates after loading. Similarly, Cdc6–Cdt1 dissociate from Mcm2–7 in *S. cerevisiae* after loading by hydrolyzing ATP.

The scenario in the cell is more complex than *in vitro* conditions for loading, in which forked junctions or single-stranded loading conformations are not readily available in the genome. Therefore, in addition to the loader, accessory

proteins (initiators) are required for origin remodeling and setting the stage for helicase loading and unwinding. Even though the complexity of genome of bacteria and yeast is incomparable, it has been observed that the accessory proteins for both function in a similar manner; they form the origin recognition complex at the origin, open up the duplex NA to form an ss bubble-like structure, followed by turning on the loader that loads the helicase at the origin to form the prereplicative complex, finally loading the polymerase and primase leading to replication. It is possible that the same mechanism could be extrapolated to metazoans as well.

Even when ss structures are available in the genome at regions of accidental replication fork arrest, the loading of helicase is regulated by specific proteins that form complexes at the site and attract the helicase and its loader. An example for this mechanism can be seen in *E. coli*, where the primosome complex (PriA, PriB, PriC, and DnaT) attracts DnaC–DnaB to the site of replication fork arrest.

The complexity of loading of hexameric helicases on NA has begun to be unveiled by biochemical and structural studies. Future probably lies in visualizing the dynamics of these events by a congruence of single-molecule and *in vivo*-imaging techniques.

Mechanisms of NTPase-Powered Translocation of Ring Helicases

Loading of the helicase is followed by its translocation on NA, which is fueled by NTP hydrolysis. The RecA or AAA⁺-like NTPase core binds and hydrolyzes NTP to nucleoside diphosphate (NDP) and inorganic phosphate (Pi) in the presence of the essential cofactor Mg²⁺. Most helicases use ATP as the substrate, unlike others such as T7 gp4 which has a specific requirement for dTTP. The first glimpse of the NTPase active site was obtained from the crystal structure of the T7 gp4 without any bound NA. This structure showed that the NTPase active sites are located at the subunit interfaces, currently recognized as a general feature of hexameric helicases. Amino acids from adjacent subunits participate in NTP binding and hydrolysis. The Walker A and B motifs from one subunit cooperate with the arginine finger from its neighboring subunit to bind

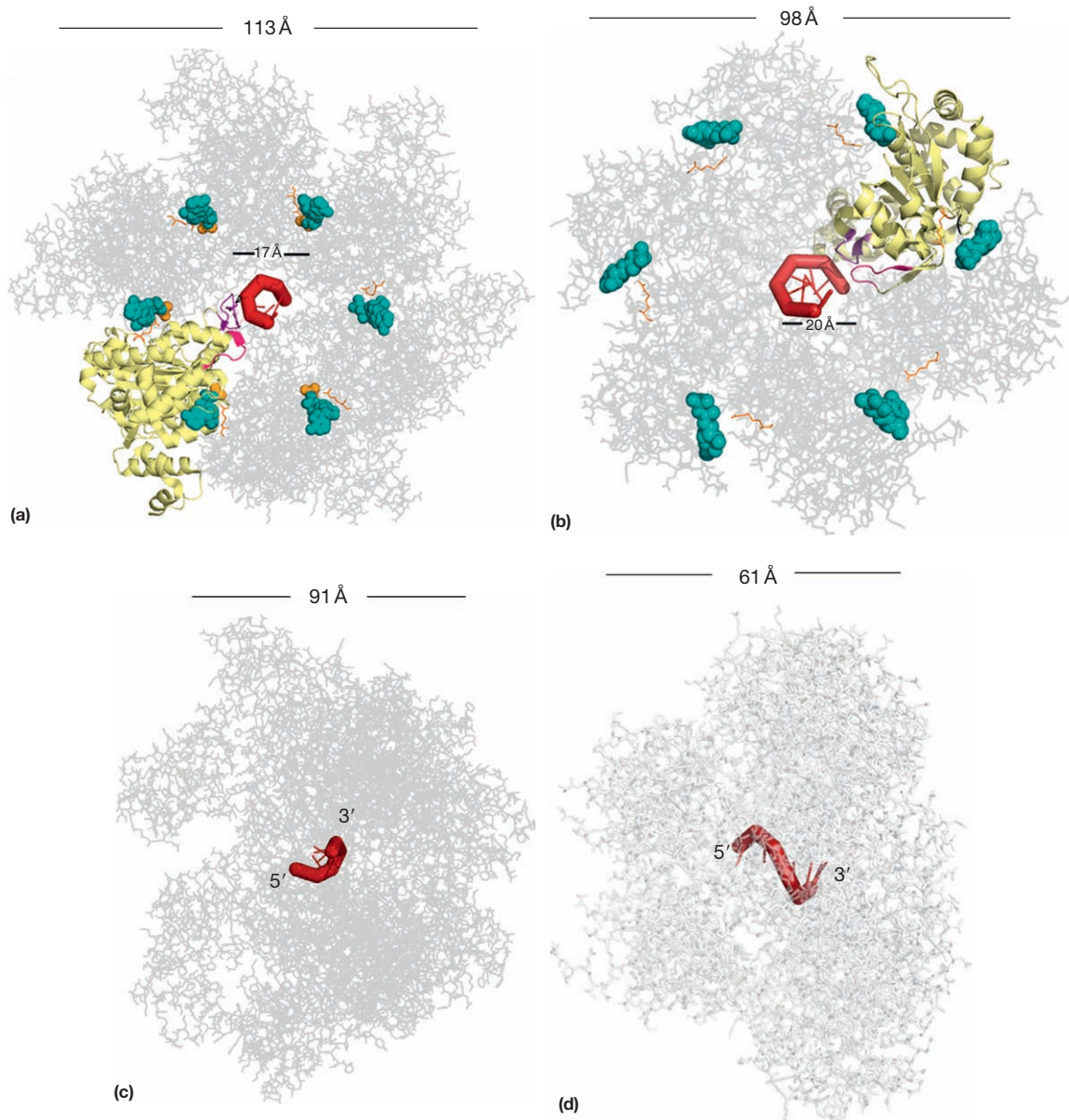


Figure 1 Structures of *E. coli* Rho and BPV E1 helicases demonstrating the structural similarity among two classes of hexameric helicases RecA and AAA⁺: (a) and (b) represent the top views of Rho (3ICE) and $\Delta 379$ E1 (2GXA), respectively. The six subunits of the hexamer assume a toroidal shape with a central channel accommodating the ssNA. A single chain of the hexamer is highlighted (yellow) to show the two loops, which forms a staircase-like structure that interacts with ssNA (brown) in the central channel. The cyan spheres represent the ADP•BeF₃ (in orange) in Rho and ADP in E1. The Arginine finger is the yellow stick interacting with the Walker B motif of the adjacent chain. (c) and (d) represent the side views of Rho and E1, respectively, with the N-terminal domain to the left and C-terminal domain to the right. The brown strand represents the ssNA in the central.

NTP and catalyze its hydrolysis (Figure 1). This interfacial mode of NTP binding provided the initial model for the structural basis for the intersubunit transmission of the changes rooting from NTP binding and hydrolysis to DNA-binding/unbinding reactions in the ring. Biochemical studies have shown that the hexamer binds NTP, NTP analogs (nucleoside 5'-(β,γ -imido) triphosphate (NMPPNP), nucleoside (β,γ , methylene) triphosphate (NMPPCP), NDP.BeF₃), and NDP with micromolar K_d values and stoichiometry of 4–6 NTPs/hexamer.

The NA affinity of the hexameric helicase depends on its nucleotide-ligation state. For example, T7 gp4 binds ssDNA tightly in the presence of dTTP without Mg²⁺, an essential cofactor for NTPase. This binding is unaffected if dTTP is replaced by its nonhydrolyzable analog deoxythymidine (β,γ , methylene) (dTTP-PCP), but becomes 500 times weaker in the presence of deoxythymidine diphosphate (dTDP). Similar observation was recorded with the DnaB helicase that binds ssDNA with 200-fold tighter affinity with adenylyl-imidodiphosphate

(AMP.PNP) as compared to ADP. From these observations, it was proposed that NTP binding and Pi release promote DNA binding and release events, respectively. The exact sequence of ATP hydrolysis around the ring, which brings about NA binding and unbinding, is not known.

Mutant poisoning experiments indicated that one or two inactive mutant subunits in the ring can inactivate the helicase. This ruled out a mechanism whereby NTPs are hydrolyzed randomly and suggested a mechanism whereby NTP hydrolysis is coordinated among the six subunits. Initial presteady-state dTTPase studies in the absence of DNA suggested that three out of the six sites of T7 gp4 were active, similar to the F1-ATPase. Presteady-state NTPase kinetics experiments of Rho and T7 gp4 with ssNA showed hydrolysis of more than three NTPs. These results suggested the involvement of all six subunits of the hexameric helicase in catalysis. Although, there is a general agreement that the NTPase reactions are coordinated among the subunits of the ring, it is not known if the sequence of events, ATP binding, hydrolysis, and product release, resulting in NA movement is conserved across helicases.

Based on presteady-state NTPase studies and protein–DNA cross-linking data of T7 gp4 that indicated binding of one to two subunits to ssNA at any time, a 2-site sequential mechanism was proposed wherein two subunits were proposed to be involved in the ssNA binding–unbinding coupled to ATP binding and hydrolysis in the following manner (Figure 2): upon binding NTP, a subunit makes tight interactions with the NA, and upon NTP hydrolysis the subunit loses its grip on the NA. The complete release of the NA is prevented by the cooperative action of two adjacent subunits. Before the NA is completely released by a subunit, the neighboring subunit assumes the NTP-liganded state and binds the NA at a forward position resulting in processive translocation of the NA. Hence, at any given point of time, only two of the six subunits are actively playing a role in NTPase-coordinated translocation.

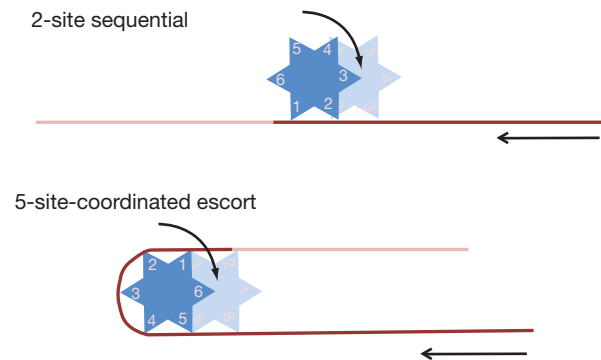


Figure 2 Models of single-stranded nucleic acid translocation by hexameric helicase. The star (blue) represents the hexamer with the inner core (central channel) exposed out. The inner NA-binding loops and the NTP-binding domains of the hexamer are splayed open to clearly show the interaction of the NA with each domain. The red strand represents NA and each subunit of the hexamer is represented as individual triangles in the star. The 2-site model postulates that at any time two subunits interact with the NA and during translocation if subunit-1 is unbinding NA, the subunit-3 is binding NA. The 5-site coordinated escort model postulates that any time five subunits interact with the NA, and during translocation as subunit-1 is unbinding NA, subunit-6 is binding NA.

An alternative coordinated escort model was proposed from a snapshot provided by the crystal structures of E1 and Rho helicases that showed that ssNA was bound to five subunits of the ring (Figure 2). The major difference between the 2-site sequential and coordinated escort model is two versus five subunits bound to ssNA at any time. However, both the models agree that while one subunit is binding NA another one is unbinding from the NA.

According to the coordinated escort model, the reactions of ATP binding and product release are coordinated with ssNA binding–unbinding by the subunits at the end of the staircase. In E1, the subunit at the top of the staircase (near the 5′ end of the NA) binds ATP and makes new interactions with an ssNA base, while the subunit near the 3′ end of the ssNA at the bottom of the staircase releases itself from the NA base. Thus, E1 moves in the 3′–5′ direction along the ssNA. The events in Rho are exactly the opposite as it moves with opposite polarity in the 5′–3′ direction. There, the subunit at the bottom of the staircase (3′ end of the RNA) makes new interactions with a NA base as it binds ATP and the subunit at the top releases a NA base as it releases Pi. The middle subunits are proposed to be in different states of prehydrolysis, hydrolysis, and post-hydrolysis states.

Based on the asymmetric binding of the ADP in E1 and ADP.BeF₃ in Rho at the six ATPase sites, the subunits are designated as ATP type, ADP type, or empty type. The subunits at one end of the staircase are ATP type (or prehydrolysis), the middle are ADP type (or hydrolysis competent), and at the other end of the staircase are open or empty type. The subunits that bind ATP tightly also appear to bind NA tightly, while the subunit that binds ATP weakly binds NA weakly as well.

The coordinated escort model is extremely appealing in terms of explaining how NTPase coordinated translocation of NA works to bring about processive movement. Future structures with ATP rather than ADP or ATP analogs will provide additional snapshots to define the proposed pathway. In addition, such intermediate structures will determine if all five subunits remain bound to ssNA through multiple cycles of ATPase.

Additional ways to verify the model are to measure the step sizes of the helicase. Both 2-site sequential model and 5-site coordinated escort model predict a mechanical step size (distance moved per ATPase) of 1 nucleotide. If a rate-limiting step occurs after each ATPase cycle, then a kinetic step size (number of bases translocated between two rate limiting steps) of one nucleotide should be observed. A kinetic step size of >1 nucleotide would indicate alternative mechanisms where multiple ATPase cycles lead to multiple base movement with intermediate rate-limiting step (similar to the stepping mechanism of Φ29 dsDNA-packaging motor, see below). Single molecule NA translocation studies are achieving high resolution to measure small steps and can distinguish between these possible mechanisms.

Translocation of dsDNA

Ring-shaped ATPases such as the packaging motors, FtsK, and RuvB bind and translocate on dsDNA without destabilizing the base pair (bp). Studies of the Φ29 dsDNA-packaging motor indicate a distinct mechanism of dsNA translocation.

The kinetics of dsDNA translocation by the $\Phi 29$ dsDNA-packaging motor were measured by single molecule optical tweezer experiments, which showed repetitive lag and burst phases of translocation. The lag phase depended on the ATP concentration and became shorter with increasing ATP, while the burst phase indicated translocation of 10 bp of dsDNA. These data appear to be inconsistent with the one ATP – one-step mechanism proposed for the ssNA translocases. Hence, a new latch-lever mechanism was proposed to explain the phenomenon, where multiple ATP molecules are loaded into the ring during the lag phase. Sequential ATP hydrolysis–product release then drives translocation of 10-bp dsDNA in steps of 2.5 bp with the hydrolysis of 1 ATP/2.5 bp. It was proposed that only a subset of the subunits binds dsDNA and translocation is brought about by global conformational changes in the ring. Many aspects of the above mechanism are yet to be tested by future ATPase kinetics and structural studies of the motor protein with NA bound to it.

Although studies of ssNA and dsNA translocases indicate more than one way of coordinating the ATPase and NA binding–unbinding cycles, it could be possible that there is a single underlying mechanism of NA translocation by ring ATPases that will gradually emerge as similar techniques of studying and interpreting the two kinds of motors happen. It is also possible that the packaging motors evolved differently from ssNA translocation motors to generate larger forces required to package the genomes in the capsids. Helicases such as T7 gp4, MCM, and DnaB can translocate on both ssDNA and dsDNA, and given the different structures of ssNA versus dsNA, it indicates that the central channel and the subunit coordination in the same helicase can be adjusted to accommodate both dsNA and ssNA translocation.

Mechanisms of NA Strand Separation

Replicative hexameric helicases typically show a high processivity in unwinding dsDNA. It is believed that helicases such

as T7 gp4, T4 gp41, *E. coli* DnaB, MCM unwinds dsDNA by translocating along ssDNA that they encircle while excluding the complementary strand of the dsDNA from the central channel at the fork junction (Figure 3). This mode of DNA binding (strand-exclusion model) minimizes immediate reannealing of the DNA strands after they are separated. Whether the excluded strand plays a steric role and simply excludes the helicase from encircling the dsDNA and translocating along it without strand separation or whether the excluded strand interacts with the outside region of the helicase ring to aid in strand separation is not established.

Many eukaryotic hexameric helicases like the viral SV40 large T antigen and bovine papilloma virus (BPV) E1, initiate unwinding by first binding to dsDNA of a specific sequence. These helicases destabilize the origin to initiate DNA replication with the help of other accessory proteins. However, it is not known if after melting the dsDNA at the origin, these helicases switch from dsDNA-binding mode to the strand-exclusion mode to catalyze replication fork movement. Models, such as ploughshare and ds melting mechanism, have been proposed as alternatives to the strand-exclusion model to explain DNA unwinding at the origin and translocation of the replication fork by helicases that do not require a fork to begin unwinding.

NA strand separation is typically measured *in vitro* using short dsDNA substrates either radiolabeled or fluorophore labeled. The strand-separation kinetics is monitored by polyacrylamide gel electrophoresis or by a stopped-flow fluorescence assay. Reannealing of the strands in the time-scale of the experiment is prevented by the use of low nM concentrations of DNA substrates in the assay and excess of the helicase protein. Recent advances in single-molecule studies have made it possible to study NA separation reaction by using fluorescence resonance energy transfers (FRETs) and magnetic or optical tweezers providing rates, sequence dependencies, pausing, and stepping behaviors of the helicase translocation.

The mechanism of NA strand separation has been investigated in some detail. NA strand separation involves breaking

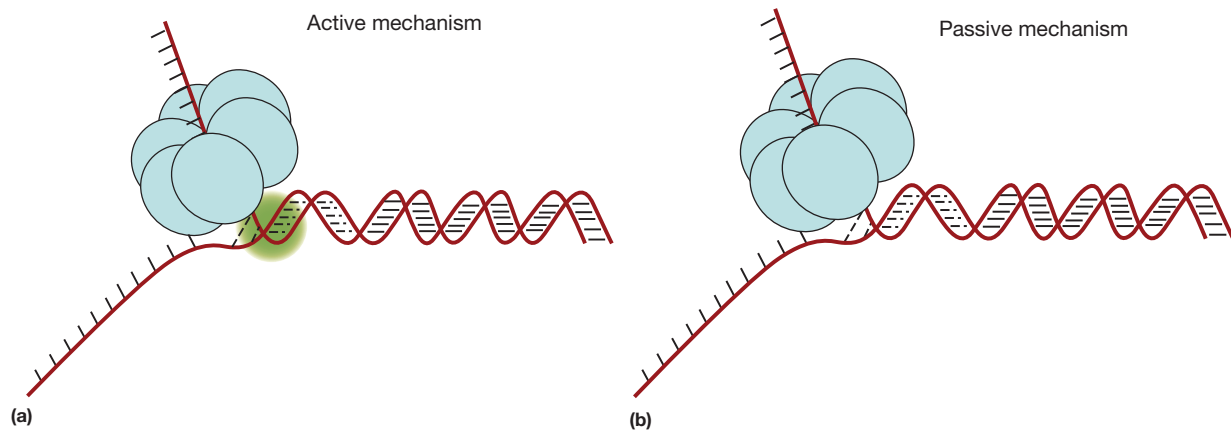


Figure 3 Cartoon representing (a) active and (b) passive mechanisms of helicase unwinding. (a) In active mechanism of unwinding the helicase exerts a force (represented by the green cloud) at the fork-junction leading to breaking of the hydrogen bonds resulting in unwinding. (b) In passive mechanism of unwinding the helicase is an opportunist waiting to stabilize the broken base pair created by NA breathing, which happens at a faster rate at the forked junction and at a slower rate in the internal base pairs. In both cases, the helicase loads on one strand while excluding the complementary strand from its central channel.

the intermolecular forces mainly hydrogen bonds that hold the two strands together. Kinetic measurements show that T7 gp4, T4 gp41, DnaB translocate at a faster rate along ssDNA compared to their rates of dsDNA strand separation. Moreover, their dsDNA strand-separation rate depends on the GC content or the DNA stability. This raises the question as to whether helicases unwind the strands of the NA by a passive or an active mechanism (Figure 3).

In the passive mechanism, the helicase translocates on the ssNA as it is generated from thermal fraying of the dsNA. At physiological temperature, NA bp at the ss/ds junction open and close spontaneously, commonly referred to as 'breathing' of the NAs. The rates of spontaneous bp opening as measured by imino proton exchange, range from 30 s^{-1} for internal bp to 1000 s^{-1} for bp at the ss/ds junction. Although, individual bp opening is a fast process, the equilibrium for the reaction, $\text{ss} \rightleftharpoons \text{ds}$, lies toward bp formation and depends on the NA sequence and stability. A helicase-unwinding NA by a purely passive mechanism would be greatly affected by the NA sequence and stability as compared to a helicase-unwinding NA by an active mechanism. In the active mechanism, the helicase actively disrupts the bp at the junction to shift the $\text{ss} \rightleftharpoons \text{ds}$ equilibrium toward the left. Typically, the free energy of an AT bp is $1.3 k_B T$ and for the GC bp is $2.9 k_B T$. A passive helicase would have destabilization energy of $0 k_B T$ and an active one greater than that.

The destabilization energy for two ring-shaped helicases has been determined from rate measurements of DNA unwinding as a function of force or DNA sequence. By applying defined forces to the ss/ds junction using optical or magnetic tweezer, the junction bp can be destabilized to different extents. Similarly, by using DNAs with different GC content, the stability of the junction bp can be controlled. These experiments showed that T7 gp4 destabilizes the junction by $1\text{--}2 k_B T \text{ bp}^{-1}$, while $3.4 k_B T$ would be required to achieve the maximal translocation rate, indicating that T7 gp4 unwinds by a partially active mechanism.

Similar experiments showed that T4 gp41 destabilizes the bp by $0.05 k_B T$ and, thus, acts largely by a passive mechanism. It appears that force production by helicases can vary depending on their functions in the cell and by their association with accessory proteins. For example, replicative helicases work with motor proteins such as the DNA polymerase, which greatly increase helicase rates of strand separation.

Conclusions

A large number of helicases and helicase-like proteins have been identified by genome-sequencing efforts. With a basic understanding of the helicase activity, it is fairly clear that they have a quintessential role in DNA metabolism. Hence, any mutations in helicases can have serious consequences, and many are known to be associated with human diseases such as cancer and premature aging. Mutations in ERCC2 gene coding a helicase leads to *Xeroderma Pigmentosum*, Bloom's syndrome is caused by mutation in a gene coding for RecQ homolog, mutations in gene coding for TWINKLE cause adPEO, Werner's syndrome is caused by a gene coding a helicase and the list continues to grow.

Many viruses code for their own helicases, which make them unique targets for antiviral therapy. Understanding the structure and mechanisms of helicases is an important means of finding the link between the complex human diseases and helicases, which, in turn, would aid the drug-discovery process. From the viewpoint of understanding how helicases work, there is still a lack of a general mechanism that would explain NA strand separation and other functions performed by helicases at the molecular level. Despite all the differences in the properties and structures of helicases, it is possible that they function by similar mechanisms. It is likely that unidirectional translocation is a basic activity of all helicases responsible for NA strand separation and all other functions of helicases. This makes helicases similar to general motor proteins that translocate on lattices whose mechanism of operation may apply to helicases as well, and given a chance all motor proteins may be able to act as a helicase under the right conditions. New methodologies such as analysis of single helicase molecules *in vitro* and *in vivo* will allow direct observation of the helicase reaction, and more detailed information from these experiments will provide physical parameters that will reveal how helicases work. Even though helicases are a crucial factor for NA metabolism, it is worth noting that they need accessory proteins to carry out their functions efficiently. Hence, studying helicase activities in the context to its accessory partners will help to define their mechanism more precisely.

See also: **Molecular Biology:** DNA Replication: Eukaryotic Origins and the Origin Recognition Complex; DNA Replication Fork, Bacterial; DNA Replication: Initiation in Bacteria; Nonhexameric SF1 DNA Helicases/Translocases; RecQ Helicase Systems; Sliding Clamps in DNA Replication: *Escherichia coli* β -Clamp and PCNA Structure; **Protein/Enzyme Structure Function and Degradation:** AAA-ATPases; Enzyme Kinetics.

Further Reading

- Adelman JL, Jeong YJ, Liao JC, et al. (2006) Mechanochemistry of transcription termination factor Rho. *Molecular Cell* 22(5): 611–621.
- Doublé S, Tabor S, Long AM, Richardson CC, and Ellenberger T (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* 391(6664): 251–258.
- Egelman EH, Yu X, Wild R, Hingorani MM, and Patel SS (1995) Bacteriophage T7 helicase/primase proteins form rings around single-stranded DNA that suggest a general structure for hexameric helicases. *Proceedings of the National Academy of Sciences of the United States of America* 92(9): 3869–3873.
- Enemark EJ and Joshua-Tor L (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* 442: 270–275.
- Hall MC and Matson SW (1999) Helicase motifs: The engine that powers DNA unwinding. *Molecular Microbiology* 34: 867–877.
- Hamdan SM and Richardson CC (2009) Motors, switches, and contacts in the replisome. *Annual review of Biochemistry* 78: 205–243.
- Johnson DS, Bai L, Smith BY, Patel SS, and Wang MD (2007) Single-molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase. *Cell* 129(7): 1299–1309.
- Levin MK and Patel SS (2003) Helicases as molecular motors. In: Schliwa M (ed.) *Molecular Motors*, pp. 179–198. Weinheim: Wiley-VCH Verlag GmbH.
- Liao JC, Jeong YJ, Kim DE, Patel SS, and Oster G (2005) Mechanochemistry of T7 DNA helicase. *Journal of Molecular Biology* 350(3): 452–475.
- Lohman TM and Bjornson KP (1996) Mechanisms of helicase catalyzed DNA unwinding. *Annual Review of Biochemistry* 65: 169–214.

- Patel SS and Donmez I (2006) Mechanisms of Helicases. *Journal of Biological Chemistry* 281: 18265–18268.
- Patel SS and Picha KM (2000) Structure and function of hexameric helicases. *Annual Review of Biochemistry* 69: 651–697.
- Pyle AM (2008) Translocation and unwinding mechanisms of RNA and DNA helicases. *Annual Review of Biophysics* 37: 317–336.
- Singleton MR, Dillingham MS, and Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annual Review of Biochemistry* 76: 23–50.
- Skordalakes E and Berger JM (2003) Structure of the Rho transcription terminator: Mechanism of mRNA recognition and helicase loading. *Cell* 114: 135–146.
- Thomsen ND and Berger JM (2009) Running in reverse: The structural basis of translocation polarity in hexameric helicases. *Cell* 139: 523–534.
- von Hippel PH and Delagoutte E (2001) A general model for nucleic acid helicases and their “coupling” within macromolecular machines. *Cell* 104: 177–190.

DNA Ligases: Mechanism and Functions

A E Tomkinson and J A Della-Maria, University of Maryland School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Alan E. Tomkinson and John P. Leppard, volume 1, pp. 632–636, © 2004, Elsevier Inc.

Glossary

Adenylation The reaction in which DNA ligase interacts with ATP or NAD to form a covalent enzyme–adenylate complex.

Motif An amino-acid sequence found to be conserved in different proteins.

Phosphodiester bond These bonds link together deoxynucleotides in DNA, forming the sugar phosphate backbone of the DNA polymer.

Phosphoramidite bond A covalent bond formed between a phosphoryl group and an amino group. For DNA ligases, the phosphoryl group of AMP is linked to the amino group of the active site lysine.

DNA ligases are a universal feature of DNA-based life forms and have almost ubiquitous involvement in DNA transactions. Using biochemical and immunological approaches, the laboratory of Tomas Lindahl provided the first evidence that mammalian cells contain more than one species of DNA ligase. This was confirmed by the cloning of three mammalian *LIG* genes. More recently, genome sequencing has revealed the presence of multiple *LIG* genes in the genome of some prokaryotes. Interestingly, some of these novel bacterial DNA ligases also possess primase, polymerase, and nuclease activities within the same polypeptide.

Reaction Mechanism

DNA ligase activity was first identified in 1967 in five different laboratories. In the years that followed, the Lehman laboratory was primarily responsible for elucidating the three-step reaction utilized by nicotinamide adenine dinucleotide (NAD)-dependent and adenosine triphosphate (ATP)-dependent DNA ligases that is described below.

Formation of Covalent Ligase–AMP Covalent Intermediate

In the first step, DNA ligase reacts with either NAD or ATP to form a covalent enzyme–adenosine monophosphate (AMP) intermediate (Figure 1). The AMP moiety is linked via a phosphoramidate bond to a conserved lysine that defines the active site of the catalytic core (Figure 2). During step 1, the catalytic region is in an extended, open conformation but the formation of the enzyme–AMP complex does result in relatively small conformational changes that are required for recognition of nicked DNA in the next step of the ligation reaction.

Formation of Covalent DNA–AMP Intermediate

The second step in the reaction involves the recognition of a DNA nick and the transfer of the AMP group to the 5' phosphate terminus of the nick in duplex DNA (Figure 1). Engagement of nicked DNA is accompanied by a major change in protein conformation. Specifically, the DNA ligase catalytic region encircles

the DNA nick, forming a compact, ring-shaped structure. Within the DNA–protein complex, the AMP moiety is positioned such that an oxygen atom in the phosphate group at the 5' phosphate terminus of the nick can attack the phosphoryl group of AMP, generating the 5' DNA–AMP intermediate.

Phosphodiester Bond Formation

The third and final step of ligation is catalyzed by nonadenylated DNA ligase that presumably retains the compact, ring-shaped conformation when it interacts with the DNA–AMP intermediate. In this reaction, nucleophilic attack by the hydroxyl group at the 3' terminus of the DNA nick results in esterification of the 5' phosphoryl group to the 3' hydroxyl group with the concomitant release of AMP (Figure 1).

Organization and Structure of DNA Ligase Catalytic Domain

Alignment of DNA ligase amino acid sequences indicates that these enzymes share a conserved minimal catalytic core that contains adenylation and oligonucleotide/oligosaccharide-binding (OB)-fold subdomains (Figure 2). Within this minimal catalytic core, there are six conserved amino-acid sequence motifs that are characteristic of the nucleotidyl transferase superfamily. In addition, all eukaryotic DNA ligases have a conserved DNA-binding domain (DBD) that is situated N-terminal to the catalytic core and is required for efficient ligation (Figure 2). The zinc finger (ZnF) and helix–hairpin–helix (HhH) motifs in the NAD-dependent *Escherichia coli* (*E. coli*) DNA ligase (Figure 2) appear to play an analogous role in DNA binding to the DBD of eukaryotic DNA ligases. A combination of mutagenesis, enzymology, and, in particular, structural biology has advanced our understanding of the mechanism of DNA joining by DNA ligases.

X-ray crystallographic studies of both NAD- and ATP-dependent DNA ligases have provided important insights into the three-dimensional structure of these enzymes and the dynamic conformational changes that are likely to occur as the enzyme interacts with the nucleotide cofactor and

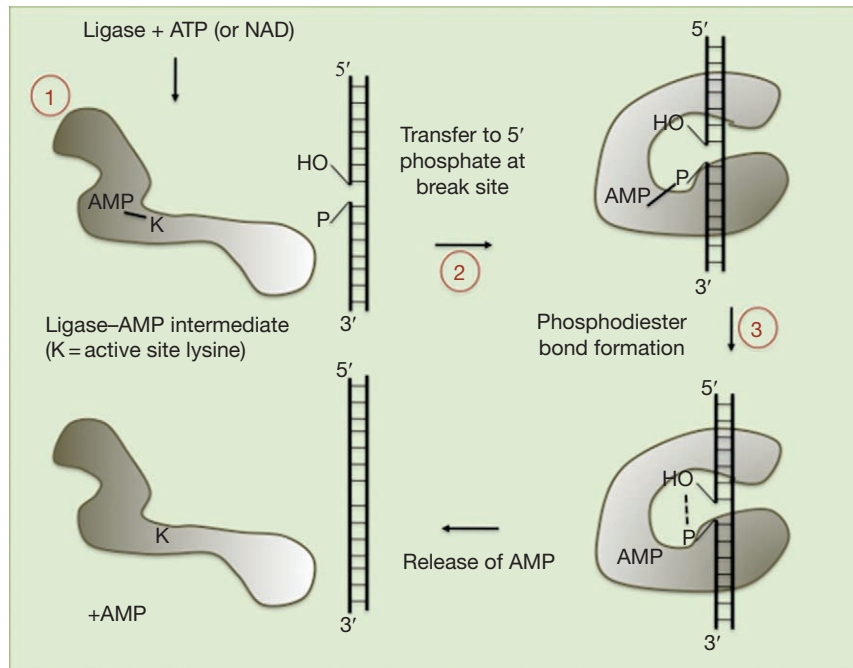


Figure 1 Mechanism of phosphodiester bond formation by DNA ligase. Step 1: DNA ligase interacts with ATP or NAD forming a covalent enzyme-AMP intermediate. Step 2: after nick recognition, the AMP group is transferred to the 5' phosphate terminus forming a high-energy phosphate bond. Step 3: DNA ligase catalyzes phosphodiester bond formation, releasing AMP.

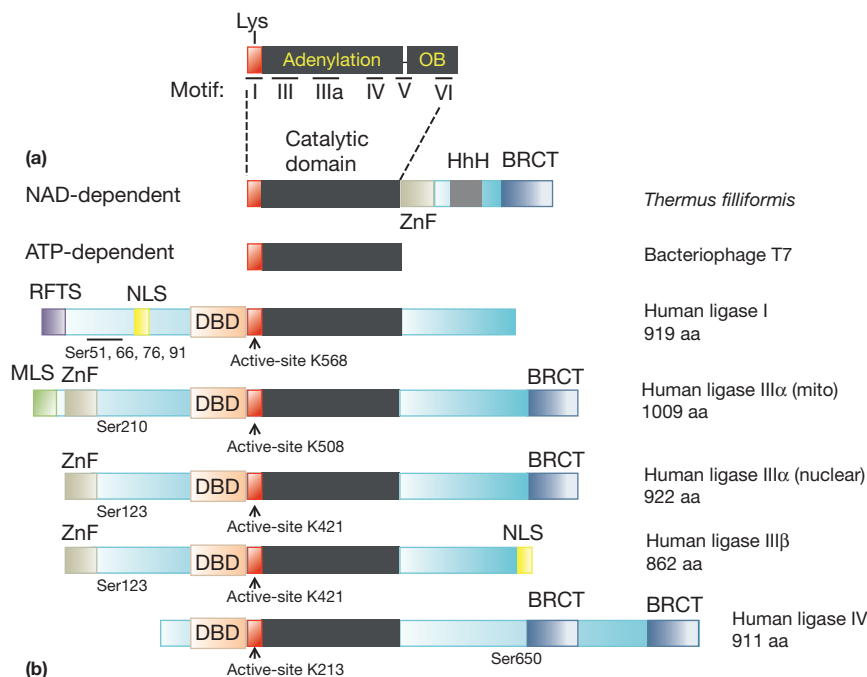


Figure 2 Features of the catalytic core and alignment of NAD- and ATP-dependent DNA ligases. (a) Schematic representation of the catalytic core shared by all DNA ligases that contains the six conserved motifs (I, III, IIIa, IV, V, and VI) characteristic of the NTases (nucleotidyl transferases); motif VI resides within the OB-fold domain that is separated from the adenylation domain by a linker. The active-site lysine is present within motif I, represented in red. (b) Comparison of a prokaryotic NAD-dependent DNA ligase (*Thermus filiformis*), and ATP-dependent DNA ligases from virus (bacteriophage T7), and humans (DNA ligases I, III α , and IV). DBD, DNA-binding domain; RFTS, replication factory-targeting sequence; ZnF, zinc finger motif; HhH, helix-hairpin-helix domain; BRCT, BRCA1 C-terminus domain; K, active lysine; OB, oligomer-binding; NLS, nuclear leader sequence; MLS, mitochondrial leader sequence; Ser, serine; aa, amino acid.

polynucleotide substrate. In 1996, the Wigley laboratory solved the first crystal structure of an ATP-dependent DNA ligase, the bacteriophage T7 DNA ligase, in a complex with ATP. This was followed by the structure of the enzyme–AMP intermediate formed by the *Chlorella* virus DNA ligase that was determined by the Shuman laboratory. A comparison of these structures revealed a conformational change in the DNA ligase catalytic core upon formation of the DNA ligase–AMP intermediate that allows binding to nicked DNA. Within the catalytic core, the larger adenylation subdomain is the minimum region required for formation of an enzyme–AMP intermediate whereas the OB-fold domain allows the enzyme to bind to DNA and coordinates the ligation event (Figure 2). Residues from motifs I through V line a cleft formed between the two subdomains to generate a positively charged nucleotide-binding pocket. The active-site lysine residue, which is contained within motif I, sits at the bottom of the cleft, close to where the adenylation domain and the OB-fold domain are linked. In the nonadenylated form, the enzyme is in an open conformation with the DNA-binding surface of the OB-fold domain rotated away from the active site thereby preventing nonproductive DNA binding. Upon adenylation, the enzyme undergoes a conformational change such that the DNA-binding surface of the OB-fold domain faces in toward the cleft, making the active site accessible.

The structure of a catalytically active fragment of human DNA ligase I determined by the Ellenberger laboratory was the first structure of a eukaryotic DNA ligase, but more importantly was the first structure of a DNA ligase complexed with nicked DNA. This revealed that the DNA ligase polypeptide encircles the DNA nick and distorts the DNA helix structure in the vicinity of the nick, exposing the nick termini. In this ring structure, the DBD, adenylation, and OB-fold domains each specifically interact with the nicked DNA. The DBD not only contacts the DNA upstream and downstream of the nick but also stabilizes the ring structure via protein–protein interaction with the adenylation and OB-fold domains. Subsequent studies have shown that the encircling of the DNA nick is a common feature of NAD-dependent DNA ligases and smaller ATP-dependent DNA ligases that lack the DBD domain.

DNA Ligase Polypeptides Encoded by the *LIG1*, *LIG3*, and *LIG4* Genes

The amino acid residues that flank the catalytic domain of ATP-dependent eukaryotic DNA ligases contain conserved domains

and/or motifs that mediate specific protein–protein interactions and protein–DNA interactions (Figure 2). These include the BRCT domain that was first identified in the breast cancer susceptibility gene 1 and is usually involved in protein–protein interactions, and the ZnF motif that is closely related to the two ZnFs that mediate DNA binding by poly (ADP-ribose) polymerase 1. The roles of the regions flanking the catalytic domain of ATP-dependent eukaryotic DNA ligases in defining the cellular activities of these enzymes are described below:

LIG1 (DNA Ligase I)

The unstructured N-terminal region of the DNA ligase I polypeptide is subjected to phosphorylation at several serine residues (Figure 2). These modifications occur during cell-cycle progression and regulate the participation of DNA ligase I in DNA replication and repair by modulating interactions with partner proteins. The N-terminal region of DNA ligase I also contains a nuclear localization signal (NLS) motif and a motif that acts as a replication factory targeting sequence (RFTS). Subsequently, it was shown that the RFTS corresponds to a proliferating cell nuclear antigen (PCNA)-binding motif (PIP box) and that the interaction with PCNA (Table 1) is critical for the participation of DNA ligase I in DNA replication and repair.

PCNA is a homotrimeric ring-shaped DNA-sliding clamp that was identified as a processivity factor for the replicative DNA polymerase, DNA Pol δ . Similarly, the interaction of DNA ligase I with PCNA rings, which are topologically linked to DNA, tethers DNA ligase I to the DNA. In addition, DNA ligase I also interacts with replication factor C (RFC), which loads the PCNA ring onto DNA and with the heterotrimeric DNA sliding clamp, Rad9–Rad1–Hus1, and clamp loader, hRad17–RFC, involved in cell-cycle checkpoints (Table 1).

LIG3 (DNA Ligase III)

Unlike DNA ligases I and IV, homologs of DNA ligase III are not present in lower eukaryotes, such as yeast, and may only be found in vertebrates. Notably, the *LIG3* gene encodes three distinct polypeptides (Figure 2). Nuclear and mitochondrial versions of DNA ligase III α are ubiquitously expressed from the same mRNA using different translation-initiation start sites. In addition, a germ-cell-specific form of DNA ligase III, DNA ligase III β , is generated by an alternative splicing mechanism in which the terminal 3'-coding exon in DNA ligase III α mRNA is replaced by a different exon, generating DNA ligase III β mRNA. The C-terminal 17–18 residues, which are unique

Table 1 Mammalian DNA ligases^a

Gene	Gene product	Interacting protein	Function
<i>LIG1</i>	DNA ligase I	PCNA, RFC, Pol β , Rad17–RFC, Rad9–Rad1–Hus1	DNA replication, BER, NER
<i>LIG3</i>	DNA ligase III α (nuclear)	XRCC1, NEIL1, TDP1, PARP1	SSBR, BER, NER, alt-NHEJ
	DNA ligase III α (mitochondrial)	Poly	Mitochondrial DNA metabolism
	DNA ligase III β	?	Postmeiotic repair and meiotic recombination?
<i>LIG4</i>	DNA ligase IV	XRCC4, APLF, Ku, XLF	NHEJ, V(D)J recombination

^aBER, base excision repair; NER, nucleotide excision repair; SSBR, single-strand break repair; NHEJ, nonhomologous end-joining; alt-NHEJ, alternative nonhomologous end-joining; PCNA, proliferation cell nuclear antigen; XRCC1, XRCC4, X-Ray cross complementing factor 1 and 4; NEIL1, nei endonuclease VIII-like 1; TDP1, tyrosyl-DNA phosphodiesterase 1; PARP1, poly (ADP-ribose) polymerase 1; APLF, aprataxin and PNKP like factor; XLF, XRCC4-like factor.

to DNA ligase III β , function as a nuclear localization signal (Figure 2), whereas the 77 residues that are unique to DNA ligase III α contribute to a BRCT domain that specifically interacts with the DNA repair protein, XRCC1 (Figure 2). The interaction with XRCC1 is required for the stability and activity of nuclear DNA ligase III α .

All the DNA ligase III isoforms have an N-terminal ZnF domain that is closely related to the two ZnFs in the DBD of PARP1. Although the DNA ligase III ZnF is not required for ligation of DNA nicks, it acts in concert with the DNA ligase III DBD to bind to DNA single-strand breaks and is required for intermolecular ligation by DNA ligase III. The region between the ZnF and DBD contains a serine residue, Ser123, which is progressively phosphorylated by Cdk2 as cells progress through S phase and is dephosphorylated in response to DNA damage in a reaction dependent upon the ataxia-telangiectasia-mutated protein (ATM). As the region of DNA ligase III containing this serine mediates interaction with different proteins (Table 1), it is possible that the phosphorylation status of Ser123 regulates these interactions.

LIG4 (DNA ligase IV)

DNA ligase IV has an extended C-terminal region that contains tandemly arrayed BRCT motifs (Figure 2). The linker region between the two BRCT motifs mediates an interaction with the DNA repair protein XRCC4 that is required for the stability and activity of DNA ligase IV. DNA ligase IV stability is also influenced by phosphorylation of Ser650 by the DNA-dependent protein kinase (DNA-PK). Unlike DNA ligases I and III, the majority of DNA ligase IV molecules in the DNA ligase IV/XRCC4 complex, after purification, are pre-adenylated and only catalyze a single ligation event. However, the DNA repair protein XLF, which interacts with XRCC4, stimulates the joining of compatible and incompatible DNA ends by DNA ligase IV/XRCC4.

Cellular Functions of Eukaryotic DNA Ligases

As DNA joining is required to complete DNA replication, DNA repair, and genetic recombination, it appears likely that the multiple species of DNA ligase have evolved to participate in specific DNA transactions. Insights into the cellular functions of the different DNA ligases have been obtained by various approaches including phenotypic analysis of DNA ligase-deficient cells and the identification of interacting proteins (Table 1). The results of these studies are summarized below.

DNA Replication

There is compelling evidence that DNA ligase I is predominantly responsible for joining Okazaki fragments generated by discontinuous DNA synthesis on the lagging strand at the replication fork. For example, DNA ligase I-deficient cells exhibit a marked defect in Okazaki fragment joining, and DNA ligase I physically and functionally interacts with the replication proteins, PCNA and RFC. The surprising ability of mouse *lig1* null cells to proliferate suggests that another DNA ligase, most likely DNA ligase III α , is able to join Okazaki fragments in the absence of DNA ligase I.

Excision Repair

Damaged and mispaired nucleotides are removed from the genome by excision repair pathways that share three common steps: (1) excision of the damaged or mispaired DNA; (2) gap-filling DNA synthesis using the undamaged strand as template; and (3) DNA ligation to complete the repair. In base excision repair, the DNA damage is removed as a nitrogenous base by DNA glycosylase. Two subpathways of BER, long patch and short patch, have been defined based on the extent of repair synthesis. Short-patch BER events are mostly completed by DNA ligase III α in a complex with its partner protein XRCC1, whereas long-patch BER is completed by DNA ligase I. The hypersensitivity to DNA alkylating agents of both DNA ligase I- and DNA ligase III-deficient cells indicates that these subpathways are not functionally redundant. It is possible that the DNA ligase III-dependent short-patch pathway, which also involves DNA polymerase β , is a housekeeping pathway whereas repair by the DNA ligase I-dependent long-patch pathway, which involves other DNA replication proteins such as RFC and PCNA, is associated with DNA replication. Recent studies suggest that a similar situation occurs in the nucleotide excision repair pathway, in which DNA damage is excised as an oligonucleotide. Once again, DNA ligase III α appears to participate in a housekeeping subpathway in combination with DNA Pol δ , whereas DNA ligase I appears to participate in a replication-associated subpathway in combination with DNA Pol ϵ . The identity of the DNA ligase(s), which is involved in the mismatch repair pathway that removes mismatched nucleotides generated as a consequence of errors by the replicative DNA polymerases, has not been definitively established.

Repair of DNA Strand Breaks

DNA ligase III α is a key player in a repair pathway for DNA single-strand breaks, which appears to be restricted to higher eukaryotes. In this pathway, the single-strand breaks are recognized by PARP-1. This activates PARP-1 polymerase activity, resulting in poly-ADP-ribosylation of PARP-1 itself and other proteins. The DNA ligase III α /XRCC1 complex is recruited to the damage site by interaction with poly (ADP-ribosylated) PARP-1. XRCC1 acts as a scaffold protein, recruiting proteins that process the single-strand break to generate a ligatable structure.

In all organisms, efficient repair of double-strand breaks (DSBs) is critical for the maintenance of genomic stability and viability. These lesions are particularly difficult to repair because both strands of the DNA duplex are broken. The repair pathways for DSBs can be divided into two groups based on whether they are dependent upon extensive DNA sequence homology or not. The major homology-dependent DSB repair pathway in mammalian somatic cells utilizes the intact sister chromatid as the template to guide repair and so this repair pathway is restricted to stages of the cell cycle when sister chromatids are available, that is, late S and G2 phases of the cell cycle. The homology-dependent repair of DSBs during meiosis is also a critical component in the generation of haploid gametes. The identity of the DNA ligase(s) that is involved in homology-dependent repair of DSBs has not been definitively established. The germ-cell specific isoform of DNA ligase III, DNA ligase III β , may be involved in meiotic recombination.

In the pathways for the repair of DSBs that are not directed by DNA sequence homology, the DNA ends are simply brought together by DNA end-bridging factors, processed, and then ligated by DNA ligase IV. A consequence of this mechanism is that unlike the faithful sister chromatid-mediated repair of DSBs, the nonhomologous end joining (NHEJ) pathways are error prone, generating genetic changes ranging from small insertions and deletions at the break site to chromosomal rearrangements. Surprisingly, the majority of DSBs are repaired by NHEJ in mammalian cells. Repair of DSBs by the major NHEJ pathway is completed by DNA ligase IV/XRCC4 complex, which is recruited to the DSBs by interactions with DNA-PK. Similar to the repair of DNA single-strand breaks, end processing occurs. As DNA ligase IV/XRCC4, in conjunction with XLF, has the ability to join mismatched termini, the DNA ligation event itself may cause changes in DNA sequence. DNA ligase IV/XRCC4 and the core components of the major NHEJ pathway are also involved in the error-prone repair of site-specific DSBs that are generated during immunoglobulin gene rearrangements in immune-system cells. There is emerging evidence for an alternative NHEJ pathway that is more error prone than the major NHEJ pathway, in particular generating chromosomal translocations. Although the components and mechanisms of alternative NHEJ are not well defined, it does appear to involve DNA ligase III α /XRCC1.

Mitochondrial DNA Metabolism

The DNA replication and repair pathways described above are all active on the nuclear genome. Studies by the Campbell group have shown that a mitochondrial form of DNA ligase III α that is generated by translation initiation (**Figure 2**) plays a key role in mitochondrial DNA metabolism. In yeast, which lacks a *LIG3* homolog, a mitochondrial version of Cdc9 (DNA ligase I) is generated by the same mechanism. As XRCC1 has not been detected in mammalian mitochondria, the mitochondrial functions of DNA ligase III α are independent of XRCC1, and instead involve interactions with mitochondrial proteins such as DNA Pol γ (**Table 1**). The participation of DNA ligase III in both nuclear and mitochondrial DNA metabolism will complicate molecular genetic analysis of the cellular functions of DNA ligase III and may explain the absence of studies describing *lig3* null cells or mice.

Concluding Remarks

Although the mechanism by which DNA ligases catalyze phosphodiester bond formation was described over four decades ago, new insights into DNA ligase functions have been obtained recently from studies of DNA ligase-deficient cell lines, the identification of DNA ligase protein partners, the reconstitution of DNA repair pathways *in vitro* and atomic resolution structures of DNA ligases at different stages of the ligation reaction. This new knowledge not only deepens our understanding of the network of pathways that protect against cancer formation by maintaining genome stability, but also provides a framework for the development of novel clinical strategies that target DNA metabolic enzymes, such as DNA ligases, in highly proliferative cancer cells with abnormalities in the DNA damage response.

See also: **Molecular Biology:** DNA Base Excision Repair Pathways; DNA Mismatch Repair in Bacteria; DNA Polymerase β , Eukaryotic; DNA Replication Fork, Eukaryotic; Eukaryotic DNA Polymerase α ; Eukaryotic DNA Polymerase δ .

Further Reading

- Ellenberger T and Tomkinson AE (2008) Eukaryotic DNA ligases: Structural and functional insights. *Annual Reviews in Biochemistry* 77: 313–338.
- Lehman IR (1974) DNA ligase: Structure, mechanism and function. *Science* 186: 790–797.
- Lieber MR (2010) NHEJ and its backup pathways in chromosomal translocations. *Nature Structural and Molecular Biology* 17: 393–395.
- Martin IV and MacNeill SA (2002) ATP-dependent DNA ligases. *Genome Biology* 3: 3005.1–3005.7.
- Pitcher RS, Brissett NC, and Doherty AJ (2007) Nonhomologous end-joining in bacteria: A microbial perspective. *Annual Review in Microbiology* 61: 259–282.
- Shuman S (2009) DNA ligases: Progress and prospects. *Journal of Biological Chemistry* 284: 17365–17369.
- Shuman S and Glickman MS (2007) Bacterial DNA repair by non-homologous end joining. *Nature Reviews in Microbiology* 5: 852–861.
- Tomkinson AE, Vijayakumar S, Pascal JM, and Ellenberger T (2006) DNA ligases: Structure, reaction mechanism, and function. *Chemical Reviews* 106: 687–699.

DNA Ligases: Structures

J M Pascal, Thomas Jefferson University, Philadelphia, PA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Adenosine triphosphate (ATP) An adenosine-containing nucleotide that is used to provide energy for biochemical processes through hydrolysis of high-energy phosphate bonds.

Covalent nucleotidyl transferase A member of a family of enzymes that reacts with nucleotides and transfers the

nucleoside moiety to an acceptor molecule through the formation of a covalent nucleoside monophosphate (NMP)-enzyme intermediate.

Nicotinamide adenine dinucleotide (NAD⁺) A coenzyme present in all cells that accepts electrons to assist in metabolic processes; provides energy for biochemical processes through hydrolysis of high-energy bonds.

Introduction

The structural integrity of the genome is continually challenged by normal cellular activities that alter DNA structure, such as the processes of DNA replication, repair, and recombination. The stability of the genome is further challenged by exogenous sources of DNA damage, such as ultraviolet radiation and chemical carcinogens. DNA ligases make an essential contribution to maintaining genome integrity by correcting breaks, or nicks, in the phosphate backbone structure of duplex DNA. The vital importance of DNA ligases to the health of an organism is underscored by the fact that deficiencies in DNA ligase activity are associated with genetic instability and predisposition to cancer. The utility of joining fragments of DNA has also made DNA ligase enzymes a vital component of molecular biology techniques, such as gene cloning.

DNA ligases carry out a three-step reaction that results in the joining of two DNA strands: one with a 3'-hydroxyl terminus (3'-OH), and one with a 5'-phosphorylated terminus (5'-PO₄). The reaction mechanism requires a high-energy nucleotide cofactor, adenosine triphosphate (ATP) or nicotinamide adenine dinucleotide (NAD⁺), and DNA ligases are classified as being ATP-dependent or NAD⁺-dependent enzymes. The reaction mechanism also requires divalent cations such as Mg²⁺ or Mn²⁺. In step 1 of the reaction (Figure 1), the epsilon nitrogen atom of a conserved active site lysine side chain attacks the phosphorus atom of the adenosine monophosphate (AMP) group of ATP or NAD⁺, forming an N-P bond and creating the covalent ligase-AMP reaction intermediate. The nucleotide cofactor is hydrolyzed during the first step of the reaction, releasing pyrophosphate (PP_i) for ATP-dependent DNA ligases, or nicotinamide mononucleotide (NMN) for NAD⁺-dependent ligases. In step 2 of the reaction, the AMP group is transferred from DNA ligase to the 5'-PO₄ end of the DNA break, creating the covalent AMP-DNA reaction intermediate. In step 3 of the ligation reaction, the 3'-OH end of the DNA break is joined to the 5'-PO₄ end of the apposing DNA strand, liberating AMP and releasing the DNA substrate with a continuous phosphodiester backbone structure.

Domain Architecture of DNA Ligases

DNA ligases employ multidomain architectures to carry out the three-step ligation reaction mechanism. There are significant variations in the number and type of structural domains that compose the different DNA ligases (Figure 2). However, there are common structural domains found in all DNA ligases. The centerpiece of the DNA ligase structure is the nucleotidyltransferase (NTase) domain. The overall structural fold of the NTase domain is composed of α -helices surrounding a central arrangement of β -strands that form the nucleotide-binding pocket. The NTase domain houses six highly conserved amino acid sequence motifs that contribute key residues to the active site and form the nucleotide-binding pocket (Figure 2). Structures of DNA ligases in complex with nucleotide cofactors have demonstrated how the conserved sequence motifs position the cofactor and select for ATP or NAD over other nucleotides.

DNA ligases have an oligonucleotide/oligosaccharide-binding (OB) domain with a five-stranded β -barrel structure that extends from the C-terminus of the NTase domain (Figure 2). The OB domain is named for the OB properties of several domains that share this fold, although, in general, OB domains found in other types of proteins can also mediate other functions, such as protein-protein interactions. The basic NTase-OB domain structure is found in all DNA ligases. For the simplest DNA ligases, such as those from bacteriophage T7 and *Chlorella* virus, a two-domain NTase-OB architecture is sufficient to carry out all steps of DNA ligation. More complex DNA ligases decorate the basic NTase-OB domain structure with N- and C-terminal appendages that (1) assist the overall ligation reaction and/or (2) provide special DNA-binding properties that dictate substrate specificity, and/or (3) bestow special protein interaction properties that designate ligase activity to a particular cellular pathway of DNA repair. Several examples of these specialized domains are described in the following sections that detail how the multidomain structures carry out the three-step ligation reaction.

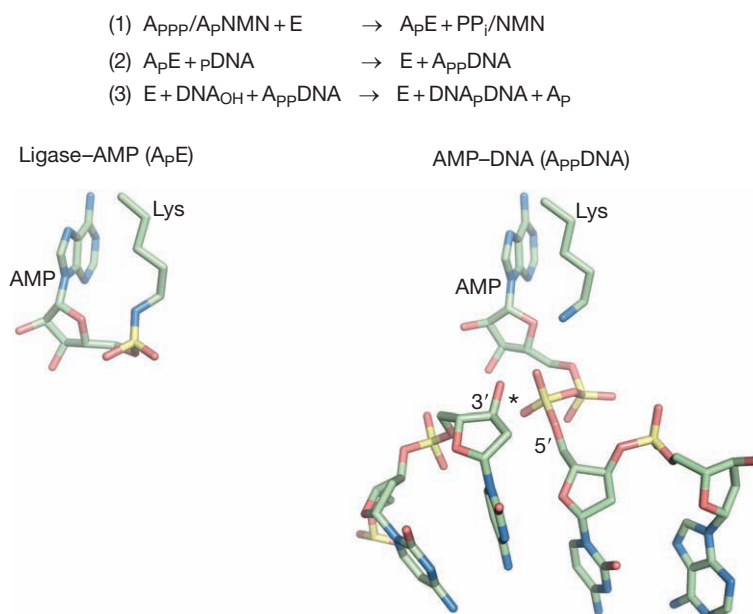


Figure 1 Three steps of the DNA ligation reaction and covalent intermediates. During step 1, the AMP (A_p) group of the nucleotide cofactor, ATP (A_{ppp}) or NAD^+ (A_pNMN), is attached to a conserved lysine of the ligase enzyme (E), forming the ligase-AMP intermediate (A_pE). During step 2, the AMP group is transferred to the DNA substrate containing a 5'- PO_4 ($pDNA$), forming the AMP-DNA intermediate ($A_{pp}DNA$). During step 3, the 3'-OH of an apposing nucleic acid strand (DNA_{OH}) attacks the 5'- PO_4 , joining the DNA ends and releasing AMP. The two covalent intermediates of the ligation reaction are drawn as sticks. The AMP-DNA intermediate is shown in the context of several nucleotides from the broken DNA strand; the complementary, intact DNA strand is not shown. The DNA break between the 3'-OH and the 5'- PO_4 is highlighted with an asterisk (*). Adapted with permission from Pascal JM (2008) DNA and RNA ligases: Structural variations and shared mechanisms. *Current Opinion in Structural Biology* 18: 96–105.

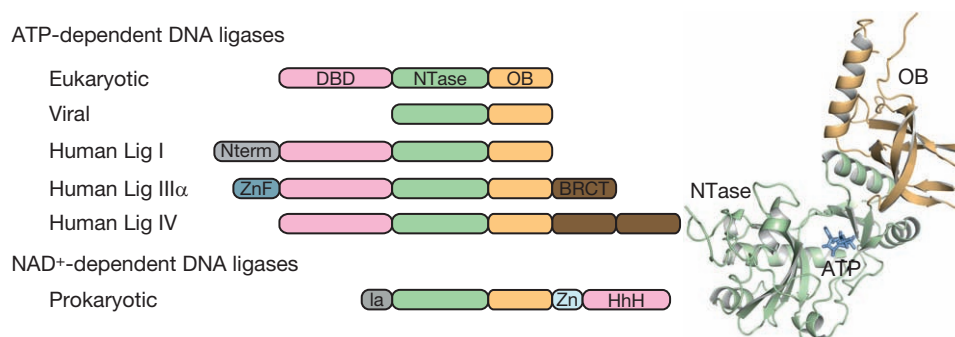


Figure 2 Multidomain, modular architectures of DNA ligases. Representative domain architectures of ATP- and NAD^+ -dependent DNA ligases. The nucleotidyltransferase (NTase) domain is the centerpiece of all DNA ligases and houses the broadly conserved residues that catalyze the three-step end-joining reaction. The OB domain extends from the C-terminal end of the NTase domain and contributes key residues that assist in DNA ligation. A diverse collection of N- and C-terminal accessory domains extends from the NTase-OB domains to promote the ligation reaction, or to provide special substrate-binding properties: DNA-binding domain (DBD), helix-hairpin-helix (HhH) domain, domain Ia (Ia), zinc-binding domain (Zn), PARP-like zinc finger (ZnF), and BRCA-1 C-terminal (BRCT) domain. The two-domain crystal structure of bacteriophage T7 DNA ligase illustrates the fold of the NTase and OB domains found in all DNA ligases. ATP bound to the nucleotide-binding pocket of the NTase domain is drawn as sticks (blue). Adapted with permission from Pascal JM (2008) DNA and RNA ligases: Structural variations and shared mechanisms. *Current Opinion in Structural Biology* 18: 96–105.

Mechanisms of Ligase-AMP Formation

In step 1 of the ligase reaction, the covalent ligase-AMP intermediate is formed. ATP-dependent DNA ligases use conserved residues of the OB domain to assist in formation of the ligase-AMP intermediate (Figure 3). The OB domain of ATP-dependent ligases closes over the active site of the NTase domain to contribute residues that position the

α -phosphate of ATP for attack by the active site lysine. Interestingly, RNA-capping enzymes use a similar mechanism and a two-domain NTase-OB structure to form a covalent-capping enzyme-guanosine monophosphate (GMP) intermediate, prior to transferring GMP to the 5' end of messenger RNA. DNA ligases, RNA ligases, and RNA-capping enzymes constitute the covalent nucleotidyltransferase superfamily of enzymes.

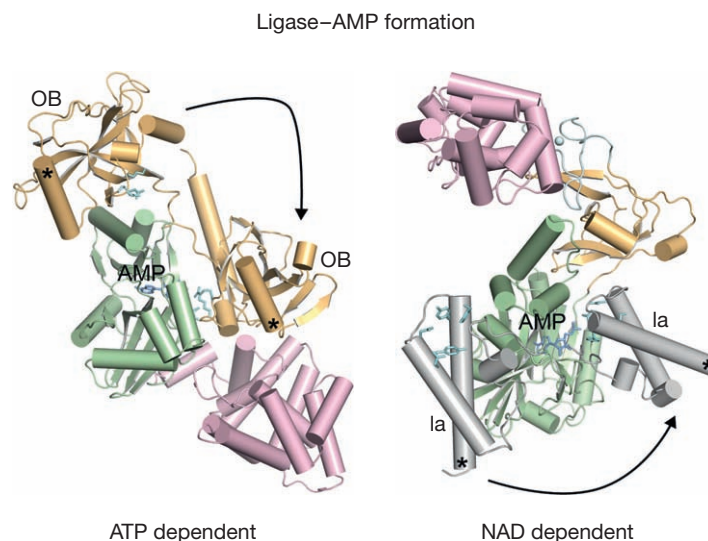


Figure 3 Formation of the ligase–AMP intermediate. For ATP-dependent DNA ligases, the OB domain contributes key residues that stimulate the step 1 reaction, ligase–AMP formation. The open, extended conformation is modeled based on the *S. solfataricus* DNA ligase structure. The closed conformation, with the OB domain positioned next to the cofactor-binding site of the NTase domain, is based on the crystal structure of *P. furiosus* DNA ligase. Conserved residues that assist in step 1 are drawn as sticks (cyan). The transition of the OB domain from the open to closed conformation is highlighted by the arrow, and can be followed by noting the relative position of the helix of the OB domain marked with an asterisk (*), and the residues involve in step 1 chemistry (cyan). The domains are colored as in **Figure 2**: DBD (pink), NTase (green), and OB (yellow). For NAD⁺-dependent DNA ligases, domain Ia contributes key residues that stimulate the step 1 reaction. The open, extended conformation of domains is based on the structure of *T. filiformis* DNA ligase, which includes all ligase domains. The position of domain Ia during the step 1 reaction is based on the crystal structure of *Enterococcus faecalis* DNA ligase. Key residues from domain Ia that position NAD⁺ in the NTase active site for the step 1 reaction are drawn as sticks (cyan). The transition of domain Ia from the open to closed conformation is highlighted by the arrow and can be followed by noting the relative position of the helix of domain Ia marked with an asterisk (*), and the residues involve in step 1 chemistry (cyan). The domains are colored as in **Figure 2**: Ia (gray), NTase (green), OB (yellow), Zn (cyan), and HhH (pink).

In contrast to ATP-dependent DNA ligases, NAD⁺-dependent DNA ligases do not use their OB domain to assist in formation of the ligase–AMP intermediate. NAD⁺-dependent DNA ligases use a different mechanism that requires residues from an N-terminal extension of the NTase domain, called domain Ia (**Figure 3**). During ligase–AMP formation, domain Ia closes over the active site of the NTase domain to bind the NMN moiety of NAD⁺ and position the nucleotide cofactor for attack by the active site lysine. Thus, both NAD⁺- and ATP-dependent DNA ligases use accessory domains to promote ligase–AMP formation, but they do so using entirely different domains and mechanisms. After formation of the ligase–AMP reaction intermediate, domain Ia and the OB domain pivot away from their respective NTase active sites to allow access to incoming DNA substrates; therefore, large-scale conformational changes accompany the binding and hydrolysis of NAD⁺ or ATP.

Mechanism of AMP–DNA Formation and DNA End Joining

Crystal structures of several DNA ligases bound to nucleic acid have provided instructive views of ligase enzymes engaging their DNA substrates and nucleotide cofactors at different stages of the ligation reaction. The structures illustrate how breaks in DNA are recognized, how AMP is used to orchestrate ligation chemistry, how novel accessory domains are utilized

in catalysis and substrate recognition, and how the shape of the DNA substrate plays an important role in the ligation reaction. The DNA ligase structures in complex with DNA indicate that large-scale changes in domain arrangements accompany nucleic acid binding. Furthermore, the structures have revealed that DNA ligases form protein clamps that encircle their DNA substrates; however, the ligase clamps are formed using distinct topological arrangements of their domains (see the following section).

DNA ligases were trapped in complex with DNA for crystallographic analysis by stalling DNA ligation at different steps along the reaction pathway. The catalytic fragment of human LIG1 was crystallized in complex with the AMP–DNA reaction intermediate. A 2'-H, 3'-H dideoxy nucleotide at the 3' end of the DNA break stalled the ligation reaction after AMP–DNA formation – the absent 3'-OH prevented the final DNA end-joining reaction from occurring. The structure of *Escherichia coli* DNA ligase A (LigA) was also determined in complex with the AMP–DNA reaction intermediate. A 2'-H, 3'-OH deoxynucleotide at the 3' end of the DNA break made the substrate competent for ligation; however, divalent metals were excluded from crystal solutions to prevent ligation of DNA ends. The ligase–AMP intermediate of *Chlorella* ligase was crystallized bound to a DNA break with 3'-OH and 5'-PO₄ ends, showing the arrangement of a ligase active site before AMP is transferred to DNA. Divalent metals were excluded from crystal solutions, and this stalled the ligation reaction prior to step 2. After the addition of divalent metal to the *Chlorella* ligase

crystals, the broken DNA ends were joined together, highlighting that the *Chlorella* ligase–DNA complex is truly paused along the ligation reaction pathway. The ligase–AMP intermediate of human DNA ligase III (LIG3) was also crystallized in complex with DNA; the DNA break was not competent for ligation due to a 2'-H, 3'-H dideoxy nucleotide at the 3' end of the DNA break.

Collectively, the DNA ligase structures in complex with DNA demonstrate that the NTase and OB domains interact to form a continuous protein surface that engages the DNA minor groove at the break in the DNA substrate (Figures 4(a) and 4(b)). The NTase–OB interaction with DNA creates a localized change in substrate conformation at the DNA break, imposing an RNA-like A-form configuration on the substrate, resulting in a widened and more shallow minor groove compared to B-form DNA (Figures 4(c) and 4(d)). The distortion of the substrate partially unwinds the DNA duplex and alters the trajectory of the helical axis at the break in the DNA. The local alteration of the substrate conformation extrudes the DNA ends from the DNA duplex and positions them in the active site for the joining reaction. Each of the DNA ligase structures in complex with DNA demonstrate a similar arrangement of their NTase and OB domains on the DNA break, although there are variations in the amino acid residues and structural elements that mediate the interactions. However, in each of the DNA ligase structures bound to DNA, the NTase

and OB domain interactions with the minor groove conspire to impose an RNA-like A-form conformation on the substrate at the DNA break.

The NTase domain primarily interacts with the substrate on the 3' side of the DNA break, making contacts to the phosphodiester backbone of the broken and intact DNA strands, and placing conserved amino acid residues in the minor groove near the DNA break. These residues position the 3'-OH end of the substrate within the active site for the ligation reaction to occur. The NTase domain positions the 5' side of the break primarily through contacts with the AMP group, which is covalently linked to the 5'-PO₄ of DNA. The OB domain primarily engages the 5' side of the break, spanning the minor groove and making extensive contacts with the phosphodiester backbone of both the intact strand and the broken DNA strand. In collaboration with the NTase domain, the OB domain contributes residues to the minor groove around the DNA break to assist in creating the local A-form distortion that positions the DNA ends in the active site. The ligase–AMP intermediates of *Chlorella* DNA ligase and human LIG3 were crystallized in complex with a DNA break containing the required 5'-PO₄ end. In these structures, the NTase and OB domains engage the DNA substrate in the same manner shown for LIG1 and LigA (Figures 4(a) and 4(b)), indicating that the NTase–OB configuration is consistent during the step 2 and step 3 reactions. Notably, the DNA-bound

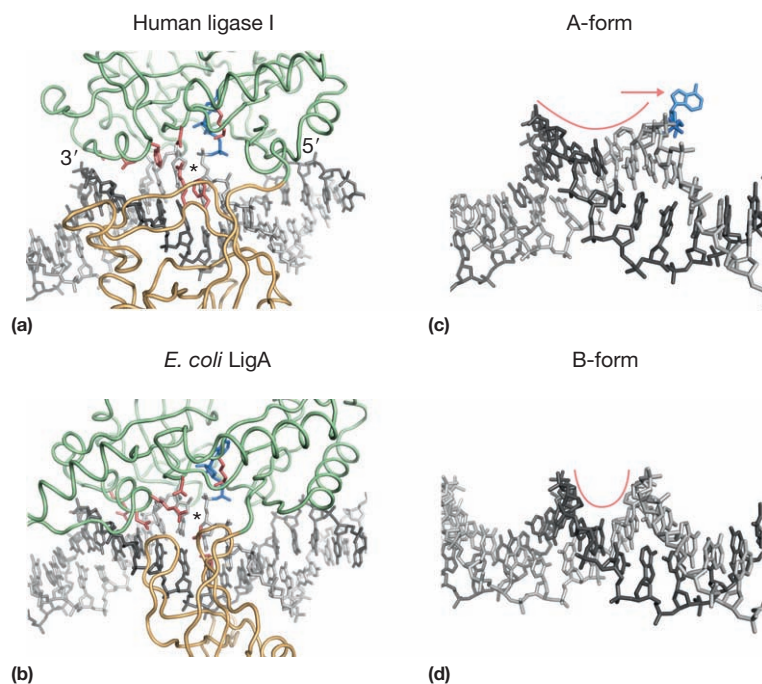


Figure 4 The NTase and OB domains engage the minor groove at the DNA break. A detailed view of NTase–OB interactions with the DNA minor groove for (a) human Ligase I and (b) *E. coli* Lig A. The NTase domain (green) and OB domain (yellow) are drawn as tubes. The AMP cofactor (blue), the broken DNA strand (light gray), the intact DNA strand (dark gray), residues involved in minor groove interactions (red), and the active site lysine (red) are drawn as sticks. The break in the DNA substrate is highlighted with an asterisk (*). The 3' and 5' sides of the DNA break are labeled for LIG1 in (a), and (b) has the same orientation. (c) DNA ligase binding to the minor groove distorts the DNA substrate near the break in the phosphodiester backbone. The minor groove is made wider and less deep, with dimensions more similar to A-form duplex. The substrate distortion positions the DNA ends toward the ligase active site (red arrow). The overall shape and dimensions for ideal B-form duplex DNA is shown in (d) for comparison. Adapted with permission from Pascal JM (2008) DNA and RNA ligases: Structural variations and shared mechanisms. *Current Opinion in Structural Biology* 18: 96–105.

orientation of the OB domain is dramatically different from the orientation of the OB domain during step 1. The transition between the two active conformations of the OB domain requires a 180° swiveling of the OB domain around its linker to the NTase domain.

DNA Ligase Protein Clamps

A remarkable feature of the DNA ligase structures in complex with DNA is that the enzymes form protein clamps that fully encircle the substrate DNA (Figure 5). Even more remarkable is that the protein clamps are formed using different strategies for different DNA ligases, as further described below. There are two key reasons why this might be important for DNA ligase function. First, engulfing the DNA ends within the ligase domain structure fully protects the DNA break from exposure to other cellular factors that might act on the DNA

ends, such as nucleases. Second, the protein clamp allows a more extensive interaction surface with the DNA substrate. This could be important for DNA ligases to properly engage DNA substrates, which are quite large relative to the size of the ligase enzyme.

As mentioned previously, ATP-dependent eukaryotic and archaeal DNA ligases have multidomain architectures. In addition to the NTase–OB catalytic core, they have an N-terminal α -helical extension termed the DNA-binding domain (DBD; Figures 5(a) and 5(b)). The DBD of human LIG1 and LIG3 stimulates the ligation activity of the NTase and OB domains, indicating that the DBD acts cooperatively with the NTase–OB catalytic core by supporting their active conformation on DNA and the distorted shape of the substrate DNA. As shown for human LIG1, the DBD binds across the minor groove of the DNA substrate at two positions, equally spaced around the break in the DNA (Figure 5(b)). The α -helical fold of the DBD has pseudo-twofold symmetry, providing similar

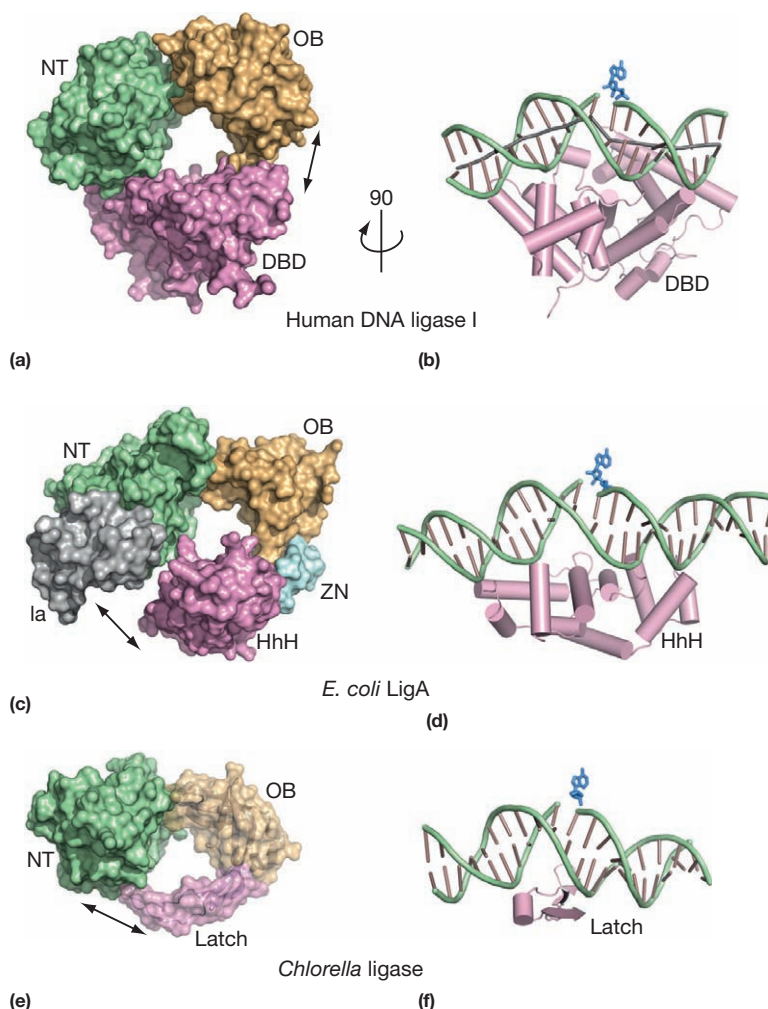


Figure 5 DNA ligase protein clamps. Surface representations of DNA ligase protein clamps: (a) Human LIG1, (c) *E. coli* LigA, and (e) *Chlorella* ligase. Double-headed arrows indicate breakpoints in the ligase clamps. Domains are labeled, and colored as in Figure 2. For clarity, DNA substrates are not shown in (a), (c), and (e). Cartoon representations of DNA substrates bound to (b) the DBD of human LIG1, (d) the HhH of *E. coli* LigA, and (f) the latch of *Chlorella* ligase. The helical axis of the DNA substrate is shown for LIG1 in (a), highlighting the local distortion that occurs at the DNA break. The orientations of (b), (d), and (e) are rotated roughly 90° relative to the orientations in (a), (c), and (e), respectively. Adapted with permission from Pascal JM (2008) DNA and RNA ligases: Structural variations and shared mechanisms. *Current Opinion in Structural Biology* 18: 96–105.

protein–DNA interaction motifs on both sides of the DNA break. The DBD physically interacts with the NTase and OB domains, establishing a ligase protein clamp that fully encircles the DNA substrate. The DBD is required for the ligation activity of several eukaryotic ligases, and the DBD residues that mediate interactions with DNA substrate are conserved among other human DNA ligases. It is, therefore, expected that all eukaryotic ligases encircle DNA when carrying out the end-joining reactions required for the replication and repair of DNA.

The eukaryotic DBD may have additional functional roles. DNA ligases from *Pyrococcus furiosus* and *Sulfolobus solfataricus* interact with the proliferating cell nuclear antigen (PCNA) DNA sliding clamps of these organisms using residues located in their respective DBDs. For these archaeal DNA ligases, the DBD interaction with PCNA stimulates ligation activity. The DBD interaction with PCNA is conserved in humans, where the DBD of LIG1 interacts with human PCNA. The function of this interaction has not been established, but it may play a role in coordinating LIG1 activity during replication and repair.

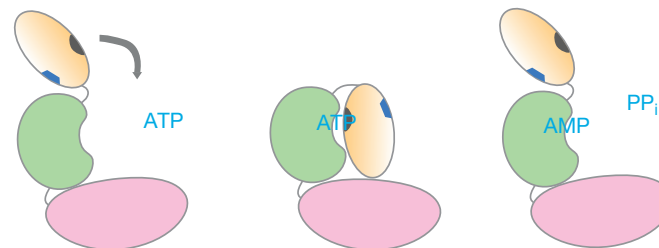
Mammalian ligases (LIG1, LIG3, and LIG4) have extra domains that extend from the N- and C-termini of their shared catalytic DBD–NTase–OB architecture, and these appended domains provide unique DNA-binding or protein–protein interaction properties that endow specificity for different DNA replication and repair pathways. Understanding the

structure and function of these distinguishing features of mammalian ligases will play an important role in determining their mechanistic roles in cellular DNA metabolism.

Bacterial DNA ligases are almost exclusively NAD^+ -dependent enzymes. They have multidomain architectures that are quite different from the eukaryotic ATP-dependent DNA ligases, although the NTase and OB domains still form the core of these enzymes. In addition to the unique N-terminal domain Ia that stimulates ligase–AMP formation, there are three C-terminal domains that extend from the OB domain: a small zinc-binding (Zn) domain, a helix–hairpin–helix (HhH) domain, and a BRCT domain (BRCA1 gene C-terminus). Biochemical analysis of *Thermus filiformis* ligase indicates that the BRCT domain is not important for the ligation reaction *in vitro*, and the *in vivo* function of the BRCT domain is not clear.

The structure of *E. coli* LigA in complex with DNA demonstrates the arrangement of bacterial ligase domains when engaging substrate nucleic acid, revealing a novel ligase protein clamp that encircles the DNA substrate (Figure 5(c)). On the side of the DNA substrate opposite to the NTase–OB domains, the HhH of LigA binds to DNA with approximate twofold symmetry, gripping the DNA across the minor groove at two positions spaced symmetrically about the break (Figure 5(d)). The HhH domain of LigA extends from the OB and Zn domains and contacts the NTase domain, creating a protein clamp that envelops the DNA substrate. The DNA-bound

Step 1, ligase–AMP formation



Steps 2 and 3, AMP–DNA formation and ligation

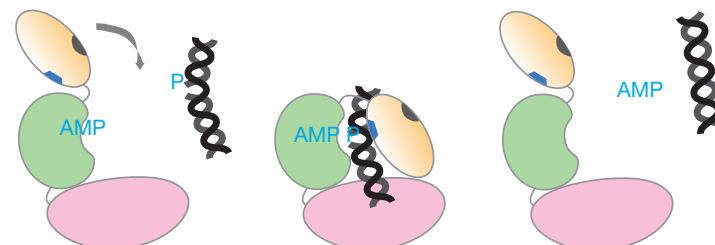


Figure 6 Conformational rearrangements during the DNA ligation reaction. A schematic model for DNA ligase conformational changes during the DNA end-joining reaction based on crystal structures of ATP-dependent DNA ligases. (Top) To form the ligase–AMP reaction intermediate during step 1, the OB domain (yellow) moves toward the NTase domain (green) to place conserved residues (brown half-oval) near the ATP-binding site. Inorganic phosphate (PP_i) is released after nucleotide hydrolysis, and ligase returns to the open conformation in order to provide access to substrate DNA. (Bottom) Upon binding to nicked DNA, the OB domain pivots and moves toward the NTase domain and the bound DNA substrate. This conformation of the OB domain places conserved residues (blue trapezoid) near the break in the DNA and contributes to AMP–DNA formation and ligation of DNA ends. AMP and DNA with a continuous backbone structure are released. The configuration of domains in the middle of the top and bottom panels represents the two active conformations of DNA ligase. The transition between the two active conformations of the OB domain requires a 180° swiveling of the OB domain around its linker to the NTase domain. The DBD domain (red) contributes to DNA-binding affinity, and supports the arrangement of the NTase–OB domains and the distorted DNA substrate conformation.

conformation of the LigA domains is remarkably different from the arrangement of domains observed in the ligase–AMP intermediate of *T. filiformis* ligase in the absence of DNA, demonstrating that large-scale alterations in domain arrangements occur upon binding to DNA breaks. The Zn domain bridges between the OB and HhH domains of LigA, most likely providing structural support for the modular ligase architecture as it reorganizes during the steps of DNA ligation.

The HhH domain of LigA and the DBD of LIG1 appear to play analogous functional roles, despite having entirely different protein folds. They both occupy similar positions on the DNA substrate in their respective protein clamps, and they both engage DNA in similar manners, primarily using polar atoms of the protein main chain to brace the phosphodiester DNA backbone. Similarly, both the DBD of LIG1 and the HhH domain of *T. filiformis* provide the major DNA-binding activity for these ligases. The distinct topology and connectivity of the LIG1 and LigA domain architectures indicate that the eukaryotic and bacterial protein clamps are independent yet strikingly similar solutions to encircling DNA breaks.

Chlorella virus PBCV-1 encodes a 298-residue DNA ligase with a two-domain NTase–OB architecture that lacks the flanking accessory domains required by the cellular eukaryotic and bacterial ligases. *Chlorella* ligase displays yet another distinct protein clamp that encircles the DNA substrate (Figure 5(e)). The OB domain of *Chlorella* ligase has a 30-residue surface loop, designated the ‘latch’, that forms extensive contacts with the phosphodiester backbone of the major groove opposite the break in the DNA (Figure 5(f)). The distal end of the latch contacts the NTase domain to complete a protein clamp that encloses the substrate DNA. The latch was disordered in an unliganded crystal structure of *Chlorella* ligase, and protease sensitive in solution in the absence of DNA, indicating that the latch only becomes ordered upon binding to DNA. The latch is substantially smaller than the DBD and HhH domains, yet it appears to play a similar role in providing the major contribution to DNA binding.

The collection of DNA ligase structures bound to nucleotide cofactors and to DNA substrates have illustrated how the multidomain architectures of DNA ligases carry out the three-step ligation reaction. A stunning level of conformational plasticity is required to perform the steps of DNA ligation, as illustrated in the cartoon depiction of the ligase reaction based on several structures of ATP-dependent eukaryotic DNA ligases (Figure 6). The variety of DNA ligase domain architectures has converged on a common principle of encircling DNA substrates to assist in manipulating the broken ends of DNA for the joining reaction. The structures have also detailed the dramatic mechanistic differences between the bacterial NAD⁺-dependent and

eukaryotic ATP-dependent ligases during the step 1 reaction, as well as key mechanistic differences between the eukaryotic and the viral ATP-dependent DNA ligases. Due to the fundamental requirement for DNA ligase activity in all organisms, and the important role of ligases in maintaining genome integrity, selective inhibitors of ligase activity are potential antimicrobial agents and novel tools for fighting cancer. The specific molecular level insights into the ligation mechanism derived from structural analysis provide new avenues for the design of selective inhibitors of DNA ligase activity.

See also: [Molecular Biology: DNA Ligases: Mechanism and Functions.](#)

Further Reading

- Cotner-Gohara E, Kim IK, Hammel M, et al. (2010) Human DNA ligase III recognizes DNA ends by dynamic switching between two DNA-bound states. *Biochemistry* 49(29): 6165–6176.
- Ellenberger T and Tomkinson AE (2008) Eukaryotic DNA ligases: Structural and functional insights. *Annual Review of Biochemistry* 77: 313–338.
- Lehman IR (1974) DNA ligase: Structure, mechanism, and function. *Science* 186(4166): 790–797.
- Nair PA, Nandakumar J, Smith P, et al. (2007) Structural basis for nick recognition by a minimal pluripotent DNA ligase. *Nature Structural and Molecular Biology* 14(8): 770–778.
- Nandakumar J, Nair PA, and Shuman S (2007) Last stop on the road to repair: Structure of *E. coli* DNA ligase bound to nicked DNA-adenylate. *Molecular Cell* 26(2): 257–271.
- Nishida H, Kiyonari S, Ishino Y, and Morikawa K (2006) The closed structure of an archaeal DNA ligase from *Pyrococcus furiosus*. *Journal of Molecular Biology* 360(5): 956–967.
- Pascal JM, O'Brien PJ, Tomkinson AE, and Ellenberger T (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature* 432(7016): 473–478.
- Pascal JM, Tsodikov OV, Hura G, et al. (2006) A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. *Molecular Cell* 24(2): 279–291.
- Shuman S and Lima CD (2004) The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases. *Current Opinion in Structural Biology* 14(6): 757–764.
- Subramanya HS, Doherty AJ, Ashford SR, and Wigley DB (1996) Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. *Cell* 85(4): 607–615.
- Tomkinson AE, Vijayakumar S, Pascal JM, and Ellenberger T (2006) DNA ligases: Structure, reaction mechanism, and function. *Chemical Reviews* 106(2): 687–699.

Relevant Website

<http://www.rcsb.org> – Protein Data Bank; Accession Codes: Human Ligase I (1X9N), Human Ligase III (3L2P), *Chlorella* Virus DNA Ligase (2Q2T), *E. coli* Lig A (2OWO), *S. solfataricus* DNA Ligase (2HIV), *P. furiosus* DNA Ligase (2CFM), *E. faecalis* DNA Ligase (1TA8), *T. filiformis* DNA Ligase (1V9P), and Bacteriophage T7 DNA Ligase (1A0I).

DNA Methyltransferases: Eubacterial GATC

M G Marinus, University of Massachusetts Medical School, Worcester, MA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

DNA methyltransferase Enzyme that transfers a methyl group to a base in DNA.

Epigenetic Pertaining to phenotypic, but not genotypic, change.

Exonuclease Enzyme that digests DNA strand from an end.

Hemimethylated Describing DNA in which one strand contains methylated residues but the complementary strand does not.

Promoter Sequence at which RNA polymerase initiates transcription.

Properties of Dam Methyltransferase

The Dam methyltransferase is encoded by the DNA adenine methyltransferase (*dam*) gene of *Escherichia coli*, which has a complex mechanism of gene regulation involving multiple promoters and terminators. The major promoter for transcription of the gene is regulated by growth rate of the cells; that is, the faster the growth rate, the greater is the level of initiation of transcription.

During chromosome replication, the replication fork is followed by a region of hemimethylated DNA which comprises new unmethylated DNA and the complementary methylated template strand. After 1 min, on average, in slow-growing cells the newly synthesized DNA is methylated so that both strands are fully methylated. This delay in methylation is due to the concentration of the Dam methyltransferase; it is present at 130 molecules per cell. Increasing the cellular concentration of the enzyme proportionately decreases the amount of hemimethylated DNA.

The enzyme is a single polypeptide chain of 278 amino acids with a molecular weight of 32 kDa and exists in solution as a monomer. The enzyme has a turnover number of 19 methyl transfers per minute and an apparent K_m of 3.6 nM for DNA; double-stranded DNA is a better methyl acceptor than denatured DNA; and there is one methyl transfer per site per binding event even if the substrate DNA is fully unmethylated. In contrast to these findings, there is also evidence that the Dam methyltransferase binds the template and slides processively along the DNA, methylating about 50 sites before dissociating. The differences in distributive and processive activity may be dependent on the length, sequence, and topology of DNA. The atomic structure of Dam complexed with DNA has been solved and shows both nonspecific backbone contacts and specific contacts with the glutamyl-tRNA (Gln) amidotransferase, subunit C homolog (GATC) bases. Importantly, the aromatic ring of Y119 intercalates into the DNA between GA and TC, thereby flipping the adenine into the enzyme's active site. The unpaired complementary T residue can adopt an intrahelical or extrahelical position. The contacts of the enzyme with GATC bases, and flanking phosphate contacts by conserved residues, position Dam on the DNA duplex.

Properties and Uses of *dam* Mutant Strains

The commonly used *dam* mutant strains have neither detectable residual methyltransferase activity nor detectable methylation at GATC sequences. *Dam* mutants are most often used to launder DNA molecules that have sites that are resistant to digestion with specific restriction endonucleases. These include AlwI, BcgI, BclI, BsaBI, BspDI, BspEI, BspHI, ClaI, HphI, NruI, TaqI, and XbaI. Resistance occurs because these enzymes have recognition sequences that overlap with the GATC tetranucleotide and almost all of these are methylated in wild-type *E. coli*. After transfer through the *dam* mutant strain, there is no methylation at GATCs and DNA molecules are now sensitive to digestion with the restriction endonucleases listed above.

Dam mutants have also been useful to define the biological roles of methyl groups on adenine in DNA. The mutant strain shows a large number of phenotypic differences compared to wild type suggesting multiple functions. All the known mutant phenotypes in *E. coli* can be explained by the involvement of GATC methylation in mismatch repair, regulation of gene expression, and initiation of chromosome replication.

Dam-Directed Mismatch Repair

During chromosome replication, errors are made at low frequency by the replicative polymerase to form base mismatches in newly synthesized DNA. Such errors need to be removed because they are potentially mutagenic. *E. coli* and most other organisms, including humans, have a highly conserved repair system that removes such mismatches in newly synthesized DNA. The Dam- (or methyl-)directed mismatch repair system in *E. coli* uses only hemimethylated DNA as a substrate, thereby restricting its activity to the region of the chromosome immediately trailing the replication fork. The repair system is outlined in [Figure 1](#). Briefly, base mismatches in hemimethylated DNA are recognized by the MutS protein followed by the formation of a ternary complex with two additional proteins, MutL and MutH. MutH is an endonuclease active on hemimethylated GATC sites but only when complexed with the other Mut proteins on DNA. Following DNA incision, the UvrD helicase unwinds DNA in

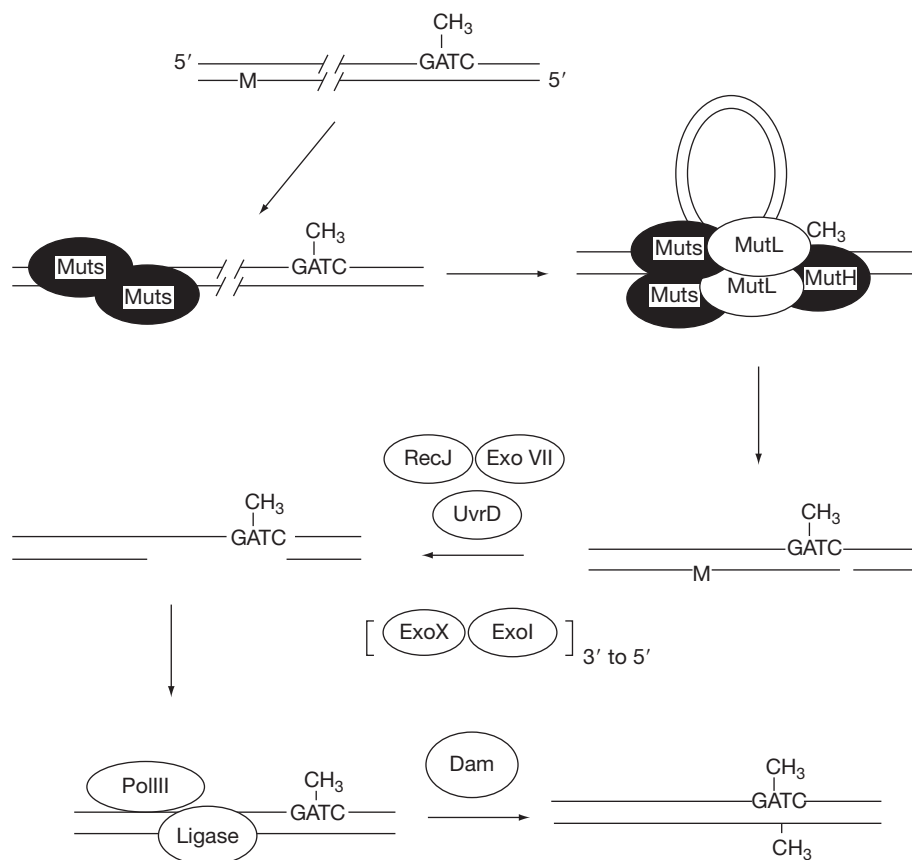


Figure 1 DNA mismatch repair. A base mismatch (M) in the newly synthesized unmethylated DNA strand is some distance from the nearest GATC sequence. MutS protein binds the mismatch and a complex with MutL and MutH is formed leading to cleavage of the unmethylated strand. The UvrD helicase unwinds the unmethylated strand which is digested by an exonuclease. Excision can occur in either the 3'–5' direction (ExoI and ExoX) or the reverse (RecJ and ExoVII), depending on the orientation of the mismatch relative to the GATC sequence. DNA polymerase III holoenzyme synthesizes a new strand and the resultant nick is ligated by DNA ligase. The GATC sequence is eventually methylated by Dam methyltransferase.

either the 5'–3' direction or the reverse. The directionality is dependent on the location of the mismatch relative to the nearest GATC sequence. The unwound nicked DNA strand is digested by one or more exonucleases, followed by resynthesis using the replicative polymerase. DNA ligation of the nick at the end of replication completes the process of repair and the hemimethylated DNA is eventually methylated.

Three lines of genetic evidence indicate the importance of DNA methylation in the process. First, overproduction of Dam methyltransferase inhibits the repair process by reducing the amount of hemimethylated DNA. Under these conditions, mismatches are not repaired and lead to an increase in mutation frequency. Second, in a *dam* mutant the repair system is active but cannot discriminate the template strand from the daughter strand. Consequently, the repair system not only removes mutations from the newly synthesized strand but also introduces mutations into the parental strand, yielding an increased mutation frequency. Third, DNA duplexes containing a mismatch can be constructed with no methylation, methylation on both strands, or on only one of the two strands. When introduced into cells the fully methylated DNA is not repaired, hemimethylated DNA is repaired using

the methylated strand as template, and the unmethylated duplex is repaired using either strand as template.

In a *dam* mutant, mismatch repair produces detectable nicks or gaps in DNA but it is not known if the nicks or gaps occur exclusively behind the replication fork as a result of replication errors or in any part of the chromosome as a result of mismatches formed spontaneously. The nicks or gaps are converted to double-strand breaks either by a replication fork encountering an unrepaired gap or by an activated MutH cleavage on complementary strands at the same GATC sequence or by a combination of these (Figure 2). It is not yet known which of these mechanisms (or both) generates double-strand breaks *in vivo*; however, *in vitro* evidence supports the MutH cleavage model.

Mismatch repair in *dam* mutants is also the basis for the phenomenon of drug tolerance. The mutants are more sensitive than wild type to certain anticancer agents, such as cisplatin, and certain methylating agents. Inactivation of mismatch repair, however, renders the cell tolerant to these agents. In other words, mismatch repair sensitizes *dam* cells to the effects of these deleterious agents. This is of interest because mammalian cells in culture respond in exactly the same way;

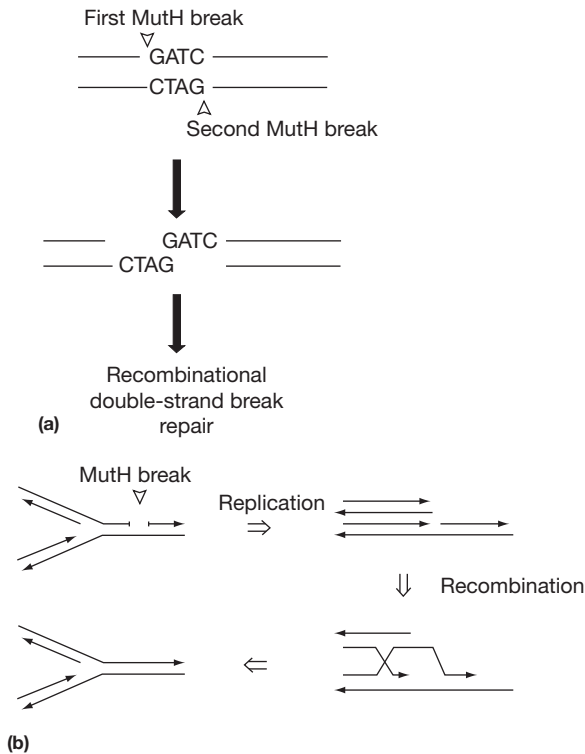


Figure 2 Generation of double-strand DNA breaks. (a) Mismatch repair endonuclease MutH introduces nicks on either side of a GATC sequence to produce a double-strand break. (b) Mismatch repair introduces a nick ahead of a replication fork producing a double-strand break when the fork encounters it. Recombination is required to restore the fork.

that is, they are sensitive to these agents but become tolerant upon loss of mismatch repair capacity. In *dam* cells exposed to a methylating agent, there is an increase in the level of DNA double-strand breaks but these are not present in *dam* cells deficient for mismatch repair. In other words, it is the action of mismatch repair that produces double-strand breaks. DNA double-strand breaks are repaired by homologous recombination but there is *in vitro* evidence that the MutS and MutL proteins inhibit the initial stage of recombination if the substrate DNA has been exposed to a methylating agent. The loss of viability in *dam* bacteria exposed to a methylating agent may, therefore, be due to both the formation of double-strand breaks and the suppression of their repair by mismatch repair enzymes. In mammalian cells exposed to a methylating agent, the molecular basis for mismatch-repair-dependent killing has yet to be defined.

Initiation of Chromosome Replication

In *E. coli*, chromosome replication is initiated only once per cell cycle and Dam methylation is involved in this process. Chromosome replication is initiated at *oriC*, a region which has a 10 times higher than expected density of GATC sequences, by the initiation protein DnaA. The promoter region of the *dnaA* gene

must be methylated for maximal expression and methylated origins are more efficiently initiated than unmethylated ones. Initiation of chromosome replication produces hemimethylated DNA but the region surrounding *oriC* remains hemimethylated for a period of time equal to about one-third of the cell cycle. During this eclipse period, the hemimethylated *oriC* and *dnaA* gene region is bound specifically and with high affinity by the SeqA protein, which prevents methylation by Dam methyltransferase and renders the origin inert to further initiation. By a process that is not understood, SeqA is released from *oriC* later in the cell cycle and methylation ensues, followed subsequently by another initiation event.

The eclipse period is dependent on the level of Dam methyltransferase and can be lengthened or shortened by decreasing or increasing the enzyme level, respectively. *In vivo*, *dam* and *seqA* mutants have an asynchronous initiation pattern and the eclipse period is shortened or absent. The *seqA* mutant, as expected, performs extra initiations at each cell cycle but the *dam* mutant does not. This indicates that in *dam* mutants there must be an alternative mechanism preventing excessive initiation.

In fast-growing cells, the time required to replicate the bacterial chromosome exceeds the culture doubling time, and initiation of replication may take place one, two, or even three generations prior to cell birth. Consequently, cells are born with replicating chromosomes and contain multiple origins of replication. Initiation of replication in such cells takes place only once at each origin within a short time interval of the cell cycle, leading to synchronous initiation. Dam methylation, despite facilitating duplex opening at *oriC*, is not essential for replication. Rather, methylation of these sites is instrumental in maintaining initiation synchrony, because it allows for discrimination between an old uninitiated origin and one that has recently been initiated.

Altered Gene Expression

In wild-type *E. coli*, GATC sequences can be methylated on both strands, hemimethylated or unmethylated. These configurations are used by specific proteins to bind to DNA. For example, as discussed, the *dnaA* gene is maximally transcribed only if the promoter region is fully methylated. This is consistent with its role in the initiation of chromosome replication. Presumably, since *dnaA* gene expression is autoregulated and that DnaA can act as an activator/repressor, fully methylated GATCs act to modulate DnaA binding. The mismatch repair protein, MutH, and the nucleoid organizing protein, SeqA, bind specifically to hemimethylated sequences. There is also evidence that unmethylated GATC sites are protected from Dam by specific binding of regulatory proteins (Fnr and Cap, see below).

The promoter for the *tsp* (transposase) gene of the mobile genetic element, *Tn10*, is active only when it is hemimethylated, thereby ensuring that transposition is coupled with chromosome duplication and occurs only once per cell cycle. A further example is the pyelonephritis-associated pilus (*pap*) operon in pathogenic *E. coli*, where a Pap protein and the transcriptional activator Lrp compete for binding to two specific symmetrically

unmethylated GATC sequences. Depending on which protein is bound to which GATC, *pap* expression is either off or on.

There are few other genes where the rationale of coupling methylation to gene expression is so clear-cut. High-density oligonucleotide array analysis of messenger RNA (mRNA) from *dam* mutants indicates that expression of many genes is altered. In several instances, this is due to competition between Dam methyltransferase and transcriptional activators such as Fnr and Cap competing for the GATC sites that overlap their recognition sequences. For other genes such as *sula* (involved in control of cell division) and *glnS* (glutamyl-tRNA synthetase), which have GATCs in their promoter regions, expression is increased when they are unmethylated. The biological rationale for this is not known. There are also many examples where GATCs in the promoter region have no effect on initiation of transcription. At present, there are no rules to predict if a GATC sequence in the promoter will increase, decrease, or have no effect on the frequency of transcription initiation.

As discussed, *dam* mutant DNA is subjected to DNA breakage *in vivo*. This results in the induction of a stress response termed International Distress Signal (SOS) in which transcription of about 50 genes is increased. The gene products include those that repair DNA damage, recombination, and control cell division. In *dam* mutants, this response is constitutively expressed because three recombination proteins (RecA, RuvA, and RuvB), which are part of the response, are required for cell viability.

Role of Dam Methylation in Bacterial Pathogenesis

Altered levels of Dam methyltransferase can modulate the virulence of certain bacterial pathogens, including *Actinobacillus*, *Aeromonas*, *Brucella*, *Campylobacter*, *Haemophilus*, *Klebsiella*, *Pasteurella*, *Salmonella*, *Vibrio*, and *Yersinia* in animal models and in *Erwinia*, a plant pathogen. The best-studied system is the *Salmonella* mouse model. *Dam* mutant *Salmonella* are 10000 times less capable of killing mice than wild type but a low level of tissue infectivity occurs. Mice that have been infected with the avirulent *dam* mutant become much more resistant to killing by wild-type *Salmonella*, suggestive of a vaccine effect. The molecular basis for the role of Dam methylation in bacterial pathogenesis has yet to be elucidated but the following virulence-related defects have been described for *dam* cells: reduced interaction with intestinal epithelium, reduced motility, envelope instability, increased Std fimbriae, and bile salt sensitivity. It is probable that not all bacterial pathogens will fit with the *Salmonella* mouse model.

Cell functions under control of Dam methylation in various pathogens include secretion of virulence effectors, toxin synthesis and export, flagella and motility, and fimbrial and nonfimbrial adhesins. It is probable that for many of these virulence factors an alteration in Dam-dependent transcription initiation is responsible.

A contrast to the *Salmonella* example is provided by *Neisseria* that can cause meningitis. These bacteria have a *dam* gene but it is not involved in pathogenesis. Rather, it is the defects in the mismatch repair system (*mutS* and *mutL* gene inactivation) that allows for frameshift mutations to occur in DNA sequences controlling expression of certain virulence genes. There is no

general rule about the effect of Dam methylation on virulence; it has to be investigated for each organism.

Posttranscriptional Regulation by Dam Methylation

As discussed, there are many examples of Dam-dependent alteration of transcription initiation of particular genes. There are also a few instances of posttranscriptional regulation affected by Dam methylation. The *dcm* gene is on the same mRNA transcript as the *vsr* gene, which is a component of the very-short-patch (VSP) repair pathway. The level of Dcm in wild-type cells is the same in logarithmic or stationary phases of growth but Vsr is produced only in stationary phase. In a *dam* mutant, there is no production of Vsr at any stage of the growth curve.

Enterohemorrhagic *E. coli* produce, and sit on, raised projections (pedestals) on the surface of mammalian cells grown in culture medium. Pedestal formation requires bacterial secreted proteins including Tir and intimin, which are encoded on the same mRNA transcript. *Dam* mutants produce a greater frequency of pedestals and contain increased levels of Tir and intimin but there is no change in the steady-state level of the mRNA encoding them. *Dam* overproduction in *Yersinia enterocolitica* causes numerous metabolic alterations, including a change in the composition of lipopolysaccharide O-antigen. However, there is no change in the mRNA levels for the genes responsible compared to wild type. The genes of pathogenicity island, SP-1, of *Salmonella enterica* are affected by Dam methylation at both the transcriptional and posttranscriptional levels. A model to explain these Dam-dependent posttranscriptional effects invokes the altered transcription of intergenic small regulatory RNA molecules which in turn regulate translation of the affected genes.

Role of Dam Methylation in Other Bacteria and Their Viruses

At present, little is known about the role of Dam methylation in other bacteria. Inactivation of the *dam* gene in *Yersinia* and *Vibrio*, however, is a lethal event, although the reason is not known. In enterohemorrhagic *E. coli*, inactivation of the *dam* gene is also lethal due to the induction of a prophage in response to the increased level of SOS regulon expression. By contrast, inactivation of the *dam* gene in *Neisseria meningitidis* is not lethal and does not lead to a mutator phenotype, and inactivation of recombination does not lead to lethality. This suggests a fundamentally different role for Dam methyltransferase in this species compared to the *E. coli* paradigm.

The enterobacterial virus P1 has its own *dam* gene that is related to that of its *E. coli* host and is essential for its life cycle. Dam methylation is required at the stage where the viral DNA is packaged into the heads. This is accomplished by filling the head with genomic DNA and then cleaving it at specific sequences, termed *pac* sites. These *pac* sites are flanked by multiple GATC sequences which must be methylated in order for cleavage to occur. In other *E. coli* bacterial viruses, however, such as T2 and T4, deletion of the *dam* gene has no discernable effect on their life cycle. *Dam* genes are also present in certain

lambdoid bacteriophages but the essentiality of these genes in the viral life cycle has not yet been determined.

See also: **Molecular Biology:** DNA Mismatch Repair in Bacteria; Mammalian DNA Methyltransferase Structural Themes.

Further Reading

Low DA and Casadesús J (2008) Clocks and switches: Bacterial gene regulation by DNA adenine methylation. *Current Opinion in Microbiology* 11: 106–112.

Marinus MG and Casadesús J (2009) Roles of DNA adenine methylation in host–pathogen interactions: Mismatch repair, transcriptional regulation, and more. *FEMS Microbiology Reviews* 33: 488–503.

Marinus MG and Løbner-Olesen A (2009) DNA methylation. In: Böck A, Curtiss R, III, Kaper JB, et al. (eds.) *EcoSal – Escherichia coli and Salmonella: Cellular and Molecular Biology*. Washington, DC: ASM Press.

Relevant Website

<http://www.ecosal.org> – The new, continually updated Web resource based on the classic ASM Press publication *Escherichia coli and Salmonella: Cellular and Molecular Biology*.

DNA Mismatch Repair and Homologous Recombination

I Matic, INSERM U1001, Université Paris Descartes, Paris, France

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Ivan Matic and Miroslav Radman, volume 1, pp. 665–670, © 2004, Elsevier Inc.

Glossary

Homologous recombination Interaction between two DNA sequences sharing extensive nucleotide sequence identity, present on a single or two different DNA molecules, which results in a generation of mixed sequences derived from two parental ones. Such recombination events can be nonreciprocal (gene conversion or ‘patch’ recombinants) or reciprocal (crossover or ‘splice’ recombinants).

Meiosis (*meion*: smaller (Greek)) Two successive nuclear and cellular divisions resulting in the reduction of the chromosome number from diploid to haploid. The products

of meiosis are haploid gametes. As chromosome segregation in meiosis requires physical attachment of homologous chromosomes (chiasmata), each chromosome in the gamete is recombinant.

Mismatch The mispaired or unpaired bases in the double-strand DNA molecule.

Mitosis (*mitos*: thread (Greek)) The nuclear and cellular division of eukaryotic somatic cell resulting in generation of two daughter cells having the same number of chromosomes as the parent cell.

Mutation (*mutare*: to change (Greek)) Any heritable modification of genetic material.

Mismatches are mispaired or unpaired bases in the double-strand DNA molecule. Mismatches are generated as a consequence of errors during DNA replication or during homologous recombination involving nonidentical sequences. Specific mismatches can be formed by DNA damage. Enzymes that recognize and process mismatches have been identified in prokaryotes and eukaryotes. Different mismatch-repair systems (MRSs) are distinguished on the basis of their mismatch specificity and the size of excision tract that varies from single nucleotide to over a kilobase. The long-patch, also called the general, MRS (because it recognizes a variety of mismatches), is highly conserved during evolution. This molecular system maintains genomic stability by controlling fidelity of chromosomal replication by eliminating DNA biosynthetic errors and by controlling meiotic and mitotic homologous recombination. The control of homologous recombination is particularly important because this fundamental biological process is essential for the maintenance of chromosomal integrity, generation of genetic diversity, proper segregation of meiotic chromosomes, and speciation. In addition, MRS is required for activation of cell cycle arrest and apoptosis in response to certain types of DNA lesions, such as alkylation DNA damage. MRS also participates in transcription-coupled repair, in nonhomologous end joining (NHEJ) and in somatic hypermutation.

Specific mismatches created by chemical modification of DNA bases (e.g., G•T mismatch due to deamination of 5-methyl-C to T or 8-oxo-G•A mismatch due to oxidation of G) are repaired by specialized short-patch MRS. When such mismatches occur in the course of recombination, their localized repair in heteroduplex DNA creates mosaic sequences, that is, an apparent hyper-recombination effect (Figure 1).

Generalized MRSs

Escherichia coli Methyl-Directed Mismatch-Repair Pathway

The best-characterized, generalized MRS is *Escherichia coli* methyl-directed MRS. *E. coli* MRS, which has been completely

reconstituted *in vitro*, involves three dedicated proteins: MutS, MutL, and MutH. Other proteins that participate in processing of mismatches, DNA helicase II (UvrD), four exonucleases (Exo I, Exo VII, Exo X, and RecJ), SSB, DNA polymerase III holoenzyme, and DNA ligase, are shared with other repair pathways. MutS, protein recognizes and binds to a mismatch as a homodimer. The mismatch-binding domains of the two MutS proteins, although having identical amino-acid sequence, are structurally and functionally different. This indicates that the MutS homodimer–DNA complex is not symmetrical. Hence, the MutS homodimer acts as a functional heterodimer when bound to a DNA mismatch.

MutL homodimer associates with the MutS homodimer–mismatch complex, and recruits and activates MutH protein. MutH, which functions as a monomer and belongs to a family of type-II restriction endonucleases, incises the newly synthesized strand at a nearby hemi-methylated 5'-GATC-3' site. Hemi-methylated 5'-GATC-3' may reside either 3' or 5' to the mismatch at distances of, as much as, 1 kb or more. The N6 position of adenine in GATC sites is transiently not methylated in the newly synthesized strand, because methylation by Dam methylase lags behind replication by ~2 min. This strand-specific nick provides the initiation site for the mismatch-repair process and directs it to the newly synthesized strand. Neither MutH nor a nonmethylated strand is required if a single-strand break is present in the substrate DNA.

MutL also stimulates the loading and the processivity of UvrD at the mismatch-repair initiation site. UvrD unwinds the DNA duplex from the nick toward the mismatch, thereby allowing degradation of the displaced single-strand DNA, assuring the irreversibility of the repair process. The involvement of different single-strand-specific exonucleases in this repair pathway depends on orientation of the nick relative to the mismatch. Exo I or Exo X (3'→5' exonuclease), or Exo VII (3'→5' and 5'→3' exonuclease), or RecJ (5'→3' exonuclease) excises the nicked strand from the nicked site up to and slightly past the mismatch. The resulting single-stranded gap undergoes

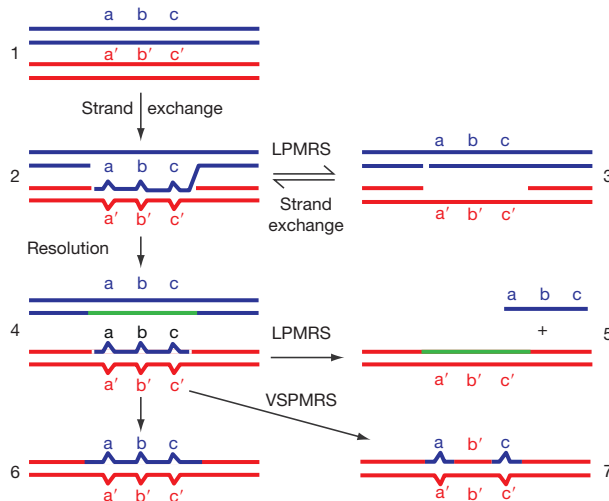


Figure 1 Anti-recombination by long patch mismatch-repair system (LPMRS) and hyper-recombination by very short patch mismatch-repair system (VSPMRS). a/a', b/b', and c/c' represent single-nucleotide polymorphisms between two recombining DNA molecules (blue and red) (1). For simplicity, only one kind of strand exchange (asymmetric strand exchange) is shown; a:a', b:b', and c:c' are mismatched bases in the heteroduplex intermediate (2). LPMRS can destroy mismatched heteroduplex before (3) or after (4 and 5) resolution of recombination intermediates. Heteroduplexes that escape LPMRS, and which are sealed by ligase (6), will be separated by subsequent replication producing one recombinant and one parental DNA molecule. If some of the mismatches (e.g., b:b') in this heteroduplex molecule are substrate for VSPMRS, its activity will generate a complex recombinant on the repaired strand (7). Green lines represent tracts of DNA repair synthesis.

repair DNA resynthesis and ligation by DNA polymerase III holoenzyme, SSB, and DNA ligase. The excision-repair tracts associated with this pathway can be a kilobase long or longer. Hence, this repair system is referred to as the 'long-patch repair system.'

The critical first step in mismatch repair is the recognition and binding of MutS to mismatches. Detection of the mismatches, which are lacking the modified chemical groups generally present in most types of DNA damage facilitating their recognition, is a challenging task for the MutS protein. Binding affinities of MutS are about 10–1500-fold higher for mismatched than for perfectly matched DNA. Mismatch recognition by MutS is based on weakened base stacking of the mismatched base pairs and susceptibility to kinking rather than a specific shape or hydrogen-bonding pattern. Base pairing at the mismatch site undergoes significant reorganization upon MutS binding resulting in the 45–60° kink in the DNA helix.

MutS protein binds to seven of eight possible base pair mismatches. C●C mismatches, which are the least frequent replication error, are refractory. In addition, MutS protein binds up to four unpaired bases allowing for repair of frameshift errors. The affinities of MutS protein for various mismatches *in vitro* reflect the efficiency of the mismatch *in vivo*. Hence, transitions G●T et A●C, and transversions G●G et A●A, are repaired with the maximal efficiency. C●C are not repaired. Efficiency of the mismatch repair of transversions T●T, T●C et G●A depends on the local sequence context. The inactivation of *mutS*, or any other of *mut* genes coding for

general mismatch repair, results in strong identical mutator phenotypes with 10^2 – 10^3 -fold increased rates of transition and frameshift mutations.

MutS has an intrinsic ATPase activity, which actively dissociates MutS from DNA and performs a proofreading function to enhance mismatch-repair specificity. In the presence of homoduplex DNA, MutS quickly hydrolyzes ATP, but in the presence of a mismatch, the ATP hydrolysis is inhibited, which allows the MutS–DNA–ATP complex to form. Only when associated with both ATP and a mismatch is MutS able to recruit MutL protein. Thus, MutS uses ATP as a high-energy cofactor to compensate its low binding specificity and increase the repair specificity of the mismatch repair. Such kinetic proofreading ensures that interactions of mismatch-repair proteins with nonsubstrate DNA result in no repair.

Generalized Mismatch Repair in Eukaryotes

The knowledge about *E. coli* general MRS greatly facilitated identification and comprehension of homologous MRS in other organisms. Core MRS functions, MutS, and MutL proteins are conserved in almost all prokaryotes and eukaryotes. In eukaryotic organisms, invertebrates, vertebrates, and plants, multiple MutS and MutL homologs have also been identified. Different species contain different sets of homologs, without an obvious correlation to their position in the evolutionary tree. However, unlike their prokaryotic homologs, these proteins function as heterodimers. The most extensively characterized eukaryotic generalized MRS is that of *Saccharomyces cerevisiae*. In this organism, six MutS homologs (Msh 1–6) have been identified. Msh1 functions in mitochondria, while others act in the nucleus. Heterodimers of the MutS homologs cooperate with heterodimers of MutL homologs to process different mismatches. MutS α (Msh2–Msh6 heterodimer) complex recognizes base pair mismatches and insertions/deletions of one or two nucleotides. MutS β (Msh2–Msh3 heterodimer) binds to up to 12 nucleotides large insertion/deletion loops.

Four MutL homologs (Mlh 1–3 and Pms1) were identified in *S. cerevisiae*. Mlh2, Mlh3, and Pms1 proteins form heterodimers with Mlh1, which is a common subunit. These complexes are referred to as MutL α (Mlh1–Pms1 heterodimer), MutL β (Mlh1–Mv2 heterodimer), and MutL γ (Mlh1–Mlh3 heterodimer). The majority of mismatch-repair activity is carried out by MutL α interacting with either MutS α or MutS β complexes. MutL β and MutL γ have distinct specialized roles in the repair of mutational intermediates. As in prokaryotes, additional factors participate in *S. cerevisiae* mismatch repair: proliferating cell nuclear antigen (PCNA; processivity factor for replication machinery), Exo1 (5'→3' exonuclease), and the DNA polymerases δ and ϵ . The involvement of MutS and MutL homolog heterodimers was also observed in other eukaryotic organisms including humans.

There are some other significant differences between *E. coli* MRS and MRSs found in other organisms. In most organisms, DNA-strand recognition is not governed by DNA methylation. Dam methylase, coupled with the presence of *mutH* gene, has been identified in only a small number of bacterial species. The mechanism of strand discrimination in other prokaryotes and eukaryotes is not yet elucidated. However, a nick or a gap can serve as a signal that directs the reaction not only in *E. coli*, but

also in eukaryotic cells. This fact, together with the observation that mismatch repair is more efficient on the lagging strand than on the leading strand, suggest that DNA termini that occur as natural intermediates during replication may suffice as strand signals to direct the correction of DNA biosynthetic errors in eukaryotic cells.

Human and *S. cerevisiae* MutL α is a latent endonuclease that incises the discontinuous strand of a nicked heteroduplex in a mismatch- MutS α -, RFC-, PCNA-, and ATP-dependent manner. The activity of this endonuclease is dependent on the integrity of a DQHA(X)₂E(X)₄E metal-binding motif located within the C-terminal portion of PMS2. This motif is highly conserved in eukaryotic PMS2 homologs and is also common in archaeobacterial and eubacterial MutL proteins, but is lacking in MutL proteins from bacteria such as *E. coli* that rely on adenine methylation in 5'-GATC-3' sequences for strand discrimination during mismatch repair. Although *E. coli* MutH and human MutL are both latent endonucleases, these activities play functionally distinct roles in the two repair systems. Whereas MutH incision provides a nick, which serves as signal that directs repair, incision by MutL depends on preexistence of a signaling strand break.

MRS and Homologous Recombination

Anti-Recombination Activity in *E. coli*

Homologous recombination can be defined as interaction between two DNA sequences sharing extensive nucleotide sequence identity, present on a single or two different DNA molecules, which results in a generation of mixed sequences derived from two parental ones. The homologous recombination process occurs in four basic steps: initiation, homologous pairing and DNA exchange, DNA heteroduplex extension, and resolution. The recombination process is initiated by the combined action of a variety of nucleases and helicases that generate single-strand DNA. Single-strand DNA is a substrate for proteins that catalyze homologous pairing and DNA exchange. The best-studied homologous pairing and DNA-exchange protein is *E. coli* RecA.

RecA protein binds to single-strand DNA, or double-strand DNA containing single-strand DNA gaps, and forms a nucleoprotein complex containing one RecA protein monomer for every three nucleotides. RecA-single-strand DNA filaments engage in the search for homologous target DNA sequences. Scanning DNA for homologous sites is very fast and highly accurate even when they are very rare in the genome. *In vitro* data show that one helical turn of a RecA nucleoprotein filament, containing approximately six RecA monomers and 15 bases of single-strand DNA, is the functional unit sufficient to carry out the homology search. *In vivo* data show that RecA-catalyzed pairing is effective provided the length of shared uninterrupted identity is at least 26 nucleotides. This minimal effective DNA sequence identity required for homologous recombination is called the minimal efficient processing segment (MEPS).

After homologous pairing is achieved, the resultant joint molecules exchange homologous strands and establish heteroduplex DNA where each parental DNA contributes one complementary strand. Pairing of nonidentical DNA sequences produces mismatched heteroduplex molecules. RecA can detect

single mismatches at the initial stages of strand exchange, whose efficiency is strongly dependent on the location and distribution of mismatches. Mismatches near the 5' end of the incoming strand have a small effect, whereas mismatches near the 3' end strongly impede strand exchange. Once established, heteroduplex DNA is extended by a unidirectional branch migration step. At this step of recombination, RecA protein tolerates DNA lesions, large heterologies, and up to 30% nucleotide sequence divergence between recombining molecules.

MRS inhibits recombination between nonidentical DNAs as mismatched heteroduplex molecules are the substrate for MRS (Figure 1). MRS is capable of impeding recombination even between sequences with very low divergence, that is, less than 1%. While MRS inhibits recombination between DNAs differing by point substitution mutations and small insertions/deletions, it cannot prevent recombination between DNAs that have insertions/deletions >4 bp because resulting loops in heteroduplex molecules are not recognized by MRS. The MRS exerts anti-recombinogenic activity at different steps of homologous recombination. First, it acts as a fidelity element, which determines the length of MEPS by imposing an increased rigor during the homology search catalyzed by the RecA filament. Second, MRS blocks the RecA-catalyzed heteroduplex elongation, which results in the reversion of this step of homologous recombination. The increased concentration of RecA protein does not prevent mismatch-repair anti-recombination activity, whereas the overproduction of MutL protein severely reduces recombination frequency between nonidentical DNAs (Figure 2). Cellular MutL level is critical for the tolerance of base-pair mismatches in heteroduplex DNA. As MutL moderates the ATP-dependent dissociation of MutS from DNA, high concentration of MutL protein results in a longer sequestration of MutS protein on mismatches and therefore more effectively blocks the strand exchange process.

Unlike the identical effect of MRS mutants on spontaneous mutagenesis, the inactivation of different MRS genes has a distinct and characteristic effect on the recombination between nonidentical sequences. Very strong hyper-recombination effect is observed upon inactivation of *mutS* and *mutL* genes (Figure 2), while the effect of inactivation of *mutU* and *mutH* genes is relatively weak. The presence of single-strand ends in recombination intermediates (Figure 1) can explain the weak effect of *mutH* gene inactivation. The helicase II can probably be replaced by other enzymes that are involved in recombination, for example, RecG.

The anti-recombination activity of MRS in bacteria controls intergenomic gene transfer mediated by conjugation, transduction, and transformation. In addition, it controls intrachromosomal recombination between repeated, interspersed DNA sequences thus preventing potentially deleterious genomic rearrangements. For example, MRS reduces by 250-fold, recombination frequency between 96% identical repeated DNA sequences. Similarly, the frequency of intrachromosomal recombination between *E. coli* *rhsA* and *rhsB* loci (99% identical) increases 10–15-fold in *mutL* or *mutS* backgrounds. MRS also reduces the rate of gene conversion 1000-fold between <1% divergent *S. enterica* *tufA* and *tufB* genes. Gene conversion is a phenomenon in which genetic information on one chromosomal site is replaced with information from another site, which remain unchanged (nonreciprocal recombination).

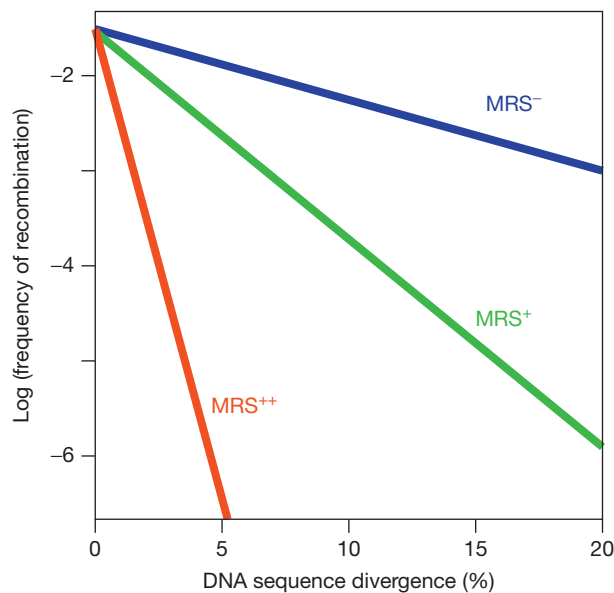


Figure 2 The relationship between DNA sequence divergence, homologous recombination, and mismatch-repair system (MRS). Conjugational crosses between enterobacterial strains and species show that linear increase of sequence divergence between recombining genomes results in the exponential decrease in recombination frequency. This relationship shows that an increase in DNA divergence increases the difficulty for RecA protein to find the sites for initiation of homologous pairing and strand exchange. The MRS activity reduces the apparent length of these sites and blocking the maturation of RecA-catalyzed recombination intermediates. Consequently, the inactivation of MRS decreases the slope of correlation between recombination and DNA divergence, i.e., increasing the frequency of recombination between diverged sequences. By contrast, the overproduction of the key MRS proteins (MutS and MutL) increases the slope (MRS⁺⁺), i.e., creates new genetic barriers.

Therefore, the degree and distribution of sequence divergence between recombining DNA molecules are structural parameters that influence recombination efficiency because they affect the activity of recombination enzymes and determine whether the anti-recombinogenic activity of MRS proteins is triggered. Studies of the effects of DNA sequence divergence on the efficiency of homologous recombination show that there is an exponential decrease in recombination frequency with a linear increase in DNA divergence. This relationship results from an exponential decrease in the number of available MEPSs due to the requirements of RecA protein for the sequence identity. The slope of this relationship is determined by MRS activity, which alters MEPS size.

Anti-Recombination Activity in *S. cerevisiae*

Recombination between sequences at identical, allelic, positions on homologous chromosomes is important for mitotic DNA repair and for meiotic chromosome disjunction in eukaryotes. In addition to the allelic interactions, eukaryotic genomes contain large numbers of repeated sequences that can serve as substrates for nonallelic, ectopic, recombination events. Ectopic recombination can lead to alterations in genome structure and must be avoided as much as possible. As in bacteria,

recombination efficiency in eukaryotes depends on the DNA sequence identity and is controlled by different molecular mechanisms. Regulation of recombination in eukaryotes is best characterized in *S. cerevisiae*. In this organism, like in bacteria, recombination frequency decreases exponentially with a linear increase in DNA divergence between recombining DNA molecules. This log-linear relationship is due to the decrease of the number of MEPSs, which is estimated to be around 250 bp for *S. cerevisiae*. Mismatched recombination intermediates are substrate for MRS in *S. cerevisiae*, where it inhibits recombination between nonidentical DNA sequences both in mitosis and in meiosis. Even a single mismatch in heteroduplex molecule is sufficient for MRS to inhibit recombination. Additional mismatches have a cumulative negative effect on recombination efficiency. When DNA sequence divergence reaches several percent, the probability of a heteroduplex recombination intermediate escaping detection by the MRS is so low that the additional mismatches fail to increase anti-recombination activity. The MutS α or MutS β complexes recognize the same types of mismatches in recombination intermediates as in replication intermediates. Although *msh2* and *mlh1* or *pms1* mutants have identical mutator phenotypes, *msh2* mutants exhibit higher recombination rates between nonidentical DNAs than do *pms1* or *mlh1* mutants. This indicates that Msh binding to mismatch may by itself inhibit the recombination process.

In *S. cerevisiae*, MRS processing of recombination intermediates results in gene conversion or in complete destruction of recombination intermediates. In *S. cerevisiae*, where it is possible to analyze all four products derived from a single meiosis, a non-Mendelian 3:1 segregation pattern of allelic sequences is diagnostic of a gene conversion event. In the absence of a functional MRS, mismatches in heteroduplex recombination intermediates are not repaired and the mismatch-containing DNA strands segregate at the first mitotic division following meiosis (postmeiotic segregation or PMS). The frequency of PMS reflects the repair efficiency for the particular mismatch. In *msh2*, *pms1*, *mlh1*, and, to a lesser extent, *msh3* and *msh6* mutants, meiotic gene conversion events are reduced while PMS events are elevated. Since the repair patch is shorter than the meiotic heteroduplex DNA, the repair process does not necessarily interfere with the formation of crossing-overs. MRS also inhibits crossing-overs between diverged chromosomes in interspecific hybrids during meiosis, and is therefore important for establishing and maintaining interspecific genetic barriers as well.

Although the same mismatch-repair machinery detects mismatches in both recombination and replication intermediates, some accessory proteins are unique to each type of repair. For example, PCNA appears to play a role in correcting DNA replication errors, but has not been implicated in the correction of heteroduplex recombination intermediates. In addition, two *S. cerevisiae* MutS homologs, Msh4 and Msh5, have completely lost the ability to participate in a standard mismatch-repair reaction, but have acquired novel, meiosis-specific roles. The loss of either MSH4 or MSH5 is associated with an approximately 50% reduction in meiotic crossing-over, in increased levels of homolog nondisjunction, and decreased viability of meiotic products. The Msh4 and Msh5 genes are in the same epistasis group and, like the other MutS homologs, physically interact to form heterodimers. As gene

conversion levels are normal in *msh4* or *msh5* mutants, the Msh4–Msh5 complex is assumed to act late in recombination. Genetic data indicate that the Msh4–Msh5 heterodimer interacts specifically with the Mlh1–Mlh3 complex to promote meiotic crossing-over. Both *mlh1* and *mlh3* mutants exhibit a reduction in meiotic crossover events, and epistasis analysis indicates that Mlh1 and Msh4 act in the same crossover pathway. Interestingly, mitotic recombination in yeast is more sensitive (in a mismatch-repair-dependent manner) to low levels of sequence divergence than is meiotic recombination, which may help reinforce the strong bias for sister chromatid versus interhomolog interactions during mitosis.

MRS was also shown to play an important role in inhibition of ectopic interactions between nonidentical, repetitive DNA sequences, during both mitosis and meiosis, and hence maintain genome stability by preventing genome rearrangements. Single-strand annealing (SSA) is a major recombination pathway for repairing spontaneous and induced double-strand breaks that arise between repeated sequences. Annealing of complementary single-stranded DNA molecules produced by the exonucleolytic processing of double-strand breaks adjacent to repetitive genomic sequences produces a variety of branched DNA structures including the 3' tails at the junction of double-stranded and single-stranded DNA. Msh2–Msh3 heterodimer binds to these structures and stabilizes them, thus allowing the removal of nonhomologous DNA tails by structure-specific nuclease, Rad1–Rad10.

During SSA, the requirement for Msh2–Msh3 depends upon the length of the annealed region. Rad1–Rad10-dependent 3' end processing requires Msh2–Msh3 only when the annealed region is lesser than 1 kb. Decreased dependence on Msh2–Msh3 is also observed with larger loop sizes during Rad1–Rad10-dependent meiotic loop repair. Msh2 physically interacts with both Rad1 and Rad10 independently of other mismatch-repair factors, and no other mismatch-repair factors are required for Rad1–Rad10-dependent 3' end processing, besides Msh3. Msh2 localizes to double-strand break sites flanked by nonhomologous sequence, and Msh2 and Msh3 physically interact with subunits of the single-strand binding protein RPA. The role of Msh2–Msh3 in nonhomologous tail removal during recombination can be distinguished from its role in DNA mismatch repair, as mutations have been isolated in *MSH2* that disrupt mismatch repair but are functional for recombination.

DNA Sequence Divergence, Mismatch Repair, Loss of Heterozygosity, and Cancer in Mammals

Except for the mitochondrial Msh1 protein, all other *S. cerevisiae* MutS and MutL proteins are present in mammalian cells. Defects in MRS in mammals are associated with genome-wide instability, predisposition to certain types of cancer, including hereditary nonpolyposis colorectal cancer, resistance to certain chemotherapeutic agents, and abnormalities in meiosis and sterility. Carcinomas evolve by accumulation of several somatic mutations, of which some are recessive (e.g., in tumor suppressor genes). Such mutations are expressed only when homozygous, that is, when both alleles in diploid cells are inactivated by mutation. The passage from the heterozygous (*m/+*) to

homozygous (*m/m*) or hemizygous (*m/o*) state usually occurs by a chromosomal rearrangement leading to the loss of the functional gene copy (Figure 3). Such rearrangement can be due to (1) the loss of the (+) allele (e.g., a small or large deletion, or the loss of the entire chromosome) or (2) a homologous recombination event, such as gene conversion, of the allele to (*m/m*) or a mitotic crossover anywhere between the centromere and the allele (Figure 3). All these events, with the exception of gene conversion, lead to the loss of heterozygosity (LOH), initially present because of the maternal/paternal sequence divergence (about 0.1% of nucleotides in human population).

Two lines of experiments suggest that the functionally largely neutral sequence polymorphism between the two parental lines effectively suppresses the mitotic recombination mechanism for LOH:

1. Gene targeting in mouse embryonic stem cells is highly sensitive to the natural sequence polymorphism of the parental strains. This suppression of mitotic recombination can be relieved by the defect in mismatch repair, that is, it appears that the mammalian mismatch-repair acts to prevent mitotic recombination between nonidentical sequences.
2. The expression of a recessive heterozygous (*m/+*) mutation in inbred, globally highly homozygous mice, occurs at elevated frequency *via* mitotic recombination, that is, LOH from the crossover point to the end of the chromosome

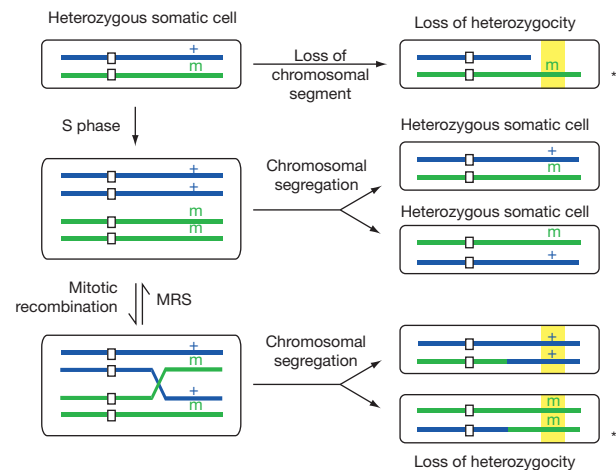


Figure 3 DNA sequence divergence, mismatch repair, and loss of heterozygosity. Only one pair of homologous chromosomes is shown. The chromosome colors (blue and green) indicate the sequence polymorphism between the maternal and paternal line (0.1% in human population). The symbol (*m*) stands for a specific recessive mutation. Only cells (*) having lost the wild type (+) allele express the (*m*) mutation. Such cells show loss of heterozygosity (LOH; yellow). In human tumors, most LOH occurs by chromosomal rearrangements (top right). However, in inbred mice, with little or no sequence divergence between homologs, most LOH occurs by mitotic recombination. The recombination-suppressing effect of sequence polymorphism is relieved in MRS-deficient mouse cells (see Figure 1). Because LOH is an obligate step in the expression of recessive mutations in tumor suppressor genes, sequence polymorphism protects against cancer due to MRS activity.

including the recessive mutation (Figure 3). In hybrids between different strains, the LOH frequency is much lower than in inbred strains. The decrease in LOH is accounted by the loss of mitotic recombination due to the activity of the MRS.

Thus, it appears that, like in bacteria and *S. cerevisiae*, recombination in mammals has same requirements for DNA sequence identity. Increase of sequence divergence between recombining DNA molecules reduces recombination efficiency by decreasing the number of available MEPSs and by the anti-recombination activity of MRS. MEPSs in mammalian cells was estimated to be about 250 bp. Such genetic barriers based on the genomic sequence polymorphism are expected to (1) delay the process of carcinogenesis by delaying the expression of relevant recessive somatic and inherited mutations and (2) enhance sympatric speciation by creating genetic isolation on the germline level.

Evolutionary Role of the Control of Homologous Recombination by Mismatch Repair

Evolutionary conservation of key MRS functions illustrates the importance of this genetic system in the preservation of genomic integrity in prokaryotes and eukaryotes. In bacteria, the loss of MRS activity can provide transient selective advantage when adaptation is limited by the supply of mutations. However, the restoration of MRS proficiency is a prerequisite for evolutionary success because upon adaptation, the load of deleterious mutations is no longer counterbalanced by the generation of beneficial mutations resulting in fitness reduction. In eukaryotes, the increased genomic instability resulting from MRS deficiency is apparently never advantageous. For example, in mammals, the loss of MRS activity is associated with male and female sterility and tumorigenesis.

Increased frequency of chromosomal rearrangements is a deleterious phenotype of MRS deficiency in prokaryotes and eukaryotes. The MRS-mediated recombination surveillance is particularly important for eukaryotes because their genomes contain a multitude of repetitive DNA sequences. MRS in *S. cerevisiae* prevents intra- and interchromosomal crossovers and gene conversions resulting from recombination between nonidentical DNA sequences.

By controlling homologous recombination, MRS also controls the process of speciation. In bacteria, MRS controls conjugational, transductional, and transformational recombination between strains and species, thus reducing gene flow between diverged populations. Similarly, by controlling crossing-over and chromosome segregation in meiosis, MRS prevents hybridization between diverged organisms resulting in reproductive isolation of closely related species.

See also: **Molecular Biology:** DNA Methyltransferases: Eubacterial GATC; DNA Mismatch Repair and the DNA Damage Response; DNA Mismatch Repair in Bacteria; DNA Mismatch Repair in Mammals; Homologous Recombination in Meiosis.

Further Reading

- Datta A, Hendrix M, Lipsitch M, and Jinks-Robertson S (1997) Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 94: 9757–9762.
- de Wind N, Dekker M, Berns A, et al. (1995) Inactivation of the mouse Msh2 gene results in the mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell* 82: 321–330.
- Delmas S and Matic I (2005) Cellular response to horizontally transferred DNA in *Escherichia coli* is tuned by DNA repair systems. *DNA Repair* 4: 221–229.
- Denamur E, Lecointre G, Darlu P, et al. (2000) Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell* 103: 711–721.
- Denamur E and Matic I (2006) Evolution of mutation rates in bacteria. *Molecular Microbiology* 60: 820–827.
- Elez M, Radman M, and Matic I (2007) The frequency and structure of recombinant products is determined by the cellular level of MutL. *Proceedings of the National Academy of Sciences of the United States of America* 104: 8935–8940.
- Friedberg EC, Walker GC, Siede, et al. (2006) *DNA Repair and Mutagenesis*, 2nd edn. Washington, DC: ASM Press.
- Hoffmann ER and Borts RH (2004) Meiotic recombination intermediates and mismatch repair proteins. *Cytogenetic and Genome Research* 107: 232–248.
- Hunter N, Chambers SR, Louis EJ, and Borts RH (1996) The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO Journal* 15: 1726–1733.
- Jones M, Wagner R, and Radman M (1987) Mismatch repair and recombination in *E. coli*. *Cell* 50: 621–626.
- Jun SH, Kim TG, and Ban C (2006) DNA mismatch repair system: Classical and fresh roles. *FEBS Journal* 273: 1609–1619.
- Kadyrov FA, Holmes SF, Arana ME, et al. (2007) *Saccharomyces cerevisiae* MutL[α] is a mismatch repair endonuclease. *Journal of Biological Chemistry* 282: 37181–37190.
- Kunkel TA and Erie DA (2005) DNA mismatch repair. *Annual Review of Biochemistry* 74: 681–710.
- Li GM (2008) Mechanisms and functions of DNA mismatch repair. *Cell Research* 18: 85–98.
- Marra G and Jiricny J (2005) DNA mismatch repair and colon cancer. *Advances in Experimental Medicine and Biology* 570: 85–123.
- Rayssiguier C, Thaler DS, and Radman M (1989) The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature* 342: 396–401.
- Shao C, Yin M, Deng L, et al. (2002) Loss of heterozygosity and point mutation at Aprt locus in T cells and fibroblasts of PMS2 $^{-/-}$ mice. *Oncogene* 21: 2840–2845.
- Surtees JA, Argueso JL, and Alani E (2004) Mismatch repair proteins: Key regulators of genetic recombination. *Cytogenetic and Genome Research* 107: 146–159.
- Vulic M, Dionisio F, Taddei F, and Radman M (1997) Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proceedings of the National Academy of Sciences of the United States of America* 94: 9763–9767.
- Waldman AS (2008) Ensuring the fidelity of recombination in mammalian chromosomes. *BioEssays* 30: 1163–1171.

DNA Mismatch Repair and the DNA Damage Response

G-M Li, University of Kentucky Medical Center, Lexington, KY, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Guo-Min Li and Steven R. Presnell, volume 1, pp. 671–674, © 2004, Elsevier Inc.

Glossary

Apoptosis An active cell death process that requires RNA and protein synthesis to specifically digest cellular DNA into nucleosomal fragments; it is also known as programmed cell death.

Cytotoxicity The degree to which an agent possesses a specific destructive action on certain cells or the possession of such action; it is often used when referring to the action of antineoplastic drugs that selectively kill dividing cells.

DNA mismatch Non-G:C or -A:T DNA base pairs or small insertion/deletion (unpaired) nucleotides in duplex DNA; it is also called a heteroduplex.

DNA mismatch repair A DNA repair pathway that specifically converts mismatched DNA bases (heteroduplexes) into normal G:C or A:T base pairs (homoduplexes).

Tumorigenesis The process that generates tumors.

DNA mismatches are base-pairing errors that deviate from the Watson–Crick rules that stipulate that A pairs with T and G pairs with C. In addition, small insertion/deletion structures, in which one DNA strand contains a small number of unpaired nucleotides, are also considered mismatches. Mismatched base pairs occur as errors during the normal course of DNA replication and recombination, and they are mutagenic if left uncorrected. DNA mismatch repair (MMR) is a cellular process that rectifies mismatches to yield correct Watson–Crick base pairs. MMR is an important genome maintenance system because defects in the system cause genome-wide instability and have been implicated in the development of certain types of human cancer. In the past, the ability of the MMR system to correct DNA heteroduplexes has been considered the primary mechanism by which it contributes to genomic stability. However, more recent studies indicate that the MMR system also contributes to genomic stability by mediating protective cellular responses to DNA damage.

MMR Proteins and Their Role in Mismatch Correction

The eukaryotic MMR pathway is similar to the *Escherichia coli* MthLS pathway with respect to features of mechanism and involvement of several homologous activities. In fact, many of the eukaryotic MMR components were initially identified by their sequence homology to the *E. coli* MMR proteins (see Table 1). The eukaryotic MMR proteins are known to include the MutS homologs (MutS α and MutS β), the MutL homologs (MutL α , MutL β , and MutL γ), exonuclease I (ExoI), replication protein A (RPA), a eukaryotic SSB (single-strand DNA-binding protein), and DNA polymerase δ . Although eukaryotic homologs of the bacterial MthH and helicase II have not yet been identified, studies in human nuclear extracts have demonstrated that the human MMR reaction possesses a mechanism similar to that of *E. coli*, that is, the repair is targeted to the newly synthesized strand and can occur bidirectionally. The eukaryotic system appears, however, to be more complex than the *E. coli* pathway. For example, whereas there is

only a single form of MutS or MutL in *E. coli*, at least two MutS homologs and three MutL homologs have been identified in eukaryotic cells, each of which is a heterodimer. For example, MutS α consists of the polypeptides MSH2 and MSH6, and MutL α comprises MLH1 and PMS2 (see Table 1). These MutS and MutL heterodimers, although redundant, play special roles in MMR subpathways. In human cells, MutS α and MutL α are the most abundant species among the MutS and MutL heterodimers, respectively.

MMR Function Is Required for Drug Cytotoxicity

Although MMR is well known for its role in correcting biosynthetic errors, recent studies have revealed another important role that the MMR system plays in genome maintenance, that is to mediate DNA damage signaling programmed cell death (or apoptosis).

Treatment of cells with chemical DNA-damaging agents, such as the alkylating agents *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG), temozolomide, and procarbazine, leads to increased amounts of cell death. For this reason, these cytotoxic agents are often used in chemotherapy to kill rapidly growing tumor cells. Interestingly, it has been found almost universally that, whereas cells that are proficient in MMR are sensitive to these agents, cells that are deficient in MMR are more resistant to killing by these agents. This was first observed with *E. coli* in the 1980s when *E. coli* MMR mutants (*mutS*[−] or *mutL*[−]) were found to be more resistant to killing by MNNG than were wild-type bacterium. Subsequently, resistance to these alkylating agents was observed in human cells that are defective in MutS α or MutL α . An interesting example is afforded by the development of the lymphoblastoid cell line MT1. This cell line was derived from the TK6 cell line by selection with a high dose of MNNG. MT1 cells are 500-fold more resistant to killing by this agent than the parental cells and exhibit an elevated spontaneous mutation rate. The MT1 cell line was subsequently found to be defective in *MSH6*, the gene that encodes a subunit of MutS α . Conversely, many human

Table 1 MMR components and their function^a

<i>E. coli</i>	Human	Function
MutS	MutS α (MSH2-MSH6) ^b	DNA mismatch/lesion sensor
MutL	MutS β (MSH2-MSH3) MutL α (MLH1-PMS2) ^b MutL β (MLH1-PMS1) MutL γ (MLH1-MLH3)	Molecular chaperon, endonuclease, and excision terminator
MutH	Not identified	Strand discriminator
Helicase II	Not identified	Unwinding DNA helix
ExoI, ExoVII, ExoX, RecJ	ExoI	Removing mispaired base
Pol III holoenzyme	Pol δ , PCNA	Repair DNA synthesis
SSB	RPA	Protecting template DNA from degradation
DNA ligase	DNA ligase I	Nick ligation

^aSSB, single-strand DNA-binding protein; ExoI, exonuclease I; PCNA, proliferating cell nuclear antigen; Pol δ , polymerase δ ; RPA, replication protein A.

^bMajor components in cells.

tumor cells that are defective in MMR have been found to be relatively resistant to alkylating agents. For example, the *MLH1*-defective HCT116 colorectal tumor cell line is resistant to killing by MNNG, but its MNNG-resistant phenotype is lost after the cell line received a wild-type copy of the *MLH1* gene by chromosome transfer. Therefore, the cytotoxicity of DNA-damaging agents depends on a functional MMR system.

MMR Proteins Promote DNA Damage-Induced Cell-Cycle Arrest and Apoptosis

Normal cells are known to undergo growth arrest at cell-cycle checkpoints when exposed to DNA-damaging agents. This response allows cells to respond to DNA damage either by repairing the damage or by committing programmed cell death (apoptosis). Interestingly, cells defective in MMR fail to arrest at crucial checkpoints after exposure to DNA-damaging agents. For example, MMR-proficient cells undergo growth arrest at the G2 phase of the cell cycle after treatment with alkylating agents, but MMR-deficient cells do not. The G2 phase arrest in MMR-proficient cells has been found to be associated with apoptosis. As illustrated in **Figure 1**, when wild-type TK6 cells and MutS α -deficient MT1 cells were treated with MNNG, apoptotic cell death was observed in TK6 cells (see the pattern of DNA fragmentation, a classical characteristic of apoptosis), but not in MT1 cells. Similarly, MutL α is also required for alkylating agent-induced apoptosis. This MMR-dependent apoptotic response can also be induced by other DNA-damaging agents, including cisplatin, certain environmental carcinogens, and ionizing radiation. In addition to cell lines, DNA-damage-induced MMR-dependent apoptosis also occurs in classic laboratory animals such as the mouse and the soil nematode *Caenorhabditis elegans*.

How is the MMR pathway involved in the apoptotic response? Although the details are far from clear, it has been established that the mismatch recognition proteins (i.e., MutS α

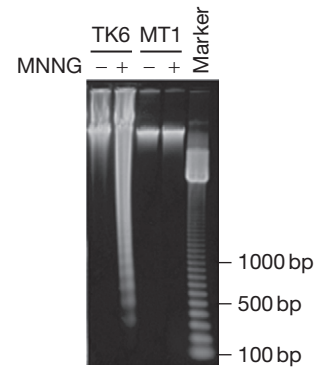


Figure 1 MNNG induces apoptosis in MMR-proficient cells.

MMR-proficient TK6 and MMR-deficient MT1 cells were treated with MNNG for 1 h and then cultured in fresh medium for 72 h before harvesting. Genomic DNAs isolated from MNNG-treated (+) and untreated cells (–) were electrophoresed in a 2% agarose gel and detected by UV-illumination in the presence of ethidium bromide. Notice that DNA fragmentation is observed in MNNG-treated TK6 cells, but not in MNNG-treated MT1 cells. This DNA fragmentation pattern is diagnostic of apoptosis, or programmed cell death.

and MutS β) recognize DNA bases damaged or modified by a variety of agents. These agents include MNNG, cisplatin, chemical carcinogens, oxidative free radicals, and ultraviolet (UV) light. These damaged DNA bases, or DNA adducts, possess structures different from regular bases, thereby causing conformational changes in the DNA duplex. For example, whereas the 8-hydroxyguanine adduct generated by ionizing radiation is approximately the same size as guanine, the guanine adduct of benzo[a]pyrene dihydrodiol epoxide, an environmental chemical carcinogen, is twice as large as guanine. However, a common feature shared by these DNA adducts, as well as DNA base–base mismatches, is that they more or less distort the structure of the DNA helix. This distortion may constitute the basis of their recognition by MutS proteins.

MMR Proteins Interact with Proteins Involved in DNA Damage Signaling

It is generally accepted that DNA damage response requires at least three types of components: damage sensors (ataxia telangiectasia, mutated (ATM), ATM and Rad3-related (ATR), and the Rad9–Rad1–Hus1 (9–1–1) complex), signal transducers (Chk1 and Chk2), and effectors (p53 and p73). Recent studies have implicated all these three components in MMR-dependent damage signaling. First, MutS α and MutL α physically interact with ATR, ATM, and the 9–1–1 complex, and these interactions activate ATR and ATM kinases to phosphorylate Chk1 or Chk2, which phosphorylates downstream components such as p53 and p73. Second, phosphorylation and stabilization of both p53 and p73 occur in an MMR-dependent manner during cellular response to MNNG- or cisplatin-induced DNA damage. These observations indicate that MMR-dependent apoptosis in response to DNA damage involves a signaling cascade, and that MMR proteins (MutS α and MutL α) are the initiator of the signaling cascade.

MMR-Mediated Apoptosis Eliminates Damaged Cells from Tumorigenesis

The molecular events involved in the MMR-dependent apoptotic response have not yet been established. However, based on what is already known, the apoptotic signaling is initiated by binding of MMR proteins to DNA adducts. Two models have been proposed to describe the role of MMR in DNA damage signaling (Figure 2). The futile cycle model (Figure 2, left) proposes that repetitive attempts by MMR to remove a DNA adduct in the template DNA strand cause cell death. DNA adducts in the template strand can pair with appropriate bases or lead to mispairs during DNA replication. MutS α , along with MutL α , recognizes these unusual base pairs as mismatches and provokes a strand-specific MMR reaction. However, because MMR is always targeted to the newly synthesized strand, adducts in the template strand cannot be removed and thus unusual base pairs re-form upon DNA resynthesis during repair. As a result, the repair cycle can be perpetually reinitiated. Such a futile repair cycle may signal cells to switch on apoptotic machinery. An alternative model, referred to as 'the direct signaling model' (Figure 2, right), suggests that the death signal could directly come from the binding of MutS α and MutL α to DNA adducts in the replication fork. These protein–DNA adduct complexes may block DNA transactions such as replication, transcription, and repair and could be recognized as a signal for cell-cycle arrest and death. In both models, the action on DNA adducts by MMR proteins activates DNA damage signaling cascade by activating ATR or ATM protein kinase and its downstream transducers and effectors, leading to apoptosis (Figure 2).

The MMR pathway is well known for its function of promoting genomic stability. The newly identified apoptotic function of MMR, however, may be as important as its replication-fidelity function for maintaining genomic stability. Normally, base excision repair and nucleotide excision repair pathways are responsible for the repair of DNA damage induced by physical and chemical agents. However, when excision repair pathways are not available or there is too much damage to be repaired, genomic DNA is in danger of accumulating a large number of mutations, a process that is tumorigenic. Eliminating these damaged cells from the body would be beneficial. It is the MMR system that initiates the elimination of these pretumorigenic cells by promoting apoptosis. The inability of the

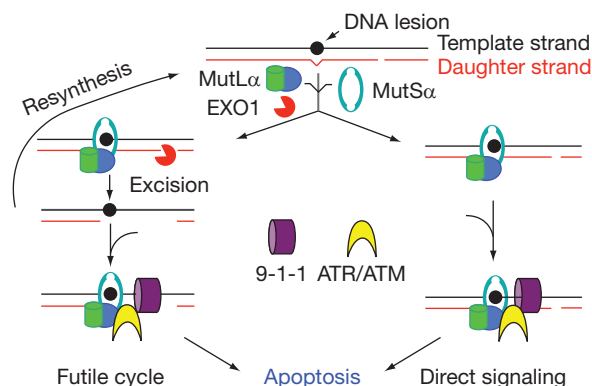


Figure 2 Proposed models of how MMR mediates apoptosis in response to DNA damage. The futile repair cycle model (left) suggests that DNA adducts (solid black circle) induce misincorporation, which triggers the strand-specific MMR reaction. Since MMR only targets the newly synthesized strand for repair, the offending adduct in the template strand cannot be removed, and will provoke a new cycle of MMR upon repair resynthesis. Such a futile repair cycle persists and activates the ATR and/or ATM damage signaling network to promote cell-cycle arrest and/or apoptosis. The direct signaling model (right) proposes that recognition of DNA adducts by the MutS α –MutL α complexes allows the proteins to recruit ATR and/or ATM to the site, activating the downstream damage signaling. Modified from Li GM (2008) Mechanism and functions of DNA mismatch repair. *Cell Research* 18: 85–98.

MMR system to commit genetically damaged cells to apoptosis may contribute to a molecular basis for cancer development.

See also: **Bioenergetics:** Cell Death by Apoptosis and Necrosis; **Molecular Biology:** Repair of G–T Mismatches by *Escherichia coli* Vsr and Eukaryotic DNA Glycosylases.

Further Reading

- Bellacosa A (2001) Functional interactions and signaling properties of mammalian DNA mismatch repair proteins. *Cell Death and Differentiation* 8: 1076–1092.
- Karran P and Bignami M (1994) DNA damage tolerance, mismatch repair and genome instability. *Bioessays* 16: 833–839.
- Li GM (1999) The role of mismatch repair in DNA damage-induced apoptosis. *Oncology Research* 11: 393–400.
- Li GM (2008) Mechanism and functions of DNA mismatch repair. *Cell Research* 18: 85–98.

DNA Mismatch Repair in Bacteria

A-L Lu, University of Maryland, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Base excision repair (BER) A DNA repair pathway that is initiated by a DNA glycosylase to remove relatively small lesions. Some glycosylases are highly specific to certain lesions or mismatches.

dam methylation The methylation of the adenine at the 6-NH₂ group in GATC sequences by *dam* methylase. Methylation on the parental strands and transient unmethylation on the newly synthesized strands provide a signal to direct the mismatch repair to the daughter strands.

Mismatch repair (MMR) A process in which the long-patch MMR removes replicative errors that are missed by DNA polymerase's proofreading activity. MMR also functions during genetic recombination and corrects certain types of DNA damage.

Mutator phenotype A phenotype that is associated with elevated genomewide spontaneous mutation frequencies and is usually caused by mutations at the DNA repair genes.

Replication fidelity The extremely low error rate of DNA replication, which is achieved by the accuracy of DNA replication and mismatch repair.

There are two major DNA mismatch-repair (MMR) systems in bacteria to correct DNA biosynthetic errors. Long-patch repair systems involve a long-patch excision and resynthesis (up to 1000 nucleotides) and are specific to a particular DNA strand. Short-patch repair systems have a repair tract shorter than 15 nucleotides, and are dictated by the nature of the mismatches. The methyl-directed MMR distinguishes the daughter strand from the parental strand by the hemi-methylated state of the DNA and enhances the fidelity of DNA replication and genetic recombination. Cells defective in MMR have several hundred-fold higher spontaneous mutation frequencies of transitions and short nucleotide insertion/deletion mutations. MMR-defective strains are resistant to some types of alkylating agents. MMR also has an anti-recombination function in limiting recombination between divergent sequences and preventing gross chromosomal rearrangement. The short-patch MutY base excision repair (BER) pathway is independent of methylation. MutY, an adenine DNA glycosylase, is responsible for the reduction of replication errors caused by oxidized guanines (8-oxoGs). MutY pathway may be coupled with DNA replication and cooperate with the mismatch repair pathway to achieve anti-mutagenic functions.

Escherichia coli Long-Patch Mismatch Repair Is Dependent on *dam* Methylation

For MMR to correct replication error, a strand-discrimination system must target the repair to the daughter DNA strand that contains the incorrect nucleotides. *Escherichia coli* MMR is dependent on DNA adenine methylation (*dam*) and is biased to the unmethylated DNA strand. The *dam* methylase modifies the adenine at the 6-NH₂ group in GATC sequences. In 1976, Wagner and Meselson proposed that transient undermethylation of GATC sites in newly synthesized DNA provides the basis to direct the MMR to the daughter strands, an idea that has been confirmed both *in vivo* and *in vitro*. As methylation is a postsynthetic process, methylation on the parental strands and transient unmethylation on the newly synthesized strands provide the

signal to direct the repair to the daughter strands. MMR can be directed by hemi-methylated sites located either 5' or 3' to the mismatch at a separation distance of up to 1 kb. Some bacteria and eukaryotes lack adenine methylation and may use strand break on the daughter strands as a discrimination mechanism. MMR that occurs during genetic recombination is also independent of DNA methylation. MMR has an anti-recombination function in limiting recombination between divergent sequences and preventing gross chromosomal rearrangement. It has been shown that MMR plays a vital role in preventing interspecies recombination between *E. coli* and *Salmonella typhimurium*.

Specificity of Mismatch Repair

Methyl-dependent MMR has a broad specificity. Of the eight possible base-base mismatches, only C-C is refractory to the repair. Heteroduplexes containing short insertion/deletion loops (IDLs) (up to five unpaired nucleotides) derived from DNA polymerase slippage can also be repaired. The efficiency of MMR is influenced by the sequence flanking the mismatch. The MMR system has also been shown to be involved in the repair of oxidative DNA damage. The major oxidative lesion, 7,8-dihydro-8-oxo-guanine (8-oxoG, G^o), can mispair with adenine during DNA replication. *E. coli* MMR can remove either adenine or G^o on the daughter strand from the A-G^o mispairs. MutS can also recognize O⁶-methyl-guanine (O⁶-meG) paired with T, U-G mispair, ultraviolet (UV) photoproducts, and cisplatin adducts. The efficiency of lesion recognition by MutS is not proportional to the repair efficiency.

Mechanism of *E. coli* Long-Patch Mismatch Repair

An *in vitro* assay developed by Lu and colleagues has provided the basis for purifying and characterizing the *E. coli* MMR proteins. The *E. coli* MMR pathway has been reconstituted *in vitro* from purified proteins in P. Modrich's laboratory. Biochemical and genetic studies have shown that *E. coli* MMR

Table 1 Functions of *E. coli* mismatch repair enzymes

Enzyme	Function
MutS	DNA mismatch/damage recognition
MutL	Molecular matchmarker; enhances mismatch recognition; stimulates MutH, UvrD, DNA polymerase III, and the β clamp activities
MutH	Strand discrimination; cleaves at unmethylated DNA strands
UvrD	DNA helicase II; unwinding DNA
ExoI, ExoX	3' to 5' exonuclease; removes nucleotides
ExoVII, RecJ	5' to 3' exonuclease; removes nucleotides
DNA Pol III	DNA resynthesis of the long-patch mismatch repair pathway
holoenzyme	A subunit of DNA Pol III holoenzyme; interacts with MutS and MutL
β sliding clamp	Protecting DNA from degradation and re-anneal
SSB	Nick ligation
DNA ligase	DNA glycosylase; excises A from A/G, A/8-oxoG, A/C, and A/5-hydroxyuracil mismatches and removes G from G/8-oxoG mismatches
MutY	Endonuclease; removes T from T-G mismatch
Vsr	Repair synthesis of short-patch mismatch repair and base excision repair pathways
DNA polymerase I	

requires 11 protein activities: MutS, MutL, MutH, UvrD (DNA helicase II), four single-strand specific exonucleases, a single-strand DNA-binding protein (SSB), DNA polymerase III (Pol III) holoenzyme, and DNA ligase (Table 1). Whereas MutH, MutL, and MutS are specific for MMR, the other proteins are involved in other DNA metabolic pathways, including replication and recombination. The mechanism of *E. coli* long-patch MMR is depicted in Figure 1.

Initiation of Mismatch Repair

MMR requires specifically MutH, MutL, and MutS proteins in the initiation steps. The first step of MMR is the recognition of mismatches by the MutS protein. MutS recognizes base–base mismatches and short IDLS. MutL serves to couple mismatch recognition by MutS to activate several downstream activities. MutH endonuclease cleaves at the 5' of G of an unmethylated GATC sequence at the transiently hemi-methylated GATC sites. This ensures that MMR is directed to the newly synthesized DNA strands.

MutS

MutS (a 95-kDa polypeptide) exists as an equilibrium of dimers and tetramers and recognizes all base–base mismatches except C–C mismatch and small IDLS up to five unpaired nucleotides. The discrimination of a mismatch-containing heteroduplex from a homoduplex by MutS protein is not great; there is only 10–20-fold preference. It has been suggested that mismatch recognition by MutS may be enhanced by MutL or by the association with replication proteins such as β -clamp. MutS is a member of the ABC transporters with adenosine triphosphatase (ATPase) activity, which is activated by DNA and is essential for the MMR function. ATP binding reduces the MutS affinity for heteroduplex DNA. Although several models have been proposed, the role of ATPase of MutS is still under

investigation. The translocation and sliding clamp models suggest that MutS moves away from the mismatch site to activate MMR. The translocation model suggests that ATP hydrolysis provides the energy and enables MutS to move like a motor protein. The sliding clamp model or molecular switch model suggests that ATP-bound MutS slides along DNA without ATP hydrolysis. There is a conformational change from adenosine diphosphate (ADP)-bound MutS to ATP-bound MutS. The third model suggests that the ATPase activity of MutS plays a kinetic proofreading role in MMR. In the presence of a mismatch, ATP hydrolysis is inhibited and MutS recruits downstream repair proteins to the sites. The co-crystal structures of MutS–DNA–ADP suggest that MutS has to bind both ATP and the mismatched DNA simultaneously to activate the repair process. Thus, the model suggests that MutS remains bound to the mismatch site and recruits MutL and MutH to initiate MMR.

The structures of truncated forms of *E. coli* and *Thermus aquaticus* MutS proteins complexed with unpaired or mispaired bases have been solved by both T. Sixma and W. Yang's groups. Although different types of mismatches were used, all MutS structures are very similar. These structures show that MutS homodimer forms an oval disk with two channels, one of which encircles the DNA at the mismatch. The bound DNA is sharply kinked by 60° toward a narrowed major groove at the mismatch. The minor groove of DNA is widened and contacted extensively by MutS. Interestingly, MutS forms an asymmetric dimer with only one monomer making direct contact with the mismatched base. MutS contacts the mismatched bases through a highly conserved N-terminal motif Phe–X–Glu. The Phe residue in this conserved motif is inserted into the double helix via the minor groove and stacks with a mismatched base. The ATP-binding site at the C terminus (domain V) of the protein is composed of residues from both subunits and is located far from the DNA-binding site. Domains II and III connect the mismatch-binding and ATPase domains. In the absence of DNA, the mismatch-binding domains (domains I and IV) are disordered in the crystal structure. The extreme C-terminus of MutS, which is not present in the structural studies, has been shown to be important for mutation avoidance, tetramer formation, and dimer stability.

MutL

MutL is a dimer of a 68-kDa polypeptide that operates as a molecular matchmaker in MMR to assemble a functional repair complex. MutL plays a key role in determining biological outcome by recruiting and/or activating downstream proteins in response to lesion recognition by MutS. In the presence of ATP, MutL binds to the MutS–heteroduplex complex and activates MutH endonuclease, UvrD helicase, DNA Pol III, and the β clamp activities. MutL can bind nonspecifically to DNA and this DNA binding is essential for MMR. Studies on a ternary complex consisting of MutS–MutL–mismatched DNA show that MutL increases the DNaseI footprint of MutS–mismatched DNA. The divergent C-terminal domain of MutL is responsible for the dimer formation in solution. The conserved N-terminal domain has weak ATPase activity that is stimulated by single-stranded DNA. The structures of both domains have been determined and both domains are predicted to be linked by a random coil. Structural and biochemical studies indicate that ATP binding and hydrolysis modulate the conformation,

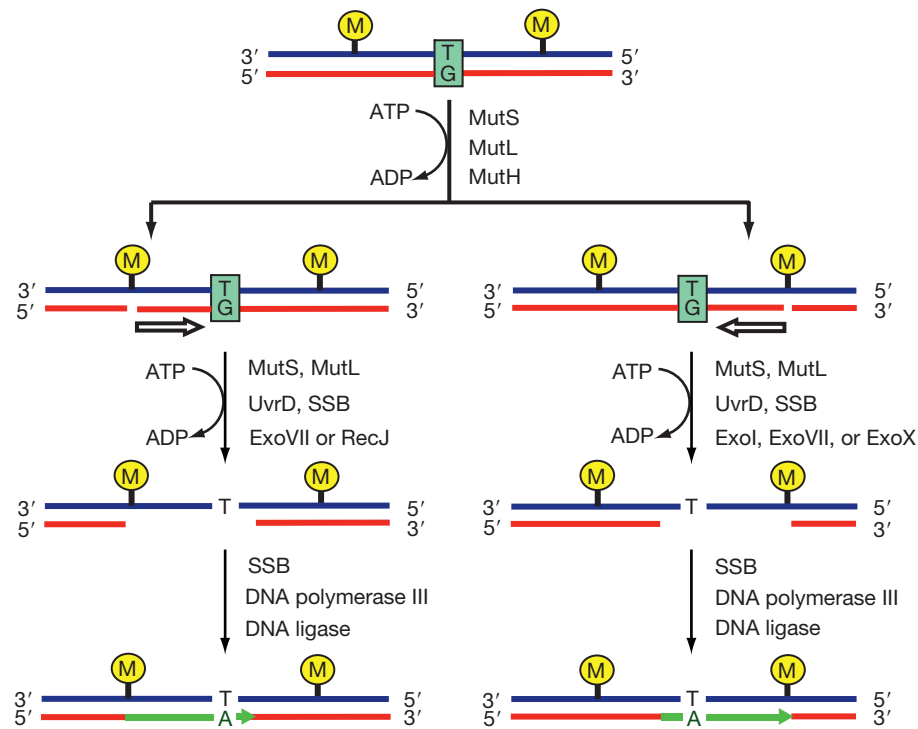


Figure 1 The long-patch methyl-dependent mismatch-repair system of *E. coli*. Mismatches (G–T, as shown) are recognized by dimeric MutS protein. After binding of MutL and MutH, the unmethylated strand in red is nicked by MutH at the hemi-methylated GATC sites (M in yellow circle represents for methyl groups). ATP hydrolysis by MutS and MutL triggers the downstream repair process. DNA unwinding by UvrD (DNA helicase II) in the presence of ATP from the nick site toward the mismatch reveals a single-stranded DNA region that is protected by SSB. Depending on the position of the nick relative to the mismatch, ExoVII or RecJ with 5' to 3' exonuclease activity and ExoI, ExoVII, or ExoX with 3' to 5' exonuclease activity degrade the nicked strand from the nick site up to and slightly past the mismatch. The resulting single-stranded gap undergoes repair resynthesis by the DNA Pol III holoenzyme and the nick is finally sealed by DNA ligase. Open arrows represent the direction of exonuclease degradation and thick green arrows show the direction of repair resynthesis.

oligomeric state of MutL, and the interactions of MutL with its partners. Upon ATP binding, the N-terminal domain undergoes conformation changes and dimerizes. In addition to function in MMR, MutL also interacts with enzymes involved in nucleotide excision repair (UvrB and UvrC), BER (MutY), and very short patch (VSP) (Vsr) pathways. C. G. Cupples proposed that stalled MutS–MutL complex may recruit these partners to the lesion sites.

MutH

The 25-kDa monomer MutH belongs to a family of type II restriction enzyme. MutH protein is absent in bacteria without *dam* methylase. MutH endonuclease activity is activated by MutS and MutL in the presence of ATP and a mismatch-containing DNA. W. Yang has proposed that MutS and MutL facilitate MutH binding to hemi-methylated DNA and overcome the inhibition of MutH by SegA (the competitor for MutH). MutH cleaves at the 5' side of the G of the GATC sequence on unmethylated strand located on either the 5' or 3' side of the mismatch. Tyr212 of MutH is important in sensing the methylation status of DNA. The structure of MutH is like a clamp with a cleft and two arms. In the complex of MutH with hemi-methylated GATC substrate, the active site is more compact and DNA cleavage is more efficient than that

of MutH complexed with unmethylated substrates. The Lys residue in the conserved DEK motif is the linchpin that couples DNA cleavage with sequence-specific recognition.

Excision and Repair Synthesis of Mismatch Repair

The strand break created by MutH at a GATC site of the unmethylated strand serves as the starting point for excision of the mispaired base. MutL interacts with and stimulates UvrD helicase, and facilitates the loading of the helicase on the nick. UvrD unwinds DNA from the nick toward the mismatch and reveals a single-stranded DNA region that is protected by SSB.

The excision of the displaced strand is then initiated from the nick by one or more of the exonucleases and is dependent on MutS, MutL, and UvrD. Depending on the position of the nick relative to the mismatch, ExoVII or RecJ with 5' to 3' exonuclease activity and ExoI, ExoVII, or ExoX with 3' to 5' exonuclease activity degrade that portion of the nicked strand displaced by the helicase up to and slightly past the mismatch (within 100 nucleotides). The resulting single-stranded gap is filled in by the DNA Pol III holoenzyme, and the nick is finally sealed by DNA ligase. MMR may be coupled with DNA replication via physical interactions of MutS and MutL with the β -clamp, the processivity factor for DNA Pol III.

Mismatch Repair Deficiency and Drug Resistance

The MMR proteins not only repair replication errors but also recognize or repair lesions that arise from DNA damage. Tolerance to methylating agents and cisplatin was observed in MMR-deficient *E. coli* and eukaryotic cells. This finding has significant impacts on drug sensitivity of cancer patients undergoing chemical therapy. *E. coli* *dam* mutant is sensitive to killing by these agents and this sensitivity is lost when *mutS* or *mutL* mutation is introduced. The mechanism of killing has been proposed by the formation of double-strand breaks generated by activated MutH on both unmethylated DNA strands. The pathway leading to cell killing in *E. coli* seems to be different from those in mammalian cells.

MutY Repair Pathway

In addition to the long-patch methyl-dependent MMR, two short-patch MMR pathways have been characterized in *E. coli*. Only the MutY repair pathway is described here and the VSP pathway initiated by Vsr endonuclease is covered elsewhere in this encyclopedia.

Oxidative damage is a major source of mutation load in living organisms. G° is one of the most stable products of oxidative DNA damage and has the most deleterious effects because it can mispair with adenine. In *E. coli*, MutT, MutM (Fpg), Nei (EndoVIII), MutS, and MutY are involved in defending against the mutagenic effects of G° lesions (Figure 2). The MutT protein eliminates 8-oxo-dGTP from the nucleotide pool with its pyrophosphohydrolase activity (Figure 2, reaction 1),

whereas the MutM glycosylase removes G° from G° -C base pairs. MutS and MutY increase replication fidelity by removing the adenines misincorporated opposite G° or G during DNA replication (Figure 2, reaction 3). Nei can excise G° from G° -C or G° -A and serve as a backup pathway to repair G° in the absence of MutM and MutY. When dGTP is incorporated opposite adenine during DNA replication, MutY repair on G° -A can cause more mutation (Figure 2, reaction 5) while G° -A repair by MutS and Nei can reduce mutation (Figure 2, reaction 6).

The *E. coli* MutY protein is a 39-kDa DNA glycosylase that removes a free base from DNA to initiate BER process. Although MutY can form a covalent Schiff base intermediate with its DNA substrates, it is controversial whether MutY has AP lyase (β -elimination) activity in addition to the DNA glycosylase activity. MutY specifically excises mispaired adenines from A-G, A- G° , A-C, and A-5-hydroxyuracil mismatches and removes guanines from G- G° mismatches. The DNA-binding affinity of MutY with G- G° is similar to that of A- G° , but the catalytic activity of MutY to G- G° is much weaker than that to A- G° . MutY binds tightly to noncatalytic DNA containing T- G° and C- G° . A- G° mismatches are particularly important biological substrates for MutY. MutY, like other DNA glycosylases, binds tightly to its DNA product containing an apurinic-apyrimidinic (AP) site. This prevents the release of toxic AP intermediates and the formation of double-strand breaks. It has been suggested that BER pathway may involve highly coordinated processes. The MutY BER pathway involves DNA PolI and DNA ligase and may involve AP endonuclease (EndoIV or ExoIII). MutY also interacts with Nei (EndoVIII) glycosylase. EndoIV, ExoIII, and Nei enhance the product

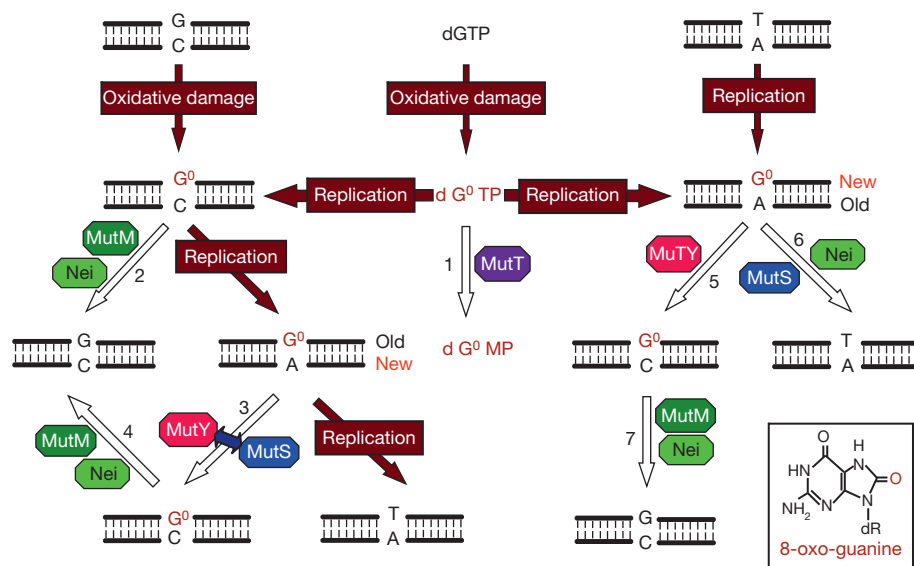


Figure 2 8-oxoG repair in *E. coli*. The MutT protein hydrolyzes 8-oxo-dGTP (d G° TP) to 8-oxo-dGMP (d G° MP) and pyrophosphate (reaction 1). The MutM glycosylase (Fpg protein) removes mutagenic 8-oxoG (G°) adducts (structure is shown in the inset) while they are paired with cytosines (reactions 2, 4, and 7). Nei can function as a backup for MutM. When C-8-oxoG is not repaired by MutM, adenines are frequently incorporated into the 8-oxoG bases during DNA replication. MutS and MutY increase replication fidelity by removing the adenines misincorporated opposite 8-oxoG (reaction 3). When d G° TP is incorporated opposite adenine during DNA replication, MutY repair on G° -A can cause more mutation (reaction 5) while G° -A repair by MutS and Nei can reduce mutation (reaction 6). MutY repair is coupled with DNA replication and MutS-dependent MMR to ensure repair fidelity. Adapted from Bai H and Lu A-L (2007) Physical and functional interactions between *Escherichia coli* MutY glycosylase and mismatch repair protein MutS. *Journal of Bacteriology* 189: 902–910, with permission from American Society for Microbiology.

release with A–G but not with A–G° substrates. Thus, it remains unclear how MutY dissociates from AP–G° *in vivo*.

The most frequent mutations observed in *mutY* mutants are G–C to T–A transversions (~50-fold higher than the wild-type cells). The *mutY* mutants also have slightly increased G–C to C–G and G–C to A–T mutations due to the defects in G–G° and A–C (or A–5-hydroxyuracil) repair, respectively. The *mutM* and *mutT* mutants have approximately 20-fold and 1000-fold higher mutation frequencies than wild-type cells, respectively. However, the mutation rate of *mutY* and *mutM* double mutants is about 1000-fold higher than the wild-type cells. By contrast, *mutY* and *mutT* double mutants have twofold lower mutation frequency than the single *mutT* mutants.

MutY glycosylase contains a [4Fe–4S] iron–sulfur cluster and it belongs to a conserved helix–hairpin–helix (HhH) family that includes several DNA glycosylases. The X-ray crystal structure of the N-terminal catalytic domain of *E. coli* D138N–MutY with bound adenine shows that adenine is buried in the active site of the catalytic domain, suggesting that the mismatched adenine must flip out of the DNA helix for the glycosylase action. The iron–sulfur cluster loop is important in substrate recognition and MutY stability, and the HhH motif is involved in binding to the phosphate backbone. The structure of MutY from *Bacillus stearothermophilus* in complex with A–G°-containing DNA was determined. Similar to several DNA glycosylases, MutY distorts the bound DNA and flips out the mismatched adenine out of the helix but the mismatched G° remains intrahelical. The C-terminal domain of MutY has been shown to play an important role in the recognition of G° lesions. The C-terminal domain of MutY has no homology to other members of the HhH superfamily, but shares structural similarities to MutT. The N-terminal catalytic domain and the C-terminal domain are connected through a flexible linker. The intact MutY encircles the bent DNA. It has been proposed that MutY scans the DNA cooperatively as a dimer or a multimeric complex to locate mismatches.

To be successful in repairing the replication errors caused by G°, MutY adenine glycosylase activity must be directed to the newly synthesized DNA strand. If MutY activity targets the adenines on the template DNA strand, A–T to C–G transversions will arise (Figure 2, reaction 5) and the repair would be mutagenic. The mechanisms that modulate the MutY glycosylase activity and allow it to target only newly synthesized DNA remain elusive. It has been suggested that coupling to DNA replication and interacting with MutS ensure that MutY BER is targeted to the daughter DNA strands but not to the parental strands. *E. coli* MutY has been shown to interact with the β -clamp and mismatch repair enzymes (MutS and MutL) (see later). A similar mechanism is also suggested in the human MutY homolog (MYH or MUTYH) repair because MYH interacts with proliferating cell nuclear antigen (PCNA), replication protein A (RPA), and MutS homologs.

Similarity and Interaction between MutS and MutY Repair Pathways

MutS-dependent MMR and MutY repair pathways share many common features. First, both pathways take place immediately after DNA replication to increase the fidelity of DNA

replication. To do so, both pathways need to distinguish daughter strands from parental strands. Both pathways may be coupled with DNA replication because MutS, MutL, and MutY interact with the β -clamp. Second, both pathways recognize base–base mismatches such as A–G and A–C. Finally, both pathways are responsible for reducing mutations caused by oxidative damage. MutY removes misincorporated adenines from A–G° while *E. coli* MMR prevents oxidative mutagenesis by removing either adenines or G° from A–G° mispairs. Wyrzykowski and Volkert have shown that when MMR-defective cells are grown anaerobically, spontaneous mutation frequencies are reduced compared with those of the same cells grown aerobically. In addition, a *dam* mutant has an increased sensitivity to hydrogen peroxide treatment. This sensitivity can be suppressed by mutations that inactivate MMR. Although the most frequent mutations observed in *mutS* mutants are A–T to G–C and G–C to A–T transitions, overexpression of MutS protein significantly decreases the rate of G–C to T–A transversions in wild-type, *mutY*, and *mutM* mutants. Thus, the MMR and MutY pathways have extensively overlapping functions.

It has been shown that *E. coli* MutS interacts with MutY by stimulating the DNA-binding activity of MutY with A–G° mismatches. Further, the expression level of MutY is upregulated in *mutS* cells as compared with the wild-type cells. Unexpectedly, the frequency of A–T to G–C mutations is reduced by two- to threefold in a *mutSmutY* mutant as compared to a single *mutS* mutant. Our data suggested a cooperation and competition model for the MutS–MutY interaction. In MutS⁺ cells, A–C mismatches generated during DNA replication are mainly repaired by MutS-dependent mismatch repair. MutS enhances MutY binding to A–G° mismatches. In *mutS* cells, MutY is overproduced to repair A–G° mismatches. However, when MutY activity is increased in the *mutS* cells to excise adenines on both daughter and parental strands, it may cause A–T to G–C transitions and A–T to C–G transversions (Figure 2, reaction 5). An alternative explanation for the decreased mutagenesis in the *mutSmutY* strain as compared to the *mutS* mutant is that MutS enhances the strand specificity of MutY. C. G. Cupples' group has also shown that MutL interacts with MutY. Overall, these findings suggest that the MutY repair pathway may cooperate with the MMR pathway to achieve anti-mutagenic functions.

See also: **Molecular Biology:** DNA Base Excision Repair Pathways; DNA Helicases: Hexameric Enzyme Action; DNA Ligases: Structures; DNA Mismatch Repair and Homologous Recombination; DNA Polymerase III, Bacterial; Nonhexameric SF1 DNA Helicases/Translocases; Repair of G–T Mismatches by *Escherichia coli* Vsr and Eukaryotic DNA Glycosylases.

Further Reading

- Bai H and Lu A-L (2007) Physical and functional interactions between *Escherichia coli* MutY glycosylase and mismatch repair protein MutS. *Journal of Bacteriology* 189: 902–910.
- Dalhous B, Laerdahl JK, Backe PH, and Bjoras M (2009) DNA base repair: Recognition and initiation of catalysis. *FEMS Microbiology Reviews* 33: 1044–1078.
- David SS, O'Shea VL, and Kundu S (2007) Base-excision repair of oxidative DNA damage. *Nature* 447: 941–950.
- Friedberg EC, Walker GC, Siede W, et al. (2005) *DNA Repair and Mutagenesis*. Washington, DC: ASM Press.

- Hsieh P and Yamane K (2008) DNA mismatch repair: Molecular mechanism, cancer, and ageing. *Mechanisms of Ageing and Development* 129: 391–407.
- Iyer RR, Pluciennik A, Burdett V, and Modrich PL (2006) DNA mismatch repair: Functions and mechanisms. *Chemical Reviews* 106: 302–323.
- Kunkel TA and Erie DA (2005) DNA mismatch repair. *Annual Review of Biochemistry* 74: 681–710.
- Li GM (2008) Mechanisms and functions of DNA mismatch repair. *Cell Research* 18: 85–98.
- Lu AL, Bai H, Shi G, and Chang DY (2006) MutY and MutY homologs (MYH) in genome maintenance. *Frontiers in Bioscience* 11: 3062–3080.
- Yang W (2008) Structure and mechanism for DNA lesion recognition. *Cell Research* 18: 184–197.

DNA Mismatch Repair in Disease and Aging

P Hsieh, National Institutes of Health, Bethesda, MD, USA

Published by Elsevier Inc.

Glossary

DNA polymerase Enzymes that synthesize DNA using one strand as a template to direct synthesis.

Epigenetic silencing The inhibition of gene expression caused by methylation of DNA sequences in the promoter region of genes.

Microsatellite instability (MSI) The accumulation of frameshift mutations in simple repeat sequences found in most eukaryotic chromosomal DNA.

Mispaired base An incorrectly pair base in DNA, for example, a G paired opposite T, or an unpaired base resulting from incorporation of an extra base during DNA replication.

MutL Bacterial mismatch repair (MMR) protein that interacts with MutS to initiate MMR; equivalent proteins that

are conserved in eukaryotic cells are MutL homologs (MLHs).

MutS Bacterial MMR protein that recognizes mismatched DNA and interacts with MutL; equivalent proteins that are conserved in eukaryotic cells are MutS homologs (MSHs).

Proliferating cell nuclear antigen (PCNA) A ring-shaped protein that keeps proteins like DNA polymerases bound to DNA.

Replication protein A (RPA) Protein complex that binds to single-strand DNA and facilitates DNA synthesis.

Tumor suppressor gene Gene whose product prevents the development of a tumor. These genes become inactivated by mutation during tumor development.

Introduction

DNA replication is normally a highly faithful process with mutations occurring at a frequency of only one in 10^9 – 10^{10} base pairs (bp) per cell division. This fidelity is maintained by three equally important processes, nucleotide selection at the active site of the DNA polymerase, the exonucleolytic proof-reading function of the polymerase, and DNA mismatch repair (MMR) functioning in a post replication fashion. The MMR pathway is highly conserved from bacteria to humans, and its loss universally confers a mutator phenotype. Early observations of a hypermutator phenotype in model microorganisms presaged the discovery in humans that mutations in MMR genes confer a heritable predisposition to colorectal cancer in which tumor cells exhibit a hypermutator phenotype. DNA MMR targets two classes of replication errors, base–base mispairs, for example, when G is misincorporated opposite T, and loops of unpaired bases that result when the template strand of the DNA slips out of alignment with the new strand that is being synthesized. If uncorrected, base–base mispairs give rise to missense mutations while loops give rise to frameshifts. Mismatches can also arise during homologous recombination, a process whereby genetic information is transferred between maternal and paternal chromosomes during meiosis, and as a result of exposure to DNA damaging agents. In this article, the consequences of loss of MMR for disease and aging are discussed.

Molecular Mechanisms of DNA Mismatch Repair

The first step in MMR is the recognition of mismatches by MutS proteins (see [Figure 1](#)). How MutS proteins find relatively rare DNA mismatches poses interesting problems as there are few chemical signatures that can serve as flags for their location. In addition, while mismatch binding by MutS proteins can be

sensitive to the DNA sequence immediately adjacent to the mismatch, so-called sequence context effects, MutS proteins find mismatches without relying on a specific DNA sequence. In prokaryotes, there is a single MutS protein that functions as a homodimer. In eukaryotes, this is the role of the MutS homologs or MSH proteins so called because they are similar in structure and function to the bacterial MutS proteins. MSH2–MSH6, referred to as MutS α , is a heterodimeric protein that recognizes base–base mispairs as well as small loops of one or a few bases. MutS β , MSH2–MSH3, preferentially targets loops. All MSH proteins are members of the adenosine triphosphate (ATP)-binding cassette (ABC) transporter superfamily of ATPases, and each heterodimeric protein has two composite nucleotide-binding sites built from amino acid residues contributed by both protein subunits. Binding and hydrolysis of adenosine triphosphate (ATP) serve to effect conformational changes that modulate both protein–DNA and protein–protein interactions. Structural and biochemical studies reveal that MutS α forms a clamp on a mismatched DNA whose interaction with the mismatch is modulated by the nucleotide cofactors ATP and adenosine diphosphate (ADP). Both MSH6 and MSH3 interact with proliferating cell nuclear antigen (PCNA), a protein required for DNA synthesis, via a conserved motif at the amino-terminus of the MSH proteins. PCNA, which forms a tripartite ring that encircles the DNA, appears to facilitate recruitment of MutS α to mismatched DNA.

A second critical participant in MMR is the MutL protein. Here again, bacteria have a homodimeric MutL protein. In eukaryotes, there are multiple MutL homologs that function as heterodimeric proteins, but the primary MMR protein in mammalian cells is a heterodimer of MLH1–PMS2 known as MutL α . All eukaryotic MutL homologs like their bacterial counterparts have an essential ATPase activity, and ATP binding and hydrolysis again play important roles in modulating the activity of MutL α . Structural studies indicate that MutL α is a very dynamic protein like MutS α , and undergoes large

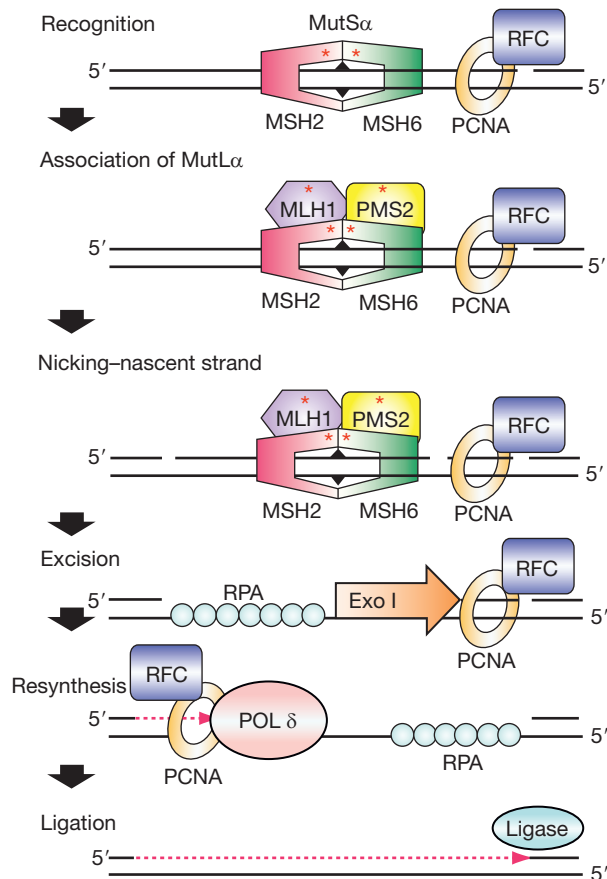


Figure 1 Cartoon Scheme for Eukaryotic DNA Mismatch Repair. Recognition of a mismatch by MutS α (MSH2–MSH6) or MutS β (MSH2–MSH3, not shown) and MutL α (MLH1–PMS2) results in the formation of a ternary complex whose protein–protein and protein–DNA interactions are modulated by ATP/ADP cofactors (indicated by red *) bound by MutS α (or MutS β) and MutL α . PCNA, a DNA replication factor, may help recruit MMR proteins to the vicinity of the advancing replication fork. Nicking by the endonuclease function of PMS2 stimulated by ATP, PCNA and RFC (symbolized by green arrows) must target the excision step to the newly synthesized strand. Excision by EXO1 in concert with PMS2 gives rise to a DNA duplex having a single-strand gap coated with RPA in which the mismatched nucleotide(s) have been excised. Resynthesis by the replicative POL δ and sealing of nicks by DNA ligase restore the integrity of the duplex. Adapted from Hsieh P and Yamane K (2008) DNA mismatch repair: Molecular mechanism, cancer, and ageing. *Mechanisms of Ageing and Development* 129: 391–407, with permission from Elsevier.

conformational changes during the course of repair. In bacteria, MutL interacts with and stimulates proteins such as DNA helicases, proteins that separate DNA strands, and exonucleases that carry out subsequent excision steps of MMR. Such protein–protein interactions are also thought to be mediated by MutL α though the exact mechanism is unknown.

The formation of a ternary complex consisting of MutS α , MutL α , and a mismatched DNA signals the start of the excision step of MMR. Details of how this ternary complex is formed and its overall architecture remain subjects of intense study. In the presence of ATP, MutS α is a sliding clamp that can diffuse away from the mismatch. *In vitro* studies of MutS and MutL bound to

mismatched DNAs suggest that multiple MMR proteins are loaded on the DNA in the vicinity of the mismatch. Although details are still missing, it is clear that both MutS α and MutL α are required to authorize the start of excision.

Excision of the mismatch in the newly synthesized DNA strand followed by gap filling by a replicative DNA polymerase and ligation complete the MMR pathway and yield an intact DNA duplex free of any replication errors. The development of *in vitro* MMR assays reconstituted with purified proteins has led to the identification of proteins required for MMR. Excision initiates at a nick in the newly synthesized DNA strand positioned on either side of the mismatch. In *Escherichia coli*, the single-strand breaks are made by the MutH endonuclease that nicks at specific methylated sequences in the *E. coli* chromosome. MutH targets the newly synthesized strand for breakage because having just been synthesized, it is temporarily missing the methylation modification unlike the fully methylated parental strand. In most other bacteria and in all eukaryotes, there is no MutH protein or corresponding DNA methylation signal to confer strand discrimination. In these organisms, MMR may utilize nicks that originate during the normal DNA replication process. For example, in the lagging strand, DNA synthesis is discontinuous leading to the formation of short Okazaki fragments that are separated by transient gaps. These gaps could serve to prime MMR. Whatever the origin of the initiating nicks, these must direct excision to the newly synthesized strand rather than the parental or template strand.

EXO1, a 5′–3′ double-strand DNA exonuclease, can utilize nicks on the 5′ side of a mismatch and is, to date, the only exonuclease that has been shown to be required for MMR. Excision from the 3′ side remained an enigma until it was discovered that PMS2, one of the two subunits of MutL α , harbors a cryptic endonuclease activity. This endonuclease activity is stimulated by the replication factor C (RFC) and PCNA, proteins that facilitate normal DNA synthesis. How PMS2 directs nicking exclusively to the newly synthesized strand is unclear at present. Thus, the combined nuclease activities of EXO1 and MutL α can explain bidirectional excision though it is still possible that other as yet unidentified proteins participate. The intermediate that is formed, a DNA duplex containing a single-strand gap coated with the single-strand DNA-binding protein RPA, is then a substrate for DNA synthesis by Pol δ aided by PCNA, a protein that improves the efficiency of DNA synthesis and RFC, a protein that loads PCNA onto DNA. In the final step, DNA ligase seals the nicks.

DNA Mismatch Repair and Cancer

The worldwide incidence of colorectal cancer is estimated at over a million cases annually. Of these, Lynch syndrome, or hereditary nonpolyposis colorectal cancer, constitutes approximately 3–5% of all colorectal cancer cases making it the most prevalent inherited colorectal cancer predisposition syndrome. Heterozygous germline mutations in the DNA MMR genes *MSH2* and *MLH1* account for the majority of Lynch syndrome families. Less common are mutations in *MSH6* and *PMS2*. Tumors result when the remaining functional copy of an MMR gene is inactivated via somatic mutation or loss of heterozygosity (LOH) giving rise to an autosomal

dominant mode of inheritance. In addition, it is estimated that approximately 15% of all spontaneous cases of colorectal cancer are characterized by microsatellite instability (MSI), a hallmark of MMR deficiency. MSI refers to contractions and expansions of regions of the genome containing simple nucleotide repeats, for example, a run of A's or repeating (AT)_n. These can occur due to frameshift mutations that accumulate in the absence of MMR as these simple repeats are prone to slippage during replication and misalignment of the template DNA strand with respect to the newly synthesized strand.

Evidence to support the idea that cancer can develop from increased mutation rates due to uncorrected replication errors stems from several findings. First, mutations in MMR genes cause Lynch syndrome. Second, approximately 15% of tumors in the case of sporadic colorectal cancer exhibit MSI at mono-, di-, tri-, and, more rarely, tetra-nucleotide repeats, and these tumors are associated with epigenetic silencing of *MLH1*, the gene encoding an MutL protein. Such epigenetic silencing results from hypermethylation of DNA bases in the promoter region of *MLH1*. Third, laboratory mice harboring disruptions of MMR genes exhibit MSI and a marked predisposition to cancer. How do high mutation rates contribute to cancer? In individuals who have lost MMR activity, genes that normally inhibit tumorigenesis, for example, tumor suppressor genes, may be inactivated, particularly if they harbor simple nucleotide repeats that are prone to contraction or expansion. Other genes, oncogenes, that are normally turned off in healthy tissue, may become activated or acquire a tumor-promoting function as a result of mutation.

Clinical studies of families with frequent diagnoses in multiple generations of colorectal cancer without polyposis as well as gastric and uterine cancer led to agreement on a set of criteria for diagnosing hereditary nonpolyposis colorectal cancer (HNPCC) known as the Amsterdam Criteria and Amsterdam Criteria II. Subsequently, the Bethesda Guidelines were developed to aid in identifying colorectal cancer (CRC) patients who should be tested. Molecular genetics studies of affected families led to the identification of cancer susceptibility loci mapping to specific regions on human chromosomes 2 and 3. The discovery of specific defects in MMR genes was achieved when it was appreciated that tumor cells from Lynch syndrome patients exhibited the same kind of DNA MSI as that found in microorganisms like *E. coli* and budding yeast that were missing MMR. Thus, germline mutations first in *MSH2* and *MLH1* and then in *MSH6* and *PMS2* were discovered.

The diagnosis of Lynch syndrome requires careful attention to family histories and documentation of all cancers in family members. Although the syndrome is defined by mutations in MMR genes, direct determination of inactivating mutations in MMR genes is not routinely done in the colorectal cancer population at large. Clinically, there is no specific list of phenotypic features, but notable features include the frequent occurrence of colorectal cancer in families, an autosomal dominant inheritance pattern, a relatively early age of onset, on average 45 years of age compared to the mid-60s in the normal population, a tendency toward proximal (right-sided) colonic cancer, and strikingly, the presence of a variety of extracolonic tumors including heightened risk of endometrial and ovarian cancer in females (see Table 1). The pathology of colorectal cancer is oftentimes poorly differentiated with an excess of

Table 1 Cardinal features of Lynch syndrome (LS)

Autosomal dominant inheritance pattern seen for syndrome cancers in the family pedigree
Earlier average age of colorectal cancer (CRC) onset than in the general population
Average age of 45 years in LS vs. 69 years in the general population
● Proximal (right-sided) colonic cancer predilection
<70% of LS CRCs are proximal to the splenic flexure
Accelerated carcinogenesis (tiny adenomas can develop into carcinomas more quickly)
Within 2–3 years in LS vs. 8–10 years in the general population
High risk of additional CRCs
25–30% of patients having surgery for an LS-associated CRC will have a second primary CRC within 10 years of surgical resection if the surgery was less than a subtotal colectomy
Increased risk for malignancy at certain extracolonic sites ^a
Endometrium (40–60% lifetime risk for female mutation carriers)
Ovary (12–15% lifetime risk for female mutation carriers)
Stomach (higher risk in families indigenous to the Orient, reason unknown at this time)
Small bowel
Hepatobiliary tract
Pancreas
Upper uro-epithelial tract (transitional cell carcinoma of the ureter and renal pelvis), especially in males with LS-MSH2 type ^b
Brain (in the Turcot's syndrome variant of the LS)
Multiple sebaceous adenomas, sebaceous carcinomas, and keratoacanthomas in the Muir-Torre syndrome variant of LS
Pathology of CRCs is more often poorly differentiated, with an excess of mucoid and signet-cell features, a Crohn's-like reaction, and a significant excess of tumor-infiltrating lymphocytes within the tumor
Increased survival from CRC
The sine qua non for diagnosis is the identification of a germline mutation in a mismatch repair gene (<i>MLH1</i> , <i>MSH2</i> , <i>MSH6</i> , or <i>PMS2</i>) that segregates in the family: that is, members who carry the mutation show a much higher rate of syndrome-related cancers than those who do not carry the mutation

^aWatson P, Vasen HFA, Mecklin J-P, et al. (2008) The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome. *International Journal of Cancer* 123: 444–449; and Barrow E, Robinson L, Alduaij W, et al. (2009) Cumulative lifetime incidence of extracolonic cancers in Lynch syndrome: A report of 121 families with proven mutations. *Clinical Genetics* 75: 141–149.

^bWatson P, Vasen HFA, Mecklin J-P, et al. (2008) The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome. *International Journal of Cancer* 123: 444–449.

Adapted from Lynch HT, Lynch PM, Lanspa SJ, et al. (2009) Review of the Lynch syndrome: History, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clinical Genetics* 76: 1–18, with permission from John Wiley and Sons.

infiltrating lymphocytes within the tumor. Individuals with Lynch syndrome are at very high risk for developing colorectal cancer necessitating intensive surveillance including regular colonoscopies starting in the mid-20s.

As discussed earlier, instability at microsatellite sequences is a hallmark for the loss of MMR. Currently, most Lynch syndrome candidates with suggestive family histories undergo MSI testing according to the Revised Bethesda Guidelines. These guidelines include (1) a diagnosis of colorectal cancer in individuals <50 years of age; (2) the presence of synchronous or metachronous colorectal, or other syndrome-associated tumors regardless of age; (3) colorectal cancer with microsatellite instability-high (MSI-H) diagnosed in a patient who is <60 years

of age; (4) colorectal cancer or syndrome-associated tumor diagnosed under age 50 years in at least one first-degree relative; and (5) colorectal cancer or syndrome-associated tumor diagnosed at any age in two first- or second-degree relatives. Syndrome-associated tumors include sebaceous adenomas, sebaceous carcinomas, and multiple keratoacanthomas associated with Muir-Torre syndrome and primary brain tumors associated with Turcot syndrome.

Following the Bethesda Guidelines, a panel of five microsatellite markers is surveyed, three with dinucleotide repeats and two with mononucleotide repeats. When two or more of these five show a change in the number of repeats in tumor tissues compared with normal tissues, this is classified as MSI-H. Approximately 20–25% of these tumor cells have mutated MMR genes. Alteration at only one of the five

microsatellite sequences is termed MSI-low and is not associated with germline mutations in MMR genes. Immunohistochemistry can be used to detect the levels of MMR proteins in tumor cells using antibodies to MSH2, MSH6, MLH1, or PMS2. In the case of MSH2, loss is usually via deletion and immunohistochemistry can readily reveal its absence. Loss of MSH2 protein also reduces the amount of MSH6 protein as this subunit appears to be unstable without MSH2. In contrast to MSH2, most Lynch syndrome mutations in *MLH1* are missense mutations in which the protein is produced, but not functional. In this case, immunohistochemistry can be uninformative. For those individuals who are MSI-H, germline mutation testing is recommended. Recent advances in technology make DNA sequencing for direct detection of germline mutations and DNA methylation much more feasible. Efforts

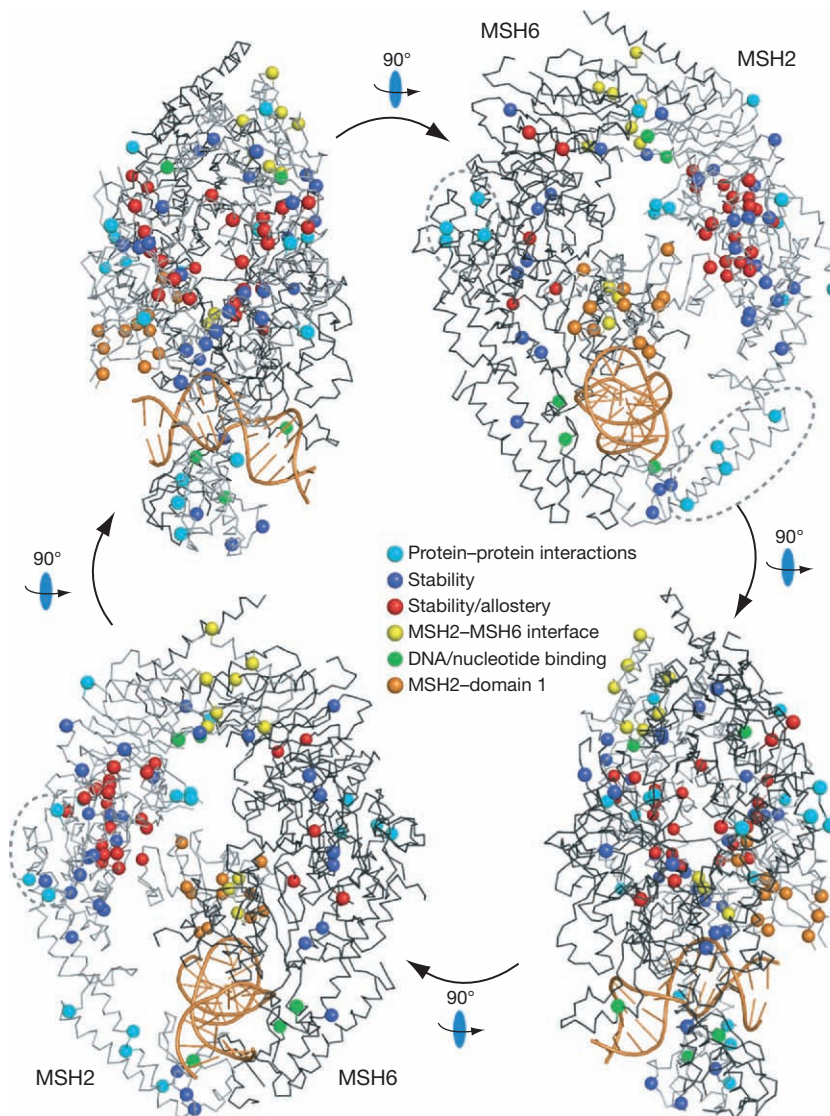


Figure 2 Structural model for human MutS α with HNPCC mutations. Four views of MutS α related by 90° rotations as indicated, with positions of HNPCC missense mutations indicated by spheres. Hypothetical functional classification of mutations is indicated by sphere color (see legend). MSH2 and MSH6 are shown as light and dark gray C α chain traces, respectively, and the DNA is colored orange. Three clusters of surface mutations, which may correspond to sites of protein-protein interactions are indicated with dashed ovals. Reproduced from Warren JJ, Pohlhaus TJ, Changela A, et al. (2007) Structure of the human mutS α DNA lesion recognition complex. *Molecular Cell* 26: 579–592, with permission from Elsevier.

are underway to improve diagnostic sequencing with regard to both sensitivity and cost.

Individuals with tumor cells classified as MSI-H tend to have better clinical prognoses overall and are less likely to benefit from adjuvant chemotherapy. Why MSI-H is associated with a better outcome is unclear, but the poor response to adjuvant chemotherapy may reflect the role of MMR proteins in the cellular response to DNA damage. A subset of DNA-damaging agents commonly used in chemotherapy are incorporated into DNA or chemically modify DNA bases and result in cell killing. In this way, tumor cells that are rapidly dividing are targeted for elimination. Examples of cytotoxic DNA modifying drugs include Temozolomide that adds a methyl group to DNA bases and flurouracil that is metabolized to fluorodeoxyuridine and incorporated into DNA. The cytotoxicity of these drugs requires functional MutS α and MutL α proteins. The noncanonical bases resulting from drug treatment cause the DNA polymerase to make errors during DNA synthesis leading to the formation of mispairs that are recognized by MutS α . Subsequently, cells arrest at strategic points in the cell cycle and activate apoptotic or programmed cell death pathways. Understanding how this happens is an area of active research. Since MSI-H cells with inactivating mutations in MSH2 or MLH1 lack MutS α or MutL α , the cytotoxic effects of these chemotherapeutic drugs may be compromised. A related issue in chemotherapy is the observation that tumors oftentimes become tolerant to treatment. The chemotherapy regimen exerts selective pressure on cells that are missing MMR. Examination of these tumor cells reveals that expression of the *MLH1* gene has been silenced by hypermethylation.

What regions of MutS α are important for its tumor suppressor function? X-ray crystallographic studies of bacterial and human MutS and MutL proteins have revealed the three-dimensional (3D) structure of these proteins with atomic resolution. One can map mutations found in Lynch syndrome/HNPCC patients that change only one amino acid residue onto representations of the 3D structure of MMR proteins. These are shown in [Figure 2](#) for human MutS α (MSH2–MSH6) bound to a mismatched DNA. It is immediately obvious that mutations that lead to cancer map all over the MutS α molecule, in all the structural domains, in regions that are at the surface, and in those that are buried. Based on structure–function studies of MMR proteins, one can surmise that HNPCC mutations can disrupt many different aspects of MMR. They can interfere with critical protein–protein interactions, alter the stability of the proteins, disrupt ATP binding or hydrolysis or interfere with DNA binding. Disease-causing mutations are invaluable in helping investigators to understand how proteins carry out their normal function and what happens when that function is disrupted. In the best scenarios, they can also direct efforts at finding new treatments.

Other Aspects of MMR in Disease and Aging

DNA MMR proteins have important roles in a number of cellular functions. It is not surprising, therefore, that inactivation or loss of these proteins has important implications for health. In the examples that follow, it is important to note that unlike Lynch syndrome where mutations in MMR genes

resulting in the loss of MMR have unequivocally been established as an underlying cause of cancer, the roles of MMR in other disease processes is inferential and has been studied almost exclusively in model experimental systems such as yeast, mouse models, and cultured human cells. These studies reveal novel functions of subsets of MMR and suggest that MMR proteins can, in certain contexts, assume functions that are quite distinct from post replication MMR.

A special case of DNA repeat instability involving repeats of specific DNA triplets causes human neurological and muscular inherited diseases such as Huntington's disease and myotonic dystrophy. The most common trinucleotide repeats implicated in disease are CAG, CTG, CCG, and GAA triplets. In unaffected individuals, the number of triplet repeats is stable. However, in affected individuals, the repeat is unstable and expands. The expanded sequence can disrupt protein or RNA expression and/or function leading to disease. In studies utilizing a mouse model having long CTG repeats, MSH2, MSH3, and PMS2, but not MSH6, appear to have an ameliorating effect on triplet expansion. The underlying mechanism is unclear at present though it is certain to involve many other processes. Work proceeds to understand repeat instability and the evolutionary importance of the repeats themselves with the goal of eventually finding therapeutic approaches.

Eukaryotic versions of bacterial MutS and MutL proteins have diverged and some now function exclusively in meiosis, a key step in sexual reproduction. The product of meiosis is haploid gametes that have only one of the two sets of chromosomes found in diploid cells. Fertilization involving male and female gametes restores the full chromosome complement. These gametes are formed in a specialized program of cell divisions that exactly halves the number of chromosomes in daughter cells. In mice that have been genetically ablated for two meiosis-specific MutS homologs, *MSH4* or *MSH5*, or for the MutL homologs, *MLH1* or *MLH3*, meiosis is arrested, and the animals are by and large infertile. Thus, the MSH4–MSH5 and MLH1–MLH3 proteins are required to carry out meiosis and appear to ensure the proper alignment and separation of paired chromosomes so that the chromosome number can be reduced by exactly half. Since these MMR proteins are conserved in humans as well, it is quite likely that they have important roles in human sexual reproduction though that remains to be confirmed.

Some MMR proteins also have functions in the generation of immuno-globulins or antibodies that function in the immune response. In most mammals, this process occurs in specialized lymphocytes in bone marrow called B cells. Paradoxically, in B cells the MMR proteins participate in a hypermutation process that stands in marked contrast to post replication repair where MMR proteins serve to keep mutations in check. Antibodies must recognize and bind a vast array of antigens and carry out different immune functions. Vertebrates create antibody diversity by combining different short protein segments in a combinatorial process of immunoglobulin assembly, but they also require a hypermutation phase. Studies of immunoglobulins in mice missing different MMR proteins indicate that MSH2–MSH6 and EXO1 are important for somatic hypermutation, and mice missing MSH2, MSH6, or EXO1, and to a lesser extent, MLH1 and PMS2, exhibit reduced repertoires of different immunoglobulin. In all these cases, the bulk of the evidence suggests

that MMR proteins are functioning in alternative pathways to canonical post replication MMR.

The accumulation of DNA damage and resulting genomic instability not only cause cancer but are thought to be contributors to aging as well. Does loss of MMR contribute to aging? This question is difficult to address as neither Lynch syndrome patients nor MMR-deficient mice show overt signs of premature ageing. However, some studies of mouse models suggest that the capacity for MMR diminishes with age, and MMR can process certain types of oxidative DNA damage that are thought to accumulate with age. In addition, data from budding yeast and mice suggests that loss of MLH1 and/or PMS2 can increase cell proliferation in special cases. The role of DNA repair in aging remains an active area of research.

Summary

DNA MMR plays a critical role in preserving the integrity of the genome in virtually all organisms. Loss of MMR proteins MSH2, MLH1, PMS2, and MSH6 lead to greatly elevated rates of mutation, and, in humans, a genetic predisposition to cancer. MMR proteins also modulate responses to DNA damaging agents that are common chemotherapeutic drugs. Studies of model organisms point to important roles for subsets of MMR proteins in trinucleotide repeat expansion diseases, sexual reproduction, and immunoglobulin diversity, and the role of MMR in the aging process remains unclear. Much work remains to understand how the MMR proteins function in so many diverse processes and to find new therapeutic avenues.

See also: **Molecular Biology:** DNA Mismatch Repair and Homologous Recombination; DNA Mismatch Repair and the DNA

Damage Response; DNA Mismatch Repair in Bacteria; DNA Mismatch Repair in Mammals; Eukaryotic DNA Polymerase δ ; Repair of G–T Mismatches by *Escherichia coli* Vsr and Eukaryotic DNA Glycosylases.

Further Reading

- Boland CR, Koi M, Chang DK, and Carethers JM (2008) The biochemical basis of microsatellite instability and abnormal immunohistochemistry and clinical behavior in Lynch syndrome: From bench to bedside. *Familial Cancer* 7: 41–52.
- Fishel R (2001) The selection of mismatch repair defects in hereditary nonpolyposis colorectal cancer: Revising the mutator hypothesis. *Cancer Research* 61: 7369–7374.
- Hsieh P and Yamane K (2008) DNA mismatch repair: Molecular mechanism, cancer, and ageing. *Mechanisms of Ageing and Development* 129: 391–407.
- Iyer RR, Pluciennik A, Burdett V, and Modrich PL (2006) DNA mismatch repair: Functions and mechanisms. *Chemical Reviews* 106: 302–323.
- Jiricny J (2006) The multifaceted mismatch-repair system. *Nature Reviews* 7: 335–346.
- Kunkel TA and Erie DA (2005) DNA mismatch repair. *Annual Review of Biochemistry* 74: 681–710.
- Li GM (2008) Mechanisms and functions of DNA mismatch repair. *Cell Research* 18: 85–98.
- Lynch HT, Lynch PM, Lanspa SJ, et al. (2009) Review of the Lynch syndrome: History, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clinical Genetics* 76: 1–18.
- Mueller J, Gazzoli I, Bandipalliam P, et al. (2009) Comprehensive molecular analysis of mismatch repair gene defects in suspected Lynch syndrome (hereditary nonpolyposis colorectal cancer) cases. *Cancer Research* 69: 7053–7061.
- Taketo MM and Edelman W (2009) Mouse models of colon cancer. *Gastroenterology* 136: 780–798.

Relevant Website

<http://www.insight-group.org> – InSight database.

DNA Mismatch Repair in Mammals

A J Plys and E Alani, Cornell University, Ithaca, NY, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by James T. Drummond, volume 1, pp. 687–690, © 2004, Elsevier Inc.

Glossary

ATP (adenosine triphosphate) A nucleotide that serves as a source of energy for physiological reactions and as a protein cofactor for a variety of processes.

ATPase Enzymatic activity associated with protein molecules and accompanied domains that hydrolyze ATP into ADP, releasing inorganic phosphate.

DNA polymerase Enzymatic activity that catalyzes the synthesis of DNA using a template DNA molecule, a primer, and deoxyribonucleotides.

DNA repair Reconstruction of an intact double-stranded DNA molecule free of damage or mismatches by a variety of enzymatic processes.

DNA replication The process of DNA synthesis that creates an identical copy from a template DNA molecule.

Endonuclease Enzymatic activity that cleaves the phosphate backbone of single- or double-stranded DNA at interior positions within a polymer.

Excision DNA repair process that removes part of a damaged DNA strand generally through the directed activity of exonucleases.

Exonuclease Enzymatic activity that removes individual nucleotides from the 3' or 5' ends of single- or double-stranded DNA polymers.

Indel mutations Insertion/deletion loops that arise from slippages that occur when DNA polymerase is replicating repetitive sequences in the genome.

Mismatch DNA base interactions that do not follow the classical Watson–Crick G–C or A–T pairing rules.

Mutation A change in a DNA sequence that can occur from a variety of unrepaired endogenous or exogenous insults.

Introduction

Mismatch repair (MMR) is an evolutionarily conserved pathway comprising components that work in concert to recognize DNA polymerase-mediated misincorporation errors and facilitate their excision. These misincorporations are primarily base–base mismatches (e.g., G–T, C–A) or insertion/deletion loops (indels) that arise from slippages that occur when DNA polymerase is replicating repetitive sequences in the genome (Figure 1). The major MMR components, Mut proteins, were identified several decades ago in the bacterium *Escherichia coli* in genetic screens for mutations that increased the cellular mutation rate. Defects in *mut* genes result in an increased rate of base substitution and indel mutations (~1000-fold), as well as increased genetic recombination between divergent DNA sequences and gene amplifications. Mutant homologs of *mut* genes in mammals also have defects in cell-mediated death and immune-system functions. In humans, the high frequency of indel mutations seen in tumors obtained from a subset of patients suffering from inherited forms of nonpolyposis colorectal cancer (2–6% of all colon cancers), was a strong hint that these individuals suffered from defects in MMR. This characteristic resulted in the identification of mutations in at least four MMR genes that are linked to HNPCC. Loss of MMR activity has also been linked to resistance to chemotherapeutic agents (Figure 1).

Mammals contain multiple homologs of the *E. coli* *mut* genes. MutS homologs (MSH family), of which there are five members (MSH2–MSH6), act as heterodimers (Figure 2). MutS homolog proteins initiate the MMR pathway by recognizing mismatches. The primary complexes are MSH2–MSH6 (MutS α) and MSH2–MSH3 (MutS β). These complexes have

partially overlapping roles in recognizing mismatch substrates. MutS α is primarily involved in repairing base–base mismatches and small indel mutations. MutS β acts primarily to repair indels that are up to 17 nucleotides in length. MSH4 and MSH5 also form a heterodimeric complex that has no apparent role in MMR but acts to promote crossing over in meiosis. The lower eukaryote baker's yeast contains an MSH1 protein involved in mitochondrial genome stability; however, no mammalian homolog of MSH1 has been identified. Similarly, there are multiple MutL homologs (MLH family), MLH1, MLH3, PMS1, and PMS2, that function as heterodimers. The function of MutL homologs is to bridge mismatch recognition with the downstream repair steps of MMR. MLH1–PMS2 (MutL α) and MLH1–MLH3 (MutL γ) also have partially overlapping roles in initiating downstream repair from MMR targets. MutL α is the predominant complex in repairing mismatches and indels while MutL γ participates in only a subset of indel repair and in promoting crossing over in meiosis. MLH1 and PMS1 also form a heterodimeric complex; however, the function of this complex in mammals is less clear. Lastly, an MLH2 homolog has been identified in baker's yeast that has been implicated in the repair of frameshift mutations; however, an equivalent factor has not been identified in mammals. The downstream components of the mammalian MMR pathway do not belong to the Mut homolog families and are not MMR specific.

A key breakthrough in understanding the mechanistic details of the MMR pathway came from the development of a reconstituted MMR system from purified *E. coli* components. The substrate for this assay was a hemi-methylated plasmid containing a single mismatch; *in vitro* repair resulted in excision of the mismatch on the unmethylated strand (see later). This achievement led to a detailed characterization of individual bacterial

MMR components, and has served as the basis for characterizing key mammalian MMR factors.

In *E. coli*, MMR is initiated by a MutS homodimer binding to a DNA mismatch (Figure 2). Interactions between MutS, MutL, and mismatch DNA result in the removal of the mismatch through an excision mechanism. Excision is initiated by activation of the MutH endonuclease in steps that require MutS binding to mismatch DNA, MutL interaction with the MutS-mismatch complex, and ATP. MutH, a single-strand DNA cleaving enzyme that acts at unmethylated GATC sites, plays a critical role in a strand-discrimination mechanism that removes the mismatch on the newly replicated strand. Strand discrimination is accomplished by a MutH cleavage activity that is inhibited by Dam, a protein that specifically methylates adenine residues at GATC sites. Methylation of newly replicated DNA by Dam occurs within a short period after replication fork passage. This postreplication function of Dam establishes a strand-discrimination signal because MutH will not cleave methylated GATC sites but will cleave the unmethylated strand of hemi-methylated GATC sites. Thus, DNA mismatches that are recognized shortly after DNA replication fork passage are directed for removal of the newly replicated strand

due to MutH-directed cleavage of hemi-methylated GATC sites. MutH incision at hemi-methylated GATC sites can occur up to several thousand base pairs from the mismatch, with repair efficiency decreasing with distance.

MutH incision creates an entry site for UvrD, a helicase that can unwind duplex DNA beginning at a nick site with the aid of single-stranded DNA-binding protein (SSB). The MutS–MutL mismatch complex provides signals to load UvrD and unwind DNA toward the mismatch. The unwound single-stranded DNA containing the nicked end is acted upon by at least four exonucleases (ExoI, ExoVII, ExoX, and RecJ). These exonucleases display different polarities (5'–3' or 3'–5') for excision. Which nuclease acts depends on the orientation of the mismatch relative to which MutH cleavage site is chosen. Thus, the MMR system is capable of bidirectional repair (5'– or 3'–directed excision) and is able to sense the location of the mismatch to direct appropriate excision. The resulting gap created by excision is filled in by DNA polymerase III with the help of the replication-processivity factor β -clamp. DNA ligase closes the nick to complete repair (Figure 3).

Mammalian MMR shares aspects of the *E. coli* MMR mechanism with some important differences. Like in the *E. coli* system, the development of a cell-free system to study MMR has been critical in terms of identifying most of the key components. The main caveat, discussed here, is the fact that the cell-free system has not been able to address how strand discrimination is set up, other than showing that a preexisting nick on a mismatch plasmid substrate is sufficient to direct strand-specific repair.

As described earlier, mammalian MSH and MLH factors act as heterodimers to carry out functions analogous to those observed for the *E. coli* proteins. However, there are at least three mechanistic differences between the bacterial and mammalian MMR pathways:

1. A methylation system that distinguishes the template and newly replicated DNA strands does not appear to function in mammals and bonafide MutH homologs appears to be missing in eukaryotes and in most bacteria. Thus, mammalian MMR uses a different mechanism for strand discrimination. Models that outline how the mammalian MMR

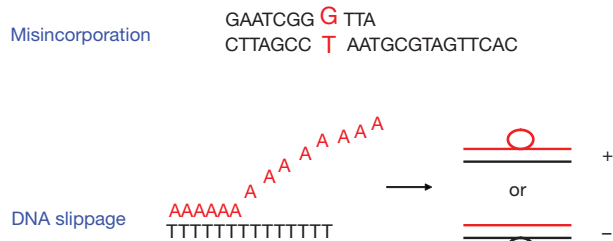


Figure 1 Mutations seen in mismatch repair (MMR)-defective strains. Misincorporation mutations arise from DNA polymerase errors that create non-Watson–Crick base pairs. DNA slippage mutations commonly occur on repetitive DNA elements by DNA strands reannealing out of register. This creates an insertion (+) if the loop is on the newly replicated strand or a deletion (–) if on the template strand.

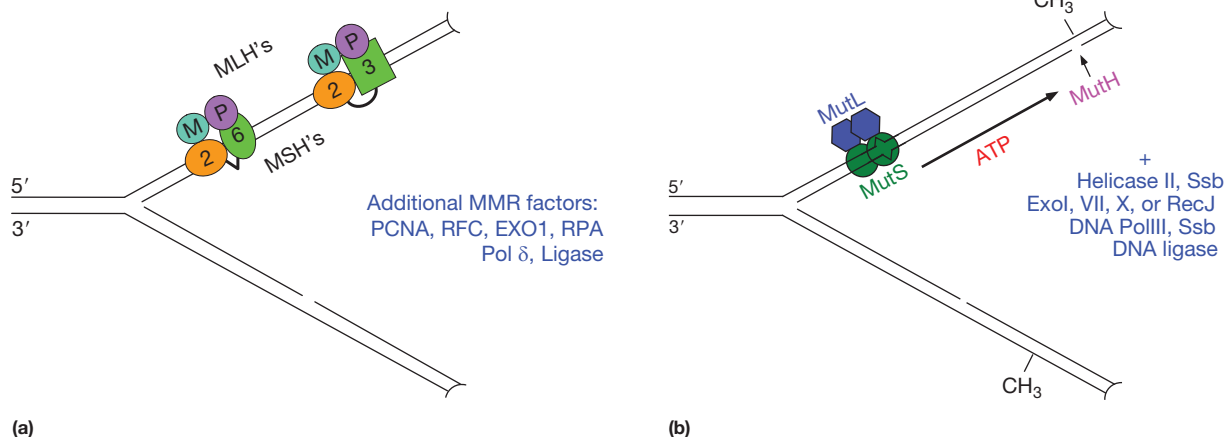


Figure 2 MMR components. (a) Mammalian mismatch repair. M = MLH1, P = PMS2, 2 = MSH2, 3 = MSH3, 6 = MSH6. A base–base mismatch is shown by the triangle and an indel created by DNA slippage is indicated by a half circle. Additional factors in mammalian MMR are shown. (b) Model replication fork from *E. coli*. A base–base mismatch is indicated by a diamond. Methylation by Dam is indicated by –CH₃.

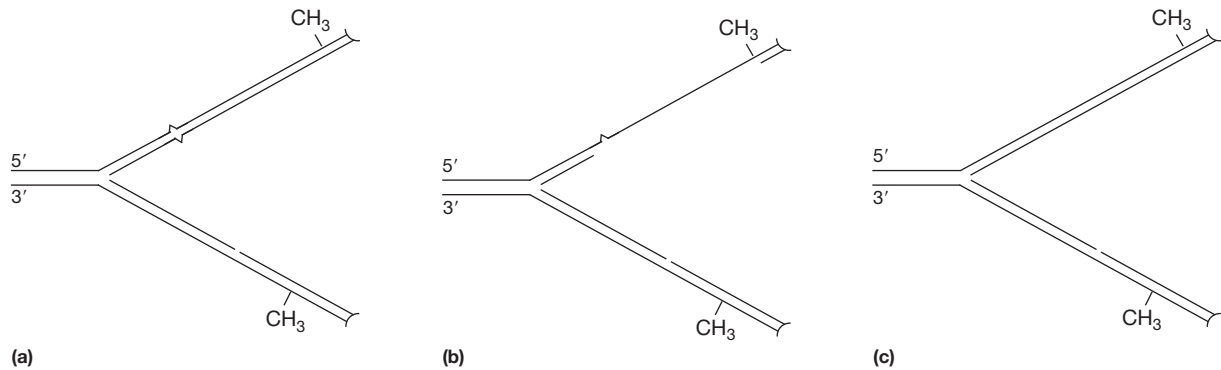


Figure 3 Basic steps in MMR. (a) DNA mismatch is created by a DNA replication error. (b) Excision of the mismatch on the newly replicated strand creates a gap. (c) Resynthesis of the gap completes repair.

pathway distinguishes template from daughter strands have been proposed and will be discussed later.

2. A helicase activity analogous to UvrD does not appear to be required in cell-free mammalian MMR reactions.
3. Exonucleolytic activities that excise the newly replicated strand appear to work differently in mammalian MMR. So far, only a single exonuclease, EXO1, has been identified, and this protein is a 5' to 3' exonuclease that acts primarily on double-stranded DNA. The fact that EXO1 can act on double-stranded DNA may preclude the need for a helicase activity.

Mismatch Search and Recognition

MutS, the Mismatch Recognition Complex

Crystallographic analysis revealed that the MutS homodimer binds to mismatch DNA in an asymmetric fashion. In the absence of DNA, MutS contains disordered DNA-binding domains. Upon mismatch binding, these domains undergo extensive local folding to form more ordered structures. Each MutS subunit contains two DNA-binding domains; however, only one domain of one subunit directly interacts with a mismatched base while a DNA-binding domain on the other subunit makes nonspecific interactions to stabilize the DNA backbone. The overall structure of MutS resembles a clamp that encircles the DNA backbone. Studies on the crystal structures of the human MutS α complex have shown similarities with bacterial MutS in terms of domain organization and DNA interactions. In *E. coli* MutS, mismatch recognition is coordinated by two amino acids, phenylalanine 36 and glutamic acid 38. In mammals, MSH6, but not MSH2 or MSH3, contains these two key residues and is the subunit responsible for direct interactions with the mismatch. In both MutS and MSH6, the mismatch base stacks with the phenylalanine side chain and hydrogen bonds with the glutamic acid side chain. Moreover, the DNA bound to MutS in the crystal structure is in a bent state that is facilitated by these interactions and with nonspecific interactions between the other DNA-binding domain and the DNA backbone. Recent atomic force microscopy studies have suggested that MutS undergoes a conformational change after binding, which results in the formation of unbent DNA containing the mismatch. This type of manipulation may allow MutS to sample DNA for mismatches in a manner that increases target-recognition specificity.

MSH complexes must identify rare misincorporation errors among a vast excess of undamaged DNA. Biochemical and single-molecule studies have shown that MSH complexes can slide along DNA by facilitated one-dimensional diffusion; such a mechanism is likely to promote efficient identification of targets because the search is confined to scanning along DNA rather than requiring random collisions within the three-dimensional space of the nucleus. MSH complexes have also been shown to interact with components of the DNA replication machinery. This feature, coupled with their ability to diffuse along DNA, suggests that MSH complexes have immediate access to DNA polymerase-mediated misincorporation errors.

ATP binding and hydrolysis by the MSH proteins appear critical for coordinating mismatch recognition with the commitment to undergo excision repair. The MSH proteins are members of the ABC transporter adenosine triphosphatase (ATPase) family and they weakly hydrolyze adenosine triphosphate (ATP). Nucleotide binding results in conformational changes in MSH proteins. The nucleotide-binding sites of MSH proteins are found at their dimerization interface. MSH2 and MSH6 in MutS α show different affinities for adenine nucleotides. MSH6 has a higher affinity for ATP whereas MSH2 has a higher affinity for ADP. Nucleotide-binding asymmetry is also seen in the *E. coli* MutS homodimer. Such asymmetries in ATP binding by MSH subunits are considered to be important to induce coordinated conformational changes in MSH-mismatch DNA complexes that signal downstream repair factors. For example, in the eukaryotic MutS α (MSH2–MSH6) complex, upon mismatch binding, MSH6 ATP hydrolysis is inhibited and ATP binding by MSH2 is favored. This ADP $\rightarrow$ ATP nucleotide exchange allows MutS α to enter a sliding-clamp diffusion mode. Implications for a MutS α sliding-clamp diffusion mechanism in searching out the unknown strand-discrimination signal is discussed later (Figure 4). ATP is also required for recruitment of MLH proteins to mismatch–MSH complexes to form a ternary complex for the transmission of the mismatch-recognition signal to the downstream repair factors.

MutL Bridges Mismatch Recognition and Excision

As outlined above, the MLH family of proteins plays critical molecular matchmaking roles by signaling MSH-dependent mismatch recognition to downstream repair factors. Crystallographic structures have been reported for individual N- and

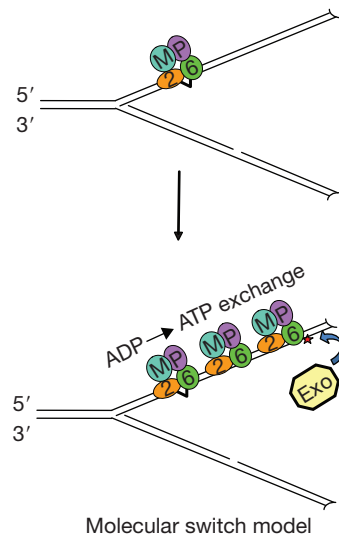


Figure 4 Model for signaling downstream repair. The molecular switch model proposes that after mismatch recognition, MSH-MLH complexes act as molecular switches triggered by nucleotide exchange. This exchange allows the complexes to enter a sliding-clamp diffusion mode in search of the unknown strand-discrimination signal. From Acharya S, Foster PL, Brooks P, and Fishel R (2003) The coordinated functions of the *E. coli* MutS and MutL proteins in mismatch repair. *Molecular Cell* 12: 233–246.

C-terminal fragments of the *E. coli* MutL protein. This analysis, described earlier, has shown that the MLH complexes undergo conformational changes that involve primarily the N-terminal domains. The existing structures lack connecting linker arms between the N- and C-terminal domains. Currently, only the N-terminal domains of the yeast and human PMS2 subunits (named PMS1 in yeast) have been solved.

The N-terminal domain of MutL family proteins contains an ATP-binding site that belongs to the GHKL family of ATPases. This domain in *E. coli* MutL dimerizes in the presence of a nonhydrolyzable ATP analog, AMP-PNP. ATP binding by GHKL family members induces large conformational changes in the proteins and such changes occur in the ATP-bound MutL proteins as well. The less-conserved C-terminal domain of MutL family proteins contains a region important for dimerization that does not appear to be modulated by nucleotide or other cofactors. The unstructured linker arms of MutL α become more ordered upon nucleotide binding and appear to promote interactions between the N-terminal domains of the two subunits.

Like the MSH family proteins, ATP hydrolytic activities appear different among the mammalian MLH subunits and are likely to be functionally relevant for coordinating interactions with downstream repair factors. MLH1 hydrolyzes ATP more proficiently than PMS2. In the absence of nucleotide, the overall structure of full-length MutL α is predicted to be a V-shaped molecule connected at the C-termini that undergoes conformational changes in the linker arms upon nucleotide binding to bring the N-termini together and form a ring-like structure. As previously mentioned, ATP has an apparent role in modulating the interactions of various components of the MMR pathway. MutS α bound to a mismatch is capable of interacting with MutL α in the presence of ATP to form a ternary complex. In the absence of a mismatch or ATP, MutS α and MutL α do not interact.

ATP-dependent dimerization of *E. coli* MutL is also a prerequisite for interactions with MutH and UvrD helicase.

What is the significance of the ATP-dependent conformational changes in MLH proteins? The ring-like structural topology of the MLH heterodimer in the presence of ATP suggests that DNA can thread through the central pore of the complex; such a structure could facilitate or regulate the sliding-clamp activities of the MSH complex (Figure 4). Consistent with this idea, biochemical analysis of MLH proteins has shown that they can bind to DNA; however, these proteins do not show an increased affinity for mismatched DNA. DNA binding by MLH proteins appears to be through nonspecific electrostatic contacts to the DNA backbone; such binding can be seen on both single-stranded and double-stranded DNA. There is also evidence for cooperative binding and polymerization along the DNA duplex.

It is unclear what role MLH DNA binding has in bridging mismatch recognition to downstream repair events. This activity, however, may be related to MutL α displaying a latent DNA endonuclease activity that is restricted to DNA near the mismatch site. This activity is essential for MMR and likely acts at the excision step of the process (see later). The endonuclease activity of MutL α resides in the PMS2 subunit and requires both ATP and heavy metal binding. Such an activity was not identified in MutL proteins from bacterial species that contain MutH and use the methylation-dependent strand-discrimination mechanism. Intriguingly, this observation hints at the possibility that the endonuclease activity of MutL α has replaced the requirement for MutH in higher organisms, though this is yet to be proven.

Downstream Steps in MMR

Downstream steps in MMR include strand discrimination, excision, resynthesis, and ligation. The factors that act in these steps do not appear to be MMR specific. They include (1) exonuclease 1 (EXO1), a factor required in multiple DNA repair pathways that acts in excision; (2) proliferating cell nuclear antigen (PCNA), the replication-processivity clamp required at multiple stages of MMR; the replication clamp loader (RFC) necessary for directing 3'-directed excision and proper loading of PCNA to initiate DNA synthesis; (3) replication protein A (RPA), an SSB that plays an important role in stimulating strand excision; (4) high mobility group protein B1 (HMGB1), a nonhistone chromatin associated protein likely having redundant functions with RPA; (5) DNA pol δ , a highly accurate polymerase used during resynthesis that fills in the gap created by excision; and (6) DNA ligase I, which seals the nick leftover after resynthesis to form an unbroken strand and complete repair. Coordination between each of these factors and each step of the MMR pathway is critical for efficient repair; the little that is known about how this is accomplished is indicated below. Mechanisms for strand discrimination, which are still being worked out, are discussed at the end of this section.

Excision

In reconstituted mammalian MMR involving a nicked mismatch substrate (typically a G–T mismatch), mismatch-dependent excision can be observed in the 5' to 3' or 3' to 5' direction, depending on the position of the strand nick in the DNA substrate relative to the mismatch. MutS α , RPA, and

the 5' to 3' double-stranded exonuclease EXO1 are sufficient for 5' to 3' excision from a nick to a mismatch; however, MutL α enhances mismatch dependence of this reaction by suppressing excision on homoduplex DNA by EXO1. Recent *in vitro* studies suggest that the addition of PCNA and RFC to a 3'-directed mismatch substrate is sufficient to promote 3' to 5' excision repair and to suppress inappropriate 5' to 3' excision. These observations raise the question of how EXO1, a 5' to 3' exonuclease, can support both 5'- and 3'-directed MMR. The finding that MutL α has a latent endonuclease activity that appears to be restricted to regions surrounding the mismatch provides an elegant solution; the endonuclease cleavages generated by MutL α are considered to provide an entry site for the 5' to 3' EXO1 exonuclease to act in bidirectional MMR.

Resynthesis

During reconstituted mammalian MMR, single-strand gaps are created during excision steps; these gaps can be as large as 1 kb in length and must be filled in by DNA polymerase and sealed by DNA ligase I to complete repair. The gaps created by excision span from a nick site introduced into a plasmid substrate (in lieu of a strand-specific repair signal) to roughly 150 base pairs beyond the mismatch site. The signaling mechanisms involved in coordinating excision and resynthesis steps are currently being worked out. Many of the MMR factors that act in early steps in MMR (MutS α , MutL α , RPA, PCNA, and RFC) have been implicated in both suppressing and terminating EXO1-mediated excision. The DNA synthesis steps in the reconstituted system are initiated by recruitment of DNA pol δ by PCNA that has been loaded by RFC.

Strand Discrimination: The Holy Grail in the MMR Field

Unlike many DNA repair events, where a lesion can provide localized information to direct repair, MMR uses information provided by the DNA replication machinery to remove the mismatch on the newly replicated strand. Perhaps the most interesting question remaining in the mammalian MMR field is how strand discrimination occurs. The nature of the strand-discrimination signal is unclear but genetic and biochemical evidence, obtained primarily from yeast and bacteria, suggest that it involves the identification of nicks on the newly replicated daughter strand (Figure 5).

The replication-processivity clamp PCNA interacts with many proteins, especially those involved in DNA repair, and this list includes the MutS α , MutS β , MutL α , and EXO1 MMR factors. In one model, PCNA provides a link between MMR and replication to allow mismatches to be identified immediately after formation by concentrating MMR factors at the replication fork. Like the hemi-methylation status of GATC sites in *E. coli*, nicks generated during replication are transient in nature. The identification of these nicks by MMR could, thus, be used as a strand-discrimination signal. PCNA is known to increase mismatch specificity of MutS α to activate EXO1 during excision and is required to initiate resynthesis by DNA polymerase. Furthermore, PCNA is always loaded onto junctions of single- and double-stranded DNA by RFC in a specific orientation. Thus, PCNA by virtue of its orientation upon loading could relay strand information through interactions with MutS α or MutL α . After MutS α or MutL α is oriented onto DNA, they could search out nearby strand discontinuities created during replication. There is also the possibility that a PCNA-independent mechanism orients MutS α or MutL α .

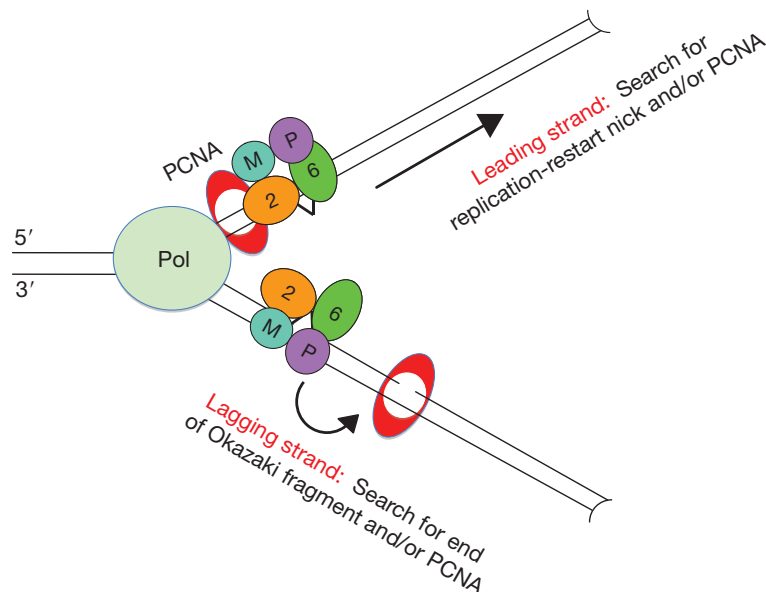


Figure 5 How might strand discrimination be accomplished during mismatch repair? Shown is a model of a replication fork with mismatches indicated by triangles. After mismatches are recognized, repair is directed to the daughter strand. Orientation of MutS α (MSH2-MSH6) or MutL α (MLH1-PMS2) to initiate excision of the newly replicated strand may be facilitated by interactions with proliferating cell nuclear antigen (PCNA). The nature of the strand-discrimination signal is unclear but is likely a nick created during replication. The lagging strand (bottom strand) contains unprocessed Okazaki fragments that provide nicks. The leading strand (top strand) would only generate nicks if the replication fork was restarted. Alternatively, it is possible that the DNA end at the elongating replication fork may serve to discriminate strands. MutL α could also be directed to cleave the correct strand to promote excision in each case.

Depending on the orientation of the strand discontinuity to the mismatch, excision can be directly started by the 5' to 3' exonucleases activity of EXO1 or, to allow 3'-directed repair by EXO1, would require a new strand break to be created on the opposite side of the mismatch through MutL α endonucleolytic cleavage. This model is supported by *in vitro* data showing that MutL α is dispensable on a substrate with a nick 5' to a mismatch, but is absolutely required for a substrate with a nick 3' to a mismatch. In order for the MMR system to utilize this mechanism, it would require relatively frequent strand discontinuities to be created during replication. Such discontinuities occur during lagging-strand synthesis where the average Okazaki fragment size is a few hundred base pairs. The fact that MMR appears more efficient on the lagging strand may reflect the high frequency of nicks that are accessible to MMR factors on this strand. There is little evidence for the presence of nicks on the leading strand at a density (every few kb) that would be utilized by MMR. However, several studies have suggested that leading-strand synthesis may not be as continuous as had been thought. One possibility is that like lagging-strand synthesis, replication is regularly restarted on the leading strand, thus creating strand discontinuities that can serve as a strand-discrimination signal. There is also the possibility that other undiscovered factors or signals play a role in this process.

Another point of contention in the field is how identification of the newly replicated strand is coordinated with mismatch recognition and downstream signaling. Recent evidence in bacterial and mammalian MMR systems suggests that MSH proteins form a sliding clamp with MLH proteins (Figure 4). Experiments in *E. coli* suggest that after mismatch recognition, MutS and MutL track along the helical contour to activate MutH to cleave the nearest hemi-methylated GATC site. These studies support a model in mammalian cells proposed by Rick Fishel and colleagues in which, following mismatch recognition, MSH-MLH complexes act as molecular switches that are triggered by nucleotide exchange to enter a sliding-clamp diffusion mode. In this model, MSH-MLH complexes then locate presently unknown strand-discrimination signals and guide excision steps that result in mismatch removal. An implication of this model is that multiple binding and release events can take place at the mismatch that may lead to polymer formation and communication between the two sites. The fact that MutL proteins can form polymers on DNA provides support for this idea. Excision needs to take place in the correct direction for competent repair to occur. Therefore, communication between the mismatch and the strand-discrimination signal is likely to be important because of the bidirectional nature of MMR.

Summary

The MMR system monitors the genome during DNA replication to ensure that mistakes made by DNA polymerases are corrected. A large group of repair factors acts in a coordinated

fashion to ensure accurate and timely repair. Some details of the repair process, such as how strand discrimination occurs to specifically remove errors on the newly replicated strand, are still unclear; however, the development of an *in vitro* mammalian MMR system from purified components, coupled with cell biological analysis of MMR factors, has provided a basic understanding of how MMR leads to mutation avoidance and how defects in this process can create a mutational load that can cause disease. The fact that mutations in many of the key recognition factors in MMR have been implicated in tumorigenesis illustrates the importance of MMR in maintaining mammalian genome stability.

See also: [Molecular Biology: DNA Mismatch Repair and Homologous Recombination](#); [DNA Mismatch Repair and the DNA Damage Response](#); [DNA Mismatch Repair in Bacteria](#); [DNA Mismatch Repair in Disease and Aging](#); [Repair of G-T Mismatches by *Escherichia coli* Vsr and Eukaryotic DNA Glycosylases](#).

Further Reading

- Acharya S, Foster PL, Brooks P, and Fishel R (2003) The coordinated functions of the *E. coli* MutS and MutL proteins in mismatch repair. *Molecular Cell* 12: 233–246.
- Ban C, Junop M, and Yang W (1999) Transformation of MutL by ATP binding and hydrolysis: A switch in DNA mismatch repair. *Cell* 97: 85–97.
- Constantin N, Dzantiev L, Kadyrov FA, and Modrich P (2005) Human mismatch repair: Reconstitution of a nick-directed bidirectional reaction. *Journal of Biological Chemistry* 280: 39752–39761.
- Gorman J, Chowdhury A, Surtees JA, et al. (2007) Dynamic basis for one-dimensional DNA scanning by the mismatch repair complex Msh2–Msh6. *Molecular Cell* 28: 359–370.
- Iyer RI, Pluciennik A, Burdett V, and Modrich PL (2006) DNA mismatch repair: Functions and mechanisms. *Chemical Reviews* 106: 302–323.
- Kadyrov FA, Dzantiev L, Constantin N, and Modrich P (2006) Endonucleolytic function of MutL α in human mismatch repair. *Cell* 126: 297–308.
- Kunkel TA and Erie DA (2005) DNA mismatch repair. *Annual Review of Biochemistry* 74: 681–710.
- Lahue RS, Au KG, and Modrich P (1989) DNA mismatch correction in a defined system. *Science* 245: 160–164.
- Lamers MH, Perrakis A, Enzlin JH, et al. (2000) The crystal structure of DNA mismatch repair protein MutS binding to a G $\times$ T mismatch. *Nature* 407: 711–717.
- Lynch HT, Lynch PM, Lanspa SJ, et al. (2009) Review of the Lynch syndrome: History, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clinical Genetics* 76: 1–18.
- Mazur DJ, Mendillo ML, and Kolodner RD (2006) Inhibition of Msh6 ATPase activity by mispaired DNA induces a Msh2(ATP)–Msh6(ATP) state capable of hydrolysis-independent movement along DNA. *Molecular Cell* 22: 39–49.
- Mechanic LE, Frankel BA, and Matson SW (2000) *Escherichia coli* MutL loads DNA helicase II onto DNA. *Journal of Biological Chemistry* 275: 38337–38346.
- Obmolova G, Ban C, Hsieh P, and Yang W (2000) Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* 407: 703–710.
- Tessmer I, Yang Y, Zhai J, et al. (2008) Mechanism of MutS searching for DNA mismatches and signaling repair. *Journal of Biological Chemistry* 283: 36646–36654.
- Warren JJ, Pohlhaus TJ, Changela A, et al. (2007) Structure of the human MutS α DNA lesion recognition complex. *Molecular Cell* 26: 579–592.
- Zhang Y, Yuan F, Presnell SR, et al. (2005) Reconstitution of 5'-directed human mismatch repair in a purified system. *Cell* 122: 693–705.

DNA Modification, Restriction Endonucleases: Type I Enzymes

D T F Dryden and G A Roberts, University of Edinburgh, Edinburgh, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Host specificity of DNA (*hsd*) In a host expressing a type I restriction/modification (R/M) system, DNA containing the specific target sequence becomes modified or restricted, depending upon the methylation status of the target sequence.

Modification The process accompanying restriction by which host DNA is maintained in a state of sequence-specific methylation. Most restriction enzymes cannot cleave DNA, which has been modified by the methyltransferase partner. Type I restriction enzymes combine both restriction endonuclease and methyltransferase in one protein complex.

Restriction The process of hindering the propagation of bacteriophage through a bacterial culture. This process requires cleavage of the bacteriophage DNA by DNA-sequence-specific endonucleases commonly referred to as restriction enzymes.

Translocation Usually, the processive motion of an enzyme along DNA is described as translocation. However, for type I restriction enzymes, which remain bound to their initial target sequence, translocation refers to the pulling in of DNA toward the enzyme.

Biochemical Function

Type I restriction enzymes are large complexes with multiple enzymatic functions, and they possess the ability to switch between these functions depending upon the chemical nature of specific nucleotides within a target DNA sequence. A demonstrated function of type I enzymes in their natural hosts, eubacteria and archaea, is to protect the host from infection by foreign DNA from bacteriophages. This protection is achieved by two processes – modification of the host DNA by methylation at the N6 position of adenine nucleotides within the recognition sequence targeted by the enzyme and destruction of invading DNA, which typically lacks the appropriate modification of the target sequence. For a host specifying a type I restriction enzyme, typically far fewer than one in a thousand phage particles successfully infect the host. Therefore, the type I restriction enzymes are a very effective defense system despite the considerable resources devoted by the host cell to their synthesis. This protective function is termed ‘restriction’ and results in the destruction of the phage DNA once it has entered the host cell. Methylation of host DNA occurs during each round of DNA replication. The newly synthesized DNA strand lacks the appropriate methylation, the type I restriction enzyme recognizes this DNA as hemimethylated and adds the methyl group to the adenine in the target sequence in the newly synthesized strand. In this manner, the type I restriction enzyme recognizes three different states of methylation of its target sequence, unmodified, hemimethylated, and fully modified, and performs a different reaction on each form of DNA (Figure 1). *In vitro*, the efficiency or accuracy of the switching between activities varies depending upon the particular type I restriction enzyme. For example, EcoKI, the type I restriction enzyme in *E. coli* K12, and EcoR124I, a plasmid-encoded type I R/M system, are very precise, methylating only hemimethylated targets and cutting only unmodified ones. In contrast, EcoAI from *E. coli* 15T[−] methylates unmodified targets as effectively as hemimethylated targets; as a result, this activity competes with the restriction of unmodified DNA. These three enzymes are the best-characterized members of the type I restriction enzymes.

Genes and Protein Structure

It has been shown that three genes are required to specify a type I R/M system: *hsdS*, *hsdM*, and *hsdR*, where *hsd* refers to host specificity for DNA and *S*, *M*, and *R* stand for specificity, modification, and restriction, respectively. Typically, the *hsdR* gene is expressed from its own promoter and the *hsdM* and *hsdS* genes from a single promoter upstream of *hsdM*. The *hsdR* gene is typically close to *hsdM/hsdS*, although some genome sequences suggest that this is not always the case. Genome rearrangements also occur around the *hsd* locus and this is used to induce some interesting switching of the expression and target-sequence specificity of alternative type I R/M systems within a cell. Most studies have concentrated upon the type I R/M systems of *E. coli* and *Salmonella enterica*, and it has been found in these hosts that type I R/M systems can be divided into at least four families based upon the following criteria: DNA hybridization, antibody cross-reactivity, and genetic complementation of the *hsd* genes from different systems. These families are known as type IA, IB, IC, and ID. Structural and bioinformatics studies of the Hsd subunits have been used to understand the basis of these families. Within a family, amino acid sequence identity of the three expressed polypeptides is very high except for two regions of HsdS involved in DNA target recognition. Amino acid sequence identity between polypeptides from different families is much lower and is usually confined to regions of HsdS involved in DNA target recognition, if the R/M systems being compared share at least part of the same DNA target sequence specificity, and to short motifs involved in catalysis.

The Sequence Specificity Subunit

HsdS, often referred to as the S subunit, is responsible for recognizing the characteristic bipartite, asymmetric DNA target sequence of type I restriction enzymes; for example, the target for EcoKI is 5′-AACNNNNNNGTGC-3′, where N is any base. S subunits have a mass of approximately 50 kDa and comprise five regions of primary structure (polypeptide sequence)

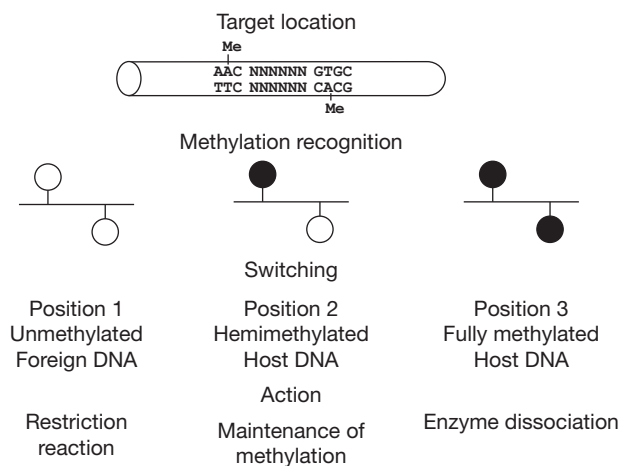


Figure 1 The operations performed by a typical type I restriction enzyme are illustrated using the EcoKI enzyme binding to its DNA target sequence AACNNNNNGTGC and determining the methylation status of the target. The sequence of operations is: target location, methylation recognition, switching between different activities, and performance of the selected action.

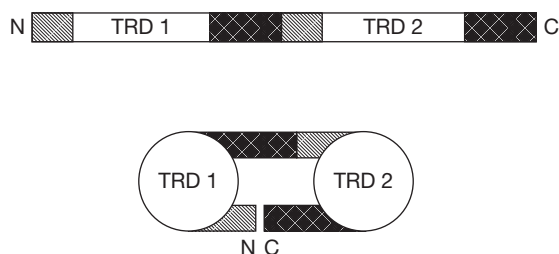


Figure 2 Diagrams of a typical S subunit of a type I restriction enzyme. The upper diagram shows the subunit as a linear representation from the N-terminus to the C-terminus. Regions of sequence conservation between different S subunits are shown hatched. Regions within the subunit with sequence conservation have the same sort of hatching. The conserved regions flank the TRDs responsible for DNA sequence recognition. Each TRD recognizes one part of the bipartite target sequence. The lower diagram depicts the folded arrangement of these regions making apparent the approximate twofold rotational symmetry axis of the S subunits.

(Figure 2). There are sequences (conserved regions) at the N-terminus, the central part of the polypeptide, and at the C-terminus that are highly conserved within a family of type I restriction enzymes and show some identity between families. Separating these three conserved regions are two longer sequences that display very limited sequence identity between different S subunits unless the different S subunits recognize the same DNA target sequence or one part of the same target sequence. These polypeptide sequences (variable regions) are associated with recognition of DNA target sequence and each region recognizes one part of the bipartite target sequence of the enzyme. These two variable regions each fold into a tertiary structure domain referred to as a target recognition domain (TRD). A weak sequence identity exists between the TRDs of the S subunits and the TRD of methyltransferases from type II R/M systems. This identity covers a region involved in DNA

sequence recognition. Indeed, extensive mutagenesis of a TRD from HsdS of EcoKI corroborated this similarity between TRDs from type I and type II R/M systems. The conserved polypeptide regions flanking the TRDs are structural components required to position the TRDs against the DNA target and to act as a scaffold for the binding of the M and R subunits. Some *hsdS* genes have been manipulated to exchange TRDs from different type I R/M systems. The resulting chimeric HsdS recognize new target DNA sequences. In addition, sequence similarity between conserved regions within an S subunit suggested an approximate twofold degree of symmetry in the subunit. A polypeptide comprising one TRD and part of the conserved regions formed a dimeric protein recognizing a twofold symmetrical DNA target. Within a family, the length of the central conserved region influences the number of nonspecific bases found in the DNA target sequence. Thus, naturally occurring variation in the structure of *hsdS* genes has led to the evolution of new target specificities. Kim and colleagues presented the first crystal structure of an S subunit from *Methanococcus jannaschii*. The essential features of the structure include two TRDs of almost identical topologies separated from one another by a coiled coil of long antiparallel α -helices. The interface between the two helical coils is made up of a plethora of hydrophobic residues. Thus, the coiled-coil region resembles a molecular ruler that facilitates recognition of the bipartite target sequences by appropriate orientation of the two TRDs. Moreover, the structure suggests that the bound DNA is subject to bending and exposure of the target adenine bases.

The Modification Subunit

Two M subunits bind to HsdS to form a core trimeric enzyme with the properties of sequence specificity, detection of DNA methylation (whether hemimethylated or unmethylated), binding of methylation cofactor S-adenosyl-methionine (SAM), and methyltransferase activity. These M subunits have an approximate molecular mass of 60 kDa. The central third of the polypeptide sequence contains a series of amino acid motifs that are found in SAM-dependent methyltransferases and are involved in binding of the cofactor, binding of the adenine base specified for methylation, and catalytic transfer of the methyl group from SAM to the N6 position of the adenine base. This region of the M subunit has been successfully modeled upon the catalytic domain of the methyltransferases of type II R/M systems. Mutations in the motifs found in this catalytic domain have deleterious effects on activity. Methylation by type I restriction enzymes almost certainly occurs via the same nucleotide flipping mechanism used by methyltransferases of type II R/M systems in which the base targeted for modification is displaced from the DNA double helix into a catalytic pocket in the SAM-binding catalytic domain. The N-terminal part of the M subunit of EcoKI appears to play a role in recognition of the methylation status of the DNA target sequence and mutations reduce the preference of EcoKI for methylating only hemimethylated DNA. The C-terminal region of M subunits appears to be important in interacting with the S subunit to form the core assembly. The polypeptide sequences of M subunits are almost identical within the same type I R/M family, but identity between families is mostly confined to the motifs in the catalytic domain.

The Restriction Subunit

The complete type I restriction enzyme is formed by adding two R subunits to the core trimer to form a complex with subunit stoichiometry $R_2M_2S_1$ and a mass of around 440 kDa, an enormous multifunctional molecular machine. As for HsdM, the R subunits show high sequence identity for members in the same family but lower identity, confined to a series of amino acid motifs, for comparisons between families. The complete restriction enzyme is still capable of methylating the DNA target sequence if it is hemimethylated and, if it is a type I R/M system of low preference such as EcoAI, of methylating unmethylated targets as well. However, unmethylated DNA target sequences primarily trigger the restriction reaction. The R subunits contain three main regions. An N-terminal domain contains an amino acid motif found in DNA endonucleases, particularly endonucleases from type II R/M systems. Mutagenesis experiments reveal that this domain is responsible for DNA cleavage by type I restriction enzymes. The central portion of the subunit contains a further set of amino acid motifs characteristic of the so-called DEAD-box helicases, enzymes that translocate or move DNA. These motifs form a multidomain structure that binds and hydrolyzes adenosine triphosphate (ATP) and couples this hydrolysis to the movement of DNA relative to the enzyme (DNA translocation). Mutation abolishes the ability to hydrolyze ATP and translocate DNA. In the absence of translocation, the enzyme cannot cut DNA and restriction activity is lost even though the endonuclease domain is unaltered. The C-terminal region of HsdR appears to be required for interaction with the core trimer. Lapkouski and colleagues presented the first crystal structure of an R subunit from EcoR124I in complex with Mg^{2+} and ATP. The overall structure is composed of four globular domains in a square-planar arrangement. Each domain is separated from its near neighbor by a prominent groove. Analysis of the structure, in particular, the shape and electrostatic character of the grooves, together with biochemical data, has allowed the authors to propose a path taken by a bound DNA molecule during translocation.

Assembly and Control of the Type I Restriction Enzyme

The assembly of the complete type I restriction enzyme is a complex affair and many type I R/M systems can form proteins with different subunit stoichiometries. The main partially assembled form of type I restriction enzymes is a trimer M_2S_1 , which forms a fully functional modification methyltransferase recognizing the same DNA target as the complete enzyme. Taylor and colleagues have recently presented small-angle neutron scattering experiments on an engineered version of the type I DNA methyltransferase of EcoR124I to obtain a low-resolution structural model. Other partially assembled forms and even individual subunits have been isolated and can display some limited activities such as sequence specificity, cofactor binding, and, for the $R_1M_2S_1$ form of EcoR124I, DNA methylation and ATP hydrolysis-driven DNA translocation. Intracellular concentrations of the Hsd subunits vary with host conditions and it has been suggested that concentration-dependent assembly may play a role in the establishment of type I R/M systems in new hosts. It has also been shown that when the host cell acquires unmodified target sequences after

suffering DNA damage, the restriction activity of type IA and IB R/M systems is controlled by proteolytic digestion of HsdR. This control process prevents the type I restriction enzyme from attacking the unmodified target sequences in the damaged host chromosome.

The Restriction Reaction

DNA Binding and Recognition of Unmodified DNA

The type I restriction enzymes display a strong preference, in the presence of their cofactors, for binding to their DNA target sequence rather than to other DNA sequences. Although sequence specificity resides within HsdS, the presence of HsdM appears to be essential for strong binding to DNA. Once bound, the HsdM plus SAM detect the presence of adenine methylation on each strand of the DNA target using the base-flipping mechanism. If the adenine is already methylated, then it will not be able to fit properly into the catalytic site as the methyl group will clash with the methyl group of SAM. This steric hindrance appears to change the conformation of the enzyme to allow the methylation reaction to proceed upon hemimethylated DNA or to dissociate from fully modified DNA. However, if both adenines are unmethylated, they can be accommodated within the HsdM along with SAM. The absence of steric hindrance for both target adenines appears to be the trigger for the highly complicated restriction reaction.

Upon recognition of an unmodified target site, the restriction enzyme becomes strongly attached to the DNA target site and does not appear to dissociate from the DNA after carrying out the restriction reaction. Strictly speaking, this lack of enzymatic turnover in the complete process means that type I restriction enzymes are not enzymes. If the DNA molecule contains more than one target site, the type I restriction enzymes bound at each site can stick to each other to form a large complex. This additional DNA-dependent dimerization of the enzyme arises from diffusional motion of the DNA molecule bringing the bound enzymes into contact and does not depend upon enzyme activity. The functional role, if any, of this dimerization is not known as it is not necessary for restriction.

ATP Hydrolysis

Unmodified DNA target sequences trigger the hydrolysis of large amounts of ATP. The ATPase activity continues throughout the DNA translocation and cleavage process and continues even after DNA cleavage. The purpose of this continuing ATP hydrolysis observed *in vitro* is not known but may represent continuing DNA translocation on the cleaved DNA substrate.

DNA Translocation

The enzyme remains bound to the target sequence and commences pulling DNA on either side of the target sequence toward itself. As it is still attached at the target sequence, loops of DNA appear to be extruded from the enzyme (Figure 3). This process requires ATP hydrolysis and generates extensive supercoiling in the DNA. The reeling in of DNA occurs at approximately 500 base pairs (bp) per second and has been shown to occur over huge distances up to approximately 50 000 bp.

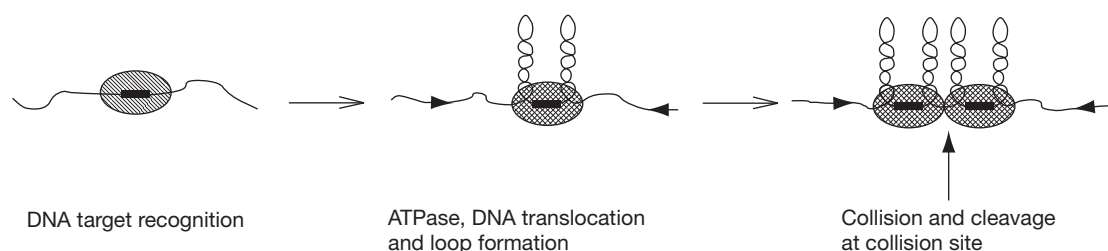


Figure 3 The sequence of events occurring after the type I restriction enzyme (oval) has bound and recognized an unmodified DNA target sequence (black box) on a DNA molecule (thin line). ATP hydrolysis drives the translocation of DNA, reeling the DNA in toward the enzyme with, as a consequence, the extrusion of loops of DNA. Eventually, translocation is blocked by, in this example, collision with a second translocating type I restriction enzyme, whereupon cleavage occurs at the collision point. After cleavage *in vitro*, the enzymes remain bound to their target sequence.

Single-molecule experiments performed by Seidel and colleagues to analyze the DNA translocation of EcoRI 24I demonstrate that the two R subunits work independently. The same study showed that, in this case, translocation was limited to ~4300 bp per binding event, which is much shorter than reported for other type I enzymes. The apparent discrepancy may be due to instability of the EcoRI 24I assembly, leading to the dissociation of the R subunit from the remainder of the enzyme.

Stalling of the movement appears to occur when the translocating enzyme collides with a blockage on the DNA. This blockage can be another translocating type I restriction enzyme if there are two or more target sequences on a particular DNA molecule, stalling on a circular DNA molecule containing only one target sequence or collision with a complex DNA structure such as a fixed Holliday junction.

Cleavage of DNA

Translocation blockage appears to be the signal for cutting of the DNA strands at the stalling location. Cutting does not occur at the original target sequence. As translocation rates of individual type I restriction enzymes on a DNA molecule may vary depending upon the particular enzyme species and upon the sequence being translocated, individual enzymes will stall or collide with each other at different locations on the DNA. Hence, in a solution containing many DNA molecules, translocation blockage and DNA cleavage occur at a range of locations. For a DNA molecule containing two or more target sequences for the same type I restriction enzyme, cleavage occurs approximately halfway between the original DNA target sequences. For a DNA molecule containing target sequences for different type I restriction enzymes, the approximate location of cleavage depends on the relative translocation speeds of the different enzymes. Thus, DNA cleavage by type I restriction enzymes does not produce DNA fragments of defined length and sequence but rather a broad range of fragment lengths. Circular DNA molecules containing one target sequence are also cleaved at an undefined distance from the target sequence. The presence of a nick in a DNA molecule appears to bias the cleavage event to be close to the nick.

Anti-Restriction

Despite their complexity, type I restriction enzymes are very effective barriers against horizontal gene transfer. However,

bacteriophages have developed a range of counter-measures to overcome this barrier. These anti-restriction strategies include selection against the presence of target sequences, modification of DNA bases, hydrolysis of the SAM cofactor, and the use of genes encoding specialized proteins that act as competitive inhibitors of DNA binding by type I restriction enzymes.

This last strategy is employed by bacteriophage T7, which produces gene 0.3 protein, also known as overcome classical restriction (Ocr), immediately upon infection of *E. coli*. Ocr binds very strongly to all characterized type I restriction enzymes, irrespective of their DNA target sequences, and prevents the enzyme from binding to DNA. The crystal structure of Ocr has been solved by Walkinshaw and colleagues to a resolution of 1.8 Å. The protein exists as a highly extended homodimer that mimics the size and shape of a bent B-form DNA duplex of ~24 bp in length. Furthermore, the protein is replete with acidic residues that mimic the negative charges of the phosphate backbone of the DNA molecule. Stephanou and colleagues have demonstrated that extensive neutralization of the acidic residues of Ocr by chemical modification, while leaving the overall protein fold largely unaffected, reduced the affinity of the molecule for the M₂S₁ methyltransferase (M.EcoKI) of EcoKI by as much as 800-fold. Electron microscopy analysis by Kennaway and colleagues of the M.EcoKI:Ocr complex, combined with modeling data indicate that Ocr is completely enveloped by M.EcoKI and fits perfectly into the DNA-binding site of type I restriction enzymes. In addition, the proposed model is entirely consistent with a large body of experimental data that has accumulated over some 40 years of research on EcoKI. Recently, McMahon and colleagues solved the crystal structure of the ArdA anti-restriction protein from the conjugative transposon Tn916. Like Ocr, the ArdA protein is a homodimer with an extended curved cylindrical structure. However, the ArdA protein is even more elongated than Ocr and mimics a ~42 bp stretch of B-form DNA – by far the longest DNA mimic yet described. Furthermore, the distribution of negative charge on the molecule shows an obvious mimicry with that of the DNA double helix.

The question of how DNA mimics, such as Ocr and ArdA, could have arisen through the process of natural selection is intriguing. Sequential acquisition of point mutations to generate codons corresponding to acidic residues seems a highly unlikely event. Nonetheless, once established, a mimic can spread efficiently throughout its biological niche. For example, ArdA is widespread in eubacteria on plasmids and transposons while Ocr-related proteins are common in bacteriophages related to the T7 phage. More research needs to be undertaken

to fully understand the evolutionary origins of these DNA mimics and their biological role in assisting horizontal gene transfer.

See also: **Molecular Biology:** DNA Polymerase I, Bacterial; Eukaryotic DNA Polymerase δ ; Repair of G–T Mismatches by *Escherichia coli* Vsr and Eukaryotic DNA Glycosylases.

Further Reading

- Bertani G and Weigle JJ (1953) Host controlled variation in bacterial viruses. *Journal of Bacteriology* 65: 113–121.
- Bickle TA and Kruger DH (1993) Biology of DNA restriction. *Microbiological Reviews* 57: 434–450.
- Burckhardt J, Weisemann J, Hamilton DL, and Yuan R (1981) Complexes formed between the restriction endonuclease EcoK and heteroduplex DNA. *Journal of Molecular Biology* 153: 425–440.
- Davies GP, Kemp P, Molineux IJ, and Murray NE (1999) The DNA translocation and ATPase activities of restriction-deficient mutants of EcoKI. *Journal of Molecular Biology* 292: 787–796.
- Davies GP, Martin I, Sturrock SS, et al. (1999) On the structure and operation of type I DNA restriction enzymes. *Journal of Molecular Biology* 290: 565–579.
- Kennaway CK, Obarska-Kosinska A, White JH, et al. (2009) The structure of M.EcoKI type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucleic Acids Research* 37: 762–770.
- Kim J-S, DeGiovanni A, Jancarik J, et al. (2005) Crystal structure of DNA sequence specificity subunit of a type I restriction–modification enzyme and its functional implications. *Proceedings of the National Academy of Sciences of the United States of America* 102: 3248–3253.
- Lapkouski M, Panjikar S, Janscak P, et al. (2009) Structure of the motor subunit of type I restriction–modification complex EcoR124I. *Nature Structural and Molecular Biology* 16: 94–95.
- McMahon SA, Roberts GA, Johnson KA, et al. (2009) Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance. *Nucleic Acids Research* 37: 4887–4897.
- Meselson M and Yuan R (1968) DNA restriction enzyme from *E. coli*. *Nature* 217: 1110–1114.
- Murray NE (2000) Type I restriction systems: Sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiology and Molecular Biology Reviews* 64: 412–434.
- Seidel R, van Noort J, van der Scheer C, et al. (2004) Real-time observation of DNA translocation by the type I restriction modification enzyme EcoR124I. *Nature Structural and Molecular Biology* 11: 838–843.
- Stephanou AS, Roberts GA, Cooper LP, et al. (2009) Dissection of the DNA mimicry of the bacteriophage T7 Ocr protein using chemical modification. *Journal of Molecular Biology* 391: 565–576.
- Studier FW and Bandyopadhyay PK (1988) Model for how type I restriction enzymes select cleavage sites in DNA. *Proceedings of the National Academy of Sciences of the United States of America* 85: 4677–4681.
- Taylor JE, Callow P, Swiderska A, and Kneale GG (2010) Structural and functional analysis of the engineered type I DNA methyltransferase EcoR124I_{NT}. *Journal of Molecular Biology*, doi:10.1016/j.jmb.2010.03.008.
- Tock MR and Dryden DT (2005) The biology of restriction and anti-restriction. *Current Opinion in Microbiology* 8: 466–472.
- Walkinshaw MD, Taylor P, Sturrock SS, et al. (2002) Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Molecular Cell* 9: 187–194.

DNA Oxidation

D O Zharkov, SB RAS Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Arthur P. Grollman and Dmitry O. Zharkov, volume 1, pp. 694–697, © 2004, Elsevier Inc.

Glossary

AP site A nucleotide in DNA that has lost the nucleobase.

DNA damage Any change in the bases (Ade, Cyt, Gua, and Thy) or phosphodiester backbone structure that comprise canonical DNA.

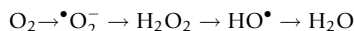
DNA repair Any enzymatic process that participates in purging DNA of damaged elements and returns it to its canonical form.

Lesion A damaged nucleotide in DNA.

Reactive oxygen species A molecule that contains an oxygen atom with a highly reactive configuration of electrons.

Oxidants and Antioxidants

Oxygen-releasing photosynthesis and aerobic respiration have provided the basic machinery of life for almost 3 billion years and have driven the evolution of complex life forms. The reduction of oxygen to water, due to spin restrictions imposed on an O_2 molecule, proceeds in sequential steps involving three reactive oxygen species (ROS): superoxide anion radical, hydrogen peroxide, and hydroxyl radical:



Additionally, molecular oxygen in the 1A_g excitation state (singlet oxygen) is included in the category of ROS. Reactive nitrogen species, such as peroxynitrite $ONOO^-$ formed upon recombination of O_2^- with an important signaling molecule, NO, are often considered together with ROS due to their similar biological effects. In cells, ROS react with many molecules, with their average lifetime and distance of diffusion from the point of generation being inversely related to their reactivity. Free radicals, such as O_2^- and $HO\cdot$, trigger chain reactions of oxidation, allowing unpaired electrons to migrate over longer distances. Unsaturated fatty-acid side chains are especially prone to reactions with free radicals; the resulting peroxy radicals propagate by a chain reaction. Given the ubiquitous presence of phospholipid membranes in cells, lipid peroxidation also plays an important role in the toxicity associated with ROS.

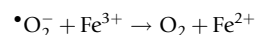
The primary source of ROS in eukaryotes is the mitochondrial respiratory chain. Approximately 0.1% (under phosphorylating respiration) to 2% (under resting respiration) of consumed oxygen is converted to ROS, primarily through auto-oxidation of the semiquinone forms of ubiquinone and flavin mononucleotides in complex I and complex III. In plants, photosystem I is also a major producer of ROS.

Other cellular redox systems are also capable of producing ROS. Important nonmitochondrial endogenous sources of ROS are reactions catalyzed by cytochromes P450, NADPH oxidase, xanthine oxidase, etc. In mammals, immunoglobulins generate ROS through oxidation of water by singlet oxygen.

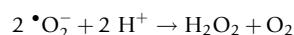
The major environmental source of ROS is ionizing radiation. Radiolysis of water releases hydroxyl radicals and other

reactive entities, such as solvated electrons, $H\cdot$ and $\cdot O^-$ radicals. Other sources of ROS include ultrasound, photosensitizing dyes (methylene blue and rose bengal), transition metal ions, ozone, a variety of salts (e.g., $KBrO_3$), and certain drugs.

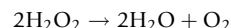
Hydroxyl radicals and singlet oxygen are the most reactive forms of ROS and, consequently, the most damaging to biomolecules. Hydroxyl radicals are generated from superoxide and H_2O_2 by Fenton chemistry when Fe^{2+} or other low-valence transition metal ions are present; such ions bind to DNA and these reactions play an important role in DNA damage:



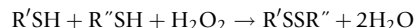
In living organisms, various mechanisms have evolved to minimize the effects of oxidative damage, including deactivation of ROS by superoxide dismutases, catalases, and peroxidases. Superoxide dismutases (SODs) are ubiquitous metal-dependent enzymes that contain Cu/Zn, Mn, Fe, or Ni ions. All forms of SOD catalyze the dismutation of superoxide radicals into hydrogen peroxide and oxygen:



At the next stage of oxygen reduction, hydrogen peroxide is converted to water and oxygen by catalase:



or peroxidase:



Almost any low-molecular-weight substance that contains a reducing moiety can react with ROS. Such nonenzymatic radical scavengers are present in significant quantities in cells or concentrated at ROS target sites. These antioxidants include lipophilic substances (α -tocopherol and β -carotene) and hydrophilic molecules (ascorbic acid, uric acid, glucose, and glutathione) that prevent transition metals from producing hydroxyl radicals. The low concentrations of free Fe^{2+} in mammalian cells are due to the presence of intra- and extracellular proteins that bind and sequester iron (ferritin and transferrin) or oxidize it (ceruloplasmin).

DNA Damage

DNA is an important target of oxidative damage. Both the sugar–phosphodiester backbone and the bases are subject to attack by ROS, producing single- and double-strand breaks, intra- and interstrand cross-links, abasic (AP) sites, cyclonucleotides, and a variety of modified bases. Lipid-derived radicals induced by ROS also damage DNA. All DNA bases are subject to direct oxidation; some of the principal lesions detected in cellular DNA are shown in **Figure 1**. A characteristic feature of ROS-mediated radiation damage to DNA is production of clustered lesions, which are separated by less than 20 base pairs and are easily converted to double-strand breaks.

8-Oxoguanine

One of the most prevalent and best-studied oxidized bases in DNA, 8-oxoguanine (8-oxoGua), is formed by oxidation of guanine at C8. Almost all endogenous and environmental sources of ROS can induce the appearance of 8-oxoGua. Of several different mechanisms of its formation, two are likely the most important in biological systems: one involves an attack by $\text{HO}^\bullet$ at C8 of Gua, another is initiated by singlet oxygen addition at C4–C8.

Several methods of detection (electrochemical, mass spectrometry, ^{32}P -postlabeling, immunoassay, and comet assay) and efficient synthetic protocols have been developed for this lesion, making it a widely used biomarker of oxidative DNA damage and the substrate of choice to study the repair of oxidative DNA damage. Under normal conditions, human

genomic DNA contains approximately one 8-oxoGua base per 10^6 Gua, and this level significantly increases in many pathologic conditions.

Like all purine nucleotides substituted at C8, 8-oxo-2'-deoxyguanosine can easily adopt the *syn* conformation. In the normal *anti* conformation, 8-oxoGua forms a Watson–Crick pair with Cyt, whereas in the *syn* conformation, it forms a Hoogsteen pair with Ade. This ability provides the basis for mutagenic properties of 8-oxoGua. Nevertheless, 8-oxoGua has almost no impact on the three-dimensional structure of DNA; both types of 8-oxoGua-containing pairs are easily accommodated in regular B-DNA. This base can also be easily oxidized further, forming several possible products (e.g., guanidinohydantoin), each having its own biological consequences.

Formamidopyrimidines and Oxidized Adenine Lesions

Although Gua is the most easily oxidized purine, Ade can be converted to 8-oxoadenine through the same hydroxyl radical chemistry. However, Ade is more prone to form another oxidative product, 2-hydroxyadenine. Another important class of purine oxidation products, formamidopyrimidines, is formed when 8-hydroxyguanylyl or 8-hydroxyadenylyl radicals are reduced rather than oxidized to 8-oxopurines. The yields of 8-oxopurines and formamidopyrimidines are similar, making the latter biologically important as well. Generally, physicochemical properties of formamidopyrimidines, their base-pairing capabilities, and their interactions with DNA polymerases and DNA repair proteins are similar to their counterpart 8-oxopurines.

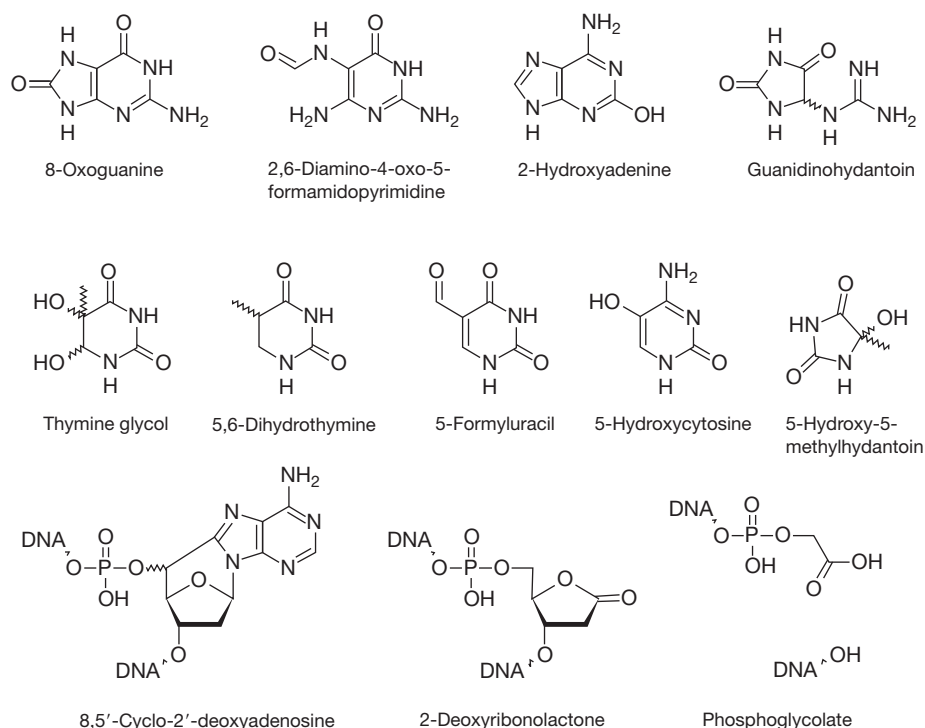


Figure 1 Examples of major oxidative lesions in DNA.

Oxidized Pyrimidines

Oxidation of pyrimidines initiated by HO[•] leads to the formation of three primary radicals (at C6, C5, and exocyclic C5 in the case of Thy) and a wide spectrum of products of their further conversion. Pyrimidine oxidation may be accompanied by deamination (as in 5-hydroxyuracil, a derivative of Cyt), ring contraction (as in 5-hydroxy-5-methylhydantoin), opening, or fragmentation. Some damaged pyrimidines that are technically not oxidized, such as 5,6-dihydropyrimidines, are also included in this class because they are produced as a part of spectrum of true oxidized pyrimidines during irradiation. Many oxidized pyrimidines lose their planarity, making them much more distorting for DNA than oxidized purines discussed earlier.

Products of Deoxyribose Oxidation

DNA oxidation involving the deoxyribose moiety often leads to base loss and DNA strand cleavage with a variety of abasic sites produced and a variety of ends generated at the site of the break. However, two classes of oxidative lesions retain the base and the integrity of the DNA strand: 8,5'-cyclopurines and α -anomers of nucleotides. Oxidation of deoxyribose forms a basis of action of important anticancer drugs, bleomycin and neocarzinostatin.

Biological Effects

Cytotoxicity and Mutagenesis

The major cytotoxic effects of ROS represent responses to DNA double-strand breaks and lesions that block replication. Certain oxidized bases are mutagenic, for example, 8-oxoGua accommodates a miscoding base (Ade) during DNA synthesis. The proofreading activity of DNA polymerases fails to remove deoxyadenosine monophosphate (dAMP) when it is incorporated opposite 8-oxoGua. Thus, mutations at 8-oxoGua are predominantly G→T transversions. DNA polymerases differ significantly with respect to nucleotide insertion opposite 8-oxoGua. Replicative polymerases (e.g., eukaryotic DNA polymerases α and δ and bacterial DNA polymerase III) are strongly blocked by 8-oxoGua and preferentially insert dAMP opposite this damaged base, while polymerases involved in DNA repair (such as eukaryotic DNA polymerase β and bacterial DNA polymerase I) mainly incorporate dCMP. DNA polymerases responsible for lesion bypass (e.g., eukaryotic DNA polymerases ζ , η , and κ , and bacterial DNA polymerases II, IV, and V) can insert both dAMP and dCMP opposite 8-oxoGua. On the other hand, oxidized pyrimidines, due to their nonplanar structure, often strongly block DNA polymerases.

Deoxynucleotide triphosphates are also subject to damage by ROS, promoting base misincorporation, which may lead to mutational events. In addition to their harmful consequences associated with replication, oxidative lesions in DNA may promote so-called transcriptional mutagenesis by instructing RNA polymerases to misincorporate dNMPs, which then may lead to amino-acid substitution or early termination of translation. The presence of oxidative lesions in DNA is known to interfere with binding of some regulatory proteins and with action of DNA-processing enzymes, such as helicases or topoisomerases.

Carcinogenesis

Oxidative stress is an important factor in tumor initiation and promotion. It is clear that somatic mutations play a central role in carcinogenesis and oxidative DNA damage has been implicated as an initiating event in this process. Its importance can be appreciated by noting that the amount of oxidized bases in DNA is significantly higher than the amount of DNA adducts formed by well-established chemical carcinogens. The role of ROS in tumor development may also be attributed, in part, to their mitogenic properties and their ability to damage cellular proteins and membranes.

8-oxoGua and thymine glycol are frequently used as biomarkers for oxidative DNA damage. DNA in tumors usually contains increased amounts of oxidized bases. As solid tumors are often hypoxic and display lower levels of general oxidative damage, this increase in DNA damage appears to be a cause rather than a consequence of neoplastic transformation. For tumor development to occur, a series of mutations are generally required to inactivate tumor suppressor genes and/or alter the status of protooncogenes. The presence of G→T transversions, characteristic of 8-oxoGua, suggests that oxidative damage may have been involved in these instances. This signature transversion is found in mutated *TP53* and *RAS* genes in many human cancers and experimental systems.

Aging and Age-Related Diseases

In 1956, D. Harman proposed a free-radical theory of aging. Oxidative DNA damage increases with age and the correlations between aging and DNA oxidation are reflected in increasing levels of 8-oxoGua and thymine glycol. The relationship between an organism's life span and its ability to repair oxidatively damaged DNA has also been reported.

High levels of 8-oxoGua are present in several age-related noncancer pathologies, such as Parkinson's disease. Although the frequency of mutations increases with age, there are no documented cases of somatic mutations in the nuclear genome that are directly involved in age-related changes or age-related diseases, with the exception of spontaneous cancers and trinucleotide repeat pathologies. Oxidation of DNA increases its immunogenicity, potentially contributing to development of autoimmune diseases. Mitochondrial DNA (mtDNA), rather than nuclear DNA, has been proposed as a principal site of age-related damage accumulation. Direct estimates of oxidative damage in mtDNA usually are much higher than those in nuclear DNA. A specific, age-related deletion in the mtDNA of postmitotic cells, as well as multiple rearrangements, has been documented.

Repair of Oxidative DNA Damage

General

DNA damage occurs to a significant degree despite the ubiquitous presence of ROS scavengers. However, several cellular repair mechanisms are available: direct reversal, base excision repair, nucleotide excision repair, mismatch repair, recombination repair, and nonhomologous end-joining. Double-strand breaks are generally repaired by recombination repair or nonhomologous end-joining; interstrand cross-links, by a

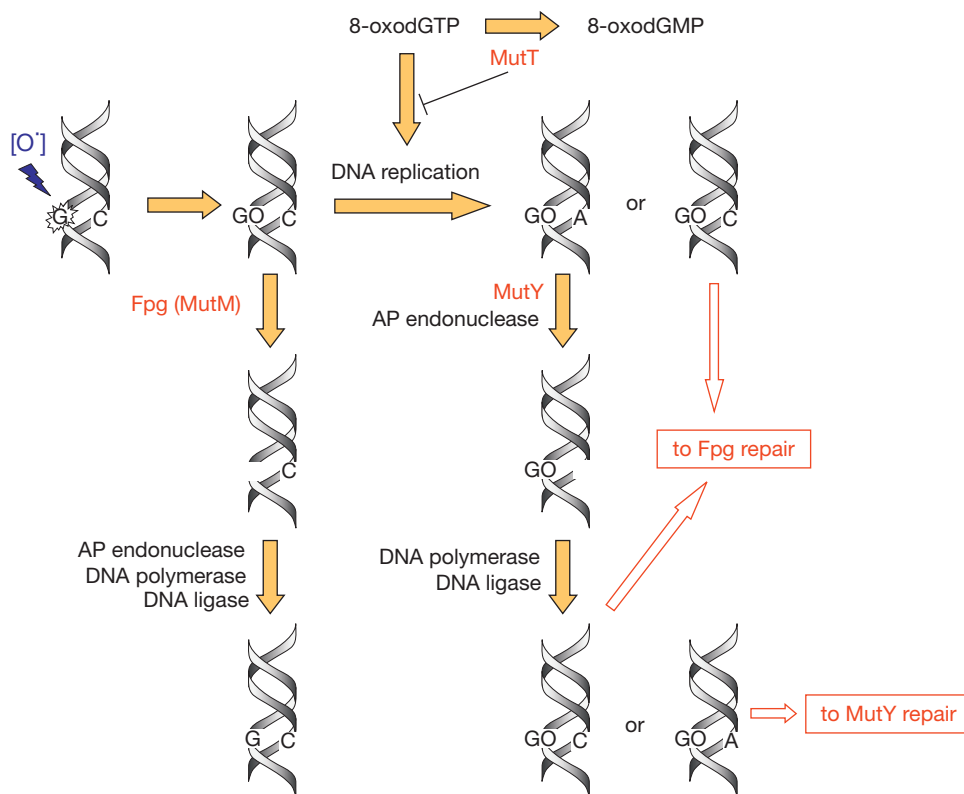


Figure 2 GO system in bacteria.

combination of nucleotide excision repair and recombination repair; and intrastrand cross-links and cyclonucleotides, by nucleotide excision repair. Most oxidative damage is dealt with by base excision repair. Here, damaged bases are excised by specific DNA glycosylases, creating an AP site, and an AP endonuclease then cleaves DNA 5' to this position. Deoxyribose phosphate lyase removes the sugar residue, creating a substrate for a DNA polymerase. After the polymerase fills the resulting gap, a DNA ligase restores the double-stranded DNA. Several variations of this essential scheme exist.

Despite the wide variety of oxidatively damaged bases found in DNA, only a few glycosylases are involved in their repair. These enzymes necessarily exhibit broad substrate specificity, with some acting primarily on oxidized pyrimidines and others on oxidized purines. In *Escherichia coli*, oxidative damage to pyrimidines is repaired by endonuclease III (Nth) and endonuclease VIII (Nei); structurally related enzymes perform similar functions in eukaryotes (NTHL1, NEIL1, NEIL2, and NEIL3 in humans). More than 20 types of oxidized bases are excised by these DNA glycosylases. Depending on the chemistry performed by the DNA glycosylases, modified nicks of different nature are generated in DNA and processed by different AP endonucleases and phosphatases to create a suitable end for the DNA polymerase.

GO System

8-oxoGua presents a special problem for DNA repair. Since either dAMP or dCMP can be incorporated opposite it, the repair system must convert both 8-oxoGua:Cyt and

8-oxoGua:Ade mispairs into a Gua:Cyt pair. At least two cycles of base excision repair are required to repair 8-oxoGua:Ade. First, Ade, the normal base, is excised, a process followed by insertion of dCMP opposite the lesion. The resulting 8-oxoGua:Cyt mispair is then repaired by excision of 8-oxoGua and subsequent insertion of dGMP.

In *E. coli*, a system of three enzymes – Fpg (MutM), MutT, and MutY (GO system) – counters the deleterious effects of 8-oxoGua (Figure 2). MutT, a nucleotide hydrolase, cleanses the cellular nucleotide pool of 8-oxodGTP. Fpg is an 8-oxoguanine-DNA glycosylase that excises 8-oxoGua from 8-oxoGua:Cyt, but not from 8-oxoGua:Ade. MutY protein is an adenine glycosylase specific for Ade:8-oxoGua and Ade:Gua pairs. If 8-oxoGua in DNA arises by spontaneous oxidation of Gua or by incorporation of 8-oxodGMP opposite Cyt, it is excised by Fpg. However, if Ade:8-oxoGua is present because of misincorporation of 8-oxodGMP opposite Ade, it becomes a substrate for MutY. If dAMP is inserted again by a DNA polymerase after excision by MutY, the cycle of MutY repair is repeated, whereas dCMP insertion converts the 8-oxoGua:Ade mispair to 8-oxoGua:Cyt, a substrate for Fpg. The GO system also operates in most eukaryotes; in humans, OGG1 protein is a functional analog of Fpg, MUTYH is a homolog of MutY, and NUDT1 is a homolog of MutT. Certain polymorphisms of OGG1 confer increased risk of lung cancer, and inactivating mutations of MUTYH cause colorectal cancer.

See also: **Bioenergetics: Mitochondria: Source and Target of Free Radicals; Reactive Oxygen and Nitrogen Species: Interactions with**

Mitochondria and Pathophysiology; Superoxide Dismutase;
Molecular Biology: DNA Base Excision Repair Pathways; DNA
 Glycosylases: Mechanisms.

Further Reading

- Barja G (2004) Free radicals and aging. *Trends in Neuroscience* 27: 595–600.
- David SS, O'Shea VL, and Kundu S (2007) Base-excision repair of oxidative DNA damage. *Nature* 447: 941–950.
- Dizdaroglu M, Jaruga P, Birincioglu M, and Rodriguez H (2002) Free radical-induced damage to DNA: Mechanisms and measurement. *Free Radical Biology and Medicine* 32: 1102–1115.
- Friedberg EC, Walker GC, Siede W, et al. (2006) *DNA Repair and Mutagenesis*. Washington, DC: ASM Press.
- Halliwell B and Gutteridge JMC (2007) *Free Radicals in Biology and Medicine*. Oxford: Oxford University Press.
- Kamiya H (2003) Mutagenic potentials of damaged nucleic acids produced by reactive oxygen/nitrogen species: Approaches using synthetic oligonucleotides and nucleotides. *Nucleic Acids Research* 31: 517–531.
- Kanvah S, Joseph J, Schuster GB, et al. (2010) Oxidation of DNA: Damage to nucleobases. *Accounts of Chemical Research* 43: 280–287.
- Lukin M and de los Santos C (2006) NMR structures of damaged DNA. *Chemical Reviews* 106: 607–686.
- Olinski R, Gackowski D, Rozalski R, Foksinski M, and Bialkowski K (2003) Oxidative DNA damage in cancer patients: A cause or a consequence of the disease development? *Mutation Research* 531: 177–190.
- von Sonntag C (2006) *Free-Radical-Induced DNA Damage and Its Repair: A Chemical Perspective*. Berlin–Heidelberg: Springer.

DNA Polymerase β , Eukaryotic

W A Beard and S H Wilson, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA

Published by Elsevier Inc.

Glossary

Apurinic/aprimidinic (AP) site DNA lesion resulting from the loss of a base (i.e., abasic site). These sites can be mutagenic since they have lost their coding potential or cytotoxic since they are quickly incised resulting in DNA strand breaks.

C-subdomain One of the three subdomains found in all DNA polymerases. This subdomain includes three active site acidic residues that coordinate two metals (Mg^{2+}) that assist nucleotidyl transfer. It is also referred to as the palm subdomain in the polymerase nomenclature that utilizes the analogy to a right hand.

D-subdomain One of the three subdomains found in all DNA polymerases. This subdomain interacts with duplex DNA upstream of the primer terminus. It is also referred to as the thumb subdomain in the polymerase nomenclature that utilizes the analogy to a right hand.

DNA polymerase β A small eukaryotic DNA polymerase involved in short-gapped DNA repair synthesis. In addition to its polymerase activity, this polymerase contributes an accessory lyase activity required to remove the backbone of an AP site (i.e., deoxyribose) during base excision repair.

Helix-hairpin-helix (HhH) The HhH structural motif binds single- or double-stranded DNA in a

non-sequence-specific manner and has been identified in a number of DNA repair proteins. Two such motifs in pol β are observed to make DNA backbone interactions with the incised DNA strand downstream and upstream of the gap, suggesting that they stabilize the pronounced DNA bend observed when gapped DNA binds to pol β .

Lyase In the context of single-nucleotide base excision repair, this reaction results in the removal of the 5'-deoxyribose phosphate backbone after incision of an AP site by an endonuclease. The reaction proceeds by β -elimination through a Schiff base intermediate. The ϵ -NH₂ group of a lysine side chain (Lys72 of pol β) serves as a nucleophile resulting in transient covalent attachment of the enzyme to its substrate. This Schiff base intermediate can be trapped by sodium borohydride, resulting in conversion to an irreversibly linked lyase-DNA complex.

N-subdomain One of the three subdomains found in all DNA polymerases. This subdomain forms one face of the nascent base pair (templating and incoming nucleotide bases) binding pocket and it is involved in selecting the correct deoxynucleotide triphosphate. It is also referred to as the fingers subdomain in the polymerase nomenclature that utilizes the analogy to a right hand.

Introduction

Endogenous and environmental agents continually threaten cellular DNA. These threats can result in physical damage (e.g., DNA strand breaks or base loss) or modification (e.g., alkylation and cross-links). Since these genetic insults lead to an altered DNA structure that can result in deleterious outcomes, cells have evolved DNA repair mechanisms to correct these abnormalities. The base-excision repair (BER) pathway is responsible for removing simple base lesions and AP sites in DNA. The repair of an AP site minimally requires four coordinated enzymatic activities: strand incision by AP endonuclease, removal of the resulting deoxyribose phosphate (dRP) backbone of the AP site by a pol β -associated lyase, single-nucleotide DNA synthesis by pol β , and ligation of the DNA nick by DNA ligase. These steps are illustrated in [Figure 1](#).

Pol β is the smallest eukaryotic cellular DNA polymerase (335 residues; 39 kDa) and it lacks a proofreading 3'→5' exonuclease activity that enhances the accuracies of replicative DNA polymerases (e.g., DNA polymerase ϵ and δ). Based on primary sequence alignments, pol β belongs to the X-family of DNA polymerases that also includes pol λ and pol μ .

Biological Role

Based upon the high level of sequence conservation of pol β among mammalian species, it seemed highly likely that pol β is conducting a role that is essential for animal survival. This is consistent with the embryonic lethality of pol β null mice. Pol β preferentially fills short DNA gaps (<6 nucleotides). Because of this attribute, it had generally been assumed that pol β is involved in short-gap DNA repair synthesis. In addition, early studies implicated pol β in gap-filling DNA synthesis during mammalian BER of alkylation damage and the repair of ultraviolet (UV) DNA damage, and inhibitor studies also implicated pol β in other types of DNA repair, such as the repair of oxidative DNA damage. Although these attributes and studies implicated pol β in BER and in other types of short-gap-filling DNA repair, pol β 's role in BER was confirmed when mouse cells lacking pol β were found to be hypersensitive to DNA-damaging agents believed to be repaired by the BER pathway.

In addition to the simple single-nucleotide BER pathway illustrated in [Figure 1](#), alternate BER pathways are needed to remove modified dRP moieties that cannot be excised by the pol β lyase activity. In this situation, pol β strand displacement

DNA synthesis creates a DNA flap of several nucleotides with a 5'-modified dRP group that can be subsequently removed by flap endonuclease-1 (FEN-1), thereby creating a nick that will be sealed by DNA ligase. This alternate pathway is referred to as 'long-patch' BER. In the absence of strand displacement DNA synthesis, FEN-1 incision excises the dRP group and one nucleotide, thereby creating a one-nucleotide gap that will be filled by pol β . Alternating single-nucleotide gap-filling DNA synthesis and FEN-1 activities can create a long BER patch and has been referred to as 'hit and run'.

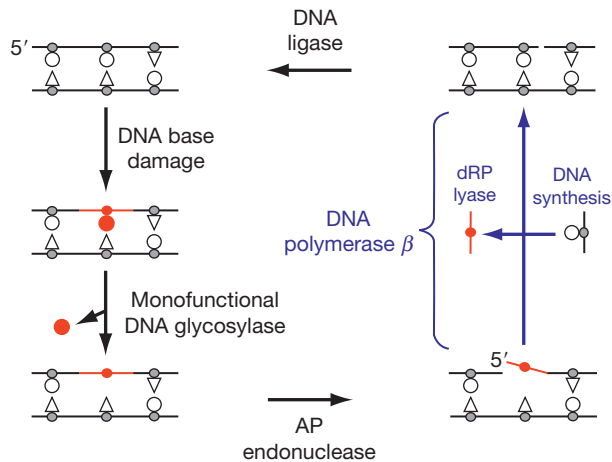


Figure 1 Single-nucleotide base excision repair. AP sites are lesions in DNA that originate from spontaneous or enzymatic (glycosylase) removal of a damaged base (large red circle). These sites are removed from DNA by four coordinated enzymatic activities, two of which are contributed by pol β (blue); DNA synthesis and dRP lyase that occur concurrently.

Pol β Domain Organization

Controlled proteolytic or chemical cleavage of pol β first demonstrated that it is folded into discrete domains. Subsequently, the X-ray crystal structure of the ternary substrate complex with DNA and an incoming nucleoside triphosphate (dNTP) defined the location of these domains in relation to the tertiary structure of the polymerase. Pol β is organized into two domains: an 8-kDa amino-terminal lyase domain and a 31-kDa carboxyl-terminal polymerase domain (Figure 2).

The structures of DNA polymerases derived from other polymerase families indicate that they also have a modular domain organization. The polymerase domain is typically composed of three functionally distinct subdomains. The catalytic subdomain coordinates two divalent metal cations that promote DNA synthesis. The other two subdomains are spatially situated on opposite sides of the catalytic subdomain. Based on an architectural analogy to a right hand, the subdomains are referred to as palm, thumb, and fingers. To highlight the intrinsic function of the polymerase subdomains of pol β , a functionally based nomenclature has been proposed. Accordingly, the pol β polymerase domain is composed of C- (catalytic), D- (duplex DNA binding), and N-subdomains (nascent base pair binding) (Figure 2). These correspond to the palm, thumb, and fingers subdomains employing the right-hand analogy.

Lyase Domain

The amino-terminal lyase domain (residues 1–90) has an associated lyase, metal-independent, activity that removes the 5'-dRP moiety generated during single-nucleotide BER (Figure 1). This activity plays a pivotal role in BER in that if the lyase reaction is inhibited (e.g., modified dRP moiety),

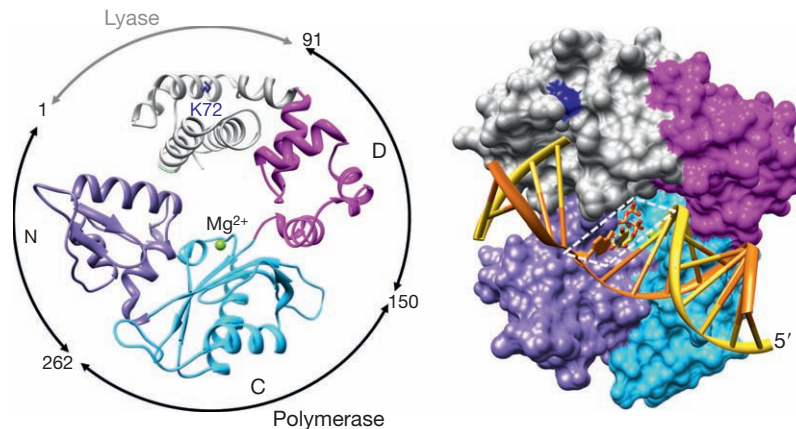


Figure 2 Domain and subdomain organization of pol β . (Left) A ribbon representation of pol β illustrating the polymerase (colored) and amino-terminal lyase (gray) domains (PDB code 2FMS). The polymerase domain is composed of three subdomains: D, magenta; C, light blue; and N, purple. These correspond to the thumb, palm, and fingers subdomains of DNA polymerases that utilize an architectural analogy to a right hand, respectively. Lys72 of the dRP lyase active site is indicated (blue), as well as the catalytic Mg^{2+} required for DNA synthesis (green sphere). (Right) The molecular surface of pol β with single-nucleotide gapped DNA and incoming nucleoside triphosphate substrates. The template strand is orange and the upstream and downstream portions of the incised strand are yellow. The duplex DNA is illustrated in a ladder representation with the nascent base pair (templating nucleotide and incoming nucleoside triphosphate) delineated with a white dashed parallelogram. The lyase domain and the D-subdomain each have a HhH motif (not shown) that interacts with the DNA backbone of the incised DNA strand downstream and upstream of the gap, respectively.

alternate BER pathways will be initiated. The lyase reaction proceeds by β -elimination through a Schiff base intermediate. A stable covalently bound intermediate can be formed between the dRP-containing DNA substrate and the enzyme by NaBH_4 trapping. Lys72 is preferentially modified by primary amine reactive agents in both full-length pol β and the isolated lyase domain. Site-directed mutagenesis of Lys72 diminishes dRP lyase activity nearly two orders of magnitude, indicating that this residue is the likely nucleophile that forms the Schiff base intermediate. The lyase domain is highly basic (net charge = +10), and the positively charged surface potential predicts that it would interact with the DNA backbone.

In addition to providing a crucial enzymatic activity for single-nucleotide BER, the lyase domain possesses a 5'-phosphate binding site that facilitates targeting of pol β into the 5'-position in a DNA gap. Importantly, pol β binds tightly to the 5'-phosphate only when there is DNA adjacent to the 5'-phosphate. This can be double-stranded or single-stranded DNA since a 3'-primer terminus is not required for optimum binding. Accordingly, pol β is expected to bind to the 5'-phosphate in a DNA gap of any size. The observation that pol β can processively (i.e., insert several nucleotides before dissociating from the DNA substrate) fill short gaps (<6 nucleotides) suggests that the lyase domain tethers the polymerase domain to the downstream position in gapped DNA. When the primer terminus (i.e., 3'-OH) is within 6 nucleotides of the 5'-phosphate on the downstream DNA strand, the proximity of the polymerase domain to the primer terminus would hasten processive DNA synthesis.

The lyase domain is composed of five α -helices (Figure 2, left panel). Structural characterization of this domain has identified a structural motif that binds monovalent metals and interacts with the DNA backbone. The helix-hairpin-helix (HhH) motif (residues 55–79) has also been identified in other DNA repair proteins as well as many proteins that bind either single- or double-stranded DNA in a non-sequence-specific manner. The crystal structure of pol β bound to a one-nucleotide gap DNA substrate indicates that the lyase domain and the N-subdomain of the polymerase domain interact (Figure 2). The interaction of the lyase domain and the N-subdomain results in a donut-like structure. The interactions between these subdomains are altered upon binding substrates resulting in a tighter complex.

Polymerase Domain

D-subdomain

The polymerase domain is composed of three functionally distinct subdomains. The lyase domain is connected to the D-subdomain (residues 91–149) with a protease-hypersensitive hinge region. The D-subdomain interacts with the DNA sugar-phosphate backbone of the duplex DNA upstream of the polymerase active site. In addition to the HhH motif found in the lyase domain, a second HhH motif (residues 92–118) is also found in the D-subdomain. This HhH motif interacts with the primer strand phosphate backbone through a monovalent metal ion. Thus, the two HhH motifs in pol β are observed to make DNA backbone interactions with each end of the incised DNA strand. In the crystal structure of pol β bound with an incoming nucleotide and a one-nucleotide

DNA gap, the DNA is bent about 90° (Figure 3, left panel). The DNA template does not travel through the hole in the donut-like structure. The sharp bend occurs at the 5'-phosphodiester bond of the templating base and is also observed in a product complex where pol β is bound to nicked DNA. The 3'-hydroxyl and 5'-phosphate groups on the nicked DNA strand, which needs to be ligated in the final step of BER, are observed to be over 27 Å apart. For pol β , the function of the HhH motifs appears to be a sequence nonspecific phosphate-backbone binding motif that stabilizes the pronounced bend observed in the pol β -gapped DNA structure. The 90° bend also exposes the terminal base pairs of each DNA duplex that is situated in the gap. His34 of the lyase domain interacts with the first base pair of the downstream duplex, whereas the N-subdomain contributes interactions with the nascent base pair (see below). The dramatic bend of the template strand as it enters the polymerase active site is a general feature observed in crystal structures of substrate complexes of most DNA polymerases.

C-subdomain

The C-subdomain (residues 150–261) of pol β includes three acidic residues (Asp190, Asp192, and Asp256) that coordinate two essential magnesium ions. One magnesium coordinates nonbridging oxygens on the three phosphates of the incoming nucleotide (nucleotide-binding metal), whereas the other coordinates oxygens on the α -phosphate of the incoming dNTP and the DNA primer terminus (catalytic metal). These metals are believed to play critical and conserved roles in catalysis (nucleotidyl transfer) in all DNA polymerases. Although the C-subdomains of X-family and bacterial pol III members are not homologous to those of many other polymerase families, crystal structures of substrate complexes of DNA polymerases from other families indicate that the reactive groups (i.e., metals, incoming nucleotide, and DNA) have a similar three-dimensional arrangement. Likewise, all DNA polymerases follow an ordered binding of substrates where DNA binds first. The polymerase then selects a nucleoside triphosphate from a pool of structurally similar molecules to preserve Watson-Crick base-pairing rules. The ability to choose between right and wrong is highly dependent on the identity of the polymerase. Pol β has a moderate ability to select the correct (i.e., right) nucleotide and capacity for DNA synthesis. Since it does not have an intrinsic proofreading activity, it has the potential to be highly mutagenic. The fidelity of pol β is too low (1 error per ~2000 nucleotides synthesized) for pol β DNA synthesis errors to be tolerated during DNA repair. It is generally believed that the proofreading activity of AP endonuclease, or some other extrinsic 3'→5'-exonuclease, may proofread base substitution errors generated by pol β .

N-subdomain

The N-subdomain (residues 262–335) contributes important interactions to the binding pocket of the nascent base pair (templating and incoming nucleotides). Comparison of DNA polymerase structures bound to DNA with those that include an incoming complementary dNTP reveals that the N-subdomain repositions itself to sandwich the nascent base pair between the growing DNA terminus and the polymerase (Figure 3). In the absence of an incoming nucleotide, the N-subdomain is open, but forms a closed complex upon

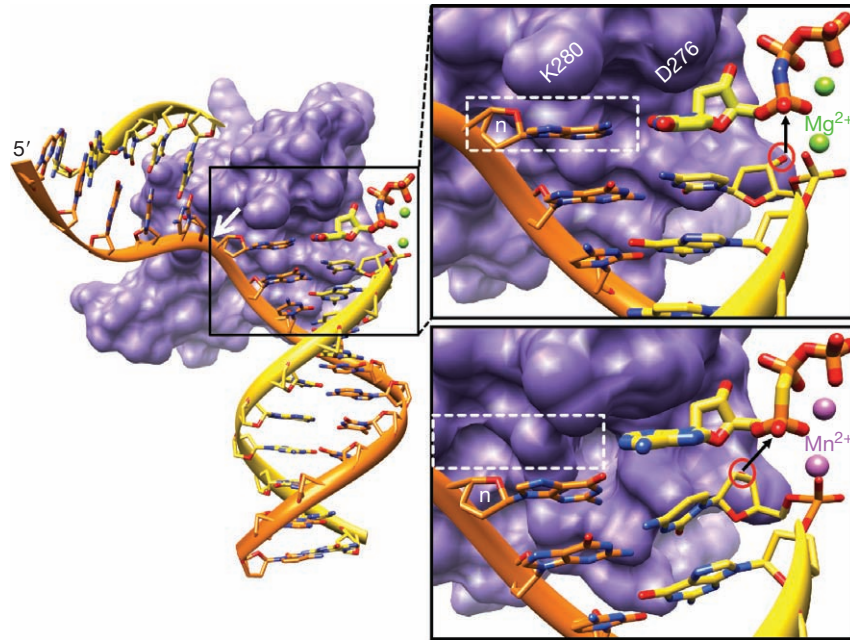


Figure 3 Nascent base-pair-binding pocket. The perspective is from the major groove edge of the nascent base pair. (Left) The N-subdomain (purple) is shown forming one face of the binding pocket for the new base pair (i.e., templating nucleotide and incoming nucleoside triphosphate). The lyase and C- and D-subdomains are not shown for clarity. The 5'-end of the template strand (orange) is indicated. The complementary DNA strands (primer and downstream oligonucleotides) are yellow. The trajectory of the template strand is altered dramatically as it enters the polymerase active site at the 5'-phosphodiester bond of the templating nucleotide (white arrow). The active site metals (Mg^{2+}) that coordinate the incoming dNTP and assist catalysis are illustrated as green spheres. (Right, upper panel) Detailed view of the nascent base-pair-binding pocket occupied by a correct base pair. The templating nucleotide (n) binding site (e.g., adenosine) is outlined with a white dashed box. The 3'-OH of the primer terminus (red circle) is in an ideal position to attack $\text{P}\alpha$ of the incoming nucleotide (black arrow). The active site metals (Mg^{2+}) that assist catalysis are shown. Asp276 (D276) and Lys280 (K280) make van der Waals contact with the bases of the incoming and templating nucleotides, respectively. Asn279 and Arg283 form the minor groove edge of the binding pocket. (Right, lower panel) Detailed view of the nascent base-pair-binding pocket during insertion of the wrong nucleotide (dAMPCPP) with a templating guanine (PDB code 3C2M). The active site metals (Mn^{2+} , pink) are shown. In this case, the template pocket (white dashed box) is empty due to an upstream shift in the template strand. The shift in template strand also displaces $\text{O}3'$ (red circle) of the primer terminus deterring misinsertion by altering the geometry (angle and distance) of attack on $\text{P}\alpha$ of the incoming nucleotide (black arrow).

binding a correct nucleotide. Thus, the dNTP-binding pocket is formed by the template base, DNA duplex terminus, and enzyme. For pol β , subdomain interactions with the nascent base pair are contributed primarily through the N-subdomain (Figure 3). These include stacking interactions with Lys280 and Asp276 with the templating and incoming nucleotide bases, respectively. Additionally, Asn279 and Arg283 contribute DNA minor groove interactions. Alanine substitution for Arg283 results in a dramatic decrease in fidelity. Since Arg283 plays a critical role in the formation of the closed complex, the low-fidelity mutant (i.e., R283A) is believed to be in an open conformation that has lost the ability to promote efficient DNA synthesis when a correct nucleotide binds. Thus, the ability to choose the correct nucleotide has been lost. As noted above, the trajectory of the duplex DNA into the polymerase active site requires that the templating strand bends by 90° (Figure 3). This bend in the template strand serves at least two functions: (1) it provides access for the N-subdomain to check whether geometrical constraints imposed by correct Watson-Crick hydrogen bonding occurs and (2) it discourages the next templating base from entering the polymerase active site prematurely which could result in the incorrect template base coding for nucleotide insertion (deletion mutagenesis). The closed polymerase conformation also discourages

misinsertion by removing the templating nucleotide from its coding positions when an incorrect nucleotide binds (Figure 3, bottom right panel). In this situation, the binding of an incorrect incoming nucleotide results in an upstream shift in the template strand near the active site. This serves at least two functions: (1) removes coding/misreading potential with the incoming nucleotide, thereby decreasing the binding affinity of the incorrect nucleotide, and (2) alters the geometry of the primer terminus relative to $\text{P}\alpha$ of the incoming nucleotide that will deter misinsertion. The removal of the templating base from its coding position has been compared to that of insertion of a nucleotide, typically dATP, opposite an abasic site. It remains to be seen whether this is a general or alternate strategy to optimize polymerase fidelity.

Pol β -BER-Protein Interactions

In addition to its catalytic function, pol β also interacts with several proteins known to be involved in BER. These interactions may facilitate the coordination of the necessary enzymatic steps required in alternate BER pathways. In this context, the observations that pol β interacts with other DNA repair proteins such as AP endonuclease, DNA ligase I, X-ray

cross-complementing factor-1 (XRCC1), poly(ADP-ribose) polymerase-1 (PARP-1), and proliferating cell nuclear antigen (PCNA), among others, suggest complex regulatory mechanisms that are not fully understood. Thus, pol β interacts with the enzymes involved in the steps immediately upstream and downstream of its own steps in single-nucleotide BER. Importantly, the formation of binary or ternary protein complexes, rather than a super BER complex, has the advantage of flexibility that can accommodate alternate pathways. The influence of these protein–protein interactions on catalytic function and processing of BER intermediates remains to be determined.

See also: [Molecular Biology: DNA Damage: Alkylation; DNA Glycosylases: Mechanisms; DNA Ligases: Mechanism and Functions; DNA Polymerases: Kinetics and Mechanism; Eukaryotic DNA Polymerase \$\delta\$; XPV Polymerase and the Bypass of Ultraviolet DNA Damage.](#)

Further Reading

- Batra VK, Beard WA, Shock DD, et al. (2006) Magnesium induced assembly of a complete DNA polymerase catalytic complex. *Structure (Camb)* 14: 757–766.
- Batra VK, Beard WA, Shock DD, Pedersen LC, and Wilson SH (2008) Structures of DNA polymerase β with active site mismatches suggest a transient abasic site intermediate during misincorporation. *Molecular Cell* 30: 315–324.
- Beard WA, Shock DD, Batra VK, Pedersen LC, and Wilson SH (2009) DNA polymerase β substrate specificity: Side chain modulation of the “A-rule.” *Journal of Biological Chemistry* 284: 31680–31689.
- Beard WA, Shock DD, and Wilson SH (2004) Influence of DNA structure on DNA polymerase β active site function: Extension of mutagenic DNA intermediates. *Journal of Biological Chemistry* 279: 31921–31929.
- Beard WA and Wilson SH (2006) Structure and mechanism of DNA polymerase β . *Chemical Reviews* 106: 361–382.
- Horton JK, Joyce-Gray DF, Pachkowski BF, Swenberg JA, and Wilson SH (2003) Hypersensitivity of DNA polymerase β null mouse fibroblasts reflects accumulation of cytotoxic repair intermediates from site-specific alkyl DNA lesions. *DNA Repair (Amst)* 2: 27–48.
- Horton JK and Wilson SH (2007) Hypersensitivity phenotypes associated with genetic and synthetic inhibitor-induced base excision repair deficiency. *DNA Repair (Amst)* 6: 530–543.
- Liu Y, Beard WA, Shock DD, et al. (2005) DNA polymerase β and flap endonuclease 1 enzymatic specificities sustain DNA synthesis for long patch base excision repair. *Journal of Biological Chemistry* 280: 3665–3674.
- Liu Y, Prasad R, and Wilson SH (2010) HMGB1: Roles in base excision repair and related function. *Biochimica et Biophysica Acta* 1799: 119–130.
- Moon AF, Garcia-Diaz M, Batra VK, et al. (2007) The X family portrait: Structural insights into biological functions of X family polymerases. *DNA Repair (Amst)* 6: 1709–1725.
- Prasad R, Batra VK, Yang X-P, et al. (2005) Structural insight into the DNA polymerase β deoxyribose phosphate lyase mechanism. *DNA Repair (Amst)* 4: 1347–1357.

DNA Polymerase I, Bacterial

C M Joyce, Yale University, New Haven, CT, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Apo-enzyme Enzyme in the absence of bound substrates.

Mispair Two opposed bases in DNA that do not conform to the Watson–Crick rules of complementarity, that is, not A–T or G–C.

Primer In duplex DNA, the strand that presents a 3' end for extension by a DNA polymerase.

Template In duplex DNA, the strand that is paired with the primer, and, therefore, provides the information for copying by DNA polymerase.

Introduction

DNA polymerase I (Pol I) is the most abundant DNA polymerase in eubacteria. Although it lacks the speed and processivity of the more complex polymerases which replicate the bacterial chromosome, it is ideally suited for the synthesis of short stretches of DNA in excision repair and in the removal of RNA primers during lagging strand replication. DNA Pol I of *Escherichia coli*, the first DNA polymerase to be discovered and studied, has long served as the prototype for this class of enzymes. Klenow fragment, which contains the polymerase activity of Pol I, was the first polymerase to be studied crystallographically. Subsequently, Klenow fragment and its homologs from bacteriophage T7 and the thermophiles *Thermus aquaticus* (*Taq* DNA pol) and *Bacillus stearothermophilus* (*Bst* DNA pol) have provided a wealth of informative co-crystal structures that are the foundation for our current understanding of the DNA polymerase reaction mechanism.

Overview of Structure

Bacterial Pol I enzymes are multifunctional and follow the pattern seen throughout the polymerase superfamily: a modular structure with a common polymerase domain (described below) and auxiliary enzyme activities located on separate protein domains. All bacterial Pol Is have an N-terminal domain with 5' nuclease activity; this activity is responsible for the removal of DNA ahead of the growing primer strand. Attachment of the 5' nuclease domain to the rest of the polymerase is protease sensitive; in *E. coli* Pol I, mild protease digestion was originally used to remove the 5' nuclease (35 kDa) from the remainder of the molecule (Klenow fragment, 68 kDa); subsequently, this was accomplished by recombinant DNA manipulations. The linkage between the 5' nuclease and Klenow fragment is almost certainly flexible, as shown by the observation that the 5' nuclease does not occupy the same position relative to the rest of the molecule in different crystal forms of *Taq* DNA pol. The Pol I homologs found in bacteriophage genomes resemble the Klenow fragment portion of the bacterial Pol Is, in that they lack the 5' nuclease portion of the sequence. In some cases, such as bacteriophage T7, the 5' nuclease is encoded as a separate gene product.

The Klenow fragment portion of Pol I also contains two domains. In about half of the bacterial Pol Is, the N-terminal of the two domains contains a 3'–5' exonuclease (3' exo) activity, which serves as a proofreader to eliminate polymerase errors. If the polymerase were to insert an incorrect nucleotide, resulting in a terminal mispair, the 3'–5' nuclease would remove the incorrect primer-terminal nucleotide and provide a second opportunity for correct insertion. Those bacterial Pol I enzymes that lack 3'–5' exonuclease activity (from species including *Thermus*, *Bacillus*, and *Rickettsiae*) nevertheless contain the domain, though key active site side chains are absent. Aside from the thermophiles, where fraying of double-stranded DNA termini at high temperatures might encourage excessive nuclease degradation, it is unclear why certain bacterial Pol I enzymes have proofreading activity and others do not.

Based on the sequence of their polymerase domains, the bacterial Pol Is are placed in the A-family, which also includes some bacteriophage DNA polymerases as well as the mitochondrial DNA polymerase gamma and two eukaryotic nuclear DNA polymerases, ν (nu) and θ (theta).

Polymerase

Domain Structure and Relation to Other Polymerases

When the structure of Klenow fragment was reported in 1985, the polymerase domain structure was described as resembling a half-open right hand, with subdomains called 'fingers', 'palm', and 'thumb' forming a cleft. Currently, structures are available for over 60 polymerases, encompassing all four biochemical categories of polymerase (DNA- or RNA-dependent, with DNA or RNA as product), and five of the six families of DNA-dependent DNA polymerases (as defined by protein sequence alignments). It is clear that the overall polymerase domain structure originally observed in Klenow fragment is a common theme (Figure 1), though a detailed examination reveals two distinct polymerase lineages which can be recognized by the structure of the palm subdomain. The larger group, comprising all polymerases except the DNA polymerase families C and X, contains a highly conserved structural motif (the polymerase fold) consisting of a three-stranded antiparallel β -sheet supported by two α -helices. This serves as the scaffold for important active site residues, in particular a pair of

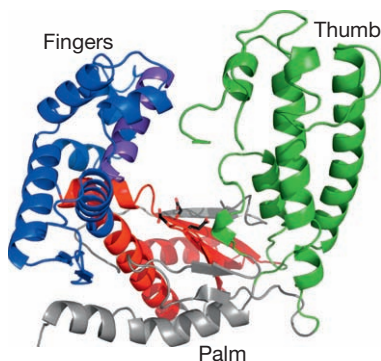


Figure 1 Three-dimensional structure of the polymerase domain of *E. coli* DNA polymerase I. The thumb and fingers subdomains are colored green and blue, respectively, with the O-helix on the fingers subdomain in purple. The palm subdomain is gray, except for the polymerase fold, a conserved structural motif found in the majority of nucleic acid polymerases, which is shown in red. Three conserved carboxylate side chains, two of which are ligands to the catalytic metal ions, mark the location of the polymerase active site. From Brautigam CA and Steitz TA (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Current Opinion in Structural Biology* 8: 54–63.

carboxylate ligands to the two divalent metal ions that catalyze the phosphoryl transfer reaction. The fingers and thumb subdomains are structurally much more divergent between polymerase families, but carry out analogous functions in all polymerases – the fingers providing important active site residues, particularly those involved in nucleotide binding, and the thumb binding the primer–template duplex.

Substrate Binding

The binding of substrates at the polymerase active site has been revealed in co-crystals of several Pol I homologs. The template–primer duplex (corresponding to the product of DNA synthesis) is bound in a shallow cleft between the thumb and 3′ exo domains, with the largely α -helical thumb providing important binding contacts to the phosphate backbone. In a polymerase–DNA binary complex, the primer terminal base pair abuts the side of the polymerase cleft formed by the fingers subdomain. This wall, which defines one side of the active site, is formed primarily by a long α -helix (the O-helix, shown in purple in **Figure 1**), which runs the length of the fingers subdomain and has a group of important and highly conserved side chains on the surface pointing into the cleft. At the C-terminus of the O-helix is an invariant Tyr side chain, which is stacked against the template side of the terminal base pair.

When the deoxyribonucleotide triphosphate (dNTP) complementary to the templating position is added so as to form a polymerase–DNA–dNTP ternary complex, co-crystal structures indicate that the polymerase must have undergone a conformational change (**Figure 2**). The novel conformation seen in ternary complex structures differs from the apo-enzyme and binary complex structures by a substantial movement of a segment of the fingers subdomain so as to close the polymerase cleft. In conjunction with the transition from the open (binary complex) to closed (ternary complex) conformation, the Tyr side chain at the C-terminus of the O-helix moves further down

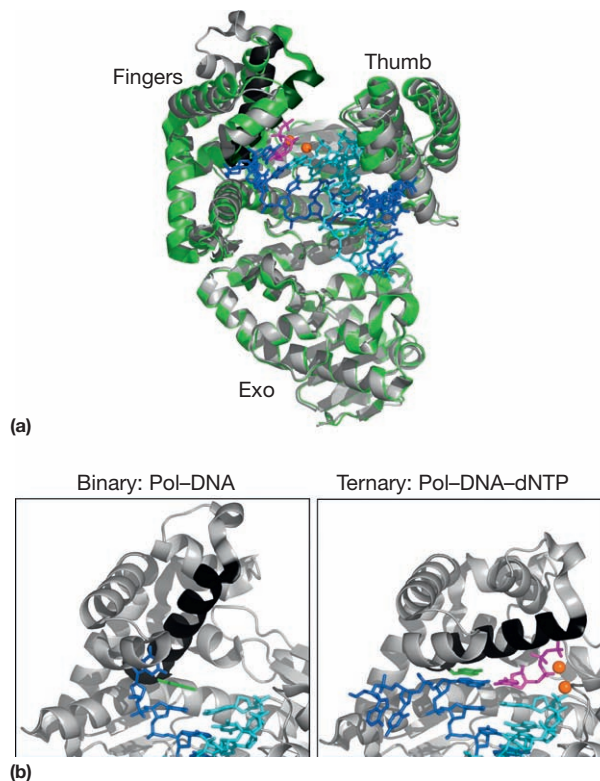


Figure 2 (a) Superposition of binary (Pol–DNA) and ternary (Pol–DNA–dNTP) complexes of the Klenow-fragment-like portion of *Bst* DNA pol. The DNA template (dark blue) and primer (cyan), incoming nucleotide (red) and active site metal ions (orange) are shown in their positions in the ternary complex. The protein backbone in the binary complex is gray, and the ternary complex green. The two protein structures are essentially superimposable except for a segment of the fingers subdomain that includes the O-helix (shown in black and dark green, respectively). The domain marked Exo corresponds to the 3′–5′ exonuclease domain in Klenow fragment, though it has no enzymatic activity in *Bst* DNA pol. (b) Detailed comparison of the active site regions in the two complexes. In each case, the protein is shown in gray, with the O-helix in black. Coloring of the DNA, nucleotide, and metal ions is the same as in the top panel. The structures are oriented so that the palm subdomain and the DNA are in similar positions in both panels, emphasizing the movement of the O-helix. The side chain of the conserved Tyr at the C-terminus of the O-helix (Tyr714 in *Bst* DNA pol, Tyr 766 in Klenow fragment) is shown in green. This side chain is stacked with the template side of the terminal base pair in the binary complex; in the ternary complex, it has moved out of the way to allow the next template base to enter the active site (note that this template base is bound in a pocket between the O and O₁ helices in the binary complex). For clarity, the top of the thumb subdomain is omitted in all the structures illustrated in this figure. From Johnson SJ, Taylor JS, and Beese LS (2003) Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proceedings of the National Academy of Sciences of the United States of America* 100: 3895–3900.

into the active site, away from its stacked position with the terminal base pair. This allows the templating base to stack with its 3′ neighbor at the duplex terminus, and to base pair with the incoming dNTP in a snug binding pocket formed by side chains of the O-helix and nearby residues on one side, and the primer-terminal base pair on the other side. Analogous conformational transitions have been inferred from crystal

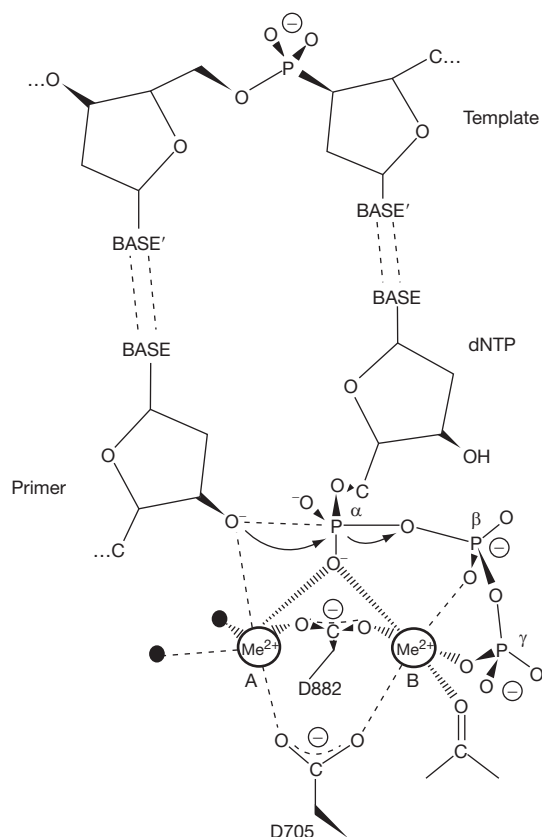


Figure 3 Mechanism of phosphoryl transfer, resulting in nucleotide addition, at the polymerase active site. Catalysis is mediated by two divalent metal ions, coordinated, in Klenow fragment, by Asp705 and Asp882. Metal ion A activates the primer 3'-OH for nucleophilic attack on the α -phosphate of the dNTP. Metal ion B stabilizes the negative charge that develops on the pyrophosphate leaving group. Both metal ions also assist the reaction by stabilizing the negative charge on the α -phosphate in the pentacovalent transition state. Reproduced from Brautigam CA and Steitz TA (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Current Opinion in Structural Biology* 8: 54–63.

structures of a diverse variety of polymerases, suggesting that the formation of a closed ternary complex may be an important feature of the polymerase reaction pathway.

DNA Polymerase Reaction Mechanism

Extensive work by Benkovic and co-workers initially defined the polymerase reaction pathway of Klenow fragment. There is an obligatory order of substrate binding – DNA before dNTP – as would be expected, given that templating information is required for nucleotide selection. As mentioned above, the chemical step of dNTP addition is catalyzed by two divalent metal ions coordinated at the active site (Figure 3). One metal ion facilitates deprotonation of the primer 3'-OH, making it a better nucleophile, while the other stabilizes the pyrophosphate leaving group. Both metals also stabilize the developing negative charge on the α -phosphate in the transition state. This two-metal-catalyzed reaction mechanism is beautifully illustrated in the ternary complex co-crystal structures, and is

widespread in phosphoryl transfer reactions, including the 3'-5' exonuclease reaction (discussed later). The work of Benkovic and colleagues indicated that the phosphoryl transfer step is rapid, but is preceded and followed by slow noncovalent steps. Subsequent studies, using fluorescence as a reporter of conformational transitions, have identified at least two noncovalent steps that precede phosphoryl transfer. These steps, which include the fingers-closing transition, are not observed in the presence of a mispaired dNTP, implying the existence of early specificity-determining steps, which reject mispaired dNTPs or other inappropriate substrates.

More recently, the application of single-molecule fluorescence techniques to Klenow fragment has allowed observation of processes previously obscured when observing large populations of molecules in traditional ensemble kinetic studies. For example, we now know that, in the absence of bound substrates, the unliganded polymerase explores the same conformational motions that are used in the reaction pathway. In another study, observations of single Klenow fragment molecules copying a tethered DNA molecule have demonstrated the effects of local DNA sequence and secondary structure on polymerase behavior and emphasized the stochastic nature of processive DNA synthesis.

DNA Polymerase Reaction Specificity

Base pairing

DNA polymerases function with extraordinary specificity, selecting the complementary dNTP and rejecting incorrect choices in the majority of cases. Klenow fragment makes approximately one error for every 10^5 nucleotides synthesized, about 1000-fold more accurate than would be predicted solely from the energetics of base pairing. Three major mechanisms have been proposed to account for the way in which the polymerase enhances the accuracy of DNA synthesis. The first is exclusion of water from the active site, which should amplify the energetic differences between correct and incorrect base pairings. The second is geometric selection in the binding pocket for the nascent base pair, such that there is good steric complementarity to the rather symmetrical shape of a Watson–Crick base pair, but exclusion of mispairs. This model is consistent with the very snug fit seen in the binding pocket of polymerase ternary complex structures. Moreover, a variety of polymerases, including Klenow fragment, incorporate template–primer pairings that mimic the shape of a Watson–Crick base pair but lack any hydrogen bonds. The first two specificity mechanisms are proposed to operate at the level of nucleotide addition and would be expected to weaken the binding and/or slow the incorporation of an incorrectly paired incoming nucleotide. As mentioned earlier, a mispaired dNTP affects early steps of the polymerase reaction pathway, well before the chemical step of nucleotide addition. Any proposed selection mechanism would therefore come into play at an early fidelity checkpoint, most probably preceding the fingers-closing conformational change, so that specificity models based solely on the geometry of the binding pocket are too simplistic. Indeed, there is increasing evidence for a hypothetical preview step, in which the incoming dNTP is tested for complementarity to the templating base, before fingers-closing.

The third selection method would operate after nucleotide addition and relies on hydrogen-bonding interactions between the polymerase and the minor groove at or close to the primer terminus. The hydrogen bond acceptors on N₃ of purines and O₂ of pyrimidines occupy similar positions in all Watson–Crick base pairs, but are mispositioned in incorrect pairings, thus providing a mechanism for scanning the minor groove to ensure that template information has been correctly read. A mispair located at the primer terminus will fail to make the minor groove interactions that are seen for correct base pairs in polymerase co-crystal structures and this, together with other geometric abnormalities, results in a much slower rate of addition of the next nucleotide. Slower polymerase addition to a mispaired primer terminus reduces the likelihood of a mispair becoming fixed as a mutation by continued synthesis, and also increases the time window during which the error may be proofread, or the mispaired DNA may dissociate.

As might be expected from the different geometries of various mispairs, a polymerase such as Klenow fragment does not make all errors at the same frequency; for example, insertion of dGTP opposite a template T is relatively facile, whereas insertion of dCTP opposite C is rare. Klenow fragment mutator mutants change not only the frequency of errors but also the specificity, and their study provides a valuable window into the recognition processes that take place at the polymerase active site.

Sugar specificity

Bacterial Pol Is also have stringent specificity for the sugar of the incoming nucleotide. A DNA polymerase must reject ribonucleotides, despite their higher concentration *in vivo*, and this is achieved by an invariant Glu residue (the ‘steric gate’) whose position in the closed polymerase–DNA–dNTP ternary complex is incompatible with any substituent on the C2′ position of the sugar. Mechanistically, an incoming ribonucleotide triphosphate (rNTP) affects the rate and extent of the fingers-closing step, as expected from the structural description. Mutation of the steric gate Glu to the smaller Ala in Klenow fragment results in a polymerase that can add a ribonucleotide almost as easily as a deoxyribonucleotide. However, the mutant enzyme does not function as a true RNA polymerase because it cannot efficiently add multiple ribonucleotides, most probably because the duplex binding site cannot accommodate a DNA–RNA product.

Pol I enzymes fall into two classes regarding their handling of nucleotide analogs which lack the sugar 3′ hydroxyl (dideoxy nucleotides) and therefore function as chain terminators. The majority, exemplified by Klenow fragment, discriminate very strongly against dideoxy nucleotides. The ternary complex co-crystal structures show a hydrogen bond between the 3′ hydroxyl and the β-phosphate of the incoming nucleotide; presumably, the loss of this interaction compromises the alignment of reactive groups in the transition state, resulting in a slower reaction rate. A minority of Pol Is, exemplified by T7 DNA pol, discriminate only slightly against dideoxy nucleotides. In these polymerases, a conserved aromatic side chain that forms part of the nucleotide-binding pocket is present as Tyr, instead of the more usual Phe, and the Tyr hydroxyl provides the missing interaction with the

β-phosphate. The ability to manipulate chain-terminator specificity via a single active-site point mutation has proved to be invaluable in DNA sequencing and related biotechnology applications.

3′–5′ Exonuclease

The 3′–5′ exonuclease, present in about half of Pol Is, carries out its editing function by means of hydrolysis of the terminal nucleotide of the DNA primer strand. The hydrolysis reaction is a phosphoryl transfer, analogous to that described above for the polymerase reaction (Figure 3), catalyzed by a pair of metal ions liganded by a cluster of conserved carboxylate side chains. The other important component of the 3′ exo site is a binding site for single-stranded DNA. In model studies, the exonuclease can be studied using single-stranded DNA as the substrate, though the natural substrate *in vivo* is a duplex DNA whose primer terminus is frayed so as to present three or four bases of single-stranded DNA for binding at the 3′ exo site. The requirement for fraying in order to bind at the 3′ exo active site provides the specificity for editing, because a mismatched primer terminus is more easily melted and is therefore a better substrate for the 3′ exo than a correctly paired DNA. Whereas a correctly paired DNA duplex is bound predominantly at the polymerase site of Klenow fragment, a single terminal mismatch results in ≈50:50 partitioning between polymerase and 3′ exo sites. The preference of the 3′ exo for a mispaired substrate is amplified by the slowing of the polymerase reaction caused by a mispaired primer terminus (discussed earlier), so that a mispair is likely to be targeted for proofreading, while a correct primer terminus will serve as a substrate for continued addition.

The contribution of proofreading to polymerase fidelity is quite variable in the Pol Is. As already noted, some Pol Is do not proofread at all. In others, such as *E. coli* Pol I, the contribution of proofreading is relatively modest, about 10-fold. Others, such as the polymerases from bacteriophages T5 and T7, have more active exonucleases. Additionally, depending on the strength of the polymerase–DNA interaction, the DNA may stay associated with the polymerase during the transfer from polymerase to editing site, as in T5 and T7, or it may reach the proofreading site via dissociation from the polymerase site and binding to the exonuclease site of another enzyme molecule. As the selectivity of editing is determined simply by the structure of the DNA substrate (paired or mispaired) and does not require a particular mode of presentation of the primer terminus by the polymerase, the intermolecular transit from polymerase to exonuclease site would not be expected to compromise the editing process.

5′ Nuclease

The 5′ nuclease of the bacterial Pol I enzymes was described as the ‘5′–3′ exonuclease’ in the earlier literature. However, it is now recognized that this activity is not a true exonuclease, but is a structure-specific nuclease that recognizes the junction between duplex DNA and a single-stranded 5′ end, and cuts between the first two paired bases (Figure 4(a)). The 5′

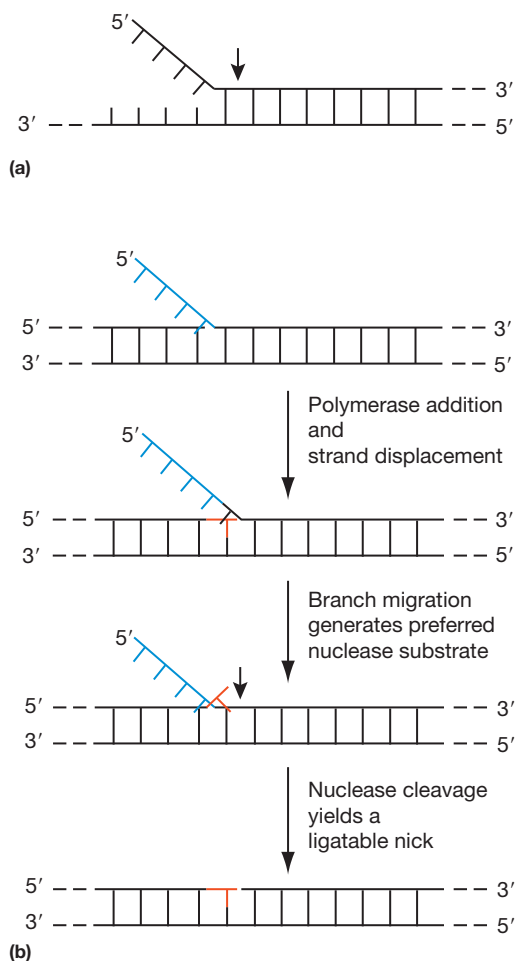


Figure 4 (a) Site of cleavage by the 5' nuclease on a model substrate containing noncomplementary 5' and 3' extensions. The nuclease cuts between the first two paired bases. (b) The polymerase and 5' nuclease activities collaborate in processing a DNA molecule with a 5' single-stranded tail so as to produce a nicked product that can be sealed by DNA ligase. Cleavage of the first structure, between the first two paired bases adjacent to the 5' tail, would yield a single-base gap. Nucleotide addition by the polymerase (in red), together with the preference of the 5' nuclease for a melted base at the 3' end of the adjacent primer strand, favors the formation of a ligatable product.

nucleases present in bacterial Pol Is are related to a larger family of autonomous structure-specific nucleases, or flap endonucleases (FENs), found in bacteria, bacteriophages, eukaryotes, and archaea.

The 5' nuclease of Pol I cooperates with the polymerase when acting on a nicked DNA molecule; nucleotide addition by the polymerase causes strand displacement which generates the substrate for the 5' nuclease, so that the net result is nick translation, the movement of a nick along a duplex DNA molecule. DNA-binding experiments show that the polymerase and 5' nuclease compete for binding to a DNA substrate; this implies a mechanism in which the DNA is passed from one active site to the other, and rules out mechanisms in which both reactions take place within a single bound species. The preferred substrate for the 5' nuclease has, in addition to the unpaired 5' strand, an unpaired base at the 3' end of the primer

strand abutting the ss-ds junction. This substrate could be formed by rearrangement of the product of polymerase extension and, when cleaved by the 5' nuclease, will leave a junction ready for ligation (see [Figure 4\(b\)](#)). The observed specificity of the 5' nuclease fits well with the *in vivo* functions of the bacterial Pol I enzymes in DNA repair and in the removal of primers from Okazaki fragments on the lagging strand. In both cases, the desired end point of Pol I action is a nick that can be sealed by DNA ligase.

Several 5' nuclease structures have been solved in the absence of substrates, and two co-crystal structures with bound DNA show the architecture of the enzyme-substrate complex. The 5' nuclease active site contains a large number of conserved carboxylate residues, inviting speculation that, like the polymerase and 3'–5' exonuclease activities, the 5' nuclease catalyzes phosphoryl transfer using bound divalent metal ions. However, where metal ions are present in 5' nuclease or FEN structures, the metal-to-metal distances are larger than would be appropriate for a simple two-metal-ion mechanism analogous to that shown in [Figure 3](#) (which requires a metal-to-metal distance of <4 Å). It is therefore possible that the 5' nuclease mechanism may differ from the polymerase and 3'–5' exonuclease reactions by requiring the direct participation of active-site side chains in the phosphoryl transfer process, and further studies will be needed to clarify this question.

See also: [Molecular Biology: DNA Polymerase III, Bacterial](#); [DNA Polymerases: Kinetics and Mechanism](#); [DNA Replication: Initiation in Bacteria](#).

Further Reading

- Benkovic SJ and Cameron CE (1995) Kinetic analysis of nucleotide incorporation and misincorporation by Klenow fragment of *Escherichia coli* DNA polymerase I. *Methods in Enzymology* 262: 257–269.
- Brautigam CA and Steitz TA (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Current Opinion in Structural Biology* 8: 54–63.
- Ceska TA and Sayers JR (1998) Structure-specific DNA cleavage by 5' nucleases. *Trends in Biochemical Sciences* 23: 331–336.
- Devos JM, Tomanicek SJ, Jones CE, Nossal NG, and Mueser TC (2007) Crystal structure of bacteriophage T4 5' nuclease in complex with a branched DNA reveals how flap endonuclease-1 family nucleases bind their substrates. *Journal of Biological Chemistry* 282: 31713–31724.
- Johnson SJ, Taylor JS, and Beese LS (2003) Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proceedings of the National Academy of Sciences of the United States of America* 100: 3895–3900.
- Joyce CM (2010) Techniques used to study the DNA polymerase reaction pathway. *Biochimica et Biophysica Acta* 1804: 1032–1040.
- Joyce CM and Benkovic SJ (2004) DNA polymerase fidelity: Kinetics, structure, and checkpoints. *Biochemistry* 43: 14317–14324.
- Kunkel TA and Bebenek K (2000) DNA replication fidelity. *Annual Reviews of Biochemistry* 69: 497–529.
- Santoso Y, Joyce CM, Potapova O, et al. (2010) Conformational transitions in DNA polymerase I revealed by single-molecule FRET. *Proceedings of the National Academy of Sciences of the United States of America* 107: 715–720.
- Schwartz JJ and Quake SR (2009) Single molecule measurement of the 'speed limit' of DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* 106: 20294–20299.

DNA Polymerase III, Bacterial

H Maki and A Furukohri, Nara Institute of Science and Technology, Ikoma, Japan

Published by Elsevier Inc.

This article is a revision of the previous edition article by Hisaji Maki, volume 1, pp. 729–733, © 2004, Elsevier Inc.

Glossary

AAA+ ATPase family An extension of the AAA (for ATPases associated with a variety of cellular activity) family, originally defined to include proteins with a common ~200-residue ATPase core.

Mismatch repair A class of DNA repair that corrects mispairs and small bulge structures in DNA, which are caused mainly by replication errors.

Mutator A phenotype showing an increased rate of spontaneous mutation, which is caused by a mutation

within one of the genes for maintenance of replicational fidelity or repair of spontaneous damage to DNA and substrate nucleotides.

Nucleotide excision repair A class of DNA repair by which a segment of DNA containing lesions is excised and replaced with resynthesized DNA.

Translesion DNA synthesis A cellular process involving specialized DNA polymerases that counteract the replication-blocking damage to DNA.

Replicative Apparatus in Bacterial Cells

In bacterial cells, the circular chromosome contains a unique origin, and DNA replication proceeds bidirectionally from the origin to the terminus. Replication of the whole bacterial genome (4700 kb for *Escherichia coli* (*E. coli*)) is continuous from the origin to the terminus, and is accompanied by movement of the replicating point, called the replication fork. Both parental DNA strands are concurrently replicated at the fork. As DNA polymerase can extend a DNA chain only in the 5'→3' direction, replication at a fork is semi-discontinuous: DNA synthesis is continuous on one strand (the leading strand) and discontinuous on the other (the lagging strand). Short pieces of DNA, called Okazaki fragments, are repeatedly synthesized on the lagging-strand template. These Okazaki fragments are a few thousand nucleotides long in bacterial cells.

DNA replication is a complex process, involving numerous enzymes at the replication fork. In *E. coli*, more than 20 different proteins participate in DNA replication. Among these proteins, DnaB, DnaG, and Pol III HE are the three basic components acting at the replication fork, forming a multiprotein complex called the replisome (Figure 1). The DnaB protein is a major replicative DNA helicase which moves along the lagging strand in the 5'→3' direction and opens up the duplex DNA at the tip of the replication fork to expose a pair of single-stranded DNA templates. The DnaG protein is a primase which synthesizes a short RNA to prime DNA chain elongation catalyzed by Pol III HE.

The highly organized and remarkably efficient process needed for replicating chromosomal DNA is achieved by physical and functional interactions among these three components. Pol III HE associates with DnaB protein and pushes the helicase forward as it proceeds along the leading strand. Thus, the velocity of replication fork movement is determined mainly by the rate of chain elongation by Pol III HE, about 1000 bp s⁻¹. DnaB interacts with DnaG and activates cyclically the primase activity as the helicase moves along the lagging strand, resulting in timely initiation of Okazaki fragment

synthesis. Finally, the proper recruitment of Pol III HE to the replisome requires an RNA primer formed within the origin and, thus, depends upon the function of DnaG.

In bacterial cells, Pol III HE and the other components of the replisome localize to discrete positions, predominantly at or near the midcell throughout the cell cycle. It is considered that the DNA is threaded through the centrally positioned replisome during duplication, and then extruded from the replisome after duplication.

Subunits and Subassembly of Pol III Holoenzyme

The machine-like action of Pol III HE at the replication fork, far more complex than what is necessary for mere DNA polymerization, is based on its subunit structure. Ten different polypeptides form an isolable 17-subunit Pol III HE complex in *E. coli*. The whole assembly is very stable while participating in the replisome at the replication fork. However, during the course of purification, Pol III HE readily disassembles into free subunits and various subassemblies, analyses of which have greatly helped clarify the biochemical function of each subunit and the architecture of the holoenzyme (Figure 2). Smaller numbers of subunits assemble to form the holoenzyme in *Bacillus subtilis* and other bacterial species, although the basic architecture is conserved in most bacteria.

Pol III Core

The α subunit, the largest polypeptide among the holoenzyme subunits, is the heart of the enzyme in the form of DNA polymerase. It exhibits DNA polymerase activity and forms a catalytic core subassembly, called Pol III core, with one molecule each of the ϵ and θ subunits. The α subunit has a characteristic structure, with a right-hand shape composed of palm, thumb, and finger domains, which is common in all other types of DNA polymerases. In the palm domain, as in other DNA polymerases, the α subunit has the polymerase

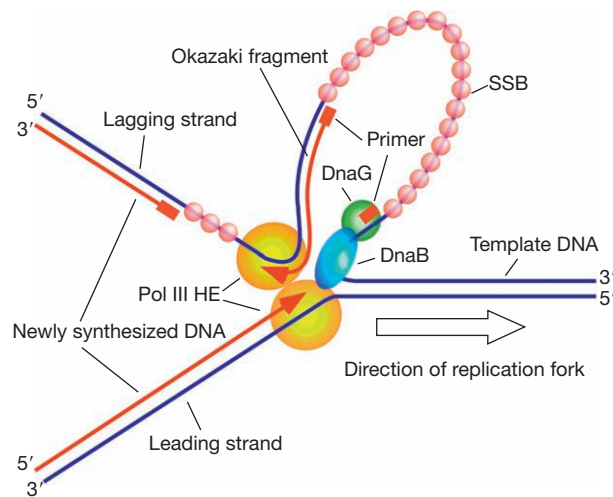


Figure 1 The replication fork and three basic components of the replicative apparatus in bacteria.

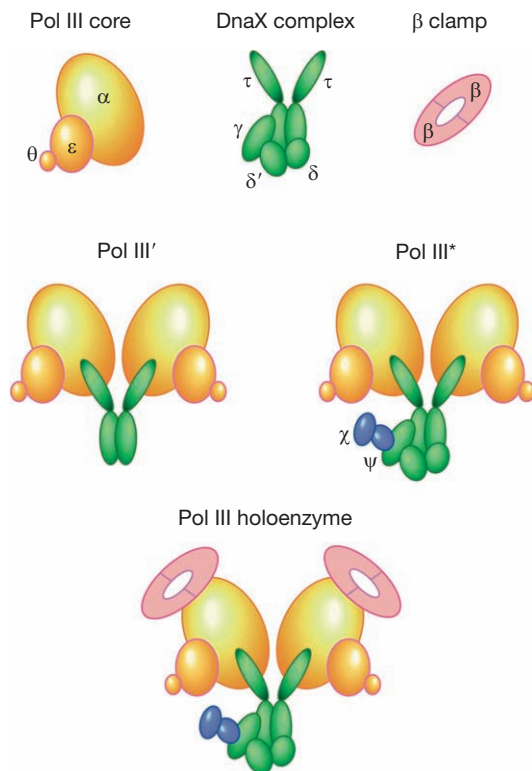


Figure 2 A schematic view of the architecture of DNA polymerase III holoenzyme.

motif, an amino acid sequence motif, which consists of six segments each containing 5–10 residues. The polymerase motif corresponds to catalytically important structural determinants for the active center of DNA polymerases.

Although the basic structure of DNA polymerases is conserved, the DNA polymerases are classified into six families, A, B, C, D, X, and Y, according to their structural similarity. All the catalytic subunits of replicative DNA polymerases in eukaryote, archaea, and bacteriophage belong to the family B,

while the α subunit of Pol III belongs to a distinct family, family C. The family C DNA polymerase is more closely related to the family X, which is known as a member of the human Pol β -like nucleotidyltransferase superfamily, rather than the family B. In contrast to the family B DNA polymerases, the α subunit of Pol III lacks the 3'→5' exonuclease activity required for the proofreading function, which removes nucleotides that have been incorrectly inserted by the polymerase. In addition, the α subunit is a distributive enzyme and it can extend fewer than 10 nucleotide residues for each DNA-binding event.

Despite the feeble activity of the α subunit, Pol III core has high catalytic efficiencies for DNA synthesis and proofreading. This is achieved by forming a complex with other subunits. In the Pol III core, the ϵ subunit acts as the editing exonuclease together with the θ subunit and ensures highly accurate DNA synthesis by the polymerase. Cells that are defective in ϵ subunit function show a strong mutator phenotype, with remarkably elevated frequencies of spontaneous mutation. Despite the catalytic activity of Pol III core being higher than that of the α subunit, the subassembly has limited processivity of DNA synthesis due to its low affinity to template DNA. Other subunits, especially the β subunit, greatly enhance and regulate the processivity of the core.

τ Subunit and Pol III'

The τ subunit is the second largest subunit and forms a stable homodimer complex when it is free in solution. An important function of the τ subunit is to make the Pol III core into a dimer. The C-terminal domain of the τ subunit interacts with the α subunit, and the dimeric nature of the τ subunit itself facilitates dimerization of the Pol III core. The resulting subassembly, called Pol III', consists of two Pol III cores and two τ subunits. The τ subunit improves the low processivity displayed by the Pol III core, resulting in a moderately increased processivity for Pol III'.

Besides dimerization of the Pol III core, the τ subunit plays multiple roles at the replication fork. In the HE complex, the τ dimer forms bridges between dimeric polymerases and other auxiliary subunits. In addition, as a scaffold of the replisome, the τ dimer connects Pol III HE and the DnaB helicase at the replication fork.

β Subunit and Pol III*

Among the holoenzyme subunits, the β subunit is the one that most easily dissociates from the holoenzyme complex. Phosphocellulose or cation-exchange column chromatography effectively separates the holoenzyme into the β subunit and Pol III*, a subassembly that retains all the subunits except β . However, the association of the β subunit with Pol III* is very stable when it forms a ternary complex with template-primer DNA.

The β subunit is a crescent-shaped polypeptide, and it forms a homodimer with a doughnut-like structure which encircles double-stranded DNA and slides along the DNA. The β dimer binds to the α subunit, even in the absence of other HE subunits, and it tethers the α subunit to template DNA during DNA synthesis. The ability of Pol III HE to perform highly processive DNA synthesis is based on the sliding-clamp nature of the β subunit and the α - β interaction. The β subunit also interacts with the δ subunit in an adenosine triphosphate

(ATP)-dependent manner when β is functioning in the HE complex. Albeit much more weakly than its interaction with the α subunit, the β dimer can associate with other proteins including all the other *E. coli* DNA polymerases, a mismatch repair protein (MutS), DNA ligase, and a replication-initiator protein (DnaA). Involvement of the β subunit in many aspects of DNA transactions other than DNA replication has been suggested.

Pol III*, the largest subassembly of Pol III HE, consists of two Pol III cores, two molecules of τ subunit, and one molecule

each of the γ , δ , δ' , χ , and ψ subunits. The architecture of Pol III* is semi-symmetrical. As in Pol III', the τ dimer connects two Pol III cores in a symmetrical configuration. On the other hand, the other five auxiliary subunits impose an asymmetry on the dimeric polymerases. Among these auxiliary subunits, γ , δ , and δ' form a pentameric circular complex (the DnaX complex) with the τ dimer, while the χ and ψ subunits form another complex that bridges the DnaX-complex and the single-stranded DNA-binding protein (SSB).

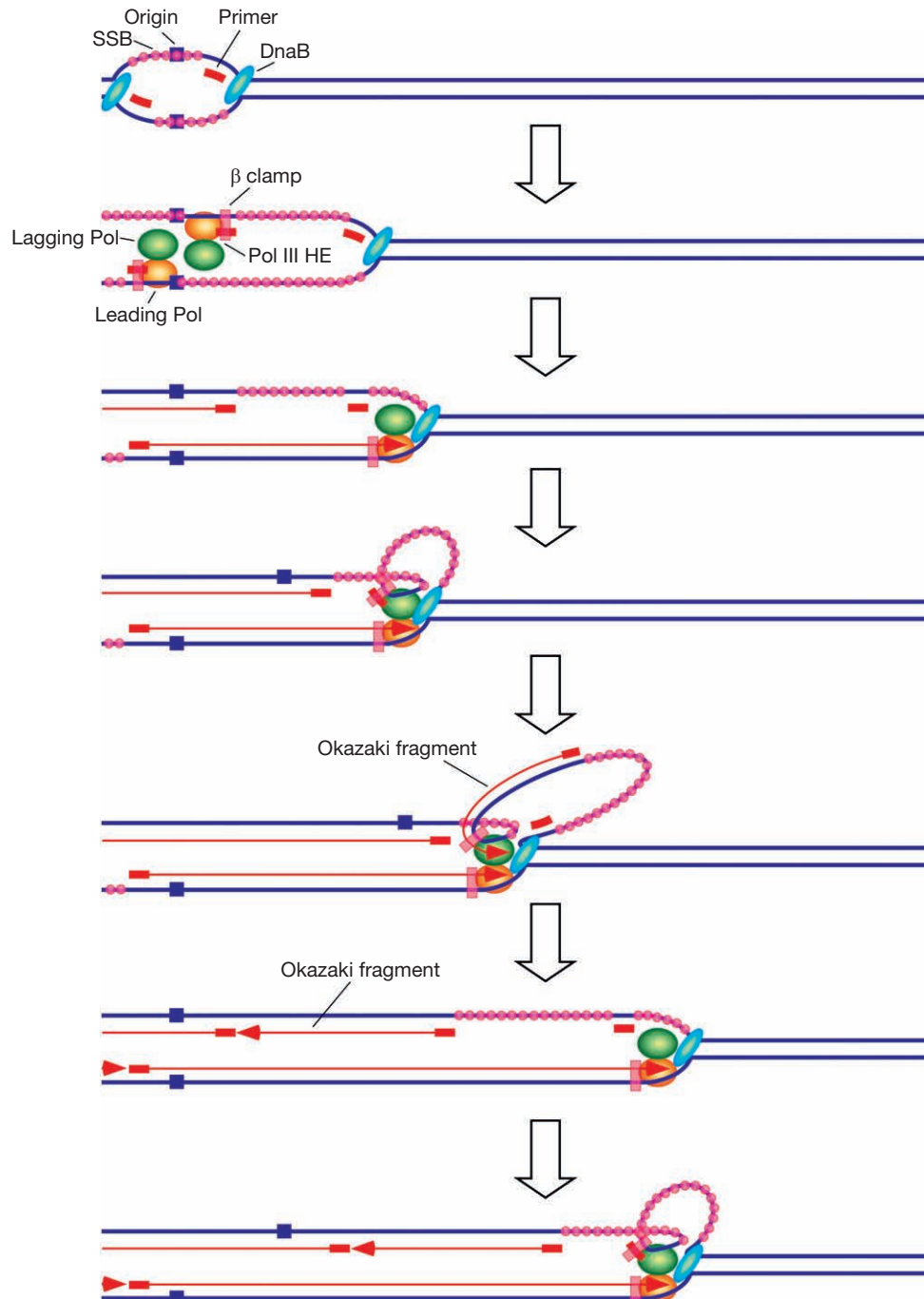


Figure 3 Concurrent DNA synthesis of leading and lagging strands by the bacterial replisome.

DnaX Complex

The biochemical function of the DnaX complex is to load the β clamp on DNA. The δ subunit directly interacts with the β subunit and opens the β ring. Other subunits assist and regulate the action of the δ subunit. The τ and γ subunits are encoded by the *dnaX* gene and, thus, are called DnaX proteins. τ is the full-length translation product, while γ is a truncated protein that arises by translational frameshifting. Hence, the N-terminal portion of τ is identical to that of the γ subunit. Furthermore, all the subunits of the DnaX complex are structurally related to the AAA+ ATPase family, although only τ and γ show ATPase activity.

ATP binding and hydrolysis of ATP are crucial events for the clamp-loading action of the DnaX complex. In the absence of ATP, the five-subunit circular complex is in a tightly closed configuration, and inert for loading the β subunit. In particular, the δ' subunit blocks the δ subunit's interaction with the β subunit. When all three DnaX proteins in the complex bind ATP, the DnaX complex attains a more relaxed form in which the δ subunit readily gains access to the β subunit. Upon association with the δ subunit, the β ring is opened and mounted on primer-template DNA. Contact between the β -DnaX complex and DNA activates hydrolysis of ATP bound to the DnaX complex, which leads to dissociation of β from the δ subunit, leaving the closed β ring on the primer-template DNA.

Pol III HE in Other Bacteria

Genes encoding the subunits of Pol III HE have been identified in the genomes of many species of eubacteria. Among these genes, those encoding the α , β , τ , δ , and δ' subunits are well conserved, while those for other subunits seem to be significantly divergent or lost. In many bacterial species, the *dnaX* gene does not show an obvious frameshifting signal sequence and, therefore, likely produces only the τ subunit as a single DnaX protein. A group of bacterial species (Gram-positive, G + C low), including *B. subtilis*, *Staphylococcus aureus*, and *Streptococcus pyogenes*, possesses two distinct family C DNA polymerases, both of which are required for chromosomal replication. Despite the apparent diversity among different species, analyses of Pol III HE in bacterial species that are evolutionarily distant from *E. coli* have demonstrated that the basic architecture and biochemical functions of Pol III HE are common throughout the eubacteria.

Concurrent DNA Synthesis of Leading and Lagging Strands

In vitro reconstitution of the replisome with purified Pol III*, β subunit, DnaB helicase, DnaG primase, and SSB has indicated that a single Pol III HE particle can simultaneously

synthesize both leading and lagging strands at a replication fork. Pol III HE is a functionally and structurally asymmetric complex with twin polymerases, one of which participates in leading-strand synthesis and the other in lagging-strand synthesis. The leading-strand polymerase remains continuously clamped to DNA, but the lagging-strand polymerase is repeatedly clamped and unclamped from DNA, to cycle from one Okazaki fragment to the next. As the lagging-strand polymerase is held at the replication fork via the τ -subunit bridge and the leading-strand polymerase, retargeting of the lagging-strand polymerase to the next primer is so efficient and rapid that the overall rate of lagging-strand synthesis matches the rate of leading-strand synthesis (Figure 3).

Pol III HE is initially loaded on the chromosome DNA after the first RNA primer is synthesized within or near the replication origin. One polymerase clamped on the initial RNA primer becomes the leading polymerase, and the other polymerase in the same HE complex serves as the lagging polymerase. An asymmetric arrangement of the DnaX complex in Pol III HE is the basis for the functional asymmetry of the replicative polymerase. It has been hypothesized that the position of the DnaX complex is biased to the lagging polymerase, which limits its clamp-loading activity exclusively to the lagging strand.

Other Biological Functions

Almost all plasmids, and the majority of bacteriophages, use Pol III HE for their DNA replication. Pol III HE is involved in several kinds of DNA transactions that require synthesis of relatively long segments of DNA. These include long-patch nucleotide excision repair, *mutS*-dependent mismatch repair, homologous recombination, and replicative translocation of transposons.

See also: **Molecular Biology: DNA Polymerase I, Bacterial; DNA Polymerases: Kinetics and Mechanism; DNA Replication Fork, Bacterial; DNA Replication: Initiation in Bacteria; Processivity Clamps in DNA Replication.**

Further Reading

- Johnson A and O'Donnell M (2005) Cellular DNA replicases: Components and dynamics at the replication fork. *Annual Reviews in Biochemistry* 74: 283–315.
- Kelman Z and O'Donnell M (1995) DNA polymerase III holoenzyme: Structure and function of a chromosomal replicating machine. *Annual Reviews in Biochemistry* 64: 171–200.
- Kornberg A and Baker T (1992) *DNA Replication*, 2nd edn. New York, NY: Freeman.
- Marians KJ (1992) Prokaryotic DNA replication. *Annual Reviews in Biochemistry* 61: 673–719.
- McHenry CS (1988) DNA polymerase III holoenzyme of *Escherichia coli*. *Annual Reviews in Biochemistry* 57: 519–550.

DNA Polymerases: Kinetics and Mechanism

K A Johnson, University of Texas, Austin, TX, USA

Published by Elsevier Inc.

This article is reproduced from the previous edition, volume 1, pp. 734–739, © 2004, Elsevier Inc.

Glossary

Discrimination The ratio of the specificity constant for a correct base pair divided by that for a particular mismatch.

Fidelity The reciprocal of the error frequency; for example, an error frequency of 0.000 002 is expressed as a fidelity of one error in 500 000.

Proofreading exonuclease An exonuclease that removes a single base from the 3' end of the primer strand.

Selectivity The fidelity contribution of individual constants; for example, the selectivity in the ground state binding is

equal to the ratio of the K_d for the incorrect deoxynucleoside triphosphate (dNTP) divided by the K_d for the correct dNTP binding.

Specificity constant The apparent second-order rate constant for nucleotide binding and incorporation, defined by the steady-state kinetic parameters, k_{cat}/K_m , but more accurately measured using pre-steady-state kinetic methods to define k_{pol}/K_d , where k_{pol} is the maximum rate of incorporation and K_d is the apparent ground state nucleotide dissociation constant.

Polymerases can replicate DNA with extraordinary speed and fidelity, copying a template strand at a rate of 300 base pairs per second and making a mistake only one time out of a million. When the polymerase does make a mistake, it stalls by slowing the incorporation of the next correct base pair on top of the mismatch, giving time for a proofreading exonuclease to remove the mismatched base. Selective removal of mismatched bases by the proofreading exonuclease contributes an additional factor of ~ 1000 , resulting in a net fidelity of approximately one error in a billion bases copied. Not all polymerases achieve such high fidelity, however. Rather, each polymerase has evolved a fidelity that balances the biological needs for stability and adaptability. Although the fidelity varies greatly among the polymerases that have been examined in detail, the pathway by which a correct base pair is selected remains invariant. This article briefly summarizes our understanding of the kinetic, structural, and thermodynamic bases governing the fidelity of DNA replication.

Elementary Steps in Polymerization

The free energy difference between a correct and incorrect base pair is quite small ($1\text{--}2\text{ kcal mol}^{-1}$), which leads to a selectivity factor of only 5–30 in favoring the correct base pair over a mismatch. Therefore, the polymerase does not simply stitch together base pairs that form in solution; rather, the fidelity of DNA replication is largely a function of the kinetics of the reactions, involving nucleotide binding, recognition, and incorporation. The polymerase uses not only the base pair free energy but also the base pair geometry in selecting the correct base pair. The new base pair is buried at the active site of the polymerase, shielding it from water molecules, which enhances the free energy difference between correct and incorrect base pairs and facilitates rapid catalysis.

DNA polymerases achieve their extraordinary fidelity by using a two-step nucleotide-binding sequence to select the proper nucleotide for incorporation, as shown in [Figure 1](#).

The deoxynucleoside triphosphate (dNTP) initially binds in a rapid equilibrium reaction to an open state of the enzyme to form the ground state complex (E-DNA-dNTP). A correct base pair enables a change in enzyme structure (at rate k_2) to a closed form in which the active catalytic residues are brought in to the proper orientation necessary for catalysis of the chemical reaction (at rate k_3). This two-step nucleotide-binding sequence is important for two reasons. First, the binding of the nucleotide to the open enzyme form allows a rapid selection of the correct base pair from the four competing nucleotides in solution, using base pair free energy to favor the correct base. Second, the conformational change in the enzyme to the closed form provides additional selectivity dependent upon the proper base pair geometry, and it leads to a close alignment of protein residues around the reactants, shielding them from solvents and bringing about rapid catalysis. A mismatched base pair is discriminated against at each step in the sequence. An incorrect base binds weaker in the ground state, inhibits the rate of the conformational change, and may lead to a misalignment of the reactive groups necessary for catalysis, leading to a much slower rate of incorporation.

Selectivity Contributions of Each Step

Fidelity varies greatly for different DNA polymerases. The DNA polymerase responsible for replicating the viral genome of human immunodeficiency virus (HIV), known as 'reverse transcriptase (RT)' because it copies RNA templates as well as DNA, has a much lower fidelity, making one mistake in only 10 000 base pairs. This higher error rate is essential for the survival of HIV by affording a fast mutation frequency that gives the virus the ability to evade the immune system and all modern drugs, by constantly changing and evolving resistance via the process of natural selection. By contrast, the human DNA polymerase responsible for replicating the mitochondrial genome has an intermediate fidelity, making one error in 300 000; with proofreading, the error frequency is increased to one error in 1.5 million.

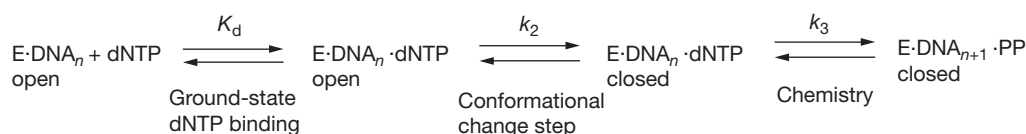


Figure 1 Pathway of nucleotide binding and incorporation.

Table 1 Fidelity contributions of ground-state binding and polymerization^a

Polymerase	$K_{d,\text{inc}}/K_{d,\text{cor}}$	$k_{\text{pol,cor}}/k_{\text{pol,inc}}$	Overall fidelity
T7 DNA polymerase	300	5000	1 500 000
Pol γ DNA polymerase	200	1500	300 000
Klenow	170	2300	210 100
Pol β	290	610	19 700
HIV RT	260	50	13 000

^aThe average contribution of ground state binding ($K_{d,\text{inc}}/K_{d,\text{cor}}$) and the maximum rate of polymerization ($k_{\text{pol,cor}}/k_{\text{pol,inc}}$) to the net fidelity are shown for various polymerases, where inc represents an incorrect while cor represents a correct nucleotide, respectively.

Table 2 Kinetics of polymerization

Polymerase	K_d (nM)	k_{pol} (s^{-1})	k_{pol}/K_d ($\mu\text{M}^{-1}\text{s}^{-1}$)	k_{off} (s^{-1})	Processivity	$K_{d,\text{DNA}}$ (nM)
T7 DNA polymerase	18	300	15	0.2	1500	18
Pol γ DNA polymerase	0.8	40	50	0.02	2250	10
Klenow	5	50	5	0.2	250	5
Pol β	10	10	1	0.3	30	50
HIV RT	4	30	8	0.2	150	5

Table 1 shows the net selectivity observed in each step of the reaction for several polymerases. It is interesting to note that each of the polymerases shows a nearly constant selectivity of approximately 200–300 for the ground state nucleotide binding, whereas the selectivity afforded by the conformational change step (and ensuing chemistry step) contributes a widely varying selectivity. For a high-fidelity enzyme such as T7 DNA polymerase, the selectivity in the second step is 5000, leading to a net fidelity of 1.5 million. By contrast, the HIV RT shows a selectivity factor of only 50 in the conformational change step, leading to an average overall fidelity of only 13 000. Effectively, HIV RT makes at least one error each time it replicates the 9 kb viral genome, allowing the virus to rapidly evolve to evade the host immune system and any single anti-viral drug. Its low fidelity is due to the small contribution of the conformational change and/or chemistry toward the net fidelity. Moreover, because HIV RT lacks a proofreading exonuclease, once a mismatch is incorporated it is not readily removed.

Measurement of the Kinetics of Incorporation

Measurement of the kinetics of DNA polymerization is complicated by the processivity of the enzyme, defined by the tendency of the polymerase to remain bound to the DNA template/primer and to continue multiple rounds of polymerization (**Table 2**). The processivity is calculated by the rate of polymerization divided by the rate of dissociation of the E-DNA complex and is equal to the average number of bases incorporated before the complex dissociates. Although polymerization can be quite fast (50–300 s^{-1}), the rate of

dissociation is slow (typically 0.02–0.2 s^{-1}). Therefore, in the measurement of the kinetics of incorporation of a single nucleotide in the steady state with an excess of DNA and limiting enzyme, the rate that is measured is due solely to the dissociation reaction and provides no information pertaining to the nucleotide incorporation reaction, unless incorporation is slower than DNA dissociation. Therefore, steady-state rate measurements are virtually meaningless. To circumvent these difficulties, the kinetics of incorporation are measured by examining the rate of extension using single-turnover kinetic methods. A stoichiometric E-DNA complex is rapidly mixed with the correct dNTP and Mg^{2+} and then quenched by mixing with ethylenediaminetetraacetic acid to chelate metal ions needed for catalysis. The time course of extension of the DNA by one base pair is then quantified after resolution on a DNA sequencing gel.

To resolve the time dependence of extension, the reaction must be examined on the timescale of a single turnover, which usually is in the range of milliseconds. Definitive experiments require a rapid mixing device to achieve rapid mixing of small volumes. Compared to experiments using steady-state kinetic methods, single-turnover kinetic experiments may be slightly more difficult to perform, but they can be interpreted directly and often unambiguously. The nucleotide concentration dependence of the rate of the incorporation in the single-turnover reaction defines the maximum rate of incorporation, k_{pol} , which may be limited by k_2 , k_3 , or a combination of the two steps (**Figure 1**) and the ground state nucleotide dissociation constant, K_d . Because the nucleotide binding is a rapid equilibrium reaction and k_{pol} is the single, rate-limiting step in polymerization, the ratio of k_{pol}/K_d is equal to k_{cat}/K_m , the specificity constant for the reaction. Thus, this one experiment

defines the key kinetic parameter governing nucleotide selectivity and resolves it into the two steps contributing to fidelity, the ground state nucleotide dissociation constant and the rate of incorporation, k_{pol} . Most importantly, these methods provide a measurement of the equilibrium constant for the binding of the substrate to the enzyme in a complex that is poised for catalysis.

Structural Determinants of Fidelity

The structure of T7 DNA polymerase in the closed complex is shown in Figure 2. The E-DNA-dNTP closed complex was formed with the correct dNTP and a dideoxy-terminated DNA primer/template to prevent the chemical reaction from occurring. It shows the tight arrangement of the protein around the DNA and the dNTP poised for the chemical reaction to proceed. The DNA lies largely on the surface of the protein, making contacts through the phosphodiester backbone. As the DNA approaches the active site, the structure changes from standard B-DNA to an A-form DNA with a more open minor groove that allows contacts with several residues that are thought to be important for sensing mismatches in the incoming base pair as well in the primer/template. Also important to note is that the next residue in the template strand is rotated out from the active site at 90° , and protein residues stack with both the templating base and the incoming dNTP. These contacts are all thought to be important in enforcing the proper base pair geometry.

Comparison of the T7 DNA polymerase closed complex with the structures of other polymerases in the E-DNA state

reveals that there is a large conformational change in a nucleotide recognition domain involving a 45° rotation to bring key residues into contact with the dNTP at the active site. Included among these residues are positively charged lysine and arginine, amino acids that bind to the β - and γ -phosphates to facilitate catalysis, and a tyrosine that stacks with the dNTP to help align the reaction center. This conformational change step provides the most significant contribution to nucleotide selectivity by sensing the proper base pair geometry and leading to fast catalysis of the chemical reaction.

Chemistry of Catalysis

A close-up view of the reaction center is shown in Figure 3. Two tetrahedrally coordinated Mg^{2+} ions are ligated by conserved acidic residues at the active site (D475 and D654), orient the reactants, and facilitate catalysis. Metal A activates the $3'$ OH and is ligated to the non-bridging oxygen of the α -phosphate of the incoming dNTP, while metal B is ligated to nonbridging oxygens of α -, β -, and γ -phosphates and stabilizes the pyrophosphate-leaving group as the reaction proceeds. This two-metal ion mechanism appears to be common for all polymerases.

Catalysis is also critically dependent upon contacts made involving residues from the recognition domain. Positively charged residues K522, H506, and R518 contact the α -, β -, and γ -phosphates, respectively; and Y526 stacks against the incoming dNTP. These contacts are formed following the conformational change and facilitate catalysis by binding and orienting the dNTP for reaction and stabilizing the

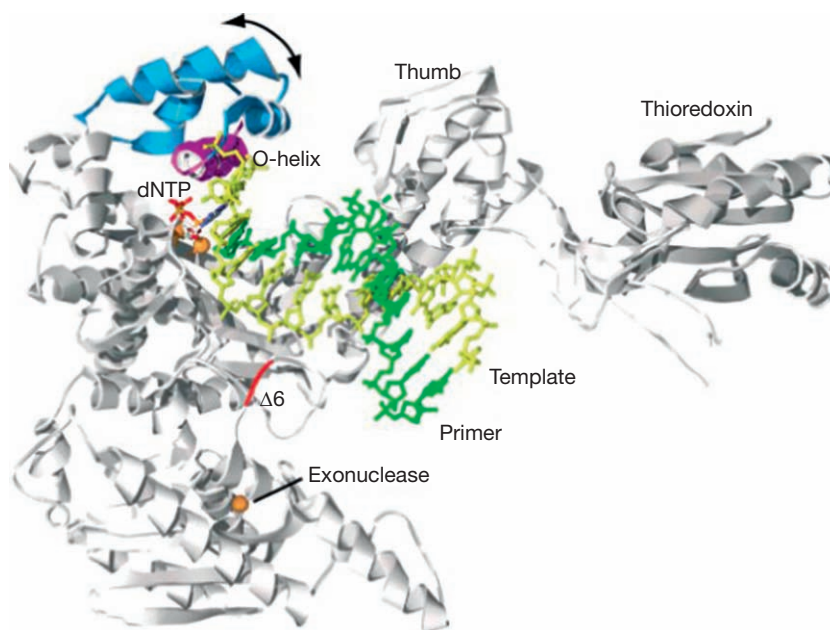


Figure 2 Structure of T7 DNA polymerase. The structure shows the closed complex resulting from the rotation (arrow) of the recognition domain (blue) to bring the O-helix (magenta) into contact with the incoming dNTP (CPK colors). The template strand is shown in yellow, the primer strand in green, and the thioredoxin accessory protein in gray. The two metal ions at the polymerase active site and the single metal ion at the exonuclease site are shown in gold. The location of the six-residue deletion ($\Delta 6$) to create the exo-mutant is shown in red. Drawn from Protein Data Bank structure file 1T7P.

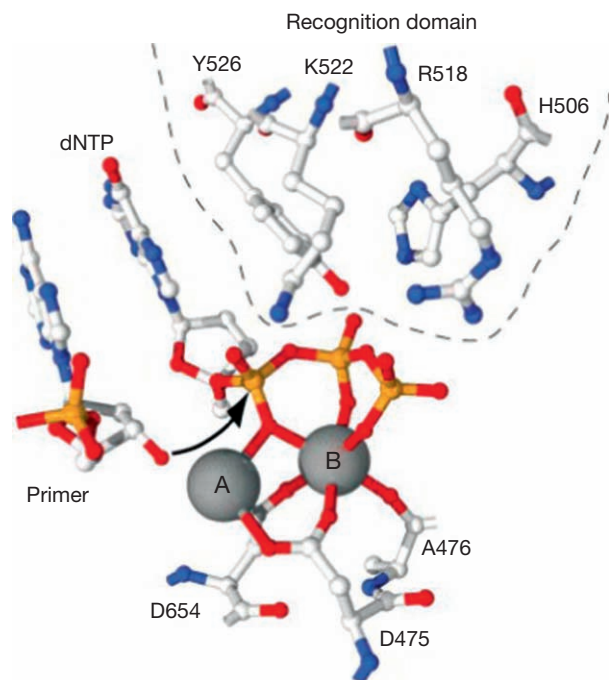


Figure 3 Two-metal ion mechanism. A close-up view of the active site shows the 3' OH of the terminal base in position to react with the incoming dNTP, bound to the two metal ions. Residues from the recognition domain contact the incoming dNTP to facilitate catalysis. Drawn from Protein Data Bank structure file 1T7P.

development of negative charge on the pyrophosphate. Thus, the structure provides a rationale for the importance of the conformational change in nucleotide selectivity and incorporation efficiency.

Selectivity of the Proofreading Exonuclease

The polymerase contains a proofreading function to efficiently remove mismatches after they are formed. This is accomplished by the polymerase recognizing it has made an error by stalling in its attempt to insert the next correct base on top of the mismatch to give time for the primer strand to flip over into the exonuclease site to have the 3'-terminal base excised. Although the exonuclease active site is 25 Å away from the polymerase site (see Figure 2), the DNA primer strand is able to flip back and forth between the two sites.

The selectivity of the exonuclease is governed by kinetic partitioning between the polymerase and exonuclease active sites, with the decision to excise being made at the polymerase site. As summarized in Figure 4, during successive correct nucleotide incorporation reactions, the polymerase continues down the DNA template, inserting bases at a rate of 300 s^{-1} . Occasionally, the DNA will dissociate from the polymerase or flip into the exonuclease site at a rate of 0.2 s^{-1} . Therefore, the cost of having a proofreading exonuclease site next to the polymerase active site is $0.2/(300 + 0.2) = 0.07\%$ of the correct nucleotides removed after correct incorporation. However, after the polymerase inserts a mismatched dNTP (occurring at an infrequent rate of 0.002 s^{-1}), the rate of incorporation of the next correct base on top of the mismatch is reduced to

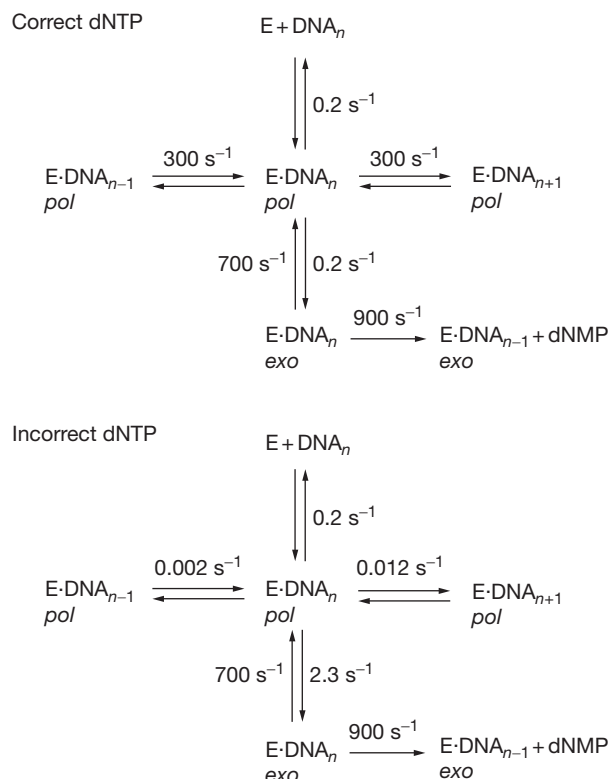


Figure 4 Kinetic partitioning for exonuclease proofreading. The changes in the rates of polymerization and exonuclease removal for correct and incorrect base pairs are summarized. Correct base pairs are incorporated sequentially at a rate of 300 s^{-1} , while occasionally the DNA will dissociate (0.2 s^{-1}) or slide into the exonuclease site (0.2 s^{-1}). Mismatches are incorporated at a slow rate (0.002 s^{-1}) and extension to incorporate the next correct base is also slow (0.012 s^{-1}), but the rate of sliding into the exonuclease site is increased to 2.3 s^{-1} . This change in kinetic partitioning leads to selective removal of mismatches based upon testing for proper base pairing at the polymerase site.

0.012 s^{-1} , while the rate at which the primer flips over into the exonuclease site is increased to 2.3 s^{-1} . The probability of removing the mismatch increases to $2.3/(2.3 + 0.012) = 99.5\%$. Thus, the kinetic partitioning defining the relative probability of reaction at the polymerase and exonuclease sites is inverted when the polymerase encounters a mismatch in the primer/template.

The dynamics of the reaction provides a perfect solution to the problem of designing a proofreading function. A single-stranded primer in the exonuclease site is needed to excise the 3'-terminal base, but the DNA must be in duplex form in order to know whether the base pair is correct. According to the kinetics, the DNA spends most of the time at the polymerase site and makes only brief excursions to the exonuclease site, by melting out a single-stranded segment of the primer strand sufficient to reach the exonuclease site 25 Å away. The decision as to whether to remove a base pair is made solely at the polymerase site, based upon the ability of the enzyme to sense a mismatch in the primer template and inhibit the rate of forward reaction, thereby changing the probability of excision versus extension.

It is interesting to note that the binding of the next correct base pair is not significantly impaired with a mismatch in the primer/template; rather, the rate of the conformational change and/or the chemical reaction is reduced by the presence of a mismatch. This again is in keeping with the postulate that one purpose of the conformational change step is to test for mismatches. The proofreading selectivity demonstrates that the polymerase is capable of sensing a mismatch not only in the incoming base pair but also at the adjacent position in the primer/template, and this ability is used to increase the selectivity at each step.

Summary

DNA polymerases are remarkable machines central to the replication of all life forms on the Earth. The fidelity of different DNA polymerases varies widely in order to meet the requirements of the biology in balancing the conflicting needs for stability and adaptability. All polymerases studied in sufficient detail show a similar two-step binding mechanism to select the correct base pair using both the base pair free energy and base pair geometry.

See also: **Molecular Biology:** DNA Polymerase I, Bacterial; DNA Polymerase III, Bacterial; DNA Polymerase β , Eukaryotic; DNA Polymerases: Reverse Transcriptase Integrase, and Retrovirus Replication; Eukaryotic DNA Polymerase α ; Eukaryotic DNA Polymerase δ ; UmuC D Lesion Bypass DNA Polymerase V.

Further Reading

- Doublie S, Tabor S, Long AM, Richardson CC, and Ellenberger T (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* 391: 251–258.
- Johnson AA and Johnson KA (2001) Fidelity of nucleotide incorporation by human mitochondrial DNA polymerase. *Journal of Biological Chemistry* 276: 38090–38096.
- Johnson AA, Ray AS, Hanes J, et al. (2001) Toxicity of antiviral nucleoside analogs and the human mitochondrial DNA polymerase. *Journal of Biological Chemistry* 276: 40847–40857.
- Johnson KA (1993) Conformational coupling in DNA polymerization. *Annual Review of Biochemistry* 62: 685–713.
- Lee H, Hanes J, and Johnson KA (2003) Toxicity of nucleoside analogs used to treat AIDS and the selectivity of the mitochondrial DNA polymerase. *Biochemistry* 42: 14711–14719.
- Spence RA, Kati WM, Anderson KS, and Johnson KA (1995) Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. *Science* 267: 988–993.

DNA Polymerases: Reverse Transcriptase Integrase, and Retrovirus Replication

M-L Andréola, V Parissi, and S Litvak, UMR 5234 CNRS, Université Victor Segalen Bordeaux 2, Bordeaux, France

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Laura Tarrago-Litvak, Marie-Line Andréola, and Simon Litvak, volume 3, pp. 708–713, © 2004, Elsevier Inc.

Glossary

Cap The structure at the 5'-end of eucaryotic mRNA introduced after transcription. It results from adding a methylated G to the terminal base of the mRNA.

DNA polymerase An enzyme that synthesizes a daughter strand of DNA under direction from a DNA or an RNA template.

Hotspot A site at which the frequency of mutation (or recombination) is increased significantly.

Integrase A retroviral encoded enzyme able to insert the viral double-stranded DNA synthesized by reverse transcriptase in the nuclear genome of the infected cell.

LTR An abbreviation for long-terminal repeat, a sequence directly repeated at both ends of a retroviral DNA.

Primer A short sequence (often of RNA) that provides a free 3'-OH end at which a DNA polymerase starts synthesis of a deoxyribonucleotide chain. The primer of retroviral reverse transcriptase is a cellular transfer RNA (tRNA) molecule.

Retrovirus An RNA virus that propagates via conversion into duplex DNA.

RNase H An endoribonuclease active on the RNA moiety of DNA–RNA hybrids.

Template A polynucleotide that furnishes the instructions for the sequence of nucleotides to be added to the primer strand during DNA polymerization.

Based on our results I proposed the DNA provirus hypothesis at a meeting in 1964: the RNA of infecting Rous sarcoma virus (RSV) acts as a template for the synthesis of viral DNA, the provirus, which acts as a template for the synthesis of progeny RSV RNA. At this meeting and for the next 6 years this hypothesis was essentially ignored.

Howard Temin (Nobel Lecture 1975)

In 1964, the University of Wisconsin virologist Howard Temin proposed the DNA provirus hypothesis to explain the mechanism by which a cancer-producing virus contains only RNA infects and transforms cells. His hypothesis reversed the flow of genetic information, following what was known as the central dogma of molecular biology. Although there was a ferocious initial opposition to his hypothesis, it was widely accepted, after the discovery of reverse transcriptase (RT) in 1970. Since its discovery, RT has been associated with retroviruses. In the following years, this RNA-dependent DNA polymerase was found in eucaryotes, procaryotes, retrotransposons, hepadnaviruses, retrans, and even humans. The role of RT, essential for viral replication in all retroviruses, is to synthesize double-stranded DNA copying the retrovirus single-stranded RNA genome. This DNA is then integrated into the host cell chromosomes as a provirus. Transcription then leads to copies of the viral RNA genome, from which the virus proteins and enzymes are formed. New viral particles then bud from the membrane of the cell.

RT performs a remarkable feat, reversing the normal flow of genetic information. The polymerases synthesizing DNA and RNA in cells are very accurate and make very few mistakes. This is essential because they are the caretakers of our genetic information, and mistakes may be passed on to the offspring. RT, on the contrary, makes mistakes very often by incorporating a noncomplementary nucleotide. One might think that this low fidelity would cause severe problems to the virus. But, in fact, this high error rate turns out to be an advantage for the virus.

The errors allow the retrovirus to mutate rapidly, in order to fit new conditions and escape, for instance, the immunological response of the host or become resistant to the effect of therapeutic agents.

Discovery of Reverse Transcriptase

The publication of the presence of RT in retroviruses in 1970, by the groups of Howard Temin and David Baltimore, in the same issue of *Nature*, must be considered as a landmark in the story of biological sciences of the second half of the twentieth century. The presence of a DNA polymerase able to copy RNA into DNA (contrary to all the enzymes of this family described before which copy a DNA template into DNA) violated the so-called central dogma of molecular biology which stated that genetic information always flows from DNA to RNA to protein. It is of interest to point out that while the laboratory of Baltimore came to this observation by searching for the mechanisms used by polymerases, involved in the replication of different viruses, the findings of Temin's group came as a strong support of his earlier experiments showing that the replication of retroviruses was sensitive to inhibitors of DNA-dependent polymerases. Years later after the discovery of RT, the finding in 1983 that the cause of acquired immune deficiency syndrome (AIDS) was a retrovirus, the human immunodeficiency virus type 1 (HIV-1), and the consequent emotion provoked by the AIDS pandemics led to an explosive burst of research on this virus. It can be considered that very few organisms have provided so much information on their structure and function in such limited time as in the case of HIV-1. Most of the information on retroviral replication in these last 25 years has come from the work on this human retrovirus.

Retrovirus Genome

All retroviral genomes consist of two identical molecules of RNA, which are single-stranded, (+) sense, and have a 5'-end cap and a 3'-end poly-(A) as in the case of eucaryotic messenger RNAs (mRNAs) (Figure 1). Retroviral genomes vary in size from 8 to 11 kb. The two RNA genomic molecules are physically linked as a dimer by hydrogen bonds. The reason for the presence of two identical RNA molecules is unclear; probably they may be an advantage for the replication of the viral genome in case of damage of one of the strands. Temin showed that only one RNA is enough for viral replication. The other distinctive feature of the retroviruses concerns the fact that their genome is produced by the cellular RNA polymerase II, without the involvement of a viral-encoded polymerase. Moreover, they are the only (+)-sense RNA viruses whose genome does not serve directly as mRNA immediately after infection. Another nucleic acid, present in all viral particles, a specific transfer RNA (tRNA), is the primer required for the initiation of replication. Most, if not all, DNA polymerases require a primer carrying a free 3'-end OH from where DNA synthesis is initiated. In the case of retroviruses this primer is a cellular tRNA that is able to anneal partially to a region near the 5'-end of the RNA viral genome, the primer binding site (PBS).

Examination of the sequences of several retroviruses has shown that the PBS is complementary to the 3'-end of a given tRNA (e.g., tRNA^{Trp} in avian retroviruses, tRNA^{Pro} in murine viruses, and tRNA^{Lys3} in HIV-1). The mechanism involved in the selection of a specific tRNA is not fully understood. While the role of the viral genome has been ruled out, early results with avian retroviruses suggested that RT was involved in tRNA packaging. Very recent work seems to indicate that in the case of HIV-1, the lysine aminoacyl-tRNA synthetase may help to

encapsidate the cognate tRNA primer in the retroviral particle. The small basic nucleocapsid (NC) protein encoded in the C-terminal region of the *gag* gene plays a role in the annealing of primer tRNA to the PBS region allowing the initiation of retroviral replication. The genomic organization in all retroviruses is invariant: 5'-*gag-pol-env*-3'. However, some complex retroviruses, like HIV-1, encode several other small regulatory proteins.

Activities Associated with the Reverse Transcriptase

With the exception of some murine retroviral polymerases, most RTs are heterodimeric enzymes. The embedded retroviral-encoded protease cleaves a polyprotein precursor leading to an RT subunit that is further cleaved by the same protease. Thus, RTs are formed by two subunits where the smallest is contained in the sequence of the largest (Figure 2). Retroviral DNA synthesis is absolutely dependent on the two enzymatic activities associated with RT: a DNA polymerase that can use either RNA or DNA as a template, and a nuclease, named ribonuclease H (RNase H), that is specific for the RNA strand of RNA-DNA hybrids. In the case of avian RTs, the β subunit of the *in vivo* active enzyme also carries the retroviral integrase domain (see below). All of the enzymatic functions required to complete the series of steps involved in the generation of a retroviral DNA can be attributed to either the DNA polymerase or the RNase H of RT as described below. However, it cannot be ruled out that certain viral proteins, such as the NC protein, may increase the efficiency of reverse transcription. RT is associated to viral and cellular proteins in the so-called reverse transcription complex (RTC) that is probably closely related to the preintegration complex (PIC) described further in this article.

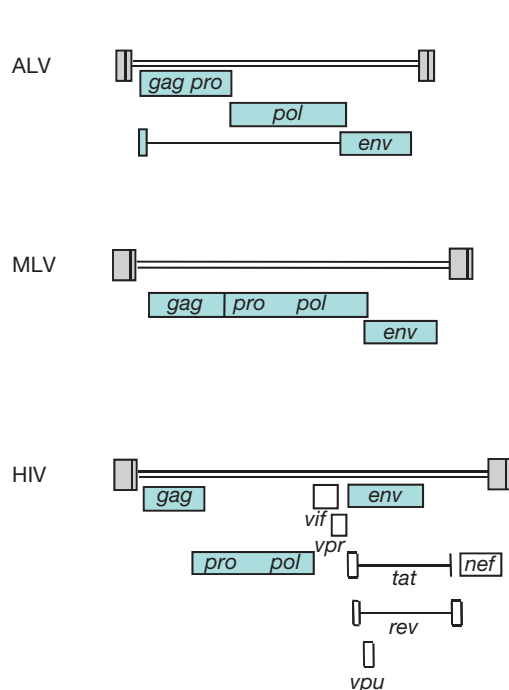


Figure 1 Genetic organization of retroviruses.

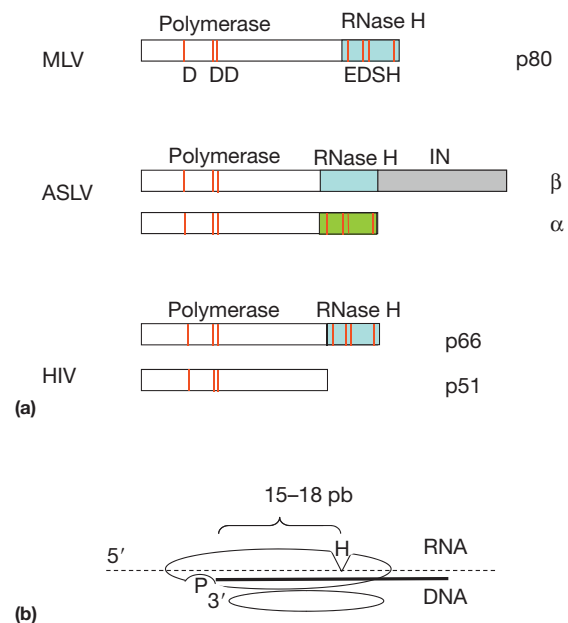


Figure 2 The subunit structure of different retroviral reverse transcriptases.

Reverse transcription of the retroviral genome depends on the coordination of the polymerase and RNase H activities of RT. The spatial arrangement between the two active sites of the enzyme (corresponding to about 15–18 bases in the template) contributes to the coordination of these two activities. While the 3'-OH of the primer strand is positioned close to the RT amino acid residues, important in the DNA polymerase active site, the template strand is positioned close to RT residues important for RNase H activity.

Biochemistry of Reverse Transcriptase

Mechanism of Action

The study of the mechanism of the RNA- and DNA-dependent DNA polymerase activity has led to a model involving a two-step binding mechanism. DNA polymerization catalyzed by RT is a multistep reaction, requiring the binding of the primer-template duplex and a deoxynucleoside triphosphate (dNTP) to RT prior to deoxy-nucleoside-5'-monophosphate addition to the 3' end of the primer. Kinetic experiments have established the order of binding of the substrates. The primer-template duplex binds prior to binding of the dNTP substrate to form the catalytically competent ternary complex (Figure 3). The rate-limiting step for nucleotide incorporation prior to dissociation of the enzyme is a conformational change. Following pyrophosphate liberation, the elongated product is released. Small changes in the rates and equilibrium constants have been observed for RNA- versus DNA-dependent DNA synthesis. All these mechanisms have been derived from pre-steady-state or steady-state kinetic assays using short primer-templates and do not always reflect the true nature of reverse transcription, although they are necessary in determining the kinetics and affinity of substrate binding to RT.

Processivity

The elongation phase of DNA synthesis by DNA polymerases may be processive or distributive. A processive DNA polymerase is able to synthesize long DNA stretches without dissociation of the enzyme from the primer-template duplex while a distributive polymerase dissociates easily after incorporation of few nucleotides. RT is considered to be poorly processive as compared with replicative DNA polymerases. However, a higher processivity of RT may be observed when copying specific regions of the viral genome or in the presence of some viral proteins.

Fidelity

All RTs show poor fidelity when compared with host DNA polymerases in *in vitro* systems. The lack of a proofreading function as well as the enzyme ability to extend mismatched primer termini probably explains much of this high error rate. The tendency of RT to switch templates, paired with its ability to extend mismatches by incorporating wrong nucleotides,

may be a significant source of replication errors. The error rates of various RTs estimated in different types of assays can be significantly affected by altering the parameters of the reaction. The error rate is quite dependent on sequence context and hot spots, where errors occur frequently, may differ according to the particular RT used in the assay. In addition, error rates may be different on RNA and DNA templates. As a consequence, it is not possible to define the error rate as a single number. These studies, taken together, do suggest that the overall error rate is high: about one per genome in a reverse transcription cycle. Cellular factors may alter error rates and some studies suggest that somewhat lower error rates occur *in vivo*. However, the *in vitro* data lead to the generally accepted impression that the average retroviral DNA genome differs from its parent by at least one mutation and provide the evidence for the existence of retroviral quasi-species.

Replication Errors *In Vivo* and the Emergence of Antiviral Resistance

The rates at which mutations accumulate in retroviral and in the host genomes are very different since the viral genomes are found to evolve at rates perhaps a million fold higher than the nuclear host genomes. The lack of RT fidelity and the accumulation of nonlethal mutations are closely related to the emergence of resistance toward antiviral therapeutic agents. This effect is especially dramatic in the case of HIV-1, the causal agent of AIDS. After short times of treatment, viral quasi-species appear with mutations in the viral target proteins (the most used targets against HIV-1 are RT and protease). The introduction of the so-called combination therapy against AIDS has greatly improved the possibility to keep the seropositive patients for several years with a low or undetectable viral charge. The combination of several anti-proteases and anti-RTs makes possible that, if a given mutation leads to the apparition of resistance against a specific drug, the other antivirals are still effective giving time to replace the one to which the target enzyme has become resistant.

Inhibitors of Reverse Transcriptase

Emphasis in the search of RT inhibitors has been obviously focused on the HIV-1 enzyme but some of the inhibitors developed against this agent may be useful in arresting human or animal retroviruses involved in other pathologies. Inhibitors of HIV-1 RT were the first therapeutic agents described against AIDS. Later appeared the protease inhibitors, and currently, inactivating agents against other viral targets (viral entry, integrase, and fusion) are used in AIDS patients resistant to the usual treatments. As mentioned above, the combined use of anti-proteases and anti-RTs has led to a great improvement in the control of HIV-1 proliferation in AIDS patients. Two main families of HIV-1 RT inhibitors are used against AIDS, the nucleoside analogs inhibitors of RT (NRTI) and the non-nucleoside inhibitors of RT (NNRTI). The first ones are

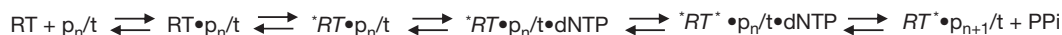


Figure 3 The kinetics of the reverse transcription reaction.

recognized and incorporated by RT in the newly synthesized DNA but as they lack, or have, a modified group in the 3' position of the ribose they cannot be linked to the following dNTP precursor and thus they act as chain terminators. To be active *in vivo*, the putative NRTI, which are active against all retroviral RTs, must be triphosphorylated by cellular kinases. The first NRTI described was azidothymidine or zidovudine (AZT), but currently more than a dozen of these inhibitors have been approved by the US Food and Drug Administration (FDA) and are used in clinics. It is important to point out that the resistance to NRTIs may appear by either mutations in the RT leading to the nonrecognition of the NRTI or a more recently identified RT activity consisting in the pyrophosphorolysis and further elimination of the terminator nucleotide. A second family, the NNRTIs, is specific of the HIV-1 RT since they inhibit the viral polymerase after binding to a hydrophobic region of this enzyme. They do not have a typical nucleoside structure, but belong to a variety of different chemical families of compounds, and in contrast to NRTIs they do not need to be metabolized inside the cell to display their antiviral effect.

Recombination

Since retroviruses ordinarily carry two identical or nearly identical RNAs, the genetic consequences of using only one versus portions of both of the RNAs to template DNA synthesis are usually the same. However, under certain conditions, two genetically distinct RNAs can be encapsidated. Portions of each template can be used to generate a single DNA, producing a recombinant virus. Recombination does occur in virions produced by a single cell that has been co-infected by two different viruses demonstrating that recombination requires co-packaging of two viral genomes into the same particle. During reverse transcription in such heterozygous virion particles, the DNA product can be generated containing portions of each of the two genomic sequences in one molecule.

The Role of RT in Reverse Transcription

Reverse transcription catalyzed by RT begins when the viral particle enters the cytoplasm of a target cell. DNA synthesis using the RNA retroviral genome as template takes place in the cytoplasm leading to a linear double-stranded DNA (dsDNA), co-linear with the viral RNA template, the proviral DNA. However, proviral DNA contains terminal duplications known as the long terminal repeats (LTRs) that are not present in viral RNA.

The following scheme is widely accepted to reflect the process of retroviral DNA synthesis catalyzed by RT (Figure 4):

1. Minus-strand DNA synthesis starts from the 3' end of a tRNA partially annealed to PBS in genomic RNA. Minus-strand DNA synthesis proceeds until the 5'-end of genomic RNA is reached, generating a short DNA intermediate termed minus-strand strong-stop DNA (–sssDNA). Since the PBS is near the 5'-end of viral RNA, the length of this strong-stop DNA is about 100–150 nucleotides.
2. Following RNase H-mediated degradation of the RNA strand of the RNA(–sssDNA) duplex, the first strand

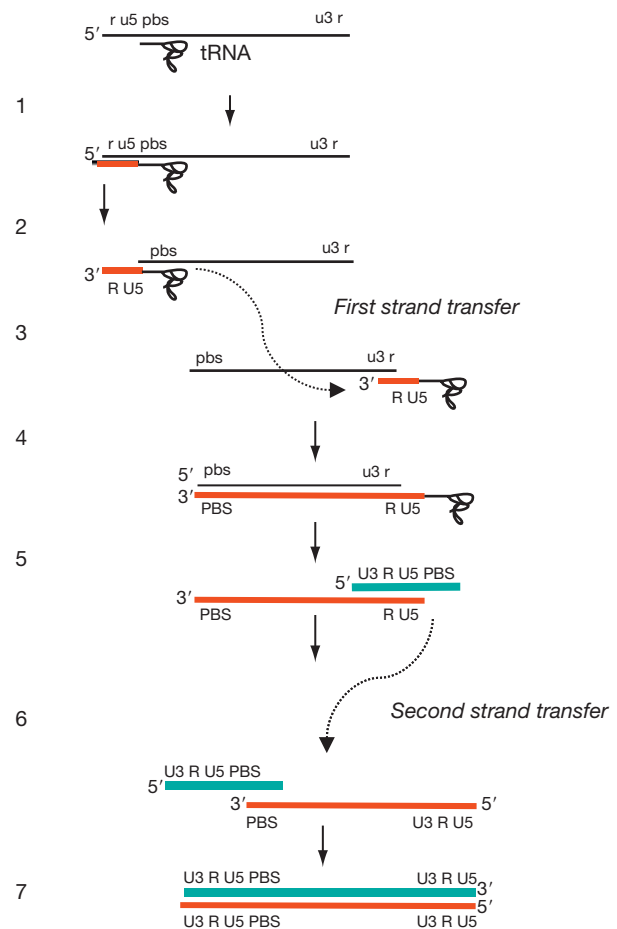


Figure 4 Reverse transcription steps.

transfer involves the annealing of the –sssDNA to the complementary 3'-end of a viral genomic RNA. This transfer is allowed by the presence of identical sequences known as the repeated (R) sequences, present at the 5'- and 3'-ends of the RNA genome. The 3'-end of –sssDNA is copied from the R sequences at the 5'-end of the viral genome and therefore contains sequences complementary to R. After the RNA template has been removed, the –sssDNA can anneal to the R sequences at the 3'-end of the RNA genome. The annealing reaction related to the first strand transfer appears to be facilitated by the NC protein.

3. Once the –sssDNA has been transferred to the 3' R region on viral RNA, minus-strand DNA synthesis may start again, followed by RNase H digestion of the template strand.
4. RNase H degradation is not complete since some hybrid regions seem to be more resistant to the nuclease. The RNA genome contains a short polypurine tract (PPT) that is relatively resistant to RNase H degradation. A defined RNA segment derived from the PPT primes plus-strand DNA synthesis. Plus-strand synthesis is halted after a portion of the primer tRNA is reverse-transcribed, leading to a DNA called plus-strand strong-stop DNA (+sssDNA). Some retroviruses generate more than one plus-strand primers from the RNA genome.

5. RNase H removes the primer tRNA, exposing sequences in +sssDNA that are complementary to sequences at or near the 3' end of plus-strand DNA.
6. Annealing of the complementary PBS segments in +sssDNA and minus-strand DNA constitutes the second strand transfer.
7. Plus- and minus-strand syntheses are then completed by RT, with the plus and minus strands of DNA each serving as a template for the synthesis of the other strand.

Ribonuclease H Activity of RT

The functions of retroviral RNase H embedded in RT are essential during DNA synthesis. Several steps of reverse transcription require this activity as seen in the retrotranscription scheme (Figure 4). Kinetic studies have shown that RT lacking RNase H activity is incapable of supporting DNA strand transfer. This has been confirmed in a variety of retroviral systems in culture. RNase H activity requires the presence of divalent cations, four conserved amino acids (one glutamic and three aspartic residues) being close to the Mg^{2+} site.

As two obligatory strand transfer events occur in the production of proviral DNA, it would be of considerable interest to identify molecules able to interfere with these steps. Inhibitors of DNA translocation and RNase H activity could provide new lead compounds for treating HIV-1 infection. Unfortunately, the *in vitro* inhibitors of retroviral RNase H studied until now have proved to be inactive *ex vivo* or *in vivo* because of minimal cell uptake or/and cytotoxicity.

The best-studied retroviral RNase H is that of HIV-1 associated with the p66 subunit of RT. Contrary to some other retrovirus, the RNase H domain of HIV-1 RT expressed as an isolated recombinant protein is inactive due to the importance for this activity of some amino acid residues present in the RT DNA polymerase domain. When the HIV-1 RNase H domain was purified and mixed with a separately expressed and purified fragment containing the DNA polymerase domain of HIV-1 RT, RNase H activity was restored. The three-dimensional (3D) structure of the isolated HIV-1 RNase H domain was solved to a resolution of 2.4 Å and compared with the *Escherichia coli* RNase H structure. Several conserved amino acids are present in both enzymes. Despite limited sequence homology, the structures of the RNase H of HIV-1 RT and the RNase H of *E. coli* are remarkably similar. During RNase H cleavage concomitant with DNA elongation, the 3'-terminus template-primer complex spans the polymerase and RNase H domains, with 18–19 nucleotide base pairs between the primer terminus and the cleavage site. This represents the distance between the polymerase and RNase H active sites of HIV-1 RT determined by biochemical and structural studies. No clear data is available on the specificity of the cleavage mechanism of RNA by the viral enzyme.

Integration and Retroviral Replication

As proposed by Temin since the early 1960s, retrovirus replication is performed via a DNA intermediate, the proviral DNA that is synthesized by the RT in the cytoplasm. Integrase (IN), the third retroviral encoded enzyme in addition to RT and

protease, is involved in the integration of the proviral DNA in the nuclear DNA of the infected cell. Integration of the proviral DNA synthesized by RT into the host genome has the advantages of increasing the long-term stability of the viral information inserted in the chromosomes and also takes profit of the cellular enzymes involved in gene expression. Thus, once integrated, the proviral DNA is recognized by the host RNA polymerase II for the efficient transcription of viral DNA into new copies of the viral genome and mRNAs. Transcription of the proviral DNA is facilitated by the viral-encoded Tat protein. These mRNAs are then translated into the viral proteins.

IN catalyzes two reactions, the first one takes place in the cytoplasm and concerns the cleavage of two terminal residues of the proviral DNA and the second step operates in the nucleus and involves the integration of the provirus in the nuclear DNA (Figure 5).

Both the central catalytic domain and C-terminal domains have been shown to bind both viral and cellular DNA. Studies of early steps of IN attachment to viral DNA showed that the specificity of IN for DNA is linked to its oligomerization state. Interestingly, this mechanism is also observed with other proteins sharing high structural analogies with retroviral IN and RNase H such as eukaryotic transposases (*mariner*) and the maize Mu transposase. As expected, the structure of HIV-1 IN has attracted special attention. At the present time, no crystal structure data of the entire HIV-1 IN or IN bound to its DNA substrates exists but biochemical and structural studies show that IN functions as a dimer or a tetramer. Very recently, the group of Cherepanov described a crystal structure of full-length integrase from the prototype foamy virus IN in complex with its cognate DNA. The structure shows the organization of the DNA-IN complex comprising a tetrameric integrase tightly associated with a pair of viral DNA ends. Foamy viruses (FVs) also known as spumavirus are complex retroviruses that are widespread in many species. Infection of natural hosts by FV leads to a lifelong persistent infection, without any evidence of pathology. While FV are ubiquitous in all nonhuman primates, they are only acquired as rare zoonotic infections in humans.

The 3D structure of the catalytic domain of HIV-1 integrase shows a remarkable similarity with the structure of the isolated HIV-1 and *E. coli* RNase H as well as with other enzymes belonging to the polynucleotide phosphorylase family (Figure 6). The latter led to the description of inhibitors able to affect simultaneously these two retroviral targets.

The transfer of the viral DNA and the IN into the nucleus takes place in a multimeric PIC composed of viral and cellular encoded proteins. The exact composition and the mechanism of nuclear entry of the PIC is still controversial but in addition to the IN and the proviral DNA, some consensus exists on the presence in the PIC of the viral proteins RT, viral protein R (VPR), NC, and cellular factors playing a role in the recognition of the chromosomal integration locus (Ini1, LEDGF/p75). Concerning the trafficking mechanism of the PIC from the cytosol to the nucleus, it has been proposed that the cytoskeleton and the centrosome may play a role in this process.

The importance of the retroviral IN in the HIV-1 replication cycle has fostered the search of specific inhibitors of this enzyme (Figure 7). Among the pharmacophores identified as integrase inhibitors by the pharmaceutical companies Merck and Shionogi were the β -hydroxyketo group, diketo group, or a

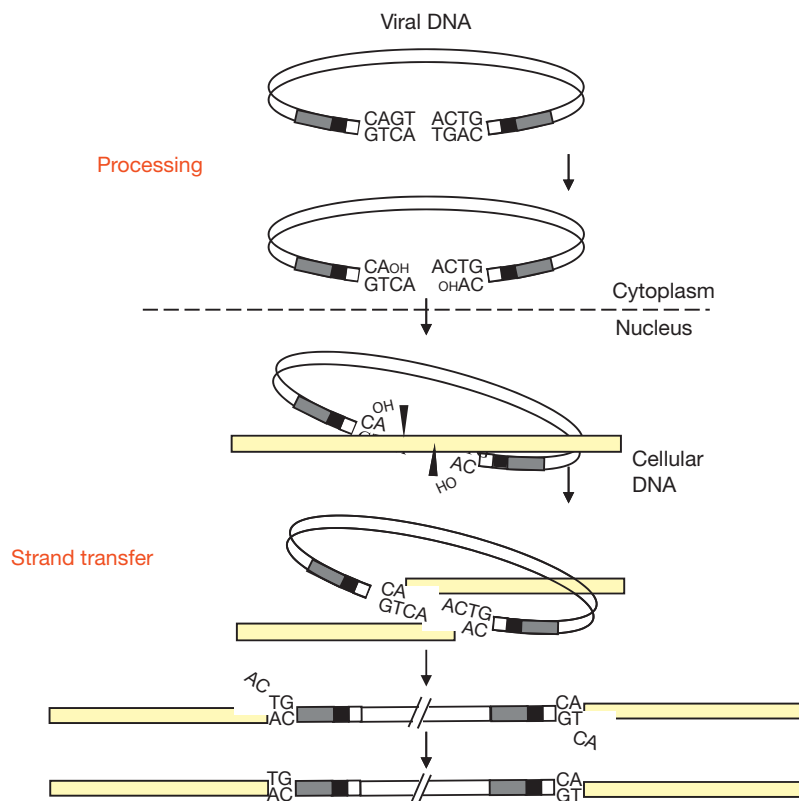
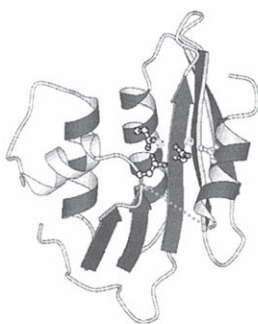
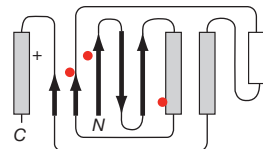


Figure 5 Scheme of the two steps of integration.



RNase H



IN

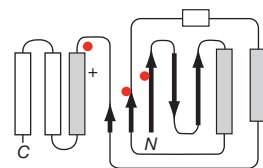


Figure 6 Comparison of retroviral RNase H and IN structures. Red dots correspond to the essential conserved amino acids at the active site. Adapted from Yang W and Steitz TA (1995) Recombining the structures of HIV integrase, RuvC and RNase H. *Structure* 3: 131–134.

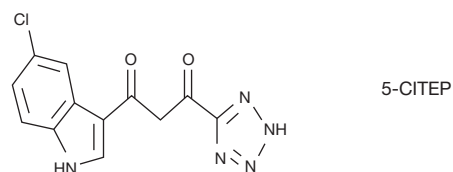
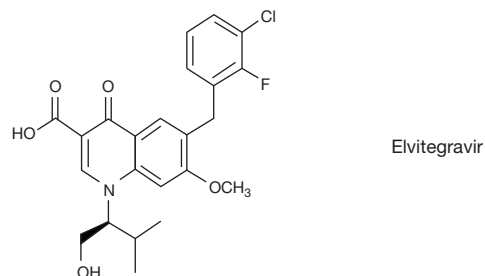
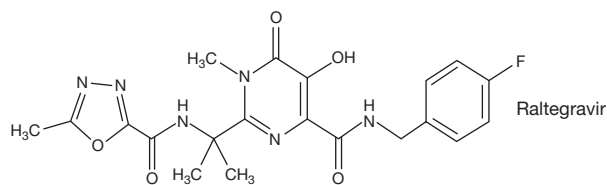


Figure 7 IN inhibitors.

carboxylate group (DKA). From these studies, a reaction in the presence of Mg^{2+} was shown to be the most stringent condition with respect to strand transfer inhibition. The first HIV integrase inhibitor, Raltegravir, was approved by the FDA on October 2007 for use in antiretroviral treatment experienced adult patients with viral resistance against other anti-HIV agents.

Role of Reverse Transcriptase in Modern Biology

Nowadays, it is known that RT not only is present in animal retroviruses but also plays a crucial role in DNA replication of procaryotic and eucaryotic cells, in the replication of some plant and animal DNA viruses, as well as in the function of several classes of transposable elements. As mentioned in the remarkable book *Retroviruses*:

it should be remembered that RT has had a critical role in the development of modern biology. The ability to convert RNA into DNA has been as important to molecular biologists as it has been to retrovirologists. It is worth taking a moment to reflect on how much more difficult it would have been to solve the puzzles posed by the organization and expression of genes in higher eucaryotes if RT did not exist.

See also: [Cell Architecture and Function: Dynein](#); [Heat/Stress Responses](#); [Microtubule-Associated Proteins](#); [Molecular Biology: DNA Polymerases: Kinetics and Mechanism](#); [Nonhomologous Recombination: Bacterial Transposons](#); [Regulation of Chromatin Dynamics](#); [RNA Polymerase II and Its General Transcription Factors](#).

Further Reading

Andréola ML (2004) Closely related antiretroviral agents as inhibitors of two HIV-1 enzymes, ribonuclease H and integrase: Killing two birds with one stone. *Current*

- Pharmaceutical Design* 10(30): 3713–3723.
- Arts EJ and Le Grice SF (1998) Interaction of retroviral reverse transcriptase with template–primer duplexes during replication. *Progress in Nucleic Acid Research and Molecular Biology* 58: 339–393.
- Champoux JJ and Schultz SJ (2009) Ribonuclease H: Properties, substrate specificity and roles in retroviral reverse transcription. *FEBS Journal* 276(6): 1506–1516.
- Coffin JM, Hughes SH, and Varmus HE (eds.) (1997) *Retroviruses*. New York, NY: Cold Spring Harbor Laboratory.
- Darlix JL, Lapadat-Tapolsky M, de Rocquigny H, and Roques BP (1995) First glimpses at structure–function relationships of the nucleocapsid protein of retroviruses. *Journal of Molecular Biology* 254: 523–537.
- De Clercq E (2009) Looking back in 2009 at the dawning of antiviral therapy now 50 years ago: An historical perspective. *Advances in Virus Research* 73: 1–53.
- Delelis O, Carayon K, Saib A, Deprez E, and Mouscadet JF (2008) Integrase and integration: Biochemical activities of HIV-1 integrase. *Retrovirology* 5: 114–127.
- Hare S, Gupta SS, Valokov E, Engelman A, and Cherepanov P (2010) Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* 464(7286): 232–236.
- Kleiman L, Jones CJ, and Musier-Forsyth K (2010) Formation of the tRNA packaging complex in HIV-1. *FEBS Letters* 584: 359–365.
- Litvak S (1996) *Retroviral Reverse Transcriptases*. Austin: Chapman and Hall, R.G. Landes Company.
- Marchand C, Maddali K, Metifiot M, and Pommier Y (2009) HIV-1 IN Inhibitors: 2010 Update and perspectives. *Current Topics in Medicinal Chemistry* 9(11): 1016–1037.
- Skalka AM and Goff SP (eds.) (1993) *Reverse Transcriptase*. New York, NY: Cold Spring Harbor Laboratory.
- Temin HM (1976) The DNA provirus hypothesis. *Science* 192(4244): 1075–1080.
- Temin HM (1993) Retrovirus variation and reverse transcription: Abnormal strand transfers result in retrovirus genetic variation. *Proceedings of the National Academy of Sciences of the United States of America* 90: 6900–6903.
- Temin HM and Baltimore D (1972) RNA-directed DNA synthesis and RNA tumor viruses. *Advances in Virus Research* 17: 129–186.
- Verma IM (1997) The reverse transcriptase. *Biochimica et Biophysica Acta* 473: 1–38.

DNA Replication: Eukaryotic Origins and the Origin Recognition Complex

I Chesnokov and A Svitin, University of Alabama, Birmingham, AL, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Melvin L. DePamphilis and Cong-jun Li, volume 1, pp. 753–760, © 2004, Elsevier Inc.

Glossary

Autonomously replicating sequence (ARS) A DNA sequence that imparts origin activity to extrachromosomal elements in the presence of appropriate replication proteins, and whose activity is sensitive to genetic alterations.

Cdc Cell division cycle protein. A large number of Cdc genes have been identified in eukaryotes that affect various stages in cell division. Cdc6 is the name used in the budding yeast, *Saccharomyces cerevisiae*; Cdc18 is the name used in the fission yeast, *Schizosaccharomyces pombe*. The nomenclature for budding yeast proteins is generally applied to other organisms.

Cdk Cyclin-dependent protein kinase. This phosphorylates specific amino acids in proteins, but only when associated with a cyclin protein.

Cdt1 A protein encoded by Cdc10-dependent transcript 1 in *S. pombe*. Cdt1 is the same as RLF-B in *Xenopus laevis*.

DNA replication origin The DNA site where replication begins; also called an origin of bidirectional replication.

Mcm Minichromosome maintenance proteins. These were identified as genes required to maintain plasmids in *S. cerevisiae*. At least seven of these proteins are involved in DNA replication.

Origin recognition complex (ORC) Six different proteins that bind to DNA replication origins and thereby initiate assembly of a pre-replication complex.

Pre-replication complex (pre-RC) Fourteen different proteins consisting of ORC, Cdc6, Cdt1, and Mcm(2–7) that form a complex with chromatin during G1 phase of the cell cycle and that become the site where DNA replication begins during S phase.

Replicator A DNA sequence that imparts origin activity when translocated to other chromosomal regions, and whose activity is sensitive to genetic alterations.

Eukaryotic cells replicate their DNA with remarkable precision during the course of growth and division. Each genomic region should be replicated once and only once per cell cycle to avoid genomic instability. To assure such an accuracy of genome duplication, eukaryotes have evolved a mechanism for the initiation of DNA replication that involves multiple origins of replication (*ori*) along the chromosomal DNA. DNA replication in eukaryotes is a highly conserved process that begins with the binding of the six-subunit origin recognition complex (ORC) to the origins of DNA replication. ORC binds to origin sites in an ATP-dependent manner and helps in recruiting and in the assembly of additional initiation factors (Cdc6, Cdt1 and mini-chromosome maintenance proteins (MCM) 2–7 helicase complex) to form pre-replicative complex (pre-RC), thus licensing origin for further replication activity. Before the S phase, the combined action of cyclin-dependent kinases (CDK) and Dbf4-dependent kinases (DDK) results in the loading of additional replication proteins to form the pre-initiation complex (pre-IC). DNA synthesis is initiated after the recruitment of DNA polymerase complexes (polymerase α -primase and polymerase ϵ).

Structure and Diversity of Eukaryotic Origins

Replicon Model

The basis of our current views on replication initiation stems from replicon model, proposed by Francois Jacob and Sydney Brenner in 1963. In this model, a replicon, defined as a DNA molecule capable of autonomous replication, is characterized by two functional components – initiator and replicator. A *trans*-acting

initiator protein binds to a specific sequence, DNA *cis*-element called replicator. Upon binding, initiator is transformed into replication-competent form which is followed by a replication initiation at the replicator site.

DNA replication starts with the separation of DNA strands (DNA unwinding) resulting in a formation of replication forks moving in opposite directions. Due to the antiparallel nature of the DNA duplex, DNA synthesis occurs continuously on one strand (leading strand) and discontinuously on the other (lagging strand). As DNA polymerases can travel along their templates only in one direction (5′–3′), the DNA synthesis on a lagging strand results in short DNA fragments (Okazaki fragments) which are joined together by DNA ligase.

Replication Origins in Prokaryotes, Viruses, and Lower Eukaryotes

Replicon model was first solidly proven in prokaryotic organisms. In *Escherichia coli*, initiator protein DnaA binds specifically to the 250-bp AT-rich replicator sequence *oriC*, resulting in a local ATP-dependent DNA unwinding and replicative DNA helicase (DnaB) loading. This is followed by the assembly of replication fork components, including DNA polymerases, and a replication initiation.

Viral origins contain short species-specific sequences. These sequences are bound by a single virus-encoded protein (such as SV40 T antigen). The binding is highly sequence-specific; even mild alterations of the viral origin strongly affect replication. Unlike most eukaryotic origins, viral origins can initiate replication several times during the same cell cycle. Viral origins are functionally independent from neighboring sequences.

This represents another difference from eukaryotic initiation sites in which origin activity may be strongly influenced by the context of a surrounding DNA.

The most well-studied and characterized eukaryotic replication origins are found in the genomes of yeast cells. In budding yeast, *Saccharomyces cerevisiae* sites of replication initiation were originally identified in a screen for genomic regions supporting replication of extrachromosomal plasmids which carry them. These regions were termed autonomously replicating sequences (ARS) and were shown to have a significant sequence similarity with each other. A typical ARS has size of approximately 100 bp and consists of 11-bp ARS-consensus sequence (ACS or A element) and another less-conserved 10- to 15-bp element (B element) essential for origin function. Mutations in these two elements affect binding of ORC to the origin. At the same time, many studies suggest that additional factors and epigenetic parameters, such as nucleosome positioning, are also important in defining of origin selection and function.

Unlike budding yeast, the origins in other eukaryotic species are significantly less strictly defined. Even in fission yeast, *Schizosaccharomyces pombe* origins are much larger (500–1000 bp) and are not conservative in sequence, except for being AT-rich. The latter feature is apparently important for ORC binding, as *S. pombe* Orc4 subunit contains multiple AT-rich sequence-binding motifs (AT-hooks) at the N-terminal part of the protein. Reduction in AT-content of origins results in delayed initiation, thus illustrating connection between sequence composition and functional characteristics of origin in *S. pombe*.

Metazoan Origins

Difference from budding yeast origin structure is even more obvious in metazoan species where origins are large and do not exhibit any sequence conservation; no sequence elements analogous to yeast ARS were found. Situation is even more complicated by a large size and complexity of the genomes in higher eukaryotes, and consequently, by a significant number of replicons and replication initiation sites (Figure 1). For example,

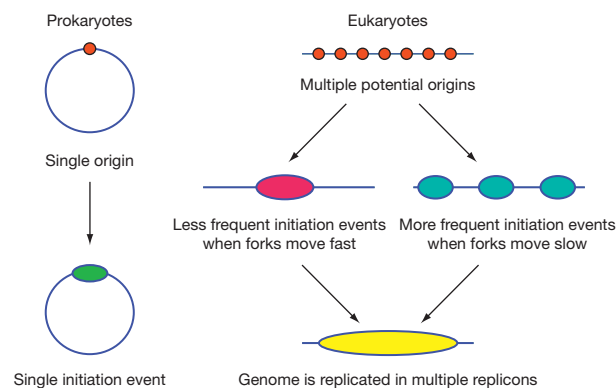


Figure 1 Replication origins in prokaryotes and eukaryotes. Bacterial chromosome contains one origin which initiates single replication fork. Multiple potential origins are found on eukaryotic chromosomes. Activation of these origins is coregulated with replication fork progression rate; if forks slow down, additional origins are activated.

an estimated 30 000 replication origins may function at different times in human cells. Distribution and localization of origins change during development in metazoans, which further obscures the study and an understanding of origin-defining factors in these species. Such flexibility of DNA replication in metazoans is well exemplified by the crucial change in this process during the development of insects and amphibians. In *Drosophila melanogaster* early embryos and *Xenopus laevis* fertilized eggs, DNA replication is extremely rapid, taking only 10–30 min for complete replication of the whole genome due to large number of functioning origins spaced as close as 4–7 kb. In these conditions, frequent initiation of replication employs practically all available origins. Later in development, origin distribution significantly changes, with the initiation occurring selectively on a fraction of all origins, and time needed for genome replication also increases up to 8–10 h.

Some of the most characterized origins are found in *D. melanogaster*. These origins control an amplification of two clusters of chorion genes in somatic follicle cells that surround developing oocyte. At a specific stage of oogenesis, these two genomic regions on the third and X chromosome are amplified up to 60-fold by supporting extra rounds of DNA replication. This process is followed by high expression of the chorion genes, required for the successful formation of the eggshell. Two *cis*-regulatory elements important for the amplification were identified within the chorion gene cluster: the 400 bp amplification control element 3 (ACE3) and ~1 kb ori β . It was shown that *Drosophila* ORC localizes to the chorion amplification gene cluster in follicle cells, and that ACE3 and ori β are sufficient to drive this localization. ORC also binds these sequence elements *in vitro*, however, in DNA-binding assays, *Drosophila* ORC shows at best sixfold preference for the chorion locus origin DNA compared to random DNA fragments. This suggests that origin determination in metazoan relies on the mechanisms other than simple initiator–replicator sequence recognition.

Many metazoan origins were mapped by different methods in mammalian cells. They can be roughly subdivided into two groups. Origins of one group localize to relatively confined locations on chromosomes. One example of this type of origins is human β -globin locus. This locus includes five genes that encode β subunit of hemoglobin. DNA replication initiation occurs within several kb region located in the intergenic space between δ - and β -globin genes. Removal of this initiation region abolishes origin activity. Under these conditions, β -globin locus is passively replicated from outside origin. Initiation region itself contains two independent nonoverlapping replicators, and each of them is sufficient to induce replication in ectopic sites upon transfer. Combination of asymmetric purine–pyrimidine sequence and several AT-rich stretches is required for replication initiation activity of these replicators.

The second type of higher eukaryotes origins is characterized by large extended initiation zones, tens of kilobases in size. Initiation in these origins occurs infrequently at multiple sites distributed across the entire region. One of the better studied origins of this kind is Chinese hamster ovary dihydrofolate reductase (*DHFR*) locus, a 240-kb chromosomal domain amplified in methotrexate-resistant CHO cell line. In this locus, replication initiation events are mapped across the whole 55-kb intergenic region. The initiation of DNA replication is sensitive

to mutations in specific sequences surrounding the initiation zone. For example, the deletion of the *DHFR* promoter allows initiation to occur within the *DHFR* gene, which does not normally initiate replication. Removal of specific sequences at the 3' end of the *DHFR* gene, causing *DHFR* transcripts to extend to the intergenic region, confines initiation to the far 3' end of the locus.

Importantly, no consensus origin sequence was found among the metazoan initiation sites. *In vitro* DNA-binding assays demonstrate that metazoan ORC preferentially binds to sequences with higher AT-content, though it is unable to differentiate between naturally occurring origin sequences derived from known initiation sites and artificial AT-rich DNA fragments. Thus, origin specification in higher eukaryotes apparently involves other factors rather than simple recognition of replicator sequence by initiator.

Specification and Selection of Eukaryotic Origins

Mechanisms or Origin Specification

Several mechanisms have been described which participate in targeting ORC to the specific sites on a DNA. First, ORC itself has a DNA-binding ability. The binding of ORC to DNA occurs with high specificity in budding yeast, however, in metazoan species, ORC does not have a strong preference for binding to a particular DNA sequence but still favors AT-rich DNA. Second, other replication initiation proteins, such as Cdc6, could enhance ORC interaction with origin sequences. Third, other factors, such as transcription factors or chromosomal proteins, which interact specifically with DNA, could be involved in recruiting ORC to specific sites in the genome. Fourth, characteristics and state of chromatin structure in specific regions of the genome could restrict the areas where ORC can function. Finally, certain specific conditions such as nucleotide pool level or DNA methylation may also influence ORC binding and assembly of the pre-RC.

DNA topology may be one of the factors determining a choice of the binding site by ORC. In *in vitro* DNA-binding experiments, *Drosophila* ORC demonstrates significant preference for negatively supercoiled DNAs, binding to them with 30-fold higher affinity as compared to linear or relaxed DNA fragments. Such topological sensitivity of ORC binding is closely related to another important factor – a chromatin structure of the binding site. There are several examples of changes in origin usage as a result of alterations in local chromatin structure. In yeast, mutations in histone deacetylase (HDAC) Sir2 facilitate initiation events, rescuing replication mutants involved in pre-RC assembly. In *Xenopus* and humans, HBO1, a histone acetyltransferase (HAT) directly interacts with the components of pre-RC and is required for MCM loading onto chromatin. Chromatin acetylation was also shown to be critical for chorion gene cluster amplification in *Drosophila* follicle cells.

Influence of chromatin structure on the replication initiation is not restricted to histone acetylation events. In *Xenopus* system, DNA methylation state was shown to affect initiation. Origin activity is repressed by highly methylated state which interferes with the ORC binding. However, if a low-methylated region is created on such DNA, ORC binding is detected, and replication initiation occurs. Also, DNA methylation was

shown to determine replication start location, either directly or indirectly, in mammalian cells.

Transcription is known to induce a negative supercoiling which facilitates DNA unwinding required for the replication initiation. This idea is supported by the data that, replication initiation frequency in fission yeast genome is twice as high in the intergenic regions of divergent transcription, compared to the ones of convergent transcription. Also, formation of active transcription complex can specify replication origin on the plasmid DNA in *Xenopus* extracts.

Preferential localization of replication origins to the actively transcribed regions, also suggests that an open chromatin state associated with these regions is favorable for the replication initiation. Many yeast origins are known to be located in the proximity of promoters. Moreover, direct interactions between ORC and transcription factors E2F1, Rb and Myb complex important for chorion gene expression, are described in *Drosophila*. Mutations in these genes affect localization of ORC and a function of chorion-cluster amplification origin. Another example of tight transcription–replication association comes from studies in human cells, where Epstein-Barr virus (EBV) transcription factor EBNA1 (EBV-coded nuclear antigen 1) participates in recruitment of ORC to the viral replication origin, *oriP*. These results could partially explain why all attempts to identify consensus sequence of higher eukaryotes origin have failed; various transcription factors could impart both developmental and DNA sequence specificity to each particular origin.

Replication factors such as Cdc6 can also participate in the origin specification. In budding yeast, interaction between ORC and Cdc6 is facilitated by origin DNA and reduces dissociation rate of ORC from origin. In metazoan species, Cdc6–ORC interaction might also enhance DNA binding by ORC. It is possible, that Cdc6 and transcription factors act in combination to specify the interactions between ORC and DNA.

Origin Plasticity

Important characteristic of the origin selection process is its dynamical regulation. This feature is illustrated by the origin plasticity phenomenon, described by J. Herbert Taylor in 1977. Under the conditions of thymidine starvation of Chinese hamster ovary cells, when rate of replication fork progression is decreased, total replication rate is compensated by the increase in the number and frequency of replication initiation events; cells with slower fork movement activate additional origins. On the other hand, acceleration of fork progression results in a decrease in the number of active origins. Thus, replication fork movement rate and distances between active origins are coordinately regulated (Figure 1).

The number of genomic sites with assembled pre-RC is higher than the number of the sites of actual replication initiation. Pre-RC sites are formed in excess; possibly, to supply a pool of additional origins which can be utilized when needed. This mechanism protects genome from being under-replicated, since the regions which would be out of reach for stalled or slowed replication forks started from usually active origins, would still be replicated due to the activity of the local dormant origins.

The mechanism of dormant origins activation could be mainly stochastic. In this model, dormant origins are repressed by passing replication forks which originated from preferred,

normal origins. However, under conditions of the replication stress, dormant origins become unrepressed since the replication forks do not reach them. Another model stipulates that stalled replication fork induces conformational changes in the surrounding chromatin and facilitates replication initiation from less preferred dormant origins. In any case, dormant origin activation would be triggered by the changes in replication fork progression.

Methods to Study DNA Replication

Current understanding of eukaryotic origin structure, selection, and DNA replication in general was achieved by the employment of a broad array of *in vitro* and *in vivo* methods. The most important of them include chromatin immunoprecipitation followed by analysis on microarrays (ChIP-Chip) or direct high-throughput sequencing (ChIP-seq) where ORC or other components of pre-RC are cross-linked to the DNA, isolated by immunoprecipitation and associated DNA is identified. DNA footprinting as well as assays based on identification of early replication intermediates, such as nascent DNA strand assay and 2D gel analysis, are also important *in vivo* approaches. Important quantitative parameters of replication, such as origin densities in the genome and rates of replication fork progression, are assessed using DNA fiber-based approaches where chromosomes are pulse-labeled with modified nucleotide precursors and newly synthesized DNA is visualized by immunofluorescence. Cell-free systems that allow studies of DNA replication *in vitro* are described for humans, *Xenopus* and *Drosophila* and can be used for functional dissection of the processes and factors involved in the initiation of DNA replication.

Origin Recognition Complex as the Core of Replication Initiation Machinery

ORC Discovery, Conservation, and Structure

The identification of the ORC in *S. cerevisiae* was an important advance in understanding eukaryotic DNA replication. It was identified by Stephen Bell and Bruce Stillman in 1992 as a factor that specifically bound to the yeast ARSs. Yeast ORC is composed of six tightly associated protein subunits, ranging from 104 kDa (Orc1) to 50 kDa (Orc6). Since its original discovery, evidence has steadily accumulated that ORC plays a central role in the initiation of DNA replication and recruitment of other essential replication factors to the Ori.

ORC has been conserved throughout eukaryotic evolution. ORC subunits and/or complete ORC complexes have been identified in *S. pombe* and various metazoans, including *D. melanogaster*, *X. laevis*, and humans. This conservation of ORC, as well as numerous other factors required for DNA replication, strongly suggests that there must be common mechanisms for the initiation of DNA replication in all eukaryotes, despite dramatic differences in the structure of eukaryotic origins of DNA replication and an absence of obvious conserved sequences among them.

ORC proteins also share homology with another component of pre-RC—Cdc6. This protein is well conserved among eukaryotes and may be a paralog of Orc1. Moreover, Orc1 can be more related to Cdc6 than to other ORC subunits. The structural

data indicate that ORC and Cdc6 may form a ring-like structure around the DNA reminiscent of MCM helicase ring.

Subunits 1–5 of ORC as well as Cdc6 contain conserved Walker A and B ATP-binding domains within the AAA+ fold. These features are characteristic of the proteins which form ring-shaped complexes and bind DNA in the central channel of the ring. The N-terminus of Orc1 contains bromo-adjacent homology (BAH) domain which is important for protein–protein interaction and provides a structural basis for ORC functions in heterochromatin. Orc6, on the other hand, does not share any of the structural features of Orc1–5 and has its own characteristic domains; unique conserved Orc6 protein fold domain at the N-terminus and the coil-coiled motif at the C-terminal part found in metazoan species. This coil-coiled region is important for cytokinetic functions of Orc6.

ORC Functions in Pre-RC Formation

Binding of ORC to DNA specifies sites where initiation of replication will occur, and this process is ATP-dependent. Experimental data indicate that ATP binding by ORC, but not the ATP hydrolysis, is important for DNA binding by the ORC. The coordination of DNA and ATP binding by ORC probably involves allosteric interactions among the subunits. It seems that ORC can exist in alternative conformational states induced by stable interaction of ORC with ATP and origin DNA. In humans, it was shown that ATP is also required for maintenance of ORC integrity and serves as a structural cofactor for ORC.

Even though ORC is bound at specific chromosomal regions containing origins of replication in both differentiated *Drosophila* and human somatic cells *in vivo*, little is known about how ORC complex finds these sequences. Several mechanisms which target ORC to the specific sites on the DNA were proposed. Among those are DNA topology and modifications, chromatin structure, transcription, transcription and replication factors, and nucleotide pool levels. These mechanisms of ORC targeting to specific DNA sites are addressed in more detail above (see section Specification and Selection of Eukaryotic Origins).

Although ORC plays a central role in the initiation of DNA replication in all eukaryotes, the interaction of this protein complex with replication origins in other species differs in some respects from that observed in *S. cerevisiae*. Fission yeast *S. pombe* has larger and more complex origins than *S. cerevisiae*. These origins contain multiple AT-rich elements important for its function. The *S. pombe* Orc4 subunit contains an N-terminal domain that binds to origin DNA even in the absence of other ORC subunits. This domain contains nine copies of the AT hook motif found in a number of DNA-binding proteins. The AT hook motif is known to bind to the minor groove of AT tracts. The N-terminal domain of *S. pombe* Orc4 may function to tether the ORC complex to origins of DNA replication. This interaction is independent of ATP. However, the tethered complex may also make ATP-dependent contacts with additional sites in the origin to nucleate formation of the initiation complex.

ATP-dependent ORC binding to DNA provides basis for the sequential recruitment of a number of additional replication factors which together form pre-RC (Figure 2). The recruitment of Cdc6 and Cdt1 occurs first, followed by the repeated

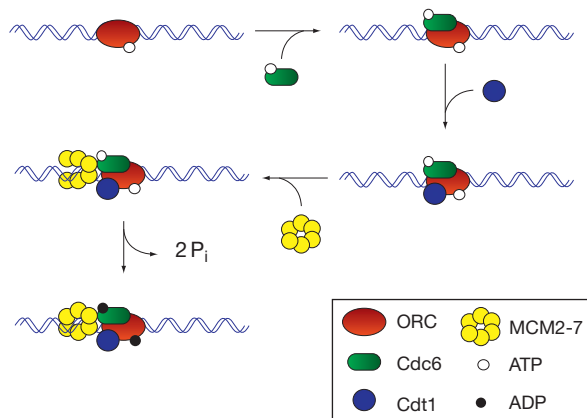


Figure 2 Assembly of the pre-RC. Binding of ORC to the replication origins marks the beginning of pre-RC assembly. On the next step, Cdc6, and then Cdt1, join origin-bound ORC. These proteins together facilitate initial association of MCM2-7 helicase with chromatin. Hydrolysis of ORC- and Cdc6-bound ATP results in loading of MCM2-7 onto DNA.

loading Mcm2-7, a hexameric-helicase complex responsible for unwinding parental DNA strands. ORC, together with Cdc6 and Cdt1, emerges as a part of a helicase loading machinery that facilitates the assembly of multiple Mcm2-7 complexes which play a crucial role in origin function. Specifically, a coordinated ATPase activity of ORC and Cdc6 facilitates tighter loading of initial MCM2-7 helicase onto the origin DNA along with recruitment and binding of additional MCM2-7 complexes.

MCM2-7 loading concludes formation of the pre-RC as shown in **Figure 2**. At this point, origin is licensed for further activity and ready to proceed to pre-IC assembly. As cells continue through the G₁ phase toward S phase transition, additional replication factors are recruited to the origin including CDKs and DDKs, Cdc45, Mcm10, GINS complex, the three eukaryotic DNA polymerases, and the eukaryotic ssDNA-binding protein, RPA. In the absence of ORC, the assembly of the pre-RC fails in both cells and cell-free systems, even so ORC by itself has only been shown to interact directly with a small subset of these factors. In these events, ORC not only serves as a landing pad but also actively participates in the assembly of pre-RC at the origins. Interestingly, the artificial recruitment of eukaryotic replication initiation factors, such as ORC and/or Cdc6, to a DNA can create a functional origin of replication.

Regulation of ORC Activity

Regulation of replication initiation and pre-RC formation is an important mechanism of preventing re-replication and avoiding genomic instability. As a critical component of these processes, ORC becomes the target of several regulatory pathways. In multicellular eukaryotes, one or more ORC subunits dissociate from the complex soon after the formation of pre-RC is complete. For example, in *Drosophila*, Orc1 subunit is selectively ubiquitinated by APC/Fzr and degraded. In mammalian cells, Orc1 is ubiquitinated through the SCF(Skp2) system and in some cases degraded, however, this subsequent degradation is not mandatory, suggesting several mechanisms of

mammalian Orc1 regulation in different cell lines/types. Cdk activities have both negative and positive roles in the initiation of DNA replication and cell-cycle progression. Numerous studies have shown the role of Cdk in the prevention of re-replication through direct inactivation of pre-RC proteins. The molecular consequences of ORC phosphorylation that inhibit pre-RC formation are not completely elucidated. It is likely, that modification directly inhibits association between ORC and other components of the pre-RC. Overall, this multi-level, cell-cycle-dependent regulation of ORC activity is a crucial step in preventing re-replication during single cell-cycle progression. ORC does not merely select the sites for assembly of pre-RC; it is also an important component of multiple cellular pathways that determines when pre-RCs are assembled. In all eukaryotic organisms, ORC subunits undergo cell-cycle-dependent modifications involving phosphorylation and ubiquitination that repress ORC activities during S, G₁, and M phases. ORC activity is restored after a completion of mitosis.

Nonreplicative Functions of ORC

Many studies indicate that the functions of ORC extend beyond DNA replication. The first discovered nonreplication function for ORC was its participation in the establishment of transcriptionally silent chromatin domains at the budding yeast silent mating type loci (HMR and HML) through the interaction with the silent chromatin protein Sir1. ORC's function in the establishment of transcriptionally repressed regions appears to be a conserved feature, as both *Drosophila* and mammalian ORC interact with HP1, a well-known modifier of position effect variegation. Mutations in the *Drosophila* Orc2 gene suppress position effect variegation and alter the localization of Hp1. siRNA knockdown of Orc2 in human cells results in delocalization of HP1. Human ORC also interacts with the histone acetyl transferase, HBO1 suggesting that histone acetylation around origins is an active process in which chromatin is remodeled by replication initiators.

Another nonreplication function of ORC is related to mitotic events in the chromosomes. Different ORC subunits in yeast, *Drosophila*, and human cells are involved in events such as sister chromatid cohesion, chromosome condensation and chromatid segregation in the mitosis, and mutations in ORC subunits cause defects in these processes.

In both *Drosophila* and human cells, the smallest subunit of ORC, Orc6, has been implicated in coordinating cytokinesis with pre-RC formation and chromosome segregation, a role that it performs independently of the rest of the complex. Orc6 had been found at the cell membranes and cytokinetic furrow in *Drosophila* and mammalian cells. RNAi/siRNA knockdown of Orc6 in either *Drosophila* or human cells results in reduction of DNA synthesis, and also in the prominent appearance of cells that have completed mitosis without cytokinesis (multi-nucleate cells), a phenotype that is not seen following knockdown of other ORC subunits. In *Drosophila*, Orc6 interacts and colocalizes with the Pnut protein, a member of the septin complex which is required for cytokinesis and other processes that involve spatial organization of the cell cortex. Experimental data suggest that Orc6 may be involved in the regulation of septin complex activities such as guanosine triphosphate hydrolysis and filament formation, and thus control dynamic

cytoskeleton events driven by septins. Also, involvement of ORC component into cytokinesis may be important for coordination between replication and subsequent mitotic events.

Concluding Remarks

Recent years of research brought significant progress in our understanding of replication–initiation events. Though basic principles of this process are universal, species-specific differences in certain components and steps of replication–initiation can be significant. This is further complicated by the plastic and adjustable character of replication. Such flexibility exists on multiple levels, from connection between replication fork progression rate and density of active origin in the genome, to diverse global patterns of replication in various tissues and on several developmental stages.

Many significant questions related to the understanding of replication–initiation mechanisms and their modifications in different species are still open. Among the most important of them are the exact criteria of origin specification, selection, and regulation. What determines whether a particular locus would be an origin and whether this origin would be active or dormant? How exactly is the activation of dormant origins achieved? What determines the timing of origin firing? The role of ORC in other than DNA replication cell functions, such as heterochromatin formation and transcriptional silencing, chromosome condensation and cohesion, and cytokinesis is an additional example of connections between different regulators of the cell cycle. In-depth study of ORC and pre-RC components functions which span beyond replication, can potentially clarify important global regulatory mechanisms connecting different seemingly isolated pathways, such as replication and chromosomal M phase events, cytokinesis, and transcriptional silencing.

See also: **Molecular Biology:** DNA Polymerase β , Eukaryotic; DNA Replication Fork, Eukaryotic.

Further Reading

- Arias EE and Walter JC (2007) Strength in numbers: Preventing rereplication via multiple mechanisms in eukaryotic cells. *Genes and Development* 21(5): 497–518.
- Bell SP (2002) The origin recognition complex: From simple origins to complex functions. *Genes and Development* 16(6): 659–672.
- Bowers JL, Randell JC, Chen S, and Bell SP (2004) ATP hydrolysis by ORC catalyzes reiterative Mcm2–7 assembly at a defined origin of replication. *Molecular Cell* 16(6): 967–978.
- Chesnokov IN (2007) Multiple functions of the origin recognition complex. *International Review of Cytology* 256: 69–109.
- Clarey MG, Botchan M, and Nogales E (2008) Single particle EM studies of the *Drosophila melanogaster* origin recognition complex and evidence for DNA wrapping. *Journal of Structural Biology* 164(3): 241–249.
- DePamphilis ML, Blow JJ, Ghosh S, et al. (2006) Regulating the licensing of DNA replication origins in metazoa. *Current Opinion in Cell Biology* 18(3): 231–239.
- Duncker BP, Chesnokov IN, and McConkey BJ (2009) The origin recognition complex protein family. *Genome Biology* 10(3): 214.
- Gilbert DM (2004) In search of the holy replicator. *Nature reviews Molecular Cell Biology* 5(10): 848–855.
- Herrick J and Bensimon A (2008) Global regulation of genome duplication in eukaryotes: An overview from the epifluorescence microscope. *Chromosoma* 117(3): 243–260.
- Machida YJ, Hamlin JL, and Dutta A (2005) Right place, right time, and only once: Replication initiation in metazoans. *Cell* 123(1): 13–24.
- Masai H, Matsumoto S, You Z, Yoshizawa-Sugata N, and Oda M (2010) Eukaryotic chromosome DNA replication: Where, when, and how? *Annual Review of Biochemistry* 79: 89–130.
- Randell JC, Bowers JL, Rodríguez HK, and Bell SP (2006) Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2–7 helicase. *Molecular Cell* 21(1): 29–39.
- Remus D and Diffley JF (2009) Eukaryotic DNA replication control: Lock and load, then fire. *Current Opinion in Cell Biology* 21(6): 771–777.
- Schepers A and Papior P (2010) Why are we where we are? Understanding replication origins and initiation sites in eukaryotes using ChIP-approaches. *Chromosome Research* 18(1): 63–77.
- Tuduri S, Tourrière H, and Pasero P (2010) Defining replication origin efficiency using DNA fiber assays. *Chromosome Research* 18(1): 91–102.

DNA Replication Fork, Bacterial

M M Spiering and S J Benkovic, The Pennsylvania State University, University Park, PA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

DNA helicase An enzyme that unwinds the two strands of a DNA duplex.

DNA ligase An enzyme that covalently joins two DNA strands together.

DNA polymerase An enzyme that synthesizes DNA by incorporating nucleotides at the 3' end of the new strand that are complementary to the nucleotides on the strand serving as the template.

Lagging strand The strand at the replication fork that is synthesized discontinuously by making shorter fragments that are sealed together by DNA ligase.

Leading strand The strand at the replication fork that is synthesized continuously as one long strand.

Primase An enzyme that synthesizes the short RNA primers to begin the discontinuous fragments on the lagging strand of a replication fork.

Two Strands of the DNA Duplex Are Copied by Different Mechanisms

DNA usually exists as a double-stranded structure, with both strands coiled together to form the characteristic double helix. The pairing of bases in DNA through hydrogen bonding means that the information contained within each strand is redundant allowing for both strands of parent DNA to serve as templates for the synthesis of new DNA. Thus, one of the strands of each daughter DNA molecule is newly synthesized and the other is passed on unchanged from the parent DNA molecule. DNA strands have directionality; the different ends of a single strand are referred to as the 3' end or the 5' end indicating which carbon atom in deoxyribose is attached to the next phosphate in the chain. In duplex DNA, the parent strands are antiparallel or oriented in opposite directions (Figure 1). This directionality has consequences for DNA synthesis, because DNA polymerase can only synthesize DNA in one direction by adding nucleotides to the 3' end of a DNA strand. The parent DNA strand oriented 3'–5' relative to the movement of the replication fork may serve as a template for the continuous synthesis of one daughter DNA strand and is called the leading-strand template. However, the other parent DNA strand is oriented 5'–3', so that the polymerase copying this strand moves away from the replication fork and is called the lagging-strand template. DNA replication on the lagging strand is accomplished by a discontinuous mechanism in which the polymerase makes short fragments that begin near the fork and extend for 1000–3000 bases (Figure 1). These fragments, known as 'Okazaki fragments' after their discoverer, are then joined together by DNA ligase to form the second daughter DNA strand. To allow for the rapid and coordinated synthesis of the leading and lagging strands, Bruce Alberts proposed that the lagging strand folds into a loop, bringing the two polymerases together (Figure 2). This has been called the trombone replication model because the expansion of the loop, as each Okazaki fragment is synthesized, resembles the slide of a trombone.

Proteins Required for DNA Replication

Much of our present understanding of bacterial DNA replication comes from the characterization of the replication system

of *Escherichia coli* and the simpler systems of *E. coli* bacteriophage T4 and T7. The fundamental components that constitute a functional replication fork in bacteriophage faithfully represent more complex systems. The types of proteins necessary for DNA replication are shown in Figure 2, and the specific proteins for each of these three systems are given in Table 1.

DNA Polymerase Holoenzymes

Replicative DNA polymerases are distinguished by their rapid and accurate DNA synthesis. The *E. coli* polymerase III, T4 polymerase gp43, and T7 polymerase gp5 are capable of extending the 3' end of an existing nucleotide chain by several hundred nucleotides every second. Replicative DNA polymerases are generally extremely accurate making only one misincorporation for every 10^7 nucleotides (nt) added. To decrease the error frequency even further, many replicative polymerases, including these three, possess a 3'–5' exonuclease proofreading activity able to remove the 3' nucleotide if a mismatch is accidentally incorporated. These polymerases alone are not processive and produce short DNA product strands each time they bind to the DNA template. To increase the processivity of these polymerases so that thousands of nucleotides can be incorporated during a single DNA-binding event, the *E. coli* and T4 polymerases interact with a circular-clamp protein that surrounds the duplex DNA and slides as the polymerase moves forward. These clamp proteins need to be loaded on the DNA by multisubunit complexes of proteins known as 'clamp loaders' using the energy from adenosine triphosphate (ATP) hydrolysis. Isolation of the *E. coli* Pol III holoenzyme yields a complex containing two polymerase cores for the leading and lagging strands, two clamps, and a single multisubunit clamp loader. The clamp loader (γ -complex) travels with the polymerases which are linked through the C-terminal domain of the two τ -subunits. The T4 proteins are less tightly associated with the holoenzyme consisting of only the polymerase and clamp; the clamp loader departs the complex once the holoenzyme is assembled. There is evidence that the two T4 polymerases at the fork interact with each other at least transiently. The T7 DNA replication system does not include a clamp or clamp loader; rather, the T7 DNA

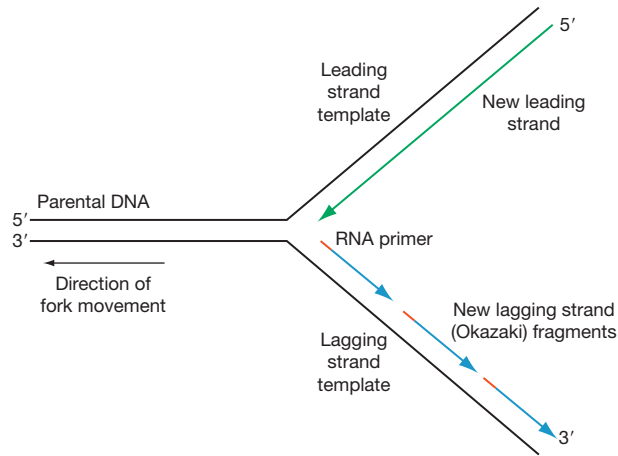


Figure 1 Schematic diagram of leading- and lagging-strand DNA synthesis at a replication fork. The parental DNA strands are antiparallel and the new DNA strands lengthen by addition of nucleotides to the 3' end causing the leading strand to be synthesized continuously, whereas the lagging strand is synthesized in short fragments known as Okazaki fragments.

polymerase is isolated as a complex with *E. coli* thioredoxin, which serves the same purpose as a clamp and increases the processivity of the polymerase.

Primosomes

A DNA polymerase can only extend an existing DNA strand paired with a template strand; it cannot begin the synthesis of a new strand. Therefore, primase proteins are required to make very short RNA chains, called 'primers', in order to initiate each Okazaki fragment for the DNA polymerase to extend on the lagging strand. The length, typically 4–12 nt, and sequences of the primers are different in each of these systems. In the bacterial replication systems, there is a close physical and functional relationship between the primase and helicase proteins. The bacterial replicative helicases are ring-shaped hexamers that surround the lagging-strand template moving 5'–3' on that strand to unwind the duplex DNA by a reaction that requires hydrolysis of nucleotide triphosphates. The DnaC helicase-loading protein promotes loading of the *E. coli* DnaB helicase by opening the preformed helicase hexamer allowing the ssDNA to bind inside. By contrast,

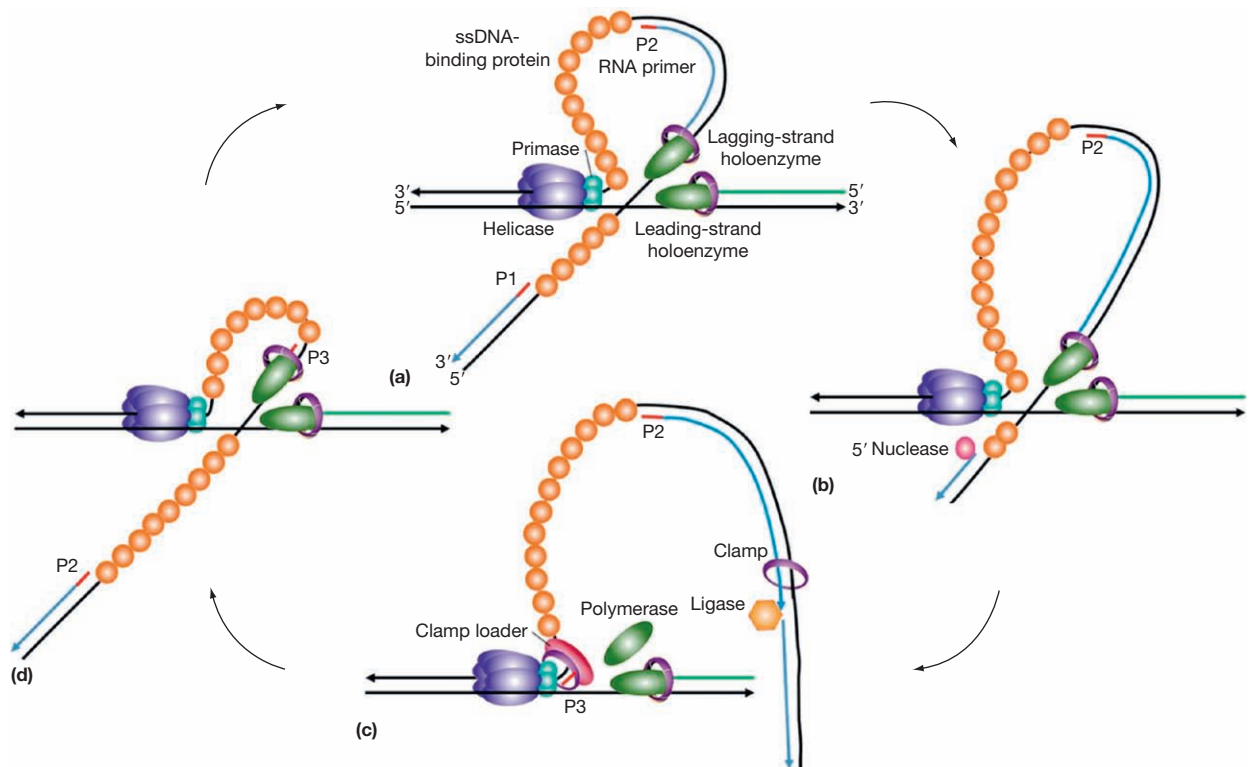


Figure 2 Cycle of steps for lagging-strand DNA synthesis at a bacterial replication fork. Depicted is the trombone model of DNA replication with the lagging strand folded to bring the leading- and lagging-strand polymerases together. (a) Extension of a lagging-strand fragment that began with primer P2. (b) The size of the loop increases as the nascent fragment is extended and the helicase continues to unwind duplex DNA. The primer P1 from the previous Okazaki fragment is hydrolyzed. (c) The polymerase is released after it completes the fragment frequently leaving behind the clamp. The nick between the fragments is sealed by DNA ligase. A new primer P3 is synthesized for the next Okazaki fragment. (d) Polymerase is transferred to the new primer P3 to begin elongation of the next fragment. The figure shows the reactions that must be completed in each cycle, but the order and timing of some of the steps, such as ligation of adjacent fragments relative to polymerase transfer, has not been established.

Table 1 Protein composition of the *E. coli* and bacteriophage T4 and T7 DNA replication systems

	<i>E. coli</i>	Phage T4	Phage T7
Polymerase holoenzyme			
Polymerase	Polymerase core (α , ϵ , θ)	gp43	gp5 + <i>E. coli</i> thioredoxin
Clamp	β (dimer)	gp45 (trimer)	
Clamp loader	γ -complex (γ , τ , τ , δ , δ' , χ , ψ)	gp44/62 complex	
Primosome			
Primase	DnaG	gp61	gp4 ^a
Helicase	DnaB (hexamer)	gp41 (hexamer)	(hexamer)
Helicase loader	DnaC (hexamer)	gp59	
ssDNA-binding protein	SSB	gp32	gp2.5
5' nuclease	DNA pol 5' nuclease	RNase H	T7 exo 6
Ligase	<i>E. coli</i> DNA ligase	T4 DNA ligase	T7 DNA ligase

^aT7 gp4 protein catalyzes both helicase and primase reactions.

the T4 helicase-loading protein binds to the DNA fork and facilitates the assembly of the gp41 helicase subunits into a hexamer surrounding the lagging-strand template. The T7 primase and helicase activities are both catalyzed by gp4, which forms a protein hexamer of stacked rings of primase and helicase domains. There is no helicase-loader protein in the T7 replication system; rather, the ssDNA binds first to a primase site and then the ring opens to allow the DNA to reach the center of the helicase ring.

Single-Stranded DNA-Binding Proteins

The single-stranded DNA of the lagging strand created by the action of the helicase is covered by tightly binding single-stranded DNA-binding proteins. In *E. coli*, this protein is a tetramer known as single-stranded-binding (SSB) protein, which the ssDNA winds itself around, while in T4 and T7, these proteins are monomers known as gp32 and gp2.5, respectively. In general, SSB proteins protect the ssDNA from nucleases until it can be replicated, prevent the formation of secondary structure in the ssDNA that can inhibit polymerases, and bind specifically to several other replication proteins, including polymerases, primases, and gp59 in the T4 system, to modulate their activities.

5'–3' Nucleases and DNA Ligases

To complete the synthesis of the lagging strand, the RNA primers of the Okazaki fragments must be removed and the fragments joined. The RNA primers are removed by 5'–3' nuclease proteins that are capable of degrading both RNA: DNA and DNA:DNA duplexes. Once the RNA primer is removed and any gap between fragments filled in by a DNA polymerase, the fragments are joined by DNA ligase.

Synthesis at DNA Replication Forks

Leading-Strand Synthesis

The continuous leading-strand DNA synthesis at the replication fork begins with an RNA primer that is made by an RNA polymerase. The leading-strand polymerase holoenzyme loads on the 3' end of the RNA primer and couples with a helicase bound to the opposite lagging-strand template. The rates of synthesis by the polymerase and duplex unwinding by the helicase are much greater when these proteins work together at a replication fork, than when these reactions occur separately. The polymerase holoenzyme and helicase are connected directly to each other in the T7 system and through the τ -subunit of the clamp loader in the *E. coli* system. The physical coupling between the polymerase holoenzyme and helicase in T4 is currently not well understood. DNA synthesis on the leading strand is extremely processive and, in principle, a single polymerase can remain bound to complete the synthesis of the chromosome of the three systems described here. However, evidence now shows that bacterial replication forks frequently stall before completing the chromosome replication and must be restarted.

Lagging-Strand Synthesis

Synthesis of the lagging strand at the replication fork begins with an RNA primer that is made by a primase. However, because the lagging strand is made in short Okazaki fragments, several RNA primers are necessary in a repeating cycle of primer initiation, elongation by DNA polymerase, and ligation to seal the fragments (Figure 2). Okazaki fragments are 1000–3000 bases in length and replication is proceeding at 400–1000 bases s^{-1} , so the cycle of reactions on the lagging strand must be completed in a few seconds. Reactions required to prime the next fragment, elongate the present fragment, and remove primers from the previous fragments can occur simultaneously allowing the lagging strand cycle to be completed rapidly.

Beginning in the elongation stage of this cycle, the lagging-strand polymerase holoenzyme is synthesizing an Okazaki fragment that began with the primer labeled P2 (Figure 2(a)). At the same time, the helicase surrounding the lagging-strand template at the fork is unwinding the duplex ahead of the leading-strand polymerase holoenzyme. The ssDNA being generated by the helicase along with the ssDNA remaining to be replicated ahead of the lagging-strand polymerase holoenzyme is coated with SSB proteins.

As the lagging strand loop continues to grow due to the activities of the helicase and lagging-strand polymerase holoenzyme, the 5'–3' nuclease is removing the RNA primer (P1) and a small amount of adjacent DNA from the previous fragment (Figure 2(b)). The removal of this DNA first added to the primer may increase the accuracy of replication.

When the lagging-strand polymerase holoenzyme completes the nascent fragment, creating a nick, the clamp releases the polymerase and stays behind on the DNA (Figure 2(c)). In *E. coli*, this release of the polymerase core from the β -clamp is facilitated by the C-terminal domain of the τ -subunit of the clamp loader acting as an unloader. There is no evidence that a clamp loader is needed to disengage the clamp and polymerase in the T4 system. If the polymerase fails to be released at

the nick, it can continue to synthesize DNA displacing the 5' end of the downstream fragment forming a flap. However, all of the prokaryotic 5' nucleases involved in primer removal also have flap endonuclease activities able to remove these displaced strands. The nicks that are formed between adjacent fragments are sealed by DNA ligase.

When the synthesis of the nascent fragment is complete, primase, associated with the helicase at the fork, makes the RNA primer (P3) that will be used to start the next fragment (**Figure 2(c)**). A new clamp is loaded on to the primer in the *E. coli* and T4 systems by the clamp loader. The *E. coli* clamp loader is bound to the two polymerases as well as the helicase at the fork, while the clamp is recruited from solution. Dilution experiments suggest that both the T4 clamp and clamp loader must be provided from solution for each cycle of lagging strand synthesis. Sometimes, the primase makes a primer before the lagging-strand polymerase holoenzyme finishes the previous Okazaki fragment. This acts as a signal that causes the lagging-strand polymerase to be released prematurely leaving a gap between fragments that must ultimately be repaired by loading another polymerase.

Finally, the lagging-strand polymerase is transferred to the clamp-loaded new primer (P3) to begin synthesis of the next fragment (**Figure 2(d)**). In the case of T7, the DNA-binding domain of the primase binds the new primer and transfers it directly to the lagging-strand polymerase. The same lagging-strand polymerase is recycled from one Okazaki fragment to the next being held at the replication fork by interactions with the clamp loader in *E. coli*, the leading-strand polymerase in T4, or the helicase in T7.

Coordination of Leading- and Lagging-Strand Synthesis

The cycle of reactions for lagging-strand synthesis must be carefully coordinated with each other as well as with the reactions on the leading strand. Experiments using synthetic templates, which allow for the separate detection of leading- and lagging-strand synthesis, have demonstrated that the leading-strand synthesis is stopped either when a chain terminator was added or by the absence of rNTPs for primer synthesis to block lagging-strand synthesis, indicating a close interaction between the polymerases on the two strands. With the complexity of the multiple steps involved in primer and Okazaki fragment synthesis on the lagging strand and the relatively straightforward continuous incorporation of nucleotides on the leading strand, two models have been proposed to maintain this coordination. In the first model, the leading-strand synthesis temporarily stops or pauses to allow additional time for primer synthesis and polymerase recycling to occur on the lagging strand, after which the coordinated synthesis on both the leading and lagging strands continues. Some evidence suggests that the T7 system may operate this way. In the second model, the leading-strand polymerase holoenzyme never pauses; rather, the lagging-strand polymerase is occasionally signaled to recycle prematurely before completing the nascent fragment, thereby catching up to the leading-strand holoenzyme. Evidence

suggests that in the *E. coli* and T4 systems, the leading-strand synthesis never stops and that lagging-strand polymerase recycling occurs prematurely up to 50% of the time.

DNA Replication within the Cell

DNA replication is initiated at particular points within the DNA known as 'origins'. The *oriC* origin in *E. coli* contains an A-T rich sequence recognized by the replication initiator protein, DnaA. DnaA subsequently recruits other proteins beginning a series of reactions that result in the separation of the DNA strands at the origin, forming a bubble with two replication forks at either end. In bacteria, which have a single origin of replication on their circular chromosome, the bidirectional DNA synthesis eventually creates a theta structure because the DNA resembles the Greek letter theta, θ . Termination of replication occurs when the two replication forks meet each other on the opposite side of the parental chromosome. *E. coli* regulate this process through the use of the Tus protein and termination sequences that allow replication forks to pass through in only one direction, but not the other. Thus, the two replication forks are constrained to always meet within the termination region of the chromosome. Regulation of DNA replication in *E. coli* is achieved through several mechanisms, including the ratio of ATP to adenosine diphosphate (ADP) and the levels of DnaA within the cell, as well as the hemimethylation and sequestering of *oriC*. DNA synthesis results in hemimethylated DNA recognized by the SeqA protein. SeqA binds the hemimethylated DNA and sequesters the origin sequence preventing newly replicated origins from immediately initiating another round of DNA replication before cell division.

See also: **Molecular Biology:** DNA Helicases: Hexameric Enzyme Action; DNA Polymerase III, Bacterial; DNA Replication: Initiation in Bacteria; Processivity Clamps in DNA Replication; Sliding Clamps in DNA Replication: *Escherichia coli* β -Clamp and PCNA Structure.

Further Reading

- Hamdan SM and Richardson CC (2009) Motors, switches, and contacts in the replisome. *Annual Review of Biochemistry* 78: 205–243.
- Kornberg A and Baker T (1992) *DNA Replication*, 2nd edn. San Francisco: W. H. Freeman.
- Manosas M, Spiering MM, Zhuang Z, Benkovic SJ, and Croquette V (2009) Coupling DNA unwinding activity with primer synthesis in the bacteriophage T4 primosome. *Nature Chemical Biology* 5: 904–912.
- Nossal NG, Makhov AM, Chastain PD, II, Jones CE, and Griffith JD (2007) Architecture of the bacteriophage T4 replication complex revealed with nanoscale biointerferometry. *Journal of Biological Chemistry* 282: 1098–1108.
- Pandey M, Syed S, Donmez I, et al. (2009) Coordinating DNA replication by means of priming loop and differential synthesis rate. *Nature* 462: 940–943.
- Spiering MM, Nelson SW, and Benkovic SJ (2008) Repetitive lagging strand DNA synthesis by the bacteriophage T4 replisome. *Molecular BioSystems* 4: 1070–1074.
- Tanner NA, Loparo JJ, Hamdan SM, et al. (2009) Real-time single-molecule observation of rolling-circle DNA replication. *Nucleic Acids Research* 37: e27.
- Yao NY, Georgescu RE, Finkelstein J, and O'Donnell ME (2009) Single-molecule analysis reveals that the lagging strand increases replisome processivity but slows replication fork progression. *Proceedings of the National Academy of Sciences of the United States of America* 106: 13236–13241.

DNA Replication Fork, Eukaryotic

L M Kelman, Montgomery College, Germantown, MD, USA

Z Kelman, Institute for Bioscience and Biotechnology Research, Rockville, MD, USA; University of Maryland, College Park, MD, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Lori M. Kelman, Jerard Hurwitz, and Zvi Kelman, volume 1, pp. 745–748, © 2004, Elsevier Inc.

Glossary

Bidirectional DNA replication DNA synthesis that originates at the origin of replication and results in the formation of two replication complexes leading to two replication forks that move in opposite directions.

DNA polymerase Enzyme that utilizes a primed single-stranded DNA (ssDNA) template to synthesize the complementary strand.

Helicase Enzymes that use the energy derived from hydrolysis of nucleoside triphosphates to break the hydrogen bonds that hold the duplex DNA together.

Primase Enzyme that synthesizes small RNA chains *de novo* on ssDNA resulting in RNA–DNA hybrids, which are used as primers for DNA polymerases.

Sliding clamp Ring-shaped protein that encircles duplex DNA and binds to the DNA polymerase, endowing it with high processivity.

Introduction

All organisms must replicate their chromosomal DNA in order to propagate their genetic information. During the S phase of the cell cycle, DNA replication duplicates chromosomes into two identical copies that segregate to daughter cells during mitosis. Chromosomal DNA replication begins at regions of the chromosomes called origins of replication and can be divided into three phases: initiation, elongation, and termination. In the initiation stage, an origin recognition protein (ORP) binds the origin of replication to form the initial replication bubble and recruits additional initiation factors to the origin. Next, the helicase is recruited and assembled on the DNA. The single-stranded DNA (ssDNA) exposed behind the helicase is coated with ssDNA-binding protein (SSB). Primase, the polymerase, and the rest of the replication machinery are associated with the replication bubble to form two replication forks and to initiate bidirectional DNA synthesis (the elongation phase). During termination, replication forks collide and are resolved, and the resulting daughter DNA molecules are completed and separated. A number of proteins and complexes that participate in replication fork progression are described in the article.

Due to the antiparallel nature of the DNA and the unidirectionality of the polymerase, one strand of the chromosome is synthesized continuously (the leading strand) while the other is copied discontinuously (the lagging strand) as a series of Okazaki fragments (Figure 1). It is believed that at the replication fork, the two polymerases responsible for replicating the leading and lagging strand are associated (either directly or indirectly via other molecules). Thus, the replication of the two strands is coupled.

Minichromosome Maintenance Complex

Minichromosome maintenance (MCM) is a family of six proteins (Mcm2–7, molecular masses of 101, 91, 97, 82, 93, and 81 kDa, respectively) with highly conserved amino acid

sequences between the six different polypeptides. The family was first identified in a screen for proteins required for the maintenance of minichromosomes in yeast. All six proteins have been identified in all eukarya. Based on genetic and biochemical studies, the MCM complex is presumed to be the helicase responsible for the separation of duplex DNA during chromosomal replication. All MCM proteins are essential for cell viability. They are needed for the initiation phase as well as for replication fork progression. In addition to the MCM heterohexamer, *in vivo* and *in vitro* studies have revealed the presence of several additional MCM complexes composed of different combinations of MCM proteins. Biochemical studies with these various complexes from different organisms showed that *in vitro* a dimeric complex of the Mcm4,6,7 heterotrimer contained 3'–5' DNA helicase activity, ssDNA and double-stranded DNA (dsDNA) binding, DNA-dependent ATPase activity, and could translocate along ssDNA and dsDNA and unwind a DNA–RNA hybrid when translocating on the DNA strand. It was also found that the heterohexamer of Mcm2–7 possesses 3'–5' DNA helicase activity *in vitro*, but only in the presence of accessory factors Cdc45 and the GINS complex [referred to as the CMG (Cdc45–MCM–GINS) complex]. It is believed that the CMG complex is the active helicase at the replication fork. The 3'–5' direction of translocation on DNA by the MCM proteins suggests that the helicase moves along the leading strand.

Replication Protein A

Replication protein A (RPA) is the eukaryotic ssDNA-binding protein. It is a heterotrimeric complex of proteins with molecular masses of 14, 32, and 70 kDa; these essential proteins have been found in all eukarya studied. SSBs are essential components of all replication systems. The protein coats the ssDNA exposed behind the helicase, protecting the DNA from attack by nucleases and chemical modification. In addition, RPA stimulates the activity of DNA polymerases by removing

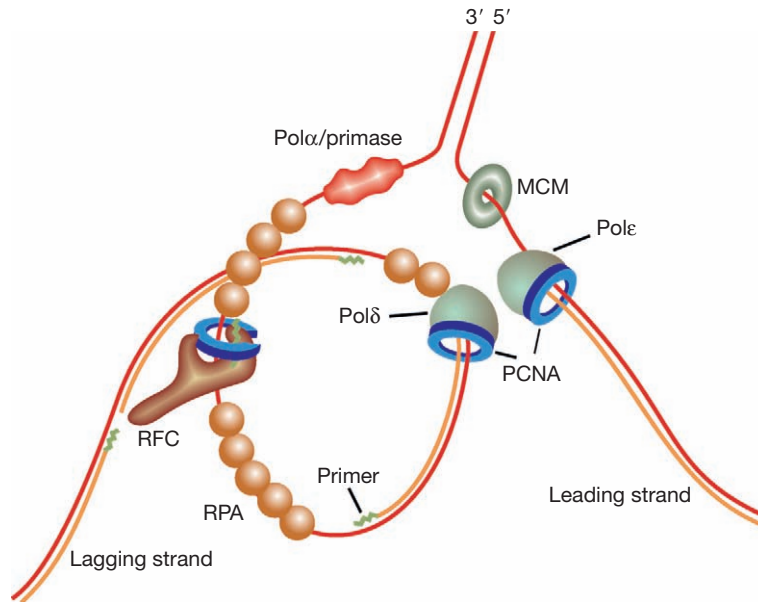


Figure 1 A schematic representation of the proteins at the eukaryotic replication fork. Although it is not shown in the figure, it is thought that all proteins shown interact with each other but the factor responsible for the interaction has not been identified.

secondary structures that interfere with polymerase movement, creating a uniform substrate for polymerase activity. RPA was shown to play an instrumental role in the coordination of DNA synthesis on the lagging strand and to interact with a large number of cellular factors needed for DNA metabolism.

DNA Polymerase α /Primase Complex

DNA polymerases are incapable of initiating DNA synthesis *de novo* and require DNA primases, which synthesize short RNA primers on template DNA that are subsequently extended by the polymerase. In eukarya, DNA primase is part of a four-subunit complex, the Pol α /primase complex, containing subunits with molecular masses of 180, 68, 58, and 48 kDa. Two of the subunits, p48 and p58, are required for primase activity, with p48 serving as the catalytic unit. Primase synthesizes short RNA primers (8–12 nucleotides (nt)), which are then elongated by polymerase α (Pol α) to ~30 nt, forming pre-Okazaki fragments. These RNA–DNA hybrid molecules are recognized by the polymerase accessory complex, replication factor C (RFC), to initiate processive DNA synthesis by the replicative polymerase. On the lagging strand, Pol α /primase activity is required at the initiation of each Okazaki fragment.

Polymerase Accessory Proteins

The processivity of DNA polymerases is low. This means that only a few nucleotides are incorporated into the newly synthesized DNA before the polymerase dissociates from the substrate. Processivity is conferred by a ring-shaped protein, referred to as the DNA polymerase sliding clamp (also called DNA polymerase processivity factor), which encircles duplex DNA and acts to tether the polymerase catalytic unit to the

primed template for processive DNA synthesis. The eukaryotic sliding clamp is the proliferating cell nuclear antigen (PCNA). PCNA, as well as the other sliding clamps, cannot assemble itself around the DNA, but must be loaded onto DNA by a protein complex, known as the clamp loader (also called polymerase accessory complex), which in eukarya is RFC. RFC recognizes the 3' end of the single-strand/duplex (primer–template) junction of the pre-Okazaki fragment and utilizes ATP hydrolysis to assemble PCNA around the primer. PCNA encircles the primer DNA and then binds the DNA polymerase, allowing rapid and processive DNA synthesis. Upon completion of an Okazaki fragment, the polymerase dissociates, leaving the clamp assembled around the duplex DNA.

Proliferating Cell Nuclear Antigen

PCNA, the eukaryotic sliding clamp, is a ring-shaped homotrimer of a 29-kDa protein, which is assembled on the leading strand by RFC and interacts with DNA Pol ϵ , the leading strand polymerase, to initiate processive DNA synthesis. On the lagging strand, PCNA must be loaded on each primer. Upon assembly around the primer of the pre-Okazaki fragment by RFC, PCNA associates with DNA Pol δ to initiate processive DNA synthesis. Upon completion of an Okazaki fragment, PCNA disengages from the polymerase and is left on the duplex DNA. The PCNA molecules left on replicated DNA play diverse roles in postreplication DNA metabolic processes by interacting with and stimulating a large number of enzymes. Examples of enzymes stimulated by PCNA include Fen1 and DNA ligase, needed for Okazaki fragment maturation. PCNA is a target for cell-cycle regulators. A number of proteins that inhibit DNA synthesis were shown to operate via interaction with PCNA, preventing its association with the polymerase.

Replication Factor C

RFC, the eukaryotic clamp loader, is a five-subunit complex of 140-, 40-, 38-, 37-, and 36-kDa proteins (for Rfc1–5, respectively). All five subunits are essential for cell viability and all belong to the AAA+family of ATPases. In addition to the conserved domains found in all members of this family, the large subunit (Rfc1) shares homology with prokaryotic DNA ligases at its N-terminal region. RFC recognizes the 3' end of the pre-Okazaki fragment and utilizes ATP hydrolysis to assemble PCNA around the duplex DNA. It was shown that the interaction between RFC and RPA plays a role in the switch that removes Pol α /primase from DNA, allowing the loading of PCNA. In addition to its role as a clamp loader, *in vitro* studies suggest that RFC also functions as a clamp unloader, that is, it removes PCNA from DNA.

Polymerase δ and Polymerase ϵ

Both Pol δ and Pol ϵ are found in all eukarya and are essential polymerases required for chromosomal DNA replication. Pol δ is a heterotetramer of proteins with molecular masses of 125-, 66-, 50-, and 12-kDa proteins, while Pol ϵ is comprised of proteins of 261, 59, 17, and 12 kDa. Both polymerases were shown to be monomeric in solution. Following the assembly of PCNA at the primer, Pol δ and Pol ϵ associate with PCNA to initiate rapid and processive DNA synthesis. *In vivo* genetic studies and *in vitro* biochemical characterization have suggested that Pol ϵ and Pol δ act on different DNA strands. These studies suggested that Pol ϵ is the leading strand polymerase while Pol δ is the lagging strand enzyme.

Topoisomerases

During replication, DNA is supercoiled in front of and behind the polymerase. Positive supercoils are produced in front of the replication fork, while negative supercoils result behind it. Topoisomerases prevent excessive supercoiling and regulate the level of supercoiling within the DNA molecule. These enzymes are essential for replication fork progression. Topoisomerases introduce transient breaks in the phosphodiester backbone of the DNA and allow supercoils to be added or removed.

Additional Proteins

The replication of eukaryotic chromosomes is a very complex and highly coordinated process. To date, dozens of proteins have been shown to participate in the replication of eukaryotic chromosomes. However, the replication fork continues to be studied, and recent work suggests that more proteins may be a part of the fork, such as the GINS complex, MCM10, and Cdc45. The roles, if any, of these enzymes at the fork remain to be elucidated.

See also: **Molecular Biology:** DNA Helicases: Hexameric Enzyme Action; DNA Polymerase β , Eukaryotic; DNA Replication: Eukaryotic Origins and the Origin Recognition Complex; Eukaryotic DNA Polymerase δ ; Processivity Clamps in DNA Replication.

Further Reading

- Bochman ML and Schwacha A (2009) The MCM complex: Unwinding the mechanism of a replicative helicase. *Microbiology and Molecular Biology Reviews* 73: 652–683.
- Bowman GD, Goedken ER, Kazmirski SL, O'Donnell M, and Kuriyan J (2005) DNA polymerase clamp loaders and DNA recognition. *FEBS Letters* 579: 863–867.
- Burgers PM (2009) Polymerase dynamics at the eukaryotic DNA replication fork. *Journal of Biological Chemistry* 284: 4041–4045.
- Costa A and Onesti S (2009) Structural biology of MCM helicases. *Critical Reviews in Biochemistry and Molecular Biology* 44: 326–342.
- Garg P and Burgers PM (2005) DNA polymerases that propagate the eukaryotic DNA replication fork. *Critical Reviews in Biochemistry and Molecular Biology* 40: 115–128.
- Hübscher U (2009) DNA replication fork proteins. *Methods in Molecular Biology* 521: 19–33.
- Kunkel TA and Burgers PM (2008) Dividing the workload at a eukaryotic replication fork. *Trends in Cell Biology* 18: 521–527.
- O'Donnell M and Kuriyan J (2006) Clamp loaders and replication initiation. *Current Opinion in Structural Biology* 16: 35–41.
- Sclafani RA and Holzen TM (2007) Cell cycle regulation of DNA replication. *Annual Review of Genetics* 41: 237–280.
- Vivona JB and Kelman Z (2003) The diverse spectrum of sliding clamp interacting proteins. *FEBS Letters* 546: 167–172.
- Yao N and O'Donnell M (2009) Replisome structure and conformational dynamics underlie fork progression past obstacles. *Current Opinion in Cell Biology* 21: 336–343.

Relevant Websites

- <http://www.hhmi.org> – Howard Hughes Medical Institute.
- <http://www.ncbi.nlm.nih.gov> – National Center for Biotechnology Information.

DNA Replication: Initiation in Bacteria

J M Kaguni, Michigan State University, East Lansing, MI, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

DNA polymerase III holoenzyme The replicative DNA polymerase of *E. coli*.

DNA replication The process of duplication of DNA.

DnaA box A 9-bp DNA sequence motif recognized by DnaA protein.

DnaA protein DnaA is the initiator of DNA replication, recognizing DNA sequence motifs named the DnaA box, the I-site, and the τ -site in the chromosomal replication origin.

DnaB protein Composed of six identical subunits, DnaB helicase is a ring-shaped protein with a central cavity through which one of the two strands of DNA passes as the helicase unwinds the parental duplex DNA.

DnaC protein DnaC forms a complex with DnaB helicase to inhibit its activity as a DNA helicase and ATPase.

Formation of the DnaB–DnaC complex is required for DnaA to load DnaB at *oriC*.

I-site A sequence within *oriC* bound by DnaA complexed to ATP.

oriC The *Escherichia coli* chromosomal replication origin where DNA replication starts.

Primase This enzyme forms primers that are extended by DNA polymerase III holoenzyme during DNA replication.

SSB SSB is an acronym for single-stranded DNA-binding protein.

τ -site A sequence within *oriC* recognized by DnaA complexed to ATP.

In free-living organisms, including *Escherichia coli*, DNA replication starts at replication origins. *E. coli* contains a single replication origin (*oriC*), which serves as the site for assembly of the replisome, a multi-protein molecular machine that duplicates the circular duplex genome. Replisome assembly is a highly regulated event that is strictly coordinated with the cell cycle.

The *E. coli* Replication Origin, *oriC*

The biochemical mechanism of DNA replication of the *E. coli* chromosome has been extensively studied. At *oriC*, DnaA directs the formation of two replisomes for bidirectional fork movement. As an adenine–nucleotide-binding protein, DnaA specifically recognizes DNA sequences in *oriC* and then loads DnaB helicase bound to its partner, DnaC. The focus of this article is on the biochemical mechanism of initiation. Recent excellent reviews offer different perspectives of this process and its regulation, and are listed in the section Further Reading.

DNA Sequence Motifs in *oriC*

E. coli oriC has been studied by a variety of approaches. A number of DNA sequence motifs that are functionally important have been identified (Figure 1). One is an AT-rich region near the left border that carries three 13-mer motifs. In the absence of magnesium ion *in vitro*, this region has a natural tendency to unwind and is also called the DNA unwinding element (DUE). In the presence of magnesium ion, DnaA is required for unwinding as described in more detail later. Between this region and the right border are discrete binding sites recognized by DnaA and other proteins. Three types of binding sites for DnaA have been identified. One is called the

DnaA or R box, whose consensus sequence is TTA/TTNCACA. Each of these boxes is designated as R1 through R5 (or M) to denote their positions within *oriC*. A domain (domain IV, see Table 1) near the C-terminus of DnaA recognizes these DnaA boxes. Other sites in *oriC* are named I- and τ -sites.

The affinity of DnaA for the DnaA boxes varies, which apparently reflects the influence of differences in both their sequence composition and flanking sequences. Their order of affinities is $R4 \geq R1 > R2 > R3 \approx R5$. As DnaA binds to adenosine triphosphate (ATP) and adenosine diphosphate (ADP), either nucleotide-bound form binds with comparable affinity to R4 and R1. For the other DnaA boxes, DNase I footprinting with a polymerase chain reaction (PCR)-amplified DNA containing *oriC* showed that DnaA–ATP binds more avidly than DnaA–ADP to R2, R3, and R5, but this difference was not observed with dimethyl sulfate as the cleaving reagent in the footprinting of DnaA bound to a supercoiled plasmid carrying *oriC*. As an explanation for the disparity in results, the difference may arise from the effect of DNA topology (supercoiled compared with a linear DNA). A second possibility is that the inclusion of poly (dI–dC) and poly (dA–dT) in the DNase I experiments and their omission in the experiments with dimethyl sulfate leads to this difference. In comparison, DnaA complexed to ATP binds with lower affinity to sites named I- and τ -sites. At present, it is not known if domain IV of DnaA interacts directly with the I- and τ -sites. However, because the I-sites differ from the R box consensus only at 3–4 bases, it is likely that domain IV of DnaA recognizes both of these types of DNA sequences. As *oriC* contains 11 GATC sites recognized by DNA adenine methylase (Dam) and several overlap the τ 1, τ 2, I2, and I3 sites, the methylated state of these sites in the *oriC*-containing plasmid compared with the nonmethylated PCR-amplified DNA may influence binding affinity.

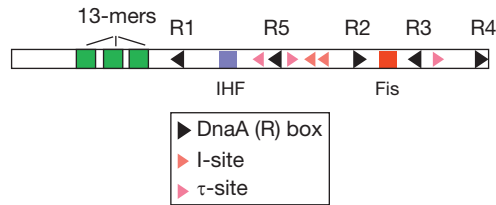


Figure 1 DNA sequence motifs in *E. coli oriC*. Three 13-mers (green boxes) are located near the left border. The black arrowheads denote the orientation of the DnaA boxes named R1, R2, R3, R4, and R5 (or M). The I- and τ -sites bound by DnaA complexed to ATP, and binding sites for IHF (purple box) and Fis (red box) are also shown.

Table 1 Functional domains of DnaA^a

Domain	Region of DnaA	Function
I	1–90	DnaA oligomerization Interaction with DnaB–DnaC, HU, DiaA, or Dps
II	91–130	Linker connecting domains I and III
III	131–347	ATP binding and ATP hydrolysis DnaA oligomerization (Box VII) Interaction with DnaB (residues 130–148)
IV	348–467	DNA binding Interaction with acidic phospholipids (residues 372–381)

^aDomain I acts in the recruitment of the DnaB–DnaC complex, in DnaA oligomerization, and in the interaction of DnaA with HU, DiaA or Dps. Distinct regions within domain I act in DnaA oligomerization and in the recruitment of the DnaB–DnaC complex. Domain II may function as a flexible linker, and participate in loading the DnaB–DnaC complex. Domain III contains the Sensor I, II (Box VIII), and Box VII motifs of the AAA⁺ protein family, and functions in ATP binding, ATP hydrolysis, and in DnaA oligomerization. A region (residues 130–148) within domain III also interacts with DnaB. Domain IV functions to recognize the DnaA box motif. A region within domain IV (residues 372–381) is implicated in membrane binding.

Of interest, the R4 box can be scrambled or the right half of *oriC* can be removed entirely from the chromosomal locus without loss of viability. However, initiation is severely asynchronous compared with the synchronous initiation at the respective origins in wild-type cells harboring replicating chromosomes. Evidently, sequences in this region that include the R4 box are needed for the temporal regulation of initiation. The importance of sequences in the right-half of *oriC* recognized by DnaA complements several lines of evidence showing that the ordered assembly of DnaA is essential for the coordination of initiation with the cell cycle.

Binding Sites for Other Proteins

Sites for the binding of integration host factor (IHF) and factor for inversion stimulation (Fis) are located between R1 and R5, and R2 and R3, respectively (Figure 1). Both nucleoid-associated proteins bend DNA upon binding. *In vivo*, footprinting studies of *oriC* minichromosomes in synchronous cultures revealed that the R3 box was occupied at the time of initiation. At other times, the Fis site was protected, but R3 was not. These results suggest that the exchange of Fis with DnaA at the

respective sites is essential for initiation to occur at the proper time in the bacterial cell cycle. Mutation of either binding site disrupted *oriC* function when carried in a plasmid, suggesting that the binding of these proteins to these sites is important for initiation. In support, Fis and IHF mutants, although viable, initiate asynchronously and fail to maintain *oriC*-dependent plasmids, but the IHF mutant requires an additional *polA* mutation to elicit the latter defect. The conundrum is that mutation of these binding sites in the bacterial chromosome does not inhibit *oriC* function *in vivo*, and plasmids with mutations in the Fis or IHF binding sites are active *in vitro*.

Replisome Assembly Is a Stepwise Process

The biochemical mechanism of initiation at *oriC* can be separated into discrete steps. Studies of these events rely in part on an *in vitro* DNA replication system that requires a supercoiled plasmid containing *oriC* and other purified components, including DnaA, DnaB, DnaC, single-stranded DNA-binding protein (SSB), HU, primase DNA polymerase III holoenzyme, and DNA gyrase. By the following criteria, this system recapitulates the *in vivo* process. First, the *in vitro* system absolutely requires the *oriC* sequence; mutations that abolish *oriC* function *in vivo* when carried in a plasmid also inactivate *oriC* in the *in vitro* system. Second, replication forks that propagate from *oriC* move bidirectionally as they do *in vivo*. Third, the *in vitro* system requires many of the same gene products that are required in living cells. This system coupled with other biochemical approaches has led a mechanistic view of the function of each required protein.

The DnaA–*oriC* Complex

As an adenine-nucleotide-binding protein, DnaA contains the amino acid sequence motifs shared by the AAA⁺ family of proteins. These motifs are the Walker A and B boxes, which respectively function in nucleotide binding and in coordinating magnesium chelated to the phosphates of the nucleotide, and sensor I, II, and box VII motifs that are considered to both promote and coordinate ATP hydrolysis with a conformational change. As predicted in the initiator titration model, initiation occurs at a particular cell mass when the abundance of DnaA reaches the level necessary for initiation. Thus, DnaA binds to the high-affinity sites early in the bacterial cell cycle. As ATP is more abundant than ADP *in vivo*, DnaA that is newly synthesized more likely binds ATP over ADP. As DnaA–ATP accumulates, it binds to the low-affinity sites, which include the R3 box and the I- and τ -sites, forming the DnaA–*oriC* complex. In support, *in vivo* footprinting experiments in synchronized cells reveal that DnaA is bound to R1, R4, and R2 throughout the cell cycle, and that R3 becomes protected at the time of initiation. When this study was performed, the other low-affinity sites had not been discovered.

In vivo and *in vitro* evidence indicates that formation of the DnaA–*oriC* complex requires an interaction between and/or among DnaA molecules. Two regions within DnaA are required. Based on studies of a mutant DnaA carrying an alanine substitution for tryptophan 6 (W6A), one region appears to be a hydrophobic surface near the N-terminus.

This mutant protein is defective in an assay (see later) that measures the assembly of DnaA at *oriC* in a supercoiled plasmid followed by the recruitment of DnaB complexed to DnaC. Other studies that rely on footprinting methods showed that this mutant DnaA fails to protect the I- and τ -sites within *oriC*. Together, these results suggest that the hydrophobic surface functions in DnaA oligomerization to stabilize the interaction between and/or among DnaA molecules bound to the I- and τ -sites. The second region was identified via the analysis of a mutant bearing an alanine substitution for a conserved arginine at residue 281 (R281A) in the box VII motif within domain III. Like the W6A-substituted protein, the R281A mutant was poorly retained on a supercoiled plasmid containing *oriC*, suggesting that this arginine is necessary for self-oligomerization of DnaA. The latter observations correlate with structural studies of *Aquifex aeolicus* DnaA (lacking residues 1–76), in which domain III of one molecule interacts with this domain in the adjacent molecule in the DnaA oligomer. In this structure, DnaA assembles as a right-handed helical filament in which a separate arginine at position 285 in box VII interacts with ATP bound between adjacent protomers.

HU or DiaA Stabilizes DnaA at *oriC*

Recent experiments indicate that other proteins stabilize the DnaA oligomer bound at *oriC*. One protein is HU, a dimer that binds nonspecifically to DNA to form the nucleoid. The composition of HU is influenced by the growth phase of a culture. In early log phase, the α_2 dimer predominates, followed by the relative prevalence of the $\alpha\beta$ heterodimer during mid-log growth. The β_2 dimer becomes the principal form of HU as cells enter stationary phase. During initiation, HU was originally found to stimulate the DnaA-dependent unwinding of the 13-mer region in *oriC*. Despite HU's nonspecific DNA-binding activity, it was localized at or near *oriC* by immunoelectron microscopy. Recent work showed that HU interacts via its α subunit with the N-terminal region of DnaA to stabilize the DnaA oligomer at *oriC*. These *in vitro* results provide an explanation for these earlier observations and are supported by *in vivo* evidence that mutants of *hupA*, which encode the α subunit of HU, initiate at abnormal times in the bacterial cell cycle. As this phenotype of asynchronous initiation is also observed with other mutations that affect the function of either DnaA or *oriC* to modulate the frequency of initiation, these results suggest that DnaA oligomerization is essential for initiation to proceed normally.

Like HU, DiaA also stabilizes DnaA at *oriC*. Apparently, each protomer of the DiaA tetramer interacts with the N-terminal region of DnaA to stabilize it at the low-affinity I- and τ -sites. DiaA was originally identified in a genetic method to obtain mutations that suppressed the lethal effect caused by DnaAcos, a mutant DnaA that hyperinitiates. A simple explanation is that the lack of DiaA function leads to the diminished stability of DnaAcos oligomerized at *oriC*, which reduces the frequency of initiation to a level that no longer interferes with viability. Similar to mutants in the α subunit of HU, *diaA* mutants initiate asynchronously, so DiaA is physiologically important. That both HU and DiaA stabilize the DnaA oligomer suggests that conditions may arise *in vivo* when HU is favored over DiaA or the converse.

DnaA-Dependent Unwinding of the AT-Rich Region of *oriC* and Proteins That Modulate This Process

Bound to *oriC*, DnaA then unwinds the AT-rich region carrying the 13-mers near the left border. As DnaA complexed to ATP γ S is as effective as DnaA–ATP, nucleotide hydrolysis is not required. However, DnaA–ADP is essentially inactive. Considering that domains III and IV of *Aquifex aeolicus* DnaA assemble into a helical filament in the presence of ATP, whereas it forms a ring-like structure with ADP, ATP bound between adjacent molecules of the DnaA oligomer alters the conformation of domain III to prevent the flat ring arrangement. As the interior of this DnaA oligomer contains positively charged surfaces, a proposed model is that one or more basic residues bind to the negatively charged unwound DNA. In support, an alanine substitution for lysine 245 that is positioned on the interior surface of the filament leads to a specific defect in unwinding. Alanine substitutions of other basic residues (lysine 223 and lysine 243) are also defective in unwinding and in self-oligomerization to support the model that the DnaA oligomer binds to and stabilizes the unwound region of *oriC*. A prediction of this model is that all mutant DnaAs that are defective in self-oligomerization should be unable to unwind *oriC*. However, mutants bearing W6A or R281A substitutions are impaired in self-oligomerization, yet they remain as active as wild-type DnaA in *oriC* unwinding.

HU, DiaA, or IHF stimulate the DnaA-dependent unwinding of *oriC*. Stimulation by HU or DiaA correlates with both their interaction with residues in the N-terminal region (domain I) of DnaA and their ability to stabilize the DnaA oligomer at *oriC*. As IHF is known to bend DNA and also stimulate the binding of DnaA to the weaker sites in *oriC*, DNA bending by IHF at the IHF site in *oriC* may stabilize the DnaA oligomer by facilitating the interaction between or among DnaA molecules. By comparison, Fis obstructs the binding of DnaA to the low-affinity sites to inhibit the unwinding of *oriC*.

DNA-binding protein from starved cells (Dps) is expressed in response to oxidative stress and accumulates in starved cells. *In vivo*, Dps causes less frequent initiations. Biochemically, this protein interacts with the N-terminal region of DnaA and also inhibits the unwinding of *oriC*. These results support a model which proposes that Dps acts as a checkpoint during oxidative stress to reduce the frequency of initiation. This checkpoint may provide time for cells to repair the oxidative damage to the genome before it is duplicated.

Other Proteins Required for Initiation

DnaC

Similar to DnaA, DnaC is a member of the AAA+ ATPase family. Although many AAA+ proteins are ring-shaped oligomers, DnaC is purified as a monomer. Several studies showed that DnaC binds ATP, albeit with low affinity ($K_D \sim 8 \mu\text{M}$), and is a weak ATPase. To support the idea that ATP binding by DnaC is essential, genetic studies showed that missense mutations in each AAA+ motif impair DnaC function. *In vitro* work indicated that substitution of a conserved lysine with arginine in the Walker A box results in undetectable ATP binding and inactivity in DNA replication of an *oriC*-containing plasmid.

Recent work revealed that one of the AAA+ motifs, named box VII, appears to coordinate ATP hydrolysis with conformational changes that lead to the dissociation of DnaC from DnaB at *oriC*. Specifically, independent alanine substitutions for two conserved arginines in the box VII motif severely impaired DnaC in DNA replication by rendering the mutant proteins unresponsive to the signal that normally induces DnaC to dissociate. This signal appears to involve primer formation, defining a discrete event in the transition from initiation to the elongation phase of DNA replication.

DnaC also binds weakly to single-stranded DNA (ssDNA). This interaction, which is stimulated by ATP, has been measured by ultraviolet (UV)-crosslinking, or sensitive methods of fluorescence anisotropy, or surface plasmon resonance (SPR). If DnaC binds to the unwound region of *oriC*, this activity may help to load DnaB at *oriC*.

DnaB and the DnaB–DnaC Complex

DnaB is the replicative helicase for the *E. coli* chromosome, and for many plasmids and bacteriophage that use *E. coli* as a host. For DnaB to load at *oriC*, the helicase must be complexed with DnaC. However, DnaC bound to DnaB inhibits its helicase and ATPase activity, so DnaC must dissociate for DnaB to become active. Bound to ssDNA, the functional form of DnaB is considered to consist of six protomers assembled into a ring-like structure through which the ssDNA passes. Recent reports suggest how members of this helicase family unwind DNA. Structural studies reveal that each DnaB protomer has a RecA-like fold contained in its larger C-terminal domain and a smaller N-terminal domain. Via a domain near its N-terminus, DnaC binds to the larger domain of DnaB to form the DnaB–DnaC complex.

Loading DnaB Helicase

Following the unwinding of *oriC*, DnaA loads a single DnaB–DnaC complex onto each separated strand within *oriC* of a supercoiled plasmid to form an intermediate named the pre-priming complex. Loading appears to be a concerted event because replication forks move bidirectionally even when DnaB is at a subsaturating level. The loading process can also be measured with an ssDNA named M13 A-site that carries a DnaA box in a hairpin structure. DNA replication of this SSB-covered DNA requires the binding of DnaA to the hairpin, followed by loading the DnaB–DnaC complex. Primer formed by primase and DNA synthesis by DNA polymerase III holoenzyme converts the ssDNA to duplex form.

Based on the following evidence, two regions of DnaA appear to interact with DnaB to load the helicase. First, a monoclonal antibody that binds to an epitope within residues 111–148 inhibits both the binding of DnaA to DnaB, and DNA replication of the ssDNA and of an *oriC*-containing plasmid. Deletion analysis further refined this region to residues 130–148 near the N-terminus of domain III. The second interacting region of DnaA is within domain I. Mutant DnaAs carrying an E21A or F46A substitution are active in assays of DnaA function that do not also involve the DnaB–DnaC complex. In contrast, these mutants are essentially inactive in DNA replication of M13 A-site DNA, supporting the conclusion that the respective wild-type residues directly contact DnaB during helicase loading. In the nuclear magnetic resonance (NMR)

structure of domain 1 of *E. coli* DnaA, these residues lie in a solvent-exposed surface formed by the second and third α helices near the N-terminus.

Of interest, mutant DnaAs bearing W6A, or R281A substitutions are as active as DnaA⁺ in binding to and unwinding *oriC*, but are inactive in initiation at *oriC* due to a specific defect in self-oligomerization. In contrast, the mutants support DNA replication of M13 A-site ssDNA, suggesting that they are able to interact with DnaB. Apparently, DnaA oligomerization is necessary to load DnaB at *oriC* but not at the hairpin, at which only one or two DnaA molecules appear to bind in order to load a single DnaB helicase complexed to DnaC onto the ssDNA. Presumably, the DnaA oligomer at *oriC* forms a structure required to load both DnaB molecules.

Earlier studies of DnaB and DnaC have assumed that the functional form of the DnaB–DnaC complex contains six DnaC molecules per DnaB hexamer. Recent work suggests that the complex assembled at *oriC* contains three DnaC monomers per DnaB hexamer. These latter results are similar to those describing the initiation of DNA replication of bacteriophage λ , which encodes λ P protein that substitutes for DnaC. This viral DnaC analog forms a complex with DnaB, and even displaces DnaC from the DnaB–DnaC complex in the process of loading DnaB at the λ replication origin. Like the DnaB₆–DnaC₃ complex at *oriC*, the complex of λ P protein bound to DnaB contains three λ P monomers per DnaB hexamer. These similarities suggest that the helicase loading process at *oriC* and at the λ replication origin is comparable.

Other experiments showed that primer formation by primase, which interacts with the smaller domain of DnaB during primer synthesis, leads to the dissociation of DnaC bound to the larger C-terminal domain of DnaB. These results suggest that the interaction of primase with DnaB at *oriC* coupled with primer synthesis induce a conformational change in the helicase that causes DnaC to dissociate. Once DnaC has left, the helicase can begin to unwind the parental DNA.

Events after Initiation

Primer Formation and DNA Replication

Following the dissociation of DnaC from DnaB at *oriC*, one helicase molecule translocates clockwise on the circular bacterial chromosome while the other moves in the opposite direction. Each DnaB helicase encircles one strand of the duplex DNA and moves in the 5'-to-3' direction relative to this ssDNA as it unwinds the parental duplex. The occasional interaction of DnaB with primase leads to primer synthesis; several recent reports suggest how DnaB and primase coordinate their functions. A dimeric DNA polymerase III holoenzyme at each replication fork then extends the primers to duplicate the chromosome. Comparing bacteria with eukaryotes, it is remarkable that organisms in separate kingdoms utilize similar biochemical mechanisms of initiation of DNA replication.

Summary and Perspective

The molecular mechanism of initiation of DNA replication in *E. coli* can be separated into discrete steps. The major steps involve the recognition of the replication origin (*oriC*) by DnaA followed by the loading of DnaB helicase. For DnaB to

begin to unwind the parental duplex DNA, DnaC must dissociate from DnaB, which leads to the assembly of the replication fork machinery that is responsible for duplication of the bacterial genome. Fifty-eight years ago, Watson and Crick predicted that each parental DNA strand would serve as the template for the synthesis of the complementary strand. We now understand some of the molecular details of this process. Interested readers are invited to read other articles that review how DNA replication is regulated at the stage of initiation so that DNA replication occurs only once per cell cycle.

See also: **Molecular Biology:** DNA Helicases: Hexameric Enzyme Action; DNA Polymerase III, Bacterial; DNA Replication: Eukaryotic Origins and the Origin Recognition Complex; DNA Replication Fork, Bacterial.

Further Reading

- Bailey S, Eliason WK, and Steitz TA (2007) Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. *Science* 318: 459–463.
- Chodavarapu S, Felczak MM, Yaniv JR, and Kaguni JM (2008) *Escherichia coli* DnaA interacts with HU in initiation at the *E. coli* replication origin. *Molecular Microbiology* 67: 781–792.
- Chodavarapu S, Gomez R, Vicente M, and Kaguni JM (2008) *Escherichia coli* Dps interacts with DnaA protein to impede initiation: A model of adaptive mutation. *Molecular Microbiology* 67: 1331–1346.
- Enemark EJ and Joshua-Tor L (2008) On helicases and other motor proteins. *Current Opinion in Structural Biology* 18: 243–257.
- Erzberger JP, Mott ML, and Berger JM (2006) Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nature Structural and Molecular Biology* 13: 676–683.
- Hansen FG, Christensen BB, and Atlung T (1991) The initiator titration model: Computer simulation of chromosome and minichromosome control. *Research in Microbiology* 142: 161–167.
- Keyamura K, Abe Y, Higashi M, Ueda T, and Katayama T (2009) DiaA dynamics are coupled with changes in initial origin complexes leading to helicase loading. *Journal of Biological Chemistry* 284: 25038–25050.
- Leonard AC and Grimwade JE (2009) Initiating chromosome replication in *E. coli*: It makes sense to recycle. *Genes Development* 23: 1145–1150.
- Makowska-Grzyska M and Kaguni JM (2010) Primase directs the release of DnaC from DnaB. *Molecular Cell* 37: 90–101.
- McHenry CS (2003) Chromosomal replicases as asymmetric dimers: Studies of subunit arrangement and functional consequences. *Molecular Microbiology* 49: 1157–1165.
- Messer W (2002) The bacterial replication initiator DnaA. DnaA and *oriC*, the bacterial mode to initiate DNA replication. *FEMS Microbiology Reviews* 26: 355–374.
- Mott ML and Berger JM (2007) DNA replication initiation: Mechanisms and regulation in bacteria. *Nature Reviews Microbiology* 5: 343–354.
- Neuwald AF, Aravind L, Spouge JL, and Koonin EV (1999) AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Research* 9: 27–43.
- Nielsen O and Lobner-Olesen A (2008) Once in a lifetime: Strategies for preventing re-replication in prokaryotic and eukaryotic cells. *EMBO Report* 9: 151–156.
- Ozaki S and Katayama T (2009) DnaA structure, function, and dynamics in the initiation at the chromosomal origin. *Plasmid* 62: 71–82.
- Pomerantz RT and O'Donnell M (2007) Replisome mechanics: Insights into a twin DNA polymerase machine. *Trends in Microbiology* 15: 156–164.
- Remus D and Diffley JF (2009) Eukaryotic DNA replication control: Lock and load, then fire. *Current Opinion in Cell Biology* 21: 771–777.
- Stepankiw N, Kaidow A, Boye E, and Bates D (2009) The right half of the *Escherichia coli* replication origin is not essential for viability, but facilitates multi-forked replication. *Molecular Microbiology* 74: 467–479.
- Thomsen ND and Berger JM (2009) Running in reverse: The structural basis for translocation polarity in hexameric helicases. *Cell* 139: 523–534.
- Tougu K and Marians KJ (1996) The interaction between helicase and primase sets the replication fork clock. *Journal of Biological Chemistry* 271: 21398–21405.

DNA Replication, Mitochondrial

D A Clayton, Howard Hughes Medical Institute, Ashburn, VA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Closed-circular DNA DNA in which the double helix forms a continuous circular structure with no ends nor any interruptions in either strand.

DNA replication The process by which a given DNA molecule duplicates itself. This process occurs in different ways depending on the form of DNA and the enzymes and factors supporting replication.

Mitochondria The organelles in cells primarily responsible for energy production.

Mitochondrial DNA (mtDNA) The genome of mitochondria; every species has a characteristic mtDNA sequence.

Origin A DNA sequence region in which new DNA synthesis begins.

Promoter A DNA sequence recognized by enzymes and factors that copy genes into RNA molecules. Like an origin, a promoter marks the beginning of a nucleic acid synthesis event.

Transcription The overall process by which DNA sequence is copied into RNA sequence. This is the first major step in gene expression.

Mitochondrial Constituents and Their Functions

Mitochondria comprise an estimated 1500 different macromolecules, including mitochondrial DNA (mtDNA), mitochondrial RNA (mtRNA), and proteins.

mtDNA

Although certain cells can exist without mtDNA under special nutrient and growth conditions, where oxidative phosphorylation activity is not essential, in all other cases, mtDNA is needed. This is due to the fact that mtDNA encodes some of the proteins necessary for mitochondrial function in energy production. The sizes of mitochondrial genomes are quite broad. Vertebrate mtDNAs are typically 16–18 kb in size, with yeast mtDNA being several times larger. Plant mtDNAs can be 30-fold larger than vertebrate mtDNA. In addition, the number and identity of mtDNA genes are not constant across all mtDNAs. Human (and vertebrate) mtDNAs contain 37 genes, of which 13 are for proteins involved in bioenergetics. Although larger mtDNAs can contain more genes, the amount of coding information is not directly linear with genome size.

RNA

The RNA inside the mitochondrion is almost exclusively that which is encoded by mtDNA. For mammals, this generally includes two ribosomal RNAs (rRNAs) and 22 transfer RNAs (tRNAs). These 24 RNAs are thought to be necessary and sufficient to support fully the mitochondrial translation machinery, responsible for decoding the additional 13 mammalian mRNAs for mtDNA-encoded proteins.

Other RNAs, encoded by nuclear genes, include additional tRNAs, in cases where mtDNA contains an insufficient number of its own tRNA genes, and 5S rRNA. In addition, certain RNA-processing activities, consisting of enzymes with nucleus-encoded RNA components, have been implicated as participants in processing mtRNA sequences. These include an RNase P

for processing tRNA precursors and an RNase that cleaves mtRNA sequences at the origin of leading-strand DNA replication. This latter activity, RNase MRP, is thought to provide 3'-primer RNA ends for elongation by mtDNA polymerase.

Protein

The protein components of mitochondria are of several classes. Although mitochondria play key roles in apoptosis, calcium regulation, iron metabolism, and the synthesis of amino acids, sterols, and heme, the classic principal function of the organelle is energy production. It is, therefore, not surprising that all of the mtDNA-encoded proteins participate in the formation of one or another of the several bioenergetic complexes that support the chain of reactions leading to adenosine triphosphate (ATP) production. Although the vast majority of mitochondrial proteins are encoded by nuclear genes, those encoded by mtDNA are essential and there are no nuclear genes present that can substitute for a loss of mtDNA.

The nucleus-encoded proteins that comprise the bulk of mitochondrial proteins fall roughly into two classes. One class is devoted to the task of energy production and mitochondrial function, including maintenance and expression of mtDNA. The second population consists of structural elements that provide the infrastructure that maintains basic mitochondrial morphology, mobility, and organelle physiology. Most mitochondrial proteins, being nuclear-gene products, are produced by translation on cellular ribosomes and then targeted to mitochondria by their sequence (the most common feature directing the process is a signal sequence at one end of the protein, the amino terminus). Physical importation of the protein involves specific portions of the mitochondrion's inner and outer membranes that form ports for entry of these proteins. Initially, a general internalizing import mechanism applies to most proteins regardless of their final destination within the mitochondria. This is followed by other mechanisms leading to sorting of a protein according to its function.

Replication of the Mitochondrial Genome

Replication of mtDNA occurs within the confines of the mitochondrial-organelle network, which is usually organized dispersively throughout the cell's cytoplasm. As such, it represents a separately managed genome, physically and functionally distinct from nuclear DNA replication. The cellular copy number of mtDNA varies according to cell type. For example, mature lymphocytes have a few hundred mitochondrial genomes, whereas mammalian oocytes contain tens of thousands. Unlike nuclear DNA replication, which is subject to cell cycle control, it appears that mtDNA is able to initiate and complete a round of replication in dividing cells without restriction. Furthermore, mtDNA needs to be continually replenished in somatic cells due to organelle turnover. Such cells normally do not divide and, hence, do not replicate their nuclear DNA.

Mammalian mtDNA

The historic replication model posits that leading-strand replication of mammalian mtDNA begins at closely spaced, defined sites located downstream from a major transcription promoter and proceeds unidirectionally with displacement of the parental leading strand until approximately two-thirds of the closed circular mtDNA have been copied (Figure 1). As a consequence, the replication fork passes a major origin for lagging-strand synthesis, leaving it in single-stranded form. Displacement as a single strand is thought to allow the characteristic secondary structure of this origin to occur, thereby permitting initiation of lagging-strand synthesis. A natural consequence of the separate and distinct locations of the two origins is that the two segregated progeny mtDNA circles are of two types: one a duplex circle with a newly synthesized leading strand and the other a gapped circle with a partial newly synthesized lagging strand. In each case, the final steps of synthesis and closure result in the mature closed circular mtDNA products.

Convincing biochemical isolation and characterization of mammalian mtDNA began in the 1960s. The key technical breakthrough was the use of dye-salt buoyant density centrifugation to isolate even relatively rare closed-circular DNAs in highly purified form. This technique was exploited to perform mtDNA analyses by sophisticated analytical centrifugation techniques, by electron microscopy, and eventually by gel electrophoresis. It is fortuitous that mammalian mtDNA molecules are small enough to be easily manipulated and analyzed by a variety of techniques.

Unique to mammalian mtDNA and a few viral DNA genomes is a bias in the content of guanosine plus thymine in one strand of the double helix compared to the other strand. A consequence of this situation is that the two strands of mammalian mtDNA can be physically separated to an extent that allows their individual identification. Therefore, one can isotopically label mtDNA and directly assay the extent of mtDNA synthesis that has occurred in each strand.

Perhaps the most signature feature of mammalian mtDNA is the displacement loop (D loop). The D-loop nomenclature follows from the fact that closed-circular mtDNA contains a region of varying size, depending on the species, in which

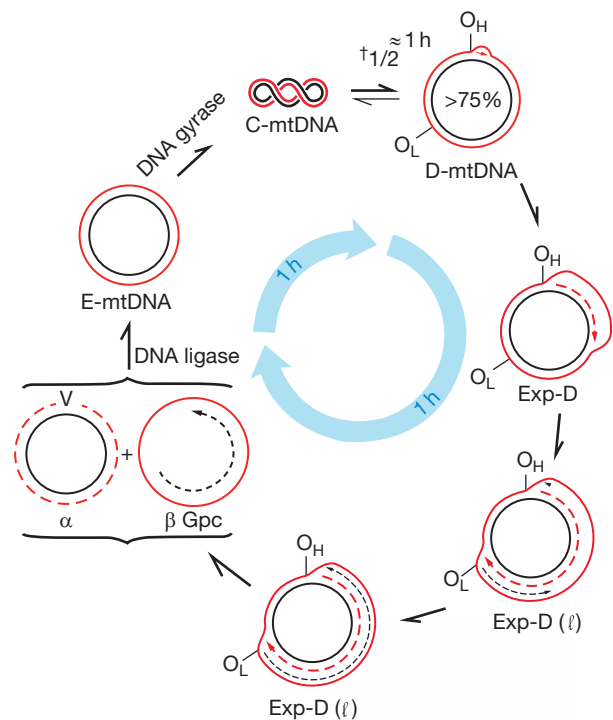


Figure 1 Diagram of mtDNA replicative intermediates. In the strand-displacement model closed-circular mtDNA (C-mtDNA) is in equilibrium with D-loop mtDNA (D-mtDNA). Productive replication is defined by elongation of the leading strand (red), which progresses around the genome (dashed red; Exp-D) until exposure of the major lagging-strand origin, which sponsors initiation of synthesis in the counterclockwise direction. End products are the new α - and β -circles, which are finished to closed circles (E-mtDNA). Superhelical turns are introduced (C-mtDNA) and finally a new D loop is synthesized. There is a rapid conversion back and forth between D-mtDNA and C-mtDNA. In mouse tissue culture cells the predominant form of mtDNA (>75%) is D-mtDNA.

leading-strand DNA synthesis has initiated and elongated to a size between 0.5 and 1 kb. Molecules exhibiting progressive growth of this nascent leading strand, approaching full genomic size, were identified and argued as support for the proposed model for replication.

Lagging-strand synthesis is revealed by the appearance of duplex DNA within the context of the ever-increasing displaced parental-single strand. A preferred initiation site is located about two-thirds of the genomic distance away from the D-loop origin of leading-strand synthesis. Recent studies have suggested that other additional sites can serve as points of lagging-strand DNA replication initiation and thereby provide a broader range of duplex replicative intermediates.

Alternative models for human mtDNA replication have been proposed mainly by inferring replicative intermediate structures based on migration patterns of mtDNA under two-dimensional gel electrophoresis. The currently postulated alternative model is one in which leading-strand synthesis is very similar to that depicted in Figure 1. The unusual feature of the model is that lagging-strand mtDNA synthesis is preceded by

extensive RNA synthesis which is somehow converted to DNA by processing or displacement of significant stretches of this RNA. Studies from other laboratories have also reported significant amounts of mtRNA sequence co-isolating with mtDNA. Some, but not all, of these RNAs have a suitable 3'-end for extension into a DNA strand by mtDNA polymerase.

Important insights required for a full elaboration of mtDNA replication mechanics will be aided by the development of robust *in vitro* replication systems using cloned and highly purified components. Indeed, recent *in vitro* work has verified, at the nucleotide level, prior studies over the last several decades that mapped the exact 5'-end positions of newly synthesized strands of mtDNA.

Future Studies

In contrast to studies on mtDNA, the mitochondrial system has been challenging with respect to identifying, purifying, and characterizing the critical proteins and other factors that sponsor mtDNA replication. A principal reason for this is that although low-abundance mammalian mtDNA can be successfully purified to a high state because of its closed circular physical form, in general, the mitochondrial proteins involved in mitochondrial nucleic acid syntheses can be similar, with regard to their purification properties, to their much more abundant nuclear counterparts. A further complication is that while most activities are unique to mitochondria, others are shared with the nucleus (e.g., RNase MRP and DNA topoisomerase III α). With the current availability of mammalian genomic sequences, and cloning and expression technologies, it should be possible to make significant advances in this area. A partial listing of known and predicted activities includes mtDNA and mtRNA polymerases, transcription factors, single-strand DNA-binding protein, specialized RNase activities, helicases, and topoisomerases.

It is now clear that mitochondrial function and the integrity of normal mtDNA are important for normal physiology in mammals. There are numerous reports of human diseases that have mtDNA mutations, and even loss of mtDNA, as the underlying basis for what is usually manifested as neuromuscular disorder phenotypes. It will be important to learn the nature and rate of mtDNA turnover in different cells and tissues. Unlike nuclear DNA, which is not replicating in non-dividing cells, mtDNA continues in its need to replicate due to continuous turnover and renewal of the mitochondrial organellar population. Thus, at least the opportunity to accumulate mutations in mtDNA due to replication errors during an individual's lifetime can be appreciated. Knowledge of the global pattern of mtDNA-turnover rates could prove valuable in predicting risk factors for human mitochondrial disease.

A full understanding of the role and dynamics of small D-loop formation at the leading-strand origin of replication remains to be determined (Figure 1). This unique, species-specific structure is aggressively synthesized and is immediately adjacent to the promoters of transcription. It is likely that both RNA processing and termination events occur in this region of the genome. The RNA processing activity that recognizes and cleaves mtRNA primer sequences contains an RNA encoded by a nuclear gene. Recent studies have begun to reveal the mechanism of RNA import.

The identification and structural determination of the mammalian mtDNA polymerase subunits that comprise the holoenzyme pave the way for the future important studies on elucidating the bases for how the many known cases of mutations in the enzyme can cause human pathological conditions.

There are now a number of studies that are identifying the key activities that regulate the fission and fusion of individual organelles. This will likely involve further consideration of how the mitochondrial network travels along microfilaments and is positioned within the cellular cytoplasm. In turn, it will be important to learn more about the precise location of intracellular mtDNA-replication sites to determine if replication is site specific or can occur generally.

Yeast and Others

Historically, the yeast system has been the most widely used to study almost all aspects of mitochondrial biogenesis. It has provided a wealth of information on bioenergetics, organelle assembly, protein import, and other areas of mitochondrial biology, including the details of mtDNA gene expression.

However, studies of mtDNA replication in yeast have been challenging due, in part, to the plasticity of the genome and the very aggressive rate of recombination of yeast mtDNA in cells. Thus, the isolation and analysis of productive replicative intermediates have proved much more difficult than for mammalian systems with their small genomes and almost complete absence of recombination, including the nucleases usually associated with breakage and rejoining of DNA strands. Nevertheless, there are a few intriguing similarities between yeast and mammalian mtDNAs with regard to potential yeast mtDNA origins resembling origin sequences in D loops. It will be interesting to learn whether origins of replication have been conserved over this period of evolution, or whether other roles for these sequences, such as attachment to membranes to facilitate segregation of mtDNA molecules after replication, are the driving force for these similarities. In this regard, recent studies have suggested that mtDNA can be isolated in close association with specific proteins, which may be key to understanding genomic placement in the organelle.

See also: [Bioenergetics: Bioenergetics and the Mitochondrial Genome](#); [Mitochondrial DNA](#); [Mitochondrial Genes and Their Expression](#); [Yeast](#); [Nuclear Genes Involved in Mitochondrial Function and Biogenesis](#).

Further Reading

- Attardi G (1986) The elucidation of the human mitochondrial genome: A historical perspective. *BioEssays* 5(1): 34–39.
- Clayton DA (1982) Replication of animal mitochondrial DNA. *Cell* 28: 693–705.
- Copeland WC (2008) Inherited mitochondrial diseases of DNA replication. *Annual Review of Medicine* 59: 131–146.
- Falkenberg M, Larsson N-G, and Gustafsson CM (2007) DNA replication and transcription in mammalian mitochondria. *Annual Review of Medicine* 76: 679–699.
- Fusté JM, Wanrooij S, Jemt E, et al. (2010) Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. *Molecular Cell* 37: 67–78.
- Graziewicz MA, Longley MJ, and Copeland WC (2006) DNA polymerase γ in mitochondrial replication and repair. *Chemical Reviews* 106: 383–405.
- Hoppins S, Lackner L, and Nunnari J (2007) The machines that divide and fuse mitochondria. *Annual Review of Medicine* 76: 751–780.

- Kaguni LS (2004) DNA polymerase γ , the mitochondrial replicase. *Annual Review of Medicine* 73: 293–320.
- Larsson N-G and Clayton DA (1995) Molecular genetic aspects of human mitochondrial disorders. *Annual Review of Medicine* 29: 151–178.
- Lee Y-S, Kennedy WD, and Yin YW (2009) Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations. *Cell* 139: 312–324.
- Scheffler IE (1999) *Mitochondria*. New York: Wiley-Liss.
- Shadel GS and Clayton DA (1997) Mitochondrial DNA maintenance in vertebrates. *Annual Review of Medicine* 66: 409–435.
- Spelbrink JN (2010) Functional organization of mammalian mitochondrial DNA in nucleoids: History, recent developments, and future challenges. *IUBMB Life* 62(1): 19–32.
- Weng G, Chen H-W, Oktay Y, et al. (2010) PNPASE regulates RNA import into mitochondria. *Cell* 142: 456–467.

DNA Restriction and Modification: Type II Enzymes

S E Halford, University of Bristol, Bristol, UK

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Darren M. Gowers and Stephen E. Halford, volume 1, pp. 772–777, © 2004, Elsevier Inc.

Glossary

Palindromic DNA A DNA in which the 5′–3′ sequence of one strand is the same as that of the complementary strand.

Recognition site A sequence of nucleotide bases recognized by a restriction enzyme.

Restriction digest The cleavage of DNA molecules into defined fragments by the action of a (Type II) restriction enzyme.

Restriction mapping Placing the series of DNA fragments from a restriction digest into the order in which they occur along the DNA.

Restriction–modification (RM) The bipartite bacterial defense system. Unmethylated phage DNA is hydrolyzed at specific sites by a restriction enzyme; host DNA containing the same sequences is guarded from destruction by methylation from a modification methyltransferase.

S-Adenosylmethionine (AdoMet) The cofactor for methyltransferases.

History

The ability of bacteria to defend themselves from viral infection was first reported by Luria and by Bertani in 1952 and 1953. The latter noted that some strains of *Escherichia coli* could restrict the growth of bacteriophage λ . This was accounted for a decade later by W. Arber, who proposed the existence of restriction–modification (RM) enzymes; the restriction enzymes were envisaged to cleave DNA in response to a particular sequence of bases, but only if the sites had not been protected by prior modification.

A landmark was attained in 1970, when H. O. Smith identified a novel restriction enzyme that cleaved DNA at a specific sequence. (The restriction enzymes identified previously were all Type I systems that cleave DNA at nonspecific sequences.) In collaboration with D. Nathans, he showed that this enzyme could be used to dissect DNA into discrete fragments. In 1978, Arber, Nathans, and Smith shared the Nobel prize in medicine. At that time, nearly 50 Type II restriction enzymes had been discovered. Today, that number stands near 4000. The early availability of restriction enzymes as off-the-shelf reagents heralded the age of recombinant DNA technology.

Classification

At least four distinct categories of RM systems have been identified from their subunit compositions and their modes of action: Types I, II, III, and IV. This article focuses on the Type II systems. In brief, most Type II systems consist of two proteins: a restriction endonuclease that, in the presence of Mg^{2+} , cleaves DNA at a particular sequence, its recognition site; and a modification methyltransferase that blocks restriction activity by transferring a methyl group from S-adenosyl methionine (AdoMet) to a particular base in the recognition site. Only the nucleases from Type II systems cleave DNA at specific positions and it is these, unlike the Type I and the Type III enzymes, which have the myriad applications in recombinant DNA technology.

The enzymes within the Type II category display a variety of reaction mechanisms and subunit compositions. Consequently, they can be categorized into various subtypes, for example, the Type IIS, the Type IIB, the Type IIE, and the Type IIF enzymes.

Discovery of New RM Systems

The value of Type II restriction enzymes for DNA manipulations has driven extensive searches for more enzymes of this sort. The standard method for identifying new restriction enzymes has been to screen bacterial lysates for endonucleases that cleave DNA into discrete fragments. Phage or adenoviral DNA are often used as test substrates. The recognition sites are then determined by sequencing methods. About one-half of the bacterial strains tested in this way are found to have an RM system, in some cases two or three such systems. But if the enzyme fails to cleave the test DNA for any reason, it will not be detected by this procedure. For example, if it was present at too low a level or if it required a special substrate or an unusual cofactor, it would escape detection.

In recent years, new RM systems have been identified by analyzing the genome sequences of bacteria. Such analyses show that almost all bacteria possess multiple candidates for RM genes, in some cases over 20 such candidates. Over one-third of the candidates are for Type II systems, with the remainder Types I, III, or IV. Subsequent studies showed that a high proportion of the candidates for the Type II RM systems did indeed encode active enzymes, though in some instances only the methyltransferase was active.

Organization of Genes

Modification and Restriction Genes

The Type II systems generally feature two separate genes: one for the restriction enzyme and one for the modification

enzyme. In all such systems examined to date, the gene for the endonuclease is adjacent to that for the methyltransferase, although the order of the genes varies between systems. In some cases, they are in head-to-tail orientation and in others head to head or tail to tail.

Modification genes from different Type II systems are often homologous to one another. In many cases, the methyltransferase genes encode 10 highly conserved blocks of amino acid sequence which are involved directly in catalysis and AdoMet binding. Another block of protein sequence, which is not conserved, is involved in DNA recognition.

The restriction genes in the Type II systems are never homologous to the modification gene from the same system, despite coding for proteins recognizing the same DNA sequence. Moreover, the restriction enzymes from different Type II systems generally have dissimilar amino acid sequences, apart from certain enzymes with the same recognition site. The only well-conserved feature is the motif Pro-Asp-X₁₀₋₂₀-(Asp/Glu)-X-Lys (where X is any amino acid), which is often found at the active site. The aspartate and glutamate residues form part of the binding sites for the Mg²⁺ ions.

There are, however, many Type II systems that deviate from the standard pattern of one modification gene and one restriction gene. For instance, some systems employ two modification enzymes, one for each strand of the DNA. Others have two restriction genes to encode an enzyme comprising two different subunits. In further cases, one gene encodes a single polypeptide with both modification and restriction activities.

Regulating RM Expression

Whenever a bacteria gains a new RM system, expression of the RM genes has to be stringently coordinated. The methyltransferase must be produced first, to protect all the recognition sites in the chromosome, before significant amounts of the endonuclease are made. To do this, some Type II RM systems, such as PvuII and BamHI, contain a gene upstream of the restriction gene that codes for a protein that regulates gene expression in this manner.

The Type II Restriction Endonucleases

Recognition Sequences

The recognition sequences for many – but by no means all – of the Type II restriction enzymes possess three attributes (Table 1). First, the sites are usually 4–8 bp long. Second, the sites are palindromic: they read the same in 5'–3' directions on both strands. Third, the sites are fully specified, in that every position in the sequence is occupied by a particular base. Yet there exist numerous exceptions.

For some enzymes, the recognition sequence is degenerate, in that certain positions can be occupied by alternate bases. For example, the recognition sequence for Cfr10I (Table 1) has either G or A at its 5'-end and either C or T at its 3'-end. Similarly, the central base pair in the target site for EcoRII can be either an A:T or a T:A. For some systems, the recognition sequence is discontinuous, for example, SfiI (Table 1). Such sites contain two blocks of specified sequence interrupted by a section of unspecified sequence but defined length that varies

from just 1 bp (e.g., HinfI at GANTC) to 9 bp (e.g., XcmI at CCA(N)₉TGG). In most cases, the specified base pairs in the degenerate and in the discontinuous sites are still symmetrical palindromes.

Some recognition sites are truly asymmetric. For example, the 5'–3' sequence in the top strand of the recognition site for BbvCI is not the same as the bottom strand (Table 1). Many enzymes that recognize asymmetric sequences cleave the DNA some distance away from the site, at fixed positions on one side of the site. These are called the Type IIS enzymes. One example is FokI at GGATG(9/13); it cleaves 9 bases downstream from its site on the top strand and 13 bases away on the bottom strand (Table 1). However, others cut just one strand, for example, Nt.BstNBI (Table 1). Conversely, certain enzymes that recognize asymmetric sequences cleave both upstream and downstream of the recognition site, for example, BcgI (Table 1). They cut four phosphodiester bonds in their reactions to liberate a small DNA fragment, 32 bp in the case of BcgI, which contains the intact recognition site. Such enzymes belong to the Type IIB category.

The ~4000 known enzymes have only ~290 different recognition sequences, so enzymes from different species often have the same recognition site. Those that cut at exactly the same position are called isoschizomers; for example, HhaI and CfoI both cut GCG↓C. Those that cut at different positions within the same sequence are called neoschizomers; HinfII, which cuts G↓CGC, is thus a neoschizomer of HhaI.

Protein Structures

The enzymes that recognize palindromic DNA sequences are generally dimers of identical subunits. They interact with their recognition sequences symmetrically so that all the contacts between one subunit of the protein and one-half of the recognition site are duplicated by the second subunit with the other half of the DNA. One active site in the dimer is positioned against the scissile phosphodiester bond in one strand of the DNA and likewise the second active site on the other strand.

The structures of numerous Type II endonucleases, and their complexes with DNA, have been determined by X-ray crystallography. As with most other proteins that act at specific DNA sequences, restriction enzymes recognize their target sequences primarily by multiple hydrogen bonds to the bases in the major groove of the DNA. However, the restriction enzymes do not usually employ the elements of protein structure that are common among DNA-binding proteins, such as the helix–turn–helix or the zinc finger. Instead, their DNA-recognition elements encompass a wide variety of nonstandard structures, including loops, sheets, and helices. These also differ considerably from one enzyme to the next. In some cases such as EcoRV, recognition of the specific sequence also depend on deformations to DNA structure.

Despite the overall dissimilarity in their amino acid sequences and their differences in their structural elements for DNA recognition, some Type II restriction enzymes have similar tertiary structures in terms of the overall fold of the polypeptide chains, for example, EcoRI has a similar fold to BamHI and likewise EcoRV to PvuII. Similar folds have since been found in many other proteins that act on DNA, including enzymes involved in replication, recombination, repair, and transposition.

Table 1 Recognition sites for Type II restriction endonucleases

Site	Enzyme	Genus	Recognition site
<i>Palindromic</i>			
<i>4 bp</i>			
5' Overhangs	Sau3AI	<i>Staphylococcus aureus</i>	5' ↓G-A-T-C 3' 3' C-T-A-G ↑ 5'
Blunt ends	AluI	<i>Arthrobacter luteus</i>	A-G ↓C-T T-C ↑G-A
3' Overhangs	HhaI	<i>Haemophilus haemolyticus</i>	G-C-G ↓C C ↑G-C-G
<i>6 bp</i>			
5' Overhangs	EcoRI	<i>Escherichia coli</i>	G ↓A-A-T-T-C C-T-T-A-A ↑G
Blunt ends	EcoRV	<i>Escherichia coli</i>	G-A-T ↓A-T-C C-T-A ↑T-A-G
3' Overhangs	PvuI	<i>Proteus vulgaris</i>	C-G-A-T ↓C-G G-C ↑T-A-G-C
<i>8 bp</i>			
5' Overhangs	NotI	<i>Nocardia otitidis</i>	G-C ↓G-G-C-C-G-C C-G-C-C-G-G ↑C-G
Blunt ends	PmeI	<i>Pseudomonas mendocina</i>	G-T-T-T ↓A-A-A-C C-A-A-A ↑T-T-T-G
3' Overhangs	SgfI	<i>Streptomyces griseoruber</i>	G-C-G-A-T ↓C-G-C C-G-C ↑T-A-G-C-G
<i>Degenerate</i>			
	Cfr10I	<i>Citrobacter freundii</i>	(G/A) ↓C-C-G-G-(C/T) (C/T)-G-G-C-C ↑(G/A)
	EcoRII	<i>Escherichia coli</i>	↓C-C-(A/T)-G-G G-G-(T/A)-C-C ↑
<i>Discontinuous</i>			
	SfiI	<i>Streptomyces fimbriatus</i>	G-G-C-C-N ₄ ↓N-G-G-C-C C-C-G-G-N ↑N ₄ -C-C-G-G
<i>Asymmetric</i>			
Cleaves within site	BbvCI	<i>Bacillus brevis</i>	C-C ↓T-C-A-G-C G-G-A-G-T ↑C-G
Cleaves outside site	FokI	<i>Flavobacterium okeanoicoites</i>	G-G-A-T-G-N ₉ ↓N-N-N-N C-C-T-A-C-N ₉ -N-N-N-N ↑
Cleaves one strand	Nt.BstNBI	<i>Bacillus stearothermophilus</i>	G-A-G-T-C-N-N-N-N ↓N C-T-C-A-G-N-N-N-N-N
Cleaves both sides	BcgI	<i>Bacillus coagulans</i>	↓N ₁₀ -C-G-A-N ₆ -T-G-C-N ₁₂ ↓ ↑N ₁₂ -G-C-T-N ₆ -A-C-G-N ₁₀ ↑

Cleavage of Specific DNA Sequences

Enzymes such as EcoRV or BbvCI find their recognition sites in long DNA molecules by first binding to the DNA anywhere along the chain and then translocating to the recognition site by intramolecular pathways. The transfer seems to occur primarily by multiple rounds of dissociation and re-association events within the same chain. However, one-dimensional diffusion along the chain may also play a role, particularly over short distances of DNA. Its role is primarily to scan the DNA over the 10–100 bp immediately adjacent to the site of each re-association event. Once at the recognition site, each active site in the dimeric enzyme catalyzes the hydrolysis of its target phosphodiester bond in independent reactions. Both reactions are usually complete within the lifetime of the DNA–protein complex, so the initial product released from the enzyme is the DNA cut in both strands. On the other hand, if the complex dissociates before both strands

are cut – for example, in reactions at high ionic strength – DNA cleaved in one strand is liberated.

With very few exceptions, the Type II endonucleases require Mg^{2+} for their reactions. Some other metal ions, such as Mn^{2+} or Co^{2+} , give low levels of activity but others, such as Ca^{2+} , give no activity at all. The Mg^{2+} ions play direct roles in the catalytic reaction, but the precise roles vary from enzyme to enzyme. Some, such as EcoRI, seem to have one metal ion at each active site, whereas EcoRV, BamHI, and many others use two ions. There exist, however, several Type II enzymes that display their maximal activities with Ca^{2+} rather than Mg^{2+} and one, BfiI, that has no requirement for any divalent metal ion.

The hydrolysis of phosphodiester bonds by Type II endonucleases yields termini with 5'-phosphate and 3'-hydroxyl groups. Some enzymes introduce staggered breaks into the DNA, leaving fragments of double-stranded DNA with single-strand extensions at either their 5'-ends (e.g., EcoRI in Table 1)

or their 3' end (e.g., PvuI in Table 1). Others cut both strands at equivalent positions, producing blunt-ended fragments (e.g., EcoRV in Table 1).

Interactions with Nonspecific DNA

The Type II restriction enzymes are extremely specific for their cognate sequences. They cleave their recognition sites over a million times more rapidly than DNA sequences that differ from the recognition site by just 1 bp. Surprisingly, this specificity in catalysis need not be accompanied by an equivalent specificity in DNA binding. In the absence of Mg^{2+} ions, many restriction enzymes bind to all DNA sequences with similar affinities. Their extraordinary specificities for DNA cleavage arise instead from the catalytic reaction, with only the cognate DNA giving rise to a catalytically competent enzyme- Mg^{2+} -DNA complex. For example, the complex of EcoRV with non-specific DNA has a low affinity for Mg^{2+} , whereas its complex with specific DNA has a high affinity for Mg^{2+} , so only the specific DNA is cleaved.

Certain reaction conditions can, however, enhance the activities of these enzymes at alternative sites. This is termed star activity. The star sites are generally sequences that differ from the recognition site by 1 bp. Sequences differing in two or more positions are normally resistant even under star conditions. The conditions that cause star activity include high pH, low ionic strength, the addition of organic solvents, and the replacement of Mg^{2+} with Mn^{2+} .

Subsets of Type II Restriction Enzymes

In recent years, many Type II enzymes have been found to have distinctive properties and have been classified into subsets, a fraction of which are noted below. A large number of these enzymes have to interact with two copies of their recognition site before cleaving DNA. The restriction enzymes that need two sites are in effect double-checking the DNA to ensure that they cleave DNA only at the correct sequence.

Type IIS

Type IIS enzymes recognize asymmetric sequences and cleave the DNA at fixed positions downstream of the recognition site, typically 1–20 bp away. FokI (Table 1) is an archetype of the Type IIS enzymes. It is a monomeric protein containing two domains connected by a flexible linker: a DNA recognition domain that makes all of the contacts to the recognition site and a catalytic domain that has an active site capable of cleaving one DNA strand. To cut both strands, two monomers of FokI have to associate to a dimer via their catalytic domains to create a unit with two active sites. The dimer is formed most readily when both monomers are bound to separate FokI sites in the same DNA molecule. FokI is thus more active on molecules with two sites than on DNA with one site. This is a common feature of the Type IIS enzymes. Most of them cleave DNA with two target sites more rapidly than DNA.

An asymmetric sequence cannot be recognized by two identical subunits making identical interactions with the DNA, with each cutting one strand. Instead, the Type IIS enzymes

have evolved various strategies to cut both strands adjacent to their asymmetric sites. The mode of action of FokI, transient dimerization, is only one example. Others include: two active sites within a single polypeptide; a single active site that moves from one strand to the other; the recruitment of a separate DNA-cleavage subunit; and a tetrameric protein with four active sites that cut both strands at two separate sites.

Type IIE

Another subset of Type II enzymes is the Type IIE group, typified by EcoRII and NaeI, which consists of dimers of identical subunits. The interface between the subunits of the NaeI endonuclease has two binding clefts for the recognition sequence rather than the single cleft in enzymes such as EcoRV and BamHI. One cleft has the catalytic functions for DNA cleavage, but these are inactive unless the other cleft has bound another copy of the recognition site. Interactions spanning two DNA sites occur more readily with sites in *cis* on the same molecule of DNA than with sites in *trans* on different molecules. The Type IIE enzymes thus cleave DNA with two recognition sites faster than DNA with one site, but only one of the two sites is cut per turnover.

Type IIF

SfiI, Cfr10I, NgoMIV, and several other restriction enzymes show further differences from the orthodox pattern, and this group is now known as the Type IIF enzymes. They function as tetramers of identical subunits, two of which interact with one copy of a (usually) palindromic recognition sequence while the other two interact with a second copy. To cleave DNA, both binding sites must be filled with cognate DNA. Hence, like the Type IIE enzymes, the Type IIF enzymes interact with two copies of the target sequence, but, in contrast to the IIE systems, the IIF enzymes cut both sites in both strands. They thus cleave four phosphodiester bonds per turnover to convert a DNA with two sites into the product cut at both sites. They display optimal activity with sites in *cis* – the concurrent binding of the enzyme to sites in the same molecule of DNA traps the intervening DNA in a loop. Nevertheless, they can also act, albeit less efficiently, with sites in *trans* by bridging separate molecules.

Type IIB

By cutting both strands of the DNA on both sides of their recognition sites, the Type IIB enzymes such as BcgI (Table 1) cut four phosphodiester bonds at each recognition site. Moreover, like many other restriction enzymes, the IIB enzymes are more active against DNA with two recognition sites than DNA with a single site. They thus have the potential to cleave eight phosphodiester bonds per turnover. The IIB enzymes also differ from the orthodox enzymes in subunit composition. For instance, BcgI contains two different subunits: one carries both the endonuclease and methyltransferase activities while the other recognizes the cognate DNA. Other Type IIB enzymes contain a single subunit with endonuclease, methyltransferase, and DNA recognition functions. Moreover,

some require AdoMet for not only the methyltransferase but also the endonuclease activity.

The Type II Modification Methyltransferases

DNA Modification

The methyltransferases of Type II RM systems are altogether different from the partner endonucleases; their only commonality is recognition of the same DNA sequence. The purpose of the modification enzyme is to protect recognition sites in the host DNA from the restriction enzyme. They achieve this by transferring a methyl group from AdoMet to a particular base in the recognition sequence. Some methylate the carbon at position 5 (C5) of the cytosine ring, some the exocyclic amino group at position 4 in cytosine, and some the exocyclic amino group at position 6 in adenine. Each turnover of a modification enzyme methylates the recognition sequence in one strand. This is sufficient to block the activity of the restriction endonuclease. Hence, following semi-conservative replication of fully methylated DNA, the hemi-methylated daughter strands are protected. The unmethylated strand must then be modified by the methyltransferase before the next round of replication. Surprisingly, the methyltransferases from certain Type IIS R–M systems target only one strand and how these protect both of the newly replicated daughter chromosomes has yet to be explained.

The Methyltransferase Enzymes

The methyltransferases from Type II RM systems are monomeric proteins. The same enzyme methylates both strands of palindromic sites, but two methyltransferase functions are needed at asymmetric sites: one for the top strand and another for the bottom strand, though both activities are sometimes present in one polypeptide. The crystal structures of these enzymes have revealed two domains, one on either side of a DNA-binding cleft. One domain contains the DNA-recognition functions, which are unique to each methyltransferase, and the other the binding site for AdoMet and the catalytic residues for the transfer reaction.

The crystal structure of the HhaI methyltransferase bound to its recognition sequence revealed a novel perturbation of the DNA. Remarkably, the target cytidine had been swiveled completely out of its stacked conformation in the DNA helix and into the active site of the enzyme, deep in the catalytic domain. Other DNA-processing enzymes also use this mechanism, for example, the uracil DNA glycosidases in DNA repair. Base flipping may be a ubiquitous device used by enzymes that require access to the nucleotide bases in DNA that are otherwise buried in the double helix.

All methyltransferases use AdoMet as the methyl donor. With the cytosine C5 methyltransferases, the carbon must first be activated. This is achieved by forming a covalent bond between the adjacent C6 position on the cytosine and the thiol group of a cysteine in the active site of the protein. The subsequent transfer of the methyl group to C5 expels the cysteine from C6, thus breaking the covalent link between

protein and DNA. With the cytosine N4 and the adenine N6 methyltransferases, the exocyclic amino group may function directly as a nucleophile to attack the transferable methyl group of AdoMet.

Restriction Enzymes in the Laboratory

Various factors have contributed to the adoption of Type II restriction enzymes as essential research tools. The foremost is their specificity in cleaving DNA at fixed positions at particular sequences; that is, their abilities to recognize simple DNA sequences within any DNA molecule and to cleave the DNA at those sites without cleaving other sites to any detectable extent. Consequently, a restriction digest yields a discrete set of defined fragments determined by the locations of the recognition sites in the DNA molecule. Moreover, the resultant fragments of duplex DNA may, depending on the restriction enzyme being used (Table 1), possess single-strand extensions that are mutually complementary. Such fragments can then be ligated to other DNA molecules with the same single-strand extensions.

The applications of restriction enzymes have led to the emergence of several companies for whom restriction enzymes are the primary product line. The availability of these enzymes permitted far-reaching technical advances, such as the construction of recombinant DNA molecules, restriction mapping, DNA sequencing, restriction-fragment-length polymorphisms (RFLPs) and DNA fingerprinting. These techniques have proved to be invaluable for gene isolation and analysis, for genetic diagnosis and for genealogical tracking.

One feature of Type II nucleases that concerns users is their star activity (noted previously). Under altered conditions, the barriers that prevent the enzymes from cleaving DNA at non-cognate sites are less severe than those under optimal conditions. There is no general remedy for this, other than avoiding the conditions that cause it: low ionic strength, high pH, organic solvents, and Mn^{2+} . Nevertheless, it is proving possible to develop mutationally altered restriction enzymes that display considerably less star activity than the native forms.

A further concern is the fact that many restriction enzymes cleave DNA only after interacting with two recognition sites. Such enzymes have low activities against DNA molecules with one target site, but they can be activated against single-site substrates by adding oligonucleotide duplexes that have the recognition sequence.

See also: [Molecular Biology: DNA Modification, Restriction Endonucleases: Type I Enzymes; DNA Restriction and Modification: Type III Enzymes.](#)

Further Reading

- Cheng X and Roberts RJ (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Research* 29: 3784–3795.
 Halford SE (2001) Hopping, jumping and looping by restriction enzymes. *Biochemical Society Transactions* 29: 363–373.
 Halford SE (2009) An end to 40 years of mistakes in DNA–protein association kinetics? *Biochemical Society Transactions* 37: 343–348.

- Halford SE and Marko JF (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Research* 32: 3040–3052.
- Halford SE, Welsh AJ, and Szczelkun MD (2004) Enzyme-mediated DNA looping. *Annual Review of Biophysics and Biomolecular Structure* 33: 1–24.
- Marshall JJT and Halford SE (2010) The Type IIB restriction endonucleases. *Biochemical Society Transactions* 38: 410–416.
- Mucke M, Kruger DH, and Reuter M (2003) Diversity of Type II restriction endonucleases that require two DNA recognition sites. *Nucleic Acids Research* 31: 6079–6084.
- Perona JJ (2002) Type II restriction endonucleases. *Methods* 28: 353–364.
- Pingoud A, Fuxreiter M, Pingoud V, and Wende W (2005) Type II restriction endonucleases: Structure and mechanism. *Cellular and Molecular Life Sciences* 62: 685–707.
- Roberts RJ (2005) How restriction enzymes became the workhorses of molecular biology. *Proceedings of the National Academy of Sciences of the United States of America* 102: 5905–5908.
- Roberts RJ, Belfort M, Bestor T, et al. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Research* 31: 1805–1812.
- Roberts RJ, Vincze T, Posfai J, and Macelis D (2010) REBASE – a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Research* 38: D234–D236.
- Tao T and Bourne JC (1991) Bourne and RM Blumenthal. A family of regulatory genes associated with type II restriction–modification systems. *Journal of Bacteriology* 173: 1367–1375.

Relevant Website

<http://rebase.neb.com> – The Restriction Enzyme Database.

DNA Restriction and Modification: Type III Enzymes

D N Rao and S Bheemanaik, Indian Institute of Science, Bangalore, India

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Desirazu N. Rao and S. Srivani, volume 1, pp. 778–781, © 2004, Elsevier Inc.

Glossary

DNA translocation Movement of DNA through a region of protein, an energy-dependent process.

Methylation Modification of target base within the recognition sequence by the transfer of a methyl group from AdoMet.

Phase variation The high-frequency, reversible ON–OFF switching of gene expression that leads to changes in the expression of surface antigens.

Restriction DNA cleavage by a restriction enzyme within or outside the recognition sequence.

Organization of Genes and Regulation of Restriction Activity

A tight control of the potentially lethal activity of the restriction enzyme is ensured for efficient establishment of type III restriction–modification (R–M) systems in a cell. Two independent post-transcriptional regulatory mechanisms control the expression of restriction activity:

1. Modification activity is expressed immediately after the *res* and *mod* genes enter the cell, whereas the expression of restriction activity is delayed until the complete protection of cellular DNA is achieved by methylation. The expression of Mod subunit regulates the amount of restriction subunit present in the cell, suggesting that the correct folding of Res into an active and stable conformation is promoted by its interaction with the Mod subunit.
2. *In vivo* restriction activity is modulated by a decrease in the efficiency of translation and by varying ribosomal fidelity. For instance, EcoP1I restriction is drastically decreased by mutation in *rpsL* gene that encodes ribosomal protein S12, probably due to reduced translation efficiency.

It has been unequivocally shown that type III restriction enzymes, R.EcoP1I, R.EcoP15I, and R.PstII, occur as R_2M_2 complexes, while the companion DNA methyltransferases (MTases) occur as a dimer (M_2). In the case of R.EcoP15I and R.EcoP1I, the tetrameric R_2M_2 complex is stable at nanomolar concentrations *in vitro*. However, R.PstII occurs as R_2M_2 complex at high enzyme concentrations, while various subassemblies exist at low enzyme concentrations. Disassembly of PstII by the loss of a Res subunit is able to control relative modification/restriction activities. While both the EcoP1I and EcoP15I R–M systems are freely transferable from cell to cell by phage infection, conjugation, and transformation, horizontal transfer of the type III restriction enzyme, StyLTI, results in cell death. It is, thus, possible that these differences may reflect alternative routes to prevent auto-restriction by type III restriction enzymes.

Cofactor Requirements

Among the type III restriction enzymes that have been characterized to some extent, R.EcoP15I, R.EcoP1I, R.HinfII, and,

more recently, R.PstII require adenosine 5'-(γ -thiotriphosphate) (ATP) for DNA cleavage and possess sequence-specific ATPase activity. Nonhydrolyzable analogs of ATP, such as adenosine (β,γ -imido)triphosphate (AMP-PNP) or ATP- γ S, cannot replace ATP in the reaction. Interestingly, R.PstII has more relaxed nucleotide specificity and can cut DNA with guanosine triphosphate (GTP) and cytosine triphosphate (CTP) but not uridine triphosphate (UTP). While AdoMet is an absolute requirement for DNA cleavage by type I restriction enzymes, the requirement of AdoMet for DNA cleavage by type III restriction enzymes has been questioned. DNA restriction by R.EcoP1I requires exogenous addition of AdoMet, while the closely related R.EcoP15I has bound AdoMet and, therefore, presumably does not require the addition of this cofactor for DNA cleavage. The presence of different monovalent cations in assay buffers seems to influence the AdoMet requirement. In the presence of sodium ions, R.EcoP1I showed an absolute AdoMet requirement for DNA cleavage, whereas this was not the case in buffers containing potassium ions.

In addition, in the absence of AdoMet, R.EcoP1I showed nonspecific endonuclease activity that was further enhanced by the presence of AdoHcy (product of methylation reaction). Under these conditions, the requirement for two inversely oriented recognition sites does not seem to be necessary. Similarly, sinefungin, analog of AdoMet, supports DNA cleavage by R.EcoP15I and interestingly, the two-site requirement is not a strict prerequisite. Another study reports that with purified R.EcoP15I sinefungin has no effect. AdoMet is postulated to have three roles as far as type III restriction enzymes are concerned: (1) as a methyl group donor for DNA methylation, (2) as an allosteric activator of DNA binding, and (3) as a specificity factor, preventing unregulated cleavage of DNA substrates. The reasons for the discrepant findings on cofactor requirements of type III restriction enzymes remain unclear. Varying buffer conditions and duration of dialysis after purification or the host strains used could influence enzyme characteristics. Further systematic studies are needed to clarify these differences.

Domain Organization in Restriction Subunit

Multiple sequence alignment of all known and putative Res subunits suggests a modular structure (Figure 1(b)).

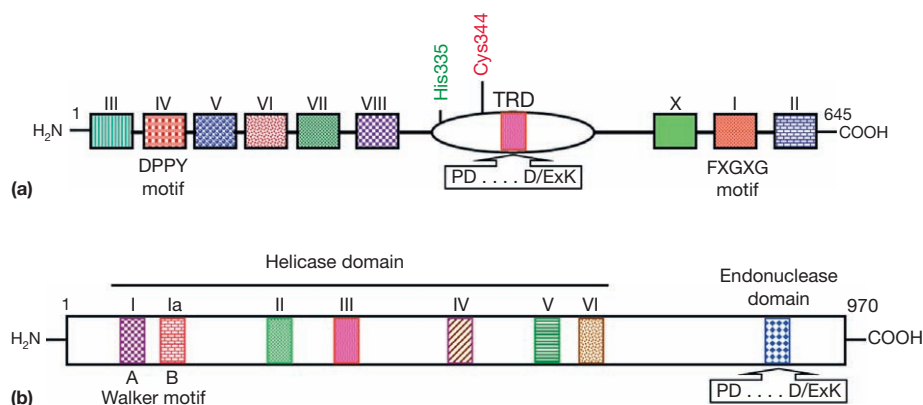


Figure 1 Modular structure of type III restriction enzyme. (a) Mod subunit showing the location of conserved amino acid sequence regions; catalytic motif, DPPY (motif IV) and AdoMet-binding motif, FXGXG (motif I). (b) Res subunit showing the location of conserved sequences, helicase motif, and putative endonuclease motif.

Table 1 Properties type III restriction enzymes

Enzyme	Source	Recognition sequence	Subunit composition	Cofactor requirement for restriction
EcoP15I	Plasmid P15B in <i>E. coli</i> 15T ⁻	5'-CAGCAG-3'	Res ₂ Mod ₂	Mg ²⁺ , ATP, AdoMet?
EcoP1I	Prophage P1	5'-AGACC-3'	Res ₂ Mod ₂	Mg ²⁺ , ATP, AdoMet?
StyLTI	<i>Salmonella typhimurium</i>	5'-CAGAG-3'	Res ₂ Mod ₂	Mg ²⁺ , ATP, AdoMet
HinfIII	<i>Haemophilus influenzae</i>	5'-CGAAT-3'	Res ₂ Mod ₂	Mg ²⁺ , ATP
PstII	<i>Providencia stuartii</i>	5'-CATCAG-3'	Res ₂ Mod ₂	Mg ²⁺ , ATP, GTP, CTP
LlaFI	<i>Lactococcus lactis</i>	ND	ND	Mg ²⁺ , ATP
BceSI	<i>Bacillus cereus</i>	ND	ND	Mg ²⁺ , ATP
PhaBI	<i>Pasteurella haemolytica</i>	ND	ND	ND
NgoAXP	<i>Neisseria gonorrhoeae</i>	5'-CCACC-3'	ND	ND

ND, not determined; A, site of methylation.

The C-terminus contains the PD(x)_n...(D/E)XK endonuclease motif that is commonly present in the catalytic center of restriction endonucleases. Sequence analysis of the Res subunit of EcoP1I and several putative Res subunits revealed the so-called DEAD box motif that is present in the helicase superfamily II. The members of the DEAD family of helicases have seven conserved motifs (motifs I, IA, and II–VI). The first motif of this family resembles the Walker A domain commonly present in ATPases. Mutational analysis of motif I resulted in a loss of DNA cleavage and ATP hydrolysis, while that of motif II significantly decreased ATP hydrolysis but had no effect on DNA cleavage. These motifs must, therefore, clearly play a role in ATP hydrolysis. Mutations in motif VI abolished both the DNA cleavage and ATPase activities, while mutations in putative endonuclease motif abolished DNA cleavage, but not ATP hydrolysis.

Domain Organization in Modification Subunit (DNA MTase)

The Mod subunit of type III restriction enzymes, EcoP1I, EcoP15I, HinfIII, and possibly other systems, alone can recognize and methylate the cognate sites. A unique feature of methylation by type III R–M systems is that only one strand of the asymmetric recognition sequence is methylated.

Sequence comparisons and biochemical studies in conjunction with site-directed mutagenesis have established the fact that the Mod subunit of the type III R–M systems contains the target recognition domain (TRD). TRD is a highly diverse and variable region in Mod subunit. So far, all the MTases belonging to this class of enzymes have been known to methylate adenine residues in their respective recognition sequences (Table 1).

Like all other N⁶-adenine MTases, the type III restriction enzymes contain two highly conserved sequences, FxGXG (motif I) and DPPY (motif IV), postulated to form the AdoMet-binding site and the catalytic site, respectively (Figure 1(a)). Extensive investigations on EcoP15I DNA MTase (M.EcoP15I) have revealed that this MTase adds a methyl group to the second adenine in the recognition sequence 5' CAGCAG 3' in the presence AdoMet and Mg²⁺. Mutations in motif I completely abrogated AdoMet binding but left the target DNA recognition unaltered. Mutations in motif IV resulted in the loss of enzyme activity but enhanced cross-linking of AdoMet and DNA, implying that DNA and AdoMet-binding sites are close to motif IV. Using chemical modification and site-directed mutagenesis, cysteine 344 and histidine 335 have been shown to play an essential role in the DNA methylation reaction catalyzed by M.EcoP15I. Both these amino acids are present in the putative TRD region. Curiously, M.EcoP15I possesses a PD(X)_n(D/E)XK like endonuclease motif (catalytic site of type II restriction endonucleases

that is shown to bind magnesium). However, the proline in this motif has been replaced by a methionine. M.EcoP15I requires magnesium ions for methylating DNA, and substitution of the acidic amino acid residues in the so-called 'endonuclease' motif abolishes magnesium binding and methylation activity. Intriguingly, substitution of methionine at position 357 by proline in the PDX_n(D/E)XK motif converted M.EcoP15I to a site-specific endonuclease. The cryptic endonuclease activity in EcoP15I DNA MTase perhaps provides an evolutionary link between restriction endonucleases and the MTase protein families.

M.EcoP15I binds about threefold more tightly to DNA containing its recognition sequence 5' CAGCAG 3' than to nonspecific sequences, in the presence of cofactors. In the presence of ATP, the discrimination between specific and nonspecific sequences increases significantly, suggesting a role for ATP in DNA recognition by these enzymes. M.EcoP15I makes contacts in the major groove of its substrate DNA. Both potassium permanganate footprinting and fluorescence spectroscopic measurements support a base-flipping mechanism for M.EcoP15I. EcoP1I DNA MTase (M.EcoP1I) methylates the second adenine in the recognition sequence 5' AGACC 3' in both single- and double-stranded DNA. The interesting property of binding and methylation of single-stranded DNA by M.EcoP1I might have an important role in the biology of single-stranded phages.

Mechanism of Action

Type III restriction enzymes show a strong preference for cleavage to occur when there are two copies of their asymmetric unmethylated recognition sites in an indirectly repeated, head-to-head orientation within a distance up to 3.5 kb. DNA cleavage occurs at one of the two sites, 25 or 26 nt downstream in the top strand and 27 or 28 nt downstream in the bottom strand. After cleavage, the enzyme remains bound to the still intact recognition site. DNA cleavage is blocked or reduced if (1) the sites are in the wrong orientation (tail to tail or head to tail), (2) there is only one site, or (3) a Lac-repressor molecule is bound to the DNA between the two head-to-head sites. Type III restriction enzymes must communicate between a pair of distant sites in order to cleave a DNA substrate. A specific protein-protein interaction is required, as two unrelated type III restriction enzymes cannot mutually activate cleavage of a linear DNA. It is proved that the type I enzymes extrude a DNA loop during translocation, and on the basis of the shared superfamily 2 helicase motifs a similar model has been suggested for the type III restriction enzymes.

The basis of long-distance communication for which a low-level ATP hydrolysis is required has not been accurately defined, and several models have been proposed, driven by either active DNA translocation or passive three-dimensional (3D) diffusion or invoking both DNA looping and limited translocation. A 'tracking-collision model' was first proposed to explain the DNA interaction and cleavage by type III restriction enzymes using R.EcoP15I as a model system (Figure 2(a)). This model states that one R.EcoP15I molecule bound to a recognition site produces a DNA loop of increasing size as it tracks along the DNA until it collides with another R.EcoP15I molecule bound to a second site tracking in the opposite

direction. ATP-hydrolysis-driven translocation thus positions two inversely oriented enzyme-site complexes for collision, which elicits conformational change of the enzyme molecules activating them for cleavage that occurs at a fixed distance away from one of the recognition sites. DNA loops originating from DNA translocation at two remote recognition sites within a DNA molecule were first observed by scanning force microscopy. To understand the mode of communication between distant recognition sites by type III restriction enzymes, atomic force microscopy (AFM) has been used to measure the DNA-protein interactions. These studies monitored the formation of translocated DNA loops in real time. The loop formation was both an ATP-dependent and -independent process. Based on the translocated DNA loops, a modified model was proposed in which R.EcoP15I initially employs diffusive looping to reduce the intersite DNA distance, followed by translocation-coupled extruded looping that brings the Res subunit into direct contact to cleave the DNA (Figure 2(a)). The low rates of ATP hydrolysis suggest that translocation-coupled looping cannot be the only mechanism, and that diffusive DNA looping must also act to reduce the lengths of DNA that the enzymes must translocate. The single-molecule resolution of fast-scan AFM further showed that the same R.EcoP15I-DNA complex is capable of both looping via extrusion coupled to translocation and diffusive looping. Thus, for two DNA-bound enzymes to communicate between two distal sites and trigger cleavage only when the sites are in a suitable orientation, enzyme dimerization, looping, and limited directional translocation driven by ATP hydrolysis are required. It thus appears that type III restriction enzymes have found the optimal strategy for their function by utilizing minimum energy for translocation prior to cleavage.

The DNA looping model based on AFM studies has been questioned. First, it was argued that it is not possible to conclude from the AFM studies that the events observed are necessary steps before DNA cleavage, because it does not measure both DNA communication and cleavage. Second, DNA attachment to a charged mica surface may have imposed a 2D geometry that can bias conformational flexibility. To monitor DNA looping in 3D environment, magnetic tweezers were used where both DNA cleavage and looping were measured simultaneously. These studies showed that type III restriction enzymes communicate between their sites in a 1D manner without loop formation. DNA cleavage rates were similar to those in bulk solutions and cleavage was not affected by magnetic force applied by magnetic tweezers. Based on these observations, an alternative DNA sliding mechanism was proposed in which restriction enzyme remains bound to DNA via weak electrostatic interactions (Figure 2(b)). Thermal activation then leads to a 1D random movement on DNA, in which both backward and forward steps occur with equal probability. The helicase domain produces a conformational switch from a DNA-recognition mode to a DNA sliding mode. The original binding orientation of enzyme to DNA is retained because enzyme is not released from DNA. If the enzyme eventually encounters a second enzyme molecule at another target site, cleavage of head-to-head DNA sites can occur. The model proposed for cleavage of linear DNA molecule explains that the enzyme moves in a direction and, upon encountering an end, dissociates from DNA, making cleavage less efficient.

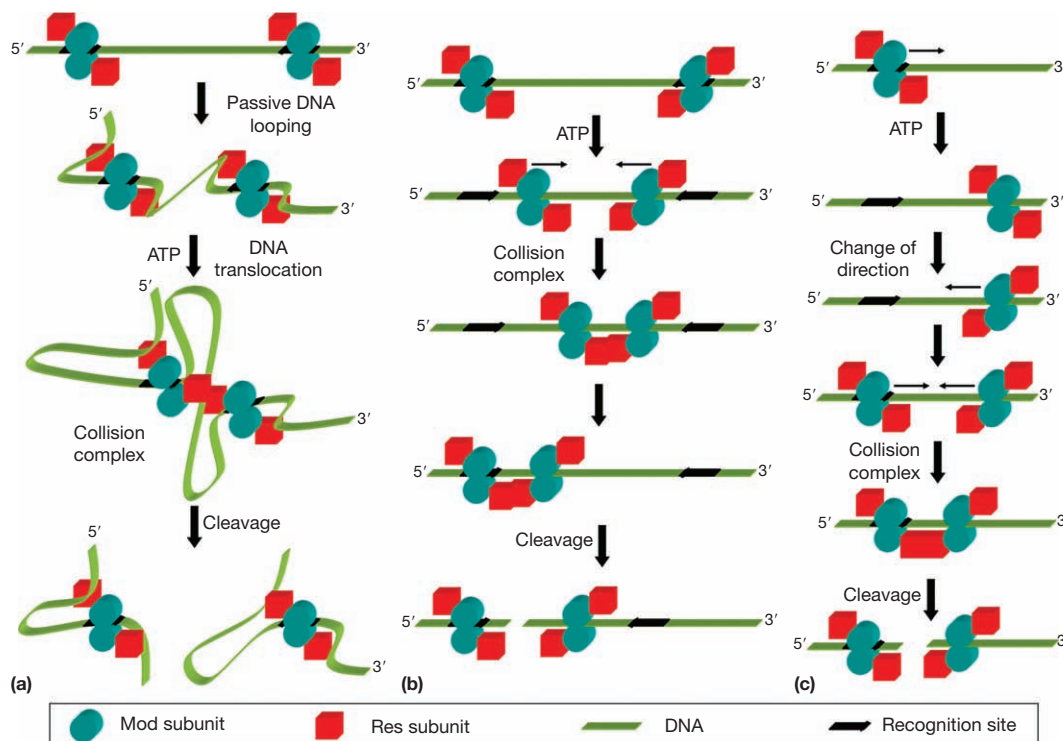


Figure 2 Schematic diagram showing the models proposed for DNA cleavage by type III restriction enzymes. (a) Tracking-collision model; (b) DNA sliding mechanism; (c) one-dimensional diffusion of enzyme along the DNA.

When the DNA ends are capped by a bulky moiety, the sliding protein will be backtracked, making cleavage as efficient as on circular DNA. The model for cleavage of linear DNA describes a dynamic binding, sliding, and release scheme, where not every sliding event would lead to cleavage. On DNA with two directly repeated (head-to-tail) sites, the enzymes load onto the DNA in the same direction and, therefore, collisions between two enzyme molecules cannot produce DNA cleavage, since the favorable interactions to cleave DNA never occur. However, on DNA with sites in tail-to-tail repeat, the sliding model predicts that cleavage could occur if the first enzyme slides past the second site before the second enzyme binds. The observations that type III restriction enzymes can cut DNA even when the head-to-head target sites are adjacent can be explained by sliding mechanism.

Yet another model has been proposed to explain the mechanism of DNA cleavage by type III restriction enzymes in which the enzyme follows 1D diffusion along the DNA (**Figure 2(c)**). Studies with R.EcoP15I and linear DNA substrates demonstrated that enzyme can cleave linear DNA substrates containing a single recognition sequence. The diffusion occurs on the DNA unidirectionally first in the 5'–3' direction from the recognition site, and then switches the direction upon reaching the 3' end of DNA. Such a switching can be stalled by binding roadblock protein such as Lac repressor at the 3' end of a linear DNA molecule but not at 5' end. The change in translocation direction is possible with R.EcoP15I either by activating ATP hydrolysis of a Res subunit or by making a 180° rotation at the DNA ends. The absence of effect of different DNA termini

suggests activation of the second Res subunit for reverse translocation. In either case, the enzyme molecule at 3' DNA end would, therefore, resemble an enzyme molecule translocating from the opposite site, whose interaction with another molecule translocating toward the 3' side of a single site results in the formation of a collision complex.

Phase Variation and Type III R–M Systems

Phase variation is an adaptive process by which pathogenic bacteria undergo frequent and reversible phenotypic changes in expression of surface antigens such as lipopolysaccharide and outer-membrane proteins resulting from genetic alterations in specific loci of their genomes. This helps in colonization of the host, adaptation to host environments, and evasion of immune responses. Phase variation in many bacterial species is mediated by frequent and reversible changes in the lengths of short DNA sequence repeats associated with a subset of specific genes. The gain or loss of repeat units in these homo- or heteropolymeric tracts is thought to involve a mechanism of slipped-strand mispairing that can occur during chromosomal replication or in the course of a variety of DNA repair and recombination processes that require DNA synthesis. While phase variation is typically associated with genes encoding surface antigens, several host-adapted bacterial pathogens have MTases (*mod* gene) associated with type III R–M systems, which contain single tandem repeats that are prone to phase variation (*Pasteurella haemolytica*, *Haemophilus influenzae*, and

Helicobacter pylori) or are predicted to have phase variable genes (*Neisseria gonorrhoeae* and *Neisseria meningitidis*).

The number of repeats within the *mod* gene influences the rate of phase variation and expression of the *mod* gene. Due to mutations within the *res* gene, restriction enzyme is not functional. Therefore, the evolution of type III R–M system into epigenetic modulators for controlling gene expression results in loss of the DNA restriction function. In *H. influenzae*, the random switching of the *modA* gene controls expression of a phase variable regulon of genes (a ‘phasevarion’) via differential methylation of the genome in the *modA* ON and OFF states. This was the first report of the coordinated random switching of a regulon of genes, and considering the wide distribution of phase variable type III R–M systems this phenomenon may represent a widely used mechanism in bacterial pathogens. *H. pylori*, *N. gonorrhoeae*, and *N. meningitidis* contain multiple phase variable *mod* genes, whereas *H. influenzae* has only a single *mod* gene. The striking feature of two type III R–M genes (*ngoAXP* and *ngoAORF641P*) is the presence of tetra- and pentanucleotide repeats at the 5′ end of their *mod* genes. Some studies with these genes showed that the changes in the tetra- and pentanucleotide repeats in the 5′ end of *mod* gene results in phase variation and may also control expression of different *N. gonorrhoeae* genes. The phylogenetic studies of *mod* genes of pathogenic *Neisseria* strains indicated that differences in the DNA recognition domain within the *mod* gene resulted in distinct *mod* alleles. Based on differences in the DNA recognition domain, several different alleles were found (*modA11*, *modA12*, *modA13*, *modB1*, and *modB2*), and this suggested the existence of multiple phase variations within the pathogenic *Neisseria*, each regulating a different set of genes. Since each strain has both *modA* and *modB*, these genes switch independently and there are thus four potential combinations of *mod* gene expression.

The presence of phase-variable type III R–M systems has led to various predictions about their role in the biology of bacterial pathogens. The role of R–M systems is to protect bacterial cells against foreign DNA. The presence of phase-variable type III R–M systems may result in the provisional removal of the protection barrier by altering methylation patterns and might allow lateral DNA transfer within or between bacterial populations. For organisms with phase-variable R–M systems, the implications of *mod* switching with respect to restriction function are unclear. If a phase-variable *mod* gene associated with an active type III R–M system switches from OFF to ON after several rounds of bacterial replication, the result may be auto-restriction and death of the bacterium, as unmethylated targets in the bacterial chromosome would be vulnerable to cleavage. However, in many phase-variable R–M systems *Res* has been found to be inactive. The phase variable type III R–M systems may epigenetically regulate genes through differential methylation of genome. Furthermore, a phase-variable MTase could be involved in pathogenicity by randomizing virulence factor expression through global changes in methylation. The list of phase-variable R–M systems is expanding and it is proposed that the phase-variable expression of the type III R–M systems has evolved owing to the selective advantage conferred by the phase variation, which enables random switching of an organism between two distinct phenotypes.

Serial analysis of gene expression (SAGE) is a method for comprehensive analysis of the cellular gene expression patterns. The key feature in SAGE is the generation of a short sequence tag containing sufficient information to uniquely identify a messenger RNA (mRNA) transcript. Quantitation of the number of times a particular tag is observed provides the expression level of the corresponding transcript. The distance of 25–27 bp between the R.EcoP15I recognition and cleavage sites in the DNA molecule is the longest defined distance known so far for any restriction endonuclease. In addition, the characteristic property of R.EcoP15I to cleave optionally at only one of two interacting DNA sites, yielding heterogeneous digestion products, makes it an ideal tagging enzyme in the SuperSAGE method of transcriptome analysis.

The cleavage of fluorescence-labeled DNA substrates by R.EcoP15I has been successfully used in determining the number of CAG triplet repeats present in the Huntington’s disease (HD) gene (30–180 CAG repeats on chromosomes of HD patients as compared to 6–37 repeats on chromosomes of unaffected individuals).

It is generally thought that types I and III restriction enzymes do not perform multiple rounds of catalysis unlike the type II restriction enzymes. No turnover of the enzyme could be measured with respect to their cleavage function. Therefore, it is interesting to determine whether one is justified calling these proteins as enzymes. Phage DNA after endonuclease restriction is further degraded by bacterial exonucleases. Type III restriction enzymes recognize specific asymmetric sequences and cut DNA 25–27 bp away from the recognition site. After cleavage, the enzyme remains bound to the DNA. It has been quite clearly shown with R.EcoP15I that the intact recognition site (5′-CAGCAG-3′) on the cleaved DNA sequesters the restriction enzyme, thus decreasing the effective concentration of the enzyme. When the restriction assay was performed in the presence of exonucleases, the removal of cleaved DNA by exonucleases released R.EcoP15I to perform further rounds of catalysis. It is, therefore, tempting to postulate that type III restriction enzymes and, possibly, other enzymes, which cleave sufficiently away from their recognition sites, depend on exonucleases to perform multiple rounds of catalysis and to act as efficient endonucleases. The importance of functional cooperation between transacting proteins in the modulation of enzyme activity cannot be, therefore, underscored.

See also: [Molecular Biology: DNA Methyltransferases: Eubacterial GATC](#); [DNA Modification, Restriction Endonucleases: Type I Enzymes](#); [Mammalian DNA Methyltransferase Structural Themes](#).

Further Reading

- Bourniquel AA and Bickle TA (2002) Complex restriction enzymes: NTP-driven molecular motors. *Biochimie* 84: 1047–1059.
- Crampton N, Roes S, Dryden DT, et al. (2007) DNA looping and translocation provide an optimal cleavage mechanism for the type III restriction enzymes. *EMBO Journal* 26: 3815–3825.
- Madhusoodanan UK and Rao DN (2010) Diversity of DNA methyltransferases that recognize asymmetric target sequences. *Critical Reviews in Biochemistry and Molecular Biology* 45: 125–145.

- Möncke-Buchner E, Rothenberg M, Reich S, et al. (2009) Functional characterization and modulation of the DNA cleavage efficiency of type III restriction endonuclease EcoP15I in its interaction with two sites in the DNA target. *Journal of Molecular Biology* 387: 1309–1319.
- Ramanathan SP, van Aelst K, Sears A, et al. (2009) Type III restriction enzymes communicate in 1D without looping between their target sites. *Proceedings of the National Academy of Sciences of the United States of America* 106: 1748–1753.
- Srikhanta YN, Fox KL, and Jennings MP (2010) The phasevarion: Phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. *Nature* 463: 196–206.
- Szczelkun MD, Friedhoff P, and Seidel R (2010) Maintaining a sense of direction during long-range communication on DNA. *Biochemical Society Transactions* 38: 404–409.

DNA Secondary Structure

A Bacolla and K M Vasquez, The University of Texas, Smithville, TX, USA

R D Wells, The Texas Medical Center, Houston, TX, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Albino Bacolla and Robert D. Wells, volume 1, pp. 782–787, © 2004, Elsevier Inc.

Glossary

Base pairs Bases associated by hydrogen bonds and hydrophobic forces.

Copy number variation (CNV) A segment of chromosomal DNA, usually 1 kb in length or more, which is present in varying number in the genome of different individuals.

DNA secondary structure A DNA conformation (non-B-DNA) different from the antiparallel right-handed double helix, B-DNA.

Slippage The annealing of an array of repeating DNA motifs on one strand of duplex DNA with the array on the complementary strand in an out-of-register manner.

Supercoiling The high-energy state generated by unwinding the helical duplex DNA when the two ends of the molecule are not free to rotate, as it occurs in the cell.

Translocation The process in which the end of a broken chromosome is joined with the end of another broken chromosome.

B-DNA

In 1980, the crystal structure of a helical turn of DNA was reported, which confirmed the model proposed by Watson and Crick in 1953. Since then, analyses of many sequences confirmed that the structure of DNA (B-DNA) is a right-handed double helix rotating about a central axis of symmetry and forming a major and a minor groove. The Watson and Crick base-pairing scheme between a purine and pyrimidine base (adenine (A) with thymine (T) and cytosine (C) with guanine (G)) was also confirmed. Deviations from these pairings (mismatches) perturb the helical structure of duplex DNA and can be corrected by cellular DNA repair mechanisms, including base excision repair (BER), nucleotide excision repair (NER), and mismatch repair (MMR).

Z-DNA

The first crystal structure of a DNA minihelix, reported in 1979, revealed a left-handed double helix formed by the self-complementary d(CG)₆ hexamer. It was called Z-DNA because of the zigzag shape of the phosphate and sugar backbone.

Chemistry

Z-DNA can form at alternating purine–pyrimidine (RY·RY) sequences (where R indicates a purine, A or G, and Y indicates a pyrimidine, C or T; the dot designates the complementary strands), such as the repeating (CG·CG)_n and (CA·TG)_n motifs (Figure 1). Structurally, the Z-DNA helix is more elongated than B-DNA; it lacks a major groove and the minor groove is very narrow and deep. The left-handedness is caused by the alternation of the N-glycosidic bonds between the *anti* and the *syn* conformations for the pyrimidines and purines, respectively. Both the B- and the Z-helices remain fully stacked since most of the helical distortion is dissipated over the two bases

flanking the Z-helix (B–Z junctions), which extrude into the solvent and become unpaired; these junctions are thought to be unstable and susceptible to oxidative damage, chemical modification, and cleavage *in vivo*. In the human genome, Z-DNA-forming tracts are often interrupted, for example, (CG·CG)₃A(CG·CG)₃. These interruptions disrupt the Z-DNA helix and create a Z–Z junction. Contrary to B–Z junctions, the bases forming a Z–Z junction are not extruded; however, they retain substantial conformational distortion that partially disrupts stacking within the overall Z-helix. Similar to other non-B-DNA structures, Z-DNA represents a high-energy state for DNA; *in vivo*, most energy required for the B to non-B transitions is thought to be supplied by negative supercoiling, a topological state of chromosomal DNA.

Biology

Estimates of ~300 000 alternating (RY·RY)_{≥6} motifs are present in the reference human genome sequence (see list of Relevant Websites), indicating that the potential sites for Z-DNA formation are higher than expected by chance alone. These sequences were probed with a peptide specific for Z-DNA for their existence in the left-handed conformation in the human A549 lung cancer cell line and ~7000 sites were found. Of these, 186 occurred at identical loci in multiple cells, suggesting that these sites may represent hot spots for Z-DNA formation *in vivo*. The most frequent hot-spot colocalization (46/186, 25%) occurred in centromeres, suggesting a putative role in kinetochore formation during mitosis or meiosis and the proper segregation of chromatids into daughter cells. Another 66 hot spots occurred within transcribed regions, mostly (49/66, 74%) in introns, supporting a role in gene transcription. No correlation was found between the presence of a Z-DNA hot spot and the level of gene transcription; thus, it remains to be determined whether Z-DNA activates or represses transcription on a genome-wide scale. Substantial evidence supports the conclusion that a Z-DNA tract is refractory to nucleosome assembly. Thus, Z-DNA formation at promoter or other regulatory sites is

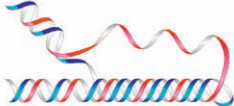
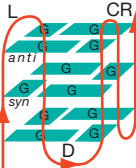
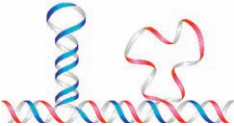
Name	Conformation	General seq. requirements	Example of sequence
Left-handed Z-DNA		$(YR \cdot YR)_n$	CGCGTGC GTGTG GCGCACGCACAC
Triplex		$(R \cdot Y)_n$ mirror repeats	<u>AAGAGG</u> _x <u>GGAGAA</u> <u>TTCTCC</u> _y <u>CCTCTT</u>
Quadruplex		Oligo (G) _n tracts	AG ₃ (T ₂ AG ₃) ₃ Single strand
Cruciform		Inverted repeats	<u>TCGGT</u> _x <u>ACCGA</u> <u>AGCCA</u> _y <u>TGGCT</u>
Slipped (hairpin) structure		Direct repeats	<u>TCGGT</u> <u>TCGGT</u> <u>AGCCA</u> <u>AGCCA</u>

Figure 1 Common types of DNA secondary structures. The DNA strands and their folding are depicted as ribbons. For quadruplex DNA, the idealized folding of single-stranded DNA into an intramolecular quadruplex is shown, highlighting the *anti* and *syn* orientation of G residues connected by lateral (*L*), diagonal (*D*) or chain reversal (*CH*) loops. For triplex DNA, one of four possible isomers is drawn, that is, the 3'-RRY isomer, in which the 3' half of the purine-rich strand of the R-Y mirror repeat engages in Hoogsteen hydrogen bonding with the 5' half of the R-Y mirror repeat. The other three isomers correspond to 5'-RRY, 3'-YRY, and 5'-YRY. For Z-DNA, the single-stranded B-Z junctions are indicated. *x* and *y*, ~1–5 of any nt combination separating inverted or mirror repeats (loops). Adapted from Zhao J, et al. (2010) Non-B DNA structure-induced genetic instability and evolution. *Cellular and Molecular Life Sciences* 67: 43–62.

expected to play a role in transcription by exposing free DNA to regulatory factors. Proteins that have been found to bind specifically to Z-DNA share a common motif, named Z α . In addition to the human RNA-editing enzyme ADAR1, the Z α domain is also found in the DNA-dependent activator of interferon regulatory factor (DAI, also called DLM-1 and ZBP1), PKR-like eIF2 α kinase (PKZ), which is functionally related to the dsRNA-dependent protein kinase PKR, and E3L, a vaccinia virus immediate early protein. These proteins are related by their activity in binding cytosolic double-stranded RNA (dsRNA) and double-stranded DNA (dsDNA), most likely as a means to elude (E3L) or trigger a cascade of signaling events in response to viral infection. Thus, the formation of Z-DNA is thought to play a role in infection-related processes.

Triplex DNA

The association of nucleic acids into three-stranded structures (triplex DNA or H-DNA) was reported in 1957 with synthetic

polymers and confirmed in 1968 using naturally occurring sequences (Figure 1).

Chemistry

In three-stranded nucleic acids, consecutive purine bases in a B-DNA duplex engage a third strand through Hoogsteen hydrogen bonding in the major groove of the underlying duplex molecule (Figure 2).

The formation of a triplex structure requires DNA to contain mostly purines on one strand and thus mostly pyrimidines on the complementary strand (R-Y), and that a mirror repeat symmetry is created. In intramolecular triplexes, one-half of the mirror repeat (the purine-rich or pyrimidine-rich strand) dissociates from the complementary strand and folds back to engage the recipient purine-rich strand of the remaining mirror repeat in the major groove. Hence, this type of structure may take place during replication or transcription, at a time when the complementary strands are separated and negative supercoiling is generated. Alternatively,

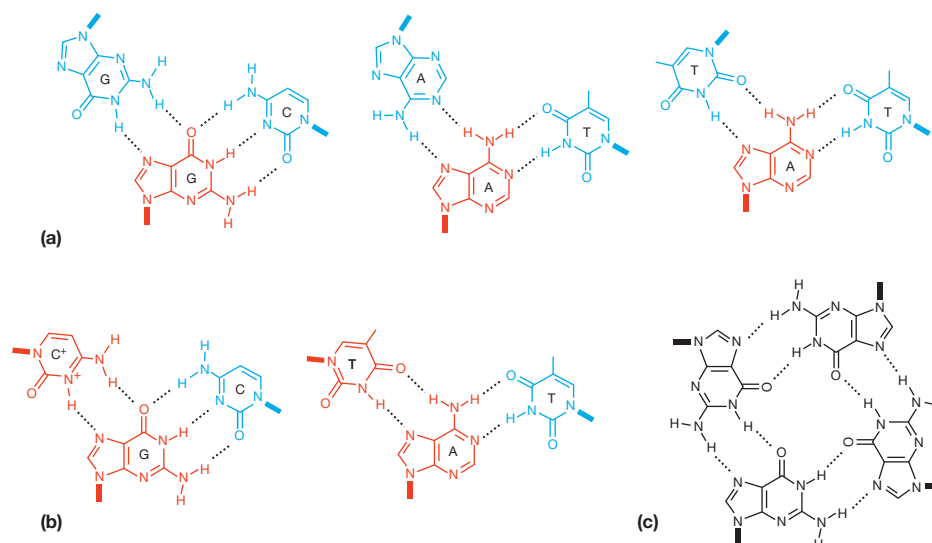


Figure 2 Hydrogen bonding schemes in triplex DNA and G-tetrads: (a) base triplets in the RRY-type of triplex DNA and (b) base triplets in the YRY-type of triplex DNA. Different colors signify antiparallel strand orientations. For each base triplet, the Hoogsteen–hydrogen bonded base is shown on the left. (c) G-tetrad; typical arrangement of G residues associated through Hoogsteen hydrogen bonds. Strand orientation and *syn/anti* conformations are not specified.

in an intermolecular triplex, an exogenous oligonucleotide (triplex-forming oligonucleotide or TFO) is added to the cell to bind a specific chromosomal (R-Y) duplex target sequence, also through Hoogsteen hydrogen bonding in the major groove. The use of TFOs is currently being explored as a therapeutic means to inhibit gene expression (oncogenes) in cancer or to directly inactivate genes by delivering a DNA damaging agent via the TFO. For any given intramolecular triplex, four isomers are possible: two in which the 3' half of the mirror repeat engages in Hoogsteen hydrogen bonding and two in which the 5' half of the mirror repeat engages in Hoogsteen hydrogen bonding. For each of these cases, either the pyrimidine-rich sequence engages as the third strand in a parallel orientation with respect to the recipient purine-rich strand or the purine-rich sequence engages as the third strand in an antiparallel orientation with respect to the recipient purine-rich strand.

Biology

Approximately 625 000 R-Y tracts with a mirror repeat symmetry formed by at least 10 bases are present in the reference human genome, a number of which (814) are embedded in very long R-Y tracts exceeding 250 nucleotides (nt) in length. These long tracts are enriched in both the PAR1 region, a segment of ~2.5 million DNA bases common to both the X- and Y-chromosomes and indispensable for meiotic recombination and chromosomal segregation and in genes expressed in the brain that function in the transmission of the nerve impulse at synapses. A large fraction of the genome is believed to undergo transcription on both strands of the DNA (sense and antisense transcription). Regions with long R-Y tracts tend to associate with poor antisense transcription and also with weaker sense transcription than genome-wide average, suggesting a genome-wide role for triplex DNA in gene function/regulation. In addition, endogenous

H-DNA-forming sequences that map to translocation break-points have been found to induce DNA double-strand breaks, leading to genomic instability. Such findings implicate these structure-forming sequences in cancer etiology.

Quadruplex DNA

The ability of G residues to assemble in tetrameric arrangements (tetraplex DNA, four-stranded DNA, or G4-DNA) was noted in 1962. In 1988, tetraplex DNA was reported in biological systems.

Chemistry

Quadruplex DNA can form at DNA sequences composed of four repeating motifs, each containing at least two G residues and separated by ~1–7 bases (spacers). The basic unit of quadruplex DNA is a G-tetrad, a planar array of four Gs, one from each of the four repeating motifs, stabilized by Hoogsteen hydrogen bonding and monovalent ($K^+ > Na^+$) cation coordination (Figures 1 and 2). G-tetrads stack on one another and are connected by the bases in the spacers, which form loops. The human telomeric sequence $(TTAGGG)_n$ has served as a prototype for structural characterization studies on quadruplex DNA. Various conformations have been reported, depending on buffer conditions and sequence lengths, in which the G-tetrad carrying strands may run either parallel or antiparallel, the Gs may adopt either the *syn* or *anti* conformations, and the loops provide connectivity in lateral, diagonal, or chain-reversal fashions.

Biology

The number of sites in the reference human genome containing predicted quadruplex-forming sequences exceeds 374 000.

As mentioned, chromosomes end with long stretches of $(TTAGGG)_n$ repeats, the last 100–200 of which are in the single-stranded form. In addition, several bioinformatic analyses have demonstrated a significant enrichment of quadruplex-forming motifs near (± 500 nt) transcription start sites and to transcription termination sites in numerous genes. Genome-wide probing of quadruplex DNA with a quadruplex-specific single-chain antibody further demonstrated that binding of the antibody to the expected DNA secondary structures caused

either an increase or a decrease in transcriptional activity in 1767 genes, supporting the conclusion that formation of quadruplex DNA plays a physiological role in gene expression regulation by (1) either inducing or repressing transcription, depending on the specific gene (Figure 3) and (2) favoring transcriptional termination. At the messenger RNA (mRNA) level, the localization of quadruplex DNA in 5-untranslated (5-UTR) regions has also been reported to elicit an inhibitory effect on translation.

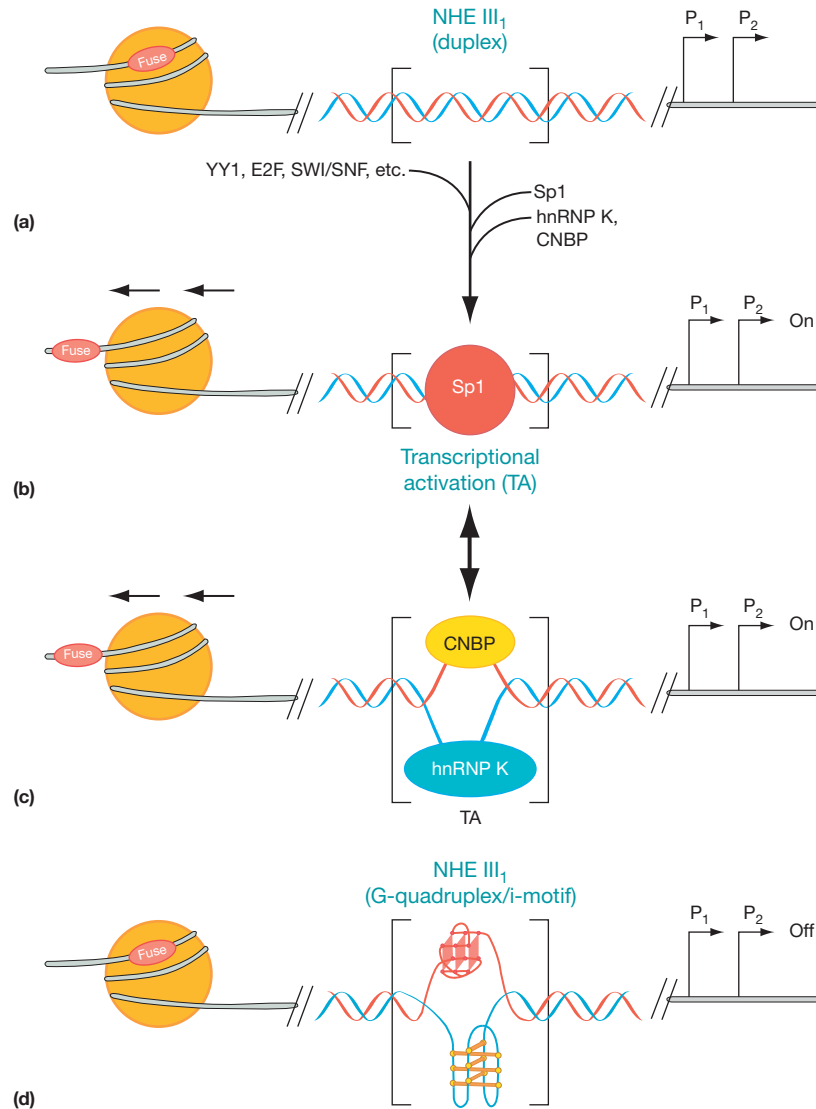


Figure 3 Quadruplex DNA and the regulation of *c-MYC* gene expression. (a) The promoter region of the *c-MYC* proto-oncogene includes four main promoters (P0-P3): two of which (P1 and P2) lie close to the main transcription start site, five nuclease hypersensitive elements (NHE), one of which (NHE III₁) contains a quadruplex-forming sequence (TGGGGAGGGTGGGGAGGGTGGGGAAGGTGGGGA), and a far upstream element (FUSE), which senses the torsional stress (negative supercoiling) accumulated in the DNA double-helix during the process of transcription. FBP (FUSE-binding protein) binds to single-stranded FUSE to activate and maintain *c-MYC* transcription. Alternatively, FIR (FBP-interacting repressor) binds FBP returning FUSE to double-stranded DNA and *c-MYC* transcription to basal levels. A number of binding sites for other transcriptional regulators are present in the promoter region. (b) Binding of Sp1 to double-stranded DNA within the NHEIII₁ region turns-on transcription. (c) Duplex DNA in the NHEIII₁ region is in equilibrium with single-stranded DNA; each strand binds a specific transcriptional activator (CNBP recognizes the G-rich strand whereas hnRNP K binds the complementary C-rich strand), yielding a complex that supports active transcription. (d) The single-stranded NHEIII₁ region folds into (a) quadruplex structure(s) on the G-rich strand and an i-motif on the C-rich strand, which prevents transcriptional activator binding and thus represses transcription. Stabilization of quadruplex DNA by intercalators, such as TMPyP4, stably represses transcription. Reprinted from Gonzalez V and Hurley LH (2010) The *c-MYC* NHE III₁: function and regulation. *Annual Review of Pharmacology and Toxicology* 50: 111–129.

Telomeres, that is, the protein-DNA complexes at the chromosomal ends, have also been probed for quadruplex formation. One of the proposed biological roles for quadruplex DNA at telomeres is related to the formation of T-loops, DNA loops where the extreme (2–3 repeats) 3′ single-stranded TTAGGG repeats fold back into duplex DNA with the assistance of proteins (such as TRF2) and engage with similar TTAGGG repeats to form quadruplex structures. The putative function of T-loops would be to provide protection for linear chromosomal ends against degradation.

Pharmacology

In the process of developing TFOs as tools for targeting specific genes, it was discovered that some G-rich oligonucleotides (GROs) could inhibit cancer cell viability without targeting the intended gene(s). Further research led to the finding that GROs assemble into stable secondary structures, including quadruplex DNA, before entering the cell, and that this folding induces their binding to nucleolin at the cell surface. The nucleolin–GRO complexes were then shown to migrate into the cell, leading to cell-cycle arrest, inactivation of the nuclear factor kappa B (NF- κ B) pathway, the induction of tumor suppressors, and induction of cell death, which ultimately resulted in an anti-proliferative effect for cancer cells. Nucleolin is present within the cell normally, but actively translocates to the plasma membrane in many cancer cell types. Thus, GRO activity appears to be selectively targeting certain cancer cells. The use of AS1411, a 26-mer GRO, is currently in phase II clinical trials to treat acute myeloid leukemia (AML) and renal cell carcinoma.

Cruciforms and Slipped Structures

Cruciforms and slipped structures are formed by inverted and direct repeats, respectively, under the influence of negative supercoiling. Contrary to the conformations described above, cruciforms and slipped structures do not require specific base composition.

Chemistry

In inverted repeats, a DNA sequence is followed by its complementary sequence, usually separated by a loop, on the same strand of DNA. Thus, each complementary strand may fold onto itself rather than binding to its complement, thus forming two fully paired intramolecular Watson–Crick duplexes, each with a single-stranded loop, that is, a cruciform structure (Figure 1). Fluorescence resonance energy transfer (FRET) studies indicate that in the presence of magnesium ions, a key cellular metal, cruciforms adopt an x-shape structure with pairwise antiparallel coaxial stacking of the neighboring helices. Two different conformations exist in equilibrium in solution, in which helices exchange partners on timescales of ~ 100 ms. Interestingly, no parallel stacked conformations have been detected at time resolutions of 1 ms; hence, cruciform arms extruding from chromosomal DNA are expected to lie in close proximity to the continuous and adjacent double helix. By contrast, tandem repeats may give rise to

bases looping out from duplex DNA (slipped DNA) when out-of-register re-annealing, such as during transcription, occurs.

Biology

Approximately 200 000 short (>10 bases) inverted repeat sequences separated by small loops are present in the reference human genome sequence. In addition, 96 large inverted repeats (arm size >8 kb with $>95\%$ sequence similarity), including nine inverted repeats with arm length exceeding 100 kb and displaying near identical ($>97\%$) sequence composition, have been identified. Five of these are present on the male-specific region of the Y-chromosome (MSY) and contain gene/gene classes that display testis-restricted expression with a function in sperm production and maturation. The palindromic architecture of the MSY gene families (amplicons) has been recognized as a critical feature for the genetic stability and gene function associated with the Y-chromosome. Indeed, the symmetric position of genes along the palindromic arms is believed to facilitate gene conversion events, which have been shown to occur frequently at these locations. For the Y chromosome, such high rates of gene conversion are believed to represent an efficient means for counteracting the threat of gene decay, which would otherwise be expected in the absence of recombination with a homologous chromosome. Gene conversion events may take place in various ways, including the formation of large cruciform structures.

Simple tandem repeat sequences are also common in the human genome reference sequence and display high rates of length variation among individuals, a feature attributed in part to the propensity of the repeat units to slip relative to one another during transcription and replication. Particularly important from physiological and pathological standpoints is the genome-wide distribution of trinucleotide repeats in exons, since their length polymorphisms in the general population also encode polymorphic homo-amino acid runs in the translated protein products, potentially altering their function. Several bioinformatic analyses concur with the conclusion that coding trinucleotide repeats are significantly overrepresented in specific classes of genes, such as those encoding transcription factors and proteins involved in transcription and its regulatory circuitries. The current view is that homo-amino acid runs contribute to protein intrinsic disorder, that is, unstructured domains in monomeric proteins that are critical for function. Indeed, intrinsically disordered protein domains are recognized as playing important roles in protein–protein and protein–nucleic acids interactions, whereby their flexibility to adopt various conformations enable high-affinity interactions upon binding with structurally diverse partners. Therefore, the ability of trinucleotide repeats to expand and contract within relatively short evolutionary times is believed to have been exploited as a means to explore and/or fine-tune a number of diverse protein–protein and protein–nucleic acids interactions during the course of evolutionary time.

Pathology

More than 30 human genetic disorders, mostly affecting the central nervous system and neuromuscular functions, have

been associated with the expansion of trinucleotide repeats and other microsatellites, in coding exons, untranslated regions, and introns of genes. In general, up to ~35–50 repeat units are present at any individual locus in normal alleles; however, large expansions of up to thousands of repeat units have been observed in affected individuals. Most of the expansion process is thought to take place early during embryonic development and/or gametogenesis, with further repeat instability continuing later during the lifetime of somatic tissues. Although the precise mechanisms leading to trinucleotide expansion remain a subject of intense investigation, the formation of aberrant DNA structures (including hairpins, stem-loops, and other intermediate structures) during replication and/or transcription is believed to play pivotal roles.

In addition to trinucleotide repeats, evidence is accumulating to suggest that other DNA repetitive elements are also involved in inducing genomic instability leading to human disease. Indeed, extensive work has been conducted on the characterization and the physical properties of DNA sequences involved in specific chromosomal translocations, such as the recurrent t(11;22) associated with the Emanuel syndrome (Figure 4). This composite work strongly suggests that large (~300–500 nt) cruciform structures may form on chromosomes 11 and 22, which are then cleaved and joined between the two chromosomes, resulting in a translocation.

A number of bioinformatic analyses aimed at analyzing the types of DNA sequences found at sites of pathological gene conversion events, gross chromosomal rearrangements and copy number variation (CNV), support the conclusion that repetitive motifs are found more often than expected by chance. These analyses indicate that DNA secondary structures, rather than the primary DNA sequences *per se*, facilitate DNA transactions leading to changes in genomic architecture that

result in disease. Analogous genomic changes, such as those associated with CNV, are also likely to contribute to phenotypic differences among individuals.

Molecular Biology

A number of studies conducted on trinucleotide repeats (such as CTG·CAG) *in vitro* and *in vivo* support the notion that the MMR proteins Msh2 and Msh3, components of the MutS β complex, are required for the expansion processes leading to their pathological expansion. Looped-out structures engineered *in vitro*, which are thought to simulate the conformations formed by the repeats *in vivo*, are indeed recognized and cleaved by MMR proteins. Additional steps involved in the expansion process appear to include interactions of the repeats (or most likely their underlying secondary structures) with incoming replication forks.

Other DNA secondary structures, such as triplexes, quadruplexes, and 4-way junctions (cruciform-like), have been shown *in vitro* to serve as specific targets for a number of DNA and RNA helicases, which resolve their folded conformations using nucleoside triphosphate (NTP) hydrolysis. Among these, WRN, BLM, and RECQL4 belong to the evolutionarily conserved family of recQ-helicases, and inactivating mutations in any of the corresponding genes cause the Werner, Bloom's, or Rothmund-Thomson syndromes, respectively, conditions characterized by widespread genetic instability and predisposition to cancer. These results prompt speculation that the maintenance of genome integrity is in part elicited by numerous helicase activities on DNA secondary structures, whose resolution alleviates their potential interference with key cellular transactions, such as DNA replication and transcription.

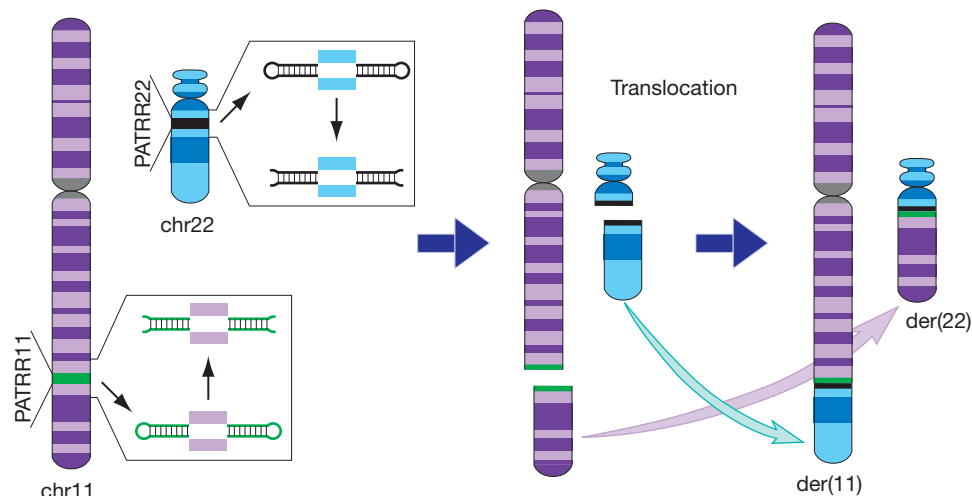


Figure 4 Cruciform DNA and the Emanuel syndrome. The human t(11;22) translocation is the most commonly recurrent non-Robertsonian translocation. In most patients, the breakpoints on chromosomes 11 and 22 cluster at the center of palindromic A/T-rich DNA (PATRR) regions, which comprise long (~350–500 nt) A/T-rich sequences with inverted repeat symmetry. The PATRR11 and PATRR22 do not share sequences homology; however, they are thought to extrude large cruciform structures (insets) that are then cleaved at their center loops (the breakpoint cluster region) and joined aberrantly, yielding the der(11) and der(22) chromosomes. The Emanuel syndrome is caused by the inheritance of a supernumerary der(22) chromosome in the offspring. Reprinted from Bacolla A, et al. (2010) Non-B DNA structure and mutations causing human genetic disease. *Encyclopedia of Life Sciences*.

Acknowledgments

This work was supported by NIH/NCI Grant CA093729 to K.M.V and with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. HHSN261200800001E to A.B. The content of this publication does not necessarily reflect the views of policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

See also: **Bioenergetics:** Friedreich's Ataxia; **Molecular Biology:** DNA Base Excision Repair Pathways; DNA Damage: Alkylation; DNA Mismatch Repair and the DNA Damage Response; DNA Mismatch Repair in Disease and Aging; DNA Supercoiling; Nonhexameric SF1 DNA Helicases/Translocases; Nonhomologous End Joining in Eukaryotes; Nucleotide Excision Repair: Biology; Nucleotide Excision Repair in Eukaryotes; Telomeres: Maintenance and Replication

Further Reading

- Bates PJ, Laber DA, Miller DM, Thomas SD, and Trent JO (2009) Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer. *Experimental and Molecular Pathology* 86: 151–164.
- Castel AL, Cleary JD, and Pearson CE (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nature Reviews Molecular Cell Biology* 11: 165–170.

- Conrad DF, Pinto D, Redon R, et al. (2010) Origins and functional impact of copy number variation in the human genome. *Nature* 464: 704–712.
- Gonzalez V and Hurley LH (2010) The c-MYC NHE III₁: Function and regulation. *Annual Review of Pharmacology and Toxicology* 50: 111–129.
- Kurahashi H, Inagaki H, Ohye T, et al. (2010) The constitutional t(11;22): Implications for a novel mechanism responsible for gross chromosomal rearrangements. *Clinical Genetics* 78: 299–309.
- Li H, Xiao J, Li J, et al. (2009) Human genomic Z-DNA segments probed by the Z α domain of ADAR1. *Nucleic Acids Research* 37: 2737–2746.
- Neidle S (ed.) (1999) *Oxford Handbook of Nucleic Acid Structure*. Oxford: Oxford University.
- Perry GH, Ben-Dor A, Tsalenko A, et al. (2008) The fine-scale and complex architecture of human copy-number variation. *American Journal of Human Genetics* 82: 685–695.
- Voineagu I, Freudenreich CH, and Mirkin SM (2009) Checkpoint responses to unusual structures formed by DNA repeats. *Molecular Carcinogenesis* 48: 309–318.
- Wang G and Vasquez KM (2009) Models for chromosomal replication-independent non-B DNA structure-induced genetic instability. *Molecular Carcinogenesis* 48: 286–298.
- Wells RD (2009) Discovery of the role of non-B DNA structures in mutagenesis and human genomic disorders. *Journal of Biological Chemistry* 284: 8997–9009.
- Zhao J, Bacolla A, Wang G, and Vasquez KM (2010) Non-B DNA structure-induced genetic instability and evolution. *Cellular and Molecular Life Sciences* 67: 43–62.

Relevant Websites

- <http://nonb.abcc.ncicrf.gov> – non-B DB: A Database for Integrating Annotations and Analysis of non-B DNA Forming Motifs.
- <http://www.ncbi.nlm.nih.gov> – OMIM, Online Mendelian Inheritance in Man.

DNA Sequence Recognition by Proteins

K Rutherford and G D Van Duyne, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Arabela A. Grigorescu and John M. Rosenberg, volume 1, pp. 788–793, © 2004, Elsevier Inc.

Glossary

Motif A discrete structural element (a mixture of α -helices, β -strands and/or loop) of a protein with a specialized function, such as interacting with DNA. For example, the helix-turn-helix motifs of $\gamma\delta$ -resolvase position one helix so that it can optimally insert into the major groove of the target DNA.

Sequence specificity Interactions that are dependent upon the protein recognizing, and forming contacts with, a sequence-specific pattern of donor and acceptor substituents within the DNA major and minor grooves. DNA-binding proteins generally display many levels of specificity, taking advantage of the overall shape of the DNA grooves as well as the nucleotide sequence.

Introduction

A protein's ability to recognize a specific deoxyribonucleic acid (DNA) sequence is essential for many biological processes, including initiation of replication and transcription and regulation of gene expression and development. Understanding the principles that govern these recognition events is therefore very important. In general, proteins that bind DNA use structural motifs that provide complementarity to the unique shape and electrostatic properties of the DNA double helix. Specificity arises partly through recognition of the pattern of functional groups presented by nucleotide bases in the DNA helix and partly through recognition of changes in helical structure caused by a given sequence. Although the number of specific protein–DNA interactions formed in a typical cell is immense, there are a relatively small number of structural motifs that are used by DNA-binding proteins. Proteins accomplish DNA recognition through a set of recurring structural themes that have been individually fine-tuned by evolution to bind to specific sequences. A number of DNA-binding proteins contain multiple repeats of these motifs, greatly increasing the binding affinity and specificity that can be achieved. This article summarizes the structural and chemical properties of proteins and DNA that contribute to sequence-specific recognition.

DNA Structure and Properties

The structural and chemical properties of DNA play an important role in defining protein–DNA interactions. The canonical DNA fold is that of a double-stranded helix with a core of stacked base pairs sandwiched between two highly electronegative phosphodiester backbone chains (**Figure 1(a)**). DNA base stacking and hydrogen bonding twist the helix, creating wide major (M) and narrow minor (m) grooves that form a repeating pattern parallel to the sugar-phosphate backbone (**Figure 1(a)**).

Major and Minor Grooves

The major groove is the primary site of sequence-specific protein–DNA interactions. The width of the major groove (~ 12 Å) matches well with the width of a protein α -helix, allowing for a highly complementary fit between the two elements. In addition, the major groove is rich in accessible polar and nonpolar groups presented along the base pair edges (**Figure 1**). These groups provide an important docking surface for protein recognition and binding. While the major and minor grooves of DNA share a similar depth, the minor groove is more narrow (~ 6 Å). Because of this difference, some distortion of the minor groove structure is usually necessary to accommodate the docking and insertion of protein structural elements. Although the DNA minor groove also contains functional groups that provide an interaction surface for proteins, there is less information about the actual sequence than is found in the major groove. As shown in **Figure 1(b)**, for example, C and T bases appear identical in the minor groove but can be distinguished in the major groove.

Unique Properties of DNA Bases

Unique DNA sequences are defined by the pairing of four nucleic acid bases: adenine (A), thymine (T), guanine (G), and cytosine (C). A–T base pairs are coordinated by two hydrogen bonds, while G–C base pairs are linked by three hydrogen bonds (**Figure 1(b)**). The additional hydrogen bond in a G–C base-pair leads to increased rigidity and thermal stability relative to the A–T base pair. DNA helices rich in A and T are therefore more flexible and melt at lower temperatures than those rich in G and C, allowing for sequence-dependent differences in malleability of the DNA structure upon protein binding.

The bulk of DNA-sequence recognition is due to interactions between protein side chains and functional groups on the base-pair edges in the major groove. Most of these interactions are through hydrogen bonds, but proteins often make hydrophobic contacts to the methyl group of thymine as well. The

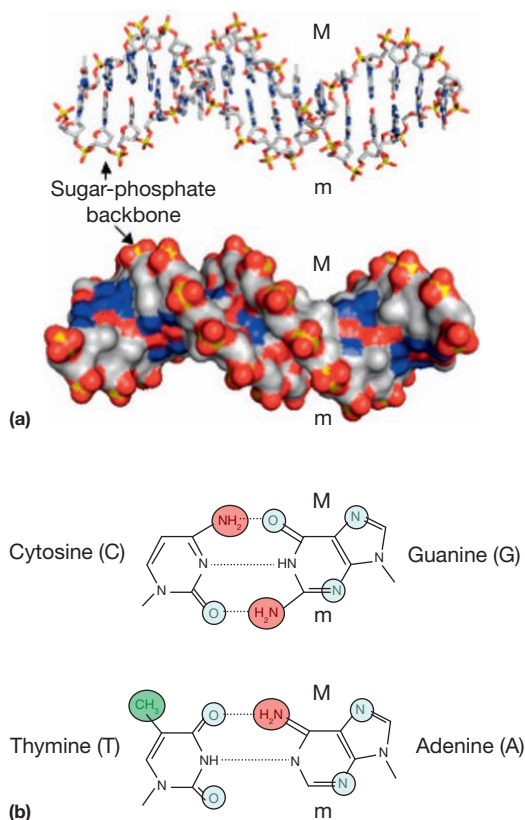


Figure 1 Structural and chemical properties of DNA. (a) DNA exists as a double-stranded helix with a core of stacked base pairs sandwiched between a negatively charged deoxyribose-phosphate backbone. Atomic and space-filling representations are shown. The DNA major (M) and minor (m) grooves are labeled. Atoms are colored by element. (b) Watson-Crick base pairing in DNA. The major and minor grooves contain hydrogen-bond donors (blue), hydrogen bond acceptors (red), and hydrophobic groups (green) that can be recognized by DNA-binding proteins.

edges of guanine and cytosine present an asymmetric pattern of four hydrogen-bond acceptor and two hydrogen-bond donor groups, enabling easy discrimination of G-C versus C-G base pairs (Figure 1(b)). Lysine and arginine side chains are often found interacting with G-C base pairs in the major groove, generally by hydrogen bonding to the N and/or O atoms of G. The positions of hydrogen bond donors and acceptors along the edges of A-T and T-A base pairs in the major groove are symmetric, making discrimination more difficult (Figure 1(b)). However, A-T and T-A base-pairs can be readily distinguished by the bulky methyl group of thymine, which forms van der Waals contacts with a number of protein side chains. Side chains of glutamine and asparagine are often used to hydrogen bond to adenine in A-T base pairs of DNA.

Nonspecific Protein-DNA Interactions

Not all DNA-binding proteins recognize and directly contact a specific sequence. Some proteins instead depend only on recognition of the general shape and properties of the DNA helix for binding. An important component of such nonspecific

interfaces involves hydrogen-bonding interactions between the protein and the phosphates in the DNA backbone. These generally involve hydrogen bond donors such as the side chains of serine, threonine, tyrosine, asparagine, glutamine, and histidine, and the amide NH groups on the polypeptide backbone. A particularly common type of polar interaction involves the strongly electrostatic hydrogen bonds formed between positively charged lysine and arginine side chains and the negatively charged phosphates.

In addition to polar interactions with the phosphates, DNA-binding proteins also make numerous van der Waals contacts to the deoxyribose groups in the DNA backbone. The deoxyribose groups are primarily hydrophobic, lining the walls of the major and minor grooves of the helix (colored gray in Figure 1(a)). Analyses of high-resolution structures of a number of protein-DNA complexes have revealed that these are the most abundant types of interactions.

Specific protein-DNA interfaces generally contain similar types of polar phosphate interactions and abundant van der Waals contacts to deoxyribose groups that are found in non-specific interfaces. However, proteins that recognize specific DNA sequences also contain additional elements responsible for recognizing correct target sequences and for rejecting incorrect sequences. The nature of these structural motifs is described in the next section.

Protein Structural Elements Used to Bind DNA

α -Helices

The α -helix is the structural element most frequently used for sequence-specific interactions in protein-DNA interfaces. The size of an α -helix matches the width of the DNA major groove, allowing them to fit together tightly while the protein side chains on the helix probe the available base-pair functional groups. The mode of helix docking to the DNA varies widely among different DNA-binding domains that use this element. The maximum number of side-chain-DNA contacts can be achieved when the α -helix docks in the major groove with its axis parallel to the phosphate backbone. However, helices also insert into the major groove in an end-on fashion and with their helical axes parallel to the base pairs. Some examples can be seen in Figures 2(a)–2(d). Helices can insert into the DNA minor groove as well, as seen with the lactose repressor. However, because the minor groove must be distorted to fit the helix, this type of interaction is not often seen and generally only one or two residues from the helix are inserted.

β -Strands

β -Strands paired together to form small β -sheets are also found in sequence-specific DNA interfaces. In these complexes, the β -sheets lie flat within the major groove, where side chains on the exposed surface of the sheet interact with functional groups on the edges of neighboring base pairs. An example is the Arc repressor-DNA complex, shown in Figure 2(e). Larger β -sheets take on a rigid, slightly twisted structure and their insertion into either the major or minor groove requires considerable

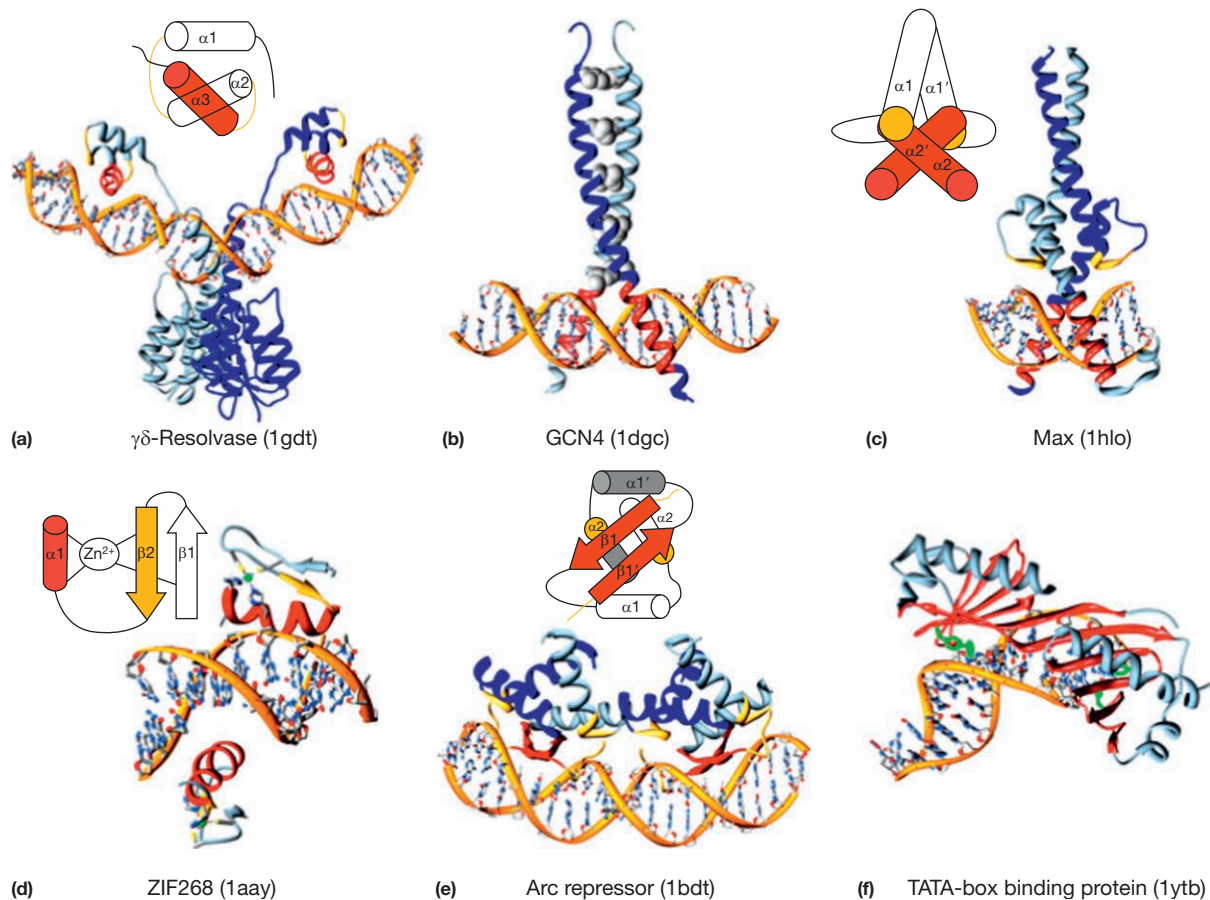


Figure 2 Examples of DNA-binding motifs bound to specific DNA sequences: (a) helix–turn–helix, (b) basic leucine zipper, (c) basic helix–loop–helix, (d) zinc finger, (e) ribbon–helix–helix, and (f) β -sheet motifs bound to DNA. α -Helices and β -strands that form sequence-specific interactions with the DNA are colored red. Protein structural elements that form nonspecific contacts with the DNA phosphate backbone are colored yellow. Intercalating phenylalanine residues positioned along the TATA-box binding protein's β -sheet edge are colored green. The DNA is shown as a ribbon structure with the sugar-phosphate backbone colored orange and base pairs colored by atom. The Protein Data Bank codes for the examples are given in parentheses.

distortion of the DNA helix, as shown for the TATA-box binding protein (TBP) in [Figure 2\(f\)](#).

Loops

Both α -helices and β -strands act as rigid scaffolds that direct residue side chains into the DNA major groove in a specific orientation. In contrast, protein loops have the versatility to adopt a wide range of conformations. This flexibility allows the loop to vary its position within the DNA major groove, altering the orientation of specific residues in order to achieve optimal side chain–base interactions. Loops are common DNA-binding elements in DNA-binding proteins with immunoglobulin-like domains, such as the tumor suppressor p53 and the transcription factor nuclear factor-kappaB (NF- κ B).

DNA-Binding Motifs

There are currently over 1500 high-resolution structures of protein–DNA complexes deposited in the Protein Data Bank,

representing approximately 10 different DNA-binding motifs. The motifs contain a mixture of α -helical, β -strand, and loop structural elements that interact via the DNA major and minor grooves. Summarized below are examples of the most common DNA-binding motifs. More in-depth descriptions can be found in the texts recommended for further reading.

Helix–Turn–Helix

The helix–turn–helix (HTH) motif consists of a right-handed three helical bundle with helices $\alpha 2$ and $\alpha 3$ arranged perpendicular to one another ([Figure 2\(a\)](#)). Sequence-specific DNA binding is mediated by the recognition helix $\alpha 3$, which docks into the DNA major groove forming hydrogen-bond and van der Waals contacts with functional groups on the exposed base pairs and the deoxyribose-phosphate backbone. The protein–DNA interaction is stabilized by nonspecific contacts between the DNA backbone and residues in $\alpha 2$ and the turn between $\alpha 2$ and $\alpha 3$. HTH motifs often exist as parts of larger domains and the orientation of $\alpha 3$ within the major groove is greatly dependent on the tertiary structure of the surrounding

elements. In prokaryotic DNA-binding proteins, HTH motifs function as homodimers with $\alpha 3$ docking parallel to the base-pairing edges within the major groove. In contrast, $\alpha 3$ from monomeric eukaryotic HTH motifs binds parallel to the DNA sugar-phosphate backbone and tracks along the major groove. The HTH–DNA interaction is elaborated in some DNA-binding domains by the insertion of additional motifs that provide auxiliary contacts to the DNA helix. For example, the winged HTH motif involves the combination of the HTH motif with a flanking β -hairpin that extends the protein–DNA interface.

Basic Leucine Zipper and Helix–Loop–Helix

Leucine zippers (bZIP) are rigid dimers of parallel extended helices. The dimer is stabilized by a coiled-coil interaction between heptad repeats of hydrophobic leucine and/or isoleucine residues along the C-terminal regions of each helix. A well-studied example is the DNA-binding domain from the transcription factor GCN4 shown in [Figure 2\(b\)](#). The two N-terminal regions in the bZIP domain splay out from the coiled coil and straddle the DNA major groove. Interestingly, the bZIP N-terminal DNA-binding regions are often unstructured in the absence of DNA. This lack of structure presumably allows the protein to sample a variety of potential interactions within the exposed DNA elements until the optimal DNA target sequence is found. Side chains on the docked helices form specific contacts with the target nucleotides and nonspecific contacts with the surrounding sugar-phosphate backbone. Coupling helical structure formation to DNA binding and sequence recognition is an interesting example of how specificity can be regulated in a DNA-binding protein.

Basic helix–loop–helix (bHLH) motifs are modifications of the extended helical structure of bZIP proteins. The bHLH monomeric unit consists of two short extended helices – a C-terminal coiled-coil dimerization helix and an N-terminal DNA-binding helix – separated by a loop. An example is shown in [Figure 2\(c\)](#). Dimerization results in the formation of a four-helix bundle near the loop, allowing for increased flexibility in the positioning of the N-terminal helices in the DNA major groove. Both the bZIP and bHLH motifs can exist as homo- and heterodimers, greatly increasing the variety of target DNA sequences recognized by these proteins.

Zinc Finger

Zinc fingers are encoded by approximately 3% of the human proteome and are the most common DNA-binding motif in eukaryotes. Transcription factors often use zinc fingers to mediate protein–DNA interactions. The most common zinc finger motif consists of a two-stranded anti-parallel β -sheet and a short α -helix ($\beta\beta\alpha$) held together by a Zn^{2+} ion tetrahedrally coordinated by conserved cysteine and histidine residues in the α -helix and two β -strands ([Figure 2\(d\)](#)). In all cases, Zn^{2+} plays a structural role and does not interact directly with the DNA. The α -helix serves as the recognition element, inserting into the DNA major groove where side chains form specific contacts with three consecutive base pairs. The overall interaction is further stabilized by nonspecific contacts between the β -sheet and the sugar-phosphate backbone of DNA.

Zinc fingers generally occur in units of 3–15 tandem repeats, where the number of DNA-binding motifs directly affects the sequence specificity of the interaction. The length of the linker region between zinc fingers determines the number and spacing of the recognition helices docked in the DNA major groove. It is also common for the linker regions and zinc fingers to form stabilizing nonspecific contacts across the DNA minor groove. The use of multiple zinc finger motifs in a single DNA-binding module allows for recognition of an extended (and therefore more specific) DNA sequence, since each finger can in principle recognize a base-pair triplet. This modular feature has also led to the development of engineered zinc finger DNA-binding proteins that can be tailored to bind specifically to alternative sequences of interest.

β -Ribbon–Helix–Helix

The ribbon–helix–helix motif consists of two intertwining monomers, each of which contains a β -strand followed by two α -helices ($\beta\alpha\alpha$)₂ ([Figure 2\(e\)](#)). Dimerization generates a two-stranded anti-parallel β -sheet, which is the primary DNA-binding element. Initial contacts between the sugar-phosphate backbone and the $\alpha 2$ peptide backbone play a critical role in orienting the β -sheet for insertion into the DNA major groove with the strands parallel to the sugar-phosphate backbone. Once docked, side chains on the exposed face of the sheet form hydrogen bonds with functional groups on the base-pair edges. The complex is stabilized through additional contacts between the helices and the sugar-phosphate backbone. The ($\beta\alpha\alpha$)₂ motif often dimerizes to form a symmetric tetramer that can recognize both symmetric and asymmetric DNA sequences, as shown in the example in [Figure 2\(e\)](#). The specificity and stability of ($\beta\alpha\alpha$)₂–DNA interactions can be further enhanced by the insertion of additional structural elements into the minimal ribbon–helix–helix fold.

β -Sheet

The TBP represents one of the few protein families that use an extended β -sheet to bind target DNA. TBP is a transcription factor that recognizes and binds to a TATA-box sequence that is extremely A/T-rich. TBP consists of two domains separated by a concave anti-parallel 10-stranded β -sheet which covers and binds at the DNA minor-groove (shown in [Figure 2\(f\)](#)). This interaction is facilitated by two phenylalanine residues from each β -sheet edge which intercalate into the DNA minor groove. This intercalation causes a broad opening of the minor groove and large-scale distortion of the DNA which can then accommodate the twisted structure of the large sheet. Sequence specificity is achieved through contacts between residues on the β -sheet surface and the edges of newly accessible base pairs and through the sequence-dependent deformability of the A/T-rich TATA-box sequence.

Conclusions

Proteins have evolved a number of structural motifs capable of binding to specific DNA sequences. Complementarity of the protein motifs with the shape and properties of the DNA double helix governs initial protein–DNA docking and

provides a substantial fraction of binding affinity. Specificity is achieved primarily through side-chain interactions with patterns of polar and nonpolar groups presented along the base-pair edges of a target DNA sequence within the major and minor grooves. Sequence-dependent deformations of the DNA helix also contribute to binding specificity. Proteins mediating tightly regulated processes often employ a combination of motifs to ensure a high level of specificity and affinity for the correct sequence. Understanding the factors that govern specific protein–DNA interactions has important implications in the identification of regulatory sites in genomes, as well as in the *de novo* design and engineering of DNA-binding proteins with altered affinities and/or specificities.

See also: **Molecular Biology:** DNA Replication: Eukaryotic Origins and the Origin Recognition Complex; DNA Replication: Initiation in Bacteria; DNA Secondary Structure; **Protein/Enzyme Structure Function and Degradation:** Zinc Fingers; **Signaling:** p53 Family.

Further Reading

- Garvie CW and Wolberger C (2001) Recognition of specific DNA sequences. *Molecular Cell* 8: 937–946.
- Luscombe NM, Austin SE, Berman HM, and Thorton J (2000) An overview of the structures of protein–DNA complexes. *Genome Biology* 1(1): reviews 001.1–001.37.
- Neidle S (2007) *Principles of Nucleic Acid Structure*. London: Academic Press.
- Norambuena T and Melo F (2010) The Protein–DNA Interface database. *BMC Bioinformatics* 11: 262–274.
- Rice PA and Correll CC (2008) *Protein–Nucleic Acid Interactions: Structural Biology*. Cambridge: RSC Publishing.
- Rohs R, Jin X, West SM, et al. (2010) Origins of specificity in protein–DNA recognition. *Annual Review of Biochemistry* 79: 233–269.
- Shrivastava T and Tahirou TH (2010) Three-dimensional structures of DNA-bound transcriptional regulators. *Methods in Molecular Biology* 674: 43–55.

Relevant Website

<http://www.rcsb.org> – RSCB Protein Data Bank: an information portal to biological macromolecular structures.

DNA Supercoiling

T-S Hsieh, Duke University Medical Center, Durham, NC, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

DNA twist The number of times that two complementary strands wind around one another in the double-helical structure.

Kilobase (kb) One thousand base pairs.

Linking number The number of times that DNA strands wind around one another in space. For a closed circular DNA, it is an integer and an invariant.

Supercoil The winding of the DNA helical axis in space.

Topoisomerase An enzyme that can change supercoiling or other aspects in the topological structure without altering any covalent structure in DNA.

Background

The beauty and elegance of the double-helical structure of DNA, as first proposed by Watson and Crick in 1953, also impart critical functions of DNA. However, the unwinding and rewinding of the DNA double helix during the process of replication, transcription, and recombination create intriguing topological problems. The potential topological problem intrinsic to a circular, double-stranded DNA molecule was first noted by Cairns in 1963. In his autoradiographic analysis of chromosome replication in the bacterium *Escherichia coli*, Cairns pointed out that the bihelical DNA structure of a circular molecule poses a topological constraint on the strand separation necessary in the process of DNA replication. About the same time, Vinograd and his associates demonstrated that an animal virus, polyomavirus, also has a circular genome. Furthermore, the helical axis itself winds and turns in space, thus producing a superhelix. Since then, circular genomes have been discovered in many organisms, including most bacteria, bacterial plasmids, and organelles such as mitochondria and chloroplasts. For example, Figure 1 shows an electron micrograph of a small bacterial plasmid DNA. The DNA superhelicity results in a duplex making crossovers (intersections) with itself. For most circular DNAs isolated from natural sources, the number of superhelical turns is directly proportional to the length of DNA. There are approximately six to eight superhelical turns per 1 kb DNA. We discuss the quantitative details in later sections.

The biological significance of DNA superhelicity was first realized when Vinograd and his co-workers discovered that there is a tight coupling between the DNA secondary structure and superhelical structure. Therefore, DNA superhelicity can influence the unwinding and rewinding of DNA duplex, and vice versa. Because such a coupling is direct, it allows one a rare opportunity in molecular biology to design experiments to gain insight into the structure of DNA and to probe the mechanism of the enzymes that affect DNA structure. The theoretical consideration of the coupling between DNA secondary and tertiary structures, based on mathematical topology, was later extended in the treatises by Fuller, White, and Crick. It should be pointed out that such a topological coupling also applies to linear DNA with anchored sites to create the loop structure that is found in the chromosomes of essentially all cells.

Nature has evolved unique enzymes to solve the topological problems that arise during the process of unwinding/rewinding DNA helix. The first such enzyme, topoisomerase I from the bacterium *E. coli*, discovered by Wang in 1970, can remove negative supercoils efficiently. Many new families of topoisomerases have since been discovered, and these enzymes are present in all organisms and have essential biological functions.

Quantitative Relationship between Twist and Supercoil

For a circular DNA without any nicks, in which both DNA strands are covalently continuous, the sum of DNA twists (Tw) and writhes of the DNA helical axis (Wr), a measure of DNA superhelical structure, is invariant. Because the algebraic sum of these two numbers is equal to the number of times that one of the DNA strands winds around its complementary strand in space, it defines the topological linkage between DNA strands, termed the linking number (Lk). We therefore have the following equation:

$$Tw + Wr = Lk$$

where Lk is a constant for a covalently closed circular DNA. Tw is related to the DNA secondary structure, the bihelical winding of two DNA strands, and it can be calculated if one knows the size of the DNA and the average helical pitch in this DNA. For example, in a 1-kb B-form DNA with a helical pitch of 10.5 bp/turn, $Tw = 95$. Wr is a measure of the superhelical structure of DNA. It relates to the number of times that the helical axis crosses itself in space. The higher the Wr , the more supertwisted the DNA. Lk is a topological quantity and defines the total linkage between the two complementary strands. Whereas Tw and Wr can be any real number, Lk is always an integer. This is because, for a closed circular DNA, strands have to wind around one another an integral number of turns to avoid nicks or gaps.

Handedness and the Sign of Tw and Wr

An important feature for all three parameters is that each has a sign, positive or negative, depending on the handedness of

the double helix and superhelix. Because most of the DNA duplexes are right handed, we define right handed as positive and left handed as negative. We can define whether the cross-over of two curves in space is left handed or right handed if an axis is clearly marked. For example, **Figure 2** shows two curves, denoted as two arrows in space, one making a left-handed cross (**Figure 2(a)**) and the other a right-handed one (**Figure 2(b)**). The axis is shown as a thin vertical line going through the crossover. The significance of the definition of the axis can be

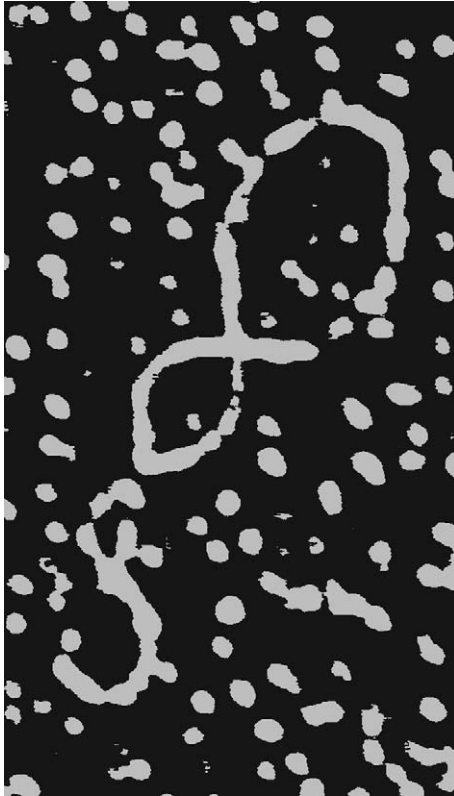


Figure 1 An electron micrograph of a bacterial plasmid DNA. The DNA has about three superhelical turns, as evidenced by the number of crossovers made by the DNA duplex.

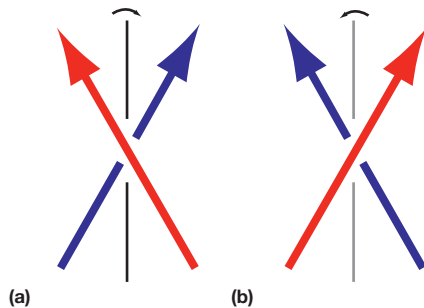


Figure 2 Diagrams of (a) a left-handed crossover and (b) a right-handed one. The arrows shown here serve to illustrate the point that the two perpendicular axes, one vertical (shown here) and one horizontal, are not equivalent. They also illustrate the rotational process to bring upper strand to the lower one. The directionality of such rotation is marked with a curved arrow around the helical axis.

visualized in this figure as well. If one chose instead a horizontal line as the axis, then the assignment of handedness is reversed. **Figure 2(b)** would then be left handed and **Figure 2(a)** would be right handed. For the duplex DNA structure, the axis definition is rather intuitive and straightforward. The helical axis is the axis around which the two strands are wrapped with respect to one another. There is an alternative method in defining handedness without the concern of marking a helical axis. It is based on the relative rotation of upper strand to the lower one (red arrow over blue in **Figure 2**). If a counterclockwise rotation is required to make the upper strand coincident with the lower one, then it is right handed; conversely, a clockwise rotation for upper strand matching the lower one will be defined as left handed.

For the DNA superhelix, it is easier to use the upper-to-lower strand rotation to define handedness. However, it is also possible to mark the helical axis for handedness definition. In this case, it is necessary to define an axis in a manner that is consistent with what we defined for secondary structure. An example is shown in **Figure 3**. A superhelical DNA with two superhelical turns is shown here. If one lays down the superhelical DNA horizontally, then the axis is the vertical line going through the crossover. With this definition of axis, the superhelix shown in **Figure 3(a)** has two left-handed crossovers, and thus it has two negative superhelical turns ($Wr = -2$); **Figure 3(b)** has two positive superhelical turns ($Wr = +2$).

Because it is not practical to measure the absolute value of Wr or Lk , they are usually determined experimentally with respect to a reference state, DNA in a relaxed state. The linking number and writhe for the relaxed state are denoted with a subscript zero.

Because $Wr_0 = 0$, the measure of DNA supercoils, Wr , is equal to the writhe difference with respect to the reference state:

$$Wr = \Delta Wr = Wr - Wr_0$$

Assuming that Tw for the supercoiled and relaxed DNA is about the same ($Tw = Tw_0$), then it follows that

$$\Delta Wr = \Delta Lk$$

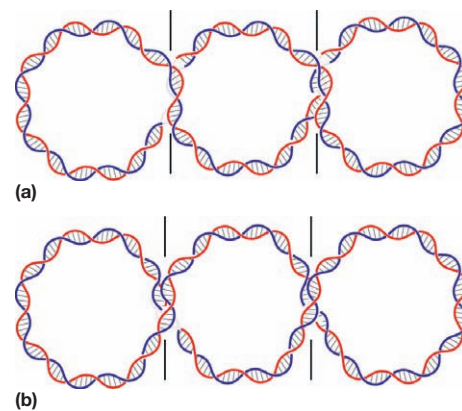


Figure 3 Diagram of a superhelical DNA with (a) two left-handed supercoils and (b) with two right-handed supercoils. Notice the two complementary strands (in red and blue) are wound in a right-handed helix. The axis for determining the handedness of the superhelical turn is shown as a vertical line going through the DNA crossovers.

Because Lk is a topological invariant, it does not change for a given DNA in different environments. It is thus a more useful parameter for tracking the structure of DNA. As a result, ΔLk is frequently used to mark the superhelicity of DNA. If ΔLk is negative, DNA is underwound; the absolute value of ΔLk in this case gives a measure of negative supercoils or linking deficiency. If ΔLk is positive, then the DNA is overwound; its absolute value relates to the number of positive superhelical turns. DNA superhelicity can also be expressed as a fraction with respect to the linking number in a relaxed DNA. This is usually referred to as superhelical density, σ :

$$\sigma = \Delta Lk / Lk_0 = \Delta Lk / Tw_0$$

because $Lk_0 = Tw_0$, when $Wr_0 = 0$.

Most of the circular DNAs isolated from natural sources are negatively supercoiled. An interesting exception is the DNA in hyperthermophilic bacteria and archaeobacteria, which is usually positively supercoiled. For the numerous negatively supercoiled DNAs studied so far, the superhelicity mostly falls within a narrow range, with σ between -0.06 and -0.08 .

Biological Effects of Supercoiling and Enzymes That Can Change DNA Supercoiling

To access the genetic information embedded within the double helix, most of the biological processes associated with critical functions of DNA involve unwinding or rewinding of the twin DNA strands. With a covalently closed circular DNA for which Lk stays constant, any change in Tw will be compensated by a change in Wr that is equal in magnitude but opposite in sign:

$$\Delta Tw = -\Delta Wr$$

This quantitative coupling between the DNA secondary structure and superhelical structure suggests that supercoiling can profoundly influence DNA structure and function. DNA with supercoils, either positive or negative, is thermodynamically less stable than its relaxed counterpart. For a negatively supercoiled DNA, any reactions associated with a reduction in the negative superhelical turns ($\Delta Wr > 0$) are energetically favored. These reactions will also be coupled with the unwinding of the DNA duplex because $\Delta Tw < 0$ if $\Delta Wr > 0$. Therefore, negative DNA supercoiling facilitates the unwinding of DNA. In many instances, it has been demonstrated that replication, transcription, and recombination, all of which require the

unwinding of DNA, are promoted by negative supercoiling. In an interesting corollary, DNA with positive supercoils will favor winding and disfavor unwinding of duplex structure (i.e., $\Delta Tw > 0$ if $\Delta Wr < 0$). It has been hypothesized that positive supercoiling found in some of the hyperthermophiles serves to stabilize the DNA double-helical structure in the extreme temperatures at which these organisms find home.

Nature has harnessed the use of DNA supercoiling for regulating the biological functions of DNA; it has also developed tools to modulate supercoiling. A ubiquitous class of enzymes, DNA topoisomerases, evolved to carry out the topological transformation in DNA. In order to break the hold of the topological constraint on closed circular DNA, these enzymes can make transient and reversible DNA breaks. There are two important features in these topoisomerase-mediated breaks in comparison with those made by nucleases. One is that a specific tyrosyl residue in the enzyme is joined covalently to the phosphate at the breakage site through a phosphodiester bond. Because this enzyme–DNA bridging bond is made at the expense of the neighboring DNA backbone bond, there is no significant energy change associated with the process of strand scission, thus assuring that the cleavage/religation is readily reversible. The enzyme-mediated DNA breakage is also accompanied by strand passage through the break or a rotation of the strand with the transient break around the intact strand. Both processes result in a change of DNA supercoiling or other topological transformation in DNA. Indeed, genetic experiments have demonstrated essential functions of these enzymes in the cell. These results also underscore the importance of supercoiling in the biological functions of DNA.

See also: **Molecular Biology:** DNA Secondary Structure; DNA Topoisomerases: Type I; DNA Topoisomerases: Type II.

Further Reading

- Wang JC (1980) Superhelical DNA. *Trends in Biochemical Sciences* 5: 219–221.
- Wang JC (1994) Appendix I: An introduction to DNA supercoiling and DNA topoisomerase-catalyzed linking number changes of supercoiled DNA. In: Liu L (ed.) *Advances in Pharmacology*, pp. 257–270. San Diego, CA: Academic Press.
- Wang JC (2009) *Untangling the Double Helix*. New York, NY: Cold Spring Harbor Laboratory Press.

DNA Topoisomerases: Type I

J J Champoux, University of Washington, Seattle, WA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Catenane Two interlocked circular DNA molecules in which the two duplexes are wound around each other one or more times.

Catenate The process whereby two circular DNAs are interlocked to form a catenane. The unlocking of catenated DNAs is referred to as decatenation.

Closed circular DNA Circular DNA in which both strands are intact.

DNA gyrase DNA topoisomerase that couples the hydrolysis of adenosine triphosphate (ATP) to the introduction of negative supercoils into a closed circular DNA.

DNA supercoiling The coiling of the axis of a DNA molecule in three-dimensional space. Supercoiling may result either from an interaction of the DNA with protein or from an inequality between the number of helical turns dictated by the structure of the DNA helix under a particular set of conditions (the twist of the DNA) and the linking number of the DNA.

Linking number Topological property of a closed circular DNA that is a measure of the fixed interwinding of the two DNA strands.

Precatenanes The interwinding of the two daughter duplexes behind a replication fork.

Reverse gyrase DNA topoisomerase that couples the hydrolysis of ATP to the introduction of positive supercoils into a closed circular DNA.

Scissile strand The strand of DNA that is cleaved by a type I topoisomerase.

Topoisomerase Enzyme that changes the linking number of a closed circular DNA by temporarily breaking one (type I) or both (type II) of the strands of the DNA.

Topoisomers Variants of a closed circular DNA that have different linking numbers.

Torsionally strained supercoils Supercoils that result from an inequality between the number of helical turns dictated by the structure of the DNA helix under a particular set of conditions and the linking number of the DNA.

Both the large size of DNA molecules and the double-helical nature of DNA create unique topological problems during replication, transcription, recombination, and chromatin remodeling, which are solved by a family of enzymes called DNA topoisomerases. Members of the type II subfamily of DNA topoisomerases alter the supercoiling of DNA and disentangle chromosomes by introducing temporary double-strand breaks into the DNA. Type I DNA topoisomerases, the subject of this article, manage DNA topology in the cell by transiently cleaving only one of the two DNA strands.

Reactions Catalyzed by Type I DNA Topoisomerases

To introduce a temporary single-strand break into duplex DNA, type I DNA topoisomerases must catalyze the cleavage and subsequent religation of a DNA strand. Since these two reactions occur without external energy sources such as adenosine triphosphate (ATP), cleavage cannot result from simple hydrolysis of a phosphodiester bond in the DNA. Instead, a covalent enzyme–DNA intermediate is generated that makes the religation step energetically feasible. The formation of the covalent intermediate involves nucleophilic attack by the O-4 atom of the active site tyrosine in the enzyme on a phosphodiester bond in the DNA to produce a phosphodiester bond between the tyrosine and the DNA and leave a free DNA hydroxyl end. DNA religation and release of the enzyme is the reverse reaction with the oxygen of the free DNA hydroxyl acting as the nucleophile. The type I enzymes display a loose preference for certain

nucleotides in the vicinity of a cleavage site; therefore, a cleavage site typically occurs every 5–20 base pairs along the DNA.

By introducing a transient nick in DNA, type I topoisomerases can act on closed circular DNAs to change the number of times one strand winds around the other, a parameter referred to as the linking number of the DNA. Changes in the linking number are reflected in either a reduction or an increase in the supercoiling of a plasmid DNA, a property that is most often measured by gel electrophoresis. In addition to altering the supercoiling of a plasmid DNA, many type I topoisomerases are capable of catalyzing a number of other transactions involving both single- and double-stranded DNAs. Most of these enzymes can catenate (interlock), decatenate, knot, and unknot single-stranded DNA circles. The same series of reactions can be carried out with duplex circular DNAs, provided at least one of the circular molecules possesses a nick or a gap. In some cases, the enzymes can facilitate the interwinding required for the renaturation of two complementary single-stranded circular DNAs, a reaction that could be important during homologous recombination. Interestingly, topoisomerase V, which has only been described in the hyperthermophilic archaeon *Methanopyrus kandleri*, possesses, in addition to the usual topoisomerase activity, an apurinic/apyrimidinic site-processing activity that would appear to implicate the enzyme in DNA repair. Finally, with certain unusual DNA substrates, a block to religation leads to permanent suicide cleavage and the enzyme remains covalently linked to the DNA. The anti-cancer drug camptothecin blocks religation and similarly leads to the accumulation of enzyme–DNA covalent complexes.

Classification, Nomenclature, and General Properties

Type I topoisomerases are classified into three distinct subfamilies (types IA, IB, and IC) based on structural considerations. In addition, the type IA subfamily differs from the IB and IC subfamilies with respect to which DNA end becomes covalently attached to the enzyme during the cleavage reaction: type IA enzymes attach via a tyrosine phosphodiester linkage to the 5' end of the DNA, whereas the enzymes in the other two subfamilies attach to the 3' end of the DNA. **Table 1** lists the known type I DNA topoisomerases in the three subfamilies with their common names and origins; representatives of the IA and IB topoisomerases are found in all three domains of life, while the only known member of the IC subfamily is found in an archaeon. The common names have generally been assigned in the order of discovery using odd Roman numerals (even Roman numerals are similarly used for type II DNA topoisomerases). Type I enzymes with unusual properties or origins have been given unique names (reverse gyrase, poxviral topoisomerase, and mitochondrial topoisomerase). The recently described IB enzymes in some eubacteria are referred to as bacterial topoisomerases IB.

The three categories of type IA enzymes listed in **Table 1** can be distinguished on the basis of the types of reactions they catalyze. Topoisomerases I relax negative but not positive supercoils in plasmid DNAs; however, since relaxation does not go to completion, some residual negative supercoils remain in the product. Topoisomerases III require hypernegatively supercoiled plasmid DNA as a substrate and again relaxation is incomplete. Interestingly, topoisomerases III are much more proficient than the topoisomerases I in DNA catenation and decatenation (see below). Reverse gyrases, which are only found in hyperthermophilic and some thermophilic eubacteria and archaeobacteria, introduce positive supercoils into plasmid DNAs at the expense of ATP hydrolysis. All of the type IA enzymes require Mg^{2+} and are monomeric with the exception of the reverse gyrase from the archaeon *M. kandleri* which is a heterodimer.

The type IB DNA topoisomerases are capable of relaxing both positive and negative supercoils in a reaction that does not require ATP or divalent cations. Although the reactions go to completion to produce a completely relaxed set of plasmid DNA topoisomers, the nuclear enzyme in eukaryotes exhibits a

preference for binding supercoiled DNAs of either sign over completely relaxed DNA. With the exception of the recently discovered heterodimeric topoisomerases I from trypanosomatids, all of the type IB enzymes are monomeric.

Type IA DNA Topoisomerases

Protein Domains

All type IA topoisomerases (**Table 1**) share a highly conserved cleavage/strand-passage domain that contains the active site tyrosine. This domain is also responsible for promoting the structural change in the DNA during the interval between the cleavage and religation reactions that results in a linking number change (see below) (**Figure 1**, red boxes). As indicated in **Figure 1**, all type IA enzymes contain a poorly conserved basic C-terminal domain (solid or hatched green boxes) and some contain a Zn(II)-binding domain as well (blue boxes). These latter two features appear to be important for the interaction of the enzyme with DNA. Finally, reverse gyrases contain an N-terminal domain, which resembles the ATPase domains of helicases (yellow box) and is connected to two domains that are structurally very similar to the cleavage/strand-passage and basic domains of the typical type IA topoisomerases.

Crystal Structure of the Conserved Cleavage/Strand-Passage Domain

The crystal structure of the cleavage/strand-passage domain of *Escherichia coli* DNA topoisomerase I shown in **Figure 2** provides key insights concerning the substrate preference of the enzyme and the mechanism of DNA relaxation. Notably, the cleavage/strand-passage domains of all type IA topoisomerases bear a strong resemblance to the *E. coli* structure. The hallmark of the crystal structure is a toroidal shape in which the diameter of the hole in the center of the torus is sufficient to accommodate either single- or double-stranded DNA.

The requirement for a negatively supercoiled substrate and the inability to completely relax negative supercoils is best explained by supposing that these enzymes will only bind an otherwise duplex DNA substrate if it contains a single-stranded region resulting from local unwinding of the helix. A plasmid

Table 1 Type I DNA topoisomerases

Subfamily	Common name	Source	Structure
IA	Topoisomerase I	All eubacteria and some archaeobacteria	Monomer
IA	Topoisomerase III	Some eubacteria and most eukaryotes	Monomer
IA	Reverse gyrase	All hyperthermophilic eubacteria and archaea	Monomer
IA	Reverse gyrase	Archaeon <i>Methanopyrus kandleri</i>	Heterodimer
IB	Topoisomerase I	Nucleus of all eukaryotes	Monomer
IB	Mitochondrial topoisomerase	Mitochondria of higher eukaryotes	Monomer
IB	Poxviral topoisomerase	All members of poxviridae family	Monomer
IB	Topoisomerase IB	Some eubacteria	Monomer
IB	Topoisomerase I	Archaea <i>Cenarchaeum symbiosum</i> and <i>Nitrosopumilus maritimus</i>	Monomer
IB	Topoisomerase I	Trypanosomatids <i>Trypanosoma brucei</i> and <i>Leishmania donovani</i>	Heterodimer
IC	Topoisomerase V	Archaeon <i>Methanopyrus kandleri</i>	Monomer

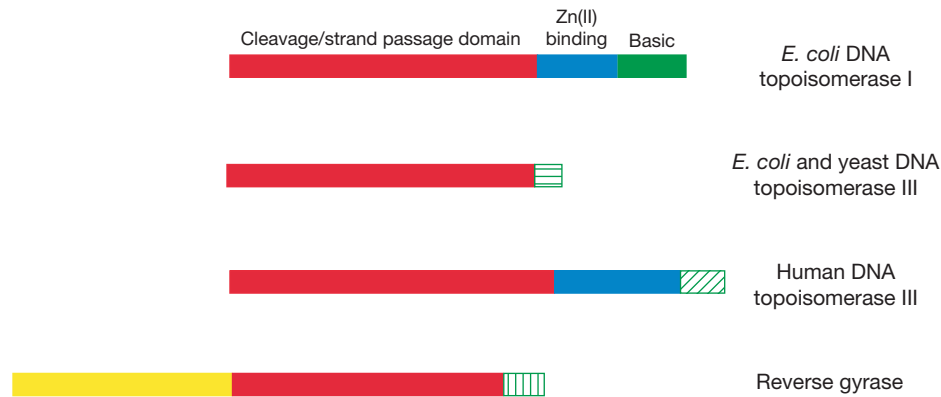


Figure 1 Domain structure and sequence relationships between type IA DNA topoisomerases. The domain structure of representative type IA topoisomerases is denoted by colored boxes. The names of the domains for the prototypic *E. coli* DNA topoisomerase I are given along the top. The cleavage/strand-passage domains shared by all the IA enzymes are shown in red. Green is used to denote the basic C-terminal domains, but the different types of fills for these boxes indicate that these domains are poorly conserved. Some type IA enzymes contain a Zn(II)-binding domain shown in blue. The helicase-like domain of reverse gyrase is shown in yellow.

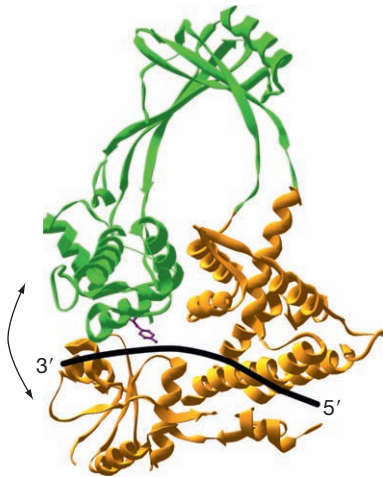


Figure 2 A ribbon diagram showing the crystal structure of the cleavage/strand-passage domain of *E. coli* DNA topoisomerase I (pdb entry 1ec1) drawn with Swiss-Pdb Viewer software (Glaxo Wellcome Experimental Research). The approximate path of the bound single-stranded substrate DNA is indicated by the solid black line, and the active site tyrosine is shown in magenta. The two regions of the protein that move relative to each other (double-headed arrow) to open and close the torus during the strand-passage reaction and to release the DNA are shown in green and orange.

DNA that is highly negatively supercoiled is energetically disposed toward helix unwinding, which explains why such a DNA is a good substrate for the enzyme. However, relaxation ceases when the negative supercoiling falls to a level below which there is insufficient energy to promote the required opening of the helix. As shown in [Figure 2](#), a single strand of DNA (solid black line) binds to a narrow groove on the cleavage/strand-passage domain of the enzyme in close proximity to the active site tyrosine (magenta).

Enzyme-Bridging Mechanism for Strand Passage

To change the linking number of a closed circular DNA during relaxation, one strand of DNA must pass through a break in the

other strand. Knotting, catenation, and decatenation of either single- or double-stranded DNAs similarly require such a strand-passage event. The strand-passage reaction for type IA topoisomerases occurs by what is referred to as an enzyme-bridging mechanism. Once the scissile DNA strand is cleaved, both DNA ends remain tightly associated with the enzyme; the 5' end is bound covalently to the active site tyrosine and the 3' end is bound noncovalently to the enzyme. To orchestrate strand passage, the enzyme undergoes a conformational change in which the top half of the protein containing the 5' end of the cleaved strand ([Figure 2](#), shown in green) lifts upward to generate a gate in the DNA through which another strand of DNA is passed. After strand passage, the broken strand is religated and the enzyme opens up a second time to release the strand that had been passed into the hole of the torus. This model predicts that the linking number can only be changed in steps of one and this prediction has been verified biochemically.

This same scheme can explain how the enzyme can catenate or decatenate a DNA containing a nick or a gap. However, it is unclear whether the DNA that passes through the temporary gate in the cleaved strand is captured in the hole of the torus before strand cleavage and is then passed out of the hole after cleavage or vice versa as described previously.

Reverse Gyrase Mechanism

Despite the presence of a helicase-like ATPase domain in the N-terminal region of reverse gyrases, these enzymes lack helicase activity when assayed under conditions that would require processive translocation along the DNA, and the N-terminal domain alone does not unwind DNA. Instead, the binding of the N-terminal domain to DNA in the context of the complete enzyme is believed to unwind a region of a closed DNA that, in turn, produces positive supercoiling within a region elsewhere in the same molecule. Subsequently, in a reaction dependent on ATP, one of the two strands in the unwound region is cleaved and the other strand is passed through the resultant gate by the cleavage/strand-passage domain present in the C-terminal half of the molecule. ATP hydrolysis releases

the enzyme, leaving behind the positive supercoils. The key to positive supercoiling is that the strand-passage event that occurs in the presence of ATP is directional such that the linking number of the DNA is increased and therefore the DNA ends up positively supercoiled.

Type IB DNA Topoisomerases

Domain Structure and Sequence Conservation

Type IB DNA topoisomerases are present in the nucleus of all eukaryotic cells and the mitochondria of higher eukaryotes, as well as in some archaea and eubacteria (Table 1). The typical eukaryotic type IB enzyme possesses the four domains shown in Figure 3. A highly charged and poorly conserved N-terminal domain (red box) is followed by the core domain (blue box), which binds DNA and contains most of the catalytic residues. The active site tyrosine is found in the C-terminal domain (yellow box) that is connected to the remainder of the protein by a poorly conserved linker region (orange box).

All of the other type IB enzymes share at least partial sequence and structural homology with the catalytically important core domain, as can be seen from the color scheme in Figure 3 (blue boxes). Where present, the sequence of the mitochondrial enzyme is very similar to the nuclear enzyme with the exception of the N-terminal region (green box), which contains the organelle targeting signals. The archaeal-type IB topoisomerases lack the N-terminal domain, but are otherwise very similar to the eukaryotic enzymes. The eubacterial IB enzyme and the vaccinia topoisomerase are very similar to each other, but they lack most of the core domain as well as the conserved C-terminal domain that is characteristic of the other IB enzymes; instead, they share unique N-terminal and C-terminal regions (white and magenta boxes). The topoisomerase I found in the trypanosomatids is a heterodimer with one subunit containing the catalytic

core (blue box) and the other subunit containing a region homologous to the C-terminal domain of the prototypic eukaryotic sequence (yellow box).

Crystal Structure of Human Topoisomerase I

Two views of the crystal structure of human topoisomerase I (missing the N-terminal domain) with a bound 22-base-pair duplex oligonucleotide are shown in Figure 4. The protein is a bilobed structure that clamps completely around the DNA with the active site tyrosine (shown in black in Figure 4(a)) juxtaposed to the scissile phosphate. The linker region comprises the coiled-coil structure that protrudes conspicuously from the bottom portion of the enzyme and has an unknown function (Figure 4(a)). To release the DNA, the top half of the protein (shown in blue) must shift upward relative to the bottom half, as indicated in Figure 4(b) by the double-headed arrow. Likewise, DNA binding requires that the protein clamp be in an open conformation. The region of the human topoisomerase I structure shared by the poxviral and eubacterial IB enzymes corresponds to the portion of the core domain, depicted in yellow in Figure 4. Tyrosine recombinases, such as the bacteriophage λ and HP1 integrases, and cre recombinase are also structurally very similar to approximately the same region.

Catalysis

The co-crystal structure of human topoisomerase I with bound DNA reveals which amino acid residues in the protein are directly involved in catalysis. It is worth noting that the nucleophilic tyrosine O-4 does not appear to be activated for cleavage by general base catalysis, although a lysine residue acts as a general acid to protonate the leaving 5' oxygen. The pentavalent transition state is stabilized by hydrogen-bonding interactions between three basic amino acid side chains and the

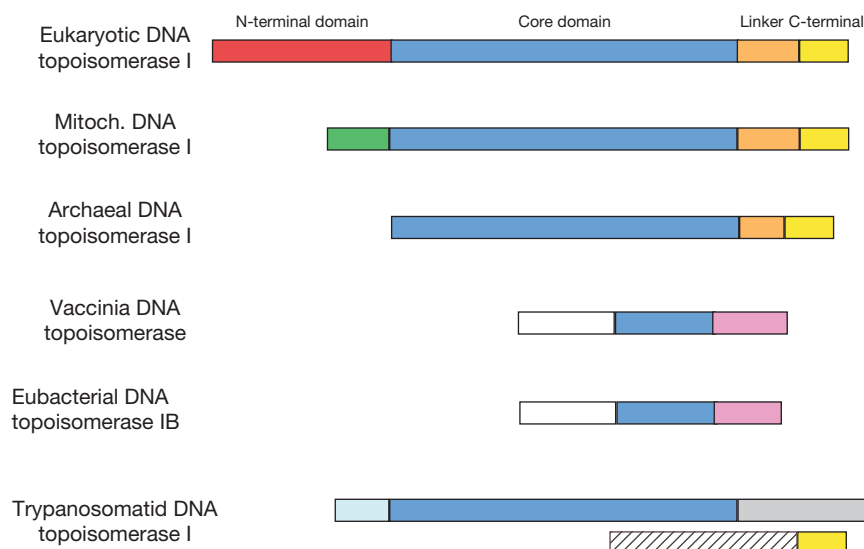


Figure 3 Domain structure and sequence relationships between type IB topoisomerases. The domain structure of the type IB topoisomerases from the indicated sources are denoted by colored boxes with similar domains aligned vertically. The names of the domains for the eukaryotic type IB topoisomerases are shown along the top. Regions that are similar in amino acid sequence share the same color; distinct sequences are assigned different colors. The two subunits of the heterodimeric topoisomerase I from trypanosomatids are shown with these same color conventions.

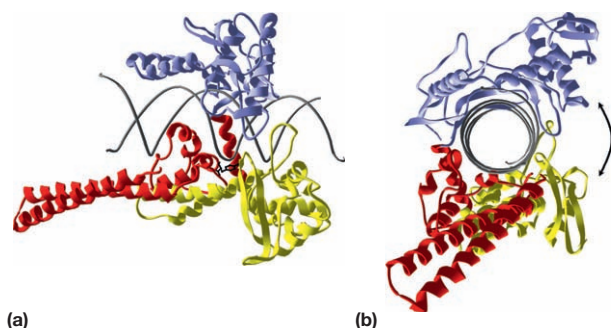


Figure 4 The ribbon diagrams show two views of the crystal structure of human topoisomerase I clamped around a 22-base-pair duplex DNA, shown in gray (pdb entry 1a36). The top lobe of the enzyme is shown in blue and the bottom lobe is shown in red and yellow. The poxviral topoisomerase and the eubacterial topoisomerases IB are structurally very similar to the region of the bottom lobe, shown in yellow. The coiled-coil linker region in bottom lobe (in red) is most easily seen in the side view shown in (a). The active site tyrosine (in black) is shown in (a). Panel (b) shows a view of the structure looking down the axis of the DNA. The double-headed arrow in (b) indicates the nature of the conformational change that is required to open and close the clamp during binding and release of the DNA.

scissile phosphate oxygens. An interaction between a lysine residue and the base of the nucleotide where cleavage occurs is also important for catalysis. Religation is likely to proceed by a pathway that is essentially the reverse of cleavage.

Rotational Mechanism for Strand Passage

Examination of the crystal structure of human topoisomerase I (Figure 4(a)) suggests that the strand-passage reaction required to change the linking number of the DNA during DNA relaxation occurs by a rotational mechanism rather than by the enzyme-bridging mechanism described previously for the type IA enzymes. However, based on the structure there appeared to be insufficient space within the confines of the enzyme to accommodate unrestricted rotation of the DNA. This prediction was confirmed by single-molecule experiments showing that rotation was hindered by friction between the DNA and the enzyme cavity and that the number of supercoils relaxed per nicking-closing cycle depended on the amount of supercoiling torque present in the DNA.

Cellular Roles

Although much is yet to be learned about how the various topoisomerases collaborate to manage DNA topology in the cell, a partial picture has emerged based on work in bacteria and simple eukaryotes. Although the type II topoisomerases are not the subject of this article, the activities of these enzymes are briefly considered in the sections to follow for the sake of completeness. The type II enzymes are important for any cellular process that requires the passage of a region of duplex DNA through a double-strand break in the same or a different DNA molecule. The allocation of functions to the known topoisomerases in setting the global levels of supercoiling, in transcription and in DNA replication, is discussed next.

Types of Supercoiling in Eukaryotic versus Prokaryotic Cells

Two different situations lead to the supercoiling of DNA *in vivo*. First, DNA will assume a supercoiled configuration through an interaction with certain proteins or other cellular components. Alternatively, a closed domain of DNA (e.g., a closed circular DNA) will spontaneously supercoil if the linking number is not the same as the helical winding (referred to as twist) of the DNA helix. A closed DNA with this latter type of supercoil is said to be torsionally strained. The chromosomal DNA of eukaryotes is wrapped into a protein-constrained solenoidal superhelix in nucleosomes and, except for the transient occurrence of torsionally strained supercoils associated with replication and transcription, is maintained in a relaxed state by DNA topoisomerases. However, in prokaryotes it appears that although some supercoils are constrained by virtue of an interaction with proteins as in eukaryotes, there exists, in addition, a fixed steady-state level of torsionally strained supercoiling generated by DNA gyrases (see below).

Generation of Supercoiling Stress in Prokaryotes

Mesophilic and psychrophilic bacteria

The DNA in all mesophilic and psychrophilic eubacteria and archaeobacteria contains torsionally strained negative supercoils that are introduced by the type II enzyme called DNA gyrase. It appears that this steady-state level of negative superhelicity is required to facilitate helix opening during the initiation of DNA replication and transcription. To prevent the introduction of excess negative supercoils by the gyrase, nearly all of these bacteria also contain one or two type IA DNA topoisomerases (topoisomerases I or III or both) to counteract the effects of DNA gyrase. The inability of the type IA enzymes to remove negative supercoils below a critical threshold level prevents these enzymes from completely negating the effects of DNA gyrase and is crucial for fine-tuning the negative supercoiling levels in these organisms. Some eubacteria also contain a type IB enzyme which could also balance the effects of DNA gyrase; however, the apparent ability of these enzymes to completely relax the DNA suggests that their activity would have to be regulated in some way to prevent the loss of the superhelicity generated by DNA gyrase. The observation that the eubacterial *topoisomerase IB* genes are widespread, but present in only a subset of the species in a variety of phyla (*Proteobacteria*, *Bacteroides*, *Actinobacteria*, *Acidobacteria*, and *Deinococci*), provides no clear insights regarding the intracellular role(s) of the enzyme, but does suggest that they may have been acquired by horizontal transmission. Those few archaea that contain a type IB topoisomerase (psychrophiles *Cenarchaeum symbiosum* and *Nitrosopumilus maritimus*, Table 1) curiously lack type IA topoisomerases of any kind, suggesting that there must be some way in these organisms for the IB enzyme to cooperate with the DNA gyrase to set the net intracellular supercoiling level.

Hyperthermophilic bacteria

All hyperthermophilic eubacteria and archaeobacteria possess a reverse gyrase that actively maintains positive supercoils in the chromosomal DNA. It seems that this positive supercoiling is necessary to stabilize the DNA helix against denaturation at the high growth temperatures of these organisms. The mechanism

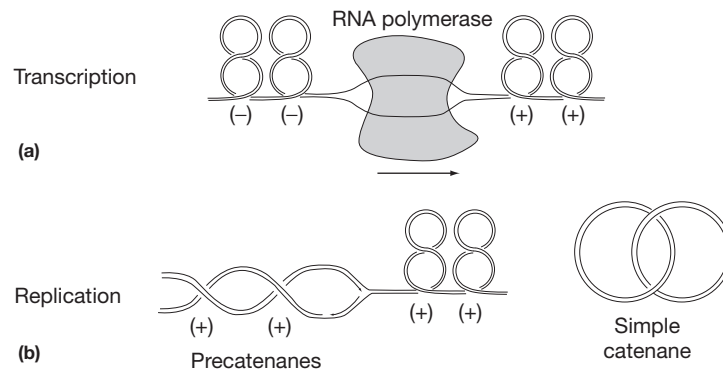


Figure 5 Topological transformations of DNA during transcription and replication. (a) The generation of positive supercoils (+) in front of and negative supercoils (–) behind a translocating RNA polymerase during transcription are depicted. (b) Replication fork movement in a closed domain results in an overwinding of the DNA ahead of the replication fork and the interwinding of the two daughter helices to form precatenanes behind the replication fork, as shown in (b). Any precatenanes remaining at the end of replication for a circular replicon will result in the catenation of the two daughter circular molecules. A simple catenane with a single interlink is also shown.

for preventing excess positive supercoiling is not known, but it is likely that a type II enzyme that can relax positive supercoils (DNA gyrase or the archaeal topoisomerase VI) counteracts the effects of reverse gyrase to set the final steady-state level of positive supercoiling.

Transcription

During transcription, the movement of RNA polymerase along a DNA that is rotationally fixed transiently generates positive supercoils in front of the translocating polymerase and negative supercoils behind the polymerase (Figure 5(a)). The type IA enzymes present in nearly all organisms relax the negative supercoils that accompany transcription, but the mechanism for the removal of the positive supercoils depends on the organism. In prokaryotes, positive supercoils are relaxed by a type II enzyme such as DNA gyrase or in the archaea by topoisomerase VI, and, in the case of select eubacteria, probably by a combination of a type II enzyme and the type IB topoisomerase.

In eukaryotes, it is likely that the type IB DNA topoisomerase I relaxes the supercoils of both signs associated with transcription. Based on the known complete genome sequences, most eukaryotes, but not fungi or *Caenorhabditis elegans*, contain two distinct genes for type IB enzymes, one for the nuclear enzyme and a second for an enzyme that is imported into the mitochondrion to presumably function in transcription. In those eukaryotic organisms with only a single type IB gene, topoisomerase I is likely targeted to both organelles.

DNA Replication

As the two DNA strands are separated during DNA replication, the DNA helix in front of the replication fork becomes at least transiently overwound or positively supercoiled (Figure 5(b)). This overwinding of the helix has been shown to be at least partially transmitted to the region behind the replication fork to cause an interwinding of the two daughter helices. These interwindings are referred to as precatenanes (Figure 5(b)), since in a circular replicon, if any remain at the end of replication, the two daughter circular molecules will be catenated. Resolution of the overwound structure of a replicating

chromosome can be accomplished by relaxing positive supercoils in front of the fork or by decatenating (or unlinking) the precatenanes behind the moving fork, or both.

As in transcription, positive supercoils can be removed in prokaryotes by a type II topoisomerase such as the DNA gyrase, the archaeal topoisomerase VI, or, for those eubacteria and archaea that have it, topoisomerase IB. Precatenanes can be resolved by a type II topoisomerase, most notably topoisomerase IV in eubacteria or topoisomerase VI in archaea. As long as gaps exist during discontinuous DNA synthesis, precatenanes can also be unlinked in some eubacteria either by topoisomerase IB or by the potent type IA decatenating enzyme, topoisomerase III. Segregation of true catenanes lacking nicks or gaps in either of the two strands can only be accomplished by a type II enzyme.

In eukaryotes, the type IB topoisomerase relaxes the positive supercoils ahead of the replication fork and, during synthesis, can probably decatenate the precatenanes by acting at gaps in the DNA. However, most of the precatenanes and terminally interlinked or catenated structures are likely resolved by a type II topoisomerase. In mitochondria, replication-associated positive supercoils are relaxed by a type IB topoisomerase as described previously for transcription. A mitochondrial type II topoisomerase and a type IA enzyme (topoisomerase III) are likely involved in unlinking precatenanes and catenated circular DNAs that occur during mitochondrial DNA replication.

See also: **Molecular Biology: DNA Supercoiling; DNA Topoisomerases: Type II.**

Further Reading

- Alexandrov AI, Cozzarelli NR, Holmes VF, et al. (1999) Mechanisms of separation of the complementary strands of DNA during replication. *Genetica* 106: 131–140.
- Champoux JJ (2001) DNA topoisomerases: Structure, function, and mechanism. *Annual Review of Biochemistry* 70: 369–413.
- Forterre P, Gribaldo S, Gadelle D, and Serre M-C (2007) Origin and evolution of DNA topoisomerases. *Biochimie* 89: 427–446.
- Wang JC (1996) DNA topoisomerases. *Annual Review of Biochemistry* 65: 635–692.
- Wang JC (2002) Cellular roles of DNA topoisomerases: A molecular perspective. *Nature Reviews Molecular Cell Biology* 3: 430–440.

DNA Topoisomerases: Type II

A C Gentry and N Osheroff, Vanderbilt University School of Medicine, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by R. Velez-Cruz and N. Osheroff, volume 1, pp. 806–811, © 2004, Elsevier Inc.

Glossary

Adenosine triphosphate (ATP) A nucleotide cofactor that supplies energy for many enzymatic processes.

Cell cycle Term used to describe the cycle by which a cell grows, replicates its genome, and divides. The cell cycle is divided into four distinct phases: G₁, a growth phase; S, the phase in which the cell duplicates (i.e., synthesizes) its genetic material; G₂, a second growth phase in which the cell prepares to divide; and M, the phase in which the cell divides (i.e., mitosis).

DNA recombination The process by which the cell reorganizes its genetic material in order to repair certain forms of DNA damage (including double-stranded DNA breaks) or promote genetic diversity.

DNA replication The process by which the cell duplicates (i.e., synthesizes) its genetic material.

DNA supercoiling The underwinding (i.e., negative supercoiling) or overwinding (i.e., positive supercoiling) of the genetic material.

Gram staining Eubacteria can be distinguished on the basis of the Gram staining test. Species that do not stain with crystal violet dye are classified as Gram negative and those that do stain with this dye are classified as Gram positive.

Topoisomerase II poison A compound that increases levels of covalent topoisomerase II-cleaved DNA complexes.

Topology A field of mathematics that deals with relationships that are not altered by elastic deformation.

Transcription The process by which the cell expresses its genetic material; the generation of messenger RNAs from a DNA template.

DNA Topology

Topological properties of DNA are defined as those that cannot be altered without breaking one or both strands of the double helix. Since DNA is comprised of two interwound nucleic acid strands and the genomes of all known organisms are very long or circular (or both), two distinct topological issues arise as a result of the double helical structure of the genetic material. Proliferating cells must be able to cope with both of these topological issues in order to survive.

The first issue is related to the torsional stress on the double helix. DNA from all species of eubacteria and eukaryotes is globally underwound ~6%. DNA under torsional stress is termed 'supercoiled', because underwound (or overwound) molecules writhe about themselves to form superhelical twists. Underwound and overwound molecules are called negatively and positively supercoiled, respectively. Negative supercoiling puts energy into the genetic material and makes it easier to separate the two strands of the double helix for replication and transcription. Thus, DNA underwinding increases the rates of these two fundamental processes. By contrast, the movement of DNA tracking systems (such as replication forks and transcription complexes) through the double helix locally overwinds the DNA ahead of their actions. Since overwinding makes it much harder to pull apart the double helix, it blocks many essential cellular processes.

The second issue is related to the extreme length of genomic DNA. Nucleic acid knots (intramolecular) and tangles (intermolecular) are formed routinely during a variety of ongoing cellular processes including DNA recombination and replication. DNA knots impair the ability to separate the two strands of the double helix and reduce the tensile strength of the

genetic material. DNA tangles must be resolved in order for daughter chromosomes to segregate properly during meiosis and mitosis. Consequently, these two topological structures can be lethal to cells if they are not resolved.

DNA Topoisomerases

In order to maintain the appropriate level of DNA supercoiling and remove knots and tangles from the genetic material, cells encode enzymes known as topoisomerases. Topoisomerases modulate these topological structures by creating transient breaks in the DNA backbone. There are two classes of topoisomerases that can be distinguished by the number of DNA strands that they cleave during their catalytic cycles. Type I topoisomerases alter DNA topology by creating a transient single-stranded break in the genetic material and facilitating controlled rotation of the double helix about (or passage of the opposite strand through) the nick. Type II topoisomerases act by generating a transient double-stranded break, through which they pass a separate intact double helix. To maintain genomic integrity during DNA cleavage, topoisomerases form covalent links between active site tyrosyl residues and the newly generated DNA termini. These covalent protein-cleaved DNA complexes, known as 'cleavage complexes', are hallmarks of all topoisomerases irrespective of enzyme classification. Since type I topoisomerases create single-stranded breaks in the genetic material, they can regulate levels of DNA supercoiling. However, since type II topoisomerases generate double-stranded breaks in the DNA backbone, they can resolve knots and tangles in addition to removing torsional stress from the genetic material.

Type II topoisomerases, which are the subject of this article, are essential to all prokaryotic and eukaryotic organisms. They are highly conserved among species, and the eukaryotic enzymes appear to be direct descendents of ancestral bacterial proteins.

Type II Topoisomerases

There are two classes of type II topoisomerases, type IIA and type IIB, which are defined on the basis of homology. With one exception, type II topoisomerases are denoted by even numerals. Bacteria encode two type IIA enzymes, DNA gyrase and topoisomerase IV. Eukaryotes, by contrast, encode only one type IIA enzyme, topoisomerase II. It should be noted, however, that vertebrate species express two closely related isoforms of the enzyme, topoisomerase II α and topoisomerase II β . Archaea and plants express the only known type IIB topoisomerase, topoisomerase VI. These individual enzymes will be discussed in detail in later sections.

Enzyme Mechanism

All type II topoisomerases interconvert different topological forms of DNA by the double-stranded DNA passage reaction depicted in [Figure 1](#), which uses eukaryotic topoisomerase II as a representative enzyme. Briefly, type II enzymes (1) bind two DNA segments, (2) create a double-stranded break in one of the segments, (3) translocate the other DNA segment through the cleaved double helix, (4) rejoin (i.e., ligate) the cleaved DNA, (5) release the translocated segment through a gate in the protein, and (6) close the protein gate, regaining the ability

to start a new round of catalysis. The scissile bonds on the two strands of the double helix that are cut by type II topoisomerases are staggered. Type IIA enzymes generate cleaved DNA molecules that contain four-base single-stranded cohesive ends at their 5'-termini, while type IIB enzymes generate molecules with two-base single-stranded ends. All of the type II topoisomerases covalently attach to the newly generated 5'-termini during the cleavage event.

Type II enzymes require two cofactors in order to carry out their catalytic double-stranded DNA passage reactions. First, they need a divalent cation for all steps beyond enzyme-DNA binding. Magnesium(II) appears to be the divalent cation that the enzymes use *in vivo*. Second, they use the energy of adenosine triphosphate (ATP) to drive the overall DNA strand passage reaction. Type IIA topoisomerases do not require ATP for either DNA cleavage or ligation. However, the binding of this nucleoside triphosphate triggers DNA translocation, and its hydrolysis to adenosine diphosphate (ADP) and inorganic phosphate (Pi) is necessary for enzyme recycling. Normally, type II enzymes bind two molecules of ATP. Although hydrolysis of the cofactor is not a prerequisite for the strand passage event, it appears that this step proceeds more rapidly if it is preceded by hydrolysis of one of the bound ATP molecules. Type IIB enzymes use ATP in a similar fashion with one important distinction: they require ATP binding as a prerequisite for DNA cleavage.

Enzyme Domain Structures

Type II topoisomerases share a number of common structural motifs that are shown in [Figure 2](#). The founding type II enzyme, bacterial DNA gyrase, is comprised of two distinct

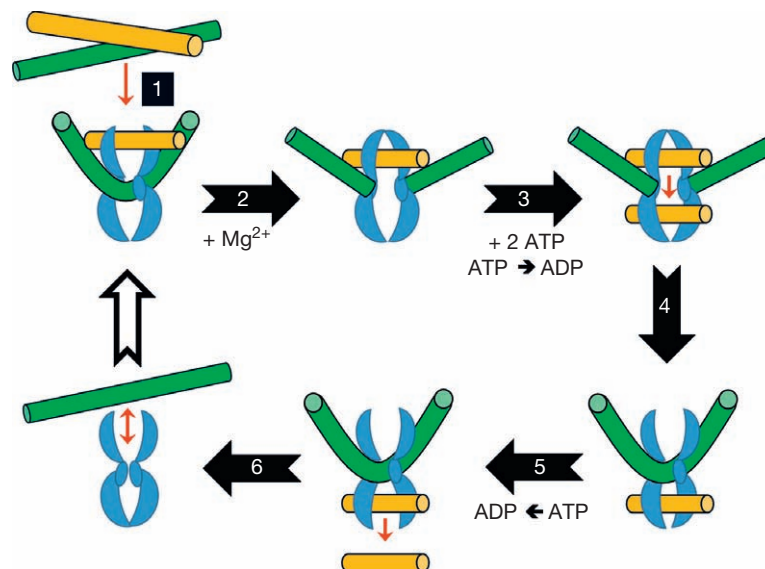


Figure 1 The double-stranded DNA passage reaction of topoisomerase II can be separated into six discrete steps. (1) DNA binding to regions of helix-helix juxtaposition. (2) Double-stranded DNA cleavage (i.e., formation of the cleavage complex), which requires the presence of Mg^{2+} (physiologically) or other divalent metal ions. (3) Double-stranded DNA passage through the DNA gate generated by cleavage. This reaction requires the binding of two ATP molecules, and strand passage proceeds more rapidly if one of the two ATP molecules is hydrolyzed. (4) Ligation of the cleaved DNA. (5) Hydrolysis of the second ATP molecule, which allows release of the DNA through a C-terminal gate in the protein and promotes (6) enzyme turnover, which allows the enzyme to initiate a new round of catalysis. The protein (shown in blue) is based on crystallographic studies of yeast topoisomerase II. Modeled DNA helices are shown in green (horizontal) and orange (coming out of the plane of the paper).

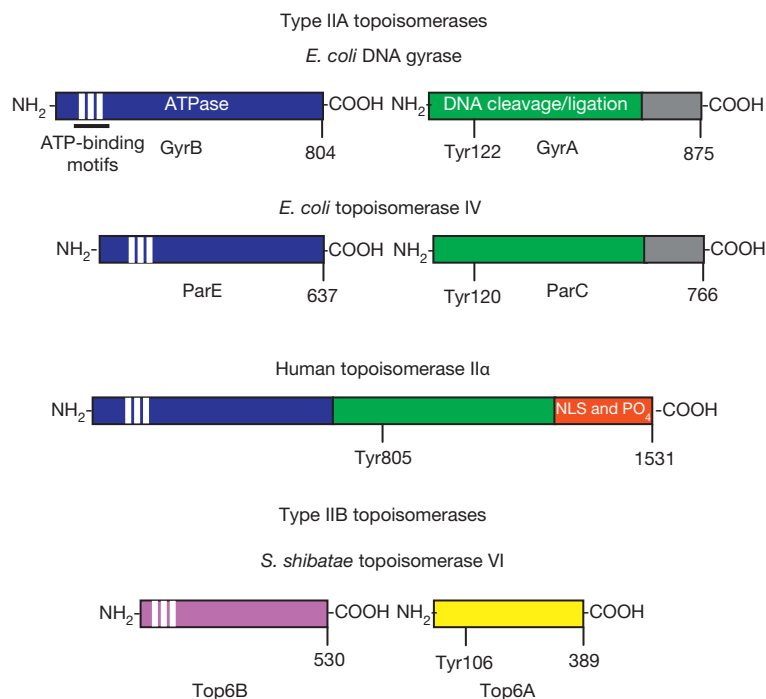


Figure 2 Domain structures of type II topoisomerases. The domain structures of three type IIA topoisomerases, bacterial (*Escherichia coli*) DNA gyrase and topoisomerase IV, and human topoisomerase II α are shown. Regions of homology among the enzymes are indicated by colors. The N-terminal (i.e., GyrB) homology domains (blue) contain the regions responsible for ATP binding and hydrolysis. The vertical white stripes represent the three conserved motifs of the 'Bergerat fold' that define the ATP-binding domain. The central (i.e., GyrA) homology domains (green) contain the active site tyrosyl residue that forms the covalent bond with DNA during scission. For human topoisomerase II α , the variable C-terminal domain (red) contains nuclear localization sequences (NLS) and phosphorylation sites (PO₄). The Top6A (purple) and Top6B (yellow) subunits of the archaeal type IIB topoisomerase, *Sulfolobus shibatae* topoisomerase VI, are shown for comparison. The active site tyrosine residue is indicated for each enzyme.

subunits, GyrA and GyrB (molecular mass ≈ 96 and 88 kDa, respectively) and acts as an A₂B₂ tetramer. GyrA contains the active site tyrosyl residue that forms the covalent bond with DNA during scission, and GyrB contains consensus sequences for ATP binding. Bacterial topoisomerase IV is also an A₂B₂ tetramer comprised of two subunits. In Gram-negative species, these subunits were first identified as proteins required for partitioning, and are designated ParC (molecular mass ≈ 88 kDa) and ParE (molecular mass ≈ 70 kDa), which are homologous to the A and B subunits of DNA gyrase. In Gram-positive species, the subunits of topoisomerase IV are designated as the gyrase-like proteins GrlA and GrlB, respectively.

Eukaryotic type IIA topoisomerases are homologous to the bacterial enzymes. However, the two subunits have fused into a single polypeptide with protomer molecular masses ranging from ~ 160 to 180 kDa (depending on the species). The eukaryotic enzymes function as homodimers. On the basis of amino acid sequence comparisons with bacterial gyrase, each topoisomerase II protomer can be divided into three distinct domains. The N-terminal domain of the enzyme is homologous to the B-subunit of DNA gyrase (GyrB) and contains the binding site for ATP. The central domain is homologous to the A-subunit of DNA gyrase (GyrA) and contains the active site tyrosyl residue. The C-terminal domain is variable and appears to have no corresponding region of homology in DNA gyrase. This region of the eukaryotic enzyme contains nuclear localization sequences as well as amino acid residues that are phosphorylated *in vivo*.

The type IIB enzyme, topoisomerase VI, has two subunits, Top6A and Top6B (molecular masses ≈ 47 and 60 kDa, respectively), and is arranged as an A₂B₂ tetramer. Both subunits are considerably smaller than those of bacterial DNA gyrase or topoisomerase IV. Although short regions of Top6B surrounding the ATP-binding domain are homologous to portions of GyrB, and Top6A contains an active site tyrosine that is required for DNA cleavage, the primary structure of topoisomerase VI displays little similarity to the type IIA enzymes.

Prokaryotic Type IIA Topoisomerases

Eubacteria contain two distinct type II topoisomerases, DNA gyrase and topoisomerase IV. In addition, some archaeal species encode DNA gyrase.

DNA Gyrase

DNA gyrase was discovered in 1976. It was the first type II topoisomerase to be described and is the only one to retain its historical name. In contrast to all other type II topoisomerases, DNA gyrase is the only enzyme that is capable of actively underwinding (i.e., negatively supercoiling) the double helix. It accomplishes this feat by wrapping DNA around itself in a right-handed fashion and carrying out its strand-passage reaction in a unidirectional manner. The C-terminal domain of the GyrA subunit is required for the enzyme to carry out this unique function.

The negative supercoiling activity of DNA gyrase far exceeds the ability of the enzyme to remove either knots or tangles from the genetic material. Consequently, the major physiological roles of DNA gyrase stem directly from its ability to underwind the double helix. DNA gyrase plays a crucial role in opening DNA replication origins and removing positive supercoils that accumulate in front of replication forks and transcription complexes. In addition, this enzyme works in conjunction with the ω protein (a type I topoisomerase that removes negative supercoils from the double helix), to maintain the global balance of DNA supercoiling in bacterial cells.

Topoisomerase IV

It had been known for several years that the ParC and ParE proteins were necessary for proper chromosome partitioning in bacteria. However, it was not discovered until 1990 that these two subunits together constituted a type II topoisomerase.

The catalytic properties of topoisomerase IV can be distinguished from those of DNA gyrase in two important ways. First, while topoisomerase IV can remove positive and negative superhelical twists from DNA, it cannot actively underwind the double helix. Second, the ability of topoisomerase IV to resolve DNA knots and tangles is considerably better than that of DNA gyrase. Because of these differences, the physiological roles of topoisomerase IV are distinct from those of DNA gyrase. The primary cellular functions of topoisomerase IV are to unlink daughter chromosomes following DNA replication and to resolve DNA knots that are formed during recombination. Recently, it was found that topoisomerase IV removes positive supercoils from DNA more efficiently than it does negative supercoils. This ability, which requires sequences in the variable C-terminal domain, has led to speculation that the enzyme also may act ahead of DNA tracking systems to help alleviate overwinding of the double helix. However, the precise role of topoisomerase IV in this process has yet to be defined.

Eukaryotic Type IIA Topoisomerases

The eukaryotic type IIA enzyme, topoisomerase II, was discovered in 1980. Topoisomerase II can remove superhelical twists from the double helix and can resolve DNA knots and tangles. Eukaryotic species, such as yeast and *Drosophila*, encode only a single type II topoisomerase (i.e., topoisomerase II). However, as discussed above, vertebrates contain two isoforms, topoisomerase II α and topoisomerase II β . These isoforms share extensive amino acid sequence identity (~70%) but are encoded by separate genes (located at chromosomal bands 17q21–22 and 3p24 in humans, respectively). Topoisomerase II α and topoisomerase II β can also be distinguished by their protomer molecular masses (~170 and ~180 kDa, respectively).

Topoisomerase II plays a number of essential roles in eukaryotic cells and participates in virtually every major process that involves the genetic material. It unlinks daughter chromosomes that are tangled following replication and resolves DNA knots that are formed during recombination. It also helps to remove the positive DNA supercoils that are generated ahead of replication forks and transcription complexes. Topoisomerase II is required for proper chromosome condensation, cohesion, and segregation, and appears to play roles in centromere function and chromatin remodeling.

Finally, topoisomerase II is important for the maintenance of proper chromosome organization and structure, and is the major nonhistone protein of the mitotic chromosome scaffold and the interphase nuclear matrix.

It is not obvious why vertebrate species encode two distinct topoisomerase II isoforms. Enzymological differences between topoisomerase II α and topoisomerase II β are subtle and relationships between these isoforms are not well defined. The only major enzymological characteristic that distinguishes topoisomerase II α and topoisomerase II β is the fact that the α isoform removes positive DNA supercoils ~10-fold faster than it does negative, while the β isoform removes both at similar rates. As with topoisomerase IV, this ability of topoisomerase II α is embodied in the C-terminal domain of the protein.

Topoisomerase II α and topoisomerase II β have distinct patterns of expression and separate cellular functions. Topoisomerase II α is essential for the survival of proliferating cells and is regulated over both cell and growth cycles. Enzyme levels increase throughout S-phase of the cell cycle and peak at the G₂/M boundary. Topoisomerase II α is found almost exclusively in proliferating tissues, is associated with replication forks, and remains tightly bound to chromosomes during mitosis. Thus, it is believed to be the isoform that functions in growth-related processes, such as DNA replication and chromosome segregation.

Topoisomerase II β is dispensable at the cellular level but is required for proper neural development in mice. In contrast to the α isoform, the concentration of topoisomerase II β is independent of the cell cycle and this isoform is found in most cell types regardless of proliferation status. Topoisomerase II β dissociates from chromosomes during mitosis and cannot compensate for the loss of topoisomerase II α in mammalian cells. Although the physiological functions of the β isoform have yet to be fully defined, recent evidence indicates involvement in the transcription of hormonally or developmentally regulated genes.

Archaeal Type IIB Topoisomerase

In 1997, a novel type IIB topoisomerase, topoisomerase VI, was discovered in hyperthermophilic archaeal species. This enzyme has been designated as the first member of the topoisomerase IIB subfamily due to its lack of homology to previously identified type II enzymes. Topoisomerase VI is also found in some plant species.

Archaeal topoisomerase VI can relax positively and negatively supercoiled DNA and can untangle interwound double-stranded DNA molecules. Although the physiological functions of topoisomerase VI have yet to be determined, the enzyme is believed to play a role in unlinking daughter chromosomes following replication in archaeal cells.

With the exception of plants, no Top6B homolog has been identified in eukaryotic species. However, a Top6A homolog, Spo11, has been found in eukaryotes ranging from yeast to humans. Spo11 generates the double-stranded DNA breaks that initiate meiotic recombination. Like its topoisomerase relatives, Spo11 forms a covalent bond between an active site tyrosyl residue and the 5'-DNA termini generated by its scission reaction. At the present time, there is no evidence that Spo11 displays topoisomerase (i.e., DNA strand passage) activity *in vivo*.

Type IIA Topoisomerases as Therapeutic Targets

In addition to their varied and critical physiological functions, the type IIA topoisomerases are targets for some of the most active antibacterial and anticancer drugs in clinical use. These agents do not act by robbing cells of essential topoisomerase activity. Rather, drugs that target type II topoisomerases kill cells by dramatically increasing the concentration of covalent enzyme–DNA cleavage complexes that are requisite intermediates formed during the double-stranded DNA passage reaction. Normally, cleavage complexes are present at low steady-state levels and are tolerated by cells. However, conditions that significantly increase either their concentration or their lifetime trigger numerous mutagenic events.

The potential lethality of cleavage complexes rises substantially when DNA-tracking enzymes, such as polymerases or helicases, attempt to traverse the covalently bound topoisomerase roadblock in the genetic material. Such an action disrupts cleavage complexes and converts transient enzyme-mediated DNA breaks to permanent DNA breaks. These permanent breaks in the genome trigger the generation of chromosomal insertions, deletions, translocations, and other aberrations and, when present in sufficient numbers, initiate a series of events that culminates in cell death. Since the drugs that target type II topoisomerases convert these essential enzymes to potent cellular toxins that fragment the genome, they are referred to as topoisomerase poisons to distinguish them from drugs that act as catalytic inhibitors.

Antibacterial Drugs

DNA gyrase and topoisomerase IV are the targets for quinolone-based antibacterial agents (Figure 3). Quinolones are the most active and broad-spectrum oral antibacterial drugs currently in clinical use. Drugs such as ciprofloxacin are prescribed routinely for a wide variety of Gram-negative bacterial infections, including gastrointestinal tract and bone and joint infections. Ciprofloxacin is also used to treat a number of sexually transmitted diseases as well as anthrax. Newly developed quinolones, such as levofloxacin, display significant efficacy against Gram-positive bacterial infections, such as those found in the respiratory tract.

DNA gyrase is the primary cytotoxic target of quinolones in Gram-negative bacteria. However, topoisomerase IV appears to be the more important target for many of these drugs in most Gram-positive species.

Anticancer Drugs

At the present time, six topoisomerase II-targeted anticancer agents (Figure 3) are approved for use in the United States. Additional drugs that target the type II enzyme are used worldwide. Drugs such as etoposide and doxorubicin are front-line therapy for breast and lung cancers, as well as a variety of leukemias, lymphomas, and germ-line malignancies.

Approximately one-half of all cancer chemotherapy regimens contain drugs targeted to topoisomerase II. Moreover, every form of cancer that can be cured by systemic chemotherapy is treated with these agents.

Due to the high concentration of topoisomerase II α in rapidly proliferating cells, this isoform probably is the more important

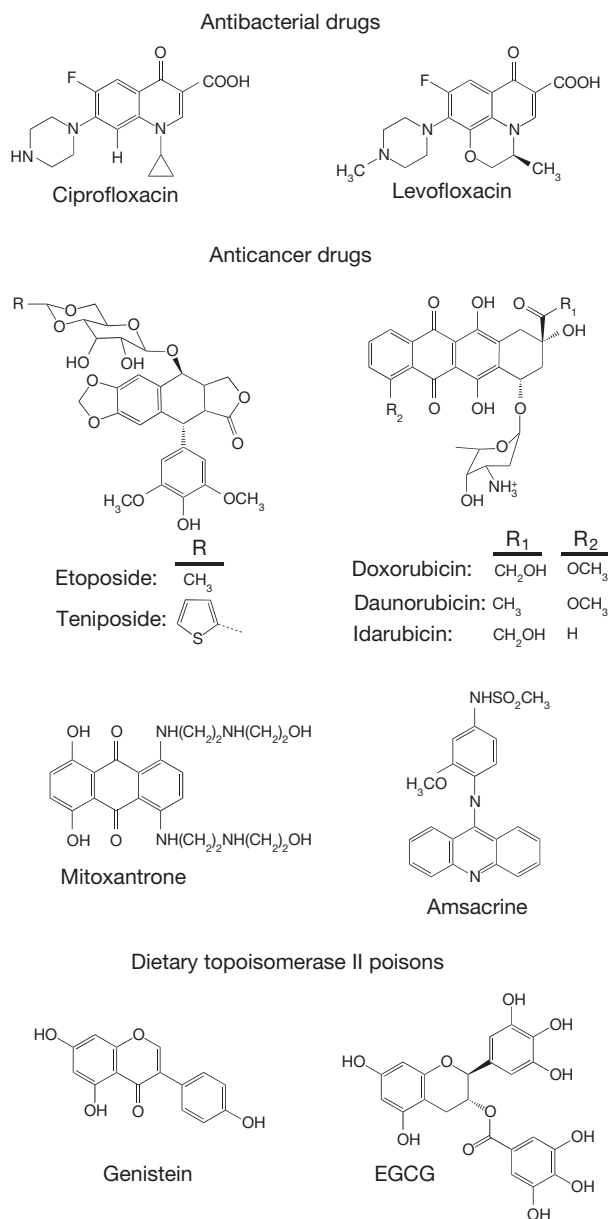


Figure 3 Structures of selected antibacterial drugs targeted to DNA gyrase and topoisomerase IV, anticancer drugs targeted to topoisomerase II, and dietary topoisomerase II poisons.

target of anticancer therapy. However, circumstantial evidence suggests that the β isoform also contributes to drug efficacy.

Dietary Topoisomerase II Poisons

The most prevalent dietary topoisomerase II poisons are the bioflavonoids (i.e., phytoestrogens) (Figure 3). This diverse group of polyphenolic compounds are components of many fruits, vegetables, legumes, and plant leaves. Bioflavonoids are an integral component of the human diet and represent the most abundant natural source of antioxidants. It is believed that the dietary intake of these compounds provides a number of health benefits, including protection against cancer, cardiovascular disease, osteoporosis, and inflammation. The mechanistic basis for the physiological actions of phytoestrogens is

complex, and they have a variety of effects on human cells. However, several common bioflavonoids, including genistein and (–)-epigallocatechin gallate (EGCG), are potent topoisomerase II poisons. Genistein is abundant in soy products and is believed to contribute to the lower incidence of specific types of cancers in the Pacific Rim. EGCG is the most abundant and biologically active polyphenol in green tea, and many of the therapeutic benefits of the beverage have been attributed to this compound.

See also: **Molecular Biology: DNA Supercoiling; DNA Topoisomerases: Type I.**

Further Reading

- Anderson VE and Osheroff N (2001) Type II topoisomerases as targets for quinolone antibacterials: Turning Dr. Jekyll into Mr. Hyde. *Current Pharmaceutical Design* 7: 337–353.
- Champoux JJ (2001) DNA Topoisomerases: Structure, function, and mechanism. *Annual Review of Biochemistry* 70: 369–413.
- Corbett KD and Berger JM (2004) Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases. *Annual Review of Biophysics and Biomolecular Structure* 33: 95–118.
- Deweese JE and Osheroff N (2009) The DNA cleavage reaction of topoisomerase II: Wolf in sheep's clothing. *Nucleic Acids Research* 37: 738–748.
- Drlica K, Malik M, Kerns RJ, and Zhao X (2008) Quinolone-mediated bacterial death. *Antimicrobial Agents and Chemotherapy* 52: 385–392.
- Forterre P and Gadelle D (2009) Phylogenomics of DNA topoisomerases: Their origin and putative roles in the emergence of modern organisms. *Nucleic Acids Research* 37: 679–692.
- Li TK and Liu LF (2001) Tumor cell death induced by topoisomerase-targeting drugs. *Annual Review of Pharmacology and Toxicology* 41: 53–77.
- McClendon AK and Osheroff N (2007) DNA topoisomerase II, genotoxicity, and cancer. *Mutation Research* 623: 83–97.
- Nitiss JL (2009a) DNA topoisomerase II and its growing repertoire of biological functions. *Nature Reviews: Cancer* 9: 327–337.
- Nitiss JL (2009b) DNA topoisomerase II in cancer chemotherapy. *Nature Reviews: Cancer* 9: 338–350.
- Wang JC (2002) Cellular roles of DNA topoisomerases: A molecular perspective. *Nature Reviews: Molecular Cell Biology* 3: 430–440.
- Wang JC (2009) *Untangling the Double Helix: DNA Entanglement and the Action of the DNA Topoisomerases*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Dopamine Receptors

K A Neve, Portland Veterans Affairs Medical Center, Portland, OR, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Affinity The term for how tightly a drug binds to a receptor, expressed as the concentration of drug that occupies half of the available receptors. High affinity means that a drug will bind at very low concentrations.

Agonist A drug that binds to and activates a receptor. Dopamine is the endogenous agonist for dopamine receptors.

Antagonist A drug that binds to a receptor without activating it and, thus, prevents the binding of an agonist.

Receptor A protein, usually associated with the cell membrane, that binds a neurotransmitter or hormone and initiates a biological response.

Second messenger A chemical signal elicited by an activated receptor that transmits information within a cell. The extracellular neurotransmitter is the first messenger, and whereas some receptors transduce information about the neurotransmitter concentration into an electrical response, other receptors convert that information into altered abundance of an intracellular second messenger such as cyclic adenosine monophosphate (AMP).

Introduction

Nobel prize-winning work by Carlsson and colleagues in 1958 first demonstrated that dopamine is a neurotransmitter, rather than being only an intermediate step in the synthesis of norepinephrine, and a neurotransmitter that plays an important role in regulating motor behavior. The first pharmacological evidence for the existence of specialized receptors for dopamine was obtained 5 years later, but it was not until 1972 that the identification of dopamine-stimulated adenylyl cyclase, the enzyme that converts adenosine triphosphate (ATP) into the second messenger cyclic adenosine monophosphate (AMP), made possible the biochemical characterization of a dopamine receptor. Later that decade, further pharmacological and biochemical characterization of dopamine receptors, including their division into D1 and D2 subtypes, was made possible by the development of radioisotopically labeled ligands, or radioligands, that bind to dopamine receptors with high affinity and selectivity. What might be considered the modern era of dopamine receptor research began in 1988 with the determination of the DNA sequence and predicted protein sequence of the D2 dopamine receptor and the subsequent identification of additional dopamine receptor subtypes.

Dopamine Receptor Subtypes

Virtually all neurotransmitters and hormones activate multiple receptor subtypes – distinct classes of receptors that share relatively high affinity for a single neurotransmitter but that differ in sequence, affinity for drugs, signaling, and distribution. It is common for the distribution of receptor subtypes to overlap, so that one organ or brain region may have multiple receptor subtypes for any neurotransmitter or hormone that is present.

D1-Like and D2-Like Receptor Subfamilies

Prior to the molecular cloning of dopamine receptors, they were divided into D1 and D2 subtypes on the basis of several criteria. D1 and D2 receptors have distinct pharmacological profiles; in particular, D2 receptors have high affinity for benzamide and butyrophenone antagonists such as sulpiride and spiperone, respectively, while D1 receptors have high affinity for benzazepine antagonists such as SCH23390. D1 receptors stimulate adenylyl cyclase and cyclic AMP accumulation, while D2 receptors inhibit the enzyme. Furthermore, the D1 and D2 receptors are physically distinct. For example, there are tissues such as the anterior and intermediate pituitary gland that have an abundance of D2 receptors but no D1 receptors. Finally, it is possible to cause selective damage to cell bodies in the substantia nigra or in the neostriatum of the rat brain, using lesions that spare axons and axon terminals, and with these lesions to cause a preferential reduction in the abundance of D2 receptors (substantia nigra) or D1 receptors (neostriatum), thus demonstrating the differential localization of the subtypes on cell bodies and axon terminals in these brain regions. The molecular cloning of additional subtypes of dopamine receptors revealed the existence of two receptor subfamilies, in each of which the subtypes share D1-like or D2-like characteristics.

Molecular Cloning of Dopamine Receptor Subtypes

Following the cloning of DNA encoding receptors determined to be identical to the pharmacologically characterized D1 and D2 receptors, several additional subtypes were identified in 1990–91. DNA sequence information was a new tool for classification of receptors, in that receptors belonging to the same subfamily are more similar in DNA and protein sequence. Thus, there are many amino acids that are common in D1 and D5 receptors (Figure 1(a)), but fewer that are shared among all the dopamine receptor subtypes (Figure 1(b)).

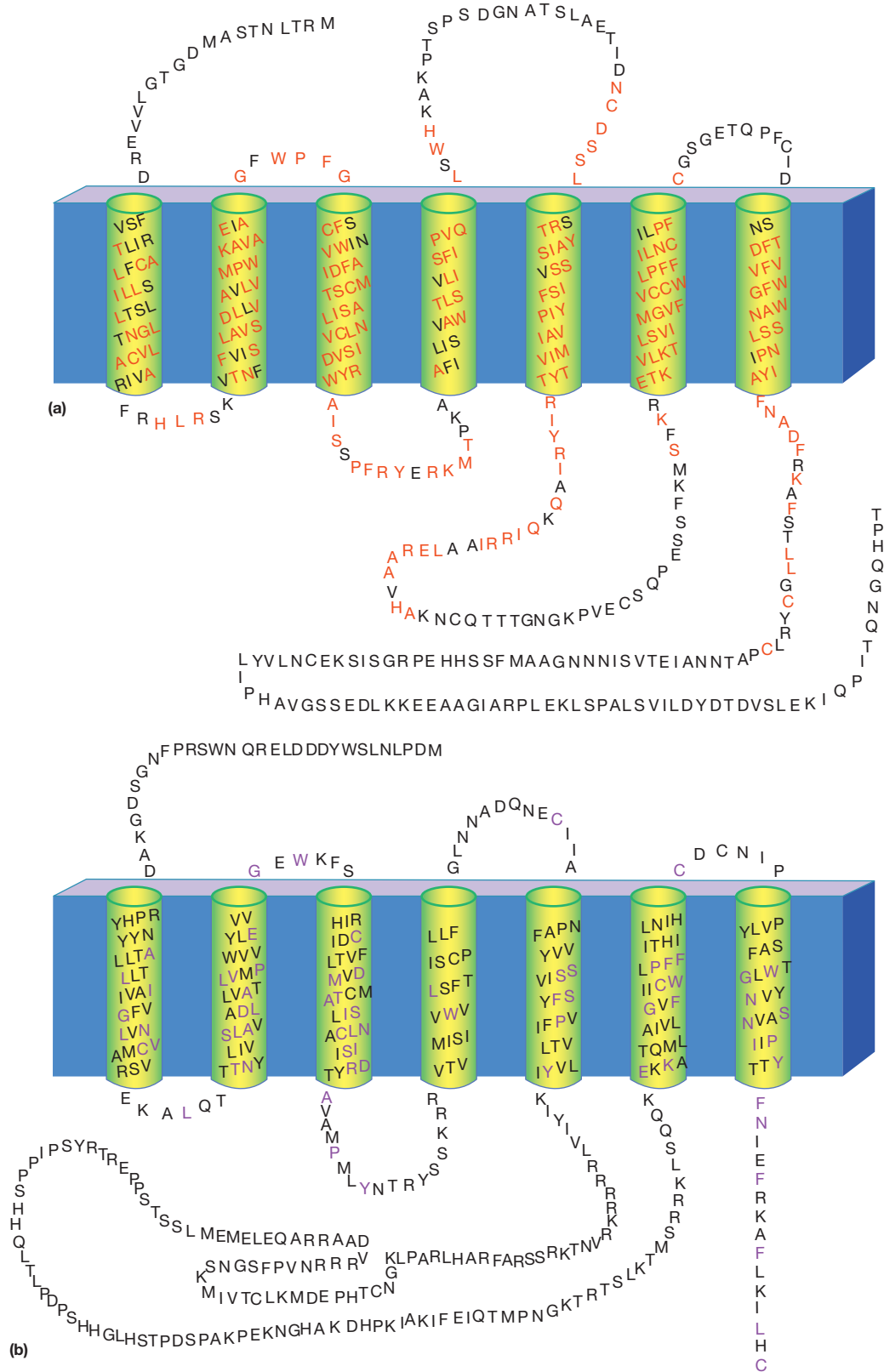


Figure 1 Predicted amino acid sequence and membrane-spanning segments of the D1 (a) and D2 (b) dopamine receptor subtypes. The figure is oriented so that the extracellular regions of each receptor are above the shaded transmembrane helices and the intracellular regions are below. Amino acids colored red in panel (a) are shared between D1 and D5 receptors, whereas amino acids colored purple in panel (b) are shared among all of the dopamine receptor subtypes.

D1 and D5 receptors are both D1-like receptors, whereas D2, D3, and D4 receptors belong to the D2-like receptor subfamily in terms of sequence identity and affinity for drugs.

Structural Characteristics of Dopamine Receptors

Most hormone and neurotransmitter receptors are integral membrane proteins that span the cell membrane at least once, so that a portion of the receptor lies outside the cell and another portion lies inside. Dopamine receptors all belong to the superfamily of proteins called seven-transmembrane receptors because they traverse the membrane 7 times, serpentine receptors because of the manner in which they wind back and forth across the membrane, or G protein-coupled receptors because most of the effects of neurotransmitter binding to the receptors are mediated by activation of G proteins. The superfamily of seven-transmembrane receptors includes receptors for light, odors, and calcium, in addition to most neurotransmitters. Although all dopamine receptors have seven membrane-spanning regions with four intracellular domains (loops 1–3 and the carboxy terminus) and four extracellular domains (the amino terminal extension and extracellular loops 1–3), the D1-like receptors have relatively short third intracellular loops and relatively long carboxy termini compared to the D2-like receptors (Figure 1).

As illustrated in Figure 1, the presence of specific amino acid residues at certain locations across receptor subtypes, or amino acid sequence homology, is highest within the membrane-spanning regions of seven-transmembrane receptors. This is primarily because maintaining the three-dimensional structure of a receptor so that it is relatively quiescent until activated by the binding of a neurotransmitter such as dopamine requires precise packing of the membrane-spanning α -helices and a network of interhelical bonds that constrain the movement of the helices. Random mutations in these regions would be often deleterious and selected against because of the relatively high likelihood of disrupting helix packing or interhelical bonds. Within a family of receptors for one neurotransmitter or for several structurally related neurotransmitters, such as the catecholamines dopamine and norepinephrine, sequence homology also reflects the conservation of amino acids that bind the neurotransmitter; for small molecule neurotransmitters such as dopamine, these amino acids are typically in the membrane-spanning helices, which form a water-accessible binding pocket within the membrane lipid bilayer.

Although depicted as single receptors in Figure 1, the functioning holoreceptor is probably at least a dimer of two dopamine receptors. Higher-order oligomers, such as tetramers, are also likely to exist. Oligomers may be homomers composed of two or more copies of the same subunit or heteromers incorporating more than one receptor subtype. Some oligomers, such as the D1:D2 oligomer, have novel properties such as unique pharmacological profiles and distinct mechanisms of signaling.

Dopamine Receptor Subtype-Selective Drugs

D1 and D2 receptor-selective drugs have been very useful for differentiating between the behavioral and biochemical effects

of D1-like or D2-like receptor stimulation. There are numerous commercially available potent and selective D2-like receptor agonists such as quinpirole and 7-OH DPAT, and antagonists such as spiperone and nemonapride (YM-09151-2). Although there are fewer D1-like receptor-selective drugs, the selective benzazepine antagonist SCH23390 and agonists such as SKF38393 and 6-chloro-APB (SKF82958) have been used to demonstrate that blockade or stimulation of D1-like receptors has significant behavioral consequences.

There is great interest in developing drugs that differentiate between D1 and D5 receptors or among the D2-like receptors, both as research tools and because of the possibility that selective blockade or stimulation of only one subtype will be a more effective treatment for a disease that is currently treated by blockade or stimulation of the D1-like or D2-like subfamily (Table 1). It is still not possible to differentiate pharmacologically between D1 and D5 receptors, two subtypes with very similar sequence in the membrane-spanning helices (Figure 1(a)). On the other hand, the use of antagonists that are highly selective for only one of the three subtypes of the D2-like receptor subfamily, particularly for the D4 receptor, has demonstrated that dopamine binding to D3 or D4 receptors in rats has behavioral effects that are very different from D2 receptor-stimulated locomotor activation. Mice that have null mutations in the gene for one receptor subtype, and thus do not express functional protein for that subtype, have also been very useful both for evaluating the *in vivo* selectivity of putative subtype-selective drugs and for determining different behavioral effects mediated by each receptor subtype.

Dopamine Receptor Signaling

Most of the effects of dopamine receptors on cellular function are mediated by activation of heterotrimeric GTP-binding proteins called G proteins. Binding of dopamine to amino acids within the membrane-spanning helices of a dopamine receptor disrupts the interhelical bonds that hold the receptor in an inactive state. The resulting movement of one or more of the helices exposes amino acids on the cytoplasmic face of the membrane that bind to and activate G proteins. D1-like and D2-like receptors have very different signaling properties as a result of differential selection of G protein subtypes.

G Protein Coupling

Heterotrimeric G proteins are composed of α , β , and γ subunits, and are generally named according to the subtype of $G\alpha$ subunit. D1-like receptors that have been stimulated by dopamine or other D1-receptor agonists activate two G proteins that stimulate adenylate cyclase, $G\alpha_s$ and the closely related G protein, $G\alpha_{olf}$ (Table 1). (The latter G protein was first identified as mediating olfactory responses to receptors for odorants.) D2-like receptors activate the $G\alpha_i$ subtypes, named after their ability to inhibit adenylate cyclase, and the closely related G protein $G\alpha_o$. That these G proteins are substrates for adenosine diphosphate (ADP)-ribosylation by cholera toxin, which persistently activates $G\alpha_s$, or pertussis toxin, which inactivates $G\alpha_{i/o}$, has contributed importantly to determining their role in dopamine receptor signaling.

Table 1 Features of human dopamine receptor subtypes

Type of feature	Dopamine receptor subtype				
	D1	D5	D2	D3	D4
Length (amino acids)	446	477	443 ^a	400	387 ^b
Introns in coding region	No	No	Yes	Yes	Yes
G protein coupling	G α_s /G α_{off}	G α_s	G $\alpha_{i/o}$	G $\alpha_{i/o}$	G $\alpha_{i/o}$
Agonists	SKF38393 6-CI-APB	SKF38393 6-CI-APB	Quinpirole 7-OH-DPAT	Quinpirole 7-OH-DPAT PD128907	Quinpirole 7-OH-DPAT PD168077
Antagonists	SCH23390	SCH23390	L-741,626	GR 103691 U99194	L-741,7141 L-745,870
Radioligands	[³ H]SCH 23390 [¹²⁵ I]SCH23982	[³ H]SCH 23390 [¹²⁵ I]SCH23982	[³ H]spiperone [³ H]YM-09151-2	[³ H]spiperone [³ H]YM-09151-2 [³ H]7-OH-DPAT	[³ H]spiperone [³ H]YM-09151-2
Localization ^c	Caudate-putamen, cerebral cortex, substantia nigra, parathyroid gland	Hippocampus, cerebral cortex	Caudate-putamen, substantia nigra, intermediate and anterior pituitary gland	Nucleus accumbens, olfactory tubercle	Cerebral cortex, hypothalamus, olfactory bulb
Splice variants	None	None	D2 _L /D2 _S	D3nf ^d	None
Allelic variants ^e	None	Leu88→Phe Ala269→Val Pro330→Gln Cys-335→stop Asn351→Asp Ser453→Cys	Val96→Ala Pro310→Ser Ser311→Cys	Ser9→Gly	Gly11→Arg 12 bp repeat in exon 1 21 bp deletion in exon 1 13 bp deletion in exon 1 Val194→Gly 48 bp repeat in exon 3

^aLength of D2_L; D2_S is 414 amino acids long.

^bLength of the variant with two repeats (D4.2) although actual length depends on the number of repeat units, which varies from 2 to 10.

^cSelected tissues and brain regions with relatively high expression of the subtype are listed.

^dD3nf is a frame-shifted variant with D3 receptor sequence only through the fifth membrane-spanning helix. Although shorter-truncated splice variants of D3 mRNA also exist, D3nf is the only one for which expression of the protein has been demonstrated.

^eOnly variants that alter the protein sequence are listed.

Signaling Pathways

As suggested by their activation of G proteins that stimulate adenylate cyclase, most effects of D1-like receptor stimulation are mediated by increased levels of cyclic AMP and cyclic AMP-dependent protein phosphorylation. D1 receptor-stimulated protein phosphorylation alters the activity of other protein kinases, receptors, protein phosphatases, ion channels, and transcription factors. Under some conditions, D1-like receptors also activate G α_q , leading to calcium mobilization. D2-like receptors inhibit the activity of adenylate cyclase via G α_i . D2-like receptors also modulate many other signaling pathways, most of them as a result of the liberation of G-protein $\beta\gamma$ subunits that occurs when seven-transmembrane receptors activate G $\alpha_{i/o}$ proteins. G $\beta\gamma$ -regulated pathways that are regulated by stimulation of D2-like receptors include several forms of adenylate cyclase, ion channels, phospholipase C, and mitogen-activated protein (MAP) kinases. Although the basic mechanism of activation of heterotrimeric G proteins by receptors is probably the same for all receptor-G protein

combinations, for reasons that have not been determined, activation of other types of G proteins such as G α_s does not produce the same G $\beta\gamma$ -mediated signaling that is observed after receptor-mediated activation of G $\alpha_{i/o}$. D2 receptors also signal via a pathway that requires the scaffolding protein arrestin3, which recruits the protein kinase Akt and protein phosphatase 2A, leading to the dephosphorylation and inhibition of Akt.

Distribution of Dopamine Receptors

The D1 and D2 dopamine receptors are overall the most abundant dopamine receptor subtypes. The D1 receptor is expressed most highly in the brain, with lower expression in peripheral tissues such as the parathyroid gland, renal, mesenteric, and coronary vascular beds, and the kidney. Within the brain, the D1 receptor is most abundant in the caudate-putamen and other basal forebrain nuclei such as the nucleus accumbens, and is also found in the cerebral cortex and the substantia nigra

pars reticulata. The D5 receptor is expressed at much lower levels in the same basal forebrain nuclei, the cerebral cortex, and hippocampus. The D2 receptor is expressed most highly in the intermediate pituitary gland, with abundant expression also observed in the anterior pituitary and in a number of brain regions including the caudate-putamen and other nuclei of the basal ganglia, the substantia nigra *pars compacta*, and the cerebral cortex. The D4 receptor is expressed in most of the same forebrain regions as the D2 receptor and in the cerebral cortex, albeit at a lower level. Whereas the D2 receptor is most abundant in dorsal areas of the striatum, the D3 receptor is also expressed in the basal forebrain but is more abundant in the ventral nuclei (nucleus accumbens and olfactory tubercle).

Dopamine Receptor Variants

Although there are only five mammalian dopamine receptor genes that give rise to D1–5 receptors, four of the genes produce multiple receptor variants (Table 1). For example, the D2 receptor has two variants that result from alternative splicing of the RNA that is transcribed from the *DRD2* gene. The long form of the D2 receptor, D2_L, contains a 29-amino-acid insert encoded by an exon that is spliced out of the short form, D2_S. Recent evidence suggests that D2_S and D2_L might serve very different functions as autoreceptors that regulate dopamine release and as postsynaptic receptors that mediate the actions of dopamine on nondopaminergic neurons, respectively. In contrast to splice variants that are derived from a single gene sequence and that are present in everybody, there are also dopamine receptor variants transcribed from distinct gene sequences, so that each individual expresses only one or two of the two or more existing variants. For example, there are at least 27 allelic variants of the human D4 receptor that have from 2 to 10 copies of an imperfectly repeated 48 nucleotide (16-amino-acid) sequence in the third intracellular loop of the receptor. There is considerable interest in the possibility that these allelic variants also vary in functional characteristics.

Therapeutic Uses for Dopamine Receptor Agonists and Antagonists

By far, the most common therapeutic purpose for drugs that act on dopamine receptors is the use of D2 receptor antagonists for

the treatment of schizophrenia and other psychoses. D2-like receptor antagonists are also used for the treatment of nausea and vomiting, for delirium or dementia of unknown cause, and for the symptomatic treatment of hyperactive movement disorder such as Tourette's syndrome and Huntington's disease. Dopamine-receptor agonists alleviate the symptoms of Parkinson's disease and are often used in combination with the dopamine precursor dihydroxyphenylalanine (L-DOPA). The D2-like receptor agonist, bromocriptine, is used to treat hyperprolactinemia because of the inhibitory effect of anterior pituitary D2 receptors on prolactin secretion. When hyperprolactinemia results from a prolactin-secreting tumor, treatment with bromocriptine also decreases the size of the tumor. Dopamine or D1 agonists such as fenoldopam cause D1 receptor-mediated vasodilation, thus increasing glomerular filtration, renal blood flow, and sodium excretion, and are used to manage types of shock associated with low cardiac output and compromised renal function.

See also: [Signaling: Gs Family of Heterotrimeric G Proteins.](#)

Further Reading

- Bunzow JR, Van Tol HHM, Grandy DK, et al. (1988) Cloning and expression of a rat D₂ dopamine receptor cDNA. *Nature* 336: 783–787.
- Carlsson A, Lindqvist M, Magnusson T, and Waldeck B (1958) On the presence of 3-hydroxytyramine in brain. *Science* 127: 471.
- Emilien G, Maloteaux JM, Geurts M, Hoogenberg K, and Cragg S (1999) Dopamine receptors – physiological understanding to therapeutic intervention potential. *Pharmacology and Therapeutics* 84: 133–156.
- Jose PA, Eisner GM, and Felder RA (1998) Renal dopamine receptors in health and hypertension. *Pharmacology and Therapeutics* 80: 149–182.
- Meador-Woodruff JA, Damask SP, Wang J, et al. (1996) Dopamine receptor mRNA expression in human striatum and neocortex. *Neuropsychopharmacology* 15: 17–29.
- Neve KA (ed.) (2010) *The Dopamine Receptors*, 2nd edn. New York, NY: Humana Press.
- Sealfon SC and Olanow CW (2000) Dopamine receptors: From structure to behavior. *Trends in Neurosciences* 23: S34–S40.
- Sidhu A, Laruelle M, and Vernier P (eds.) (2003) *Dopamine Receptors and Transporters: Function, Imaging, and Clinical Implication*. New York, NY: Marcel Dekker.
- Vallone D, Picetti R, and Borrelli E (2000) Structure and function of dopamine receptors. *Neuroscience and Biobehavioral Reviews* 24: 125–132.

Dynactin

T A Schroer, Johns Hopkins University, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Centrosome A complex organelle located at the center of many cells that is the major site of microtubule nucleation and in some cells, focusing on minus ends.

Dynein A massive, multisubunit mechanoenzyme that powers movement toward the minus ends of microtubules.

Microtubule A 25-nm-diameter hollow cytoskeletal polymer comprising 11–15 parallel protofilament strands of α/β tubulin heterodimers.

Motor An adenosine triphosphatase that performs mechanical work as part of its adenosine triphosphate hydrolytic cycle.

Processivity The ability of an enzyme to catalyze a reaction on multiple, successive subunits of a polymeric substrate.

Overview of Dynactin Composition and Structure

Dynactin contains a total of 20 individual polypeptide subunits that are encoded by 11 different genes (Figure 1). The locations and nearest neighbors of most dynactin subunits have been identified using antibody labeling and biochemical analysis (Figure 2). Most of dynactin's mass is contained in a brick-like structure that is a pentameric polymer of the actin-related protein, Arp1. The subunits p62, Arp11, p27, and p25 are organized into a cup-shaped complex that sits at one end of this polymer. The actin-capping protein (CapZ) α/β heterodimer is present at the other end of the Arp1 polymer. This end of the structure terminates in cap that most likely contains portions of the p150^{Glued}, dynamitin, and/or p24 subunits. Dynactin also contains a single monomer of conventional cytoplasmic β -actin, the precise location of which has not been rigorously defined. Dynactin uses different subunits to mediate binding to a wide variety of cellular structures.

A major structural domain of dynactin is the shoulder and flexible arm which projects approximately 20 nm from the Arp1 rod. The point at which the arm connects to the shoulder is highly flexible. This part of the dynactin molecule contains the largest dynactin subunit, p150^{Glued}, dynamitin (p50), and p24 in a stoichiometric ratio of 2:4:2. The p150^{Glued} subunit mediates interactions between dynactin, motors, and microtubules and may bind other proteins as well.

Dynactin structure has been imaged in the electron microscope (EM) by a variety of methods, including negative-staining EM and scanning transmission EM. Detailed structural information has come from platinum replicas of rotary shadowed, quick-frozen, deep-etched molecules and also averaged images of negatively stained particles. The dynactin structures obtained in this manner are the basis of the cartoon shown in Figure 2. Till the time of preparation of this entry, no high-resolution structural information about dynactin as a whole or any of its individual subunits (aside from actin, CapZ and the CAP-Gly domain of the p150^{Glued} subunit) was available.

Dynactin Functions

Functions Associated with Dynein-Based Motility

Dynactin was first discovered as a large (20S) protein complex that copurified with the minus end-directed motor, cytoplasmic dynein. Dynactin was found to be required for long-range, dynein-driven translocation of membrane vesicles *in vitro*. The name dynactin derives from this dynein-activating activity.

Adapter Function

An important function of dynactin is to serve as an adapter that allows cytoplasmic dynein to bind a variety of cargoes. This adapter function highlights the importance of dynactin's distinct dynein and cargo-binding domains. Dynein is the predominant minus end-directed motor activity in most cells, so virtually any subcellular structure or macromolecular complex that moves toward the cell center, or toward microtubule minus ends in a noncentrosomal array, is translocated by dynein in conjunction with dynactin. Dynein cargoes include membranous organelles such as Golgi-derived membranes, the ER–Golgi intermediate compartment (ERGIC), early, late, and recycling endosomes, lysosomes, and mitochondria. Dynein can also move lipid droplets, aggresomes, intermediate filaments, RNA–protein particles, viral capsids, and centrosome components toward microtubule minus ends.

Dynactin also contributes to the binding of dynein to large cellular structures such as centrosomes, mitotic spindle poles, chromosomes, the nucleus, and the plasma membrane. Dynein associates with centrosomes beginning in S-phase and remains at spindle poles during mitosis. Centrosomal dynactin is required for dynein binding. The persistent action of dynein at spindle poles is required for the retention and organization of microtubules. Dynactin is also required for proper binding and function of dynein at chromosome kinetochores, which is important for capture of spindle microtubules in prometaphase and poleward movement of kinetochore-associated mitotic regulators. Dynein may also power lateral sliding of chromosomes along microtubules and help maintain end-on attachment of chromosomes to microtubules that undergo

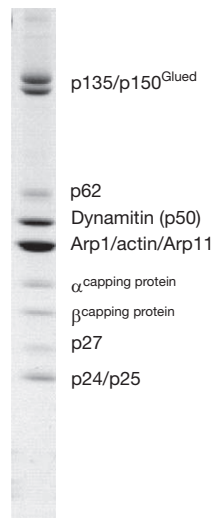


Figure 1 Dynactin composition. Dynactin's 11 distinct polypeptide subunits are shown on this Coomassie blue-stained, SDS polyacrylamide gel.

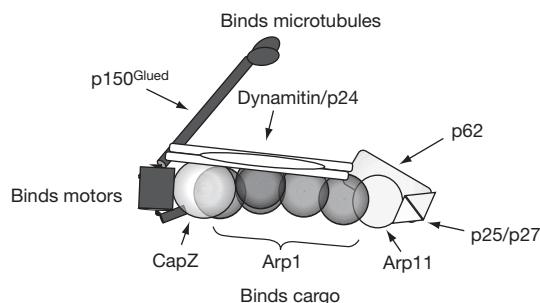


Figure 2 Subunit organization in dynactin. The three different binding domains are indicated. The precise location of conventional actin is unknown.

disassembly at their plus ends during prometaphase and anaphase. In some cell types, dynactin and dynein are recruited to the outer nuclear membrane just prior to mitosis. Dynein-driven movement of the nuclear envelope along microtubules pulls the entire centrosomal microtubule array toward the nucleus. This causes centrosomes to become closely associated with the nuclear surface and contributes to nuclear envelope rupture.

Dynein binds to the cell cortex via dynactin. As in the case of dynein bound to the nuclear envelope, the force exerted by cortical dynein 'reels in' microtubules, pulling with them any structures to which they might be attached. This process can reorient the entire microtubule array relative to the cell surface and contributes to the long-range movements of nuclei that occur during mitosis and cellular morphogenesis.

Processivity Enhancement

In the absence of dynactin, individual cytoplasmic dynein molecules are able to travel only short distances along the microtubule lattice. Dynactin's p150^{Glued} subunit provides additional microtubule binding sites that help hold dynein

near the microtubule. This allows dynein to move over longer distances, making it a more processive motor. Dynactin acts in a similar manner to enhance the processivity of kinesin-2, and may also contribute to the activity and function of other kinesin family motors.

Dynein-Independent Functions

Dynactin has the potential to support nonmotile interactions of microtubules with cellular components in a manner that is independent of motor binding. This cross-linking function depends on dynactin's distinct microtubule and cargo-binding domains. Dynactin associates with microtubules at their plus ends, their minus ends, and may even bind along their length. At plus ends, dynactin may help control the rate of microtubule growth and shrinkage by recruiting or working with other plus end-binding factors. Dynactin is also seen at sites where microtubule minus ends converge, such as centrosomes, at times when dynein is not. At this site, dynactin contributes to microtubule anchoring. Dynactin may interact with microtubule minus ends directly, or it might bind other minus end-binding or centrosome-associated proteins. Subunits of dynactin's cargo-binding domain are able to associate with a wide array of intracellular molecules and macromolecular complexes that are still being identified. This may allow dynactin to tether these structures to microtubules in a transient manner, suppressing free diffusion and helping localize them at particular sites within the cell. This may be a mechanism for holding signaling molecules in the cytoplasm and out of the nucleus.

Structure and Function of Individual Dynactin Subunits

Structure

Dynactin subunits can be grouped into four structural families: (1) proteins whose structures resemble actin (actin itself, plus the actin-related proteins Arp1 and Arp11) or actin-binding proteins (CapZ), (2) elongated, largely α -helical proteins (p150^{Glued}, dynamitin, and p24), (3) β -helix proteins (p25 and p27), and (4) Cys/His RING/LIM domain proteins (p62).

Actin-Related or Actin-Binding Proteins

Dynactin contains three proteins of the actin superfamily: β -(cytoplasmic) actin, Arp1, and Arp11. Arp1 is very similar (~65% identical) to conventional actin and is the only Arp that has been shown to assemble into filaments. Arp11 is less similar to actin but can copolymerize with actin and bind Arp1 *in vitro*. Sequence comparisons suggest that Arp11 is most divergent at its pointed end face. Arp11 is bound directly to the pointed end of the Arp1 filament to block further subunit addition. The CapZ species found in dynactin contains the β_2 isoform only. Among the dynactin subunits, these five dynactin polypeptides are the most highly conserved between species.

Elongated α -Helical Proteins

The polypeptides p150^{Glued}, dynamitin, and p24 are the least highly conserved of all the dynactin subunits, suggesting that

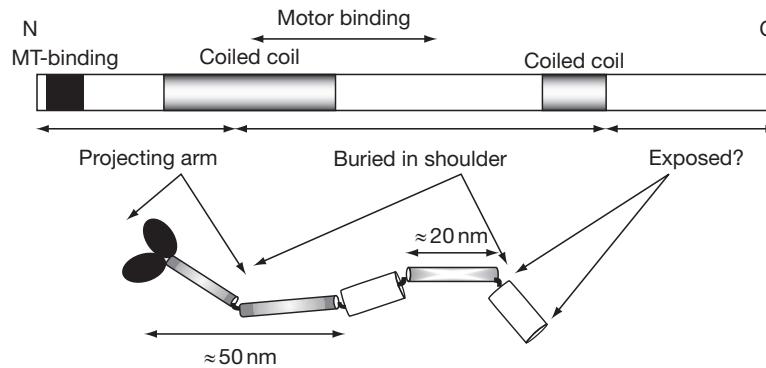


Figure 3 Two representations of the domain organization of p150^{Glued}. Top: a schematic of the primary sequence with predicted structural features and known and potential interaction sites indicated. The microtubule-binding (MT) region contains a CAP-Gly motif and a basic motif that have both been implicated in microtubule binding. Bottom: a cartoon illustrating the N-terminal globular heads plus the location and approximate lengths of the two coiled-coil domains. The structural features of the open-boxed regions have not been determined.

their overall structures play a more important role in dynactin function than their specific sequences. p150^{Glued} forms a parallel homodimer that is stabilized by two coiled coils, one ≥ 200 amino acids long and the other ~ 125 amino acids long (Figure 3). The 200-amino-acid N-terminal domains comprise the globular heads at the tip of the projecting arm. This part of p150^{Glued} supports microtubule binding via a CAP-Gly (i.e., cytoskeleton-associated protein, glycine-rich) domain and basic motif. The structure of the CAP-Gly domain (AA ≈ 40 –110) has been determined and is similar to other CAP-Gly domains. Only AA ≈ 1 –350 are exposed in the projecting arm, leaving the rest of the p150^{Glued} molecule to associate with other dynactin subunits to form the shoulder that sits on one side of the Arp1 filament. The structural features of the p150^{Glued} central region and C terminus have not yet been defined, but the middle portion (AA ≈ 520 –1050) is relatively hydrophobic and likely to be buried. The p150^{Glued} messenger RNA (mRNA) is subject to alternative splicing which yields proteins that show differential microtubule-binding properties. p150^{Glued} splice isoforms may also interact differentially with motors and other cellular components.

Dynamitin is a major structural component of the dynactin shoulder and is fundamentally important to the structural integrity of dynactin. Its sequence (AA ≈ 100 –406) predicts multiple α -helices that participate in intra- and intermolecular coiled-coil and multicoil interactions with other shoulder subunits. By itself, recombinant dynamitin forms elongated, flexible oligomers (trimers or tetramers), but it adopts a more compact structure when associated with the dynactin subunit, p24. The dynamitin N-terminus binds to Arp1 directly. Free dynamitin is able to displace p150^{Glued} from the Arp1 filament and is, thus, a powerful experimental tool that is widely used to inhibit dynactin-based phenomena *in vivo*.

β -Helix Proteins: p25 and p27

The opposite end of the dynactin molecule terminates in a complex of four proteins, Arp11, p62, p27, and p25. p27 and p25 are evolutionarily related and contain a repeated polypeptide motif that is predicted to fold into a β -helix, a structure not commonly found in animal proteins. Known β -helix proteins

are elongated triangular prisms that possess an extended binding face or groove on one face of the prism. The short β -helices in p25 and p27 may dimerize via side-to-side association to form a compact structure, or they may associate with end-to-end to form longer-binding face. Both p27 and p25 contain a loop that is predicted to project away from the β -helix prism and a C-terminal extension. p27 is subject to mitotic phosphorylation which may be a means for regulating dynactin function across the cell cycle.

RING/LIM Domain Protein: p62

Very little is known about p62 structure, but it has been shown to associate with actin, Arp11, and Arp1 and is necessary for dynactin's structural integrity. p62 contains a Cys/His-rich motif of the RING/LIM type that may indicate an associated E3 ubiquitin ligase activity. p62 contributes to the mechanism by which dynactin binds the nuclear envelope at the onset of mitosis.

Binding Activities

Dynactin can interact with a wide variety of cellular components. Its binding partners can be grouped into three major categories: (1) microtubules, (2) motors, and (3) cargoes that undergo microtubule-dependent movement or tethering. Each type of interaction is supported by a different part of the dynactin molecule (Figure 2).

Microtubule Binding

The ability of dynactin to bind microtubules is believed to contribute to its activity as processivity factor or dynein and other motors and may also allow it to serve as a tethering factor for nonmotile structures. Two distinct motifs present in the N-terminal head at the tip of the p150^{Glued} arm (Figure 3) contribute to microtubule binding. The well-known CAP-Gly motif found in a number of microtubule-binding proteins allows p150^{Glued} to associate with EB-family proteins that associate with microtubule plus ends. The adjacent basic motif is thought to support low-affinity direct interactions

with microtubules. Binding may be regulated by phosphorylation at a site that lies between these two motifs.

Motor Binding

p150^{Glued} is also the major motor-binding subunit. Its best-characterized binding partner is cytoplasmic dynein, but p150^{Glued} has been reported to interact with the kinesin family motors, Eg5, and kinesin-2. Motors bind to the central portion of p150^{Glued} (amino acids ~430–800; **Figure 3**) and phosphorylation may play a regulatory role. It is not known whether this element contains one or more motor-binding site and whether other dynactin subunits are involved.

Binding to Other Cellular Components

Membranes

Dynactin localizes to and contributes to the motility of a wide variety of cellular membranes, including the Golgi complex and associated vesicles, ERGIC, TGN, numerous elements of the endocytic pathway, the nuclear envelope, and the plasma membrane. Membrane binding involves Arp1 filament components. Arp1 associates with the actin-binding region of at least one spectrin isoform, which might provide a general binding mechanism. Binding to recycling endosomes requires the p27 and p25 subunits, whereas nuclear envelope binding utilizes p62 and Arp11.

Other Cytoplasmic Structures

The association of dynactin with lipid droplets, intracellular pathogens, RNA–protein particles, and cytoplasmic signaling complexes also appears to involve the Arp1 filament domain. In most cases, binding mechanisms have not yet been defined.

Chromosomes

Dynactin is bound to chromosome kinetochores during mitosis. As in the case of other structures that bind dynactin, kinetochore binding probably involves multiple dynactin subunits. One known contributor to this interaction is dynamitin, which has been reported to bind the kinetochore component, ZW10, directly.

See also: **Cell Architecture and Function:** Actin-Related Proteins; Dynein; Microtubule-Associated Proteins.

Further Reading

- Kardon JR and Vale RD (2009) Regulators of the cytoplasmic dynein motor. *Nature Reviews Molecular Cell Biology* 10: 854–865.
- Schroer TA (2004) Dynactin. *Annual Review of Cell and Developmental Biology* 20: 759–779.
- Steinmetz MO and Akhmanova A (2008) Capturing protein tails by CAP-Gly domains. *Trends in Biochemical Sciences* 33: 535–545.

Dynamic Behavior of Photosystem II Light Harvesting

P Horton, University of Sheffield, Sheffield, UK

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Peter Horton and Alexander Ruban, volume 3, pp. 357–362, © 2004, Elsevier Inc.

Glossary

Grana membrane The characteristic organization of the chloroplast thylakoids in which regions of membrane become tightly appressed together to form stacks of membranes, associated with the lateral segregation of the photosystems.

Nonphotochemical quenching The quenching of chlorophyll fluorescence that arises because of an increase in thermal dissipation of excitation energy, contrasted to the photochemical quenching that arises because energy is used to drive photosynthetic electron transfer.

Photoinhibition The inhibition of photosynthesis that arises during excess illumination, either from

sustained quenching of antenna or reaction center complexes or from accumulation of damaged photosystem II D1 proteins.

State transitions The reversible adaptation of the light-harvesting system to differential excitation of photosystem I (PSI) and PSII; in state 1, induced by over excitation of PSI, energy transfer from LHCII to PSII is maximum, and in State 2, induced by overexcitation of PSII, energy transfer from LHCII to PSI is maximum.

Xanthophyll cycle The reversible de-epoxidation of thylakoid-bound violaxanthin to zeaxanthin, via the intermediate antheraxanthin.

The light-harvesting proteins of higher plants play a vital role in the regulation of photosynthesis. These proteins show striking structural and functional flexibility. This dynamic behavior provides a key physiological function to plants, which adapts them to continuously changing environmental conditions. In particular, dynamic properties of the light-harvesting proteins of photosystem II (PSII) promote the efficient collection of sunlight when the intensity is limiting photosynthesis and effective photoprotection when the intensity is excess.

Photosynthetic Light Harvesting

PSII is the multisubunit chloroplast membrane-associated pigment–protein complex that uses the energy of sunlight to drive the oxidation of water, evolving oxygen, donating electrons into the photosynthetic electron transfer chain, and depositing protons into the thylakoid lumen. Along with the PSI, it forms the electron–proton transfer chain, which drives the synthesis of adenosine triphosphate (ATP) and reduced nicotinamide adenine dinucleotide phosphate (NADPH) (Figure 1). Vitrally important components of both photosystems are the light-harvesting antennae, the light-collecting units (mainly LHCI and LHCII for PSI and PSII, respectively) that ensure high rates of energy input into the photosynthetic reaction centers (RCI and RCII, respectively) by intercepting large numbers of light quanta of various energies/colors.

The light-harvesting antenna of PSII consists of several proteins, which together bind around 300 chlorophyll molecules. Associated tightly with the D1/D2 reaction center are the core antenna complexes CP47 and CP43. The remainder of the antenna consists of the Lhcb proteins, Lhcb1–6. These bind chlorophyll *a*, chlorophyll *b*, and xanthophylls to form several

different complexes – LHCII (Lhcb1–3), CP29 (Lhcb4), CP26 (Lhcb5), and CP24 (Lhcb6). LHCII is the main complex and contains about 40% of the PSII chlorophyll – it is the most abundant chlorophyll protein in nature. Lhcb1, Lhcb2, and Lhcb3 associate in different combinations to form a population of heterotrimeric LHCII, which show different strengths of binding to the PSII core complex.

A single Lhcb monomeric unit is a relatively small protein, ~25 kDa, containing three transmembrane alpha-helical structures (A, B, and C helices) and binding up to 12 molecules of chlorophyll (up to 7 chlorophyll *a* and 5 chlorophyll *b*) and up to four xanthophylls. In the case of LHCII, two xanthophylls, lutein 1 and 2, are associated with the helices A and B. A third carotenoid, neoxanthin, is associated with the helix C and the trimer also binds peripherally the carotenoids violaxanthin or zeaxanthin (Figure 2). *In vivo*, two PSII reaction center complexes form the dimeric PSII core complex, which may bind up to four trimeric LHCII and six monomeric CP24, CP26, and CP29. In the photosynthetic membrane, these PSII units are sometimes seen as ordered arrays, their frequency reflecting the regulation of PSII function as explained below. The LHCII antenna forms a dynamic network, or macrodomain, of monomeric and trimeric subunits, associated with each other and the reaction-center complex. Associated with this macrostructure is the stacking of the complexes together in the characteristic grana membranes.

Three major parameters determine the efficiency of light harvesting: (1) the absorption cross section (or number of pigments) and their ability to intercept light quanta with the broad range of energy; (2) the time during which energy of light can be kept in antenna so it can be funneled into the reaction center; and (3) rate of the funneling. Thus, a large cross section, long-excitation-energy lifetime, and high rate of energy transfer to the reaction center are the attributes of an efficient antenna and productive photosynthetic unit. Hence,

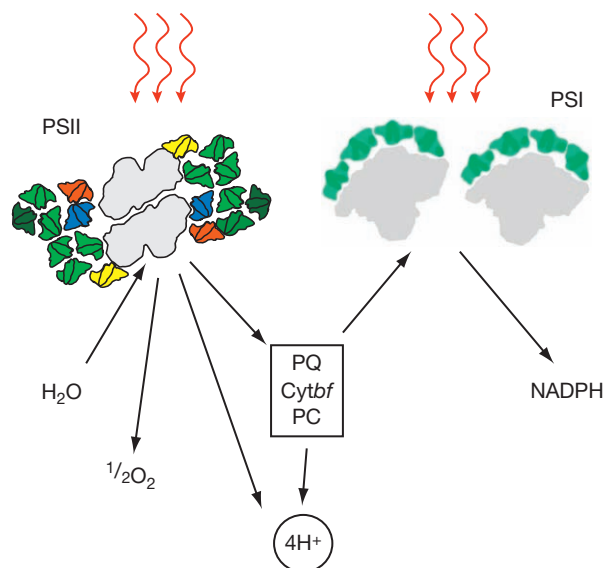


Figure 1 The photosynthetic light-harvesting and electron transfer system. PSII and PSI both served by light-harvesting antenna drive the oxidation of water and the reduction of NADP, connected by an electron transfer chain consisting of plastoquinone (PQ), the cytochrome *bf* complex (Cytbf) and plastocyanin (PC). H^+ release into the thylakoid lumen results in formation of a ΔpH , the driving force for ATP synthesis. Shown are the dimeric and monomeric core complexes of PSII and PSI (gray), the antenna protein of PSI, Lhca1-4 (mixed green), and PSII, Lhcb1, Lhcb2 (pale green), Lhcb3 (dark green), Lhcb4 (blue), Lhcb5 (yellow), Lhcb6 (red). Two types of LHCII trimers are shown – strongly bound S-trimers of Lhcb1, Lhcb2 and more weakly bound M-trimers containing Lhcb3.

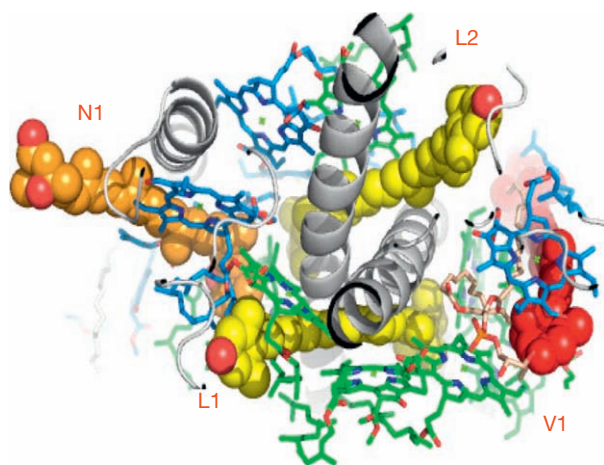


Figure 2 Structural model of LHCII monomer as revealed by X-ray crystallography. The dominant feature of the structure are the three α -helices (gray), which provide the scaffold for the binding of chlorophyll *a* (green) and chlorophyll *b* (blue) molecules. The four xanthophylls are shown: N1 (neoxanthin), L1 and L2 (luteins), and V1 (violaxanthin).

the efficiency of light harvesting in PSII is determined by the number of LHCII subunits, the pigment order within them, the interaction between subunits, and their closeness to the reaction center complex.

The Need for Control of the Light Reactions of Photosynthesis

The light and dark reactions of photosynthesis are tightly coupled, the ATP and NADPH provided by the electron-proton transfer system driving the fixation of CO_2 into carbohydrate. There are dramatic diurnal changes in the spectrum of sunlight reaching the leaf surface due to filtering by atmosphere, clouds, and other plants. As LHCI and LHCII have different capacities to absorb light of different wavelengths, there will be a frequent imbalance in light energy input into PSI and PSII reaction centers. This imbalance would decrease the efficiency of electron transfer, and hence the overall quantum yield of photosynthesis. The light intensity is also a very changeable parameter, varying a 1000-fold. Whenever light input exceeds the capacity of electron transport and CO_2 assimilation, there is an excess of light, which may cause photoinhibition – the destabilization (by overreduction and overenergization) and even damage (by photooxidation and generation of reactive-oxygen species) to photosynthetic components, mainly in PSII. The effects of such changes in light are amplified by alterations in photosynthetic capacity brought about by changes in other environmental factors (such as low temperature or drought), or by changes in metabolic demand, which result in variation in the rate of turnover and the stoichiometry of ATP and NADPH.

To achieve balance and stability of the photosynthetic process, regulatory mechanisms are necessary. The plant cell first has to assemble a chloroplast with the correct composition, and then various parts of the photosynthetic process need to be able to adjust their activities in response to internal and external information. Thus, light harvesting, electron transport, and carbon metabolism can be considered to be linked by a feedback and feedforward control network, which operates at the levels of gene expression, protein turnover, covalent modification, and protein activity. There is a compromise between maximizing the collection and utilization of light, and the avoidance of instability when light is in excess. The most appropriate way of looking at the regulatory mechanisms is that, given a fixed composition and a fluctuating environment, they extend the range of conditions over which photosynthesis can remain in balance – they provide homeostasis of excitation energy level, redox state, and ΔpH .

Two distinct regulatory strategies optimize light-harvesting function. The first balances the input of light energy to PSI and PSII reaction centers in order to optimize electron transfer rate – this is called the state transition. The second regulates the amount of light energy directed to PSII – this is called nonphotochemical energy dissipation or quenching (NPQ). Both regulatory mechanisms take place within the LHCII system and rely on its abilities to sense the state of the light-energy balance and to structurally respond, being capable of dynamic behavior.

State Transitions

The Lhcb1 and Lhcb2 polypeptides of LHCII are reversibly phosphorylated by a redox-regulated thylakoid-associated protein kinase, STN7. Phosphorylations at threonine residues near the N terminus of LHCII weaken the association between

PSII and LHCII, as a result of electrostatic and conformational changes (Figure 3). Phospho-LHCII partitions in favor of PSI, this association depends on the presence of the PSI-H subunit – in mutants without PSI-H phospho-LHCII remains associated with PSII. These changes in association between LHCII and the two photosystems may take place primarily (or only) at the grana periphery, and may not involve long-distance migration between appressed and unappressed membranes. Conversely, limited diffusion of complexes in the plane of the thylakoid membrane may be promoted under conditions of LHCII phosphorylation. Certain specific LHCII trimers seem to be preferentially dissociated from the PSII complex and the kinetics of the state transition are correspondingly influenced by the exact composition of the Lhcb proteins. The kinase is activated when the plastoquinone pool is reduced, this being sensed by the plastoquinol oxidation site on the cytochrome *b₆f* complex. The kinase, therefore, monitors the relative rates of delivery of excitation energy to PSII and PSI; overexcitation of PSII (state 1) activates the protein kinase, decreases delivery of excitation energy to PSII, and increases transfer to PSI (state 2). Levels of LHCII phosphorylation are lower at high light compared to low light, showing that the state transitions optimize photosynthesis in limiting light. This includes not only optimizing linear electron transport, but also the provision of optimal ΔpH and ATP/NADPH ratios by controlling the proportion of cyclic electron transfer around PSI.

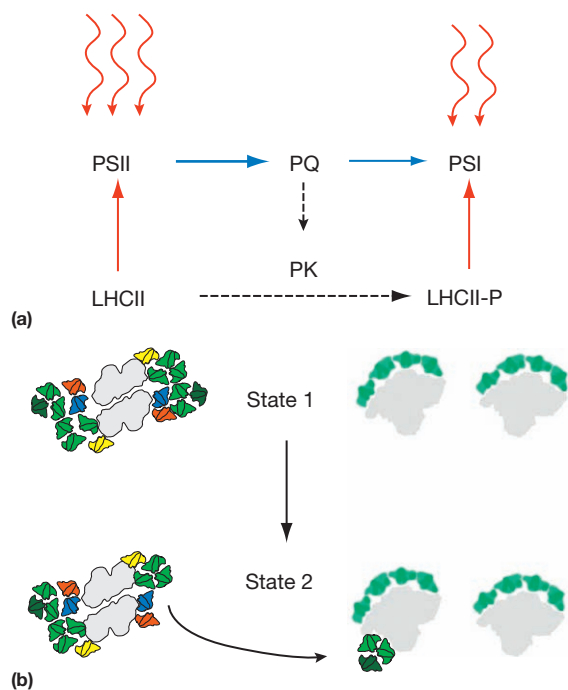


Figure 3 The state transitions. (a) Kinetic model showing how overexcitation of PSII relative to PSI in state 1 leads to electron buildup in the PQ pool (blue arrows), activating the thylakoid STN7 protein kinase (PK), which phosphorylates LHCII. Phosphorylated LHCII (LHCII-P) transfers its energy to PSI (state 2), thereby correcting the imbalance. (b) Structural basis of the state transitions. Phosphorylation of a subpopulation of LHCII trimers, leads to its dissociation from the PSII complex and its attachment to PSI, thereby increasing its absorption cross section. Color coding of subunits as in Figure 1.

Nonphotochemical Energy Dissipation

Energy-Dependent Quenching, qE

Dissipation, or quenching, of excess excitation energy absorbed by LHCII occurs by two processes, distinguished by the speeds with which they are induced upon exposure to illumination, their relaxation times in darkness, and their light intensity thresholds. The major process under normal conditions is called qE, because it depends upon the energization of the thylakoid membrane as a result of ΔpH formation (Figure 4). Thus, the progressive increase in ΔpH that occurs as light saturation of photosynthesis is reached is the trigger for energy dissipation. It can, therefore, be referred to as feedback de-excitation. The formation of qE occurs within seconds to minutes of exposure to excess light, and relaxes with a similar rate in darkness. By contrast, the sustained qI-type of quenching may take several minutes or even hours to appear and relax, and under some conditions (such as low temperature) may be stable for days.

The Site of Quenching

Application of mathematical models for PSII energy transfer to simple chlorophyll fluorescence measurements showed that quenching associated with qE occurred in the light-harvesting antenna. Analysis of the chlorophyll fluorescence lifetimes showed that qE is best explained by the transition between two states of the antenna, with different rate constants for energy dissipation. Spectroscopic data and biochemical analysis confirmed that the PSII antenna was the site of quenching. A powerful way to determine which proteins are involved in qE is to investigate plants deficient in specific components. Experiments using plants deficient in all of the Lhcb proteins suggested that qE was not a property of the core antenna. However, plants in which Lhcb1 and Lhcb2, Lhcb3, Lhcb4, Lhcb5, and Lhcb6 have been separately reduced to less than 5% wild-type levels of protein by antisense technology or to zero by knockout mutation all show qE, albeit with reduced levels and/or altered kinetics. It is concluded that qE is not confined to a single complex, but rather is a property of several working together. A major landmark in qE research was the identification of the *npq4* mutant, a mutant lacking qE and which is deficient in the Lhc-related protein PsbS. This four-helix protein is located in PSII, and contains protonatable amino acid residues that are thought to play some part in sensing the thylakoid lumen pH. It remains unclear how PsbS carries out its function in qE, but there is evidence that it acts indirectly, controlling the transition of the antenna from the unquenched to quenched states.

The Xanthophyll Cycle

Under conditions of excess light, violaxanthin is de-epoxidized to zeaxanthin via the monoepoxide antheraxanthin due to the activation of violaxanthin de-epoxidase, a thylakoid lumen enzyme. Another enzyme, zeaxanthin epoxidase reverses the reaction in low light. A vast amount of data, from a wide variety of plant species under many different environments, showed correlations between the extent of qE and the deepoxidation state of the xanthophyll cycle pool. Zeaxanthin is an

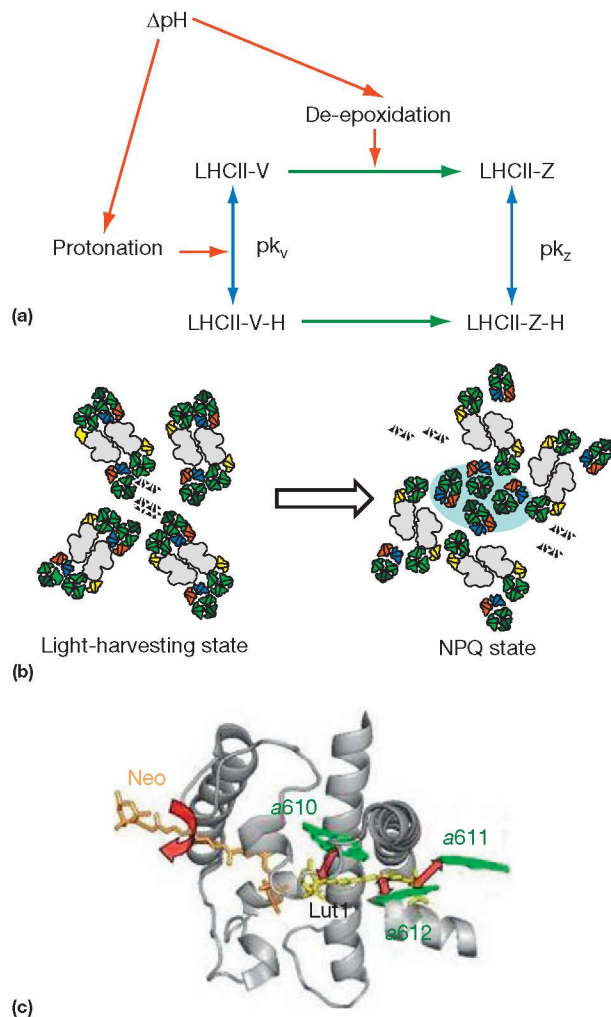


Figure 4 Nonphotochemical quenching. (a) Kinetic model showing how the ΔpH , the signal of excess light, triggers the protonation of the light-harvesting antenna (resulting in the NPQ state, qE) and also stimulates the de-epoxidation of bound violaxanthin to zeaxanthin. The pK_a of the V-form is lower than that of the Z-form, so that V and Z are allosteric regulators of qE. (b) A possible structural basis of qE regulation: under the influence of the PsbS protein (black), partial dissociation of the PSII complex occurs, leading to aggregation of Lhc subunits (pale blue) in which quenching centers are created. De-epoxidation then favors the formation of the NPQ state. The proportion of PSII complexes forming semi-crystalline domains (not shown) may control how many PSII complexes are available for switching into the NPQ state. (c) A quenching center in LHCII, depicting the role of Lutein 1 as the quencher within a chlorophyll *a* domain (a610, a611, and a612). A conformational change in neoxanthin occurs as the complex switches from the light-harvesting state into the NPQ state.

activator of qE in that the apparent pK_a of the protonation reaction inducing qE shifts from near 4.5 to well over 5.0 in the presence of zeaxanthin. As the ΔpH *in vivo* does not fall below 5.5, qE will depend almost completely on de-epoxidation of violaxanthin. Therefore, qE is an allosteric process, under the control of the interacting effects of protonation and zeaxanthin binding, similar to a regulated enzyme (Figure 4(a)). This mode of regulation explained how the chloroplast could have a ΔpH high enough in limiting light to allow ATP

synthesis without qE, yet in saturating light, how to have maximum electron transport rates and high qE simultaneously. Similarly, the variable sensitivity of qE to ΔpH explains the kinetics of NPQ when plants are transferred from dark adaptation into light: an initial rapid phase of quenching reflecting the baseline response to ΔpH , is followed by a slow phase as violaxanthin is de-epoxidated and qE activated; then a second illumination after brief dark adaption sees NPQ form rapidly in one phase, reflecting the activated response of qE to ΔpH .

It has proved difficult to completely distinguish between the proposal that zeaxanthin was directly involved in quenching chlorophyll-excited states by chlorophyll/zeaxanthin energy transfer, and the notion that zeaxanthin was working indirectly, modifying or inducing a quenching process intrinsic to the PSII antenna.

The Nature of the Quenched State of LHCII

Changes in the absorption spectra of chlorophyll and carotenoid occur upon formation of qE. A band with a maximum at 535 nm (ΔA_{535}) has attracted most attention, because its appearance is perfectly correlated to the amount of quencher. Resonance Raman spectroscopy of leaves and chloroplasts has shown that ΔA_{535} is electronic in origin, and arises from a pool of one or two zeaxanthin molecules per PSII, which undergo a strong red shift in the presence of ΔpH .

Analysis of fluorescence lifetimes shows that the qE state of the PSII antenna has a lifetime of around 0.5 ns and the unquenched state 2.0 ns. Experiments with isolated light-harvesting complexes have shown how such a change can arise. Detergent-solubilized LHCII has a lifetime of 4 ns, but reduction of the detergent concentration causes formation of strongly quenched oligomers with an average lifetime of 0.3 ns. This quenching has many features that are similar or identical to qE, and it is thought that the local environment in the grana membrane tunes the state (and lifetime) of the antenna complexes to give rise to those found *in vivo*. The spectral changes in chlorophyll accompanying this *in vitro* quenching, indicate that a particular LHCII domain is involved, and transient absorption spectroscopy has shown that protein-bound xanthophyll participates in the quenching mechanism (Figure 4(c)). The *in vitro*-quenched state can also be investigated by reconstitution of light-harvesting complexes with altered carotenoid content. Studies on such complexes again point to a quenching mechanism involving xanthophylls.

In general, the ease with which LHCII *in vitro* can adopt a quenched state reflects the high density of pigment within it. Its design as a light-harvesting complex means it maximizes the amount of chlorophyll per unit volume. Specific features of protein structure are needed to control the interactions between pigments, to promote energy transfer and to prevent formation of quenchers – dimers or excimers of chlorophylls and/or chlorophyll–xanthophyll associates. Therefore, only rather small changes in structure are needed to allow a quencher to form. Indeed the activation energy for the formation of a quencher in LHCII indicates the breakage of just a couple of H bonds. Thus, LHCII is uniquely poised to allow a switch between an unquenched state functioning in

photosynthetic energy capture and a quenched state, dissipating excess energy as heat. Single-molecule spectroscopy of LHCII has shown how this intrinsic switching arises from inherent molecular disorder that allows structurally very similar but functionally very different states.

The Role of Thylakoid Suprastructure in LHCII Regulation

The dynamic behavior of LHCII occurs within the grana membranes of the chloroplast. Here, there is a flexibility in the macroorganization of the LHCII–PSII supercomplexes. Complexes can associate into areas of regular crystalline arrays or of more random organization. There appears to be sufficient mobility of complexes in the plane of the membrane to allow a dynamic equilibrium between different associations between complexes and their subunits. For example, qE may require the complexes to be in a random organization in which clustering and associations between Lhcb subunits can take place under the influence of the protonation exposed residues on thylakoid lumen surface (**Figure 4(b)**), just as the state transitions are regulated by changes in stromal surface charge due to phosphorylation (**Figure 3(b)**).

Biological Diversity of qE

The amount of qE in different plants is highly variable. High-light-grown plants may have 2–3 times more qE capacity and plant species adapted to growing in stressful environments have a much higher qE capacity than those inhabiting milder conditions. Genetic manipulation to increase the level of the PsbS protein results in an increase in qE, suggesting that the level of this protein is an important determinant of qE capacity.

Algae exhibit different qE properties to higher plants. In *Chlamydomonas*, qE is smaller and forms more slowly and mutation of an LHCII subunit decreases the capacity of qE, suggesting that here a less-efficient process is controlled directly by the antenna and not by PsbS. In diatoms, which have light-harvesting proteins of a different type, qE can be much larger than in higher plants, and its formation and relaxation totally depend upon the de-epoxidation of the violaxanthin analog, diadinoxanthin. It seems that some features of the molecular mechanism for the regulation of light harvesting have diverged during evolution of classes of organisms.

Other Responses of LHCII to Excess Light

The sustained qI-type of quenching appears to also result, in part at least, from alteration in the antenna of PSII. At low temperature, strong quenching is induced by light which has some features diagnostic of an increase in aggregation state of LHCII – formation of a red-shifted fluorescence emission band at 77 K. This state of LHCII seems to be semi-permanent in some evergreen plants during winter. The mechanism by which such quenching is formed is currently unknown. At elevated temperature, increased light-induced quenching arises from a different process – LHCII dissociates from PSII to be quenched

by PSI, in a process that resembles the state transition, and, in fact, phosphorylation of LHCII and elevated temperature act synergistically.

When plants grown under low light are exposed to a sustained increase in light intensity, LHCII degradation is induced. Only monomers can be the substrates for proteolysis, and so this process must be preceded by a light-induced breakdown of trimers. Little is known about the biochemistry or regulation of this proteolytic process. Under extreme conditions, not only LHCII but also complete photosystems are degraded as the plants seek to minimize oxidative damage.

The content of LHCII varies significantly depending on the light intensity under which the plants are grown. Typically, a low-light-grown plant may have four to five trimers per PSII core complex, whereas in high light this reduces to one to two. The extra LHCII in low-light plants appears to be present in membrane domains deficient in PSII core complexes. These LHCII domains can still be involved in efficient energy transfer, even transferring energy to pigments on opposite membranes in the grana.

The adjustment of the composition of photosystem II upon growth in high light compared to low light maintains the redox potential and ΔpH at the correct level, optimizing photosynthesis and avoiding oxidative stress. At higher growth irradiance, when LHCII cannot decrease any more, there are increases in the capacity for photoprotection (e.g., through an increase in xanthophyll-cycle activity or PsbS content) and, eventually, decrease in the chlorophyll content of the leaf in order to lower light absorption. Such acclimation to light conditions involves gross changes in chloroplast composition and is costly in terms of energy and time. Thus, one may look upon the short-term dynamics of the LHCII antenna as buffering the effect of environmental perturbation, extending the range over which photosynthetic efficiency and photoprotection can be maintained without alteration in composition.

Even macroscopic events such as chloroplast movements or changes in leaf orientation can be similarly viewed as contributing to the optimization of light-harvesting and counteracting photodamage. Thus, plants do whatever they can, by a multitude of mechanisms, to limit the level of excitation energy in the PSII antenna, giving balance with the demands of photosynthesis, and optimizing the redox state and ΔpH . Such regulation is driven by the survival and fitness of plants: it has been questioned whether the optimization observed in economically important crop plants is adapted to high productivity. Thus, there is interest in exploring whether retuning the dynamics of LHCII could bring about increases in crop yield.

See also: **Bioenergetics:** Chloroplasts; Photoinhibition and Photoprotection in Plants, Algae, and Cyanobacteria; **Metabolism** **Vitamins and Hormones:** Photosynthesis.

Further Reading

- Blankenship RE (2001) *Molecular Mechanisms of Photosynthesis*, p. 328. Oxford: Blackwell Science.
- Demmig-Adams B and Adams WW (2002) Antioxidants in photosynthesis and human nutrition. *Science* 298: 2149–2153.

- Green BR and Dunford DG (1996) The chlorophyll–carotenoid proteins of oxygenic photosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology* 47: 685–714.
- Haldrup A, Jensen PE, Lunde C, and Scheller HV (2001) Balance of power: A view of the mechanism of photosynthetic state transitions. *Trends in Plant Science* 6: 301–305.
- Horton P, Johnson MP, Perez-Bueno ML, Kiss AZ, and Ruban AV (2008) Photosynthetic acclimation: Does the dynamic structure and macro-organisation of photosystem II in higher plant grana membranes regulate light harvesting states? *FEBS Journal* 275: 1069–1079.
- Horton P, Murchie EH, Ruban AV, and Walters RG (2001) Increasing rice photosynthesis by manipulation of the acclimation and adaptation to light. Rice biotechnology: Improving yield, stress tolerance and grain quality. *Novartis Foundation Symposium* 236: 117–134.
- Horton P, Ruban AV, and Walters RG (1996) Regulation of light harvesting in green plants. *Annual Reviews of Plant Physiology and Plant Molecular Biology* 47: 655–684.
- Kargul J and Barber J (2008) Photosynthetic acclimation: Structural reorganisation of light harvesting antenna – role of redox-dependent phosphorylation of major and minor chlorophyll *a/b* binding proteins. *FEBS Journal* 275: 1056–1068.
- Kirchhoff H (2008) Molecular crowding and order in photosynthetic membranes. *Trends in Plant Science* 13: 201–207.
- Müller P, Li XP, and Niyogi KK (2001) Non-photochemical quenching. A response to excess light energy. *Plant Physiology* 125: 1558–1566.
- Niyogi KK (1999) Photoprotection revisited: Genetic and molecular approaches. *Annual Review of Plant Physiology and Plant Molecular Biology* 50: 333–359.

Dynein

K K Pfister, University of Virginia, Charlottesville, VA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

AAA+ ATPase Superfamily of ATPases associated with diverse cellular activities found in all kingdoms of living organisms.

Axoneme Cytoskeletal structure of the cilia and flagella composed of nine outer doublet microtubules with dynein arms and radial spokes, and two central microtubules with central pair projections.

Head Globular C-terminal portion of the dynein heavy chain containing six AAA domains and the microtubule binding stalk. It is the smallest portion of the heavy chain that has ATP-dependent microtubule binding.

Linker The region of the N-terminus of the heavy chain that links the Tail/stem to AAA domain 1 of the head. Movement of the linker plays an important role in dynein movement along microtubules.

Microtubule A cytoskeletal polymer of tubulin dimers in approximately 13 protofilaments arranged as a hollow tube with two distinct ends, the plus and minus ends.

Tail/stem N-terminal region of the dynein heavy chain binds most other subunits of the dynein complex and can participate in dimerization with one or more other heavy chains.

Characteristics of Dynein Complexes

Dyneins are found in many eukaryotes, including fungi, worms, insects, and vertebrates, but genome sequence analyses indicate that they are not present in flowering plants. Dynein complexes are composed of one to three heavy chains, and each complex also has various smaller accessory subunits (Tables 1 and 2). Dyneins are classified as either cytoplasmic or axonemal. The green alga, *Chlamydomonas*, which has two flagella, has been a useful model system for studies of axonemal dyneins. These dyneins are the arms which project from the doublet axonemal microtubules of flagella and cilia. The axonemal dyneins generate the sliding force between outer doublet microtubules that is converted by other axonemal structures into the bending of cilia and flagella. The axonemal dyneins are further divided into the outer arm and inner arm dyneins defined by their position of the outer doublet microtubules. Of the more than ten different dynein complexes identified in animals, most are components of the flagellar axoneme. Various organisms, from yeast to mouse, have been utilized for studies on cytoplasmic dynein. Cytoplasmic dynein 1 moves membranous organelles, kinetochores, and viruses along microtubules. It is also the motor for retrograde axonal transport and is involved in the assembly and function of the mitotic spindle. The second type of cytoplasmic dynein, cytoplasmic dynein 2, is responsible for a component of intraflagellar transport (IFT), specifically the transport of protein complexes from the tips of cilia and flagella to their base, that occurs between the axonemal microtubules and the flagellar membrane. This dynein is often referred to as IFT dynein.

The different dynein complexes hydrolyze MgATP at different rates, for example, the microtubule stimulated ATPase activity of bovine brain cytoplasmic dynein is $0.3 \text{ mol min}^{-1} \text{ mg}^{-1}$ and the activity of axonemal outer arm dynein is $2.0 \text{ mol min}^{-1} \text{ mg}^{-1}$. Different dynein complexes also move along microtubules at different rates. An *in vitro* motility assay is used to quantify the velocity of dynein-based movement

by using computer-enhanced video microscopy. Dynein is attached to a glass coverslip and when microtubules are added, the dynein motor domains bind the microtubules. In the presence of MgATP, the dynein moves microtubules across the coverslip with the microtubule-plus ends leading. Cytoplasmic dynein from bovine brain moves microtubules at the rate of approximately 1 $\mu\text{m/s}$ while sea urchin outer arm dynein moves microtubules at the rate of 3.5 $\mu\text{m/s}$.

The Dynein Heavy Chain

The central component of the dynein complex is the heavy chain, an approximately 530,000 Da polypeptide, and it is a relative of the AAA+ ATPase superfamily. Sometimes the heavy chain alone is mistakenly referred to as a dynein. The C-terminal two-thirds of the heavy chain, approximately 350,000 Da, is necessary for ATP-dependent microtubule binding. Structural analyses indicate that the C-terminal motor portion forms a large globular domain known as the head (Figure 1) made up of six sequential AAA domains. Projecting from the head as a stalk between AAA domains 4 and 5 is the microtubule binding region, approximately 350 amino acids long, consisting of a globular region between two antiparallel coiled coils. AAA domains 1 through 4 contain consensus nucleotide-binding sites known as Walker A motifs, or P loops, while AAA domains 5 and 6 lack nucleotide-binding motifs. AAA 1 is the site of ATP hydrolysis for force generation. In the presence of vanadate and inorganic phosphate, the heavy chains are cleaved by photolysis with ultraviolet light into two fragments in AAA 1. Experimental evidence indicates that the nucleotide-binding activities associated with AAA domains 2, 3, and 4 are involved in the regulation of the motor activities, including microtubule binding and ATPase activity. The extreme C-terminus of the heavy chain spans a region from AAA domain 6 toward the stalk. The N-terminal portion of the heavy chain extends from the head at AAA

domain 1 as a flexible linker that continues as the stem or tail. Movements of the lever arm-like linker and the stalk play key roles in the conversion of the energy released by ATP hydrolysis at AA1 into dynein movement along a microtubule.

Two or three dynein heavy chains can interact via their N-terminal tails to form a common base. Most of the other subunits of the dynein complexes bind to the N-terminus of the heavy chain at the tail and these subunits contribute to the motor protein's cargo binding domain, or base. The axonemal and cytoplasmic dynein heavy chains can be distinguished by comparing the Walker A motifs of AAA domains 1–4. Genome analyses indicate that generally there are approximately 14

axonemal dynein heavy chain genes, one cytoplasmic dynein 1 heavy chain gene, and one cytoplasmic dynein 2 heavy chain gene. Axonemal outer arm dynein complexes are composed of two or three different heavy chains depending on the species. Each axoneme has many different inner arm dyneins. One inner arm dynein is a heterodimer of two different heavy chains, and there are at least six different monomeric inner arm dyneins each with one copy of a unique heavy chain. The cytoplasmic dynein complexes are homodimers of their unique heavy chains.

Cytoplasmic Dyneins

Cytoplasmic Dynein 1

The mammalian cytoplasmic dynein 1 (Figure 1 and Table 1) has a native molecular weight of 1.5×10^6 Da. It is a homodimer of two identical heavy chains (DYNC1H1) associated with two approximately 74 kDa intermediate chains (DYNC1I1), two approximately 55 kDa light intermediate chains (DYNC1LI1), and three pairs of light chains (8–12 kDa), the Tctex1 (DYNLT), roadblock (DYNLRB), and LC8 (DYNLL) families. The heavy chain is encoded by a single gene and in vertebrates all the other subunits are encoded by two genes. Additional intermediate chain and light intermediate chain isoforms are also generated by mRNA alternative splicing. The multiple gene products enable the cell to produce functionally different cytoplasmic dynein complexes; for example, in neurons, signaling endosomes bind to dynein complexes with a specific intermediate chain isoform.

The heavy chain dimerization domain is a large portion of its N-terminal tail. Also in this region are the overlapping

Table 1 Composition of cytoplasmic dynein complexes^a

Subunit	Cytoplasmic dynein 1	Cytoplasmic dynein 2 (IFT)
Heavy chain	DYNC1H1 (HC)	DYNC2H1
Intermediate chain	DYNC1I1 (IC1) DYNC1I2 (IC2)	DYNC2I1
Light intermediate chain	DYNC1LI1 (LIC1) DYNC1LI2 (LIC2)	DYNC2LI1
Light chains		
Tctex1 family	DYNLT1 (Tctex1) DYNLT3 (Rp3)	None identified
Roadblock family	DYNRB1 (RB1) DYNRB2 (RB2)	None identified
LC8 family	DYNLL1 (LC8-1) DYNLL2 (LC8-2)	DYNLL1

^aAll subunits are present as (homo)dimers. The formal subunit name is followed by a descriptive abbreviation in parentheses.

Table 2 Composition of major Chlamydomonas axonemal dynein complexes

Subunit	Outer arm	Inner arms ^a			
		<i>f</i> (I1)	<i>a/c</i>	<i>d</i>	<i>b/e/g</i>
Number of heavy chains	3	2	1	1	1
Heavy chains	∇, ∃, & (	HC1 (I∇) & HC10 (I∃)	HC6/ HC9	HC2	HC5/ HC8/ HC7
Intermediate chains	IC1 IC2	IC140 IC 138 IC 97 ODA7	None	None	None
Light intermediate chain	None	none	None	None	None
Light chains	LC1/Leucine rich LC2/ Tctex2 LC3/ Thioredoxin LC4/ Ca+2 binding LC5/ Thioredoxin LC6/ LC8 homolog LC7a/ Roadblock homolog LC7b/ Roadblock homolog LC8 LC9/ Tctex1 homolog LC10/ LC8 homolog	Tctex1 Tctex2b LC7a LC7b LC8	Actin p28	Actin p44 p38 p28	Actin Centrin

^aOf the several different nomenclatures used to identify inner arm dyneins, only one is presented here, except for that of the dimeric (two-headed) inner arm dynein. Monomeric inner arm dyneins with different heavy chains, but shared light chains, are grouped for simplicity.

From King SM and Kamiya R (2009) Axonemal dyneins: Assembly, structure, and force generation. In: Witman GW and Harris EH (eds.) *The Chlamydomonas Source Book*, Vol. 3, *Cell Motility and Behavior*, 2nd edn., pp. 31–208. Oxford: Elsevier/Academic Press.

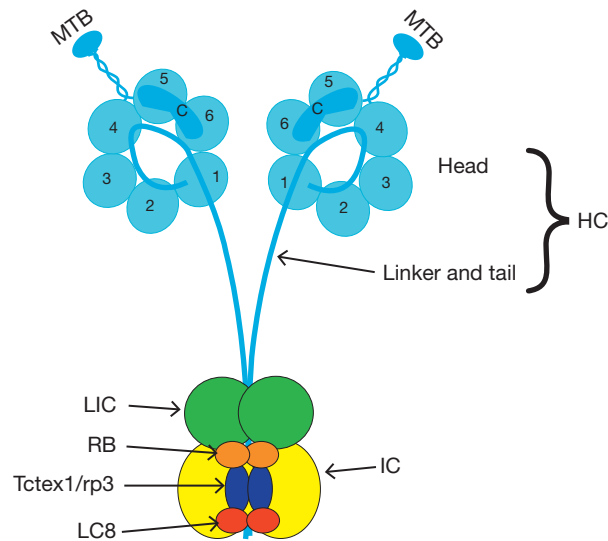


Figure 1 Model for the structure of cytoplasmic dynein 1. The heavy chains (blue) have a C-terminal head (light blue) with six AAA domains (numbered); the far C-terminus (C dark blue) forms a separate seventh domain of the head. A microtubule binding domain (MTB) projects from each head at the tip of a stalk (dark blue). The N-terminus of the heavy chain extends from the head at AAA domain 1, as the linker region (dark blue) along the ring of AAA domains; the N-terminal region then continues as the tail/stem which dimerizes and binds the light intermediate chains (LICs) and the intermediate chains (ICs). The three light chain dimers bind the intermediate chains.

binding sites for the light intermediate chain dimers and the intermediate chain dimers. The intermediate chain plays an important scaffold role in the structure of the cytoplasmic dynein complex. In addition to binding the heavy chains, the intermediate chain binds the three light chain families, as dimers. It also binds other proteins including the p150 (Glued) subunit of dynactin, a protein that links dynein to its cargo. The intermediate chain has seven WD domains at its C-terminus which are necessary for binding to the heavy chain. The intermediate chain N-terminus is a region predicted to form a coiled coil and following it are the alternative splicing regions. The binding site for the dynactin subunit, p150, is in the N-terminal coiled coil region. The Tctex1 and LC8 light chains bind the intermediate chain just down from the alternative splicing region, while the Roadblock light chain binds upstream of the first of the WD domains. The intermediate chain dimerization domain is between the LC8 and Roadblock light chain-binding regions. The LC8 light chain is also a component of the axonemal outer arm dynein and the two-headed inner arm dynein. LC8 is not only associated with dyneins. It is also a component of myosin V, axonemal radial spokes, and it is important for the proper folding and function of many cytoplasmic proteins.

Cytoplasmic dynein 1 has many functions in cells. It is important for the transport of many different membranous organelles, maintaining Golgi position near the centrosome, endosome movement, and nuclear migration. During mitosis, dynein assists in the assembly of spindle poles, is used by kinetochores to move along microtubules, and generates pulling force from the cell cortex to separate the spindle poles.

In axons, cytoplasmic dynein is essential for retrograde membrane-bound organelle movement in fast transport and for the movement of microtubules and neurofilaments. Several viruses including adenovirus, herpes, and HIV also utilize cytoplasmic dynein to move along microtubules during the infection process. The cytoplasmic dynein cargo-binding domain at the far N-terminus of the heavy chain includes the intermediate chains, light intermediate chains, and the three light chain families. The best-characterized protein responsible for linking dynein to cargo is the dynactin complex; it is important for dynein binding to kinetochores and most membrane-bound organelles. Dynactin also increases cytoplasmic dynein processivity, the distance a dynein travels along a microtubule before detaching. Cytoplasmic dynein may also transport individual proteins which bind directly to the various subunits. The cytoplasmic dynein heavy chain, light intermediate chain, and intermediate chain subunits are phosphorylated *in vivo*. Light intermediate chain phosphorylation is correlated with cell-cycle regulation of dynein binding to membrane-bound organelles. Intermediate chain phosphorylation reduces its binding affinity for the p150 subunit of dynactin *in vitro*.

In fungi, deletion mutations of dynein subunit genes are not lethal, but mitosis is disrupted. Deletion of the cytoplasmic dynein heavy chain is lethal in multicellular organisms such as mouse and *Drosophila*. Nonlethal mutations of cytoplasmic dynein subunits result in axonal pathfinding defects and other pleiotropic developmental defects in *Drosophila*. A dominant mutation in the mouse heavy chain results in age-dependent progressive degeneration of motor neurons as well as pathfinding defects in heterozygotes. The protein encoded by human genetic brain disorder lissencephaly, Lis1, binds to and regulates the function of cytoplasmic dynein.

Cytoplasmic Dynein 2 (IFT Dynein)

The heavy chain of cytoplasmic dynein 2 was initially identified by molecular studies as a heavy chain closely related to the cytoplasmic dynein 1 heavy chain, yet one whose expression level was unregulated during flagellar synthesis. Further study showed that the primary function of cytoplasmic dynein 2 is to be the motor for one direction of IFT. It moves protein complexes from the flagellar tip to the base in the space between the outer double microtubules and the flagellar membrane. It is thus required for the formation of most flagella including sperm flagella. The cytoplasmic dynein 2 structure is similar to that of cytoplasmic dynein 1. It is homodimer of the heavy chain and has its own intermediate and light intermediate chains, although the only light chain component identified to date is LC8. The green alga *Chlamydomonas*, which does not contain a cytoplasmic dynein 1, but does have two flagella, has been a useful model system for the study of cytoplasmic dynein 2.

Axonemal Dyneins

Dyneins are the motor proteins that drive the bending of cilia and flagella. The axonemal dyneins are attached to the outer doublet microtubules of axonemes and generate the sliding

force between neighboring outer doublet microtubules that is the basis for the propulsive bending of flagella that move cells through fluids. The base of an axonemal dynein binds to one of the axonemal outer doublet microtubules, the A microtubule. A dynein head binds to the B microtubule of the neighboring outer doublet in an ATP-sensitive manner and utilizes the hydrolysis of ATP to push the B microtubule toward the tip of the axoneme. The arms are arranged in two rows along the microtubules, the outer and inner arms, and the dyneins in each row are different and have different functions. The outer arms specify the flagellar beat frequency and the inner arms specify wave form. *Chlamydomonas* has 14 axonemal dynein heavy chain genes, and eight different axonemal dynein complexes have been purified from their flagella to date. There is one outer arm dynein complex and seven well-characterized different inner arm complexes (Table 2).

Outer Arm Dyneins

The *Chlamydomonas* outer arm dynein is a heterotrimer of three different heavy chains; however in some species, such as sea urchin, the outer arm is a heterodimer. The outer arm dynein also has two different intermediate chains and up to 11 light chains. Outer arm dynein binds to the microtubule through the outer arm docking complex. Although they bind to different cargo, the intermediate chains of outer arm dynein are closely related to that of cytoplasmic dynein 1. The C-termini of the intermediate chains, the region which interacts with the heavy chains, are well conserved. However, their N-termini are different and that is thought to allow the binding to the different cargos. Also present in the outer arm dynein are: a member of the Tctex1 light chain family, LC9; the Roadblock family light chains, LC7a & b; and LC8 light chain family members, LC6, LC8, and LC10. The other outer arm dynein light chains have also been well characterized. LC3 and LC5 are thioredoxin homologs. LC4 is involved in the regulation of dynein by Ca^{+2} and LC2 is Tctex2. LC1 is a member of a leucine-rich repeat family, and unlike the other light chains, it binds to the motor domain of the gamma heavy chain, not to its N-terminus and it functions as a conformational switch to control outer arm activity.

Inner Arm Dyneins

The inner arm dyneins are very heterogeneous (Table 2). Inner arm dynein 1 has two different heavy chains, three different intermediate chains, and three light chains, including Tctex1 and LC8. The other six inner arm dyneins have a single unique heavy chain and associate with different sets of light chains. All associate with conventional actin and several associate with p28 and centrin. The functional significance of the association of actin with these dyneins is unknown.

To convert from dynein-driven microtubule sliding to flagellar bending, only a subset of the axonemal dyneins functions at one time. The axonemal dyneins must be regulated. The phosphorylation state of the inner arm dynein intermediate chain IC138 modulates this regulation. Genetic analysis has also identified several proteins that make up a dynein regulatory complex, though the mechanism of the complex remains unknown.

See also: [Cell Architecture and Function: Cell Migration; Cytoskeletal Motors: General Principles; Dynactin; Kinesin Superfamily Proteins; Kinesins as Microtubule Disassembly Enzymes; Microtubule-Associated Proteins; Mitosis; Myosin Motors; Tubulin and Its Isoforms.](#)

Further Reading

- Gibbons IR (1988) Dynein ATPases as microtubule motors. *Journal of Biological Chemistry* 263: 15837–15840.
- King SM and Kamiya R (2009) Axonemal dyneins: Assembly, structure, and force generation. In: Witman GW and Harris EH (eds.) *The Chlamydomonas Source Book. Vol. 3, Cell Motility and Behavior*, 2nd edn., pp. 31–208. Oxford: Elsevier/Academic Press.
- Pazour GJ, Dickert BL, and Witman GB (1999) The DHC1b (DHC2) isoform of cytoplasmic dynein is required for flagellar assembly. *Journal of Cell Biology* 144: 473–481.
- Pfister KK, Shah PR, Hummerich H, et al. (2006) Genetic analysis of the cytoplasmic dynein subunit families. *PLoS Genetics* 2: e1.
- Porter ME and Sale WS (2000) The 9 + 2 axoneme anchors multiple inner arm dyneins and a network of kinases and phosphatases that control motility. *Journal of Cell Biology* 27: F37–42.
- Roberts AJ, Numata N, Walker ML, et al. (2009) AAA+ ring and linker swing mechanism in the dynein motor. *Cell* 136: 485–495.
- Vallee RB, Shpetner HS, and Paschal BM (1989) The role of dynein in retrograde axonal transport. *Trends in Neuroscience* 12: 66–70.

Eicosanoid Receptors

R M Breyer and M D Breyer, Vanderbilt University, Nashville, TN, USA

Published by Elsevier Inc.

This article is reproduced from the previous edition, volume 2, pp. 6–9, © 2004, Elsevier Inc.

Glossary

Autacoid A class of physiologically active substances that act upon the cell that elaborated this compound or on adjacent cells.

Cyclooxygenase (COX) Enzymes that catalyze the oxidation of arachidonic acid. Also known as ‘prostaglandin H (PGH) synthase’. COX catalyzes two sequential reactions, a *bis* oxygenase, or cyclooxygenase, reaction leading to the formation of PGG₂ and a subsequent peroxidase activity at the C15 position leading to the conversion of PGG₂ to PGH₂.

Eicosanoid Any of the collections of oxygenated metabolites of the 20-carbon fatty acid, 5,8,11,14-eicosatetraenoic acid (arachidonic acid) that are the products of cyclooxygenase, cytochrome P450, lipoxygenase, and nonenzymatic pathways.

G-protein-coupled receptors (GPCRs) A class of integral membrane protein cell-surface receptors that change conformation upon binding cognate agonists. This conformational change causes the dissociation of heterotrimeric G proteins leading to the initiation of the signal transduction cascade.

Leukotriene Monooxygenated metabolites of arachidonic acid formed by the action of lipoxygenases.

Nonsteroidal anti-inflammatory drugs (NSAIDs) A class of drugs that effect their action by inhibiting the activity of cyclooxygenase. They have potent analgesic, antipyretic, and anti-inflammatory properties.

Prostanoids Oxygenated metabolites of the 20-carbon essential fatty acid arachidonic acid. They are analogs of the 20-carbon unnatural fatty acids, prostanoic, and thrombanoic acid, produced by the action of cyclooxygenase.

Eicosanoids are the oxygenated metabolites of the 20-carbon polyunsaturated fatty acid arachidonic acid. These compounds are locally acting autacoids and are rapidly metabolized with a half-life of minutes to seconds. Once formed, eicosanoids exert their physiologic effects via interaction with specific receptors. The best characterized of these receptors belong to the superfamily of G-protein-coupled receptors (GPCRs), although some eicosanoids are thought to interact with nuclear hormone receptors as well.

Eicosanoid Biosynthesis Pathways

The 5,8,11,14-eicosatrienoic acid, a 20-carbon polyunsaturated fatty acid with the trivial name arachidonic acid, is esterified in the lipid bilayer of cell membranes. Upon its enzymatic release by the action of phospholipase, free arachidonate is rapidly metabolized. Oxygenated metabolites of arachidonic acid are collectively known as the ‘eicosanoids’. These

compounds are formed by the action of three distinct enzymatic pathways.

The Lipoxygenase Pathway

This pathway leads to the formation of the monooxygenated compounds such as hydroperoxytetraenoic acids (HPETEs), hydroxytetraenoic acids, and trihydroxylated metabolites known as ‘lipoxins’ via three principal enzymes: 5-lipoxygenase (LO), 12-LO, and 15-LO. The 5-HPETEs are substrates for the leukotriene synthases that lead to the formation of a series of key inflammatory mediators, the leukotrienes, including the unstable intermediate LTA₄, which is converted to the potent mediator LTB₄ and the cysteinyl leukotrienes LTC₄, LTD₄, and LTE₄. Sequential oxidation of arachidonic acid by the action of both 12-LO and 5-LO or 12-LO and 15-LO leads to the formation of the lipoxins LXA₄ and LXB₄. The lipoxins are structurally related to the leukotrienes, but appear to act through a distinct set of receptors and have very different actions *in vivo*.

Finally, the 5-LO metabolite HPETE can be further metabolized to the 5-oxo-EETE (5-oxo-6E,8Z,11Z,14Z-eicosatetraenoic acid), for which a unique receptor has recently been cloned.

The Cyclooxygenase Pathway

Cyclooxygenase (COX) metabolism leads to the production of the five principal prostanoids via two distinct isozymes, COX-1 and COX-2. Four of the principal prostaglandins (PGs), PGE₂, PGD₂, PGF_{2α}, and PGI₂ are analogs of the 20-carbon unnatural fatty acid prostanoid acid, which is distinguished by its five-carbon prostane ring group comprised of carbons five through eight. The fifth prostanoid, thromboxane, has an inserted ether oxygen and thus has a six-member ring structure and is an analog of the unnatural fatty acid thrombanoic acid.

The Epoxygenase Pathway

Oxidation of arachidonic acid by the cytochrome P450 pathway leads to the formation of epoxytrienoic acids (EETs). Although physiologic effects have been attributed to these compounds, no specific EET receptors have been identified to date.

Eicosanoid Action

As a class, eicosanoids mediate a wide array of physiologic effects including pain, inflammation, and modulation of smooth muscle tone. Many of these effects are receptor-mediated by cell surface GPCRs. There is recent evidence that some PG metabolites can activate nuclear hormone receptors of the peroxisome-proliferation-activated receptor family (PPARs). In addition to the enzymatic products of arachidonic acid metabolism, prostanoid-like products resulting from nonenzymatic oxidation have also been described and designated isoprostanes. These products are structurally related to the COX-derived prostanoids, but are the product of nonenzymatically free radical oxidation. Isoprostanes are markers of oxidative stress and evidence suggests that these metabolites can also evoke receptor-mediated physiologic effects.

Receptors of Lipoxigenase Metabolites

The action of LO metabolites of arachidonic acid is mediated by a distinct family of GPCRs. Six GPCRs have been identified that bind LO metabolites of arachidonic acid: four leukotriene receptors designated BLT₁, BLT₂, CysLT₁, and CysLT₂ have been described, a lipoxin receptor designated ALX, and the receptor for 5-oxo-EETE.

Leukotriene Receptors

The leukotriene receptors fall into two groups of structurally related GPCRs: the chemoattractant-like BLT receptors and the CysLT receptors, which are related to the family of

nucleotide-binding receptors. Of the four cloned leukotriene receptors, BLT₁ and BLT₂ receptors have highest affinity for the leukotriene LTB₄ with BLT₁ binding LTB₄ with much higher affinity as compared to BLT₂. The BLT receptors are predominantly expressed in the peripheral blood including leukocytes, granulocytes, macrophages, and eosinophils. The CysLT₁ and CysLT₂ receptors in contrast each bind the cysteinyl leukotrienes LTC₄ and LTD₄ with high affinity. CysLT receptors are expressed in peripheral blood leukocytes, spleen, and lung as well in a number of other tissues at lower levels. The CysLT₁ receptor has been implicated as an important mediator in asthma, and antagonists of this receptor such as montelukast have been used clinically in the treatment of asthma.

Lipoxin Receptors

The lipoxin A₄ receptor (ALX), is a member of the GPCR family of chemoattractant receptors, as are the BLT receptors discussed above. This receptor was identified as an orphan receptor complementary DNA (cDNA) sharing significant homology with the formyl-methionine-leucine-phenylalanine (FMLP) chemoattractant receptor and was designated FMLP-like receptor 1 (FPLR1). Subsequent identification of this receptor as a mediator of LXA₄ action resulted in its designation as the ALX receptor. Activation of this receptor mediates the anti-inflammatory actions of lipoxin LXA₄. Northern blot analysis of ALXR messenger RNA (mRNA) in mouse tissues demonstrates that the receptor is most highly expressed in neutrophils, lung, and spleen with lower levels detected in the liver and heart. Some actions of lipoxins cannot be accounted for by the ALXR, and the existence of other lipoxin receptors has been postulated, but existence of other lipoxin receptor cDNAs has not been confirmed.

The 5-oxo-EETE Receptor

The orphan GPCR TG1019 and closely related R527 have been identified as receptors for 5-oxo-EETE. These receptors have no official designation at this time. These receptors have been shown to be highly expressed in the kidney and may mediate the observed hemodynamic effects of 5-oxo-EETE described in this tissue.

Prostaglandin Receptors

The local action of PGs depends in part, on activation of a family of specific GPCRs, designated EP for E-prostanoid receptors, FP, DP, IP, and TP receptors respectively, for the other prostanoids. The EP receptors are unique in that four receptors, designated EP1 through EP4, have been described for PGE₂ each encoded by a distinct gene. A second class of prostaglandin D receptor designated chemoattractant receptor-homologous molecule expressed on Th2 cells (CRTH2) has been identified, which has no sequence homology to the remaining PG receptors. This receptor has been unofficially designated

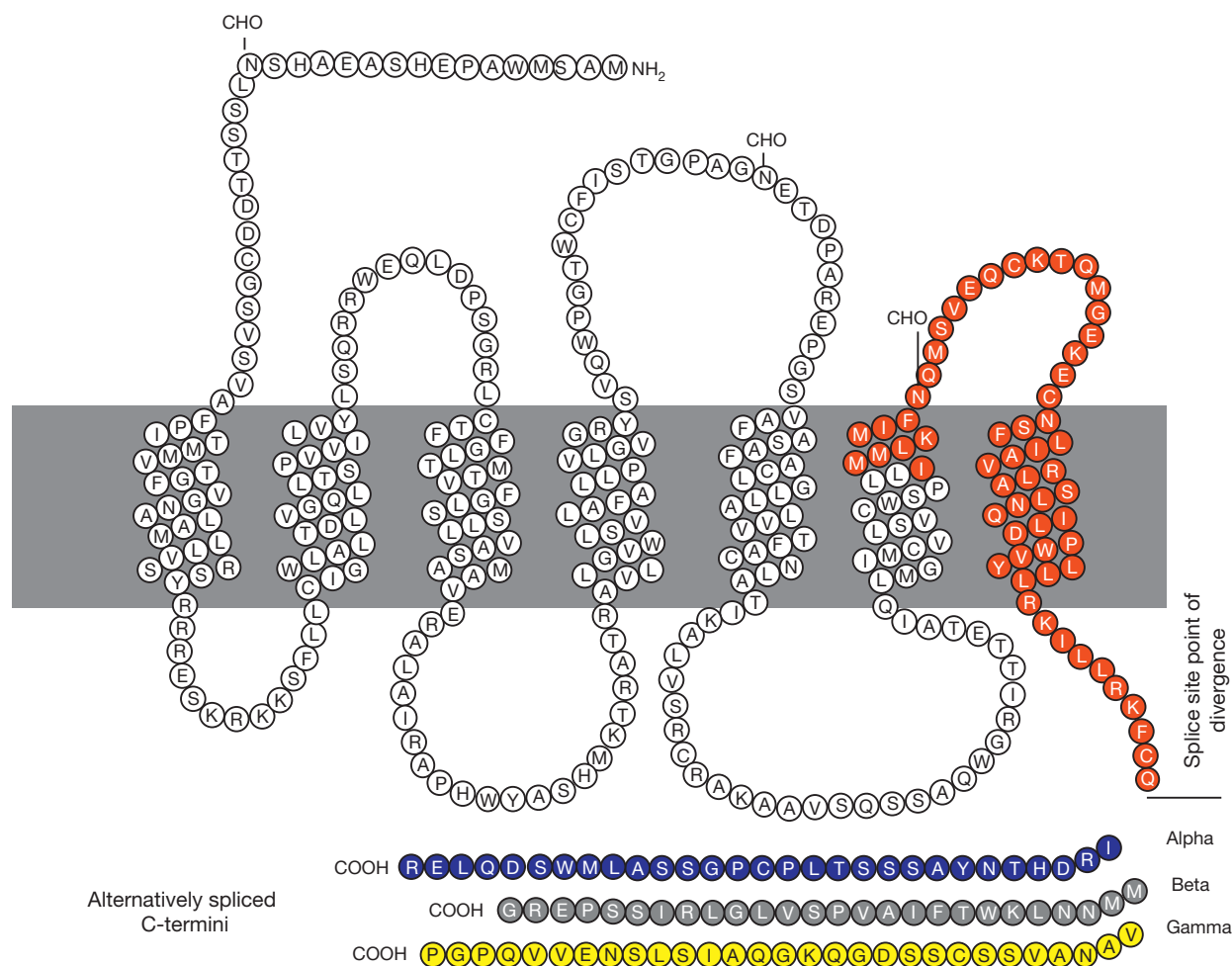


Figure 1 EP₃ receptor sequence of three mouse EP₃-receptor splice variants differing only in their intracellular carboxyl termini. The predicted amino acid sequences of each splice variant is represented by the one-letter amino acid code. Exons are color coded. The common region is comprised of two exons (in white and red circles) which are spliced to three possible C-terminal tails. The carboxyl variable tails are designated alpha (blue), beta (gray), and gamma (yellow), each encoded by distinct exons.

DP2. Each of the other PGs has a single receptor, and taken together there are nine PG GPCRs, each encoded by distinct genes. Alternative mRNA splice variants have been cloned for the EP1, EP3, TP, and FP receptors. In each case, these splice variants generate receptor sequence diversity in the intracellular C-terminal tail of the receptor protein (Figure 1). Functionally, these splice variants appear to modulate the specificity of G-protein coupling, as well as regulation of receptor desensitization by encoding alternate phosphorylation sites. Pharmacologically, the PG receptors are distinguished by their ligand-binding selectivity as well as the signal transduction pathway they activate. In general, PG receptors may have significant affinity for more than one prostanoid ligand. Moreover, multiple PG receptors are frequently coexpressed in a single cell type or tissue. COX activation and resulting PG production may lead to complex effects in the target tissue by activation of multiple PG receptor subtypes. Thus, a given PG ligand may elicit multiple, and at times apparently opposing, functional effects on a given target tissue. For example, PG receptors were initially characterized by their actions on smooth muscle,

where they may lead to either smooth muscle contraction or relaxation. The vasodilator effects of PGE₂ have long been recognized in both arterial and venous beds. Smooth muscle relaxation by PGE₂ is, however, not uniformly observed, and PGE₂ is a potent constrictor in other smooth muscle beds, including trachea, gastric fundus, and ileum. Importantly, some structural analogs of PGE₂ are capable of reproducing the dilator effects of PGE₂, but are inactive on tissues where it is a constrictor. Conversely, analogs that reproduce the constrictor effects of PGE₂ may fail to affect tissues where PGE₂ is a dilator. The EP receptor mRNAs exhibit differential expression in a number of tissues with distinct functional consequences of activating each receptor subtype. Functional antagonism among PGs can also be observed in platelet aggregation, where TXA₂ activation of the TP receptor causes platelet aggregation. Conversely, prostacyclin (PGI₂) activates platelet IP receptors which opposes this platelet aggregation. Thus, the balance of PGs synthesized as well as the complement of PG receptors expressed determines the net effect of COX metabolism on platelet function.

PPARs

In addition to the GPCR-mediated effects, there is evidence that PG metabolites are capable of activating some nuclear transcription factors. This action is exemplified by the cyclopentenone PGs of the J-series, for example, 15-deoxy- Δ 12, 14-PGJ₂ (15d-PGJ₂) which are derived from PGD₂. 15d-PGJ₂ activates a nuclear hormone receptor designated peroxisome proliferation activated receptor gamma. There is also evidence that PGI₂ also activates another member of this family designated PPARdelta. It is unclear whether the PGs represent true endogenous ligands of these receptors, as their affinity is generally in the micromolar range, which is two to three orders of magnitude lower than the nanomolar affinities observed for PGs at the GPCRs; nonetheless, these concentrations could be achieved in the intracellular environment.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Prostaglandins and Leukotrienes; **Molecular Biology:**

Eukaryotic DNA Polymerase δ ; Processivity Clamps in DNA Replication; **Signaling:** Peroxisome Proliferator-Activated Receptors.

Further Reading

- Breyer RM, Bagdassarian CK, Myers SA, and Breyer MD (2001) Prostanoid receptors: Subtypes and signaling. *Annual Review of Pharmacology and Toxicology* 41: 661–690.
- Brink C, Dahlen SE, Drazen J, et al. (2003) International union of pharmacology XXXVII. Nomenclature for leukotriene and lipoxin receptors. *Pharmacological Reviews* 55: 195–227.
- Forman BM, Tontonoz P, Chen J, et al. (1995) 15-Deoxy-delta 12, 14-prostaglandin J2 is a ligand for the adipocyte determination factor PPAR gamma. *Cell* 83: 803–812.
- Hosoi T, Koguchi Y, Sugikawa E, et al. (2002) Identification of a novel eicosanoid receptor coupled to Gi/o. *Journal of Biological Chemistry* 277: 31459–31465.
- Jones CE, Holden S, Tenaillon L, et al. (2003) Expression and characterization of a 5-oxo-6E,8Z,11Z,14Z-eicosatetraenoic acid receptor highly expressed on human eosinophils and neutrophils. *Molecular Pharmacology* 63: 471–477.
- Morrow JD and Roberts LJ (2001) Lipid-derived autacoids. In: Hardman JG, Limbird LE, and Gilman AG (eds.) *Goodman & Gilman's the Pharmacological Basis of Therapeutics*, 2nd edn. New York, NY: McGraw-Hill.

Elastin

J A Foster, Boston University School of Medicine, Boston, MA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 10–12, © 2004, Elsevier Inc.

Glossary

Cyanogen bromide A chemical reagent used to cleave peptide bonds adjacent to methionine residues.

Desmosine A pyridinium ring alkylated in four positions; it serves as a cross-link derived from four lysine residues.

Lathyrism A disease caused by ingestion of seeds of *Lathyrus odoratus*, a sweet pea. The active agent

is β -aminopropionitrile which inhibits lysyl oxidase.

PolyA tail A string of adenine nucleotides added to the 3' end of most eukaryotic messenger RNAs.

Promoter A DNA sequence that binds RNA polymerase resulting in transcription initiation.

Svedberg unit (S) A unit used for the sedimentation coefficient; equivalent to 10–13 s.

Elastin is a protein that exists as fibers in the extracellular spaces of many connective tissues. Elastin derives its name from its ability to act as an elastic band, that is, to stretch and recoil with transient force. It is located throughout many tissues and organs of higher vertebrates and plays an important functional role in maintaining pressures associated with liquid and air flow in the cardiovascular and pulmonary systems. Elastogenic cells synthesize and secrete a soluble monomeric form of elastin into the extracellular space. The enzyme, lysyl oxidase, initiates cross-linking of the soluble monomers into insoluble fibers. Extracellular elastin associates closely with other proteins in the matrix, including microfibrillar proteins and collagens. Once laid down in the matrix, the insoluble protein is very stable and resistant to degradation. Because of its critical role in the normal development and function of vital organs, either impairment of elastin synthesis or proteolytic degradation of the insoluble fibers results in major clinical pathologies.

Composition and Primary Sequence

Elastin, as any protein, possesses a unique amino acid composition and a unique sequence of those amino acid residues along the polypeptide chain.

Amino Acid Composition

The soluble form of elastin, sometimes referred to as tropo-elastin, contains a preponderance of uncharged and nonpolar amino acids. The polypeptide chain contains approximately 850 amino acids and a molecular mass of 65–72 kDa. The actual size varies somewhat among different species. Glycine (33%), alanine (18%), proline (13%), valine (17%), and leucine (5%) residues represent 86% of the total amino acids. Prolyl hydroxylase converts 1–2% of the proline residues to hydroxyproline but it is unclear as to the significance of this cotranslational modification. There are ~36–38 lysine residues per molecule, so the overall charge of monomer is basic with an isoelectric point over pH 10. Once lysyl oxidase deaminates the epsilon amino group of most lysine residues in the protein, the resulting semi-aldehydes undergo a series of aldol

condensations and Schiff bases to form the desmosine cross-links. At this stage, the protein is irreversibly insoluble and any success in extracting soluble fragments requires cleavage of peptide bonds.

Amino Acid Sequence

Unlike the classical triple repeat of collagen wherein glycine residues occupy every third position of the triplet, elastin does not contain a uniform repeating structure. Instead, elastin contains two broad sequence motifs that accommodate much subdivision. Most lysine residues segregate together with alanine residues to create pairs of lysines separated by two or three alanine residues. These sequences exist within stretches of uncharged, nonpolar amino acid residues such as glycine, valine, and proline. The latter amino acid residues occur together with other nonpolar, hydrophobic, and polar residues, form tri, tetra, penta, and hexa repeats with not a single consistent motif. The carboxy terminal sequence is unique and highly conserved among different species. It contains the only cysteine residues within the molecule.

Isolation of Elastin

Insoluble Elastin

An early operational definition of elastin was that it represented the protein remaining, after one extracted all other connective tissue proteins. Consequently, the insolubility of elastin serves as the basis for all isolation methods. Isolation procedures involve very harsh treatments such as autoclaving, extraction with hot alkali or strong denaturants, and exposure to cyanogen bromide or collagenases. Amino acid analysis of the material remaining insoluble after these treatments confirms the presence and homogeneity of the elastin.

Soluble Elastin

In vivo the conversion of soluble to insoluble elastin occurs rapidly, making the isolation of soluble elastin very difficult. In order to isolate quantitative amounts of soluble elastin from a tissue, one must block the conversion of the soluble to the

insoluble form. Inhibition of the cross-linking enzyme, that is, lysyl oxidase, is the most effective approach for isolating intact elastin monomers. As lysyl oxidase is a copper-dependent enzyme that is specific for lysine residues, it can be inhibited by removing copper or adding a noncompetitive inhibitor. In practice, this involves rendering an animal copper deficient by dietary restriction or by adding β -aminopropionitrile, a lathyrogen, to the diet of young animals. The procedure for isolating soluble elastin is based primarily on its unique solubility in organic solvents.

The Elastin Gene

Gene Structure

The single-copy, elastin gene spans ~35–40 kb of DNA but only 7% of the sequences represent exons that are transcribed into the messenger RNA (mRNA). The huge amount of intron sequences in the gene (19:1, intron to exon) makes this one of the most disperse genes reported. The short exon sequences (<190 bp) encode separate cassettes for the nonpolar and lysine/alanine residues of the protein. Many of the intron sequences encode elastin-like sequences.

Gene Promoter

The elastin gene promoter contains features of a constitutively expressed (housekeeping) gene. It lacks a classic TATA box, is very GC rich (67%), has two CAAT boxes, and possesses multiple transcription start sites. There are also several Sp1 sites as well as a cryptic GC box and several consensus AP2 sequences within the proximal promoter. In addition, consensus sequences for glucocorticoid, TPA-inducible, and CRE response elements reside in the distal promoter regions.

Elastin Gene Expression

Elastin mRNAs

Transcription of the elastin gene results in multiple mRNAs. All the multiple forms possess a size in the 24S range arising from alternate splicing of several different exons. All of the transcripts contain ~3.5 kb that includes a large untranslated 3' region of ~1.0 kb in addition to the polyA tail.

Elastin Protein Isoforms

Translation of multiple mRNA results in different isoforms of soluble elastin that reflect insertion of different exon sequences. Although there is variance in the presence of isoforms dependent on cell type and age, the function of the different variations of soluble elastin is still not clear. There is speculation that different isoforms play a major role in the early phase of elastic fiber assembly by providing interactive sites for specific binding to other components of the extracellular matrix.

Regulation of Elastin Gene Expression

The expression of elastin is high in developing tissues and through early growth to maturity. In the adult animal, elastin

synthesis is very low and its reinitiation occurs only in normal injury/repair situations or in some pathological conditions. Many studies have focused on understanding how elastin expression is regulated in development and disease conditions. The results show that expression is regulated at both the transcriptional and the posttranscriptional levels.

Transcriptional Regulation

Many studies have shown that elastin expression in developing pulmonary and cardiovascular tissues is controlled primarily at the transcriptional level. These results have been confirmed in transgenic animals carrying the transgene with a reporter driven by the elastin promoter and by *in vitro* tissue transfections of the same reporter gene. In addition, primary cultures of elastogenic cells demonstrate that elastin transcription plays a major role in the action of elastin modulators, such as basic fibroblast growth factor and insulin-like growth factor. Putative *cis*-acting elements in the 5' flanking region of the gene have been broadly defined through transient transfections of various cell types with a series of deletion reporter constructs. Of the *cis*-acting elements identified within the elastin promoter, few have been functionally shown to convey an increase in elastin transcription. In fact, majority were found to be repressors. These latter observations suggest that a possible route to increase elastogenesis is to block the repressor complex from binding via the disruption of the receptor and/or ligand binding, as well as blocking the resultant signal pathway.

Posttranscriptional Regulation

Studies on the expression of elastin in cultures of elastogenic cells challenged with biologically significant factors show that elastin mRNA stability is important in regulating expression, especially in the adult animal.

See also: **Metabolism Vitamins and Hormones: Amino Acid Metabolism; Protein/Enzyme Structure Function and Degradation: Collagens.**

Further Reading

- Franzblau C and Lent RW (1968) Studies on the chemistry of elastin. *Structure, Function, and Evolution of Proteins Brookhaven Symposia in Biology* 21: 358–377.
- Gallo PM and Paz MA (1975) Posttranslational protein modifications, with special attention to collagen and elastin. *Physiological Reviews* 55: 418–487.
- Gray WR, Sandberg LB, and Foster JA (1973) Molecular model for elastin structure and function. *Nature* 246: 461–466.
- Indik Z, Yeh H, Ornstein-Goldstein N, et al. (1989) Structure of the elastin gene and alternative splicing of elastin mRNA: Implications for human disease. *American Journal of Medical Genetics* 34: 81–90.
- Lansing AI, Rosenthal TB, Alex M, and Dempsey EW (1952) The structure and chemical characterization of elastic fibers as revealed by elastase and by electron microscopy. *Anatomical Record* 114: 555–575.
- Partridge SM, Davis HF, and Adair GS (1955) The chemistry of connective tissues. 3. Composition of the soluble proteins derived from elastin. *Biochemical Journal* 61: 11–21.
- Perrin S and Foster JA (1997) Developmental regulation of elastin gene expression. *Critical Reviews in Eukaryotic Gene Expression* 7: 1–10.
- Sandberg LB, Weissman N, and Smith DW (1969) The purification and partial characterization of a soluble elastin-like protein from copper-deficient porcine aorta. *Biochemistry* 8: 2940–2945.

Endocannabinoids

D Piomelli, University of California, Irvine, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 13–15, © 2004, Elsevier Inc.

Glossary

Arachidonic acid Common name for eicosatetraenoic acid (eicosatetraenoic acid, 20:4 $\Delta^{5,8,11,14}$), an essential fatty acid that serves as metabolic precursor for eicosanoids and endocannabinoids.

Axon terminal Specialized structure of a neuron that secretes neurotransmitters.

G proteins Heterotrimeric proteins with GTPase activity that link the occupation of certain cell surface receptors to cellular responses.

Neurotransmitter Substance secreted by a neuron at a synapse.

Phospholipase A group of enzymes that catalyze the hydrolysis of phospholipids at their glycerol ester (PLA) of phosphodiester (PLC, PLD) bonds.

Protein kinase Enzyme that transfers a phosphate group from adenosine triphosphate to a protein.

Stereospecific numbering (sn) A convention on how to designate the stereochemistry of glycerol-based lipids. When the glycerol moiety is drawn with the secondary hydroxyl to the left, the carbons are numbered 1,2,3 from top to bottom.

Striatum A subcortical brain structure involved in the control of movement, habit learning, and the rewarding properties of drugs of abuse.

Synapse Specialized junction between the ending of the presynaptic neuron and the dendrite, cell body, or axon of a postsynaptic neuron.

Vanilloid receptor A receptor channel permeable to monovalent cations and activated by heat, acid, and capsaicin, the active ingredient of chili peppers.

The endocannabinoids are a family of biologically active lipids that bind to and activate cannabinoid receptors, the G-protein-coupled receptors targeted by Δ^9 -tetrahydrocannabinol in marijuana. The term encompasses several derivatives of arachidonic acid, which are generated on demand by neurons and other cells in response to physiological or pathological stimuli. The two best-characterized endocannabinoids are anandamide (arachidonylethanolamide) and 2-arachidonoylglycerol (2-AG). Others are noladin ether (2-arachidonoyl glyceryl ether) and virodhamine (O-arachidonoyl ethanolamine).

Synthesis

Anandamide

Anandamide is produced from the hydrolysis of an N-acylated species of phosphatidylethanolamine (PE), called N-arachidonoyl-PE. This reaction is initiated by activating neurotransmitter receptors and/or by elevating intracellular levels of Ca^{2+} ions and is probably catalyzed by phospholipase D. The anandamide precursor, N-arachidonoyl-PE, is present at low levels in nonstimulated cells, but its formation can be stimulated by Ca^{2+} and occurs simultaneously with that of anandamide. N-Arachidonoyl-PE synthesis is catalyzed by membrane-bound N-acyltransferase, which has been partially purified (see [Figures 1](#) and [2](#)).

2-Arachidonoylglycerol

In brain neurons, 2-AG formation is probably initiated by the activation of phospholipase C, which cleaves membrane phospholipids (e.g., phosphatidylinositol-4,5-bisphosphate) at the proximal phosphate ester bond, producing 1,2-diacylglycerol.

This intermediate is broken down by diacylglycerol lipase to yield 2-AG and free fatty acid. Another pathway of 2-AG release might involve the hydrolytic cleavage of a phospholipid at the *sn*-1 position of the glycerol backbone, catalyzed by phospholipase A_1 . This reaction yields a *sn*-2 lysophospholipid, which can be further hydrolyzed to produce 2-AG ([Figure 3](#)). Finally, 2-AG might be formed by hormone-sensitive lipase acting on triacylglycerols or by lipid phosphatases acting on lysophosphatidic acid. However, as these enzymes preferentially target lipids enriched in saturated or monounsaturated fatty acids, they are unlikely to play a role in the synthesis of a polyunsaturated species such as 2-AG.

Physiological Regulation of Endocannabinoid Synthesis

Anandamide

Anandamide synthesis is initiated by intracellular Ca^{2+} rises and/or by activation of G-protein-coupled receptors. For example, activation of vanilloid receptors elevates intracellular Ca^{2+} levels and stimulates anandamide synthesis in rat sensory neurons in culture. In addition, activation of dopamine D_2 -receptors enhances anandamide release in the brain striatum of the rat *in vivo*. The molecular steps involved in these effects have not yet been clarified.

2-Arachidonoylglycerol

2-AG formation is also linked to intracellular Ca^{2+} rises. For example, in freshly dissected slices of rat hippocampus, electrical stimulation of the Schaffer collaterals (a glutamatergic fiber tract that projects from CA3 to CA1 neurons) produces a

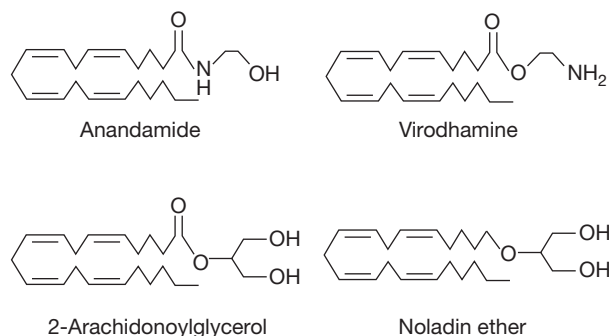


Figure 1 Chemical structures of the endocannabinoids anandamide (arachidonylethanolamide), 2-arachidonoylglycerol (2-AG), noladin ether, and virodhamine.

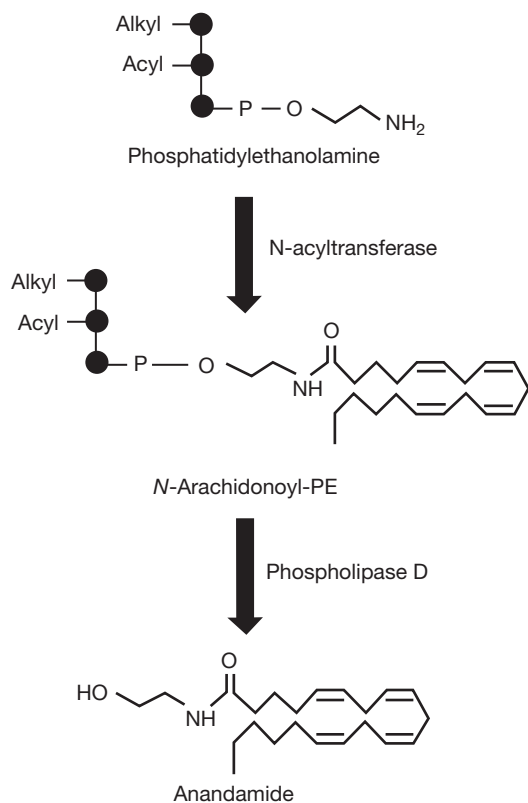


Figure 2 Anandamide biosynthesis. PE, phosphatidylethanolamine.

fourfold increase in 2-AG levels, which is prevented by the Na^+ channel blocker tetrodotoxin or by removing Ca^{2+} from the medium. It is notable that anandamide levels are not changed by the stimulation, suggesting that the syntheses of 2-AG and anandamide can be independently regulated. This idea is supported by the fact that activation of D_2 receptors, a potent stimulus for anandamide release in the rat striatum, has no effect on striatal 2-AG levels.

Deactivation

In the brain and other tissues, anandamide and 2-AG are rapidly eliminated through a two-step process consisting of

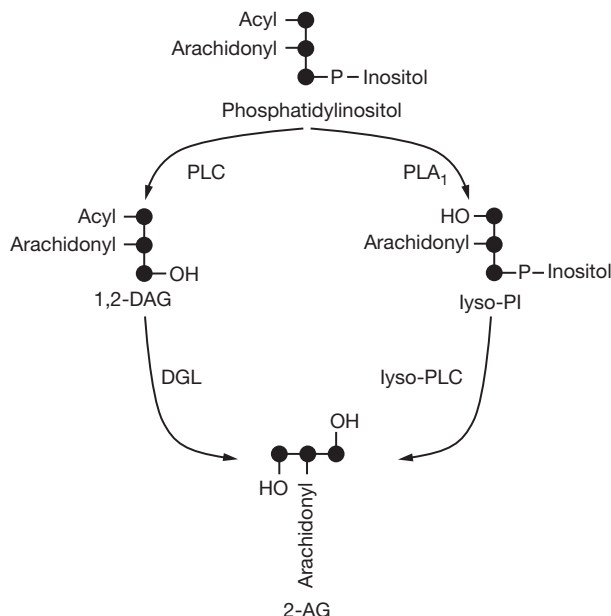


Figure 3 2-AG biosynthesis. 1,2-DAG, 1,2-diacylglycerol; DGL, diacylglycerol lipase; PL, phospholipase.

uptake into cells and enzymatic hydrolysis. The two endocannabinoids share a functionally similar transport mechanism, but they follow distinct routes of intracellular degradation.

Transport into Cells

The transport of anandamide and 2-AG into neurons and astrocytes is structurally specific, displays classical saturation kinetics, and is selectively inhibited by compounds such as *N*-(4-hydroxyphenyl)-arachidonamide (AM404). The putative transporter involved has not yet been identified, but transport has been shown to be Na^+ independent, which is suggestive of a facilitated diffusion mechanism.

Intracellular Hydrolysis

Inside cells, anandamide is metabolized by fatty acid amide hydrolase, a membrane-bound intracellular serine hydrolase that also cleaves oleoylethanolamide, an endogenous satiety factor, and other lipid amides. Hydrolysis of 2-AG is catalyzed instead by monoglyceride lipase, a cytosolic serine hydrolase that converts 2- and 1-monoglycerides into fatty acid and glycerol. The contribution of other lipases to 2-AG degradation cannot be excluded at present.

Cannabinoid Receptors

The endocannabinoids regulate the function of multiple organs and tissues of the body. These regulatory effects are primarily mediated by two G-protein-coupled receptors: CB_1 and CB_2 . CB_1 is highly expressed in the central nervous system, but is also present at lower levels in a variety of peripheral

tissues. By contrast, CB₂ is mostly found in immune cells such as lymphocytes. Both subtypes are linked to G_i/G_o proteins and can initiate signaling events typical of these transducing proteins, which include inhibition of adenylyl cyclase activity, opening of K⁺ channels, closing of Ca²⁺ channels, and stimulation of protein kinase activities. Nevertheless, CB₁ and CB₂ are structurally different (they have only 44% sequence homology), which has allowed the development of subtype-selective ligands such as the CB₁ antagonist SR141716A (rimonabant) and the CB₂ antagonist SR144528. There is evidence for the existence of at least two additional cannabinoid-sensitive sites in the brain and vasculature, which remain however uncharacterized.

Functions

In broad terms, the endocannabinoids are considered paracrine mediators, substances that act on cells near their sites of synthesis without entering the bloodstream. For example, they are formed by circulating leukocytes and platelets during hypotensive shock and induce the vascular relaxation that accompanies this phenomenon by activating CB₁ receptors on the surface of smooth muscle cells. Similar paracrine actions occur

in the brain, where the endocannabinoids mediate a localized signaling mechanism through which neurons modify the strength of incoming inputs. The endocannabinoids are generated by neuronal depolarization and travel backward across the synapse to regulate the release of neurotransmitters from neighboring axon terminals, a process called 'retrograde signaling'.

See also: **Signaling:** Adenylyl Cyclases; G-Protein-Coupled Receptor Kinases and Arrestins; Neurotransmitter Transporters; Phospholipase C.

Further Reading

- Devane WA, Hanus L, Breuer A, et al. (1992) Isolation and structure of a brain constituent that binds to the cannabinoid receptor. *Science* 258: 1946–1949.
- Di Marzo V, Fontana A, Cadas H, et al. (1994) Formation and inactivation of endogenous cannabinoid anandamide in central neurons. *Nature* 372: 686–691.
- Piomelli D (2003) The molecular logic of brain endocannabinoid signaling. *Nature Neuroscience Reviews* 4: 873–884.
- Wagner JA, Varga K, and Kunos G (1998) Cardiovascular actions of cannabinoids and their generation during shock. *Journal of Molecular Medicine* 76: 824–836.

Endocytosis

J G Donaldson, National Institutes of Health, Bethesda, MD, USA

Published by Elsevier Inc.

Glossary

Caveolae Flask-shaped invaginations from the plasma membrane that are enriched in cholesterol and glycosphingolipids and are coated with the protein caveolin.

Clathrin Cytosolic protein that polymerizes into basket-like coat complex on membranes, facilitating deformation of the plasma membrane to form a vesicle.

Dynamin GTP-binding protein that assembles at the necks of forming vesicles and facilitates the fission of membranes to release the vesicle.

Phagocytosis Endocytic process involving particle ingestion through extension of cell-surface membranes around particle (typically >250 nm in diameter) and membrane fusion to form phagosome.

Pinocytosis Endocytic process involving invaginations of the plasma membrane to generate small (100 nm in diameter) intracellular vesicles.

Endocytosis is a process carried out by all eukaryotic cells that involves the invagination of the cell-surface membrane and constricted closure to form a vesicle that enters the cell interior. Included in the internalized vesicle are extracellular fluid and plasma-membrane proteins and lipids. Endocytosis enables the cell to take up extracellular nutrients, remove activated receptors from the cell surface, and turn over cell-surface proteins and lipids. Endocytosis is also used for the removal of extracellular debris such as dead cells and bacteria, and as a cellular port of entry for infectious bacteria and viruses. Endocytosis can be broadly divided into pinocytosis (cell drinking) and phagocytosis (cell eating) (Figure 1). There are several types of pinocytosis. The best-characterized type is that involving endocytosis of vesicles coated on the cytoplasmic surface with a protein called clathrin. Phagocytosis involves ingestion of particles by extension and wrapping of the plasma membrane around the particle, bringing it into the cell. After internalization, the endocytosed material meets different fates: degradation, recycling back to the cell surface, or routing to other destinations within the cell.

Clathrin-Mediated Endocytosis

Endocytosis that is associated with clathrin coating enables certain plasma-membrane proteins to be concentrated and efficiently internalized (Figure 1(a)). The transferrin receptor and low-density lipoprotein (LDL) receptor are cell-surface proteins that are internalized by this mechanism, carrying iron-loaded transferrin protein and cholesterol-bearing LDL particles bound to these receptors into the cell. The clathrin coat that assembles on these structures forms a distinctive basket-like structure that can be visualized by electron microscopy. The polymerization of the clathrin coat is believed to facilitate changes in membrane curvature, deformation of the surface membrane, and formation of the vesicle. Once the vesicle has separated from the cell surface (fission), the clathrin coat is rapidly shed and then reutilized

for additional endocytic events. Clathrin-coated pits cover 2% of the plasma membrane and since they are believed to have a lifetime of ~1 min, ~2% of the cell-surface membrane is internalized each minute in the average cell. Clathrin-mediated endocytosis accounts for the bulk of pinocytosis in most cells. Clathrin assembly onto membrane requires the prior binding of a set of cytosolic adaptor proteins onto the plasma membrane.

Adaptor Proteins

The clathrin adaptor protein 2 complex (AP2) is composed of four different cytosolic proteins that bind to specific lipids and proteins at the plasma membrane and to clathrin, thus initiating clathrin assembly at distinct sites along the cell surface. There are three additional AP complex proteins in cells that associate with other cellular membranes, such as the Golgi complex, where they also facilitate clathrin assembly. One of the subunits of AP2 specifically binds to tyrosine-based, amino-acid sorting signals in the cytoplasmic regions of plasma-membrane proteins. The transferrin receptor contains such a tyrosine signal that sorts it into forming clathrin-coated PM vesicles, enabling the transferrin receptor to be efficiently and rapidly internalized into cells. In addition, there are other clathrin adaptor proteins that function at the PM, including disabled 2 (Dab2), and the autosomal recessive hypercholesterolemia protein (ARH), which is involved in the internalization of the LDL receptor. The ability to concentrate specific surface cargo proteins into clathrin-coated pits is the hallmark of clathrin-mediated endocytosis (Figure 2). There are other aminoacid sorting signals that allow for rapid internalization via clathrin endocytosis, including paired leucine and/or isoleucine residues. Plasma-membrane proteins that contain these sorting sequences are continually brought into cells by clathrin-mediated endocytosis. Some hormone receptors such as the β 2-adrenergic receptor lack these sorting sequences and do not appreciably enter clathrin-coated vesicles in the absence of hormonal stimulation. However, when the receptor is activated

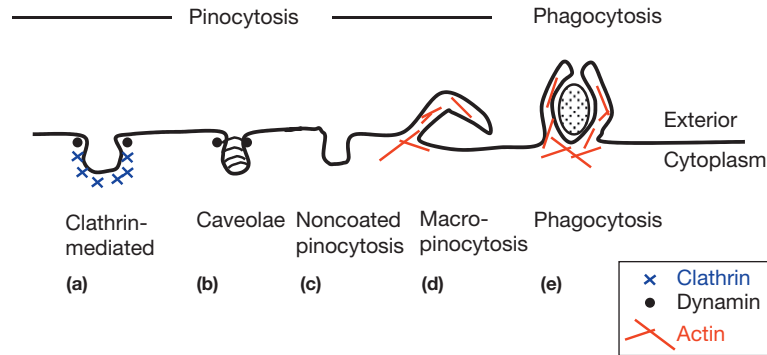


Figure 1 Types of endocytosis. Features of the various types of endocytosis (pinocytosis and phagocytosis) are shown. (a) Clathrin-mediated pinocytosis is associated with a cytosolic coat of clathrin (X) and uses dynamin (black spots) for vesicle fission. (b) Caveolae are flask-shaped structures coated with caveolin that also use dynamin for vesicle fission. (c) Noncoated pinocytosis involves invagination of membrane that does not have a recognizable protein coat. (d) Macropinocytosis is associated with cell-surface actin-dependent protrusions and forms by fusion of the protruding membrane back onto the cell surface. (e) Phagocytosis is a process that is stimulated by an extracellular particle that induces actin-dependent cell extensions to envelope the particle.

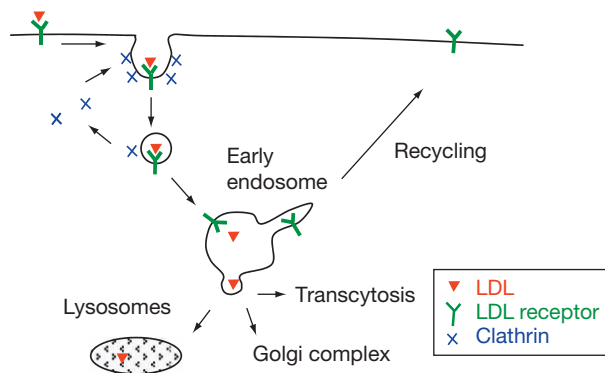


Figure 2 Itinerary of internalized cell-surface receptors and vesicle contents. Shown here is the pathway followed by internalized LDL bound to the LDL receptor. Cell-surface LDL receptor binds to LDL and is sorted into forming clathrin-coated vesicles at the plasma membrane. After vesicle fission, the clathrin coat is rapidly released from the vesicle so it can assemble again onto the plasma membrane. The vesicle membrane then fuses with the early endosome, a membrane compartment that serves as a sorting station for membrane proteins and lipids. The low pH of the early endosome causes release of LDL from its receptor. The LDL particle and other molecules present in the fluid portion of the early endosome can then be routed to lysosomes for degradation. The LDL receptor enters regions of the early endosome that break off and carry these receptors and membrane back to the plasma membrane. Some membrane proteins leave the early endosome and are transported to other regions of the plasma membrane (transcytosis) or to the Golgi complex.

by a hormone, a special adaptor protein called β -arrestin binds to the receptor, allowing the receptor–arrestin complex to associate with AP2 and be rapidly internalized by the clathrin-dependent pathway.

Dynamin

The final step in endocytosis is membrane fission that allows the vesicle to detach from the cell surface. A guanosine triphosphate (GTP)-binding protein called dynamin facilitates this

process by assembling into ring-like structures and constricting at the necks forming clathrin-coated vesicles. Mutations in dynamin, originally described for *Drosophila melanogaster* mutants, freeze clathrin-forming vesicles at the constricted state and the vesicles are not released. All clathrin-mediated and some other types of endocytosis are dependent upon the function of dynamin. Dynamin also associates with other factors that may facilitate the fission process. Rearrangements in membrane lipids and changes in lipid composition also contribute to the changes in membrane curvature required for vesicle budding and vesicle fission.

Clathrin-Independent Endocytosis

Several other types of pinocytosis have been identified that are not associated with clathrin coats. We know little about the mechanisms of these processes and how they are regulated. Plasma-membrane proteins that lack the cytoplasmic amino-acid sequences specifying association with adaptor proteins and clathrin can be endocytosed into cells by these processes.

Caveolae

Some proteins associate with cholesterol and glycosphingolipid-enriched regions of the plasma membrane referred to as membrane raft domains. In many cells, especially in endothelial cells that line blood vessels, these raft domains are present in flask-shaped invaginations of the plasma membrane. These flask-shaped invaginations are called caveolae (caves), and the cytoplasmic coat protein associated with these structures that give them their distinctive shape is called caveolin (Figure 1(b)). Caveolar endocytosis is dependent upon dynamin and is used as an endocytic mechanism for the transport of materials across endothelial cells. Interestingly, simian virus 40 stimulates caveolar endocytosis and enters many types of cells through this mechanism. Finally, there are cells that lack caveolin and, hence, caveolae and yet endocytose proteins that associate with membrane raft domains.

Noncoated Endocytosis and Macropinocytosis

Plasma-membrane proteins and lipids are also observed in invaginations budding from the membrane that lack any identifiable protein coat (Figure 1). This process may or may not require dynamin function depending upon the cell type. This type of endocytosis is only beginning to be studied. Although this process may account for a smaller fraction of cell-surface endocytosis than clathrin-mediated endocytosis, it can be greatly stimulated when cells become activated and form membrane ruffles and protrusive structures. In this case, the membrane ruffle can fuse with the plasma membrane forming a large macropinosome that internalizes a large amount of surrounding fluid. Dendritic cells of the immune system are very active in macropinocytosis as a means of sampling foreign material in the extracellular fluid. Macropinosome formation is actin dependent. Some bacteria, notably *Salmonella typhi*, stimulate host cells to form membrane ruffles and macropinosomes, and then use these structures as a port of entry into the cells.

Phagocytosis

In lower eukaryotes, such as protozoa, this process of particle ingestion is used for nutrient uptake. However, in mammals, phagocytosis is observed in specialized cell types, namely, macrophages and neutrophils that ingest bacteria and dead cells and degrade them. When a macrophage encounters a bacterium that has been coated with antibodies generated by the immune system, cell-surface receptors that recognize the Fc portion of the antibodies are activated. This stimulates changes in the macrophage actin cytoskeleton and leads to the extension of membrane around the bacteria, and final closure of the structure through membrane fusion (Figure 1(e)). Macrophages are also active in phagocytosing senescent-red blood cells and other cells that have undergone programmed cell death. It is believed that macrophages can detect alterations on the cell surface of the senescent cell, such as appearance of a phospholipid on the outer surface that is normally present only on the cytoplasmic side of the plasma membrane. After phagocytosis, the resulting phagosome usually fuses with lysosomes and membrane-bound organelles that contain hydrolytic enzymes that degrade the bacteria. However, some bacteria such as *Mycobacterium tuberculosis* can actively prevent the fusion of the phagosome with lysosomal membranes, and hence escape degradation.

Fate of Internalized Membrane and Content

After endocytosis, the endocytic vesicle usually fuses with a membrane-bound structure called the early endosome (Figure 2). The early endosome accepts newly endocytosed material and serves as a sorting station that directs incoming proteins and lipids to their final destination. The early endosomal membrane contains a special membrane lipid, phosphatidylinositol 3-phosphate, which recruits cytosolic proteins onto these membranes and facilitates the fusion of incoming vesicles with the early endosome. The interior, luminal environment of the early endosome is moderately acidic (pH 6.5–6.0). The low pH facilitates the sorting process, releasing bound proteins

from their receptors and sorting membrane proteins and lipids into different domains. A good example of how low pH serves this function is the itinerary of the transferrin receptor that carries iron-saturated transferrin into the cell. The acidic environment of the early endosome causes the release of iron from transferrin. The iron is then transferred across the membrane into the cytosol, while the transferrin bound to the transferrin receptor is recycled back to the cell surface via membrane carriers that are released from the early endosome. Another example of this sorting of incoming material is the fate of internalized LDL particles bound to the LDL receptor (shown in Figure 2). LDL is released from the receptor in the early endosome and is transferred with other luminal material to late endosomes and lysosomes. There the LDL is degraded, causing cholesterol release inside the cell. Meanwhile, the LDL receptor returns to the cell surface to carry out additional rounds of endocytosis. For the transferrin and LDL receptors, endocytosis is a mechanism for the delivery of iron and LDL into the cell. By contrast, for signaling receptors, such as those activated by hormones and growth factors, endocytosis is generally used to halt the signaling process by the removal of the receptor from the cell surface. However, recent observations suggest that some signaling receptors, such as the epidermal growth factor receptor, require endocytosis for certain signaling activities.

Endocytosed membrane proteins can be transported to different locations in the cell (Figure 2). Some membrane proteins are sorted into vesicles that fuse to another region of the plasma membrane, a process called transcytosis. Mucosal immunoglobulins such as IgA are transcytosed across intestinal epithelia via IgA receptors that ferry IgA from the basal (serosal) side to the apical (mucosal) surface of the epithelia. Other proteins leave the early endosome in vesicles that fuse with membranes adjacent to the *trans*-Golgi complex.

Regardless of how a membrane is brought into the cell, it must be balanced by the addition of membrane back to the plasma membrane. This is accomplished by endosomal-membrane recycling and also through the contribution of newly synthesized membranes from the secretory pathway. The cell has to regulate and coordinate the flux of membrane into and out of the cell to maintain cell-surface area and for proper localization of important cell-surface receptors.

See also: Lipids Carbohydrates Membranes and Membrane Proteins: Glycoprotein Folding and Processing Reactions.

Further Reading

- Alberts B, Johnson A, Lewis J, et al. (2002) *Molecular Biology of the Cell*, 4th edn. New York, NY: Garland Science.
- Conner SD and Schmid SL (2003) Regulated portals of entry into the cell. *Nature* 422: 37–44.
- Marsh M (ed.) (2001) *Endocytosis*. New York, NY: Oxford University Press.
- Scita G and Di Fiore PP (2010) The endocytic matrix. *Nature* 463: 464–473.
- Slepnev VI and DeCamilli P (2000) Accessory factors in clathrin-dependent synaptic vesicle endocytosis. *Nature Reviews. Neuroscience* 1: 161–172.
- Sorkin A and von Zastrow M (2002) Signal transduction and endocytosis: Close encounters of many kinds. *Nature Reviews. Molecular Cell Biology* 3: 600–614.
- Traub LM (2009) Tickets to ride: Selecting cargo for clathrin-regulated internalization. *Nature Reviews. Molecular Cell Biology* 10: 583–596.

Endoplasmic Reticulum-Associated Protein Degradation

R Bernasconi, Institute for Research in Biomedicine, Bellinzona, Switzerland

M Molinari, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Maurizio Molinari, volume 2, pp. 20–23, © 2004, Elsevier Inc.

Glossary

Calnexin and calreticulin Calnexin is a nonglycosylated type I protein of the endoplasmic reticular (ER) membrane with 573 amino acids. Most of them are present in the lumen of the ER and comprise a glucose-binding site for transient association with folding substrates and a P-domain for noncovalent association with the co-chaperone ERp57; 89 residues are cytosolic and contain a RKPRRE ER-retention sequence. Calreticulin is the soluble homolog of calnexin; it has 400 amino acids with a KDEL (Lys–Asp–Glu–Leu) ER-retrieval sequence.

EDEM proteins EDEM1 (for ER degradation enhancing alpha mannosidase-like protein), EDEM2, and EDEM3 are ER-resident soluble glycoproteins involved in the disposal of misfolded polypeptides from the ER lumen. They belong to the glycosyl hydrolase 47 (GH47) family. Their intracellular level increases in response to accumulation of misfolded proteins in the ER lumen. Some controversy exists as to

their function in the ER lumen. In particular, it is disputed if they are mannose-binding lectins or mannose-processing enzymes. So far, it has been demonstrated that at least EDEM1 is an active mannosidase.

ERAD ER-associated protein degradation collectively describes the processes leading to recognition of misfolded proteins present in the ER lumen and their dislocation into the cytosolic compartment for proteasome-mediated destruction.

Lectin Lectins are carbohydrate-binding proteins, which are often classified based on their binding specificity. They are involved in many biological processes, including polypeptide folding, transport, and ERAD.

Quality control It describes the capacity of cells to distinguish and retain in the ER lumen, to be eventually targeted for degradation, proteins that have structural defects.

For proper functioning, a protein has to fold into a characteristic three-dimensional structure and may have to form a multimeric complex. The endoplasmic reticulum (ER) provides an environment optimized for the folding and assembly of proteins that will then be either secreted (hormones, antibodies, pancreatic enzymes, etc.) or displayed at the plasma membrane (receptors, channels, etc.), or that reside in intracellular compartments, such as the lysosomes and the Golgi. Only native proteins leave the ER to be transported along the secretory pathway to their final destination. A quality-control system is in place to prevent forward transport of polypeptides for which folding and assembly have failed. To prevent their accumulation in the lumen of the ER, terminally misfolded polypeptides are dislocated into the cytosol to be degraded by 26S proteasomes. This occurs in a series of events collectively named 'ER-associated protein degradation (ERAD)'. This process is regulated by luminal, membrane-embedded, and cytosolic components that interrupt futile folding attempts, direct misfolded proteins at specific sites in the ER membrane, facilitate their dislocation across the membrane, and insure polyubiquitylation that facilitates extraction from the ER and proteasomal intervention. The aim of this article is to summarize recent advances in understanding how misfolded glycoproteins are cleared from the ER lumen of mammalian cells.

Protein Synthesis and Folding in the ER

In mammalian cells, secretory proteins are synthesized by ribosomes associated at the cytosolic face of the ER at an average

rate of 3–5 amino acids per second. Nascent chains are translocated into the ER lumen and most of them become N-glycosylated upon addition of pre-assembled, branched oligosaccharides (N-glycans) at asparagines in Asn–Xaa–Ser/Thr sequons. The highly hydrophilic N-glycans preserve the solubility of yet unstructured chains likely to expose hydrophobic, aggregation-prone patches. Most importantly, they act as ligands for calnexin and calreticulin, two lectin chaperones that make nascent chains accessible to ERp57. ERp57 is an oxidoreductase that catalyzes the formation of disulfide bonds, a rate-limiting step of the folding process peculiar for proteins expressed in the ER. As thoroughly reviewed elsewhere (see section Further Reading), calnexin, calreticulin, and ERp57, together with the sugar-processing enzymes glucosidase I and II and uridine diphosphate (UDP)-glucose:glycoprotein glucosyltransferase (GT) compose the calnexin/calreticulin cycle, a chaperone network that assists folding and quality control of glycoproteins.

The kinetics and the efficiency of the folding process are highly variable and depend on the substrate–protein considered and on the cell type where folding occurs. Somewhat surprisingly, the folding of heterologous proteins might be faster and significantly more efficient compared to the maturation of cell self-proteins. Hence, it may take only 10 min for a viral glycoprotein (e.g., influenza virus hemagglutinin) to complete the folding and oligomerization processes and to leave the ER with an efficiency approaching 100%. On the other hand, it may take several hours before roughly 20% of the cystic fibrosis transmembrane conductance regulator (CFTR) expressed in the ER lumen becomes native and leaves the compartment for the cell surface. About 80% of the wild-type

CFTR never attains the native structure and is degraded shortly after synthesis. A relatively low folding efficiency is a common feature for several eukaryotic gene products and has no consequences. However, folding efficiency and transport of the polypeptide at the site of activity may drop substantially, and can be the cause of impaired cell, organ, and/or organism viability when genetic mutations occur that change or delete amino acids. If the mutated protein cannot fold properly, it will not fulfill the quality-control standard for ER exit and will eventually become a substrate of the ERAD machinery or will accumulate intracellularly in toxic aggregates. For example, deletion of phenylalanine at position 508 of CFTR (Δ Phe508 CFTR) results in 99% of the protein being incapable of attaining the native structure. This results in a loss-of-function disease (cystic fibrosis) caused by the absence of functional CFTR at the cell surface.

Protein disposal from the ER also plays an important regulatory role, as exemplified by hydroxymethylglutaryl (HMG)-CoA reductase, a rate-limiting enzyme for the production of sterols and a variety of essential isoprenoids, whose intracellular level is controlled, at least partially, by regulated degradation.

Diseases Related to Failure of Protein Folding and Quality Control in the ER

An increasing number of strongly debilitating human diseases have been ascribed to defective protein folding and quality control. Together with cystic fibrosis, one could mention the hereditary lung emphysema and liver failure caused by α 1-antitrypsin deficiency, diabetes mellitus caused by retention and degradation from the ER of mutant insulin receptors, familial hypercholesterolemia (low-density lipoprotein receptor), osteogenesis imperfecta (type I procollagen), retinitis pigmentosa, and several neurodegenerative diseases. In these so-called 'conformational diseases', mutated proteins that do not acquire their native structure are retained in the ER and are diverted to ERAD. If their disposal is not efficient, they accumulate intra- or extracellularly, triggering severe damages to cells and tissues. In some cases, such as the hereditary lung emphysema caused by α 1-antitrypsin deficiency (α 1-antitrypsin is the principal blood-borne inhibitor of the destructive neutrophil elastase in the lungs), the lack-of-function phenotype observed at the level of patient's lungs is accompanied by a gain-of-toxic-function phenotype at the level of the patient's liver because the mutated protein accumulates in the ER of hepatocytes.

Degradation of Aberrant Proteins Expressed in the ER: The Mannose Timer

How are polypeptides recognized as being misfolded, retained in the ER, and targeted for ERAD? First, it has to be mentioned that the quality-control machinery of the ER checks the structure and not the function of newly synthesized proteins. Therefore, function-competent proteins can be retained in the ER and destroyed because they are structurally imperfect. This is the case, as one example, of the Δ Phe508 CFTR.

Second, modification of protein-bound oligosaccharides determines the fate of newly synthesized glycoproteins within

the ER. N-Glycans are added as pre-assembled units composed of two *N*-acetyl-glucosamine, nine mannose, and three glucose residues. The ER-resident glucosidase I and glucosidase II sequentially remove the two outermost glucose residues in a matter of seconds after addition of the N-glycan to nascent polypeptide chains. In the minutes following the synthesis of a glycoprotein, the number of glucose residues on N-glycans can vary from 0 to 1 in response to the counteracting enzymatic activities of glucosidase II and GT. These modifications control cycles of dissociation from/association with the lectin chaperones calnexin and calreticulin. How long newly synthesized proteins are retained in the calnexin/calreticulin cycle depends on their rate of folding. As GT recognizes only pseudo-native structure, immature polypeptides are repeatedly re-glucosylated and retained in the calnexin/calreticulin cycle, whereas native glycoproteins exit the folding cycle, leave the ER, and are secreted.

In this model, proteins unable to fold properly persist in the calnexin/calreticulin cycle. To prevent accumulation of aberrant polypeptides in the ER, terminally misfolded polypeptides have to be released from the calnexin/calreticulin cycle and to be targeted for disposal and degradation. After several unsuccessful folding attempts, N-glycans become substrates of several slow-acting ER-resident mannosidases. Therefore, long retention in the ER as signaled by a reduced number of mannose residues on N-glycans is a clear symptom of a problematic folding. It was originally proposed by Helenius in 1994 that a timer based on mannose cleavage exists in the ER that decides when it is time to stop folding attempts and to direct folding-defective polypeptides for degradation. The involvement of mannosidases in ERAD is supported by findings showing that cleavage of mannose residues from N-glycans facilitates extraction of misfolded polypeptides from the calnexin/calreticulin cycle and is essential for disposal of misfolded polypeptides.

ER-Resident α 1,2-Mannosidases

In mammalian cells, extensive de-mannosylation is performed by one or more members of the glycosyl hydrolase family 47 (GH47) of α 1,2 mannosidases. This family comprises the ERManI, EDEM (ER degradation enhancing α mannosidase-like protein) family members (EDEM1-3), and a collection of Golgi-resident enzymes. Recent models claim that ERManI removes the terminal mannose from the middle branch of N-glycans and EDEM proteins (at least EDEM1 and EDEM3) further process the oligosaccharide to eventually generate a Man₅₋₆ structure that does not support binding to the folding lectins calnexin and calreticulin anymore. Misfolded polypeptides displaying de-mannosylated oligosaccharides are therefore extracted from the folding machinery and redirected into the disposal machinery.

Distinct ER Chaperone Networks Are Involved in the ERAD Process

Involvement of at least two distinct chaperone networks, the calnexin/calreticulin cycle and the BiP/protein disulfide isomerase (BiP/PDI) system, has been shown to regulate degradation of ERAD candidates in mammalian cells. Interestingly, misfolded glycoproteins remain in the calnexin/calreticulin

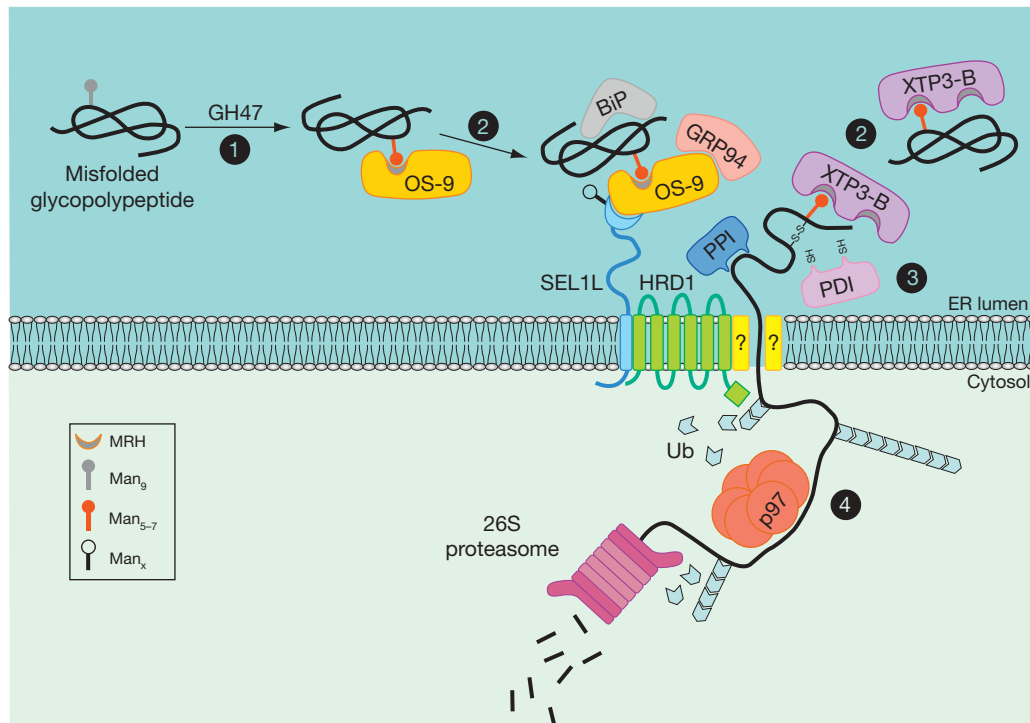


Figure 1 The ERAD mechanism as depicted in this figure can be divided into four steps: first, the folding-incompetent glycoprotein is extracted from the calnexin cycle through the concerted activity of several mannose-trimming enzymes (GH47); second, the lectin ERAD shuttles OS-9 and XTP3-B transport the misfolded glycoprotein from the ER lumen to the membrane-anchored adaptor protein SEL1L, which is associated with the E3 ubiquitin ligase HRD1; third, the misfolded glycoprotein has to be unfolded (e.g., reduction of aberrant disulfide bonds and *cis/trans* isomerization of peptidyl–prolyl bonds) in order to; and fourth, allow dislocation across the ER membrane through an elusive channel, polyubiquitylation and degradation by the 26S proteasome. The different mannose composition of the N-glycan is depicted in color code: gray, oligosaccharide with nine mannose residues (Man₉); red, five to seven mannose residues (Man₅₋₇); and white, oligosaccharide with an undefined number of mannose residues.

cycle when degradation is blocked by mannosidase inhibitors, whereas they are released from the lectin chaperones to remain trapped in the BiP/PDI system when the proteasome activity is specifically inhibited. These data show that the two chaperone networks work sequentially and offer the appealing view that ERAD candidates are first subjected to folding attempts in the calnexin/calreticulin cycle (lag between synthesis and degradation onset) to be then transferred to the BiP/PDI system that would regulate unfolding and dislocation of the aberrant polypeptide into the cytosol for proteasome-mediated degradation.

ERAD Receptors: The Mannose-Binding Lectins

It has been recently shown by the groups of Markus Aebl and Jon Weissman that substrate de-mannosylation generates a signal for disposal. In fact, de-mannosylated oligosaccharides on ERAD substrates expose terminal α 1,6-bonded mannose residues. These recruit a series of ERAD lectins containing mannose 6-phosphate receptor homology domains (MRH), namely the OS-9 and the XTP3-B splice variants. OS-9 and XTP3-B are stress-inducible genes whose expression is regulated by the ER-stress-activated transcription factor Xbp1. They play a crucial role in retention-based ER quality control and ERAD. Current models propose that OS-9 and XTP3-B, as

well as other ER chaperones, such as BiP, GRP94, and PDI, shuttle misfolded polypeptides from the ER lumen to retrotranslocation (dislocation) complexes in the ER membrane. They deliver the misfolded polypeptides to the transmembrane protein SEL1L, which is associated with HRD1, a membrane-embedded E3 ubiquitin ligase. Dislocons also contain an elusive proteinaceous pore that serves as the conduit for the dislocation of ERAD substrates to the cytoplasm (see Figure 1). The identity of the actual channel remains controversial as a number of ER membrane proteins have been proposed to serve this function including Sec61, Derlins, and E3 ubiquitin ligases themselves.

Further Reading

- Aebi M, Bernasconi R, Clerc S, and Molinari M (2010) N-glycan structures: Recognition and processing in the ER. *Trends in Biochemical Sciences* 35: 74–82.
- Bernasconi R, Galli C, Calanca V, Nakajima T, and Molinari M (2010) Stringent requirement for HRD1, SEL1L, and OS-9/XTP3-B for disposal of ERAD-L_S substrates. *Journal of Cell Biology* 188: 223–235.
- Cabral C, Liu Y, Moremen KW, and Sifers RN (2002) Organizational diversity among distinct glycoprotein ER-associated degradation programs. *Molecular Biology of the Cell* 13: 2651–2663.
- Cabral CM, Liu Y, and Sifers RN (2001) Dissecting glycoprotein quality control in the secretory pathway. *Trends in Biochemical Sciences* 26: 619–624.
- Ellgaard L, Molinari M, and Helenius A (1999) Setting the standards: Quality control in the secretory pathway. *Science* 286: 1882–1888.

- Hampton RY (2002) ER-associated degradation in protein quality control and cellular regulation. *Current Opinion in Cell Biology* 14: 476–482.
- Hosokawa N, Kamiya Y, Kamiya D, Kato K, and Nagata K (2009) Human OS-9, a lectin required for glycoprotein ERAD, recognizes mannose-trimmed N-glycans. *Journal of Biological Chemistry* 284: 17061–17068.
- Molinari M, Calanca V, Galli C, Lucca P, and Paganetti P (2003) Role of EDEM in the release of misfolded glycoproteins from the calnexin cycle. *Science* 299: 1397–1400.
- Molinari M, Galli C, Piccaluga V, Pieren M, and Paganetti P (2002) Sequential assistance of molecular chaperones and transient formation of covalent complexes during protein degradation from the ER. *Journal of Cell Biology* 158: 247–257.
- Oda Y, Hosokawa N, Wada I, and Nagata K (2003) EDEM as an acceptor of terminally misfolded glycoproteins released from calnexin. *Science* 299: 1394–1397.
- Parodi AJ (2000) Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochemical Journal* 348: 1–13.
- Quan EM, Kamiya Y, Kamiya D, et al. (2008) Defining the glycan destruction signal for endoplasmic reticulum-associated degradation. *Molecular Cell* 32: 870–877.

Energy Transduction in Anaerobic Bacteria

G Uden, Johannes-Gutenberg-Universität Mainz, Mainz, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

ATP synthase (F_0F_1 -ATPase) An enzyme in mitochondria and bacteria for the phosphorylation of adenosine diphosphate (ADP), which is driven by the proton potential over the membrane.

Chemiosmotic hypothesis Mechanisms for the coupling of the endergonic adenosine triphosphate (ATP) synthesis by ATP synthase to the exergonic electron transport via a proton potential over the membrane.

Denitrification A form of nitrate respiration that is performed by bacteria in soil and water such as *Paracoccus* and *Pseudomonas* strains, resulting in the formation of N_2 as the product (nitrite, NO, and N_2O as the intermediates).

Nitrate ammonification A form of nitrate respiration producing ammonia as the end product (nitrite as the intermediate); typical for enteric bacteria such as *Escherichia coli*.

Oxygen sensors Facultatively anaerobic bacteria such as *E. coli* or *Pseudomonas* repress the synthesis of pathways for anaerobic respiration or fermentation by transcriptional regulators responding directly (e.g., fumarate nitrate reductase regulator FNR of *E. coli* with an O_2 -sensitive [4Fe4S] cluster) or indirectly to O_2 .

Prokaryotes Microorganisms, mostly unicellular, in which the chromosome is organized in a nucleoid (rather than a nucleus) which is not separated from the cytoplasm by a membrane. Bacteria and Archaea, two of the three major domains of living organisms, are prokaryotes.

Redox half-loop enzymes Enzymes or mechanism for H^+ potential formation that is common in anaerobic respiration (e.g., nitrate reductase, formate dehydrogenase, and hydrogenases). The H^+ potential is formed by the consumption and the release of H^+ on opposite sides of the membrane and e^- transfer (generally $2 H^+ / 2 e^-$) without direct H^+ translocation across the membrane.

Many prokaryotes (bacteria and archaea) are able to grow in the absence of molecular O_2 . Under anaerobic conditions, metabolic energy is gained by the catabolic reactions of fermentation, anaerobic respiration, anoxygenic photosynthesis, decarboxylation of dicarboxylic acids, or electrogenic transport of substrate and product molecules. Many of the reactions are unique for prokaryotes. Anaerobic metabolism is essential for biotopes, which are permanently or occasionally anoxic. The bacteria responsible for anaerobic metabolism and mineralization are strictly anaerobic and are restricted to anaerobic biotopes, or are facultatively anaerobic, using either aerobic or anaerobic metabolism, depending on the O_2 supply. The facultative bacteria switch between aerobic and anaerobic metabolism by using molecular O_2 sensors, which control the expression of aerobic or anaerobic metabolic systems mostly at the transcriptional level.

Adenosine Diphosphate Phosphorylation Driven by Chemical Coupling or by the Membrane Potential

Synthesis of adenosine triphosphate (ATP) by fermentation relies on substrate-level phosphorylation, whereas the other reactions for anaerobic energy conservation use a membrane potential as the coupling device between energy-supplying and energy-consuming reactions, in particular adenosine diphosphate (ADP) phosphorylation (Figure 1). Phosphorylation of ADP for generation of ATP requires a free energy ($\Delta G_0'$) of approximately $+50 \text{ kJ mol}^{-1}$, or higher, under cellular conditions. This amount of free energy has to be supplied in fermentation by direct coupling in one chemical reaction. In the metabolic systems using the membrane or proton potential

(Δp) as the coupling device (Figure 1), the energy-supplying reaction, on the other hand, has to supply only $12\text{--}17 \text{ kJ mol}^{-1}$ per reaction, which is the amount required for translocating one H^+ (or Na^+) ion across the membrane. Three or four H^+ ions are used then to drive ADP phosphorylation. Therefore, storing the free energy in a membrane potential is favorable for anaerobic catabolic processes, which often provide only small amounts of free energy.

The overview will demonstrate, by examples, the functioning of anaerobic metabolism of bacteria, in particular, fermentation, and reactions using a membrane potential for coupling between energy-supplying and energy-consuming reactions.

Fermentation

Fermentation plays a major role in the carbon cycle of anaerobic sediments in freshwater, sewage fermenters, or the digestive apparatus of vertebrates and insects. Fermentation means growth in the absence of external electron acceptors or light. ADP is phosphorylated under these conditions by substrate-level phosphorylation, or direct chemical coupling, without the involvement of membranes or membrane potential. The most common substrates for fermentative metabolism are sugars and amino acids. The substrates are first degraded in an oxidative part of metabolism yielding pyruvate or related compounds and reduced nicotinamide adenine dinucleotide (phosphate) (NAD(P)H). For oxidation of the sugars, most bacteria use the glycolytic (Embden–Meyerhof–Parnas, EMP) pathway to pyruvate ($\text{glucose} + 2\text{NAD}^+ + 2\text{ADP} + 2\text{P}_i \rightarrow 2\text{pyruvate} + 2\text{NADH} + 2\text{H}^+ + 2\text{ATP}$). Some bacteria use the phosphoketolase (modified pentosephosphate) pathway

(glucose + 3NAD(P)⁺ + ADP + P_i + HSCoA → pyruvate + acetyl-CoA + 3NADH + 3H⁺ + 1ATP) for pyruvate formation, whereas the Entner–Doudoroff (or 2-keto-3-deoxy-6-P-gluconate, KDPG) pathway (glucose + 2NAD(P)⁺ + ADP + P_i → 2pyruvate + 2NAD(P)H + 2H⁺ + 1ATP) is found only rarely in anaerobic bacteria.

In simple fermentations, pyruvate is used as the electron acceptor for reoxidation of the NAD(P)H formed during sugar oxidation, yielding lactate or ethanol as the fermentation products. In more complicated fermentation reactions (mixed acid fermentation, butyrate fermentation, and others), pyruvate is further oxidized yielding acetate and additional ATP. In these pathways, butyrate or ethanol (plus acetate) can be obtained as fermentation products, among others.

Typical fermentation products derived from sugars and amino acids are carboxylic acids (formate, acetate, propionate, butyrate, and lactate), alcohols (ethanol, 2-propanol, *n*-butanol, and 2,3-butanediol), but also polyols (glycerol, erythritol, and mannitol), CO₂, and H₂ (Table 1). The well-known fermentation reactions of this type are homo- and heterofermentative lactic acid, propionic acid, butyric acid, mixed acid, or ethanolic fermentations, which are named according to the major or characteristic products (Table 1). The ATP yields of the fermentation reactions are generally in the range of 1–3

ATP mol⁻¹ of hexose (compared to up to 38 ATP mol⁻¹ glucose in aerobic metabolism).

The most important energy-conserving reactions in fermentation are the oxidation of aldehyde to carbonates (thioester formation by glyceraldehyde-3-phosphate dehydrogenase), the oxidative decarboxylation of α-ketoacids (thioester formation by pyruvate-ferredoxin oxidoreductase), and the phosphoroklastic cleavage of ketuloses (acetyl-phosphate formation by phosphoketolase), which are all located in the oxidative part of fermentative metabolism. Reductive deamination of amino acids (acetyl-phosphate formation by glycine reductase) is an example of an energy-conserving reaction in the reductive branch of Stickland fermentation. The energy-rich products obtained from these reactions are then used finally for the phosphorylation of ADP, yielding ATP.

Organic matter (polysaccharides, proteins, fats, and monomers derived thereof) can be mineralized in the absence of external electron acceptors to methane and CO₂ by the interaction of fermenting and methanogenic prokaryotes in an anaerobic food chain (Figure 2). Primary fermenters degrade sugars and amino acids by typical fermentation reactions to propionate, butyrate, succinate, lactate, acetate, ethanol, CO₂, H₂, and NH₃. Ethanol and fatty acids (short-chain fatty acids, such as propionate, up to C₁₈) are then degraded by secondary fermenters to acetate, H₂, and CO₂. H₂, CO₂, and acetate from the former fermentation reactions are converted by homoacetogenic and methanogenic prokaryotes to methane and CO₂. Secondary fermentation is mostly an endergonic process under standard conditions (e.g., butyrate + 2H₂O → 2acetate + H⁺ + 2H₂; ΔG₀' = +48.2 kJ mol⁻¹ butyrate) and becomes exergonic only in the presence of syntrophic prokaryotes, which consume H₂ and other products to keep them at very low levels. By this, the reaction becomes exergonic under environmental conditions (ΔG_{env} = -18 kJ mol⁻¹ butyrate), enabling growth of the bacteria. In this way, fatty acids, alcohols, and aromatic compounds are mineralized yielding CO₂ and methane as the final products.

Fermentation reactions are used on a large scale for the preparation of food and beverages. The fermentation reactions and products improve food (nutritional) quality, taste, and durability. For this purpose, mainly primary fermentation reactions such as ethanol (beverages), lactic acid (dairy products, vegetables, and sour dough), propionic acid fermentations (cheese), and others are used. Solvent production (ethanol, propanol, butanol, and acetone) by Clostridia also depend on type I fermentations. Biotechnological processes for

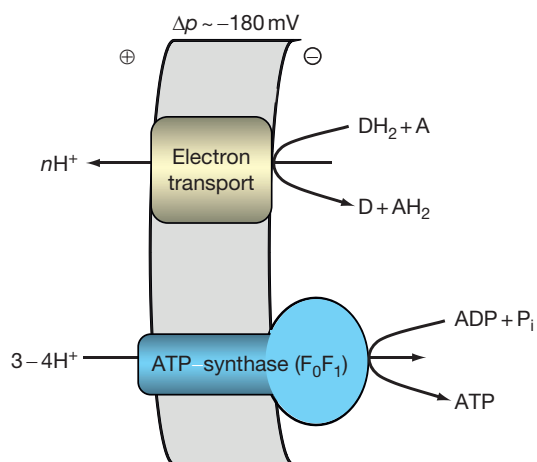


Figure 1 Generation of a proton potential (Δp) by translocation of protons across a cytoplasmic membrane by anaerobic respiration and by coupling to ADP phosphorylation by ATP synthase. DH₂, electron donor; A, electron acceptor.

Table 1 Fermentation products of sugars (primary fermentations)

Fermentation (pathway for sugar degradation)	Products per mol glucose (ATP yield per glucose)	Bacterium or microorganism
Lactic acid, homofermentative (EMP)	2 Lactate (+2ATP)	<i>Lactobacillus lactis</i>
Lactic acid, heterofermentative (PK)	1 Lactate + 1 ethanol + 1CO ₂ (+1ATP)	<i>Leuconostoc</i>
Propionic acid (EMP)	2 Propionate (+2–3ATP)	<i>Propionibacterium</i>
Butyric acid (EMP)	1 Butyrate + 2CO ₂ + 2H ₂ (+3ATP)	<i>Clostridium butyricum</i>
Mixed acid (EMP)	1 Acetate + 1 ethanol + 2 formate (+3ATP)	<i>Escherichia coli</i>
Ethanolic (ED)	2 Ethanol + 2CO ₂ (+1ATP)	<i>Zymomonas mobilis</i>
Ethanolic (EMP)	2 Ethanol + 2CO ₂ (+2ATP)	<i>Saccharomyces cerevisiae</i> (yeast)

EMP, Embden–Meyerhof–Parnas pathway (glycolysis); PK, phosphoketolase pathway; ED, Entner–Doudoroff pathway.

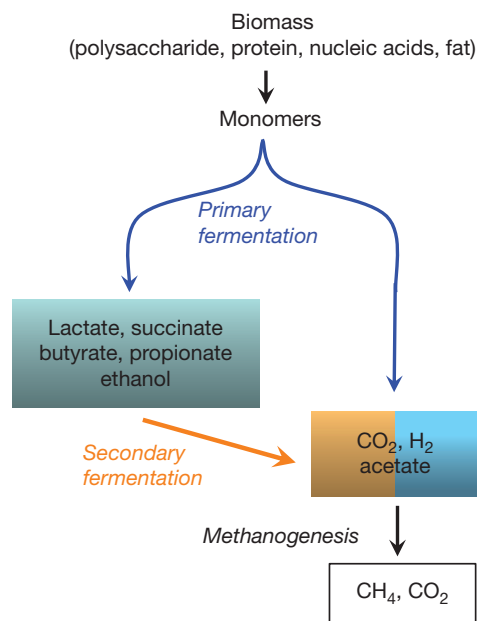


Figure 2 Role of primary and secondary fermentation, and of methanogenesis in the anaerobic degradation of organic matter in the absence of electron acceptors. The food chain shown here is found in anaerobic sediments, sewage sludge, and the digestive apparatus of vertebrates and insects. Most reactions of secondary fermentation are endergonic under standard conditions, but become exergonic under environmental conditions ($\Delta G'_{\text{env}}$) when the levels of H_2 , CO_2 , and acetate are kept low by methanogenic archaea or by bacteria.

microbiological production of amino acids, citrate, and others, on the other hand, rely on an overflow metabolism of central pathways and anabolic reactions.

Electron Transport Phosphorylation in Anaerobic Respiration

Anaerobic respiration is able to generate an electrochemical proton potential ($\Delta p \sim -0.18 \text{ V}$) over the membrane, which is driven by the oxidation of an electron donor by an (mostly externally added) electron acceptor. The free energy of the redox reaction is conserved in a proton potential (Figure 1). The proton potential then drives phosphorylation of ADP and formation of ATP by ATP synthase. The Δp -generating respiratory chain and the Δp -consuming ATP synthase are thus coupled only by the Δp . The free energy of the redox reaction of the respiratory chain under standard conditions and pH 7 is directly related to the difference in the redox potentials of the electron donors and acceptors ($\Delta E'_0$), the Faraday constant ($F = 96.5 \times 10^3 \text{ J mol}^{-1} \text{ V}^{-1}$), and the number of participating electrons (n):

$$\Delta G'_0 = -nF\Delta E'_0$$

Figure 3 gives a selection of important electron donors and acceptors used by prokaryotes in anaerobic respiration and the corresponding midpoint potentials. The $\Delta E'_0$ value sets the maximal (theoretical) amount of electrons (H^+/e^- ratio),

which can be translocated across the membrane for the given Δp value (-0.18 V):

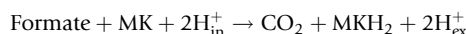
$$(\text{H}^+/\text{e}^-)_{\text{max}} = \Delta E'_0 / \Delta p$$

The actual H^+/e^- ratio generally is lower than the calculated value due to lower coupling efficiency, and since some electron transport enzymes lack coupling mechanisms for H^+ translocation. From the above equations, it follows that the minimal amount of free energy required for the translocation of one H^+ ion across membranes ($\Delta p = -180 \text{ mV}$) is $-17.4 \text{ kJ mol}^{-1}$. Thus, free energy as low as $\Delta G'_0 = -17 \text{ kJ mol}^{-1}$ can be conserved in a membrane potential by anaerobic respiration (or other Δp -generating reactions) compared to the $\Delta G'_0$ value of at least -50 kJ mol^{-1} required in substrate-level phosphorylation. Electron transport therefore allows use of energy-supplying reactions with low $\Delta G'_0$ values, since three to four translocated H^+ or Na^+ ions (derived from up to four energy-supplying reactions) are used for phosphorylation of one ADP molecule. This coupling mechanism is of particular significance for reactions of anaerobic metabolism, which are only weakly exergonic.

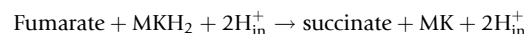
In anaerobic respiration, a large number of different electron acceptors can be used, including S^0 , CO_2 , sulfate, fumarate, nitrate, tetrachlorethen, and Fe^{3+} (Figure 3). The free energy, and thus the ATP yields, are distinctly lower than in aerobic respiration due to the more negative redox potentials (and thus $\Delta E'_0$ values) of the electron acceptors. Most of the anaerobic respiratory systems are composed of two enzymes, a dehydrogenase and a terminal reductase. The enzymes are linked by diffusible low-molecular-weight redox mediators (quinones). In anaerobic respiration, menaquinone (MK, a lipophilic naphthoquinone derivative with $E'_0 = -80 \text{ mV}$) replaces ubiquinone (Q, a lipophilic benzoquinone derivative with $E'_0 = +110 \text{ mV}$) from aerobic respiratory chains. Figure 4 gives a schematic presentation of the topology and arrangement of enzymes in fumarate and nitrate respiration, with formate as the electron donor. The biochemistry and energetics of the enzymes are known in detail, and the structures of the enzymes have been solved by X-ray crystallography. Each of the enzymes is anchored in the membrane by a hydrophobic anchor protein, which contains the active site for the lipophilic quinone. The second active site for the hydrophilic substrates, formate, fumarate, or nitrate, has either cytoplasmic or extracellular topology.

In formate–fumarate respiration, the enzymes catalyze the following reactions:

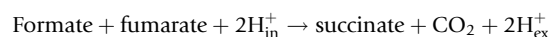
Formate dehydrogenase (or formate:quinone oxidoreductase):



Fumarate reductase (or menaquinol:fumarate reductase):



Formate–fumarate reduction:



The overall reaction of formate–fumarate respiration results in the release of two H^+ ions at the periplasmic, and consumption of two H^+ ions at the cytoplasmic side of the membrane

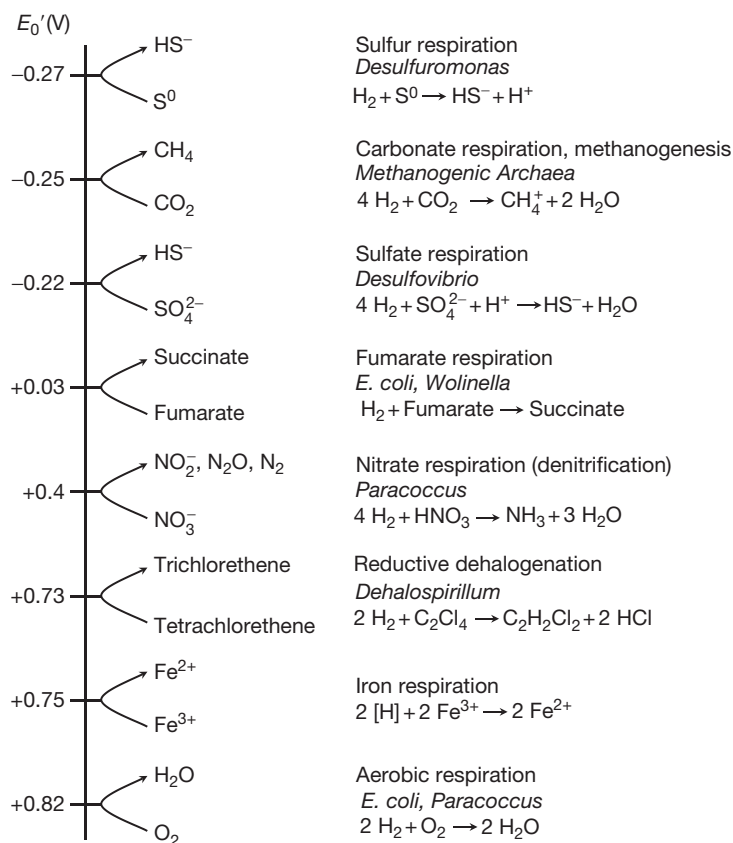


Figure 3 Important electron acceptors of anaerobic respiration by prokaryotes. The scheme gives the redox potentials of the redox pairs, the type of respiration, the metabolic balance using H_2 as the electron donor, and examples for prokaryotes catalyzing the reaction. The free energy of the reaction can be calculated from $\Delta G'_0 = -nF\Delta E'_0$, and is, for example, for fumarate respiration ($H_2 + \text{fumarate} \rightarrow \text{succinate}$) $\Delta G'_0 = -87 \text{ kJ mol}^{-1}$.

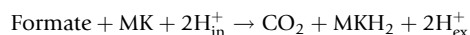
resulting in a $H^+/2e^-$ ratio of 2, which is caused by the action of formate dehydrogenase. Formate dehydrogenase has active sites for formate oxidation and menaquinone reduction on opposite sides of the membrane, and translocates two electrons from the formate to the quinone site, resulting in an overall translocation of two charges across the membrane. The enzyme releases and consumes H^+ ions on opposite sides of the membrane (without actual translocation of H^+), resulting in Δp production. Many anaerobic respiratory enzymes generate Δp in the same way. This mechanism of Δp generation by redox half-loop enzymes has to be distinguished from proton-pumping redox enzymes, such as cytochrome *c* oxidase, which transfer protons across the membrane in a reaction, which is driven by a redox reaction (proton pump).

In fumarate reductase, the protons are released (menaquinol site) and consumed (fumarate site) on the same (cytoplasmic) side of the membrane. The reaction, therefore, is not electrogenic and does not contribute to the formation of Δp , and fumarate reductase functions only as an electron sink. The overall $H^+/2e^-$ ratio of formate–fumarate reduction is 2. The big difference in the redox midpoint potentials of formate and menaquinone in formate:menaquinone reduction ($\Delta E'_0 = -340 \text{ mV}$) allows the generation of a proton potential ($H^+/2e^-$ ratio ≤ 3.8). For fumarate reductase, the $\Delta E'_0$ ($\Delta E'_0 = -110 \text{ mV}$) is not sufficient for generation of a proton

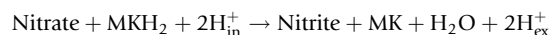
potential. Similarly, other terminal reductases of bacterial respiratory chains function without direct contribution to Δp production.

Formate–nitrate respiration is catalyzed by an electron transport chain consisting of formate dehydrogenase and nitrate reductase:

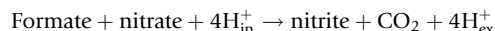
Formate dehydrogenase (or formate:quinone oxidoreductase):



Nitrate reductase (or quinol:nitrate reductase):



Formate–nitrate reduction:



In quinol–nitrate reduction, $\Delta E'_0$ is -510 mV with menaquinol, or -350 mV with ubiquinol as the electron donor, which allows conservation of the redox energy in a proton potential (Figure 4). Accordingly, nitrate reductase functions as a redox half-loop enzyme very similar to (though mirror-inverted) formate dehydrogenase. Overall, formate–nitrate respiration has a $H^+/2e^-$ ratio of 4.

A similar organization is found for many enzymes of anaerobic respiratory chains, which are composed of one respiratory

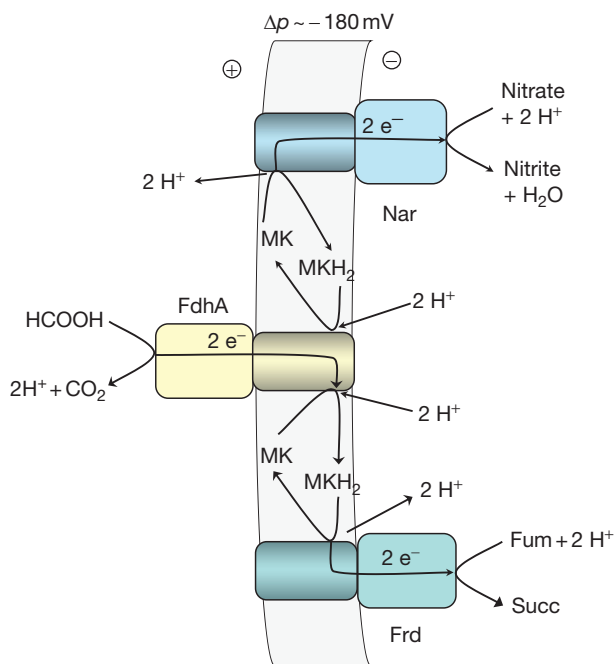


Figure 4 Formate–fumarate and formate–nitrate respiration. The figure shows the topology and reactions of formate dehydrogenase (formate: quinol reductase, Fdh), fumarate reductase (quinol:fumarate reductase, Frd), and nitrate reductase (quinol:nitrate reductase, Nar). Note the topology of the active sites for formate, fumarate, nitrate, and the quinones, the release and consumption of H^+ at the active sites, and transmembrane transfer of electrons in formate dehydrogenase and nitrate reductase. Frd, fumarate reductase; Nar, Nitrate reductase; Fdh, formate dehydrogenase. Examples of this type of fumarate respiration are found in *E. coli* and *Wolinella*, and of the nitrate respiration in *E. coli*. Formate dehydrogenase and nitrate reductase represent redox half-loop enzymes. The latter enzymes generate a proton potential by transfer of electrons across the membrane and by consuming and releasing protons on opposite sides of the membrane. Due to the topology of the substrate and quinone sites, a H^+ potential is produced without H^+ translocation.

dehydrogenase and terminal reductase each, and which are linked by the quinones. In this way, a large number of different respiratory dehydrogenases and terminal reductases can be connected in branched respiratory chains.

Decarboxylation of Dicarboxylic Acids

Some anaerobic bacteria grow at the expense of decarboxylation of dicarboxylic acids, such as oxaloacetate or succinate. The free energy of this reaction is $\sim -20 \text{ kJ mol}^{-1}$ and is used for translocation of one Na^+ ion across the membrane by a membranous decarboxylase as the energy-conserving enzyme. The sodium motive force generated in this way is used to drive secondary substrate transport, flagella, motility, or phosphorylation of ADP, if the bacteria contain an Na^+ potential-driven ATP synthase. The Na^+ -dependent ATP synthase is similar in function and composition to the H^+ -translocating F_1F_0 -ATP synthase. In this way, three to four decarboxylation reactions are required for phosphorylation of one ADP molecule. *Propionigenium modestum* grows by conversion of succinate to

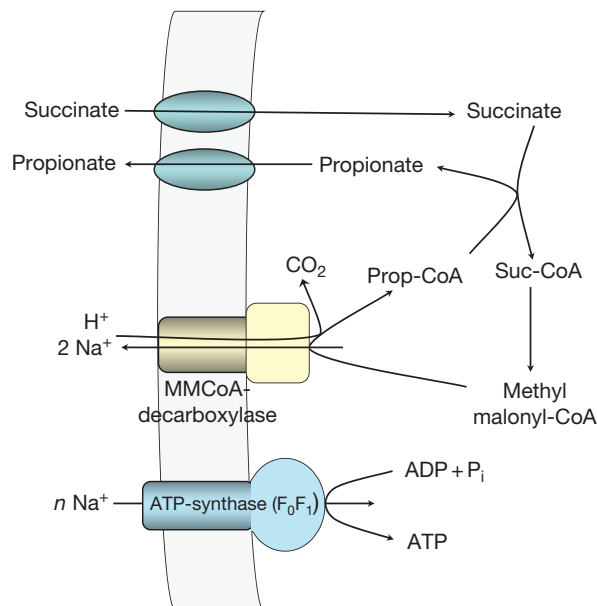


Figure 5 Anaerobic growth on succinate by *P. modestum* (succinate + $H_2O \rightarrow$ propionate + CO_2 ; $\Delta G_0' = -21 \text{ kJ mol}^{-1}$) and generation of an Na^+ potential by decarboxylation. The membranous methylmalonyl-CoA decarboxylase couples decarboxylation to translocation of two Na^+ ions over the membrane. H^+ ions for the decarboxylation reaction are derived from the outside.

propionate, including decarboxylation of methylmalonyl-CoA by a membranous decarboxylase (Figure 5). Succinate is converted to methylmalonyl-CoA, which is then decarboxylated in a reaction coupled to Na^+ ion translocation. The enzyme, in addition, consumes one H^+ ion from the periplasmic side of the membrane for the formation of propionyl-CoA and CO_2 . Thus, the free energy of decarboxylation is used for generating an electrochemical Na^+ potential. The decarboxylation represents an intramolecular redox reaction producing an oxidized (CO_2) and a reduced (propionate) product. The reaction depends only on the membranous decarboxylase for generation of the electrochemical potential. In *Klebsiella*, an enterobacterium closely related to *Escherichia coli*, a membranous Na^+ translocating decarboxylase (oxaloacetate decarboxylase), is part of citrate fermentation. Oxaloacetate decarboxylation yields pyruvate, which can be further converted to acetate and formate. In citrate fermentation, ATP is formed in addition by substrate-level phosphorylation, or a H^+ potential is generated in respiratory pathways. Therefore, the ATP synthase of *Klebsiella* is H^+ -ion dependent, and the Na^+ potential is used mainly for secondary substrate transport.

Generation of an H^+ Potential by Transport of Charged Substrate and Product Molecules

Some bacteria use the (electrogenic) transport of substrate and product molecules for the generation of a proton potential. *Oxalobacter formigenes* and lactic acid bacteria such as *Lactococcus lactis* or *Oenococcus oeni* are able to generate a proton potential by decarboxylation of the dicarboxylic acids oxalate or malate. The decarboxylation is catalyzed by cytoplasmic

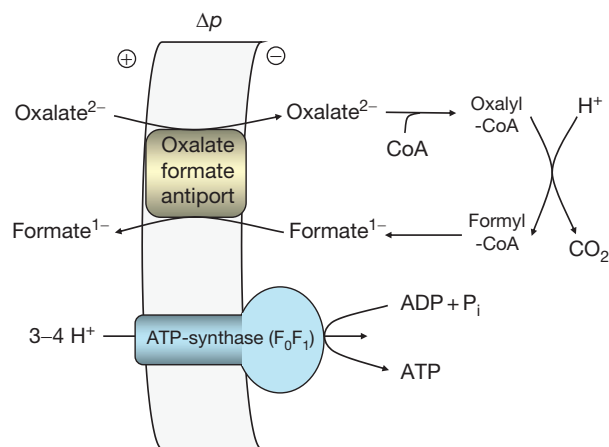


Figure 6 Anaerobic growth on oxalate by *Oxalobacter formigenes* (oxalate $\rightarrow$ formate + CO_2 ; $\Delta G_0' = -27 \text{ kJ mol}^{-1}$) and generation of a H^+ potential by electrogenic oxalate $^{2-}$ /formate $^{1-}$ antiport and decarboxylation of oxalate by a cytoplasmic (H^+ -consuming) decarboxylase.

enzymes, and the free energy of the reaction is not directly stored in a membrane potential as in the membranous decarboxylases. In *Oxalobacter*, oxalate is taken up by an electrogenic antiporter with formate, the end product of the pathway (Figure 6). Decarboxylation of activated oxalate (oxalyl-CoA) produces CO_2 and consumes one H^+ . The antiport of the dianionic oxalate against the monoanionic formate generates an electrical gradient (positive on the outside), and consumption of one H^+ ion by the decarboxylation generates a ΔpH (alkaline on the inside). The resulting proton potential is used to drive ADP phosphorylation by an H^+ -ATP synthase.

Decarboxylation of malate by lactic acid bacteria (malolactic fermentation: malate + $\text{H}^+ \rightarrow$ lactate + CO_2) uses a similar mechanism of energy conservation. In *Lactococcus lactis*, malate is taken up and converted by a cytoplasmic decarboxylase to lactate + CO_2 , consuming H^+ as a cosubstrate. Malate and lactate are transported by an electrogenic antiporter, generating an electrical potential (positive on the outside), and H^+ ion consumption in the cytoplasm during decarboxylation generates a ΔpH value (alkaline on the inside). The proton potential can be used for driving ADP synthesis by a membranous ATP synthase.

Anoxygenic Photosynthetic Prokaryotes

Anoxygenic photosynthetic bacteria use only one photosystem for converting light energy into an electrochemical proton

potential, which is then used for driving ADP phosphorylation. Anoxygenic photosynthetic bacteria are found in three different phylogenetic groups, which contain different photosynthetic systems: the purple bacteria, the green phototrophic bacteria with the subgroups of green sulfur bacteria (*Chlorobiaceae*) and *Chloroflexus*, and the Gram-positive *Heliobacteria*. The three groups show differences in the type of photosynthetic reaction center, photosynthetic electron transfer and the electron donors, pigments, and the pathway for CO_2 fixation. In anoxygenic photosynthesis, light is used to raise electrons to a more electronegative redox potential and to feed the electrons into a cyclic electron transport to generate a proton potential. The bacteria contain only one reaction center in contrast to oxygenic photosynthesis in cyanobacteria, green algae, or plants. The photosystems of anoxic photosynthesis transfer the electrons either to quinones or to ferredoxin (quinone-type- or FeS-type reaction centers). The archaeon *Halobacterium salinarum*, on the other hand, contains a proton pump that is directly driven by light to translocate protons over the membrane without involvement of photosynthetic electron transport. In contrast to the photosynthetic bacteria that use chlorophyll containing proteins for photosynthesis and energy conversion, *Halobacterium* contains bacteriorhodopsin with retinal as a chromophore for light absorption and conversion to an electrochemical proton potential.

See also: **Bioenergetics:** Chemiosmotic Theory; Membrane-Associated Energy Transduction in Bacteria and Archaea; Respiratory Chain and Adenosine Triphosphate Synthase; **Metabolism Vitamins and Hormones:** Pentose Phosphate (Hexose MonoPhosphate) Pathway; Photosynthesis.

Further Reading

- Dimroth P, Jockel P, and Schmid M (2001) Coupling mechanism of the oxaloacetate decarboxylase Na^+ pump. *Biochimica et Biophysica Acta* 1505: 1–14.
- Jormakka M, Byrne B, and Iwata S (2003) Proton motive force generation by a redox loop mechanism. *FEBS Letters* 545: 25–30.
- Madigan MT, Martinko JM, Dunlap PV, and Clark DP (2009) *Brock Biology of Microorganisms*, 12th edn. San Francisco, CA: Pearson Education.
- Mitchell P (1961) Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Naturwissenschaften* 191: 144–148.
- Udden G and Bongaerts J (1997) Alternative respiratory pathways of *Escherichia coli*: Energetics and transcriptional regulation in response to electron acceptors. *Biochimica et Biophysica Acta* 1320: 217–234.
- White D (2000) *The Physiology and Biochemistry of Prokaryotes*, 2nd edn. New York, NY: Oxford University Press.

Enzyme Inhibitors

V L Schramm, Yeshiva University, Bronx, NY, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 31–37, © 2004, Elsevier Inc.

Glossary

Biological efficiency The ability of an enzyme inhibitor to achieve significant inhibition at the physiological conditions of substrates and cofactors.

Catalytic site A cavity on the enzymatic surface that binds substrates and causes them to be converted to product by increasing the probability that they will achieve the transition state.

Inhibitor design Selection or synthesis of stable molecules intended to disrupt catalytic function of a specific enzyme.

Mechanism-based inhibitor A molecule resembling a substrate that is chemically activated by the action of

the enzyme to form a stable covalent link with the enzyme – also known as suicide substrates or suicide inhibitors.

Substrate analog A molecule with shape and charge features similar to a substrate of the enzyme that binds at the catalytic site and prevents substrate binding.

Transition-state analog A chemically stable molecule with shape and electrostatic features that mimic those of the transition state and that inhibit at the catalytic site by converting energy of catalysis to energy of binding.

Reversible Catalytic-Site Inhibitors

Reactant and Product Analogs

Molecules with shape and charge similar to the substrate or products may bind to the catalytic site similar to the normal reactants. When they are bound, the substrate cannot bind and the enzyme is prevented from catalyzing the reaction. The relative affinity of these competing molecules for the catalytic site and their abundance relative to the substrate will determine if they prevent substrate binding (Figures 1 and 2). Thus, if reactants and inhibitors bind with equal affinity, at equal concentrations, the enzyme catalytic site will be equally occupied with reactants and inhibitor resulting in 50% inhibition. This relationship is described in quantitative terms by the competition between an inhibitor (I) and a reactant (A) for a simple enzyme with one reactant by the equation $v = (k_{\text{cat}})(A)/(K_m + A(1 + I/K_i))$ where v is the enzymatic reaction rate, k_{cat} is the maximum catalytic rate of the enzyme, K_m is related to the binding affinity between enzyme and substrate, defined as the concentration of substrate to achieve 50% of k_{cat} in the absence of inhibitor, and K_i is the dissociation constant for the enzyme–inhibitor complex, defined by the equation $K_i = (E)(I)/(EI)$, and A and I are reactant and inhibitor concentrations, respectively. Diagrams of substrate and product inhibition illustrate the mutually exclusive or competitive inhibition relative to the catalytic site (Figure 2). Complete graphical and mathematical analyses of enzyme inhibition cases are available.

Bisubstrate Mimics

Many enzymatic reactions combine two different molecules at the catalytic site to transfer a chemical group between them. An example is the enzyme adenylate kinase that transfers the terminal phosphoryl group of adenosine triphosphate (ATP) to adenosine monophosphate (AMP) to generate two molecules of adenosine diphosphate (ADP; Figure 3). The catalytic

site must accommodate ATP and AMP at the same time in preparation for group transfer. The bisubstrate mimic for the adenylate kinase reaction is chemically similar to ATP combined with AMP to make a bisubstrate molecule called AP₅A (Figure 3). Such molecules compete for both substrate sites, since they prevent either substrate from binding by competition for the same sites. Bisubstrate mimics can bind tightly to the enzyme since connected molecules bind with affinity equal to the product of the affinity of the individual molecules. For example, if individual dissociation constants (K_i values) for ATP and AMP are both 10^{-4} M, the dissociation constant for the bisubstrate mimic would be expected to be related to the product of these constants or $10^{-4} \times 10^{-4}$ M = 10^{-8} M.

Transition-State Analogs

A special class of competitive inhibitors resembles the reactant molecule at the very point where it crosses the chemical energy equivalence point of bond breaking/bond making called the transition state. These inhibitor molecules must be chemically stable, but have close similarity to the unstable transition state. Based on the theory of transition-state stabilization for enzyme-catalyzed reactions, the perfectly designed transition-state analog is capable of binding to the enzyme by the factor: K_i for transition-state inhibitor = K_m /catalytic rate enhancement. Since enzymatic catalytic rate enhancements are typically 10^{10} – 10^{15} , perfectly designed transition-state inhibitors can bind 10^{10} – 10^{15} times better than substrate. These incredibly large affinities have been rarely achieved. It is impossible to match the nonequilibrium bond lengths and partial charges present at the precise moment of the transition state with chemically stable molecules. However, incorporation of even modest similarities between the actual transition state and transition-state analogs can achieve binding affinities 10^5 – 10^7 times that of reactants, and this improvement is adequate to achieve inhibition of the targeted enzymes in biological systems.

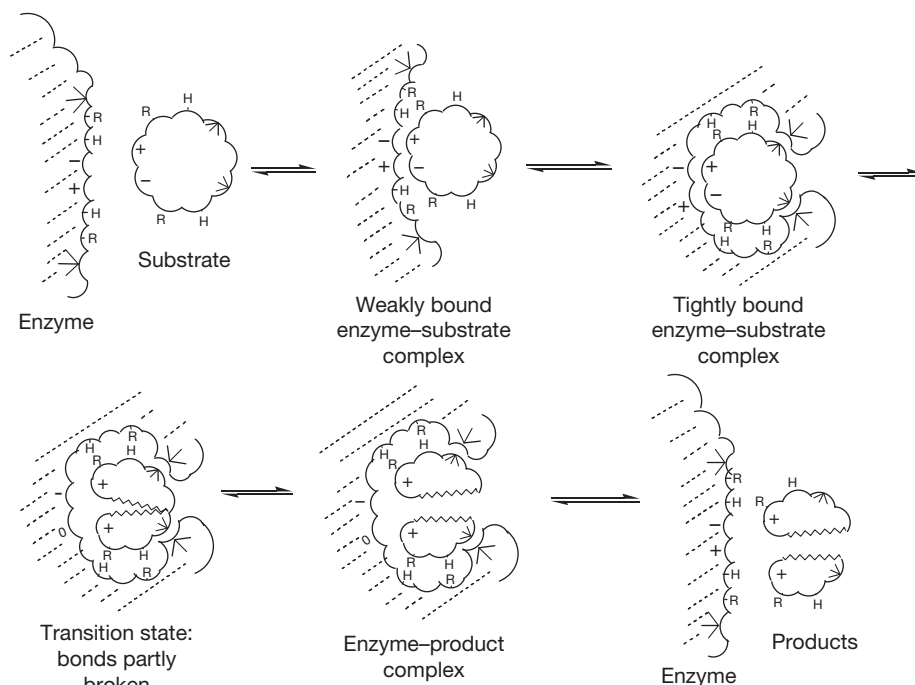


Figure 1 A schematic for the steps in enzyme-catalyzed reactions. In the first step, diffusion causes collision between the substrate and an open catalytic site. The weakly bound enzyme–substrate complex causes the protein to close to form the tightly bound enzyme–substrate complex. Hydrogen bonds (H–R), ionic bonds (+–), and hydrophobic interactions (><) in this closed complex at the transition state are shown. Following bond changes, the enzyme–product complex relaxes to release product and regenerate the original form of the enzyme to repeat the cycle.

Enzymatic transition states

This point of the chemical reaction is often the most energetic point in the reaction cycle, and therefore the rarest and the most difficult to detect. The exact shape and charge of the substrate molecule at the moment of the transition state is of great interest, since chemically stable molecules that resemble the substrate molecule at the transition state bind extremely tightly to the enzyme, as discussed above. This concept was proposed early in the history of biochemistry by Linus Pauling, and was formalized several decades later by Richard Wolfenden. An example of the substrate, enzyme-stabilized transition state and a transition analog for purine nucleosidase phosphorylase is shown (Figure 4). These inhibitors commonly bind to the catalytic sites of enzymes millions of times more tightly than the substrates. Many of the antibiotics isolated in nature have been found to be transition-state analogs. Recently, molecules designed to be transition-state analogs are being tested in medical applications such as anticancer agents. Transition-state inhibitors compete at the catalytic site and cause the enzyme to fold into its tightly binding mode and keep it closed for long times, preventing catalysis (see Figure 1).

Design of transition-state analogs

Both the shape and charge of reactant molecules change as they reach the transition state (see Figure 1). The structure of the transition state can be experimentally measured by using kinetic isotope effects, combined with computational chemistry to locate the enzymatic transition state. The electron distribution at the van der Waals surface of a molecule is called its

molecular electrostatic potential surface. This can be described for substrate and transition states for enzyme reactions provided the transition-state structure is known, and can also be established for proposed inhibitor molecules. Electrostatic matches to the transition state make powerful inhibitors by capturing some of the transition-state energy usually applied to catalysis (Figure 4).

Natural product transition-state analogs

Many antibiotics isolated from natural sources are transition-state inhibitors. One example is an antibiotic formed in the culture medium of *Streptomyces antibioticus* that is apparently made as an agent of microbial warfare between species of soil organisms. R-coformycin binds to the catalytic sites of adenine nucleoside and nucleotide deaminases millions of times tighter than the reactant molecules. The transition-state structure for deaminase enzymes has been solved and the similarity of transition-state analog and transition state is greater than that of either to the substrate (Figure 5).

Noncatalytic Site Inhibitors

Noncatalytic site inhibitors include all inhibitors binding at sites other than the catalytic site. These inhibitor molecules can cause several different types of enzymatic inhibition, both competitive and noncompetitive. This diversity arises as a result of the inhibitor binding distant to the catalytic site, which changes the enzyme conformation that prevents the enzyme from functioning properly, either in binding its substrate or in forming products.

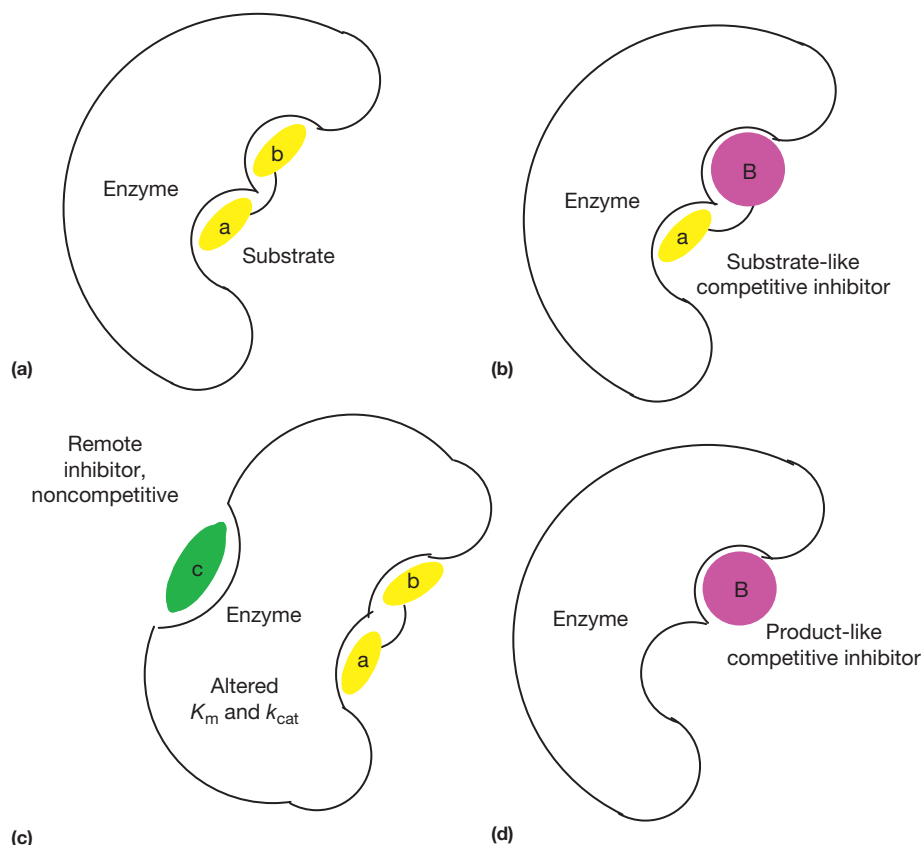


Figure 2 Enzyme A is a normal Michaelis complex with substrate a–b at the catalytic site (yellow). The bond to be broken by the enzyme connects a and b. The substrate-like competitive inhibitor (pink, enzyme B) has a difference in the B part of the bound molecule and is not a substrate but competes directly for a–b at the catalytic site. Since this is a competition, large amounts of a–b will displace a–B. A noncompetitive inhibitor binds to influence catalysis but does not directly compete for the catalytic site. There are many variants of noncompetitive inhibition and binding at a remote site to change features of the catalytic site is one example (green in enzyme C). Competitive inhibitors can also be product-like as shown in enzyme D (pink).

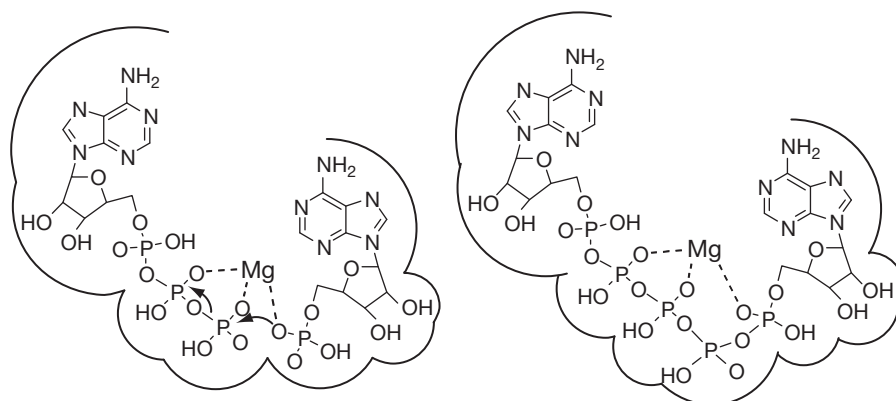


Figure 3 The Michaelis complex for adenylate kinase ($\text{ATP} + \text{AMP} \rightarrow 2\text{ADP}$) is shown on the left with the arrows indicating electron flow for the phosphoryl group transfer. The inhibited complex with adenosyl-pentaphosphate-adenosine (A-P5-A) is on the right. The bisubstrate inhibitor competes for both ATP- and AMP-binding sites. Ionic charges are omitted for clarity and the ribosyl groups are all β -D-ribose in stereochemistry.

Prevent Catalytic Site Binding

An inhibitor may bind remote from the catalytic site to change the protein conformation and cause mutual exclusion between

reactant binding at the catalytic site and inhibitor binding at the remote site. For this interaction, kinetic analysis of the type shown above will be kinetically indistinguishable from competitive inhibition. The surfaces of proteins contain highly irregular

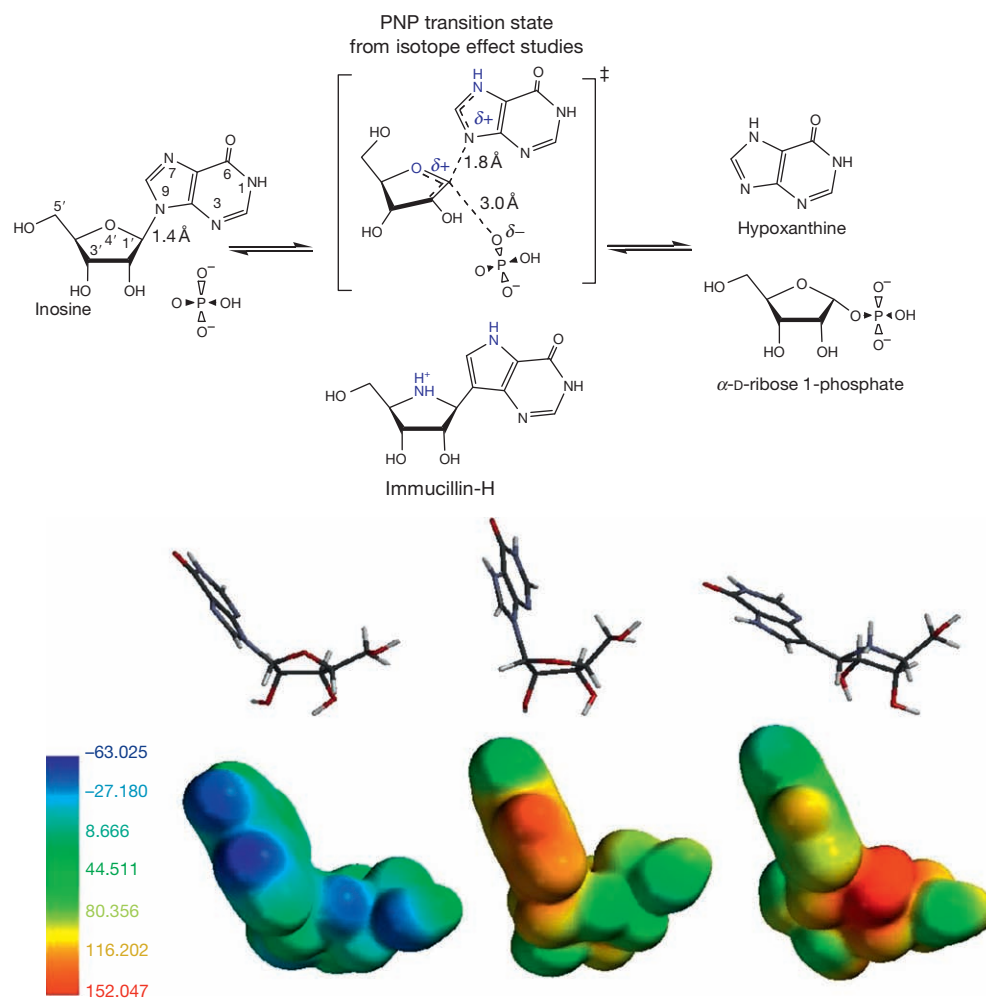


Figure 4 The reaction catalyzed by bovine purine nucleoside phosphorylase (PNP), transition-state structure, and a transition-state analog inhibitor (immucillin-H). Novel features of the transition state are protonation at N7 and the partial positive charge on the ribosyl group. Both are represented in the inhibitor, shown in blue. Immucillin-H binds approximately 1 million times more tightly to PNP than does the substrate inosine. Shown below are the electronic van der Waals surfaces of inosine (left), transition state (middle), and immucillin-H (right). Note the electrostatic similarity between transition state and Immucillin-H. Reproduced from Schramm VL (2002) Development of transition state analogues of purine nucleoside phosphorylase as anti-T-cell agents. *Biochimica et Biophysica Acta* 1587: 107–117.

and charged surfaces that interact with specific molecules. Current practice of testing libraries of diverse chemical compounds against a specific enzyme to find any type of interaction that can cause inhibition often discovers this class of inhibitor.

Prevent Subunit Formation

Most enzymes have multiple protein subunits that must interact to provide a functional catalytic unit. An example is the ribonucleotide diphosphate reductase required to form deoxynucleotides for the biological production of DNA. During cellular protein synthesis of these enzymes, the protein subunits are formed separately and must then combine to make the functional catalyst. In the cell, these subunits are thought to separate and recombine many times, providing an opportunity to insert an inhibitory molecule in the protein interface between subunits. Selection of such molecules from chemical libraries or designing them from known

protein–protein interfaces prevents formation of the active enzyme molecule in the cell. Inhibitors of this type have been designed for ribonucleotide diphosphate reductase and show promise in this new area of specific inhibitor design.

Prevent Conformational Change for Catalysis

A related inhibitor type is one binding remote from the catalytic site to prevent an enzymatic conformational change that is required for catalysis. Most enzymes require protein motion to close the catalytic site prior to the formation of the transition state. These loops or domain motions involve physical motion of segments of the proteins, often over relatively large distances. A consequence of the motion is to position amino acid residues in the catalytic site where they are required for catalysis. An inhibitor that binds outside the catalytic site and prevents this protein motion required for catalysis will provide complete loss of enzyme function without blocking access of

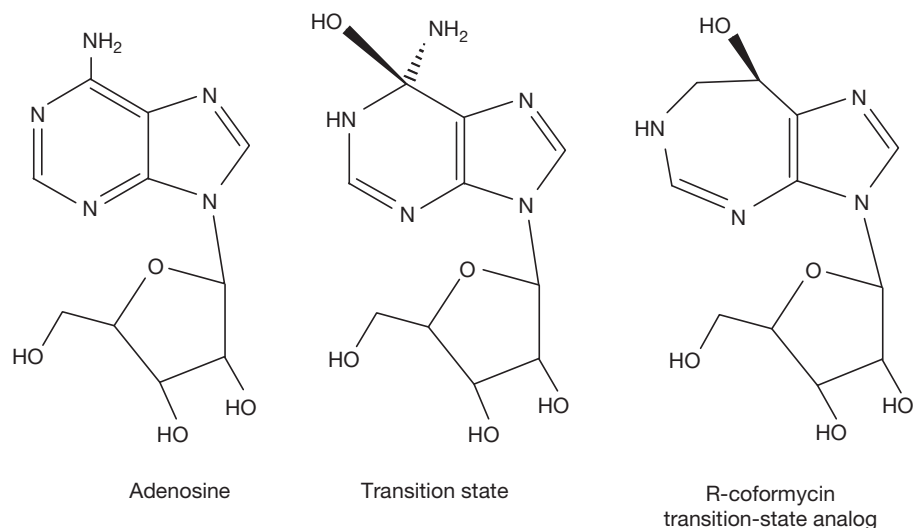


Figure 5 Reactant (adenosine), transition state, and transition-state analog inhibitor for adenosine deaminase. R-coformycin is a natural product transition-state analog inhibitor that binds 10 million times tighter to adenosine deaminase than does adenosine.

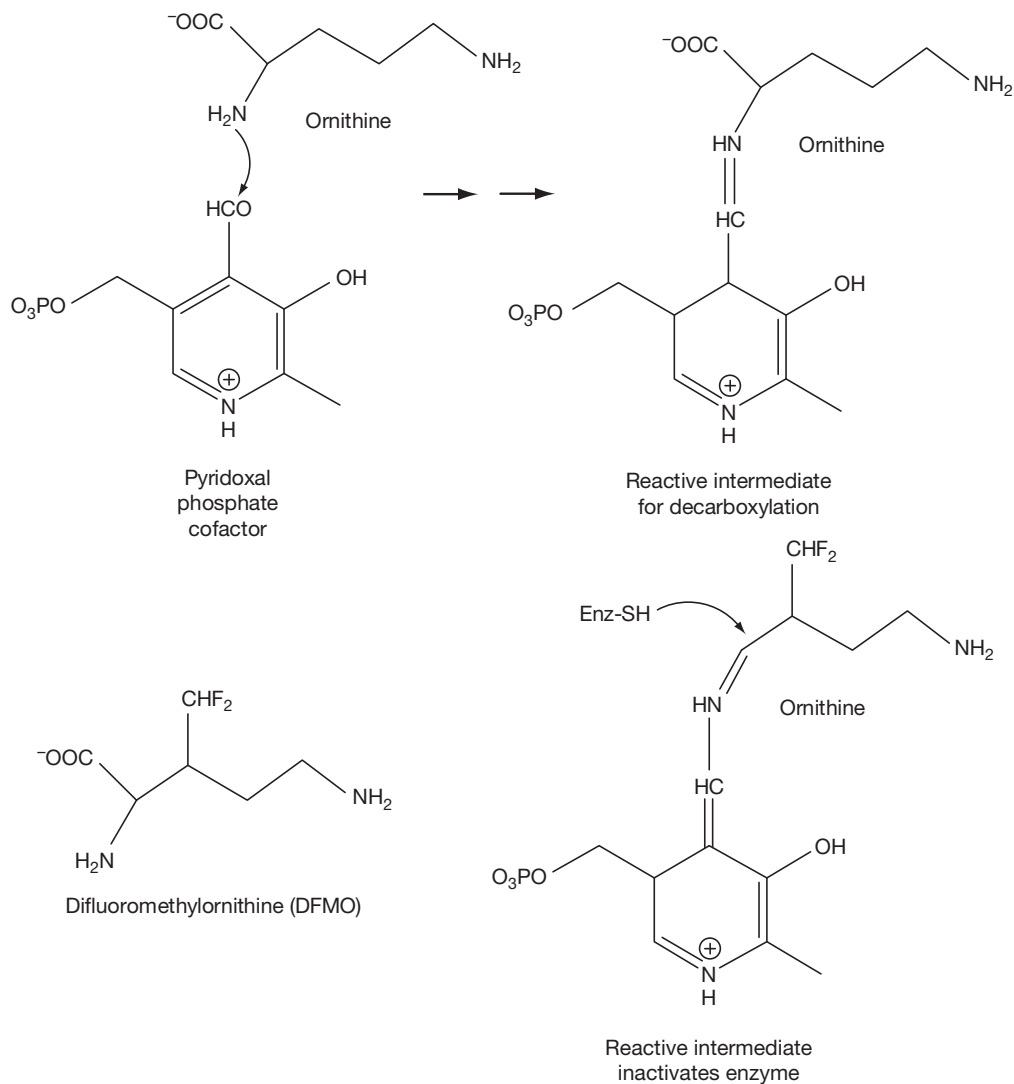


Figure 6 Covalent suicide inhibition of ornithine decarboxylase by DFMO. In the normal reaction, the enzymatic cofactor, pyridoxal phosphate, forms a Schiff base favoring decarboxylation of ornithine. With DMFO, the same steps occur but decarboxylation generates a reactive difluoro-conjugated imine. A nearby cysteine nucleophile reacts to form a stable, inactive enzyme.

reactants to the catalytic site. As knowledge of the protein surfaces becomes more complete, it will be increasingly possible to design molecules that participate in these novel inhibitory roles.

Mechanism-Based Inhibitors

Mechanism-based enzymatic inhibitors are designed to be stable molecules that resemble the substrate. They bind at the catalytic site of the target enzyme and are converted by the enzyme to a chemically reactive form that covalently reacts with the catalytic site of the enzyme. This covalent intermediate of enzyme inhibitor differs from normal catalytic intermediates by its chemical stability and long lifetime. Trapping the enzyme in a stable covalent complex blocks the catalytic site from normal interaction and provides long-term inhibition of the enzyme. Mechanism-based inhibitors are in broad use in biology, and our example of this chemical inhibitor logic is in the pathway of polyamines (Figure 6).

Polyamine Synthesis

The polycationic polyamines are essential counterions for DNA replication in dividing cells. The first step in polyamine synthesis is catalyzed by ornithine decarboxylase which uses a transient covalent intermediate to the enzymatic cofactor pyridoxal phosphate. Difluoromethylornithine (DFMO) is decarboxylated in similar fashion to ornithine, but the reaction creates a highly reactive intermediate that reacts with a nearby group on the enzyme to form a stable covalent intermediate, resulting in an inactive ornithine decarboxylase. It therefore stops polyamine biosynthesis. It has been used as an anticancer agent and an antibiotic against *Trypanosoma* species that cause sleeping sickness. Specificity is provided by the binding of DFMO only to the target enzyme and long biological action is provided by the stable covalent interaction with the enzyme. These are features of most mechanism-based enzyme inhibitors.

Care is needed that chemical reactivity causing enzymatic inactivation is revealed only to the target enzyme.

Other Mechanism-Based Inhibitors

Many of the common anti-metabolites are mechanism-based inhibitors. Thus, certain insecticides form covalent interactions with the catalytic site of enzymes involved in insect neurotransmission. Deprenyl is a drug used to treat Parkinson's disease and mental depression by increasing brain dopamine levels. It functions by suicide inhibition at the catalytic site of monoamine oxidase, the enzyme that degrades dopamine.

Conclusion

The intricate steps in enzymatic action can be disrupted in numerous ways, each causing inhibition of catalytic steps. Enzymes have evolved to function in the presence of the complex mixture of normal cellular components, but inhibitors can be designed to specifically target a single enzyme of the thousands that function in coordination in a normal cell.

See also: **Protein/Enzyme Structure Function and Degradation:** Enzyme Kinetics; Enzyme Reaction Mechanisms: Stereochemistry; Substrate Binding, Catalysis, and Product Release.

Further Reading

- Abeles RH and Alston TA (1990) Enzyme inhibition by fluoro compounds. *Journal of Biological Chemistry* 265: 16705–16708.
- Northrop DB (1999) Rethinking fundamentals of enzyme action. *Advances in Enzymology and Related Areas of Molecular Biology* 73: 25–55.
- Schramm VL (1998) Enzymatic transition states and transition state analog design. *Annual Review of Biochemistry* 67: 693–720.
- Schramm VL (2003) Enzymatic transition state poise and transition state analogues. *Accounts of Chemical Research* 36: 588–596.
- Wolfenden R (1976) Transition state analog inhibitors and enzyme catalysis. *Annual Review of Biophysics and Bioengineering* 5: 271–306.

Enzyme Kinetics

I H Segel, University of California, Davis, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 38–44, © 2004, Elsevier Inc.

Glossary

Catalytic rate constant (k_{cat}) Moles of substrate converted to product per second per mole of enzyme active site; it is also called active-site turnover number.

Kinetic mechanism The order in which substrates add and products are released in an enzyme-catalyzed reaction.

Maximal velocity (V_{max}) Maximum reaction rate (v observed at saturating substrate concentrations) for a given concentration of enzyme: $V_{\text{max}} = k_{\text{cat}}[E]_{\text{t}}$.

Michaelis constant (K_{m}) Kinetic constant equivalent to the concentration of the varied substrate that yields half-maximal velocity when all other substrates are saturating.

Rapid equilibrium Condition where the ligands and enzyme species present are at chemical equilibrium.

Steady state Condition where the rate of formation of an intermediate is equal to the rate of its utilization.

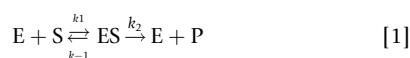
Unireactant Enzymes and the Michaelis–Menten Equation

By the mid- to late 1800s, most of the laws governing the effect of reactant concentrations on the rates of chemical reactions were known. However, enzyme-dependent reactions did not always follow these simple rate laws. In 1902, Victor Henri proposed the first useful equation for an enzyme-dependent reaction. Henri's equation took into account the known properties of enzymes, including that (1) an enzyme was highly specific for a particular substrate (or class of substrates), (2) the substrates often protected an enzyme from inactivation by heat, and (3) a plot of initial velocity versus substrate concentration was hyperbolic. (These properties all pointed to the participation of an enzyme–substrate complex.) Subsequent workers, including Michaelis and Menten (1913), Van Slyke and Cullen (1914), and Briggs and Haldane (1925), confirmed and/or extended Henri's equation, which is derived below.

The Rapid Equilibrium Assumption

The following assumptions and procedures led Henri and others to the relationship between substrate concentration and reaction rate, which is now known as the Michaelis–Menten equation:

1. The overall reaction occurs in two steps, as shown below:



This sequence recognizes that the enzyme is not consumed in the reaction, but rather is a recycling catalyst. The first step is the rapid equilibration of the free enzyme (E) and the free substrate (S) to form an ES complex. This is followed by the unidirectional, rate-limiting decomposition of ES to E + P. The rapid equilibrium between E, S, and ES is maintained throughout the course of the reaction. (This condition requires that $k_{-1} \gg k_2$.) The equilibrium is described by the dissociation constant, K_s , of the ES complex:

$$K_s = [E][S]/[ES] = k_{-1}/k_1 \quad [2]$$

Equation [2] allows the concentration of the ES complex to be expressed in terms of the concentration of free E, free S, and K_s :

$$[ES] = [S][E]/K_s \quad [3]$$

2. At any time, the total enzyme, $[E]_{\text{t}}$, is present either as free E or as the ES complex. Thus, the mass balance (conservation) equation is

$$[E]_{\text{t}} = [E] + [ES] \quad [4]$$

The total substrate is also distributed between the free and complexed species:

$$[S]_{\text{t}} = [S] + [ES] \quad [5]$$

However, for most enzyme-dependent reactions that are studied *in vitro*, the enzyme is present at a very low (catalytic) level (i.e., $[E]_{\text{t}} \ll [S]_{\text{t}}$), so that the maximum $[ES]$ that can form will be much smaller than the $[S]_{\text{t}}$ added. Consequently, it can be assumed that the free S concentration is essentially identical to the total added S (i.e., $[S] \approx [S]_{\text{t}}$). This assumption is not valid for many enzymes inside the cells, where $[E]_{\text{t}}$ may be in the same range as $[S]_{\text{t}}$.

3. The instantaneous or initial rate, $d[P]/dt$ or $-d[S]/dt$ (usually indicated as v for velocity), is proportional to the concentration of the ES complex:

$$v = k_2[ES] \quad [6]$$

4. It is assumed that the accumulating product has no effect on the reaction. That is, the P concentration remains close to zero throughout the course of the reaction, or P has no affinity for the enzyme.
5. The final velocity equation is obtained by dividing eqn [6] by eqn [4] and substituting for $[ES]$ from eqn [3]:

$$\frac{v}{[E]_{\text{t}}} = \frac{k_2[ES]}{[E] + [ES]} \quad \text{or} \quad v = \frac{k_2[E]_{\text{t}} \frac{[S][E]}{K_s}}{[E] + \frac{[S][E]}{K_s}} \quad [7]$$

$k_2[E]_t$ is defined as $V_{\max}$ because when all the enzyme is in the ES form, the observed velocity will be maximal. This substitution and canceling $[E]$ in eqn [7] yields the final Michaelis–Menten equation:

$$v = \frac{V_{\max} \frac{[S]}{K_s}}{1 + \frac{[S]}{K_s}} = \frac{V_{\max}[S]}{K_s + [S]} \quad [8]$$

The Steady-State Assumption

By the 1920s, researchers realized that the rapid equilibrium assumption was too restrictive and would not be valid for enzymes where the ES complex proceeds on to E + P much faster than it dissociates back to E + S. In such cases, E, S, and ES would not attain equilibrium. In 1925, Briggs and Haldane offered a more general method for deriving velocity equations that did not require a rapid equilibrium assumption. Their steady-state approach recognized that under most assay conditions, $[E]$ and $[ES]$ would attain certain levels very quickly after mixing E and S, but that the concentrations of these enzyme species would change very little thereafter. In other words, for the duration of the assay, the rate at which ES is formed (from E + S) remains identical to the rate at which it decomposes (back to E + S, plus forward to E + P). Similarly, the rate at which E is produced from ES equals the rate at which E is used to form ES. Equation [1] is still applicable, but now rate constants for individual steps must be taken into account. Applying the steady-state assumption we can write

$$d[ES]/dt = 0 \text{ or } k_1[E][S] = (k_{-1} + k_2)[ES] \quad [9]$$

and

$$d[E]/dt = 0 \text{ or } (k_{-1} + k_2)[ES] = k_1[E][S] \quad [10]$$

Solving either equation for $[ES]$ yields

$$[ES] = k_1[E][S]/(k_{-1} + k_2) \text{ or } [ES] = [S][E]/K_m \quad [11]$$

where in the latter expression, the three rate constants are combined into a single kinetic constant, K_m , as shown below:

$$K_m = (k_{-1} + k_2)/k_1 \quad [12]$$

The initial velocity (v) at any $[S]$ still equals $k_2[ES]$. So, proceeding as described earlier, but substituting for $[ES]$ from eqn [11], we obtain

$$v = \frac{V_{\max}[S]}{K_m + [S]} \quad [13]$$

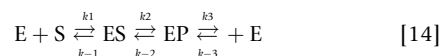
Thus, the velocity equation is the same as that derived for rapid equilibrium conditions, except that the constant in the denominator (now called a Michaelis constant) is more complex than a simple dissociation constant. The Briggs–Haldane approach is more general than the rapid-equilibrium approach because it makes no assumptions about the relative magnitudes of the various rate constants. If $k_2 \ll k_{-1}$, K_m reduces to k_{-1}/k_1 , which is equivalent to the simpler K_s .

Equation [13] is valid throughout the experimental $[S]$ range provided that the rate of P formation is measured before $[S]$ changes appreciably. (In practice, substrate use should be restricted to less than 5% (or at most 10%) of $[S]_t$.) When

$[S] \ll K_m$, eqn [13] reduces to $v = V_{\max}[S]/K_m = k[S]$, that is, the reaction is first order with respect to $[S]$ with a rate constant equivalent to $V_{\max}/K_m$. When $[S] \gg K_m$, v is essentially constant at $V_{\max}$ and independent of $[S]$. (The reaction is zero order with respect to $[S]$.)

A More Realistic Unireactant Reaction Sequence

Equation [1] contains the minimum number of steps that must be considered in the description of a uni–uni (one substrate, one product) reaction. A more realistic scheme takes into account the conversion of enzyme-bound S to enzyme-bound P (before P is released) and the reversibility of each step:



A steady-state treatment of the above reaction sequence once again yields the familiar Michaelis–Menten equation when $[P] = 0$, except that now

$$K_m = \frac{k_{-1}k_3 + k_{-1}k_{-2} + k_2k_3}{k_1(k_2 + k_{-2} + k_3)} \quad [15]$$

and

$$V_{\max} = \frac{k_2k_3[E]_t}{k_2 + k_{-2} + k_3} \quad [16]$$

So, in the absence of other information, the rate constant compositions of the two kinetic constants, K_m and $V_{\max}$, are not usually known. But regardless of its makeup, K_m is a valuable characteristic because it is equivalent to the substrate concentration that yields half-maximal velocity. $V_{\max}$ is not a true constant because it depends on the enzyme concentration. However, $V_{\max}/[E]_t$, expressed as moles of S converted to P per second per mole of enzyme active site (which reduces to per second, or s^{-1}) is the catalytic rate constant (or turnover number) of the enzyme, k_{cat} . The ratio k_{cat}/K_m has been used to compare the action of an enzyme on a series of related substrates. The higher the value of this specificity constant, the better the substrate. If an enzyme is evolutionarily perfected so that the catalytic and product release rate constants are very much larger than any of the reverse first-order rate constants, the ratio k_{cat}/K_m reduces to k_1 , the second-order rate constant for the interaction of E and S to form ES. In aqueous solution, this constant has a maximum, diffusion-limited value of $\sim 10^9 \text{ M}^{-1} \text{ s}^{-1}$. Many enzymes have k_{cat}/K_m values that approach this maximum.

Reversible Reactions

In the presence of both the substrate and product, the net steady-state velocity is given by the difference between the forward and reverse rates of any step, for example,

$$v = k_2[ES] - k_{-2}[EP] \quad [17]$$

or

$$\frac{v}{[E]_t} = \frac{k_2[ES] - k_{-2}[EP]}{[E] + [ES] + [EP]} \quad [18]$$

After substituting for $[ES]$ and $[EP]$ and grouping rate constants into kinetic constants, the equation is

$$v = \frac{V_{\max, f} \frac{[S]}{K_{mS}} - V_{\max, r} \frac{[P]}{K_{mP}}}{1 + \frac{[S]}{K_{mS}} + \frac{[P]}{K_{mP}}} \quad [19]$$

where $V_{\max, f}$ and $V_{\max, r}$ are the forward and reverse maximal velocities, respectively; K_{mS} and K_{mP} are the Michaelis constants of S and P, respectively. Equation [19] can also be written as

$$v = \frac{V_{\max, f} [S]}{K_{mS} \left(1 + \frac{[P]}{K_{mP}}\right) + [S]} - \frac{V_{\max, r} [P]}{K_{mP} \left(1 + \frac{[S]}{K_{mS}}\right) + [P]} \quad [20]$$

which shows that in this unireactant process, P acts as a competitive inhibitor of the forward reaction and S acts as a competitive inhibitor of the reverse reaction. (In reactions catalyzed by multireactant enzymes, individual products might be competitive, uncompetitive, or noncompetitive with respect to different substrates.)

The Velocity Curve and Its Linear Forms

Figure 1 shows the typical hyperbolic dependence of the velocity of an enzyme-catalyzed reaction on the substrate concentration. The kinetic constants, K_m and $V_{\max}$, can be obtained by fitting the v versus $[S]$ data directly to the Michaelis–Menten equation using an appropriate computer application. Alternatively, K_m and $V_{\max}$

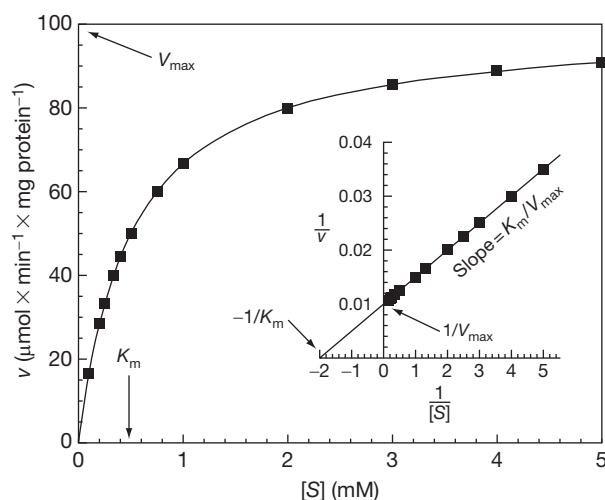


Figure 1 Theoretical plot of initial velocity vs. substrate concentration in the absence of the product for an enzyme with a K_m of 0.5 mM and a $V_{\max}$ of 100 $\mu\text{mol min}^{-1} \text{mg}^{-1}$ protein. The $[S]$ at which v is 50% of $V_{\max}$ (i.e., $[S]_{0.5}$) is equivalent to K_m . The curvature of the plot is the same for all enzymes obeying Henri–Michaelis–Menten kinetics, regardless of the absolute values of the kinetic constants. For example, the ratio of substrate concentration that yields 90% of $V_{\max}$ ($[S]_{0.9}$) to that yielding 10% of $V_{\max}$ ($[S]_{0.1}$) is always 81. Similarly, the $[S]_{0.75}/[S]_{0.5}$ ratio is always 3. Note that $V_{\max}$ is difficult to determine visually. But a computer-assisted fit to $v = V_{\max} [S] / (K_m + [S])$ will return both $V_{\max}$ and K_m . (Inset) Double-reciprocal plot of the data. A wide range of substrate concentrations is needed to construct both the primary plot and the double-reciprocal plot because $[S]$ values that are evenly spaced on one plot do not space evenly on the other.

can be obtained by plotting the data in a linear form such as $1/v$ versus $1/[S]$ (usually called the Lineweaver–Burk or double-reciprocal plot), $v/[S]$ versus v (usually called the Scatchard plot), v versus $v/[S]$, or $[S]/v$ versus $[S]$. The most popular of these is the double-reciprocal plot, which is based on the linear transformation of the Michaelis–Menten equation:

$$\frac{1}{v} = \frac{K_m}{V_{\max}} \frac{1}{[S]} + \frac{1}{V_{\max}} \quad [21]$$

A major advantage of a linear plot is that data obtained at several different fixed concentrations of a second ligand (e.g., a co-substrate, product, inhibitor, or activator) can be displayed simultaneously and the resulting slope and intercept characteristics used to diagnose the kinetic mechanism or the mode of inhibition or activation.

Multireactant Enzymes

Common Mechanisms

Most enzymes catalyze reactions between two or more substrates to yield two or more products. The substrates bind to the enzyme either in a random manner or in a compulsory order. Similarly, product release is random in some cases, and ordered in the others. Mechanisms in which catalysis occurs only after two (or more) substrates are brought together on the enzyme are called sequential. Some enzymes catalyze the formation and release of a product before all the substrates have bound. For example, an enzyme might capture a substrate, A, and retain part of that molecule (sometimes as a covalent adduct with the enzyme) releasing the rest of the molecule as product P. Only then does substrate B bind and combine with the part of A that was left on the enzyme to form product Q, which then leaves. This type of substituted enzyme or ping-pong mechanism is common for transaminases; the amino group of substrate A is left temporarily on the enzyme. A reaction catalyzed by a terreactant (three-substrate) enzyme might proceed by a combination of sequential and ping-pong steps. The order of addition of substrates and release of products is called the kinetic mechanism. Four bireactant examples are shown in Figure 2.

Velocity Equations for Some Bireactant Mechanisms

In some cases, different kinetic mechanisms are described by different velocity equations. (see standard enzyme kinetics texts for the procedures used to derive these equations). Consequently, it is sometimes possible to distinguish between mechanisms from the characteristics of the double-reciprocal plots. For example, the velocity equation for an ordered bireactant mechanism under steady-state conditions in the absence of products is

$$v = \frac{V_{\max} [A][B]}{K_{ia} K_{mB} + K_{mB} [A] + K_{mA} [B] + [A][B]} \quad [22]$$

where K_{mA} and K_{mB} are the Michaelis constants of the two substrates and K_{ia} is the simple dissociation constant of the EA complex. (Velocity equations for multireactant mechanisms often contain both types of constants.) When $[A]$ is varied at different fixed concentrations of B, the velocity

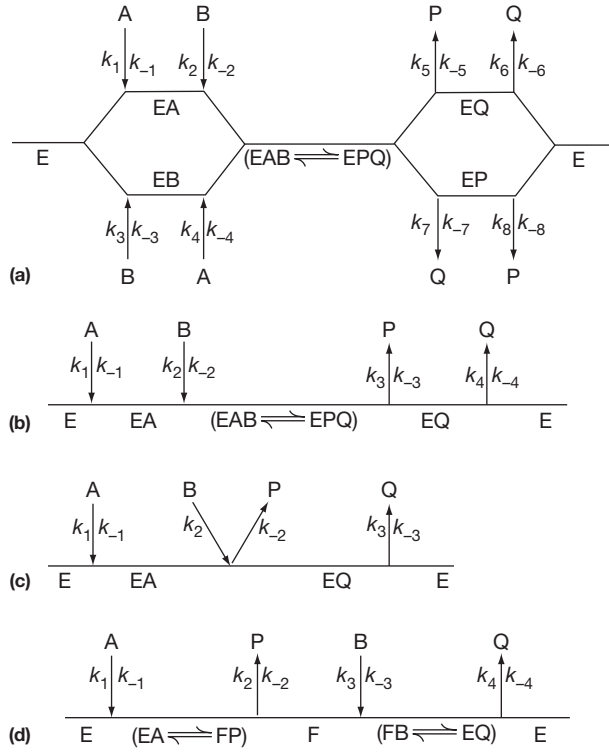


Figure 2 Four common kinetic mechanisms for bireactant enzymes are shown in the Cleland shorthand notation. Substrates are indicated above a horizontal line as A, B, C, etc., in the order in which they add to the enzyme. Products are indicated as P, Q, R, etc., in the order in which they dissociate from the enzyme. Arrows are shown pointing in a single direction, but all the reactions are reversible. Rate constants for each reversible reaction are shown alongside arrows. For computer-based derivations of the velocity equation, it is usually more convenient to indicate all rate constants with positive subscripts (odd numbers for the forward steps; even numbers for the reverse direction). The various enzyme species present in the mechanism are placed below the horizontal line (in sequence of their formation). Species that are interconverted by a unimolecular step are usually grouped together within parentheses. (a) Random mechanism, (b) ordered mechanism, and (c) Theorell–Chance mechanism. The last is a limiting case of an ordered mechanism in which the EAB + EPQ central complex does not account for a significant fraction of the total enzyme in the steady state. (d) Ping-pong mechanism where stable enzyme forms F may be a covalent adduct of E with the part of substrate A left behind.

dependence can be written in the form of the Michaelis–Menten equation:

$$v = \frac{V_{\max}[A]}{K_{mA}\left(1 + \frac{K_{ia}K_{mB}}{K_{mA}[B]}\right) + [A]\left(1 + \frac{K_{mB}}{[B]}\right)} \quad [23]$$

or, in double reciprocal form:

$$\frac{1}{v} = \frac{K_{mA}}{V_{\max}} \left(1 + \frac{K_{ia}K_{mB}}{K_{mA}[B]}\right) \frac{1}{[A]} + \frac{1}{V_{\max}} \left(1 + \frac{K_{mB}}{[B]}\right) \quad [24]$$

When [B] is varied at different fixed concentrations of A, the equations are

$$v = \frac{V_{\max}[B]}{K_{mB}\left(1 + \frac{K_{ia}}{[A]}\right) + [B]\left(1 + \frac{K_{mB}}{[A]}\right)} \quad [25]$$

and

$$\frac{1}{v} = \frac{K_{mB}}{V_{\max}} \left(1 + \frac{K_{ia}}{[A]}\right) \frac{1}{[B]} + \frac{1}{V_{\max}} \left(1 + \frac{K_{mB}}{[A]}\right) \quad [26]$$

At a fixed concentration of one substrate, for example, A, the velocity dependence on the concentration of substrate B can also be described by a simplified form of eqn [25]:

$$v = \frac{V_{\max, \text{app}}[B]}{K_{mB, \text{app}} + [B]} \quad [27]$$

where the apparent constants are related to the true or limiting constants:

$$V_{\max, \text{app}} = \frac{V_{\max}}{\left(1 + \frac{K_{mB}}{[A]}\right)} \quad \text{and} \quad K_{mB, \text{app}} = \frac{\left(1 + \frac{K_{ia}}{[A]}\right)}{\left(1 + \frac{K_{mB}}{[A]}\right)} \quad [28]$$

In other words, the experimentally determined K_m and $V_{\max}$ depend on the concentration of the co-substrate and will not equal the limiting values unless that concentration is infinitely high (saturating). In practice, the limiting kinetic constants are obtained from appropriate replots of the slopes and $1/v$ -axis intercepts of the double-reciprocal plots. (The former is a replot of $K_{m, \text{app}}/V_{\max, \text{app}}$; the latter is a replot of $1/V_{\max, \text{app}}$.) For example, the replots for the family of $1/v$ versus $1/[B]$ plots at different fixed [A] are described by the following linear equations:

$$\text{int} = \frac{K_{mB}}{V_{\max}} \frac{1}{[A]} + \frac{1}{V_{\max}} \quad [29]$$

$$\text{slope} = \frac{K_{mB}K_{ia}}{V_{\max}} \frac{1}{[A]} + \frac{K_{mB}}{V_{\max}} \quad [30]$$

$V_{\max}$, K_{mB} , and K_{ia} can be determined from the intercepts of the replots.

The equation for the substituted enzyme (ping-pong) bireactant mechanism in the absence of products is

$$v = \frac{V_{\max}[A][B]}{K_{mB}[A] + K_{mA}[B] + [A][B]} \quad [31]$$

When [A] is varied at different [B], the Michaelis–Menten and double reciprocal equations are

$$v = \frac{V_{\max}[A]}{K_{mA} + \left(1 + \frac{K_{mB}}{[B]}\right)} \quad [32]$$

and

$$\frac{1}{v} = \frac{K_{mA}}{V_{\max}} \frac{1}{[A]} + \frac{1}{V_{\max}} \left(1 + \frac{K_{mB}}{[B]}\right) \quad [33]$$

The equations for varied [B] at different [A] are symmetrical to those shown for varied [A]. **Figure 3** shows the families of $1/v$ versus $1/[A]$ plots for an ordered and a ping-pong mechanism. In both mechanisms different fixed [B] values affect the vertical-axis intercepts. In the ordered mechanism, changing [B] also affects the slopes of the $1/v$ versus $1/[A]$ plots. However in the ping-pong mechanism, changing the fixed [B] has no effect on the slopes. (Note that eqns [24] and [33] predict these effects.) The absence of a slope effect is often sufficient to identify the ping-pong mechanism. But the presence of both the slope and intercept effects (as shown in **Figure 3(a)**) is not

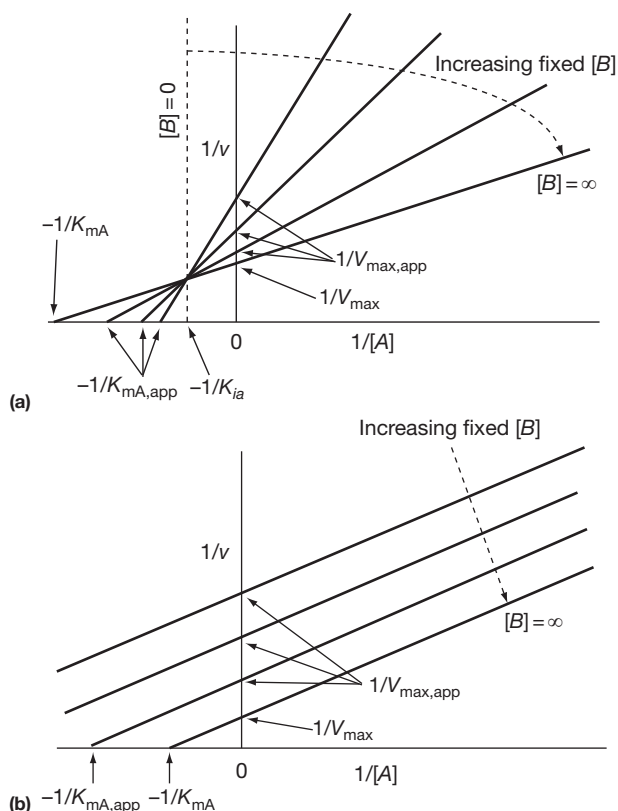


Figure 3 Double-reciprocal plots of $1/v$ vs. $1/[A]$ at different fixed concentrations of B. A is called the varied substrate; B is called the changing fixed substrate. (a) A sequential mechanism. The plots could be for a steady-state ordered, Theorell–Chance, or a rapid-equilibrium random mechanism. (b) Ping-pong mechanism. For the above mechanisms, the plots of $1/v$ vs. $1/[B]$ at different fixed $[A]$ have the same appearance as for varied $1/[A]$. The limiting kinetic constants, $K_{m,A}$, $K_{m,B}$, $K_{i,A}$, V_{max} , etc., are obtained by replotting the slopes and/or $1/v$ -axis intercepts of the double-reciprocal plots against the reciprocal of the changing fixed substrate concentration.

sufficient to diagnose the mechanism as ordered because in the absence of products, a steady-state ordered, a Theorell–Chance, and a rapid-equilibrium random mechanism all have the same velocity equation. Consequently, the double-reciprocal plot patterns of the three different mechanisms are the same. Additional measurements are needed to distinguish between these mechanisms. Product inhibition studies can be informative. For example, in a steady-state ordered mechanism, one of the products (P, the first to dissociate from the enzyme) is noncompetitive with respect to both substrates, but in a Theorell–Chance and random mechanism, each product is competitive with at least one substrate. Dead-end inhibition studies may also be useful. For example, a nonreactive analog that is competitive with B will be uncompetitive with A in the ordered and Theorell–Chance mechanisms, but noncompetitive with A in the random mechanism. (This assumes that the inhibitor forms an EAI complex in the Theorell–Chance mechanism and mimics substrate B completely in the random mechanism by binding to both E and EA.)

Cleland's Slope and Intercept Rules

WW Cleland has presented a series of rules for predicting the effect of changing the fixed concentration of a co-substrate or product on the slopes and vertical-axis intercepts of $1/v$ versus $1/[\text{varied substrate}]$ plots. The slope and intercept effects are predicted separately and then combined to deduce the appearance of the double-reciprocal plot family. Conversely, the slope and intercept patterns can help diagnose the kinetic mechanism. The rules are as follows.

Intercept effect

A changing fixed substrate or product will cause an intercept effect if it and the varied substrate combine with different enzyme forms. However, there are exceptions: an intercept effect will not be seen for a changing fixed ligand that adds to the enzyme before the varied substrate under rapid-equilibrium conditions, or in a Theorell–Chance hit-and-run sequence between a substrate and the immediately released product (B and P in Figure 2(c)).

Slope effect

A changing fixed substrate or product will cause a slope effect if it and the varied substrate combine with the same enzyme form, or with different forms that are reversibly connected in the upstream or downstream direction. Two enzyme forms will not be reversibly connected if either (1) a product-release step occurs between them and that product is at zero concentration in the assay solution, or (2) a substrate-addition step occurs between the two enzyme forms and that substrate is present at a saturating concentration. Both the upstream and downstream directions must be blocked to eliminate reversibility.

A product will have no effect on slopes or intercepts if the steady-state level of the enzyme form with which it combines is zero because of saturation with a competitive substrate. Similar rules predict the slope and intercept effects produced by dead-end inhibitors.

See also: Protein/Enzyme Structure Function and Degradation: Allosteric Regulation; Enzyme Inhibitors.

Further Reading

- Cleland WW (1963) The kinetics of enzyme-catalyzed reactions with two or more substrates or products. *Biochimica et Biophysica Acta* 67: 104–137; 173–187; and 188–196.
- Cornish-Bowden A (1995) *Fundamentals of Enzyme Kinetics*. London: Portland Press.
- Fromm HJ (1975) *Initial Rate Enzyme Kinetics*. Berlin: Springer.
- Gutfreund H (1995) *Kinetics for the Life Sciences*. Cambridge: Cambridge University Press.
- Kuby SA (1991) *A Study of Enzymes, Vol. 1: Enzyme Catalysis, Kinetics, and Substrate Binding*. Boca Raton, FL: CRC Press.
- Purich DL and Allison RD (2000) *Handbook of Biochemical Kinetics*. San Diego, CA: Academic Press.
- Segel IH (1976) *Biochemical Calculations*, 2nd edn., ch. 4. New York, NY: Wiley.
- Segel IH (1993) *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady State Enzyme Systems*. Wiley Classics Library. New York, NY: Wiley-Interscience.

Enzyme Reaction Mechanisms: Stereochemistry

M-D Tsai, L Zhao, and B J Lamarche, The Ohio State University, Columbus, OH, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 45–50, © 2004, Elsevier Inc.

Glossary

Enzyme-catalyzed reaction Although usually stereospecific, many of them are also stereoselective. The stereochemical property of an enzymatic reaction is useful in elucidating its reaction mechanism.

Stereoselectivity The ability of an enzyme to choose from two or more possible stereoisomers as a preferred substrate, to choose from two enantiotopic groups or from two

diastereotopic groups, or to produce one stereoisomer as a preferred product when more than one stereoisomeric products are possible.

Stereospecificity The ability of an enzyme to convert a particular stereoisomeric substrate to a specific stereoisomeric product. For a substitution reaction, the stereochemical course of a stereospecific reaction can be retention or inversion.

Explanation of Stereochemical Terms

Figure 1 uses chemical structures to illustrate key stereochemical terms. Stereoisomers (or stereo-isomers) are molecules that have the same molecular formula and the same order of attachment of atoms, but they differ in the way their atoms are oriented in space. Any molecule that cannot be superimposed on its mirror image is said to be chiral. Though chirality can result from a number of molecular properties, it is usually due to the presence of one or more carbon or phosphorus atoms, called chiral centers, which are surrounded by four unique substituents. Chiral molecules are classified as either enantiomers (stereoisomers that are mirror images of each other) or diastereomers (stereoisomers that are not mirror images of each other).

Configuration specifies how substituents are oriented around a chiral center and is determined by the *R*–*S* notation system, where *R* and *S* refer to the Latin terms *rectus* (right) and *sinister* (left). To assign *R*/*S* configuration, a priority rating is given to each substituent at a chiral center using the following rules: priority is first assigned on the basis of atomic number with highest priority given to the atom with the highest atomic number. When identical atoms are directly attached to the chiral center, the next atoms on the substituent are compared, until a difference is noted. Once priority has been established, the molecule is oriented so that the lowest priority substituent is pointing away from the viewer. The direction of rotation observed when moving from the highest to lowest priority will now be either clockwise or counterclockwise and the configuration is designated *R* or *S*, respectively.

An atom that is attached to two unique substituents and two identical substituents is said to be prochiral. The two identical groups at a prochiral center can be designated as pro-*R* and pro-*S* following a prochirality rule. As shown in Figure 1(c), the pro-*R* hydrogen (H_R) is the one that leads to *R* configuration if it is given a higher priority than the other hydrogen atom. When a chiral center or prochiral center is phosphorus instead of carbon, a subscript *p* is added (i.e., R_p , S_p , pro- R_p , and pro- S_p).

Stereoselectivity, Stereospecificity, and Stereochemical Course of Enzymatic Reactions

If exposed to a racemic mixture of substrates, most enzymes utilize one enantiomer or diastereomer preferentially. In reactions where chemistry is occurring at a prochiral center, an enzyme is also likely to act on just one of the two available enantiotopic or diastereotopic groups. The enzyme is said to be stereoselective in both cases.

Enzymatic reactions are usually stereospecific, which means that they convert one stereoisomer of substrate to one stereoisomer of product. For substitution reactions, stereochemical course refers to the change (or lack thereof) in configuration at the reaction center in a stereospecific reaction. While retention means the new group occupies the same position as the displaced group, inversion means it occupies the opposite position. If the reaction is nonstereospecific, it leads to racemization, with both isomers being formed at a ratio close to 1:1. Note that an *R*-substrate does not necessarily give an *S*-product in an inversion reaction, since the priority of the substituents at the reaction center can be altered on going from substrate to product. Also note that there is a steric course (inversion or retention) even when the displacement occurs at a prochiral center or a pro-prochiral center (a methyl group or phosphoryl group). The three hydrogen atoms of a methyl group and the three oxygen atoms of a phosphoryl group are homotopic and cannot be differentiated from one another by an enzyme. In these cases, there is no issue of stereoselectivity, only steric cause. Experimentally, pro-prochiral centers can be made chiral via isotope substitution, which in turn allows elucidation of the steric course of the reaction as illustrated in Figure 2.

In addition to the use of oxygen isotopes, an oxygen atom at a prochiral or pro-prochiral center can also be replaced by a sulfur atom. These phosphorothioate analogs have found broad applications in the study of diverse biological systems. The stereochemical course of phosphatases, which convert a phosphomonoester to inorganic phosphate, have been deciphered in this manner, as illustrated in Figure 2(d).

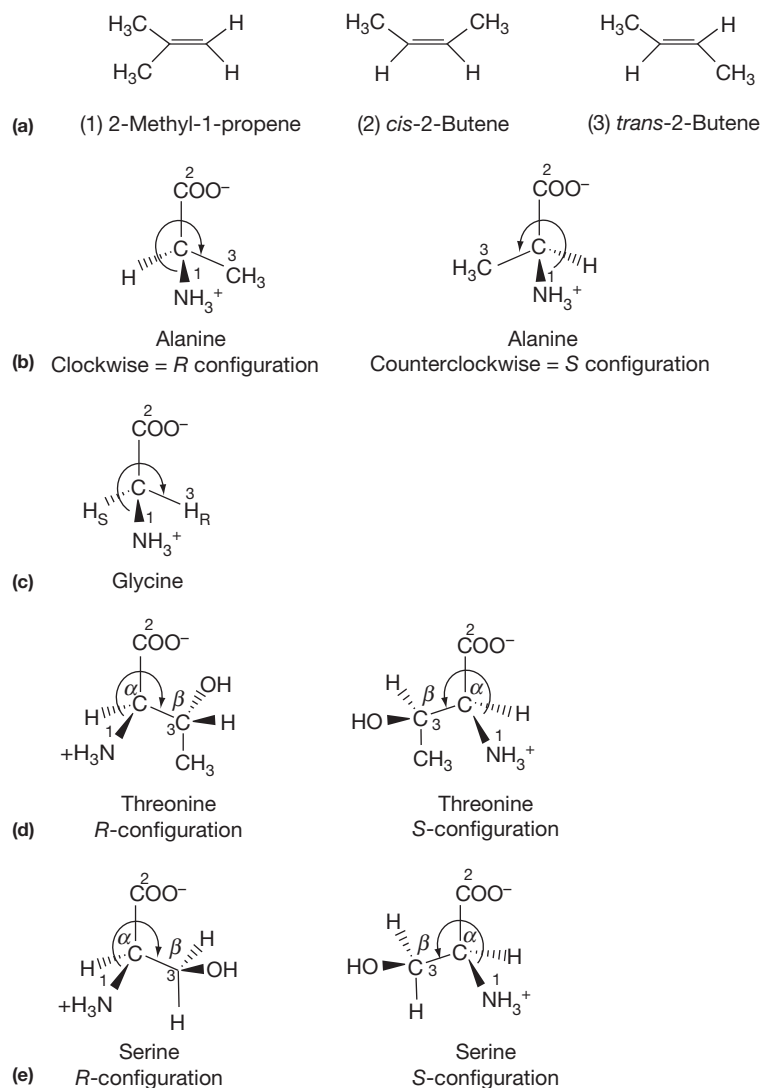


Figure 1 Demonstration of stereochemical terms using structures. (a) Though molecules 1, 2, and 3 have the same molecular formula, only 2 and 3 are stereoisomers. Molecule 1 has a different pattern of bonding (i.e., a different linkage of atoms) and is therefore a structural isomer of molecules 2 and 3. (b) Alanine stereoisomers are enantiomers since they are mirror images of each other. Note that the different configuration of substituents about the central carbon makes it impossible to superimpose the two isomers. (c) Glycine is achiral and therefore does not have stereoisomers. However, the central carbon is a prochiral center since it can be converted to a chiral center by replacing one of the two hydrogens with a different group. The two hydrogens are said to be enantiotopic since the separate replacement of each hydrogen results in a pair of enantiomers. (d) The two stereoisomers of threonine shown are diastereomers, that is, stereo-isomers that are not complete mirror images of each other. Note that this pair of molecules results by altering the configuration at C_α while leaving the configuration at the C_β unmodified. (e) Serine contains a chiral center at C_α and a prochiral center at C_β . Unlike glycine, the two hydrogen atoms at the prochiral center of serine are diastereotopic groups since the separate replacement of each hydrogen results in a pair of diastereomers.

Key steps in conducting stereochemical experiments are the synthesis of isomers of chiral substrates and/or substrate analogs and the analysis of a compound's stereochemical configuration before and after being processed by an enzyme. Since different substrate analogs are required for each enzyme studied, and since isotopes are frequently used, these studies are particularly challenging.

Pioneering Study: Alcohol Dehydrogenase

Yeast alcohol dehydrogenase (ADH) catalyzes the reversible transfer of a hydrogen atom between a molecule of ethanol

and a molecule of the coenzyme nicotinamide-adenine dinucleotide (NAD^+). By the use of deuterated ethanol or coenzyme, ADH was shown to abstract the pro-*R* hydrogen of ethanol and adding it to the *re* face of NAD^+ , as illustrated in Figure 3.

Analysis of Steric Course to Probe an Enzymatic Reaction Pathway

In nonenzymatic reactions, a bimolecular S_N2 reaction leads to inversion of configuration of the chiral carbon, while a unimolecular S_N1 reaction generates a planar carbonium ion that

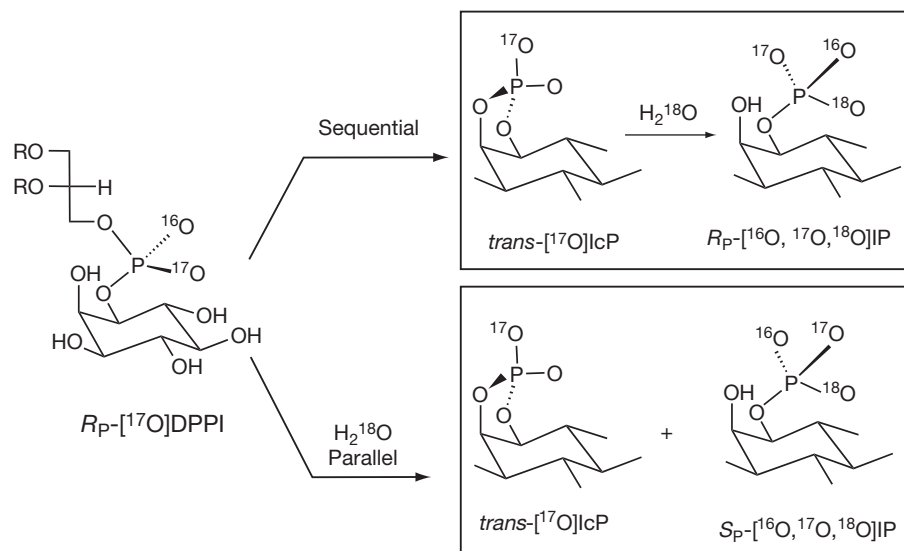


Figure 4 Proposed sequential and parallel mechanisms for the formation of IcP and IP catalyzed by mPI-PLC. In the sequential mechanism, intramolecular attack by the axial ring hydroxyl forms the cyclic intermediate (IcP), which is then broken down by hydrolysis (H_2^{18}O) to give R_P -[^{16}O , ^{17}O , ^{18}O] IP. In the parallel mechanism, intramolecular attack by the axial ring hydroxyl and direct water (H_2^{18}O) attack gives S_P -[^{16}O , ^{17}O , ^{18}O] IP, respectively.

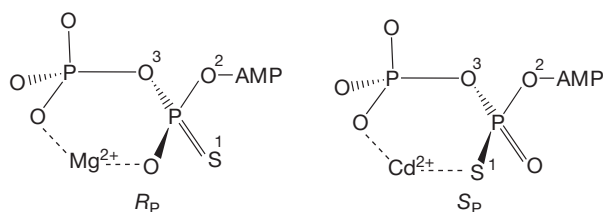


Figure 5 Metal-nucleotide interactions showing that substitution of Mg^{2+} by Cd^{2+} leads to a change in the stereoselectivity of AK from R_P -ATP β S to S_P -ATP β S.

interactions. A reversal (or a dramatic change) of stereoselectivity in the presence of an unnatural metal ion or upon mutating an active site residue provides strong evidence that the metal ion or mutated residue interacts with that particular phosphate moiety. Studies of adenylate kinase (AK) described below further illustrate this point.

AK catalyzes the reversible phosphoryl transfer reactions: $\text{Mg}^{2+} \cdot \text{ATP} + \text{AMP} \rightleftharpoons \text{Mg}^{2+} \cdot \text{ADP} + \text{ADP}$. Of the R_P and S_P isomers of ATP β S shown in Figure 5, AK prefers the S_P isomer by a factor of 9 in the presence of Mg^{2+} . However, this stereoselectivity is reversed (R_P isomer is preferred by a factor of 10) when Mg^{2+} is substituted by Cd^{2+} . Since Cd^{2+} preferentially coordinates sulfur over oxygen, which is opposite to the ligand preference of Mg^{2+} , the switch of stereospecificity upon Cd^{2+} substitution is strong evidence for direct coordination of the metal ion by the pro-S oxygen of the β -phosphate of ATP. Additional mechanistic information for AK was derived using active site mutants. Wild-type AK catalyzes the stereoselective conversion of adenosine monophosphates (AMPs) to S_P -ADP α S. Such stereoselectivity could be achieved by restricting orientation of the P-S and P-O bonds at the active site as shown in Figure 6. Experimental results demonstrated that

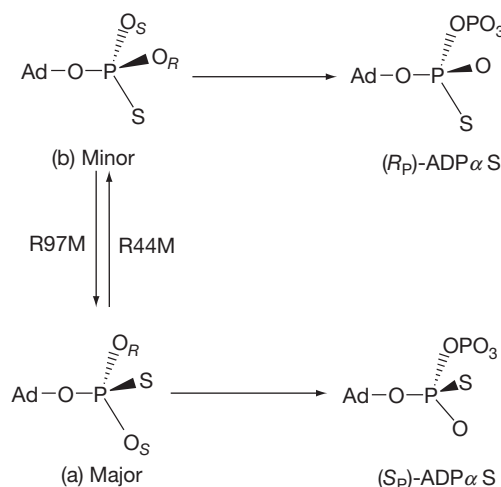


Figure 6 Enhancement and reversal of adenylate kinase stereoselectivity by site-directed mutagenesis (Ad, adenosine). In the reaction catalyzed by WT protein, phosphate is preferentially attached to the pro- R oxygen of AMPs. The R97M and R44M mutations enhance and reverse this stereoselectivity, respectively. This provides strong evidence that R97 and R44 interact with the phosphate moiety of AMP at the transition state. (a) and (b) represent two conformers of AMPs when bound to the active site of the enzyme.

the stereoselectivity of this reaction is further enhanced by the R97M mutant, but is completely reversed in the reaction catalyzed by the R44M mutant. These results indicate that the conformational equilibrium of the AMP substrate in the active site is perturbed in both mutants, suggesting that the Arg97 and Arg44 residues interact with the phosphate moiety of AMPs.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Carbohydrate Chains: Enzymatic and Chemical Synthesis; **Protein/Enzyme Structure Function and Degradation:** Enzyme Inhibitors; Substrate Binding, Catalysis, and Product Release.

Further Reading

Bruzik KS and Tsai M-D (1991) Phospholipase stereospecificity at phosphorus. *Methods in Enzymology* 197: 258–269.

Fersht A (1998) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. New York, NY: Freeman.

Floss HG and Tsai M-D (1979) Chiral methyl groups. *Advances in Enzymology* 50: 243–302.

Knowles JR (1980) Enzyme-catalyzed phosphoryl transfer reactions. *Annual Review of Biochemistry* 49: 877–919.

Tomasselli AG and Noda LH (1983) Baker's yeast adenylate kinase. Evidence of conformational change from intrinsic fluorescence and difference spectra. Determination of the structure of enzyme-bound metal-nucleotide by use of phosphorothioate analogues of ATP. *European Journal of Biochemistry* 132: 109–115.

Tsai M-D (1982) Use of $^{31}\text{P}(^{18}\text{O})$, $^{31}\text{P}(^{17}\text{O})$, and ^{17}O NMR methods to study enzyme mechanisms involving phosphorus. *Methods in Enzymology* 87: 235–279.

Tsai M-D, Jiang RT, Dahnke T, and Shi Z (1995) Manipulating phosphorus stereospecificity of adenylate kinase by site-directed mutagenesis. *Methods in Enzymology* 249: 425–443.

ER/SR Calcium Pump: Function

G Inesi, California Pacific Medical Center Research Institute, San Francisco, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Long-range linkage A conformational mechanism whereby phosphorylation and Ca^{2+} -binding domains are functionally coupled through a large intramolecular distance, within the adenosine triphosphatase (ATPase) molecule.

Partial reactions The sequential steps of the sarcoendoplasmic reticulum Ca^{2+} ATPase (SERCA) catalytic and transport cycles, including Ca^{2+} and ATP binding to the enzyme, formation of phosphoenzyme intermediate,

translocation of bound Ca^{2+} against a concentration gradient, and hydrolytic cleavage of phosphate.

Sarcoendoplasmic reticulum Ca^{2+} ATPase (SERCA) The membrane-bound ATPase involved in Ca^{2+} transport and refilling of intracellular Ca^{2+} stores.

Trigger points The phosphorylation domain and Ca^{2+} -binding site II where phosphorylation and occupancy by a Ca^{2+} trigger the long-range linkage for interconversion of phosphorylation and Ca^{2+} -binding potentials.

ATP Use for Ca^{2+} Transport

The sarcoendoplasmic reticulum calcium adenosine triphosphatase (ATPase) (SERCA) is the Ca^{2+} pump of intracellular membranes. The pump plays a prominent role in numerous cytosolic-signaling mechanisms that require the sequestration of intracellular Ca^{2+} . The SERCA pump was first noted in a microsomal fraction of rabbit skeletal muscle homogenates, which was found to prevent activation of glycerinated muscle fibers or isolated contractile models upon the addition of adenosine triphosphate (ATP). The microsomes were then identified with vesicular fragments of sarcoplasmic reticulum (SR), and the relaxing effect was attributed to the sequestration of Ca^{2+} from the medium through active transport by a membrane-bound ATPase. Complementary DNA (cDNA) cloning identified three gene products (SERCA1, SERCA2, and SERCA3) with two to three splice variants for each primary transcript. The related isoforms display specific tissue distribution, functional roles, and regulatory mechanisms.

The SR vesicles obtained from skeletal muscle contain a high quantity of Ca^{2+} ATPase (SERCA1A isoform), which accounts for approximately 50% of the total protein and which is densely spaced within the plane of the membrane (Figure 1). These vesicles provide a very useful experimental system inasmuch as they allow parallel measurements of ATPase activity and Ca^{2+} transport, as well as control of the ionic environment in compartments delimited by the native membrane. For this reason, they have been used extensively in biochemical and structural studies to clarify the mechanism of ATP use for Ca^{2+} transport.

Steady-State Behavior

Active transport of Ca^{2+} into the lumen of SR vesicles occurs with a stoichiometric ratio of two Ca^{2+} per hydrolyzed ATP under optimal steady-state conditions. Acetyl phosphate or *p*-nitrophenyl phosphate can be used as a substrate by the enzyme instead of ATP, although with lesser kinetic competence. The maximal levels of ATP-dependent Ca^{2+} uptake by

the vesicles is increased by Ca^{2+} binding to a native acidic protein, calsequestrin, in the lumen of the vesicles. It can also be increased by complexation with anions such as oxalate or phosphate in the lumen of the vesicles.

Independent of the substrate, the ATPase has a strict requirement for Ca^{2+} activation. In fact the enzyme is activated by a medium containing free Ca^{2+} within the micromolar range and is inhibited by lumenal-free Ca^{2+} in the millimolar range. This indicates that ATP is used to change the affinity and orientation of the Ca^{2+} -binding sites, resulting in a three-orders-of-magnitude Ca^{2+} gradient across the membrane. Furthermore, in reconstituted ATPase preparations that are deprived of pathways for passive leakage, it was shown that Ca^{2+} transport is accompanied by H^{+} countertransport at a 1:1 ratio and that the pump is electrogenic.

Considering these functional features, the free energy required for active transport of Ca^{2+} into the vesicles can be defined, at first approximation, as:

$$\Delta G = RT \ln ([\text{Ca}_{\text{in}}^{2+}]/[\text{Ca}_{\text{out}}^{2+}]) + zF\Delta V$$

where $[\text{Ca}_{\text{in}}^{2+}]$ and $[\text{Ca}_{\text{out}}^{2+}]$ refer to the Ca^{2+} concentration in the lumen of the vesicles and in the outer medium, z is the electrical charge of the transported species, R and F are the gas and Faraday constants, respectively, and ΔV is the transmembrane electrical gradient. The free-energy requirement, estimated in this manner, provides a satisfactory match of the experimentally observed gradients with the chemical potential of ATP.

The Partial Reactions of the Catalytic and Transport Cycles

The Ca^{2+} ATPase catalytic cycle includes a phosphorylated intermediate formed by covalent transfer of the ATP terminal phosphate to an aspartyl residue (Asp351) at the catalytic site. Phosphoryl transfer requires activation by the same Ca^{2+} concentration as ATP hydrolysis, indicating that enzyme activation by Ca^{2+} must occur early in the catalytic cycle. In fact, cooperative binding of two Ca^{2+} per ATPase in the absence of ATP, with

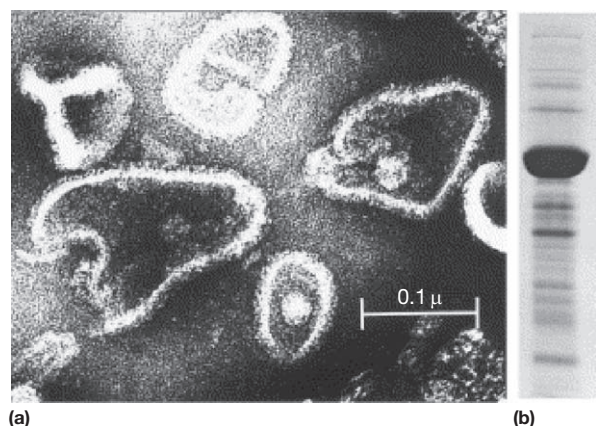


Figure 1 Structural characterization of SR vesicles. (a) Negatively stained SR vesicles, obtained from rabbit skeletal muscle, are visualized by electron microscopy. The densely spaced granules on the surface of the membrane correspond to Ca^{2+} ATPase molecules. (b) Electrophoretic analysis demonstrating that the Ca^{2+} ATPase separates as a prominent band that accounts for the major portion of the membrane protein. Adapted from Scales D and Inesi G (1976) Assembly of ATPase protein in sarcoplasmic reticulum membranes. *Biophysical Journal* 16: 735–751.

an apparent K_d in the micromolar Ca^{2+} range, was demonstrated by direct equilibrium measurements (Figure 2(a)).

The sequence of partial reactions making up the catalytic cycle was unveiled by transient-state experiments showing that addition of ATP to SR vesicles preincubated with micromolar Ca^{2+} is followed by the rapid formation of phosphorylated intermediate and vectorial displacement of two bound Ca^{2+} per ATPase. These initial events are then followed by hydrolytic cleavage of inorganic phosphate (Pi) after a time lag for protein conformational changes involved in the energy transduction mechanism (Figure 2(b)). On the other hand, the ATPase cycle is highly reversible, as demonstrated by ATP synthesis coupled to the Ca^{2+} efflux from loaded vesicles. The partial reactions of the reverse cycle were further clarified by the finding that, in the absence of bound Ca^{2+} , the enzyme can be phosphorylated with Pi to yield adenosine diphosphate (ADP)-insensitive phosphoenzyme (E-P). This phosphoenzyme can then be made ADP-sensitive by the addition of millimolar Ca^{2+} , whereupon it reacts with ADP to form ATP.

These functional findings have been attributed to a mechanism based on the interconversion of two enzyme states. One state (E_1), stabilized by Ca^{2+} binding, has the Ca^{2+} sites in high affinity and cytosolic orientation. The alternate state (E_2), stabilized by enzyme phosphorylation, has the Ca^{2+} sites in low affinity and luminal orientation. A more explicit view of the coupling mechanistic can be obtained by considering a minimal number of partial reactions and their equilibrium constants, as follows:

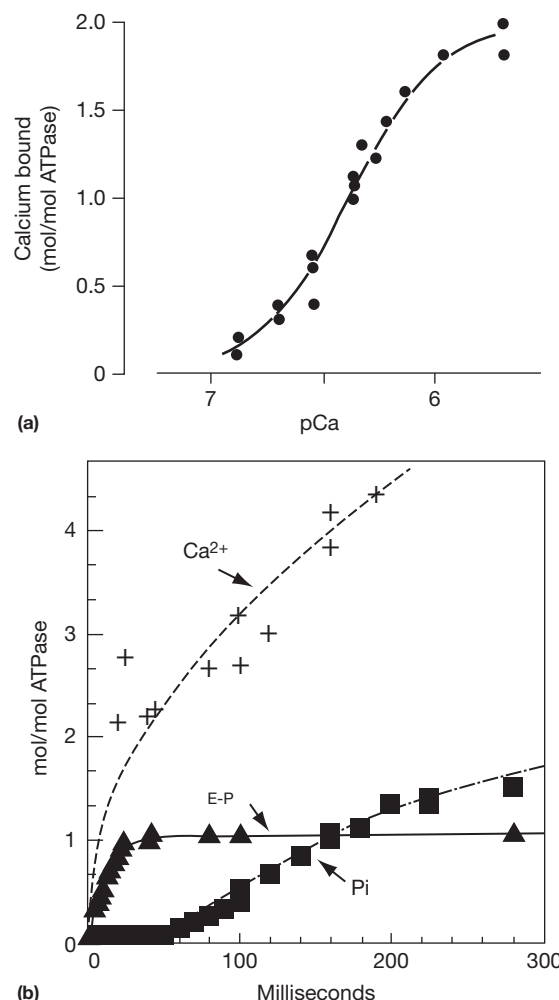
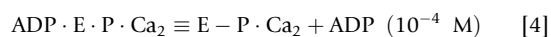
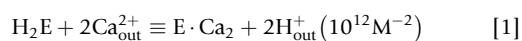
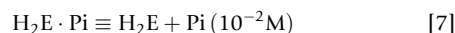
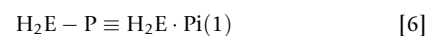
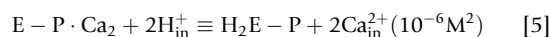


Figure 2 Functional characterization of SR vesicles. (a) Calcium binding at equilibrium in the absence of ATP, demonstrating cooperative binding of two Ca^{2+} per ATPase. (b) The catalytic cycle upon addition of ATP begins with enzyme phosphorylation and inward displacement of the two bound Ca^{2+} . Hydrolytic cleavage of inorganic phosphate (Pi) is the final step of the cycle. E-P, phosphoenzyme. (a) From Inesi G, Kurzmack M, Coan C, and Lewis D (1980) Cooperative calcium binding and ATPase activation in sarcoplasmic reticulum vesicles. *Journal of Biological Chemistry* 255: 3025–3031. (b) From Inesi G, Watanabe T, Coan C, and Murphy A (1978) The mechanism of sarcoplasmic reticulum ATPase. *Annals of the New York Academy of Sciences* 402: 515–534.



where the equilibrium constants (listed after each reaction) relate to conditions allowing constant temperature (25 °C) and pH (7.0).

In the reaction sequence given, the initial high-affinity binding (reaction 1) of two Ca^{2+} activates the enzyme, permitting the use of ATP and the formation of a phosphorylated intermediate (reactions 2–4). In turn, enzyme phosphorylation destabilizes and changes the vectorial orientation of the bound Ca^{2+} , thereby increasing the probability of its

dissociation into the lumen of the vesicles (reaction 5). Note that the equilibrium constant for the enzyme phosphorylation by ATP is nearly 1, indicating that the free energy of ATP is conserved by the enzyme and used to change the Ca^{2+} -binding characteristics. Finally, the phosphoenzyme undergoes hydrolytic cleavage and releases P_i (reactions 6 and 7) before entering another cycle. The reaction sequence shows clearly that the direct mechanistic device for translocation of bound Ca^{2+} is enzyme phosphorylation rather than hydrolytic cleavage of P_i .

The Coupling Mechanism

Free-Energy Use

Because the catalytic and transport cycles are likely to include conformational transitions, it should be noted that such transitions are coupled intrinsically with the chemical reactions subjected to experimental measurement and that their influence is reflected by the equilibrium constants just listed. In fact, the standard free energies ($-RT \ln K$) of the partial reactions add up to the standard free energy of ATP hydrolysis (γ -phosphate), as expected. Most interestingly, the standard free-energy diagram for the partial reactions reveals that the chemical potential of ATP does not manifest itself in the phosphoryl transfer or hydrolytic cleavage reactions (K_4 and K_6 near 1) but rather in the drastic reduction of the enzyme affinity for Ca^{2+} (compare K_1 to $1/K_5$). We can, then, write that, under standard conditions:

$$\Delta G = nRT \ln (K_a^{\text{CaE-P}} / K_a^{\text{CaE}}) \quad [8]$$

per n calcium ions (2 in our case) transported per cycle, and where the equilibrium constant (listed after the reaction) relates to conditions allowing constant temperature (25 °C and pH 7.0). $K_a^{\text{CaE-P}} / K_a^{\text{CaE}}$ are the association constants of the enzyme for Ca^{2+} in the ground state and following activation by ATP. With reference to the reaction scheme given previously, the two relevant constants are K_1 and $1/K_5$. Note that, as written in the scheme and in the forward direction of the cycle, the equilibrium constant for reaction 1 is K_a , while it is K_d for reaction 5.

It is then apparent that the basic coupling mechanism of catalysis and transport consists of the mutual destabilization of Ca^{2+} and phosphorylation sites. Analysis of the amino acid sequence, electron microscopy of ordered ATPase arrays, and diffraction patterns obtained from three-dimensional (3D) crystals have shown that the SERCA enzyme is a 100-kDa protein, folded into a membrane-bound region, an extramembranous (cytosolic) region, and a short stalk between them. The membrane-bound region includes 10 helical segments (M1, ..., M10) and the calcium-binding domain. On the other hand, the extramembranous region or headpiece, comprises the nucleotide-binding (N) domain and a phosphorylation (P) domain, in addition to a smaller actuator (A) domain. It is clear that the Ca^{2+} -binding domain and the catalytic domain are separated by a large distance (~ 50 Å). Therefore, the interconversion of cation-binding and phosphorylation potentials requires a long-range intramolecular linkage, rendered possible by protein conformational changes. It is, then, useful to relate the binding and covalent reactions to

conformational changes that may occur through the catalytic and transport cycles.

Ca^{2+} Binding and Catalytic Activation

Binding of two calcium ions per ATPase molecule is an absolute requirement for enzyme activation. The cooperative character of binding is consistent with a sequential binding mechanism to two interdependent sites (I and II). Spectroscopic studies provided early suggestions of Ca^{2+} -induced conformational effects, accounting for binding cooperativity and enzyme activation. A detailed characterization of the large conformational changes produced by Ca^{2+} binding was then obtained by high-resolution diffraction studies.

The functional role of the Ca^{2+} -induced conformation change lies in its absolute requirement for enzyme activation, rendered possible by the long-range intramolecular linkage. The requirement for Ca^{2+} involves both ATP use for phosphoenzyme formation in the forward direction of the cycle and the formation of ATP upon addition of ADP to phosphoenzyme formed with P_i in the reverse direction of the cycle. Activation is not obtained by Ca^{2+} occupancy of the first binding site only but requires the occupancy of the second site, which may then be considered a Ca^{2+} trigger point for enzyme activation. Ca^{2+} -binding affects directly the M4, M5, and M6 transmembrane helices and is then transmitted to the extramembranous region, resulting in the large displacement of the headpiece domains and catalytic activation. Single mutations of several residues in the segments connecting the Ca^{2+} -binding region with the phosphorylation domain, such as M4, M5, and the M6–M7 loop, interfere with the phosphorylation reactions.

Nucleotide Binding and Substrate-Induced Conformational Fit

Nucleotide protection of the ATPase from digestion with proteinase and the mutational analysis of this phenomenon suggest that ATP binding produces a conformational change that includes repositioning of the A domain, concomitant with approximation of the N and P domains to match the molecular geometry of ATP. This substrate effect appears similar to the approximation of nucleotide binding (fingers) and catalytic (palm) domains that occurs in DNA polymerases. Although ATP is likely to form an initial complex with Mg^{2+} due to the cation-binding property of the nucleotide, ATP and Mg^{2+} reach the catalytic site through a random mechanism. Accordingly, the initial nucleotide binding at the substrate site does not include Mg^{2+} , even though the subsequent phosphoryl transfer and hydrolytic reactions require Mg^{2+} . The stabilization provided by Mg^{2+} to the transition state, relative to that of the enzyme–substrate complex, is of definite kinetic advantage.

Phosphoryl Transfer and Hydrolytic Cleavage

The Ca^{2+} ATPase mechanism includes a covalent intermediate, general acid–base catalysis and metal ion assistance. The covalent intermediate is formed through an attack on the electrophilic ATP terminal phosphate by a nucleophilic carboxylate group (D531) within the catalytic site. The phosphoenzyme intermediate reacts, then, with water to yield P_i .

It is useful to consider the Ca^{2+} ATPase catalytic mechanism in the light of the phosphorylserine phosphatase (PSPase), due to a remarkable structural and functional analogy and the high definition available for the PSPase catalytic intermediates. Based on this analogy, and on ATPase mutational analysis, it appears that the terminal phosphate of ATP comes in close proximity to Asp351, stabilized by neighboring residues such as K352, T353, G626, K684, and N706. Mg^{2+} is then coordinated by D703, T353, D351, and phosphate oxygen, participating in the transition state and in all subsequent steps. Furthermore, T353 is likely to serve as a general acid for proton donation to the leaving ADP. The same residue then acts as a general base, extracting a proton from nucleophilic water for hydrolytic cleavage of the phosphoenzyme.

Interconversion of Phosphorylation and Ca^{2+} -Binding Potentials

The specific step related to interconversion of phosphorylation and Ca^{2+} -binding potentials is often referred to as the $\text{E}_1\text{P}\cdot\text{Ca}_2 \rightleftharpoons \text{E}_2\text{P}\cdot\text{Ca}_2$, to indicate the transition of an intermediate of high phosphorylation potential and high Ca^{2+} affinity to an intermediate of low phosphorylation potential and low Ca^{2+} affinity. The functional and structural information described so far indicates that small conformational changes occurring in concomitance with ATP use at the catalytic site (phosphorylation trigger point) are amplified and transmitted to the Ca^{2+} -binding domain through a long-range relay mechanism. The inversion of phosphoryl oxygen atoms during the transition state of the phosphorylation reaction and the displacement of residues interacting with these oxygen atoms may be the triggering perturbation. This is then followed by large motions of the headpiece domains, which are transmitted to the Ca^{2+} -binding domain in the membrane-bound region. Accordingly, the ability of headpiece domains to chemically cross-link changes drastically as they are displaced by intermediate reactions of the catalytic cycle.

In conclusion, the transduction mechanism relies on a long-range linkage of transmembrane helices and cytosolic domains, whereby large-scale and mutually exclusive motions are triggered by Ca^{2+} binding at one end and ATP use at the other. This is rendered possible by the ability of the protein to use free energy derived from binding and phosphorylation reactions. The cycle proceeds forward to Ca^{2+} release or in reverse to ATP synthesis, depending on the concentrations of ligands, substrate, and products.

Physiological Implication and Inhibitory Drugs

The SERCA2 isoform of cardiac muscle is subject to physiological regulation by phospholamban (PLB), a 22-kDa pentameric protein undergoing reversible association and dissociation. The PLB monomer interacts with the ATPase at the level of both the membrane-bound and extramembraneous regions of the enzyme. In the presence of PLB, the ATPase activation curve is displaced to a higher Ca^{2+} concentration range. The velocity at nonsaturating Ca^{2+} concentrations is thereby

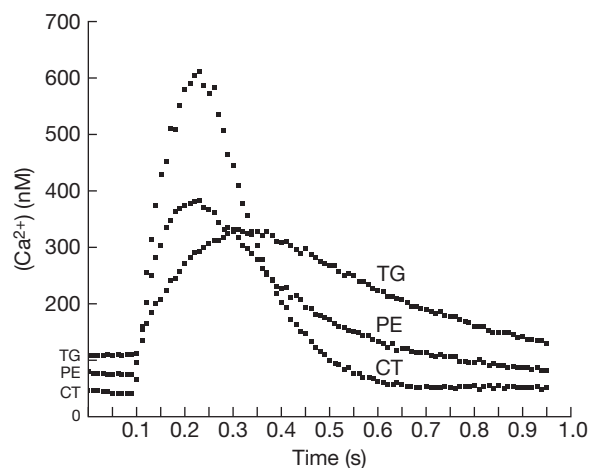


Figure 3 Ca^{2+} signaling in cardiac muscle cells following electrical stimuli. Size and kinetics of the normal (CT) myocytes signal are strongly altered in adrenergic hypertrophy (PE) due to SERCA protein downregulation, and following exposure to thapsigargin (TG) due to SERCA inhibition. From Prasad AM and Inesi G (2009) Effects of thapsigargin and phenylephrine on calcineurin and protein kinase C signaling functions in cardiac myocytes. *American Journal of Physiology* 296: C992–C1002.

reduced, although the maximal velocity at saturating Ca^{2+} concentrations remains unchanged. This inhibition is reversed when PLB undergoes phosphorylation catalyzed by cAMP and calmodulin-dependent kinases. The phosphorylation-dependent relief of inhibition by PLB is related to the physiological mechanism whereby adrenergic stimuli increase Ca^{2+} storing by SR and improve the contractile performance of the heart muscle *in vivo*.

Experimental and clinical studies have uncovered important aspects of SERCA involvement in cardiac physiology and pathology. In fact, the SERCA protein level is reduced in cardiac hypertrophy, relative to the cardiac muscle augmentation. The consequent impairment of Ca^{2+} signaling then contributes to development of heart failure (Figure 3).

Thapsigargin (TG), a plant-derived sesquiterpene lactone, is a very potent and specific SERCA inhibitor. Although TG is not likely to be a physiologic inhibitor, it is nevertheless an extremely useful experimental tool. TG interferes with all partial reactions of the catalytic cycle and induces stabilization of ordered ATPase arrays, suggesting a general effect on the enzyme. Mutational and structural studies indicate that the TG-binding site resides in a cavity delimited by the M3, M5, and M7 helices near the cytosolic surface of the membrane. It is likely that the presence of TG in this cavity confers structural stabilization to the enzyme, thereby preventing conformational responses to Ca^{2+} binding and phosphorylation reaction, as required for the progression of the catalytic cycle. Specific delivery of TG to neoplastic tissue causes apoptosis and may be useful in cancer treatment. Cyclopiazonic acid is another SERCA inhibitor that may be useful for its specificity and reversibility.

See also: **Bioenergetics:** Calcium-Binding Proteins: Cytosolic (Annexins, Gelsolins, and C₂-Domain Proteins); Calcium/Calmodulin-Dependent Protein Kinase II; ER/SR Calcium Pump: Structure; **Signaling:** Calcium/Calmodulin-Dependent Protein Kinases.

Further Reading

- Andersen JP (1995) Dissection of the functional domains of the sarcoplasmic reticulum Ca²⁺ ATPase by site-directed mutagenesis. *Bioscience Reports* 15: 243–261.
- East JM (2000) Sarco(endo)plasmic reticulum Ca²⁺ pumps: Our advances in understanding of structure/function and biology. *Molecular Membrane Biology* 17: 189–200.
- Green NM and Stokes DL (2003) Structure and function of the calcium pump. *Annual Review of Biophysics and Biomolecular Structure* 32: 445–468.
- Inesi G, Kurzmack M, Coan C, and Lewis D (1980) *Journal of Biological Chemistry* 255: 3025–3031.
- Inesi G, Watanabe T, Coan C, and Murphy A (1978) *Annals of the New York Academy of Sciences* 402: 515–534.
- MacLennan DH, Rice WJ, and Green NM (1997) The mechanism of Ca²⁺ transport by sarco(endo)plasmic reticulum ATPase. *Journal of Biological Chemistry* 272: 28815–28818.
- Moller JV, Juul B, and le Maire M (1996) Structural organization, ion transport and energy transduction of P-type ATPases. *Biochimica et Biophysica Acta* 1286: 1–51.
- Prasad AM and Inesi GAJ (2009) Effects of thapsigargin and phenylephrine on calcineurin and protein kinase C signalling functions in cardiac myocytes. *American Journal of Physiology* 296: C992–C1002.
- Scales D and Inesi G (1976) Assembly of ATPase protein in sarcoplasmic reticulum membranes. *Biophysical Journal* 16: 735–751.
- Toyoshima C, Nomura H, and Sagita Y (2003) Structural basis for ion pumping by Ca²⁺-ATPase of sarcoplasmic reticulum. *FEBS Letters* 555: 106–110.

ER/SR Calcium Pump: Structure

C Toyoshima, The University of Tokyo, Tokyo, Japan

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Chikashi Toyoshima and Yuji Sugita, volume 2, pp. 61–65, © 2004, Elsevier Inc.

Glossary

EF-hand motif The most common calcium-binding motif that has a helix–loop–helix structure.

HAD superfamily A large family found by Aravind et al. in 1998, which includes P-type ion-transporting adenosine triphosphatases (ATPases), haloacid

dehalogenase (HAD), and phosphoserine phosphatase, among others.

Rossmann fold A supersecondary structure, consisting of a pair of $\beta\alpha\beta$ structures arranged so that at least four β -strands form a single parallel β -sheet flanked by two layers of α -helices.

Architecture of Ca^{2+} -Adenosine Triphosphatase

Sarco(endo)plasmic reticulum calcium adenosine triphosphatase 1a (SERCA1a) consists of a single polypeptide of 994 amino acid residues that constitute three cytoplasmic domains (A, actuator; N, nucleotide binding; and P, phosphorylation), 10 transmembrane helices (M1–M10), and small luminal loops (Figure 1). The A-domain, which includes the N-terminal ~50 residues that form two short α -helices and ~110 residues between the M2 and M3 helices, is the smallest of the three cytoplasmic domains. It works as the actuator of the transmembrane gating mechanism that regulates Ca^{2+} binding and release. The A-domain is connected to the M1 and M3 helices with rather long flexible linkers, the lengths of which are of critical importance. It contains one of the signature sequences $^{181}\text{TGES}$, which plays an important role in dephosphorylation. The P-domain takes a Rossmann fold that consists of a central parallel β -sheets and associated α -helices. It contains the phosphorylated residue Asp351, a Mg^{2+} coordinating residue Asp703, and other critical residues that characterize the P-type ATPase as a member of the haloacid dehalogenase superfamily. The N-domain, a long insertion (~240 residues) between the two parts forming the P-domain, is the least conserved of the three cytoplasmic domains, but contains the residues for adenosine binding (e.g., Phe487) and those critical for bridging adenosine triphosphate (ATP) and the P-domain (e.g., Arg560). The hinge between the N- and P-domains is double stranded and allows a change of 60° in inclination.

Of the 10 transmembrane helices, four of them (M2–M5) have long cytoplasmic extensions (Figure 1). In particular, M5 is 60-Å long and extends from the luminal surface to the end of the P-domain, working as the spine of the molecule. Two (M4 and M6) helices have proline residues at the middle and are partly unwound throughout the reaction cycle. M6 and M7 are far apart and connected by a cytosolic loop (L6/7) of ~20 residues. L6/7 runs along the bottom of the P-domain and is hydrogen bonded to M5 (Figure 1(a)). This loop appears to be very rigid and presumably restricts the movement of the M5 helix. M4–M6 and M8 contain the residues directly coordinating two Ca^{2+} (Figure 2). All helices from M1 to M6 move considerably during the reaction cycle,

whereas M7–M10 helices do not, apparently acting as a membrane anchor. In SERCA1, the middle part of M7 containing a GxxG motif makes a tight couple with M5 and allows the bending of M5. A short helical segment between M9 and M10 appears to serve as a socket for the luminal end of the M2 helix. M7–M10 have a specialized function in each subfamily of P-type ATPases, and are absent in heavy metal pumps. In the Na^+K^+ -ATPase, M9 contributes to the association with the regulatory FXYD protein, whereas M10 to the association of the β subunit.

It is important to note that the cytoplasmic end of M5 is integrated into the P-domain near the phosphorylation site as a part of the Rossmann fold and clamped with the M4 helix through a short β -strand (β_0 and β_7 ; Figure 3). The P-domain is also connected to the M3 helix with critical hydrogen bonds. Thus, the events that occur at the phosphorylation site or Ca^{2+} -binding sites can be mechanically transmitted to the other site.

Transmembrane Ca^{2+} -Binding Sites

SERCA1 has two high-affinity transmembrane Ca^{2+} -binding sites (I and II), which are located side by side near the cytoplasmic surface of the lipid bilayer (Figures 1 and 2). The binding of two Ca^{2+} is, however, sequential and cooperative. Site I, the binding site for the first Ca^{2+} , is located in a space surrounded by M5, M6, and M8 helices, and formed by entirely side-chain oxygen atoms and two water molecules (Figure 2). Asn768, Glu771 (M5), Thr799, Asp800 (M6), and Glu908 (M8) contribute to this site. Site II is nearly 'on' the M4 helix and located ~3 Å closer to the cytoplasmic surface than site I (Figures 1 and 2). The M4 helix is partly unwound (between Ile307 and Gly310) and provides three main-chain carbonyls for coordination (Figure 2). Asn796 and Asp800 (M6) provide one side-chain oxygen, whereas Glu309, the gating residue, provides two to cap the bound Ca^{2+} . No water molecule contributes to site II. This arrangement of oxygen atoms is reminiscent of the classical EF-hand motif. Thus, both sites have seven coordination but of different characteristics. Asp800, on the unwound part of M6, is the only residue that contributes to both sites (Figure 2).

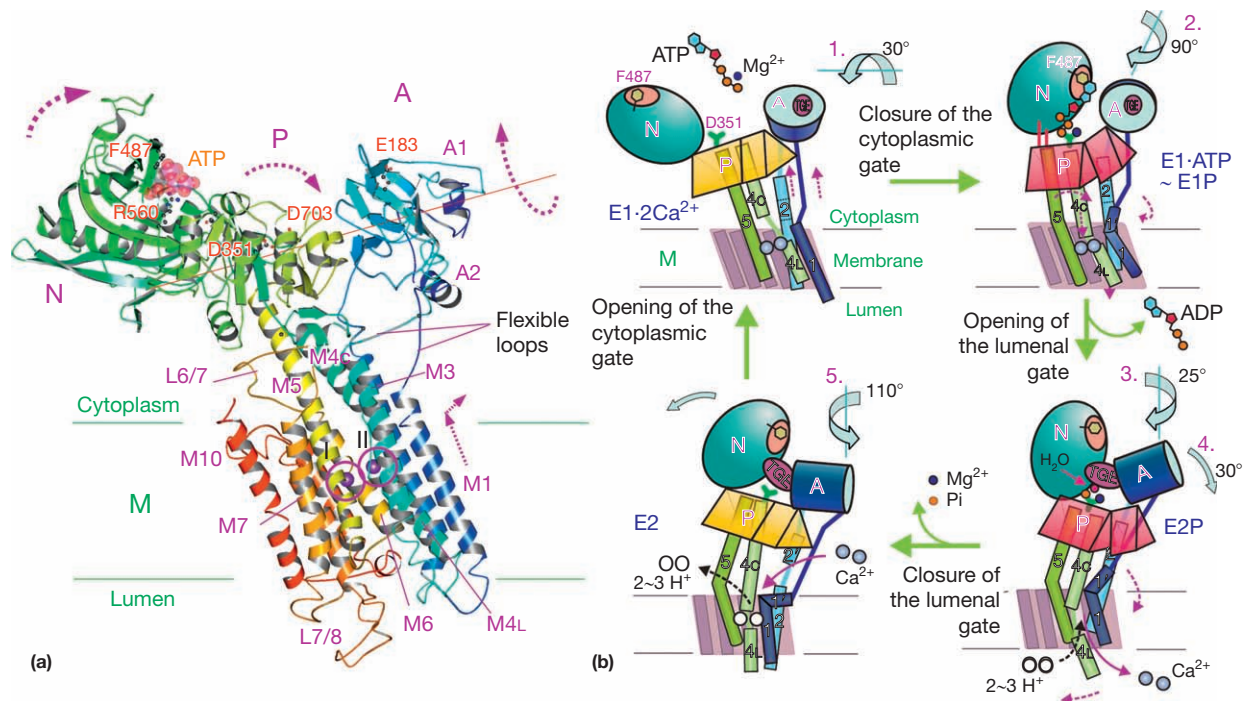


Figure 1 Architecture of Ca²⁺-ATPase and its ion-pumping mechanism. (a) A ribbon representation of Ca²⁺-ATPase in the E1·2Ca²⁺ state, viewed parallel to the membrane plane. Colors change gradually from the amino terminus (blue) to the carboxy terminus (red). Two purple spheres (numbered and circled) represent bound Ca²⁺. Three cytoplasmic domains (A, N, and P), the α -helices in the A-domain (A1–A2), and those in the transmembrane domain (M1–M10) are indicated. M1' is an amphipathic part of the M1 helix lying on the bilayer surface. Docked ATP is shown in transparent space fill. Several key residues – E183 (A), F487 and R560 (N, ATP binding), D351 (phosphorylation site), and D703 (Mg²⁺ binding, P) – are shown in ball and stick. Axis of rotation (or tilt) of the A-domain is indicated with thin orange line. PDB accession code is 1SU4 (E1·2Ca²⁺). (b) A cartoon illustrating the structural changes of Ca²⁺-ATPase during the reaction cycle, based on the crystal structures in nine different states.

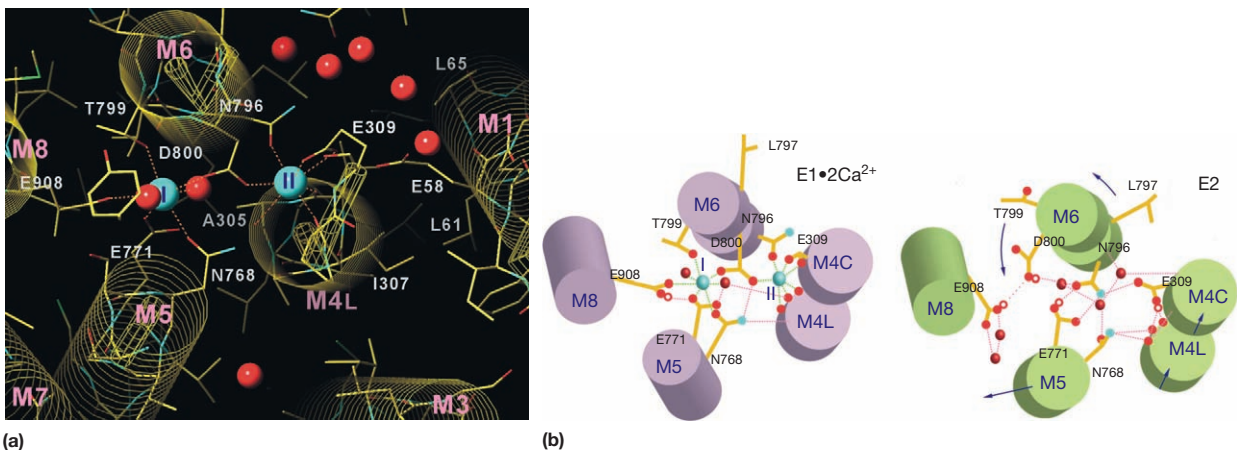


Figure 2 Structure of the transmembrane Ca²⁺-binding sites. (a) Atomic model in the E1·2Ca²⁺ state. (b) Cartoons illustrating the binding sites in E1·2Ca²⁺ and E2. Viewed approximately normal to the membrane. Ca²⁺ appears as cyan spheres and water molecules as red spheres. Note that site II Ca²⁺ is exchangeable by conformation change of the E309 side chain in E1·2Ca²⁺. The arrows in (b) indicate the movements of the helices in the transition from E2 → E1·2Ca²⁺. Red circles represent protonation of the oxygen atoms. Dotted lines in (b) indicate hydrogen bonds (pink) and coordination of Ca²⁺ (light green).

ATP-Binding Site

ATP binds near the hinge between the N- and P-domains so that it is able to cross-link the two domains. The adenine ring of ATP is inserted in a hydrophobic pocket in the N-domain,

sandwiched by Phe487 and Leu562 (Figure 3). A hydroxyl of the ribose is stabilized by a hydrogen bond with Arg489, and the β -phosphate with Arg560. In the absence of Ca²⁺, the triphosphate chain of ATP takes a folded conformation with Mg²⁺, and interactions with the ATPase are limited to residues

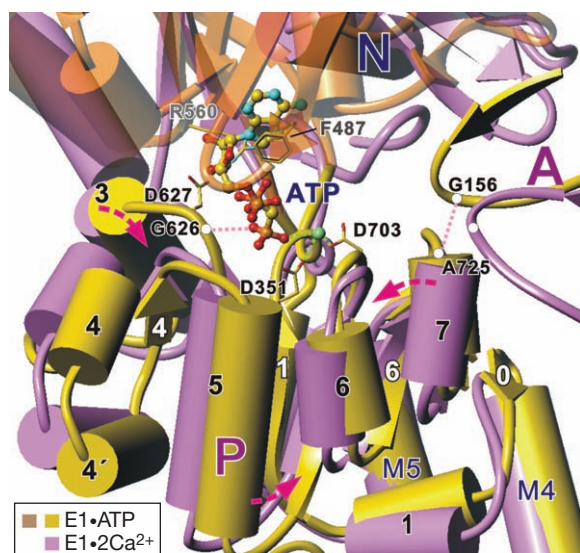


Figure 3 Structural changes caused by the binding of ATP and Mg^{2+} to the P-domain. The bound metal (small green sphere) is most likely Ca^{2+} rather than Mg^{2+} in the crystal structure. The N-domain in E1-ATP appears transparent. Note that the P-domain is bent (arrows in red broken lines) by coordination of the metal by the γ -phosphate, D351 and D703 (shown in stick model) and that the A-domain tilts because of the inclination of the P7 helix.

in the N-domain. When phosphoryl transfer occurs, the Mg^{2+} bound to ATP is released and the triphosphate chain takes an extended zigzag conformation stabilized further by hydrogen bonds with residues in the P-domain (Figure 3).

Synopsis of Ion Pumping

E2 $\rightarrow$ E1 $\rightarrow$ E1·2 Ca^{2+} : Binding of Ca^{2+}

The state after release of Ca^{2+} into the lumen and after dephosphorylation, namely, the E2 state, is considered to be the ground state of the ATPase (Figure 1(b)). The cytoplasmic headpiece is compact with the N- and A-domains tightly associated with several hydrogen bonds between them, primarily to prevent delivery of the γ -phosphate of ATP to Asp351. The P-domain is free from distortion by bound phosphate. The transmembrane Ca^{2+} -binding cavity is filled with water molecules and all the carboxyl groups there are predicted to be protonated to compensate for the empty space and charge imbalance created by the release of Ca^{2+} (Figure 2(b)). However, this structure can persist only at low pH, apparently because the charge compensation by H^+ is insufficient at pH 7. Although the A- and N-domains are most closely associated in this state, thermal agitation opens the headpiece and releases bound protons into the cytoplasm. Released protons are dissipated away through the SR membrane, placing most of the ATPase molecules, at pH 7, into the E1 state. In E1, the carboxyl groups in the Ca^{2+} -binding cavity are not protonated and, therefore, have high affinity to Ca^{2+} .

ATP can bind to the ATPase in the E2 state with a μM K_d , but delivery of the γ -phosphate to Asp351 is hindered by the closed configuration of the cytoplasmic domains. Yet the

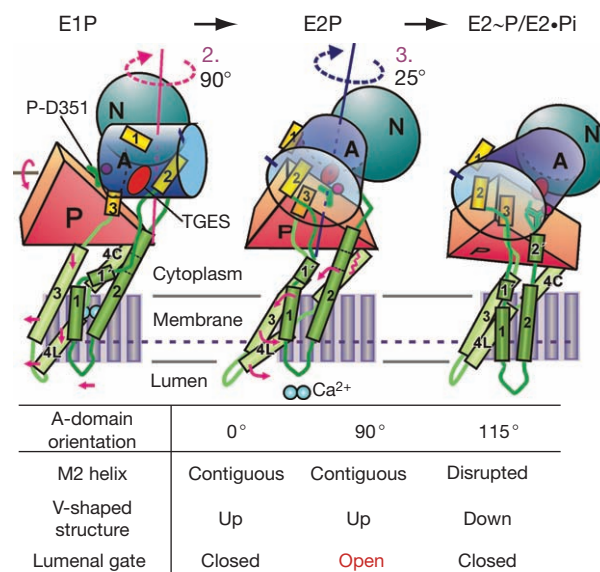


Figure 4 A cartoon illustrating two-step rotation in the processing of aspartylphosphate and luminal gating of the ion pathway. Small arrows indicate the movements of the transmembrane helices. M1–M4 (green) and A1–A3 (yellow) helices are numbered. P-D351 refers to phosphorylated Asp351.

binding of ATP in this state facilitates the transition into E1 presumably by opening the headpiece.

The first Ca^{2+} will enter the binding cavity through improperly formed site II (otherwise it will be trapped there) and binds to site I. This will rotate M6 and position Asp800 properly. The M5 helix will straighten (Figure 1(b)) and alter the arrangement of oxygen atoms in site II to form a higher-affinity-binding site. Finally, the carboxyl of Glu309 side chain will cap site II Ca^{2+} , and the binding signal is transmitted to the phosphorylation site some 50 Å away. Because there is enough space around Glu309, the cytoplasmic gate remains unlocked, and site II Ca^{2+} is exchangeable with those in the cytoplasm (Figure 2(a)). The M1 helix is deeply embedded in the lipid bilayer, stabilized indirectly by the bound Ca^{2+} .

Straightening of the M5 helix breaks the closed configuration of the three cytoplasmic domains to allow more inclination of the N-domain for delivery of the ATP γ -phosphate to Asp351, the phosphorylation residue.

E1·2 Ca^{2+} $\rightarrow$ E1P Transition: Formation of the Occluded State

ATP cross-links the P- and N-domains, so that the γ -phosphate of ATP and a Mg^{2+} (or Ca^{2+}) bind to the P-domain to bend it in two directions (Figure 3). The loop containing Asp703, an absolutely conserved residue in the P-domain, is pulled toward Asp351 by Mg^{2+} binding and, consequently, the P6 and P7 helices. This bending of the P-domain tilts the A-domain, which sits on the P7 helix by $\sim 30^\circ$ (1 in Figure 1(b)). This tilting places strain on the link between the A-domain and the M3 helix. Because the M3 helix is connected to the A-domain at the opposite end of the A-domain to the P7 helix, the movement of the P7 helix is highly amplified. This strain appears to be the driving force for the A-domain rotation in the next step (Figures 1(b) and 4). At the same time, the

M1 helix is pulled up so that the top of the transmembrane part (in particular Leu65; **Figure 2(a)**) fixes the side-chain conformation of Glu309 by occupying the space around it. Thus, the cytoplasmic gate is locked and two Ca^{2+} are occluded in the transmembrane-binding sites. M1 and M2 now form a V-shaped structure, which moves as a rigid body until the end of the reaction cycle, and transmits the movements of the A-domain to other transmembrane helices, primarily the luminal half of M4 (M4L).

Phosphoryl transfer from the γ -phosphate to Asp351 fixes the N-domain in a highly inclined position so that a mechanical couple is formed with the A-domain in preparation for a 90° rotation of the A-domain in the E1P–E2P transition. The interface between the N- and A-domains is different from that in E2, as the phosphate and Mg^{2+} bends the P-domain so that the N-domain inclines in a different orientation (**Figure 3**).

E1P → E2P Transition: Release of Ca^{2+} into the Lumen of SR

Phosphoryl transfer to Asp351 allows the dissociation of adenosine diphosphate (ADP), which triggers the opening of the N- and P-domain interface. The A-domain rotates 90° around an axis $\sim 25^\circ$ inclined from the membrane normal (2 in **Figures 1(b)** and 4) and brings the $^{181}\text{TGES}$ loop of the A-domain deep into the gap between the N- and P-domains above the aspartylphosphate (**Figure 5**). The rotation stops when the Thr181 carbonyl makes van der Waals contacts with the Gly626 C α in the conserved ^{625}TGD motif in the P-domain. This position of the TGES loop is stabilized by several hydrogen bonds, presumably to make the resident time in this state long enough for releasing the bound Ca^{2+} into the lumen of SR. The TGES loop occupies the space where ADP was in E1P–ADP to prevent the binding of ADP and to shield the aspartylphosphate from bulk water, as the Gly182 C α makes van der Waals contacts with the aspartylphosphate (**Figure 5**).

This A-domain rotation causes a 30° change in inclination of the P-domain toward M1, which, in turn, causes a drastic rearrangement of the transmembrane helices M1–M6, including a large downward movement of M4, sharp bending of M5 toward M1 (**Figure 1(b)**), and rotation of M6 (**Figure 2(b)**), which destroy the Ca^{2+} -binding sites. The V-shaped structure formed by the M1 and M2 helices pushes against M4L, opening the luminal gate and releasing the bound Ca^{2+} into the lumen (**Figure 4**). This will allow protons and water molecules to enter and stabilize the empty Ca^{2+} -binding sites (**Figure 2(b)**).

In the conversion of the A-domain rotation to the downward movement of the M4 helix, the wedge shape of the P-domain and the flexible link between the A-domain and the M1 helix (**Figure 1(a)**) play a critical role.

E2P → E2 Transition: Hydrolysis of Aspartylphosphate and Closing of the Luminal Gate

In the E2P → E2•Pi transition, the A-domain rotates further by 25° around an axis near Thr181, different from the previous one (3 in **Figures 1(b)** and 4). This rotation places further strain on the M2 helix, which, as a result, partly unwinds and releases the V-shaped structure formed by the M1 and M2 helices so that it takes a $\sim 5\text{-}\text{\AA}$ lower position toward the

luminal side. This movement imposes more upright position on M4L and closes the luminal gate.

This rotation of the A-domain is related to the introduction of one water molecule in the phosphorylation site. Glu183 in the TGES loop now fixes the water molecule and catalyzes its attack on the aspartylphosphate, presumably by withdrawing a hydrogen from the water molecule (**Figure 5**). Then, the inorganic phosphate and Mg^{2+} are released and the P-domain becomes relaxed. This, in turn, releases the M1 and M2 helices to lock the luminal gate and places the ATPase into the E2

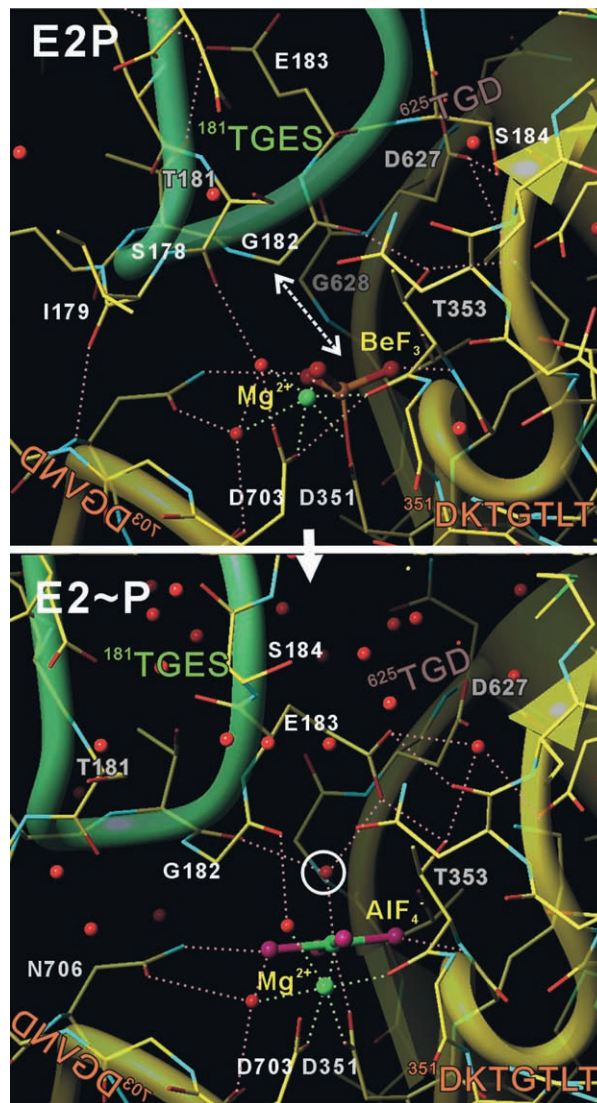


Figure 5 Phosphorylation site in the E2P ground and E2 ~P transition states. BeF_3^- and AlF_4^- are phosphate analogs of the ground state and transition state, respectively. Secondary structure motifs are superimposed on the atomic models. Conserved sequence motifs are labeled. Water molecules appear as small red spheres. Dotted lines represent hydrogen bonds and coordination of Mg^{2+} . Double-headed arrow in E2P shows that Gly182 C α and aspartylphosphate form van der Waals contacts. The white circle in E2 ~P identifies the water molecule attacking the aspartylphosphate.

state. The top amphipathic part of M1 (M1') forms a part of a cytoplasmic access funnel leading to Glu309, the side chain of which becomes stabilized by hydrogen bonding with Asn796 (Figure 2(b)). By release of protons to the cytoplasm, Glu309 is supposed to change the side-chain conformation and catch Ca^{2+} .

Summary

As described, Ca^{2+} -ATPase undergoes a series of very large movements of the cytoplasmic domains and rearrangements of the transmembrane helices. Such large movements are necessary to cope with thermal fluctuations and yet to utilize thermal energy effectively for ion pumping. The P- and N-domains change interfaces and thereby control the orientation of the A-domain, the actuator of transmembrane gates. ATP, phosphate, Mg^{2+} , and Ca^{2+} are the modifiers of the interfaces (Figure 1(b)). Here, ATP works as a cleavable cross-linker of the N- and P-domains. Free energy provided by the hydrolysis of ATP is used in a very implicit way. It seems that the overall reaction is made possible by suppressing backward reaction and ATP works as a loose coupler that connects thermal energy and electrochemical work.

Atomic coordinates of the aligned structures, movies on the structural changes during the reaction cycle, and some results of molecular dynamics simulations are available on the Internet.

See also: **Bioenergetics:** Calcium-Modulated Proteins (EF-Hand); ER/SR Calcium Pump: Function; Membrane Transport, General Concepts; Plasma-Membrane Calcium Pump: Structure and Function; P-Type Pumps: Copper Pump; P-Type Pumps: H^+/K^+ Pump; P-Type Pumps: Na^+/K^+ -ATPase; P-Type Pumps: Plasma-Membrane H^+ Pumps; Structure of P-Type Adenosine Triphosphatases.

Further Reading

- Aravind L, Galperin MY, and Kunin EV (1998) The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. *Trends in Biochemical Sciences* 23: 127–129.
- Møller JV, Juul B, and le Maire M (1996) Structural organization, ion transport, and energy transduction of P-type ATPases. *Biochimica et Biophysica Acta* 1286: 1–51.
- Toyoshima C (2008) Structural Aspects of ion pumping by Ca^{2+} -ATPase of sarcoplasmic reticulum. *Archives of Biochemistry and Biophysics* 476: 3–11.
- Toyoshima C (2009) How Ca^{2+} -ATPase pumps ions across the sarcoplasmic reticulum membrane. *Biochimica et Biophysica Acta* 1793: 941–946.
- Toyoshima C, Nakasako M, Nomura H, and Ogawa H (2000) Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* 405: 647–655.
- Toyoshima C and Nomura H (2002) Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* 418: 605–611.

Relevant Website

<http://www.iam.u-tokyo.ac.jp> – Institute of Molecular and Cellular Biosciences, University of Tokyo.

Eukaryotic Chemotaxis

C P McCann, University of Maryland College Park, College Park, MD, USA

C A Parent, Center for Cancer Research, Bethesda, MD, USA

Published by Elsevier Inc.

Glossary

Actin Protein active in muscular contraction, cellular movement, and maintenance of cell shape. Actin can be in either globular (G-actin) or filamentous (F-actin) form.

Cytoskeleton Three-dimensional network of scaffolding proteins that provides internal support for cells, anchors internal cell structures, and functions in cell movement and division.

Dictyostelium discoideum Soil-living migratory amoeba that transitions from a group of unicellular amoebae into

a multicellular organism composed of spores atop a stalk of vacuolated cells.

Myosin Motor protein that combines with actin to form actomyosin, a contractile complex.

Neutrophil The most abundant white blood cell and key player in innate immunity.

Polarity The asymmetric organization of cellular components and structures.

Pseudopod Temporary cellular protrusion process that functions during migration.

Introduction

Directed cell migration in response to external chemical gradients, or chemotaxis, is of fundamental importance in many physiological and pathological processes. For example, in mammals, chemotaxis plays a key role in embryogenesis, tissue maintenance, and the immune response, and has also been implicated in metastasis, arthritis, and asthma. Different cell types display distinct migration behaviors depending on the strength of adhesion forces they impose on neighboring cells and/or on the surfaces on which they migrate. This, in turn, affects the shape of the cells and the speed at which they move. Mesenchymal migration is characterized by strong cell–substrate adhesion forces, which give rise to relatively slow migration speeds and low cell deformability. On the other hand, amoeboid motion is characterized by relatively few and weak cell–substrate adhesions, greatly deformable cell shapes, and fast motion (often reaching in excess of $10\ \mu\text{m min}^{-1}$). In this article, we describe the signal transduction mechanisms that regulate amoeboid-like motion as displayed by neutrophils and the soil-dwelling amoeba *Dictyostelium discoideum*, two systems that use common molecular mechanisms to sense chemical gradients.

During chemotaxis, both *Dictyostelium* and neutrophils are able to transduce shallow external chemical gradients into a highly polarized shape consisting of F-actin rich pseudopods at the front and myosin II contractions at the back, leading to cell elongation and directional migration. The sensing abilities of these cells are impressive, as they can (1) detect and respond to an external gradient as small as 2% across their length and (2) follow gradients that are rapidly changing in time. Many signaling molecules involved in this process have been identified through decades of research; however, motion is a space- and time-dependent process. Therefore, knowledge of the spatio-temporal dynamics of these molecules is of critical importance. Aiding in our understanding of this process is the fact that, to a large extent, the chemical gradient-sensing abilities of cells can be decoupled from their cytoskeletal polarities and motion capabilities. This allows a detailed investigation of each part of the chemotactic system independently.

Gradient Sensing

The first step in the chemotaxis response is the sensing of an external chemical or chemoattractant (Figure 1). In establishing polarity, the up-gradient side of the cell determines the front or leading edge, and the down-gradient side determines the back or trailing edge. In both neutrophils and *Dictyostelium*, gradient sensing is transmitted through transmembrane G-protein-coupled receptors (GPCRs). These receptors are associated with internal membrane-bound heterotrimeric G proteins containing α , β , and γ subunits. The binding of chemoattractants to GPCRs induces a conformational change that results in the dissociation of the coupled G proteins into G α and G $\beta\gamma$ subunits, which activate a wide array of downstream effectors (Figure 1). The fundamental question is: how do biochemical signals transduce external chemical gradients into highly polarized responses and migration? It has been shown that GPCRs and G proteins remain uniformly distributed on the plasma membrane during chemotaxis (Figure 2) and that the distribution of chemoattractant-bound GPCRs closely matches the external gradient. These findings show that signals downstream of receptor binding and G-protein dissociation lead to a polarized morphology.

In both *Dictyostelium* and neutrophils, chemoattractant stimulation of GPCRs leads to the rapid activation of Ras, which, in turn, activates phosphatidylinositol 3 kinase (PI3K). Activated Ras is the most upstream molecule identified to date that is localized primarily at the front of chemotaxing cells, indicating that the onset of polarity occurs between G proteins and Ras activation (Figure 2). Once activated, PI3K converts membrane-bound phosphatidylinositol-4,5-bisphosphate (PIP₂) into phosphatidylinositol-3,4,5-trisphosphate (PIP₃). PIP₃ acts as a membrane-docking site for molecules containing pleckstrin homology (PH) domains, such as Akt/protein kinase B (PKB), as well as nucleotide exchange factors for Ras (such as RasGEFs), and other small GTP binding proteins. These molecules in turn recruit actin polymerization/stabilization proteins to the front of cells, leading to pseudopod extension. While this pathway has been viewed for many years as the key gradient-sensing pathway,

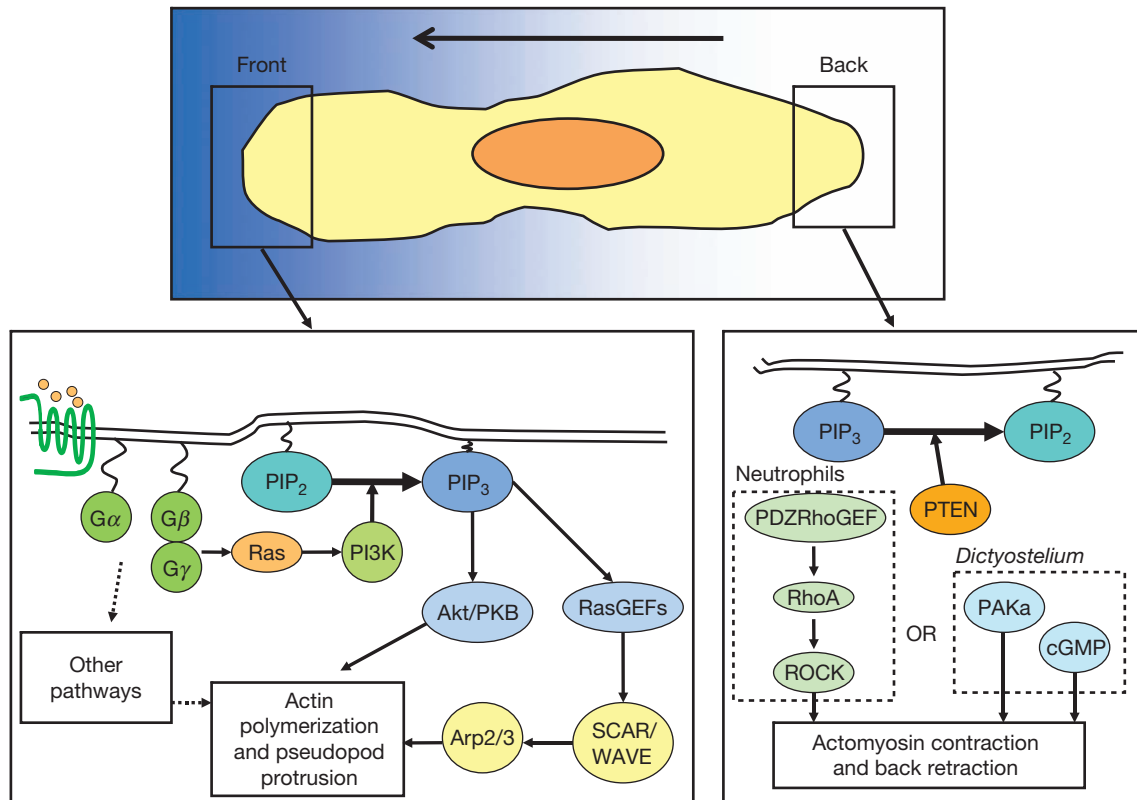


Figure 1 The chemotaxis signaling pathway. (Top) Scheme showing a polarized cell in a chemoattractant gradient. The front of the cell is oriented toward the higher concentration of chemoattractant. The black arrow indicates the direction of motion. (Bottom, left) At the front of the cell, GPCR activation initiates an intracellular signaling cascade that results in actin polymerization and pseudopod protrusion. The PI3K pathway, in which PIP₂ is converted into PIP₃, is shown in detail; other pathways are explained in the text and are not shown. (Bottom, right) At the back of the cell PIP₃ signaling molecules are converted back into PIP₂. In addition, further molecules are recruited to cause actomyosin contraction leading to retraction of the cell back. Note that the pathways at the cell back are only now becoming understood and, therefore, no links are drawn between the components.

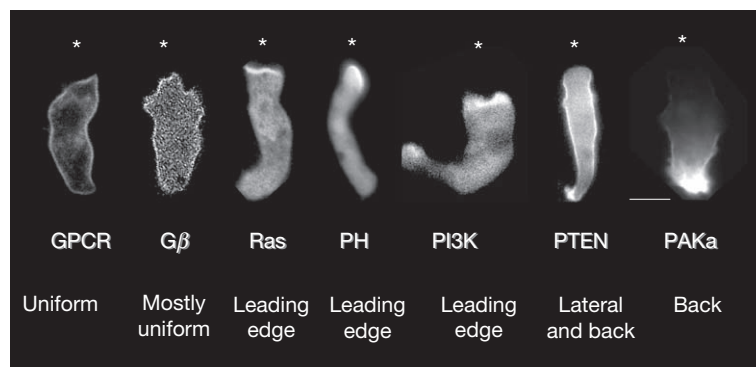


Figure 2 Fluorescence images depicting the cellular distribution of various GFP-labeled signaling components during chemotaxis. The up-gradient side is depicted by the stars. Ras is the most upstream molecule in the chemoattractant-sensing cascade to display spatial localization. From Comer FI and Parent CA (2002) PI3Kinases and PTEN: How opposites chemoattract. *Cell* 109: 541–544.

recent work in *Dictyostelium* has indicated that other effectors may be acting in parallel. These include phospholipase A₂ (PLA₂), guanylyl cyclase (GC), and target of rapamycin complex 2 (TORC2). Interestingly, these pathways may not have identical roles. For example, the PI3K pathway has been proposed to specifically confer sensitivity to shallow extracellular

gradients as PI3K-inhibited cells can still perform chemotaxis toward steep gradients.

In *Dictyostelium*, the phosphoinositide 3' phosphate phosphatase (PTEN), which de-phosphorylates PIP₃ into PIP₂, is recruited at the back and side of chemotaxing cells. This further ensures that PIP₃ is localized exclusively at the cell front,

effectively eliminating the binding and activation of PH domain-containing proteins at the back and sides. Cells lacking PTEN have multiple fronts and exhibit severe chemotaxis defects. In neutrophils, genetic knockout of the phosphoinositide 5' phosphate phosphatase (SHIP1), which also dephosphorylates PIP_3 , leads to unregulated PIP_3 levels and chemotaxis defects. SHIP1 therefore acts homologously to PTEN in these cells. The phosphorylation state of $\text{PIP}_2/\text{PIP}_3$ can therefore be viewed as an effective means to determine the gradient-sensing state of a cell (Figure 2).

Cytoskeletal Polarization

The gradient-sensing machinery described in the previous section drives directed migration by regulating where specific cytoskeletal elements should be active (Figure 1). For both *Dictyostelium* and neutrophils, signals downstream of PIP_3 bring the suppressor of cAR1 (SCAR)/Wiskott-Aldrich syndrome protein (WASP)-family verprolin homologous protein complex to the cell front. This complex then recruits Arp2/3, an actin-binding protein complex that promotes the nucleation of new actin filaments and pseudopod extension. In the back of the cell, actomyosin complexes are assembled and regulated by a variety of proteins (RhoA, ROCK, and PDZ/RhoGEF in neutrophils; PAKa and cGMP in *Dictyostelium*), and myosin-II based contraction pulls the back of the cell forward. These structures do not form in the front of cells, ensuring that contraction moves the entire cell forward.

Intiguently, cells with an already highly polarized cytoskeleton tend to have enhanced polarization of signaling molecules, with sharp distinctions between front and back. Indeed, an elongated cell, through its increased surface area, can better sample the external chemoattractant environment than a rounded cell, leading to better chemotaxis. In addition, there are positive feedback loops between gradient sensing and cytoskeletal proteins – that is, gradient-sensing molecules drive cytoskeletal polarization, which in turn further enhance the spatial localization of front and back signaling molecules. Exactly how this feedback is achieved is not known. One possibility is that cytoskeletal polarity stabilizes the gradient-sensing molecules; another is that polarity actually drives further recruitment. Importantly, the stimulation of cells with a uniform dose of chemoattractant can cause cells to polarize and randomly migrate (a process referred to as 'chemokinesis') in the absence of an external gradient. This further shows that the polarization of the cytoskeleton is not strictly dependent on the gradient-sensing machinery detecting an external gradient.

Finally, the role of the environment on chemotaxis is now beginning to be understood. Most chemotaxis experiments are carried out on simple two-dimensional (2D) surfaces, such as agar or glass for *Dictyostelium* and collagen or fibronectin for neutrophils. These types of substrates are, in many cases, an oversimplification of the types of three-dimensional (3D) environments that amoeboid cells encounter in their natural settings. It has been shown that patterning a 2D surface, either chemically or with physical etching, can itself direct migration. Recent work has also shown large differences in cell morphology between migration in 2D and 3D environments, and it

is becoming clear that various signaling pathways can play critical roles depending on the type of environment encountered. For example, neutrophil migration through a tight 3D matrix depends on one type of PI3K ($\text{PI3K}\gamma$) that is not as critical for 2D surface migration.

Mathematical Models

Although many of the key players regulating chemotaxis have been identified, a thorough understanding of the system requires knowledge of the timescales, binding rates, binding locations, and regulation of each of the molecules. To clarify this process, mathematical modeling has been used to determine if experimentally observed cell behavior can be entirely understood using known molecules and their spatiotemporal dynamics. This approach has many advantages, including suggesting quantitative experiments, classes of new molecules to look for, or even what measurements to take. By bringing together the approaches of quantitative experimental biology

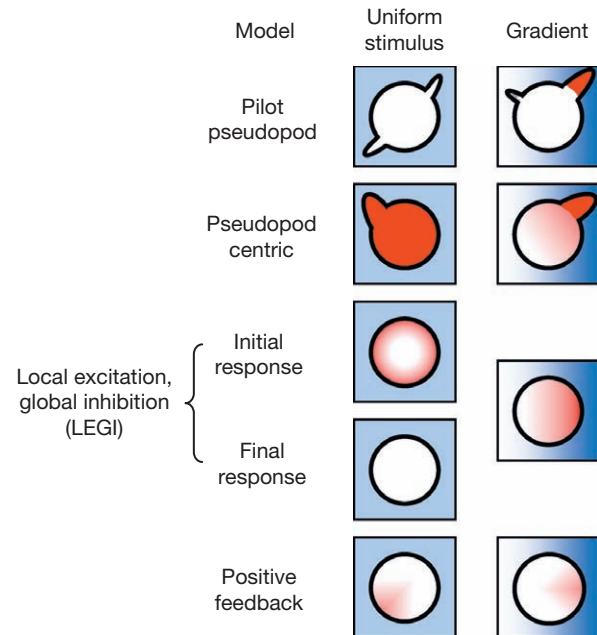


Figure 3 Schematic representation of mathematical models for gradient sensing. Cells (round) are exposed to either a uniform stimulus (left column) or a gradient (right column). The resulting spatial intracellular stimulation level of the cell is shown in red. In the pilot pseudopod model, cells detect a gradient by extending small pseudopods that sense a temporal change in chemoattractant receptor occupancy. In the pseudopod-centric model, stimulated cells continually extend large pseudopods, and in a gradient the internal sensing machinery biases the location of new pseudopods. In both of these models, a uniform stimulus does not change cell behavior. In the local excitation, global inhibition (LEGI) model, the response of the cell to a uniform stimulus is initially along the periphery but diminishes with time. In a gradient, the response matches the strength of the external gradient. In the positive feedback model, the cell has a region of strong excitation. When uniformly stimulated, this region is randomly located, whereas in a gradient it follows the direction of, but is much stronger than, the external gradient.

and mathematical modeling, researchers will eventually be able to fully understand the gradient sensing and motility networks of chemotaxing cells (see [Box 1](#) for an overview of how to measure various migration metrics).

Current mathematical models of gradient sensing generally follow the PI3K pathway and model the cell as a 2D circular object. [Figure 3](#) illustrates some of these mathematical models. One model is the pilot pseudopod model, where a cell continuously extends small pseudopods in many directions, only keeping and enlarging those that experience an overall positive change in chemoattractant-bound receptor. While this can account for motion in response to a gradient, it does not explain the observed enrichment of PH domain-containing proteins at the front of cells treated with agents that inhibit pseudopod formation. A pseudopod-centric model has also been proposed, where a cell continually extends large pseudopods and chemotaxes, but it biases the formation of new pseudopods with its gradient-sensing apparatus. This model also does not properly

account for all chemotactic behavior. It does not explain the absence of migration when no chemoattractant is present, although it is possible that the motion cycle has to be activated by the addition of chemoattractant.

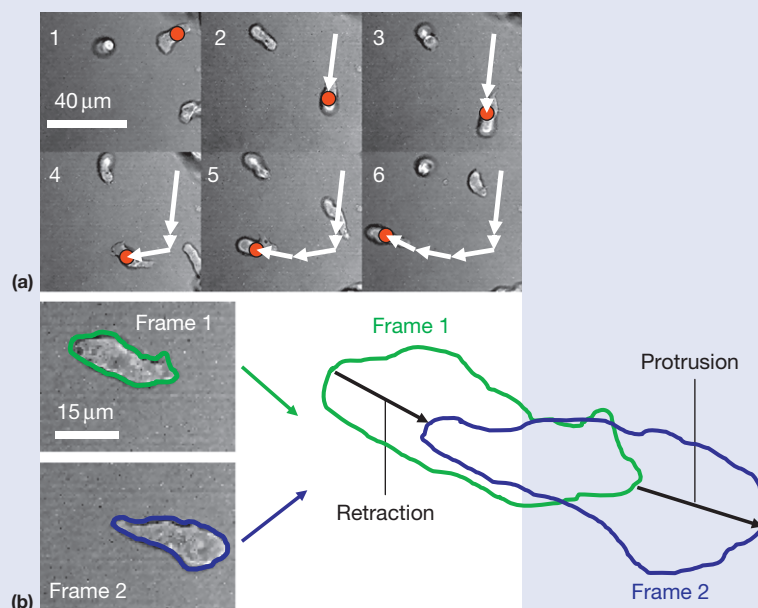
To address the gradient-sensing experimental findings, the local excitation, global inhibition (LEGI) model has been proposed. In this model, there are two chemical species: one is a stationary membrane-bound excitation molecule that is quickly activated by the binding of chemoattractants to GPCRs; the other is a diffusible slowly activating global inhibitor whose concentration is set by the average cell exposure to chemoattractants. A uniform stimulus of chemoattractant causes a rapid activation of the excitation molecule all around the cell periphery, which is dampened by the slow activation of the inhibitor molecule, which eventually deactivates the excitation molecule back to baseline. A steady gradient of chemoattractant, however, causes the excitation molecule to be more active at the front, while the inhibitor is present

Box 1 Metrics of Chemotaxis

There are several quantitative metrics that measure the ability of cells to perform chemotaxis. Often these metrics are used to determine the phenotype of a mutant in which a signaling pathway component has been suppressed or overexpressed. The metrics fall broadly into two types: center-of-mass and shape.

Center-of-mass statistics focus on the translocation of the center of a cell as it migrates (a). Such metrics are relatively easy to compute and understand; the center of a cell is followed in time as it moves, and various questions can be asked about its trajectory. Some software packages exist to automate the tracking of cell centers. One metric of motion is the speed of a cell, and another is the chemotaxis index, which is a measure of the fidelity with which a cell moves up a chemotactic gradient. Both speed and chemotactic index can be measured instantaneously (i.e., from one image frame to the next) or as a track total (i.e., from the start point to the end point). Instantaneous measures, if taken over too small a time, suffer from large uncertainties induced by the imaging noise and/or small variations in the center-finding procedure. Measures using only the start and end points of the tracks are less sensitive to noise but generally underestimate the speed and miss details in the chemotactic index. Other noninstantaneous metrics can also be useful. One example is the persistence of a cell's path, which measures the wiggleness or randomness with which a cell moves. This metric is less susceptible to noise as it involves a cell's entire path. However, it is still important to have enough data points to avoid overestimating the persistence. Therefore, for center-of-mass metrics, choosing an appropriate time frame for measurement is essential.

Shape metrics generally ignore overall cell motion and instead focus on how a cell's membrane deforms in time (b). These metrics tend to be more complicated and harder to compute, as they require accurate tracing of the membrane edge, though various software packages exist to assist in this process. A cell's overall shape and polarization (as measured by its circularity) are one metric, and another is the curvature of the membrane at every membrane point. These metrics can also be measured as functions of time, giving timescales for the cell's motion cycle. The location and time dynamics of protrusions and retractions are two important metrics. Finally, a powerful technique is to correlate shape behaviors with the spatiotemporal dynamics of fluorescently tagged gradient-sensing or cytoskeletal components, allowing further dissection of the chemotaxis process.



equally everywhere. This leads to net excitation at the front of the cell and net inhibition at the cell back (**Figure 3**). While Ras or PI3K can play the role of the local exciter, a candidate molecule for the global inhibitor has yet to be identified. This model is able to explain how cells treated with actin polymerization inhibitors are capable of sensing external gradients. However, it is not able to address how shallow external gradients are transduced into highly polarized responses (**Figure 2**), as the excitation in this model is proportional to the external gradient. Adding positive feedback loops via actin-binding proteins can introduce amplification of the chemotactic signal, but this causes spontaneous activation problems as well as insensitivity to changing gradients. Many other types of models exist but, in short, no single model has been able to accurately capture all of the complex dynamics cells display during chemotaxis. Most likely, a mix of the above models will accurately explain the biology of chemotactic sensing and motion.

See also: **Cell Architecture and Function:** Actin Assembly/Disassembly; Actin-Capping and -Severing Proteins; Actin-Related

Proteins; Cell Migration within Three-Dimensional Matrices; Cell Migration; Cytoskeletal Motors: General Principles; Myosin Motors; Rho GTPases and Actin Cytoskeleton Dynamics.

Further Reading

- Bagorda A and Parent CA (2008) Eukaryotic chemotaxis at a glance. *Journal of Cell Science* 121: 2621–2624.
- Iglesias PA and Devreotes PN (2008) Navigating through models of chemotaxis. *Current Opinion in Cell Biology* 20: 35–40.
- Insall RH (2010) Understanding eukaryotic chemotaxis: A pseudopod-centred view. *Nature Reviews Molecular Cell Biology*.
- Janetopoulos C and Firtel RA (2008) Directional sensing during chemotaxis. *FEBS Letters* 582: 2075–2085.
- Onsum MD and Rao CV (2009) Calling heads from tails: The role of mathematical modeling in understanding cell polarization. *Current Opinion in Cell Biology* 21: 74–81.
- Rappel W-j and Loomis WF (2009) Eukaryotic chemotaxis. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 1: 141–149.
- Smirnova T and Segall JE (2007) Amoeboid chemotaxis: Future challenges and opportunities. *Cell Adhesion and Migration* 1: 165–170.
- Swaney KF, Huang CH, and Devreotes PN (2010) Eukaryotic chemotaxis: A network of signaling pathways controls motility, directional sensing, and polarity. *Annual Review of Biophysics* 39: 265–289.

Eukaryotic DNA Polymerase α

C Lee and B K Tye, Cornell University, Ithaca, NY, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Teresa S.-F. Wang, volume 1, pp. 703–707, © 2004, Elsevier Inc.

Glossary

Cell cycle or cell division cycle Refers to the four sequential stages (G1, S, G2, and M) that are repeated in each cell division.

Cyclin-dependent kinases (CDKs) Phosphorylation enzymes whose activities are specified by their cyclin partners, which are periodically synthesized in the cell division cycle.

DNA polymerases Enzymes that join nucleotides together via phosphodiester bonds in a template-dependent manner.

Lagging strand Template for the DNA strand that is synthesized discontinuously due to the polarity of DNA polymerases.

MCM helicase Eukaryotic counterpart of the DnaB helicase of prokaryotes. It is a ring-shaped structure consisting of six nonidentical but highly conserved subunits (Mcm2, -3, -4, -5, -6, and -7) that encircles and unwinds DNA at the replication fork.

Okazaki fragments Short DNA fragments of about 200 nucleotides (~2000 nuc in bacteria) synthesized on the lagging strand in a polarity opposite to the unwinding of the replication fork.

Proofreading Refers to the editing function of the 3' to 5' exonuclease activity that removes mispaired nucleotides as they are being incorporated. It is associated with DNA polymerases that synthesize the bulk of the genomic DNA to provide accuracy to the copying mechanism.

Of the numerous DNA polymerases that replicate and repair DNA in eukaryotic cells, polymerase α -primase (Pol α) is the only enzyme capable of *de novo* DNA synthesis. This unique property is essential for initiating DNA synthesis in providing the primers for leading- and lagging-strand synthesis. The priming of Okazaki fragments in lagging-strand DNA synthesis is critical for coordinating DNA replication with repair and cell-cycle progression. Pol α has also been implicated in other cellular processes such as telomere metabolism, sister chromatid cohesion, and epigenetic regulation of nucleosome structure for transcriptional silencing. The article focuses on the roles of Pol α at DNA replication forks.

Structure and Enzymatic Function

Pol α is composed of four subunits of 180, 68, 58, and 48 kDa according to their molecular mass in SDS gel. In yeast, all four subunits are essential for viability and the basic 4-subunit structure is conserved in all eukaryotes. The largest 180-kDa subunit can be divided into three domains each with a specific function. The central domain encodes the catalytic DNA polymerase activity and the C-terminal domain is responsible for the subunits assembly. Consistent with the many functions of Pol α , the N-terminal domain has been shown to interact with several factors/structures with unrelated functions including Mcl1(And1/Ctf4), PP2A, concanavalin A, SV40 large T antigen, and Mcm10. The smallest p48 has the primase activity that synthesizes short RNA primers *in vitro*. However, the full primase activity requires p48 to be in a complex with p58. p68, also known as the B subunit, which regulates the nuclear activity of Pol α , is phosphorylated by cell-cycle-dependent kinases and its interaction with p180 is important for the nuclear import of the polymerase.

An important feature of the 4-subunit complex is that both catalytic activities are involved in initiating DNA replication by the synthesis of a chimeric 5' RNA–DNA 3' primer (Figure 1). The transition of RNA to DNA synthesis involves the initial synthesis of 7–10 bases of RNA oligonucleotide by the primase (p48), followed by an extension of the RNA primer by the DNA polymerase activity of the p180 subunit to 30–40 nt. The switch between the RNA primase and the DNA polymerase activity within Pol α is believed to be regulated by the unit length of the primer acting as a termination signal. The tight association of the primase and DNA polymerase within the same protein complex facilitates direct transfer of the RNA 3' end to p180 without dissociation from template.

In eukaryotes, replication initiates from certain sites on the chromosomes where the prereplication complex is assembled. Upon cell-cycle-dependent cues, unwinding of the DNA and assembly of the elongation complex occurs, signaling the start of DNA replication. At the leading strand, Pol α synthesizes the primers for extension by DNA polymerase ϵ . However, lagging-strand synthesis is more complex. The RNA polymerase activity of Pol α initiates synthesis of the primers that are completed by the DNA polymerase activity. Further elongation of the Okazaki fragments is achieved by a switching of polymerases from Pol α to Pol δ mediated by replication factor-C (RF-C) and replication protein A (RPA). Pol α is a low-fidelity enzyme with an incorporation error rate of 10^{-4} – 10^{-5} compared to the error rates of 10^{-6} – 10^{-7} of the main replicative DNA polymerases, Pol δ or Pol ϵ . As a result, primers synthesized by Pol α on the lagging strand are mutagenic if not removed. This problem is circumvented by the proof-reading property of Pol δ , which interacts with Pol α and collaborates with FEN1 and DNA ligase to remove the primers during the maturation of Okazaki fragments.

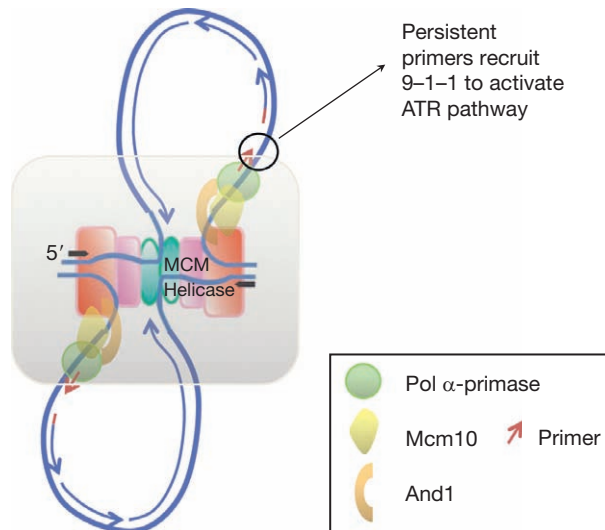


Figure 1 A model for the interaction between Pol α and the MCM helicase through the linkage of other factors such as Mcm10 and And1 on the lagging strand at the replication fork. The MCM helicase is a double hexamer with the N-terminal domains organized in a head-to-head configuration. Not shown are other polymerases of the replisome (shaded gray) such as Pol δ and Pol ϵ and the associated factors on the leading strand.

Though the main role of Pol α is to carry out DNA replication, there is evidence that Pol α has other roles. In conditional mutants of p180 and p48 subunits, replication of the chromosomal DNA is impaired, but DNA is still synthesized at the restrictive temperature. In addition to phenotypes such as increased mutation rate (measured by mitotic recombination, gene conversion, excision of large inverted regions, spontaneous mutations, etc.) and sensitivity to alkylating reagents, some Pol α mutants affect telomere length. Other mutants of Pol α are defective in localization of the major heterochromatin protein Swi6 at centromeres and telomeres suggesting a role for Pol α in epigenetics. Likely, the function of Pol α is not limited to DNA replication, but may include a broader role in chromatin biology.

Regulation of Pol α Activity

In budding yeast, the genes encoding the Pol α subunits are transcribed at the G1/S boundary. However, the protein subunits are stable as the cells are able to undergo several divisions after inhibition of new Pol α synthesis. Therefore, the initiation of DNA synthesis at the beginning of S phase is unlikely to be regulated by fluctuations in the Pol α level. However, cell-cycle-dependent phosphorylation of the two largest subunits has been observed at different stages of the cell-division cycle suggesting that the cell-cycle-dependent activity of Pol α may be regulated by posttranslational phosphorylation. Consistent with this idea, Pol α from human cells has the greatest capacity to replicate SV40 DNA when it is purified during the G1/S transition and this activity decreases as the cells reach G2/M. Furthermore, in human cells, p180 and p68 are phosphorylated by cyclin-dependent kinase (CDK) kinases

during the G1/S and G2/M transitions. In the fission yeast, p180 is phosphorylated during S phase. In the budding yeast, p68 is phosphorylated during the G1/S transition and dephosphorylated as the cells exit mitosis. The phosphorylation event, concomitant with RNA synthesis by Pol α , is dependent on the activation of the MCM helicase, melting of the origin DNA, and recruitment of topoisomerase I, Cdc45, and RPA, culminating in the initiation of DNA replication.

The cellular level and integrity of Pol α are critical for genome stability. Aberrant mutant activity causes chromosomal damage such as DNA breaks as well as changes in telomeric length and chromatin structure. Reduced level of Pol α results in elevated chromosome translocation and loss. Proper binding of p180 to p68 is required for nuclear localization, and misfolded p68 or p180 is barred from nuclear entry because of altered interaction. Consequently, misfolded Pol α is sequestered in the cytoplasm and channeled for proteasome-dependent degradation.

Pol α and the Helicase

The two main molecular motors in the replication elongation machinery are the DNA polymerases and the helicase. Coordination of these two activities on both the leading and lagging strands is important for the simultaneous replication of both strands as well as the prevention of excessive single-strand exposure due to unwinding. Moreover, coupling of the helicase and Pol α on the discontinuous strand ensures the proper spacing and timing for the *de novo* synthesis of Okazaki primers. The importance of an intimate association between primase and helicase activity is best illustrated in bacteriophage and prokaryotes. In bacteriophage T7 and p4, the primase and helicase are encoded in a single polypeptide. In *Escherichia coli*, the primase (DnaG) and helicase (DnaB) interact to form a complex. This interaction stimulates both the primase and helicase activities and relaxes the sequence specificity of DnaG, allowing DnaG to successively prime Okazaki fragment synthesis during lagging-strand synthesis. The hexameric DnaB helicase forms a ring structure that encircles the 5' lagging strand as it translocates and unwinds, while DnaG uses the newly exposed single-stranded DNA (ssDNA) as template. The topological linking of the helicase to DNA enhances processivity of the helicase and facilitates translocation of the primase to the priming sites.

In eukaryotes, the MCM helicase is considered the main replicative helicase. Like DnaB, the MCM helicase is a hexameric enzyme that forms a ring structure, but unlike DnaB the six subunits are nonidentical though highly conserved. The MCM helicase is loaded onto duplex DNA as a double hexamer, encircling the duplex DNA suggesting that the unwinding mechanism may be different from that of DnaB. It has been proposed that the MCM helicase may act as a double-strand DNA translocase that pumps duplex DNA into the double hexamer in a manner analogous to that of the SV40 large T antigen or the bacterial RuvAB Holliday junction branch-migration enzyme (Figure 1). Under this scenario, connecting the MCM helicase that is positioned ahead of the fork on the duplex DNA with Pol α that initiates on the unwound single strand must rely on other factors. Indeed, direct interaction

between Pol α and the MCM helicase has not been observed. A candidate for the linker protein is Mcm10 in budding yeast, which is required for both the initiation of DNA synthesis at the G1 to S phase transition and throughout the elongation process. Mcm10 interacts with both Pol α and MCM helicase and is required for Pol α stability in the budding yeast and humans (Figure 1). Interaction between the subunits of the MCM helicase and Mcm10 has been confirmed in *Xenopus*, and physical interaction between human Pol α and Mcm10 has been determined by crystallography. In *Xenopus* and humans, the cohesion establishment factor Ctf4/And1 also interacts with Pol α . Furthermore, interaction with RPA and Ctf4/And1 stimulates the primase activity and renders the enzymes more processive.

Coordinating DNA Replication with Checkpoint Response

The surveillance of vital cellular processes and the coordination of their ordered progression are important functions of the checkpoint pathways in eukaryotic cells. Under DNA replication stress or in the presence of DNA damage, the S-phase checkpoint system ensures the integrity of the replication fork and the proper firing of origins by delaying S-phase progression and by preventing exit from mitosis. The S- and G2/M-phase delay facilitates DNA repair and stress recovery. The key players in the budding yeast S-phase checkpoint pathway are Mec1 and Rad53 and the main replication stress signal that activates the checkpoint is thought to be RPA-coated ssDNA. The exposure of ssDNA occurs when unwinding of the double-stranded DNA by the helicase continues while the polymerase stalls. Inhibition of DNA synthesis as a result of checkpoint activation is important and many lines of evidence suggest that Pol α may be involved. The *pri1-M4* primase mutant is defective in DNA damage response and other Pol α mutants display synthetic lethality with mutations in the key checkpoint genes. These observations suggest that Pol α may function in coupling DNA replication to checkpoint activation as well as serving as one of the final targets of the checkpoint pathway. This scenario is conceptually attractive as inhibition of the primase activity due to DNA lesion will slow down replication and allow repair. Failure to do so will increase mutation rates. However, Pol α is not a direct target of Rad53, but rather, it is suggested that ssDNA-bound RPA, which physically interacts with Pol α , may be the mediator that transduces the signal from Rad53 to Pol α .

In addition to its role in initiation of DNA synthesis, RNA primers may also function in checkpoint activation as the type of chromatin structure required for checkpoint activation included the synthesis of an RNA primer. Specifically, DNA unwinding, Pol α loading, and RNA synthesis, but not new DNA synthesis, are required for intra-S-phase checkpoint activation. Blocking primase activity by the use of primase inhibitors or by utilizing Pol α mutants defective only in RNA synthesis inhibited replication but did not induce checkpoint activation. This observation suggested that RNA primer, in addition to RPA-coated ssDNA, is a component of the signal for checkpoint activation.

A detailed mechanism of how stalled forks are sensed by the Ataxia Telangiectasia-mutated and Rad3-related (ATR)

(homolog of Mec1) pathway remains to be understood, but Pol α seems to play an important role in providing the substrate for the activation of the ATR pathway (Figure 1). Tethering ATR to RPA-coated ssDNA through interaction with ATR interacting protein (ATRIP) is required for ATR activation. However, recruitment of ATR to RPA ssDNA is not sufficient, TOPBP1 (Mus101/Rad4/ Cut5/Dpb11) and 9–1–1 (Rad9, Rad1, Hus1) complex are also required. TOPBP1 has an ATR activation domain that interacts with both ATR and ATRIP and stimulates ATR kinase activity. This ATR activation requires the interaction of TOPBP1 with Rad9. It has been suggested that 9–1–1 helps to position TOPBP1 to stimulate ATR kinase activity. The current model for checkpoint activation is that under replication stress RPA-coated ssDNA recruits ATR–ATRIP, while 9–1–1, loaded onto DNA, associates with TOPBP1, and facilitates the activation of ATR–ATRIP–RPA. The role of Pol α in this process is to recruit TOPBP1 and the 9–1–1 complex by providing a specific chromatin substrate for 9–1–1, which specifically binds 5' DNA junction. Primer synthesis by Pol α creates just such a structure. The idea that persistent primers are the direct targets of ATR activation is corroborated by the observation that primer synthesis is required for checkpoint activation. These observations paint an intricate relationship between primer synthesizing activity on the lagging strand and checkpoint surveillance.

The vital role that Pol α seems to play at the fork during elongation suggests a possible advantage in having a discontinuous lagging-strand synthesis composed of multiple Okazaki fragments. The persistent presence of Pol α at the fork places Pol α in a position to directly regulate DNA synthesis on the lagging strand. As one of the final targets of checkpoint activation Pol α is poised to coordinate DNA replication with checkpoint surveillance. However, this important role would require the coordinated modulation of both primase and helicase activity because uncoupling of the two motors would exacerbate ssDNA exposure. Unlike the leading strand where DNA unwinding is immediately followed by DNA synthesis, lagging-strand synthesis involves significant exposure of unwound ssDNA for retrograde primer synthesis by Pol α . Therefore, modulation of Pol α alone or uncoupling of the helicase and primase activity of the lagging strand would present an increasingly dire situation for recovery. As the level of checkpoint activation increases with the amount of ssDNA accumulation this uncoupling would further stimulate checkpoint activation. To compensate, additional modifications of the helicase to slow down DNA unwinding or alternative mechanism of coupling the primase and helicase would be required under replication stress. Accordingly, MCM helicase subunits are phosphorylated by the checkpoint pathway during replication stress conditions. In the budding yeast, Pol α is destabilized in the conditionally lethal *mcm10-1* mutant presumably as a result of an unrestrained helicase activity that leads to extended exposure of ssDNA and collapse of the replication fork. A recent study showed that suppressor mutations in Mcm2, which compromise the MCM helicase activity, reduce ssDNA exposure and stabilize Pol α suggesting that fork stability is restored. In addition to the suppressor mutation in the MCM helicase, stabilization of Pol α requires the function of the Mec1 checkpoint protein suggesting that Mec1 provides an alternative mechanism for stabilizing the replication fork

under replication stress. The dynamic relationship between the replication and checkpoint factors underscores the measures that eukaryotic cells must take to provide a failsafe mechanism for keeping Pol α and the MCM helicase intact at the fork for the resumption of DNA synthesis once repair is complete.

See also: **Cell Architecture and Function:** Cell Cycle: DNA Damage Checkpoints; **Molecular Biology:** DNA Polymerase I, Bacterial; DNA Polymerase III, Bacterial; DNA Polymerase β , Eukaryotic; DNA Replication Fork, Bacterial; DNA Replication Fork, Eukaryotic; Eukaryotic DNA Polymerase δ .

Further Reading

- Bermudez VP, Farina A, Tappin I, and Hurwitz J (2010) Influence of the human cohesion establishment factor Ctf4/AND-1 on DNA replication. *Journal of Biological Chemistry* 285(13): 9493–9505.
- Burgers PMJ (2009) Polymerase dynamics at the eukaryotic DNA replication fork. *Journal of Biological Chemistry* 284(7): 4041–4045.
- Corn JE and Berger JM (2006) Regulation of bacterial priming and daughter strand synthesis through helicase-primase interactions. *Nucleic Acids Research* 34(15): 4082–4088.
- Fioani M, Lucchini G, and Plevani P (1997) The DNA polymerase α -primase complex couples DNA replication, cell-cycle progression and DNA-damage response. *Trends in Biochemical Sciences* 22(11): 424–427.
- Frick DN and Richardson CC (2001) DNA primases. *Annual Review of Biochemistry* 70(1): 39–80.
- Gambus A, van Deursen F, Polychronopoulos D, et al. (2009) A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase α within the eukaryotic replisome. *EMBO Journal* 28(19): 2992–3004.
- Hubscher U, Maga G, and Spadari S (2002) Eukaryotic DNA polymerases. *Annual Review of Biochemistry* 71(1): 133–163.
- Lee C, Liachko I, Bouten R, Kelman Z, and Tye BK (2010) Alternative mechanisms for coordinating polymerase alpha and MCM helicase. *Molecular and Cellular Biology* 30(2): 423–435.
- Li D, Zhao R, Lilyestrom W, et al. (2003) Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. *Nature* 423: 512–518.
- Ricke RM and Bielinsky AK (2004) Mcm10 regulates the stability and chromatin association of DNA polymerase-alpha. *Molecular Cell* 16(2): 173–185.
- Shiotani B and Zou L (2009) ATR signaling at a glance. *Journal of Cell Science* 122(3): 301–304.
- Sugino A (1995) Yeast DNA polymerases and their role at the replication fork. *Trends in Biochemical Sciences* 20(8): 319–323.
- Warren EM, Huang H, Fanning E, Chazin WJ, and Eichman BF (2009) Physical Interactions between Mcm10, DNA, and DNA Polymerase α . *Journal of Biological Chemistry* 284(36): 24662–24672.
- Yan S and Michael WM (2009) TopBP1 and DNA polymerase α -mediated recruitment of the 9–1–1 complex to stalled replication forks: Implications for a replication restart-based mechanism for ATR checkpoint activation. *Cell Cycle* 8(18): 2877–2884.
- Zhou Y and Wang TS-F (2004) A coordinated temporal interplay of nucleosome reorganization factor, sister chromatin cohesion factor, and DNA polymerase α facilitates DNA replication. *Molecular and Cellular Biology* 24(21): 9568–9579.

Eukaryotic DNA Polymerase δ

P M J Burgers, Washington University School of Medicine, St. Louis, MO, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Antero G. So and Kathleen M. Downey, volume 1, pp. 713–715, © 2004, Elsevier Inc.

Structure

Catalytic Subunit

Pol δ is a multisubunit complex, consisting of a catalytic subunit, and, depending on the organism, two or three accessory subunits (Table 1). The catalytic subunit is essential for yeast cell growth and belongs to the B-family of DNA polymerases found in all three kingdoms of life.

The simplest B-family DNA polymerase is the replicative DNA polymerase of bacteriophage RB69. Its structure shows three distinct domains: a large N-terminal domain, the 3'-exonuclease domain, and the polymerase domain. The polymerase domain shows a structure that is typical for DNA polymerases with the template primer junction and the dNTP binding into a broad cleft that closes up into a closed complex prior to the chemical polymerization step. The single-stranded template DNA enters the active site from a cleft situated between the N-terminal domain and the exonuclease domain, whereas the double-stranded DNA exits the active toward the bottom in Figure 1(a). The active site of the polymerase domain is precisely defined to promote favorable binding of that incoming dNTP that will form a proper Watson–Crick base pair with the next template nucleotide to be replicated, whereas non-Watson–Crick base pairs show a poor fit and are, therefore, disfavored. These geometric considerations, together with the favorable energetics obtained from Watson–Crick base pairing, result in a high fidelity of replication. Yeast Pol δ has an error rate of about 10^{-4} – 10^{-5} , that is, there is one misincorporation event for every 10^4 – 10^5 nucleotides (nt) replicated. The fidelity of DNA replication by Pol δ is enhanced about 10-fold through proofreading by its 3'-5'-exonuclease activity.

The 3'-exonuclease domain of Pol δ belongs to the same class as those identified in other proofreading DNA polymerases. Two divalent metals in the exonuclease domain coordinate with aspartates/glutamates to orient a water molecule as a metal-hydroxide ion for attack of the internucleotide phosphodiester bond to be cleaved. The active site of the 3'-exonuclease domain is ~ 35 Å distant from the active site of the polymerase domain. Misincorporation by the polymerase results in stalling of further DNA synthesis because of a poor geometric fit of the mismatched template–primer junction in the active site of the polymerase domain. This stalling promotes dissociation of the primer terminus from the polymerase domain and binding in the exonuclease domain. Although it is yet unclear how this translocation from the polymerase to the exonuclease domain occurs, it is an important step in maintaining replication fidelity, and it is just as important that this translocation is reversible so that polymerization can resume after hydrolysis of the misincorporated nucleotide. The relative orientation of the polymerase and exonuclease domains in B-family DNA polymerases is completely different from that in the A-family

enzymes; yet both types of enzymes evolved a pathway of facile shuttling of the primer terminus between the two domains.

The overall structure of the N-terminal domain is also largely conserved between various B-class DNA polymerases. It forms a three-lobed structure, with one lobe showing similarity with single-stranded DNA-binding motifs and a second lobe showing similarity with ribonucleic acid (RNA)-recognition motifs. Given that all crystal structures of polymerase–DNA complexes show the single-stranded template DNA extruding toward a cleft between the N-terminal domain and the exonuclease domain, a likely function of the N-terminal domain is to bind the downstream single-stranded template DNA. In addition, Pol δ is involved in several pathways that require gap-filling reactions including those that occur during Okazaki fragment maturation on the lagging strand of the replication fork. Therefore, the N-terminal domain may recognize the downstream 5'-single-/double-stranded DNA junction during gap filling.

In addition to these three conserved domains shared with all B-family DNA polymerases, the cellular B-family enzymes uniquely contain a highly conserved C-terminal domain. This domain shows a typical structural arrangement of conserved cysteine residues that chelate divalent metals. In the case of Pol α , two Zn ions are found at these sites, and it is these motifs that mediate interactions with the other subunits of the complex. In addition to the structural functions ascribed to this C-terminal domain, it may also play a specialized function in mediating damage-induced mutagenesis in the cell as mutations in the domain are specifically defective for this process.

Associated Subunits

The second subunit of Pol δ shows considerable structural homology with the second subunits of Pol α and Pol ϵ . Given the high conservation of the metal-binding domains of the C-termini of the catalytic subunits, it is likely that subunit–subunit interactions between each of the three catalytic subunits with their respective second subunits are also conserved. The proposed subunit–subunit interface between yeast Pol3 and Pol31 in Figure 1(b) is based on the crystal structure of the homologous subunit interactions between the catalytic and second subunit of Pol α . A possible function for an OB-fold domain (oligonucleotide/oligosaccharide binding) in Pol31, beyond its involvement in subunit interactions, remains to be established. OB-fold domains are often active in binding single-stranded DNA.

The third subunit of Pol δ consists of a ~ 15 kDa N-terminal domain that mediates interactions with the second subunit, followed by a largely unstructured domain of 20–45 kDa, and a proliferating cell nuclear antigen (PCNA)-binding domain at the extreme C-terminus of the subunit. In addition, interactions with Pol α and with Rev1, a subunit of the cellular

mutagenesis complex, have been identified. The presence of the middle domain is responsible for the highly elongated structure of this subunit, and consequently of the entire Pol δ complex. *Saccharomyces cerevisiae* Pol32 is dispensable for growth, but the deletion shows defects in DNA replication, mutagenesis, and in break-induced DNA replication. By contrast, the *S. pombe* *Cdc27* gene is essential.

A small fourth subunit of Pol δ has been identified in human and *S. pombe*, but appears to be lacking from *S. cerevisiae* Pol δ . This subunit enhances the stability of the complex. It may also play a regulatory role during damage response in the cell as this subunit is specifically degraded after DNA damage.

PCNA as Processivity Factor

Pol δ has a very low processivity of DNA synthesis, which would make it an unsuitable enzyme for replicating long templates with high efficiency. However, its interaction with PCNA transforms Pol δ into a highly processive enzyme, capable of replicating thousands of nucleotides without dissociating from the DNA. PCNA is a donut-like homotrimer that encircles dsDNA and binds proteins, thereby locking them onto the DNA. In addition to Pol δ , almost a hundred other PCNA-interacting proteins have been identified that function in DNA metabolism and cell-cycle control.

Table 1 Subunit structure of Pol δ

Genes and subunit sizes			Function
<i>S. cerevisiae</i>	<i>S. pombe</i>	Human	
<i>POL3</i> 125	<i>Pol3</i> 124	POLD1 124	Polymerase, 3'-exonuclease
<i>POL31</i> 55	<i>Cdc1</i> 51	POLD2 51	Required for PCNA-dependent replication
<i>POL32</i> 40	<i>Cdc27</i> 42	POLD3 66	PCNA binding
-	<i>Cdm1</i> 19	POLD4 12	Complex stability

Gene designations and subunit sizes in kDa are given.

The PCNA-binding domain at the C-terminus of the third subunit of Pol δ is responsible for DNA-independent interactions with PCNA. When both PCNA and Pol δ are assembled onto DNA, additional PCNA-interaction domain(s) become active in promoting processivity. The precise domain(s) in Pol δ responsible for these DNA-dependent interactions with PCNA still remain to be mapped. However, PCNA-dependent processive DNA replication requires at a minimum both the catalytic and second subunit of Pol δ .

Functions of Pol δ

The two catalytic functions of Pol δ , that is, polymerization and 3'-exonuclease, support high-fidelity DNA replication, a necessary function for genome maintenance. In addition, these two catalytic functions have been adapted for specialized functions of Pol δ , in particular those involving gap-filling reactions. During gap filling, when the downstream 5' ds/ss junction is reached, Pol δ can invade the double-stranded region and, thus, carry out strand-displacement synthesis. Remarkably, however, strand-displacement synthesis is limited to only a few nucleotides producing a very short 5'-flap that, depending of the pathway involved, is cleaved by an appropriate 5'-exonuclease. In the absence of this processing by the 5'-exonuclease, Pol δ switches the primer terminus to its 3'-exonuclease domain that degrades the extended terminus back to the nick position. The recurring process of extension and degradation, called idling, allows Pol δ to dynamically maintain the primer terminus either at the nick position, to promote ligation of the nick, or at a 5' flap position, for degradation by a 5'-exonuclease (Figure 2(a)). Idling prevents the generation of long flaps that can be a source for genomic instability.

DNA Replication

Recent studies have established that leading-strand DNA replication is carried out by Pol ϵ , while Pol δ is the lagging-strand

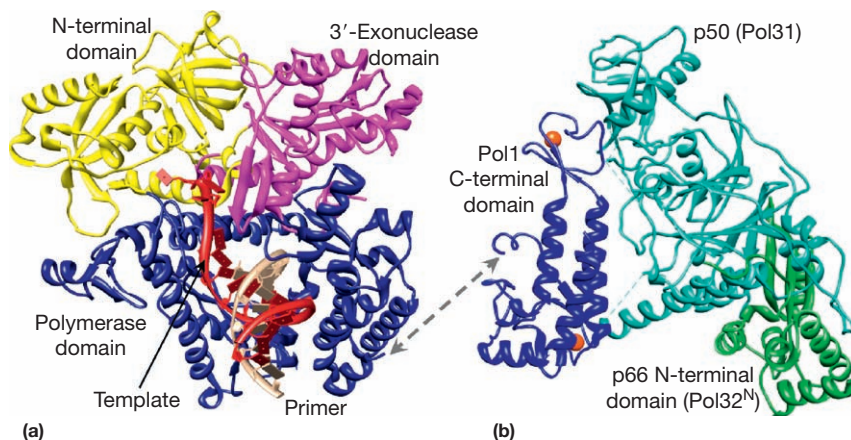


Figure 1 Structural components of Pol δ . All structures were rendered using Chimera, from coordinates deposited in the PDB database. (a) The *S. cerevisiae* Pol3 polypeptide, aa 95–985, in a complex with DNA (3IAY). The N-terminal, exonuclease and polymerase domains are indicated. The C-terminal domain of Pol3 is missing from this structure. (b) The C-terminal domain (3FLO) of *S. cerevisiae* Pol 1 (catalytic subunit of Pol α), which is highly homologous to that of Pol3, is docked onto the structure (3EOJ) of the human heterodimeric complex of p50 with p66 (N-terminal domain). The proposed connectivity between the N-terminal and C-terminal domains of Pol3 is indicated with a double gray arrow.

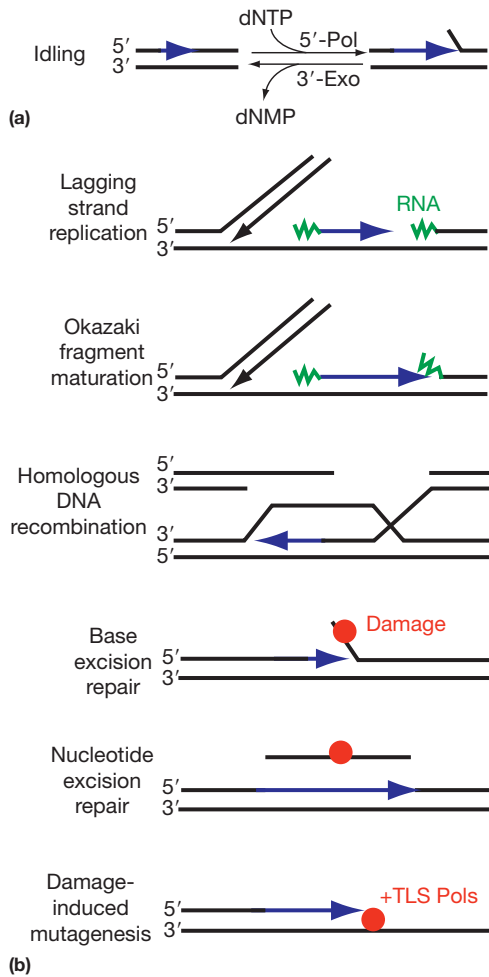


Figure 2 Activities and pathways of Pol δ . (a) Idling is the result of recurring cycles of polymerization and exonuclease activity. (b) Pathway involvement for Pol δ . See text for details. TLS, translesion synthesis.

DNA replicase. However, in the case of dysfunction of the leading strand replicase Pol ϵ , Pol δ can, in addition, perform leading-strand DNA replication. The ability of Pol δ to function on both strands may be important in some DNA damage-response pathways that operate at damaged or unusual DNA replication forks. Each Okazaki fragment on the lagging strand is initiated by the primase activity of Pol α -Primase. After a switch to Pol δ , the enzyme can correct replication errors that were made by Pol α during initiation. Pol δ elongates Okazaki fragments, which are ~150 nt in length, and also matures these fragments. The predominant pathway for maturation proceeds in cooperation with the 5'-exonuclease FEN1 and DNA ligase I. It is during maturation that the strand-displacement activity of Pol δ becomes important. By carrying out strand displacement synthesis, Pol δ dynamically generates the substrate for the FEN1 5'-flap exonuclease ensuring degradation of the RNA primer (Figure 2(b)). Once the RNA is degraded, DNA ligase I can seal the nick. All of these reactions are coordinated by interactions of each of these factors with PCNA.

When replication forks stall because of the presence of template DNA damage, several mechanisms are available to deal with this damage. One of these mechanisms is to

switch to a mutagenic form of DNA replication that involves low-fidelity DNA polymerases that replicate across the damaged template, a process called translesion synthesis (TLS). The activity of the low-fidelity TLS Pol ζ forms the basis for damage-induced mutagenesis in eukaryotic cells. Mutagenesis is a highly controlled process that involves the participation of several other factors. Genetic studies have shown that Pol δ is one of the factors essential for damage-induced mutagenesis. This is rather unexpected because Pol δ is a high-fidelity enzyme that would be very inefficient in TLS. Possibly, Pol δ has a role in organizing the mutagenic complex during TLS.

DNA Recombination

The repair of double-stranded breaks in the cell by homologous recombination is the most accurate repair mechanism as it uses the information on the homologous chromosome to repair the information lost at the break site. Homologous recombination is initiated by strand-specific resection from the double-stranded break in order to generate long 3'-single-stranded tails. These 3'-tails invade the intact homologous chromosome using standard leading to the formation of a D-loop. DNA synthesis using the 3'-terminated strand of the D-loop as a primer is an essential step in homologous DNA recombination as it restores the information lost at the initial break site. The strand-displacement synthesis in this step is carried out by Pol δ and its interaction with PCNA is essential in the initiation of this process. Resolution of recombination intermediates restores the proper genetic identity of the broken chromosome.

Double-stranded breaks can also be generated by collapse of stalled replication forks or by gradual degradation of the telomeric ends of chromosomes. Such breaks have only one broken end and do not undergo classical recombination. Rather, after strand invasion, they establish a replication fork that proceeds to the end of the invaded homologous chromosome, hence break-induced replication (BIR). During the initial stages of BIR, replication is carried out by Pol δ . Remarkably, in the absence of the nonessential Pol32 subunit of Pol δ , BIR is largely defective. Only at later stages of BIR is a more standard replication fork generated that also contains Pol ϵ .

DNA Repair

Among the various DNA repair pathways, both base excision repair (BER) and nucleotide excision repair (NER) involve a DNA synthesis step. Two distinct pathways exist for removal of damaged nucleotides in BER. After 5'-incision of the damaged strand, repair can proceed either through a Pol β -dependent pathway or through a Pol δ -dependent pathway. The latter pathway is also called long-patch or PCNA-dependent BER. The damage is removed by PCNA-dependent strand-displacement synthesis by Pol δ coordinate with PCNA-promoted excision of the damaged strand by FEN1. This pathway is comparable to that of RNA degradation during Okazaki fragment maturation. In contrast to BER, it is less clear which DNA polymerases carry out the required synthetic step during NER. NER proceeds by excision of a ~30-nt section of

single-stranded DNA surrounding the DNA damage leaving a gapped DNA intermediate. Genetic studies in yeast have implicated both Pol δ and Pol ϵ as enzymes capable of carrying out gap-filling synthesis. However, there is evidence that in human cell, the low-fidelity Pol κ may also serve a role in gap filling.

See also: **Molecular Biology:** DNA Mismatch Repair in Disease and Aging; DNA Polymerase I, Bacterial; DNA Replication: Initiation in Bacteria; Repair of G–T Mismatches by *Escherichia coli* Vsr and Eukaryotic DNA Glycosylases.

Further Reading

Burgers PM (2009) Polymerase dynamics at the eukaryotic DNA replication fork. *Journal of Biological Chemistry* 284: 4041–4045.

Hitomi K, Iwai S, and Tainer JA (2007) The intricate structural chemistry of base excision repair machinery: Implications for DNA damage recognition, removal, and repair. *DNA Repair* 6: 410–428.

Jain R, Hammel M, Johnson RE, et al. (2009) Structural insights into yeast DNA polymerase delta by small angle X-ray scattering. *Journal of Molecular Biology* 394: 377–382.

Johansson E and Macneill SA (2010) The eukaryotic replicative DNA polymerases take shape. *Trends in Biochemical Sciences* 35: 339–347.

Kunkel TA and Burgers PM (2008) Dividing the workload at a eukaryotic replication fork. *Trends in Cell Biology* 18: 521–527.

Lydeard JR, Jain S, Yamaguchi M, and Haber JE (2007) Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* 448: 820–823.

Relevant Website

<http://www.cgl.ucsf.edu> – UCSF Chimera.

Eukaryotic Protein Biosynthesis: The Elongation Cycle

A A Komar, D J Taylor, and W C Merrick, Cleveland State University, Cleveland, OH, USA; Case Western Reserve University, Cleveland, OH, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

E, A, P, and F sites Physical locations in the intersubunit space of the ribosome that are occupied by aminoacyl-, peptidyl-tRNA, or free tRNA.

Elongation (of protein synthesis) The sequential steps that lead to the addition of one amino acid at a time to the growing polypeptide chain.

Elongation factor (EF) A nonribosomal protein that facilitates the process of elongation only. (note: eukaryotic elongation factors are designated eEF where the lower case 'e' signifies 'eukaryotic').

Initiation (of protein synthesis) The required steps that lead to the placement of the initiator transfer RNA (tRNA) in the P site of the ribosome, correctly base paired with the initiating AUG codon.

Protein synthesis The process of joining amino acids in a specific sequence through the α carbonyl and α amino groups via a peptide bond that is templated by an mRNA molecule.

Termination (of protein synthesis) The codon-directed (UAA, UAG, or UGA) process of cleavage (and therefore, release) of the polypeptide chain from the tRNA in the P site of the ribosome.

Introduction

Protein synthesis (i.e., translation) is a complex event catalyzed by over 20 factors in eukaryotes and regulated by many more. That said, the purpose of this article is to provide a basic understanding of the process of translation elongation, a central step in the protein's synthesis, leading to the sequential addition of amino acids to the polypeptide chain. Although the focus will be on eukaryotic systems, the factors involved and the general mechanism of protein synthesis elongation are essentially conserved in eukaryotes and prokaryotes (see Table 1). This is perhaps not surprising given the ease with which one can trace the evolution of the eukaryotic factors from their bacterial predecessors. At present, the most complete biochemical and structural data on the elongation cycle comes from bacterial systems and includes crystallographic structures of the 70S ribosome in complex with messenger RNA (mRNA) and transfer RNAs (tRNAs). Recently, X-ray structures of the 70S ribosome in complex with individual elongation factors and their various substrates and cofactors (in particular, EF1A-GDP, EF2-GDP, and EF1A-GDPNP-Phe-tRNA^{Phe}) have emerged. In addition, valuable structural information and many mechanistic insights have been obtained from the cryo-electron microscopy structures of ribosomes (from both prokaryotes and eukaryotes) in complex with initiation, elongation, termination, and recycling factors. The regulation of the elongation cycle has been studied in the eukaryotic system with several posttranslational modifications of the elongation factors that appear to influence the rate of peptide synthesis. Much of the kinetic information regarding the elongation cycle has been obtained from the prokaryotic system. These data have most recently come from sophisticated experiments employing pre-steady state kinetic measurements and single-molecule fluorescence resonance energy transfer (FRET) analysis utilizing labeled components of the reaction.

The Traditional Eukaryotic Elongation Cycle: The A and P sites

The elongation cycle is defined as a process leading to the sequential addition of amino acids to the polypeptide chain. It is catalyzed by a set of extremely conserved elongation factors, whose function is depicted in Table 1. After initiation, the 80S ribosome is bound to mRNA (or AUG codon) correctly base paired with the initiator tRNA (Met-tRNA_i) in the peptidyl or P site of the ribosome (see Figure 1). This placement of the initiator tRNA in the P site sets the reading frame for all subsequent aminoacyl-tRNAs (aa-tRNAs). The next event is the binding of the ternary complex (eEF1A-GTP-aa-tRNA) to the aminoacyl site (A site) (see Figure 1). This binding is directed by the interaction of the anticodon of the tRNA and the three nucleotides of mRNA in the A site. The specificity is such that perfect Watson-Crick base pairs are observed for the first two nucleotides in the codon (i.e., only A-U or G-C base pairs), but altered base pairing is possible at the third position (as originally proposed by Crick). The Genetic Code dictates which aa-tRNA will bind to the A site in response to the 61 possible triplet codons that specify an amino acid (note: three codons of the possible 64 are stop codons: UAA, UAG, and UGA). This process, called decoding, is regulated by the ribosomal RNA of the 40S subunit. If a correct codon-anticodon base pair forms between the mRNA and the incoming aa-tRNA, a conformational change occurs which triggers the hydrolysis of guanosine triphosphate (GTP) leading to the release of eEF1A-GDP from the ribosome.

Decoding is followed by peptide bond formation. The synthesis of the peptide bond is catalyzed by the peptidyl transferase center in the large ribosomal subunit (60S). *In vivo*, this step appears to occur instantaneously and is likely the most rapid step in the entire process of elongation. Decoding and peptide bond formation (as should be noted in Figure 1) lead to the movement of the upper halves of the two tRNAs such

that the tRNA originally in the P site is now half in the E site (aminoacyl end) of the 60S subunit, while the lower half of the tRNA remains in the P site (the anticodon end) of the 40S subunit. The partial translocation of the tRNA molecule results in a hybrid, P/E-site tRNA. In a similar manner, the tRNA originally in the A site now resides in an A/P-hybrid state. The hybrid tRNA states are formed in the absence of eEF2 and are authentic intermediates in the elongation process.

The final step in the elongation cycle is translocation. Translocation is driven by the binding of eEF2-GTP to the A site of the ribosome. Hydrolysis of the GTP in the eEF2-GTP complex, which is stimulated by the ribosomal stalk, triggers the

movement of the peptidyl-tRNA in the A/P-hybrid site into the canonical, P/P site. By moving the anticodon of the peptidyl-tRNA fully into the P site (P/P), the mRNA is moved along the ribosome precisely by three nucleotides (or one codon). In this manner, the ribosome maintains the reading frame of the mRNA. It should be noted that translocation is accompanied by a rotational movement (ratchet) of the ribosomal subunits relative to each another. These conformational changes involve a ratchet-like rotation of the 40S subunit with respect to the 60S subunit. The combination of eEF2 binding, GTP hydrolysis, and ratcheting of the ribosomal subunits results in two distinct states: the posttranslocational state, where the ribosome has a peptidyl-tRNA in the P site and a vacant A site ready to decode an incoming aa-tRNA and the pretranslocational state, where the ribosome has a deacylated tRNA in the P site and a peptidyl-tRNA in the A site.

After each round of elongation, the entire process is repeated for each amino acid encoded by the mRNA transcript until the mRNA reaches a STOP codon. During the elongation cycle, and as noted above, eEF1A is released as an eEF1A-GDP complex. The binding affinity of eEF1A for GDP is about 0.1 μM . Thus, in biological terms, this is a rather stable complex. To enhance the conversion of eEF1A-GDP into the biologically active eEF1A-GTP form, the eEF1A-GDP complex binds to eEF1B (a nucleotide exchange factor). eEF1B is composed of two subunits, eEF1B α and eEF1B γ , in yeast and three

Table 1 Eukaryotic and prokaryotic components of the elongation cycle

Prokaryotes	Eukaryotes	Function
30S, 50S subunits	40S, 60S subunits	Peptide bond formation
EF1A	eEF1A	Binding of aa-tRNA to ribosomes
EF1B	eEF1B	Nucleotide exchange factor
EF2	eEF2	Translocation
tRNA	tRNA	Carrier of the activated amino acid (as aa-tRNA, recognition of the codon in the A site)

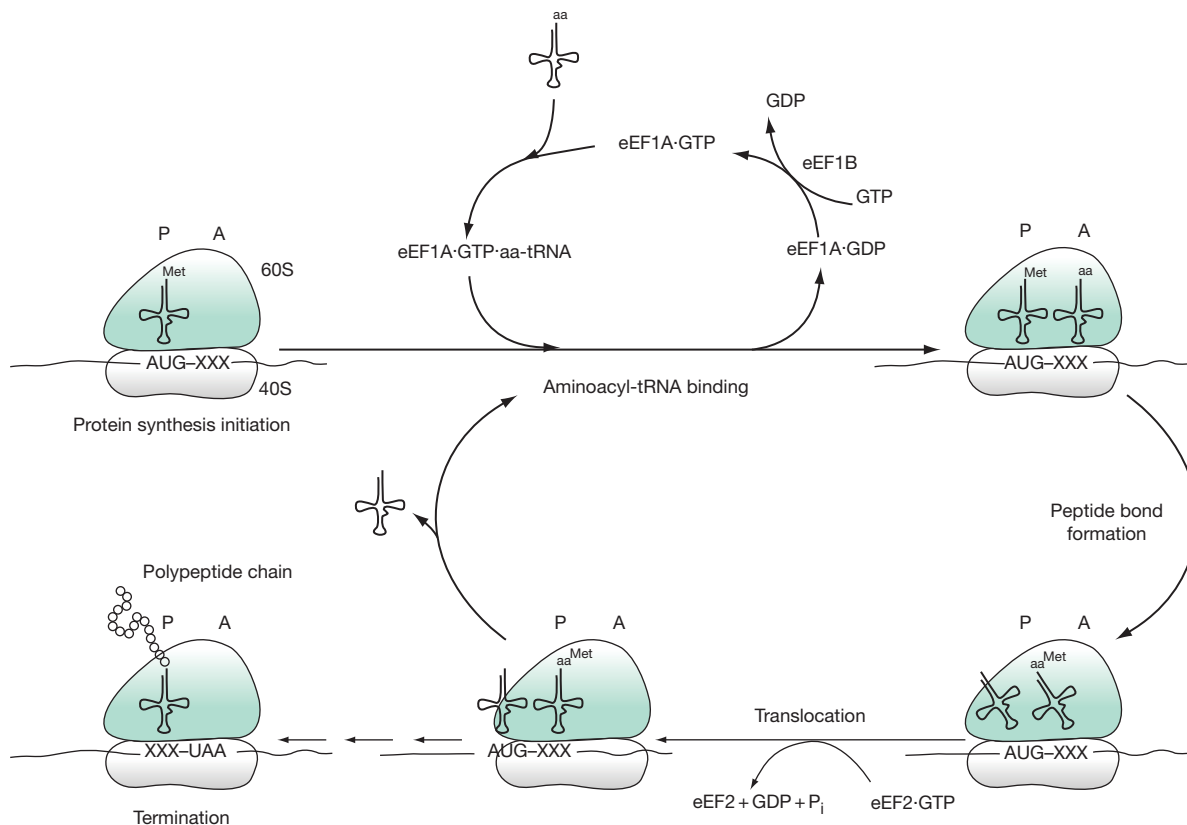


Figure 1 The traditional elongation cycle: the A and P sites. Shown above is the elongation cycle with the traditional A and P sites (aminoacyl and peptidyl sites, respectively). Following initiation, subsequent steps in the elongation cycle ensue with aminoacyl-tRNA binding, peptide bond formation, and then translocation. This process continues until a stop codon (UAA, UAG, or UGA) appears in the A site at which point termination occurs.

subunits, eEF1B α , eEF1B γ , and eEF1B β , in mammals. The catalytic domain necessary for nucleotide exchange is located on the eEF1B α subunit. Interaction of eEF1A-GDP with eEF1B leads to the release of GDP and the formation of an eEF1A-1B complex. Subsequent binding of GTP yields eEF1A-GTP + eEF1B. By contrast, eEF2 does not require an exchange factor as it has a much poorer affinity for guanosine diphosphate (GDP) and thus GDP is rapidly released from eEF2 following GTP hydrolysis and subsequent dissociation from the ribosome.

There are a few points regarding the thermodynamics and kinetics of the elongation cycle. First, the energy cost would appear to be 2 high-energy phosphates per cycle; however, the actual value is 4 when one considers that 2 high-energy phosphates were required for the attachment of the amino acid to the tRNA (i.e., charging) that was then placed in the A site. Second, on careful examination, it becomes apparent that the synthesis of the peptide/protein is from the amino terminus to the carboxy terminus of the growing polypeptide, while the mRNA codons are read in a 5'-3' direction. It should also be noted that movement of the ribosome on the mRNA is not a uniform process. Translation rates of individual codons might differ substantially (5- to 25-fold) and mainly depend on the concentration(s) (which appear to be different) of cognate

aa-tRNAs surrounding the ribosome during translation. It is assumed that aa-tRNAs (in complex with eEF1A and GTP) randomly encounter the A site and become temporally trapped. Trapped species are then tested for proper pairing with the corresponding codon in the A site (i.e., decoded). If the pairing is correct, the transpeptidation reaction becomes irreversible; if not, the ternary complex (eEF1A-GTP-aa-tRNA) is then rejected and another complex is processed in turn.

Elongation Cycle Chemistry

Shown in **Figure 2** is the chemistry that occurs with peptide bond formation. Although most of the details described below were elaborated for the prokaryotic system, it is believed that the general mechanism is the same for all the kingdoms of life. The central event in peptide bond formation is the nucleophilic attack of the amino group of the second amino acid (R2) on the ester carbon of the initiating amino acid methionine (R1), or the amino acid of the peptidyl-tRNA in the P site, in case the ribosome already moved down the mRNA by one or more codons (**Figure 2(a)**). The ribosome accelerates this reaction approximately $\sim 10^5$ -fold. It was originally thought that

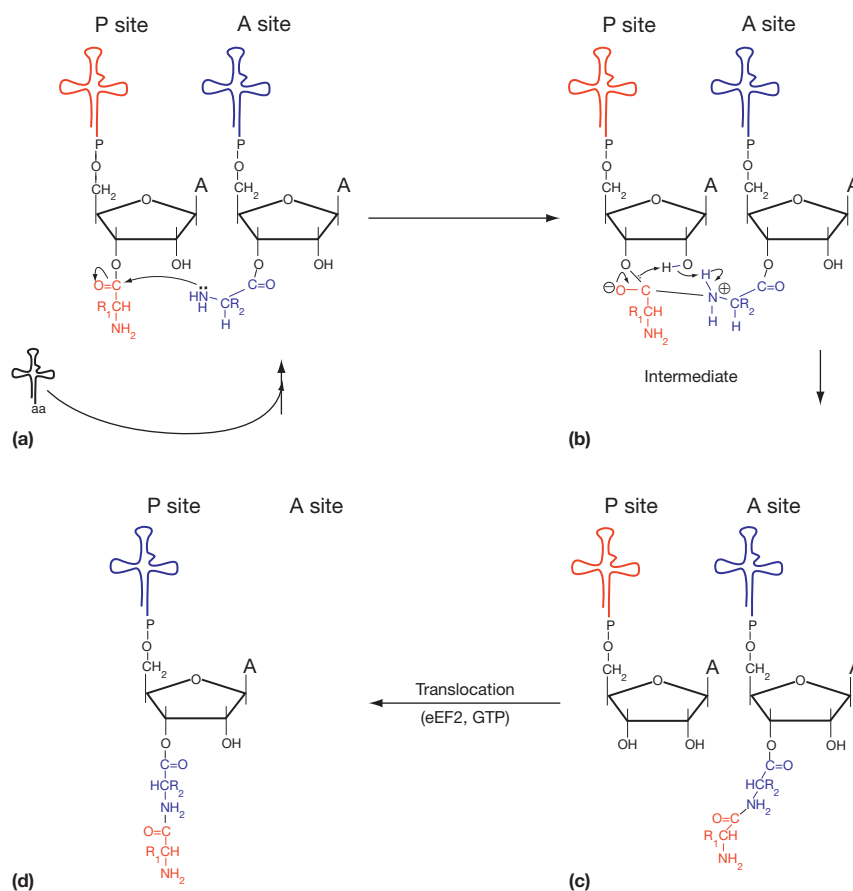


Figure 2 Elongation cycle chemistry. Shown above is the chemistry thought to occur during the process of elongation. Nucleophilic attack by the amino group in the A site on the carbonyl bond of the amino acid in the P site leads to an amino-tetrahedral intermediate (b). The primary role of the ribosome in peptide bond formation is to facilitate the forward reaction by the use of proton-shuttling mechanism. The chemistry is completed by subsequent cleavage of the ester bond to the tRNA in the P site (c). The cycle is further followed by translocation (d). Binding of a new aminoacyl-tRNA (as an eEF1A-GTP-aminoacyl-tRNA complex) starts the beginning of a new cycle (a).

the peptide bond formation on the ribosome involves an acid/base mechanism of catalysis; however, recent experiments clearly indicate that this is not the case and there is no general acid/base catalysis involving any of the ribosomal counterparts. Instead, a proton-shuttling mechanism was proposed, whereby the 2' OH of A76 of the peptidyl-tRNA simultaneously accepts a proton from the α -amino group and donates one to the 3' O⁻ leaving group (Figure 2(b)). An extensive network of hydrogen bonds has therefore been proposed to allow proper positioning of the substrates for reaction. The bond between the 3' oxygen of the terminal adenosine in the tRNA in the P site and the methionine is then cleaved, completing the transfer of the methionine to the amino group of the second amino acid with formation of the peptide bond (Figure 2(c)). Finally, with eEF2 catalyzed translocation, the peptidyl-tRNA is moved completely into the P site (Figure 2). In the overall thermodynamics of the reaction, the resulting peptide bond is of low energy while the original aminoacyl linkage is high energy (equivalent to that in adenosine triphosphate (ATP)) and, thus, the net ΔG for peptide bond formation is quite negative (i.e., extensive heat is released with the completion of the reaction). Although most primary amines would be expected to have a pK_a of about 9–10, that of the α amino groups is about 7.5 due to the inductive effects of the ester linkage of the amino acid to the tRNA (in a similar manner, the pK_a of the amino group at the amino terminus of the peptide/protein is also about 7.5). Thus, at physiologic pH values, the nucleophilic amino group is positively charged only about 50% of the time. A reactive N3 amine of the conserved nucleotide A2451 in ribosomal RNA, which is in close proximity to the CCA ends of aminoacyl and peptidyl-tRNAs, was originally suggested to be indispensable for the general acid/base mechanism of catalysis. However, as indicated above, recent evidence proved its dispensability. Thus, the exact details of the catalytic mechanism are not fully known and the ribosome is suggested to work mainly by properly orienting the substrates and providing an environment that reduces the free energy of formation of the polar transition state.

Molecular Mimicry

High-resolution crystal structures of prokaryotic ribosomes as well as cryo-EM structures of eukaryotic ribosomes have contributed a great deal to our understanding of the mechanism of protein synthesis. The structure of individual elongation factors, alone and in complex with the ribosome, has also been solved providing further insights into the molecular basis of the protein synthesis. The crystal structures of eukaryotic eEF1A and eEF2 revealed high similarity to their bacterial homologs. Comparison of the structures to each other and to the bacterial proteins indicated that to a first approximation they have a similar size, shape, and electronic distribution. While this might have been expected for the GTP-binding domain of EF2, EF2 has a unique protein finger that appears to be the mimic of the anticodon stem and loop of a tRNA by overall size and charge (rich in aspartic and glutamic acid residues); this initially was not expected for the other factor(s). However, subsequently it has been found that other factors

(including termination/release factors) that are expected to encounter and bind the ribosome at the A site have very similar shapes overall approaching the shape of tRNA. The ability of a protein to take on the apparent characteristics of RNA is known as 'molecular mimicry'. Although not discussed here, it has been further proposed that all other factors that bind to this region of the ribosome are also molecular mimics and will yield structures similar to that (at least in some parts) of the L-shaped tRNA molecule.

Influence of Other Sites

In the discussion above, reference was made to an (exit) E site (see also Figure 3). This site, to the 5' side of the P site, has been visualized by X-ray crystallography (in prokaryotes), cryo-electron microscopy (in both pro- and eukaryotes) and inferred by nuclease footprinting. The most important effect of the E site is that it appears to increase the stringency for the correct matching of the codon and anticodon in the A site of the ribosome. It should be noted that for the original binding of the first aa-tRNA, Met-tRNA_i, the E site was unoccupied. One might anticipate that this would allow for less stringent recognition of the initiating AUG codon; however, this strict recognition appears to result from the unique set of factors associated with the initiation process. A second proposed site is the F site or entry site which is just 3' of the A site. This may be an initial test-binding site (check of codon/anticodon match) such that correctly matching complexes would be pulled into the A site while incorrectly matching complexes would dissociate. The advantage of an F site would be that it would be more accessible to ternary complexes of eEF1A-GTP-aa-tRNA as both the A site and the P site appear to be partially occluded at the interface of the large and small ribosomal subunits. However, the biochemical evidence supporting the existence of the F site is weak compared to that for the E site.

Assuming that all four of the ribosomal sites are functional and exist, they can be characterized by positions of the tRNAs on either the large or small ribosomal subunits. Thus, following initiation, the Met-tRNA_i may be described as being in the P/P position (both the aminoacyl end and the anticodon end of the tRNA correspond to the P site; see Figure 3(a) or 3(c)). In the next step, the initial binding of the aa-tRNA is to the F/A site where the anticodon is in the A site, but the aminoacyl end is in the F site and this end of the aa-tRNA is still bound to eEF1A-GTP (Figure 3(d)). With correct codon/anticodon recognition, GTP is hydrolyzed and eEF1A-GDP is released from the ribosome concomitant with the movement of the aminoacyl end of the tRNA into the A site (now in the A/A configuration, Figure 3(a)). At the same time, any tRNA in the E site (as would be true for most elongation steps) would be released from the ribosome. Subsequent reactions would cyclically yield the other states of the ribosome (P/P A/A, A/P, P/E, and then E/E P/P). Additional intermediate states have also been suggested, but are not discussed here for simplicity and because the states noted above have been suggested to represent the major configurations. 'Snapshots' of the translocation cycle as determined by cryo-electron microscopy (some for the yeast *Saccharomyces cerevisiae* 80S ribosome, and others for *Escherichia coli* 70S complexes) are shown in Figure 4.

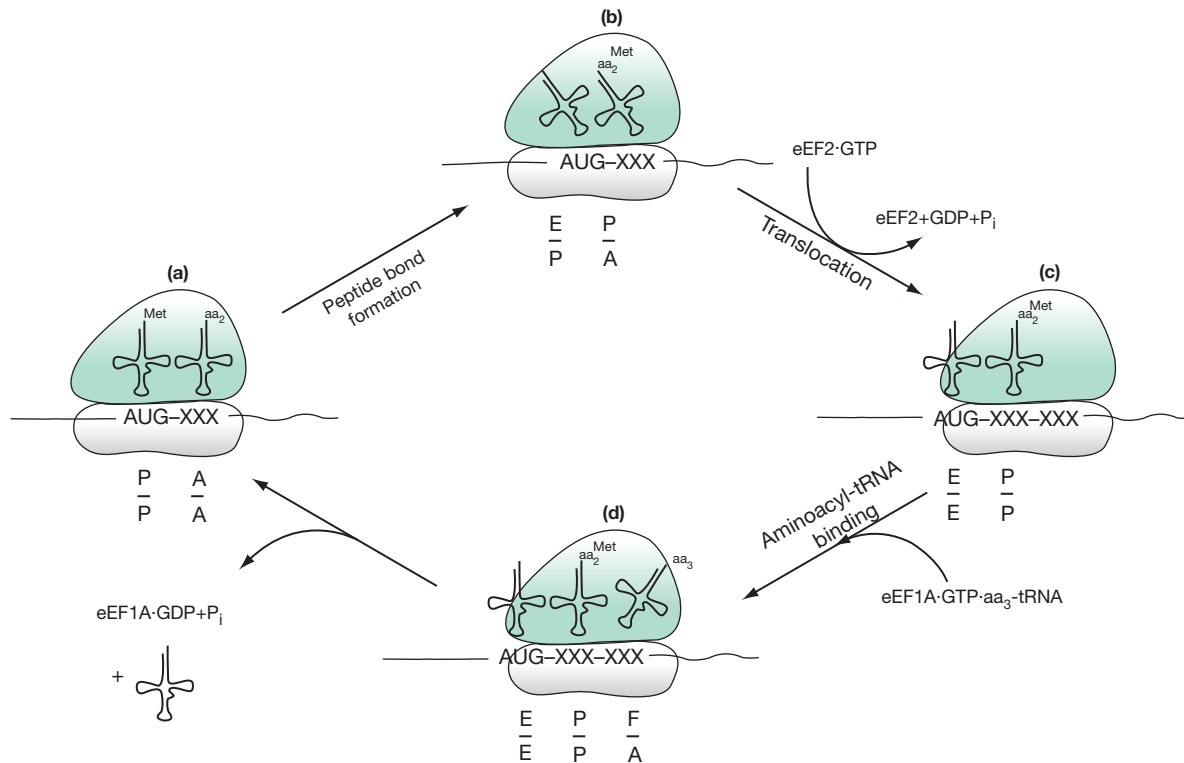


Figure 3 Elongation cycle: influence of other sites. Illustrated in this figure are the four proposed tRNA-binding sites in the eukaryotic ribosome. The A and P sites are the traditional tRNA binding sites on the ribosome (see Figure 1). The E site (exit site) and the F site (entry site) are also shown. Note that a portion of each site is located on the small (40S) and the large (60S) ribosomal subunit. In the designation for the tRNA molecules, the upper letter indicates position on the 60S subunit and the lower letter indicates position on the 40S subunit.

The Other Elongation Factors, eEF3, eEFSec

Protein synthesis in yeast and fungi requires an additional unique translation elongation factor, eEF3. This protein, which is an ATPase, contains two nucleotide-binding sites, and appears to be required for the nucleotide-dependent release of the non-acylated tRNA from the ribosomal E site. As this protein is an essential gene product in yeast, it is surprising that an equivalent activity has not been identified in higher eukaryotes. However, it has been noted *in vitro* that only elongation reactions using yeast ribosomes demonstrate the eEF3 requirement and thus, this requirement for eEF3 would appear to reflect unusual properties of the yeast and fungal ribosome compared to other eukaryotic ribosomes. Recently, the crystal structure of the yeast *S. cerevisiae* eEF3 was solved and the position of the factor on the surface of the 80S ribosome was mapped using cryo-EM. In contrast to other elongation factors, eEF3 was found to occupy an unusual place on the 80S ribosome surface in close proximity to the E site. Although all other elongation factors bind 80S ribosome at the A site, this unusual location of eEF3 coincides well with its function.

A 21st amino acid (i.e., Selenocysteine (Sec)) is incorporated at certain UGA codons in mRNAs and requires a specific translation elongation factor (eEFSec), a functional homolog of eEF1A, capable of recognizing the unique Sec-tRNA^{Sec} (anticodon UCA)

and delivering the Sec-tRNA^{Sec} to the ribosome in a ternary complex with GTP. In eukaryotes, the incorporation of selenocysteine also requires a SECIS-binding protein 2 (SBP2) capable of recognizing both the sec insertion sequence (SECIS) element in the 3'-untranslated region of mRNAs and eEFSec. In prokaryotes, the recognition of SECIS element and Sec-tRNA^{Sec} delivery is performed solely by elongation factor SelB. The requirement for an additional protein SBP2 in eukaryotes for incorporation of the Sec-tRNA^{Sec} has been attributed to the difference in the structural organization of the eukaryotic and prokaryotic proteins, respectively. Prokaryotic SelB is a multidomain protein consisting of an N-terminal RNA-binding domain and a C-terminal eEF1A-like domain. Eukaryotic SelB contains only the eEF1A-like domain and lacks the RNA-binding domain and, therefore, relies on the additional SelB-binding protein, SBP2, for its interaction with mRNA. The crystal structure of the archaeal SelB has been recently determined, confirming its similarity to other elongation factors in the family.

Regulation of the Elongation Cycle

The major regulation of gene expression at the translational level occurs through the covalent modification of translation initiation factors or regulatory proteins that bind to the

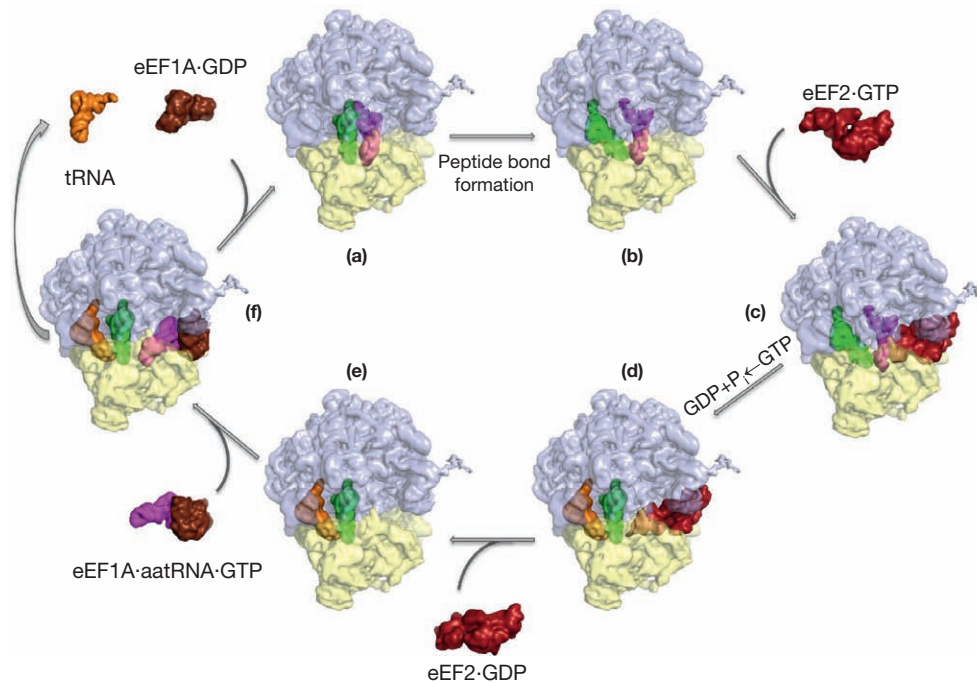


Figure 4 Structural models of the translocation cycle. The structure of the yeast 80S ribosome, as determined by cryo-electron microscopy, is shown with the 60S in blue and the 40S in yellow. (a) The ribosome contains A and P site tRNAs shown in magenta and green, respectively. (b) After peptide-bond formation, the upper half of the tRNAs migrate to the next binding site on the 60S subunit, while the lower halves remain stationary on the 40S subunit. This results in hybrid-state A/P- and P/E-site tRNAs. (c) eEF2 binds the ribosome in its GTP-bound conformation. (d) Binding of eEF2 to the ribosome catalyzes GTP hydrolysis, which assists in translocating the lower portions of the tRNA molecule to the next site on the 40S subunit. (e) Dissociation of eEF2 in its GDP-bound conformation leaves the ribosome with P and E site tRNAs and a vacant A site. (f) eEF1A bound to GTP delivers the next aa-tRNA to the A site, which is decoded by the ribosome. GTP-hydrolysis, again induced by binding to the ribosome, releases eEF1A from aa-tRNA and the ribosome. Occupation of the newest tRNA in the A site completes the cycle.

initiation factors. However, there is ample evidence that regulation of translation also occurs at the level of elongation although the degree of regulation (the fold change in the elongation rate) is not as great as seen with regulation of the initiation step. While eEF1A, eEF1B, and eEF2 are all phosphorylated under different conditions, eEF1A and eEF2 also contain unique posttranslational modifications. eEF1A contains methylated lysines (mono-, di-, or trimethyl lysine) and glycerylphosphorylethanolamine. eEF2 contains a hypermodified histidine residue (histidine #714 in mammalian eEF2's) that is found in all eukaryotic eEF2's and is called diphthamide. The diphthamide residue is a known inactivation site for ADP-ribosylation catalyzed by either *Diphtheria* or *Pseudomonas* A toxins with nicotinamide adenine dinucleotide (NAD) serving as the donor of the adenosine diphosphate (ADP) group. There is tentative evidence that this modification may be part of the normal cellular regulation of eEF2 activity as well.

The best-studied regulation of any of the elongation factors is via phosphorylation. In general, phosphorylation of eEF1A and eEF1B correlates well with increases in the rate of elongation noted *in vivo* with either insulin or phorbol ester treatment. Additionally, most phosphorylations of translation factors are associated with an increased rate of protein synthesis. By contrast, the phosphorylation of eEF2 leads to its inactivation. Under most circumstances, changes in the elongation

rate due to changes in covalent modifications of eEF1A, eEF1B, or eEF2 are associated with a coordinate change in the rate of initiation of protein biosynthesis.

The Ribosome and the Nobel Prize

The 2009 Nobel prize in chemistry was awarded to Drs. Venkatraman Ramakrishnan, Thomas A Steitz, and Ada E Yonath for the 'studies of the structure and function of the ribosome'. The work is undoubtedly unique, as for the first time the structure of such large, highly complex, and asymmetrical particles, consisting of both proteins and RNA has been solved by X-ray crystallographic analysis. The Nobel-winning crystal structures of the prokaryotic ribosomes clarified many aspects of the ribosome work and provided answers to many long-standing questions. In particular, analysis of codon-anticodon interactions within the ribosome provided new insights into codon reading and showed that the ribosome is a thermodynamic machine strengthening codon/anticodon interactions and thus providing stabilization above the ΔG of simple base pairing. Tantalizing efforts to solve the crystal structure of eukaryotic ribosomes are bringing first results and will undoubtedly provide additional insights into the mechanism of eukaryotic protein synthesis and its regulation.

See also: **Molecular Biology:** Ribosome Structure; Translation Elongation in Bacteria; Translation Initiation in Bacteria: Factors and Mechanisms; Translation Initiation in Eukaryotes: Factors and Mechanisms.

Further Reading

- Ban N, Nissen P, Hansen J, et al. (1999) Placement of a protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit. *Nature* 400: 841–847.
- Burkhardt N, Junemann R, Spahn CMT, and Nierhaus KH (1998) Ribosomal tRNA binding sites: Three-site models of translation. *Critical Reviews in Biochemistry and Molecular Biology* 33: 95–149.
- Clark BFC, Grunberg-Manago M, Gupta N, et al. (1996) Prokaryotic and eukaryotic translation factors. *Biochimie* 78: 1119–1122.
- Clemons WM, Jr., May JLC, Wimberly BT, et al. (1999) Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature* 400: 833–840.
- Frank J, Gao H, Sengupta J, Gao N, and Taylor DJ (2007) The process of mRNA–tRNA translocation. *Proceedings of the National Academy of Sciences of the United States of America* 104: 19671–19678.
- Jørgensen R, Merrill AR, and Andersen GR (2006) The life and death of translation elongation factor 2. *Biochemical Society Transactions* 34(Pt 1): 1–6.
- Merrick WC and Nyborg J (2000) The protein biosynthesis elongation cycle. In: Sonenberg N, Hershey JWB, and Mathews MB (eds.) *Translational Control of Gene Expression*, pp. 89–125. New York, NY: Cold Spring Harbor Laboratory.
- Moazed D and Noller HF (1989) Intermediate states in the movement of tRNA in the ribosome. *Nature* 342: 142–148.
- Nissen P, Kjeldgaard M, Thirup S, et al. (1995) Crystal structure of the ternary complex of Phe-tRNA^{Phe}, EF-Tu, and a GTP analog. *Science* 270: 1464–1472.
- Noble CG and Song H (2008) Structural studies of elongation and release factors. *Cellular and Molecular Life Sciences* 65: 1335–1346.
- Proud C (2000) Control of the elongation phase of protein synthesis. In: Sonenberg N, Hershey JWB, and Mathews MB (eds.) *Translational Control of Gene Expression*, pp. 719–739. New York, NY: Cold Spring Harbor Laboratory.
- Rodnina MV and Wintermeyer W (2001) Fidelity of aminoacyl-tRNA selection on the ribosome: Kinetic and structural mechanisms. *Annual Review of Biochemistry* 70: 415–435.
- Rodnina MV and Wintermeyer W (2009) Recent mechanistic insights into eukaryotic ribosomes. *Current Opinion in Cell Biology* 21: 435–443.
- Schmeing TM and Ramakrishnan V (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461: 1234–1242.
- Stark H (2002) Three-dimensional electron cryomicroscopy of ribosomes. *Current Protein and Peptide Science* 3: 79–91.

Evolving Concepts of Leptin

H Münzberg and T W Gettys, Pennington Biomedical Research Center, Baton Rouge, LA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by T. W. Gettys, volume 2, pp. 541–545, © 2004, Elsevier Inc.

Glossary

Brown adipocyte Adipose cell type with a large number of mitochondria whose primary function is to conduct fat oxidation that is uncoupled from adenosine triphosphate synthesis via mitochondrial protein, uncoupling protein 1 (UCP1).

Energy homeostasis Physiological state when energy intake and energy utilization are matched to produce stable body weight.

Leptin A 16-kDa hormone secreted by adipose tissue that regulates the balance between food intake and energy utilization.

Obesity Excessive fat storage resulting from chronic positive energy balance.

Sympathomimetic Endogenous or synthetic agonist of adrenoceptors.

Transdifferentiation A programmatic change in gene expression in white adipocytes which transforms their morphology and functionality from fat storage to fat oxidation.

Uncoupling protein 1 Specialized mitochondrial protein that short-circuits the proton gradient across the mitochondrial matrix, uncouples oxidative phosphorylation, and transforms electrochemical energy into heat energy.

Leptin is a 16-kDa peptide hormone released from adipose tissue that regulates the balance between energy intake and energy utilization by binding to hypothalamic receptors and modulating appetite control centers in the brain. Occupancy of central leptin receptors also increases activity of the sympathetic nervous system, which stimulates energy expenditure in peripheral target tissues through a process called 'adaptive thermogenesis'. Through coordinated regulation of these two systems, leptin functions to match rates of energy expenditure with energy intake. Leptin also functions to preserve energy reserves during starvation by lowering body temperature, basal metabolic rate, and limiting hypothalamic support for reproductive function. As such, leptin is a peripheral signal that acts centrally to integrate a complex array of behavioral and metabolic systems that function to maintain energy homeostasis when food supplies are adequate and ensure survival when they are not.

Historical Perspective of Leptin Discovery

The search for the primary genetic defect giving rise to the obese/diabetic syndrome of *ob/ob* mice (Figure 1) was guided by the early work of Douglas Coleman, who postulated that *ob/ob* mice were missing a circulating satiety signal. To test this hypothesis, he surgically joined *ob/ob* mice with lean littermates, *ob/ob* mice with *db/db* (diabetes) mice, and lean mice with *db/db* mice so that the parabiotic pairs shared blood supplies. The *db/db* mice were included in his experiments because they also developed morbid obesity after weaning but differed from *ob/ob* mice in that the *db/db* mutation resides on a different chromosome. This landmark experiment showed that the obesity syndrome of *ob/ob* mice, but not *db/db* mice, was corrected by a circulating factor from lean mice. The obesity syndrome of *ob/ob* mice was also reversed by parabiotic

pairing with *db/db* mice, but, in this case, the *ob/ob* mouse eventually died from starvation. The results indicate that *ob/ob* mice were missing a satiety factor but retained the ability to respond to the missing factor. By contrast, *db/db* mice were unresponsive to the satiety factor, but apparently produced it in abundant quantities. Nearly 30 years later, the missing factor (and receptor) inferred from the work of Coleman was identified by Jeffrey Friedman at Rockefeller University with the cloning of the *ob* gene (now officially called '*lep*'), in late 1994. The cloning of the *ob* receptor gene (now officially called '*lepr*') by Louis Tartaglia followed soon thereafter. The product of the *ob* gene was named leptin, from the Greek leptos, meaning thin.

The Obesity Syndrome Produced by Leptin Deficiency in *ob/ob* Mice

The cloning of the *ob* gene identified leptin as a lipostat, which regulates the balance between energy intake and utilization to maintain body weight homeostasis. Adipocytes of *ob/ob* mice do not produce leptin, and its absence produces a complex metabolic syndrome characterized by hyperphagia, morbid obesity, and diabetes. Hypertrophy of adipose tissue is noted prior to the hyperphagia that develops after weaning in *ob/ob* mice, and obesity occurs even when *ob/ob* mice are pair-fed with lean littermates. Thus, their thriftiness is a significant component of their propensity to become obese. Treatment of *ob/ob* mice with exogenous leptin corrects their obesity, and part of the weight loss can be accounted for by reductions in food intake. However, even after controlling for differences in food intake between saline and leptin-injected *ob/ob* mice, leptin-injected mice still lose more weight. These results illustrate that an important component of leptin's physiological actions is to regulate the metabolic efficiency of fuel oxidation.

Central Mechanisms of Leptin Action

Leptin is a key component of a homeostatic system that functions to match energy intake with energy expenditure. The adipocyte-derived hormone produces these effects through coordinated but reciprocal regulation of food intake and energy expenditure in peripheral tissues, and both sides of this balance are regulated via the neuroendocrine axis and autonomic nervous system. Neurons expressing the leptin receptor are found in many areas of the brain, and the current consensus is that integration of all signals from these neurons is necessary for the full effect of leptin on energy homeostasis (Figure 2).

The *ob* receptor is encoded by the *lepr* gene, which is translated into six distinguishable receptor isoforms (LepRa–f) by alternative splicing, but only LepRb – also referred to as the ‘long LepRb’ isoform – consists of a long intracellular signaling domain. LepRb is the main receptor that mediates leptin action, demonstrated in *db/db* mice. *Db/db* mice lack the long intracellular LepRb domain and retain normal expression of other LepR isoforms, which is sufficient to recapitulate the phenotype of leptin deficient *ob/ob* mice. LepRb is expressed

throughout the central nervous system and even though there is evidence for the existence of peripheral leptin targets, central leptin action adequately imitates phenotypes observed in *ob/ob* and *db/db* mice.

Leptin action in the hypothalamic arcuate nucleus (ARC) has been particularly well studied with the focus on two distinct neuronal populations: one population co-expressing the anorexigenic neuropeptides pro-opiomelanocortin (POMC) and cocaine and amphetamine-regulated transcript (CART) and another population co-expressing the orexigenic neuropeptides agouti-related peptide (AgRP) and neuropeptide Y (NPY). Leptin induces the transcription of CART and POMC messenger RNA (mRNA) and stimulates POMC/CART neuronal activity to increase the release of CART and the POMC cleavage product α -melanocyte stimulating hormone (α -MSH). By contrast, leptin decreases AgRP and NPY transcripts and inhibits NPY/AgRP neuronal activity to decrease their neuropeptide release. Thus, leptin differentially acts on these inversely acting neuronal populations.

Both neuronal populations have been well documented to project into the paraventricular hypothalamus – a major neuroendocrine output site – as well as to preganglionic neurons of the sympathetic nervous system, where the differential action of leptin on POMC/CART and NPY/AgRP neurons activates melanocortin and CART receptors to decreased food intake and increased sympathetic outflow. The resulting increase in sympathetic stimulation of liver, muscle, and adipose tissue promotes fat oxidation through targeted effects on futile energy cycles in the respective tissues. The effector systems in target tissues involve signaling through both β -adrenoceptors and adenosine monophosphate (AMP) kinase to produce an integrated increase in fuel oxidation. Thus, it is through reciprocal regulation of several leptin-responsive neurons that leptin produces its full effect on energy homeostasis (see Figure 2).

In addition to LepRb neurons in the ARC, a large network of LepRb-expressing neurons is found in several central sites. LepRb neurons are aggregated into populations that outline several hypothalamic nuclei as well as midbrain and brainstem structures (Figure 3) and support an overall role of leptin in



Figure 1 *ob/ob* mouse and lean littermate (+/?) 12 weeks after weaning.

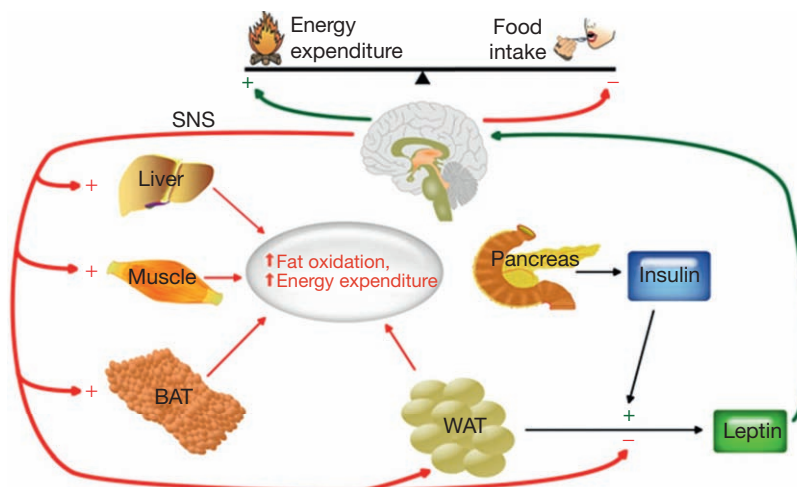


Figure 2 Integrated model of leptin synthesis, secretion, and function. Leptin regulates its own expression and secretion from white adipose tissue and peripheral fatty acid oxidation through central modulation of sympathetic outflow to target tissues.

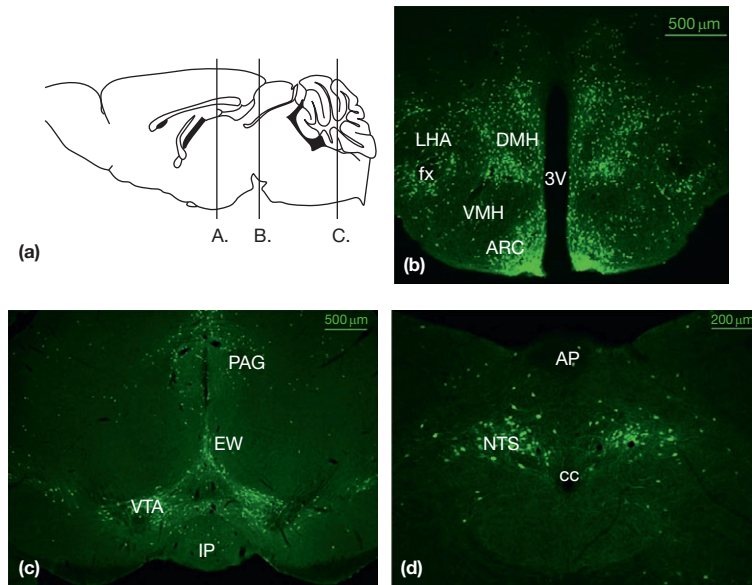


Figure 3 The schematic drawing of a mouse brain (a) shows the relative brain levels of fluorescent images depicted from coronal sections of a reporter mouse with green fluorescent protein expression in LepRb neurons. Examples of LepRb-expressing sites in the hypothalamus (b), midbrain (c), and brainstem (d) are shown to demonstrate the outlining of specific LepRb populations. LHA, lateral hypothalamic area; DMH, dorsomedial hypothalamus; 3V, 3rd ventricle; VMH, ventromedial hypothalamus; ARC, arcuate nucleus; PAG, periaqueductal gray; EW, edinger westphal nucleus; VTA, ventral tegmental area; IP, interpeduncular nucleus; AP, area postrema; NTS, nucleus of the solitary tract; cc, central canal.

neuroendocrine, autonomic, and limbic functions. Indeed, in addition to leptin's obvious role in the control of food intake and energy expenditure, other physiologic functions have been well documented, including reproduction, thyroid and immune function, glucose homeostasis, and growth.

However, the specific physiological function of leptin in these distinct LepRb populations outside the ARC is largely unknown. Even though the ARC is an important contributor of leptin action, recent studies using conditional deletion of LepRb in POMC and/or AgRP neurons showed a surprisingly moderate effect on body weight. Other central leptin targets in the hypothalamus (ventromedial hypothalamus, lateral hypothalamic area), midbrain (ventral tegmental area), and the brainstem (nucleus of the solitary tract) have been found to contribute to leptin action on body weight regulation. By contrast, LepRb expressing neurons in the hypothalamic pre-mammillary nucleus have been mainly associated with reproductive function. Further studies are under way to dissect the neuronal circuits and physiologic function of distinct LepRb populations.

Adipose Tissue as a Target of the Sympathetic Nervous System

Mitochondrial oxidation of fatty acids creates a proton electrochemical gradient that is used to drive the conversion of adenosine diphosphate (ADP) to adenosine triphosphate (ATP) via ATP synthase. Brown adipose tissue (BAT) mitochondria possess an alternative pathway that allows protons to re-enter the mitochondrial matrix without coupling to ATP synthesis, and this pathway is mediated by uncoupling protein 1 (UCP1). UCP1 transforms electrochemical energy into heat, enabling

small mammals to survive exposure to cold. The sympathetic nervous system and UCP1 are essential components of this thermogenic response system. Brown adipocytes are also unique in the sense that they are capable of simultaneously oxidizing glucose and fatty acids and conducting *de novo* lipogenesis. In liver and muscle, increased glycolytic flux increases formation of the substrate malonyl-coenzyme (malonyl-CoA) needed for the first committed step of *de novo* lipogenesis. In these tissues, the formation of malonyl-CoA potentially inhibits fatty acid oxidation by inhibiting carnitine palmitoyltransferase-1 β (CPT-1 β), which transports fatty acids into mitochondria for oxidation. These functions appear to be compartmentalized in brown adipocytes, so that increased glucose uptake and oxidation during cold exposure does not inhibit the high rate of fatty oxidation needed to support the thermogenic response to cold exposure. Maximal thermogenesis requires an adequate pool of long-chain fatty acids, which serve to stimulate their own uptake and oxidation within the mitochondria. However, increased cytosolic long-chain fatty acids inhibit fatty acid synthesis in liver and muscle, but this is not seen in BAT during cold exposure, where a high rate of *de novo* lipogenesis is required to support the high rate of mitochondrial fat oxidation. Although the mechanistic details are not yet firmly established, it is possible that high flux rates of inhibitory molecules (e.g., malonyl-CoA and long-chain acyl CoAs) mitigate their inhibitory effects on pathways that are normally mutually exclusive. The extent to which this paradoxical biochemistry occurs in physiological contexts other than cold exposure is also not known. However, it is clear that the simultaneous conduct of glucose oxidation, fat oxidation, and *de novo* lipogenesis constitutes an energetically inefficient substrate cycle that would waste energy, produce heat, and reduce the lipid content of tissues.

The conventional wisdom has been that UCP1 expression is restricted to brown adipocytes, but recent evidence has clearly established that sympathomimetics can induce ectopic expression of UCP1 in white adipose tissue (WAT) depots. The origin of the UCP1-expressing cells was originally a mystery, but evidence has emerged to support the concept that mature adipocytes are capable of significant functional plasticity, and readily transdifferentiate between cell types whose primary function is to store triglyceride (WAT) and multilocular, mitochondria-rich adipocytes capable of significant fat oxidation (BAT). It is clear that intense sympathetic nervous system (SNS) stimulation can also promote the differentiation of preadipocytes within white depot sites into brown adipocytes. The unexpected appearance of UCP1 in WAT was originally observed in mice, where cold exposure produced the emergence of islands of multilocular, mitochondria-rich brown adipocytes within WAT depots primarily composed of unilocular white adipocytes. Cold exposure produces these responses through increased sympathetic outflow to adipose tissue and induces the transcriptional program of differentiation or transdifferentiation within WAT depots through a norepinephrine, β -adrenoceptor, and cyclic adenosine monophosphate (cAMP)-dependent mechanism. SNS stimulation of adipose tissue is also subject to regulation by leptin, and in this case, modulation of SNS outflow has the effect of increasing thermogenic potential of both BAT and WAT.

The potential for emergence of brown adipocytes within WAT is depot-site specific and has a strong genetic component in both mice and humans. The significance of this process was demonstrated in elegant studies analyzing backcrosses of parental mouse strains with widely varying ability of WAT to form brown adipocytes. Using UCP1 as a BAT-specific indicator, Leslie Kozak showed that recombinant inbred strains from the crosses reduced fat deposition after adrenergic stimulation in direct proportion to their induction of UCP1 in convertible WAT depots. These findings indicated that the relative ability of a mouse strain to resist obesity is a function of the ability of WAT to increase UCP1 in response to adrenergic stimulation. Thus, an important component of leptin's ability to regulate energy utilization in peripheral tissues is conferred by its ability to induce UCP1 expression and enhance fat oxidation in adipose tissue. Some WAT depots do not respond to SNS stimulation with increases in UCP1 expression and fat oxidation. In this case, SNS stimulation activates lipolysis and mobilizes lipid for transport to more oxidatively active sites.

Regulation of Leptin Expression

Well before the discovery of leptin, George Bray proposed that most obesities known are low in sympathetic activity (MONA-LISA hypothesis). This hypothesis stemmed in part from the observation that *ob/ob* mice were cold intolerant, but after leptin and its receptor were cloned, it was noted that leptin mRNA was significantly upregulated in adipose tissue from both models. This indicated the absence of an efferent signal from the SNS in both models which was important in regulating leptin expression. Circulating leptin concentrations are the product of transcriptional regulation of the gene and release of

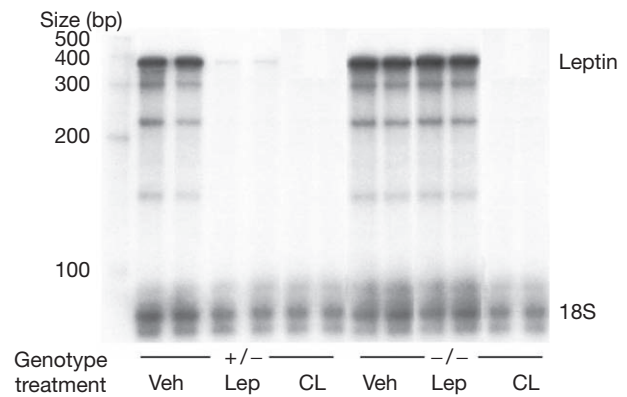


Figure 4 Ribonuclease protection assay of leptin mRNA in total RNA from EWAT of control (*Dbh*^{+/+}) and dopamine β -hydroxylase (*Dbh*^{-/-}) knockout mice injected with either saline, leptin, or the β 3-adrenergic receptor agonist, CL-316,243 for 3 days. The blot shows that the ability to synthesize catecholamines is required for leptin to downregulate its own expression in white adipose tissue. Reproduced from Commins SP, Marsh DJ, Thomas SA, et al. (1999) Norepinephrine is required for leptin effects on gene expression in brown and white adipose tissue. *Endocrinology* 140: 4772–4776, with permission from Copyright 1999, The Endocrine Society.

the expressed protein from adipocytes. Work in the author's laboratory established that leptin secretion from isolated adipocytes was inhibited by activation of β -adrenoceptors in parallel with the activation of cAMP-dependent protein kinase. As shown in Figure 4, leptin also regulates its own expression via the SNS, as mice lacking the gene which converts dopamine into norepinephrine (dopamine β -hydroxylase) were unable to downregulate leptin mRNA in WAT after treatment with exogenous leptin. Thus, the SNS is the efferent arm of a feedback system which negatively regulates leptin expression and release.

Leptin is positively regulated by insulin and glucocorticoids. Both hormones produce rapid increases in leptin secretion from adipose tissue in conjunction with a coordinated increase in transcription of the gene. These hormonal and neurotransmitter-dependent pathways represent acute regulatory systems that are superimposed onto a diurnal pattern of leptin secretion, which, in humans, produces lowest concentrations at night and highest during the day. Serum leptin levels are also strongly correlated to percentage body fat and body mass index, but the relationship is dichotomous between male and female patients. For instance, as body mass index increases, serum leptin levels increase threefold faster in females than males. Testosterone inhibits leptin expression and likely explains the gender difference in circulating leptin between males and females. Lastly, leptin is also responsive to nutrient-sensing systems that are regulated by meal composition. For example, increases in amino acids and insulin after a meal activate the mammalian target of rapamycin (mTOR) pathway which rapidly increases leptin synthesis and release. Considered together, leptin expression and release are regulated by a complex network of nutritional, hormonal, and autonomic inputs which are integrated to regulate transcription of the leptin gene and secretion of expressed protein into the bloodstream.

Leptin Resistance and Obesity

Diet-induced obesity in mice and humans is commonly associated with insensitivity to leptin, with high-circulating leptin levels and reduced anorexigenic leptin effects, and has been termed 'leptin resistance'. The development of leptin resistance may involve several levels and stages, including leptin transport mechanisms, cellular signaling mechanisms, as well as reorganization of axonal processes.

Leptin is too large a protein to cross the blood–brain barrier by perfusion, and a saturable, regulated transporter is needed for leptin access into the brain. Obese individuals show a reduced transport capacity of serum leptin into the cerebrospinal fluid, suggesting that the brain is exposed to lower leptin levels than available in the serum of obese individuals. This is further supported by the fact that leptin resistance in rodents can be improved (even though not entirely recovered) by central leptin application, thus bypassing the blood–brain barrier and the need for a transporter system and indicating that defective leptin transport depicts one component in the development of leptin resistance.

Leptin resistance is also manifested at the cellular level and a severe reduction in leptin-induced signal transducer and activator of transcription-3 (STAT3) phosphorylation. A deamplification of LepRb signaling can occur on several levels, for example, a decrease in LepRb availability or negative-feedback mechanisms via inhibitory molecules like suppressor of cytokine signaling-3 (SOCS-3) as well as protein tyrosin phosphatase PTP1B. SOCS-3 expression is induced by leptin signaling, and thus high circulating leptin levels may chronically increase SOCS-3 levels resulting in a decreased LepRb signal. PTP1B is also upregulated in response to leptin, resulting in decreased LepRb signaling; even though the specific molecular mechanisms are less well understood. Furthermore, even though leptin is able to induce the expression of SOCS-3 and PTP1B, they both are also regulated by other cytokines that are likely induced in obese subjects. Whole body deletion of either SOCS-3 or PTP1B, as well as neuron-specific deletion of SOCS-3 in POMC neurons, attenuates body weight gain on a high-fat diet and increased leptin sensitivity. Thus, inhibitory molecules such as SOCS-3 and PTP1B in specific neuronal populations like POMC neurons present a further level in the development of leptin resistance (Figure 5).

The concept that leptin induces its own resistance by inducing SOCS-3 and PTP1B has been discussed as a mechanism for increased weight gain on a high-fat diet. Indeed, chronic and central infusions of leptin cause leptin resistance over time. By contrast, low leptin levels have been commonly associated with increased leptin sensitivity, for example, fasted mice or leptin-deficient *ob/ob* mice are known to respond more sensitively to leptin treatment. The evolutionary importance of a system with major anorexigenic leptin effects when body weight is already low does not seem to be favorable. However, to evaluate the capability to survive energy-demanding periods, such as reproduction, a sensitive system to communicate body fat stores would be particularly important during famine and thus when fat stores are low.

Leptin resistance does not affect all physiological end points of leptin action. Whereas leptin resistance clearly blunts anorexigenic leptin effects, the sympathetic tone to the kidney

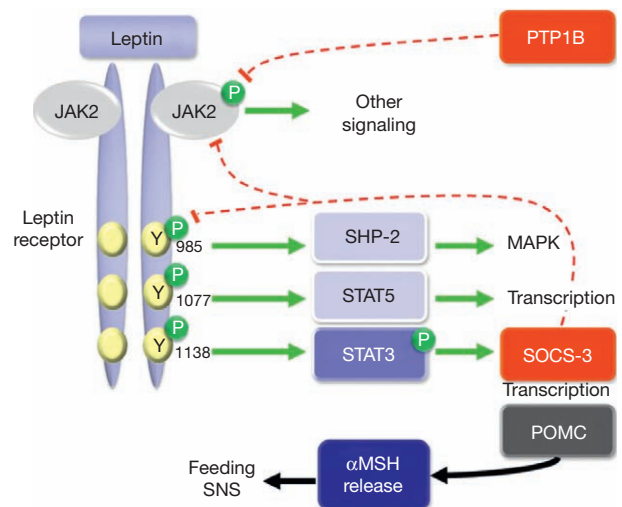


Figure 5 Schematic model of leptin signaling through LepRb, resulting in the activation of several signaling pathways as well as interaction with PTP1B and SOCS-3, which causes the deamplification of leptin signaling.

remains leptin sensitive in some obese mice, resulting in increased blood pressure in obese individuals with high circulating leptin levels. A selective leptin resistance has also been found in diet-induced obese mice and seasonal mammals with upregulated SOCS-3 and decreased leptin-dependent STAT3 phosphorylation specifically in the ARC, but not other hypothalamic sites. Neuronal processes of ARC LepRb neurons can access the circulation directly and respond more sensitively and faster to changes in circulating leptin as well as other obesity-induced cytokines compared to other hypothalamic sites. Therefore, ARC neurons would be more susceptible to increase SOCS-3 levels in response to high circulating leptin levels than other hypothalamic sites and could explain the appearance of ARC-specific leptin resistance. By contrast, pregnancy has been associated with leptin resistance confined to the ventromedial hypothalamus (VMH). Several LepRb neurons in the VMH also express estrogen receptors, so that interactions of the estrogen-signaling pathway with LepRb signaling might provoke VMH-specific leptin resistance during pregnancy, indicating that leptin resistance can be differentially modulated depending on the physiological context.

Leptin resistance could also occur at the level of translation of SNS input in peripheral tissues. In adipose tissue, activation of β -adrenoceptors initiates the transcriptional program of adaptive thermogenesis in a cAMP-dependent manner. Analysis of the promoter structure of many genes transactivated during this response shows that the nuclear receptor, peroxisome proliferator-activated receptor γ (PPAR γ), is an essential component of the transcriptional complex. Although the endogenous ligand for this receptor has yet to be identified, *in vitro* studies have shown that long-chain-polyunsaturated fatty acids and their metabolites can act as partial agonists. The relevance of PPAR γ to adaptive thermogenesis is apparent from work with obesity-prone strains of mice, where it is known that high-fat diets compromise the ability of sympathetic stimulation to initiate thermogenic responses in adipose tissue. Remarkably, simultaneous provision of synthetic PPAR γ

agonists with sympathomimetics rescues the compromised adaptive thermogenic response of obesity-prone mice and produces a significant loss of WAT. Coadministration of PPAR γ agonists with sympathomimetics is not required in obesity-resistant mice, as high-fat diets do not compromise their adaptive thermogenic responses. Although the underlying mechanism is not yet known, high-fat diets accentuate leptin resistance by altering the requirement for endogenous PPAR γ agonists to support transcriptional activation of genes involved in adaptive thermogenesis. The consensus at present is that changes in dietary fat content and composition alter synthesis of the endogenous ligands for PPARs in a manner that produces coordinated, integrated transcriptional responses in and among metabolically active target organs. Collectively, the PPARs are viewed as a lipid-based nutrient-sensing system. In the case of adipose tissue, the productive integration of signaling inputs from leptin-induced SNS outflow and lipid-sensitive PPAR activation is required for a full and effective adaptive thermogenic response. A better understanding of the signaling systems, which interact to regulate and limit leptin signaling, is an important missing component in our overall understanding of the mechanisms of energy homeostasis.

See also: **Bioenergetics:** Uncoupling Proteins; **Metabolism**
Vitamins and Hormones: Fatty Acid Oxidation.

Further Reading

- Bates SH, Stearns WH, Dundon TA, et al. (2003) STAT3 signalling is required for leptin regulation of energy balance but not reproduction. *Nature* 421: 856–859.
- Bray GA (1991) Obesity, a disorder of nutrient partitioning: The MONA LISA hypothesis. *Journal of Nutrition* 121: 1146–1162.
- Commins SP, Marsh DJ, and Thomas SA (1999) Norepinephrine is required for leptin effects on gene expression in brown and white adipose tissue. *Endocrinology* 140: 4772–4776.
- Friedman JM (2002) The function of leptin in nutrition, weight, and physiology. *Nutrition Reviews* 60: S1–S14.
- Gettys TW, Harkness PJ, and Watson PM (1996) The β_3 -adrenergic receptor inhibits insulin-stimulated leptin secretion from isolated rat adipocytes. *Endocrinology* 137: 4054–4057.
- Guerra C, Koza RA, Yamashita H, Walsh K, and Kozak LP (1998) Emergence of brown adipocytes in white fat in mice is under genetic control – effects on body weight and adiposity. *Journal of Clinical Investigation* 102: 412–420.
- Halaas JL, Gajiwala KS, and Maffei M (1995) Weight-reducing effects of the plasma protein encoded by the *obese* gene. *Science* 269: 543–546.
- Zhang Y, Proenca R, and Maffei M (1994) Positional cloning of the mouse obese gene and its human homologue. *Nature* 372: 425–432.

Exosomes and Microvesicles

S J Gould, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Published by Elsevier Inc.

Introduction

Animal cells secrete a variety of vesicular particles, from the micron-sized platelets released by megakaryocytes, to smaller vesicles that have the approximate size of virus particles. Several names have been used to describe these smaller vesicles, primarily exosomes and microvesicles. This article discusses our current understanding of the structure of these vesicles, their biogenesis, and their functional significance. While some investigators differentiate between exosomes and microvesicles based on where they are made within the cell, that view is fraught with conceptual, mechanistic, and technical difficulties. For this reason, exosomes and microvesicles will be referred to by the collective acronym of exosome and microvesicle (EMV) in this article, though the term exosome will be used when discussing the classic model of exosome biogenesis.

EMV Structure

EMVs have a limiting bilayer membrane and a proteinaceous lumen. EMVs are commonly claimed to have a density of $\sim 1.15 \text{ g ml}^{-1}$. However, the density of an EMV is primarily a function of their protein:lipid ratio. Not surprisingly, the density of EMVs released by a single cell can be altered by changing the array of proteins being targeted to EMVs. EMVs also lack a unique, defining morphology. For example, EMVs released by human T-cells are fairly uniform in size and shape, but these parameters change dramatically upon expression of a single EMV cargo protein, AcylTyA. In short, EMVs cannot be strictly defined on the basis of density, size, or shape.

EMV Composition: Lipids, Proteins, Carbohydrates, and Nucleic Acids

EMV Lipids

EMVs have a limiting membrane with a bilayer appearance and a high degree of curvature. The phospholipid composition of mammalian EMV membranes is somewhat similar to that of mammalian plasma and endosome membranes, with high levels of phosphatidylcholine ($\sim 50\%$) and phosphatidylethanolamine ($\sim 20\text{--}30\%$), significant amounts of sphingomyelin ($\sim 10\text{--}20\%$), with the remaining portion coming from phosphatidylserine and phosphatidylinositols ($\sim 5\text{--}10\%$). EMVs are also rich in cholesterol, which comprises $\sim 30\%$ of EMV membrane lipid. It has been suggested by some that EMV biogenesis occurs at lipid rafts. However, a direct test of this hypothesis found that EMVs were devoid of several lipid raft markers, undermining the validity of this idea.

EMV Proteins

Like all biological membranes, the limiting membrane of EMVs is rich in proteins (Figure 1). The tetraspanins CD63, CD81, and CD9 are enriched in EMV membranes, as are major

histocompatibility (MHC) proteins, a subset of integrins, and other adhesion molecules such as intercellular adhesion molecule-1 and, in some cases, CD44. EMVs are also enriched for certain GPI-anchored cell-surface proteins such as the complement inhibitory factors CD55 and CD59. EMV membrane proteins also include cytoplasmically oriented proteins anchored to the inner leaflet of the EMV membrane via an acyl moiety (e.g., N-terminal myristoylation) or protein domains that bind to the plasma membrane-enriched phosphoinositides phosphatidylinositol-(4,5)-bisphosphate (PI(4,5)P2) and phosphatidylinositol-(3,4,5)-trisphosphate (PI(3,4,5)P3). These include the Gag proteins of exogenous and endogenous retroviruses, as well as certain synthetic EMV cargoes that have been developed. The outer surface of EMVs can also be decorated with peripheral membrane proteins such as fibronectin, which binds to EMV-anchored $\alpha 4 \beta 1$ integrin, and lactoferrin/MFG-E8, which binds to the phosphatidylserine that is enriched in the outer leaflet of EMVs.

EMVs also include cytoplasmic proteins. Actin and many actin-binding proteins have been identified in purified EMVs, raising the possibility that there might be a role for the cytoskeleton in EMV biogenesis. However, there is as yet no evidence for such a role. In fact, it is not even clear whether these cytoskeletal proteins are selectively targeted to EMVs, or merely incorporated passively during vesicle formation. A similar phenomenon of unselective incorporation might explain the detection of many other cytoplasmic (and membrane) proteins identified in proteomic analyses of purified EMVs (>2000 at last count). One of the more intriguing sets of cytoplasmic EMV cargoes are members of the endosomal sorting complexes required for transport (ESCRT) machinery, as TSG101, VPS28, and Alix are significantly enriched in EMVs. Although there is no set of proteins that defines exosome or microvesicles from one another or from the cell (a result that likely reflects the stochastic nature of the EMV biogenesis pathway), several proteins serve as reasonably reliable markers of EMVs. These include the tetraspanins CD63 and CD81 and the ESCRT-associated proteins TSG101 and Alix.

EMV Carbohydrates

The carbohydrate composition of EMVs is presumed to reflect that of the glycoproteins and glycolipids found on their surface. Unfortunately, the high degree of cell-to-cell variation in carbohydrate moieties, even among human T-cell lines, results in a corresponding cell-to-cell variation in the lectin-binding properties of EMVs. Thus, while there does appear to be a specific enrichment of certain carbohydrates in EMVs (e.g., high mannose), EMVs lack a unique carbohydrate signature.

EMV Nucleic Acids

A major advance in EMV biology was the finding that secreted vesicles contained significant amounts of nucleic acids,

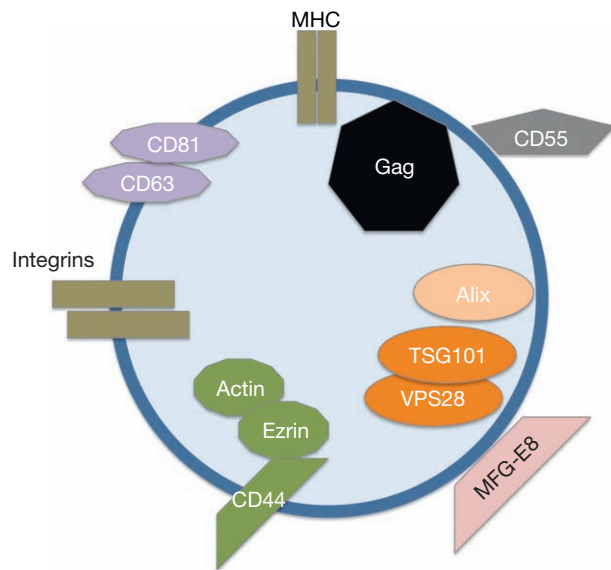


Figure 1 Major proteins of the EMV. EMVs contain a diverse array of proteins, including integral membrane proteins (e.g., CD63, CD81, MHC proteins, integrins, CD44), GPI-anchored proteins (e.g., CD55), outer peripheral membrane proteins (e.g., MFG-E8), inner peripheral membrane proteins (e.g., Gag), and cytosolic proteins that associate directly or indirectly with the inner leaflet of the plasma and endosome membranes (e.g., ezrin, actin, TSG101, VPS28, Alix).

particularly RNAs. Importantly, it has been demonstrated that EMV-enclosed RNAs can be transferred from cell to cell and that EMV-delivered messenger RNAs (mRNAs) and microRNAs (miRNAs) can alter gene expression in the recipient cell. As for the profile of RNAs present in secreted vesicles, there is a strong bias toward small RNAs (miRNAs and other small noncoding RNAs) and away from mRNAs and large ribosomal RNAs. There is also emerging evidence that some cells target specific miRNA families to EMVs.

EMV Biogenesis: A Prominent Role for the Plasma Membrane

The classic model of exosome biogenesis invokes (1) the selective targeting of exosomal proteins to the limiting membrane of endosomes, (2) the enrichment of exosomal proteins at endosomal sites of vesicle budding, (3) the inclusion of exosomal proteins in the intraluminal vesicles of discrete endosomes/multivesicular bodies (MVBs), and (4) the subsequent vesicular secretion of exosomal proteins when MVBs fuse with the plasma membrane. However, several tests of this hypothesis have generated anomalous rather than supporting data for this classical model of exosome biogenesis. For example, a wide variety of exosomal proteins and exosomal lipids are enriched at discrete domains of the plasma membrane in human T-cells, which also serve as sites for exosome budding. In addition, a search for *cis*-acting signals that target proteins to exosomes revealed that a wide variety of plasma membrane anchors can target proteins to exosomes, but that endosomal membrane anchors cannot. Furthermore, inhibitors of ESCRT function block MVB biogenesis but have no effect on the budding of exosomal cargoes. Finally, it has been established

that macrophages (and other cell types thought to bud exosomes from MVBs) possess deep invaginations of the plasma membrane that have an MVB-like morphology, and that budding into these structures might be mistaken for budding into bona fide MVBs. Based on these and other observations, it appears that the plasma membrane plays a more prominent role in exosome biogenesis than previously believed, and that the endosome membrane might play a less important role than currently assumed.

EMV Biogenesis: *trans*-Acting Factors

Another way to shed light on the mechanisms of EMV biogenesis is to identify *trans*-acting factors that are essential for vesicle production. Over the past decade, several proteins have been implicated in the production of secreted vesicles. These include Rab 11, 35, and 27, though the exact roles of these proteins in EMV biogenesis remain obscure. Ceramide has also been implicated in the vesicular budding of some EMV cargoes, raising the possibility that lipid composition might also contribute to EMV biogenesis. However, it is still unclear whether these factors play direct roles in EMV biogenesis or whether they act upstream of EMV biogenesis in the sorting of cargo proteins to sites of EMV biogenesis.

Cellular Functions of EMV Biogenesis

A major facet of EMV research is to understand how the formation of EMVs benefits the EMV-producing cell. Perhaps the most obvious role of EMV biogenesis is in protein (and lipid?) quality control at the plasma membrane. This involves both the targeted elimination of specific plasma membrane proteins during specific developmental processes and the continual elimination of aggregated plasma membrane proteins under steady-state conditions. The EMV biogenesis pathway also appears to play critical roles in cell polarity: in both human lymphocytes and protozoan social amoeba, the EMV protein-sorting pathway is responsible for targeting proteins to their posterior pole, or uropod. Furthermore, it has been shown that elevated levels of EMV cargo proteins polarize the EMV protein-sorting pathway, drive the formation of a nascent posterior pole, and prime the cell for the subsequent morphological and functional development of the posterior pole, as occurs during uropod formation.

EMV-Mediated Effects on Neighboring Cells

The physiological significance of EMV biology extends to neighboring cells. For example, EMVs have been reported to play numerous roles in immunological signaling, such as during antigen presentation, and the induction of dietary tolerance. EMVs have also been implicated in signaling from somatic epithelium to developing germ cells during sperm maturation, in the transport of morphogens during development, in neuronal signaling, and in the modulation of stem cell fate. The molecular mechanisms underlying these signaling events are still under investigation. EMVs transmit some of their signals by contact between EMV surface molecules and

their cognate receptors on the surface of neighboring cells. However, there is increasing evidence that EMVs can transmit signals to neighboring cells through a process of EMV:cell fusion, delivering molecules from the membrane and cytoplasm of donor cells into the membrane and cytoplasm of recipient cells. Clear evidence for this comes from the EMV-mediated transfer of activated epidermal growth factor receptor, as well as from the EMV-mediated transfer of RNAs.

EMVs and Human Disease

EMVs have been implicated in a wide range of human diseases, including human immunodeficiency virus (HIV) infection. Specifically, it has been proposed that HIV (and other retroviruses) use the preexisting, nonviral mechanism of EMV biogenesis for the production of infectious particles. In support of this hypothesis, it has been shown that: (1) EMVs and retrovirus particles have a similar molecular composition; (2) retroviral proteins are targeted to sites of EMV budding and secreted from the cell in EMVs; (3) higher-order oligomerization and plasma membrane binding target proteins to both EMVs and to HIV particles; (4) higher-order oligomerization and plasma membrane binding are the primary budding information in HIV Gag; and (5) retroviruses use the same types of plasma membrane anchors that are the most efficient at inducing the EMV targeting of nonviral proteins (i.e., an N-terminal myristoylation tag or a PIP2-binding domain).

The roles of EMVs in cancer are less direct but potentially more complex. Cancer-derived EMVs are thought to be mediators of immunosuppressive environments, potentiating the apoptosis of activated, cytotoxic, anti-cancer T-cells, impairing the differentiation of monocytes into dendritic cells, and inducing myeloid-suppressive cells and T-regulatory cells. Moreover, such EMVs are found not only within the cancer but also in the circulating fluids of the cancer patient, raising the possibility that tumor-derived EMVs might prime the body for subsequent spread of metastatic tumor cells. In addition, cancer-derived EMVs have been shown to be enriched in doxorubicin and perhaps other anti-cancer drugs, raising the possibility that EMVs might be a significant route for cancer drug resistance.

The role of EMVs in human disease also extends to several neurological disorders. The most obvious is in prion-based amyloidopathies where the normal cellular prion protein, PrP^C, is normally sorted to sites of EMV biogenesis and secreted from the cell in EMVs. Intriguingly, this appears to also hold true for the pathogenic form of the protein, PrP^{Sc}, suggesting that EMVs might facilitate the spread of disease within the body, and perhaps between individuals. In addition, it has been proposed that EMVs might also be involved in the release and spread of 'prionoids' that are implicated in Alzheimer's disease, Parkinson's disease, Huntington's disease, etc.

Conclusion

This article demonstrates the significant strides that have been made in understanding the structure and function of exosome and microvesicles, the fascinating biological potential of the intercellular vesicle trafficking pathway defined by these secreted organelles, and the continuing efforts to elucidate the modes and mechanisms of vesicle secretion. The accomplishments of the past decade bode well for a rapid resolution of the most pressing questions in the field as well as for new discoveries about how EMVs contribute to the biology of animal cells and the complex organisms they comprise.

Further Reading

- Aguzzi A and Rajendran L (2009) The transcellular spread of cytosolic amyloids, prions, and prionoids. *Neuron* 24: 783–790.
- Fang Y, Wu N, Gan X, et al. (2007) Higher-order oligomerization targets plasma membrane proteins and HIV Gag to exosomes. *PLoS Biology* 5: e158.
- Gould SJ, Booth A, and Hildreth J (2003) The Trojan exosome hypothesis. *Proceedings of National Academy of Sciences of the United States of America* 100: 10592–10597.
- Simons M and Raposo G (2009) Exosomes – vesicular carriers for intercellular communication. *Current Opinion in Cell Biology* 21: 575–581.
- Shen B, Fang Y, Wu N, and Gould SJ (2011) Biogenesis of the posterior pole is mediated by the exosome/microvesicle protein-sorting pathway. *Journal of Biological Chemistry* 286: 44162–44176.
- Taylor DD and Gercel-Taylor C (2011) Exosomes/microvesicles: Mediators of cancer-associated immunosuppressive microenvironments. *Seminars in Immunopathology* 33: 441–454.

FAK Family

S K Hanks, Vanderbilt University School of Medicine, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Activation loop A region within protein kinase catalytic domains that is often phosphorylated to promote catalytic activity.

Integrins A large family of cell surface adhesion receptors, composed of α - and β -subunits, which link the extracellular matrix to the actin cytoskeleton and transmit biochemical signals across the plasma membrane.

Src-homology 2 (SH2) A domain that mediates protein–protein interactions by binding to sites of tyrosine phosphorylation.

Src-homology 3 (SH3) A domain that mediates protein–protein interactions by binding to proline-rich motifs with a core sequence of Pro-X-X-Pro.

Tyrosine kinase A kinase that phosphorylates protein substrates on tyrosine residues.

Focal Adhesion Kinase

Focal adhesion kinase (FAK) ([Figure 1](#)) was the first member of the family to be described, in 1992, and has been most extensively studied. Its name derives from its prominent localization to focal adhesions observed in cultured fibroblasts ([Figure 2](#)). Focal adhesions are sites where cell surface integrins cluster and make strong adhesive contact with components of the extracellular matrix (ECM) such as fibronectin, vitronectin, and collagen. Focal adhesion localization was the first indication that FAK has a signaling function in regulating aspects of cell behavior in response to integrin-mediated adhesion. The focal adhesion targeting (FAT) domain is necessary and sufficient for focal adhesion localization. The FERM domain (4.1 protein, ezrin, radixin, and moesin (FERM)) acts as a negative-regulatory element by interacting with and inhibiting the kinase domain. Signaling by FAK requires release of this autoinhibition, which involves conformational change triggered by contractile tension generated at the focal adhesions. Hence, FAK can act in the process of mechanotransduction whereby cellular signaling events are initiated in response to mechanical forces.

FAK Expression and Roles in Development

FAK is widely expressed throughout embryogenesis and functions in many developmental processes. FAK-deficient mouse embryos fail to develop past late stages of gastrulation, and exhibit dramatically retarded development of the antero-posterior axis, suggestive of defects in mesodermal migration. Studies using late-stage embryos from mouse, *Xenopus*, and zebrafish indicate high FAK expression associated with several developing systems, including the central nervous system,

somites, arterial walls, myotendinous junctions, and germ cells. FAK is ubiquitously expressed in adult tissues, with high levels found in brain, lung, and testis. In neuronal cells, FAK is enriched in growth cones. Neuronal FAK is expressed from alternatively spliced transcripts resulting in short insertions in the region between the FERM and kinase domains. Targeted knockout studies in mice have demonstrated critical roles for FAK in development of the heart, vascular network, and nervous system.

FAK Activation and Signaling

Like most tyrosine kinases, FAK is in its active signaling state when phosphorylated on tyrosine residues. FAK tyrosine phosphorylation results from integrin-mediated cell adhesion to the ECM and is maintained by Rho-mediated contraction of the actin cytoskeleton. The initial step in FAK activation is phosphorylation of Tyr-397, which occurs through intermolecular autophosphorylation. Neuronal-specific isoforms of FAK have a higher intrinsic capacity to autophosphorylate. A crucial aspect of FAK signaling is the recruitment of Src-family kinases, which bind via their SH2 domains to the phosphorylated Tyr-397 site. This leads to Src-mediated phosphorylation of other FAK tyrosine residues: 576, 577, 861, and 925 ([Figure 1](#)). Tyrosines 576 and 577 reside in the kinase domain activation loop, and phosphorylation of these sites elevates catalytic activity. Phosphorylated Tyr-925 can bind the SH2 domain of the adaptor protein Grb2, suggesting a mechanism for integrin stimulation of extracellular-signal regulated kinase (ERK) via a FAK > Grb2/SOS > Ras pathway. Tyr-925 lies buried in the FAT domain such that phosphorylation of this residue may occur only upon a structural alteration of the FAT

domain that would also release FAK from focal adhesions. The signaling function of the Tyr-861 site is uncertain.

Src-family kinases recruited to the Tyr-397 site also phosphorylate two FAK-associated proteins: Crk-associated substrate (p130Cas) and paxillin. P130Cas, via its SH3 domain, interacts with either of two proline-rich motifs (PR1 and PR2) that reside in the region linking the kinase and FAT domain. Paxillin interacts with the FAT domain. For both p130Cas and paxillin, the major tyrosine phosphorylation sites are in Tyr-X-X-Pro (YxxP) motifs. Phosphorylation of YxxP sites promotes downstream signaling events through recruitment of additional SH2-containing effector molecules, notably adaptor proteins of the Crk and Nck families. In addition to Src family kinases, the phosphorylated Tyr-397 site mediates interactions with SH2 domains of other signaling proteins including phosphatidylinositol 3-kinase (PI3K), the γ 1 isoform of phospholipase C (PLC- γ 1), Shc, Grb7, and Nck-2. Thus, FAK likely promotes assembly of several distinct signaling complexes within a larger integrin-associated protein network.

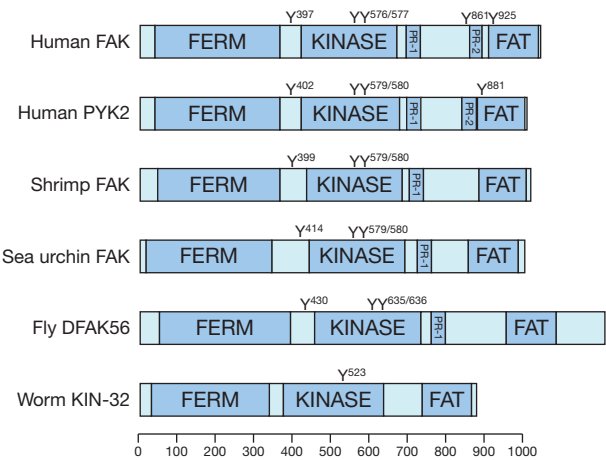


Figure 1 Representative members of the FAK family. FAK and PYK2 (represented by the human proteins) are the two vertebrate family members. Also shown are the homologs from *Marsupenaeus japonicus* (Shrimp FAK), *Lytechinus variegatus* (Sea urchin FAK), *Drosophila* (Fly DFak56), and *Caenorhabditis elegans* (Worm KIN32). Shown in the linear structures are major conserved domains, protein interaction motifs (PR1 and PR2), and sites of tyrosine (Y) phosphorylation. The scale bar represents number of amino acid residues.

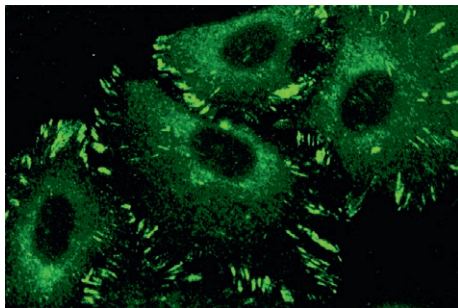


Figure 2 Localization of FAK in focal adhesions. Primary cultures of mouse keratinocytes were plated on fibronectin and stained using an antibody against FAK. The highly immunoreactive structures observed at the cell peripheries are focal adhesions.

Cellular Responses to FAK Signaling

FAK signaling promotes cell migration. For example, fibroblasts derived from *FAK*^{-/-} mouse embryos exhibit poor motility compared to cells from normal embryos, and the defect can be rescued by re-expressing wild-type FAK, but not by signaling-deficient FAK mutational variants (i.e., lacking the Tyr-397 autophosphorylation site). The cell motility response to FAK signaling is due to defects in focal adhesion disassembly and organization of lamellipodia at the protruding leading edge. FAK signaling also promotes activation of matrix-degrading metalloproteinases, thus stimulating the ability of cells to invade through the ECM.

In addition to motility and invasion, FAK functions as a positive regulator of cell growth by triggering both survival and proliferative signals. FAK-mediated signaling confers resistance to apoptosis resulting from various stimuli, including loss of anchorage to the ECM (anoikis), oxidative stress, ultraviolet (UV) irradiation, and exposure to anticancer drugs. FAK survival signals can result from multiple downstream events, including p130CAS/Crk complex formation, resulting in PI3K-mediated activation of Akt. FAK may also help drive progression through G1 phase of the cell cycle through its ability to activate ERKs and promote transcriptional activation of the gene encoding cyclin D1. FAK expression and activity are commonly elevated in human cancers, making FAK a potentially important target in the therapeutic treatment of these diseases.

Proline-Rich Tyrosine Kinase 2

Proline-rich tyrosine kinase 2 (PYK2) is also known as cellular adhesion kinase β (CAK- β), related adhesion focal tyrosine kinase (RAFTK), and calcium-dependent tyrosine kinase (CADTK). The FAT and kinase domains of PYK2 each exhibit 60% amino acid identity with the homologous domains of FAK, while the FERM domain is less conserved (Table 1). Both PR motifs in the kinase-FAT linking region are present in PYK2 (and capable of mediating interactions with p130Cas), but otherwise this region is poorly conserved, as is the short region amino-terminal to the FERM domain. Four tyrosine residues corresponding to sites of FAK phosphorylation are functionally conserved (Figure 1): phosphorylated Tyr-402 mediates interactions with Src-family kinases, phosphorylated activation loop tyrosines 579 and 580 stimulate PYK2 catalytic activity, and phosphorylated Tyr-881 binds Grb2.

PYK2 Activation and Signaling

Like FAK, PYK2 activation can result from integrin stimulation. For example, PYK2 undergoes tyrosine phosphorylation

Table 1 Conservation of FAK family domains^a

	PYK2	DFak56	KIN-32
FERM	45	29	25
KINASE	60	63	51
FAT	60	44	19

Human PYK2 was used for the comparison.

^aPercent amino acid identities are shown compared to human FAK.

following plating of megakaryocytes onto fibronectin and upon antibody stimulation of B-cell integrins. In fibroblasts, PYK2 is detected in focal adhesions, consistent with the high conservation of the FAT domain. It is notable, however, that the FAT domain of PYK2 targets to focal adhesions inefficiently compared to that of FAK. PYK2 activation in response to integrin-mediated adhesion has been observed in *FAK*^{−/−} fibroblasts. But re-expression of FAK in these cells effectively suppresses the PYK2 response, probably by displacing PYK2 from the adhesion sites. These observations suggest that PYK2 may not play as widespread a role in integrin signaling as does its sister molecule, FAK.

In contrast to FAK, PYK2 can be activated to autophosphorylate on Tyr-402 by extracellular signals that elevate levels of intracellular calcium and stimulate calcium-dependent isoforms of protein kinase C. These include agonists for G-protein-coupled receptors, voltage-gated calcium channels, and the nicotinic acetylcholine receptor. The mechanisms by which calcium and/or protein kinase C activate PYK2 remain poorly understood. The downstream signaling events resulting from PYK2 activation are generally similar to those of FAK, involving recruitment of Src family kinases to the autophosphorylation site, and Src-mediated phosphorylation of other sites on PYK2 as well as PYK2-associated p130Cas and paxillin (Figure 3). The interaction of Grb2 with the PYK2 Tyr-881 site provides a link between G-protein-coupled receptors and ERK activation.

Cellular Responses to PYK2 Signaling

Unlike FAK, PYK2 overexpression does not efficiently promote fibroblast motility, and results in apoptosis rather than cell survival. Nevertheless, a crucial role for PYK2 in the motility and migration of macrophages and B cells has been demonstrated using cells derived from PYK2-deficient mice. Defects in

lymphocyte motility could explain the absence of marginal zone B cells in these animals. Through its ability to be activated by both integrin- and calcium/PKC-dependent mechanisms, PYK2 can function in numerous and diverse signaling pathways regulating a wide variety of cellular activities. In neuronal cells, PYK2 has been implicated as a positive regulator of neurite outgrowth and *N*-methyl-D-aspartate (NMDA) channel activity. Other consequences of PYK2 signaling include inhibition of voltage-gated potassium channels leading to membrane depolarization of vascular smooth muscle cells, translocation of the glucose transporter type 4 (GLUT4) to the plasma membrane of adipocytes, stimulation of Na-HCO^{3−} cotransporter activity in renal epithelial cells, and angiogenesis of vascular endothelial cells. In small-cell lung cancer, activation of PYK2 is involved in neuropeptide-stimulated cell survival and proliferation.

Invertebrate Members of the FAK Family

Drosophila and *Caenorhabditis elegans* genomes each contain a single member of the FAK family: DFak56 in *Drosophila* and KIN32 in *C. elegans*. Sea urchin and shrimp FAK homologs have also recently been described. FAK and PYK2 are more closely related to one another than they are to any of the invertebrate homologs, indicating that the two vertebrate family members arose through gene duplication after evolutionary separation from the invertebrates.

Drosophila DFak56 shows overall conservation with the two vertebrate family members, but is slightly more similar to FAK than to PYK2. The kinase domain of DFak56 is most highly conserved in comparison to FAK, followed in order by the FAT domain and the FERM domain (Table 1). Within the kinase-FAT linker region, DFak56 contains a proline-rich motif corresponding to the PR1 site, but lacks the PR2 site. Of the six sites of FAK tyrosine phosphorylation, DFak56 lacks only the site corresponding to FAK Tyr-861. However, DFak56 Tyr-962, which corresponds to the Grb2-binding sites in FAK and PYK2, lacks the asparagine residue at the +2 position, which is part of the consensus sequence for Grb2 SH2 binding. A unique structural feature of DFak56 is an extended carboxyl-terminal region. Like the vertebrate members of the FAK family, DFak56 exhibits autophosphorylation activity and undergoes tyrosine phosphorylation in response to integrin-mediated adhesion. DFak56 is broadly expressed during embryogenesis, with strongest expression in the developing gut, central nervous system, and muscle. Null mutations of the *DFAK56* gene have been linked to defects in optic stalk formation by glial cells of the visual nervous system and to overgrowth of larval neuromuscular junctions. Given the embryonic lethal phenotype of the mouse FAK knockout, it is somewhat surprising that flies lacking DFak56 do not have more dramatic developmental defects. The activity of DFak56 in restricting neuromuscular junctions' synaptic elaboration was associated with ERK inhibition, further indicating differences in signaling functions of vertebrate and invertebrate FAK-family kinases.

C. elegans KIN32 contains a well-conserved kinase domain, while both the FAT and FERM domains are rather poorly conserved in comparison to other members of the FAK family

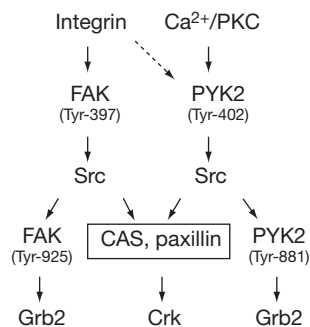


Figure 3 Signaling by FAK and PYK2. Both FAK and PYK2 can be activated by integrin-mediated adhesion, while PYK2 is uniquely activated through a calcium/PKC-dependent mechanism. A critical element of FAK and PYK2 signaling involves the recruitment of Src family kinases to sites of autophosphorylation (Tyr-397 for FAK and Tyr-402 for PYK2). The associated Src family kinases may then phosphorylate a conserved FAK/PYK2 tyrosine site to promote Grb2 binding, and also phosphorylate FAK/PYK2-associated proteins including p130Cas and paxillin to promote downstream signals such as those mediated by Crk SH2/SH3 adaptor proteins. Grb2 and Crk interact with guanine-nucleotide exchange factors leading to the activation of small G proteins Ras and Rac, respectively.

(Table 1). Interestingly, KIN32 lacks key motifs associated with signaling by FAK-family kinases including a tyrosine corresponding to the autophosphorylation site and proline-rich motifs in the kinase-FAT linker region, suggesting a divergent signaling function. KIN32 is expressed in all larval stages and is prominent in body wall muscle cells. As is the case for *Drosophila* DFak56, there is no obvious phenotype associated with a deletion of the gene encoding *C. elegans* KIN32.

See also: **Cell Architecture and Function:** Focal Adhesions and Related Integrin Contacts; **Signaling:** Integrin Signaling; Src Family of Protein Tyrosine Kinases.

Further Reading

Avraham H, Park SY, Schinkmann K, and Avraham S (2000) RAFTK/Pyk2-mediated cellular signaling. *Cellular Signaling* 12: 123–133.

- Cram EJ, Fontanez KM, and Schwarzbauer JE (2008) Functional characterization of KIN-32, the *Caenorhabditis elegans* homolog of focal adhesion kinase. *Developmental Dynamics* 237: 837–846.
- Girault J-A, Costa A, Derkinderen P, Studler J-M, and Toutant M (1999) FAK and PYK2/CAK β in the nervous system: A link between neuronal activity, plasticity, and survival? *Trends in Neurosciences* 22: 257–263.
- Hanks SK, Shin N-Y, Ryzhova L, and Brabek J (2003) Focal adhesion kinase signaling activities and their implications in the control of cell survival and motility. *Frontiers in Bioscience* 8: d982–d996.
- Ilic D, Furuta Y, Kanazawa S, et al. (1995) Reduced cell motility and enhanced focal adhesion contact formation in cells from FAK-deficient mice. *Nature* 377: 539–544.
- McLean GW, Carragher NO, Avizienyte E, et al. (2005) The role of focal adhesion kinase in cancer – a new therapeutic opportunity. *Nature Reviews. Cancer* 5: 505–515.
- Mitra SK, Hanson DA, and Schlaepfer DD (2005) Focal adhesion kinase: In command and control of cell motility. *Nature Reviews. Molecular Cell Biology* 6: 56–68.
- Parsons JT (2003) Focal adhesion kinase: The first ten years. *Journal of Cell Science* 116: 1409–1416.
- Schaller MD (2001) Biochemical signals and biological responses elicited by the focal adhesion kinase. *Biochimica et Biophysica Acta* 1540: 1–21.
- Tsai P-I, Kao H-H, Grabbe C, et al. (2008) Fak56 functions downstream of integrin α PS3betanu and suppresses MAPK activation in neuromuscular junction growth. *Neural Development* 3: 26.

F-ATPases

J E Walker, The Medical Research Council Mitochondrial Biology Unit, Cambridge, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Oligomycin and the OSCP subunit Oligomycin is an antibiotic that inhibits the proton-pumping activity of the F-ATPase in mitochondria, but not in eubacteria or chloroplasts. In *Saccharomyces cerevisiae*, oligomycin-resistant mutants map to the F_o subunits, a and c, indicating that the site of action of the antibiotic is probably in the region of their interface. However, the name of the OSCP subunit, or in full, the oligomycin sensitivity conferral protein, often leads to the erroneous assumption that this subunit provides the site of binding of the antibiotic. The name of the OSCP derives historically from reconstitution experiments in which the separate mitochondrial F_1 domain was reassembled onto inner membrane vesicles from bovine heart mitochondria lacking the F_1 domain. It was found that in order to reassemble a fully functional proton-pumping F-ATPase and thereby regain the sensitivity of its activity to the action of oligomycin, other protein factors had to be added also. One of these factors was named the OSCP in recognition of the necessity for its presence in the reassembled enzyme in order to regain sensitivity to the antibiotic, and the other was named simply as factor 6, or F_6 . Today, we know that both of these proteins are essential components of the peripheral stalk of the mitochondrial F-ATPase. In an enzyme assembly lacking either of them, ATP synthesis or hydrolysis is uncoupled from proton translocation through the F_o membrane domain, and is insensitive to the action of oligomycin in the F_o domain.

Proton-motive force The term trans-membrane proton-motive force (pmf or Δp) was coined by the

British biochemist and Nobel Laureate, Peter Mitchell, as a way of describing the proton electrochemical gradient, $\Delta \tilde{\mu}_{H^+}$, that is built up across an energy-transducing membrane by respiration or photosynthesis. It has two components, the membrane potential difference across that membrane, $\Delta\psi$, and the pH difference across that membrane, ΔpH . The pmf is expressed in units of voltage. At 25 °C: Δp (in mV) = $\Delta\psi - 59 \Delta pH$.

Subunit equivalence The nomenclatures of subunits of F-ATPases from eubacteria, chloroplasts, and mitochondria have evolved somewhat independently, creating the potential for confusion and misunderstanding. The terms α -, β -, and γ -subunits denote equivalent subunits in all F-ATPases. However, the equivalent subunits in mitochondria of the eubacterial and chloroplast δ - and ϵ -subunits are, respectively, the OSCP and the δ -subunit. Eubacterial and chloroplast F-ATPases do not contain a subunit that is equivalent to the mitochondrial ϵ -subunit. In chloroplast F-ATPases, subunits a, b, b', and c are sometimes referred to as subunits IV, II, I, and III, respectively, and in the mitochondrial enzyme subunit a is sometimes referred to as ATPase-6 or A6. For details of nomenclature of peripheral stalk subunits and supernumerary subunits in mitochondrial F-ATPases, see also the legend to [Figure 1](#).

The P/O ratio By convention in bioenergetics, O refers to $1/2O_2$, which is equivalent to two electrons. The P/O ratio is the number of moles of ADP phosphorylated to ATP per two electrons transferred to oxygen.

Introduction

The F-ATPases, also known as F_1F_o -ATPases, are multisubunit enzyme complexes found in energy-transducing membranes in eubacteria, mitochondria, and chloroplasts. Their role is to synthesize adenosine triphosphate (ATP) from adenosine diphosphate (ADP) and phosphate under aerobic conditions using a proton-motive force, Δp , generated by respiration or photosynthesis, as a source of energy. The ATP hydrolase activities of the enzymes from mitochondria, chloroplasts, and some eubacteria are inhibited, and these enzymes can only synthesize ATP. However, some other eubacterial enzymes can hydrolyze ATP, made by glycolysis under anaerobic conditions, to generate the Δp that is required for other essential cellular functions, such as chemotaxis and transmembrane transport processes. The F-ATPases are close relatives of the V-ATPases (vesicular ATPases) found in secretory membranes that use the energy from ATP hydrolysis to generate ion gradients across those membranes, and of the A-ATPases that make ATP in the archaea. This article focuses on the structure and

function of the F-ATPases. Some eubacteria generate a sodium ion-motive force that is coupled to ATP synthesis, and the structures and functions of these F-ATPases are broadly similar to the F-ATPases that are driven by Δp .

Structure of F-ATPases

Irrespective of their source, F-ATPases have related structures, as summarized in [Figure 1](#). They consist of two major functional domains, a membrane extrinsic F_1 sector and a membrane intrinsic F_o sector joined together by central and peripheral stalks. The F_1 domain is the catalytic part of the enzyme where ATP, the energy currency of the biological world, is formed from ADP and inorganic phosphate. The F_o domain contains a motor, which generates rotation using the potential energy stored in Δp , as explained in [Figure 2](#). The rotational energy of the motor is transmitted to the catalytic domain by the central stalk, which is attached directly to the rotary motor. Thus, the enzyme is analogous to a turbine in an

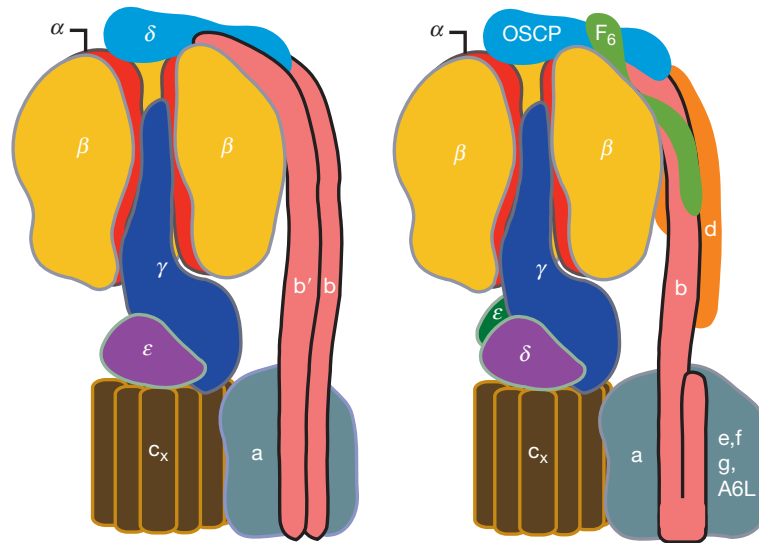


Figure 1 The organization of protein subunits in F-ATPases. The eubacterial and chloroplast F-ATPases are depicted on the left, and the more complex mitochondrial F-ATPase is on the right. The upper part of each model contains the subunits in the F_1 catalytic domain. One of the three α -subunits (red) has been removed to expose the elongated α -helical structure in the γ -subunit (dark blue), which lies approximately along the central axis of the $\alpha_3\beta_3$ domain. The γ -subunit (and associated subunits) is in contact with the F_0 membrane domain, which contains the c-ring (brown) and the associated a-subunit (gray). The number of c-subunits in the c-ring differs between species. The rotor of the enzyme consists of the ensemble of the c-ring and the γ -subunit (and associated subunits). The pathway for protons through the F_0 domain is in the vicinity of the interface between the c-ring and a-subunit. The peripheral stalk is on the right of each model. In some eubacterial enzymes, it consists of the δ -subunit (light blue) and two identical b-subunits (pink). In other eubacterial enzymes, and in the chloroplast enzyme, the two b-subunits are replaced by single copies of homologous, but nonidentical, subunits b and b'. The N-terminal domain of the δ -subunit binds to the N-terminal region of one of the three α -subunits, and the b- and b'-subunits interact with the a-subunit via their N-terminal transmembrane α -helices. In the mitochondrial enzyme, the peripheral stalk consists of single copies of subunits OSCP, b, d, and F_6 . The OSCP is the homolog of the bacterial δ -subunit. The sequences of mitochondrial subunits b, d, and F_6 are not evidently related to those of the bacterial b- and b'-subunits. Their structures also differ significantly, although they are both dominated by α -helices, except for the C-terminal domain of the OSCP (and presumably the eubacterial and chloroplast δ -subunit), which contain β -structures. The membrane domains of the mitochondrial enzyme contain a number of membrane subunits with single transmembrane spanning α -helices that are not found in eubacteria and chloroplasts. These supernumerary subunits have no known roles in the generation of ATP; subunits e, f, g, and A6L are indicated, and subunits DAPIT and the 8 kDa proteolipid, which are found in the complex when phospholipids are maintained during the purification of the complex, are not shown.

electric power station, which uses the potential energy of water stored in a dam to generate rotation in order to produce electricity, the energy currency of the man-made world.

The F_1 domain of the F-ATPase is an assembly of five globular proteins, α , β , γ , δ , and ϵ with the stoichiometry $\alpha_3\beta_3\gamma_1\delta_1\epsilon_1$. Their combined molecular weight is around 350 kDa. The $\alpha_3\beta_3\gamma_1$ subassembly is common to all F-ATPases. The three α - and the three β -subunits form an approximately spherical $\alpha_3\beta_3$ structure about 100 Å in diameter with the six subunits arranged in alternation around an elongated α -helical structure found in the γ -subunit. This part of the γ -subunit is completely enveloped in the $\alpha_3\beta_3$ domain and occupies its central axis. Thus, the $\alpha_3\beta_3$ structure resembles an orange made of six segments arranged around the central pith stalk of the fruit. The rest of the γ -subunit protrudes about 30 Å beyond the $\alpha_3\beta_3$ domain and forms part of a foot that attaches the $\alpha_3\beta_3\gamma_1$ complex firmly to the F_0 domain. This attachment is augmented by another protein, known for historical reasons as the ϵ -subunit in eubacteria and chloroplasts and as the δ -subunit in mitochondria. Despite their different names, these ϵ - and δ -subunits are folded and probably bind in a similar way, and their functions are probably related also, although they are not identical. In mitochondria, the attachment of the foot of the γ -subunit to the F_0 domain is

augmented by a second subunit. Known confusingly as the ϵ -subunit, this mitochondrial protein has no counterpart in eubacterial and chloroplast enzymes. The γ -subunit, and the associated ϵ -subunits in the eubacterial and chloroplast enzymes, and the associated δ - and ϵ -subunits in the mitochondrial enzyme, are bound firmly to one end of a hydrophobic cylindrical structure in the F_0 membrane domain of the enzyme. This cylinder, together with the γ -subunit and its associated ϵ - and δ -subunits, forms the rotor of the enzyme.

The cylinder is made of c-subunits, simple membrane proteins folded into two transmembrane α -helices, with the loop regions joining the α -helices exposed on the same side of the membrane as the F_1 domain, and with some of the loops at least in contact with, and binding to, the foot of the central stalk. The N-terminal α -helices of c-subunits form the core of the cylinder, and the C-terminal α -helices provide its external annulus. Most importantly for the generation of rotation of the rotor from the proton-motive force, each c-subunit contains an essential acidic amino acid residue (aspartate or glutamate depending on the species) placed at the mid-point of the membrane domain of the enzyme. The number of c-subunits that form the hydrophobic cylinder is not constant throughout the biological world. Vertebrates (about 50 000 species) have eight c-subunits arranged in their cylindrical ring structures,

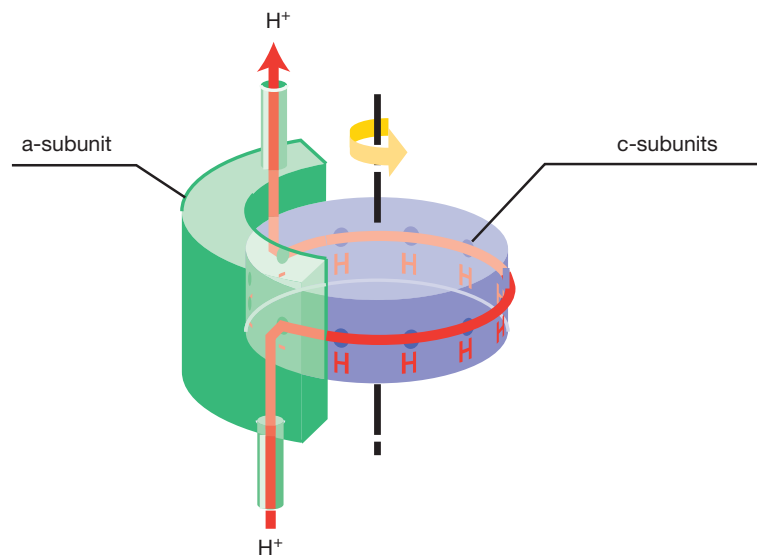


Figure 2 The generation of rotation in the F_0 membrane domain of F-ATPases. The a-subunit (green) is shown in contact with the cylindrical ring of c-subunits (blue). The C-terminal α -helix of each c-subunit contains a carboxyl (in the side chain of an aspartate or glutamate residue; dark blue circles) exposed on the external circumference of the cylinder. From the bottom of the figure, protons from the side of the membrane distal from the F_1 domain of the F-ATPase are shown entering a half-channel and neutralizing a negatively charged carboxylate in the interface between the a-subunit and the c-ring. Once neutralized, this residue moves by Brownian motion to the more hydrophobic environment of the lipid bilayer generating a rotation in the direction indicated. This rotational step brings another negatively charged carboxylate into the lower half-channel allowing another proton to be consumed, generating another rotational substep. Thus, as more negative charges are neutralized by protons, the protons are carried around on the external surface of the ring by the neutralized carboxyl, until they reach a second site in contact with the a-subunit, where the local environment reionizes the carboxyl, releasing the proton through a second half channel on the opposite of the membrane from which it entered. The direction and magnitude of Δp (~ 180 mV inwards in mitochondria) ensure that the rotation is unidirectional, and there is no requirement for a ratchet mechanism. The number of protons required to generate each 360° rotation of the c-ring corresponds to the number of c-subunits that form the ring. The γ -subunit (and associated subunits) are attached firmly to the ring, and each 360° rotation of this part of the rotor provides the energy to generate three molecules of ATP from the F_1 domain (see below).

and invertebrates (about 2 million species) probably have c_8 -rings as well. Hence, the c_8 -feature is probably found in F-ATPases throughout the multicellular biological world. By contrast, unicellular organisms, including fungi and eubacteria (and also chloroplasts from green plants), have a variety of different numbers of c-subunits in their c-rings, and values of 9, 10, 11, 13, 14, and 15 have been reported. These values are highly significant in the consideration of the energy cost to the cell in making each molecule of ATP (see later).

During ATP synthesis, the rotation of the ring is driven by the proton-motive force in a clockwise direction (as viewed from the membrane domain toward the F_1 domain), with an estimated rotary speed of about 100–150 revolutions per second, depending on the species. In the F-ATPases that hydrolyze ATP under anaerobic conditions, the energy to drive rotation is now provided by the hydrolysis of ATP, which drives the rotor in a counterclockwise direction, and pumps protons across the energy transducing membrane in the outward direction (away from the F_1 domain of the enzyme), thereby generating the trans-membrane Δp . The c-ring is in contact with, and rotates against, the surface of another hydrophobic membrane protein, known variously as subunit a (in eubacteria), ATPase-6 or A6 (in mitochondria) and subunit IV (in chloroplasts). For simplicity, here it will be referred to as subunit a. Little is known about the structure of this protein, except that it contains a basic amino acid (arginine-210 in the protein from *Escherichia coli*) that participates in the translocation of protons

through the F_0 membrane domain of the enzyme. It is likely that this protein provides a pathway (a channel or two half channels) to allow protons to access the negatively charged carboxylate residues on the C-terminal α -helices of c-subunits on the external surface of the c-ring in the rotor, and, following neutralization of the successive carboxylates, rotation of the ring, and successive reionization of the carboxylates, to allow the protons to be released on the opposite side of the F_0 domain. The protein is joined to the external surface of the $\alpha_3\beta_3$ domain by the peripheral stalk, an elongated and largely α -helical structure that extends from the F_0 membrane domain to the topmost extremity of the F_1 -domain, where it binds to the N-terminal region of an α -subunit. The role of the peripheral stalk is to counter the tendency of the $\alpha_3\beta_3$ domain to follow the rotation of the rotor, and its integrity is essential for Δp to be coupled to ATP synthesis.

How ATP Is Made in the F_1 Domain

The understanding of the mechanism of how ATP is synthesized in the catalytic F_1 domain of the enzyme is based on a series of high-resolution structures determined by X-ray crystallography with the bovine and yeast enzymes. Many of them represent the ground state in the catalytic cycle either of ATP hydrolysis or synthesis, or of the closely related ADP-inhibited state. One such structure is summarized in [Figure 3](#).

In these structures, the three noncatalytic α -subunits and the three catalytic β -subunits of the enzyme are arranged in alternation around an asymmetric α -helical structure in the single γ -subunit. The α - and β -subunits are similar, each consisting of an N-terminal domain with six β -strands, a central nucleotide-binding domain made of both α -helices and β -strands, and an α -helical C-terminal domain containing six α -helices in β -subunits and seven in α -subunits. Because the γ -subunit is asymmetrical, F₁-ATPase is an inherently asymmetrical enzyme, and a structure of yeast F₁-ATPase without bound nucleotides demonstrated that the asymmetry is derived from the protein assembly itself. The asymmetry of the single γ -subunit obliges each of the three catalytic β -subunits to adopt a different conformation with different nucleotide binding and catalytic properties. Two of them have similar conformations, but in one structure, now recognized as the ADP-inhibited state of the enzyme (see section [Regulatory Mechanisms](#)), one subunit designated as β_{DP} contains bound ADP-Mg, and the second subunit, β_{TP} , has bound AMP-PNP-Mg (adenylylimidodiphosphate is a nonhydrolyzable analog of ATP). However, in other structures, which represent the ground state of the active catalytic cycle, two molecules of AMP-PNP-Mg, ADP-BeF-Mg, or ADP-Mg are bound to both β_{TP} - and β_{DP} -subunits in the same structure. Hence, the β_{DP} - and β_{TP} -subunits can bind either ADP or ATP. In all of these structures, the third subunit, the β_E -subunit, adopts a radically different conformation, in which part of the nucleotide-binding domain and the attached C-terminal domain have hinged outward together, in response to the curvature of the γ -subunit. In these structures, this β -subunit has no bound nucleotide, and so it is known as the empty or open state, designated as β_E . None of the α -subunits open in a similar way. All three of them remain closed, each binding a magnesium ion and a nucleotide, and they remain bound throughout the catalytic cycle. The nucleotides bound to α -subunits have no direct role in the formation of ATP. Other intermediate states in the catalytic cycle have been resolved by X-ray crystallography most notably, a transition-state analog structure, where the catalytic cycle has been arrested before the release of ADP-Mg from the β_E -subunit.

In order to explain the interconversion of catalytic sites through tight, loose, and open states required by a binding change mechanism of catalysis of ATP hydrolysis by F₁-ATPase, it was proposed that the interconversion of sites was effected by a mechanical rotation of the γ -subunit, each 360° rotation taking each β -subunit through the three states represented by the β_{TP} -, β_{DP} -, and β_E -subunits, thereby hydrolyzing three ATP molecules. It was shown subsequently by biophysical experiments conducted with single enzyme complexes that during ATP hydrolysis, in either a bacterial $\alpha_3\beta_3\gamma$ complex or the intact F₁F₀-ATPase, the direction of rotation was counterclockwise (as viewed from the membrane domain of the enzyme) and that the rotation proceeds in 120° steps which were resolved subsequently into 90° and 30° substeps (or 80° and 40° substeps in other experiments). The ground-state structure probably represents the state of the enzyme at the end of the complete 120° rotary step. During ATP synthesis, the direction of rotation is reversed, and it is assumed that the order of the structural changes accompanying ATP hydrolysis is also reversed. One significant question that arises is how is the stepping action

generated from the turning of the rotor? The most likely explanation is that before a 120° step occurs energy is stored from the turning rotor by an elastic element, and then released in packet (or two subpackets corresponding to the substeps) in order to generate the stepping action. This elastic element is likely to reside in the α -helical coiled-coil region of the γ -subunit, but it remains possible that the peripheral stalk participates in transient energy storage also. However, the main role of peripheral stalk is probably to help the $\alpha_3\beta_3$ domain to resist the rotational torque of the rotor.

The main features of the rotary action of the F-ATPase and the synthesis of ATP are shown in [Movie 1](#).

Bioenergetic Cost of Making an ATP Molecule

Knowledge of the main architectural features of the F-ATPases and their functions allows a fundamental deduction to be made about energy transduction in the organisms in which these enzymes are found, namely the number of protons that are required to be translocated through the energy-transducing membrane for each ATP molecule that is synthesized by the F-ATPase from ADP and inorganic phosphate. This parameter is the bioenergetic cost to the organism of making each ATP molecule. Since each 360° rotation of the rotor of the F-ATPase produces three ATP molecules from the F₁ domain of the enzyme, and each 360° rotation of the c-ring requires the translocation of the same number of protons as there are c-subunits in the ring, the bioenergetic cost to the F-ATPase is the number of c-subunits in the c-ring divided by 3. The c₈-rings in the bovine enzyme are the smallest c-rings that have been observed hitherto, and it is likely that they occur in all vertebrate F-ATPases, and probably in the enzymes from invertebrates also. The associated bioenergetic cost to these F-ATPases is 2.7 protons/ATP. In other organisms where ring sizes of 9, 10, 11, 13, 14, and 15 have been observed, the bioenergetic cost to the F-ATPase will be 3.0, 3.3, 3.7, 4.3, and 5.0 protons/ATP, respectively. In mitochondria, ATP is produced in the matrix of the organelle, and it is made available as a source of cellular energy by exchanging it for external ADP. Inorganic phosphate is also actively transported back into the matrix of mitochondria. The combination of the electrogenic exchange of an internal ATP for an external ADP and the nonelectrogenic symport of phosphate and a proton adds one proton to the total required to provide ATP to the cellular cytoplasm, and so the cost to mitochondria containing an F-ATPase with a c₈-ring will be 3.7 protons. In mammalian mitochondria for each two electrons transferred to oxygen or succinate, 10 or 6 protons, respectively, are translocated out of the matrix. Therefore, the number of moles of ADP phosphorylated to ATP per two electrons transferred to oxygen, known as the P/O ratio, will be 10/3.7 and 6/3.7 or 2.7 and 1.6, respectively, for NADH and succinate, similar to the experimental values of 2.5 and 1.5. P/O ratios have been less studied in other species, and there is uncertainty about the proton stoichiometries of their electron transfer complexes. Nonetheless, similar considerations apply to their F-ATPases, except that in eubacteria there is no requirement to translocate ATP, and ADP and phosphate across the energy-transducing membrane.

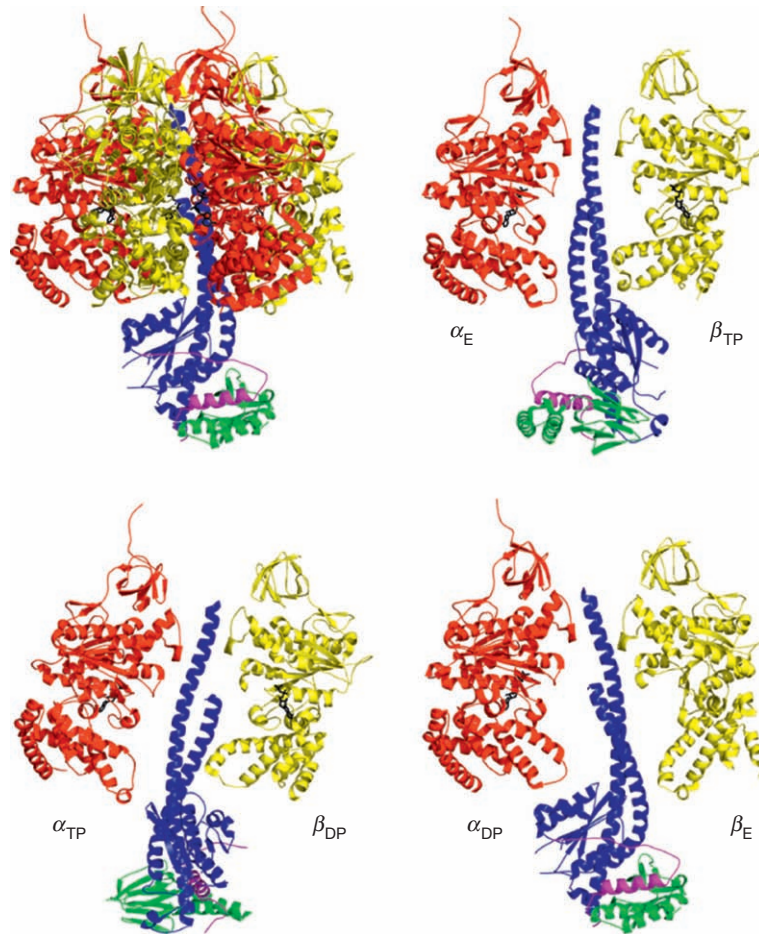


Figure 3 The structure of the F_1 catalytic domain bovine F-ATPase in the ground state of the catalytic cycle. The subunits of the protein are depicted in ribbon representation. The α -, β -, γ -, δ -, and ϵ -subunits are red, yellow, blue, green, and purple, respectively. Top left, the complete F_1 domain with three α - and three β -subunits and single copies of γ -, δ -, and ϵ -subunits. In the intact F-ATPase, the rotor of the enzyme consists of the central stalk interacting via its foot with loop regions between the N- and C-terminal α -helices of the eight c-subunits in the F_0 membrane domain. In the top right, bottom left and right diagrams, the three different conformations of the catalytic β -subunits that are present in the structure in the top left panel are shown, together with the central stalk (γ -, δ -, and ϵ -subunits) and an α -subunit, for reference. Each β -subunit has three domains, the N-terminal domain (top) is made of β -sheets, the nucleotide-binding domain (middle) is a mixed α - β structure, and the C-terminal domain (bottom) is a bundle of six α -helices. Each of the three β -subunits has been obliged by the asymmetry of the α -helical coiled-coil region of the γ -subunit to adopt a different conformation, denoted β_{TP} , β_{DP} , and β_E , with different affinities for nucleotides, the β_{TP} - and β_{DP} -subunits can each bind either ADP or ATP, and the β_E -subunit is unable to bind any nucleotide. During the catalytic cycle of ATP synthesis or hydrolysis, the rotation of the central stalk takes each β -subunit through each of the three conformations. As the rotor turns, this action leads to the binding and entrapment of ADP and phosphate at the β_E -site as it closes, ATP formation at the β_{DP} -site, and ATP release from the β_{TP} site as it opens and converts back to the β_E site. Each 360° rotation of the central stalk (clockwise as viewed from beneath) leads to the formation of an ATP molecule from each of the three β -subunits.

Regulatory Mechanisms

Different types of F-ATPases have different mechanisms of inactivation and reactivation of ATP synthesis under conditions of low and increasing available energy. In mitochondria and chloroplasts, when the pmf is low, ADP-Mg (without inorganic phosphate) remains bound to one of the three catalytic sites of the enzyme (E) forming an inactive complex, E-ADP. In chloroplasts, it is thought that the inactive E-ADP state is stabilized during the hours of darkness by formation of an intramolecular disulfide linkage in the γ -subunit of the F-ATPase. When daylight is restored, the enzyme is reactivated by thioredoxin-regulated reduction of the disulfide and

energy-dependent dissociation of the ADP-Mg. The role of the bound ADP-Mg in the physiological regulation of the mitochondrial enzyme is less clear. At a pH of about 6.7 or below, the mitochondrial F-ATPase forms a 1:1 complex with an inhibitory protein, IF_1 . The formation of the inhibited complex requires the hydrolysis of two ATP molecules, and in the inhibited complex, the inhibitor is bound in a complex-binding site at a catalytic interface between the α_{DP} - and β_{DP} -subunits. The inhibition is relieved by restoration of an adequate membrane potential difference, allowing the enzyme to rotate the rotor in the synthetic direction, with the ejection of the inhibitor from its binding site and the resumption of the synthesis of ATP.

Despite similarities between the subunits of the F_1 domains of mitochondrial and eubacterial enzymes, for reasons that are not understood, mitochondrial IF_1 has no effect on eubacterial F-ATPases. However, the eubacterial F-ATPase from *Paracoccus denitrificans* and other α -proteobacteria have their own inhibitor protein, known as the ξ -subunit, that is thought to prevent ATP hydrolysis. However, the protein is not obviously closely unrelated to mitochondrial IF_1 , and its mode of action is not known. Other eubacterial F-ATPases, exemplified by the enzymes from *E. coli* and from the moderate thermophile, *Bacillus PS3*, appear to be able both to synthesize and to hydrolyze ATP, depending on whether they are operating under aerobic or anaerobic conditions. In addition, an inhibitory mechanism has been demonstrated *in vitro* for the enzymes from *E. coli* and *Bacillus PS3* involving the ϵ -subunit which is bound both to the γ -subunit in the central stalk of the F_1 -domain and to the c-ring in the membrane domain of the enzyme (see Figure 1). In a structure of the isolated ϵ -subunit, the N-terminal domain of the protein is folded into a 10-stranded β -sandwich, and the C-terminal region consists of two side-by-side α -helices. In the intact enzyme, the β -sandwich domain is involved in binding the subunit to the γ -subunit and the c-ring, and the α -helical C-terminal region has been proposed to adopt two different conformations, down and up. In the down conformation, the two α -helices are held associated closely with the β -sandwich by an ATP molecule bound (evidently without a magnesium ion) between the two domains. In the absence of bound ATP, the two α -helices are thought to assume the up position, where it has been proposed that they could interact with the $\alpha_3\beta_3$ complex and inhibit the enzyme in the ATP hydrolysis direction. However, the up position has not been captured in a high-resolution structure of the intact F_1 -domain, and a coherent explanation of the physiological context for such an inhibitory mechanism is lacking. Other eubacterial F-ATPases, exemplified by the enzymes from *Caldalkalibacillus thermarum* and *Mycobacterium tuberculosis*, synthesize ATP in the usual way under aerobic conditions, but they are not able to hydrolyze ATP under physiological conditions. Their latent ATP hydrolase activities can be activated only under nonphysiological conditions. Currently, there is no adequate molecular explanation for these observations.

Conclusions

The general structural and mechanistic molecular principles by which the F-ATPases from eubacteria, chloroplasts, and

mitochondria make ATP are now well established. However, it is apparent that the F-ATPases from these three sources differ substantially in the details of their operation, for example, in the amount of energy required to generate an ATP molecule, in the structures and possibly functions of their peripheral stalks, and in their modes of regulation. Thus, in order to gain a complete understanding of these enzymes, it will be necessary to establish separately the details of their operation in each class.

See also: **Bioenergetics:** ATP in Plant Mitochondria: Substrates, Inhibitors, and Uncouplers; P-Type Pumps: Na^+, K^+ -ATPase; Structure of P-Type Adenosine Triphosphatases; V-ATPases.

Further Reading

- Abrahams JP, Leslie AGW, Lutter R, and Walker JE (1994) Structure at 2.8 Å resolution of F_1 -ATPase from bovine heart mitochondria. *Nature* 370: 621–668.
- Boyer PD (1993) The binding change mechanism for ATP synthase – some probabilities and possibilities. *Biochimica et Biophysica Acta* 1140: 215–250.
- Gibbons CMG, Montgomery MG, Leslie AGW, and Walker JE (2000) Structure of the central stalk in bovine F_1 -ATPase. *Nature Structural and Biology* 7: 1055–1061.
- Gledhill JR, Montgomery MG, Leslie AGW, and Walker JE (2007) How the regulatory protein, IF_1 , inhibits F_1 -ATPase from mitochondria. *Proceedings of the National Academy of Sciences of the United States of America* 104: 15671–15676.
- Itoh H, Takahashi A, Adachi K, et al. (2004) Mechanically driven ATP synthesis by F_1 -ATPase. *Nature* 427: 465–468.
- Junge W, Sialaff H, and Engelbrecht S (2009) Torque generation and elastic power transmission in the rotary F_0F_1 -ATPase. *Nature* 459: 364–370.
- Nicholls DG and Ferguson SJ (2002) *Bioenergetics 3*. San Diego, CA: Academic Press.
- Rees DM, Leslie AGW, and Walker JE (2009) The structure of the membrane extrinsic region of bovine ATP synthase. *Proceedings of the National Academy of Sciences of the United States of America* 106: 21597–21601.
- Stock D, Leslie AGW, and Walker JE (1999) Molecular architecture of the rotary motor in ATP synthase. *Science* 286: 1700–1705.
- Watt IN, Montgomery MG, Runswick MJ, Leslie AGW, and Walker JE (2010) Bioenergetic cost of making an adenosine triphosphate molecule in animal mitochondria. *Proceedings of the National Academy of Sciences of the United States of America* 107: 16823–16827.
- Yagi H, Kajiwaru N, Tanaka H, et al. (2007) Structures of the thermophilic F_1 -ATPase ϵ -subunit suggesting ATP regulated arm motion of its C-terminal in F_1 . *Proceedings of the National Academy of Sciences of the United States of America* 104: 1123–1128.

Fatty Acid Metabolism and Cancer

F P Kuhajda[†], Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

AMP-activated kinase (AMPK) AMPK is phosphorylated (activated) by AMP-activated kinase when the AMP/ATP ratio increases. pAMPK phosphorylates and reduces acetyl-CoA carboxylase activity reducing fatty acid synthesis.

Enzyme-linked immunosorbent assay (ELISA) A commonly used immunological method to measure a target antigen in blood or any other biological sample. There are variations of the ELISA format but, generally, an antibody absorbed to a plastic well used to capture the antigen and a second antibody, which binds to a distinct epitope, is utilized as a detector. The detector antibody is often biotinylated to be visualized by an avidin-based colorimetric system.

Gleason grade A morphology-based grading system of prostate-cancer grading, based upon the growth pattern of the tumor, which is nearly universally applied to the clinical management of prostate cancer.

HER2/*neu* A proto-oncogene encoded by the ERBB2 gene; it has also been designated as CD340. Approximately,

15–20% of breast cancers amplify the HER2/*neu* gene or otherwise overexpress the protein. Patients whose breast cancers overexpress HER2/*neu* have a worse prognosis.

There is a relationship between HER2/*neu* and fatty acid synthase expression in human breast cancer.

Tumor heterogeneity When human cancers are studied with immunohistochemistry for biomarker expression, nearly any antigen will show varied amounts of expression in regions of the tumor and variation from cell to cell, which is defined as tumor heterogeneity.

Xenograft A cancer treatment model where human cancer cells are injected into an immunodeficient animal – usually athymic or severe combined immunodeficient (SCID) mouse. The cells are usually injected subcutaneously in the flank where growth and response to treatment can be measured with calipers. Orthotopic xenografts describe a model where the human cancer cells are injected into the same primary tissue in the mouse, that is, human lung cancer cells are injected into the trachea of mice, and human ovarian cancer cells are injected into mouse ovaries.

One of the most consistent biochemical changes associated with cancer is its reliance upon aerobic glycolysis described in the 1920s by Otto Warburg. Glucose metabolism continues to be exploited as a treatment strategy for cancer, based upon his discovery. More recently, cancer-related changes in fatty acid metabolism have provided an additional biochemical pathway replete with enzyme targets showing promise for both cancer diagnosis and therapy. Moreover, changes in fatty acid metabolism appear early during cancer development. Thus, these biochemical changes may play a role in promoting cellular transformation. In this article, fatty acid metabolism is reviewed as it relates to human cancer, highlighting fatty acid synthesis, acyl-CoA synthetase (ACS), and racemase. These have been selected because they represent targets for cancer diagnosis or treatment.

Overview of the Fatty Acid Synthesis Pathway

Within the last decade, most human cancers have been found to undergo high levels of *de novo* fatty acid synthesis compared to surrounding normal tissues. Enzymes in this pathway have become targets for both cancer therapy and diagnosis. At first, cancer cells undergoing high levels of fatty acid synthesis was a surprising and controversial concept, because our understanding of the role of fatty acid synthesis was borne out of studies in liver and other normal lipogenic organs.

In lipogenic organs, such as liver and adipose tissue, fatty acid

synthesis converts excess carbohydrate calories to fatty acids, which are ultimately stored as triglycerides. The triglyceride droplets in liver can grow to occupy a substantial mass of the hepatocyte, the so-called large droplet steatosis seen in fatty liver disease. In contrast, most cancer cells do not store appreciable amounts of fat as triglyceride. Renal cell carcinoma and liposarcomas are the only two human cancers with appreciable storage triglyceride, a feature used in their diagnosis. Although the basic biochemical pathway is present in both transformed and normal cells, fatty acid synthesis functions to store energy in the liver, while in cancer cells, it appears to maintain energy balance.

More than 20 enzymes contribute to the synthesis of fatty acids from glucose to palmitate. [Figure 1](#) outlines the pathway, highlighting the key enzymes of interest to oncology research. A brief summary of pathway function in lipogenic cells provides the background for its function in cancer. It is to be borne in mind, however, that in lipogenic tissues such as liver, this pathway is highly regulated by diet, whereas in the cancer cell, dietary regulation is minimal.

In the fed state, when energy is surfeit, glucose enters the mitochondria and exits as citrate to provide carbon for fatty acid synthesis. In the cytoplasm, citrate is converted to acetyl-CoA, a key substrate for fatty acid synthesis, by citrate lyase. Malonyl-CoA, the main substrate providing carbon for fatty acid synthase (FASN), is synthesized by acetyl-CoA carboxylase (ACC), in an adenosine triphosphate (ATP)-dependent reaction. ACC is the pace-setting enzyme of the fatty acid synthesis pathway. ACC activity is rapidly regulated by a number of substrates and hormones allowing the organism to adapt rapidly to

[†]Deceased.

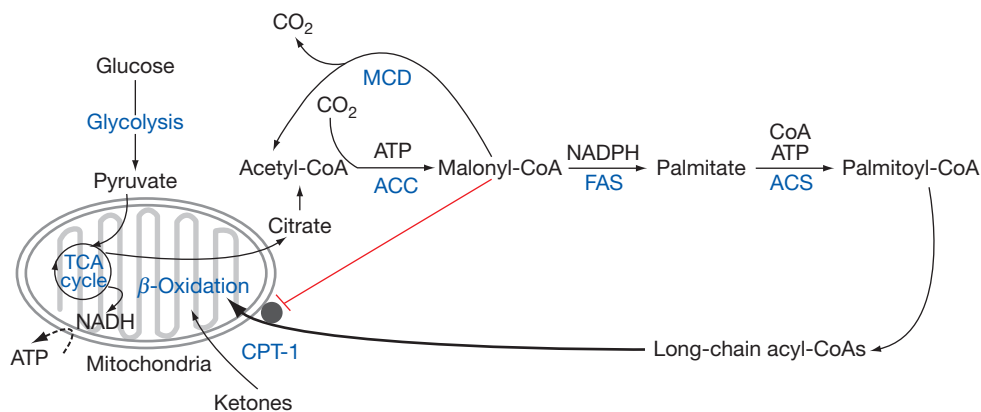


Figure 1 Fatty acid synthesis pathway from glucose to palmitate. This pathway schematic highlights enzymes and substrates that are of interest to fatty acid metabolism and cancer. See text for details.

the energy demand. ACC activity is increased by citrate, insulin, and ATP, and is decreased by palmitoyl CoA, glucagon, and adenosine-monophosphate-activated protein kinase (AMPK), which phosphorylates ACC to inactivate it. During times of energy excess, citrate leads to an aggregation of ACC molecules, which provides a positive feedforward stimulation of ACC activity. When energy is in surplus, the decreased AMP/ATP ratio inactivates AMPK, preventing the phosphorylation and inactivation of ACC, and allowing fatty acid synthesis to proceed.

FASN is the enzyme responsible for the reductive synthesis of fatty acids from acetyl-CoA and malonyl-CoA. Its predominant (80%) product is the 16-carbon saturated free fatty acid, palmitate, which must be esterified to CoA prior to its conversion to triglyceride or its use in other lipid moieties. ACS esterifies free fatty acids to CoA whether derived from the diet or *de novo* synthesis. There is a family of ACS enzymes, which can esterify short- to very-long-chain fatty acids to CoA, which is discussed in more detail in another section. The final fate of most of the endogenously synthesized fatty acids in liver and adipose tissue is esterification to glycerol as storage triglycerides. In mammary gland fatty acid synthesis, there is a specialized thioesterase that cleaves the fatty acids from FASN to produce, predominantly, the 10-carbon free fatty acid, decanoate. Decanoate is present in human breast milk and is easier for the infant to digest than long-chain fatty acids such as palmitate.

As a consequence of a high-fat diet, vigorous exercise, prolonged fasting, or outright starvation, fatty acid synthesis, along with other forms of reductive synthesis, must be reduced to conserve energy. Rapid regulation of the pathway occurs through citrate and AMPK. As ingestion of carbohydrates is reduced, citrate efflux from the citric acid cycle to the cytoplasm declines, leading to decreased ACC activity. As the AMP/ATP ratio rises, AMPK is activated by phosphorylation. pAMPK, in turn, phosphorylates and inactivates ACC, further diminishing fatty acid synthesis. If starvation continues, FASN expression would be highly downregulated in the liver within 24 h.

Whereas the starvation state requires fatty acid synthesis to cease, it concomitantly entails an increase in fatty acid oxidation. The rapid regulation of fatty acid oxidation is conveniently provided by malonyl-CoA, the substrate of FASN. Malonyl-CoA, the product of ACC, plays an important role in

cellular energy regulation. In addition to acting as a substrate for FASN, malonyl-CoA regulates fatty acid oxidation through its allosteric inhibition of carnitine palmitoyl-transferase-1 (CPT-1), the rate-limiting transporter of long-chain fatty acids into the mitochondria for oxidation. Thus, when the fatty acid synthesis pathway is active, the high steady-state levels of malonyl-CoA inhibit CPT-1, preventing the futile cycle of oxidation of newly synthesized fatty acids. To further maintain physiological levels of malonyl-CoA, malonyl-CoA decarboxylase (MCD) catabolizes malonyl-CoA back to acetyl-CoA, preventing accumulation of malonate, which is toxic to cellular respiration.

The fatty acid synthesis pathway acts as a fulcrum between states of energy abundance and starvation. It is therefore plausible that it could act to maintain energy balance in the cancer cell with its dysregulated drive to replicate, survive, and metastasize.

FASN Function in Cancer

FASN is the most studied enzyme of this pathway in cancer cells. Cytoplasmic or mammalian type I FASN contains seven distinct enzyme activities on a single ~250-kDa polypeptide, which functions as a dimer producing two molecules of free palmitate. FASN is represented by a single gene without isoforms and catalyzes the nicotinamide adenine dinucleotide phosphate (NADPH)-dependent synthesis of palmitic acid, a 16-carbon saturated free fatty acid. One acetyl-CoA and seven malonyl-CoA contribute the carbon for palmitate synthesis. The basic structure and function of this enzyme do not appear to be altered in cancer cells; the specific activity of FASN purified from cancer cells is similar to that from nontransformed cells, and palmitic acid is its predominant product.

In normal subjects, FASN expression and activity are relatively restricted to organs without rapidly proliferating cellular compartments such as liver, adipose tissue, lactating mammary gland, and brain. The exception is the high levels of FASN expression in the endometrium during the proliferative phase. FASN is not expressed in skeletal, cardiac, or smooth muscle; only limited expression has been identified in other organs with more rapidly proliferating cellular compartments such as bone

marrow and gastrointestinal tract. The relative lack of expression in normal proliferating cellular compartments makes it an attractive target for cancer treatment.

FASN expression in human cancer

About 15 years ago (mid-1990s), high levels of FASN were reported in human breast cancer cells. Since then, high levels of FASN expression have been found in most human carcinomas, with expression also detected in sarcomas and some hematological malignancies, using a variety of techniques including immunohistochemistry and messenger RNA (mRNA) detection. Immunohistochemistry demonstrates cytoplasmic localization of FASN in human colon cancer as seen in [Figure 2](#). Not all primary cancers of a particular cancer subtype type express FASN, nor do all the cells in a particular tumor express the same amount of the enzyme which is characteristic of tumor heterogeneity. Comparing FASN expression in cancers to their biological behavior has shown a relationship between FASN expression and tumor virulence, which is discussed next.

The regulation responsible for the high levels of FASN expression in cancer is complex and incompletely understood. Cancer cells do not exhibit the dramatic dietary regulation of FASN expression as the liver. For example, at autopsy, FASN expression is often low to absent in the liver due to cachexia and inanition, but metastatic tumor deposits show abundant FASN expression by immunohistochemistry. In tumors whose growth is dependent upon androgens or estrogen/progesterone, such as breast or prostate cancer, FASN expression is largely driven by sex steroid hormones. In prostate cancer and human prostate cancer cell lines, FASN is positively regulated by androgens. In xenograft studies of human prostate cancer, high levels of FASN expression in the xenografts fell following castration. When the animals were treated with exogenous testosterone, FASN expression increased to precastration levels. In long-term castrated animals with relapsed androgen-independent tumors, FASN expression again increased. These data in xenografts were reflected in studies of human disease. Following therapeutic orchiectomy, tumor metastases shrink

and FASN expression is markedly reduced in the remaining prostate tissue consistent with androgen control of FASN expression in the prostate. However, when the tumor reappeared years later as androgen-independent disease, FASN expression was high in metastatic deposits studied at autopsy. Thus, it is likely that FASN expression in lethal prostate cancer, following androgen ablation, reappears under control of androgen-independent mechanisms.

In cancers whose growth is independent of sex steroid hormones, such as lung or colon cancer, the transcriptional regulation of FASN is less well understood. In these tumors, fatty acid synthesis is increased likely as a result of the actions of oncogenes. For example, phosphorylated (activated) Akt phosphorylates and activates ATP citrate lyase, which provides acetyl-CoA for fatty acid synthesis. In addition, pAkt through its inhibition of the tuberous sclerosis complex (TSC1/TSC2) and subsequent inhibition of Rheb, increases mammalian target of rapamycin (mTOR) activity. Activated mTOR increases glycolysis and protein synthesis. As a secondary consequence of increased flux through glycolysis, the pentose phosphate pathway provides ribose for nucleic acid synthesis and NADPH for enhanced fatty acid synthesis. Thus, the combination of pAKT and mTOR leads to a cascade of events including increased glycolysis (aerobic glycolysis), protein and fatty acid synthesis. Thus, FASN expression, in some instances, is increased because of a programmatic shift in cellular metabolism, as a result of Akt activation. Alternatively, in a study of human prostate cancer, fluorescence *in situ* hybridization identified a gene copy gain for FASN of 24% in all prostate cancers. In cases where FASN expression was increased as determined by immunohistochemistry, an increase in gene copy number occurred in 59% of the specimens. Thus, in addition to androgens, increased gene copy number may also contribute to increased FASN expression in prostate and other cancers.

FASN and cancer diagnostics

Although high levels of FASN expression occur commonly in human cancers, not all cancers have uniformly high levels of expression. FASN expression levels in breast, prostate, pancreas, and ovarian cancer tissues have been shown to have prognostic value, as increased expression was associated with more clinically virulent disease. In prostate cancer, high levels of FASN expression were associated with extension of cancer beyond the prostate and into the seminal vesicle, which indicates a high likelihood of systemic disease that is not yet clinically detectable. Others have found a relationship between FASN expression and prostate cancer high Gleason grade and disease progression.

FASN has also been detected in the blood of patients with breast, colon, prostate, ovarian, and pancreas cancer at levels 5–60-fold of those found in normal subjects using an enzyme-linked immunosorbent assay (ELISA). FASN in the blood originates from cancer cells as blood levels decline postresection of the cancer. The mechanism whereby FASN is elaborated from the cancer cell into the blood is unknown. FASN is not a secretory protein, as it does not contain the requisite signal peptide. Recently, however, cancer cells have been shown to shed exosomes, 40–100-nm-diameter nanovesicles of endocytic origin. Proteomic analysis of exosomes from human colon and bladder cancer has detected FASN sequences, which may

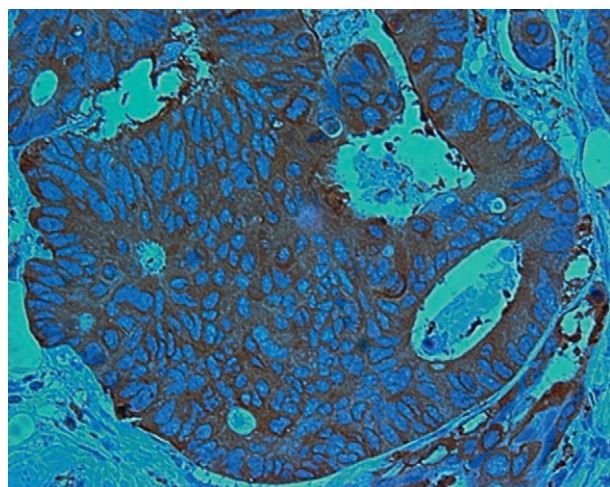


Figure 2 FASN expression in human colon cancer. Immunohistochemistry localizes FASN to the cancer cell cytoplasm (brown staining) in human colon cancer (magnification $\times 400$).

account for its presence in the serum. FASN has also been detected in bovine and human milk as part of the milk fat globule; this form of protein elaboration may occur in human breast cancers, which produce milk proteins. Alternatively, FASN may also be elaborated from apoptotic cancer cells and nonspecifically enter the circulation.

Thus, in the realm of cancer diagnosis, FASN expression in tissues may aid in the pathological diagnosis of cancer and assist in determining prognosis. The clinical utility of serum FASN determination is under active investigation. Serum FASN levels may prove to be helpful in monitoring cancer therapy or perhaps serve to aid in cancer detection in the population at large.

FASN as a therapeutic target

As it was believed that cancer cells have access to adequate amounts of fatty acid from the blood, the detection of robust FASN activity in cancer cells was surprising. As the pathway consumes ATP and large amounts of anabolic reducing equivalents, it must provide some vital function for cancer cells. Otherwise, FASN-negative clones would likely have both an energetic and a synthetic advantage over FASN-positive clones, leading ultimately to predominantly FASN-negative tumors. Indeed, quite the opposite is true; as cancers progress, they often express higher levels of FASN.

Thus, the combination of both basic biochemical and clinical observations led to studies of fatty acid synthesis inhibition in human cancer cells. With its seven distinct enzymes, FASN has been the focus of small-molecule pharmacological inhibition. Early studies of FASN inhibition employed a natural product, cerulenin, which inhibits both mammalian (type I) and prokaryotic (bacterial) type II FASN. Cerulenin forms a covalent adduct with the ketoacylsynthase domain of both type I FASN and type II KAS, acting as a suicide inhibitor with a low micromolar IC_{50} against purified FASN from human cancer cells. As might be predicted from FASN expression studies, most human cancer cell lines undergo brisk fatty acid synthesis in culture, even in the presence of 10% fetal calf serum, which provides sufficient fatty acid for cancer cell growth. Nonetheless, cerulenin was cytotoxic to human cancer cells *in vitro* with LC_{50} s in the low micromolar range for most cell lines. Many types of human cancer cells were susceptible to cerulenin-induced apoptosis including breast, prostate, colon, and ovarian cancer cell lines. In addition to cultured cells, cerulenin had activity *in vivo* against human cancer cell grown in athymic mice. However, the chemical instability and reactivity of cerulenin precluded its further clinical development.

C75 was the first synthetic small-molecule inhibitor of FASN. C75 was cytotoxic to many types of human cancer cell lines *in vitro* and showed significant antitumor activity against human cancer xenografts. However, substantial weight loss in the mice limited its utility, and it was found that the weight loss was due to C75 stimulation of CPT-1 activity leading to enhanced fatty acid oxidation. More recently, pharmacological inhibition of FASN has employed new chemical scaffolds and other enzyme targets on FASN. Thiolactomycin, another prokaryotic natural product FASN inhibitor, which occupies a site on the beta ketoacylsynthase distinct from cerulenin, has led to a wealth of new FASN inhibitors. Compounds targeted to the enoyl reductase domain on FASN have also afforded new

opportunities for synthetic chemistry. Drugs approved for other uses, such as orlistat, a pancreatic lipase inhibitor, have been shown to have anti-FASN activity as does triclosan, a type II FASN inhibitor used widely in personal care products. Although there is great variation in potency, these myriad of FASN inhibitors have shown activity *in vitro*, against human cancer cell lines derived from diverse primary sites. In addition to pharmacological inhibition, siRNA against FASN has been shown to be cytotoxic to human cancer cells *in vitro*.

A variety of pharmacological FASN inhibitors have shown activity *in vivo* against human cancer xenografts in athymic or SCID mice. There have been at least 12 reports of FASN inhibitors exhibiting antitumor activity against the following human cancer xenografts: breast, lung, prostate, colon, ovary, thyroid, malignant mesothelioma, and leukemia. In a study of mice bearing human non-small-cell carcinoma xenografts, oral treatment with a synthetic FASN inhibitor for 2 weeks led to substantial reduction in tumor growth. Similar results were obtained with orthotopic (lung) xenografts. These FASN inhibitors did not increase fatty acid oxidation and as such did not induce substantial weight loss. Signs of toxicity in the mice were not identified. FASN inhibition has led to the reduction of tumor growth of a number of diverse human cancer xenografts suggesting its utility as a target for cancer therapy.

In addition to established cancers, high levels of FASN expression have been found in a number of premalignant lesions in the breast, prostate, colon, and oropharynx. Using the prostate as an example, high FASN expression was found in prostatic intraepithelial neoplasia (PIN), thought to be a precursor lesion to invasive prostate cancer. Moreover, FASN expression has also been identified in histologically normal prostatic epithelium adjacent to invasive cancer, which may indicate a field effect of normal-appearing tissue at risk for development of cancer. Based on these and other observations, animal models of genetically induced cancer or chemical carcinogenesis have been treated with FASN inhibitors to prevent the development of neoplasia. FASN inhibitors substantially reduced mammary cancer development in transgenic HER2/*neu* mice. In this model, HER2/*neu*, known to upregulate FASN expression, was restricted to the mammary epithelium. FASN inhibition delayed the growth of transformed breast epithelium and led to complete abrogation of cancer growth in about 25% of the mice. The effects on the mice were limited to the mammary gland. In mice treated with chemical carcinogens to induce lung cancer, FASN inhibitors were administered orally twice a day for 4 months. These daily treatments led to a dramatic reduction of lung cancers in both urethane and NKK [4-(methylnitrosamino)-1-(3-pyridyl)-1- butanone] induced tumors. Interestingly, pAkt expression was reduced in the tumors of these FASN-inhibitor-treated animals compared to controls. In addition, lung cancer incidence was significantly reduced in rat models of chemical carcinogenesis. Thus, FASN inhibition is cytotoxic to established cancers and is able to prevent the development on cancers known to have high levels of FASN expression.

Mechanism of action of FASN inhibition

The *in vitro* and *in vivo* anticancer activity of FASN inhibition is well documented; however, the mechanism of action surrounding its cytotoxicity is under active investigation. What

survival advantage FASN expression and activity affords the cancer cell is poorly understood and likely differs depending on the primary site and morphological type of cancer. It has been postulated that reductive fatty acid synthesis helps the cancer cell maintain redox balance in the low oxygen tension found in human tumors by transferring electrons to fatty acids rather than oxygen. Indeed, FASN expression is upregulated by hypoxia, and FASN is highly expressed in hypoxic regions of tumors. Caveolin-1 (Cav-1) is a palmitoylated lipid raft protein that has been implicated in prostate cancer progression. Cav-1 and FASN have been shown to be coordinately regulated in prostate cancer and FASN may form a complex with Cav-1 based on co-immunoprecipitation. In Cav-1^{-/-}/TRAMP mice, FASN was not upregulated in the prostate tumors and the cancers were smaller and less virulent based upon a reduction in their ability to metastasize. Recent studies have found a relationship between high levels of FASN expression and reduced level of intrinsic apoptosis in human prostate cancer. In breast cancer, HER2/*neu* has been shown to be coordinately expressed with FASN, which may act together to support tumor growth.

Pharmacological studies have studied the mechanism of action responsible for the antitumor activity of FASN inhibition. There are a myriad of events that occur following FASN inhibition related to cell proliferation, survival, and apoptosis. Many of these effects occur hours following treatment. Within minutes of pharmacological FASN inhibition, there is an increase in the AMP/ATP ratio and reduction of cellular redox power. These changes in cellular bioenergetics lead to a rapid phosphorylation of AMPK, which attempts to reduce cellular proliferation and conserve energy. Although many of the changes of FASN inhibition in cancer cells are likely contextual to the cell type and its culture conditions, it appears that the FASN pathway functions to maintain energy balance in the cancer cell. Once the energy balance is disrupted, the dysregulation of cellular proliferation inherent to many types of cancer cells leads to eventual cell death.

If FASN inhibition triggers energy imbalance in the cancer cell, what is the trigger? Some clues have emerged from pharmacological studies of ACC inhibition. It has been found that 5-tetradecyloxy-2-furoic acid (TOFA) is a competitive inhibitor of ACC, which is used to block the synthesis of malonyl-CoA. TOFA is a potent inhibitor of the FASN pathway. If cancer cells are treated with concentrations of TOFA or an FASN inhibitor that lead to a similar amount of pathway inhibition, the FASN inhibitor is far more toxic than the ACC inhibitor. This suggests that the elevated concentration of malonyl-CoA following FASN inhibition may be involved in its cytotoxic effect. In patients with an absence or inactivating mutation of MCD, there are high levels of free malonate, which are toxic to cellular metabolism. These infants fail to thrive and suffer from fatal malonic aciduria, and seizures. Malonate is a potent inhibitor of succinyl-CoA dehydrogenase, a member of both the citric acid cycle and complex II of the electron transport chain. Thus, one hypothesis is that pharmacological FASN inhibition could lead to a rapid but transient elevation of malonyl-CoA that overwhelms the baseline activity of MCD. The resulting malonate toxicity could lead to ATP reduction and subsequent metabolic failure in the proliferating cancer cell. Based on this hypothesis, MCD has been suggested as another potential

anticancer target. Both siRNA and pharmacological inhibition of MCD have been shown to be cytotoxic to human breast cancer cells *in vitro*. Moreover, pharmacological MCD inhibition potentiates FASN inhibition in human breast cancer cells. Although the understanding of the biochemical mechanism is yet incomplete, there is increasing evidence that FASN inhibition has selective antitumor activity, which likely involves energy balance disruption in the cancer cell.

Acyl-CoA Synthetase

Fatty acids must be esterified to CoA for any further metabolism, which is the function of the ACS. In humans, the CS family contains 26 genes, which activate short-, medium-, long-, or very-long-chain fatty acids. Of the six-member very long chain ACS family, ACSVL3 is present at low levels in neurons, but cannot be detected in glia. Recently, however, human malignant gliomas have been found to express high levels of ACSVL3. Knockdown of ACSVL3 expression has been shown to decrease the growth of malignant glioma cells *in vitro* and in xenografts. This effect is likely mediated through modulation of Akt function. As ACSVL3 expression has not been detected in normal murine cells, it is hypothesized that it might represent another metabolic target for cancer therapy.

Recently, long-chain fatty acyl-CoA synthetase 4 (ACSL4) has been found to be highly expressed in subsets of breast cancer and in prostate cancer. ACSL4 expression is overexpressed in estrogen-receptor negative breast cancers. The inverse relationship between sex-steroid-hormone-receptor expression and ACSL4 was also found in prostate cancer, where androgen-receptor negative cancers had upregulated ACSL4 expression. The function of this isoform in cancer has not yet been determined, but sensitivity of cancer cells to inhibitors of ACS was increased threefold in cancer cells with low levels of ACSL4 expression.

A-Methylacyl-CoA Racemase

A-Methylacyl-CoA racemase (AMACR) functions in normal cells to enable oxidation of branched-chain fatty acids. AMACR catalyzes the chiral inversion of a variety of 2-methyl fatty acyl-CoAs and thereby regulates their entry into the peroxisomal or mitochondrial oxidation pathways. Interestingly, AMACR is involved in the metabolism of phytanic acid, a dietary branched-chained fatty acid associated with the development of prostate and other cancers. AMACR was identified as a gene highly upregulated in prostate cancer through analysis of microarray data sets. A large multi-institutional study demonstrated that AMACR helped distinguish benign from cancer prostate tissue with 97% sensitivity and 92% specificity. AMACR has become a mainstay of diagnostic pathology along with the basal cell marker p63, to aid in the identification of small foci of prostate cancer on needle biopsies. A recent study has shown that AMACR and FASN immunostaining was superior to AMACR alone to detect cancer in needle biopsies of the prostate. In addition, AMACR has been shown to be prognostic in prostate cancer, as low levels of expression are found in more-aggressive disease.

See also: **Metabolism Vitamins and Hormones:** Fatty Acid Oxidation; Fatty Acid Structure and Synthesis; **Signaling:** Fibroblast Growth Factor Receptors and Cancer-Associated Perturbations; Tumor Necrosis Factor Receptors.

Further Reading

- Alli PM, Pinn ML, Jaffee EM, McFadden JM, and Kuhajda FP (2005) Fatty acid synthase inhibitors are chemopreventive for mammary cancer in neu-N transgenic mice. *Oncogene* 24: 39–46.
- Di Vizio D, Sotgia F, Williams TM, et al. (2007) Caveolin-1 is required for the upregulation of fatty acid synthase (FASN), a tumor promoter, during prostate cancer progression. *Cancer Biology and Therapy* 6: 1263–1268.
- Furuta E, Pai SK, Zhan R, et al. (2008) Fatty acid synthase gene is up-regulated by hypoxia via activation of Akt and sterol regulatory element binding protein-1. *Cancer Research* 68: 1003–1011.
- Jiang Z, Wu CL, Woda BA, et al. (2004) Alpha-methylacyl-CoA racemase: A multi-institutional study of a new prostate cancer marker. *Histopathology* 45: 218–225.
- Kuhajda FP (2006) Fatty acid synthase and cancer: New application of an old pathway. *Cancer Research* 66: 5977–5980.
- Kuhajda FP, Jenner K, Wood FD, et al. (1994) Fatty acid synthesis: A potential selective target for antineoplastic therapy. *Proceedings of the National Academy of Sciences of the United States of America* 91: 6379–6383.
- Kuhajda FP, Pizer ES, Li JN, et al. (2000) Synthesis and antitumor activity of an inhibitor of fatty acid synthase. *Proceedings of the National Academy of Sciences of the United States of America* 91(97): 3450–3454.
- Menendez JA and Lupu R (2007) Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nature Reviews* 7: 763–777.
- Migita T, Ruiz S, Fornari A, et al. (2009) Fatty acid synthase: A metabolic enzyme and candidate oncogene in prostate cancer. *Journal of the National Cancer Institute* 101: 519–532.
- Orita H, Coulter J, Lemmon C, et al. (2007) Selective inhibition of fatty acid synthase for lung cancer treatment. *Clinical Cancer Research* 13: 7139–7145.
- Orita H, Coulter J, Tully E, Kuhajda FP, and Gabrielson E (2008) Inhibiting fatty acid synthase for chemoprevention of chemically induced lung tumors. *Clinical Cancer Research* 14: 2458–2464.
- Pei Z, Sun P, Huang P, et al. (2009) Acyl-CoA synthetase VL3 knockdown inhibits human glioma cell proliferation and tumorigenicity. *Cancer Research* 69: 9175–9182.
- Pizer ES, Pflug BR, Bova GS, et al. (2001) Increased fatty acid synthase as a therapeutic target in androgen-independent prostate cancer progression. *Prostate* 47: 102–110.
- Pizer ES, Thupari J, Han WF, et al. (2000) Malonyl-coenzyme-A is a potential mediator of cytotoxicity induced by fatty-acid synthase inhibition in human breast cancer cells and xenografts. *Cancer Research* 60: 213–218.
- Wakil S (1989) Fatty acid synthase, a proficient multifunctional enzyme. *Biochemistry* 28: 4523–4530.

Fatty Acid Oxidation

H Schulz, City University of New York, New York, NY, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp 90–94, © 2004, Elsevier Inc.

Glossary

Activation of fatty acids The conversion of fatty acids into thioesters with coenzyme A.

Peroxisome Subcellular organelle that is enclosed by a single membrane.

Unsaturated fatty acids Fatty acids that contain one or more double bonds.

β -Oxidation Degradation of fatty acyl-coenzyme A thioesters by the sequential removal of two carbon atoms at a time in the form of acetyl coenzyme A.

Cellular Uptake and Activation of Fatty Acids

The uptake of fatty acids by animal cells seems to be a facilitated process even though fatty acids can diffuse unassisted across biological membranes. A number of proteins of the cellular membrane are thought to be involved in this process. However, their specific functions in fatty acid uptake have not been elucidated. Once fatty acids have crossed the cellular membrane, they either diffuse or are transported to mitochondria, peroxisomes, and the endoplasmic reticulum where they are activated by esterification with coenzyme A (CoA). The transfer of fatty acids from the cellular membrane to sites of their activation may be facilitated by fatty-acid-binding proteins (FABPs), a group of low-molecular-weight (14–15 kDa) proteins that are present in the cytosol of various animal tissues. The function of FABP as an intracellular carrier of fatty acids has not been proved nor other suggested roles of this protein in the uptake and storage of fatty acids have been established.

The cellular uptake of fatty acids is tightly coupled to their intracellular esterification with CoA. Acyl-CoA synthetases catalyze the adenosine triphosphate (ATP)-dependent esterification of fatty acids and CoA to produce fatty acyl-CoA, AMP, and pyrophosphate. Long-chain-specific acyl-CoA synthetases are associated with the outer mitochondrial membrane and the membranes of the endoplasmic reticulum and peroxisomes. Acyl-CoA synthetases with a preference for medium-chain fatty acids are mostly located in the matrix of mitochondria while short-chain specific acyl-CoA synthetases are found in the mitochondrial matrix and/or in the cytosol depending on the organ of the animal.

Fatty Acid Oxidation in Mitochondria

In animal cells, fatty acids are degraded both in mitochondria and peroxisomes, whereas in lower eukaryotes fatty acid oxidation is generally confined to peroxisomes. Acetyl-CoA formed by mitochondrial fatty acid oxidation is further oxidized to release energy that drives the formation of ATP by oxidative phosphorylation. In liver, acetyl-CoA is also utilized for the synthesis of ketone bodies.

Mitochondrial Uptake of Fatty Acids

Fatty acids are activated by their conversion to fatty acyl-CoA thioesters at the outer mitochondrial membrane while their degradation by β -oxidation takes place in the mitochondrial matrix. Because acyl-CoAs cannot efficiently cross the inner mitochondrial membrane, an uptake system exists that shuttles fatty acyl residues across the inner membrane. As illustrated in [Figure 1](#), the acyl residue of fatty acyl-CoA is transferred from CoA to carnitine at the outer mitochondrial membrane catalyzed by carnitine palmitoyltransferase I (CPT I). The resultant fatty acylcarnitine can cross the inner mitochondrial membrane in exchange for carnitine. This exchange is catalyzed by carnitine:acylcarnitine translocase (T). Once fatty acylcarnitine is in the mitochondrial matrix, it is converted back to fatty acyl-CoA by carnitine palmitoyltransferase II (CPT II), which is associated with the inner mitochondrial membrane. Medium-chain and short-chain fatty acids are not activated in the cytosol but directly enter the mitochondrial matrix where they are converted to their acyl-CoAs by medium-chain and short-chain acyl-CoA synthetases, respectively.

β -Oxidation in Mitochondria

Fatty acyl-CoAs in the mitochondrial matrix are substrates of β -oxidation. This cyclic process consists of four reactions that shorten the acyl chain of fatty acyl-CoA by two carbon atoms at a time and produce acetyl-CoA. As shown in [Figure 2](#), reaction 1a, β -oxidation begins with the removal of two hydrogens from carbon atoms 2 (α -carbon) and 3 (β -carbon) of acyl-CoA to form the corresponding 2-*trans*-enoyl-CoA. The hydrogens are initially accepted by the flavine adenine dinucleotide (FAD) cofactor of acyl-CoA dehydrogenase, the enzyme that catalyzes this reaction, then are transferred to the FAD of a soluble matrix enzyme named electron transferring flavoprotein, and finally are fed into the mitochondrial electron transport chain, which supplies the energy for ATP formation by oxidative phosphorylation. In the second β -oxidation reaction, water is added across the double bond of 2-*trans*-enoyl-CoA by enoyl-CoA hydratase to yield

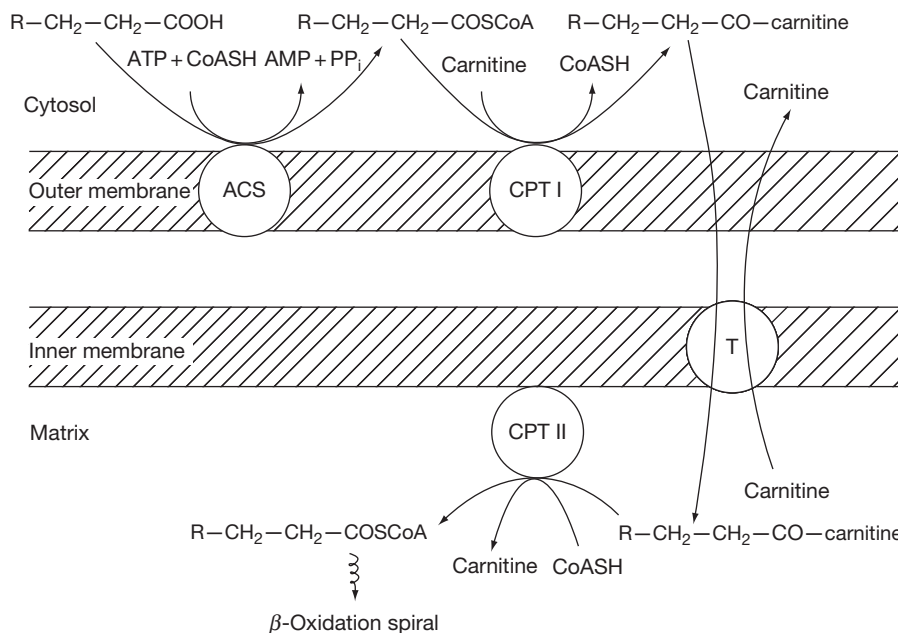


Figure 1 Carnitine-dependent transfer of fatty acyl groups across the inner mitochondrial membrane. ACS, acyl-CoA synthetase; CPT I and CPT II, carnitine palmitoyltransferase I and II, respectively; T, carnitine:acylcarnitine translocase.

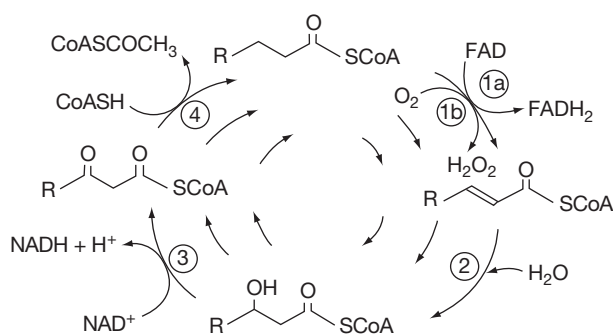


Figure 2 β -Oxidation of fatty acyl-CoA. Reactions 1a, 1b, 2, 3, and 4 are catalyzed by acyl-CoA dehydrogenase, acyl-CoA oxidase, enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase, and 3-ketoacyl-CoA thiolase, respectively. Abbreviations: FAD, oxidized form of flavine adenine dinucleotide; FADH_2 , reduced form of flavine adenine dinucleotide; NAD^+ , nicotinamide adenine dinucleotide.

L-3-hydroxyacyl-CoA (see **Figure 2**, reaction 2). The third reaction is the nicotinamide adenine dinucleotide (NAD^+)-dependent dehydrogenation of L-3-hydroxyacyl-CoA catalyzed by L-3-hydroxyacyl-CoA dehydrogenase (see **Figure 2**, reaction 3). The product of this reaction, 3-ketoacyl-CoA, is the substrate of the fourth and last reaction of β -oxidation catalyzed by 3-ketoacyl-CoA thiolase, which cleaves the compound in the presence of CoA between carbons 2 and 3 to yield an acyl-CoA shortened by two carbon atoms and acetyl-CoA (see **Figure 2**, reaction 4). Repetitive cycles of β -oxidation result in the complete degradation of fatty acids to acetyl-CoA and also propionyl-CoA, if fatty acids with an odd-numbered number of carbon atoms are involved.

Two or more enzymes with different acyl chain length specificities participate in each of the four reactions. Four acyl-CoA dehydrogenases are required for the β -oxidation of fatty acids. Their names, very long chain acyl-CoA dehydrogenase, long-chain acyl-CoA dehydrogenase, medium-chain acyl-CoA dehydrogenase, and short-chain acyl-CoA dehydrogenase, reflect their preferences for substrates of different acyl chain lengths. Together they efficiently dehydrogenate all acyl-CoAs that are formed during the β -oxidation of most dietary fatty acids. Two enzymes, one with a preference for short-chain and medium-chain substrates and another most active with long-chain substrates, cooperate in each of the three other reactions to metabolize a range of β -oxidation intermediates of different acyl chain lengths.

β -Oxidation of Unsaturated Fatty Acids

Unsaturated and polyunsaturated fatty acids also are degraded by β -oxidation. However, additional reactions are required to metabolize pre-existing double bonds that would otherwise interfere with the complete β -oxidation of unsaturated fatty acids. All double bonds found in unsaturated fatty acids can be classified as either odd- or even-numbered double bonds. Both classes are present in linoleic acid, which contains an odd-numbered double bond at position 9 and an even-numbered double bond at position 12 (see **Figure 3**). The degradation of linoleic acid therefore illustrates the β -oxidation of all unsaturated fatty acids. As outlined in the figure, linoleoyl-CoA (I; all Roman numerals refer to compounds), after passing 3 times through the β -oxidation cycle, yields 3-*cis*,6-*cis*-dodecadienoyl-CoA (II). This metabolite cannot enter another cycle of β -oxidation due to interference by the double bond at position 3.

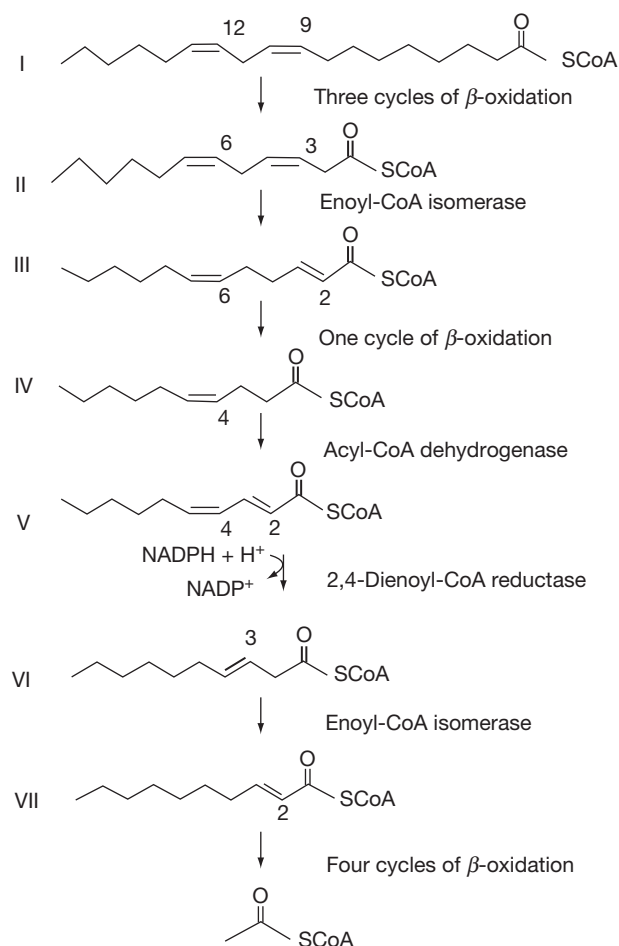


Figure 3 Pathway for the β -oxidation of linoleoyl-CoA.

However, an additional (auxiliary) enzyme, Δ^3, Δ^2 -enoyl-CoA isomerase (enoyl-CoA isomerase), converts 3-*cis*,6-*cis*-dodecadienoyl-CoA (II) to 2-*trans*,6-*cis*-dodecadienoyl-CoA (III) that can complete its pass through one cycle of β -oxidation to yield 4-*cis*-decenoyl-CoA (IV). 4-*cis*-Decenoyl-CoA is dehydrogenated by acyl-CoA dehydrogenase to 2-*trans*,4-*cis*-decadienoyl-CoA (V), which is not a substrate of β -oxidation but which is reduced by nicotinamide adenine dinucleotide phosphate (NADPH) in the presence of 2,4-dienoyl-CoA reductase, the second auxiliary enzyme, to 3-*trans*-decenoyl-CoA (VI). The latter compound is converted by enoyl-CoA isomerase to 2-*trans*-decenoyl-CoA (VII) that can be completely degraded by β -oxidation. A third auxiliary enzyme, $\Delta^{3,5}, \Delta^{2,4}$ -dienoyl-CoA isomerase (dienoyl-CoA isomerase), catalyzes the isomerization of 3,5-dienoyl-CoA to 2,4-dienoyl-CoA. This reaction facilitates the complete β -oxidation of 3,5-dienoyl-CoAs that are minor metabolites of unsaturated fatty acids with odd-numbered double bonds.

Regulation of Mitochondrial Fatty Acid Oxidation

The rate of fatty acid oxidation changes in response to the nutritional and hormonal state of the animal. The rate of

fatty acid oxidation is high during fasting but low in the fed animal. One cause for this change is the higher concentration of unesterified (free) fatty acids in the circulation of the fasting animal as compared to the concentration in the fed animal. An increased concentration of free fatty acids results in higher rates of cellular uptake and oxidation of fatty acids. In liver, which has high capacities for both synthesizing and oxidizing fatty acids, a reciprocal relationship exists between these two processes. After feeding, when carbohydrates are converted to triacylglycerols, the rate of fatty acid synthesis is high because acetyl-CoA carboxylase is active. This enzyme catalyzes the carboxylation of acetyl-CoA to malonyl-CoA, the first committed intermediate in fatty acid synthesis. Malonyl-CoA binds to and effectively inhibits CPT I that initiates the uptake of fatty acids by mitochondria. Thus, when fatty acids are rapidly synthesized, the cytosolic concentration of malonyl-CoA is high and the mitochondrial uptake and oxidation of fatty acids are inhibited. This situation is reversed during fasting when lower blood glucose levels cause the plasma concentration of the hormone glucagon to increase and that of insulin to decrease. Glucagon promotes the phosphorylation and inactivation of acetyl-CoA carboxylase with the result that the cytosolic concentration of malonyl-CoA declines. The lower concentration of malonyl-CoA causes fatty acid synthesis to decrease and fatty acid oxidation to increase. The same regulatory mechanism may be effective in tissues such as heart and skeletal muscle that oxidize fatty acids but do not synthesize them. Although malonyl-CoA is generated in these tissues by acetyl-CoA carboxylase, it seems to be metabolized by decarboxylation catalyzed by malonyl-CoA decarboxylase.

Fatty Acid Oxidation in Peroxisomes

Peroxisomes of animal cells contain a β -oxidation system that partially degrades fatty acids in contrast to mitochondria that break down fatty acids completely. The enzymes of peroxisomal and mitochondrial β -oxidation are not the same. Different genes encode most of the peroxisomal and mitochondrial enzymes. Substrates of peroxisomal β -oxidation are the CoA derivatives of fatty acids and carboxylic acids that are poorly or not at all taken up by mitochondria as, for example, very long chain fatty acids, methyl-branched carboxylic acids (e.g., pristanic acid), prostaglandins, dicarboxylic acids, xenobiotic compounds (e.g., phenyl fatty acids), and hydroxylated cholestanoic acids that are precursors of cholic acid. However, peroxisomes of lower eukaryotes catalyze the complete β -oxidation of fatty acids, whereas mitochondria in the same organisms generally are devoid of a functional β -oxidation system.

The uptake and activation of fatty acids by peroxisomes remain poorly understood. While some fatty acids such as very long chain fatty acids may be activated by conversion to CoA esters in the peroxisomal matrix after directly entering animal peroxisomes, other fatty acids or carboxylic acids seem to be activated outside of peroxisomes. The uptake of fatty acyl-CoAs by peroxisomes is suspected to take place but has not been demonstrated. Carnitine does not seem to be required for this uptake.

The four reactions of β -oxidation in peroxisomes are similar to the corresponding mitochondrial reactions (see Figure 2). An exception is the first reaction, in which fatty acyl-CoA is dehydrogenated to 2-*trans*-enoyl-CoA by acyl-CoA oxidase while oxygen is reduced to H_2O_2 (see Figure 2, reaction 1b). The latter product is converted to H_2O and O_2 by catalase that is present in peroxisomes. At least two acyl-CoA oxidases exist in animal peroxisomes; one named palmitoyl-CoA oxidase is active with long-chain and medium-chain, but not short-chain acyl-CoAs that have straight acyl chains while the second acyl-CoA oxidase also acts on branched-chain acyl-CoAs. Since the acyl-CoA oxidase activity of lower eukaryotes extends over the whole spectrum of acyl-CoAs of various chain lengths, peroxisomes in these organisms, in contrast to those in animals, can degrade fatty acids completely. The second and third reactions of peroxisomal β -oxidation are catalyzed by enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase, respectively, that are associated with a same polypeptide. Two such multifunctional enzymes, MFE 1 and MFE 2, are present in animal peroxisomes. They differ from each other in that the 3-hydroxyacyl-CoA intermediate formed and acted upon by MFE 1 has the L configuration, whereas the intermediate of MFE 2 is the D isomer. Peroxisomes of yeast and fungi only contain the type 2 multifunctional enzymes. Two types of 3-ketoacyl-CoA thiolases are present in rat peroxisomes. One of the thiolases, referred to as SCP_x -thiolase, is unusual because it is part of a bifunctional protein that additionally contains a lipid-binding segment named sterol carrier protein (SCP). Studies with β -oxidation mutants have led to the conclusion that the partial β -oxidation of straight-chain fatty acyl-CoAs, like very long chain fatty acyl-CoAs, in animal peroxisomes requires palmitoyl-CoA oxidase, MFE 2, and 3-ketoacyl-CoA thiolase, whereas the partial degradation of branched-chain acyl-CoAs, such as pristanoyl-CoA, involves branched-chain acyl-CoA oxidase, MFE 2, and SCP_x -thiolase. Unsaturated and polyunsaturated fatty acids also are partially degraded in peroxisomes. The necessary additional enzymes, enoyl-CoA isomerase, 2,4-dienoyl-CoA reductase, and even dienoyl-CoA isomerase, are present in this organelle. The products of peroxisomal β -oxidation are chain-shortened acyl-CoAs, NADH, and acetyl-CoA. In addition, propionyl-CoA is formed when methyl-branched-chain substrates are degraded. Acyl-CoAs may be converted to the corresponding acylcarnitines before they exit from peroxisomes. The necessary carnitine acyltransferases active with medium-chain and short-chain acyl-CoAs have been identified in peroxisomes. However, little is known about transporters that facilitate the efflux of products from peroxisomes.

α -Oxidation and ω -Oxidation

α -Oxidation is the process that results in the oxidative removal of the first carbon atom, the carboxyl group, of a fatty acid or carboxylic acid to yield CO_2 and a fatty acid or carboxylic acid shortened by one carbon atom. In animals, this process occurs in peroxisomes where, for example, phytanic acid, which is derived from phytol of chlorophyll and which cannot be degraded by β -oxidation, is chain-shortened by one carbon atom to pristanic acid that can be degraded by β -oxidation. ω -Oxidation is the process by which fatty acids are oxidized at their terminal (ω) or penultimate ($\omega - 1$) carbon atom to form dicarboxylic acids. This process occurs at the endoplasmic reticulum and requires molecular oxygen. When β -oxidation of fatty acids is impaired, more fatty acids are converted to dicarboxylic acids by ω -oxidation.

See also: **Bioenergetics:** Anaplerosis; **Cell Architecture and Function:** Peroxisomes; **Lipids Carbohydrates Membranes and Membrane Proteins:** Lipases; **Metabolism Vitamins and Hormones:** Carnitine and β -Oxidation; Cholesterol Synthesis and Regulation; Fatty Acid Structure and Synthesis; Gluconeogenesis; Insulin- and Glucagon-Secreting Cells of the Pancreas; **Protein/Enzyme Structure Function and Degradation:** Lipid Modification of Proteins: Targeting to Membranes; Protein Palmitoylation.

Further Reading

- Eaton S (2002) Control of mitochondrial β -oxidation flux. *Progress in Lipid Research* 41: 197–239.
- Kunau W-H, Dommes V, and Schulz H (1995) Beta oxidation of fatty acids in mitochondria, peroxisomes, and bacteria. *Progress in Lipid Research* 34: 267–341.
- McGarry JD (2001) Travels with carnitine palmitoyltransferase: I. From liver to germ cell with stops in between. *Biochemical Society Transactions* 29: 241–245.
- Ruderman NB, Saha AK, Vavas D, and Witters LA (1999) Malonyl-CoA, fuel sensing, and insulin resistance. *American Journal of Physiology* 276: E1–E18.
- Schulz H and Kunau W-H (1987) Beta-oxidation of unsaturated fatty acids: A revised pathway. *Trends in Biochemical Sciences* 12: 403–406.
- Van Veldhoven PP, Casteels M, Mannaerts GP, and Baes M (2001) Further insight into peroxisomal lipid breakdown via α - and β -oxidation. *Biochemical Society Transactions* 29: 292–298.
- Wanders RJA, Vreken P, Ferdinandusse S, et al. (2001) Peroxisomal fatty acid α - and β -oxidation in humans: Enzymology peroxisomal metabolite transporters, and peroxisomal diseases. *Biochemical Society Transactions* 29: 250–267.
- Watkins PA (1997) Fatty acid activation. *Progress in Lipid Research* 36: 55–83.

Fatty Acid Structure and Synthesis

S D Clarke, City University of New York, Brooklyn, NY, USA

M T Nakamura, University of Illinois at Urbana–Champaign, Urbana, IL, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

CHORE Carbohydrate response element, a DNA sequence that binds a unique carbohydrate response protein whose DNA-binding and transactivation activities are upregulated by glucose ingestion.

HUFA Highly unsaturated omega-6 and -3 fatty acids that uniquely regulate gene expression, cell differentiation, and are required for cognition and neural function.

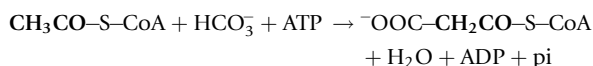
LXR (liver-X receptor) A nuclear transcription factor that is activated by the binding of oxysterols and stimulated by the transcription of lipogenic genes.

SREBP-1 Sterol regulatory element binding protein-1, required for the transcription of genes encoding proteins of fatty acid biosynthesis.

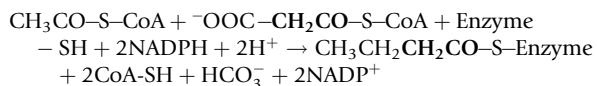
De Novo Fatty Acid Biosynthesis of Saturated Fatty Acids

Saturated fatty acids are categorized as short (C2–C6), medium (C8–C12), long (C14–C18), and very long (C20–C24) fatty acids. Short-chain fatty acids are derived largely from bacterial fermentation such as that which occurs in the gut or rumen. Medium-chain fatty acids are characteristic of milk fat and are absorbed from the intestine directly into the portal blood, and subsequently metabolized largely by the liver. Fatty acids containing 14 or more carbons are absorbed from the intestine and transported to the periphery as chylomicrons. Very long chain fatty acids are largely found in neural tissue and used for myelin formation. The liver and adipose tissue of humans are the primary sites of conversion of excess carbohydrate to fatty acids, but fatty acid synthesis is more than just for conserving excess glucose as fat. Fatty acid synthesis is essential for cell membrane production, lung surfactant function, milk fat production, brain myelination, and fat storage. Consequently, nearly all tissues have the capacity to synthesize fatty acids.

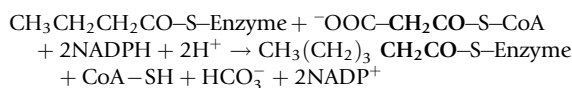
De novo fatty acid biosynthesis in humans occurs in cytosol. Two enzymes, acetyl-coenzyme A (CoA) carboxylase and fatty acid synthase, are required for this process. The first and rate-limiting step is the conversion of acetyl-CoA to malonyl-CoA by acetyl-CoA carboxylase (bold letters show the carbon unit in malonyl-CoA that is derived from acetyl-CoA):



Fatty acid synthase catalyzes all of the subsequent steps for the synthesis of long-chain saturated fatty acids. Fatty acid synthase is made of a single peptide chain that possesses seven distinct catalytic domains. Fatty acid synthase first adds two carbons to acetyl-CoA using malonyl-CoA. Nicotinamide adenine dinucleotide phosphate (NADPH) is used as a reducing power in this reaction:



Fatty acid synthase repeats the acyl chain elongation, two carbons at a time, using malonyl-CoA as a substrate (bold letters show the carbon unit that is newly added from malonyl-CoA to the elongating carbon chain after one cycle of reaction):



When the carbon chain reaches 16, the thioesterase domain of fatty acid synthase releases free palmitate, completing fatty acid biosynthesis:



Regulation of Pyruvate Dehydrogenase in Mitochondria

Malonyl-CoA is the substrate used by fatty acid synthase to assemble the acetate units into fatty acids. The first step in malonyl-CoA production involves the mitochondrial formation of acetyl-CoA by pyruvate dehydrogenase's decarboxylation of pyruvate. This is followed by the intra-mitochondrial synthesis of citrate from acetyl-CoA and oxaloacetate. Citrate is transported from the mitochondria to the cytosol where it is cleaved by citrate lyase to cytosolic acetyl-CoA and oxaloacetate. The acetyl-CoA is available for carboxylation to malonyl-CoA by the biotin-containing enzyme acetyl-CoA carboxylase. The liver and adipose tissue of humans are the primary sites of conversion of excess carbohydrate to fatty acids. The source of hepatic or adipose pyruvate for fatty acid biosynthesis may be derived from the metabolism of glucose via the glycolytic pathway with the liver or adipocyte itself, or it may be provided to the liver and/or adipose tissues as lactate or alanine that is produced by glucose metabolism in peripheral tissues, such as the small intestine and muscle. Entry of pyruvate into the tricarboxylic acid (TCA) pathway of the mitochondria is determined by the activity of pyruvate dehydrogenase. The concentration of pyruvate dehydrogenase does not widely fluctuate,

but its catalytic activity is suppressed by phosphorylation and enhanced by dephosphorylation of the enzyme. Insulin and glucose stimulate dephosphorylation while glucagon and fatty acids enhance the phosphorylation. Thus, pyruvate entry into the TCA cycle is highly dependent upon the macronutrient composition of the diet, and the ratio of insulin to glucagon.

Malonyl-CoA and Fatty Acid Metabolism

Malonyl-CoA plays a pivotal role in energy metabolism. As the substrate for fatty acid synthase, the cytosolic concentration of malonyl-CoA determines maximum rates of *de novo* fatty acid synthesis. As an inhibitor of carnitine palmitoyltransferase, malonyl-CoA controls the rate of fatty acid entry into the mitochondria, and hence is a key determinant of the rate of fatty acid oxidation. Thus, conditions that lead to high levels of malonyl-CoA (e.g., high carbohydrate intakes) suppress fatty acid entry into the mitochondria and increase their flux to triglycerides. On the other hand, conditions that result in low cellular concentrations of malonyl-CoA favor fatty acid oxidation because the inhibition of carnitine palmitoyltransferase is released. The synthesis of malonyl-CoA is catalyzed by the biotin-containing enzyme, acetyl-CoA carboxylase. Acetyl-CoA carboxylase exists as two isoforms: a liver or type I and a muscle or type II. The two isoforms are distributed in the cell in a manner that results in an intracellular compartmentalization of malonyl-CoA (i.e., malonyl-CoA produced by acetyl-CoA carboxylase type I is the substrate for fatty acid synthase, while malonyl-CoA generated by the type II enzyme governs fatty acid entry into the mitochondria via regulation of carnitine palmitoyltransferase). Like pyruvate dehydrogenase, the catalytic activity, and hence the production of malonyl-CoA, by both acetyl-CoA carboxylase isoforms is regulated by phosphorylation and dephosphorylation of the protein. Acetyl-CoA carboxylase I and II are substrates for AMP-activated protein kinase (AMPK) and cyclic AMP-dependent protein kinase (PKA). AMPK activity is enhanced by leptin and adiponectin, two hormones that stimulate fatty acid oxidation and inhibit fatty acid biosynthesis. The dephosphorylation of acetyl-CoA carboxylase is carried out by protein phosphatase 1 and 2, and phosphatase activity appears to be stimulated under conditions where glucose flux through glycolysis is high. In addition to its rate of production by acetyl-CoA carboxylase, the concentration of malonyl-CoA is also dependent upon its rate of decarboxylation by malonyl-CoA decarboxylase. Malonyl-CoA decarboxylase is also a substrate for AMPK, and its phosphorylation leads to an increase in decarboxylation activity and a loss of malonyl-CoA. Thus, AMPK activation governs fatty acid partitioning between triglyceride synthesis and oxidation by coordinately inhibiting malonyl-CoA synthesis and stimulating its degradation. In addition to having its activity acutely regulated via phosphorylation and dephosphorylation, the amount of hepatic acetyl-CoA carboxylase, like other lipogenic enzymes, increases and decreases in concentration depending upon nutritional conditions, dietary composition, and hormonal milieu. For example, the hepatic abundance of acetyl-CoA carboxylase is low during fasting and diabetes, but is greatly induced by refeeding glucose or administering insulin. On the other hand, muscle type II acetyl-CoA carboxylase is not an adaptive enzyme. Thus, its production of malonyl-CoA

is solely dependent upon the phosphorylation state of acetyl-CoA carboxylase type II. In addition to playing a key role in the partitioning of fatty acids between storage and oxidation, malonyl-CoA may also be a central metabolite in appetite regulation. Thus, even though net fatty acid synthesis by humans is relatively small, the fatty acid biosynthetic pathway plays an instrumental role in human lipid metabolism. Since excessive cellular triglyceride production and accumulation is causatively linked to the development of insulin resistance and type II diabetes, factors regulating the fatty acid biosynthetic pathway will also exert a direct influence on the development of diabetes.

Nutritional Regulation of Lipogenic Gene Expression

Hepatic and adipose fatty acid biosynthesis is increased by the ingestion of a high carbohydrate diet and decreased by the consumption of a high-fat diet. Dietary regulation of lipogenesis is not simply a product of nutritional manipulation of hormonal balance. Rather, nutrients exert a specific influence on the expression of glycolytic and lipogenic proteins. Nutrients exercise their influence by modulating gene transcription, messenger RNA (mRNA) processing, mRNA decay, and by stimulating posttranslational protein modifications.

Carbohydrate Induction of Lipogenic Gene Transcription

Work over the past decade has identified a collection of transcription factors that are specifically targeted by glucose and fatty acids. Dietary carbohydrate induces lipogenic gene transcription in two ways. First, glucose increases the nuclear concentration of the potent lipogenic transcription factor SREBP-1. SREBPs are a family of transcription factors that were first isolated as a result of their properties for binding to the sterol regulatory element of genes encoding proteins of sterol metabolism. SREBP-2 is a regulator of genes encoding proteins of cholesterol metabolism. SREBP-1 exists in two forms, 1a and 1c. SREBP-1a is the dominant form in cell lines and is a regulator of genes encoding proteins involved in fatty acid biosynthesis as well as cholesterolgenesis. SREBP-1c constitutes 90% of the SREBP-1 found *in vivo* and is a determinant of lipogenic gene transcription. SREBP-1 is synthesized as a 125-kDa precursor protein that is anchored in the endoplasmic reticulum membrane. Proteolytic release of the 68-kDa mature SREBP-1 occurs in the Golgi system, and movement of SREBP-1 from the endoplasmic reticulum to the Golgi requires the trafficking protein SREBP cleavage-activating protein. Once released, mature SREBP-1 translocates to the nucleus and binds to the classic sterol response element and/or to a palindrome CATG sequence.

Glucose increases the abundance of precursor SREBP-1 by inducing SREBP-1 gene transcription. In addition, glucose elevates the nuclear concentration of SREBP-1 by stimulating the proteolytic release of mature SREBP-1 from its membrane-anchored precursor. These glucose effects are largely mediated by insulin although the signaling pathway involved is yet to be fully elucidated. Activation of phosphatidylinositol 3-kinase and protein kinase C λ (PKC λ) by insulin is required for both transcriptional induction and proteolytic activation of SREBP-1c.

A regulatory protein Insig2a is expressed in liver, and retains precursor SREBP-1 in endoplasmic reticulum, thus preventing SREBP-1 from proteolytic activation. Insulin stimulates proteolytic activation of SREBP-1 by degrading Insig2a mRNA.

A second effect of glucose is to stimulate the DNA-binding activity of select transcription factors involved with the transcription of lipogenic genes, notably a unique carbohydrate response factor (CHORE). Glucose exerts its effect by elevating the cellular concentration of the pentose shunt metabolite xylulose-5-phosphate. Xylulose-5-phosphate is a positive metabolite effector of phosphatase 2A, and increased phosphatase activity leads to the dephosphorylation of CHORE, increased translocation of CHORE from the cytosol to the nucleus, and a subsequent increase in its DNA binding to a specific carbohydrate response element (CHORE) in select glycolytic (e.g., pyruvate kinase) and lipogenic (e.g., fatty acid synthase) genes. The consequence is an increase in gene transcription followed by a rise in the abundance of glycolytic and lipogenic enzymes.

Suppression of Lipogenic Gene Expression by Omega-6 and -3 Fatty Acids

Glucose stimulation of lipogenesis is a nice example of a feed-forward regulatory mechanism that functions to enhance glucose storage as fatty acids. Interestingly, the saturated fatty acid products of the *de novo* fatty acid synthetic pathway do not feed back and suppress glycolytic and lipogenic gene expression or activity. Only omega-6 and -3 fatty acids (e.g., linoleic (C18:2n-6) and α -linolenic (C18:3n-3)) interfere with glycolytic and lipogenic gene expression (Figure 1). Highly unsaturated 20- and 22-carbon omega-6 and -3 fatty acids (HUFAs) inhibit glycolytic and lipogenic gene transcription through two different types of DNA regulatory response regions (HUFA-RR). One type

overlaps with the insulin response element (HUFA-RR/IRE), while the second region co-localizes with the carbohydrate response element (HUFA-RR/CHORE).

Co-localization of the HUFA-response region with the insulin response element

The HUFA-RR/IRE contains recognition sequences for SREBP-1c, NF-Y, Sp1, and USF. HUFAs inhibit the activity of promoters containing the HUFA-RR/IRE by decreasing SREBP-1c and NF-Y interactions with their respective recognition sequences. HUFAs lower the concentration of SREBP-1 precursor protein as well as inhibit the proteolytic release and nuclear localization of mature SREBP-1. In addition to inhibiting the proteolytic release of SREBP-1, HUFAs reduce the cellular content of precursor SREBP-1 protein by reducing the cellular abundance of SREBP-1 mRNA. HUFAs decrease the hepatic abundance of SREBP-1 mRNA, and consequently the hepatic content of precursor SREBP-1 protein, by accelerating SREBP-1 mRNA decay and by suppressing SREBP-1 gene transcription. The mechanisms by which HUFAs accelerate SREBP-1 mRNA decay remain unclear, but the mechanisms by which HUFAs suppress SREBP-1 gene transcription are beginning to emerge. The SREBP-1 gene contains two response elements for the lipogenic transcription factor liver-X receptor (LXR). Oxysterol ligand activation of LXR induces the transcription of SREBP-1. HUFAs reportedly displace oxysterol from LXR and, in this way, antagonize the transactivation activity of LXR. In addition, HUFA may stimulate the trapping of LXR as a PPAR α /LXR heterodimer. Thus, when the ratio of HUFA to oxysterol is low LXR/RXR heterodimers favor lipogenic gene expression, but when the ratio of HUFA to oxysterol is high the PPAR α /LXR and PPAR α /RXR heterodimer mixtures are poised toward a gene profile that favors fatty acid oxidation over fatty acid synthesis. Although SREBP-1 metabolism is highly responsive to HUFA,

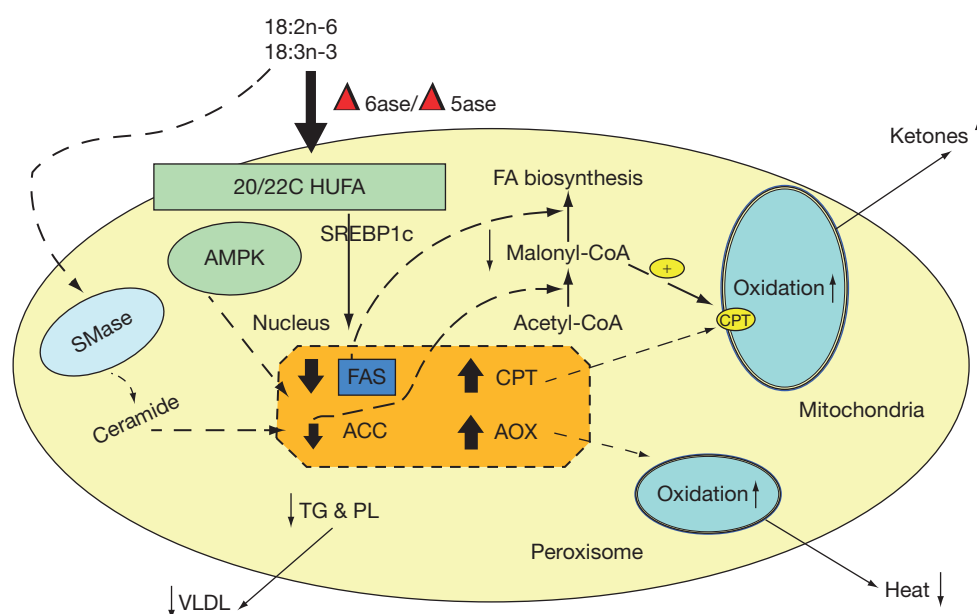


Figure 1 Regulation of hepatic gene expression and lipid metabolism by HUFAs. ACC, acetyl-CoA carboxylase; AMPK, AMP-activated protein kinase; AOX, acyl-CoA oxidase; CPT, carnitine palmitoyltransferase; FA, fatty acid; FAS, fatty acid synthase; HUFAs, highly unsaturated fatty acids; PL, phospholipid; SMase, sphingomyelinase; TG, triglyceride; VLDL, very low density lipoproteins; $\Delta 5$ ase, Δ -5 desaturase; $\Delta 6$ ase, Δ -6 desaturase.

it is only one of several transcription factors targeted by HUFA. In fact, >50% of the HUFA control of fatty acid synthase gene transcription can be attributed to the NF-Y site at -99/-93. NF-Y is an abundant heterotrimeric nuclear protein that interacts with histone acetyltransferases, and consequently plays a role in modifying histone structure. HUFAs do not reduce the nuclear abundance of NF-Y but rather suppress NF-Y transactivation activity by decreasing its DNA-binding activity.

A CHO-response element within an HUFA-response region

Part of the HUFA-RR of certain lipogenic genes, notably fatty acid synthase, contains DNA-binding sites for several transcription factors (e.g., CHORF and HNF-4). The DNA-binding activities of CHORF and HNF-4 are dependent upon their state of phosphorylation. Phosphorylation of CHORF inhibits its binding to the CHORE and prevents HNF4 from forming the homodimers required for its DNA binding. AMPK phosphorylation of HNF4 also accelerates the decay of HNF4 protein. HUFA suppression of lipogenic gene transcription involves an inhibition of the translocation of CHORF from the cytosol to the nucleus. However, whether or not HUFAs exert their effects via AMPK activation remains equivocal; but HUFAs' ability to coordinately suppress lipogenesis and stimulate fatty acid oxidation is consistent with the established role that AMPK plays in the partitioning of fatty acids between triglyceride storage and fatty acid oxidation.

Unsaturated Fatty Acid Synthesis

Oleic and Palmitoleic Acid Synthesis

Palmitate (C16:0) and stearate (C18:0) are rapidly desaturated to mono-unsaturated palmitoleic (C16:1) and oleic (C18:1) acids by the stearoyl-CoA desaturase (also known as Δ -9 desaturase) system (Figure 2). Stearoyl-CoA desaturase exists in multiple isoforms that are derived from separate genes. The dominant form within a cell depends upon the origin of the cell. Stearoyl-CoA desaturase inserts a double bond between carbons 9 and 10. This desaturation reaction requires three different components: the desaturase itself, cytochrome *b5*, and cytochrome *b5* reductase. Oleic acid synthesis is required for a variety of reasons. In particular, it appears to be required for the hepatic secretion of very low density lipoproteins.

Arachidonic and Docosahexaenoic Acid Synthesis

While oleic acid clearly plays an important role in overall energy metabolism, it cannot fulfill the biological requirements for growth, reproduction, cell signaling, and cell differentiation. These requirements can only be fulfilled by long-chain fatty acids from the omega-6 and -3 fatty acid family, that is, linoleic (C18:2n-6) and α -linolenic (C18:3n-3), respectively (Figure 2). Linoleate and α -linolenate undergo desaturation and elongation to 20- and 22-carbon highly unsaturated fatty acids. Prostaglandins and leukotrienes produced from arachidonic acid (C20:4n-6) are required for blood clotting, cardiac function, reproduction, inflammatory response, and cell differentiation. Docosahexaenoic acid (C22:6n-3) is required for

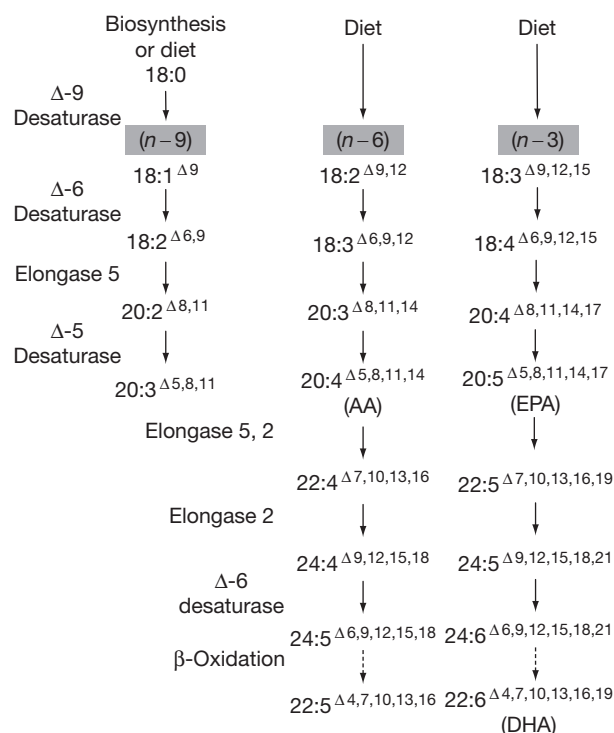


Figure 2 Synthesis of highly unsaturated fatty acids.

cognition, brain development, and sperm biogenesis. Formation of bioactive 20- and 22-carbon highly unsaturated omega-6 and -3 fatty acids requires that the dietary essential fatty acids, linoleate and α -linolenate, undergo two steps of desaturation that are catalyzed by the Δ -6 desaturase (FADS1) and Δ -5 desaturase (FADS2). Both desaturases are expressed in all tissues. The human desaturase genes are members of a gene cluster on chromosome 11 that includes a third desaturase of unknown function (FADS3), and a noncoding RNA gene that encodes an anti-sense transcript for Δ -5 desaturase. Nearly all human tissues express the products of the Δ -6 desaturase gene cluster. Unlike the stearoyl-CoA desaturase, the Δ -6 and Δ -5 desaturases contain a cytochrome *b5* domain within the desaturase peptide. The rate of production of 20- and 22-carbon highly unsaturated fatty acids is dependent upon the tissue content of the two desaturases, and this in turn is determined by nutritional and hormonal factors that are similar to those that regulate the expression of lipogenic genes. The one notable difference between lipogenic and desaturase genes is that while HUFAs suppress both gene families, activators of peroxisome proliferator activator receptor- α induce desaturase genes but they suppress lipogenic genes. The increase in desaturase expression may be in response to a greater need for unsaturated fatty acids for the production of phospholipids, particularly in the liver for the secretion of very low density lipoproteins.

See also: **Cell Architecture and Function:** Unfolded Protein Responses; **Metabolism Vitamins and Hormones:** Carbohydrate-Responsive Element Binding Protein; Coenzyme A.

Further Reading

- Berger J and Moller DE (2002) The mechanisms of action of PPARs. *Annual Review of Medicine* 53: 409–435.
- Clarke SD, Gasperikova D, Nelson C, et al. (2002) Fatty acid regulation of gene expression: A genomic explanation for the benefits of the Mediterranean diet. *New York Academy of Science* 967: 283–298.
- Horton JD, Goldstein JL, and Brown MS (2002) SREBPs: Activators of the complete program of cholesterol and fatty acid synthesis in the liver. *Journal of Clinical Investigation* 109: 1125–1131.
- Loewen CJR and Levine TP (2002) Cholesterol homeostasis: Not until the SCAP lady INSIGs. *Current Biology* 12: R779–R781.
- Puigserver P and Spiegelman BM (2003) Peroxisome proliferator-activated receptor-gamma coactivator 1 alpha (PGC-1 alpha): Transcriptional coactivator and metabolic regulator. *Endocrine Reviews* 24: 78–90.
- Vessby B, Gustafsson IB, Tengblad S, Boberg M, and Andersson A (2002) Desaturation and elongation of fatty acids and insulin action. *New York Academy of Science* 967: 183–195.
- Viollet B, Foretz M, Guigas B, et al. (2006) Activation of AMP-activated protein kinase in the liver: A new strategy for the management of metabolic hepatic disorders. *Journal of Physiology* 574: 41–53.

Fatty Acyl-CoA Synthetases

P A Watkins, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Activated fatty acid The fatty acyl-CoA product of the acyl-CoA synthetase reaction. Activation allows the fatty acid to participate in most subsequent metabolic processes.

Acyl-CoA synthetase(s) Enzyme(s) found in all organisms that catalyzes the adenosine triphosphate (ATP)-dependent formation of a thioester bond between the carboxyl carbon atom of a fatty acid and the sulfhydryl sulfur atom of coenzyme A. Different enzymes show a general preference

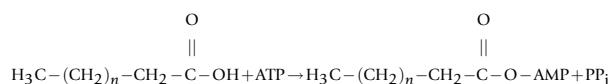
for fatty acids of differing chain length, including short-chain, medium-chain, long-chain, and very long chain fatty acids.

Bubblegum acyl-CoA synthetase An acyl-CoA synthetase encoded by the defective gene in the *Drosophila melanogaster* mutant *Bubblegum*, and structurally related acyl-CoA synthetases found in other organisms.

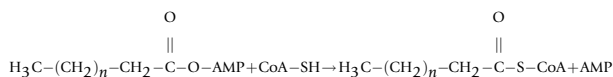
Motif An amino acid sequence within a protein that is evolutionarily conserved. All known acyl-CoA synthetases possess two such motifs.

Fatty acids (FAs) are extremely diverse biomolecules. They serve, along with carbohydrates and proteins, as important metabolic fuels. They are constituents of numerous types of complex lipid, including triacylglycerol, phospholipids, plasmalogens, sphingolipids, glycolipids, and cholesterol esters. FAs can be converted to aldehydes and alcohols, and can be covalently attached to proteins. FAs can be remodeled by elongation, shortening, and the insertion or removal of double bonds. For essentially all of the reactions in which an FA participates, it must first be activated by conversion to its coenzyme A (CoA-SH) thioester (Figure 1). This activation reaction is catalyzed by a family of enzymes known as fatty acyl-CoA synthetases, or simply acyl-CoA synthetases (ACSs; EC 6.2.1.x).

The ACS reaction is a two-step reaction that proceeds by a unidirectional, bi uni bi ping-pong mechanism. In the first half-reaction, FA and adenosine triphosphate (ATP) react to form an acyl-adenylate, releasing inorganic pyrophosphate:



Hydrolysis of PP_i by inorganic pyrophosphatases effectively prevents reversal of this reaction. In the second half-reaction, CoA-SH displaces adenosine monophosphate (AMP), forming the acyl-CoA product:



The thioester linkage between the acyl group and CoA-SH is a high-energy bond that energetically facilitates subsequent reactions. Because AMP requires two phosphorylation steps to regenerate ATP, the ACS reaction consumes the equivalent of two ATP molecules.

A large number of different FAs, and FA-like compounds, that are ACS substrates exist in nature. The carbon chain length of FA in animals varies from 2 carbons (acetate) to about 28 carbons; some plant lipids contain FAs with more than 30 carbons. The FA carbon chain can be saturated, or contain one

or more double bonds. FAs can contain one or more methyl groups, forming branched-chain FAs. FA-like molecules that are ACS substrates include the bile acids and many xenobiotic compounds (e.g., nonsteroidal anti-inflammatory drugs). Because these FA substrates are molecularly diverse, multiple ACS isoforms have evolved to accommodate them.

ACS Families and Nomenclature

ACSs have been found in all organisms examined for their presence, including archaea, bacteria, yeasts, fungi, plants, invertebrates, fish, birds, and mammals. Early on, it was evident that there were distinct ACSs capable of activating FAs based roughly on their carbon chain length. They were described primarily as short-, medium-, and long-chain ACSs. It was subsequently recognized that ACS activities differed somewhat among the various tissues (e.g., liver vs. brain). Within a single tissue, distinct ACS activities were associated with different subcellular compartments; while the endoplasmic reticulum seemed to prefer long-chain FA substrates, mitochondria avidly utilized medium-chain FAs. Differing preferences for saturated versus mono- or polyunsaturated FAs of the same carbon chain-length were also reported. Thus by the beginning of the genomics era, evidence was emerging for the existence of a large number of distinct ACSs within a given organism.

Currently, the widespread availability of protein sequence databases has greatly facilitated the identification of ACSs. Many ATP-dependent enzymes whose reaction mechanism involves formation of an adenylated intermediate and liberation of PP_i have a characteristic amino acid sequence called the 'AMP-binding domain' (motif I; Figure 2). The presence of this sequence along with another more recently identified conserved amino acid sequence (motif II; Figure 2) predicts that a protein will have ACS activity with a high degree of accuracy. For example, firefly (*Photinus pyralis*) luciferase, well known for catalyzing an ATP-dependent bioluminescence reaction, contains motifs I and II. Luciferase has now been shown to be a bifunctional enzyme that also catalyzes the ATP-dependent synthesis

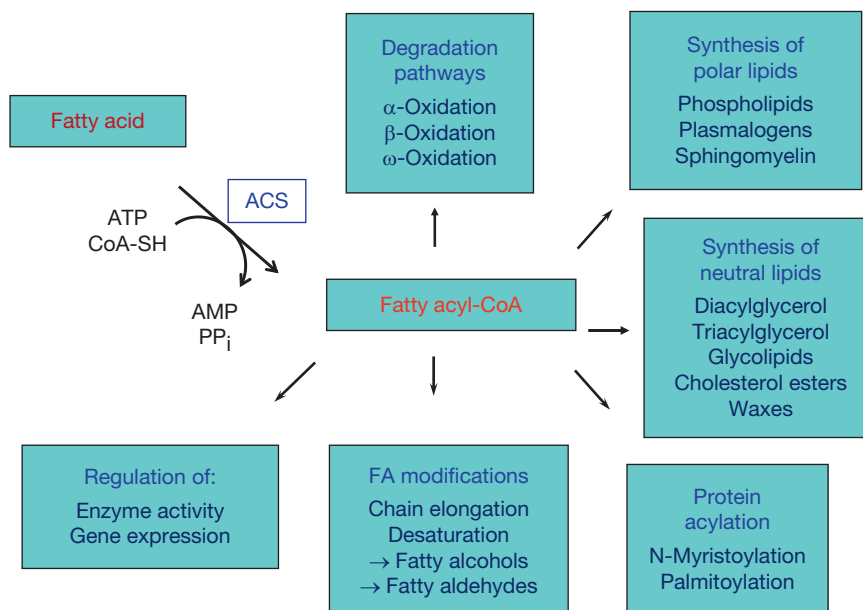


Figure 1 The fatty acyl-CoA synthetase (ACS) reaction. The ACS reaction catalyzes the ATP- and coenzyme A-dependent formation of fatty acyl-CoA. This activation process is necessary for the participation of FAs in the downstream metabolic processes illustrated. FA-CoAs are also regulators of enzyme activity and gene expression. One of the few reactions of FAs that does not require formation of the CoA-SH derivative is the conversion of the highly polyunsaturated FAs arachidonate and docosahexaenoate to bioactive eicosanoids and docosanoids, respectively.

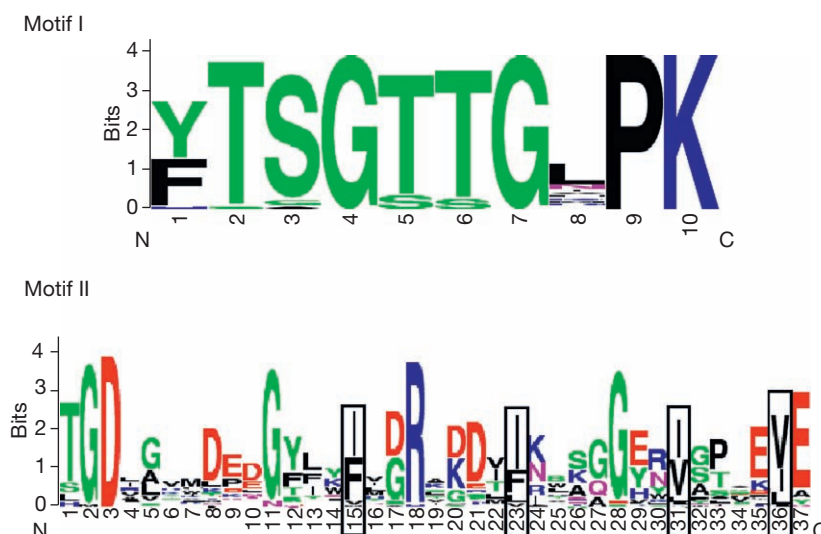


Figure 2 Conserved amino acid sequence motifs in ACSs. All known ACSs from all species contain two highly conserved domains termed motif I and motif II. Motif I is an AMP-binding domain containing 10 amino acids; motif II contains 36–37 amino acids. To illustrate the evolutionary conservation of these domains, sequence logos were created from 36 ACS sequences: 26 were from humans and 10 were from the archaeobacterium *Archaeoglobus fulgidus*. In this representation, the height of each letter is proportional to the amino acid sequence conservation at that position. Thus, in motif I, both glycine (G) residues, as well as the proline (P) and lysine (K) are invariant in the 36 ACS sequences. Motif II is longer and shows more variation. The aspartate (D) at position 3 is invariant. The arginine (R) at position 18 is conserved in all but one of the 36 sequences, where it is a histidine (H) in human ACSBG2; this residue is otherwise 100% conserved in nature. Only hydrophobic amino acid residues are found at four positions in motif II, which are enclosed in boxes. For the 36 ACSs compared, consensus sequences are: motif I, (Y/F)TSGTTGPK, and motif II, TGD(X)₇G(X)₃hX(D/G)R(X)₄h(X)₃₋₄G(X)₂h(X)₃EhE, where X is any amino acid and h indicates any hydrophobic residue. Sequence logos were created using the Weblogo program, found at <http://weblogo.berkeley.edu>.

of long-chain fatty acyl-CoA. Motifs I and II are found in all known ACSs from archaeobacteria to man.

The genomes of humans and mice each encode 26 proteins predicted to be ACSs (Table 1). Enzyme activity has been

verified experimentally for nearly all of these proteins. Alignment of these ACS sequences – either the full amino acid sequence or only motif II – revealed that there are distinct families of ACSs in an organism. Three structurally related

Table 1 Human ACS families and nomenclature

Human ACS family	FA chain length preference	Members
ACSS	Short (2–4 carbons)	3
ACSM	Medium (4–12 carbons)	6
ACSL	Long (8–20 carbons)	5
SLC27A (ACSVL)	Very long (16–26 carbons)	6
ACSBG	16–18 carbons	2

families identified with no *a priori* knowledge of function contain enzymes historically identified as short-, medium-, or long-chain-specific ACSs. Additional ACS families in humans and mice include the very long chain and the BG ACS families.

The current nomenclature of ACSs is, in general, based on the chain-length preference of families of structurally related proteins (Table 1). For most enzymes, the root designation ACS is followed by a family designation (e.g., L for long chain) and a number that distinguishes the different members of a family. Thus, the three-member human short-chain ACS family includes ACSS1, ACSS2, and ACSS3. ACSs belonging to the very long chain family do not conform to the uniform nomenclature. These enzymes are also known as FA transport proteins (FATPs; see below), and their official symbol (SLC27Ax) and name (solute carrier family 27 (FA transporter), member x) is based on this. Not all ACSs in an organism are structural members of a specific family. Humans and mice each have four ACSs that fall into this category. They have been called ACS family member 1 (ACSF1) through ACSF4, although this is the official designation for only ACSF2 and ACSF3 (see below).

Conversely, not all organisms contain representatives of all ACS families. Roundworms (*Caenorhabditis elegans*) and yeast (*Saccharomyces cerevisiae*) do not have enzymes belonging to the ACSBG family. Fruit flies and roundworms have a large ACS family that is not found in humans, mice, zebrafish (*Danio rerio*), or yeast. Yeast (*S. cerevisiae* and *Schizosaccharomyces pombe*) have a unique enzyme that is also found in plants (*Arabidopsis thaliana*) and bacteria (*Mycobacterium tuberculosis*), but not in higher organisms.

Short-Chain ACSs

This and subsequent sections focus primarily on human and rodent ACSs; their genomes encode three ACSS family members. These enzymes generally activate FAs such as acetate, propionate, and butyrate that contain two to four carbon atoms. ACSS1 is a mitochondrial acetyl-CoA synthetase found in heart, skeletal muscle, kidney, lung, spleen, brain, and testis. ACSS2 is also expressed in many tissues and activates acetate, but is found in the cytosol. Both enzymes can activate the three-carbon FA, propionate, but to a lesser extent than acetate. ACSS3 has not been characterized, although it is predicted to be widely expressed in mammalian tissues. The likely function of these enzymes is to reactivate acetate (or propionate) produced primarily via the action of cellular acyl-CoA thioesterases.

Medium-Chain ACSs

Humans have six ACSM genes. Five are found in a cluster on chromosome 16, while the other is located on chromosome 12. Substrates for these enzymes generally contain 4–12 carbon atoms. All of the ACSM family members are found mainly in mitochondria. ACSM1, which is found in liver and kidney, activates FAs containing 4–14 carbons. However, the enzyme has maximal activity toward six- and eight-carbon saturated acids, and also activates several xenobiotic compounds such as benzoate and valproate. ACSM2 was also characterized as a xenobiotic kidney-specific ACS, although transcripts are also found in liver. In addition to xenobiotics, it activates FAs containing 3–10 carbons, and is maximally active with the six-carbon FA, hexanoate. ACSM2 actually refers to two closely related proteins named ACSM2A and ACSM2B. ACSM2A and 2B are encoded by distinct genes, one located on the (+) strand of chromosome 16 and the other on the (–) strand; however, their amino acid sequences are nearly identical, making it difficult to distinguish their individual biochemical properties.

ACSM3 was originally described the SA protein which is expressed at high levels in spontaneously hypertensive rats. Like ACSM1 and 2, ACSM3 is found mainly in liver and kidney. FA substrates ranging from three to eight carbons, including isobutyrate, are reportedly activated by this enzyme. The relevance of this enzyme to hypertension remains questionable, although a specific polymorphism has been related to blood pressure, serum cholesterol, and serum triglycerides in a population study. ACSM4 is an olfactory epithelium-specific medium-chain enzyme that activates FAs containing 6–12 carbons. The human ACSM4 gene resides on chromosome 12, apart from the cluster on chromosome 16 containing the other five ACSM genes. In contrast, the mouse ACSM4 gene is clustered with several other ACSM genes. The role of this enzyme in olfaction has not been established. Little is known about ACSM5. Transcripts are found mainly in liver and kidney, but also to a lesser extent in heart, muscle, mammary gland, testis, and other tissues.

Long-Chain ACSs

The most abundant FAs in nature contain 16–18 carbons; thus, many ACS studies have focused on the long-chain (ACSL) enzyme family members. These enzymes typically utilize FA substrates containing 8–20 carbon atoms. It should be recognized that there can be significant overlap in the substrate specificities of ACSL enzymes with both ACSM and ACSVL family enzymes, making assignment of ACS activity to a specific enzyme family difficult. Humans and mice express five different ACSL enzymes, ACSL1 and ACSL3–6. The designation ACSL2 is no longer used in order to reconcile the current nomenclature with historical references to specific long-chain ACSs in the literature.

ACSL1 has been widely studied. It is most abundant in liver, muscle, adipose, and heart. ACSL1 activates saturated FAs containing 10–18 carbons, as well as many long-chain mono- and polyunsaturated FAs. The protein is found mainly in endoplasmic reticulum and mitochondria-associated membranes (MAMs); plasma membrane and cytosol may contain some ACSL1 as well. Functionally, ACSL1 has been implicated

in hepatic triacylglycerol synthesis and mitochondrial FA β -oxidation; overexpression of this enzyme in heart also increased triacylglycerol synthesis. In adipose tissue, ACSL1 participates in FA reesterification following lipolysis, thereby affecting lipid efflux from fat cells.

ACSL3 is most highly expressed in brain, but is also found at lower levels in many other tissues. Its substrate specificity is somewhat similar to that of ACSL1, but the enzyme shows a high degree of specificity for the polyunsaturated FA, arachidonate. The function of this enzyme in brain is poorly understood. The enzyme is also found associated with lipid droplets in human hepatoma cells and mouse 3 T3-L1 adipocytes. In hepatoma cells, ACSL3 has been implicated in both very low density lipoprotein secretion and mitochondrial FA β -oxidation.

Like ACSL3, ACSL4 shows a high level of substrate specificity for arachidonic acid; this enzyme also activates other polyunsaturated FAs. The enzyme is highly expressed in heart, brain, placenta, muscle, kidney, spleen, adrenal gland, testis, and ovary. Also similar to ACSL3, ACSL4 was found associated with lipid droplets in a human liver cell line. Subcellular fractionation of rat liver showed that the enzyme is also located in peroxisomes and MAMs. The latter location suggests a role for ACSL4 in triacylglycerol synthesis. This enzyme has also been implicated in steroidogenesis in several tissues. Mutations in ACSL4 are associated with X-linked mental retardation in some patients.

The mitochondrial enzyme ACSL5 is expressed in several tissues, including liver, spleen, uterus, intestine, colon, and leukocytes. Its substrates include saturated 10- to 20-carbon FAs, as well as long-chain monounsaturated and polyunsaturated FAs. ACSL5 is variably associated with some human cancers; its expression was markedly increased in human glioblastomas and glioma cell lines, but was decreased in small intestine epithelial tumors.

In rodents, ACSL6 was formerly known as ACSL2. This enzyme is capable of activating long-chain saturated, monounsaturated, and polyunsaturated FAs. The main human tissues expressing ACSL6 are brain and erythrocyte precursor cells. A differential splicing event produces two variants of ACSL6 that differ in the amino acid sequence of one exon; one variant is brain specific, while the other is found in blood cell precursors. The brain variant of ACSL6 is upregulated in the frontal cortex in a rodent model of depression; this enzyme may participate in neuronal cell proliferation and differentiation. Mutations in the hematopoietic variant of ACSL6 have been associated with acute myelogenous leukemia in several patients.

Very Long Chain ACSs

The human and mouse very long chain ACS families each have six members. In general, these enzymes are capable of activating long- to very long chain FAs (containing 16–26 carbons); thus, there is significant substrate specificity overlap with ACSL family enzymes. The designations of these ACSs in the literature using the ACSVL nomenclature have been inconsistent; therefore, they are described herein using their official gene/protein names. SLC27A1 (also known as FATP1 and ACSVL4) was identified in a screen for proteins that facilitated uptake of long-chain FAs by cultured cells. Thus, it was initially called FATP. This protein was subsequently shown to have ACS

activity. It activates FAs containing 16–24 carbons, and is found in muscle, heart, brain, and adipose tissue. Subcellular locations include endoplasmic reticulum, mitochondria, plasma membrane, and cytoplasmic vesicles. Functionally, SLC27A1 has been implicated in triglyceride synthesis and mitochondrial β -oxidation. Expression is upregulated during adipocyte differentiation. Heart-specific overexpression of SLC27A1 in mice resulted in the increased uptake of long-chain FAs and lipotoxicity, suggesting a role in maintaining normal FA levels in cardiomyocytes.

SLC27A2 (also known as FATP2 and ACSVL1) is a liver- and kidney-specific ACS. This enzyme was the first very long chain specific ACS isolated from a tissue source – rat liver peroxisomes – using classical protein purification methodology. It activates FAs containing 16–24 carbons, and is also found in the endoplasmic reticulum. In peroxisomes, SLC27A2 activates FAs to provide substrates for the β -oxidation pathways found in that organelle. Bile acid precursors derived from cholesterol, which undergo one cycle of peroxisomal β -oxidation in their conversion to bile acids, are also SLC27A2 substrates. The function of this enzyme in the endoplasmic reticulum is unknown.

SLC27A3 (also known as FATP3 and ACSVL3) expression is most abundant in brain and steroidogenic tissues. Its substrate specificity is similar to that of SLC27A1. The enzyme is found in small cytoplasmic vesicles that have not yet been characterized. SLC27A3 expression is highly upregulated in human malignant gliomas and cultured glioma cells; knockdown of the enzyme by RNA interference reduced the malignant properties of tumors and cells.

Brain, adipose tissue, skeletal muscle, heart, and small intestine are the primary tissues expressing SLC27A4 (also known as FATP4 and ACSVL5). Substrates activated by SLC27A4 are similar to those activated by SLC27A1 and A3, but this enzyme seems to have a greater preference for the very long chain FAs. SLC27A4 and SLC27A1 share considerable amino acid identity, and both have been implicated in the cellular uptake of long-chain FAs. Unlike SLC27A1, however, which is partially located in plasma membrane, SLC27A4 is in peroxisomes, MAMs, and endoplasmic reticulum, but not in plasma membrane, reducing the likelihood of a role in FA uptake. In skin fibroblasts, this enzyme is required for incorporation of very long chain FAs (but not long-chain FAs) into triglycerides and phospholipids, and is needed for the peroxisomal β -oxidation of very long chain FAs. Polymorphisms in SLC27A3 have been associated with obesity and insulin resistance in some human population studies.

Bile acids, and not FAs, are the primary substrates for the liver-specific enzyme SLC27A5 (also known as FATP5, ACSVL5, and BACS). For conjugation to taurine or glycine, bile acids must first be activated to their CoA-SH derivatives. In liver hepatocytes, SLC27A5 is found in the basal plasma membrane. Although this enzyme only weakly activates FA substrates, mice deficient in SLC27A5 have reduced triglyceride and free FA levels in addition to reduced levels of conjugated bile acids. SLC27A5-deficient mice also have decreased food intake and increased energy expenditure.

SLC27A6 (also known as FATP6 and ACSVL2), characterized initially as a heart-specific protein, is also detected in hair follicle epithelia. The protein is found on the sarcolemma of cardiomyocytes near small blood vessels. Like most other

members of the family, long-chain to very long chain FAs are the primary substrates of SLC27A6. The function of this enzyme is not known.

The Bubblegum or BG ACS Family

The fruit fly (*D. melanogaster*) "Bubblegum" mutant, characterized by neurodegeneration and elevated levels of saturated very long chain FAs, has a defect in the gene encoding an ACS that does not belong to any of the families listed above. Humans, mice, and fruit flies each have two ACSBG family enzymes. Human ACSBG1 is found only in neurons in the brain and spinal cord, and in steroidogenic cells within the adrenal cortex, testis, and ovary. Like SLC27A3, ACSBG1 is contained within an uncharacterized population of small cytoplasmic vesicles. While most ACSs activate a range of FAs both shorter and longer than their optimal substrate, ACSBG1 has a unique preference for the 16-carbon saturated FA, palmitate. A role for this enzyme in the human disorder, X-linked adrenoleukodystrophy, has been hypothesized but not proven. There is some evidence for a role of ACSBG1 in the regulation of testicular steroidogenesis.

The tissue expression of ACSBG2 is limited to the testis, particularly Sertoli cells, and large motoneurons of the brainstem and cervical spinal cord. Preferred FA substrates are 18-carbon FAs containing one (oleate) or two (linoleate) double bonds; these are the most abundant FAs in testis. A role for ACSBG2 in spermatogenesis has been suggested.

Other Human and Mouse ACSs

Four proteins containing structural features of ACSs, including motifs I and II, but that do not belong to any family above, are found in mice and men. These are referred to as ACSF (ACS family) proteins. ACSF1 (official symbol, AACS) is acetoacetyl-CoA synthetase. It is a cytosolic protein found mainly in brain and kidney, with moderate levels in heart, skeletal muscle, and placenta. The proposed function of this enzyme is in cholesterol biosynthesis. Little is known about ACSF2 and ACSF3; the former prefers medium-chain FA substrates, whereas the latter activates the 24-carbon very long chain FA, lignocerate. ACSF4 (official symbol, AASDH) is amino adipate-semialdehyde dehydrogenase. Its reaction mechanism involves an adenylated amino adipate intermediate, and the enzyme itself is covalently modified by the addition of a phosphopantetheine group donated by CoA-SH. However, the ability of AASDH to function as an ACS and activate FA substrates has not been demonstrated.

ACS Structure–Function Considerations

Human ACSs, excluding the unusually long putative ACS, ACSF4/AASDH, contain 576–739 amino acids (average: 656). The relative positions of motifs I and II (Figure 2) within ACS amino acid sequences are highly conserved. The initial residue of motif I is found at approximately amino acid 261 (range: 201–334). The average distance between this residue and the

conserved arginine (R) of motif II is 263 amino acids (range: 241–293). Motifs I and II are found in all known ACSs and are highly conserved throughout evolution, suggesting that they are necessary for catalytic function. Mutagenesis of specific residues in motifs I and II of ACSs from yeast, bacteria, and plants confirmed that several amino acids in these domains are critical for enzyme activity. In addition to motifs I and II, there are at least three additional conserved domains that are present in subsets of ACSs. For example, a motif with consensus sequence LPLXH (where X is any amino acid) is found in all human ACSL, SLC27A, and ACSBG family members, but not in any of the ACSS or ACSM enzymes.

No crystal structure for a mammalian ACS was available until 2009, when the structure of the human medium-chain enzyme, ACSM2A, was published. Crystal structures of three ACSs from lower organisms have also been determined. Two are short-chain ACSs, one from the yeast *S. cerevisiae* and the other from the bacterium *Salmonella enterica*. The structure of a long-chain ACS from the extreme thermophile *Thermus thermophilus* has also been solved.

Human ACSM2A crystallizes as a monomer, whereas the yeast and bacterial enzymes crystallize as dimers or trimers. All monomeric units have a characteristic two-domain structure, consisting of a larger N-terminal portion and a smaller C-terminal domain linked by a hinge region. Catalysis occurs at the interface of the two domains. The C-terminal domain assumes different conformations during the catalytic cycle, rotating 135–140° between the first (acyl-adenylate forming) and second (thioester forming) half-reactions. Release of PP_i by the first half-reaction is thought to trigger this conformational switch.

Motifs I and II are located in the larger N-terminal domain of ACSs. Crystal structures indicate that residues comprising motif I are in close proximity to the adenosine moiety of AMP, maintaining it in the proper orientation. Similarly, the highly conserved arginine (R) residue at the center of motif II is in proximity to ribose hydroxyl groups in AMP, again maintaining the spatial orientation of the adenylate group.

The LPLXH motif of human ACSs described above is located between motifs I and II, and is thus also in the large N-terminal domain. This motif is homologous to a region described as the gate motif in the *T. thermophilus* long-chain ACS. FA binding to the enzyme promotes opening of the gate, allowing access to the ATP-binding site for acyl-adenylate formation. This motif is not present in human ACSS or ACSM enzymes, nor is it found in the yeast or bacterial acetyl-CoA synthetases noted above, suggesting that it is more critical to enzymes that prefer longer-chain FA substrates.

There is a highly conserved lysine residue found near the end of the small C-terminal domain of ACSs. This residue is critical for both catalysis and regulation of enzyme activity. In the first half-reaction, a salt bridge between the epsilon amino group of the conserved lysine and the β-phosphate oxygen atoms of ATP helps position the latter compound for nucleophilic attack by the FA substrate. Following rotation of the C-terminal domain, the lysine residue is located on the exterior surface of the protein, where it remains during the second half-reaction. In some short-chain ACSs, acetylation of the epsilon amino group occurs, effectively rendering the enzyme catalytically inactive. Deacetylation catalyzed by sirtuins restores activity. Although

all ACSs seem to retain the conserved lysine near the C-terminus, regulation of ACSs that activate longer chain FAs by this mechanism has not been demonstrated.

Concluding Remarks

The ACSs comprise a large family of enzymes that must accommodate the diverse nature of both the FA substrates they activate, and the downstream processes in which their products participate. The presence of multiple members of the ACSs, ACSM, ACSL, and SLC27A suggests the possibility that each serves at unique positions in metabolism, providing fatty acyl-CoAs for specific anabolic or catabolic pathways. This is accomplished via substrate specificity, tissue and cellular expression patterns, and subcellular distribution. The specific functions of ACSs in metabolism are only beginning to be understood. As new observations on ACSs are reported, it is clear that these enzymes perform many unique functions, and are critical for the maintenance of normal cellular FA and lipid homeostasis.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Glycerolipid Structure, Function, and Synthesis in Eukaryotes; Phospholipid Synthesis in Yeast; Sphingolipid

Biosynthesis; Metabolism Vitamins and Hormones: Bile Salts; Coenzyme A; Fatty Acid Metabolism and Cancer; Fatty Acid Oxidation; Fatty Acid Structure and Synthesis; **Protein/Enzyme Structure Function and Degradation:** Lipid Modification of Proteins: Targeting to Membranes.

Further Reading

- Black PN and DiRusso CC (2007) Yeast acyl-CoA synthetases at the crossroads of fatty acid metabolism and regulation. *Biochimica et Biophysica Acta* 1771: 286–298.
- Coleman RA, Lewin TM, Van Horn CG, and Gonzalez-Baró MR (2002) Do long-chain acyl-CoA synthetases regulate fatty acid entry into synthetic versus degradative pathways? *Journal of Nutrition* 132: 2123–2126.
- Gimeno RE (2007) Fatty acid transport proteins. *Current Opinion in Lipidology* 18(3): 271–276.
- Knights KM and Drogemuller CJ (2000) Xenobiotic-CoA ligases: Kinetic and molecular characterization. *Current Drug Metabolism* 1: 49–66.
- Kochan G, Pilka ES, von Delft F, Oppermann U, and Yue WW (2009) Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-coenzyme A synthetase ACSM2A. *Journal of Molecular Biology* 388: 997–1008.
- Soupe E and Kuypers FA (2008) Mammalian long-chain acyl-CoA synthetases. *Experimental Biology and Medicine (Maywood)* 233: 507–521.
- Watkins PA (1997) Fatty acid activation. *Progress in Lipid Research* 36: 55–83.
- Watkins PA (2008) Very-long-chain acyl-CoA synthetases. *Journal of Biological Chemistry* 283: 1773–1777.
- Watkins PA, Maiguel D, Jia Z, and Pevsner J (2007) Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome. *Journal of Lipid Research* 48: 2736–2750.

Ferredoxin

G Zanetti and V Pandini, Università degli Studi di Milano, Milan, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Iron-sulfur cluster A prosthetic group containing iron and sulfur atoms liganded by protein-cysteine residues.

Isoform Any of multiple forms of a protein differing in primary structure but having similar function.

Photosynthesis A process that converts the light energy into the chemical energy of nicotinamide adenine dinucleotide phosphate, reduced form (NADPH) and adenosine triphosphate (ATP).

Thioredoxin A small protein containing an oxidoreducible disulfide.

2Fe Ferredoxins

The 2Fe ferredoxins (Fds) are hydrophilic, acidic proteins of 11–15 kDa which harbor a single 2Fe–2S cluster coordinated by four cysteine residues. They function as single-electron shuttle in fundamental metabolic processes. The family can be subdivided into two major classes, the mitochondrial-type and the plant-type Fds, based on the cysteine spacing in the iron-sulfur binding motif, the three-dimensional (3D) structure, and function. A third class of bacterial 2Fe-Fds has recently been recognized on the basis of sequence and spectroscopic properties.

Plant-Type 2Fe Fds

More than a hundred sequences of 2Fe Fds have been deposited in sequence databases. The polypeptide chain is 93–98 residues long, yielding a protein of molecular mass of ~11 kDa. Plants as well as the single-celled green algae contain several Fd isoforms (up to six) at variance with cyanobacteria which host just one Fd in the vegetative cells and an isoform in the heterocysts. In photosynthetic cells, the main role of Fd is the electron transfer from photoreduced photosystem I (PSI) to Fd-NADP⁺ reductase (FNR), as well as a distribution of electrons to a large number of Fd-dependent enzymes involved in regulatory and assimilatory pathways (Figure 1). The archetype of this class is the chloroplast ferredoxin I (FdI) which is encoded by the PET F gene. Under normal growth conditions, the PET F transcript in the green alga *Chlamydomonas reinhardtii* constitutes 98% of the total Fd-encoding transcript.

Molecular Structure

The characteristic spacing of the cysteinyl residues in the iron-sulfur-cluster-binding motif for this class is C-X₄-C-X₂-C, which provides three of the four thiolate ligands to the two iron atoms, with the fourth ligand 29 residues downstream in the primary structure. Several Fds have been structurally characterized by either X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy. They show a folding motif, dubbed β -grasp, which is similar to that of ubiquitin. It is constituted by a

four-stranded antiparallel β -sheet and an α -helix lying perpendicularly to the sheet. A representation of the structure of recombinant spinach FdI is depicted in Figure 2 to show the secondary structure elements. The central core (β -grasp) holds the loop which harbors the Cys motif for binding the iron-sulfur cluster. In addition, two short β -strands, a second α -helix, and a short C-terminal helical turn are present. The binuclear iron cluster is located near the surface of the protein with Fe1 atom more exposed to solvent. Upon protein reduction, Fe1 has been shown to become reduced. The surface of the Fd molecule shows an asymmetry in distribution of the many acidic residues, which cluster in two distinct areas around the FeS prosthetic group. This feature common to all plant-type Fds has been shown to be important for interaction with protein partners.

Physiological Function

The nuclear-encoded FdI is imported in the chloroplast as an apoprotein, where it acquires the prosthetic group. FdI biosynthesis is under the control of light. As depicted in Figure 1, FdI forms transient electron transfer complexes with several different proteins. First, it is reduced through electron transfer from the 4Fe–4S center F_B of the subunit PsaC of PSI, by docking to both the PsaD and the PsaE subunits at the stromal side of the thylakoid membrane. Then, the reduced protein donates its electron to either one of the various enzymes known to be Fd dependent. It is conceivable that some of these functions may be performed by a specific isoform of Fd. NADPH production through Fd-NADP⁺ reductase (FNR) is the main role of FdI. Several studies employing a number of different techniques have attempted to define the binding region of FdI for its various redox partners. The 3D structure of the Fd-FNR complexes from both *Anabaena* spp. and maize has recently been obtained. Overall, the picture which comes out is that FdI acts as a mobile electron shuttle between PSI and the enzyme partners, utilizing for the interaction the same acidic surface area surrounding the iron-sulfur cluster. The binding sites need not be exactly the same but may be just overlapping.

Fd isoforms are also present in plastids of nongreen plant tissues and, recently, such a type of Fd has been found in the organelle called 'apicoplast' of protozoan parasites Apicomplexa (*Plasmodium* spp., *Toxoplasma gondii*, etc.). These Fds, which have a more positive redox potential than the photosynthetic

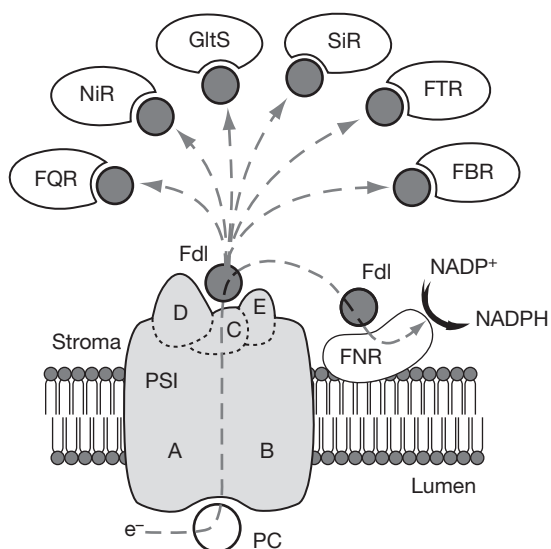


Figure 1 Schematic representation of the photosynthetic chain indicating the electron transfer pathways from photosystem I (PSI) to several metabolic processes of chloroplast by FdI servicing. FNR, Fd-NADP⁺ reductase (carbon assimilation); FBR, Fd-bilin reductase (phytochromobilin synthesis); FTR, Fd-thioredoxin reductase (regulation of Calvin cycle enzymes); SiR, sulfite reductase (sulfur assimilation); GlTS, glutamate synthase (amino acid synthesis); NiR, nitrite reductase (nitrogen assimilation); FQR, hypothetical Fd-quinone reductase (cyclic electron flow). PC, plastocyanin. A, B, C, D, and E indicate the PsaA, PsaB, PsaC, PsaD, and PsaE subunits of PSI, respectively.

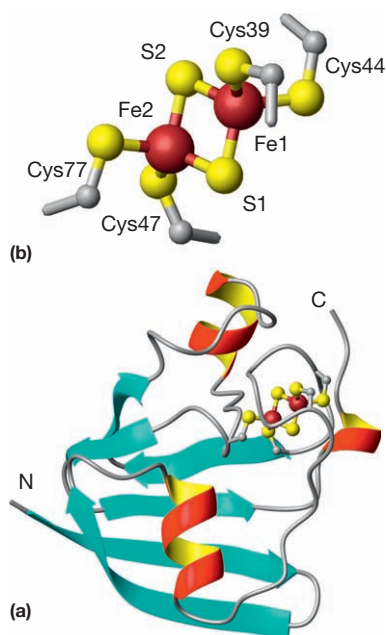


Figure 2 Ribbon representation of the three-dimensional structure of spinach FdI showing secondary structure elements. (a) N and C define the N terminus and C terminus, respectively. (b) Detail of the iron-sulfur cluster showing the side chains of the cysteines (residues 39, 44, 47, and 77) and the two sulfide sulfur atoms which coordinate the two iron atoms. The figure was prepared using Molmol.

ones, are reduced by FNR isoforms at the expense of NADPH and act as reductant of Fd-dependent pathways in these organelles.

Mitochondrial-Type 2Fe Fds

Mammalian adrenodoxin (Adx) is the prototype of this class; other well-known members are putidaredoxin of *Pseudomonas putida*, yeast Fd, and *Escherichia coli* Fd. The first two Fds have the role of providing electrons to cytochrome P450-catalyzed hydroxylation reactions, the latter two have been more recently implicated in iron-sulfur cluster biogenesis. They are 105–128 residues long, thus are up to 4 kDa larger than plant Fds. The typical Fe-S-binding sequence is C-X₅-C-X₂-C, which provides three of the four thiolate ligands to the iron atoms, with the fourth ligand 35–37 residues downstream in the sequence. The 3D structures of these Fds show high similarity to those of the plant Fds in the fold of the core involving the N-terminal moiety (i.e., the β -grasp), but they have an additional C-terminal lobe, called ‘interaction domain’, which has been shown to be involved in binding both the redox partners, Adx reductase and cytochrome P450. As in the case of FdI, the binding sites may be separate but overlapping.

Thioredoxin-Type 2Fe Fds

Thioredoxin-like Fds (also dubbed Shethna subtype) were first identified in *Azotobacter vinelandii* and *C. pasteurianum*. They seem to be present only in bacteria. The polypeptide chain length is that usually reported for 2Fe Fds, that is, 100–120 residues, but these Fds, being dimer, have a larger molecular mass. The cysteine motif for binding the iron-sulfur cluster is rather unique: C-X₁₀₋₁₂-C-X₂₉₋₃₄-C-X₃-Cys and the crystal structure of *Aquifex aeolicus* Fd revealed a fold unusual for Fds and quite similar to that of thioredoxin. The function of these Fds is not yet known, although there are some indications for their involvement in nitrogen fixation.

Fds Containing 4Fe and 3Fe Iron-Sulfur Clusters

This class comprises 3Fe, 4Fe, 7Fe, and 8Fe Fds, which are small, bacterial proteins containing one or two FeS clusters of the cubane type. The iron and sulfide atoms alternate on the corners of a distorted cube. In the case of the 3Fe cluster, one corner is devoid of the fourth iron atom. The two types of clusters are in some cases interconvertible. The cysteine motif for the coordination of these iron-sulfur centers is C-X₂-C-X₂-C with the middle cysteine replaced by aspartic acid or any residue in the case of 3Fe clusters, in a polypeptide chain of 60–100 residues. The 8Fe Fds constitute the largest family of this class of Fds. Their structure highlights a pseudo-twofold symmetry of the molecule which reflects an ancient sequence duplication. All 7Fe and 8Fe Fds are phylogenetically related. They all share a very similar [Fe-S] cluster-binding core of about 55 amino acids folded according to a conserved ($\beta\alpha\beta$)₂ structure topology. Nevertheless, outside the cluster-binding core, the Fds show a high level of structural variability. All these Fds generally function as electron carriers in several

metabolic pathways, including pyruvate metabolism, nitrogen fixation, sulfate reduction, hydrogen production, and cytochrome P450 hydroxylations.

See also: **Bioenergetics:** Ferredoxin-NADP⁺ Reductase; Iron–Sulfur Proteins; **Metabolism Vitamins and Hormones:** Photosynthesis.

Further Reading

- Beinert H, Holm RH, and Münck E (1997) Iron–sulfur clusters: Nature's modular, multipurpose structures. *Science* 277: 653–659.
- Grinberg AV, Hannemann F, Schiffler B, et al. (2000) Adrenodoxin: Structure, stability, and electron transfer properties. *Proteins* 40: 590–612.
- Knaff DB (1996) Ferredoxin and ferredoxin-dependent enzymes. In: Ortand DR and Yocum CF (eds.) *Oxygenic Photosynthesis: The Light Reactions*, vol. 4, pp. 333–361. Dordrecht: Kluwer.
- Meyer J (2008) Iron–sulfur protein folds, iron–sulfur chemistry, and evolution. *Journal of Biological Inorganic Chemistry* 13: 157–170.
- Tagawa K and Arnon DI (1962) Ferredoxins as electron carriers in photosynthesis and in the biological production and consumption of hydrogen gas. *Nature* 195: 537–543.
- Terauchi AM, Lu S-F, Zaffagnini M, et al. (2009) Patterns of expression and substrate specificity of chloroplast ferredoxins from *Chlamydomonas reinhardtii*. *Journal Biological Chemistry* 284: 25867–25878.
- Zanetti G, Binda C, and Aliverti A (2001) The [2Fe–2S] ferredoxins. In: Messerschmidt A, Huber R, Poulos T, and Wieghardt K (eds.) *Handbook of Metalloproteins*, vol. 1, pp. 532–542. Chichester: Wiley.

Ferredoxin-NADP⁺ Reductase

G Zanetti and A Aliverti, Università degli Studi di Milano, Milan, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Domain An independently folded structural region of a protein molecule.

Flavoprotein A protein that contains a flavin nucleotide (flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN)) as prosthetic group.

Photosynthesis A process that converts the light energy into chemical energy of NADPH and adenosine triphosphate (ATP).

Prosthetic group A tightly bound cofactor that contributes to the biological activity of a protein.

Rossmann fold A protein structural motif for nucleotide binding.

Introduction

Ferredoxin-NADP⁺ reductase (FNR) is the common name of members of a broad group of flavoenzymes (classified as EC 1.18.1.2) sharing the ability to catalyze the transfer of reducing equivalents between NADP(H) and ferredoxins (Fds). The best-known example is the photosynthetic FNR, the terminal component of the photosynthetic electron transport chain, in both eukaryotic and prokaryotic phototrophic organisms. In addition to this photosynthetic oxidoreductase, isoforms exist in nonphotosynthetic cells. The subdivision between photosynthetic and nonphotosynthetic FNRs is functional, and is based on the physiological direction of the catalyzed reaction, that is, toward NADP⁺ reduction in the case of photosynthetic FNR and toward Fd reduction for heterotrophic FNRs (Figure 1). Besides such functional classification, a structural/phylogenetic one is possible, which distinguishes between plant-type and glutathione reductase (GR)-type FNRs. While all photosynthetic FNRs are plant-type, nonphotosynthetic reductases are distributed between both structural classes.

Furthermore, the plant-type FNR is the prototype of a large superfamily of homologous enzymes, which display a great variety of catalytic functions. All members of this family possess the two-domain FNR unit and additional domains that modify and extend their functional properties.

Photosynthetic FNR

FNR was first discovered in 1956 by Avron and Jagendorf, who isolated it from pea chloroplasts as an NADPH-dependent diaphorase (i.e., a flavin-containing enzyme that catalyzes the transfer of reducing equivalents from NADPH to various electron acceptors). This chloroplast oxidoreductase was later shown by Shin and Arnon to mediate the transfer of electrons from ferredoxin to NADP⁺.

Molecular Structure

Photosynthetic FNRs are monomeric proteins of about 35-kDa molecular mass, containing one molecule of noncovalently but tightly bound flavin adenine dinucleotide (FAD) as prosthetic group. FNRs from different eukaryotic sources show a

high sequence similarity (~80% identity), which is spread over the entire polypeptide chain of ~300 amino acid residues. Cyanobacterial FNR possesses an additional N-terminal sequence of ~10 kDa involved in binding to phycobilisomes, light-harvesting structures of the thylakoid membrane of cyanobacteria. The three-dimensional (3D) structures of photosynthetic FNRs have been determined at high resolution (up to 1.7 Å) by X-ray crystallography for the proteins from spinach, pea, paprika, and maize leaves, as well as from the cyanobacterium *Anabaena* spp. The enzyme molecule consists of two structural domains, each of ~150 residues (Figure 2). The N-terminal domain, comprising a six-stranded antiparallel β-barrel, a small β-sheet, and an α-helix, provides most of the residues involved in FAD binding and is thus also known as the FAD-binding domain. The C-terminal domain, topologically similar to the supersecondary structure named 'Rossmann fold', binds NADP(H) and is thus indicated as the NADP-binding domain. However, both dinucleotides interact also with residues belonging to the other domain. Indeed, FAD- and NADP(H)-binding sites are located within a deep cavity at the interface of the two domains. The isoalloxazine ring of FAD, as in many flavoproteins, is sandwiched between the side chains of two aromatic residues. In particular, the C-terminal tyrosine residue makes an aromatic stacking interaction with the *re*-face of the isoalloxazine ring.

Subcellular Localization and Physiological Role

FNR is present in the chloroplasts of algae and higher plants. It is an extrinsic membrane protein localized on the side facing the stroma of the thylakoid membrane. FNR is the catalyst of the last step in the light-driven electron transport. FNR is thus directly responsible for producing the NADPH required in the Calvin cycle for carbohydrate biosynthesis. Reducing power is in turn provided to FNR by Fd, which acts as an electron shuttle by diffusing in the stroma between photosystem I and FNR and being reduced by the former and oxidized by the latter.

Catalytic Mechanism

FNR catalyzes the reversible transfer of reducing equivalents between the obligatory one-electron carrier ferredoxin (Fd)

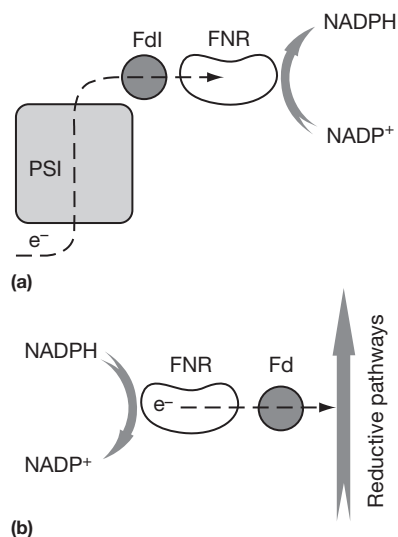


Figure 1 Comparison of the physiological reactions catalyzed by photosynthetic and nonphotosynthetic FNR isoforms. (a) Scheme of the reaction catalyzed by FNR in chloroplasts. PSI, photosystem I; FdL, photosynthetic isoform of Fd. (b) Scheme of the reaction catalyzed by nonphotosynthetic FNR isoforms. Fd indicates any nonphotosynthetic Fd isoform, including adrenodoxin. The dashed line indicates the electron transfer path.

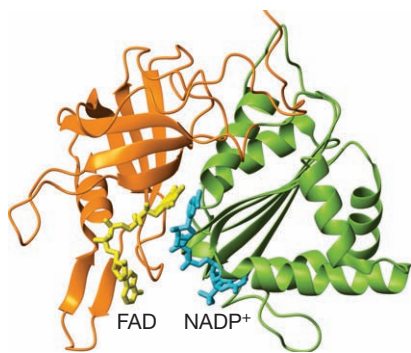
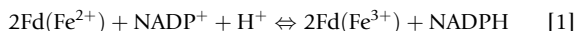


Figure 2 Ribbon model of the 3D structure of the pea photosynthetic FNR mutant with a serine replacing the C-terminal tyrosine residue. The FAD-binding and the NADP-binding domains are depicted in orange and green, respectively. The wireframes of bound FAD and NADP⁺ are shown in yellow and cyan, respectively. The figure was prepared using Molmol.

and the obligatory two-electron carrier NADP(H), according to the following balanced chemical equation:



The ability of FNR to pair single electrons in order to transfer them as a couple to a two-electron acceptor is a feature of the dehydrogenase-electron transferases class of flavoproteins to which FNR belongs, and is mediated by the FAD prosthetic group. FAD can exist in three redox states: oxidized, one-electron reduced (semiquinone form), and two-electron reduced (dihydroquinone form). During turnover, enzyme-bound FAD cycles among the three redox states described above, according to the reaction scheme shown in Figure 3.

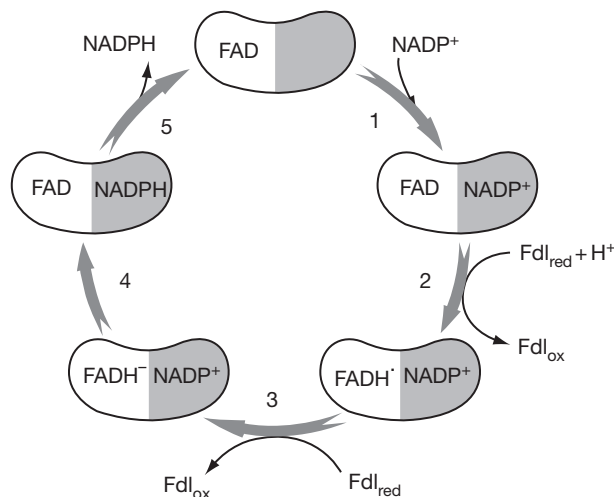


Figure 3 Simplified scheme of the mechanism of the reaction catalyzed by the photosynthetic FNR isoform. FdL indicates the photosynthetic Fd isoform present in chloroplasts. FdL_{ox} and FdL_{red} indicate oxidized and reduced forms of FdL, respectively. FADH[•] and FADH⁻ indicate the neutral semiquinone and the hydroquinone anion of FAD, respectively.

Two reduced Fd molecules sequentially transfer one electron at a time to FAD, before FNR can complete the catalytic cycle transferring a couple of electrons to NADP⁺. Electron transfer is strictly coupled to proton (H⁺) transfer, at least in steps 2 and 4. FAD protonation in step 2 is required to yield the neutral semiquinone form observed during turnover. In step 4, two electrons are transferred simultaneously with one H⁺ in the form of a hydride ion (H⁻). Catalysis proceeds according to a sequential-ordered mechanism, through the formation of ternary complexes of FNR with both substrates. The redox potential of the FAD/FADH[•] couple in FNR is -360 mV, a value intermediate between that of FdL (-400 mV) and that of the NADP⁺/NADPH couple (-320 mV).

Interaction with NADP(H)

FNR is a remarkably specific enzyme, showing a strong preference for NADP⁺ over NAD⁺. Substrate discrimination is due to a number of interactions that FNR establishes with the 2'-phosphate group of NADP(H). NADP(H) interactions with FNR have been described in terms of a bipartite binding mode, where the adenylate moiety of the dinucleotide binds independently of the nicotinamide one. Binding of the adenylate moiety occurs with high affinity and is the leading event in substrate recognition. In order to allow hydride transfer to occur between the reduced isoalloxazine ring and the nicotinamide, FNR has to undergo an induced-fit process where the nicotinamide ring of NADP⁺ displaces the phenol ring of the C-terminal tyrosine residue from its stacking interaction with *re*-face of the isoalloxazine (Figure 2). A detailed picture of the enzyme-NADP(H) interaction has been obtained by producing an engineered FNR, where the C-terminal tyrosine was changed to a nonaromatic residue. This greatly stabilized the interaction between the flavin and the nicotinamide ring. Thus, X-ray crystallography of the NADP(H) complex with the mutant FNR revealed the hydride-transfer-competent-binding mode of the substrate (Figure 2).

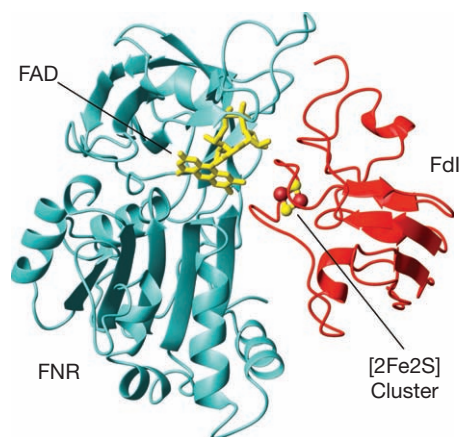


Figure 4 Ribbon model of the binary complex between maize photosynthetic FNR and FdI. FNR is depicted in cyan, while FdI is in red. The wireframe of FAD is shown in yellow. The iron–sulfur cluster is represented as a ball and stick model, with the iron atoms in brown and the sulfur atoms in yellow. The figure was prepared using Molmol.

Interaction with Fd

Crystals of the complexes between FNR and Fd, both in the oxidized state, have been obtained for the eukaryotic (maize leaf and root) and the prokaryotic (*Anabaena* spp.) protein couples. X-ray structures are similar (Figure 4). Fd binds to a large concave region of FNR, which is contributed by both structural domains of the reductase, with the FAD-binding domain providing most of the interactions. Complex formation places the redox active prosthetic group at a distance (6.0–7.4 Å) compatible with high-rate electron transfer. Protein–protein complex formation is thought to be initially driven by the mutual orientation of the dipole moments of the two molecules as they approach. After initial docking, minor adjustments are expected to occur to reach a conformation optimal for intermolecular electron transfer, with the formation of a network of intermolecular weak bonds (salt bridges, hydrogen bonds, van der Waals, and hydrophobic interactions). Charge–charge interactions, which play a major role in the stabilization of the Fd–FNR complex, are formed between basic groups provided by FNR and acidic groups of Fd.

Nonphotosynthetic Plant-Type FNRs

Root FNR

Nonphotosynthetic plastids of higher plants contain an isoform of FNR, also known as ‘root FNR’. Root and leaf isoforms of FNR are clearly homologous, showing a sequence identity of ~48%, which corresponds to highly similar 3D structures. Isoform-specific differences are limited to the structure of surface loops of the molecules, the presence of a conserved disulfide bridge in the FAD-binding domain of the root isoform, a higher redox potential for the bound FAD of the root isoform, and different affinities for the Fd isoforms. Many of these structural and functional differences are instrumental for the diverse physiological role played by the two FNR isoforms. Indeed, the role of root FNR is opposite to that of the leaf isoform, that is, transfer

of reducing equivalents from NADPH to a root-specific Fd isoform to yield reduced Fd for several Fd-dependent pathways.

Bacterial and Apicomplexan FNR

FNRs belonging to the FNR structural family are present in some prokaryotic organisms such as *Escherichia coli* and *Azotobacter vinelandii*. A plant-type FNR has been recently identified in different species belonging to the *Apicomplexa*, a phylum of obligatory eukaryotic unicellular parasites, including the causative agents of malaria, toxoplasmosis, and coccidiosis. In the apicomplexan cell, the FNR is localized, together with a plant-type Fd, in a peculiar organelle, the apicoplast. Apicomplexan FNR is thus expected to play a metabolic function similar to that of root FNR, that is, production of reduced Fd to be used in reductive biosynthetic pathways. In particular, the apicoplast FNR/Fd electron transport system has been experimentally shown to be involved in the biosynthesis of isoprenoid precursors and in heme degradation. Due to the increasing importance attributed to the apicoplast metabolism in parasite pathogenicity, apicomplexan FNR is an attractive target for novel antiparasitic drugs.

Heterotrophic GR-Type FNRs

In addition to plant-type FNRs, various FNRs exist that – while being structurally and phylogenetically distinct from the former – show essentially the same catalytic activity, thus representing a noteworthy example of convergent evolution. Such FNRs show a significant sequence identity. The 3D structures of two of them (bovine adrenodoxin reductase and *Mycobacterium tuberculosis* FprA) have been determined and shown to be similar to that of GR. FNRs of the GR-like class thus have both domains based on the Rossmann fold motif for binding the nucleotides (FAD and NADPH). The prototype of such enzymes is adrenodoxin reductase (AdR), found in vertebrate mitochondria, where it transfers electrons from NADPH to an Fd known as adrenodoxin (Adx). AdR and Adx form an electron supply system for the cytochrome P450-catalyzed hydroxylation reactions to produce steroid hormones, vitamin D metabolites, and bile acids. An AdR-homologous protein (Arh1) has been recently found in *Saccharomyces cerevisiae*. By now, AdR and Adx have been shown to be ubiquitous in mitochondria of eukaryotes, where they play a role in the biogenesis of iron–sulfur clusters. More recently, an AdR-like enzyme (FprA) has been identified in *M. tuberculosis*, showing that they are not restricted to eukaryotes. FprA is thought to play a crucial role in the complex lipid metabolism of this pathogen as a source of reducing power for the unusually large numbers (~20) of cytochromes P450 present in this organism. This observation could eventually lead to novel pharmacological treatment for tuberculosis.

See also: **Bioenergetics:** Ferredoxin; F_X, F_A, and F_B Iron–Sulfur Clusters in Type I Photosynthetic Reaction Centers; **Metabolism** **Vitamins and Hormones:** Photosynthesis; **Protein/Enzyme** **Structure Function and Degradation:** Protein Folding and Assembly.

Further Reading

- Aliverti A, Pandini V, Pennati A, de Rosa M, and Zanetti G (2008) Structural and functional diversity of ferredoxin-NADP⁺ reductases. *Archives of Biochemistry and Biophysics* 15: 283–291.
- Ceccarelli EA, Arakaki AK, Cortez N, and Carrillo N (2004) Functional plasticity and catalytic efficiency in plant and bacterial ferredoxin-NADP(H) reductases. *Biochimica et Biophysica Acta* 1698: 155–165.
- Deng Z, Aliverti A, Zanetti G, et al. (1999) A productive NADP⁺ binding mode of ferredoxin-NADP⁺ reductase revealed by protein engineering and crystallographic studies. *Nature Structural Biology* 6: 847–853.
- Hurley JK, Morales R, Martínez-Júlvez M, et al. (2002) Structure–function relationships in *Anabaena* ferredoxin/ferredoxin: NADP⁺ reductase electron transfer: Insights from site-directed mutagenesis, transient absorption spectroscopy and X-ray crystallography. *Biochimica et Biophysica Acta* 1554: 5–21.
- Karplus PA, Daniels MJ, and Herriott JR (1991) Atomic structure of ferredoxin-NADP⁺ reductase: Prototype for a structurally novel flavoenzyme family. *Science* 251: 60–66.

Fibroblast Growth Factor Receptors and Cancer-Associated Perturbations

M Kornmann, University of Ulm, Ulm, Germany

M Korc, Dartmouth Medical School, Hanover, NH, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Malignant transformation Process of exhibiting qualities of malignancies including uncontrolled growth, anchorage-independent cell growth, and *in vivo* tumor formation.

mRNA splice variant An isoform of a protein resulting from the processing of different exons of a gene.

Fibroblast growth factors (FGFs) constitute a family of homologous heparin-binding polypeptides that presently consists of at least 23 members. FGFs play important roles in many biological functions, including development, pattern formation, cellular differentiation, metabolic regulation, tissue repair, angiogenesis, and mitogenesis. FGFs act by binding to a family of specific, high-affinity FGF receptors (FGFRs) that possess a discontinuous intracellular kinase domain. Perturbations in FGF–FGFR expression, localization, and signaling have been implicated in several pathological processes, including malignant transformation and tumor spread and metastasis.

Fibroblast Growth Factors

Genes

In vertebrates, 23 *Fgf* genes have been identified. Human *Fgf15* has not been described. Human *Fgf19* is most closely related to mouse *Fgf15* raising the possibility that human *Fgf19* may be identical to mouse *Fgf15*. Both genes are closely linked to the *Fgf3* and *Fgf4* genes. According to their evolutionary relationship, the 22 *Fgfs* in humans can be classified into several subgroups (Figure 1). Members of each subgroup may also share similar biochemical and developmental properties. For example, members of the *Fgf8* subgroup display similar receptor-binding properties and overlapping, yet distinct expression patterns during development, underscoring both their commonality and uniqueness.

Proteins

Among vertebrates, FGFs share 13–71% of their amino acid sequence homology. Most notably, there is an internal core region consisting of six identical and 28 highly conserved amino acids, which include the 10 amino acid residues interacting with the high-affinity FGFRs within an approximately 120 amino acid region. In FGF-1 and FGF-2, this area consists of 12 anti-parallel beta strands forming a structure exhibiting a threefold internal symmetry called a 'beta-trefoil'.

FGF-4–8, -10, -17–19, -21, and -23 all possess classical N-terminal leader sequences that may allow for rapid and efficient secretion. This sequence is also present in FGF-3 and -22. However, they are not efficiently secreted. By contrast, FGF-1, -2, -9, and -11 to -14, -16 and -20 lack a leader sequence for secretion. Nonetheless, FGF-9, -16, and -20 have a

hydrophobic region in the core domain allowing for secretion. Moreover, FGF-1 and FGF-2 may be released by an exocytotic mechanism, while FGF-11–14 are thought to remain intracellularly. As these four family members do not function as FGF ligands, they are also referred to as 'FGF homologous factors'. Posttranslational glycosylation, alternative splicing, and initiation of synthesis from N-terminal cytosine–uridine–guanine (CUG) sequences can result in several different molecular weight forms in the case of several FGFs.

Physiologic Actions

FGFs regulate diverse processes during developmental processes and exert a broad variety of functions in the adult organisms, including growth, survival, apoptosis, motility, and differentiation.

Fibroblast Growth Factor Receptors

Genes

FGFRs are glycoproteins encoded by four distinct *Fgfr* genes. A special feature of the FGFR family is the existence of several receptor isoforms for FGFR-1, -2, -3, and -4 that are generated by alternative messenger RNA (mRNA) splicing and result in different ligand-binding specificities. A fifth related receptor, *FGFR5* (also known as 'FGFRL1'), can bind FGFs, but is devoid of an intracellular kinase domain. The function of FGFR-5 which is also devoid of intracellular signaling capacity is presently unknown, but it may provide a binding site for FGFs.

Proteins

FGF signaling is generally mediated by a dual-receptor system, consisting of the high-affinity FGFRs and low-affinity heparan sulfate proteoglycan (HSPG) chains enhancing ligand presentation to the receptors. The structure of an FGF–FGFR complex is composed of two FGFR chains, two FGFs, and an HSPG chain. FGFRs are related to vascular endothelial growth factor receptors and platelet-derived growth factor receptors. The extracellular domain is usually composed of three immunoglobulin (Ig)-like domains (I–III), including a stretch of acidic residues (acidic box) between domain I and II, which is unique for the FGFRs. The hydrophobic transmembrane region is followed by a juxtamembrane domain and a split tyrosine kinase catalytic

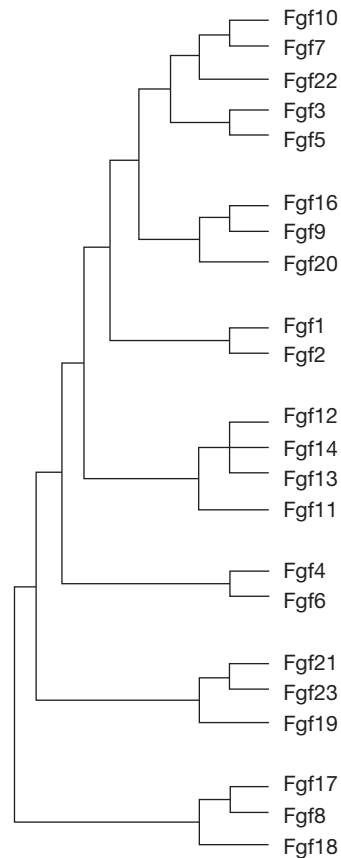


Figure 1 Evolutionary relationships within the human FGF family. Adapted from Ornitz DM and Itoh N (2001) Protein family review: Fibroblast growth factors. *Genome Biology* 2: 3005.1–3005.12.

domain (Figure 2). Most notably, alternative splicing of the second half of the Ig-like domain III results in three receptor variants, termed IIIa, IIIb, and IIIc (Figure 2). Due to the fact that the ligand-binding pocket is formed by Ig-like domains II and III, these splice are important for defining ligand-binding specificity. The first half of Ig-like domain III is encoded by exon IIIa yielding a secreted receptor that is devoid of any signaling capacity. In case exon IIIa is spliced to either exon IIIb or IIIc, transmembrane receptors are formed. The expression of the IIIb variants is believed to be restricted to epithelial cell types, whereas the expression of the IIIc variants, especially in the case of FGFR-2 and FGFR-3, is believed to be restricted to mesenchymal cell types.

FGF binding results in receptor oligomerization, activation of the cytoplasmic tyrosine kinase domains, and receptor autophosphorylation. Intracellular signaling is then mediated through the tyrosine phosphorylation of key substrates and the activation of downstream pathways such as the mitogen-activated protein kinases (MAPKs) extracellular signal regulated kinase 1 (ERK-1) and ERK-2 (Figure 3). The existence of at least 18 human FGF ligands, four transmembrane FGFRs including a tyrosine kinase domain, and multiple FGF isoforms and FGFR splice variants can potentially result in a large combinatorial set of interactions. This further increases the complexity, diversity, and functional specificity of FGFs.

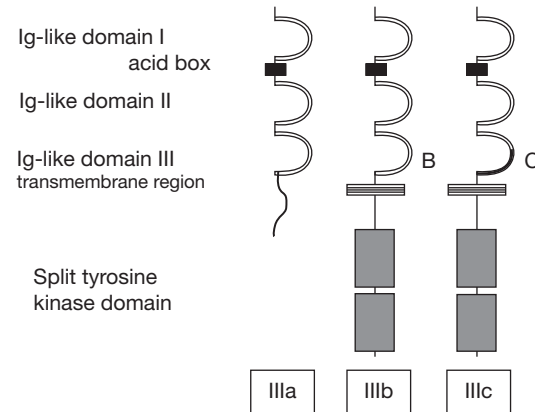


Figure 2 Schematic diagram of structural features of FGF receptors. FGFRs are usually composed of three extracellular immunoglobulin (Ig)-like domains (I–III), including a stretch of acidic residues (acidic box) unique for FGFRs, a hydrophobic transmembrane region followed by a juxtamembrane, and a split tyrosine kinase catalytic domain. The IIIa isoform is devoid of any signaling capacity. In the case of the IIIc isoform the second half of the Ig-like domain III (solid) corresponds to the exon 7 sequence of the *FGFR-1* gene, and in the case of IIIb the second half of the Ig domain III (open) corresponds to the exon 6 sequence of the *Fgfr1* gene.

Physiologic Actions

The expression of FGFRs seems to be fundamental for many processes, including limb, craniofacial, and lung development, hepatogenesis, osteogenesis, and inner ear formation. Mutations of *Fgfr1-3* genes have been linked to syndromes characterized by specific craniofacial and limb malformations. For example, mutations of *Fgfr2* were found in patients with Apert syndrome and mutations of *Fgfr3* in patients with achondroplasia.

FGF Actions in Cancer

Many studies demonstrate that alterations of the expression of FGFs and FGFRs play an important role in the development and progression of various malignant diseases. It is now well established that overexpression of FGFs can induce carcinogenesis through autocrine and paracrine mechanisms. The FGFR-mediated MAPK pathway is an important inducer of growth and has been implicated in tumor cell proliferation. By contrast, tumor-suppressive roles were described for FGFR1-IIIb and FGFR2-IIIb in the pancreas and prostate. FGFs can also regulate tumor cell survival pathways by increasing the expression of anti-apoptotic proteins such as XIAP and Bcl-X_L and triggering chemoresistance. Overexpression of FGFs was also associated with invasion and metastasis in a variety of human tumors. Tumor–stroma interactions are also of essential importance for tumor growth and extent of metastasis in several malignancies. Normal homeostasis of many organs depends upon the communication of epithelial and stromal compartments. This bidirectional signaling is mediated and controlled by multiple FGFs. In the case of the development of a malignant disease, specific ligands may gain growth-promoting functions by the expression of an FGFR variant

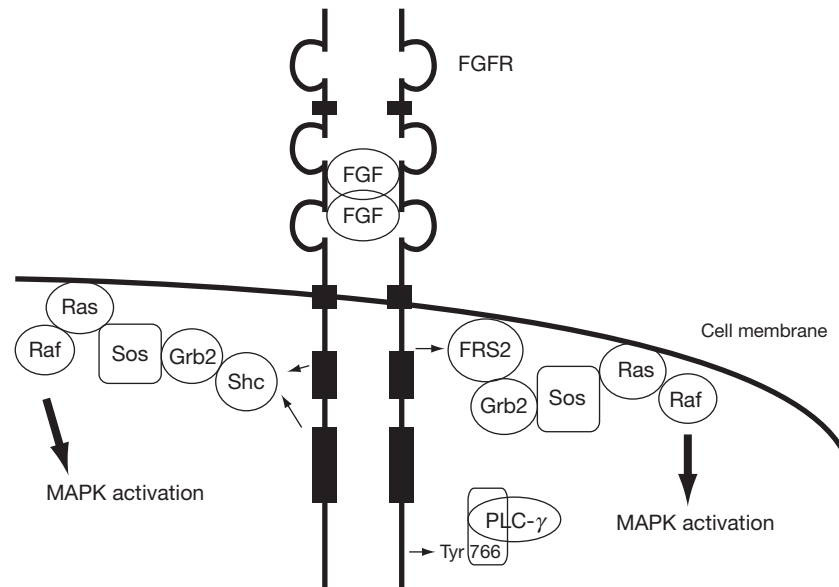


Figure 3 Schematic diagram of intracellular FGF signaling. FGF binding between Ig domain II and III results in receptor oligomerization, activation of the split cytoplasmic tyrosine kinase domains (rectangles) and receptor autophosphorylation. This is followed by binding, phosphorylation, and activation of fibroblast growth factor receptor substrate-2 (FRS2) or phospholipase C- γ (PLC- γ), which binds to tyrosine residue 766. Activation of the downstream mitogen-activated protein kinases (MAPK) extracellular signal regulated kinase -1 (ERK-1) and ERK-2 can occur via FRS2 or bypassing FRS2 and includes the signaling molecules Shc, Grb-2, SOS, ras, and raf.

usually not expressed in normal cells of the same origin. For example, human pancreatic stellate cells are able to induce FGFR1-IIIc and inhibit FGFR1-IIId expression in human pancreatic cancer cells, thereby enhancing the malignant phenotype. FGF-FGFR pathways are also involved in the regulation of epithelial-mesenchymal transition (EMT). EMT is a process where epithelial cells acquire properties of a motile and invasive mesenchymal-like phenotype. Epithelial cell characteristics are altered during EMT with changes in the expression of extracellular matrix proteins, proteins involved in cell adhesion and cell-cell interactions playing crucial roles in cancer progression and metastasis. FGF signaling also has direct effects on tumor angiogenesis by promoting cell proliferation in FGFR-expressing endothelial cells. Moreover, FGFs and other proangiogenic factors have also been implicated in the emerging phenomenon of resistance to vascular endothelial growth factor receptor-2 (VEGFR-2) inhibition.

FGF-2

FGF-2 exerts mitogenic effects on various endothelial, mesothelial, and epithelial cell types. It is overexpressed in a great portion of human tumor cell lines and various human malignancies. High levels of FGF-2 in serum of patients or in tumor specimens have been correlated with advanced tumor stage and poor prognosis in several gastrointestinal and other cancers, including thyroid, lung, and renal cell cancer. However, FGF-2 expression may be associated with favorable prognosis in ovarian and breast cancer.

FGF-1

FGF-1 is also a potent mitogenic factor for many normal and malignant cells and also expressed in a variety of cultured cell

lines and tissues. However, despite the fact that FGF-1 is present in many malignancies and can display similar biological activities as FGF-2, FGF-1 expression has not been clearly demonstrated to correlate with clinico-pathological parameters in human cancers.

FGF-3

The *Fgf3* gene was initially identified as murine mammary tumor virus integration site oncogene homolog (*int2*) implicated in mouse mammary tumors. *Fgf3* is localized on chromosome 11q13, a locus frequently amplified in malignant diseases. Amplification of *Fgf3* was associated with poor prognosis in breast cancer and biological aggressiveness in ovarian and esophageal cancer. The overexpression of FGF-3 in normal mouse mammary cells resulted in the development of orthotopically implanted tumors in nude mice, emphasizing the important role of FGF-3 in mammary tumorigenesis.

FGF-4

Testing the capacity of DNA from stomach tissues for their malignant transformation of fibroblasts resulted in the identification of *Fgf4*, also called *human stomach cancer transforming factor-1* (*hst-1*). The same oncogene was isolated by transfection of Kaposi sarcoma DNA and is therefore also called *kaposi sarcoma Fgf* (*kFgf*). *Fgf4* is also located on chromosome 11q13 together with *Fgf3*. Amplification of *Fgf4* correlates with increased risk of systemic recurrence after resection of esophageal cancer and with nodal involvement in head and neck cancers. FGF-4 overexpression increases the metastatic potential of breast cancer cells in association with altered expression of matrix metallo-proteinases.

FGF-5

Fgf-5 was also initially identified as a transforming proto-oncogene. The effectively secreted FGF-5 is produced by stromal cells (fibroblasts and pancreatic stellate cells) and acts primarily via FGFR-1 IIIc. FGF-5 is also expressed in breast cancer, several cancers of the gastrointestinal and urinary tract, and in glioblastoma multiforme. FGF5 can contribute to the malignant progression of human astrocytic brain tumors and pancreatic cancer by both autocrine and paracrine effects.

FGF-6

FGF-6 or human stomach transforming factor-2 also displays transforming potential and was found in prostate cancer in conjunction with FGFR-4. Exogenous FGF-6 enhances the proliferation of primary prostatic epithelial and stromal cells and prostate cancer cell lines, suggesting that FGF-6 may act as paracrine and autocrine factor in prostate cancer.

FGF-7

FGF-7 is an important mitogen for a variety of epithelial and cancer cells, including liver, pancreas, gastrointestinal tract, lung, and breast. Inhibition of FGF-7 signaling results in reduced mobility and invasion of breast and gastric cancer cells. FGF-7 is overexpressed in many cancers, but is expressed at lower levels in squamous cell carcinomas of the head and neck and endometrial cancers by comparison with the corresponding normal tissue. A similar observation has been made for serum levels of FGF-7, which were lower in patients with prostate cancer in comparison to patients with benign prostatic hyperplasia. Together, these observations suggest that FGF-7 may localize in different cell types within these pathological tissues and may differentially modulate tumor growth depending on the expression and localization of the FGF receptors in these tissues.

FGF-8

FGF-8 or androgen-induced growth factor was isolated from SC-3 androgen-stimulated mouse mammary carcinoma cells and found to be overexpressed in breast cancer. *Fgf-8* has been proposed to enhance murine mammary tumorigenesis in cooperation with the *Wnt-1* proto-oncogene. It was found that the four known human FGF-8 isoforms (a, b, e, and f) display different proliferating and transforming potencies, the FGF-8b isoform being the most potent. Thus, constitutive expression of FGF-8b resulted in enhanced invasion, angiogenic capacity, and tumor growth of breast cancer cells. FGF-8b initiated prostatic intraepithelial neoplasia and prostate cancer when expressed in a *Pten* haplosufficient background, underscoring its potential role for these malignancies.

FGF-9 to -23

Expression of FGF-9 or glia-activating factor, FGF-10, and FGF-19 was detected in various cancer tissues. These FGFs are also able to act as mitogens. Thus, FGF-10 was able to induce by autocrine and paracrine mechanisms pulmonary and prostate cancers. FGF-19 transgenic mice that overexpressed FGF-19 in skeletal muscle exhibited elevated hepatic α -fetoprotein levels at 2 months of age and developed hepatocellular

carcinomas by 10 months of age. In this context, the effects of FGF-19 are hormonal. The effects of FGF-19 on the development of hepatocellular carcinomas are mediated almost exclusively via FGFR-4.

FGFR Actions in Cancer

Aberrant FGF signaling in cancer underlies tumor-specific mechanisms. They can be divided into genomic FGFR alterations responsible for ligand-independent receptor signaling and into changes in receptor isoform expression supporting ligand-dependent activation. The ligand-dependent signaling has a similarly important role in the pathogenesis of cancer, by either autocrine or paracrine production of ligands from cancer or stromal cells, respectively.

FGFR-1

FGFR-1 is expressed in a wide variety of cell types and tissues, including many malignancies. Amplification of the *Fgfr1* locus on chromosome 8p12 is one of the most common focal amplifications in breast cancer (~10%), predominantly in estrogen-receptor-positive cancers. It was also reported in oral squamous carcinoma, ovarian and bladder cancer, and rhabdomyosarcoma and is associated with nodal involvement in breast cancer and poor prognosis of patients with bladder cancer. An atypical stem-cell myeloproliferative disorder evolving toward acute myelogenous leukemia is characterized by a 8p12–p11 breakpoint and a t(8;13) translocation resulting in an aberrant formation of a tyrosine kinase with the catalytic domain of FGFR-1 and constitutive kinase activity. This translocation was also found in breast cancer. Expression of *Fgfr1* in transgenic mice suggests that FGFR-1 plays an important role in mammary gland tumorigenesis. Deletions of 8p12 have been reported for choriocarcinoma, and frequent disturbances, including amplification or deletions, have been reported in prostate cancer. Overexpression of wild-type FGFR1 was also reported in cancer. FGFR-1 expression can be induced by overexpression of cyclin D1 in fibroblasts via the pRb/E2F pathway. Cyclin D1 is known to be overexpressed in many malignancies and correlates with poor prognosis in several malignancies. Recent evidence also suggests that N-cadherin expression can enhance tumorigenesis via FGFR-1. FGFR-1 ligand-induced internalization was reduced by N-cadherin mediated by the first two Ig-like domains and resulted in increased FGFR-1 stability and sustained activation of MAPK.

A potent ligand of FGFR-1, FGF-2, has been shown to be an important prognostic factor for many malignancies. Nevertheless, high FGFR-1 expression was only associated with poor differentiation in prostate cancer, advanced stage in head and neck squamous cell carcinoma, and nonsmall cell lung cancer. In addition, recent evidence suggests that alterations in the expression of certain splice variants may be responsible for malignant transformation. Thus, expression of FGFR-1 IIIc in normal ductal pancreatic epithelial cells resulted in the establishment of an autocrine loop and cellular transformation with increased proliferation and *in vivo* tumor formation. Expression of FGFR-1 IIIc in the cancer cells, which are of epithelial origin, has been linked to the pathogenesis of pancreatic and

prostate cancers. Additionally, altered expression of FGFR-1 splice variants has been reported for breast cancer. Especially, the 2-Ig-like isoform (also called FGFR-1 β) is upregulated in pancreatic, prostate, and breast cancers, and in astrocytomas and glioblastomas in comparison to the 3-Ig form (FGFR-1 α). More recent data suggest that besides enhancing proliferation, FGFR1 signaling can initiate cancer development and promote invasion and metastasis.

Inhibition of FGFR-1 signaling using dominant-negative or soluble forms of the receptor resulted in significant growth inhibition in pancreatic, breast, prostate, and several other cancer cell lines. Using vaccination with *Xenopus* FGFR-1 in a murine tumor model resulted in the production of FGFR-1-specific auto-antibodies and effective anti-tumor activity and suppression of angiogenesis. By contrast, over-/re-expression of FGFR1-IIIb in ductal pancreatic cancer cells resulted in the reversion of the malignant phenotype, also pointing to tumor suppressive effects of FGF-FGFR pathways in cancer. Recent studies in pancreatic cancer about the expression of the tumorigenic FGFR1-IIIc isoform and the tumor-suppressive FGFR1-IIIb isoform revealed that several growth factors can induce IIIc expression via cyclin D1 and inhibit IIIb expression. Tumor-suppressive functions of FGFR1-IIIb were associated with increased expression of osteonectin, a matricellular glycoprotein important for cell-cell contact and proliferation. The change from FGFR1-IIIb to FGFR1-IIIc expression in pancreatic cancer may also be a result of the EMT similar to FGFR-2 in prostate cancer.

FGFR-2

FGFR-2 is also found in a wide variety of cell types and tissues and was first identified as an amplified gene from a human gastric cancer cell line. Amplification of *Fgfr2* on chromosome 10q26 is observed in about 10% of gastric cancers and also in breast and oral squamous cell carcinomas. One important isoform, keratinocyte growth factor (KGFR), is devoid of the Ig-like domain I and the acidic box and contains sequences encoding the IIIb variant. Overexpression of KGFR occurs in gastric and pancreatic cancers, and is associated with poor differentiation in prostate cancer and poor prognosis in pancreatic cancer, whereas KGFR expression is lower in endometrial cancer in comparison to the normal endometrium.

An essential feature of *Fgfr2* is its proposed strictly tissue-specific expression using exon IIIb in epithelial cells and IIIc in mesenchymal cells. Exon switching from IIIb to IIIc in conjunction with the presence of stromal cell-derived FGF-7 was accompanied with malignant progression in a prostate cancer model, downregulation of IIIb with malignant progression of the prostate, and de-differentiation of normal oral keratinocytes into a malignant phenotype. This exon switch from FGFR-2 IIIb to FGFR-2 IIIc can be induced by the presence of FGF-1 and -2 in keratinocytes, fibroblasts, and bladder carcinoma cells. Re-expression of KGFR by transfection, in human salivary gland adenocarcinoma cells which lost KGFR expression in the process of malignant transformation, resulted in growth inhibition, induction of differentiation, and apoptosis. *Fgfr2* mutations linked to craniosynostosis syndromes and resulting in constitutive receptor activation with transforming potential were also detected in endometrial and gastric cancers,

suggesting also that gain-of-function mutations of *Fgfr2* can play a role in tumorigenesis.

FGFR-3

Mutations of *Fgfr3* were identified in a great portion of bladder carcinomas and at low frequency in cervical, urothelial cell, prostate, and colorectal carcinomas, multiple myeloma, and seminomas. Many of these mutations occurred at highly conserved sequences in the Ig-like domain III, highlighting the functional importance of this domain. Interestingly, the presence of FGFR-3 mutations in bladder and urothelial cell carcinoma was associated with better outcome. Amplifications of *Fgfr3* in cancer are rare events.

Nuclear accumulation of FGFR-3 has been reported for breast cancer. This is probably related to a splice variant in which exons 7 and 8 are deleted resulting in the translation of a soluble intracellular FGFR-3 missing the transmembrane domain, but with an intact kinase domain. Using other splice variants that are devoid of the acidic box, it could be shown that inhibitory functions of FGFR-3 can be abolished. Other variants with altered ligand-binding specificities were described in osteosarcoma and squamous carcinoma cell lines. Thus, depending on the mutations and the receptor splice variant expressed, FGFR-3 can display transforming potential, supporting the possible role of FGFR-3 in the pathogenesis of some malignancies. Moreover, *Fgfr3* translocations have been described for hematological malignancies.

FGFR-4

A G to A conversion resulting in a substitution of glycine by arginine at position 388 in the transmembrane domain of the *Fgfr4* gene has been described. Expression of this receptor in mammary tumor cells increased cell mobility and the presence of this allele was associated with lymph node metastases and more advanced disease in colon cancer and reduced disease-free survival in breast cancer. Blocking FGF19 antibodies further suggested that proliferation of colon cancer cells may be driven by an autocrine loop involving FGFR4. Amplification of *Fgfr4* on 5q35.1-qter has been described for some breast and ovarian cancers, leukemias, and lymphomas. FGFR-4 overexpression was reported in endocrine tumors of the digestive system, several gastrointestinal and urinary tract cancers as well as breast and lung cancers. FGFR-4 has also been correlated with low differentiation and tends to be associated with shorter patient survival in astrocytomas.

Furthermore, a truncated form of FGFR-4 lacking the first two Ig-like domains was identified in human pituitary tumors. Because the N-terminus of this truncated receptor lacks a signal peptide, it is not inserted into the plasma membrane and resides in the cytoplasm. The truncated receptor is constitutively phosphorylated and exhibits transforming potential in fibroblasts. Its expression in transgenic mice results in the formation of pituitary tumors suggesting that this isoform may play an important role for the development of pituitary tumors. Experiments using dominant-negative FGFR-4 also suggest that FGFR-4 may be involved in matrix adhesion and, therefore, tumor dissemination. Expression of FGFR-4 induced membrane ruffling, a sign of transformation of breast cells, in

the presence of FGF-1 or -2. Furthermore, differentiation of teratocarcinoma cells was associated with loss of FGFR-4 expression.

Summary

Besides their involvement in many biological processes during development and in the adult, FGFs and FGFRs are involved in the pathobiology of many malignancies. One important reason for this role seems to be the altered expression of FGFs and FGFRs and their isoforms resulting in the aberrant activation of autocrine and paracrine signaling loops that eventually promote tumorigenesis. Other changes include the expression of abnormal FGFs or FGFRs with constitutive activation of FGFR tyrosine kinase signaling. Thus, the discovery and functional analysis of the role of FGF–FGFR alterations in specific cancer entities and other malignancies may result in the treatment

strategies in the future targeting these pathways or using these pathways for specific treatment.

See also: **Metabolism Vitamins and Hormones:** Fatty Acid Metabolism and Cancer.

Further Reading

- Beenken A and Mohammadi M (2009) The FGF family: Biology, pathophysiology and therapy. *Nature Review Drug Discovery* 8: 235–253.
- Grose R and Dickson C (2005) Fibroblast growth factor signaling in tumorigenesis. *Cytokine Growth Factor Reviews* 16: 179–186.
- Korc M and Friesel RE (2009) The role of fibroblast growth factors in tumor growth. *Current Cancer Drug Targets* 9: 639–651.
- Ornitz DM (2000) FGFs, heparan sulfate and FGFRs: Complex interactions essential for development. *BioEssays* 22: 108–112.
- Ornitz DM and Itoh N (2001) Protein family review: Fibroblast growth factors. *Genome Biology* 2: 3005.1–3005.12.
- Turner N and Grose R (2010) Fibroblast growth factor signaling: From development to cancer. *Nature Reviews Cancer* 10: 116–129.

Flavins

B Entsch and D P Ballou, University of Michigan, Ann Arbor, MI, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Fatty acyl-CoA Derivatives of fatty acids (CoA thioesters) formed in cells to initiate oxidation of the fatty acid.

Glyoxisome/peroxisome Common membrane-enclosed compartments in plant and animal cells.

Hydroperoxide A derivative of hydrogen peroxide where one of the H atoms is replaced by a chemical group.

Isoalloxazine Fused three-ring structure that forms the core chemical function of flavins.

Redox Shorthand for reduction and oxidation processes.

The word flavin in biochemistry comes from the Latin word meaning yellow. The root word flavin was used in the early twentieth century to refer to yellow compounds extracted from tissue samples. Richard Kuhn is credited with the preparation of pure riboflavin (which he called lactoflavin – from milk) and its structural determination. Kuhn received the Nobel prize for this and other work in 1938. Today, the term ‘flavin’ is reserved for a group of compounds that contain the fused three-ring structure called isoalloxazine, which is responsible for the yellow color.

The most important flavins in nature are riboflavin, flavin mononucleotide (FMN), and flavin adenine dinucleotide (FAD) as shown in [Figure 1](#). Riboflavin is often known as vitamin B2 (one of the soluble vitamins), which is an essential requirement in the diet of mammals and many other organisms. Riboflavin is manufactured by most microorganisms and plants. There is now a great deal of information on the process of biosynthesis. Organisms that require riboflavin as a vitamin use specific enzymes and adenosine triphosphate (ATP) to convert the compound into both FMN and FAD. These nucleotides are used as cofactors (or prosthetic groups) in the structure and function of many proteins. The relationship with proteins is so important that we normally think of flavins in the context of protein structures. Thus, we refer constantly to flavoproteins as proteins with a flavin cofactor. This article describes the fundamentally important functions of these compounds in a wide variety of chemical processes in nature.

Chemistry

The chemistry of flavins has been studied extensively. Riboflavin is a stable compound with a variety of interesting properties. FMN and FAD are much more important in cells, but are less stable in solution because of the presence of phosphate ester and anhydride bonds. The isoalloxazine is responsible for the important biological chemistry of all flavins. Only the reactions most relevant to life are mentioned below. One reason that flavoproteins have been so extensively analyzed is that they can be studied by examining the properties of their cofactor that is nearly always involved in the core function of each protein. Another reason is the variety of chemical properties that have been exploited for important biological functions of considerable diversity.

Redox Properties

Oxidized flavins (called flavoquinones) can be reduced by the addition of two electrons and two protons, and the redox potential for reduction is -0.2 V at pH 7 ([Figure 2](#)). Reduced flavins (flavohydroquinones) are thus moderate reducing agents and can occur as either a neutral or an anionic form. Part of the versatility of flavins comes from their ability to be reduced in one-electron steps because flavins are capable of forming stable free radicals (flavosemiquinones in [Figure 2](#)). A neutral and an anionic free radical can be formed by flavins, and both occur in nature. Proteins modify these redox properties by the interactions formed with the flavin cofactor to enable a wide variety of biological processes. Flavins can pass electrons to transition metal ions. Many proteins make use of this property to link stable reducing compounds (e.g., NADH) to reactive metal ions and one-electron processes in electron transport chains. Flavins are almost always involved as the transforming molecule for linking two-electron to one-electron processes.

Oxygen and Other Derivatives

Flavins, unlike most organic molecules, react with oxygen. This reactivity is a product of the redox properties mentioned above. Oxygen is unreactive unless free radical intermediates are formed. Reduced flavins provide the capacity for one-electron reductions and, in addition, have the ability to form reactive covalent adducts with oxygen and other molecules. Bruice and colleagues have shown that the reactions between reduced flavins and oxygen involve covalent adducts and free-radical intermediates. Many proteins take advantage of the oxygen adducts ([Figure 3](#)) to help carry out chemical reactions. A wide variety of biological processes depend upon the specific interactions between particular proteins and these flavin adducts. A small number of proteins form stable covalent bonds to the flavin through the 8- and/or the 6-positions ([Figure 1](#)).

Spectroscopic Properties

Oxidized flavins (the stable form) are normally yellow due to a major electronic transition observed at 450 nm ([Figure 4](#)). Full reduction of flavins (two equivalents) gives a structure that has

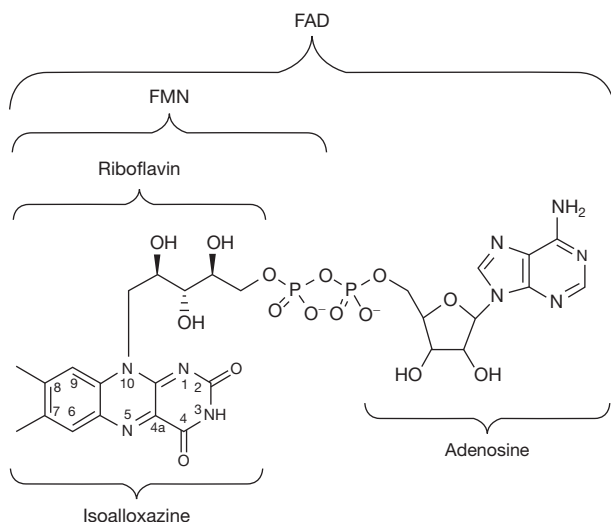


Figure 1 The structures of flavins discussed in this article. The numbering system shown for the isoalloxazine ring is used in the text.

a low absorbance beyond 400 nm, and is almost colorless (pale yellow) by comparison to oxidized flavins. Free-radical flavins (flavosemiquinones) have complex spectra with spectacular colors – either red or blue (**Figure 4**), depending on the state of protonation. Flavins also form charge-transfer complexes with a variety of other compounds. These complexes have a range of different visible spectra with a corresponding variety of colors. Such complexes generally have weak absorbance bands between 500 and 800 nm. Various proteins can stabilize one or more of these chemical species, so that flavo-proteins can have many colors, besides changing color during catalytic function.

Flavins have important luminescent spectral properties. Many flavins are fluorescent, and the fluorescence properties are strongly influenced by association with proteins and by the oxidation state or formation of derivatives. These properties can be studied in biological systems because the flavin can be excited by light (between 350 and 480 nm) and the emitted light is a broadband with a peak near 530 nm. This property provides a sensitive assay of flavins down to 10^{-9} M. The

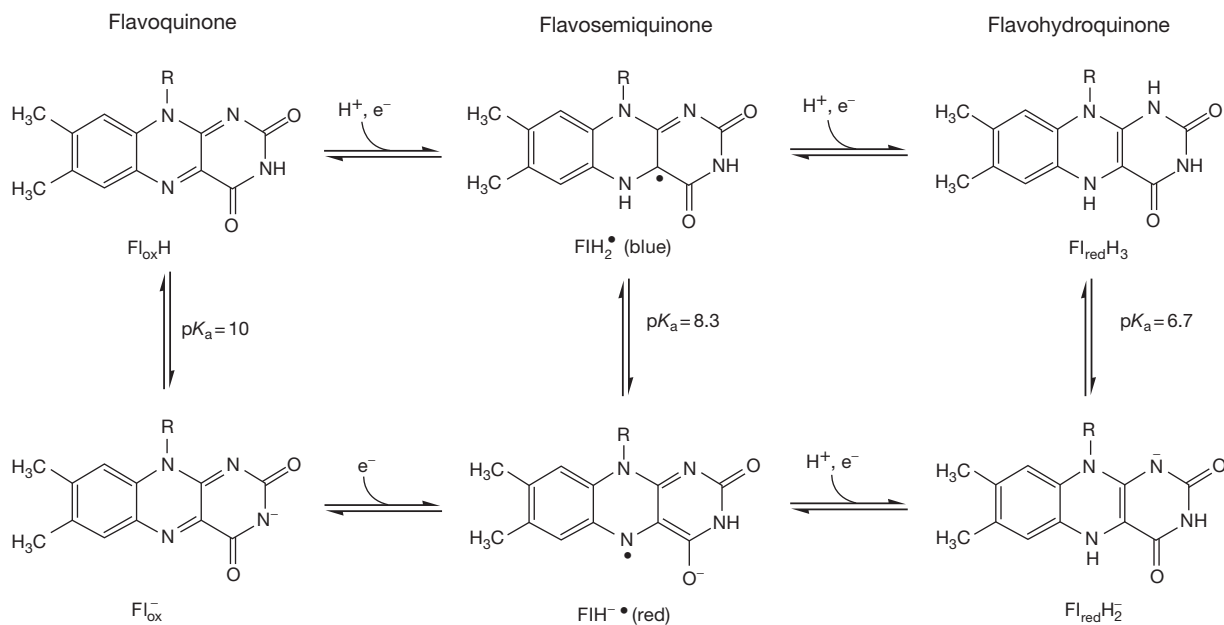


Figure 2 The redox states of the isoalloxazine ring that are observed in natural systems.

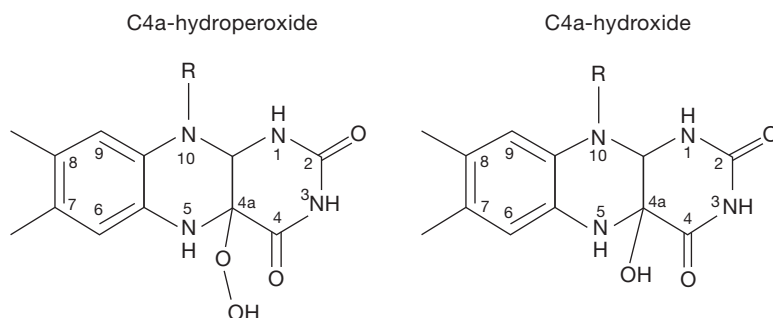


Figure 3 The structure of transient oxygen derivatives of the isoalloxazine ring that are important in protein function. Note that thiol groups can form a similar structure to the C4a-hydroxide. The biological function of the sulfur derivative is discussed in this article.

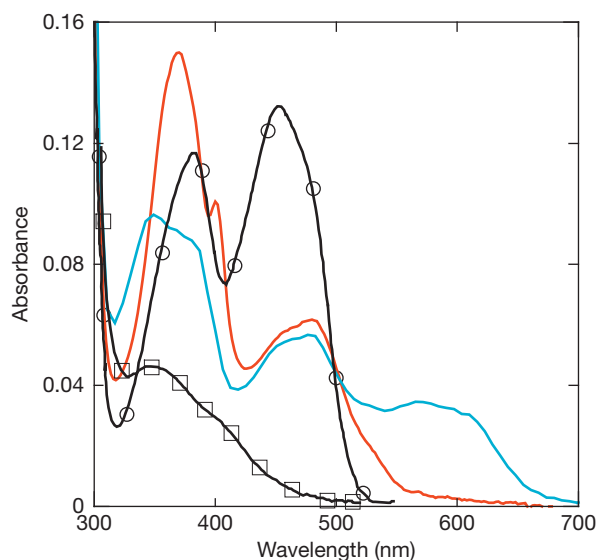


Figure 4 Absorption spectra of the redox states of FAD in glucose oxidase. A solution of 9.5 μM oxidized enzyme (solid line/open circles) was photoreduced in anaerobic conditions at pH 9.2 to form the red, anionic semiquinone. The solution was shifted to pH 5.0 with citric acid to form the blue semiquinone, and then reduced with glucose to form the hydroquinone (solid line/open squares).

fluorescent properties of flavoproteins often provide information not available from absorbance properties. Flavins can even be trapped in excited states and this has been exploited in living systems. Nature has also utilized the enhanced reactivity of the excited states of flavins that last long enough to be chemically useful. For example, the 4a-hydroxyflavin is often fluorescent, whereas the hydroperoxyflavin is not, even though they have nearly identical absorbance spectra.

Types of Flavoproteins

To try to display the diversity and utility of flavin function in living systems, we adapt below the classification system for flavoproteins developed by Massey.

Pyridine Nucleotide Oxidoreductases

There is a diverse group of proteins that connects the stable reducing currency of cells (NADH/NADPH) with other redox processes required for cellular function. For example, glutathione reductase is one of a group of related disulfide reductases that contain both FAD and a pair of reactive cysteine residues in the active site. The function of this enzyme is to help maintain a reducing environment within cells by keeping glutathione reduced. Glutathione then carries out many important functions such as reducing disulfides in proteins and removing hydrogen peroxide from cells (catalyzed by glutathione peroxidase). Through the flavin buried in the protein, the enzyme can use the reducing power of NADPH to reduce the disulfide of glutathione, because the flavin reacts rapidly with thiols

by the formation of a transient covalent adduct (Figure 3). Another example is ferredoxin reductase, the enzyme principally associated with photosynthesis that enables reactive, strongly reducing proteins such as ferredoxin to provide stable reducing equivalents in the cell as NADPH. Through the FAD in the active site, the enzyme can accept single reducing electrons from ferredoxin (electrons collected from the reactive free-radical photochemical process) that can then be safely stored as NADPH.

Dehydrogenase/Electron Transferase

There are a limited number of enzymes that use flavins to link oxidation of a stable metabolite to electron transport chains for energy conservation in cells. For example, succinate dehydrogenase is a component of the citric acid cycle in mitochondria and it links directly into the respiratory complexes through an iron/sulfur cluster complex. The FAD in this enzyme is covalently linked to the protein where it accepts a pair of electrons from the oxidation of succinate and then transfers single electrons into the respiratory complex II. The acyl-CoA dehydrogenases in animals have a similar fundamental role in obtaining energy from the oxidation of fats. These enzymes contain FAD, which accepts electrons and the latter accepts electrons from a fatty acyl-CoA to initiate oxidation of a fatty acid. In this case, the electrons are transferred to the respiratory chain in mitochondria through another soluble flavoprotein called the electron transferring flavoprotein. In plants, the process of fatty acyl-CoA oxidation occurs in glyoxisomes (rather than mitochondria) by the use of similar enzymes that are oxidases. That is, after being reduced by the fatty acyl-CoA, the enzyme flavin reacts with oxygen to dispose of the reducing electrons. The difference between these enzymes in animals and plants is a subtle change in the protein interactions with the FAD that changes the course of the reaction.

Electron-Transfer Proteins

There is a small group of proteins that act to provide single reactive reducing electrons for a variety of biological reactions. Most of these proteins contain iron/sulfur complexes, but a few are flavoproteins that use the ability of flavins to form stable free radicals. For example, flavodoxins are widespread in microorganisms and have been extensively studied. The protein in flavodoxins makes a specific interaction with the isoalloxazine ring that stabilizes the blue neutral radical (Figures 2 and 4), and makes further reduction difficult. Thus, when reduced, flavodoxins are strong one-electron reducing agents. Electron-transferring flavoprotein referred to earlier is a different member of this family of proteins.

Oxygenases

Flavins are widely employed in cells to unlock the reactivity of the oxygen molecule. Oxygenases not only react with oxygen, but also incorporate oxygen into metabolites. There are

many types of flavoprotein oxygenases that tame oxygen for useful purposes. For example, *p*-hydroxybenzoate hydroxylase adds one atom of oxygen to an aromatic ring so that it can eventually be used for carbon and energy supply for a microorganism. Oxygenated aromatic rings are then converted to aliphatic compounds by subsequent metabolic steps. *p*-Hydroxybenzoate hydroxylase contains FAD in the structure and has been studied extensively to understand this type of reaction. In catalysis, the reduction of FAD by NADPH is tightly controlled in a remarkable process. The enzyme senses the presence of the substrate *p*-hydroxybenzoate in the active site and responds by swinging the isoalloxazine ring to the surface of the protein where it contacts the NADPH. This process makes use of the flexibility of the ribityl side chain of FAD. Then, after reduction, the isoalloxazine swings back into the interior of the protein where it forms a transient covalent adduct (Figure 3, hydroperoxide) in the controlled environment of the enzyme. The peroxide provides the reactive oxygen to oxygenate the substrate. Cyclohexanone monooxygenase provides a different example where the FAD, after reduction by NADPH, reacts with oxygen to form a peroxide anion that reacts rapidly with the carbonyl group of cyclohexanone. Another variation is found in lactate monooxygenase. Here, the enzyme flavin (FMN) is reduced by the substrate lactate to form pyruvate that is held by the enzyme and oxygenated by the peroxide formed from the reaction between reduced FMN and oxygen.

Oxidases

Flavoproteins can also use oxygen as a sink for removing unwanted reducing electrons. This process was briefly referred to above in relation to fatty acyl-CoA oxidase in plants. Another important example is glycolate oxidase in plants. During photosynthesis, glycolate is formed as an unwanted byproduct and must be removed. Glycolate oxidase uses FMN to oxidize glycolate to glyoxalate, which is then aminated to form glycine, a normal amino acid. This process occurs in peroxisomes where oxygen is used to react with reduced FMN to remove the reducing electrons from glycolate, rather than through an electron transport chain. The oxygen is reduced to hydrogen peroxide (similar to the oxygenases), but the peroxide is released from the enzyme and must be removed from the cell by other enzymes such as catalase to prevent dangerous side reactions. Other examples of this use of oxygen by flavoproteins are amino acid oxidases in kidneys and glucose oxidase in fungi. It is the absence of a flavoprotein oxidase function (gulonolactone oxidase) that explains why humans require vitamin C in their diets, when most animals can make this molecule. In humans, the gene for this enzyme has mutated and no longer codes for a functional protein.

Photochemistry

Flavoproteins show some really novel functions when the properties of their excited states are utilized in cells. A classic example is bacterial luciferase, which is an enzyme that

produces lime-green light from chemical energy. The light-producing process is an outgrowth of the reactions of the oxygenases. NADH is used to reduce FMN by one enzyme and then the reduced FMN binds to luciferase where it reacts with oxygen to form the flavin C4a-hydroperoxide (Figure 3). The novel reaction occurs when the hydroperoxide oxidizes a fatty aldehyde to form the corresponding fatty acid. In this reaction, the protein traps the FMN in an excited state of the flavin C4a-hydroxide (Figure 3), which can decay with the release of a photon. Many marine animals in the deep oceans form symbiotic relationships with the bacteria that make luciferase. The animal can then form light-emitting organs essential for communication, hunting, or defense.

Just as remarkable as luciferase is DNA photolyase, which is found in all cells. An essential part of cellular survival is the maintenance of the integrity of genetic information in the face of damage by environmental factors. One source of damage is UV light, which can cause cross-linking of bases in the DNA helix. It is the task of DNA photolyase to repair the damage often by using the same source of light that caused the damage. This enzyme contains both an FAD and an accessory pigment. The flavin functions in the protein in the reduced state, where it is converted to a powerful reducing agent when excited by energy from light. The excited reduced state is powerful enough to break down the unwanted covalent bonds in damaged DNA and thus repair the integrity of the genetic information. Electrons used for this purpose are then recovered to reform the reduced FAD.

Another developing story about flavoprotein photochemistry involves flavins as blue-light photoreceptors. For example, phototropins are photoreceptors for some plant responses to light. These proteins contain FMN that forms a transient and reversible covalent bond to an active site cysteine upon illumination. This complex is thought to form through a free-radical mechanism of excited state flavin chemistry. In addition to these, cryptochromes are flavoproteins similar in structure to DNA photolyase (above), and they provide modulation of circadian rhythms in animals upon irradiation with blue light.

Medical Significance

It is extremely rare for riboflavin deficiency to occur in humans because of its widespread occurrence in food. However, some flavoproteins are important in medical practice. A few examples are mentioned below.

A common medication for gout (caused by excess uric acid) is allopurinol. This drug inhibits xanthine oxidase, a complex flavoprotein involved in the formation of the uric acid that causes gout. Uric acid is the end product of the natural breakdown of purines in the body.

Another complex flavoprotein, monoamine oxidase, is present as two different proteins in the nervous system. Both are targets for modulation by drug therapy, particularly for depression, because this enzyme degrades some of the potent natural neurotransmitters involved in mental disorders.

The flavoprotein, methylenetetrahydrofolate reductase, is involved in the formation of the amino acid methionine. About 12% of the human population has a form of the enzyme

that leads to elevated homocysteine in the blood, a major risk factor in cardiovascular disease.

Some medications for stomach ulcers work because they are substrates for the bacterial flavoprotein, nitroreductase, produced by the ulcer-causing *Helicobacter pylori*. The reduced products formed from the drug by the enzyme kill the bacterial cells.

See also: **Cell Architecture and Function:** Peroxisomes; **Metabolism Vitamins and Hormones:** Fatty Acid Structure and Synthesis; Tricarboxylic Acid Cycle.

Further Reading

- Medina Trullenque MM, Gómez-Moreno Calera C, and Frago Moreno S (eds.) (2008) *Flavins and Flavoproteins 2008*, Proceedings of the International Symposium on Flavins and Flavoproteins 2008, 628pp. Zaragoza: Prensas Universitarias de Zaragoza.
- Massey V (1994) Activation of molecular oxygen by flavins and flavoproteins. *Journal of Biological Chemistry* 269: 22459–22462.
- Massey V and Ghisla S (1974) Role of charge–transfer interactions in flavoprotein catalysis. *Annals of the New York Academy of Sciences* 227: 446–465.
- Müller F (ed.) (1991) *Chemistry and Biochemistry of Flavoenzymes*, vols. I–III. Boca Raton, FL: CRC Press.
- Palfey BA and Massey V (1998) Flavin-dependent enzymes. In: Sinnott M (ed.) *Comprehensive Biological Catalysis. Radical Reactions and Oxidation/Reduction*, vol. III, pp. 83–154. London: Academic Press.

Flippases

C J Waechter, University of Kentucky College of Medicine, Lexington, KY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Dolichyl monophosphate An eukaryotic glycosyl carrier lipid containing 15–20 isoprene units in which the α -isoprene unit is saturated.

Flippase A general term for a class of membrane proteins proposed to facilitate the transbilayer movement of the polar headgroup of a phospholipid, glycosphingolipid, or a polyisoprenyl-P(-P)-linked glycosyl intermediate.

Scramblase Bidirectional, adenosine triphosphate (ATP)-independent transporters that randomly redistribute polar

lipids formed *de novo* in the endoplasmic reticulum. This class also facilitates the nonspecific, bidirectional movement of phospholipids in the plasma membrane.

Transverse diffusion The flip-flopping or transbilayer movement of a polar membrane lipid from one leaflet to the opposite monolayer of a cellular membrane.

Undecaprenyl phosphate A fully unsaturated bacterial glycosyl carrier lipid containing 11 isoprene units.

Introduction

During the bioassembly of N-linked oligosaccharides, C- and O-linked mannosyl units in proteins, glycosylphosphatidylinositol (GPI) anchors, glycosphingolipids, and newly formed phospholipid bilayers in eukaryotic cells, charged hydrophilic headgroups must diffuse transversely through the hydrophobic core of the endoplasmic reticulum (ER) and an early Golgi compartment. Similarly, in prokaryotes, phospholipids and undecaprenyl pyrophosphate-linked glycosyl building blocks must be translocated from the inner leaflet to the outer monolayer during the expansion of the cytoplasmic membrane and the biosynthesis of cell-wall components. Biochemical and biophysical model systems indicate that the unassisted transbilayer movement of polar lipids with charged headgroups is highly unfavorable thermodynamically and much too slow to allow cells to synthesize the essential glycoconjugates in a physiologically requisite time frame. Thus, it has been proposed that the flip-flopping of the amphipathic precursors/intermediates is mediated by a class of membrane proteins referred to as flippases. The term flippase was originally suggested by Mark Bretscher during the 1970s.

Three types of the transbilayer movement of lipids with the generalized structure of glycerophospholipids and of the glycolipid intermediate, mannosylphosphoryldolichol (Man-P-Dol), mediated by flippases are illustrated in [Figures 1](#) and [2](#). [Table 1](#) summarizes several examples of polar lipids whose transverse diffusion is believed to be facilitated by flippases in prokaryotic and eukaryotic cells. Despite the critical function of flippases in membrane biology, there are still very large gaps in the information about their structures, the genes encoding this vital class of proteins, and the mechanism(s) by which they shield and permit the passage of the polar moieties through the hydrophobic core of the ER and other cellular membranes.

during active membrane biogenesis, ~50% of the newly formed phospholipids must be translocated to the lumenal monolayer. The landmark study of Kornberg and McConnell proved that the unassisted movement of spin-labeled phosphatidylcholine (PC) molecules between the two leaflets of pure lipid vesicles did not occur fast enough to maintain bilayer expansion at physiological rates. Further emphasizing the essential role of flippases, the classic experiments of James Rothman and Eugene Kennedy demonstrated that movement of phosphatidylethanolamine molecules, newly synthesized on the cytoplasmic leaflet of a *Bacillus megaterium* cell, to the extracellular monolayer was remarkably faster than the rate of spontaneous diffusion in artificial lipid bilayers. Primarily due to severe technical difficulties inherent in studying the phospholipid flippase(s), there is still very little known about the structure and mechanism of this eminently important class of membrane proteins. Bishop and Bell proposed the novel idea of following the transport of a water-soluble analog (diC₄PC) of PC by sealed microsomal vesicles to implicate one or more ER proteins in the transbilayer movement of the major membrane phospholipid. The PC flippase is depicted in [Figure 1\(a\)](#). The basic principle of this experimental approach is to design a water-soluble form of the polar lipid in which the hydrophobic groups are reduced in size to increase their solubility in aqueous solutions while retaining the essential structural features recognized by the proteins mediating their translocation. Although this experimental approach and alternate methods have been utilized by other researchers more recently to make progress toward these goals, the PC and other phospholipid flippases in biogenic membranes remain to be identified, isolated, and cloned. Neutral lipids such as cholesterol, coenzyme Q, and neutral glycerides may diffuse transversely in cellular membranes by a protein-unassisted process because they lack charged, hydrophilic groups.

Expansion of the ER Bilayer during Membrane Biogenesis

It is well established that the major site of phospholipid biosynthesis *de novo* in eukaryotes is on the cytoplasmic leaflet of the ER membrane. Clearly, to have uniform expansion of the bilayer

Aminophospholipid Flippases (P4-Adenosine Triphosphatase Family)/Translocases, Floppases, Scramblases, and Phospholipid Asymmetry

In most, if not all, cellular membranes phospholipids are distributed asymmetrically between the two leaflets, a property

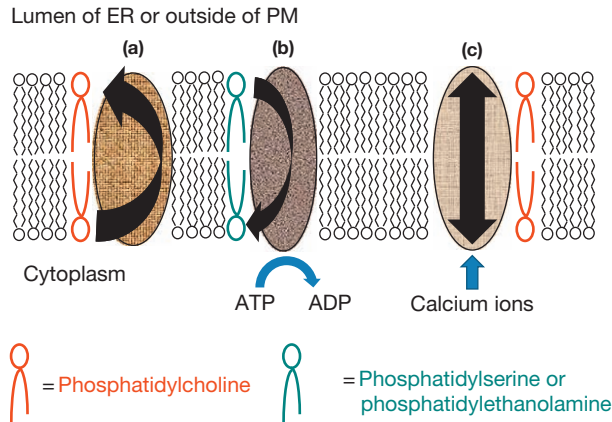


Figure 1 Illustration of three types of flippase-mediated transbilayer movement of membrane phospholipids in the ER or Golgi. The transbilayer movement of phosphatidylcholine during membrane biogenesis in the ER (a); the ATP-dependent flip-flopping of phosphatidylserine/phosphatidylethanolamine from the outer surface to the inner leaflet of the plasma membrane (PM, C, left) by the P_4 -ATPase family/aminophospholipid flippases (b); and the bidirectional, ATP-independent, Ca^{2+} -activated movement of a polar lipid mediated by scramblases is depicted in model (c). Although the translocation in examples in (a) and (b) is illustrated as being unidirectional, these processes appear to be bidirectional.

that is critical to membrane function. Thus, in addition to the flippases crucial to uniform expansion of the bilayer in the ER during membrane biogenesis, another set of flippases is required to maintain the proper distribution of phospholipids between the two leaflets of membranes in the various cellular compartments. The combined action of aminophospholipid flippases (P_4 -adenosine triphosphatase (ATPase) family members), two types of outwardly driven floppases (multidrug resistance, MDR, P -glycoprotein, other members of the MDR family, and members of the multidrug resistance-associated protein, MRP, subfamily) and possibly the adenosine triphosphate (ATP)-independent, Ca^{2+} -activated, bidirectional scramblases (Figure 1(c)) are believed to play a role in the asymmetric distribution of PC and other membrane phospholipids between the two leaflets of the plasma membrane (PM) and other cellular membranes.

The best characterized of the candidate flippases is the ATP/ Mg^{2+} -dependent aminophospholipid flippase (P_4 -ATPase family). The properties of this type of flippase have been studied extensively, and progress has been made in purifying the protein component(s). It has been found in the PM of erythrocytes and other mammalian cells, synaptosomes, chromaffin granules, and secretory vesicles. This flippase appears to play a direct role in maintaining the enrichment of phosphatidylserine (PS) and phosphatidylethanolamine (PE) in the inner leaflet of the PM by moving the two aminophospholipids from the outer to the inner leaflet of the PM against a concentration gradient at the expense of ATP. In Figure 1(b) (left) the lipid moiety would be diacylglycerol and the polar headgroup (O) would be either phosphorylserine or phosphorylethanolamine. The loss of this activity due to diminishing ATP pools or the loss of the protein(s) allows PE/PS to diffuse back to the outer monolayer where the surface alteration is recognized by the immune system as an early signal in apoptosis.

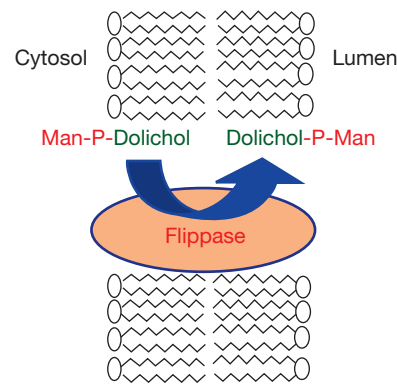


Figure 2 Illustration of the transbilayer movement of mannose-P-dolichol in the ER of eukaryotic cells mediated by a putative flippase composed of one or more membrane proteins. Man-P-dolichol is an intermediate mannosyl donor in protein N-glycosylation, C- and O-mannosylation, as well as in GPI anchor assembly. The transbilayer movement of Glc-P-dolichol and $Man_5GlcNAc_2$ -P-P-dolichol, two other glycolipid intermediates in the biosynthesis of asparagine-linked oligosaccharides, in the ER is also proposed to be facilitated by flippases.

However, inhibition of the flippase or loss of ATP is insufficient to cause a rapid loss of PM PS asymmetry. Rapid PS externalization requires the additional activation of a (Ca^{2+} -dependent) scramblase. The bidirectional, ATP-independent, Ca^{2+} -activated scramblase (Figure 1(c)) plays an important role in platelet activation, blood clotting, and apoptosis.

There is good evidence that floppases mediate the ATP-dependent translocation of cholesterol, PC, and sphingolipids from the cytoplasmic to the external leaflet of the PM.

Protein N-Glycosylation

N-linked oligosaccharide chains are initially synthesized as the dolichyl pyrophosphate (Dol-P-P)-linked precursor oligosaccharide by a two-stage process in the ER. In the first stage, three lipid intermediates, mannosylphosphoryldolichol (Man-P-Dol), glucosylphosphoryldolichol (Glc-P-Dol), and a heptasaccharide intermediate ($Man_5GlcNAc_2$ -P-P-Dol) are synthesized in the ER by enzymes with active sites facing the cytoplasm. In the next stage, these three glycolipid intermediates diffuse transversely to the luminal monolayer where four more mannosyl units are added to $Man_5GlcNAc_2$ -P-P-Dol with Man-P-Dol (Figure 2) functioning as the mannosyl donor. Finally, three glucosyl units are donated by Glc-P-Dol, completing the synthesis of the oligosaccharyl donor, $Glc_3Man_9GlcNAc_2$ -P-P-Dol. Based on evidence obtained from biophysical and enzymological topological studies, it is very likely that the transverse diffusion of the three intermediates requires ER proteins functioning as flippases to maintain lipid intermediate biosynthesis and protein N-glycosylation at physiologically relevant rates. Evidence has been obtained for the presence of ER proteins that mediate the transbilayer movement of short-chain, water-soluble analogs of Man-P-Dol and Glc-P-Dol by sealed vesicles from rat liver and brain by a variation of the experimental approach pioneered by Bishop and Bell. Preliminary studies using this transport technique indicate that the flippases are bidirectional and ATP

Table 1 Summary of specific flip-flopping events in prokaryotic and eukaryotic cells

Polar lipid	Membrane site	Transbilayer movement from
Aminophospholipids (PS/PE)	Plasma membrane	Outer leaflet to inner monolayer
PC and other glycerophospholipids, GlcN-PI, Man-P-Dol, Glc-P-Dol, and Man ₅ GlcNAc ₂ -P-P-Dol	ER	Cytoplasmic to luminal leaflet
Glucosylceramide	Early Golgi	Cytoplasmic to luminal leaflet
Glycerophospholipids and undecaprenyl-P(P)-linked cell-wall intermediates	Bacterial cytoplasmic membrane	Cytoplasmic leaflet to exterior monolayer (or outer membrane)

independent. It is plausible that this class of flippase acts by facilitated diffusion, and the transbilayer movement of the lipid intermediates is driven by mass action as they are consumed during Glc₃Man₉GlcNAc₂-P-P-Dol synthesis on the luminal surface. Hopefully, the *in vitro* assays will provide a means to isolating the ER proteins involved in this flip-flop-ping process. More recently, genetic and biochemical studies have implicated the *Rft1* protein in the transbilayer movement of Man₅GlcNAc₂-P-P-Dol in the yeast, *Saccharomyces cerevisiae*, although the precise function of this protein has yet to be determined. The structure of the yeast protein mediating the flipping of Man₅GlcNAc₂-P-P-Dol may provide clues to the identity of the gene encoding the corresponding protein in mammalian cells. The identification of the flippase proteins involved in the dolichol pathway and the corresponding genes should also ultimately be relevant to the diagnosis of related genetic defects in patients with as-yet uncharacterized forms of congenital disorders of glycosylation.

Relationship of a Man-P-Dol Flippase to the Biosynthesis of Other Mannose-Containing Glycoconjugates

Protein O- and C-Mannosylation

In yeast and mammalian cells, many proteins are modified by the addition of mannosyl units from Man-P-Dol to either serine/threonine in O-mannosidic linkage or to the indole ring of tryptophan residues in a C-mannosidic linkage. Since Man-P-Dol is believed to be formed only on the cytosolic face of the ER (Figure 2), and these modifications appear to occur on the luminal side of the ER, a Man-P-Dol flippase is essential for O- and C-mannosylation of proteins, as well as completing the synthesis of Glc₃Man₉GlcNAc₂-P-P-Dol and to maintain normal rates of protein N-glycosylation.

GPI Anchors

One class of proteins that is associated with membranes, despite not having any transmembrane domains, is attached by an amide linkage to GPI anchors. The biosynthesis of the GPI anchors theoretically requires at least two ER flippases. The best current evidence indicates that the early intermediate, GlcN-PI, is formed on the cytoplasmic face of the ER, and is subsequently translocated to the luminal monolayer where the GPI anchor is completed and the appropriate polypeptides are covalently attached to the ethanolamine moiety by a transpeptidation reaction. The completion of the anchor on the luminal

surface also requires Man-P-Dol as a mannosyl donor, and the transverse diffusion of the mannosyl intermediate requires a flippase that is discussed above (Figure 2).

Glucosylceramide

The enzyme that synthesizes glucosylceramide (Glc-Cer) from UDP-glucose and ceramide in an early Golgi compartment has an active site that is oriented toward the cytoplasm. As more complex cerebroside and gangliosides are elaborated on the interior of the Golgi compartments, it is necessary for the Glc-Cer formed on the cytosolic side to be translocated to the luminal leaflet where the glucosyl unit can be elongated by lumenally oriented glycosyltransferases. Thus, in this flippase-mediated event, the lipid moiety would be ceramide and the polar headgroup would be a β -linked glucosyl group, as illustrated in Figure 1(a). The biosynthesis of the complex glycosphingolipids also requires a set of Golgi proteins to mediate the transbilayer movement of the sugar nucleotides required for ganglioside synthesis. As there is experimental evidence that some fraction of the Glc-Cer formed on the cytoplasmic leaflet of an early Golgi compartment can be translocated directly to the inner surface of the PM, there may also be a PM flippase that mediates the movement of Glc-Cer to the outer leaflet.

Bacterial Cell Walls

In Gram-positive and Gram-negative bacteria, some of the building blocks for the assembly of peptidoglycan, lipopolysaccharide, and several other cell-wall complex glycoconjugates are synthesized on the inner face of the cytoplasmic membrane while attached to undecaprenyl pyrophosphate (Und-P-P). There is solid genetic evidence for the presence of membrane proteins that mediate the transbilayer movement of the bacterial lipid intermediates so the assembly of the cell-wall components can be completed on the exterior face of the cytoplasmic membrane, or in the case of Gram-negative bacteria, on the outer membrane. The Wzx family has been proposed to encode proteins involved in the translocation of Undec-P-P-linked intermediates in bacteria. There is recent evidence that the *wzx* gene encodes a protein involved in the transbilayer movement of a trisaccharide lipid intermediate in the biosynthesis of the enterobacterial common antigen in *Escherichia coli*. Although there is very limited information about their structures at this time, it is quite possible that the prokaryotic flippases resemble their eukaryotic counterparts. The intriguing question of why the polyisoprenoid glycosyl carrier lipids became longer (C55 for prokaryotes to C75–95

for eukaryotes) and acquired the saturated α -isoprene unit in dolichols during evolution remains to be explained.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Lipid Bilayer Structure; Glycosylphosphatidylinositol Anchors.

Further Reading

- Bishop WR and Bell RM (1985) Assembly of the endoplasmic reticulum phospholipid bilayer: The phosphatidylcholine transporter. *Cell* 42: 51–60.
- Bratton DL, Fadok VA, Richter DA, et al. (1997) Appearance of phosphatidylserine on apoptotic cells requires calcium-mediated nonspecific flip-flop and is enhanced by loss of the aminophospholipid translocase. *Journal of Biological Chemistry* 272: 26159–26165.
- Daleke DL (2003) Regulation of transbilayer plasma membrane phospholipid asymmetry. *Journal of Lipid Research* 44: 233–242.
- Daleke DL (2007) Phospholipid flippases. *Journal of Biological Chemistry* 282: 821–825.
- Dolis D, Moreau C, Zachowski A, and Devaux PF (1997) Aminophospholipid translocase and proteins involved in transmembrane phospholipid traffic. *Biophysical Chemistry* 68: 221–231.
- Helenius J, Ng DTW, Marolda CL, et al. (2002) Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature* 415: 447–450.
- Kornberg RD and McConnell HM (1971) Inside–outside transitions of phospholipids in vesicle membranes. *Biochemistry* 10: 1111–1120.
- McCloskey MA and Troy FA (1980) Paramagnetic isoprenoid carrier lipids: 2. Dispersion and dynamics in lipid membranes. *Biochemistry* 19: 2061–2066.
- Rick K, Barr K, Sankaran J, et al. (2003) Evidence that the *wzxE* gene of *E. coli* K-12 encodes a protein involved in the transbilayer movement of a trisaccharide-lipid intermediate in the assembly of enterobacterial common antigen. *Journal of Biological Chemistry* 278: 16534–16542.
- Rothman JE and Kennedy EP (1977) Rapid transmembrane movement of newly synthesized phospholipids during membrane assembly. *Proceedings of the National Academy of Sciences of the United States of America* 74: 1821–1825.
- Sanyal S, Frank CG, and Menon AK (2008) Distinct flippases translocate glycerophospholipids and oligosaccharide diphosphate dolichols across the endoplasmic reticulum. *Biochemistry* 47: 7937–7946.
- Sanyal S and Menon AK (2009) Flipping lipids: Why an' what's the reason for? *ACS Chemical Biology* 4(11): 895–909.
- Schenk B, Fernandez F, and Waechter CJ (2001) The ins(ide) and outs(ide) of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum. *Glycobiology* 11: 61R–70R.
- Sprong M, van der Sluijs P, and van Meer G (2001) How proteins move lipids and lipids move proteins. *Nature Reviews Molecular Cell Biology* 2: 504–513.

Focal Adhesions and Related Integrin Contacts

E Zamir, Max Planck Institute of Molecular Physiology, Dortmund, Germany

B Geiger, The Weizmann Institute of Science, Rehovot, Israel

© 2013 Elsevier Inc. All rights reserved.

Glossary

Adhesome The set of proteins and other molecules that are localized in adhesion sites (stably or transiently) and their direct interacting partners and regulators. The integrin adhesome is the subset of these components that is related to integrin-mediated cell–matrix adhesion sites.

Fibrillar adhesions Elongated or dot-like adhesions that are associated with fibronectin fibrils, mostly along the central areas of the cell.

Focal adhesions Flat elongated adhesions associated with prominent actin bundles (stress fibers); also known as ‘focal contacts’. They are usually located near the periphery of the cell.

Focal complexes Small dot-like adhesions located mainly at the edge of lamellipodium. They are induced by Rac, are associated with cell migration, and can develop into focal adhesions.

Podosomes Small cylindrical adhesions found in osteoclasts, macrophages, and various malignant cells.

Rac A Rho family guanosine triphosphatase (GTPase) that stimulates the extension of the lamellipodium and formation of focal complexes.

Rho A Rho family GTPase that stimulates stress-fiber and focal-adhesion formation.

Focal adhesions (also known as ‘focal contacts’) are integrin-dependent adhesions in which the cell membrane is attached at its external surface to the extracellular matrix (ECM), and at its internal aspect to the actin cytoskeleton. These transmembrane interactions are mediated via a complex network of cytoplasmic proteins, forming a submembrane plaque. Additional ECM–integrin–actin-mediated adhesion sites, termed ‘focal complexes’, fibrillar adhesions, three-dimensional (3D)-matrix adhesions, and podosomes, are related to focal adhesions, although the details of their molecular structures and functions differ. In addition to playing central roles in cell migration and morphogenesis, focal adhesions and related structures convey adhesion-triggered signals across the cell membrane, regulating cell proliferation, differentiation, and death.

Cell Adhesion

In multicellular organisms of the animal kingdom, cells adhere to their neighbors either directly, or via a complex meshwork of fibers known as the ECM. This adhesion is not a passive attachment but rather a complex process, mediating specific and dynamically regulated cell interactions as well as crucial signaling cross-talk between the cell and its environment. Indeed, cell adhesion plays a fundamental role in embryogenesis and organogenesis as well as in cellular processes such as cytoskeletal organization, cell motility, and regulation of the cell cycle and differentiation.

Moreover, cell adhesion is not a single, molecularly uniform, event but rather a highly diversified process at the cellular, molecular, and functional levels. For example, cell-to-cell adhesion occurs at tight junctions, gap junctions, chemical synapses, adherens junctions, desmosomes, and various carbohydrate-mediated contacts. Cell–ECM adhesions involve ECM proteins and glycosaminoglycans and their respective receptors, and occur at specialized cellular sites such as the

intermediate-filament-associated hemidesmosomes, and the widely occurring, actin-associated focal adhesion type, which is addressed here.

Focal Adhesions – A Structural Perspective

Specialized adhesions between cultured fibroblasts and the underlying substrate were revealed during the 1970s by interference–reflection microscopy and electron microscopy (Figure 1). These studies showed that the attached cells contain many specialized and defined regions (commonly measuring 0.25–2 by 2–10 μm) along the ventral plasma membrane, which are in very tight contact (10–15 nm gap) with the substrate. Moreover, these sites, which were termed ‘focal contacts’ or ‘focal adhesions’, appeared to be associated, within the cells, with the termini of actin microfilaments. These early studies revealed that focal adhesions are responsible for tight ECM contacts in which the extracellular substrate and the actin cytoskeleton are physically linked.

Following these early structural studies, the molecular composition of focal adhesions was addressed by immunofluorescence and immunoelectron microscopy, yielding a long list of diverse proteins, starting with α -actinin and vinculin. Biochemical examination of these proteins suggested that they are capable of establishing a complex network of protein–protein interactions with both cytoskeletal and signaling partners, and implicated focal adhesions in both mechanical and signaling processes. Studies carried out, in culture and *in vivo*, with different cell types exposed to a variety of conditions, established that focal adhesions are not the only form of ECM–integrin adhesions: included in this group are focal complexes, fibrillar adhesions, 3D-matrix adhesions, and podosomes (Figure 1).

Correlated fluorescence microscopy and cryo-electron tomography revealed the 3D features of focal adhesions,

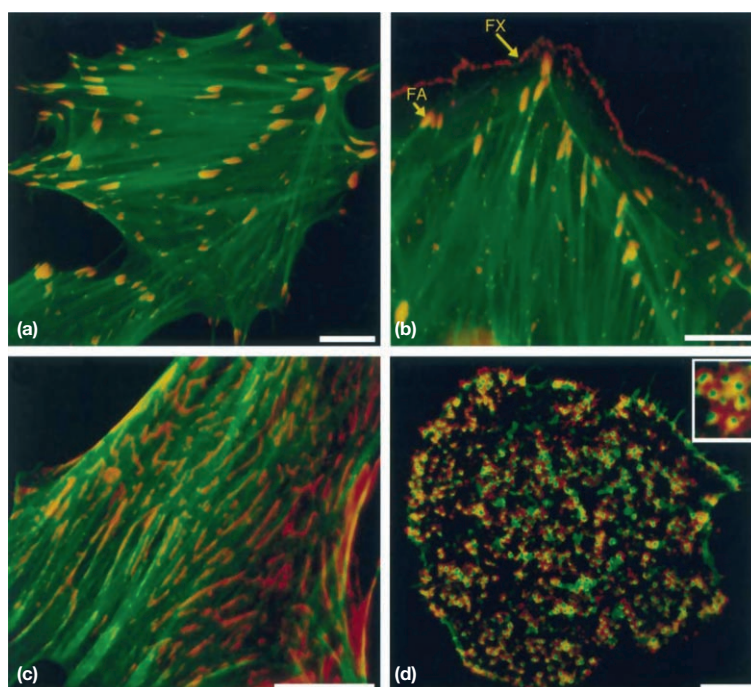


Figure 1 Focal adhesions and related adhesion sites. (a) Stationary human fibroblasts, stained for actin (green) and vinculin (red), having typical oval focal adhesions located at the termini of actin stress fibers. (b) Migrating endothelial cells (porcine aortic endothelial cells, PAEC) stained for actin (green) and phosphotyrosine (red), displaying both focal adhesions (FA) and focal complexes (FX). Focal complexes have a dot-like morphology and are located at the edge of the lamellipodium, where actin is organized as a dense network of highly branched fibers. (c) Human fibroblasts stained for actin (green) and for the fibronectin receptor $\alpha_5\beta_1$ integrin (red), which is enriched in fibrillar adhesions. (d) A mouse osteoclast, stained for actin (green) and paxillin (red), showing typical podosomes – doughnut-shaped adhesions with an actin core. The insert shows a cluster of podosomes at a higher magnification. Scale = 10 μm .

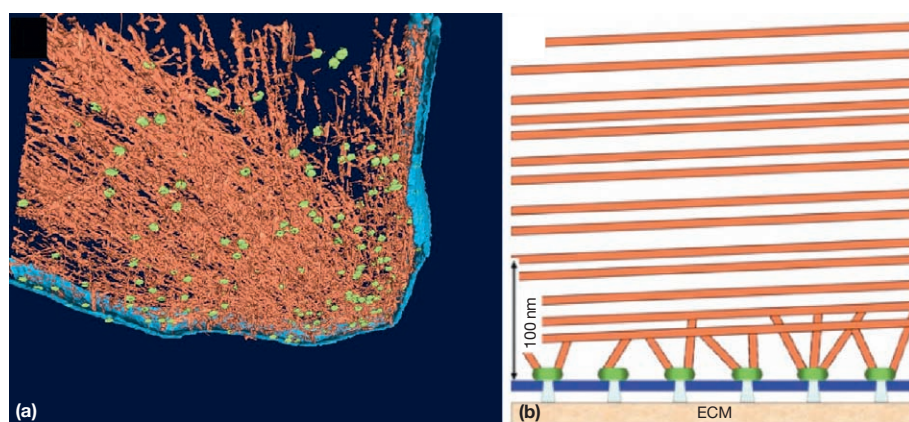


Figure 2 Cryo-electron tomography (cryoET) of focal adhesions carried out by correlated fluorescence-cryoET. (a) REF52 cells expressing a fluorescent paxillin were cultured on EM grids, fixed, frozen, and areas identified by fluorescence as focal adhesions were examined using a cryo-electron microscope. The 3D rendered image provides a bottom view of a focal adhesion from which the ventral membrane (blue) was removed, in order to visualize the membrane-actin interface. The image shows an array of parallel actin filaments (tan), as well as doughnut-shaped particles (light green) located at the interface between the cytoskeletal filaments and the plasma membrane. (b) A scheme showing a lateral view of a focal adhesion, including the extracellular matrix (ECM), transmembrane integrins, the plasma membrane, the particles, short interconnecting filaments, and the parallel array of actin filaments (stress fibers). Based on Patla I, Volberg T, Elad N, et al. (2010) Dissecting the molecular architecture of integrin adhesion sites by cryo-electron tomography. *Nature Cell Biology* 12: 909–915.

showing that the actin stress fibers are attached to the plasma membrane via short filaments linked to membrane-attached, doughnut-shaped particles of around 25 nm in diameter (Figure 2). The molecular nature of these particles

is still unknown; however, they appear to interact directly with actin filaments and the plasma membrane, and thereby to anchor the actin cytoskeleton to the ECM-bound membrane.

The Molecular Organization of Focal Adhesions

More than 180 distinct proteins and other molecules – collectively termed the ‘integrin adhesome’ – were reported to be localized in, or directly associated with, focal adhesions and related adhesion sites (Figure 3). Constituent components are largely segregated into the three structural domains within these sites: the transmembrane domain, the cytoskeletal domain, and the interconnecting submembrane plaque (Figure 2).

The Receptors and Their ECM Ligands

The membrane domains of ECM adhesions contain specific integrins, which are heterodimers of α - and β -subunits that bind the ECM via a large extracellular domain. Integrins span the membrane and contain a cytoplasmic region through which they interact with plaque proteins. The most common integrins found in focal adhesions and related ECM adhesions are the fibronectin receptor, $\alpha_5\beta_1$, and the vitronectin receptor, $\alpha_v\beta_3$. Additional, somewhat less characterized membrane-bound components of focal adhesions include potentially adhesive molecules such as syndecan-4 (SDC4, Figure 3) and the hyaluronan-binding protein layilin (LAYN), as well as the tyrosine phosphatase LAR (PTPRF) and the glycoprotein SHPS1 (SIRPA), a substrate of the tyrosine phosphatase SHP-2 (PTPN11).

The Cytoskeletal Domain

Focal adhesions are associated with the termini of actin bundles termed ‘stress fibers’. The polarity of the attached actin filaments

is apparently uniform, with their barbed ends pointing to the membrane. That area of the actin network is particularly enriched with actin-associated proteins such as the actin-bundling protein α -actinin (ACTN1, Figure 3). Actin is also associated with the other forms of integrin-mediated adhesions, although not in the form of stress fibers (Figure 1). In addition to the actin system, cell–matrix adhesion sites can also be associated with, and regulated by, the microtubule cytoskeleton (TUBA1B) and intermediate filaments (vimentin; VIM).

The Submembrane Plaque

The submembrane plaque is a multiprotein complex that interconnects actin to the membrane in focal adhesions and related adhesions (Figure 3). Some of the plaque components contain binding domains for both actin and integrins, including talin, α -actinin, tensin (TNS1, Figure 3), and filamin (FLNA), and may function as direct integrin–actin linkers. Additional integrin-associated molecules do not bind directly to actin, and may link the cytoskeleton to the membrane indirectly, via other plaque components. Some of them, such as focal adhesion kinase (FAK; PTK2), the LIM-domain protein DRAL (FHL2), and integrin-linked kinase (ILK), are signaling molecules (Figure 3).

Yet another group of focal adhesion-associated proteins includes actin-binding proteins that were not shown to interact directly with integrins, such as vinculin (VCL), VASP/Ena (EHAH), and ezrin–radixin–moesin (ERM) proteins (VIL2, RDX, MSN) (Figure 3). VCL plays a pivotal role as a central linker interacting with many plaque proteins – such as talin (TLN1), α -actinin, VASP/Ena, ponsin, and vinexin (SORBS3) – as well as with acidic phospholipids and actin. Finally, a very

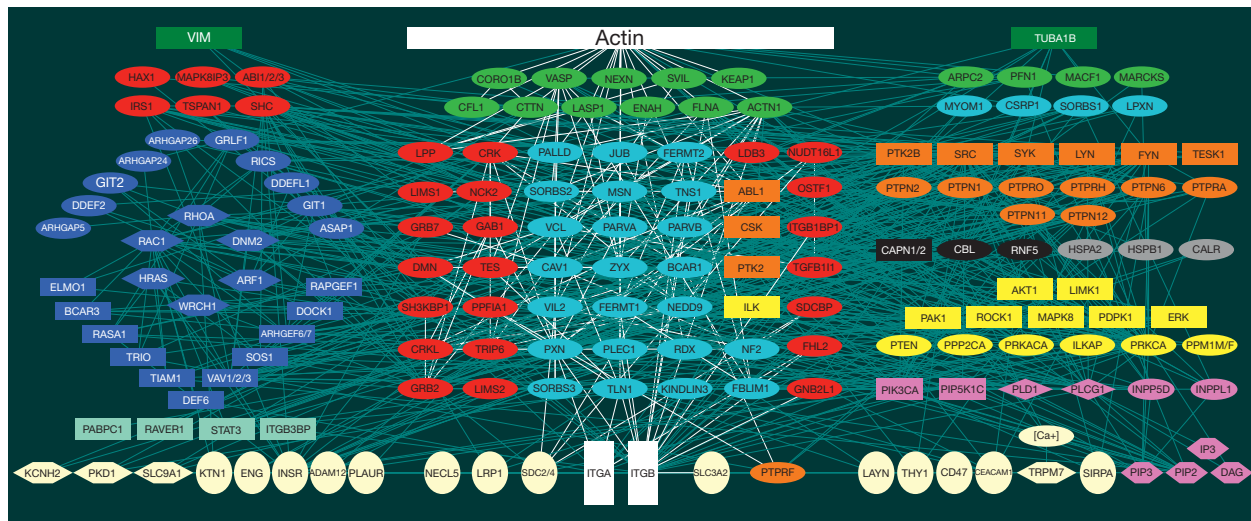


Figure 3 The integrin adhesome network. This network is divided here into structural (center) and regulatory/signaling (left and right) sub-networks. The structural part consists of actin, integrins (ITGA, ITGB) and other adhesion receptors (light brown ellipses), cytoskeletal proteins (cyan ellipses), actin modulators (light green), and multidomain adaptor proteins (red). The regulatory/signaling part includes small GTPase proteins and their activators and inhibitors (blue hexagons, rectangles, and ellipses, respectively), tyrosine kinases and phosphatases (orange rectangles and ellipses), serine/threonine kinases and phosphatases (yellow rectangles and ellipses), phospholipids (and their hydrolysis products), phosphatidylinositol kinases, phosphatidylinositol phosphatases and phospholipases (pink hexagons, rectangles, ellipses, and diamonds), chaperones and mediators of protein degradation (black and gray), and regulators of transcription and translation (mint green rectangles). Integrin adhesome data are largely based on Zaidel-Bar R and Geiger B (2010) The switchable integrin adhesome. *Journal of Cell Science* 123: 1385–1388, and <http://www.adhesome.org>.

large group of proteins consists of adapter proteins that apparently interact with actin-bound and integrin-bound components and link them to one another.

The components of cell–matrix adhesions have an unusually wide range of intrinsic activities. Beyond their protein–protein binding specificities, many of these proteins are enzymes, such as tyrosine kinases, serine/threonine kinases, tyrosine phosphatases, inositol 5′-phosphatase, modulators, and adapters of small guanosine triphosphatases (GTPases), and other enzymes such as PI 3-kinase (PIK3CA) and protease calpains (CAPN1/2) (Figure 3). It should be pointed out that the true molecular complexity of focal adhesions is probably greater than that depicted in Figure 3 because many of these components represent products of multigene families, and may undergo posttranslational modification and proteolytic cleavage.

A striking characteristic of many focal-adhesion components is that they are multidomain proteins that can interact with several distinct partner molecules. For example, vinculin, FAK, Src family kinases, and paxillin (PXN) can each bind to more than 10 different partners. Thus, the theoretical number of different combinations of molecular interactions that might be involved in linking integrins to actin is enormous. Importantly, most interactions between focal-adhesion components are not constitutive, but can be tightly regulated by local intra- and extracellular cues. Therefore, the pool of adhesome

components within a cell can give rise to a variety of adhesion sites with diverse structures, molecular compositions, and functions.

Diversity of Cell–Matrix Adhesion Sites

Adhesions with the ECM are formed by essentially all adherent animal cells, in culture and *in vivo*. Yet their morphology, size, molecular composition, and subcellular distribution can be quite heterogeneous. These adhesions, however, share two common features: they are mediated by integrins, and they interact with the actin cytoskeleton. Focal adhesions, the best-characterized form of these sites, are flat elongated structures, several square microns in area, and are usually located near the periphery of cells, associated with bundles of actin microfilaments (Figures 1 and 3). The development of focal adhesions is stimulated by the small GTPase Rho, and requires actomyosin contractility. Characteristic focal-adhesion plaque proteins include vinculin, talin, paxillin, and tyrosine-phosphorylated proteins. Focal adhesions are particularly prominent in cultured cells growing on solid surfaces; yet structures with similar molecular properties are also found *in vivo*.

Another group of matrix adhesions is the focal complexes, which are small, dot-like adhesions mainly found close to the edges of the lamellipodium (Figures 1 and 4). These sites can

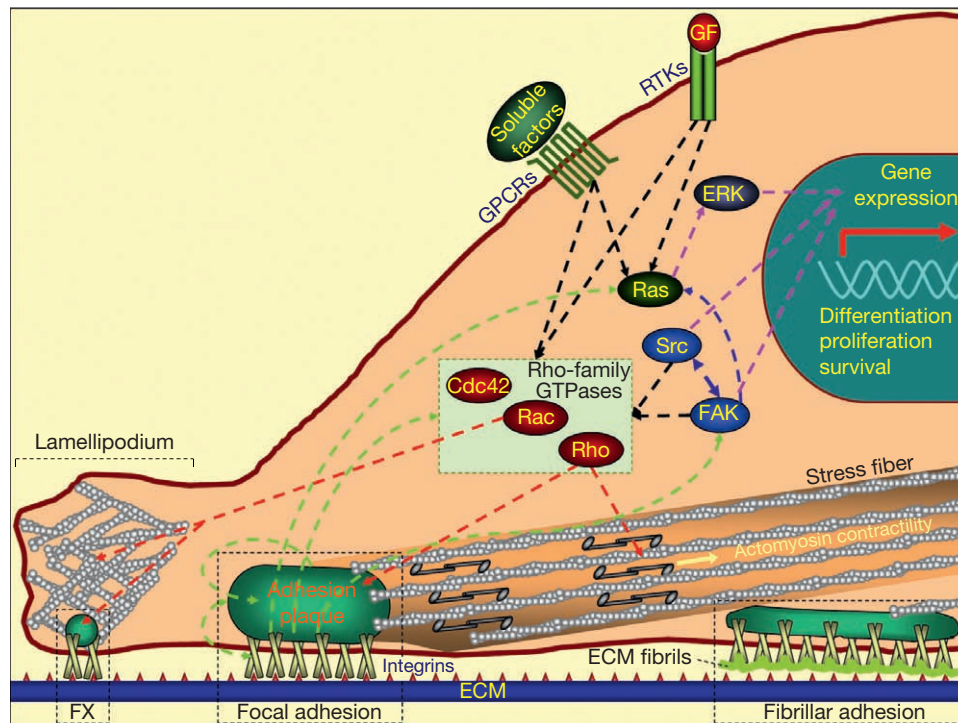


Figure 4 A simplified scheme highlighting major pathways of focal adhesion-mediated signaling. Rho family GTPases, including Rho, Rac, and Cdc42, are regulated by integrins, growth factors (GFs), receptor tyrosine kinases (RTKs), and G-protein-coupled receptors (GPCRs). Cdc42 induces the formation of filopodia (not shown) and activates Rac. Rac induces the formation of a meshed network of branched actin filaments, extension of lamellipodia, and formation of focal complexes (FX). Rho induces actomyosin contractility and organizes actin in contractile bundles called ‘stress fibers’. Rho activation drives the development of focal adhesions from focal complexes. Rho-induced contractility is also essential for the development of fibrillar adhesions, and their translocation toward the cell center. In adhesion sites, integrins locally activate the tyrosine kinases, FAK and Src, which then regulate the enzymatic activities and interactions of many other focal adhesion components. Integrins also affect gene expression, mainly through the activation of the Ras–ERK pathway, and thereby regulate differentiation, proliferation, and survival.

be associated with cell migration or serve as precursors of focal adhesions. Their formation is induced by the Rho family GTPase Rac. In the more central locations of many cell types, there are the fibrillar adhesions (also referred to as ECM contacts), elongated or dot-like structures associated with ECM fibrils (Figures 1 and 4). The typical components of fibrillar adhesions are extracellular fibronectin fibrils, the fibronectin receptor $\alpha_5\beta_1$ integrin, and the cytoplasmic protein tensin. Fibrillar adhesions emerge from focal adhesions in an actomyosin-dependent fashion. Another form of ECM adhesions is the podosomes, small (around 0.5 μm in diameter) cylindrical structures containing typical focal adhesion proteins such as vinculin and paxillin (Figure 1(d)). Podosomes are found in a variety of malignant cells and in some normal cells such as macrophages and osteoclasts. Characteristic and indispensable podosome proteins are gelsolin and dynamin, which localize to tubular invaginations of the plasma membrane typical for these structures. Finally, another type of adhesion is the 3D-matrix adhesion, formed with 3D ECM, which differ from focal adhesions in their overall morphology, fine molecular composition, and function.

Signaling

Focal adhesions and related structures are involved in adhesion-mediated signaling affecting various cellular processes such as proliferation, differentiation, anchorage-dependent survival, spreading, and migration. The main signaling paths include the stimulation of local tyrosine phosphorylation driven by FAK and Src, activation of small GTPases of the Rho and Ras families, and the extracellular signal-regulated kinase (ERK) pathway (Figure 4). It is proposed that tethering of FAK to clustered integrins leads to its autophosphorylation, recruitment of Src, and phosphorylation of their downstream targets. The presence of a variety of kinases and their substrates in the submembrane plaque suggests that additional signaling processes are activated by matrix adhesions; yet definitive information about these signaling processes is still scarce. Focal adhesions themselves are regulated by several signaling systems. In addition to Rho and Rac, excessive Src-mediated phosphorylation and stimulation of appropriate cells by growth factors (e.g., epidermal growth factor (EGF) or neuregulin), phosphatases, and proteases have a profound effect on the organization of cell–matrix adhesions.

Focal Adhesions and Mechanosensitivity

Mechanical forces play a key role in the development and fates of focal adhesions and focal complexes. The transition from focal complexes to focal adhesions and the growth of focal adhesions require local, actomyosin-generated tension, and are dramatically suppressed following treatment with inhibitors of Rho kinase or of myosin light-chain kinase (Figure 5). The mechanisms underlying adhesion-mediated mechanosensing are still poorly understood; yet it is likely that forces applied to the adhesomal network affect the conformation of some of the scaffolding or signaling constituents, thereby blocking or enhancing signaling events, thereby altering the

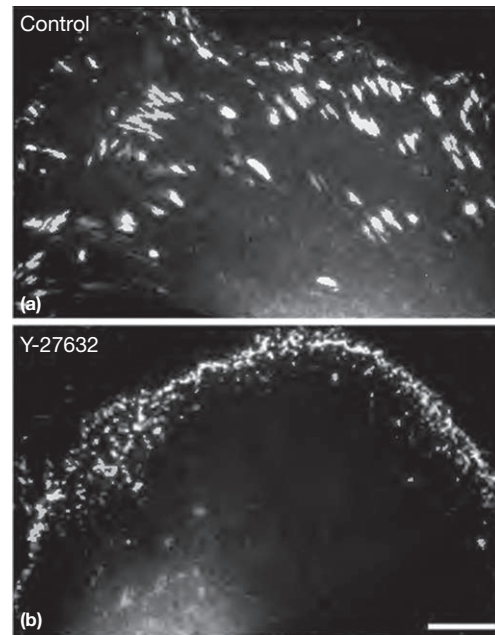


Figure 5 The key role of actomyosin contractility in the organization of adhesion sites. Rat embryonic fibroblasts (REF52), stably expressing β_3 -integrin-GFP, were either (a) untreated (control) or (b) treated with the Rho-kinase inhibitor Y-27632, prior to fixation. Inhibition of actomyosin contractility by Y-27632 causes disassembly of focal adhesions, and accumulation of focal complexes. Scale = 10 μm .

scaffold's organization. Recent data indicated, for example, that force stretches some of the plaque proteins (such as Crk-associated substrate (CAS) and talin) and thereby regulates their interactions. Focal complexes do not appear to be stimulated by actomyosin contractility, and even become abundant upon inhibition of contraction (Figure 5). Similarly, the maintenance of fibrillar adhesions is not affected by the inhibition of contractility; yet their emergence from peripheral focal adhesions requires actomyosin-driven tension.

See also: Cell Architecture and Function: Actin Assembly/Disassembly; Actin-Capping and -Severing Proteins; Actin-Related Proteins; Rho GTPases and Actin Cytoskeleton Dynamics; Lipids Carbohydrates Membranes and Membrane Proteins: Cell–Matrix Interactions; Signaling: Integrin Signaling; Src Family of Protein Tyrosine Kinases.

Further Reading

- Alberts B, Johnson A, Lewis J, et al. (2002) *Molecular Biology of the Cell*. New York: Garland Publishing.
- Beckerle MC (2001) *Frontiers in Molecular Biology, Vol. 39. Cell Adhesion*. New York: Oxford University Press.
- Geiger B, Bershadsky A, Pankov R, and Yamada KM (2001) Extracellular matrix–cytoskeleton cross-talk. *Nature Review Molecular Cell Biology* 2: 793–805.
- Kreis T and Vale R (eds.) (1999) *Guidebook to the Extracellular Matrix, Anchor, and Adhesion Proteins*. New York: Oxford University Press.

- Patla I, Volberg T, Elad N, et al. (2010) Dissecting the molecular architecture of integrin adhesion sites by cryo-electron tomography. *Nature Cell Biology* 12: 909–915.
- Zaidel-Bar R and Geiger B (2010) The switchable integrin adhesome. *Journal of Cell Science* 123: 1385–1388.
- Zaidel-Bar R, Itzkovitz S, Ma'ayan A, Iyengar R, and Geiger B (2007) Functional atlas of the integrin adhesome. *Nature Cell Biology* 9: 858–867.
- Zamir E and Geiger B (2001) Molecular complexity and dynamics of cell–matrix adhesions. *Journal of Cell Science* 114: 3583–3590.
- Zamir E, Katz M, Posen Y, et al. (2000) Dynamics and segregation of cell–matrix adhesions in cultured fibroblasts. *Nature Cell Biology* 2: 191–196.

Relevant Websites

- <http://www.adhesome.org> – ADHESOMEFAnetwork.
- <http://www.weizmann.ac.il> – Benny Geiger's Lab.
- <http://www.cellmigration.org> – Cell Migration Gateway.
- <http://www.cell-adhesion.com> – Eli Zamir's Lab: Systems Biology of Cell–Matrix Adhesion.
- <http://www.cytoskeletonforum.org> – European Cytoskeleton Forum.
- <http://www.mechanicalbiology.org> – Nanomedicine Center for Mechanobiology: Directing the Human Response.

Folate and Vitamin B₁₂ Function

B Shane, University of California, Berkeley, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Cobalamin Another name for vitamin B₁₂.

Coenzyme B₁₂ Also called 5'-deoxyadenosylcobalamin.

H₄PteGlu_n A polyglutamate form of tetrahydrofolate (tetrahydropteroylglutamate).

Megaloblastic anemia A large cell anemia.

Neural tube defects Birth defects such as spina bifida and anencephaly due to incomplete closure of the neural tube during development.

Pteroylmonoglutamate Another name for folic acid.

Folate Structure and Chemistry

Folate is a generic term for a family of chemically and functionally related compounds based on the folic acid (pteroylmonoglutamate) structure (Figure 1). Folic acid is an oxidized, synthetic form of the vitamin and is reduced within the cell to the metabolically active 5,6,7,8-tetrahydrofolate form. Folates in tissues act as donors and acceptors of one-carbon units in metabolic reactions known as 'one-carbon metabolism'. These one-carbon units can be at the oxidation level of methanol (5-methyl-tetrahydrofolate), formaldehyde (5,10-methylene-tetrahydrofolate), or formate (5- or 10-formyl-tetrahydrofolate or 5,10-methenyl-tetrahydrofolate).

Practically all tissue folates are polyglutamate forms in which the glutamate tail is extended via an unusual peptide bond via the γ -carboxyl of glutamate. Glutamate chain lengths can vary from about 4 to 10 in human tissues. Metabolism of folates to polyglutamate forms is required for their biological activity as the polyglutamate forms are much more effective substrates for folate-dependent enzymes than are the monoglutamate derivatives, which are the transport forms of the vitamin. Conversion of folates to polyglutamates of chain length greater than three or more is also required for effective retention of folate by tissues.

Absorption and Transport

Food folates usually occur in a reduced, polyglutamyl form. Before absorption, they are cleaved to their monoglutamyl forms by a brush border glutamylhydrolase (glutamate carboxypeptidase II). Folates are absorbed in the proximal small intestine via the proton-coupled folate transporter (PCFT), an apically expressed transmembrane, pH-sensitive carrier that transports both oxidized and reduced folates with similar efficacy under the slightly acidic conditions of the gut wall. Folate monoglutamate in plasma is transported into peripheral tissues via the reduced folate carrier (RFC). This transmembrane carrier is specific for reduced folates with affinities in the low micromolar range while the affinity for folic acid is about 100-fold lower. The liver carrier transports both folic acid and reduced folates and is thought to be PCFT. Folate monoglutamate transported into cells is metabolized to polyglutamate forms by the enzyme folylpolyglutamate synthetase, which

catalyzes the adenosine triphosphate (ATP)-dependent sequential addition of glutamate moieties. Mitochondrial and cytosolic isozymes of folylpolyglutamate synthetase are encoded by a single nuclear gene.

A second family of closely related folate transporters known as 'folate-binding proteins' or 'folate receptors' (α , β , and γ) are expressed in a more limited range of epithelial tissues. The affinities of reduced folate monoglutamates for these receptors vary by receptor but are very high, usually in the low nanomolar range. This transporter is responsible for reabsorption of folate in the kidney by a receptor-mediated endocytotic process and is believed to play a similar role in folate transport in other tissues.

Folate Biochemistry and Functions

Folate coenzymes are involved in three major interrelated metabolic cycles in the cytosol of cells. These cycles are required for the synthesis of thymidylate and purines, precursors for DNA and RNA synthesis, and for the synthesis of methionine from homocysteine and the interconversion of serine and glycine (Figure 2).

5,10-Methylene-tetrahydrofolate plays a central role in these cycles as it can be used directly for thymidylate synthesis or it can be reduced to 5-methyl-tetrahydrofolate in the methionine synthesis cycle or can be oxidized to 10-formyl-tetrahydrofolate to be used in purine synthesis. Although the enzymes involved in these synthetic cycles are normally located in the cytosol, during the S phase of the cell cycle the enzymes involved in thymidylate synthesis translocate to the nucleus. Mammalian cells also contain a large, separate mitochondrial folate pool, which is involved in serine and choline metabolism and in the provision of one-carbon precursors for cytosolic one-carbon metabolism.

Serine-Glycine Interconversion

Serine is the major provider of one-carbon units for folate-dependent one-carbon metabolism. It donates its β -carbon to tetrahydrofolate to generate glycine and 5,10-methylene-tetrahydrofolate in a reaction catalyzed by the reversible pyridoxal 5'-phosphate (PLP)-dependent enzyme serine hydroxymethyltransferase (SHMT). Mammalian tissues contain

two distinct isozymes of SHMT, one cytosolic and one mitochondrial, encoded by different genes. The mitochondrial isozyme may be responsible for the generation of the majority of one-carbon units required for cytosolic one-carbon metabolism. 5,10-Methylene-tetrahydrofolate generated via the mitochondrial SHMT reaction can be oxidized to 10-formyl-tetrahydrofolate and then hydrolyzed to formate. The formate produced exits the mitochondria and the one carbon is reincorporated into the cytosolic one-carbon pool as 10-formyl-tetrahydrofolate. 10-Formyl-tetrahydrofolate can be reduced to 5,10-methylene-tetrahydrofolate in the cytosol via the action of a reversible trifunctional C1 synthase enzyme.

Nucleotide Synthesis

Folate is required for the synthesis of thymidylate, a nucleotide required specifically for DNA synthesis and repair. Thymidylate

synthase catalyzes the transfer of the one-carbon group from 5,10-methylene-tetrahydrofolate to the 5'-position of deoxyuridine monophosphate (dUMP) and its reduction to a methyl group to generate deoxythymidine monophosphate (dTMP) (Figure 2). The folate molecule also provides the reducing component in this reaction and the tetrahydrofolate is oxidized to dihydrofolate. The dihydrofolate generated has to be reduced back to tetrahydrofolate before it can be reutilized in one-carbon metabolism in a reaction catalyzed by dihydrofolate reductase (DHFR). Thymidylate synthase (TS) activity is primarily expressed in replicating cells during the S phase of the cell cycle and is highest in fast-growing cells. Consequently, drugs targeted to DHFR, such as methotrexate, have proven to be effective chemotherapeutic agents as they are selective inhibitors of rapidly growing cell. Some of the cytosolic folate enzymes involved in provision of thymidylate are sumoylated and translocate to the nucleus during the S phase of the cell cycle (3). These include TS, DHFR, and cytosolic SHMT.

Cytosolic folate coenzymes are also used in two steps of *de novo* purine biosynthesis (Figure 2). The C-8 and C-2 positions of the purine ring are derived from 10-formyl-tetrahydrofolate in reactions catalyzed by glycineamide ribonucleotide (GAR) transformylase and 5-amino-4-imidazolecarboxamide ribonucleotide (AICAR) transformylase. The enzymes involved in *de novo* purine are multifunctional and form a complex in the cytosol called the 'purinosome'. This complex remains in the cytosol during the S phase of the cell cycle.

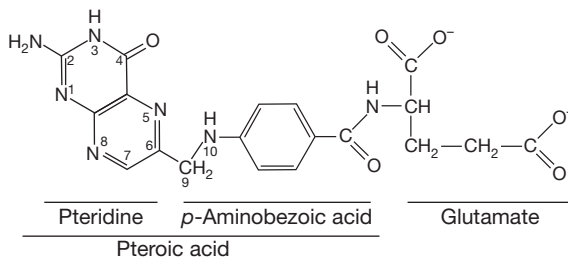


Figure 1 Folic acid (pteroylmonoglutamate, PteGlu).

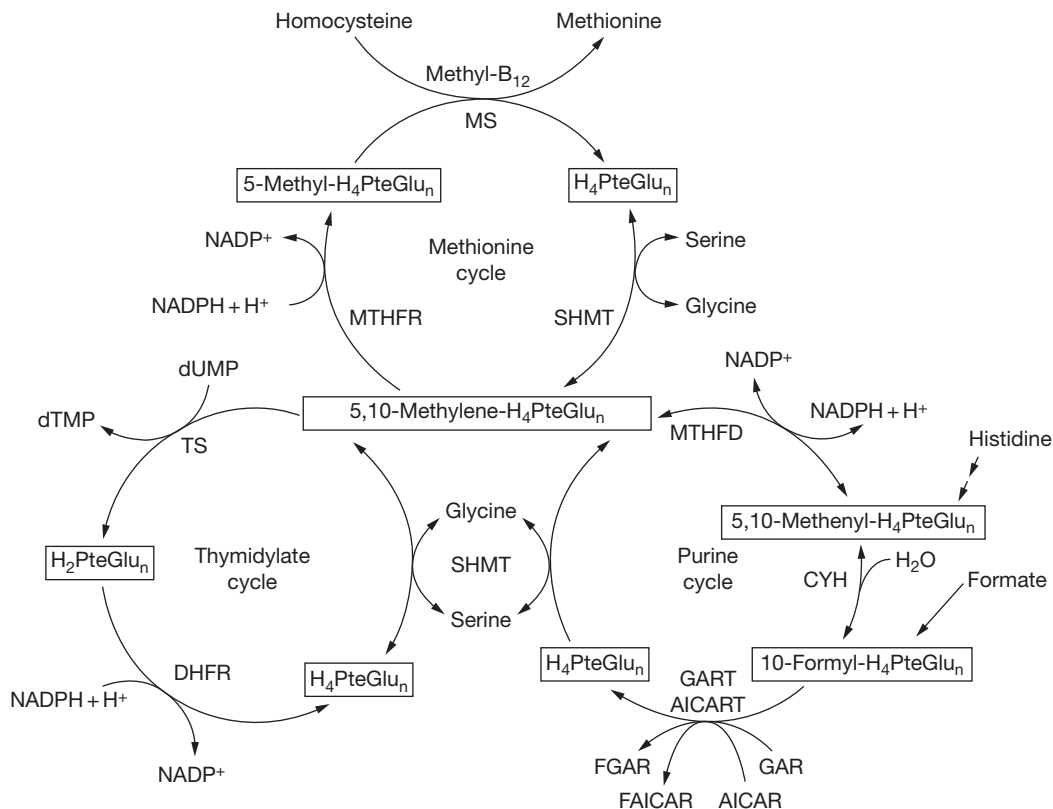


Figure 2 Folate-dependent cytosolic one-carbon metabolism.

Methionine Synthesis

The methylation of homocysteine to produce methionine uses 5-methyl-tetrahydrofolate as the methyl donor in a reaction catalyzed by methionine synthase, one of only two vitamin B₁₂-dependent enzymes in mammals (Figure 2). 5-Methyl-tetrahydrofolate is generated from 5,10-methylene-tetrahydrofolate in a reaction catalyzed by the flavoprotein methylenetetrahydrofolate reductase (MTHFR). Methionine can be metabolized to S-adenosylmethionine, which acts as the methyl donor in many reactions, including the methylation of DNA, histones and other proteins, neurotransmitters and phospholipid synthesis, and the synthesis of creatine. These methylation reactions play important roles in development, gene expression, and genomic stability. S-Adenosylhomocysteine, the product of methylation reactions, is a potent inhibitor of many methyltransferases and is catabolized by hydrolysis to adenosine and homocysteine. In liver and kidney, homocysteine can be further metabolized to cysteine via the transsulfuration pathway utilizing the PLP-dependent enzymes cystathionine- β -synthase (CBS) and cystathioninase. In most other tissues, homocysteine is exported to the circulation or is reconverted back to methionine via the folate-dependent methionine synthase reaction.

Homocysteine can also be remethylated to methionine by cytosolic betaine homocysteine methyltransferase (BHMT), a folate-independent enzyme that is expressed in high levels in liver and in some species, including humans, in kidney. Betaine (trimethylglycine) is derived by mitochondrial oxidation of choline. Dimethylglycine, the product of the BHMT reaction, is oxidized in the mitochondria to sarcosine (methylglycine) and then to glycine by the folate-dependent

flavoproteins dimethylglycine dehydrogenase and sarcosine dehydrogenase. Each oxidation results in the formation of a molecule of 5,10-methylene-tetrahydrofolate from tetrahydrofolate.

Methionine is a dietary essential amino acid and it is often the most limiting amino acid in human diets. Methylation reactions account for a large proportion of the methyl group intake in humans and the methionine synthase and BHMT reactions allows salvage of its backbone after its use for methylation.

Chemistry of Vitamin B₁₂

Vitamin B₁₂ (cobalamin) consists of a central cobalt atom surrounded by a heme-like planar corrin ring structure (Figure 3), with the four pyrrole nitrogens coordinated to the cobalt. It contains a phosphoribo-5,6-dimethylbenzimidazolyl side group, with one of the nitrogens linked to the cobalt by coordination at the bottom position. The lower axial ligand is usually replaced by an active site histidine residue when bound to protein. The upper axial position can be occupied by a number of different ligands, including methyl, hydroxyl, and 5'-deoxyadenosyl groups, and by cyanide in what is commonly known as vitamin B₁₂ (cyanocobalamin). In cyanocobalamin and the naturally occurring hydroxy cobalamin forms, the cobalt atom is trivalent Co³⁺ (cob(III)alamin), the most oxidized form. The coenzyme forms of the vitamin are the unliganded, fully reduced Co¹⁺ derivative or cob(I)alamin; 5'-deoxyadenosylcob(III)alamin, also known as coenzyme B₁₂; and methylcob(III)alamin.

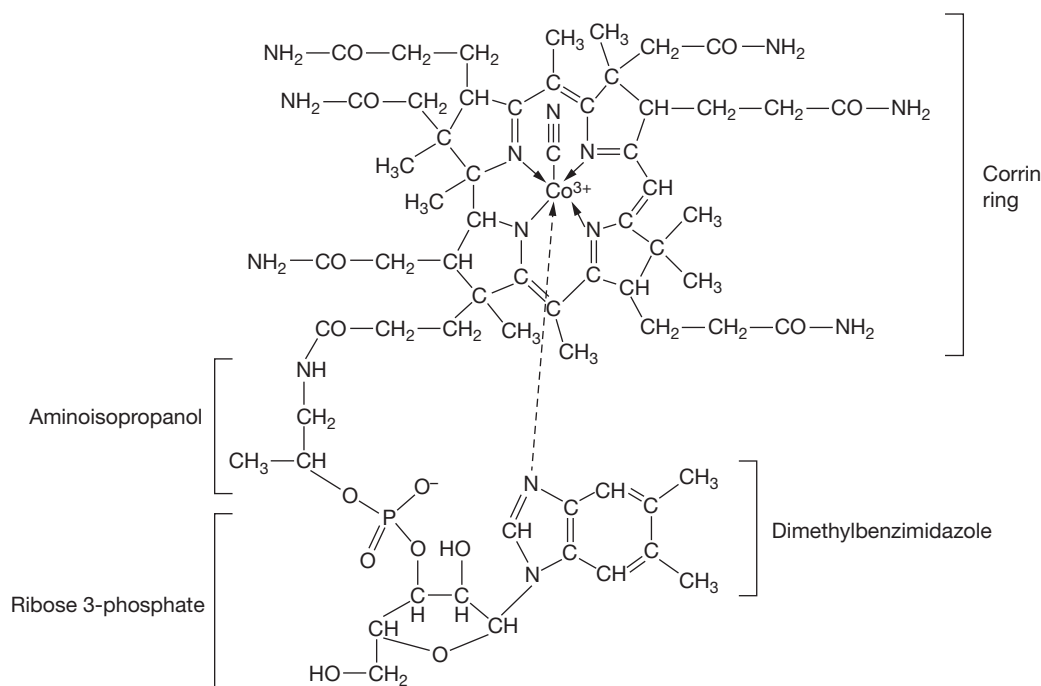


Figure 3 Structure of vitamin B₁₂ (cyanocobalamin).

Absorption and Transport of Vitamin B₁₂

Vitamin B₁₂ in food is bound to protein and is released in the stomach by the acid environment and by proteolysis of binders by pepsin. It then binds to other proteins known as R binders and these complexes have to be hydrolyzed by proteases in the small intestine before the B₁₂ is available for absorption. The parietal cells in the stomach secrete a 50-kDa glycoprotein called 'intrinsic factor (IF)' that can bind vitamin B₁₂ after its release from the R binders. Vitamin B₁₂ is absorbed via receptors located at the distal ileum at the end of the small intestine. In the presence of Ca²⁺, the IF-B₁₂ complex binds to the IF receptor (cubilin), and the complex is internalized by a receptor-mediated endocytotic process. The IF is degraded and vitamin B₁₂ is released into the cytosol and into the circulation, and binds to a 38-kDa protein called 'transcobalamin II (TC-II)'. Vitamin B₁₂ can also be absorbed by a diffusion-like process, but this is very inefficient (less than 1%). TC-II-B₁₂ is transported into tissues by receptor-mediated endocytosis, in this case via the TC-II receptor.

Vitamin B₁₂ Biochemistry and Functions

Mammals use vitamin B₁₂ as a cofactor for two enzymes, cytosolic methionine synthase (described above), which utilizes methylcob(III)alamin as the cofactor, and mitochondrial methylmalonyl CoA mutase, which uses 5'-deoxyadenosylcob(III)alamin as its cofactor. Vitamin B₁₂ is released into the cytosol of cells as hydroxycob(III)alamin and is reduced to cob(I)alamin. Cob(I)alamin is then methylated to methylcob(III)alamin after binding to methionine synthase. Alternatively, vitamin B₁₂ can be transported into the mitochondria and reduced, and the 5'-deoxyadenosyl ligand added from ATP in a reaction catalyzed by a deoxyadenosyltransferase. Rare human genetic defects in many of these steps in vitamin B₁₂ metabolism have been described.

Methionine Synthase

Methylcobalamin is a cofactor for the previously described folate-dependent methionine synthase involved in homocysteine remethylation (see [Figure 2](#)). Methionine synthases are large, monomeric Zn metalloproteins and consist of three domains: a catalytic domain containing the binding sites for 5-methyl-tetrahydrofolate and homocysteine, a B₁₂ domain in which the B₁₂-cofactor binds, and an accessory protein domain. The cob(I)alamin cofactor is methylated by 5-methyl-tetrahydrofolate, generating enzyme-bound methylcob(III)alamin and releasing tetrahydrofolate, and then methylcob(III)alamin transfers its methyl group to homocysteine to generate methionine and regenerate the enzyme-bound cob(I)alamin cofactor. The cofactor is occasionally oxidized to the nonfunctional cob(II)alamin form during catalysis. The enzyme is reactivated by methionine synthase reductase, an accessory protein that catalyzes the adenosylmethionine and NADPH-dependent reductive methylation of enzyme-bound cob(II)alamin to methylcob(III)alamin. The gene (*MTRR*) for a single human methionine synthase reductase protein that

contains binding sites for nicotinamide adenine dinucleotide phosphate reduced (NADPH), flavin adenine dinucleotide (FAD), and flavin mononucleotide (FMN) has been cloned. It belongs to the P450 reductase protein family. Human mutations that result in loss of methionine synthase activity and early onset megaloblastic anemia fall into two complementation groups. The *cblG* complementation group results from mutations in the structural gene for methionine synthase (*MTR*) whereas mutations in methionine synthase reductase (*MTRR*) are responsible for the *cblE* complementation group.

Methylmalonyl CoA Mutase

Mitochondrial β -oxidation of dietary odd-chain fatty acids produces propionyl CoA in addition to acetyl CoA. Propionyl CoA is converted to D-methylmalonyl CoA in a reaction catalyzed by the biotin-dependent propionyl CoA carboxylase. Propionyl CoA and methylmalonyl CoA can also arise during catabolism of isoleucine, valine, methionine, and threonine. A racemase converts D-methylmalonyl CoA to L-methylmalonyl CoA, and the 5'-deoxyadenosylcobalamin-dependent methylmalonyl CoA mutase catalyzes the conversion of L-methylmalonyl CoA to succinyl CoA. All these reactions occur in the mitochondria, and the succinyl CoA formed has several potential fates, including entry into the citric acid cycle and heme biosynthesis.

Folate Deficiency

Folate deficiency is usually due to a dietary insufficiency although it can arise from other causes such as malabsorption syndromes. The classical symptom of folate insufficiency is megaloblastic anemia, a condition reflecting deranged DNA synthesis in the erythropoietic cells. Blood cells are enlarged and often multinucleated. Megaloblastic cells contain close to twice the normal DNA content and the DNA is partially fragmented. Many cells are arrested in the G2 phase just prior to mitosis. The defect in DNA synthesis in folate deficiency has been ascribed to defective thymidylate synthesis under these conditions with a resulting increase in uracil misincorporation into DNA. Removal of uracil by the repair enzyme uracil DNA glycosylase, and a decreased repair of the gaps produced by this enzyme, would lead to an increase in double-stranded DNA breaks under these conditions.

Neural Tube Defects

In the early 1990s, double-blind, randomized trials confirmed that supplemental folic acid given prior to conception and during early pregnancy significantly reduced by over 70% the occurrence and recurrence of neural tube defects (NTDs), the most common birth defects in humans. It is thought that some women and/or their offspring have a higher requirement for folate due to genetic reasons. The current fortification of the food supply with folic acid in the US and over 50 other countries arose because those individuals at risk for having an NTD child cannot be identified at the moment. NTD rates have fallen dramatically after fortification with the largest decreases in populations with the highest pre-fortification rates.

Vitamin B₁₂ Deficiency

Vitamin B₁₂ deficiency is rarely caused by a dietary insufficiency and most commonly arises from a defect in vitamin B₁₂ absorption. The classical manifestations are pernicious anemia, which is a megaloblastic anemia identical to that observed with folate deficiency, and a severe and often irreversible neurological disease called subacute combined degeneration of the spinal cord, which is characterized by demyelination and peripheral neuropathy.

In B₁₂ deficiency, the B₁₂-dependent methionine synthase enzyme is inactive and cytosolic folate is trapped as 5-methyl-tetrahydrofolate at the expense of other folate coenzyme forms required for one-carbon metabolism such as thymidylate synthesis, leading to a functional folate deficiency in the cell. As the defective DNA synthesis in pernicious anemia is caused by an induced secondary folate deficiency, high levels of folate can cause a hematological response in patients with megaloblastic anemia due to B₁₂ deficiency, but folate is ineffective in preventing the severe neurological pathologies associated with B₁₂ deficiency.

Defects in vitamin B₁₂ absorption affect a significant proportion of elderly people. Classical pernicious anemia is caused by an inability to absorb vitamin B₁₂ as a result of lack of production of IF. The disease is age related and is usually due to destruction of the parietal cells in the stomach, which is caused by an autoimmune disease in which antibodies to the H⁺,K⁺-adenosine triphosphatase (ATPase) or to IF are produced. Because body vitamin B₁₂ stores are usually ample at the onset of the disease and the turnover of body stores is slow, it can take many years before deficiency symptoms become apparent.

Malabsorption of dietary vitamin B₁₂ also can arise from any condition of decreased acid production or pancreatic

insufficiency. The prevalence of atrophic gastritis in subjects 50 years and older has been estimated at 10–30%, and many elderly people suffer from *Helicobacter pylori* infection. These conditions result in an impairment of food vitamin B₁₂ absorption because of an impaired ability to release the vitamin from protein binders, while absorption of vitamin B₁₂ in the absence of food is unimpaired.

The mechanism behind the neurological symptoms of vitamin B₁₂ deficiency is not currently understood.

See also: [Metabolism Vitamins and Hormones: Amino Acid Metabolism; Biochemistry of Hematopoiesis; Protein/Enzyme Structure Function and Degradation: B₁₂-Containing Enzymes.](#)

Further Reading

- Bailey LB (ed.) (2009) *Folate in Health and Disease*, 2nd edn. Boca Raton, FL: CRC Press, Taylor and Francis Group.
- Chanarin I (1979) *The Megaloblastic Anemias*, 2nd edn. Oxford: Blackwell Scientific Publications.
- Fox JT and Stover PJ (2008) Folate-mediated one-carbon metabolism. *Vitamins and Hormones* 79: 1–44.
- Green R and Miller JW (2007) Vitamin B12. In: Zemleni J and Rucker RB (eds.) *Handbook of Vitamins*, 4th edn., pp. 413–457. Boca Raton, FL: CRC Press.
- Scott JM, Weir DG, and Kirke PN (1995) Folate and neural tube defects. In: Bailey LB (ed.) *Folate in Health and Disease*, pp. 329–360. New York, NY: Marcel Dekker.
- Shane B and Stokstad ELR (1985) Vitamin B₁₂–folate interrelationships. *Annual Review of Nutrition* 5: 115–141.
- Zemleni J, Rucker RB, Suttie J, and McCormick DB (eds.) (2011) *Handbook of Vitamins*, 4th edn. New York, NY: CRC.
- Zhao R, Matherly LH, and Goldman ID (2009) Membrane transporters and folate homeostasis: Intestinal absorption and transport into systemic compartments and tissues. *Expert Reviews in Molecular Medicine* 11: e4.

Free Oligosaccharides: Formation and Degradation (Free Oligosaccharides Related to N-Linked Glycans)

T Suzuki, RIKEN Advanced Science Institute, Wako, Japan

© 2013 Elsevier Inc. All rights reserved.

Glossary

Chaperones Molecules that, by associating with substrates, prevent them from aggregation and assist their folding or assembly to form macromolecular structures.

Complex-type glycans Vertebrate *N*-glycans can be classified into three types based on their structures: high mannose, hybrid, and complex-type. In complex-type glycans, both α 1,3- and α 1,6-linked core mannoses are modified by GlcNAc residues, in most cases followed by galactose and sialic acid residues. Most cell-surface- or secreted glycoproteins contain complex-type *N*-glycans.

N-linked glycan An oligosaccharide linked by asparagine in the protein/peptide, a common

modification in all eukaryotic proteins that go through the secretory pathway.

***N,N'*-diacetylchitobiose** A dimer of *N*-acetylglucosamine with β 1–4 linkage.

Proteasome A cytoplasmic gigantic protease complex involved in various cellular processes, including the degradation of misfolded proteins.

Quality-control machinery A system that maintains the quality of product, that is, newly synthesized proteins in the endoplasmic reticulum.

Sialic acid An acidic monosaccharide often found in nonreducing termini of vertebrate complex-type *N*- and *O*-glycans.

There is growing evidence that *N*-linked glycans play pivotal roles in protein folding and intra- or intercellular trafficking of *N*-glycosylated proteins. It has been known that during the *N*-glycosylation of proteins, significant amounts of free oligosaccharides (fOSs) are generated in the lumen of the endoplasmic reticulum (ER) by an unclarified mechanism. fOSs are also formed in the cytosol by enzymatic deglycosylation of misfolded glycoproteins destined to be degraded. Although the precise fate of intracellular fOSs remains obscure, recent studies have revealed that novel cellular machinery exists to process/degrade these glycans in a well-ordered, sophisticated fashion.

Free Oligosaccharides Formed in the Endoplasmic Reticulum

N-Glycosylation is one of the most common co- and posttranslational modifications of eukaryotic proteins occurring in the lumen of the endoplasmic reticulum (ER). During the translocation of proteins in the ER, an enzyme complex called oligosaccharyl transferase (OST) transfers an oligosaccharide (OS) moiety from the dolichol-linked donor substrate to asparagine residues located within the consensus sequence –Asn-Xaa-Ser/Thr– (where Xaa can be any amino acid except Pro) to form *N*-linked glycans on the nascent polypeptide chains. Although the biosynthesis of lipid-linked OSs, as well as processing of *N*-linked glycan chains on glycoproteins, is understood in detail, certain aspects of *N*-glycosylation reaction remain unknown. For instance, biochemical studies have shown that a significant pool of fOSs does occur in the lumen of the ER in mammalian cells. Although the exact molecular mechanism of the formation of these luminal fOSs still remains unknown, this release is clearly independent from the cytoplasmic peptide: *N*-glycanase (PNGase)-dependent process (see later; [Figure 1](#)).

These luminal glycans are believed to bear the *N,N'*-diacetylchitobiose structure at their reducing termini and are therefore called Gn2-type free glycans.

Free OS Transport System from the ER to the Cytosol

Although it is still not known whether fOSs have a physiological role in the ER, one would imagine that the accumulation of vast amounts of fOSs in the ER could interfere with the glycan-based quality-control system for nascent luminal proteins. Therefore, it is not surprising that cells have machinery to eliminate fOSs from the ER lumen. Using permeabilized cells and/or microsomes from mammalian sources, it has been demonstrated that fOSs in the ER can be exported into the cytosol in an adenosine triphosphate (ATP)-dependent manner. A gene encoding the transporter responsible for the export has not been identified. The OS export machinery appears to recognize the nonreducing mannose residues, but not the *N*-acetylglucosamine (GlcNAc) residue at the reducing terminus. The major structure of fOSs released from the ER to the cytosol has been shown to be $\text{Man}_{8-9}\text{GlcNAc}_2$ ([Figure 1](#)).

Free OSs Formed in the Cytosol: Its Connection with ER-Associated Degradation

Recent evidence clearly shows that the ER has quality-control machinery that differentiates between misfolded (glyco)proteins and correctly folded ones, so that only the latter exit from the ER to be delivered to their respective destinations. On the other hand, proteins that fail to fold or form functional subunit structures are retained in the ER and interact with various luminal chaperones that assist them to mature into their

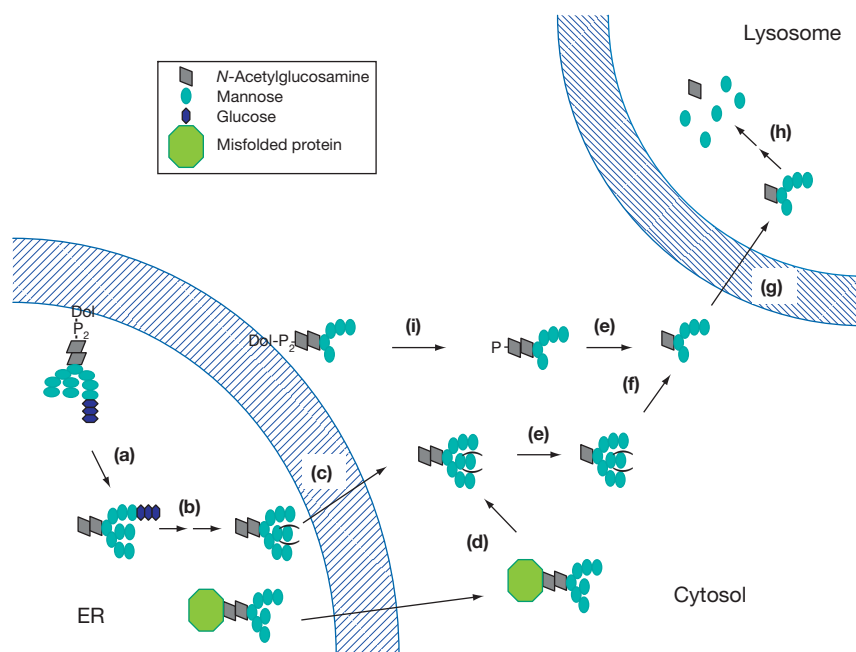


Figure 1 Current proposed model for the fate of free oligosaccharides (fOSs) formed in and outside of the ER in mammalian cells. fOSs generated in the lumen of the ER by an undefined mechanism bear *N,N'*-diacetylchitobiose structures at their reducing termini (Gn2 glycans) (step (a)). After quick deglycosylation by α -glucosidases I/II (and sometimes ER α -mannosidase I) (step (b)), $\text{Man}_{8-9}\text{GlcNAc}_2$ is transported into the cytosol by an oligosaccharide transporter on the ER membrane (step (c)). fOSs can also be generated by the action of the cytoplasmic PNGase on misfolded glycoprotein, generating Gn2 glycans (step (d)). Putative pyrophosphatase, the activity of which is reported on the cytosolic side of the ER membrane, releases phosphorylated Gn2 glycans (fOS-P) from the dolichol-linked oligosaccharides (step (e)). The fOS-P can be converted into Gn2 fOS by the putative fOS-P phosphatase (step (f)). In the cytosol, ENGase acts on Gn2 glycans (and possibly fOS-P or misfolded glycoproteins) to form Gn1 glycans (step (g)). Gn1 glycans are now susceptible for the action of Man2C1, giving rise to specific $\text{Man}_5\text{GlcNAc}$ structures (step (h)). The isomeric structure of Man_5 is identical to that of the last biosynthetic intermediate on dolichol oriented to the cytosolic side of the ER. The $\text{Man}_5\text{GlcNAc}$ glycan (and possibly more trimmed $\text{Man}_{3-4}\text{GlcNAc}$ glycans) may be transported into the lysosomes by a specific oligosaccharide transporter, which remains to be identified (step (i)). The fOSs, incorporated into the lysosomes, are now hydrolyzed into monomeric sugars by lysosomal glycosidases for reutilization (step (j)). Note that this scheme cannot be applied to other organisms; for example, budding yeast *Saccharomyces cerevisiae* does not possess ENGase/Man2C1, and the occurrence of a totally distinct pathway for fOSs generation/degradation is reported.

functional structures. When proteins consistently fail to acquire the correct state, however, they eventually are degraded by the system called 'ER-associated degradation' or ERAD. It is now clear that this degradation process involves retrotranslocation of the defective proteins from the ER to the cytosol, followed by their degradation by the 26S proteasome. Evidence shows that glycan chains on glycoproteins play critical roles in monitoring the folding state of glycoproteins in the ER that involves various intracellular lectins.

When the misfolded glycoproteins are retrotranslocated to the cytosol for proteasomal degradation, the cytoplasmic PNGase can remove *N*-glycans from these glycoproteins before or during proteasomal degradation (Figure 1).

Another enzyme that can generate fOSs in the cytosol is a putative pyrophosphatase to generate phosphorylated fOSs (fOS-Ps) (Figure 1). Such enzyme activity has been detected mainly in mammalian cells. It has also been suggested that there may be fOS-P phosphatase activity to generate Gn2-type fOSs, but the molecular nature of these enzymes remains unknown.

Irrespective of their source, the cytosolic Gn2-type glycans can be converted into Gn1-type fOSs, bearing only a single GlcNAc residue at their reducing termini (Figure 1). This

reaction can be catalyzed by the action of endo- β -*N*-acetylglucosaminidase (ENGase). There is also a possibility that, in some cases, ENGase may release Gn1-type glycans directly from glycoproteins or dolichol-linked OS, while such reaction has not been unequivocally demonstrated. It should be noted, however, that a number of proteins with *N*-linked single GlcNAc, which is potentially formed by the action of ENGase, have been identified through proteomics approach. In mammalian cells, most of glycans accumulated in the cytosol are Gn1-type glycans, suggesting that ENGase action is not a rate-limiting step for fOS-catabolism.

Gn1 glycans formed in the cytosol can be further catabolized by the cytoplasmic α -mannosidase, Man2C1 (Figure 1). It has been shown that Man2C1 activity prefers Gn1 over Gn2 glycans as substrates, suggesting that reactions involving ENGase and Man2C1 are well ordered, that is, Man2C1 action follows ENGase. The final product of Man2C1 primarily is $\text{Man}_5\text{GlcNAc}$, which possesses the same isomeric structure as the biosynthetic intermediate of dolicholpyrophosphoryl OS (Figure 1). The $\text{Man}_5\text{GlcNAc}$ glycan is believed to be delivered into the lysosomes by a putative transporter for further degradation into monomeric sugars for reutilization. The nature of the lysosomal OS transporter is still unknown.

Species Specificity of Formation/Degradation of fOSs

Thus far, most of the studies on fOSs are based on ones using mammalian cells. However, there is accumulating evidence that the formation/processing mechanism of fOSs is distinct between different species. For instance, budding yeast (*Saccharomyces cerevisiae*) release fOSs almost exclusively by the PNGase-dependent pathway, while a PNGase-independent pathway appears to be dominant in mammalian cells. Also, in this yeast, all the detected fOSs were Gn2 type, and consistent with this observation, no ENGase activity was found in *S. cerevisiae*. In the worm *Caenorhabditis elegans*, neither enzyme activity nor orthologous gene for Man2C1 was found. Now, it is also clear that, while the gene for PNGase is well conserved among eukaryotes, the ortholog proteins in some species, such as *Neurospora crassa*, apparently have no enzyme activity as they retain critical mutations in the amino acid residues. Therefore, one could imagine that the formation/degradation process for fOSs should be determined in various distinct species to unveil the whole picture of this nonlysosomal glycan-processing pathway.

Formation of Complex-Type (Golgi-Type) Glycans

It should also be noted that processes depicted in [Figure 1](#) are only for high mannose-type glycans, and occurrence of fOSs with more processed glycans, that is, complex-type glycans in the case of mammalian cells, is just starting to be recognized. As a result, the molecular mechanism of their formation/degradation is completely unknown. In budding yeast, occurrence of PNGase-dependent, Golgi-modified fOSs are observed, suggesting that these fOSs should be derived from misfolded glycoproteins once they reach the Golgi apparatus. Because the Golgi-type (complex-type) glycans in mammalian cells contain not only mannose but also other sugars, such as sialic acid or galactose, the nonlysosomal catabolic pathway for complex-type glycans, if any, should be literally 'complex' than the case with the one for high mannose-type glycans that needs to be explored in the future.

Concluding Remarks

Cells possess a variety of enzymes/transporters to handle fOSs generated by several distinct mechanisms. Cytosolic fOSs, whether directly formed by the action of the cytoplasmic PNGase or transported into the cytosol from the ER lumen, are processed by ENGase and Man2C1 before being incorporated into the lysosomes for degradation. Why this rather complicated pathway is necessary for fOS catabolism is unknown. One notion is that this trafficking route enables them to be removed quickly from the ER, thereby preventing them from interfering with the efficient glycan-dependent quality-control system for newly synthesized glycoproteins in

the ER. However, it is still possible that a particular fOS itself may play a role in certain cellular processes. Indeed, fOS derivatives of *N*-glycan chains are suggested to play roles in the growth, senescence, and ripening of fruit in plants. Although similar bioactivity in other organisms still remains to be clarified, the identification of molecules involving formation, processing, and transport of fOSs should provide us more insight into the mechanism and function of this novel pathway.

See also: [Lipids Carbohydrates Membranes and Membrane Proteins: Glycoprotein Folding and Processing Reactions; Glycoproteins, N-Linked; N-Linked Glycan-Processing Glucosidases and Mannosidases.](#)

Further Reading

- Chalkley RJ, Thalhammer A, Schoepfer R, et al. (2009) Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proceedings of the National Academy of Sciences of the United States of America* 106: 8894–8899.
- Chantret I, Fasseu M, Zaoui K, et al. (2010) Identification of roles for peptide: *N*-glycanase and endo- β -*N*-acetylglucosaminidase (Engase1p) during protein glycosylation in human HepG2 cells. *PLoS One* 5: e11734.
- Chantret I and Moore SEH (2008) Free oligosaccharide regulation during mammalian protein N-glycosylation. *Glycobiology* 18: 210–224.
- Funakoshi Y and Suzuki T (2009) Glycobiology in the cytosol: The bitter side of a sweet world. *Biochimica et Biophysica Acta – General Subject* 1790: 81–94.
- Funakoshi Y, Negishi Y, Gergen JP, et al. (2010) Evidence for an essential deglycosylation-independent activity of PNGase in *Drosophila melanogaster*. *PLoS One* 5: e10545.
- Haga Y, Totani K, Ito Y, and Suzuki T (2009) Establishment of a real-time analytical method for free oligosaccharide transport from the ER to the cytosol. *Glycobiology* 19: 987–994.
- Hirayama H, Seino J, Kitajima T, Jigami Y, and Suzuki T (2010) Free oligosaccharides to monitor glycoprotein endoplasmic reticulum-associated degradation in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* 285: 12390–12404.
- Ishizuka A, Hashimoto Y, Naka R, et al. (2008) Accumulation of free complex-type *N*-glycans in stomach cancer cells, MKN7 and MKN45. *Biochemical Journal* 413: 227–237.
- Kato T, Kitamura K, Maeda M, et al. (2007) Free oligosaccharides in the cytosol of *Caenorhabditis elegans* are generated through endoplasmic reticulum–Golgi trafficking. *Journal of Biological Chemistry* 282: 22080–22088.
- Kato H, Ashida H, and Yamamoto K (2009) Generation and metabolism of cytosolic free oligosaccharides in *C. elegans*. *Trends in Glycoscience and Glycotechnology* 21: 163–177.
- Maerz S, Funakoshi Y, Negishi Y, Suzuki T, and Seiler S (2010) The neurospora peptide: *N*-glycanase orthologue PNG1 is essential for cell polarity despite its lack of enzymatic activity. *Journal of Biological Chemistry* 285: 2326–2332.
- Peric D, Durrant-Arico C, Delanda C, et al. (2010) The compartmentalization of phosphorylated free oligosaccharides in cells from a CDG Ig patient reveals a novel ER-to-cytosol translocation process. *PLoS One* 5: 311675.
- Suzuki T (2007) Cytoplasmic peptide: *N*-glycanase and catabolic pathway for free *N*-glycans. *Seminars in Cell and Developmental Biology* 18: 762–769.
- Suzuki T (2009) Introduction to "Glycometabolome". *Trends in Glycoscience and Glycotechnology* 21: 219–227.
- Suzuki T, Hara I, Nakano M, et al. (2006) Man2C1, an α -mannosidase is involved in the trimming of free oligosaccharides in the cytosol. *Biochemical Journal* 400: 33–41.
- Vleugels W, Duvel S, Peanre R, et al. (2011) Identification of phosphorylated oligosaccharides in cells of patients with a congenital disorders of glycosylation (CDG-I). *Biochimie* 93: 823–833.

Friedreich's Ataxia

P E Hart and A H V Schapira, University College London, London, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Antioxidant Any substance that delays or prevents the progress of oxidation.

Ataxia Incoordination of voluntary muscle action.

Mitochondria Cytoplasmic organelle, consisting of an outer membrane and an inner folded membrane, and containing its own DNA.

Mitochondrial DNA A circular DNA molecule of 16.5 kb contained within the mitochondrion and inherited maternally.

Mitochondrial respiratory chain A series of enzyme complexes located on the mitochondrial inner membrane responsible for generating a proton gradient, which is itself required for the conversion of adenosine diphosphate to adenosine triphosphate (ATP) by complex V (ATP synthase) of this chain.

Trinucleotide repeat A three-nucleotide sequence, present in normal individuals as a set number of repeats, but which causes disease when a threshold number of repeats are exceeded.

Friedreich's ataxia (FRDA) is an autosomal recessive disorder that causes ataxia, sensory loss, cardiomyopathy, skeletal abnormalities, and, in a proportion of patients, diabetes and optic atrophy. It is the most common inherited ataxia with an estimated prevalence of 1:29 000. Carrier prevalence is between 1:60 and 1:90. The responsible gene and its gene product, frataxin, have both been identified. Ongoing investigations into the function of frataxin and the pathophysiology of FRDA are opening up new avenues for therapeutic intervention in this devastating neurodegenerative disorder.

Historical Perspective

In a series of five papers published between 1863 and 1877, Nicholas Friedreich reported nine members of three families who had an onset in puberty of ataxia, dysarthria, sensory loss, muscle weakness, scoliosis, foot deformity, and cardiac symptoms. Areflexia was only described in the final two reports, but was later adopted as one of the clinical diagnostic criteria. Over a century later, the elucidation of the molecular mechanism of FRDA, and the availability of genetic testing for diagnosis, allowed greater clarity regarding the phenotypic spectrum of this disorder.

Clinical Features

Disease onset ranges from 1.5 to 51 years in reported series. Loss of ambulation has been reported to occur at a mean of 15.5 ± 7.4 years after disease onset but ranges between 3 and 44 years. Death is most commonly as a consequence of cardiomyopathy and is at a mean age of 37.5 ± 14.4 years. Retrospective analysis of disease progression has reported a mean time to wheelchair confinement of 11 years. Atypical presentations include onset over the age of 25 years (late-onset Friedreich's ataxia – LOFA), the preservation of lower limb reflexes (Friedreich's ataxia with retained reflexes – FARR), and presentation as a pure spastic paraparesis, spastic ataxia (Acadian FRDA), pure sensory ataxia, or chorea.

Pathology

Pathological studies of FRDA report changes maximal in dorsal root ganglia, dorsal columns, corticospinal tracts, and heart. Macroscopically, the spinal cord is atrophic, with the posterior and lateral columns particularly affected. Changes within the nervous system are thought to be the consequence of a dying-back process from the periphery affecting the longest and largest myelinated fibers that show changes of an axonopathy. Demyelination is seen in the dorsal columns. The cerebellar cortex shows only mild neuronal loss, but the dentate nucleus and Clarke's column in the cord show marked changes. The cerebellar and occipital cortex show reduced phospholipid levels in the absence of neuronal loss.

Genetics

In 1988, the FRDA gene was mapped to chromosome 9. The gene was linked to 9q13-21.1 in 1990 and cloned in 1996. Ninety-eight percent of cases are now known to be the result of a homozygous GAA triplet repeat expansion in intron 1 of the FRDA gene. This is a unique trinucleotide repeat (TNR) disorder in that its inheritance is autosomal recessive, its location is intronic, and it involves a GAA trinucleotide as opposed to CAG repeat as found in the majority of TNR disorders. The repeat length in normal individuals is 6–34, but is expanded in patients and carriers to between 67 and 1700. Repeat lengths are unstable. Paternal transmission is associated with a reduction in repeat length, an effect that increases with increasing paternal age. Maternal transmission can cause an increase or decrease in repeat length, and expansions are greater with increasing maternal age. The remaining 2% of patients are compound heterozygotes harboring an expanded repeat on one allele and a point mutation on the other. Twenty-three different point mutations are described to date and include missense, frameshift, splice site, initiation codon, and non-sense mutations. The former are found only in the C-terminal, suggesting that functional domains reside in this part of the protein. The location of these mutations in highly or poorly

conserved amino acids correlates with the severity of the phenotype and the presence of atypical features. No cases resulting from homozygous point mutations have been described, although the population incidence of such individuals has been calculated to be $1:100 \times 10^6$. Clinically, typical FRDA may occur in individuals in the absence of linkage to chromosome 9, suggesting that a second locus may exist.

Phenotype–Genotype Correlations

Clinical parameters correlate with the repeat-length size. The size of GAA1 (the larger repeat allele) accounts for between 33% and 73% of the variation in the age of onset in various studies. GAA2 (the smaller allele) accounts for less than 20% of this variability. Indeed, GAA1 correlated better than GAA2 for several disease parameters and disease complications, including cardiomyopathy, in all studies except one. By contrast, the development of diabetes mellitus in FRDA correlated in only one study.

Further variation may result from other factors. The GAA repeat length may vary in a tissue-specific pattern due to mitotic instability, and peripheral blood samples may be a poor indicator of repeat lengths in pathologically affected tissues. GAA1 has been shown to differ in different brain regions and in various tissues. *cis*-Acting factors may also influence the phenotype. Potential mechanisms would include effects upon the stability of the trihelix structure by sequence alterations within or flanking the GAA expansion. Other genetic or environmental factors may also exert an influence on the phenotype. This is illustrated by the finding that the age of onset in sibs correlates strongly regardless of the degree of difference between their repeat lengths.

Gene and Gene Product

The FRDA gene contains seven exons within 80 kb of nuclear DNA. Transcription most commonly generates a 1.3-kb product translated into a 210-amino-acid protein named 'frataxin'. Alternative splicing generates a 171-amino-acid protein of uncertain significance. Frataxin and messenger RNA (mRNA) levels are maximal in tissues of high mitochondrial content (heart, pancreas, liver, and skeletal muscle), although not all of these exhibit obvious clinical involvement in FRDA. Within the central nervous system, mRNA levels are highest in the cord, low in the cerebellum, and very low in the cerebral cortex. Frataxin appears to play a role in development; its homozygous knockout is lethal at an early embryological stage. Frataxin mRNA levels are high in fetal spinal cord, dorsal root ganglion, heart, liver, skeletal muscle, and skin. Lymphocytes from patients with homozygous expansions contain low levels of frataxin, and this level and that of the mRNA are inversely related to the size of the smaller repeat. The block is thought to occur at the level of transcript elongation, and the mechanism for this is believed to be through the formation of unusual DNA structures, such as DNA triplexes, and sticky DNA by the GAA/TTC repeats. Several point mutations appear to alter the secondary structure of the protein, thus influencing mitochondrial uptake or cleavage.

The function of frataxin is incompletely understood. The amino acid sequence shows no strong homology to any proteins of known function. It has, however, been shown to contain an N-terminal mitochondrial targeting sequence in its first 55 amino acids. The predicted mitochondrial location was confirmed when tagged expressed frataxin was shown to co-localize with mitochondrial markers in HeLa and COS cells. An inner membrane and a matrix location within the mitochondrion have both been proposed. X-Ray crystallography studies reveal similarities to ferritin, a site for protein–protein interaction, and the ability to bind one molecule of iron. Point mutations in the protein core cause more severe phenotypes than those within the ferritin-like anionic patch or flat external protein interaction surface. Studies have confirmed that frataxin is critical for the formation of iron–sulfur (Fe–S) clusters and in their protection against oxidation, and in heme biosynthesis.

A number of animal and cell models of FRDA, including transgenic mice, have provided insights into the pathogenesis of FRDA. Increased susceptibility to oxidative stress, increase in cellular and mitochondrial iron content, defects of the Fe–S center-containing complexes of the mitochondrial respiratory chain, and reduced mitochondrial DNA levels have all been reported. Some of these changes have also been reported in human tissue. A role for frataxin in mitochondrial Fe–S center synthesis has been proposed. The exact relationship between these biochemical perturbations remains incompletely understood.

Therapeutic Intervention

Based on the improved understanding of the pathogenesis of FRDA, a number of novel therapeutic strategies have been evaluated. Iron chelation restores mitochondrial iron levels and prevents mitochondrial respiratory chain dysfunction in yeast models, but their *in vivo* use has been problematic. Antioxidants, including idebenone, coenzyme Q₁₀, vitamin E, N-acetyl cysteine, and selenium, have also been used. In some studies, idebenone has resulted in a decrease in cardiac hypertrophy and decreases in surrogate markers of oxidative damage to DNA. Other studies have failed to show a benefit. Combined therapy with coenzyme Q₁₀ and vitamin E has shown improvement in magnetic resonance spectroscopy and echocardiographic parameters and possibly slowing of neurological progression in a proportion of patients – particularly those with low coenzyme Q₁₀ levels at baseline.

Future therapeutic trials in FRDA may make use of antioxidants or other agents coupled to the triphenylphosphonium cation, which facilitates mitochondrial targeting of the agent. Potentials for gene therapy are also being explored, and may utilize drugs that interfere with the sticky DNA structures that are thought to be the cause of the transcription blockade that occurs in FRDA.

See also: **Bioenergetics:** Mitochondrial DNA; **Metabolism**
Vitamins and Hormones: Diabetes.

Further Reading

- Campuzano V, Montermini L, Lutz Y, et al. (1997) Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. *Human Molecular Genetics* 6(11): 1771–1780.
- Campuzano V, Montermini L, Molto MD, et al. (1996) Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271(5254): 1423–1427.
- Cooper JM, Korlipara LV, Hart PE, Bradley JL, and Schapira AH (2008) Coenzyme Q₁₀ and vitamin E deficiency in Friedreich's ataxia: Predictor of efficacy of vitamin E and coenzyme Q₁₀ therapy. *European Journal of Neurology* 15(12): 1371–1379.
- Durr A, Cossee M, Agid Y, et al. (1996) Clinical and genetic abnormalities in patients with Friedreich's ataxia. *New England Journal of Medicine* 335(16): 1169–1175.
- Hart PE, Lodi R, Rajagopalan B, et al. (2005) Antioxidant treatment of patients with Friedreich ataxia: Four-year follow-up. *Archives of Neurology* 62(4): 621–626.
- Lodi R, Hart PE, Rajagopalan B, et al. (2001) Antioxidant treatment improves *in vivo* cardiac and skeletal muscle bioenergetics in patients with Friedreich's ataxia. *Annals of Neurology* 49(5): 590–596.
- Mancuso M, Orsucci D, Choub A, and Siciliano G (2010) Current and emerging treatment options in the management of Friedreich ataxia. *Journal of Neuropsychiatric Disease and Treatment* 6: 491–499.
- Muhlenhoff U, Richhardt N, Ristow M, Kispal G, and Lill R (2002) The yeast frataxin homolog Yfh1p plays a specific role in the maturation of cellular Fe/S proteins. *Human Molecular Genetics* 11(17): 2025–2036.
- Pandolfo M (2008) Drug insight: Antioxidant therapy in inherited ataxias. *Nature Clinical Practice Neurology* 4(2): 86–96.
- Santos R, Lefevre S, Sliwa D, et al. (2010) Friedreich ataxia: Molecular mechanisms, redox considerations, and therapeutic opportunities. *Antioxidants and Redox Signaling* 13(5): 651–690.

F_X, F_A, and F_B Iron–Sulfur Clusters in Type I Photosynthetic Reaction Centers

B Jagannathan and J H Golbeck, The Pennsylvania State University, University Park, PA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Mikhail L. Antonkine and John H. Golbeck, volume 3, pp. 348–356, © 2004, Elsevier Inc.

Glossary

Cofactor Organic or inorganic molecule bound (covalently or through noncovalent interactions) to the protein that is necessary for/aiding its function. For example, chlorophylls, quinones, and iron–sulfur clusters are cofactors in photosystem I.

F_A and F_B Low-potential [4Fe–4S] clusters that act as the terminal electron acceptors in type I photosynthetic reaction centers.

F_X A low-potential, interpolypeptide [4Fe–4S] cluster that acts as an intermediate electron

acceptor in type I photosynthetic reaction centers.

High-potential iron–sulfur proteins (HiPIP) These contain [4Fe–4S] clusters that operate between oxidation states of 3⁺/2⁺.

Low-potential iron–sulfur proteins These contain [4Fe–4S] clusters that operate between oxidation states of 2⁺/1⁺.

Type I Photosynthetic reaction centers that use [4Fe–4S] clusters as the terminal electron acceptors.

Type II Photosynthetic reaction centers that use a mobile quinone as the terminal electron acceptor.

Photosynthesis and Types of Photosynthetic Reaction Centers

Photosynthesis is the process by which solar energy is converted into chemical bond energy by plants and photosynthetic bacteria. Two types of photosynthetic reaction centers have evolved that are capable of performing light-induced charge separation. Type I reaction centers utilize iron–sulfur clusters as terminal electron acceptors, and type II reaction centers utilize a mobile quinone as the terminal electron acceptor. Type I reaction centers are present in green-sulfur bacteria, heliobacteria, and the recently discovered phototrophic acidobacteria, and as photosystem I (PS I) in cyanobacteria, algae, and higher plants. Type II reaction centers are present in purple bacteria and green-nonsulfur bacteria, and as photosystem II (PS II) in cyanobacteria, algae, and higher plants.

Oxygenic photosynthesis can be conveniently divided into two parts, the light reactions and the dark reactions. The light reactions employ four major membrane-bound multisubunit complexes in the energy-conversion process: PS I, PS II, the cytochrome *b₆f*-Rieske iron–sulfur complex, and adenosine triphosphate (ATP) synthase. In cyanobacteria and plants, the two reaction centers operate in series, producing oxygen as a by-product of the water-splitting reaction in PS II, and nicotinamide adenine dinucleotide phosphate, reduced form (NADPH) as a product of the reduction of nicotinamide adenine dinucleotide phosphate (NADP⁺) in PS I. A proton gradient is generated, in part, by the cytochrome *b₆f*-Rieske iron–sulfur complex and utilized by ATP synthase for the generation of ATP. During the dark reactions, carbon dioxide is enzymatically converted into carbohydrates at the expense of ATP and NADPH via the Calvin–Benson cycle. The chemical bond energy of ATP and NADPH is also used for cellular processes such as metabolism, growth, and reproduction.

Green-sulfur bacteria and heliobacteria contain only a type I reaction center and hence do not evolve oxygen. These bacteria are termed ‘anoxygenic phototrophs’ and thrive only

under low concentrations of molecular oxygen. The recently discovered *Candidatus Chloracidobacterium thermophilum* similarly contains only a type I reaction center but lives in an oxygen-containing environment.

Iron–Sulfur Clusters in Biology

Iron–sulfur clusters are among the most versatile cofactors in biology. They are involved in a wide variety of roles, such as imparting structural stability (endonuclease III and archaeal RNA polymerase), performing catalysis (radical S-adenosylmethionine (SAM) enzymes and aconitase), and carrying out the regulation of gene expression (SoxR). The most common function of iron–sulfur clusters, however, is in mediating biological electron transfer. Proteins that contain iron–sulfur clusters can be as simple as a bacterial dicluster ferredoxin or as sophisticated as the mitochondrial complex I.

There are four basic classes of iron–sulfur clusters. Each differs in the number of iron and sulfur atoms (**Figure 1**), resulting in a wide variety of cluster geometries, reduction potentials, and biological functions. (1) A single iron atom tetrahedrally coordinated by four cysteine ligands is present in rubredoxin. (2) A [2Fe–2S] cluster containing two iron atoms bridged by two sulfur atoms in a rhombic structure is present in plant ferredoxins. The iron atoms in a [2Fe–2S] cluster are typically ligated by cysteine residues, although Rieske iron–sulfur proteins employ two histidine ligands to one of the irons. (3) A [4Fe–4S] cluster containing four iron atoms and four sulfur atoms in alternate corner positions of a cube is present in bacterial ferredoxins. Each iron atom is ligated by a cysteine residue, although in rare instances one position is occupied by water or aspartate. (4) A [3Fe–4S] cluster with a geometry similar to that of a [4Fe–4S] cluster, except that one of the iron atoms is missing from the corner of the cube (**Figure 1**), is present in aconitase.

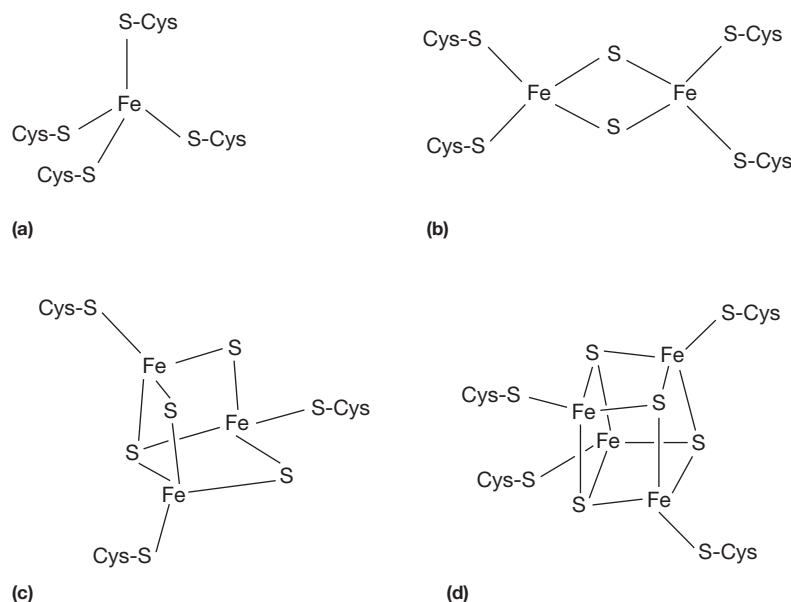


Figure 1 Schematic representation of the different types of simple iron–sulfur clusters found in biology. (a) Rubredoxin-like Fe center, (b) [2Fe–2S] cluster, (c) [3Fe–4S] cluster, and (d) [4Fe–4S] cluster.

Among these classes of iron–sulfur clusters, the [4Fe–4S] clusters are the most versatile because they can exist in three different oxidation states: 3^+ , 2^+ , and 1^+ . However, in any given protein, only one of the two oxidation–reduction pairs is used. High-potential iron–sulfur proteins (HiPIPs) employ the $3^+/2^+$ redox couple and exhibit midpoint potentials ranging from +50 mV to +450 mV. Low-potential iron–sulfur proteins employ the $2^+/1^+$ couple and exhibit midpoint potentials ranging from –200 mV to –700 mV. The protein environment plays a crucial role in determining which oxidation–reduction pair is used and in modulating the midpoint potential of the oxidation–reduction pair. Factors such as the immediate electrostatic environment, the number and strengths of H-bonds, solvent accessibility, and polarizability of the protein play important roles in modulating the thermodynamic properties of a given cluster.

Type I reaction centers contain three low-potential [4Fe–4S] $^{2+/1+}$ clusters termed F_X , F_A , and F_B . In the oxidized state, each [4Fe–4S] $^{2+}$ cluster contains two Fe^{2+} and two Fe^{3+} ions, which are manifest as two $Fe^{2.5+}$ – $Fe^{2.5+}$ mixed-valence pairs due to valence delocalization. As a result of paramagnetic coupling between the two $Fe^{2.5+}$ – $Fe^{2.5+}$ mixed-valence pairs, the net spin is zero and hence [4Fe–4S] $^{2+}$ clusters are diamagnetic. In the reduced state, a low-potential [4Fe–4S] $^{1+}$ cluster contains three Fe^{2+} ions and one Fe^{3+} ion, which are manifest as an equal-valence Fe^{2+} – Fe^{2+} pair and a mixed-valence $Fe^{2.5+}$ – $Fe^{2.5+}$ pair. The position of both pairs, with respect to protein ligands, can change at room temperature, and as a result, the mixed- and equal-valence pairs are free to migrate around the cube. However, because [4Fe–4S] $^{2+,1+}$ clusters are slightly distorted due, in part, to the interactions with an inhomogeneous matrix, preferential positions for both pairs exist. As a result of paramagnetic coupling between the equal-valence Fe^{2+} – Fe^{2+} pair and the mixed-valence $Fe^{2.5+}$ – $Fe^{2.5+}$ pair, the net spin is 1/2 and hence the [4Fe–4S] $^{1+}$ clusters are paramagnetic. The latter allows detection by low-temperature

electron paramagnetic resonance (EPR) spectroscopy. The EPR properties of a paramagnetic iron–sulfur cluster are additionally sensitive to the immediate protein environment, hence, a change in structure typically leads to a change in spectrum. EPR spectroscopy is especially advantageous for the study of cofactors in large protein complexes, wherein it is difficult to apply simpler techniques due to the presence of interfering chromophores or to the size of the biomolecule.

Type I Reaction Centers

Photosystem I

From a structural point of view, PS I is the best characterized of the type I reaction centers. A model based on the 2.5 Å resolution X-ray crystal structure of cyanobacterial PS I is available, providing detailed structural information on the three-dimensional (3D) arrangement of pigments, proteins, and cofactors (Figure 2). Each monomer of the cyanobacterial PS I complex contains 12 protein subunits, 96 chlorophyll *a* (Chl *a*) molecules, 22 carotenoids, four lipids, two phytylquinones, and three [4Fe–4S] clusters. PsaA (83 kDa) and PsaB (83 kDa) form the membrane-embedded heterodimeric core and contain the majority of the antenna molecules and redox cofactors. PsaC (9 kDa) is a membrane-extrinsic (stromal) subunit that contains the two terminal electron acceptors, F_A and F_B . PsaD (15 kDa) and PsaE (8 kDa) are stromal subunits and, along with PsaC, provide a docking site for the soluble electron acceptors ferredoxin and flavodoxin. Cyanobacterial PS I incorporates seven additional subunits (PsaF, PsaI, PsaJ, PsaK, PsaL, PsaM, and PsaX), many of which serve as chlorophyll-binding proteins.

The electron transfer chain includes a special pair of Chl *a* molecules termed ' P_{700} ', named for its peak absorbance in the visible region. According to current understanding, when P_{700} becomes excited to the singlet state, its electron is transferred to

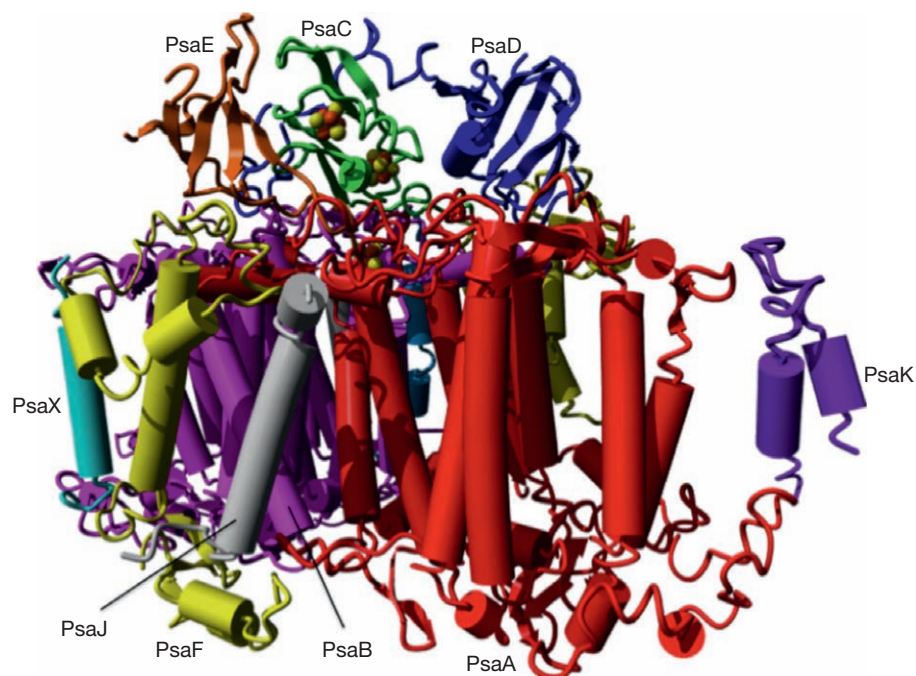


Figure 2 Side view of the arrangement of all proteins in one monomer of cyanobacterial PS I (PDB entry 1JB0), with the main subunits indicated.

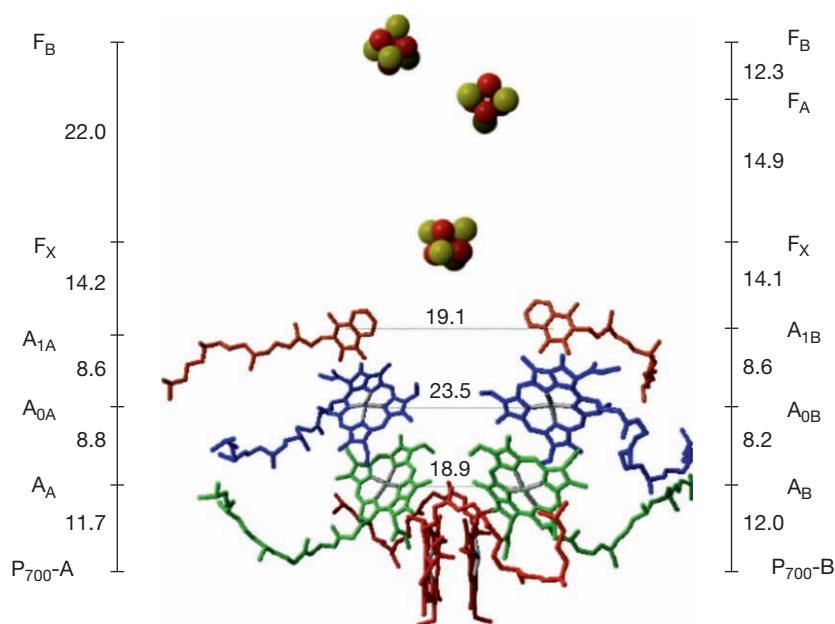


Figure 3 The arrangement of redox cofactors in cyanobacterial PS I with center-to-center distances depicted in Å (PDB entry 1JB0).

A_0 , a Chl a monomer (Figure 3). An additional molecule of Chl a (A) is located between P_{700} and A_0 that may act as an electron transfer intermediate. The initial $P_{700}^+A_0^-$ charge-separated state is stabilized by rapid transfer of the electron to a bound phyloquinone, termed A_1 , and then to a series of $[4Fe-4S]$ clusters, termed F_X , F_A , and F_B . The latter functions as a molecular wire, shuttling the electron to the stromal surface and making it accessible to ferredoxin or flavodoxin. Because the terminal electron acceptors are iron–sulfur clusters, and

because the core is composed of PsaA and PsaB subunits, PS I is termed a ‘heterodimeric type I reaction center’.

The Heliobacterial and Green-Sulfur Bacterial Reaction Centers

Although heliobacteria and green-sulfur bacteria employ a similar contingent of electron transfer cofactors, the polypeptides that bind the cofactors are significantly different.

A (PshA)₂ homodimer and a (PscA)₂ homodimer form the membrane-embedded cores of the heliobacterial and green-sulfur bacterial reaction centers, respectively. PshB in heliobacteria and PscB in green-sulfur bacteria bind the terminal F_A and F_B clusters.

The primary donor is a special pair of bacteriochlorophyll molecules termed P_{798} in heliobacteria and P_{840} in green-sulfur bacteria. The primary electron acceptor is considered to be a chlorophyll *a* derivative in both species. Similar to PS I, two branches of electron transfer cofactors are thought to exist that converge at the F_X cluster, with subsequent linear electron transfer to F_A and F_B . The participation of a quinone in forward electron transfer is still under debate in both heliobacterial and green-sulfur bacterial reaction centers.

The absence of a high-resolution crystal structure for a homodimeric type I reaction center from either heliobacteria or green-sulfur bacteria precludes visualization of the 3D arrangement of the redox cofactors and the polypeptide subunits. Structural inferences from biochemical and spectroscopic data rely heavily on using PS I as a model type I reaction center.

The F_X Cluster in Type I Reaction Centers

The F_X Cluster in Photosystem I

F_X is a interpolypeptide, low-potential $[4Fe-4S]^{2+,1+}$ cluster bound at the interface of the PsaA and PsaB polypeptides in PS I. It is one of the most reducing $[4Fe-4S]$ clusters known in biology, with a measured midpoint potential of -705 mV.

F_X is ligated by two cysteine residues from PsaA and two cysteine residues from PsaB that are part of highly conserved loop regions connecting the transmembrane α -helices *h* and *i* in both proteins (Figure 4). In addition to coordinating F_X , the loop regions shield the $[4Fe-4S]$ cluster from molecular oxygen. A well-protected environment around F_X is critical for stability, particularly during the assembly of PS I. Otherwise, the bridging sulfides (S^{2-}) of an iron–sulfur cluster would

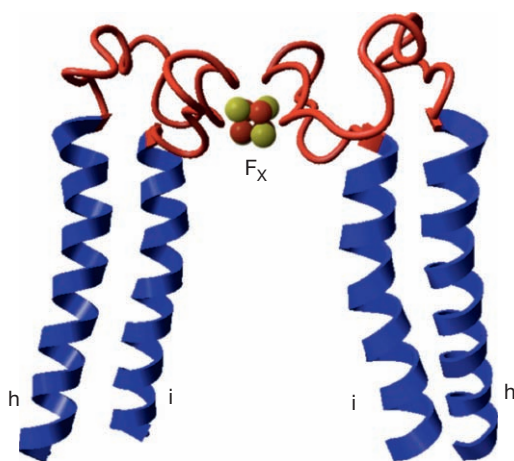


Figure 4 A view of the loop regions that connect transmembrane helices *h* and *i* on PsaA (A-*hi*) and PsaB (B-*hi*) (PDB entry 1JB0). The loops contain the Cys residues that coordinate the F_X cluster, protect it from oxidative denaturation, and provide the Asp residues that form symmetric ionic bonds with PsaC.

become oxidized to the elemental S^0 state, and a persulfide bridge would be formed between the cysteine residues, causing loss of iron and degradation of the cluster. The loop regions on PsaA and PsaB also contain the negatively charged Asp residues that provide a portion of the binding surface for the PsaC subunit. The binding interface between PsaC and the PsaA/PsaB heterodimer will be discussed later in this article.

F_X exists in an $S=1/2$ ground-spin state and exhibits characteristic low-temperature EPR resonances at *g*-values 2.06, 1.86, and 1.76. Prior to the determination of the X-ray crystal structure of cyanobacterial PS I, the identity of F_X as a $[4Fe-4S]$ cluster was established by EPR and Mössbauer spectroscopy.

The F_X Cluster in the Heliobacterial and Green-Sulfur Bacterial Reaction Centers

The F_X -binding motif (FPCXGPXXGGTC, where X is a noncysteine amino acid) occurs in the PscA protein of green-sulfur bacteria and the PshA protein of heliobacteria. Although X-ray crystal structures are not available, there is little doubt that these two cysteine residues coordinate F_X in the (PscA)₂ homodimer of green-sulfur bacteria and the (PshA)₂ homodimer of heliobacteria in a manner similar to that in the PsaA/PsaB heterodimer of PS I. The F_X -coordinating loops in the PscA and PshA proteins do not contain the negatively charged Asp residues that are involved in PsaC–PsaA/PsaB interactions in PS I. It therefore appears that homodimeric type I reaction centers adopt a different mode of interaction with the F_A and F_B -containing protein. Interestingly, the F_X cluster appears to be relatively insensitive to oxidative denaturation in heliobacteria and green-sulfur bacteria, implying that the cysteine-containing loop regions shield the iron–sulfur cluster in a manner similar to that in PS I.

The magnetic properties of F_X are somewhat different in homodimeric type I reaction centers. It has long been known that the F_X cluster in PS I exists in an $S=1/2$ ground-spin state, whereas it has recently been shown that the F_X cluster in the green-sulfur bacterial and in the heliobacterial reaction centers exists in either an $S=1/2$ or an $S=3/2$ ground-spin state (or in a mixture of the two). The $S=1/2$ spin state is detected by EPR as rhombic resonances around $g=2$, and the $S=3/2$ spin state is characterized by low-field turning points between $g=4$ and $g=5$. Although the factors that force the coupling of the ferrous and ferric ions in cubane $[4Fe-4S]$ clusters into a particular spin state are not completely understood, a highly symmetrical environment surrounding an interpolypeptide $[4Fe-4S]$ cluster appears to be a common feature in proteins that fail to form an exclusive $S=1/2$ ground-spin state.

The F_A and F_B Clusters in Type I Reaction Centers

The terminal F_A and F_B acceptors in all type I reaction centers are low-potential $[4Fe-4S]^{2+,1+}$ clusters located on an extrinsic polypeptide. In PS I, the F_A cluster accepts an electron from F_X , and the F_B cluster donates the electron to soluble proteins such as ferredoxin and flavodoxin. The midpoint potentials of F_A and F_B are -530 mV and -580 mV, making

the $F_A \rightarrow F_B$ electron transfer step thermodynamically unfavorable. Although the presence of an uphill electron transfer step seems counterintuitive, the inverted midpoint potentials may play a role in preventing the formation of harmful oxygen-free radicals. Because the electron resides on the buried F_A cluster most of the time, the possibility that the exposed F_B cluster will transfer an electron by a collisional event with oxygen is minimized. In green-sulfur bacteria, which can survive only under low concentrations of oxygen, the F_A cluster is considered to be more reducing than F_B . Although this arrangement favors $F_A \rightarrow F_B$ electron transfer, the absence of oxygen renders the formation of harmful radicals moot. The precise midpoint potentials of the F_A and F_B clusters in the heliobacterial reaction center are yet to be measured.

The F_A and F_B clusters can be reduced either by the addition of a strong chemical reductant such as sodium hydrosulfite or by illuminating reaction centers at low temperature. In PS I, freezing a sample in the dark with subsequent illumination at 15 K promotes one electron from P_{700} to either F_A or F_B in a given PS I reaction center. Under these conditions, the resulting EPR spectrum is a superimposition of the resonances from F_A^- , with g -values of 2.05, 1.95, and 1.85 and F_B^- , with g -values of 2.09, 1.93, and 1.88. Freezing PS I complexes during illumination promotes multiple electron transfers, leading to the reduction of both F_A and F_B in a given reaction center. Here, the magnetic coupling between the reduced clusters results in a so-called interaction spectrum, which shows resonances at apparent g values of 2.05, 1.94, 1.92, and 1.88.

In the green-sulfur bacterial reaction center, the one-electron transfer illumination protocol leads to the preferential accumulation of F_B^- , and the two-electron illumination protocol results in additional accumulation of F_A^- . Because the midpoint potentials of the F_A and F_B clusters are not known with certainty, it is not yet possible to assign the EPR resonances in heliobacteria.

The F_A - and F_B -containing Proteins in Type I Reaction Centers

The PshB Polypeptide in the Heliobacterial Reaction Center

The identity of the protein that binds the terminal iron–sulfur clusters in heliobacteria remained elusive until it was discovered that the F_A/F_B -containing polypeptide, named PshB, was dissociated from the heliobacterial reaction center at relatively low ionic strength. The PshB protein was subsequently cloned and expressed in *Escherichia coli*, and the iron–sulfur clusters were chemically inserted *in vitro*. The recombinant PshB holoprotein was found to be capable of binding to the heliobacterial reaction center core and of accepting electrons from the F_X cluster.

The amino acid sequence of PshB contains two [4Fe–4S] cluster-binding motifs, one of which is a traditional CxxCxxCxxxCP motif and the other is a novel CxxCxxCxxxCC motif (Figure 5). It is not known whether the presence of a Cys residue instead of a Pro adjacent to the fourth Cys has functional or structural significance.

The PscB Polypeptide in the Green-Sulfur Bacterial Reaction Center

The PscB polypeptide in the green-sulfur bacterial reaction center is significantly different from its homologs, PshB and PsaC. First, the 23-kDa polypeptide has a unique, positively charged N-terminal domain, which is highly enriched in Lys, Pro, and Ala residues. A role in protein binding has been proposed for this domain, although there is no experimental evidence to support this conjecture. Second, although the C-terminal domain of PscB is similar to that of bacterial dicluster ferredoxins, it contains a novel CxxCxxCxxxCP Fe/S cluster-binding motif that ligates the putative F_A cluster (Figure 5). This motif

PscB	MAEPVENKNQAPAPGAKVPPKGAPAAPKAGAPAAPKGPVAPKAGAPAAKTGASAAKQAGK
PsaC	-----
PshB	-----
PscB	PRLASLGVTLGRSGVRQESALPYVKPKAVPPPKPAAPAAKGAPAPKAPKAPAAKAAAPGA
PsaC	-----
PshB	-----
PscB	PVAKAAPKAKKHYFIIENLCVGCGLCLDKCPPKVNAIGYKFYGDVQEGGFRCYIDQAACI
PsaC	-----MAHTVKIYDTCIGCTQCVRACP--TDVLEMVPWDGCKAGQIASSPRTEDCV
PshB	-----MAYKITDACTACGACMDGCC-----VGAIVEG--KKYSITSDCV
PscB	SCSACFSGDECPGAL-IEVLDPGEVLDFSYPPTPERLDFDLRFLHRFHREAR
PsaC	GCKRC--ETACPTDFLSIRVYLGAETTRSMGLAY-----
PshB	DCGVC--ADKCPVDAL-I---PG-----

Figure 5 Alignment of the amino acid sequences of PscB (green-sulfur bacteria), PsaC (cyanobacteria), and PshB (heliobacteria). The Cys residues that ligate the [4Fe–4S] clusters are indicated by arrows.

is not found in any other known dicluster ferredoxins. EPR and Mössbauer spectroscopic studies reveal that unlike recombinant PshB and PsaC, the two $[4Fe-4S]$ clusters cannot be properly inserted into recombinant PscB *in vitro*. Thus, PscB represents one of the few instances in which iron–sulfur clusters fail to become properly inserted into an expressed apo-ferredoxin.

The PscB polypeptide in the green-sulfur bacterial reaction center can be dissociated from the membrane core under conditions of moderate ionic strength. Thus, a loosely bound F_A/F_B -containing protein appears to be a common feature of homodimeric type I reaction centers. This finding indicates that in spite of the differences in the mass and composition of PscB and PshB, their mode of binding to the homodimeric type I reaction is likely to share more similarities than previously thought.

The PsaC Polypeptide in Photosystem I

PsaC is the best characterized of the F_A/F_B -containing proteins in type I reaction centers. The amino acid sequence of this 9-kDa polypeptide is typical of bacterial dicluster ferredoxins in that it contains two traditional CxxCxxCxxxCP cluster-ligating motifs and an amino acid sequence that is highly conserved in cyanobacteria, algae, and higher plants (Figure 5). The most similar residues are located at the PsaC–PsaA/PsaB binding interface, whereas the variable residues are usually located on the stromal face. Although PsaC is tightly bound to the membrane-associated PsaA/PsaB heterodimer, it is highly soluble in its unbound state. The ability to express large amounts of unbound PsaC via recombinant technology has allowed the determination of its 3D solution structure by nuclear magnetic resonance (NMR) spectroscopy. As expected, the iron–sulfur core of unbound PsaC shows the highest similarity to low-molecular-weight dicluster bacterial ferredoxins. The position, distance, and relative orientation of the two $[4Fe-4S]$ clusters as well as the conformation of two-cluster consensus binding sites are similar in PsaC and bacterial ferredoxins. In PsaC, the F_A cluster is ligated by Cys48, Cys51, Cys54, and Cys21, and the F_B cluster is ligated by Cys11, Cys14, Cys17, and Cys58. A fixed distance between F_A and F_B

is insured by the presence of a rigid one-turn α -helix between the third and fourth cysteine residues in each of the CxxCxxCxxxCP iron–sulfur cluster-binding motifs.

The presence of a high-resolution X-ray crystal structure of cyanobacterial PS I has provided structural details on the bound state of PsaC. PsaC is one of the very few membrane-associated proteins to have its structure solved in its unbound and membrane-bound states. The crystal structure of PS I-bound PsaC shows significant structural differences in the N- and the C-terminal regions when compared to unbound PsaC (Figure 6). The iron–sulfur core part of the protein largely remains unaltered, which is expected, given that clusters impart a high degree of rigidity to the central protein structure. The N-terminus, by contrast, is positioned closer to the F_A cluster and forms an anti-parallel β -sheet with the pre C-terminus. This conformation is significantly different than that observed in unbound PsaC wherein the N-terminus pushes the pre C-terminus away from the iron–sulfur core part of the protein. The C-terminus of PS I-bound PsaC assumes an extended conformation, which is strikingly different from the coiled-up structure encountered in unbound PsaC. The C-terminus is located in the binding interface of the PsaD and PsaA subunits; consequently, the high degree of structural flexibility of the C-terminus in unbound PsaC is altered to a fixed and extended form in the bound state of the protein. The uncoiling of the C-terminus is considered to play a crucial role during the orientation of PsaC on the PS I membrane core.

The structural differences between unbound and PS I-bound PsaC also have an effect on the midpoint potentials of the F_A and F_B clusters. In the unbound form, both the clusters have equivalent midpoint potentials of ~ -500 mV. On binding to membrane core, the midpoint potentials of F_A and F_B in PS I, respectively, become -530 mV and -580 mV.

The Tight Binding of PsaC with the PsaA/PsaB Heterodimer in Photosystem I

PsaC binds directly above the F_X -binding site, with the F_A cluster proximal to F_X . The binding site consists of an intricate network of ionic bonds (Figure 7) including one Lys and two

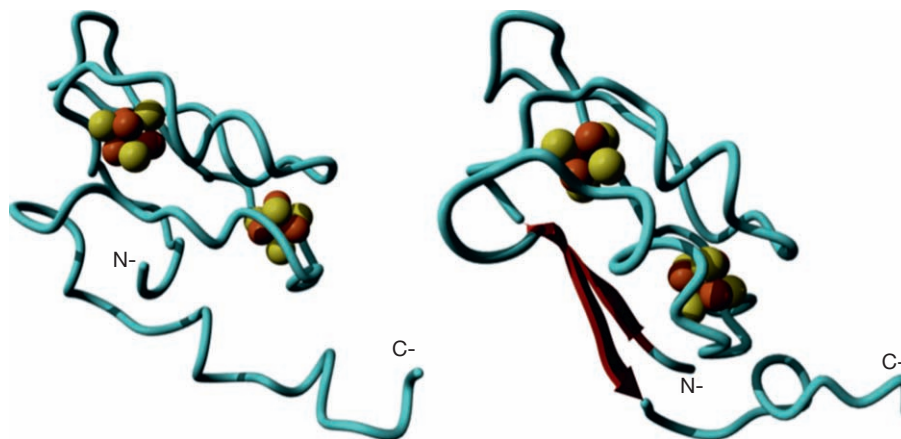


Figure 6 Structural differences between unbound PsaC (left, PDB entry 1K0T) and PS I-bound PsaC (right, PDB entry 1JB0). The N- and C-termini are indicated.

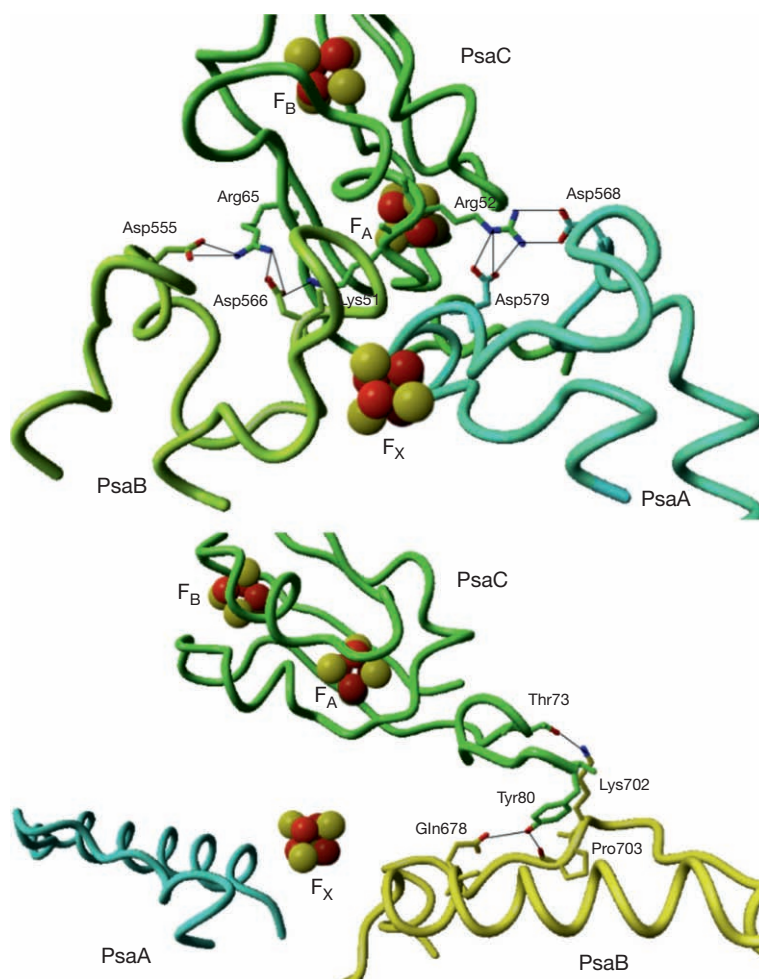


Figure 7 The network of ionic (*top*) and hydrogen bonds (*bottom*) between PsaC and the PsaA/PsaB heterodimer in photosystem I (PDB entry 1JB0).

Arg residues on PsaC, and four Asp residues on PsaA/PsaB, two of which are part of the extrinsic loop that provides the Cys residues that coordinates the F_X cluster. The network of Asp-Arg/Lys ionic contacts is highly C_2 -symmetric; R52_{PsaC} forms five ionic bonds with D568_{PsaA}/D579_{PsaA}, and K51_{PsaC}/R65_{PsaC} form five ionic bonds with D555_{PsaB}/D566_{PsaB}. If PsaC were rotated (*in silico*) 180° about the C_2 axis of symmetry, R52_C would still form five ionic bonds, but with D555_B/D566_B, and K51_{PsaC}/R65_{PsaC} would still form five ionic bonds, but with D568_{PsaA}/D579_{PsaA}. The equal division of the 10 ionic contacts between PsaA and PsaB represents a highly symmetric and extremely tight binding surface for the PsaC protein.

PsaC also forms three H-bonds (Figure 7) in a region of the P₇₀₀- F_X core distant from the ionic contacts, which breaks the otherwise perfect C_2 symmetry. These bonds are formed between T73_{PsaC}/Y80_{PsaC} on the C-terminus of PsaC and Q678_{PsaB}/K702_{PsaB}/P703_{PsaB} on a surface-located segment of PsaB. Unlike the symmetric ionic contacts, a symmetry-related binding pocket for the C-terminus of PsaC does not exist on the comparable surface-located segment of PsaA. Thus, the three H-bonds between the C-terminus of PsaC and PsaB break the binding symmetry, locking the PsaC protein into the correct orientation on the PsaA/PsaB heterodimer.

PsaC can be dissociated from PS I with high concentrations of chaotropes such as 6.8 M urea. This is in contrast to the reaction centers in heliobacteria and in green-sulfur bacteria wherein the F_A / F_B -protein can be dissociated with the use of 0.1–0.2 M NaCl. When released from the PsaA/PsaB heterodimer, the F_A and F_B clusters in unbound PsaC degrade due to the presence of oxygen, resulting in the loss of tertiary structure of the protein. A recent study in cyanobacteria showed that the extensive network of ionic and hydrogen bonds between bound PsaC and the PsaA/PsaB heterodimer are present to protect the F_A and F_B clusters against degradation by oxygen in organisms that contain PS II.

The terminal F_A and F_B clusters in the reaction centers of heliobacteria and green-sulfur bacteria do not require a comparable level of protection because the bacteria live in anoxic environments. By contrast, the F_A and F_B clusters in PS I have to function in a cell that is supersaturated with photosynthetically generated oxygen, and hence protection against oxidative degradation is required. It has been proposed that the extensive contacts between PsaC and the PS I membrane core developed as an evolutionary response to the onset of oxygenic photosynthesis ca. 3.5 billion years ago when the ancestor to PsaC may have been a loosely bound protein similar to PshB in heliobacteria.

See also: **Bioenergetics:** Green Bacteria: Secondary Electron Donor (Cytochromes); Iron–Sulfur Proteins; Photosystem I; **Metabolism Vitamins and Hormones:** Photosynthesis.

Further Reading

- Antonkine ML, Jordan P, Fromme P, et al. (2003) Assembly of protein subunits within the stromal ridge of Photosystem I. Structural changes between unbound and sequentially PS I-bound polypeptides and correlated changes of the magnetic properties of the terminal iron sulfur clusters. *Journal of Molecular Biology* 327: 671–697.
- Antonkine ML, Liu G, Bentrup D, et al. (2002) Solution structure of the unbound, oxidized Photosystem I subunit PsaC, containing [4Fe–4S] clusters F_A and F_B: A conformational change occurs upon binding to photosystem I. *Journal of Biological Inorganic Chemistry* 7: 461–472.
- Bentrup D, Capozzi F, and Luchinat C (2001) Iron–sulfur proteins. In: Bertini I, Sigel A, and Sigel H (eds.) *Handbook on Metalloproteins*, pp. 357–460. New York: Marcel Dekker.
- Chamorovsky SK and Cammack R (1982) Direct determination of the midpoint potential of the acceptor X in chloroplast photosystem I by electrochemical reduction and ESR spectroscopy. *Photobiochemistry and Photobiophysics* 4: 195–200.
- Golbeck JH (1994) Photosystem I in cyanobacteria. In: Bryant DA (ed.) *The Molecular Biology of Cyanobacteria*, pp. 179–220. The Netherlands: Kluwer Academic Publishers.
- Golbeck JH (2003) The binding of cofactors to photosystem I analyzed by spectroscopic and mutagenic methods. *Annual Review of Biophysics and Biomolecular Structure* 32: 237–256.
- Heinnickel M and Golbeck JH (2007) Heliobacterial photosynthesis. *Photosynthetic Research* 92: 35–53.
- Jagannathan B and Golbeck JH (2008) Unifying principles in homodimeric type I photosynthetic reaction centers: Properties of PscB and the F_A, F_B, and F_X iron–sulfur clusters in green sulfur bacteria. *Biochimica et Biophysica Acta* 1777: 1535–1544.
- Jagannathan B and Golbeck JH (2009) Breaking biological symmetry in membrane proteins: The asymmetrical orientation of PsaC on the pseudo-C₂ symmetric photosystem I membrane core. *Cellular and Molecular Life Sciences* 66: 1257–1270.
- Jagannathan B and Golbeck JH (2009) Understanding of the binding interface between PsaC and the PsaA/PsaB heterodimer in photosystem I. *Biochemistry* 48: 5405–5416.
- Jordan P, Fromme P, Witt HT, et al. (2001) Three dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* 411: 909–917.
- Li N, Zhao J, Warren PV, et al. (1991) PsaD is required for the stable binding of PsaC to the photosystem I core protein of *Synechococcus* sp. PCC 6301. *Biochemistry* 30: 7863–7872.
- Vassiliev IR, Antonkine ML, and Golbeck JH (2001) Iron–sulfur clusters in type I reaction centers. *Biochimica et Biophysica Acta* 1507: 139–160.

G

G₁₂/G₁₃ Family

S Offermanns, Max-Planck-Institute for Heart and Lung Research, Bad Nauheim, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

Actin stress fiber Bundles of actin microfilaments, which are found on the ventral side of cells cultured on artificial surfaces. They contain myosin and α -actin as well as a variety of other structural and regulatory proteins that allow them to contract and to exert tension. The ends of stress fibers terminate at specific sites at the plasma membrane (focal adhesions), which are involved in cell adhesion. Activation of RhoA has been shown to induce actin stress fiber formation.

Angiogenesis Growth of new blood vessels by sprouting from existing ones.

Cadherins Calcium-dependent adhesion proteins, characterized by the presence of cadherin repeats, which are present in the extracellular part of the protein. Cadherins mediate Ca²⁺-dependent homophilic adhesion. They are

divided into two subfamilies, the classic cadherins and the protocadherins. Classic cadherins are subdivided into four subfamilies, the type I classic cadherins, the type II classic cadherins, the desmosomal cadherins, and the so-called 'other classic cadherins'.

Guanine nucleotide exchange factors (GEFs) Proteins that catalyze the release of nucleotide bound to small guanosine triphosphatases (GTPases) such as RhoA. In most cases, GEFs bind to the guanosine diphosphate (GDP)-bound GTPase, causing dissociation of the GDP. GTP, which is present at higher concentrations in the cell than GDP, then binds to the GTPase, and the GEF is released.

RhoA Small GTP-binding protein (small GTPase). Being one of the best-studied members of the Rho-family of small GTPases, it has been shown to play an important role in the regulation of the actin cytoskeleton.

General Properties and Modifications

The α -subunits of G₁₂ and G₁₃, G α_{12} , and G α_{13} share 67% sequence identity and have a length of 377–380 amino acids. Biochemical analysis of purified G α_{12} and G α_{13} has revealed a relatively low rate of guanosine triphosphate (GTP) hydrolysis, as well as a relatively slow guanine nucleotide exchange rate. Their expression levels in most cells appear to be lower than those of other G-protein α -subunits. Both, G α_{12} and G α_{13} , undergo palmitoylation, a reversible posttranslational modification, which occurs at cysteine residues close to their N-termini. Palmitoylation localizes G₁₂ and G₁₃ to the inner side of the plasma membrane and is required for proper signaling via G₁₂/G₁₃. *In vitro* experiments have shown that purified G α_{12} is a substrate for protein kinase C, whereas G α_{13} can be phosphorylated by protein kinase A. The physiological significance of these phosphorylations is unclear.

Receptor-Mediated Activation of G₁₂ and G₁₃

It has been difficult to provide direct evidence for the coupling of a defined receptor to G₁₂/G₁₃. However, for a variety of GPCRs coupling to G₁₂/G₁₃ could be demonstrated using

different methods including the immunoprecipitation of receptor-activated G proteins, as well as genetic approaches. These studies revealed that some lipid mediators such as lysophospholipids or thromboxane A₂, various peptides such as endothelin, angiotensin II, or substance P as well as the protease thrombin act on receptors, which are able to couple to G₁₂/G₁₃ (Table 1). G₁₂ and G₁₃ appear to be activated by GPCRs, which also couple to G_q/G₁₁ and, in some cases, to G_i/G_o. However, not all G_q/G₁₁-coupled receptors are also able to activate G₁₂/G₁₃.

Cellular Functions of G₁₂/G₁₃

Most information on the cellular functions regulated via G₁₂/G₁₃ came from indirect experiments employing constitutively active mutants of G α_{12} /G α_{13} that lack the intrinsic GTPase activity due to an exchange of a glutamine residue (Q231 and Q226 in human G α_{12} and G α_{13} , respectively) to a leucine residue. Several laboratories could show that transfection of these mutants into different cell types leads to cellular transformation. Interestingly, human G α_{12} was first cloned in a screen for transforming oncogenes from an Ewing sarcoma complementary DNA library. In subsequent studies, the

Table 1 Some GPCRs shown to couple to G₁₂/G₁₃

Endogenous ligand(s)	Receptor	G-protein subclass(es)
Angiotensin II	AT ₁	G ₁₂ /13, G _q /11, G _{i/o}
Bradykinin	B ₂	G ₁₂ /13, G _q /11
Endothelin-1, -2	ET _A	G _q /11, G ₁₂ /13, G _s
Endothelin-1, -2, -3	ET _B	G ₁₂ /13???
GALP, galanin	GAL2	G _{i/o} , G _q /11, G ₁₂ /13
Serotonin	5-HT _{2C}	G ₁₂ /13, G _q /11
Lysophosphatidic acid	LPA ₁ , LPA ₂ , LPA ₄ , LPA ₅	G _i , G _q /11, G ₁₂ /13
Substance P	NK ₁	G ₁₂ /13, G _q /11
Thrombin and others	PAR-1 PAR-3/4	G ₁₂ /13, G _q /11, G _i
Spingosine-1-phosphate	S1P ₂ , S1P ₃ , S1P ₄ , S1P ₅	G ₁₂ /13, G _i , G _q /11G ₁₂ /13, G _i , G _q /11
Thromboxane A ₂	TP	G _q /11, G ₁₂ /13
Thyrotropin	TSH	G ₁₂ /13, G _s , G _q /11, G _i
Vasopressin	V1A	G ₁₂ /13, G _q /11

Summarized are findings from various laboratories, which are based on the immunoprecipitation of receptor activated G_α₁₂/G_α₁₃ using photoaffinity labeling or a GDP/[³⁵S]GTPγS exchange assay as well as on genetic evidence from G_α₁₂/G_α₁₃ deficient cells. Please note that this list is far from being complete.

expression of constitutively active mutants of G_α₁₂/G_α₁₃ was shown to induce a variety of signaling pathways leading to the activation of various downstream effectors including phospholipase A₂, Na⁺/H⁺ exchanger, or c-Jun N-terminal kinase.

Another important cellular function of G₁₂/G₁₃ is their ability to regulate the formation of actomyosin-based structures and to modulate their contractility by increasing the activity of the small GTPase RhoA. The RhoA-mediated formation of actin stress fibers in fibroblasts activated by various GPCR agonists is one of the best-described cellular paradigms for G₁₂/G₁₃-mediated Rho activation and subsequent rearrangement of the actin cytoskeleton. This phenomenon involves a Rho-induced bundling of actin filaments into stress fibers and the clustering of integrins and associated proteins into focal adhesion complexes. The ability of G_α₁₂/G_α₁₃ to induce actin stress fiber formation in fibroblasts was first described by the group of Gary Johnson. Regulation of actin-based structures via a G₁₂/G₁₃-mediated and Rho-dependent pathway has also been shown to occur in many other eukaryotic cells. For instance, in neuronal cells, activation of Rho through lysophosphatidic acid or thrombin receptors leads to the formation of contractile actomyosin filaments, eventually leading to the induction of neurite retraction, cell rounding, or axonal growth cone collapse. In vascular smooth muscle cells, the G₁₂/G₁₃-Rho-mediated pathway has been shown to contribute to the vasoconstrictor-induced actomyosin-based cell contraction, and the same pathway appears to be involved in the platelet shape change response.

Proteins Directly Interacting with G_α₁₂/G_α₁₃

The activation of Rho by exchange of guanosine diphosphate for GTP is catalyzed by guanine nucleotide exchange factors (GEFs), which, in turn, are regulated by various, mostly ill-defined mechanisms in the cell. First indications that a Rho-GEF protein is regulated by G₁₂/G₁₃ came from studies on early

gastrulation events in *Drosophila*. Genetic analysis indicated that DRhoGEF2, a *Drosophila* RhoGEF protein, functions downstream of the *Drosophila* G_α₁₂/G_α₁₃ ortholog, the *concertina* gene product. Subsequently, three mammalian RhoGEF proteins (p115RhoGEF, post synaptic density protein (PSD95), *Drosophila* disc large tumor suppressor (DlgA), and zonula occludens-1 protein (zo-1)-domain (PDZ)-RhoGEF, and leukemia-associated RhoGEF (LARG)) were described that are able to interact with activated G_α₁₂/G_α₁₃. This family of RhoGEF proteins is characterized by the presence of a 'regulator of G-protein signaling' (RGS)-domain, which mediates the interaction with the α-subunits of G₁₂/G₁₃. Although the RhoGEF activity of PDZ-RhoGEF and LARG appears to be activated by both, G_α₁₂ and G_α₁₃, p115 RhoGEF activity is stimulated only by G_α₁₃. In addition, the RGS domains of these RhoGEF proteins accelerate the intrinsic GTP hydrolysis activity of G_α₁₂/G_α₁₃ *in vitro* and, thus, function as 'GTPase-activating proteins' (GAPs). Although it is now well established that RhoGEF proteins, which contain an RGS-domain, interact with G_α₁₂/G_α₁₃, the precise mode of this interaction is still not completely clear. In addition, several groups have found evidence for an involvement of tyrosine kinases in the regulation of Rho via G₁₂/G₁₃. In any case, regulation of RhoGEF proteins by G_α₁₂/G_α₁₃ provided the missing link between G₁₂/G₁₃ and their ability to activate the small GTPase RhoA, which then mediates the effects of G₁₂/G₁₃ on the cytoskeleton and may also be involved in the transforming activity of these G proteins (Figure 1).

Active G_α₁₂ and G_α₁₃ interact with the cytoplasmic domain of some type I and type II class cadherins, such as E-cadherin, N-cadherin, or cadherin-14, causing the release of β-catenin from cadherins (Figure 1). This downregulates cadherin-mediated cell activation, and the release of β-catenin from cadherins may be a mechanism by which G_α₁₂/G_α₁₃ induces cellular transformation.

Various other proteins including Bruton's tyrosine kinase, the Ras GAP Gap1m, radixin, heat-shock protein 90, or protein phosphatase type 5 have been shown to interact with G_α₁₂ and/or G_α₁₃ (Figure 1). However, the biological significance of these interactions is not yet clear.

In Vivo Functions of G₁₂/G₁₃

The fact that G₁₂/G₁₃ appears to be the major mediators of the activation of RhoA via GPCRs suggests that they are involved in multiple biological processes including morphogenetic processes, regulation of cell shape, or regulation of cell movements. Data on *in vivo* functions of G₁₂/G₁₃ came initially from the analysis of phenotypical changes of null-mutants of these G-protein α-subunits in mice, flies, and worms. Experiments in *Drosophila* have provided genetic evidence for a physiological role of the G₁₂/G₁₃/RhoGEF-mediated pathway leading to Rho activation. By gene inactivation and epistasis experiments, it could be shown that a signaling pathway, consisting of the G₁₂/G₁₃ ortholog *concertina*, the *Drosophila* RhoGEF protein DRhoGEF2, as well as of Rho, is involved in the actin/myosin-mediated cellular shape change, which results in the ventral furrow formation of the *Drosophila* embryo. Interestingly, loss of DRhoGEF2 results in a more severe phenotype than observed in *concertina* mutant embryos, suggesting that the *concertina* gene product conveys only part of the upstream regulation of DRhoGEF2.

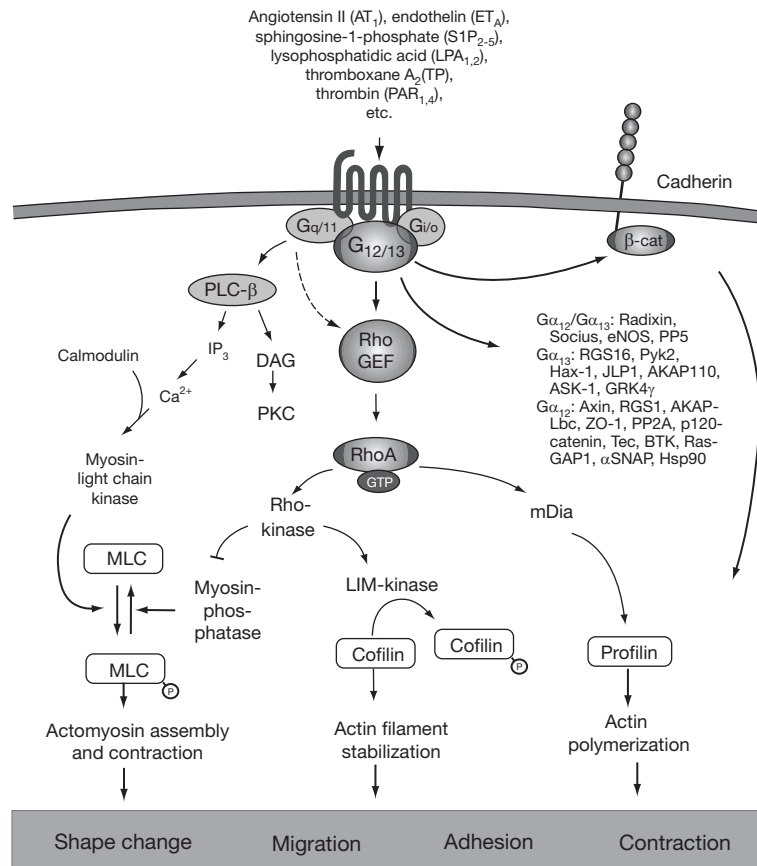


Figure 1 Major signaling pathways regulated by G₁₂/G₁₃. Most receptors coupling to G₁₂/G₁₃ also couple to G_q/G₁₁ and, in some cases, also to G_i-type G-proteins. AT₁, ET_A, S1P₂₋₅, LPA_{1,2}, TP, PAR_{1,4} are some of the receptor subtypes that have been shown to signal through G₁₂/G₁₃ family G-proteins. Abbreviations: PLC-β, β-isoform of phospholipase C; IP₃, inositol-1,4,5-trisphosphate; DAG, diacylglycerol; PKC, protein kinase C; MLC, myosin light chain; RhoGEF, Rho-specific guanine nucleotide exchange factor; GTP, guanosine triphosphate; β-cat, β-catenin; eNOS, endothelial nitric oxide synthase; PP, protein phosphatase; RGS, regulator of G-protein signaling; AKAP, A-kinase anchoring protein; Pyk, proline-rich tyrosine kinase; JLP, JNK-interacting leucine zipper protein; ASK, apoptosis signal-regulating kinase; GRK, G-protein-coupled receptor kinase; ZO, zonula occludens; BTK, Bruton's tyrosine kinase; RasGAP, Ras-specific GTPase-activating protein; SNAP, soluble NSF attachment protein; Hsp90, heat-shock protein 90.

Table 2 *In vivo* functions of mammalian G₁₂/G₁₃ as revealed by conditional mutagenesis in the indicated cell types

Cell type	Function of G ₁₂ /G ₁₃
Platelets	G ₁₃ mediates ligand-induced shape change, contributes to aggregation and degranulation, is required for primary hemostasis and arterial thrombosis
Vascular smooth muscle	G ₁₂ /G ₁₃ are critically involved in the development of salt-dependent hypertension; not required for maintenance of basal blood pressure
Endothelium	G ₁₃ function in endothelial cells is required for embryonic angiogenesis
Neurons	G ₁₂ /G ₁₃ regulates radial migration of cortical neurons
Neural crest	G ₁₂ /G ₁₃ regulates cardiac morphogenesis
B lymphocytes	Humoral responses are impaired in the absence of G ₁₂ /G ₁₃ either due to reduced numbers of splenic marginal zone B cells or due to inappropriate trafficking of MZB-cells during immune responses
T lymphocytes	Increased susceptibility toward T-cell-mediated diseases

Although mice lacking the Gα₁₂ gene (*gna12*) are phenotypically normal, deletion of the Gα₁₃ gene (*gna13*) results in embryonic lethality at embryonic day 9.5 due to a defect in angiogenesis. Chemokinetic effects of thrombin were completely abrogated in fibroblasts lacking Gα₁₃, indicating that

Gα₁₃ is required for full migratory responses of cells to certain stimuli. The defects observed in Gα₁₃-deficient embryos and cells occurred in the presence of Gα₁₂, and loss of Gα₁₂ did not result in any obvious defects. However, Gα₁₂-deficient mice, which carry only one intact Gα₁₃ allele, also die *in utero*, and

G α_{12} /G α_{13} double-deficient mice die even earlier than G α_{13} single-deficient animals. This genetic evidence indicates that G α_{13} and its closest relative, G α_{12} , fulfill at least partially nonoverlapping cellular and biological functions, which are required for proper mammalian development.

In recent years, mouse lines carrying conditional null alleles of the genes encoding G α_{12} and G α_{13} have been generated (Table 2).

The conditional inactivation of G α_{12} and G α_{13} in glial and neuronal cells results in the overmigration of cortical neurons in particular regions of the cerebral and cerebellar cortices. This eventually leads to neuronal ectopia of the cerebral cortex and malformations of the rostral part of the cerebellar cortex, suggesting a role of G₁₂/G₁₃ in the radial migration of cortical neurons. In platelets lacking G α_{13} , but not G α_{12} , severe defects in the responsiveness to various platelet activators were observed, resulting in increased bleeding times and a protection of animals from the formation of arterial thrombi in a carotid artery thrombosis model. Several lines of evidence indicate a role of G₁₂/G₁₃ in the regulation of vascular tone. Although smooth-muscle-specific G α_{12} /G α_{13} deficiency did not affect basal blood pressure, it protected mice from the development of salt-induced hypertension, indicating that the increase in vascular tone responsible for salt-dependent hypertension involves G α_{12} /G α_{13} -mediated signaling in vascular smooth muscle cells. G₁₂/G₁₃-mediated signaling has also been shown to play particular roles in various immune cells. B-lymphocyte-specific G α_{12} /G α_{13} deficiency results in altered trafficking of marginal zone B cells during immune responses with defects in cellular polarization and migration. Lack of G α_{12} /G α_{13} in T-lymphocytes leads to increased activity of leukocyte integrins and abnormal proliferative responses resulting eventually in an increased susceptibility toward T-cell-mediated diseases.

Although G α_{12} and G α_{13} have been shown to possess oncogenic potential (see above), the *in vivo* relevance of this finding is still unclear. Recent *in vivo* data indicate that blockade of G₁₂/G₁₃-mediated signaling in a murine breast cancer cell line

did not affect tumor growth but rather reduced the metastatic spreading in a xenograft model, strongly indicating a role of G₁₂/G₁₃-mediated signaling in tumor progression.

See also: **Bioenergetics:** P-Type Pumps: Na⁺/K⁺-ATPase; **Cell Architecture and Function:** Rho GTPases and Actin Cytoskeleton Dynamics; **Signaling:** Cadherin Signaling; G-Protein-Coupled Receptor Kinases and Arrestins; Small GTPases.

Further Reading

- Buhl AM, Johnson NL, Dhanasekaran N, and Johnson GL (1995) G α_{12} and G α_{13} stimulate Rho-dependent stress fiber formation and focal adhesion assembly. *Journal of Biological Chemistry* 270: 24631–24634.
- Fukuhara S, Chikumi H, and Gutkind JS (2001) RGS-containing RhoGEFs: The missing link between transforming G proteins and Rho? *Oncogene* 26: 1661–1668.
- Juneja J and Casey PJ (2009) Role of G12 proteins in oncogenesis and metastasis. *British Journal of Pharmacology* 158: 32–40.
- Kelly P, Casey PJ, and Meigs TE (2007) Biologic function of the G12 subfamily of heterotrimeric G proteins: Growth, migration and metastasis. *Biochemistry* 46: 6677–6687.
- Kozasa T, Jiang X, Hart MJ, et al. (1998) p115 RhoGEF, a GTPase activating protein for G α_{12} and G α_{13} . *Science* 280: 2109–2111.
- Meigs TE, Fields TA, McKee DD, and Casey PJ (2001) Interaction of G α_{12} and G α_{13} with the cytoplasmic domain of cadherin provides a mechanism for beta-catenin release. *Proceedings of the National Academy of Sciences of the United States of America* 98: 519–524.
- Offermanns S, Mancino V, Revel J-P, and Simon MI (1997) Vascular system defects and impaired cell chemokinesis as a result of G α_{13} deficiency. *Science* 275: 533–536.
- Riobo NA and Manning DR (2005) Receptors coupled to heterotrimeric G proteins of the G12 family. *Trends in Pharmacological Sciences* 26: 146–154.
- Singer WD, Miller RT, and Sternweis PC (1994) Purification and characterization of the α -subunit of G₁₃. *Journal of Biological Chemistry* 269: 19796–19802.
- Strathmann MP and Simon MI (1991) G α_{12} and G α_{13} subunits define a fourth class of G protein α -subunits. *Proceedings of the National Academy of Sciences of the United States of America* 88: 5582–5586.
- Suzuki N, Hajicek N, and Kozasa T (2009) Regulation and physiological functions of G12/13-mediated signaling pathways. *Neurosignals* 17: 55–70.
- Worzel T, Wettschreck N, and Offermanns S (2008) G(12)/G(13)-mediated signalling in mammalian physiology and disease. *Trends in Pharmacological Sciences* 29: 582–589.

GABA_A Receptor

R W Olsen and G-D Li, University of California, Los Angeles, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Richard W. Olsen and Gregory W. Sawyer, volume 2, pp. 162–166, © 2004, Elsevier Inc.

Glossary

Allosteric Literally, at another site; proteins have domains, or sites, that carry out their functions, termed 'active' sites, such as the binding site on a receptor for the neurotransmitter ligand. Other ligands bind at different, or allosteric, sites, to modulate the activity of the protein.

Ligand-gated ion-channel receptor Some receptors are coupled to effector systems in the cells, possibly a cascade of events, while other receptors are themselves ion channels, regulated by binding of the neurotransmitter ligand.

Neurosteroids Steroids that are found endogenously in the nervous system and have some functions there. Neuroactive steroids have some action in the nervous system but may originate there or have exogenous sources, elsewhere in the body, or be administered as drugs.

Receptor proteins Proteins that recognize the signal molecule, the neurotransmitter, bind it, and trigger a response are called receptors.

Transgenic mouse An animal that has been genetically engineered to express a foreign gene (a 'trans' gene) or a mutated version of its own gene (knockout or knockin).

γ -Aminobutyric acid (GABA) is the major inhibitory neurotransmitter, mediating synaptic and tonic inhibition at virtually all nerve cells in the central nervous system (CNS). This amino acid is synthesized in one step from L-glutamate by the enzyme glutamic acid decarboxylase (GAD). GAD is present only in GABAergic neurons, comprising up to 30% of those in the nervous system of all organisms, and is a marker for GABA synapses. Dysfunction of GABA, including GAD, is implicated in a variety of neuropsychiatric disorders, including epilepsy, anxiety, depression, and drug dependence.

Fast inhibitory synaptic transmission is mediated primarily by the GABA type A receptor (GABA_A receptor = GABAR), a ligand-gated chloride ion channel. Some slow inhibition is mediated by GABA at the GABA type B (GABA_B) receptor, a G-protein-coupled receptor. GABARs are members of a neurotransmitter receptor super-family that includes the nicotinic acetylcholine receptors, glycine receptors, and serotonin 5HT₃ receptors. The nicotinic acetylcholine receptors and the GABARs are families of heteropentameric isoforms, or receptor subtypes. These have different age-dependent and brain regional localization, circuit participation, and, thus, involvement in different functions and behaviors. Despite the lack of X-ray crystallographic data on the structure of GABAR, considerable information about functional domains within the protein is evolving. GABARs are the targets of important drugs, including general anesthetics and agents used to treat epileptic seizures, anxiety, and sleep disorders, such as the benzodiazepines (BZs), and also mediate some effects of alcohol at low doses. BZs are synthetic compounds with important clinical utility (neuroactive drugs). The BZ receptor, or site that mediates their pharmacological effects, is a portion of the GABAR protein itself; indeed, the binding site appears to be a modified version of the binding site for GABA (see below).

Molecular Structure

Heteropentamers

The members of the super-family share a pseudo-symmetric, pentameric membrane-spanning structure, with all subunits

contributing equally to the ion channel (**Figure 1**). Each member of the super-family is made from a family of homologous subunits of 45–67 kDa, with a conserved topology. Each of these subunits possesses a long extracellular N-terminus, which in some cases carries the neurotransmitter-binding site, four membrane-spanning domains, including the ion channel wall in M2, and a large variable sequence intracellular loop between M3 and M4. Although the structure of these complicated oligomeric receptor-channel membrane proteins has so far not been solved by crystallographic methods, some structural information has been obtained for other kinds of membrane receptors, ion channels, and, particularly, the extracellular domain of the nicotinic acetylcholine receptor, found to be exactly homologous to a snail acetylcholine-binding protein. Additional structural information on the acetylcholine receptor has been obtained using computer-assisted electron microscope imaging.

A Family of Isoforms

The heteropentameric GABAR protein is most often made from α , β , and γ subunits. The most common combination contains two copies of one α subunit, two copies of one β subunit, and one copy of γ 2, arranged as in **Figure 2**. The isoform most abundantly expressed in the CNS contains α 1 β 2 γ 2. Six α subunits define the major subtypes. Those containing α 2 or α 3 are less abundant, and α 4, α 5, and α 6-containing GABAR are even less abundant, with very restricted regional expression. Each α has a preferential but not obligatory β partner (β 1, β 2, β 3, or θ , similar to avian β 4), because some α 's can combine with different β subunits in different regions. Some GABAR isoforms appear to contain two different α subunits, but these are relatively nonabundant. Additional rare isoforms have subunits substituted for γ 2. For example, the α 4 and α 6 subunits can combine with γ 2 or with δ in cerebellar and dentate gyrus granule cells and thalamic relay cells. Rarely, γ 1, γ 3, or ϵ (related to avian γ 4) can substitute for γ 2 in pentamers containing certain $\alpha\beta$ combinations.

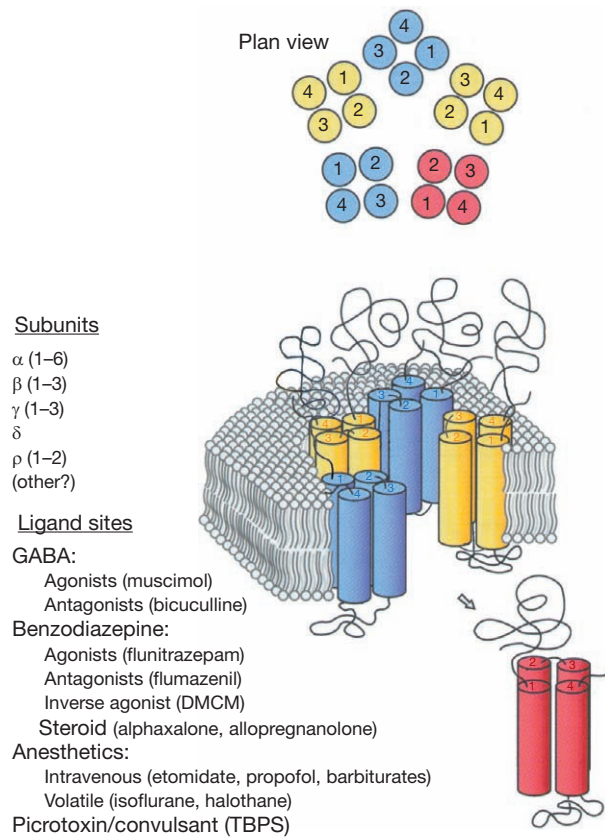


Figure 1 Schematic of GABA_A receptors. The protein is shown as a pseudo-symmetric membrane-spanning ion-channel protein made of five homologous subunits, each of which has four membrane-spanning regions as shown in the pull-out subunit. The view from outside the cell (plan view) shows the arrangement around the central core, the chloride ion channel. Also indicated are the subunit families that can be utilized in composing each receptor, and the ligand-binding sites present on the receptor.

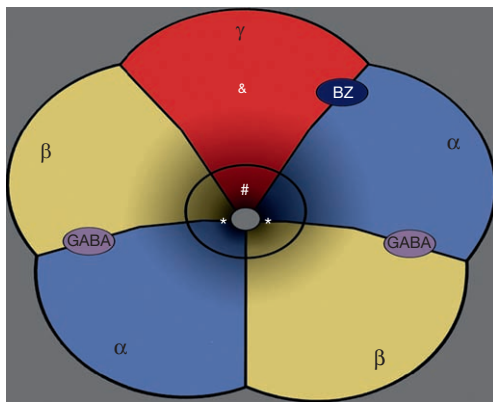


Figure 2 A three-dimensional-view donut model of GABA_A receptor heteropentamer with central pore and ligand-binding sites at subunit interfaces. The five subunits (β – γ – α – β – α) and five subunit interfaces are shown, each subunit having an extracellular domain (&) and a transmembrane domain (#). GABA and benzodiazepine (BZ) sites are indicated at β – α and α – γ subunit interfaces, respectively. The etomidate sites (*) are located in the transmembrane domain at β – α subunit interfaces, the same subunit interface as the GABA-binding site in the extracellular domain, around 50 Å away.

Anatomical Localization and Functional Heterogeneity

Each GABA_A subunit shows a specific anatomic and cellular localization as determined by assay of messenger RNA (mRNA) and polypeptides, indicating tissue-specific gene expression. Subunit-subunit associations and heteropentamer localization have been revealed by co-localization, co-purification, or co-immunoprecipitation studies. Furthermore, there is specific localization to subcellular sites on the plasma membrane, such as synaptic versus nonsynaptic regions and cell bodies versus axons and dendrites, including positioning at postsynaptic locations apposed to specific cellular inputs. For example, the abundant isoform $\alpha 1\beta 2\gamma 2$ is found primarily on inhibitory interneurons, thus inhibiting inhibition and producing a net excitatory circuit. Alternatively, many isoforms containing $\beta 3$ are partnered with $\alpha 2$, $\alpha 3$, and $\alpha 5$ subunits and localized on cells that use neurotransmitters other than GABA. These include principal excitatory cells, in which case GABA_ARs inhibit a local circuit. The $\alpha 1$ subunit appears to occur at any and all subcellular locations within a given cell, while the $\alpha 2$ is limited to axo-axonal synapses in some cells. Thus, the nature of the α subunit, or perhaps specific partnering, may play a role in localization. In addition, heterogeneity has been demonstrated in the developmental profile of individual subunits in brain regions and cells.

The heterogeneity in subunit composition for the GABA_A isoforms provides an explanation for pharmacological subtypes based primarily on ligand affinities, and secondarily on other factors such as sensitivity to regulatory systems. GABA_ARs in certain brain regions, with certain subunit compositions, vary in their sensitivity to modulatory drugs such as general anesthetics, including barbiturates, and BZs, as well as the biologically endogenous modulators, the neurosteroids. The neurosteroids are hormone metabolites that have rapid and direct effects on neurons, such as modulation of GABA_ARs. Metabolites of the female sex hormone progesterone (allopregnanolone), corticosteroids (tetrahydrocorticosterone), and, possibly, androgens enhance GABA_A currents and appear to reach effective concentrations during certain physiological and pathological conditions. This field of research is just in its infancy.

Insights Gained from Transgenic Mice

Interestingly, the specific anatomical location and circuitry of GABA_A subtypes dictate the specificity of functions and involvement in specific behaviors. Thus, transgenic and gene-targeted mice have demonstrated a role for certain subunits in particular functions and abnormalities, as well as pharmacology. For example, a mouse knockout for the $\gamma 2$ subunit reveals a role for this subunit in proper synaptic localization and clustering of GABA_ARs, and a role in anxiety, stress, and fear conditioning. Knockout of the $\beta 3$ subunit produces a severe phenotype with epilepsy, motor incoordination, hyperactivity, and cognitive defects. These observations suggest that decreased levels of $\beta 3$ subunit expression may contribute to Angelman syndrome, a human genetic disease with similar phenotype. Knockout of the common $\alpha 1$ and $\beta 2$ subunits has surprisingly little phenotype, perhaps due to compensatory increases in other subunits. Knockout of the $\alpha 6$ subunit reveals its obligatory partnership with the δ subunit in the cerebellum, because no δ peptide is expressed in surface receptors. Knockout or

knockdown of the $\alpha 5$ subunit yields an animal with improved spatial memory acquisition, implicating $\alpha 5$ -containing GABAR in the CA1 region in mechanisms that are inhibitory for learning. Genetic knockins of point mutations for drug-binding sites have revealed subunit and subtype functions such as the differential importance of the $\alpha 1$ subunit and its sites of expression in sedative/hypnotic functions, the $\alpha 2$ subunit in anxiety, and the $\beta 3$ subunit more so than $\beta 2$ subunit in the action of the intravenous anesthetic etomidate. Aberrant plasticity of GABAR in which a subunit switch occurs, such as $\alpha 1$ to $\alpha 4$, appears to contribute to drug withdrawal and seizure susceptibility. A summary of current (incomplete) knowledge of subtype subunit composition, localization, and function can be found in a recent review.

Subcellular Localization, Associated Proteins, and Synaptic Plasticity

The localization of GABAR likewise involves events other than gene expression, which include intracellular trafficking, surface membrane insertion, synaptic clustering, removal of GABAR from synapses, and recycling or degradation. For example, the nature of the β subunit seems to determine partially the subcellular location of GABAR. Regulation of these activities allows some plasticity of subunit composition of adult brain.

Several associated proteins have been identified in GABAR cell biology. These include especially gephyrin, originally isolated as a glycine receptor-associated synaptic clustering protein, which also appears to co-localize at synapses with some isoforms of GABAR. Likewise, GABARAP, a microtubule-associated protein, is involved in intracellular membrane vesicle trafficking and targeting of GABAR. GABARAP binds to the intracellular loop of the γ subunits but not others. As the $\gamma 2$ subunit is needed to direct GABAR to synapses, this interaction, as well as those with other proteins, may contribute to the process. Some aspects of channel function, allosteric modulation, and cell biology are regulated by posttranslational mechanisms including protein phosphorylation. Many such topics are still controversial.

Functional Domains

Assembly

Rules determining which pentamers are produced are built into the sequences of the subunits, primarily in the extracellular N-terminal domains. These residues involved in subunit partnering, assembly, and stoichiometry have been identified by subunit sequence scanning, site-directed chimeras, and mutagenesis. They involve homologous positions in different subunits and are located at regions of subunit interfaces (Figure 2). Thus, approximately two dozen isoforms are significantly present in the nervous system, out of the literally thousands of possible permutations of heteropentamers.

Ligand-Binding Sites

Although the presence of a γ subunit (generally $\gamma 2$) in the pentamer is required for BZ sensitivity, the nature of the α

subunit determines the BZ site selectivity. The δ subunit is selectively associated with the $\alpha 6$ subunit in cerebellar granule cells and with the $\alpha 4$ subunit elsewhere. Although isoforms containing $\alpha 4/6$ and $\gamma 2$ can bind BZ inverse agonists, $\alpha 4/6$ – δ isoforms are totally insensitive to traditional BZ, and they are excluded from synapses. The δ -containing isoforms are involved in producing tonic inhibition involving spillover of synaptic transmitter or other long-lasting pools of extrasynaptic GABA. These tonic currents are, therefore, affected by the activity of GABA transporters in nerve endings, postsynaptic cell bodies, and surrounding glia. The extrasynaptic δ receptors appear to have higher affinity for GABA than synaptic receptors, slower desensitization, and they are the main target of endogenous neurosteroid modulation, and certain exogenous modulatory drugs.

Inspection of sequences for the BZ-sensitive and insensitive α subunits suggested amino acid residues that are critical for BZ binding. The contribution of these residues was subsequently verified by site-directed mutagenesis and functional analysis by recombinant expression in heterologous cells. Additional information on ligand-binding site residues was obtained by microsequencing of peptides covalently attached by radioactive photoaffinity-labeled BZ ligand. Figure 2 summarizes results from the two approaches, indicating the general position of the BZ-binding site, one per pentamer; this is localized at the interface of α and γ subunits.

Similar approaches have identified residues involved in the agonist (GABA)-binding sites (two per pentamer), which are localized at the interface between β and α subunits (Figure 2). In fact, the residues in the β subunit that participate in the GABA-binding pocket are part of the BZ-binding pocket in the α subunit. This suggests that the binding site for BZ, a synthetic drug, is an evolutionary development of an agonist site. In addition, it is clear that the same residues in homologous sequences of other members of the receptor super-family are involved in the neurotransmitter-binding pocket, involving five to six small loops in the extracellular domain.

Ion Channel/Sites for Action of Allosteric Blockers and Enhancers

As in other members of the super-family, the second transmembrane domain M2 has been shown to form the ion channel in GABAR. Residues in this domain affect ion selectivity, gating, conductance, kinetics, and desensitization of the channel. This area is also involved in the action of picrotoxin and cage convulsants that functionally block the channel, as well as certain allosteric modulators such as general anesthetics that enhance GABA or activate the channel. Residues have been identified within the mouth of the ion channel in M2, as well as on the backside of the M2 helix, which are required for general anesthetic modulation of GABAR, suggesting the possibility of an anesthetic binding pocket, although this area may be involved only in allosteric coupling. Other residues near the extracellular ends of M3 and M1 have been identified as an actual binding site for the intravenous general anesthetic etomidate (two per pentamer) at the β/α interface (Figure 2) using photoaffinity labeling with an etomidate analog.

See also: **Bioenergetics:** Ligand-Operated Membrane Channels: GABA; **Signaling:** GABA_B Receptor.

Further Reading

- Belelli D and Lambert JJ (2005) Neurosteroids: Endogenous regulators of the GABA_A receptor. *Nature Reviews Neuroscience* 6: 565–575.
- Franks NP (2008) General anaesthesia: From molecular targets to neuronal pathways of sleep and arousal. *Nature Reviews Neuroscience* 9: 370–386.
- Li G-D, Chiara DC, Sawyer GW, et al. (2006) Identification of a GABA_A receptor anesthetic binding site at subunit interfaces by photolabeling with an etomidate analog. *Journal of Neuroscience* 26: 11599–11605.
- Macdonald RL and Kang J-Q (2009) Molecular pathology of genetic epilepsies associated with GABA_A receptor subunit mutations. *Epilepsy Currents* 9: 18–23.
- Mihic SJ, Ye Q, Wick MJ, et al. (1997) Sites of alcohol and volatile anaesthetic action on GABA_A and glycine receptors. *Nature* 389: 385–389.
- Mody I and Pearce RA (2004) Diversity of inhibitory neurotransmission through receptors. *Trends in Neurosciences* 27: 569–575.
- Moss SJ and Smart TG (2001) Constructing inhibitory synapses. *Nature Reviews Neuroscience* 2: 240–250.
- Olsen RW and Sieghart W (2008) International Union of Pharmacology LXX. Subtypes of γ -aminobutyric acid_A receptors: Classification on the basis of subunit composition, pharmacology, and function. Update. *Pharmacological Reviews* 60: 243–260.
- Rudolph U and Möhler H (2004) Analysis of GABA_A receptor function and dissection of the pharmacology of benzodiazepines and general anesthetics through mouse genetics. *Annual Review of Pharmacology and Toxicology* 44: 475–498.
- Wallner M, Hancher HJ, and Olsen RW (2006) Low dose acute alcohol effects on GABA_A receptor subtypes. *Pharmacology and Therapeutics* 112: 513–528.

GABA_B Receptor

S J Enna, University of Kansas Medical Center, Kansas City, KS, USA

N G Bowery, University of Birmingham Medical School, Birmingham, UK

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by S.J. Enna, volume 2, pp. 167–170, © 2004, Elsevier Inc.

Glossary

Allotsterism A change induced in the properties and function of a receptor by attachment of a ligand to a site other than that utilized by the endogenous substrate.

Heterodimer A complex consisting of two different proteins.

Ligand A drug, hormone, or neurotransmitter that attaches to a receptor.

Recognition site That portion of a receptor to which the neurotransmitter or hormone attaches.

GABA_B-Receptor Structure and Function

The γ -aminobutyric acid_B (GABA_B) receptor was first identified in 1981. Studies revealed that (*R,S*)- β -aminomethyl- β -(4-chlorophenyl)-propanoic acid (baclofen), a drug used to treat spasticity, and GABA inhibit depolarization-induced neurotransmitter release and compete for a specific binding site in brain tissue. As neither the functional response to baclofen nor its binding is inhibited by bicuculline, a competitive GABA_A-receptor antagonist, it was proposed that baclofen interacts with a different class of receptors, which were designated GABA_B sites. Subsequent work demonstrated further that GABA_A and GABA_B receptors are distinct entities. Thus, although baclofen, like GABA_A-receptor agonists, hyperpolarizes neurons, the baclofen-sensitive receptors are coupled to G proteins, unlike the ionotropic GABA_A site. Activation of the GABA_B receptor decreases Ca²⁺ and increases K⁺ membrane conductance, whereas the GABA_A receptor regulates chloride ion flux. Although baclofen inhibits forskolin-stimulated adenylyl cyclase activity in brain tissue, it enhances cyclic adenosine monophosphate (cAMP) production in brain slices if they are exposed simultaneously to agents, such as β -adrenoceptor agonists, that activate receptors coupled to G_s. Thus, depending on the conditions, GABA_B-receptor stimulation either enhances or inhibits cAMP accumulation, suggesting an effector system associated with both G_i and G_o proteins. Furthermore, GABA_B receptor stimulation induces nuclear accumulation of transcription factors suggesting genomic consequences to activation of this site.

Definitive proof that the GABA_B receptor is molecularly distinct from the GABA_A site came from expression-cloning experiments. These studies revealed that the GABA_B receptor is a heterodimer composed of two seven-transmembrane-spanning proteins (Figure 1). The first protein to be identified was designated GABA_{B(1)}, and the second GABA_{B(2)}. These subunits, which have 54% amino acid similarity and 35% sequence homology, are characterized by a large extracellular N-terminal domain. Heterodimerization, which is essential for receptor trafficking to the membrane as well as receptor function, occurs predominately through association of the intracellular α -helical portions of the C termini (Figure 1).

The orthosteric-binding site for GABA is located within a conserved Venus flytrap region of the GABA_{B(1)}, but not GABA_{B(2)}, subunit N-terminal domain, whereas the affiliated G proteins are coupled to the GABA_{B(2)} component (Figure 1). The GABA_B site is a Group 3 receptor, a category that includes vomeronasal organ, metabotropic glutamate (mGluR), and calcium-sensing receptors. Indeed, calcium is absolutely required for agonist binding to the GABA_B site. Although several splice variants have been identified for the GABA_{B(1)}-receptor subunit, not all are capable of forming functional receptors when coupled to GABA_{B(2)}. The most extensively studied GABA_B-receptor subunit combinations are GABA_{B(1a)}/GABA_{B(2)} and GABA_{B(1b)}/GABA_{B(2)}. A variety of studies, including those with GABA_B subunit deletion mutant mice, have demonstrated that heterodimerization of these two distinct gene products is absolutely essential for a functional receptor, with homodimers being biologically inert. There is crosstalk between GABA_B-receptor signaling and other G-protein-coupled sites, such as mGluRs.

GABA_B receptors are located pre- and postsynaptically throughout the central nervous system (CNS) and in some peripheral organs. Even though heterodimerization is obligatory for GABA_B-receptor function, the distribution of GABA_{B(1)} and GABA_{B(2)} expression differs among brain regions, as do drug- or physiologically induced changes in subunit expression. This suggests a differential regulation of these genes, pointing to the possibility that the protein products may serve functions other than formation of GABA_B receptors. For example, it has been reported that GABA_B-receptor subunits can directly associate with other proteins, including transcription factors.

The restricted number of GABA_B-receptor subunits, the 1:1 stoichiometry of the system, and the conserved nature of the GABA_{B(1)}-neurotransmitter recognition site, limits the possibilities for pharmacologically distinct GABA_B receptors. Although there is indirect evidence that certain GABA_B receptors may be more responsive to some agents than to others, there is no direct proof of molecularly or pharmacologically distinct sites. Because the neurotransmitter recognition site on the GABA_{B(1)} splice variants is identical, any difference in their responsiveness to a ligand is most likely due to an allosteric modification mediated by attachment of the

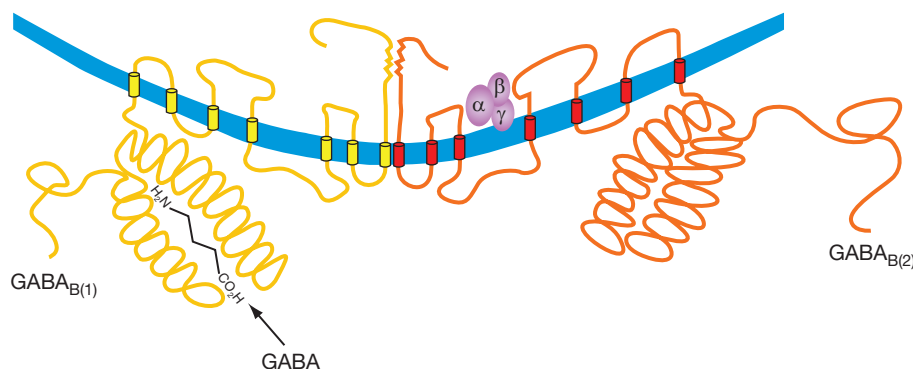


Figure 1 Structure and function of the GABA_B receptor. Attachment of GABA to the receptor recognition site located on the GABA_{B(1)} component causes dissociation of the α - and $\beta\gamma$ -subunits from the G-protein affiliated with the GABA_{B(2)} component of the receptor dimer. The liberated G-protein subunits influence cellular activity by regulating Ca^{2+} and K^{+} flux, adenylyl cyclase activity, and transcription factors. Reproduced from Enna S (2001) A GABA_B mystery: The search for pharmacologically distinct GABA_B receptors. *Molecular Interventions* 1: 208–218, with permission from American Society for Pharmacology and Experimental Therapeutics.

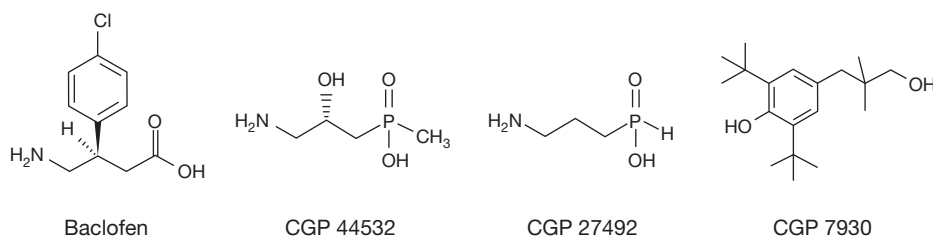


Figure 2 Chemical structures of some GABA_B-receptor agonists and a positive allosteric modulator.

agonist or antagonist to some portion of the receptor complex separate from the recognition site. The lack of evidence supporting pharmacologically distinct GABA_B-receptor subtypes has limited the clinical development of GABA_B-receptor agonists and antagonists because dose-limiting side effects accompany a generalized modification of GABA_B receptor function.

GABA_B-Receptor Agonists and Allosteric Modulators

Baclofen is the prototypical GABA_B-receptor agonist (Figure 2). In ligand-binding assays with rat brain membranes its affinity for the receptor is approximately 40 nM, with its EC_{50} values in functional assays being in the low- μM range. Baclofen has been employed for years as a treatment for spasticity associated with multiple sclerosis and spinal cord injuries, for nocturnal myoclonus, and for trigeminal neuralgia. It is the only GABA_B-receptor agent approved for clinical use.

A series of phosphinic acid analogs of GABA has yielded a number of potent and selective GABA_B-receptor agonists, including CGP 44532 and CGP 27492, which have been used for experimental purposes (Figure 2). Other compounds in this series include CGP 34938, the less-active (R)-(+)-enantiomer of CGP 34938, of which CGP 44532 is the more active (S)-(–) enantiomer, and SKF 97541, also known as CGP 35024. Rat brain membrane GABA_B-receptor-binding affinities for CGP 44532 and CGP 27492 are ~ 45 and 6.0 nM, respectively, with the corresponding values for CGP 44533 and SKF

97541 being 150 and 1.0 nM, respectively. Both tritiated baclofen and tritiated CGP 27492 are available as radioligands for labeling GABA_B receptors.

In vivo studies indicate that GABA_B agonists display sedative, muscle relaxant, and antinociceptive properties. They also impair memory performance, reduce the craving for cocaine, and can display both pro- and anticonvulsant activities. Tolerance develops to some of these responses upon repeated administration of agonist, and sedation precludes their use in many situations. Attempts to minimize these problems include modifications in the pharmacokinetic properties of baclofen preparations to enhance absorption and prolong duration of action.

As there are GABA_B receptors in the periphery, the therapeutic potential of GABA_B agonists is not limited to its actions in the CNS. For this reason, agents that do not cross the blood–brain barrier are being developed to treat conditions such as gastroesophageal reflux disease and pancreatic cancer.

Another approach to activating GABA_B receptors is with selective allosteric modulators, as exemplified by CGP 7930 (Figure 2). Modulators differ from conventional agonists in that they do not directly interact at the GABA_B-receptor-recognition site, but rather attach to some other portion of the receptor complex causing an allosteric modification in the neurotransmitter-binding site, or enhancing the efficiency of the receptor coupling with the effector system. Allosteric modulators, such as CGP 7930, GS 39738, and BHF177, increase both receptor affinity and the maximal response to the neurotransmitter. Certain arylalkylamines, some L-amino

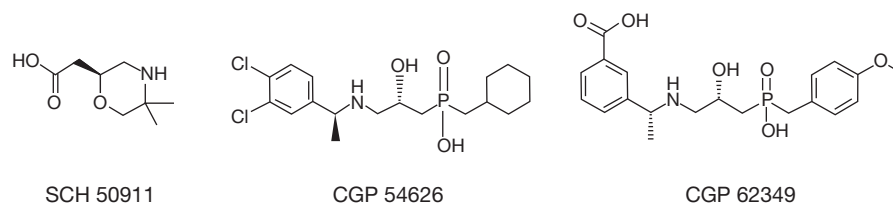


Figure 3 Chemical structures of some GABA_B-receptor antagonists.

acids, and selected dipeptides have been proposed as GABA_B-receptor allosteric modulators, although these claims have been disputed. Nonetheless, because allosteric modulators indirectly influence receptor sensitivity by acting at a site distinct from that utilized by the neurotransmitter, such agents may be capable of distinguishing subtypes of GABA_B receptors, making it possible to more selectively modify this system for therapeutic gain.

GABA_B-Receptor Antagonists

The absence of selective receptor antagonists initially hindered the pharmacological characterization of the GABA_B site. The first antagonists developed for this purpose were phaclofen, a phosphonic acid analog of GABA, and saclofen, a sulfonic acid analog. Although both are highly selective as GABA_B-receptor antagonists, their affinities for this site are quite low, being greater than 100 μ M. Subsequently, a series of selective, high-affinity, and, in some cases, systemically active phosphinic acid derivatives were developed as GABA_B-receptor antagonists. Included in this group are CGP 54626 and CGP 62349 (Figure 3). Both of these compounds display K_d values at the GABA_B-binding site of 10 nM or less, and both have been radiolabeled for use as ligands for studying the localization, molecular properties, and pharmacological selectivity of this receptor. Other notable phosphinic acid GABA_B-receptor antagonists are CGP 36742, CGP 35348, CGP 55845, CGP 46381, and CGP 51176.

Morpholine derivatives have also been found to display GABA_B-receptor antagonist properties. This is exemplified by SCH 50911, a systemically active agent with a K_i of $\sim$ 300 nM at the GABA_B receptor (Figure 3).

Although no GABA_B-receptor antagonists have been approved for clinical use, laboratory animal studies suggest they may be effective in treating absence epilepsy, cognitive impairments, and affective illness. Clinical trials have been undertaken to examine their efficacy as treatments for cognitive impairments and depression.

See also: [Bioenergetics: Ligand-Operated Membrane Channels: GABA](#); [Signaling: GABA_A Receptor](#).

Further Reading

- Alstermark C, Amin K, Dinn S, et al. (2008) Synthesis and pharmacological evaluation of novel gamma-aminobutyric acid type B (GABAB) receptor agonists as gastroesophageal reflux inhibitors. *Journal of Medicinal Chemistry* 51: 4315–4320.
- Bowery N, Bettler B, Froestl W, et al. (2002) Mammalian γ -aminobutyric acid B receptors: Structure and function. *Pharmacological Reviews* 54: 247–264.
- Bowery N and Enna S (2000) γ -Aminobutyric acid_B receptors: First of the functional metabotropic heterodimers. *Journal of Pharmacology and Experimental Therapeutics* 292: 2–7.
- Bythyn D, Kuo S, Shue H, and McPhail A (1996) Substituted morpholine-2S-acetic acid derivatives: SCH 50911 and related compounds. *Bioorganic and Medicinal Chemistry Letters* 6: 1529–1534.
- Enna S (2001) A GABA_B mystery: The search for pharmacologically distinct GABA_B receptors. *Molecular Interventions* 1: 208–218.
- Froestl W, Mickel S, Hall R, et al. (1995) Phosphinic acid analogues of GABA. 1. New potent and selective GABA agonists. *Journal of Medicinal Chemistry* 38: 3297–3312.
- Froestl W, Mickel S, von Sprecher G, et al. (1995) Phosphinic acid analogues of GABA. 2. Selective, orally active GABA_B antagonists. *Journal of Medicinal Chemistry* 38: 3313–3331.
- Hill D and Bowery N (1981) ³H-Baclofen and ³H-GABA bind to bicuculline-insensitive GABA_B sites in rat brain. *Nature* 290: 149–152.
- Kaupmann K, Malitschek B, Schuler V, et al. (1998) GABA_B-receptor subtypes assemble into functional heteromeric complexes. *Nature* 396: 683–687.
- Kerr D and Ong J (2003) Potentiation of metabotropic GABA_B receptors by L-amino acids and dipeptides in rat neocortex. *European Journal of Pharmacology* 468: 103–108.
- Lal R, Sukbunthorn J, Tai E, et al. (2009) Arbaclofen placarbil, a novel R-baclofen prodrug: Improved absorption, distribution, metabolism, and elimination properties compared to R-baclofen. *Journal of Pharmacology and Experimental Therapeutics* 330: 911–921.
- Prezeau L, Rives M, Comps-Agar L, et al. (2010) Functional crosstalk between GPCRs: With and without oligomerization. *Current Opinion in Pharmacology* 10: 6–13.
- Sands S, McCarron K, and Enna S (2003) Differential regulation of GABA_B receptor subunit expression and function. *Journal of Pharmacology and Experimental Therapeutics* 305: 191–196.
- Schuller H, Al-Wadei H, and Majidi M (2008) GABA B receptor is a novel drug target for pancreatic cancer. *Cancer* 112: 767–778.
- Urwiler S, Pozza M, Lingenhoehl K, et al. (2003) N,N'-Dicyclopentyl-2-methylsulfany-5-nitor-pyrimidine-4,6-diamine (GS39783) and structurally related compounds: Novel allosteric enhancers of γ -aminobutyric acid_B receptor function. *Journal of Pharmacology and Experimental Therapeutics* 307: 322–330.

Gastrointestinal Digestion and Absorption

J Keller, Israelitic Hospital, Hamburg, Germany

© 2013 Elsevier Inc. All rights reserved.

Introduction

Digestion and absorption of nutrients within the gastrointestinal tract require a complex interaction between secretory, motor, and absorptive functions, on the one hand, and between physical and biochemical events, on the other hand. Secretion of digestive enzymes starts in the oral cavity with salivary amylase, and nutrient digestion is continued in the stomach, mainly by the actions of gastric acid and pepsins. The pancreas delivers the most important and effective enzymes for macronutrient digestion. However, the final breakdown of oligosaccharides and oligopeptides takes place at the level of the brush border. While some drugs and potential constituents of human food, for example, urea and ethanol, may already be absorbed via the oral or gastric mucosa, the vast majority of nutrients are absorbed in the small intestine. Although nutrient digestion and absorption are extremely efficient in healthy humans, minor amounts are physiologically malabsorbed. These exert regulatory, mainly inhibitory effects on upper gastrointestinal secretory and motor activities and serve to nourish the colonic flora. The main function of the colon is absorption of water and electrolytes and under physiological circumstances the colon contributes little to overall calorie uptake. However, in case of carbohydrate malabsorption, colonic absorption of short-chain fatty acids that result from bacterial carbohydrate metabolism may salvage substantial amounts of nutrients that would otherwise be excreted with feces.

While nutrients are digested by secreted or membrane-bound enzymes and absorbed by the gastrointestinal mucosa, coordinated gastrointestinal motility is also one of the basic requirements for undisturbed and efficient digestion and absorption of a meal as well as excretion of indigestible residuals. Accordingly, milder motility disturbances can induce or contribute to symptoms such as diarrhea, constipation, and abdominal pain, severe motility disturbances may also impair nutrient absorption.

Digestive and absorptive mechanisms and in particular the regulation of the highly coordinated events that are needed for efficient digestion and absorption of meal nutrients are extremely complex. Thus, given the limitations of this article, it is impossible to give a detailed report of contemporary knowledge of each step of gastrointestinal digestion and absorption. Likewise, absorptive mechanisms responsible for micronutrient absorption, for example, the complex regulation of iron absorption and absorption of inorganic food components such as water and electrolytes will not be described in detail. Instead, this chapter will summarize what is currently known about the luminal and mucosal processes required for nutrient absorption with only brief reference to regulatory mechanisms and predominant limitation to human data. The article is organized in a way that the reader can follow the transit of a meal through the digestive tract. It serves to prepare the ground for the understanding and classification of the

information on individual nutrient components, the modes of action of specific enzymes, or dedicated metabolic pathways as described in other articles of this encyclopedia.

Mechanical and Enzymatic Breakdown of Nutrients in the Oral Cavity

In the oral cavity, ingested nutrients are fragmented by mastication and are mixed with saliva to form a food bolus that can easily be swallowed. The relevance of mastication for nutrition is illustrated by reduced appetite, weight loss, and increased mortality observed in people with chewing problems, especially edentulous subjects without dentures. The masticatory process appears to be crucial for optimal absorption of nutrients from some essential food components such as meat and vegetables but not from softer foods such as bread, cheese, rice, fish, and egg.

In addition to mechanical breakdown of food components, enzymatic digestion by salivary enzymes also starts within the oral cavity. The most important enzyme of saliva is α -amylase (molecular weight 54–57 kDa depending on the degree of glycosylation), which breaks carbohydrates (starches) down to maltoses by cleaving the α -1–4 glycosidic bonds. However, salivary α -amylase is rapidly inactivated at acidic pH in the stomach and therefore its contribution to carbohydrate digestion in healthy individuals is limited. Throughout the digestive period, salivary amylase comprises about 15% of total amylase output, the rest is attributable to pancreatic α -amylase. Autonomic nerve stimulation induces secretion of salivary α -amylase mainly from the parotid gland and to a lesser extent from the submandibular gland. In saliva, specifically collected from the parotid gland α -amylase may account for up to about one-third of the total protein content, while in mixed saliva the content is much lower. Since production and release of salivary α -amylase are controlled by the autonomic nervous system, there has been renewed interest in this enzyme as a surrogate marker of autonomic/sympathetic activity.

There is conflicting data whether in humans relevant amounts of lipolytic activity are secreted into saliva. While some researchers describe lipolytic activity in human tongue preparations, particularly in homogenates of the glandular region (Ebner) and in secretions collected from the trough of the papillae, this could not be confirmed by others. Anyway, there appear to be marked species differences with high activities of lingual lipase in homogenates of lingual serous glands from mouse and rat and only trace amounts in humans. Considering these levels of enzyme activity, it appears questionable whether lingual lipase has a physiological role in humans, at all.

While several drugs and some potential constituents of human food (e.g., urea) may be absorbed through the oral mucosa, oral absorption plays no relevant role for overall nutrient absorption in humans.

Esophageal Transport

The esophagus has no direct digestive function. The esophageal mucosa secretes low amounts of mucus that contains no enzymes but serves to facilitate the bolus transport and to protect the esophageal mucosa against acid and other aggressive components of gastroesophageal refluxate such as pepsin and potentially pancreatic enzymes. The main function of the esophagus is to transport ingested material to the stomach. During swallowing, coordinated activities of lingual and pharyngeal muscles transport the bolus toward the upper esophageal sphincter whereas the nasopharynx and larynx need to be occluded. The upper esophageal sphincter relaxes briefly in order to allow entrance of the bolus into the esophagus. Within seconds, it closes again and the bolus is then propelled downward through the tubular part of the esophagus which is formed by striated muscles in the upper third and by smooth muscles in the lower parts. At each time during the peristaltic wave, a segment several centimeters in length is contracted. The food bolus is always pushed several centimeters ahead of the zone of maximal contractility that is located in the middle of the contracted segment. In the distal part of the tubular esophagus, intraluminal pressure may rise physiologically to up to 180 mmHg. Normal propagation velocity of the peristaltic wave ranges from 2 to 8 cm s⁻¹. The lower esophageal sphincter, which normally maintains a resting pressure of about 15–30 mmHg that serves to prevent reflux of gastric contents, relaxes before the peristaltic wave arrives and, thus, lets the bolus pass into the stomach.

Gastric Functions

The stomach has important motor, secretory, and digestive functions, while gastric absorption of food components only plays a marginal role.

Gastric Transit of a Meal

Food ingestion induces a vagally mediated reflex which relaxes the proximal stomach; arrival of food into the stomach enhances this process of accommodation so that a meal can be ingested without inducing increased intragastric pressure or tension that would cause nausea and discomfort. During the early postprandial period, solid food is retained in the proximal part of the stomach. By contrast, liquids are distributed more widely within all of the stomach. The pylorus maintains only a small opening area throughout the digestive period. This allows rapid gastric emptying of liquids. However, with increasing calorie content of the liquid phase of a meal, there is a deceleration from an exponential to a more linear-emptying pattern induced by inhibitory duodenal feedback mechanisms. The release of cholecystokinin from duodenal enterochromaffin cells likely plays a major role not only for prandial inhibition and regulation of gastric emptying but also for integration of digestive upper gastrointestinal functions (Figure 1).

Solids are retained selectively and undergo trituration by antral contractions as they are propelled toward the almost closed pylorus. Acid and peptic digestion of meal components supports this process. Digestible food particles are emptied into the duodenum after their size has been reduced to about 2 mm. Under physiological circumstances, gastric emptying of nutrients takes place at a rate of 2–3 kcal min⁻¹. In consequence, complete gastric emptying of a regular meal approximately takes 3–4 h. Food components that cannot be reduced to small pieces remain within the stomach and are emptied into the duodenum at the very end of the digestive period when the fasting motility pattern (phase III-motility) recurs.

Gastric Secretory and Digestive Functions

Gastric motility does not only serve to grind solids and to transport gastric contents toward the duodenum but also

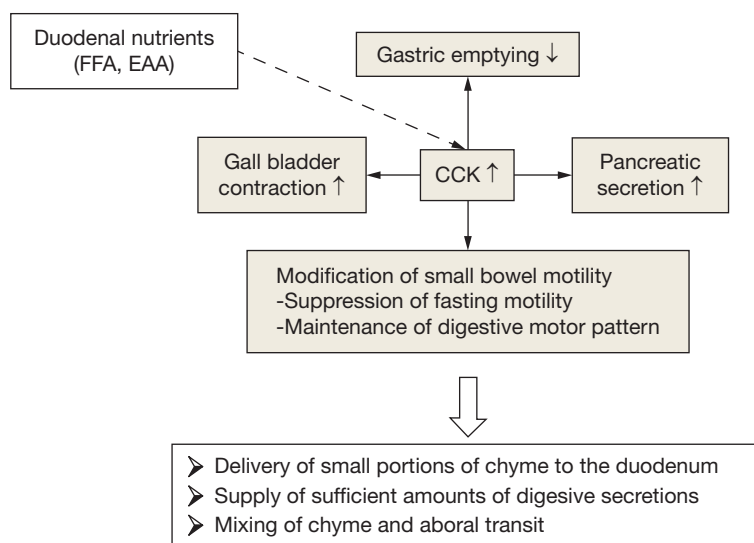


Figure 1 Nutrients within the duodenal lumen, in particular, free fatty acids (FFA) and specific essential amino acids (EAA), release cholecystokinin (CCK) from enterochromaffin cells of the mucosa. CCK exerts inhibitory effects on gastric emptying and at the same time stimulates pancreatic enzyme secretion, gall bladder contraction, and digestive small intestinal motility. Accordingly, gastric chyme is delivered in small portions to the duodenum with simultaneous supply of large amounts of digestive enzymes and biliary secretions and extensive mixing of chyme and slow aboral transit.

mixes the food with gastric secretions, and, thus, improves digestion.

Digestive role of gastric secretions

Gastric acid secretion and intragastric pH

Hydrochloric acid is the main component of gastric juice and is secreted by the parietal cells of the gastric mucosa in the fundus and corpus. In healthy adults, intragastric pH ranges between 1.5 and 2.5 in the fasting state. Intragastric nutrients are the most potent stimulators of gastric acid secretion; however, due to dilution and buffering effects there is usually a transient rise in gastric pH postprandially to values between 4.5 and 6.0. Extent and duration of the rise in pH depend not only on meal properties but also on gastric-emptying velocity. Mean postprandial hydrochloric acid output is about 20–30 mmol h⁻¹.

Gastric protein digestion

Gastric secretions facilitate the digestion of protein which commences in the stomach by acid denaturation and hydrolysis by gastric proteases or pepsins. These enzymes are secreted by gastric mucosal cells in precursor form (pepsinogens) and are activated by exposure to acid and autocatalysis. Pepsins are endopeptidases that cleave internal peptide bonds, in particular those adjacent to hydrophobic amino acids.

Gastric acid secretion and absorption of micronutrients

While gastric secretions contribute to protein digestion, gastric acidity improves iron absorption. Gastric acid specifically facilitates the dissociation of iron salts from food. Gastric acid also reduces ferric iron to the more soluble ferrous form and allows formation of complexes with ascorbate, sugars, and amines. These further enhance iron absorption in the duodenum.

Moreover, absorption of minerals in the small intestine usually requires soluble molecules. Gastric acid can release minerals, almost exclusively cations, from organic food matrices. Since solubility of positively charged ions is higher in more acidic solutions, gastric acid improves absorption of minerals. In particular, gastric acid transforms the relatively insoluble calcium salts (e.g., calcium carbonate) to more soluble forms and thus facilitates calcium absorption.

Vitamin B 12 ingested with meals is bound to food proteins. The vitamin needs to be released from these proteins and then bound to R-proteins and intrinsic factor in order to be absorbed in the terminal ileum. Gastric acid and pepsin facilitate the proteolytic process involved in releasing the vitamin from the proteins in ingested food. Accordingly, short-term treatment of healthy subjects with an acid-suppressive drug (proton pump inhibitor) was associated with a more than 70% reduction of vitamin B 12 absorption. Long-term studies in patients have shown that vitamin B 12 concentrations remain within the normal limit within the first 3 years of treatment, whereas a longer duration of therapy was associated with a small but significant downward trend.

Impairment of vitamin C absorption in case of decreased gastric acid secretion is less well established. However, it has been shown that treatment with proton pump inhibitors may produce a slight reduction in the serum concentration of vitamin C. This was attributed to the relative instability of vitamin C at higher pH and luminal metabolism to 2,3-diketogulonic acid which cannot be converted back to bioactive vitamin C.

Gastric lipase

Pepsins are not the only enzymes that are produced by the gastric mucosa in humans. In addition, gastric lipase is secreted by the chief cells of the fundus in response to neurohormonal stimuli elicited by food intake, in particular, gastrin and cholinergic mechanisms, and inhibited by cholecystokinin and glucagon-like peptide-1. It has been estimated that 10–20% of meal triglycerides may be hydrolyzed by gastric lipase, with diacylglycerol and free fatty acids as the main metabolites. This mechanism is essential for digestion of human milk fat globules in newborn infants. By contrast, in healthy adults, the physiological role of gastric lipase may be rather regulatory than quantitatively contributing to overall lipid digestion: enhanced intragastric emulsification and predigestion of dietary fat results in early postprandial duodenal delivery of free fatty acids as the most potent stimulators of postprandial pancreatic enzyme secretion.

Digestion and Absorption of Nutrients in the Small Intestine

Intestinal Transit of Meal Nutrients

The transportation rate of the small intestine is comparable to that of the stomach. Mainly due to faster gastric emptying, liquids typically arrive in the colon before solids and the head of the column of liquid chyme may pass the small intestine and reach the cecum within 30 min after oral ingestion. However, it takes much longer, namely about 2.5 h for half the liquid chyme to traverse the small bowel, and after grinding of solid components, intestinal transit time appears to be similar for solid and liquid food components.

The proximal small intestine absorbs nutrients extremely efficiently: Following perfusion of nutrients into the proximal duodenum at physiological postprandial rates, about 80% of triglycerides, 60% of carbohydrates, and 50% of proteins have been shown to be absorbed before reaching the distal duodenum 20 cm apart. However, even in healthy humans small amounts of nutrients are not absorbed during small intestinal transit but reach the terminal ileum and are delivered to the colon. The extent of this so-called physiological malabsorption depends on type and preparation of a meal. With the exception of starch from rice, most complex carbohydrates result in malabsorption rates of about 10%. Lipid malabsorption may reach 5% of the amount administered. Physiologically malabsorbed nutrients do not only serve as energy source for colonic bacteria but also exert regulatory, mainly inhibitory effects on upper gastrointestinal functions (ileal brake mechanism) and, thus, likely contribute to restitution of the interdigestive secretory and motor pattern at the end of the digestive period. Glucagon-like peptide-1 and peptide YY are important neurohormonal mediators implicated in the ileal brake mechanism in humans.

Irregular, partly segmental, partly propagated contractions of the small intestinal wall induce slow aboral transport of chyme and are also responsible for extensive mixing of nutrients with digestive secretions and distribution of chyme over the mucosal surface. Intestinal motility is based on contractions of intestinal smooth muscle cells. However, coordinated motility also requires a complex regulation by neurohormonal mediators, particularly via the enteric nervous system, the

brain–gut axis, and gastrointestinal peptide hormones. These regulatory mechanisms provide the coordination of smooth muscle cells among themselves, with secretory and absorptive functions of the gastrointestinal tract and with the superordinate requirements of the body.

Intestinal pH

The chyme that reaches the duodenal bulb in small portions usually has an acidic pH. Both, the pancreatic duct and the common bile duct lead into the proximal duodenum at the Papilla of Vater, and, mainly due to bicarbonate secretion from the pancreas and to a far lesser extent from biliary secretions or from the intestinal mucosa itself, the pH in the descending duodenum is maintained at a considerably higher level. The pH of aspirates from the distal duodenum of healthy subjects is 6–7 in the fasting state. Following ingestion of a normal meal, duodenal pH is around 6 early postprandially, drops toward 5–5.5 during the second and third postprandial hour, and reaches preprandial values at the end of the digestive period. The high duodenal pH is a prerequisite for efficient digestion of complex carbohydrates, proteins, and lipids because all pancreatic enzymes have their pH optima in the alkaline range. Lipase in particular is even irreversibly destroyed at pH levels below 4.

Due to impaired accessibility of the more distal parts of the human small intestine, pH profiles of the jejunum and the ileum are less well established. However, it has been shown that pH increases toward pH 7–8 and is more stable in the distal small intestine.

Pancreatic Exocrine Secretion

Sufficient absorption of macronutrients to maintain energy supply depends on prior hydrolysis by pancreatic enzymes. Thus, untreated complete lack of pancreatic enzymes (e.g., due to pancreatectomy) is not compatible with life. Normally, however, in healthy individuals, the cumulative postprandial pancreatic enzyme response exceeds by far (10–15-fold) the quantity required to prevent overt maldigestion. Depending on nutrient intake, about 3 l of pancreatic juice are secreted into the duodenum per day. The most important inorganic components of pancreatic juice apart from bicarbonate are water, potassium, and chloride. Moreover, pancreatic juice contains high amounts of hydrolytic enzymes or their precursors for digestion of complex carbohydrates, proteins, and lipids. The most important ones are α -amylase, trypsin and chymotrypsin, lipase, and colipase. Furthermore, elastases, carboxypeptidases, phospholipase A₂, carboxylesterase lipase, and ribo- and deoxyribonucleases are also secreted. Some of the enzymes such as trypsinogen are present in more than one form (cationic and anionic form). Proteases enter the duodenum as inactive zymogens; trypsinogen is then partly converted to trypsin by the duodenal enzyme enterokinase. Subsequently, trypsin exerts an autocatalytic effect and also activates the other proteases (Figure 2). In contrast to proteases, α -amylase and lipase are secreted in active form.

Like the other gastrointestinal organs, the pancreas is not inactive during the fasting state but maintains a cyclical secretory pattern with maximal interdigestive secretion rates at

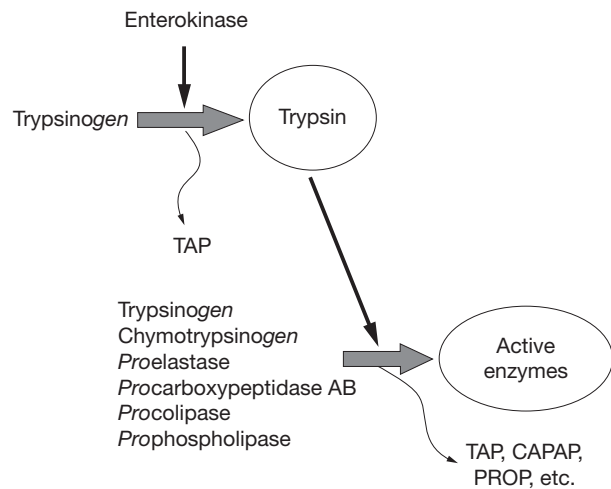


Figure 2 The pancreatic enzyme cascade: pancreatic proteases enter the intestinal lumen as inactive zymogens. Within the duodenum a specific enzyme of the duodenal mucosa, enterokinase, activates trypsinogen by releasing the trypsinogen activation peptide (TAP). Subsequently, active trypsin activates the other zymogens and acts autocatalytically.

about 2-h intervals that is closely coordinated with gastrointestinal motor functions. However, in response to ingestion of a regular meal, enzyme output increases considerably with a short phase of maximal enzyme secretion. This means a three- to sixfold increase above interdigestive levels of all major enzymes. Overall, maximal postprandial outputs reach about 3–6 kU min^{−1} for lipase, 0.5–1 kU min^{−1} for amylase, and 0.2–1 kU min^{−1} for trypsin. Thereafter, 2–4 kU min^{−1} of lipase, 0.5 kU min^{−1} of amylase, and 0.15–0.5 kU min^{−1} of trypsin are stably released into the duodenum for 2 h or longer. Late postprandially, secretory rates return to the interdigestive range. Both, degree and duration of the digestive enzyme response are determined by the caloric content, nutrient composition, and physical properties of a meal. Meals containing at least about 500 kcal induce maximal enzyme responses. Lipids are the strongest stimulants of pancreatic enzyme secretion, but, in humans, duodenal free fatty acids rather than triglycerides appear to be responsible for release of cholecystokinin from the duodenal mucosa which in turn stimulates enzyme output. The stimulatory potency of proteins is limited to specific essential amino acids (phenylalanine, valine, methionine, and tryptophane). Carbohydrates are the weakest stimulants.

Following submaximal digestive stimulation by a test meal devoid of exogenous proteins, cumulative postprandial duodenal protein delivery mainly represents pancreatic enzyme output and was demonstrated to be about 5 g. Maximal duodenal protein delivery was achieved 1 h postprandially and reached 30 g min^{−1}.

The most important stimulatory hormones are cholecystokinin and secretin. Both are peptide hormones which are produced and released by the enterochromaffin cells of the proximal small intestinal mucosa. Cholecystokinin is the strongest physiological stimulator of enzyme secretion. By contrast, secretion of bicarbonate is predominantly stimulated

by secretin which is released in response to duodenal acid exposure. Apart from this, vagal efferents are important regulators of both, enzyme and bicarbonate secretion. Moreover, in humans, part of the hormonal effects appear to be mediated via receptors on vagal nerve fibers instead of receptors on pancreatic cells.

Late postprandially, stimulatory mechanisms caused by nutrients in the proximal small bowel and inhibitory mechanisms induced by nutrient exposure of the distal small bowel (ileal brake mechanism) seem to interact to coregulate pancreatic exocrine secretion during the functional interface between the late postprandial and the subsequent interdigestive period.

Bile Acid Secretion and Absorption

In parallel with pancreatic secretion, basal bile acid output (about 10–20 mmol min⁻¹) is increased markedly by meal ingestion (three- to sixfold) leading to a short-lived peak of bile acid output followed by a period of stable or slowly declining biliary secretion. Bile acids are produced by the liver and represent the water-soluble end products of cholesterol metabolism. After their synthesis from cholesterol, they are conjugated with glycine or taurine. This conjugation makes them impermeable to cell membranes; in consequence, high concentrations are achieved in bile and intestinal chyme. Bile acids are secreted into the proximal duodenum and absorbed actively from the terminal ileum by the so-called apical bile salt transporter (sodium-dependent cotransporter) and a basolateral transporter (anion exchanger). According to animal studies, there may be a second transporter in the proximal small intestine that transports dihydroxy-conjugates preferentially. Each bile acid molecule undergoes multiple enterohepatic circulations before it is finally excreted. Unconjugated bile acids formed by bacterial deconjugation are passively absorbed in the distal intestine and promote diarrhea.

Lipid Digestion and Absorption

Triglycerides

Long-chain triglycerides are the most important dietary lipids. Their digestion depends on an intricate interplay among pancreatic lipase, colipase, and bile acids. Pancreatic lipase hydrolyzes a triglyceride molecule to two fatty acid molecules and 2-monoacylglycerol. For this, lipase binds to the oil–water interface of oil droplets. Both, bile acids and colipase are important to achieve its full lipolytic activity: Bile acids support the emulsification of triglycerides and form micelles with fatty acids and monoglycerides to remove them from the oil–water interface. Colipase forms a complex with lipase in the presence of bile salts. This complex anchors lipase and allows its action in a more hydrophilic environment. Further, hydrolysis of a minor fraction of the 2-monoacylglycerol molecules by pancreatic lipase can result in the formation of glycerol and free fatty acids; cholesterol esterase can also hydrolyze the acyl group at the *sn*-2 position to form the same end products.

Long-chain free fatty acids and 2-monoacylglycerol are absorbed after incorporation into micelles. A protein-independent diffusion model and protein-dependent mechanisms have been proposed for the uptake and transport of fatty acids across

the apical membrane of the enterocyte. Within the enterocytes, the products of intraluminal triglyceride digestion are resynthesized to form triglycerides before they are transported to the liver.

Triglyceride digestion in healthy humans is highly efficient and up to 80% of lipids may be absorbed during duodenal transit, that is, even before reaching the ligament of Treitz. In exocrine pancreatic insufficiency, on the other hand, elevated fecal fat excretion is only expected if prandial secretory rates are less than 5–10% of normal.

Phospholipids

Phosphatidylcholine is the predominant phospholipid in the lumen of the small intestine. Phospholipids are digested by pancreatic phospholipase A2 and other pancreatic lipases. These interact with phospholipids at the *sn*-2 position to yield free fatty acids and lysophosphatidylcholine. These metabolites are removed from the water–oil interface when they are incorporated into the mixed micelles.

Cholesterol

Cholesterol in the intestinal lumen is derived from both, endogenous sources and dietary sources: The human diet provides about 400 mg of cholesterol daily, and the liver secretes 1 g daily. About 50% of the cholesterol in the intestine is absorbed; the remainder is excreted with feces. Most dietary cholesterol exists in the form of the free sterol, with only 10–15% existing as the cholesteryl ester. The latter must be hydrolyzed by cholesterol esterase to release free cholesterol for absorption. Moreover, cholesterol must pass through a diffusion barrier at the intestinal lumen–enterocyte membrane interface before it can interact with the transporter proteins responsible for its uptake and subsequent transport across the cellular brush border. Bile salt micelles facilitate the transfer of cholesterol across the unstirred water layer. Thus, cholesterol absorption depends on the presence of bile acids in the intestinal lumen.

Protein Digestion and Absorption

Digestion of dietary and endogenous proteins by pepsins and pancreatic proteases produces small polypeptides and free amino acids; these are further hydrolyzed by brush border peptidases. Absorption mainly takes place in the proximal jejunum but other parts of the small intestine also have significant transport capacity. The transport into the enterocytes occurs by major active carrier transporters for neutral, dibasic, and dicarboxylic amino acids.

Small peptides consisting of 2 or 3 amino acids can be absorbed across the apical membrane of the enterocytes lining the small intestine via the proton-driven peptide transporter PEPT-1. Small peptides are then mostly broken down intracellularly into single amino acids. However, intact polypeptides can also be transported across the enterocyte membrane. Thus, in individuals with amino acid transport defects, peptide absorption can be normal. The nutritional importance of oligopeptide absorption may be considerable because the total absorption of amino-nitrogen is greater from such peptides that are more rapidly absorbed than amino acids even in diseased intestine.

Gastric acid suppression therapy hampers protein assimilation. However, the relative importance of gastric protein digestion (by acid and pepsin) is generally thought to be limited because of the high proteolytic activity delivered by the pancreas. Conversely, decreased pancreatic protease secretion may partly be compensated by gastric and small intestinal compensatory mechanisms.

Carbohydrate Digestion and Absorption

Dietary carbohydrate is also predominantly absorbed within the proximal small intestine. Before small intestinal absorption can occur, complex carbohydrates need to be hydrolyzed by salivary and pancreatic amylases to glucose, maltose, maltotriose, and oligosaccharides. Together with dietary disaccharides (mainly sucrose and lactose), these metabolites are further hydrolyzed by the brush border enzymes maltase, sucrase-isomaltase, and lactase resulting in monosaccharide production. Luminal monosaccharides such as glucose and galactose are transported by carrier-mediated mechanisms across the enterocyte brush border. Up to 50% of this transport depends on a sodium ion gradient. This active process requires the presence of a sodium gradient between the exterior and interior of the cells. The low intracellular sodium concentration is maintained by the basolaterally located Na-K-adenosine triphosphatase, which extrudes accumulated sodium. Transport of sodium ions down their electrochemical gradient provides the driving force for glucose cotransport and allows the uptake of glucose from a low extracellular concentration into the cells. The main sodium-glucose cotransporter present in the human intestine is SGLT-1.

Apart from this, five functional mammalian-facilitated hexose carriers (GLUTs) have been characterized by molecular cloning. Of these, three are high-affinity transporters (GLUT-1, GLUT-3, and GLUT-4) and one is a low-affinity transporter (GLUT-2) for glucose. GLUT-5 is primarily a fructose carrier. The expression of these transporters is regulated by glucose and several hormones. The transporters most relevant to monosaccharide absorption in the intestine are GLUT-2 and GLUT-5, which support facilitated rather than active transport.

Contribution of the Colon to Nutrient Absorption

Under physiological circumstances, the colon contributes very little to overall nutrient absorption. By contrast, in states of carbohydrate malabsorption, such as severe pancreatic exocrine insufficiency or short bowel syndrome, the colon can become a significant source of energy salvage: Up to about 80% of carbohydrates not absorbed by the small bowel can be absorbed after bacterial metabolism to short-chain fatty acids and thus contribute markedly to energy supply (500–1000 kcal).

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Lipases; Sugar Nucleotide Transporters; **Metabolism Vitamins and Hormones:** Amino Acid Metabolism; Bile Salts; Folate and Vitamin B₁₂ Function; Glucose/Sugar Transport in Mammals; Vitamin C.

Further Reading

- Broer S (2008) Amino acid transport across mammalian intestinal and renal epithelia. *Physiological Reviews* 88: 249–286.
- Camilleri M (2006) Integrated upper gastrointestinal response to food intake. *Gastroenterology* 131: 640–658.
- Drozdowski LA and Thomson ABR (2006) Intestinal sugar transport. *World Journal of Gastroenterology* 12(11): 1657–1670.
- Iqbal J and Hussain MM (2009) Intestinal lipid absorption. *American Journal of Physiology. Endocrinology and Metabolism* 296: E1183–E1194.
- Keller J and Layer P (2005) Human pancreatic exocrine response to nutrients in health and disease. *Gut* 54(supplement VI): vi1–vi28.
- Kenneth EL and McCo MD II (2009) Effect of proton pump inhibitors on vitamins and iron. *American Journal of Gastroenterology* 104: S5–S9.
- Layer P and Keller J (2005) Gastric lipase and pancreatic exocrine insufficiency. *Clinical Gastroenterology and Hepatology* 3(1): 25–27.
- Mittal RK and Bhalla V (2004) Oesophageal motor functions and its disorders. *Gut* 53: 1536–1542.
- Monte MJ, Marin JGJ, Antelo A, and Vazquez-Tato J (2009) Bile acids: Chemistry, physiology, and pathophysiology. *World Journal of Gastroenterology* 15(7): 804–816.
- Pedersen AM, Bardow A, Beier Jensen S, and Nauntofte B (2002) Saliva and gastrointestinal functions of taste, mastication, swallowing, and digestion. *Oral Diseases* 8: 117–129.

Gene Expression in Bacterial Systems: The *trp* Operon and Attenuation

P Gollnick, University at Buffalo, Buffalo, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Intrinsic terminator A signal in the nascent RNA transcript that signals RNA polymerase to halt transcription and dissociate from the DNA template. Intrinsic terminators consist of a short base-paired stem-loop structure followed by a short stretch of U residues in the RNA. Also called factor-independent terminator or Rho-independent terminator.

Ribosome The large RNA-protein complex that translates messenger RNA (mRNA) into protein. It consists of two subunits, termed small and large.

RNA polymerase (RNAP) The enzyme that transcribes DNA into RNA. In bacteria, there is only one version of this enzyme.

Transcriptional pausing In response to signals in the RNA, RNAP will pause and discontinue transcription but not dissociate from the DNA template. Either after some period of time, or in response to a signal, a paused RNAP will resume transcription.

Transfer RNA (tRNA) The adaptor RNA molecule that reads the genetic code in the mRNA by base pairing with the codon triplets. An amino acid is attached to the 3' end of the last base in the tRNA by an enzyme called an aminoacyl tRNA synthetase in a process called 'aminoacylation' or 'charging'. There are tRNAs corresponding each of the 20 amino acids.

Transcription attenuation is defined as any mechanism that utilizes transcriptional pausing or termination to control expression of downstream genes. This mechanism of regulation is used to control expression of many genes in bacteria in responses to changes in their environment. Moreover, bacteria have evolved several elaborate and elegant variations on the attenuation theme, mainly based on the mechanism used to control formation of the RNA structures that signal RNA polymerase (RNAP) to pause or terminate transcription.

Attenuation

The first attenuation mechanism to be thoroughly described was that which controls transcription of the *Escherichia coli* tryptophan biosynthetic (*trpEDCBA*) operon. When Charles Yanofsky and co-workers elucidated this mechanism, it was the first demonstration that organisms can utilize changes in RNA structure to regulate gene expression. Subsequently, many other examples of transcription attenuation have been discovered. In each case, the *cis*-acting genetic information is contained within a 150–300-base region of RNA located after the start of transcription and prior to the start of the coding sequence for the first structural gene of the operon. This segment of RNA is called the leader region but is analogous to 5' untranslated regions (5'UTRs) seen in many eukaryotic messenger RNAs (mRNAs). The article illustrates the features of transcription attenuation control of the *E. coli trp* operon and then describes the variations to this classic system that allow attenuation to be adapted to control other amino acid biosynthetic operons, then goes on to describe attenuation control of the *trp* operon in the Gram-positive bacterium *Bacillus subtilis*, which involves larger variations on the attenuation theme. There are several recent reviews that cover the attenuation in more depth as well as describe attenuation control of other bacterial genes.

The *E. coli trp* Operon: Attenuation Based on Translation of a Leader Peptide

The *E. coli trpEDCBA* operon encodes the enzymes required to synthesize *L*-tryptophan from chorismic acid. Transcription of the *trp* operon is regulated in response to changes in the levels of tryptophan inside the cell. When cells contain adequate amounts of tryptophan, for example, when it is present in the growth medium, transcription of the operon is downregulated. By contrast, when the cell has low levels of tryptophan, the *trp* operon is actively transcribed in order to express the enzymes required for synthesizing more. Initiation of transcription is regulated by the *trp* repressor, a DNA-binding protein encoded by the *trpR* gene. In addition, after transcription has initiated, the elongating transcription complex is subject to regulate by attenuation. Together, repression (80-fold) and attenuation (eightfold) serve to allow approximately 600-fold overall control of transcription of the *trp* operon in response to various levels of tryptophan availability.

Attenuation Control of the *E. coli trp* Operon

The *E. coli trp* operon contains a 162-nucleotide leader region prior to the start of the *trpE* coding sequence. There are several inverted repeats in the *trp* leader mRNA transcript, which are composed of the segments labeled 1–4 in Figure 1. These segments can base-pair to form three different overlapping base-paired RNA secondary structures (Figure 1). These structures include an intrinsic transcription terminator (3:4), an overlapping antiterminator (2:3), and a pause structure (1:2). In addition, the leader transcription contains a small open reading frame (ORF), which encodes a 14-amino-acid leader peptide. There are two critical tandem UGG Trp codons within this ORF. How efficiently these two Trp codons can be translated determines which RNA structure forms in the nascent leader transcript, which in turn controls whether transcription

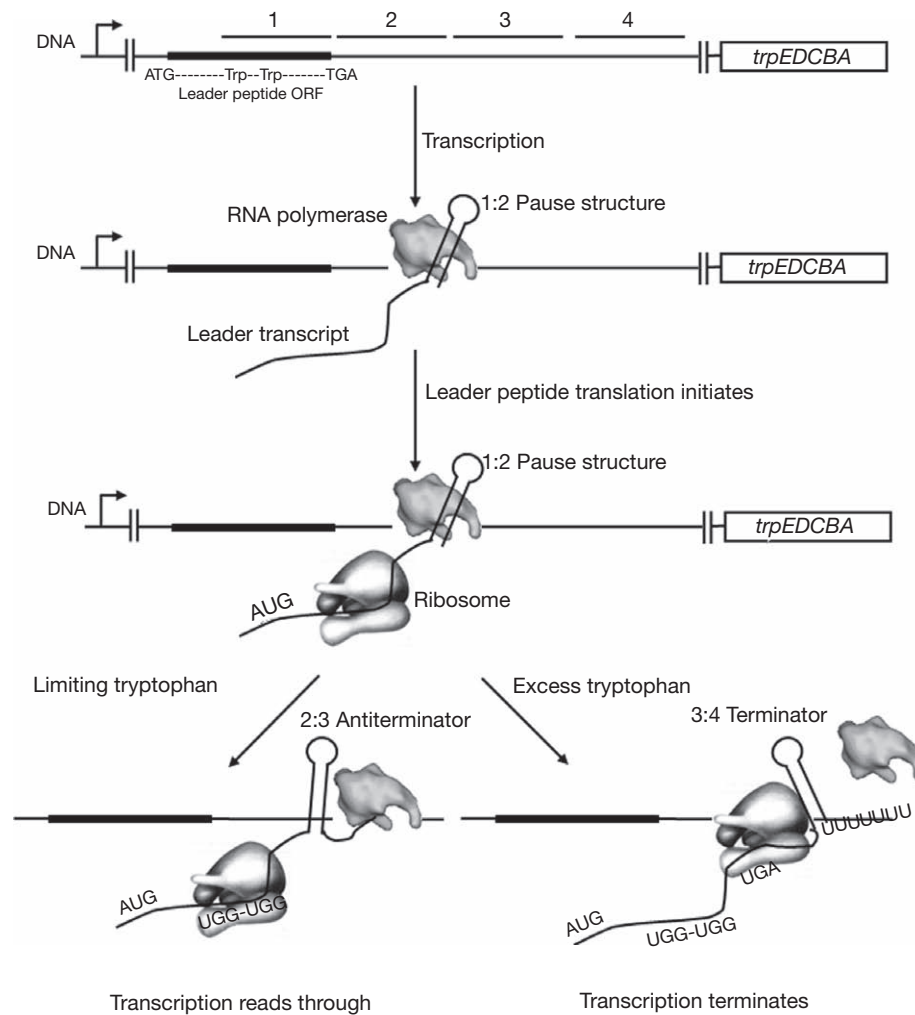


Figure 1 Model of transcription attenuation of the *E. coli trp* operon. RNA polymerase pauses following formation of the pause structure. This provides time for a ribosome to initiate translation of the leader peptide. Under tryptophan limiting conditions, the ribosome stalls at the tandem Trp codons, resulting in transcription read through. Under conditions of tryptophan excess, the ribosome reaches the leader peptide stop codon. This ribosome position blocks formation of the antiterminator leading to terminator formation and transcription termination.

halts in the leader region or continues into the structural genes of the operon.

Shortly after transcription initiates from the *trp* promoter, the 1:2 pause hairpin forms in the nascent RNA (Figure 1). This structure signals RNAP to suspend transcription after nucleotide 92. Pausing of RNAP is a critical feature of the attenuation mechanism because it allows time for a ribosome to initiate translation of the leader peptide. When a ribosome begins translating the leader peptide, this releases the paused RNAP to resume transcription. Transcription and translation are now coupled, with the ribosome closely following RNAP. This situation is essential to allow events involving the ribosome to affect transcription by the associated RNAP.

At this point, there are two possible pathways for the attenuation mechanism to follow depending on the level of tryptophan in the cell. The choice depends on how efficiently the tandem Trp codons in the leader peptide are translated. This efficiency reflects the availability of aminoacylated $tRNA^{Trp}$ in the cell. Under conditions where tryptophan is limiting, the

amount of charged $tRNA^{Trp}$ is low. As a result of the low concentration of tryptophanyl- $tRNA^{Trp}$, translation of the tandem Trp codons is inefficient and the ribosome stalls at one of these two codons. The associated RNAP continues to transcribe through the *trp* leader region and transcription and translation become uncoupled. As RNAP continues transcribing through segments 2 and 3 of the leader region, the antiterminator structure (2:3) forms, which prevents formation of the overlapping intrinsic terminator (3:4) structure. Hence, transcription continues through the leader region and into the structural genes of the operon which encode the tryptophan biosynthetic enzymes.

When tryptophan is plentiful, the level of charged $tRNA^{Trp}$ in the cell is high. This situation allows efficient translation of the tandem Trp codons and hence the ribosome proceeds rapidly to the end of the leader peptide. When the ribosome reaches the leader peptide stop codon, it covers part of RNA segment 2 and prevents formation of the 2:3 antiterminator structure as transcription proceeds. Thus, RNA segment 3 is free to base-pair with segment 4 and form the transcription

terminator. Under these conditions, transcription terminates in the leader region prior to the *trp* structural genes, which are therefore not expressed.

In this attenuation mechanism the regulatory signal is the level of aminoacylated tRNA^{Trp} and the sensory event is the efficiency with which the ribosome can translate the tandem UGG Trp codons in the leader peptide. This system can easily be adapted to regulate other amino acid biosynthetic operons. The only modification that is needed is to change the identity of the critical codons in the leader peptide, which controls the response of the system to a specific amino acid. There are numerous examples of such adaptations of this attenuation mechanism, particularly in enteric bacteria, including the *his*, *phe*, and *leu* operons, which contain seven His, seven Phe, and four Leu codons in their respective leader peptides.

The *B. subtilis trp* Operon: Attenuation Mediated by an RNA-Binding Protein

The transcription attenuation mechanism that regulates the *trp* operon in the Gram-positive bacterium *B. subtilis* differs more dramatically from that of the *E. coli trp* operon than those described above for other amino acid biosynthetic operons in Gram-negative bacteria. Most notably there is no leader peptide, and ribosomes are not involved in this attenuation mechanism. Instead, an RNA-binding protein senses the level of tryptophan in the cell and determines whether transcription terminates in the leader region or continues into the structural genes.

The *B. subtilis trp* Leader

Expression of the *trpEDCFBA* operon in *B. subtilis* and related bacteria is regulated by the *trp* RNA-binding attenuation protein (TRAP). Transcription of the operon initiates from the *trp* promoter 203 nucleotides upstream of the start codon of the first structural gene, *trpE* (Figure 2). There is no evidence for any regulation of initiation of transcription from this promoter. Hence, there is no homolog, either functional or based on amino acid sequence, of the DNA-binding *trp* repressor from *E. coli* in *B. subtilis*. The *B. subtilis trp* leader transcript contains inverted repeats that can form an intrinsic transcription terminator (C:D) and an antiterminator (A:B) RNA secondary structure (Figure 2). These two structures overlap by four nucleotides and hence their formation is mutually exclusive.

In the presence of excess tryptophan, TRAP binds to a series of 11 trinucleotide repeats in the *trp* leader transcript consisting of seven GAGs and four UAGs (Figure 2). These (G/U)AG repeats are separated from each other by two or three spacer nucleotides, whose sequence is not conserved. The triplet repeats partially overlap the 5' portion of the antiterminator structure and hence TRAP binding interferes with formation of the antiterminator (A:B) thus favoring formation of the terminator (C:D) hairpin. Therefore, under these conditions, transcription halts in the leader region prior to the *trp* structural genes, which are not expressed.

When tryptophan is limiting, TRAP does not bind to the *trp* leader RNA, and hence formation of the A:B antiterminator is favored. This situation allows transcription to continue into the *trp* operon structural genes, which can now be expressed to produce the enzymes to synthesize tryptophan.

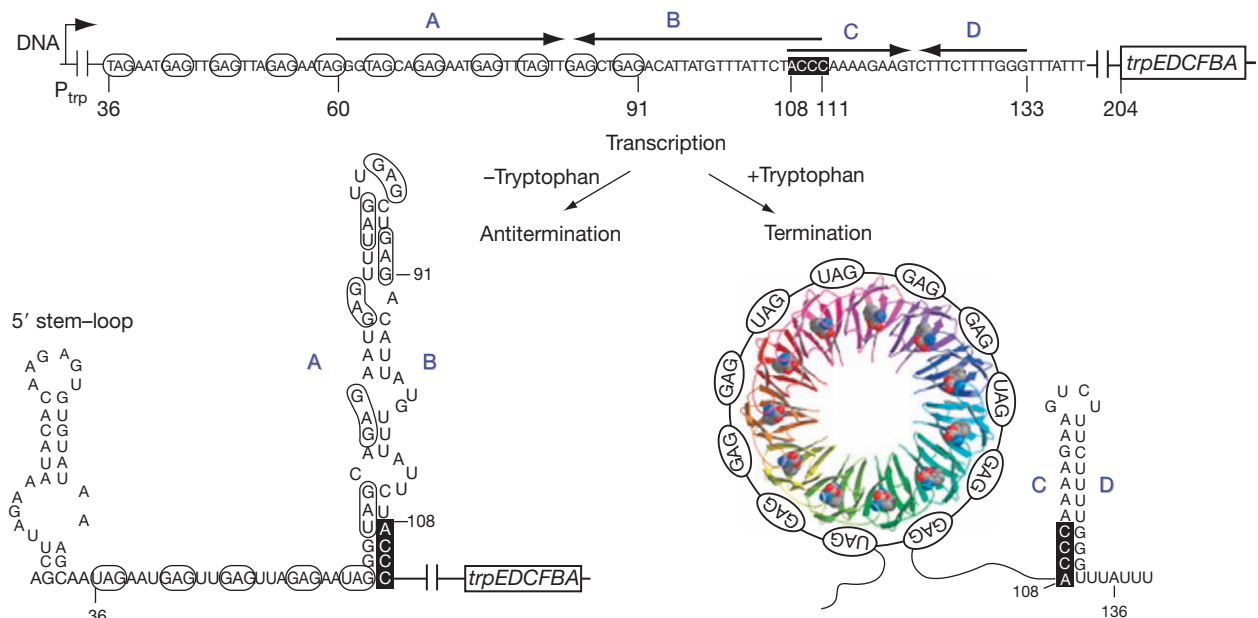


Figure 2 Model of transcription attenuation of the *B. subtilis trpEDCFBA* operon. The large blue letters indicate the complementary strands of the terminator and antiterminator RNA structures. The TRAP protein is shown as a ribbon diagram with each of the 11 subunits as a different color and the bound RNA is shown forming a matching circle upon binding to TRAP. The bound tryptophans are shown as space filling models. The GAG and UAG repeats involved in TRAP binding are shown in ovals and are also outlined in the sequence of the antiterminator structure. Numbers indicate the residue positions relative to the start of transcription. Nucleotides 108–111 overlap between the antiterminator and terminator structures and are shown as darkly outlined letters.

The TRAP Protein

In this system, the basis for attenuation control is the ability of TRAP to sense the level of intracellular tryptophan and then to influence the structure of the nascent *trp* leader mRNA transcript. TRAP is a ring-shaped protein composed of 11 identical subunits (Figure 3). TRAP is activated to bind its RNA targets by binding 11 molecules of *L*-tryptophan in clefts between adjacent subunits. The detailed mechanism by which tryptophan binding activates TRAP to bind RNA is not known; however, it is clear that the protein is an 11mer in the absence or presence of bound tryptophan. Recent studies have shown that tryptophan binding induces a conformational change in TRAP as well as reduces the flexibility of the protein.

Upon activation by tryptophan, TRAP binds to its RNA target consisting of the 11 (G/U)AG repeats by wrapping the single-stranded RNA around the outer perimeter of the protein ring (Figure 3). The phosphodiester backbone is on the outside of the RNA ring and the bases point in toward the protein. The specificity of this interaction derives from hydrogen bonds between Glu36, Lys37, Lys56, and Arg58 of each TRAP subunit and the bases on the (G/U)AG repeats. The spacer nucleotides between the triplet repeats do not contact the protein, and their identity is not crucial for TRAP recognition and binding. This mechanism of wrapping the RNA around the protein ring explains both the specificity of the TRAP–RNA interaction as well as how TRAP binding alters the structure of the *trp* leader RNA to regulate attenuation.

The Role of Timing in TRAP-Mediated Attenuation

As described above for the *E. coli trp* attenuation system, the timing of altering the folding of the leader region RNA is crucial for proper regulation of transcription. In this case, TRAP must bind to its target and induce formation of the transcription terminator before RNAP transcribes beyond the terminator segment of the leader region. If TRAP binds too late,

altering the folding pattern of the RNA will not influence the activity of the transcribing RNAP. There are several features of the *B. subtilis trp* system that serve to ensure proper timing of the regulatory factors.

There are several features of the *trp* attenuation system that allow TRAP to bind to the nascent transcript early during transcription of the leader region, before all of the 11 triplet repeats of the binding site are exposed from RNAP. The first feature is the multi-subunit nature of TRAP combined with the multiple small repeated elements in the RNA binding site. TRAP has been shown to bind with high affinity to RNAs with as few as 4–5 (G/U)AG repeats. In addition, an RNA hairpin forms at the 5' end of the *trp* leader mRNA (Figure 2, 5'SL), which influences TRAP binding to the RNA. The presence of the 5'SL increases the affinity of TRAP for the leader region RNA when fewer than the full complement of 11 triplets have been synthesized.

Recent studies also show that RNAP pauses at U107 (Figure 2) during transcription of the *trp* leader region, and that this pausing is dependent on the NusA transcription factor. This pause provides additional time for TRAP to bind to the nascent *trp* leader RNA and promote termination.

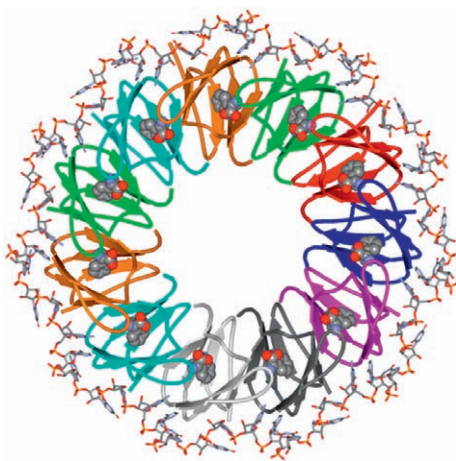


Figure 3 Ribbon diagram of TRAP complexed with an RNA containing 11 GAG repeats separated by AU spacers. The TRAP protein is shown in ribbon diagrams with each subunit depicted in a different color. The RNA is shown in stick models and the 11 molecules of bound *L*-tryptophan are shown in space filling models.



Figure 4 Ribbon diagram of anti-TRAP (AT) and AT in complex with TRAP. The proteins are shown as ribbon with each subunit depicted in a different color. The 11 molecules of *L*-tryptophan bound to TRAP and the zinc atoms bound to AT are shown in space filling models.

Anti-TRAP and the Role of tRNA^{Trp} in Regulating the *B. subtilis trp* Genes

In the *B. subtilis trp* attenuation system, the regulatory signal is the level of the amino acid tryptophan free in the cell, which activates TRAP to bind RNA. In the previously described *E. coli trp* attenuation system, free tryptophan is sensed by the DNA-binding *trp* repressor protein and the regulatory signal for attenuation is the availability of aminoacylated tRNA^{Trp}. The level of charged tRNA^{Trp} also influences attenuation control of the *B. subtilis trp* operon, although by a very different mechanism than in the *E. coli* system. In *B. subtilis*, the *rtpA* gene encodes a protein called anti-TRAP (AT) (Figure 4). AT influences expression of the *trp* genes by binding to tryptophan-activated TRAP and preventing it from binding RNA, thus elevating expression of *trp* genes. Transcription of *rtpA*, which is in a two-gene operon together with *ycbK* (unknown function), is regulated in response to changes in the level of uncharged tRNA^{Trp} by a mechanism known as T-box antitermination. High levels of uncharged tRNA^{Trp} induce transcription of AT, which prevents TRAP from downregulating expression of the *trp* genes.

Translation of *rtpA* is also regulated in response to the level of aminoacylation of tRNA^{Trp}. This regulation is based on the cells' ability to translate a 10-codon leader peptide sequence located just upstream from the *rtpA* gene. The leader peptide-coding segment contains three consecutive Trp codons. If there is insufficient aminoacylated tRNA^{Trp} to allow efficient translation, the translating ribosome stalls at one of these three Trp codons. This situation exposes the ribosome-binding site for the *rtpA* gene and allows efficient translation of rtpA. However, if there is sufficient charged tRNA^{Trp} to allow efficient translation of rtpLP, when the ribosome reaches the stop codon, its presence blocks the ribosome-binding site for *rtpA* and inhibits translation initiation at the *rtpA* start codon.

Together, these systems ensure that even if free tryptophan levels are high enough to activate TRAP, if the level of charging of tRNA^{Trp} is low, then the *trp* genes will be expressed. Thus,

while both *E. coli* and *B. subtilis* regulate expression of the *trp* genes in response to changes in levels of both free tryptophan and aminoacylated tRNA^{Trp}, they have evolved very different mechanisms to do so.

Occurrences of TRAP-Mediated Attenuation in Bacteria

TRAP-mediated attenuation control of the *trp* operon has to date been observed mainly in low GC Gram-positive bacteria and some related species. In contrast to the leader-peptide-dependent mechanism described previously, the RNA-binding protein-dependent attenuation mechanism has never been characterized for any other amino acid biosynthetic operons, besides tryptophan, even though it would seem to be easily adaptable by simply changing the amino acid that activates the protein to bind its RNA target. Perhaps with the explosion of bacterial genomic information currently becoming available, we will soon discover such a system.

See also: [Molecular Biology: Riboswitches; RNA Editing.](#)

Further Reading

- Antson AA, Dodson EJ, Dodson GG, et al. (1999) Structure of the *trp* RNA-binding attenuation protein, TRAP, bound to RNA. *Nature* 401: 235–242.
- Antson AA, Otridge JB, Brzozowski AM, et al. (1995) The three dimensional structure of *trp* RNA-binding attenuation protein. *Nature* 374: 693–700.
- Babitzke P and Gollnick P (2001) Posttranscriptional initiation control of tryptophan metabolism in *Bacillus subtilis* by the *trp* RNA-binding attenuation protein (TRAP), anti-TRAP, and RNA Structure. *Journal of Bacteriology* 183: 5795–5802.
- Gollnick P, Babitzke P, Antson A, and Yanofsky C (2005) Complexity in regulation of tryptophan biosynthesis in *Bacillus subtilis*. *Annual Review of Genetics* 39: 47–68.
- Valbuzzi A and Yanofsky C (2001) Inhibition of the *B. subtilis* regulatory protein TRAP by the TRAP-inhibitory protein, AT. *Science* 293: 2057–2059.
- Yanofsky C (2000) Transcription attenuation: Once viewed as a novel regulatory strategy. *Journal of Bacteriology* 182: 1–8.

Gene Expression in Eukaryotes: RNA Polymerase II Structure

P Cramer, Ludwig-Maximilians-Universität München, Munich, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

Backtracking Reverse movement of Pol II along DNA and RNA.

DNA–RNA hybrid Eight-to-nine base pair (bp) double-stranded heteroduplex formed between the DNA template strand and the RNA transcript within the transcription bubble.

Downstream DNA DNA that is going to be transcribed.

Rpb1–Rpb12 RNA polymerase B (II) polypeptide subunits 1–12.

Transcription bubble Open DNA region in the active center of Pol II.

Translocation The movement of Pol II by a step of 1 bp along the DNA template.

Upstream DNA DNA that has been transcribed.

Architecture

A 10-polypeptide polymerase II (Pol II) core enzyme is sufficient to elongate the RNA transcript. For transcription initiation, an additional two-subunit Rpb4/7 complex and several general transcription factors are required. In the crystal structures of the complete 12-subunit Pol II, the two large subunits, Rpb1 and Rpb2 (Table 1), form the central mass of the enzyme, and opposite sides of a positively charged cleft that contains the active center (Figure 1). The two large subunits are bridged on one side by a module of subunits Rpb3, Rpb10, Rpb11, and Rpb12. Around the periphery of the enzyme, Rpb5, Rpb6, and Rpb8 bind to Rpb1, and Rpb9 binds to both Rpb1 and Rpb2. The Rpb4/7 complex binds to the Pol II core via subunit Rpb7 and protrudes from the Pol II core surface near subunit Rpb6.

Structural Elements

The Rpb1 side of the Pol II cleft is formed by a mobile clamp, whereas the Rpb2 side consists of the lobe and protrusion domains. The entrance to the cleft is formed between the upper and the lower jaw, which include subunits Rpb9 and Rpb5, respectively. The end of the cleft is blocked by a protein wall. The active center is at the floor of the cleft, between the protrusion, the wall, and the clamp. Before the active center and opposite of the wall, a long bridge helix spans the cleft. The bridge partially lines a pore in the active center, which widens toward the other side of the enzyme, creating an inverted funnel. The rim of the pore includes the highly conserved aspartate loop of Rpb1 that binds a catalytic Mg^{2+} ion, termed metal A. A second metal ion, metal B, is weakly bound next to metal A further in the pore. Both metal ions are accessible from one side.

The clamp is trapped in two different open states in the free core structures, but is rotated and closed in the structure of the Pol II transcription elongation complex (EC) that comprises template DNA and the RNA transcript. In this structure, the clamp binds the DNA template strand with three out of five switch regions. The switch regions form the base of the clamp, connecting to the remainder of Pol II. Upon clamp closure,

the switches change conformation or undergo folding transitions. The closed conformation of the clamp is also observed in the complete Pol II that contains the Rpb4/7 complex. The Rpb4/7 complex appears to form a wedge between the clamp and subunit Rpb6, thus preventing clamp opening.

The Rpb1 polypeptide chain protrudes below the Rpb4/7 complex on the outside of Pol II, and gets disordered after a few residues. These residues of Rpb1 constitute the beginning of a flexible linker that connects to the mobile C-terminal repeat domain (CTD). The CTD, a unique feature of Pol II, consists of repeats of a heptapeptide with the consensus sequence Tyr–Ser–Pro–Thr–Ser–Pro–Ser. A total of 26 and 52 CTD repeats is found in yeast and human Rpb1, respectively.

Mechanism

Transcription Cycle

The transcription cycle is divided into three phases: initiation, elongation, and termination. Steps during initiation include promoter binding, DNA melting, and synthesis of short RNA transcripts. The transition from initiation to elongation is referred to as promoter escape, and results in a stable EC that is characterized by an open DNA region, the transcription bubble. The bubble contains the DNA–RNA hybrid, a heteroduplex of 8–9 base pairs (bp). The growing RNA 3'-end is located at the end of the hybrid that is engaged with the Pol II active site. Pol II alone can maintain an open transcription bubble, translocate along the DNA template, synthesize RNA, and detect errors in the nascent RNA. For all other steps, however, Pol II requires the help of additional proteins.

Several steps of the transcription cycle are accompanied by phosphorylation or dephosphorylation of the Pol II CTD. During initiation, the CTD becomes phosphorylated, and the CTD phosphorylation pattern changes during elongation. CTD phosphorylation patterns govern specific interaction with RNA processing factors, thereby coupling transcription to RNA processing. The CTD is flexibly linked to a region beyond the saddle, from which RNA exits, consistent with its role in coupling transcription to messenger RNA (mRNA) processing. Several kinases and at least one phosphatase control the phosphorylation state of the Pol II CTD.

Table 1 Polypeptide composition of cellular RNA polymerases

<i>Pol I</i>	<i>Pol II</i>	<i>Pol III</i>	<i>Archaea</i>	<i>Bacteria</i>	<i>Class^a</i>
A190	Rpb1	C160	A' + A''	β'	Core
A135	Rpb2	C128	B (B' + B'')	β	Core
AC40	Rpb3	AC40	D	α	Core
AC19	Rpb11	AC19	L	α	Core
Rpb6	Rpb6	Rpb6	K	ω	Core, common
Rpb5	Rpb5	Rpb5	H		Common
Rpb8	Rpb8	Rpb8			Common
Rpb10	Rpb10	Rpb10	N		Common
Rpb12	Rpb12	Rpb12	P		Common
A14	Rpb4	C17	F		Rpb4/7
A43	Rpb7	C25	E		Rpb4/7
A12.2	Rpb9	C11	X		Specific
A34.5		C53			Specific
A49		C37			Specific
		C82			Specific
		C34			Specific
		C31			Specific

^aCore, sequence partially homologous in all RNA polymerases; common, shared by all eukaryotic RNA polymerases; and Rpb4/7, Rpb4/7 heterodimer and its structural counterparts.

Initiation

Transcription initiation begins with formation of the pre-initiation complex (PIC) on promoter DNA. The PIC contains Pol II and the general transcription factors TFIIB, -D, -E, -F, and -H. The general factors are involved in sequence-specific promoter recognition (TFIIB and TFIID), prevention of nonspecific DNA binding (TFIIF), and DNA melting and CTD phosphorylation (TFIIE and TFIIH). Recently, the structure of RNA Pol II in complex with TFIIB was presented. The structure shows that to initiate gene transcription, promoter DNA is positioned over the Pol II active center cleft with the B-core domain that binds the wall at the end of the cleft. DNA is then opened with the help of the B-linker that binds the Pol II rudder and clamp coiled coil at the edge of the cleft. The DNA template strand slips into the cleft and is scanned for the transcription start site with the help of the B-reader that approaches the active site. RNA synthesis initiates within the bubble. The early transcribing complex is functionally unstable. Short RNAs are frequently released and Pol II has to restart transcription (abortive cycling).

Synthesis of the RNA chain and rewinding of upstream DNA displace the B-reader and B-linker, respectively, to trigger TFIIB release and EC formation. The elongating RNA extends over the dock domain, preventing reassociation of the TFIIB ribbon with Pol II during elongation.

Elongation

During elongation, downstream DNA enters Pol II at two mobile jaws, and extends through the cleft toward the active site. Beyond the active site, a 9-bp DNA–RNA hybrid extends upward, toward the wall. The axes of the downstream DNA duplex and the DNA–RNA hybrid heteroduplex enclose an angle of almost 90°. The growing RNA 3'-end is located above the pore, which allows entry of nucleoside triphosphate

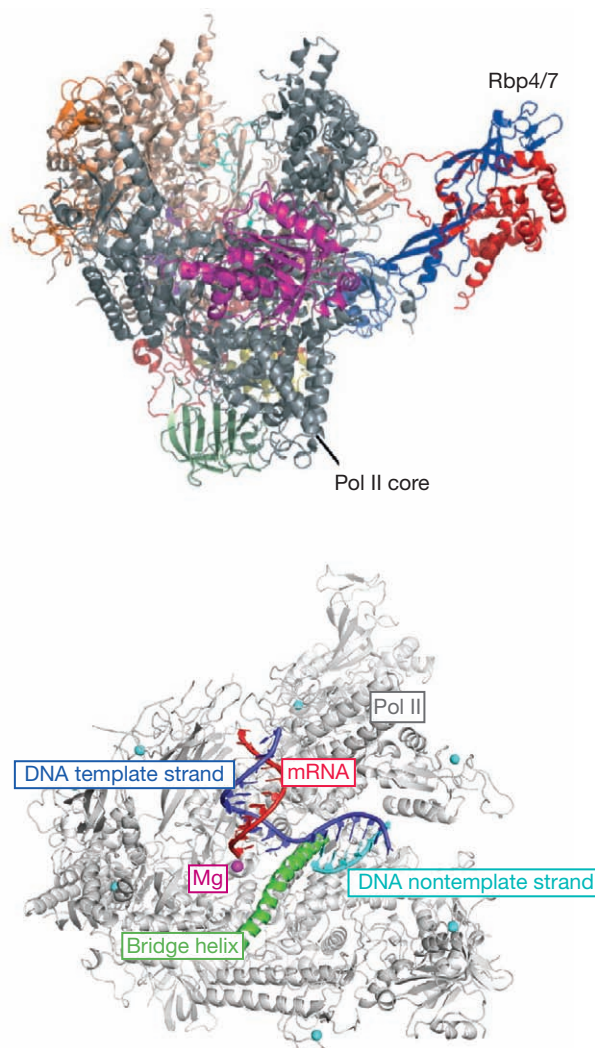


Figure 1 RNA polymerase II structures. (a) Ribbon model of the complete yeast RNA polymerase II crystal structure. The 12 subunits are shown in different colors. (b) Ribbon model of the RNA polymerase II EC. The DNA template and nontemplate strands are in blue and cyan, respectively, and the RNA is in red. Metal ion A is depicted as a magenta sphere. Zinc ions are shown as cyan spheres.

(NTP) substrates from below. In the crystal structure of the Pol II EC, the incoming DNA duplex is mobile but is bound for strand separation closer to the active site. Downstream DNA runs along the Rpb1 face of the cleft. Upstream of the point of DNA-strand separation, the DNA template strand positions a base in the active site that directs RNA synthesis.

The property of the polymerase to stay attached to the template, even during transcription of long genes, is often referred to as 'processivity'. The major cause of processivity is the high stability of the Pol II EC, caused by tight binding of the DNA–RNA hybrid. This stability can be accounted for by the highly complementary hybrid-binding site on Pol II, which is partially created upon clamp closure and folding of the switches. Several Pol II structural elements maintain the hybrid and the transcription bubble during elongation, including the

fork loops, the top of the wall, and three loops protruding from the edge of the clamp, called 'rudder', 'lid', and 'zipper'.

Extension of the RNA chain is achieved during the nucleotide addition cycle (NAC). The NAC begins with binding of an NTP substrate to the transcription EC. Catalytic addition of the nucleotide to the growing 3'-end of the RNA leads to the formation of a pyrophosphate ion. Pyrophosphate release leads to the pre-translocation state, in which the newly added, 3'-terminal RNA nucleotide remains in the substrate site. Finally, translocation of DNA and RNA results in the post-translocation state, in which the substrate site is free for binding the next NTP, and the NAC can be repeated. Recent structural studies of ECs of the eukaryotic Pol II and the bacterial RNA polymerase have revealed different functional states that reflect intermediates in the NAC.

Specificity for synthesizing RNA rather than DNA is achieved by at least three mechanisms. First, the discriminating 2'-OH group of the incoming NTP may be hydrogen-bonded by a conserved Pol II residue. Second, 2'-OH groups of the last few nucleotides that were incorporated into the growing RNA are directly hydrogen-bonded by Pol II residues. Finally, the active center of Pol II is complementary to the resulting DNA–RNA hybrid duplex that adopts a conformation intermediary between canonical A and B forms.

It is a mystery of the transcription mechanism how rapid translocation of nucleic acids can be achieved while nucleic acids are bound tightly by Pol II. Hints for understanding translocation are however provided by the Pol II structures. First, nucleic acids are only contacted via their backbones; base interactions that would impede translocation are not observed. Second, long-range electrostatic attraction of the nucleic acid backbones by positively charged protein groups in a second shell may enable tight binding of nucleic acids without restricting their movement. Finally, translocation is accompanied by conformational changes in Pol II that maintain some of the protein–nucleic acid contacts during translocation. Structural analysis of Pol II ECs in pre-translocation, post-translocation, and in an intermediate state show that such conformational changes include folding of the trigger loop and bending of the bridge helix.

Pol II elongation is inhibited by the cyclic octapeptide α -amanitin, the toxin of the death cap mushroom. α -Amanitin does not greatly influence NTP binding, and a phosphodiester bond can still be formed when the toxin is added to an EC. However, the rate of transcription is dramatically reduced. Only a handful of nucleotides is incorporated per minute. In the structure of an α -amanitin containing Pol II EC, α -amanitin binds to the trigger loop and to the bridge helix. The structure suggests that α -amanitin inhibits Pol II by trapping the trigger loop and shifted bridge helix, thereby stabilizing a conformation of the EC that represents a translocation intermediate.

Pol II does not move along the DNA template in a unidirectional manner; rather, it oscillates between movements forward and backward (backtracking). During backtracking, RNA is extruded through the Pol II pore into the funnel. Backtracking can lead to transcriptional pausing and arrest. Pausing is defined as a temporary block to elongation, from which Pol II can escape by itself. By contrast, Pol II cannot escape from arrest without the help of the transcript cleavage factor TFIIS,

which stimulates a weak intrinsic nuclease activity of Pol II, which cleaves RNA. Access of TFIIS to the Pol II active site is apparently provided via the funnel and pore. Backtracking and TFIIS action may also underlie proofreading of the nascent transcript. During proofreading, a misincorporated nucleotide results in a mismatch base pair within the DNA–RNA hybrid. The induced distortion of the hybrid destabilizes the EC and can lead to immediate cleavage of a mononucleotide, or it can trigger in backtracking, which results in threading of a short stretch of RNA (comprising the misincorporated nucleotide) into the pore and cleavage with the help of TFIIS. Any proofreading reaction creates a new RNA 3'-end at the active site, from which transcription can resume.

Termination

Transcription termination occurs in a reaction coupled to RNA 3'-end processing. Most eukaryotic mRNA precursors are cleaved in a site-specific manner in the 3'-untranslated region, followed by polyadenylation of the upstream cleavage product. A large number of proteins are involved in these reactions. The exact mechanism of coupling between 3'-end processing and transcription termination remains unclear. Termination is accompanied by dephosphorylation of the Pol II CTD, but the precise timing of Pol II dephosphorylation is also unclear. Dephosphorylation is required for the reinitiation of transcription, because Pol II can only join an initiation complex in its unphosphorylated form. The CTD phosphatase Fcp1 plays a key role in Pol II dephosphorylation and recycling. Fcp1 binds to Pol II via Rpb4. Rpb4 apparently recruits Fcp1 to the vicinity of the CTD, because the Rpb4/7 complex binds near the linker that connects the Pol II core to the CTD.

Conservation

Pol II belongs to the family of cellular RNA polymerases, which also includes two other eukaryotic RNA polymerases, Pol I and Pol III, and the bacterial and archaeal RNA polymerases. All three eukaryotic RNA polymerases share five common subunits (Table 1). Four core subunits of Pol II, Rpb1, Rpb2, Rpb3, and Rpb11, as well as subunit Rpb9, all have close homologs in Pol I and Pol III. The Rpb4/7 complex of Pol II also has structural and functional counterparts in Pol I and Pol III, and in the archaeal RNA polymerase. As the 12 subunits of Pol II are either identical or homologous in all three eukaryotic enzymes, the Pol II structure is a good model for all eukaryotic RNA polymerases. Comparison of Pol II with a bacterial RNA polymerase structure revealed that five core subunits underlie a general RNA polymerase architecture with an active center cleft. Twenty-two regions of sequence homology cluster around the active site and generally adopt the same structure in all cellular RNA polymerases. The structurally conserved core includes the functional elements of the active center, indicating that all cellular RNA polymerases share common mechanistic features. Homologs for all Pol II subunits (except Rpb8) are found in archaeal RNA polymerases. Thus, the overall structure of archaeal RNA polymerases is very similar to that of Pol II except for several external domains.

The Pol II structure is strikingly different from structures of the many single subunit DNA and RNA polymerases. However,

in functional complexes of these diverse enzymes, nucleic acids take a similar course through the active center. The entering DNA duplex encloses an angle of almost 90° with the exiting template–product duplex. At the location of the bend, subsequent DNA template bases are twisted. This twist aligns the coding base with the binding site for the incoming NTP substrate. The NTP enters through an opening that is found in all polymerases, and generally binds between an α -helix and two catalytic metal ions. The exiting template–product duplex is bound from the minor groove side in all polymerases. Conformational changes upon nucleic acid binding have been detected for several different polymerases, but the nature of this induced fit differs.

See also: **Molecular Biology:** RNA Polymerase I and RNA Polymerase III in Eukaryotes; RNA Polymerase II and Its General Transcription Factors; RNA Polymerase II Elongation Control in Eukaryotes; RNA Polymerase Structure, Bacterial; T7 RNA Polymerase; Transcription Termination.

Further Reading

- Armache K-J, Kettenberger H, and Cramer P (2003) Architecture of the initiation-competent 12-subunit RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* 100: 6964–6968.
- Brueckner F and Cramer P (2008) Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nature Structural and Molecular Biology* 15: 811–818.
- Bushnell DA and Kornberg R (2003) Complete 12-subunit RNA polymerase II at 4.1 Å resolution: Implications for the initiation of transcription. *Proceedings of the National Academy of Sciences of the United States of America* 100: 6969–6973.
- Cramer P (2002) Multisubunit RNA polymerases. *Current Opinion in Structural Biology* 12: 89–97.
- Cramer P, Bushnell DA, Fu J, et al. (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288: 640–649.
- Cramer P, Bushnell DA, and Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* 292: 1863–1876.
- Gnatt AL, Cramer P, Fu J, Bushnell DA, and Kornberg RD (2001) Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292: 1876–1882.
- Kostrewa D, Zeller ME, Armache KJ, et al. (2009) RNA polymerase II–TFIIB structure and mechanism of transcription initiation. *Nature* 462: 323–330.
- Woychik NA and Hampsey M (2002) The RNA polymerase II machinery: Structure illuminates function. *Cell* 108: 453–463.

Genome-Wide Analysis of Gene Expression

D-W Doug Chung and K G Le Roch, Institute for Integrative Genome Biology, University of California Riverside, Riverside, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

cDNA DNA that has been created from RNA using reverse transcription.

Microarray A device typically used for monitoring gene expression for a large number of genes. They consist of an ordered array of different elements (usually nucleic acids complementary to the genes of interest) that are usually printed or synthesized *in situ* on a solid surface. Microscopic beads can also be used.

Probe The individual elements bound to a microarray consisting of long oligonucleotides, short oligonucleotides (25mers), fragments of genomic DNA or cDNAs.

Read The sequence of bases determined from a single segment of sample nucleic acid by a sequencing instrument.

Read length The number of individual bases identified in a continuous read.

RNA Ribonucleic acid. Single-stranded molecules transcribed from DNA.

RNA-seq Whole transcriptome analysis based on high-throughput sequencing technologies.

Second-generation sequencing (SGS) High-throughput sequencing based on short polymerase chain reaction (PCR) amplified DNA fragments, usually based on wash-and-scan techniques.

Serial analysis of gene expression (SAGE) A method used in the genome-wide monitoring of gene expression in which short sequence tags from different cDNAs are isolated, sequenced, and counted.

Third-generation sequencing (TGS) High-throughput direct sequencing of single molecules of nucleic acid without the need of amplification or halting between read steps.

Gene Expression Data for Discovery and Classification

Gene expression is the process by which gene sequences are transcribed into functional gene products such as proteins or functional RNAs (e.g., rRNA, tRNA, small RNA). The process of gene expression is used by all known organisms and can be regulated and modulated at several levels (transcription initiation, splicing, alternative splicing, mRNA stability, post-transcriptional regulation, and eventually translational and posttranslational regulation mechanisms). Gene expression and regulation are the basis of cell development and differentiation. They also allow the cell to adapt to different conditions. By controlling the time, location, and expression level, gene transcripts can have a profound effect on the functions of genes within cells or in multicellular organisms. Genome-wide analysis of gene expression allows for the detection and the quantification of transcript levels for every known gene in the genome of the cell type and organism of interest. The primary use for whole-genome expression analysis is the identification of cell or tissue phenotypes as well as the discovery of novel genes or pathways involved in particular molecular processes and disease development. For example, genome-wide analysis of gene expression can be used to distinguish two different biological states, such as, acute myoblastic leukemia and acute lymphoblastic leukemia, which are relatively difficult to distinguish from one another using conventional molecular and cytological approaches. New genes or pathways induced or repressed in the condition of interest, relative to the reference condition, may provide a better description of the biological processes under study, present new molecular reporters and novel drug targets, predict patient's disease classifications, and also optimize treatments. To analyze gene expression in a genome-wide manner, total RNA is collected from the condition of interest

and from a control condition and analyzed by different genome-wide methodologies (e.g., microarray or sequencing technologies – see described methods below). Newly developed software are then used to examine these large genome-wide expression data projects.

Expressed Sequence Tag

In 1991, the primary tool employed for human gene discovery was a method using expressed sequence tags (ESTs). ESTs are single-pass sequence reads derived from complementary DNA (cDNA). ESTs are typically around 100–800 base pairs (bp) in length, unedited, and randomly selected. Because the ESTs are sequenced only once, the reads are susceptible to errors. Often times, the quality of the base reads for each EST was poor at the beginning of the sequence, improved significantly in the middle, and then diminished again at the end. Phred scores provided a measure of sequence quality of ESTs and were used to extract sequences of a minimal specified quality. Today, EST is no longer widely used because of its inherent drawbacks, which are low-throughput capabilities, redundancy, under-representation, and over-representation.

Serial Analysis of Gene Expression

Serial analysis of gene expression (SAGE) was developed in 1995 by Dr. Victor Velculescu and provides a snapshot of transcript population within a sample by analyzing small tags that correspond to the 3' fragments of messenger RNA (mRNA) (Figure 1). First, cDNA is synthesized from mRNA using biotinylated oligo(dT) primer and then is cleaved with a

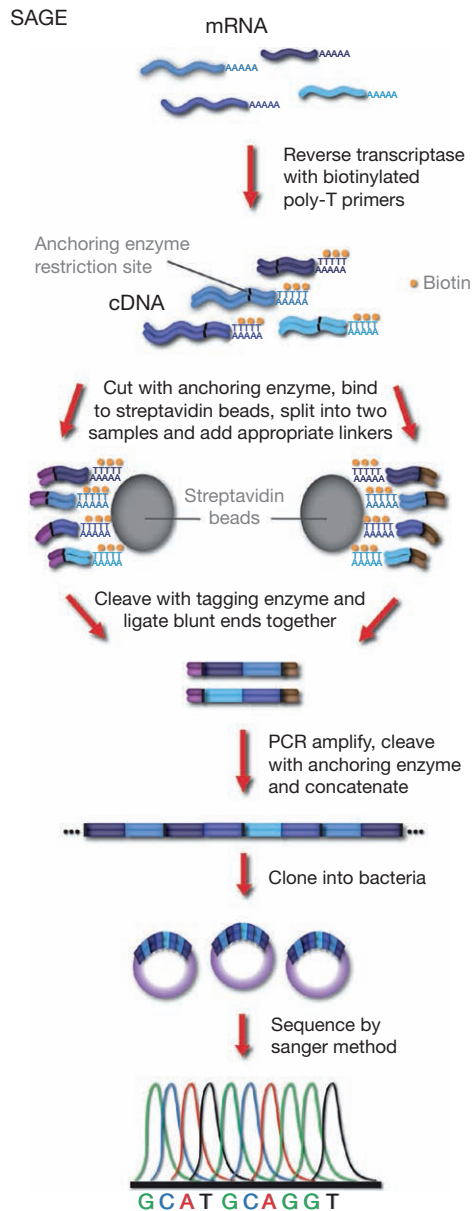


Figure 1 Serial analysis of gene expression (SAGE). RNA is extracted and purified from desired samples. mRNA is converted into double-stranded cDNA by reverse transcriptase with biotinylated poly-T tails as primers. The biotinylated cDNA is fragmented with anchoring enzymes that recognize and cut at their respective restriction sites that are usually found naturally in every few hundred basepairs. The biotinylated ends of the fragmented cDNA are bound to streptavidin beads and split into two portions. Appropriate adaptors with an endonuclease site for tagging enzymes are ligated to each cDNA fragment at the cleaved end. Tagging enzymes are added to cleave a small distance from their endonuclease sites found at each adaptor, releasing the fragmented cDNA (up to 26 bp) from the beads. Blunt ends from each of the two portions of fragmented cDNA are ligated together, forming a ditag, and then amplified by PCR. After PCR, ditags are cleaved at their adaptors, concatenated, and then cloned into plasmids. Once cloned, plasmids are then sequenced, mapped back to the gene and subsequently analyzed for gene expression.

restriction enzyme (anchoring enzyme) that is expected to cut each transcript at least once. The most 3' fragment of the cleaved cDNA is bound to streptavidin beads at its poly(A) tail that is complemented by biotinylated thymidines. Bound cDNA fragments are then ligated at the anchoring cleavage site with linkers that contain a type IIS restriction site. Type IIS endonucleases (tagging enzymes) cleave at a specific distance up to 20 bp away from the recognition site. This process is designed so that cleavage with the tagging enzymes results in the release of the linker along with a short piece of cDNA (tag) of around 9 bp. Released tags are then ligated together at their blunt ends to form a ditag, which serves as a template for polymerase chain reaction (PCR) amplification using primers specific to the linkers. Following PCR amplification, the flanking linkers are cleaved from the ditags using the anchoring enzyme. Cleaved ditags are subsequently concatenated, cloned into a plasmid vector, and then sequenced.

With SAGE, up to 20 000 tags can be sequenced in a single experiment. With full-genome sequence information, the sequenced SAGE tags can be mapped back to the gene. This approach provides for a highly quantitative analysis since the number of tags for a particular gene can be counted. Since the advent of SAGE, several more robust variants, LongSAGE, RL-SAGE, and SuperSAGE, have been developed. These newer variants allow up to 100 000 sequenced tags per trial, thus providing a more comprehensive profile of the transcriptome, and capture longer tags, up to 26 bp, enabling more confident identification of the source gene. Though there have been improvements, SAGE is no longer widely implemented because it is highly labor intensive and is relatively low throughput compared to the newly developed sequencing technologies described below.

DNA Microarray

DNA microarrays consist of an arrayed series of thousands of microscopic spots of DNA oligonucleotides, each containing specific DNA sequences, called 'probes' or 'reporters'. Genes or DNA fragments of interest are amplified with PCR, cleaned up, and printed as a two-dimensional (2D) array using robotic microarraying devices that are able to dispense small quantities of solution at a precise location on a solid support, typically a glass slide due to its low inherent fluorescence. These glass slides are usually coated with hydrophobic poly-lysine or aminosilanes, which limit the spread of DNA on the slides. Other microarray platforms, such as Illumina, use microscopic beads instead of a solid support.

Alternatively, solid-phase *in situ* synthesis of DNA probes from genes of interest can also be implemented in constructing a microarray using a combination of light-directed oligonucleotide chemistry and photolithography. This technology was first developed by Affymetrix but is now available from various other companies such as Agilent. Nucleic acid arrays manufactured from this method allows for high information content with up to 1 000 000 oligonucleotides selected from sequencing databases and precisely synthesized on a small platform (Figure 2). Since these arrays are designed *in silico*, intermediates such as cDNA libraries, PCR, and clones are unnecessary

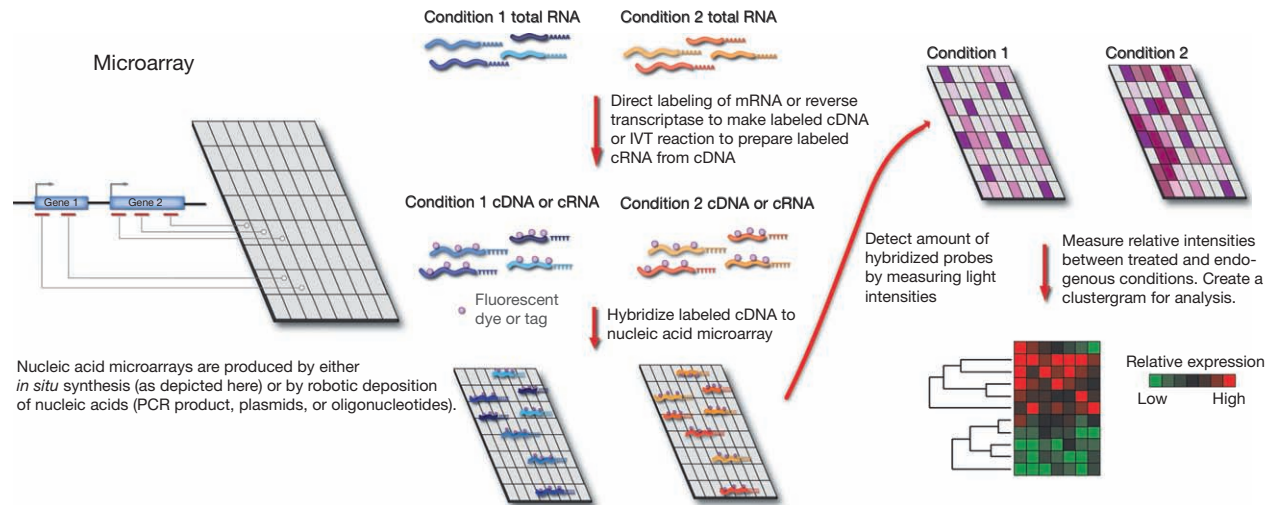


Figure 2 Microarray. Nucleic acid microarrays must first be produced by either *in situ* synthesis or by robotic deposition of nucleic acids. For *in situ* synthesis, multiple probes per gene are placed on the array, while in the case of robotic deposition longer (up to 1000 bp) double-stranded DNA probe(s) is used for each gene. RNA from various samples is extracted and purified. The RNA can be converted into cDNA by reverse transcriptase in order to produce labeled cRNA by *in vitro* transcription (IVT) or have the cDNA labeled with a tag such as biotin or fluorescent dye. In addition, RNA can also be directly labeled. Labeled RNA/cDNA/cRNA are then hybridized to the arrays. After washing, and appropriate secondary fluorescent dyes (if necessary) are added, the intensities of light from the hybridized dyes are detected and recorded. Clustergrams and relative expression levels for each gene are extrapolated by comparing microarray data between various conditions.

and thus the risk of mislabeling is reduced. In addition, this method provides more control over the sample amount deposited onto each location on the array. In order to gain confidence, different multiple probes are synthesized for a given gene.

Transcripts from the sample of interest are converted into cDNA or cRNA (target) and typically labeled with fluorophores or biotin, but chemiluminescent molecules or radioactive isotopes can also be used. Labeled cDNA or cRNA is then hybridized to the DNA microarray under high-stringency conditions. The underlying principle of gene expression level analysis using DNA microarrays is that the transcript level for a particular gene will directly correspond to the amount of hybridization detected on the probe representing that same gene. Probe-target hybridization is detected and quantified using various methods depending on the application. Usually, the background-subtracted hybridization levels from probes representing the same gene are averaged because different oligonucleotide probes have distinct hybridization properties. This approach allows for uses such as distinguishing genes that are highly expressed from those that are lowly expressed.

Tiling arrays are a subtype of microarrays that use short probes that cover the entire length of the genome. Depending on the probe length and spacing, different resolutions can be achieved. The advantage of tiling arrays over traditional microarrays is that the detection of unidentified genes, alternative splice sites, single-feature polymorphisms, and transcript variants are made possible.

Microarray Analysis

High-throughput techniques such as DNA microarrays can produce a large amount of raw data. With proper analytical

approaches, a significant amount of biological information can be extrapolated from microarray data. Clustering is a method of analysis that organizes data and groups genes according to their similarity in gene expression patterns. Since genes with similar biological roles will often have similar patterns of expression, genes involved in related processes are often clustered together. Pairwise average-linkage cluster analysis is a widely used tool that creates hierarchical clusters. Relationships between genes are represented by trees, branch lengths of which reflect the degree of similarity between genes.

Another technique for analyzing microarray data is the ontology method that implements the guilt-by-association approach by inferring possible functions to genes by using prior biological knowledge of other co-expressed genes. Often times, known inhibitors of a biological pathway can be used to elicit the up- or downregulation of clusters of genes, and thus uncharacterized genes can be investigated by their association with other previously known genes that they are co-regulated with. Gene expression analysis by microarray is still being implemented today and is still contributing significantly to our understanding of gene expression and transcript profiles. However, with the advent of newer, cheaper, and more powerful tools such as next-generation sequencing technologies, microarray technology may be phased out in the near future.

Next-Generation Sequencing Technologies

In 1977, Sanger sequencing became available and, since then, has undergone numerous improvements that allowed for automation. Up until recent years, large sequencing projects (e.g., whole genome sequencing of species) have depended on Sanger sequencing. However, in the past decade or so, there

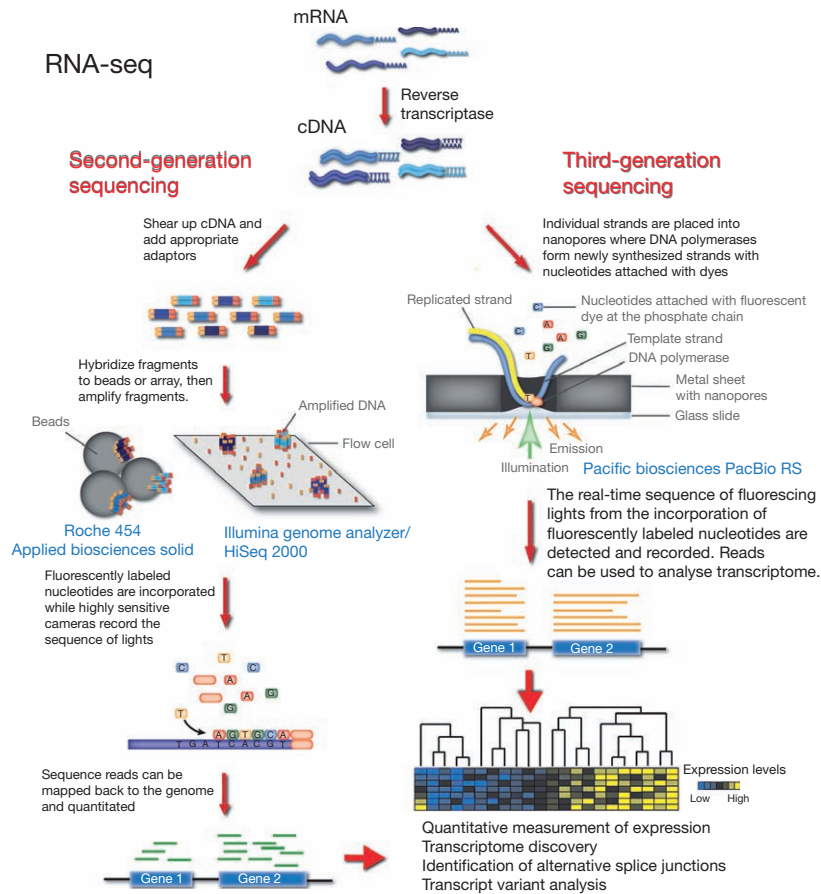


Figure 3 RNA-Seq with second-generation and third-generation sequencing technology. Whole RNA extracts are purified and converted into cDNA. For RNA-Seq using second-generation sequencing (SGS) technology, cDNA is sheared into fragments and have appropriate adaptors added. cDNA fragments are added to either DNA capture beads or flowcells and then PCR amplified creating colonies. Using a series of wash-and-scan cycles, fluorescently labeled nucleotides are incorporated or ligated according to the sequence of the template strand. Highly sensitive cameras are able to detect and record which nucleotide was incorporated by the light signal emitted. Sequence reads are then mapped to the genome for analysis. For RNA-seq using third-generation sequencing (TGS) (only Pacific Biosciences technology is shown here), whole cDNA strands are individually placed into nanowells, and are replicated with fluorescently labeled nucleotides by DNA polymerases. A different color light signal is emitted for each type of nucleotide as they are incorporated into the replicated strand. The series of light signals is detected and recorded by cameras, and then converted into sequence reads, which are then mapped back to genes for analysis.

have been several newly developed sequencing methods, collectively called next-generation sequencing (NGS) (Figure 3), that have both transformed the field of whole genome sequencing and also opened up new applications for sequencing-based approaches such as transcriptome analysis.

In 2005, second-generation sequencing technology (SGS) had become commercially available. Unlike Sanger sequencing, cloning the DNA template into bacterial vectors is unnecessary. On top of saving time and labor, the advantage of bypassing the need for bacterial cloning is that cloning biases, such as cloning difficulties of AT-rich regions and genes that are toxic to bacteria, are circumvented. In addition, SGS provides for a much higher throughput than automated Sanger methods by achieving a tremendous degree of parallelization where up to billions of sequencing reactions within small volumes are taking place at the same time. In most SGS methods, template DNA is sheared into fragments, bound to adaptors, and PCR amplified in order to generate clonal representations of the original template DNA. Tens of thousands of identical PCR-amplified fragments

are usually spatially immobilized on beads or on a given location on an array making up a PCR colony (polony) or clusters. These clusters are read by numerous cycles of washing and scanning operations where sequencing is accomplished by employing various enzymes (e.g., ligases and polymerases) and reagents (e.g., fluorescently labeled nucleotides) to produce a series of fluorescent signals representing the average sequence of an individual cluster. The number of cycles of wash-and-scan determines the length of each read and is usually limited by the amount of noise produced by the incorporation of out-of-phase nucleotides as the read becomes longer. Because of the numerous scan-and-wash cycles, the length of time required to complete a run usually lasts for several days. The massive parallel series of light signals are recorded by highly sensitive detection systems and subsequently decoded to determine the base sequence of the DNA. During the decoding process, also known as 'base calling', each called base is generally given a quality score. Timely analysis and reconstruction of the large amount of sequenced reads are made possible by significant

advancements in computational power and data storage, along with a number of software tools that are available.

Roches' 454 genome sequencer was the first commercially available SGS. Its average read length is around 400–500 bp, which is the longest among the SGS systems available today. It is able to produce 400–600 million bases per run or about 1 billion bases per day. The 454 sequencing method starts with the generation of a library containing single-stranded template DNA with attached adaptors. This library is then immobilized onto DNA capture beads with only one single-stranded DNA fragment per bead. By emulsion PCR, the DNA library is amplified on each bead and then placed into individual wells on a PicoTiterPlate. With the supplementation of sequencing enzymes and materials, the addition of nucleotides complementary to the template DNA results in a base-by-base chemiluminescent light that is detected and recorded by the 454 genome sequencer.

Applied Biosystems' SOLiD system is another SGS system that has a shorter read length of 35–50 bp but is able to generate a high throughput of about 100 gigabases per run (or 8 gigabases per day). Like the 454 genome sequencer, the SOLiD system uses beads to immobilize and PCR-amplify template strands. Once the template strands are amplified and denatured, their 3' ends are modified to allow for covalent bonding to a glass slide. The 3' modified beads are then deposited onto glass slides which allow for high densities of beads per slide and thus a high level of throughput. Primers are hybridized to adaptors at the ends of each template strand, and a set of four fluorescently labeled di-probes is allowed to compete for the primers by interrogating every first and second base of each ligation cycle. Multiple rounds of ligation, detection, and cleavage are the driving mechanism for this sequencing-by-ligation approach. After the first series of ligation cycles, the extension product is removed and the template is reset with the hybridization of primers that are offset by one nucleotide from the previous primer and a subsequent series of ligation reactions takes place. This process is repeated with a total of five primer resets for each sequence tag. For each template fragment, the multiple series of fluorescence produced by the ligation cycles are detected, recorded, and threaded together in order to generate a complete read.

Illumina's (formally known as Solexa) sequencers, such as the Genome Analyser and HiSeq 2000, are capable of producing high-throughput sequencing, up to 200 gigabases per run (or up to 25 gigabases a day) with a read length of 35–150 bp. Template strands are first fragmented and ligated with adaptors. Then the template fragments are hybridized to a flat, optically transparent surface, where they are bridge-amplified to create a highly dense sequencing flowcell containing tens of millions of clusters each containing around 1000 copies of the same template fragment. Illumina implements a sequencing-by-synthesis approach by using reversible terminators with removable four-color fluorescent dyes to generate a base-by-base sequence. Each type of nucleotide is distinguished by a different color. Fluorescence detection is accomplished with laser excitation and total internal reflection optics.

The Polonator is an NGS that is based upon sequencing-by-ligation that has an open-source platform where users can use off-the-shelf reagents and freely download protocols and software. The open-source approach allows for users to modify protocols, reagents, and software to custom-fit their sequencing

application. However, the Polonator has the shortest read lengths when compared to other NGS systems.

In addition to genome sequencing, NGS systems are applied to sequence cDNA derived from the whole transcriptomes of an organism in order to analyze expression profiles, transcript boundaries, intron/exon junctions, alternative expression, and also discover novel transcripts in a high-throughput manner. The obtainment of transcriptomics data by NGS is collectively known as deep RNA sequencing or RNA-seq. In 2008, the implementation of RNA-seq was first reported in yeast, and then was quickly adopted for use of transcriptome analysis in several other organisms. The advantages of RNA-seq over previous analytical sequencing tools such as EST and SAGE are that RNA-seq has a much higher throughput and increased sequencing depth, which leads to better resolution of expression profiles. Though microarray has a genome-wide coverage, nucleic acid arrays are not available for all organisms, while RNA-seq can be used on any organism. In addition, microarray expression data can be biased by the binding affinities of each individual probe. Another advantage is that RNA-seq is able to identify transcript boundaries and intron/exon junctions at a single-nucleotide resolution.

Third-Generation Sequencing

More recently, a few new companies, such as Helicos Biosciences and Pacific Biosciences, are developing advanced sequencing methods that do not require template amplification but rather directly sequence from single template molecules. These systems are commonly referred to as third-generation sequencing (TGS) platforms (Figure 3). The advantage of directly sequencing from single template molecules is that clonal duplication and amplification biases due to difficulties in amplifying certain regions of template strands are avoided. Also, the amplification process can introduce errors in the replicated template strands.

In the Helicos system, DNA or cDNA is sheared and turned into single-stranded molecules. Fluorescently labeled poly-A adaptors are added to the ends of the sheared strands and then hybridized to poly-T oligonucleotide capture-strands bound to a flowcell. A highly sensitive camera that is able to detect single-molecule fluorescence captures the locations of each of the billions of hybridized template strands. After the fluorescent label is removed from the adaptors, a series of wash-and-scan cycles is initiated; a single type of fluorescently labeled nucleotides and polymerases are washed over the flowcell, followed by fluorescence detection by imaging and removal of the fluorescent group. The wash-and-scan cycle is then repeated with the other three nucleotides. Multiple four-base cycles result in reads of around 25 bp.

Pacific Biosciences has developed a real-time single-molecule sequencing system, PacBio RS (Figure 3), that is capable of 30 minute sequencing runs because it does not require the conventional wash-and-scan method, which can normally take up to several days. A thin 100-nm-thick metal film is perforated with an array of nanoscopic holes and bound to a glass substrate, creating wells that are tens of nanometers in diameter. A laser light is shined through the glass substrate, illuminating only the bottom 30 nm of each well. Each well

typically has a single DNA polymerase anchored to the glass bottom and is small enough to allow for only one template strand. Fluorescently labeled nucleotides that have different fluorescent colors depending on the nucleotide are washed over the nanoscopic wells and are incorporated into the sequencing strand by the DNA polymerase. These nucleotides have fluorescent dyes attached to the phosphate chain instead of the base, thus preventing the inactivation of the DNA polymerase. As the fluorescently labeled nucleotides are incorporated one-by-one via the DNA polymerase, the fluorescent dye is excited by the laser, transmits a color signal that lasts milliseconds, and then is cleaved off by the DNA polymerase to allow for additional rounds of nucleotide incorporation and fluorescence. The series of color fluorescence is detected and decoded to produce reads of over 1000 bp.

Analytical Tools for NGS Expression Data

The high throughput and deepness of quantitative measurements produced by NGS technology comes at the price of examining millions to billions of reads. Though the first users have created their own computer programs to study the numerous reads, a new breed of sophisticated algorithms and software tools has emerged to help analyze the massive data sets produced by RNA-seq. These analytical tools are used for transcript discovery, RNA quantification and mapping, and also identification of transcript variants and splice sites.

With a given gene model, the total number of mapped reads can be tallied for a particular gene as a measure of gene expression. The number of reads is naturally a function of the molar concentration of mRNA but must be normalized according to the length of each mRNA. A common approach is to normalize the read count by the length of the transcript and the number of millions of mappable reads to obtain reads per kilobase per million (RPKM) values. Within a given sample, RPKM values for genes are directly comparable to give relative expression values. However, RPKM values vary depending on the software package employed and can also be influenced by experimental factors such as quality of input RNA, the amount of ribosomal RNA remaining in the sample, size selection procedures, and the accuracy of the gene model used. Amplification biases can also obstruct uniform sequence coverage across the transcriptome, though TGS technologies circumvent this problem by bypassing the amplification process altogether.

In addition to quantifying gene expression, RNA-seq allows for the investigation of the correspondence between gene and

transcript model. With alternative splicing, alternative promoter employment, and differing 3' poly (A) tail sites, various types of transcripts can be derived by a single gene. RNA-seq is capable of using reads that map across splice junctions to identify splice sites and also discover alternatively spliced transcript isoforms. Using more advanced software, users are capable of extracting more information from their RNA-seq data, such as analysis of transcript modeling, quantifying specific isoforms, and investigating sequence variance due to single-nucleotide polymorphisms, private mutation, or RNA editing. Bioinformatic tools are constantly being developed in order to match the quickly evolving sequencing technology.

Though current analytical tools of investigating transcripts on a genome-wide level have their own distinct challenges, they offer more advantages than previous methods such as higher-throughput, wider-coverage, and increased sensitivity. Today, with RNA-seq technology, we are able to quantify differences in gene expression at a single-nucleotide resolution across multiple samples and conditions. Though we still have much to learn and discover, we draw closer to understanding how genomes are capable of encoding diverse patterns of gene expression that define each cell type/stage and also elicit a distinct response from various stimuli/conditions.

See also: **Molecular Biology:** Alternative Splicing; Gene Expression in Eukaryotes; RNA Polymerase II Structure; RNA Polymerase II and Its General Transcription Factors; T7 RNA Polymerase.

Further Reading

- Adams MD, Kelley JM, Gocayne JD, et al. (1991) Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science* 252: 1651–1656.
- DeRisi JL, Iyer VR, and Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278: 680–686.
- Nagaraj SH, Gasser RB, and Ranganathan S (2007) A hitchhiker's guide to expressed sequence tag (EST) analysis. *Briefings in Bioinformatics* 8: 6–21.
- Nagarajan N and Pop M (2010) Sequencing and genome assembly using next-generation technologies. *Methods in Molecular Biology* 673: 1–17.
- Pepke S, Wold B, and Mortazavi A (2009) Computation for ChIP-seq and RNA-seq studies. *Nature Methods* 6: S22–S32.
- Schadt EE, Turner S, and Kasarskis A (2010) A window into third-generation sequencing. *Human Molecular Genetics* 19: R227–R240.
- Schena M, Shalon D, Davis RW, and Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467–470.
- Velculescu VE, Zhang L, Vogelstein B, and Kinzler KW (1995) Serial analysis of gene expression. *Science* 270: 484–487.

Giant Mitochondria (Megamitochondria)

B Tandler, Veterans Affairs Medical Center, Cleveland, OH, USA

C L Hoppel, Case Western Reserve University School of Medicine, Cleveland, OH, USA; Veterans Affairs Medical Center, Cleveland, OH, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 186–188, © 2004, Elsevier Inc.

Glossary

Ariboflavinosis A deficiency of riboflavin in the diet.

Crista (pl. cristae) Any of numerous infoldings of the inner mitochondrial membrane in eukaryotic cells.

Nebenkern A globular structure in a spermatid that results from interlocking and fusion of the mitochondria during spermatogenesis.

Megamitochondrial Structure

Under some conditions, the megamitochondria are virtual simulacra of their normal size counterparts; that is, on a unit mitochondrial area basis they display the same number and arrangement of cristae as do normal mitochondria (Figure 1), although in some cases they may contain within their matrix a small stack of closely packed, short cristae. In other conditions, the megamitochondria have very short cristae that protrude inward from the boundary membrane in a manner not unlike a picket fence. In still others, the cristae are very rare, with the megamitochondria consisting of a morphologically conventional outer membrane and a smooth inner boundary membrane that is almost devoid of cristae. In all types of megamitochondria, the inner compartment contains moderately dense, homogeneous material. In some cases of human

thyroid tumors, the matrix compartment contains a jumble of dense ribbons of undefinable length. In certain types of giant mitochondria, an expanded crista contains a bundle of helical filaments. Filaments of this type occurring in normal hepatic mitochondria consist of three distinct, but otherwise uncharacterized, proteins.

It is well established that normal mitochondria replicate by division, but the morphological pathway by which this process is accomplished was unclear. The megamitochondrial model has been used successfully to settle this point. Based on restoration to normal size of experimentally induced megamitochondria, there appear to be two distinct pathways in mitochondrial division; during recovery, only one of the two mechanisms is called upon, depending on the agent that evoked the

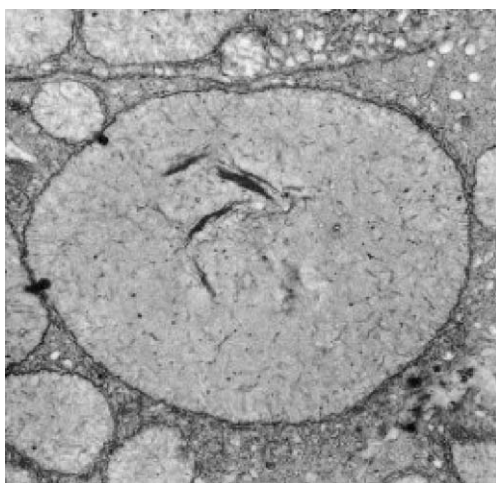


Figure 1 Transmission electron micrograph of a hepatic megamitochondrion induced in a mouse by riboflavin deficiency. Note the typical distribution of cristae and the presence of dense matrix granules. The diameter of this organelle is approximately 6 times as large as that of normal hepatic mitochondria. At the left, the large mitochondrion is linked to two smaller organelles by means of small myelin figures, indicating that it is still in the process of enlargement. Original magnification, $\times 9000$.

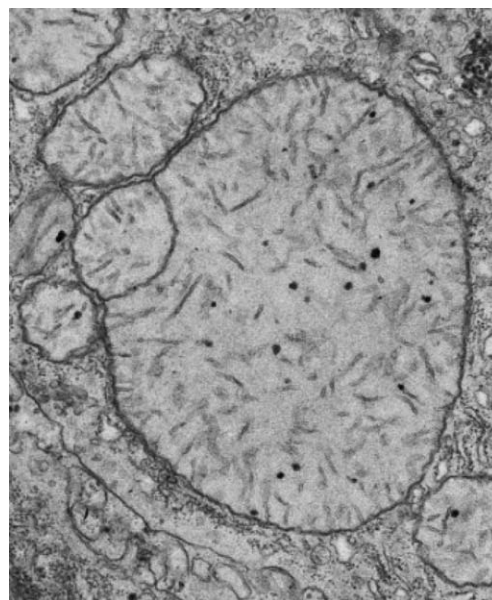


Figure 2 Transmission electron micrograph of a dividing hepatic megamitochondrion during recovery from ariboflavinosis. A small bud is subtended by a membranous partition; ingrowth of the outer membrane ultimately leads to the separation of the bud, which is equal in size to a typical mitochondrion, from the enlarged organelle. Repetition of this process results in mitochondrial normalization in terms of size. Original magnification, $\times 34\,000$.

megamitochondria in the first place. The first mechanism involves the formation of a membranous partition, probably of boundary membrane origin, that spans the organelle, frequently at the base of a bud (Figure 2). This is followed by ingrowth of the outer membrane into the partition. Like a closing iris diaphragm, the outer membrane eventually completes its incursion, and two daughter mitochondria result. Repetition of this process ultimately leads to the disappearance of megamitochondria and to the appearance of organelles of normal size. The second pathway involves a simple elongation and attenuation of the equator of the megamitochondrion until membranes from opposite sides of the organelle touch and fuse. In effect, the giant mitochondrion has been pulled apart. The progeny of such division repeatedly undergoes the same process to finally restore the mitochondria to their accustomed dimensions. What controls the selection of a particular pathway for mitochondrial division is completely unknown.

See also: **Bioenergetics:** Cell Death by Apoptosis and Necrosis; **Protein/Enzyme Structure Function and Degradation:** Flavins

Further Reading

- Hoppel CL and Tandler B (1993) Megamitochondria. *Methods in Toxicology* 2: 191–206.
- Karbowski M, Kurono C, Wozniak M, et al. (1999) Free radical-induced megamitochondria formation and apoptosis. *Free Radical Biology and Medicine* 26: 396–409.
- Tandler B, Dunlap M, Hoppel CL, and Hassan M (2002) Giant mitochondria in a cardiomyopathic heart. *Ultrastructural Pathology* 26: 1–7.
- Tandler B, Erlandson RA, Smith AL, and Wynder EL (1969) Riboflavin and mouse hepatic cell structure and function. II. Division of mitochondria during recovery from simple deficiency. *Journal of Cell Biology* 41: 477–493.
- Tandler B, Erlandson RA, Wynder EL, et al. (1968) Riboflavin and mouse hepatic cell structure and function. I. Ultrastructural alterations in simple deficiency. *American Journal of Pathology* 52: 69–95.
- Tandler B and Hoppel CL (1973) Division of giant mitochondria during recovery from cuprizone intoxication. *Journal of Cell Biology* 56: 266–272.
- Tandler B and Hoppel CL (1986) Studies on giant mitochondria. *Annals of the New York Academy of Sciences* 488: 65–81.

Gi Family of Heterotrimeric G Proteins

M E Linder, Cornell University, Ithaca, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Chemokine Small, secreted proteins that guide the migration of white blood cells.

Effector A protein whose activity is changed in response to a stimulus.

Hormone A circulating peptide or chemical that acts on target organs at a distance from its site of synthesis.

Neurotransmitter A chemical messenger released by a nerve cell at a synapse.

Receptor A protein at the cell surface or in the cytoplasm that binds a specific small molecule or protein and activates other proteins to produce a cellular response.

Second messenger A small molecule synthesized intracellularly in response to an extracellular stimulus that propagates the signal by binding to an effector and modulating its activity.

Overview of the Gi Family

G_i was discovered in the early 1980s as the G protein that mediates hormonal inhibition of cyclic adenosine monophosphate (cAMP) production. Its discovery followed that of G_s, the G protein that stimulates cAMP production and G_t, the G protein that regulates cyclic guanosine monophosphate (cGMP) levels in the retina. Characterization of G_s, G_i, and G_t revealed a structurally homologous family of guanine-nucleotide-binding proteins that are heterotrimers with subunits designated α , β , and γ . The α subunit contains the guanine-nucleotide-binding site. G protein α subunits act as molecular switches between an active and inactive state depending on whether GDP or GTP is bound. GTP-bound α is the active form and regulates downstream targets. The GDP-bound form of the α subunit is the inactive state.

In mammals, there are 16 different genes that encode G protein α subunits that range in size from 39 to 45 kDa. G _{α} proteins encoded by these genes can be divided into four families based on function and primary amino acid sequence homology. These are G_s, G_i, G_q, and G_{12/13}. The G_i family is the most functionally diverse of the G-protein families. It can be subdivided into five groups – G_{i1}, G_o, G_{i2}, G₁₂, and G₁₃ – according to the identity of the α subunit. The evolutionary relationship among the G_i family α subunits is shown in Figure 1. G _{α} subunits associate with G _{$\beta\gamma$} dimers composed of combinations of the five G _{β} and 12 G _{γ} subunits. Although not all combinations are possible, a large number of heterotrimers can be generated within this family. Like all heterotrimeric G proteins, G_i family members undergo a cycle of nucleotide exchange and hydrolysis when activated by receptors (Figure 2). Binding of GTP to G _{α} induces dissociation from G _{$\beta\gamma$} and both G _{α} -GTP and G _{$\beta\gamma$} can interact with effector molecules and regulate their activity. Although, in principle, all G-protein families can signal through G _{$\beta\gamma$} subunits, this mode of signaling is most prevalent with signaling pathways regulated by G_{i1}, G_o, and G₁₂.

G-protein pathways are deactivated when G _{α} hydrolyzes GTP to GDP. In the GDP-bound form, G _{α} binds efficiently to G _{$\beta\gamma$} and restores the heterotrimer to its inactive state. The length of time that G _{α} is in its GTP-bound state determines the duration of the

signal. Regulators of G-protein signaling (RGS) bind to G _{α} and accelerate the rate of GTP hydrolysis by acting as GTPase activating proteins (GAPs). Thus, RGS proteins are important regulators of the timing of G-protein responses. Members of the G_i family are targets for RGS proteins.

Signaling through G _{$\beta\gamma$}

Many receptors coupled to G_i and G_o exert their effects through signals propagated to downstream effectors by G _{$\beta\gamma$} subunits. Similar to G _{α} subunits, G _{$\beta\gamma$} subunits regulate the activity of enzymes involved in the generation of second messengers. These are adenylyl cyclase, phospholipase C β , and phosphatidylinositol 3-kinase. G _{$\beta\gamma$} exhibits type-specific regulation of adenylyl cyclase, inhibiting some isoforms, but activating others synergistically with G _{α} . G _{$\beta\gamma$} activates certain isoforms of phospholipase C β , an enzyme that converts phosphatidylinositol into diacylglycerol and inositol trisphosphate. The second messenger inositol trisphosphate induces the release of calcium from intracellular stores, whereas diacylglycerol recruits protein kinase C to the plasma membrane. Phosphatidylinositol 3-kinase (PI3K) is also activated by G _{$\beta\gamma$} subunits. The generation of phosphatidylinositol 3,4,5-trisphosphate (PIP₃), the product of PI3K, mediates important cellular responses. Production of PIP₃ in white blood cells is an essential step in inducing the cells to migrate to sites of cell injury. PI3K activation by G _{$\beta\gamma$} is also associated with signaling pathways that regulate cell survival.

In addition to regulating second messengers, G _{$\beta\gamma$} subunits modulate the activity of numerous ion channels. The best characterized are potassium channels in the heart which slow heart rate when activated. G _{$\beta\gamma$} subunits bind directly to the channel and cause it to open (Figure 2). Certain voltage-gated calcium channels are inhibited by binding G _{$\beta\gamma$} subunits.

The effectors described above are regulated through direct binding of the G _{$\beta\gamma$} subunits. G _{$\beta\gamma$} is also an important regulator of mitogen-activated protein (MAP) kinase pathways. The direct target of G _{$\beta\gamma$} in these pathways has not been identified.

Gi

The founding member of the family, G_i , is named for its ability to inhibit the enzyme adenylyl cyclase. There are three forms of $G_{i\alpha}$, designated $G_{i\alpha1}$, $G_{i\alpha2}$, and $G_{i\alpha3}$, each encoded by a separate

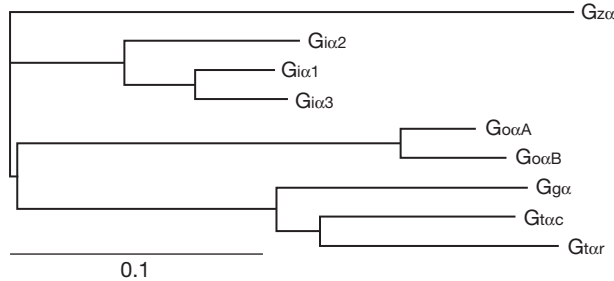


Figure 1 Sequence comparison of $G_{i\alpha}$ family members. A nearest-neighbor dendrogram of $G_{i\alpha}$ family sequences. The scale bar represents a measure of evolutionary distances between the sequences analyzed.

gene. The three isoforms share many functions and will be considered together here. Heterotrimeric G_i couples to many different receptors, including those for catecholamines, neurotransmitters, and chemokines. Many of the effects associated with activation of G_i by receptors are mediated by $G_{\beta\gamma}$ subunits as discussed above. The best-characterized effector for $G_{i\alpha}$ is adenylyl cyclase. $G_{i\alpha}$ binds to this enzyme and inhibits its activity, thereby reducing intracellular levels of the second messenger cAMP (Figure 2). $G_{i\alpha}$ (and $G_{o\alpha}$) can propagate signals to protein kinases (e.g., the MAP kinase pathway or c-Src), but the direct target of $G_{i\alpha}$ in these pathways is unknown.

$G_{i\alpha}$'s are ubiquitously expressed and most cell types contain more than one form. Eliminating the gene for $G_{i\alpha1}$ or $G_{i\alpha3}$ in mice does not have any obvious effects on the organism's ability to develop or to reproduce. Mice lacking a functional gene for $G_{i\alpha2}$ appear to develop normally, but exhibit growth retardation, inflammatory bowel disease, and are prone to developing adenocarcinomas, demonstrating that $G_{i\alpha2}$ is essential for normal function of the intestine. The relatively mild phenotypes associated with inactivation of the individual

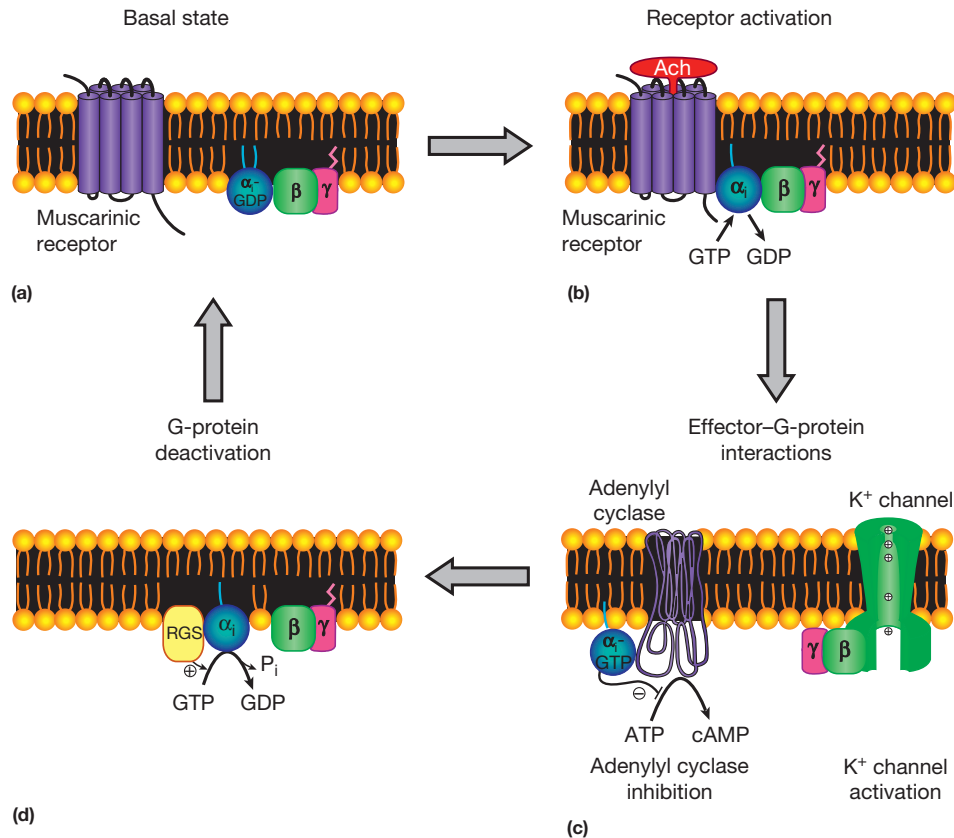


Figure 2 G_i signaling in the heart. In heart cells, the neurotransmitter acetylcholine elicits a slowing of the heart rate by activating a G_i -coupled signaling pathway. (a) In the basal state, $G_{i\alpha}$ is in the GDP-bound state and associated with $G_{\beta\gamma}$. The receptor is unoccupied and effectors (not shown) are inactive. (b) The neurotransmitter acetylcholine is released at the terminal of the vagus nerve that innervates the heart. Binding of acetylcholine activates the receptor, which, in turn, stimulates the release of GDP from $G_{i\alpha}$. GTP is present intracellularly at high concentration and quickly diffuses into the guanine-nucleotide-binding pocket on $G_{i\alpha}$. GTP binding induces a conformational change in $G_{i\alpha}$, causing it to dissociate from $G_{\beta\gamma}$. (c) $G_{i\alpha}$ -GTP interacts with adenylyl cyclase in the plasma membrane and inhibits its activity. $G_{\beta\gamma}$ subunits bind to the K^+ channel, causing it to open. The cell membrane becomes hyperpolarized, slowing action potential frequency and heart rate. (d) $G_{i\alpha}$ hydrolyzes its GTP to GDP to deactivate the pathway. The rate of GTP hydrolysis is accelerated by $G_{i\alpha}$ interaction with an RGS protein. Once in its GDP-bound state, $G_{i\alpha}$ reassociates with $G_{\beta\gamma}$ and the system is returned to the basal state.

Table 1 $G_{i\alpha}$ family members

Protein	Gene name	% Amino acid identity ^a	PTX substrate ^b	Fatty acylation ^c	Major tissue distribution
$G_{i\alpha 1}$	GNAI-1	100	Yes	Myr, palm	Nearly ubiquitous
$G_{i\alpha 2}$	GNAI-2	87	Yes	Myr, palm	Ubiquitous
$G_{i\alpha 3}$	GNAI-3	93	Yes	Myr, palm	Nearly ubiquitous
$G_{o\alpha A}$	GNAO	70	Yes	Myr, palm	Brain, neuroendocrine, heart
$G_{o\alpha B}$	GNAO	69	Yes	Myr, palm	Brain, neuroendocrine, heart
$G_{t\alpha f}$	GNAT1	67	Yes	Myr	Retinal rods
$G_{t\alpha c}$	GNAT2	69	Yes	Myr	Retinal cones
$G_{g\alpha}$	GNAT3	68	Yes	Myr	Taste buds
$G_{z\alpha}$	GNAZ	66	No	Myr, palm	Brain, adrenal, platelets, pancreas

^aAmino acid sequence (human) comparison with $G_{i\alpha 1}$.

^bPTX, pertussis toxin, see text.

^cMyr, amide-linked myristate; palm, thioester-linked palmitate; see text.

$G_{i\alpha}$ genes probably reflect an ability of one $G_{i\alpha}$ subtype to compensate for the function of another.

G_o

$G_{o\alpha}$ is a close relative of the $G_{i\alpha}$'s, with 70% sequence identity to $G_{i\alpha 1}$ (Table 1). Two forms of $G_{o\alpha}$ ($G_{o\alpha A}$ and $G_{o\alpha B}$) are generated by alternative splicing of messenger RNA (mRNA). G_o is very abundant in the nervous system where it represents 0.5% of total membrane protein. It is also expressed in neuroendocrine tissues and, at low levels, in the heart. In general, receptors that are coupled to G_o bind neurotransmitters that mediate inhibitory, rather than excitatory, responses in the central nervous system. The best-studied effectors regulated by G_o are voltage-gated calcium channels, which are inhibited by $G_{\beta\gamma}$ subunits released upon receptor activation of G_o . Direct effectors of $G_{o\alpha}$ are not as well characterized. Inactivation of the $G_{o\alpha}$ gene in mice results in central nervous system dysfunction and abnormal regulation of calcium channels in the nervous system and heart.

G_z

$G_{z\alpha}$ is the most distantly related member of the $G_{i\alpha}$ family based on sequence comparison (Figure 1). Its expression pattern is more restricted than $G_{i\alpha}$, with abundant expression in brain, neuroendocrine tissues, and platelets. G_z has several properties that distinguish it from the other members of the G_i family. It is the only family member that is not a substrate for adenosine diphosphate (ADP)-ribosylation by pertussis toxin (see below). $G_{z\alpha}$ also has an intrinsic rate of GTP hydrolysis that is much slower than G_i or G_o . Thus, G_z pathways may be slow to turn off unless regulated by RGS proteins. $G_{z\alpha}$ interacts with several RGS family members that show marked specificity for $G_{z\alpha}$ over other G-protein α subunits. The complement of receptors that G_z responds to overlaps extensively with those that couple to $G_{i\alpha}$ and $G_{o\alpha}$. Given its abundant expression in brain, G_z is coupled to many neurotransmitters and can regulate ion channels, adenylyl cyclase, and MAP kinase pathways similar to G_i and G_o . $G_{z\alpha}$ -GTP binds directly to Rap1GAP, a GAP for the monomeric GTPase Rap1, an interaction that may modulate neuronal differentiation.

The gene for $G_{z\alpha}$ has been knocked out in mice. The $G_{z\alpha}$ -deficient mice develop normally and are fertile. However, they exhibit altered responses to psychoactive drugs and a mild impairment in platelet function. These phenotypes are consistent with the restricted distribution of $G_{z\alpha}$ in neural tissues and platelets. Mice deficient for $G_{z\alpha}$ also display increased pancreatic islet insulin secretion and correspondingly increased glucose clearance following glucose challenge, likely due to a constitutive increase in islet cAMP levels.

G_t (Transducin)

G_t is the G protein that mediates the visual signal transduction cascade. It is also known as 'transducin'. There are two closely related forms of $G_{t\alpha}$ that are encoded by separate genes. $G_{t\alpha p}$ is expressed in retinal rods, whereas $G_{t\alpha c}$ is found in cone cells. $G_{t\alpha p}$ and $G_{t\alpha c}$ both associate with specific $G_{\beta\gamma}$ complexes: $G_{t\alpha p}$ with $G_{\beta 1\gamma 1}$ and $G_{t\alpha c}$ with $G_{\beta 1\gamma 14}$. The receptor in the visual signal transduction cascade is rhodopsin, which harbors a covalently attached 11-*cis*-retinal. When a photon of light is absorbed by the retinal, it undergoes a *cis-trans* isomerization that induces a conformational change in rhodopsin. The activated receptor then promotes nucleotide exchange and subunit dissociation of G_t . $G_{t\alpha}$ -GTP propagates the signal by stimulating cGMP phosphodiesterase (PDE), an enzyme that hydrolyzes cGMP. PDE consists of two catalytic subunits and two inhibitory subunits (called PDE_{γ}). $G_{t\alpha}$ binds PDE_{γ} and relieves the inhibitory constraint on the catalytic subunits. Reduction of cellular levels of cGMP causes closure of cGMP-gated ion channels in the plasma membrane. This reduces the influx of Na^+ and Ca^{2+} , increasing the negative charge inside the cell, and causing the cell to become hyperpolarized. The net effect is a change in the signals that photoreceptors send to the brain.

An important mechanism for deactivating the phototransduction process is the hydrolysis of GTP by $G_{t\alpha}$. The intrinsic rate of GTP hydrolysis measured for $G_{t\alpha}$ *in vitro* is too slow to mediate the physiological response. *In vivo*, the GTPase rate of $G_{t\alpha}$ is accelerated by the GAP activity of a protein complex of the retinal RGS protein RGS9 and its binding partner $G_{\beta 5L}$. RGS9 has a G_{γ} -like domain that mediates its association with $G_{\beta 5L}$. The effector molecule PDE_{γ} also promotes GAP activity

by increasing the affinity of G_{α} for the RGS9/ $G_{\beta 5L}$ complex. Thus, timely inactivation of G_i is achieved through cooperation between RGS9- $G_{\beta 5L}$ and PDE $_{\gamma}$.

Mutations in the genes that encode $G_{\alpha_{tr}}$ and $G_{\alpha_{tc}}$ result in visual impairment in humans. A form of congenital night blindness is caused by a missense mutation in GNAT1 ($G_{\alpha_{tr}}$). Achromatopsia, also referred to as 'rod monochromacy', is an autosomal recessive disorder characterized by total colorblindness, low visual acuity, photophobia, and involuntary eye movement. GNAT2 ($G_{\alpha_{tc}}$) is the third gene identified with mutations associated with this disorder.

G_g (Gustducin)

G_g is a signal transducer for signaling pathways that mediate taste. Its expression is restricted primarily to taste buds. $G_{g\alpha}$ is closely related to $G_{\alpha_{tr}}$ and $G_{\alpha_{tc}}$ (Figure 1). Gustducin is important in recognition of sweet, bitter, and umami, the taste associated with savory meat. G_g couples receptors for bitter and sweet tastants to specific taste cell PDEs. Transducins and gustducins therefore share a common family of effectors. Taste receptors also elicit the activation of phospholipase C- $\beta 2$ through the action of $G_{\beta\gamma}$ subunits. In mice deficient for $G_{g\alpha}$, the animals show impaired responses to bitter and sweet agents and umami, but are normal in their responses to sour and salty tastants. Thus, animals use a different signal transduction pathway to experience sour and salty taste.

Posttranslational Modifications

Members of the G_i family are subject to a number of posttranslational modifications (Table 1). All family members except $G_{\alpha_{zz}}$ are modified with ADP-ribose at a cysteine residue near the C-terminus by a bacterial toxin from *Bordetella pertussis*. The C-terminus of G_{α} is an important site of interaction with G-protein-coupled receptors. Modification of G_{α} with the bulky ADP-ribose group interferes with receptor-G-protein coupling. Therefore, cells treated with pertussis toxin cannot respond to ligands that bind to G_i -coupled receptors. Pertussis toxin has been a useful experimental tool to define cellular responses that are mediated by members of the G_i family.

All G-protein α subunits are modified with fatty acids at or near the amino terminus. Fatty acylation provides a hydrophobic anchor that mediates binding to the plasma membrane. Members of the G_i family (but not G_s , G_{12} , or G_q families) are N-myristoylated. Myristic acid is added through an amide linkage to a glycine residue exposed after removal of the

initiator methionine. This is the only modification found on $G_{\alpha_{tr}}$ and $G_{\alpha_{g\alpha}}$. However, $G_{\alpha_{tr}}$, $G_{\alpha_{g\alpha}}$, and $G_{\alpha_{zz}}$ undergo a second type of fatty acylation, S-palmitoylation, which occurs at a cysteine residue adjacent to the N-myristoylated glycine. Whereas N-myristoylation is a stable modification, S-palmitoylation is reversible and can be regulated.

Heterotrimeric G proteins are generally not regulated by protein phosphorylation and dephosphorylation, in contrast to many signaling proteins. $G_{\alpha_{zz}}$ is an exception. It is phosphorylated by protein kinase C and p21-activated kinase 1. Phosphorylation of $G_{\alpha_{zz}}$ reduces its ability to bind $G_{\beta\gamma}$ and interact with RGS proteins, thereby prolonging $G_{\alpha_{zz}}$ activation.

See also: **Signaling: Adenylyl Cyclases; Cyclic GMP Phosphodiesterases; G Protein Signaling Regulators; G₁₂/G₁₃ Family; G_q Family.**

Further Reading

- Arshavsky VY, Lamb TD, and Pugh EN (2002) G proteins and phototransduction. *Annual Review of Physiology* 64: 153–187.
- Clapham DE and Neer EJ (1997) G protein $\beta\gamma$ subunits. *Annual Review of Pharmacology and Toxicology* 37: 167–203.
- Gilman AG (1995) Nobel lecture. $G_{\beta\gamma}$ proteins and regulation of adenylyl cyclase. *Bioscience Reports* 15: 65–97.
- Gomperts BD, Kramer IM, and Tatham PER (2009) *Signal Transduction*, 2nd edn. New York: Academic Press.
- Ho MKC and Wong YH (2001) G α signaling: Emerging divergence from Gi signaling. *Oncogene* 20: 1615–1625.
- Ishimaru Y (2009) Molecular mechanisms of taste transduction in vertebrates. *Odontology* 97: 1–7.
- Krauss G (2008) *Biochemistry of Signal Transduction and Regulation*, 4th edn. New York: Wiley.
- Margolske RF (2002) Molecular mechanisms of bitter and sweet taste transduction. *Journal of Biological Chemistry* 277: 1–4.
- Marks F, Klingmüller U, and Müller-Decke K (2009) *Cellular Signal Processing: An Introduction to the Molecular Mechanisms of Signal Transduction*. Oxford: Garland Science, Taylor and Francis Group LLC.
- Nelson J (2008) *Structure and Function in Cell Signaling*. New York: Wiley.
- Neves SR, Ram PT, and Iyengar R (2002) G protein pathways. *Science* 296: 1636–1639.
- Offermans S (2001) *In vivo* functions of heterotrimeric G proteins: Studies in G α -deficient mice. *Oncogene* 20: 1635–1642.

Relevant Websites

- <http://stke.sciencemag.org> – Science Signaling: The Signal Transduction Knowledge Environment.
- <http://www.signaling-gateway.org> – thesignalinggateway: Molecule Pages.

Gluconeogenesis

R W Hanson, Case Western Reserve University School of Medicine, Cleveland, OH, USA

O E Owen[†], Southern Illinois University School of Medicine, Springfield, IL, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Anaplerotic reactions Enzymatic reactions that produce citric acid cycle anions to replace those that leave the cycle for the synthesis of compounds such as glucose and fatty acids. The major anaplerotic enzyme is pyruvate carboxylase.

Cataplerotic reactions Enzymatic reactions that remove citric acid cycle anions from the mitochondria for biosynthetic processes (i.e., gluconeogenesis) to insure the balance of intermediates in the cycle. The major cataplerotic enzyme is phosphoenolpyruvate carboxykinase.

Cori cycle Process in which the lactate generated in the red blood cells and muscle is converted into glucose in the liver and cycled back to these tissues for reconversion into lactate.

Gluconeogenesis The *de novo* synthesis of glucose from nonhexose precursors.

Ketone bodies Water-soluble fuels produced by the liver from the acetyl coenzyme A (CoA) that is generated from the β -oxidation of fatty acids. The three ketone bodies, acetoacetate, β -hydroxybutyrate, and acetone, accumulate in the blood of mammals during periods of fasting.

Lipolysis The degradation of triglyceride to free fatty acids and glycerol; this occurs primarily in the white adipose tissue.

Proteolysis The breakdown of proteins to their component amino acids.

β -Oxidation The oxidation of fatty acids to acetyl CoA in the mitochondria.

History

Claude Bernard first noted the synthesis of glucose in his famous description of glucose turnover during fasting. However, it is the work of Carl and Gerty Cori that established gluconeogenesis as critical for the synthesis of glucose from lactate as part of what is now known as the Cori cycle. Gluconeogenesis is not a reversal of glycolysis, although the two pathways share a number of enzymes. A major advance in the understanding of the specific reactions in the pathway of glucose synthesis took place when Merton Utter and his colleagues discovered two of the major enzymes of the pathway, phosphoenolpyruvate carboxykinase (PEPCK) and pyruvate carboxylase (PC). In the 1960s and 1970s, the research of Hans Krebs, Charles Park and John Exton, John Williamson and Henry Lardy, among others, set out the metabolic parameters of the pathway as it is known today. Finally, the elevated rate of synthesis and release of glucose by the liver and kidney cortex during diabetes mellitus is a major factor in the pathogenesis of this disease.

Overview of Gluconeogenesis

Gluconeogenesis is defined as the *de novo* synthesis of glucose from nonhexose precursors. Gluconeogenesis does not include the conversion of fructose or galactose into glucose in the liver or the generation of glucose from glycogen via glycogenolysis. The pathway of gluconeogenesis ([Figure 1](#)) occurs mainly in the liver and kidney cortex and to a lesser extent in the small intestine. The major substrates for gluconeogenesis include lactate, pyruvate, propionate, glycerol, and 18 of the 20 amino acids (the exceptions are leucine and lysine). Glucose cannot be synthesized from fatty acids, since they are converted by β -oxidation into acetyl coenzyme A (CoA), which subsequently enters the citric acid cycle and is oxidized to CO₂.

[†]Deceased.

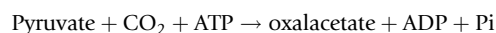
The three-carbon fatty acid, propionate, is an exception since it is carboxylated, converted into succinyl-CoA, and enters the citric acid cycle as a four-carbon intermediate, not as acetyl CoA; acetone, which can be converted into propanediol, is a very minor gluconeogenic precursor. In addition, the last three carbon atoms of the odd-chain fatty acids generate propionyl CoA during β -oxidation and are thus partly gluconeogenic. There are 14 enzymes involved in the conversion of lactate into glucose; three of these enzymes are classified as gluconeogenic (PEPCK, fructose-1,6-bisphosphatase (FBPase), and glucose-6-phosphatase (G6Pase)) and one is anaplerotic (PC), since it is important in both gluconeogenesis and lipogenesis. The remainder of the pathway is simply a reversal of the enzymes of glycolysis, which is responsible for the breakdown of glucose. Gluconeogenic enzymes are present in the cytosol, mitochondria, and endoplasmic reticulum (ER) of the tissues in which this pathway is present. Net gluconeogenesis occurs during starvation and after a meal high in fat and protein without carbohydrate.

The Enzymes of Gluconeogenesis

The unique properties of gluconeogenesis are attributable to the one anaplerotic and the three gluconeogenic enzymes in the pathway. The following discussion describes each of these enzymes and reviews their regulatory properties.

Pyruvate Carboxylase

PC is found in the mitochondria and catalyzes the following reaction:



PC is a multisubunit enzyme that has acetyl CoA as a positive allosteric regulator. The enzyme is critical for both gluconeogenesis and fatty acid synthesis, since it provides oxalacetate which is a precursor of malate and citrate, two citric acid cycle

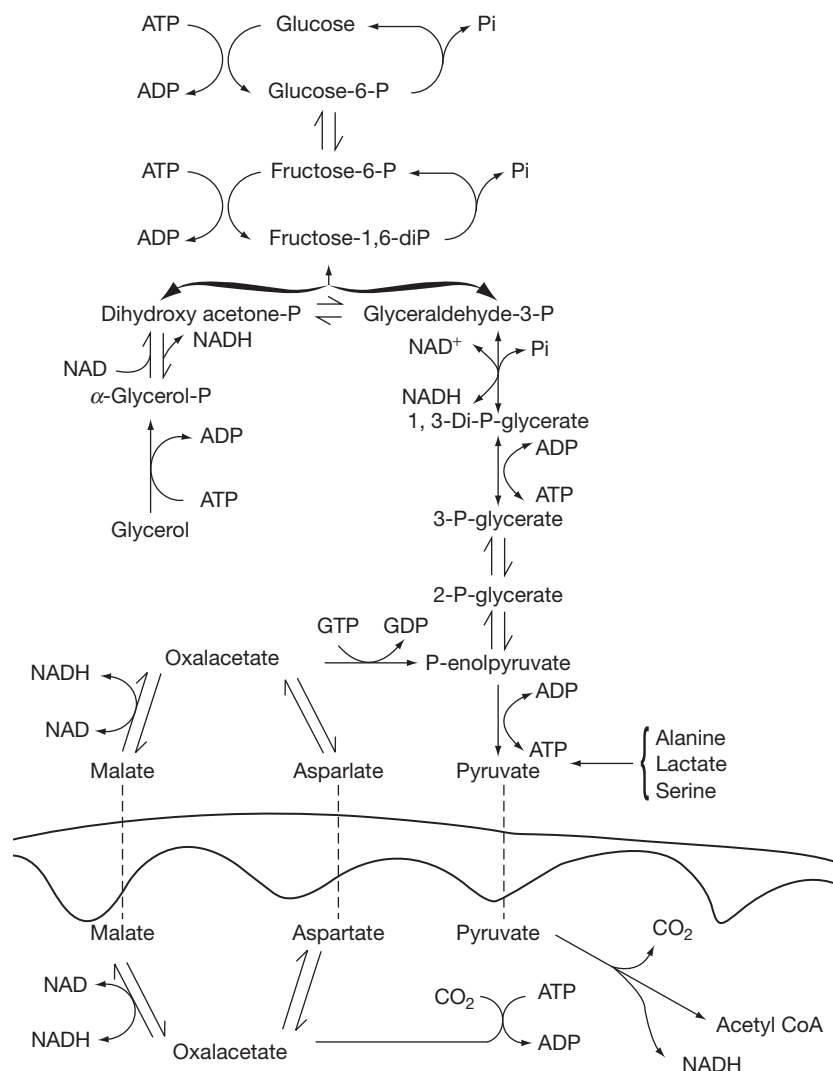
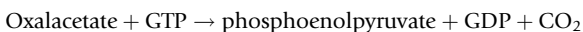


Figure 1 The pathway of hepatic gluconeogenesis. The figure demonstrates the reactions involved in gluconeogenesis in the liver starting with alanine, lactate, or serine as precursors. The movement of aspartate and malate from the mitochondria is also indicated to demonstrate the redox state balance that occurs between the mitochondria and the cytosol during gluconeogenesis.

intermediates that leave the mitochondria as part of biosynthetic processes. Since it is essentially irreversible under physiological conditions, PC does not generate pyruvate from oxalacetate. PC is generally considered an anaplerotic (Greek: to fill up) enzyme since it functions to replace the oxalacetate that is reduced to malate in the citric acid cycle and used in the synthesis of glucose. Since citric acid cycle intermediates are important in both gluconeogenesis and lipogenesis, PC plays a role in both the pathways.

Phosphoenolpyruvate Carboxykinase

PEPCK catalyzes the following reaction:



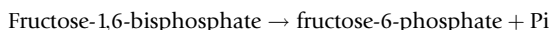
It is found in both the mitochondria (PEPCK-M) and the cytosol (PEPCK-C). Two different nuclear genes encode these two isoforms of PEPCK. Expression of the gene PEPCK-C is inducible by diet and hormones, while PEPCK-M is largely constitutive.

Species vary in the amount of the two isoforms expressed in gluconeogenic tissues. Rodent species such as the rat and mouse have 90% PEPCK-C in their liver and kidney cortex, while adult birds have 100% PEPCK-M. Humans and most other mammalian species studied to date have 50% of both isoforms. The exact metabolic significance of having two isoforms of PEPCK is not entirely clear but it has been proposed that PEPCK-M allows the direct synthesis of PEP in the mitochondria, thus bypassing the need to transport reducing equivalents out of the mitochondria. This point is discussed in more detail later. It is important to note that while PEPCK is a gluconeogenic enzyme, it is also involved in the synthesis of 3-glycerol phosphate from citric acid cycle intermediates for the synthesis of triglyceride in a process termed 'glyceroneogenesis'. In this regard, PEPCK is a cataplerotic enzyme, since it removes citric acid cycle anions that could accumulate when the carbon skeletons of amino acids enter the mitochondria for further metabolism. Thus, anaplerosis is balanced by cataplerosis to insure the appropriate level of citric acid cycle anions.

Despite the fact that PEPCK is a key enzyme in gluconeogenesis, it has no known allosteric regulators. However, recent studies have shown that PEPCK-C is acetylated at several lysines, one at a residue near the active site, which increases the activity of the enzyme; the importance of this covalent modification in the enzyme in its overall regulation of gluconeogenesis remains to be established. PEPCK-C has a relatively short half-life of 6 h and alterations in the rate of gene transcription are rapid. For example, the rate of PEPCK-C gene transcription in the liver is increased 10-fold in 30 min after the administration of cyclic adenosine monophosphate (cAMP) to animals. Insulin administration causes an equally rapid decrease in the transcription of the gene. It has thus been suggested that the level of enzyme protein in the liver is a critical determinant in the overall activity of the enzyme. Transcription of the gene for PEPCK-C has been studied in considerable detail, and many of the transcription factors that regulate the tissue-specific expression of the gene have been identified. A detailed discussion of this area is beyond the scope of this article but the reader is directed to several reviews on the subject listed in the section Further Reading.

Fructose-1,6-Bisphosphatase

FBPase is found in the cytosol. In mammalian liver, the enzyme is a homotetramer with a subunit molecular weight of 38–41 kDa, depending on the species. FBPase catalyzes the following reaction:



The enzyme is regulated allosterically by a number of small molecules including AMP and fructose-2,6-bisphosphate, which are negative regulators, and adenosine triphosphate (ATP) that is a positive regulator. The enzyme is also covalently modified by phosphorylation by cAMP-dependent protein kinase (PKA), which decreases its activity. The concentration of fructose-2,6-phosphate is determined by a unique bifunctional enzyme, 6-phosphofructo-2-kinase-2,6-bisphosphatase, which, when phosphorylated by protein kinase A, functions as a phosphatase and converts fructose-2,6-phosphate into fructose-6-phosphate, thereby lowering the concentration of this key allosteric regulator. This has the effect of decreasing flux through the glycolytic enzyme, phosphofructokinase, while increasing the activity of FBPase, since fructose-2,6-phosphate is a negative regulator of that enzyme. This type of reciprocal regulation of enzymes that share common intermediates is critical in controlling the rate of futile cycling in the pathways of glycolysis/gluconeogenesis, since it insures carbon flow toward the synthesis of glucose during fasting and toward pyruvate (glycolysis) when dietary carbohydrate is available.

Glucose-6-Phosphatase

The role of this enzyme in the synthesis of glucose was first suggested by the Coris in 1938 and finally demonstrated by Marjorie Swanson in the United States and Christian deDuve and colleagues in Belgium in the early 1950s.

G6Pase is a multisubunit enzyme that is present in the ER of the liver, kidney cortex, and small intestine, and comprises

several integral membrane proteins: the catalytic subunit (G6PaseC) and transporters for glucose-6-phosphate (for transport into the ER) and inorganic phosphate and glucose (GLUT2) for transport out. Pancreatic islets contain an isoform of G6PaseC, termed G6PaseC2, which is 50% identical with G6PaseC; while its function is not definitively established (the pancreas does not make and release glucose), it is suggested to be involved in glucose sensing by the β -cells of the pancreas. A third isoform, called G6PaseC3, which is 30% identical with G6PaseC, was recently identified and shown to be present in a number of mammalian tissues. Its metabolic role is not clear at present.

G6Pase is not solely a gluconeogenic enzyme, since it also functions as the last step in glycogenolysis, releasing glucose from the liver into the blood. Thus, mutations in any of the enzymes in the multienzyme alter not only gluconeogenesis but also glycogenolysis. Glycogen storage disease 1a (commonly called von Gerke's disease) is a mutation in the gene for G6PaseC, while glycogen-storage disease 1b involves mutations in the gene for the glucose-6-phosphate transporter. Mutations in the genes for this multisubunit complex result in profound hypoglycemia during the postabsorptive state and lactic acidemia, among other severe metabolic defects in hepatic and renal function.

Factors that regulate the expression of the gene for G6PaseC have been studied in detail. The activity of the enzyme can be induced by starvation, cAMP, fatty acids, and glucocorticoids. Somewhat paradoxically, glucose at high concentrations also stimulates transcription of the gene for this enzyme. It has been suggested that the induction of G6Pase gene transcription by glucose may prevent excessive hepatic glucose storage as glycogen.

Physiological Control of Gluconeogenesis

Dietary Status Determines the Extent of Gluconeogenesis

There are two normal metabolic situations during which gluconeogenesis occurs in mammals. In addition, gluconeogenesis increases during diabetes, resulting in an increased rate of glucose output by the liver. The first physiological situation involves the gluconeogenesis, characteristic of starvation, and, the second, the gluconeogenesis, that occurs after a meal that contains minimal carbohydrate. During starvation, the source of carbon for gluconeogenesis is body protein, largely from the muscle and lactate and pyruvate from glycolysis and glycerol from lipolysis. The muscle undergoes proteolysis, largely due to the fall in insulin. Proteolysis in muscle results in the generation of 20 amino acids. However, reactions in the muscle convert these amino acids mainly into alanine and glutamine, which are leased for subsequent metabolism. Both alanine and glutamine can be converted into glucose in the liver but in starvation it is the kidney cortex that utilizes glutamine. The kidney synthesizes glucose from the carbon skeleton of glutamine and uses the ammonia derived from this process to maintain the acid–base balance of the tubular urine. Later in this article, the metabolic role of renal gluconeogenesis is discussed in detail. The gluconeogenesis that occurs after a meal containing protein and fat, without carbohydrate, involves the 18 gluconeogenic amino acids found in dietary protein

(degradation of the amino acids leucine and lysine contributes no carbon for glucose synthesis). The carbon skeletons of most of the amino acids enter the citric acid cycle, proceed forward to malate, and are subsequently converted into glucose via the gluconeogenesis as described in detail below.

The Role of Fatty Acid Oxidation in the Control of Gluconeogenesis

Gluconeogenesis occurs during fasting but the amount of glucose supplied by this process is far less than can be generated by glycogenolysis during the immediate postabsorptive period. Blood glucose homeostasis is possible only because other substrates (fatty acids and ketone bodies) are used as fuels by peripheral tissues, thus sparing the use of glucose. In general, the rate of gluconeogenesis is increased in the liver during starvation due to a fall in the concentration of insulin and a rise in glucagon; glucocorticoids play a permissive role in stimulating this process. This change in hormone status results in profound alterations in the metabolism of both skeletal muscle and adipose tissue. There is an increased rate of lipolysis in the adipose tissue and a net protein breakdown (proteolysis) in muscle. This results in an elevated rate of delivery of fatty acids to the liver from adipose tissue and primarily alanine from muscle. Fatty acid oxidation by hepatic mitochondria raises the concentration of acetyl CoA and shifts the redox state of that organelle toward reduction; there is a concomitant increase in ATP synthesis. In the rat, for example, the hepatic nicotinamide adenine dinucleotide (NAD)/NADH ratio shifts from 700 to 550 after 48 h of starvation. The net effect of this increase in fatty acid oxidation is a stimulation of gluconeogenesis and an inhibition of glycolysis. This stimulatory effect is exerted at several levels. First, the rise in NADH and ATP inhibits multiple steps in the citric acid cycle, including citrate synthase, isocitrate dehydrogenase, and α -ketoglutarate dehydrogenase complex. Most importantly, the increased concentrations of ATP, NADH, and acetyl CoA that occur in the liver during fasting inhibit flux through the pyruvate dehydrogenase complex. The importance of the latter step is emphasized upon subsequently.

The Pathway of Gluconeogenesis

During starvation, alanine is transported from the muscle to the liver, where it is converted into pyruvate. As mentioned above, during fasting the pyruvate dehydrogenase complex is inhibited by the products of fatty acid oxidation. At the same time, acetyl CoA activates PC, thus converting pyruvate into oxalacetate, which is then reduced to malate by the mitochondrial isoform of NAD malate dehydrogenase. Malate leaves the mitochondria and is oxidized to oxalacetate via the cytosolic isoform of NAD malate dehydrogenase; the oxalacetate is decarboxylated to PEP by PEPCK-C, and the PEP is further metabolized via a reversal of glycolysis to fructose-1,6-bisphosphate. This intermediate is then converted into fructose-6-phosphate by the gluconeogenic enzyme, FBPase. Finally, G6Pase, which is present in the ER, generates free glucose from glucose-6-phosphate; GLUT2 transports the glucose from the liver to the blood.

Redox Balance for Gluconeogenesis

It is important to note that gluconeogenesis requires two molecules of NADH for every molecule of glucose that is synthesized. The NADH is required to reverse the glyceraldehyde-3-phosphate dehydrogenase reaction. The source of this NADH varies according to the substrate used for gluconeogenesis. When oxidized compounds such as pyruvate and alanine are used for glucose synthesis, the pathway involves malate transport from the mitochondria, essentially generating reducing equivalents in the cytosol for gluconeogenesis, as described above. There is a slight variation in this pathway of glucose synthesis when the starting substrate is lactate. In this case, the conversion of lactate into pyruvate in the cytosol generates NADH, which obviates the necessity of transporting malate from the cytosol to produce NADH. There are two possible routes for the movement of intermediates from the mitochondria that is not linked to the transport of reducing equivalents. The first involves the transamination of mitochondrial oxalacetate to aspartate via aspartate aminotransferase. The aspartate is transported to the cytosol and converted into oxalacetate by the cytosolic isoform of aspartate aminotransferase (Figure 1). The second variation in metabolite transfer involves the generation of oxalacetate in the mitochondria via PC, followed by the synthesis of PEP directly in the mitochondria via PEPCK-M. PEP is readily transported from the mitochondria on the tricarboxylate carrier in the inner mitochondria membrane. This role of PEPCK-M in avian metabolism supports the physiological role of the two isoforms of PEPCK. In birds, the liver uses only lactate for gluconeogenesis, and the avian liver contains only PEPCK-M; gluconeogenesis from amino acids occurs in the kidney.

Regulation of Gluconeogenesis

Gluconeogenesis is regulated at a few fundamental levels, which are discussed in the following.

Substrate Delivery to the Liver and Kidney

The rate of delivery of substrates such as lactate, which is derived from glycolysis in skeletal muscle (Cori cycle) or glycolysis in erythrocytes, glycerol from the breakdown of triglycerides in adipose tissue, and amino acids from proteolysis in skeletal muscle and other protein-rich tissues, greatly influences the level of hepatic gluconeogenesis. Thus, the control of the release and delivery of precursors from peripheral tissues is an important controlling factor in the rate of gluconeogenesis in the liver and kidney cortex. For example, insulin controls the rate of gluconeogenesis, in part, by regulating the delivery of substrates from muscle and adipose tissue to the liver.

Mass Action Regulation of Enzymes

As an example of this type of regulation, the activity of glyceraldehyde-3-P dehydrogenase (in the direction of gluconeogenesis) is stimulated by a rise in NADH in cytosol (the NAD/NADH ratio). The activity of G6Pase is increased by a rise in the concentration of G6Pase, and PEPCK activity is stimulated by a rise in the concentration of oxalacetate.

Allosteric Activation of Key Enzymes

As an example of this type of regulation of gluconeogenesis, acetyl CoA allosterically activates PC, while FBPase is activated by ATP and inhibited by AMP and fructose-2,6-bisphosphate. In addition, during gluconeogenesis, pyruvate kinase must be inhibited in order to prevent the conversion of PEP back to pyruvate (futile cycling) by liver-type pyruvate kinase (PK-L). A rise in the concentration of alanine or ATP and a decrease in the level of fructose-1,6-bisphosphate that occurs in the liver during fasting result in the allosteric inhibition of the activity of this enzyme (Figure 2).

Covalent Modification

Phosphorylation of gluconeogenic enzymes by PKA regulates the rate of key steps in gluconeogenesis. PK-L (not the muscle isozyme, PK-M) is inhibited by cAMP-induced phosphorylation via PKA. This, together with the allosteric regulation mentioned above, prevents the conversion of PEP into pyruvate in the liver and thus limits possible futile cycling of carbon and resulting energy loss. Despite this coordinated regulation, it has been estimated that about 25% of the carbon recycles at that step. The PKA regulation of hepatic gluconeogenesis is controlled in part by changes in the concentration of fructose-2,6-bisphosphate. As mentioned above, this compound allosterically regulates two critical enzymes, phosphofructokinase (the rate-controlling enzyme in glycolysis) and FBPase (the key gluconeogenic enzyme). Fructose-2,6-bisphosphate is both synthesized and degraded by a single, bifunctional enzyme 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase, which can act as a kinase (it synthesizes fructose-2,6-bisphosphate using ATP as a source of phosphate) or as a phosphatase (it converts fructose-2,6-bisphosphate into fructose-6-phosphate plus inorganic phosphate). Phosphorylation of this enzyme by PKA induces the phosphatase function of the enzyme. This results in a decrease in the concentration of fructose-2,6-bisphosphate, thereby inhibiting the activity of phosphofructokinase (the rate-limiting enzyme in glycolysis), which requires fructose-2,6-bisphosphate as an allosteric activator and stimulating the activity of FBPase, for which fructose-2,6-bisphosphate is an inhibitor (Figure 2). Thus, the covalent modification of the bifunctional enzyme has a reciprocal effect on the gluconeogenesis (stimulation) and glycolysis (inhibition).

Alteration in Gene Expression

The expression of several of the genes in the gluconeogenic pathway is modified by starvation, diabetes, or by diets rich in carbohydrate. Transcription of the genes coding for PEPCK-C, FBPase, and G6Pase is increased by fasting and during diabetes, while transcription of the genes for several of the glycolytic enzymes, such as PK-L, phosphofructokinase, and glucokinase, is decreased. Alterations in the rate of gene transcription change the levels of messenger RNA (mRNA) for a specific enzyme, resulting in a change in the rate of synthesis for the enzyme. An increase or decrease in enzyme content usually occurs over a period of several hours and is mediated by hormones such as glucagon (acting via cAMP), insulin, and glucocorticoids that stimulate the transcription of the genes for

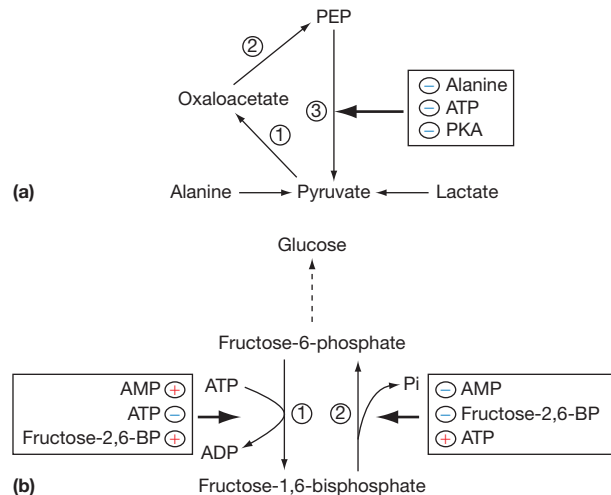


Figure 2 Regulation of potential futile cycles during gluconeogenesis. (a) The control of the oxaloacetate/PEP/pyruvate futile cycle. The enzymes shown in this panel include: (1) pyruvate carboxylase, (2) PEP carboxykinase, and (3) liver-type pyruvate kinase. This cycle is regulated by inhibition of liver-type pyruvate kinase by allosteric regulation by alanine and ATP and by phosphorylation of the enzyme by PKA. In addition, the decreased glycolytic flux that occurs during periods of gluconeogenesis (starvation) reduces the concentration of fructose-6-phosphate, a positive allosteric regulator of liver-type pyruvate kinase. (b) The control of the fructose-1,6-bisphosphate/fructose-6-phosphate futile cycle. The enzymes shown in this panel include: (1) phosphofructokinase and (2) fructose-1,6-bisphosphatase (FBPase). Phosphofructokinase is activated by AMP and fructose-2,6-bisphosphate and inhibited by ATP, while fructose-1,6-bisphosphatase is regulated in the opposite manner by the same intermediates. The result is a coordinated control of carbon flux via glycolysis and gluconeogenesis in the liver.

these enzymes. Metabolites such as glucose and fatty acids also exert control of the expression of specific genes involved in their metabolism. Several of the references listed in the section Further Reading review the details of the control of transcription of genes involved in gluconeogenesis.

Renal Gluconeogenesis

The synthesis of glucose in the kidney cortex is directly related to the loss of ketone bodies in the urine. During periods of fasting, the kidney excretes large amounts of ketone bodies (weak acids), but produces urine that is near neutrality. The relative acidity of the tubular urine is maintained at about pH 6.0 by the generation of ammonia from the metabolism of glutamine that has been mobilized from the muscle during starvation. Glutamine is converted into glutamate by glutaminase and glutamate to α -ketoglutarate by NAD glutamate dehydrogenase; this generates two molecules of ammonia that are released into the urine to maintain the neutrality of the urine. The α -ketoglutarate is then converted into glucose. Thus, ammoniogenesis in the kidney is linked to gluconeogenesis (Figure 3). Interestingly, acidosis induces transcription of the gene for PEPCK-C, increasing its concentration in the kidney cortex. This elevation in the activity of PEPCK-C is critical for the increased rates of gluconeogenesis noted during acidosis.

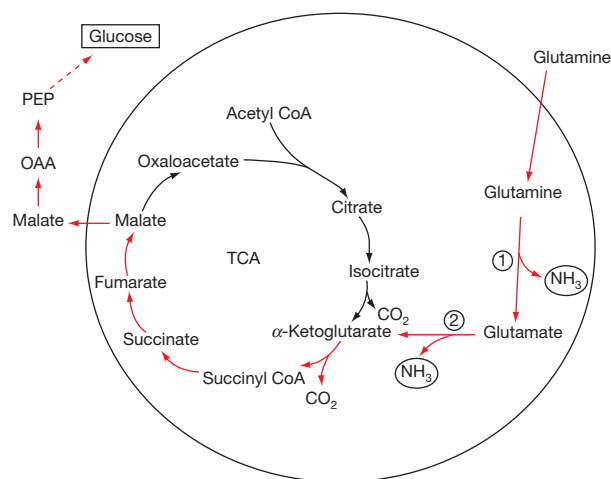


Figure 3 Gluconeogenesis from glutamine in the kidney. Glutamine from the muscle is metabolized by the kidney cortex to generate ammonia (circled) that is used to titrate the acidity of the tubular urine. This involves two enzymatic reactions, both of which are in the renal mitochondria, (1) glutaminase and (2) NAD glutamate dehydrogenase. The α -ketoglutarate, produced by the removal of the two amino groups of glutamine, enters the citric acid cycle, is oxidized to malate, and then proceeds to glucose via gluconeogenesis. The large circle in the figure represents the renal mitochondria.

Final Words

Glucose oxidation is essential for human metabolism and thus gluconeogenesis is a critical element in the maintenance of glucose homeostasis in all mammals. In humans, this process is continuous and augmented after the depletion of liver glycogen. Gluconeogenesis is the sole source of glucose during starvation. Since gluconeogenesis from amino acids results in a depletion of lean body mass, a number of metabolic accommodations occur to minimize the use of glucose by tissues such as the brain and skeletal muscle (fuel sparing). The details of these processes appear elsewhere in this encyclopedia. However, available data from human studies suggest that the rate of hepatic gluconeogenesis does not change appreciably in nondiabetic individuals; the mechanisms for this conclusion are presently not clear. A dysfunction in gluconeogenesis is a critical element in diabetes, a disease that is increasing at an alarming rate worldwide. Drugs aimed at inhibiting specific steps in gluconeogenic pathway hold promise in blocking the markedly elevated rate of hepatic glucose output that occurs in

this disease due in part to increased glucose synthesis by the process discussed in detail in this article. It is impossible to reference all of the research responsible for our understanding of the pathway of gluconeogenesis and its regulation. A select number of more general articles that reflect the field are included in the section Further Reading.

See also: [Metabolism Vitamins and Hormones: Diabetes](#); [Fatty Acid Oxidation](#); [Glucose/Sugar Transport in Bacteria](#); [Glucose/Sugar Transport in Mammals](#); [Pyruvate Kinase](#); [Starvation](#); [Protein/Enzyme Structure Function and Degradation](#); [Regulated Intramembrane Proteolysis](#); [Signaling: Diacylglycerol Kinases and Phosphatidic Acid Phosphatases](#).

Further Reading

- Granner DK and O'Brien RM (1992) Molecular physiology and genetics of NIDDM. *Diabetes Care* 15: 369.
- Hanson RW and Patel YM (1994) P-enolpyruvate carboxykinase: The gene and the enzyme. In: Meister A (ed.) *Advances in Enzymology*, vol. 69. New York, NY: Wiley.
- Hutton JC and O'Brien RM (2009) The glucose-6-phosphatase catalytic subunit gene family. *Journal of Biological Chemistry* 284: 29241–29245.
- Nordlie RC, Foster JD, and Lange AJ (1999) Regulation of glucose production by the liver. *Annual Review of Nutrition* 19: 379.
- Nuttall FQ, Ngo A, and Gannon MC (2008) Regulation of hepatic glucose production and the role of gluconeogenesis in humans: Is the rate of gluconeogenesis constant? *Diabetes/Metabolism Research and Reviews* 24: 438.
- Owen OE, Felig P, Morgan AP, Wahren J, and Cahill GF, Jr. (1969) Liver and kidney metabolism during prolonged starvation. *Journal of Clinical Investigation* 48: 584.
- Pilkis SJ and Claus TH (1991) Hepatic gluconeogenesis/glycolysis: Regulation and structure/function relationships of substrate cycle enzymes. *Annual Review of Nutrition* 11: 465.
- Pilkis SJ, Claus TH, Kurland IJ, and Lange AJ (1995) 6-Phosphofructo-2-kinase/fructose-2,6-bisphosphatase: A metabolic signaling enzyme. *Annual Review of Biochemistry* 64: 799.
- Rider MH, Bertrand L, Vertommen D, et al. (2004) 6-Phosphofructo-2-kinase/fructose-2,6-bisphosphatase: Head to head with the bifunctional enzyme that controls glycolysis. *Biochemical Journal* 381: 561.
- Solting H-D and Kleineke J (1976) Species dependent regulation of gluconeogenesis in higher animals. In: Hanson RW and Mehlman MA (eds.) *Gluconeogenesis, Its Regulation in Mammalian Species*, pp. 368–462. New York, NY: Wiley.
- Utter MF and Kolenbrander HM (1972) Formation of oxalacetate by CO_2 fixation on P-enolpyruvate. In: Boyer PD (ed.) *The Enzymes*, vol. VI, p. 117. New York, NY: Academic Press.
- Watford M, Hod Y, Chiao Y-B, Utter MF, and Hanson RW (1981) The unique role of the kidney in gluconeogenesis in the chicken: The significance of a cytosolic form of PEPCK. *Journal of Biological Chemistry* 256: 10023.
- Yang J, Kalhan SC, and Hanson RW (2009) What is the metabolic role of PEPCK? *Journal of Biological Chemistry* 284: 27025.

Glucose/Sugar Transport in Bacteria

L Guan, Texas Tech University Health Sciences Center, Lubbock, TX, USA

H R Kaback, University of California, Los Angeles, Los Angeles, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Active transport A process requiring input of energy where solute is generally accumulated against a concentration gradient.

Alternating access A mechanism in which both co-substrate-binding sites are exposed alternatively to the each side of the membrane by reciprocal opening and closing of outward- and inward-facing cavities.

Electrochemical ion gradient ($\Delta\tilde{\mu}_X^Z$; where X is an ion and Z is its valence) An electrochemical ion concentration gradient across a membrane is composed

of an electrical potential (ΔY) and a chemical concentration gradient (ΔY) (e.g., $\Delta\tilde{\mu}_{H^+}$ is composed of a ΔY and a pH).

Primary active transport Chemical energy or light is used directly to energize pumping of solute.

Secondary active transport Free energy stored in an electrochemical ion gradient is transduced into a solute concentration gradient.

Symport Two co-substrates are transported across the membrane in a same direction.

Symporter A transport protein that catalyzes symport.

Sugar transport in bacteria involves highly specific membrane proteins that catalyze translocation and accumulation of various sugars from the outside environment across the cytoplasmic membrane and into the cell where they are metabolized. As with other membrane proteins that catalyze transport of metabolites other than sugars, these proteins are called porters, permeases, or carriers. When energy is involved in the transport, the process is called active transport. In bacteria, there are three types of active transport systems involved in sugar transport, and each one utilizes different energy sources.

The Phosphoenolpyruvate: Sugar Phosphotransferase System

Classification

According to the Transport Commission (TC) classification, the phosphotransferase system (PTS) represents a distinctive family found only in certain bacteria. It is a bacterial sugar transport system, which also plays an essential role in the phenomena of catabolite repression and inducer exclusion, particularly in *Escherichia coli* (*E. coli*) and *Salmonella typhimurium*.

Functional Properties

The PTS is a multicomponent system that catalyzes vectorial phosphorylation of various sugars. In *E. coli*, glucose and other sugars are translocated across the membrane by PTS-catalyzed vectorial phosphorylation (Figure 1) (i.e., glucose from outside appears in the cytoplasm as glucose-6-phosphate). Therefore, translocation across the membrane and the first step in metabolism are one and the same. Many sugars are transported by PTS systems (glucose, mannitol, fructose, mannose, galactitol, sorbitol, xylitol, and *N*-acetylglucosamine), and specificity for each sugar is provided by the membrane-embedded component of the system. In *E. coli*, only monosaccharides and their derivatives are substrates for PTS systems, but in other bacteria, this type of system also utilizes disaccharides.

The PTS is also involved in regulating uptake and metabolism of many other sugars, as discussed here, as well

as bacterial chemotaxis toward these sugars, which is not covered here.

Assembly and Phosphoryl Transfer Reaction

The PTS includes three essential enzymes, termed Enzyme I (EI), Enzyme II (EII), and Histidine protein (HPr) (Figure 1). EI is a soluble protein kinase encoded by *ptsI* gene that catalyzes the phosphorylation of HPr from phosphoenolpyruvate (PEP). HPr, encoded by the *ptsH* gene, is a small monomeric protein that is phosphorylated on a histidine residue at position 15. Both proteins are involved only in phosphoryl transfer reactions and are not sugar specific. PEP autophosphorylates EI on His189, and the phosphoryl group is sequentially transferred to HPr. EII consists of three functional domains – IIA, IIB, and IIC. The organization of these domains depends on the specific sugar transported. In the case of glucose transport in *E. coli* (Figure 1), IIA^{Glc} , encoded by the *crr* gene, exists as a separate polypeptide, which enables it to perform multiple functions. IIA^{Glc} receives a phosphoryl group at His90 from HPr and transfers the phosphoryl group to Cys421 of IIB^{Glc} , which is a part of the membrane protein complex $IIBC^{Glc}$ (Figure 1). IIB^{Glc} directly donates the phosphoryl group to sugar to catalyze translocation across the membrane. IIC^{Glc} consists of multiple transmembrane helices and is likely responsible for allowing access of the sugar to IIB^{Glc} for phosphorylation and release on the inner surface of the membrane (i.e., vectorial phosphorylation). With some PTS systems in *E. coli*, the three domains are contained within a single polypeptide (e.g., mannitol) or domains IIA and IIB are in a single polypeptide, but domain C consists of two polypeptides (e.g., mannose).

Preferential Utilization of Carbohydrate

In *E. coli*, there is a preferential growth on glucose via the PTS in the presence of several sugars (diauxie). This phenomenon involves regulatory functions of the PTS with respect to transcription of a wide variety of inducible proteins that are

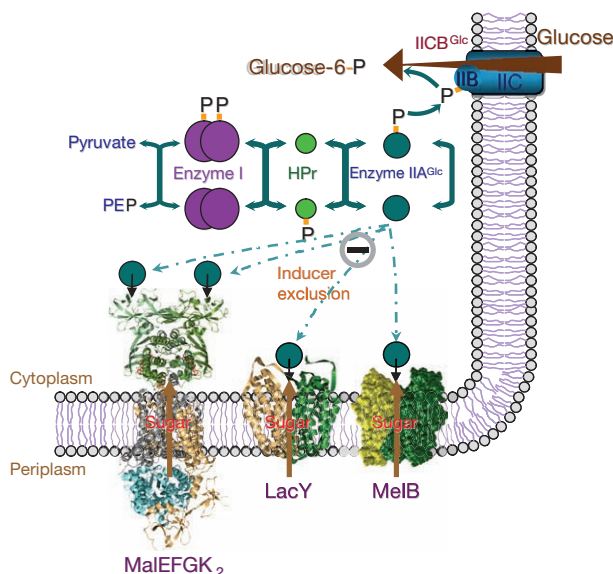


Figure 1 The glucose PTS system in *E. coli*. Shown is the sequence of phosphoryl transfer reactions through $\text{PEP} \rightarrow \text{EI} \rightarrow \text{HPr} \rightarrow \text{EIIA} \rightarrow \text{EIIB}$ domain $\rightarrow$ glucose during vectorial phosphorylation across the membrane through the EIIC domain. Glucose enters the cells as glucose-6-phosphate. Dephosphorylated IIA^{Glc} binds to non-PTS permeases – MalK in MalEFGK₂ (PDB ID:2r6g), LacY (2v8n), and MelB – and inhibits transport of non-PTS sugars.

dependent upon cyclic-AMP (cAMP; catabolite repression), as well as regulation of the activity of endogenous levels of inducible permeases (inducer exclusion). Both regulatory activities are properties of EIIA^{Glc} . In the presence of glucose, the steady-state level of phosphorylated enzyme IIA^{Glc} is low, which leads to decreased adenylate cyclase activity, low intracellular levels of cAMP, and low levels of the catabolite receptor protein–cAMP complex. Therefore, when glucose is present, transcription cannot be activated for most inducible PTS systems and non-PTS sugar permeases.

Another way to ensure preferential utilization of glucose is by inducer exclusion (Figure 1). In the presence of glucose, not only is the expression level of inducible transport systems dramatically reduced due to decreased transcription, but also the activity of the non-PTS sugar permeases expressed at basal levels is inhibited by binding of unphosphorylated IIA^{Glc} . These proteins include lactose permease (LacY), melibiose permease (MelB), and MalK (a component of the maltose transport complex). For example, with LacY, binding of IIA^{Glc} blocks entry of lactose and subsequent formation of allolactose, the true inducer of the lactose operon.

Electrochemical Cation-Gradient-Driven Symporters

Classification

Most cation gradient-driven symporters that catalyze sugar transport belong to the large, ancient, and diverse major facilitator superfamily (MFS), members of which are found ubiquitously in all living organisms. Within the MFS, the sugar porter family (e.g., galactose permease and arabinose permease), the oligosaccharide/proton symporter family (e.g., LacY and

sucrose permease), and the glycoside–pentoside–hexuronide: cation symporter family (e.g., MelB) are major subfamilies involved in bacterial sugar transport. In addition, the solute sodium symporter (SSS) family (e.g., sodium/galactose transport (SgIT)) also catalyzes monosaccharide accumulation.

Structural Features

Cation gradient-driven symporters involved in sugar transport are composed of a single polypeptide with 12–14 membrane-spanning helices, and for those studied intensively, a single polypeptide is able to catalyze translocation and accumulation. X-ray crystal structures of LacY from *E. coli* indicate that the protein is pseudo-symmetrically organized by the N- and C-terminal six-helix bundles (Figure 2(a)). A large central internal hydrophilic cavity in LacY is open to the cytoplasm, closed to the periplasm, and extends halfway into the membrane. Both sugar- and H^+ -binding sites lie at the apex of the cavity in the approximate middle of the protein. The inward-facing conformation represents the major population of LacY in the membrane, and the fold is postulated to represent most members of the MFS. The X-ray crystal structure of a SgIT from *Vibrio parahaemolyticus*, a member of the SSS family, reveals a different overall fold (Figure 2(b)), which contains an inverted repeat of five-transmembrane helices, a common structural core shared by many other genetically unrelated Na^+ symporters.

Functional Properties

The sugar permeases in the MFS or SSS families catalyze the transport of the mono- and disaccharides and their derivatives. They do not utilize PEP or adenosine triphosphate (ATP), but are driven by the free energy released from the energetically downhill movement of a cation (H^+ , Na^+ , or Li^+) in response to an electrochemical ion gradient to catalyze sugar translocation across the membrane. Unlike the PTS, with cation-gradient-driven transport, sugar enters the cell in an unchanged form and accumulates inside the cell against a concentration gradient and/or is metabolized. Mechanistically and thermodynamically, it is noteworthy that MFS or SSS sugar permeases catalyze sugar/ion symport in both directions across the membrane (influx and efflux), unlike the other transport systems described. An alternating access model describes the overall conformational change during transport: both co-substrate binding sites are exposed alternatively to the each side of the membrane by reciprocal opening and closing of outward-facing and inward-facing cavities. Turnover occurs only when co-substrate binding sites are concurrently occupied.

Representatives

LacY

LacY, encoded by the *lacY* gene of the *lac* operon, consists of 417 amino acid residues and is solely responsible for all the translocation reactions catalyzed by the β -galactoside transport system in *E. coli*. LacY is a particular well-studied representative of the oligosaccharide/ H^+ symport family of the MFS. Like many other MFS transporters, LacY utilizes free energy released from the downhill translocation of H^+ in response to an electrochemical H^+ gradient ($\Delta\mu_{\text{H}^+}$: interior negative and/or alkaline) to drive the stoichiometric translocation of a

galactopyranoside. In the absence of the electrochemical proton gradient, LacY catalyzes the converse reaction, utilizing free energy released from downhill translocation of sugar to drive uphill stoichiometric transport of H^+ . Intracellular β -galactosidase catalyzes the cleavage of lactose into galactose and glucose, which are metabolized via the glycolytic pathway. LacY is selective for disaccharides containing a D-galactopyranosyl ring, as well as D-galactose, but has no affinity for D-glucopyranosides or D-glucose. The K_m of transport is dramatically decreased in the presence of $\Delta\mu_{H^+}$, although the K_D remains unchanged.

By regulation of both transcription of *lacY*, as well as LacY activity by IIA^{Glc} of the PTS, *E. coli* grown in the presence of a mixture of glucose and lactose will utilize the lactose only when the glucose in the medium is completely exhausted, leading to diauxic or biphasic growth.

MelB

MelB of *E. coli*, encoded by the *melB* gene, consists of 473 amino acid residues and is a well-studied member of the glycoside–pentoside–hexuronide:cation symporter family. MelB, containing 12-transmembrane domains, is suggested to share a similar overall fold with LacY (Figure 1). MelB catalyzes the accumulation of galactopyranosides by utilizing the free energy from the energetically downhill inward movement of Na^+ , Li^+ , or H^+ to drive the stoichiometric accumulation of sugar. α -Galactopyranosides (melibiose, raffinose, and *p*-nitrophenyl- α -galactoside) are co-transported with Na^+ , H^+ , or Li^+ , while the β -galactopyranosides (lactose, methyl-1- β -D-galactopyranoside, or *p*-nitrophenyl- β -galactoside) are co-transported with Na^+ or Li^+ but not H^+ . Intracellular α -galactosidase catalyzes the cleavage of melibiose into galactose and glucose, which are metabolized via the glycolytic pathway. Expression and function of MelB are also regulated by the PTS.

SglT

SglT of *V. parahaemolyticus*, encoded by the *sglS* gene, consisting of 543 amino acid residues, has 14-transmembrane helices with extracellular N and C termini. The bacterial SglT catalyzes Na^+ -dependent sugar transport, with sugar preferences

(galactose > glucose > fucose). The structural core is formed from two five-helix inverted repeats, helices II–IV and VII–XI. Different from the MFS, the co-substrate binding sites in the SglT are located in two connected pockets in the middle of the molecule (Figure 2). Expression and function of SglT is not regulated by PTS. Human homologs of SglT are expressed primarily in the epithelia of the small intestine and the kidney where they play important roles in the physiology of these organs.

Binding Protein-Dependent ATP-Binding Cassette Transport System

Classification

Although ABC transporters are broadly distributed in archaea, prokaryotes, and eukaryotes, ABC transport systems that require binding proteins for function are found only in archaea and prokaryotes. All members share the common feature that turnover involves the utilization of energy released from the hydrolysis of ATP to translocate solute in one direction across the membrane against concentration gradient.

Assembly and Structural Features

The binding protein-dependent ABC transporters have a common global organization: two transmembrane domains or subunits, two cytoplasmic nucleotide-binding domains or subunits, also called ATP-binding cassettes, and one substrate-binding protein (Figure 3). Both the integral membrane protein and the cytoplasmic ATP-binding protein may be present as homodimers or heterodimers. Both ABCs are required for transport activity, and the sequences are highly conserved among different ABC transporters. In Gram-negative bacteria, each member of this family possesses a periplasmic binding protein, and in Gram-positive bacteria, the substrate-binding protein is a membrane-bound lipoprotein. Although these binding proteins have diverse specificities, the X-ray structures of the sugar-binding proteins exhibit a similar bilobate architecture composed of two symmetrical lobes surrounding substrate-

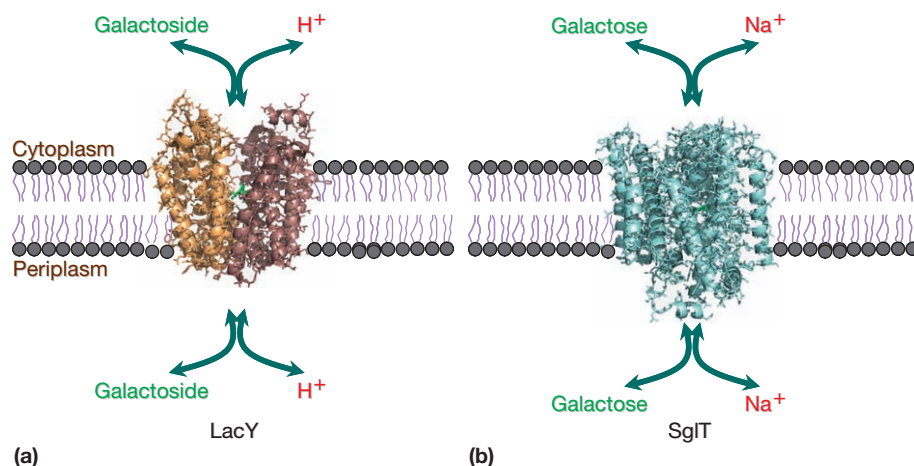


Figure 2 Crystal structures of LacY of *E. coli* and SglT of *V. parahaemolyticus*. (a) The X-ray crystal structure of the C154G LacY mutant (1pv7) viewed parallel to the membrane with a bound sugar (green sticks) at the bottom of the large cytoplasmic cavity formed by N- and C-terminal pseudo-symmetric domains. As shown the protein catalyzes the stoichiometric translocation of sugar and H^+ in either direction across the membrane. (b) The X-ray crystal structure of the SglT (3dh4) viewed parallel to the membrane with a bound galactose molecule (green sticks) in the middle of protein.

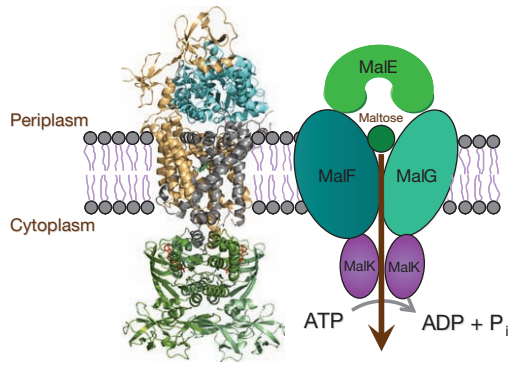


Figure 3 The X-ray crystal structure of MalEFGK₂ complex (2r6g) of *E. coli* with a bound maltose (green sticks) on MalF and two ATP (red sticks) on MalK. The periplasmic binding protein MalE tightly docks to the membrane proteins MalF and MalG, delivering maltose and stimulating ATPase activity of the cytoplasmic protein MalK. The energy released from the hydrolysis of ATP drives the translocation of maltose across the membrane. The sugar enters the cells in an unchanged form.

binding site. The X-ray crystal structure of the maltodextrin/maltose transport complex of *E. coli* (MalEFGK₂) represents a typical assembly and architecture for this class of transport systems (Figure 3). The two MalK subunits with the ATP-binding cassettes form a closed dimer at the cytoplasmic side, with two ATP molecules bound at the interface. The MalF and MalG subunits with the transmembrane domains consist of 8- and 6-transmembrane helices, respectively, and form a large water-filled central cavity open to the periplasmic side and closed on the cytoplasmic side. A maltose molecule is bound at the bottom of the cavity about half way into the cytoplasmic membrane. The periplasmic MalE subunit or maltose-binding protein (MBP) docks onto MalF and MalG in a sugar-free open conformation and serves as a cap to ensure unidirectional translocation of the sugar. These five subunits assemble into a tight complex through extensive interactions among each subunit.

Functional Properties

The binding protein selectively binds the substrate with high affinity and delivers it to the cavity formed by the cognate MalF and MalG subunits. Interaction of the sugar-bound periplasmic binding protein with MalF and MalG stimulates the ATPase activity of the cytoplasmic ATP-binding domains and initiates the transport process. The sugar enters the cell in an unchanged form. This system has the capacity to accumulate sugars to extremely high levels, and translocation is unidirectional.

Maltodextrin/Maltose Transport

Maltodextrin/maltose transport MalEFGK₂ (Figure 3), encoded by the maltose operon of *E. coli*, is one of the most extensively studied representatives of the binding protein-dependent sugar

transporters. MalE, which is the primary determinant of substrate specificity, recognizes and binds maltodextrin/maltose with high affinity. Sugar-bound MalE docks on the top of the large hydrophilic cavity formed by MalF and MalG complex, allowing direct delivery of the cargo sugar into the central cavity and triggers ATPase activity in MalK. There is a single maltose-binding site on MalF at the bottom of the cavity, which is closed at the cytoplasmic side by tight contact between MalF and MalG, and represents an outward-facing conformation. An alternating access model is also suggested for transport by the MalEFGK₂ complex. In addition, the C-terminal fragment of MalK forms a regulatory domain for EIIA^{Glc} regulation (Figure 1).

See also: **Bioenergetics: ABC Transporters; Lipids Carbohydrates Membranes and Membrane Proteins: Carbohydrate Chains: Enzymatic and Chemical Synthesis.**

Further Reading

- Abramson J, Smirnova I, Kasho V, et al. (2003) Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* 301(5633): 610–615.
- Cordat E, Mus-Veteau I, and Leblanc G (1998) Structural studies of the melibiose permease of *Escherichia coli* by fluorescence resonance energy transfer. II. Identification of the tryptophan residues acting as energy donors. *Journal of Biological Chemistry* 273(50): 33198–33202.
- Faham S, Watanabe A, Besserer GM, et al. (2008) The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na⁺/sugar symport. *Science* 321(5890): 810–814.
- Fetsch EE and Davidson AL (2003) Maltose transport through the inner membrane of *Escherichia coli*. *Frontiers in Bioscience* 8: d652–d660.
- Guan L and Kaback HR (2004) Binding affinity of lactose permease is not altered by the H⁺ electrochemical gradient. *Proceedings of the National Academy of Sciences of the United States of America* 101(33): 12148–12152.
- Guan L and Kaback HR (2006) Lessons from lactose permease. *Annual Review of Biophysics and Biomolecular Structure* 35: 67–91.
- Guan L, Mirza O, Verner G, Iwata S, and Kaback HR (2007) Structural determination of wild-type lactose permease. *Proceedings of the National Academy of Sciences of the United States of America* 104(39): 15294–15298.
- Kaback HR (1968) The role of the phosphoenolpyruvate–phosphotransferase system in the transport of sugars by isolated membrane preparations of *Escherichia coli*. *Journal of Biological Chemistry* 243(13): 3711–3724.
- Kaback HR, Dunten R, Frillingos S, et al. (2007) Site-directed alkylation and the alternating access model for LacY. *Proceedings of the National Academy of Sciences of the United States of America* 104(2): 491–494.
- Oldham ML, Khare D, Quirocho FA, Davidson AL, and Chen J (2007) Crystal structure of a catalytic intermediate of the maltose transporter. *Nature* 450(7169): 515–521.
- Postma PW, Lengeler JW, and Jacobson GR (eds.) (1996) *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, pp. 1149–1174. Washington, DC: American Society for Microbiology.
- Smirnova I, Kasho V, Choe JY, et al. (2007) Sugar binding induces an outward facing conformation of LacY. *Proceedings of the National Academy of Sciences of the United States of America* 104(42): 16504–16509.
- Yousef MS and Guan L (2009) A 3D structure model of the melibiose permease of *Escherichia coli* represents a distinctive fold for Na⁺ symporters. *Proceedings of the National Academy of Sciences of the United States of America* 106(36): 15291–15296.

Glucose/Sugar Transport in Mammals

S Mora, University of Liverpool, Liverpool, UK

J Pessin, Albert Einstein College of Medicine, Bronx, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Insulin resistance A state in which the same concentration of insulin elicits a decreased biological response compared to normal conditions.

Isoform A protein with similar amino acid sequence and same function.

Splice variant Different messenger RNA sequences that originate from the same gene.

Sodium-Dependent Glucose Transporters

The Na⁺/glucose cotransporters are responsible for the absorption of glucose against its concentration gradient by utilizing the energy generated by the electrochemical Na⁺ gradient (Figure 1(a)). These transporters are primarily located in the brush border (apical) membranes of intestinal and kidney epithelial cells and are responsible for the absorption of glucose from the lumen of the small intestine and proximal tubule of the kidney. In the intestines, this transporter functions in the absorption of dietary glucose from the lumen and is then transported across the basolateral membrane into the circulation by the facilitative glucose transporters. Similarly, in the kidneys glucose is filtered into the lumen of the proximal tubule and is reabsorbed by the Na⁺/glucose cotransporters such that under normal physiological conditions, urine is completely depleted of glucose. However, in states of insulin resistance and diabetes, the concentration of luminal glucose may reach such high levels that it exceeds the capacity of the Na⁺/glucose cotransporters to fully reabsorb the filtered glucose.

There are three different isoforms of sodium/glucose cotransporters, namely SGLT1, SGLT2, and SGLT3. Amino acid sequence alignments show that SGLT3 and SGLT2 have identities of 70% and 59% to SGLT1, respectively. The secondary structure model predicted for SGLT1 contains 14 transmembrane α -helices (TMHs) with both the hydrophobic NH₂ terminus and the COOH terminus of TMH 14 facing the extracellular space (Figure 2(a)). All SGLTs probably use the consensus N-linked glycosylation site (N248) between TMH 6 and 7. The predicted molecular weight of these transporters is 73 kDa with the proteins having 659–672 amino acids. The major differences in function between SGLT1 and SGLT2/SGLT3 resides in the selectivity for sugar transport. SGLT1 accepts the natural sugars – glucose and galactose – with comparable affinities and a variety of synthetic sugars such as 3-O-methyl-glucose. SGLT2 and SGLT3 are much more restrictive in that they prefer glucose to galactose or 3-O-methyl-glucose. The affinity of SGLT1 for glucose is 10-fold higher than that of SGLT2 or SGLT3.

Facilitative Glucose Transporters

The facilitative glucose transporters comprise a relatively large family of structurally related proteins that are encoded by distinct genes and are expressed in an overlapping yet tissue-specific manner (Table 1). All the facilitative glucose

transporters are predicted to have 12-membrane-spanning domains and allow for the stereospecific transport of glucose through the lipid bilayer. In contrast to the SGLT1 family, the GLUT4 family has both the NH₂ terminus and the COOH terminus facing the intracellular space. The transport of glucose is energy independent and is solely driven by the relative glucose concentration across the plasma membrane (Figure 1(b)). As under most conditions, the cytoplasmic glucose is rapidly phosphorylated by hexokinase or glucokinase, the relative levels of intracellular glucose are compared to the extracellular concentration resulting in net inward flux of glucose. However, in some instances, glucose can be secreted into the circulation under normal physiological conditions. For example, in hepatocytes facilitative glucose transporters participate in the uptake of glucose from the portal circulation following a meal and in the release of glucose generated by glycogenolysis or gluconeogenesis in the postabsorptive state and during fasting.

Currently, there are 14 identified facilitative hexose transporters in mammalian cells, which have been subdivided into three classes according to the similarity of their amino acid sequences: class I comprises GLUT1, GLUT2, GLUT3, GLUT4, and GLUT14; class II comprises GLUT5, GLUT7, GLUT9, and GLUT12; and class III comprises GLUT6, GLUT8, GLUT10, GLUT12, and the H⁺/myo-inositol transporter (previously termed GLUT13). Each of these transporters exhibit different transport properties and have distinct but overlapping tissue distributions, which underscores their specific physiologic function. For example, GLUT1 is generally expressed and is thought to be responsible for the basal uptake of glucose. GLUT2 is predominantly expressed in the liver and pancreatic β -cells where it functions as part of a sensor-mediating hepatic glucose output during states of fasting and insulin secretion in the postprandial absorption state. In contrast, neurons primarily express the relatively high-affinity GLUT3 isoform necessary to maintain high rates of glucose metabolism for energy production. GLUT8 appears to provide important function during early embryogenesis, whereas GLUT4 is exclusively expressed in insulin-responsive tissues, adipose, and striated muscle. These latter tissues provide the key functional elements responsible for the insulin stimulation of glucose uptake and are key targets for dysregulation in states of insulin resistance and diabetes. The latest GLUT member identified, GLUT14, is present in the testis and seems to have evolved from a gene duplication of GLUT3.

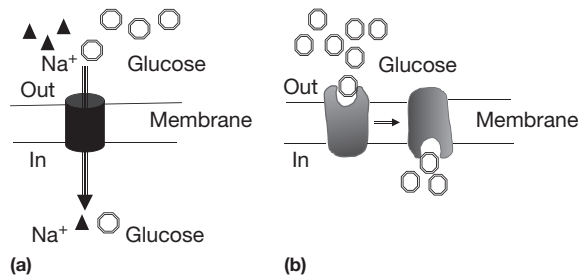


Figure 1 Schematic representation of the glucose transport process mediated by sodium-dependent glucose transporters (a) and facilitative glucose transporters (b).

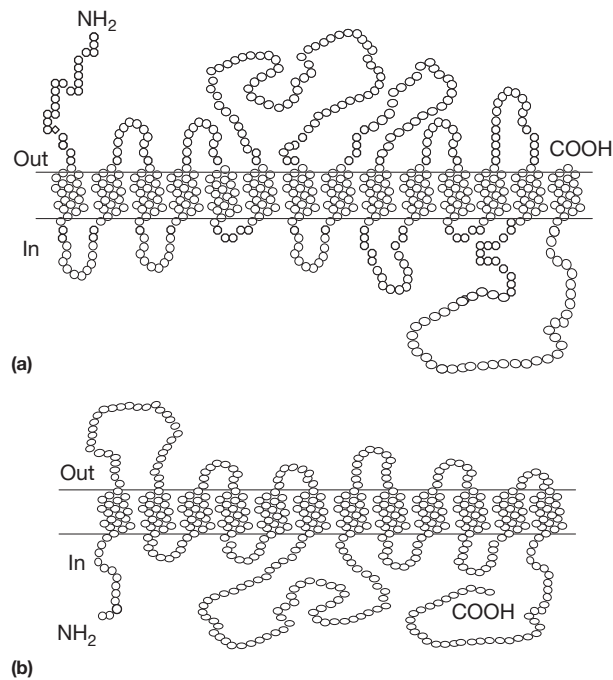


Figure 2 Schematic representation of the predicted structure of glucose transporters: (a) SGLT-1 and (b) GLUT-1.

GLUT1

GLUT1 was originally identified as the erythrocyte-type facilitative glucose transporter and was the first isoform to be purified and eventually cloned. The amino acid sequence of GLUT1 is highly conserved among different species (i.e., 98% identity between human and rat). Computer analysis of the primary sequence of GLUT1 has revealed that GLUT1 is a hydrophobic protein, with 50% located within the lipid bilayer. Several lines of evidence have supported a structural model for GLUT1 that consists of 12-membrane-spanning α -helical segments and an intracellular located NH_2 and COOH termini. According to this model, a large intracellular segment of 33 amino acids connects transmembrane domains M1 and M2, and an extremely hydrophilic intracellular segment of 65 amino acids joins M6 and M7. Short regions of 7–14 amino acids connect the remaining membrane-spanning regions. There are two potential N-linked glycosylation sites in GLUT1 at Asn-45 and Asn-411, but only the former site in the large extracellular loop has been shown to be glycosylated (**Figure 2(b)**). Although other models have been proposed, structural analysis of GLUT1 topography has confirmed the originally proposed 12-transmembrane-spanning model. The other members of the glucose transporter family are similar in size (between 492 and 524 amino acids) and all are predicted to maintain the 12-transmembrane-spanning structure, although the sites of glycosylation and length of the different intracellular and extracellular loops appear to vary, classes I and II are glycosylated in loop 1, whereas class III is glycosylated in loop 9. The N-glycosylation appears to be required to maintain high-affinity transport for glucose.

GLUT1 is highly expressed in placenta, brain, epithelial cells of the mammary gland, transformed cells, and fetal tissues. Because of its ubiquitous expression it has been proposed that GLUT1 might be responsible, at least in part, for basal glucose uptake. Kinetic studies in erythrocytes have demonstrated that GLUT1 has a K_m of 2 mM for α -glucose, whereas the K_m is greater than 3 M for β -glucose. In addition, it has been postulated that GLUT1 functions in either a dimeric or a tetrameric oligomeric state. This appears to account for the various kinetic properties of glucose transport and, in addition to expression levels and

Table 1 Mammalian facilitative sugar transporters

<i>Isoform</i>	<i>Tissue distribution</i>	<i>Function</i>
GLUT-1	Brain, endothelial cells, placenta, erythrocyte, mammalian gland	Basal transport, transport in fetal tissues
GLUT-2	Liver, small intestine, pancreatic cells, kidney	Basolateral transport in epithelial cells, glucose release from liver
GLUT-3	Neurons, sperm, white blood cells, placenta	Basal transport
GLUT-4	Muscle, heart, white and brown adipose tissue	Insulin-stimulated glucose transport
GLUT-5	Small intestine, kidney, testis, sperm	Fructose transporter
GLUT-6	Spleen, leukocytes, brain	Unknown
GLUT-7	Intestine, prostate	Unknown
GLUT-8	Testis, blastocyst, brain	Unknown
GLUT-9	GLUT9a: liver, kidney, placenta GLUT9b: kidney, placenta	Hexose and urate transporter
GLUT-10	Liver, pancreas	Unknown
GLUT-11	Heart, muscle	Unknown
GLUT-12	Heart, prostate	Unknown
GLUT-13	Brain	H^+ /myo-inositol transporter
GLUT-14	Testis	Unknown

localization, provides another level of control of transporter function. Homozygous deletion of the Glut1 in mice results in embryonic lethality.

GLUT2

The low-affinity GLUT2 facilitative glucose transporter functions in the β -cells of the pancreas to transport glucose in direct proportion to the circulating glucose concentration. This transporter functions in concert with glucokinase such that small increases in glucose influx can signal changes in the adenosine triphosphate/adenosine diphosphate (ATP/ADP) ratio, resulting in a highly sensitive mechanism regulating insulin secretion. The central importance of GLUT2 in the regulation of pancreatic insulin secretion was established with the identification of a patient expressing a defective GLUT2 mutation. This transporter also functions in the liver to allow for shuttling of increased glucose load into glycogen. GLUT2 also appears to play a significant role in the reabsorption of filtered glucose in the kidney and in the basolateral transport of glucose from the digestive tract epithelium into the circulation.

GLUT3

GLUT3 was originally isolated from a fetal skeletal muscle library, but is not significantly expressed in adult muscle. This isoform is primarily expressed in neurons, where it is located in cell processes (axons and dendrites), and it has also been detected in the placenta, endothelial cells, human skeletal muscle, and chondrocytes; and in murine sperm, where it is concentrated in the plasma membrane of the flagellum. GLUT3 has also been found in blastocysts, and is essential for embryonic development and survival. Deletion of GLUT3 in the mouse results in increased apoptosis of blastocysts and loss at embryonic day 8.5. GLUT3 has been detected in white blood cells, in human lymphocytes, monocytes/macrophages, neutrophils, and platelets. Interestingly, GLUT3 appears to have an intracellular localization in these cells, and it is recruited to the cell surface upon cell activation.

GLUT4

In 1989, five research groups independently cloned the insulin-responsive glucose transporter isoform GLUT4. This isoform is predominantly expressed in striated muscle (skeletal and cardiac) and adipose tissue (white and brown), the major tissues that display insulin-stimulated glucose uptake. Sequence analysis of GLUT4 suggests that its secondary structure and orientation in the plasma membrane is similar to GLUT1 with an overall 65% amino acid identity.

In contrast to the other glucose transporters under basal conditions, GLUT4 is predominantly intracellular localized. The primary mechanism responsible for the increase in glucose uptake in insulin-sensitive tissues upon insulin treatment is the recruitment or translocation of glucose transporters from these intracellular storage sites to the plasma membrane. Although the molecular mechanisms that regulate GLUT4 translocation are poorly understood, the trafficking of the GLUT4 protein has several similarities with the trafficking, docking, and fusion of synaptic vesicles in neurons responsible for neurotransmitter release.

The long-term expression of the GLUT4 gene is also controlled during development and during various hormonal, nutritional, and metabolic states. For example, fetal tissues contain very little GLUT4 and expression of this transporter increases after birth. In addition, during states of insulin deficiency such as fasting and following destruction of the insulin secreting β -cells by streptozotocin (STZ)-induced diabetes, GLUT4 expression levels are severely reduced in adipose, heart, and skeletal muscle. Similarly, obesity, high-fat feeding, and muscle denervation also result in a decrease in GLUT4 expression and subsequent insulin resistance. By contrast, GLUT4 is upregulated by exercise training, thyroid hormone treatment, and insulin therapy – treatments that increase insulin sensitivity.

GLUT5

The GLUT5 fructose transporter is highly conserved among species with a 69–88% amino acid identity between mice, human, rat, and rabbit. In mice, expression is highest in small intestine, kidney (proximal tubule), and testis. In humans, it is also present in skeletal muscle and in small amounts in adipocytes. Expression of GLUT5 in adipocytes increases under conditions of hypoxia and obesity. Studies in rats and rabbits have shown that the expression of this transporter is highest in the upper part of the small intestine (duodenum and proximal jejunum). Expression of GLUT5 in the intestine is highly regulated, showing circadian rhythm, dependence on dietary fat and carbohydrate content, and insulin levels. Interestingly, fructose does not regulate GLUT5 expression in adipose cells. GLUT5 has been found in the sperm of numerous species, although its role in sperm metabolism remains obscure.

GLUT6

The GLUT6 gene encodes a complementary DNA (cDNA), which exhibits significant sequence similarity with the other members of the glucose transporter family, particularly GLUT8. GLUT6 protein has been detected in spleen, peripheral leukocytes, and brain. When heterologously expressed in 3T3L1 adipocyte cells this transporter is intracellular localized, but unlike GLUT4 does not translocate to the plasma membrane in response to insulin.

GLUT7

The GLUT gene encodes for a protein of 512 amino acids, which has a sequence identity of 58% with GLUT5. When expressed in *Xenopus* oocytes, this protein transported both glucose and fructose at much lower concentrations than those found for GLUT5. The messenger (mRNA) for GLUT7 (but not yet the protein) has been detected mainly in the small intestine and colon, and to lesser amounts in prostate and testis. The role of GLUT7 in these tissues remains to be elucidated.

GLUT8

GLUT8, previously termed as GLUTX1, is a newly characterized glucose transporter isoform that is expressed at high levels in the testis and the brain, and at lower levels in several other

tissues including spleen, liver, skeletal muscle, heart, and adipose tissue. In the brain, GLUT8 is expressed in the neuronal cell bodies of excitatory and inhibitory neurons in the rat hippocampus. It was found in hippocampal and dentate gyrus neurons as well as in amygdala and primary olfactory cortex. In these neurons, its location was close to the plasma membrane of cell bodies and sometimes in proximal dendrites. High GLUT8 levels were detected in the hypothalamus, supraoptic nucleus, median eminence, and the posterior pituitary. In the testis, it is found in differentiating spermatocytes of type 1 stage and in the acrosomal region it is found in mature spermatozoa. When expressed in *Xenopus* oocytes it exhibits high-affinity transport activity for glucose, which is specifically inhibited by fructose and galactose, suggesting that it may be a multifunctional transporter.

GLUT8 contains a dileucine intracellular localization motif; however, its precise subcellular distribution has been controversial. In some cells, it has been found to be associated with the endoplasmic reticulum, whereas in others it has been detected in late endosomal fractions suggesting that GLUT8 may have a role in the transportation of sugars obtained from lysosomal degradation of glycoproteins. GLUT8-deficient mice display normal growth but with subtle modifications of behavioral traits, and reduced sperm mobility.

GLUT9

GLUT9 has a high-sequence similarity to the fructose transporter GLUT5 and is mainly expressed in the kidney (both proximal and convoluted tubule), liver, and placenta. It has two splice variants GLUT9a and GLUT9b, which differ in the NH₂ terminal part of the molecule and subcellular distribution. GLUT9a is present in the basolateral membrane of epithelial cells whereas GLUT9b is located in the apical membrane. Expression of GLUT9 is regulated by starvation and hyperglycemia. The substrate specificity of this transporter is unique in that it transports hexoses (both glucose and fructose) and uric acid. Since GLUT9 has been found to be present in articular cartilage, it is possible that this transporter may play a role in gout disease.

GLUT10–14

Other than sequence analysis and tissue distribution of mRNA expression, little information is available about the function and properties of GLUT10–14. GLUT10 is primarily expressed in human heart, lung, brain, liver, skeletal muscle, pancreas, placenta, and kidney. GLUT11 gene in humans has three splice variants: GLUT11a is most abundant in skeletal muscle and

heart, although it is also present in kidney; GLUT11b is present in placenta, adipose tissue, and kidney; and GLUT11c is present in adipose tissue, heart skeletal muscle, and pancreas. All GLUT11 isoforms transport glucose and fructose. Interestingly, the GLUT11 gene has not been found in the mouse or rat genome. GLUT12 is expressed in heart, prostate, and has also been detected in breast cancer cells and in epithelial cells of rat mammary gland during pregnancy and lactation. It has dileucine motifs in both NH₂- and COOH-terminal parts of the molecule. In humans, GLUT12 expression is restricted to insulin-sensitive tissues: skeletal muscle, heart, and adipose; and it has been postulated to be another insulin-responsive glucose transporter. GLUT13 is predominantly expressed in the brain and displays an H⁺/myo-inositol transporter activity. It does not transport glucose. GLUT14 is expressed in the testis and is 95% identical to GLUT3.

See also: **Bioenergetics:** Membrane Transport, General Concepts; **Lipids Carbohydrates Membranes and Membrane Proteins:** Sugar Nucleotide Transporters; **Metabolism Vitamins and Hormones:** Gluconeogenesis; Glycogen Metabolism; Insulin- and Glucagon-Secreting Cells of the Pancreas.

Further Reading

- Carruthers A, Dezutter J, Ganguly A, and Devaskar SU (2009) Will the original glucose transporter isoform please stand up! *American Journal of Physiology – Endocrinology and Metabolism* 297: E836–E848 (doi:10.1152/ajpendo.00496.2009).
- Cheeseman C (2009) Solute carrier family 2, member 9 and uric acid homeostasis. *Current Opinion in Nephrology and Hypertension* 18(5): 428–432 (review).
- Karnieli E and Armori M (2008) Transcriptional regulation of the insulin-responsive glucose transporter GLUT4 gene: From physiology to pathology. *American Journal of Physiology – Endocrinology and Metabolism* 295: E38–E45.
- Klip A (2009) The many ways to regulate glucose transporter 4. *Applied Physiology, Nutrition, and Metabolism* 34(3): 481–487 (review).
- Koistinen H and Zierath J (2003) Regulation of glucose transport in human skeletal muscle. *Annals of Medicine* 34: 410–418.
- Marsenic O (2009) Glucose control by the kidney: An emerging target in diabetes. *American Journal of Kidney Diseases* 53(5): 875–883 (review).
- Purcell SH and Moley KH (2009) Glucose transporters in gametes and preimplantation embryos. *Trends in Endocrinology and Metabolism* 20(10): 483–489.
- Schmidt S, Joost HG, and Schürmann A (2009) GLUT8, the enigmatic intracellular hexose transporter. *American Journal of Physiology – Endocrinology and Metabolism* 296(4): E614–E618 (review).
- Uldry M and Thorens B (2004) The SLC2 family of facilitated hexose and polyol transporters. *Pflügers Archiv* 447(5): 480–489. (Erratum (2004) *Pflügers Archiv* 448(2): 259–260.)
- Zhao F-Q and Keating AF (2007) Functional properties and genomics of glucose transporters. *Current Genomics* 8: 113–128.

Glutamate Receptors, Metabotropic

K J Gregory and P J Conn, Vanderbilt University Medical Center, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by P. Jeffrey Conn, volume 2, pp. 220–223, © 2004, Elsevier Inc.

Glossary

Allosteric site Small-molecule-binding site that is topographically distinct from the endogenous ligand (orthosteric) binding site.

Gamma-aminobutyric acid (GABA) An amino acid that serves as the primary inhibitory neurotransmitter in the central nervous system.

Glutamate An amino acid that serves metabolic functions and as a component of proteins and polypeptides in all cells. In addition, glutamate serves as a major excitatory neurotransmitter in the central nervous system.

G protein Membrane-associated protein that interacts with a particular class of cell-surface receptors and participates in

receptor-induced activation of intracellular enzymes, ion channels, and other signaling molecules. It is also known as guanosine triphosphate (GTP)-binding protein.

Negative allosteric modulator Compound that inhibits the binding and/or activity of an orthosteric ligand via interaction with an allosteric site.

Orthosteric site The binding pocket on a protein to which the endogenous neurotransmitter/hormone binds to activate the receptor.

Positive allosteric modulator Compound that enhances the binding and/or activity of an orthosteric ligand by binding to an allosteric site.

Fast Synaptic Transmission and Neuromodulation in the Central Nervous System

The majority of connections that make up a neural network in the mammalian central nervous system are synaptic connections in which a neurotransmitter released from the presynaptic neuron activates postsynaptic receptors that are neurotransmitter-gated ion channels. Opening of these ion channels results in either excitation or inhibition of a receptive neuron; these processes are, respectively, known as fast excitatory and inhibitory synaptic transmission. Normal function of the central nervous system (CNS) requires a delicate balance between excitatory and inhibitory transmission. In mammalian brain, two major neurotransmitters are responsible for most fast inhibitory and excitatory synaptic responses. Gamma-aminobutyric acid (GABA) is the principal inhibitory neurotransmitter, whereas the amino acid glutamate is the primary excitatory neurotransmitter. Transmission through a neuronal circuit can be fine-tuned or regulated by activation of another class of receptor that couples to activation of second-messenger systems and various biochemical processes through guanosine triphosphate (GTP)-binding proteins (G proteins). This form of synaptic transmission is often referred to as slow synaptic transmission or neuromodulation to differentiate it from fast excitatory and inhibitory transmission.

Fast Excitatory and Neuromodulatory Actions of Glutamate

As the primary fast excitatory neurotransmitter, glutamate plays an important role in virtually all major functions of the CNS. Traditionally, the actions of glutamate were thought to be mediated exclusively by activation of glutamate-gated cation channels termed ionotropic-glutamate (iGlu) receptors.

Fine-tuning or modulation of transmission through glutamatergic circuits was thought to require neuromodulators from extrinsic afferents such as dopamine, acetylcholine, serotonin, and norepinephrine that activate G protein-coupled receptors (GPCRs). However, in the mid-1980s, evidence for the existence of glutamate receptors directly coupled to second-messenger systems through G proteins began to appear with the discovery of glutamate receptors coupled to activation of phosphoinositide hydrolysis. A key turning point being the discovery of *trans*-1-amino-1,3-cyclopentanedicarboxylate that activated glutamate receptors coupled to phosphoinositide hydrolysis while having no effect on iGlu receptors. Since that time, it has become clear that glutamate activates a large family of receptors, termed metabotropic glutamate (mGlu) receptors, which signal by coupling to G proteins.

The discovery of mGlu receptors dramatically altered the traditional view of glutamatergic neurotransmission because activation of mGlu receptors can modulate activity in glutamatergic circuits in a manner previously associated only with neuromodulators from non-glutamatergic afferents. However, unlike receptors for monoamines and other neuromodulators, the mGlu receptors provide a mechanism by which glutamate can modulate or fine-tune activity at the same synapses at which it elicits fast synaptic responses. Because of the ubiquitous distribution of glutamatergic synapses, mGlu receptors participate in a wide variety of functions of the CNS.

Eight mGlu Receptor Subtypes

A major breakthrough in the understanding of mGlu receptors occurred in 1991 when groups led by Shigetada Nakanishi at Kyoto University in Japan and by Eileen Mulvihill and F.S. Hagen at ZymoGenetics, Inc., in Seattle, Washington, independently cloned the first gene encoding an mGlu receptor.

A search for related complementary DNA (cDNA) resulted in the isolation of seven other genes encoding different mGlu receptor subtypes, named mGlu1 through mGlu8. Also, with the exception of mGlu2, several alternatively spliced forms or structural variants have been isolated for the mGlu receptor subtypes. These include splice variants for mGlu1 (a–d), mGlu3 (GRM3Δ2, GRM3Δ4, GRM3Δ2Δ3, and GRM3Δ4) mGlu4 (a and b), mGlu5 (a and b), mGlu6 (a–c), mGlu7 (a–e), and mGlu8 (a–c). Thus, it is now clear that there is tremendous diversity within the mGlu receptor family.

The mGlu receptors belong to family C GPCRs, which also include GABA_B receptors, calcium-sensing receptors, pheromone receptors, and taste receptors. Although the predicted topology of these receptors is similar to rhodopsin receptors and other GPCRs in that they contain seven predicted transmembrane-spanning domains, the primary structure of this family shares very little sequence homology with other GPCRs. With the exception of orphan receptors, all family C GPCRs are characterized by a large extracellular N-terminus, referred to as the Venus flytrap (VFT) domain, which contains the endogenous ligand-binding site. Numerous structure–function studies suggest that the VFT domain of mGlu receptors is made up of two globular domains with a hinge region. This forms a clam-shell-shaped structure, with the glutamate-binding site residing between the two globular domains (see Figure 1). Evidence suggests that when glutamate binds, the globular domains close into a stable conformation with glutamate inside. The conformation changes induced by glutamate binding are transmitted via the cysteine-rich domain (CRD) to the transmembrane-spanning α -helices, subsequently promoting coupling to intracellular G proteins and activation of second-messenger pathways. The CRD contains nine cysteine residues and X-ray crystallographic studies have shown that eight of these form disulfide bonds within this region, while

the ninth forms a disulfide bond with a cysteine in the VFT. The transmembrane-spanning α -helical domains contain small-molecule-binding sites (allosteric sites) that depending upon the ligand can either inhibit (negative allosteric modulators) or enhance (positive allosteric modulators) the function of the receptor, while the intracellular loops and large C terminal domain of the mGlu receptors are important determinants for G protein-coupling specificity as well as regions subject to alternate splicing and regulation by phosphorylation and protein–protein interactions.

The cloned mGlu receptors have been classified into three major groups based on sequence homology, pharmacological properties, and coupling to different second-messenger pathways. Group I includes mGlu1 and mGlu5, group II, mGlu2 and mGlu3, and group III, all the others. mGlu receptors of the same group show ~70% sequence identity, whereas between groups this percentage falls to ~45% (Figure 2). Group I mGlu receptors preferentially couple to activation of the G_{q/11} family of G proteins activating phosphoinositide hydrolysis as the major signaling mechanism. By contrast, group II and group III mGlu receptors preferentially couple to G_{i/o} and inhibition of adenylyl cyclases. Members of each group have a unique pharmacological profile and can be selectively activated by specific agonists that have no effects on members of the other groups of mGlu receptors. Also, in recent years, progress has been made in developing ligands, predominantly allosteric modulators, which differentiate between members of a single group. These ligands are having a major impact on the understanding of the physiological roles and pharmacological properties of the different mGlu receptor subtypes.

mGlu Receptors Play Diverse Roles in Regulating Neuronal Excitability and Synaptic Transmission

With the exception of mGlu6, which is localized to the retina, mGlu receptors are ubiquitously expressed throughout the

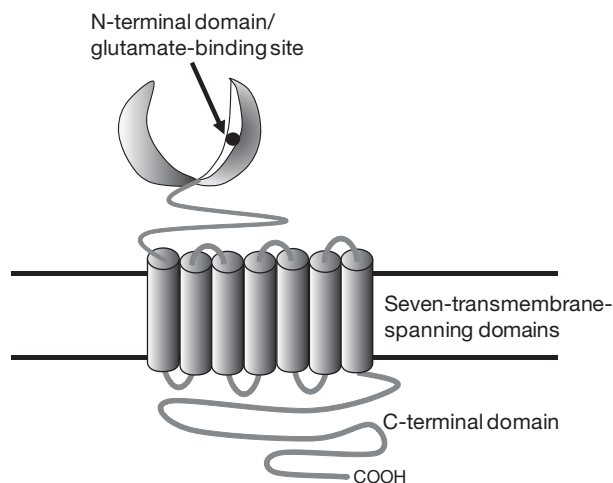


Figure 1 Schematic representation of an mGlu receptor. The mGlu receptors contain a large extracellular N-terminal domain, seven-transmembrane-spanning regions, and an intracellular C-terminal domain. Glutamate binds in the large extracellular domain. Reproduced from Conn PJ (2004) Glutamate receptors, metabotropic. In: *Encyclopedia of Biological Chemistry*, 1st edn. Oxford: Elsevier, with permission from Elsevier.

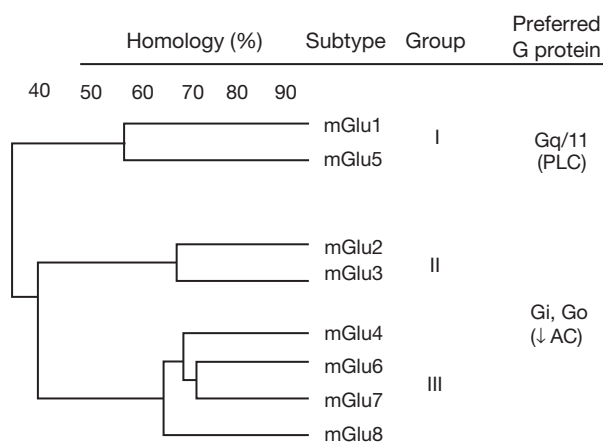


Figure 2 Dendrogram and classification of the members of the mGlu receptor family. The mGlu receptors are divided into three major groups, based on sequence homology, pharmacological profile, and second-messenger coupling. Reproduced from Conn PJ (2004) Glutamate receptors, metabotropic. In: *Encyclopedia of Biological Chemistry*, 1st edn. Oxford: Elsevier, with permission from Elsevier.

CNS. Anatomical studies using probes that specifically interact with individual subtypes coupled with electrophysiological studies have revealed that each mGlu receptor subtype is differentially localized in different brain regions. In recent years, all mGlu receptor subtypes have been genetically deleted in mice; studies using these animals have yielded further insights into the biological functions of mGlu receptors as well as potential diseases where mGlu receptors may be a viable target for therapeutic intervention. Although it is clear that the specific mGlu receptor subtype involved in mediating a given effect varies in different brain regions, some generalizations can be made regarding the major roles of different groups of mGlu receptors. Elucidation of these roles of mGlu receptors suggests that selective agonists and antagonists of specific mGlu receptor subtypes could subtly alter transmission in glutamatergic circuits in a way that would provide a therapeutic benefit without eliciting the side effects commonly associated with drugs that interact with members of the iGlu receptor family.

One of the most common physiological functions of mGlu receptors seen throughout the CNS is as autoreceptors present at presynaptic terminals at glutamatergic synapses. Activation of these presynaptic receptors results in a reduction of glutamate release at that synapse. It is likely that this provides a negative feedback mechanism for tightly controlling transmission at glutamatergic synapses. This presynaptic function of mGlu receptors is usually served by a group II mGlu receptor or by one of the group III mGlu receptors. Although multiple mGlu receptor subtypes can serve this role, the specific mGlu receptor subtype(s) involved varies from synapse to synapse. mGlu receptors also exist postsynaptically where they play an important role in regulating cell excitability. The most common postsynaptic effect of mGlu receptor activation is induction of an increase in neuronal excitability. Again, the specific mGlu receptor subtype involved varies from synapse to synapse but typically involves one of the group I mGlu receptors. In some cells, activation of postsynaptic mGlu receptors can lead to hyperpolarization and a reduction in neuronal excitability. This has been observed for both group I and group II mGlu receptors. Finally, a commonly observed effect of mGlu receptor activation is regulation of transmission at inhibitory (GABAergic) and neuromodulatory (acetylcholine, monoamines, and peptides) synapses. In this case, group II and III mGlu receptors are often found localized at presynaptic terminals or preterminal axons where they regulate neurotransmitter release. In addition to these effects on neuronal function, mGlu receptors can also exist on glial cells, the most prominent being mGlu3 and mGlu5. Although mGlu receptors can have a number of effects on glial function, Ferdinando Nicoletti and co-workers have reported a series of studies revealing that one important effect of mGlu receptors in glia is the induction of release of neuroprotective factors that reduce excitotoxicity and protect neurons from damage.

Therapeutic Potential of mGlu Receptor Ligands

Given the wide diversity, heterogeneous distribution, and diverse physiological roles of mGlu receptor subtypes, the opportunity exists for developing therapeutic agents that selectively interact with mGlu receptors involved in only one or a

limited number of CNS functions. A potential advantage of developing new therapeutic agents that specifically target mGlu receptors is that it provides a rational approach for subtle fine-tuning of activity at glutamatergic synapses in the circuits implicated in a particular disease. Because of this, a great deal of effort has been focused on developing small molecules that selectively activate or inhibit specific mGlu receptor subtypes. One of the most important and earliest breakthroughs in developing drugs that target mGlu receptors was the development of highly selective agonists for the group II mGlu receptors by Jim Monn, Darryle Schoepp, and their co-workers at Eli Lilly. Indeed, preclinical studies revealed robust anxiolytic and antipsychotic effects of these agonists, and subsequent clinical studies have shown that selective group II agonists have efficacy in the treatment of anxiety, panic attacks, and schizophrenia.

Preclinical studies of compounds selective for other mGlu receptor subtypes have also shown much promise for drug development. Hyperactivity of glutamatergic neurotransmission is thought to underlie a number of different CNS disorders and, in certain brain regions, mGlu5 activation induces the activity of the *N*-methyl-D-aspartate (NMDA) subtype of iGlu receptors. Thus, inhibition of mGlu5 function by selective negative allosteric modulators is an attractive avenue for therapeutic development for fragile X syndrome, anxiety disorders, chronic pain, addiction, depression, migraine, and gastroesophageal reflux disorder. Conversely, hypofunction of glutamatergic synapses has been implicated in the pathophysiology of schizophrenia and cognitive disorders, such that activation of positive allosteric modulators that enhance mGlu5 activity resulting in the potentiation of NMDA currents is a promising novel approach for therapeutic intervention. Another potential target is mGlu4, as activation of mGlu4 reduces GABAergic transmission at synapses that have been implicated in the pathology of Parkinson's disease. Indeed, selective enhancement of mGlu4 activity by positive allosteric modulators has been shown to have anti-parkinsonian effects in preclinical animal models. Over the past decade, our understanding of mGlu receptor biology has grown considerably, with the discovery of compounds that selectively activate/inhibit individual mGlu subtypes, providing many insights into the biological functions and pathologies associated with these receptors.

See also: **Bioenergetics:** Ligand-Operated Membrane Channels: Calcium (Glutamate); **Signaling:** GABA_B Receptor; Gi Family of Heterotrimeric G Proteins; G_q Family.

Further Reading

- Conn PJ, Lindsley CW, and Jones CK (2009) Activation of metabotropic glutamate receptors as a novel approach for the treatment of schizophrenia. *Trends in Pharmacological Sciences* 30: 25–31.
- Conn PJ and Pin JP (1997) Pharmacology and functions of metabotropic glutamate receptors. *Annual Reviews of Pharmacology and Toxicology* 37: 205–237.
- Coutinho V and Knopfel T (2002) Metabotropic glutamate receptors: Electrical and chemical signaling properties. *Neuroscientist* 8: 551–561.
- D'Antoni S, Berretta A, Bonaccorso CM, et al. (2008) Metabotropic glutamate receptors in glial cells. *Neurochemical Research* 33: 2436–2443.
- Fagni L, Chavis P, Ango F, and Bockaert J (2000) Complex interactions between mGluRs, intracellular Ca²⁺ stores and ion channels in neurons. *Trends in Neuroscience* 23: 2380–2388.

- Houamed KM, Kuijper JL, Gilbert TL, et al. (1991) Cloning, expression and gene structure of a G protein-coupled glutamate receptor from rat brain. *Science* 252: 1318–1321.
- Kim CH, Lee J, Lee JY, and Roche KW (2008) Metabotropic glutamate receptors: Phosphorylation and receptor signalling. *Journal of Neuroscience Research* 86: 1–10.
- Masu M, Tanabe Y, Tsuchida K, Shigemoto R, and Nakanishi S (1991) Sequence and expression of a metabotropic glutamate receptor. *Nature* 349: 760–765.
- Niswender CM and Conn PJ (2010) Metabotropic glutamate receptors: Physiology, pharmacology, and disease. *Annual Reviews in Pharmacology and Toxicology* 50: 295–322.
- Pinheiro PS and Mulle C (2008) Presynaptic glutamate receptors: Physiological functions and mechanisms of action. *Nature Reviews in Neuroscience* 9: 423–436.
- Swanson CJ, Bures M, Johnson MP, et al. (2005) Metabotropic glutamate receptors as novel targets for anxiety and stress disorders. *Nature Reviews in Drug Discovery* 4: 131–144.

Glutathione Peroxidases

F Ursini and M Maiorino, University of Padua, Padua, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Antioxidant A molecule or an enzymatic system capable of slowing or preventing the oxidation of other molecules. Oxidation is a chemical reaction that transfers electrons from a substance to an oxidizing agent. Oxidation reactions can produce free radicals, which start chain reactions that damage cells. Antioxidants terminate these chain reactions by removing free radical intermediates, and inhibit other oxidation reactions by being oxidized themselves.

Apoptosis The process of programmed cell death that occurs in multicellular organisms. Programmed cell death involves a series of biochemical events leading to change of cell morphology and death.

Glutathione A cysteine-containing tripeptide (γ -glutamyl-cysteinyl-glycine) found in most forms of aerobic life. Glutathione exists in reduced (GSH) and oxidized (GSSG) states. GSH is involved in antioxidant protection and regulation of the redox status of the cells. Here it is usually found almost exclusively in its reduced form, since the enzyme that reverts GSSG in a NADPH-dependent reaction, glutathione reductase, is constitutively active and inducible upon oxidative stress.

Lipid peroxidation The oxidative degradation of lipids. It is the process whereby free radicals steal electrons from the lipids (mainly poly-unsaturated) in cell membranes, resulting in cell damage. This process proceeds by a free radical chain reaction mechanism and produces lipid hydroperoxides and their degradation products.

Mitochondrial capsules A keratinous-like structure containing a large amount of disulfides, wrapping the mitochondria in the mid-piece of mammalian spermatozoa.

Peroxidase An enzyme belonging to the class of oxidoreductases catalyzing the reduction of a hydroperoxide to the corresponding alcohol and water by using the reducing potential of a specific substrate.

Peroxiredoxins A ubiquitous family of peroxidases not homologous to the glutathione peroxidase (GPx) family of

proteins. Peroxiredoxins use Cys catalysis and thioredoxins or other redoxins as donor substrate for hydroperoxide reduction. Prxs are divided into three classes: typical 2-Cys Prxs, atypical 2-Cys Prxs, and 1-Cys Prxs. The atypical 2-Cys Prxs use the same mechanism as typical 2-Cys Prxs but, differently to these, are functionally monomeric proteins. In atypical 2-Cys Prxs, both the peroxidatic Cys and its corresponding resolving Cys are contained within the same polypeptide, with the condensation reaction resulting in the formation of an intramolecular disulfide bond.

Phosphatidylcholine hydroperoxide A membrane lipid hydroperoxide, which is a nonradical intermediate of membrane lipid peroxidation.

Selenium A chemical element represented by the chemical symbol Se. It is chemically related to sulfur and tellurium. Although toxic in large amount, Se is nutritionally essential in traces for animals, being necessary for the functions of specific selenoenzymes (at least 25 in humans), some of which are essential for life. Selenoenzymes contain Se in the form of selenocysteine (three letter code: Sec, one letter code: U) as the redox active moiety. Sec is inserted in the protein during translation and it is therefore the 21st naturally occurring protein amino acid.

Selenoproteins Proteins that contain at least one selenocysteine residue. Selenocysteine has a structure similar to cysteine, but with an atom of selenium substituting for sulfur.

Sperm mitochondria cysteine-rich protein (SMCP)

A protein that is a structural component of the sperm mitochondrial capsule. It contains unusual motifs made of adjacent Cys residues (CC).

Thiol A compound that contains the functional group composed of a sulfur-hydrogen bond ($-SH$).

Thioredoxins and redoxins Proteins that act as reductants by reducing disulfides by a cysteine thiol-disulfide exchange reaction. They are characterized by the CXXC motif.

They exhibit a typical tertiary structure named the 'thioredoxin fold'.

The term 'glutathione peroxidase' (GPx) describes a family of phylogenetically related proteins. It was coined after GPx-1, the mammalian tetrameric selenoenzyme first described and characterized.

GPx-1 reduces various hydroperoxides by glutathione (GSH):



So far, the GPx family encompasses more than 700 members spread over all domains of life, mostly only known as deduced sequence.

History

In 1957, G. C. Mills described glutathione peroxidase (GPx) activity in red blood cells and showed that it could protect hemoglobin from oxidative breakdown. In 1973, the group of Hoekstra in Wisconsin reported that in a model of rat Se-deficiency reproducing the human pathological condition of favism, GPx activity was low and radioactive selenium coeluted with GPx activity. Almost at the same time, the team of Flohé in Germany showed that the purified enzyme contained one selenium atom per subunit and studied the catalytic mechanism

and the peculiar kinetics. Soon the tetrameric enzyme, today known as GPx-1, emerged as the most efficient H_2O_2 removing system present in the cytosol of almost every mammalian cell.

The selenium moiety of GPx-1 was later identified as selenocysteine (Sec) in the laboratories of Al Tappel (USA) and A. Wendel (Germany), while, few years later, protein sequencing demonstrated that the Sec residue was inserted in the GPx-1 peptide chain as the 21st of the naturally occurring amino acids. The discovery of Se in GPx-1 enforced a revision of the appreciation of selenium as a just toxic, carcinogenic, and teratogenic element.

GPx-1 remained the only glutathione peroxidase known until 1982, when the team of F. Ursini in Italy described a monomeric glutathione peroxidase that, differently from GPx-1, protects membrane peroxidation by reducing phospholipid hydroperoxides. The new enzyme was named 'phospholipid hydroperoxide glutathione peroxidase' (PHGPx). PHGPx was fully acknowledged as a new glutathione peroxidase only in 1991, when primary structure studies showed that it was a new selenoprotein poorly related to GPx-1. Notably, the strongly conserved peptide loops constituting the active site of GPx-1 were retained in PHGPx. PHGPx is known today as GPx-4. In the meanwhile, other GPx homologs were described in mammals.

To date, the list of mammalian GPxs has grown up to eight. Most of them are selenoproteins (SecGPxs) often coexisting in different cell types (GPx-1, GPx-2, GPx-4, and GPx-6, which is indeed mainly expressed in embryo and olfactory epithelium). One of them, GPx-3, is extracellular. In the remaining variants, the active site Sec residue is replaced by Cys (CysGPxs) (GPx-5, GPx-6 in some species, GPx-7, and GPx-8). Only GPx-4, GPx-7, and GPx-8 are monomeric while the others are tetrameric. The alternative use of Cys or Sec as the peroxidatic residue ($\text{C}_\text{P}/\text{U}_\text{P}$) brought into the focus the still challenging question of the biological advantage of having sulfur rather than selenium in these enzymes.

Accumulation of genomic databases allowed the analysis of homologous proteins from diverse taxonomical groups. From these studies it emerges that the SecGPxs are just a minority, only found in mammalian and in nonmammalian vertebrates, in parasitic trematodes, and, sporadically, in protists and bacteria. On the other hand, monomeric CysGPxs are the most abundant type of glutathione peroxidases, found in terrestrial plants, insects, bacteria, fungi, and protozoa. Most of these monomeric homologs contain a nonaligned second Cys residue that represents a second redox center (resolving Cys, C_R) conferring specificity for redoxins as reducing substrate. Therefore, the historical term 'glutathione peroxidase' turns today to be a misnomer that when all domains of life are considered, it functionally describes just a small subgroup of GPxs.

Enzymology and Kinetics

The kinetic mechanism of GPxs has been first worked out for the Sec-containing homologs, GPx-1, GPx-3, and GPx-4. The reaction encompasses two independent events: oxidation of the reduced enzyme by a hydroperoxide (ROOH) and reduction of the oxidized enzyme by GSH (Figure 1).

In the oxidative phase of the catalytic cycle, the selenolate at the active site is oxidized by the hydroperoxide. This first step is

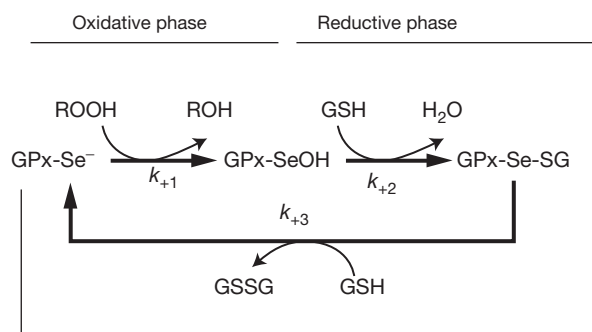


Figure 1 Scheme of the kinetic mechanism of SecGPxs. In the oxidative phase, the reduced enzyme is oxidized to form a putative selenenic acid derivative. For this reaction, the rate constant k_{+1} is identical to that obtained by the kinetic analysis ($k_{+1} = k'_{+1}$). In the reductive phase, the putative selenenic acid derivative of the enzyme is reduced back in two steps by GSH. The product of the first reaction is a glutathionylated intermediate that is resolved by a second molecule of GSH. The rate constant for these two steps, k_{+2} and k_{+3} , cannot be separated by steady-state kinetic analysis. This indeed restitutes the cumulative k'_{+2} value, where $k'_{+2} = k_{+2} + k_{+3}$. For simplicity, the enzyme–substrate complexes are not depicted, and only forward reactions are considered, although enzyme–substrate complexes may be postulated at least in the reducing phase of catalysis and the two reducing steps may be reversible (see text for details).

assumed to occur without formation of a typical enzyme–substrate complex, being the active site selenolate accessible and reactive enough to promptly reduce any hydroperoxide group that comes close. Accordingly, for SecGPxs the apparent rate constant for this step (k_{+1} in Figure 1) is among the fastest ever determined for bimolecular enzymatic reactions (around $10^8 \text{ M}^{-1} \text{ s}^{-1}$). The product is assumed a selenenic acid derivative, although a chemical evidence is missing, due to the extreme reactivity of the group.

The reductive phase is more complex. The overall reduction of the oxidized selenium proceeds by two subsequent steps, the first yielding the glutathionylated enzyme, that is, a selenodisulfide. In the second reduction step, a second GSH molecule forms the stable end product GSSG and releases the reduced selenium for the next cycle.

As in the oxidative step, in the reductive steps the kinetic pattern did not reveal any enzyme–substrate complex. Thus, even if an enzyme–substrate complex maybe invoked for each of the two reductive steps, its decay is not rate limiting. In summary, the kinetics of GPxs is described by a ping-pong mechanism which does not comply with the Michaelis–Menten assumption, where the decay of an enzyme–substrate complex is the rate-limiting step and leads to a saturation kinetics.

The pertinent Dalziel equation that describes this kinetic behavior is the following:

$$[\text{E}_0]/v_0 = 1/k'_{+1} \cdot [\text{ROOH}] + 1/k'_{+2} \cdot [\text{GSH}]$$

where k'_{+1} and k'_{+2} are the apparent rate constants for the net forward oxidative and reductive part of the catalytic cycle, respectively.

Ping-pong kinetics with substrate saturation have been reported for the Cys homolog of *Plasmodium falciparum* and more recently also for other CysGPxs, but not with the GPx

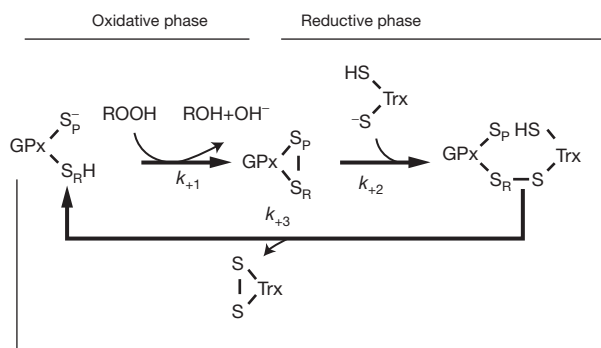


Figure 2 Scheme of the kinetic mechanism of the monomeric CysGPxs carrying the resolving Cys residue. Typically, these proteins mimic the atypical 2-Cys peroxidoredoxin catalysis and use redoxins as substrate for the reductive phase (see text). Here, the reaction with Trx is depicted. S_P's indicate the thiol group of the GPx peroxidatic Cys and S_R's the thiol group of the GPx resolving Cys. See legend of **Figure 1** for details.

from *Drosophila melanogaster*. These are enzymes that, like many other nonmammalian GPxs, mimic the catalysis of atypical 2-Cys peroxidoredoxins in terms of mechanism and specificity (**Figure 2**).

After oxidation of the peroxidatic Cys (C_P), a disulfide bridge is formed with the C_R, which is attacked at the sulfur of C_R by the reducing substrate, which here is Trx or other redoxins. Thus, in GPxs, the switch to saturation kinetic is not necessarily associated with the peroxidoredoxin-like mechanism and therefore the kinetic mechanism may change with species, the substrates involved, or even testing conditions.

Rate constants (i.e., k'_{+1} and k'_{+2} values) calculated from steady-state kinetics for some SecGPxs, for the chimeric cysteine variant of GPx-4 and for some naturally occurring CysGPxs, reveal that while the SecGPxs are more efficient, also the CysGPxs can deal with a physiologically relevant hydroperoxide removal.

In the SecGPx homologs, the reductive part of the catalytic cycle is the rate-limiting step of the overall reaction. In addition, k'_{+2} values decline from GPx-1 to GPx-3 and GPx-4, suggesting that among the SecGPxs, GPx-4 has the lowest reactivity with GSH. The nonmammalian cysteine variants are specifically and efficiently reduced by redoxins and show only marginal activities with GSH.

Substrate Specificity

Glutathione peroxidases are poorly specific in respect to the oxidizing substrates. However, differences do exist and have important biological outcomes. GPx-1 has been shown to reduce a large variety of organic hydroperoxides, although complex lipid hydroperoxides, such as phosphatidylcholine (PCOOH), cholesterol (ChOOH), and cholesterol ester hydroperoxide (CEOOH), are not substrates. In mammals, PCOOH is preferentially reduced by GPx-4 (k'_{+1} values of $10^7 \text{ M}^{-1} \text{ s}^{-1}$ for human and porcine GPx-4 have been measured), which reduces ChOOH and CEOOH as well. Reduction of PCOOH has also been observed for the mammalian extracellular GPx-3, but the rate constant for this reaction is two orders of magnitude lower than that of GPx-4. Furthermore, GPx-3 does not react with CEOOH. Activity with PCOOH is also generally

observed in nonmammalian CysGPxs that, like GPx-4, are monomeric. In mammals, the reduction of complex hydroperoxides is relevant on the light of the anti-apoptotic role of GPx-4 (see below).

Taking into account all life domains, the prevailing reducing substrates in the GPx family are thioredoxins (Trxs) or other redoxins. GSH, instead, is the preferred substrate of the mammalian SecGPxs. GPx-4 reacts 5 times as fast with dithiothreitol than with GSH, almost equally fast with mercaptoethanol and significantly also with cysteine, when none of them is accepted by GPx-1. The ability of GPx-4 to react with a large variety of protein thiols is linked to its function in spermatogenesis (see below). These protein substrates include chromatin, fragments of the sperm mitochondrion-associated cysteine-rich protein (SMCP) where adjacent Cys residues are oxidized.

Structure

Altogether, almost 20 structures of GPx homologs are deposited in the data bank. They display a tertiary structure similar to that first observed for Trx, and shared with many families of oxidoreductases, consisting of a four-stranded β -sheet and three flanking α -helices that yield a two-layer $\alpha/\beta/\alpha$ sandwich with a $\beta_1\alpha_1\beta_2\alpha_2\beta_3\beta_4\alpha_3$ secondary structure pattern. In GPxs the Trx-like fold contains additional stretches of amino acids (**Figure 3**).

Two additional β -sheets are located at the N-terminus, one α -helix and one β -sheet between β_2 and α_2 . Notably, within this additional α -helix lies the C_R in the majority of the Cys homologs and, in the tetrameric homologs only, this stretch of amino acids extends a little further to form an inter-monomer interface. Another additional stretch of amino acids is located between α_2 and β_3 . This participates to the tetrameric interface of GPx-1 and therefore it is deleted in the monomeric homologs. A typical feature of Trx-like fold is a conserved Cys-X-X-Cys active site motif that in the majority of GPx is changed to a Sec/Cys-X-X-Thr (**Figure 4**). This strongly conserved Thr residue appears to be important for structural stability.

The active site pocket of GPxs lies in a flat cleft of the protein surface. The redox active Se, or S, is at hydrogen-bonding distances from three strongly conserved residues, a Gln, a Trp, and an Asn located in distant loops in the primary structure (**Figures 4** and **5**). These residues represent the signature of the entire family and contribute to catalysis.

Considering that the bimolecular rate constants for the oxidation of a fully dissociated Cys by H_2O_2 do not exceed $21 \text{ M}^{-1} \text{ s}^{-1}$, and that theoretical predictions of the unknown rate constants for free Sec oxidations yields similar values, to explain the surprising velocity of the enzymatic reduction of ROOH by GPx, the tetrad of the GPxs catalytic center must play a peculiar role. Tentatively it was assumed that the active site has to fulfill the following requirements: (1) dissociation of the active site (seleno)cysteine to make it as nucleophilic as possible, (2) polarization of the substrate hydroperoxy group, and (3) proton donation for generating an ROH as a good leaving group.

Comparison of the active sites of the two extremes, human GPx1 and human GPx-4, in respect to accessibility and

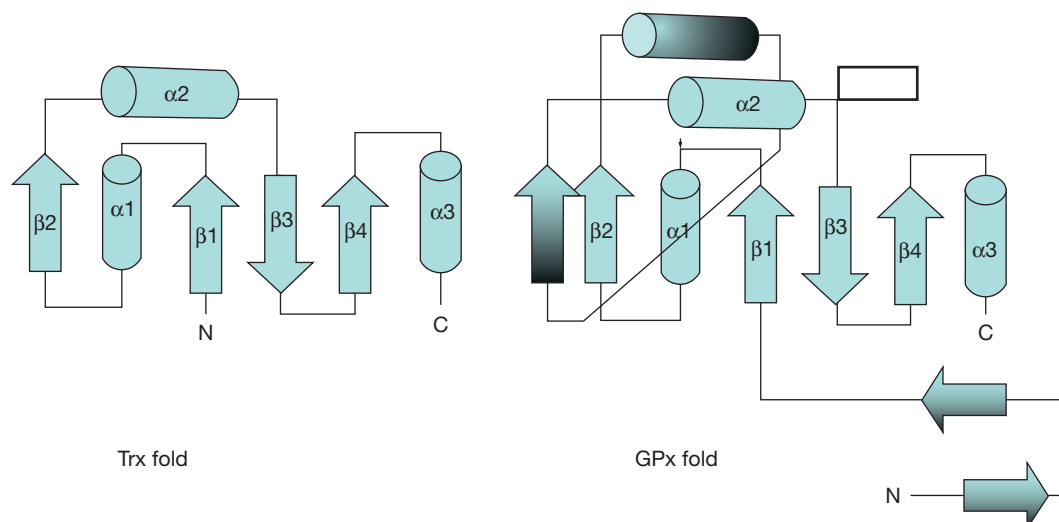


Figure 3 Topological diagram for Trx fold (left) and GPx fold (right). The additional sheets and helices of the GPx fold are shown in dark. A tiny arrow indicates the position of the selenocysteine residue in SecGPxs. The bold line represents the stretch of amino acids composing the tetrameric interface that is deleted in the monomeric members. In the invertebrate and plant CysGPxs, the C_R residue is located typically in a nonconserved position within the additional α -helix (not shown).

GPx-1_Hs	1	-----MCAARLAAAAAQA	SVYA	FSARPLAGGEP	VSLGSLRGKVL	LI	ENVAS	L48
GPx-2_Hs	1	-----MAFIAKSFYDLSA	ISLDG	EKVD	FNTRFRGRAV	LI	ENVAS	L39
GPx-4_Hs	1	-----MCASRDDWRCAR	SMHE	FSAKDIDG	HMVNLDKYRGFVC	IV	TNVAS	Q45
GPx-7_Hs	1	MVAATVAAAWLL	LWAAACAQQEQDF	YDFKAVNIRG	KLVSLEKYRGSVS	LV	VNVAS	E56
GPx3_S. cerevisiae	1	-----MS-----	EFYK	LAPVDKKG	QPF	FPDQLKGKVV	LI	VNVASK35
GPx2_A. thaliana	1	-----MADE-----	SPKSI	YDFTVKDIG	NDVSLDQYK	GKTL	LV	VNVASK40
GPx-1_Hs	49	UGTTVRDY	TQMNEL	QRRRLGPRGL	VVLGF	FPCNQFGHQENAK	NEEI	LNLSLKYVRPGGGF105
GPx-2_Hs	40	UGTTTRDFTQ	LNELQCRF	-PRRLVVLG	FPCNQFGHQENCQ	NEEI	LNLSLKYVRPGGGY95	
GPx-4_Hs	46	UGKTEVNY	TQLVDLHARYAEC	GLRILAF	FPCNQFGKQEPGS	NEEI	KEFAAG	-----Y96
GPx-7_Hs	57	CGFTDQHYRA	LQQLQRDLGPHHF	NVLA	FPCNQFGQQEPDS	NKEI	IESFARRT	-----Y108
GPx3_S. cerevisiae	36	CGFT	-PQYKELEALYKRYK	DEGFTI	IGFPCNQFGHQEPGS	DEEI	IAQFCQLN	-----Y86
GPx2_A. thaliana	41	CGLTDANY	KELVNLYE	KYKEQGLEI	LA	FPCNQFLGQEPGN	NEEI	QQTVCTR-----F92
GPx-1_Hs	106	EPNFMLFE	KCEVNGAGAHPL	FAFL	REALPAPSD	DDATALMTDPKL	ITWSPVCRND	VAW162
GPx-2_Hs	96	QPTFTLVQK	CEVNGQNEHPV	FAYLKD	KLPYPYDD	PFSLMTDPKL	IWSPVRRSD	VAW152
GPx-4_Hs	97	NVKFDMFS	KICVNGDDAHL	LWKWMKI	QP	-----	KGKGI	LGNAIKW136
GPx-7_Hs	109	SVSFPMFS	KIAVTGTGAHP	AFKYLA	QT	-----	-----	SGKEPTW142
GPx3_S. cerevisiae	87	GVTFPIMKK	IDVNGGNEDPV	YKFLKSQ	-----	-----	KSGMLGLRG	IKW125
GPx2_A. thaliana	93	KAEFP	IFDKVDVNGKNTA	PLYKYLKAE	-----	-----	KGGLL	-IDAIKW130
GPx-1_Hs	163	NFEKFLVGP	DGVPLRRYSRR	FQTD	IEPDIEA	LLSQGPSCA	----	203
GPx-2_Hs	153	NFEKFLIG	PEGEFFRRYSRT	FPTINI	EPDIKRL	LKVAI	-----	190
GPx-4_Hs	137	NFTKFL	IDKNGCVVKRYG	PMEEPLVIE	KDLPHY	-----	-----	169
GPx-7_Hs	143	NFWKYLVA	PDGKVVGAWDPT	VSVEEVRPQ	ITALVRKLI	LLKREDL	-----	187
GPx3_S. cerevisiae	126	NFEKFLVD	KKGKVYERYSSL	TKPSSLS	ETIEELL	KEVE	-----	183
GPx2_A. thaliana	131	NFTKFLVS	PDGKVLQRYSPRT	SP	LQFEKDI	QTALGQASS	-----	169

Figure 4 Comparison of some GPx homologs. The sequence of some SecGPxs and CysGPxs – that is, the GPx-1, GPx-2, and GPx-3 from *Homo sapiens* and GPx-7, GPx3, and GPx2 from *Homo sapiens*, *Saccharomyces cerevisiae*, and *Arabidopsis thaliana*, respectively – have been taken as examples. U indicates the selenocysteine residue. The arrows at the bottom indicate the four residues involved in the catalysis. Note the C/UxxT motif typical of GPxs. The arrows at the top indicate the basic residues that, in GPx-1, have been proposed to direct GSH toward the catalytic center. The gray and the black thick lines indicate the dimer interface and the tetramer interfaces, respectively. The thin line indicates the block that typically contains the resolving Cys (C_R) in the invertebrate CysGPx. Note that the amino acids composing the dimer and tetramer interfaces are deleted in the monomeric homologs. See text for details.

electrostatic environment of the peroxidatic Sec residue (U_P), might offer explanation for the peculiar reactivity with PCOOH of GPx-4. First, as in the other monomeric isoforms, the tetrameric interface is deleted. Since this stretch of amino acids partially shields the active site Trp residue in the

tetrameric homologs, it is believed that its absence facilitates accessibility to complex lipid substrates. Consistently, reactivity with PCOOH and other complex lipid substrates is generally a shared feature of the monomeric GPxs. Second, in both human GPx-1 and GPx-4, the selenium protrudes from a flat

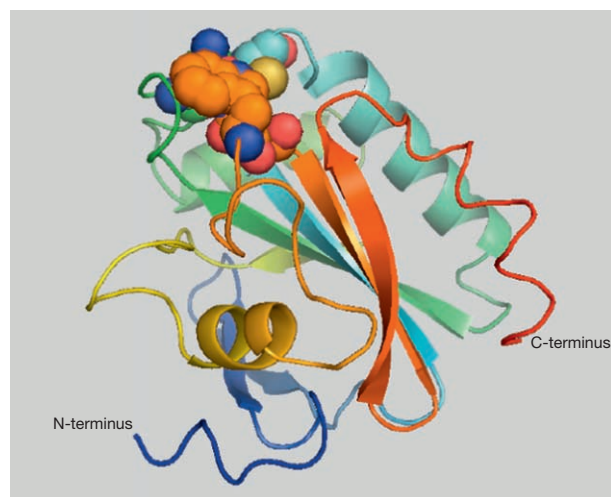


Figure 5 Ribbon representation showing the overall structure of the U46C mutant of the human GPx-4. The active site residues, including the peroxidatic residue here mutated into Cys (Cys 46), Gln 81, Trp 142, and Asn 143, are shown as spheres. The yellow sphere represents the sulfur of the Cys replacing the selenocysteine.

area made up by some hydrophobic and basic residues that lend a cationic character to the active site. Besides the positive charges in close proximity to U_p , a second large positively charged area is present in human GPx-4. In the tetrameric congeners, this part is largely covered by a loop that forms the subunit interface. In GPx-4, this second cluster of positive charges is expected to contribute to the binding of exposed polar heads of phospholipids, conferring to GPx-4 the ability to catalyze interfacial reactions.

Within the mammalian SecGPxs, the specificity for the reducing substrate may be explained also on the ground of structural considerations. It has been proposed that in GPx-1 four Arg and one Lys residues provide an electrostatic architecture that directs the donor substrate GSH toward the catalytic center, in a way that its sulfhydryl group contacts the oxidized selenium. In the other GPxs, the peculiar architecture of the GSH binding site is conserved only partially in GPx-2 and GPx-3 and not at all in GPx-4 (Figure 4); therefore, the specificity of GPxs toward GSH should decline in the following order: GPx-1 > GPx-2 > GPx-3 > GPx-4.

Thioredoxin specificity appears to be the domain of CysGPxs that, containing an intramolecular C_R , mimics the atypical 2Cys peroxiredoxin catalysis. Besides C_R , the other peculiar feature of GPxs to determine Trx specificity is deletion of the tetramerization interface, leading to flexibility of the C_R -containing loop. The disulfide bridge in the oxidized enzyme provides the canonical motif for the reduction by a CXXC protein. An intriguing mechanistic variation of the CysGPx homologs is that of the *Saccharomyces cerevisiae* GPx-3, that *in vivo* can oxidize a specific transcription factor, as discussed in the next section.

Biological Functions

Hydroperoxides are unstable compounds produced in cells either enzymatically or as byproducts of oxygen metabolism.

Their homo- or heterolytic decomposition, in many inflammatory diseases, produces reactive species damaging tissues. Therefore, by removing hydroperoxides, GPxs have for long been considered for their antioxidant and anti-inflammatory function.

More recently, it has been recognized that hydroperoxides are also involved in cell growth, differentiation, proliferation, and apoptosis, most likely by interacting with different signaling mechanisms where oxidation-prone cysteines undergo a functional redox switch. Thus, GPxs play a role in these processes by dampening the oxidant signal and/or targeting the oxidative signal to specific protein thiols. An interesting example is that of *S. cerevisiae* GPx-3. In this specie, the Yap1 transcription factor regulates hydroperoxide homeostasis and is activated when hydroperoxide level increases. When oxidized by H_2O_2 , GPx-3 C_{p36} bridges Yap1 Cys598 by a disulfide bond. This intermolecular disulfide bond is then reshuffled into a Yap1 intramolecular disulfide, giving rise to the activated form of the regulator. Thioredoxin turns off the pathway by reducing both sensor and regulator. Thus, yeast GPx3 works as H_2O_2 receptor and redox transducer in gene activation by oxidizing specific thiols of a transcription factor.

The role of mammalian SecGPx as 'effector' of oxidative stress has been only proved for the mitochondrial and nuclear forms of GPx-4. The oxidation of these proteins within spermatozoa results indeed in oxidation of critical thiols relevant to the process of sperm cell maturation (see below).

On the other hand, reverse genetics suggest that GPx-2 and the cytosolic form of GPx-4 may be involved in regulating signaling cascades depending on hydroperoxides, through still not fully defined mechanisms.

What is presently known about the biological functions of mammalian SecGPxs is briefly summarized below.

H_2O_2 metabolism is the domain of GPx-1. This enzyme has been found practically in all investigated tissue, although very low in male germ cells. GPx-1 knockout (k.o) mice model confirms the antioxidant role. These mice are vital, grow and develop normally, but are extremely sensitive to oxidative challenges (i.e., exposure to redox cyclers or induction of oxidative burst in the phagocytic system). Worthy of note is the evidence that the phenotype is not affected by selenium status. This indicates that none of the other SeGPxs, nor any other selenoprotein or peroxiredoxins – indeed functionally depending on the selenoproteins thioredoxin reductases – can compete with GPx-1 in the defense against a generalized oxidative challenge.

Interestingly, over-expression of GPx-1 in mice leads to obesity and insulin resistance, which underscores the role H_2O_2 in the signaling cascade of insulin.

GPx-2, the glutathione peroxidase known as 'gastrointestinal', is mainly expressed in the whole gastrointestinal tract, including the squamous epithelium of the esophagus. In humans, it is also detectable in the liver. The enzyme is part of the battery of genes that are up-regulated by the Nrf2/Keap1 system and it is over-expressed in cancer cells, where this pathway is often activated. Similarly to GPx-1, also GPx-2 k.o mice do not display any significant phenotype in unstressed conditions. The k.o of both these enzymes, however, yields a model of inflammatory bowel disease. Although GPx-2 activity in the gut makes up at best 50% of the total GPx activity, only one allele of GPx-2, but not of GPx-1, is enough to rescue

the phenotype. Clearly, a specific role has to be attributed to GPx-2 that is not substituted by GPx-1. This seems to be somehow related to the control of apoptosis. Indeed, shortened intestinal villi have been observed in GPx-2 k.o mice, suggesting an increased but still tolerated apoptosis rate in these animals. Furthermore, there is less apoptosis in GPx-2 over-expressing MCF7 cells, indicating that GPx-2, by down-regulating apoptosis, promotes cell growth.

GPx-3 is also known as plasma glutathione peroxidase. The GPx-3 transcript contains a signal peptide sequence that is cleaved off in the mature protein. Accordingly, GPx-3 messenger RNA (mRNA) has been found in several tissues and the protein in extracellular compartments such as amniotic fluid, lining fluid of the lung, thyroid lumen, and chamber water of the eye. The major source of plasma GPx-3 is, however, the kidney. GPx-3 is secreted primarily by proximal convoluted tubule cells and binds to the extracellular matrix and to the basement membrane of tubules. It has been proposed that bound GPx-3 is a reservoir of enzyme, mobilized when needed to contrast an oxidative challenge in plasma. However, the role of GPx-3 is not obvious, since reducing equivalents are extremely limited in plasma and extracellular fluids.

GPx-3 k.o mice do not display any significant phenotype, at least under unstressed conditions, also when deletion is associated with deletion of GPx-1.

GPx-4 is an intracellular enzyme expressed at low level almost ubiquitously, with the exception of male germ cells of testis where expression is high, and the glial cells of the brain, where, instead, it is undetectable. Notably, the absence of GPx-4 mRNA has also been reported in the islets of Langerhans, in the adenohypophysis, and in the lymphoid follicles of the white pulp of the spleen.

GPx-4, also known as PHGPx, performs the broadest variety of functions. As mentioned above, GPx-4 is expressed as a mitochondrial, cytosolic, or nuclear isoform. Deletion of the whole GPx-4 gene is lethal: GPx-4 k.o mice die *in utero* at midgestation. Individual deletions of the mitochondrial and nuclear GPx-4 allowed establishing that the nuclear GPx-4 has a marginal role in fertility, contributing to sperm chromatin condensation. The mitochondrial GPx-4 instead is more relevant in this process, the k.o of this form resulting in infertility. By exclusion, therefore, the cytosolic GPx-4 emerges as the essential form for embryonic development and apoptosis regulation.

The role in fertility of the nuclear and mitochondrial GPx-4 is related to its limited specificity for the reducing substrate. During late spermatogenesis, when the GSH levels fall down, mitochondrial GPx-4 uses the thiol groups of SMCP, a Cys-rich protein of the sperm mitochondrial capsule, as alternate reducing substrate. This yields GPx-4 as an enzymatically inactive structural component of the sperm mitochondrial capsule, a process indispensable for appropriate sperm function. Nuclear GPx-4 works similarly on chromatin, but the consequences are less dramatic, nuclear GPx k.o only exhibiting sperm head instability.

On the other hand, lethality due to GPx-4 deletion is apparently related to enzymatic activity on complex lipid substrates. Inducible GPx-4 inactivation reveals 12/15-lipoxygenase-derived lipid peroxidation as the specific downstream event, triggering apoptosis-inducing factor (AIF)-mediated cell death. In agreement with the anti peroxidative functions of GPx-4 first observed *in vitro*, cell death could be prevented by α -tocopherol,

12/15-lipoxygenase inhibitors, or small interfering RNA (siRNA)-mediated AIF silencing. Furthermore, neuron-specific GPx-4 depletion causes neurodegeneration *in vivo* and *ex vivo*, highlighting the primary role of GPx-4 in cell survival.

Conclusions and Future Perspectives

The growing information on the physiological function of GPxs has challenged the concept that proteins of similar sequence and conserved active site are redundant proteins playing similar physiological roles. While the expected antioxidant function of GPx-1 could be validated by k.o animal models, other family members exhibited a series of surprising functions: GPx-4, for instance, resulted to be relevant for fertility by oxidizing specific protein thiols and crucial for survival by inhibiting a lipoxygenase.

Unexpectedly, CysGPxs resulted to be almost as efficient as their Sec-containing homologs in reducing hydroperoxides. The majority of these enzymes, which typically are monomeric and contain a C_R residue within the functional helix, are linked to the thioredoxin and not to the GSH system. The discovery of changes in substrate specificity and mechanism was paralleled by unexpected novel roles of GPxs such as regulation of growth, peroxide sensing, and gene activation. These evidences focus GPxs as crucial players of redox homeostasis and contribute to highlight redox transition as a complementary mechanism to other posttranslational modifications. In the present scenario, understanding the molecular details of these signaling cascades and defining the specific role of the individual GPxs are expected to be extremely rewarding and full of surprises. Further, it appears indeed to be the ground for developing effective strategies to counteract major human pathological processes such as cancer and cerebral degeneration.

See also: [Bioenergetics: Cell Death by Apoptosis and Necrosis; Mitochondria: Source and Target of Free Radicals; Protein/Enzyme Structure Function and Degradation: Selenoprotein Synthesis.](#)

Further Reading

- Brigelius-Flohé R and Kipp A (2009) Glutathione peroxidases in different stages of carcinogenesis. *Biochimica et Biophysica Acta* 1790: 1555–1568.
- Delaunay A, Pflieger D, Barrault MB, Vinh J, and Toledano MB (2002) A thiol peroxidase is an H₂O₂ receptor and redox-transducer in gene activation. *Cell* 111: 471–481.
- Donovan J and Copeland PR (2010) Threading the needle: Getting selenocysteine into proteins. *Antioxidants and Redox Signaling* 12: 881–892.
- Flohé L (2009) The labour pains of biochemical selenology: The history of selenoprotein biosynthesis. *Biochimica et Biophysica Acta* 1790: 1389–1403.
- Kryukov GV, Castellano S, Novoselov SV, et al. (2003) Characterization of mammalian selenoproteomes. *Science* 300: 1439–1443.
- Maiorino M, Ursini F, Bosello V, et al. (2007) The thioredoxin specificity of *Drosophila* GPx: A paradigm for a peroxiredoxin-like mechanism of many glutathione peroxidases. *Journal of Molecular Biology* 365: 1033–1046.
- Toppo S, Flohé L, Ursini F, Vanin S, and Maiorino M (2009) Catalytic mechanisms and specificities of glutathione peroxidases: Variations of a basic scheme. *Biochimica et Biophysica Acta* 1790: 1486–1500.
- Tosatto SCE, Bosello V, Fogolari F, et al. (2008) The catalytic site of glutathione peroxidases. *Antioxidants and Redox Signaling* 10: 1515–1526.
- Wood ZA, Schroder E, Robin Harris J, and Poole LB (2003) Structure, mechanism and regulation of peroxiredoxins. *Trends in Biochemical Sciences* 28: 32–40.

Glycation

M Lima and J W Baynes, University of South Carolina, Columbia, SC, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Suzanne R. Thorpe and John Baynes, volume 2, pp. 229–236, © 2004, Elsevier Inc.

Glossary

Advanced glycation end product (AGE) Irreversible end product of nonenzymatic reaction of carbohydrates with protein; it includes fluorescent and nonfluorescent adducts and cross-links in protein.

Advanced lipoxidation end product (ALE) Stable end product of chemical modification of protein by reactive carbonyl intermediates formed during lipid peroxidation reactions.

Amadori adduct or compound First stable product of glycation of protein, a ketoamine.

Glycation Enzymatic or nonenzymatic adduction of a carbohydrate to a biomolecule.

Glycoxidation product AGE formed by a combination of glycation and oxidation chemistry.

RAGE Receptor for advanced glycation end products.

Introduction

Glycation is the most general term describing the adduction of a carbohydrate to another biomolecule, such as a protein, lipid, or DNA. Glycation may occur either enzymatically or nonenzymatically. The common term for enzymatic glycation is glycosylation, for example, formation of a glycosidic bond using a sugar nucleotide donor during synthesis of glycoproteins. The terms nonenzymatic glycation, nonenzymatic glycosylation, or glycation (without a modifier) are commonly used in reference to direct chemical reactions of reducing sugars with proteins, illustrated by the reaction of glucose with lysine residues in protein to form a ketoamine (Amadori) adduct (Figure 1). Glucation, fructation, ribation, etc. are used in reference to glycation by specific sugars.

Although glycation is a reversible reaction, it is considered a first step in the Maillard or browning reaction, which leads to irreversible chemical modification, browning, generation of fluorescence, and cross-linking of proteins during cooking. These same reactions proceed in the body at lower temperature and slower rates during normal aging. The irreversible adducts and cross-links in tissue protein are known as advanced glycation end products (AGEs). Most AGEs are formed by a combination of glycation and oxidation reactions and are termed glycoxidation products. AGEs accumulate gradually with age in long-lived extracellular proteins, such as collagen, and are also formed from glycolytic intermediates on shorter-lived intracellular proteins. Increased rates of formation of AGEs in tissue proteins during hyperglycemia and oxidative stress or inflammation are implicated in the pathophysiology of aging, and diabetes and other chronic diseases.

Glycation of Proteins in Blood

Glycated Hemoglobin

Interest in the Maillard reaction in biochemistry began with the identification of HbA_{1c}, an anodic variant of adult hemoglobin (Hb), discovered during screening of diabetic patients.

HbA_{1c} contains glucose bound to the amino-terminal valine residues of the Hb β -chains. The glucose–valine adduct was characterized as an aminoketose or ketoamine (fructosevaline), formed by nonenzymatic reaction of glucose with protein (Figure 1). The high reactivity of the β -chain valine residues of Hb results from catalysis of the Amadori rearrangement by phosphate or 2,3-bisphosphoglycerate bound near these valine residues in the allosteric site regulating the oxygen affinity of Hb. The α -chain valines, which have similar exposure to solvent and the same low pK_a , but are not in the allosteric site, are only $\sim 10\%$ as glycated as the β -chain valines. Hb has 66 lysine residues, but there are only about twice as many fructoselysine as fructosevaline residues in the protein. The more reactive lysines are adjacent to basic or acidic amino acid in the primary or three-dimensional structure of the protein, which either affect the nucleophilicity of the lysine amino group or catalyze the Amadori rearrangement. The primary sites of glycation of ribonuclease by glucose *in vitro* are at lysine residues in the active site. Glycation of some, but not all, of these residues is catalyzed by phosphate and other anionic buffers that bind in the basic, RNA-binding region of the enzyme. Thus, glycation, although not as specific as enzymatic glycosylation of protein, is not a random modification of protein, but is affected by the pK_a of the amino acid, amino acid sequence, protein conformation, and ligand binding.

Clinical Significance of Glycated Hb

Assays of glycated Hb (GlcHb) are a powerful tool for the clinical management of diabetes. Glycation is a slow reaction. The rate of formation of HbA_{1c} is $\sim 0.1\%$ of HbA molecules per day at normal blood glucose concentration, and is first order in glucose concentration. Since the red cell is long lived (life span ~ 120 days) and freely permeable to glucose, the mean extent of glycation of Hb varies with mean blood glucose concentration. The average percent glycation of Hb ranges from 4% to 6% HbA_{1c} in a control population, to 16% or higher in poorly controlled diabetic patients. Based on clinical studies, assays of percent HbA_{1c} integrate the mean blood

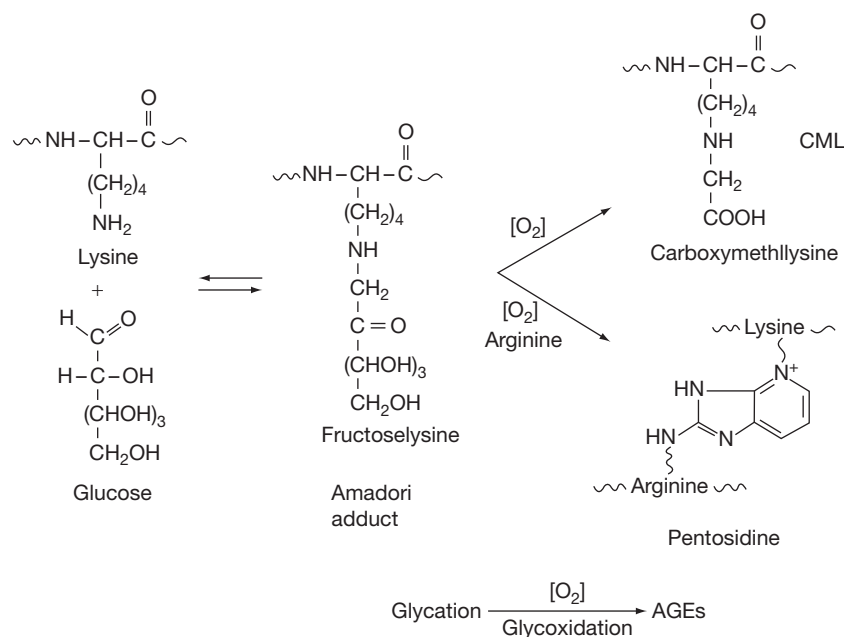


Figure 1 Pathway for glycation of protein. Glucose reacts reversibly with amino groups in protein (e.g., lysine), forming a Schiff base (not shown), which rearranges to the Amadori compound (ketoamine), fructoselysine. Oxidation of the Amadori compound is one route to formation of advanced glycation or glycoxidation end products, such as CML and pentosidine.

glucose concentration during the previous 4–6 weeks. These assays complement acute blood glucose measurements for assessment of long-term glycemic control in diabetes.

Assays for HbA_{1c} use an anion-exchange procedure that depends on the shift in pK_a of the protein, which occurs on glycation of the valine residues. Alternatively, total GlcHb may be measured by an affinity chromatography assay that depends on interaction of Amadori adducts on the surface of the protein with phenylboronate resins; this assay measures glycation of both valine and lysine residues. While assays of HbA_{1c} and GlcHb actually measure different aspects of GlcHb, the two assays yield similar quantitative results and correlate strongly with one another. Both the ion-exchange and affinity chromatography assays are available as mini-column kits. A high-performance liquid chromatography (HPLC) method is commonly used for the measurement of HbA_{1c} in clinical chemistry laboratories, and electrophoretic, nephelometric, and enzyme-linked immunosorbent assay (ELISA) assays are also used.

Glycation of Extracellular and Plasma Proteins

Glucose is ubiquitous in body fluids, and therefore all tissue proteins are subject to glycation. The steady-state extent of glycation of a protein depends on the inherent reactivity of the protein with glucose under physiological conditions, the protein's biological half-life or life span, and the ambient glucose concentration. The fractional glycation of lysine in lens proteins is ~40% that of skin collagen in humans, reflecting the lower glucose concentration in the lens, compared to extracellular fluids. Collagen accounts for about one-third of total body protein mass and is accessible by biopsy. Like glycation of Hb, the extent of glycation of skin collagen may be used as an index of very long term hyperglycemia.

Assays for glycated plasma proteins, which have relatively short half-lives compared to Hb or collagen, provide an index of intermediate-term blood glucose concentration, during the previous 7–10 days. In addition to ELISA and phenylboronate affinity chromatography assays for glycated albumin, total Amadori adducts on plasma proteins can be measured by the fructosamine assay. This assay measures the reduction of a tetrazolium dye by the Amadori compound on plasma proteins. The Amadori adduct has much stronger reducing activity than glucose at pH 10, so that the reducing activity of Amadori adducts on glycated plasma proteins can be measured without significant interference by blood glucose.

Glycation of Other Amino Acids and Biomolecules

Glucose appears to react only with primary amino groups on protein. However, Amadori compounds undergo facile rearrangement and oxidation reactions. They may rearrange nonoxidatively to form deoxyglucosones, regenerating free lysine, or undergo autooxidative decomposition to form glucosone. These dicarbonyl compounds react with the side chains of lysine, arginine, cysteine, and histidine residues, expanding the range of glycative damage to protein. Amadori adducts to phosphatidylethanolamine and phosphatidylserine have also been measured in tissues. The AGE, N²-carboxyethyl-2'-deoxyguanosine, has been detected in DNA of cells grown in high glucose concentration, and was recently reported to be increased in urine of diabetic versus control mice.

Reactivity of Glucose

The rate of reaction of a sugar with protein depends on the fraction of the sugar in the open-chain, aldehyde conformation.

Glucose, among all sugars, exists to the greatest extent in the cyclic, hemiacetal conformation. It is therefore the least reactive in glycation proteins, in comparison to other hexoses, pentoses, ketoses, or phosphorylated metabolic intermediates. Another unique feature of glucose is that it is the least active among all aldoses and ketoses in oxidation or autooxidation (oxidation by molecular oxygen) reactions – again, because it exists largely in a hemiacetal rather than aldehyde conformation. Its low reactivity with protein and its resistance to autooxidation may have been important in the evolutionary selection of glucose as blood sugar in mammals. Despite its chemical inertness, however, glucose does react with proteins at a measurable rate *in vivo* and glycation increases in concert with hyperglycemia in diabetes.

Advanced Glycation and Glycoxidation Reactions

Advanced Glycation

Food scientists learned early in the twenty-first century that glycation was one of the early steps in the Maillard or browning reaction of proteins during cooking, part of the process of caramelization that enhances the color, taste, aroma, and texture of cooked foods. The Maillard reaction also contributes to the loss in nutritional value of cooked foods, because of the chemical modification of essential amino acids, such as lysine and arginine. When biomedical scientists discovered that Hb was glycated *in vivo*, it was reasonable to speculate that advanced stages of the Maillard reaction should also proceed *in vivo* – the human body could be viewed, at one level, as a low-temperature oven (98.6 °F) with a long cooking cycle (~80 years). The Maillard reaction was seen as a possible source of the brown color and fluorescence that develops with age in long-lived proteins, such as lens crystallins and skin collagens, and as a contributing factor in both the normal aging of tissue proteins and the acceleration of age-like pathology by hyperglycemia in diabetes.

More than 20 AGEs have now been detected in tissue proteins (Figure 2), including amines, amides, pyrroles, imidazolones, imidazolium salts, and several bicyclic heteroaromatic compounds. The first two AGEs to be identified *in vivo* were N^ε-(carboxymethyl)lysine (CML) and pentosidine (Figure 1). Because of their stability during acid hydrolysis of proteins, these compounds are still the most frequently measured AGEs. CML is measured by GC/MS (GC, gas chromatography; MS, mass spectrometry) or ELISA, while pentosidine, which is present at much lower concentration, is highly fluorescent and is measured by HPLC. Other AGEs, such as the imidazolones and glucosepane, are present at concentrations comparable to CML, but are measured less frequently because they are labile to acid or base hydrolysis and must be measured following enzymatic digestion of proteins. The known AGEs are only trace components of proteins *in vivo*. Even in skin collagen and lens proteins of elderly diabetic patients, CML and N^ε-(carboxyethyl)lysine (CEL) (Figure 2) typically account for the modification of less than 1% of the lysine residues in the protein. However, these biomarkers are thought to represent only a fraction of the total chemical modification of amino acids and cross-linking of proteins *in vivo*.

Glycoxidation

Although oxygen is not required for glycation, formation of CML and pentosidine from the Amadori compound requires oxidation, for example, oxidative cleavage between C-2 and C-3 of the carbohydrate to form the lysine adduct CML or loss of one carbon from glucose during the formation of pentosidine. These oxidative reactions are catalyzed by trace amounts of free (decompartmentalized) redox-active, metal ions, such as iron and copper. The term glycoxidation was coined in recognition of the importance of both glycation and oxidation in the formation of AGEs. Oxidation reactions often have high-negative free energies, and, especially in oxidative cleavage reactions, oxygen can be viewed as the fixative of irreversible damage to protein during the Maillard reaction. However, some AGEs, such as pyrraline and the crosslines, may be derived from glucose under nonoxidative conditions from deoxyglucosones formed from Amadori compounds under nonoxidative conditions.

Alternative Pathways to AGEs and Glycoxidation Products

In the early 1980s, the Amadori adduct was considered an obligate intermediate in the formation of AGEs (Figure 1). However, it is now clear that there are multiple precursors and pathways to the formation of the same AGE (Figure 3). CML, for example, may be formed from glyoxal or glycolaldehyde which are produced on autooxidation of many carbohydrates, or by autooxidation of Schiff base intermediates and Amadori adducts, or from ascorbate and dehydroascorbate and phosphorylated intermediates in glycolysis. To complicate the issue further, glycolaldehyde formed on oxidation of serine by myeloperoxidase and glyoxal formed during lipid peroxidation reactions may also contribute to the formation of CML. Thus, the measurement of CML provides little information regarding its origin. Lipid-derived adducts to protein, such as malondialdehyde and hydroxynonenal adducts to protein, are known as advanced lipoxidation end products (ALEs), while CML and glyoxal lysine dimer (GOLD), and the homologous compounds, CEL and methylglyoxal lysine dimer (MOLD), are now considered AGEs/ALEs, reflecting the fact that they may be derived from a variety of carbohydrate and noncarbohydrate precursors *in vivo*. Some AGEs, such as pentosidine and the crosslines, which contain five to six carbon rings or side chains (Figure 2), are clearly derived from carbohydrates, but may be formed from a variety of carbohydrate precursors, including reducing sugars, sugar phosphates, and ascorbate. In general, AGEs, ALEs, and other protein oxidation products accumulate together in proteins, for example, in amyloid plaque in neurodegenerative disease.

AGE Receptors, Inhibitors, and Breakers

AGE Receptors

AGEs bind to several proteins, including lactoferrin, macrophage scavenger receptors (MSRs), oligosaccharyl transferase-48 (AGE-R1), 80 K-H phosphoprotein (AGE-R2), galectin-3 (AGE-R3), and the receptor for AGEs (RAGE). Some receptors, such as AGE-R1, AGE-R2, AGE-R3, and MSR, do not transduce

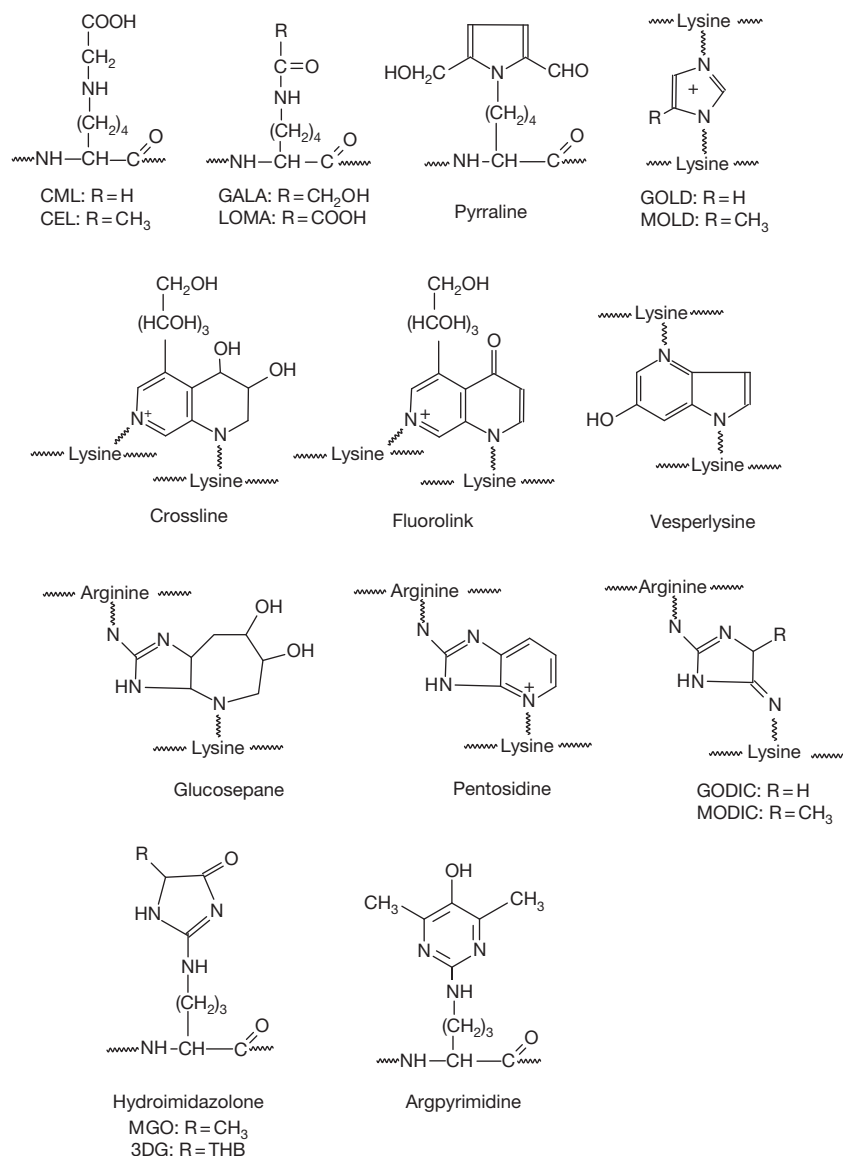


Figure 2 Structure of AGEs that have been detected in tissue proteins by chemical and/or immunological methods: (a) nonfluorescent lysine adducts and cross-links, including amides, amines, and imidazolium salts; (b) fluorescent lysine-lysine cross-links; (c) fluorescent arginine-lysine cross-links; (d) arginine derivatives. CML, N^ε-(carboxymethyl)lysine; CEL, N^ε-(carboxyethyl)lysine; GALA, glycolic acid lysine monoamide; LOMA, lysine oxalic acid monoamide; GOLD and MOLD, GO and MGO lysine dimer (imidazolium salts); and GODIC and MODIC, GO and MGO dihydroimidazolylidene cross-links.

cellular signals after ligation; instead, they cause clearance and detoxification of AGEs. In contrast, binding of AGEs to RAGE initiates proinflammatory signaling cascades. The best-characterized AGE receptor, RAGE, is a member of the immunoglobulin superfamily of cell-surface molecules. It is highly expressed in lung tissue from which it was originally isolated; however, it is also found at lower levels in a wide range of differentiated adult cells. The extracellular domain of RAGE consists of one V-type domain and two C-type domains; the latter form the ligand binding site. The intracellular, cytosolic domain mediates signal transduction. Ligation of RAGE by AGEs results in activation of cell-signaling events, involving mitogen-activated protein kinase (MAPK) and nuclear factor

kappa B (NF-κB). Various members of the MAPK family such as p44/p42 (ERK) MAPK, JNK MAPK, and p38 MAPK induce transcription of proinflammatory target genes. AGE-RAGE interaction also triggers generation of reactive oxygen species, mediated by activation of NF-κB.

RAGE is a cationic, pattern-recognition receptor, probably recognizing both the hydrophobic and anionic character of AGE proteins. Other ligands, such as high-mobility group box-1 protein (HMGB1; known as amphoterin), amyloid-β-protein, and S100/calgranulins also bind to RAGE. Under normal conditions, oxidative stress and protein turnover induced by ligation of RAGE may have a role in the remodeling and rejuvenation of tissues; however, this process may become

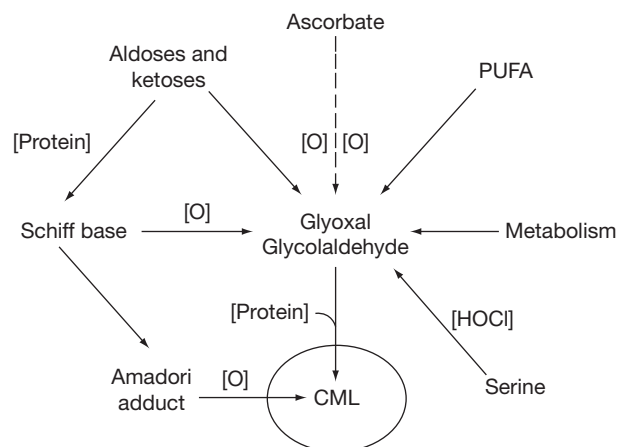


Figure 3 Alternative sources of the AGEs/ALEs. CML may be formed from reducing sugars, either directly by oxidative cleavage of Schiff base or Amadori adducts on protein, or indirectly through glyoxal or glycolaldehyde formed on oxidation of glucose or glucose adducts to protein. CML may also be formed from ascorbate, through glyoxal, glycolaldehyde, or tetroses formed on oxidative degradation of dehydroascorbate. Glyoxal or glycolaldehyde are also formed during peroxidation of polyunsaturated fatty acids, or during myeloperoxidase-mediated degradation of amino acids.

pathogenic in chronic disease. RAGE is upregulated when its ligands accumulate in tissue and its expression is increased in pathological conditions, such as diabetes, vascular and neurodegenerative disease, consistent with a role for overstimulation of RAGE in the pathogenesis of chronic inflammatory diseases.

Both protein and small molecule RAGE antagonists are being evaluated as therapeutic agents for limiting inflammatory damage in chronic disease. Small molecule inhibitors include compounds that block AGE–RAGE interaction, down-regulate RAGE expression, or block RAGE-mediated signal transduction in model systems. Binding of AGEs to RAGE, both *in vivo* and *in vitro*, can also be inhibited by an excess of soluble RAGE (sRAGE), the extracellular domain of RAGE. Because it lacks the cytosolic and transmembrane domains of RAGE, sRAGE traps ligands in the intracellular space, inhibiting cell binding and signal transduction. None of these agents has completed evaluation in clinical trials.

AGE Inhibitors and Breakers

AGE inhibitors do not affect glycation of proteins, but inhibit formation of AGEs. These compounds, such as aminoguanidine and pyridoxamine, generally have a reactive amino or thiol group that traps reactive carbonyl intermediates in AGE formation. AGE breakers are more complex nucleophilic compounds that are designed to cleave pre-existing AGEs in proteins. There is some controversy about their mechanism of action since AGE breakers also have AGE inhibitor and chelating (antioxidant) activity. However, the beneficial effects of AGE inhibitors and breakers are perhaps the most convincing evidence for the role of AGEs in disease – these compounds protect against the full range of renal, retinal, neural, and

vascular complications in animal models of diabetes. AGE breakers also reverse vascular stiffness in elderly persons with cardiovascular disease, independent of diabetes. Other activities of AGE breakers and inhibitors may contribute to their efficacy *in vivo*, since some of them inhibit amine oxidases, including some isoforms of nitric oxide synthase, have chelating activity, and also have antihypertensive and hypolipidemic effects *in vivo*. Some AGE inhibitors do not have reactive functional groups that would trap carbonyl intermediates, suggesting that chelation may be their primary mechanism of action. Indeed, the chelators trientine (triethylenetetramine) and citrate have been shown to protect against diabetic complications in animal models.

The Maillard Reaction in Aging and Disease

Aging

The age-dependent browning of tissue proteins is most apparent in the lens, which becomes visibly yellow with age. This change in color is accompanied by insolubilization, aggregation, and precipitation of lens crystallins, which may proceed to the formation of cataracts. Collagens and other long-lived proteins in the body undergo similar changes with age. The increased cross-linking of collagen contributes to the decrease in elasticity and resiliency of the extracellular matrix with age. By interfering with turnover of matrix components by metalloproteinases, browning reactions may also contribute to the age-dependent thickening of microvascular basement membranes, for example, in the retina and kidney, affecting the permeability and transport properties of the extracellular matrix. The accumulation of AGEs in long-lived proteins (Figure 4) is accompanied by an age-dependent increase in protein fluorescence and by a parallel increase in other chemical modifications of proteins, including products of both nonoxidative (e.g., aspartate racemization and spontaneous deamidation) and oxidative damage (*o*-tyrosine, nitrotyrosine, dityrosine, and methionine sulfoxide). Research on the Maillard hypothesis today is focused on increasing evidence that AGEs have effector functions, that is, they activate or participate in physiological responses to stress, including inflammation, hyperglycemia, hyperlipidemia, and oxidative stress. Although proteins serve as the register or accumulator of Maillard reaction damage *in vivo*, glycoxidative and oxidative modifications of DNA may have significant effects on the integrity of the genome during aging.

Diabetes

The glycation, AGE, or Maillard hypothesis identifies glucose as the culprit in diabetes, and proposes that diabetic complications result from accelerated Maillard reaction damage to tissue protein. Consistent with the hypothesis, both the extent of glycation of protein and the rate of accumulation of AGEs are accelerated by hyperglycemia in diabetes (Figure 4). The damage, such as hyperglycemia, is systemic, and age-adjusted levels of AGEs in skin collagen are correlated with the severity of renal, retinal, and vascular complications of diabetes. The Maillard hypothesis is a chemical rather than metabolic theory on the origin of diabetic complications; it explains the development

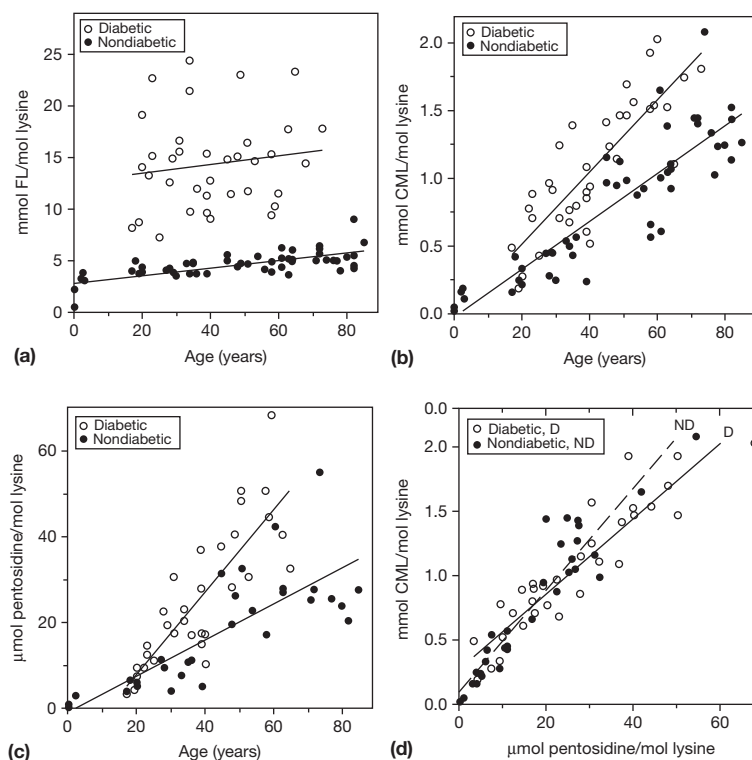


Figure 4 Influence of age and diabetes on glycation and glycoxidation of skin collagen: (a) glycation of protein is reversible and relatively constant with age. Glycation of collagen increases with mean blood glucose concentration and correlates with GlcHb in diabetic patients; (b) and (c) the concentrations of the AGE/ALE, CML, and of the AGE, pentosidine, increase with age in the nondiabetic population and at an increased rate in diabetic patients; (d) despite differences in origin and mechanism of formation and significant differences in their concentrations, the levels of CML and pentosidine in skin collagen correlate more closely with one another than with age in either control or diabetic populations. Reproduced from Dyer DG, Dunn JA, Thorpe SR, et al. (1993) Accumulation of Maillard reaction products in skin collagen in diabetes and aging. *Journal of Clinical Investigation* 91: 2463–2469, with permission.

of similar complications in both type 1 (juvenile-onset, insulin-dependent) and type 2 (adult-onset, noninsulin-dependent) diabetes, despite the differences in etiology of these diseases (insulin deficiency vs. insulin resistance). It also explains the development of complications in the kidney, nerve, retina, and vasculature, since intracellular glucose concentration in the endothelial and parenchymal cells in these tissues tracks extracellular glucose concentration. These tissues are also rich in long-lived extracellular proteins, such as collagens, elastin, and myelin. In addition to the gradual accumulation of AGEs on extracellular proteins, AGEs are also formed on intracellular proteins in these tissues, probably from glycolytic intermediates, such as the triose phosphates. Research on intracellular glycation and AGE formation is still at an early stage, but is essential for a comprehensive understanding of the role of the Maillard reaction in the pathogenesis of diabetic complications.

Other Diseases

AGEs and ALEs are increased in tissues in a wide range of age-related, chronic diseases, including diabetes, atherosclerosis, dialysis-related amyloidosis, arthritis, and neurodegenerative diseases. The increase in glycoxidative and lipid-derived (lipoxidative) modifications reflects, in part, uncontrolled chemistry occurring, either locally or systemically, in biological systems as a result of increased substrate (carbohydrate or lipid),

decreased antioxidant defenses, and/or increased decompartmentalization of metal ions. Although the AGEs/ALEs are not the primary source of pathology in these diseases, they are macrophage chemoattractants and proinflammatory molecules, acting as both intermediaries in pathogenic processes and biomarkers of resultant tissue damage. AGE/ALE inhibitors may eventually prove useful for limiting damage or complications from a wide range of chronic, age-related diseases. Their effects on health and longevity are being explored in studies in animal models.

See also: **Bioenergetics:** Reactive Oxygen and Nitrogen Species: Interactions with Mitochondria and Pathophysiology; **Metabolism** **Vitamins and Hormones:** Adipocyte Fat Mobilization: The Regulatory Roles of Perilipin 1, Adipose Triglyceride Lipase, and Hormone-Sensitive Lipase; **Protein/Enzyme Structure Function and Degradation:** Collagens.

Further Reading

Calcutt NA, Cooper ME, Kern TS, and Schmidt AM (2009) Therapies for hyperglycaemia-induced diabetic complications: From animal models to clinical trials. *Nature Reviews Drug Discovery* 8: 417–429.
 Fayle SE and Gerrard JA (2002) *The Maillard Reaction*. London: The Royal Society of Chemistry.

- Negre-Salvayre A, Salvayre R, Auge N, Pamplona R, and Portero-Odin M (2009) Hyperglycemia and glycation in diabetic complications. *Antioxidants and Redox Signaling* 11: 3071–3109.
- Nursten H (2005) *The Maillard Reaction: Chemistry, Biochemistry and Implications*. London: The Royal Society of Chemistry.
- Schleicher E, Somoza V, and Schieberle P (2008) *Annals of the New York Academy of Sciences. Vol. 1126: The Maillard Reaction: Recent Advances in Food and Biomedical Sciences*. New York, NY: Wiley-Blackwell.
- Tessier F (2010) The Maillard reaction in the human body. The main discoveries and factors that affect glycation. *Pathologie Biologie (Paris)* 58: 214–219.
- Thomas MC and Forbes J (eds.) (2010) *The Maillard Reaction: Interface between Aging, Nutrition and Metabolism*. London: The Royal Society of Chemistry.
- Thornalley PJ (2008) Protein and nucleotide damage by glyoxal and methylglyoxal in physiological systems – role in aging and disease. *Drug Metabolism and Drug Interactions* 23: 125–150.

Relevant Website

<http://www.imars.org> – International Maillard Reaction Society (IMARS)

Glycerolipid Structure, Function, and Synthesis in Eukaryotes

D R Voelker, National Jewish Health, Denver, CO, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Acyltransferases Enzymes that transfer fatty acids from thioester linkage in acyl-coenzyme A to O-ester linkage in molecules such as glycerol phosphate.

Anionic lipids Lipid molecules with a net negative charge.

Ether lipids Lipid molecules characterized by ether rather than ester linkages between glycerol and hydrophobic substituents.

Glycerolipids Lipid molecules consisting of a glycerol backbone structure to which are coupled in ester, or ether, or glycosidic linkages, fatty acids, fatty alcohols, saccharides, and phosphate esters.

Neutral uncharged lipids Lipids without substituents having a formal charge.

Zwitterionic lipids Lipid molecules with multiple functional groups having both negative and positive charges.

Introduction

Lipids, along with proteins, carbohydrates, and nucleic acids, constitute one of the major groups of biochemicals found in all eukaryotic cells. Among the lipids, glycerolipids are the most abundant family and these molecules have numerous functions within cells that range from energy storage, to structural compartmentation of cells and organelles, to key signaling activities that are essential for regulating cell growth and development. This article focuses upon the structure, functions, and biosynthesis of the glycerolipids, and identifies the interfaces of this information with other articles in the encyclopedia.

Glycerolipid Structure

The structures of all eukaryotic glycerolipids are fundamentally based upon the structure of *sn*-glycerol-3-phosphate, which is shown in [Figure 1](#) and illustrates the stereospecific numbering (*sn*) of each carbon. The *sn*-glycerol-3-phosphate molecule is asymmetric and optically active; when drawn in a Fisher projection, the chiral *sn*-2 position has the OH group on the left side of the molecule. The absolute stereoconfiguration assigned using the Cahn, Ingold, and Prelog method is 2-*R*-glycerol-3-phosphate. With few exceptions, in animal eukaryotic cells that are not specialized in fat storage, or predisposed to fat storage due to developmental programming, or pathological abnormalities, the major glycerolipids are phosphate containing and typically referred to as 'glycerophospholipids', or more simply as 'phospholipids'. In cells specialized in fat storage, such as mammalian adipocytes, the major glycerolipids are triacylglycerols (TAGs).

The general structural features of glycerophospholipids are outlined in [Figures 1](#) and [2](#). Esterification of fatty acids to the *sn*-1 and *sn*-2 carbons of *sn*-glycerol-3-phosphate produces the lipid phosphatidic acid (PtdOH). Phosphatidic acid is a minor phospholipid constituent of many eukaryotic cells and is used as a metabolic intermediate in the synthesis of additional glycerolipids. Phosphatidic acid is a phosphomonoester that can condense with a variety of alcohols to produce a phosphodiester linkage commonly found in the more complex phospholipids

(see [Figure 2](#)). The main alcohols that give rise to these structures are choline, ethanolamine, serine, inositol, and glycerol. The mature phospholipids that arise from these condensations are phosphatidylcholine (PtdCho), phosphatidylethanolamine (PtdEtn), phosphatidylserine (PtdSer), phosphatidylinositol (PtdIns), and phosphatidylglycerol (PtdGro). The terminal hydroxyl group of PtdGro can also form a phosphodiester with another molecule of PtdOH to produce a Ptd₂Gro, a lipid with the common name of cardiolipin. Cardiolipin is also referred to as diphosphatidylglycerol; however, a more accurate name is *bis*-phosphatidyl-glycerol, although it is not commonly used.

Additional forms of glycerolipids, which lack phosphate, are also found throughout eukaryotic organisms. In TAG, the phosphate substituent of PtdOH is replaced by a fatty acid esterified to the *sn*-3 position of glycerol (see [Figure 3](#)). Variations upon this latter structural theme also occur in which a glycerol contains only two fatty acids (diacylglycerols (DAGs)) or one fatty acid (monoacylglycerol). One major grouping of glycerolipids (glycosyl-glycerolipids) is highly abundant in eukaryotic plants and consists of a DAG moiety conjugated to a galactose monomer, a galactose dimer, or a sulfo-galactose monomer.

The structures of glycerolipids confer important physical properties upon the molecules. The fatty acid substituents contain relatively long hydrocarbon chains (typically 16–22 carbons, with 0–6 double bonds), which are hydrophobic and sequestered away from aqueous environments. The phosphate and polar alcohol, or sugar constituents are hydrophilic and readily interact with aqueous solvents. This dual or amphipathic property of glycerolipids causes them to form large, non-covalent complexes, such as micelles and bilayer sheets, when the lipids are introduced into aqueous environments. Molecules such as TAG have no hydrophilic domains and these lipids efficiently form phases that completely segregate away from aqueous environments, with minimal interaction. The fatty acids esterified to the glycerol backbone of phospholipids can dramatically affect the physical properties of these molecules in membranes. Membranes harboring lipids containing fatty acids with few double bonds tend to be more rigid, whereas those with fatty acids containing large numbers of double bonds tend to be more fluid. These bulk properties can significantly alter

the interactions of membrane proteins with each other, and with substrates or ligands the proteins recognize.

A number of variations occur to the basic structures of the common phospholipids described above and are important elements in the biology and function of the molecules. One unusual structural variant contains long-chain fatty alcohols in place of the long-chain fatty acids that occur in the hydrophobic portions of lipid molecules (see Figure 4). In these molecules, the ester linkage that usually couples fatty acids to the glycerol is replaced by an ether linkage. From a simple chemical perspective, this ether linkage imparts much greater chemical stability to the chemical bond. Within the class of ether lipids, a subset of the molecules contains a double bond adjacent to the ether linkage, and lipids with this characteristic are generally named 'plasmalogens', for historical reasons. In another type of structural variation, hydroxyl groups present in PtdIns are available for further modification. One of the most important modifications to PtdIns is the attachment of phosphate groups to the 3, 4, and 5 positions of the inositol ring. Although these phosphorylated forms of PtdIns are present at very low levels within cells, they play crucial roles in

regulating membrane recognition events and controlling a variety of physiological processes within cells. The hydroxyl groups of PtdIns can also be modified by the attachment of fatty acids, and these modifications also play an important role in regulating the behavior of specific membrane proteins.

Another structural variant of the phospholipids shown in Figure 2 is built upon the same *sn*-glycerol-3-phosphate backbone, but contains only one fatty acid esterified to either the *sn*-1 or *sn*-2 hydroxyl group. This modification occurs in conjunction with any of the aforementioned polar alcohols or saccharides described earlier. Broadly, these lipids are referred to as lyso-phospholipids. As an example, a glycerolipid with a fatty acid esterified to the *sn*-1 position, an unesterified hydroxyl at the *sn*-2 position, and a choline moiety esterified to the *sn*-3 phosphate is named 'lyso-phosphatidylcholine' (see Figure 5). The lyso moniker historically derives from fractions of phosphatidylcholine that were isolated by column chromatography and shown to be capable of lysing red blood cells. Many lyso-phospholipids have significant detergent activity that renders them lytic to cell membranes when the lipids are present in sufficiently high concentrations. The generic term 'lyso-phospholipid' does not specify which hydroxyl on the glycerol phosphate backbone is esterified, and forms with acyl groups esterified to either the *sn*-1 or *sn*-2 hydroxyls are commonplace. Lyso-PtdOH, lyso-PtdCho, lyso-PtdEtn, lyso-PtdSer, lyso-PtdIns, and lyso-PtdGro and lyso-Ptd₂Gro are all present in biological systems, usually at low levels. Lyso-Ptd₂Gro can occur as a variety of structural forms, with either one (monolyso-) or two (dilyso-) unesterified hydroxyls. Another variation upon the lyso-phospholipid structures occurs in which two monoacylglycerols are coupled in a single phosphodiester linkage. The resultant molecule is a *bis*-monoacylglycerol-phosphate. Many lyso-phospholipids are generated by the actions of phospholipase A enzymes, which hydrolyze the ester bonds of fatty acids esterified to the *sn*-1 or *sn*-2 positions of glycerol phosphate backbones of

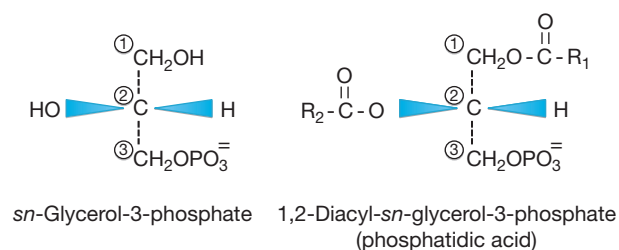


Figure 1 Structures of *sn*-glycerol-3-phosphate and phosphatidic acid. *sn*-glycerol-3-phosphate forms the structural scaffold for the glycerolipids and phosphatidic acid is the major root structure for all of the glycerolipids. The *sn*-1, *sn*-2, and *sn*-3 positions are designated within the circles.

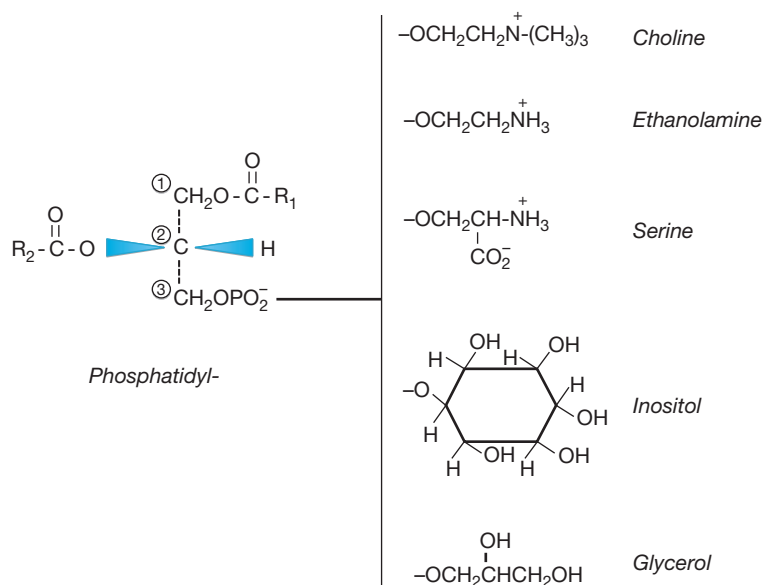


Figure 2 Glycerophospholipids share common structural features in their phosphatidic acid moiety. The coupling of the phosphatidic acid structure with the alcohols, choline, ethanolamine, serine, inositol, and glycerol defines the majority of glycerophospholipids in eukaryotic cells.

phospholipids. Lysophospholipids can occur either as biosynthetic intermediates or as degradation products in biochemical pathways.

Glycerolipid Functions

A complete list of specific functions of glycerolipids would greatly exceed the space available for this article. A few examples will be provided for each of the major and minor glycerolipids. One of the most far-reaching aspects of phospholipid function is their intrinsic ability to form membrane bilayer structures in aqueous environments. This membrane-forming action is essential for cells to define the boundaries that separate them from their environment and demarcate intracellular compartments (organelles) for specialized intracellular function.

PtdOH is a minor anionic phospholipid present in cell membranes that plays crucial roles in regulating the activity of enzymes involved in controlling cell growth. Many of these proteins play crucial roles in promoting the growth of cancer cells. PtdOH has also been identified as a lipid that sequesters regulatory proteins necessary for turning specific genes on and off. Increases or decreases of the PtdOH levels within the cell can have profound effects upon the activation of specific genes.

PtdCho is the most abundant phospholipid in cell membranes of many eukaryotes ranging from yeast to mammals. PtdCho is zwitterionic and typically functions as the major structural lipid of most membranes. A specialized molecular species of PtdCho containing palmitic acid esterified to both the *sn*-1 and *sn*-2 positions of the molecule (dipalmitoyl-PtdCho) is present in pulmonary surfactant, and forms a film at the air tissue interface of the lung, in the alveolar region where oxygen absorption occurs, and greatly reduces the work

of breathing. The lack of dipalmitoyl-PtdCho can cause respiratory failure in premature newborns.

PtdEtn typically constitutes the second most abundant phospholipid in eukaryotic cell membranes. This zwitterionic lipid plays an important role in cell division and is associated with the process whereby the membrane of a parent cell pinches off to form two daughter cells. In the mammalian eye, PtdEtn becomes modified during the visual cycle, and abnormal processing of this modified lipid can lead to macular degeneration.

PtdSer is a minor cellular lipid with anionic character. Although its total cellular concentration is not high, specific membrane compartments, such as the inside surface of the plasma membrane, are highly enriched in this lipid. PtdSer plays an important role in the early steps of blood clotting. Platelets within the blood expose PtdSer on the outside of their surface membrane to initiate the blood clotting process. In addition, the appearance of elevated levels of PtdSer on the outside of the surface membrane of numerous types of cells functions as a biochemical signal that attracts immune cells to the site for the purpose of engulfing and destroying these cells.

PtdIns is also a minor cellular lipid, which is anionic. This lipid can bind several regulatory proteins and influence the synthesis of PtdCho. PtdIns serves as an important precursor for the formation of phosphorylated derivatives known as 'polyphosphoinositides', especially PtdIns-3-P, PtdIns-4-P, PtdIns-4,5-P₂, and PtdIns-3,4,5-P₃. The different polyphosphoinositides play important roles in regulating intracellular membrane traffic, by binding specific proteins that regulate dynamic aspects of membrane structure, function, and interaction.

PtdGro is also a minor anionic, cellular lipid, which, in most cells, functions primarily as a precursor for the synthesis of cardiolipin. In the lung, PtdGro is an important component of pulmonary surfactant, where it functions as an inhibitor of inflammatory processes elicited by a variety of pathogens.

Ptd₂Gro is a lipid that is restricted to mitochondria where it functions as a cofactor in the association of proteins in the respiratory chain to form large macromolecular complexes. A specific molecular species of cardiolipin containing four linoleic acids esterified to the backbone *sn*-glycerolphosphates is markedly reduced in patients with the X-linked disease known as Barth syndrome. The mitochondria from individuals afflicted with this disorder exhibit abnormal structure and function.

TAG is the major storage form of fat. Within adipocytes, TAGs essentially phase-separate from the cytosol and form oil droplets. The limiting membrane of lipid droplets plays a crucial role in orchestrating the formation of the droplet and regulating the metabolism of the stored TAGs.

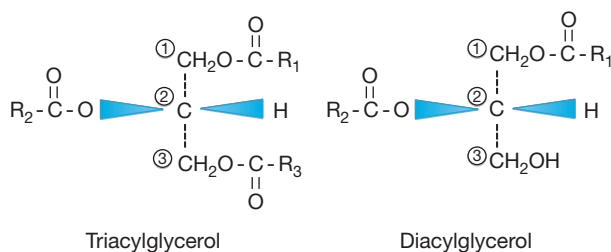


Figure 3 Glycerolipids without phosphate are of major importance. Two examples of glycerolipids without phosphate are shown. Triacylglycerol is the major storage form of fat and diacylglycerol is a crucial intermediate in glycerolipid synthesis and cell signaling pathways.

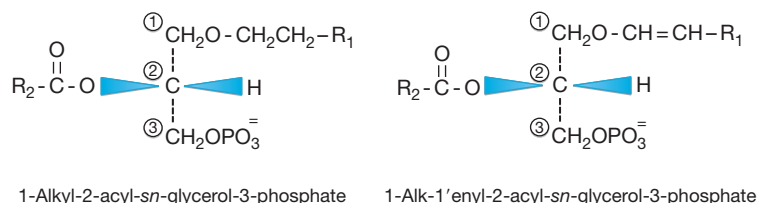


Figure 4 Structural features of ether linked glycerolipids. Two examples of ether-linked variants of phosphatidic acid are shown. The alkyl lipids have an ether bond with the *sn*-1 carbon of the *sn*-glycerol-3-phosphate scaffold. The alk-1'-enyl lipids have an ether bond with the *sn*-1 carbon of the scaffold and also contain a double bond adjacent to the ether linkage.

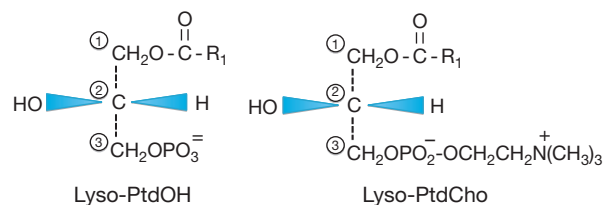


Figure 5 Structural features of lyso-phospholipids. Lyso-phospholipids lack one of the two acyl chains typical of glycerophospholipids, and lyso-PtdOH and lyso-PtdCho are shown for structural comparison.

DAGs are usually present in only trace quantities within cells. Much of the DAG pool within cells serves as a precursor for the synthesis of lipids such as PtdCho, PtdEtn, and TAG in mammalian cells. However, there is also a highly dynamic and rapidly metabolized pool of DAG associated with signaling events at the plasma membrane that occur in conjunction with hormone or growth-factor activation of cell surface receptors. This localized DAG pool plays an important role in activating protein kinases that regulate cell growth.

Bismonoacylglycerolphosphate (BMP) is an unusual lipid that is highly enriched in endosomes, multivesicular bodies, and lysosomes. The biophysical properties of this lipid indicate that it plays an important role in membrane fusion events within the endosomal system.

Glycerolipid Synthesis

Bioenergetic Considerations

The stepwise assembly of glycerolipids from precursor molecules, like all biosynthetic reactions, requires chemical energy. Ultimately, this energy source can be traced back to adenosine triphosphate (ATP), but this article focuses upon the energetics of the reactions proximal to the formation of end products. The formation of esters between fatty acids and the hydroxyl groups of glycerol utilizes the chemical potential energy stored in the thioester bond of acyl-coenzyme A (CoA) derivatives. The free energy of hydrolysis of the thioester bond between fatty acids and CoA is nearly equivalent to that of ATP.

The formation of phosphoester linkages at the *sn*-3 position of glycerol comes from ATP for the formation of PtdOH. The formation of PtdCho and PtdEtn utilizes the energy of pyrophosphate cleavage of the cytidine 5'-diphosphate (CDP) conjugated alcohols CDP-choline and CDP-ethanolamine to form the *sn*-3 phosphodiester linkages. The synthesis of anionic lipids, PtdSer, PtdIns, PtdGro, and Ptd₂Gro, utilizes the energy of pyrophosphate cleavage of CDP-diacylglycerol to form the *sn*-3 phosphodiester linkages. Some eukaryotes execute hydrophilic head group exchanges for the synthesis of PtdSer, PtdEtn, and PtdCho. The polyphosphoinositides use ATP for the direct formation of phosphomonoesters at positions 3, 4, and 5 of PtdIns.

Phosphatidic Acid Synthesis

The initiation of glycerolipid synthesis begins at a branch-point from glycolysis and utilizes the common intermediate, dihydroxyacetone phosphate (DHAP), which is derived from the aldolase cleavage of fructose-1,6-bis-phosphate. The

major synthetic reactions for the glycerolipids are summarized in Figure 6. DHAP is reduced to *sn*-glycerol-3-phosphate by glycerol-3-phosphate dehydrogenase. The first fatty acid is attached to the glycerol backbone by *sn*-glycerol-3-phosphate acyltransferase (GPAT), using an acyl-CoA substrate and yielding the product lyso-PtdOH. Some forms of this enzyme in mammalian systems show a preference for saturated fatty acids and they are located in different organelles including endoplasmic reticulum, mitochondria, and peroxisomes.

An alternative route to the formation of lyso-PtdOH commences with acylation of the available hydroxyl of DHAP, also using an acyltransferase enzyme and an acyl-CoA substrate, yielding the product acyl-DHAP. This acyl-DHAP can next be reduced by nicotinamide adenine dinucleotide phosphate (NADPH)-dependent oxidoreductase to produce lyso-PtdOH. The acylation of DHAP and its subsequent reduction occur in the peroxisome. A second fatty acid is esterified to lyso-PtdOH by the action of another family of acyltransferases, using an acyl-CoA substrate. PtdOH is a biosynthetic intermediate at a crucial juncture of glycerolipid metabolism. One branch of the pathway gives rise to PtdCho, PtdEtn (together comprising the neutral zwitterionic glycerolipids), TAG, and glycosyl-DAG (together comprising the neutral uncharged glycerolipids) and the other branch gives rise to the anionic phospholipids, PtdSer, PtdIns (and its phosphorylated derivatives), PtdGro, and Ptd₂Gro. There are important differences between lower eukaryotes (e.g., *Saccharomyces cerevisiae*) and mammals in the synthetic pathways for the anionic glycerophospholipids.

Anionic Glycerophospholipid Synthesis

The committed step for diverting PtdOH to the anionic glycerophospholipid biosynthetic pathway is the synthesis of cytidine 5'-diphosphate-diacylglycerol (CDP-DAG) (see Figure 6). Pyrophosphate cleavage of cytidine 5'-triphosphate (CTP) provides the chemical energy to couple the cytidine 5'-monophosphate (CMP) nucleotide to PtdOH, and the reaction is catalyzed by CDP-DAG synthase. The enzyme shows dual localization to both the endoplasmic reticulum and the mitochondria. In the endoplasmic reticulum, the CDP-DAG is a substrate for the synthesis of PtdIns catalyzed by PtdIns synthase. Upon cleavage of the pyrophosphate bond, the PtdOH moiety forms a phosphate ester linkage with the C1-position hydroxyl of inositol. The resultant PtdIns accounts for approximately 90% of the inositol lipids in mammalian cells. Further metabolism of PtdIns can proceed via the ATP-dependent enzymes, PtdIns-3kinase, PtdIns-4kinase, and PtdIns-5kinase reactions, yielding at least five different types of polyphosphoinositides. These reactions are highly regulated and additional biochemical details are given elsewhere in this encyclopedia.

The mitochondrial pool of CDP-DAG is used for the synthesis of PtdGro and Ptd₂Gro. The synthesis of PtdGro occurs in a two-step reaction at the inner mitochondrial membrane. In the first step, the PtdGro synthase enzyme utilizes CDP-DAG and *sn*-glycerol-3-phosphate as substrates for the synthesis of PtdGroP. In this reaction, the PtdOH moiety is esterified to the *sn*-1 position of *sn*-glycerol-3-phosphate. In the second step, the terminal phosphate of the PtdGroP is hydrolyzed to yield PtdGro. The steady-state level of PtdGro in mitochondria

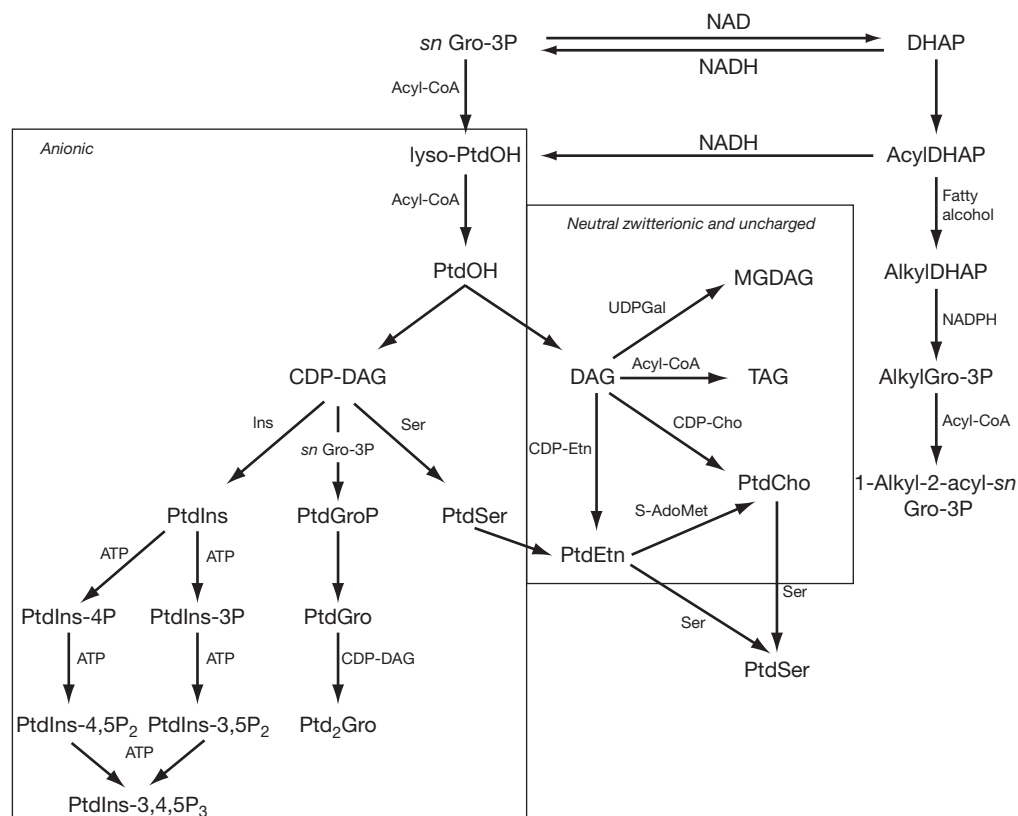


Figure 6 Major synthetic pathways for the synthesis of glycerophospholipids. DHAP derived from glycolysis generates the scaffold for both diacyl and ether lipid synthesis. PtdOH resides at a crucial branchpoint in lipid synthesis segregating anionic glycerolipids from neutral zwitterionic and uncharged lipid synthesis. CDP-DAG functions as the central precursor for anionic glycerophospholipids and DAG functions as the central precursor for zwitterionic glycerophospholipids and glycerolipids without charged groups. This schematic is a composite of the important reactions occurring in lower and higher eukaryotes and plants.

is low because most of this lipid is used for the synthesis of Ptd₂Gro via a reaction catalyzed by cardiolipin synthase, which utilizes PtdGro and CDP-DAG as substrates. In this latter reaction, the cleavage of the pyrophosphate of CDP-DAG is coupled to the esterification of the PtdOH group to the free sn -3 hydroxyl group present in PtdGro.

In many eukaryotes, but not mammalian systems, CDP-DAG is also used as a precursor for PtdSer. As in the other CDP-DAG reactions, the cleavage of the pyrophosphate linkage is accompanied by transfer of the PtdOH moiety to the alcohol acceptor, which is the primary hydroxyl group of serine. The reaction catalyzed by PtdSer synthase occurs in the endoplasmic reticulum. This reaction has been most completely studied in yeast. Mammalian cells use a different strategy for the synthesis of PtdSer in which the alcohol head groups of PtdCho and PtdEtn are exchanged for serine. The regulation of the mammalian pathways is not well understood, but may be coupled to localized calcium flux. Two enzymes have been identified, one of which uses both PtdCho and PtdEtn as sources of the PtdOH moiety and the second which preferentially uses PtdEtn.

The anionic glycerophospholipid family of eukaryotes contains 13 prominent members as shown in Figure 6. CDP-DAG serves as the essential biosynthetic precursor for the production of 10 different classes of anionic glycerophospholipids.

Despite this large number of different types of anionic glycerophospholipids, the entire family constitutes a relative minority of lipids within cells usually comprising less than 25% of total cellular glycerophospholipids. The physical characteristics of these lipids, their documented specific interactions with a variety of regulatory proteins, and their specialized localization within specific organelles all contribute to the biological importance of these lipids.

Neutral Zwitterionic and Uncharged Glycerolipid Synthesis

The second arm of the branchpoint occurring from the central PtdOH intermediate in animal eukaryotes gives rise to the neutral zwitterionic glycerolipids, PtdEtn and PtdCho, as well as the uncharged glycerolipid, TAG (see Figure 6). In plants, the dominant glycerolipids are monogalactosyl- and digalactosyl-DAG. This latter point merits particular emphasis, since digalactosyl-DAG is estimated to be the single most abundant lipid class present on the Earth. The committed metabolic step that shunts the PtdOH intermediate away from anionic glycerolipid synthesis is catalyzed by PtdOH phosphatase. The PtdOH phosphatase is a highly regulated enzyme because it not only directs biosynthesis in the direction of the major structural lipids of all eukaryotic cell membranes (PtdEtn and PtdCho) but also provides the necessary precursor for TAGs,

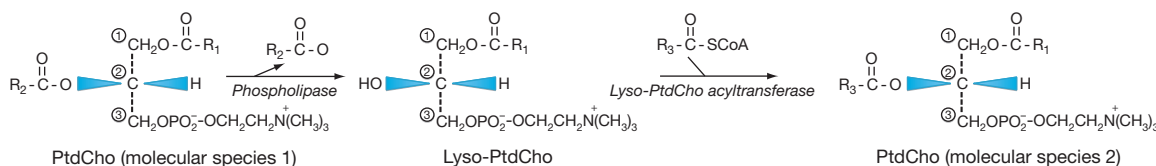


Figure 7 Remodeling reactions adjust the fatty acids attached to phospholipids. The fatty acid content of phospholipids is not static. A combination of phospholipase removal of a fatty acid (R_2 in this example) and lysophospholipid acyltransferase-catalyzed replacement with a different fatty acid (R_3 in this example) functions to adjust phospholipid structure to meet specific needs of different organelles, and the changing environmental and developmental demands.

which are the major and most efficient storage form of calories in nearly all eukaryotes. The product of the PtdOH phosphatase reaction is DAG. Similar to the role of CDP-DAG as a central precursor of anionic glycerolipid synthesis, DAG functions as a central intermediate supplying multiple crucial pathways for the nonanionic glycerolipids.

In the biosynthetic reactions utilizing DAG, the chemical energy driving the reactions is derived from a variety of substrates that participate in the transfer of groups to the primary alcohol of this lipid. For PtdEtn synthesis, the substrate is CDP-Etn. The enzyme ethanolamine phosphotransferase catalyzes the cleavage of the pyrophosphate bond in CDP-Etn and transfers the P-Etn moiety to the *sn*-3 hydroxyl position of DAG. This reaction occurs principally in the endoplasmic reticulum. An alternative route for the synthesis of PtdEtn utilizes the anionic phospholipid, PtdSer, as a substrate for the enzyme PtdSer decarboxylase. Both pathways for the synthesis of PtdEtn are active across the phylogenetic spectrum of eukaryotes. For many organisms, the PtdSer decarboxylase enzyme is localized to the mitochondria and plays a major role in providing PtdEtn for this organelle.

The synthesis of PtdCho from DAG is analogous to that for PtdEtn and utilizes the substrate CDP-Cho. The choline phosphotransferase also uses pyrophosphate cleavage to transfer the P-Cho from the water-soluble substrate, to the *sn*-3 hydroxyl of the DAG lipid. In general, this reaction is limited by the availability of CDP-Cho. The choline phosphotransferase is principally localized to the endoplasmic reticulum. Many eukaryotes are also capable of synthesizing PtdCho by methylating the primary amine of PtdEtn catalyzed by PtdEtn methyltransferases. In yeast, these methyltransferase reactions are highly active and can produce all of the PtdCho required for cell growth. The methyl donor in these reactions is S-adenosyl-methionine (SAdoMet). In mammalian systems, the PtdEtn methyltransferase is most active in liver tissue, and either marginally active or undetectable in other tissues.

TAG is also derived from DAG and this constitutes the major pathway for storing calories in the form of fat. The enzyme DAG acyltransferase utilizes an acyl-CoA substrate, and cleavage of the thioester bond of the substrate is coupled to esterification of the fatty acid to the *sn*-3 hydroxyl of the glycerolipid. Two DAG acyltransferases are the major contributors to TAG synthesis in mammalian systems in most tissues. One of these enzymes, DGAT1, is more promiscuous in its use of acceptors and can utilize monoacylglycerols and retinols in addition to DAG as substrates for transfer of acyl groups. The DAG acyltransferases localize principally to the endoplasmic reticulum, but may also be found in regions of the organelle that give rise to lipid droplets. DAG acyltransferase in

plants also gives rise to TAG, which is of significant economic importance in the generation of seed oils that have major importance in nutritional, industrial, and biofuel applications.

The DAG intermediate in plants also functions as the major metabolic precursor of the glycosyl-DAGs. The conversion of DAG to galactosyl-DAG is catalyzed by the corresponding galactosyl transferase. This reaction uses uridine 5'-diphosphate (UDP)-galactose as the water-soluble substrate, and the cleavage of the UDP moiety is coupled to the transfer of the galactose to the *sn*-3 hydroxyl of the DAG. Typically, the one carbon of galactose is coupled in β linkage to *sn*-3 hydroxyl of the glycerol backbone. Synthesis of higher-order galactolipids, such as digalactosyl-DAG, occurs through a similar mechanism of transfer of galactose from the nucleotide sugar to the galactosyl-DAG, with the terminal galactose in α -1,6 linkage to the first galactose. Although plants also synthesize the major phospholipids found in eukaryotes, the bulk of their membrane lipids consists of the galactolipids, a situation which probably evolved as a mechanism for plants to survive in phosphate-limited environments.

Ether Lipid Synthesis

As described earlier in this article, the lipid repertoire of some organisms, including mammals, contains ether-linked glycerolipids. These lipids are synthesized in a manner similar to that of the more abundant ester-linked glycerolipids, but with several unique steps. The early steps in ether lipid synthesis are confined to the DHAP backbone rather than the *sn*-glycerol-3-phosphate backbone. Acyl-DHAP is the required substrate for the synthesis of ether lipids, and the enzyme alkyl-DHAP synthase catalyzes an exchange of a long-chain fatty alcohol for the fatty acid esterified at the *sn*-1 position (see Figure 6). The resultant alkyl-DHAP product is reduced by an oxidoreductase that uses NADPH as a coenzyme. The next step in the pathway introduces a fatty acid into the *sn*-2 position by the action of an acyl transferase. The final product from the above steps is an analog of PtdOH that differs from the di-ester form, by the presence of an ether linkage connecting the hydrophobic aliphatic chain at the *sn*-1 position of the molecule (see Figure 4). This ether version of PtdOH is further processed for the synthesis of major phospholipids by many of the same pathways used for the di-ester version of the molecule. One additional reaction unique to ether lipid synthesis is the introduction of a double bond in the hydrocarbon chain adjacent to the ether linkage at the *sn*-1 position. The substrate for this reaction is 1-alkyl-2-acyl-*sn*-glycerol-3-phosphoethanolamine, and a desaturase enzyme introduces a double bond between the first and second carbons of

the alkyl chain using electrons derived from NADPH. The product is 1-alk-1'-enyl-2-acyl-*sn*-glycerol-3-phosphoethanolamine, and this molecule is often referred to as an ethanolamine plasmalogen for historical reasons. Further metabolism of this unsaturated ether lipid can give rise to the PtdCho version of the molecule, which is also generically referred to as a choline plasmalogen. Choline plasmalogens constitute a significant portion of the total phospholipids present in heart tissue.

Phospholipid Remodeling

The lipid synthetic reactions outlined in [Figure 6](#) are largely focused upon the pathways that alter the substituents attached to the *sn*-3 position of the glycerol backbone, and give rise to the major classes of the glycerolipids. From the scheme, it might appear that the acyl groups found esterified to the *sn*-1 and *sn*-2 positions at the level PtdOH might be the same found for all of the other lipid classes. However, this is not the case, since the fatty acids found in the *sn*-1 and *sn*-2 positions of PtdOH are not identical to those found in PtdCho, PtdEtn, PtdSer, PtdIns, and the polyphosphoinositides, PtdGro, Ptd₂Gro, TAG, and MGDAG. A similar situation applies to the *sn*-2 fatty acids in ether lipids. One major contributor to these differences is the process of lipid remodeling (see [Figure 7](#)). In its simplest form, remodeling is a process in which a molecule such as PtdCho undergoes a cycle of removal of one type of acyl group (perhaps a mono-unsaturated fatty acid) from the *sn*-2 position by a phospholipase A₂ to form a lyso-PtdCho, followed by attachment of a different acyl group (perhaps a polyunsaturated fatty acid) at the *sn*-2 position catalyzed by a lyso-PtdCho acyltransferase. Remodeling of phospholipids is one mechanism of ensuring that specific organelles contain the precise types (molecular species) of phospholipids to execute unique functions. The remodeling process is not restricted to the *sn*-2 position of phospholipids, but can also occur at the *sn*-1 position. Furthermore, the remodeling activities occurring in one cellular membrane (e.g., mitochondria) are not the same as remodeling reactions occurring at different membranes (e.g., Golgi). Current information suggests that remodeling occurs for all different types of phospholipids and that specific classes of phospholipids, such as PtdCho, can be remodeled by multiple acyltransferases with different selectivity for specific fatty acids.

Acknowledgments

Support for this work was provided by the National Institutes of Health Grants 5R37 GM32453 and GM081461.

See also: **Bioenergetics:** Mitochondrial Membranes, Structural Organization; Phosphatidylinositol-3-Phosphate; **Lipids Carbohydrates Membranes and Membrane Proteins:** Lipases; Lipid Bilayer Structure; Phospholipid Synthesis in Yeast; **Signaling:** Diacylglycerol Kinases and Phosphatidic Acid Phosphatases; Phosphatidylinositol Bisphosphate and Trisphosphate; Phosphoinositide 4- and 5-Kinases and Phosphatases.

Further Reading

- Carman GM and Han GS (2009) Regulation of phospholipid synthesis in yeast. *Journal of Lipid Research* 50: S69–S73.
- Carman GM and Han GS (2009) Phosphatidic acid phosphatase, a key enzyme in the regulation of lipid synthesis. *Journal of Biological Chemistry* 284: 2593–2597.
- Choi JY and Wu W (2005) Phosphatidylserine decarboxylases as genetic and biochemical tools for studying phospholipid traffic. *Analytical Biochemistry* 347: 165–175.
- Kuge O and Nishijima M (2003) Biosynthetic regulation and intracellular transport of phosphatidylserine in mammalian cells. *Journal of Biochemistry* 133: 397–403.
- Lemmon MA (2008) Membrane recognition by phospholipid-binding domains. *Nature Reviews Molecular Cell Biology* 9: 99–111.
- Michell RH (2008) Inositol derivatives: Evolution and functions. *Nature Reviews Molecular Cell Biology* 9: 151–161.
- Molday RS and Zhang K (2010) Defective lipid transport and biosynthesis in recessive and dominant Stargardt macular degeneration. *Progress in Lipid Research* 49: 476–492.
- Riekhof WR and Voelker DR (2009) The yeast plasma membrane P4-ATPases are major transporters for lysophospholipids. *Biochimica Biophysica Acta* 1791: 620–627.
- Schlame M and Ren M (2006) Barth syndrome, a human disorder of cardiolipin metabolism. *FEBS Letters* 580: 5450–5455.
- Shindou H and Shimizu T (2009) Acyl-CoA: Lysophospholipid acyltransferases. *Journal of Biological Chemistry* 284: 1–5.
- Shirai Y and Saito N (2002) Activation mechanisms of protein kinase C: Maturation, catalytic activation, and targeting. *Journal of Biochemistry* 132: 663–668.
- Snyder F (1999) The ether lipid trail: A historical perspective. *International Journal of Biochemistry, Biophysics, and Molecular Biology* 1463: 265–278.
- van Meer G, Voelker DR, and Feigenson GW (2008) Membrane lipids: Where they are and how they behave. *Nature Reviews Molecular Cell Biology* 9: 112–124.
- Vance DE and Vance JE (2008) *Biochemistry of Lipids, Lipoproteins and Membranes*. Amsterdam: Elsevier.
- Vance DE and Vance JE (2009) Physiological consequences of disruption of mammalian phospholipid biosynthetic genes. *Journal of Lipid Research* 50: S132–S137.
- Veldhuizen R and Possmayer F (2004) Phospholipid metabolism in lung surfactant. *Subcellular Biochemistry* 37: 359–388.
- Wendel AA, Lewin TM, and Coleman RA (2009) Glycerol-3-phosphate acyltransferases: Rate limiting enzymes of triacylglycerol biosynthesis. *Biochimica Biophysica Acta* 1791: 501–506.
- Yen CL, Stone SJ, Koliwad S, Haris C, and Farese RV (2008) Thematic review series: Glycerolipids, DGAT enzymes and triacylglycerol biosynthesis. *Journal of Lipid Research* 49: 2283–2301.
- Zwaal RF and Comfurius P (2005) Surface exposure of phosphatidylserine in pathological cells. *Cellular and Molecular Life Sciences* 62: 971–988.

Glycine Receptors

T Lynagh and B Laube, Technical University of Darmstadt, Darmstadt, Germany

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Bodo Laube and Heinrich Betz, volume 2, pp. 237–243, © 2004, Elsevier Inc.

Glossary

Gephyrin This is derived from the Greek word for bridge and alludes to the fact that the protein anchors the glycine receptor (GlyR) to the subsynaptic cytoskeleton.

Ligand-gated ion channels (LGICs) Family of receptor proteins composed of homologous subunits, whose intrinsic ion channel opens after binding of neurotransmitter.

Startle disease Human hereditary disorder characterized by an exaggerated startle reaction in response to unexpected stimuli.

Strychnine A plant alkaloid derived from the Indian tree *Strychnos nux vomica*. Potent convulsant antagonizing glycinergic inhibition.

Glycinergic inhibition of neuronal excitability in spinal cord and brain stem is mediated via inhibitory glycine receptors (GlyRs). These receptors are composed of two types of homologous membrane-spanning subunits (α and β) and belong to the pentameric nicotinic acetylcholine receptor superfamily of ligand-gated ion channels. Binding of presynaptically released glycine to postsynaptically enriched GlyRs prevents membrane depolarization of nerve cells by opening receptor-intrinsic chloride channels and is specifically antagonized by the convulsant alkaloid strychnine. Mutations in GlyR genes give rise to neuromotor and startle disorders in humans and animals and underline the important roles of GlyRs in motor coordination and sensory processing.

Glycine, the simplest of all amino acids, has diverse functions within the mammalian central nervous system (CNS). It inhibits postsynaptic neurons via strychnine-sensitive glycine receptors and, together with glutamate, enhances neuronal excitation by the activation of excitatory *N*-methyl-D-aspartate receptors (NMDARs). Currents of both receptors are potently modulated by the divalent cation Zn^{2+} , an endogenous modulator of inhibitory glycinergic neurotransmission. Recent studies indicate that a distinct combination of NMDAR subunits is activated by glycine alone, and thus functions as an excitatory glycine receptor. Our results establish complex roles of glycine and Zn^{2+} at this unconventional NMDAR family member. In summary, glycine regulates neuronal excitability via different receptors located at inhibitory and excitatory synaptic connections.

Distribution of Glycine in the Brain

Although glycine is the simplest of all amino acids, it has diverse functions within the CNS. It serves as an inhibitory neurotransmitter at the GlyR and as a co-agonist essential for the activation of excitatory NMDARs, in addition to being involved in the metabolism of proteins and nucleotides in all cells. The regional levels of glycine in the CNS mainly reflect its role in inhibitory neurotransmission, being highest in the medulla oblongata, pons, and spinal cord, and lowest in the cerebral hemispheres and the cerebellum.

Physiological Role and Functional Properties of Glycine Neurotransmission

Principal Properties of Glycinergic Inhibition

Excitation of glycinergic neurons in the CNS causes Ca^{2+} triggered release of presynaptic glycine into the synaptic cleft and results in the activation of postsynaptic GlyRs, which mediate an increase in the chloride conductance of the postsynaptic plasma membrane (Figure 1). These glycinergic synaptic currents have a complex time course with a fast rising phase and an exponential deactivation, which reflect the synchronous activation of GlyR channels by glycine released during exocytosis and the subsequent closure of individual channels, respectively. After receptor unbinding, glycine is removed from the synaptic cleft by rapid reuptake into presynaptic nerve terminals or surrounding glial processes to allow termination of the transmission process (Figure 1).

Roles in Spinal Control of Motor Activity and Sensory Processing

Inhibition at central synapses is mediated by both glycine and γ -aminobutyric acid (GABA). Whereas GABA is primarily used in higher brain regions, glycinergic synapses are particularly abundant in spinal cord and brain stem, that is, structures involved in the control of motor rhythm generation, the coordination of spinal reflex responses, and the processing of sensory signals. Ia glycinergic interneurons in spinal cord mediate reciprocal inhibition in stretch reflex circuits, allowing relaxation of the antagonist muscle, and thus contribute to the coordination of opposing muscles. Renshaw interneurons regulate motoneuron excitability by producing a recurrent inhibition via a negative feedback system. In sensory processing, glycine modulates neuronal circuits in the auditory system and in the retina and suppresses nociceptive signals in the spinal cord.

Synaptic versus Extrasynaptic Inhibition and the GlyR in Higher Brain Regions

Although neurotransmission between nerve cells is mainly mediated by synaptic contacts, GlyRs are also found

using adsorption of the detergent-solubilized receptor on aminostrychnine-agarose followed by elution with the agonist glycine. The resulting preparation contained the pentameric GlyR composed of α and β -subunits and the associated anchoring protein gephyrin.

Glycine Receptor Isoforms

Cloning of GlyR subunit complementary DNAs (cDNAs) revealed that the GlyR belongs to the same superfamily of group I ligand-gated ion channels as the nicotinic acetylcholine receptor (nAChR), the GABA type A (GABA_AR), the serotonin type 3 receptor (5-HT₃R), and the glutamate-gated-chloride channel receptor (GluClR). The GlyR α - and β -subunits share significant sequence homology with each other (>40%) and, to a lesser extent, with other proteins of the nAChR family. Homology is particularly high within the four transmembrane domains and a cysteine-rich loop located within the N-terminal extracellular region. Four different genes encoding GlyR α -subunits (α 1–4) have been identified in vertebrates with an overall sequence identity of >80%. Alternative splicing in the extracellular N-terminal region and in the intracellular loop between transmembrane domains 3 and 4 generates further diversity. So far, only a single GlyR β -subunit gene has been found.

Spatial and Developmental Expression of GlyR Subunits in the Brain

In situ hybridization studies revealed that α 1 GlyR subunit mRNA is mainly found in spinal cord, brain stem, and the colliculi, whereas GlyR α 2-transcripts are found in the hippocampus, cerebral cortex, and thalamus (see also Figure 2). Low levels of α 3-mRNA are detected in spinal cord, cerebellum, olfactory bulb, and hippocampus. The β -subunit gene is expressed throughout the entire CNS. Expression of α -subunits is subject to developmental regulation; α 1-transcripts increase, and α 2-mRNA decreases, after birth. This change in mRNA expression is reflected in an altered GlyR subunit composition and a faster decay time of inhibitory glycinergic postsynaptic currents. The neonatal form primarily composed of α 2-homooligomers exhibits long channel open times, while the adult isoform is a α 1 β -heterooligomer characterized by fast channel kinetics.

Heterologous Expression of the GlyR

After heterologous expression, all GlyR α -subunits form functional homooligomeric channels with properties resembling those of native GlyRs. Single expression of the β -subunit does not result in either glycine-activated currents or [³H]strychnine binding. Coexpression of α - and β -subunits generates heterooligomeric GlyRs that differ from homooligomeric α GlyRs by their insensitivity to the channel blocker picrotoxinin and a reduced main state conductance.

Structural Features of the GlyR

Cross-linking of native receptors, biochemical results, and pharmacological responses of heterologously expressed receptors

have shown that GlyR pentamers are composed of two α 1 and three β -subunits, which have molecular masses of 48 kDa and 58 kDa, respectively. All subunits share a large and highly conserved N-terminal extracellular domain, four putative transmembrane domains (M1–M4), and a short extracellular C-terminus (Figure 2). All four transmembrane segments are α -helices. Structural models of the GlyR have been aided by crystal structures of pentameric ligand-gated ion channels ELIC (RCSB Protein Database entry 2VL0) and GLIC (PDB 3EAM) from bacteria and of the GluClR from *Caenorhabditis elegans* (PDB 3RHW).

Molecular Model of the Ligand-Binding Region

Sequence homology within the group I ligand-gated ion channel superfamily and the crystal structure of the homologous soluble acetylcholine binding protein from the snail *Lymnaea stagnalis* (PDB 1I9B) allows for experimentally validated models of conserved ligand-binding sites. These are located at the extracellular interface between subunits and involve amino acid residues that belong to distinct loops of two adjacent subunits. This site in the GlyR has been described in reasonable detail.

Ion Channel Function

The second transmembrane domain M2 forms the ion channel of the GlyR. It is highly anion-selective, with a halide selectivity sequence of $I^- > Br^- > Cl^-$, and is significantly permeable to bicarbonate. A characteristic of GlyRs is that they do not exhibit a single conductance state but instead display a range of conductance states, with residue Gly 254 within the second transmembrane domain constituting a major determinant of ion flux (Figure 2). Heterooligomeric α 1/ β GlyRs show lower subconductance state levels than α 1 homomeric receptors, and in this respect resemble the adult native receptors.

The Cytoplasmic Loop of the β -Subunit Is Crucial for Postsynaptic Receptor Clustering

A protein of 93 kDa, named gephyrin, copurifies with the GlyR and is enriched on the cytoplasmic face of the postsynaptic membrane. A short amino acid motif within the cytoplasmic loop linking the third and fourth transmembrane domains of the GlyR β -subunit forms the binding site for gephyrin. Thus, synaptically localized GlyR is bound via its β -subunit to the submembraneous scaffold protein gephyrin, which interacts with cytoskeletal and other synaptic proteins, and thereby targets and anchors the GlyR to postsynaptic membrane specializations (Figure 1).

Pharmacological Properties of the GlyR

The Classical Antagonist Strychnine

The alkaloid strychnine constitutes a unique tool in the investigation of the GlyR. In electrophysiological studies, it is the most reliable antagonist to distinguish glycinergic from GABAergic inhibition. Biochemists use glycine-displaceable [³H]strychnine binding to quantify the GlyR in

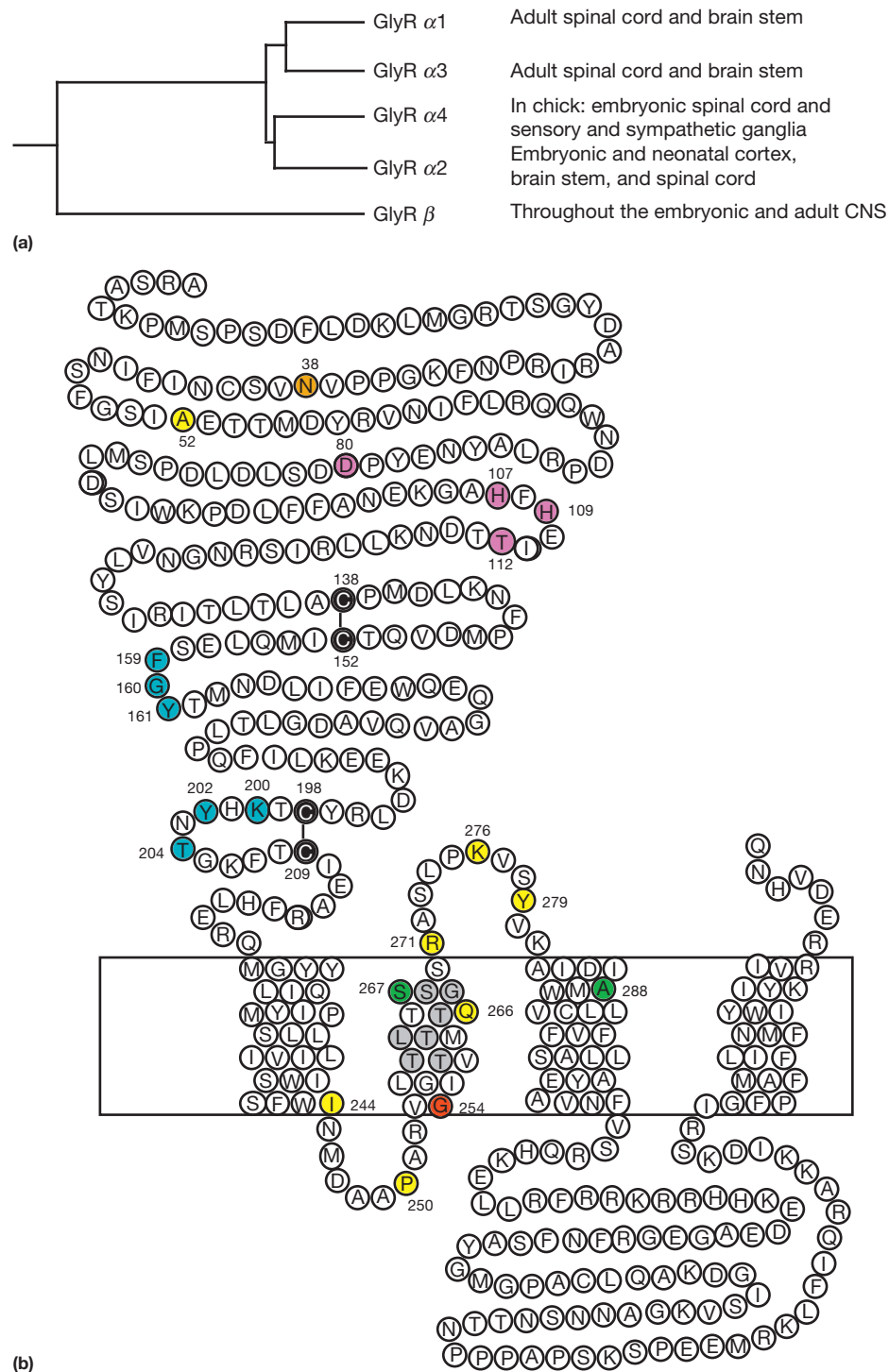


Figure 2 Phylogenetic relationships between and transmembrane topology of vertebrate GlyR subunits. (a) Evolutionary relationships between GlyR subunits. In addition to individual subunits, the chromosomal localizations of the human genes and their major sites of expression are indicated. (b) Transmembrane topology of the human GlyR α 1 subunit, with amino acid residues mutated in human hyperekplexia (gray residues) being indicated. Conserved cysteine residues forming disulfide bridges are shown in black. Residue N38 constitutes the sole N-glycosylation site. Mutation A52S is found in the *spasmodic* mouse. G254 in M2 of the α 1-subunit determines the main conductance and sensitivity to the channel blockers CTP, ginkgolides, dihydropyridines, and chlorinated insecticides. Residue D80 is an important determinant of Zn^{2+} potentiation, whereas residues H107, H109, and T112 are essential for Zn^{2+} inhibition. Secondary structure predictions for the extracellular and transmembrane domains are indicated by amino acid residues arranged in β -sheet (β 1– β 10), or α -helical (M1–M4), structures, respectively.

radioligand-binding assays. The physiological symptoms of strychnine poisoning emphasize the importance of glycinergic inhibition in the control of both motor behavior and sensory processing. Due to its high toxicity, which causes convulsions and, at higher doses, death, strychnine is of low therapeutical value but serves as a rat poison.

Agonistic Amino Acids

In addition to glycine, the endogenous amino-acids β -alanine and taurine display inhibitory activity when applied to neurons. Taurine released nonsynaptically from immature cortical neurons has been found to influence cortical development by activating extrasynaptic GlyRs. Antagonism of taurine reuptake increases GlyR currents, whereas antagonism of β -alanine reuptake has little effect, and thus the role of β -alanine is unclear. Both β -alanine and taurine are partial agonists, which can elicit only submaximal currents.

Antagonism by Channel Block

The alkaloid picrotoxinin acts as a use-dependent but voltage-independent open-channel blocker of the GlyR and can be used to discriminate homo- from heterooligomeric receptors. Hetero-oligomeric GlyRs composed of α and β -subunits are relatively resistant to block by picrotoxinin, whereas homo-oligomeric GlyRs containing only α -subunits are sensitive to low micromolar alkaloid concentrations. Several compounds block the GlyR channel apparently by binding deep within the pore. These include the bulky organic anion cyanotriphenylborate (CTB), plant-derived ginkgolides, dihydropyridines (typically used as calcium channel blockers), and chlorinated compounds used as insecticides (that target insect GABA receptors).

Allosteric Modulation of the GlyR

The number of agents known to potentiate GlyR function is increasing. Therapeutically, potentiation of glycinergic inhibition in the mammalian CNS has substantial promise for both normalizing pathological states and for lowering sensory and motor activity under different conditions including narcosis. The selectivity of allosteric modulators is as yet insufficient to target specific GlyR subtypes.

Anesthetics and alcohol

Since decades, anesthetics and alcohols are known to potentiate GlyR responses. Anesthetic concentrations of propofol and volatile halogenated hydrocarbons such as halothane, enflurane, isoflurane, methoxyflurane, and sevoflurane enhance the effects of low concentrations of glycine at the GlyR. The potencies of n-alkohols increase with chain length of their alkyl groups, up to a cutoff value of twelve carbon atoms. Using site-directed mutagenesis, residues within the transmembrane domains have been shown to be important for the potentiating effect of anesthetics and alcohol.

Glutamate

Paradoxically, the classic excitatory neurotransmitter glutamate has been shown to enhance the response of GlyR to glycine.

This is perhaps of particular significance in the brain, where high levels of glutamate might increase the effect of otherwise insignificant concentrations of glycine. Together with the agonist activity of glycine at certain NMDARs, this complicates the simple idea of excitatory receptors for glutamate and inhibitory receptors for glycine and GABA.

Zinc

The divalent cation Zn^{2+} exhibits biphasic effects on both native and recombinantly expressed GlyRs. Low concentrations of Zn^{2+} (0.1–10 μ M) potentiate submaximal glycine-induced currents about threefold, while higher concentrations cause inhibition. Since Zn^{2+} is present within presynaptic secretory vesicles and is coreleased with the neurotransmitter upon stimulation, this modulation might be of considerable physiological significance. Mutational studies indicate that the Zn^{2+} binding sites for potentiation and inhibition are located at subunit interfaces, as are the agonist binding sites (see also Figure 2).

Neurosteroids, anesthetics, and cannabinoids

Steroids markedly affect glycinergic transmission; however, their actions are variable and show a strong dependence on subunit composition. This does, however, provide an interesting therapeutic lead. *In vivo*, steroid effects on GlyR activity may be particularly important around birth and during adolescence. Anesthetics depress the nervous system partly by enhancing inhibitory neurotransmission. For example, certain effects of propofol require GlyRs, and this correlates with the enhancement of recombinant GlyRs by low concentrations of propofol. Similarly, analgesic effects of endogenous and exogenous cannabinoids (including Δ^9 -tetrahydrocannabinol from cannabis) are due to their enhancement of GlyR function. All three of these classes seem to interact with the GlyR membrane-bound helices.

Ivermectin

The large macrocyclic lactone ivermectin is an anthelmintic that activates invertebrate GluClRs and to a lesser extent mammalian GlyRs. The effects of ivermectin are interesting, because it seems to activate the chloride channel by binding in the transmembrane domain, thus forgoing the canonical extracellular ligand-binding domain. This may help describe what conformational changes are required in channel opening, and the structure–activity relationships of ivermectin at different Cys-loop receptors provides potential leads for subunit-selective macrocyclic lactones.

Posttranslational modifications

GlyR subunits contain consensus sequences for glycosylation (see Figure 2), phosphorylation (protein kinases A and C, Ca^{2+} -dependent kinases), and ubiquitination. Although a regulation of GlyR channel activity by phosphorylation was observed in different preparations, consistent with a potential role of protein kinases in the modulation of synaptic efficacy, the functional significance of GlyR phosphorylation is far from being clear. Also, ubiquitin-mediated internalization and degradation of the GlyR have been suggested to constitute important mechanisms for regulating receptor numbers in the neuronal plasma membrane. Editing of $\alpha 3$ GlyRs in the

hippocampus results in highly sensitive glycine receptors that alter neuronal excitability and may contribute to diseases such as temporal lobe epilepsy.

Disorders of Glycinergic Neurotransmission

Mutations in GlyR subunits causing an impairment of glycinergic transmission result in neurological malfunction and hereditary neuromotor disease.

Human Disorders

The first disorder associated with a mutation in a neurotransmitter receptor was found by positional cloning. Hyperekplexia (startle disease, Kok disease) is characterized by muscle rigidity of nervous system origin, particularly in neonates, and by an exaggerated startle response to unexpected auditory or tactile stimuli. The underlying genetic defect was mapped to chromosome 5q by genetic linkage using DNA markers defining the glycine receptor $\alpha 1$ gene GLRA1 as the prime candidate. Substitutions of the charged residue Arg 271 by uncharged amino acids (Leu or Gln) were identified in four families with autosomal dominant hyperekplexia. These substitutions result in a reduction of agonist sensitivity and glycine-activated current. Recombinant expression of the mutant GlyRs revealed that the reduced glycine current is due to a decrease in single-channel conductance levels. Recessive forms of hyperekplexia have subsequently been identified in other families, which carry other missense or null mutations in the GLRA1 locus (see [Figure 2](#)). However, in only 40–50% of the human patients displaying hyperekplexia-like symptoms GLRA1 mutations have been found, suggesting that genes other than GLRA1 could be involved in hyperekplexia.

Animal Mutants

The syntenies between human chromosome 5q and mouse chromosome 11 led to the discovery that recessive mutations in mouse GLRA1 underlie two mouse neurological mutants, *spasmodic* and *oscillator*. *Spasmodic* displays a phenotype similar to human hyperekplexia and is caused by missense mutation of Ala 52 to Ser (see [Figure 2](#)). *Oscillator* has a more severe phenotype, including progressive tremor and muscle spasms, leading to death by 3 weeks of age. Lethality is caused by a

frameshift-producing deletion predicted to eliminate the large cytoplasmic loop and the fourth transmembrane domain. Homozygous *spastic* mice develop at about 2 weeks postnatally a hyperekplexic phenotype that is associated with a marked reduction in the density of GlyRs in brain stem and spinal cord. The disorder results from the insertion of a LINE-1 element into an intronic region of the gene encoding the β -subunit of the GlyR (GLRB), thereby causing delayed and aberrant splicing of the β -subunit mRNA. A disease resembling the human and murine startle diseases is found in Poll-Hereford cattle and Peruvian Paso horses. These animals display strong myoclonus resulting from reduced or absent glycine receptor expression in the spinal cord. In the Poll-Hereford cattle, a missense mutation at Thr24 of the GlyR $\alpha 1$ gene has been shown to cause translational stop, and thus a complete $\alpha 1$ subunit deficiency in homozygous animals.

See also: **Bioenergetics: Ligand-Operated Membrane Channels: GABA.**

Further Reading

- Becker CM (1992) Convulsants acting at the inhibitory glycine receptor. In: Herken H and Hucho F (eds.) *Handbook of Experimental Pharmacology*, pp. 539–575. Berlin: Springer.
- Fritschy JM, Harvey RJ, and Schwarz G (2008) Gephyrin: Where do we stand, where do we go? *Trends in Neurosciences* 31: 257–264.
- Grudzinska J, Schemm R, Betz H, and Laube B (2005) The β subunit determines the ligand binding properties of synaptic glycine receptors. *Neuron* 45: 727–739.
- Haeger S, Kuzmin D, and Detro-Dasse S (2010) An intramembrane aromatic network determines pentameric assembly of Cys-loop receptors. *Nature Structural and Molecular Biology* 17: 90–98.
- Kuhse J, Betz H, and Kirsch J (1995) The inhibitory glycine receptor: Architecture, synaptic localization and molecular pathology of a postsynaptic ion channel complex. *Current Opinion in Neurobiology* 5: 318–323.
- Laube B and Betz H (1999) Purification and heterologous expression of inhibitory glycine receptors. *Methods in Enzymology* 294: 260–273.
- Schofield P (2002) The role of glycine and glycine receptors in myoclonus and startle syndromes. *Advances in Neurology* 89: 263–274.
- Yevenes GE and Zeilhofer HU (2011) Allosteric modulation of glycine receptors. *British Journal of Pharmacology* 164: 224–236.

Relevant Website

<http://www.pasteur.fr> – EMBL-EBI: Ligand-Gated Ion Channel Database.

Glycogen Metabolism

P J Roach, Indiana University School of Medicine, Indianapolis, IN, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Enzyme A protein that catalyzes or accelerates chemical reactions, such as those involved in metabolism, to interconvert compounds in cells.

Hormone A compound in the blood released from specialized organs, called endocrine glands, which acts as a messenger within the body to regulate the function of individual target tissues.

Metabolism The process by which molecules are interconverted within cells among different related chemical forms, called metabolites.

Pancreas An organ with multiple functions related to the utilization of foods. The endocrine pancreas responds to the nutrient levels in the blood to release the hormones insulin and glucagon.

Glycogen Structure

Many nutrients contain either glucose or compounds that are easily converted to glucose, so that it is an important nutritional currency for many organisms. Glycogen is a glucose polymer, meaning that individual glucose units are joined together by repeated chemical linkages to form large molecules containing thousands of glucose residues (Figure 1). Two types of linkages are present in glycogen causing a branched polymeric structure. These same two linkages are present in plant starches which serve a similar function as glycogen. When needed, the glucose units can be recovered by breaking down the glycogen. A possible explanation for the storage of glucose as glycogen rather than glucose monomers is due to the lower osmotic activity of the polymer. Some minor constituents of glycogen, such as covalently attached phosphate, have also been noted.

Glycogen Metabolism

Glycogen exists as a particle in which the carbohydrate core is associated with several of the proteins needed to make it and degrade it. The glycogen molecule expands or contracts due to growth or breakdown of the outer branches of the molecule (Figure 1).

Synthesis

Glycogen in cells is synthesized either directly from glucose transported into the cells by transporter proteins or from other cellular metabolites that are converted to the activated sugar precursor, uridine diphosphate (UDP)-glucose in animals and adenosine diphosphate (ADP)-glucose in bacteria (or for starch biosynthesis in plants). Three enzymes are needed to synthesize glycogen: an initiator protein called glycogenin that attaches glycogen residues to itself, the enzyme glycogen synthase responsible for incorporating most of the glucose moieties of glycogen from UDP-glucose, and a special enzyme, branching enzyme, to introduce the branch points (Figure 2).

Degradation

Controlled degradation of glycogen requires two enzymes (Figure 2). Glycogen phosphorylase releases a glucose unit from the polymer as a metabolite, called glucose-1-phosphate, that can be used by the cell. Since phosphorylase stalls at the branch points, a second specialized enzyme, the debranching enzyme, is needed to degrade past them. In muscle, the process stops at another metabolite of glucose, called glucose-6-phosphate, which is used in the muscle cell to provide energy or building blocks. In liver, the glucose-6-phosphate can be converted, by an enzyme called glucose-6-phosphatase, to free glucose that is exported to the bloodstream. There is a second pathway to degrade glycogen. In most cell types, there is a house-cleaning process whereby a portion of the cellular content is continuously being delivered to an organelle, called lysosome or vacuole, to recycle the building blocks of the macromolecules, for example, amino acids from proteins and glucose from glycogen. A specialized enzyme, called an α -glucosidase, is present in the lysosome to break glycogen down to glucose.

Glycogen Function and Control

Metabolic Role

Whether in bacteria or mammals, glycogen serves as a reservoir of energy and/or building blocks that can be called upon in times of need. In microorganisms, glycogen is generally made in response to impending nutritional deprivation, as judged by sensing the composition of the environment, and stored for use late in starvation. In mammals, the nutritional cues that govern glycogen metabolism in tissues are indirect and are borne by the blood levels of glucose itself and several important regulatory molecules called hormones. After a meal, nutrients enter the gut and lead to increased blood glucose, which triggers the pancreas to release the hormone insulin, a signal of the fed condition. Insulin-sensitive tissues – liver, fat, and especially skeletal muscle – respond by increasing transport of glucose into the cells (Figure 3(a)). Insulin also causes activation of the enzyme glycogen synthase, enhancing the conversion of glucose to glycogen. This is part of the process

of blood glucose homeostasis whereby the body seeks to maintain the blood glucose level within safe limits. As the time after the last meal increases, the blood glucose level normalizes; the pancreas delivers less insulin and begins to secrete the hormone glucagon, which signals the unfed state (**Figure 3(b)**). The primary target is the liver which responds by breaking down glycogen to produce glucose that can be delivered to the bloodstream, another contribution to blood glucose homeostasis.

Role in Muscle or Exercise

Unlike the liver, muscle cannot break down glycogen all the way to free glucose for export to the bloodstream. Muscle glycogen is instead broken down to metabolites that can be converted to energy to fuel muscular contraction. The stimulus for the breakdown of muscle glycogen is the same as for contraction, and nerve signals instructing the muscle to contract concurrently activate glycogen phosphorylase, which

breaks down glycogen. Most muscles do not rely entirely on glycogen and other fuels, such as fat, can also be utilized. Increased muscular activity is often associated with increased levels of the stress hormone, epinephrine (also called adrenaline). In response to anxiety, which in evolution might have involved fleeing an aggressor, the adrenal gland produces epinephrine which activates phosphorylase in muscle and liver, providing, respectively, enhanced muscle performance and increased fuel to the bloodstream. Muscular activity also sensitizes the tissue to insulin and activates glycogen synthase, so that it is ready to reform glycogen once contraction ceases.

Glycogen and Disease

Glycogen Storage Diseases

A number of genetic diseases are known in which genes encoding various enzymes involved in glycogen metabolism are mutated (**Table 1; Figure 2**). These diseases range in severity from relatively mild to early death. Their frequency is for the most part relatively low. Most of these diseases are associated with overaccumulation of glycogen or the formation of abnormally branched glycogen. One disease, glycogen storage disease 0, is associated with mutations that impair liver glycogen synthesis and patients are prone to low blood glucose (hypoglycemia). Lafora disease, a progressive myoclonus epilepsy, is characterized by the formation of Lafora bodies in neurons and other tissues. Lafora bodies contain large amounts of poorly branched glycogen and may be caused by impaired metabolism of the small amount of phosphate found in glycogen.

Diabetes Mellitus

Diabetes occurs when blood glucose homeostasis fails and blood glucose levels are elevated. High blood glucose (hyperglycemia) leads to many complications including heart,

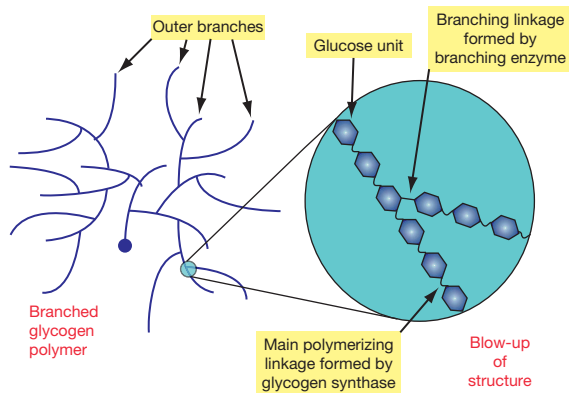


Figure 1 Glycogen is a polymer of glucose units.

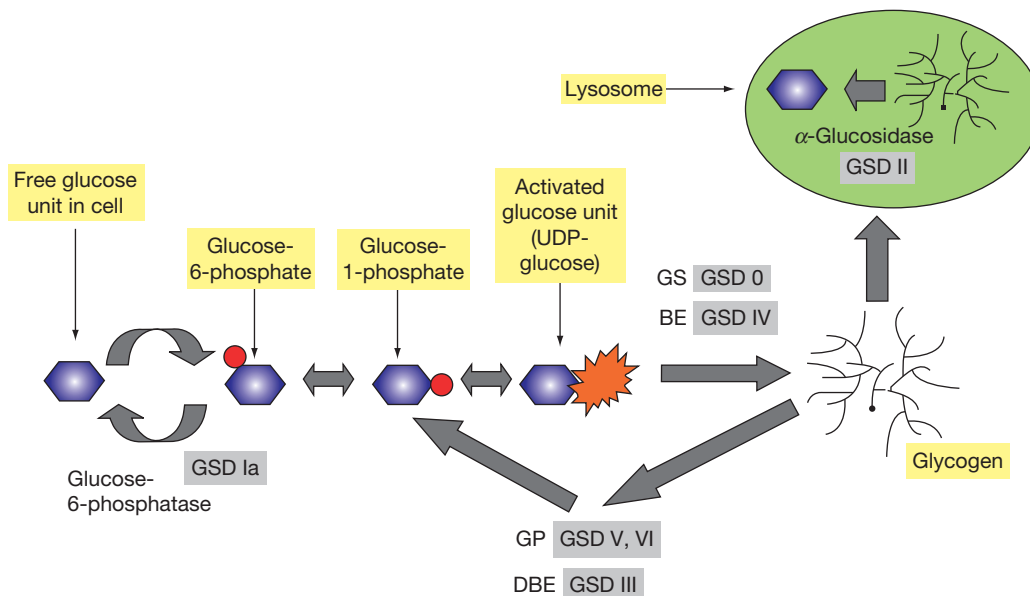


Figure 2 The pathways of glycogen production and degradation. BE, branching enzyme; DBE, debranching enzyme; GP, glycogen phosphorylation; GS, glycogen synthase; GSD, glycogen storage disease.

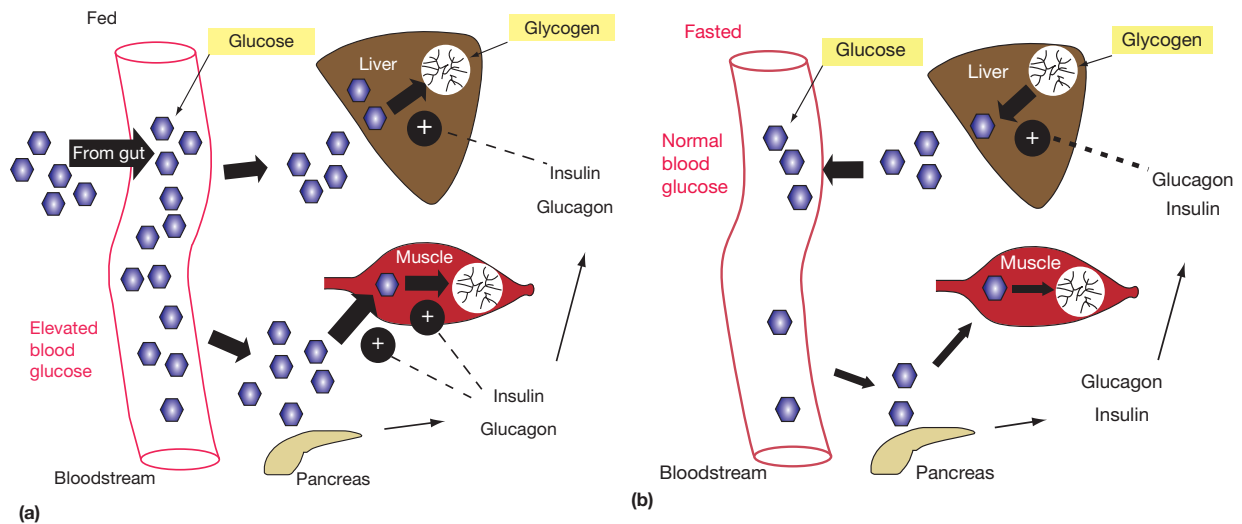


Figure 3 The interconversion of glucose and glycogen in the fed (a) and fasted (b) states.

Table 1 Selected glycogen storage diseases (GSD)

GSD type	Name	Enzyme affected	Affected organ(s)	Glycogen	Description
0					
Ia	Von Gierke disease	Glycogen synthase Glucose-6-phosphatase	Liver Liver	Reduced amount Normal structure, increased amount	Hypoglycemia Massive liver enlargement, multiple metabolic problems, treatment improving
II	Pompe disease	Lysosomal α -glucosidase	All	Increased glycogen in lysosomes	Usually death before age of 2
III	Cori disease	Debranching enzyme	Muscle, liver	Short outer branches, increased amount	Similar to type I, but less severe
IV	Andersen disease	Branching enzyme	Liver	Long outer branches, normal amount	Usually death from liver failure before age of 2
V	McArdle disease	Muscle glycogen phosphorylase	Muscle	Normal structure, moderate increase in amount	Impaired exercise capability
VI	Hers disease	Liver glycogen phosphorylase	Liver	Increased amount	Similar to type I but less severe
	Lafora disease	Laforin, malin	Brain, many others	Reduced branching	Epilepsy, fatal

kidney, nerve, and eye disease. This disease is both extremely common and growing in epidemic fashion as enrichments in diets around the world are coupled with diminished physical activity and obesity. The more prevalent form, type 2 diabetes, has a genetic component, likely linked to multiple genes, and an environmental component, lack of exercise and unhealthy diet. Type 2 diabetes is associated with an impaired ability to store glucose as glycogen, but whether there is a causal relationship with defective glycogen metabolism is still not certain.

See also: Metabolism Vitamins and Hormones: Diabetes; Glycogen Storage Diseases.

Further Reading

- Chen Y-T (2001) Glycogen storage diseases. In: Scriver CR, Beaudet AL, Sly WS, and Valle D (eds.) *The Metabolic and Molecular Bases of Inherited Disease*, 8th edn., vol. 1, pp. 1521–1551. New York, NY: McGraw-Hill.
- Harris RA (2002) Carbohydrate metabolism I: Major metabolic pathways and their control. In: Devlin TM (ed.) *Textbook of Biochemistry with Clinical Correlations*, 5th edn., pp. 597–664. New York, NY: Wiley-Liss.
- Harris RA and Crabb DW (2002) Metabolic interrelationships. In: Devlin TM (ed.) *Textbook of Biochemistry with Clinical Correlations*, 5th edn., pp. 861–902. New York, NY: Wiley-Liss.
- Preiss J and Walsh DA (1981) The comparative biochemistry of glycogen and starch. In: Ginsburg V and Robbins P (eds.) *Biology of Carbohydrates*, vol. 1, pp. 199–314. New York, NY: Wiley.
- Roach PJ, DePaoli-Roach AA, Hurley TD, and Tagliabracci VS (2012) Glycogen and its metabolism: Some new developments and old themes. *Biochemical Journal* 441: 763–787.

Glycogen Storage Diseases

G H Sack Jr. Johns Hopkins University, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Glycogen A complex, branched polymer of glucose that serves as a site of short-term glucose storage as well as a source of blood glucose during fasting.

Hypoglycemia The clinical condition of low blood glucose levels with physiologic consequences including seizures caused by brain dysfunction.

Myopathy Disease of the muscle, in this context reflecting inadequate glucose (and, hence, energy) supply or mobilization from glycogen stores.

Phosphorolysis The use of phosphate to remove successive glucose residues from the reducing end of the glycogen polymer.

Glycogen Structure

As shown in **Figure 1**, glycogen is composed of a large number of glucose polymers with ($\alpha 1 \rightarrow 4$) linkages that are branched with ($\alpha 1 \rightarrow 6$) linkages every 4–10 units. The resulting particles, called β -particles, can become large and contain 6×10^5 individual glucose monomers; these can be grouped into even larger particles called α -particles. The β -particles are organized around a central protein called glycogenin, upon which the initial glucose polymerization is based and thus glycogen is technically a large glycoprotein in which the carbohydrate predominates. There is also a thin bounding membrane that appears to contain most of the enzymes of glycogen metabolism. These particles often can be seen in electron micrographs. Glycogen is found mainly in liver and muscle.

Glycogen Function

Glycogen is a source of glucose for both muscle and liver. As such, it serves as a short-term buffer for glucose homeostasis (liver) and as a source of immediate glucose for energy (muscle). The synthesis and degradation of glycogen are closely regulated. Consistent with glycogen's central role in homeostasis, these processes are controlled by different enzyme systems and hormones. Synthesis is not the reverse of degradation and most of the enzymes catalyze reactions strongly in favor of their products (i.e., they are not at equilibrium). Several of the enzymes are regulated by covalent modification.

Glycogen Metabolism

Figure 2 presents a summary of the pathways involved in glycogen metabolism. Note that several are unique to either liver or muscle. Considerable metabolic control is also exerted on these enzymes as shown.

Glycogen Synthesis

The first committed step in glycogen synthesis is the formation of uridine diphosphate (UDP) glucose by UDP glucose pyrophosphorylase. The glycogen polymer begins with the attachment of glucose to the protein glycogenin. UDP glucose is the substrate for polymerization via ($\alpha 1 \rightarrow 4$) linkages catalyzed by glycogen synthase. The branching enzyme is bifunctional, cleaving the linear chain and creating a branch via an ($\alpha 1 \rightarrow 6$) linkage. This leads to a reduced length of the final polymer and the production of more termini – leading to a more compact particle and providing a substrate suitable for very rapid degradation.

Glycogen Degradation

Most individual glucose residues are removed from the glycogen polymer by phosphorolysis. Different phosphorylase enzymes catalyze this reaction in muscle and liver. These proteins are genetically unique – synthesized from different loci. Muscle phosphorylase has been studied in particularly great detail. It functions as a dimer and contains sites for both covalent and allosteric modifications. Phosphorylase must itself be phosphorylated to be active, a reaction catalyzed by phosphorylase kinase (which is dependent on modification by a cyclic adenosine monophosphate (cAMP)-dependent kinase and calcium).

Due to the branched nature of mature glycogen, either muscle or liver phosphorylase by itself can only partially degrade the polymer. Phosphorylase cannot cleave beyond four glucose residues from the branch point. Glycogen subjected to treatment with phosphorylase alone is converted into limit dextran – which is a still very large molecule. The action of the debranching enzyme eliminates the barrier to phosphorolysis by transferring the three distal glucose residues from the side chain to the nonreducing terminus of the parent chain. This leaves a single glucose remaining from the side chain attached by an $\alpha 1 \rightarrow 6$ linkage. The debranching enzyme also hydrolyzes this remaining side chain residue freeing a single glucose molecule and leaving a continuous

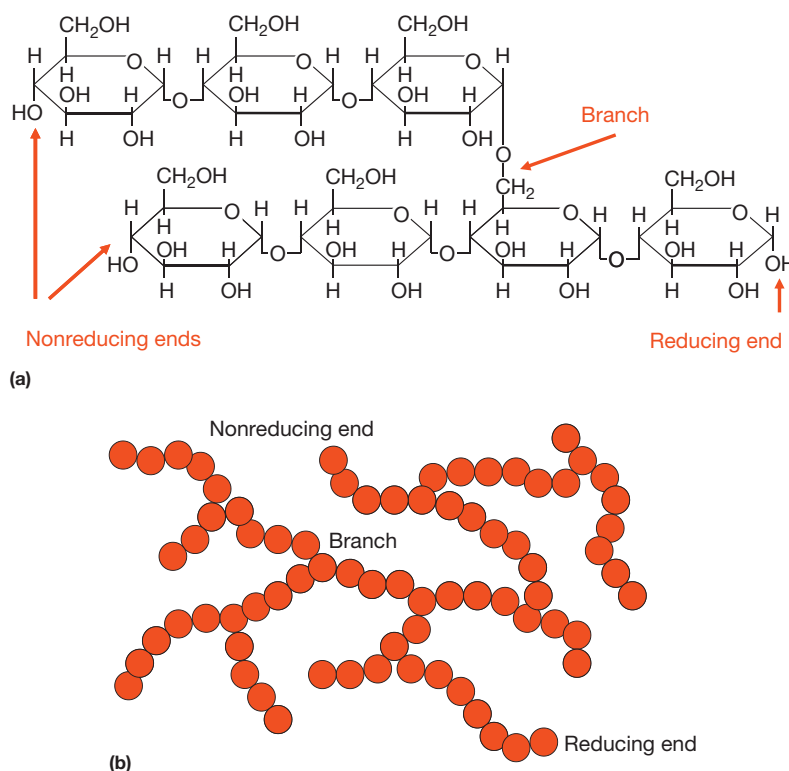


Figure 1 (a) Details of glycogen polymer showing linear ($\alpha(1\rightarrow4)$) and branch ($\alpha(1\rightarrow6)$) linkages. (b) Example of glycogen at lower resolution showing branches and reducing and nonreducing ends.

chain connected by $\alpha(1\rightarrow4)$ linkages, a substrate suitable for phosphorylase.

The products of glycogen degradation are thus predominantly glucose-1-phosphate and a small amount of free glucose from hydrolysis of the single remnant glucose of the side chain. The reversible phosphoglucomutase converts glucose-1-phosphate to glucose-6-phosphate. Because glucokinase is irreversible, the glucose-6-phosphate enters the endoplasmic reticulum (ER) by its own transporter. Within the liver ER, the glucose-6-phosphate is hydrolyzed and the free glucose is then transported out of the cell. Glucose-6-phosphatase is largely confined to the liver and thus the degradation of glycogen in muscle is not a source of blood glucose.

Control of Glycogen Metabolism

Because of the irreversibility of many of the enzymes involved and consideration of the details of **Figure 2**, the synthesis and degradation of glycogen do not use the same pathways. Furthermore, the enzymes mediating both sets of reactions are under tight control. This control is often mediated by adding or removing phosphate to and from the proteins as shown. Several critical control points are

1. The phosphorylated form of glycogen synthase (noted as "b" in the figure) is activated by removing the phosphate to give the "a" form by the active form of phosphoprotein phosphatase. The latter is produced by phosphorylation using adenosine triphosphate (ATP), stimulated by insulin.
2. Phosphorylase is active in its phosphorylated form, produced by the active form of phosphorylase kinase and ATP.

3. Phosphorylase kinase itself is activated by phosphorylation by cAMP-dependent protein kinase A.
4. The active, phosphorylated forms of both phosphorylase kinase and phosphorylase are inactivated through removal of the activating phosphate by the active form of phosphoprotein phosphatase as shown.

Thus, modification of existing enzymes regulates glycogen metabolism. This exquisitely sensitive system is subject to hormonal control. For example, a rise in insulin (the "fed" state with high blood glucose levels) activates phosphoprotein phosphatase (by phosphorylation). This enzyme then activates glycogen synthase and inactivates both phosphorylase and phosphorylase kinase, resulting in the synthesis and accumulation of glycogen.

By contrast, in the "fasting" state (low blood glucose levels) insulin levels are low, inactivating phosphoprotein phosphatase. Other hormones – glucagon and epinephrine – activate cAMP-dependent protein kinase A which phosphorylates (thereby activating) phosphorylase kinase. The latter, in turn, activates phosphorylase itself by phosphorylation leading to glycogen degradation.

This complex but sensitive system provides a critical buffer for blood glucose under common, relatively short-term conditions.

Defects in Glycogen Metabolism and Their Consequences

The remarkably sensitive metabolic system based on glycogen is central to normal physiology and is an essential defense

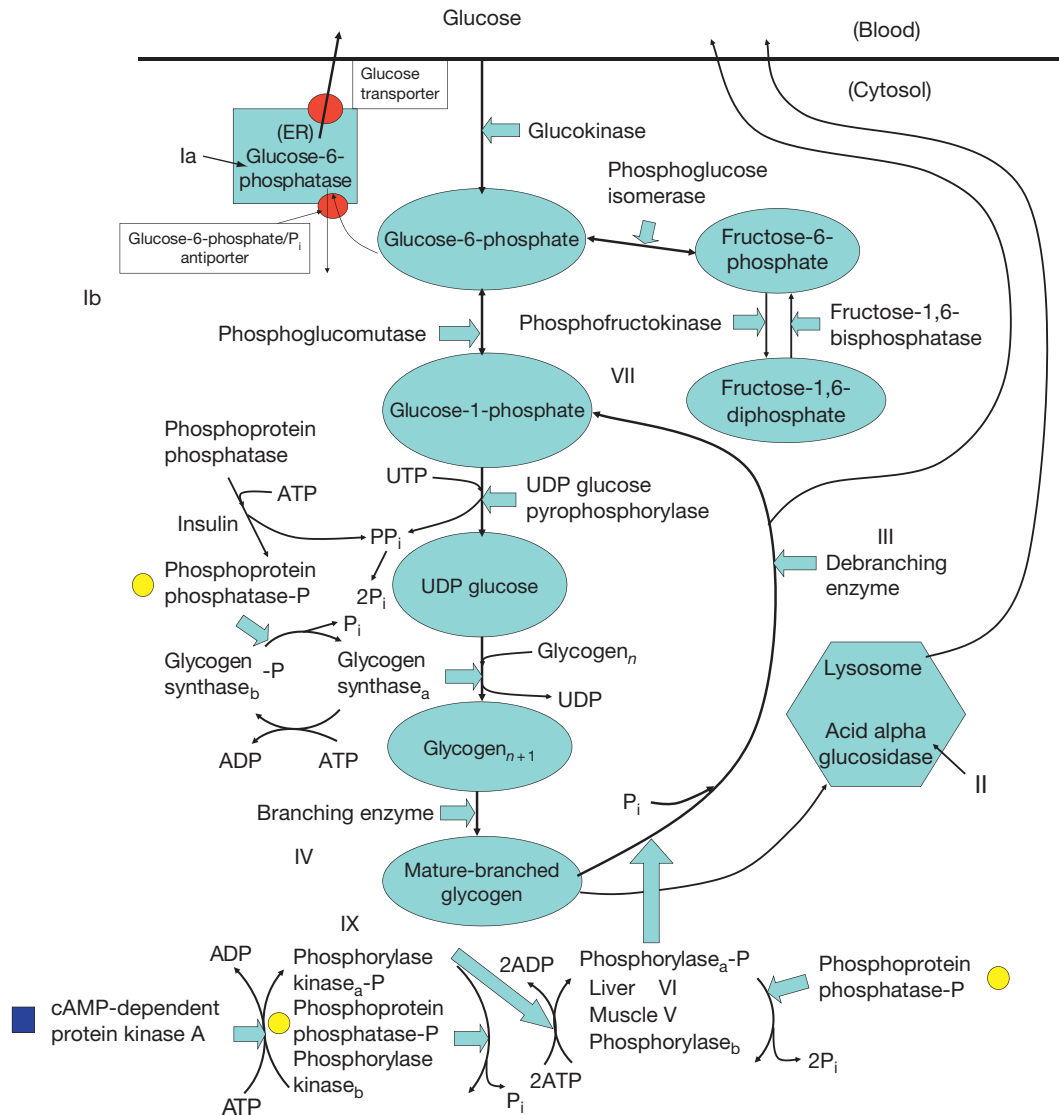


Figure 2 Metabolic pathways directly involved in the synthesis and degradation of glycogen. Note that blood glucose (top) is both the source of substrates for glycogen synthesis as well as the product of glycogen degradation. Compounds in ovals are biochemical intermediates. The rectangle represents the endoplasmic reticulum (ER) and the hexagonal structure represents a lysosome. Small closed circles show three sites for participation by the phosphorylated form of phosphoprotein phosphatase, formed by insulin stimulation. The small, closed square shows the site of stimulation by glycogen and epinephrine through cAMP-dependent protein kinase A. The sites of enzyme defects in GSDs are indicated by the bold roman numerals. ADP, adenosine diphosphate; ATP, adenosine triphosphate; cAMP, cyclic adenosine monophosphate; P_i, inorganic phosphate; PP_i, inorganic pyrophosphate; UDP, uridine diphosphate; UTP, uridine triphosphate.

against wide fluctuations in blood glucose levels. Therefore, defects in the pathway are often manifested by failure of glucose homeostasis with the frequent development of symptomatically low blood glucose levels (hypoglycemia). Furthermore, the sheer mass of glycogen polymers that develop when its degradation is inadequate can lead to organ enlargement and, ultimately, dysfunction. Inadequate mobilization of glucose from glycogen can also limit muscle contractility, especially in acute events.

The central position of blood-borne glucose in the energy metabolism of many organs (especially the brain) has meant that careful management of glucose levels is often a prime consideration in the care for affected individuals. We will

consider this group of disorders through reference to [Figure 2](#) and [Table 1](#). It is of historic interest that the definition and biochemical characterization of many of the enzymes in this system were based on detailed study of affected individuals. Although general features of individual types are presented, it is important to remember that the genetic changes underlying the disorders both within and among clinical categories are not uniform. Much clinical heterogeneity likely reflects the fact that different mutations are often responsible for the failure or dysfunction of a given enzyme in different kindreds. This latter fact also complicates diagnostic studies that rely on mutation analysis alone. Several of the categories are often recognized by eponyms as shown.

Table 1 The glycogen storage diseases

<i>Designation</i>	<i>Enzyme</i>	<i>Eponym</i>	<i>OMIM#</i>	<i>Map location</i>
Ia	Glucose-6 phosphatase	von Gierke	232200	17q21
Ib	Glucose-6-phosphate/P _i antiporter		232220	11q23
II	Lysosomal (α -1,4-glucosidase)	Pompe	232300	17q25.2-q25.3
III	Debranching enzyme	Cord/Forbes	232400	1p21
IV	Branching enzyme	Andersen	232500	3p12
V	Muscle phosphorylase	McArdle	232600	11q13
VI	Liver phosphorylase	Hers	232700	14q21-q22
VII	Phosphofructokinase	Tarui	232800	12q13.3
IXa	Phosphorylase kinase α -subunit		306000	Xp22.2-p22.1
IXb	Phosphorylase kinase β -subunit		261750	16q12-q13
IXc	Phosphorylase kinase γ -subunit		172471	16p12.1-p11.2

Data from Mendelian Inheritance in Man catalog (OMIM).

Type I (von Gierke)

Originally described in 1929, individuals with this group of disorders have early difficulties with low blood glucose levels (hypoglycemia). In the neonatal period, this can lead to seizures. Slightly later, liver enlargement is detected. Early biochemical study of the liver of affected individuals showed defective activity of glucose-6-phosphatase. Subsequent study of more individuals with similar manifestations revealed more details of the pathway.

Type Ia

This, the largest fraction of affected individuals, is notable for defects in the gene and protein for glucose-6-phosphatase, which have been confirmed by molecular study. Subsequently, however, it was determined that the active site of glucose-6-phosphatase lies within the ER and that hepatic microsomes from some affected individuals could degrade glucose-6-phosphate after detergent treatment. Thus, a transport process was implicated.

Type Ib

These individuals not only have aberrant glucose homeostasis but also have susceptibility to certain infections due to neutrophil and macrophage dysfunction. Cloning and characterizing mutations in the glucose-6-phosphate translocase identified the problem in this group. Although it had been proposed that a third group had defective phosphate transport (and had been given the designation Ic), recent analysis has shown that most of these individuals have mutations already identified in the same translocase as in Ib. The translocase itself is now recognized to have a dual role as an antiporter – mediating both heterologous (glucose-6-phosphate: inorganic phosphate (P_i)) and homologous (P_i:P_i) exchanges. This is consistent with the fact that hydrolysis of glucose-6-phosphate within the ER leads to a local increase in P_i as a product which can then be exchanged for more glucose-6-phosphate substrate to continue the process.

Although the possible existence of another, independent phosphate transporter has been proposed to account for another way of achieving the same clinical picture, there is yet no convincing evidence for this. Thus, a putative type Id has not been conclusively identified.

The care of individuals affected with glycogen storage disease (GSD) Ia/b is similar and is initially based on prevention of hypoglycemia by frequent feedings, often with the use of a slowly hydrolyzable glucose polymer (e.g., cornstarch). Liver transplantation has also been used. The long-term complications of this group of disorders are still being determined, but the prognosis has been considerably improved with adequate control of blood glucose levels.

Type II (Pompe)

This type of GSD is unique because it involves dysfunction of an intralysosomal enzyme – α -1,4-glucosidase (also called acid maltase). The characteristic feature of this group of individuals is accumulation of glycogen within lysosomes. Historically, type II GSD was the first of the so-called “lysosomal storage diseases” to be defined. It is important to recall that glycogen storage and the enzymes for its synthesis and degradation (except for glucose-6-phosphatase) are found in the cytoplasm. It has been proposed that lysosomes take up cytoplasmic glycogen and hydrolyze it within their unique low pH environment.

Intralysosomal accumulation of glycogen in GSD II individuals is more notable in muscle than in liver and, thus, many of the manifestations are related to muscle dysfunction. The disease often presents as a ‘myopathy’ with symptoms and signs of muscle abnormalities. The heart muscle can also be involved. Muscle biopsy usually shows the characteristic glycogen accumulation within vacuoles. The severity and rate of progression of the disorder vary widely from individuals with severe disease presenting in infancy to those with a more gradual adult onset. In general, these clinical differences reflect different mutations in the same gene, but a direct correlation is not always possible. Blood glucose levels are not usually affected and liver enlargement is not common. Treatment is now approached by enzyme replacement using alglucosidase-alfa (Myozyme®).

Type III (Cord/Forbes)

The absence or dysfunction of the debranching enzyme leads to the accumulation of glycogen with an abnormal structure

as noted above. This 'limit dextran' represents the product of phosphorolysis up to the limits imposed by the branches. Such products are bulky with short outer branches.

As infants, affected individuals resemble those with type I GSD – they have hypoglycemia and liver enlargement. In general, the manifestations become less severe with age and individuals can survive to adulthood. Because both liver and muscle forms of the enzyme exist, rare individuals can show selective organ involvement. Early reports suggest that there may be rare individuals who lack either of the two functions of this bifunctional enzyme, but no recent studies are available.

Treatment is generally the same as for individuals with type I GSD, although the requirements are generally less stringent. Liver transplantation has also been used, but does not affect muscle dysfunction.

Type IV (Andersen)

Consistent with the structure and synthesis of mature glycogen, defective formation of branches leads to very long polymers and bulky particles. The resulting molecules resemble amylopectin and the manifestations are related to their accumulation in the liver with attendant dysfunction of metabolic activity and, particularly, derangement of liver architecture known as cirrhosis. Complications of cirrhosis ensue in many individuals leading to liver failure and death. Since these polymers are degradable, low blood glucose complications are generally not found. A rare variety appears to largely involve muscle. Liver transplantation has been successful in selected individuals.

Type V (McArdle)

Types V and VI represent the consequences of organ-specific dysfunction of phosphorylase, the primary enzyme for glycogen degradation. Type V, the muscle-specific form, is recognized by intolerance to prolonged muscle activity. In most individuals, quiet activities are managed well and no symptoms develop, but rapid exercise can cause severe, acute muscle pain and the loss of muscle tissue integrity. Muscle cramps after exercise are common and may progress to muscle destruction – rhabdomyolysis. Usually, symptoms develop in adolescence or young adulthood. Because muscle glycogen is the source of energy for acute contraction, the manifestations of type V GSD often are related to the intensity of the exercise.

Because of the reduced availability of glucose for muscle contraction, blood levels of lactic acid fail to rise after ischemic exercise. This can be tested by measuring lactic acid levels in the forearm following repeated exercise. Although useful, these measurements are not specific for type V and can be found in other disorders affecting these pathways. Another useful correlative test is ^{31}P MRI, which shows a lack of the expected decrease in intracellular pH with exercise (as well as an exaggerated reduction in phosphocreatine).

The phosphorylase found in muscle is unique to it, and thus its mutations do not directly affect other tissues. Ingesting sucrose prior to exercise may be helpful, but the most effective treatment is avoidance of intense exercise; this often can lead to a relatively stable clinical course.

Type VI (Hers)

The mutations in this disorder are the hepatic counterpart of those in type V GSD. In general, the manifestations are milder than in individuals with type I(a,b), but poor blood glucose control and liver enlargement due to glycogen storage are observed. Muscle problems are not encountered (cf. type V). The prognosis is usually good if control of blood glucose levels can be maintained; often this can be accomplished through diet alone.

Type VII (Tauri)

Dysfunction of the muscle form of phosphofructokinase leads to manifestations similar to those in individuals with GSD type V – exercise-related muscle pain and fatigue. However, the symptoms are generally less severe. The likely pathophysiology involves elevation of the level of glucose-6-phosphate in muscle cells leading to increased glycogen synthesis. Giving glucose to affected individuals leads to lower levels of fatty acids, which normally would serve as an important energy source for muscle during contraction. Thus, while (as noted above) administering glucose can be helpful to muscle cells in type V GSD, it can aggravate the consequences of the mutation in type VII. There is no specific treatment for this form of GSD. Avoidance of strenuous exercise (similar to the advice for individuals with type V GSD) is generally effective.

Type IXa

Disorders in this category are grouped by effects on a single function. Phosphorylase kinase is a heterotetramer (subunits α , β , γ , and δ). Of several different disorders, the most frequently recognized is the X-linked form, affecting only males (it is the only X-linked form of GSD). Several mutations have been identified in the α -subunit of phosphorylase b kinase. Two α -subunits (for muscle and liver, respectively) are found at nonadjacent loci on the X-chromosome, but only the hepatic form has been associated with this disorder (GSD IXa).

These individuals usually show moderate liver enlargement and liver-function abnormalities early in life; growth is often delayed. Problems due to low blood sugar levels are not common. In general, the manifestations become less prominent with age. Treatment is not generally required once the diagnosis has been recognized and adults generally manage well.

Type IXb

Defects in the β subunit of phosphorylase kinase lead to clinical manifestations similar to those in α -subunit mutation patients. However, muscle weakness may be found; and both males and females can be affected, consistent with the genetic location for the gene. Treatment is not usually needed.

Type IXc

Mutations in the testis/liver isoform of the γ -subunit of phosphorylase kinase also affect both sexes. Although most clinical

manifestations are similar to those seen in GSD IXa, hepatic cirrhosis may develop in these individuals.

Summary

Although GSDs are rare, as a group they have led to the understanding of many aspects of glycogen metabolism. Clinical difficulties develop based on the site of the underlying defect(s). The care of affected individuals emphasizes controlling blood glucose levels and minimizing organ damage that can result from inadequate glucose mobilization and/or glycogen accumulation. Liver transplantation and gene-product administration are treatment options for selected patients.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Carbohydrate Chains: Enzymatic and Chemical Synthesis; **Metabolism Vitamins and Hormones:** Gluconeogenesis; Glucose/Sugar Transport in Mammals; Glycogen Metabolism; Glycolysis Overview; **Protein/Enzyme Structure Function and Degradation:** Allosteric Regulation; **Signaling:** Glycogen Synthase Kinase-3.

Further Reading

- Chen S-Y, Pan C-J, Nandigama K, et al. (2008) The glucose-6-phosphate transporter is a phosphate-linked antiporter deficient in glycogen storage disease type Ib and Ic. *FASEB Journal* 22: 2206–2213.
- Chen Y-T (1999) Glycogen storage diseases. In: Scriver CR, Beaudet AL, Sly WS, et al. (eds.) *Metabolic and Molecular Basis of Inherited Disease*, 8th edn., vol. I, ch. 71, pp. 1521–1551. New York, NY: McGraw-Hill.
- Chou JY and Mansfield BC (1999) Molecular genetics of type I glycogen storage disease. *Trends in Endocrinology and Metabolism* 10: 104–113.
- Cori G and Cori C (1952) Glucose-6-phosphatase of the liver in glycogen storage disease. *Journal of Biological Chemistry* 199: 661–667.
- Hirschhorn R and Reuser AJJ (1999) Glucogen storage disease type II: Acid α -glucosidase (acid maltase) deficiency. In: Scriver CR, Beaudet AL, Sly WS, et al. (eds.) *Metabolic and Molecular Basis of Inherited Disease*, 8th edn., vol. III, ch. 135, pp. 3389–3420. New York, NY: McGraw-Hill.
- Kazemi-Esfarjani P, Skomorowska E, Jensen TD, Haller RG, and Vissing J (2002) A nonischemic forearm exercise test for McArdle disease. *Annals of Neurology* 52: 153–159.

Relevant Website

<http://www.ncbi.nlm.nih.gov> – National Library of Medicine; OMIM (Online Mendelian Inheritance in Man).

Glycogen Synthase Kinase-3

J R Woodgett, Samuel Lunenfeld Research Institute, Toronto, ON, Canada

© 2013 Elsevier Inc. All rights reserved.

Glossary

Adenomatous polyposis coli (APC) A large protein that is frequently mutated in colorectal cancer, which is a component of the β -catenin destruction complex.

Glycogen synthase (GS) The rate-limiting enzyme of glycogen deposition that is inhibited by phosphorylation.

Nuclear factor of activated T cells (NF-AT) A family of transcription factors first characterized in T lymphocytes as a critical regulator of cytokine expression, including interleukin-2.

Nuclear factor of κ B (NF- κ B) An important transcription that binds specific sequences in DNA and plays roles in regulating inflammatory responses and in deciding whether a cell will live or die following stress.

Protein kinase An enzyme that catalyzes the transfer of phosphate from adenosine triphosphate to the serine, threonine, or tyrosine residues.

Wnt An evolutionarily conserved family of extracellular proteins that regulate developmental processes by binding to a specific receptor (Frizzled) and initiating a cascade of events within cells that includes the accumulation of β -catenin.

GSK-3 Genes and Proteins

Mammalian genomes harbor two genes that encode for three variants of glycogen synthase kinase-3 (GSK-3) (Figure 1). The GSK-3 α gene is located on human chromosome 19q13.2 (mouse chromosome 7) and codes for a 51-kDa protein. The GSK-3 β gene is located on human chromosome 3q13.33 (mouse chromosome 16) and expresses a 47-kDa protein that is >95% identical to GSK-3 α in the kinase catalytic domain. Through differential splicing, the GSK-3 β gene also encodes a variant termed β' in which a 39-bp exon is included, adding a 13-amino-acid loop within the catalytic core. Both genes are widely expressed in all mammalian tissues and the crystal structure of GSK-3 β has been solved. GSK-3 α and GSK-3 β differ considerably in their noncatalytic regions (shown in green in Figure 1).

GSK-3 is an ancient and conserved protein, with clear homologs in all eukaryotes. Much of the functional knowledge of the enzyme has been derived from analysis in genetically tractable organisms such as fruit flies, slime mold, and yeast. With the recent characterization of mouse conventional and conditional knockouts of the two GSK-3 genes, the role of this protein kinase in mammals has been significantly advanced.

Regulation of GSK-3

Protein kinases are typically subject to intricate modes of regulation – in most cases, their activity is not expressed unless a specific signal is received by a cell. GSK-3 displays the rare behavior of being highly active in the absence of cellular stimulation. Thus, in resting cells, the protein phosphorylates a large catalog of substrate proteins. Upon stimulation by several classes of growth factors, hormones, and other agents, GSK-3 activity is repressed or diverted, relieving phosphorylation of its targets. In general, the functions of substrate proteins of GSK-3 are inhibited by phosphorylation. Hence, GSK-3 acts like a general cellular brake to suppress the activity of its

many targets (see Figure 2). This strategy allows cells to quickly respond to stimuli by activating specific cohorts of pre-existing proteins through deactivation of just one enzyme. This also raises the important question of signaling specificity. If multiple signals target the same regulatory molecule, how does a cell discriminate between signals? The answer to this conundrum lies in fractionation of GSK-3 into differentially controlled pools. Notably, however, genetic inactivation and small molecule inhibitors of this protein kinase short-circuit this organizational sequestration and cause indiscriminate inactivation.

There are two primary mechanisms by which GSK-3 activity is regulated. Both GSK-3 α and β are inhibited by phosphorylation of a serine residue located within the N-terminal regulatory domain (serine 21 in the α -isoform and serine 9 in β). Several protein kinases phosphorylate these regulatory residues including protein kinase B (PKB)/Akt, cyclic adenosine monophosphate (AMP)-dependent protein kinase, p70 S6 kinase, and several protein kinase C isoforms. As a consequence, a wide variety of stimuli result in GSK-3 inactivation including those that induce receptor-tyrosine kinases, such as the insulin receptor, which activate phosphatidylinositol 3' kinase (which activates PKB/Akt), and agonists of G-protein-coupled receptors that result in elevated cyclic AMP (Figure 2).

GSK-3 activity is also modulated by the Wnt signaling pathway by a completely distinct and insulated mechanism (Figure 3). Wnt-mediated inhibition of GSK-3 specifically targets a small and select fraction of GSK-3 molecules that are associated with the Axin/adenomatous polyposis coli (APC) destruction complex. In resting cells, this high-molecular-mass complex acts as an efficient machine to tag cytoplasmic β -catenin, an oncoprotein associated with colon cancer and other neoplasms, for proteosomal degradation. In the absence of a Wnt signal, cytoplasmic β -catenin molecules are captured by a scaffolding protein termed 'Axin1'. Axin1 also sequesters a portion of GSK-3 molecules within the cell – likely only 5–10% of the total GSK-3. Although GSK-3 cannot phosphorylate soluble β -catenin, when both molecules are coordinately bound to Axin1, GSK-3 efficiently targets three residues (Ser33,

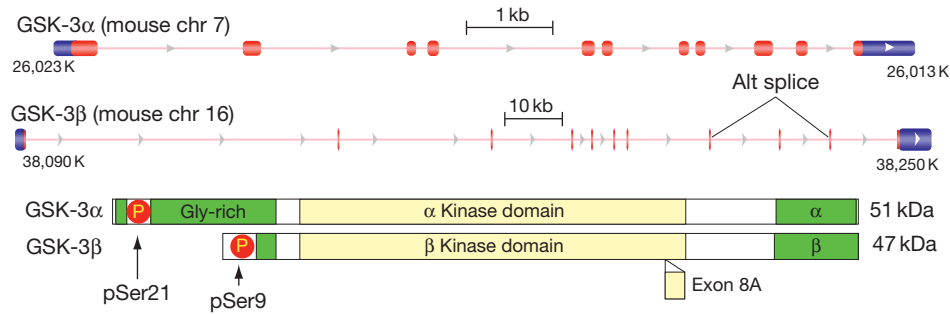


Figure 1 Domain and intron/exon structure of the two mouse GSK-3 genes. The regulatory phosphorylation sites are indicated. Variable regions are shown in green and the kinase domain in yellow.

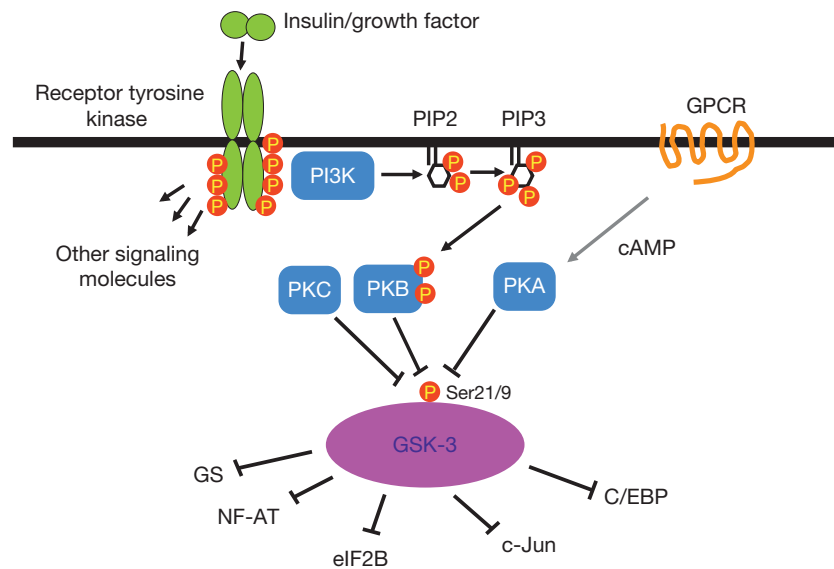


Figure 2 Simplified map of the regulation of GSK-3 by growth factors and mitogens. Several GSK-3 target proteins are indicated to represent the different classes of molecules regulated by GSK-3.

Ser37, and Thr41) within the amino-terminal domain of β -catenin. This phosphorylation is ordered (41 to 37 to 33) and requires phosphorylation of Ser45 by casein kinase 1, although the level of phosphorylated Ser45 does not appear to be regulated and is not limiting. The N-terminal pair of these phosphorylated residues (Ser33 and Ser37) form a recognition motif for specific E3 ligases such as β -TrCP that conjugate ubiquitin to β -catenin, which signals its rapid destruction by the 26S proteasome.

In the presence of certain members of the Wnt family (which comprises 15 secreted glycoproteins), a signaling cascade is initiated through engagement of the Frizzled class of receptors and the low-density lipoprotein receptor-related protein-5/6 (LRP5/6) coreceptors. Ligand binding recruits an adaptor molecule termed 'dishevelled' (there are three mammalian dishevelled proteins) that promotes re-organization of the Axin/APC destruction complex, whereby LRP5/6 becomes phosphorylated by GSK-3 and Casein kinase-1, which creates a high-affinity binding site for Axin. This disrupts the functioning of the Axin/APC complex allowing β -catenin to escape phosphorylation by GSK-3. As a consequence, β -catenin

accumulates within the cytoplasm, translocates to the nucleus, binds to the T-cell transcription factor (TCF)/lymphocyte enhancer-binding factor (LEF) family of DNA-binding proteins and modulates expression of genes such as *cyclin D*, *c-myc*, and *axin2*. The protein product of this latter gene facilitates reassembly of the APC destruction complex and thereby shuts down Wnt signaling by restoring degradation of β -catenin. Mutations in APC or the phosphorylation domain of β -catenin result in highly elevated levels of β -catenin within cells and loss of growth control. Indeed, mutations in one of these two molecules are found in the majority of colon cancers.

Although two distinct pathways converge to inhibit the same signaling molecule, GSK-3, the cellular consequences of inactivation are specific to the signal. Thus, inactivation of GSK-3 via N-terminal phosphorylation of serine 21/9 does not result in stabilization of β -catenin; nor does Wnt signaling impact metabolic targets that are regulated by insulin or growth factors. The precise mechanisms underlying the specificity determinants remain unclear but are likely related to sequestration of a subset of GSK-3 molecules in the APC/Axin complex, effectively shielding this subset from external

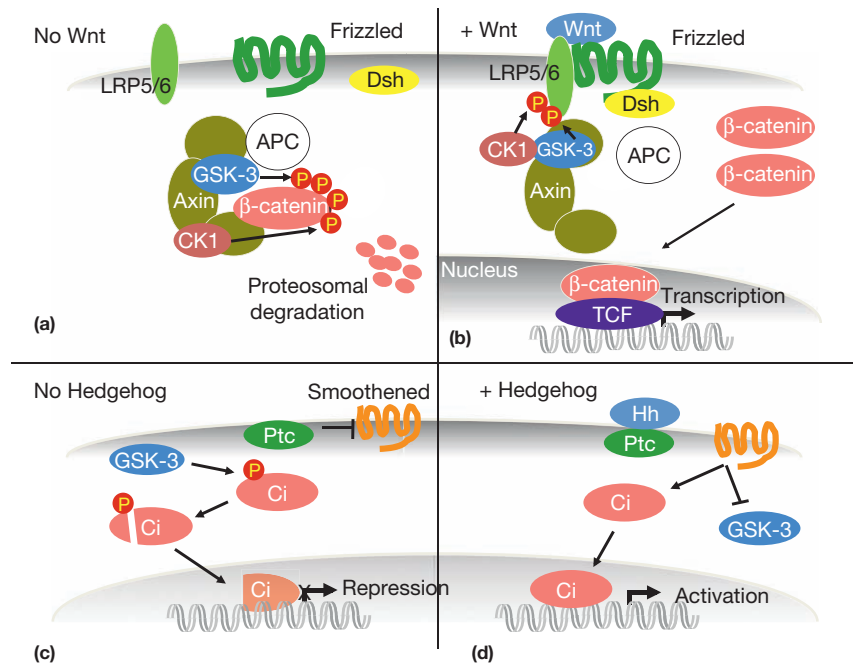


Figure 3 Regulation of GSK-3 by the Wnt and Hedgehog signaling proteins. The resting and stimulated states of these pathways are indicated for comparison.

regulation. Moreover, N-terminal phosphorylation of GSK-3 results in only partial inhibition (~50%) which is likely insufficient to impact the highly efficient phosphorylation/degradation of β -catenin, but has significant mass action consequences in metabolism.

Both isoforms of GSK-3 are also phosphorylated on a tyrosine residue within the P-loop of the catalytic center. Although phosphorylation of this site is required for full activity, the level of phosphorylation does not appear to be regulated (unlike, e.g., the analogous tyrosine residues in mitogen-activated protein (MAP) kinases). A protein-tyrosine kinase has been identified (Zak1) that targets the tyrosine in slime mold. However, in mammals, the residue is modified by autophosphorylation during the initial translational process, similar to a process initially observed in dual specificity tyrosine phosphorylation regulated kinase (DYRK) protein kinases.

As an aside, although Wnt signaling results in a decrease of GSK-3 phosphorylation of β -catenin and this underpins activation of this pathway, Wnt actually increases GSK-3 phosphorylation of LRP5/6. Hence, Wnt does not inactivate GSK-3 *per se*, rather, it uses a bait-and-switch mechanism to divert the action of GSK-3 away from β -catenin.

A Primer on GSK-3 Substrates

Many (but not all) of the proteins that are phosphorylated by GSK-3 require prior phosphorylation by a distinct protein kinase (Table 1). The primary substrate recognition sequence of GSK-3 is Ser/Thr-X-X-Ser/Thr-phosphate, where X is any amino acid but with a preference for proline. In theory, this prerequisite for prior phosphorylation provides an additional

Table 1 Known GSK-3 substrate proteins classified by function

<i>Metabolism</i>	<i>Structural</i>	<i>Transcriptional</i>
Amyloid precursor protein	Adenomatous polyposis coli	cJun (AP-1)
ATP-citrate lyase	CRMP2/4	β -Catenin
Axin	Dynamin-like protein	C/EBP
CDGAP	Dystrophin	CREB
Cubitus interruptus	Kinesin Light Chain	DF3/MUC
Cyclin D1, E	MAP 1B, 2, 2C	GATA4
eIF2B	N-CAM	Glucocorticoid receptor
Focal adhesion kinase	Neurofilaments	HIF-1
Glycogen synthase	Ninein	HSF-1
hnRNP	Paxillin	MITF
Inhibitor-2	Stathmin	c-Myc
Insulin receptor substrate 1	Synphilin-1	NF-AT
Mcl-1	Tau	NF-kB
NGF receptor	Telokin	Notch
P21 WAF/CIP		P53
Presenilin 1		Snail
R11 subunit of PKA		TCF
Tuberous sclerosis 2		vHL

regulatory step but, in practice, many of the primed phosphorylation sites are unregulated and fully phosphorylated. Instead, the priming mechanism appears to allow a distinct form of regulatory control at the level of GSK-3 itself. The crystal structure of GSK-3 β revealed the presence of several positively charged residues including a critical arginine (Arg 96) within the substrate-binding domain in a location that would permit

interaction with the priming phosphorylated amino acid in the incoming substrate. In other words, the primed phospho-amino acid is an important contributor to substrate binding. However, when GSK-3 is phosphorylated within its N-terminal domain (serine 21 or 9 in GSK-3 α or β , respectively), this phosphorylated serine competes for the primed phosphorylation-site-docking motif and when bound sterically occludes the active site. Hence, the protein kinase activity is effectively inhibited.

Substrates that require priming include glycogen synthase, cyclic AMP response element-binding protein (CREB), inhibitor-2, cyclin E, Tau, and β -catenin. Enzymes that catalyze priming of GSK-3 substrates include casein kinases 1 and 2, DYRK, and cyclic-AMP-dependent protein kinase. As can be seen from (Table 1), many GSK-3 substrates are transcriptional regulators. These include nuclear factor of activated T cells (NF-AT), a key regulator of cytokine expression. GSK-3 phosphorylation of NF-AT causes its transport from the nucleus, effectively inactivating the protein. This is reversed by the phosphatase, calcineurin, which is the target of the immunosuppressive agent, cyclosporine. GSK-3 is also implicated in other immune responses such that small molecule GSK-3 inhibitors can suppress chronic inflammation in several animal models.

Physiological Functions of GSK-3

GSK-3 has been implicated in a wide variety of cellular processes and pathologies. The first genetic analysis of the protein kinase occurred in the fruit fly where mutants of the single *Drosophila melanogaster* GSK-3 gene, termed 'Zeste-White3' or 'Shaggy', were found to display a phenotype similar to continuous signaling of the Wingless pathway, the fly counterpart of the mammalian Wnt signaling system. Wnts are a family of glycoprotein ligands that signal through a canonical pathway that acts on β -catenin or through a noncanonical pathway that involves calcium and guanosine triphosphatase (GTPase) signals. The canonical pathway plays an important role in controlling the activity of the β -catenin proto-oncogene that has been found to be mutationally activated in several cancers. Since GSK-3 negatively regulates β -catenin (in addition to other onco-proteins such as the c-Jun and c-Myc transcription factors and cyclin D, a CDK activator), it is a candidate tumor-suppressor gene. However, mutations in GSK-3 have not been found in human cancers. The most likely reason is that the two mammalian isoforms of GSK-3 are encoded by distinct genes and many (but not all) of their functions are redundant. Thus, both alleles of both genes would have to be inactivated to effect complete disruption of function – an unlikely event given that loss of the first would confer no selective advantage. A second possibility is that mutational inactivation of GSK-3 is incompatible with basic cellular properties that are requisite for tumor growth, such as proliferation. However, embryonic stem cells engineered to lack both copies of both GSK-3 genes are viable and proliferate normally.

A developmental pathway that often works in concert with the Wnt pathway is the Hedgehog pathway. Hedgehog, like Wnt, is a secreted factor but instead of targeting β -catenin, Hedgehog signaling prevents the cleavage of a protein termed Gli in mammals and Cubitus interruptus (Ci) in

flies (Figure 3). In resting cells, Ci is processed into two fragments, one of which acts as a transcriptional repressor. This cleavage requires phosphorylation of Ci by GSK-3. Upon binding of Hedgehog to its receptor, (patched) Gli cleavage is inhibited. The pathway, thus, has several parallels to Wnt signaling, although it is unclear whether GSK-3 is directly regulated by Hedgehog signaling or is, instead, a passive player that maintains low amounts of the uncleaved Ci protein under resting conditions. GSK-3 has also been implicated in Notch signaling since inactivation of the kinase results in increased amounts of the intracellular domain of Notch following Notch ligand engagement (Delta, Jagged).

Neurological Diseases

Dysregulation of GSK-3 has been implicated in neurological diseases such as Alzheimer's disease and bipolar disorder. This protein kinase is one of several that targets phosphorylation sites on the microtubule-associated protein Tau that are found to be hyperphosphorylated in insoluble structures termed 'neurofibrillary tangles'. These deposits are found in the brains of patients with Alzheimer's disease and are associated with neuronal death. Transgenic mice expressing a mutant of GSK-3 β that cannot be inhibited by N-terminal phosphorylation demonstrate higher than normal Tau phosphorylation. Activation of GSK-3 in neuronal cells has also been correlated with induction of apoptosis such that death induced by serum deprivation can be suppressed by GSK-3 inhibition.

GSK-3 is inhibited by lithium ions at concentrations (1 mM) that are therapeutically effective for the treatment of bipolar disorder. GSK-3 has also been implicated in schizophrenia. Variants in disrupted in schizophrenia 1 (DISC1) gene were identified in a Scottish family with a high incidence of this disorder. GSK-3 binds to DISC1 and this affects the functions of each protein. For example, inhibition of GSK-3 alleviates the behavioral changes associated with DISC1 mouse mutants. Inhibitors of GSK-3, therefore, have potential therapeutic value in the control of at least three neurological disorders.

Metabolism and Diabetes

Although the initial identification of GSK-3 was rooted in the study of glycogen metabolism, study of this kinase has largely focused on its roles in other processes. GSK-3 is the predominant regulator of glycogen synthase inhibition. This role has been reinforced by studies assessing the effects of GSK-3 inhibitors on glucose uptake and metabolism in response to insulin. Insulin acts to activate phosphatidylinositol 3' kinase and PKB leading to phosphorylation and inactivation of GSK-3 (Figure 2). Inhibitors of this protein kinase may, therefore, be predicted to act as insulin sensitizers. Such drugs are highly sought in the treatment of type 2 diabetes, a growing health problem as populations increase their degree of obesity and age. Small molecule inhibitors of GSK-3 have been shown to act as effective insulin sensitizers in several animal models of diabetes, reducing circulating blood glucose and enhancing glycogen deposition. In addition to negatively regulating

glycogen synthase, GSK-3 also phosphorylates and inhibits other proteins important to insulin's mechanism of action including IRS1, ATP-citrate lyase, and eIF2B. Thus, GSK-3 inhibition promotes anabolic processes such as fatty acid and protein synthesis.

One potential spoiler to the use of GSK-3 inhibitors as insulin sensitizers is the concern that chronic inhibition of this protein kinase may lead to inappropriate accumulation of β -catenin, a potential oncogene. However, cells appear to have a built-in fail-safe mechanism that normally prevents unscheduled induction of β -catenin. For example, loss of 75% of total GSK-3 activity results in virtually no increase in noncadherin-associated β -catenin. Inhibition of greater than 75% of GSK-3 activity is required for substantial stabilization in the absence of a specific signal (one of the Wnts). The reason for this is at least partially related to the relative levels of the Wnt signaling components compared to GSK-3. Axin is present at significantly lower concentrations than GSK-3 ($\alpha + \beta$). Hence, only when total GSK-3 levels decrease below that of Axin will processing of β -catenin be affected in the absence of a Wnt signal. Since most therapeutics are used at concentrations that dampen rather than eradicate activity, a therapeutic window may exist that allows tempering of aberrant GSK-3 activity without impacting stabilization of β -catenin. However, this Damocles sword will perch over the use of GSK-3 inhibitors until at least preclinical trials demonstrate no increase in tumor incidence upon chronic administration of the drugs. Given rapid progress in the development of potent and specific inhibitors of GSK-3, of which some 5000 now exist, an answer to this question should be forthcoming.

Knockout Mouse Models

Mice lacking GSK-3 β are susceptible to embryonic death due to liver failure. The defect was traced to aberrant induction of the NF- κ B transcription factor by tumor necrosis factor- α (TNF- α) in the liver that leads to hepatocyte cell death. TNF- α induced by opportunistic infection of the mother typically induces both a proapoptotic and an NF- κ B-mediated survival signal within cells. However, in the cells lacking GSK-3 β , induction of survival genes by NF- κ B was found to be suppressed, leading to dominance of the proapoptotic signal and the demise of the hepatocytes. The mechanism by which GSK-3 β influences NF- κ B is unclear, although there is evidence of both positive and negative regulation. Chemical inhibition of GSK-3 also leads to sensitization of cells to TNF- α . Of note, although the embryonic lethality of the mice lacking GSK-3 β can be rescued either by injection of anti-TNF- α antibodies or by breeding with mice that lack the 55-kDa receptor for TNF- α , the mice still die shortly after birth due to a patterning defect in the heart that resembles a human condition termed 'double-outlet right ventricle (DORV)' in which pulmonary circulation is ineffective leading to an inability to oxygenate blood via the lungs. The behaviors of GSK-3 β heterozygous mice are similar to those treated with lithium. The drug treatment leads to ~25% inhibition of total GSK-3, which is comparable to

the levels of the total kinase in the GSK-3 β heterozygous animals (since they retain GSK-3 α).

Mice lacking GSK-3 α are viable but display increased insulin sensitivity as well as behavioral differences and susceptibility to cardiomyopathy when stressed. Nonetheless, these animals exhibit a normal life span under typical conditions.

To allow examination of the role of GSK-3 β in adult animals, conditional alleles have been generated and tissue-specific inactivation of GSK-3 β (and GSK-3 α) achieved by crossing with selective Cre recombinase deleter strains. Through this methodology, selective inactivation of GSK-3 β in skeletal muscle was found to sensitize animals to insulin. Inactivation of GSK-3 β in insulin-producing β -islet cells of the pancreas has been reported to prevent insulin resistance caused by inactivation of IRS2 or haploinsufficiency of the insulin receptor. These results suggest tissue-dependent roles of the two GSK-3 isoforms, although this property is unlikely to be exploitable given that the only means to selectively inactivate either isoform is via RNA interference.

Life without GSK-3

Genetic inactivation of all four alleles of GSK-3 in embryonic stem cells results in maintenance of expression of pluripotency factors such as Nanog and Oct4 and suppression of differentiation potential. Conditional inactivation of total GSK-3 during brain development results in continued proliferation of neuronal precursors and radial glial cells and inhibits differentiation into functional neurons. These cells exhibit elevated cMyc, β -catenin, Notch, and Hedgehog signaling as might be predicted from the multiple suppressive roles of this protein kinase. These findings raise the possibility that reversible inactivation of GSK-3 may allow transient expansion of progenitor cell populations for regenerative medicine purposes.

See also: [Metabolism Vitamins and Hormones: Diabetes; Glycogen Metabolism; Glycogen Storage Diseases.](#)

Further Reading

- Beurel E, Michalek SM, and Jope RS (2010) Innate and adaptive immune responses regulated by glycogen synthase kinase-3 (GSK3). *Trends in Immunology* 31: 24–31.
- Dajani R, Fraser E, Roe SM, et al. (2001) Crystal structure of glycogen synthase kinase 3 beta: Structural basis for phosphate-primed substrate specificity and autoinhibition. *Cell* 105: 721–732.
- Doble BW, Patel S, Wood GA, Kockeritz LK, and Woodgett JR (2007) Functional redundancy of GSK-3 α and GSK-3 β in Wnt/ β -catenin signaling shown by using an allelic series of embryonic stem cell lines. *Developmental Cell* 12: 957–971.
- Hernández F, Gómez de Barreda E, Fuster-Matanzo A, Lucas JJ, and Avila J (2010) GSK3: A possible link between beta amyloid peptide and tau protein. *Experimental Neurology* 223: 322–325.
- Kim WY, Wang X, Wu Y, et al. (2009) GSK-3 is a master regulator of neural progenitor homeostasis. *Nature Neuroscience* 12: 1390–1397.
- MacDonald BT, Tamai K, and He X (2009) Wnt/ β -catenin signaling: Components, mechanisms, and diseases. *Developmental Cell* 17: 9–26.
- Wu D and Pan W (2010) GSK3: A multifaceted kinase in Wnt signaling. *Trends in Biochemical Sciences* 35: 161–168.

Glycolipid-Dependent Adhesion Processes

S Hakomori, University of Washington, Seattle, WA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a reproduced from the previous edition, volume 2, pp. 261–265, © 2004, Elsevier Inc.

Glossary

Glycosphingolipid Carbohydrate glycosidically bound to primary hydroxyl group of *N*-fatty acyl sphingosine.

Glycosynapse The assembly of glycosphingolipids and other glycoconjugates in membrane, associated with cytoplasmic signal transducers having long-chain acyl group, and with hydrophobic transmembrane proteins.

The assembly performs carbohydrate-dependent cell adhesion with concurrent signaling.

Glycosphingolipid structures Gb4, GalNAc β 3Gal α 4Gal β 4Glc β 1Cer; Gb5, Gal β 3GalNAc β 3Gal α 4Gal β 4Glc β 1Cer; Gg3, GalNAc β 4Gal β 4Glc β 1Cer; GM3, NeuAc α 3Gal β 4GlcCer; LacCer, Gal β 4Glc β 1Cer; LacNAc, Gal β 4GlcNAc β →R; Lc $_4$, Gal β 3GlcNAc β 3Gal β 4GlcCer; Le x , Gal β 4[Fuc α 3]GlcNAc β 3Gal β →R; nLc $_4$, Gal β 4GlcNAc β 3Gal β 4Glc β 1Cer.

GSL Assembly in Membrane as GSL-Enriched Microdomain

Cluster Formation by GSL-to-GSL *cis*-Interaction

Ceramide, the lipid moiety of glycosphingolipids (GSLs), has aminoacyl (*N*-acyl) residue and hydroxyl group that act as hydrogen bond donors, whereas glycerides and glycerophospholipids have only *O*-acyl group that act as hydrogen bond acceptor. In addition, GSLs have a number of hydroxyl groups and *N*-acyl residues at the sugar moiety, that may provide a much stronger hydrogen bond donor (Figure 1(a)). At the plane of lipid bilayer, the presence of sphingolipids and GSLs therefore provides higher and stronger *cis*-interaction (i.e., GSL-to-GSL side-by-side interaction). Glycerides and glycerophospholipids in the same lipid bilayer plane, although they constitute the major lipid components, show much weaker *cis*-interaction. By a similar mechanism, glycoproteins show strong *cis*-interaction, possibly through their hydrogen bonds, to form clusters separable from GSL clusters. Thus, GSLs and glycoproteins are clustered to form independent patches or islands separated from the sea of glycerophospholipids on the cell surface, as shown in Figure 1(b). Sphingomyelin interacts with cholesterol but is separated from glycerophospholipids in gel state, as evidenced by electron spin resonance and infrared spectroscopy. GSL clusters occur in the absence of cholesterol.

Association of Clustered GSLs with Lipophilic Signal Transducers and Proteolipids: Assembly of Glycosynapse

Clustered GSLs in cell membranes can be separated as insoluble low-density membrane by homogenization in high-salt, or in lysis buffer containing neutral detergent, followed by density-gradient ultracentrifugation. The GSL-enriched buoyant fraction contains >95% of GSLs originally present in cells, but <1% of total cellular protein. The proteins found in this fraction, from many cell types, are cSrc, other Src family kinases, and small G proteins (Ras, RhoA, RAC), which contain

long-chain aliphatic acyl groups and are closely associated with GSLs. In addition, the fraction has proteolipid proteins and tetraspanins, which are also closely associated with GSLs and signal transducers. Molecular assembly of these components in microdomain forms a functional unit in membrane essential for cell adhesion with concurrent signal transduction; the unit is collectively termed as 'glycosynapse'.

Cell Adhesion through Clustered GSLs

Two types of GSL-dependent cell adhesion processes can be considered.

1. GSL cluster is recognized by or bound to complementary GSL cluster expressed on counterpart cell-surface glycosynapse (i.e., GSL-to-GSL *trans*-interaction). Such interaction takes place through specific complementary structures. The binding between carbohydrate moieties of GSLs is catalyzed, in many cases, by Ca²⁺. There is a steadily increasing number of examples of such complementary GSL pairs, providing the basis of GSL-dependent adhesion. A model of such interaction involving single molecules of Gb4 and nLc $_4$, as a basis of cell adhesion, is shown in Figures 2(a)–2(c). In reality, however, GSLs are not present as single molecules, but are clustered on membrane. Cell adhesion occurs through such GSL clusters in association with signal transducers (glycosynapse), to activate or inhibit signaling. A model is shown in Figure 3.
2. GSL cluster, particularly ganglioside cluster, is bound to ganglioside-binding lectin (siglec) such as siglec-4 or -7. Alternatively, GSL having β Gal residue (e.g., GM1) is recognized by galectin, or GSL having SLe x or SLe a residue is recognized by E-selectin.

Processes 1 and 2 are shown schematically in Figure 4. In either process, structure of GSLs is the target of cell adhesion, recognized by complementary GSL (Figure 4(a)) or by specific binding protein (Figure 4(b)), which triggers activation or inhibition of signal transducers.

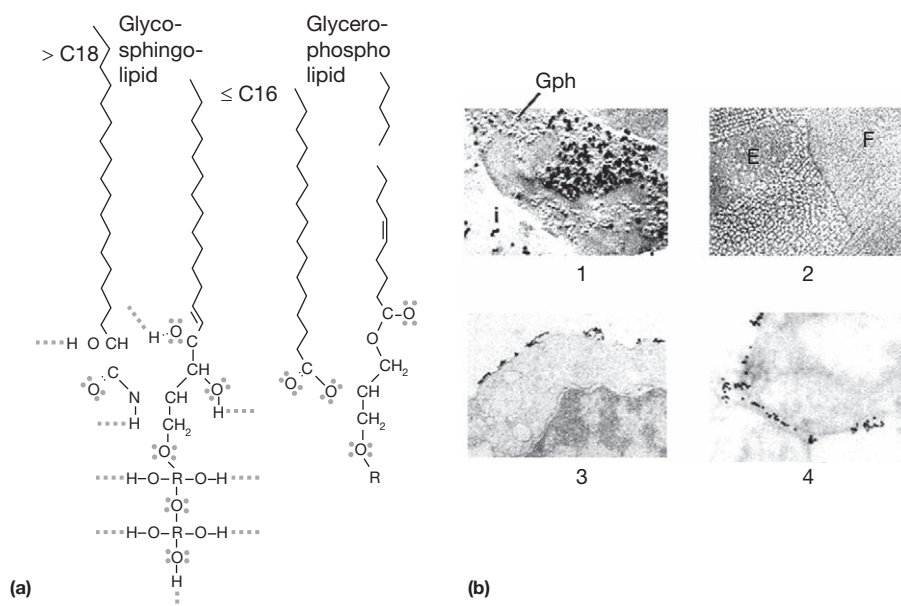


Figure 1 Clustering of glycosphingolipids and glycoproteins on membrane due to *cis*-interaction. (a) Contrasting properties of GSLs and glycerophospholipids in ability for hydrogen bond formation. GSLs are both hydrogen bond donors and acceptors (indicated respectively by $\cdots\text{H}-\text{O}$ and $\text{O}=\text{C}$), whereas glycerophospholipids only accept hydrogen bond, due to absence of hydroxyl group (OH). (b) Clustering is visualized by scanning electron microscopy following freeze-fracture technique: glycoprotein (Gph) clustering, and globoside (Gb4) clustering on erythrocyte membrane (in 1), GM1 clustering on dipalmitoyl phosphatidylcholine liposome surface (in 2). Cell-surface GSL clustering is also observed by transmission electron microscopy using specific antibody-gold sol conjugates: GM3 clustering on T-lymphocyte surface (in 3), and polysialoganglioside clustering in neuronal cells (in 4).

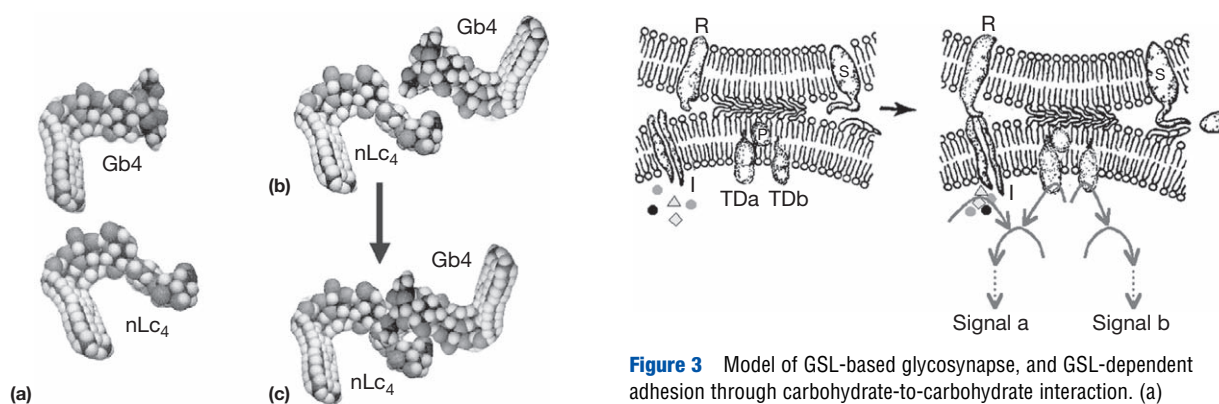


Figure 2 *trans*-Interaction of GSLs as a basis of glycolipid-dependent cell adhesion. (a) Minimum-energy conformational model of Gb4 and nLc₄. Note that axis of carbohydrate is perpendicular to that of sphingosine and fatty acid (ceramide). (b) Gb4 on the surface of one cell and nLc₄ on the surface of counterpart cell have complementary surface profile. (c) Gb4 and nLc₄ adhere to each other, by carbohydrate-to-carbohydrate interaction, on the basis of cell adhesion.

Figure 3 Model of GSL-based glycosynapse, and GSL-dependent adhesion through carbohydrate-to-carbohydrate interaction. (a) Clustered GSLs on the surface membrane of one cell are bound to a cluster of another type of GSL on surface membrane of counterpart cell, through *trans*-interaction of GSLs as in Figure 2. (b) Such adhesion induces change (activation or inhibition) of signal transducers (TDa and TDb) to send signals (a and b). In some cases, this process is synergistic with another adhesion system, such as integrin receptor (I) and its binding counterpart (R, e.g., ICAM1), or other type of carbohydrate-dependent receptor (S, e.g., CD44).

Types of Cell Adhesion Mediated by GSL-to-GSL Interaction

Mouse Melanoma B16 Cell Adhesion to Mouse Endothelial Cells through GM3-to-LacCer or GM3-to-Gg3Cer Interaction

B16 cells are characterized by predominance of GM3 ganglioside. Their adhesion to mouse or human endothelial cells

(ECs) is based on interaction of GM3 (on melanoma cells) with LacCer or Gg3Cer (on ECs). LacCer is present on essentially all types of ECs, whereas Gg3Cer has a more restricted distribution on lung microvascular ECs, where metastasis may occur preferentially. The degree of this GM3-dependent adhesion is closely correlated with the metastatic and invasive

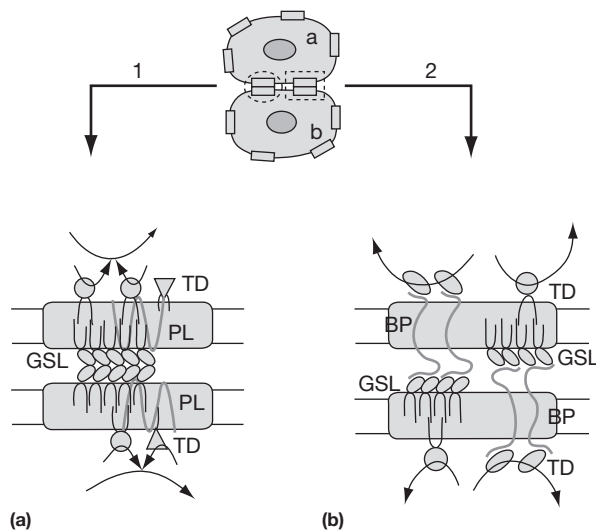


Figure 4 Two types of GSL clusters in glycosynapse function in GSL-dependent adhesion. Adhesion of cell a to cell b occurs via adhesion processes 1 and 2. (a) Process 1 is based on GSL clusters associated directly with signal transducers (TDs), and stabilized by proteolipids (PLs). Adhesion proceeds through interfacing of the two glycosynapses, and carbohydrate-to-carbohydrate interaction, and concurrent activation or inhibition of signal transducers. (b) Process 2 is based on GSL clusters and their binding protein (BP). Through interfacing of the two glycosynapses, GSLs are bound to BP, whereby TD is activated or inhibited.

potential of B16 cells, based on the following observations: (1) *in vitro* GM3-dependent adhesion was inhibited by liposomes containing GM3, LacCer, or Gg3Cer; (2) *in vivo* metastatic potential of B16 cells in terms of lung colonization was inhibited by administration of GM3- or Gg3Cer-liposomes; and (3) enhancement of B16 cell motility and invasiveness following GM3-dependent adhesion was indicated by activation of cSrc, RhoA, and FAK at glycosynapse. GM3-to-Gg3Cer and -LacCer interactions were well studied recently by surface plasmon resonance spectroscopy.

F9 Cell Autoaggregation through Le^x-to-Le^x Interaction

Mouse embryonal carcinoma F9 cells show strong autoaggregation, simulating morula-stage preimplantation embryo. Both F9 cell autoaggregation and compaction of morula are inhibited by a trivalent Le^x-lysyllysine conjugate. The molecular mechanism of Le^x-dependent autoaggregation was found to be Le^x-to-Le^x interaction in the presence of Ca²⁺. The process was recently well elucidated by molecular force microscopy and by surface plasmon resonance spectroscopy.

Human Embryonal Carcinoma Autoaggregation through Gb4-to-nLc₄ or Gb4-to-Gb5 Interaction

Gb4, Gb5, and nLc₄ are highly expressed, in addition to Le^x, in mouse preimplantation embryo. Expression of Gb5 is much higher than that of Le^x in human embryonal carcinoma and in primate preimplantation embryo, although the expression

status of these GSLs and carbohydrate antigens in human embryo has not been studied for ethical reasons. Autoaggregation of human embryonal carcinoma, which may mimic the compaction process of human embryo, is mediated by Gb4-to-nLc₄ or Gb4-to-Gb5 interaction (Gb4-dependent adhesion). These GSLs are closely associated with cSrc, RhoA, and Ras in glycosynapse, and may induce differentiation following Gb4-dependent adhesion.

Rainbow Trout Sperm Binding to Egg Mediated by (KDN)GM3 Interaction with Gg3-Like Epitope

A novel deaminated analog of sialic acid, 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN), is found in fish egg and sperm. The head of rainbow trout sperm has a high content of (KDN)GM3, as indicated by reactivity with specific anti-(KDN)GM3 mAb kdn3G, and binds strongly to plates coated with Gg3, but not other GSLs. This binding is inhibited by kdn3G and by anti-Gg3 mAb 2D4. Vitelline envelope membrane isolated from rainbow trout egg has a glycoconjugate containing Gg3-like epitope, defined by antibody 2D4. Immunoelectron microscopy revealed that the Gg3-like epitope is highly expressed in exposed second layer of vitelline envelope surrounding the micropyle, a narrow channel through which a sperm enters the egg cell. Thus, interaction of (KDN)GM3 at sperm head with Gg3-like epitope surrounding the micropyle is a crucial event in the process of rainbow trout fertilization.

Types of Cell Adhesion Mediated by Interaction of GSL with GSL-Binding Protein

Neuronal Axon Binding to Myelin through Interaction of Ganglioside with Siglec-4

Neuronal cell membranes are characterized by the presence of ganglio-series gangliosides. Gangliosides expressed at neuronal cell axons interact with myelinating glial cells through myelin-associated glycoprotein (MAG), which was known previously as sialic acid-binding protein siglec4. GD1a and GT1b, which have sialyl $\alpha 2 \rightarrow 3$ Gal terminal residue, are specific functional ligands of MAG that inhibit nerve regeneration and neurite formation. Inhibition of ganglioside synthesis by P4, or blocking of ganglioside interaction with MAG by anti-ganglioside IgG antibodies, reversed the MAG-dependent inhibitory effect on neurite formation.

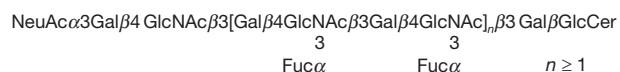
Aggregation of Renal Cell Carcinoma TOS1 Cells with Peripheral Blood Mononuclear Cells through Interaction of Disialoganglioside with Siglec-7

Renal cell carcinoma (RCC) has high metastatic potential to lung, and the degree of this potential is correlated with RCC expression of disialogangliosides, which are ligands of sialic acid-binding protein siglec-7, the major siglec expressed in peripheral blood mononuclear cells (PBMCs). These disialogangliosides expressed in RCC were identified as disialyl-Lc₄ (defined by mAb FH9), GalNAc-disialyl-Lc₄ (defined by mAb RM2), disialyl-Gb5 (defined by mAb 5F3), and GD3 (defined

by mAb R24). RCC cell line TOS1 is characterized by expression of disialyl-Lc₄, GalNAc-disialyl-Lc₄, and GD3, and interacts with siglec-7 on the surface of PBMC, causing aggregation of TOS1 cells with PBMCs. Large aggregates of this type may result in microembolisms that initiate metastasis, particularly in lung.

Adhesion of Neutrophils or Myelogenous Leukemia HL60 Cells to E-Selectin through Fucosyl-Poly-LacNAc Ganglioside (Myeloglycan)

Sialyl-Le^x (SLe^x) has been regarded as the major ligand of human neutrophils, or myelocytic, or monocytic leukemia cell lines (e.g., HL60 and U937) that interact with E-selectin expressed on ECs. Extensive studies on human neutrophils and HL60 cells indicated that SLe^x epitope is absent from GSLs and gangliosides of these cells. Instead, E-selectin-binding epitopes were identified as poly-LacNAc with multiple $\alpha 1 \rightarrow 3$ fucosylation and terminal $\alpha 2 \rightarrow 3$ sialylation:



These GSLs display strong E-selectin-dependent adhesion, particularly under dynamic flow conditions, and are collectively termed 'myeloglycan'.

Human Neuroblastoma Adhesion through GM1-to-Galectin-1 Interaction

Human neuroblastoma expresses ganglio-series gangliosides and galectin-1. Neuroblastoma cell contact activates endogenous sialidase, which causes conversion of GD1a, GD1b, GT1b, etc., to GM1, and consequent accumulation of GM1. GM1 is one of the ganglio-series gangliosides having terminal

β Gal residue that is recognized by galectin-1; therefore, GM1 at the neuroblastoma cell surface promotes cell-to-cell adhesion. Galectin-mediated cell adhesion has been proposed, but the target carbohydrate was identified as glycoprotein, and little attention has been paid to GSLs.

See also: **Metabolism Vitamins and Hormones:** Glycolysis Overview; **Protein/Enzyme Structure Function and Degradation:** Lipid Modification of Proteins: Targeting to Membranes.

Further Reading

- Hakomori S (2000) Traveling for the glycosphingolipid path. *Glycoconjugate Journal* 17: 627–647.
- Hakomori S (2002) The glycosynapse. *Proceedings of the National Academy of Sciences of the United States of America* 99(1): 225–232.
- Hakomori S and Handa K (2002) Glycosphingolipid-dependent cross-talk between glycosynapses interfacing tumor cells with their host cells: Essential basis to define tumor malignancy. *FEBS Letters* 531(1): 88–92.
- Handa K, Stroud MR, and Hakomori S (1970) Sialosyl-fucosyl poly-LacNAc without the sialosyl-Le^x epitope as the physiological myeloid cell ligand in E-selectin-dependent adhesion: Studies under static and dynamic flow conditions. *Biochemistry* 36: 12412–12420.
- Ito A, Handa K, Withers DA, Satoh M, and Hakomori S (2001) Binding specificity of siglec7 to disialogangliosides of renal cell carcinoma: Possible role of disialogangliosides in tumor progression. *FEBS Letters* 504(1–2): 82–86.
- Kopitz J, von Reitzenstein C, Burchert M, Cantz M, and Gabius HJ (1998) Galectin-1 is a major receptor for ganglioside GM1, a product of the growth-controlling activity of a cell surface ganglioside sialidase, on human neuroblastoma cells in culture. *Journal of Biological Chemistry* 273: 11205–11211.
- Rojo J, Morales JC, and Penades S (2002) Carbohydrate–carbohydrate interactions in biological and model systems. *Topics in Current Chemistry* 218: 45–92.
- Vyas AA, Patel HV, Fromholt SE, et al. (2002) Gangliosides are functional nerve cell ligands for myelin-associated glycoprotein (MAG) an inhibitor of nerve regeneration. *Proceedings of the National Academy of Sciences of the United States of America* 99(12): 8412–8417.

Glycolysis Overview

R A Harris, Indiana University School of Medicine, Indianapolis, IN, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Cancer Unlike normal tissue or a benign tumor, cancer cells exhibit the property of invasion of other organs of the body.

Hypoxia Condition in which the oxygen content of a tissue is so low that it limits electron transport activity by mitochondria.

Na⁺,K⁺-ATPase Enzyme complex present in the plasma membrane of all cells that utilizes the energy provided by adenosine triphosphate (ATP) hydrolysis to drive the transport of Na⁺ out and K⁺ into the cells.

Oxidation–reduction reactions Reaction in which one compound gets oxidized (loses electrons) and another compound gets reduced (gains electrons).

PET imaging A technique for creating an image of a tissue that relies upon the use of positron emission tomography (PET) to measure a signal produced by a short-lived radionuclide such as ¹⁸F-fluorodeoxyglucose.

Reactive oxygen species Small highly reactive oxygen-containing compounds that include the superoxide anion (O₂[−]) and hydrogen peroxide (H₂O₂). At low concentrations, they play important roles in signal transduction. At high concentrations, they damage cells and induce a condition called oxidative stress.

Role of Glycolysis

Adenosine Triphosphate Production

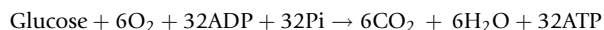
In the absence of either oxygen or mitochondria, glycolysis operates as a self-contained pathway with lactate as its end product (**Figure 1(a)**). Energy released by exergonic reactions is captured with the formation of two molecules of adenosine triphosphate (ATP) from two molecules of adenosine diphosphate (ADP) and inorganic phosphate (Pi):



Production of ATP by this overall sum equation for the pathway is the most important function of glycolysis. Some cells entirely depend on the ATP produced by glycolysis. For example, mature red blood cells obtain all of their ATP from glycolysis because they lack mitochondria, the oxygen-consuming organelles responsible for production of practically all of the ATP in most cells. Uninterrupted production of ATP is required for proper function and survival of all cells. The simplest of all cells in the body, red blood cells, still require the continuous operation of the glycolytic pathway to produce ATP to power their Na⁺,K⁺-ATPase and other energy-requiring processes.

Preparatory Pathway for Glucose Oxidation

When oxygen is available and mitochondria are present, glucose is completely oxidized to carbon dioxide and water (**Figure 1(b)**) by the sum reaction:



Under these conditions, glycolysis sets the stage for the complete oxidation of glucose, a process that produces 16 times more ATP per glucose molecule than anaerobic glycolysis. In the absence of either oxygen or mitochondria, glycolysis yields two molecules of pyruvate that must be reduced to lactate for glycolysis to proceed (**Figure 1(a)**). When oxygen

is present, pyruvate is taken up by mitochondria and the entire molecule is oxidized to carbon dioxide and water by the sequential actions of the pyruvate dehydrogenase complex, citric acid cycle, and the electron transport chain (**Figure 1(b)**).

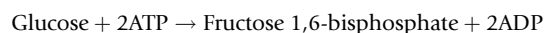
Because mitochondria are present in most cells and there is continuous supply of oxygen from the blood, virtually every tissue of the body uses glycolysis as a preparatory pathway, which provides them with much greater amounts of ATP from each glucose molecule than anaerobic glycolysis. The human brain, for example, completely oxidizes about 120 g of glucose each day, all via pyruvate molecules produced from glucose by glycolysis. Lactate is present in the brain, but the amount is small and none is released into the blood because the brain is well oxygenated and mitochondria are abundant.

Three Stages of the Glycolytic Pathway

Conversion of glucose to lactate takes place in three stages in the cell cytoplasm (**Figure 2**).

Priming Stage

The first three reactions of the pathway prime the glucose molecule for use by subsequent enzymes (see **Figure 2**). ATP molecules are invested in the reactions catalyzed by hexokinase and 6-phosphofructo-1-kinase. Phosphorylation traps glucose inside the cell and sets the stage for subsequent cleavage of the glucose into two three-carbon fragments. The priming stage reactions sum up to be simply



Splitting Stage

Fructose 1,6-bisphosphate undergoes cleavage by aldolase between carbons 3 and 4 in the first step of the splitting

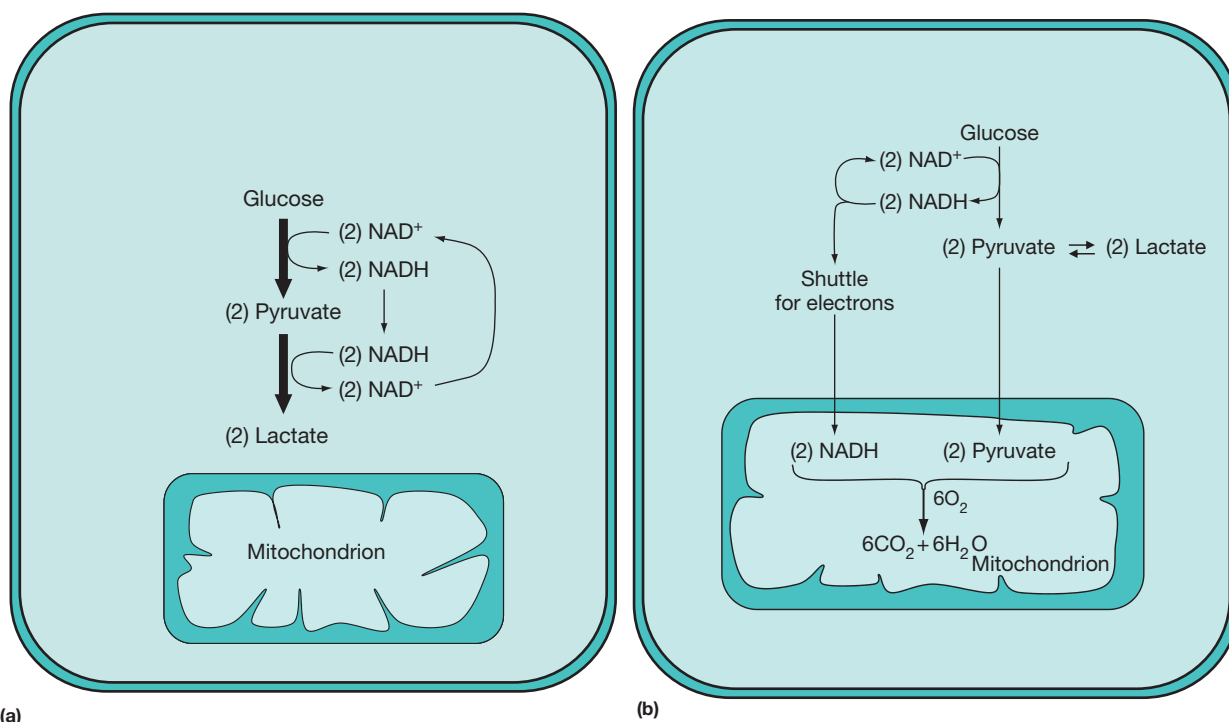
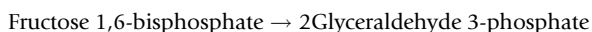


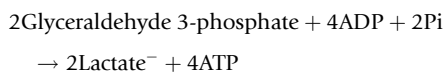
Figure 1 Overviews of glycolysis: (a) lactate is the product of glycolysis in the absence of oxygen and (b) glycolysis serves as a preparatory pathway for the complete oxidation of glucose in the presence of oxygen.

stage. Dihydroxyacetone phosphate and glyceraldehyde 3-phosphate, known collectively as triose phosphates, are produced by an aldol cleavage of fructose 1,6-bisphosphate. Inter-conversion of the triose phosphates by an isomerase completes the splitting stage, which sums up to be

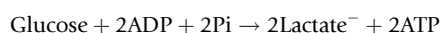


Energy-Trapping Stage

Six enzymes are required to capture the energy made available in this stage. All of the ATP produced by glycolysis is generated by the reactions catalyzed by 3-phosphoglycerate kinase and pyruvate kinase in this stage. Since two glyceraldehyde 3-phosphate molecules are converted to two lactate molecules, a total of four ATPs is produced in this stage, which sums up to be



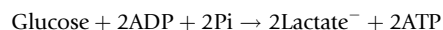
The sum of the three stages yields the overall equation for anaerobic glycolysis in which one molecule of glucose is converted to two molecules of lactate with the production of two molecules of ATP:



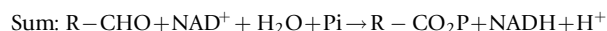
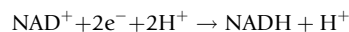
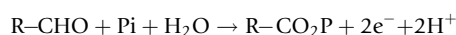
Anaerobic Glycolysis Is a Type of Fermentation

Fermentations are pathways that (1) produce ATP, (2) do not involve oxygen, (3) involve an intermediate that becomes

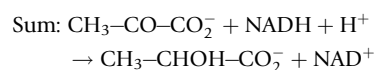
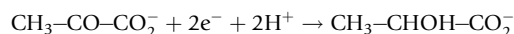
oxidized, and (4) involve an intermediate that becomes reduced. That ATP is produced and oxygen is not required can be seen from the overall balanced equation for glycolysis:



That an intermediate becomes oxidized and another reduced is not shown by this equation. However, during the first reaction of the energy-trapping stage of glycolysis (**Figure 2**), the aldehyde group ($-\text{CHO}$) of glyceraldehyde 3-phosphate is oxidized to a carboxylic acid ($-\text{CO}_2^-$) and then phosphorylated to give a mixed anhydride of a carboxylic acid and phosphoric acid (1,3-bisphosphoglycerate):



In the last reaction of the energy-trapping stage of glycolysis (**Figure 2**), the α -keto group of pyruvate is reduced to give the α -hydroxyl group of lactate:



NAD^+ serves as a bridge compound for these reactions (**Figure 3**), accepting two electrons and one H^+ to become NADH during the oxidation of the aldehyde group of glyceraldehyde 3-phosphate and subsequently donating two electrons and one H^+ when it

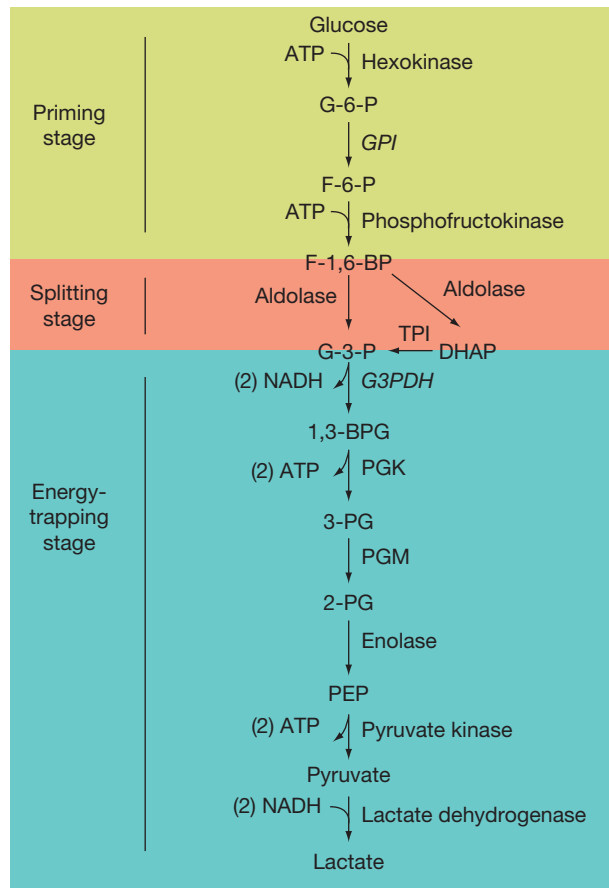


Figure 2 The three stages of the glycolytic pathway: priming, splitting, and energy trapping. Glycolytic intermediates: G-6-P, glucose 6-phosphate; F-6-P, fructose 6-phosphate; F-1,6-BP, fructose 1, 6-bisphosphate; G-3-P, glyceraldehyde 3-phosphate; DHAP, dihydroxyacetone phosphate; 1,3-BPG, 1,3-bisphosphoglycerate; 3-PG, 3-phosphoglycerate; 2-PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate. Glycolytic enzymes: GPI, glucose phosphate isomerase; TPI, triose phosphate isomerase; G3PDH, glyceraldehyde-3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; PGM, phosphoglycerate mutase.

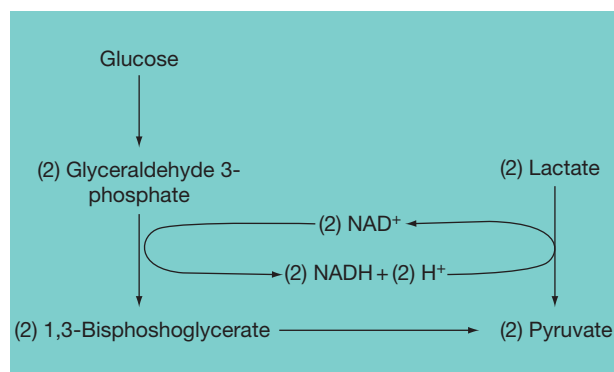


Figure 3 NAD^+ serves as a bridge molecule to couple the oxidation of glyceraldehyde 3-phosphate to the reduction of pyruvate during anaerobic glycolysis.

returns to being NAD^+ during the reduction of the α -keto group of pyruvate. The result is a perfect balance between the production of NADH in an oxidation step of glycolysis and its utilization in a reduction step of glycolysis. Anaerobic glycolysis is therefore a type of fermentation.

Control of Glycolysis

The rate at which ATP is hydrolyzed back to ADP and P_i by energy-requiring reactions is a major determinant of flux through the glycolytic pathway. In other words, ATP turnover caused by energy required to perform the work being done by a cell greatly affects the rate of glycolysis. The presence or absence of oxygen also has a profound effect upon the rate of glycolysis.

ATP Turnover

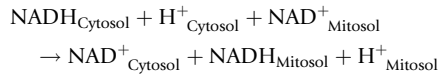
Two molecules of ATP are produced for each molecule of glucose utilized by glycolysis. However, for glycolysis to continue unabated the ATP produced by glycolysis must be continuously hydrolyzed back to ADP and P_i . This is because the amount of ADP present in the cells is quite small relative to the large amount of glucose that is utilized to meet the energy needs of cells. Glycolysis can therefore not proceed without hydrolyzing ATP to make ADP available for the steps catalyzed by phosphoglycerate kinase and pyruvate kinase. Hydrolysis of ATP to ADP plus P_i by energy-requiring processes is usually rate limiting for the process of glycolysis. Thus, an increase in ATP turnover caused by an increase in the amount of work being done by a cell has a profound effect upon the rate of glycolysis. For example, a marked increase in the rate of glycolysis occurs in response to strenuous exercise because of ATP hydrolysis by the contractile machinery of muscle.

Oxygen

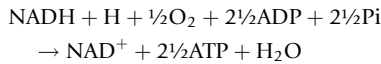
A remarkable decrease in the rate of glycolysis occurs when cells that lack oxygen are transferred into an environment in which oxygen is abundantly available. The production of carbon dioxide starts immediately but cells use much less glucose and accumulate little or no lactate. This illustrates the Pasteur effect, which refers back to Louis Pasteur's observation (150 years ago) that oxygen inhibits the fermentation of glucose by yeast. The Pasteur effect reflects the marked difference in cellular capacity for ATP production in the presence and absence of oxygen. Relative to anaerobic glycolysis, 16 times more ATP is produced when glucose is completely oxidized to carbon dioxide and water by the combined actions of glycolytic and mitochondrial enzymes. The rate of glycolysis decreases greatly in the presence of oxygen because the ATP required to meet the energy needs of a cell can be supplied by the catabolism of much less glucose.

Lactate is not normally produced in large amounts when oxygen is available to cells. Less glucose has to be used to meet the need for ATP, which automatically means less pyruvate is produced and available for reduction to lactate by NADH. The pyruvate produced by glycolysis is also rapidly oxidized by the mitochondria, leaving less available for reduction to lactate.

by NADH. Finally, less NADH is available for the reduction of pyruvate to lactate because shuttle mechanisms operate in the presence of oxygen to transport the electrons of NADH into mitochondria (Figure 1(b)) by the net equation:



NADH generated in the mitochondria by this shuttle mechanism is oxidized by the electron transport chain, which in turn produces the energy necessary for the synthesis of ATP by oxidative phosphorylation according to the net equation:



Removal of oxygen from the environment of cells has the opposite effect. Glycolysis is greatly stimulated as evidenced by increased glucose utilization and lactate production. Anaerobic glycolysis can sometimes meet the need of cells for ATP, but often it is inadequate, in spite of an increase in its rate. Cells die when ATP is not generated fast enough to keep them viable, which is what happens in the heart during a myocardial infarct or in the brain as a consequence of a stroke. Cells that survive in an oxygen-poor environment are able to elevate their capacity for glycolysis by increasing the amounts of the glycolytic enzymes present. This in turn increases the activities of the glycolytic enzymes and therefore the capacity for ATP synthesis by glycolysis. Lack of oxygen signals an increase in the transcription of the genes that encode the glycolytic enzymes. This occurs because lack of oxygen (hypoxia) causes an increase in the amount of a protein called hypoxia-inducible factor (HIF) (Figure 4). HIF is a transcription factor that promotes transcription when it binds to a specific sequence of nucleotides in the promoter region of genes encoding glycolytic enzymes. Oxygen stimulates proteolytic degradation of HIF (Figure 4), which keeps the amount of HIF low which, in turn, keeps the amounts and therefore the activities of glycolytic enzymes low in cells. Hypoxia maximizes the amount of HIF, which greatly increases expression of the glycolytic enzymes and therefore cellular capacity for glycolysis.

Cancer

Malignant tumors make good use of glycolysis as a source of ATP. Rates of glycolysis are often faster in malignant tumors than in normal tissues. Radiologists and oncologists take advantage of this phenomenon in the diagnosis of cancer by positron emission tomography (PET) imaging with a radioactive glucose derivative. Cancer cells often accumulate larger amounts of the radioactive glucose derivative because of their greater capacity for glucose uptake and glycolysis. Surprisingly, tumors often metabolize glucose to lactate even in the presence of oxygen. This is unusual and stands in contrast to normal cells that only metabolize glucose to lactate in the absence of oxygen, as discussed previously. The conversion of glucose to lactate in the presence of oxygen is called aerobic glycolysis, to distinguish it from anaerobic glycolysis which refers to the conversion of glucose to lactate in the absence of oxygen. The high rate of aerobic glycolysis exhibited by some cancer cells is

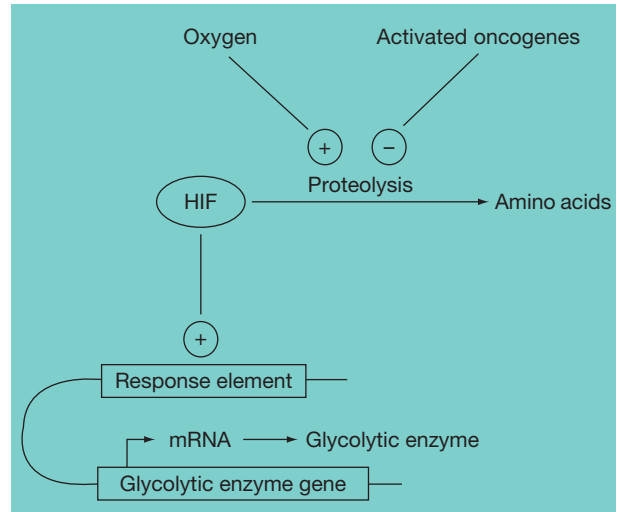


Figure 4 Oxygen controls glycolytic enzyme expression by controlling the amount of hypoxia inducible factor (HIF), a transcription factor that regulates the transcription of genes that encode glycolytic enzymes.

called the Warburg effect, in recognition of Otto Warburg's discovery (80 years ago) that tumor cells produce lactate in the presence of oxygen. Based on this observation, Warburg championed the idea that aerobic glycolysis is the cause of cancer. Now, however, it is appreciated that upregulation of the enzymatic capacity of glycolysis occurs frequently in cancer cells because of a higher concentration of HIF, the transcription factor responsible for regulation of the transcription of genes encoding glycolytic enzymes (Figure 4). HIF is highly expressed in many tumor cells, even in the presence of oxygen, because of the mutations that have occurred in genes that either prevent (tumor suppressors) or promote (oncogenes) the development of cancer. Since tumors sometimes outgrow their blood supply and cancer cells frequently migrate into areas of the body where the oxygen concentration is low, an exceptionally high capacity for ATP production that does not require oxygen may provide cancer cells with a metabolic advantage for survival and growth. Furthermore, mitochondrial respiration is associated with the production of reactive oxygen species that can damage cellular components, inhibit growth, and cause cell death. Since the low-oxygen environment that can occur in tumors can promote the formation of reactive oxygen species by the mitochondrial electron transport chain, the increased capacity for glycolysis and reduced reliance upon oxidative metabolism may serve to protect cancer cells from being killed by increased levels of reactive oxygen species.

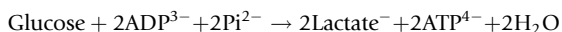
Lactic Acidosis

Regardless of whether anaerobic or aerobic, glycolysis produces acid if lactate is the end product of the pathway. The acid produced by glycolysis lowers the pH both inside cells where lactate is produced as well as outside where protons can diffuse. Since the pH range in which cells can function is quite narrow (pH 7.0–7.6), uncontrolled glycolysis can lead

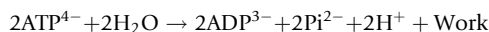
to cell death. This is the Achilles' heel of glycolysis. Indeed, in the final analysis it is overproduction of acid and lowering of the pH by glycolysis that kills most organisms, including humans. Cells incubated under anaerobic conditions produce large amounts of acid by anaerobic glycolysis. Likewise, forcing an area of the heart to obtain all of its energy from glycolysis by occluding a coronary artery causes rapid production of large amounts of acid, which lowers the pH, activates the nerve endings, and registers as pain. That the conversion of glucose to lactate produces acid is apparent when we write the balanced overall equation for glycolysis in the following manner:



Since the empirical formula for glucose is $\text{C}_6\text{H}_{12}\text{O}_6$, and there are six carbons, 12 hydrogens, and six oxygens in the products, this equation is balanced for mass and charge. Thus, two protons are produced for every glucose molecule converted to lactate molecules by glycolysis. Since glycolysis produces two ATPs per glucose, the equation seems incomplete, and in one sense it is incomplete. Expanding the equation to include ADP, P_i , and ATP in their predominant ionization states at physiological pH yields

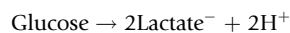


If this is accepted as the appropriate equation for glycolysis, balanced as it is for mass and charge, the pathway does not produce acid and therefore should have no effect on cellular pH. However, anaerobic glycolysis can clearly be shown to produce acid experimentally, and it does so because the pool size of ATP is small compared to the amount of glucose that is converted to lactate to meet the energy needs of a cell. For every glucose molecule converted to lactate, two ATP molecules have to be hydrolyzed according to the equation

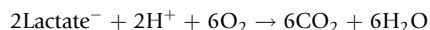


Work refers to many energy-requiring processes that can only occur as a consequence of ATP hydrolysis, such as muscle contraction, Na^+, K^+ -ATPase activity. Summing up the last

two equations brings us back to the overall balanced equation that shows acid production by glycolysis:



Anaerobic glycolysis therefore produces acid. Conditions in humans that greatly increase anaerobic glycolysis because of a shortage of oxygen, for example, failure of the respiratory system or the blood circulatory system, often cause the production of more acid than can be handled by the buffering systems of the body. The consequence is lactic acidosis, a life-threatening condition. Lactic acidosis can be dealt with most effectively by re-establishing the supply of oxygen. Reinstating ATP synthesis by oxidative phosphorylation will inhibit the production of lactic acid by glycolysis and also promote the oxidation of lactate as well as the consumption of the excess acid (H^+ 's) by the sum reaction:



See also: [Metabolism Vitamins and Hormones: Pyruvate Kinase](#).

Further Reading

- Berg JM, Tymoczko JL, and Stryer L (2007) *Biochemistry, Glycolysis*, 6th edn., pp. 433–474. New York NY: W.H. Freeman.
- Gibson DM and Harris RA (2002) *Metabolic Regulation in Mammals*. London: Taylor and Francis.
- Harris RA (2011) Carbohydrate metabolism I: Major metabolic pathways and their control. In: Devlin TM (ed.) *Textbook of Biochemistry with Clinical Correlations*, 7th edn., pp. 591–646. New York, NY: Wiley-Liss.
- Harris RA and Crabb DW (2011) Metabolic interrelationships. In: Devlin TM (ed.) *Textbook of Biochemistry with Clinical Correlations*, 7th edn., pp. 839–888. New York, NY: Wiley-Liss.
- King MW (2010) The medical biochemistry page. <http://web.indstate.edu/thcme/mwking> (accessed May 2010).
- Seagroves TN, Ryan HE, Lu H, et al. (2001) Transcription factor HIF-1 is a necessary mediator of the Pasteur effect in mammalian cells. *Molecular and Cellular Biology* 21: 3436–3444.
- Semenza L (2002) Signal transduction to hypoxia-inducible factor 1. *Biochemical Pharmacology* 64: 993–998.

Glycoprotein Folding and Processing Reactions

J J Caramelo and A J Parodi, Fundación Instituto Leloir, Buenos Aires, Argentina

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Armando J. Parodi, volume 2, pp. 272–276, © 2004, Elsevier Inc.

Glossary

BiP Classical endoplasmic reticulum-resident molecular chaperone belonging to the Hsp70 family.

Calnexin A membrane-bound endoplasmic reticulum-resident lectin that binds monoglucosylated *N*-glycans.

Calreticulin The soluble homolog of the former.

ERp57 Protein of the thioredoxin family that associates with calnexin and calreticulin and promotes proper disulfide bonding in monoglucosylated glycoproteins.

Glucosidase I A membrane-bound endoplasmic reticulum-resident glucosidase that removes the external α -(1,2)-linked Glc unit from Glc₃Man₉GlcNAc₂.

Glucosidase II A soluble endoplasmic reticulum-resident glucosidase that removes α -(1,3)-linked Glc units from Glc_{1,2}Man₇₋₉GlcNAc₂.

Glucosylation Term indicating transfer of Glc units.

Glycoproteins Proteins having covalently attached carbohydrates.

Glycosylation Generic term indicating transfer of monosaccharides or oligosaccharides.

Mannosidase I A membrane-bound, endoplasmic reticulum-resident enzyme that removes Man units from *N*-glycans.

N-glycans Glycans (or oligosaccharides) linked to Asn residues in proteins.

Quality control Mechanisms by which cells ensure that only properly folded proteins are produced by the secretory pathway.

UDP-Glc:glycoprotein glucosyltransferase A soluble, endoplasmic reticulum-resident enzyme that exclusively glucosylates glycoproteins displaying non-native conformations.

Glycoprotein Folding, Processing Reactions, and Quality Control

The term glycoprotein folding and processing reactions refer to the mechanisms by which cells ensure that only properly folded glycoproteins follow the secretory pathway. The concept of quality control of protein folding first appeared in the late 1970s and early 1980s when it was noticed that not in all cases insertion of proteins or glycoproteins in the endoplasmic reticulum (ER) lead to their expected final destination, either intracellular (e.g., the lysosomes) or secretion. Several experimental results showed that cells displayed mechanisms that ensured that only proteins in their native conformations could be produced by the secretory pathway. Those mechanisms received the collective denomination of quality control. By far, the best-studied quality control mechanism is that occurring in the ER but more recent studies revealed that it also takes place at the Golgi apparatus. The scope of this article is limited to the quality control occurring on glycoproteins in the ER.

Protein Folding in the ER

Nearly 40% of the proteins synthesized by eukaryotic cells belong to the secretory pathway. Of these, approximately 80% are *N*-glycosylated in the ER. These proteins enter the ER either post- or cotranslationally. The ER folding quality control system is adapted to handling virtually any condition that results in a protein conformation other than the native state. Accordingly, multiple scenarios of quality control arise from many ways in which proteins can fail to achieve their native conformation. All types of polypeptides inserted into the ER are subjected to

quality control, including soluble and membrane proteins (type I, II, and multispans). It applies to proteins regardless of their quaternary structure and whether they are glycosylated or not or contain intra- or intermolecular bonds. Proteins can fail the quality control due to multiple factors, such as truncations or mutations that interfere with proper folding, as well as orphan subunits of heterooligomeric complexes or homooligomers that cannot reach their proper structure. In addition, the lack of ligand binding or posttranslational modifications (signal sequence cleavage, disulfide bond formation, *N*-glycosylation, or glycosylphosphatidylinositol (GPI) anchor addition) can prevent proteins from folding correctly. Moreover, normal proteins that do not have any of the mentioned defects can also be targets of the quality control system, as protein folding is an extremely hazardous process whose efficiency heavily depends not only on particular environmental conditions but also on the amino acid primary sequence. In fact, almost never do all normal protein molecules manage to reach their proper native structure; there is always a variable fraction of newly synthesized proteins that do not fold efficiently.

A number of abundant ER-resident proteins that participate in retention of non-native conformers have been identified as chaperones and folding-assisting proteins. Protein disulfide isomerase (PDI), one of the first facilitators of protein folding described, is an abundant ER chaperone. PDI and other members of the thioredoxin family participate in disulfide bond formation as well as retention and degradation of targets of the quality control system. Of particular interest is one member of this family (ERp57) which, as will be described below, specifically associates with two ER-resident lectins, calnexin, (CNX), and calreticulin (CRT), and contributes to proper glycoprotein

folding. Another abundant ER protein, BiP (for 'binding protein'), that was originally identified as an ER retention factor for immunoglobulin heavy chains, participates in assisting protein folding through cycles of adenosine triphosphate (ATP) hydrolysis that regulate substrate binding and release.

N-Glycan Processing Reactions in the ER

Glycosylation of asparagine residues (N-glycosylation) in the mammalian cell ER presents two basic differences with respect to glycosylation of Ser/Thr units (O-glycosylation) in either the cytosol or the Golgi apparatus: (1) in the former case a glycan containing 14 monosaccharide units in most cells types is transferred *en bloc* to the polypeptide chain, whereas in O-glycosylation either a variable number of monosaccharide units are sequentially transferred (Golgi O-glycosylation) or a single of such units is added (cytosolic O-glycosylation) and (2) N-glycosylation occurs on not yet properly folded polypeptide moieties whereas in all cases O-glycosylation is a post-folding event. After being transferred to nascent chains in the ER, N-glycans are processed all along the secretory pathway. Whereas processing in the ER is basically similar in all eukaryotic cells for all N-glycans, Golgi processing reactions present a broad variation not only among different cell types but also different N-glycans belonging to the same or different glycoproteins may be differently processed within the same cell, thus yielding mature glycoproteins bearing glycan moieties with widely different structures. This diversity presents certain logic as the roles of protein-linked glycans in mature glycoproteins are mainly related to recognition phenomena. On the contrary, glycan processing in the ER is basically similar in all cells because, as it will be described below, the role of those reactions directly relates to a process common to all glycoproteins, namely the acquisition of their proper tertiary and quaternary structures.

Addition of N-glycans has several effects on protein folding, at least two of which may be observed not only *in vivo* but also in folding assays performed in test tubes with pure glycoproteins: (1) N-glycans provide bulky, highly hydrophilic groups that help to maintain glycoproteins in solution during the folding process and (2) the presence of glycans modulates protein conformations by forcing amino acids close to the linking Asn unit to be in the proximity of the water-glycoprotein interphase. This article does not deal with those effects but with *in vivo* folding facilitation and ER retention of improperly folded structures mediated by the interaction of a specific glycan structure (monoglucosylated glycans) with the ER lectins CNX and CRT.

The saccharide transferred in most wild-type cells from a lipid (dolichol) derivative to polypeptide chains as they enter the ER lumen has the composition $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ and the structure depicted in Figure 1. This is a well-conserved structure in nature as the same glycan is transferred in wild-type mammalian, plant, and fungal cells. There are some exceptions to this common pattern in certain protozoa, in which truncated N-glycans are transferred. The consensus glycosylation sequence (Asn-X-Ser/Thr, where X may be any amino acid except Pro) is not necessarily glycosylated. Apparently, N-glycosylation, that occurs when the involved Asn unit is about 10–12 amino acids far from the inner ER membrane surface, is hindered if the polypeptide chain rapidly adopts a secondary structure, in

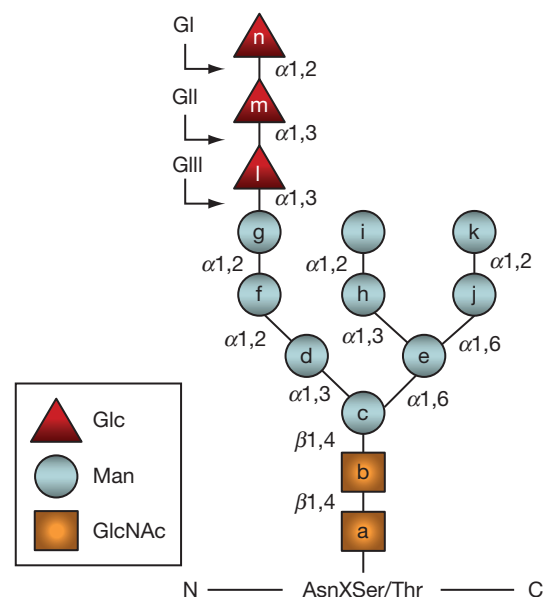


Figure 1 Glycan structures. Lettering (a–n) follows the order of addition of monosaccharides in the synthesis of $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ -P-P-dolichol. GI removes residue n and GII residues l and m. UGGT adds residue l to residue g. The $\text{Man}_8\text{GlcNAc}_2$ isomer formed by mammalian cell or *S. cerevisiae* ER α -mannosidase I lacks residues i, and l–n. The $\text{Man}_7\text{GlcNAc}_2$ isomer formed by the same enzyme lacks, additionally, residue k.

particular α -helices. Cell-free assays showed that the oligosaccharyltransferase transfers the triglucosylated glycan about 10–15-fold faster than $\text{Man}_9\text{GlcNAc}_2$. This property, as well as the strict distance between the receptor amino acid and the ER inner surface required for glycan transfer, determines a severe underglycosylation of glycoproteins synthesized by cell mutants unable to glucosylate the lipid-linked glycan. Folding is severely impaired in many, but not all, underglycosylated glycoproteins. The external Glc unit (residue n, Figure 1) is immediately removed from the protein-linked glycan by glucosidase I (GI), a membrane-bound α -(1,2)-glucosidase (see mammalian cell ER processing of N-glycans in Figure 2). Further deglycosylation mediated by glucosidase II (GII, an α -(1,3)-glucosidase) may remove the remaining Glc units (residues l and m, Figure 1). GII is a soluble heterodimer composed of catalytically active and inactive subunits. The latter, but not the former, contains an ER retrieval sequence at its C-terminus. One or two Man units (residues i and k, Figure 1) may be removed by ER α -mannosidase I.

An additional N-glycan-processing reaction occurring in the ER lumen is the transient glucosylation of glycoproteins: glucose-free glycans may be reglucosylated by the UDP-Glc:glycoprotein glucosyltransferase (UGGT) yielding $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2$, $\text{Glc}_1\text{Man}_8\text{GlcNAc}_2$, and $\text{Glc}_1\text{Man}_7\text{GlcNAc}_2$ as reaction products. The structure of the first one is identical with that produced by partial deglycosylation of $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ (Figure 1) and GII is also responsible for the *in vivo* deglycosylation of the monoglucosylated compounds formed by UGGT. This enzyme is present not only in mammalian but also in plant, fungal, and protozoan cells. Interestingly, UGGT is absent from *Saccharomyces cerevisiae* and in some protozoa such as *Tetrahymena thermophila*, *Toxoplasma gondii*, and *Cryptosporidium parvum*.

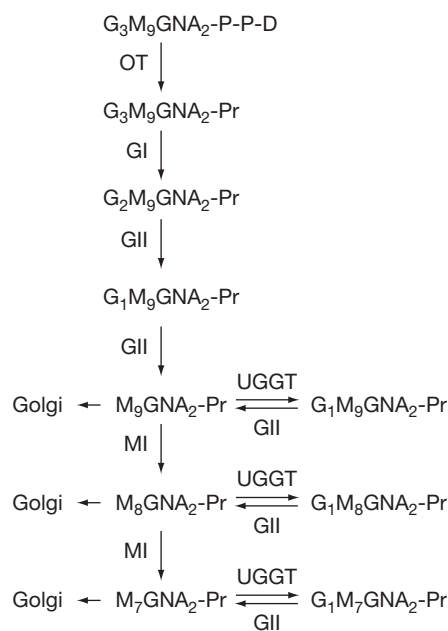


Figure 2 Glycan processing in the ER. G, M, GNA, D, and Pr stand for Glc, Man, GlcNAc, dolichol, and protein, respectively. OT, GI, GII, UGGT, and MI stand for oligosaccharyltransferase, glucosidase I, glucosidase II, UDP-Glc:glycoprotein glucosyltransferase and mannosidase I, respectively.

The Golgi apparatus of species belonging to the chordate phylum display an endomannosidase activity that degrades monoglycosylated *N*-glycans yielding Glc₁Man (plus Man₆₋₈ GlcNAc₂, depending on the substrate used). This enzyme has been described to occur also in the mammalian cell ER-Golgi intermediate compartment. It may be speculated that the role of this enzyme is to ensure a complete removal of Glc units, as their presence would hinder formation of complex-type glycans in the Golgi apparatus.

The UDP-Glc:Glycoprotein Glucosyltransferase

The transient addition of a terminal glucose residue to protein-linked glycans was discovered while studying *N*-glycan processing in trypanosomatid cells pulse-chased with [¹⁴C] Glc. These *in vivo* observations were then extended to other eukaryotes and it was soon established that transient reglucosylation of completely deglycosylated, protein-bound high mannose-type glycans was a general glycoprotein-processing reaction. Since glucose is not normally found on mature glycoproteins, transient reglucosylation was intriguing and the function of such apparent futile cycle of glucose addition and removal was not understood at the time. The reglucosylation of endogenous glycoproteins observed *in vivo* was reproduced in cell-free systems and the activity was localized by subcellular fractionation to the ER, where as mentioned above, deglycosylated glycoproteins that are subject to reglucosylation are formed by the action of GI and GII. As an *in vitro* assay for exogenous glycoprotein reglucosylation was developed, it was noticed that the substrate glycoprotein had to be denatured in order to be an acceptor in the reglucosylation reaction. This

observation raised the radical notion that glycoproteins had to be devoid of a native conformation to be recognized as substrates by UGGT. Based on this unique property, it was suggested that the enzyme (and monoglycosylated glycans) could be somehow involved in the so-called quality control of glycoprotein folding in the ER.

The involvement of protein recognition in glycosylation reactions was not without precedent. While most glycosyltransferases display specificity at the carbohydrate level (sugar acceptor, sugar donor, and linkage formed), some glycosyltransferases also recognize the protein to which the substrate glycan is bound to. The UDP-N-acetylglucosamine:lysosomal enzyme *N*-acetylglucosamine phosphotransferase selectively initiates formation of the Man 6-P epitope in lysosomal glycoproteins. In addition, the UDP-GalNAc:glycoprotein hormone *N*-acetylgalactosaminyltransferase selectively adds a GalNAc residue to secreted hormones. The substrates that reacted with UGGT *in vitro* did not appear to share any apparent common feature, except for their non-native conformation, as opposed to the other two glycosyltransferases, which maintain their rather narrow selectivity toward their respective glycoprotein substrates (lysosomal enzymes or glycoprotein hormones) when assayed *in vitro*. The requirement for non-native conformations implied that a wide range of substrates could be reglucosylated *in vivo*, since a large number of glycoproteins could potentially be recognized as substrates during their folding in the ER. Indeed, it is now accepted that many glycoproteins, independently from their final destination or of their soluble or membrane-bound status, behave as UGGT substrates during their folding process.

Purification of the enzyme from rat liver or *Schizosaccharomyces pombe* showed that activity resided in a single polypeptide of about 170 kDa, significantly larger than most glycosyltransferases. While most glycosyltransferases and glycosidases in the secretory pathway are type II membrane proteins, UGGT and its counteracting enzyme GII are soluble proteins. The purified UGGT has a neutral pH optimum, requires millimolar Ca²⁺ concentrations for activity, and uses UDP-Glc as substrate donor.

The ER membrane contains an integral protein responsible for the specific transport of UDP-Glc from its site of synthesis (the cytosol) into the lumen of the ER, where the sugar nucleotide serves as substrate for glycoprotein reglucosylation. Further, two GDPase/UDPas have been described in the mammalian cell ER lumen. Their role is probably to allow transport of UDP-Glc into the lumen by the known antiport mechanism by which entrance of sugar nucleotides into the ER or Golgi lumen is coupled to exit the corresponding nucleoside monophosphates.

The most significant finding in the isolation of UGGT was that the purified enzyme maintained its selectivity for denatured substrates. While it may still cooperate with other factors to recognize non-native conformations *in vivo*, the finding that pure UGGT exclusively reglucosylates non-native glycoproteins strongly suggests that the conformational sensing is performed primarily by the UGGT itself. No cofactors were found so far to modify the activity of isolated UGGT *in vitro*, and expression of UGGT cDNA is sufficient for glycoprotein reglucosylation in cells devoid of such activity (*S. cerevisiae*). It is now accepted that UGGT recognizes hydrophobic amino acid patches exposed in molten globule-like folding intermediates.

The Quality Control Mechanism of Glycoprotein Folding

Information mentioned above supports the mechanism depicted in **Figure 3** for the quality control mechanism of glycoprotein folding occurring in the ER. Accordingly, monoglucosylated glycoproteins generated either by partial deglucosylation of the transferred glycan or by UGGT-mediated glucosylation interact with CNX/CRT, two ER resident lectins that specifically recognize that structure. This interaction is followed by a shuttle between glucosylated (CNX/CRT-bound) and deglucosylated (CNX/CRT-free) forms catalyzed by the opposing activities of UGGT and GII. After acquiring their

proper tertiary structure, glycoproteins are deglucosylated by GII but not reglucosylated by UGGT and no longer bind to CNX/CRT. Permanently misfolded molecules continue to interact with the lectins, which are not sensing the conformation of their glycoprotein ligands. This task, all over the quality control process, is reserved to the reglucosylating enzyme, and cycles of binding and release to the lectins are mediated by independently acting enzymes (UGGT and GII) that covalently modify glycans. The CNX/CRT–glycoprotein interaction not only prevents secretion of folding intermediates and irreparably misfolded molecules, but also enhances folding efficiency as it prevents protein aggregation, premature oligomerization

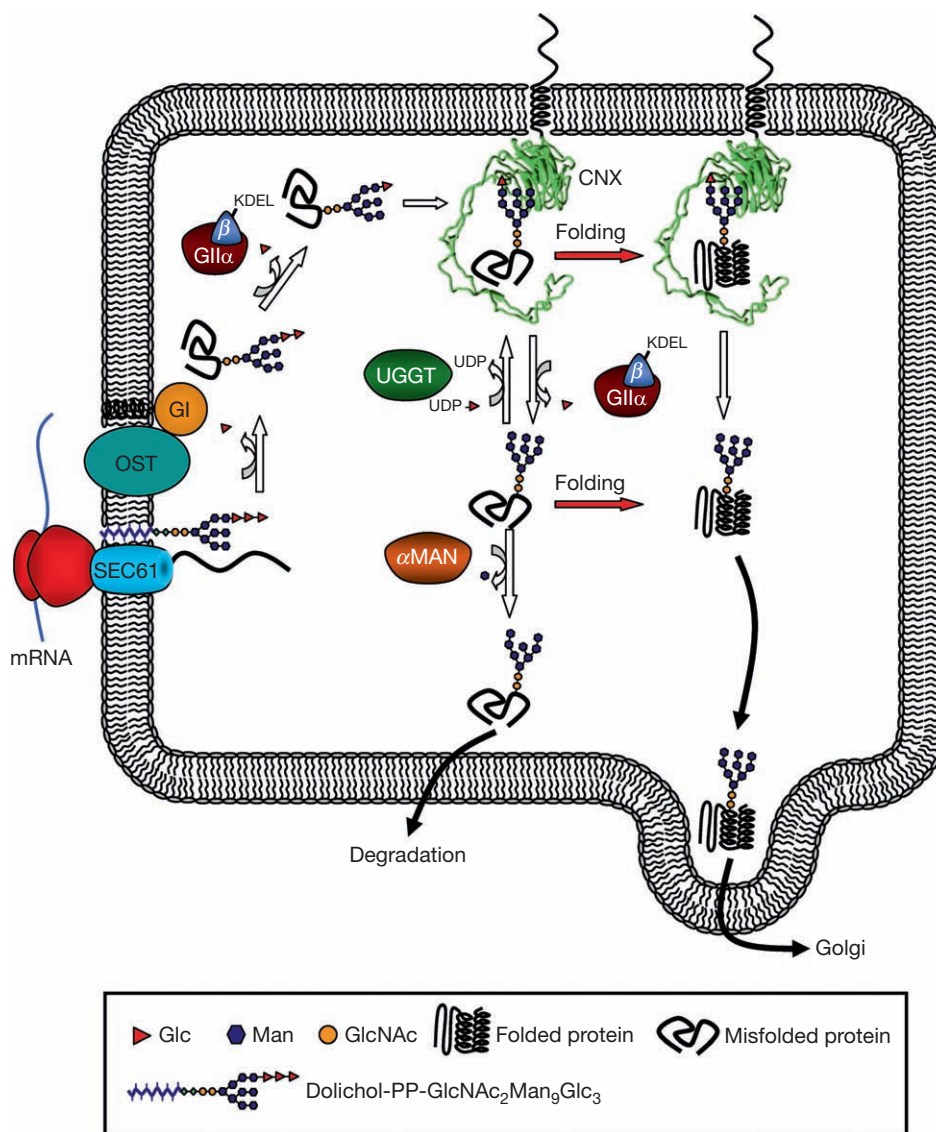


Figure 3 Model proposed for the quality control of glycoprotein folding. Proteins entering the ER are N-glycosylated by the oligosaccharyltransferase (OST) as they emerge from the translocon (Sec61). Two glucoses are removed by the sequential action of GI and GII to generate monoglucosylated species that are recognized by CNX and/or CRT (only CNX is shown) that are associated with ERp57. The complex formed by the lectins and folding intermediates/misfolded glycoproteins dissociates upon removal of the last glucose by GII and is reformed by UGGT activity. Once glycoproteins have acquired their native conformations, either free or complexed with the lectins, GII hydrolyzes the remaining glucose residue and releases the glycoproteins from the lectin anchors. These species are not recognized by UGGT and are transported to the Golgi. Glycoproteins unable to properly fold are retro-translocated to the cytosol where they are degraded by the proteasome. One or more mannose residues may be removed by the ER α-mannosidase I (α-Man) during the whole folding process.

and promotes formation of the correct disulfide bonds by the CNX/CRT-ERp57 complex mentioned above. It is worth mentioning that UGGT and BiP recognize different folding conformers. The latter specifically binds heptapeptides with bulky hydrophobic amino acids in alternating positions exposed in early, extended conformations. On the contrary, UGGT glucosylates collapsed, molten globule-like conformers. As predicted by the different recognition pattern, sequential BiP–CNX/CRT interaction has been described for the *in vivo* folding of several glycoproteins. Folding facilitation mediated by CNX/CRT–glycoprotein interaction is not essential for the eukaryotic cell viability. Hindering monoglucosylated *N*-glycan formation either by disrupting *GI*-, *GII*-, or UGGT-encoding genes or by adding cell permeable inhibitors of the first two enzymes leads to the upregulation of other chaperones and folding assisting proteins that compensate the absence of lectin–glycoprotein interaction (unfolded protein response). Finally, it should be mentioned that irreparably misfolded glycoproteins are retained in the ER and eventually transported through a complex and not-yet fully understood process to the cytosol to be degraded in the proteasomes.

See also: **Cell Architecture and Function:** Unfolded Protein Responses; **Protein/Enzyme Structure Function and Degradation:** Proteasomes, Overview.

Further Reading

- Caramelo JJ and Parodi AJ (2007) How sugars convey information on protein conformation in the endoplasmic reticulum. *Seminars in Cell and Developmental Biology* 18: 732–742.
- Caramelo JJ and Parodi AJ (2008) Getting in and out of calnexin and calreticulin cycles. *Journal of Biological Chemistry* 283: 10221–10225.
- D'Alessio C, Caramelo JJ, and Parodi AJ (2010) UDP-Glc:glycoprotein glucosyltransferase-glucosidase II: The ying-yang of the ER quality control. *Seminars in Cell and Developmental Biology* 21: 491–499.
- Ellgaard L and Helenius A (2003) Quality control in the endoplasmic reticulum. *Nature Reviews in Molecular and Cell Biology* 4: 181–191.
- Ellgaard L, Molinari M, and Helenius A (1999) Setting the standards: Quality control in the secretory pathway. *Science* 286: 1882–1888.
- Hammond C and Helenius A (1997) Quality control in the secretory pathway. *Current Opinion in Cell Biology* 7: 523–529.
- Helenius A (1994) How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Molecular Biology of the Cell* 5: 253–265.
- Helenius A and Aebi M (2004) Roles of N-linked glycans in the endoplasmic reticulum. *Annual Review of Biochemistry* 73: 1019–1049.
- Parodi AJ (2000) Protein glucosylation and its role in protein folding. *Annual Review of Biochemistry* 69: 69–95.
- Trombetta ES and Parodi AJ (2002) *N*-glycan processing and glycoprotein folding. *Advances in Protein Chemistry* 59: 303–344.
- Trombetta ES and Parodi AJ (2003) Quality control and protein folding in the secretory pathway. *Annual Review in Cell and Developmental Biology* 19: 649–676.
- Trombetta ES and Parodi AJ (2005) Quality control in glycoprotein folding. In: Buchner J and Kiefhaber T (eds.) *Handbook of Protein Folding Part II*, vol. 2, pp. 617–648. Weinheim: Wiley-VCH.

Glycoprotein-Mediated Cell Interactions, Nonmucin-Type O-Linked

R S Haltiwanger, Stony Brook University, Stony Brook, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

α -Dystroglycan An essential component of the link between the cytoskeleton and the extracellular matrix in sarcolemma.

Fringe A β 1,3-*N*-acetylglucosaminyltransferase that modifies *O*-fucose residues on Notch. Alteration of the *O*-fucose structures results in modulation of Notch activity.

Notch A genetic locus originally identified in *Drosophila* that encodes a large cell surface receptor essential for many

stages of development. The Notch receptor is modified with both *O*-fucose and *O*-glucose.

O-linked carbohydrates Carbohydrate covalently linked to a protein through a hydroxyl group, usually of serine or threonine.

Rumi A protein *O*-glucosyltransferase that adds *O*-glucose to epidermal growth factor (EGF)-like repeats in the extracellular domain of Notch.

The addition of *O*-glucose appears to be essential for Notch activity.

Classes of O-Glycans

O-linked carbohydrate modifications are defined as any carbohydrate covalently linked to protein through a hydroxyl group, most commonly the hydroxyl of a serine or threonine residue. The classes of O-linked carbohydrate modifications are defined by the sugar directly linked to the hydroxyl group. In mammalian systems, at least six classes of O-glycans exist: *O*-GalNAc (*N*-acetylgalactosamine), *O*-GlcNAc (*N*-acetylglucosamine), *O*-xylose, *O*-fucose, *O*-glucose, and *O*-mannose. With the exception of *O*-GlcNAc, all of these modifications are found on proteins in extracellular spaces. *O*-GlcNAc is uniquely found on nuclear and cytoplasmic proteins. The *O*-GalNAc and *O*-xylose modifications constitute the largest and most structurally heterogeneous classes of O-linked modifications. The *O*-GalNAc modifications are usually referred to as mucin-type O-glycosylation because they are prevalent on mucins, and the *O*-xylose modifications are the glycosaminoglycans, the carbohydrate portion of proteoglycans. Two other O-linked modifications found on the hydroxyls of amino acids other than serine or threonine also exist in mammals: *O*-galactose on hydroxylysine of collagen and *O*-glucose on tyrosine of hyalogenin.

between the second and third conserved cysteine of the EGF-like repeat at the sequence $C^2XXXX(S/T)C^3$, where X can be any amino acid and S/T is the modified amino acid. Database searches show that over 100 proteins in mouse or human genomes contain this consensus sequence with an EGF-like repeat including all Notch receptors.

The *O*-fucose is added to EGF-like repeats by protein *O*-fucosyltransferase I (Pofut1, [Figure 2\(a\)](#)). It will only transfer fucose to serine or threonine in a consensus sequence within a properly folded EGF-like repeat, suggesting that the three-dimensional structure of the EGF-like repeat is necessary for recognition of the site. The gene for Pofut1 is highly conserved in metazoans, and mutations in the gene result in embryonic lethality in mice and fruit flies. The *O*-fucose can be elongated with a β 1,3-GlcNAc by a β 1,3-*N*-acetylglucosaminyltransferase called Fringe ([Figure 2\(a\)](#)). The GlcNAc can be further elongated with a galactose and sialic acid to form a mature tetrasaccharide in mammals. Elongation of the *O*-fucose monosaccharide is dependent on both the protein being modified and the tissue-dependent expression of the Fringe proteins. Fringe will modify *O*-fucose on some EGF-like repeats but not others, suggesting it recognizes not just the *O*-fucose, but also the surrounding protein structure.

O-Fucose on Epidermal Growth Factor-Like Repeats

O-Fucose modifications were initially identified on a number of serum glycoproteins involved in blood clot formation or dissolution including urinary-type plasminogen activator (uPA), tissue-type plasminogen activator, and factors VII, IX, and XII. Analysis of the sites of fucosylation on these proteins revealed that the fucose was attached to a serine or threonine in a consensus sequence within a protein module known as an epidermal growth factor (EGF)-like repeat. EGF-like repeats are small protein modules of ~ 40 amino acids found in dozens of cell-surface and secreted proteins ([Figure 1](#)). They are defined by the presence of six conserved cysteine residues that form three disulfide bonds. The *O*-fucose modification occurs

O-Glucose on EGF Repeats

O-Glucose modifications were also initially described on EGF-like repeats of serum glycoproteins such as factors VII and IX. Analysis of the sites of glycosylation revealed that the *O*-glucose modifies serine residues between the first and second conserved cysteine of the EGF-like repeat at the sequence: C^1XSXPC^2 ([Figure 1](#)). Over 50 proteins in mouse or human genomes are predicted to bear *O*-glucose modifications based on databases using this sequence, including all Notch receptors. *O*-Glucose is added to EGF-like repeats by a protein *O*-glucosyltransferase (Poglut) called Rumi, and the *O*-glucose can be further elongated with xyloses ([Figure 2\(b\)](#)). Like Pofut1, the gene encoding Rumi is highly conserved in

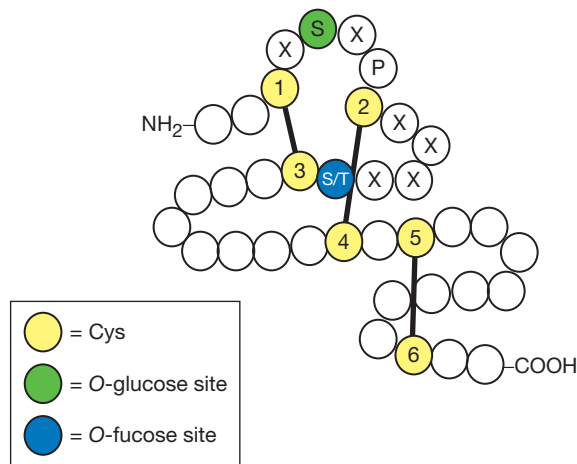


Figure 1 *O*-Fucose and *O*-glucose modifications occur within epidermal growth factor (EGF)-like repeats. A schematic representation of an EGF-like repeat is shown. Each circle represents an amino acid. Cysteines are shown in yellow, the *O*-glucose modification site is shown in green, and the *O*-fucose modification site is shown in blue. Amino acids important for defining the glycosylation sites are indicated. Disulfide linkages are represented by lines.

metazoans, and mutations in Rumi result in developmental defects in both flies and mice.

O-Fucose on TSRs

O-Fucose modifications have also been identified in a separate sequence context, that of thrombospondin type 1 repeats (TSRs). TSRs are a protein module found in numerous cell surface and secreted proteins of ~60 amino acids (Figure 3). TSRs are identified by the presence of several conserved residues, including six conserved cysteines that form internal disulfide bonds. The *O*-fucose modification on TSRs was first identified on human thrombospondin 1, and based on comparison of sequences surrounding the site of glycosylation on several TSRs, a consensus sequence for modification has been proposed: C¹X₂₋₃(S/T)C²X₂G. Database searches suggest that over 50 proteins in mouse or human genomes contain *O*-fucosylated TSRs. Modification of TSRs with *O*-fucose in several of these proteins, such as thrombospondin 2, properdin, F-spondin, ADAMTS13, and ADAMTS-like 1, has been confirmed using mass spectrometry. The *O*-fucose is added to the TSR by protein *O*-fucosyltransferase 2 (Pofut2), a distinct gene product from Pofut1 with nonoverlapping specificity (Figure 2(c)). The *O*-fucose can be elongated with a β 1,3-glucose by the action of an *O*-fucose- β 1,3-glucosyltransferase.

O-Mannose

O-Mannose modifications of proteins were originally described in yeast, but more recently have been shown to occur on mammalian proteins as well. The first description of *O*-mannose glycans in mammals was on brain proteoglycans.

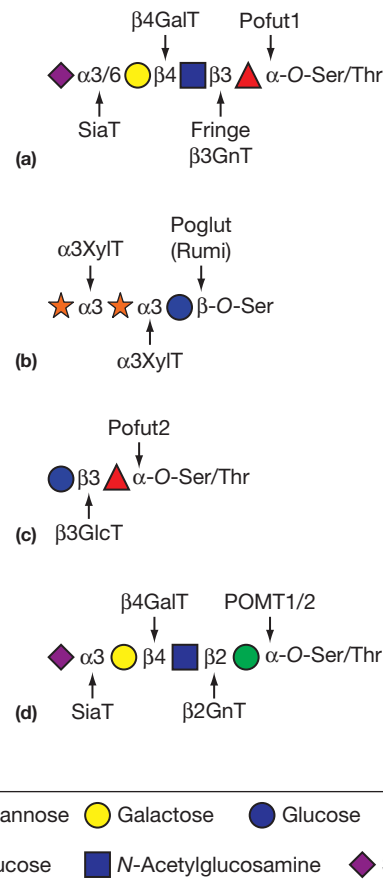


Figure 2 Structures of *O*-linked carbohydrate modifications. For each modification, the potential structures are shown. Each monosaccharide is represented by the symbols defined at the bottom. The enzyme responsible for addition of each sugar is indicated with an arrow. (a) *O*-Fucose on EGF-like repeats. (b) *O*-glucose on EGF-like repeats. (c) *O*-fucose on TSRs. (d) *O*-mannose on mucin-like domain of α -dystroglycan. Pofut1, protein *O*-fucosyltransferase-1; Poglut (Rumi), protein *O*-glucosyltransferase; POMT, protein *O*-mannosyltransferase; β 3GnT, β 1,3-*N*-acetylglucosaminyltransferase; β 4GalT, β 1,4-galactosyltransferase; SiaT, sialyltransferase; α 3XylT, α 1,3-xylosyltransferase; β 3GlcT, β 1,3-glucosyltransferase; β 2GnT, β 1,2-*N*-acetylglucosaminyltransferase.

More recently, *O*-mannose glycans were detected on a specific protein, α -dystroglycan. Numerous *O*-mannose chains are attached to α -dystroglycan in a mucin-like domain. This is a region of the protein rich in serine, threonine, and proline where many of the serines and threonines appear to be modified. The mannose is added to the mucin-like domain by a protein *O*-mannosyltransferase (POMT). Two closely related genes encoding protein *O*-mannosyltransferases, POMT1 and POMT2, are required for POMT activity. Mutations in either of these genes cause a loss of *O*-mannosylation in mammals. The *O*-mannose can be elongated with a β 1,2-GlcNAc by an *O*-mannose- β 1,2-GlcNAc transferase, and the GlcNAc can be further elongated to a mature tetrasaccharide by the addition of galactose and sialic acid (Figure 2(d)). Other elongated forms of *O*-mannose are known to exist.

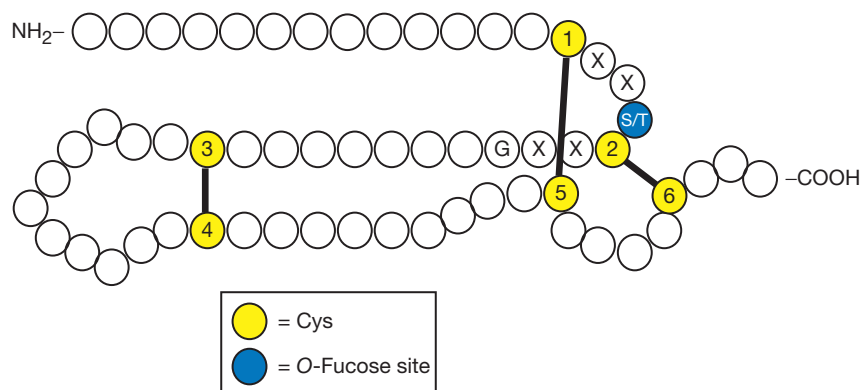


Figure 3 *O*-Fucose modifications occur on thrombospondin type 1 repeats (TSRs). A schematic representation of a TSR is shown. Each circle represents an amino acid. Cysteines are shown in yellow and the *O*-fucose site is shown in blue. Amino acids important for defining the glycosylation sites are indicated. Disulfide linkages are represented by lines.

Functions of *O*-Glycans

Determining the function of protein glycosylation has remained a problem for many years, partly because so many different proteins bear carbohydrate modifications. Carbohydrate modifications differ in function, depending on the protein they modify. Nonetheless, recent work has revealed specific functions for some of the *O*-linked modifications. The fact that these modifications are found on a smaller subset of proteins than more common types of glycosylation (e.g., *N*-glycosylation, mucin-type *O*-glycosylation, and glycosaminoglycans) has made evaluation of their functions somewhat more straightforward. Still, the principles learned from studying these modifications should be applicable to our understanding of other forms of glycosylation.

The Effects of *O*-Fucose in uPA–uPA Receptor Signaling

A role for the *O*-fucose modification in modulation of signal transduction was first proposed for the activation of the uPA receptor. As mentioned above, the EGF-like repeat of uPA was one of the first proteins to be demonstrated to bear an *O*-fucose modification. The uPA receptor is a mitogenic receptor, and it becomes activated upon binding to the EGF-like repeat of uPA. Interestingly, removal of *O*-fucose from the EGF-like repeat abrogates uPA receptor activation, although it has little or no effect on binding. Thus, the *O*-fucose appears to be essential for activation of the uPA receptor, and alterations in the level of *O*-fucose on uPA could potentially serve as a mechanism for modulating this activation.

The Effects of *O*-Fucose and *O*-Glucose in Notch Signaling

The concept that *O*-fucose and *O*-glucose modifications can regulate signal transduction events is strengthened by work on the Notch receptor. Notch is a cell-surface signaling receptor that plays many essential roles in the development of metazoans. Defects in the Notch signaling pathway lead to a number of human diseases, including several types of cancer, cerebral autosomal dominant arteriopathy with subcortical infarcts and

leukoencephalopathy (CADASIL, an inherited disease that results in recurrent strokes and early death), and a group of developmental disorders termed Alagille syndrome. Recent work has also implicated Notch in the pathogenesis of multiple sclerosis. Notch becomes activated upon binding to its ligands (members of the Delta family or the Serrate/Jagged family of proteins). Interestingly, the ligands are also cell-surface, transmembrane proteins. Thus, Notch becomes activated when an adjacent cell expresses ligand. Ligand binding causes the activation of proteases, releasing the cytoplasmic domain of Notch into the cytoplasm. The cytoplasmic domain then moves into the nucleus, where it associates with other proteins to regulate the transcription of a number of genes.

The extracellular domain of the Notch receptor is composed largely of tandem EGF-like repeats (36 consecutive EGF-like repeats in mammalian Notch1), many of which are modified with *O*-fucose and/or *O*-glucose. Both the *O*-fucose and *O*-glucose modifications are essential for Notch function. Abrogation of Pofut1 activity in *Drosophila* (by RNA interference) or in mice (by gene knockout) results in Notch phenotypes (i.e., the same phenotype observed when mutations are made in the Notch gene itself). Similarly, mutations in the Rumi gene (encoding Poglut) cause temperature-sensitive Notch phenotypes in flies. These results imply that if Notch is not modified by either *O*-fucose or *O*-glucose, it does not function properly. Since most Notch phenotypes arise early in development, these defects result in embryonic lethality.

In addition to being essential for Notch function, the *O*-fucose modification on Notch serves a regulatory role. The *O*-fucose on the EGF-like repeats of Notch can exist in either the monosaccharide or the tetrasaccharide form, depending on whether Fringe is expressed in the same cell. Fringe is a known modifier of Notch function, potentiating the signal received from Delta-like ligands, but inhibiting the signal received from Serrate/Jagged-like ligands. Although the details of how this modulation occurs are not known, these results indicate that the activation of Notch can be regulated by altering the structure of the *O*-fucose glycans on Notch. These results provide one of the clearest examples known of how a signal transduction pathway can be regulated by alterations in the glycosylation of the receptor.

The Role of *O*-Mannose in α -Dystroglycan Function

The recent demonstration of *O*-mannose on α -dystroglycan has led to the discovery of an essential role for these glycans in the musculoskeletal system. α -Dystroglycan forms a portion of an essential link between the cytoskeleton and the extracellular matrix in the sarcolemma. Specifically, α -dystroglycan binds to β -dystroglycan, a transmembrane protein that in turn interacts with cytoskeletal components. In addition, α -dystroglycan binds to laminin, an extracellular matrix component. Interestingly, the interaction with laminin appears to be mediated through the *O*-mannose oligosaccharides on α -dystroglycan, and alteration in the *O*-mannose tetrasaccharide appears to disrupt this interaction. Defects in any of the components in the link between the cytoskeleton and laminin in the muscle are known to cause various muscular dystrophies. For instance, defects in the enzyme that adds a GlcNAc to the *O*-mannose on α -dystroglycan (β 2GnT in [Figure 2\(d\)](#)), preventing formation of the mature tetrasaccharide, result in a form of congenital muscular dystrophy: muscle–eye–brain disease. Similarly, mutations in either POMT1 or POMT2 reduce *O*-mannosylation of α -dystroglycan, resulting in another congenital muscular dystrophy: Walker–Warburg syndrome. These results demonstrate that the *O*-mannose saccharides of α -dystroglycan play an essential role in the function of the dystroglycan–laminin interaction.

Potential Roles for *O*-Fucosylation of TSRs

Peters plus syndrome is an autosomal recessive human genetic disorder caused by mutations in the β 3GlcT that adds glucose to *O*-fucose on TSRs ([Figure 2\(c\)](#)). The patients display a

number of developmental defects, including anterior eye chamber defects, short stature, developmental delay, and increased incidence of cleft lip/palate. Although it is not known which protein(s) are responsible for these defects, these findings clearly demonstrate that *O*-fucose glycans on TSRs play very important roles in biology. Similarly, elimination of the gene encoding Pofut2 ([Figure 2\(c\)](#)) in mice causes early embryonic lethality. Together, these observations suggest that *O*-fucosylation of one or more of the 50 or more proteins predicted to be modified by Pofut2 is essential for their proper function.

See also: [Lipids Carbohydrates Membranes and Membrane Proteins: Glycoprotein Folding and Processing Reactions; O-Linked GlcNAc Biosynthesis and Function; N-Linked Glycan-Processing Glucosidases and Mannosidases.](#)

Further Reading

- Hewitt JE (2009) Abnormal glycosylation of dystroglycan in human genetic disease. *Biochimie et Biophysica Acta* 1792: 853–861.
- Jafar-Nejad H, Leonardi J, and Fernandez-Valdivia R (2010) Role of glycans and glycosyltransferases in the regulation of Notch signaling. *Glycobiology* 20: 931–949.
- Luther KB and Haltiwanger RS (2008) Role of unusual *O*-glycans in intercellular signaling. *International Journal of Biochemistry and Cell Biology* 41: 1011–1024.
- Stanley P (2007) Regulation of Notch signaling by glycosylation. *Current Opinion In Structural Biology* 17: 530–535.
- Takeuchi H and Haltiwanger RS (2010) Role of Glycosylation of Notch in Development. *Seminars in Cell and Developmental Biology* ePub.

Glycoproteins, N-Linked

M A Lehrman, University of Texas Southwestern Medical Center, Dallas, TX, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Asparagine-linked glycan An oligosaccharide covalently attached to the nitrogen atom of an asparaginyl residue of a protein or peptide.

Dolichol A polyisoprenoid lipid composed of five-carbon isoprene units and terminating with a hydroxyl that can be modified with groups such as phosphate, sugar phosphate, or oligosaccharyl pyrophosphate. The business-end (α) isoprene unit is reduced, while

the remaining isoprene units in dolichol retain double bonds.

Glycosidase An enzyme that can release a monosaccharide or oligosaccharide by cleavage of a glycosidic linkage.

Glycosyltransferase An enzyme that transfers a sugar residue from a donor substrate to a saccharide, protein, or lipid acceptor.

Lectin A protein with a binding site that is specific for sugars.

Biosynthesis and Structures of N-Linked Oligosaccharides

Synthesis of a Dolichol Pyrophosphate-Linked Oligosaccharide

N-linked glycosylation begins not with a polypeptide substrate, but rather with the polyisoprenoid lipid dolichol phosphate (Dol-P; also referred to as dolichyl phosphate by many workers). Dolichols are the longest lipids in humans, most commonly consisting of 95 carbons (19 isoprenyl units). Originating from the terminal phosphate of Dol-P, a 14-sugar oligosaccharide unit with a pyrophosphate linkage is assembled in a stepwise manner (Figure 1). The initial seven sugar additions of this pathway utilize enzymes and substrates present on the cytoplasmic leaflet of the endoplasmic reticulum (ER) membrane. First, a residue of GlcNAc-1-P is transferred from UDP-GlcNAc to Dol-P with release of UMP. The enzyme catalyzing this reaction, GlcNAc-1-P-transferase, is selectively inhibited by tunicamycin (TN). The resulting GlcNAc-P-P-Dol is sequentially modified with an additional residue of GlcNAc from UDP-GlcNAc and five residues of mannose from GDP-mannose. The product, $\text{Man}_5\text{GlcNAc}_2\text{-P-P-Dol}$, then flips to the luminal leaflet of the ER membrane. The subsequent extension of this molecule to $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-P-P-Dol}$, the completed dolichol-linked oligosaccharide (Figure 1, violet), requires a second function of Dol-P. On the cytoplasmic leaflet, Dol-P is modified with single residues of either mannose (from GDP-mannose) or glucose (from UDP-glucose). The resulting mannose-P-Dol (MPD) and glucose-P-Dol (GPD) then flip to the luminal leaflet where they serve as enzymatic donors for the final seven sugar residues of $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-P-P-Dol}$. The specific arrangement of sugars in $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-P-P-Dol}$, including linkages and anomeric conformations, is highly conserved among most eukaryotes and can be essential for the functions of the oligosaccharide. Exceptions include most protists and some fungi, which synthesize smaller versions of the Dol-P-P-linked glycan.

Genetic defects in the synthesis of dolichol-P-P-linked oligosaccharides are responsible for the type-I congenital disorders of glycosylation (CDG). To date, 14 distinct genetic disorders (CDG types Ia–In) have been reported, all but type Ia

since 1998. The type-I CDGs affect specific sugar transferases as well as enzymes and factors involved in the synthesis of their donor substrates (Figure 1). Given the diversity of functions of N-linked glycans, it is not surprising that type-I CDG patients suffer from abnormalities of many organ systems. In contrast, the type-II CDGs affect remodeling of the oligosaccharide once transfer to newly synthesized protein has occurred, and may be caused by defects in processing glycosidases, glycosyltransferases, and nucleotide-sugar transporters.

Transfer of the Preformed Oligosaccharide Unit to Newly Synthesized Protein

The lumenally oriented oligosaccharide unit of $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-P-P-Dol}$ is transferred to newly synthesized polypeptides that are also within the lumen of the ER, with release of Dol-P-P (Figure 1). The reaction is catalyzed by the multi-subunit enzyme oligosaccharyltransferase (OT). OT transfers oligosaccharide to sterically accessible asparaginyl residues in the tripeptide context asparagine-X-serine or threonine, where X can be any residue except proline. Transfer generally occurs cotranslationally; that is, once the polypeptide has emerged inside the luminal space but before its synthesis by the ribosome is completed. However, a small portion of acceptor sites become N-glycosylated posttranslationally, after the polypeptide chain is completed. It is now thought that two separate forms of OT may handle these distinct tasks. The resulting molecule of Dol-P-P is believed to be recycled back to Dol-P and returned to the cytoplasmic face to participate in additional synthesis of $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-P-P-Dol}$, MPD, and GPD.

Glycosidase Digestion in the ER Lumen

Within minutes after transfer to protein, glycosidic digestion of the N-linked $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ (Figure 2, structure A) unit begins by the action of a family of ER-resident glycosidases. In a sequential manner, the terminal glucose residue is removed by ER glucosidase I, the remaining two glucose residues are removed by ER glucosidase II, and a single mannose is removed by ER mannosidase I. The resulting $\text{Man}_8\text{GlcNAc}_2$

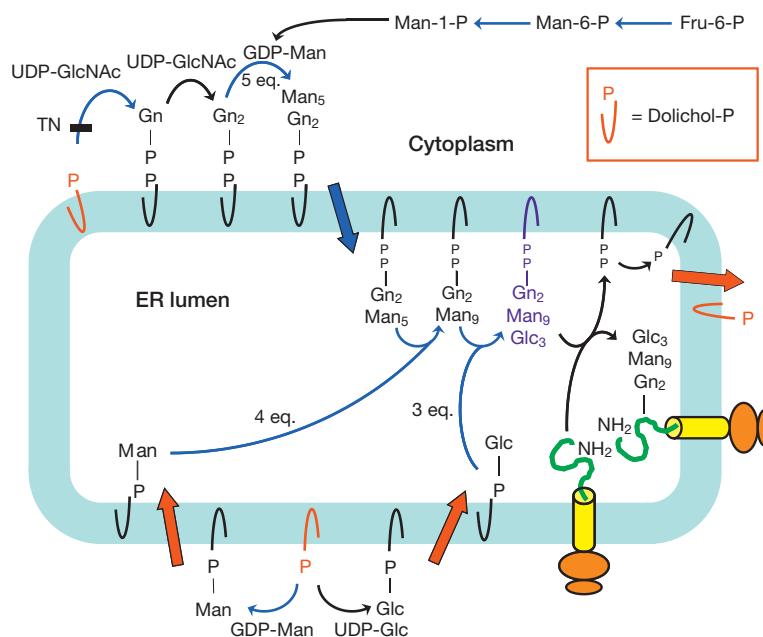


Figure 1 Biosynthesis of the lipid-linked oligosaccharide $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-PP-Dol}$ in the endoplasmic reticulum. A molecule of Dol-P (upper left-hand corner) is converted into a dolichol pyrophosphate-linked 7-sugar saccharide ($\text{Man}_5\text{GlcNAc}_2\text{-PP-Dol}$) at the cytoplasmic face of the ER membrane. It is then extended in the lumen to the 14-sugar $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2\text{-PP-Dol}$ (violet; the detailed structure of the glycan is identical to that shown in **Figure 2**, structure A). The glycan unit is transferred to polypeptides synthesized by membrane-associated ribosomes (orange) as the polypeptides (green) emerge from the translocation pore (yellow). The resulting Dol-PP is converted to Dol-P and recycled to replenish the cytoplasmically oriented pool of Dol-P (red) used for the synthesis of GlcNAc-PP-Dol , Man-P-Dol , and Glc-P-Dol . The latter two molecules are synthesized cytoplasmically and translocated to the luminal face. Dol-P resulting from their consumption is not shown. Blue arrows indicate steps believed to be defective in type I congenital disorders of glycosylation. Bold red or blue arrows represent flipping reactions. The site of action of tunicamycin (TN) is also shown.

oligosaccharide is depicted in **Figure 2** (structure B). As discussed below, the intermediates $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2$, $\text{Man}_9\text{GlcNAc}_2$, and $\text{Man}_8\text{GlcNAc}_2$ are particularly significant for folding and processing of glycoproteins. Cell-permeant inhibitors are available that block both of the ER glucosidases (castanospermine (CSN) and deoxynojirimycin (DN)), or ER mannosidase I (kifunensine (KF) and deoxymannojirimycin (DMN)). All of the N-linked oligosaccharide intermediates normally found within the ER are termed high mannose type to reflect the high proportion of mannosyl residues. Glycoproteins exported to the Golgi apparatus typically have glycans with eight mannosyl residues and no glucosyl residues, but in the event that a glycoprotein is exported before glucosidase digestion is completed, a Golgi endomannosidase exists that can remove any attached glucosyl residues plus the underlying mannosyl residue as a single unit (**Figure 3**).

Remodeling by Golgi Apparatus Glycosidases and Glycosyltransferases

The steps reviewed to this point are similar for most eukaryotes. However, within the Golgi apparatus N-linked oligosaccharides are remodeled with extensive variations in structure. These variations occur between different species, different organ systems of an individual species, different cell types of an organ system, different glycoproteins of a particular cell, and even among

different subpopulations (glycoforms) of a given glycoprotein. Many features of remodeling of N-glycans by the Golgi apparatus are covered elsewhere in this encyclopedia.

The remainder of this article is limited to remodeling in mammalian cells. It is useful to recall that the mammalian Golgi apparatus can be divided structurally and functionally into three components, known as the *cis*-, *medial*-, and *trans*-Golgi compartments, and is associated with the *trans*-Golgi network (TGN). The key enzymes involved in glycan remodeling are usually found in specific Golgi compartments or the TGN, and their physical locations in these secretory pathway organelles reflect both their order of action on secreted glycoproteins and their distinct substrate specificities (**Figure 2**). Enzymes involved in specialized remodeling (**Figures 3** and **4**) are also optimally located for accessing the correct oligosaccharide intermediate.

An elegant example of the orchestration of enzyme location and specificity comes from reactions that occur in the *cis*- and *medial*-Golgi compartments (**Figure 2**). Recall that $\text{Man}_8\text{GlcNAc}_2$ is the typical oligosaccharide intermediate on glycoproteins leaving the ER (structure B). Golgi mannosidase I, a resident of the *cis*-Golgi compartment, digests this oligosaccharide to form $\text{Man}_5\text{GlcNAc}_2$ (structure C; a different isomer than $\text{Man}_5\text{GlcNAc}_2$ formed as a dolichol-P-P-linked intermediate). Upon arrival at the *medial*-Golgi compartment, the $\text{Man}_5\text{GlcNAc}_2$ is modified by the first in a series of Golgi-associated

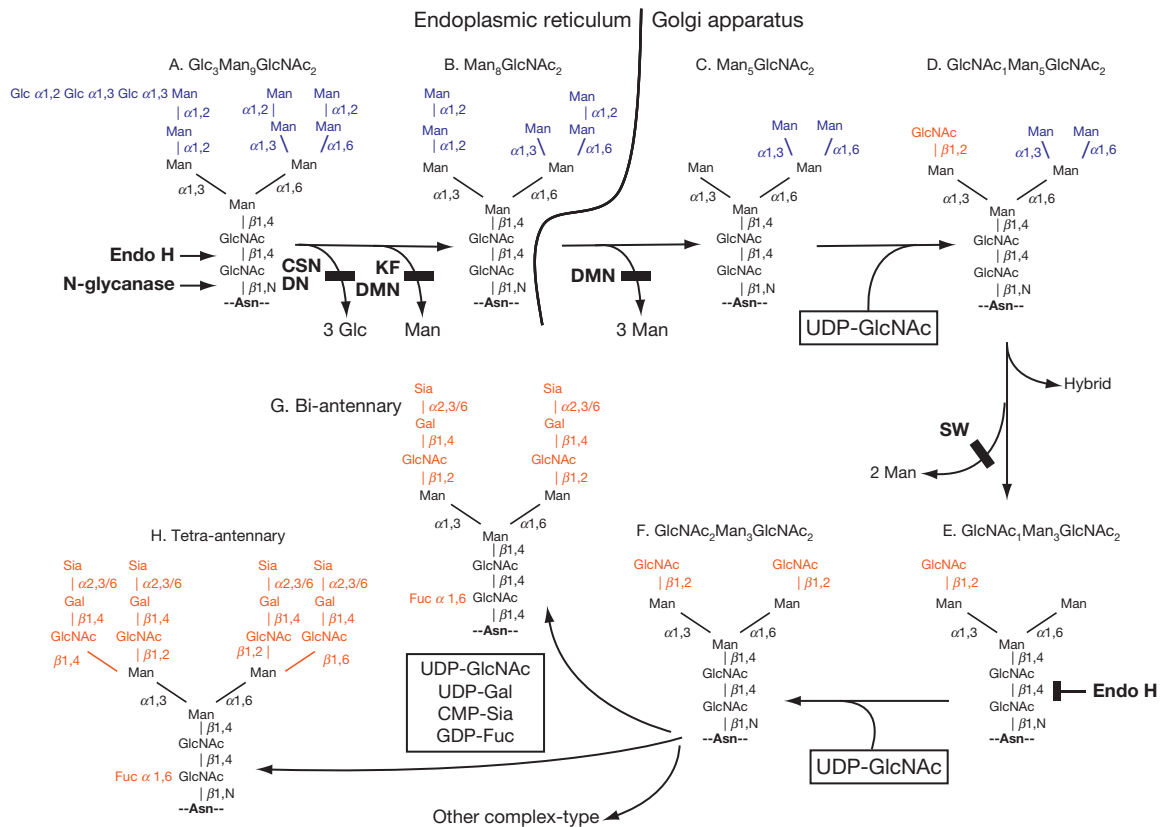


Figure 2 Remodeling of mammalian N-linked glycans in the endoplasmic reticulum, Golgi apparatus, and *trans*-Golgi network. Sugars in black form the invariant core that is retained after remodeling of N-linked glycans is completed; those in blue are removed by processing glycosidases; and those in red are added by Golgi-associated glycosyltransferases. The anomeric conformation of each glycosidic linkage is denoted α or β , with the linking group of the donor sugar preceding the comma and the linking group of the acceptor following the comma (GlcNAc, *N*-acetyl-D-glucosamine; Gal, D-galactose; Sia, sialic acid; Fuc, L-fucose). Processing of structure A ($\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$) begins by digestion with ER glucosidases I and II. Structures B through F are each formed by specific enzymes: B, ER mannosidase I; C, Golgi mannosidase I; D, GlcNAc transferase I; E, Golgi mannosidase II; F, GlcNAc transferase II. Structures A and B are associated with the ER; structure C is associated with the *cis*-Golgi apparatus; structures D–F are associated with the *medial*-Golgi apparatus; and structures G and H are associated with the *trans*-Golgi apparatus and the *trans*-Golgi network. Note that structure D can be converted to a hybrid-type glycan or, via structure E, a complex-type glycan such as those shown here (G and H). Sialic acid is commonly found in both $\alpha 2,3$ and $\alpha 2,6$ linkages on complex-type structures. The sites of cleavage by N-glycanase and endoglycosidase H (endo H) are indicated on structure A. Structure E and its products are resistant to endo H. Sites of action of the inhibitors castanospermine (CSN), deoxynojirimycin (DN), kifunensine (KF), deoxymannojirimycin (DMN), and swainsonine (SW) are shown.

glycosyltransferases, GlcNAc transferase I. Like all Golgi-associated glycosyltransferases, this enzyme uses a nucleotide-sugar donor. GlcNAc transferase I transfers a residue of *N*-acetylglucosamine (GlcNAc) from UDP-GlcNAc, forming a $\beta 1,2$ linkage with a specific mannosyl residue to yield $\text{GlcNAc}_1\text{Man}_5\text{GlcNAc}_2$ (structure D). GlcNAc transferase I is unable to recognize $\text{Man}_8\text{GlcNAc}_2$, and therefore the prior action of Golgi mannosidase I is essential. After addition of the GlcNAc residue, in most cases the $\text{GlcNAc}_1\text{Man}_5\text{GlcNAc}_2$ oligosaccharide is digested by another *medial*-Golgi enzyme, Golgi mannosidase II (which is inhibited by swainsonine (SW)), to generate $\text{GlcNAc}_1\text{Man}_3\text{GlcNAc}_2$ (structure E). Golgi mannosidase II is unable to recognize $\text{Man}_5\text{GlcNAc}_2$, and therefore requires the prior action of GlcNAc transferase I. Thus, it is clear that the subcellular locations of these enzymes must reflect their substrate specificities. Consequently, each enzyme is membrane associated and contains appropriate sorting/retention signals.

It is important to note that not all N-linked glycans are digested by Golgi mannosidases, and thus pass through the secretory pathway without further processing. This can be due to steric inaccessibility of the oligosaccharide or, as discussed later in the article, specialized competing reactions that supercede the typical processing reactions.

Digestion of $\text{GlcNAc}_1\text{Man}_5\text{GlcNAc}_2$ by Golgi mannosidase II uncovers a key mannosyl residue (structure E), allowing it to be modified by the addition of a residue of GlcNAc by GlcNAc transferase II (Figure 2). The resulting $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ oligosaccharide (structure F) has two exposed GlcNAc residues, and extension of these by additional transferases associated with the Golgi complex results in complex-type N-linked glycans such as structures G and H. As indicated earlier, the exact structures of complex-type glycans depend heavily on the organism, tissue, cell type, and protein, and are governed by which glycosyltransferases are expressed. These transferases

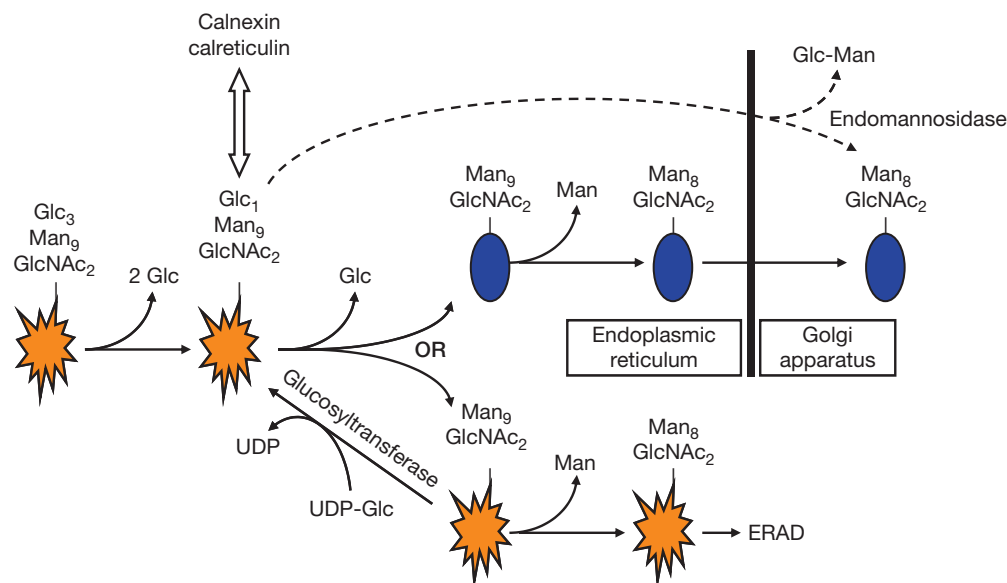


Figure 3 The roles of calnexin, calreticulin, unfolded glycoprotein glucosyltransferase, and endomannosidase in the formation and function of glycosylated N-linked oligosaccharides. A newly synthesized glycoprotein with a $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ glycan is shown in its unfolded state (orange). Removal of two glucose units forms a monoglucosylated glycan which can interact with the lectin sites of the chaperones calnexin and calreticulin. Subsequent dissociation from the lectin site allows the third glucose residue to be removed by glucosidase II. At this point, if the glycoprotein is properly folded (blue) an additional mannose unit will be removed and the glycoprotein will be exported to the Golgi apparatus. However, if the glycoprotein remains misfolded, its $\text{Man}_9\text{GlcNAc}_2$ glycan can be recognized by the unfolded glycoprotein glucosyltransferase to regenerate the monoglucosylated glycan and allow additional interaction with calnexin or calreticulin. The cycle of glucose removal and re-addition can be repeated many times for glycoproteins that are unable to fold. Eventually, on such misfolded glycoproteins one or more mannosyl units are removed from the $\text{Man}_9\text{GlcNAc}_2$ glycan to trigger destruction of the glycoprotein by a process termed ER-associated degradation (ERAD). If a glycoprotein folds properly, but exits the ER before glucose removal is completed, a Golgi-associated endomannosidase exists presumably to remove a disaccharide containing the remaining glucose unit and its underlying mannose unit (dashed arrows). Two distinct $\text{Man}_8\text{GlcNAc}_2$ isomers are formed by ER mannosidase I and Golgi endomannosidase, but each can be remodeled by Golgi-associated enzymes to form complex-type structures.

differ with regard to the sugar residue transferred (galactose, GlcNAc, fucose, GalNAc, and sialic acid), the anomeric conformation of the new linkage (α or β), the acceptor saccharide, and the hydroxyl on the acceptor that participates in the formation of the glycosidic linkage. As expected from the topological locations of the glycoprotein substrates, these transferases have catalytic sites facing the luminal spaces of Golgi compartments. The donor requirements of these transferases are met by specific transporters that import nucleotide sugars from the cytoplasm. These transporters also export the nucleotide mono- or di-phosphate by-products back to the cytoplasm. For a small fraction of oligosaccharides that fail to be digested by Golgi mannosidase II, the single exposed GlcNAc is available for extension to form hybrid-type N-linked glycans. These are so-named because they retain elements of both high-mannose-type and complex-type oligosaccharides.

A key variable among complex-type oligosaccharides is the number of branches, or antennae, that are formed. Each branch is initiated by a residue of GlcNAc attached to one of mannosyl residues in the $\text{Man}_3\text{GlcNAc}_2$ core. Complex-type structures with two to five branches have been reported, termed bi-antennary (Figure 2, structure G), tri-antennary, tetra-antennary (structure H), and penta-antennary. In most cases, sialic acid is the final sugar attached to N-linked glycans, introducing negative charges onto each branch of the complex-type oligosaccharide. However, negative charges can also result

from sulfation. It is with these final modifications that the glycoprotein is secreted or retained at the cell surface. The extent of branching appears to help regulate the functions of some surface membrane-bound glycoproteins. With more branches, cell-surface glycoproteins are better able to become engaged in lattices with lectins which also have access to the cell surface. The interactions within such lectin lattices retain glycoprotein components at the cell surface by hindering internalization. Thus, more branching may increase the surface residence time of a glycoprotein, and thereby regulate its ability to interact with hormones and other extracellular factors.

Methods for Analysis of N-Linked Glycoprotein Structure and Function

Binding to Plant Lectins

Plant lectins with a diverse range of specificities for the exposed monosaccharide, anomeric linkage, and other features of the oligosaccharide unit are commercially available. These can be used in many different ways: for example, as immobilized adsorbents attached to column matrices, as histochemical reagents when tagged with a reporter such as horseradish peroxidase, or to isolate subpopulations of cells when conjugated with fluorescent probes for flow sorting. Perhaps the best-known plant lectin is concanavalin A, which is selective for

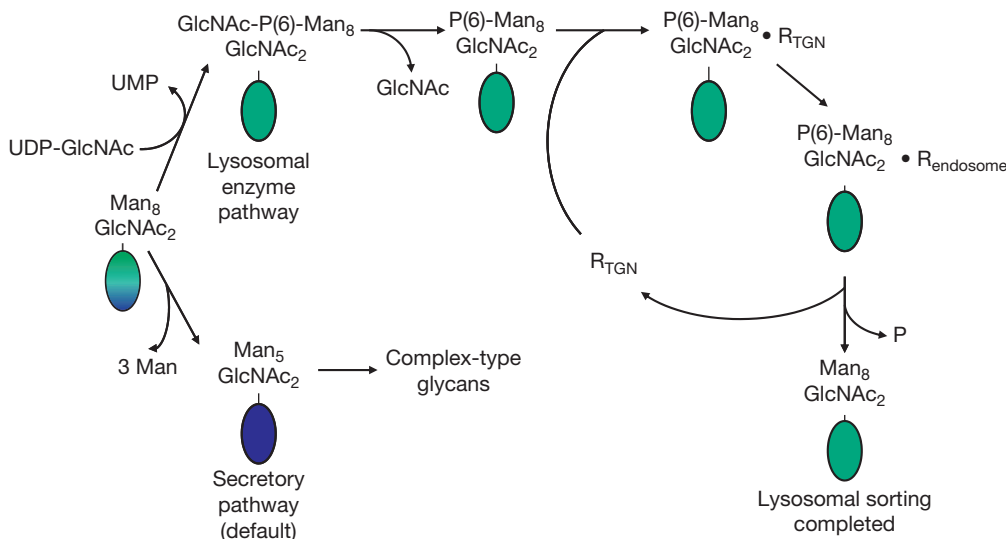


Figure 4 Formation of the mannose-6-phosphate sorting signal for lysosomal enzymes. Glycoproteins enter the *cis*-Golgi compartment with $\text{Man}_8\text{GlcNAc}_2$ glycans (Figure 2, structure B). The majority of glycoproteins (blue) are processed normally and enter the secretory pathway by default. However, a specific UDP-GlcNAc:lysosomal enzyme GlcNAc-1-P transferase recognizes protein determinants that are common to lysosomal enzymes (green), and modifies the 6-hydroxyls of mannosyl residues of the $\text{Man}_8\text{GlcNAc}_2$ glycan with GlcNAc-1-P. An uncovering enzyme then removes the GlcNAc unit, leaving a mannosyl unit with a 6-phosphate modification. In the *trans*-Golgi network (TGN), enzymes with mannose-6-P units are then bound by mannose-6-P receptors (R), which escort the enzymes to endosomes. The acidic pH of the endosome allows the enzyme-receptor complex to dissociate. The receptors recycle to the TGN, and normal endocytic flow transports the released enzymes to lysosomes where the phosphate modification is removed. Thus, mature lysosomal enzymes have no trace of the mannose-6-P sorting signal. In certain disorders, addition of the GlcNAc-1-P unit is defective, resulting in the generation of a complex-type glycan and secretion of enzymes rather than delivery to lysosomes.

oligosaccharides with exposed α -mannosyl or α -glucosyl residues. Concanavalin A binds tightly to high-mannose-type and hybrid-type oligosaccharides, and weakly to bi-antennary complex-type oligosaccharides, but not to complex-type oligosaccharides with three or more branches.

Radiolabeling of Oligosaccharides

Incorporation of an isotopic label is a simple and highly sensitive way to permit study of the oligosaccharides themselves. In principle, to label the oligosaccharide unit in intact cells a radiolabeled metabolic precursor of any of the sugar residues can be used, but in practice the choice of precursor is limited by the specific radioactivity, intracellular pool dilution, and competing metabolic fates of precursors. $[2\text{-}^3\text{H}]\text{mannose}$ is frequently used because it can label all classes of N-linked oligosaccharides with high specific activity, and labeling is selective for glycoconjugates because diversion of the mannose to glycolysis by phosphomannose isomerase results in loss of the tritium from the sugar ring. *In vitro*, oligosaccharides are typically radiolabeled by incubation with an appropriate glycosyltransferase and a nucleotide sugar with an isotopically labeled sugar residue, or by selective oxidation of hydroxyls to aldehydes followed by reduction with $[^3\text{H}]\text{NaBH}_4$.

Digestion with Purified Endo- and Exo-Glycosidases

A useful selection of glycosidases that can be used to characterize N-linked glycans is now commercially available. Endoglycosidases cleave the bonds of unexposed sugar residues; two widely

used ones are endoglycosidase H and peptide:N-glycosidase F (also known as N-glycanase) as indicated by Figure 2. N-glycanase cleaves the asparaginyl-GlcNAc bond of almost all N-linked glycans, generating a free oligosaccharide with conversion of the asparaginyl residue to aspartic acid. Endoglycosidase H cleaves the β -linkage between the two GlcNAc residues. The asparaginyl-GlcNAc bond remains intact, and an oligosaccharide released by endoglycosidase H will have one fewer GlcNAc residue than the same oligosaccharide released by N-glycanase. Importantly, endoglycosidase H only cleaves high-mannose-type and hybrid-type oligosaccharides that have not been digested by Golgi mannosidase II (Figure 2).

Exoglycosidases cleave the linkages of exposed sugar residues. Specific glycosidases are available that can remove essentially every sugar found in N-glycans, and are often used sequentially as one glycosidase exposes the substrate of another. For example, the presence of sialic acid residues on an N-glycan can be determined with enzymes called sialidases or neuraminidases. As for other families of glycosidases, these can be obtained with broad or narrow specificities for the type of glycosidic linkage. After removal of sialic acid, the nature of the underlying sugar can be deduced with other exoglycosidases.

Inhibitors of Glycan Assembly and Processing

TN, a specific inhibitor of initiation of synthesis of dolichol-P-P-linked oligosaccharides, and CSN, DN, KF, DMN, and SW, inhibitors of N-glycan processing, can be valuable tools for determining the presence or class of an N-glycan, as well as for assessing N-glycan function. These inhibitors are effective

with intact cells and purified enzymes. Their sites of action are shown in **Figures 1** and **2**.

Introduction of Mutations in Genes Encoding Glycosidases and Glycosyltransferases

A very powerful approach for the assessment of N-glycan function and structure is the use of cell lines and rodent strains in which one or more glycosidases or glycosyltransferases have been inactivated by a spontaneous mutation, a chemically induced mutation, or a targeted gene disruption. The extensive collection of N-glycosylation mutations in Chinese hamster ovary cells has proved valuable in elucidating key steps in both the biosynthetic pathway discussed earlier in the article and the secretory pathway, in addition to testing the functions of specific intermediates and production of useful glycoproteins with defined N-glycans. Many of these cell lines were isolated by their resistance to toxicity of specific plant lectins. For example, cells of the *Lec1* complementation group lack GlcNAc transferase I activity. These have been used to establish the order of reactions discussed previously, to produce N-linked glycoproteins with Man₅GlcNAc₂ glycans instead of complex-type glycans, and to establish *in vitro* conditions for transfer of glycoprotein cargo between Golgi-derived vesicles.

Chemical and Physical Analysis of N-Glycans

The combined uses of glycosylation inhibitors, genetic mutations, and glycosidases can reveal a great deal of information about the number and types of N-linked glycans attached to a protein and their functions. Detailed structural information requires removal of the oligosaccharide unit from the polypeptide, followed by chemical and physical analysis. Conventional liquid chromatography, high-pressure liquid chromatography, and gel electrophoresis have been used extensively for both purification of oligosaccharides and determination of structural features. These are effective for separation of oligosaccharides by size, charge, or differences in column adsorption due to the particular arrangements of functional groups in the molecule. Several strategies are available to determine the monosaccharide compositions of purified oligosaccharides, including gas chromatography and liquid chromatography. The actual arrangements of the sugar residues, including linkage assignments, are achieved through a combination of glycosidase digestion, gas chromatography, mass spectrometry, and/or nuclear magnetic resonance spectroscopy.

N-Linked Oligosaccharides as Experimental Aids

Purification of N-Linked Glycoproteins

The presence of an N-linked glycan gives biological chemists a simple and effective way to enrich a glycoprotein of interest, by the use of lectin-affinity chromatography. A knowledge of the subcellular location of the glycoprotein allows the likely identity of the glycan class to be surmised, and an appropriate immobilized lectin to be chosen. A column prepared with such an immobilized lectin can then be used to isolate glycoproteins bearing the appropriate N-glycan from remaining

cellular components. These glycoproteins can then be gently eluted from the adsorbent by the addition of the correct competing saccharide. For example, chromatography with immobilized concanavalin A will strongly retain glycoproteins with high-mannose-type or hybrid-type glycans, which can be recovered by elution with a solution of α -methylmannoside. A similar approach can be used to prepare an enriched fraction of proteins that have traversed the Golgi apparatus, and therefore bear complex-type glycans, by chromatography with immobilized wheat germ agglutinin and elution with the competitor disaccharide chitobiose.

Radiolabeling and Tagging of N-Linked Glycoproteins

The methods described earlier for radiolabeling of N-glycans to facilitate glycan analysis are also useful for incorporating a radiolabel for other experimental purposes, without modification of the peptide backbone. A classic example is the sequential treatment of ceruloplasmin with neuraminidase and galactose oxidase, followed by reduction of the 6-aldehyde with [³H]NaBH₄ to generate an exposed tritium-labeled β -galactosyl residue. The resulting labeled asialo-ceruloplasmin is found to be rapidly cleared from the circulation by the liver. This result allows the identification of a novel β -galactoside selective asialoglycoprotein receptor in the liver. It has become an important model for receptor-mediated endocytosis studies.

Trafficking through the Secretory Pathway

As discussed earlier, N-linked glycan structures are modified as glycoproteins pass through the various compartments of the mammalian secretory pathway (**Figure 2**). This feature makes N-linked glycans extremely useful for monitoring the subcellular location of a glycoprotein of interest. The presence of a glycoprotein in almost any compartment along the secretory pathway can be determined by information gained from binding to plant lectins, reactivity with glycosidases, and/or labeling with sugar precursors. The most common use of these techniques takes advantage of the sensitivity of high-mannose-type glycans to endoglycosidase H, and the resistance to this enzyme that occurs after digestion with Golgi mannosidase II. Thus, the acquisition of endoglycosidase H sensitivity by a newly synthesized protein signifies that it is in fact a glycoprotein, and that a portion or all of its sequence is present in the ER luminal space. The subsequent loss of endoglycosidase H sensitivity reflects the arrival of the protein at the medial compartment of the Golgi apparatus (for most glycoproteins, 30–60 min after synthesis).

Functions of Selected N-Linked Glycans and Their Specific Cognate Lectins

Since the 1990s, there has been an explosion of discoveries of specific functions for protein-linked glycans, both asparagine-linked and serine/threonine-linked. Some functions of N-linked glycans are covered, in detail, elsewhere in this encyclopedia. A central concept revealed by these studies is that glycan function usually requires recognition by a specific cognate lectin. The ability to identify biologically significant glycan-lectin pairs

has been aided tremendously in recent years by the advent of glycan arrays, a powerful technology which involves large numbers of different chemically defined glycans immobilized into sectors on slides. These arrays are then queried with a fluorescently tagged lectin of interest, sometimes produced recombinantly from genetic material and with no other information other than that deduced from the predicted amino acid sequence. Glycan-lectin pairs are then identified with specialized fluorescence detection equipment which determines the sectors that have bound the lectin. By similar strategies, arrays of immobilized lectins can be queried with fluorescently labeled glycans to also identify glycan-lectin pairs.

In this section, two functions (albeit discovered prior to the advent of arrays) that highlight reactions covered in earlier sections are selected for discussion. These two examples illustrate how certain glycosyltransferases may recognize features of the polypeptide in addition to the glycan, and thus restrict modification to only special subclasses of glycoproteins. They also show how certain key modifications are transient, and are not found on the mature glycoproteins.

Monoglucosylated Oligosaccharides in ER Quality Control

As discussed earlier, removal of two glucosyl residues from the $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ N-linked oligosaccharide by the CSN-sensitive glucosidases I and II results in a monoglucosylated-processing intermediate, $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2$. This intermediate is recognized by calnexin and calreticulin, resident ER chaperones with lectin activities specific for monoglucosylated oligosaccharides (Figure 3). Interactions with one of these lectin-chaperones can be crucial for folding and processing of nascent N-linked glycoproteins. Calnexin and calreticulin have essentially negligible recognition of $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ (as well as oligosaccharides with zero or two glucose residues). Thus, treatment of cells with CSN blocks glycan-dependent interactions of chaperones with misfolded nascent N-linked glycoproteins because it prevents the formation of $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2$ N-linked oligosaccharides. Calnexin and calreticulin bind to $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2$ reversibly, and once dissociated from the lectin site of calnexin or calreticulin, the $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2$ glycan can be digested to $\text{Man}_9\text{GlcNAc}_2$ by glucosidase II according to the scheme discussed earlier. After release of the glycan from the lectin site, calnexin and calreticulin appear to mediate further folding of their glycoprotein substrates due to lectin-independent chaperone activities.

In addition to processing of $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$, $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2$ can be formed by reglucosylation of N-linked $\text{Man}_9\text{GlcNAc}_2$. This is achieved by a remarkable enzyme, the UDP-glucose:unfolded glycoprotein glucosyltransferase (Figure 3). The glucosyltransferase has two functional domains. A catalytic domain mediates the transfer of glucose from UDP-glucose to $\text{Man}_9\text{GlcNAc}_2$, forming the same $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2$ structure as achieved by processing of $\text{Glc}_2\text{Man}_9\text{GlcNAc}_2$. A second recognition domain is able to identify physical features characteristic of misfolded proteins, presumably exposed hydrophobic surfaces. The binding of a $\text{Man}_9\text{GlcNAc}_2$ -bearing misfolded glycoprotein to the recognition domain allows reglucosylation by the catalytic domain. The regenerated $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2$ glycan then gives the misfolded glycoprotein another opportunity to interact with calnexin/

calreticulin and be folded properly. The repeated actions of the lectin-chaperones calnexin and calreticulin, of glucosidase II, and of the ER glucosyltransferase account for the observation that certain glycoproteins with mutations that affect folding may retain the $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2$ glycan for extended periods. Properly folded glycoproteins can be exported to the Golgi apparatus, while glycoproteins unable to achieve a properly folded conformation are destroyed by a process termed ER-associated degradation (ERAD) (Figure 3).

High-Mannose Oligosaccharides Bearing Mannose-6-Phosphate Residues as Sorting Signals for Lysosomal Hydrolases

The processing of $\text{Man}_8\text{GlcNAc}_2$ to $\text{Man}_5\text{GlcNAc}_2$ by the medial-Golgi enzyme mannosidase I was discussed earlier. However, hydrolases destined for the luminal space of the lysosome undergo a very different fate (Figure 4). On these proteins, the 6-hydroxyl groups of mannosyl residues are phosphorylated, producing high-mannose oligosaccharides containing mannose-6-phosphate. The mannose-6-phosphate then serves as a specific trafficking signal and allows the hydrolase to be deposited in the lysosome.

The mannose-6-phosphate recognition marker is created in two steps. First, GlcNAc-1-P is transferred to the 6-hydroxyl of a mannosyl residue of $\text{Man}_8\text{GlcNAc}_2$ by a specialized transferase, the UDP-GlcNAc:lysosomal enzyme GlcNAc-1-P transferase. Like the UDP-glucose:unfolded glycoprotein glucosyltransferase, this enzyme has two functional domains. The catalytic domain mediates transfer of GlcNAc-1-P to high-mannose-type glycans. The recognition domain binds to lysosomal hydrolases due to a special arrangement of amino acid side chains found only on these enzymes. Thus, the GlcNAc-1-P transferase is effective only with lysosomal hydrolases bearing high-mannose-type glycans. The GlcNAc-1-P transferase is found in the cis-Golgi compartment; thus, its reaction with $\text{Man}_8\text{GlcNAc}_2$ glycans can take place before digestion with Golgi mannosidase I. In the second step, the product is the substrate for another enzyme, known as the uncovering enzyme, which removes the GlcNAc residue and exposes the phosphate. Unlike the GlcNAc-1-P transferase, the uncovering enzyme has a catalytic domain but no known polypeptide recognition domain. The resulting mannose-6-phosphate-labeled hydrolase continues its movement along the secretory pathway, but undergoes no additional modifications of its glycan.

In the TGN, receptors for glycans with mannose-6-phosphate bind the hydrolases and escort them to endosomes. Usually, at least two mannose-6-phosphates must be generated per hydrolase for effective receptor binding. The acidic pH of the endosome causes release of the hydrolase from the receptor, which then recycles back to the TGN for additional rounds of sorting. The hydrolase is then deposited in the lysosome by normal vesicular traffic, and phosphatases remove the phosphate modification. Thus, mature lysosomal hydrolases bear high-mannose-type glycans without phosphate. Elucidation of this unique process has been aided by the discovery that it is disrupted in patients with I-cell disease and pseudo-Hurler polydystrophy. In these patients, heritable mutations cause loss of UDP-GlcNAc:lysosomal enzyme GlcNAc-1-P transferase activity. As a result, the mannose-6-phosphate sorting signal

is not formed and lysosomal enzymes are secreted instead of being deposited in lysosomes. This accounts for the storage of pathological waste products in the lysosomes of I-cell and pseudo-Hurler patients.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Glycosylation, Congenital Disorders of; Sugar Nucleotide Transporters; Glycoproteins, Plant.

Further Reading

- Beeley JG (1985) *Glycoprotein and Proteoglycan Techniques*. New York, NY: Elsevier.
- Lehrman MA (2001) Oligosaccharide-based information in endoplasmic reticulum quality control and other biological systems. *Journal of Biological Chemistry* 276: 8623–8626.
- Lennarz WJ and Hart G (eds.) (1994) *Methods in Enzymology*, vol. 230. New York, NY: Academic Press.
- Varki A, Cummings R, and Esko J, et al. (eds.) (1999) *Essentials of Glycobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Voet D and Voet JG (1995) *Biochemistry*, 2nd edn. New York, NY: Wiley.

Glycoproteins, Plant

C J Schultz, University of Adelaide, Glen Osmond, SA, Australia

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 293–296, © 2004, Elsevier Inc.

Glossary

Arabinogalactan protein (AGP) A heavily glycosylated cell-surface glycoprotein composed of large branched polysaccharide chains (containing predominantly arabinosyl and galactosyl residues) attached to hydroxyproline residues.

Extensin A moderately glycosylated cell-surface glycoprotein containing short linear arabinosyl chains attached to hydroxyproline residues and which may also contain single galactosyl residues attached to Ser residues.

Glycoprotein A compound containing carbohydrate (or glycan) covalently linked to protein (International Union of Pure and Applied Chemistry). The carbohydrate may be in the form of a monosaccharide, disaccharide(s), oligosaccharides, and polysaccharides.

Glycosylphosphatidylinositol (GPI) anchor A lipid modification added to the C terminus of selected proteins that anchors them in the outer layer of the lipid bilayer of the plasma membrane.

Types of Glycosylation

The classification of glycans as *N*-linked or *O*-linked reflects the type of covalent linkages between the amino acid residue in the core protein, and the first sugar of the glycan chain.

N-Linked Glycans

N-linked glycans are attached to core proteins through an amide nitrogen (N) on the side chain of selected asparagine (Asn) residues. They are attached to some, but not all, Asn residues in the motif Asn–Xaa–(Ser/Thr), where Xaa can be any amino acid except proline (Pro). Although the motif for the attachment of *N*-linked glycans is conserved in eukaryotes, the structure of the *N*-linked glycans differs in different organisms.

Attachment of an *N*-linked glycan is a multistep process. The first step is the addition of a 14-sugar preformed precursor oligosaccharide. This oligosaccharide contains *N*-acetylglucosamine (GlcNAc), mannose (Man), and glucose (Glc) and is the same in most eukaryotes (Figure 1(a)). The oligosaccharide is subsequently modified by glycosidases that remove selected sugars and by glycosyltransferases that selectively add sugars to specific positions.

The modifications can be temporally and spatially regulated, and are different in different organisms. An example of a complex-type *N*-linked glycan found in plants is shown in Figure 1(b). The Man₃GlcNAc₂ core (boxed) is found in all eukaryotic *N*-linked glycans; however, the terminal α -1,3-linked fucose and β -1,2 linked xylose residues attached to the Man₃GlcNAc₂ core are specific to plants. These plant-specific sugars on *N*-linked glycan are believed to contribute to the response experienced by people who are allergic to pollen and certain plant foods.

N-linked glycosylation is a common motif occurring on many plant proteins. *N*-linked glycans assist protein folding and are important for ensuring that only correctly folded proteins pass through the secretory system.

O-Linked Glycans

O-linked glycans are attached to the core protein through the oxygen (O) moiety of selected serine (Ser), threonine (Thr), and hydroxyproline (Hyp) residues. Hyp is a posttranslational modification of Pro that is commonly glycosylated in plants. Although Hyp is found in animal proteins (e.g., collagen), glycosylation of Hyp does not occur.

Glycosylation of Hyp exists in two different forms that depend on the arrangement of Hyp residues in the core protein. The first type of *O*-linked glycans attached to Hyp are short Ara chains (one to four) and are typically found on extensins (Figure 1(b)). In contrast, the arabinogalactan proteins (AGPs) contain large arabinogalactan (AG) polysaccharides attached to Hyp residues (Figure 1(c)). *O*-linked glycosylation of Thr is rare in plants, although it is common in animals. Glycosylation of Ser in plants is also relatively rare and usually consists of the addition of a single galactosyl residue.

In general, glycoproteins that contain *O*-linked glycans are destined for the cell surface and are found in the cell-wall/extracellular matrix (ECM), or in vesicles in transit to/from the cell surface. The addition of carbohydrate to proteins provides unique structural properties and informational complexity to the cell. The remainder of this article focuses on the *O*-linked glycosylation of Hyp residues.

Synthesis of *O*-Linked Glycans

Hydroxylation of Pro

For proteins to be glycosylated, they must enter the secretory system via the endoplasmic reticulum (ER) and subsequently the Golgi apparatus. Targeting is achieved by an N-terminal signal sequence of 20 amino acids in length and this signal is subsequently cleaved. In plants, the hydroxylation of Pro occurs in the ER as it does for animals.

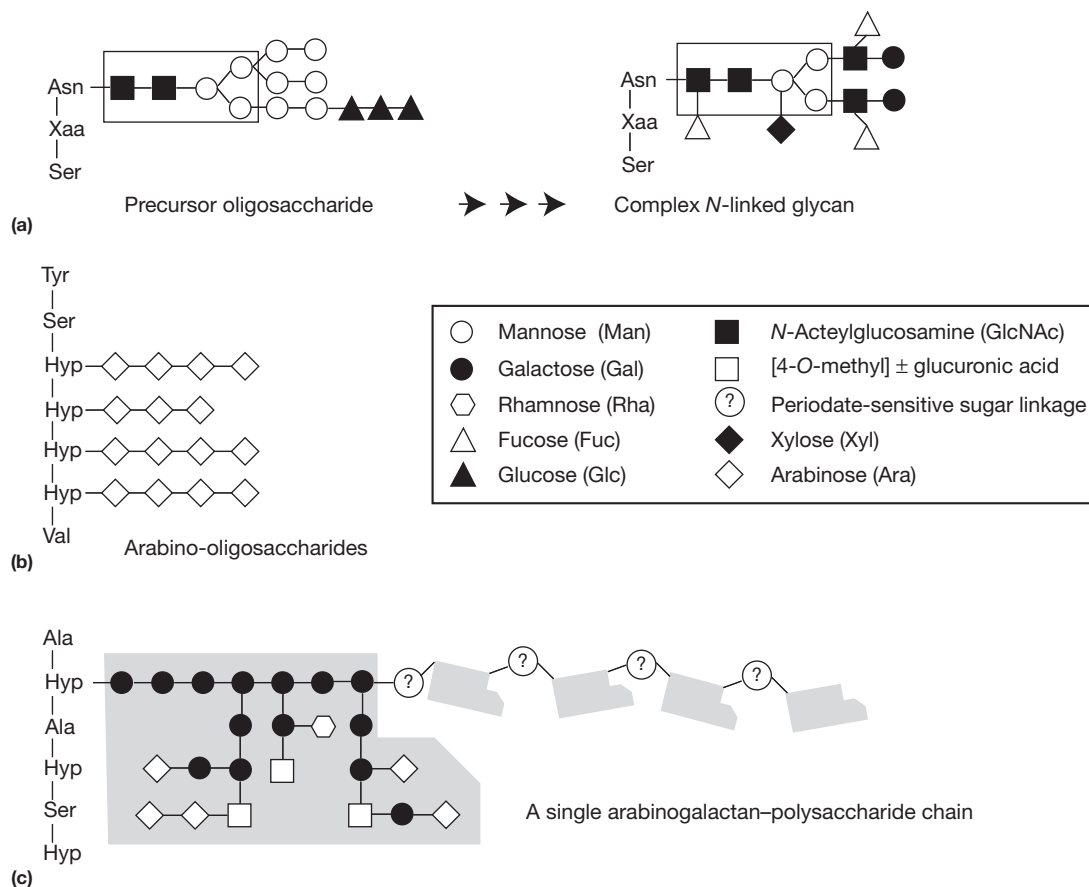


Figure 1 (a) The first step in *N*-linked glycosylation is the addition of a 14-sugar preformed precursor oligosaccharide. This is subsequently modified by a series of glucanases and glycosyltransferases to produce a variety of *N*-linked glycans. An example of a complex-type *N*-linked glycan from plants is shown. The five core sugars (boxed) are conserved in plants, insects, and mammals. (b) An example of the *O*-linked glycosylation that occurs on extensins. Attached to the contiguous (uninterrupted) Hyp residues are one to four (1→2) α -L-Araf residues. (c) An example of the *O*-linked glycosylation found on AGPs. Attached to the predominantly noncontiguous Hyp residues are repeats of (1→3) β -linked Galp oligosaccharides with a degree of polymerization of about seven (large shaded region), separated by a periodate-sensitive linkage. The branching from the (1→3) β -linked Galp oligosaccharide backbone (as seen in the large shaded region) can occur from any Gal in the backbone. There may be as many as 10 repeats (only 5 are shown here). For simplicity, only a single Hyp residue has a polysaccharide chain attached. Reprinted from Gaspar YM, Johnson KL, McKenna JA, Bacic A, and Schultz CJ (2001) The complex structures of arabinogalactan-proteins and the journey towards a function. *Plant Molecular Biology* 4: 161–176.

Glycosyltransferases

O-linked glycosylation is poorly understood in plants. It is believed to occur sequentially with the addition of single sugars by a multitude of glycosyltransferases, starting in the ER and then further elaborated in the Golgi apparatus. The recent sequencing of several plant genomes has identified hundreds of plant glycosyltransferases. The lack of knowledge of these enzymes and the abundant heterogeneity of the *O*-linked glycans in plants means that it will be many years before we have a good understanding of the structure and synthesis of these glycans in plants.

Arrangement of Hyp Residues Dictates the Type of Glycans

Our understanding about the arrangement of glycans on the Hyp-rich glycoproteins is more advanced than our understanding of the individual glycan structures. In 1994,

Marcia Kieliszewski and Derek Larpport developed the Hyp contiguity hypothesis. This hypothesis suggests that contiguous (uninterrupted) Hyp residues are glycosylated with arabino-oligosaccharides (Figure 1(b)) and noncontiguous Hyp residues are glycosylated with AG polysaccharides (Figure 1(c)). Experimental support for this hypothesis continues to grow and recent work by Kieliszewski and colleagues shows that *O*-linked oligosaccharides and AG polysaccharides can be found in the same molecules, for example, LeAGP1 from tomato and gum arabic from *Acacia senegal*.

The Hyp contiguity hypothesis is useful for predicting *O*-linked glycosylation, although the precise rules governing the minimum length of clustered noncontiguous Hyp residues is not yet established. This is useful for modern biology where genes are generally characterized before the corresponding protein. Figure 2 shows the predicted arrangement of glycosylation of the major classes of Hyp-rich glycoproteins in *Arabidopsis thaliana*.

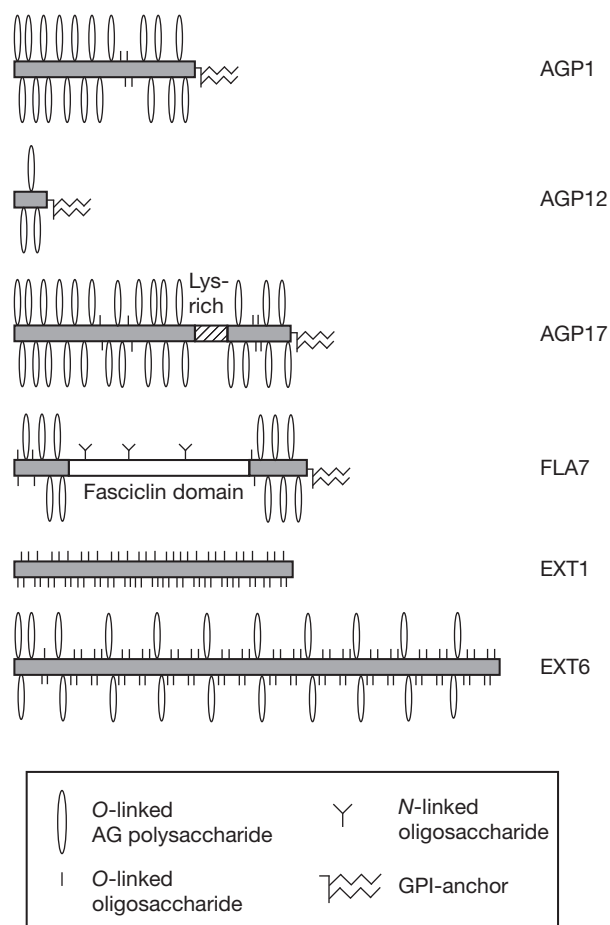


Figure 2 Schematic representation of the major classes of glycoproteins from *A. thaliana*. The models are based on AGP1 (At5g64310), AGP12 (At3g13520), AGP17 (At2g23130), FLA7 (At2g04780), EXT1 (At1g76930), and EXT6 (At2g24980), as downloaded from the Internet. For the AGPs and FLA7, all of the predicted Hyp glycosylation sites are shown. For the longer extensin protein backbones, the relative placement of the predicted arabinosaccharides and the AG polysaccharides is shown, but the number of glycan chains is not accurate. Not drawn to scale.

Hyp-Rich Glycoproteins Belong to Multigene Families

It is 10 years since the first extensin and AGP core protein genes were cloned from plants. The completion of the sequencing of the first plant genome, from *A. thaliana*, has provided a better understanding of the size and complexity of the Hyp-rich glycoprotein gene families. For example, the AGP gene family consists of 50 different genes in four distinct subclasses, and there are at least 20 different extensin genes. A major challenge for the future is determining the function of each family member.

Approaches for Determining the Function of Plant Glycoproteins

Structural Studies

The two best-studied classes of plant glycoproteins, the extensins and AGPs, have different structures and it is widely

accepted that they have different functions. Extensins are an integral part of the insoluble plant cell wall. The addition of arabinosaccharides to the extensin core proteins induces an extended polyproline-II helical structure giving extensins a rod-like property. The Ser-Hyp₃₋₄ regions of extensins are interspersed with amino acid residues (e.g., Tyr and Lys) that are important for cross-linking.

In contrast, AGPs are highly soluble components of the ECM. Rather than forming rods, they are believed to be almost spherical. In addition to polysaccharides, many AGPs also possess a lipid modification at their C terminus, in the form of a glycosylphosphatidylinositol (GPI) anchor. The finding that AGPs are GPI anchored helped explain their frequent association with the plasma membrane and provided additional clues to how they might be involved in signaling.

GPI anchors confer many different properties to animal proteins. These include (1) regulated release from cell surfaces, (2) inclusion in lipid rafts where they may be important components of signaling cascades, and (3) polarized targeting of proteins to specific cell surfaces. Although none of these features have been confirmed for AGPs, there is recent evidence suggesting that a GPI-anchored plant protein, COBRA, is polarized in specific cells of the root.

Immunohistochemistry

Much of the original interest in AGPs came from their abundance in female reproductive tissues. The two most commonly used tools for detecting AGPs are monoclonal antibodies that recognize carbohydrate epitopes, and the artificial carbohydrate antigen, β -glucosyl Yariv reagent (β -glcY). A variety of experiments using these tools suggest that AGPs are important for cell expansion and somatic embryogenesis. However, because these tools interact with carbohydrate epitopes present on multiple AGPs, they have limited use in determining the function of individual AGPs.

Genetic Approaches

Genetic approaches offer the best hope for finding the function(s) of individual plant glycoproteins. Identification of insertion (knockout) mutants in specific genes by sequencing the DNA flanking the site of DNA insertions is well advanced in *A. thaliana*. Current research efforts around the world aim to identify insertion mutants for every gene by 2010. By studying the phenotype of knockout mutants for genes involved in the synthesis and assembly of plant glycoproteins, it should be possible to determine the function of individual family members.

Future Studies

Once we know the function of individual glycoproteins, we can begin to answer the key question of whether the function of each glycoprotein resides in the protein core or in the microheterogeneity of the glycan epitopes, or in both. It is now widely recognized that carbohydrate epitopes on mammalian, *Drosophila*, and protozoan parasite glycoproteins are key determinants of biological specificity/function. The carbohydrate component of plant glycoproteins is expected to be equally important.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Free Oligosaccharides: Formation and Degradation (Free Oligosaccharides Related to N-Linked Glycans); Glycoprotein-Mediated Cell Interactions, Nonmucin-Type O-Linked; Glycosylphosphatidylinositol Anchors; N-Linked Glycan-Processing Glucosidases and Mannosidases.

Further Reading

- Buchanan BB, Gruissem W, and Jones RL (eds.) (2000) *Biochemistry and Molecular Biology of Plants*. Rockville: American Society of Plant Physiologists.
- Fincher GB and Stone BA (1983) Arabinogalactan-proteins: Structure, biosynthesis and function. *Annual Review of Plant Physiology* 34: 47–70.
- Gaspar YM, Johnson KL, McKenna JA, Bacic A, and Schultz CJ (2001) The complex structures of arabinogalactan-proteins and the journey towards a function. *Plant Molecular Biology* 47: 161–176.
- Haltiwanger RS (2002) Regulation of signal transduction pathways in development by glycosylation. *Current Opinion in Structural Biology* 12: 593–598.

- Johnson KJ, Jones BJ, Schultz CJ, and Bacic A (2003) Non-enzymic cell wall (glyco) proteins. In: Rose J (ed.) *The Plant Cell Wall*. Oxford: Blackwell.
- Lerouge P, Cabanes-Macheteau M, Rayon C, et al. (1998) N-glycoprotein biosynthesis in plants: Recent developments and future trends. *Plant Molecular Biology* 38: 31–48.
- Schindelman G, Morikami A, Jung J, et al. (2001) COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in *Arabidopsis*. *Genes and Development* 15: 1115–1127.
- Schultz CJ, Rumsewicz MP, Johnson KL, et al. (2002) Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case. *Plant Physiology* 129: 1448–1463.
- Wilson IBH (2002) Glycosylation of proteins in plants and invertebrates. *Current Opinion in Structural Biology* 12: 569–577.
- Zhao ZD, Tan L, Showalter AM, Lamport DT, and Kieliszewski MJ (2002) Tomato LeAGP-1 arabinogalactan-protein purified from transgenic tobacco corroborates the Hyp contiguity hypothesis. *Plant Journal* 31: 431–444.

Relevant Website

<http://www.tigr.org> — J. Craig Venter Institute.

Glycosylation, Congenital Disorders of

B G Ng and H H Freeze, Sanford – Burnham Medical Research Institute, La Jolla, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Hudson H. Freeze, volume 2, pp. 302–307, © 2004, Elsevier Inc.

Glossary

Congenital disorders of glycosylation Inherited disorders of protein glycosylation involving the synthesis and transfer of the Dol-PP-oligosaccharide precursor sugar chain to protein or the subsequent processing of these *N*-linked oligosaccharides. The term can also apply to the addition of other types of sugar chains to proteins.

α -Dystroglycan An extracellular dystrophin–glycoprotein complex (DGC) component which is one of the major proteins known to contain O-mannose-based oligosaccharides. Mutations in genes that encode proteins needed for this type of glycosylation cause several types of muscular dystrophy.

DGC A large protein complex found in muscle, nerve, heart, and brain. It is composed of extracellular, transmembrane, and intracellular proteins. Mutations in different proteins of this complex cause different types of muscular dystrophy.

Glycosaminoglycan (GAG) A large complex sugar chain composed of alternating amino-sugar and uronic acid units.

Glycosylation The process of adding single (monosaccharides) or multiple sugars to proteins or lipids.

Glycosyltransferase An enzyme that transfers a sugar from an activated nucleotide donor to an acceptor protein, carbohydrate, or lipid.

Lipid linked oligosaccharide (LLO) Structure that is the precursor for *N*-linked glycans; it is composed of a sugar chain linked to dolichol through a pyrophosphate linkage.

Multiple hereditary exostosis (MHE) A dominant inherited disorder caused by mutations in two genes (EXT1 and EXT2) that encode a polymerase needed for the synthesis of heparan sulfate. Mutations cause painful recurrent outgrowths of bone and must be surgically removed.

Oligosaccharyl transferase complex A protein complex found in the endoplasmic reticulum where it recognizes the LLO sugar chain and attaches it to the newly synthesized proteins having the N-X-Thr/Ser consensus sequence for *N*-glycosylation.

Transferrin (Tf) An iron-carrying plasma glycoprotein that is especially susceptible to incomplete *N*-glycosylation caused by an insufficient amount of LLO or one that is inefficiently transferred to proteins.

All cells are covered with a dense forest of glycosylated molecules that consist of glycoproteins and glycolipids. These frostings often dictate and monitor the stability of newly synthesized proteins and glycolipids within the cell. On the surface, they help cells communicate with their neighbors to assess the state of the local neighborhood. Hundreds of enzymes are used in making the sugar chains and adding them to thousands of proteins. Inherited mutations in some of the biosynthetic enzymes cause human diseases that affect the function of many organs during embryonic development, childhood, and even as adults. This article covers these inherited protein glycosylation diseases and show where glycoobiologists studying sugar chain biosynthesis and function team up with physicians to define and understand the basis of these diseases.

Setting the Stage

About 1–2% of the human genome encodes proteins that synthesize or recognize sugar chains. Hundreds of different sugar chains decorate the surface of eukaryotic cells. The immense variety of sugar chains stems from variability in length (number of monosaccharides), kinds of linkage (alpha versus beta), spatial orientation (linkage) of the sugar, branching patterns (two or more sugars attached to another sugar),

and modifications by nonsugar elements (e.g., phosphate or sulfate). Adding each of these sugars or modification requires one or more separate enzymes. Sometimes coordinated groups of enzymes work in a specific order to add some sugars, or cleave others from the maturing chain. Cells express different complements of sugar biosynthetic enzymes (glycosyl transferases and glycosidases) and glycosylation of a single protein can vary in different cells. Moreover, the same sugar chain can have different functions depending on the protein it is linked to, and the cell expressing it. Loss of glycosylation capacity can have serious effects in some cells and much milder effects in others. This complexity means that the function of sugar chains has to be investigated case by case, protein by protein.

The linkage joining the sugar chain to the protein defines the type of glycosylation, as shown in [Table 1](#). Most of the glycosylation disorders have been identified in the *N*-linked oligosaccharide biosynthetic pathway since 1995. In 2002, incomplete glycosylation of α -dystroglycan with O-mannose-based sugar chains was found to be the cause of rare forms of muscular dystrophy. Defects in the biosynthesis of chondroitin and dermatan sulfate chains also cause diseases that affect growth of bone and connective tissues. Abnormalities in O-GalNAc modification can give rise to the autoimmune disease Tn syndrome and familial tumoral calcinosis. The failure to add phosphate or sulfate to sugar chains also causes glycosylation-based disorders.

Table 1 Major types of glycan–protein linkages

Glycan type	Linkage	Typical proteins	Location
N-linked	GlcNAc- β -Asn	Cell-surface receptors secreted proteins	All cells, blood, subcellular compartments (e.g., lysosomes)
O-linked	GalNAc- α -Ser/Thr	Secreted and cell-surface mucins	Cell-surface, gastrointestinal, and reproductive tracts
O-linked	Xyl- β -Ser	Chondroitin/dermatan sulfate	Extracellular matrix, cartilage
O-linked	Man- α -Ser/Thr	Heparan sulfate/heparin α -dystroglycan	Muscle, nerve cells

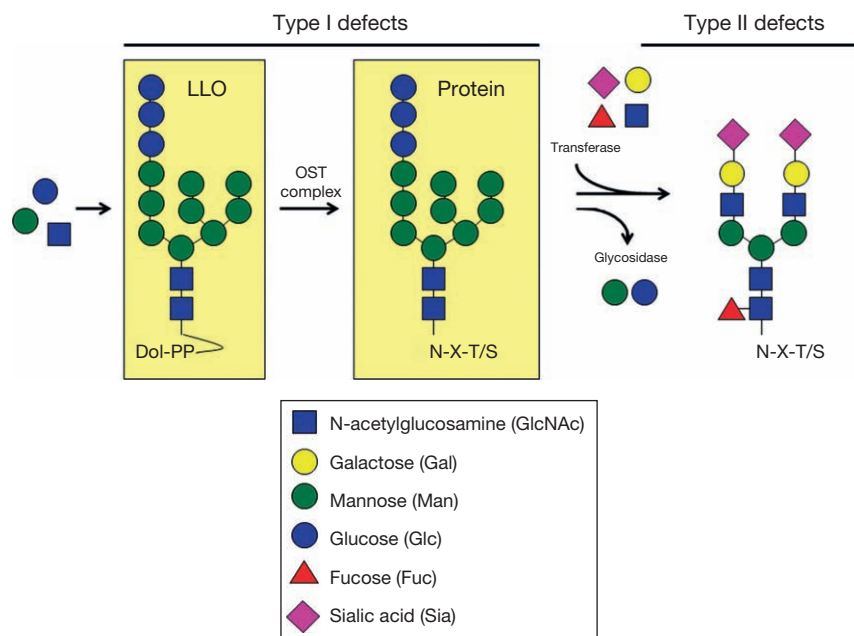


Figure 1 The types of congenital disorders of glycosylation (CDG). CDG types are based on finding mutations in genes whose products are involved in glycosylation. Type-I CDGs are defined by mutations in steps leading to the assembly and transfer of the lipid linked oligosaccharide chain (LLO) from the carrier lipid to potential N-glycosylation sites (N-X-T/S) on newly synthesized proteins in the ER. At the extreme left, the various monosaccharides (see symbol key) are first activated, and then they are incorporated one-by-one into a growing chain assembled on the carrier lipid, Dol-PP. The complete sugar chain is then transferred to the protein in a single step. Approximately, 40 genes are needed to carry out the first stage. Type-II CDG defects are defined as those that involve the sequential, highly ordered removal, and addition of individual sugars on protein-bound N-linked sugar chains. More than 20 additional genes are needed to synthesize a sugar chain typical of mammalian plasma glycoprotein, shown on the extreme right side. It is unclear exactly how many genes may truly be involved in this process, as the list continues to grow each year. To date, 11 separate type-II CDGs have been identified, and it is likely that many more will be found.

Congenital Disorders of Glycosylation

N-Glycosylation Overview

All N-linked chains are derived from a common dolichyl pyrophosphate (Dol-PP)-linked 14-sugar oligosaccharide (lipid-linked oligosaccharide (LLO)). It is composed of three glucose (Glc), nine mannose (Man), and two N-acetylglucosamine (GlcNAc) residues, forming a $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ chain that is transferred *en bloc* from the lipid carrier to proteins. Addition of each sugar unit occurs in a preferred order and requires at least 13 glycosyltransferases. The completed oligosaccharide chain is then transferred to the nascent polypeptide chain by the oligosaccharyl transferase (OST) complex located in the endoplasmic reticulum (ER) membrane. After the transfer, the sugar chain is processed. Specific glycosidases in the ER trim the $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ chain, removing all Glc and some Man. As the proteins move to the Golgi, additional Man is often carved away and the chain is extended once again by addition

of two to four branches composed of GlcNAc, galactose (Gal), and sialic acid (Sia) to form complex-type sugar chains (Figure 1). Literally, hundreds of different sugar chain structures can be made by variations of this pathway.

Scope of the Disorders

Congenital disorders of glycosylation (CDG) result from genetic defects in the assembly, transfer, and processing of N-linked oligosaccharides. For convenience, the CDGs are divided into groups I and II. Originally, defective genes were lettered in chronological order of their discovery. More recently, a new nomenclature has been suggested to accommodate the rapidly rising number of new CDG genes (see, Table 2). The first group includes defects in the assembly of the LLO and its transfer to proteins, which probably requires at least 40 nonredundant genes. Defects in 18 of these genes have been shown to cause CDG. These include defects in the mobilization

Table 2 Inherited glycosylation disorders

<i>Disorders</i>		<i>Enzymatic or protein defect</i>	<i>Gene</i>	<i>OMIM^a</i>	<i>Frequent clinical features</i>
<i>CDG type</i>	<i>Formerly</i>				
PMM2-CDG	CDG-Ia	Phosphomannomutase Man-6-P→Man-1-P	<i>PMM2</i>	212065	Variable intellectual disability, hypotonia, seizures, coagulopathy, feeding problems, and liver fibrosis.
MPI-CDG	CDG-Ib	Phosphomannose isomerase Fru-6-P⇌Man-6-P	<i>MPI</i>	601785	Normal development, hypoglycemia, coagulopathy, diarrhea protein-losing enteropathy, vomiting, and liver fibrosis.
ALG6-CDG	CDG-Ic	Alpha-1,3-glucosyltransferase	<i>ALG6</i>	603147	Moderate intellectual disability, hypotonia, epilepsy, and esotropia
ALG3-CDG	CDG-Id	Alpha-1,3-mannosyltransferase	<i>ALG3</i>	601110	Severe psychomotor delay, optic atrophy, microcephaly, iris coloboma, and hypsarrhythmia
DPM1-CDG	CDG-Ie	Dolichol-P-Man synthase	<i>DPM1</i>	603503	Severe intellectual disability, hypotonia, epilepsy, mild dysmorphism, and coagulopathy
MPDU1-CDG	CDG-If	Dolichol-P sugar utilization protein	<i>MPDU1</i>	604041	Intellectual disability, short stature, ichthyosis, and pigmentary retinopathy
ALG12-CDG	CDG-Ig	Alpha-1,6-mannosyltransferase	<i>ALG12</i>	607143	Intellectual disability, hypotonia, facial dysmorphism, acquired microcephaly, and frequent infections
ALG8-CDG	CDG-Ih	Alpha-1,3-glucosyltransferase	<i>ALG8</i>	608104	Hepatomegaly, protein-losing enteropathy, renal failure, hypoalbuminaemia, oedema, and ascites
ALG2-CDG	CDG-Ii	Alpha-1,3-mannosyltransferase	<i>ALG2</i>	607906	Intellectual disability, hypomyelination, intractable seizures, iris colobomas, hepatomegaly, and coagulopathy
DPAGT1-CDG	CDG-Ij	GlcNAc-1-P transferase	<i>DPAGT1</i>	608093	Severe intellectual disability, hypotonia, seizures, microcephaly, and exotropia
ALG1-CDG	CDG-Ik	Beta-1,4-mannosyltransferase	<i>ALG1</i>	608540	Severe intellectual disability, hypotonia, intractable seizures, nephrotic syndrome, acquired microcephaly, fever, coagulopathy, and early death
ALG9-CDG	CDG-IL	1,2-mannosyltransferase	<i>ALG9</i>	608776	Hepatomegaly, hypotonia, seizures, and severe microcephaly
DOLK-CDG	CDG-Im	Dolichol kinase	<i>DOLK</i>	610768	Ichthyosis, baldness, hypotonia, dialyated cardiomyopathy, and hypoketotic hypoglycemia
RFT1-CDG	CDG-In	Man(5)GlcNAc flippase	<i>RFT1</i>	612015	Hepatomegaly, developmental delay, hypotonia, seizures, and sensorineural deafness
DPM3-CDG	CDG-Io	Dolichol-P-Man synthase	<i>DPM3</i>	612937	Normal development, dystroglycanopathies, and stroke like episode
TUSC3-CDG		Subunit of the OST complex	<i>TUSC3</i>	611093	Nonsyndromic intellectual disability
IAP-CDG		Subunit of the OST complex	<i>IAP</i>	–	X-linked nonsyndromic intellectual disability
ALG11-CDG	CDG-Ip	Alpha-1,2-mannosyltransferase	<i>ALG11</i>	–	Develop intellectual disability, muscular hypotonia, seizures, early death
MGAT2-CDG	CDG-IIa	N-acetylglucosaminyltransferase II	<i>MGAT2</i>	212066	Intellectual disability, dysmorphism, stereotypies, seizures
GCS1-CDG	CDG-IIb	α-1,2-glucosidase	<i>GCS1</i>	601336	Dysmorphism, hypotonia, seizures, hepatomegaly, hepatic fibrosis; death at 2.5 months
SLC35C1-CDG	CDG-IIc	GDP-fucose transporter	<i>FUCT1</i>	266265	Severe intellectual disability, hypotonia, elevated peripheral neutrophils.
B4GALT1-CDG	CDG-IId	Beta-1,4-galactosyltransferase	<i>B4GALT1</i>	607091	Hypotonia (myopathy), spontaneous hemorrhage, Dandy– Walker malformation
COG7-CDG	CDG-IIe	Conserved oligomeric Golgi subunit 7	<i>COG7</i>	608779	Early Fatality, dysmorphism, hypotonia, intractable seizures, hepatomegaly, progressive jaundice, recurrent infections, cardiac failure, excessive skin

(Continued)

Table 2 (Continued)

<i>Disorders</i>		<i>Enzymatic or protein defect</i>	<i>Gene</i>	<i>OMIM^a</i>	<i>Frequent clinical features</i>
<i>CDG type</i>	<i>Formerly</i>				
SLC35A1-CDG	CDG-II _f	CMP-sialic acid transporter	<i>SLC35A1</i>	605634	Thrombocytopaenia, no neurological symptoms; normal transferrin, abnormal platelet glycoproteins
COG1-CDG		Conserved oligomeric Golgi subunit 1	<i>COG1</i>	606973	Mild MR, hypotonia, growth retardation, progressive microcephaly, hepatosplenomegaly
COG8-CDG	CDG-II _h	Conserved oligomeric Golgi subunit 8	<i>COG8</i>	611182	MR, hypotonia, and encephalopathy
COG5-CDG		Conserved oligomeric Golgi subunit 5	<i>COG5</i>	—	Moderate intellectual disability, cerebellar atrophy, and hypotonia,
COG4-CDG		Conserved oligomeric Golgi subunit 4	<i>COG4</i>	—	Mild intellectual disability, mild dysmorphia, epilepsy, recurrent respiratory infections, and mild ataxia
ATP6V0A2-CDG		Golgi pH regulator	<i>ATP6V0A2</i>	219200	Cutis laxa, wrinkly skin, connective tissue weakness, large fontanelle, and —/+ variable intellectual disability
Achondrogenesis type 1A		Golgi structure	<i>GMAP-210</i>	200600	Infants have soft skull bones, short ribs that fracture easily, extremely short limbs, and lack normal ossification in the spine and pelvis
Congenital dyserythropoietic anemia (CDA II)	Hempas	ER–Golgi protein trafficking	<i>SEC23B</i>	224100	Anemia, jaundice, splenomegaly, and gall bladder disease
Cranio-lenticulo-sutural dysplasia (CLSD)		ER–Golgi protein trafficking	<i>SEC23A</i>	607812	Facial dysmorphisms, sutural cataracts late-closing fontanels, and skeletal defects
<i>Muscular dystrophy</i> Walker–Warburg syndrome		Putative O-mannosyltransferase	<i>POMT1</i>	268870	Severe muscle weakness, death in infancy, absent psychomotor development, neuronal migration disorder, and ocular abnormalities
Muscle–eye–brain disease (MEB)		O-mannosyl glycan synthesis	<i>POMGnT1</i>	253280	Severe muscle weakness, intellectual disability, epilepsy, neuronal migration disorder, and ocular abnormalities
Fukuyama-type congenital muscular dystrophy (FCMD)		Putative glycosyltransferase	<i>Fukutin</i>	253800	Severe proximal and axial weakness, intellectual disability, and epilepsy neuronal migration disorder
Congenital muscular dystrophy type 1C (MDC1C)		Fukutin-related protein, a putative glycosyltransferase	<i>FKRP</i>	606612	Hypotonia, impaired motor development with respiratory muscle weakness
Congenital muscular dystrophy type 1D (MDC1D)		Putative glycosyltransferase	<i>LARGE</i>	608840	Muscular dystrophy with profound intellectual disability
Hereditary inclusion body myopathy / distal myopathy with rimmed vacuoles		UDP-GlcNAc epimerase/kinase	<i>GNE</i>	600737	Adult onset myopathy that spares quadriceps
<i>Other types</i>					
1-cell disease		GlcNAc-1-P transferase	<i>GNPTA</i>	252500	Severe developmental abnormalities
Multiple hereditary exostoses		Heparan sulfate co-polymerase	<i>EXT1, EXT2</i>	133701	Bony outgrowths
Macular corneal dystrophy		GlcNAc-6-sulfotransferase	<i>CHST6</i>	605294	Progressive corneal opacity
Adducted thumb-clubfoot syndrome		N-acetylgalactosamine 4-O-sulfotransferase 1	<i>CHST14</i>	601776	Congenital contractures of thumbs and feet with joint instability, facial clefting, coagulopathy, distinct facial appearance, thin and translucent skin, heart, kidney, or intestinal defects

(Continued)

Table 2 (Continued)

<i>Disorders</i>	<i>Enzymatic or protein defect</i>	<i>Gene</i>	<i>OMIM^a</i>	<i>Frequent clinical features</i>
<i>CDG type</i>	<i>Formerly</i>			
Ehlers–Danlos syndrome progeroid form	Xylosylprotein β -1,4-galactosyltransferase	<i>B4GALT7</i>	130070 604327	Hypotonia, delayed development, connective tissue abnormalities, and loose skin
Diastrophic dysplasia achondrogenesis	Anion (sulfate) transporter	<i>DTDST</i>	606718 600972	Scoliosis, club foot, and premature calcification
Spondylo-epimetaphyseal dysplasia	3'-phosphoadenosine-5'-phosphosulphate synthase (PAPS)	<i>ATPSK2</i>	603005	Abnormal skeletal development and linear growth
Familial tumoral calcinosis	GalNAc transferase	<i>GALNT3</i>	211900	Massive calcium deposits in skin and tissue
Tn syndrome	Chaperone of β -1,3GalT	<i>COSMC</i>	230430	Anemia, leukopenia, and thrombocytopenia (somatic mutation)
Schneckenbecken dysplasia	UDP-GlcA / UDP-GalNAc Golgi Transporter	<i>SLC35D1</i>	610804	Severely shortened long bones with bowing of limb bones and unossified vertebral bodies
Peters Plus syndrome	B-1,3 Glucosyltransferase specific for O-Fucose on thrombospondin type 1 repeats	<i>B3GALT1</i>	261540	Intellectual disability with prenatal growth delay, short stature, brachymorphism, short limbs, and sometimes cleft palate
<i>Defects in glycolipid synthesis</i>				
Paroxysmal nocturnal haemoglobinuria	PI-GlcNAcT	<i>PIGA</i>	311770	Complement-mediated hemolysis (somatic mutation)
Amish infantile epilepsy	Sia2,3Gal β 1,4Glc-Cer synthase (GM3)	<i>SIAT9</i>	609056	Tonic-clonic seizures, arrested development, and neurological decline
Autosomal recessive GPI anchor deficiency	1st Mannosyltransferase in GPI biosynthesis	<i>PIGM</i>	610273	Venous thrombosis and seizures caused by SP1 promoter mutation

^aOMIM: Online Mendelian Inheritance in Man.

and interconversion of monosaccharide precursors or defective glycosyltransferases used to link these sugar units. The second group involves processing of the sugar chains. It is uncertain how many genes may be involved in the step because various genes encode enzymes that may catalyze the same or similar sugar-transfer reactions. In addition, genes from this group involve glycosidases, glycosyl transferases, and Golgi-associated donor nucleotide transporters that deliver substrates to the Golgi lumen and Golgi trafficking proteins. Thus far, defects in 14 genes from this group cause CDG. It is becoming clear that defects in Golgi maintenance and trafficking represent a growing number of CDG cases. Currently, there are ~1000 confirmed cases of all types of CDG worldwide, which probably represent a small percentage of the total patients. It is safe to say that the entire group of CDGs is severely under diagnosis, and that we are now only seeing the 'tip of the iceberg'.

Laboratory Diagnosis of CDG

CDG patients are often misdiagnosed early on because their symptoms resemble other generic disorders. CDG-Ia patients are sometimes mistaken for having inherited mitochondrial disorders or cerebral palsy. This is understandable since glyco-biology is not part of a typical medical education or training program and most practicing physicians have a limited exposure to glycobiology. Fortunately, most CDG patients can be

diagnosed by a simple blood test to analyze the glycosylation status of transferrin (Tf). Abnormal Tf is detected by isoelectric focusing, or by electrospray ionization-mass spectrometry. Many other proteins are misglycosylated, but Tf appears to be especially sensitive to limitations in the amount of LLO precursor. These biochemical analyses provide clues to the defect, but they cannot pinpoint the defective gene, since many defects produce the same altered pattern.

Biochemical Overview

All of the CDGs are autosomal recessive disorders and they result from an insufficient amount of precursor molecules or from deficient glycosyltransferases or sugar-processing glycosidases. For a complete listing, see Table 2. By far the most common type, CDG-Ia (OMIM 212066), is caused by mutations in the gene encoding phosphomannomutase (Man-6-P $\rightarrow$ Man-1-P). Mutations reduce the amount of GDP-Man, resulting in insufficient LLO. This leaves some of the normal glycosylation sites on many proteins unoccupied, which can lead to their misfolding and degradation. Sometimes the misfolded proteins accumulate in the ER, activating cell apoptosis. CDG-Ib (OMIM 602579) results from mutations in the *MPI* gene encoding phosphomannose isomerase (PMI) (fructose-6-P $\rightarrow$ Man-6-P), just one metabolic step away from phosphomannomutase. The clinical pictures of CDG-Ib and CDG-Ia

patients are quite different. CDG-Ic (OMIM 603147) is caused by mutations in *ALG6*, which encodes an α -1,3-glucosyltransferase used to add the first Glc to the immature LLO precursor. Chains lacking glucose are transferred to proteins less efficiently than full-sized ones. Finally, the OST complex mediates the transfer of the final mature glycan product to the protein. Mutations already identified in two of the OST subunits will provide insight into understanding how the glycan chains are transferred to proteins. But identification of these patients will be more challenging since they have normal Tf.

Group II CDGs are defined as those that affect the processing of the protein-bound sugar chains. Only a few patients have been identified with each of these disorders, and most, but not all of them can be diagnosed by analysis of Tf. Thus far, the most common group II CDG is composed of patients with defects in the conserved oligomeric Golgi (COG) complex. The COG complex is an octameric structure whose stability is dependent on the presence of functional subunits; mutations in various subunits destabilize complex. The stabilized complex is required for intracellular transport of the glycosylation machinery. To date, mutations have been identified in five of the eight subunits (see [Table 2](#)).

Several other genetic disorders have provided key insight into the roles of protein trafficking and its effects on glycosylation. For example, mutations in Golgi pH regulator ATP6V0A2 have shown the role in glycosylation and trafficking play in maintaining extracellular matrix integrity as mutations result in autosomal recessive cutis laxa type II. Mutations in SEC23A, a COPII component, results in cranio-lenticulo-sutural dysplasia, a rare disorder with defects in clearing secreted proteins resulting in a dysfunctional ER. Congenital dyserythropoietic anemia (CDA-II/HEMPAS) results in ineffective erythropoiesis due to mutations in the COPII component SEC23B, which appears to be involved in erythroid differentiation.

Many of these proteins have several homologs whose function is conserved within similar pathways; but it is becoming clearer that certain protein components are required for non-redundant functions.

Clinical Features of CDG

The most common clinical features in various types of CDG are shown in [Table 2](#). There is considerable clinical heterogeneity even within a specific type of CDG. Intellectual disability ranging from mild to severe and hypotonia are consistent features in all patients except type-Ib. Other neurological findings include ataxia (Ia and Ic), seizures, and stroke-like episodes. Cerebellar hypoplasia (Ia and Ic), delayed myelination (Ie, IIa, and IIb), microcephaly, and atrophy of the cerebrum (Ia, Ic, Id, and Ie) are seen. Sometimes the cognitive deficiency can be mild. Nearly all patients have feeding problems and fail to thrive. Strabismus, abnormal fat distribution, and retracted nipples are common, especially in the CDG-Ia group.

Mortality in CDG-Ia children is about 20% during the first few years, but they stabilize after childhood. Substantial survival means that many adult CDG patients remain undiagnosed.

CDG-Ib, caused by deficiency in MPI, is substantially different from all the others; in that case mental development and

the nervous system are normal. The reason for this is unknown. These patients have serious gastrointestinal problems, protein-losing enteropathy (loss of proteins through the intestine), low plasma albumin, coagulopathy (fast clotting), and hypoglycemia. A substantial number of patients died before they received a diagnosis, which is unfortunate, since this type of CDG can be treated effectively.

Less common clinical features can periodically help to define a CDG subtype include ichthyosis, ocular coloboma, cataracts, abnormal genitalia, and muscular dystrophy.

Therapy for CDG

The effective therapy for CDG-Ib is oral mannose. Mannose bypasses the fructose-6-P $\rightarrow$ Man-6-P block to replenish the depleted GDP-Man pools. A mannose transporter delivers mannose to the cells and hexokinase converts it into Man-6-P. Mannose therapy reverses the hypoglycemia and coagulopathy within a few weeks, and within 1–2 months plasma protein levels become normal and protein-losing enteropathy disappears. Some patients have been treated for over 12 years without side effects. Growth during childhood is metabolically demanding and mannose supplements are most important then. It is encouraging that none of the adult patients with proven CDG-Ib is currently taking mannose, suggesting that it will not be a lifelong requirement. It is important to point out that there is no evidence that healthy people require mannose supplements. Vendors of complex nutraceutical mannose fail to point out that the natural plant polysaccharides they sell are non-digestible in the human intestine to produce bio-available mannose. Fucose supplements have also been used to treat patients with CDG-IIc who have a defective GDP-Fucose transporter. Fucose enters the cell and apparently increases the GDP-Fucose pool driving it into the Golgi through the impaired transporter with some residual activity. Within a day, fucose treatment corrects the patients' elevated circulating neutrophil by generating a selectin ligand needed for neutrophil rolling on the endothelium prior to their transmigration into the tissues. Infections cease and health improves. Unfortunately, fucose does not improve or reverse the intellectual disability.

Congenital Muscular Dystrophies

α -Dystroglycan is a peripheral membrane component of the dystrophin–glycoprotein complex (DGC) found in muscle, nerve, heart, and brain. This protein binds to merosin in extracellular matrix bridging it to the cytoskeleton molecules that include dystrophin and actin. Mutations in dystrophin cause Duchenne and Becker muscular dystrophies, while assorted mutations in merosin and sarcoglycans cause other inherited muscular dystrophies. Total systemic loss of α -dystroglycan is embryonic lethal. In addition, this protein is the major carrier of O-mannose-based sugar chains. In α -dystroglycan, these sugar chains cluster together in a small mucin-like domain of where they mediate many of the critical stabilizing interactions with the other matrix molecules such as merosin, neurexin, and agrin, depending on which cells express α -dystroglycan. Several kinds of muscular dystrophies

including muscle–eye–brain (MEB) disease, Fukuyama-type congenital muscular dystrophy (FCMD), Walker–Warburg syndrome, and limb–girdle muscular dystrophy, involve mutations in the genes needed for the biosynthesis of these chains (Table 2). These mutations also appear to affect neuronal migration in the developing brain, thus accounting for the combined effects on muscle and brain development. Studying the glycosylation of α -dystroglycan will be important to understand how these pathologies develop. However, α -dystroglycan is not the only protein that carries these sugar chains. In fact, one out of four O-linked chains in the brain is O-mannose based.

One form of adult onset muscular dystrophy, hereditary inclusion body myopathy type II, is due to mutations in an enzyme involved in the biosynthesis of CMP-sialic acid, the universal activated donor for addition of sialic acid. The mutations appear to have only moderate (50–60%) reduction in the activity of the enzyme (UDP-GlcNAc epimerase-kinase) which is lethal when totally deleted. Preliminary studies in fibroblasts and muscle cells from the patients show that addition of sialic acid is reduced and that supplying sialic acid or N-acetylmannosamine (the product of the impaired enzyme) can correct the glycosylation. This suggests that patients might benefit from supplements of these sugars. In fact, mice carrying patient-like mutations in this enzyme respond well to diets supplemented with ManNAc, sialic acid, or even sialyllactose.

Defects in Proteoglycan Biosynthesis Cause Disease

Incomplete or impaired synthesis of glycosaminoglycan (GAG) chains also cause inherited disorders. The bony outgrowths found in patients with the dominant disorder, multiple hereditary exostoses (MHE) result from mutations in the copolymerase that assembles alternating units of glucuronic acid and GlcNAc in heparan sulfate. Ehlers–Danlos syndrome, which causes a form of progeria is caused by mutations a galactosyltransferase used for assembly of the tetrasaccharide core linkage common to heparin, heparan sulfate, chondroitin, and dermatan sulfate. Macular corneal dystrophy that leads to corneal opacity is caused by loss of a specific sulfotransferase that modifies a special type of keratan sulfate found in the cornea. Sulfate is such an important component for making extracellular matrices that cells have a cell-surface transporter that delivers sulfate to the cytoplasm. Mutations in the transporter cause multiple cartilage and bone malformations and in

the most severe forms prenatal lethality results from respiratory insufficiency.

Glycosylation Disorders: The Next Generation

The very complexity of sugar chain assembly practically guarantees that more glycosylation disorders will be discovered. Finding defects in the N-linked pathway has been relatively easy because many of the steps involve only a single enzyme, many of which have strict substrate requirements. In addition, Tf provides an unusually robust assay for defective N-glycosylation. Except for α -dystroglycan, the other pathways lack a similar diagnostic champion. New disorders that compromise glycosylation will probably be discovered in the synthesis of O-GalNAc-linked sugar chains, GAGs, and in the assembly, organization, or recycling of proteins in the Golgi apparatus. These causes may be more difficult to identify because many of these pathways are composed of redundant, sometimes, overlapping enzymes and proteins with tissue specific distributions.

Identification of known CDG subtypes and new causative genes will certainly be enhanced by newer technologies such as high-density linkage analysis, high-throughput sequencing, and greater awareness. But in the end, CDG subtype also needs to be confirmed biochemically.

Understanding the role of sugar chains in building cell-surface signaling complexes and finding how to boost sugar donors in cells may help to understand and treat these pathologies. Interactions among basic scientists focused on glycobiology and glycosylation-savvy clinicians will be the key to further discoveries.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins: Proteoglycans.**

Further Reading

Freeze HH (2006) Genetic defects in the human glycome. *Nature Reviews Genetics* 7: 537–551.

Relevant Website

<http://www.ncbi.nlm.nih.gov> – NCBI: National Center for Biotechnology Information; OMIM: Online Mendelian Inheritance in Man.

Glycosylphosphatidylinositol Anchors

A K Menon, Weill Cornell Medical College, New York, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Endoplasmic reticulum (ER) Intracellular meshwork of membranes in eukaryotic cells; the ER is endowed with the ability to synthesize lipids and is a departure point for the transport of proteins to the cell surface and extracellular space.

Glycoconjugate Macromolecule containing sugars.

Glycolipid Lipid molecule to which sugars are covalently attached.

Lipid-modified proteins Proteins, including glycosylphosphatidylinositol (GPI)-anchored proteins, to which lipids (fatty acids, isoprenoids, or GPI) are covalently attached.

Phospholipid Amphipathic molecule with a hydrophobic tail consisting typically of two fatty acids linked to glycerol, and a hydrophilic head also attached to a glycerol backbone.

Glycosylphosphatidylinositol (GPI)-anchored proteins are a class of lipid-anchored membrane proteins that are ubiquitously expressed at the surface of eukaryotic cells. GPI proteins represent roughly 1% of all proteins encoded by eukaryotic genomes. They differ from traditional membrane proteins in that they rely on a complex glycolipid, GPI, rather than a hydrophobic transmembrane sequence for their association with membranes. GPI anchors are found on a variety of functionally diverse proteins (as well as glycoconjugates) including cell-surface receptors (folate receptor, CD14), cell adhesion molecules (neural cell adhesion molecule isoforms and carcinoembryonic antigen variants), cell surface hydrolases (5'-nucleotidase, acetylcholinesterase), complement regulatory proteins (decay accelerating factor), and protozoal surface molecules (*Trypanosoma brucei* variant surface glycoprotein, *Leishmania* lipophosphoglycan). Along with serving to attach proteins to the cell surface, GPI-anchored proteins appear to be markers and major constituents of detergent-resistant lipid rafts, the sphingolipid- and sterol-rich domains in membranes that are postulated to play an important role in the activation of signaling cascades.

GPI Structure

The structure of a GPI-anchored protein is shown in [Figure 1](#). The carboxy terminus of the protein is linked via an ethanolamine residue to GPI. The core GPI structure consists of ethanolamine phosphate linked via a linear sequence of three mannose residues and a single glucosamine residue to the phospholipid phosphatidylinositol (PI).

GPI Biosynthesis

Outline of the Biosynthetic Pathway

The GPI structure is assembled in the endoplasmic reticulum (ER) via a complex metabolic pathway composed of at least 10 steps and requiring the participation of at least 20 distinct gene products. Synthesis occurs by sequential addition of monosaccharides to PI yielding the core GPI structure that is then added to proteins. [Figure 2](#) illustrates the sequence of reactions involved in the assembly of a GPI protein anchor precursor in mammalian cells; variations of this general sequence are found in all

organisms studied thus far. GPI biosynthesis is initiated on the cytoplasmic face of the ER via the addition of *N*-acetylglucosamine (GlcNAc) from UDP-GlcNAc to PI yielding GlcNAc-PI ([Figure 2](#), step 1). The enzyme that catalyzes this first step is an unusual complex consisting of six membrane proteins in mammalian cells. GlcNAc-PI is subsequently de-*N*-acetylated to generate GlcN-PI ([Figure 2](#), step 2). The third step in biosynthesis is the acylation of the 2-OH of the inositol residue of GlcN-PI ([Figure 2](#), step 3). The acylation step is required prior to the addition of three mannose residues to GlcN-(acyl)PI ([Figure 2](#), step 5), and modification of the mannose residues with phosphoethanolamine (EtNP) side chains ([Figure 2](#), step 6). In contrast to the first two steps of synthesis that occur on the cytoplasmic leaflet of the ER, the mannosylation reactions are thought to occur lumenally based on the predicted membrane topology of the mannosyl transferases. This indicates that the substrate (GlcN-PI or GlcN-(acyl)PI) must be flipped across the ER membrane bilayer ([Figure 2](#), step 4). The mannose and EtNP residues are derived from lipids. Dolichol-P-mannose serves as the mannose donor for all three mannosylation steps and EtNP residues are added from the phospholipid phosphatidylethanolamine. The terminal EtNP residue is critical, since it is this residue that is involved in the linkage of GPI to the carboxy terminus of proteins ([Figure 2](#), step 7). Fully assembled GPI structures, or the GPI moiety in GPI-anchored proteins, are frequently subject to lipid remodeling reactions in which fatty acids or the entire lipid moiety is replaced with different fatty acids or lipids.

GPI Attachment to Proteins

Newly synthesized proteins are attached to pre-existing GPIs in the lumenal leaflet of the ER by a GPI:protein transamidase complex. Proteins to be GPI-anchored contain an N-terminal signal sequence that targets them to the ER and a C-terminal signal sequence that directs GPI anchoring. The N-terminal signal sequence is cleaved by signal peptidase during or after translocation of the nascent polypeptide into the ER lumen. The C-terminal GPI signal sequence is then cleaved and replaced with a GPI anchor through the action of GPI transamidase.

The C-terminal GPI-signal peptide, which is cleaved at what is referred to as the ω site, is necessary and sufficient for designating that a protein becomes GPI anchored. The GPI

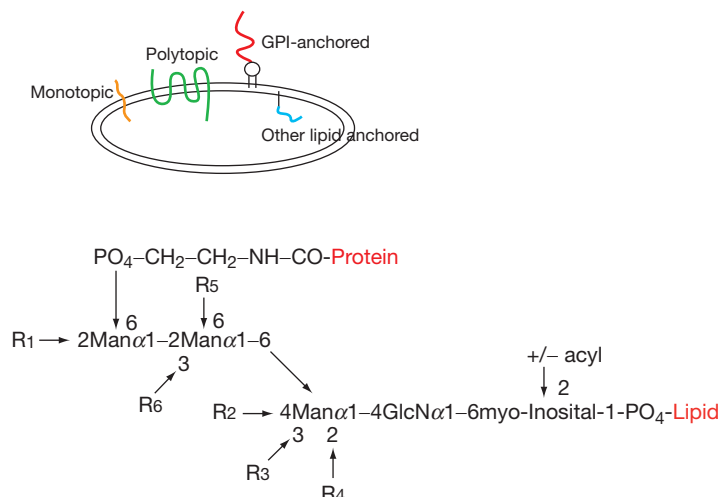


Figure 1 The top schematic illustrates different membrane and membrane-associated proteins in the cell-surface membrane of a eukaryotic cell (internal membranes of the cell are not shown). Proteins containing hydrophobic stretches of amino acids may span the membrane once (monotopic membrane proteins) or several times (polytopic). GPI-anchored proteins are firmly associated with the cell-surface membrane, but confined to the outer leaflet of the membrane bilayer since the anchor cannot span the membrane. Other lipid-anchored proteins (such as N-myristoylated, palmitoylated, and prenylated proteins) are attached to the cytoplasmic leaflet of the cell-surface membrane. GPIs attached to protein have the conserved structure shown in the lower part of the figure. Variations of the structure include substituents on each of the mannose residues (phosphoethanolamine groups, monosaccharides, etc.; indicated R1–R6), differences in the lipid moiety (diacylglycerol, alkylacylglycerol, ceramide, etc.), and the presence of a fatty acyl chain attached to the 2-position of the myo-inositol ring. From Ferguson MAJ, Brimacombe JS, Brown JR, *et al.* (1999) The GPI biosynthetic pathway as a therapeutic target for African sleeping sickness. *Biochimica et Biophysica Acta* 1455: 327–340.

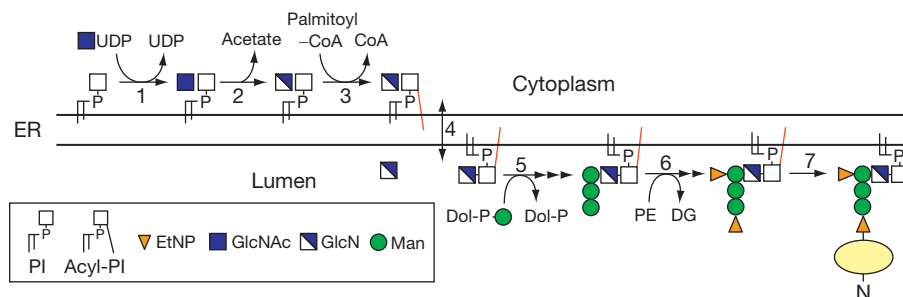


Figure 2 GPI biosynthesis in the ER. The pathway is a simplified version of the biosynthetic process in mammalian cells and is broadly similar in all other eukaryotic organisms. Biosynthesis is initiated on the cytoplasmic face of the ER by the transfer of GlcNAc from UDPGlcNAc to PI (step 1). GlcNAc-PI is de-N-acetylated (step 2), then acylated on the inositol residue (step 3) to yield GlcN-acyl PI. This lipid is flipped across the ER membrane (step 4), triply mannosylated (step 5) (dolichol-phosphate-mannose (Dol-P-Man) contributes each of the mannose residues) and capped with a phosphoethanolamine residue (step 6) (donated by the phospholipid phosphatidylethanolamine (PE)). Other phosphoethanolamine residues may also be added, some prior to the completion of mannosylation. It is possible that step 3 occurs in the ER lumen and that GlcN-PI, rather than GlcN-acyl PI is flipped across the ER membrane. The mature GPI structure is then attached to ER-translocated proteins bearing a C-terminal, GPI-directing signal sequence, to yield a GPI-anchored protein (step 7).

transamidase complex responsible for attaching the protein to the GPI anchor consists of at least five subunits. The enzyme recognizes a protein containing a GPI-anchoring signal sequence, then cleaves the signal sequence and attaches the new carboxy terminus to the terminal EtNP of a pre-existing GPI.

GPI Anchoring in Mammals, Parasitic Protozoa, and Yeast

GPI-deficient mammalian cells are viable in tissue culture, but a GPI defect has clear consequences for multicellular organisms. Transgenic mouse embryos lacking the ability to

initiate GPI biosynthesis (defective in PIG-A) do not develop beyond the ninth day of gestation. The inability of certain blood cells to express GPI-anchored proteins results in the rare human disease, paroxysmal nocturnal hemoglobinuria (PNH), characterized by intravascular hemolysis, thrombosis, and bone marrow failure. The disease is caused by a somatic mutation of the X-linked PIG-A gene (encoding the catalytic component of the enzyme responsible for the first step in GPI biosynthesis) in hematopoietic stem cells. Cells defective in PIG-A either are unable to synthesize GPIs or exhibit a significant decrease in the synthesis of GPIs and thus decreased expression of GPI-anchored proteins. Red blood cells no longer expressing GPI-anchored complement regulatory proteins

(e.g., CD55 and CD59) become susceptible to complement-mediated lysis, resulting in the release of heme and hemoglobin into the blood, filtering by the kidney, and excretion in the urine. Interestingly, while the GPI-deficient phenotype of hematopoietic stem cells makes the cells more sensitive to complement-mediated lysis, the defective PNH clone still persists. This implies that there must be other factors that select for clonal expansion and maintenance of GPI-deficient blood cells. One proposal is that PNH patients possess autoreactive T cells that target GPI on the surface of hematopoietic stem cells; thus, PNH cells would evade damage because they lack surface GPI molecules.

GPI anchoring is the most prominent mode of attachment for cell-surface proteins and glycans in parasitic protozoa. Pathogenic protozoa, including species of the genera *Trypanosoma*, *Leishmania*, and *Plasmodium*, display abundant GPI-anchored cell-surface macromolecules that play crucial roles in parasite infectivity and survival. An example is the GPI-anchored variant surface glycoprotein of bloodstream forms of *Trypanosoma brucei*, the causative agent of African sleeping sickness.

GPIs serve a unique function in yeast (*Saccharomyces cerevisiae*). In addition to anchoring secretory proteins to the cell surface, GPIs play an additional role in yeast cell-wall biosynthesis. The yeast cell wall consists of a fibrous lattice of mannoproteins, β 1,3-glucan, β 1,6-glucan, and chitin. During cell-wall biosynthesis, GPI-anchored mannoproteins are transported through the secretory pathway to the cell surface. After arrival at the plasma membrane, a transglycosylation reaction results in cleavage of the GPI moiety between GlcN and the first mannose residue and formation of a glycosidic linkage between the mannoprotein-GPI-remnant and β 1,6 glucan. Mutations in GPI biosynthesis are lethal in yeast and decreased levels of GPI biosynthesis cause growth defects and aberrant cell-wall biogenesis. However, not all GPI-anchored proteins become cross-linked to β 1,6 glucan; the amino acids within four or five residues upstream of the ω site determine whether the protein becomes incorporated into the cell wall or remains anchored to the plasma membrane. A unique characteristic of yeast GPIs is the addition of a fourth mannose residue to the core GPI structure (via an α 1–2 linkage to the third mannose). Most GPI-anchored proteins in yeast also undergo lipid remodeling, replacing the glycerolipid backbone with ceramide. The remodeling occurs after the attachment of protein to the GPI and can occur in both the ER and the Golgi.

Functions of GPIs

GPIs function as membrane anchors for secretory proteins and are believed to provide targeting signals that influence the intracellular trafficking of these proteins. GPI-anchored proteins have been shown to coalesce with sphingolipids and cholesterol into detergent-insoluble membrane domains, or lipid rafts. Association of molecules in rafts at the plasma membrane, including GPI-anchored proteins (in the exoplasmic

leaflet) and acylated signaling molecules (in the cytoplasmic leaflet), is postulated to play an important role in the activation of signaling cascades. In some polarized epithelial cells, GPI-anchored proteins and many glycosphingolipids are sorted in the *trans*-Golgi network and specifically targeted to the apical membrane. Once at the plasma membrane, GPI-anchored proteins can undergo endocytosis and recycling back to the cell surface. However, uptake of GPI-anchored proteins is $\sim$ 5 times slower than that of receptor-mediated endocytosis and recycling to the plasma membrane is $\sim$ 3 times slower than that of recycling receptors. The current view is that GPI anchoring of proteins directs their segregation into lipid rafts and affects their sorting in both the exocytic and endocytic pathways. A GPI anchor may also allow a protein to be selectively released from the cell surface upon hydrolysis by a GPI-specific phospholipase (e.g., PI-PLC or GPI-PLD) or to transfer between cells and stably insert in the external leaflet of the acceptor cell's plasma membrane. Although the biological significance of this latter phenomenon is unclear, however, the ability of GPI-anchored proteins to transfer between cells has implications for the expression of foreign proteins on the cell surface.

Acknowledgments

Sections of this article are reproduced and/or adapted with permission from an article by Baumann and Menon published in 2002 that provides a summary of all types of lipid anchored proteins.

See also: [Protein/Enzyme Structure Function and Degradation: Protein N-Myristoylation; Protein Palmitoylation.](#)

Further Reading

- Baumann NA and Menon AK (2002) Lipid modifications of proteins. In: Vance DE and Vance J (eds.) *Biochemistry of Lipids, Lipoproteins and Membranes*, ch. 2, pp. 37–54. Amsterdam: Elsevier.
- Eisenhaber B, Bork P, and Eisenhaber F (2001) Post-translational GPI lipid anchor modification of proteins in kingdoms of life: Analysis of protein sequence data from complete genomes. *Protein Engineering* 14: 17–25.
- Ferguson MAJ (2000) Glycosylphosphatidylinositol biosynthesis validated as a drug target for African sleeping sickness. *Proceedings of the National Academy of Sciences of the United States of America* 97: 10673–10675.
- Ferguson MAJ, Brimacombe JS, Brown JR, et al. (1999) The GPI biosynthetic pathway as a therapeutic target for African sleeping sickness. *Biochimica et Biophysica Acta* 1455: 327–340.
- Karadimitris A and Luzatto L (2001) The cellular pathogenesis of paroxysmal nocturnal hemoglobinuria. *Leukemia* 15: 1148–1152.
- Lingwood D and Simons K (2010) Lipid rafts as a membrane-organizing principle. *Science* 327: 46–50.
- Orlean P and Menon AK (2007) GPI anchoring of protein in yeast and mammalian cells, or: How we learned to stop worrying and love glycopospholipids. *Journal of Lipid Research* 48: 993–1011.
- Pittet M and Conzelmann A (2007) Biosynthesis and function of GPI proteins in the yeast *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta* 1771: 405–420.

Golgi Complex

M Starnes, University of Iowa, Iowa City, IA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Cisternae Flattened membrane-limited structures that stack together to form the Golgi complex.

Coatomer A seven-subunit transport vesicle protein coat complex involved in trafficking at the Golgi complex.

Endoprotease An enzyme that cleaves a protein into two or more fragments.

Glycoprotein A protein that is attached to complex sugar molecules called 'oligosaccharides'.

Golgi matrix A protein network that binds to the cytosolic face of Golgi cisternae and contributes to Golgi morphology and positioning.

Golgin A large family of rod-shaped proteins that bind to the cytosolic face of Golgi membranes. Golgins contribute to the formation of the stacked ribbon-like morphology.

Secretory vesicles Specialized transport vesicles that transport proteins from the *trans*-Golgi network to the plasma membrane. These vesicles can sometimes be regulated to allow release of proteins such as hormones from cells only under certain conditions.

Transport vesicle A small spherical structure that mediates protein transport between organelles. It is formed by the assembly of a protein coat.

The Golgi complex or Golgi apparatus is an organelle of eukaryotic cells involved in the processing and transport of proteins. The Golgi complex is one of a group of cellular organelles including the endoplasmic reticulum, endosomes, and lysosomes that compose the secretory and endocytic pathways. Proteins are transported through the organelles of the secretory pathway en route to the plasma membrane for secretion into the extracellular space or insertion into the membrane. This pathway also transports and sorts the resident proteins of many cellular organelles. Secreted proteins such as hormones enter the secretory pathway by translocation into the endoplasmic reticulum. They are then concentrated and transported via vesicle intermediates to the Golgi complex. At the Golgi complex, proteins can undergo processing events such as proteolytic cleavage and the removal or addition of sugar moieties. Proteins are finally sorted into various classes of post-Golgi vesicles for transport to the proper location in the cell. Hence, the Golgi complex plays a central role in the function of the secretory pathway and is the subject of intense investigation by cell biologists. Although important aspects of Golgi function remain to be fully characterized, previous studies have revealed many interesting features of its morphology and its role in the transport and processing of proteins.

Morphology of the Golgi Complex

The Golgi complex is often found as a set of three to five stacked saccules or cisternae (Figure 1(a)). The cisternae are organized with a polarity such that there is a *cis*-cisterna where proteins arriving from the endoplasmic reticulum enter and a *trans*-cisterna where proteins exit the Golgi complex. Proteins pass through the medial cisternae that are present between the *cis*- and *trans*-cisternae. The *cis*- and *trans*-faces of the Golgi apparatus are often reticulated and thus referred to as the *cis*-Golgi network and the *trans*-Golgi network. Although most eukaryotic cells have stacked Golgi cisternae, this configuration is not universal. For example, brewer's yeast, *Saccharomyces cerevisiae*,

has distinct *cis*-, medial-, and *trans*-Golgi cisternae that remain dispersed from each other in the cytoplasm.

Stacked Golgi Morphology Is Conferred by the Golgin Proteins

Golgins are rod-shaped proteins that localize to the Golgi apparatus in all eukaryotic cells. They play important roles in establishing and maintaining Golgi morphology. The Golgi reassembly and stacking proteins, GRASPs, are golgins that contribute to the stacked morphology observed in most eukaryotes. Vertebrates have two GRASP proteins, GRASP55 and GRASP65, that assemble into membrane-bound oligomers that then tether the individual cisternae into stacks. GM130 is another golgin protein that acts as a tether to assist with Golgi stacking and binding interactions between transport vesicles and the cisternal membrane.

Golgins are also known to be components of the Golgi matrix, an assembly of peripheral membrane-binding proteins that contribute to Golgi morphology. The Golgi matrix persists near the nucleus even when membrane components of the Golgi complex are removed. In vertebrate cells, the Golgi matrix can be observed by electron microscopy as a region surrounding the Golgi apparatus from which cytosolic ribosomes are excluded.

Golgin proteins of the Golgi matrix and the Golgi membranes are regulated by the Rab family of guanosine triphosphate (GTP)-binding proteins. Rabs comprise a large protein family that regulates many aspects of vesicular trafficking, membrane fusion, and organelle morphology. There are over 15 Rab proteins that localize to the Golgi apparatus in mammalian cells.

The Subcellular Localization of Golgi Complexes

The subcellular distribution of the Golgi complex varies among eukaryotic organisms. In plants and invertebrates the stacked Golgi complexes are typically dispersed throughout the cell. In these cell types, the dispersed Golgi membranes are often positioned near transport-vesicle-exit sites on the endoplasmic

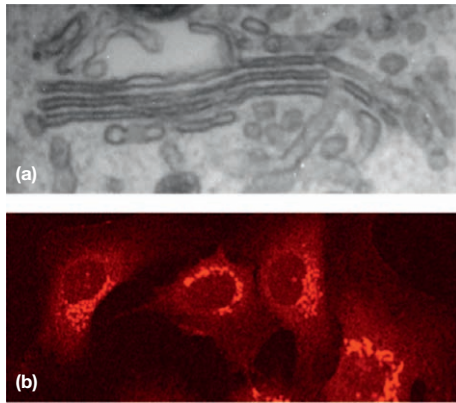


Figure 1 The morphology and localization of the mammalian Golgi complex. (a) A transmission electron micrograph of a Golgi complex from a rat kidney cell. Note that it is composed of three stacked cisternae: *cis*, medial, and *trans*. Transport vesicles are apparent near the Golgi complex as circular profiles. (b) A fluorescence micrograph of normal rat kidney cells that are decorated with an antibody against the Golgi enzyme, mannosidase II. Note that the labeled Golgi stacks (red) are concentrated in a region next to the cell nucleus.

reticulum. The Golgi complex in plants is dispersed through cytoplasmic streaming that results from myosin motor-based motility along actin microfilaments. In vertebrate cells, the stacked Golgi complexes are frequently localized together at a single site adjacent to the cell nucleus (**Figure 1(b)**). This site is near the centrosome, a cellular structure involved in the assembly and organization of microtubules, which are one of the key components of the cell cytoskeleton. In mammalian cells, the juxtanuclear localization of the Golgi complex involves the translocation of pre-Golgi vesicles along microtubules using a molecular motor protein called 'dynein'. Dynein motor protein also plays a role in organizing the Golgi cisternae into the interconnected ribbon-like structures.

Golgi positioning is related to cell polarity

In polarized cells of vertebrates, such as intestinal epithelial cells, the Golgi complex is localized in a polarized manner between the nucleus and the apical cell surface. Similarly, the Golgi apparatus is localized between the nucleus and the leading edge of motile cells. The advantage of maintaining a compact Golgi complex adjacent to the nucleus in vertebrate cells is still not fully understood, but it is likely to facilitate the polarized delivery of post-Golgi transport vesicles toward sites of rapid membrane growth such as the leading edge of a motile cell.

The Vertebrate Golgi Apparatus Serves as a Microtubule-Organizing Center

The morphology and positioning of the Golgi apparatus in vertebrate cells are highly dependent on the microtubule cytoskeleton. There are only a few sites called 'microtubule-organizing centers' where microtubule assembly is initiated in cells. By contrast, actin microfilament assembly can be initiated at many different sites within the cell and at the plasma membrane. Historically, the nucleation of new microtubules was thought to be restricted to the centrosomes and

basal bodies, cellular structures characterized by the presence of centrioles. It is now clear the Golgi complex is also able to nucleate microtubule assembly. Both centrosomes and Golgi membranes nucleate microtubules through a mechanism that requires the gamma-tubulin ring complex, γ -TURC. Nevertheless, Golgi-derived and centrosome-derived microtubules are distinct from each other. Golgi-derived microtubules are able to bind to distinct isoforms of the cytoplasmic linker associated protein (CLASP) family of microtubule plus-end-binding proteins. One role for the Golgi-apparatus-derived microtubules is to establish membrane connections between Golgi stacks and maintain the ribbon-like morphology. Although there is variability in the morphology and organization of the Golgi complex among eukaryotic cells, in all cases, the Golgi is a critical site for processing and sorting of proteins that are transiting the secretory pathway.

Protein Processing in the Golgi Complex

Protein Glycosylation

Many proteins, especially those localized to the cell surface, are covalently attached to complex sugar moieties called 'oligosaccharides' to form glycoproteins. One of the best-characterized functions for the Golgi apparatus is in the addition and modification of oligosaccharides on glycoproteins. Many proteins are modified at specific asparagine amino acid residues in the endoplasmic reticulum in a process called 'N-linked glycosylation', which is carried out by the enzyme oligosaccharyl transferase. In the Golgi complex, additional oligosaccharide chains can be attached to certain serine or threonine amino acid residues via the enzyme peptidyl N-acetylgalactosamine-transferase. This process is referred to as 'O-linked glycosylation'. Both N-linked and O-linked oligosaccharides are further modified in the Golgi complex through both the removal and addition of specific sugars. The processing of N-linked oligosaccharides by the Golgi complex is understood in some detail. In the Golgi complex, mannose is trimmed from N-linked oligosaccharides through the action of mannosidase II. This is followed by the addition of N-acetylglucosamine, galactose, and sialic acid by sugar-specific transferases. The enzymes that catalyze the various processing steps are spatially organized throughout the Golgi stacks in the order in which they function. For example, N-acetylglucosamine transferase is mostly enriched in the *cis*-Golgi cisterna, while sialic acid transferase is found predominantly in the *trans*-cisterna. The role of protein glycosylation is still not fully understood, but one described function includes protein sorting.

Proteolytic Processing

Some proteins must be cleaved within the secretory pathway before they are functional. Examples of such proteins include various proteases, hormones, and transcription factors.

Prohormone processing

Many peptide hormones are first synthesized as part of a larger inactive polypeptide, which is then cleaved by enzymes called 'endoproteases' into a smaller active form. For example, the metabolic hormone, insulin, folds first as a long polypeptide,

which is then cleaved at two sites releasing an inactive fragment, the C-peptide, to form the fully active insulin molecule. Many of these proteolytic processing steps occur in the *trans*-Golgi network and in post-Golgi secretory vesicles. The endoproteases that carry out this processing are generally referred to as prohormone convertases. One *trans*-Golgi-network-localized endoprotease, furin, has been extensively characterized and is involved in processing a wide variety of receptors and hormones.

Proteolytic regulation of transcription factors

Regulation of a transcription factor involved in cholesterol metabolism

In some cases, Golgi-localized endoproteases are used for the activation of proteins that function as part of regulatory signaling pathways. A very interesting example of this type of regulation is the proteolytic activation of sterol-response-element-binding protein (SREBP). SREBP is a transcription factor that activates genes that are involved in cholesterol uptake and biosynthesis. It resides as an integral membrane protein on the endoplasmic reticulum when cellular cholesterol levels are elevated. However, when cholesterol levels in the cell are low, SREBP is transported out of the endoplasmic reticulum to the Golgi complex where it is cleaved by two endoproteases. One of these enzymes is unique since it cleaves the transmembrane domain of SREBP within the plane of the membrane. The result of this proteolysis in the Golgi complex is the release of an active transcription factor from the membrane, which is then translocated into the nucleus to activate genes involved in sterol metabolism.

A Golgi-localized endoprotease is involved in Alzheimer's disease

The toxic peptide β -amyloid accumulates in plaques found within the brain of individuals with Alzheimer's disease. β -Amyloid is generated through inappropriate proteolytic cleavage of amyloid precursor protein. A Golgi-localized endoprotease called 'presenilin' is important for normal processing of amyloid precursor protein. Presenilin is similar to the Golgi-localized endoprotease that activates SREBP since it cleaves the amyloid precursor protein within the membrane.

The examples of endoprotease function described here are illustrative of the wide variety of proteins and cellular processes that utilize proteolytic processing at the Golgi complex.

Mechanisms of Protein Transport at the Golgi Complex

Vesicular Transport at the Golgi Complex

One important mechanism for moving proteins and phospholipids to, from, and within the Golgi complex is via transport vesicles. These small, spherical, membrane-limited structures are formed on organelles through the assembly of a protein coat. The coat deforms the membrane into a vesicle and plays a role in selecting the appropriate cargo proteins. Once formed, the vesicles are translocated to the target organelle where they fuse with the membrane and deliver their contents. Several types of transport vesicles, each with a distinct protein coat, function at the Golgi complex. Two well-studied examples are clathrin-coated vesicles that mediate transport of proteins leaving the *trans*-Golgi network, and coat-protomer I (COPI)

vesicles that mediate transport among the cisternae and backwards transport out of the Golgi complex.

The clathrin complex binds to membranes via adaptor proteins during the formation of transport vesicles. The clathrin adaptor proteins adaptor protein 1 (AP-1) and Golgi-localized, gamma adaptin ear-containing, ARF binding (GGA) that bind to the *trans*-Golgi network for anterograde post-Golgi trafficking are distinct from the AP-2 adaptor protein used for clathrin-mediated endocytosis from the plasma membrane. COPI vesicles form on Golgi membranes using the seven-subunit protein complex called 'coatamer'. The recruitment of both clathrin adaptors and coatamer is regulated by the small GTP-binding protein called 'ARF' (ADP-ribosylation factor). The ARF1 isoform localizes at the Golgi apparatus and can trigger the assembly of both clathrin and COPI vesicles. The GTP exchange reaction that activates ARF1 is inhibited by the fungal metabolite brefeldin A. Brefeldin A is commonly used in experimental cell biology to study the consequences of disrupting the secretory pathway.

Cisternal Progression

A second mechanism for moving proteins through the Golgi complex is cisternal progression. Here, the *cis*-cisternae are transformed into medial cisternae, and medial cisternae are transformed into *trans*-cisternae. New *cis*-cisternae are created by vesicles arriving from the endoplasmic reticulum. The *trans*-cisternae are expended by the formation of transport vesicles involved in post-Golgi trafficking. Progressive changes in the repertoire enzymes within the cisternae occur through vesicle-mediated protein transport and sorting of the resident Golgi proteins. During cisternal progression, forward-moving cargo proteins are carried along with the cisternae as they progress from the *cis* to medial to *trans* positions.

Protein Sorting at the Golgi Complex

Proteins must be properly sorted and targeted as they traverse and exit the Golgi complex. At the *trans*-Golgi network, proteins are destined for several different locations. Many proteins are sent to the cell surface for insertion into the plasma membrane or for release into the extracellular space. Other proteins must be targeted to intracellular organelles such as late endosomes or lysosomes. In addition to the sorting of forward-moving proteins, some proteins are retrieved or recycled back to earlier compartments. For example, proteins that function within the Golgi complex are recognized and retained at or retrieved to the proper cisterna. Two sorting mechanisms that have been characterized at the Golgi complex are receptor-mediated packaging into transport vesicles and protein aggregation.

Receptor-mediated sorting into transport vesicles

A key aspect of protein sorting throughout the secretory pathway is the selective packaging of cargo proteins into transport vesicles. This involves either direct or indirect (receptor-mediated) interactions with the vesicle-coat proteins. Thus, a vesicle targeted for a specific organelle in the cell can select cargo proteins that are destined for the same location. Discussed next are two receptor proteins that function in cargo selection during protein transport from the Golgi complex, namely the mannose-6-phosphate receptor and the KDEL receptor.

The mannose-6-phosphate receptor

The mannose-6-phosphate receptor sorts proteins into clathrin-coated vesicles that are leaving the *trans*-Golgi network and are destined for organelles called lysosomes involved in breaking down cellular waste products. Mannose-6-phosphate is an N-linked oligosaccharide generated by enzymes present in the *cis*-Golgi cisterna. It is formed specifically on proteins destined for the lysosome. The mannose-6-phosphate receptor spans the Golgi membrane and binds to coat proteins on the cytosolic side of the membrane and to the mannose-6-phosphate-containing protein on the luminal side of the membrane. Thus, mannose-6-phosphate-containing proteins are sorted and concentrated into the correct clathrin-coated vesicles for transport to the lysosome.

The KDEL receptor

Many resident proteins of the endoplasmic reticulum are retrieved from the Golgi complex if they escape. The KDEL receptor functions in this retrieval process by sorting proteins that belong in the endoplasmic reticulum into COPI vesicles for backwards transport. KDEL refers to the single-letter code for the amino acid residues lysine, aspartic acid, glutamic acid, and leucine, which are present on many endoplasmic reticulum-resident proteins. Similar to the mannose-6-phosphate receptor, the KDEL receptor binds to COPI-coat proteins on the cytosolic side of the membrane and to KDEL-containing protein on the luminal side. This allows for sorting and retrieval of endoplasmic reticulum-localized proteins via COPI vesicles.

Protein aggregation as a sorting mechanism

Another way that proteins are sorted in the Golgi complex is through aggregation. Some resident Golgi proteins, such as the enzyme mannosidase II, form aggregates with themselves and other Golgi-resident proteins. Golgi proteins may be retained at the proper cisternae through the exclusion of the large aggregates from transport vesicles. By contrast, some proteins – for example, insulin and many other hormones – may cluster within the *trans*-Golgi network to facilitate their concentration and packaging into vesicles. Aggregation continues in the secretory vesicles before its delivery to the plasma membrane and release into the extracellular space.

The Golgi Cisternae Serve as an Intracellular Calcium Store

Together with the endoplasmic reticulum and mitochondria, the Golgi apparatus acts as a calcium store. The spatially and temporally regulated release of calcium from intracellular stores contributes to a broad range of cellular signaling observed in all eukaryotes including plants, fungi, and vertebrates. Cell metabolism, hormone secretion, and neuronal transmission are a few examples of processes that are regulated by agonist-triggered release of calcium. Calcium enters the lumen of Golgi cisternae through the action of adenosine triphosphate (ATP)-dependent pump proteins. Golgi-specific members of the sarcoplasmic/endoplasmic reticulum calcium ATPase (SERCA)-family are one class of proteins that pump calcium into the Golgi lumen. A second family of Golgi pump proteins belong to the secretory-pathway calcium ATPase (SPCA) family that catalyzes the active transport of both

calcium and manganese ions. PMR1 is an example of a Golgi-specific SPCA pump that has been characterized in yeast. Calcium is buffered in the lumen of the Golgi apparatus by Golgi-specific calcium-binding proteins such as CALNUP. Calcium release is triggered from both the Golgi apparatus and the endoplasmic reticulum by the second messenger inositol trisphosphate (IP₃). IP₃ binds to calcium-specific channel proteins, allowing calcium to exit the Golgi lumen into the cytosol.

The familial skin disorder, Hailey–Hailey disease, occurs in humans with only one allele of the Golgi calcium pump SPCA1. Hailey–Hailey disease is characterized by skin blistering and erosion. The disruptions in intracellular calcium dynamics that occur from decreased calcium ATPase activity at the Golgi complex causes defects in cell adhesion among skin keratinocytes leading to the painful disease phenotypes.

Summary

The Golgi apparatus plays a central role in the function of the secretory pathway. In eukaryotic cells, thousands of proteins can be properly transported, processed, and sorted as they move through the Golgi complex. The Golgi complex adopts a complex morphology and can be precisely positioned to carry out these tasks. The ability of the Golgi apparatus to act as a microtubule-organizing center and the complex connections between Golgi stacks to form the stacked ribbon-like morphology are recently studied features of the Golgi. Future studies by cell biologists are certain to uncover unexpected features of this complex cellular organelle at the center of the secretory pathway.

See also: [Cell Architecture and Function: Dynein](#); [Lipids](#); [Carbohydrates](#); [Membranes and Membrane Proteins](#); [Free Oligosaccharides: Formation and Degradation \(Free Oligosaccharides Related to N-Linked Glycans\)](#); [N-Linked Glycan-Processing Glucosidases and Mannosidases](#); [Secretory Pathway](#); [Metabolism](#); [Vitamins and Hormones](#); [Cholesterol Synthesis and Regulation](#); [Protein/Enzyme Structure Function and Degradation](#); [Amyloid](#).

Further Reading

- Beck R, Rawet M, Wieland FT, and Cassel D (2009) The COPI system: Molecular mechanisms and function. *FEBS Letters* 583: 2701–2709.
- Emr S, Glick BS, Linstedt AD, et al. (2009) Journeys through the Golgi – taking stock in a new era. *Journal of Cell Biology* 187: 449–453.
- Goud B and Gleeson PA (2010) TGN golgins, Rab and cytoskeleton: Regulating the Golgi trafficking highways. *Trends in Cell Biology* 20: 329–336.
- Lowe M (2011) Structural organization of the Golgi apparatus. *Current Opinion in Cell Biology* 23: 85–93.
- Mironov AA and Pavelka M (eds.) (2010) *The Golgi Apparatus: State of the Art 110 Years after Camillo Golgi's Discovery*. New York: Springer.
- Mowbray K and Dacks JB (2009) Evolution and diversity of the Golgi body. *FEBS Letters* 583: 3738–3745.
- Nakano A and Luini A (2010) Passage through the Golgi. *Current Opinion in Cell Biology* 22: 471–478.
- Pizzo P, Lissandron V, Capitanio P, and Pozzan T (2011) Ca(2+) signalling in the Golgi apparatus. *Cell Calcium* 305: 1802–1803.
- Ungar D (2009) Golgi linked protein glycosylation and associated diseases. *Seminars in Cell and Developmental Biology* 20: 762–769.
- Wei JH and Seemann J (2010) Unraveling the Golgi ribbon. *Traffic* 11: 1391–1400.
- Wilson C, Venditti R, Rega LR, et al. (2011) The Golgi apparatus: An organelle with multiple complex functions. *Biochemical Journal* 433: 1–9.

G-Protein-Coupled Receptor Kinases and Arrestins

N Zhukovskaya, B Barker, D S Kang, and J L Benovic, Thomas Jefferson University, Philadelphia, PA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Jeffrey L. Benovic, volume 2, pp. 152–157, © 2004, Elsevier Inc.

The GRK Family

GRKs specifically phosphorylate activated GPCRs and initiate the recruitment of arrestins, which mediate receptor desensitization, endocytosis, and downregulation (Figure 1). GRKs are serine/threonine protein kinases found in metazoans. In mammals, there are seven GRKs that can be divided into three subfamilies based on overall structural organization and homology: GRK1 (also termed 'rhodopsin kinase') and GRK7; GRK2 (β ARK1) and GRK3 (β ARK2); and GRK4, GRK5, and GRK6. GRK1 and 7 are expressed in visual rod and cone cells, respectively, GRK4 is primarily expressed in testes, kidney, and brain, and GRK2, 3, 5, and 6 are ubiquitously expressed. *Drosophila melanogaster* (GPRK1 and GPRK2) and *Caenorhabditis elegans* (GRK-1 and GRK-2) each express a GRK2 (GPRK1 and GRK-2) and GRK4 (GPRK2 and GRK-1) subfamily ortholog.

GRKs have a tripartite modular structure. A central ~330 amino acid catalytic domain, most related to those of other AGC kinases such as PKA and PKC is flanked by an ~180 residue N-terminal region that contains a regulator of G-protein signaling homology (RH) domain, and an ~80–180 amino acid C-terminal lipid-binding domain that varies in structure (Figure 2(a)).

Role of GRKs in GPCR Regulation

GRKs specifically phosphorylate agonist-occupied GPCRs and initiate the process of desensitization. While no clear consensus sequence for GRK-mediated phosphorylation has been identified, specific sites for GRK-mediated phosphorylation have been mapped to either the carboxyl-terminal tail or the third intracellular loop of many GPCRs. A wide variety of strategies have been used to define the specificity of GPCR phosphorylation by GRKs including *in vitro* analysis as well as intact cell studies using overexpression, dominant negative mutants, inhibitory antibodies, antisense constructs, and RNA interference. While these efforts have started to dissect the specificity of GRK/GPCR interaction, the fact that most cell lines express a large number of GPCRs as well as several GRKs makes it virtually impossible to completely define such specificity. Nevertheless, some receptors appear to be primarily phosphorylated by a single GRK while others are regulated by multiple GRKs in a cell-type-dependent manner.

Insight into GRK specificity/function has also been gained from *in vivo* strategies. For example, transgenic mice with cardiac-specific overexpression of GRK2 lead to increased desensitization of the β -adrenergic receptor (β AR) and an associated decrease in left ventricular contractility while mice overexpressing a GRK2 inhibitor demonstrate increased left ventricular contractility even in the absence of β -agonist. Cardiac-specific overexpression of GRK5 in transgenic mice has a similar effect on β AR-mediated signaling

and contractility, while the angiotensin II-mediated contractile response is attenuated only in the GRK2-overexpressing mice. Further demonstration of specificity is suggested in transgenic mice with myocardial overexpression of GRK3 where no effect on β AR response is observed. Moreover, studies involving cardiac overexpression of individual GRKs in mice revealed that GRK3 > GRK5 > GRK2 at desensitizing α_{1B} -adrenergic receptor signaling in the heart. These studies suggest some level of GRK specificity is evident *in vivo* even when the GRKs are overexpressed.

Perhaps the most insight into GRK specificity and function has been gained from functional knock-outs. GRK1 knock-out mice have a prolonged response to light and undergo apoptotic retinal degeneration due to the loss of rhodopsin phosphorylation and also have a slowed cone recovery. Disruption of the GRK2 gene in mice results in embryonic lethality as a consequence of hypoplasia of the ventricular myocardium, suggesting a role for GRK2 in cardiac development, although studies that selectively ablated GRK2 in cardiomyocytes produced viable mice with a normal heart structure and function. Disruption of the mouse GRK3 gene results in normal development and physiology, although specific defects in desensitization of odorant-induced cyclic adenosine monophosphate (cAMP) responses in olfactory epithelium, cholinergic response of airway smooth muscle, and κ -opioid receptor-mediated responses are observed. A GRK4 knock-out mouse exhibits no obvious phenotypes while a GRK5 knockout results in muscarinic supersensitivity and impaired receptor desensitization in the lung but not in the heart. Disruption of the GRK6 gene results in altered CXCL12-promoted chemotaxis of lymphocytes, supersensitivity to the locomotor-stimulating effects of cocaine and amphetamine, and enhanced coupling of postsynaptic D2-like dopamine receptors.

GRKs have also been characterized in lower eukaryotic animals. For example, a mutation that disrupts expression of GPRK2 leads to specific defects in *Drosophila* embryogenesis, possibly due to a role in Hedgehog signaling, while defects in Hedgehog signaling are also observed with knockdown of GRK2/3 in zebrafish. A mutation that results in reduced activity of GRK-2 in *C. elegans* results in a defect in chemosensory responses. Taken together, these findings reveal that GRKs are involved not only in regulating signaling but may also have critical roles in regulating growth and development.

GRKs Have Diverse Functional Roles

While GRKs specifically phosphorylate activated GPCRs, they also phosphorylate a variety of other proteins including Smoothed, the epidermal growth factor (EGF) and platelet-derived growth factor (PDGF) receptors, tubulin, synucleins, phosphodiesterase γ (PDE γ), phosphducin, ezrin, R-Smads, downstream

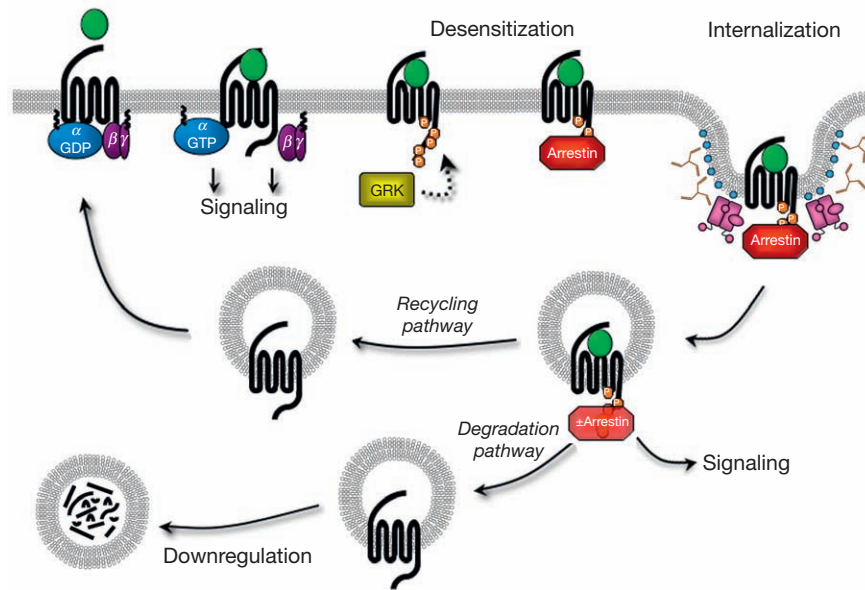


Figure 1 Schematic representation of GPCR trafficking. Agonist-activated receptors are phosphorylated by a GRK leading to the recruitment of arrestins. Arrestins serve as adaptor proteins by virtue of their ability to link receptors to components of the transport machinery such as clathrin, AP-2, and phosphoinositides leading to internalization. Additional interactions with GRKs and arrestins appear to play a role in this process. Once in endosomes, recycling receptors (recycling pathway) are readily segregated from receptors destined for lysosomes (degradation pathway). Although poorly understood, the sorting decision may be regulated by specific and distinct protein interactions or modification by ubiquitin. Receptors that enter the recycling pathway are sorted into recycling endosomes and traffic back to the cell surface resulting in resensitization. Receptors that enter the degradation pathway are trafficked to lysosomes where they are proteolyzed leading to a loss in the total cellular complement of the receptor, a process known as 'downregulation'. Internalized GPCR–arrestin complexes have also been shown to regulate intracellular signaling pathways such as MAP kinases.

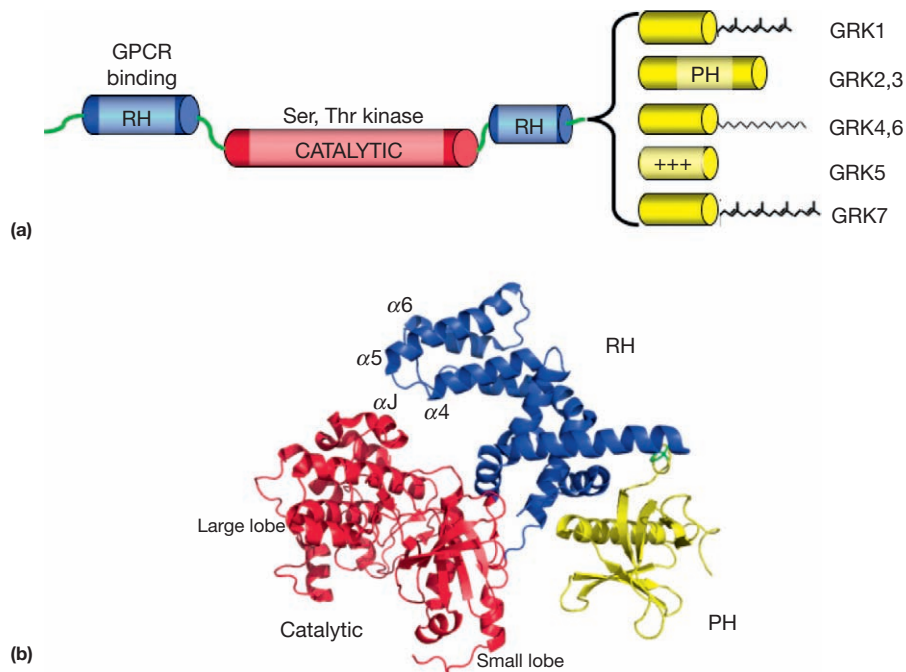


Figure 2 The structure of GRK. (a) Schematic representation of the GRK structure. GRKs have a conserved regulator of G-protein signaling homology (RH) domain (shown in blue). The conserved catalytic domain is shown in red. In yellow is the C-terminal domain: GRK2 and GRK3 possess pleckstrin homology (PH) domain whereas GRK5 contains a polybasic region involved in phospholipid binding. Other GRKs are posttranslationally modified: GRK1 is farnesylated, GRK4 and 6 are palmitoylated, and GRK7 is geranylgeranylated. (b) Crystal structure of GRK2. The structure was solved as part of the G_{α_q} -GRK2- $G_{\beta\gamma}$ complex. The RH domain is shown in blue, the catalytic domain in red, and the PH domain in yellow.

regulatory element antagonist modulator (DREAM), histone deacetylase 5 (HDAC5), arrestin2, p38, p53, p105, and I κ B α (Table 1). The ability of GRK2 to phosphorylate Smoothed appears to regulate arrestin binding and function in a positive manner in Hedgehog signaling while GRK2 and GRK5 can phosphorylate and desensitize the PDGF receptor in a cell-type-dependent manner. GRK2 phosphorylation of ezrin, a protein involved in cytoskeletal reorganization, has been linked to β_2 AR endocytosis. The synucleins serve as effective

GRK substrates *in vitro* and in cells although it is currently unclear whether GRKs regulate synuclein phosphorylation *in vivo*. GRK2 has also been shown to phosphorylate Thr-123 in p38 resulting in inhibition of p38 activity while GRK2 phosphorylation of DREAM at Ser-95 attenuates expression of the Kv4.2 potassium channel. GRK2 can also phosphorylate R-Smads upon transforming growth factor β (TGF- β) stimulation resulting in inhibition of TGF- β signaling. Phosducin and phosducin-like protein are both phosphorylated by GRK2 *in vitro* resulting in a reduced

Table 1 GRK non-GPCR substrates and interacting targets

Target	GRK subtype	Functional role
<i>Substrates^a</i>		
Arrestin2	GRK5 > GRK2	Phosphorylation inhibits G-protein-independent ERK1/2 signaling downstream of the 5-HT ₄ receptor
DREAM	GRK2 > GRK6	Phosphorylation reduces membrane expression of Kv4 potassium channels
EGFR	GRK2	Unknown
Ezrin	GRK2	Phosphorylation regulates actin rearrangement and receptor internalization
HDAC5	GRK5	Phosphorylation regulates the transcription factor MEF2 and alters gene transcription
NF κ B1 p105	GRK5 > GRK2	Enhances phosphorylation by I κ B kinases, thus promoting p105 degradation
p38	GRK2	Reduces activation of p38 by MKK6, thus reducing p38 activity and p38-dependent cell differentiation and cytokine release
p53	GRK5	Phosphorylation regulates p53 degradation and inhibits radiation-induced apoptosis
PDE γ	GRK2	Promotes EGF- and thrombin-dependent ERK1/2 activation
PDGFR	GRK2 and GRK5	Attenuates PDGF-dependent signaling responses
Phosducin	GRK2 = GRK3 = GRK5	Reduces interaction with G $\beta\gamma$ <i>in vitro</i>
R-Smads	GRK2	Inhibits TGF β -promoted gene transcription, cell growth, and apoptosis
Smoothed	GRK2	Promotes arrestin binding and positively influences Hedgehog signaling
Synucleins	α – GRK2 = GRK5	Reduces binding to phospholipids and antagonizes PLD2 inhibition <i>in vitro</i>
(α , β , and γ)	β – GRK2 > GRK5	
	γ – GRK5 > GRK2	
Tubulin	GRK5 > GRK2 > GRK1	Unknown
<i>Interacting proteins and molecules</i>		
α -Actinin	GRK1, GRK2, and GRK5	Inhibits GRK1, GRK2, and GRK5
Actin	GRK5	Inhibits GRK5 activity
Akt	GRK2	GRK2 inhibits Akt
Caveolin	GRK2	Inhibits GRK2
Ca ²⁺ /CaM	GRK5 > GRK2	Inhibits GRK2 with IC ₅₀ = 2 μ M and GRK5 with an IC ₅₀ ~40 nM
Clathrin	GRK2	Regulates GRK2-mediated β_2 AR phosphorylation and internalization
EGFR	GRK2	Phosphorylates and activates GRK2
ERK1/2	GRK2	Inhibits GRK2 activity
G α q	GRK2	GRK2 binds and sequesters activated G α q
G $\beta\gamma$	GRK2	Recruits GRK2 to the plasma membrane
GIT1/2	GRK2	Not fully elucidated; may regulate cell migration
HSP90	GRK2, GRK3, GRK5, and GRK6	Required in GRK folding and maturation
MEK	GRK2	MEK/GRK2 interaction affects downstream signaling
NOS	GRK2	Inhibition of GRK2 by S-nitrosylation of Cys-340
Phospholipids	GRK2 and GRK5	Stimulates GRK2 and GRK5 activity
PDGFR	GRK2 and GRK5	Phosphorylates and activates GRK2 and GRK5
PI3K	GRK2	GRK2 recruits PI3K to the plasma membrane
PIP ₂	GRK2	Activates GRK2
PKA	GRK1, GRK2, and GRK7	Phosphorylates and attenuates activity of GRK1 and GRK7: enhances activity of GRK2
PKC	GRK2 and GRK5	Phosphorylation reverses Ca ²⁺ /CaM inhibition of GRK2 and inhibits GRK5 activity
PTCH1	GRK2	Regulates cell-cycle progression
Recoverin	GRK1	Inhibits GRK2 activity
RKIP	GRK2	Inhibits GRK2 activity
c-Src	GRK2	Phosphorylates and activates GRK2

^aDREAM: downstream regulatory element antagonist modulator; EGFR: epidermal growth factor receptor; ERK: extracellular signal-regulated kinase; HDAC5: histone deacetylase 5; NF κ B: nuclear factor κ B; MEF2: myocyte enhancer factor-2; MKK6: mitogen-activated protein kinase kinase 6; PDE γ : phosphodiesterase γ ; PDGFR: platelet-derived growth factor receptor; PTCH1: patched homolog 1; TGF β : transforming growth factor β ; PLD2: phospholipase D2; CaM: calmodulin; GIT1/2: G-protein-coupled receptor kinase-interacting protein 1/2; HSP90: heat shock protein 90; MEK: mitogen-activated protein kinase 1; NOS: nitric oxide synthase; PI3K: phosphoinositide 3 kinase; PIP₂: phosphatidylinositol 4,5-bisphosphate; PKA: protein kinase A; PKC: protein kinase C; RKIP: Raf kinase inhibitor protein.

ability of these proteins to bind $G\beta\gamma$ subunits. HDAC5 is a histone deacetylase that can be phosphorylated by GRK5 resulting in the inhibition of transcriptional repressor activity. GRK5 can also phosphorylate the nuclear factor κB (NF- κB) protein p105, resulting in decreased phosphorylation of p105 by $I\kappa B$ kinase (IKK) and attenuated signaling, while both GRK2 and GRK5 phosphorylate $I\kappa B\alpha$ and enhance TNF α -induced NF- κB signaling. In addition, GRK5 has recently been implicated in genome stability via the ability to phosphorylate Thr-55 in the tumor suppressor p53 and mediate p53 degradation.

While GRKs mediate many of their effects via protein phosphorylation, GRKs also interact with additional proteins and regulate their function (Table 1). For example, the N-terminal RH domain of GRK2 and GRK3 interact with and effectively inhibit $G\alpha_q$, $G\alpha_{11}$, and $G\alpha_{14}$ while GRK2 interacts with and inhibits MEK1. GRK2 also interacts with several proteins implicated in the endocytic process including clathrin, GRK interactor 1 (GIT1), and PI 3-kinase. GIT1, an ARF guanosine triphosphatase (GTPase)-activating protein, has been implicated in regulating the trafficking of receptors internalized via an arrestin- and dynamin-dependent pathway and its interaction with GRK2 may regulate cell migration. GRK2 interaction with PI 3-kinase regulates the recruitment of PI 3-kinase to the plasma membrane, where it helps modulate GPCR endocytosis. The GRK2/PI 3-kinase pathway appears to involve enhanced recruitment of AP-2 to the receptor via the increased production of D-3 phospholipids. Taken together, these studies suggest that GRKs orchestrate the formation of protein complexes that regulate receptor signaling and trafficking.

Regulation of GRK Function

A critical component in generating specificity is the molecular mechanism responsible for regulating the activity and cellular localization of GRKs. For example, phosphorylation appears to play an important role in regulating GRK activity. GRK2 phosphorylation by ERK1/2 inhibits GRK2 activity while phosphorylation by PKA, Src, PKC, and the PDGF receptor results in increased activity (Table 1). GRK5 activity is attenuated by PKC phosphorylation but stimulated by autophosphorylation and PDGF receptor phosphorylation. Interestingly, GRK2 is also inhibited by nitrosylation of a cysteine residue within the activation loop of the catalytic domain. GRK function is also regulated by interaction with additional proteins including $G\alpha$ and $\beta\gamma$ subunits, caveolin-1, tubulin and actin, Raf kinase inhibitor, and various calcium binding proteins. Phospholipids also play an important role in regulating GRK function since $G\beta\gamma$ -mediated GRK2 activation is dependent on negatively charged phospholipids including phosphatidylinositol 4,5-bisphosphate (PIP₂). Interestingly, while GRK4 subfamily members do not bind $G\beta\gamma$ subunits, these kinases share a putative amino-terminal PIP₂-binding domain that may facilitate receptor phosphorylation. These kinases also have the ability to associate with phospholipids via a carboxyl-terminal domain that is either palmitoylated (GRK4 and 6) or can directly bind phospholipids (GRK5) (Figure 2(a)). Thus, the immediate phospholipid environment may have a general and critical role in modulating GRK function.

While GRKs have been implicated in the phosphorylation of plasma-membrane-localized receptors as well as various cytoplasmic proteins, several studies reveal that some GRKs can also localize in the nucleus. GRK5 contains a nuclear localization sequence and all GRK4 family members (GRK4, 5, and 6) can localize in the nucleus. Moreover, the ability of GRK6 to localize in the nucleus appears to be regulated by its palmitoylation state. While the functional role of GRKs in the nucleus is not fully understood, a broad-based shRNA screen for proteins involved in mitotic progression identified GRK2 and GRK5 as playing a potential role in HT29 cell mitosis. Recent studies also suggest that GRK2 degradation plays an important role in regulating normal cell-cycle progression. Moreover, GRK2 interaction with patched homolog 1 (PTCH1) inhibits cyclin B1 interaction with PTCH1 and regulates the cell cycle and early development in zebrafish. Taken together, multiple mechanisms including GPCR, G protein, and phospholipid binding, post-translational modification, and intracellular localization play an important role in regulating GRK activity and function. Moreover, some GRKs may play a role in regulating nuclear events such as transcription and cell-cycle regulation.

Structural Basis of GRK Function

The X-ray crystal structures of GRK1, GRK2, GRK2/ $G\beta_1\gamma_2$, GRK2/ $G\beta_1\gamma_2$ / $G\alpha_q$, and GRK6 have provided significant insight into GRK function. Interestingly, the crystal structure of GRK2 reveals that the N-terminal RH domain, central catalytic domain, and C-terminal pleckstrin homology (PH) domain form an equilateral triangle that is ~ 80 Å on a side (Figure 2(b)). The RH domain contacts both the kinase and PH domains. Moreover, the RH domain consists of two discontinuous regions with the characteristic nine-helix bundle in the N-terminal region and two additional helices following the kinase domain. The interaction surface of GRK2 with $G\alpha_q$ primarily resides on the α_5 and α_6 helices of the RH domain. The kinase domain is most similar to that of PKA, PKB, and PDK1 but appears to be in an inactive conformation in the crystal structure. This inactive conformation is also observed in the structure of GRK6 even when bound to AMPPNP. Interestingly, the structure of the RH and kinase domain cores of GRKs appear to have similarities to the inactive structure of Src with the α_{10} helix of the RH domain potentially functioning to regulate GRK activation. Similarly, the interface between the α_4 - α_5 loop of the RH domain and the α_j helix of the kinase large lobe might also modulate GRK activity by influencing the orientation of the large and small kinase lobes. The C-terminal PH domain of GRK2 mediates phospholipid binding, primarily via the inner face of the β_1 - β_2 loop, and $G\beta\gamma$ binding via four different regions including strands β_1 - β_4 and the extended α -helix. Docking analysis of the GRK2/ $G\beta\gamma$ complex with rhodopsin as well as the co-crystal structure of GRK2/ $G\beta\gamma$ / $G\alpha_q$ suggests that the receptor, $G\alpha_q$, and $G\beta\gamma$ should be able to bind to GRK2 simultaneously. This would represent an effective way of turning off signaling by phosphorylating receptor and sequestering $G\alpha_q$ and $G\beta\gamma$ thereby preventing interaction with effector molecules.

Since the GRK catalytic domains are in an inactive conformation in the crystal structures, it is important to understand

how these enzymes are activated. Interestingly, GPCR interaction appears to regulate the catalytic activity of the GRKs. While the mechanism of this activation is unclear, there is some evidence suggesting that the GRK N-terminal region functions as a switch to regulate the activation state of the GRK. However, since the N-terminal domains of GRK1, GRK2, and GRK6 are not fully observed in the crystal structures, it is not completely understood how this region functions to regulate GRK activation.

Role of GRKs in Disease

Various experimental approaches have led to a growing appreciation of the potential pathophysiological roles of GRKs in human disease. GRK1 mutation has been identified in Oguchi's disease (a form of night blindness) while several studies have demonstrated increased expression of GRKs associated with congestive heart failure (CHF), hypertension, and myocardial ischemia. In experimentally induced models of CHF, both GRK2 and GRK5 expression and activity were enhanced two- to threefold in the left ventricular myocardium early in CHF progression. While it is not clear whether this increase in GRK levels is causal or a byproduct of elevated hormones secondary to disease, it is interesting that myocardial ischemia, which is associated with a large local release of noradrenaline, also produces a rapid increase in GRK2 messenger RNA levels. Moreover, chronic stimulation with a β AR agonist also leads to a corresponding increase in myocardial GRK2 expression in mice, while infusion of angiotensin II into the aortas of hypertensive rats induces a rapid threefold increase in GRK5 expression as well as an increase in GRK2 expression. Conversely, a reduction in β -agonist levels results in decreased myocardial GRK2 expression in mice. Interestingly, lymphocyte GRK2 levels may provide a marker of myocardial GRK2 levels as well as a diagnostic in the treatment of heart failure patients. Moreover, lymphocyte GRK2 levels are also increased in Alzheimer's disease patients and the increase was correlated with the degree of cognitive decline. By contrast, lower levels of GRK2 and GRK6 are observed in lymphocytes from rheumatoid arthritis patients. Interestingly, GRK3 and GRK5 levels appear to be increased in Parkinson's disease with dementia while GRK3 levels may be decreased in some patients with bipolar disorder or with an immune deficiency disease called 'warts, hypogammaglobulinemia, infections, and myelokathexis (WHIM) syndrome'. Collectively, these findings support the notion that hormonal signals can directly regulate GRK expression, altered GRK activity may contribute to the disease state, and lymphocyte GRK2 levels may be diagnostic in cardiovascular and neurological disease.

Several studies have also shown that GRK expression is selectively regulated as a function of the hypertensive state. In a comparison of normotensive and mildly hypertensive subjects, GRK2 protein expression in lymphocytes was positively correlated with blood pressure and negatively correlated with β_2 AR-mediated adenylyl cyclase activity. These results suggest that GRK2 expression may underlie reduced β_2 AR responsiveness characteristic of the hypertensive state. Another form of hypertension (human essential hypertension) is genetic and involves impairment in dopamine promoted urinary sodium excretion. Defective dopamine receptor/G-protein

coupling in the kidney proximal tubule is the cause of the impaired renal dopaminergic action in this form of hypertension. Interestingly, recent studies suggest that single nucleotide polymorphisms of an isoform of GRK4 (GRK4 γ) increase GRK activity and cause dopamine receptor/G-protein uncoupling in the renal proximal tubule. Moreover, expression of polymorphic GRK4 γ -A142V in transgenic mice produces hypertension and studies suggest a positive correlation between certain GRK4 polymorphisms and human hypertension. Taken together, these findings suggest that GRKs play a role in vascular and renal control and may represent novel therapeutic targets for the treatment of human hypertension.

The Arrestin Family

The first arrestin (termed 'visual arrestin' or 'arrestin1') was identified in the visual system as a 48-kDa protein that redistributed from the cytoplasm to the disk membrane following light activation of rod cell outer segments. Arrestin1 is a 404-amino-acid protein and binds to the light-receptor rhodopsin in a light- and phosphorylation-dependent manner resulting in suppression of G-protein signaling. Gene targeting has demonstrated a clear physiological role for arrestin1 in quenching phototransduction in mice while mutations in human arrestin1 have been identified in Oguchi's disease.

Evidence for a broader involvement of arrestins in the regulation of GPCRs was initially suggested by the finding that purified arrestin1 could impair β_2 AR/G-protein coupling in a phosphorylation-dependent manner. Subsequent studies identified a 418-amino-acid arrestin1 homolog (termed 'arrestin2' or ' β -arrestin1') that could effectively uncouple phosphorylated β_2 AR from G protein. Arrestin2 is widely expressed and has been broadly implicated in regulating GPCR desensitization. Arrestin2 knock-out mice appear normal although they are more sensitive to β -agonist stimulation suggesting an *in vivo* role in β AR desensitization.

A third mammalian arrestin, termed 'arrestin3' or ' β -arrestin2', is 409 amino acids and is ubiquitously expressed. Arrestin3 has also been implicated in regulating the desensitization of numerous GPCRs. Similarly, knockout studies in mice suggest a role for arrestin3 in μ -opioid receptor desensitization while a positive role in CXCR4-mediated chemotaxis has been suggested. Surprisingly, a double knock-out of arrestin2 and 3 results in an embryonic lethal phenotype suggesting an important developmental role for these proteins. Taken together, these results suggest that arrestin2 and 3 have partially overlapping specificities for regulating development and GPCR signaling.

A fourth mammalian arrestin, termed 'arrestin4' or 'X-arrestin', is 388 amino acids and is expressed specifically in cone cells. The function of arrestin4 has not been characterized. In addition to the mammalian arrestins, numerous lower eukaryotic arrestins have been identified including three in *Drosophila*, one in *C. elegans*, and at least one in zebrafish. It is worth noting that arrestins have not been identified in yeast or *Dictyostelium* although there does appear to be a family of arrestin-related proteins termed ' α -arrestins' that are widely expressed. Arrestins have no enzymatic activity and primarily function as scaffolding proteins to bring various receptors, endocytic proteins, and signaling molecules together (Table 2).

Table 2 Diverse functions of nonvisual arrestins

<i>Interacting protein^a</i>	<i>Arrestin subtype</i>	<i>Functional role</i>
Receptors		
IGF-1R	Arrestin2	Scaffolding Mdm2/intracellular trafficking/MAPK signaling
TGFβRIII	Arrestin3	Receptor internalization/TGFβ signaling
Frizzled	Arrestin3	Receptor internalization/Wnt signaling
Smoothed	Arrestin3	Activation of Gli transcriptional factor
Notch	Kurtz (fruit fly)	Receptor degradation
Trafficking		
β2 Adaptin	Arrestin2 and 3	Clathrin-mediated endocytosis of receptors
Clathrin	Arrestin2 and 3	Clathrin-mediated endocytosis of receptors
ARF6	Arrestin2 and 3	Receptor endocytosis
ARNO	Arrestin2 and 3	Regulating ARF6 activity
NSF	Arrestin2	Clathrin-mediated endocytosis of receptors
eNOS	Arrestin3	S-nitrosylation/receptor endocytosis
Signaling		
ERK1/2	Arrestin2 and 3	MAPK signaling/reorganization of actin cytoskeleton/arrestin phosphorylation
JNK3	Arrestin2 and 3	MAPK signaling/shuttling between nucleus and cytosol
MKP7	Arrestin3	Inhibiting JNK3 signaling
Akt	Arrestin3	Activation of GSK3
c-Raf	Arrestin2 and 3	MAPK signaling
c-Src	Arrestin2 and 3	ERK1/2 activation and signaling
NFκB	Arrestin2 and 3	Inhibition of NFκB-dependent signaling
IκBα	Arrestin2 and 3	Sequestering IκBα–NFκB complex
TRAF6	Arrestin2 and 3	Inhibition of IκBα and NFκB activity
DGK	Arrestin2 and 3	Diacylglycerol degradation/modulating receptor-stimulated signaling
PDE4D	Arrestin2 and 3	cAMP degradation/modulating receptor-stimulated signaling
PP2A	Arrestin2 and 3	Dephosphorylation of arrestin2, receptors and Akt/inactivation of Akt
Yes	Arrestin2	Gαq stimulation and glucose transport
Dishevelled	Arrestin2 and 3	Frizzled internalization/Wnt signaling
Ubiquitination		
Mdm2	Arrestin3	Nuclear shuttling/arrestin ubiquitination/receptor endocytosis
Nedd4	Arrestin3	Arrestin ubiquitination/receptor degradation
AIP4	Arrestin2	Degradation of CXCR4 receptor
USP33 and USP20	Arrestin3	Arrestin deubiquitination/receptor recycling
Deltex	Kurtz (fruit fly)	Ubiquitination and degradation of Notch
Others		
Ral-GDS	Arrestin2 and 3	Actin cytoskeletal rearrangement
RhoA	Arrestin2	RhoA activation and stress-fiber formation
Filamin A	Arrestin2	Change in cell shape and mobility
Cofilin	Arrestin2 and 3	Chemotaxis/actin cytoskeletal rearrangement
Kif3A	Arrestin3	Smoothed translocation to the primary cilium
NHE5	Arrestin2 and 3	NHE5 endocytosis
Histone acetylase P300	Arrestin2	Chromatin reorganization and increased transcription of <i>c-fos</i> and <i>p27</i>
Hck	Arrestin2	Granule exocytosis in granulocytes
c-Fgr	Arrestin2	Granule exocytosis in granulocytes
MPZ-1	ARR-1 (worm)	IGF-1R signaling and longevity
DAF-18 (PTEN)	ARR-1 (worm)	IGF-1R signaling and longevity

^aIGF-1R: insulin-like growth factor 1 receptor; MAPK: mitogen-activated protein kinase; TGFβRIII: type III transforming growth factor-β receptor; ARF6: ADP-ribosylation factor 6; ARNO: ARF nucleotide exchange factor; NSF: *N*-ethylmaleimide-sensitive fusion protein; eNOS: endothelial nitric oxide synthase; ERK: extracellular signal-regulated kinase (MAPK); JNK3: c-Jun N-terminal kinase 3; MKP7: MAPK phosphatase 7; c-Src: sarcoma Schmidt-Ruppin A-2 viral oncogene homolog; NFκB: nuclear factor kappa-light-chain-enhancer of activated B cells; IκBα: NFκB inhibitor α; TRAF6: E3 ligase tumor necrosis factor receptor-associated factor 6; DGK: diacylglycerol kinase; PDE4D: cAMP-specific phosphodiesterase 4D; PP2A: protein phosphatase 2; Mdm2: murine double mutant E3 ubiquitin ligase; Nedd4: HECT ubiquitin ligase neural precursor cell expressed, developmentally downregulated 4; AIP4: HECT ubiquitin ligase atrophin-interacting protein 4; USP: ubiquitin-specific processing protease; Ral-GDS: Ral GDP dissociation stimulator; RhoA: Ras homolog gene family A; Kif3A: kinesin family member 3A; NHE5: Na⁺/H⁺ exchanger isoform 5; Hck: hemopoietic cell kinase; c-Fgr: Gardner-Rasheed feline sarcoma viral oncogene homolog; MPZ-1: multi-PDZ domain containing protein; PTEN: phosphatase and tensin homolog.

Molecular and Structural Nature of Arrestin Interaction with GPCRs

Initial structural insight on arrestins was provided by the X-ray crystal structure of bovine arrestin1 while the crystal structures of C-terminally truncated and wild-type bovine arrestin2 and salamander arrestin4 have also been solved. In general, arrestins have an elongated shape and are almost exclusively made up of β sheets and connecting loops with the exception of one short α helix (Figure 3). Arrestins are composed of two major domains termed the 'N-domain' and 'C-domain' that are held together by a polar core of buried salt bridges. The N- and C-domains are connected by a short hinge while the C-tail is connected by a flexible linker to the C-domain and contains a short β strand that interacts with a lateral β strand of the N-domain. A primary region that has been suggested to undergo substantial conformational rearrangement after receptor binding is the C-tail of arrestin. In addition, the hinge connecting the N- and C-domains may play an important role in this conformational rearrangement since reducing the length of the hinge results in reduced activation of arrestin. This is consistent with the notion that the N- and C-domains move relative to each other in the process of arrestin's transition into an active conformation.

An important feature of arrestin interaction with GPCRs is the ability of arrestins to recognize both the activation and

phosphorylation state of the receptor. The ability to discriminate between agonist-activated and non-activated GPCRs suggests that arrestins contain domains that specifically contact regions of the receptor exposed following receptor activation. The N-domain of arrestins retain the ability to recognize the agonist-activated state of GPCRs, suggesting that the activation-recognition region of all arrestins is largely contained within the N-terminal half. This N-domain receptor binding region has been localized to β -strands 5 and 6 and adjacent loops while a secondary receptor-binding domain is present within β -strands 15 and 16 in the C-domain (Figure 3). As expected, swapping these two receptor-binding elements between arrestin1 and arrestin2 reverses their receptor specificity.

The importance of receptor phosphorylation in arrestin binding has also been extensively explored. The primary phosphorylation-recognition domain within arrestin1 is localized near the middle portion of the protein, while positively charged residues within β -strand 1 also participate in phosphate binding and may function to facilitate binding within the primary phosphorylation-recognition domain. Previous studies also identified a phosphorylation switch residue (Arg-175) that plays a critical role in sensing the phosphorylation state of the receptor. The conserved nature of this mechanism is supported by the observation that mutation of the comparable residue in arrestin2 (Arg-169) and arrestin3 (Arg-170) also results in an activated arrestin that binds GPCRs in a

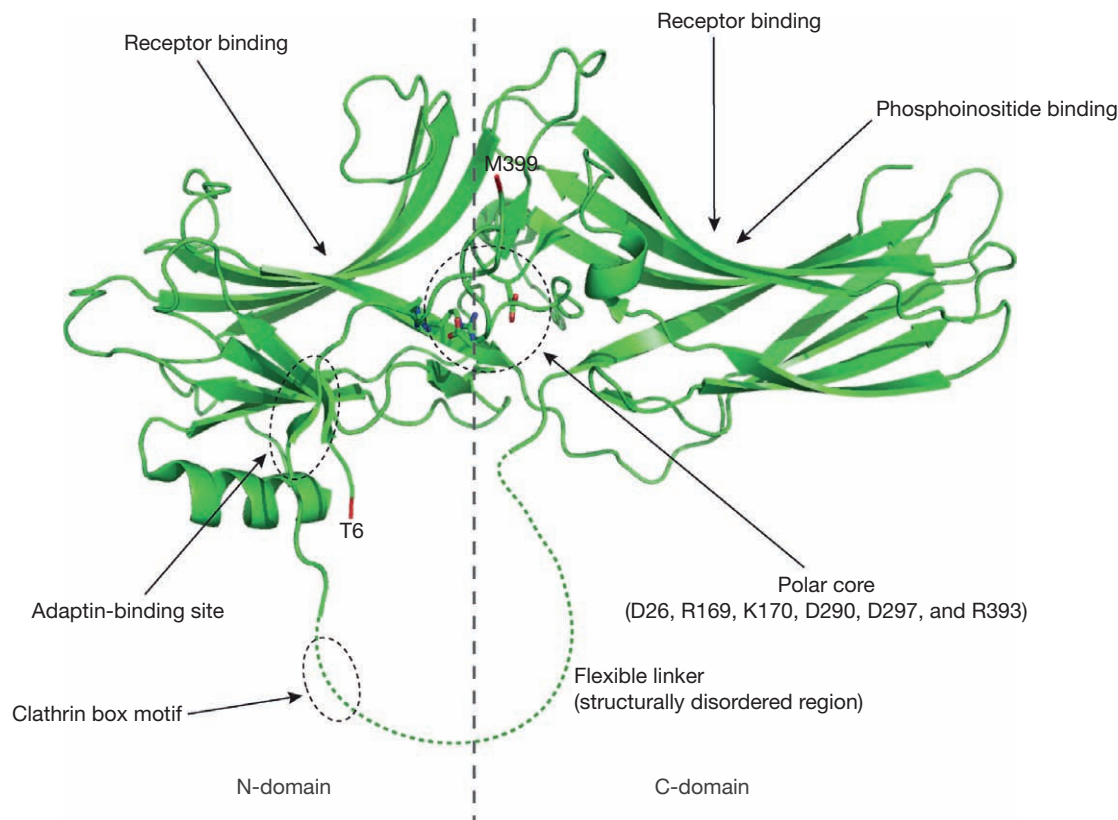


Figure 3 Structure of arrestin2 with major interfaces for protein-protein interaction. Crystallographic structure of arrestin2 is depicted as a green cartoon model from the N-terminus (Thr-6) to the C-terminus (Met-399) together with largely disordered C-terminal flexible linker (green dotted line). Receptor-binding sites and phospholipid-binding sites are shown as arrows in this structure, together with adaptin- and clathrin-binding sites and polar core region (black dotted circles).

phosphorylation-independent manner. The phosphorylation switch residue plays a central role in the formation of a polar core comprised of charged residues from the amino-terminus (Asp-29 in arrestin2), N-domain (Arg-169 and Lys-170), C-domain (Asp-290 and Asp-297), and C-tail (Arg-393), thus bringing different parts of the protein together to maintain a basal conformation (Figure 3). These residues are highly conserved suggesting that this structural element is conserved in all arrestins and critical for function. Because the buried side chains of the polar core achieve neutrality by an elaborate network of electrostatic interactions, it has been suggested that disturbance of the polar core by introduction of a phosphate group from the receptor promotes structural changes that result in an active conformation of the protein. Indeed, conformational changes induced by incorporation of an R175Q mutation in arrestin1 have been detected by small-angle X-ray scattering. Similarly, conformational changes induced by activating mutations in arrestin2 (e.g., R169E) have been detected by hydrogen exchange mass spectrometry. In addition, bioluminescence resonance energy transfer has been used to detect conformational changes in arrestin2 induced by receptor binding. Taken together, these studies suggest that activation of arrestin results in significant structural changes that will play an important role in mediating the various protein interactions and cellular functions of arrestin.

Based on the culmination of biophysical and biochemical data, a model for arrestin-receptor interaction has been proposed. Initially, arrestin is thought to probe the functional state of the receptor and bind via its activation and phosphorylation-recognition domains. Only when the receptor is activated and phosphorylated can arrestin achieve a high affinity interaction where the activation and phosphorylation-recognition sites simultaneously engage the receptor. Lysine residues (Lys-14 and Lys-15 and perhaps others in arrestin1) located in the N-terminus are suggested to guide receptor-attached phosphates toward the phosphorylation-recognition domain and trigger the phosphate-sensitive residue, Arg-175, initiating arrestin activation. The highly charged phosphate groups are proposed to disrupt the buried polar core of arrestin releasing the constraints imposed by the C-tail on both the N- and C-domains. Phosphate binding to Lys-14 and Lys-15 is also proposed to disrupt intramolecular contacts between β -strand I, α -helix I, and the C-tail further destabilizing the basal conformation of arrestin. The various domains then rearrange and transform arrestin into the activated state optimizing arrestin-receptor interactions and facilitating arrestin interaction with endocytic proteins.

Role of Arrestins in Receptor Trafficking

Radioligand binding and immunocytochemical techniques have demonstrated that many GPCRs undergo internalization. Multiple studies have implicated an important role for GRKs and arrestins in GPCR internalization and have revealed that arrestin interaction with clathrin and AP-2, the two major proteins in clathrin-coated pits, plays an important role in this process (Table 2). Nonvisual arrestins interact specifically, stoichiometrically, and with high affinity with clathrin, and immunofluorescence analyses demonstrate that activated

GPCR, arrestin, and clathrin all colocalize upon agonist addition. The primary clathrin binding determinants in arrestins are localized to hydrophobic and acidic residues in the carboxyl-terminal region (Figure 3). Interestingly, these residues constitute a clathrin box motif, $L\phi X\phi D/E$ (where ϕ is a bulky hydrophobic residue and X represents any polar amino acid), which is a consensus sequence found in clathrin-binding proteins. A similar mutagenesis study on clathrin localized the arrestin binding site to the extreme N-terminus of the heavy chain of clathrin. Interestingly, recent structural analysis of an arrestin2/clathrin complex confirms the importance of the clathrin-binding box in this interaction but also identified a secondary binding region that contributes to the dynamics of arrestin/clathrin interaction. Nonvisual arrestins also directly interact with the endocytic adaptor AP-2, which normally functions to promote the assembly of clathrin lattices and targets receptors to clathrin-coated pits. The domain structure of AP-2 consists of four subunits (α , β 2, μ 2, and σ 2) where the β 2-subunit (β 2-adaptin) was shown to interact with an α -helical domain ($[DE]_nX_{1-2}FXX[FL]XXXR$) in nonvisual arrestins. Interestingly, the potential transition of the C-terminal β 2-adaptin binding region in arrestin from a β -sheet in the holo-protein to an α -helix when associated with β 2-adaptin has been suggested to function as a molecular switch to stabilize the activated conformation of arrestin.

Since nonvisual arrestins bind to clathrin and AP-2 with high affinity *in vitro* but appear to be primarily localized in the cytoplasm in cells, this raises questions as to what regulates arrestin association with clathrin-coated pits *in vivo*. Previous studies have suggested that the phosphorylation state of arrestin2 regulates clathrin-mediated endocytosis of the β 2AR. Upon receptor stimulation, arrestin is recruited to the plasma membrane where it is dephosphorylated, which appears to be required for targeting the activated receptor-arrestin signaling complex to clathrin-coated pits. As discussed above, intramolecular contacts of the β 2-adaptin binding domain in arrestin may effectively attenuate arrestin/AP-2 association until arrestin is activated. Indeed, activation of arrestin3 by mutation of Arg-170 significantly enhances arrestin3 binding to clathrin and β 2-adaptin. Similarly, addition of a phosphorylated receptor peptide to arrestin3 enhances arrestin3 binding to clathrin. Thus, arrestin binding to a phosphorylated activated GPCR may be the primary switch to promote conformational changes that unmask clathrin and AP-2 binding sites.

Arrestins mediate GPCR regulation by interacting with a diverse array of intracellular proteins; however, like other endocytic machinery such as AP-2 and AP180, arrestins also bind phosphoinositides (Figure 3). Arrestin2 and 3 contain a high-affinity phosphoinositide-binding site located in the C-domain that, when mutated, fails to support β 2AR internalization and receptor recruitment to clathrin-coated pits suggesting that phosphoinositides are important in delivering the receptor-arrestin complex to sites of endocytosis. Interestingly, the major visual arrestin in *Drosophila*, Arr2, also contains a highly electropositive patch of residues located in its C-domain that is involved in phosphoinositide binding. Alteration of this site interfered with Arr2 trafficking in photoreceptor cells effecting light adaptation. While PIP₂ and PIP₃ are the proposed physiological ligands for arrestins at the plasma

membrane, inositol hexakisphosphate (IP₆) displays a higher binding affinity for nonvisual arrestins than either PIP₂ or PIP₃. IP₆ is abundant in cells with concentrations ranging between 15 and 100 μM and has been proposed to regulate receptor endocytosis and receptor signaling. Interestingly, the ability of IP₆ to bind to two distinct sites on nonvisual arrestins appears to mediate homo- and hetero-oligomerization of arrestin2 and 3. Mutation of either IP₆-binding site disrupts oligomerization, and subcellular localization studies show that arrestin2 oligomers are primarily cytoplasmic whereas arrestin2 monomers display increased nuclear localization. This suggests that IP₆ binding to arrestin may regulate arrestin localization and function as a negative regulator of arrestin interaction with plasma membrane and nuclear-signaling proteins.

Role of Arrestins as Signaling Scaffolds

Numerous studies suggest that arrestins have roles beyond regulating GPCR desensitization and trafficking. Initial insight into this was provided by the finding that arrestin2 binds directly to c-Src (Table 2). This interaction appears to be mediated via proline-rich regions in arrestin2 interacting with the SH3 domain in Src although the catalytic domain may also contribute to binding. Interestingly, arrestin2/Src interaction may be regulated by phosphorylation since mutation of Ser-412 in arrestin2, the primary site of ERK1/2 phosphorylation, to an aspartate to mimic phosphorylation disrupts binding to c-Src. Thus, the current model proposes that arrestin2 binding to receptor promotes arrestin dephosphorylation enabling c-Src to bind and be recruited to clathrin-coated pits as part of a receptor/arrestin/Src complex. This may regulate Src-mediated phosphorylation of components (e.g., dynamin) that may, in turn, contribute to the regulation of GPCR trafficking and/or signaling. Interestingly, c-Src also appears to play a role in regulating dissociation of a GPCR/arrestin/AP-2 complex although the mechanism for this has not been delineated. Additional evidence for arrestin interaction with nonreceptor tyrosine kinases was found in neutrophils where treatment with IL-8 (which activates CXCR1) promotes interaction of arrestin2 with the tyrosine kinases Hck and c-fgr. Interestingly, an arrestin2 mutant defective in tyrosine kinase interaction blocks IL-8-promoted granule release suggesting that formation of an arrestin/Hck complex regulates exocytosis in neutrophils.

Additional studies have suggested a role for arrestins as mitogen-activated protein (MAP) kinase scaffolds (Table 2). Initial work focused on the potential involvement of arrestin2 in GPCR-mediated activation of ERK1/2 and demonstrated formation of complexes that contain receptor, arrestin2, ERK1/2, and either raf-1 or Src. Interestingly, inhibition of complex formation by expression of dominant negative arrestin2 or a truncated receptor that does not interact with arrestin2 inhibited both receptor endocytosis and activation of ERK1/2. Studies have also demonstrated that arrestin3 can directly bind to

JNK3 and ASK1, a JNK kinase kinase, suggesting a role for arrestin3 as a MAP kinase scaffold. Receptor stimulation results in JNK3 activation and triggers the colocalization of arrestin3 and active JNK3 with intracellular vesicles. Arrestin3 has also been implicated in activation of p38 MAP kinase although direct interaction of these proteins has not been established. It is interesting to note that for some GPCRs, arrestin2 and 3 may have differential roles in regulating MAP kinase signaling. For example, arrestin2 desensitizes angiotensin AT_{1A} receptor activation of ERK1/2 while arrestin3 plays a stimulatory role in ERK1/2 activation. By contrast, arrestin2 appears to stimulate CXCR4-mediated activation of ERK1/2 while arrestin3 desensitizes calcium flux. Thus, arrestins serve as scaffolds to mediate GPCR activation of multiple MAP kinase complexes.

See also: **Molecular Biology:** DNA Helicases: Hexameric Enzyme Action; **Signaling:** Adrenergic Receptors; Angiotensin Receptors; Bradykinin Receptors; Calcitonin Gene-Related Peptide and Adrenomedullin Receptors; Chemokine Receptors; Chemotactic Peptide/Complement Receptors; Dopamine Receptors; Eicosanoid Receptors; G Protein Signaling Regulators; GABA_A Receptor; GABA_B Receptor; Glutamate Receptors, Metabotropic; Muscarinic Acetylcholine Receptors; Opioid Receptors; P2Y Receptors; Parathyroid Hormone/Parathyroid Hormone-Related Protein Receptor; Photoreceptors; Serotonin Receptor Signaling; Tachykinin/Substance P/Neurokinin-1 Receptors; Taste Receptors; Vasopressin/Oxytocin Receptor Family.

Further Reading

- Defea K (2008) β -Arrestins and heterotrimeric G proteins: Collaborators and competitors in signal transduction. *British Journal of Pharmacology* 153: S298–S309.
- DeWire SM, Ahn S, Lefkowitz RJ, and Shenoy SK (2007) Beta-arrestins and cell signaling. *Annual Review of Physiology* 69: 483–510.
- Dorn GW II (2009) GRK mythology: G-protein receptor kinases in cardiovascular disease. *Journal of Molecular Medicine* 87: 455–463.
- Gurevich VV and Gurevich EV (2006) The structural basis of arrestin-mediated regulation of G-protein-coupled receptors. *Pharmacology and Therapeutics* 110: 465–502.
- Kendall RT and Luttrell LM (2009) Diversity in arrestin function. *Cellular and Molecular Life Sciences* 66: 2953–2973.
- Kovacs JJ, Hara MR, Davenport CL, Kim J, and Lefkowitz RJ (2009) Arrestin development: Emerging roles for β -arrestins in developmental signaling pathways. *Developmental Cell* 17: 443–458.
- Krupnick JG and Benovic JL (1998) The role of receptor kinases and arrestins in G-protein-coupled receptor regulation. *Annual Review of Pharmacology and Toxicology* 38: 289–319.
- Lodowski DT, Pitcher JA, Capel WD, Lefkowitz RJ, and Tesmer JJ (2003) Keeping G proteins at bay: A complex between G-protein-coupled receptor kinase 2 and G $\beta\gamma$. *Science* 300: 1256–1262.
- Moore CAC, Milano SK, and Benovic JL (2007) Regulation of receptor trafficking by GRKs and arrestins. *Annual Review of Physiology* 69: 451–482.
- Premont RT and Gainetdinov RR (2007) Physiological roles of G protein-coupled receptor kinases and arrestins. *Annual Review of Physiology* 69: 511–534.
- Ribas C, Penela P, Murga C, Salcedo, et al. (2007) The G-protein-coupled receptor kinase (GRK) interactome: Role of GRKs in GPCR regulation and signaling. *Biochimica et Biophysica Acta* 1768: 913–922.

G Protein Signaling Regulators

J H Exton, Vanderbilt University, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

G protein A signal-transducing protein that binds guanosine triphosphate (GTP; active form) or guanosine diphosphate (GDP; inactive form). It occurs in monomeric and heterotrimeric forms. Heterotrimeric G proteins are comprised of α -, β -, and γ -subunits. The α -subunit binds the guanine nucleotides and interacts with effector enzymes or ion channels. Heterotrimeric G proteins are activated by receptors with seven-transmembrane-spanning domains termed G-protein-coupled receptors.

GAP A GTPase-activating protein that acts on monomeric and heterotrimeric G proteins to inactivate them. GAPs for heterotrimeric G proteins are RGSs and certain effectors.

Phosphorylation The modification of proteins by protein kinases and lipids by lipid kinases involving the addition of phosphate groups. This results in changes in the activity of the proteins or in the generation of new lipids with different functions.

Receptor A membrane protein that binds and is activated by hormones, neurotransmitters, growth hormones, and cytokines.

Regulators of G protein signaling (RGSs) are important controllers of the system by which many agonists (hormones and neurotransmitters) transmit their signals across the cell membrane to influence intracellular processes. This system involves heterotrimeric G proteins, which are activated when the agonists bind to their specific receptors, and effectors (enzymes and ion channels), which are the targets of the G proteins. Activation of the G proteins by the receptors involves the release of guanosine diphosphate (GDP) from their α -subunits and the subsequent binding of guanosine triphosphate (GTP), and the dissociation of the $\beta\gamma$ -subunits from the G proteins. The GTP-bound α -subunits and the free $\beta\gamma$ -subunits then act on their target effectors. Turning off the system requires the hydrolysis of GTP to GDP by the intrinsic GTPase of the α -subunits and this is a relatively slow process. RGSs accelerate the GTPase activity, thereby terminating the stimulation of effectors and allowing the G protein to revert to its GDP-liganded heterotrimeric state where it can activate again by agonist binding to receptor.

Discovery of Regulators of G Protein Signaling

RGSs were discovered as negative RGSs in *Saccharomyces cerevisiae* and *Caenorhabditis elegans*. In *S. cerevisiae*, a mutation termed SST2 caused the yeast to become supersensitive to mating pheromone. However, at the time the SST2 gene was identified, knowledge of signaling pathways was rudimentary and the connection to G proteins could not be made. Later, it became clear that signaling in yeast was similar to that in mammals, that is, the topology of pheromone receptors was similar to that of G-protein-coupled receptors and the yeast G_{α} -subunit shared considerable identity with mammalian $G_{i\alpha}$. Later, genetic evidence indicated a direct interaction between SST2 and G_{α} , and studies of mutations in G_{α} and SST2 led to the conclusion that SST2 acted as a GTPase-activating protein (GAP) for G_{α} .

Genetic analysis of *C. elegans* revealed a gene (*egl-10*) that encoded a protein that was similar in its C-terminal region to

SST2. Similar to what was found in *S. cerevisiae*, the function of EGL-10 (to increase the frequency of egg laying via serotonergic motor neurons) depended on the presence of a G_{α} protein (GOA-1) that was homologous to mammalian G_{α} .

Multiple homologs of SST2 and EGL-10 have been identified in higher eukaryotes, utilizing yeast two-hybrid screens, database searches for related sequences, polymerase chain reaction, and rescue of the SST2-deficient phenotype in yeast. A large family of RGS proteins (more than 30 members) that show activity to different G proteins has now been identified in mammals (Table 1). They differ significantly in size, but all have a diagnostic RGS core domain and exhibit GAP activity toward heterotrimeric G proteins.

Structure of RGS Proteins

RGS proteins have a conserved RGS domain of ~ 130 amino acids (Figure 1). This domain is capable of binding G_{α} subunits and stimulating GTP hydrolysis. In addition, certain RGSs have other domains in their N- and C-terminal regions. These may be PDZ, PTB, GGL, PH, DH, PX, and DEP domains, transmembrane spans, cysteine strings, and domains that bind the adenomatous polyposis coli (APC) protein, GSK3, β -catenin, PKA, GRK, and $G\beta\gamma$ (Figure 1, see legend for definitions). The presence of these domains illustrates that RGSs have more functions than GAP activity and are targeted to membranes by multiple mechanisms.

The RGS domain has been defined by X-ray crystallography as a complex between RGS4 and $G_{\alpha i1}$, and by nuclear magnetic resonance (NMR) analysis of a fragment of GAIP. The domain is globular and mostly helical in structure, based on a four-helix bundle and a second domain composed of the N- and C-terminal helices. Many of the most conserved residues face into the helical bundles, and others are involved in G_{α} binding. Three interhelix loops make contact with the three switch regions of G_{α} (those regions that show the greatest conformational changes when GDP is replaced by GTP), but there is no contact with bound GTP.

Table 1 Mammalian RGS proteins

Protein	G α interactions	Tissue distribution
hRGS1	G α_{zi} family, G α_{zq}	Activated B cells
hRGS2	G α_{zq} > G α_{zi1}	Ubiquitous
hRGS3	G α_{zi} family, G α_{zq}	Ubiquitous
rRGS4	G α_{zi} family, G α_{zq}	Brain
mRGS5	G α_{zi} family	Heart, lung, brain, muscle
hRGS6	G α_{zo}	Brain
hRGS7	G α_{zi} family, G α_{zq}	Brain, retina
rRGS8	G α_{zi} family	Brain
hRGS9	G α_{zi} family, G $\alpha_{t\alpha}$	Brain, retina
hRGS10	G α_{zi} family	Brain
hRGS11	G α_{zo}	Brain, retina, pancreas
rRGS12	G α_{zi} family, G α_{z12} , G α_{z13}	Brain, lung, liver, heart, spleen
hRGS13	N.D.	N.D.
rRGS14	G α_{zi} family, G α_{z12} , G α_{z13}	Brain, lung, spleen
hRGS15	G α_{z12} , G α_{z13}	N.D.
mRGS16	G α_{zi} family, G α_{zq}	Retinal, pituitary, liver
hRGS-GAIP	G α_{zi} family, G α_{zq}	Heart, lung, liver
bRET-RGS1	G α_{zt}	Retina
hRGSZ1	G α_{zz}	Brain
hRGSZ2	G α_{zz}	N.D.
h115RhoGEF	G α_{z13} , G α_{z12}	Ubiquitous
mLscRhoGEF	N.D.	N.D.
hRhoGEF	N.D.	N.D.
mD-AKAP2	N.D.	Ubiquitous
rAxin	N.D.	Ubiquitous
mConductin	N.D.	Brain, lung, liver
bGRK2	G α_{zq}	Brain

N.D., no data.

Mechanisms of RGS Proteins

RGS proteins interact with active (GTP-bound) G α -subunits and accelerate the hydrolysis of GTP. In the absence of RGSs and certain effectors, the rate of hydrolysis of GTP by G α -subunits is very slow and inappropriate for rapid on/off signaling. RGSs can increase the rate of GTP hydrolysis by up to 2000-fold. RGSs act by altering the conformation of the G α -GTP complex, thus causing G α to be a more efficient GTPase. The transition state intermediate involved in the hydrolysis of GTP can be mimicked by the complex G α -GDP-A1F4, and this has helped elucidate the mechanism of action of RGSs through crystallographic and other studies. As noted above, RGS proteins have a broad interface with G α -subunits that includes the three switch regions. In the case of RGS4, the major change exerted on G α_{zi1} is decreased mobility of switch II, and it has been suggested that switch II is stabilized in a catalytically active conformation. Due to the conservation of key residues in the contact site of the RGS domain, this mechanism is probably general. Other residues that have been shown by mutation to affect the GAP activity of the RGSs are probably involved in stabilizing the catalytic conformation of the G α -subunit or altering the affinity of RGS for G α . Although the usual mode of action of RGSs is to interact with G α -subunits, some of them can also interact directly with effectors. For example, RGS2 can inhibit type V adenylyl cyclase.

From their structure, RGS proteins are predicted to be soluble. However, their mechanism requires that they act at the plasma membrane. The generation of GTP-liganded G α -subunits in response to receptor activation would be expected to recruit RGSs to the membrane. However, transmembrane segments, lipid modifications, for example, palmitoylation, and domains that interact with membrane lipids may also be involved, and there is evidence that some RGSs can be recruited selectively to the plasma membrane by receptors.

Selectivity of RGS Action

In view of the large number of RGSs, some selectivity in their interactions with G α -subunits would be expected (Table 1). Since certain receptors can couple to more than one G protein, the selective modulation of one G α by an RGS protein can change the output from such receptors. However, selectivity has been hard to define *in vitro* and may involve restrictive patterns of cellular expression of the RGS and G α proteins. *In vitro* analysis of RGS selectivity has involved single-turnover GAP assays, studies of competition with substrate, and co-immunoprecipitation and co-adsorption assays. For example, GAIP displays activity toward G α_i and G α_q class α -subunits, but the closely related RGSZ1 and RGSZ2 are selective for G α_{zz} and display low activity toward G α_{zq} . RGS6 and RGS11 are selective for G α_{zo} compared with other G α_{zi} class subunits, and RGS2 prefers G α_{zq} . Surprisingly, only one RGS (sorting nexin 13) has been reported to act on G α_{zs} .

In a cellular setting, the selectivity of RGSs for certain G α -subunits is less readily defined than *in vitro* since the readouts are quite distal from the interactions. RGS2 displays selectivity for G α_{zq} in accordance with *in vitro* studies. RGS16 inhibits the activation of p38 MAP kinase by PAF more effectively than the activation of ERK2, whereas RGS1 has the opposite effects. In dorsal root ganglion cells, α 2-adrenergic activation of G α_{zi} and G α_{zo} was enhanced by antibodies against RGS4 and GAIP, whereas antibodies to other RGSs had no effect. Microinjection of the RGS proteins confirmed these results.

Concerning the receptor selectivity of RGS proteins, there have been few studies. In pancreatic acinar cells, RGS4 blocked G α_q -mediated PLC activation initiated by M $_3$ muscarinic, cholecystokinin, and bombesin receptors. However, it was more potent against the muscarinic receptors. RGS1 and RGS16 were also more selective against M $_3$ muscarinic receptors. Surprisingly, RGS2, which has activity toward G α_{zq} , did not show the same selectivity. The molecular basis for this receptor selectivity is presently unknown.

Significance of RGS Action

The primary action of RGSs is to induce rapid signal termination upon removal or inactivation of agonists in G-protein-coupled receptor systems. The knockout of RSG9 prolongs photon responses because G α_{zt} remains active. In addition, this approach has been used to show that RGS3 and RGS5 selectively suppress MAP kinase activation by M $_3$ muscarinic

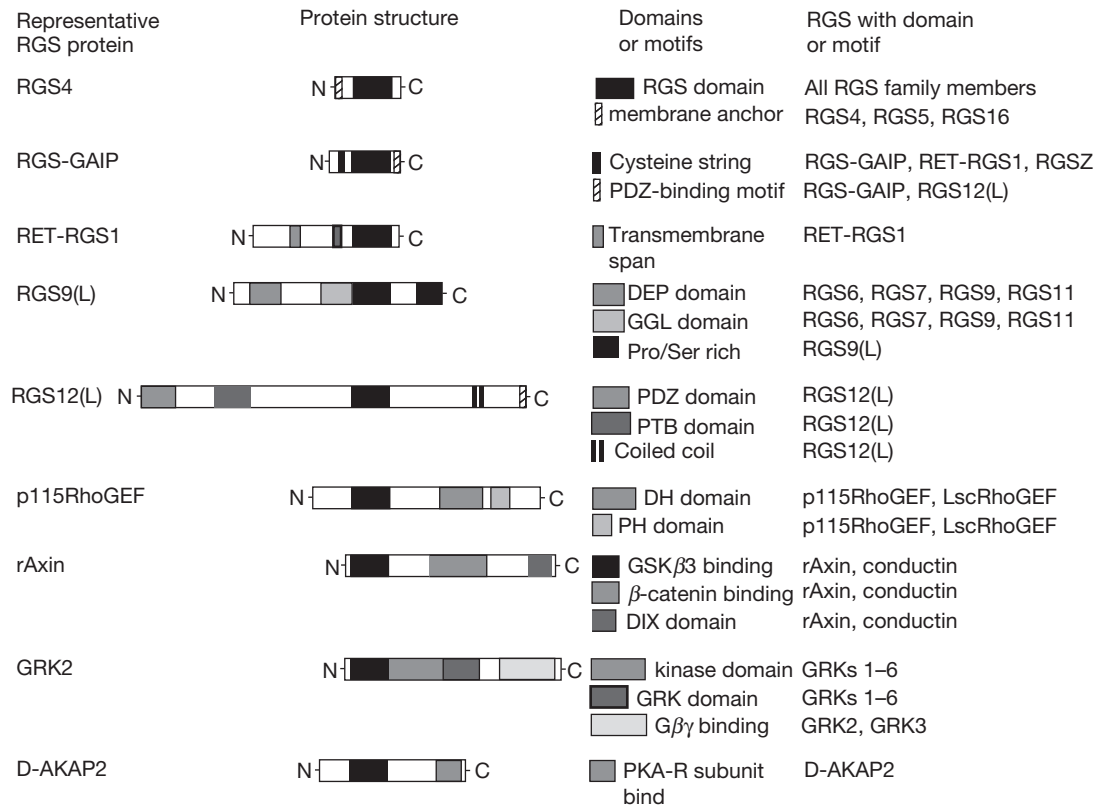


Figure 1 Structural organization and motifs in RGS proteins. PDZ, PSD-95, disk-large, and zo-1; DEP, disheveled, egl-10, and pleckstrin; GGL, G γ subunit-like; PTB, phosphotyrosine binding; DH, Dbl homology; PH, pleckstrin homology; GSK β 3, glycogen synthase β 3-binding; DIX, disheveled homology; GRK, G-protein-coupled receptor kinase; PKA-R, regulatory subunit of cAMP-dependent protein kinase. Modified from Hepler JR (1999) Emerging roles for RGS proteins in cell signalling. *Trends in Physiological Science* 20: 376–382.

and ATIA–angiotensin receptors, respectively, in vascular smooth muscle. On the other hand, transgenic mice overexpressing RGS4 in ventricular tissue develop less ventricular hypertrophy in response to pressure overload, apparently because of reduced $G_{\alpha q}$ action. It should also be noted that RGSs can inhibit effectors by mechanisms not involving their GAP activity. This may be because the RGSs compete with the effectors for binding the G_{α} -subunits or because the RGSs inhibit the effectors directly.

Although signal termination is an important component of RGS action, this is not the whole story. There are interesting kinetic aspects to RGS action, which arise from the cycle of activation/deactivation that G-protein-coupled receptor systems undergo. For example, the activating receptors can only interact with the GDP-liganded form of G_{α} . Thus, RGS activity can promote the association of G_{α} with the receptor through the generation of G_{α} -GDP, and permit G protein activation to continue.

Through their actions, RGS proteins modulate a variety of hormone- and neurotransmitter-stimulated responses in cells. These include the activity of adenylyl cyclase, the mitogen-activated protein kinase (MAP kinase) family, phospholipase C (with attendant changes in inositol trisphosphate and Ca^{2+} ions), G protein-gated inward rectifying K $^{+}$ (GIRK) channels, and cGMP phosphodiesterase in the retina (with altered phototransduction). Studies of the GIRK channels have

revealed that RGS proteins can alter the timing, amplitude, and duration of the K $^{+}$ currents. An interesting group of RGSs is the RhoGEFs. These are guanine-nucleotide-exchange factors for the small G protein Rho. The first of these was discovered in *Drosophila*, where genetic analysis indicated that a RhoGEF (DRhoGEF2) acted downstream of the concertina gene, which is the *Drosophila* homolog of $G_{\alpha 12}$. Several laboratories then showed that $G_{\alpha 12}$ acted directly on three mammalian homologs of DRhoGEF2 (p115 RhoGEF, PDZ-RhoGEF, and LARG). In particular, the N-terminal region of p115 was observed to be similar to the conserved region of RGSs, and p115 was found to associate with $G_{\alpha 12}$ and $G_{\alpha 13}$ through its RGS domain. Even more interesting were the findings that $G_{\alpha 13}$ could stimulate the GEF activity of p115 toward Rho, and that the RGS domain of p115 displayed GAP activity toward $G_{\alpha 12}$ and $G_{\alpha 13}$. There is now much evidence that p115 and related RhoGEFs mediate the activation of Rho induced by agonists whose receptors are coupled to G13. RGS16 inhibits $G_{\alpha 13}$ -mediated activation of Rho by binding directly to $G_{\alpha 13}$ and disrupting its interaction with p115. The binding does not involve the RGS domain of RGS16.

Although they do not contain the RGS domain, many effectors of G proteins act as GAPs toward G_{α} -subunits. Examples are phospholipase C, which acts on $G_{\alpha q}$, and the γ -subunit of cGMP phosphodiesterase, which acts on transducin. In the case of transducin, RGS proteins and the phosphodiesterase act synergistically to inactivate the G protein.

Regulation of RGS Proteins

RGS proteins can be regulated by various mechanisms. One is by transcriptional regulation of their cellular levels. Other mechanisms include covalent modification, cellular relocation, and interaction with regulatory proteins and lipids. The yeast RGS SST2 can be regulated at the transcriptional level by mating factor. Almost all RGSs are expressed in brain, although they are restricted to certain regions. They can be regulated by neurotransmitters through changes in their messenger RNA (mRNA) levels. In other tissues, there is transcriptional regulation involving growth factors and other peptides.

Regulation of RGSs also occurs at the posttranscriptional level. For example, phosphorylation can alter their degradation through ubiquitin-dependent proteolysis or by affecting a PEST sequence(s). For example, tumor necrosis factor- α (TNF α) can protect RGS7 from proteolysis through phosphorylation of a Ser/Thr sequence near the GGL and RGS domains. RGS activity can also be modified by phosphorylation of its G_x targets. Phosphorylation of $G_{\alpha z}$ by protein kinase C inhibits the GAP activity of RGS5. Another modification is reversible palmitoylation of either the RGS or its substrate G_x . The effects on the RGS may be positive or negative, depending on the assay used, but palmitoylation of G_x -subunits usually results in inhibition of the GAP activity of the RGS.

RGS proteins can bind to a variety of cellular proteins (Figure 1). As noted in the section on structure, they contain many different protein-binding motifs. They can also bind lipids, for example, PIP₃, Ca²⁺-calmodulin, and the APC protein. RGS12 is the only RGS that contains a PDZ domain and thus can interact with proteins that contain a PDZ-binding motif, for example, IL-8. AKAP2 binds the regulatory subunit of cAMP-dependent protein kinase (PKA). RGS6, RGS7, RGS9, and RGS11 contain a DEP domain and this is involved in membrane localization. Axin and conductin, which are scaffolding proteins, block signaling by Wnt proteins, which are involved in development. The RGS domains of these proteins bind the APC protein, whereas other domains bind β -catenin and GSK3.

Important interacting partners of the RGSs are the $\beta\gamma$ -subunits of G proteins ($G_{\beta\gamma}$). These bind to G_x cooperatively with respect to GDP and suppress receptor-independent activation of G proteins. They also anchor G_x to the plasma membrane and regulate many effector proteins, for example, adenylyl cyclase, phospholipase C, ion channels, and lipid kinases. With respect to RGS proteins, $G_{\beta\gamma}$ inhibits the GAP activity of many of these proteins. It appears that $G_{\beta\gamma}$ interacts with both the RGS and its target G_x -subunits. Several RGSs contain a GGL domain that is similar in sequence to $G_{\beta\gamma}$. It has also been shown that RGS11 binds $G_{\beta 5}$, and that the binding involves the GGL domain. Binding of $G_{\beta 5}$ to other RGSs has been demonstrated *in vivo*. In addition, RGSs can

inhibit the effects of $G_{\beta\gamma}$ -subunits on effectors by virtue of the generation of G_x -GDP which recombines with $G_{\beta\gamma}$ to reform inactive heterotrimeric G proteins. In the case of RGS3, there is another mechanism, namely its binding to $G_{\beta 1\gamma 2}$ -subunits to inhibit signaling to effectors. This effect involves two regions of RGS3 separate from the RGS domain.

RGS Proteins as Drug Targets

Due to the importance of G-protein-coupled receptors in regulating a large variety of physiological processes, the RGSs represent very attractive targets for therapeutic intervention. Alterations in their expression indicate that they regulate cardiac function, immune responses, neuronal function, behavior, vision, and embryonic development. There is also evidence that they may be involved in the diseases of retinitis pigmentosa, schizophrenia, Parkinson's disease, chronic heart failure, drug addiction, and prostate cancer. Drugs could be targeted to RGSs in several ways, and could act as agonists or inhibitors. The first approach is to target the RGS/ G_x interface, which involves certain amino acids that are essential for the contact. The second is through modulation of the proteins and lipids that act as allosteric modifiers of RGS action. This also includes covalent modifications. The third is modification of the membrane attachment of the RGSs since it is essential for them to be recruited to the plasma membrane in order to exert their effects. In all cases, determination of the amino-acid residues involved in RGS selectivity and in their interactions with allosteric regulators will be essential for the development of drugs of sufficient specificity to be therapeutically useful.

See also: **Signaling: Adenylyl Cyclases; G_{12}/G_{13} Family; Gi Family of Heterotrimeric G Proteins; Mitogen-Activated Protein Kinase Family; Phospholipase C.**

Further Reading

- Berman DM and Gilman AG (1998) Mammalian RGS proteins: Barbarians at the gate. *Journal of Biological Chemistry* 273: 1269–1272.
- Dohlman HG and Thorner J (1997) RGS proteins and signaling by heterotrimeric G proteins. *Journal of Biological Chemistry* 272: 3871–3874.
- Fukuhara S, Chikumi H, and Gutkind JS (2001) RGS-containing RhoGEFs: The missing link between transforming G proteins and Rho? *Oncogene* 20: 1661–1668.
- Hepler JR (1999) Emerging roles for RGS proteins in cell signalling. *Trends in Physiology and Science* 20: 376–382.
- Hollinger S and Hepler JR (2002) Cellular regulation of RGS proteins: Modulation and integrators of G protein signaling. *Pharmacological Reviews* 54: 527–559.
- Ross EM and Wilkie TM (2000) GTPase-activating proteins for heterotrimeric G proteins: Regulators of G protein signaling (RGS) and RGS-like proteins. *Annual Reviews in Biochemistry* 69: 795–827.

G_q Family

W Yang and J D Hildebrandt, Medical University of South Carolina, Charleston, SC, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Ataxia An inability to coordinate muscle activity during voluntary movement. It is often caused by disorders of the cerebellum.

Cardiac hypertrophy Enlargement of cardiomyocytes and thickening of the walls of the heart in response to hemodynamic stress and myocardial injury.

Primary hemostasis The formation of a primary platelet plug and involving platelets, the blood vessel wall, and von Willebrand factor. The process includes contraction of injured vessel, platelet adhesion, activation, aggregation, and secretion. Platelet secretion triggers the coagulation process and proceeds in a positive feedback loop to further stop bleeding.

Guanine-nucleotide-binding proteins (G proteins) are heterotrimeric proteins composed of α -, β -, and γ -subunits. Activated by G-protein-coupled receptors (GPCRs) in the plasma membrane, G proteins play pivotal roles in signal transduction from hormones, neurotransmitters, and drugs to cellular responses. The human genome has genes that code for 16 α -, 5 β -, and 12 γ -subunits. Based on functional and structural similarities, G proteins are grouped into four families named by the α -subunits they contain: G_s, G_i/G_o, G_q, and G₁₂/G₁₃. Members of the G_q family of G proteins are best characterized for their direct activation of phospholipase C β (PLC β) isozymes to produce the second messengers inositol 1,4,5-trisphosphate (IP₃) and diacylglycerol (DAG). Unlike G_s or G_i/G_o families, the G_q family proteins are insensitive to cholera toxin and pertussis toxin. G_q is activated by *Pasteurella multocida* toxin. This potentially provides a tool for studying G_q signaling, especially since the toxin does not activate G₁₁. Its use is complicated by the fact that it also activates G_{12/13}, and possibly G_i. Its mechanism of activation is not clear, but may require proteolytic activity of the toxin, and its effects on G₁₄ and G₁₅ are not known.

G_q Family of G Proteins

Phosphatidylinositol metabolites are primary intracellular regulators of eukaryotic cells. Studies of hormone and neurotransmitter regulation of phosphatidylinositol turnover implicated the frequent involvement of a specific class of G proteins. The subsequent purification of pertussis toxin-insensitive G_z subunits led to the demonstration of guanine nucleotide regulation of PLC β isoforms through G_q. The G_q family of proteins includes G_{zq}, G_{z11}, G_{z14}, and G_{z15}, which are expressed from individual genes with different expression patterns. The human genes are designated GNAQ, GNA11, GNA14, and GNA15, respectively. G_{zq} and G_{z11} are the most similar (89% identical amino acids), whereas G_{z15} shares only about 56% identities with other family members (Figure 1). The human ortholog of G_{z15} was originally called G_{z16} because it diverged so greatly from its mouse counterpart (only 84% identities at the amino acid level compared to 96–99% identities between the other human and mouse members of this family). The four genes encoding these proteins may have arisen from a pair of gene duplications, since G_{zq} and G_{z14} are within 70 kb of each other

on human chromosome 9 or mouse chromosome 19, and G_{z11} and G_{z15} are within 7 kb of each other on human and mouse chromosomes 19 and 10, respectively. All four genes have similar structures and contain seven coding exons. Proteins likely related to the G_q family are found in the genome of many species throughout the animal kingdom, but their relationship to G_z subunits in other kingdoms is unclear.

G_q Is Activated by Hormones, Neurotransmitters, and Drugs through Multiple GPCRs

The G_q family of G proteins is responsible for cellular responses to a plethora of hormones, neurotransmitters, and drugs, and is involved in the corresponding physiological, pathophysiological, and pharmacological processes (Table 1). In the CNS, M₁, M₃, and M₅ muscarinic receptors, group I metabotropic glutamate receptors (mGluR₁ and mGluR₅), and serotonin 5-HT₂ receptors, all couple to G_q family of G proteins and thus activate the PLC β signal transduction pathway. Activation of this pathway in the CNS is involved in synaptic transmission and modulation, neuronal development, and long-term depression (LTD) and long-term potentiation.

In the cardiovascular system, α_1 adrenergic receptors, endothelin receptors, angiotensin receptors, and prostaglandin F receptors modulate cardiovascular function through activation of G_q with prominent effects on blood pressure and cardiac function. G_q signaling may play a particularly important role in the pathophysiology of cardiac hypertrophy and congestive heart failure, a compensation process induced by hemodynamic stress and myocardial injury and subsequent decompensation after sustained stimulation. The fundamental molecular mechanisms mediating cellular responses to pressure or wall stress within the heart and vessels have been elusive, but may involve G_q-mediated generation of DAG for the direct activation of a subset of transient receptor potential channels (TRPCs), particularly TRPC6. Evidence supports the idea that G_q-coupled receptors, particularly for angiotensin II (AT1R), act as pressure sensors in cell membranes even in the absence of their activating ligand. Support for this idea comes from the observation that angiotensin II receptor blockers such as losartan that are inverse agonists block stress-induced activation of TRPC6, while receptor blockers that are neutral antagonists do not.

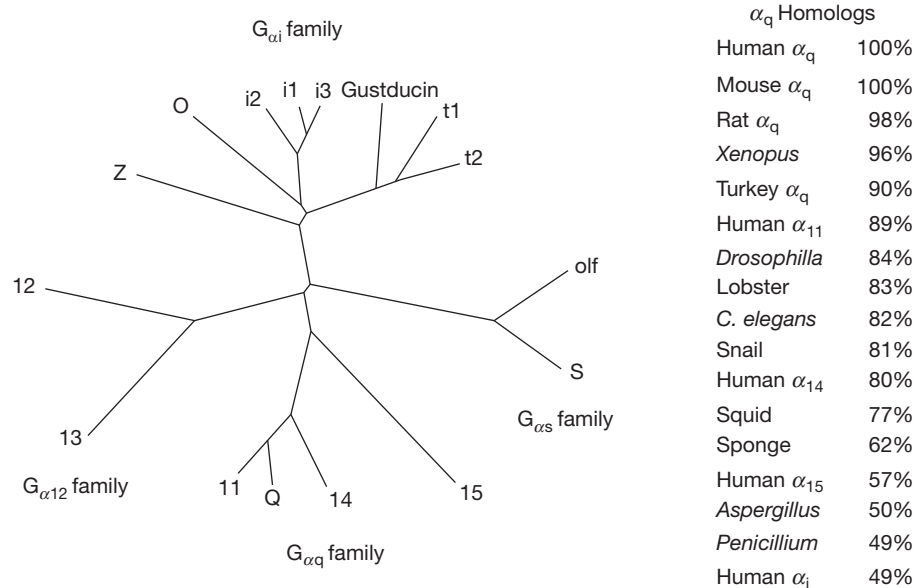


Figure 1 Homology of G_{αq} to other human G_α proteins and to related proteins in other species. The phylogenetic tree diagram was generated by TreeView from human G protein α-subunit protein sequences aligned with ClustalX. The length of the bars connecting different G_α proteins is proportional to the similarity of their aligned sequences. The α_q homologs listed are based upon percent identical amino acids in the aligned sequences.

Table 1 Receptors that couple to the G_q family of G proteins and their ligands and physiological functions

Ligands	Receptors	Function
Acetylcholine carbachol	M ₁ , M ₃ , and M ₅ muscarinic	Memory and cognition, smooth muscle contraction, and salivary secretion
Adenine nucleotides	P ₂ Y purinergic	Vasodilation
Angiotensin II, losartan	AT _{1a} and AT _{1b}	Modulation of cardiovascular activities
Bradykinin	B ₂ bradykinin	Hyperalgesia and smooth muscle contraction
C _{5a} anaphylatoxin	C _{5a}	Chemoattraction
Cholecystikinin (CCK)	CCK _a and CCK _b	Gastrointestinal activities
Endothelins	ET _a and ET _b	Regulation of cardiovascular activities
Glutamate	mGluR ₁ and mGluR ₅	Modulation of neuronal activities
Gonadotrophin-releasing hormone (GnRH)	GnRH	Synthesis and release of FSH and LH
Histamines	H ₁	Inflammation and allergy
5-Hydroxytryptamine, α-methyl-5-HT	5-HT _{2a} , 5-HT _{2b} , and 5-HT _{2c}	Modulation of neuronal activities of the brain
Neurotensin	Neurotensin	Modulation of gastrointestinal and brain activities
Norepinephrine (noradrenaline), epinephrine (adrenaline), phenylephrine, prazosin	α _{1a} , α _{1b} , α _{1c} , and α _{1d} adrenergic	Modulation of cardiovascular activities
Platelet activating factor	PAF	Platelet activation and aggregation
Prostaglandin F _{2α}	Prostaglandin F (FP)	Bronchoconstriction and luteolysis
Thromboxane A ₂	TXA ₂ (TP)	Smooth muscle contraction and platelet aggregation
Tachykinins	NK ₁ , NK ₂ , and NK ₃	Smooth muscle contraction and transmission of sensory information
Thrombin	Thrombin	Platelet aggregation
Thyrotrophin-releasing hormone (TRH)	TRH	Synthesis and release of thyroidstimulating hormone
Vasopressin	V _{1a} and V _{1b}	Smooth muscle contraction, platelet activation, glycogenolysis, and ACTH release
Leukotrienes	CysLT ₁ and CysLT ₂	Bronchoconstriction and vasodilation

Thromboxane A₂ receptors, thrombin receptors, and purinergic receptors induce platelet aggregation and granule secretion through activation of the G_q/PLCβ pathway. The G_q family of G proteins is also involved in hormone secretion, inflammation, and allergy through the activation of many other receptors including those for thyrotrophin-releasing hormone, gonadotrophin-releasing hormone, leukotrienes, bradykinin, and histamine.

PLCβ Activation and Calcium Mobilization Are Downstream Responses to G_q Activation

Phospholipases (PLs) are membrane-associated proteins involved in the biosynthesis and degradation of membrane lipids. They include PL A₁, A₂, C, and D (PLA₁, PLA₂, PLC, and PLD). PLC contains three subclasses on the basis of size and amino acid sequences: PLCβ, PLCγ, and PLCδ. PLCβ

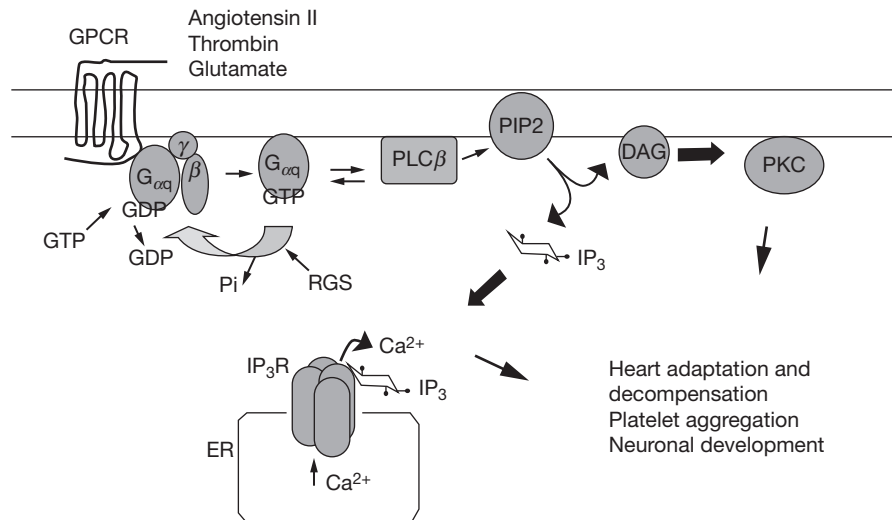


Figure 2 The G_q signal transduction pathway. Activation of many GPCRs leads to activation of G_q . $G_{\alpha q}$ activates $PLC\beta$ that subsequently cleaves PIP_2 to generate DAG and IP_3 . IP_3 binds to the IP_3 receptor (IP_3R) on the endoplasmic reticulum (ER) and induces release of calcium. DAG remains membrane bound and activates PKC. Both calcium and PKC are important signaling molecules that are involved in many cellular functions including cell growth and differentiation. $PLC\beta$ and regulators of G protein signaling (RGS) proteins can work as GAPs to facilitate the termination of G_q signaling by enhancing its GTPase activity.

isotypes ($\beta 1$ –4) are activated by GPCRs via G_q family and/or $\beta\gamma$ -subunits primarily from activation of the G_i/G_o family. Activated α -subunits of the G_q family stimulate all four isoforms of $PLC\beta$ with the potency rank order of $\beta 1 \geq \beta 3 \geq \beta 4 \gg \beta 2$ (Figure 2), but with the exception of $G_{\alpha 15}$, which preferentially stimulates $\beta 1$, $\beta 2$, and $\beta 3$. $PLC\beta$ hydrolyzes the membrane phospholipid phosphatidylinositol 4,5-bisphosphate (PIP_2) and generates two signal molecules: DAG and IP_3 . While DAG remains membrane associated, IP_3 is released into cytosol and binds to IP_3 receptors, which are ligand-gated Ca^{2+} channels, in the endoplasmic reticulum or sarcoplasmic reticulum (cardiac muscle cells). Binding of IP_3 induces calcium release from intracellular stores and causes a rise in cytosolic-free calcium concentration. One result of the increase in calcium concentration is its binding to calmodulin. Calcium–calmodulin complexes bind to other proteins and change their functional activities. Isozymes of protein kinase C (PKC) translocate from cytosol to plasma membrane and bind to DAG in the presence of calcium. Activated PKC phosphorylates other proteins and alters their functional state. PKC is known to have a wide variety of effects including receptor desensitization, activation of gene transcription, immune regulation, and regulation of cell growth.

The regulation of gene expression by cell-surface receptors often involves the stimulation of signaling pathways including one or more members of the mitogen-activated protein kinases (MAPKs). The MAPKs are a family of evolutionarily conserved serine/threonine kinases that transmit externally derived signals regulating cell growth, division, differentiation, and apoptosis. MAPKs are activated in response to stimulation of GPCRs that couple to the G_q family of G proteins such as α_{1b} adrenergic receptors or M_1 muscarinic cholinergic receptors. This process involves $PLC\beta$, intracellular Ca^{2+} , and PKC. PKC activation leads to MAPK activation through both Ras-dependent and Ras-independent mechanisms. PKC α can

apparently activate Raf-1 by direct phosphorylation. In many cases, however, $G_{q/11}$ activation causes MAPK activation that is unaffected by PKC inhibition.

G_q Uses Multiple Downstream Effector Mechanisms

In addition to its canonical stimulation of $PLC\beta$, G_q proteins appear to regulate other downstream effectors accounting, in part, for G_q actions not explained by effects on phosphatidylinositol metabolism. One important target appears to be p63RhoGEF, which is a guanine nucleotide exchange factor (GEF) for the activation of the small GTPase RhoA. Rho proteins regulate cytoskeletal-related processes such as secretion, cell migration, and smooth muscle contraction, as well as specific gene transcription events. A crystal structure of a $G_{\alpha i/q}$ chimera protein complexed with p63RhoGEF, and its RhoA target, confirms that G_q does not directly contact Rho but instead interacts with p63RhoGEF primarily through a region analogous to the effector-binding sites of other $G\alpha$ proteins in complex with their effector targets.

Another possible effector of G_q signaling is G protein-coupled receptor kinase 2 (GRK2). GRKs are protein kinases that participate in receptor desensitization, mediated in the case of GRK2 by its binding to $G\beta\gamma$. GRK2, and its related GRK3, also has a separate and distinct binding site for $G_{\alpha q}$ which, like p62RhoGEF, interacts with the effector binding site of the activated $G_{\alpha q}$ protein. This interaction may in part be a component of the desensitization process following G_q activation, but the GRK2- $G_{\alpha q}$ complex has also been shown to phosphorylate proteins such as insulin receptor substrate-1 (IRS-1) and a cytoskeletal regulator called ezrin. Thus, GRK2 and GRK3 have been proposed as possible downstream effectors for G_q , in addition to $PLC\beta$ and p63RhoGEF.

G_q in Cardiovascular Disease, Platelet Aggregation, and Cerebellum Development

The study of the functions of the G_q family of proteins has been greatly facilitated by the use of various animal models that overexpress the wild-type G_{αq} protein or that express a constitutively active form (gain of function mutations). Also useful have been animal models deficient in the protein, or those that overexpress an inhibitor (loss of function mutations). Some well-characterized functions of the G_q proteins include roles in the development of cerebellum and motor coordination, platelet aggregation, blood pressure responses, heart development and cardiac adaptation, and development of heart failure.

G_{αq} is essential for the signaling processes used by different platelet activators. It has been reported that a defect in human platelet G_q function leads to impaired platelet aggregation, secretion, and calcium mobilization, in response to a number of agonists. Platelets from G_{αq}-deficient mice are unresponsive to a variety of physiologic platelet activators. As a result, these mice have increased bleeding times, which is an *in vivo* measure of platelet function.

Mice with a null mutation in the G_{αq} gene suffer from ataxia with typical signs of motor coordination defects. In the cerebellum of newborn rodents, each Purkinje cell (PC) is innervated by multiple climbing fibers (CFs). Massive elimination of supernumerary CF-PC synapses occurs during the first three postnatal weeks, and by about postnatal day 20, most PCs are innervated by a single CF. In G_{αq}-deficient mice, 40% of adult PCs remain innervated by multiple CF because of a defect in regression of supernumerary CFs. PLCβ₄, which shows predominant expression in cerebellar PCs, and mGluR₁ may be involved in this function of G_{αq} since in mGluR₁- or PLCβ₄-deficient mice, a similar phenotype was observed. LTD of the PF-PC synapse in mice lacking G_{αq} was also deficient. Cerebellar LTD is a form of synaptic plasticity believed to be related to motor learning.

The G_q family of G proteins has been implicated in signaling pathways regulating cardiac growth under physiological and pathological conditions. The ability of cardiomyocytes to undergo hypertrophic growth is an important adaptive response to a wide range of conditions that require the heart to work more effectively. Norepinephrine (noradrenaline) (NE) and other hormones or regulators such as endothelin, prostaglandin F_{2α}, and angiotensin II induce cardiomyocyte growth through activation of G_q-coupled receptors. A relatively modest degree of wild-type G_{αq} overexpression in transgenic mice produced many of the features of compensated left ventricular hypertrophy. Conversely, expression of a G_{αq} inhibitor peptide that prevents receptor coupling to G_{αq} in mouse myocardium, or RGS4, which binds to and activates the GTPase activity of G_{αq}, significantly attenuated hypertrophy induced by pressure overload. The ability of cardiomyocytes to function at high capacity under increased workload usually cannot be sustained and ventricular failure develops after prolonged stimulation. One event associated with this transition to heart failure is the apoptotic death of cardiomyocytes. It has been shown that sustained or excessive activation of the G_q signaling pathway results in apoptotic loss of cardiomyocytes. Expression of a constitutively activated mutant of G_{αq}, which

further increased G_q signaling, produces initial hypertrophy, which rapidly progresses to apoptotic cardiomyocyte death. Conversely, inhibition of specific components of the pathway prevents or ameliorates heart failure. Compatible with these experimental findings, clinical studies have associated polymorphisms of the GNAQ promoter region thought to increase G_{αq} expression with outcomes or incidence of heart failure.

Modulation of G_q Functions

The G_q family of G proteins undergoes palmitoylation at N-terminal cysteine residues. Mutation of cysteines 9 and 10 in G_{αq} to serine profoundly alters the behavior of the subunit. Mutant G_{αq} cannot couple a coexpressed receptor to stimulation of PLCβ. Cysteine substitution also prevented a constitutively active form of G_{αq} from stimulating PLC directly. However, a substitution of a single cysteine in G_{αq} did not alter its activity.

Long-term activation of GPCRs coupled to G_q induces downregulation of the G_{αq} subunit. For example, prolonged exposure of αT3-1 pituitary cells to a gonadotrophin-releasing hormone receptor agonist results in marked downregulation of G_{αq} and G_{α11}. Chinese hamster ovary cells stably transfected with human M₃ muscarinic acetylcholine receptors show a 40–50% reduction in G_{αq} and G_{α11} when stimulated with the cholinergic agonist carbachol. Usually, levels of mRNA encoding G_{αq} and G_{α11} were not greatly altered, suggesting a change in protein degradation.

The intrinsic GTPase activity of G_α-subunits acts as a timer to limit the duration of G protein activation. PLCβ works as a GAP (GTPase-activating protein) to increase the intrinsic GTPase activity of G_{αq}. The ability of PLCβ to act as a GAP for G_{αq} provides negative feedback, limiting PLC activity. In addition, a family of proteins known as the regulators of G protein-signaling (RGS) proteins has been identified as GAPs for G_α-subunits. RGS4 and GAIP are known to act as GAPs for the G_q family. Furthermore, these RGS proteins block activation of PLCβ by G_{αq}, which apparently results from occlusion of the effector-binding site on G_{αq}.

The portion of GRK2 that binds G_q is an RGS homology (RH) domain with the potential to act as a GAP for terminating G_q activation. However, the GAP activity of GRK2 is very low and the crystal structure of a GRK2-G_{αi/q} chimera indicates that the GRK2 RH domain binds the effector region of G_{αq} rather than the GAP region of the protein.

The most obvious physiologic role of proteins with GAP activity would be to limit the time of action of the G_q proteins. However, GAP activity is also thought to increase the selectivity of GPCRs that couple to G_q by filtering out low-efficacy responses.

G_q Family Members Have Both Unique and Functionally Complementary Roles

Mammals express four G_q class α-subunits of which two, G_{αq} and G_{α11}, are widely expressed and are primarily responsible for coupling receptors to PLCβ. In a few cases, as in platelets, or

certain brain regions, such as the cerebellum, G_{zq} is expressed alone or in great excess over G_{z11}. In contrast to these widely expressed isoforms, G_{z14} and G_{z15} are restricted to certain tissues. G_{z14} is predominantly expressed in spleen, lung, kidney, liver, pancreas, and testis, whereas G_{z15} is primarily found in hematopoietic lineages. In most tissues, G_{zq} and G_{z11} are functionally complementary to each other. Receptors activating G_q family members in mammalian systems do not discriminate between G_{zq} and G_{z11}. Similarly, there appears to be little difference between the abilities of these G proteins in regulating PLC β functions.

Functional redundancy of G_q family proteins is also demonstrated by knocking out individual genes that code for the proteins of this family. Only G_{zq}-deficient mice revealed obvious phenotypic defects, including cerebellar ataxia and deficiencies in primary hemostasis due to a defect in platelet activation. The observed phenotypes correlated with preferential or exclusive expression of G_{zq} in the affected cell types of the cerebellum and platelets. Despite the specific expression patterns of G_{z14} and G_{z15}, which may indicate tissue-specific functions for these proteins, no obvious phenotypic changes were observed in mice carrying inactivating mutations of the G_{z14} and G_{z15} genes. G_{z11}-deficient mice were also viable and fertile with no apparent behavioral or morphologic defects.

A gene dosage effect between G_{zq} and G_{z11} was observed after breeding G_{zq}-deficient mice with G_{z11}-deficient mice. Embryos lacking both genes died in uterus with heart malformations. Mice inheriting a single copy of either gene died within hours of birth with craniofacial and/or cardiac defects. Apparently, two active alleles of these genes are required for life after birth. Analyses of intercross offspring inheriting different combinations of these two mutations indicated that G_{zq} and G_{z11} have overlapping functions in embryonic cardiomyocyte proliferation and craniofacial development.

Although studies of genetically engineered mice lacking selective G_{zq} isoforms reveal specific phenotypes only for the canonical G_{zq} isoform, analysis of signaling properties of cells derived from these animals suggests unique roles for each of the different isoforms. Although the details and consequences of these differences have been hard to elucidate based upon our current understanding of G_q signaling, one study suggested that each of the different isoforms regulates transcription of both shared and unique populations of genes in vascular smooth muscle cells. Expression of constitutively active forms of G_{zq}, G_{z14}, or G_{z15} showed that 20–34% of the genes regulated by each isoform were unique to it and not shared by any of the others. Studies such as this suggest both complementary and unique roles for each G_q protein that will likely be elucidated as more information is accumulated about recently described properties and roles of the proteins.

See also: Bioenergetics; IP₃ Receptors; Protein/Enzyme Structure Function and Degradation: Protein Palmitoylation; Signaling: G Protein Signaling Regulators; G₁₂/G₁₃ Family; Gi Family of Heterotrimeric G Proteins; Gs Family of Heterotrimeric G Proteins; Phospholipase C; Protein Kinase C Family.

Further Reading

- Adams JW and Brown JH (2001) G proteins in growth and apoptosis: Lessons from the heart. *Oncogene* 20: 1626–1634.
- Cockcroft S and Gomperts BD (1985) Role of guanine nucleotide binding protein in the activation of polyphosphoinositide phosphodiesterase. *Nature* 314: 534–536.
- Exton J (1996) Regulation of phosphoinositide phospholipases by hormones, neurotransmitters, and other agonists linked to G proteins. *Annual Review of Pharmacology and Toxicology* 36: 481–509.
- Hermans E and Challiss RA (2001) Structural, signaling and regulatory properties of the group I metabotropic glutamate receptors: Prototypic family C G-protein-coupled receptors. *Biochemistry Journal* 359: 465–484.
- Hubbard KB and Hepler JR (2006) Cell signalling of the Gq α family of heterotrimeric G proteins. *Cellular Signal* 18: 135–150.
- Iyengar J and Hildebrandt JD (eds.) *G Protein Pathways, Part B, G Proteins and Their Regulators, Methods in Enzymology*, vol. 344. New York, NY: Academic Press.
- Litosch I, Wallis C, and Fain JN (1985) 5-Hydroxytryptamine stimulates inositol phosphate production in a cell-free system from blowfly salivary glands. Evidence for a role of GTP in coupling receptor activation to phosphoinositide breakdown. *Journal of Biological Chemistry* 260: 5464–5471.
- Lutz S, Shankaranarayanan A, Coco C, et al. (2007) Structure of Gq α -p63RhoGEF-RhoA complex reveals a pathway for the activation of RhoA by GPCRs. *Science* 318: 1923–1927.
- Mizuno N and Itoh H (2009) Functions and regulatory mechanisms of Gq-signaling pathways. *Neurosignals* 17: 42–54.
- Offermanns S (2001) *In vivo* functions of heterotrimeric G proteins: Studies in G α -deficient mice. *Oncogene* 20: 1635–1642.
- Pang IH and Sternweis PC (1990) Purification of unique alpha subunits of GTP-binding regulatory proteins (G proteins) by affinity chromatography with immobilized beta gamma subunits. *Journal of Biological Chemistry* 265: 18707–18712.
- Peavy RD, Hubbard KB, Lau A, et al. (2005) Differential effects of Gq α , G14 α , and G15 α , on vascular smooth muscle cell survival and gene expression profiles. *Molecular Pharmacology* 67: 2102–2114.
- Sharif-Naeini R, Folgering JHA, Bchet D, et al. (2010) Sensing pressure in the cardiovascular system: Gq-coupled mechanoreceptors and TRP channels. *Journal of Molecular and Cellular Cardiology* 48: 83–89.
- Simon MI, Strathmann MP, and Gautam N (1991) Diversity of G proteins in signal transduction. *Science* 252: 802–808.
- Smrcka AV, Hepler JR, Brown KO, and Sternweis PC (1991) Regulation of polyphosphoinositide-specific phospholipase C activity by purified G α . *Science* 251: 804–807.
- Sternweis PC and Smrcka AV (1993) G proteins in signal transduction: The regulation of phospholipase C. *Ciba Foundation Symposium* 176: 96–111.
- Strathmann M, Wilkie TM, and Simon MI (1989) Diversity of the G-protein family: Sequences from five additional alpha subunits in the mouse. *Proceedings of the National Academy of Sciences of the United States of America* 86: 7407–7409.
- Taylor SJ, Chae HZ, Rhee SG, and Exton JH (1991) Activation of the 1 isozyme of phospholipase C by subunits of the G_q class of G proteins. *Nature* 350: 516–518.
- Taylor SJ, Smith JA, and Exton JH (1990) Purification from bovine liver membranes of a guanine nucleotide-dependant activator of phosphoinositide-specific phospholipase C. Immunologic identification as a novel G-protein alpha subunit. *Journal of Biological Chemistry* 265: 17150–17156.
- Tesmer VM, Kawano T, Shanaranarayanan A, Kozasa T, and Tesmer JJG (2005) Snapshot of activated G proteins at the membrane: The G α q-GRK2-G β γ complex. *Science* 310: 1686–1690.
- Tikhonoff V and Casiglia E (2008) Evolving concepts of left ventricular hypertrophy. *European Heart Journal* 29: 846–848.
- Uhing RJ, Popic V, and Exton JH (1986) Hormone-stimulated polyphosphoinositide breakdown in rat liver plasma membranes. Roles of quanine nucleotides and calcium. *Journal of Biological Chemistry* 261: 2140–2146.

Relevant Website

<http://www.signaling-gateway.org> — Alliance for Cell Signaling/Nature, The Signaling Gateway.

Green Bacteria: Chlorophyll Biosynthesis, Light-Harvesting, Reaction Centers, and Electron Transport

D A Bryant, The Pennsylvania State University, University Park, PA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Chlorophototroph A bacterium that obtains its energy for growth by chlorophyll-based light energy capture and conversion.

Chlorosome The light-harvesting antenna complexes of green bacteria. They are envelope-enclosed sacs containing up to 250 000 bacteriochlorophyll *c*, *d*, or *e* molecules that self-organize into supramolecular structures.

Energy transfer The transfer of excited-state energy between two pigment molecules by nonradiative, dipole-induced dipole resonance transfer (also called Förster resonance energy transfer). In more strongly coupled pigment systems (e.g., chlorosomes), energy transfer is better described by exciton theory.

Filamentous anoxygenic phototrophs (FAPs) Formerly green nonsulfur bacteria, these bacteria are chlorophototrophic members of the phylum *Chloroflexi*. Most FAPs are photoheterotrophs or photomixotrophs, but some are able to oxidize reduced sulfur compounds or hydrogen and grow autotrophically by fixing carbon dioxide by the 3-hydroxypropionate pathway.

Fenna–Matthews–Olson protein (FMO protein) It was the first (B)Chl protein to have its structure determined by X-ray crystallography. It is a water-soluble, BChl *a*-binding protein found in members of the green sulfur bacteria and *Candidatus* Chloracidobacterium thermophilum.

Green bacteria Bacteria belonging to three phyla within the Domain *Bacteria*, which synthesize BChl *c*, *d*, or *e* and have chlorosomes as light-harvesting antennae.

Green sulfur bacteria Chlorophototrophs belonging to the order *Chlorobiales* within the phylum *Chlorobi*. These bacteria have type-1 reaction centers, chlorosomes, and FMO in their photosynthetic apparatus. These obligately photoautotrophic organisms oxidize reduced sulfur compounds, hydrogen, or ferrous iron and fix carbon dioxide by the reverse (reductive) tricarboxylic acid (TCA) cycle.

Reaction center A (B)Chl (and carotenoid) containing protein complex that performs light-activated electron transport. Absorption of a photon leads to oxidation of a special pair of (B)Chls and reduction of an electron acceptor. Secondary electron transfer events stabilize the primary charge separation by transferring the electron to additional acceptors, which are increasingly distant from the special pair.

Type-1 reaction center Reaction centers that terminate in a chain of three [4Fe–4S] acceptors and that produce a weak oxidant and a strong reductant. Type-1 reaction centers can be homodimeric (green sulfur bacteria, heliobacteria, and *Candidatus* Chloracidobacterium thermophilum) or heterodimeric (photosystem I of cyanobacteria).

Type-2 reaction center Reaction centers that terminate in a quinone acceptor, which acts as a two-electron acceptor. Type-2 reaction centers have heterodimeric cores and produce strong oxidants and weak reductants. Bacterial type-2 reaction centers are found in purple bacteria and FAPs. Photosystem II is a water-oxidizing, type-2 reaction center found in cyanobacteria, algae, and plants.

Green Bacteria: Introduction

Green bacteria are anoxygenic chlorophototrophic bacteria: they do not produce oxygen, they use light as their primary energy source for growth, and they use (bacterio)chlorophylls ((B)Chls) for light capture and conversion into chemical energy (as adenosine triphosphate (ATP) and reducing power). Green bacteria occur as members of three phyla, *Chlorobi*, *Chloroflexi*, and *Acidobacteria*, within the Domain *Bacteria*, and the members of these three taxa differ significantly in many physiological and biochemical properties. Green sulfur bacteria (GSB; *Chlorobiales*) are obligately anaerobic photolithoautotrophs, which oxidize reduced sulfur compounds, ferrous iron, or hydrogen and reduce carbon dioxide by the reverse tricarboxylic acid (TCA) cycle. Green nonsulfur bacteria are more commonly described today as filamentous anoxygenic phototrophs (FAPs; *Chloroflexales*). Although some can oxidize sulfide and/or hydrogen and fix carbon dioxide by the 3-hydroxypropionate cycle, these organisms are principally photoheterotrophs or photomixotrophs.

Candidatus Chloracidobacterium thermophilum was only discovered very recently, and this organism is currently the only described chlorophototrophic member of the phylum *Acidobacteria*. *Candidatus* C. thermophilum does not fix carbon dioxide and is an aerobic photoheterotroph, but it has a photosynthetic apparatus very similar to that of GSB.

Chlorophototrophs convert light energy into chemical energy by using specialized, multi-subunit, pigment–protein complexes known as photochemical reaction centers (RCs). Two major classes of RCs are known in nature: type-1 RCs have three [4Fe–4S] clusters as terminal electron acceptors and produce weak oxidants and strong reductants; type-2 RCs have quinones terminal electron acceptors and produce strong oxidants and weak reductants. All RCs utilize a dimer of (B)Chl molecules, the so-called special pair, to initiate light-dependent electron transfer events. All GSB and *Candidatus* C. thermophilum (as well as Heliobacteria (phylum *Firmicutes*), which are not considered here) have homodimeric type-1 RCs that bind about 20 BChl/Chl molecules (see below), while FAPs have

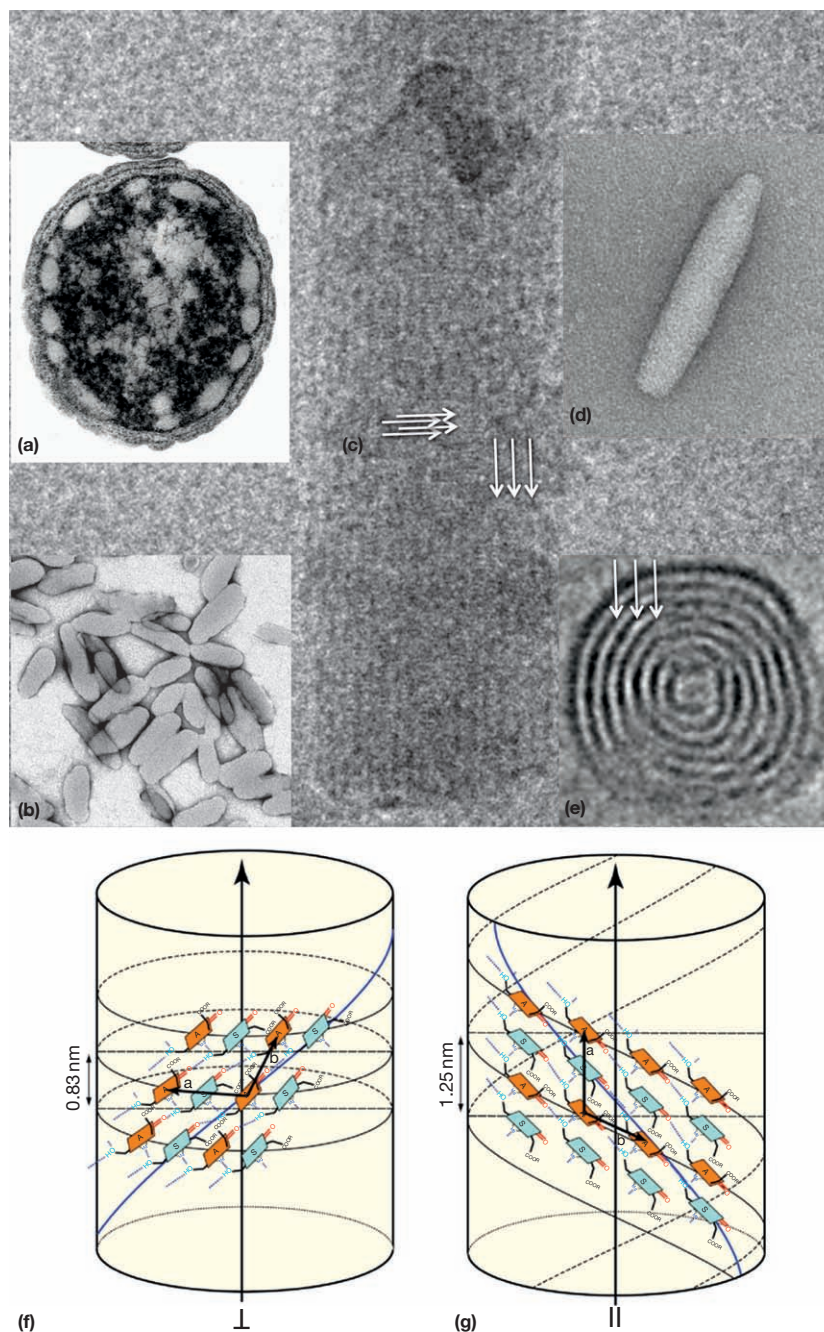


Figure 1 Structure of chlorosomes from the green sulfur bacterium *Chlorobaculum tepidum*. (a) Transmission electron micrograph showing chlorosomes, electron transparent ovoids appressed to the inner surface of the cytoplasmic membrane. (b) Negatively stained transmission electron micrograph of isolated chlorosomes from *C. tepidum*. (c) Cryo-electron micrograph showing the structure of BChl *d* molecules in chlorosomes of a *bchQ bchR bchU* mutant of *C. tepidum*. The three horizontal white arrows indicate the 2.1-nm spacings between the surfaces that form coaxial nanotubules. The four horizontal arrows indicate the 0.83-nm repeats that arise from the *syn-anti* monomer stacks of BChl *d* molecules. (d) Transmission electron micrograph of a single, negatively stained chlorosome from the *bchQ bchR bchU* mutant of *C. tepidum*. (e) Cryo-electron micrograph showing an end view of a single chlorosome from the *bchQ bchR bchU* mutant of *C. tepidum*. The three vertical white arrows show the 2.1-nm spacings between the surfaces formed from the *syn-anti* monomer stacks of BChl *d*. (f, g) Scheme showing the *syn-anti* monomer stacks of BChl *d* in the *bchQ bchR bchU* mutant of *C. tepidum* (f) and of BChl *c* in wild-type *C. tepidum* (g).

heterodimeric type-2 RCs that are similar to the type-2 RCs of purple bacteria; they bind six total pigment molecules: three (or four) BChl *a* molecules and three (or two) bacteriopheophytin (BPhe *a*) molecules.

Although some green bacteria are actually green in color, these organisms can also be brown, orange, or reddish-brown in color. Thus, the defining property of these organisms is not their color, but two related biochemical properties. All green

bacteria synthesize BChl *c*, *d*, or *e*, and these BChl molecules are used to produce chlorosomes, which are literally sacs containing self-organized BChls, as the principal light-harvesting antenna complexes. Chlorosomes are very large light-harvesting structures (up to 30–60 nm in diameter, 100–200 nm in length) (Figure 1), and each chlorosome can contain up to 250 000 BChl *c*, *d*, or *e* molecules (Figure 2). The very high cellular content of BChl allows some green bacteria to grow at extraordinarily low-light intensities – conditions under which no other type of chlorophototroph can grow. Chlorosomes are connected to the inner surface of the cytoplasmic membrane through a structure known as the ‘baseplate’, which is a paracrystalline array of the small protein, CsmA. CsmA binds a single BChl *a* molecule and one or two carotenoid molecules (Figure 3). In GSB and *Candidatus C. thermophilum*, the baseplate is in turn attached to another BChl *a*-binding protein, the Fenna–Matthews–Olson (FMO) protein (Figures 3 and 4). FMO is a water-soluble, peripheral membrane protein that serves as the point of interconnection between the chlorosomes and the type-1 RCs in the cytoplasmic membrane. In green members of the *Chloroflexales*, such as *Chloroflexus aurantiacus*, FMO does not occur; instead, a ring of membrane-intrinsic, heterodimeric BChl *a*-binding proteins surrounds the type-2 RCs. This complex, known as B808–B866, resembles the light-harvesting-I (LH-I) complexes of purple bacteria.

Bacteriochlorophyll Biosynthesis in Green Bacteria

All GSB and *Candidatus C. thermophilum* synthesize three types of Chls: BChl *a_P*, Chl *a_{PD}*, and either BChl *c*, *d*, or *e* (the subscripts indicate the esterifying alcohol: P, phytol; PD, Δ²,6-phytyadienol; F, farnesol; S, stearyl (octadecanol)). The structure of BChl *c_F*, which is actually a family of compounds with different numbers of methyl groups at the C-8² and C-12¹ positions, is shown in Figure 2. Chlorosome-producing

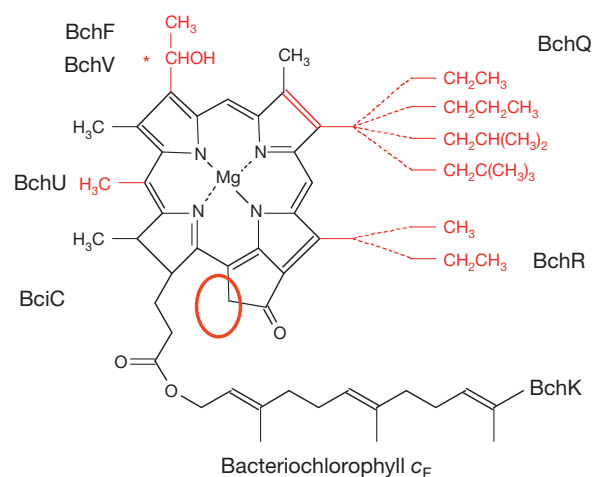


Figure 2 The structure of bacteriochlorophyll *c* esterified with farnesol. The protein designations at the periphery provide the names of the gene products responsible for the modifications that lead from chlorophyllide *a* to bacteriochlorophyll *c* (also see Figure 5). The asterisk indicates the C-3¹ chiral carbon. The circle indicates the absence of a methylcarboxyl group at C-13².

Chloroflexi, for example, *Chloroflexus aurantiacus*, synthesize BChl *a_P* and BChl *c* but do not synthesize Chl *a*. Figure 5 shows the biosynthetic pathway for the synthesis of these (B)Chls. The synthesis of all Chls proceeds through a series of core reactions that convert protoporphyrin IX (1; the numbers in parentheses in bold indicate the numbered compounds in Figure 5) into chlorophyllide *a* (6), the immediate precursor of Chl *a* (18) (Figure 5). The first committed step in (B)Chl biosynthesis is the ATP-dependent insertion of Mg into protoporphyrin IX to produce Mg-protoporphyrin (2), a reaction that is catalyzed by Mg-chelatase (BchHID/ChlHID). The subsequent reaction is the methylation of the C-13² carboxyl group by BchM/ChlM. The conversion of Mg-protoporphyrin IX monomethyl ester (3) into 3,8-divinyl-protochlorophyllide *a* (4) is catalyzed by two different enzymes, BchE in anaerobes and AcsF in aerobes. The former enzyme is an oxygen-sensitive member of the radical-S-adenosylmethionine (SAM) superfamily, while the latter is an oxygen-dependent di-iron oxidase. Both introduce the C-13¹ keto group and isocyclic ring while performing a six-electron oxidation of the substrate. At least three different enzymes, BciA, BciB, and a still unidentified enzyme, also reduce the C-8 vinyl group to produce protochlorophyllide *a* (8). BciA of GSB is very similar to the NADPH-dependent enzyme of plants, while BciB is similar to the enzyme found in cyanobacteria. BciB is found in many GSB, in some members of the *Chloroflexi*, and in *Candidatus C. thermophilum*. *Roseiflexus* spp., which synthesize BChl *a* and must reduce the C-8 vinyl group, lack homologs of both BciA and BciB and therefore must have a third type of 8-vinyl reductase. In all anoxygenic bacteria, the reduction of the C-17,18 double bond of protochlorophyllide *a* (5) to produce chlorophyllide *a* (6) is catalyzed by an enzyme related to nitrogenase. This enzyme is encoded by the *bchB*, *bchN*, and *bchL* genes and, like the reduction of dinitrogen, requires the hydrolysis of ATP. Chlorophyllide *a* (6), which can directly be converted into Chl *a* (18) by chlorophyll synthase (ChlG) and BchP (geranylgeranyl reductase), is the hub compound of (B)Chl biosynthesis. This intermediate can be converted into 11 of the 14 BChl and Chl derivatives currently known to occur in chlorophototrophic bacteria.

The first committed step in the synthesis of BChl *a* is the reduction of the C-7,8 double bond of the B-ring by chlorophyllide reductase to produce 3-vinyl-bacteriochlorophyllide *a* (15). Chlorophyllide reductase is closely related to protochlorophyllide reductase by gene duplication (and by extension to nitrogenase) and is encoded by the *bchX*, *bchY*, and *bchZ* genes. The remaining steps in BChl *a* synthesis are the hydration of the C-3¹ vinyl group by BchF and subsequent oxidation of the resulting hydroxyl group to a keto moiety by BchC. BchG adds the phytol side chain to the C-17 propionate group to produce BChl *a* (19).

Chlorophyllide *a* (6) is also the immediate precursor required for the synthesis of BChl *d* (12), BChl *c* (13), and BChl *e* (14). The first committed step in the synthesis of these molecules, the removal of the C-13² methylcarboxyl group to produce 3-vinyl-bacteriochlorophyllide *d* (7), requires BciC, but the mechanism of this reaction is not yet known. Two related radical-SAM methyl transferases, BchQ and BchR, add methyl groups from S-adenosyl-L-methionine to the C-8² and

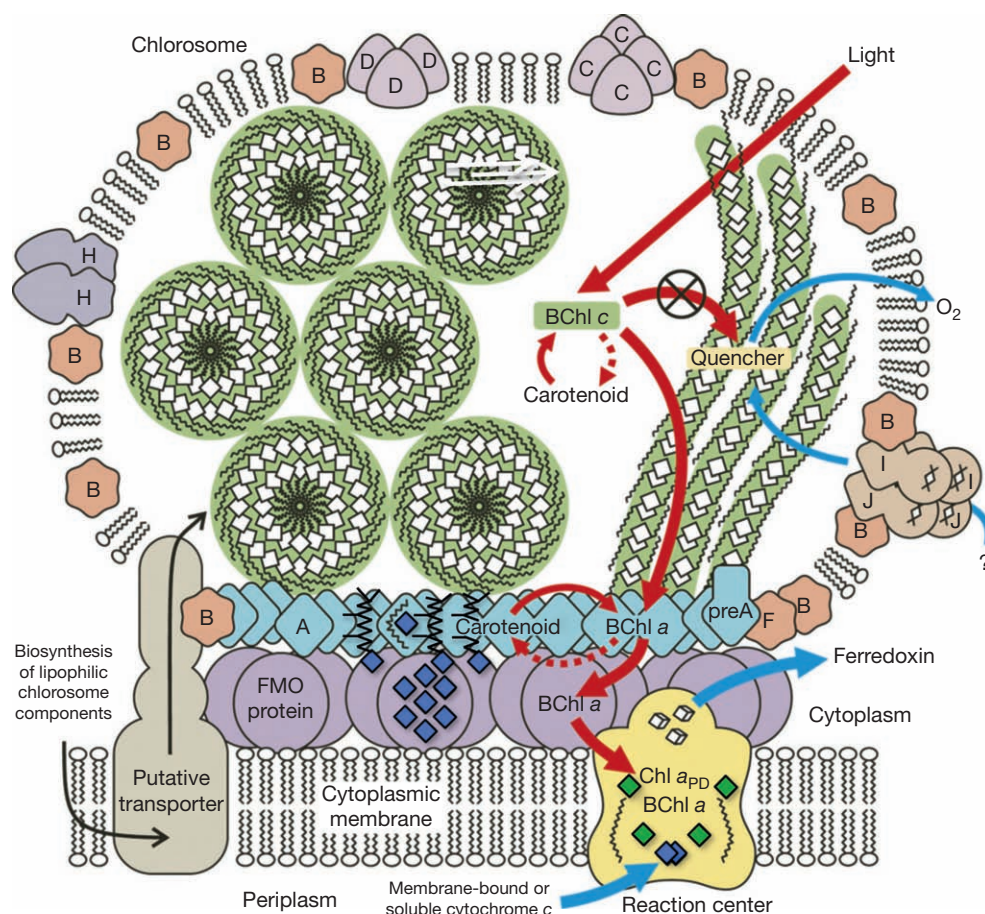


Figure 3 Scheme showing the structural organization of chlorosomes, FMO protein, and the type-1 reaction centers of *C. tepidum*. Energy transfer pathways are indicated by red arrows, and electron transfer reactions are indicated by blue arrows. The quencher is mostly chlorobiumquinone and menaquinone. Ten proteins occur in the chlorosome envelope, and the paracrystalline array of CsmA is shown in the baseplate. Blue diamonds indicate BChl a_P , and green diamonds in the reaction center indicate Chl a_{PD} molecules. In *C. tepidum*, the BChl c (white diamonds with green colored halos) form *syn-anti* monomer stacks that self-organize into coaxial nanotubes. In *Candidatus C. thermophilum*, the BChl c_s molecules appear to form folded or twisted sheets (right side of the scheme). For additional details, see text.

C-12¹ positions, to produce a family of compounds differing in their degree of methylation (see Figure 2). BchR can add a single methyl group at C-12¹, and in the GSB *Chlorobaculum tepidum*, ~95% of the BChl c molecules have the resulting ethyl group at this position. BchQ can add one, two, or three methyl groups at the C-8² position, producing *n*-propyl, isobutyl, or neopentyl side chains at this position. BchQ and BchR are not found in *C. aurantiacus*, and thus this FAP can only synthesize [8-Et, 12-Me]-BChl c . However, the BchQ and BchR methyltransferases do occur in other FAPs and are also found in *Candidatus C. thermophilum*. Depending on the degree of methylation at C-8² and C-12¹, hydration of the C-3 vinyl group, to produce a chiral center at C-3¹, is catalyzed by BchF or BchV. Highly methylated BChl c molecules have predominantly *S*-chirality at this position; [8-Et, 12-Me]-BChl c has almost exclusively *R*-chirality. BChl d (9) is methylated at C-20 by BchU, producing bacteriochlorophyllide c (10). The mechanism and gene products responsible for the oxidation of the C-7 methyl group to produce the formyl side chain of bacteriochlorophyllide e (11) are currently unknown. BChl c , d , and e are produced when BchK esterifies the corresponding

bacteriochlorophyllides with alcohols such as farnesol, geranylgeraniol, hexadecanol, and octadecanol (stearol).

Assembly of BChls in Chlorosomes

BChl c , d , and e assemble into suprastructures in chlorosomes (Figures 1(a), 1(b), 1(d), and 3), which can contain up to 250 000 BChl molecules, without the participation of proteins. Cells of the model GSB, *Chlorobaculum tepidum*, contain about 200 chlorosomes per cell (Figure 1(a)) and thus can contain up to 50 million BChl c molecules per cell, which is about 20 times greater than the number of protein molecules in such cells. *C. tepidum* chlorosomes are large and rather heterogeneous in size: length 100–200 nm and diameter 30–60 nm (Figures 1(b) and 1(d)). In addition to BChl c , the chlorosomes of *C. tepidum* contain about 2500 BChl a molecules, 5000 protein molecules (all are found in the envelope, and half of which are CsmA), 20 000 carotenoids, 18 000 chlorobiumquinone and menaquinone molecules, and 20 000 lipid molecules (glycolipids, phospholipids, and wax esters). The

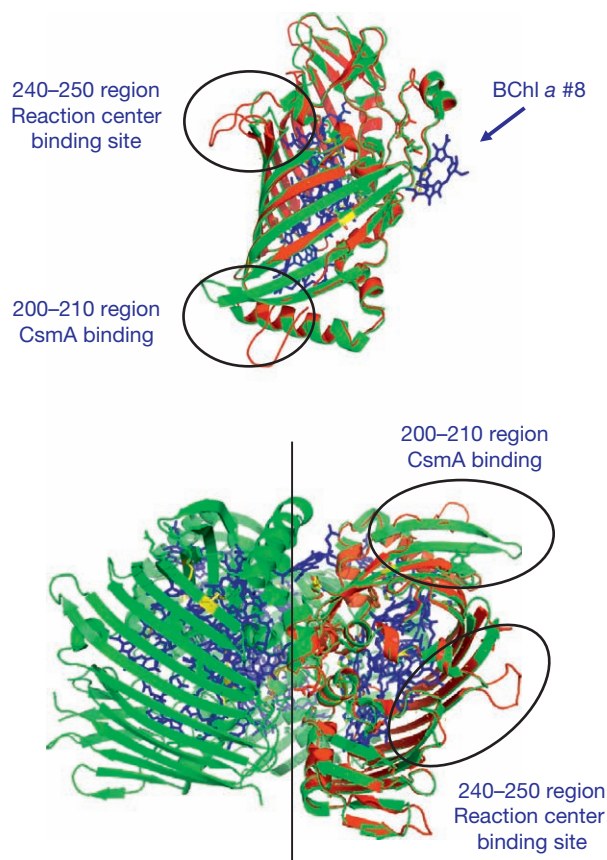


Figure 4 Comparison of the X-ray structure of the FMO protein from *C. tepidum* (green) to the predicted structure of the highly divergent FMO protein from *Candidatus C. thermophilum* (red). BChl *a* molecules are shown in blue. FMO from *Candidatus C. thermophilum* is predicted to differ most from that of *C. tepidum* in the regions that interact with the reaction center (amino acids 240–250) and the chlorosome baseplate protein CsmA (amino acids 200–210). For other details, see text.

chlorosome envelope is best described as an asymmetric bilayer membrane, in which glycolipids and proteins form the outer leaflet and the esterifying alcohol tails of BChls form the inner leaflet (Figure 3). The envelope is stabilized and mostly comprised of 10 different proteins, which belong to four structure motif families: CsmA/CsmE, CsmB/CsmF/CsmH, CsmC/CsmD/CsmI, and CsmJ/CsmK/CsmL. Mutational studies in *C. tepidum* showed that CsmA is the only essential envelope protein; it forms a paracrystalline array known as the ‘baseplate’ that attaches the chlorosomes to either the FMO protein or the B808–866 complexes and RCs in the cytoplasmic membrane (see below).

Cryo-electron microscopy of chlorosomes has demonstrated that not only are chlorosomes heterogeneous in size, but also the overall organization of BChl *c* molecules inside each chlorosome is unique – that is, no two chlorosomes are structurally alike. End-on views of chlorosomes showed that the BChl *c* molecules form coaxial, concentric nanotubular structures, which are sometimes interconnected by arched lamellae (Figures 1(c) and 1(e)). The surface layers formed by the BChl tetrapyrrole head groups are separated by spacings of ~2.1 nm (vertical arrows in Figures 1(c) and 1(e)), which

are established by the hydrophobic interactions of esterifying alcohols of two adjacent layers, in much the same way a lipid bilayer membrane is stabilized by hydrophobic interactions of fatty acids. Solid-state nuclear magnetic resonance (NMR) studies showed that the BChls form stacks of *syn-anti* dimers, in which the esterifying alcohol of one molecule points above and one points below the plane formed by the head groups. The *syn-anti* dimer stacks are actually arranged as shallow helices, and the 0.83-nm repeats observed in the cryo-electron micrographs (Figure 1(c), horizontal arrows) are formed from the repeats between these *syn-anti* stacks (Figure 1(f)). In wild-type *C. tepidum* chlorosomes, these stacks run parallel to the long axis of the chlorosome (Figure 1(g)), but in a highly ordered *bchQ bchR bchU* mutant, these stacks are approximately perpendicular to the long axis of the chlorosome (Figure 1(f)).

The chlorosomes of *C. aurantiacus* are much smaller than those of *C. tepidum* (140–220 nm in length, 30–60 nm in width, and 10–20 nm in height) and have also been extensively characterized. The baseplate of these chlorosomes has a distinctive paracrystalline lattice with a 3.3-nm spacing and is probably formed from CsmA dimers. However, the suprastructural organization of the BChls in chlorosomes of *C. aurantiacus* has not yet been determined. The proteins of the chlorosome envelope of *C. aurantiacus* are very different from those of *C. tepidum*, but the same four structure-motif families are present. The chlorosomes of *Candidatus C. thermophilum* are structurally distinct from those of *C. tepidum* and *C. aurantiacus*. Cryo-electron microscopy showed that 2.3-nm lamellar spacings are present, similar to other chlorosomes, but it is clear from cryo-electron microscopy that the BChls do not form concentric nanotubes. Instead, the BChls may form folded or twisted sheets. CsmA is found in the envelopes of these chlorosomes, together with several other proteins, one of which harbors a [2Fe–2S] cluster like CsmI/CsmJ/CsmX of *C. tepidum* and CsmP of *C. aurantiacus*. The other proteins share no sequence homology with chlorosome envelope proteins in *Chlorobi* or *Chloroflexi*.

The self-assembly of BChl *c*, *d*, or *e* into suprastructures causes a significant red-shift in their absorption spectrum. For BChl *c*, the absorption shifts from about 670 nm to about 750 nm. The methylation at C-8² and C-12¹ causes the bandwidth of the absorption band to increase because of inhomogeneous broadening, and these methylations also help to tune the absorption maximum. These effects increase the absorption cross section of chlorosomes and reduce the overlap with Chl *a*, which is found in plants, algae, and cyanobacteria, and which filter the light reaching these anaerobes because of their occurrence in the upper, oxic layer of stratified aquatic environments. The high density of BChl molecules in chlorosomes produces very short inter-BChl distances and leads to very strong excitonic coupling of the BChls. Excited states due to the absorption of photons are thus delocalized over many BChls and migrate very rapidly with sub-picosecond, energy-transfer times.

The type-1 RCs of GSB and *Candidatus C. thermophilum* produce strong reductants, which can potentially produce highly toxic, reactive oxygen species when cells are exposed to oxygen. Exposure to oxygen rapidly quenches energy transfer in chlorosomes, probably by oxidation of some of the numerous quinones found in the chlorosomes. Reactivation of

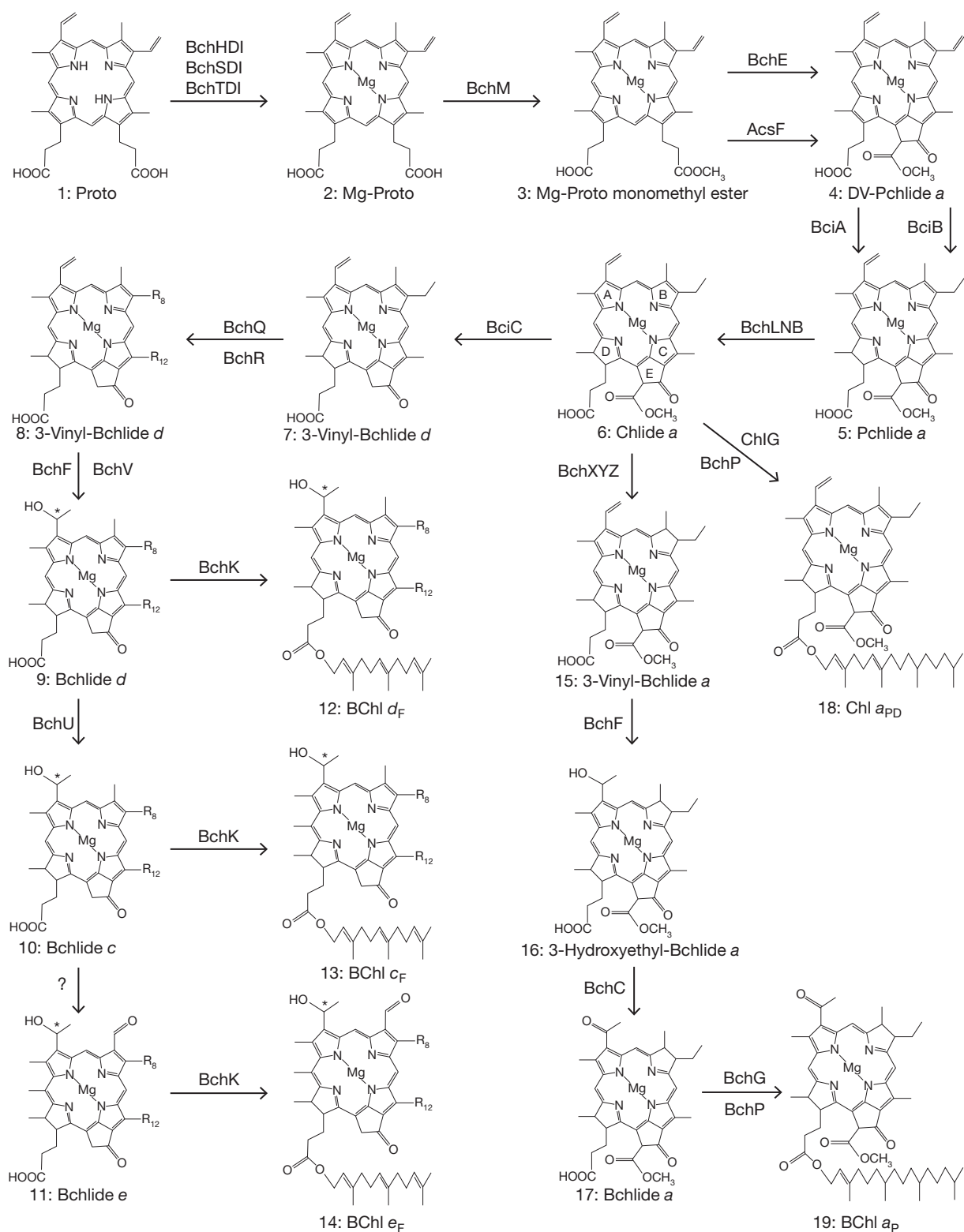


Figure 5 Scheme showing the biosynthetic pathway for the synthesis of BChls *a*, *c*, *d*, and *e* as well as Chl *a* from protoporphyrin IX. The compounds are numbered, and the gene products responsible for known reactions are indicated. For further details, see text.

energy transfer requires that the oxidized quenchers are reduced again, and this appears to be the role of heterotetrameric complexes of the [2Fe-2S] proteins (e.g., CsmI, CsmJ, CsmX, or CsmP) that occur in the chlorosome envelope. The chlorosomes of *Candidatus C. thermophilum* are also highly susceptible to quenching under oxidizing conditions, and these chlorosomes contain large amounts of menaquinone. Although an Fe-S protein related to CsmI/CsmJ/CsmX is present in the chlorosome envelopes of *Candidatus C. thermophilum*, the chlorosomes of this bacterium additionally include a type-2 NADH dehydrogenase, which could transfer electrons from NADH to oxidized menaquinone to reactivate energy transfer.

The FMO BChl *a*-Binding Protein and B808-866

Light energy captured by the BChls in chlorosomes is trapped BChl *a* molecules in the baseplate; these BChl *a* molecules are bound to CsmA (Figure 3) and have an absorption maximum at about 795 nm. In all GSB and *Candidatus C. thermophilum*, this energy is then transferred to the FMO protein, a water-soluble BChl *a*-binding protein encoded by the *fmoA* gene and absorbing maximally around ~800–810 nm. FMO was the first (B)Chl protein to be crystallized, and it is a trimer composed of taco-shaped monomers that are mostly composed of beta sheets (Figures 3 and 4). The BChl *a* molecules form the filling of the taco and are covered by α -helices. Relatively long-lived quantum coherence has recently been detected in the FMO protein, which may play an important role in energy transfer. For decades it was believed that the ~40-kDa apoprotein bound seven BChl *a* molecules, but recent studies of the native protein by mass spectrometry, as well as atomic-resolution X-ray structures, have strongly suggested that an eighth BChl *a* molecule is bound to the monomer (Figure 4; also see blue diamonds in Figure 3). Mass spectrometric analyses were also used to determine the orientation of the FMO trimers relative to the RCs and chlorosome baseplate. Based upon the FMO orientation and the position of the binding site for the eighth BChl *a* molecule, it is very likely that this additional BChl *a* molecule plays an important role in energy transfer from the BChl *a* associated with the chlorosome baseplate and CsmA to FMO. It has been suggested that the esterifying alcohol of this additional BChl *a* molecule might insert into the chlorosome envelope, which could stabilize the association between these two light-harvesting structures (see Figure 3). Energy transfer from the BChl *a* pool associated with FMO to that associated with RCs is also thought to be controlled by a redox mechanism, but even under reducing, anoxic conditions, energy transfer from FMO to RCs is very inefficient *in vitro*. The biochemical basis for this inefficiency is not understood. The FMO protein of *Candidatus C. thermophilum* is very similar to that of *C. tepidum* (Figure 4; *C. tepidum* FMO structure is shown in green; the modeled *Candidatus C. thermophilum* FMO is shown in red; BChl *a* molecules are in blue) and differs significantly only in the domains predicted to contact CsmA and the RC (See Figure 4).

In *C. aurantiacus*, the BChl *a*-binding B808-866 complex replaces FMO. This complex is very similar to the LH-I complexes of purple bacteria and probably forms a ring-like

structure that surrounds the type-2 RC. Excitation energy collected by the baseplate must be transferred into the B808-866 ring, which in turn delivers the energy to the P865 special pair of the RC.

RCs in Green Bacteria

The RCs of GSB are made up of four protein subunits: PscA, PscB, PscC, and PscD, and about 16 BChl *a*, four Chl *a*, two menaquinone-7, at least one carotenoid, and three [4Fe-4S] clusters. The RC core is a homodimer of PscA molecules, which are large, 82-kDa membrane-intrinsic polypeptides with 11 transmembrane α -helices. This homodimer binds most of the electron transport cofactors and all of the pigments associated with the RCs. In addition to the core homodimer, all type-1 RCs have a second, membrane-extrinsic subunit, which is protein related to bacterial ferredoxins with two [4Fe-4S] clusters that serve as the terminal electron acceptors, usually denoted F_A and F_B . Electrons from these strongly reducing Fe/S clusters ($E_0' \sim -600$ mV) are transferred to a water-soluble electron carrier protein that is usually a soluble ferredoxin (see below). The RCs of GSB additionally include two proteins, PscC and PscD, not found in the type-1 RCs of *Candidatus C. thermophilum*. PscC encodes a membrane-anchored *c*-type cytochrome, Cyt c_Z (also known as Cyt c_{551}), which is the immediate donor of electrons to the oxidized special pair $P840^+$. There are apparently two molecules of Cyt c_Z per RC complex. PscD is a nonessential, 16-kDa subunit, which is associated with the cytoplasmic surface of the RC and is not always present in highly purified RCs. This subunit may interact with PscB and may participate in the docking of soluble electron carriers to the RC. Two FMO trimers are also tightly associated with each RC complex.

Direct absorption of a photon by the special pair, or its excitation by energy transfer from chlorosomes and/or FMO, results in the formation of an excited state, denoted as $P840^*$, of the special pair of BChl *a* molecules. The formation of this excited state causes photochemical charge separation to occur within a few picoseconds, producing $P840^+ A_0^-$, which results in an absorption decrease at 840 nm due to oxidation of the special pair. Interestingly, the primary electron acceptors, A_0 , are two of the four Chl *a* molecules found in the RC. $P840$ has a weakly oxidizing midpoint potential, $E_0' = +240$ mV, and $P840^+$ is rapidly reduced by electron transfer from Cyt c_Z ($E_0' = +180$ mV), which helps to prevent charge recombination from the reduced primary acceptor, A_0^- . Although the RCs of GSB contain two molecules of menaquinone-7, the participation of one or both of these quinones in the electron transfer chain has not been conclusively demonstrated. The terminal electron acceptor for the core complex is the F_X [4Fe-4S] cluster, which is ligated by two cysteine residues contributed by each PscA polypeptide and has a midpoint potential lower than -650 mV.

The type-1 RCs of *Candidatus C. thermophilum* have recently been isolated and characterized. The genome of this organism predicted very simple RCs composed of only two polypeptides, PscA and PscB, but because of necessity to use chaotropic agents to remove chlorosomes from the cytoplasmic membranes, PscB has not yet been observed in purified RC

preparations. The PscA polypeptide (865 amino acids) is larger than the apo-proteins of most other type-1 RCs because of the insertion of ~ 165 amino acids between trans-membrane helices 6 and 7. The PscA homodimers copurify with a carotenoid-binding protein, which is related to a type-IV pilus subunit, as a 480-kDa complex. The RC core complex had absorption maxima at 812 nm and 670 nm, and high-performance liquid chromatography (HPLC) analyses suggested the presence of 12 BChl a_P molecules and eight Chl a_{PD} molecules per P840. Surprisingly, no quinones were detected in this RC preparation. Whole cells, cell membranes, and isolated RCs all showed very similar light-induced difference spectra with maximal bleaching at 840 nm. The observed difference spectrum was very similar to that for RC preparations from *C. tepidum*.

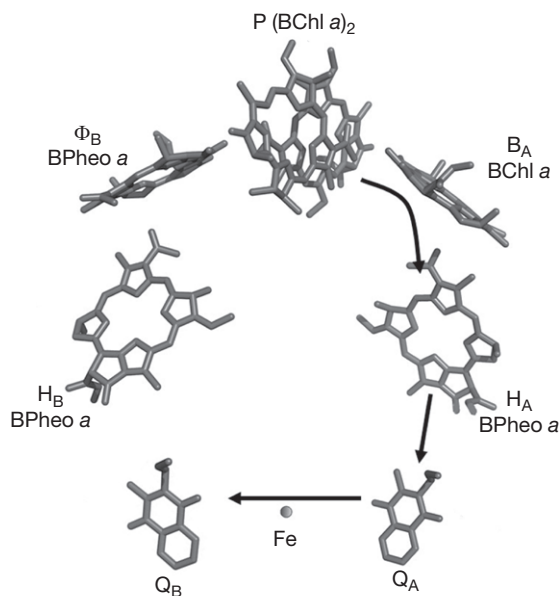


Figure 6 Arrangement of electron transport cofactors in the type-2 reaction centers of FAPs. Each BChl a and BPheo a molecule is identified. The arrows indicate the pathway for electrons from the special pair (P865) to Q_B . For additional details, see text.

All FAPs studied to date have heterodimeric type-2 RCs, which contain subunits similar in sequence to subunits PufL and PufM of bacterial RCs. A membrane-associated tetraheme cytochrome, similar to cytochrome $c-554$, is also associated with these RCs. Type-2 RCs in characterized RCs of FAPs contain three BChl a , three BPheo a , and two menaquinones (see Figure 6). The additional BPheo a molecule, indicated as Φ_B in Figure 6, is believed to be associated with the PufM on branch of cofactors that normally is not active in electron transport (The active branch of cofactors is indicated by the arrows in Figure 6). After two photons have been absorbed, Q_B would be fully reduced and would gain two protons and would dissociate from the RC and enter the quinone pool, from which it can be oxidized by alternative complex III (see below).

Electron Transport Chains in Green Bacteria

GSB are obligately photoautotrophic, and thus they must produce both ATP and strong reductants required for carbon dioxide fixation by the reverse TCA cycle. Practically, this means that they use light energy and their type-1 RCs to produce strong reductants, that is, reduced ferredoxin and NADPH by oxidizing reduced sulfur compounds. Electrons from the F_A and F_B [4Fe-4S] clusters on PscB are presumably transferred to the small, abundant bacterial ferredoxins found in GSB cells. The redox potentials for two of these ferredoxins in *C. tepidum* are highly reducing (ca. -514 and -584 mV) and can readily support the reductive carboxylation reactions required for carbon dioxide fixation. GSB also have a ferredoxin:NADP $^+$ oxidoreductase that is more closely related to thioredoxin reductase than to the enzyme of cyanobacteria algae, and plants.

Although it seems extremely likely that GSB can perform cyclic electron transport for ATP production, the mechanism by which this occurs is presently not clear. GSB strains have an incomplete type-1 NADH dehydrogenase, which lacks the diaphorase subunits (NuoE, NuoF, and NuoG) that would be required if this complex were to oxidize pyridine nucleotides. This suggests that reduced ferredoxin, or possibly an uptake hydrogenase (or both), might deliver electrons to this complex.

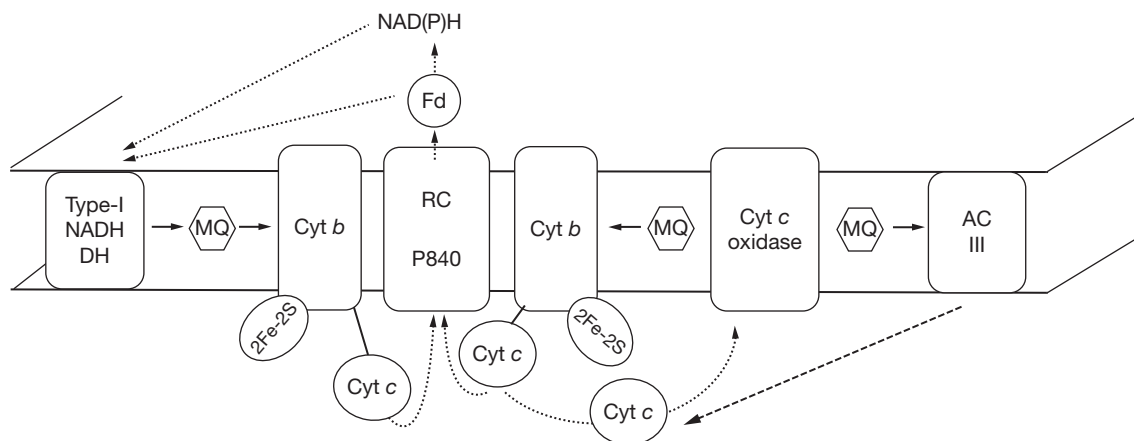


Figure 7 Scheme showing the electron transport complexes identified in the genome of *Candidatus C. thermophilum*. Dotted arrows indicate possible pathways of electron transport. MQ, menaquinone/menaquinol; AC III, alternative complex III.

Electrons from menaquinol would still have to be returned to the RC by some route, and how this is accomplished is still unknown. Although GSB have genes encoding Cyt *b* (PetB) and a Rieske Fe/S protein (PetC), they seemingly do not produce a menaquinol:Cyt *c* oxidoreductase. No *c*-type cytochrome similar to Cyt *c*₁ or Cyt *f* is encoded in the genomes of GSB. It is possible that PscC/Cyt *c*₂ or another small membrane-bound, *c*-type cytochrome performs a function analogous to that of Cyt *c*₁. However, complexes capable of menaquinol oxidation have proven to be difficult to isolate and characterize.

The electron transport chain of *Candidatus C. thermophilum* has not yet been characterized biochemically, but inferences from the genome sequence suggest that it has some similarities to that of GSB. This photoheterotroph must have some mechanism to produce protonmotive force by cyclic electron transport, but it must also have a respiratory electron transport chain. It is assumed that electrons are transferred from PscB to a soluble ferredoxin and from there to ferredoxin:NADPP⁺ oxidoreductase to produce NADPH (Figure 7). Either NADPH or reduced ferredoxin must return electrons to a membrane-bound complex so that proton translocation can occur for ATP synthesis. This could occur via type-I NADH dehydrogenase, which is encoded in the genome. Cytochrome oxidase is also encoded in the genome and, surprisingly, the genome includes genes for three different complexes that could oxidize menaquinol (Figure 7). Unlike GSB but similar to FAPs, *Candidatus C. thermophilum* has alternative complex III (AC III) for oxidation of menaquinol and reduction of Cyt *c*. In addition, two Cyt *b*/Rieske complexes are encoded in the genome. These complexes have tethered Cyt *c* domains that may transfer electrons to the RCs or to cytochrome oxidase. A small soluble Cyt *c*, which could also transfer electrons among these complexes, is also encoded in the genome.

FAPs are predominantly photoheterotrophs, and they use their photosynthetic apparatus to generate protonmotive force for ATP synthesis. However, unlike purple bacteria, FAPs do not have any electron transfer complex similar to the cytochrome *bc*₁ complex (ubiquinol:cytochrome *c* oxidoreductase) that is widely distributed in bacteria. Instead, *C. aurantiacus* has two distinct alternative complex III enzymes. One may be used preferentially during growth under anoxic, phototrophic conditions and the other may be used under oxic, respiratory conditions. Exactly how electrons are transferred between these complexes and cytochrome oxidase and the RCs is still not clear. Two copper-containing proteins, auracyanin-A and auracyanin-B, may play a role in these processes.

See also: **Bioenergetics:** Chlorophylls and Carotenoids; F_x, F_A, and F_B Iron–Sulfur Clusters in Type I Photosynthetic Reaction Centers; ORAI Store-Operated Calcium Channel.

Further Reading

- Bryant DA and Frigaard N-U (2006) Prokaryotic photosynthesis and phototrophy illuminated. *Trends in Microbiology* 14: 488–496.
- Bryant DA, Garcia Costas AM, Maresca JA, et al. (2007) “*Candidatus Chloracidobacterium thermophilum*”: An aerobic phototrophic acidobacterium. *Science* 317: 523–526.
- Bryant DA, Liu Z, Li T, et al. (2012) Comparative and functional genomics of anoxygenic green bacteria from the taxa *Chlorobi*, *Chloroflexi*, and *Acidobacteria*. In: Burnap RL and Vermaas W (eds.) *Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration*, vol. 35, pp. 47–102. Dordrecht, The Netherlands: Springer.
- Frigaard N-U and Bryant DA (2006) Chlorosomes: Antenna organelles in green photosynthetic bacteria. In: Shively JM (ed.) *Complex Intracellular Structures in Prokaryotes. Microbiology Monographs*, vol. 2, pp. 79–114. Berlin: Springer.
- Frigaard N-U, Li H, Gomez Maqueo Chew A, Maresca JA, and Bryant DA (2003) *Chlorobium tepidum*: Insights into the physiology and biochemistry of green sulfur bacteria from the complete genome sequence. *Photosynthesis Research* 78: 93–117.
- Ganapathy S, Oostergetel GT, Wawrzyniak PK, et al. (2009) Alternating *syn*–*anti* bacteriochlorophylls form concentric helical nanotubes in chlorosomes. *Proceedings of the National Academy of Sciences of the United States of America* 106: 8525–8530.
- Gomez Maqueo Chew A and Bryant DA (2007) Chlorophyll biosynthesis in bacteria: The origins of structural and functional diversity. *Annual Reviews of Microbiology* 61: 113–129.
- Hauska G, Schödl T, Remigy H, and Tsiotis G (2001) The reaction center of green sulfur bacteria. *Biochimica et Biophysica Acta* 1507: 260–277.
- Larson CR, Seng CO, Lauman L, et al. (2011) The three-dimensional structure of the FMO protein from *Pelodictyon phaeum* and the implications for energy transfer. *Photosynthesis Research* 107: 139–150.
- Liu Z and Bryant DA (2011) Biosynthesis and assembly of bacteriochlorophyll *c* in green bacteria: Theme and variations. In: Kadish KM, Smith KM, and Guillard R (eds.) *Handbook of Porphyrin Science*, set IV-20, in press. Hackensack, NJ: World Scientific Publishing.
- Panitchayangkoon G, Hayes D, Fransted KA, et al. (2010) Long-lived quantum coherence in photosynthetic complexes at physiological temperature. *Proceedings of the National Academy of Sciences of the United States of America* 107: 12766–12770.
- Pedersen MØ, Linnanto J, Frigaard N-U, Nielsen NC, and Miller M (2010) A model of the protein–pigment baseplate complex in chlorosomes. *Photosynthesis Research* 104: 233–243.
- Romberger SP and Golbeck JH (2010) The bound iron–sulfur clusters of type-I homodimeric reaction centers. *Photosynthesis Research* 104: 333–346.
- Sadekar S, Raymond J, and Blankenship RE (2006) Conservation of distantly related membrane proteins: Photosynthetic reaction centers share a common structural core. *Molecular Biology and Evolution* 23: 2001–2007.
- Tronrud D, Wen J, Gay L, and Blankenship RE (2009) The structural basis for the difference in absorbance spectra for the FMO antenna protein from various green sulfur bacteria. *Photosynthesis Research* 100: 79–87.

Green Bacteria: Secondary Electron Donor (Cytochromes)

H Oh-oka, Osaka University, Osaka, Japan

R E Blankenship, Washington University in St. Louis, St. Louis, MO, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

c-Type cytochrome A class of cytochromes in which the heme cofactor is covalently attached to the protein via thioether linkages to cysteine amino acid residues.

Cytochrome A heme-containing protein that serves as an electron transfer cofactor in biological systems.

Fe-S protein An iron-sulfur protein which holds a [2Fe-2S]-, [3Fe-4S]-, or [4Fe-4S]-type cluster and functions as an electron transfer cofactor in biological systems.

Flavoproteins A flavin adenine dinucleotide (FAD)-containing protein that serves as an electron transfer cofactor in biological systems.

P840 A special pair of bacteriochlorophyll *a* in green sulfur bacterial reaction center (RC), which serves as the primary electron donor.

P870 A special pair of bacteriochlorophyll *a* in green nonsulfur bacterial RC, which serves as the primary electron donor.

Reaction center (RC) A pigment-protein complex that converts light energy into chemical energy in photosynthetic organisms.

Green photosynthetic bacteria are a class of photosynthetic organisms that carry out light-driven electron transfer that leads to long-term energy storage. A reaction center (RC), pigment-protein complex, mediates the electron transfer process. Cytochromes, proteins that contain Fe-containing heme prosthetic groups, are an essential part of the photosynthetic electron transfer system. The green photosynthetic bacteria contain a rich variety of different cytochromes. Various cytochromes act as either soluble or membrane-bound electron carriers and are involved in substrate oxidation processes, as well as proton translocation across the membrane.

Cytochromes in Green Bacteria

There are two phyla of green photosynthetic bacteria, green sulfur bacteria and green nonsulfur bacteria, and the latter group is now more correctly called filamentous anoxygenic phototrophic (FAP) bacteria. Both groups of organisms contain cytochromes, although the patterns and types of cytochromes in the two phyla are very different.

Green sulfur bacteria (chlorobiaceae) are strictly anaerobic, photosynthetic organisms whose RC complex shares common structural and functional features with the photosystem (PS) I RC complex of plants and cyanobacteria or the heliobacterial RC complex. This type-1 (also called Fe-S-type) RC complex consists of five subunits, including a homodimeric core RC (with two identical PscA subunits with a molecular mass of 65 kDa), Fenna Matthews Olson (FMO) antenna protein (41 kDa), Fe-S protein (PscB) that includes the early electron acceptors F_A and F_B (31 kDa), cytochrome *c_z* (PscC) (22 kDa), and 18-kDa protein (PscD), respectively. Cytochrome *c_z* serves as a secondary electron donor to the primary electron donor P840. The green sulfur bacteria utilize sulfur compounds as a source of reducing power and therefore also contain additional electron transfer components, primarily soluble cytochromes *c*, which are involved in the metabolic pathway of sulfur oxidation (Figure 1).

FAP bacteria are not closely related organisms to the green sulfur bacteria, with the exception of the primary shared characteristic, the chlorosome antenna complex. Many of these organisms are facultatively aerobic photosynthetic bacteria that are often found in microbial mat environments with cyanobacteria. The only member of this group that has been well studied is *Chloroflexus aurantiacus*, which is found in hot spring environments all over the world. FAP bacteria contain a type-2 (also called quinone-type) RC complex that includes bacteriopheophytins and quinones as early electron acceptors. The type-2 RC is similar to that found in purple photosynthetic bacteria and photosystem II of oxygenic photosynthetic organisms. The FAP RC consists of a heterodimeric complex of two proteins, L and M, plus a tightly bound tetra-heme *c*-type cytochrome, *c*-554. Cyt *c*-554 is a membrane-associated cytochrome of molecular mass 48 kDa, which serves as a secondary electron donor to the primary electron donor P870. FAP bacteria have not been reported to contain any soluble *c*-type cytochromes. They also do not contain a membrane-bound cytochrome *bc*₁ complex.

Types of c-Type Cytochromes in Green Sulfur Bacteria

Recent genomic analyses of green sulfur bacteria have enabled us to see the whole picture of the photosynthetic electron transport chain linked to sulfur oxidation pathways. A number of *c*-type cytochromes are supposed to be involved in their relevant reactions as electron carriers. These include soluble cytochromes *c*-554/555, flavocytochrome *c*, and SoxAX, the membrane-bound cytochrome *c_z*, and the membrane-bound cytochrome *bc*₁ complex. The gene locus encoding each protein is presented as CT#### according to the annotation listed in the genomic database of *Chlorobium tepidum* (syn. *Chlorobaculum tepidum*). The pattern described here is similar in most species of green sulfur bacteria.

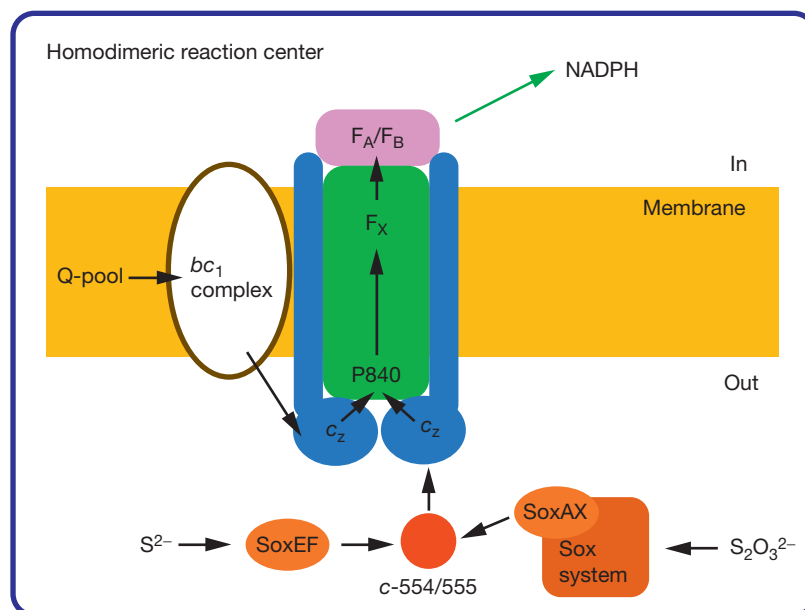


Figure 1 Schematic structural model for the reaction center and associated cytochromes involved in electron transport in photosynthetic green sulfur bacteria. Details of the various cytochromes can be found in the text.

Cytochrome c_z (PscC)

This cytochrome c_z (CT1639) is a monoheme type with a molecular mass of 23 kDa and is tightly bound to the RC complex with putative amino-terminal helices that span the membrane three times. Two molecules of cytochrome c_z are present in one RC complex. Each of them donates an electron to the P840 in a reaction mode with high viscosity dependence. The carboxyl-terminal heme-containing hydrophilic moiety seems to fluctuate on the RC surface and search for an appropriate site for the reaction to proceed. The cytochrome c_z was designated in order to characterize its functional similarity to the membrane-bound cytochrome c_y in *Rhodobacter capsulatus* and mediates the electron transfer between menaquinol:cytochrome c oxidoreductase and P840 RC complex. The exact α -peak wavelength of cytochrome c_z varies from 553 nm in membranes to 552–551 nm in a solubilized RC complex, which appears to be ascribed to some structural modification of cytochrome c_z induced by different solubilization procedures and/or a species-dependent difference.

Cytochrome c -554/555

This water-soluble cytochrome c -554/555 (CT0075) has a molecular mass of 10 kDa and gives a characteristic asymmetric α -absorption peak, similar to that of algal water-soluble cytochrome c_6 that serves as an electron donor to PSI RC. Its α -peak wavelength is species dependent: it exhibits 554 nm in *Cba. tepidum* and 555 nm in *Cba. parvum*. A locus accession number of cytochrome c -554 in *Cba. tepidum* is CT0075. The cytochrome c -554/555 donates an electron directly to the membrane-bound cytochrome c_z and never shuttles between menaquinol:cytochrome c oxidoreductase and P840 RC complex as cytochrome c_2 in purple bacteria.

Flavocytochrome c

The flavocytochrome c (FccAB/SoxEF) consists of an 11-kDa cytochrome c -553 (FccA/SoxE, CT2080) that binds a single heme plus a 47-kDa flavoprotein (FccB/SoxF, CT2081) subunit and serves as a sulfide:cytochrome c reductase. The flavoprotein subunit contains a flavin adenine dinucleotide (FAD) that is covalently bound to a cysteine residue through a thioether linkage. The electron transfer is considered to proceed as follows: $S^{2-} \rightarrow$ flavocytochrome c (FAD $\rightarrow$ heme c -553) $\rightarrow$ cytochrome c -554/555 $\rightarrow$ cytochrome c_z .

SoxAX

Although a previous observation was the involvement of dimeric cytochrome c -551 in thiosulfate oxidation, it has recently been clarified that SoxAX, which is a heterodimer of distinct monoheme-type cytochromes c (CT1019/1016) with apparent molecular masses of 30 and 10 kDa, respectively, makes a ternary complex along with SoxK (CT1020, 9 kDa) and reduces cytochrome c -554/555 in the presence of SoxB (CT1021, 60 kDa) *in vitro* by oxidizing thiosulfate. A complex probably composed of SoxYZ (CT1017/1018, 13 and 9 kDa, respectively) and SoxB is assumed to be a thiosulfate:cytochrome c reductase (TCR, 80 kDa), which had been isolated in the early 1970s but not identified biochemically.

Cytochrome bc_1 Complex

The cytochrome bc_1 complex is a multisubunit membrane-bound quinol:cytochrome c oxidoreductase found in many bacteria and in mitochondria, where it is also called complex III. It consists of a b -type cytochrome, a c -type cytochrome (cytochrome c_1), a Rieske Fe–S protein, and in some cases additional subunits. It oxidizes reduced quinone (quinol)

and transfers the electrons to soluble *c*-type cytochromes, at the same time pumping protons across the membrane and creating a proton-motive force that is used to make adenosine triphosphate (ATP). The cytochrome *bc*₁ complex from green sulfur bacteria is not as well characterized as those from other organisms; no *fbcC* (*petC*) gene coding for cytochrome *c*₁ has been found either downstream or upstream of the unit housing the *fbcF* (*petA*, CT0302) and *fbcB* (*petB*, CT0303) genes coding for the Rieske Fe-S protein and cytochrome *b*, respectively (see below). An unusual aspect of it is that it uses menaquinone, which is lower in redox potential than ubiquinone, which is used in the better-characterized complexes.

Other Cytochromes in Green Sulfur Bacteria

The genome project of *Chl. tepidum* has revealed five more *c*-type cytochromes (CT0073, CT0188, CT1704, CT1734, and CT2026) whose functions are still unknown (protein length is ~130–150 amino acids). The spectroscopic data have suggested the presence of cytochrome *c*-556, which is assumed to be a counterpart of cytochrome *c*₁ in quinol:cytochrome *c* oxidoreductase. The most probable candidate for this cytochrome *c* is considered to be CT0073, whose gene is located adjacent of CT0075 encoding cytochrome *c*-554/555.

Cytochromes in FAP Bacteria

The pattern of cytochromes in FAP bacteria is strikingly different from that found in green sulfur bacteria. The only cytochrome that has been characterized in detail is the RC-associated cytochrome *c*-554.

Cytochrome *c*-554

Cytochrome *c*-554 is a membrane-bound tetraheme cytochrome in which all four of its heme groups have α -band maxima ~554 nm. This cytochrome is the secondary electron donor to the RC and bears considerable similarity to the tetraheme cytochrome that is found in many purple photosynthetic bacteria, the best characterized is the tetraheme cytochrome from *Rhodospseudomonas viridis* (recently renamed *Blastochloris viridis*). The four heme groups in *c*-554 have redox potentials of 0 to +300 mV (vs. normal hydrogen electrode (NHE)).

Other Cytochromes in FAP Bacteria

The genomes of FAP bacteria contain other open reading frames that code for *c*-type cytochromes, although most have not been characterized. There is no cytochrome *bc*₁ complex present, as indicated by both genetic and biochemical analysis.

Alternative Complex III

Instead of the cytochrome *bc*₁ complex, another complex, now named alternative complex III is present. It contains seven distinct protein subunits and has two *c*-type cytochromes. One is a monoheme cytochrome and one is a pentaheme cytochrome. This complex has menaquinol:auracyanin oxidoreductase activity. Characterization of this complex is underway.

See also: **Bioenergetics:** Cytochrome *bc*₁ Complex (Respiratory Chain Complex III); Respiratory Chain Complex II and Succinate: Quinone Oxidoreductases.

Further Reading

- Azai C, Kim K, Kondo T, et al. (2011) A heterogeneous tag-attachment to the homodimeric type I photosynthetic reaction center core protein in the green sulfur bacterium *Chlorobaculum tepidum*. *Biochimica et Biophysica Acta* 1807: 803–812.
- Blankenship RE, Madigan MT, and Bauer CE (eds.) (1995) *Anoxygenic Photosynthetic Bacteria*. Dordrecht: Kluwer.
- Dracheva S, Williams JC, Van Driessche G, Van Beeumen JJ, and Blankenship RE (1991) The primary structure of cytochrome *c*-554 from the green photosynthetic bacterium *Chloroflexus aurantiacus*. *Biochemistry* 30: 11451–11458.
- Frigaard NU and Bryant DA (2004) Seeing green bacteria in a new light: Genomics-enabled studies of the photosynthetic apparatus in green sulfur bacteria and filamentous anoxygenic phototrophic bacteria. *Archives of Microbiology* 182: 265–276.
- Gao X, Xin Y, and Blankenship RE (2009) Enzymatic activity of the alternative complex III as a menaquinol: Auracyanin oxidoreductase in the electron transfer chain of *Chloroflexus aurantiacus*. *FEBS Letters* 583: 3275–3279.
- Hauska G, Schoedl T, Remigy H, and Tsiotis (2001) The reaction center of green sulfur bacteria. *Biochimica et Biophysica Acta* 1507: 260–277.
- Ogawa T, Furusawa T, Nomura R, et al. (2008) SoxAX binding protein, a novel component of the thiosulfate-oxidizing multienzyme system in the green sulfur bacterium *Chlorobium tepidum*. *Journal of Bacteriology* 190: 6097–6110.
- Oh-oka H, Iwaki M, and Itoh S (1998) Membrane-bound cytochrome *c*₂ couples quinol oxidoreductase to the P840 reaction center complex in isolated membranes of the green sulfur bacterium *Chlorobium tepidum*. *Biochemistry* 37: 12293–12300.
- Tsukatani Y, Azai C, Kondo T, Itoh S, and Oh-oka H (2008) Parallel electron donation pathways to cytochrome *c*₂ in the type I homodimeric photosynthetic reaction center complex of *Chlorobium tepidum*. *Biochimica et Biophysica Acta* 1777: 1211–1217.
- Yanyushin MF, del Rosario M, Brune DC, and Blankenship RE (2005) A new class of bacterial membrane oxidoreductases. *Biochemistry* 44: 10037–10045.

Green Bacteria: The Light-Harvesting Chlorosome

J M Olson, University of Massachusetts, Amherst, MA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 325–330, © 2004, Elsevier Inc.

Glossary

Chemoheterotroph An organism that requires organic compounds and a chemical source of energy.

Excitonic interaction An interaction in which the excited state is shared by more than one molecule.

Förster energy transfer Nonradiative resonance transfer.

Heliobacteria Photosynthetic bacteria that contain bacteriochlorophyll *g* (BChl *g*) and an iron–sulfur-type reaction center (RC) similar to those found in green sulfur

bacteria and photosystem I of cyanobacteria and chloroplasts.

Homolog A molecule that is identical to another molecule except for one substituent.

Photoautotroph An organism that requires light and inorganic carbon.

Proteobacteria A subgroup of Gram-negative bacteria that includes purple photosynthetic bacteria.

Chlorosomes are the main light-harvesting structure of green filamentous bacteria and green sulfur bacteria and are filled with aggregated bacteriochlorophyll (BChl) *c*, *d*, or *e*. BChl *a* is associated with a small protein in the envelope of the chlorosome, whereas the other BChls are organized in rod-like structures located in the interior of the chlorosome and contain little or no protein. Chlorosomes are appressed to the cytoplasmic side of the cytoplasmic membrane, and the main path of excitation energy transfer can be written as

BChl *c*, *d*, or *e* (chlorosome) → BChl *a* (chlorosome)
→ BChl *a* (Fenna – Matthews – Olson,
FMO or membrane proteins)
→ BChl *a* (reaction center, RC)

Chlorosomes are the characteristic light-harvesting complexes of green filamentous bacteria (green nonsulfur bacteria, Chloroflexaceae) and green sulfur bacteria (Chlorobiaceae). Chlorosomes are filled with BChl *c*, *d*, or *e* molecules in a highly aggregated state and are appressed to the cytoplasmic side of the cytoplasmic membrane. BChl *c*, *d*, and *e* form a family of chlorophylls (chlorosome chlorophyll) that contain a hydroxyethyl group at position 3 as shown in Figure 1. The main esterifying alcohol is farnesol in green sulfur bacteria and stearyl in filamentous bacteria, which contain only BChl *c*. In some green sulfur bacteria, there are at least four chemically distinct homologs of BChl *c* that differ in the degree of alkylation on the chlorin ring at positions 8 and 12 (see Figure 1 and Table 1). In addition to chlorosome chlorophyll, all chlorosomes contain a small amount of BChl *a*, the same chlorophyll found in proteobacteria (purple bacteria). BChl *a* is associated with a small protein in the envelope of the chlorosome, whereas chlorosome chlorophyll is located in the interior and is organized in rod-like structures containing little or no protein.

Phylogeny

From a phylogenetic viewpoint, green filamentous bacteria and green sulfur bacteria are two distinct groups based on

16S rRNA, RC type, and physiology. Green filamentous bacteria contain a quinone-type RC similar to that found in proteobacteria and similar to the RC of photosystem II in cyanobacteria and plant chloroplasts as well. Green sulfur bacteria contain an iron–sulfur-type RC similar to those found in heliobacteria and in photosystem I of cyanobacteria and chloroplasts. The filamentous bacteria live either as facultative photoautotrophs or as respiring chemoheterotrophs. They are found mostly in hot springs, often in mixed populations with cyanobacteria. The sulfur bacteria are obligate photoautotrophs and strict anaerobes that grow in dim light in sulfide-rich environments such as effluents of sulfur springs and the lower layers of stratified lakes and in marine habitats. In one extreme case, green sulfur bacteria have been found at a depth of 80 m in the Black Sea.

Structural Features

Size and Internal Structure

Chlorosomes are mostly made up of chlorophyll, but lipid, protein, carotenoid, and quinone are also present. Typically, the chlorosomes from filamentous bacteria (designated F-chlorosomes) are ~100 nm long, 20–40 nm wide, and 10–20 nm high. Chlorosomes from sulfur bacteria (designated S-chlorosomes) are considerably larger with lengths 70–260 nm and widths 30–100 nm. All chlorosomes consist of a core and a 2–3-nm envelope as shown in Figure 2. The core consists of rod elements made up of aggregated chlorosome chlorophyll, while the envelope is a monolayer of mostly galactolipid and some protein. Each chlorosome may contain 10–30 rod elements. The rod elements of S-chlorosomes are 10 nm in diameter, whereas those of F-chlorosomes are only 5.2–6 nm. The envelopes of both F- and S-chlorosomes contain BChl *a* associated with a specific protein, CsmA. The molar ratio of chlorosome chlorophyll to BChl *a* in S-chlorosomes is ~90/1, whereas in F-chlorosomes it is ~25/1.

Proteins

In highly purified S-chlorosomes, the mass ratio of protein to chlorophyll is very low – only 0.5 compared to 6.3 for the FMO BChl *a* protein found in green sulfur bacteria. The proteins of chlorosomes are localized in the envelope. In F-chlorosomes, these proteins are called CsmA, CsmM, and CsmN. In S-chlorosomes, they are called CsmA, CsmB, CsmC, CsmD, and CsmE. The major component in both kinds of chlorosomes is CsmA with a molecular mass of 5.7 kDa in F-chlorosomes and 6.3 kDa in S-chlorosomes. This protein binds BChl *a* in both kinds of chlorosomes, and the BChl *a*–CsmA complexes are thought to be concentrated in a base-plate that is in contact with BChl *a* proteins associated with the cytoplasmic membrane. Some chlorosome proteins may be responsible for maintaining the native shape.

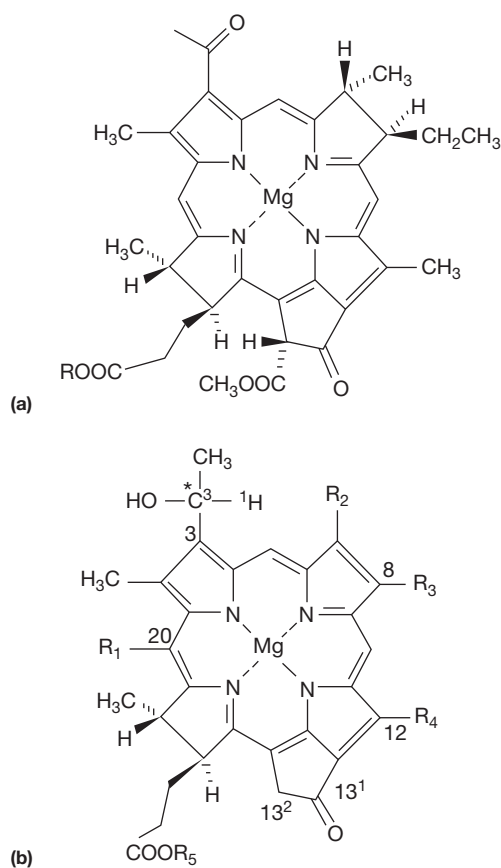


Figure 1 Structures of (a) BChl *a* and (b) the chlorosome chlorophylls (BChl *c*, *d*, and *e*). For BChl *a*, R = phytol (see Table 1 for details about the chlorosome chlorophylls).

Carotenoids

All chlorosomes contain carotenoids which function as light-harvesting pigments and triplet-state quenchers.

Spectral Properties

In Figure 3 the absorbance and fluorescence emission spectra of F-chlorosomes indicate the presence of a large amount of BChl *c*, a small amount of BChl *a*, and energy transfer from BChl *c* to BChl *a*. The absorbance and fluorescence peaks at 740 and 750 nm, respectively, belong to BChl *c*, whereas the peaks at 795 and 801 nm belong to BChl *a*. The circular dichroism (CD) spectrum (not shown) demonstrates a strong excitonic interaction among the BChl *c* molecules. All these properties of BChl *c* in both F- and S-chlorosomes are completely different from those of monomeric BChl *c* dissolved in polar solvents. Monomeric BChl *c* has absorption and fluorescence peaks at ~660 and 665 nm, and a CD spectrum with a very small negative peak at 660 nm. Pure BChl *c* forms aggregates in dried films or when dissolved in nonpolar solvents such as hexane and carbon tetrachloride. The spectral properties of these aggregates are similar to those of chlorosomes and suggest that in chlorosomes BChl *c* exists in an aggregated state. The absence of proteins from the aggregates *in vitro* supports the idea that BChl *c* does not require protein to form aggregates in the chlorosome.

Models of Chlorophyll Organization

Although it is not yet possible to present a definitive structure for the chlorophyll aggregates in the chlorosome, several characteristics have been established:

1. The aggregates of chlorosome chlorophyll (BChl *c*, *d*, or *e*) are the result of chlorophyll–chlorophyll interactions. The molecules are close together and oriented so as to give a strong excitonic interaction.
2. The chlorophyll molecules also exhibit long-range order with their Q_y transitions parallel to the long axis of the chlorosome.
3. The 13¹-carbonyl of each chlorophyll molecule interacts with a functional group on another chlorophyll molecule.
4. The 3¹-hydroxyethyl group is coordinated directly to a Mg atom in a second chlorophyll molecule and is H-bonded to the 13¹-carbonyl of a third chlorophyll molecule.
5. Essentially, all the Mg atoms in the chlorophyll molecules are five-coordinate.

Using computer modeling, Prokhorenko and colleagues have developed the useful model of chlorophyll organization

Table 1 Ring substituents found on chlorosome chlorophylls (see Figure 1)

Chlorophyll	R ₁	R ₂	R ₃	R ₄	R ₅
BChl <i>c</i> (F-chlorosomes)	Methyl	Methyl	Ethyl	Methyl	Stearyl (mostly)
BChl <i>c</i> (S-chlorosomes)	Methyl	Methyl	Ethyl <i>n</i> -propyl Isobutyl	Methylethyl	Farnesyl (mostly)
BChl <i>d</i> (S-chlorosomes)	H	Methyl	Ethyl <i>n</i> -propyl Isobutyl	Methylethyl	Farnesyl (mostly)
BChl <i>e</i> (S-chlorosomes)	Methyl	Formyl	Ethyl <i>n</i> -propyl Isobutyl	Ethyl	Farnesyl (mostly)

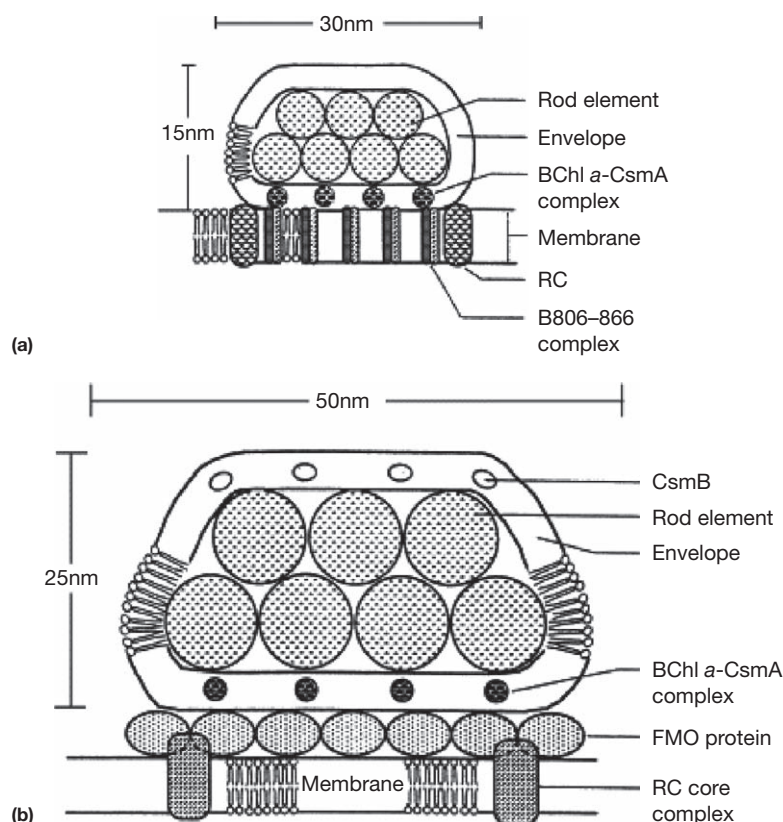


Figure 2 Cross-sectional views of antenna system models for (a) green filamentous bacteria and (b) green sulfur bacteria. Relatively small chlorosomes are shown in both cases. Adapted from Olson JM (1998) Chlorophyll organization and function in green photosynthetic bacteria. *Photochemistry and Photobiology* 67: 61–75.

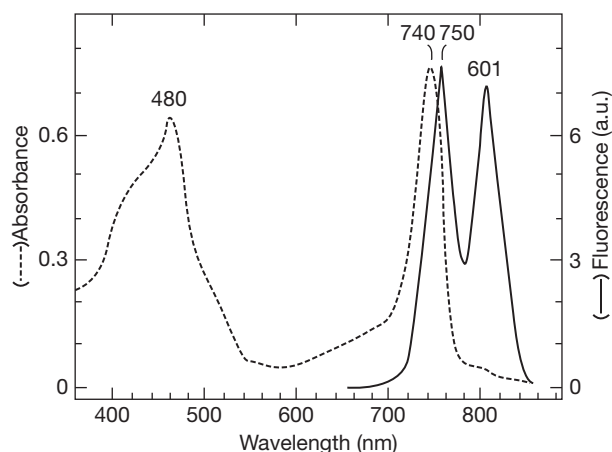
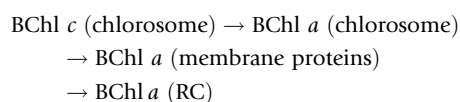


Figure 3 Absorption and fluorescence emission spectra of F-chlorosomes. Fluorescence was excited at 460 nm. Adapted from Brune DC, Gerola PD, and Olson JM (1990) Circular dichroism of green bacterial chromosomes. *Photosynthesis Research* 24: 253–263.

shown in Figure 4. The large-scale structure is a tube with a diameter of 4.6 nm (with respect to Mg atoms), corresponding to the diameter of the rod elements in F-chlorosomes. The chlorophyll macrocycles are perpendicular to the surface of the tube, and the hydrocarbon tails (not shown) are oriented toward the outside of the tube.

Energy Transfer

The overall pathway of excitation energy transfer upon irradiation of filamentous bacteria with red or far-red light can be written as follows:



The membrane proteins are labeled B806–866 in Figure 2(a). In sulfur bacteria, the FMO protein replaces the membrane proteins as shown in Figure 2(b). Energy transfer between chlorosome chlorophyll molecules takes place by a relatively fast exchange mechanism, whereas energy transfer between chlorosome chlorophyll and BChl *a* in the chlorosome takes place by a relatively slow Förster mechanism. The chlorosome light-harvesting system is an example of an energy funnel in which each successive pigment pool has progressively red-shifted absorption and fluorescence emission spectra. The descending energy levels thus provide an excitation gradient into the RC.

Energy transfer in the green sulfur bacteria is regulated by the redox potential *in vivo*. Under oxidizing conditions, the overall energy-transfer efficiency drops from ~100% to 10% or less. Redox titration of fluorescence in isolated chlorosomes gives a midpoint potential of –146 mV at pH 7.0. (This regulation is mediated by the quinone in the chlorosomes.)

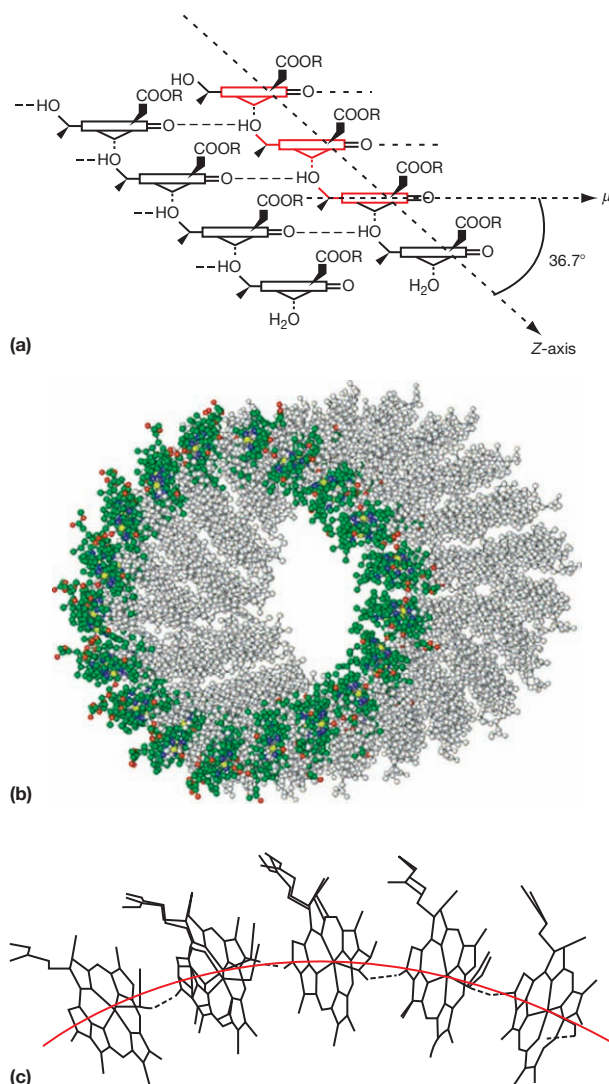


Figure 4 (a) The arrangement of chlorosome chlorophylls in a stack. The dashed Z-axis connects the Mg atoms of the chlorophylls in a single stack. (b) Space view of the tubular chlorophyll aggregate obtained by molecular modeling. The farnsyl groups, pointing to the outside of the rod, are not shown. (c) Arrangement of chlorophyll molecules at the outside edge of the chlorosome. Reproduced from Prokhorenko VI, Steensgaard DB, and Holzwarth AR (2000) Exciton dynamics in the chlorosomal antennae of the green bacteria *Chloroflexus aurantiacus* and *Chlorobium tepidum*. *Biophysical Journal* 79: 2105–2120.

The formation of fluorescence quenchers at high redox potentials may protect the bacteria from lethal photo-oxidation reactions such as superoxide formation resulting from the reduction of ferredoxin in the presence of oxygen. These redox effects are largely missing from the green filamentous bacteria, whose RCs do not reduce ferredoxin.

Evolutionary Considerations

The presence of chlorosomes in both green filamentous bacteria and green sulfur bacteria is the only justification for lumping them together as green bacteria. Although the two types of chlorosomes (F and S) show differences in size and rod element morphology, the organization of the chlorosome chlorophylls appears to be basically similar, and the *csmA*/CsmA genes and proteins appear to be homologous. This suggests that the F- and S-chlorosomes are descended from a common ancestor coded for by an ancestral genome containing the *csmA* gene.

See also: **Bioenergetics:** Chlorophylls and Carotenoids; **Green Bacteria:** Chlorophyll Biosynthesis, Light-Harvesting, Reaction Centers, and Electron Transport; **Green Bacteria:** Secondary Electron Donor (Cytochromes); **Metabolism Vitamins and Hormones:** Photosynthesis; Photosynthetic Carbon Dioxide Fixation.

Further Reading

- Blankenship RE and Matsuura K (2003) Antenna complexes from green photosynthetic bacteria. In: Green BR and Parsons WW (eds.) *Light-Harvesting Antennas*, pp. 195–217. Dordrecht: Kluwer.
- Blankenship RE, Olson JM, and Miller M (1995) Antenna complexes from green photosynthetic bacteria. In: Blankenship RE, Madigan MT, and Bauer CE (eds.) *Anoxygenic Photosynthetic Bacteria*, pp. 399–435. Dordrecht: Kluwer.
- Brune DC, Gerola PD, and Olson JM (1990) Circular dichroism of green bacterial chromosomes. *Photosynthesis Research* 24: 253–263.
- Ke B (2000) The green bacteria: I. The light-harvesting complex, the chlorosomes. In: *Photosynthesis: Photochemistry and Photobiophysics*, pp. 147–158. Dordrecht: Kluwer.
- Oelze J and Golecki JR (1995) Membranes and chlorosomes of green bacteria. In: Blankenship RE, Madigan MT, and Bauer CE (eds.) *Anoxygenic Photosynthetic Bacteria*, pp. 259–278. Dordrecht: Kluwer.
- Olson JM (1998) Chlorophyll organization and function in green photosynthetic bacteria. *Photochemistry and Photobiology* 67: 61–75.
- Prokhorenko VI, Steensgaard DB, and Holzwarth AR (2000) Exciton dynamics in the chlorosomal antennae of the green bacteria *Chloroflexus aurantiacus* and *Chlorobium tepidum*. *Biophysical Journal* 79: 2105–2120.

Gs Family of Heterotrimeric G Proteins

L S Weinstein, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA

Published by Elsevier Inc.

Glossary

Albright hereditary osteodystrophy (AHO)

A congenital syndrome due to heterozygous $G_{s\alpha}$ mutations characterized by short stature, neurocognitive defects, subcutaneous ossifications, and brachydactyly.

Genomic imprinting An epigenetic phenomenon associated with differential DNA methylation of the two parental alleles of a gene leading to partial or complete silencing of one of the parental alleles.

Heterotrimeric G protein A protein complex consisting of an α -, β -, and γ -subunit, each the product of separate genes, which transduces signals from cell surface receptors to generate intracellular second messengers.

McCune–Albright syndrome A sporadic condition caused by widespread distribution of a somatic activating $G_{s\alpha}$ mutation.

Pseudohypoparathyroidism Conditions characterized by parathyroid hormone resistance in the renal proximal tubule.

Introduction

Heterotrimeric G proteins are primarily involved in signaling at the plasma membrane and other intracellular sites. G proteins are generally defined by their specific α -subunit, which binds guanine nucleotide and mediates signals from seven transmembrane receptors to downstream effectors. The Gs family of heterotrimeric G proteins include two closely related α -subunits that couple receptors to adenylyl cyclase and generation of cyclic adenosine monophosphate (cAMP): $G_{s\alpha}$ (stimulatory) encoded by the *GNAS* gene at 20q13 and $G_{olf\alpha}$ (olfactory) encoded by *GNAL* at 18p11. $G_{s\alpha}$ is ubiquitously expressed and is responsible for receptor-mediated cAMP responses in most tissues while $G_{olf\alpha}$ has a much more restricted tissue distribution, including the olfactory epithelium and several other neuroendocrine tissues. This article primarily focuses on $G_{s\alpha}$ biochemistry and genetics and summarizes the clinical disorders associated with $G_{s\alpha}$ defects.

General Organization of the *GNAS* Gene

The human and mouse $G_{s\alpha}$ genes (*GNAS* and *Gnas*, respectively) are located within syntenic regions (20q13.2–13.3 and distal chromosome 2, respectively) and have similar overall structure. Both orthologs also undergo similar patterns of genomic imprinting, an epigenetic phenomenon associated with allele-specific differences in DNA methylation, which results in partial or complete suppression of gene expression from one parental allele. *GNAS/Gnas* generates multiple gene products through the use of alternative promoters and first exons that splice onto a common exon (exon 2, [Figure 1](#)). The most downstream alternative exon (exon 1) generates transcripts encoding $G_{s\alpha}$. The $G_{s\alpha}$ coding region spans 13 exons (12 exons in mice). Two long and two short forms of $G_{s\alpha}$ protein are generated by alternative splicing of exon 3. The use of an alternative terminal exon (N1) located between $G_{s\alpha}$ exons 3 and 4 generates a neural-specific splice variant of unknown function. Although the $G_{s\alpha}$ promoter is not methylated, $G_{s\alpha}$ is imprinted in a tissue-specific manner, being expressed

primarily from the maternal allele in several tissues, including pituitary, thyroid, renal proximal tubules, and gonads, but biallelically expressed in most other tissues.

The most upstream alternative promoter (*Nesp* located ~45 kb upstream of $G_{s\alpha}$ exon 1 in mouse; [Figure 1](#)) generates transcripts that encode the neuroendocrine-specific protein of 55 kDa (NESP55), a chromogranin-like protein totally unrelated to $G_{s\alpha}$. NESP55 is only transcribed from the maternal allele as its promoter region is methylated on the paternal allele. Another alternative promoter (*Gnasxl* located ~30 kb upstream of $G_{s\alpha}$ exon 1 in mouse, [Figure 1](#)) generates transcripts encoding the $G_{s\alpha}$ isoform XLzs. The specific XLzs first exon encodes a long amino-terminal extension while the remainder of the protein is identical to $G_{s\alpha}$ and encoded by $G_{s\alpha}$ exons 2–13. A neural-specific truncated form of XLzs (XLN1, presumably nonfunctional) is generated by alternative splicing to the alternative terminal exon N1. In contrast to NESP55, XLzs is only expressed from the paternal allele as its promoter is methylated on the maternal allele. The same differentially methylated region also generates antisense transcripts (NESPAS) that are important for establishing imprinting at the NESP55 promoter.

Just upstream of the $G_{s\alpha}$ promoter is an alternative promoter and first exon (exon 1A, also referred to as exon A/B in humans; [Figure 1](#)) that is methylated on the maternal allele and that generates untranslated transcripts from the paternal allele. The exon 1A region is also a primary imprint control center that is required for tissue-specific imprinting of $G_{s\alpha}$, and loss of exon 1A imprinting (DNA methylation) is the underlying molecular mechanism for the development of pseudohypoparathyroidism type 1B (PHP1B).

$G_{s\alpha}$ Isoforms Encoded by *GNAS*

$G_{s\alpha}$

$G_{s\alpha}$ is ubiquitously expressed and is targeted to the inner leaflet of both plasma and intracellular membranes by palmitoylation at its amino terminus. Its major function is to couple seven transmembrane receptors to adenylyl cyclase and mediate

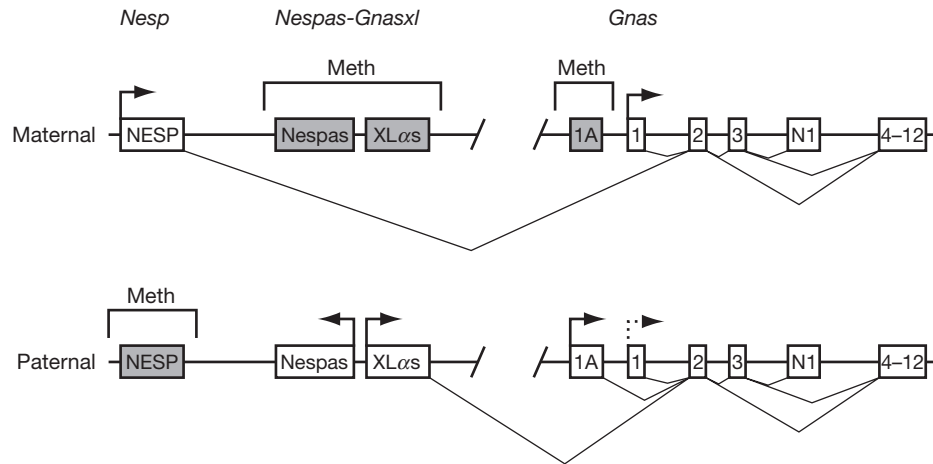


Figure 1 Organization of the *Gnas* locus. Schematic drawing of the *Gnas* maternal (top) and paternal (bottom) alleles with methylated regions shown above and splicing patterns shown below each allele. First exons for transcripts encoding NESP55 (NESP), Nespas, XL α s, and G α s (exon 1) are shown along with downstream exons 3–12 (13 exons in human *GNAS*) and the alternative neural-specific terminal exon N1 (exons 4–12 are shown as a single box). The mouse regions designated *Nesp*, *Nespas-Gnasxl*, and *Gnas* are indicated at the top. Horizontal arrows show active promoters while shaded boxes indicate inactive promoters. The dashed horizontal line at the paternal G α s promoter indicates that transcription is suppressed due to imprinting in a tissue-specific manner. For clarity, only the first exon for Nespas is shown. The figure is not drawn to scale. Reproduced from Weinstein LS, Xie T, Qasem A, Wang J, and Chen M (2010) The role of GNAS and other imprinted genes in the development of obesity. *International Journal of Obesity (London)* 34: 6–17.

receptor-stimulated cAMP generation. However there is evidence that G α s also directly activates other effectors, including cardiac Ca $^{2+}$ channels and Src tyrosine kinases and may be activated by the epidermal growth factor and basic fibroblast growth factor receptors. G α s may also play a role in membrane trafficking. cAMP, the major second messenger of G α s signaling, mediates its effects through direct binding and stimulation of protein kinase A, cAMP-regulated guanine nucleotide exchange factors (cAMP-GEFs), and ion channels. Protein kinase A is a serine/threonine kinase which phosphorylates many proteins, including enzymes involved in intermediary metabolism and transcription factors such as the cAMP response element binding protein (CREB). cAMP-GEFs are expressed at high levels in neuroendocrine cells and lead to activation of the ras family member Rap1.

Like all G α -subunits, prior to activation G α s has guanosine diphosphate (GDP) bound to its guanine nucleotide binding pocket and associates with $\beta\gamma$ dimers to form a heterotrimer. Upon interaction with an activated receptor, GDP is released and is replaced by ambient guanosine triphosphate (GTP), allowing G α s to achieve an active conformation leading to depalmitoylation, release from $\beta\gamma$ and the plasma membrane, and activation of adenylyl cyclase. Deactivation occurs via an intrinsic guanosine triphosphatase (GTPase) activity within G α s that hydrolyzes bound GTP to GDP.

G α s has two structural domains: a Ras-like GTPase domain, which includes sites for guanine nucleotide binding and effector interaction, and a helical domain which forms a lid over the guanine nucleotide binding site. Interactions between the two domains may be important to prevent constitutive guanine nucleotide release. Alternative splicing of exon 3 leads to two long and two short forms of G α s with helical domains of differing length. Although the ratio of long to short forms of G α s varies between cell types, all of these forms are biologically active with only subtle biochemical differences between them.

GTP binding alters the conformation of three regions (switch 1, 2, and 3 regions) within the GTPase domain allowing interactions between switch 2 and 3 and between switch 3 and the helical domain that are important to maintain the active conformation. Two highly conserved residues (Arg 201 and Gln 227 in the long form of G α s) are critical for the intrinsic GTPase catalytic activity, and therefore adenosine diphosphate (ADP)-ribosylation of Arg 201 by cholera toxin or mutations encoding substitutions of Arg 201 or Gln 227 lead to constitutive activation by disrupting the GTPase turn-off mechanism. The amino terminus and switch 2 regions directly interact with β -subunit within the heterotrimer, while the carboxyl terminus is important for receptor interactions.

XL α s

In XL α s, the first 47 amino acids of G α s are replaced by a large amino terminal domain with multiple repeat regions (the XL domain) encoded by the XL α s-specific first exon which targets XL α s to the plasma membrane and possibly to the Golgi. XL α s has been shown in cultured cells to be able to bind to $\beta\gamma$ and mediate receptor-stimulated cAMP production with the same efficacy and potency as G α s. XL α s knockout mice die soon after birth and the small number of mice that survive have increased metabolic rate and sympathetic nerve activity (SNA) as well as improved glucose tolerance and insulin sensitivity. However, it remains unclear whether XL α s has an important physiological function in humans, as patients with paternal *GNAS* mutations predicted to disrupt XL α s expression do not develop a specific phenotype.

Unlike G α s, XL α s is only expressed from the paternal *GNAS* allele and its expression is primarily limited in neuroendocrine tissues. In rats, XL α s is present in the melanotrophs within the pars intermedia of the posterior lobe, and to a lesser extent in brain, adrenal, heart, and pancreatic islets. In mouse neonates,

XLzs was shown to be expressed in brown and white adipose tissue (BAT, WAT), adrenal medulla, intermediate lobe of the pituitary, brain (hypothalamus, locus ceruleus, and the laterodorsal tegmental, hypoglossal, trigeminal, ambiguous, medullary reticular, raphe obscurus, and facial nuclei), stomach, heart, pancreas, endochondral growth plates, and kidney. XLzs expression in mouse adipose tissue is lost during the first two weeks of postnatal development.

Large amounts of a truncated XLzs (XLN1) are produced in the central nervous system by splicing to the alternative terminal exon N1, which may potentially act as a dominant negative regulator of $G_s\alpha$ signaling. It has been reported that the first XLzs exon has an alternate more upstream translational start site producing a longer product called XXLzs, which also is a functional $G_s\alpha$ isoform. XLzs exon 1 may also use an alternate reading frame to generate an unrelated basic protein named Alex, which may bind to XLzs. One study suggested that polymorphisms within XLzs exon 1 generate alternative forms of Alex and XLzs that fail to interact.

G $_s\alpha$ Defects in Human Disease

Posttranslational Modifications

From a public health standpoint, the most significant $G_s\alpha$ defect is a posttranslational ADP-ribosylation of $G_s\alpha$ residue Arg²⁰¹ catalyzed by cholera toxin, an exotoxin released by *Vibrio cholerae* during intestinal cholera infection. As Arg²⁰¹ is critical to the intrinsic GTPase turn-off mechanism, this leads to constitutive $G_s\alpha$ activity and excess cAMP in intestinal lining cells, resulting in the severe secretory diarrhea characteristic of intestinal cholera.

Activating G $_s\alpha$ Mutations

Missense mutations leading to substitution of Arg²⁰¹, or less often Gln²²⁷, lead to constitutive activation as these residues are necessary for catalyzing the GTPase turn-off reaction and therefore $G_s\alpha$ remains in the GTP-bound active form for a very extended period relative to wild-type $G_s\alpha$. As these mutations are lethal in the germline, they are always found as somatic mutations and are never transmitted to the next generation. These residues are highly conserved among $G\alpha$ subunits, and mutation of these residues leads to constitutive activation in all G proteins.

Activating $G_s\alpha$ mutations were first identified in a subset of growth hormone-secreting (somatotroph) pituitary tumors in which basal adenylyl cyclase activity was high even in the absence of stimulating hormones, such as growth hormone-releasing hormone. This subset comprises about a third of patients with growth hormone-secreting tumors (a condition known as acromegaly). Because of its ability to form tumors, the activated $G_s\alpha$ allele has also been referred to as the *gsp* oncogene. In almost all *gsp*-bearing acromegalic tumors, the activating mutation is on the maternal allele due to the fact that $G_s\alpha$ is imprinted in the pituitary somatotroph and therefore it is normally only expressed from the maternal allele in these cells. Subsequently, similar activating $G_s\alpha$ mutations were identified in a small subset of other endocrine tumors (thyroid and other tumors) and nonendocrine tumors

(e.g., ovarian tumors and myxomas) and in premature thelarche (breast development).

Similar activating $G_s\alpha$ mutations involving Arg²⁰¹ with a more widespread tissue distribution lead to McCune–Albright syndrome (MAS). Classically MAS was defined as the triad of sexual precocity (primarily in females), café-au-lait skin lesions, and polyostotic fibrous dysplasia of bone. It is now clear that these patients may also have many other features, both endocrine (adrenal adenoma and/or hyperplasia with hypercortisolism, thyroid nodules with hyperthyroidism, pituitary tumors with acromegaly, and excess prolactin secretion) and nonendocrine (e.g., biliary liver disease, cardiomyopathy with rare occurrence of sudden death, duodenal polyps, and intramuscular myxomas).

Consistent with constitutive activation of $G_s\alpha$ /cAMP pathways, the hormonal abnormalities result from overgrowth of and hypersecretion from peripheral endocrine organs independent of their respective pituitary stimulatory hormones, which mediate their effects by stimulating $G_s\alpha$ /cAMP. In fact, the pituitary hormones (thyrotropin, adrenocorticotrophic hormone, and gonadotropins) are suppressed due to negative feedback from high levels of thyroid hormone, cortisol, and estrogens or androgens, respectively. Hyperpigmentation of café-au-lait skin lesion is due to constitutive stimulation of melanin production in melanocytes via a pathway normally stimulated by melanocyte-stimulating hormone. Fibrous dysplasia of bone is characterized by focal fibrous bone lesions originating from the central marrow cavity and primarily affecting the head and axial skeleton. These lesions arise from the effect of excess cAMP in preosteoblastic mesenchymal precursor cells to inhibit differentiation and increase proliferation.

Based on the fact that MAS patients are widespread mosaics and the distribution of the café-au-lait lesions along developmental dermatomes, it is presumed that this disorder is the result of a somatic $G_s\alpha$ mutation occurring during early post-fertilization development leading to widespread distribution of cells bearing activating $G_s\alpha$ mutations. Therefore, the specific clinical presentation of each individual case is dependent upon the original site of the mutation and how late in development the mutation occurred. In fact, patients may have only 1–2 features of the classic triad, such as patients with polyostotic or monostotic fibrous dysplasia or a single endocrinopathy.

Inactivating G $_s\alpha$ Mutations

Heterozygous inactivating $G_s\alpha$ mutations lead to Albright hereditary osteodystrophy (AHO), an autosomal dominant congenital syndrome characterized by short stature, brachydactyly (shortening and widening of bones in hands and feet including metacarpals and metatarsals), neurocognitive abnormalities, subcutaneous ossifications, and characteristic facial features. Patients who inherit the mutation from the mother or have a *de novo* mutation on the maternal allele, in addition, develop early-onset obesity as well as resistance to some, but not all, hormones that activate $G_s\alpha$ in their respective target tissues (parathyroid hormone, thyrotropin, gonadotropins, and growth hormone-releasing hormone), which leads to hypocalcemia and hyperphosphatemia, hypothyroidism, reproductive abnormalities, and growth hormone deficiency, respectively. This is also referred to as pseudohypoparathyroidism type 1A (PHP1A).

Although about half of PHP1A patients have growth hormone deficiency, the short stature (as well as the brachydactyly) is due to rapid maturation and early closure of skeletal growth plates due to loss of responsiveness of growth plate chondrocytes to the paracrine factor parathyroid hormone-like peptide (PTHrP), which normally activates $G_{s\alpha}$. Patients with $G_{s\alpha}$ mutations on the paternal allele develop AHO alone in the absence of hormone resistance or obesity, a condition also referred to as pseudopseudohypoparathyroidism (PPHP). While the ectopic ossifications are generally focal and limited to the subcutaneous layer, in rare cases, more often with paternal $G_{s\alpha}$ mutations, the ossifications coalesce to form larger plate-like structures and invade the deeper soft tissues, a condition known as progressive osseous heteroplasia (POH). Overall, the AHO features that are common to both PHP1A and PPHP patients are presumed to be the result of $G_{s\alpha}$ haploinsufficiency resulting from a mutation on one parental allele (maternal in PHP1A or paternal in PPHP).

The presence or absence of hormone resistance in PHP1A versus PPHP, respectively, relates to the fact that $G_{s\alpha}$ is imprinted in the various hormone target tissues (kidney proximal tubule, thyroid, gonads, and pituitary somatotrophs) with $G_{s\alpha}$ being expressed primarily from the maternal allele in these tissues. Therefore, mutation of the active maternal allele will lead to $G_{s\alpha}$ deficiency and hormone resistance while mutation of the relatively inactive paternal allele will have little effect on $G_{s\alpha}$ expression or hormone sensitivity.

AHO is one of a small number of monogenic obesity disorders, with the caveat that obesity only occurs when the mutation is on the maternal allele (PHP1A). Mouse models have provided important insights into the likely cause of obesity in PHP1A. As in AHO patients, mice with maternal, but not paternal, germline $G_{s\alpha}$ mutation (deletion of exon 1) develop severe obesity and, in addition, develop diabetes, insulin resistance, and hyperlipidemia. In these mice, obesity is primarily due to reduction in SNA and energy expenditure rather than increased food intake. Although $G_{s\alpha}$ plays many roles in regulating peripheral metabolism in various tissues (e.g., mediating hormone-stimulated lipolysis in adipose tissue, stimulating glucose production in liver), mouse models knocking out $G_{s\alpha}$ in liver, adipose tissue, and muscle did not produce an obese phenotype. In addition, molecular studies did not show evidence for $G_{s\alpha}$ imprinting in these tissues.

Similar to what was observed with germline $G_{s\alpha}$ mutation, maternal deletion of $G_{s\alpha}$ exon 1 limited to the central nervous system (mBrGsKO) also resulted in severe obesity, diabetes, insulin resistance, and reduced SNA and energy expenditure. Central melanocortins promote negative energy balance (reduced food intake, increased SNA and energy expenditure) via receptors (MC3R, MC4R) that activate $G_{s\alpha}$ /cAMP pathways. Homozygous MC4R mutations are the most common cause of human monogenic obesity and MC4R knockout mice have a metabolic phenotype similar to that of mBrGsKO, except that MC4R mice have increased food intake as well as reduced energy expenditure. Acute administration of an MC4R agonist to mBrGsKO mice showed that these mice have reduced stimulation of energy expenditure in response to MC4R stimulation while the ability of the agonist to reduce food intake remained intact. In addition, there was molecular evidence for $G_{s\alpha}$ imprinting in the paraventricular nucleus of the hypothalamus (PVH), a site known to express MC4R receptors and to be

important in regulation of food intake and energy expenditure. It therefore appears that maternal $G_{s\alpha}$ mutations in mice (and likely in PHP1A patients as well) lead to $G_{s\alpha}$ deficiency in one or more brain regions (PVN and probably other sites) due to site-specific imprinting, and that this leads to obesity primarily via loss of melanocortin signaling to stimulate SNA and energy expenditure. While studies in XL α s knockout mice suggest that this isoform plays a role opposite to $G_{s\alpha}$ in metabolic regulation, its role, if any, in humans is unclear as PPHP patients with paternal $G_{s\alpha}$ mutations that should also disrupt XL α s do not present with a significant metabolic phenotype.

GNAS Imprinting Defects

PHP1B is a related disorder in which patients present with renal parathyroid hormone resistance in the absence of AHO features or other evidence of hormone resistance (except for borderline thyrotropin resistance in about half of the cases). Responsiveness of the skeleton to PTH is relatively maintained and therefore some PHP1B patients may present with skeletal changes resulting from chronically elevated levels of serum PTH. This condition may be familial or sporadic.

In all cases of PHP1B, methylation of the exon 1A region on the maternal allele is absent, indicating a defect in establishment or maintenance of the methylation imprint, and therefore at this site both parental alleles have a paternal (unmethylated) epigenotype. Most cases of familial PHP1B are also associated with a maternal deletion in the linked *STX16* gene located upstream of *GNAS* with no other *GNAS* methylation defects. In these families, abnormal methylation and renal PTH resistance only occur when the *STX16* deletion is inherited maternally, while those who inherit the deletion paternally are silent carriers. A small number of familial PHP1B cases are associated with maternal deletion of the NESP55 upstream region associated with loss of maternal methylation at both the XL α s and 1A promoter regions. Sporadic cases are often associated with more global methylation changes in the *GNAS* locus with no associated mutations. There is some evidence that patients with 1A methylation defects reminiscent of PHP1B may occasionally present with some features of AHO.

The absence of 1A methylation in PHP1B suggested a model by which this region contained a negative control element for the $G_{s\alpha}$ promoter that is both methylation-sensitive and tissue-specific. We proposed that the region contains a *cis*-acting suppressor element that can bind a tissue-specific repressor protein on the paternal allele, leading to suppressed $G_{s\alpha}$ expression from the paternal allele in tissues where the repressor is expressed (e.g., renal proximal tubules, thyroid, somatotrophs, etc.), but no suppression (and therefore biallelic expression) in most other tissues where the repressor is not expressed. On the maternal allele, the repressor cannot bind due to methylation and therefore the $G_{s\alpha}$ promoter remains active in the maternal allele in all tissues. In PHP1B, loss of 1A methylation allows the repressor to now bind to both alleles in specific tissues (e.g., renal proximal tubules) leading to $G_{s\alpha}$ deficiency and PTH resistance. Support for this model came from the observation that $G_{s\alpha}$ imprinting is reversed in mice with deletion of the 1A region on the paternal allele and in fact these mice show increased PTH sensitivity due to increased $G_{s\alpha}$ expression in proximal tubules.

Table 1 Phenotypes of $G_{s\alpha}$ knockout mouse models

Tissue or cell type	Phenotype
Germline (heterozygote)	Maternal – reduced survival, obesity, glucose intolerant, insulin resistant, and reduced energy expenditure Paternal – minimal phenotype
Brain (heterozygote)	Maternal – obesity, diabetes, insulin resistant, and reduced energy expenditure/sympathetic nerve activity Paternal – none
Liver (homozygote)	Hypoglycemia Increased insulin sensitivity and secretion Reduced fat mass Pancreatic α -cell hyperplasia and hyperglucagonemia
Skeletal muscle (homozygote)	Reduced muscle mass and oxidative capacity Switch to type 2 muscle fibers Improved fatigue resistance
Adipose tissue (homozygote)	Early lethality Impaired adipogenesis (white adipose tissue) Inactive brown adipose tissue and cold intolerance Increased diet-induced thermogenesis
Pancreatic islets (homozygote)	Severely reduced β cell mass and insulin-deficient diabetes Reduced body mass
Chondrocytes (homozygote)	Growth plate defects (accelerated chondrocyte proliferation) Ectopic chondrocyte formation
Osteoblasts (homozygote)	Reduced trabecular bone formation Increased cortical bone thickness Reduced B lymphocyte production
Hematopoietic stem cells (homozygote)	Impaired homing and engraftment to bone marrow
Juxtaglomerular granule (JG) cells (homozygote)	Impaired renin synthesis and release Renal failure

Conclusion

The G_s family of G proteins includes two forms, which primarily act by activating adenylyl cyclase, the ubiquitously expressed G_s and the more limited G_{olf} , both of which are defined by their specific α -subunits encoded by distinct genes. $G_{s\alpha}$ is encoded by a complex locus *GNAS* which also encodes a

paternal-specific $G_{s\alpha}$ isoform *XL α s*. $G_{s\alpha}$ defects, including post-translational modifications, activating and inactivating mutations, and abnormal imprinting, lead to many clinical disorders resulting from excess of or loss of $G_{s\alpha}$ signaling. In addition, a number of germline and tissue-specific knockouts of $G_{s\alpha}$ and *XL α s* (Table 1) have provided other important insights into the role of these signaling molecules in development, cell biology, and physiology.

See also: **Signaling:** Adenylyl Cyclases; G_{12}/G_{13} Family; Gi Family of Heterotrimeric G Proteins; G_q Family; Ras Family; Small GTPases.

Further Reading

- Chen M, Nemecek NM, Mema E, Wang J, and Weinstein LS (2011) Effects of deficiency of the G protein $G_{s\alpha}$ on energy and glucose homeostasis. *European Journal of Pharmacology* 660: 119–124.
- Chen M, Wang J, Dickerson KE, et al. (2009) Central nervous system imprinting of the G protein $G_{s\alpha}$ and its role in metabolic regulation. *Cell Metabolism* 9: 548–555.
- Hayward BE, Moran V, Strain L, and Bonthron DT (1998) Bidirectional imprinting of a single gene: *GNAS1* encodes maternally, paternally, and biallelically derived proteins. *Proceedings of the National Academy of Sciences of the United States of America* 95: 15475–15480.
- Jones DT, Masters SB, Bourne HR, and Reed RR (1990) Biochemical characterization of three stimulatory GTP-binding proteins. The large and small forms of Gs and the olfactory-specific G-protein, Golf. *Journal of Biological Chemistry* 265: 2671–2676.
- Jones DT and Reed RR (1989) Golf: An olfactory neuron-specific G protein involved in odorant signal transduction. *Science* 244: 790–795.
- Landis CA, Masters SB, Spada A, et al. (1989) GTPase inhibiting mutations activate the α chain of Gs and stimulate adenylyl cyclase in human pituitary tumours. *Nature* 340: 692–696.
- Liu J, Litman D, Rosenberg MJ, et al. (2000) A *GNAS1* imprinting defect in pseudohypoparathyroidism type 1B. *Journal of Clinical Investigation* 106: 1167–1174.
- Lyons J, Landis CA, Harsh G, et al. (1990) Two G protein oncogenes in human endocrine tumors. *Science* 249: 655–659.
- Shore EM, Ahn J, Jan de Beur S, et al. (2002) Paternally inherited inactivating mutations of the *GNAS1* gene in progressive osseous heteroplasia. *New England Journal of Medicine* 346: 99–106.
- Weinstein LS, Shenker A, Gejman PV, et al. (1991) Activating mutations of the stimulatory G protein in the McCune–Albright syndrome. *New England Journal of Medicine* 325: 1688–1695.
- Weinstein LS, Xie T, Qasem A, Wang J, and Chen M (2010) The role of *GNAS* and other imprinted genes in the development of obesity. *International Journal of Obesity (London)* 34: 6–17.
- Weinstein LS, Xie T, Zhang QH, and Chen M (2007) Studies of the regulation and function of the $G_{s\alpha}$ gene *Gnas* using gene targeting technology. *Pharmacology and Therapeutics* 115: 271–291.

Heat/Stress Responses

G Thibault, Temasek Life Sciences Laboratory, National University of Singapore, Singapore

D T W Ng, Temasek Life Sciences Laboratory and Department of Biological Sciences, National University of Singapore, Singapore

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Davis T.W. Ng, volume 2, pp. 343–347, © 2004, Elsevier Inc.

Glossary

Endoplasmic reticulum First organelle of the secretory pathway. All secreted and resident proteins of the pathway are first synthesized in this organelle.

Kinase Enzyme that transfers phosphate groups, usually from ATP, to substrates. On protein substrates, it is usually added to the hydroxyl group of serine, threonine, or tyrosine side chains.

Molecular chaperone These molecules function by preventing inappropriate interactions of other molecules. Chaperones play many cellular roles including protein

folding, protein translocation, signal transduction, and protein degradation.

Proteasome Large multisubunit protein complex that mostly degrades proteins modified by poly-ubiquitin.

Transcription factor Proteins that regulate the synthesis of messenger RNA at gene promoter elements.

Ubiquitin A highly conserved 76-amino-acid protein that functions through covalent conjugation of proteins (either a target or another ubiquitin molecule) via an isopeptide bond to lysyl side chains. Ubiquitin plays roles in protein degradation, regulation of protein activity, and protein localization.

The Stress of Unfolded Proteins

Proteins are the class of molecules that perform most cellular functions. First synthesized as chains of amino acids encoded by genes, nascent proteins must fold into their correct three-dimensional structures to be active. The folded state is labile and dependent on the thermodynamic nature of each molecule. Fully mature proteins are the product of unique amino acid sequences that specify numerous noncovalent interactions within each molecule. Abrupt changes can create unfavorable local environments that weaken such interactions or prevent their correct formation. Conditions that cause the unfolding or disrupt the proper synthesis of proteins are major forms of cellular stress. Left unchecked, the inherent toxicity of unfolded proteins or their loss of function could leave cells irreversibly damaged.

The Heat Shock Response

The first stress response to unfolded proteins was reported over 40 years ago. Examination of fruit fly salivary chromosomes revealed morphological puffs along their lengths after treatment by heat, dinitrophenol, or sodium salicylate. Termed the heat shock response (HSR), it was later shown that the puffs signified activation of genes encoding heat shock

proteins (Hsps). For virtually all organisms, exposure to 5–10 °C above the optimal growth temperature leads to a robust HSR. Within the cell, an increase of this magnitude can have many deleterious effects including increased membrane fluidity and the partial unfolding of proteins.

Inputs

The HSR can be activated by wide variety of conditions beyond temperature change. A few examples include exposure to ethanol, heavy metals, amino acid analogs, mechanical damage, and conditions of hypoxia, ischemia, and osmotic stress. As temperature increases, these conditions are believed to increase the concentration of unfolded proteins. In addition, the non-stress conditions of cells undergoing rapid proliferation or differentiation also induce the HSR. Here, the physiological need is to boost the capacity for protein synthesis.

Stress Detection and Response

The HSR relies on the activity of molecular chaperone proteins to detect stress. Chaperones bind to unfolded proteins and facilitate folding by preventing inappropriate interactions. Chaperones also bind misfolded proteins to neutralize their toxicity and promote their destruction. As stress conditions increase the concentration of unfolded proteins, the available

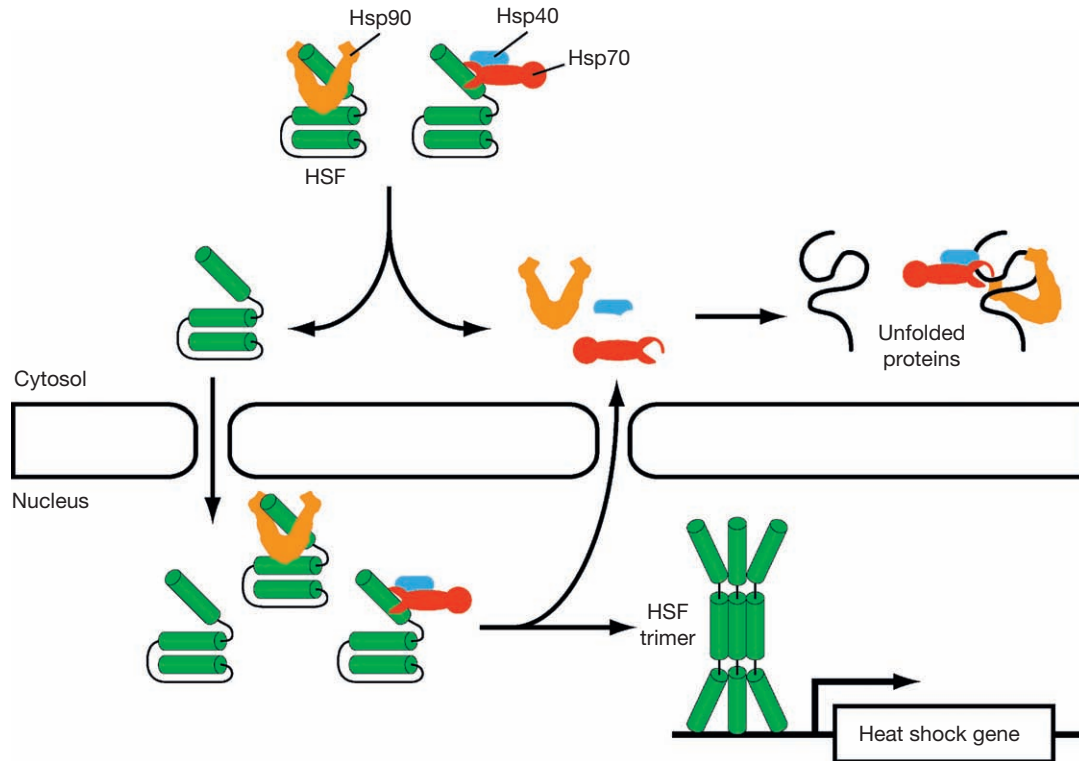


Figure 1 Activation of the heat shock response is a multistep mechanism. The increase of unfolded proteins stimulates the binding of cytosolic chaperones and reduces the free chaperone pool. The reduction in chaperone binding to HSF monomers promotes assembly into HSF trimers (and nuclear localization of cytosolic HSF). Initial binding of trimeric HSF to heat shock promoters is in the inert state. Transcriptional activation of the target gene requires HSF phosphorylation. In yeast, HSF is similarly regulated except that it always exists as a trimer.

pool of free chaperones diminishes. Specialized sensor proteins can detect the change and initiate a response. In the eukaryotic HSR (prokaryotes employ a different mechanism), the heat shock factor (HSF) plays both stress sensor and transcription activator for the response output. In metazoan organisms, the inert state of HSF is a monomer that is maintained through associations with the chaperones Hsp70, Hsp90, and Hsp40 (Figure 1). As the demand for chaperones increases, release from chaperone complexes permits conversion of HSF to a homotrimer (and translocation into the nucleus if found in the cytosol). This form of HSF can bind to heat shock elements of heat shock genes but does so in the inert state. Transcriptional activation requires HSF phosphorylation and a conformational change. Subsequently, a dramatic but transient rise in the synthesis of Hsps is accompanied by a reduction in overall protein synthesis.

The HSR Output

Protein repair and preservation

The most notable output of the HSR is the induction of chaperone genes. Chaperones are well adapted to resist stress. Beyond their established role in the folding of most newly synthesized proteins, direct biochemical analyses have shown that some chaperones disaggregate and refold misfolded proteins. Organisms harboring mutations in chaperone genes are less tolerant to various forms of stress. Taken together, the evidence supports the view that chaperones protect cells from

stress by maintaining the folded states of proteins and repairing those that are damaged.

Degradation of aberrant proteins

Since they can be toxic, damaged proteins that cannot be repaired must to be destroyed. Fittingly, the HSR regulates key factors of protein degradative pathways. For example, another facet of chaperone function is their requirement in the degradation of misfolded proteins. In eukaryotes, offending proteins are conjugated with ubiquitin, a highly conserved 76-amino-acid protein. Poly-ubiquitin chains serve as signals for the 26 S proteasome, which degrades the attached protein. The factors responsible for the recognition and modification of substrates belong to two large enzyme families that include the ubiquitin conjugating enzymes (E2s) and the ubiquitin ligases (E3). The role of protein degradation pathways in stress tolerance is well established. For example, yeast strains that lack two heat shock genes, *UBC4* and *UBC5* (E2s), are hypersensitive to elevated temperatures and other forms of stress. A major role of the HSR in stress tolerance is to increase the cellular capacity to rapidly degrade damaged proteins.

The environmental stress response

The development of DNA microarray technology has made analysis of whole genome expression patterns possible. The application of this technology revealed that the transcriptional

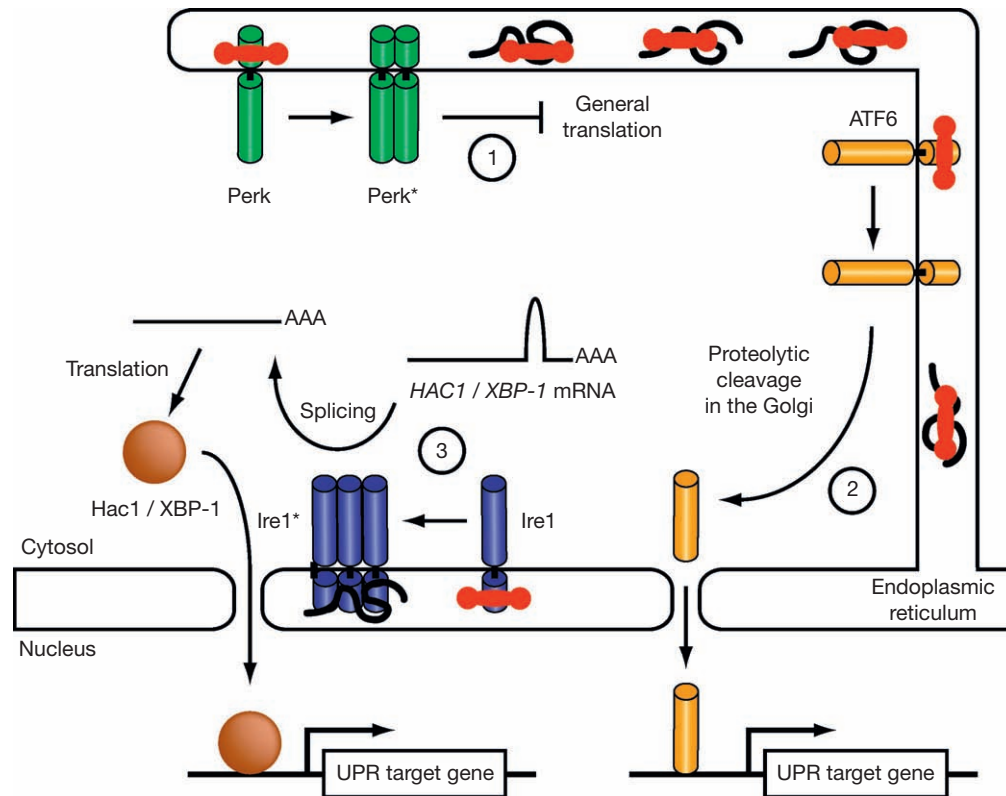


Figure 2 Activation of the unfolded protein response. The accumulation of unfolded proteins in the ER reduces the free pool of BiP and activates three distinct UPR sensor/transducers. (1) PERK dimers form following the reduction of BiP (shown in red) bound to their luminal domains. Activated PERK phosphorylates eIF2 α resulting in the temporary attenuation of general translation. (2) The reduction of BiP binding to ATF6 luminal domains stimulates their relocalization to the Golgi apparatus where the transcription factor is released by intramembrane proteolysis. Soluble ATF6 translocates to the nucleus where it activates expression of a subset of UPR target genes. (3) Ire1 is activated by BiP release and direct binding of unfolded proteins which promotes its oligomerization and trans-autophosphorylation. The activated ribonuclease domain cleaves introns of messages encoding UPR-specific transcription factors. Newly synthesized Hac1 (yeast) or XBP-1 (metazoans) activates the transcription of UPR target genes. In yeast, mechanism 3 is the only known UPR pathway. Asterisks indicate activated forms of enzymes.

output of the HSR is remarkably broad. Although the HSR has its own molecular signature, a large subset of its genes is regulated similarly among disparate environmental stress conditions. These genes, comprising the environmental stress response (ESR), represent many stress-related functions and a variety of metabolic pathways. Thus, any specific stress response can be thought of as comprising the ESR plus additional outputs tailored to their specific circumstances.

The Unfolded Protein Response

The HSR primarily monitors the state of the cytosol and nucleoplasm. For the endomembrane system of the secretory pathway, the health of proteins in the endoplasmic reticulum (ER) is monitored by the unfolded protein response (UPR). Circumstances leading to ER disequilibrium and activation of the UPR include exposure to reducing agents, incomplete protein glycosylation, glucose deprivation, excess nitrogen, calcium depletion, and proteasome dysfunction. The UPR is a signal transduction pathway that relays information on the condition of the ER lumen to regulatory networks of the nucleus and cytosol.

Stress Sensors and Transmembrane Signaling

As with the HSR, increasing unfolded proteins in the ER reduces the free pool of luminal chaperones. However, since these chaperones are physically separated from downstream regulatory mechanisms, the UPR employs strategies distinct from the HSR to monitor stress. The key sensor, found in all eukaryotes, is Ire1 (**Figure 2**). Ire1 is a transmembrane protein with three functional domains. Binding of the ER chaperone BiP to the luminal domain keeps Ire1 in the inactive monomeric state. Upon stress, the luminal region of Ire1 senses the accumulation of unfolded proteins by binding them directly, which drives its oligomerization. Oligomerization stimulates trans-autophosphorylation resulting in the activation of its RNase activity. The RNase, located on the cytosolic domain of Ire1, unconventionally splices messenger RNAs encoding UPR-specific transcription factors. In budding yeast, splicing enables *HAC1* mRNA to be translated. In metazoans, the Ire1 target is *XBP-1*. Spliced *XBP-1* directs the expression of a functional UPR transcription factor that is larger and more stable from the version synthesized from the unspliced transcript. Metazoan organisms also have additional levels of UPR regulation. A membrane-bound transcription factor, ATF6 (α and β

isoforms), is released by regulated intramembrane proteolysis upon ER stress. Soluble ATF6 then translocates into the nucleus to activate UPR target genes. A third sensor is the ER transmembrane protein PERK. Like Ire1, PERK contains a kinase that is regulated by BiP occupancy of its luminal domain. By contrast, PERK targets the translation factor eIF2 α under ER stress. Phosphorylation of eIF2 α attenuates general translation.

Outputs of the UPR

The transcriptional program of the UPR shows functional similarities to the HSR but it is customized for the secretory pathway. Cells lacking an intact UPR are hypersensitive to ER stress. Under stress, such cells fail to properly fold and modify secretory proteins, trafficking functions are shut down, and misfolded proteins accumulate unabated. Accordingly, the UPR program is complex with nearly half the regulated genes having functions throughout the secretory pathway. Similar to the HSR, the deployment of chaperone proteins preserves the integrity of existing proteins, aids the folding of new proteins, and promotes the degradation of misfolded proteins. Misfolded secretory and membrane proteins are degraded by a specialized arm of ubiquitin/proteasome pathway called ER-associated protein degradation (ERAD). The ERAD pathway serves to recognize, translocate, and ubiquitylate proteins for degradation. Studies have shown that all levels of the ERAD pathway are regulated by the UPR and this regulation plays an essential role in ER stress tolerance. In metazoans, the three ER sensors appear to act coordinately to resist stress. The activity of PERK is the first line of defense. Activation of ATF6 is similarly rapid and precedes the mobilization of XBP-1. The available data indicate that key chaperone genes are ATF6 regulated while key genes of the ERAD pathway are regulated by XBP-1. Thus, a model has emerged (Figure 2) that has PERK rapidly attenuating general translation to decrease the load of unfolded proteins in the ER. This gives the existing machinery in the ER an opportunity to begin restoring homeostasis. Those efforts

are augmented by the increased synthesis of chaperone proteins directed by ATF6. If stress persists, the ERAD pathway is upregulated by the activity of XBP-1. This serves to clear those proteins that cannot be refolded in the initial response. In cells experiencing stress that cannot be ameliorated, ER-associated apoptotic pathways are activated to sacrifice afflicted cells. The phenomena may contribute to a number of prevalent diseases, including diabetes, neurodegenerative diseases, atherosclerosis, and renal disease.

See also: [Cell Architecture and Function: Endoplasmic Reticulum-Associated Protein Degradation; Unfolded Protein Responses; Lipids Carbohydrates Membranes and Membrane Proteins: Secretory Pathway; Protein/Enzyme Structure Function and Degradation: Proteasomes, Overview; Protein Degradation; Protein Folding and Assembly; Ubiquitin System.](#)

Further Reading

- Akerfält M, Morimoto RI, and Sistonen L (2010) Heat shock factors: Integrators of cells stress, development and lifespan. *Nature Reviews Molecular Cell Biology* 11: 545–555.
- Buchberger A, Bukau B, and Sommer T (2010) Protein quality control and the endoplasmic reticulum: Brothers in arms. *Molecular Cell* 40: 238–252.
- Hotamisligil GS (2010) Endoplasmic reticulum stress and the inflammatory basis of metabolic disease. *Cell* 140: 900–917.
- Jarriel-Encontre I, Bossis G, and Pechaczyk M (2008) Ubiquitin-independent degradation of proteins by the proteasome. *Biochimica et Biophysica Acta* 1786: 153–177.
- Ron D and Walter P (2007) Signal integration in the endoplasmic reticulum unfolded protein response. *Nature Reviews Molecular Cell Biology* 8: 519–529.
- Tabas I and Ron D (2011) Integrating the mechanisms of apoptosis induced by endoplasmic reticulum stress. *Nature Cell Biology* 13: 184–190.
- Vembar SS and Brodsky JL (2008) One step at a time: Endoplasmic reticulum-associated degradation. *Nature Reviews Molecular Cell Biology* 9: 944–957.
- Xie W and Ng DTW (2010) ERAD substrate recognition in budding yeast. *Seminars in Cell and Developmental Biology* 21: 533–539.

Hematopoietin Receptors

B A Miller, Pennsylvania State University College of Medicine, Hershey, PA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Barbara A. Miller and Joseph Y. Cheung, volume 2, pp. 348–353, © 2004, Elsevier Inc.

Glossary

Colony stimulating factor (CSF) CSF stimulates hematopoietic progenitor/precursor cells. The prefix denotes the hematopoietic lineage stimulated by the factor, that is, M-CSF denotes macrophage or monocyte colony-stimulating activity; G-CSF stimulates granulocyte colony formation; and GM-CSF stimulates granulocyte and macrophage/monocyte colony formation.

Cytokine An extracellular signaling peptide or protein that mediates cell–cell communication, activating a

signaling cascade after interaction with a cell-surface receptor.

Hematopoiesis The process of development of mature blood cells from stem cells or precursors cells committed to blood cell formation.

Interleukin (IL) One of a number of secreted peptides or proteins, produced by lymphocytes or monocytes, which are involved in regulation of the immune system or leukocyte interactions. The number behind IL is the designation of that particular interleukin (i.e., IL-3).

The hematopoietic system consists of multipotential hematopoietic stem and progenitor cells that express hematopoietin (cytokine) receptors. These cells selectively lose or gain growth-factor receptors, which is key in the production of mature cells of specific lineages. Hematopoietin receptors comprise a superfamily of glycoproteins with structural similarities including single transmembrane spanning domains. They are activated by the binding of specific hematopoietic growth factors or cytokines, initiating a signaling cascade that results in responses including cell commitment, stimulation of proliferation or differentiation, enhancement of survival, and/or functional activation. Conserved domains in the extracellular region include cysteine pairs and a tryptophan–serine–x–tryptophan–serine (WSXWS) motif for class I receptors (see below), which are important in ligand binding. Cytoplasmic domains are diverse but have some limited similarities in membrane proximal regions known as box 1 and box 2. Box 1 and box 2 are essential for transducing mitogenic signals, and for association with and activation by Janus family (JAK) kinases. No intrinsic kinase or enzyme motif has been identified in the cytoplasmic domains of hematopoietin-receptor superfamily members. Instead, signal transduction through these receptors is regulated by their interaction with kinases of the JAK family, other cytosolic kinases, and protein messengers. Cytokine receptor signaling for all of these factors is initiated by ligand binding to the receptor, resulting in receptor oligomerization; some receptors form homodimers while others are active as hetero-oligomers. Following receptor oligomerization, the receptor-associated JAKs are phosphorylated, typically through transphosphorylation, resulting in their activation. JAK kinases subsequently phosphorylate tyrosines on the associated receptor, which then serve as docking sites for other signal transducers including the signal transducer and activator of transcription (STAT) proteins, Src kinases, protein phosphatases, and other proteins including Shc, Grb2, and phosphatidylinositol 3-kinase (PI 3-kinase), initiating a signaling cascade. Activation of hematopoietin receptors regulates responses through both transcription-dependent

and -independent pathways. Negative regulators of hematopoietin receptor signaling include tyrosine phosphatases that dephosphorylate the receptor and/or JAKs, and proteins which are members of the suppressors of cytokine-signaling family (SOCS). Hematopoietin receptor family members share many intracellular signaling molecules. The physiological function of hematopoietin receptors is complex. Some hematopoietin receptors are expressed on very early hematopoietic progenitors resulting in multilineage stimulation; others are expressed on a restricted subset or terminally differentiated cells. Many cells express more than one receptor and respond to more than one cytokine. The precise mechanisms responsible for signaling specificity remain uncertain. The *in vivo* importance of many hematopoietic receptors has been unequivocally demonstrated through deletion or over expression of the receptor or its agonist in mice, and the use of recombinant receptor agonists in humans.

Classification

Members of the hematopoietin receptor superfamily, also referred to as the cytokine receptor superfamily, can be separated into class I and class II receptors based on structural features. Class I includes: (1) receptors consisting of a single subunit which forms homodimers including those for erythropoietin (Epo), thrombopoietin (Tpo), granulocyte colony stimulating factor (G-CSF), growth hormone, and prolactin; (2) receptors consisting of two subunits, a unique α subunit and a common β subunit including those of the IL-3, GM-CSF, and IL-5 subfamily that share β_c , and the IL-6 subfamily that share the β chain gp130; and (3) heteromeric receptors for several interleukins which share the common gamma chain (γ_c) and consist of two to three subunits, α , β , and γ . A schematic representation of class I hematopoietin receptors is presented in [Figure 1](#). Class II consists of receptors for the interferons (IFNs) and certain interleukins (IL-10, IL-22, IL-24, IL-28, and IL-29). Soluble forms of cytokine receptors are formed by limited proteolysis of membrane-bound

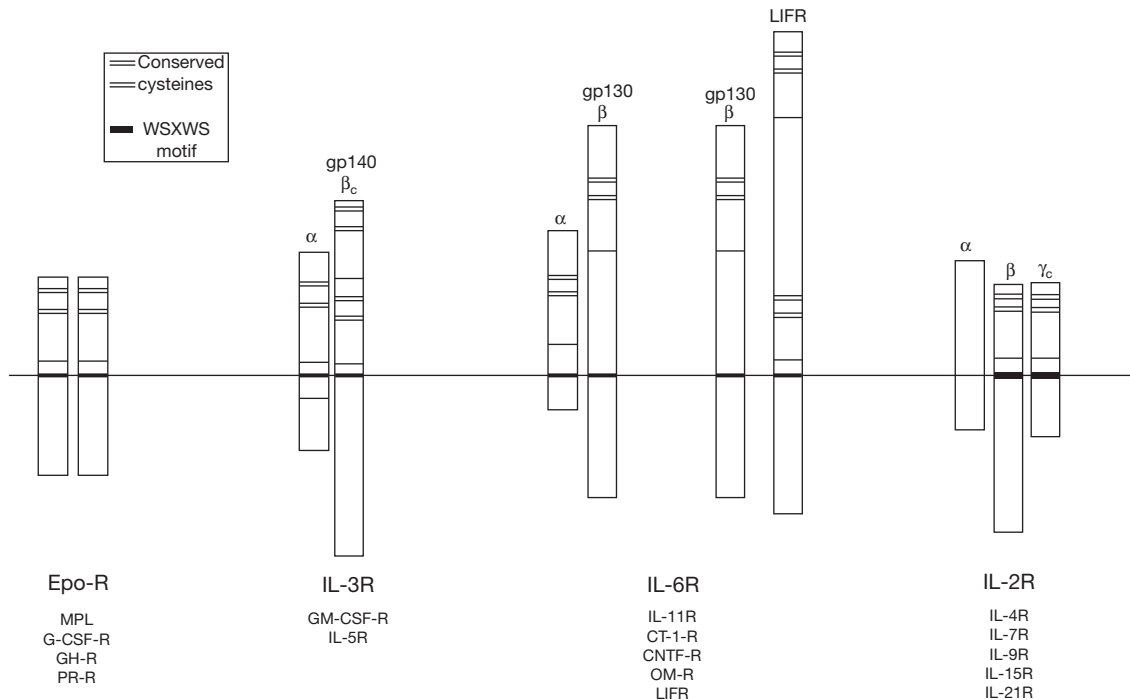


Figure 1 Schematic representation of class I hematopoietin receptor families. The single-chain receptor (receptors for erythropoietin, thrombopoietin, G-CSF, growth hormone, and prolactin) and multiple-chain IL-3, IL-6, and IL-2 receptor families are shown.

receptors or alternative messenger RNA (mRNA) splicing. These can form a complex with specific cytokines and act antagonistically or agonistically depending on the complex involved. A second group of receptors through which hematopoietic growth factors signal, unrelated to the hematopoietin receptor family, is called the receptor tyrosine kinase gene family. This group of receptors has intrinsic tyrosine kinase activity. Examples include c-kit, the receptor for stem cell factor, which acts synergistically in culture with many other hematopoietic growth factors, and the M-CSF receptor, which is critical for monocyte/macrophage growth. This family is not discussed here. Representative members of the hematopoietin receptor family are presented as model systems for signaling of other members.

Single Subunit Hematopoietin Receptors

Epo Receptor

Expression

Epo is the primary regulator of erythropoiesis and promotes the survival, proliferation, and differentiation of erythroid cells. Epo is synthesized primarily in the kidney, and also in the liver, and its level is regulated by the arterial oxygen tension. It mediates its effects through interaction with its receptor, Epo-R, a 66-kDa protein expressed on the surface of progenitor and precursor cells committed to the erythroid lineage. The essential function of Epo-R is demonstrated by the severe anemia and death at 11–15 days gestation following homozygous deletion of the Epo or Epo-R genes in mice. The earliest cell in the erythroid lineage, the burst-forming unit erythroid, has only low numbers of Epo-R. Epo-R expression peaks at the later erythroid progenitor (colony-forming unit

erythroid) and proerythroblast stage, followed by a decline during maturation to undetectable levels in mature erythroid cells. Epo-R is also expressed on megakaryocytes and several nonhematopoietic cells including endothelial cells, myoblasts, and neuronal cells. In these nonhematopoietic cells, the biological significance of receptor expression has not been clarified. Concerns about cardiovascular and thromboembolic events, as well as promotion of tumor growth when Epo is used in cancer patients have resulted in modified use guidelines.

Signaling pathways activated by Epo

Epo-R consists of a single subunit which forms a homodimer in the presence or absence of Epo. Other members of the hematopoietin receptor family that form homodimers are receptors for Tpo (MPL), G-CSF, growth hormone, and prolactin. These receptors transduce signals for a specific cytokine but do not cross-react with other ligands of the same family. Epo binding results in a conformational change in Epo-R, allowing two JAK2, a tyrosine kinase associated with the cytoplasmic domain of Epo-R, to come into close proximity with each other. JAK2 is activated by transphosphorylation of its active site, and subsequently phosphorylates some or all of the eight tyrosine residues in the intracellular domain of Epo-R. The essential role of JAK2 in red blood cell production has been demonstrated in knockout mice. Phosphorylation of tyrosine residues on Epo-R provides docking sites for intracellular proteins to bind to Epo-R via SH2 domains, and many of these are then also tyrosine phosphorylated. However, for both Epo-R and other hematopoietin receptors, the physiologic role of tyrosines in the cytoplasmic domain in regulation of proliferative, differentiating, and survival signals requires further definition. A representation of major signaling pathways activated

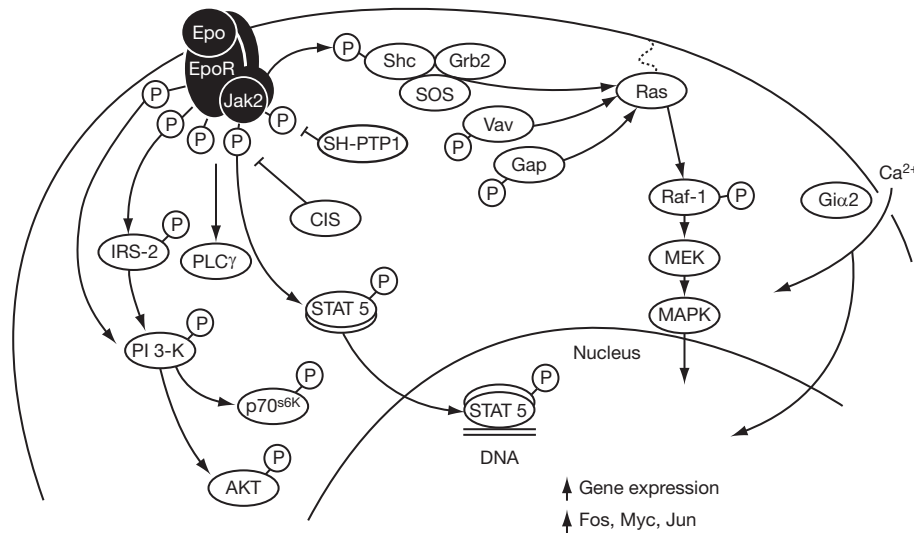


Figure 2 Pathways in erythropoietin receptor signaling. Erythropoietin binding leads to activation of associated JAK2 by autotransphosphorylation. Activated JAK2 phosphorylates tyrosines in the cytoplasmic region of Epo-R as well as associated proteins. This results in activation of STAT5, Ras/Raf/MAPK, and PI 3-kinase pathways, leading to transcription and activation of genes involved in cell cycling, differentiation, and survival.

following Epo-R stimulation is shown in **Figure 2**. One of the pathways activated by Epo is STAT5. STAT5 binds to phosphorylated Epo-R and is phosphorylated by JAK2. Phosphorylated STAT5 dissociates from Epo-R and forms dimers through SH2-phosphotyrosine interactions. These dimers then translocate to the nucleus and bind specific promoter elements to activate expression of target genes. Epo-R activation by its receptor also results in activation of Ras through several pathways. Through a number of signal adapters and transducers including Grb2, Shc, GAB1, Ras guanine nucleotide exchange factors called son of sevenless (SOS), and/or a hematopoietic-specific protein with Ras/guanosine diphosphate (GDP)/guanosine triphosphate (GTP) nucleotide exchange activity called Vav, Ras is activated by increased exchange of GDP with GTP. Epo stimulation also results in inhibition of GTPase-activating protein by Epo-induced tyrosine phosphorylation. Following Ras activation, a downstream cascade proceeds including phosphorylation and activation of Raf-1, and of mitogen-activated protein kinases (MAPKs), with subsequent induction of immediate-early genes including c-fos, c-myc, and c-jun. PI 3-kinase is another important kinase activated by Epo-R, either through direct association of the SH2 domain of the 85-kDa PI 3-kinase subunit with activated Epo-R or by binding to Epo-R through IRS-2 (insulin receptor substrate-2). PI 3-kinase can activate downstream targets including Akt, a kinase with crucial anti-apoptotic activity, and p70^{S6K}, a kinase that phosphorylates ribosomal protein S6 and plays an important role in transcription, translation, and cell-cycle progression. Ligand binding of Epo-R also results in an increase in intracellular calcium, a universal second messenger that influences many cell functions, through a mechanism involving activation of the transient receptor potential (TRP) channel TRPC3. Tyrosine phosphorylation in erythroid cells is down modulated by the hematopoietic protein tyrosine phosphatase SH-PTP1 (SHP1), which acts as a negative regulator. SH-PTP1 is recruited via its SH2 domains to tyrosine-phosphorylated Epo-R, dephosphorylates JAK2 and terminates proliferative signals.

Negative regulation is also mediated through SOCS proteins. One example is the SOCS protein CIS (cytokine-inducible SH2-containing protein), which may compete with STAT5 for binding sites on Epo-R and through this mechanism modify JAK/STAT signaling. CIS is also ubiquitinated and along with SOCS3 promotes Epo-R proteasomal degradation. A number of other nonreceptor protein kinases, phosphatases, and phospholipases have also been shown to have important roles in Epo-R signal transduction, as well as additional adapters and transducers, including phospholipase C γ , Cbl, and SHIP. Together, these pathways influence the expression and activity of a variety of proteins including transcription factors (GATA-1, NF-E2, SCL, NF- κ B, and Forkhead) and proto-oncogenes that participate in regulation of cell survival, proliferation, and differentiation.

Tpo Receptor

Expression

Tpo is the primary physiological hormone, which regulates all aspects of megakaryocyte and platelet development. Tpo interacts with its receptor, Mpl, on the cell surface of megakaryocyte precursors, mature megakaryocytes, and platelets to stimulate megakaryocyte progenitor proliferation and maturation, megakaryocyte polyploidy, and induction of platelet-specific proteins. The biological activity of the Mpl receptor is not limited to the megakaryocytic lineage. Unlike Epo or G-CSF, Tpo exerts profound effects on the proliferation of primitive hematopoietic cells through expression of Mpl on early hematopoietic progenitor cells of all lineages.

Signaling pathways activated by Tpo

Upon binding of Tpo to its receptor, Mpl, a number of signaling events are triggered, which are very similar to those of activated Epo-R. Binding of Tpo to Mpl induces the formation of an active dimer, enabling the associated JAK2 kinases to transphosphorylate each other. Activation of JAK2 results in

phosphorylation of two of five tyrosines in the cytoplasmic region of Mpl. Proteins recruited to these sites include STAT3 and STAT5, Shc, Grb2, SOS, and Vav. Signaling pathways that are activated include those of Ras, MAPKs, and PI 3-kinase/Akt, as well as negative regulatory pathways through tyrosine phosphatases including SH-PTP1 and the SOCS family. The distinction between intracellular signals of Tpo which promote hematopoietic stem cell expansion versus megakaryocyte development and platelet production are unknown.

G-CSF Receptor

Expression

The receptor for G-CSF, G-CSF-R, regulates the proliferation and differentiation of myeloid progenitors to form neutrophilic granulocytes, enhances the effector functions of mature neutrophils, and promotes their survival. In the hematopoietic system, G-CSF-R is expressed on myeloid progenitor cells, mature neutrophils, monocytes, and platelets. G-CSF-R expression increases with neutrophil differentiation. G-CSF-R is also expressed on endothelial cells, where it is thought to induce proliferation as well as migration, and on other nonhematopoietic cells including placenta, neuronal cells, activated T-lymphocytes, and cardiomyocytes.

Signaling pathways activated by G-CSF

G-CSF-R binds G-CSF as a homodimer. Unlike the other single-chain cytokine receptors, G-CSF-R activates multiple JAK kinases including JAK1, JAK2, and TYK2. Once activated, the JAK kinases induce phosphorylation of receptor tyrosines, which become docking sites for multiple SH2-containing signaling proteins as well as STAT1, STAT3, and STAT5. Pathways activated by G-CSF-R include those leading to activation of Ras, MAPKs, and Akt. Negative regulation occurs through the phosphatase SH-PTP1 and through SOCs. Experiments with G-CSF-R lacking one or more of the four cytoplasmic tyrosines suggest that pathways that regulate proliferation, differentiation, or survival differ, and characterization of these differences have important implications for understanding disorders of granulopoiesis.

GM-CSF/IL-3/IL-5 Receptors

Expression

The granulocyte/macrophage CSF (GM-CSF), interleukin-3 (IL-3), and IL-5 receptors are a subfamily of class I hematopoietin receptors. They have unique ligand-binding alpha chains but share a common beta chain (β_c), also called gp140. Many factors contribute to specificity in signaling between these receptors; one of these is that they are differentially expressed. Both the GM-CSF-R and the IL-3R are expressed on hematopoietic stem cells, and on the progenitors/precursors of neutrophils, macrophages, erythrocytes, megakaryocytes, and eosinophils. The IL-3R is expressed earlier in myeloid development than the GM-CSF-R. IL-3 promotes the expansion of the CD34⁺ multipotential progenitors and hematopoietic proliferation to a greater extent than GM-CSF, and promotes the development of a broader range of lineages that GM-CSF including monocytes, eosinophils, basophils megakaryocytes, mast cells, and

lymphoid cells. The GM-CSF receptor preferentially induces differentiation. By contrast, IL-5R expression is restricted to eosinophils, basophils, and mouse B cells. The IL-3 receptor subfamily is expressed on nonhematopoietic cells, but the *in vivo* significance is uncertain. Study of the crystal structure of the GM-CSF-stimulated receptor revealed a hexameric complex consisting of two GM-CSF molecules, two GM-CSF receptor α chains, and two β_c chains, which associate through a sequential process that is conserved among many other heterodimeric hematopoietin receptors.

Signal Transduction through the GM-CSF/IL-3/IL-5 Receptor

The common β_c subunit does not bind ligand by itself. Instead, β_c executes two critical functions following ligand binding. It enhances the ligand- α receptor binding from a low-affinity to a high-affinity interaction, and it functions as a major signaling component. The α chain may preassociate with β_c , but increased association occurs in the presence of ligand. Ligand binding increases oligomerization, creating clusters of two α and two β_c chains, the functional complex, or even larger complexes. Ligand binding facilitates formation of a disulfide linkage between α and β_c , not observed with the single-chain receptors. The β_c subunit elicits many signaling events, although the α chain is also required for signal transduction. There are eight tyrosine residues present on human β_c and proteins with SH2-binding domains that bind to the phosphorylated tyrosine residues largely determine the function of different receptor regions. The membrane proximal cytoplasmic region of β_c constitutively or inducibly associates with JAK kinases, predominantly JAK2, and is involved in induction of proliferation. The mid-region is necessary for promotion of viability, and the distal region is involved in growth inhibition. Receptor binding results in rapid activation/tyrosine phosphorylation of JAK2, β_c , and other JAK substrates. Downstream pathways activated include multiple STATs (1,3,5,6), Src kinases, Ras, MAPKs, PI 3-kinase, and Akt. Differences in signaling which result in the stimulation of different lineages, and proliferation versus differentiation or survival signals still require definition. For GM-CSF-R, which may serve as a model for some other hematopoietin receptors, exposure to low concentration of GM-CSF resulted in β_c receptor serine phosphorylation and initiates cell survival, whereas higher concentrations promoted tyrosine phosphorylation of β_c associated with proliferation as well as enhanced survival.

Receptors in the IL-6 SubFamily

Receptors for IL-6, IL-11, cardiotropin-1, ciliary neurotrophic factor, oncostatin M (OM), and leukemia inhibitory factor (LIF) represent another class I hematopoietin subfamily. They have major roles in hematopoiesis and in acute-phase and immune responses. They consist of a unique, ligand-binding α subunit and a common β chain called gp130, which oligomerize in response to ligand binding to α to initiate signaling. Although gp130 is ubiquitously expressed, the types of cells that respond to a specific IL-6 family cytokine is limited by the restricted and tightly regulated expression of the α receptor subunit. IL-6R and IL-11R have their own α subunits which complex with gp130, whereas other IL-6 subfamily cytokine

receptor complexes signal through different combinations of gp130, LIFR, and OM-R. A soluble form of the IL-6R α chain can be produced following limited proteolysis of membrane bound IL-6R or alternative splicing of mRNA. This receptor (sIL-6R) is unusual because unlike most other cytokine receptors, it can bind to and activate gp130 on cells that do not express IL-6R α chain after IL-6 exposure, thereby initiating signaling. Early hematopoietic cells are among the cells which can respond to IL-6 in the presence of only sIL-6R. By contrast, a soluble receptor form of gp130 detected in human serum inhibited signaling triggered by sIL-6R/IL-6 but not membrane-bound IL-6R, and thus is hypothesized to function as a natural inhibitor able to prevent constitutive signaling from sIL-6R. Similar to the common IL-3 β_c and the IL-2R γ subunit, the shared gp130 subunit enhances the affinity of receptor binding and serves as a major signaling component. Like other members of the hematopoietin family, these receptors signal through JAK/STAT, Ras, MAPK, and PI 3-kinase cascades, and terminate or modulate signals through tyrosine phosphatases, the SOCS, and PIAS (protein inhibitors of activated STAT) proteins.

Receptors in the IL-2 Subfamily

The IL-2 family cytokines play a critical role in the development of the immune system and in modulation of lymphocyte activation including maintenance of tolerance. The IL-2 receptor itself is critical in development and homeostasis of regulatory T cells. The IL-2 receptor subfamily includes receptors for IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21, which share the γ_c chain of IL-2R. Receptors for IL-4, IL-7, and IL-9 consist of two subunits, a unique α subunit in addition to γ_c . Receptors for IL-2 and IL-15 have three subunits, a unique α subunit and shared IL-2R β and γ_c . The differential expression of IL-2-receptor subunits, mechanisms of activation, and *in vivo* function in immune regulation are complex and have been described in detail. The α subunit acts as the ligand-binding subunit, whereas the β and γ subunits function to enhance receptor affinity and play a predominant role in intracellular signal transduction. Many of the signal transducers utilized by these receptors are the same as for other hematopoietin receptors, except that this subfamily is one of the few to utilize JAK3. Understanding the dynamics of γ_c -containing receptor expression and their function is important in lymphocyte biology and in the use of cytokine therapies targeting these receptors, which may have a number of different clinical applications in immune modulation.

The Class II Hematopoietin Receptor Family

IFN receptors have been divided into several major groups including type I (IFN- α , IFN- β , IFN- ω , and others), which are vital for cells to initiate an antiviral response and innate resistance to infection; type II (IFN- γ), which are involved in regulation of specific immune responses; and type III (IFN- λ). As an example of interferon receptor activation, IFN- γ binds to ligand-binding IFN γ R1 subunits, followed by recruitment of IFN γ R2 subunits to initiate signal transduction events through

activation of associated JAKs. The active IFN- γ receptor consists of two IFN γ R1 and two IFN γ R2 subunits. IFN receptors signal through distinct combinations of JAKs through signaling cascades involving the STATs, MAP kinases, PI 3-kinase, and additional pathways. Many of the pathways leading to specificity of response need to be defined.

Human Diseases Associated with Defect in Hematopoietin Receptors or Their Signaling Pathways

Naturally occurring defects in hematopoietin receptors are associated with a number of human diseases, resulting in either reduction or enhancement of cell production depending on the genetic location. Natural mutations in the negative regulatory region of cytoplasmic domain of Epo-R have been identified in benign erythrocytosis, and mutations in MPL (Tpo receptor) in hereditary thrombocytosis. By contrast, defective expression or function of MPL is the cause of the majority of cases of congenital amegakaryocytic thrombocytopenia, a severe inherited thrombocytopenia in infants with absent megakaryocytes in bone marrow and severely low circulating platelet counts. Congenital neutropenia is a heterogeneous disorder of myelopoiesis associated with maturation arrest at the promyelocyte/myelocyte stage and with distinct molecular mechanisms involving defects in genes including those for neutrophil elastase and Hax-1. Acquired mutations in the G-CSF receptor define a group of patients with a significantly increased likelihood of developing leukemia. The mutations, often resulting in truncation of the G-CSF receptor C-terminal cytoplasmic region, are thought to disrupt granulocytic maturation. Defects in transducers in the signaling pathways of hematopoietin receptors can also result in significant clinical phenotypes. Loss-of-function mutations of the γ_c (IL-2 receptor subfamily), JAK3, or TYK2 genes result in immunodeficiency. Mutations and translocations in JAK genes leading to constitutive activation are associated with a number of hematological malignancies including myeloproliferative disorders, acute lymphoblastic leukemia, and acute myeloid leukemias. Aberrant expression of IL-6 or increased expression of sIL-6R is now recognized to play an important role in a number of inflammatory and autoimmune disorders, as well as cancer. Blocking the IL-6R with targeted antibody is a promising new treatment for rheumatoid arthritis and potentially certain other inflammatory diseases. Mutations of the γ_c chain in the subfamily of IL-2 receptors have been shown to be the cause of X-linked severe combined immunodeficiency. All mutant receptors are defective in activation of JAK3.

See also: **Signaling: Cytokines; Interferon Receptors; T-Cell Antigen Receptor.**

Further Reading

- Baker SJ, Rane SG, and Reddy EP (2007) Hematopoietic cytokine receptor signaling. *Oncogene* 26: 6724–6737.
- Bezbradica JS and Medzhitov R (2009) Integration of cytokine and heterologous receptor signaling pathways. *Nature Immunology* 10: 333–339.

- Burchill MA, Yang J, Vang KB, and Farrar MA (2007) Interleukin-2 receptor signaling in regulatory T cell development and homeostasis. *Immunology Letters* 114: 1–8.
- Hercus TR, Thomas D, Guthridge MA, et al. (2009) The granulocyte–macrophage colony-stimulating factor receptor: Linking its structure to cell signaling and its role in disease. *Blood* 114: 1289–1298.
- Jelkmann W, Bohlius J, Hallek M, and Sytkowski AJ (2008) The erythropoietin receptor in normal and cancer tissues. *Critical Reviews in Oncology/Hematology* 67: 39–61.
- Kaushansky K (2009) Molecular mechanisms of thrombopoietin signaling. *Journal of Thrombosis and Haemostasis* 7(supplement): 235–238.
- Malek TR (2008) The biology of Interleukin 2. *Annual Review of Immunology* 26: 453–479.
- Metcalf D (2008) Hematopoietic cytokines. *Blood* 111(2): 485–491.
- Overwijk WW and Schluns KS (2009) Functions of γ C cytokines in immune homeostasis: Current and potential clinical applications. *Clinical Immunology* 132: 153–165.
- Panopoulos AD and Watowich SS (2008) Granulocyte colony-stimulating factor: Molecular mechanisms of action during steady state and 'emergency' hematopoiesis. *Cytokine* 42: 277–288.
- Percy MK and Rumi E (2009) Genetic origins and clinical phenotype of familial and acquired erythrocytosis and thrombocytosis. *American Journal of Hematology* 84: 46–54.
- Quelle FW (2007) Cytokine signaling to the cell cycle. *Immunology Research* 39: 173–184.
- Robb L (2007) Cytokine receptors and hematopoietic differentiation. *Oncogene* 26: 6715–6723.
- Scheller J and Rose-John S (2006) Interleukin-6 and its receptor: From bench to bedside. *Medical Microbiology Immunology* 195: 173–183.
- Vainchenker W, Dusa A, and Constantinescu SN (2008) JAKs in pathology: Role of Janus kinases in hematopoietic malignancies and immunodeficiencies. *Seminars in Cell and Developmental Biology* 19: 385–393.

Heme Proteins

J Everse, Texas Tech University Health Sciences Center, Lubbock, TX, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 354–361, © 2004, Elsevier Inc.

Glossary

Biradical bond A chemical bond characterized as a biradical, three-center four-electron bond.

Prosthetic group A chemical compound present in and is an integral part of an enzyme or protein that is not composed of amino acid residues.

Singlet oxygen An oxygen molecule having one electron in an excited state. Chemical notation: $O_2(^1\Delta g)$.

Soret band The most intense absorption band in heme proteins. Also referred to as the γ band.

The class of proteins known as heme proteins includes all proteins that contain a heme moiety as a prosthetic group, which may or may not be covalently bound to the protein. The class encompasses proteins with a wide range of functions. Based on their properties, they may be separated into the following three subclasses: (1) Proteins that bind oxygen reversibly. These proteins are designed to transport and/or store oxygen among or within various body tissues. (2) Proteins that bind oxygen and activate it. These proteins serve as enzymes, capable of catalyzing an extensive list of reactions. This group also includes the enzymes that react with hydrogen peroxide. (3) Proteins that are unable to bind oxygen, and serve as electron carriers or transporters.

Structural Properties of Heme Proteins

Heme proteins are strongly colored proteins, usually reddish-brown, which is due to the presence of the heme moiety. The heme moiety consists of a substituted protoporphyrin ring, containing a liganded iron atom. In many mammalian heme proteins, the protoporphyrin ring is protoporphyrin IX, shown in [Figure 1](#). Chelated iron, whether ferrous or ferric iron, preferentially forms six liganded structures. In the heme proteins, four of these ligands are formed with the nitrogen atoms of the four pyrroles in the protoporphyrin ring. The fifth ligand, often referred to as the proximal ligand, is formed by an amino acid residue of the protein. This residue is a histidine residue in the oxygen-carrying proteins – myoglobin and hemoglobin, where the bond with the oxygen molecule forms the sixth or distal ligand. In heme proteins that serve as enzymes, the fifth ligand may be formed by a histidine, cysteine, methionine, or tyrosine residue. In the electron-carrying cytochromes, both fifth and sixth ligands are occupied by amino acid residues.

Heme Proteins That Serve as Oxygen Carriers

Several heme proteins are specifically designed to bind oxygen reversibly. In higher animals these proteins are hemoglobin

and myoglobin. The function of hemoglobin is to take up oxygen in the lungs, and transport it to the various tissues in the body. Myoglobin, present in muscle tissues, takes up the oxygen from the hemoglobin and stores it until it is needed for the production of adenosine triphosphate (ATP) in the mitochondria.

One oxygen molecule binds to the iron atom of the heme moiety at the sixth position which is vacant in the deoxygenated form of these proteins. The nature of the Fe–oxygen bond has been the subject of many investigations, and is still a matter of some controversy. At the present time, it appears that the properties of the oxygenated forms of hemoglobin and myoglobin are best explained by assuming that the bond between the oxygen molecule and the heme iron is a biradical bond, stabilized by d–p orbital overlap.

Myoglobin

Function

Myoglobin serves as an oxygen storage protein in muscle tissues, where it also aids in the diffusion of oxygen into the tissues. Together with the cytochromes, it is responsible for the red-brown color of muscle. The protein is abundant in the heart and skeletal muscles of animals; and it is especially abundant in the muscles of diving mammals, such as whales and seals, where a high capacity for the storing of oxygen is essential.

Structure

The protein is a monomeric heme protein, consisting of a single chain of 153 amino acids, and a non-covalently bound heme moiety. The structure is that of a compact, globular protein, consisting of eight helices, with the heme moiety located in a cleft formed between helices C, E, and F ([Figure 2](#)). The fifth ligand to the heme iron is donated by the imidazole nitrogen of histidine residue F8 of the F helix. The sixth ligand is formed by the bound oxygen, whereas in deoxymyoglobin the sixth ligand is vacant.

The iron atom in myoglobin is in the ferrous state. Oxidation of the iron to the ferric form yields metmyoglobin, which is unable to bind oxygen, and is therefore no longer functional.

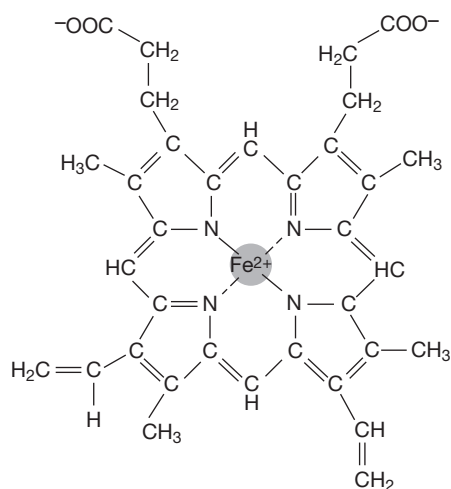


Figure 1 Chemical structure of iron-protoporphyrin IX (hemin).

Oxygen binding

Myoglobin has a high affinity for oxygen, as shown in [Figure 3](#). This fact allows muscle tissues to extract oxygen from the bloodstream.

Hemoglobin

Structure

Hemoglobin is one of the most widely studied proteins. Its structure consists of four subunits, normally an $\alpha_2\beta_2$ tetramer, and each containing a heme moiety with the iron in the ferrous state.

Each α chain consists of 141 amino acid residues, and the β chains have 146 residues each. The three-dimensional structure of the individual subunits is very similar to that of myoglobin, although the amino acid compositions are very different.

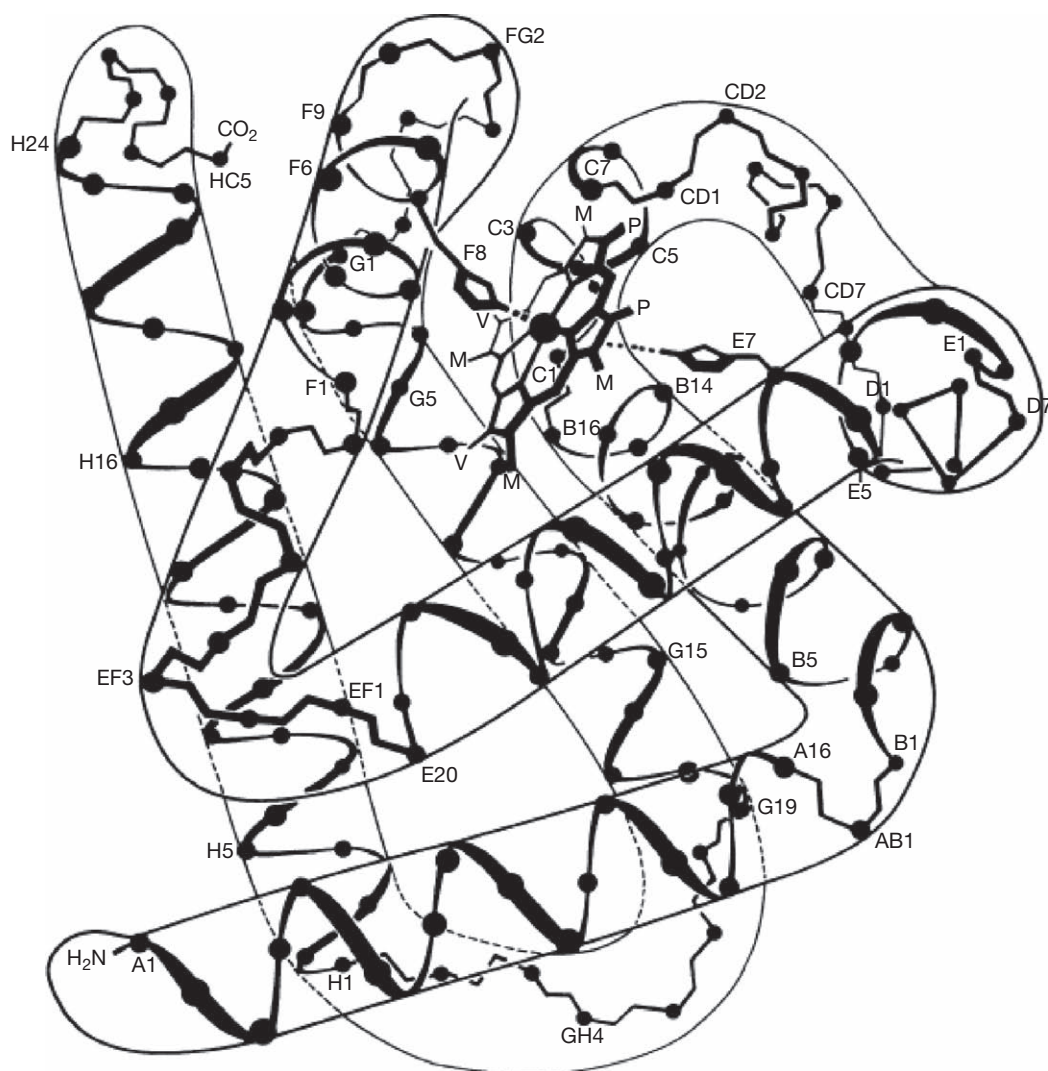


Figure 2 Molecular structure of myoglobin. The polypeptide chain consists of eight helices, designated by the letters A through H, linked by short, unordered regions. Amino acids are numbered by their relative position in the chain. Reprinted from Dickenson RE (1964) *The Proteins*, 2nd edn., vol. II, p. 634. San Diego, CA: Academic Press, with permission from Elsevier.

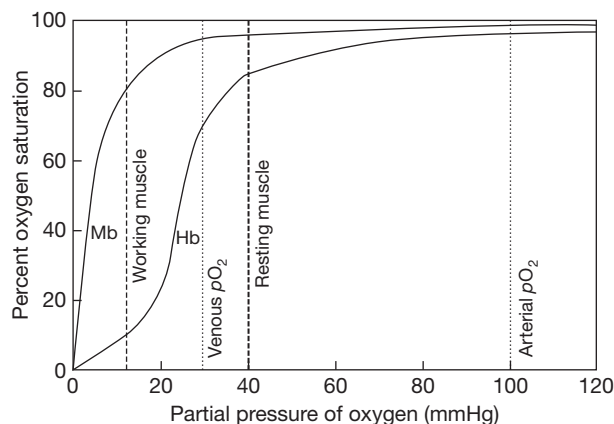


Figure 3 Oxygen saturation curves of hemoglobin (Hb) and myoglobin (Mb).

Oxygen binding

Allosteric effects

The tetrameric structure of the molecule is crucial to its biological function. Oxygen does not have a high affinity for deoxyhemoglobin. However, binding of a molecule of oxygen to one of the hemes causes the Fe atom, which is normally located somewhat above the plane of the heme ring, to be drawn into the plane, a distance of 0.04 nm. This relatively slight movement triggers a change in the protein conformation that is transmitted to the adjacent subunits, which in turn dramatically enhances the affinity of their heme groups for oxygen. The result is that the oxygen saturation curve of hemoglobin has a sigmoidal form, as shown in **Figure 3**. This strong cooperative effect among the four subunits assures that almost total oxygenation occurs when the hemoglobin passes through the lungs, and that the oxygen is effectively released in the body tissues. A similar change in the position of the heme iron occurs in myoglobin upon the binding of an oxygen molecule; in myoglobin, however, this change in position appears to be nonconsequential.

Bohr effect

The affinity for oxygen is also dependent on pH. Lower pH decreases the affinity for oxygen. This is known as the Bohr effect. This effect causes a better unloading of the oxygen in tissues with a high metabolic rate. Such tissues usually have a lower pH, because of the presence of acidic products, such as lactate, citrate, and malate.

Regulation by 2,3-diphosphoglycerate

Oxygen binding to hemoglobin is also regulated by the metabolite 2,3-diphosphoglycerate (2,3-DPG). 2,3-DPG is present in erythrocytes at a concentration similar to that of hemoglobin. Hemoglobin binds one molecule of 2,3-DPG per tetramer. Its regulatory effect on oxygen binding by hemoglobin is most evident when there is a change in altitude. At higher altitudes, where the partial pressure of oxygen is lower, 2,3-DPG binding compensates by shifting the oxygen saturation curve to the right.

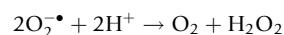
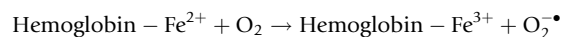
Interaction with other compounds

In addition to oxygen, other compounds bind to the heme iron of hemoglobin and myoglobin. Among them are the

notorious compounds, that is, carbon monoxide and cyanide. Both compounds bind to the heme iron much more strongly than oxygen, thus preventing oxygen from binding. This is the basis for the high toxicity of these compounds, even at low concentrations.

Methemoglobin formation

The oxygen in oxyhemoglobin promotes a spontaneous oxidation of the ferrous hemoglobin to ferric hemoglobin or methemoglobin, which is unable to bind oxygen. During this reaction the oxygen is reduced to the superoxide radical, which subsequently dismutates to hydrogen peroxide:



Fortunately, an enzyme present in red cells, methemoglobin reductase, reduces the methemoglobin back to normal hemoglobin, whereas catalase, which is also abundant in erythrocytes, catalyzes the dismutation of hydrogen peroxide to oxygen and water. This mechanism, together with the fact that the oxidation to methemoglobin is rather slow, assures that under normal conditions the amount of methemoglobin present in the red cell is only a few percent of the total hemoglobin present.

Hemoglobin mutants

Fetal hemoglobin

A large number of mutant hemoglobins are known to exist, most of which are harmless modifications. Several, however, are of physiological importance. Most important among them is fetal hemoglobin or hemoglobin F (HbF), which contains two γ chains instead of two β chains. Its tetrameric constitution is thus $\alpha_2\gamma_2$, instead of the $\alpha_2\beta_2$ configuration of normal hemoglobin (HbA). The γ chain differs from the β chain in that a serine residue is present at position 143 instead of a histidine. This one amino acid substitution causes a shift in the oxygen affinity curve to the left, especially at the upper portion of the curve. Thus, HbF has a higher affinity for oxygen than HbA. This allows the fetus to extract oxygen from its mother's blood in the placenta. HbF is gradually replaced by HbA during the first several months after birth.

Sickle cell hemoglobin

Another well-known hemoglobin mutant is HbS, or sickle cell hemoglobin. In HbS the amino acid residues at position 6 in the β chains are valines rather than glutamic acid residues. This change causes the deoxy form of HbS to aggregate into insoluble fibers. The presence of these fibers distorts the shape of the erythrocytes from a disk-shaped into a sickle-shaped cell. Sickle cells are considerably more fragile than normal erythrocytes, which accounts for the observed anemia in sickle cell patients.

It is estimated that in the USA about 3 million people suffer from the disease, mostly people of African descent. The reason for the high survival rate of the mutant gene is that HbS provides protection against the malaria parasite, the greatest killer of all times. This parasite spends part of its life cycle in human erythrocytes. However, the HbS is apparently much

less suitable for the proper development of the parasite than HbA, thus protecting the sickle cell patients when malaria epidemics occur.

Blood substitutes

A considerable amount of research has been done recently on the modification of human and bovine hemoglobins in order to produce a suitable product to serve as a blood substitute. This requires a covalent cross-linking of the α and β chains in order to prevent dissociation of the tetramer. Initial clinical trials, however, revealed that hemoglobin can be quite toxic when it is present in large quantities in the circulation outside the confines of the red cell.

Heme Proteins That Serve as Enzymes

A considerable number of heme proteins are found in nature that serve as enzymes. These include most peroxidases and catalases, as well as a large group of monooxygenases generally referred to as the cytochrome P₄₅₀ enzymes. In addition, a few oxidases, hydrolases, and dioxygenases are heme enzymes. Peroxidases and catalases use hydrogen peroxide as the oxidizing agent. The cytochrome P₄₅₀ enzymes as well as the other enzymes catalyze oxidizing reactions that involve molecular oxygen, of which either one or both atoms are incorporated into a substrate.

Peroxidases

Structure and function

Peroxidases are a group of enzymes that catalyze the oxidation of a substrate by hydrogen peroxide or an organic peroxide. Most peroxidases are ferric heme proteins – one notable exception being the glutathione peroxidase, which is a selenium-containing enzyme. They are present in virtually all living species.

Most peroxidases are able to oxidize a variety of substrates of widely different structures, varying from halide ions to hydroquinones to aromatic azo compounds. A binding site for the substrate to be oxidized is therefore virtually absent in most of these enzymes. In fact, simple molecules consisting of 8–12 amino acids and a covalently linked heme moiety, with a histidine residue forming the proximal ligand, have considerable peroxidase activity. These compounds are known as microperoxidases. Furthermore, methemoglobin and metmyoglobin can effectively function as peroxidases, having about 5–10% of the activity of horseradish peroxidase.

Mechanism of action

The mechanism of action of the heme peroxidases is shown in Figure 4. The first step involves an oxidation of the enzyme by the peroxide, leading to the formation of a ferryl (Fe^{IV}) heme moiety and the removal of an electron from either the porphyrin ring or an amino acid residue of the protein. This intermediate is commonly referred to as compound I. Compound I can then react with a suitable substrate and promote a one-electron oxidation of the substrate. A one-electron oxidation of the substrate (such as the oxidation of a hydroquinone to a semiquinone) converts the enzyme to compound II, in which the

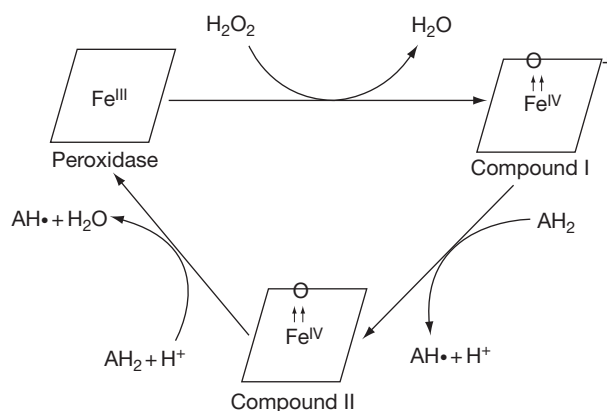


Figure 4 Mechanism of action of peroxidases.

heme is still in the ferryl form, but the original structure of the porphyrin ring or the amino acid residue is restored. Transfer of a second electron from compound II to a substrate restores the native ferric enzyme.

The two arrows in compounds I and II represent the two unpaired electrons in the biradical bond between the iron and the oxygen atom. The '+' mark in compound I represents the positive charge on the porphyrin ring. AH_2 and $\text{AH}\cdot$ represent the substrate and the radical product, respectively.

Of the plant peroxidases, the enzyme from horseradish is perhaps the most studied. It is a 40-kDa monomeric enzyme, with a histidine residue as the proximal ligand. The functions of the peroxidases in plants appear to be the generation of free-radical intermediates that promote the polymerizations and cross-linkings necessary to form the cell-wall structures in plants. Enzymes present in the roots of plants are thought to protect the host against invading fungi and other microorganisms.

In mammals, the most studied peroxidases are myeloperoxidase present in neutrophils, eosinophil peroxidase present in eosinophils, thyroid peroxidase present in the thyroid gland, and lactoperoxidase present in milk and saliva. Except for the thyroid peroxidase, the function of all these peroxidases seems to involve a defense against invading microorganisms. The function of the thyroid peroxidase is the iodination of tyrosine residues during the synthesis of the hormone thyroxine.

Haloperoxidases

The mechanism by which myeloperoxidase exerts its antimicrobial activity has been largely elucidated by Seymour Klebanoff and his colleagues. The phagocytosis of a microorganism by a neutrophil activates a membrane-bound NADPH oxidase that excretes large amounts of hydrogen peroxide into the phagosome. The myeloperoxidase reacts with the H_2O_2 to form a compound I-type structure, which in turn oxidizes a chloride ion to hypochlorite (HOCl) in a reaction involving a two-electron transfer, similar to that of catalase. The hypochlorite formed is lethal to the microorganism. Lactoperoxidase uses a similar mechanism to kill bacteria present in milk or saliva; however, it uses either iodide or isocyanate as the ion to be oxidized. *In vitro*, these enzymes exert a cytolytic activity against most cells, prokaryotic as well as eukaryotic.

Other peroxidases that are capable of oxidizing halide ions include chloroperoxidase present in certain molds. This enzyme carries out organic chlorination reactions, presumably through the formation of an intermediate enzyme-bound form of hypochlorite.

Reactive oxygen species

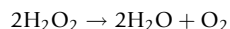
An increasing number of the heme enzymes have recently attracted attention, because of their alleged involvement in the generation of reactive oxygen species (ROSs). ROSs are implicated in causing damage to DNA, RNA, lipids, and proteins, and may be directly or indirectly involved in the initiation of cancer, as well as diseases such as Alzheimer's, Parkinson's, multiple sclerosis, and other diseases involving tissue degeneration and cell death. ROSs include superoxide, hydrogen peroxide, the hydroxyl radical, singlet oxygen, and ozone.

It should be clear from the above description that peroxidases are enzymes specifically designed to generate free-radical products. Compounds I and II have oxidative energies comparable to that of the hydroxyl radical, and the products generated by these enzymes often are free radicals. These products, in turn, can promote polymerization reactions of proteins and lipids. Furthermore, it is known from *in vitro* experiments that most, if not all, peroxidases are able to exert cytolytic activities under appropriate conditions. It is also known that both myeloperoxidase and chloroperoxidase are able to generate singlet oxygen. These observations support the hypothesis that peroxidases may be involved in at least some of the tissue damage that is presently ascribed to the actions of ROSs.

Catalases

Structure and function

Catalases are enzymes designed to neutralize the hydrogen peroxide that is formed *in vivo* by various oxidases and other enzymes. The enzyme catalyzes the dismutation of 2 mol of hydrogen peroxide into 1 mol of oxygen and 1 mol of water:



Catalases are found in all aerobic cells; in fact, in some bacteria catalase may account for as much as 1% of their total dry weight. High concentrations are also present in erythrocytes, where it serves to neutralize the hydrogen peroxide formed during the autoxidation of oxyhemoglobin to methemoglobin. Catalases are among the enzymes with the highest turnover rates known: under optimal conditions each subunit can dismutate 2×10^5 mol of hydrogen peroxide per second.

Catalases in eukaryotes are tetramers, each subunit containing a ferric heme and a bound molecule of NADPH. The reason for the presence of this coenzyme is still unclear at this time. In the catalases, a tyrosine residue forms the proximal ligand to the heme iron.

Mechanism of action

Their mechanism of action is somewhat similar to that of the peroxidases. Upon binding of a mole of H_2O_2 , a compound I-type intermediate is formed, and a mole of water is released. The enzyme then reacts with a second mole of H_2O_2 to form a mole of oxygen and the native enzyme is regenerated (Figure 5). No radical intermediates are formed by catalases.

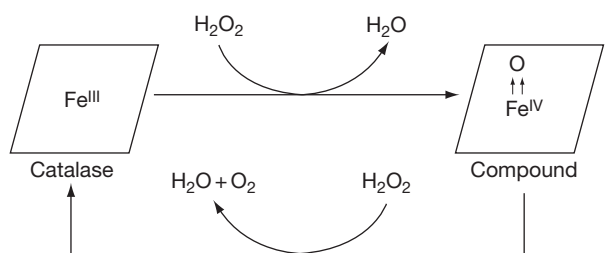


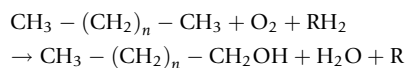
Figure 5 Mechanism of action of catalases.

Oxygenases

Oxygenases are a group of heme enzymes that catalyze the oxidation of various substrates by molecular oxygen. Certain oxygenases catalyze the incorporation of both oxygen atoms into the substrate; these are referred to as dioxygenases. Others catalyze the incorporation of one atom of oxygen into the substrate, while the other atom is reduced to water. This group constitutes the monooxygenases.

Almost all of the dioxygenases do not contain heme; non-heme iron sulfur groups form their active center. An example of a heme-containing dioxygenase is tryptophan dioxygenase.

Monooxygenases are also referred to as mixed-function oxygenases or hydrolases. In addition to molecular oxygen, this group of enzymes requires the presence of a co-substrate capable of donating a pair of electrons to reduce the second atom of oxygen to water. A typical monooxygenase reaction is the hydroxylation of an alkane to an alcohol:



in which R represents the co-substrate. Many compounds can serve as a co-substrate for the monooxygenases; these include the pyridine nucleotides, flavins, ferredoxins, hydroquinones, ascorbate, and others.

Cytochromes P₄₅₀ Enzymes

The cytochrome P₄₅₀ enzymes are a special group of monooxygenases, all of which use NADPH as the co-substrate. The name of this group of enzymes is associated with their discovery, when it was noted that, in the presence of carbon monoxide, their absorption spectrum resembles that of the known cytochromes, with the maximum absorbance of the Soret band located at about 450 nm. Functionally, however, they are not related to the group of proteins normally known as cytochromes.

Occurrence

The cytochrome P₄₅₀ family of enzymes is present in all living species. They have been isolated from bacteria, yeasts, plants, insects, as well as from vertebrate animals, where they are present in a large number of tissues. In fact, a list of the sequences of 325 different cytochrome P₄₅₀'s is presented in Ortiz de Montellano's book *Cytochrome P450*, and many more have been identified since then.

Functions

The cytochromes P₄₅₀ comprise a large group of enzymes that are able to incorporate one of the two atoms of an oxygen

molecule into a broad variety of substrates, with the concomitant reduction of the other oxygen atom to water. Members of the family are known to catalyze hydroxylations, epoxidations, N-oxidations, N-, S-, and O-dealkylations, sulfoxidations, dehalogenations, and other reactions.

The two main functions of the cytochrome P₄₅₀ enzymes appear to include first the control of the steady-state level of endogenous effectors of growth and differentiation. As such, they are important in the metabolism of numerous physiological substrates, including prostaglandins, steroids, cytokines, bile acids, fatty acids, and biogenic amines. Second, they metabolize a wide range of foreign chemicals, including environmental pollutants, drugs, various natural plant products, and many other foreign materials that enter the body.

Mechanism of action

The enzymes contain a heme moiety with iron in the ferric state, and with a cysteine residue as the proximal ligand. Oxygen binds as the sixth ligand. The principal catalytic cycle of the cytochrome P₄₅₀ enzymes for a hydroxylation reaction is shown in **Figure 6**. The first step involves a binding of the substrate XH to the enzyme. In step 2, the ferric enzyme is reduced to the ferrous form with an electron obtained from the co-substrate NADPH. The ferrous enzyme then binds a molecule of oxygen in step 3. In step 4, a second electron from NADPH is used to reduce the oxygen molecule, which leads to the uptake of two protons, the release of a molecule of water, and the formation of a compound I-type ferryl enzyme. In the final step, the oxygen atom of compound I is incorporated into the substrate,

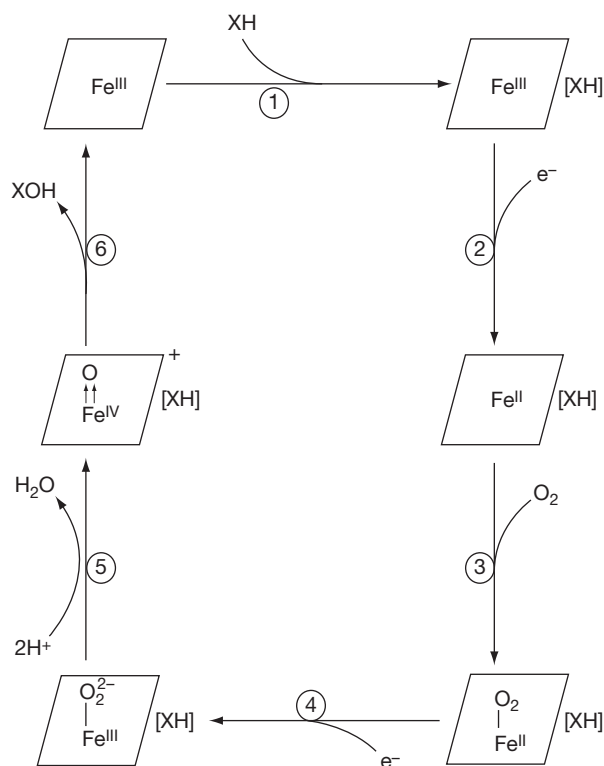


Figure 6 Principal catalytic cycle of cytochrome P₄₅₀ enzymes. XH, substrate; XOH, product.

and the product is released. Steps 4 and 5 probably constitute a series of individual steps; however, they occur so fast that they cannot be observed with present-day's techniques.

Heme Proteins That Serve as Electron Carriers

Cytochromes

Function

The cytochromes, with the exception of the members of the cytochrome P₄₅₀ family, are heme proteins, designed for the transport of individual electrons. Their heme moiety is normally enveloped by the protein portion, excluding it from contact with the outside solvent. The transport of individual electrons is facilitated by an oscillation of the heme iron between the ferrous and the ferric forms. The flow of electrons from the outer edge of the cytochrome to the heme iron may involve the participation of one or more amino acid residues (e.g., tyrosines). The driving force for the movement of electrons is provided by the differences in redox potential between the electron donor and the acceptor.

Cytochromes are identified by the type of heme moiety present (a, b, c, or d), and the wavelength of maximal absorbance of the Soret band (e.g., cytochrome b₅₅₉). The four heme moieties differ in the structure of the side chains of the porphyrin ring.

Cytochromes c

c-Type cytochromes are found in animals, plants, and eukaryotic microorganisms. They are the only type cytochromes that are water soluble. They are monomeric proteins, containing a heme c moiety, with the proximal and distal ligands formed by histidine and methionine residues. They usually contain a single heme per molecule; however, molecules containing up to eight heme residues are found in certain photosynthetic bacteria. The best-known and most-studied member of this group is the mitochondrial cytochrome c, which transports electrons from complex II to complex III in the electron transport system.

Cytochromes b

b-Type cytochromes are widely distributed. They are present in bacteria, chloroplasts, and mitochondria. The mitochondrial b-type cytochromes are normally embedded in membranes as part of complex II of the electron transport system. They contain a protoheme with histidine in both the proximal and distal ligand positions. The function of all b-type cytochromes appears to be to transport electrons from dehydrogenases to cytochrome c-type proteins or to iron-sulfur proteins.

Cytochromes a

a-Type cytochromes in mitochondria are an important part of the electron transport chain. They form an integral part of complex III, where they take up electrons from cytochrome c and pass them on to the copper ions and the bound oxygen in the cytochrome oxidase system.

See also: **Bioenergetics**; Cytochrome c; Cytochrome P-450; Heme Synthesis; Iron-Sulfur Proteins.

Further Reading

- Alayash AI (2000) Hemoglobin-based blood substitutes and the hazards of blood radicals. *Free Radical Research* 33: 165–175.
- Alayash AI, Patel RP, and Cashion RE (2001) Redox reactions of hemoglobin and myoglobin: Biological and toxicological implications. *Antioxidants and Redox Signaling* 3: 313–327.
- Brunori M (2000) Structural dynamics of myoglobin. *Biophysical Chemistry* 86: 221–230.
- Dickenson RE (1964) *The Proteins*, 2nd edn., vol. II, p. 634. San Diego, CA: Academic Press.
- Everse J (1998) The structure of heme proteins compounds I and II: Some misconceptions. *Free Radical Biology and Medicine* 24: 1338–1346.
- Everse J, Everse KE, and Grisham MB (eds.) (1991) *Peroxidases in Chemistry and Biology*, vols. I and II. Baton Rouge, FL: CRC Press.
- Everse J, Vandegriff KD, and Winslow RM (eds.) (1994) *Hemoglobins Methods in Enzymology*, vols. 231 and 232. San Diego, CA: Academic Press.
- Garrett RH and Grisham CM (1994) *Biochemistry*. Orlando, FL: Saunders College Publishing.
- Ortiz de Montellano PR (ed.) (1995) *Cytochrome P450: Structure, Mechanism, and Biochemistry*. New York, NY: Plenum Press.
- Ortiz de Montellano PR and De Voss JJ (2002) Oxidizing species in the mechanism of cytochrome P450. *Natural Product Report* 19: 477–493.

Heme Synthesis

G C Ferreira, University of South Florida, Tampa, FL, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Erythroblast Erythroid cell precursor; nucleated cell in bone marrow that develops into an erythrocyte.

Erythrocyte A red blood cell; non-nucleated, disk-shaped blood cell that contains hemoglobin.

Erythroid Erythrocyte related.

Erythropoiesis Formation and production of erythrocytes.

Tetrapyrrole A general term that refers to molecules with four rings of the pyrrole type, most often linked together by single-atom bridges between the α -positions of the five-membered pyrrole rings. The arrangement of the four rings is macrocyclic in porphyrins.

Introduction

Heme, a four-ring organic compound with a central iron atom, plays multiple roles in cellular processes. For example, it is the crucial component in hemoglobin, a red blood cell protein responsible, in vertebrates, for the transport of oxygen from the lungs to the different tissues and the transport of carbon dioxide from the tissues to the lungs. It is heme that gives the characteristic red color to blood. Given its importance in many vital functions in organisms ranging from bacteria to man, it is not surprising that most cells synthesize heme. However, in humans, differentiating erythrocytes represent the major site of heme production due to the synthesis of hemoglobin, which accounts for ~85% of the total heme.

Heme Structure

The observation that hemoglobin can be split into two components, with the smaller unit corresponding to the compound which today is recognized as heme, was made by L R LeCanu in 1837. While the correct structure of heme was initially proposed by Kuster in 1912, it only received general acceptance 16 years later upon its chemical synthesis. The Nobel prize-winning organic chemist Hans Fischer developed syntheses leading to the assembly of four rings, or pyrroles, into a macrocyclic structure with a purple-red color, which he appropriately named 'porphyrin' (from the Greek porphuros, meaning purple). The chemically synthesized porphyrin, which had properties identical to those of the natural porphyrin (derived from heme upon the release of the iron atom), was designated protoporphyrin IX. Thus, heme consists of four rings (pyrroles), composed of carbon, nitrogen, and hydrogen atoms, linked together in a cyclic array and containing an atom of iron in its center (Figure 1).

Heme is now recognized as a member of a large family of metalated tetrapyrroles, which are essential for life on Earth. They include chlorophylls, cobalamin (vitamin B₁₂), siroheme, and coenzyme F₄₃₀. Chlorophylls trap the light energy required in photosynthesis. Cobalamin-containing enzymes catalyze intramolecular rearrangements, methylations, and reduction of ribonucleotides to deoxyribonucleotides. Siroheme is an essential component of enzymes involved in

sulfur and nitrogen metabolisms. Coenzyme F₄₃₀, present in methanogenic bacteria, functions in methane formation. While the central metal ions in these tetrapyrroles differ – Fe²⁺ in the case of heme and siroheme, Mg²⁺ in chlorophyll and bacteriochlorophyll, Co²⁺ in cobalamin, and Ni²⁺ in coenzyme F₄₃₀ – the unifying feature is the presence of a complexed metal ion, which promotes a remarkable number of different biochemical reactions. Further, the protein scaffold binding the metalloporphyrin modulates the chemistry (redox state and coordination properties) of the metal ion, such that the spectrum of the reactions expands beyond the normal properties of the metal ion. It is this interplay between the protein moiety and the metalloporphyrin that provides the plethora of functions associated with metalloporphyrin-containing proteins.

The Heme Biosynthetic Pathway

Formation of 5-Aminolevulinate

The biosynthesis of tetrapyrroles is initiated from simple molecules and comprises only a few enzymatic steps (Figure 2). The first committed precursor, 5-aminolevulinic acid (ALA), is synthesized from either succinyl-CoA or glutamate, and thus the names C4 and C5 (based on the number of carbon atoms in the precursor), for the two ALA biosynthetic pathways evolved in nature. In most nonphotosynthetic eukaryotes and a few prokaryotes (i.e., the α -proteobacteria), ALA results from the condensation between glycine and succinyl CoA (C4 pathway), while in plants and most bacteria, ALA is derived from glutamate (C5 pathway). The C4 pathway requires a single enzyme, ALA synthase (ALAS); by contrast, three enzymes are involved in the conversion of glutamate into ALA. Specifically, glutamyl-transfer RNA (tRNA) synthetase converts glutamate to glutamyl-tRNA, which is reduced by glutamyl-tRNA reductase to glutamyl-1-semialdehyde (GSA); this product is then transformed by GSA aminomutase or GSA aminotransferase (GSA-AT) to ALA. Pyridoxal 5'-phosphate, a vitamin B₆-derivative, is an essential component (cofactor) of both ALAS and GSA-AT, as these enzymes cannot function relying solely on the protein moiety. Among prokaryotes, ALAS is only found in the α -proteobacteria, which are considered to be the most closely related, free-living organisms to the

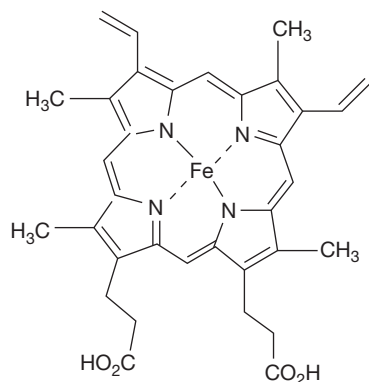


Figure 1 Iron-protoporphyrin IX, which is commonly designated as heme. Iron-protoporphyrin IX (or heme) has an iron ion complexed with the conjugated, aromatic porphyrin ring and an asymmetrical side-chain substitution. Note the $-\text{CH}_3$ (methyl), $-\text{CH}=\text{CH}_2$ (vinyl), and $-\text{CH}_2-\text{CH}_2-\text{COOH}$ (propionate) groups as side chains.

ancestral progenitor of the mitochondrion. Thus, it is conceivable that eukaryotic ALASs were acquired from a free-living α -proteobacterium during its transformation, first to an endosymbiont and subsequently to a mitochondrion.

From ALA to Uroporphyrin III

In eukaryotes, the first and three final enzymatic steps of heme biosynthesis occur within mitochondria, while the intervening four occur in the cytosol (**Figure 2**). Once ALA is made and transported to the cytosol, through most likely a specialized mitochondrial transporter, two molecules of ALA are asymmetrically condensed to form the monopyrrole porphobilinogen (PBG) by PBG synthase (PBGs; also known as ALA dehydratase). PBG constitutes the building unit of tetrapyrroles and hence eight molecules of ALA are required to form one heme molecule. Crystal structures of PBGs from different species have confirmed the presence of two distinct binding sites for each ALA molecule, termed A and P to distinguish the two ALA substrate molecules. The A-side of ALA forms the acetyl half of PBG, leaving its amino group intact, while the P-side of ALA

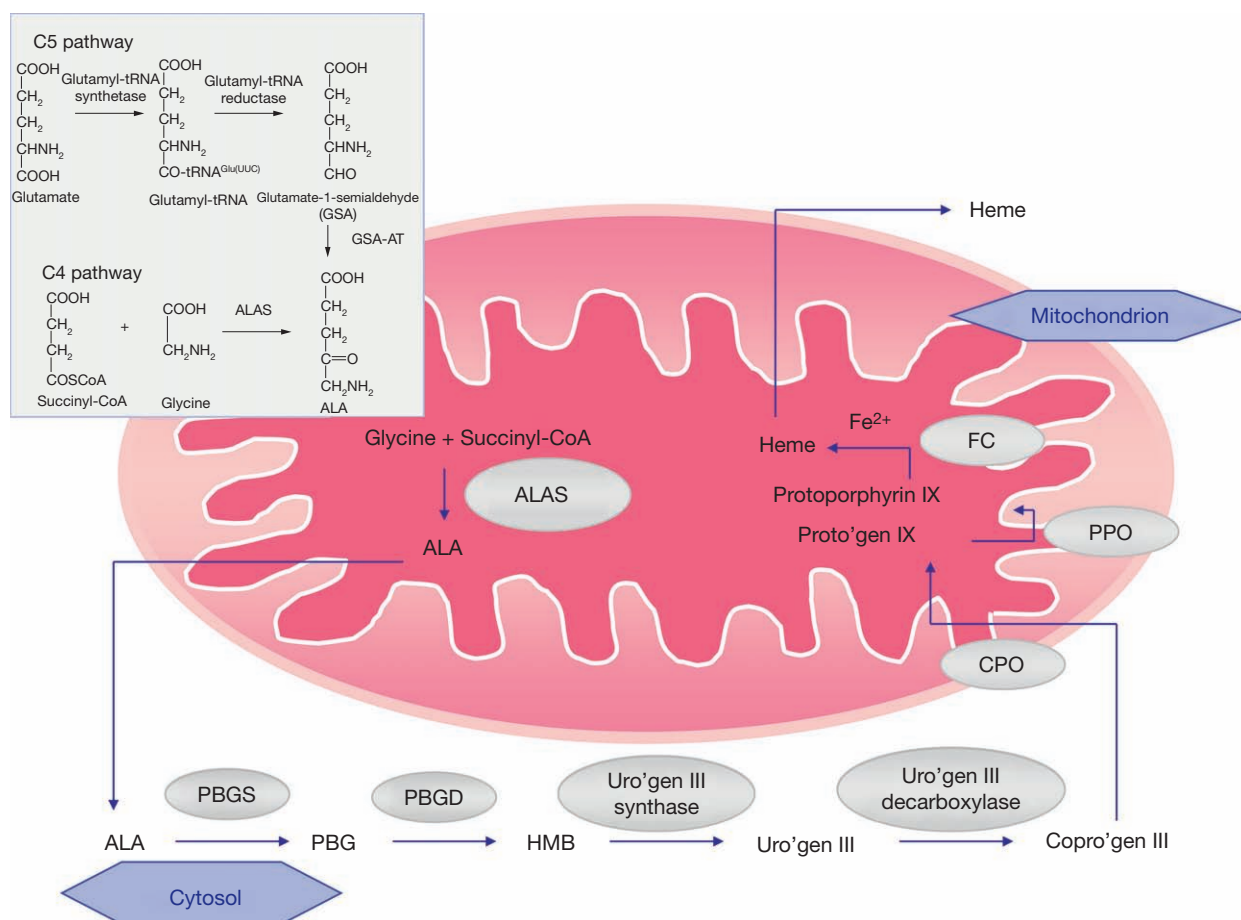


Figure 2 Diagrammatic representation of the heme biosynthetic pathway in eukaryotic animal cells. ALA, 5-aminolevulinate; ALAS, ALA synthase; PBG, porphobilinogen; PBGS, PBG synthase; PBGD, PBG deaminase; HMB, hydroxymethylbilane; Uro'gen III, uroporphyrinogen III; copro'gen III, coproporphyrinogen III; CPO, coproporphyrinogen III oxidase; proto'gen IX, protoporphyrinogen IX; PPO, protoporphyrinogen IX oxidase; FC, ferrochelatase. Inset: The two paths to ALA.

forms the propionyl half of PBG with the amino group becoming part of the PBG ring. PBGS has a homo-octomeric structure composed of four dimers, and each subunit contains an active site metal ion (zinc or magnesium and/or potassium depending on the species) and in many cases an allosteric metal ion (magnesium). The next step in heme biosynthesis involves the deamination and stepwise polymerization of four PBG molecules by PBG deaminase to give the linear tetrapyrrole 1-hydroxymethylbilane. Then, uroporphyrinogen III synthase catalyzes the cyclization of the linear tetrapyrrole, whereby the D ring of hydroxymethylbilane is inverted to generate the asymmetrical macrocycle, uroporphyrinogen III. In the absence of uroporphyrinogen III synthase, 1-hydroxymethylbilane cyclizes into the nonphysiological relevant tetrapyrrole, uroporphyrinogen I. The incredible acrobatics that must occur with the flipping of the D ring and closure of hydroxymethylbilane speak for the importance of uroporphyrinogen III synthase in making uroporphyrinogen III synthesis reaction. In fact, disastrous consequences emerge with the malfunction of uroporphyrinogen III synthase, as observed in patients with the blood disorder congenital erythropoietic porphyria, who typically have high levels of uroporphyrinogen I and derivatives thereof.

From Uroporphyrinogen III to Protoporphyrinogen IX

For most eukaryotes, the last cytosolic step is the sequential decarboxylation of the four acetate side chains of uroporphyrinogen III to methyl groups by uroporphyrinogen decarboxylase (Figure 2). This enzyme is a $(\alpha/\beta)_8$ barrel protein, which appears to be functional as a homodimer and not to require prosthetic groups or cofactors. Remarkably, uroporphyrinogen decarboxylase enhances the rate of substrate decarboxylation by a factor of 1.2×10^{17} , one of the largest rate enhancements reported for an enzyme that functions without the help of cofactors. The uroporphyrinogen decarboxylase-catalyzed reaction gives coproporphyrinogen III, which is the substrate of coproporphyrinogen oxidase (CPO). The oxidative decarboxylation of the propionate side chains of the rings A and B of coproporphyrinogen III yields protoporphyrinogen IX. While eukaryotic CPO activity depends on oxygen, prokaryotic organisms, under anaerobic conditions, use alternative electron acceptors in the oxidative decarboxylation reaction. Hence, oxygen-independent CPOs (HemN and HemZ, i.e., the products of the *hemN* and *hemZ* genes) are found in bacteria; indeed, facultatively anaerobic bacteria, such as *Escherichia coli*, possess both oxygen-dependent CPO (HemF) and oxygen-independent CPO (HemN). HemN contains an iron-sulfur center ($[4\text{Fe}-4\text{S}]$), which is essential for the activity of the enzyme in conjunction with nicotinamide adenine dinucleotide (phosphate), reduced (NAD(P)H) and S-adenosyl-L-methionine (SAM). As one would expect, the enzyme appears to have a distinct reaction mechanism with the involvement of SAM in the formation of a radical, and thus HemN has been aptly identified as a radical SAM enzyme. As far as the subcellular location of the oxygen-dependent CPO goes, it seems to vary! In mammals and other higher eukaryotes, it appears to be associated with the inner side of the mitochondrial outer membrane, whereas in the yeast *Saccharomyces cerevisiae* and bacteria it is found in the cytosol.

From Protoporphyrinogen IX to Heme

In the penultimate step, protoporphyrinogen oxidase (PPO) catalyzes the complete aromatization of protoporphyrinogen IX ring into protoporphyrin IX, which involves a six-electron oxidation (Figure 2). Similar to CPO, anaerobic and facultative anaerobic bacteria have an oxygen-independent PPO; however, in contrast to CPO, facultative anaerobic bacteria possess solely an oxygen-independent PPO (and not in complementation with an oxygen-dependent PPO). All other organisms appear to have PPOs, which use oxygen as the terminal electron acceptor. Ferrochelatase catalyzes the terminal step of the heme biosynthetic pathway, the insertion of ferrous iron into the protoporphyrin IX macrocyclic ring. Eukaryotic ferrochelatases associate with the inner side of the mitochondrial inner membrane, and vertebrate enzymes have a $[2\text{Fe}-2\text{S}]$ cluster (although this is not an exclusive structural feature of vertebrate ferrochelatases). The mechanism of the ferrochelatase reaction involves an out-of-plane distortion of the porphyrin, which facilitates the reaction by exposing the pyrrole nitrogens to the incoming ferrous iron substrate.

Regulation of the Heme Biosynthetic Pathway/Heme Biosynthesis

There is not a single ubiquitous regulatory mechanism for heme biosynthesis, and so the discussion in this article focuses solely on the regulation of the pathway in mammalian cells. Considering that heme is predominantly synthesized in bone marrow (i.e., in erythroblasts and reticulocytes that still contain mitochondria) and liver cells, most of the studies on regulation have focused on these two types of cells. Predominance of heme synthesis in bone marrow and liver cells reflects the greater demands due to the syntheses of the heme-containing proteins, hemoglobin, and cytochrome P-450 enzymes, respectively. The major regulatory and rate-limiting step of the pathway corresponds to the reaction catalyzed by ALAS, the first enzyme of the pathway. Two genes, located in two different chromosomes in humans, encode the two ALAS forms of the enzyme (or isozymes): ALAS1 and ALAS2. The gene encoding ALAS1 is expressed in every cell, whereas that for ALAS2 is specifically expressed in erythroid cells. Importantly, the regulatory mechanisms for the expression of the two ALAS genes and the posttranscriptional events seem to be distinct in differentiating erythrocytes from those in nonerythroid cells. ALAS1 controls the production of heme for basic cellular functions and is feedback regulated by heme through (1) repression of transcription of the ALAS1 gene, (2) inhibition of translation of the ALAS1 message, (3) destabilization of the ALAS1 message, and (4) inhibition of import of the cytosolic precursor form of the enzyme into mitochondria. By contrast, certain drugs and hormones induce liver cells to make more ALAS, and consequently heme and cytochrome P-450, by activating transcription of the ALAS1 gene.

In differentiating erythroid cells, the syntheses of heme and globin need to be coordinated, as the production of hemoglobin requires that neither the protein (globin) component nor the nonprotein component (heme) will be in excess. And indeed, identical transcriptional elements have been identified

in the genes of ALAS2 and globin. Another level of regulation of the erythroid heme biosynthetic pathway is centered at the stage of translation, which responds to the cellular iron levels. This response is mediated through specific interactions between an iron-responsive element (IRE), present in the ALAS2 messenger RNA (mRNA), and IRE-binding proteins, such that under iron limitation, the translation of ALAS2 mRNA is inhibited, while the presence of iron above the cellular threshold relieves the repression caused by IRE-binding proteins. The essential requirement for ALAS2 in heme biosynthesis and erythropoiesis was clearly demonstrated with targeted disruption of ALAS2 in the mouse, which led to no hemoglobin in cells, the accumulation of iron (albeit in the cytoplasm and not in mitochondria) and, ultimately, embryonic death.

Heme Function

The known roles of heme are many and variable, and most probably many other functions await discovery. As referred to above, heme is a prosthetic group of proteins involved in oxygen transport such as hemoglobins and myoglobins, which are the major blood and muscle proteins, respectively. In addition, heme participates in other cellular processes including anaerobic and aerobic respiration, sensing of diatomic gases (e.g., nitric oxide and carbon monoxide), detoxification of foreign substances (e.g., drugs) by oxidative metabolism, and signal transduction. Some of these functions might also be coupled with the regulatory roles of heme in gene transcription, translation, microRNA (miRNA) processing, mitochondrial protein import, protein stability, and differentiation. A transcription factor that regulates the mammalian biological clock was also discovered to use heme as a cofactor to sense, and react to, changes in the cell, thus linking cellular metabolism with the biological clock. Another area of research has led to the emergence of a new proposal that heme deficiency or dysregulation is intimately associated with the aging process, including neural decay.

Disorders of the Heme Biosynthetic Pathway

Malfunction of the heme biosynthetic pathway can have disastrous consequences, and in humans it can be manifested in porphyrias. Briefly, porphyrias are disorders that result from inherited or acquired defects in the heme biosynthetic pathway. The first reports of porphyrias were in 1874. Since then, eight types of porphyrias associated with defects in each of the eight enzymes of the pathway have been identified (Table 1), and depending on the major site of expression of the enzymatic defect, they have been classified into hepatic or erythropoietic porphyria. Although the molecular mechanisms behind the clinical manifestations are not well understood, the two major symptoms include cutaneous photosensitivity and central nervous system disturbances, and thus the other classification of either photosensitive or neurological porphyrias. Some porphyrias, however, manifest overlapping symptoms. Photosensitivity is a result of porphyrin accumulation, which typically occurs when the defective enzyme is

Table 1 Heme biosynthesis-associated disorders

Disorder	Defective enzyme
X-linked sideroblastic anemia	5-Aminolevulinate synthase (erythroid)
X-linked dominant protoporphyria	5-Aminolevulinate synthase (erythroid)
ALA dehydratase, or Doss, porphyria	Porphobilinogen synthase (or ALA dehydratase)
Acute intermittent porphyria	Porphobilinogen deaminase
Congenital erythropoietic porphyria	Uroporphyrinogen III synthase
Porphyria cutanea tarda	Uroporphyrinogen decarboxylase
Coproporphyria	Coproporphyrinogen oxidase
Variegate porphyria	Protoporphyrinogen oxidase
Erythropoietic protoporphyria	Ferrochelatase

anywhere in the pathway beyond the third step of the pathway (i.e., the PBG deaminase-catalyzed reaction), whereas the neurologic manifestations are associated with overproduction of ALA and PBG. However, the first porphyria associated with erythroid ALAS, named X-linked dominant protoporphyria (XLDPP), was recently identified, and elevated concentration of ALA-derived protoporphyrin IX and Zn-protoporphyrin were observed in the erythrocytes of the XLDPP patients. Since the 1980s, many of the genetic mutations responsible for the enzymatic defects have been identified and mouse models, targeted to correct the defective enzyme, have been developed by engineering functional enzyme expression in bone marrow cells. Zebrafish have also provided powerful molecular models for erythropoietic porphyrias. For example, mutations in the ferrochelatase gene resulted in photosensitive zebrafish with autofluorescent blood. Finally, mutations in the gene encoding for the erythroid ALAS can cause X-linked sideroblastic anemia (Table 1), a blood disorder characterized by mitochondrial iron accumulation due to a reduction in heme biosynthesis destined for hemoglobin production in erythroblasts.

See also: **Metabolism** Vitamins and Hormones: Porphyrin Metabolism; **Protein/Enzyme Structure Function and Degradation**: Heme Proteins; **Signaling**: Hematopoietin Receptors.

Further Reading

- Ajioka RS, Phillips JD, and Kushner JP (2006) Biosynthesis of heme in mammals. *Biochimica et Biophysica Acta* 1763: 723–736.
- Bottomley SS (2008) Sideroblastic anaemias. In: Greer JP, Foerster J, Rodgers GM, et al. (eds.) *Wintröbe's Clinical Hematology*, 12th edn., pp. 835–873. Philadelphia, PA: Lippincott Williams and Wilkins.
- Ferreira GC (1999) 5-Aminolevulinate synthase and mammalian heme biosynthesis. In: Ferreira GC, Moura JG, and Franco R (eds.) *Iron Metabolism. Inorganic Biochemistry and Regulatory Mechanism*, pp. 15–34. Weinheim, Germany: Wiley-VCH.
- Ferreira GC and Hunter GA (2011) Ferrochelatase structure and reaction mechanism. In: Kadish KM, Smith KM, and Guillard R (eds.) *Handbook of Porphyrin Science*, pp. 49–121. Hackensack, NJ: World Scientific Publishing.
- Latunde-Dada GO, Simpson RJ, and McKie AT (2006) Recent advances in mammalian haem transport. *Trends in Biochemical Sciences* 31: 182–188.
- Severance S and Hamza I (2009) Trafficking of heme and porphyrins in metazoa. *Chemical Reviews* 109: 4596–4616.

Hexokinase/Glucokinase

E Van Schaftingen, Université Catholique de Louvain, Brussels, Belgium

© 2013 Elsevier Inc. All rights reserved.

Glossary

Glucokinase Hexokinase that acts specifically on glucose; common designation of mammalian hexokinase IV.

Hexokinase Enzyme catalyzing the adenosine triphosphate-dependent phosphorylation of glucose and related hexoses (mannose, 2-deoxyglucose, and fructose) to their 6-phospho-derivative.

Mnemonic and slow transition models Kinetic models allowing one to explain the nonhyperbolic saturation curve of some monomeric enzymes, most particularly the sigmoid saturation curve of mammalian glucokinase. These models assume the existence of two conformations with distinct affinities for the substrate and that they interconvert slowly. Only the high-affinity conformation is endowed with catalytic activity in the mnemonic model.

Role of Phosphorylation in Sugar Utilization

In order to be utilized by cells, monosaccharides need to be transported across the plasma membrane and phosphorylated. In many bacteria, transport of glucose and other sugars is coupled to their phosphorylation by complex systems utilizing the high-energy phosphoryl donor phosphoenolpyruvate. By contrast, transport and phosphorylation – in this case, at the expense of adenosine triphosphate (ATP) – are catalyzed by separate proteins in eukaryotic cells and also to some extent in bacteria. Sugars are most often phosphorylated on their last carbon (D-glucose to glucose 6-phosphate; ribose to ribose 5-phosphate, etc.), although, in some cases, phosphorylation takes place on the first carbon (galactose, L-fucose, and fructose in liver).

Phosphorylation provides several advantages: it is an exergonic process, capable of concentrating metabolites inside the cells; it prevents leakage of metabolites across the plasma membrane and intracellular membranes; and it increases the potential binding energy of substrates to enzymes, thereby facilitating catalysis. The name ‘hexokinase’ (from hexose and the Greek word, *κινεειν*, to move) was coined by Otto Meyerhof to designate the enzyme that activates glucose, making it ready for fermentation by muscle extracts. Kinase was to become the word used to designate all ATP-dependent phosphorylating enzymes.

Reaction Catalyzed by Hexokinases

Hexokinases (EC 2.7.1.1) are defined as enzymes phosphorylating glucose and other closely related hexoses (mannose, 2-deoxyglucose, glucosamine, and fructose) on their sixth carbon. Glucokinases (2.7.1.2) catalyze the same reaction but with only glucose as a substrate. Glucokinases using adenosine diphosphate (ADP) or polyphosphate as phosphoryl donor have been described in archaeobacteria (*Pyrococcus furiosus* and *Thermococcus litoralis*) and even in mammals; these will not be discussed further here.

Hexokinases and glucokinases catalyze the transfer of the terminal phosphoryl group of ATP to the oxygen borne

by the sixth carbon of glucose. As with other kinases, the true phosphoryl donor is the ATP–Mg complex. Except for inosine triphosphate (ITP), other triphosphonucleotides are usually much poorer substrates than ATP. ADP is generated in the reaction, as well as a hexose 6-phosphate. The reaction is highly exergonic ($\Delta G^{\circ'} \approx -16.7 \text{ kJ mol}^{-1}$), which means that in practical terms it is irreversible, most particularly under intracellular physiological conditions ($\Delta G < -30 \text{ kJ mol}^{-1}$). However, reversibility of the reaction can be demonstrated *in vitro* provided the generated ATP and glucose are continuously consumed.

The reaction mechanism involves the formation of a ternary complex and an in-line transfer of the phosphoryl group from the donor to the acceptor without formation of an enzyme-bound covalent intermediate. Many studies have been devoted to the determination of the order of substrate addition and product release, without any emerging consensus. It is likely that most hexokinases and glucokinases may bind the substrates in any order to produce a reactive ternary complex, although some enzymes display a preferred order (e.g., glucose before ATP in the case of mammalian hexokinase IV).

Diversity of Hexokinases and Glucokinases

Eukaryotes usually have several hexokinases. The yeast *Saccharomyces cerevisiae* has two hexokinases (P_I and P_{II} or A and B) as well as a glucokinase (Table 1). All three enzymes have subunits of 50 kDa, the two hexokinases being dimers of 100 kDa. The two hexokinases share ≈75% sequence identity with each other and only ≈37% sequence identity with glucokinase from the same species. All three enzymes display hyperbolic saturation curves for their substrates and a K_m for glucose <1 mM. Both hexokinases A and B are inhibited by trehalose 6-phosphate, an intermediate in the formation of trehalose, which appears to play the role of a feedback inhibitor for this enzyme (equivalent to the role of glucose 6-phosphate with mammalian hexokinases).

Mammals are classically said to have four distinct hexokinases, designated I, II, III, and IV, or A, B, C, and D. A fifth one, named ‘HKDC1’ (hexokinase domain containing 1), is an

Table 1 Some well-studied hexokinases and glucokinases

	Approximate subunit size (kDa)	Number of subunits	K_m for glucose	Inhibitor	Tissular distribution
<i>Mammals</i>					
Hexokinase I (A)	100	1	$\approx 10 \mu\text{M}$	Glucose 6-phosphate	Brain Erythrocytes (all tissues)
Hexokinase II (B)	100	1	$\approx 100 \mu\text{M}$	Glucose 6-phosphate	Heart Skeletal muscle Adipocytes
Hexokinase III (C)	100	1	$\approx 1 \mu\text{M}$	Glucose 6-phosphate Glucose	White blood cells
Hexokinase IV (D) (glucokinase)	50	1	$\approx 8 \text{ mM}$	Glucokinase regulatory protein (fructose 6-phosphate)	Liver β - and α -cells of pancreatic islets hypothalamic neurons
<i>Invertebrates</i>					
<i>Asterias amurens</i> (Starfish) hexokinase	50	1	$\approx 50 \mu\text{M}$	Glucose 6-phosphate	
<i>Saccharomyces cerevisiae</i>					
Hexokinase A (or PI)	50	2	$\approx 100 \mu\text{M}$	Trehalose 6-phosphate	
Hexokinase B (or PII)	50	2	$\approx 100 \mu\text{M}$	Trehalose 6-phosphate	
Glucokinase	50	(aggregates)	$\approx 30 \mu\text{M}$		

uncharacterized protein sequence deduced from genomes. Hexokinases I–III are ≈ 100 -kDa monomeric proteins, which resulted from the duplication of an ancestral gene encoding a 50-kDa hexokinase. They are characterized by low K_m values for glucose ($<0.2 \text{ mM}$) and by their sensitivity to inhibition by physiological concentrations of glucose 6-phosphate. Hexokinase III has an extremely low K_m for glucose and is inhibited by high concentrations of this substrate. Only the C-terminal half of hexokinase I appears to be catalytically active, whereas both amino and carboxyl halves of hexokinase II are endowed with catalytic activity.

Hexokinase IV or D is a ≈ 50 -kDa protein that has a low affinity for glucose, a sigmoidal saturation curve for this substrate and is not inhibited by physiological concentrations of glucose 6-phosphate. It is, however, inhibited by glucokinase regulatory protein. It does catalyze the phosphorylation of other hexoses than glucose, but due to its low affinity for all hexoses (including glucose), its physiological action is restricted to the most concentrated one, glucose, which is also its best substrate. Furthermore, there is ample evidence that hexokinase D plays the role of a glucose sensor. This may explain the popularity of its designation as glucokinase, which however does not mean that this enzyme is closer to true (microbial) glucokinases than to hexokinases. As a matter of fact, hexokinase IV is closer to mammalian low K_m hexokinases than to any other hexokinases or glucokinases.

Distinct hexokinase isozymes are also found in insects (e.g., in *Drosophila melanogaster*) and plants. The starfish enzyme is a 50-kDa protein that is inhibited by physiological concentrations of glucose 6-phosphate, indicating that this property is not limited to the 100-kDa enzymes.

Prokaryotes typically contain a series of specific hexokinases that each act on one hexose, normally glucose, mannose, or fructose. They are often composed of two subunits with a smaller size (20–35 kDa) than the eukaryotic enzymes.

There are, however, examples of true hexokinases in bacteria (e.g., the hexokinase of the archaeobacterium *Thermoproteus tenax*) or of enzymes catalyzing the phosphorylation of two hexoses but not glucose (e.g., mannofructokinase).

Classification and Evolution of Hexokinases

Sugar kinases can be classified into three distinct families (hexokinase, ribokinase, and galactokinase) of enzymes that differ in their three-dimensional structures and show no significant similarity. In the hexokinase family, the eukaryotic enzymes form a distinct subfamily of enzymes sharing $>30\%$ sequence identity. The hexokinase family also comprises bacterial enzymes acting on various substrates (hexokinases, glucokinases, mannokinases, fructokinases, and *N*-acetylmannosamine kinases) that are very distantly related to eukaryotic hexokinases.

The ancestral eukaryotic hexokinase was probably a 50-kDa protein (Figure 1). Early and late duplication events in the fungi lineage led to the appearance of yeast glucokinase and of the more closely related hexokinases A and B. In vertebrates, hexokinase IV (glucokinase) separated from hexokinases I–III before the gene duplication and fusion event that led to the 100-kDa enzyme. Hexokinase III branched off earlier than the duplication event that led to hexokinases I and II, consistent with the closer kinetic properties of the latter two enzymes compared to hexokinase III.

Three-Dimensional Structure

Yeast hexokinase A and B structures were among the first such structures determined during the 1970s. Yeast hexokinase has a bilobar structure with a large (primarily C-terminal) and a small (primarily N-terminal) lobe that are connected by a

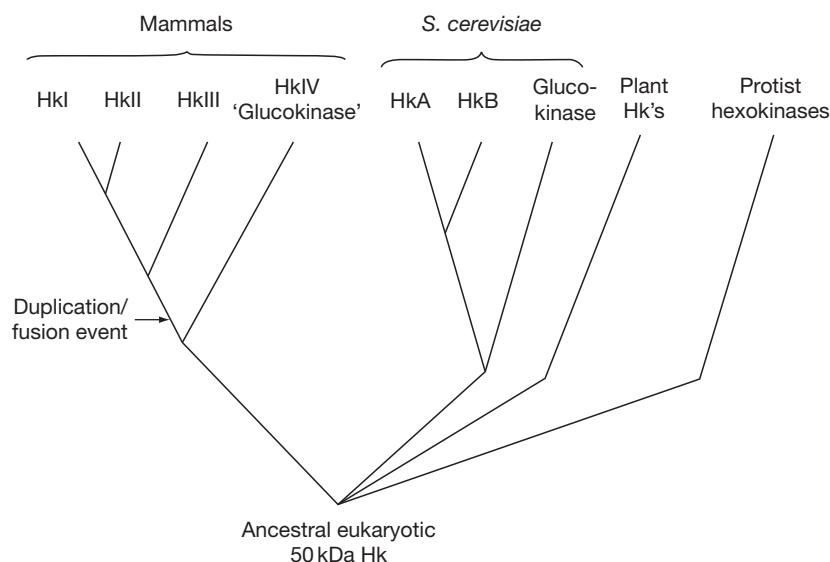


Figure 1 Evolutionary relationship between some eukaryotic hexokinases and glucokinases. Hk, hexokinase.

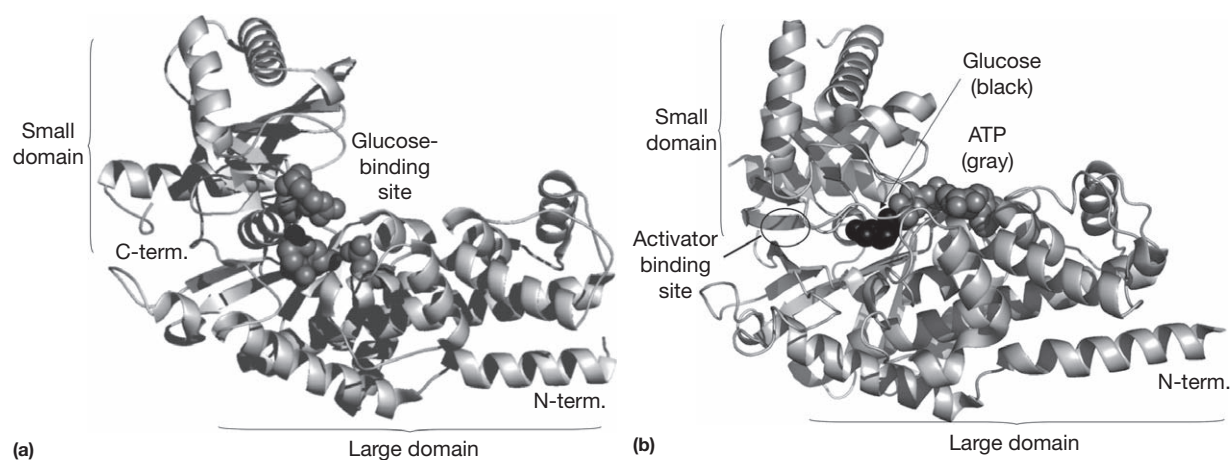


Figure 2 Three-dimensional structures of yeast hexokinase (a) and human glucokinase (b). The structures shown are those of *Saccharomyces cerevisiae* hexokinase B (P-II) isozyme in its open conformation and human glucokinase in its closed conformation. Glucose-binding residues are shown as space-filling models in yeast hexokinase. Glucose (black) and an ATP analog (gray) are shown as space-filling models in human glucokinase. The location of the binding site for glucokinase activators is also indicated. From Kuser PR, Krauchenco S, Antunes OA, and Polikarpov I (2000) The high resolution crystal structure of yeast hexokinase PII with the correct primary sequence provides new insights into its mechanism of action. *Journal of Biological Chemistry* 275: 20814–20821.

hinge region. Both domains are made of β -sheets of similar topologies, flanked by α -helices (Figure 2). These are more abundant and larger in the C-terminal domain, accounting for its larger size. Between the two lobes is a deep crevice where the catalytic site resides. The enzyme exists under two distinct conformations, described as open and close, which differ by the diedral angle between the two lobes.

The glucose-binding site comprises side chains of extremely conserved residues of the hinge region (Asp211 in hexokinase B), the large lobe (Asn237, Glu269, and Glu302) and the small lobe (Ser158, Thr172, and Arg173). Binding of glucose to the large lobe induces closure of the catalytic cleft, allowing glucose to

contact the small lobe as well. The conformational change induced by this substrate is thought to facilitate the subsequent phosphoryl group transfer, being a nice illustration of the induced-fit theory. This theory is also supported by the observation that lyxose, a five-carbon substrate analog lacking the hydroxymethyl group that normally serves as phosphoryl acceptor, stimulates by ≈ 30 -fold the hydrolysis of ATP to ADP and Pi catalyzed by yeast hexokinase.

The binding site of hexokinases for ATP has been initially inferred from structural analogy of hexokinase with glycerokinase, HSP-ATPase, and actin, enzymes that have a similar bilobar architecture as hexokinase. It was confirmed by

site-directed mutagenesis and later by crystal structures with ATP analogs. ATP binds mainly to the larger domain, almost parallel to its main axis, with the gamma-phosphate group pointing to the hinge region.

The crystal structures of human hexokinase I and of the enzyme from *Schistosoma mansoni*, a glucose 6-phosphate-sensitive enzyme, are fundamentally the same as the structure of yeast hexokinase, repeated twice in the case of the mammalian enzyme. This is also true for mammalian glucokinase (Figure 2). The human and *S. mansoni* hexokinases were also crystallized as complexes with glucose 6-phosphate, showing that the binding site for this inhibitor overlaps with the putative ATP-binding site.

Mammalian Low- K_m Hexokinases

The activity of the low K_m hexokinases cannot be regulated by the glucose concentration in tissues or cell types (human erythrocytes) where the glucose concentration is well above their K_m . In other cell types (e.g., the adipocytes), their activity may well be controlled by the glucose concentration and therefore by glucose transport. There is, however, still a large degree of uncertainty as to the glucose concentration and the rate-limiting aspect of glucose transport in several cell types.

Inhibition by Glucose 6-Phosphate

Hexokinases I–III are inhibited by glucose 6-phosphate in a noncompetitive manner with respect to glucose, but competitive versus ATP. Inhibition by glucose 6-phosphate takes place at physiological concentrations, which argues for its physiological significance. It is antagonized by inorganic phosphate in the case of hexokinase I, but not of hexokinase II. This inhibition involves a site distinct from the glucose-binding site, as indicated by the finding that the phosphorylation products (mannose 6-phosphate and 2-deoxyglucose 6-phosphate) of several good substrates for hexokinase I are poor inhibitors compared to glucose 6-phosphate. At one time, it was thought that the (catalytically inactive) aminoterminal half of hexokinase contained the glucose 6-phosphate allosteric site; however, the realization that the isolated carboxyterminal half of the enzyme, as well as starfish hexokinase (a 50-kDa enzyme), was sensitive to glucose 6-phosphate inhibition argued against this hypothesis. Structural studies indicate that glucose 6-phosphate binds to the two halves of hexokinase I, in positions that overlap the putative ATP-binding sites. This agrees with the fact that the inhibition by glucose 6-phosphate is competitive versus ATP but not versus glucose. Site-directed mutagenesis experiments suggest, however, that the binding of glucose 6-phosphate to both sites may contribute to inhibition in hexokinase I.

Subcellular Localization of Mammalian Hexokinases

Hexokinases are cytosolic enzymes, but hexokinases I and II bind to mitochondria through a N-terminal hydrophobic region that interacts with voltage-dependent anion channel (VDAC). VDAC, a porin present in the outer mitochondrial membrane, is involved in the transport not only of

low-molecular-mass solutes, but also of proteins such as cytochrome *c*, which escape mitochondria to elicit apoptosis. Hexokinase II, which is highly expressed in numerous cancer cells, exerts antiapoptotic effects through its binding to VDAC. Their proximity with the mitochondrial pore makes also that hexokinases I and II are closer to the main source of ATP and to the main sink for ADP, thus facilitating diffusion. Furthermore, binding of hexokinase I to mitochondria decreases its affinity for the inhibitor glucose 6-phosphate.

Hexokinase IV (Glucokinase)

Sigmoidal Kinetics

One intriguing aspect of glucokinase, a monomeric enzyme, is that it displays a sigmoidal saturation curve for glucose, with a Hill coefficient of 1.6. This is not due to the presence of a second, allosteric binding site for glucose on the glucokinase monomer, but because of the existence of an additional conformation compared to hexokinases with a hyperbolic saturation curve. This additional conformation is a super-open state (Figure 3), in which the small domain is disorganized. The super-open state presumably binds glucose with very low affinity and only slowly reverts to the open and closed states. Its formation from the open state also occurs slowly and only when glucose is not bound. This denotes that the low-affinity conformation predominates at low glucose concentration. Raising the glucose concentration has a twofold effect: (1) it increases the proportion of enzyme that is able to bind glucose with high affinity (i.e., the enzyme in the open conformation) and (2) it promotes binding of glucose to the open conformation as it would do for any other hexokinase. This twofold effect is at the basis of the sigmoidal saturation curve. The slowness of the conformational changes is critical in this model. Kinetic models of cooperativity based on slow conformational changes are known as 'mnemonical models' and 'slow-transition models'.

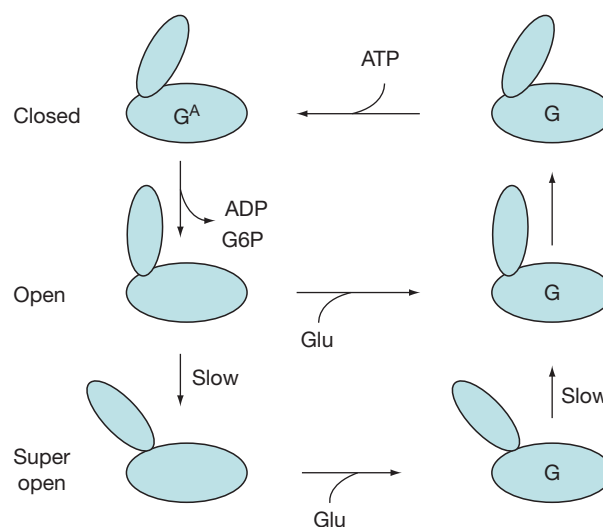


Figure 3 Kinetic model of glucokinase. For details, see text. A, ATP; G or Glu, glucose; G6P, glucose 6-phosphate.

The existence of a super-open state, with no or little affinity for glucose, explains also how glucokinase has a much lower apparent affinity for glucose compared to other hexokinases, while all its glucose-binding residues are the same as those present in low K_m hexokinases. The best proof for this is the identification of a series of synthetic glucokinase activators that decrease the K_m value from ≈ 8 mM to about 0.1 mM, increase V_{max} , and suppress cooperativity. These activators bind to the open and closed states of glucokinase, though not to the super-open state, such that the latter no longer exists when the activators are present at saturating concentrations. Thus, the activators make glucokinase behave as a low K_m hexokinase. Their binding site has been identified in the hinge region of this enzyme (Figure 2). Because of their effect on islet cells and liver glucokinase, they stimulate *in vivo* both insulin secretion and glucose uptake by the liver, making them potentially useful therapeutic agents for the treatment of diabetes.

Liver Glucokinase and Its Regulatory Protein

One of the main sites of glucokinase expression is the liver. This is in relation with the ability of this organ to store glucose as glycogen. Glucose transport across the plasma membrane of liver parenchymal cells, carried out by glucose transporter 2 (GLUT-2), is extremely rapid. Glucokinase may therefore play a regulatory role and indeed the low affinity that glucokinase displays for glucose is one of the mechanisms by which glucose uptake can be regulated by glucose concentration. In addition to this, glycogen metabolism is under the direct control of glucose, which binds to glycogen phosphorylase, thereby affecting the phosphorylation state of this enzyme and glycogen synthase.

Hepatic glucokinase is controlled by a glucokinase regulatory protein (GCKR), which binds to glucokinase and inhibits it in a competitive manner with respect to glucose. Association of the two proteins is stimulated by fructose 6-phosphate, a GCKR ligand, which behaves therefore as an inhibitor of glucokinase, and antagonized by fructose 1-phosphate, another GCKR ligand. Regulation by GCKR makes therefore that glucokinase in intact liver cells has a lower affinity for glucose than the purified enzyme, and that low concentrations of fructose stimulate glucose utilization by the liver. Interestingly, a sequence variant that is commonly found in the human population alters the affinity of GCKR for fructose 6-phosphate and is reproducibly associated with differences in triglyceride and fasting plasma glucose, indicating the importance of GCKR in blood glucose and lipid control.

GCKR also appears to control the subcellular localization of glucokinase: this enzyme is predominantly in the nucleus at low concentration of glucose and in the absence of fructose, but moves to the cytosol when the glucose concentration increases or if fructose is added. The physiological meaning of glucokinase translocation is not known, but it may be to protect glucokinase against degradation by cellular proteases. This would be consistent with the finding that mice devoid of GCKR have lower amounts (and total activities) of glucokinase than control mice. A surprising feature of GCKR is that it

evolved from a prokaryotic enzyme involved in peptidoglycan recycling.

Role of Glucokinase in β -Cells

The low affinity that this enzyme displays for glucose (half-saturation is observed at 8 mM with the human enzyme), the sigmoidal saturation curve, and the lack of inhibition by physiological concentrations of glucose 6-phosphate make this enzyme well suited to respond to changes in glucose concentration and to act thereby as a glucose sensor. This role is well established in the case of the insulin-secreting cells (β -cells) of pancreatic islets. Glucose appears to be one of the main fuels for β -cells. Raising the glycolytic flux (e.g., by raising the glucose concentration) leads to an increase in the cytosolic [ATP]/[ADP] ratio, which in turn results in closure of a nucleotide-sensitive K^+ channel. The ensuing depolarization of the plasma membrane causes an increase in the cytosolic Ca^{2+} concentration, which triggers insulin secretion. Metabolic analyses indicate that glucokinase, not glucose transport or a step beyond glucose phosphorylation, plays a major role in the control of this system. Accordingly, a form of diabetes known as maturity-onset diabetes of the youth type II is caused by glucokinase mutations resulting in decreased enzymatic activity. The disease is dominantly inherited, which means that a mutation in one copy of the gene (resulting at most in a 50% decrease of glucokinase activity) is sufficient to perturb the glucose-sensing apparatus of the endocrine pancreas. Conversely, glucokinase mutations that increase the activity (mostly by increasing the affinity for glucose) result in familial hypoglycemia.

See also: [Metabolism Vitamins and Hormones: Glucose/Sugar Transport in Mammals](#); [Glycogen Metabolism](#); [Glycolysis Overview](#).

Further Reading

- Agius L (2008) Glucokinase and molecular aspects of liver glycogen metabolism. *Biochemistry Journal* 414: 1–18.
- Aleshin AE, Zeng C, Bartunik HD, Fromm HJ, and Honzatko RB (1998) Regulation of hexokinase I: Crystal structure of recombinant human brain hexokinase complexed with glucose and phosphate. *Journal of Molecular Biology* 282: 345–357.
- Beer NL, Tribble ND, McCulloch LJ, et al. (2009) The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. *Human Molecular Genetics* 18: 4081–4088.
- Cárdenas ML, Cornish-Bowden A, and Ureta T (1998) Evolution and regulatory role of the hexokinases. *Biochimica et Biophysica Acta* 1401: 242–264.
- Kamata K, Mitsuya M, Nishimura T, Eiki J, and Nagata Y (2004) Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase. *Structure* 12: 429–438.
- Matschinsky FM (2002) Regulation of pancreatic beta-cell glucokinase: From basics to therapeutics. *Diabetes* 51: S394–S404.
- Matschinsky FM (2009) Assessing the potential of glucokinase activators in diabetes therapy. *Nature Reviews Drug Discovery* 8: 399–416.
- Mulichak AM, Wilson JE, Padmanabhan K, and Garavito RM (1998) The structure of mammalian hexokinase-1. *Nature Structural and Molecular Biology* 5: 555–560.
- Van Schaftingen E, Detheux M, and Veiga da Cunha M (1994) Short-term control of glucokinase activity: Role of a regulatory protein. *FASEB Journal* 8: 414–419.
- Wilson JE (1995) Hexokinases. *Reviews of Physiology, Biochemistry and Pharmacology* 126: 65–198.

HIV-1 Reverse Transcriptase Structures

K Das and E Arnold, Rutgers University, Piscataway, NJ, USA
S H Hughes, National Institutes of Health, Frederick, MD, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Kalyan Das, Stefan G. Serafianos, Eddy Arnold, and Stephen H. Hughes, volume 2, pp. 388–392, © 2004, Elsevier Inc.

Glossary

Diarylpyrimidine (DAPY) A class of compounds that are potent non-nucleoside inhibitors; etravirine (TMC125) and rilpivirine (TMC278) are DAPY non-nucleoside RT inhibitors (NNRTIs).

Excision Related to the inverse of the polymerization reaction. Reverse transcriptase (RT) uses excision to remove a nucleoside RT inhibitor (NRTI) from the DNA primer terminus.

Non-nucleoside RT inhibitors (NNRTIs) Inhibitors that bind in a hydrophobic pocket near the polymerase active site and block DNA polymerization. Nevirapine, delavirdine, efavirenz, and etravirine are four clinically approved NNRTI drugs for treating human immunodeficiency virus 1 (HIV-1) infection.

Nucleoside RT inhibitors (NRTIs) Analogs of normal nucleosides, which lack a 3'-OH and act as terminators of DNA elongation. AZT (zidovudine) and 3TC (lamivudine) are among the approved NRTI drugs for treatment of HIV-infected patients.

Resistance mutation A change in the amino acid sequence that reduces the susceptibility of the enzyme to an inhibitor.

Reverse transcriptase (RT) A viral enzyme that converts single-stranded genomic RNA into double-stranded DNA.

Thymidine analog mutations (TAMs) A set of RT mutations, also known as AZTr, which facilitate excision to cause NRTI resistance.

Introduction

Human immunodeficiency virus 1 (HIV-1) reverse transcriptase (RT) is a heterodimer consisting of p66 (66 kDa) and p51 (51 kDa) subunits. Both subunits share a common N-terminal domain; p51 lacks the C-terminal RNase H domain present in p66. The p66 subunit of RT has a structure that has been compared to a right hand, having fingers, palm, thumb, and connection subdomains (**Figure 1**). While both the subunits contain the same subdomains, the arrangement of the subdomains is different in p66 and p51. The p51 subunit is not directly involved in DNA polymerization; however, it plays a structural role and forms part of the nucleic-acid-binding cleft that extends from the polymerase active site to the RNase H active site in p66. Non-nucleoside RT inhibitors (NNRTIs) bind to a hydrophobic pocket near the RT polymerase active site and act as allosteric inhibitors that block the chemical step of DNA polymerization. Over the past 20 years, significant progress has been made in understanding the structure and functions of RT. The functional states of RT that have been captured in structural studies can be broadly categorized into: (1) apo HIV-1/HIV-2 RT (in the absence of any bound substrate or inhibitor); (2) HIV-1 RT in complexes with NNRTIs; (3) RT–nucleic acid (double-stranded DNA (dsDNA) or RNA:DNA) binary complexes (with and without the Fab of a monoclonal antibody); (4) RT–nucleic acid–nucleoside triphosphate ternary complexes; (5) RT–nucleic acid–excision product (AZTppppA) complexes; and (6) RT–RNase H inhibitor (RNHI) complexes. Here, we discuss how these structures that have helped elucidate the enzymatic functions of RT, the molecular basis of drug resistance, and aided drug design.

Functional Implications of RT Structures

Comparison of the structures of RT–nucleic acid complexes with apo RT reveals that the thumb of p66 folds down onto the nucleic-acid-binding cleft in apo RT, and binding of a nucleic acid substrate causes the thumb to move out of the cleft to form a clamp that holds the nucleic acid (dsDNA/RNA:DNA). The NNRTI-binding site is in the palm subdomain of p66 near the base of the thumb; the bound NNRTI locks the p66 thumb in an upright position even in the absence of nucleic acid.

The structure of the RT–dsDNA complex revealed details of nucleic acid recognition including the template grip and the primer grip. The polymerase and RNase H active sites, defined by the catalytic carboxylates (Asp110, Asp185, and Asp186 for the polymerase and Asp443, Glu478, Asp498, and Asp549 for the RNase H active sites), both of which are Mg^{2+} dependent, are responsible for the nucleotidyltransferase that polymerizes DNA and the endonuclease that degrades the RNA strand after the complementary DNA strand has been synthesized. The two active sites are separated by 17 DNA base pairs; the nucleic acid substrate is bent at the region that interacts with p66 thumb. The structure of RT complexed with an RNA:DNA duplex led to the identification of the RNase H primer grip, a set of amino acids that interact with the DNA primer strand and help to determine the specificity of RNase H cleavage. The structure of the RT–RNA:DNA complex was determined using an RNase H-resistant RNA, known as the polypurine tract (ppt). During viral replication, the ppt is not cleaved by RNase H and the ppt serves as a primer for the synthesis of the second DNA strand. The RNase H cleavage specificity is presumably

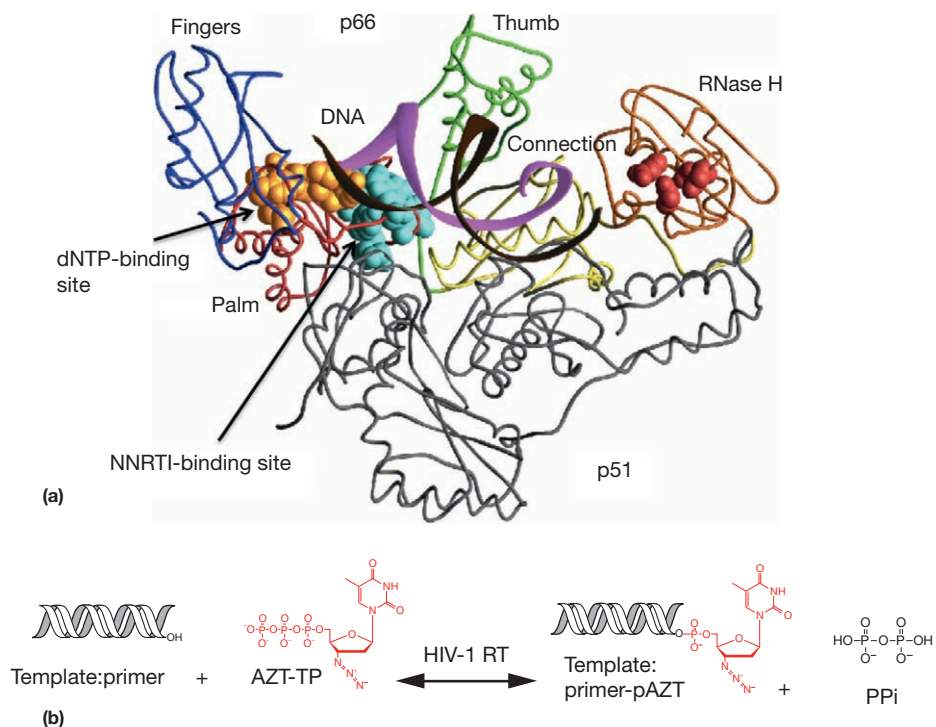


Figure 1 Structure and function of HIV-1 RT. (a) A cartoon representation of HIV-1 reverse transcriptase (RT). The subdomains in the p66 subunit are color-coded: fingers (blue), palm (red), thumb (green), connection (yellow), and the RNase H domain is orange; the p51 subunit is gray. The RNase H active site is shown in red, the site for binding incoming dNTP substrates is gold, and the site for binding NNRTIs is blue. The double-stranded (ds) DNA is shown as a double-helical ribbon with the template (brown) and the primer (purple), extending from the polymerase active site/dNTP-binding site to the RNase H active site. (b) A nucleotide triphosphate (or analog) binds RT and is incorporated at the 3'-end of the DNA primer strand releasing pyrophosphate; the drawing shows the incorporation of AZT-MP.

controlled by the width of the minor groove and the trajectory of the RNA:DNA duplex.

To obtain a structure of a ternary complex of RT with dsDNA and deoxyribose-nucleotidetriphosphate (dNTP), the groups of Stephen Harrison and Gregory Verdine developed a technique in which a disulfide bond was used to covalently link dsDNA to the protein at position 258 of the p66 thumb. Use of a dideoxy-nucleotide at the 3'-end of the primer prevented the incorporation of an incoming dNTP, trapping RT and the substrates in a state preceding DNA-dependent DNA polymerization. In this structure, a portion of the p66 fingers folds down to form part of the dNTP-binding site; the $\beta 3$ – $\beta 4$ loop of the fingers plays a critical role in dNTP binding. An analogous movement occurs in many other DNA polymerases (such as bacterial DNA polymerase I and murine leukemia virus RT, T7 RNA polymerase, and in the RNA-dependent RNA polymerases of poliovirus, foot-and-mouth disease virus, hepatitis C virus, rabbit hemorrhagic disease virus, reovirus, and bacteriophage $\phi 6$). The structural and functional characteristics of these enzymes are useful in understanding the general mechanism of nucleic acid polymerization. The RT–DNA cross-linking technique has been subsequently used to obtain crystals of various RT binary and ternary complexes that are important in understanding the mechanisms of nucleoside RT inhibitor (NRTI) drug resistance.

NNRTIs are highly specific inhibitors of HIV-1 RT; most do not inhibit HIV-2 RT. Four NNRTIs (nevirapine, delavirdine, efavirenz, and etravirine) have been approved for treating

HIV-1 infections. The structure of HIV-1 RT in a complex with nevirapine, from Thomas Steitz and co-workers, was the first structure of RT bound to an NNRTI. Since then, a large number of crystal structures of wild-type and drug-resistant RT–NNRTI complexes have been determined independently by several groups including ours and that of David Stammers and David Stuart. Included among the structures are of RT in complex with α -APA, TIBO, delavirdine, efavirenz, HBY 097, HEPT, PETT, TSAO, DATA, and DAPY derivatives. The conformation of RT in all these structures is similar, having an open nucleic-acid-binding cleft; however, there are significant variations in the shape and size of the NNRTI-binding pocket (NNIBP) related to the size and shape of the bound NNRTI.

Analysis of RT–NNRTI structures has helped in understanding the mechanism of RT inhibition by NNRTIs. NNRTIs are noncompetitive inhibitors that do not interfere with the binding of either the nucleic acid or dNTP substrates. No predefined NNIBP exists in the structures of RT that do not contain a bound NNRTI. Opening of the NNIBP involves large torsional displacements of the aromatic side chains of Y181 and Y188 and a rotation of the $\beta 12$ – $\beta 13$ – $\beta 14$ sheet that moves the primer grip away from the polymerase active site. Once bound, an NNRTI restricts the flexibility of the enzyme. This has two important consequences: (1) primer grip restriction that prevents RT from properly positioning the 3'-end of the DNA primer relative to the polymerase active site and (2) catalytic triad distortion, that is, distortion of the

catalytic triad (Asp110, Asp185, and Asp186) that prevents coordination of catalytically relevant metal ions (Mg^{2+}); there are no crystal structures of an RT–NNRTI complex that have a divalent metal ion bound at the polymerase active site.

NRTI Resistance

The most effective strategy for reducing HIV-1 viral load and keeping it at undetectable levels in patients is by using combinations of drugs (this treatment strategy is called highly active antiretroviral therapy, HAART) that usually include one or two RT inhibitors, sometimes two NRTIs. However, because treatment is lifelong, side effects and the emergence of drug-resistant HIV-1 strains are serious concerns. Resistant strains emerge because of the virus' rapid rate of replication and high rate of mutation. Most of the sites where there are mutations that are associated with drug resistance have the potential to influence (directly or indirectly) the binding of the drug or specific functional steps in reverse transcription. In a number of cases, structural studies in combination with biochemical data and clinical results have helped explain the molecular basis of drug resistance.

Even though all NRTIs lack a 3'-OH and are incorporated at the 3'-end of the DNA primer to inhibit DNA polymerization, there are specific NRTI-resistance mutations associated with treatments with different drugs (Figure 2). There are two primary classes of NRTI-drug resistance mechanisms: (1) exclusion – in which the mutation reduces the ability of RT to incorporate the triphosphate form of the NRTI and (2) excision – in which the mutation enhances the ability of RT to remove an incorporated NRTI. M184I/V mutations are exclusion mutations that emerge in patients treated with lamivudine (3TC), a nucleoside analog that contains a β -L-pseudo-ribose-ring. The structure of the RT–DNA–dTTP complex and the structure of the lamivudine-resistant M184I mutant RT–DNA complex have suggested that replacement of methionine at position 184 with a β -branched amino acid residue (V, I, or T) causes resistance to lamivudine and emtricitabine (FTC) by a steric hindrance mechanism, a proposal that is supported by biochemical experiments. Structures of HIV-1 RT were used to create a structural model of hepatitis B virus (HBV) polymerase. This model has been used to interpret the available biochemical and clinical data describing the resistance/susceptibility of HBV polymerase to nucleotide analogs such as 3TC, FTC, and adefovir (PMEA).

Another exclusion mutation, K65R, was originally isolated from patients treated with ddC or ddI, and subsequently emerged as a primary mutation in patients treated with tenofovir (TFV). The ternary structures of K65R RT in complexes with TFV-DP and the corresponding normal nucleotide deoxyadenosine triphosphate (dATP) revealed that the guanidinium stacking of Arg65 with Arg72 forms a molecular platform (Figure 2(b)). The platform enhances the ability of the mutant RT to discriminate NRTI triphosphates from normal triphosphates. There are other exclusion mutations, including L74V and Q151M (which is usually accompanied by secondary mutations).

AZT treatment frequently selects a set of RT mutations (M41L, D67N, K70R, T215F/Y, and K219Q) that confer

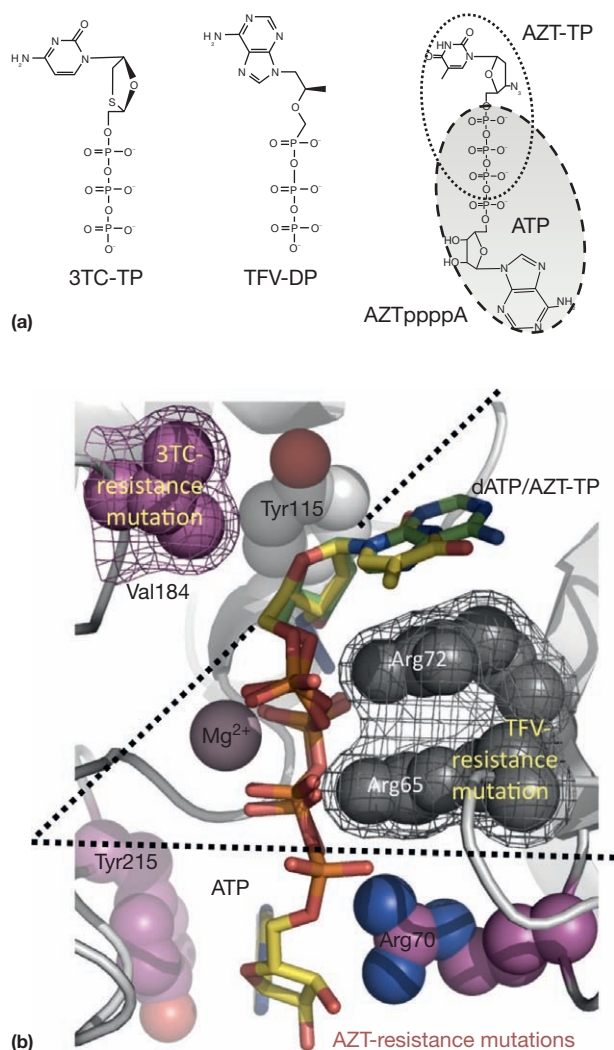


Figure 2 NRTI triphosphates and the locations of NRTI-resistance mutation sites. (a) Chemical structures of NRTIs: 3TC-TP (lamivudine triphosphate), TFVDP (tenofovir diphosphate), and AZTppppA – the product of the ATP-mediated excision reaction that removes AZTMP from the end of the primer strand. Both the AZT-TP and ATP moieties are circled. (b) Close-up of the RT active site showing the superposition of a bound AZT-TP (yellow C-atoms) and dATP (green C-atoms), one of the two active site Mg^{2+} ions that was observed in the crystal structures, and the sites of the primary resistance mutation to 3TC (M184V), TFV (K65R), and AZT (T215Y and K70R). The molecular surface of the Arg72 + Arg65 platform is shown in gray mesh and Val184 in magenta mesh. The primary AZTr mutations T215Y base-stacks and K70R has polar interactions with ATP. Reproduced from Das K, Bandwar RP, White KL, et al. (2009) Structural basis for the role of the K65R mutation in HIV-1 reverse transcriptase polymerization, excision antagonism, and tenofovir resistance. *Journal of Biological Chemistry* 284: 35092–35100.

resistance to AZT; these mutations are referred to as AZTr/TAMs (thymidine analog mutations). However, recombinant HIV-1 RT containing these mutations remained susceptible to inhibition by AZTTP in conventional polymerase assays. The molecular mechanism of AZT resistance was solved in part when Michael Parriak and colleagues proposed that RT performs a pyrophosphate-mediated excision of AZTMP after the

drug has been incorporated. Walter Scott and colleagues suggested that the pyrophosphate donor in the *in vivo* excision reaction is adenosine triphosphate (ATP). We used a combination of biochemical analysis and modeling to propose that the primary AZTr mutations enhance the ability of the mutant RT to bind ATP. Our recent crystal structures of wild-type and AZTr RT-ternary complexes with the ATP-mediated AZT-excision product (AZTppppA, Figure 2) revealed the RT-ATP interactions in atomic detail: the ATP part of AZTppppA binds differently to wild type and AZTr RTs. Wild-type RT has weak interactions with the AMP-part in the wild-type RT-dsDNA-AZTppppA structure, whereas the AZTr mutations create a new high-affinity ATP-binding site. In this site, ATP is predominantly held by stacking the adenine base with Tyr215 and hydrogen bonding of the ribose and α -phosphate with Arg70 (Figure 2). RT can use either pyrophosphate (PPi) or ATP as a pyrophosphate donor to excise the AZT-MP from the primer terminus.

If PPi is the substrate, the excision reaction is the exact inverse of polymerization. However, to generate resistance, the mutant RT must develop a mechanism that allows for efficient excision without perturbing polymerization. Using ATP as the excision substrate allows the enzyme to solve this problem. The excision product, AZTppppA, is a better substrate for AZTr RT than AZTTP, which suggests that the excision-product (AZTppppA)-binding site is a potential target for designing excision-resistant RT inhibitors.

There exist complex relationships among the NRTI-resistance mutations; some co-emerge to enhance resistance and some rarely or never co-emerge. The K65R mutation enhances the resistance to 3TC caused by the M184V mutation; however, the K65R, M184V, or K70E mutations reduce excision by AZTr RT. Some, but not all, of these effects are understood; however, knowing which NRTIs select for antagonistic mutations is an important part of planning optimal HAART strategies.

NNRTI Resistance

Structural studies have shown that NNRTI-resistance mutations are located in and around the NNIBP. Commonly observed resistance mutations in NNRTI-treated patients include K103N, Y181C, L100I, Y188L, and G190A (Figure 3(a)); these can be present singly or in combination. The structural data, supported by clinical and biochemical evidence, suggest that the mutations have different resistance mechanisms. The Y181C and Y188L mutations cause decreased NNRTI binding by the loss of aromatic ring interactions with inhibitors, this is a loss-of-contact mechanism; L100I and G190A/S can cause steric hindrance with NNRTIs; and the K103N mutation appears to affect the entry of an NNRTI into the NNIBP. However, the presence of the K103N mutation can also favor the binding of certain NNRTIs such as TMC125-R165335 (etravirine) (intence).

RT Structures in Drug Design

Variations in RT sequence and the emergence of drug-resistant strains present serious challenges for developing potent

anti-AIDS drugs (acquired immune deficiency syndrome, AIDS). New anti-AIDS drugs should be able to overcome the effects of the commonly observed drug-resistance mutations. A structural understanding of the target enzyme, mode of drug action, and effects of drug resistance can aid the design of potent anti-AIDS drugs. Together with Dr. Paul Janssen and colleagues, we used systematic structural and molecular modeling studies of RT-NNRTI complexes to design new NNRTIs that are effective against a broad spectrum of resistant strains. This multidisciplinary approach helped to develop the novel diarylpyrimidine (DAPY) NNRTIs; the DAPY derivative etravirine has been approved to treat AIDS, and another derivative, rilpivirine (TMC278), has shown positive results in phase III clinical trials. These next-generation DAPY NNRTIs have broad efficacy against commonly observed NNRTI-resistant HIV-1 RT strains. In general, more than two mutations are required to cause significant resistance against the DAPY NNRTIs. Torsional flexibility and the compactness of the DAPY backbone enable these compounds to undergo conformational changes (wiggling) and repositioning (jiggling), allowing these compounds to bind successfully to a variety of NNRTI-resistant mutant RTs (Figure 3(b)). Our high-resolution structures of wild-type and two double-mutant RT-TMC278 complexes (Figure 3(a)) showed that TMC278 can wiggle and jiggle in response to pocket mutations, thus maintaining its binding affinity. The elasticity of the binding pocket permits the protein to shrinkwrap itself around the inhibitor, which can help binding of certain inhibitor particularly to drug-resistant mutants; however, the flexibility of the protein makes it quite difficult to accurately model drug-RT interactions. For this reason, high-resolution structures of RT-NNRTI complexes are valuable guides for NNRTI drug design. Strategic flexibility is emerging as a general principle that can be used in designing drugs against other rapidly evolving targets.

Additional Drug Targets in RT

The second enzymatic activity of RT, RNase H, has not yet been targeted by any drugs. Two types of RNase H specific inhibitors (RNHIs) have been described: (1) active-site metal-ion chelators and (2) nonchelators. The development of anti-HIV-1 integrase drugs that are metal-ion chelators suggests that this is a viable approach. The recent crystal structures of metal-ion chelating RNHIs in complexes with RT and with an isolated RNase H domain have defined the inhibitor-protein interactions and the mode of metal chelation at the RNase H active site – an important step toward developing RNHIs as anti-HIV drugs.

RT is a flexible enzyme with multiple subdomains. In addition to the RNase H active site, the p66/p51-subunit interface and hinges at subdomain interfaces are potential sites for new classes of RT inhibitors. Our recently engineered HIV-1 RT crystals consistently reproduce high-resolution RT structures (better than 2-Å resolution), which opens up new opportunities to screen for RT inhibitors that bind at novel sites by using a crystallographic drug-like fragment screening approach.

Crystallographic studies have played a crucial role in characterizing the molecular interactions of drugs with wild-type

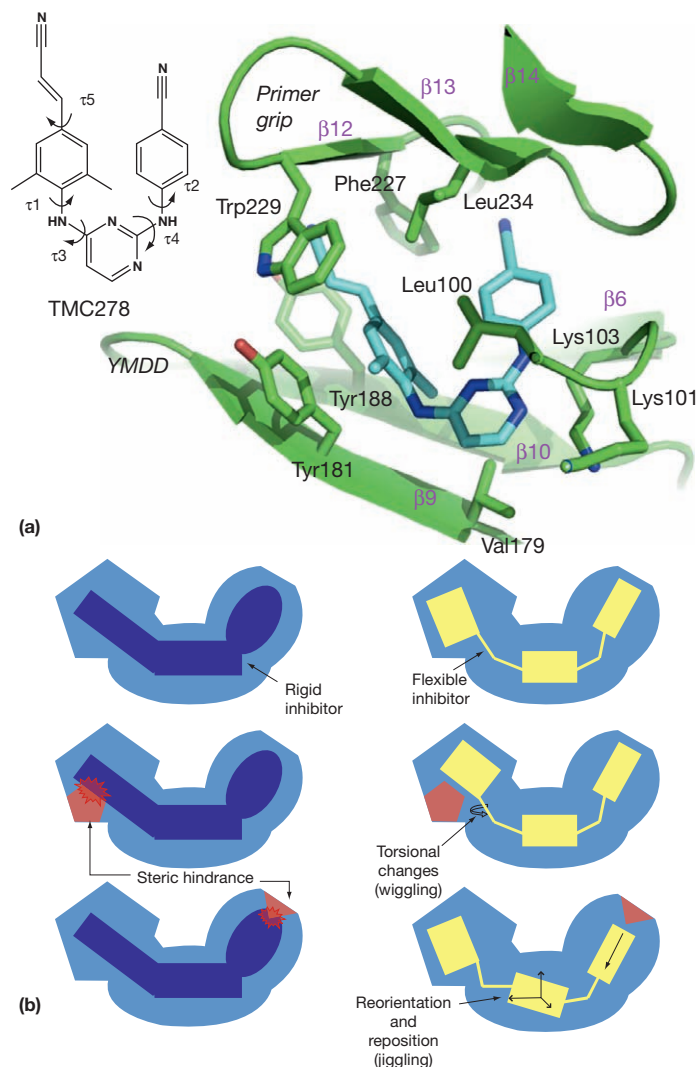


Figure 3 Designing NNRTIs to overcome known resistance mutations. (a) The chemical structure and binding of TMC278 to wild-type HIV-1 RT, as obtained by high-resolution (1.8 Å) crystal structure determination; the torsional flexibility of TMC278 is represented by τ -angles (left) and the non-nucleoside inhibitor-binding pocket (NNIBP) residues are labeled (right). (b) A schematic representation showing how a flexible and compact inhibitor can adapt to changes in the binding pocket (NNIBP) caused by mutations. A rigid inhibitor (left panel), which can bind strongly to the wild-type NNIBP, fails to retain its binding affinity to a mutated NNIBP. A flexible inhibitor (right panel) can adapt to changes in the NNIBP by its torsional rearrangements (wiggling) and repositioning (jiggling). (a) Reproduced from Das K, Clark AD, Lewi PJ, et al. (2004) Roles of conformational and positional adaptability in structure-based design of TMC125-R165335 (etravirine) and related non-nucleoside reverse transcriptase inhibitors that are highly potent and effective against wild-type and drug-resistant HIV-1 variants. *Journal of Medicinal Chemistry* 47: 2550–2560, with permission from American Chemical Society.

and drug-resistant HIV-1 RTs. We believe that structural approaches will become increasingly important in understanding existing and emerging multidrug resistance and in using that information to design improved anti-AIDS drugs and to develop new treatment strategies.

See also: **Molecular Biology:** DNA Polymerases: Reverse Transcriptase Integrase, and Retrovirus Replication; **Protein/Enzyme Structure Function and Degradation:** HIV Protease.

Further Reading

Acosta-Hoyos AJ and Scott WA (2010) The role of nucleotide excision by reverse transcriptase in HIV drug resistance. *Viruses* 2: 372–394.

- Bauman JD, Das K, Ho WC, et al. (2008) Crystal engineering of HIV-1 reverse transcriptase for structure-based drug design. *Nucleic Acid Research* 36: 5083–5092.
- Boyer PL, Sarafianos SG, Arnold E, and Hughes SH (2001) Selective excision of AZTMP by drug-resistant human immunodeficiency virus reverse transcriptase. *Journal of Virology* 75: 4832–4842.
- Coffin JM, Hughes SH, and Varmus HE (1997) *Retroviruses*. Plainview, NY: Cold Spring Harbor Laboratory Press.
- Das K, Bandwar RP, White KL, et al. (2009) Structural basis for the role of the K65R mutation in HIV-1 reverse transcriptase polymerization, excision antagonism, and tenofovir resistance. *Journal of Biological Chemistry* 284: 35092–35100.
- Das K, Bauman JD, Clark AD, Jr., et al. (2008) High-resolution structures of HIV-1 reverse transcriptase/TMC278 complexes: strategic flexibility explains potency against resistance mutations. *Proceedings of the National Academy of Sciences of the United States of America* 105: 1466–1471.
- Das K, Lewi PJ, Hughes SH, and Arnold E (2005) Crystallography and the design of anti-AIDS drugs: Conformational flexibility and positional adaptability are important

- in the design of non-nucleoside HIV-1 reverse transcriptase inhibitors. *Progress in Biophysics and Molecular Biology* 88: 209–231.
- Huang H, Chopra R, Verdine GL, and Harrison SC (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: Implications for drug resistance. *Science* 282: 1669–1675.
- Jacobo-Molina A, Ding J, Nanni RG, et al. (1993) Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proceedings of the National Academy of Sciences of the United States of America* 90: 6320–6324.
- Janssen PAJ, Lewi PJ, Arnold E, et al. (2005) In search of a novel anti-HIV drug: Multidisciplinary coordination in the discovery of 4-[[4-[[[(1E)-2-cyanoetenyl]-2,6-dimethylphenyl]amino]-2-pyrimidinyl]amino]-benzonitrile (R278474, rilpivirine). *Journal of Medicinal Chemistry* 48: 1901–1909.
- Kohlstaedt LA, Wang J, Friedman JM, Rice PA, and Steitz TA (1992) Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* 256: 1783–1790.
- Ren J, Esnouf R, Garman E, et al. (1995) High resolution structures of HIV-1 RT from four RT inhibitor complexes. *Nature Structural and Molecular Biology* 2: 293–302.
- Ren J and Stammers DK (2008) Structural basis for drug resistance mechanisms for non-nucleoside inhibitors of HIV reverse transcriptase. *Virus Research* 134: 157–170.
- Sarafianos SG, Marchand B, Das K, et al. (2009) Structure and function of HIV-1 reverse transcriptase: Molecular mechanisms of polymerization and inhibition. *Journal of Molecular Biology* 385: 693–713.
- Tu X, Das K, Han Q, et al. (2010) Structural basis of HIV-1 resistance to AZT by excision. *Nature Structural and Molecular Biology* 17: 1202–1209.

HIV Protease

B M Dunn, University of Florida College of Medicine, Gainesville, FL, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

AIDS Acquired immunodeficiency syndrome, a devastating disease caused by the HIV.

Enveloped virus A virus that has an outer covering consisting of lipid and protein derived from host cells.

Gag polyprotein A 55-kDa protein produced by translation of HIV messenger RNA.

HIV Human immunodeficiency virus.

Myristolation The addition of a fatty acid chain to the amino terminus of a protein, altering the properties of the protein.

Polyprotein A protein that contains several functional elements attached head-to-tail, which must be separated in order to express their function.

Progeny virus New viral particles produced as a result of replication within an infected cell.

Protease An enzyme that cuts other proteins into pieces.

Retrovirus A virus containing RNA as the genetic element and which must be converted into a DNA copy as part of the replicative cycle.

Virus The simplest form of life, a virus depends on host cells in order to reproduce and continue infection.

Human immunodeficiency virus (HIV) protease is a proteolytic enzyme and is part of the protein component of the HIV. HIV protease plays a critical role in the life cycle of the virus and for this reason has become an important target for drug development in the battle against HIV infection. HIV protease is a member of the aspartic protease family of proteolytic enzymes and discoveries in the general field have provided clues to effective drug design. Subsequently, HIV protease has become one of the most thoroughly studied proteins and most likely has been studied by structural analysis more than any other protein.

Human Immunodeficiency Virus

HIV is an enveloped virus with ribose nucleic acid (RNA) as the molecule used for storage of genetic information. HIV is a retrovirus, which means that the RNA is converted into a deoxyribose nucleic acid (DNA) copy that becomes integrated into the genomic DNA of a host cell. The genomic DNA is transcribed into RNA by the normal cellular machinery and this RNA is used as both messenger RNA for production of proteins by cellular ribosomes and as daughter RNA for assembly into new viral particles. HIV infects cells of the human immune system, leading to death of the cell and, thus, depletion of the immune constituents of the bloodstream with a consequent fall in immune function. Hence, the name of the disease caused by HIV is acquired immunodeficiency syndrome (AIDS). This disease is a scourge upon humankind and, as of early 2010, has killed more than 30 million people worldwide and infected as many as 40 million more.

Proteins of HIV

Gag Proteins

Translation of the gag/pol messenger RNA of HIV leads to the production of two proteins: Gag and Gag/Pol. The Gag polyprotein has a molecular weight of ~55 kDa and is called a polyprotein because it contains several proteins connected head-to-tail. The Gag–Pol polyprotein is produced about 5% of the time when the ribosome reaches a stop codon and slips

back one nucleotide to read in a different frame, thus passing through the stop codon. Three of the proteins of the Gag and Gag–Pol polyprotein (Figure 1) are important constituents of the new viral particles: (1) matrix, or MA, a 17-kDa protein that forms a spherical shell around other proteins, thus providing one level of packaging for the viral particle; (2) capsid, or CA, a 24-kDa protein that forms a conical structure within the matrix shell and contains the daughter RNA, providing a second layer of protection; and (3) nucleocapsid, or NC, a 7-kDa protein that binds to the RNA to provide a third level of packaging. When the new viral particle buds off from an infected cell, it takes some of the cell envelope along with it to complete the packaging of the new virion.

Pol Proteins

The *Pol* gene, which is the second part of the Gag/Pol fusion protein, codes for a polyprotein containing three enzymatic functions: (1) reverse transcriptase (RT), which copies the viral RNA into DNA; (2) protease (PR), which acts to cleave the Gag and Gag/Pol proteins into their separate pieces; and (3) integrase (IN), which catalyzes the insertion of the DNA copy of the viral genetic information into the host cell genome. In addition, several regulatory proteins are produced, including *vif*, *rev*, and *tat*.

Structure of HIV Protease

HIV protease begins as a component of the Gag/Pol fusion polyprotein. As a single molecule, the Gag/Pol-embedded protease is inactive, because dimer formation is required to create the active site and catalytic machinery. Understanding the structure of the Gag–Pol fusion protein is an important objective of current research.

Gag/Pol Polyprotein and Viral Assembly

Following ribosomal synthesis, Gag/Pol fusion proteins become myristolated and assemble on the inner surface of the cell membrane. This concentration of the protein leads to dimerization at

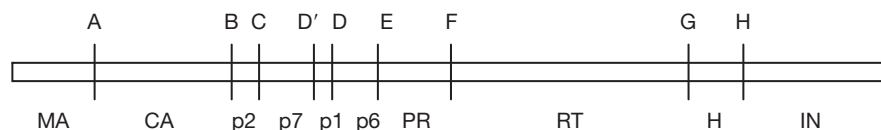


Figure 1 Diagram illustrating the arrangement of the Gag–Pol polyprotein. The letters above the boxes indicate the cleavage sites within the Gag–Pol polyprotein. The codes below the boxes indicate the designations for the protein products of the cleavages.

several points along the polyprotein, including the HIV protease. The dimerization of two protease components of the polyprotein may be assisted by prior assembly of other regions of the protein, such as the proteins in the Gag portion. The viral particle, or virion, buds off from the cell, taking some of the plasma membrane along. At this point, the virus is immature and noninfectious. Upon a change in conditions, possibly caused by an influx of protons into the virion with concomitant decrease in pH, the protease begins its function of cutting apart Gag and Gag/Pol fusion proteins. Little is known about the structure of the Gag/Pol fusion protein in the immature virion. It has been demonstrated that the cleavage of the viral polyprotein into smaller units proceeds by a series of steps that could be controlled by the organization of the polyprotein or by the selectivity of the protease for the different cleavage site junctions.

Structure of HIV Protease as an Isolated Entity

Once released from attachment to the other components of the viral Gag/Pol polyprotein, or when expressed from the 297 bp gene encoding the sequence, HIV protease can be isolated and crystallized. The three-dimensional structure reveals a molecule with structural similarities to other members of the aspartic protease family, but with some significant differences. The enzyme is a dimer of two identical 99 residue monomers and has a folding pattern that contains all the elements seen in the pepsin-like aspartic proteases, including a deep active site cleft on one side of the molecule (**Figure 2**) with both monomers contributing contact sites for bound substrates or inhibitors. Each monomer provides one aspartic acid residue and the carboxylic acid side chains of each of these come together at the bottom of the active site and bind a water molecule, which completes the catalytic machinery. This machinery catalyzes the attack of the water molecule upon the peptide bond of a bound substrate.

Inhibitor Binding

Due to the desire to create potent and selective inhibitors of the HIV protease, many different laboratories, in pharmaceutical companies, in government institutes, and at universities, have studied complexes of HIV protease with small molecule inhibitors. In fact, HIV protease was the first case where structural biology provided key insights that were used to design new, effective compounds. To date, the Food and Drug Administration (FDA) has approved 11 drugs for use in humans and these have had an enormous impact on therapy of AIDS patients, especially when used in combination with inhibitors of the viral proteins RT and IN or an inhibitor of viral entry into cells. The approved compounds provide a longer life span and an improved quality of life. However, problems with the development of resistant variants dampen the success.

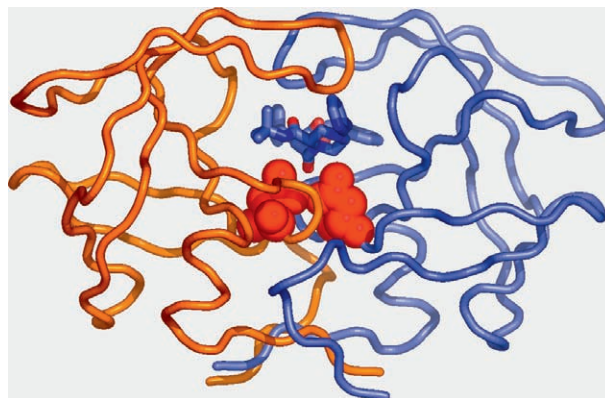


Figure 2 Representation of the three-dimensional structure of HIV protease. Two identical domains are shown in orange (left monomer) and blue (right monomer). Each domain contributes one aspartic acid residue to the catalytic machinery at the bottom of the active site cleft. These are shown in space-fill representation near the center of the figure. A bound inhibitor is shown above the two aspartic acids and directly in the middle of the active site cleft. Above the inhibitor, one can see the two prominent β -hairpin flaps, one from each monomer. The flaps cover the inhibitor in the bound state depicted here.

Mode of Inhibitor Binding

Inhibitors bind within the active site cavity of HIV protease, similar to what has been shown for binding of inhibitors against other members of the aspartic protease family. The compounds are arranged in an extended conformation (**Figure 3**) with hydrogen bonding to backbone atoms of the protease and with interactions between groups of the inhibitor and hydrophobic regions of the active site cleft. These interactions provide the specificity and potency of binding necessary to have a compound that will not cause harm to the aspartic proteases of the human host. The FDA-approved drugs all meet these criteria; however, some have side effects that are not completely understood at this time.

Development of Resistant HIV Protease Variants

As a retrovirus, replication of the virus begins with the creation of a DNA copy of the viral RNA genome. The enzyme that performs this reaction, HIV RT, is coded for by the virus and is present in the infectious virion particle. The process does not take place until the virion enters a new cell and the four different barriers protecting the viral RNA are stripped off. HIV RT makes errors while copying the viral genetic information and, unlike DNA polymerases of human cells, HIV RT does not have a proofreading function. This results in nucleotide changes during replication, so that the progeny virus will contain variations on the sequence of the infecting virus. On an

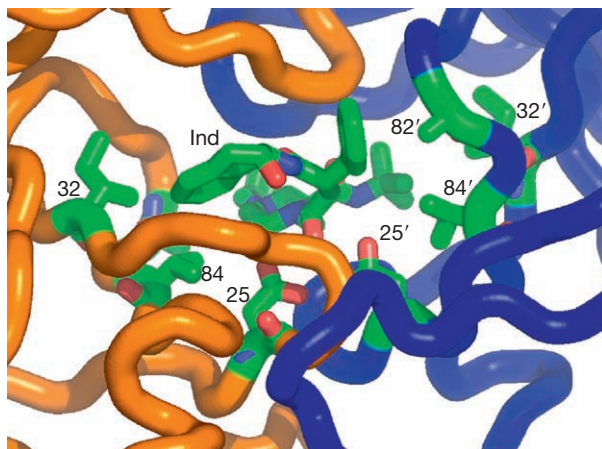


Figure 3 Close-up view of a bound inhibitor (shown in capped-sticks representation at the center of the picture) within the active site cleft of HIV protease. The two catalytic aspartic acids are seen just below the inhibitor. Several amino acids of the HIV protease are shown around the inhibitor. Each monomer contributes several amino acids that make hydrophobic interactions with the bound inhibitor. This view shows a drug-resistant form of protease, with Val32Ile, Val82Ala, and Ile84Val mutations shown in this view.

average, one nucleotide out of the ~9000 nucleotide genome of the virus is altered in each round of replication of virus. As replication takes ~2 days, the number of mutations that can arise over a 6-month period is quite large. Some of these nucleotide alterations will not change the amino acid (silent mutations), but some will. What will be the properties of the new viral protein? In some cases, the progeny virus will be defective and noninfectious, as function could be affected by the mutations. In other cases, the new virus will have favorable properties. This is how drug-resistant variants arise during therapy, especially if the patient does not follow the drug regimen prescribed. Virus that is sensitive to the drug will be killed off, but virus that has properties that include decreased sensitivity to the effects of the antiviral drugs will survive and multiply.

The Spectrum of Drug-Resistant Variants

Out of the 99 amino acid HIV PR, as many as half of the amino acids can change without losing all catalytic activity. In many cases, the efficiency of the enzyme is reduced, but as long as a threshold of activity is maintained (believed to be 10% of starting level), the virus can survive and continue to replicate. Some of the mutations that occur are within the active site cleft and this can change the binding of inhibitors to reduce the effectiveness of antiviral therapy. Discovering new drugs that do not lead to the development of resistance remains a major research objective in antiviral therapy. On the other hand, knowledge of the genetic drift of the virus population into sequences that are resistant to a certain therapeutic regimen can predict when the patient could be switched to another set of antiviral drugs.

Structure of HIV PR Variants

In addition to studies of the structure of HIV protease from laboratory strains of the virus, with the discovery of drug-resistant forms of the virus, it became important to discover how the amino acid substitutions, coded for by the altered forms of the viral genetic information, altered the activity of the protein. Thus, considerable effort has been expended to study the structure of drug-resistant forms of HIV protease. In general, it can be stated that changes in the enzyme structure are small, with obvious changes in the size of an amino acid that interacts with a bound inhibitor in the wild-type enzyme, leading to an obvious hole or protrusion, depending on the direction of the change. Even subtler are changes that occur when amino acids outside of the active site cleft occur, leading to an enhancement of resistance in some cases. A picture is emerging from studies of variant forms of HIV protease that the enzyme is very plastic or malleable, meaning that changes at any single position can be transmitted throughout the enzyme to have an impact on binding within the active site cleft and enzymatic function.

HIV PR as a Model System for Analysis of Protein Structure/Function

The classical history of protein biochemistry has involved studies of a few key proteins: the oxygen-binding proteins, hemoglobin and myoglobin, the proteolytic enzymes trypsin and chymotrypsin, and the nucleic acid degrading enzymes, ribonuclease and staphylococcal nuclease. In the 1990s, HIV protease became a new classical protein that proved to be amenable to detailed analysis. The protein is relatively easy to produce by recombinant means and purify, it has a robust activity and can be analyzed by several different methods, and variants can be created by current methods of molecular biology. Consequently, the HIV protease has been studied by nearly every method of analysis available to the protein chemist. Due to its small size, computational approaches have been applied and HIV protease has become an excellent system for theoretical studies of protein structure/function.

See also: [Molecular Biology: Catalysis by RNA](#), [Structural Themes: Group I Introns](#); [DNA Polymerases: Reverse Transcriptase Integrase, and Retrovirus Replication](#).

Further Reading

- Dunn BM (ed.) (1999) *Proteases of Infectious Agents*. San Diego, CA: Academic Press.
 Kuo LC and Shafer JA (eds.) (1994) *Retroviral Proteases*, vol. 241. San Diego, CA: Academic Press.
 Levy JA (ed.) (1994) *The Retroviridae*, 3 vols. New York, NY: Plenum.
 Wlodawer A and Vondrasek J (1998) Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annual Reviews in Biomolecular Structures* 27: 249–284.

Homologous Recombination in Meiosis

N M Hollingsworth, Stony Brook University, Stony Brook, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Crossover homeostasis The maintenance of a certain number of crossovers (COs) at the expense of non-crossovers (NCOs).

Holliday junction A mobile junction between four strands of DNA.

Homologs Chromosomes with highly similar, but not necessarily identical, DNA sequences.

Hot spot Region on a chromosome that is preferentially cleaved by Spo11.

Interference A CO in one region decreases the probability of a second CO in an adjacent region.

Recombination Includes both crossing over, which is the reciprocal exchange of segments of DNA between homologous chromosomes, and gene conversion, the process by which the sequence of one allele is changed to the sequence of the other allele in the cell.

Sister chromatids Two identical copies of a chromosome created by DNA replication.

Synapsis The physical association of homologous chromosomes during meiosis by the formation of the synaptonemal complex.

Meiosis

Sexually reproducing organisms face a daunting challenge: for the chromosome number of the species to remain constant, the gametes (germ cells) used for reproduction must contain only half the number of chromosomes as the diploid parent. This reduction in chromosome number is achieved by a specialized cell division, meiosis, in which two rounds of chromosome segregation follow a single round of chromosome duplication. DNA replication produces pairs of sister chromatids held together by multi-subunit protein complexes called cohesins (Figure 1, 1). Reciprocal exchange of DNA specifically between nonsister chromatids, in combination with sister chromatid cohesion, physically connects homologous chromosomes, enabling them to orient properly at metaphase of the first meiotic division (meiosis I) (Figure 1, 2). Anaphase of meiosis I is triggered by the removal of cohesins specifically from chromosome arms, allowing homologous pairs of sister chromatids to segregate to opposite poles (Figure 1, 3). At anaphase of the second meiotic division (meiosis II), centromere cohesion is eliminated and sister chromatids segregate to opposite poles, similar to mitosis (Figure 1, 4). The fate of the four haploid products varies in different organisms. In mammals, for example, only one haploid genome from a single meiosis is packaged into an egg, while all four are matured into sperm. In yeast, the four haploid gametes are called spores and they are packaged together into a sac called an ascus (Figure 1, 5).

The Role of Recombination in Meiosis

In addition to independent assortment, crossing over between homologous chromosomes produces new combinations of alleles that provide the grist for evolution. Recombination also plays a critical mechanical role in meiosis by creating

physical connections between homologs that are crucial for proper meiosis I segregation (Figure 1). In the absence of recombination, chromosomes segregate randomly, thereby creating gametes with too few or too many chromosomes. Meiotic recombination is therefore critical to fertility and the production of healthy offspring.

A Molecular Model for Meiotic Recombination

Initiation

The double-strand break repair (DSBR) model for recombination was first proposed in 1983. Since then, many of the molecular intermediates predicted by the DSBR model have been observed experimentally in yeast and the model has gained wide acceptance. The DSBR model proposes that meiotic recombination is initiated by the introduction of a double-strand break (DSB) on one of the four chromatids of a pair of homologs (Figure 2, 1). The formation of DSBs is catalyzed by a meiosis-specific topoisomerase-like protein called Spo11. Spo11 introduces DSBs nonrandomly throughout the genome at places called 'hot spots'. Spo11 cleavage is not sequence dependent, but instead appears to occur primarily in nucleosome-free areas. Yeast cells lacking the Spo11 protein fail to make breaks, undergo no recombination, and exhibit high levels of chromosome missegregation. Meiotic defects resulting from mutation of *SPO11* have also been observed in fruit flies, plants, nematodes and mice, indicating that the mechanism for initiating meiotic recombination is evolutionarily conserved.

Resection

Once a DSB is made, the 5' ends on each side of the break are resected to produce 3' single-stranded (ss) tails (Figure 2, 2). Resection requires a highly conserved protein complex containing Rad50, Xrs2 (Nbs1 in mammals), and Mre11. In yeast,

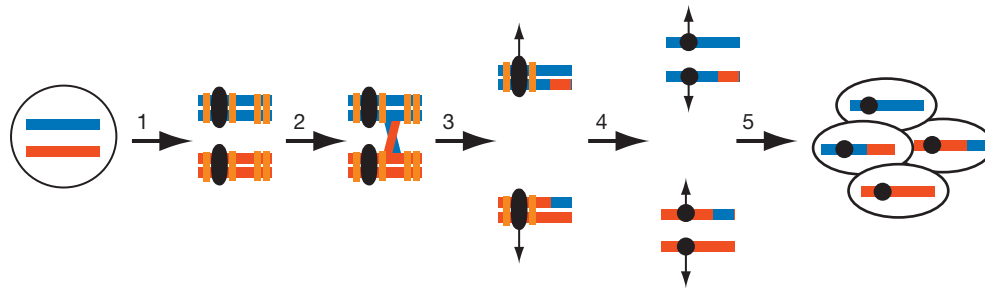


Figure 1 Chromosome behavior during meiosis. A diploid cell containing a pair of homologous chromosomes (one blue and one red) is shown. Black-filled ovals and circles represent meiosis I and meiosis II centromeres, respectively, to which spindle fibers attach. Arrows connected to centromeres indicate the direction of segregation. (1) Chromosomes replicate to form sister chromatids that are held together by cohesin complexes (indicated by orange bars). (2) Reciprocal recombination, or crossing over, connects homologous chromosomes, enabling their proper alignment at metaphase of meiosis I. Note that sister centromeres function as a single unit. (3) At anaphase I, arm cohesion is removed, and homologous pairs of sister chromatids segregate to opposite poles. (4) At anaphase II, loss of centromere cohesion allows sister chromatids to segregate to opposite poles, similar to mitosis. (5) Four haploid products are packaged into gametes. In yeast, these gametes are called spores.

the absence of any one of these genes prevents DSB formation. However, specific mutant versions of the genes encode proteins that make DSBs, but the breaks remain unprocessed because Spo11 remains covalently bound to the 5' ends of the break. Therefore, DSB formation and 5' end processing are coupled together by these proteins. Resection is initiated when Spo11 is removed by cleavage of the ssDNA on either side of the break, releasing Spo11 covalently attached to a short oligonucleotide.

Invasion

There are two potential sources of homology that can be used as templates for repair of DSBs, sister chromatids and the nonsister chromatids of the homologous chromosomes. Strand invasion is the step at which this decision is made. Although repair occurs preferentially between sister chromatids in mitotically dividing cells, in meiosis this bias is shifted so that interhomolog recombination is favored (Figure 2, 3 and 2–4). This bias is important, as intersister recombination events do not facilitate meiosis I chromosome segregation.

In yeast and mammals, strand invasion is mediated by two recombinases, Rad51 and Dmc1, both highly conserved proteins that are members of the RecA family of strand transferases. Rad51 is the primary recombinase in mitotically dividing cells, whereas Dmc1 is meiosis specific. The proteins form filaments on the ss ends of DSBs and, in conjunction with other proteins, denature homologous duplexes, displacing strands of like polarity to form displacement or D-loops (Figure 2, 3 and 2–4).

Double Holliday Junction Formation and Resolution

After strand invasion, DNA synthesis extends the invading strand, thereby increasing the size of the D-loop until the displaced single strand can complementary base pair with the 3' tail on the other side of the break (Figure 2, 5). A recombination intermediate containing two cross-shaped structures called 'Holliday junctions' (HJs) is then created. To separate the two recombining chromatids from each other, each HJ must be resolved by nicking the two strands of like polarity, followed by religation. To generate a crossover (CO), the two

HJs must be resolved in opposite directions by nicking one pair of strands in one junction and the complementary pair of strands in the second junction (Figure 2, 6).

Crossovers versus Non-Crossovers

Originally, the DSBR model proposed that double HJs could be resolved to yield either CO chromosomes (different strands are cut and ligated in the two different junctions so that the parts of each chromatid distal to the crossover are exchanged) or non-crossover (NCO) chromosomes (the same strands are cut and ligated in the two different HJs so that the arms of the two chromatids are not exchanged) and that this could occur with equal frequency. Subsequent studies have shown, however, that DSBs processed to form double HJs are resolved solely as COs. Instead, NCO chromosomes are proposed to occur by a mechanism called synthesis-dependent strand annealing or SDSA. SDSA results when the extended invading strand is displaced from the D-loop and anneals to the single-stranded end from the other side of the break (Figure 2, 7 and 8).

Studies in yeast have shown that there are approximately 160 DSBs in each meiosis, ~90 of which are processed into COs, ~20 are processed into NCOs, with the remaining presumably repaired using sister chromatids. The relationship between COs and NCOs exhibits 'crossover homeostasis' in that reducing the number of breaks reduces NCOs to a greater extent than COs. This suggests that the cell can preferentially channel breaks into the CO pathway to ensure that homologous chromosomes are physically connected. How the CO/NCO decision is made is not yet understood.

Gene Conversion

Mendel's first law states "alleles segregate equally into gametes." Therefore, when a diploid that contains two different alleles of a gene (+/-) is put through meiosis, half of the gametes should contain the '+' allele and half should contain the '-' allele (indicated by 2⁺:2⁻). The ability in fungi to analyze all four products from a single meiosis (called spores)

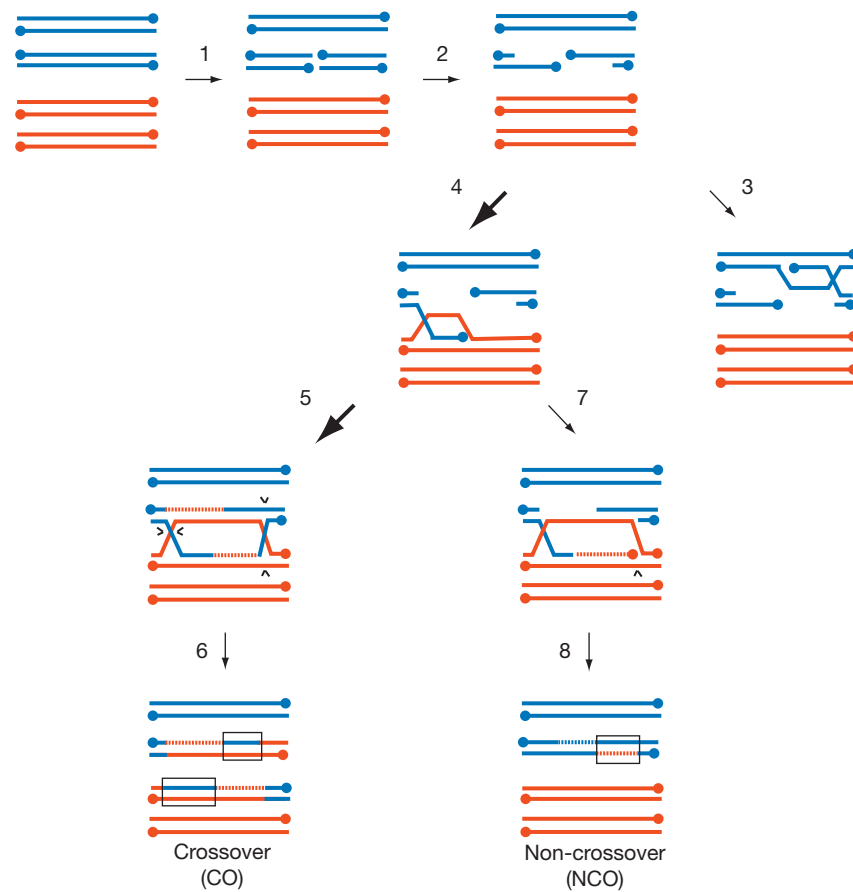


Figure 2 Double-strand break repair by meiotic recombination. The DNA duplexes of two homologous chromosomes (one blue and one red) are drawn after DNA replication has produced sister chromatids. The 3' ends of the DNA strands are indicated by circles. (1) Recombination is initiated after DNA replication by the introduction of a DSB on one of the four chromatids by Spo11. (2) The 5' ends of the breaks are resected by Mre11/Xrs2 (Nbs1)/Rad50 to generate 3' single-stranded tails. These tails are coated with the meiosis-specific recombinases Dmc1 and/or Rad51 and the resulting protein/DNA filaments mediate strand invasion of either sister chromatids (3) or the nonsister chromatids of homologous chromosomes (4) to produce D-loops. The size of the arrows indicates a bias for interhomolog recombination at the strand invasion step. Strand invasion of the homolog can result in either CO or NCO chromosomes. For COs, the invading 3' end is extended by DNA replication (indicated by dashed lines), thereby increasing the size of the D-loop (5). Eventually, the D-loop is large enough that the second end of the DSB can be captured by annealing to the displaced single strand of the D-loop, forming two Holliday junctions. Resolution of the two Holliday junctions in opposite directions (indicated by carats) results in two CO chromatids (6). NCOs are created by synthesis-dependent strand annealing (SDSA). The invading DNA is extended (7) but is then removed from the D-loop and the newly synthesized region of DNA anneals with the 3' tail on the other side of the break (8). This end then acts as a primer for DNA synthesis to fill in the gap. Both CO and NCO pathways produce heteroduplex DNA (indicated by black boxes). Correction of mismatches contained in the heteroduplex DNA may result in gene conversion. Although not shown, Holliday junction formation and resolution as well as SDSA can occur between sister chromatids also.

led to the discovery that a small percentage of the time Mendel's first law was broken. In these cases, meiosis produces spores in a ratio of $3^+:1^-$ or $1^+:3^-$ by a process called 'gene conversion'.

Gene conversion changes the sequence of one allele precisely to the sequence of the other allele in the cell. This transfer of sequence information can occur due to the formation of heteroduplex DNA during recombination. Heteroduplex DNA results when a strand of DNA from one homolog complementary base pairs with a strand of DNA from the other homolog. According to the DSBR model, there are two times when heteroduplex is initiated – after strand invasion and after the displaced single strand of DNA in the D-loop anneals to the 3' end on the other side of the break (Figure 2, 6). Because the HJs can branch-migrate along the chromosome, heteroduplex

DNA can be formed some distance from the DSBs. Heteroduplex may also be generated by SDSA (Figure 2, 8). Because the DNA sequences of homologous chromosomes, while highly similar, are not necessarily identical, base-pair mismatches can occur in heteroduplex DNA. These mismatches are then corrected by a highly conserved mismatch repair machinery. A mismatch can be corrected so that the original base pair is restored, thereby producing $2^+:2^-$ spores or it can be corrected so that the sequence of one allele is changed to the sequence of the other allele ($3^+:1^-$ or $1^+:3^-$). Gene conversion has been observed on both CO and NCO chromosomes, consistent with the formation of heteroduplex prior to double HJ resolution or SDSA. Gene conversion has been proposed to play an important role in maintaining the sequence homogeneity of multi-gene families.

The Meiotic Recombination Checkpoint

The introduction of DSBs into chromosomes by Spo11 puts the cell in grave peril. Proceeding ahead through meiosis I with unrepaired DSBs would be fatal, as pieces of broken chromosomes would become irrevocably separated from each other. Once recombination is initiated, therefore, the process is carefully monitored such that meiosis I occurs only after recombination is complete. This surveillance process has been called the 'meiotic recombination checkpoint' or the 'pachytene checkpoint'. For example, budding yeast lacking *DMC1* arrest prior to meiosis I. Examination of the DNA indicates that DSBs are made and resected, but because there is no Dmc1 to promote strand invasion into the homolog, recombination cannot proceed and the meiotic recombination checkpoint is triggered. The cells remain arrested, waiting for the break to be repaired. The meiotic recombination checkpoint also functions in mammals with mutants such as *dmc1Δ* triggering apoptosis in the germ cells of mice. Interestingly, this checkpoint appears to be more robust in males than in females.

Genetic Interference

It is crucial that each pair of homologs receives at least one CO to prevent missegregation at meiosis I. Because in many organisms there is a finite number of COs in each meiosis, COs are distributed so that large chromosomes do not sustain a disproportionate number of COs at the expense of small chromosomes. This distribution is accomplished by a phenomenon known as 'genetic interference'. Interference was first observed at the beginning of the twentieth century when fruit fly geneticists were mapping genes on the same chromosomes. These geneticists found that the presence of a CO in one interval decreases the chance that a second CO occurs nearby. Mutation of genes required for interference in budding yeast increases meiosis I missegregation, especially of small chromosomes. Interference is calculated as $1 - cc$, where cc stands for the 'coefficient of coincidence'. The coefficient of coincidence is the ratio of observed double COs to expected double COs based on the assumption that the two COs occur independently. The strength of interference varies in different parts of the genome, as well as between organisms. COs in nematodes, for example, exhibit interference values close to 1, while fission yeast cells exhibit no interference. Most organisms, such as yeast, fruit flies, and mammals, fall between these two extremes. Despite intense research, the mechanism by which the cell distributes COs through interference remains obscure.

Two Pathways of Crossing Over

Two different classes of interhomolog COs have been observed in budding yeast, plants, and mammals. The majority pathway is mediated by the meiosis-specific MutS homolog pair, Msh4 and Msh5, and these COs exhibit interference. A second, interference-independent set of COs is mediated by a structure-specific endonuclease, Mus81/Mms4 (also known as Eme1).

The Mus81/Mms4 complex has HJ resolution activity *in vitro*, but is not used for resolving HJs in the Msh4/Msh5-dependent COs. Examination of these CO pathways in different organisms has revealed interesting evolutionary divergence. The genome of the nematode, *Caenorhabditis elegans*, is missing *MUS81/MMS4*, and disruption of the *MSH4/MSH5* pathway eliminates all COs. As expected, these COs exhibit very strong interference. By contrast, the genome of the fission yeast, *Schizosaccharomyces pombe*, lacks *MSH4/MSH5*. COs exhibit no interference and are dependent on *MUS81/EME1* (the fission yeast name for *MMS4*). Although it appears that Mus81/Eme1 acts as a Holliday resolvase in fission yeast, the resolvase used for the Msh4/Msh5 pathway remains to be found.

The Synaptonemal Complex

A unique feature of meiosis is the physical association of homologous chromosomes by a meiosis-specific structure called the synaptonemal complex (SC) that is formed by a process called 'synapsis' (Figure 3). After DNA replication, sister chromatids condense along protein cores called 'axial elements' (AEs) that are comprised, at least in part, of meiosis-specific proteins. AEs from homologous chromosomes then become glued together by insertion of the central region to form SCs. In the context of the SC, AEs are referred to as lateral elements. Associated with SCs are densely staining bodies called 'recombination nodules' (RNs). RNs are thought to contain the proteins that mediate recombination. The relationship between the SC and meiotic recombination varies depending upon the organism. In nematodes and fruit flies, SC formation is independent of recombination, whereas in yeast and mammals, recombination is a prerequisite for SC formation. The absence of AEs greatly decreases the frequency of both gene conversion and crossing over between homologous chromosomes, demonstrating that there is an important connection between meiotic chromosome structure and recombination. Furthermore, studies in yeast and nematodes indicate that meiosis-specific components of axial elements are important for promoting interhomolog recombination by suppressing strand invasion of sister chromatids.

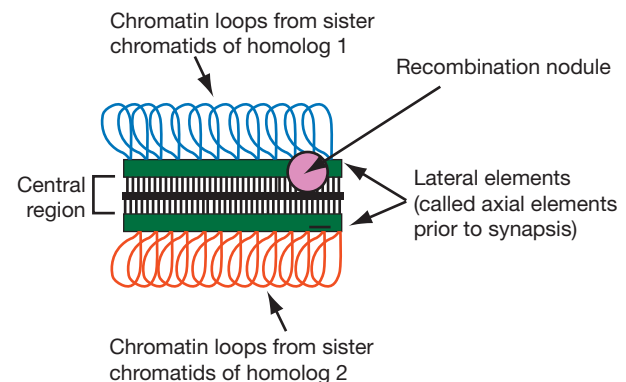


Figure 3 The synaptonemal complex. Synapsis of two homologous chromosomes (one blue and one red) is shown.

Mitotic versus Meiotic Recombination

Meiotic recombination exhibits a number of important differences from mitotic recombination. First, meiotic recombination is several orders of magnitude more frequent. This is presumably due to the programmed induction of meiosis-specific recombination proteins such as Spo11. In mitotically dividing cells, recombination is used to repair lesions in DNA and therefore recombination events occur rarely and are located randomly along the chromosomes. In meiotic cells there are 'hot spots' of recombination – these are places on chromosomes where Spo11 prefers to act. In mitotic cells, recombination occurs most frequently between sister chromatids, which act as a template for DNA repair. By contrast, crossing over in meiotic cells is targeted between homologs. This bias is key because only interhomolog crossovers promote proper segregation at meiosis I. Finally, meiotic recombination occurs between homologous chromosomes that are physically associated by a meiosis-specific structure, the SC.

See also: [Cell Architecture and Function: Meiosis](#); [Molecular Biology: DNA Mismatch Repair and Homologous Recombination](#).

Further Reading

- Hassold TJ and Hunt P (2001) To err (meiotically) is human: The genesis of human aneuploidy. *Nature Reviews* 2: 280–291.
- Hollingsworth NM and Brill S (2004) The Mus81 solution to resolution: Generating meiotic crossovers without Holliday junctions. *Genes and Development* 18: 117–125.
- Hunt PA and Hassold TJ (2002) Sex matters in meiosis. *Science* 296: 2181–2183.
- Hunter N (2008) Meiotic recombination. In: Aguilera A and Rothstein R (eds.) *Molecular Genetics of Recombination, Topics in Current Genetics*, pp. 1–62. Heidelberg: Springer.
- Keeney S (2001) Mechanism and control of meiotic recombination initiation. *Current Topics in Developmental Biology* 52: 1–53.
- Pan J and Keeney S (2007) Molecular cartography: Mapping the landscape of meiotic recombination. *PLoS Biology* 5: 2774–2777.
- Roeder GS and Bailis JM (2000) The pachytene checkpoint. *Trends in Genetics* 16: 395–403.
- Villeneuve AM and Hillers KJ (2001) Whence meiosis? *Cell* 106: 647–650.

Hydrogenases, Structure and Function

W Lubitz and H Ogata, Max-Planck-Institut für Bioanorganische Chemie, Mülheim an der Ruhr, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

Electron paramagnetic resonance (EPR) A spectroscopic method used to detect and characterize paramagnetic species like transition metal complexes, radicals, or triplet states.

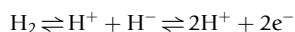
FeGP cofactor An iron guanylylpyridone cofactor which is light- and temperature-sensitive.

Fourier transform infrared spectroscopy (FTIR) A spectroscopic method to detect vibrational frequencies of molecules or complexes from different origins.

H-cluster The active site of [FeFe] hydrogenase containing a binuclear iron site ([FeFe] subcluster) and a [Fe₄S₄] subcluster bridged by the thiolate of cysteine residue.

Introduction

Hydrogenases catalyze the reversible oxidation of dihydrogen in nature according to the following simple reaction:



The cleavage of dihydrogen is strongly endergonic – homolytic in the gas phase by +436 kJ mol⁻¹ and heterolytic in water at 20 °C by about +200 kJ mol⁻¹. Heterolytic cleavage is thus energetically favored. The low acidity of H₂ (pK_a ≈ 35) is strongly increased by binding to a metal center. Abundant transition metals, such as Fe and Ni, are used for this purpose in nature by the hydrogenases, whereas precious metals such as Pt are employed in chemical catalysis. The sustainable production and conversion of hydrogen are highly interesting for future technologies.

Hydrogenases are mainly found in microorganisms such as archaea, bacteria, and also in some eukaryotes. They usually live in anoxic environments such as freshwater ponds, marine sediments, swamps, wetlands, and also in hot springs, hydrothermal vents, and even in the intestinal tracts of ruminants and termites. Most hydrogenases are oxygen-sensitive – but in organisms living in aerobic environments some oxygen-tolerant enzymes have been described; examples are found in the Knallgas bacteria (e.g., of the genus *Ralstonia*) or certain thermophiles (e.g., *Aquifex*). Hydrogenases can be classified according to their redox partner, their location in the cell (soluble or membrane-bound), or their function (catalytic or regulatory/sensory hydrogenases). Structurally hydrogenases are divided into three major groups depending on the metal components of their active sites. These are phylogenetically distinct classes and are known as the [NiFe] (subclass [NiFeSe]), [FeFe], and [Fe] hydrogenases:

1. [NiFe] hydrogenases are composed of two subunits and contain the [NiFe] active center, three iron–sulfur clusters, and a magnesium ion (an iron ion in case of [NiFeSe] hydrogenase).
2. [FeFe] hydrogenases are composed of one or two subunits and they consist of at least one iron–sulfur cluster and a di-iron metal center connected via the thiolate of a cysteine residue (so-called H-cluster).
3. [Fe] hydrogenases harbor a unique iron–guanylylpyridinol cofactor (FeGP).

All hydrogenases possess at least one iron in the active site that carries unusual nonprotein ligands such as CO or CN⁻. Their metal active sites are shown in [Figures 1–3](#). Most of the hydrogenases function bi-directionally in H₂-uptake and H₂-production. [NiFe] hydrogenases are usually more active in H₂ oxidation and [FeFe] hydrogenases in the production of H₂. The catalytic activity is often very high. Some [FeFe] hydrogenases are, for example, extremely active in H₂-generation, each molecule of enzyme can produce up to 10⁴ molecules of H₂ per second at ambient temperature. In view of the prospective applications of hydrogenases in biotechnology, for example, for producing dihydrogen or for employing it as an energy source, the interest in these enzymes and their properties has been constantly growing over the past few years. Since the crystal structures of the hydrogenases and their unusual catalytic active centers have been revealed, many bio-inspired synthetic models have been developed to mimic the function of the hydrogenases.

[NiFe] Hydrogenase

Structure

Single crystals and structural information have so far only been obtained from the catalytic standard [NiFe] and [NiFeSe] hydrogenases from sulfate-reducing bacteria, *Desulfovibrio* species, that are discussed here. These are comprised of two subunits ([Figure 1\(a\)](#)). The small subunit harbors three iron–sulfur clusters (2 [Fe₄S₄] and 1 [Fe₃S₄]) that are active in the electron transfer pathway. An exception is the [NiFeSe] hydrogenase that has three [Fe₄S₄] clusters. These metal centers are located within distances of ~13 Å, optimized for efficient electron transfer. The [Fe₄S₄] cluster most distal from the [NiFe] active site is coordinated by three cysteines and one histidine; the others have only cysteine residues. The heterolytic cleavage of dihydrogen takes place at the [NiFe] active site in the large subunit ([Figure 1\(b\)](#)). The iron is six-coordinated, whereas the nickel is only carrying five ligands. The open coordination site at the Ni is believed to be the active site, where the substrate dihydrogen is attached and inhibitors such as carbon monoxide bind ([Figure 1\(b\)](#)). The iron atom is coordinated to three diatomic nonprotein ligands (1 CO and 2 CN⁻). Furthermore, two thiolates of cysteine residues are bridging the iron and the nickel. In addition, two terminal cysteine residues are

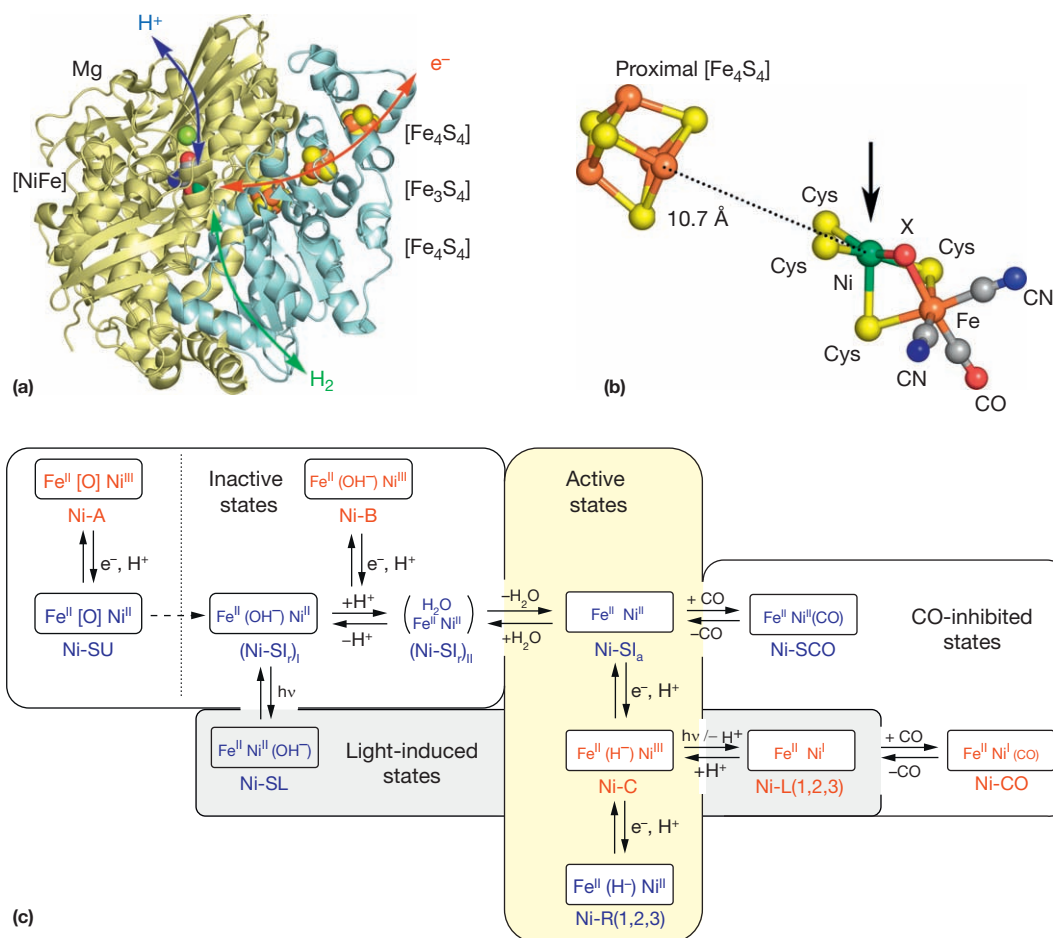


Figure 1 Crystal structure of [NiFe] hydrogenase from the sulfate-reducing bacterium *Desulfovibrio vulgaris* Miyazaki F (PDB ID 1WUI). (a) Overall structure: the large and the small subunits are shown (yellow and light blue, respectively) and the [NiFe] active site (green/orange), the Mg ion (light green) active in proton transfer, and the three iron–sulfur centers of the electron transfer chain; the hydrophobic gas-access channel is also indicated. (b) [NiFe] active site with its ligands (see text) and the location of the proximal [Fe₄S₄] cluster. The open coordination site is indicated with an arrow. (c) Scheme detailing the various states involved in the enzyme activation (inactive states), in the catalytic cycle (active states), in the CO-inhibition and in the light sensitivity of the [NiFe] hydrogenase.

bound to the Ni atom. In the case of [NiFeSe] hydrogenase, one of the terminal cysteine residues is replaced by a selenocysteine. The third bridging ligand X between Ni and Fe changes identity during the catalytic cycle (Figure 1(c)). It is an oxygenic species in the oxidized state of the enzyme and a hydride in the reduced state. Spectroscopic techniques such as electron paramagnetic resonance (EPR) have shown that the Ni ion is redox active and varies the valence between Ni³⁺ and Ni²⁺ in the catalytic cycle, whereas the Fe atom remains in the low-spin Fe²⁺ state, which is caused by the strong ligand field.

In the C-terminus of the large subunit, a Mg²⁺ ion is found coordinated to the terminal histidine residue, a leucine residue, an oxygen atom from the peptide-backbone, and three water molecules. In the case of [NiFeSe] hydrogenase, an Fe ion is present at the corresponding position. These metals are proposed to play a role in the proton transfer pathway to and from the active site. Since the [NiFe] active site is buried deeply in the molecule, the substrate dihydrogen has to transfer via

gas-access channels. There are several entrances of the gas-access channels but it is a narrow channel in the vicinity of the active site that selects the small molecules, such as H₂, CO, and O₂.

Catalytic Cycle of [NiFe] Hydrogenase

Based on spectroscopic and electrochemical data a number of intermediates in processes related to hydrogenase function could be identified and characterized. This knowledge is summarized in Figure 1(c), which is divided into the inactive, catalytically active, CO-inhibited, and the light-induced states. Nickel is the redox active metal ion and thus the functional scheme consists of a number of EPR-active and EPR-silent states. The enzyme cycles between formal Ni³⁺ and Ni²⁺ states as indicated (a formal Ni¹⁺ state is only present at cryogenic temperatures under illumination). The iron is not redox active, it is always diamagnetic (low-spin Fe²⁺) coordinated by strong π- and σ-ligands, that is, CO and CN⁻.

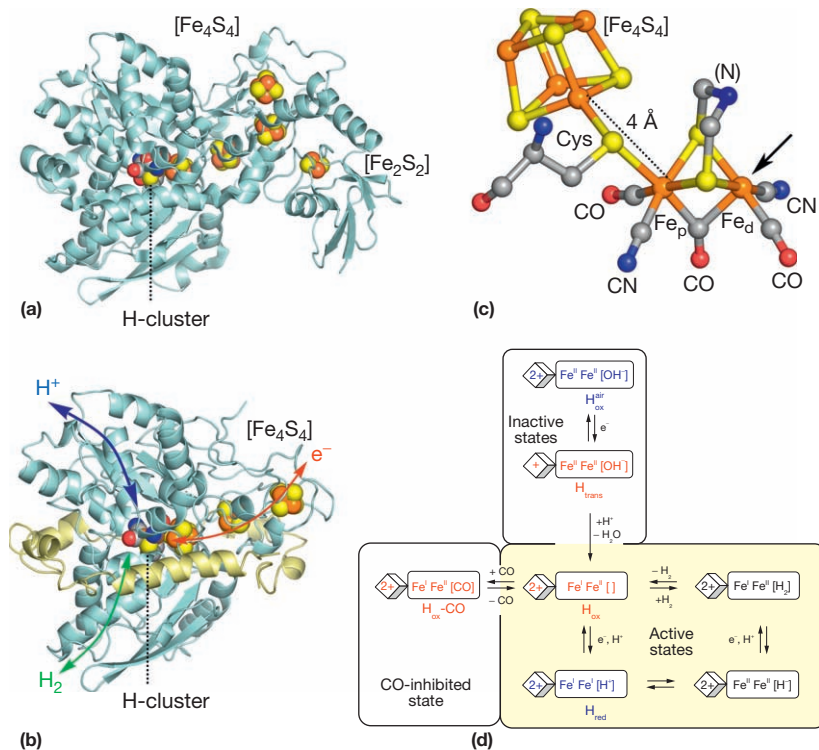


Figure 2 (a) Three-dimensional representation of [FeFe] hydrogenase from Cpl (PDB ID 3C8Y) and (b) from *D. desulfuricans* (PDB ID 1HFE) showing the active site (H-cluster) and the [FeS] clusters. In (b) the transfer pathways for H⁺, e⁻, and H₂ are indicated. (c) Structure of the H-cluster (see text). The open coordination site is indicated with an arrow. (d) Scheme detailing a model for the various states involved in the reactivation (inactive states), in the catalytic cycle (active states), and in the CO-inhibition of the [FeFe] hydrogenases.

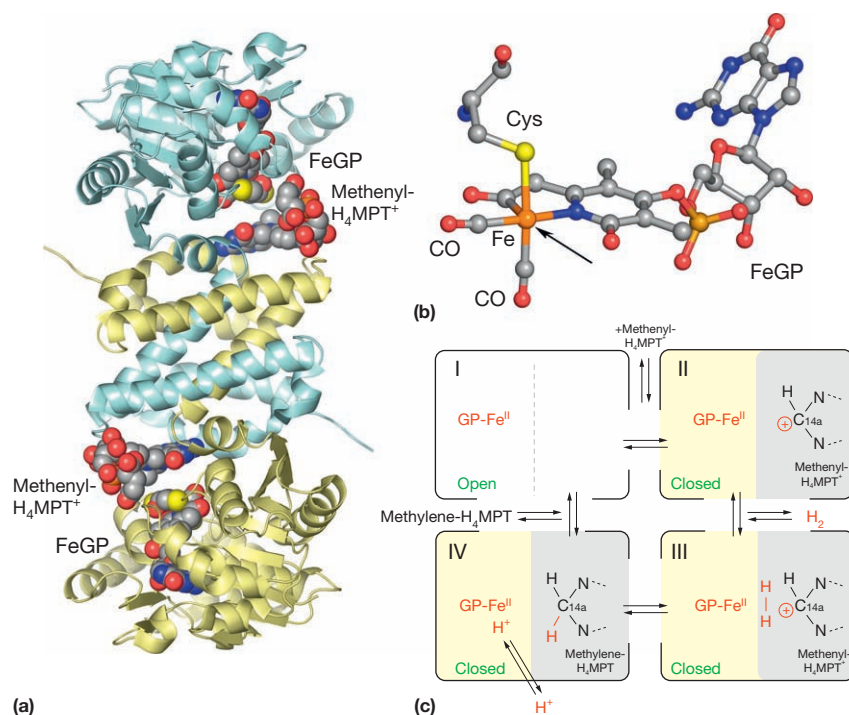


Figure 3 (a) Crystal structure of [Fe] hydrogenase from *Methanocaldococcus jannaschii* (PDB ID 3H65). (b) Structure of the FeGP cofactor. The open coordination site is indicated with an arrow. (c) Schematic representation of the catalytic reaction mechanism of [Fe] hydrogenase (see text).

In the presence of O_2 , catalytic ability is lost. Two paramagnetic inactive states (Ni^{3+} , $S = 1/2$) of the enzyme (Ni-A and Ni-B) have oxo-based ligands bound to the active site bridging Ni and Fe. The Ni-B (ready) state has a μ -hydroxo bridge; this state is quickly activated, a process that finally leads to a loss of the bridging ligand in the active Ni-SI_a state of the catalytic cycle. The Ni-A state (unready) is slowly activated. It is absent in all oxygen-tolerant [NiFe] hydrogenases. A final conclusion on the bridging ligand has not been made.

One-electron reduction of Ni-B and Ni-A leads to the formation of the EPR-silent (Ni^{2+}) catalytically inactive states Ni-SI_r and Ni-SU, respectively. Ni-SI_r exists in two forms that are in an acid-base equilibrium: the (Ni-SI_r)_I and the more protonated (Ni-SI_r)_{II}. In this process, the μ -hydroxo bridge is protonated (to form water) prior to converting to the active Ni-SI_a state, which lacks the bridge X.

The physiological states related to dihydrogen catalysis are Ni-SI_a, Ni-C, and Ni-R (Figure 1(c)). Ni-SI_a is the most oxidized active intermediate and is EPR-silent (Ni^{2+}). It has an open bridging position and the nickel is thus only four-coordinated. It is suggested that this state attaches the dihydrogen, which is polarized and heterolytically split. The Ni is probably the site of H_2 binding since CO is also attached to this metal and the gas-access channel ends close to the nickel ion (Figure 1(a)). The open bridging position between Ni and Fe has also been discussed as the primary site of dihydrogen attack.

Proton-coupled one-electron reduction leads to the catalytic Ni-C state. In this step, base-assisted heterolytic cleavage of the H_2 molecule takes place, leaving a hydride ligand (H^-) in the bridging position between Ni and Fe, and an electron is released to the proximal $[Fe_4S_4]$ cluster. One of the candidates for the respective base is a terminal cysteine that can then transfer the proton to a nearby glutamic acid. Alternatively, the proton could be taken up by another cysteine residue or even by a water molecule loosely bound to the Fe, as has been proposed in another mechanism.

Proton-coupled one-electron further reduction of Ni-C leads to the EPR-silent state Ni-R. However, Ni-R can also be formed directly from the Ni-SI_a state after its reaction with H_2 if all three clusters are already reduced, without the Ni-C being observed as a detectable intermediate. Ni-R is the most reduced state of the enzyme and can exist in three isoelectronic forms that differ in the protonation degree either of the coordinating cysteine residues or of the substrate ligands (e.g., H_2 and H^-) bound to the active site. All details of the mechanism are not available experimentally, a fact that has led to several mechanistic proposals. An interesting alternative to the above-discussed scheme has been postulated, in which only the Ni-R and Ni-C states are involved in catalysis.

Whereas the fully oxidized states Ni-B and Ni-A are not light sensitive, the Ni-SI_r state upon illumination forms an EPR-silent species denoted as Ni-SL, for which a displacement of the μ -hydroxo ligand has been postulated. Illumination at low temperatures also affects the Ni-C state, which converts to a paramagnetic state termed Ni-L, with a concomitant loss of the hydride bridge in form of a proton. Ni-L is described by a formal Ni^{1+} state. Several Ni-L forms have been reported, which most likely differ with respect to the location of the proton. Light-induced structural changes have also been reported for the Ni-R states.

It has been demonstrated by X-ray crystallography that externally added carbon monoxide binds terminally to the nickel ion and inhibits the enzyme. Two distinct CO-inhibited redox states have been characterized by spectroscopy: a paramagnetic Ni-CO and an EPR-silent Ni-SCO state. Both are light sensitive, leading to photodissociation of the extrinsic CO ligand; Ni-CO converts to the same Ni-L state as Ni-C, while Ni-SCO converts to Ni-SI_a. Only four-coordinated nickel in a low oxidation state, either Ni^{2+} as in Ni-SI_a or Ni^{1+} as in Ni-L, has the affinity to bind carbon monoxide.

[NiFe] hydrogenases have the advantage that they are less oxygen sensitive than [FeFe] hydrogenases, with the O_2 inhibition being often reversible, while there are a few hydrogenases, for example, from Knallgas or hyperthermophilic bacteria, which are even oxygen tolerant. Models for their oxygen tolerance are currently being developed.

[FeFe] Hydrogenase

Structure

[FeFe] hydrogenases are composed of several subunits depending on the organisms. The core structure contains the H-cluster that consists of a bimetallic [FeFe] center, which is connected to an $[Fe_4S_4]$ cluster by a thiolate of a cysteine residue (Figure 2(c)). Most of [FeFe] hydrogenases contain additional $[Fe_4S_4]/[Fe_2S_2]$ clusters, whereas the [FeFe] hydrogenase from green algae contains only the H-cluster and no other iron-sulfur cluster.

The crystal structures of the periplasmic [FeFe] hydrogenase from *Desulfovibrio desulfuricans* and the cytoplasmic [FeFe] hydrogenase from *Clostridium pasteurianum* (CpI), have been determined at atomic resolution (Figures 2(a) and 2(b)). They revealed the unique features of this type of enzyme. CpI comprises only one amino acid chain and possesses an H-cluster, three $[Fe_4S_4]$ and one $[Fe_2S_2]$ cluster. The [FeFe] hydrogenase from *D. desulfuricans* consists of two subunits harboring the H-cluster and two $[Fe_4S_4]$ clusters; the structure of the active site is very similar to CpI.

As found for the active site of [NiFe] hydrogenases, the irons are also coordinated by nonprotein ligands, such as CO and CN^- (Figure 2(c)), which stabilize the iron in low oxidation states (here $Fe^{1+/2+}$). One CO is bridging the two iron ions. The diatomic ligands have been assigned on the basis of the surrounding amino acid residues: the CO ligands are located in a hydrophobic pocket and the CN^- ligands form hydrogen bonds to nearby amino acid residues. This assignment is corroborated by Fourier transform infrared spectroscopy (FTIR) spectroscopy. Furthermore, the two iron ions are bridged by a dithiolate unit. The apical atom was identified as a nitrogen atom by pulse EPR techniques. A hydrophobic gas-access tunnel has been found connecting the molecular surface and the distal iron Fe_d of the H-cluster. This gas-access tunnel is much shorter than that of [NiFe] hydrogenases. Fe_d is proposed as the initial site of dihydrogen conversion. In CpI the $[Fe_4S_4]$ clusters, serving as an electron transfer chain, are located within a distance of approximately 11 Å. The distal $[Fe_4S_4]$ cluster in CpI is coordinated by three cysteine and one histidine residues, as found in the case of [NiFe] hydrogenases. The other $[Fe_4S_4]$ clusters are coordinated by four cysteine residues.

Catalytic Cycle of [FeFe] Hydrogenase

Application of spectroscopic and electrochemical techniques, together with quantum chemical calculations, has led to the proposal of a reaction scheme for [FeFe] hydrogenases (**Figure 2(d)**). The most oxidized state, H_{ox}^{air} , is inactive and probably has an OH^- (or H_2O) ligand bound to the open coordination site at the distal Fe_d in the di-iron subcluster; in this state both irons are low-spin Fe^{2+} . One-electron reduction of the H_{ox}^{air} state leads to the H_{trans} state in which the $[Fe_4S_4]^{2+}$ subcluster is reduced to $[Fe_4S_4]^+$, but both irons of the binuclear [FeFe] subcluster remain Fe^{2+} . As a result of conversion of the H_{trans} state, the enzyme reaches the catalytically active state H_{ox} . This state is EPR-active with an oxidized $[Fe_4S_4]^{2+}$ and a mixed $Fe^{1+}Fe^{2+}$ subcluster ($S=1/2$), in which the terminal OH^-/H_2O ligand at Fe_d is removed. The substrate H_2 can now bind to the active site, that is, at Fe_d . In this process, the dihydrogen bond is polarized and heterolytic splitting occurs, with the nitrogen in the dithiolate bridge acting as a suitable base. In the most reduced state, H_{red} , the bridging CO could break the bond to the proximal Fe_p and move toward the distal Fe_d site, allowing the hydride to occupy the rather stable bridging position between the two iron ions. Alternatively, the hydride could remain bound end-on at the distal iron. In the catalytic cycle, the valence of the $[Fe_4S_4]^{2+}$ subcluster is retained, whereas the [FeFe] subcluster is proposed to change the valence from $Fe^{2+}Fe^{2+}$ to $Fe^{1+}Fe^{1+}$ (**Figure 2(d)**). It is not yet entirely clear if the proton reduction (H_2 production) and H_2 oxidation (uptake reaction) will follow the same mechanism.

The [FeFe] hydrogenase is (reversibly) inhibited by carbon monoxide. The external CO binds to the open coordination site of the distal Fe_d in the H_{ox} state resulting in an EPR-active state, termed $H_{ox}-CO$. The enzyme is very sensitive to molecular oxygen, which restricts its use in biotechnological processes. The oxygen has been proposed to initially bind to the open coordination site (e.g., in H_{ox}) before further reacting and destroying the H-cluster. The mechanism of the oxygen inhibition and destruction of the cluster has not yet been elucidated.

[Fe] Hydrogenase

Structure

[Fe] hydrogenases are found only in methanogenic archaea, which convert carbon dioxide with H_2 to methane in a complex reaction. [Fe] hydrogenases catalyze the reversible reduction of the substrate methenyltetrahydromethanopterin (methenyl- H_4MPT^+) with H_2 to methylene- H_4MPT and a proton by transferring a hydride ion to the *proR* position of the C14a carbon of methylene- H_4MPT . Its structure and catalytic reaction mechanism differ from those of [NiFe] or [FeFe] hydrogenases but it shows the typical reactions of hydrogenase (e.g. H^+/H_2 exchange; para/ortho H_2 conversion).

The crystal structures of [Fe] hydrogenase from *Methanocaldococcus jannaschii* and *Methanopyrus kandleri* have been determined at atomic resolution. Furthermore, the hydrogenase-substrate complex has also been described (PDB ID 3 H65). [Fe] hydrogenase is a homodimeric protein and comprises two peripheral units bound to the central globular unit in a linear manner (**Figure 3(a)**). These subunits form an intertwined

intersubunit helix bundle. Structural change from the open to the closed form occurs when the substrate methenyl- H_4MPT^+ binds to the enzyme. The cleft between the peripheral unit and the central globular unit in the open form has a depth of 13 Å. The enzyme harbors a unique iron-cofactor (iron-guanylylpyridinol, FeGP) and contains no iron-sulfur clusters. The iron is five-coordinated; ligands are a thiolate from a cysteine, two CO molecules, a nitrogen (sp^2) and an acyl-carbon from the pyridinol of the GP (see **Figure 3(b)**). The open coordination position is supposed to be the site of dihydrogen activation.

Catalytic Mechanism of [Fe] Hydrogenase

The catalytic mechanism of [Fe] hydrogenase differs fundamentally from that of the other type of hydrogenases. **Figure 3(c)** shows the proposed reaction scheme. In the first step, the methenyl- H_4MPT^+ substrate binds to the open form and the cleft between the two domains is closed (**Figure 3(c)** I/II). The substrate is probably activated in this process leading to a carbocationic character of the carbon C14a. The dihydrogen molecule binds side-on to the open coordination site of the iron of the activated FeGP complex (**Figure 3(c)** III). The oxidation state of the iron is proposed to be low-spin Fe^{2+} . The hydrogen is heterolytically cleaved by the adjacent C14a carbocation acting as a Lewis acid and the hydride is transferred to the methenyl- H_4MPT^+ (**Figure 3(c)** IV). The proton derived from the heterolytic cleavage of H_2 is accepted by a base (a cysteine thiolate or the pyridinol hydroxylate group) and is expected to exchange quickly with protons of the bulk solvent.

The dinuclear C-Fe arrangement of the [Fe] hydrogenase resembles that of the dinuclear metal centers of the [NiFe] and [FeFe] hydrogenases with respect to arrangement and the presence of a low-spin Fe^{2+} . The only variable is the first cation which can be a nickel, an iron, or even a carbon. This suggests a convergent evolution of the H_2 -activation machinery.

See also: **Bioenergetics:** Algal Hydrogen Production; Bioenergetics: General Definition of Principles; Cytochrome *c*; Iron-Sulfur Proteins; Renewable Hydrogen from Biomass; **Protein/Enzyme Structure Function and Degradation:** Enzyme Inhibitors; Enzyme Kinetics; Substrate Binding, Catalysis, and Product Release.

Further Reading

- Armstrong FA, Belsey NA, Cracknell JA, et al. (2009) Dynamic electrochemical investigations of hydrogen oxidation and production by enzymes and implications for future technology. *Chemical Society Reviews* 38: 36–51.
- Cammack R, Frey M, and Robson R (2001) *Hydrogen as a Fuel*. London: Taylor and Francis.
- Fontecilla-Camps JC, Amara P, Cavazza C, Nicolet Y, and Volbeda A (2009) Structure–function relationships of anaerobic gas-processing metalloenzymes. *Nature* 460: 814–822.
- Hiromoto T, Warkentin E, Moll J, Ermler U, and Shima S (2009) The crystal structure of an [Fe] hydrogenase–substrate complex reveals the framework for H_2 activation. *Angewandte Chemie International Edition* 48: 6457–6460.
- Lenz O, Ludwig M, Schubert T, et al. (2010) H_2 conversion in the presence of O_2 as performed by the membrane-bound [NiFe] hydrogenase of *Ralstonia eutropha*. *ChemPhysChem* 11: 1107–1119.

- Lubitz W and Tumas W (guest eds.) (2007) Hydrogen (100th Thematic Issue). *Chemical Reviews* 107: 3899–4435.
- Ogata H, Lubitz W, and Higuchi Y (2009) [NiFe] hydrogenases: Structural and spectroscopic studies of the reaction mechanism. *Dalton Transactions*, 7577–7587.
- Ogo S (guest ed.) (2010) Bioinspired catalysis (special issue). *Dalton Transactions*, 2949–3136.
- Pandelia ME, Ogata H, and Lubitz W (2010) Intermediates in the catalytic cycle of [NiFe] hydrogenase: Functional spectroscopy of the active site. *ChemPhysChem* 11: 1127–1140.
- Pickett C and Best S (guest eds.) (2005) Hydrogenases (special issue). *Coordination Chemistry Reviews* 249: 15–16, 1517–1690.
- Tard C and Pickett CJ (2009) Structural and functional analogues of the active sites of the [Fe], [NiFe], and [FeFe] hydrogenases. *Chemical Reviews* 109: 2245–2274.
- Thauer RK, Kaster A-K, Goenrich M, et al. (2010) Hydrogenases from methanogenic archaea, nickel, a novel cofactor and H₂ storage. *Annual Review of Biochemistry* 79: 507–536.

Immunoglobulin (Fc) Receptors

M Hogarth, Burnet Institute, Prahran, VIC, Australia

Published by Elsevier Inc.

Glossary

Autoimmune disease Diseases where the immune system attacks the hosts' tissues and/or organs. Examples include rheumatoid arthritis and systemic lupus erythematosus.

Hypersensitivity Powerful and frequently destructive inflammatory reaction often induced by antibodies in autoimmune diseases.

Leukocyte White blood cell including eosinophil, polymorphonuclear leukocyte or neutrophil, macrophage, monocyte, mast cell, basophil, and lymphocyte.

Phagocyte Cell that actively ingests particles, typically macrophage or neutrophil.

Phagocytosis Process distinct from pinocytosis by which particles are taken into the cell. Unlike pinocytosis, considerable intracellular and membrane reorganization is required for the uptake of the particle.

Pinocytosis Process by which cells sample small amounts of fluid from their surrounding environment. This usually involves internalization of small vesicles formed from the invagination of the cell membrane.

Immunoglobulin Fc receptors (FcRs) are cell surface molecules that bind the Fc portion of immunoglobulin. Receptors have been identified for all immunoglobulin classes and subclasses with the exception of IgD. Note that microbial proteins that can also bind the Fc portion of immunoglobulin are not receptors but should be regarded only as Fc-binding proteins. Such proteins include, for example, staphylococcal protein A.

General Comments

When antibodies contact their specific antigen, they form an immune complex and in the case of most pathogens, many antibodies will coat a single macromolecule or pathogen fragment leading to the formation of multimeric immune complexes. These multimeric, and therefore oligovalent, complexes of antibody and antigen then interact with specific cell surface Fc receptors and this interaction subsequently initiates signaling cascades inside cells.

The affinity of the interaction between the different Ig subclasses and specific Fc receptors varies by as much as 1000-fold, ranging from low micromolar affinity for the so called 'low-affinity' IgG receptors to nanomolar affinity for the high-affinity IgG receptor. Furthermore, the high affinity IgE receptor binds IgE with femtomolar affinity. Such a vast range of affinities has specifically evolved to suit the biological activities of particular receptor types.

The nomenclature of FcRs is, as the name suggests, based on the capacity to bind the Fc portion of a particular class of antibody, for example, IgM, Fcμ; IgG, Fcγ. The letter R designates

a receptor and a Roman numeral designates a particular subtype of FcR, for example, FcγRII. These can be further subclassified by letters a, b, c, etc., for example, FcγRIIIa. Note also that several receptors are also given an alternative designation as part of the cluster of differentiation (CD) system of classifying leukocyte surface molecules; thus, FcγRI is known also as CD64, FcγRII as CD32, FcγRIII as CD16, and FcεRII as CD23.

The vast majority of FcRs are found on leukocytes FcεR and FcαR where they are involved in the regulation of cellular activity by antibody. Most leukocyte FcRs are primarily designed to activate cells and with one notable exception, are multisubunit complexes. However, a discrete and unique class of FcR is able to downregulate leukocyte activity following their binding of immune complexes.

Fc receptors are largely found on leukocytes, which have been the principal subject of most research. However, several other FcRs have been identified, present on epithelial and endothelial cells with specific functions related to transport and recovery of immunoglobulins and are considered separately below.

Leukocyte FcRs

Most leukocyte FcRs are multisubunit complexes requiring a ligand-binding chain noncovalently associated with a signaling chain and in some cases with an additional signal amplifying subunit (Figures 1 and 2). The biochemical and functional characteristics of the most frequently studied FcR of leukocytes are detailed below.

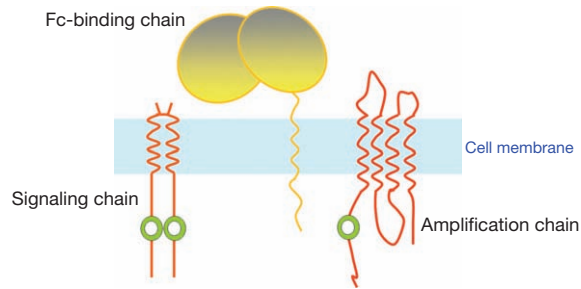


Figure 1 Schematic representation of a typical multisubunit Fc receptor. The typical components of many Fc receptors found on leukocytes. A ligand binding chain, usually a type-1 transmembrane glycoprotein, is noncovalently associated with other subunits: a signaling chain, often the FcR- γ chain, which is a 24-kDa covalent homodimer wherein each monomer contains an ITAM, a specialized tyrosine-containing signaling motif (green annulus). In at least two cases the ligand-binding chain is further associated with a tetra transmembrane signal-amplifying chain.

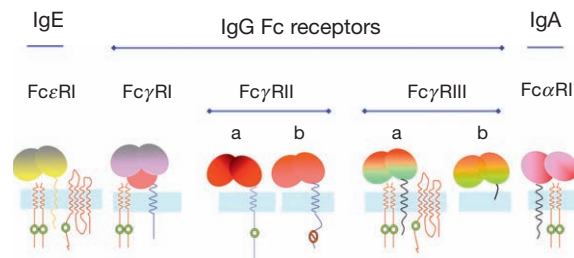


Figure 2 Schematic representation of leukocyte FcR showing configuration of signaling-competent receptor complexes. FcγRII (not shown). ITAMs are shown as green annulus; the FcγRIIb ITIM as a red crossed circle.

IgE FcRs

The high-affinity IgE receptor FcεRI binds monomeric IgE with an affinity of greater 10^{10} M^{-1} and is found principally on mast cells and basophils and in a modified form on eosinophils and monocytes where it lacks the amplifying beta chain (see below). The binding of IgE to this receptor and the subsequent cross-linking of the bound IgE by antigen induces one of the most potent pharmacological responses in biology. In adverse situations this results in a classic type-I hypersensitivity reaction clinically manifested as allergy, for example, hay fever, penicillin, or food allergies. The FcεRI signaling is mediated via a noncovalent association with a low-molecular-weight transmembrane dimer called 'Fc gamma chain (FcR γ)'. This chain is shared by several other Fc receptors. Unlike the high-affinity IgE receptor, the FcεRII (CD23) binds IgE with a lower affinity (10^7 M^{-1}). This is a type-II membrane glycoprotein with lectin-like activity and is found on many leukocytes including T-cells, B-cells, and monocytes. Its function appears largely to regulate activity of immune responses.

These receptors are biochemically distinct. FcεRI is a transmembrane glycoprotein and is a member of the Ig superfamily with two extra cellular domains, each of which belong to the Ig super family. By contrast, FcεRII is a type-II membrane glycoprotein with lectin-like activity.

IgG Receptors

In humans, there are three main subtypes of leukocyte IgG Fc receptors – FcγRI, FcγRII, and FcγRIII (Figures 2, 3(a), and 3(b)). Splice variants and multiple gene products are known for FcγRII and FcγRIII. (Note that while the mouse and humans share many features of IgG Fc receptors, there are significant differences, i.e., mice do not have orthologs of FcγRIIa, FcγRIIc, or FcγRIIb; humans lack an ortholog of mouse FcγRIV.) The FcγR are widespread and every leukocyte with the exception of $\alpha\beta\text{T}$ lymphocytes expresses one or more of these receptors. The different receptors can be distinguished by amino acid sequence and by the use of specific monoclonal antibodies and by differences in affinities for IgG.

FcγRI (also known as CD64) is the high-affinity receptor for IgG (10^8 M^{-1}) and is restricted largely to monocytes, macrophages, and neutrophils.

FcγRII (CD32) is composed of two major subtypes: FcγRIIa and FcγRIIb, which are separate gene products and are biochemically distinct. The extracellular domains are highly related in sequence but their transmembrane and cytoplasmic tails are distinct, which reflects their distinct signaling activities of activation or inhibition, respectively. Indeed, FcγRIIb is the only inhibitory FcγR. (see below). Both bind monomeric immunoglobulin IgG with very low affinity (one micromolar) evident by the rapid on and off rate; however, since these low-affinity receptors decorate the cell membrane they can effectively, and very avidly, bind oligovalent complexes of antigen and antibody. FcγRIIa is found on virtually every leukocyte including platelets. FcγRIIb, similarly, is found on virtually every leukocyte including B-cells but not platelets. Note that a third form, FcγRIIc, has been described but little is known of its function. Its extracellular domains are closely related to those of the inhibitory receptor FcγRIIb whereas its cytoplasmic domain contains an immunoreceptor tyrosine-based activation motif (ITAM), as does FcγRIIa.

FcγRIII (CD16) also has two subtypes: FcγRIIIa and FcγRIIIb. Both are low-affinity receptors though they bind with IgG with 10-fold higher affinity than the FcγRII low-affinity receptors. FcγRIIIa is found on natural killer cells, macrophages, and some monocytes, whereas FcγRIIIb is found only on neutrophils.

All of these FcRs belong to the Ig super-family. Each has two extracellular Ig-like domains with the exception of FcγRI which has three (Figure 2). Another interesting difference is that while FcγRI, FcγRIIa, and FcγRIIa are transmembrane glycoproteins, FcγRIIb is a glycosylphosphatidylinositol (GPI)-anchored receptor (Figure 2).

IgA and IgM FcRs of Leukocytes

A single receptor specific for IgA has been defined – FcαRI which is found on neutrophils and monocytes (Figure 2). The affinity for IgA is approximately 10^7 M^{-1} , but immune complexes of IgA binding to FcαRI are potent inducers of cell activation. A receptor that is able to bind both IgA and IgM, the FcαR of Fcα μ R, has also been defined though there has been relatively little characterization of this.

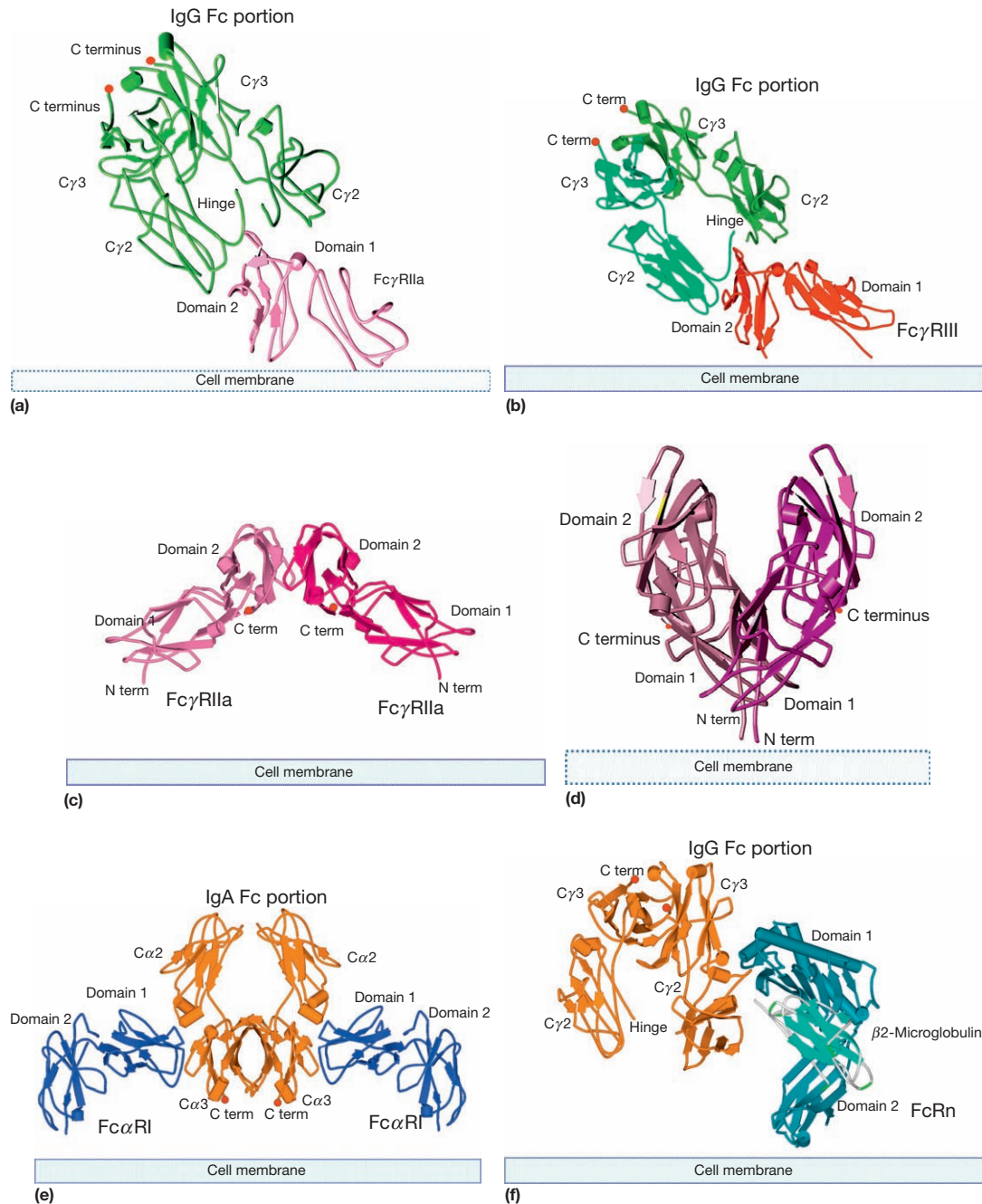


Figure 3 FcR structure. Shown are the alpha carbon backbones of selected Fc receptors and ligands. (a) Comparison of the interaction of IgG with Fc γ RIIa or Fc γ RIIIa; from the database: PDB 3RY6; PDB, 1E4K. The receptors' second domain and interdomain region interacts with the lower hinge region and hinge proximal parts of the IgG Fc giving the characteristic 1:1 stoichiometry. (b) Comparison of the two dimeric forms of Fc γ RIIIa (PDB 1FCG) and the second derived from PDB 3RY6. This receptor is unique as it forms a noncovalent homodimer. In the case of the left dimer, binding of IgG is not possible because steric clashes between the dimer and C γ 2 prevent proper interaction of domain 2 with the lower hinge of IgG. However, in the second dimer the binding surfaces are sufficiently separated to allow interaction with an IgG Fc for each receptor in the dimer, that is, one dimer interacts with one or two Fcs. (c) Complex of Fc γ RI and IgA (PDB, 1OW0). Interaction between the tip of the first domain of Fc α RI and the C α 2/C α 3 junction of each Fc chain gives the 2:1 stoichiometry. (d) Structure of FcRn and IgG complex (PDB, 111A). Interaction of the FcRn first domain with amino acids in the junction of C γ 2 and C γ 3 of a single Fc chain. The β 2 microglobulin typically associated with MHC class I molecules is shown associated with FcRn ectodomains. In all figures the polypeptide chains are represented as α -carbon backbones; strands as arrows and α -helices as barrels. The C terminus of each polypeptide chain is shown in red. Constant regions of IgA or IgG are indicated – C α or C γ and the position of the hinge are shown. Orientation of the polypeptides to the membrane is assumed.

Of most recent interest is the identification of an IgM receptor, FcμR, which had historically been very difficult to identify but relatively little biology is known of this receptor.

Sites of Interaction with Ig Ligands

The Ig binding sites of FcγRI, FcγRII, and FcγRIII are located in similar regions in these structurally related but functionally distinct receptors (Figures 3(a) and 3(b)). The identity of the Ig-binding sites of many of the receptors has been extensively investigated and defined by mutagenesis and by X-ray structures. In the case of the IgG receptors FcγRII and FcγRIII – and probably FcγRI – the site is located in the second domain together with sequences at the junction of the first and second domains as seen in the similar crystallographic structures (Figure 3(a)). Interestingly, FcγRIIa can form distinct dimers only one of which is able to interact with IgG (Figures 3(c) and 3(d)). The stoichiometry of FcγR:IgG interaction for all IgG receptors is 1:1 (Figure 3(d)).

The interaction of IgE and the high-affinity IgE receptor FcεRI occurs in the second domain of FcεRI and the stoichiometry is 1:1. Crystallographic studies show the interaction is similar to that of the related FcγR and IgG. In the case of the low-affinity IgE receptor, FcεRII, the head groups of the triple helix interact with IgE also in a 1:1 stoichiometry.

Although distantly related to the FcγR and FcεRI, the IgA receptor, FcαRI, binds IgA to the extreme tip of its first domain which allows it to form a rather unique 2:1 stoichiometry with IgA (Figure 3(d)).

Receptor Binding Sites in Immunoglobulin

FcγRI, FcγRII, and FcγRIII interact with sequences in the lower hinge and adjacent areas of Cγ2 of the IgG Fc (Figure 3(a)). Differences in the nature and extent of these interactions explain the differences in affinity of FcR:Ig interaction. FcεRI interacts with sequence between Cε2 and Cε3, which is distinct from the area of the sites of interaction of FcεRII. The nature of these interactions clearly defines the stoichiometry. FcαRI interacts with the area between Cα2 and Cα3. Currently, there is no information on the interactions of FcαμR and IgA/IgM or for FcμR and IgM.

Signal Transduction Complexes

With the exception of FcεRII, there has been a great deal of investigation of the signal transduction by leukocyte Fc receptors which have become one of the most widely studied groups of receptors. The binding of antibody or immune complexes to these can be a potent inducer of biochemical signals, but FcγRIIa and FcγRIIb have additional unique biochemistry. The structurally related leukocyte receptors – FcγRI, FcγRII, FcγRIIIa, and FcεRI – as well as the more distantly related FcαRI all signal by very similar mechanisms dependent on ITAM which involve the *syk* and *src* family protein tyrosine kinases. To signal, these receptors must form multisubunit

signaling complexes composed minimally of the Fc-binding chain (all of which are the Ig superfamily members), noncovalently associated with a low-molecular-weight hetero- or homodimeric signaling chain – the common FcR-γ chain or in the case of FcγRIIIa in NK cells the CD3-ζ chain (Figures 1 and 2).

The ITAM motif that is necessary for recognition by *syk* and *src* kinases is contained in these associated low-molecular-weight dimers. Cross-linking of the FcR by immune complexes, or by antigen to cell-bound Ig, aggregates the receptors leading to phosphorylation of tyrosine in the ITAMs and the subsequent induction of the signaling cascades. Similar mechanisms of leukocyte activation are well described for the antigen receptors of T- and B-lymphocytes.

FcγRIIa stands out as an exception to the multisubunit paradigm for ITAM-dependant immunoreceptor signaling (Figures 2 and 3(d)). While aggregation of this receptor leads to *syk* and *src* kinase activation, this receptor does not require association with the common FcR-γ chain. Since a noncanonical, the ITAM is contained within its cytoplasmic tail. Indeed, FcγRIIa is one of the few immunoreceptors to have the ITAM motif contained in the ligand-binding chain. This receptor forms either of two unique dimers on the cell surface; one dimer form is incapable of binding IgG (Figure 3(c)). However, a second dimer of receptors is formed with either one or both receptors liganded with IgG (Figure 3(d)). As a result, the cytoplasmic ITAM motifs are juxtaposed thereby forming the necessary dimeric ITAM configuration for signal transduction, equivalent to that of the FcRγ chain that associates with other activating FcR.

FcγRIIb is not an activating Fc receptor; indeed, its role is to inhibit and regulate the activities of other ITAM-containing activating Fc receptors and the antigen receptor of B-cells. The cytoplasmic tail of FcγRIIb contains an immunoreceptor tyrosine inhibitory motif (ITIM) which on co-aggregation with ITAM-containing activating receptors recruits phosphatases which in turn downregulate or eliminate ITAM-dependent signal transduction. Its best-characterized role is its co-engagement with the antigen receptor of B-lymphocytes where it regulates B-cell activation by antigen.

Other Leukocyte FcRs

In recent years, especially with the output of the human and mouse genome projects, new cell surface molecules with FcR-like activity have been described though they are currently poorly characterized. Nonetheless, for the sake of completeness they have been included here. The most advanced characterizations have involved the leukocyte in receptor cluster, which is a large family of cell surface molecules present on leukocytes but most of these do not show immunoglobulin binding activity. Biochemically, these molecules include members with multiple Ig-like domains and with tissue distributions ranging from the highly restricted to the very broad. The precise biochemical nature and biological roles are poorly understood but will no doubt be revealed in time, although from sequence comparisons some members are likely to have ITIM-like inhibitory activity and others ITAM-dependant activation roles.

Nonleukocyte Fc Receptors

Two major Fc receptors have been defined: the poly Ig receptor (poly IgR) and FcRn.

The poly Ig receptor is responsible for the binding and transport of the secretory immunoglobulins – IgA and IgM. It is found on secretory epithelium, notably in the lactating breast, as well as mucosal surfaces. This is a cell surface molecule composed of five domains: the IgA and IgM-binding sites occur in the first two domains which when attached to sIgA or sIgM are cleaved from the receptor and thereby form the so-called ‘secretory component’ associated with secretory IgA and IgM.

FcRn was originally described as a receptor on epithelial and endothelial cells. In humans, its main function is the recycling of immunoglobulins to/from the circulation which ensures the prodigious half-life of IgG. Intact and undenatured IgG which is taken into endothelial cells by pinocytosis is recycled to the circulation following its pH-dependent attachment to FcRn in the acidifying pinocytotic vesicle where it undergoes acid-dependent attachment to FcRn which results in recycling to the endothelial cell surface and release at the neutral pH of the circulation. Recently, it has become apparent that FcRn is expressed on leukocytes where it internalizes IgG including IgG:antigen complexes; such an internalized antigen can then enter antigen-processing pathways of the leukocyte.

FcRn is biochemically distinct from all other leukocyte and nonleukocyte Fc receptors (Figure 3(f)). It is a transmembrane glycoprotein but it is more structurally related to the major histocompatibility molecules (MHC class I) whose principal function is antigen presentation to T-cells. However, it is sufficiently biochemically different from these which makes it unable to be involved directly in presentation.

Biological Role of Leukocyte FcRs

The leukocyte receptors FcεRI, FcεRII, FcγRI, FcγRII, and FcγRIII participate in a range of normal functions most of which is associated with the activation of leukocytes and the modulation of immunity which then leads to specialized functions associated with different cell types, for example, the degranulation of mast cells and the release of inflammatory mediators caused by aggregation of the high-affinity IgE receptor by IgE: antigen complexes. However, for the most part, FcγRI, FcγRII, and FcγRIII appear to be mostly concerned with antibody-dependent activation and modulation of innate and adaptive immunity rather than the historical view that Fc receptors on cells such as macrophages, dendritic cells, and other phagocytes serve the principal role for the phagocytic removal of immune complexes.

It is now clear that these receptors play a far more active and critical role rather than one of simple garbage disposal of immune complexes, a function which, in humans, is largely filled by the complement system and, in particular, CR-1 on erythrocytes.

This capacity to bind immune complexes, though not in itself pathogenic, can in the presence of aberrant immune complexes, such as autoantibody, lead to the potent and continuous stimulation of leukocytes, leading in turn to pathological

inflammation and tissue destruction. This is clearly the case in diseases such as systemic lupus erythematosus, rheumatoid arthritis, and glomerulonephritis, allergies which are associated with major morbidity. Attempts to interfere with the interaction of immune complexes with Fc receptors are now forming the basis of novel approaches to the treatment of those diseases.

Key Events in the Analysis of FcR

The existence of specific receptors for immune complexes was noted in the early 1970s by Metzger and by Basten and their co-workers. The observation by Metzger and Bevan that engagement of these receptors could induce a measurable intracellular biochemical event was an exciting observation of importance, not only for Fc receptors but also for immunologically important receptors generally. There have been several major milestones in the application of new technologies to the characterization of Fc receptors – Jay Unkeless produced the first monoclonal antibody defining an Fc receptor and Hibbs and colleagues defined first function associated genetic polymorphism using monoclonal antibodies. The genetics revolution led to the first cloning in the mid-1980s of the FcR gene simultaneously and independently by Hibbs, Ravetch, Mellman, and Kinet. By the early 1990s, Ierino and coworkers had established that blockade of FcR may provide useful therapeutic approaches for the treatment of disease. Finally, in more recent times, the discovery of the elusive IgM receptor by Kubagawa and co-workers as well as the three-dimensional structures of FcR have become available for the first time through the work of Maxwell and Powell, Jardetzky, Sun, Sonderrmann, and others.

Conclusion

The FcRs are a large family of immunoregulatory molecules that play crucial roles in antibody effector systems, leading in normal immune responses to the activation or regulation of leukocytes.

See also: **Signaling:** Cytokines; Fibroblast Growth Factor Receptors and Cancer-Associated Perturbations.

Further Reading

- Hamburger AE, West AP, Jr., and Bjorkman PJ (2004) Crystal structure of a polymeric immunoglobulin binding fragment of the human polymeric immunoglobulin receptor. *Structure* 11: 1925–1935.
- Hulett MD and Hogarth PM (1994) Molecular basis of Fc receptor function. *Advances in Immunology* 57: 1–127.
- Ierino F, Powell MS, McKenzie IFC, and Hogarth PM (1993) Recombinant soluble human FcγRII: Production, characterization and inhibition of the Arthus reaction. *Journal of Experimental Medicine* 178: 1617–1628.
- Kubagawa H, Oka S, Ohno H, and Wang JY (2009) Identity of the elusive IgM Fc receptor (FcμR) in humans. *Journal of Experimental Medicine* 206(12): 2779–2793.
- Lu J, Ellsworth JL, Hamacher N, Oak SW, and Sun PD (2011) Crystal structure of Fcγ receptor I and its implication in high affinity γ-immunoglobulin binding. *Journal of Biological Chemistry* 286(47): 40608–40613 [Epub September 29].
- Metzger H (1999) It's spring and thoughts turn to... allergies. *Cell* 97: 287–290.

- Nimmerjahn F and Ravetch JV (2010) Antibody-mediated modulation of immune responses. *Immunological Reviews* 236: 265–275.
- Ramsland PA, Farrugia W, Bradford TM, et al. (2011) Structural basis for Fc gammaRIIIa recognition of human IgG and formation of inflammatory signaling complexes. *Journal of Immunology* 187: 3208–3217.
- Sondermann P, Huber R, Oosthuizen V, and Jacob U (2000) The 3.2-A crystal structure of the human IgG1 Fc fragment-Fc gammaRIII complex. *Nature* 406: 267–273.
- Ward ES and Ober RJ (2009) Multitasking by exploitation of intracellular transport functions: The many faces of FcRn. *Advances in Immunology* 103: 77–115.

Indicators of Intracellular Calcium

M Giacomello and M Campeol, University of Padua and Venetian Institute of Molecular Medicine, Padua, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

BAPTA 1,2-bis(*o*-aminophenoxy)ethane-*N,N,N',N'*-tetraacetate is a calcium and magnesium chelator.

EF hand An helix–loop–helix motif present in many Ca^{2+} -binding proteins. As a rule, this domain consists of two alpha helices orthogonal to one another and linked by a stretch of about 12 amino acids that bind Ca^{2+} ions.

EGTA The Ca^{2+} chelator ethylene glycol-*O,O'*-bis (2-amino-ethyl)-*N,N,N',N'*-tetraacetic acid.

Fluorescence resonance energy transfer (FRET) Known also as Förster resonance energy transfer. It is the nonradiative energy transfer between a donor and an acceptor chromophore. This process requires a partial overlap in the emission spectrum of the donor and the excitation spectrum of the acceptor. In addition, the two chromophores should be positioned close enough (usually less than 10 nm) and have the appropriate relative orientation.

Introduction

Many physiological and pathological processes are linked to variations of the intracellular Ca^{2+} homeostasis. In addition to its structural role (e.g., in bones and teeth), the Ca^{2+} ion has a crucial role as a cell messenger. Its signaling ability relies on the possibility to shape the changes of its concentration in terms of amplitude, kinetics, and spatial pattern. This is accomplished by a wide number of proteins located in the cell membranes, in the cytosol, and in the subcellular compartments, such as the mitochondria and the endoplasmic reticulum (ER). The need to monitor the variations of this highly dynamic system has led to the development of specific indicators with different affinities for Ca^{2+} , adequate to analyze a wide range of its intracellular free concentration ($[\text{Ca}^{2+}]$). The possibility to choose molecules with various affinities (K_d) is particularly important, as $[\text{Ca}^{2+}]$ levels are highly heterogeneous not only inside the cytoplasm, where the presence of micropools with high $[\text{Ca}^{2+}]$ has been demonstrated, but also within organelles as the ER and Golgi, that are considered the major intracellular Ca^{2+} stores.

Ca^{2+} probes bind the ion reversibly and, as a consequence, vary their physicochemical characteristics. Among the physicochemical properties that characterize Ca^{2+} sensors, the most common is the difference in the emission and/or absorbance of light induced by the binding of Ca^{2+} . Thus, the changes in the intracellular $[\text{Ca}^{2+}]$ are defined by the respective amount of the two forms of the sensor, that is, the Ca^{2+} -bound and the Ca^{2+} -free form.

Ca^{2+} indicators can be classified according to different criteria: they could be nonratiometric and ratiometric, chemical (synthetic; Table 1) and genetically encoded (Table 2).

In nonratiometric molecules only one of their properties, for example, their fluorescence intensity, varies as a consequence of Ca^{2+} binding. Experiments with them are simple: the acquisition system needed to collect data involves a single excitation/emission setting. However, nonratiometric molecules fail to correct for the different absolute amount of dye loading that can vary from cell to cell or for focal plane shifts that could be present. On the contrary, ratiometric probes correct for these drawbacks: upon interaction with Ca^{2+} two

parameters, most commonly the excitation and emission spectra, vary simultaneously. In ratiometric measurements, the variations of $[\text{Ca}^{2+}]$ are proportional to the changes of the ratio between the two fluorescence intensity values.

The analysis of the synthetic and the genetically encoded Ca^{2+} indicators demands a more extensive discussion.

Chemical Indicators

The first sensors used to monitor cellular $[\text{Ca}^{2+}]$ variations have been organic compounds like murexide and azo dyes. Murexide, also known as ammonium purpurate or MX, is the ammonium salt of purpuric acid. This metallochromic indicator undergoes a color change in the presence of calcium. A similar behavior characterizes the azo dyes, that contain nitrogen in the form of an azo group ($-\text{N}=\text{N}-$) as part of their molecular structure. These two classes of molecules have been used for some years for the titration of Ca^{2+} , but have poor emission intensity and signal-to-noise ratio. Chlortetracycline, an antibiotic also used as a Ca^{2+} sensor, had higher changes in fluorescence intensity but its poor selectivity (it also interacts with Mg^{2+}) and its sensitivity to alterations of the membrane potential and pH limited its use.

In the 1980s, Roger Y Tsien synthesized the first fluorescent polycarboxylate Ca^{2+} indicators. These probes were designed starting from the structure of the selective Ca^{2+} chelator EGTA. The replacement of two methylene groups with two benzene rings leads to the building of a chromophore moiety. In these indicators, Ca^{2+} binding to the carboxyl groups changes the optical properties of the molecule. The parent compound of these polycarboxylate probes, BAPTA, displays large spectral shifts on Ca^{2+} binding and requires ultraviolet excitation, which is more cytotoxic than longer wavelengths. In addition, ultraviolet irradiation causes emission of intrinsic fluorescence by endogenous cell molecules, such as NADH (reduced nicotinic adenine dinucleotide; excitation at 340 nm and emission peaks at 440–470 nm). The presence of autofluorescence leads to inaccuracy in the fine estimation of the variations of free $[\text{Ca}^{2+}]$.

Quin 2, a tetracarboxylic acid derivative of quinoline, is one of these first-generation Ca^{2+} probes. As is the case for the

Table 1 Synthetic fluorescent Ca^{2+} indicators

Calcium indicator	K_d for Ca^{2+} (nM)	λ of excitation Ca^{2+} bound (nm)	λ of emission Ca^{2+} bound (nm)	Properties
Quin 2	115	339	492	Nonratiometric
Indo 1	230	330	401	Dual emission, ratiometric indicator
Fura 2	224	335 (340)	510	Dual excitation, ratiometric indicator
Bis-Fura 2	370	338	504	Dual excitation, ratiometric indicator
Mag Fura 2	25×10^3	329	508	Dual excitation, ratiometric indicator
Fluo3	400	506	526	Nonratiometric
Fluo4	345	494	516	Nonratiometric
Calcium green 1	190	504	532	Nonratiometric
Calcium green 2	550	503	536	Nonratiometric
Calcium green 5 N	14×10^3	506	532	Nonratiometric
Oregon green BAPTA-1	170	494	523	Nonratiometric
Oregon green BAPTA-2	580	494	523	Nonratiometric
X-Rhod 1	700	580	602	Nonratiometric
Rhod 2	570	552	581	Nonratiometric

Table 2 Genetically encoded Ca^{2+} indicators

Calcium indicator	K_d (μM) for Ca^{2+}	λ of excitation Ca^{2+} bound (nm)	λ of emission Ca^{2+} bound (nm)	Properties
Aequorin	~ 7		466	Bioluminescent
Camgaroo-1	7	490	514	Nonratiometric
Camgaroo-2	5.8	490	514	Nonratiometric
Ratiometric Pericam	1.7	490	514	Ratiometric
Inverse Pericam	0.2	490	514	Nonratiometric
Cameleon				
YC 2	1.2			
YC 2.1	0.1; 4.3			
YC 2.60	0.04			
YC 3.1	1.5	430	525–580 (YFP)	Ratiometric (FRET)
D1	0.8; 60			
D2	0.8; 26			
D2cpv	0.03; 3			
D3	1.2			
D3cpv	0.6			
D4	195			
D4cpv	64			
Tn-L15	0.72	430	525–580 (YFP)	Ratiometric
Tn-XL	2.5	430	525–580 (YFP)	Ratiometric
Tn-XXL	0.8	430	525–580 (YFP)	Ratiometric

other sensors, binding of Ca^{2+} leads to changes of some of the spectral properties of the compound, that is, to an emission maximum at 492 nm and to a considerable enhancement of the fluorescence emission intensity. Quin 2 has a very high affinity for Ca^{2+} (K_d 115 nM) making it ideal for studying $[\text{Ca}^{2+}]$ variations in intracellular resting concentration (100–200 nM). However, the low K_d affects the reliability of $[\text{Ca}^{2+}]$ measurements at μM levels as they may cause saturation of the sensor. In addition, the excitation wavelength of Quin 2 peaks at 339 nm and causes the appearance of significant autofluorescence. In conjunction with its low absorption coefficient, this leads to a poor signal-to-noise ratio, unless the cells are loaded with high amounts of the probe. For these reasons, Quin 2 started to be used more as an intracellular Ca^{2+} buffer rather than a real Ca^{2+} indicator. To overcome the problems of Quin 2, Tsien and coworkers synthesized new Ca^{2+} probes that combined a stilbene fluorophore with the structure of

the Ca^{2+} chelators EGTA and BAPTA. A new family of Ca^{2+} -binding molecules were thus obtained, among them Indo 1 and Fura, that had much higher brightness of fluorescence (about 30 times higher than Quin 2), with wavelength shifts upon Ca^{2+} binding and/or weaker affinity for Ca^{2+} .

Indo 1 is a BAPTA derivative with an increased fluorescence quantum yield with respect to Quin 2. $[\text{Ca}^{2+}]$ measurements with Indo 1 exploit the shift of emission wavelength from 475 to 401 nm upon Ca^{2+} binding. The fluorescence intensity ratio at the two wavelengths is proportional to the intracellular $[\text{Ca}^{2+}]$ changes and corrects for unequal dye loading or changes in the volume and shape of the cells. Three major problems unfortunately complicate the use of Indo 1: its rapid photobleaching, its photodegradation, and significant autofluorescence of cell components due to the excitation wavelengths used. While photobleaching of the indicator causes a decrease in the signal-to-noise ratio, the degradation

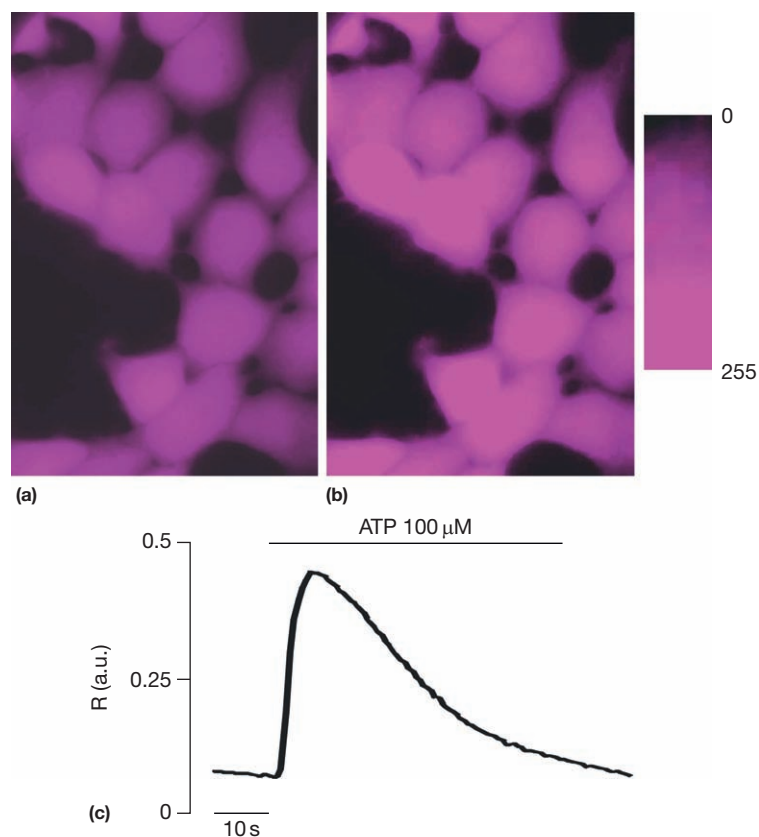


Figure 1 (a) Ratiometric image of HeLa cells loaded with 3 μM FURA2-AM under resting conditions. (b, c) In cells challenged with an agonist coupled to InsP3 production (as the purinergic agonist adenosine triphosphate (ATP)), Ca^{2+} is released from the endoplasmic reticulum (ER), leading to an increase of the cytosolic $[\text{Ca}^{2+}]$, which is reflected by the increased ratio. The trace in panel c represents the typical response of a cell loaded with FURA2-AM and challenged with 100 μM ATP, in a standard medium containing 1 mM Ca^{2+} .

by ultraviolet illumination renders the fluorescence of Indo 1 Ca^{2+} insensitive. This results in an underestimation of the true intracellular $[\text{Ca}^{2+}]$. The autofluorescence issue is also important, as both the excitation and the emission spectra of NADH overlap with those of Ca^{2+} -free Indo 1, leading to the increase of autofluorescence background.

Fura is also excited by ultraviolet illumination. Three different Fura molecules were generated: Fura 1, Fura 2, and Fura 3: the most commonly used is Fura 2 (Figure 1). This ratiometric probe has a K_d for Ca^{2+} of 224 nM and shifts absorption peak upon Ca^{2+} binding from 363 to 335 nm, the emission peak remaining unvaried. Dual excitation of Fura 2 is achieved at 340 and 380 nm since the absorption peaks at these wavelengths are closer to its isosbestic point. Other Fura 2 variants have been created in order to study signal transduction processes that involve Ca^{2+} spikes exceeding 1 μM values. One of these compounds, bis-Fura, was obtained by joining together one Ca^{2+} -binding site and two Fura 2 fluorophores. The lower K_d for Ca^{2+} of this probe decreases its Ca^{2+} -buffering activity. In addition, the higher changes in the ratio of bis-Fura 2, with respect to Fura 2, allows to lower the amount of dye to be loaded into the cells.

Furaptra, also called Mag Fura 2, had been originally designed to report changes in intracellular Mg^{2+} concentration (it binds it with a K_d of 1.9 mM). However, Mag Fura 2 also binds Ca^{2+} with a K_d of 25 μM . Since its K_d is well above the

physiological basal Ca^{2+} levels of most cells, Furaptra has been widely used to monitor intracellular Ca^{2+} variations in the range between 1 and 100 μM or to evaluate fast changes in the free $[\text{Ca}^{2+}]$, for instance in skeletal muscle during electrical stimulation.

In addition to ultraviolet light excited indicators, some compounds excited by visible wavelength have also been generated. Visible light causes lower cytotoxicity and lower background scattering than ultraviolet light. Among the visible light excited Ca^{2+} probes, Fluo3 and Fluo4, Calcium green and Oregon green BAPTA are the most commonly used. Fluo3 and Fluo4 share a similar structure. They are nonratiometric Ca^{2+} sensors that increase fluorescent emission intensity on Ca^{2+} binding. They have a relatively high K_d (400 and 345 nM, respectively) that accounts for the absence of Ca^{2+} -buffering activity under resting conditions. They differ in pH sensitivity and fluorescence, Fluo4 being more photostable and brighter than Fluo3. The Calcium green family of Ca^{2+} probes (that comprises Calcium green 1, 2, and 5 N, having different Ca^{2+} affinities) has higher quantum yield and brighter fluorescence at saturating amounts of Ca^{2+} than Fluo3 and Fluo4. Oregon green BAPTA is another family of indicators based on BAPTA (including Oregon green BAPTA-1 and -2) excited by a visible-wavelength Ca^{2+} indicator. These molecules have considerable photostability and low pH sensitivity, which affect other indicators like Fluo3. In addition to the problems of cytotoxicity,

pH sensitivity, and photodegradation, the main drawback in the use of chemical Ca^{2+} indicators is the method to introduce them inside living cells and the inability to target them to particular subcellular compartments. Synthetic Ca^{2+} probes have been introduced into the cytoplasm by techniques such as microinjection, transfer through liposomes, and electroporation. However, the transformation of virtually all the Ca^{2+} probes cited above into the acetoxymethyl (AM) ester form has greatly increased their popularity. Once in the cytoplasm, the AM forms of the probes undergo cleavage by intracellular esterases, leading to the release of the original form of the indicator in the cytoplasm, where it remains trapped since it is membrane impermeable. It is widely assumed that the AM forms of the dyes are completely cleaved but incomplete de-esterification cases have been reported, causing the permanence of Ca^{2+} -insensitive fluorescent molecules as well as the accumulation of the probe inside cell compartments as the mitochondria or the ER. The finding that AM dyes could be in part trapped into organelles has been actually exploited to address them to specific subcellular localizations. Blum and coworkers have, for instance, targeted AM-linked Ca^{2+} indicators to the ER using the targeted-esterase induced dye loading (TED) technique. The technique requires the expression of recombinant esterases into the ER and the loading of the cells with a Ca^{2+} indicator in the AM form: upon cleavage, the probe remains in the lumen of the organelle. In some cases, the chemical properties of the Ca^{2+} -binding molecule allow for its accumulation into specific cellular compartments. For example, the AM forms of X-Rhod 1 and Rhod 2, two members of the Rhod family of Ca^{2+} indicators, are positively charged and thus tend to be sequestered inside mitochondria. Furthermore, their low affinity (K_d of about 700 and 570 nM, respectively) makes them ideal for the measurement of $[\text{Ca}^{2+}]$ changes inside the mitochondrial matrix.

Genetically Encoded Ca^{2+} Indicators

Genetically encoded Ca^{2+} indicators (Table 2) can be easily introduced into cells and can be selectively targeted to specific subcellular compartments. They can be divided into two main families: bioluminescent and green fluorescent protein (GFP)-based probes. Both groups are based on proteins derived from biological organisms: for example, Obelin was found in the hydrozoa *Obelia geniculata*, while aequorin and the GFP have been extracted from the jellyfish *Aequorea victoria*. Molecular biology techniques have made possible to clone the coding sequence of these proteins, to modify their properties by introducing point mutations, and to overexpress them in model cells. They have also made it possible to modulate the pattern and the timing of their expression by tissue-specific or inducible promoters and to target them to specific organelles by coupling them to appropriate lead sequences.

Bioluminescent Probes

Bioluminescent probes are based on the production of light by an intramolecular reaction. They do not suffer the photobleaching problems that affect chemical probes (and also the

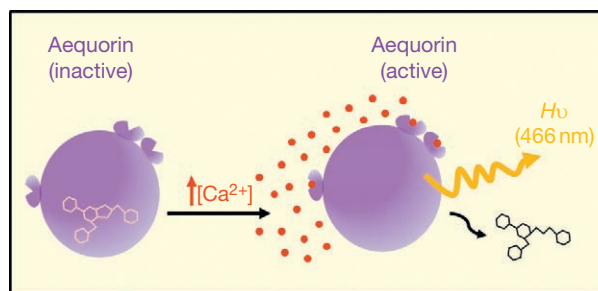


Figure 2 Schematic representation of photon emission from aequorin upon Ca^{2+} binding. The chemical structure depicted within the Aequorin molecule is the prosthetic group coelenterazine. The receptor-like structures on the surface of Aequorin represent the EF hand motifs that bind calcium.

GFP-based Ca^{2+} probes, see below). Obelin and aequorin, the most commonly used bioluminescent proteins, consist of a single polypeptide chain that hosts three EF hands Ca^{2+} -binding domains. They are apoproteins that need binding of a cofactor, coelenterazine, to become active. Following interaction with Ca^{2+} , conformational changes in the protein trigger the peroxidation of coelenterazine, which is oxidized to coelenteramide and is released with the concomitant emission of a photon (Figure 2). This reaction inactivates the photoprotein irreversibly. The luminescent signal supplies the raw data from which the $[\text{Ca}^{2+}]$ levels are then estimated with an algorithm that considers the exponential correlation between $[\text{Ca}^{2+}]$ and the amount of emitted light. Obelin and aequorin differ in their emission wavelength (466 nm for aequorin; 495 for obelin), in their sensitivity for Ca^{2+} , which is lower in the former, and in the speed in the emission of light, which is faster in obelin (3 ms vs. 10 ms). These properties make obelin more suitable for studies that require high temporal resolution, while aequorin is more adequate for $[\text{Ca}^{2+}]$ measurements in 0.5–10 μM Ca^{2+} concentration range. Mutated forms of aequorin have been created to study Ca^{2+} variations in the 10–100 μM range. Furthermore, the use of a coelenterazine analog (coelenterazine n) makes it possible to monitor Ca^{2+} concentrations up to a millimolar. By fusing its complementary DNA (cDNA) with specific intracellular targeting signals, it has been targeted to the nucleus, the ER, the Golgi apparatus, the plasma membrane, the peroxisomes, and even to different positions inside mitochondria, for example, the matrix and the intermembrane space.

In general, since each bioluminescent Ca^{2+} -binding protein only emits one photon when it binds Ca^{2+} , the emission intensity is quite low, ruling out single-cell measurements and only allowing cell-population studies.

GFP-Based Ca^{2+} Probes

Soon after the discovery of the GFP, its potential to detect intracellular events became obvious. GFP is composed of 238 amino acids organized in an 11-stranded beta barrel structure. The chromophore of the protein is not an independent prosthetic group, but it is formed by amino acids of the polypeptide chain itself. These residues were identified as Serine 65,

Tyrosine 66, and Glycine 67, arranged in the center of the beta barrel structure. The fluorophore emits green fluorescence when challenged with blue light. The introduction of point mutations in the fluorophore moiety of GFP produces changes in the spectral properties of the fluorophore and has led to the production of numerous variants of the fluorescent protein. For example, the Y66H and Y66W replacement lead to the generation of a blue fluorescent protein (BFP) and ciano fluorescent protein (CFP), respectively. The T203Y replacement caused a shift of the excitation and emission spectra to higher wavelengths: the protein carrying this mutation is the yellow fluorescent protein (YFP). YFP is peculiar since it is more sensitive to photobleaching, pH, and chloride ions than GFP. Some commonly used analogs of YFP are Citrine and Venus. These variants have a lower pH and chloride sensitivity and faster maturation rates.

The fluorescent variants of GFP have set the basis for the development of other Ca^{2+} indicators, among which the most commonly used are the Camgaroos, the Pericams, and the Cameleons.

The Ca^{2+} indicators known as 'Camgaroos' comprise Camgaroo-1 and -2. They are based on a single fluorescent protein: YFP. Camgaroos were generated by exploiting the finding that the introduction of peptides at the level of amino acid 145 of GFP does not interfere with its ability to fluoresce. Thus, YFP was split into an N-terminal moiety, comprising amino acids 1–144, and a C-terminal domain, from amino acid 145 to the end of the polypeptide chain. These two portions of the YFP molecule were then fused to the N and C terminus of the Ca^{2+} -binding protein Calmodulin, respectively. Ca^{2+} binding to the grafted Calmodulin induces the deprotonation of the chromophore, leading to the increase in the amplitude of both excitation and emission spectra but without shifts in peak wavelength.

Another group of Ca^{2+} sensors, ratiometric Pericams, instead take advantage of another type of modification in the YFP structure, that is, circular permutation. The latter involves the interchange of the N and C terminus of the protein: in the case of YFP, the circularly permuted fluorophore is not only able to fluoresce as in the original protein, but it also shows efficient maturation and higher stability in acidic pH. In Pericams, Calmodulin is fused in tandem to the C terminus of the circularly permuted YFP (cpYFP), while a peptide derived from the Calmodulin-binding domain of the skeletal muscle myosin light chain kinase (known as M13) is present ahead of the N terminus of the cpYFP. When $[\text{Ca}^{2+}]$ levels rise, the M13 peptide interacts with Calmodulin, causing a conformational change of the whole fusion protein that induces variations in the fluorescence of cpYFP.

The most popular members of the Pericam family are ratiometric Pericams and inverse Pericam. Ratiometric Pericam has a bimodal excitation spectrum (415 and 494 nm), and the relative intensities of the emission light change in the transition from the Ca^{2+} -bound to the Ca^{2+} -free forms, allowing for ratiometric measurement. Inverse Pericam instead is a nonratiometric Ca^{2+} probe that decreases the fluorescence intensity on Ca^{2+} binding. One important problem linked to the use of Pericams in living cells is their pH sensitivity. Another is the possibility that endogenous Calmodulin partners could recognize the exogenous Calmodulin moiety of the Ca^{2+} probe and interact with it.

The third class of GFP-based Ca^{2+} indicators, yellow Cameleons (YC), includes two different GFP variants (typically CFP and YFP) linked by Calmodulin and M13. The ability of these ratiometric probes to measure Ca^{2+} levels relies on a fluorescence resonance energy transfer (FRET), which defines the mechanism of nonradiative energy transfer between two fluorophores. In particular, the excited donor chromophore (which is CFP in the Cameleon probe) transfers energy to the acceptor chromophore (YFP). In the case of Cameleons, Ca^{2+} -binding causes Calmodulin and M13 to wrap around each other, bringing CFP and YFP closer and allowing FRET to occur. In these conditions, the fluorescence intensity emission of CFP decreases and the emission of YFP increases: the ratio between the two is proportional to $[\text{Ca}^{2+}]$ levels. Also with Cameleons, the main problem is their interaction with endogenous Calmodulin/M13 partners, that could cause artifacts in the evaluation of intracellular Ca^{2+} changes.

Heim and coworkers have developed a particular class of FRET-based Ca^{2+} probes, in which the molecular switch, instead of Calmodulin and M13, is Troponin c. This muscle protein is predicted to have no interacting partners in other cell types. Among the Troponin c-based Ca^{2+} indicators, Tn-XXL seems particularly promising since it has been successfully expressed both *in vivo* and *in vitro*.

To solve the problem of interactions between endogenous proteins and Ca^{2+} probes, a new generation of Cameleons has been created, in which the interaction domains of M13 and calmodulin have been point mutated. The amino acid substitutions have led to the development of novel variants of the probes with different Ca^{2+} affinities. To increase the brightness of these new Ca^{2+} probes, YFP has been replaced by its circularly permuted analog, Venus, providing an increased dynamic range. By means of targeting sequences, Cameleons have been addressed to various organelles, such as the ER, mitochondria, peroxisomes, nucleus, as well as to the space beneath the plasma membrane.

One last advance in the development of GFP-based Ca^{2+} probes has been recently reported by J J Yang. The addition of negatively charged residues, such as those that typically compose Ca^{2+} -binding sites, to the GFP chromophore affects both its maturation and fluorescence properties. J J Yang and coworkers have introduced Ca^{2+} -binding motifs into the structure of GFP, laying the foundations for the development of a novel class of GFP-based Ca^{2+} indicators.

Conclusions

Ca^{2+} indicators have come a long way from the development of the first Ca^{2+} probes. The availability of genetic and molecular biology techniques coupled to cell biology strategies to deliver recombinant probes inside cell compartments have greatly advanced the study of Ca^{2+} signaling. While the first compounds had to be microinjected into cells, loading of the cells can now be easily obtained by transfection or viral infection. The GFP-based probes have allowed the analysis of the contribution of single intracellular compartments to the overall cellular Ca^{2+} homeostasis process. A caveat is, however, necessary: Ca^{2+} measurement experiments must be carefully

designed in order to choose the more appropriate indicator. Each Ca^{2+} probe has both advantages and disadvantages and the ideal Ca^{2+} indicator for living cells or tissue slices has not been developed yet. The necessity to monitor $[\text{Ca}^{2+}]$ variations in whole organs or tissue will require the development of novel sensors, suitable for the measurements of $[\text{Ca}^{2+}]$ in living organisms.

See also: **Bioenergetics: Ryanodine Receptor Calcium Ion Channels.**

Further Reading

- Baird GS, Zacharias DA, and Tsien RY (1999) Circular permutation and receptor insertion within green fluorescent proteins. *Proceedings of the National Academy of Sciences of the United States of America* 96(20): 11241–11246.
- Carafoli E (2002) Calcium signaling: A tale for all seasons. *Proceedings of the National Academy of Sciences of the United States of America* 3: 1115–1122.
- Demaurex N (2005) Calcium measurements in organelles with Ca^{2+} -sensitive fluorescent proteins. *Cell Calcium* 38(3–4): 213–222.
- Heim N, Garaschuk O, Friedrich MW, et al. (2007) Improved calcium imaging in transgenic mice expressing a troponin C-based biosensor. *Nature Methods* 4(2): 127–129.
- Holder AN, Ellis AL, Zou J, Chen N, and Yang JJ (2009) Facilitating chromophore formation of engineered Ca^{2+} -binding green fluorescent proteins. *Archives of Biochemistry and Biophysics* 486(1): 27–34.
- Jaepel J and Blum R (2011) Capturing ER calcium dynamics. *European Journal of Cell Biology* 90(8): 613–619.
- Magalhães PJ and Rizzuto R (2001) Mitochondria and calcium homeostasis: A tale of three luminescent proteins. *Luminescence* 16(2): 67–71.
- McCombs JE and Palmer AE (2008) Measuring calcium dynamics in living cells with genetically encodable calcium indicators. *Methods* 46(3): 152–159.
- Miyawaki A, Llopis J, Heim R, et al. (1997) Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature* 388: 882–887.
- Miyawaki A, Nagai T, and Shimosono S (2002) Circularly permuted green fluorescent proteins engineered to sense Ca^{2+} and their application to imaging of subcellular Ca^{2+} dynamics. *RIKEN Review* No. 49.
- Ogawa Y, Harafuji H, and Kurebayashi N (1980) Comparison of the characteristics of four metallochromic dyes as potential calcium indicators for biological experiments. *Journal of Biochemistry* 87(5): 1293–1303.
- Palmer AE, Giacomello M, Kortemme T, et al. (2006) Ca^{2+} indicators based on computationally redesigned calmodulin–peptide pairs. *Chemistry and Biology* 13(5): 521–530.
- Rudolf R, Mongillo M, Rizzuto R, and Pozzan T (2003) Looking forward to seeing calcium. *Nature Reviews Molecular Cell Biology* 4(7): 579–586.
- Tsien RY (1980) New calcium indicators and buffers with high selectivity against magnesium and protons: Design, synthesis, and properties of prototype structures. *Biochemistry* 19: 2396–2404.

Inositol Phosphate Kinases and Phosphatases

S B Shears, NIEHS, Research Triangle Park, NC, USA

Published by Elsevier Inc.

Glossary

Diphosphoinositol polyphosphate A type of inositol phosphate that has either one or two diphosphate (pyrophosphate) groups.

Inositol phosphate A compound with a six-member carbon ring structure to which single phosphates (and sometimes diphosphates) are attached at various positions.

Inositol phosphate kinase An enzyme that transfers a phosphate group from adenosine triphosphate to the inositol ring.

Inositol phosphate phosphatase An enzyme that removes a phosphate group from the inositol ring.

Inositol phosphate phosphotransferase An enzyme that transfers a phosphate from one inositol phosphate to another – a specific property of ITPK1.

Intracellular signal/messenger One member of what is usually a functionally coupled series of molecules or ions that together comprise a chemical pathway which conveys and amplifies a cell's biological response to a specific extracellular stimulus, such as a hormone or neurotransmitter.

Phospholipase C A specialized enzyme that releases Ins(1,4,5)P₃ from its membrane tether.

Phytic acid The nonsystematic but widely used alternate name for InsP₆; the molecule formed by filling all six carbons of the inositol ring with single phosphate groups.

Turtle A marine reptile that provides a visual mnemonic for easy recall of the nomenclature for numbering the carbon atoms that comprise the inositol ring.

Introduction

Information transfer in biological systems – a process frequently termed signal transduction – requires sensor proteins to detect and respond to regulatory input from an effector, which can be either another protein or a small molecule. The ability of a sensor to perceive an effector is frequently dependent upon a recognition pattern generated by the presence or absence of phosphate groups. When one or more of these phosphate groups are added (phosphorylation) or removed (dephosphorylation) from the effector, such a process acts as a molecular switch that determines whether the regulator is 'on' or 'off'. Inositol (Figure 1) is an example of a small molecule that is functionally modified in this manner. Each of the six hydroxyl groups that are attached to the carbons of the *myo*-inositol ring (Figure 1(a)) are sites for phosphorylation. Enzymes that add phosphate groups to the inositol moiety are known as inositol phosphate kinases. A separate family of enzymes, the inositol phosphate phosphatases, removes phosphate groups from the inositol ring. Phosphate-transfer reactions catalyzed by these phosphatases and kinases are networked in serial and parallel configurations, permitting the cell to add and subtract phosphate groups in a combinatorial manner, thus generating an array of inositol phosphate molecules. Recognition patterns, which arise from specific topographical arrangements of phosphate groups around the inositol ring, endow certain inositol phosphates with specific, biologically important functions, often in the context of signal transduction. Thus, inositol phosphate kinases and phosphatases control the synthesis and metabolism of important effector molecules.

Inositol Phosphate Nomenclature

To appreciate the nature of phosphate recognition patterns around the inositol ring, it is necessary to understand the nomenclature that distinguishes the different carbons that comprise the inositol moiety. Fortunately, Agranoff's turtle (Figure 1) provides us with a timeless, visual mnemonic. In this *aide mémoire*, the hydrogens are conveniently ignored and the inositol ring is depicted in its thermodynamically stable so-called chair conformation (Figure 1(b)). The structure's resemblance to a turtle is then apparent (Figure 1(c)). As a consequence of deliberations by international nomenclature committees, the turtle's appendages are numbered in an anti-clockwise direction (when viewed from above), commencing with the front right flipper. The 2-hydroxyl (the turtle's head) is axial to the plane of the ring, whereas the other five hydroxyls (the four flippers and the tail) are equatorial (i.e., approximately in the same plane as the ring). This numbering system allows us to readily and unambiguously specify the location of phosphate groups that are added to the inositol ring. Ins(1,4,5)P₃, for example (see Figure 2), is an abbreviation that describes the addition of three single phosphate groups (P₃) to the inositol (Ins) ring, at positions 1, 4, and 5. Ins(1,2,3,4,5,6)P₆ describes the molecule in which all six of the carbons are attached to single phosphate groups, but since this feature makes the numbering superfluous, InsP₆ is the acceptable definition (Figure 2). This particular inositol phosphate is also widely known as phytic acid. Some inositol phosphates even contain diphosphate (i.e., PP) groups (Figure 2), the so-called 'inositol pyrophosphates' or, more correctly, diphosphoinositol polyphosphates.

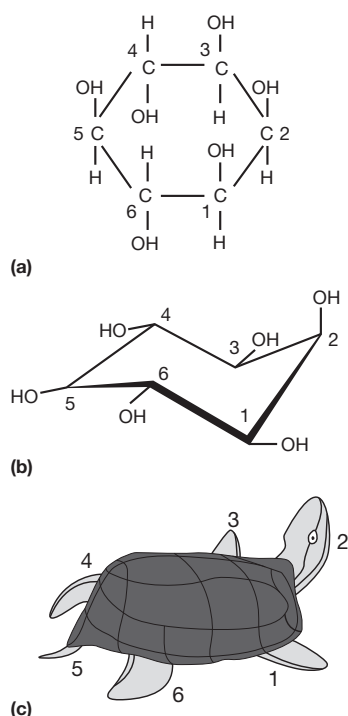


Figure 1 Inositol ring nomenclature. (a) Complete chemical structure of inositol; (b) naturally occurring conformation of inositol, *sans* hydrogens; (c) the numbering of the turtle's appendages in an anti-clockwise direction (beginning with its right flipper) provides an *aide m  moire* for inositol phosphate nomenclature.

Enzyme Nomenclature and Metabolic Interrelationships

Figure 2 provides a metabolic map showing the participation of those inositol phosphate kinases and phosphatases that are most commonly found across the phylogenetic spectrum. Unfortunately, these enzymes are not always named either in a manner that is both universally accepted or by a system that unambiguously describes their function(s). For example, the names IPK1 and IPK2 do not designate the positional specificities of these kinases, but instead merely reflect the chronological order in which a particular team of researchers identified these kinases. There is another enzyme that removes the 5-phosphate from both $\text{Ins}(1,4,5)\text{P}_3$ and $\text{Ins}(1,3,4,5)\text{P}_4$. However, inositol polyphosphate 5-phosphatase is a popular terminology for this enzyme, despite the fact that this exaggerates its promiscuity, by erroneously implying that the enzyme removes the 5-phosphate from other inositol phosphates. The preferred, unambiguous nomenclature for this enzyme is $\text{Ins}(1,4,5)\text{P}_3/\text{Ins}(1,3,4,5)\text{P}_4$ 5-phosphatase (**Figure 2**). Nevertheless, such a systematic approach does become unwieldy when a single enzyme metabolizes three or more substrates, especially when the positional specificity of the enzyme varies with the nature of the substrate. For example, there is a single kinase that adds a 3-phosphate to $\text{Ins}(1,4,5)\text{P}_3$, a 6-phosphate to $\text{Ins}(1,3,4,5)\text{P}_4$, a 5-phosphate to $\text{Ins}(1,3,4,6)\text{P}_4$, and a

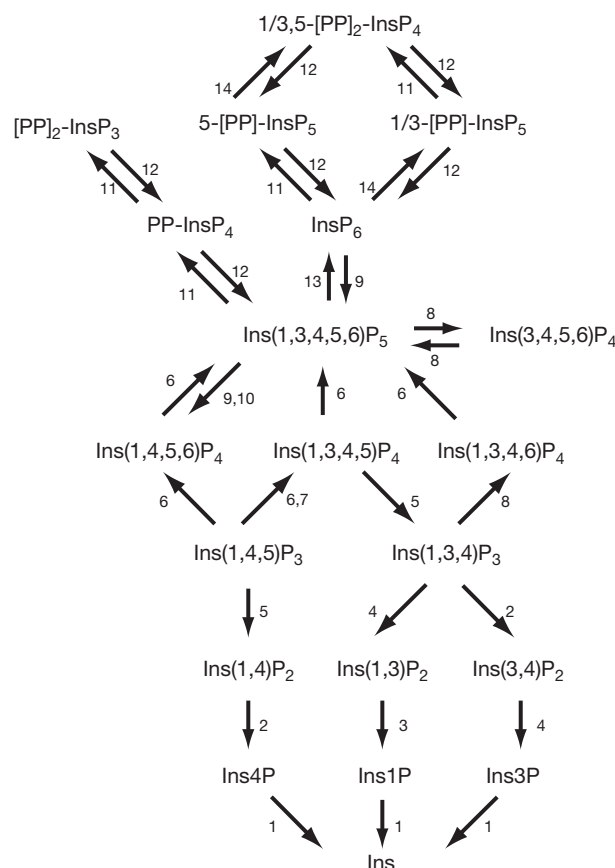


Figure 2 Metabolic interrelationships of inositol phosphate kinases and phosphatases. The figure provides a metabolic map showing the participation of some important inositol phosphate kinases and phosphatases that are widely distributed across the phylogenetic spectrum. The numbers associated with each arrow represent the following enzymes: 1, inositol monophosphatase; 2, $\text{Ins}(1,4)\text{P}_2/\text{Ins}(1,3,4)\text{P}_3$ 1-phosphatase; 3, $\text{Ins}(1,3)\text{P}_2$ 3-phosphatase; 4, $\text{Ins}(3,4)\text{P}_2/\text{Ins}(1,3,4)\text{P}_3$ 4-phosphatase; 5, $\text{Ins}(1,4,5)\text{P}_3/\text{Ins}(1,3,4,5)\text{P}_4$ 5-phosphatase; and 6, IPMK, for inositol polyphosphate multikinase. Also known as IPK2; 7, $\text{Ins}(1,4,5)\text{P}_3$ 3-kinase; 8, ITPK1; 9, MIPP, for multiple inositol polyphosphate phosphatase; 10, PTEN (for phosphatase and tensin homolog, deleted on chromosome ten); 11, IP6K, for inositol hexakisphosphate kinase; 12, DIPP, for diphosphoinositol polyphosphate phosphohydrolase; and 13, IP5K, for $\text{Ins}(1,3,4,5,6)\text{P}_5$ 2-kinase. Also known as IPK1; 14, PPIP5K (for diphosphoinositol pentakisphosphate kinase), and sometimes described as VIP.

3-phosphate to $\text{Ins}(1,4,5,6)\text{P}_4$. For brevity, this protein is often known as the 'inositol polyphosphate multikinase' (**Figure 2**). Equally, the phosphatase that specifically removes the 3-phosphate from $\text{Ins}(1,3,4,5,6)\text{P}_5$ and the 6-phosphate from $\text{Ins}(1,4,5,6)\text{P}_4$ is also capable of removing any of the phosphate groups from InsP_6 ; this enzyme is known as the 'multiple inositol polyphosphate phosphatase'. ITPK1 is another enzyme that deserves special mention. It was originally named from its ability to phosphorylate $\text{Ins}(1,3,4)\text{P}_3$. The name has remained unaltered even as our insight into the enzyme's full catalytic repertoire has considerably expanded.

Now we know it also phosphorylates $\text{Ins}(3,4,5,6)\text{P}_4$ and, perhaps most remarkably of all, ITPK1 also acts as a phosphotransferase, exchanging phosphate groups between $\text{Ins}(1,3,4,5,6)\text{P}_5$ and $\text{Ins}(1,3,4,6)\text{P}_4$. Finally, the name of one of this family of enzymes – PTEN (phosphatase and tensin homolog deleted on chromosome 10) – does not specifically address its ability to metabolize an inositol phosphate. That nomenclature reflects some other characteristics of this enzyme which were established before its $\text{Ins}(1,3,4,5,6)\text{P}_5$ 3-phosphatase activity was clarified (see [Figure 2](#)).

Although the inositol phosphates are generally considered to be freely diffusible molecules, some of the enzymes that metabolize them can often be found to be concentrated in specific regions of the cell. This creates a cell-signaling paradox. Why segregate one component of a cell-signaling apparatus, in order to generate a second messenger that then freely diffuses throughout the entire cell? Providing an answer to this question is one of the challenges for future work in this field. In part, enzyme compartmentalization may reflect some additional, noncatalytic role, for example, as a molecular scaffold around which higher-order multimeric complexes can be constructed.

Regulatory Input into the Inositol Phosphate Pathway

There are two separate metabolic routes that feed into the pathway shown in [Figure 2](#): the first is by enzymatic conversion of glucose-6-phosphate into $\text{Ins}3\text{P}$ and the second is synthesis of $\text{Ins}(1,4,5)\text{P}_3$. The latter is initially attached through its 1-phosphate (i.e., its front right flipper, [Figure 1\(c\)](#)) to the lipid material that represents the bulk phase of cellular membranes. This attachment is severed by a specialized enzyme – a phospholipase C – that is stimulated when a hormone or neurotransmitter binds to appropriate receptors on the cell surface. The $\text{Ins}(1,4,5)\text{P}_3$ released by phospholipase C has

the role of liberating cellular stores of calcium, which, in turn, regulates many calcium-dependent cellular processes. The efficacy of $\text{Ins}(1,4,5)\text{P}_3$ depends upon its intracellular concentration, which, in turn, is dictated by a dynamic balance between competing rates of $\text{Ins}(1,4,5)\text{P}_3$ synthesis (by phospholipase C) and metabolism (by inositol phosphate kinases and phosphatases; [Figure 2](#)). Hormones and neurotransmitters also use cell-surface receptors to activate the synthesis of many other intracellular chemical messengers, and these can lead to changes in the activities of several inositol phosphate kinases and phosphatases. The activities of some of these enzymes can also respond to stimuli that are generated from within cells, such as those that arise following a decline in bioenergetic health. Through these various regulatory processes, the cell can carefully regulate metabolic flux through specific areas of the inositol phosphate pathway, thereby manipulating levels of inositol phosphates which are themselves intracellular signals (see section [Further Reading](#)).

Further Reading

- Alcazar-Roman AR and Wente SR (2008) Inositol polyphosphates: A new frontier for regulating gene expression. *Chromosoma* 117: 1–13.
- Berggren PO and Barker CJ (2008) A key role for phosphorylated inositol compounds in pancreatic beta-cell stimulus–secretion coupling. *Advances in Enzyme Regulation* 48: 276–294.
- Berridge MJ (2009) Inositol trisphosphate and calcium signaling mechanisms. *Biochimica et Biophysica Acta* 1793: 933–940.
- Irvine RF (2003) 20 years of $\text{Ins}(1,4,5)\text{P}_3$ and 40 years before. *Nature Reviews Molecular Cell Biology* 4: 586–590.
- Michell RH (2008) Inositol derivatives: Evolution and functions. *Nature Reviews Molecular Cell Biology* 9: 151–161.
- Raboy V and Bowden D (2006) Genetics of inositol polyphosphates. *Subcellular Biochemistry* 39: 71–101.
- Shears SB (2009) Diphosphoinositol polyphosphates: Metabolic messengers? *Molecular Pharmacology* 76: 236–252.
- Shears SB (2009) Molecular basis for the integration of inositol phosphate signaling pathways via human ITPK1. *Advances in Enzymology* 49: 87–96.

Inositol Trisphosphate and Calcium Signaling

M J Berridge, The Babraham Institute, Cambridge, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

2-Aminoethoxydiphenyl borate (APB) A pharmacological agent used to inhibit the release of Ca^{2+} by the InsP_3 receptors.

Ca^{2+} -induced Ca^{2+} release (CICR) A regenerative mechanism whereby Ca^{2+} triggers release of further Ca^{2+} .

Endoplasmic reticulum (ER) Internal membrane sacs that store Ca^{2+} .

G-protein-coupled receptors (GPCRs) Cell-surface receptors that responds to external stimuli such as hormones and neurotransmitters.

Inositol 1,4,5-trisphosphate (InsP_3) A Ca^{2+} -mobilizing second messenger.

Phospholipase C (PLC) An enzyme that hydrolyzes a membrane phospholipid to release the Ca^{2+} -mobilizing second-messenger inositol 1,4,5-trisphosphate (InsP_3).

Ryanodine receptor (RyR) Calcium release channels locate in the ER.

Cellular Ca^{2+} signaling systems are characterized by their variety and specificity. Each cell type has a unique system specifically designed to control its particular function. All of these Ca^{2+} control systems are variations on a basic theme where the level of Ca^{2+} is low in resting cells and is rapidly elevated in response to external stimuli. Much of the specificity of Ca^{2+} signaling depends on how this surge in Ca^{2+} is organized in both time and space. For example, a brief microsecond pulse of Ca^{2+} restricted to the synaptic endings of neurons triggers the release of neurotransmitters. Ca^{2+} sparks, located specifically at the junctional zones in cardiac cells, are fired synchronously in less than a second to produce the localized pulses of Ca^{2+} to drive cardiac contraction. At the other end of the temporal scale, much slower global Ca^{2+} signals, often spreading as regenerative Ca^{2+} waves, function over many seconds to minutes to control the cell cycle as occurs during mammalian fertilization. This great variety and versatility of the Ca^{2+} signaling systems depends on an extensive Ca^{2+} signaling toolkit. During cell differentiation, different components of this toolkit are selected to create specific signaling systems. One of the basic modules found in many different cell types is the inositol 1,4,5-trisphosphate (InsP_3)/ Ca^{2+} signaling system.

InsP_3 Formation

The main mechanisms responsible for forming the Ca^{2+} -mobilizing second messenger InsP_3 are shown in **Figure 1**. External stimuli (e.g., hormones, neurotransmitters, and growth factors) gain access to this signaling pathway by activating cell-surface receptors. The latter fall into two main classes: the G-protein-coupled receptors (GPCRs) and the protein-tyrosine kinase-linked receptors (PTKRs). During the transduction process, these receptors activate phospholipase C (PLC) that hydrolyses the precursor lipid phosphatidylinositol 4,5-diphosphate ($\text{PtdIns}4,5\text{P}_2$) to generate InsP_3 , which then diffuses into the cytosol to activate the InsP_3 receptors to release Ca^{2+} stored in the endoplasmic reticulum (ER).

There is a family of PLCs that can be distinguished by the way they are coupled to cell-surface receptors. In general, the

GPCRs use the PLC β isoforms, whereas the PTKRs are coupled to the PLC γ isoforms. Various other input signals are responsible for activating some of the other PLC isoforms. For example, PLC δ 1 appears to be activated by an elevation in cytosolic Ca^{2+} , and PLC ϵ is activated by GTP-binding proteins such as Ras and Rap1A. The second messenger cyclic adenosine monophosphate (cAMP) acts through the exchange protein activated by cAMP (EPAC) protein to stimulate Rap1 to activate PLC ϵ . The PLC ζ isoform, which is expressed in mammalian sperm, is injected into eggs following sperm-egg fusion to generate the InsP_3 that activates oocytes during fertilization.

InsP_3 Metabolism

The InsP_3 that is formed during signaling is removed by various inositol phosphate metabolic pathways from which it emerges either as free inositol or as various inositol polyphosphates, such as InsP_4 , InsP_6 , InsP_7 , and InsP_8 (**Figure 2**). The InsP_3 that enters these various pathways is metabolized initially by two different enzymes. It is dephosphorylated by type I inositol polyphosphate 5-phosphatase to form $\text{Ins}1,4\text{P}_2$. This InsP_3 phosphatase is directed to the membrane by isoprenylation of the C-terminal region. If these membrane-attachment sites are removed, allowing the enzyme to disperse into the cytosol, InsP_3 metabolism appears to be severely curtailed. It seems that much of the InsP_3 is metabolized close to its site of action at the plasma membrane. The $\text{Ins}1,4\text{P}_2$ is then dephosphorylated further to produce free inositol that is used to resynthesize PtdIns , which is the lipid precursor used to generate the $\text{PtdIns}4,5\text{P}_2$ – the substrate for PLC to generate InsP_3 . Cells have access to three separate sources of inositol: recycling the second messenger InsP_3 (as just described), *de novo* synthesis from glucose 6-phosphate, or uptake of dietary inositol circulating in the plasma. Drugs, such as lithium and valproate, may control manic-depressive illness by reducing the supply of inositol by inhibiting inositol monophosphatase and inositol synthase, respectively. The supply of free inositol is one of the essential precursors for the synthesis of PtdIns , which is transferred from its site of synthesis on the ER to

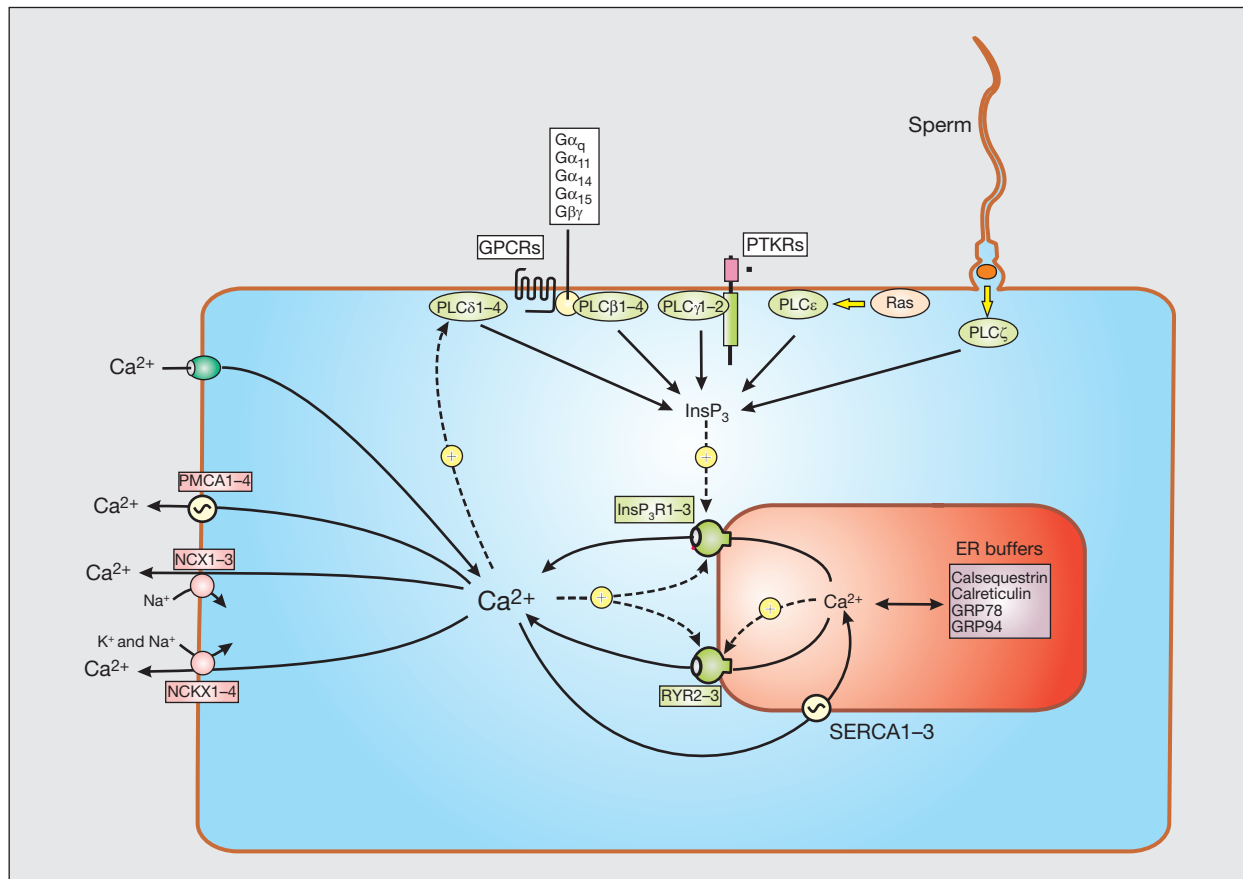


Figure 1 Inositol 1,4,5-trisphosphate (InsP₃) and Ca²⁺ signaling. G-protein-coupled receptors (GPCRs) and the protein tyrosine kinase-linked receptors (PTKRs) activate different phospholipase C (PLC) isoforms. InsP₃ activates the InsP₃ receptor (InsP₃R) to release Ca²⁺ from the endoplasmic reticulum (ER). During the recovery phase, Ca²⁺ is pumped out of the cell or is returned to the ER.

the plasma membrane by PtdIns transfer protein (PITP) (Figure 2).

Second, InsP₃ is phosphorylated by Ins1,4,5P₃ 3-kinase to form Ins1,3,4,5P₄. There are three isoforms of this enzyme (InsP₃ 3-kinase A, B, and C). Both types A and B are activated by Ca²⁺/calmodulin, with the type B being the most sensitive. Ins1,3,4,5P₄ can either be dephosphorylated to inositol or can be phosphorylated further to generate an assortment of inositol phosphates, such as IP₆, IP₇, and IP₈, that may have additional signaling functions.

InsP₃ and Ca²⁺ signaling

The primary function of InsP₃ is to function as a second messenger to release Ca²⁺ from the internal stores. Release from internal stores represents a major source of signal Ca²⁺ for many cells. The major Ca²⁺ store, which is located in the extensive ER network that extends throughout the cell, contains both the ryanodine receptor (RyR) and inositol 1,4,5-trisphosphate receptor (InsP₃R) families (Figure 1). A critical feature of these channels is that they are activated by Ca²⁺. In the case of InsP₃R, therefore, release is triggered by both InsP₃ and Ca²⁺. In effect, InsP₃R functions as a coincident detector, and this property may account for the role of the

InsP₃/Ca²⁺-signaling system in learning and memory. The process of Ca²⁺-induced Ca²⁺ release (CICR) displayed by the RyRs and InsP₃Rs plays a central role in the way many Ca²⁺ signals are generated. This positive feedback mechanism has two important functions in cells. It enables Ca²⁺ entering across the plasma membrane to function as a messenger to release Ca²⁺ from the internal store. This function of CICR was first described in cardiac cells, where the L-type channel Ca²⁺ channel provides an influx of trigger Ca²⁺ that then diffuses into the cell to activate RyR2. The other important function of CICR is to link together these intracellular channels so that the Ca²⁺ being released from one channel diffuses across to neighboring channels that are excited to release further Ca²⁺, thereby setting up regenerative waves. This mechanism is particularly important with regard to the way in which InsP₃ acts to release internal Ca²⁺ (Figure 3).

Spatial and Temporal Aspects of InsP₃/Ca²⁺ Signaling

Those cells that use this InsP₃/Ca²⁺-signaling system usually respond to agonists by generating Ca²⁺ signals with characteristic spatial and temporal patterns (Figure 3). Under resting conditions, the InsP₃Rs are inactive and the Ca²⁺ concentration is held at a steady low level of approximately 0.1 μM.

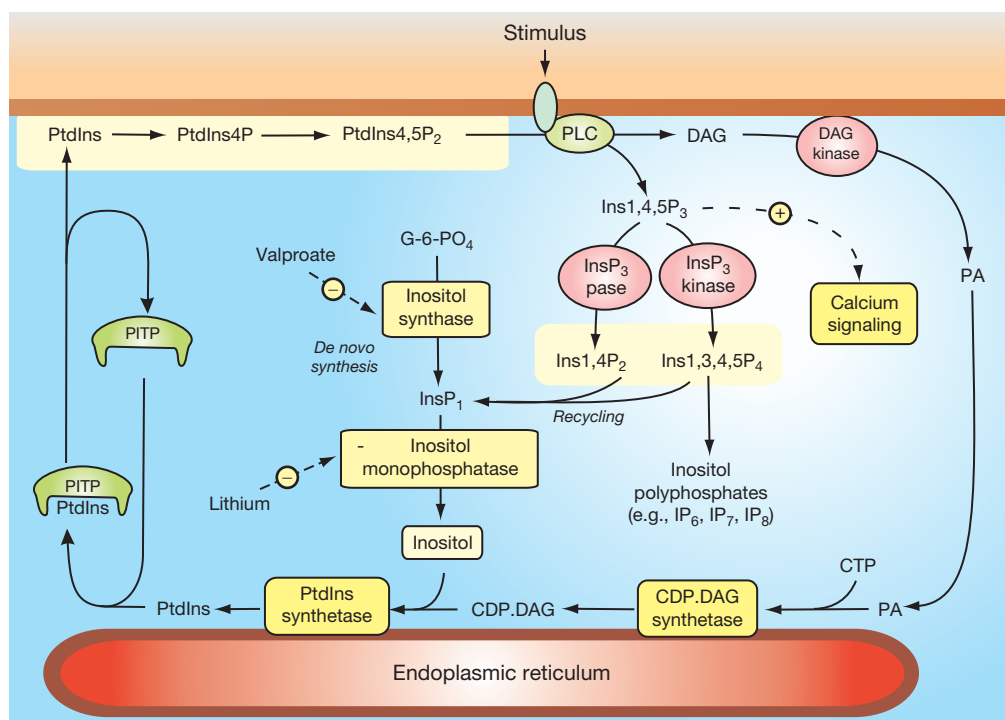


Figure 2 InsP_3 metabolism. The inositol lipids that function in signaling are embedded in the inner leaflet of the plasma membrane. The precursor lipid is phosphatidylinositol (PtdIns), which is synthesized from inositol and CDP.DAG by a PtdIns synthetase located on the endoplasmic reticulum. PtdIns is transported to the plasma membrane by a PtdIns transfer protein (PITP). The PtdIns in the plasma membrane is successively phosphorylated, first on the 4-position to form PtdIns4P, and then on the 5-position to form PtdIns4,5P₂. Activated cell-surface receptors stimulate phospholipase C (PLC) to hydrolyze PtdIns4,5P₂ to generate inositol 1,4,5-trisphosphate (InsP_3) and diacylglycerol (DAG). InsP_3 activates Ca^{2+} signaling.

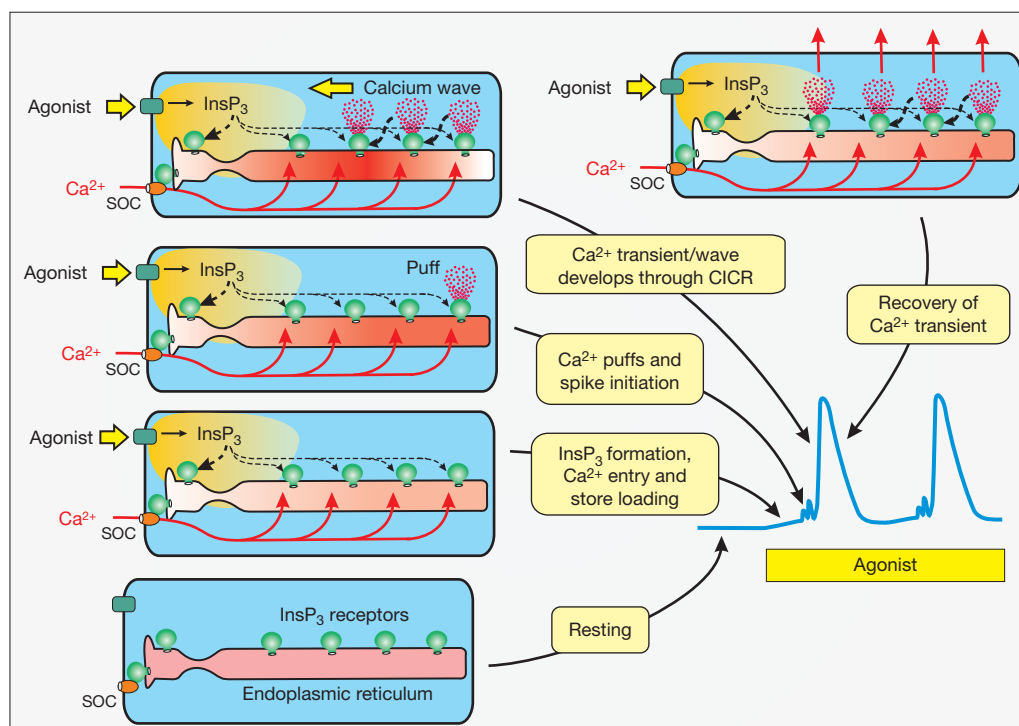


Figure 3 The spatial and temporal aspect of InsP_3 -dependent Ca^{2+} signals (see text for further details).

Upon addition of an agonist capable of activating PLC (as described earlier), there is an increase in the formation of InsP_3 that may have two functions. First, the high level of InsP_3 near the membrane may stimulate InsP_3Rs to release Ca^{2+} from stores near the membrane, and this will promote the entry of Ca^{2+} through store-operated channels (SOCs). The InsP_3Rs located deeper within the cell will be progressively sensitized by the InsP_3 that diffuses into the cell and they will begin to release brief Ca^{2+} bursts. Such Ca^{2+} bursts are the elementary events of Ca^{2+} signaling and are the basic building blocks used to create global Ca^{2+} signals. Since these InsP_3Rs are sensitive to both InsP_3 and Ca^{2+} , these initial puffs will be able to recruit neighboring InsP_3Rs to set up a regenerative Ca^{2+} wave that sweeps through the cell and is responsible for the rapid rising phase of a typical Ca^{2+} transient (Figure 3). When Ca^{2+} is released from an individual channel, it feeds back to inhibit further release, and this negative feedback effect limits the amount of Ca^{2+} that is released during each transient. Once the release has switched off, the recovery phase begins and the level of Ca^{2+} returns to resting levels as it either is extruded from the cell or is pumped back into the store. The sequence of events then repeats because the constant influx of Ca^{2+} will load up the store again, and this sets the stage for the next transient, thereby, setting up a regular Ca^{2+} oscillation.

A characteristic feature of most Ca^{2+} signals is that they are presented as a brief Ca^{2+} transient. These transients can either be produced on demand by periodic stimulation as occurs in muscle or neurons, or they can appear as part of a Ca^{2+} oscillation as described in the previous section. InsP_3 acts in many cells to set up Ca^{2+} oscillations through the periodic release of internal Ca^{2+} through the operation of a cytosolic oscillator as described in the previous section (Figure 3). An interesting feature of these oscillations is that their frequency varies with the strength of the external agonist and such frequency modulation represents a mechanism to encode information. Such agonist-dependent cytosolic Ca^{2+} oscillations are a major feature of $\text{InsP}_3/\text{Ca}^{2+}$ signaling in many cell types such as liver cells, mammalian oocytes, vascular and airway smooth muscle cells, astrocytes, and osteocyte precursor cells during their differentiation to osteoclasts.

Inositol 1,4,5-Trisphosphate Receptors

The second messenger InsP_3 functions to mobilize Ca^{2+} from the ER by acting on InsP_3 receptors (InsP_3Rs) (Figures 1 and 3). There are three different InsP_3Rs ($\text{InsP}_3\text{R1}$, $\text{InsP}_3\text{R2}$, and $\text{InsP}_3\text{R3}$), which have fairly similar primary structures and properties. However, they are sufficiently different for cells to express them in varying proportions and in different cellular locations. With regard to their basic function of releasing Ca^{2+} , all behave similarly. Four subunits come together to form functional channels, and the similarity between the three isoforms is reflected in the fact that they can form both homo- and heterotetramers.

The domain structure of one of the subunits reveals that the receptor is organized into three main regions. The transmembrane (TM) and pore (P) domain embeds the receptor in the ER. There are six TM domains, with the pore loop located between TM5 and TM6 (Figure 4). The luminal loops of the pore are glycosylated. There is a very long N-terminal region that has an InsP_3R -binding domain at the end, which is connected to TM1

by the modulatory domain. This modulatory loop has two important functions. First, it has to transmit the conformational change induced by the binding of InsP_3 and Ca^{2+} down to the TM region to open the pore. Second, it is the site where many of the modulators act to alter the release of Ca^{2+} .

As described earlier, the primary regulators of InsP_3R are InsP_3 and Ca^{2+} . The way in which these two messengers cooperate with each other is complicated, but the basic mechanism is that InsP_3 binds to the InsP_3 -binding domain to induce a conformational change, which sensitizes a Ca^{2+} -binding site, and it is the subsequent binding of Ca^{2+} that then acts to open the channel (Figure 4). The effect of Ca^{2+} on the InsP_3R is bimodal; it is stimulatory at low levels of Ca^{2+} , but then becomes inhibitory at concentrations above approximately 300 nM. This dual control of InsP_3R by both InsP_3 and Ca^{2+} is central to its multiple functions in cell signaling. It is central to the mechanism of Ca^{2+} oscillations described earlier (Figure 3). The fact that the opening of InsP_3R depends upon the presence of both InsP_3R and Ca^{2+} has suggested a possible role as a neuronal coincident detector in neurons during learning and memory.

In addition to InsP_3 and Ca^{2+} , there are a large number of other physiological and pharmacological regulators that are able to modulate the activity of the InsP_3Rs .

InsP_3R Agonists

The fungal product adenophostin is 100-fold more active than InsP_3 . It activates the receptor by binding to the InsP_3 -binding site. The InsP_3R is also sensitive to reactive oxygen species (ROS), such as hydrogen peroxide (H_2O_2), which may act by oxidizing the two highly conserved cysteine residues located in the C-terminal region of the molecule. Similarly, the oxidizing agent thimerosal can also activate InsP_3R . When added to intact cells, thimerosal can reproduce Ca^{2+} signals produced by normal stimuli, such as fertilization of mouse oocytes. Adenosine triphosphate (ATP) binds to a site in the modulatory domain where it potentiates the effect of InsP_3 . This stimulatory effect of ATP could provide a mechanism for shaping Ca^{2+} signals to the metabolic state of the cell.

InsP_3R Antagonists

Xestospingins are potent membrane-permeant blockers of InsP_3Rs . However, they can also inhibit the ER Ca^{2+} pump and this may explain some of their inhibitory actions on InsP_3 -induced Ca^{2+} release. Caffeine is an effective reversible inhibitor of InsP_3 -induced Ca^{2+} release. Caffeine may act by competing with ATP, which is an essential co-activator of InsP_3R . The inhibitory effect can be overcome by increasing the concentration of InsP_3 . 2-Aminoethoxydiphenyl borate (APB) is a membrane-permeant inhibitor of the InsP_3R that has proved to be an effective tool for inhibiting Ca^{2+} release.

InsP_3R Modulation

The activity of InsP_3Rs is modulated by protein phosphorylation driven by a number of signaling pathways. The effects of phosphorylation are complex because it can result in either an increase or a decrease in InsP_3 -induced Ca^{2+} release. In the parotid gland, acetylcholine-evoked secretion is potentiated

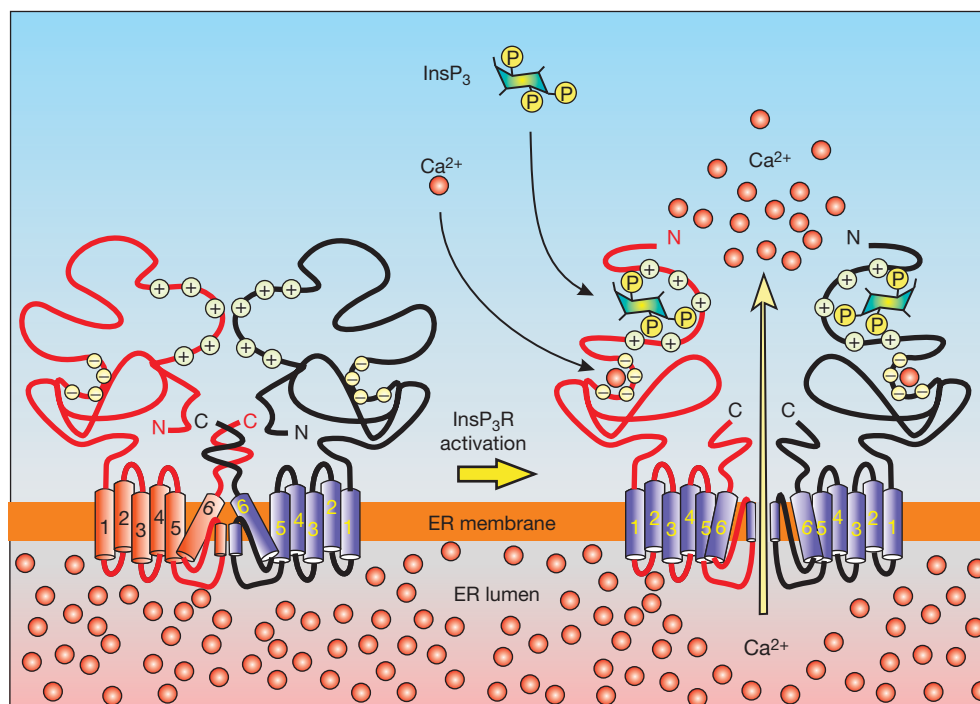


Figure 4 Activation of the inositol 1,4,5-trisphosphate receptor (InsP₃R) by the cooperative action of InsP₃ and Ca²⁺. Two of the channel subunits (colored red and blue) that make up the tetramer are shown with their pore regions coming together to form the channel. In the resting state (shown on the left), the pore may be held in a closed state by the N-terminal region of one receptor interacting with the C-terminal region of a neighboring receptor. Upon binding first InsP₃ and then Ca²⁺, there is a marked conformational change that removes this N-terminal/C-terminal interaction, and the gate region opens to allow Ca²⁺ to flow out of the endoplasmic reticulum (ER) into the cytoplasm.

by agents that elevate cAMP, which acts by enhancing the degree of Ca²⁺ signaling by phosphorylating InsP₃R. Similar effects occur in liver cells, where both cAMP and cyclic guanosine monophosphate (cGMP) enhance the phosphorylation of the InsP₃R1 and increase the release of Ca²⁺. However, there are other cell types where phosphorylation of the receptor results in a decrease in release. For example, phosphorylation of InsP₃R in vascular smooth muscle cells by cGMP-dependent kinase (cGK) causes a rapid inhibition of Ca²⁺ mobilization. This action of cGMP plays a role in smooth muscle relaxation in response to nitric oxide (NO). This action of cGMP is mediated by an InsP₃R-associated cGMP kinase substrate (IRAG), which is phosphorylated by cGK in smooth muscle cells and is an essential component for the regulation of InsP₃-induced Ca²⁺ release.

Phosphorylation of InsP₃R by protein kinase B (PKB) functions to inhibit InsP₃-induced Ca²⁺ release and this could provide a possible mechanism for the inhibition of Ca²⁺-dependent glycogenolysis by insulin in liver cells.

One of the anti-apoptotic actions of Bcl-2 may depend on the reduction of Ca²⁺ release by InsP₃R.

The Role of the InsP₃/Ca²⁺ Signaling Pathways in Cellular Control Mechanisms

A remarkable feature of the InsP₃/Ca²⁺ signaling pathway is that it participates in the control of a large number of cellular processes (Figure 5). In some cases, it plays a direct role in generating Ca²⁺ signals, whereas, in other cases, it functions

to modulate the Ca²⁺ signals produced by other signaling pathways. The following examples will illustrate the versatility of this Ca²⁺ signaling mechanism. As described earlier, the InsP₃-dependent Ca²⁺ signal appears as regular oscillations as occur during fertilization, the control of metabolism in liver cells, contraction in smooth muscle cells, and the function of the interstitial cells of Cajal. In mammalian oocytes, there is an unusual activation mechanism that depends upon the sperm fusing with the oocyte to introduce the sperm-specific phospholipase C ζ (PLC ζ), which then begins to hydrolyze PtdIns4,5P₂ to generate InsP₃ that initiates Ca²⁺ oscillations (Figure 1). These oscillations then control various oocyte responses, such as cortical granule release, the block to polyspermy, the completion of meiosis, and the initiation of the cell-cycle program leading to embryonic development.

Cell Proliferation

The InsP₃/Ca²⁺ signaling pathway also has a role to play in the control of cell proliferation in many cell types. This role is well established in lymphocyte activation, which is initiated by antigen binding to the T-cell receptor (TCR) that then recruits PLC γ 1 to produce InsP₃ that acts through InsP₃R to release Ca²⁺ from the internal store. As the ER empties, it sends a signal to the Ca²⁺ release-activated Ca²⁺ (CRAC) channel responsible for maintaining Ca²⁺ entry for the 2-h period required to induce lymphocytes to grow. The coupling between store emptying and activation of the CRAC channel depends

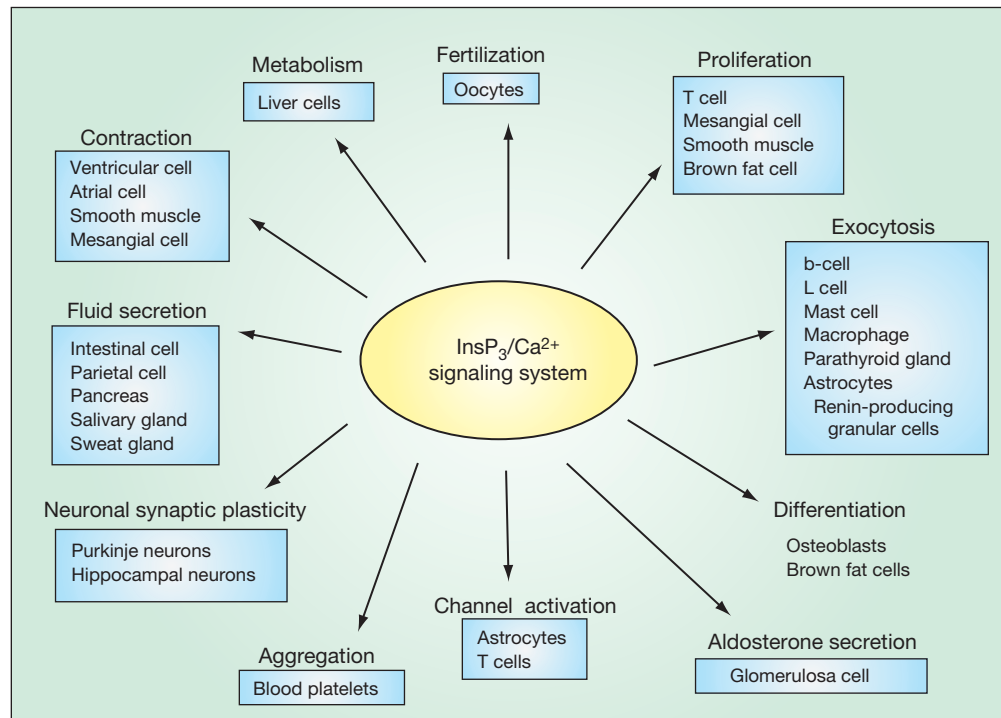


Figure 5 Diverse functions of the $\text{InsP}_3/\text{Ca}^{2+}$ signaling pathway. The mobilization of Ca^{2+} by inositol 1,4,5-trisphosphate (InsP_3) functions in the control of many different cellular processes in a wide range of cell types. Details of some of these functions are outlined in the text.

on an interaction with the stromal interaction molecule 1 (STIM1) located in the ER. Both the release and entry processes generate the Ca^{2+} signal to stimulate calcineurin (CaN) to dephosphorylate the nuclear factor of activated T cells (NFAT), which enters the nucleus where it initiates the transcriptional events that lead to cell proliferation.

Neuronal Ca^{2+} Signaling

The brain is particularly rich in the components that function in the $\text{InsP}_3/\text{Ca}^{2+}$ signaling pathway, which may play an important role in contributing to various neuronal processes, such as the changes in synaptic plasticity that underlie learning and memory. In the case of Purkinje neurons, InsP_3R are located on the ER that extends into the multitude of spines that decorate the dendritic tree. The metabotropic receptors, which are also located on the spines, are responsible for generating InsP_3 that contributes to the process of long-term depression (LTD) that is responsible for motor learning. Dysregulation of the $\text{InsP}_3/\text{Ca}^{2+}$ signaling system has been implicated in Alzheimer's disease (AD).

Cardiac Cells

Cardiac cells provide another example of how the $\text{InsP}_3/\text{Ca}^{2+}$ signaling pathway can modify the ongoing Ca^{2+} signaling system that drives contraction. In both atrial and ventricular cells, microdomains of Ca^{2+} known as ' Ca^{2+} sparks' create the global signal that drives contraction. This basic signaling system can be modified by the $\text{InsP}_3/\text{Ca}^{2+}$ signaling system in both atrial and ventricular cells. Under normal conditions, the

atrial sparks are confined to a region near the plasma membrane because they fail to breach a mitochondrial firewall. Stronger contractions are induced following stimulation of the atrial cells by endothelin, and this large positive inotropic response depends on the ability of InsP_3 to enhance the cell-surface sparks so that they breach the mitochondrial firewall to stimulate the nonjunctional RyR2s to ignite the Ca^{2+} wave that produces the larger global signals. Excessive activation of the $\text{InsP}_3/\text{Ca}^{2+}$ signaling pathway might be the cause of atrial arrhythmias.

In the case of ventricular cells, the $\text{InsP}_3/\text{Ca}^{2+}$ signaling pathway may also contribute to positive inotropic responses to agonists, such as endothelin, by enhancing the sparks that are responsible for excitation-contraction coupling. The spatial organization of the enhanced Ca^{2+} signal responsible for this positive inotropic response may be particularly important for understanding the processes responsible for cardiac hypertrophy and heart disease. During normal transients, each Ca^{2+} spark is located primarily in the cytoplasm where it functions to drive contraction. However, in the presence of agonists such as endothelin that increase the level of InsP_3 , this cytosolic signal may be amplified by the release of Ca^{2+} by InsP_3Rs located near the nucleus to give a nuclear-cytosolic signal capable of stimulating the transcriptional events responsible for hypertrophy.

See also: **Bioenergetics:** Phosphatidylinositol-3-Phosphate; Ryanodine Receptor Calcium Ion Channels; **Metabolism Vitamins and Hormones:** Chemistry of Alzheimer Disease; **Signaling:** Phospholipase C.

Further Reading

- Berridge MJ (1998) Neural calcium signaling. *Neuron* 21: 13–26.
- Berridge MJ (2006) Calcium microdomains: Organization and function. *Cell Calcium* 40: 405–412.
- Berridge MJ (2008) Smooth muscle cell calcium activation mechanisms. *Journal of Physiology* 586: 5047–5061.
- Berridge MJ (2010) Calcium hypothesis of Alzheimer's disease. *Pflügers Archiv: European Journal of Physiology* 459: 441–449.
- Berridge MJ, Bootman MD, and Roderick HL (2003) Calcium signalling: Dynamics, homeostasis and remodelling. *Nature Reviews Molecular Cell Biology* 4: 517–529.
- Bezprozvanny I and Mattson MP (2008) Neuronal calcium mishandling and the pathogenesis of Alzheimer's disease. *Trends in Neuroscience* 31: 454–463.
- Carafoli E (2004) Special issue: Calcium signaling and disease. *Biochemical and Biophysical Research Communications* 322: 1097.
- Clapham DE (2007) Calcium signaling. *Cell* 131: 1047–1058.
- Feske S (2007) Calcium signalling in lymphocyte activation and disease. *Nature Reviews Immunology* 7: 690–702.
- Hamilton SL (2005) Ryanodine receptors. *Cell Calcium* 38: 253–260.
- Kocksämper J, Zima AV, Roderick HL, et al. (2008) Emerging roles of inositol 1,4,5-trisphosphate signalling in cardiac myocytes. *Journal of Molecular and Cellular Cardiology* 45: 128–147.
- Lewis RS (2007) The molecular choreography of a store-operated calcium channel. *Nature* 446: 284–287.
- Mikoshiha K (2007) IP₃ receptor/Ca²⁺ channel: From discovery to new signaling concepts. *Journal of Neurochemistry* 102: 1426–1446.
- Rizzuto R and Pozzan T (2006) Microdomains of intracellular Ca²⁺: Molecular determinants and functional consequences. *Physiological Reviews* 86: 369–408.
- Stutzmann GE (2007) The pathogenesis of Alzheimer's disease – is it a lifelong “calciumopathy”. *Neuroscientist* 13: 546–559.
- Taylor CW, da Fonseca PCA, and Morris EP (2004) IP₃ receptors: The search for structure. *Trends in Biochemical Sciences* 29: 210–219.

Relevant Website

<http://www.cellsignallingbiology.org> – Cell Signalling Biology: Homepage.

Insect Biochemistry/Hormones

P Y Scaraffia and R L Miesfeld, University of Arizona, Tucson, AZ, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Gonotrophic cycle Reproductive process of ovarian development and egg laying in blood-feeding insects.

Hemolymph The circulatory fluid of insects, similar to blood or lymphatic fluid in other organisms.

Holometabolous insects Insects that undergo complete metamorphosis from eggs to larvae to pupae to adults.

Overview of Insects

Insects are the most abundant multicellular organisms on planet Earth and are thought to account for >70% of all species. So far, over 1 million insect species have been identified, with estimates putting the total number of insect species worldwide at 6–10 million. Insects are invertebrates that represent a class of arthropods characterized by three pairs of jointed legs; a three-part body consisting of a head, thorax, and abdomen; compound eyes; a chitin exoskeleton; and two antennae. Holometabolous insects hatch from eggs and progress through three distinct life stages. Newly hatched eggs produce larvae that go through metamorphosis, which is separated by molts. Since the chitin exoskeleton cannot accommodate the growing insect, molting provides a way for insects to shed their exoskeleton multiple times until they reach the pupal stage in which they complete metamorphosis in a protected shell called a pupal case. In the final developmental stage, the adult insect emerges from the pupal case, and within a few hours or days, females and males mate to produce fertilized eggs. In some other insects (hemimetabolous), the adult form from the hatching can be reached gradually with successive molts during a process called incomplete metamorphosis (cockroaches and grasshoppers). In this type of development, the immature stages (nymphs, or when aquatic, naiads) are like the adults except in size, flight, and sexual maturity.

Endocrine regulation in insects relies on a large number of peptide and nonpeptide hormones, several of which have been studied in great detail. Three of the best-characterized insect hormones are ecdysone, juvenile hormone (JH), and prothoracicotropic hormone (PTTH), all of which play important roles in insect metamorphosis and reproduction. Ecdysone, and most likely JH, bind to receptor proteins that function as transcription factors, which directly or indirectly regulate the expression of specific target genes. In contrast, PTTH binds to a membrane-bound receptor tyrosine kinase in the prothoracic gland of immature insects and initiates an intracellular phosphorylation cascade that activates ecdysone synthesis. Ecdysone synthesis in adult insects occurs primarily in the ovaries in response to the brain neuropeptide ovarian ecdysteroidogenic hormone.

Insect Biochemistry

Insect Metabolomics

Insects provide an ideal model system to investigate metabolic regulation in multicellular organisms because they can often

be easily manipulated for biochemical studies, and for some of the most important insect species, whole genome sequence information is readily available, which greatly facilitates molecular genetic approaches. Metabolomics is the systematic study of small molecules that are utilized or generated by metabolic processes in living cells. The most common type of metabolomic study involves collecting biochemical samples from cell or tissue cultures, or whole organisms, under various conditions that can be quantitatively and qualitatively analyzed by high-performance liquid chromatography (HPLC) and mass spectrometry. Based on bioinformatic analysis of genome sequences, it is often possible to predict what metabolic pathways in insects are responsible for changes in metabolite profiles as a consequence of environmental signals or genomic variations in metabolic enzymes.

One example of how insect biochemistry can be investigated using a metabolomics approach is the mass spectrometry-based study of the *rosy* (*ry*) mutation in *Drosophila melanogaster*. The *ry* mutation maps to the xanthine oxidase gene that encodes an enzyme required to convert xanthine, a product of purine degradation, to uric acid. In *D. melanogaster*, loss of xanthine oxidase causes a reddish brown eye color due to a deficiency in red pigment, which is how this mutation was first identified. In humans, xanthine oxidase deficiency leads to a condition called xanthinuria that is characterized by xanthine accumulation in the urinary tract and kidney. Using whole-body extracts of *D. melanogaster* wild-type and *ry* mutant flies, mass spectrometric analysis revealed significant increases in xanthine and hypoxanthine, as well as higher levels of guanine metabolites and bipterins. Surprisingly, the *ry* mutant defect in xanthine oxidase was associated with a 15-fold increase in glycerophosphocholine levels, presumably due to osmotic stress caused by the inability to excrete uric acid.

Another example of how metabolomics have been used to investigate biochemical processes in insects, was a recent study in which larvae of the Antarctic midge *Belgica antarctica* were exposed to high heat, freezing temperatures, or desiccation, and the levels of 75 metabolites were analyzed by mass spectrometry. It was found that the steady-state level of a variety of sugars, amino acids, and citric acid cycle intermediates were differentially affected by these extreme environmental conditions. Specifically, levels of alanine, aspartate, mannitol, and urea were elevated after 6 h of freezing conditions (−10 °C), whereas glycine and serine levels were decreased. Similarly, it was found that levels of α -ketoglutarate and putrescine were elevated after 1 h at 30 °C, but glycerol, serine, and glucose

levels were decreased after a 1 h heat shock. Since serine levels decreased under both conditions of extreme cold and heat, as well as when the larvae were subjected to desiccating conditions, it was proposed that decreased serine levels may be a marker of environmental stress in this insect species.

Blood Meal Metabolism in Mosquitoes

Because of their role as biological vectors for many of the arthropod-borne infectious diseases in humans, mosquitoes have been extensively studied at the biochemical level. Mosquitoes are responsible for transmitting not only malarial parasites and a variety of arboviruses, but also nematode worms that cause lymphatic filariasis. It is estimated that 40% of the world's population is at risk of being infected by malaria, with 300 million people being infected each year and over 1 million deaths. Malaria is primarily transmitted by the mosquito species *Anopheles gambiae* in Africa, *Anopheles albimanus* in South America, and by *Anopheles stephensi* in Asia. The other major mosquito-borne disease affecting humans is Dengue fever, which can develop into Dengue hemorrhagic fever and be fatal. *Aedes aegypti* mosquitoes transmit Dengue virus and are one of the most rapidly spreading human disease vectors in the world. It is estimated that up to 3 billion people in the Americas, Africa, and Asia are at risk of being infected by Dengue virus, and that of the nearly 50 million new Dengue infections each year, 20 000 deaths will occur from Dengue hemorrhagic fever.

Newly emerged *Ae. aegypti* female mosquitoes feed on nectar for several days until they are able to take their first blood meal (males do not blood feed). The blood meal is required for *Ae. aegypti* egg development and results in the deposition of ~100 fertilized eggs within 72 h of feeding. This process of blood digestion, oocyte maturation in the ovaries, and the laying of fertilized eggs is called the gonotrophic cycle. In the course of each gonotrophic cycle, blood meal metabolism requires efficient retrieval of nutrients, and rapid excretion of toxic products such as ammonia (defined here as NH_3 , NH_4^+ , or combination of both), in order to produce viable eggs. This is an amazing accomplishment considering the mosquito's size. A typical female mosquito weighs ~2 mg and can consume a blood meal of ~2 μl in few minutes. Taking into account the mass of water (~1.6 mg), protein (~400 μg), lipid (~10 μg), and carbohydrate (~1 μg) in this 2- μl blood meal, it can be

calculated that the female mosquito consumes her own body weight in a single feeding. This feat would be equivalent to a 125 lb woman drinking a 12 gallon smoothie that contains 25 lbs of hamburger meat, 0.5 lb of butter, and 2 tbs of sugar. As metabolically challenging as blood meal digestion must be, the female *Ae. aegypti* mosquito is highly capable and typically completes 4–7 gonotrophic cycles to produce up to 400 mosquito progeny over a 1-month period.

The vast majority of nutrients, in a typical mosquito blood meal, are derived from blood meal proteins, of which three constitute 80% of the total protein. These are hemoglobin (~330 μg), serum albumin (~50 μg), and IgG antibodies (~15 μg). Biochemical and molecular genetic studies have shown that mosquito blood meal digestion is initiated by endoprotease and exopeptidase enzymes that are secreted into the lumen by midgut epithelial cells within 30 min of feeding. Although mosquitoes require 10 of the 20 naturally occurring amino acids in their diet (essential amino acids), metabolic labeling studies using ^{14}C -protein revealed that only ~14% of the ingested protein is recycled into newly synthesized maternal proteins or egg proteins (Figure 1). In contrast, ~70% of the amino acids retrieved from the blood meal are deaminated to generate reduced carbon that is used for immediate energy needs or excreted as waste. Blood meal-derived amino acids are also used for lipogenesis to supply stored energy for the female mosquito (~10%), and converted to egg lipids that are used to complete oocyte maturation (~6%).

Urea Synthesis in *Ae. aegypti*

Excess nitrogen cannot be stored in cells, and therefore, nitrogen balance must be highly regulated to prevent ammonia toxicity, while at the same time, providing sufficient nitrogen for the biosynthesis of amino acids, nucleotide bases, and coenzymes. Aquatic organisms excrete most of their excess nitrogen directly into water as ammonia, whereas animals that spend all or part of their life cycle on land excrete nitrogen as uric acid or urea. Mammals synthesize urea via a metabolic pathway called the urea cycle that derives the two nitrogen atoms in urea from carbamoyl phosphate and aspartate. Certain amphibians and some fish do not express urea cycle enzymes; however, the urea they produce is the result of uric acid degradation by the uricolytic pathway (Figure 2). The

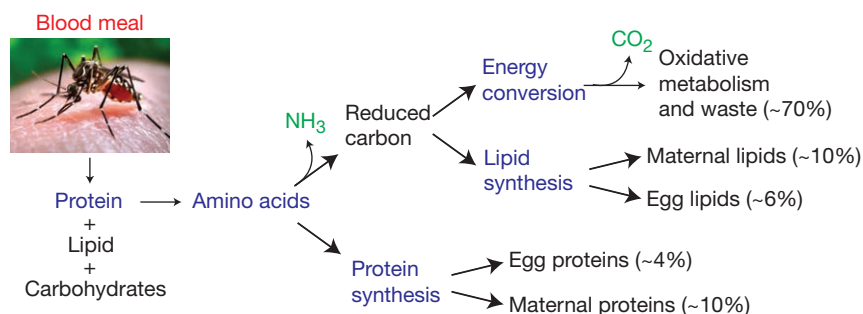


Figure 1 Blood meal metabolism in female *Aedes aegypti* mosquitoes. The major nutrient in the mosquito blood meal is protein, which is converted to amino acids that are then deaminated and used as reduced carbon for energy conversion and lipid synthesis, or used directly for the synthesis maternal and egg proteins.

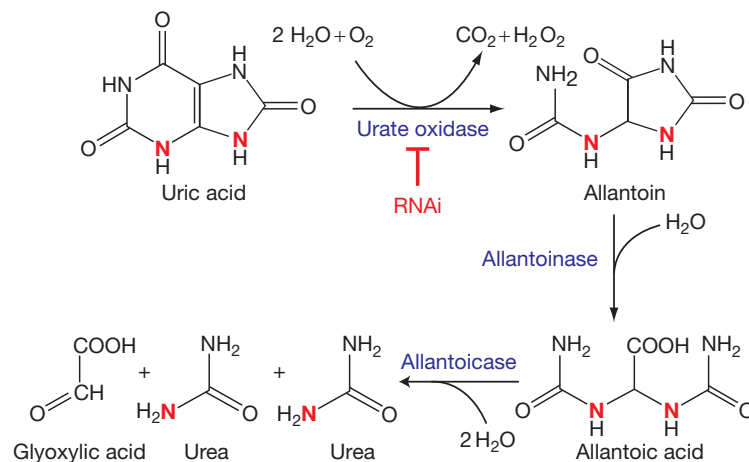


Figure 2 The uricolytic pathway converts uric acid into urea. Feeding experiments in *Aedes aegypti* mosquitoes with ¹⁵NH₄Cl have been used to quantitate uricolytic pathway intermediates by mass spectrometry (bolded N atoms were ¹⁵N-labeled in these experiments). RNAi-mediated knockdown of urate oxidase expression in blood-fed mosquitoes confirmed that the uricolytic pathway is functional under normal physiological conditions.

uricolytic pathway converts uric acid to urea and glyoxylic acid through a three-step process requiring the enzymes urate oxidase, allantoinase, and allantoicase.

Early metabolic studies of nitrogen metabolism in blood-fed *Ae. aegypti* mosquitoes showed that excess nitrogen produced by blood meal digestion is excreted as ammonia, amino acids, uric acid, and urea. Based on studies in vertebrates, it was assumed that the urea present in mosquito feces was produced by converting excess dietary arginine into urea via the arginase reaction. However, by using a combination of ¹⁵N-isotope labeling experiments, bioinformatic analysis of the *Ae. aegypti* genome sequence, and RNAi-mediated knockdown of urate oxidase expression, it was recently shown that a portion of this excreted urea is in fact derived from the uricolytic pathway. As illustrated in **Figure 2**, mass spectrometric analysis of uric acid containing ¹⁵N-labeled in two positions revealed that uric acid is metabolized to 2 mol of ¹⁵N-labeled urea, thereby providing biochemical evidence that the uricolytic pathway is functional in *Ae. aegypti* mosquitoes. In addition, RNAi-mediated knockdown of urate oxidase expression in ¹⁵NH₄Cl-fed mosquitoes was found to significantly decrease the amount of ¹⁵N-urea in the feces. Similarly, knocking down urate oxidase expression in blood-fed mosquitoes resulted in a buildup of uric acid in mosquito tissues. Given the unusually high-protein content of the mosquito blood meal, it is likely that the uricolytic pathway provides a mechanism for the adult female mosquito to maximize nutrient uptake while minimizing the risk of ammonia toxicity.

Insect Hormones

Endocrine regulation in insects is complex with many different types of hormones involved, including steroids (ecdysone), lipid compounds (JH), multiple peptide hormones (PTTH; ovarian ecdysteroidogenic hormone, OEH; adipokinetic hormone, AKH), and biogenic amines (serotonin, DOPA). The discussion here focuses on the two major gonadotropins, ecdysone and JH, and the peptide hormone PTTH, which

regulates ecdysteroidogenesis in the larval and pupal prothoracic gland. Insect endocrinology has been extensively studied in *D. melanogaster*, *Manduca sexta*, and *Ae. aegypti*, primarily with regard to the structure and function of 20-hydroxyecdysone (20E), the biologically active form of ecdysone, JH-III, the most abundant JH found in insects, and a variety of peptide hormones, of which PTTH is one of the best characterized at the biochemical level.

20-Hydroxyecdysone

The steroid hormone ecdysone is major regulator of molting, metamorphosis, and reproduction in insects. The site of ecdysone synthesis in insect nymphs or larvae is the prothoracic gland in response to PTTH signaling, whereas in the adult the majority of ecdysone is synthesized in the follicle cells of the ovaries. Ecdysone is secreted into the hemolymph and absorbed by fat body cells where it is converted to 20E by the enzyme ecdysone 20 monooxygenase, a P450 enzyme expressed in fat body tissue. The 20E product of this reaction is released into the hemolymph where it travels to target tissues by classic endocrine mechanisms. Physiological studies of 20E function in insects have shown that it controls the molting process during metamorphosis, as well as a number of other physiological processes including the wandering behavior of larvae, development of complex neuronal networks, and oocyte maturation in adult female insects.

In adult female mosquitoes, 20E is required for completion of the gonotrophic cycle and production of viable eggs. Shortly after taking a blood meal, the neuropeptide OEH is released from the brain and is transported by the hemolymph to the ovaries where it stimulates the synthesis of ecdysone and 20E via a cyclic-AMP second-messenger signaling cascade. The biochemical response to 20E production is yolk protein synthesis in the fat body, and egg maturation in the ovaries, a process known as vitellogenesis (the various yolk proteins are named vitellogenins). The yolk proteins are secreted by fat body cells into the hemolymph where they are recovered by receptor-mediated endocytosis in the developing oocytes. Within 48h

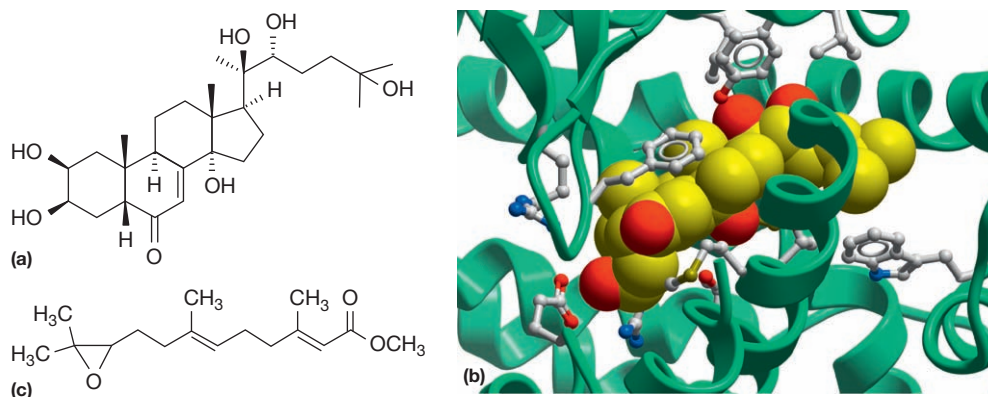


Figure 3 20-Hydroxyecdysone (20E) and juvenile hormone III (JH-III) control metamorphosis and reproduction in insects. (a) Chemical structure of 20E showing the four ring cholesterol core. (b) 20E binds to a hydrophobic pocket in the ligand-binding domain of the ecdysone receptor protein. Molecular structure based on PDB file 2R40. (c) Chemical structure of the terpenoid-derived hormone JH-III.

of blood feeding, 20E concentrations in the hemolymph decrease to previtellogenic levels, and JH titers rise sharply in preparation for the next blood meal and gonotrophic cycle.

The chemical structure of 20E is shown in **Figure 3(a)** where it can be seen that it consists primarily of a central four-ring core derived from cholesterol, as is true of all steroidal hormones that have been characterized biochemically. Insects cannot synthesize cholesterol *de novo*, and therefore, ecdysone biosynthesis is dependent on dietary cholesterol uptake during the larval and adult stages. The mechanism of action of 20E is mediated through binding and activation of a heterodimeric transcription factor consisting of two DNA-binding nuclear receptor proteins called the ecdysone receptor (ECR) and ultraspiracle (USP). The ligand-binding domain of the ECR protein from *Heliothis virescens* (Tobacco budworm) contains numerous hydrophobic amino acids, such as phenylalanine, tyrosine, and tryptophan, that line the 20E-binding pocket (**Figure 3(b)**).

Studies of the *D. melanogaster* ECR protein have shown that dimerization with USP occurs in the absence of 20E, but upon 20E binding, the ECR–USP complex is translocated to the nucleus where it interacts with specific DNA sequences that function as ecdysone response elements. The molecular structure of the DNA-binding domains of the *D. melanogaster* ECR and USP proteins bound to a pseudopalindromic ecdysone response element located upstream of the hsp27 gene promoter is shown in **Figure 4**. It can be seen that the USP-subunit binds within the major groove of DNA to the 5' half-site of the hsp27 response element and makes contacts through an α -helical region of the protein with nucleotides on both DNA strands. Similarly, the ECR subunit binds to the 3' half-site sequence that also lies within the major groove of the DNA helix and makes several direct nucleotide contacts. The net result of this sequence-specific DNA-binding event is the recruitment of auxiliary transcription factors that bind to either adjacent DNA sequences, or through protein–protein interactions, to the ECR–USP complex. One of the most thoroughly characterized examples of 20E-mediated regulation of gene expression in insects is the induced expression of multiple vitellogenin genes in the fat body of blood-fed female *Ae. aegypti* mosquitoes. Using quantitative real-time polymerase chain reaction assays, it was shown that 20E signaling through

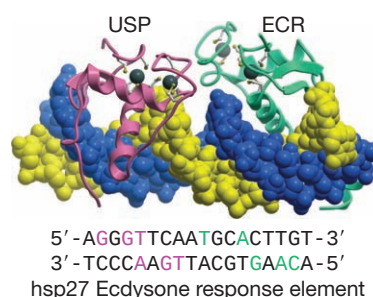


Figure 4 20E signaling is mediated by a nuclear receptor heterodimer consisting of the ecdysone receptor (ECR) and the ultraspiracle (USP) proteins. The DNA-binding domains of ECR and USP make direct contacts to nucleotide bases contained in the two half-site sequences of the ecdysone response element in the *Drosophila melanogaster* hsp27 gene (bolded nucleotides). Molecular structure based on PDB file 2HAN.

the ECR–USP heterodimeric complex induces *Ae. aegypti* vitellogenin gene expression by as much as 10000-fold within 24 h of blood feeding.

Juvenile Hormone (JH-III)

Insects have a class of terpenoid compounds, collectively called JHs that regulate metamorphosis. The best-characterized and most abundant of these hormones is JH-III, the structure of which is shown in **Figure 3(c)**. JHs were first discovered in the kissing bug *Rhodnius prolixus* in 1934 by Vincent Wigglesworth who described them as compounds that inhibit the final stage of metamorphosis when applied to the insect exoskeleton. He also showed that JH was required in adult insects to promote egg maturation. JH-III is synthesized in a pair of glands in the insect head called the corpora allata and functions in much the same way as 20E in that it regulates gene expression through a receptor-mediated signaling process. In fact, both 20E and JH-III have been found to regulate the same developmental process, but often in an opposing way, with JH-III antagonizing, or modulating, 20E signaling functions. Peak levels of JH-III and 20E in the hemolymph of female mosquitoes are diametrically opposed depending on whether the mosquito is sugar fed or blood fed, respectively.

Although much is known about the synthesis and biological function of JH-III in several insect species, it is still not exactly clear how JH-III regulates gene expression in target tissues. The search for a JH receptor in *D. melanogaster* has so far led to two candidate proteins, both of which are DNA-binding transcription factors. One is the USP protein that forms a complex with ECR and has been found to bind JH-III as a ligand with low affinity ($\sim 10^{-7}$ M). More recent studies have reported that USP also binds the metabolic precursor to JH-III, called methyl farnesoate, with high affinity ($\sim 10^{-9}$ M), which is closer to what would be expected for a physiological ligand of a typical nuclear receptor protein.

The second candidate JH receptor is a basic helix-loop-helix protein encoded by the methoprene-tolerant gene (*Met*) that binds JH-III with affinities in the nanomolar range. *Met* contains a PAS domain that is 40% identical to the dioxin receptor translocator protein ARNT which binds to the dioxin receptor, another ligand-activated transcription factor. Based on studies showing modulation of 20E-mediated transcription by increased levels of JH-III, it has been proposed that ECR-USP heterodimers might form complexes with *Met* in a way that facilitates cross-talk modulation of gene expression by E20 and JH-III at specific promoters. In addition, several other DNA-binding proteins in *D. melanogaster*, called Chd64 and FKBP39, have been shown by yeast two-hybrid assays to be interaction partners with *Met* and to bind with high affinity to DNA sequences that function as JH response elements.

Prothoracicotropic Hormone

The neuropeptide PTTH is synthesized in the corpora allata gland in the insect brain and functions in metamorphosis as a hormonal activator of ecdysone synthesis in the prothoracic gland. The existence of PTTH was first described in the 1920s by Stefan Kopec, a Polish biologist, who used decapitation experiments to show that the insect head is required for molting. Later, Vincent Wigglesworth confirmed the presence of a brain hormone distinct from JH that was required for insect metamorphosis. The active form of PTTH is homodimer of identical subunits, each containing a proteolytically processed peptide of 109 amino acids. Because of its similarity to other growth hormones, it was hypothesized that PTTH most likely binds to a growth hormone-like receptor present in prothoracic gland cells during the molting cycle. A bioinformatics approach was recently used to identify receptor tyrosine kinase transcripts that are expressed in the *D. melanogaster* prothoracic gland at times when PTTH activity is known to be required for metamorphosis. One of the candidate genes identified was the Torso receptor tyrosine kinase, which was subsequently shown by genetic and biochemical analyses to be the long sought PTTH receptor. It was found that PTTH binding to the *D. melanogaster* Torso protein on the surface prothoracic gland cells stimulates the Ras → Raf → ERK intracellular signaling pathway and leads to the activation of ecdysone biosynthesis (Figure 5). Although several molecular details of ecdysteroidogenesis in the insect prothoracic gland still need to be resolved, at least six different P450 enzymes have recently been identified as essential components of the pathway in *D. melanogaster*. The genes encoding these P450 enzymes are called the Halloween genes because of their lethal mutant

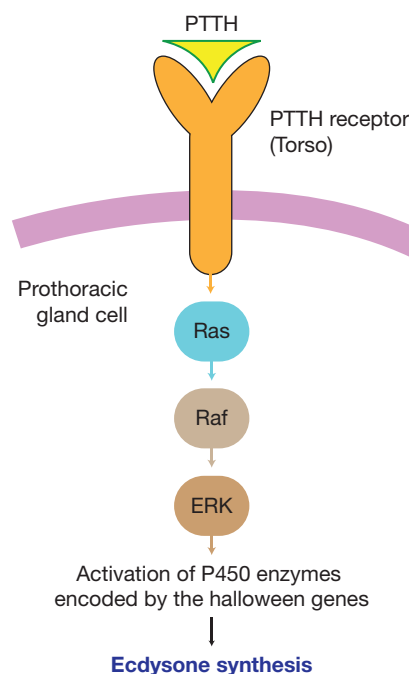


Figure 5 The insect neuropeptide hormone PTTH induces ecdysone synthesis in prothoracic gland cells by binding to and activating the receptor tyrosine kinase protein Torso. Activation of the Ras → Raf → ERK signaling pathway by Torso leads to stimulation of P450 enzyme activity and initiation of the ecdysteroidogenesis pathway. The Halloween genes discovered in *Drosophila* encode six of these p450 enzymes.

phenotype that is characterized by a defective exoskeleton. The *D. melanogaster* Halloween genes were first described by Eric Wieschaus and Christiane Nusslein-Volhard using a genetic screen for developmental mutants and creatively named *spook*, *spookier*, *phantom*, *disembodied*, *shadow*, and *shade*. The biochemical function and enzymatic properties of these six P450 enzymes are currently being investigated.

See also: **Bioenergetics:** Cytochrome P-450; **Metabolism Vitamins and Hormones:** Urea Cycle: Disease Aspects; **Signaling:** Vitamin D Receptor.

Further Reading

- Brown MR, Sieglaff DH, and Rees HH (2009) Gonadal ecdysteroidogenesis in arthropoda: Occurrence and regulation. *Annual Review of Entomology* 54: 105–125.
- Browning C, Martin E, Loch C, et al. (2007) Critical role of desolvation in the binding of 20-hydroxyecdysone to the ecdysone receptor. *Journal of Biological Chemistry* 282: 32924–32934.
- Grimmelikhuijzen CJ, Cazzamali G, Williamson M, and Hauser F (2007) The promise of insect genomics. *Pest Management Science* 63: 413–416.
- Jakob M, Kolodziejczyk R, Orlowski M, et al. (2007) Novel DNA-binding element within the C-terminal extension of the nuclear receptor DNA-binding domain. *Nucleic Acids Research* 35: 2705–2718.
- Kamleh MA, Hobani Y, Dow JA, and Watson DG (2008) Metabolomic profiling of *Drosophila* using liquid chromatography Fourier transform mass spectrometry. *FEBS Letters* 582: 2916–2922.

- Law JH and Wells MA (1989) Insects as biochemical models. *Journal of Biological Chemistry* 264: 16335–16338.
- Li Y, Zhang Z, Robinson GE, and Palli SR (2007) Identification and characterization of a juvenile hormone response element and its binding proteins. *Journal of Biological Chemistry* 282: 37605–37617.
- Michaud RM, Benoit JB, Lopez-Martinez G, et al. (2008) Metabolomics reveals unique and shared metabolic changes in response to heat shock, freezing and desiccation in the Antarctic midge, *Belgica antarctica*. *Journal of Insect Physiology* 54: 645–655.
- Rewitz KF, O'Connor MB, and Gilbert LI (2007) Molecular evolution of the insect Halloween family of cytochrome P450s: Phylogeny, gene organization and functional conservation. *Insect Biochemistry and Molecular Biology* 37: 741–753.
- Rewitz KF, Yamanaka N, Gilbert LI, and O'Connor MB (2009) The insect neuropeptide PTTH activates receptor tyrosine kinase Torso to initiate metamorphosis. *Science* 326: 1403–1405.
- Riddiford LM (2008) Juvenile hormone action: A 2007 perspective. *Journal of Insect Physiology* 54: 895–901.
- Scaraffia PY, Tan G, Isoe J, et al. (2008) Discovery of an alternate metabolic pathway for urea synthesis in adult *Aedes aegypti* mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America* 105: 518–523.
- Spindler KD, Honl C, Tremmel C, et al. (2009) Ecdysteroid hormone action. *Cellular and Molecular Life Sciences* 66: 3837–3850.
- Warren JT, O'Connor MB, and Gilbert LI (2009) Studies on the black box: Incorporation of 3-oxo-7-dehydrocholesterol into ecdysteroids by *Drosophila melanogaster* and *Manduca sexta*. *Insect Biochemistry and Molecular Biology* 39: 677–687.
- Zhou G, Flowers M, Friedrich K, et al. (2004) Metabolic fate of [^{14}C]-labeled meal protein amino acids in *Aedes aegypti* mosquitoes. *Journal of Insect Physiology* 50: 337–349.

Insulin- and Glucagon-Secreting Cells of the Pancreas

F M Matschinsky, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

Published by Elsevier Inc.

This article is reproduced from the previous edition, volume 2, pp. 430–435, © 2004, Elsevier Inc.

Glossary

Beta-cell threshold for glucose-stimulated insulin release

The level of glucose required to initiate insulin release, that is, about 5 mM. This beta-cell threshold is the critical determinant of the glucose setpoint of the body.

Glucokinase (GK) disease Hypo- and hyperglycemia syndromes in humans caused by activating and inactivating mutations of glucokinase. This experiment of nature is the most compelling evidence for the glucokinase glucose sensor concept and its importance in understanding glucose homeostasis.

Glucokinase glucose sensor Glucokinase (also known as hexokinase IV or D) phosphorylates D-glucose using adenosine triphosphate (ATP), forming glucose-6-phosphate. The enzyme serves as a glucose sensor of the pancreatic beta cells because it is rate limiting for glucose metabolism (i.e., glycolysis and oxidation), which is an absolute prerequisite for the triggering of insulin secretion.

Glucose setpoint of the body Blood glucose level maintained precisely by the fine-tuning of opposing processes that either produce or consume glucose. In humans, this setpoint is close to 5 mM.

Alpha and beta cells are the predominant cell types in microscopic endocrine organs called the 'islets of Langerhans', which are scattered throughout the exocrine pancreas and produce, store, and secrete the hormones glucagon and insulin, respectively. Secretion of these hormones is regulated by fuels (glucose, amino acids, and fatty acids) or through neuroendocrine input. Insulin and glucagon in turn regulate the metabolism of these fuels. Loss of beta cells by autoimmune destruction or impairment of their function by poorly understood mechanisms are the causes of type I or type II diabetes mellitus (T1DM or T2DM), respectively. Individuals with T1DM require insulin substitution therapy for survival, and those with T2DM can manage their disease with diet restrictions, exercise, and a variety of drugs with different mechanisms of action. It is thus of critical importance to develop a comprehensive understanding of insulin- and glucagon-producing cells of the pancreas.

Introduction

Pancreatic alpha and beta cells play the predominant role in maintaining glucose homeostasis in humans and many laboratory animals and keep the blood glucose in the narrow range of 5–8 mM in response to the demands of feeding and fasting. Both cells have a threshold of approximately 5 mM glucose for stimulation of hormone release. The alpha cell is triggered to release glucagon when the glucose falls below this level, whereas the beta cell is activated to secrete insulin when this level is exceeded. Expressed differently, the inhibitory glucose dependency curve for alpha cells and the stimulatory glucose dependency curve for beta cells have their crossover point at 5 mM, which is the setpoint of the system for glucose homeostasis. (Note that alpha-cell deinhibition with falling glucose manifests itself most clearly when these cells are also under the influence of physiological stimuli, e.g., amino acids, a condition that is met *in situ*.) Auxiliary

neuroendocrine mechanisms involving acetylcholine, catecholamines, glucagon-like peptide 1 (GLP-1), and gastric inhibitory peptide (GIP) enhance or blunt the response of the pancreatic endocrine cells to fuel stimulation. The other two fuel classes, amino acids (AAs) and fatty acids (FAs), stimulate insulin release in a strictly glucose-dependent manner. AAs are the most effective fuel stimulant of alpha cells at biologically relevant concentrations (e.g., 3.5–7.0 mM of the physiological mixture of 20 AA enhances glucagon release markedly when glucose falls below 5.0 mM).

This article focuses on the mechanisms of glucose-stimulated insulin release (GSIR) because this issue is central to beta-cell function and because knowledge of this aspect of beta-cell biology and glucose homeostasis is well established, much less tenuous or controversial than current understanding of how the other fuels act on alpha and beta cells. However, when important and possible, how glucose and other fuels or neuroendocrine factors interact with these two types of pancreatic endocrine cells is discussed.

The Beta-Cell Glucokinase Glucose Sensor and the Threshold for GSIR

The glucokinase glucose sensor concept and the threshold concept for GSIR explain a great deal about beta cells and the nature of glucose homeostasis in humans and most laboratory animals. Glucokinase (GK) serves as the beta cell's cytosolic glucose-sensing device and contributes a critical element to the regulatory feedback loop or signaling chain that interconnects blood glucose, the beta cell, the hormone insulin, and the major insulin target tissues (i.e., the liver, muscle, and adipose tissue). The kinetic characteristics of GK are ideally suited for this purpose: the k_{cat} for glucose phosphorylation is about 60 s^{-1} ; the glucose $S_{0.5}$ (i.e., the glucose concentration resulting in half-maximal enzymatic activity) is about 8.0 mM; the glucose dependency curve of GK action is sigmoidal as

indicated by a Hill coefficient (n_h) of 1.7; the adenosine triphosphate (ATP) K_m is about 0.4 mM, indicating that the enzyme operates *in vivo* near saturation with its second substrate; finally, there is no feedback inhibition by glucose-6-P in contrast to the highly effective inhibition of other hexokinases by this metabolite. Of great theoretical interest and practical importance is the recent recognition of an allosteric activator site on GK that explains the mechanism of action and efficacy of a newly discovered class of antidiabetic agents (GK activators (GKAs); see following discussion). Beta-cell GK is inducible by glucose (maximally five- to tenfold compared to basal levels at 1.0 mM glucose) through mechanisms yet to be fully clarified. Note that hepatic GK is induced by insulin in a glucose-independent manner and is also activated by GK activators (GKAs). Note also that hepatic GK is regulated by the inhibitory GK regulatory protein (GKRP), which sequesters the enzyme to the nuclear compartment from which GK is released by glucose during hyperglycemia. By contrast, GKRP seems to be of little importance for beta-cell GK function. Differential expression control of islet and hepatic GK is explained by the presence of two tissue-specific promoters in one single gene. Due to these characteristic features, GK determines the rate of beta-cell glycolysis with a control strength approaching unity, outperforming all other metabolic steps. Glycolytic flux of beta cells is therefore a function of substrate pressure rather than of

energy demand, which determines glucose use of many other tissues, including brain and muscle.

These features are a prerequisite for the glucose sensor role of GK; they do not, however, explain the 5-mM setpoint for glucose homeostasis and threshold for GSIR. This threshold is a function of the biophysical characteristics of the K^+ and Ca^{2+} channels of the beta-cell membrane (Figure 1). The K -ATP channel is a hetero-octamer composed of four subunits each of the Kir6.2 inward rectifier and the drug target SUR-1 and is regulated by the phosphate potential of the beta cells. Important for physiology, it is inhibited by ATP^{4-} and activated by $MgADP^{2-}$ such that a decrease of the cytosolic ATP/ADP ratio determines K^+ conductivity and consequently the membrane potential. Enhanced glucose metabolism increases this ratio and depolarizes the beta cell. Alterations of the membrane potential are transmitted to the voltage-sensitive L-type Ca^{2+} channel. Depolarization stimulates Ca^{2+} influx and leads to elevation of the free intracellular calcium ($[Ca^{2+}]_i$). The combination of high ATP and $[Ca^{2+}]_i$ elicits insulin release. Important for understanding the threshold phenomenon for GSIR is the fact that the decrease of the K^+ conductivity is a graded function of glucose metabolism (i.e., the GK rate), whereas the L-type Ca^{2+} channel has a well-defined triggering potential of -60 to 50 mV, which happens to be reached when GK or glycolysis operates at 25–30% of

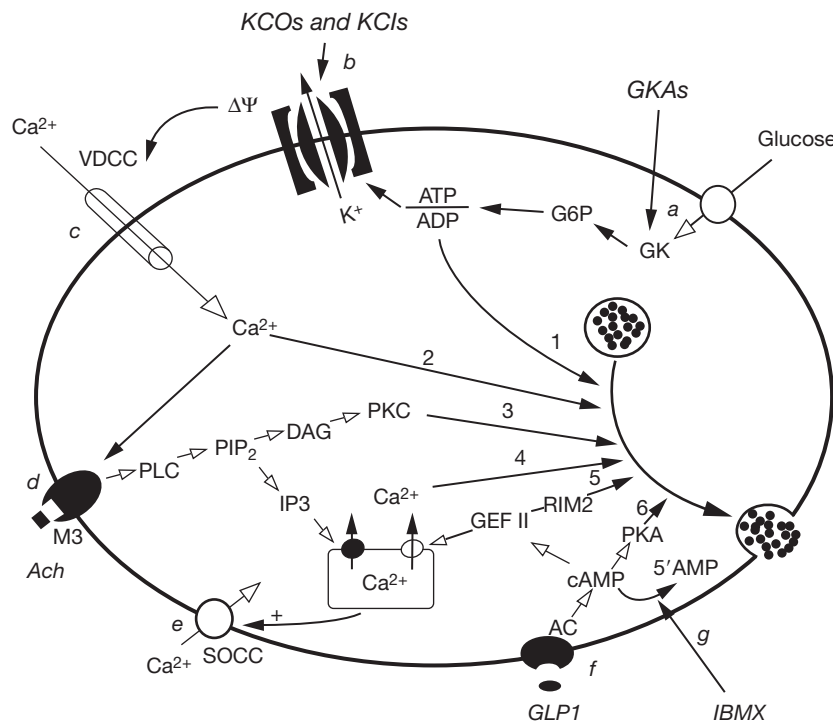


Figure 1 Signaling pathways in pancreatic beta cells. Six distinct signaling pathways in pancreatic beta cells are depicted (1–6). Abbreviations are as follows, with a–g identifying initial or other critical steps for these pathways: GK, glucokinase (a); the K_{ATP} channel complex, consisting of ATP-sensitive K^+ channels and sulfonyl urea receptor type 1 (SUR-1) subunits (b); VDCC, voltage-dependent Ca^{2+} channel or L-channels (c); acetylcholine (ACh) and muscarinic receptor type 3 (m3) (d); SOCC, store-operated calcium channels (e); GLP-1, glucagon-like peptide-1 and receptor (f); IBMX, 3-isobutyl-1-methylxanthine, a phosphodiesterase inhibitor (g). Other abbreviations are as follows: AC, adenylate cyclase; cAMP, cyclic AMP; PKC, protein kinase C; PKA, protein kinase A; DAG, diacylglyceride; IP3, inositol triphosphate; GEFII, cAMP-guanidine nucleotide exchange factor (or Epac2); $\Delta\Psi$, change of cell membrane potential. Additional abbreviations are explained in the text. Modified from Doliba NM, Qin W, Vatamaniuk MZ, et al. (2004) Restitution of defective glucose-stimulated insulin release of sulfonylurea type 1 receptor knockout mice by acetylcholine. *American Journal of Physiology – Endocrinology and Metabolism* 286(5): E834–E843.

its normal capacity, that is, at 5.0 mM glucose. Note that extra- and intracellular glucose are virtually identical at all times as a result of high-capacity glucose transport by Glut-2. This basic signaling pathway from glucose to insulin release as previously outlined is termed the 'triggering pathway (TPW)'. It is complemented by the so-called augmentation pathway (APW), which potentiates glucose action, but the molecular basis of the latter remains to be defined in detail.

Model Experiments to Illustrate the Operation of the TPW and the APW

The APW comes into play when $[Ca^{2+}]_i$ and ATP/ADP are elevated by the TPW and when the beta cells are activated in addition by acetylcholine, GLP-1, or GIP. Model experiments with perfused islets isolated from control mice and from mice lacking functional K^+ channels (i.e., from SUR-1^{-/-} knockout (KO) mice) illustrate many of the points discussed in the preceding section (Figure 2). First, beta cells from SUR-1^{-/-} do not show GSIR when acetylcholine is absent because the TPW is blocked in contrast to controls. Second, the distinctly different secretion profiles for GSIR evoked in control and SUR-1^{-/-} islets in the presence of low acetylcholine illustrate that the glucose response caused by the combined TPW and

APW in the control has a clear threshold and is blocked by diazoxide (a potassium channel opener (KCO)) as expected, whereas that of the APW alone as seen in the SUR-1^{-/-} islets is clearly graded (lacking a threshold) and is not blocked by diazoxide, again as expected. Both signaling pathways operate during a meal and result in the robust, threshold-based GSIR as mimicked by the experiment shown in Figure 2 for control islets.

AAs and FAs as Glucose-Dependent Stimuli of Beta Cells

Stimulation of insulin release by a physiological mixture of 20 AAs (20-AASIR) is biologically highly relevant but less well understood than GSIR. It is established, though, that 20-AASIR of normal islets has an absolute glucose requirement, that is, at least 2.5 mM. It is not clear which signaling pathways are involved, but it is likely that the TPW as sketched in the preceding discussion plays a role. Assuming a major role for metabolism in 20-AASIR, it must be postulated that transport into the cell and subsequent transamination are efficient processes. AA metabolism requires transamination as the first step, generating glutamate and alanine as the major products, followed by oxidative deamination of glutamate with

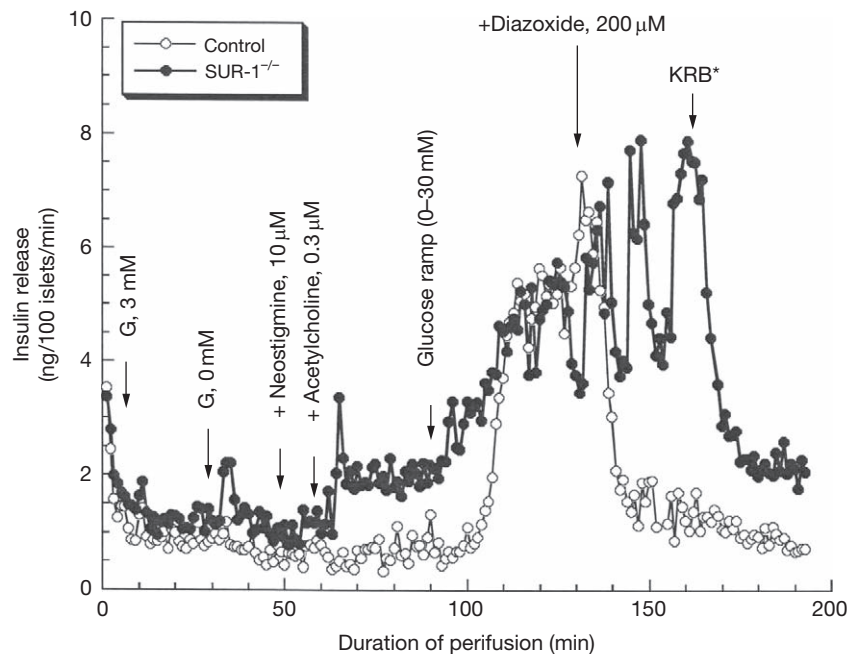


Figure 2 Experimental dissection of signaling pathways in glucose-stimulated insulin release (GSIR). Acetylcholine increases the insulin release in SUR-1^{-/-} islets in the absence of glucose, demonstrating marked hypersensitivity to the transmitter compared to controls. Acetylcholine also reconstitutes defective GSIR by these islets. After stable insulin secretion rates were achieved at 30 min, islets were perfused in the absence of glucose for 60 min (t-30–t-90). Following addition of neostigmine (a cholinesterase inhibitor) and acetylcholine, a glucose ramp (from 0 to 30 mM with an 0.8 mM increment per min) was applied for 37.5 min to control and SUR-1^{-/-} islets, and glucose was then maintained constant at 30 mM for about 30 min longer in the continued presence of acetylcholine (t-90–t-160). The experiment was completed by perfusing islets with Krebs–Ringer bicarbonate (KRB) solution lacking glucose and drugs. The rise in the glucose concentration in the perfusate led to an increase in hormone secretion in SUR-1^{-/-} islets as well as in the controls. However, note the different patterns of the release profiles. Also, appreciate that in the absence of acetylcholine and neostigmine, SUR-1^{-/-} islets did not respond to glucose in contrast to the controls (not shown). Diazoxide, a K_{ATP} channel activator, immediately decreased insulin release to baseline in control but was unable to block hormone release in SUR-1^{-/-} islets. Reproduced from Doliba NM, Qin W, Vatamaniuk MZ, et al. (2004) Restitution of defective glucose-stimulated insulin release of sulfonylurea type 1 receptor knockout mice by acetylcholine. *American Journal of Physiology – Endocrinology and Metabolism* 286(5): E834–E843.

glutamate dehydrogenase (GDH) to form nicotinamide adenine dinucleotide phosphate-oxidase (NAD(P)H), which is converted to ATP. The need for the α -ketoacids, α -ketoglutarate, and pyruvate as acceptors for the transamination reaction may explain, at least in part, the absolute requirement for glucose. It is also possible that AA metabolism generates metabolic coupling factors in addition to energy, with glutamate and glutamine as possibilities. It is indeed remarkable that more than half of the stimulatory effect of 20 AA can be attributed to glutamine, which contributes only 15% to the mixture. Studies of leucine-stimulated insulin release have been very extensive but have resulted in much controversy. It is now understood that leucine has two actions: it is converted to acetyl-CoA and acetoacetate, serving as direct fuel stimulant, or it stimulates GDH allosterically, thus enhancing glutaminolysis with the production of α -ketoglutarate, NAD(P)H, and ammonia. However, beta cells respond to this indirect allosteric effect of leucine only when GDH is sensitized, that

is, when it is not fully inhibited by guanosine triphosphate (GTP) (or ATP) as a result of enhanced glucose metabolism. 20-AASIR may also be mediated, at least in part, by depolarization due to increased Na^+ entry involving Na^+ -dependent transporters, a possibility that has received little attention. Some of this discussion is illustrated in studies with isolated perfused islets by the differential effect that energy run-down has on stimulation of insulin release by glucose, leucine, and α -ketoisocaproate (Figure 3). Isolated rat islets that had been cultured in 10 mM glucose and perfused with 2 mM glutamine show a remarkable change of chemosensitivity of their beta cells as a function of energy depletion with time. They respond well to glucose and α -ketoisocaproate at an early time point when they are totally blind to leucine, but as energy depletion proceeds they respond briskly to all three stimuli. FAs are effective glucose-dependent stimuli of insulin release (FASIR). It is now widely believed that specific metabolic products of FA metabolism play a more important role

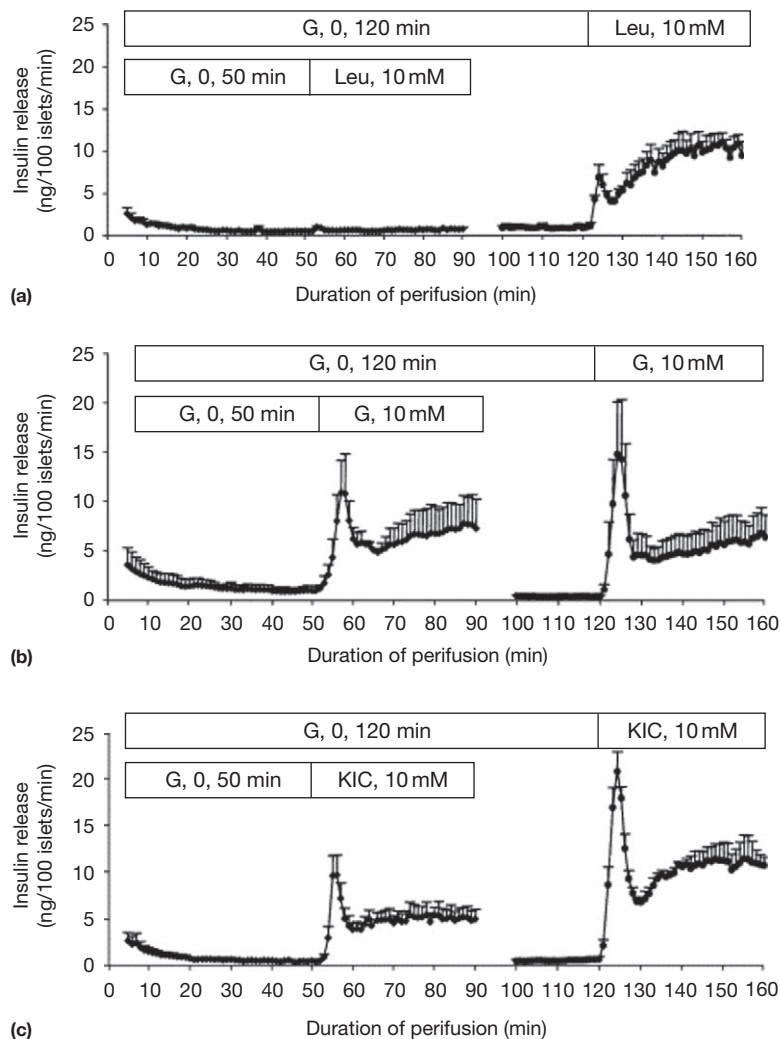


Figure 3 Effect of energy run-down on islet chemosensitivity. Isolated rat islets were cultured with 10 mM glucose for 3 days and then perfused with 2 mM glutamine in the absence of glucose for energy run-down periods of 50 min (diamonds) or 120 min (circles) prior to stimulation with (a) 10 mM leucine (Leu), (b) 10 mM glucose (G), and (c) 10 mM α -ketoisocaproate (KIC). Values represent the means $\pm$ SE for 100 islets from three separate perfusions. Reproduced from Li C, Najafi H, Daikhin Y, et al. (2003) Regulation of leucine-stimulated insulin secretion and glutamine metabolism in isolated rat islets. *Journal of Biological Chemistry* 278: 2853–2858.

as signals or coupling factors for secretion than increased generation of ATP by their catabolism. For example, long-chain acyl-CoA and/or diacylglycerol may serve as activators of protein kinase C (PKC). This idea has been expanded, and it is proposed that these and possibly other lipid-derived metabolic coupling factors are critical for GSIR as well. This notion is captured by speaking of the essentiality of FA for GSIR.

Neuroendocrine Modification of Fuel-Stimulated Insulin Release

Eating and digestion of food results in the production of the enterohormones GLP-1 and GIP and also enhances vagal activity, all of which augments fuel-stimulated insulin release. Catecholamines, on the other hand, block secretion very effectively. Recall that neuroendocrine stimulation of beta cells requires glucose. The signaling pathways are well established (Figure 1). The effects are quantitatively very marked, resulting in a left shift and increased magnitude of the fuel dose–response curves for activators and a right shift and decreased magnitude for inhibitors. It is not surprising then, that the pharmaceutical industry is making great strides in taking advantage of these pathways therapeutically.

A Brief Note on Alpha-Cell Function

AA and catecholamines are powerful physiological stimulators of glucagon release from alpha cells, and glucose inhibits these actions. The signaling pathways that operate here are not well understood but may involve steps similar to those described for beta cells, including nucleotide-controlled K^+ channels and L-type Ca^{2+} channels. Much controversy exists about the mechanism(s) of glucose inhibition of glucagon secretion. It has not been determined whether the glucose effects are direct or are mediated by the paracrine actions of insulin, GABA, Zn^{2+} , or other signals. The probable reason for this scanty knowledge lies in the great difficulties that hinder metabolic and other functional studies of alpha cells.

Glucokinase Disease and GDH or K^+ Channel-Linked Hypoglycemia

The biochemical genetic characterization of GK-linked hyper- and hypoglycemia syndromes (collectively termed ‘glucokinase disease’) and of GDH or K^+ channel-linked hyperinsulinemic hypoglycemia has strongly reinforced basic scientific concepts about the role of GK as glucose sensor and other crucial steps in GSIR and AASIR. Nearly 200 GK mutations have been discovered in patients with GK disease, usually inherited as an autosomal dominant trait. Activating mutations cause hyperinsulinemic hypoglycemia because the threshold for GSIR is left shifted as expected, while inactivating mutations cause a mild form of diabetes mellitus (DM) in the case of haploinsufficiency due to a right shift of the glucose threshold. However, severe ketotic permanent neonatal DM (PNDM) ensues when both alleles are strongly affected. In the latter case, beta cells seem to be totally nonfunctional,

and the patients require full insulin replacement therapy right after birth. The syndromes of GK disease are nature’s compelling proof of the GK glucose sensor concept. Activating mutations of GDH, usually explained by loss of GTP (ATP) inhibition of the enzyme, cause hyperinsulinemic hypoglycemia because beta-cell glutaminolysis is enhanced. These patients are also leucine hypersensitive, in agreement with the findings of basic research (see model experiment in Figure 3). Inactivating mutations of one of the subunits of the K^+ channel are a frequent cause of autosomal recessive hyperinsulinemic hypoglycemia. These mutations depolarize the beta cells and result in elevated $[Ca^{2+}]_i$, rendering them supersensitive to AA stimulation and neuroendocrine activation. The hyperresponse to protein meals as observed in these patients is reproduced in isolated perfused islets of sulfonyl urea receptor type 1 (SUR-1) $^{-/-}$ mice that are stimulated by a physiological 20 AA mixture, in contrast to control islets that are refractory to this stimulus in the absence of glucose. There is also increasing support for the view that the highly common E23K polymorphism of the Kir6.2 channel causes a predisposition for the development of T2DM. The E23K mutation increases K^{2+} conductivity and hyperpolarizes beta cells, rendering them less sensitive to physiological stimuli. These examples illustrate how basic and clinical sciences cooperate in their quest to understand glucose homeostasis and its disturbances.

Beta-Cell Therapy in T2DM and Hyperinsulinemia

Our knowledge of physiological chemistry and biochemical genetics of the beta cell has its corollary in beta-cell-based therapeutics of T2DM and hyperinsulinemic hypoglycemia, which contributes substantially to our understanding of islet cell function and glucose homeostasis. The following drugs will be discussed in brief: K^+ channel inhibitors and activators (KCI and KCOs), GK activators (GKAs), GLP-1, GIP, methylxanthines, and muscarinic agonists (see Figure 1). KCIs (i.e., sulfonyl urea derivatives and other chemicals that inhibit the K^+ channel) reduce K^+ conductivity, depolarize the beta cell, and sensitize it to all known physiological stimuli. For decades, they have been the drugs of choice for the treatment of T2DM. KCOs (e.g., diazoxide) hyperpolarize beta cells and have been useful in the treatment of mild forms of hyperinsulinemic hypoglycemia (e.g., due to mutations of GK or GDH). GKAs have been discovered only recently and hold considerable promise for a new approach to the treatment of T2DM. They have effects very similar to those of activating mutations of GK, and it seems that they act by binding to an allosteric activator site of the enzyme (Figure 4). They cause a left shift of the concentration–dependency curve of GSIR and have the additional advantage that they activate hepatic GK and augment glycogen synthesis. It is hoped that ongoing trials in humans will demonstrate the clinical usefulness and safety of this novel, theoretically very attractive, approach to treatment of T2DM. GLP-1 and GIP are obvious candidates for antidiabetic agents as is apparent from Figure 1, and promising trials are underway. M3 agonists that could reproduce the effects of acetylcholine, although potentially useful, have not been explored, most likely because lack of beta-cell specificity might become an issue. Phosphodiesterase inhibitors

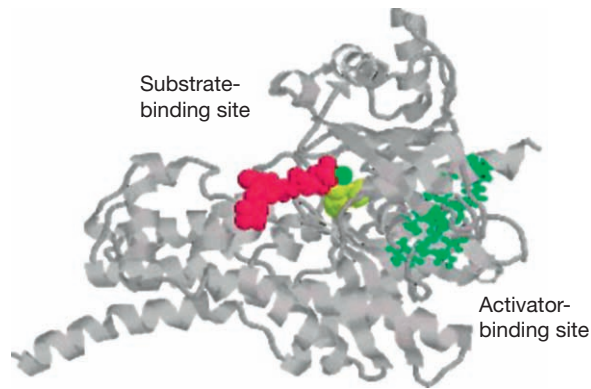


Figure 4 Allosteric activator site of the glucokinase glucose sensor molecule: ribbon drawing of the model structure of GK, with the N-terminal to the left and the C-terminal to the right. Glucose, ATP, and Mg^{2+} are located in the substrate cleft of the enzyme, and the allosteric activator site is indicated by several amino acids. The mutation of any one of these (e.g., V62M, I211A, Y214C, M235A, V455M, and A456V) activates the enzyme markedly. Glucokinase activators bind to this site.

(e.g., 3-isobutyl-1-methylxanthine (IBMX)) are powerful potentiators of fuel-stimulated insulin release, but here again it is apparently difficult to design a beta-cell-specific drug. Note that drug treatment of beta cells seems to reduce the diabetes-associated hyperfunction of the glucagon-producing alpha cells as well, thus indirectly enhancing the benefit of antidiabetic drugs with beta-cell action.

Conclusion

This article illustrates that experimental work and mutually corrective or reinforcing dialog between biochemists, geneticists, pharmaceutical chemists, pharmacologists, and the practicing

diabetologists has resulted in a deep understanding of the role that pancreatic endocrine cells play in glucose homeostasis and the pathogenesis of DM and how various therapies might ameliorate the dysfunction of these cells. This remarkable track record bodes well for the future, holding great promise for even deeper insights and the development of new antidiabetic agents and drugs that combat severe hyperinsulinemia.

See also: **Metabolism Vitamins and Hormones:** Fatty Acid Oxidation; Fatty Acid Structure and Synthesis; Hexokinase/Glucokinase; **Signaling:** Cytokines; Protein Kinase C Family.

Further Reading

- Doliba NM, Qin W, Vatamaniuk MZ, et al. (2004) Restitution of defective glucose-stimulated insulin release of sulfonylurea type 1 receptor knockout mice by acetylcholine. *American Journal of Physiology – Endocrinology and Metabolism* 286(5): E834–E843.
- Grimsby J, Sarabu R, Corbett WL, et al. (2003) Allosteric activators of glucokinase: Potential role in diabetes therapy. *Science* 301: 370–373.
- Li C, Najafi H, Daikhin Y, et al. (2003) Regulation of leucine-stimulated insulin secretion and glutamine metabolism in isolated rat islets. *Journal of Biological Chemistry* 278: 2853–2858.
- Matschinsky FM (1996) Banting Lecture 1995: A lesson in metabolic regulation inspired by the glucokinase glucose sensor paradigm. *Diabetes* 45: 223–241.
- Matschinsky FM (2002) Regulation of pancreatic beta-cell glucokinase: From basics to therapeutics. *Diabetes* 51(supplement 3): 394–404.
- Newgard CB and Matschinsky FM (2001) Substrate control and insulin release. In: Jefferson L and Cherrington A (eds.) *Handbook of Physiology*, vol. 2, pp. 125–151. Oxford: Oxford University Press.
- Prentki M and Matschinsky FM (1987) Ca^{2+} , cAMP and phospholipids derived messengers in coupling mechanisms of insulin secretion. *Physiological Reviews* 67: 1185–1248.

Insulin Mechanisms/Metabolic Actions

E J Gallagher and D LeRoith, Mount Sinai Medical Center, New York, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Anabolism The metabolic process through which cells synthesize complex molecules (e.g., proteins) from more simple molecules (e.g., amino acids), usually requiring the consumption of energy, in the form of adenosine triphosphate (ATP).

Autoimmune A state in which the body produces an immune response directed against its own antigens.

Catabolism The metabolic process in which cells break down complex molecules (e.g., proteins) into more simple molecules, with the release of energy.

Cathepsin An intracellular proteolytic enzyme (protease) involved in cellular degradation.

Gluconeogenesis The cellular production of glucose from noncarbohydrate precursors.

Insulin resistance The condition in which cells do not respond to normal concentrations of insulin, resulting in increased secretion of insulin (hyperinsulinemia) to overcome the insulin resistance.

Lipogenesis The cellular synthesis of fatty acids.

Mitogenesis The process of inducing cell division (mitosis).

Transcription factors Proteins that bind to regulatory regions of DNA and regulate the expression of specific genes.

Introduction

Insulin is synthesized and secreted by the pancreatic β cells in response to a variety of signals from nutrients, hormones, and neurotransmitters. Approximately 30–50 units of insulin are secreted daily by the human pancreas. The basal serum insulin concentration is on average 60 pmol l⁻¹, rarely rising over 610 pmol l⁻¹ in normal individuals after eating. Endogenous insulin has a half-life of 3–5 min. It is an anabolic hormone that regulates glucose, lipid, and protein metabolism as well as growth (Figure 1). It exerts these effects by binding to the insulin receptor (IR) that leads to the phosphorylation of molecules and activation of signaling pathways involved in metabolism and mitogenesis. Type 1 diabetes mellitus (T1DM) is a condition of insulin deficiency, due to autoimmune destruction of the pancreatic β cells. A lack of insulin leads to protein catabolism and lipolysis, which is seen in patients with T1DM before commencement of insulin therapy. Conversely, hyperinsulinemia and insulin resistance are seen in the initial stages of type 2 diabetes mellitus (T2DM). Insulin resistance is associated with abdominal obesity, glucose intolerance, dyslipidemia, and hypertension, which are all components of the metabolic syndrome. In addition, insulin resistance is associated with cardiovascular disease, nonalcoholic fatty liver disease, polycystic ovarian syndrome, and certain forms of cancer.

In this article, we focus on the metabolic effects of insulin on protein, lipid, and glucose metabolism. We describe insulin signaling at a molecular level and how defects in insulin signaling can lead to insulin resistance and T2DM.

Insulin Signaling

The Insulin Receptor

The IR is a transmembrane tyrosine kinase receptor. It is a tetramer with two monomers each of which has an α and β subunit linked by disulfide bonds. Each monomer is encoded by one gene that forms a single-chain proreceptor that is

proteolytically cleaved at a furin cleavage site to form the α and β subunits. The α subunits are located extracellularly and are the location of insulin binding. The β subunits span the cell membrane and, on their intracellular portion, contain tyrosine kinase domains flanked by two regulatory regions containing phosphotyrosine-binding (PTB) sites, through which the insulin-signaling cascade is initiated. In its inactive state, the tyrosine kinase activity of the IR is inhibited by an activation loop that prevents substrates from accessing the catalytic site.

The gene for the IR is located on the short arm of chromosome 19. It contains 22 exons encoding the α and β subunits. Transcription of the gene results in two messenger RNA (mRNA) products, due to alternative splicing of exon 11. The translated product without exon 11 is IR-A and the product containing exon 11 is IR-B. IR-A is more prevalent in leukocytes, the fetus, and certain cancers, while IR-B dominates in the liver; both isoforms are found in skeletal muscle and adipose tissue. Signaling through the IR-A receptor leads to cell growth and proliferation, while signaling through IR-B has more metabolic effects. The IR is similar in structure to the insulin-like growth factor-1 receptor (IGF-1R). Although insulin can bind to either the IR or IGF-1R, its affinity for the IR is approximately 1000-fold greater than for the IGF-1R. Of note, hemireceptors (or hybrid receptors) also exist. These receptors are made up of one α and one β subunit from the IR, bound to an α and β subunit from the IGF-1R. These receptors have a low affinity for insulin (Figure 2).

Binding of insulin to the α subunit of the IR leads to a configurational change in the receptor and autophosphorylation of the β subunit tyrosine kinase. In its inactive state, the Tyr 1162 in the activation loop blocks adenosine triphosphate (ATP) binding. Upon insulin binding to the IR, the conformational change in the β subunit allows ATP to bind, leading to phosphorylation of Tyr 1162, followed by phosphorylation of Tyr 1158, and finally Tyr 1163. Phosphorylation of all three tyrosines causes a configurational change that allows ATP free access to the receptor and exposes various sites for interactions with IR substrates (IRSs) and adaptor proteins.

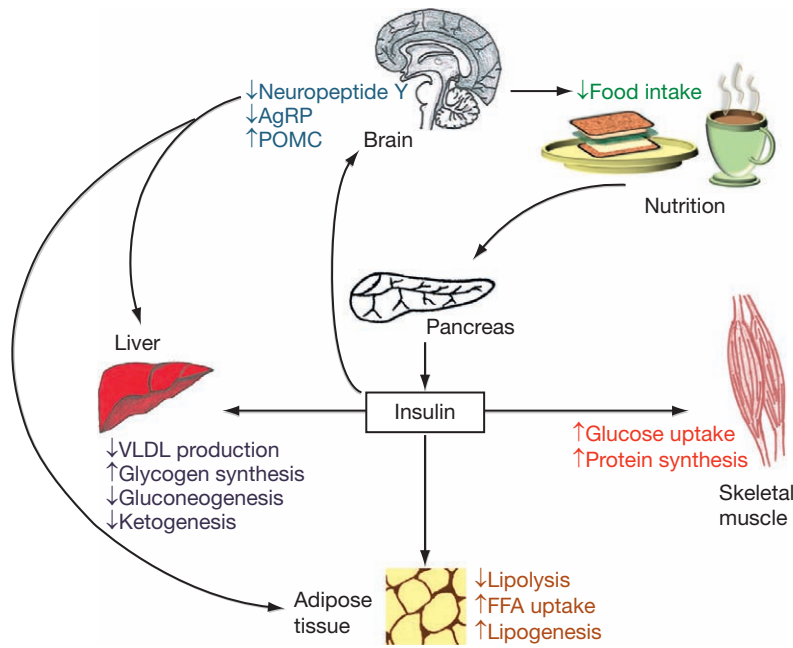


Figure 1 The metabolic effects of insulin acting on the brain, liver, adipose tissue, and skeletal muscle. AgRP, Agouti-related protein; POMC, proopiomelanocortin.

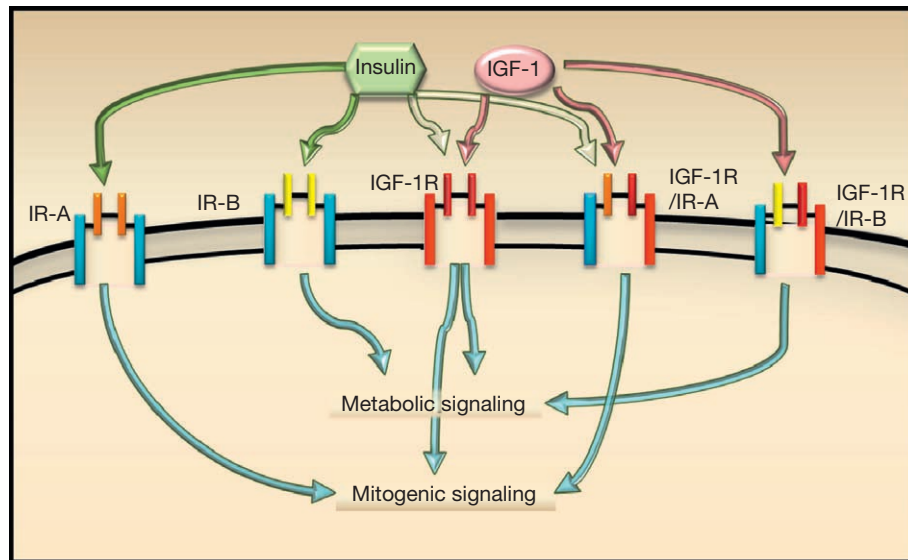


Figure 2 The activation of the insulin receptor splice variants, IGF-1 receptor and insulin/IGF-1 receptor, hybrids by insulin and IGF-1 and the resultant activation of metabolic or mitogenic pathways. Dark green and pink arrows represent strong affinity for the receptor. Pale green arrows represent weak affinity of insulin for receptors.

There are two major pathways that are activated by insulin signaling, the phosphatidylinositol 3-kinase (PI3K) pathway and the mitogen-activated protein kinase (MAPK) pathway; other less well-defined pathways are also activated by the IR tyrosine kinase. The PI3K and Cbl/Cbl-associated protein (CAP) pathways mostly mediate the metabolic effects of insulin, while the MAPK pathway is involved in mitogenesis and growth.

Insulin Signal Transduction

Autophosphorylation of the IR leads to activation of the receptor tyrosine kinase and the phosphorylation of IRS. There are four IRSs, named IRS-1 to IRS-4. The IRSs have a conserved amino terminal region which contains a pleckstrin homology (PH) domain and a PTB domain. These regions interact with the IR. When the IRS proteins are phosphorylated, they attract

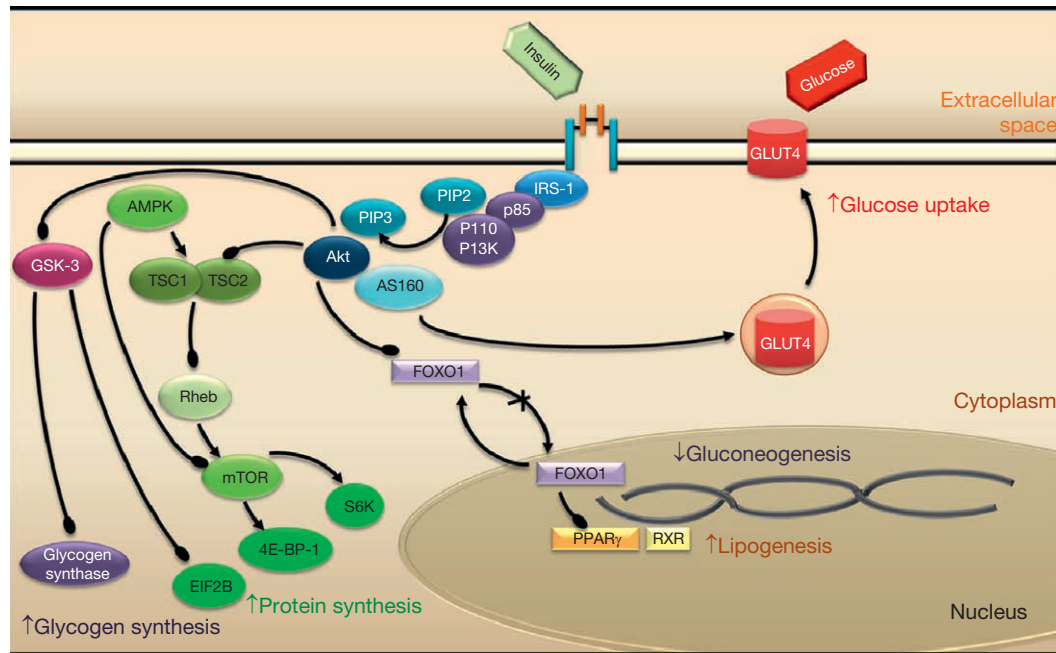


Figure 3 Metabolic signaling pathways activated by insulin binding to the insulin receptor. Arrows represent pathways that are activated. Lines ending in circles represent factors that are inhibited by the preceding component in the pathway.

adaptor proteins that contain Src homology-2 (SH2) domains, including growth factor receptor-bound protein 2 (Grb-2), the p85 subunit of PI3K, the tyrosine phosphatase SHP-2, and Grb-associated protein (Gab1), among others. These adaptor proteins then act as docking sites for further downstream effectors (Figure 3).

PI3K signaling

PI3K is composed of a regulatory subunit and a catalytic subunit. The regulatory subunit is 85 kDa and contains SH2 domains that bind to phosphorylated IRS. The catalytic subunit is 110 kDa. Binding of PI3K to IRS leads to the formation of phosphatidylinositol 4,5-bisphosphate (PIP2) and phosphatidylinositol 3,4,5-triphosphate (PIP3). PIP3 is the major product formed. It activates phosphoinositide-dependent kinase (PDK1), the serine/threonine kinase Akt (also called protein kinase B, PKB), and atypical protein kinase C (PKC).

Effects of Akt phosphorylation

Phosphorylated Akt mediates the metabolic effects of insulin signaling through the PI3K pathway. It achieves this by regulating glucose uptake through translocation of glucose transporter 4 (GLUT4) to the cell membrane of skeletal muscle and adipocytes. In addition, Akt alters hepatic glucose synthesis and adipocyte lipogenesis through its effects on certain transcription factors, such as the forkhead box class O1 (FOXO1) transcription factors, peroxisome proliferator-activated receptor gamma (PPAR γ), and the enzyme glycogen synthase kinase-3 (GSK-3). It affects protein synthesis through its interaction with the mammalian target of rapamycin (mTOR) and inhibits apoptotic proteins such as Bcl-2 and Bcl-X_L.

GLUT4 is found in intracellular vesicles of skeletal muscle and fat cells. Translocation of these vesicles to the cell

membrane not only occurs under the influence of insulin, by activation of the PI3K pathway and the Cbl/CAP pathway, but also may occur by mechanisms independent of insulin. Increased GLUT4 expression and translocation to the cell surface enhances the insulin sensitivity of the cell by increasing glucose disposal. The protein Akt substrate of 160 kDa (AS160) has been identified as a Rab guanine nucleotide exchange factor (GAP). The specific Rabs that are substrates for AS160 have also been associated with GLUT4 containing vesicles. In the absence of insulin signaling, the Rab proteins remain inactive, linked to guanosine diphosphate (GDP). When it is phosphorylated by Akt, the GAP activity of AS160 is inhibited and Rab-GTP complexes form. This has been shown in rodent studies to result in GLUT4 vesicle translocation to the cell membrane, through interaction with insulin-regulated aminopeptidase (Irap).

The transcription factor FOXO1 activates genes involved in hepatic gluconeogenesis and inhibits those involved in adipose tissue lipogenesis. In its dephosphorylated state, FOXO1 can enter the cell nucleus and instigates the transcription of genes encoding enzymes essential for gluconeogenesis, such as phosphoenolpyruvate carboxykinase and glucose-6-phosphatase. Therefore, in the fasting state, FOXO1 allows glucose production by the liver. Insulin signaling leads to the phosphorylation of FOXO1 by Akt, preventing its entry into the nucleus leading to downregulation of gluconeogenic genes. PPAR γ coactivator 1 (PGC1 α) is a coactivator of gluconeogenic genes, along with FOXO1 and is also inhibited by insulin signaling.

Lipogenic gene transcription in adipose tissue may be stimulated by PPAR γ . The function of PPAR γ is affected by ligand binding. In the presence of ligand binding, PPAR γ is an activator of gene transcription, but is a repressor in its absence.

The transactivation activity of PPAR γ results from the formation of a complex between PPAR γ and the nuclear receptor retinoid X receptor (RXR). This complex then binds to PPAR response elements on target genes that lead to lipogenesis. FOXO1 inhibits the activity of PPAR γ by two mechanisms: first, by blocking its promoter and second by preventing binding of the PPAR γ /RXR complex to target genes. Phosphorylation of FOXO1, as a result of insulin signaling, releases its inhibition of PPAR γ and results in the transcription of lipogenic genes. In the absence of ligand binding, PPAR γ may cause transrepression of other genes, by interfering with other transcription factor pathways, such as those related to nuclear factor kappa B (NF- κ B) mediating the anti-inflammatory effects of PPAR γ .

Hepatic and skeletal muscle glycogen synthesis is also affected by Akt phosphorylation. Glycogen synthase kinase-3 (GSK-3) is a constitutively active kinase that is inactivated by insulin signaling. GSK-3 causes phosphorylation of serine residues at the carboxyl terminus of glycogen synthase, inhibiting its ability to convert glucose residues to glycogen. Insulin activation of Akt leads to dephosphorylation and inactivation of GSK-3, preventing phosphorylation of glycogen synthase. In addition, insulin signaling activates protein phosphatase 1 (PP1), which dephosphorylates glycogen synthase. Increased GSK-3 expression and activity in skeletal muscle has been reported in insulin resistance and T2DM. Pharmacological inhibitors of GSK-3 are under investigation as treatment options for T2DM. GSK-3 also regulates protein synthesis by phosphorylating elongation initiation factor 2B (eIF2B). During initiation of translation, eIF2 brings tRNA_m^{Met} (methionine transfer RNA) to the 40S subunit of the ribosome. eIF2 is associated with GTP, which is hydrolyzed to GDP on binding to tRNA_m^{Met} to the start codon. eIF2B is the guanine nuclear exchange factor for eIF2, converting eIF2-GDP to eIF-GTP. Phosphorylation of an amino terminus serine on eIF2B by GSK-3 inhibits the conversion of eIF2-GDP to eIF2-GTP; thereby, GSK-3 prevents the initiation of translation and protein synthesis. By phosphorylating GSK-3, insulin signaling disrupts the inhibitory activity of GSK-3 on protein synthesis.

Insulin-mediated protein synthesis is also mediated by mTOR in the presence of amino acids. mTOR is a serine/threonine kinase that targets the ribosomal protein kinase S6 kinase (S6K) as well as the eukaryotic initiation factor 4E (eIF4E)-binding protein1 (4E-BP1). S6K and 4E-BP1 regulate the initiation of ribosomal translation of specific mRNA populations. mTOR is activated by insulin signaling and phosphorylation of Akt in the presence of nutrients, such as amino acids. Its activity is inhibited by the tuberous sclerosis 1 (TSC1) or hamartin and TSC2 or tuberlin protein complex (TSC1/TSC2). The TSC1/TSC2 complex has GTPase activity toward the small GTPase Rheb, which inhibits its activity. Rheb activates mTOR. Akt phosphorylates TSC2 at Ser939 and Thr1462, leading to loss of the inhibition of the TSC1/TSC2 complex on Rheb and therefore mTOR is activated. Increased mTOR activity allows phosphorylation of S6K and 4E-BP1 and protein synthesis. The absence of cellular nutrition leads to an increase in the adenosine monophosphate (AMP)-to-ATP ratio, with a decrease in protein, lipid, and glycogen synthesis and an upregulation of catabolic pathways with the purpose of generating ATP.

AMP activates the activity of AMP-activated protein kinase (AMPK) by phosphorylation of Thr172 in its catalytic α subunit. AMPK possesses regulatory β and γ subunits in addition to the catalytic subunit; it phosphorylates TSC2 at Thr1227 and Ser1345, sites distinct from those phosphorylated by Akt. Phosphorylation of TSC2 by AMPK increases the GTPase activity of the TSC1/TSC2 complex on Rheb, inhibiting its activity and preventing mTOR activation. AMPK may also phosphorylate mTOR directly at Thr2445, therefore decreasing downstream activity of S6K and 4E-BP1. Activated mTOR and S6K also inhibit IRS-1 and IRS-2 through serine phosphorylation, providing a negative feedback control. Certain tumor suppressor genes such as phosphatase and tensin homolog deleted on chromosome 10 (PTEN) and serine/threonine kinase 11 (STK11) regulate this pathway. Loss of their activity can lead to constitutive activity of mTOR signaling and tumor formation.

Signaling through the PI3K pathway also leads to the phosphorylation of elements involved in the inhibition of apoptosis. Bcl-2 and Bcl-X_L are anti-apoptotic proteins. When they are complexed to Bcl-2/Bcl-X_L antagonist of cell death (BAD), their anti-apoptotic effect is inhibited and apoptosis is promoted. BAD is serine phosphorylated both by PI3K activation of Akt and through the MAPK pathway. Phosphorylation of BAD leads to dissociation from Bcl-2 and Bcl-X_L, thereby inhibiting apoptosis. Therefore, insulin may inhibit cellular apoptosis.

The Cbl/CAP pathway

Glucose uptake has also been reported to be affected by the Cbl/CAP pathway. The IR recruits the adaptor protein CAP, which in turn binds to the oncogene Cbl. After phosphorylation of Cbl, the CBL-CAP complex dissociates from the IR and forms a ternary complex with the caveolar protein flotillin, leading to localization of the CAP/Cbl complex to a lipid raft domain of the plasma membrane. This translocation leads to recruitment of the adaptor protein CrkII, which then forms a complex with the guanyl nucleotide exchange protein C3G. C3G comes into proximity with the G protein, TC10, and catalyzes the exchange of GTP for GDP. This process leads to GLUT4 vesicle translocation to the cell membrane through its activity on actin. The strength of this activated pathway on GLUT-4 translocation compared with the canonical Akt pathway is not well defined.

MAPK signaling

The MAPK pathway is activated by growth factor Grb2 binding to phosphorylated Shc or IRS. Grb2 is bound to the mammalian Son of Sevenless (mSOS), which results in the activation of Ras. Ras (like Rab) is a GTP-binding protein that transmits signals when bound to GTP and is inhibited by GTPase. Activated Ras binds to the inner plasma membrane and the serine/threonine kinase Raf, leading to the phosphorylation of Raf, which in turn leads to activation of the MAP/extracellular signal-regulated kinase (ERK) kinases MEK1 and MEK2, in turn causing phosphorylation of ERK1 and ERK2. ERK1 and ERK 2 phosphorylate transcription factors, including cAMP response element binding protein (CREB), c-Myc, and c-Fos, leading to the growth effects of insulin. In addition to the canonical MAPK pathway, other MAPK pathways such as the

p38 and c-jun N-terminal kinase (JNK) pathways may also be important in insulin's biological effects, but are less well defined.

Regulation of Insulin Signaling

The PI3K and MAPK pathways can interact via their downstream signaling mediators and are also regulated by other molecules, such as cytokines. PI3K and MAPK pathways can activate each other as Akt can activate Raf kinase and Ras can activate PI3K. However, PI3K signaling can also feed back on IRS-1, to inhibit further insulin signaling.

Serine phosphorylation of IRS-1 is an important regulatory step in inhibition of insulin signaling, in normal and pathological circumstances. In rodents, Ser307 is part of the PTB domain of IRS-1 and equates to the Ser312 in humans. Phosphorylation of Ser307 in rodent studies has been seen to prevent tyrosine phosphorylation of IRS-1. This is a mechanism by which mTOR, atypical PKC, GSK-3, tumor necrosis factor- α (TNF α), and JNK inhibit insulin signaling. It is also the method by which PI3K can negatively feed back on IRS-1. IRS-1 may also undergo serine phosphorylation at the SH2 domain, under the influence of Erk1/2, disrupting its activity. Cytokines and the suppressor of cytokine signaling (SOCS) can also increase proteasomal degeneration of IRS-1. IRS-1 activity may also be inhibited by Raf, which displaces IRS-1 bound to PI3K, inhibiting downstream signaling. In addition to their effects on IRS-1, cytokines can interfere with insulin signaling by their action on the IR. SOCS-3 can bind to the IR-binding site for IRS-1, therefore inhibiting its association with the IR, where inhibitory feedback mechanisms may be involved in the development of impaired insulin signaling in insulin resistance.

Metabolic Effects of Insulin

The metabolic effects of insulin are primarily seen in the liver, skeletal muscle, and adipose tissue. In adipose tissue and skeletal muscle, insulin increases glucose uptake through GLUT4, which as described above is under the influence of insulin signaling, primarily through the PI3K pathway. In the liver, the GLUT2 transporter is not under the influence of insulin. Instead, the rate of glucose uptake is determined by the rate of glucose disposal through glycolysis and glycogen synthesis. Insulin promotes both of these pathways by activating the enzymes glucokinase, glycogen phosphorylase, and glycogen synthase. At the same time, gluconeogenesis gene transcription is inhibited by Akt phosphorylation of FOXO1. In skeletal muscle, as in the liver, insulin promotes glycogen synthesis through dephosphorylation of glycogen synthase. In adipose tissue, glucose is used to form glycerol-3-phosphate which is involved in the esterification of free fatty acids (FFAs) to triglycerides (TGs).

Insulin inhibits lipolysis in adipose tissue by dephosphorylating and inactivating hormone-sensitive lipase which is responsible for hydrolysis of stored TG into FFA and glycerol. Dephosphorylation of hormone-sensitive lipase is mediated by protein phosphatase and phosphodiesterase, under the influence of insulin. At the same time, insulin promotes the

activities of endothelial-bound lipoprotein lipase that leads to removal of FFA from circulating very low density lipoprotein (VLDL) and activates lipogenesis, causing FFA to be esterified into TG in the presence of glycerol-3-phosphate. Overall, these effects of insulin in adipose tissue decrease circulating FFA and therefore limit the substrate provided to the liver for VLDL and ketone formation. In addition to decreasing the FFA supply to the liver, ketogenesis is also inhibited by the actions of insulin on hepatic acetyl coenzyme A (CoA) carboxylase. Dephosphorylation of hepatic acetyl CoA carboxylase leads to increased malonyl CoA formation, which inhibits carnitine palmitoyl transferase-1 (CPT-1), the shuttle for long-chain fatty acids across the outer mitochondrial membrane. Inhibition of CPT-1 prevents long-chain fatty acids from being transported into the mitochondria for β oxidation and ketogenesis. Insulin also activates the sterol regulatory element-binding protein (SREBP) 1c which is a transcription factor involved in hepatic FFA and cholesterol synthesis.

In skeletal muscle, insulin signaling in the presence of nutrients promotes protein synthesis through mTOR signaling. Insulin also inhibits protein catabolism by decreasing the activity of cathepsins, proteases that are found in lysosomes and by also stabilizing lysosomal membranes.

Insulin signaling in the brain leads to alterations in energy intake and body weight. This has been demonstrated in rodents and primates, where intracerebral insulin infusions led to decreased food intake and reduced body weight, attributed to insulin signaling in the hypothalamus. IRs in the brain are concentrated in the arcuate nucleus of the hypothalamus which contains neurons that produce neuropeptide Y (NPY), Agouti-related protein (AgRP), and pro-opiomelanocortin (POMC). Activation of the hypothalamic IRs leads to signaling through the PI3K pathway and activation of K_{ATP} channels, resulting in decreased production of NPY and AgRP and increased production of POMC. POMC is a precursor of α -melanocyte-stimulating hormone that binds to melanocortin receptor 3 (MCR3) and melanocortin receptor 4 (MCR4) and causes decreased appetite. NPY activates the Y receptors, while AgRP antagonizes MCR3 and MCR4, both of which stimulate appetite. Cerebral insulin signaling through the PI3K pathway also increases hepatic glucose uptake and decreases hepatic gluconeogenesis. The pathways through which this occurs are undergoing investigation, but it appears to be partly by the autonomic nervous system, in addition to increased interleukin-6, as a result of decreased AgRP, leading to decreased glucose-6-phosphatase expression. Rodent models also suggest that cerebral insulin signaling promotes adipocyte lipogenesis and is the subject of ongoing research (Figure 1).

Insulin Resistance

Insulin resistance is associated with obesity, dyslipidemia, hypertension, and T2DM. Hyperinsulinemia usually precedes the development of insulin resistance. In normal individuals, insulin is released in two phases after ingestion of nutrients. There is a first-phase insulin response which occurs rapidly and a second-phase insulin response which occurs after about 20 min, when the first-phase response disappears. Early in development of insulin resistance and T2DM, loss of the first-phase insulin response is seen, followed by an exaggerated

second-phase response. This subsequently leads to chronic hyperinsulinemia and insulin resistance develops when tissues demonstrate a decreased response to insulin. Insulin resistance results in increased circulating levels of glucose and FFA, dyslipidemia, and increased inflammatory cytokine activity.

Rodent studies of insulin resistance and T2DM have suggested that insulin-signaling abnormalities may occur due to decreased cellular IR content, an increased number of IR/IGF-IR hybrids, and increased serine/threonine phosphorylation of IRS-1. Hyperinsulinemia has also been shown to increase proteasomal degradation of IRS-1.

Insulin resistance in adipose tissue leads to increased lipolysis and FFA production, with loss of inhibition of hormone-sensitive lipase and decreased activity of lipoprotein lipase. Increased FFAs have been seen to cause defects in IRS-1 activation of the PI3K pathway. Inability of the adipose tissue to esterify FFA leads to accumulation of metabolites involved in TG formation in the liver and skeletal muscle. These metabolites include long-chain acyl CoA, diacylglycerol (DAG), and ceramide. The final step in TG formation is the conversion of DAG to triacylglycerol by DAG acyltransferase-1. DAG is a potent activator of PKC θ which impairs insulin signaling in skeletal muscle. Rodent studies have also demonstrated that after lipid infusion, PKC θ activity was associated with decreased tyrosine phosphorylation of IRS-1. Inactivation of PKC θ in mice prevents lipid-induced defects in insulin signaling and glucose transport in skeletal muscle. In studies on the Zucker rat, ceramide has been shown to inhibit insulin signaling and Akt activation; it also inhibits insulin-mediated glucose uptake and glycogen synthesis. Although in humans, lower ceramide is seen in insulin-sensitive individuals, and elevated levels of DAG have been seen in obese diabetic individuals, their exact association with and role in insulin resistance in humans are unclear.

FFAs also increase inflammatory cytokines by signaling through the toll-like receptor 4 (TLR4). TLR4 is a transmembrane receptor with an intracellular domain that is homologous to the interleukin-1 receptor (IL-1R). Signaling through the TLR4 and IL-1R leads to activation of the transcription factor NF- κ B pathway and transcription of proinflammatory cytokines and chemokines, such as JNK and inhibitor of NF- κ B kinase (IKK). Inflammatory cytokines such as TNF α and SOCS proteins can lead to defects in the insulin-signaling pathway by serine phosphorylation of IRS-1 and interference with IR binding to IRS-1. JNK and IKK also cause serine phosphorylation of IRS-1 and contribute to insulin resistance.

Inhibition of insulin signaling leads to increased activation of the transcription factors FOXO1 and SREBP1c, with increased expression of genes for gluconeogenesis and lipogenesis. In the liver, this results in increased glucose and VLDL formation in the presence of elevated FFAs. PPAR γ expression is then inhibited by the activation of FOXO1 and TNF α . These abnormalities

in glucose and lipid formation result in hyperglycemia and elevated FFAs that have been shown to be toxic to the pancreatic β cell and lead to impaired insulin secretion and the development of T2DM.

Conclusions

This article describes the importance of insulin signaling in the regulation of human metabolism. Insulin is a key hormone in integrating and controlling the responses of organs to nutritional intake. We discuss the signaling pathways involved in normal insulin action and the abnormalities that may develop and lead to the development of insulin resistance.

See also: **Bioenergetics:** Phosphatidylinositol-3-Phosphate; **Metabolism Vitamins and Hormones:** Carbohydrate Metabolism in the Central Nervous System; Carnitine and β -Oxidation; Diabetes; Fatty Acid Structure and Synthesis; Gluconeogenesis; Glucose/Sugar Transport in Mammals; Insulin- and Glucagon-Secreting Cells of the Pancreas; **Signaling:** Glycogen Synthase Kinase-3; Mitogen-Activated Protein Kinase Family; Serine/Threonine Phosphatases; Src Family of Protein Tyrosine Kinases.

Further Reading

- Armoni M, Harel C, and Karnieli E (2007) Transcriptional regulation of the GLUT4 gene: From PPAR- γ and FOXO1 to FFA and inflammation. *Trends in Endocrinology and Metabolism* 18(3): 100–107.
- Boura-Halfon S and Zick Y (2009) Phosphorylation of IRS proteins, insulin action and insulin resistance. *American Journal of Physiology – Endocrinology and Metabolism* 296(4): E581–E591.
- Dann SG and Thomas G (2006) The amino acid sensitive TOR pathway from yeast to mammals. *FEBS Letters* 580(12): 2821–2829.
- Grunberger G and Zick Y (2002) *Insulin Signaling. From Cultured Cells to Animal Models*. *Frontiers in Animal Diabetes Research*, vol. 3. New York, NY: Taylor and Francis.
- Ishii N, Matsumura T, Kinoshita H, et al. (2006) AMPK and cell proliferation – AMPK as a therapeutic target for atherosclerosis and cancer. *Journal of Physiology* 574: 63–71.
- LeRoith D, Taylor SI, and Olefsky JM (2004) *Diabetes Mellitus. A Fundamental and Clinical Text*, 3rd edn. Philadelphia, PA: Lippincott Williams and Wilkins.
- LeRoith D and Zick Y (2001) Recent advances in our understanding of insulin action and insulin resistance. *Diabetes Care* 24(3): 588–597.
- Pirola L, Johnston AM, and Van Obberghen E (2004) Modulation of insulin signaling. *Diabetologia* 47: 170–184.
- Porte D, Baskin DG, and Schwartz MW (2005) Insulin signaling in the central nervous system. *Diabetes* 54(5): 1264–1276.
- Saltiel AR and Kahn CR (2001) Insulin signaling and the regulation of glucose and lipid metabolism. *Nature* 414: 799–806.
- Stratford S, Hoehn KL, Liu F, and Summers SA (2004) Regulation of insulin action by ceramide. *Journal of Biological Chemistry* 279: 36608–36615.
- Virkamaki A, Ueki K, and Kahn CR (1999) Protein–protein interaction in insulin signaling and the molecular mechanisms of insulin resistance. *Journal of Clinical Investigation* 103(7): 931–943.

Integrin Signaling

L E Goldfinger, Temple University School of Medicine, Philadelphia, PA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Lawrence E. Goldfinger and Mark H. Ginsberg, volume 2, pp. 441–445, © 2004, Elsevier Inc.

Glossary

Cytoplasmic tail Intracellular portion (20–60 amino acids) of integrins involved in signaling.

Cytoskeleton Structural network of actin microfilaments, keratin intermediate filaments, and tubulin microtubules in mammalian cells.

Extracellular matrix (ECM) Dense array of fibrillar proteins, enzymes, and other molecules that provides a scaffold for cells and tissues, and contains binding sites for integrins.

Focal adhesion Dash-like cell–matrix adhesion site consisting of a cytoplasmic multiprotein complex surrounding a core of clustered integrins.

Integrin ligand ECM protein that binds the extracellular head domain of integrins.

Metal ion-dependent adhesion site (MIDAS) Array of key residues in some integrin head domains which coordinates a divalent cation during ligand binding.

NPxY Conserved sequence in most integrin β -tails that serves as a docking site for signaling molecules.

Phosphotyrosine-binding (PTB) domain Structurally conserved domain in integrin β -tail-binding proteins, which binds the β -tail to regulate signaling.

Integrin Overview

Integrins are transmembrane receptors that promote cellular attachment to the extracellular matrix (ECM) and counter-receptors on the surfaces of neighboring cells. Integrins are responsible for mediating cell adhesion and migration, and for maintaining tissue integrity. They are expressed at the surface of virtually all multicellular animal cells as noncovalent heterodimers of α - and β -subunits. Integrins are named based on their α - and β -subunit composition, for example, $\alpha_1\beta_1$ and $\alpha_{IIb}\beta_3$.

The integrin family in humans consists of 18 α - and 8 β -subunits, each of which is encoded by unique genes. Each α -subunit pairs with a restricted set of β -subunits, resulting in a total of 24 known functional integrin heterodimers. Each pairing, in turn, is expressed on a limited array of cell types. Thus, each cell type has its own particular repertoire of integrins. Since different integrins display distinct ligand-binding preferences, these repertoires specify adhesive preferences of different cell types. In some cases, multiple integrins can recognize the same ligand. For example, many integrins bind to the ECM protein, fibronectin. Even within the same ECM protein, integrins can vary in recognition specificity. Integrin $\alpha_v\beta_3$, for example, binds to an Arg–Gly–Asp (RGD) motif in fibronectin, while integrin $\alpha_4\beta_1$ binds to a Leu–Asp–Val (LDV) motif present in a different part of fibronectin. Furthermore, these short peptide motifs can occur in evolutionarily unrelated proteins, where they can serve as recognition sites for many different integrins. The RGD motif occurs in many proteins and can bind to many integrins (e.g., $\alpha_{IIb}\beta_3$, $\alpha_v\beta_3$, $\alpha_v\beta_5$, and $\alpha_5\beta_1$). This integrin-binding motif interaction is conserved in phylogeny; for example, the *Drosophila* PS2 integrin recognizes an RGD sequence in tiggrrin, a matrix protein. The occurrence of integrin-binding motifs, such as RGD, is an example of convergent evolution dictated by the recognition specificity of adhesion receptors.

All integrin subunits possess a single transmembrane domain (TMD), which separates the large extracellular, N-terminal ligand-binding head domain from the typically short (less than 60 amino acids) C-terminal cytoplasmic tail domain. Besides acting as cellular glue to support adhesion, and as feet to facilitate migration, integrins initiate transmembrane signals, which regulate processes such as embryogenesis, cell development, proliferation, apoptosis, and the immune response.

General Features of Integrin Signaling

Integrins derive their name from their role of integrating the ECM with the insoluble cellular skeleton (cytoskeleton). Binding of ligands to the extracellular portion of integrins leads to both conformational changes in the receptors and to receptor clustering. This combination of events initiates intracellular signals such as protein tyrosine phosphorylation, activation of small GTPases (guanosine triphosphate-hydrolyzing enzymes), and changes in phospholipid biosynthesis. This process is commonly referred to as ‘outside-in’ signaling. In addition, integrins are subject to the so-called ‘inside-out’ signaling. Integrins can fluctuate between three functional states – an active state that manifests high affinity for binding ligands, a high-affinity state that is bound to ligand, or in a distinct inactive state, which has a low affinity for ligands. Inside-out signaling is defined as modulation of the functional state of integrins via intracellular signaling processes. It is frequently mediated by interactions of intracellular components with the integrins’ cytoplasmic tails. Furthermore, signals from one integrin can regulate the activation state of another in a phenomenon referred to as ‘integrin transregulation’ or ‘cross-talk’. Through transregulation, multiple integrins can be linked via a combination of both outside-in and inside-out signaling.

Inside-Out Signaling

Modulation of integrin function can occur by changes in ligand-binding affinity or by changes in the clustering of these receptors. The former mechanism is often referred to as integrin activation and the latter as a change in avidity. These functional changes can be caused by signaling pathways initiated by agonist–receptor interactions at the plasma membrane. The appropriate regulation of integrin activation is required for the formation and maintenance of normal tissue architecture.

Physiological Roles of Inside-Out Signaling

Integrin activation in hematopoietic cells

Integrin activation plays a crucial role in the biology of blood cells. For example, the platelet integrin, $\alpha_{IIb}\beta_3$, is usually maintained in the inactive state, with very low affinity for its extracellular ligands, fibrinogen, fibronectin, and von Willebrand factor. Upon platelet stimulation, the integrins are quickly activated, leading to fibrinogen binding and consequent platelet aggregation, an early step in hemostasis. Thus, the loss of $\alpha_{IIb}\beta_3$ integrin activation impairs hemostasis. Conversely, platelet aggregation is central in arterial thrombosis that is responsible for the majority of heart attacks and strokes. Pharmacological blockade of activation of integrin $\alpha_{IIb}\beta_3$ by agents such as aspirin or ticlopidine is therefore often used to treat patients at risk for these diseases. Similarly, rapid activation of dormant integrins on leukocytes allows these cells to attach firmly to the endothelial vessel wall and resist strong shear forces from the flowing blood. This step is essential for their migration into sites of inflammation. For this reason, loss of leukocyte integrin activation can lead to enhanced susceptibility to infections, and its blockade may be useful in anti-inflammatory therapies.

Integrin activation in tissues

For adherent cells in tissues, such as fibroblasts or epithelial cells, stable cellular attachment to the ECM is necessary for maintaining tissue integrity. Many of the best-studied integrins in these cells, such as the α_5 , α_v , and α_6 integrins, are indispensable components of cell–substrate attachment structures – focal adhesions, hemidesmosomes, and podosomes. In addition to attaching to the ECM, integrins participate in formation and remodeling of the ECM. Activation of tissue integrins can regulate the assembly and structure of the ECM and control the stability of integrin–matrix attachments. In addition, the migration of these cells in tissues is regulated by the ability of integrins to become activated. Furthermore, integrin activation may be topographically localized across the basal cell surface. For example, activated $\alpha_v\beta_3$ integrins are preferentially localized to the leading edge of migrating endothelial cells and during neurite outgrowth.

Integrin Activation

The switch from an inactive to an active state is the result of alterations in the conformation of the integrin extracellular domains. These conformational changes can be triggered by protein–protein interactions at the integrin cytoplasmic face.

These protein–protein interactions are regulated by different signaling pathways in different cells. For example, activation of Raf-1 kinase often suppresses integrin activation, whereas activated R-Ras GTPase can antagonize Raf-1-mediated suppression. However, in many contexts, activation of key intermediates such as the GTPase Rap1, or protein kinase C, promotes increased integrin affinity. In this model, protein kinase C or another upstream intermediate induces the conversion of Rap1 to its active state and its enrichment at the plasma membrane, nucleating an integrin-activating complex. Membrane-tethered Rap1 then binds Rap1-interacting adapter molecule (RIAM1), a scaffolding molecule, which also binds the cytoskeletal linker protein talin. In this fashion, talin is recruited to the plasma membrane where it participates directly in integrin activation (see section [Tail-Binding Proteins](#)). Whatever the upstream events, it is now clear that the final common events involve changes in interactions at the integrin cytoplasmic domain ([Figure 1](#)).

Contribution of cytoplasmic domains: membrane-proximal residues

Deletion of the tail sequences adjacent to the putative transmembrane regions in either the α - or β -tail can activate integrins, indicating that these residues are essential for maintaining integrins in an inactive state. One model for activation entails breaking of a salt bridge in this region that holds the tails together, leading to an opening of the tails via a scissor-like mechanism ([Figure 1](#)). This leads to an allosteric rearrangement in the extracellular domain and a shift to the active state.

Transmembrane domain conformational changes

The allosteric rearrangements of integrin extracellular domains, which are initiated by interactions at the tail domains, are propagated through the integrin TMDs. Thus, the TMDs play critical roles in controlling the activation states. For most integrins in the low-affinity state, the α -subunit TMD is maintained orthogonal to the plane of the plasma membrane, whereas the β -subunit TMD is held at a 25° tilt angle. Electrostatic interactions between these domains, particularly involving conserved

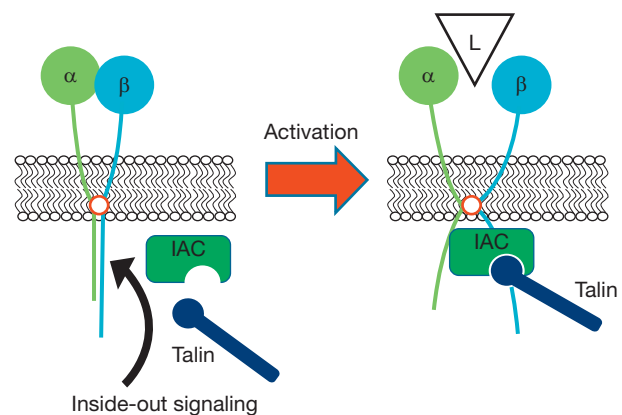


Figure 1 Inside-out signaling activates integrins. Binding of an integrin-activating complex (IAC), which includes talin, to the cytoplasmic tails of an integrin heterodimer produces conformational changes in the extracellular head domains, leading to activation and high-affinity binding to ligand (L).

clasp residues, appear important for keeping integrin head domains in the inactive conformation. Inside-out signaling causes a breaking of the inner-membrane clasps, resulting in shifting of the relative orientations of the TMDs and subsequent allosteric rearrangements of the extracellular domains.

Extracellular domain conformational changes

Precise conformational shifts at the ligand-binding face of integrins result in high-affinity binding to ligands. Integrin–ligand interactions require an essential Asp or Glu residue in the integrin recognition sequence in the ligand, as well as the presence of a divalent cation. Some of the integrin α -subunits have an inserted or I/A domain near the N-terminus; these integrins contact ligand solely through the α -subunit, although the β -subunit plays an accessory role. The I/A domain manifests a metal ion-dependent adhesion site (MIDAS). Key residues in the ligand cooperate with the MIDAS to coordinate the cation and allow integrin–ligand interaction. Conformational changes in the MIDAS during cation coordination and ligand binding subsequently result in secondary changes in the rest of the integrin, including a downward shift of a α -helix C-terminal to the I/A domain, thereby switching the integrin from the closed inactive state to the open active state. For integrins with an α -subunit that does not contain an I/A domain, the I/A-like domain in the β -subunit cooperates with the α -subunit in a somewhat different fashion to mediate ligand interaction, and also through coordination of a divalent cation.

The global structure of an inactive integrin extracellular domain may be a bent conformation, in which the I/A or I/A-like domain and associated head region are held in close proximity to the outer membrane, due to an acute bend in the outer membrane-proximal segments of the integrin subunits. Inside-out signaling appears to induce a straightening of the extracellular domains into an extended structure, which likely represents the intermediate, active state in the absence of bound ligand. The fully extended, open conformation is finally adopted as the active integrin engages the ligand molecule.

Tail-binding proteins

Numerous cytoplasmic integrin-binding proteins have been described, and several have been reported to contribute to activation through binding to the integrin tails. As mentioned above, the best-characterized example is talin, which promotes integrin activation by binding to a subset of integrin β -subunit tails at a highly conserved NPxY motif. Mutational evidence suggests that integrin α – β -subunit tail interactions inhibit activation. The ability of talin to activate the integrins has been attributed to competition with the α -subunit for β -tail binding, whereby talin forms a salt bridge with the β -tail, displacing the α -subunit tail. A segment of the talin head domain binds to the membrane-distal NPxY motif via its phosphotyrosine-binding (PTB) domain, which, combined with a secondary binding step at a membrane-proximal site, allows talin to disrupt tail α – β -subunit interactions to stimulate integrin activation.

Talin binding to the integrin β -tail is both necessary and sufficient for integrin activation. However, in addition to talin, other integrin- and talin-binding proteins contribute to integrin activation in the cellular context. Kindlins are a small family of integrin-binding proteins which interact with β -tails at an NxxY/F motif distinct from the NPxY talin-binding motif.

Kindlin binding to β -tails can enhance talin-mediated integrin activation, and other as yet unidentified proteins may be found to associate with this integrin-activating complex.

Negative regulation

Whereas the talin-centered integrin-activating complex promotes formation and maintenance of the high-affinity state of integrins, competing inside-out mechanisms suppress activation or cause reversion to the inactive state, principally by preventing or disrupting talin–integrin interactions. Integrin-binding proteins such as filamin, and a class of PTB-domain-containing proteins including Numb, Dok-1, and ICAP1 (integrin cytoplasmic domain-associated protein-1) may turn integrins off by binding the β -tail and displacing talin; similarly, phosphatidylinositol phosphate kinase type 1 γ -90 can affect the same outcome by binding to talin. Other mechanisms, such as β -tail phosphorylation by Src kinase, also contribute to negative regulation. Furthermore, endocytic recycling of integrins provides an additional layer of regulation of the concentration of available receptors at the plasma membrane.

Avidity modulation

In addition to affinity modulation through conformational changes in individual integrin molecules, cell adhesion can also be enhanced by increasing the density of ligand-binding sites through integrin clustering, via an affinity-independent mechanism known as ‘avidity modulation’. Many integrin ligands are multivalent; thus, integrin clustering can promote adhesion through cooperative ligand binding. In many cases, a combination of both affinity and avidity modulation contributes to the regulation of cell adhesion.

Outside-In Signaling

Occupancy of integrins leads to conformational changes that can propagate over long distances in the extracellular domain. It has been proposed that these allosteric rearrangements can be transmitted across the plasma membrane, changing the interactions of the α and β cytoplasmic domains with each other and, thus, the integrin’s interactions with intracellular signaling molecules. Furthermore, occupancy leads to integrin clustering, which can also change the physical relationships of the integrin cytoplasmic domains with each other. This combination of occupancy and clustering is believed to initiate signals from integrins. These signals regulate the formation and strengthening of adhesion sites, the dynamics of cytoskeletal structure, cell shape and size, and cell polarity and cell migration. In addition to controlling the physical activities of cells, signaling through integrins also regulates gene expression, cell proliferation, and apoptosis, and many integrins are required for the development of particular lineages.

Integrins in Adhesion Sites

Most integrins are physically linked to the actin cytoskeleton via interactions of the β -tail with actin-binding proteins such as α -actinin, talin, and filamin. The $\alpha_6\beta_4$ integrins associate with the intermediate filament cytoskeleton through other linker proteins – BP180 (BPAG2), BP230 (BPAG1), and plectin. By

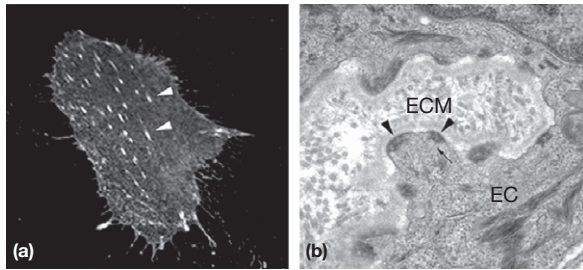


Figure 2 Focal contacts and hemidesmosomes. Two types of integrin–ECM contacts are shown. (a) Focal contacts are visualized by labeling $\alpha_{11b}\beta_3$ integrin clusters in a fibroblast expressing $\alpha_{11b}\beta_3$ attached to fibrinogen. Arrowheads indicate two focal contacts. (b) Transmission electron micrograph of a cross section of bovine tongue. Hemidesmosomes (arrowheads) are seen as electron-dense structures at the basal surface of the tongue epithelial cells (ECs), connecting the intermediate filaments (arrow) to the ECM. Courtesy of Gregory W. deHart and Jonathan C.R. Jones.

connecting the ECM with the cytoskeleton, clustered integrins form the essential core of complexes of structural and signaling molecules, such as focal complexes and adhesions (contacts), hemidesmosomes (specific to epithelial cells, they contain $\alpha_6\beta_4$ integrin and are linked to intermediate filaments), osteoclast and macrophage structures called podosomes (which have many of the same protein components as focal adhesions), and adhesion sites in invadopodia in cancer cells (Figure 2).

Formation of adhesion sites

Ligand binding and integrin clustering nucleate nascent cell adhesion sites. Numerous cytoskeletal adapters and signaling molecules are subsequently recruited to the clustered integrin tails. These complexes serve as the initiation points for signaling by integrins to affect cellular phenotypes. Focal complexes mature to become focal adhesions (focal contacts) as more integrins and signaling molecules associate, thereby strengthening cellular attachments to the ECM.

Regulation of adhesion dynamics

Adhesion sites are dynamic structures which undergo remodeling to fit the needs of the cell or of the region of the cell in contact with the ECM. Focal complexes form at the leading edges of lamellipodia and filopodia in migrating cells, where they mediate new attachments, but do not mature to focal adhesions. Instead, these smaller complexes are subject to rapid turnover, as the cell continues to extend protrusions in the direction of migration. Concomitantly, integrins in focal adhesions at the rear of the cell are recycled to the front via a treadmill-like mechanism. Focal adhesions, hemidesmosomes, and podosomes break apart and reform by severing and re-establishing integrin–ECM contacts, under conditions in which cells are motile, such as during development and in wound healing.

Contributions to Cellular Phenotypes

Connections to intracellular signaling pathways

Although integrins themselves do not manifest any catalytic activity, they transmit signals through binding partners such as integrin-associated protein (IAP, CD47), integrin-linked kinase (ILK), focal adhesion kinase (FAK), Src and Syk kinases, and

members of the tetraspanin transmembrane protein family. Integrin ligation activates members of the Ras family of GTPases, and can also lead to activation of kinases such as Jun-kinase (JNK), p21-associated kinase (PAK), and extracellular signal-regulated kinase (ERK) (microtubule-associated proteins) kinase – the latter two are localized to integrin-rich adhesion sites. These enzymes can subsequently regulate the expression of genes involved in cell proliferation and differentiation. Furthermore, integrins play a role in the inhibition of detachment-induced apoptosis, called ‘anoikis’, through regulation of the PI3-kinase/Akt signaling pathway.

Effects on the cytoskeleton

By connecting the ECM to the intracellular cytoskeleton, integrins also act as mechanosensors, transmitting information about the physical state of the ECM into the cell and altering cytoskeletal dynamics. Integrin ligation can lead to activation of small GTPases of the Rho family, particularly RhoA, Cdc42, and Rac, which promote cytoskeletal rearrangements leading to the formation of actin stress fibers, filopodia, and lamellipodia, respectively. Within focal adhesions, integrins can activate kinases important for downstream signaling, such as FAK and Src tyrosine kinases. Activation of these enzymes results from signals initiated at the integrin cytoplasmic tails. Src interacts directly with the β_3 -tail and is a key player in outside-in signaling. By binding and phosphorylating its proximal substrates such as FAK, p130Cas, Syk, and paxillin, Src acts as a master conduit for outside-in integrin signaling to regulate Rho and Ras GTPases and their effects on the cytoskeleton, adhesion sites, and gene expression. Integrin tails also affect cytoskeletal rearrangements through other mechanisms, for example, activation of Syk tyrosine kinase through its direct binding to the β_3 -tail, which promotes Vav1-induced GTP loading of Rac and subsequent lamellipodial formation. Through such mechanisms, integrins regulate cell polarity, migration, shape and size, and ECM assembly.

Integrin Cross-Talk

Ligand binding to one subtype of integrins can affect the activation state of another integrin subtype on the same cell, by modulating the ligand-binding affinity and/or avidity of that integrin. For example, outside-in signaling through ligand binding of $\alpha_4\beta_1$ integrin can affect $\alpha_1\beta_2$ -dependent lymphocyte adhesion and migration by regulating signaling through $\alpha_1\beta_2$. Similarly, cross-talk signaling between $\alpha_v\beta_3$ and $\alpha_5\beta_1$ integrins in endothelial cells regulates angiogenic responses. Furthermore, integrins are involved in cooperative signaling with other nonintegrin receptors, such as the epidermal growth factor (EGF) receptor and the T-cell receptor complex.

Summary

Integrins have lived up to their name, acting as a unique class of receptors which integrate bi-directional informational cues between the extra- and intracellular environments. On both sides of the plasma membrane, dynamic protein interactions with the integrin head and tail domains combine with mechanical forces to couple outside-in with inside-out

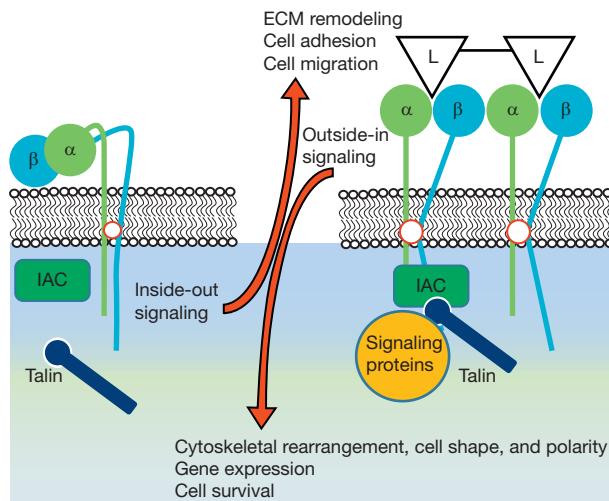


Figure 3 Interplay between outside-in and inside-out integrin signaling. Inside-out signaling converts integrins from a low-affinity to a high-affinity state by interactions of the tail domains with intracellular proteins such as talin. Through activation of the head domains, binding to ligand (L), and receptor clustering, integrins remodel the ECM and dynamically regulate cell adhesion and migration. Conversely, outside-in signaling creates binding sites on the cytoplasmic face of activated integrins to bind proteins and transmit signals to the cell interior, resulting in changes to the cytoskeleton, gene expression, and cell survival.

signaling. Structural rearrangements in the ligand-binding or cytoplasmic faces of integrins are propagated allosterically through shifts in the TMDs. Precise control of the integrin activity and avidity states allows these signaling conduits to regulate cellular and physiological responses in a wide variety of contexts (Figure 3). As the list of integrin-binding proteins

and mechanistic insights grows, the scope of outside-in and inside-out integrin signaling continues to expand.

See also: **Cell Architecture and Function:** Cell Migration; Cell Migration within Three-Dimensional Matrices; Focal Adhesions and Related Integrin Contacts; Intermediate Filament Linker Proteins: Plectin and BPAG1; Rho GTPases and Actin Cytoskeleton Dynamics; **Lipids Carbohydrates Membranes and Membrane Proteins:** Cell–Matrix Interactions; **Protein/Enzyme Structure Function and Degradation:** Allosteric Regulation; **Signaling:** FAK Family; Protein Kinase C Family; Ras Family; Small GTPases; Src Family of Protein Tyrosine Kinases.

Further Reading

- Beckerle MC (2001) *Frontiers in Molecular Biology: Cell Adhesion*. New York, NY: Oxford University Press.
- Gonzalez AM, Bhattacharya R, deHart GW, and Jones JCR (2010) Transdominant regulation of integrin function: Mechanisms of crosstalk. *Cell Signal* 22: 578–583.
- Humphries JD, Byron A, and Humphries MJ (2006) Integrin ligands at a glance. *Journal of Cell Science* 119: 3901–3903.
- Huveneers S and Danen EH (2009) Adhesion signaling – crosstalk between integrins, Src and Rho. *Journal of Cell Science* 122: 1059–1069.
- Hynes RO (2002) Integrins: Bidirectional, allosteric signaling molecules. *Cell* 110: 673–687.
- Legate KR, Wickström SA, and Fässler R (2009) Genetic and cell biological analysis of integrin outside-in signaling. *Genes and Development* 23: 397–418.
- Petrich BG (2009) Talin-dependent integrin signalling *in vivo*. *Thromb Haemost* 101: 1020–1024.
- Shattil SJ, Kim C, and Ginsberg MH (2010) The final steps of integrin activation: The end game. *Nature Reviews Molecular Cell Biology* 11: 288–300.
- Zamir E and Geiger B (2001) Molecular complexity and dynamics of cell–matrix adhesions. *Journal of Cell Science* 114: 3583–3590.

Intercellular Ca^{2+} Waves: Mechanisms of Initiation and Propagation

L Leybaert, Ghent University, Ghent, Belgium

M J Sanderson, University of Massachusetts Medical School, Worcester, MA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Connexon One of six protein subunits that aggregate to form a central channel and one-half of a gap junction (hemi-channel) in the cell membrane. There are a variety of different connexons and their varied structure determines the overall properties of the hemi-channel. A complete gap junction spans across the membrane of two closely adjacent cells by the alignment of two hemi-channels.

Electrotonic communication The passive spread of an electrical charge (without amplification) from the source of the electrical change (e.g., membrane depolarization of a cell) through adjacent cells. The distance and speed of transmission depend on the electrical resistance of the tissue. Open gap junctions between cells have low electrical resistance and facilitate the spread of electrical signals.

Ionotropic receptor A receptor that has ion channel properties and responds to agonist stimulation by the opening of the channel component to induce changes in membrane potential.

Luciferase-based luminometry A bioassay designed to measure adenosine triphosphate (ATP) concentrations based on the numbers of photons released by the enzyme luciferase during the utilization of ATP in chemical reactions.

Metabotropic receptors Receptors that do not directly open ion channels but often respond to agonist stimulation by the formation of a second intracellular messenger.

Paracrine communication The release of an extracellular signaling molecule by one cell in the local vicinity of the target cell that has the appropriate membrane receptors to respond to the signal. Diffusion carries the signal from the releasing cell to the target cell.

Photolysis The use of ultraviolet light to break chemical bonds. The molecule that absorbs the ultraviolet light is often called a 'caged compound' and has no biological action. The ultraviolet light cleaves the molecule into an active compound ('uncaged') and inactive remnant.

Introduction

Intercellular calcium (Ca^{2+}) waves consist of increases in cytoplasmic Ca^{2+} ion concentration ($[\text{Ca}^{2+}]_i$) that are communicated between cells and appear as waves that radially spread out from an initiating trigger cell. These Ca^{2+} waves propagate for variable periods (up to tens of seconds) with speeds of approximately $10\text{--}20\ \mu\text{m s}^{-1}$ and may involve tens to hundreds of contiguous cells. The actual speed and size of the Ca^{2+} wave depend on the nature and strength of the initiating stimulus and the mechanism of propagation. Ca^{2+} waves that only propagate within a single cell are called 'intracellular Ca^{2+} waves'.

The first major reports of intercellular Ca^{2+} waves were made in 1990 which described Ca^{2+} waves propagating either through cultures of astrocytes in response to extracellular glutamate or through airway epithelial cells following mechanical stimulation of a single cell. Subsequently, intercellular Ca^{2+} waves have been found to be initiated by a variety of stimuli in a wide range of cell types, including bone, pancreatic, epithelial, endothelial, and smooth muscle cells.

Intercellular Ca^{2+} waves are complex spatio-temporal events that involve active signal transduction within and between cells and are limited by passive physical characteristics of diffusion and cell architecture. As a result, our understanding of Ca^{2+} waves has arisen from investigations with real-time microscopy, molecular biology, and mathematical modeling. Interestingly, the use of mathematical models to understand inter- and intracellular Ca^{2+} waves has itself driven the development of two- and three-dimensional models that previously

only included a time domain (i.e., mechanisms of Ca^{2+} oscillations). From these studies, it is now recognized that there are three key parameters, namely the mechanisms of wave initiation, wave propagation, and messenger regeneration, which characterize the varying behavior of intercellular Ca^{2+} waves in different cell types.

Ca^{2+} Wave Initiation

The variety of stimuli that can trigger an intercellular Ca^{2+} wave ranges from a focal stimulus of a single cell to a global stimulation of a population of cells with selected agonists. Focal initiation is commonly achieved by mechanical stimulation of a cell with a micropipette. Alternatively, local stimulation may involve electrical stimulation with a microelectrode, the photoactivation of a Ca^{2+} ionophore to transiently increase plasma membrane Ca^{2+} permeability, the photolytic release of inositol 1,4,5-trisphosphate (IP_3) within a cell to elevate Ca^{2+} , or the extracellular iontophoresis of adenosine triphosphate (ATP) for partial P2 receptor activation. Global stimuli usually involve a uniform exposure of the cells to a neurotransmitter (i.e., glutamate) or to a reduced extracellular Ca^{2+} concentration. Mechanical stress may also be applied globally.

In general, stimuli that bring about a local increase in $[\text{Ca}^{2+}]_i$ are often wave-triggering events. However, in early work, it was found that mechanical stimulation in Ca^{2+} -free conditions initiated a propagating intercellular Ca^{2+} wave even though the stimulated cell did not display increases in $[\text{Ca}^{2+}]_i$. This response indicated that a Ca^{2+} increase within the

stimulated cell was not essential to trigger a Ca^{2+} wave and this led to the idea that an alternative messenger was involved. It is now well accepted that this alternative signal is the upstream second-messenger IP_3 ; this messenger serves as a powerful trigger of Ca^{2+} release from the endoplasmic reticulum (ER) and an instigator of many intra- and intercellular Ca^{2+} waves.

Ca^{2+} Wave Propagation

Intercellular Ca^{2+} wave propagation involves two steps: the propagation of the Ca^{2+} wave across a cell, followed by the transmission of the Ca^{2+} wave between cells. While both these processes involve the diffusion of messengers, they are mediated by different mechanisms.

Intracellular Propagation

The mechanism for the propagation of Ca^{2+} waves through cells is strongly dependent on the Ca^{2+} excitability of the cytosol (i.e., the ability with which a small, local transient elevation in $[\text{Ca}^{2+}]_i$ can be amplified into a large spreading Ca^{2+} pulse) and the diffusing messenger. To appreciate these mechanisms, the characteristics and spatial distribution of the Ca^{2+} release channels of the ER must be understood.

Ca^{2+} -permeable receptor/channels occur within the ER membrane of most cells in two forms: the IP_3 receptor (IP_3R) and the ryanodine receptor (RyR – named for its affinity for the plant alkaloid). The open probability of the IP_3R is primarily increased by the binding of IP_3 . Similarly, the open probability of the RyR is increased by the binding of Ca^{2+} . In addition, the RyR may also be sensitized by cyclic adenosine diphosphate-ribose (cADP-ribose). Importantly, the IP_3R and RyR are also functionally similar in that their Ca^{2+} permeability is also regulated via a process known as ' Ca^{2+} -induced Ca^{2+} release' (CICR). CICR is mediated by a positive-feedback mechanism in which the local increase in Ca^{2+} acts on the receptors to further increase their open probability and enhance Ca^{2+} efflux (Figure 1).

Although the IP_3R and RyR are believed to be distributed throughout the cell (in the ER), their distribution may not be uniform and it is believed that these channels are arranged in clusters, most likely, containing both IP_3R and RyR s.

The mechanism of intracellular Ca^{2+} wave propagation that was first experimentally described is based on the diffusion of IP_3 from a restricted stimulation site. This type of Ca^{2+} wave was hypothesized to be initiated by mechanical stimulation and is reproduced by a transient, localized iontophoretic or photolytic release of IP_3 . As expected, inhibition of IP_3Rs abolished these Ca^{2+} waves. Following liberation, IP_3 immediately diffuses in all directions. In doing so, it encounters, binds to, and activates IP_3Rs to release Ca^{2+} from the ER. As mentioned earlier, the IP_3R is also sensitive to Ca^{2+} . Consequently, a self-amplifying increase in $[\text{Ca}^{2+}]_i$, mediated by CICR via the IP_3R , follows the diffusive spread or wave of IP_3 (Figure 1).

Because Ca^{2+} also activates CICR via RyRs , it is likely that Ca^{2+} wave propagation can be enhanced by local increases and diffusion of Ca^{2+} , especially if the IP_3R and RyR are co-localized in receptor clusters. Under special circumstances, such as global agonist stimulation, where the IP_3 concentration is likely to be uniform throughout all cells or where the RyR has been sensitized with cADP-ribose, Ca^{2+} diffusion alone may propagate the Ca^{2+} wave. Thus, the rate of Ca^{2+} wave propagation is a complex function of IP_3 and Ca^{2+} diffusion, sequential CICR via sensitized IP_3Rs and RyRs , and the distribution and composition of these ion/receptor clusters (Figure 1).

Because CICR through the IP_3R and RyR is a self-limiting process, the $[\text{Ca}^{2+}]_i$ increase associated with a Ca^{2+} wave often has a similar amplitude at all locations across the cell. By contrast, the concentration of the underlying diffusive IP_3 wave is believed to decrease with distance from the initiation site (proportional to the inverse distance squared). Unfortunately, unlike the changes in $[\text{Ca}^{2+}]_i$, these changes in $[\text{IP}_3]_i$ cannot be as easily visualized by imaging techniques because of the limitations of reporter dyes for IP_3 . However, it is important to point out that as long as the $[\text{IP}_3]_i$ at the diffusive wave front remains greater than the threshold for activation of the

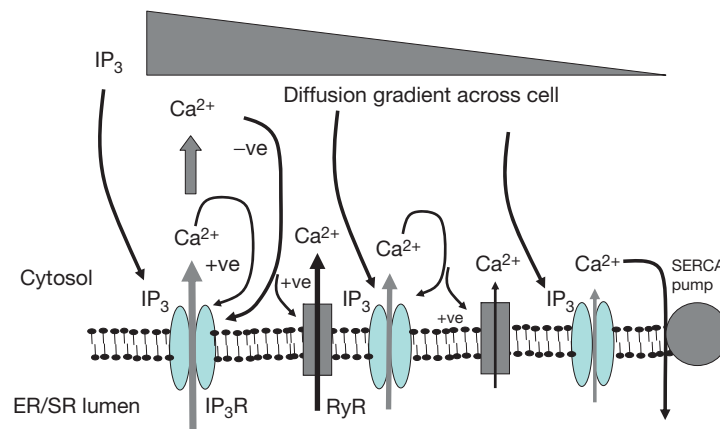


Figure 1 Basic mechanisms of a diffusive Ca^{2+} wave across a cell (either initiating or communicating). Inositol trisphosphate (IP_3) diffuses across the cell and binds to an IP_3 receptor (IP_3R) to stimulate Ca^{2+} release from the endoplasmic reticulum (ER). The positive feedback (+ve) of Ca^{2+} on the IP_3R initiates Ca^{2+} -induced Ca^{2+} release (CICR) through the IP_3R . The elevated Ca^{2+} has a negative feedback (–ve) on the IP_3R to terminate CICR and limit the Ca^{2+} increase. Some Ca^{2+} may diffuse to adjacent ryanodine receptors (RyR) to initiate CICR via these receptors. However, IP_3 diffusion occurs more quickly than Ca^{2+} diffusion (due to cytosol protein buffers) and propagates the Ca^{2+} wave. IP_3 may have been produced in response to cell stimulation or by entering the cell via a gap junction.

IP_3 ($\sim 10\text{--}30\text{ nM}$), Ca^{2+} release will continue to be initiated and a Ca^{2+} wave will appear to propagate without a degradation. In large cells, such as oocytes (diameter $>200\text{ }\mu\text{m}$), the Ca^{2+} waves have a trailing edge of decreasing $[\text{Ca}^{2+}]_i$ which represents the return of the Ca^{2+} to the ER by the action of sarcoplasmic/endoplasmic reticulum Ca^{2+} (SERCA) pumps (Figure 1). However, in small cells (diameter $10\text{--}20\text{ }\mu\text{m}$), Ca^{2+} waves often do not have a trailing edge because the Ca^{2+} release dynamics persist for periods that are longer than the time taken for the Ca^{2+} waves to travel across the cell.

Intercellular Propagation

At the cell border, the cell membrane limits both IP_3 and Ca^{2+} diffusion and hence CICR. Consequently, the transmission of the Ca^{2+} wave to the next cell must occur by a modified or alternative mechanism. There are two major scenarios for this process. The first mechanism, originally characterized, occurs by cell-cell communication via gap junctions (GJs). The second mechanism, subsequently discovered, occurs by paracrine signaling with an extracellular messenger (Figure 2).

GJ communication

GJs are large arrays of transmembrane channels that connect the cytoplasm of two cells and thereby provide a direct diffusion pathway between cells. Cells sharing a GJ each provide a hemi-channel (or connexon) that interact to form the complete GJ channel. Each hemi-channel is composed of proteins called connexins of which there are more than 20 different isoforms (i.e., Cx 23 to 62). By labeling connexin-43 with green fluorescent protein (GFP) to localize GJs, it has been demonstrated that Ca^{2+} waves in transfected HeLa cells propagated across cell borders exactly at the site of the GJ. The types of isoforms forming each GJ determine the properties of the GJ

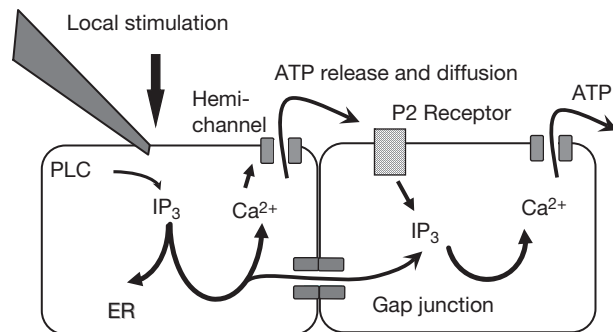


Figure 2 Overall hypothesis for the communication of intercellular Ca^{2+} waves: local stimulation can lead to elevated IP_3 in the initiator cell. As IP_3 diffuses across this cell, it generates an intracellular Ca^{2+} wave by the release of Ca^{2+} through IP_3Rs with amplification from RyRs . Diffusion of IP_3 through a gap junction to an adjacent cell initiates a second intracellular Ca^{2+} wave, via IP_3Rs and RyRs . In addition, or alternatively, the stimulated cell releases ATP (in either a Ca^{2+} -dependent or -independent manner) via hemi-channels in the plasma membrane. The extracellular diffusion of ATP (or similar messenger) to adjacent cells activates P2 receptors which, in turn, stimulate IP_3 production to generate a Ca^{2+} signal in the adjacent cell. Propagation occurs by the regenerative release of ATP. Each mechanism can occur in isolation or synergistically.

and the maximum permeability of a GJ channel, in terms of molecular weight (MW), is approximately $1\text{--}1.5\text{ kDa}$. This has the implication that GJs are readily permeable to Ca^{2+} ions ($\text{AW} = 40$) and IP_3 (MW 486).

Although both IP_3 and Ca^{2+} have the ability to diffuse to adjacent cells via GJs, the relative importance of each messenger in the propagation of intercellular Ca^{2+} waves varies with circumstances. Because Ca^{2+} , but not IP_3 , is strongly buffered by cytoplasmic proteins, Ca^{2+} movement within a cell is much slower than that of IP_3 (effective diffusion coefficients of ~ 13 and $\sim 280\text{ }\mu\text{m}^2\text{ s}^{-1}$, respectively). In addition, the number of moles of Ca^{2+} required to spread over a similar distance is much greater than for IP_3 due to Ca^{2+} attrition associated with the buffers. This implies that after permeation of the GJ, IP_3 is more likely to reach an IP_3R in the connected cell than Ca^{2+} . Because IP_3 stimulates the release of Ca^{2+} from IP_3Rs , IP_3 diffusion can reignite CICR in the adjacent cell to drive the propagation of a second intracellular Ca^{2+} wave (Figure 2). If the diffusion distance from GJ channel to the first cluster of $\text{IP}_3\text{R/RyRs}$ is small, the difference in diffusion time for Ca^{2+} and IP_3 will be less and Ca^{2+} ions by themselves may trigger an intracellular Ca^{2+} wave in the connected cell. This is likely to be the case in the presence of a global agonist and a uniform distribution of IP_3 . The possibility that Ca^{2+} can serve as the primary messenger is also enhanced by increased GJ coupling. This allows greater amounts of Ca^{2+} to reach the adjacent cell, which is necessary to overcome the buffering capacity of the cytoplasm, in order that Ca^{2+} can reach an ER receptor. Examples of this type of communication have been reported in pancreatic acinar cells, hepatocytes, and blowfly salivary gland cells. A combination of the two diffusive processes may also occur in which IP_3 initially arrives at receptor clusters, but at concentrations too low to trigger significant Ca^{2+} release. However, this IP_3 concentration is sufficient to sensitize the IP_3Rs to the following Ca^{2+} that can initiate CICR. These various transmission modes have been reported to occur in different cell types.

A frequently observed characteristic of intercellular Ca^{2+} wave propagation via GJs is a delay ($\sim 0.5\text{--}1\text{ s}$) between the Ca^{2+} wave reaching the cell membrane and the continuation of the wave in the adjacent cell. This delay is believed to represent the diffusion time for IP_3 and/or Ca^{2+} from one cell to the next. Mathematical modeling of wave propagation predicts that there is a minimal GJ permeability required to allow wave transmission between cells and that the GJ permeability required for the propagation of Ca^{2+} waves is inversely related to the effective diffusion constant of the propagating messenger and the Ca^{2+} excitability of the cytosol.

In addition to the diffusion of IP_3 or Ca^{2+} between cells, cADP ribose (MW 541 Da), which can also sensitize RyRs to Ca^{2+} , may mediate GJ-propagated Ca^{2+} waves in lens cells and astrocytes. However, these intercellular Ca^{2+} waves have slower kinetics and propagate smaller distances as compared to Ca^{2+} waves propagated by IP_3 diffusion. This might be expected for waves relying on Ca^{2+} diffusion.

Paracrine communication

Ca^{2+} waves may also be propagated by paracrine signaling between adjacent cells. In this process, a second messenger, released by one cell, diffuses across the extracellular space between cells. After binding to membrane receptors on neighbor

cells, the messenger triggers a Ca^{2+} elevation. Early evidence for this mechanism came from the ability of intercellular Ca^{2+} waves to propagate across a cell-free zone and by the biasing of the propagation direction by an extracellular fluid flow. It has now been demonstrated that the primary extracellular messenger is ATP in many cell types, including hepatocytes, keratinocytes, mast cells, bone cells, and various types of epithelial, endothelial, and glial cells (Figure 2). Indeed, with advanced imaging techniques using bioluminescent probes to detect ATP, intercellular Ca^{2+} waves have been observed to be associated with a cloud of extracellular ATP.

While ATP can activate ionotropic (ligand-gated ion channels, i.e., P2X family) and metabotropic receptors (i.e., P2Y family), Ca^{2+} waves appear to arise most commonly from metabotropic stimulation. This involves the activation of G-protein-coupled receptors, stimulation of phospholipase C, and the generation of IP_3 followed by Ca^{2+} release from the ER via the IP_3R (Figure 2). However, extracellular ATP is prone to rapid degradation by ectonucleotidases with sub-second kinetics, and the resulting metabolites adenosine diphosphate (ADP), adenosine monophosphate (AMP), and adenosine may also contribute to Ca^{2+} wave propagation. The consequence of this scenario is that the characteristics of paracrine Ca^{2+} wave propagation is a complex function of the rates of ATP and metabolite release, the spectrum of the purinergic metabotropic receptors present, and the varying affinities of ATP and its metabolites for these receptors (e.g., ADP is more potent than ATP on P2Y1 receptors).

After the initial reports that intercellular Ca^{2+} waves involved both GJs and extracellular ATP, it was quickly discovered, with the use of transgenic animals and gene silencing techniques, that ATP release was connexin dependent. In order to form GJs, connexins normally associate into hexameric assemblies in the plasma membrane called 'hemi-channels'. However, a fraction of these hemi-channels do not pair up to form the GJ and remain in the membrane. These connexin hemi-channels are normally closed but can open in response to a variety of stimuli, including changes in membrane voltage, phosphorylation, redox state, $[\text{Ca}^{2+}]_i$, metabolism, and hypoxia/ischemia, thereby bringing about ATP release (Figure 2). A similar effect may occur through hemi-channels composed of pannexins, proteins that are structurally similar to connexins but without sequence homology. Because pannexin hemi-channels are externally glycosylated and less likely to form GJs, these channels might be more likely to be involved in ATP release. Pannexin-1 hemi-channels also contribute to the large conductance pore recruited upon long-term exposure of P2X₇ receptors to ATP. P2X₇ receptors are thus equipped with both an ATP-receptor and an ATP-release mechanism.

In addition to this transient flow of ATP through an open channel or pore (diffusive release), it appears that ATP in other non-neuronal cells occurs as discrete packets associated with vesicle fusion with the membrane (vesicular release). However, the vesicles storing ATP do not appear to belong to the classical clear (glutamatergic) or dense-core (noradrenergic) vesicles, but may involve organelles such as lysosomes.

Although ATP appears to be the most common paracrine messenger, some intercellular Ca^{2+} waves are communicated by the release of glutamate and nitric oxide, while others rely on changes in external Ca^{2+} ion concentration. During the

recovery from a cytoplasmic Ca^{2+} transient, Ca^{2+} is pumped both back into the ER by SERCA pumps and out of the cell by plasma membrane Ca^{2+} (PMCA) pumps. Because of the small spaces between cells, this can result in a localized elevation of the extracellular (or interstitial) $[\text{Ca}^{2+}]$ that can stimulate Ca^{2+} -receptors of neighbor cells, thereby triggering phospholipase C (PLC) activation, IP_3 formation, and a Ca^{2+} wave.

Spatial characteristics of GJ and paracrine Ca^{2+} waves

In addition to the observation that GJ-based Ca^{2+} waves typically experience a delay in propagating from one cell to the next, these Ca^{2+} waves frequently follow a convoluted pathway through multiple cells because of heterogeneities in the degree and spatial distribution of cell-cell coupling. By contrast, paracrine-based Ca^{2+} waves propagate in a more direct or homogeneous manner (circular wave front), but may be influenced by heterogeneities in plasma membrane receptor densities. It is not uncommon for individual cells, probably because of a lack of appropriate receptors, to be bypassed and not exhibit a Ca^{2+} wave, a behavior reiterating the ability of paracrine-based Ca^{2+} waves to communicate across cell-free zones.

These different behaviors are also reflected at the subcellular level. In GJ-based Ca^{2+} waves, the intracellular Ca^{2+} waves initiated in the receiving cell always start from a site of cell membrane contact with the donor cell; this represents the location of the GJs and the point of entry of the diffusible messenger (IP_3 , Ca^{2+} , etc). In paracrine-based Ca^{2+} waves, the Ca^{2+} signals in the receiving cell can occur anywhere within the cell because of the random location of the receptor and the subsequent process of signal transduction; frequently, Ca^{2+} signals were first observed in the vicinity of the peri-nuclear ER. Interestingly, despite these differences in propagation pathways, the propagation velocity of GJ- or paracrine-based Ca^{2+} waves is approximately similar and in the range of 10–20 $\mu\text{m s}^{-1}$. This is presumably because the MWs (486, 507, and 292 Da, respectively), and, therefore, the diffusion constants, of IP_3 , ATP, and glutamate are approximately similar.

For clarity, we have addressed the propagation mechanisms of GJ- and paracrine-based Ca^{2+} separately. However, it is important to emphasize, that in many cells, these two mechanisms are active simultaneously and synergistic. Thus, a propagating Ca^{2+} wave can be driven by IP_3 diffusing via GJs, while ATP is released via Ca^{2+} -dependent or -independent mechanisms. Ca^{2+} waves that rely exclusively on IP_3 diffusion via GJs propagate only limited distances, generally up to five cells, although this depends on the amount of IP_3 generated and the degree of GJ coupling. Paracrine Ca^{2+} waves propagate for much greater distances, mainly because they also have a regenerative process regulating the release of ATP (Figure 2).

Passive and Active Waves

Intercellular Ca^{2+} waves that rely on the diffusion of IP_3 via GJs or extracellular ATP from a single initiator cell are sometimes termed 'diffusive Ca^{2+} waves'. They involve a decreasing concentration gradient of messenger from the stimulation point that provokes Ca^{2+} increases (and wave propagation) as long

as the messenger concentration (in either the intra- or extracellular space) is above a threshold level for receptor stimulation. Diffusive Ca^{2+} waves are essentially passive Ca^{2+} waves and analogous to the electrotonic spread of changes in membrane potential between cells. When passive waves are enhanced by messenger (intra- or extracellular) regeneration, the wave propagation becomes active. In principle, pure active waves should never stop propagating. However, in reality, Ca^{2+} waves often rely on a combination of passive and active mechanisms and even active Ca^{2+} waves terminate.

Messenger Regeneration in Active Intercellular Ca^{2+} Waves

Internal Messengers

Because phospholipase C_δ (PLC_δ) is Ca^{2+} dependent, its activation has been implicated in GJ-based Ca^{2+} waves. The influx of IP_3 triggers a Ca^{2+} increase that, in turn, activates further IP_3 synthesis to enhance Ca^{2+} wave velocity and propagation distance by maintaining a sustained and increased $[\text{IP}_3]$ difference between adjacent coupled cells. The extent of the contribution of IP_3 regeneration by PLC_δ is not well defined but inclusion of a limited regenerative process in mathematical models of Ca^{2+} waves often improves the fit to experimental data.

External Messengers

An early observation of paracrine-based Ca^{2+} waves that cross cell-free zones was that the distance between the initiation site and the border of the cell-free zone was not critical for wave propagation on the distal side of the cell-free area. This implied that the concentration of the extracellular messenger at the cell border was dependent on messenger release by cells near the border. In addition to Ca^{2+} -dependent release of ATP or glutamate, as previously discussed, a Ca^{2+} -independent ATP-induced ATP release occurs in astrocytes. It is proposed that ATP binding to P2X_4 or P2X_7 receptors results in the opening of large associated pores that allows diffusive ATP release. This would be expected to result in faster and more extensive Ca^{2+} waves, and mathematical models that include Ca^{2+} -dependent and -independent extracellular messenger confirm this prediction.

Within a uniform cell layer, extracellular messenger regeneration across cells might be expected to generate a homogeneous wave front. However, using luciferase-based luminometry, it was demonstrated that only specific cells within the Ca^{2+} wave (identified by the uptake of connexin hemi-channel-permeable fluorescent dyes) generated a burst of ATP. This behavior suggests that intercellular Ca^{2+} waves may propagate in a saltatory manner; that is, the waves jump between trigger cells, perhaps because only a certain population of cells have the necessary receptors to take part.

An important additional point is that extra/intracellular messenger regeneration is often counterbalanced by degradation and this can limit or slow propagation. For ATP, rapid degradation occurs via ectonucleotidases to various adenosine analogs. For IP_3 , the active concentration can be reduced by de-phosphorylation to IP_2 by a 5-phosphatase or by phosphorylation to IP_4 by a 3-kinase.

Phase Waves

A unique form of Ca^{2+} signaling, which has been characterized in hepatocytes, appears as an intercellular Ca^{2+} wave, but is, in fact, a phase wave – an ordered but independent temporal sequence of Ca^{2+} events. If a row of cells is exposed to conditions that generate a progressive decrease in $[\text{IP}_3]_i$ in each cell, it is possible to get the impression of a cell-to-cell propagating wave because each cell shows a Ca^{2+} spike with a slight delay with respect to the neighboring cells. Radial rows of hepatocytes have a gradient of plasma membrane G-protein-coupled receptors responsive to the hormones noradrenalin or vasopressin. Uniform exposure to these agonists establishes an $[\text{IP}_3]_i$ gradient within the cells. However, this gradient can be modulated by GJ communication. Ca^{2+} phase waves have been observed both *in vitro* and *in vivo* with intact perfused liver preparations.

Intercellular Ca^{2+} Waves in Isolated Tissues and *In Vivo*

Because most of our knowledge of intercellular Ca^{2+} waves is derived from *in vitro* monolayers of cultured cells, it is important to ask if such Ca^{2+} waves exist *in vivo* and what is their function? Similarly, it is important to consider if the stimuli used to evoke waves *in vitro* are relevant (in both strength and form) to *in vivo* conditions.

Intercellular Ca^{2+} waves have been demonstrated in glial cells, more specifically astrocytes and Müller cells in acutely isolated rat retina in response to local mechanical stimulation. Similarly, large intercellular Ca^{2+} waves, in response to mechanical stimulation, have been found in radial glial cells in the neocortex in partial brain preparations from embryonic rats. In mouse cortical brain slices, step increases in Ca^{2+} induced by photolysis of caged Ca^{2+} and electrical stimulation induced intercellular Ca^{2+} waves that propagated for distances greater than 100 μm . Ca^{2+} waves in neocortex slice preparations were found to be dependent on GJs, based on their sensitivity to the GJ blocker carbenoxolone or conditional knockout of connexin 43 expression in astrocytes. By contrast, intercellular Ca^{2+} waves triggered by electrical stimulation in the CA1 region of hippocampal slices are unaffected by carbenoxolone but significantly reduced by suramin, a blocker of P2 receptors, results indicating a paracrine-purinergeric communication. It is also important to emphasize that the properties of intercellular Ca^{2+} waves in nervous tissues and brain slices are influenced by neuronal involvement. When the neuronal contribution is abolished by low extracellular Ca^{2+} conditions (inhibition of presynaptic Ca^{2+} entry) or tetrodotoxin (blockage of voltage-sensitive sodium channels), the propagation speed of and distance traveled by Ca^{2+} waves are comparable to that observed in pure astrocyte cultures. With the neuronal network active, Ca^{2+} waves propagate faster and further, presumably as a result of glutamate leakage from synapses acting to influence the astrocytes. *In vivo* recordings in the brain of anesthetized rodents with two-photon microscopy have revealed spontaneous Ca^{2+} changes in cortical astrocytes, with some degree of coordinated wave-like organization of the signal. Increasing brain activity induced with bicucullin, which induces seizures, in turn, increases the coordination and

wave-like behavior. From these examples, it appears that intercellular Ca^{2+} waves *in vivo* are smaller as compared to those *in vitro*, irrespective of the preparation. One reason for this is the fact that *in vivo* conditions are three-dimensional and thereby contribute to reduced diffusion distance for the intra- or extracellular messengers and increased metabolism of the messenger. In addition, an intact vasculature will also facilitate the clearance of the extracellular messenger from the interstitial space. If messenger clearance is faster than the latency associated with Ca^{2+} signal induction and subsequent intracellular Ca^{2+} wave generation, the intercellular wave propagation will be limited. Despite the fact that intercellular Ca^{2+} waves can be triggered by exogenous stimulation, there is currently little evidence supporting the presence of such waves during normal brain activity. Recent work indicates that intercellular Ca^{2+} waves may occur spontaneously in the hypothalamus and cerebellum, and in the vicinity of amyloid plaques in animal models of Alzheimer's disease.

Other *in vivo* data on intercellular Ca^{2+} waves are available from isolated and blood-perfused lung preparations. Photolysis of caged Ca^{2+} in a single endothelial cell of alveolar capillaries has been reported to trigger Ca^{2+} changes up to 150 μm away from the stimulation site and to propagate from capillaries to venous blood vessels. Spatially propagating Ca^{2+} changes were absent in mouse lungs lacking connexin 43 and inhibited by peptide blockers of GJs. A consequence of the endothelial propagating Ca^{2+} wave was an increase in the P-selectin expression and potentiation of thrombin-induced microvascular permeability.

The final major question is what is the major function of intercellular Ca^{2+} waves? This still remains difficult to answer, but the ubiquitous nature of Ca^{2+} waves in diverse tissues with alternative mechanisms of action underscores the idea that they are important. The least attractive idea would be that Ca^{2+} waves, while exploiting inherent cell Ca^{2+} signaling mechanisms, are simply an excessive response to experimental stimulation. By contrast, activities known to be stimulated by Ca^{2+} increases, such as vesicle secretion, would be expected to be enhanced by a Ca^{2+} wave. This might be the case in blowfly salivary glands following local neural stimuli to salivate. However, a strong correlation between cell activity and Ca^{2+} waves has only been reported for airway epithelial cells. The passage of a Ca^{2+} wave initiated an increase in ciliary activity in adjacent cells. Because cilia are more effective when beating in coordinated groups, this example readily supports the idea that Ca^{2+} waves serve to coordinate tissue, rather than cell, responses. These ideas are supported by data pointing to a role of Ca^{2+} waves in coordinating the metabolic activity of glial cells and also apply for the communication between different cell types such as endothelial–glial cells at the blood–brain barrier, glial–neuron interactions in the brain, or endothelial–smooth cells in the vasculature. An alternative idea, in view of the fact that mechanical stimuli often initiate prominent Ca^{2+} waves, is that

this is a response to local injury, with the propagating waves serving to stimulate the cells near the injury site to initiate a healing process. The Ca^{2+} waves would provide a direction clue as the injury site. Other intriguing possibilities are that Ca^{2+} waves are associated with spreading depression of neural activity in the brain or the infarcts associated with ischemic events.

In conclusion, intercellular Ca^{2+} waves form a common dynamic signaling event expressed by diverse cell types in response to a variety of stimuli. Moreover, the multiple mechanisms used to propagate intercellular Ca^{2+} waves that exploit specific signaling and structural architectures of different cells suggest an essential redundancy that is common to vital activities. Consequently, Ca^{2+} waves are likely to be important in coordinating cell functions in multicellular tissues, an idea that will be explored by our increasing desire and ability to understand intact systems.

Further Reading

- Charles A (2005) Reaching out beyond the synapse: Glial intercellular waves coordinate metabolism. *Science's Signal Transduction Knowledge Environment* 2005: pe6.
- Cotrino ML, Lin JH, Lopez-Garcia JC, Naus CC, and Nedergaard M (2000) ATP-mediated glia signaling. *Journal of Neuroscience* 20: 2835–2844.
- Dupont G, Combettes L, and Leybaert L (2007) Calcium dynamics: Spatio-temporal organization from the subcellular to the organ level. *International Review of Cytology* 261: 193–245.
- Gaspers LD and Thomas AP (2005) Calcium signaling in liver. *Cell Calcium* 38: 329–342.
- Haas B, Schipke CG, Peters O, et al. (2006) Activity-dependent ATP-waves in the mouse neocortex are independent from astrocytic calcium waves. *Cerebral Cortex* 16: 237–246.
- Hirase H, Qian L, Bartho P, and Buzsaki G (2004) Calcium dynamics of cortical astrocytic networks *in vivo*. *PLoS Biology* 2: E96.
- Hofer T, Venance L, and Giaume C (2002) Control and plasticity of intercellular calcium waves in astrocytes: A modeling approach. *Journal of Neuroscience* 22: 4850–4859.
- Iacobas DA, Suadicani SO, Spray DC, and Scemes E (2006) A stochastic two-dimensional model of intercellular Ca^{2+} wave spread in glia. *Biophysical Journal* 90: 24–41.
- Paemeleire K, Martin PE, Coleman SL, et al. (2000) Intercellular calcium waves in HeLa cells expressing GFP-labeled connexin 43, 32, or 26. *Molecular Biology of the Cell* 11: 1815–1827.
- Sanderson MJ, Charles AC, Boitano S, and Dirksen ER (1994) Mechanisms and function of intercellular calcium signaling. *Molecular and Cellular Endocrinology* 98: 173–187.
- Scemes E and Giaume C (2006) Astrocyte calcium waves: What they are and what they do. *Glia* 54: 716–725.
- Sneyd J, Wetton BT, Charles AC, and Sanderson MJ (1995) Intercellular calcium waves mediated by diffusion of inositol trisphosphate: A two-dimensional model. *American Journal of Physiology* 268: C1537–C1545.
- Suadicani SO, Flores CE, Urban-Maldonado M, Beelitz M, and Scemes E (2004) Gap junction channels coordinate the propagation of intercellular Ca^{2+} signals generated by P2Y receptor activation. *Glia* 48: 217–229.
- Webb SE and Miller AL (2007) Ca^{2+} signalling and early embryonic patterning during zebrafish development. *Clinical and Experimental Pharmacology and Physiology* 34: 897–904.
- Zimmermann B and Walz B (1997) Serotonin-induced intercellular calcium waves in salivary glands of the blowfly *Calliphora erythrocephala*. *Journal of Physiology (London)* 500: 17–28.

Interferon Receptors

C P Elco and G C Sen, Cleveland Clinic Foundation, Cleveland, OH, USA; Case Western Reserve University, Cleveland, OH, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 446–451, © 2004, Elsevier Inc.

Glossary

Cytokine receptor A cell-surface protein that binds to a distinct extracellular polypeptide hormone and induces a response within the cell.

Interferon Cytokines involved in regulating antiviral and immune responses.

Kinase An enzyme that phosphorylates (adds a phosphate group to) a protein.

Signal transduction The conversion of an extracellular signal to an intracellular response.

Transphosphorylation The reaction by which one protein causes the addition of a phosphate group to another protein.

Interferons (IFNs) are a family of cytokines, or protein hormones, which modulate the immune response and provide resistance to viral infection. The effects of IFN are mediated through cell-surface receptors, which recognize extracellular IFN and activate cellular signaling pathways, ultimately leading to gene induction and repression. IFNs are divided into two classes, type-I and type-II IFNs. There are multiple members of the type-I IFN family, including IFN- α and IFN- β , whereas IFN- γ is the only known type-II IFN. The two types of IFNs bind to two distinct receptors and have very different cellular outcomes. Type-I IFN receptors predominantly activate an antiviral, innate immune response. Signaling through type-II IFN receptors, on the other hand, is involved in modulating the adaptive immune response. Both types of receptors share, in common, the ability to, directly and quickly, induce new gene transcription through the activation of JAK/STAT signaling pathways.

IFN receptors are classified as type-II cytokine receptors. Both type-I and type-II cytokine receptors have extracellular and intracellular domains connected by a single transmembrane domain. The extracellular domains of cytokine receptors contain a conserved tandem fibronectin type-III (FNIII) motif, which is distinguished by unique cystine pairings at both ends of the domain. Type-II cytokine receptors lack an additional conserved Trp-Ser-X-Trp-Ser motif, found in the extracellular domain of type-I receptors. Ligand binding to the receptor leads to receptor oligomerization and activation of a signaling cascade. In addition to type-I and type-II IFN receptors, the receptors for tissue factor and the cytokine IL-10 belong to this class. The IFN receptors and the IL-10 receptor are composed of two different subunits, both of which are required for proper functioning of the receptors.

IFN- γ Receptors

Expression of IFN- γ and Its Receptor

Expression of the IFN- γ receptor is common to almost all cell types with the exception of mature red blood cells. Two genes, IFN- γ R1 and IFN- γ R2, encode the receptor components.

The human IFN- γ R1 gene is located on the long arm of chromosome 6. The expression of IFN- γ R1 messenger RNA

(mRNA) is not regulated by any known stimuli, but appears to be under standard housekeeping controls. Although the IFN- γ receptor functions on the cell surface, a large intracellular reserve of IFN- γ R1 is present in some cells at levels up to 2–4 times the amount of receptor present on the cell surface. This reserve serves to keep the surface level of IFN- γ R1 constant.

The human gene for IFN- γ R2 is located on chromosome 21. The mouse gene for IFN- γ R2 is on chromosome 16. Study of the mouse gene promoter has revealed that it contains binding sites for a number of activated transcription factors, suggesting that unlike IFN- γ R1, IFN- γ R2 expression can be regulated. Both proteins must be expressed on the cell surface and associate in order for IFN- γ signaling to occur, but there is evidence that some T cells control their responsiveness to IFN- γ by modulating the cell-surface expression of IFN- γ R2.

Unlike the receptors, IFN- γ is expressed only in response to stimulus by antigen and only in cells of lymphoid origin, in particular, natural killer cells, macrophages, and some T cells. Once synthesized, IFN- γ is secreted and signals as a homodimer.

Receptor Structure

Both IFN- γ R1 and IFN- γ R2 have extracellular and intracellular domains, along with a single transmembrane domain, common to all type-II cytokine receptors. The sequences of extracellular domains can vary greatly among species causing highly species-specific IFN- γ binding. By contrast, intracellular domains of the IFN- γ receptor subunits are more conserved among species, and can be substituted freely. The intracellular domains mainly serve as a scaffolding structure for the interaction of downstream-signaling proteins and contain no enzymatic activity of their own. The IFN- γ R1 is a 90-kDa glycoprotein consisting of 472 amino acid residues. The first 228 residues of the protein comprise the extracellular domain. IFN- γ R1 has a 24-amino-acid transmembrane domain and an intracellular domain that spans the C terminus from residues 252 to 472. IFN- γ R2 is a 62-kDa, 315-amino-acid residue protein. Both the extracellular and transmembrane domains of IFN- γ R2 are comparable in length to those of IFN- γ R1. Residues 1–226 form the extracellular domain and 227–250 are the transmembrane domain. At only 65 amino

acid residues, the intracellular domain of IFN- γ R2 is much smaller than that of IFN- γ R1.

The crystal structure of a soluble, truncated version of the extracellular domain of human IFN- γ R1 is known. It contains two distinct domains, both consisting of two β -sheets and displaying fibrillin-3 topology. An 11-amino-acid linker connects the two domains. The binding of ligand is accomplished through the interaction of both domains.

Ligand Binding

IFN- γ binds to the extracellular domain of IFN- γ R1 only. In the absence of IFN- γ R1, IFN- γ R2 cannot bind IFN- γ . However, the presence of IFN- γ R2 does strengthen the association of IFN- γ and IFN- γ R1 threefold, though the two subunits of the receptor do not associate in the absence of IFN- γ . Further, IFN- γ R2 is vital for IFN- γ signal transduction, and for this a highly specific interaction must occur between the extracellular domains of IFN- γ R1 and IFN- γ R2.

An active signaling complex consists of an IFN- γ homodimer and two chains each of IFN- γ R1 and IFN- γ R2. Each subunit of the IFN- γ homodimer binds an IFN- γ R1 chain, which in turn binds IFN- γ R2. In the absence of IFN- γ R2, IFN- γ R1 molecules bound to an IFN- γ homodimer are separated by over 27 Å and, therefore, are unable to interact with each other.

Signal Transduction

Janus tyrosine kinase (Jak)/signal transducer and activator of transcription (STAT) signaling is carried out through a series of tyrosine phosphorylation events, at the receptor, mediated by members of the Jak family. Signaling culminates with the activation and nuclear translocation of STAT proteins, where they function to induce new mRNA synthesis. The IFN receptors serve as docking sites for both Jak1 and Jak2. With the formation of IFN-receptor complex, Jak proteins are brought into close proximity, which enables them to transphosphorylate and thereby activate each other. Independent of IFN binding to the extracellular domain of the receptor, the Jak proteins, Jak1 and Jak2, associate with the intracellular domains of IFN- γ R1 and IFN- γ R2, respectively (Figure 1(a)). Upon the formation of a receptor complex with IFN- γ , Jak2 transphosphorylates Jak1 (Figure 1(b)). Jak1 then transphosphorylates Jak2, and either Jak phosphorylates IFN- γ R1 at Tyr440 (Figure 1(c)). Phosphorylation of IFN- γ R1 in turn leads to the recruitment and binding of a STAT1 protein to each IFN- γ R1 chain. The bound STAT1 proteins are phosphorylated by the Jaks, and they disassociate from the receptor and form a homodimer (Figure 1(d)). This complex then translocates to the nucleus where it can induce the transcription of many genes.

STAT-dependent signaling is the best-studied IFN- γ signaling pathway, but there is evidence that STAT1-independent

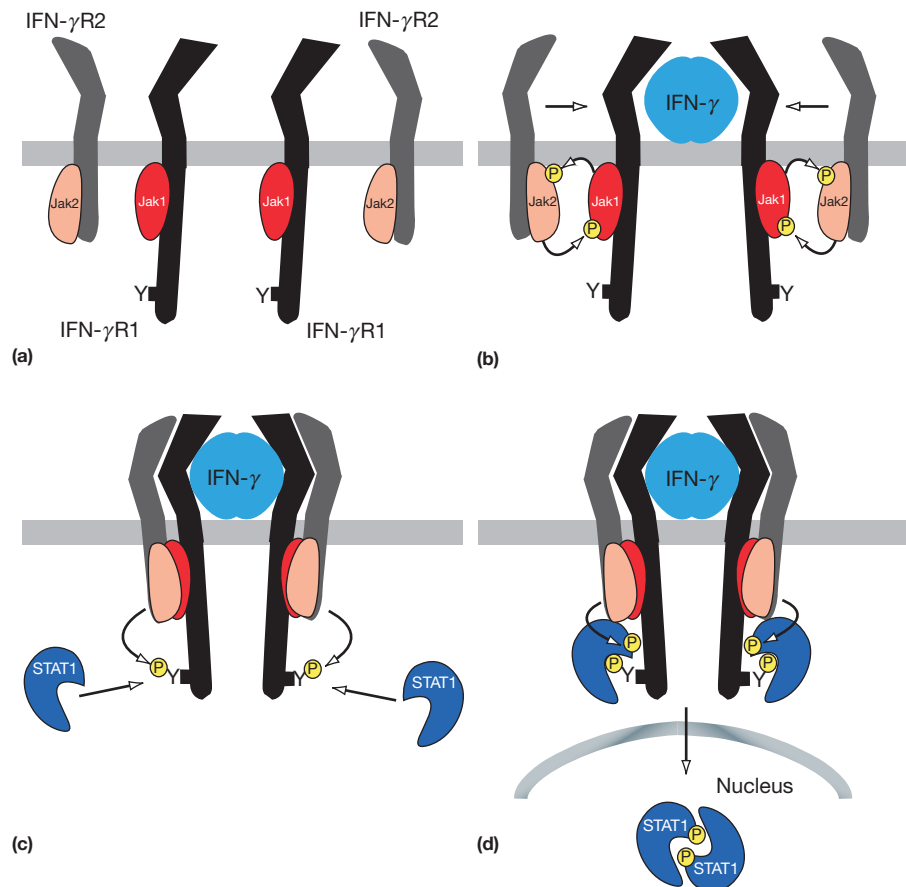


Figure 1 Kinetics of the IFN- γ signaling cascade (a–d), showing the proteins and tyrosine phosphorylation events involved in signal transduction by the IFN- γ receptor.

signaling pathways can also be activated by IFN- γ through IFN- γ R1 and IFN- γ R2. In cell lines lacking STAT1, activation of the IFN- γ receptor by IFN still causes the induction of a set of genes.

In order for signal transduction to occur normally, two regions of the intracellular domain of IFN- γ R1 and one region of the intracellular domain of IFN- γ R2 are required. The association of Jak1 with IFN- γ R1 requires the amino acid residues from 266 to 269, LPKS, located in the membrane proximal region. Distal to the membrane at amino acids 440–444, YDKPH comprise the STAT1-binding site. Within the intracellular domain of IFN- γ R2, a proline-rich 12-amino-acid sequence P²⁶³PSIPLQIEEYL²⁷⁴ is needed for Jak2 binding. Disruption of any of these three regions abolishes normal signaling.

Upon the binding of IFN- γ , the receptor complex is internalized. Internalized receptor–ligand complex is trafficked to an acidic compartment. IFN- γ disassociates from the receptor, goes to the lysosome, and is degraded, whereas the IFN- γ R1 subunit of the receptor is recycled into the cytoplasmic reserve. Mutation of either Leu 270 or Ile 271 inhibits the internalization of the receptor, leading to the accumulation of IFN- γ ligand–receptor complexes on the surface of the cell.

Effects of Signaling through the IFN- γ Receptor

IFN- γ can be thought of as the antimicrobial interferon. Signaling through type-II IFN receptors has pro-inflammatory effects and antibacterial effects. Mice lacking type-II IFN receptors have a much higher susceptibility to infection by bacteria and intracellular microbes. The antibacterial effects of IFN- γ are mediated through multiple pathways. IFN- γ leads to the activation of neutrophils, natural killer cells, and, in particular, macrophages, by increasing their ability to recognize, kill, and digest foreign material or microbes. In T cells, IFN- γ promotes the maturation of cytotoxic CD8 + T cells, as well as the T_H1 subset of CD4 + T cells. IFN- γ also enhances the humoral immune response by upregulating genes such as class-I and class-II major histocompatibility complex (MHC) molecules, along with cell-surface adhesion molecules. Together, these effects serve to augment the adaptive immune response to pathogens.

As would be expected, loss of the IFN- γ receptor genes leads to increased susceptibility to infection by microbes. Mice lacking the IFN- γ receptor genes are affected much more severely by infections with lysteria and mycobacteria. Otherwise, the receptor knockout mice display no developmental defects, have normal T cell responses, and are no more susceptible to most viral infections than wild-type mice. These effects are seen in humans as well. Children susceptible to infections with weakly pathogenic mycobacteria species, such as *M. avium*, were found to have mutations in the IFN- γ receptor genes. The susceptibility was mycobacteria specific, as these children did not have increased difficulties fighting off infections from other bacteria, viruses, or fungi.

An unexpected consequence of IFN- γ receptor loss is an increased susceptibility to tumor development. IFN- γ receptor knockout mice develop tumors at a much higher frequency than their IFN- γ responsive counterparts do. Compared to IFN- γ responsive tumors, IFN- γ unresponsive tumors are less prone to rejection when transplanted into immunocompetent wild-type

mice. Studies of human carcinomas have revealed that loss of responsiveness to IFN- γ is not uncommon. Taken together, these facts suggest that loss of IFN- γ signaling enables tumors to evade the normal surveillance mechanisms of the immune system. Normal expression of the IFN- γ receptor proteins is therefore important in preventing the development of cancer.

The IFN- α/β Receptor

Expression of IFN- α/β and Their Receptor

Like the IFN- γ receptor, the IFN- α/β receptor is composed of two polypeptides encoded by two genes. Both IFN- α/β receptor genes, IFN- α R1 and IFN- α R2, are clustered on human chromosome 21 (or chromosome 16 in mice) with IFN- γ R2 and IL-10R2 (also a member of the class-II cytokine family). Expression of the receptor proteins is thought to be ubiquitous, but it can also be modulated.

The gene encoding IFN- α R1 is 32 kb and contains 11 exons. The second through ninth exons encode the extracellular domain of the protein, while the last two exons encode the transmembrane and intracellular domains.

In the case of IFN- α R2, there are three splice variants of the protein encoded by a nine-exon gene. The 515 aa, IFN- α R2c, encoded by a 4.5-kb mRNA, is the only functional splice variant and has an extended intracellular domain compared to the other two variants. IFN- α R2b has a shorter cytoplasmic domain and cannot signal through the Jak–STAT pathway, and IFN- α R2a is a secreted form of the extracellular domain of the receptor only. Studies in mice, which express only homologs of the 2a and 2c splice variants, have shown that the 2a variant is expressed in greater abundance in most tissues, except in those of lymphoid origin where the two isoforms are expressed in a 1:1 ratio. It is not clear if the truncated splice variants play a functional role in modulating responsiveness to IFN; however, IFN- α R2a can act as a dominant-negative inhibitor for type-I IFN signaling.

There are many members of the type-I IFN family. The most well known among them are IFN- α and IFN- β , which share little structural homology with each other. IFN- β was first isolated from fibroblasts, while IFN- α was discovered in lymphoid cells. Within the IFN- α family alone, there are over 13 subtypes encoded by different genes. The proteins they synthesize can vary in size from 15 to 21 kDa, and contain different degrees of posttranslational modification. Despite the apparent diversity in type-I IFNs, they all bind to the same receptor composed of IFN- α R1 and IFN- α R2c. There is, however, variation in the type and degree of response generated.

Receptor Structure

The extracellular domain of IFN- α R1 spans the first 409 aa residues of the protein, and the intracellular domain contains the final 100 residues from amino acid 431 to 530. IFN- α R1 is unique among class-II cytokine receptors in that its extracellular domain contains four tandem FNIII domains, compared to the two contained by other IFN receptor proteins. When compared to IFN- α R1, the 217-amino-acid extracellular domain of IFN- α R2 is relatively short. By contrast, the intracellular

domain of IFN- α R2c is much longer than that of IFN- α R1, covering the C-terminal 251 amino acids. There is not nearly as much known about the structure of IFN- α/β receptors as that of IFN- γ receptors.

Ligand Binding

Both IFN- α R1 and IFN- α R2c bind type-I IFNs, cooperatively. This is different from the IFN- γ receptor where only IFN- γ R1 had ligand-binding activity. In addition, the signaling complex for the α/β -receptor consists of a single pair of receptor proteins as opposed to the two pairs that compose the γ -receptor complex.

Another feature of the IFN- α/β receptors is their ability to bind multiple ligands. Unlike IFN- γ receptors, which only bind

IFN- γ , the α/β -receptors can bind all subspecies of IFN- α , along with IFN- β and other type-I IFN species. Ligand-receptor interactions for the different classes of IFN are unique. By mutating residues in the extracellular portion of either receptor protein, it is possible to generate mutants, which bind and respond to IFN- β , but not all IFN- α species. The exact structural mechanism for how the differences in ligand binding translate into differences in downstream signaling effects is still unknown.

Signal Transduction

Signaling again occurs principally through the Jak-STAT pathway, but involves different members of the two families of proteins. Independent of ligand binding, Tyk2 a member

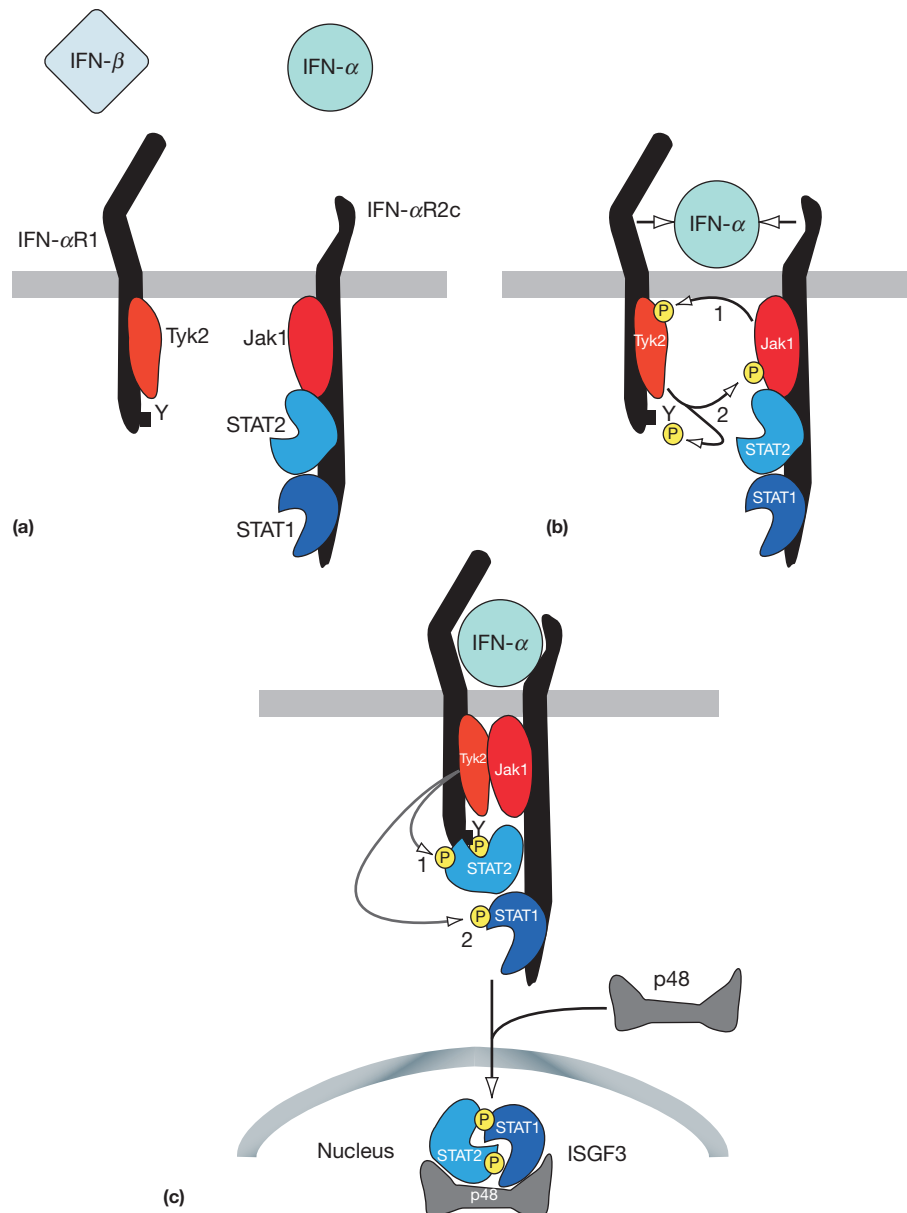


Figure 2 Signal transduction through the IFN- α/β receptor, showing the protein-protein associations and tyrosine phosphorylation steps involved. Numbers indicate the order in which tyrosine phosphorylation events (P) occur.

of the Jak family associates with IFN- α R1, while Jak1, STAT1, and STAT2 associate with IFN- α R2c (**Figure 2(a)**). The binding of ligand brings both receptor subunits into close proximity with each other. This allows Jak1 to transphosphorylate Tyk2 (**Figure 2(b)**). Tyk2 in turn phosphorylates Jak1 along with Tyr466 of IFN- α R1. The phosphorylation of Tyr466 of IFN- α R1 allows STAT2 to bind to it instead of IFN- α R2c. The IFN- α R1-bound STAT2 is phosphorylated on Tyr690 (**Figure 2(c)**). This leads to the STAT2-dependent phosphorylation of STAT1 on Tyr701. STAT1 and STAT2 dissociate from the receptor and form a heterodimer. The STAT1–STAT2 dimer associates with p48, or IRF9, a member of the interferon regulatory factor family, to form the transcription complex interferon-stimulated gene factor 3 (ISGF3). ISGF3 translocates to the nucleus, binds to interferon-stimulated regulatory element (ISRE) sequences present in the promoters of IFN-stimulated genes (ISGs) and induces transcription.

The binding sites for Jak and STAT proteins on IFN- α/β receptors have not been characterized as well as their IFN- γ receptor counterparts. Tyr 466 of IFN- α R1, along with surrounding residues, is needed for STAT2 binding to the subunit. In the case of IFN- α R2c, the constitutive STAT2 binding site has been mapped between amino acids 404 and 462. Further, multiple tyrosine residues throughout the intracellular domain of the IFN- α R2c are required for full signaling activation, but their precise role is still not fully known.

Effects of Signaling through the IFN- α/β Receptor

Type-I IFNs are considered to be antiviral interferons. They provide a vital signal for cells to mount a state of antiviral

defense. The antiviral state is achieved through the diverse actions of newly translated proteins made in response to the induction of hundreds of ISGs by IFN signaling. ISGs encode proteins that have effects ranging from cell-cycle arrest to the inhibition of translation and RNA degradation. Together, these effects serve to make the cell a very unfavorable environment for viral replication. Not surprisingly, cell lines and mice defective in type-I IFN receptors are much more susceptible to viral infection.

See also: **Cell Architecture and Function:** Cytokinesis;
Molecular Biology: Messenger RNA Processing in Eukaryotes;
Signaling: Cytokinin.

Further Reading

- Abbas AK, Lichtman AH, and Pober JS (1997) *Cellular and Molecular Immunology*, 3rd edn., pp. 250–277. Philadelphia, PA: WB Saunders.
- Bach EA, Aguet M, and Schreiber RD (1997) The IFN γ receptor: A paradigm for cytokine receptor signaling. *Annual Review of Immunology* 15: 563–591.
- Ikeda H, Old LJ, and Schreiber RD (2002) The roles of IFN γ in protection against tumor development and cancer immunoediting. *Cytokine and Growth Factor Review* 13: 95–109.
- Mogensen KE, Lewerenz M, Reboul J, Lutfalla G, and Uze G (1999) The type I interferon receptor: Structure, function, and evolution of a family business. *Journal of Interferon and Cytokine Research* 19: 1069–1098.
- Sen GC (2001) Viruses and interferons. *Annual Review of Microbiology* 55: 255–281.
- Stark GR, Kerr IM, Williams BR, Silverman RH, and Schreiber RD (1998) How cells respond to interferons. *Annual Review of Biochemistry* 68: 227–264.

Intermediate Filament Linker Proteins: Plectin and BPAG1

P Fuchs and G Wiche, University of Vienna, Vienna, Austria

© 2013 Elsevier Inc. All rights reserved.

Glossary

Alternative splicing Mechanism to create different proteins (protein isoforms) from a common transcript by splice events using different exons or exon combinations.

Cytoskeleton Network of protein filaments in eukaryotic cells, which determines and influences cell shape and enables movement. Main components are actin filaments, MTs and IFs.

Knockout mouse Mutated mouse line harboring one or more defined mutations in a gene that lead to the inactivation of the gene product. Knockout mice are generated by embryonic stem (ES) cell and recombinant DNA technology.

Plakin protein family The plakin protein family, also referred to as the plakins, currently comprises plectin, BPAG1, MACF1, desmoplakin, envoplakin, periplakin, and

epiplakin. Plakins are structurally and functionally related proteins that, in most cases, can interact with more than one of the principal cytoskeletal filament systems. Common to all family members is the presence of at least one of two structurally highly conserved protein domains, the so-called plakin domain and the plakin repeat domain. The alternative designation suggested for plakins, cytolinkers (an abbreviation for cytoskeletal linker protein), is now being used to more broadly specify proteins with functions similar to those of plakins independently of sequence homology.

Protein isoforms Variant forms of a protein either generated by alternative splicing of a common transcript or as products of different members of a set of closely related homologous genes.

Plectin

The first identification of plectin as an interaction partner of intermediate filaments (IFs) was made three decades ago. The choice of its name (from *πλεκτή*, the ancient Greek word for net- or meshwork) was motivated by its microscopic visualization as part of a dense cytoplasmic network array in cultured cells.

Gene Locus and Isoform Diversity

In humans, the plectin gene is located on chromosome 8 and the murine ortholog on chromosome 15. The exon-intron organization of the plectin gene has been determined for man, mouse, and rat. Several plectin isoforms could be identified differing in their amino (N) termini. In the case of the mouse gene, alternative splicing of 11 different first exons into one common exon (exon 2) gives rise to a multiplicity of plectin variants (Figure 1). Additional alternative splicing of two small exons (2 α and 3 α) in the 5' area within the region coding for the actin-binding domain (ABD) increases the number of possible plectin isoforms. Furthermore, plectin transcript variants were identified that lack a single large exon encoding almost the entire α -helical coiled-coil rod domain of the molecule. Plectin isoforms are named after their alternative starting exons (e.g., plectin 1 and plectin 1a). However, the exact composition of the subsequent exons transcribed in each of the very large plectin messenger RNAs (mRNAs; ~11500–15500 nucleotides), especially in their very complex 5' part, has not yet been determined with absolute certainty.

Expression and Subcellular Localization

Plectin has been found to be expressed in virtually all mammalian tissues and cell types. It decorates different types of IFs, shows codistribution with their plasma membrane anchoring structures, and colocalizes with microfilaments (actin structures), primarily in peripheral areas of cells. It is associated with Z-lines of striated muscle, dense plaques of smooth muscle, intercalated discs of cardiac muscle, the basal cell-surface membrane (specifically hemidesmosomes) of keratinocytes, desmosomes of epithelial cells, and focal and fibrillary adhesions of various cultured cell types. Plectin has been identified also in cell layers separating tissues from fluid-filled cavities such as liver bile canaliculi, gut villi, kidney glomeruli, and endothelial cells of blood vessels. The expression patterns of plectin isoforms can vary considerably among different tissues and cell types, and certain isoforms are expressed in a tissue-specific or dominant manner, as revealed by quantitation of mRNAs isolated from different mouse organs. So, exon 1d-containing plectin transcripts were exclusively found in skeletal and heart muscle, whereas transcripts containing exon 1a are dominant in organs rich in epithelial cell types, such as lung, small intestine, and, in particular, skin. Intriguingly, the short alternative N-terminal sequences encoded by the different first exons were found to target isoforms to distinct subcellular locations. For instance, plectin 1a specifically associated with hemidesmosome-like structures, plectin 1b exclusively with mitochondria, and plectin 1f with vinculin-positive structures at actin stress fiber ends. Thus, it seems that a selection of plectin isoforms can be allocated to each cell type or tissue, providing a custom-made set of plectin-mediated functions to meet specific requirements of the particular cells. Consistent with such

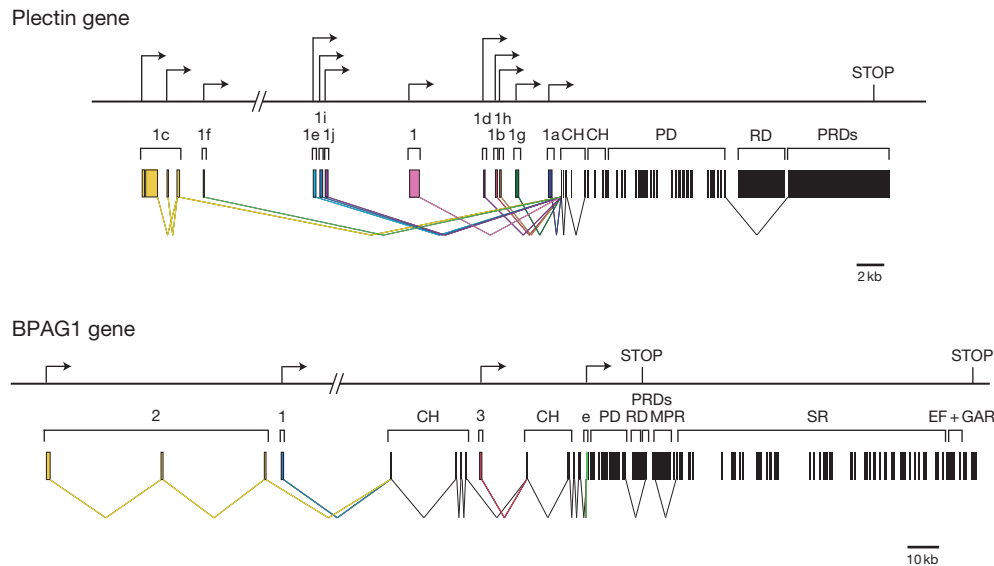


Figure 1 Plectin and BPAG1 gene structures and alternative splicing of transcripts. Schematics shown represent mouse (plectin) and human (BPAG1) species. Alternative transcriptional start sites (arrows) and stop codons are indicated. Alternative start exons are shown as colored boxes, other exons as black boxes. Splicing is indicated by kinked lines for alternative exons or when exons are bypassed. Alternative single first exons and combinations of starting exons are designated with numbers and/or small letters, protein domains with capital letters. CH, calponin-homology domains forming the ABD; PD, plakin domain; RD, coiled-coil rod domain; PRDs, plakin repeat domains; MPR, modules of plectin repeats; SR, spectrin repeats; EF + GAR, EF-hand and GAR domain. Note that in the case of plectin isoform 1c, two transcriptional start sites at the beginning of noncoding exons are used with only one 1c-specific coding exon preceding exon 2 (common to all isoforms). Reproduced from Röper K, Gregory SL, and Brown NH (2002) The 'spectraplakins': Cytoskeletal giants with characteristics of both spectrin and plakin families. *Journal of Cell Science* 115: 4215–4225, with permission from The Company of Biologists.

a model, there is increasing evidence for individual isoforms carrying out distinct and specific functions.

Structural Properties

Considering just full-length versions of the protein, plectin has a molecular mass of 506–535 kDa depending on its first coding exon (Figure 2). The molecule has been visualized by electron microscopy as a dumbbell-shaped structure composed of a central rod domain flanked by two large globular domains. In these structures, two plectin polypeptide chains, most likely, are arranged in parallel, forming an α -helical coiled-coil central rod. Within the N-terminal part of each chain, plectin contains an ABD of the 'classical' CH1–CH2-type (i.e., containing two calponin-homology domains). The so-called plakin domain, which represents the defining feature of plakin protein family members, contains an SH3-like domain flanked by spectrin-type repeats. A rod-forming, ~1280-amino-acid-long, mostly α -helical segment precedes the C-terminal part of plectin, which comprises six tandemly arranged so-called plakin-repeat domains each containing a globular core structure, or module, and a surface-exposed linker domain. Plectin modules, similar to those of other plakins, consist of four complete and one incomplete tandem copies of a sequence motif, in databases referred to as plectin repeat. Forming a β -hairpin followed by two antiparallel α -helices, the plectin repeat structurally resembles a motif called ankyrin repeat that is found in a large number of proteins. It should be noted that plectin isoforms that are encoded by transcripts starting with noncoding first

exons, such as plectin 1h, 1i, and 1j, contain only truncated versions of plectin's ABD (Figure 2).

Molecular Interactions

Plectin binds to cytoplasmic IFs of different types via direct interaction with their subunit proteins, including vimentin, desmin, glial fibrillary acidic protein (GFAP), neurofilament proteins, and cytokeratins; binding has also been demonstrated to the nuclear IF protein lamin B. The major binding site for IF subunit proteins has been mapped to a short sequence in the linker segment between modules 5 and 6 in plectin's C-terminal domain. Plectin's highly conserved ABD close to its N terminus mediates binding not only to actin but also to the integrin subunit β 4, nesprin-3, dystrophin, utrophin, and vimentin. Additionally, the signaling molecule phosphatidylinositol-4,5-bisphosphate (PIP2) binds within the ABD, regulating plectin's interaction with actin. Actin (and integrin β 4) binding of plectin isoform 1a is also affected by calmodulin, which binds in a Ca^{2+} -dependent manner to the CH1 domain of its ABD. Additional integrin β 4- and β -dystroglycan-binding sites have been ascribed to plectin's plakin domain and to its C-terminal globular domain. The plakin domain of plectin also harbors a binding site for the hemidesmosomal protein BPAG2/BP180. Additionally, direct interactions of plectin with the membrane skeleton proteins fodrin and α -spectrin, the desmosomal protein desmoplakin, MT-associated proteins of the MAP2 and MAP1 subtypes, and BRCA2, a centrosome-associated tumor-suppressor protein have been reported. In line with the emerging

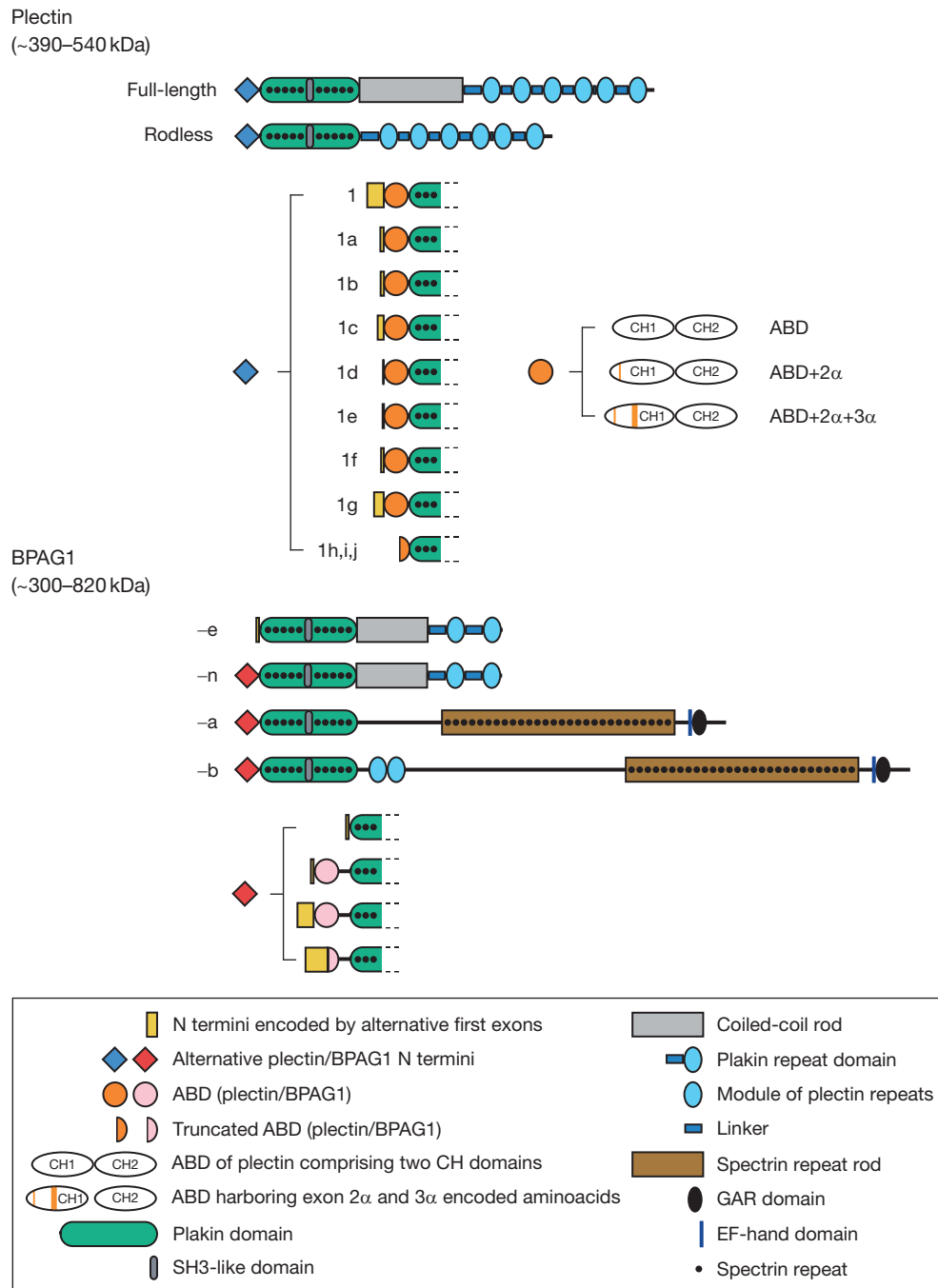


Figure 2 Schematic diagram of plectin and BPAG1 protein isoforms (mouse). Designation of isoforms (e.g., plectin 1a and BPAG1-e) and approximate molecular mass ranges are given (drawings are to scale). The blue diamond in front of full-length and rodless plectin stands for nine alternative N termini identified, the red diamond in front of BPAG1-n, -a, and -b for four alternative N termini identified. Note, that in case of BPAG1-n, one of the N-terminal variants symbolized by the red diamond is identical to BPAG1-e. For plectin three ABD variants exist, which are generated by alternative splicing of two additional exons (2 α and 3 α). In spite of the fact that only certain combinations of alternative first exons with exons 2 α and 3 α have been reported, it cannot be ruled out that all possible variants are existing. Note that in some isoforms of either one of the proteins, the ABD (typically consisting of two CH domains) can be truncated due to alternative splicing, resulting in a single CH domain. The plakin domains of plectin and BPAG1 are interaction domains consisting of ~10 spectrin repeats interrupted by an SH3-like domain. The plakin repeat domains each contain a globular module and a linker domain. Modules consist of four complete and one incomplete plectin repeats forming a complex globular domain. Linkers are considered as surface-exposed interaction domains, such as plectin's IF-binding site residing between plectin modules 5 and 6. Rod-like structures formed of spectrin repeats are present in BPAG1-a and -b. The GAS2-related (GAR) domain, named after the growth arrest specific protein 2, together with a glycine-serine-arginine (GSR)-rich domain (not indicated), has been suggested to mediate MT-binding.

concept that cytolinkers act as a platform for the assembly of complex protein machineries required for cytoskeletal rearrangements and intracellular signaling, plectin has been shown to be a target for a variety of serine/threonine kinases, including cell-cycle kinase Cdk1, and to interact with the receptor for activated C kinase 1 (RACK1), with the nonreceptor tyrosine kinase Fer, and energy homeostasis-controlling adenosine monophosphate-activated protein kinase. Whether the isoform-specific protein sequences located at the very N terminus of plectin variants can specifically select distinct binding partners or influence the affinities of other more distant modular-binding sites remains an open question.

Plectin Deficiency in Man and Mouse

In humans, mutations in the plectin gene lead to three forms of the disease epidermolysis bullosa simplex (EBS). Two of these, EBS with muscular dystrophy (EBS-MD) and EBS with pyloric atresia (EBS-PA) are autosomal recessive, one, EBS-Ogna, is dominant (Table 1). EBS-MD patients suffer from skin blistering and late-onset muscular dystrophy frequently leading to premature death in their third decade of life. In case of EBS-MD, only full-length plectin variants seem to be affected, whereas rodless variants are expressed at near-normal levels. In EBS-PA patients, where expression of all plectin isoforms is affected, the disease takes a more severe course. In this case, massive neonatal skin blistering and gastric abnormalities (pyloric or duodenal atresia) lead to early postnatal demise. EBS-Ogna was first described for a family from the Norwegian municipality of Ogna. It is caused by a site-specific mutation resulting in an exchange of a single amino acid residue (W to R) in plectin's ~1280-amino-acid-long rod domain. Patients suffering from this disease show skin fragility localized mainly to the limbs but no muscular symptoms.

Several mouse models have been generated to study the role of plectin *in vivo* and to provide animal models mimicking the clinical features of plectin-related EBS (Table 2). A first plectin-deficient mouse line, where all isoforms were inactivated in all cell types and tissues (plectin-null mouse), was followed by

several conditional knockout lines, in which the ablation of plectin was restricted to certain tissues and cell types, and by isoform-specific knockout lines. Homozygous plectin-null mice show a phenotype similar to EBS-MD, except that they die already 2–3 days after birth. In addition to severe skin blistering, compromised myofibril integrity in skeletal and heart muscle was detected already at this early stage.

In conditional knockout mice, where the ablation of plectin is restricted to keratin 5-expressing epithelia (K5-Cre cKO), skin blistering develops with a similar onset, body distribution, and severity as in plectin-null mice. Mice from both plectin knockout strains seem to die from malnutrition as a consequence of blistering of the oral epithelia, probably reducing the pups' desire for food due to pain, or obstruction of the oral cavity due to large blisters or cellular debris, or both. Mice with a conditional deletion of plectin in striated muscle (MCK-Cre cKO), develop progressive degenerative alterations in skeletal muscle, including aggregation and partial loss of desmin IF networks, detachment of the contractile apparatus from the sarcolemma, and decreased number and impaired function of mitochondria, thus, mimicking the late-onset muscular dystrophy typical of EBS-MD. Inactivation of plectin in nerve precursor cell-specific conditional knockout mice (nestin-Cre cKO) led to reduced motor nerve conduction velocity in sciatic nerve.

To reveal isoform-specific functions, a series of knockout mice was generated, each lacking just one of the isoforms, by deleting the corresponding N-terminal alternative isoform-specific sequences from the gene. All of these mouse lines turned out to be viable and fertile enabling comprehensive phenotyping. The absence of plectin isoform 1, an inner membrane-associated variant of plectin, led to alterations of the actin and IF cytoskeleton, impaired migration potential of fibroblasts and T cells, and reduced leukocyte infiltration during wound healing. Ablation of the outer mitochondrial membrane-associated plectin isoform 1b led to changes in the shape of the organelle probably by deregulation of the mitochondrial fusion–fission machinery. Inactivation of plectin 1c, the variant dominating in neural tissues, led to reduced motor

Table 1 Human disorders caused by plectin and BPAG1 mutations

<i>Genetic disease</i>	<i>Inheritance</i>	<i>Isoforms affected</i>	<i>Clinical signs</i>
<i>Plectin:</i>			
EBS-MD	Recessive	Full-length only ^a	Neonatal skin blistering and late onset, progressive muscular weakness
EBS-PA	Recessive	All ^a	Severe neonatal skin blistering and gastric abnormalities, early postnatal demise
EBS-Ogna	Dominant	Full-length only ^b	Localized skin fragility
<i>BPAG1:</i>			
Esophageal atresia and psychomotor retardation	Dominant, <i>de novo</i> reciprocal chromosomal translocation ^c	BPAG1-a, BPAG1-b	Encephalopathy, severe motor and mental retardation and delayed visual maturation
EBS	Recessive	BPAG1-e, BPAG1-n	Lifelong-generalized trauma-induced blisters and erosions. Neurological abnormalities? ^d

^aNot experimentally clarified for all mutations reported.

^bA heterozygote missense mutation in the rod-encoding exon 31 apparently derogates plectin functions in a dominant manner. Rodless variants should be unaffected.

^cSymptoms were first thought to be due to BPAG1-a and -b haploinsufficiency. As BPAG1 heterozygous mice show no symptoms it is more likely that dominant negative protein variants cause the symptoms observed.

^dWhether the observed neurological symptoms are a consequence of the reported BPAG1 mutation is uncertain.

Table 2 Mouse lines harboring plectin mutations

Designation of lines	Knockout type	Phenotypes (detected)
Plectin KO	Null ^a	Postnatal death at day 1–3, signs of starvation and growth retardation, internal and external skin blistering, skeletal muscle myopathy, and disintegration of intercalated disks in the heart
K5-Cre cKO, (<i>plec</i> ^{loxP/loxP} ; <i>Krt5-Cre</i>)	Conditional (stratified epithelia)	Postnatal death at day 1–3, signs of starvation and growth retardation, internal and external skin blistering
MCK-Cre cKO, (<i>plec</i> ^{loxP/loxP} ; <i>Ckm-Cre</i>)	Conditional (skeletal and heart muscle)	Progressive degeneration of skeletal muscle, aggregation of desmin IF networks, detachment of contractile apparatus from sarcolemma, decreased number and impaired function of mitochondria
Nestin-Cre cKO (<i>plec</i> ^{loxP/loxP} ; <i>Nes-Cre</i>)	Conditional (nerve precursor cells)	Reduced motor nerve conduction velocity in sciatic nerve
Plectin 1 KO	Isoform-specific ^b	Abnormalities of actin cytoskeleton and impaired migration of dermal fibroblasts, diminished chemotactic migration of T cells, reduced leukocyte infiltration during wound healing
Plectin 1b KO	Isoform-specific ^b	Shape changes of mitochondria in fibroblasts and myoblasts
Plectin 1c KO	Isoform-specific ^b	Reduced motor nerve conduction velocity in sciatic nerve, increased number of motor nerve fibers with small cross-sectional areas
Plectin 1d KO	Isoform-specific ^b	Necrotic, centrally nucleated, hypertrophic and atrophic skeletal muscle fibers; aggregation of desmin IFs and mitochondrial networks, compromised activities and lysis of mitochondria
Plectin 1d/desmin double KO	Double	Phenotypic characteristics similar to those of plectin 1d KO mice (partly more severe)

^aAll isoforms are inactivated in all cell types and tissues.

^bIsoform-specific KO mice lack only those transcript variants that start with a certain alternative first exon; expression of other isoforms is supposed to be unaffected.

nerve conduction velocity in sciatic nerve, reminiscent of the nestin-Cre cKO mice. Mice lacking plectin 1d, the Z-disk-associated isoform of plectin, showed all the characteristics of myofibrillar myopathies, including muscle fiber degeneration and aggregation of the desmin IF and mitochondrial networks, combined with compromised respiratory activity and lysis of mitochondria.

BPAG1

BPAG1 (alias BP230) and BPAG2 (alias BP180 or type XVII collagen) were originally identified in search of antigens with immunoreactivity to autoantibodies of patients suffering from bullous pemphigoid, an autoimmune subepidermal skin-blistering disease. Both proteins are components of hemidesmosomes but BPAG2, in contrast to BPAG1, harbors a collagenous extracellular domain and is not a member of the plakin protein family. The mouse ortholog of BPAG1 is also called dystonin in consideration of the fact that mutations in this gene cause sensory neuron degeneration in the mouse mutant *dystonia musculorum* (*dt*).

Gene Locus and Isoform Diversity

The BPAG1 gene in humans is localized on chromosome 6, and in mice, where it is referred to as dystonin, on chromosome 1. Similar to plectin, multiple exons in the BPAG1 gene are the basis of several protein isoforms generated by alternative splicing of transcripts (Figure 1). Four major variants (BPAG1-a, -b, -e, and -n) were specified showing clear differences in length and structure (Figure 2). In addition to this diversity, which is based on several combinations of structural domains, the variety of BPAG1 isoforms is further increased by

the existence of four alternative translation initiation sites for BPAG1-a, -b, and -n. Whether all variants of BPAG1, especially BPAG1-n, can be found in combination with all of these alternative N-terminal sequences is a matter of controversy. Moreover, like in the case of plectin, the exact composition of transcribed exons in BPAG1 mRNAs is often difficult to define due to the very large size of some of the transcripts.

Expression and Subcellular Localization

The epithelial isoform BPAG1-e is expressed primarily in basal epithelial cells where it colocalizes with hemidesmosomes. BPAG1-a is expressed at high levels in the nervous system, especially in dorsal root ganglia and spinal cord, while BPAG1-n, previously thought to be the dominant isoform in this tissue, seems to be minor compared to BPAG1-a. Isoform BPAG1-b is predominantly expressed in skeletal and cardiac muscle where it is found associated with Z-disks, and in developing mouse embryos it was also found in bone cartilage of the vertebrae. Of the four alternative N-terminal sequences so far identified, one is expressed only in epithelia whereas the expression profiles of the other three strongly varies in different tissues. There is some evidence that the very N-terminal alternative sequences of BPAG1 can determine the protein's localization within cells, similar to the case of plectin.

Structural Properties

The molecular masses estimated for BPAG1 isoforms vary from 302 kDa for BPAG1-e, 344 kDa for BPAG1-n, 615 kDa for BPAG1-a, to 824 kDa for BPAG1-b (Figure 2). Unlike plectin, all four isoforms of BPAG1 show considerable structural divergence. BPAG1-e contains an α -helical coiled-coil rod domain flanked by an N-terminal plakin domain and a C-terminal

region comprising two plakin-repeat domains. BPAG1-n overall is similar to BPAG1-e, except for containing an additional N-terminal ABD of the CH1-CH2-type, which in one variant is truncated, comprising only the CH2 domain. BPAG1-a, starting with an ABD followed by a plakin domain, resembles BPAG1-n at its N terminus, but differs regarding the rest of the molecule. Its rod domain consists of up to 28 spectrin repeats and its C-terminal domain contains EF-hand Ca^{2+} -binding motifs followed by a region including a GAS2-related (GAR) protein domain. BPAG1-b exhibits structural features similar to BPAG1-a, except for an additional pair of closely adjoined plectin repeat modules of unknown function located between the plakin domain and the spectrin repeats.

Molecular Interactions

Because of its large size and structural diversity, BPAG1 can be expected, like plectin, to have not only a variety of interaction partners but also different sets of binding partners for each isoform. Via their N-terminal CH1-CH2-type ABD, BPAG1 isoforms can associate with filamentous actin. However, some of its isoforms, which due to alternative splicing within the ABD, express only one of the two CH domains (Figure 2), lack this ability. This truncation has been suggested to activate a MT-binding site located within the plakin domain of BPAG1 that otherwise remains cryptic. In addition, the C-terminal GAR and the GSR-containing domains have been suggested to mediate interactions with microtubules. The location of IF-binding sites within the C-terminal plakin-repeat domains of BPAG1-e and -n has been confirmed for keratin and neurofilaments, and BPAG2 and ERBIN, a specific Erb-B2 tyrosine kinase receptor-binding protein, have been shown to interact with BPAG1's plakin domain. Furthermore, BPAG1-a was found to directly interact with dynactin p150^{Glued} and the endosomal vesicle protein retrolinkin, while BPAG1-b was reported to interact via its N terminus with plectin and α -actinin.

BPAG1 Deficiency in Man and Mouse

In humans, so far only a couple of mutations have been identified in the BPAG1 gene (Table 1). In one reported case, a 4-year old, suffering from a defect in the BPAG1 gene caused by a *de novo* reciprocal translocation, showed signs of nonprogressive encephalopathy, severe motor and mental retardation, and delayed visual maturation. The BPAG1 isoforms affected in this case probably were BPAG1-a and -b, but as the chromosomal alteration was restricted to only one allele, it is not clear whether the symptoms observed were due to a gain or loss of function. In a second case reported, a 38-year-old patient with a homozygous nonsense mutation within the exon encoding the coiled-coil rod domain of isoforms BPAG-e and -n suffers from lifelong-generalized trauma-induced blisters and erosions, a condition which is now classified as a new form of autosomal recessive EBS. Whether the additional neurological abnormalities of these patients are caused by the BPAG1 mutation remains unclear. The role of autoantibodies to BPAG1 found in patients suffering from bullous pemphigoid could be further elucidated by showing that BPAG2, not BPAG1, is the primary autoantigen critical for disease development.

Table 3 Mouse lines harboring BPAG1 mutations

Designation of lines	Knockout type	Phenotypes detected
<i>Dystonia musculorum</i> (<i>dt</i>) ^a	Natural and spontaneous mutants ^b	Sensory neuron degeneration culminating in death 3–5 weeks after birth, defects in muscle and neuron cytoarchitecture and in skin integrity, accumulation of organelles in focal axonal swellings ^c
BPAG1 KO	Null ^d	Deterioration in motor function, sensory neuron degeneration, skin blistering upon mechanical trauma; life expectancy similar to that of <i>dt</i> mice

^aSeveral allelic variants.

^bMutations in the different *dt* mouse lines most likely affect different BPAG1 isoforms.

^cDifferent *dt* lines show similar, but not identical phenotypes.

^dIn the BPAG1-null line, all major isoforms of BPAG1 are affected.

In mice, deletions in the BPAG1 gene cause the disorder *dt*, an autosomal recessive neuropathy (Table 3). At the age of 1–2 weeks, homozygous *dt* mice display progressive loss of limb coordination caused by degeneration of sensory neurons. Dorsal root ganglia axons exhibit disorganization of neuronal IFs and MT networks accompanied by abnormal axonal myelination and axonal swellings. It has been suggested that swellings result from defective axonal transport caused by MT disorganization. Additional defects in Schwann cell and myocyte functions have been reported for these mice. Not all *dt* mouse strains, however, lack the BPAG1-e isoform. Independent of the *dt* mice, most of which carry spontaneously arisen BPAG1 mutations, a BPAG1-defective mouse strain was generated using gene-targeting techniques. This mutation eliminates the exons encoding the plakin domain of BPAG1 (common to all isoforms) and, therefore, is considered as null allele. In addition to the phenotypic characteristics of some *dt* mouse strains, BPAG1-null mice show fragility of skin upon mechanical stress. Based on these findings it is assumed that the absence of BPAG1-e is responsible for the observed skin abnormalities, deficiency in BPAG1-a accounts for the neuronal deficits, and lack of BPAG1-b could be the reason for muscle defects.

See also: **Cell Architecture and Function: Actin-Related Proteins; Cytoskeletal Motors: General Principles; Intermediate Filaments; Keratins and the Skin; Microtubule-Associated Proteins.**

Further Reading

Abrahamsberg C, Fuchs P, Osmangic-Myers S, et al. (2005) Targeted ablation of plectin isoform 1 uncovers role of cytolinker proteins in leukocyte recruitment. *Proceedings of the National Academy of Sciences of the United States of America* 102: 18449–18454.

- Boyer JG, Bernstein MA, and Boudreau-Lariviere C (2010) Plakins in striated muscle. *Muscle and Nerve* 41: 299–308.
- Fuchs P, Zörer M, Reipert S, et al. (2009) Targeted inactivation of a developmentally regulated neural plectin isoform (plectin 1c) in mice leads to reduced motor nerve conduction velocity. *Journal of Biological Chemistry* 284: 26502–26509.
- Goryunov D, Adebola A, Jefferson JJ, et al. (2007) Molecular characterization of the genetic lesion in *Dystonia musculorum (dt-Alb)* mice. *Brain Research* 1140: 179–187.
- Groves RW, Liu L, Dopping-Hepenstal PJ, et al. (2010) A homozygous nonsense mutation within the dystonin gene coding for the coiled-coil domain of the epithelial isoform of BPAG1 underlies a new subtype of autosomal recessive epidermolysis bullosa simplex. *Journal of Investigative Dermatology* Feb. 18 (Epub ahead of print).
- Jefferson JJ, Leung CL, and Liem RK (2004) Plakins: Goliaths that link cell junctions and the cytoskeleton. *Nature Reviews Molecular Cell Biology* 5: 542–553.
- Konieczny P, Fuchs P, Reipert S, et al. (2008) Myofiber integrity depends on desmin network targeting to Z-disks and costameres via distinct plectin isoforms. *Journal of Cell Biology* 181: 667–681.
- Konieczny P and Wiche G (2008) Muscular integrity – a matter of interlinking distinct structures via plectin. In: Laing NG (ed.) *The Sarcomere and Skeletal Muscle Disease*, vol. 642, pp. 165–175. New York, NY: Springer, Landes Bioscience.
- Kostan J, Gregor M, Walko G, and Wiche G (2009) Plectin isoform-dependent regulation of keratin–integrin $\alpha 6 \beta 4$ anchorage via Ca^{2+} /calmodulin. *Journal of Biological Chemistry* 284: 18525–18536.
- Liu JJ, Ding J, Kowal AS, et al. (2003) BPAG1n4 is essential for retrograde axonal transport in sensory neurons. *Journal of Cell Biology* 163: 223–229.
- Liu JJ, Ding J, Wu C, et al. (2007) Retrolinkin, a membrane protein, plays an important role in retrograde axonal transport. *Proceedings of the National Academy of Sciences of the United States of America* 104: 2223–2228.
- Osmanagic-Myers S, Gregor M, Walko G, et al. (2006) Plectin-controlled keratin cytoarchitecture affects MAP kinases involved in cellular stress response and migration. *Journal of Cell Biology* 174: 557–568.
- Osmanagic-Myers S and Wiche G (2004) Plectin-RACK1 (receptor for activated C kinase 1) scaffolding. *Journal of Biological Chemistry* 279: 18701–18710.
- Reznicek GA, Konieczny P, Nikolic B, et al. (2007) Plectin 1f scaffolding at the sarcolemma of dystrophic (*mdx*) muscle fibers through multiple interactions with β -dystroglycan. *Journal of Cell Biology* 176: 965–977.
- Reznicek GA, Walko G, and Wiche G (2010) Plectin gene defects lead to various forms of epidermolysis bullosa simplex. In: Bruce HT and Murrell DF (eds.) *Epidermolysis bullosa Part I – Pathogenesis and Clinical Features, Dermatologic Clinics*, vol. 28, pp. 33–41. Saunders, Philadelphia, PA: Elsevier.
- Sonnenberg A and Liem RK (2007) Plakins in development and disease. *Experimental Cell Research* 313: 2189–2203.
- Wiche G (1998) Role of plectin in cytoskeleton organization and dynamics. *Journal of Cell Science* 111: 2477–2486.
- Wilhelmsen K, Litjens SH, Kuikman I, et al. (2005) Nesprin-3, a novel outer nuclear membrane protein, associates with the cytoskeletal linker protein plectin. *Journal of Cell Biology* 171: 799–810.
- Winter L, Abrahamsberg C, and Wiche G (2008) Plectin isoform 1b mediates mitochondrion – intermediate filament network linkage and controls organelle shape. *Journal of Cell Biology* 181: 903–911.
- Young KG and Kothary R (2007) Dystonin/BPAG1 – a link to what? *Cell Motility and the Cytoskeleton* 64: 897–905.

Intermediate Filaments

K M Bernot and P A Coulombe, Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Intermediate filament (IF) Filaments of 10–12 nm width and several microns in length found in the nucleus and the cytoplasm, where they form the cytoskeleton along with actin microfilaments and microtubules.

Nuclear lamina A scaffold-like network of lamin intermediate filaments apposed against the inner membrane of the nuclear envelope.

Trigger motif Small group of amino acids required for stabilization of an alpha helix and nucleation of dimer formation through coiled-coil interactions.

Unit length filament (ULF) Late-stage intermediate during the assembly of cytoplasmic IFs, which is ~60 nm in length and consists of eight tetramers (32 monomers) annealed laterally.

Intermediate filaments (IFs) are intracellular fibrous polymers that became a prominent constituent of the cytoskeleton coinciding with the appearance of multicellular life. They are readily distinguishable from actin microfilaments (6–8 nm) and microtubules (25 nm) by virtue of their diameter (10–12 nm). In mammals, >70 genes encode proteins able to self-assemble into IFs. A small subset of these genes encodes lamins, the building blocks of the nuclear lamina, whereas all others encode cytoplasmic proteins. A major function shared by all IFs is to endow cells and tissues with a unique ability to resist trauma – this contribution is important in structures (e.g., axons and nuclei), cells, and tissues exposed to significant stress. Additional important roles in regulating cell growth and proliferation, cell death, cell migration, and response to stresses are manifested in a IF protein- and context-specific manner.

Features of IF Proteins

As a group, IF proteins are very heterogeneous in size (40–240 kDa), charge, and various other properties (Table 1). The property of self-assembly is hard-wired into the characteristic tripartite domain structure shared by all IF proteins (Figures 1(a) and 1(b)). The central domain, an extended α -helix featuring long-range heptad repeats interrupted at three conserved locations, is the main determinant of self-assembly. This rod domain comprises 310 and 352 amino acids in cytoplasmic and nuclear IF proteins, respectively, and is bordered by invariant, signature ~15-mer sequences. Flanking sequences at the N- and C-termini of the central rod domain are key in conferring each IF protein its individuality. Most IFs are heteropolymeric *in vivo*, although several IF proteins can *de facto* self-assemble into filaments as homopolymers *in vitro*.

Almost all metazoans (animals) contain both nuclear and cytoplasmic IFs, with the notable exception of arthropods, which contain only lamins. The presence of an exoskeleton, along with an increased density of microtubules in cells subjected to stress, may account for the absence of cytoplasmic IFs in arthropods. Phylogenetic analyses suggest that IFs initially appeared in the nucleus and later moved to the cytoplasm during evolution.

IFs have not been found in plants, fungi, or protists, with one exception: there is evidence for an IF-like system in *Caulobacter crescenti*. Most IFs are intracellular proteins; rare exceptions include the hagfish IFs, which are secreted polymers.

Assembly, Structure, and Regulation of IFs

Assembly and Structure

Owing to the long-range heptad repeats present within their central α -helical domain, IF proteins readily dimerize through coiled-coil interactions. In the dimer, individual rod domains are aligned in parallel and in register (owing to controlled nucleation by a trigger motif) (Figure 1(b)). Nuclear and cytoplasmic IFs differ in the formation of tetramers. Lamin dimers associate in a head-to-tail fashion to form thin longitudinal strands, several of which laterally associate with an antiparallel orientation to form mature filaments. Dimers of cytoplasmic IFs, on the other hand, interact along their lateral surfaces with an antiparallel orientation to form compact tetramers. These tetramers continue to favor lateral interactions as they further polymerize to form an ~60-nm long unit-length filament (ULF). Annealing of ULFs produces mature 10–12 nm filaments whose length can reach >10 μ m. Mature IFs lack an obvious structural polarity (see Figure 1(c)), reflecting the antiparallel orientation of dimers within them. While models for idealized IFs usually include 32 monomers per filament cross-section, mass per-unit length determinations revealed that in actuality, this number varies a fair amount depending on the source and type of IFs. This heterogeneity is still unexplained. There is as yet no satisfactory model for the high-resolution structure of IFs.

Unlike actins and tubulins, IF proteins do not bind or metabolize nucleotides. They do not require an external input of energy or cofactor for assembly *in vitro*. They readily assemble into filaments when ectopically expressed in a virtually any cell type in culture, suggesting that the process does not require specialized factors *in vivo*. Cytoplasmic IF networks often extend from the perinuclear area to the cell periphery (Figure 1(d)), although there are many instances in which IFs are concentrated in specialized regions of the cell. Such is the case for keratins in

Table 1 The intermediate filament family

IF type	Protein name	No. genes	Protein MW	Assembly group ^a	Tissue distribution	Association with human disease ^b
I	Keratins (acidic)	28	40–64	A	Epithelia	Multiple skin fragility diseases Oral and cornea fragility diseases Hair disorders Inflammatory bowel disease ^c Cryptogenic cirrhosis ^c Nearly identical to type I keratins
II	Keratins (basic)	26	52–68	A	Epithelia	
III	Vimentin	1	55	B	Mesenchyme (esp. early develop.)	None known
	Desmin	1	53	B	Muscle	Distal myopathy; cardioskelet. Myop.
	GFAP	1	50–52	B	Astrocytes/Glia	Alexander's disease
	Peripherin	1	54	B	PNS neurons	None known
	Syncoilin	1	54	B	Muscle	None known
IV	NF-L	1	62	B	CNS neurons	Charcot-Marie-tooth disease
	NF-M	1	102	B	CNS neurons	Parkinson's disease ^d
	NF-H	1	110	B	CNS neurons	Amyotrophic lateral sclerosis
	α -Internexin	1	66	B	CNS neurons	None known
V	Lamin A/C	1	72/62	C	Differentiated tissue (nucleus)	Muscular dystrophy variants
	Lamin B1	1	65	C	Ubiquitous (nucleus)	Dilated cardiomyopathies
	Lamin B2	1	78	C	Ubiquitous (nucleus)	Familial partial lipodystrophy Mandibular acrodysplasia Charcot-Marie-tooth disease Premature aging conditions Others
Orphan	Nestin	1	240	B	Heterogeneous	None known
	Synemin	1	182	B	Muscle	None known
	Desmuslin	1	140	B	Muscle	None known
	Paranemin	1	178	B	Muscle	None known
	Phakinin / CP49	1	46	D	Lens	Juvenile and congenital cataracts
	Filensin / CP115	1	83	D	Lens	None known

^aWithin each group, proteins heteropolymerize in either an obligatory or facultative fashion to form filaments.

^bMutations affecting IF-coding sequences are causative in the diseases listed unless indicated otherwise.

^cMutations in either K8 or K18 have been identified as risk factors for these conditions.

^dAssociation needs to be confirmed in additional patients.

polarized epithelial sheets, desmin in myocytes, and neurofilaments in axons. In the cytoplasm, IFs can interact with F-actin, microtubules, and specific adhesion-mediating complexes. While the relative importance of these interactions varies between IF proteins and cell types, their dominant influence on IF organization has been established through experiments in which these elements were disrupted.

Regulation

Both nuclear and cytoplasmic IFs are disassembled and/or profoundly reorganized during mitosis. Both of them can also be dynamically remodeled under steady-state conditions in noncycling cells. *In vivo* experiments utilizing green fluorescent protein (GFP)-tagged vimentin or keratin suggest that filament formation is initiated at the cell periphery whereas the subunit exchange fueling steady-state dynamics possibly occurs along the filament wall. More information about these crucial aspects of IF biology is needed. Posttranslational modifications play a key role toward the regulation of the assembly and disassembly of IFs *in vivo*, and of their interaction with

associated proteins. Thus, all IF proteins are subject to phosphorylation at multiple sites, and the kinase, target site(s), and significance have been deciphered in many instances. The molecular basis of steady-state polymer dynamics, which have been tied to nucleotide metabolism for F-actin and microtubules, remain unresolved for IFs. Possibly, rapid cycles of IF protein phosphorylation–dephosphorylation could enact the reversible conformational change needed. In addition, specific IF proteins have been shown to be O-glycosylated, ubiquitinated, and/or sumoylated, or modified in other ways, though the underlying significance remains unclear in many cases. Lamins undergo lipid modification and enzymatic processing of their C-terminus, reflecting their association with the inner nuclear membrane.

Dissecting the IF Superfamily

Most of the >70 genes forming the IF superfamily can be easily classified into one of five sequence types based on gene

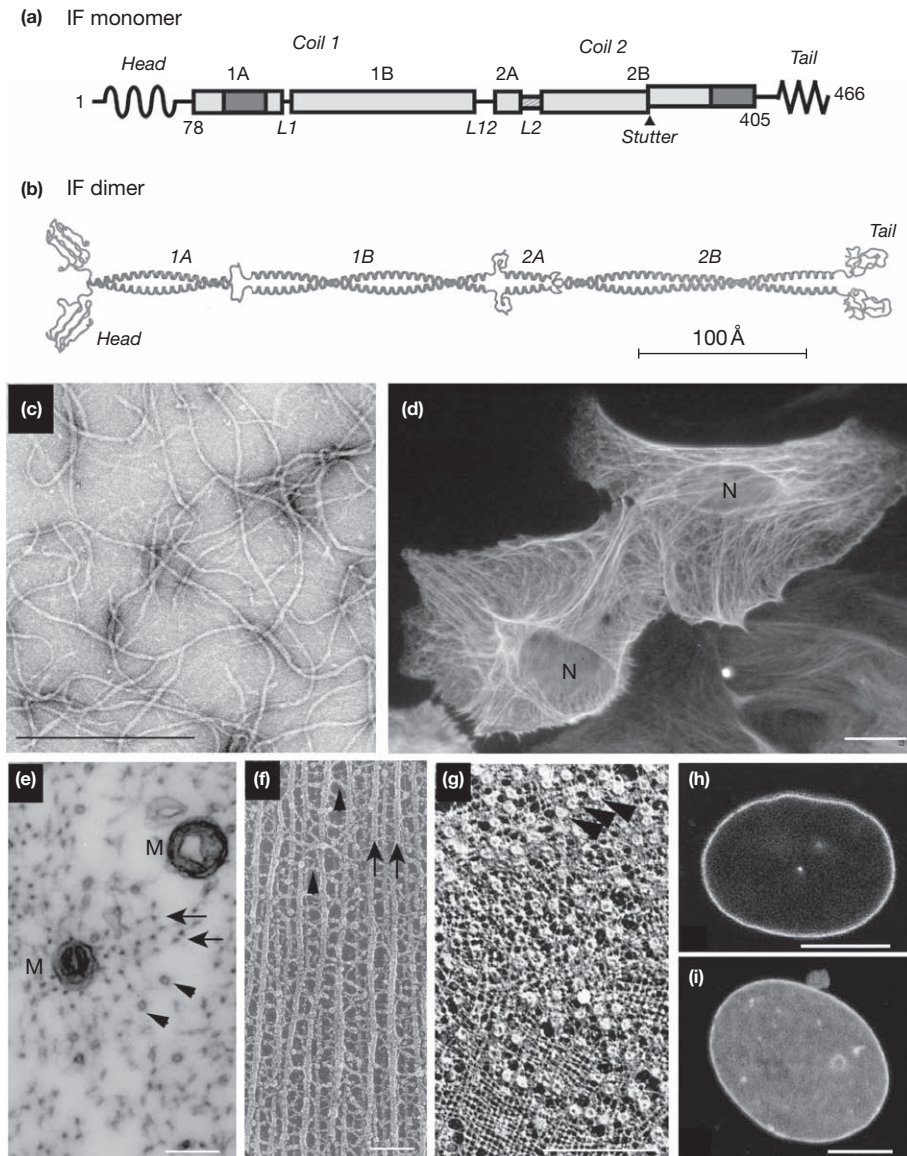


Figure 1 (a) Diagram illustrating the domain substructure of IF proteins. Rectangles depict central α -helical rod domain (1A, 1B, 2A, and 2B) separated by short, nonhelical linkers (L1, L12, and L2). The rod domain is flanked by nonhelical head and tail domains. Two highly conserved regions critical for the nucleation of coiled-coil dimer formation (trigger motifs) are darkly shadowed. The numbers represent amino acid residues in human vimentin (Strelkov et al. 2003 *Bioessays* 25: 243–251.) (b) Diagram illustrating the architecture of the human vimentin dimer, as determined by X-ray crystallography (1A and 2B subdomains, dark gray) or modeling (light gray) (Strelkov et al. 2003 *Bioessays* 25: 243–251.) (c) Keratin IFs reconstituted *in vitro* from purified proteins and visualized by negative staining and electron microscopy. These filaments appear as smooth-surfaced, apolar, flexible polymers of ~ 11 nm in width and several microns in length. Scale = 250 nm. (d) Visualization of keratin IFs by indirect immunofluorescence in primary cultures of mouse skin keratinocytes. The filaments form a network extending from the surface of the nucleus (N) to the cytoplasmic periphery in these two cells. Scale = 30 μ m. (e and f) Visualization of neurofilaments in axon processes. (e) Thin-section transmission electron microscopy, showing neurofilaments viewed in cross-section. Neurofilaments (arrows), microtubules (arrowheads), and mitochondria (M) can be distinguished. Scale = ~ 250 nm. (Garcia et al. 2003 *Journal of Cell Biology* 163: 1011–1020.) (f) Quick-freeze deep-etch electron microscopy of axoplasm along its main axis. Neurofilaments (arrows) fill the axoplasm, and numerous cross-bridges (arrowheads) between the filaments can be seen. Scale = 100 nm. (Hirokawa et al. 1984 *Journal of Cell Biology* 98: 1523–1536.) (g–i) Visualization of lamins in the nucleus. (g) Native nuclear lamina of *Xenopus* oocytes. Preparation of Triton X-100-extracted nuclear envelope by freeze-dried/metal-shadowing reveals the meshwork basket weave of the nuclear lamina and the array of nuclear pore complexes (arrowheads). Scale = 1 μ m. (Aebi et al. 1986 *Nature* 323: 560–564.) (h and i) GFP-tagged Lamin B (h) or A (i) was introduced into mouse embryonic fibroblasts. Live cells in interphase show bright fluorescence around the nuclear rim while a fainter veil of fluorescence is visible in the nucleoplasm. Scale = 10 μ m. (Moir et al. 2000 *Journal of Cell Biology* 151: 1155–1168.)

structure and nucleotide homology over the central alpha-helical domain. Individual members within each sequence type are often related in their tissue distribution. The remaining sequences defy classification and have been assigned to a sixth, orphan group of sequences (see below).

Type I and Type II Keratins

Type I and type II IFs are part of the keratin (or cytokeratin) family of proteins found in all epithelia. The human genome features 54 functional keratin genes, with 28 type I and 26 type II keratin genes (see [Table 1](#)). Type I keratins tend to be smaller and acidic compared to the larger, neutral–basic type II keratins. Keratin assembly proceeds strictly from type I to type II heterodimers, accounting for the duality of sequences as well as their pairwise regulation *in vivo*. While any combination of purified type I and type II keratin proteins yields a fibrous polymer *in vitro*, specific pairs are coregulated in a fashion directly related to epithelial tissue type and/or differentiation state *in vivo*. In mammals, specific type I–type II gene pairings occur in simple (e.g., liver and GI tract), complex (e.g., skin and oral mucosa), and hard epithelia (e.g., hair and nail), with little overlap. Quite remarkably, moreover, specific type I–type II gene pairs are differentially and sequentially expressed in progenitor (mitotically active), early, and late differentiation stages in a host of epithelia. The conservation of keratins in terms of primary structure and distribution suggests an important role in epithelial diversity – this role as determinant of tissue specificity is emerging from detailed studies of transgenic models in which select IF genes are manipulated.

Type III IFs

The type III group consists of vimentin, desmin, glial fibrillary acidic protein (GFAP), peripherin, and syncoilin ([Table 1](#)). Vimentin is widely expressed during development, but typically gives way to other IF proteins during differentiation. Accordingly, vimentin, which is an assembly competent on its own, is usually found in co-polymers with other type III, type IV, or orphan sequences in mature cell types. Vimentin persists as the dominant cytoplasmic IF protein in fibroblasts, endothelial cells, adipocytes, macrophages, lymphocytes, and neutrophils. Desmin is found in the cytoplasm of smooth muscle and is concentrated at the Z lines of skeletal and cardiac muscle. Several variants of GFAP arise from a single precursor messenger RNA (mRNA), and occur in the astrocytes of the central nervous system and in Schwann cells of the peripheral nervous system (PNS). Peripherin occurs in the post-mitotic neurons of the PNS, whereas syncoilin is an intrinsic component of the dystrophin-associated protein complex in skeletal and cardiac muscle.

Type IV Neurofilaments

This group occurs specifically in neurons and consists of the neurofilament triplet proteins NF-L, -M, and -H, and α -internexin. NF-L (light), NF-M (medium), and NF-H (heavy) co-polymerize obligatorily in a 5:3:1 ratio to form filaments with a core diameter of 10–12 nm and unusual side arms projecting >20 nm away ([Figure 1\(f\)](#)). These side arms

represent the extended C-terminal tails of NF-M and NF-H, which feature a large number of Lys–Ser–Pro repeats that adopt an elongated shape when stoichiometrically phosphorylated. Neurofilament proteins are synthesized in the cell body of neurons and transported to the axon, where they contribute to the determination of axon caliber owing to their side arm projections ([Figure 1\(f\)](#)). α -Internexin, on the other hand, is expressed at an earlier stage of neuronal differentiation and homopolymerizes into IFs that may thereafter scaffold the assembly of NF-L, M, and H, synthesized late during differentiation.

Type V Lamins

Nuclear lamins constitute the type V IFs. Consistent with their unique distribution within the cell, lamin proteins contain a canonical nuclear localization signal within their C-terminal tail and a CAAX-box at their C-terminus which, after farnesylation and carboxyl-methylation, mediates membrane anchorage. The C-terminus of lamin A is a novel β immunoglobulin-like fold that acts as a docking site for several lamin-associated proteins, and its alteration causes disease ([Table 1](#)). In addition, lamins feature six extra heptad repeats within the α -helical subdomain 1B of the rod, the significance of which is unclear. The presence of these six extra heptads in the cytoplasmic IF proteins of invertebrates has lent strong support to a proposal that the ancestral IF gene was lamin-like in character.

There are two major types of lamins, designated A and B. B-type lamins, which are essential in all species tested to date, are found in all somatic and germ cells. The mammalian *Lmna* locus yields four distinct proteins; the two major ones, A (664 a.a.) and C (573 a.a.), are identical over residues 1–566 and differ at their C-termini owing to alternative splicing of a common precursor mRNA. The *Lmna* products occur preferentially in differentiated cells. Lamins form a meshwork structure called the ‘nuclear lamina’ just beneath the inner nuclear membrane ([Figure 1\(g\)](#)), where they interact with several key constituents of that membrane. In addition, lamins also occur in the nucleoplasm, with a mixed organization as a diffuse network and in focal clusters ([Figures 1\(h\)](#) and [1\(i\)](#)).

Orphan IFs

Several orphan IFs exist that do not easily fit into any other group ([Table 1](#)). These tend to show a very restricted tissue distribution. Exclusive to differentiated lens fiber cells, phakinin (also called CP115) and filensin (also called CP49) coassemble in a 3:1 molar ratio to produce a unique structure called ‘beaded filaments’, where 20-nm-wide beads decorate the otherwise 8-nm-wide filament in a regularly spaced pattern. Nestin was originally discovered as a marker of neuroepithelial stem cells, but is now known to exhibit a broader distribution including various multipotent progenitor cells. It is incorporated into filaments at very low stoichiometry compared to others IF proteins. Synemin, desmuslin, and paranemin coassemble with type III IFs (notably desmin) in muscle, with varying stoichiometry. Just like NF-M and NF-H, nestin, synemin, and paranemin have unusually long C-terminal tail domains that influence interaction between IFs and with other elements of the cytoskeleton.

Functions of IFs and Their Involvement in Disease

Resistance to Stress

A major, family-wide function of IFs is to endow cells and tissues with a resilient and yet flexible scaffold needed to withstand mechanical, and in some cases nonmechanical, stresses. This concept first came to light from studies in which transgenic mice expressing a mutant epidermal keratin developed severe epidermal blisters. Subsequent analysis of DNA from patients with epidermolysis bullosa simplex, an epidermal blistering disease, revealed mutations in epidermal keratins, confirming the link between mutant IF proteins and inability of cells to withstand mechanical stress. Since then, the importance of IFs and resistance to various trauma has been extended to a variety of cell types including myocytes, fibroblasts, astrocytes, and trophoblasts. In the absence of stress, these cells usually tolerate the expression of a mutated, defective IF protein even though it may lead to disorganization of their IF network. In the event of shear stress, however, the intracellular rupture of such mutant-expressing cells reveals their unusual fragility. These features can also be observed in the absence of one or more keratins, as is evidenced by the fragility of the oral mucosa in *Krt6 α / β* null mice (Figures 2(a) and 2(b)) and disruption of the keratin IF network in cultured primary keratinocytes from these mice (Figures 2(c) and 2(d)). Studies of the micromechanical properties of IF assemblies *in vitro* and in live cells in culture showed that they exhibit properties that are uniquely suited to fulfill this crucial role, and

that inherited mutations compromise these properties. The diversity of IF sequences and their differential expression may serve, at least in part, the purpose of achieving the right compromise between properties of pliability and resilience given the context-specific demands placed upon a cell or its constituents.

Insight from Laminopathies

Nuclear lamins offer tantalizing clues for additional functions. Multiple lines of evidence, adeptly reviewed in recent texts, show that the nuclear lamina participates in determining the shape, size, and integrity of the nucleus, along with the number and positioning of nuclear pores. For example, *Lmna* null mice exhibit severe growth retardation and features of muscular dystrophy (Figures 2(e)–2(h)). Inherited mutations in the *Lmna* gene are now known to give rise to a variety of clinical disorders (Table 1), which from a bird's eye view seem very distinct from one another. Premature aging conditions such as Hutchinson–Gilford progeria syndrome and atypical Werner syndrome are examples of relatively recent, and unanticipated, entries in the realm of laminopathies. While fragility of the nuclear envelope could possibly account, at least in part, for the pathogenesis of the muscle-based laminopathies, other mechanisms must be invoked for the premature aging conditions and related disorders. Among the likely possibilities are alterations in the control of gene expression and perturbations in the composition and function of the endoplasmic

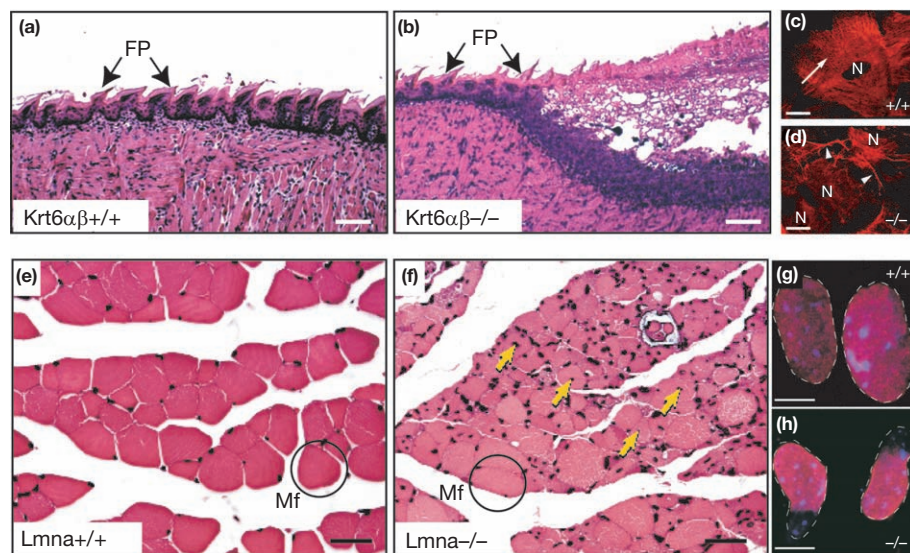


Figure 2 (a and b) Hematoxylin-eosin stained, paraffin-embedded section of wild-type (a) and *K6 α /K6 β* double null (b) mouse tongues. Tissue is oriented along its anterior-posterior axis. In the absence of the K6 proteins α and β , the posterior region of the tongue epithelium develops severe blistering owing to the fragility of filiform papillae (FP). Scale = 200 μ m. (Wong et al. 2000 *Journal of Cell Biology* 150: 921–928.) (c and d) Visualization of keratin IF networks by indirect immunofluorescence in primary culture of skin keratinocytes from wildtype (c) and *K6 α /K6 β* null mice (d). Several null cells (–/–) feature disrupted keratin IF networks (see arrowheads). Scale = 50 μ m. (Wong and Coulombe 2003 *Journal of Cell Biology* 163: 327–37.) (e and f) Aberrant localization of nuclei in muscle from *Lmna* null mice. (e) Hematoxylin-eosin stained, paraffin-embedded sections through wild-type perivertebral muscle show the peripheral localization of the nuclei in the muscle fibers. (f) In *Lmna* null muscle, myofibrils (Mf) are irregularly sized, and there is an increase in the number of nuclei, many of which are centrally located within the muscle fibers (arrows). Scale = ~15 μ m. (Sullivan et al. 1999 *Journal of Cell Biology* 147: 913–920.) (g and h) Immunohistochemical analysis of wild type and *Lmna* null mouse embryonic fibroblasts in culture. After fixation with 3% paraformaldehyde, the nuclei were labeled with DAPI (blue) and indirect immunofluorescence of Lamin B (red). Note the uniform distribution of this protein in wild-type nuclear envelopes (g), but its loss from one pole of the irregularly shaped nuclei in the *Lmna* null cells (h). Dashed line indicates boundary of nucleus. Scale = ~10 μ m. (Sullivan et al. 1999 *Journal of Cell Biology* 147: 913–920.)

reticulum, which extends into and communicates with the nuclear envelope. An impact on gene expression could result from lamins' potential involvement in basic nuclear processes such as chromatin organization, DNA replication, RNA processing, and nucleocytoplasmic exchange.

Apoptosis

Just like their nuclear counterpart, cytoplasmic IFs have evolved the ability to organize, through scaffolding, proteins and factors involved in signaling and other metabolic processes. Such interactions, for instance, clearly modulate the response of various types of cells to pro-apoptotic signals. Absence or disruption of one or more IFs in several cell types (simple epithelial cells lacking K8, keratinocytes lacking K17, and neurons containing peripherin aggregates) render cells significantly more likely to undergo apoptosis in response to tumor necrosis factor- α (TNF α) or Fas. In simple epithelial cells lacking K8, this occurs in part because K8/18 IFs contribute to the targeting of Fas receptors to the cell surface, and bind to the cytoplasmic tail domain of TNFR2, and to death domain-containing proteins such as TRADD and DEDD. In addition to influencing the decision to undergo apoptosis, IF proteins are also key in its proper execution through their caspase-mediated degradation.

Spatial Organization

Recent evidence points to a role for IFs toward spatial organization within the cell. Keratin IFs can influence the sorting events leading to the organization of the apical pole in polarized epithelial cells, and vimentin IFs promote the association of the sodium–glucose cotransporter SGLT1 with lipid rafts, and its activity, in renal tubular cells. The contribution of neurofilaments toward the radial growth of neurons, itself a crucial determinant of conduction velocity, represents another example of how IFs can promote an asymmetric cytoarchitecture. The notion that cytoplasmic IFs promote some form of spatial organization is quite surprising given their nonpolar character – the spatial code required might be provided by site-specific, dynamic phosphorylation events, and regulated interactions with associated proteins. Future efforts should lead to a broader appreciation of the many metabolic processes influenced by IFs and underlying mechanisms of how inherited IF mutations cause disease.

Cell Proliferation, Growth, and Migration, and Tissue Inflammation

The ability of several IF proteins to interact with various classes of cellular proteins (e.g., 14-3-3 adaptor proteins) accounts for their impact, again in a sequence-dependent and

context-specific fashion, toward the regulation of cell-cycle progression as well as protein synthesis, which represent key determinants of tissue growth. Yet another way in which IF proteins, and keratin in particular, can impact cell and tissue growth is via their ability to modulate, in a cell-autonomous fashion, the expression of select pro-inflammatory cytokines and chemokines. Finally, several IF proteins including vimentin, nestin, and various keratins, have been shown to impact cell migration in the physiological settings of tissue repair and/or inflammation, and in diseases such as cancer.

See also: **Cell Architecture and Function:** Desmosomes and Hemidesmosomes; Intermediate Filament Linker Proteins: Plectin and BPAG1; Keratins and the Skin; Neuronal Intermediate Filaments; Nuclear Lamina.

Further Reading

- Cabeen MT and Jacobs-Wagner C (2010) The bacterial cytoskeleton. *Annual Review of Genetics* 44: 365–392.
- Coulombe PA, Bousquet O, Ma L, Yamada S, and Wirtz D (2000) The 'ins' and 'outs' of intermediate filament organization. *Trends in Cell Biology* 10: 420–428.
- Dechat T, Pflieger K, Sengupta K, et al. (2008) Nuclear lamins: Major factors in the structural organization and function of the nucleus and chromatin. *Genes and Development* 22: 832–853.
- Erber A, Riemer D, Bovenschulte M, and Weber K (1998) Molecular phylogeny of metazoan intermediate filament proteins. *Journal of Molecular Evolution* 47: 751–762.
- Fuchs E and Cleveland DW (1998) A structural scaffolding of intermediate filaments in health and disease. *Science* 279: 514–519.
- Fuchs E and Weber K (1994) Intermediate filaments: Structure, dynamics, function, and disease. *Annual Review of Biochemistry* 63: 345–382.
- Herrmann H, Bar H, Kreplak L, Strelkov SV, and Aebi U (2007) Intermediate filaments: From cell architecture to nanomechanics. *Nature Reviews Molecular Cell Biology* 8: 562–573.
- Hyder CL, Isoniemi KO, Torvaldson ES, and Eriksson JE (2011) Insights into intermediate filament regulation from development to ageing. *Journal of Cell Science* 124: 1363–1372.
- Kim S and Coulombe PA (2007) Intermediate filament scaffolds fulfill mechanical, organizational, and signaling functions in the cytoplasm. *Genes and Development* 21: 1581–1597.
- Kim S and Coulombe PA (2010) Emerging role for the cytoskeleton as an organizer and regulator of translation. *Nature Reviews Molecular Cell Biology* 11: 75–81.
- Omary MB, Coulombe PA, and McLean WHI (2004) Intermediate filament proteins and their associated diseases. *New England Journal of Medicine* 351: 2087–2100.
- Omary MB, Ku NO, Tao GZ, Toivola DM, and Liao J (2006) "Heads and tails" of intermediate filament phosphorylation: Multiple sites and functional insights. *Trends in Biochemical Sciences* 31: 383–394.
- Szeverenyi I, Cassidy AJ, Chung CW, et al. (2008) The human intermediate filament database: Comprehensive information on a gene family involved in many human diseases. *Human Mutation* 29: 351–360.
- Worman HJ, Fong LG, Muchir A, and Young SG (2009) Laminopathies and the long strange trip from basic cell biology to therapy. *Journal of Clinical Investigation* 119: 1825–1836.

Intracellular Calcium Channels: NAADP⁺-Modulated

A A Genazzani and M Debidia, University of Cambridge, Cambridge, UK

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 469–472, © 2004, Elsevier Inc.

Glossary

cADPR An intracellular messenger derived from NAD⁺ acting on the ryanodine receptor as an agonist or a modulator.

IP₃ A second messenger derived from the cleavage of PIP₂ (phosphatidylinositol 4,5-bisphosphate) by phospholipase C which interacts with IP₃-sensitive Ca²⁺ channels causing calcium release.

NAADP An intracellular messenger derived from NADP⁺ acting on a yet unknown intracellular Ca²⁺-release channel.

Second messenger An intracellular molecule whose concentration is regulated in response to binding of a specific ligand to an extracellular receptor to transduce information.

Structure of NAADP

Nicotinic acid adenine dinucleotide phosphate (NAADP) is a pyridine nucleotide similar to nicotinamide adenine dinucleotide phosphate (NADP), except for the amide group of the nicotinamide ring, which in the NAADP molecule is replaced by a carboxyl group (Figure 1). Despite this small difference, in sea urchin eggs (the model system used to characterize this molecule), NADP is devoid of any Ca²⁺-release efficacy, suggesting that the binding site specifically recognizes this region of the molecule as a discriminator and that NAADP is not just a cellular surrogate for NADP.

NAADP-Induced Ca²⁺-Release in Invertebrates and Plants

Although the chemical existence of NAADP was reported by Bernofsky in 1980, the possibility that NAADP could be a putative intracellular messenger was hypothesized by Lee and co-workers. In 1987, they described the presence of a contaminant in commercially available NADP that could release Ca²⁺ in sea urchin eggs. Only in 1995, this contaminant was shown to be NAADP. NAADP has now been found to be a potent Ca²⁺-releasing agent in various invertebrate systems including starfish, sea snails, and sea squirts (see Table 1). Furthermore, NAADP has also been shown to release Ca²⁺ from microsomes prepared from plant preparations. Although the NAADP receptor has not yet been cloned, it presumably is a Ca²⁺ channel distinct from those described to date, as indicated by numerous biochemical evidences. Nonetheless, in a few systems, such as the skeletal muscle, NAADP might also use the ryanodine receptor as a mediator of its effect. In sea urchin eggs, the best characterized model to date, NAADP releases Ca²⁺ via an intracellular channel that is distinct from those that are responsive to inositol 1,4,5-triphosphate (IP₃) and cyclic ADP ribose (cADPR). NAADP-induced Ca²⁺ release does not exhibit any cross-desensitization with the other two mechanisms. Furthermore, it is not sensitive to heparin, ruthenium red or high concentrations of ryanodine, antagonists of the IP₃, and ryanodine channels. According to the concentrations, Ca²⁺ can both

activate and inhibit IP₃ and ryanodine receptors, providing a means of modulation of the channel independent of ligands. A major difference between the NAADP mechanism and the other two is the insensitivity of the NAADP-induced Ca²⁺ release to both cytosolic and vesicular Ca²⁺ concentrations. Furthermore, other differences exist in the channels in regard to the pH sensitivity, ion dependency, endogenous modulators, and kinetic features. Finally, specific NAADP-binding sites, insensitive to IP₃ and cADPR, have been shown in sea urchin egg microsomes.

All the efforts done to characterize the pharmacology have failed so far to produce a valid antagonist of the NAADP-sensitive mechanism. In echinoderms, it has been shown that L-type Ca²⁺-channel antagonists (and BayK8644, an agonist) are able to inhibit NAADP-induced Ca²⁺ release, but this occurs just at concentrations, which are approximately 100-fold higher than those able to inhibit voltage-operated Ca²⁺ channels. Nonetheless, it has been found that NAADP itself can be used as an antagonist since in sea urchin eggs subthreshold concentrations of this nucleotide are able to completely abrogate the calcium release caused by further NAADP elevations. Neither IP₃ nor cADPR are regulated by this mechanism and so far the significance of this process is unknown.

In sea urchin eggs, the NAADP-sensitive channel does not appear to reside on the same intracellular organelles compared to IP₃ and ryanodine receptors, as shown by membrane separation experiments. To support this, while IP₃- and cADPR-induced Ca²⁺ release are completely abolished by pretreatment of homogenates with thapsigargin, a poison of the Ca²⁺ pump that is responsible for the accumulation of Ca²⁺ in the endo(sarco)plasmic reticulum, NAADP-induced Ca²⁺-release appears to be unaffected. Recently, it has been shown that, at least in this system, the organelle on which the mechanism might reside is the reserve granule, the egg equivalent to lysosomes. Whether this will be a general feature of this mechanism across species remains to be established.

In sea urchins, starfish, and sea squirts (invertebrate systems) where the properties of NAADP on Ca²⁺ release have been investigated in intact cells, it has been shown that this Ca²⁺-release mechanism does not act in isolation from the others, but instead appears to work in concert to elicit specific responses. For example, in starfish NAADP-induced Ca²⁺ release

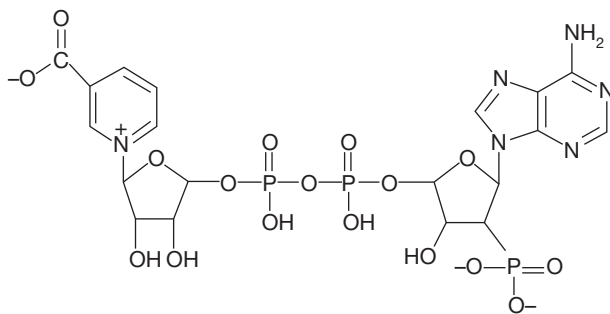


Figure 1 Structure of nicotinic acid adenine dinucleotide phosphate.

Table 1 Models described to be responsive to NAADP between 1995 and 2002

Species	Model	NAADP effect
Sea urchin	Intact egg, microsomes	Inactivation (<1 nM, IC ₅₀ 200 pM), agonist (>3 nM, EC ₅₀ 25 nM), inactivation (<1 nM, IC ₅₀ 200 pM), agonist (>3 nM, EC ₅₀ 25 nM)
Starfish	Intact egg	Agonist (1–2 μM)
Sea squirt	Intact oocyte	Agonist (EC ₅₀ 30 nM)
Sea snail ganglia	Presynaptic neuron	Agonist (10–50 μM)
Red beet	Microsomes	Agonist (96 nM)
Cauliflower	Microsomes	Agonist (1 μM), inactivation (3 nM)
Frog neuromuscular junction	Nerve terminal	Agonist (10 μM to 1 nM; liposomal delivery)
Human β pancreatic cells	Intact cells	Agonist (100 nM), inactivation (10 μM)
Human T-lymphocytes	Intact cells	Agonist (max 100 nM), inactivation (10 μM)
Mouse pancreatic acinar cells	Intact cells	Agonist (50 nM), inactivation (1–100 μM)
Rat pulmonary smooth muscle cells	Intact cells	Agonist (10 nM), inactivation (100 μM)
Rat fibroblast cells	Microsomes	Agonist (10 μM)
Rat brain	Microsomes	Agonist (EC ₅₀ 1 μM)
Rabbit heart	Microsomes	Agonist (EC ₅₀ 320 nM)
Rat mesangial kidney cells	Microsomes	Agonist (EC ₅₀ 4 μM)
A7r5, HL-60, H9c2 cell lines	Microsomes	Agonist (10 μM)

appears to be partially inhibited by the co-injection of IP₃ and ryanodine receptor antagonists, suggesting that the NAADP-induced cytosolic calcium elevations are due to both release from the NAADP channel and to amplification through IP₃ and ryanodine receptors. The extent and the manner of the cross-talk appears to be different from one system to another, but it is one of the conserved features of NAADP signaling through evolution, since it has been observed also in vertebrate systems.

NAADP-Induced Ca²⁺-Release in Vertebrates

Both microsomal and intact systems from vertebrate models have been shown to respond to NAADP (Table 1).

Furthermore, binding sites with nanomolar affinity for NAADP have been described in brain and cardiac tissues. In similar fashion to what is observed in invertebrate systems, NAADP-induced Ca²⁺ release is distinct from the IP₃ and ryanodine-mediated mechanisms when it is investigated in microsomal preparations, while in intact systems there appears to be a high degree of overlap. One of the differences observed between vertebrates and invertebrates is that the unusual desensitization property observed in sea urchin eggs does not seem to have been conserved in vertebrates. Instead, in many of the systems described (e.g., pancreatic acinar and β-cells, T-lymphocytes), NAADP appears to release Ca²⁺ in an inverted bell-shaped curve, with high concentrations unable to have any effect. Observations of cross-talk between NAADP-induced Ca²⁺ release and IP₃ and ryanodine receptors in pancreatic acinar cells and T-lymphocytes have led to the hypothesis that NAADP might act as a trigger by releasing enough Ca²⁺ to induce and/or sensitize the other two mechanisms to amplify its signal. This cross-talk means that cells might encode signals via increasing the three second messengers to different levels, instead of having the option to modulate just one messenger to different levels. The net outcome of this could be the possibility to use Ca²⁺ to specifically and unambiguously encode a higher number of intracellular messages.

At present, no extracellular message has been linked unequivocally with NAADP signaling. Nonetheless, the capacity of high concentrations of NAADP to block the release mechanism has been used experimentally to suggest that NAADP might be involved, among other functions, inolecystokinin-mediated signaling in pancreatic acinar cells, in insulin signaling in β-cells, and in the T-cell receptor/CD3 signaling.

Metabolism of NAADP

It has been shown that cADPR, a derivative of nicotinamide adenine dinucleotide (NAD), in which the nicotinamide has been cleaved and the terminal ribose is cyclized with the adenine group at the N1 position, can be synthesized by a subfamily of NAD⁺ glycohydrolases. Members of this family have been characterized in *Aplysia* and in mammalian systems. These enzymes are peculiar since they are able to catalyze both the cyclization and the degradation of cADPR. Surprisingly, in the presence of nicotinic acid, these enzymes are also capable of catalyzing a base-exchange reaction that replaces the nicotinamide on NADP with nicotinic acid to produce NAADP. Therefore, it is likely that these enzymes are central to calcium signaling, since they can catalyze the production of two signaling molecules with two distinct targets, as well as the degradation of one of these. For this mechanism to produce diversity, it could be hypothesized that the different reactions should be modulated differently by intracellular pathways. Recent evidence suggests that this might be the case. First, the base-exchange reaction seems to be favored over the cyclization in acidic conditions, although substantial amounts of NAADP are generated at physiological pH. Second, both the laboratories led by Lee and by Galione have separately shown that cyclic nucleotides can differentially regulate the production of cADPR and NAADP. In particular, in sea urchin eggs elevations of cAMP appear to favor NAADP production while elevations of

cGMP favor production of cADPR. Nonetheless, this could be due to either a differential modulation of the same enzyme or modulation of two separate enzymes. Among the enzymes described to date that can catalyze these reactions, the mammalian members, the most important of which is the lymphocyte antigen CD38, are membrane-bound ectoenzymes, with extracellular catalytic activity. Although there are some descriptions of CD38 present on intracellular membranes such as the nuclear envelope or the mitochondrion, it is likely that other members of this family will be described in the future that possess intracellular activity. Alternatively, the group led by De Flora has postulated a mechanism by which the extracellularly produced messengers can be internalized.

NAADP degradation has been shown to occur in both invertebrate and vertebrate cells. The mechanism by which this occurs is at present unknown, although it has been recently postulated that in mouse tissues a Ca²⁺-dependent phosphatase, via conversion to NAAD, might be involved in the termination of signaling by NAADP.

In most systems evaluated to date, NAADP induces Ca²⁺ release at low nanomolar concentrations, making its detection extremely difficult. It has recently been shown that sea urchin sperm contains micromolar levels of NAADP, together with a membrane-bound high-capacity enzyme for its synthesis. It is

likely that these observations will be extended to other systems with the characterization of more sensitive detection methods.

See also: **Bioenergetics:** IP₃ Receptors; Phosphatidylinositol-3-Phosphate; Ryanodine Receptor Calcium Ion Channels; Voltage-Gated Ca²⁺ Channels; **Signaling:** Phosphatidylinositol Bisphosphate and Trisphosphate; Phospholipase C.

Further Reading

- Cancela JM, Van Coppenolle F, Galione A, Tepikin AV, and Petersen OH (2002) Transformation of local Ca²⁺ spikes to global Ca²⁺ transients: The combinatorial roles of multiple Ca²⁺ releasing messengers. *EMBO Journal* 21: 909–919.
- Churchill GC, Okada Y, Thomas JM, et al. (2002) NAADP mobilizes Ca²⁺ from reserve granules, lysosome-related organelles, in sea urchin eggs. *Cell* 111: 703–708.
- Clapper DL, Walseth TF, Dargie PJ, and Lee HC (1987) Pyridine nucleotide metabolites stimulate calcium release from sea urchin egg microsomes desensitized to inositol trisphosphate. *Journal of Biological Chemistry* 262: 9561–9568.
- Genazzani AA and Billington RA (2002) NAADP: An atypical Ca²⁺-release messenger? *Trends in Pharmacological Sciences* 23: 165–167.
- Lee HC (2002) *Cyclic ADP-Ribose and NAADP Structures, Metabolism and Functions*. Boston: Kluwer.
- Patel S, Churchill GC, and Galione A (2001) Coordination of Ca²⁺ signalling by NAADP. *Trends in Biochemical Sciences* 26: 482–489.

Intracellular Calcium Waves

J T Chun and L Santella, Stazione Zoologica Anton Dohrn, Villa Comunale, Napoli, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Action potential A momentary reversal in electrical potential across a plasma membrane, as of neurons, muscles, and fertilized eggs, that occurs when a cell has been activated by a stimulus.

Calcium release-activated calcium channel protein 1 (ORAI1) A calcium selective ion channel in the cell plasma membrane that is activated upon the depletion of endoplasmic reticulum (ER) calcium stores.

Glutamate receptors Specific receptors that bind to the most common neurotransmitter in the brain, L-glutamate. There are two broad categories of glutamate receptors: ionotropic receptors and metabotropic receptors. The ionotropic glutamate receptors are ligand-gated ion channels and are further subdivided based on their ligand

specificity: α -amino-3-hydroxyl-5-methyl-4-isoxazole-propionate (AMPA), kainite, and N-methyl-D-aspartate (NMDA) types. On the other hand, metabotropic glutamate receptors transmit the signals by the action of the coupled G-proteins.

Sarco/endoplasmic reticulum Ca^{2+} -ATPase (SERCA) pump An ion pump that transfers Ca^{2+} from the cytosol to the lumen of the sarcoplasmic or endoplasmic reticulum of the cell at the expense of ATP hydrolysis.

Stromal interaction molecule 1 (STIM1) A transmembrane protein enriched in the ER that functions as a sensor for the decrease of Ca^{2+} concentration in the ER lumen.

Synaptotagmin A family of membrane-trafficking proteins that function as calcium sensors and regulate neurotransmitter and hormone release.

Phenomena

The concept that Ca^{2+} ion transduces extracellular stimuli into intracellular reaction dates back to the seminal work by Sydney Ringer in 1883. He observed that an isolated heart can beat only when Ca^{2+} is present in the medium. The significance of this observation, confirmed and appreciated decades later, is that a simple substance such as Ca^{2+} can substitute the external stimuli to activate an intracellular process. Since the same external stimuli actually induce an increase of intracellular Ca^{2+} , Ca^{2+} must have worked as a second messenger. In addition to its role in muscle contraction, Ca^{2+} serves as an intracellular regulator of diverse cell activities such as fertilization, neurotransmission, exocytosis, endocytosis, cell division, and cell death, to name a few.

That Ca^{2+} is implicated in these processes has been demonstrated either by substituting Ca^{2+} ions in the media or by using specific chelators or pharmacological agents that respectively reduce or increase the intracellular Ca^{2+} levels. Furthermore, application of electrophysiological methods measuring the membrane potential and ion currents across the membrane made it possible to study Ca^{2+} influx or efflux, while a more recent development of fluorescent Ca^{2+} indicators and the imaging technology such as rapid scanning confocal microscopy or video fluorescence microscopy has provided spatiotemporal information on Ca^{2+} mobilization inside the cell. Thus, the wave of Ca^{2+} propagating inside a living cell was first visualized in the fertilized eggs of medaka (a freshwater fish) and sea urchins by the use of photoprotein called aequorin in the late 1970s. These cells were large enough (1 mm and 80 μm , respectively) to facilitate microinjection and the imaging procedure that provided relatively low spatiotemporal resolution at that time. However, with more recent development of fluorescent protein probes being targeted to specific organelles and the chelator-type sensitive Ca^{2+} indicators (Fura, Fluo, Indo,

Calcium Green, etc.), whose emission of fluorescence is greatly enhanced upon binding to free Ca^{2+} , the intracellular Ca^{2+} wave has been observed not only in excitable cells such as neurons and myocytes but also in a variety of cell types, including pancreatic acinar cells, astrocytes, hepatocytes, fibroblasts, endothelial cells, and embryonic cells in many different physiological contexts.

Fertilization exemplifies one of the most intense intracellular Ca^{2+} signalings, as both sperm and eggs are activated by the wavelike rises of intracellular Ca^{2+} . Owing to the exceptionally large cell size, intracellular Ca^{2+} waves have been best demonstrated in the fertilized eggs. In the eggs of all animal and plant species tested, fertilization is immediately accompanied by the generation of the Ca^{2+} wave that initiates at the sperm entry site and propagates through the cytoplasm. The fertilization Ca^{2+} wave can produce either single (echinoderms, amphibians, echinuran worms, and mollusks) or multiple peaks (mammals, ascidians, and polychaete worms), depending on the species. However, the physical characteristics of the fertilization Ca^{2+} wave are highly conserved, as the average speed of its propagation (2–10 $\mu\text{m s}^{-1}$) and the amplitude of the local Ca^{2+} rise (1–5 μM) at the wave front are estimated to be in the same order of magnitude across a wide spectrum of animal species ranging from sponges to mammals. Although Ca^{2+} can also enter the fertilized egg from the extracellular space, the major part of the Ca^{2+} rise in the cytoplasm is due to the release of the sequestered free Ca^{2+} from the intracellular stores such as the lumen of the endoplasmic reticulum (ER). The wavelike propagation of the Ca^{2+} signals in the cortex of the fertilized egg is essential for the postfertilization egg activation. In echinoderm (sea urchin and starfish) eggs, the Ca^{2+} waves precede the arrival of the surface wave of the egg activation which is characterized by the gradual elevation of the vitelline layer due to the Ca^{2+} -dependent exocytosis of cortical granules. This leads to the formation of the fertilization

envelope in echinoderm eggs, which may serve as a mechanical block against polyspermy (Figures 1 and 2).

Another clear example of intracellular Ca^{2+} wave has been demonstrated in the context of muscle contraction. The three types of muscles, that is, skeletal, cardiac, and smooth muscles, all utilize Ca^{2+} as a messenger that converts depolarization of the membrane potentials to the generation of contractile force. The common theme is that an increase of cytosolic Ca^{2+} is required for myosin-based sliding of actin filaments. In skeletal muscles, nerve endings of somatic motor neurons make neuromuscular junction in which nerve-released acetylcholine binds to nicotinic receptors on the muscle membrane and produce sufficiently large ($\Delta \sim 60$ mV) depolarization at the motor end plates to ensure generation of action potentials on the sarcolemmal (plasma) membrane of the muscle.

The plasma membrane of skeletal and heart ventricular myocytes has deep invaginations called transverse tubules (t-tubules) that convey action potentials toward the cell interior. Moreover, the t-tubular membrane is in close contact with terminal cisternae of the sarcoplasmic (endoplasmic) reticulum, forming diad or triad structures. The t-tubule membrane is studded with voltage-sensitive L-type Ca^{2+} channels called dihydropyridine (DHP) receptors which are thought to be physically linked to the ryanodine receptors (RyRs) on the sarcoplasmic reticulum. Hence, the arrival of action potentials on the myocyte membrane opens not only the DHP receptor channels but also the RyRs via protein–protein interaction or calcium-induced–calcium-release mechanism (CICR), respectively (see below). The liberated Ca^{2+} in cytosol then binds to troponin and dislocates it from the actin filaments inside

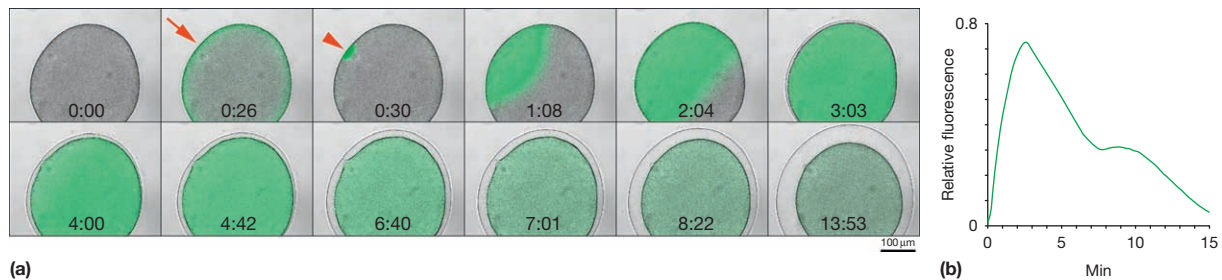


Figure 1 Propagation of the intracellular Ca^{2+} wave in the fertilized egg of starfish (*Astropecten aranciatus*) and the consequent elevation of the fertilization envelope. Mature eggs microinjected with fluorescent Ca^{2+} indicators were fertilized to monitor the changes of the cytosolic Ca^{2+} levels. (a) Transmission photomicrographs of a representative fertilized egg were superimposed with the corresponding fluorescent images of the Ca^{2+} indicator (green) at each time point. (b) Quantification of the intracellular Ca^{2+} levels in reference to the basal level. In less than a minute after the arrival of the fertilizing sperm at the jelly coat ($t = 0:00$, min:sec), a quick Ca^{2+} increase takes place in the cortex (cortical flash; arrow at 0:26), and a wave of Ca^{2+} is generated from the sperm interaction site (arrowhead) and traverses the entire cytoplasm of the egg. By the end of Ca^{2+} wave propagation, a thick wall of fertilization envelope is formed over the egg surface as a result of cortical granule exocytosis. Reproduced from figure 1 in Puppo A et al. (2008) Alteration of the cortical actin cytoskeleton deregulates Ca^{2+} signaling, monospermic fertilization, and sperm entry. *PLoS ONE* 3:e3588.

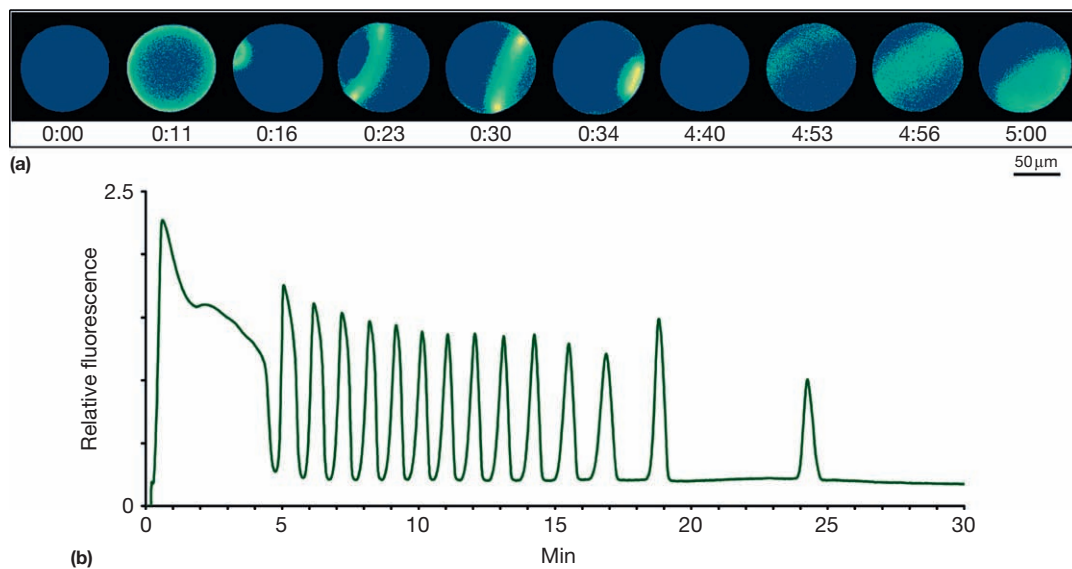


Figure 2 Propagation of the oscillating Ca^{2+} waves in the fertilized egg of ascidian (*Ciona intestinalis*). (a) Fluorescent images of instantaneous increments ($F_{\text{inst}} = [(F_t - F_{t-1})/F_{t-1}]$) of intracellular Ca^{2+} levels constituting the first two Ca^{2+} waves after fertilization. Note the cortical flash at 0:11 that was successfully visualized because of the use of the intact eggs with the chorion. After the generation of the first Ca^{2+} wave at the sperm interaction site, the following repetitive waves start from the nearby spots, but not exactly from the original spot. (b) Quantification of the intracellular Ca^{2+} levels in reference to the basal level.

sarcomere so that myosin can gain an access to exert a power stroke. In this way, the electric signal in the muscle fibers is rapidly converted to the Ca^{2+} signals to activate a contraction-relaxation cycle of the muscle. Since the propagation speed of the action potential could be on the order of a few m s^{-1} , the excitation-contraction coupling is an example of a very fast way of increasing intracellular Ca^{2+} levels.

The sheer propagation of intracellular Ca^{2+} waves in muscle cells has often been demonstrated either in dissected muscles or in cultured myocytes independent of the activities of innervating nerves. When exposed to hypertonic sucrose-Ringer solution, the frog skeletal muscle fibers display Ca^{2+} oscillation that originates from specific sites and propagates over a few hundred micrometers at the speed of $30\text{--}40\ \mu\text{m s}^{-1}$ and with the frequency of about 7 waves per minute. Once initiated, this wave is independent of extracellular Ca^{2+} but can be blocked by ryanodine, the agonist/antagonist of RyRs, implying that the intracellular wave is mediated by the internal Ca^{2+} stores. Similarly, myocytes from guinea pig ventricle have displayed submicromolar Ca^{2+} waves that rapidly propagate ($\sim 100\ \mu\text{m s}^{-1}$) down the longitudinal axis of the cell. Unlike ventricular myocytes, atrial myocytes lack t-tubules, and the main source for priming Ca^{2+} waves here is the large influx from the plasma membrane by voltage-gated calcium channels (VGCCs). The resulting Ca^{2+} sparklet diffuses and activates further Ca^{2+} release from RyRs on the sarcoplasmic reticulum through the CICR mechanism. Indeed, the cat atrial myocytes produce Ca^{2+} waves 70–100 ms after electrical stimulation, but blockade of RyRs with ryanodine restricts the Ca^{2+} increase mainly to the cell periphery. Reflecting the anatomical differences with respect to the presence of the t-tubule, the Ca^{2+} wave in rat atrial myocytes requires longer lag time (30 ms) for the Ca^{2+} level to rise in the entire cell than in ventricular myocytes, although the velocity of the Ca^{2+} wave itself is virtually the same ($>100\ \mu\text{m s}^{-1}$).

Mammalian smooth muscles are markedly different from skeletal and cardiac muscles. First of all, smooth muscles are not made of sarcomeres but are arranged in oblique bundles. Their structural organization is as diverse as their specialized functions in vascular, gastrointestinal, respiratory, ocular, urinary, and reproductive systems. Second, smooth muscles are associated with tropomyosin but lack troponin that binds Ca^{2+} and releases the inhibitory mechanism of sarcomere contraction seen in skeletal muscles. Inevitably, Ca^{2+} should initiate contraction of smooth muscles in a different way. Increased by both influx and the internal release from the sarcoplasmic reticulum, cytosolic Ca^{2+} binds to calmodulin and activates a host of proteins including the enzyme that phosphorylates myosin, for example, myosin light chain kinase. This cascade of change initiated by the Ca^{2+} signal eventually enhances myosin's adenosine triphosphatase (ATPase) activity and increases muscle contraction.

Ca^{2+} is instrumental in contraction of all smooth muscle cells, but the way intracellular Ca^{2+} level is increased and how it contributes to muscle contraction are widely variable in different cell types. Smooth muscle cells in the vascular system (blood flow) can exert prolonged tonic contraction, while other cells in bladder (micturation), uterus (labor contraction), and vas deference (ejaculation) require rapid and strong responses to the stimuli with a short period of excitation

(phasic contraction). The increase of cytosolic Ca^{2+} levels and the subsequent contraction of the smooth muscles are controlled not only by neurotransmitters but also by hormones and paracrine mechanisms. Hence, smooth muscle cells integrate multiple (sometimes contradictory) influences and produce a modulated response. For example, detrusor smooth muscle cells in the bladder display intrinsic capability of spontaneous contraction, but such endogenous contraction needs to be synchronized by the parasympathetic nerves in order to produce sufficient intravisceral pressure for micturation. During bladder emptying, the binding of neurotransmitters (acetylcholine and extracellular adenosine triphosphate (ATP)) increases the frequency of the action potentials. As each action potential is widely spread and is followed by contraction, the bladder now produces a concerted phasic contraction. This action potential and muscle contraction depend on Ca^{2+} influx, as they are prevented by inhibitors of L-type Ca^{2+} channels.

Most visceral smooth muscle cells surrounding internal organs (viscera), blood vessels, and intestinal tracts are electrically interconnected through gap junctions, which can rapidly spread action potentials to induce a unitary contraction of the entire sheets of tissues. For instance, the spontaneous Ca^{2+} transient generated in the smooth muscles of the guinea pig urinary bladder can traverse the muscle bundles as rapidly as $1.5\ \text{mm s}^{-1}$, but the propagation of this Ca^{2+} wave can be specifically disrupted by 18β -glycyrrhetic acid, a gap junction blocker. The propagation of action potentials and Ca^{2+} signals by gap junctions can thus be used as a highly efficient strategy to contract the whole tissue by using a few specialized pacemaker cells. For example, interstitial cells of Cajal are specialized smooth muscle cells of the gastrointestinal tract that generate periodic Ca^{2+} pulses and the consequent slow wave potentials to activate local population of smooth muscle cells. This slow wave potential is much more infrequent (3–12 waves per minute) than the myocardial pacemaker potentials (60–90 waves per minute). However, when the membrane depolarization surpasses the threshold, VGCC opens and the entering Ca^{2+} fires one or more action potentials. The spontaneous repetitive Ca^{2+} transients then travel through the myenteric region as a wave at the velocity of $2\ \text{mm s}^{-1}$ and activate the attached muscle fibers. The blockade of gap junctions disrupts the rhythmicity and synchronicity of the Ca^{2+} oscillations, while individual cells continue to pace independently with no reduction in the amplitude of the transients.

Intracellular Ca^{2+} increase has also been demonstrated in neurons, as it is associated with neuronal activities. In the context of excitation-secretion coupling, a large increase of cytosolic Ca^{2+} ($\sim 200\ \mu\text{M}$) was demonstrated at the synapse of squid giant axons. The firing of action potentials leads to rapid rise of this highly localized Ca^{2+} surge at the presynaptic terminals, and this Ca^{2+} increase is essential to the secretion of neurotransmitters. While squid giant axons represent a special type of axon that is formed by syncytial fusion of neurons to increase the axon diameter and electric signal conduction, similar excitation-secretion coupling involving Ca^{2+} increases also takes place in other ordinary neurons. Furthermore, intracellular Ca^{2+} can also be increased in postsynaptic neurons following neurotransmission, as has been demonstrated in a variety of cells including pyramidal neurons of the

hippocampus and neocortex, dopaminergic neurons of the midbrain, and Purkinje cells of cerebellum. In neocortical pyramidal neurons, repetitive electric stimulation at the apical dendritic shaft produces a sharp Ca^{2+} spike ($1\ \mu\text{M}$) that is followed by a larger but delayed (by 500 ms) Ca^{2+} wave ($\sim 5\ \mu\text{M}$). This latter Ca^{2+} wave (estimated $5\text{--}10\ \mu\text{m s}^{-1}$) propagates about $10\ \mu\text{m}$ in both directions and is sensitive to drugs blocking RyRs and inositol 1,4,5-trisphosphate (InsP_3) receptors (InsP_3Rs), suggesting that it is released from the intracellular stores. This slow but longer-lasting ($>1\ \text{s}$) Ca^{2+} release is not evocable in the peripheral regions beyond $150\ \mu\text{m}$ from the soma. Hence, the slow Ca^{2+} increase does not occur synchronously in all location of the cell but instead spreads as a wave over a restricted region. By contrast, the Ca^{2+} increase induced by back-propagating action potentials is much faster in response (2 ms) and propagates more than 1 mm from the soma and from the stimulation site, albeit smaller in amplitude ($<1\ \mu\text{M}$). Therefore, there are clearly distinct ways of delivering Ca^{2+} signals in neurons.

Intracellular Ca^{2+} waves were observed not only in neurons but also in astrocytes, the prevailing cell type in the brain. Filopodial extension of astrocytes envelops synaptic cleft that is occasionally flooded with neurotransmitters such as glutamate. Recent studies showed that astrocytes not only release glutamate but also propagate intracellular Ca^{2+} wave in response to extracellular glutamate or ATP. Although the exact physiological concentration of neurotransmitters in the synaptic cleft is hard to estimate, submicromolar level of glutamate and ATP can generate localized Ca^{2+} waves. At higher concentration ($10\text{--}100\ \mu\text{M}$) of glutamate, Ca^{2+} propagates not only within a single cell with a velocity of $20\ \mu\text{m s}^{-1}$ but also through adjacent cells with a slightly lowered velocity ($15\ \mu\text{m s}^{-1}$). This intercellular propagation of Ca^{2+} wave was observed to propagate nearly $770\ \mu\text{m}$ spanning more than 50 cells. The latter process is believed to be mediated through an electrically coupled syncytium, as blockers of gap junctions can inhibit the Ca^{2+} propagation. However, this type of intercellular Ca^{2+} signaling can also be mediated by extracellular ATP that serves as a diffusible messenger (Figure 3).

Similarly, a coordinated intracellular propagation of Ca^{2+} signals has been demonstrated in tracheal epithelial cells, where Ca^{2+} signaling plays a crucial role in maintaining mucociliary clearance. A mechanical distortion of the apical membrane of the tracheal epithelial cell generates a Ca^{2+} wave in the cell within 3 s, which then spreads to neighboring cells as far as $300\ \mu\text{m}$ with the velocity of $15\ \mu\text{m s}^{-1}$. The arrival of the oscillating Ca^{2+} waves then increases and sustains the ciliary beat frequency. As in astrocytes, this intercellular propagation of Ca^{2+} depends on gap junction and can also be induced by extracellular ATP that mediates autocrine or paracrine actions. Ca^{2+} waves characterized by mechanical induction and subsequent intracellular propagation via gap junctions and/or extracellular ATP have been actually reported in various mammalian cells including thymic epithelial cells ($6\text{--}10\ \mu\text{m s}^{-1}$), mast cells ($5\text{--}10\ \mu\text{m s}^{-1}$), prostate cancer cells ($15\ \mu\text{m s}^{-1}$), endothelial cells of arteries ($10\ \mu\text{m s}^{-1}$), retinal pigment epithelial cells ($10\text{--}30\ \mu\text{m s}^{-1}$), and the osteoblastic cell line ($10\ \mu\text{m s}^{-1}$).

The pancreas plays both endocrine and exocrine roles in the body of vertebrate animals. It not only releases hormones such

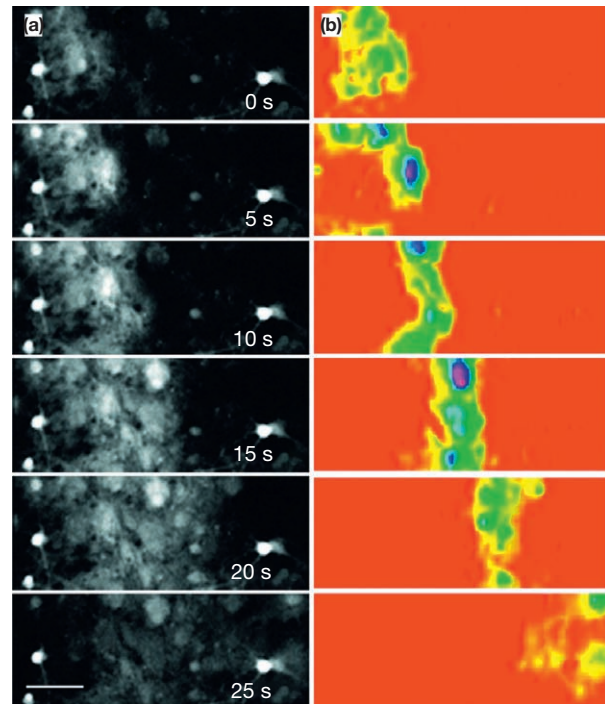


Figure 3 An example of intercellular propagation of Ca^{2+} waves in astrocytes following the stimulation of the co-cultured neurons. (a) Transmission photomicrographs of the areas confluent with astrocytes. (b) Fluorescent images of instantaneous increments of intracellular Ca^{2+} levels that propagate as a sweeping wave across the layer of astrocytes. Scale bar = $50\ \mu\text{m}$. Reproduced from figure 1 in Hung J and Colicos MA (2008) Astrocytic Ca^{2+} waves guide CNS growth cones to remote regions of neuronal activity. *PLoS ONE* 3:e3692.

as insulin, glucagon, and somatostatin (endocrine) but also excretes bicarbonate and digestive enzymes into the duct that drains to the lumen of the duodenum (exocrine). While bicarbonate is mainly provided by the cells forming the wall of the duct, the digestive enzymes are excreted by a specialized cluster of cells that resemble a lobe of berry (*acinus* in Latin) and are thus called acinar cells. This morphologically polarized cell generates and propagates intracellular Ca^{2+} wave when stimulated with acetylcholine (neurotransmitter) and cholecystokinin (peptide hormone). The wave displays agonist-specific oscillation patterns, presumably due to the distinct distribution of the receptors having different physical characteristics. In general, acetylcholine evokes sinusoidal oscillations with repetitive Ca^{2+} spikes, whereas cholecystokinin produces short-lasting (2 s) Ca^{2+} spikes with long intervals. The latter agonist produces Ca^{2+} spikes that are followed by a global and long-lasting (30 s) Ca^{2+} waves much more often than acetylcholine does. The velocity of the Ca^{2+} wave generated by acetylcholine is increased with the increasing agonist concentration and could reach nearly $100\ \mu\text{m s}^{-1}$, while the observed maximal velocity of the cholecystokinin-evoked Ca^{2+} wave is about $60\ \mu\text{m s}^{-1}$. Similar to these oscillating patterns of Ca^{2+} waves, patch pipette-injected InsP_3 produces repetitive Ca^{2+} spikes whose amplitude and frequencies can be modulated by the agonist concentration. In most cases, these waves have been observed to travel in an apical (lumen side) to basal direction. The Ca^{2+}

waves in pancreatic acinar cells are instrumental both in exocytosis of enzymes and in the activation of Cl^- channels to extrude the ions to the lumen.

An earlier example of nonexcitable cells propagating or oscillating Ca^{2+} waves was found in isolated hepatocytes of rats. When exposed to adrenergic agonists and vasopressin, these cells respond with oscillatory Ca^{2+} waves that initiate always at the specific subplasmalemmal region and propagate to the restricted adjacent subcellular regions with the speed of $20\text{--}25\ \mu\text{m s}^{-1}$. The dose of hormones does not affect the speed and amplitude (about $700\ \text{nM}$) of the Ca^{2+} wave but alters the frequency of the Ca^{2+} oscillations ($0.5\text{--}0.9$ peaks per minute).

Finally, it should be noted that different types of Ca^{2+} waves have also been observed whose propagation velocities are at least one or two orders of magnitude lower than those of the aforementioned examples. These slow waves are more easily observed along the cleavage furrows of the cells during cytokinesis in early embryos at the cleavage stage. During the first mitosis of *Xenopus* zygote, a slow-moving Ca^{2+} wave ($3\ \mu\text{m s}^{-1}$) travels along the cleavage furrows but not in the direction orthogonal to it. In medaka fish eggs, the corresponding wave apparently travels even more slowly ($0.5\ \mu\text{m s}^{-1}$). Since the cleavage furrow is formed and operated by extremely thin actinomyosin band (contractile ring), the Ca^{2+} wave may be related to the generation of the mechanical force required for cell cleavage. In addition, it may induce exocytosis to provide increased amount of plasma membrane to support cytokinesis.

Mechanisms of the Ca^{2+} Waves

The concerted increase of intracellular Ca^{2+} concentration has been often referred to as a wave in a sense that it is mobilized or periodically alternates its amplitude in space and time. However, it is rather an intuitive definition. Unlike a ripple that propagates over the water surface or the electromagnetic radiation that travels in free space, an intracellular Ca^{2+} wave is generated and propagated inside a living cell, traveling through a highly heterogeneous medium of cytoplasm that is crowded with organelles, the cytoskeleton, and numerous particles like proteins that may participate in shaping the Ca^{2+} wave by buffering or regenerating the wave. Thus, its description in mathematical terms is not a straightforward task. Despite the numerous unpredictable variables, however, several calcium formulas have been proposed to account for the particular patterns of Ca^{2+} mobilization in a given cell such as a *Xenopus* oocyte. Though contributions from minor variables are inevitably neglected, simulation of the wave by these models often approximates the experimental observation. While making a unifying model that perfectly fits all cells would be simply impossible, one could consider several basic strategies that a cell may use to propagate a Ca^{2+} wave.

First of all, Ca^{2+} can be suddenly brought from outside of the cell. To maintain cytosolic Ca^{2+} at low concentration, cells continuously pump out Ca^{2+} into extracellular space, but the trend could be reversed upon certain external cues. Depolarization of the cell membrane can quickly activate two Ca^{2+} influx pathways that either activate VGCC or reverse the action mode of $\text{Na}^+\text{--}\text{Ca}^{2+}$ exchangers on the plasma membrane.

Since depolarization is the initial cause of action potentials, Ca^{2+} influx is often associated with the propagation of action potentials, which could be as fast as $20\ \text{m s}^{-1}$ (estimation in squid giant axons). Hence, Ca^{2+} influx can provide an extremely fast way of sending Ca^{2+} signals to the other side of the cell, which is remote from the site of stimulus. However, the Ca^{2+} imported in this way usually does not propagate deep into the cytoplasm albeit its high lateral speed. One good example of this extremely fast mode of increasing Ca^{2+} concentration inside a cell is found in the fertilized eggs of echinoderms. The physical contact of the fertilizing sperm induces the depolarization of the egg membrane potential, and this immediately leads to a rise of the Ca^{2+} level in the tight subplasmalemmal regions of the entire egg surface. The rise and fall of the Ca^{2+} signal is so fast that it appears almost synchronized and is thus referred to as a cortical flash. Likewise, wiring myofibrils with t-tubules serves as a mechanism to expedite the Ca^{2+} response and synchronize the muscle contraction, although the speed of the Ca^{2+} signal itself is impeded by the diffusion step before reaching the downstream targets. In neurons, Ca^{2+} entry can also take place through ligand-gated channels such as *N*-methyl-D-aspartate (NMDA)-type glutamate receptors, which are selectively localized in dendritic spines. Whereas the Ca^{2+} influx through VGCC is fast but too short-lived to reach a high concentration, the highly localized Ca^{2+} transients fluxed through NMDA receptor last longer and thus exceed more than $1\ \mu\text{M}$ due to the long open time and the small volume of dendritic spines. On the other hand, Ca^{2+} can also enter through Ca^{2+} -permeable α -amino-3-hydroxy-5-methyl-4-isoxazole-propionate (AMPA) channels, which are localized in aspiny dendrites. Hence, Ca^{2+} ions have different physiological meanings depending on how they have entered the cell.

Intuitively, the most basic element of the intracellular Ca^{2+} wave would be diffusion. Once liberated from the mouth of ion channels, Ca^{2+} is free to diffuse to the areas of lower concentration until the dynamic equilibrium is reached. The speed of diffusion down the concentration gradient is proportional to the concentration difference between the two spots, dependent on temperature, and is inversely proportional to the square of the distance. Hence, a drop of ink dripped into a glass of water will diffuse fast in the beginning, but the speed will gradually diminish. If the propagation of Ca^{2+} in cells is solely based on diffusion, one would predict that the amplitude and the speed of the Ca^{2+} transient will dwindle as the wave propagates. However, the Ca^{2+} waves propagating in fertilized eggs and other cells can travel long distances from the origin without losing the velocity and the amplitude of the wave fronts. Hence, besides diffusion, the intracellular Ca^{2+} wave must propagate by some kind of autocatalytic reaction that replenishes the signal and facilitates its propagation.

In this regard, cytoplasm can be considered an excitable medium where the wave is amplified in the propagating field before being diminished. The cytoplasm is packed with extensive network of endoplasmic (or sarcoplasmic) reticulum that serves as the major store of Ca^{2+} inside the cell. Through SERCA pumps, Ca^{2+} is continuously pumped from cytosol into these organelles to be loosely ($K_d=0.5\text{--}2\ \text{mM}$) bound to the storage proteins inside the lumen. As a result, Ca^{2+} in lumen may reach $>1\ \text{mM}$, and the concentration of free Ca^{2+}

is at least 100 times higher than that in cytosol ($0.1\text{--}1\text{ }\mu\text{M}$). Hence, the luminal Ca^{2+} is ready to diffuse out whenever the ion channels are open on the ER membrane. The important aspect is that these ion channels (InsP₃Rs, for example) are rapidly (within $10\text{--}50\text{ ms}$) activated by Ca^{2+} itself, once the ion reaches a threshold level ($0.2\text{--}0.4\text{ }\mu\text{M}$). This cooperative positive feedback provides an essential mechanism by which a local increase of Ca^{2+} can ignite further release of Ca^{2+} ions from the nearby InsP₃Rs. As the local Ca^{2+} concentration exceeds a certain limit ($>1\text{ }\mu\text{M}$), the ion channels slowly ($1\text{--}10\text{ s}$) close down, passing the torch of Ca^{2+} signals to the next InsP₃Rs in line. This spatiotemporally coordinated opening of several channels by Ca^{2+} ion is thus termed CICR and may serve as a fundamental mechanism by which to propagate Ca^{2+} waves inside the cell without causing excessive elevation of cytosolic Ca^{2+} concentration. CICR was initially discovered as a mechanism to mobilize Ca^{2+} waves through RyRs, which is thought to be the major mediator of CICR in animal cells. As mentioned above, an antagonist (ryanodine) of this receptor blocks propagation of the Ca^{2+} waves in the frog muscles, but other cells may employ InsP₃Rs for CICR.

Besides Ca^{2+} , physiological gating agents (ligands) can also activate RyRs and InsP₃Rs. Whereas the question of cyclic ADP ribose (cADPr) being a true ligand or a modulator of RyRs may be disputable because of its delayed effect on the receptor ion channel activities, InsP₃ binds to its receptor as a ligand and directly opens the channel or allosterically modulates its sensitivity to Ca^{2+} inhibition so that the open probability of the channel is increased. Since the enzyme that hydrolyzes phosphatidylinositol 4,5-bisphosphate (PIP₂) to produce InsP₃ and diacylglycerol (DAG), that is, phospholipase C (PLC), is activated by a Ca^{2+} rise itself, this three-way causal relationship among InsP₃, Ca^{2+} , and PLC constitutes another important positive feedback loop that sustains the regenerative Ca^{2+} waves on a slower timescale of $1\text{--}2\text{ s}$. Hence, the latter mechanism termed InsP₃- Ca^{2+} coupling (ICC) could set the tone to the slower but more global Ca^{2+} response of cells.

The Ca^{2+} -releasing second messenger InsP₃ has another crucial added value, that is, InsP₃ is readily diffusible inside the cell. Whereas an individual Ca^{2+} ion does not diffuse much inside the cell due to the hundreds of binding molecules, InsP₃ is relatively free from these buffering effects. Indeed, experimental data from the cytosol of *Xenopus* oocytes estimated that the diffusion coefficient of InsP₃ ($D = 280\text{ }\mu\text{m}^2\text{ s}^{-1}$) far exceeds that of Ca^{2+} ions ($D = 13\text{ }\mu\text{m}^2\text{ s}^{-1}$ for 90 nM Ca^{2+} in cytosol). Hence, InsP₃ can diffuse $5\text{--}30$ times faster than Ca^{2+} , and its function as a mobile igniter of intracellular Ca^{2+} signaling has been clearly demonstrated in astrocytes where InsP₃ diffuses through gap junctions and releases Ca^{2+} in an array of neighboring cells. Since InsP₃ synthesis takes place in the plasma membrane, while the major internal Ca^{2+} store is the ER, the mobility InsP₃ entitles it as a traveling messenger that bridges the gap between the two sites.

Ca^{2+} transients in cells have been referred to as puff (a group of InsP₃Rs), spark (cardiac RyRs), sparklet (VGCC), and so on, depending on the nature of the elementary releasing source. For these elementary Ca^{2+} transients to develop into a traveling wave, they should be sustained by CICR. The liberated Ca^{2+} ions diffuse to the vicinal channels (RyRs and

InsP₃Rs) on the ER and make them release more Ca^{2+} ions, which diffuse further and open more channels to continue a chain reaction of regenerative Ca^{2+} release from the internal stores. As a result, Ca^{2+} propagates like toppling dominos. While RyRs underlie CICR in skeletal muscles, InsP₃Rs also contribute to CICR in many other cell types through its dual dependency on InsP₃ and Ca^{2+} . Since Ca^{2+} has limited capability of diffusion in cells, it is critical that the discrete Ca^{2+} release sites should be spaced at short intervals. This spacing of ion channels and the presence of faster moving second messenger would profoundly affect the spreading range and speed of the Ca^{2+} wave. Indeed, at low InsP₃ concentration, the release sites separated more than $6\text{ }\mu\text{m}$ tend to fail in propagating a Ca^{2+} wave. In this regard, it is worth noting that the t-tubules of muscle cells are aligned in regular arrays at $2\text{-}\mu\text{m}$ intervals and that InsP₃Rs are tightly localized at the luminal side of the pancreatic acinar cells where Ca^{2+} signals start to propagate. Likewise, during meiotic maturation of *Xenopus* eggs, InsP₃Rs on the ER become more clustered to boost the Ca^{2+} release and wave propagation.

Whereas the spatial propagation of Ca^{2+} waves has been more easily demonstrated in large cells such as oocytes, the oscillatory aspect of the Ca^{2+} waves was recognized in small cells where the length of the wave front is comparable to the cell size. In principle, periodic opening of Ca^{2+} channels in the plasma membrane can be sustained by CICR or Ca^{2+} -dependent synthesis of InsP₃ to provide a cell-wide propagation of Ca^{2+} waves. If this is timely coordinated with depletion of the internal stores or with the mechanisms that sequester or terminate Ca^{2+} transients, a Ca^{2+} oscillation can take place. Indeed, CICR can be terminated by Ca^{2+} -dependent inhibition of InsP₃Rs, and Ca^{2+} can also be quickly sequestered by mitochondria that cooperate with the ER by serving as a temporary Ca^{2+} reservoir. Thus, the interval between transients may be decided by how quickly the InsP₃-sensitive internal stores can be recharged. Like a standing wave or a blinking semaphore, this type of Ca^{2+} oscillation often contributes to biorhythmicity, for example, the way that Ca^{2+} influx contributes to creating action potentials has been well described in cardiac muscles. As in skeletal muscles, contraction of the heart muscle is initiated by action potentials. Uniquely, however, this action potential is spontaneously generated by pacemaker cells and spread into contractile cells via gap junctions, while the neurotransmitters from sympathetic and parasympathetic nerves only modulate the pacemaker rate. In the latter (myocardial contractile) cells, Na^+ -dependent membrane depolarization slowly opens VGCC, and thus the depolarized phase is much prolonged (over 200 ms , forming a plateau) to squeeze out blood in a more efficient way. On the other hand, myocardial autorhythmic (pacemaker) cells have innately unstable resting membrane potential (pacemaker potential) due to the presence of special ion channels that conduct so-called funny current, which is made of both Na^+ and K^+ currents. As Na^+ influx exceeds K^+ efflux, slow depolarization occurs. This begins to open VGCC to let in Ca^{2+} , while funny channels gradually close. Thus, it is mainly the Ca^{2+} current (not Na^+ current) that pushes the membrane potential above the threshold and causes action potentials. In these ways, Ca^{2+} is involved in shaping the characteristic trajectories of action potentials in both cardiac cell types.

Whether taking place spontaneously or induced by external stimuli, the intracellular Ca^{2+} waves usually display unique patterns characteristic of the cell type. These signature Ca^{2+} responses are shaped by the concerted action of ion channels and the mechanisms that amplify or absorb the Ca^{2+} signal. Hence, precise subcellular localization of the specific proteins being involved in these processes would be the key parameter deciding the pattern of the Ca^{2+} response. For example, neurotransmitter glutamate can either directly bind to the receptors that conduct Ca^{2+} influx (NMDA receptor) or indirectly induce Ca^{2+} influx through VGCC by depolarizing the membrane potential. Furthermore, the same glutamate can also bind to the metabotropic receptors, which activate G-proteins and ultimately release Ca^{2+} from internal stores by stimulating PLC activity and InsP_3 production. Depending on the receptor types and the downstream G-protein subunits that activate different PLC subtypes, the same glutamate produces either single Ca^{2+} transient or an oscillatory Ca^{2+} response. Obviously, the spatiotemporal dynamics of these Ca^{2+} responses from the same agonist would be widely variable depending on the pathways.

It has been demonstrated that the major part of Ca^{2+} signals in fertilized eggs comes from internal stores. However, small cells depend more on external Ca^{2+} . In cells with relatively small capacity for intracellular Ca^{2+} storage (e.g., lymphocytes), depletion of the Ca^{2+} stores in the ER is sensed by the Ca^{2+} -binding protein STIM1 that forms a Ca^{2+} channel complex with ORAI1, and it leads to Ca^{2+} entry from the extracellular space. This slow (<100 s) but sustained store-operated Ca^{2+} (SOC) entry is the major pathway for Ca^{2+} influx in nonexcitable cells and provides a prolonged intracellular Ca^{2+} increase that may be used for slow cell responses such as changes of gene expression. Hence, cells have diverse repertoire of increasing and mobilizing intracellular Ca^{2+} , and the spatiotemporal profiles of the resulting Ca^{2+} waves are as diverse.

Physiological Significance of the Ca^{2+} Waves

Why does a Ca^{2+} signal carry such a specific pattern of wave propagation or oscillation? As an answer, it is generally conceived that the physiological meaning of external stimuli is often encoded in the magnitude, frequency, duration, and spatial organization of the Ca^{2+} signals. As mentioned earlier, the concentration of hormone is reflected by the frequency of Ca^{2+} oscillations in hepatocytes. For another example, the Ca^{2+} concentration inside the proximal dendrites of pyramidal neurons is quickly elevated to a certain level by a train of action potentials, and here again the amplitude of the Ca^{2+} level and its fluctuation frequency at the steady state are proportional to the frequency of the action potential. The physiological significance of the amplitude and frequency of Ca^{2+} transients has been demonstrated in neurites and growth cones. Whereas a burst of Ca^{2+} (~1 μM) virtually stops neurite elongation and growth cone motility, a more moderate physiological Ca^{2+} transients modulate both neurite outgrowth and path-finding activities of growth cones in a frequency-dependent manner. Hence, fine-tuning of Ca^{2+} transients has a fundamental effect on cell activity.

To produce a specific physiological response from a particular Ca^{2+} signal, the specific spatiotemporal pattern of the

Ca^{2+} profile has to be translated to a biochemical language. In cells, deciphering these codes is usually done by Ca^{2+} -binding proteins having a wide spectrum of binding affinity toward Ca^{2+} . For a protein such as calmodulin that has slow kinetics of Ca^{2+} binding, a Ca^{2+} transient decaying too fast would not be recognized at all, while it may be audible to certain other proteins that are more Ca^{2+} sensitive. On the other hand, a prolonged exposure to Ca^{2+} would activate multiple effectors, some of which might be even toxic to the cell. In this regard, delivering Ca^{2+} signals by oscillation might serve as a way to activate some exquisite effectors having slow Ca^{2+} -binding kinetics without causing such a deleterious effect. Thus, with the arrival of each repetitive Ca^{2+} spike, calmodulin would have more probability to bind Ca^{2+} and activate kinases and other specific targets. Discerning Ca^{2+} signals in one way or another would be utterly important because Ca^{2+} does so many diverse things in a cell that may be even contradictory sometimes. cAMP-response element-binding protein (CREB), a transcription factor implicated in learning and memory, is affected by Ca^{2+} since it is a substrate of Ca^{2+} /calmodulin-dependent kinases. Interestingly, CREB is activated only by Ca^{2+} pulses of certain temporal patterns (e.g., 1.8 s bursts separated by 1 min or 9 s bursts separated by 5 min). As Ca^{2+} can stimulate phosphatases as well, the temporal pattern of the Ca^{2+} signal might work as a specific code to phosphorylate and thereby activate CREB in the face of a counteracting pathway that dephosphorylates it.

The Ca^{2+} wave traveling from one point to another is a reasonably fast way of intracellular communication. The arrival of a fertilizing spermatozoon has to be heralded to the other parts of the egg so that it can start to coordinate the subsequent cell-wide changes such as wavelike exocytosis and rearrangement of the actin cytoskeleton. The Ca^{2+} surge in the fertilized eggs activates Ca^{2+} /calmodulin-dependent kinases and calcineurin, launching cascades of fine-tuning phosphorylation events that lead to cell-cycle progress and proper embryonic development. Hence, the changes in intracellular Ca^{2+} concentration can be converted to other forms of signaling in cells. The Ca^{2+} wave can even go beyond the cell boundary through gap junctions, as was exemplified in astrocytes. The velocity of intracellular Ca^{2+} waves could vary several orders of magnitude depending on the underlying mechanisms. Ca^{2+} signaling associated with action potentials and membrane depolarization provides the fastest method of delivering Ca^{2+} to the remote but superficial subcellular domains. On the other hand, the mechanism based on CICR promotes further progression of the Ca^{2+} waves into the deeper cytoplasmic territories.

Ironically, however, the most fundamental character of Ca^{2+} as a second messenger might lie in the fact that Ca^{2+} does not diffuse well inside a cell. In a typical cell with 300- μM binding protein concentration, it is estimated that an average Ca^{2+} ion migrates no further than 0.1–0.5 μm , lasting only about 50 μs before encountering a binding protein. As a substance, Ca^{2+} is an extremely localized second messenger despite the wave propagation. Owing to the tendency of low diffusion, the concentration of Ca^{2+} at the mouth of ion channels reaches over 100 μM , but the radius of its effect range is limited to merely 0.1 μm for 30 μs . Though spatially restricted, this signal is strong enough to activate proteins with low Ca^{2+} affinities such as synaptotagmin at the nerve

terminals or the vicinal ion channels responding to Ca^{2+} . The Ca^{2+} ions may diffuse further from the sphere of the primary impact, but would become low in concentration and be able to activate only those proteins having relatively higher Ca^{2+} affinities. Hence, how and where the Ca^{2+} puff was generated in a cell, and what target molecules were present in that micro- or nano-domain, would decide the physiological consequence of the Ca^{2+} wave. In this regard, the Ca^{2+} release machinery and the downstream effectors need to be strategically distributed in the proper subcellular microdomains in order to carry out the intended action.

Concluding Remarks

As was briefly reviewed in this article, signaling through a local or cell-wide increase of intracellular Ca^{2+} concentration has been observed in a wide variety of cells. Several molecular mechanisms explaining the phenomena of Ca^{2+} influx from the extracellular space and the efflux from the internal stores have been postulated mainly based on voltage-gated and ligand-gated Ca^{2+} channels. In parallel to the identification of ion channels, a number of agonists and second messengers that regulate the activity of the ion channels have been identified and characterized. Interestingly, molecules that initially appeared as simple metabolites such as InsP_3 , cADPr, and nicotinic acid adenine dinucleotide phosphate (NAADP) turned out to be diffusible second messengers that release Ca^{2+} from internal stores. While cADPr regulates the channel activity of RyRs and thereby contributes to the Ca^{2+} wave, NAADP is known to release Ca^{2+} from a different kind of internal store or from extracellular space. The NAADP-dependent Ca^{2+} signaling mechanism is not sensitized by Ca^{2+} , but the Ca^{2+} ions released from this route can still induce CICR from RyRs and InsP_3 Rs. Hence, it bears emphasis that any mechanism that causes Ca^{2+} influx or efflux counts in making an intracellular Ca^{2+} wave.

With no doubt, many things still remain to be learned about the generation and propagation of intracellular Ca^{2+} waves. Recently, novel classes of ion channels that are gated by mechanical stretch have been reported, but not yet characterized in great detail. Implying a potential involvement of such an intriguing mechanism in intracellular Ca^{2+} signaling, it has been demonstrated that mechanically stimulated myocytes produce Ca^{2+} influx that evolves into a wave and propagates even to the adjacent cells through gap junctions. On other occasions, it was also suggested that stretch-activated ion channels might be accountable for a certain class of Ca^{2+} waves whose velocity does not fall into the typical range that can be explained by VGCC or CICR mechanisms. Adding more interest to the topic, it has been suggested that the actin cytoskeleton might be able to modulate ion channel activities and intracellular Ca^{2+} signaling. In support of the idea, it has been reported that InsP_3 Rs are linked to actin filaments in mammalian cells and that the amplitude and the kinetics of Ca^{2+} rises in starfish eggs are modulated by structural alteration of the actin cytoskeleton. Being highly concentrated in cells ($>100 \mu\text{M}$) and organized as a dense meshwork subjacent to the plasma membrane, the actin filaments not only serve as a mechanical barrier for Ca^{2+} diffusion but might also modulate the activity of mechanosensitive ion channels in the vicinity. Furthermore, in a more radical view, it has also been suggested that actin tread-milling

could be used as a backup mechanism to temporarily store and release Ca^{2+} in a certain circumstance. Actin may bind substantial amount of Ca^{2+} when the local concentration of Ca^{2+} becomes high enough to be comparable with Mg^{2+} (near the mouth of ion channels). Once the Ca^{2+} -bound actin is incorporated into actin filaments, the Ca^{2+} -binding site is tucked in the groove of the F-actin helix, and thereby Ca^{2+} is protected from dissociation. While this process of actin polymerization can thus serve as a way to store Ca^{2+} , depolymerization of the microfilaments reverses the process and releases Ca^{2+} . Since an increase of Ca^{2+} can activate actin-binding proteins such as gelsolin that sever actin filaments, a wave of actin depolymerization, once started, might lead to the release of more Ca^{2+} and further depolymerization of actin. Hence, if Ca^{2+} release and actin depolymerization can form a positive loop similar to what we saw in RyRs and InsP_3 Rs, in theory, such a mechanism might also contribute to CICR in a cell.

See also: **Bioenergetics: Calcium Signaling by Cyclic ADP-Ribose and NAADP.**

Further Reading

- Allbritton NL and Meyer T (1993) Localized calcium spikes and propagating calcium waves. *Cell Calcium* 14: 691–697.
- Berridge MJ (2006) Calcium microdomains: Organization and function. *Cell Calcium* 40: 405–412.
- Carafoli E, Santella L, Branca D, and Brini M (2001) Generation, control, and processing of cellular calcium signals. *Critical Reviews in Biochemistry and Molecular Biology* 36: 107–260.
- Clapham DE (2007) Calcium signaling. *Cell* 131: 1047–1058.
- Chun JT and Santella L (2009) Roles of the actin-binding proteins in intracellular Ca^{2+} signalling. *Acta Physiology (Oxford)* 195: 61–70.
- Gilliot I, Payan P, Girard JP, and Sardet C (1990) Calcium in sea urchin egg during fertilization. *International Journal of Developmental Biology* 34: 117–125.
- Hung J and Colicos MA (2008) Astrocytic Ca^{2+} waves guide CNS growth cones to remote regions of neuronal activity. *PLoS ONE* 3: e3692.
- Jaffe LF (1991) The path of calcium in cytosolic calcium oscillations: A unifying hypothesis. *Proceedings of the National Academy of Sciences USA* 88: 9883–9887.
- Jaffe LF (1993) Classes and mechanisms of calcium waves. *Cell Calcium* 14: 736–745.
- Jessup C and Haccard O (2007) Fertilization: Calcium's double punch. *Nature* 449: 297–298.
- Keizer J, Li YX, Stojilković S, and Rinzl J (1995) InsP_3 -induced Ca^{2+} excitability of the endoplasmic reticulum. *Molecular Biology of the Cell* 6: 945–951.
- Krebs J and Michalak M (eds.) (2007) *New Comprehensive Biochemistry, Vol. 41: Calcium: A matter of Life and Death*. Amsterdam: Elsevier.
- Lange K (1999) Microvillar Ca^{2+} signaling: A new view of an old problem. *Journal of Cellular Physiology* 180: 19–34.
- Meyer T (1991) Cell signaling by second messenger waves. *Cell* 64: 675–678.
- Puppo A, et al. (2008) Alteration of the cortical actin cytoskeleton deregulates Ca^{2+} signaling, monospermic fertilization, and sperm entry. *PLoS ONE* 3: e3588.
- Ross WN, Nakamura T, Watanabe S, Larkum M, and Lasser-Ross N (2005) Synaptically activated Ca^{2+} release from internal stores in CNS neurons. *Cellular and Molecular Neurobiology* 25: 283–295.
- Santella L (2005) NAADP: A new second messenger comes of age. *Molecular Interventions* 5: 70–72.
- Silverthorn DU (2010) *Human Physiology – An Integrated Approach*, 5th edn. San Francisco, CA: Pearson Education.
- Sneyd J, Keizer J, and Sanderson MJ (1995) Mechanisms of calcium oscillations and waves: A quantitative analysis. *FASEB Journal* 9: 1463–1472.

Relevant Websites

- <http://www.youtube.com> – Propagation of calcium waves in cultures glial cells.
- <http://www.plosone.org> – The image and video animation of a calcium wave at the fertilization site of the starfish egg.

Ion Channel Protein Superfamily

W A Catterall, University of Washington, Seattle, WA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Gating The intrinsic regulatory process through which ion channels are opened and closed by conformational changes of the ion channel protein.

Ion channel A transmembrane protein that forms a pore through which ions move down their electrochemical gradients into or out of cells or intracellular organelles.

Rectifying An ion channel in which ion flow is favored in one direction, for example, inwardly rectifying potassium channels that conduct potassium ions inward more readily than outward, even though outward potassium movement is the physiologically relevant direction.

Voltage-gated channels Ion channels whose gating is regulated by changes in transmembrane voltage.

Electrical signals control contraction of muscle, secretion of hormones, sensation of the environment, processing of information in the brain, and output from the brain to peripheral tissues. In excitable cells, electrical signals also have an important influence on intracellular metabolism and signal transduction, gene expression, protein synthesis and targeting, and protein degradation. In all these contexts, electrical signals are conducted by members of the ion channel protein superfamily, a set of more than 140 structurally related pore-forming proteins. In addition, members of this protein family are crucial in maintaining ion homeostasis in the kidney and in many different cell types, and they also participate in calcium-signaling pathways in nonexcitable cells. In this article, the founding members of each family of ion channel proteins that are structurally related to the voltage-gated ion channels are introduced, and an overview of their structure, function, and physiological roles is presented. References are made to one or two recent review articles on each channel family to provide a broad introduction to the literature.

Voltage-Gated Sodium and Calcium Channels

The founding members of this superfamily in terms of its discovery as a separate protein are the voltage-gated sodium channels. These channels are specialized for electrical signaling and are responsible for the rapid influx of sodium ions that underlies the rising phase of the action potential in nerve, muscle, and endocrine cells. Neurotoxin labeling, purification, and functional reconstitution showed that sodium channels from mammalian brain contain voltage-sensing and pore-forming modules in a single protein complex of one principal α -subunit of 220–260 kDa and one or two auxiliary β -subunits of 36 kDa. The α -subunits of sodium channels contain four homologous domains, each containing six hydrophobic transmembrane segments. A membrane-reentrant loop between the fifth and sixth (S5 and S6) transmembrane segments forms the narrow extracellular end of the pore while the S6 segments form the intracellular end (Figure 1, red). The pore is formed in the center of a pseudosymmetric array of the four domains, and a single α -subunit containing four domains is able to receive voltage signals and activate its intrinsic pore. The channel responds to voltage by virtue of its S4 segments (Figure 1, green), which

contain repeated motifs of a positively charged amino acid residue followed by two hydrophobic residues, and move outward under the influence of the membrane electric field to initiate a conformational change that opens the pore. Drugs that block the pore of sodium channels are important as local anesthetics, antiarrhythmic drugs, and antiepileptic drugs. The four β -subunits consist of an amino-terminal extracellular immunoglobulin-like domain, a single transmembrane segment, and a short intracellular segment. They both modulate the gating of the α -subunits and serve as cell adhesion molecules by interaction with extracellular matrix, membrane, and cytoskeletal proteins. Nine voltage-gated sodium channel α -subunits, designated $\text{Na}_v1.1$ – 1.9 , have been functionally characterized. They comprise a single family of proteins with greater than 70% amino acid sequence identity in their transmembrane segments. One additional related α -subunit, which defines a second subfamily, $\text{Na}_v2.1$, is known but is apparently not voltage gated.

Voltage-gated calcium channels are the key signal transducers of electrical signaling, converting depolarization of the cell membrane to an influx of calcium ions that initiates contraction, secretion, neurotransmission, and other intracellular regulatory events. Skeletal muscle calcium channels first identified by drug labeling, purification, and functional reconstitution have a principal $\alpha1$ -subunit of 212–250 kDa, which is similar to the sodium channel α -subunit but is associated with auxiliary $\alpha2\delta$ -, β -, and γ -subunits that are unrelated to the sodium channel auxiliary subunits. Complementary DNA (cDNA) cloning and sequencing showed that the $\alpha1$ -subunit of calcium channels is analogous to the sodium channel α -subunits in structural organization (Figure 1) and is 25% identical in amino acid sequence in the transmembrane regions. As for the sodium channel α -subunit, the calcium channel $\alpha1$ -subunit is sufficient to form a voltage-gated calcium-selective pore by itself. Ten functional calcium channel $\alpha1$ -subunits are known in vertebrates, and they fall into three subfamilies that differ in function and regulation. The Ca_v1 family ($\text{Ca}_v1.1$ – 1.4) conducts L-type calcium currents that initiate contraction, endocrine secretion, and synaptic transmission at the specialized ribbon synapses involved in sensory input in the eye and ear. L-type calcium currents are also important regulators of gene expression and other intracellular processes. Blockers of $\text{Ca}_v1.2$ channels are important in the therapy of cardiovascular diseases, including hypertension, cardiac arrhythmia, and angina

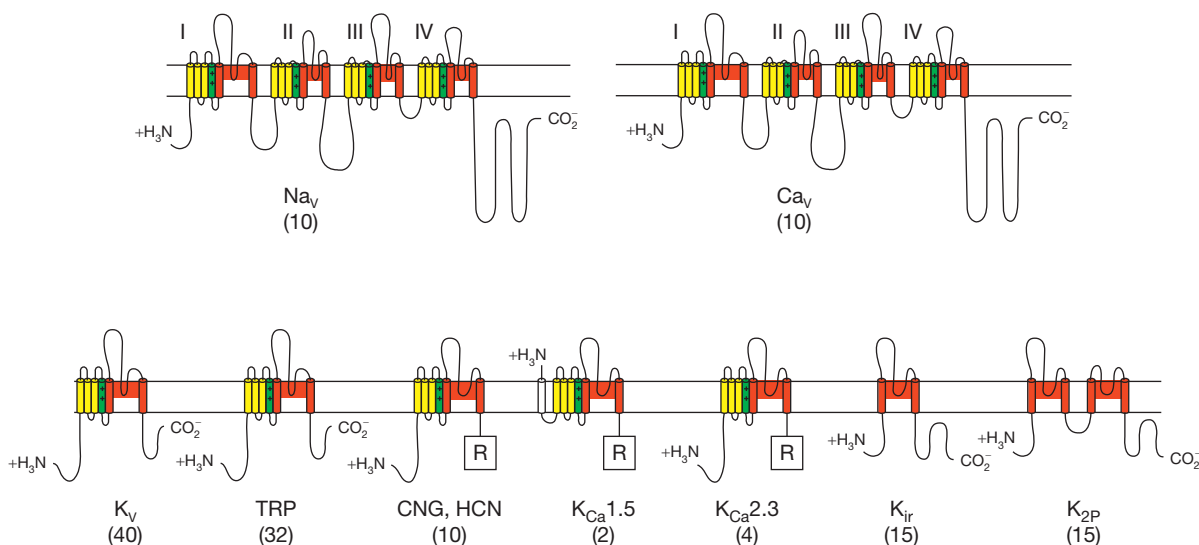


Figure 1 The voltage-gated ion channel protein family. The different members of the ion channel protein family structurally related to the voltage-gated ion channels are illustrated as transmembrane folding diagrams in which cylinders represent probable transmembrane α -helices. Red, S5–S6 pore-forming segments; green, S4 voltage sensor; yellow, S1–S3 transmembrane segments; R, regulatory domain that binds intracellular second messengers. The numbers of family members in the human genome are given in parentheses.

pectoris. The Ca_v2 family of calcium channels ($\text{Ca}_v2.1$ – 2.3) conducts N-, P/Q-, and R-type calcium currents that initiate fast synaptic transmission at synapses in the central and peripheral nervous systems and are blocked specifically by peptide neurotoxins from spider and cone snail venoms. The Ca_v3 family of calcium channels ($\text{Ca}_v3.1$ – 3.3) conducts T-type calcium currents that are important for repetitive action potential firing of neurons in the brain and in the pacemaker cells of the sino-atrial node in the heart. The functional and regulatory properties and protein–protein interactions of these channels are adapted to their different roles in electrical signaling and cellular signal transduction.

Voltage-Gated Potassium Channels

Voltage-gated potassium channels are activated by depolarization, and the outward movement of potassium ions through them repolarizes the membrane potential to end action potentials, hyperpolarizes the membrane potential immediately following action potentials, and plays a key role in setting the resting membrane potential. In this way, potassium channels control electrical signaling and regulate ion flux and calcium transients in nonexcitable cells. The first voltage-gated potassium channels were cloned from *Drosophila* based on a mutation that causes the *Shaker* phenotype. They are composed of four transmembrane subunits, each of which is analogous to a single domain of the principal subunits of sodium or calcium channels (Figure 1). The voltage-gated potassium channels are remarkable for their diversity. They include 40 different channels that are classified into 12 distinct groups based on their amino acid sequence homology (K_v1 – K_v12). These α -subunits can assemble into homo- and heterotetramers within each subfamily, leading to a wide diversity of different channel complexes. The K_v1 channels also associate with $\text{K}_v\beta$ -subunits, which are intracellular proteins. The K_v4 subunits associate

with KChIPs, intracellular calmodulin-like calcium-binding proteins that regulate channel gating. Some of the K_v7 and K_v11 α -subunits associate with single membrane-spanning subunits (minK, MiRP1, and MiRP2), which regulate their gating properties. The diversity of potassium channels allows neurons and other excitable cells to precisely tune their electrical signaling properties by expression of different combinations of potassium channel subunits.

Second-Messenger-Gated and Sensory Ion Channels

In addition to the voltage-gated sodium, calcium, and potassium channels, the voltage-gated ion channel family includes additional members that are gated by a combination of voltage and binding of second messengers, including calcium, cyclic nucleotides, and lipid messengers. These channels have six transmembrane segments and a generally similar structure to voltage-gated potassium channels, but contain in addition a regulatory domain attached at their carboxy terminus.

Calcium-activated potassium channels ($\text{K}_{\text{Ca}}1$ – 3) have an ion channel domain attached to an intracellular calcium-dependent regulatory domain in their C-terminus. They are activated by a combination of depolarization and binding of calcium ions. For the $\text{K}_{\text{Ca}}1$ channels, calcium binding to a regulatory domain in the C-terminus (R, Figure 1) enhances activation of the channels by depolarization, resulting in synergistic gating of channel activity by voltage and intracellular calcium concentration. $\text{K}_{\text{Ca}}1$ channels have an extra transmembrane segment at the N-terminus (S0) and associate with $\text{K}_{\text{Ca}}\beta$ -subunits, which are 2-TM intrinsic membrane proteins. For $\text{K}_{\text{Ca}}2$ or $\text{K}_{\text{Ca}}3$ channels, calcium binds to a constitutively associated calmodulin molecule bound to the C-terminal regulatory domain and activates the channels without a voltage stimulus. The $\text{K}_{\text{Ca}}1$ – 3 channels serve to repolarize excitable cells following single-action potentials or trains of action

potentials, during which calcium entry increases the cytosolic calcium levels and activates them. One of their prominent roles is to define the length of action potential trains in neurons by activating in response to calcium influx through voltage-gated calcium channels and thereby repolarizing and hyperpolarizing the cell. Similarly, in smooth muscle cells they hyperpolarize the membrane and relax the smooth muscle. The K_{Ca4} and K_{Ca5} channels resemble K_{Ca1} channels in structure, but are not regulated by intracellular Ca^{2+} . $K_{Ca4.1}$ is regulated by Na^+ , $K_{Ca4.2}$ is regulated by Cl^- , and K_{Ca5} is regulated by OH^- . The nomenclature of these channels will likely be changed to reflect their different mode of regulation in the future.

The structurally related cyclic nucleotide-gated (CNG) ion channels (six family members) and hyperpolarization-activated CNG (HCN) ion channels (four family members) are composed of four subunits that each have a 6-TM voltage-gated ion channel domain attached to an intracellular cyclic nucleotide-binding domain at their C-terminus. Like K_v channels, they form heterotetramers within each structural subfamily, and this contributes to their diversity of function. These channels were first purified and cloned from retinal rods, where they conduct the dark current that underlies visual signal transduction. Upon capture of photons by rhodopsin, activation of a G-protein-signaling pathway leads to the activation of a cyclic nucleotide phosphodiesterase that hydrolyzes cyclic guanosine monophosphate (cGMP). Reduction in the cGMP concentration in the rod outer segment closes these dark-current channels, causing hyperpolarization and initiation of visual signal transduction to the neural cells of the retina. Closely related CNG channels are responsible for olfactory signal transduction in the nasal epithelium. The more distantly related HCN channels are involved in cyclic nucleotide-regulated pacemaking in cells that generate repetitive action potentials, like the sino-atrial node pacemaker cells of the heart and the persistently firing neurons of the thalamus in the brain. They are also involved in sensation of sour tastes in taste buds. Remarkably, they are activated by hyperpolarization (rather than depolarization) via their S4 voltage sensor, and their opening by hyperpolarization is greatly enhanced by simultaneous binding of cyclic adenosine monophosphate (cAMP) and/or cGMP.

The transient receptor potential (TRP) channels were discovered first in *Drosophila*, where the founding member is the site of mutations that cause the TRP phenotype that alters visual signal transduction in the eye. Subsequent to that work, a family of 32 related channels, which can be divided into six subfamilies, have been defined in mammals. These channels are weakly voltage sensitive, are either calcium selective or nonselective cation channels that conduct calcium, and are activated by a range of signals including heat, cold, and intracellular lipid messengers. The TRPC subfamily of channels is most closely related to the classical *Drosophila* TRP channel. The TRPV subfamily of channels is related to the vanilloid receptor (TRPV1), a channel protein that serves as a receptor for the hot chili pepper ingredient capsaicin, a vanilloid compound. It also responds to heat and mediates noxious thermal pain sensations. The TRPM subfamily is most closely related to melastatin, first discovered as a tumor marker in melanoma. TRPM7, the first member of this family to be functionally

expressed, is slightly calcium selective and contains an intrinsic protein kinase domain that may autophosphorylate to activate its ion channel domain. The TRPP, CatSper, and TRPML subfamilies of ion channel subunits are also related to TRP channels. CatSper plays an essential role in calcium signaling in sperm. The TRPP channels are involved in sensing fluid transport in the kidney and other epithelial tissues. TRPML channels may be involved in auditory signaling in the ear and in calcium transport regulating lysosome function.

Inwardly Rectifying Potassium Channels and Their Relatives

The inwardly rectifying potassium channels (K_{ir}) are the structurally simplest ion channels, containing four subunits, each having two transmembrane segments and a membrane-reentrant pore loop between them (Figure 1). As their name implies, they conduct potassium more effectively inward than outward, but their outward conductance of potassium is physiologically important to polarize cells near the resting membrane potential. They are structurally related to the S5 and S6 segments of the voltage-gated ion channels, and their pore loop forms the extracellular end of the pore as in the K_v family of proteins. The two transmembrane segments and the pore loop of a related bacterial potassium channel have been shown directly by X-ray crystallography to form an ion-selective pore (Figure 2(a)). Many inwardly rectifying potassium channels are gated by intracellular divalent cations that bind in their pore, including Mg, spermine, and spermidine. The X-ray structure of a related vertebrate inwardly rectifying potassium channel reveals an extended pore region that projects into the cytoplasm and contains divalent cation-binding sites that may control ion flow through the pore. Other K_{ir} channels are gated by direct binding of G protein $\beta\gamma$ -subunits and inositol phospholipid messengers. Remarkably, adenosine triphosphate (ATP)-sensitive potassium channels are composed of an inwardly rectifying potassium channel core in association with the much larger sulfonylurea receptor (SUR) protein, a member of the ABC transporter family that binds adenosine diphosphate and ATP and regulates the function of the inwardly rectifying potassium channel. These channels are crucial regulators of insulin secretion from pancreatic beta cells controlled by intracellular glucose and ATP levels, and SUR is the molecular target for the sulfonylurea drugs used in treatment of diabetes.

Two-Pore-Motif Potassium Channels

The most recently discovered family of ion channels are potassium channels that have two repeats of the inward rectifier pore motif and are designated K_{2P} (Figure 1). The channels function as open rectifiers; that is, they are constitutively open and constantly drive the membrane potential toward the potassium equilibrium potential by mediating outward potassium conductance. They are thought to be important for setting the resting membrane potential of excitable cells at a negative value near the equilibrium potential for potassium of -80 mV. Although their physiological regulatory mechanisms

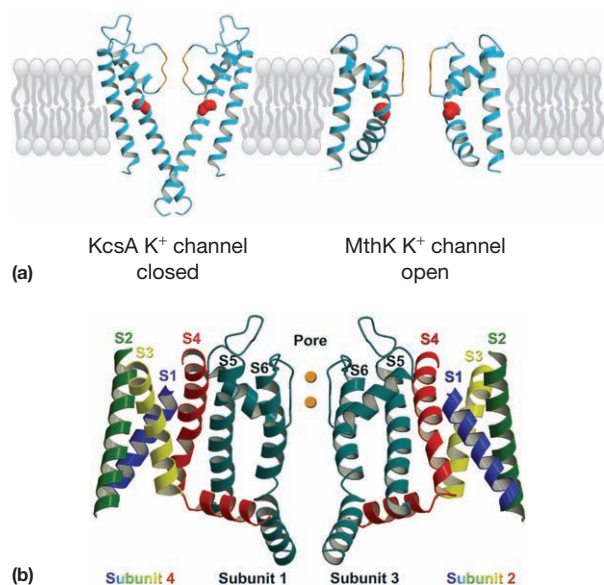


Figure 2 Structures of K⁺ channels. (a) Open and closed conformations of K⁺ channel pore structures. Two subunits of the bacterial potassium channel KcsA representing a closed conformation and MthK representing an open conformation are shown in the figure. The selectivity filter is orange and the outer helix (M1) is depicted adjacent to the lipid bilayer. The inner helix (M2) is marked with space-filling model of the conserved glycine residue (red), which is thought to be critical for the bending of M2 helix in the open conformation. (b) Cross section through the three-dimensional structure of a tetrameric K_v1.2 channel. Center, the pore regions S5-P-S6 of two subunits, designated 1 and 3, in blue-green. Left, S1 through S4 of subunit 4, whose S5-P-S6 segments project in front of the plane of the figure and form the near side of the pore. Right, S1 through S4 segments of subunit 2, whose S5-P-S6 segments project behind the plane of the figure and form the far side of the pore. As illustrated, the voltage-sensing domain composed of the S1–S4 segments is positioned to interact with the S5-P-S6 segments of the adjacent subunit located clockwise (as seen from the extracellular side) in the symmetrical tetrameric array.

are not yet fully defined, they may be an important target for the actions of inhalation anesthetics, which activate these channels and thereby could prevent action potential firing in the brain during general anesthesia.

Structure and Function

Viewed from the perspective of the entire ion channel protein family, the functions of these proteins can be divided into three complementary aspects: ion conductance, pore gating, and regulation. All members of the family share a common pore motif, with variations appropriate to determine their different ion selectivity. It is likely that this pore motif first evolved in the bacterial ion channels that resemble inward rectifiers. The X-ray crystal structure of one of these channels has shown that the narrow outer mouth of the pore is formed by the pore loops between the M1 and M2 segments that are analogous to the S5 and S6 transmembrane segments of the 6-TM family members while the length of the pore is formed by a tilted bundle, inverted teepee arrangement of the M2 segments (Figure 2(a)). This structure suggests that the pore is closed

at its intracellular end and discriminates ions at the narrow ion selectivity filter at its extracellular end. This model of pore formation fits well with a wealth of structure–function data on the voltage-gated ion channels and the CNG channels. The closure of the pore at the intracellular end suggests that this is the site of pore gating. Support for this idea comes from classical biophysical studies and from more recent work. The structure of a calcium-activated bacterial potassium channel with calcium bound shows that the crossing of the M2 segments at the intracellular end of the pore is opened by bending of the helix, consistent with the hypothesis that this is the mechanism of pore opening (Figure 2(a)).

The different protein families within the ion channel superfamily are regulated by different mechanisms. The simplest family members, the inwardly rectifying potassium channels, are gated extrinsically by divalent cations binding in the pore or by binding of other regulators. The CNG ion channels and the calcium-activated potassium channels are gated by ligand binding to their carboxy-terminal regulatory domains. These ligand-induced conformational changes are thought to exert a torque on the connected S6 segments and force the opening of the pore bundle of S6 helices. In addition, the HCN channels and the K_{Ca}1 channels are also gated synergistically by changes in membrane potential and are highly sensitive to small voltage changes.

The voltage-gated sodium, calcium, and potassium channels are primarily gated by changes in membrane potential. Their regulation depends on the S1–S4 segments of these channels, which can be viewed as an evolutionary addition to the pore-forming segments. The structure of this bundle of four alpha helices is illustrated in Figure 2(b). The detailed mechanism of voltage-dependent gating is the subject of intense research. However, there is clear structure–function evidence for an outward and rotational movement of the positively charged S4 segments during the gating process, and it is assumed that the conformational change in the voltage-sensing module accompanying this outward movement exerts a torque on the bundle of four S6 segments at the intracellular end of the pore forcing the bundle to open. This conformational change is thought to be mediated by bending of the S6 helix at conserved glycine or proline hinge regions.

Voltage-dependent gating of the sodium and calcium channels is also modulated by binding of calcium/calmodulin and other effectors to the C-terminal domain, suggesting that these interactions may also exert a torque on the S6 segments and thereby alter the energetics of gating.

Conclusion

The voltage-gated ion channel superfamily is one of the largest families of signaling proteins, following the G-protein-coupled receptors and the protein kinases in the number of family members. The family is likely to have evolved from 2-TM ancestors such as the modern inwardly rectifying potassium channels and bacterial potassium channels. Modular additions of intracellular regulatory domains for ligand binding and a 4-TM transmembrane domain for voltage gating have produced extraordinarily versatile signaling molecules with capacity to respond to voltage signals and intracellular effectors and to

integrate information coming from these two types of inputs. The resulting signaling mechanisms control most aspects of cell physiology and underlie complex integrative processes such as learning and memory in the brain and coordinated movements of muscles. The evolutionary appearance and refinement of these signaling mechanisms is one of the hallmark events allowing the development of complex multicellular organisms.

See also: **Bioenergetics: Voltage-Dependent K⁺ Channels; Voltage-Gated Ca²⁺ Channels.**

Further Reading

- Catterall WA (2000) From ionic currents to molecular mechanisms: The structure and function of voltage-gated sodium channels. *Neuron* 26: 13–25.
- Catterall WA (2000) Structure and regulation of voltage-gated calcium channels. *Annual Review of Cell and Developmental Biology* 16: 521–555.
- Goldstein SA, Bockenhauer D, O’Kelly I, and Zilberberg N (2001) Potassium leak channels and the KCNK family of two-P-domain subunits. *Nature Reviews Neuroscience* 2: 175–184.
- Jan LY and Jan YN (1997) Cloned potassium channels from eukaryotes and prokaryotes. *Annual Review of Neuroscience* 20: 91–123.
- Long SB, Campbell EB, and MacKinnon R (2005) Crystal structure of a mammalian voltage-dependent Shaker family K⁺ channel. *Science* 309: 897–903.
- Matulef K and Zagotta WN (2003) Cyclic nucleotide-gated ion channels. *Annual Review of Cell and Developmental Biology* 19: 23–44.
- Montell C (2001) Physiology, phylogeny, and functions of the TRP superfamily of cation channels. *Science’s STKE* 90: RE1.
- Patel AJ, Lazdunski M, and Honore E (2001) Lipid and mechano-gated 2P domain K⁺ channels. *Current Opinion in Cell Biology* 13: 422–428.
- Reimann F and Ashcroft FM (1999) Inwardly rectifying potassium channels. *Current Opinion in Cell Biology* 11: 503–508.
- Vergara C, Latorre R, Marrión NV, and Adelman JP (1998) Calcium-activated potassium channels. *Current Opinion in Neurobiology* 8: 321–329.
- Yu F, Yarov-Yarovoy V, Gutman GA, and Catterall WA (2005) Overview of molecular relationships in the voltage gated ion channel superfamily. *Pharmacological Reviews* 57: 387–395.
- Zhou Y, Morais-Cabral JH, Kaufman A, and MacKinnon R (2001) Chemistry of ion coordination and hydration revealed by a potassium channel–Fab complex at 2.0 Å resolution. *Nature* 414: 43–48.

IP₃ Receptors

K Mikoshiba and C Hisatsune, RIKEN Brain Science Institute, Saitama, Japan

© 2013 Elsevier Inc. All rights reserved.

Glossary

Agonist A substance that binds to a receptor and causes its activation.

Ataxia A neurological symptom that shows dysfunction of the coordinate movement.

Ca²⁺ oscillation Fluctuation of intracellular Ca²⁺ concentration.

Endoplasmic reticulum An intracellular organelle that stores Ca²⁺.

Phosphoinositides A glycerol backbone attached to two fatty acids (stearoyl at position 1 and arachidonyl at position 2) and inositol (linked at position 3 by a

phosphodiester bond). The inositol headgroup can also be further phosphorylated.

Phospholipase C A class of enzymes that catalyze cleavage of the diacylglycerolphosphate bond of a glycerophospholipid. In the process, phosphatidylinositol 4,5 bisphosphate (PIP₂) is cleaved into diacylglycerol (DAG) and IP₃.

Synaptic plasticity Change of synaptic transmission efficacy between neurons. At long-term potentiation, the efficacy is increased for long time; at long-term depression, the efficacy is decreased for long time. Both are important neurochemical basis of learning and memory.

A Brief History

The importance of intracellular Ca²⁺ concentration had been noticed since 1883 when Sidney Ringer discovered the requirement of Ca²⁺ in the contraction of frog hearts. From that time, the crucial roles of cytosolic Ca²⁺ elevation in various physiological events such as contraction, secretion, and development had been noticed, and the mechanisms by which extracellular multiple stimuli induce the cytosolic Ca²⁺ dynamics had been investigated. In the early 1950s, Hokin and Hokin found that acetylcholine increases the turnover of phosphoinositides in the pancreatic acinar cells, and, several years later, Bob Michell proposed a physiological link between extracellular stimuli-induced phosphoinositide hydrolysis and intracellular Ca²⁺ elevation. After a long debate on whether Ca²⁺ signals were causes or results of phosphoinositide hydrolysis, the production of IP₃ from PIP₂ by phospholipase C was confirmed, and Micheal Berridge, Robin Irvine, Iren Schulz, and their colleague demonstrated that IP₃ induces Ca²⁺ release from the internal Ca²⁺ stores in 1983. An attempt to purify the IP₃ binding protein to characterize the molecule was initiated by many research groups including universities and pharmaceutical companies. The IP₃ receptor protein was also characterized as a phospho- and glycoprotein abundant in the cerebellum, and named PCPP-260, GP-A, or P₄₀₀ which was absent in the cerebellum of ataxic mutant mice. General molecular properties of IP₃ receptor were made clear when the whole sequence of P₄₀₀ was determined and P₄₀₀ was found to be the IP₃ receptor by Katsuhiko Mikoshiba in 1989.

Since then, countless researchers have analyzed the nature of IP₃ receptors such as the subtypes in diverse organisms, the affinity for IP₃, the expression patterns in tissues, conformation, and regulatory mechanisms by Ca²⁺, adenosine triphosphate, and numerous proteins. Now, it is well established that IP₃ receptor is a cation-selective channel localized at endoplasmic reticulum and generates complex patterns of intracellular Ca²⁺ signals, which is crucial for many physiological events (Figure 1).

Structure of IP₃ Receptors

IP₃ receptor has been found in organisms from nematode to human. Three subtypes of IP₃ receptors encoded by distinct genes (~2700 amino acid residues) are reported in vertebrates, whereas only a single gene is reported in organisms such as *Drosophila melanogaster* and *Caenorhabditis elegans*. The three subtypes of IP₃ receptors have distinct features in the affinity for IP₃, tissue distribution, the Ca²⁺ sensitivity, and the modulation by various associated proteins. IP₃ receptor is constituted of four subunits forming a Ca²⁺-permeable pore at the center. Each subunit of IP₃ receptor has three domains, the N-terminal cytoplasmic domain, a transmembrane domain near the C-terminus, and the short C-terminal cytoplasmic domain (also named as a gate keeper domain) (Figure 2). The large N-terminal cytoplasmic domain is further divided into three domains, the N-terminal coupling domain, the IP₃ binding domain, and the internal coupling domain.

Recently, crystal structures of the N-terminal coupling domain and the IP₃ binding domain were resolved (Figure 3). The N-terminal coupling domain (residues 1–225) has a globular head domain containing the β -trefoil fold and an arm domain constituted of a helix–turn–helix (Figure 3(a)). Interestingly, the deletion of the N-terminal coupling domain resulted in no functional Ca²⁺ channels. Further, the N-terminal coupling domain determines the specific affinity for IP₃ by suppressing the IP₃-binding core region: without this region, the three subtypes of IP₃ receptors show enhanced IP₃ affinity and presented the similar affinities for IP₃ among three isoforms. Therefore, the region is originally named as ‘N-terminal suppressor domain’. On the other hand, the IP₃ binding domain (residues 224–604) forms the L-shaped structure composed of the three domains: the N-terminal β -trefoil domain, the C-terminal α -helical domain, and a hinge region linking the N- and the C-terminal domains (Figure 3(b)). The cleft between the N- and the C-terminal domains contains a cluster of basic residues on the surface, which contributes to the coordination with the three phosphoryl groups of IP₃.

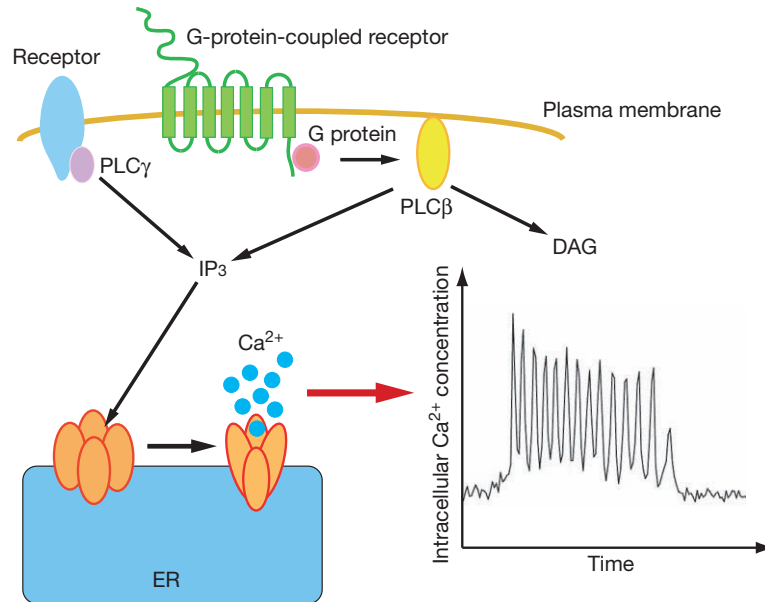


Figure 1 The signal transduction cascades for Ca^{2+} mobilization via IP_3 receptor. Extracellular stimuli stimulate phospholipase C ($\text{PLC}\beta$ or $\text{PLC}\gamma$) activities through two different types of receptors. Seven membrane-spanning G-protein-coupled receptors activate $\text{PLC}\beta$ through the dissociated α -subunit of G proteins bound to the receptors. $\text{PLC}\gamma$ is activated when it associates with and phosphorylated by receptors with either intrinsic or associated protein tyrosine kinase activity. Phospholipase C hydrolyzes PIP_2 in the plasma membrane and produces IP_3 and diacylglycerol. Then, the produced IP_3 binds to and activates the IP_3 receptor on the endoplasmic reticulum (ER) and generates the complex patterns of intracellular Ca^{2+} concentration.

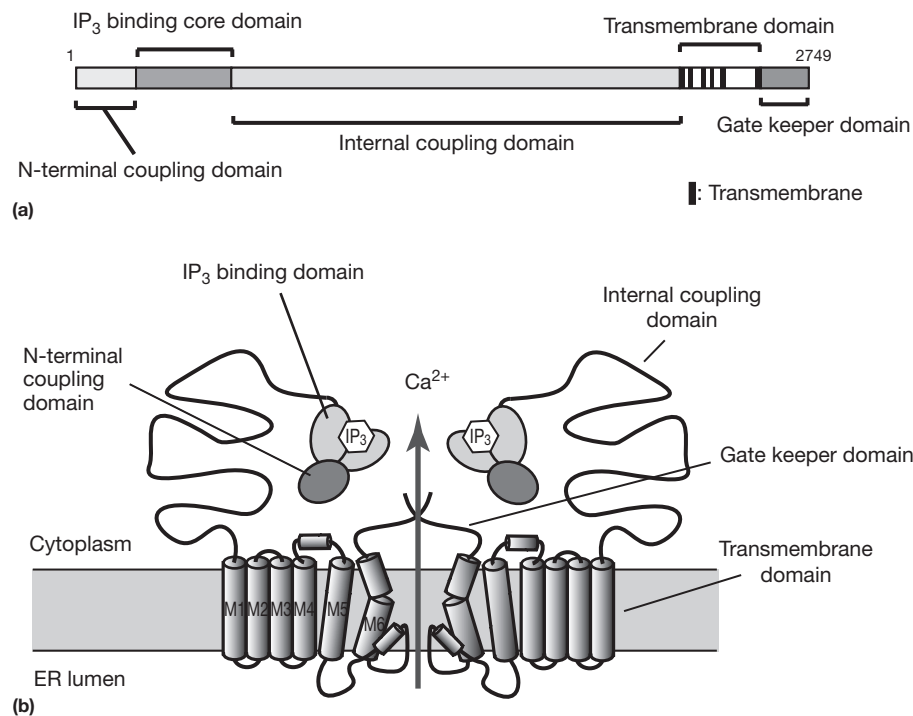


Figure 2 (a) Structure of the mouse IP_3 receptor type 1. (b) Topology of the IP_3 receptor. IP_3 receptor is a large protein and is constituted from the four homo- or hetero-tetramers. Two monomers of the four subunits are illustrated. Each subunit has six transmembranes (TM1–6) and the loop domain between M5 and M6 contributes to a pore formation of IP_3 receptor. IP_3 binds to the IP_3 binding domains near the NH_2 -terminus, which somehow triggers the gating of the channel pore localized near the COOH -terminus. Modified from figure 24-1 in Lajtha A (ed.) (2009) *Handbook of Neurochemistry and Molecular Neurobiology. Neural Signaling Mechanisms*, 3rd edn., ch. 24, pp. 442–455 and ch. 31, pp. 566–575. New York: Springer.

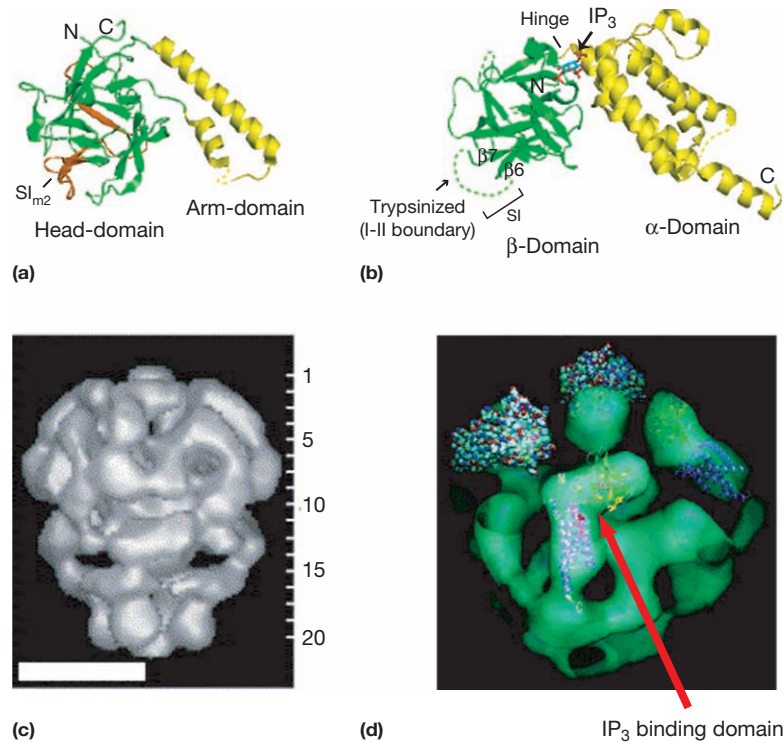


Figure 3 3D structure of mouse IP₃ receptor type 1. (a) Crystal structure of the N-terminal coupling domain. (b) Crystal structure of the IP₃ binding domain in the presence of IP₃. (c) Three-dimensional structure of mouse IP₃ receptor type 1 in the absence of IP₃ and Ca²⁺. The scale bar represents 100 Å. (d) Fitting of the modified crystal structure of IP₃ binding domain into the density map of the whole receptor. The IP₃ binding region is shown in both a ribbon diagram and a space-filled model. (a) Reproduced from figure 2(a) in Bosanac I, Yamazaki H, Matsu-Ura T, et al. (2005) Crystal structure of the ligand binding suppressor domain of type I inositol 1,4,5-triphosphate receptor. *Molecular Cell* 17: 193–203. (b) Reproduced from figure 2(a) in Bosanac I, Alattia JR, Mal TK, et al. (2002) Structure of the inositol 1,4,5-triphosphate receptor binding core in complex with its ligand. *Nature* 420: 696–700, with permission. (c, d) Reproduced from figures 2(a) and 4d from Sato C, Hamada K, Ogura T, et al. (2004) Inositol 1,4,5-triphosphate receptor contains multiple cavities and L-shaped ligand-binding domains. *Journal of Molecular Biology* 336: 155–164, with permission.

The transmembrane domain has six membrane-spanning helices (TM1–TM6) and is responsible for the tetramer formation of four subunits of IP₃ receptor (Figure 1(b)). The TM5–TM6 region of each subunit of IP₃ receptor contributes to a pore formation. High homologies of the pore region of IP₃ receptor with other cation channels suggest the conservation of structural organization.

Three-dimensional structure of IP₃ receptor in the IP₃-unbound state has also been resolved at high resolution (15 Å) using cryo electron microscopic images of the receptor. The tetrameric IP₃ receptors have a hot-air balloon-like shape (total height of 231 Å) with the spherical cytoplasmic domain (diameter of 175 Å) and the square-shaped luminal domain (side length of 96 Å) (Figure 3(c)). The transmembrane domain is 30 Å in thickness. Prominent L-shaped densities were found in the spherical cytoplasmic region of the three-dimensional structure, and estimated as the L-shaped IP₃ binding domains determined by X-ray described above (Figure 3(d)).

Regulation of IP₃ Receptors by IP₃ and Ca²⁺

The accumulating evidence suggests that IP₃ receptor requires not only IP₃ but also Ca²⁺ to open. Therefore, both IP₃ and Ca²⁺ are regarded as co-agonists of the receptor. However, the

effect of Ca²⁺ on the channel activities is more complex than IP₃: Ca²⁺ affects IP₃ receptor as an agonist at the lower concentration, while it inhibits the channel activity at the higher concentration. Thus, the effect of Ca²⁺ on the IP₃ receptor is biphasic, and this property is called as a bell-shaped Ca²⁺ dependence of IP₃ receptor. Further, IP₃ concentration also modulates the effect of Ca²⁺: at the lower IP₃ concentration, IP₃ receptors are sensitive to Ca²⁺ and tends to close before full activation, while a higher concentration of IP₃ relieves the inhibition by Ca²⁺. These fundamental complex regulations of IP₃ receptor by Ca²⁺ and IP₃ are thought to generate the various intracellular Ca²⁺ patterns, such as Ca²⁺ wave and Ca²⁺ oscillation. How Ca²⁺ affects the channel activities of the receptor is not clear, but the eight putative Ca²⁺ binding sites and several calmodulin/Ca²⁺-binding sites are proposed in the IP₃ receptor.

Physiological Roles of IP₃ Receptors

The three types of mammalian IP₃ receptors (IP₃R1, IP₃R2, and IP₃R3) show 60–70% homologies, but their distributions in tissues are quite different. While IP₃R1 is predominantly expressed in the central nervous system, especially in Purkinje cells in the cerebellum, IP₃R2 and IP₃R3 are rather broadly

expressed in almost all tissues with different ratios. Using the pharmacological tools, the functional inhibitory antibody, and a gene knockout technology, the physiological roles of IP₃Rs in various phenomena have been revealed including development, contraction, proliferation, and apoptosis. During development, IP₃R is involved in fertilization. In many species, when the sperm is attached to the unfertilized eggs, a large Ca²⁺ transient occurs at the contact site and it propagates to the opposite side across, like a wave. IP₃R is essential for achieving the Ca²⁺ waves, and, if inhibited, some early and late events of egg's activation are disturbed. IP₃R is also involved in determination of dorso-ventral axis formation in *Xenopus*: injection of functionally inhibitory antibodies for IP₃R into ventral side of early embryos induces an ectopic dorsal axis. Deletion of IP₃R in *Drosophila* leads to the lowered levels of a hormone ecdysone, and the mutant results in lethal because of the delayed molting.

In the mammalian brain, IP₃R1 plays crucial roles in synaptic plasticity and neuronal development. The mice deficient in IP₃R1 represent tremor and cerebellar ataxia, and show the deficit of the cerebellar long-term depression, a form of synaptic plasticity. On the other hand, long-term potentiation in hippocampus is enhanced. The importance of IP₃R1 in the human brain function is also demonstrated by recent studies, which point out mutation or deletion of the IP₃R1 gene underlies the cause for patients of spinocerebellar ataxia type 15 and 16. Further, the disturbed regulation of IP₃R activity by associated proteins including Huntingtin is proposed to be involved in the onset of disease in mice model.

IP₃R2 and IP₃R3 in mice play important roles in exocrine secretion, such as saliva, pancreatic juice, and olfactory mucus. The phenotype is like human Sjogren syndrome, where IP₃ receptor antibodies are detected in the serum of nearly 50% of the patients. IP₃R3 contributes to the signal transduction of taste perception in taste buds.

See also: **Bioenergetics:** Calcium Oscillations; Inositol Trisphosphate and Calcium Signaling; Nuclear Calcium in Synapse-to-Nucleus Communication: Common Regulator of Persistent Adaptations in the Nervous System.

Further Reading

- Berridge MJ (1997) Elementary and global aspects of calcium signalling. *Journal of Physiology* 499: 291–306.
- Berridge MJ and Irvine RF (1989) Inositol phosphates and cell signalling. *Nature* 341: 197–205.
- Bosanac I, Alattia JR, Mal TK, et al. (2002) Structure of the inositol 1,4,5-trisphosphate receptor binding core in complex with its ligand. *Nature* 420: 696–700.
- Bosanac I, Yamazaki H, Matsu-Ura T, et al. (2005) Crystal structure of the ligand binding suppressor domain of type 1 inositol 1,4,5-trisphosphate receptor. *Molecular Cell* 17: 193–203.
- Foskett JK, White C, Cheung KH, and Mak DO (2007) Inositol trisphosphate receptor Ca²⁺ release channels. *Physiological Reviews* 87: 593–658.
- Furuichi T, Yoshikawa S, Miyawaki A, et al. (1989) Primary structure and functional expression of the inositol 1,4,5-trisphosphate-binding protein P400. *Nature* 342: 32–38.
- Lajtha A (ed) (2009) *Handbook of Neurochemistry and Molecular Neurobiology. Neural Signaling Mechanisms*, 3rd edn, chapters 24 and 31, pp. 442–455, pp. 566–575. New York, NY: Springer.
- Michell RH (1975) Inositol phospholipids and cell surface receptor function. *Biochimica et Biophysica Acta* 415: 81–147.
- Mikoshiba K (2007) IP₃ receptor/Ca²⁺ channel: From discovery to new signaling concepts. *Journal of Neurochemistry* 102: 1426–1446.
- Sato C, Hamada K, Ogura T, et al. (2004) Inositol 1,4,5-trisphosphate receptor contains multiple cavities and L-shaped ligand-binding domains. *Journal of Molecular Biology* 336: 155–164.
- Streb H, Irvine RF, Berridge MJ, and Schulz I (1983) Release of Ca²⁺ from a nonmitochondrial intracellular store in pancreatic acinar cells by inositol-1,4,5-trisphosphate. *Nature* 306: 67–69.
- Taylor CW and Laude AJ (2002) IP₃ receptors and their regulation by calmodulin and cytosolic Ca²⁺. *Cell Calcium* 32: 321–334.
- Van de Leemput J, Chandran J, Knight MA, et al. (2007) Deletion of ITPR1 underlies ataxia in mice and spinocerebellar ataxia 15 in humans. *PLoS Genetics* 3: e108.

Iron–Sulfur Proteins

R Cammack, King's College London, London, UK

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Helmut Beinert[†], Jacques Meyer, and Roland Lill, volume 2, pp. 482–489, © 2004, Elsevier Inc.

Glossary

Fe–S active sites Sites in proteins containing iron ligated to inorganic sulfide and/or cysteine sulfur.

Iron–sulfur clusters The form in which irons and sulfides are organized in Fe–S proteins.

Iron–sulfur proteins Proteins that contain iron and sulfur, active mostly (even if not exclusively) in electron transfer.

Iron–Sulfur Clusters

Nomenclature and Shorthand Designations

The conventions for the nomenclature of iron–sulfur (Fe–S) clusters and Fe–S proteins are found in the 1989 recommendations of the Nomenclature Committee of the International Union of Biochemistry (IUB), now International Union of Biochemistry and Molecular Biology (IUBMB). The components of the cluster core are placed into brackets, and the oxidation level of the core (not including that of ligands) is given as a superscript $[2\text{Fe–}2\text{S}]^{2+}$. The possible oxidation states of the species that can be formed are indicated as shown: $[2\text{Fe–}2\text{S}]^{2+,+}$. If the kind of ligands to this cluster is to be specified, the alternative designation $[\text{Fe}_2\text{S}_2(\text{RS})_2(\text{L}_2)]^{2-}$ should be used.

Structures

The simplest Fe–S protein and also the simplest building block of Fe–S proteins is the mononuclear thiolate complex $[\text{Fe}(\text{Cys})_4]^{2-/-}$ as it occurs in rubredoxins (Rds). However, in the great majority of Fe–S proteins the iron is present in clusters, with the irons linked by sulfides. The cluster types are shown in [Figure 1](#). The coordination is tetrahedral throughout. As evident from the figure, the S–S distances are ~ 1.3 times longer than the Fe–Fe distance, which produces the distorted cubic structure typical for $[4\text{Fe–}4\text{S}]$ clusters. The $[2\text{Fe–}2\text{S}]$ and $[4\text{Fe–}4\text{S}]$ clusters are the predominant species, and the $[3\text{Fe–}4\text{S}]$ cluster can be considered as a 4Fe cluster with one Fe missing. The cluster structures shown are, to some extent, interconvertible and the 3Fe cluster is central to all cluster interconversions. By partial denaturation of the protein, it can also be converted to a linear form: $(\text{Cys})_2\text{FeS}_2\text{FeS}_2(\text{Cys})_2$. The 3Fe cluster can serve as starting material for producing hetero-metal clusters with metals such as Zn, Ni, Co, Cd, Ga, Tl, and Mo; however, such clusters are not found in nature. Those that occur have modified structures and are formed by specific synthetic systems. The 4Fe to 3Fe cluster conversion is also used in nature to condition clusters for use as active sites in enzymes and for sensing functions.

Stability

Sulfur is a weak-field ligand, allowing the iron to remain in the high-spin state, which makes the ligands relatively labile as

compared to strong ligands such as cyanide. Fe–S clusters are rather malleable, and from high-resolution X-ray structures it is apparent that they can be distorted by influences from the protein and can be adapted for various conditions and functions. The isolated clusters undergo facile ligand exchange with thiol compounds, a process that is relevant to the interconversions that occur during the assembly of Fe–S proteins. Fe–S clusters are sensitive to oxidants such as oxygen and its reduction products, and to NO, and also to an excess of their own constituents, namely S^{2-} and RS^- , to dithiothreitol and dithionite, and toward hydrolysis at high or low pH. As mentioned above, Cys ligands belonging to protein frameworks are required for the formation of stable clusters. Replacement of one Cys with another functional residue is allowed in some clusters; however, the only clusters that are known to be stable with two non-Cys ligands are those with two histidine (His) ligands, as they occur naturally in the so-called ‘Rieske’ proteins. While polypeptide chains are mandatory for the stabilization of Fe–S clusters, their protecting effect is very variable; the lifetime of Fe–S proteins in air may vary from seconds to weeks. As naturally occurring Fe–S clusters require protein ligands for support, their functions are considered later in this article (see section [Iron–Sulfur Proteins](#)).

Electronic Structure and Magnetism

The electronic charge in Fe–S clusters is concentrated at their sulfides and Cys ligands. It has been shown by X-ray near-edge spectroscopy (XANES) that all bonds involving sulfur are highly covalent, with those involving sulfide exhibiting 2–3 times the covalency of the thiolate bonds. By contrast, the unpaired spin density resides primarily on the irons. Fe–S clusters owe many of their characteristic properties to the ordering of their spin systems. This ordering is dominated by three forces: (1) the tendency to pair the spins antiparallel, which is expressed in the magnitude of the spin coupling parameter J ; (2) the tendency to pair the spins parallel, expressed in the parameter B , the resonance integral (this tendency arises when an extra electron is present, leading to spin-dependent electron delocalization (SDD)); and (3) the dynamic trapping energy, arising through vibronic coupling with groups in the neighborhood. On one-electron reduction of $[2\text{Fe–}2\text{S}]^{2+}$ (2Fe^{III}) clusters, the trapping energy leads almost invariably

[†]Deceased.

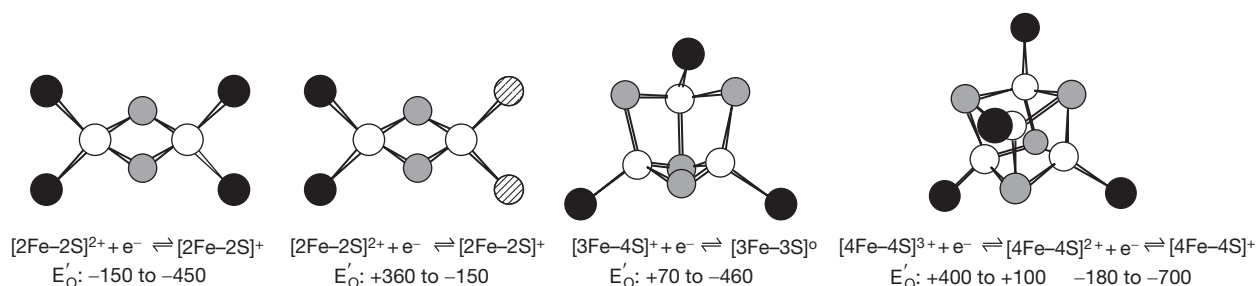


Figure 1 Models for structures of 2Fe, 3Fe, and 4Fe clusters. White: Fe; black: cysteine; gray: sulfide sulfurs; and hatched: histidine nitrogens. Modified from Holm RH, Kennepohl P, and Solomon EI (1996) Structural and functional aspects of metal sites in biology. *Chemical Reviews* 96: 2239–2314.

to antiparallel coupling and localization of the electron at one Fe (1Fe^{III} , 1Fe^{II}) so that the system spin $S = 5/2 - 4/2 = 1/2$. In $[4\text{Fe}-4\text{S}]^{2+}$ clusters, there are two pairs of parallel-coupled (according to SDD) mixed valence irons (2Fe^{II} , 2Fe^{III}), both with $S = 5/2 + 4/2 = 9/2$, which then couple antiparallel to $S = 0$. On addition of an electron, there will be one mixed valence pair and two Fe^{II} ; thus, $S = 9/2 - 2 \times 4/2 = 1/2$, that is, paramagnetism. If instead one electron is removed from the $2+$ level to form $[4\text{Fe}-4\text{S}]^{3+}$ (3Fe^{III} , 1Fe^{II}), there are one mixed valence pair and two Fe^{III} ; thus, $S = 2 \times 5/2 - 9/2 = 1/2$ and paramagnetism. The $[4\text{Fe}-4\text{S}]^{3+}$ level has been first observed in so-called high-potential iron protein (HiPIP); these proteins are now counted among ferredoxins (Fds) with high redox potentials. Occasionally, $[4\text{Fe}-4\text{S}]^+$ clusters with an $S = 3/2$, a mixed $1/2$ and $3/2$, or higher ground-state spins have been observed.

Iron-Sulfur Proteins

Nomenclature

Small proteins, whose function is electron transfer, are called rubredoxins (Rds) if they contain only mononuclear $[\text{Fe}(\text{Cys})_4]$ complexes and Fds if they contain only $[2\text{Fe}-2\text{S}]$ or $[4\text{Fe}-4\text{S}]$ clusters. Many Fe-S clusters also occur in large and sometimes membrane-bound proteins, together with other prosthetic groups and/or metal cofactors. These proteins are designated complex Fe-S proteins.

Primary Structure

In the amino-acid sequences of many Fe-S proteins, the active sites can be recognized from characteristic patterns of Cys ligands. For instance, the Fe ion in Rd is ligated by two $-\text{C}-\text{x}_2-\text{C}-$ motifs (x standing for residues other than Cys) separated by 20–30 residues. In plant- and mammalian-type $[2\text{Fe}-2\text{S}]$ Fds, the pattern is $-\text{C}-\text{x}_{4,5}-\text{C}-\text{x}_2-\text{C}-\text{x}_n-\text{C}-$, with $n \approx 30$. In Rieske-type $[2\text{Fe}-2\text{S}]$ proteins, the two Cys ligands of one iron and the two His ligands of the other one occur in a $-\text{C}-\text{x}-\text{H}-\text{x}_{15-17}-\text{C}-\text{x}_2-\text{H}-$ pattern. These proteins are found in membrane-bound enzymes in conjunction with *b*-type cytochromes in mitochondria ($b_{\text{c}1}$, or complex III) and chloroplasts (b_6/f), as well as in soluble bacterial dioxygenases. In $[4\text{Fe}-4\text{S}]$ Fds, the canonical pattern is $-\text{C}-\text{x}_2-\text{C}-\text{x}_2-\text{C}-\text{x}_n-\text{C}-$, with $n \approx 20-40$. This pattern is duplicated in Fds containing two $[4\text{Fe}-4\text{S}]$ clusters. Omission of one Cys from this sequence

(usually the third) gives rise to a $[3\text{Fe}-4\text{S}]$ cluster. Similar patterns containing the ligands in Rd and Fd are observed in the larger sequences of many redox enzymes. Significant diversification has occurred, through insertions or deletions of intervening loops of residues. In addition to these canonical sequences, an increasing number of novel Cys patterns are found to furnish ligands to $[2\text{Fe}-2\text{S}]$ and $[4\text{Fe}-4\text{S}]$ sites.

Even though non-Cys ligation of Fe-S active sites is less common, it is functionally important as a means of modulating chemical reactivity or redox potential. Several crystal structures of Fe-S active sites with His, Ser, or carboxylate ligands have now been determined. In some of these, non-Cys ligands are naturally present; in others they have been introduced by site-directed mutagenesis.

Secondary and Tertiary Structure

Many three-dimensional structures of Fe-S proteins have now been determined by X-ray crystallography. At an early stage, the structures of the small Rds and Fds revealed characteristic folds, each specific for a given Fe-S structural framework. The Rd molecule is composed of a three-stranded pleated sheet and two loops that bear a $-\text{C}-\text{x}_2-\text{C}-$ ligand pair each. These two loops are related by a pseudo-twofold symmetry axis passing through the iron atom. Some zinc-finger modules in which the zinc ion has four Cys ligands are structurally very similar to Rd. Plant- and mammalian-type $[2\text{Fe}-2\text{S}]$ Fds assume a structure named β -grasp, of which the main framework is a five-stranded antiparallel β -sheet with an α -helix lying on top of it. The remainder of the structure consists of short secondary elements and loops. Prominent among the latter is a surface loop, including three of the four Cys ligands of the cluster. Another type of $[2\text{Fe}-2\text{S}]$ Fd has a structure closely related to that of thioredoxin. The protein fold of Rieske proteins around their $[2\text{Fe}-2\text{S}]$ active site is superimposable on the Rd fold; the His-ligated iron is the one closest to the protein surface. In $2[4\text{Fe}-4\text{S}]$ Fds, the two $[4\text{Fe}-4\text{S}]$ clusters are surrounded by two pairs of antiparallel β -strands on one side and two α -helices on the other. The structure displays a conspicuous twofold symmetry as a consequence of the sequence duplication. In high-potential Fd (HiPIP), a sizeable part (40%) of the structure is composed of hairpin turns while the remainder consists of several short strands and helices. Consequently, the $[4\text{Fe}-4\text{S}]$ cluster is deeply buried in a hydrophobic cavity that stabilizes its oxidized $3+$ state. The structural frameworks

of the [2Fe–2S] Fd and low-potential [4Fe–4S] Fd are ubiquitous and occur not only in small soluble Fds but also in subunits and domains of many redox enzymes. The low-potential [4Fe–4S] Fd-fold is particularly versatile, and insertions and deletions inferred from sequence alignments have been substantiated by crystal structures. In addition to variations in sequence length, cases of cluster modification or loss have been observed.

Higher-Order Structures

While the canonical Fe–S protein folds described above have long dominated the landscape, numerous novel folds are revealed by the rapid growth of structural databases. It was predictable that polypeptide chains can fold in many different ways to position Cys residues appropriately for the binding of [2Fe–2S] or [4Fe–4S] active sites. Indeed, for the accommodation of [4Fe–4S] clusters alone, well over 30 distinct protein folds have now been structurally characterized. In AdoMet radical proteins, a specific pattern including the sequence C–x3–C–x2–C– binds a [4Fe–4S] cluster with three Cys residues and one open iron site. There are cases of [4Fe–4S] clusters held in proteins between two subunits, each of which provides two of the four Cys ligands, as in nitrogenase Fe-protein and photosystem I centre F_X .

Complex Fe–S Proteins with Clusters of High Nuclearity and Hetero-Metal Clusters

A yet higher level of complexity is displayed by proteins containing large metaloclusters, which, in several cases, belong to the Fe–S family (Figure 2). Some of these superclusters can be considered as derived from [4Fe–4S] clusters without (e.g., P clusters of nitrogenase) or with a metal atom other than Fe (Mo or V in nitrogenase, and Ni in carbon monoxide

dehydrogenase). In nitrite and sulfite reductases, a [4Fe–4S] cluster is covalently bound through a cysteine ligand to the central iron of siroheme. Protein ligands other than Cys (His, Ser, peptidyl-N) or nonprotein ligands (homocitrate, dithiols, CO, and CN^-) are also involved. The protein frameworks that accommodate these large clusters are generally bigger and more sophisticated than those of the usual Fe–S clusters. Nevertheless, it is noted that a number of them are assemblages of smaller substructures such as Rossmann folds.

Structural Flexibility in Relation to Function

Atomic resolution (1.2 Å or better) X-ray crystallography has become more accessible through technical improvements (synchrotron radiation and low temperature) and allows the detection of minute structural deviations that are imposed by the polypeptide chain or occur upon redox transitions. Crystal structures may convey the misleading impression that proteins are rigid structures; however, Fe–S proteins have malleable metal sites associated with flexible polypeptide chains. While essential functions of Fe–S proteins, electron transfer in particular, are indeed optimized through minimization of structural flexibility, many other functions, such as cluster assembly or response to oxidative stress, require structural reorganizations of the polypeptide chain. Such modifications have long been known to occur upon degradation (oxidative denaturation and loss of iron) of some Fe–S proteins. A notable exception is the 3Fe/4Fe interconversion, which may occur with minimal protein reorganization. Other modifications of cluster nuclearity are expected to be associated with significant protein movements, since the Cys ligand sets of the [1Fe], [2Fe–2S], and [4Fe–4S] sites have very different spatial distributions. Flexibility of both the metal site and the polypeptide chain is a necessity in proteins committed to sensing and transducing signals which may depend on redox state or metal concentration. Such

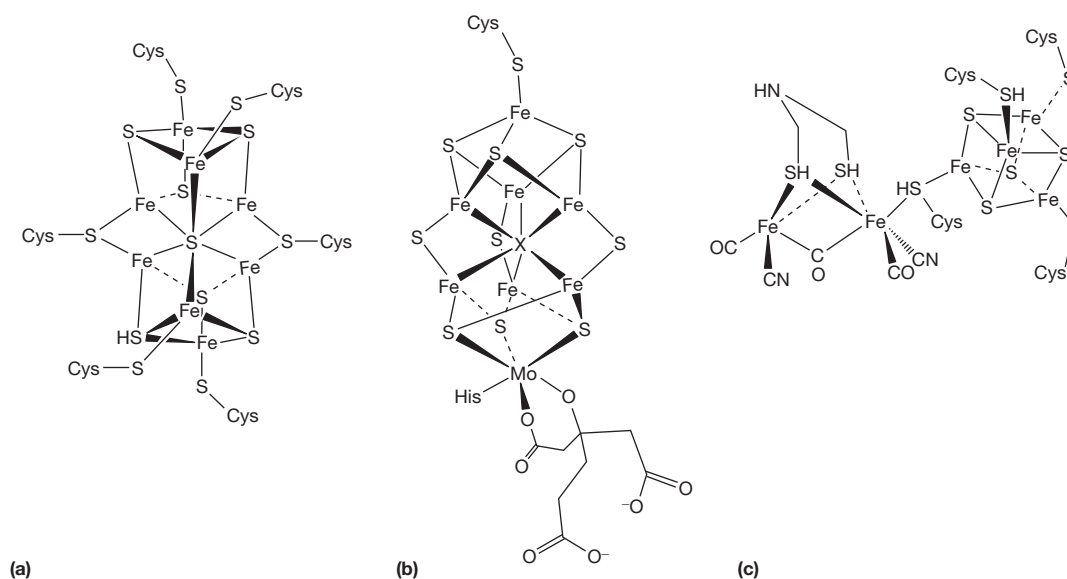


Figure 2 Superclusters: (a) the P cluster in the Fe–Mo protein of nitrogenase; (b) MoFe-cofactor, with bound homocitrate; and (c) H-cluster of the FeFe-hydrogenase.

structural changes may be classified in two main groups, according to whether it is the polypeptide chain (protein-driven reorganization of the metal site) or the metal site (cluster-driven protein movements) that assumes the leading role. Examples of the latter, where polypeptide chain movements are driven by the avidity of Fe–S clusters for Cys ligands, have been structurally characterized.

Functions

Stabilization of Protein Structures

In some enzymes, the network of ligands around a Fe–S cluster is used to stabilize specific structures. The covalent structures, particularly of [4Fe–4S] clusters, can improve protein rigidity. This provides an explanation for the existence of some highly conserved clusters in proteins, such as some DNA-repair helicases. However, in most cases, another distinct function can be found for the existence of the clusters.

Electron Transfer

Many complex Fe–S proteins contain clusters, arranged in chains, separated 1–1.4 nm apart, which allow facile electron transfer across considerable distances. Other electron-transferring groups such as heme or flavin may also be present. Examples include the electron transport chains of mitochondrial respiration (complexes I, II, and III) and photosynthesis (photosystem I), and enzymes such as hydrogenase, nitrogenase, and xanthine dehydrogenase.

Oxidation and reduction of clusters occur with minimal reorganization energy, which means that very rapid electron transfer (in less than microseconds) is possible, for example, in the primary electron-transfer processes of photosynthesis. The oxidation–reduction potentials of the Fe–S clusters determine the overall direction of electron transfer. However, in some electron-transfer chains, an uphill step from a more positive to a more negative potential center does not appear to act as a barrier as long as the overall electron transfer is favorable. Redox potentials reported in other compendia have been collected, and representative ranges for various types of Fe–S clusters are summarized in [Figure 1](#). Multiple factors influence the redox potential range of Fe–S proteins. First, the electronic structure of the Fe–S framework. Second, as each iron ion can be in either the Fe(III) or the Fe(II) state, clusters of n Fe atoms are in principle endowed with $n - 1$ redox transitions. This is observed in nature only in the case of the [4Fe–4S] clusters (3+/2+, 2+/+, and 1+/0 transitions). Third, the redox potentials of Fe–S clusters can be modified by several parameters that are adjustable by the polypeptide chain. These include cluster ligation (sulfur, nitrogen, or oxygen), polarity of the medium, and neighboring charges. Studies of synthetic analogs and proteins, native as well as mutated, have shown that variations of each of these parameters can cause redox potential shifts of 50 mV or more. For instance, in Rieske-type proteins, the presence of two His ligands on the same Fe raises the redox potential of this Fe sufficiently so that it becomes reducible at a much higher potential than that of all Cys-ligated [2Fe–2S] active sites ([Figure 1](#)). Moreover, for some of these proteins, reduction is accompanied by protonation, which compensates

for the electron charge; as a result, the redox potential of Rieske clusters is pH dependent.

Generation of unusual reducing power

In a number of enzymes, a reduced 4Fe cluster, in conjunction with the hydrolysis of MgATP, is used to achieve reductions requiring potential differences of more than 1 V, for the reduction of dinitrogen by nitrogenase, and of a porphyrin ring by the light-independent protochlorophyllide reductase. In these enzymes, the [4Fe–4S] cluster is located between two subunits of a protein, each contributing two Cys ligands, and each binding one molecule of adenosine triphosphate (ATP) which is hydrolyzed during electron transfer. The resultant conformational change in the cluster can generate very low redox potentials, with [4Fe–4S] clusters even being reduced by two electrons to the all-ferrous [4Fe–4S]⁰ state.

Catalytic Fe–S Clusters

As already noted, the most common function of Fe–S proteins is electron transfer, and the clusters are fully coordinated by ligands such as Cys. However, some enzymes, probably originating from more primitive forms of life, use Fe–S clusters directly in catalysis, and these often have a free iron position for binding to substrates or other ligands. Some examples of these are shown in [Figure 3](#).

In aconitase, a nonredox enzyme, the unique iron is able to extend its coordination from four to six coordinate and bind the substrates, citrate, or isocitrate ([Figure 3\(a\)](#)). It can thus serve as a Lewis acid in the dehydration of the substrate to *cis*-aconitate and the rehydration to the alternate substrate with inversion of the positions of the H and OH substituents.

Initiation of radical reactions

There are a number of enzymes that use free-radical chemistry supported by adenosylmethionine to achieve substitutions on unactivated carbon atoms. These so-called ‘radical SAM’ enzymes contain a [4Fe–4S]⁺ cluster with only three Cys ligands, which binds S-adenosylmethionine (AdoMet) ([Figure 3\(b\)](#)). Electron transfer from the cluster splits adenosylmethionine and generates a 5′deoxyadenosyl radical, which is analogous to those produced by cobalamin (vitamin B₁₂)-dependent enzymes. The radical is then used to catalyze difficult synthetic reactions in a wide variety of metabolic processes. In some cases, the AdoMet radical acts catalytically in the enzyme, for example, in lysine 2,3-aminomutase. In others, the radical is employed to generate another protein-bound radical, which is required for catalysis; for example, AdoMet-dependent activating enzymes generate essential glycol radicals in pyruvate formate-lyase and the ribonucleoside triphosphate reductase III of anaerobic bacteria. Other enzymes consume AdoMet in the transfer of sulfur from cysteine into biotin and lipoic acid, and the thiomethylation of transfer RNA.

Reduction of disulfide bonds

During redox reactions, Fe–S clusters usually transfer one electron at a time, but reduction of disulfide bonds requires two reducing equivalents. Plant ferredoxin:thioredoxin reductase achieves the reduction of the disulfide in thioredoxin by means of a unique [4Fe–4S] cluster extended by two cysteinate

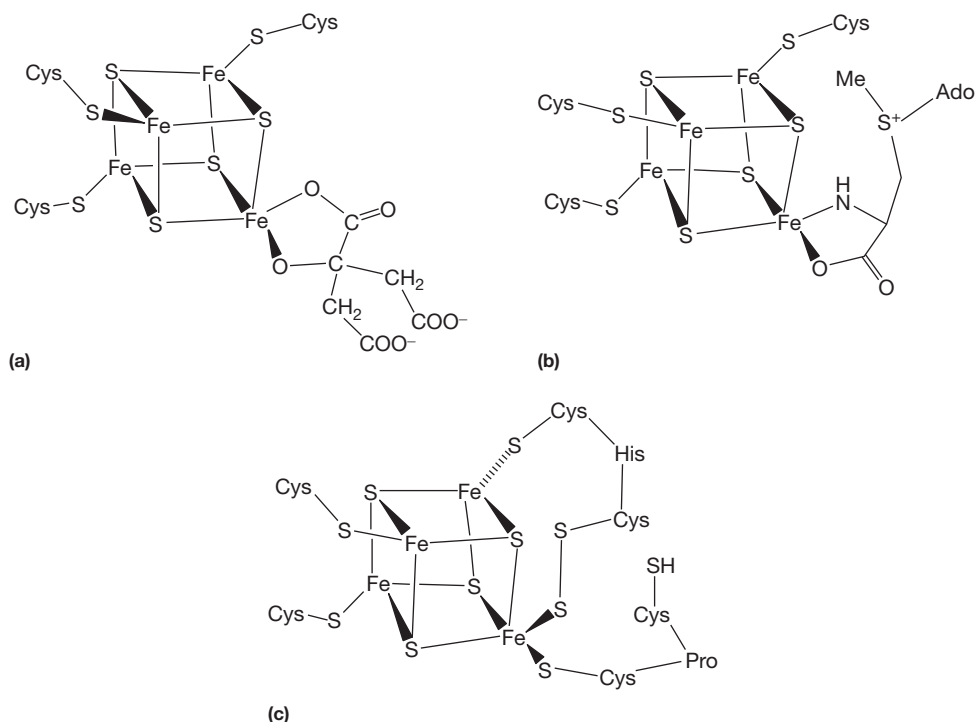


Figure 3 [4Fe-4S] clusters in the catalytic centres of enzymes: (a) aconitase (citrate bound); (b) a representative AdoMet radical enzyme (AdoMet bound); and (c) ferredoxin:thioredoxin reductase.

ligands. These form a disulfide which extends the coordination sphere of one of the iron ions (Figure 3(c)). The cluster transiently assumes the $[4\text{Fe}-4\text{S}]^{3+}$ level and generates a sulfur radical, as indicated by electron paramagnetic resonance (EPR).

Sensing and regulatory functions

Fe-S clusters are labile toward oxygen and reactive oxygen species, and toward nitric oxide. Fe-S clusters can, therefore, also be used in nature as signaling or regulatory agents. While the Fe-S clusters of proteins that belong to this group may undergo one electron oxidation or complete oxidative destruction, the ensuing regulatory function involves specific interactions with DNA (SoxR and FNR) or RNA (iron regulatory protein (IRP)), which require either the presence (SoxR, FNR) or the complete absence (IRP) of the Fe-S cluster. Examples of this are the $[2\text{Fe}-2\text{S}]$ SoxR and $[4\text{Fe}-4\text{S}]$ FNR protein in bacteria and the IRP in eukaryotes. SoxR is a sensor for the superoxide anion that binds to DNA upstream of the *soxS* gene. Its oxidation-reduction potential is such that it normally occurs in the reduced $1+$ form in bacteria. Upon oxidation to the $2+$ level by the superoxide anion, a structural modification of the soxR-soxS nucleoprotein complex triggers expression of *soxS* and induction of the synthesis of defensive enzymes such as superoxide dismutase and other proteins. The FNR protein acts as oxygen sensor. Its 4Fe cluster is rapidly destroyed in the presence of oxygen, hence its ability to induce the synthesis of enzymes required for anaerobic respiration with fumarate or nitrate as oxidants instead of oxygen. IRP of eukaryotes, on the contrary, is unable to serve as iron regulator unless its cluster-binding site (3Cys) is unoccupied and contains free thiols. However, when it acquires a Fe-S cluster as $[4\text{Fe}-4\text{S}]$ protein, it takes

on the function of a cytoplasmic aconitase. Both cytoplasmic and mitochondrial aconitases are only active in their 4Fe form and are inactivated by loss of the unique iron that lacks a Cys ligand. As the 3Fe cluster is relatively stable, it has been proposed that the switch between 4Fe (active) and 3Fe (inactive) enzyme is used in nature to protect the Fe-S cluster from irreversible further oxidation and for metabolic control, because aconitase is necessary for the functioning of the tricarboxylic acid cycle. As nitric oxide reacts with Fe-S clusters, with the main product formed being the dinitrosyl-iron complex (DNIC), NO has to be considered as an alternative messenger in the mentioned sensing pathways. In intact cells, the Fe-S clusters can eventually be rebuilt, as described later. The interrelation of the pathways for sensing iron, oxygen, or its reduction products, and nitric oxide represents an interesting example of crossroads in natural signaling systems.

Methods for Obtaining Information on Fe-S Proteins

While gross information on Fe-S proteins is available from chemical approaches such as analysis for sulfide, iron, protein, and amino acids, spectroscopic methods have furnished a wealth of information on the more detailed and subtle properties of these fascinating, multipurpose tools of nature. Historically, EPR was the first decisive method. It was soon followed by a considerable arsenal of diverse methods. Many of them probe the electronic and magnetic properties of the metal site in various ways that render them complementary to each other and to EPR. The most important of these are Mössbauer, magnetic circular dichroism (MCD), nuclear magnetic resonance (NMR), and electron-nuclear double resonance (ENDOR)

spectroscopies. Other methods, such as resonance Raman and X-ray absorption spectroscopy (EXAFS and XANES), provide specific information relevant to the structural framework of the metal site. Computational approaches have been of great help for unraveling details of the electronic and magnetic structures of Fe-S clusters.

Biogenesis of Iron-Sulfur Proteins

General Overview

The *in vitro* reconstitution of Fe-S clusters on apoproteins, using iron salts, thiols, and inorganic sulfide, was demonstrated in the 1960s. The Fe-S compounds were also synthesized in the absence of protein in order to explore cluster chemistry. However, it was not until the late 1990s that the biosynthesis of Fe-S clusters in a living cell was shown to be a highly evolved process, requiring the assistance of complex assemblies of proteins, which became a major field of interest and experimentation. At present, three different biogenesis systems have been identified. The nitrogen fixation (NIF) system was first discovered from genetic studies for the specialized assembly of the complex Fe-S protein nitrogenase and also occurs in nondiazotrophic ϵ -proteobacteria. The ISC (iron-sulfur cluster) system is a more general assembly machinery present in α -, β -, and γ -proteobacteria as well as in mitochondria of eukaryotes. The most recently discovered system is the SUF (sulfur mobilization) machinery, which is found in a wide range of bacteria including cyanobacteria, archaea, and plastids of plants. In some organisms, the NIF, ISC, or SUF system is responsible for assembly of cellular Fe-S proteins, whereas in other species they are present in parallel, and may be responsible for the assembly of different subsets of Fe-S proteins.

The three systems have in common two proteins. A cysteine desulfurase (NifS, IscS, or SufS) is a pyridoxal-phosphate-dependent enzyme which converts the amino acid Cys to Ala, liberating the sulfur which is covalently bound as an S-sulfide at a conserved Cys residue prior to insertion into the Fe-S cluster. A scaffold protein (NifU, IscU, or SufU), containing Cys residues, serves as a platform for the assembly of Fe-S clusters. This structure is labile and can be transferred to target proteins. In addition, there are iron-transfer proteins, electron-transfer proteins (Fd), and specialized proteins such as chaperones that facilitate the assembly of particular clusters and their stable integration into finished Fe-S proteins.

Biogenesis of Fe-S Proteins in Bacteria

The ISC assembly machinery

In prokaryotes, the operon for the ISC assembly machinery encodes the proteins IscS, IscU, IscA, HscA, HscB, and Fdx. In addition to these proteins directly required for biogenesis, IscR is involved in the regulation of the *isc* gene expression. Biogenesis starts with the production of an S-sulfide sulfur from cysteine by the cysteine desulfurase IscS (Figure 4). The sulfur remains transiently bound to IscS and, upon interaction of IscS with the dimeric form of IscU, the sulfur moiety is transferred to IscU, where it is also bound as a persulfide. The binding of ferrous iron to IscU leads to the formation of a [2Fe-2S] cluster, which then can give rise to the generation

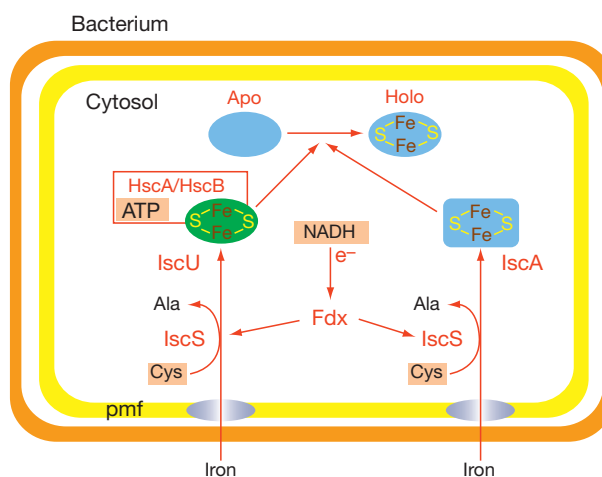


Figure 4 A working model for biosynthesis of Fe-S proteins in bacteria. For details, see text. Abbreviation: pmf, proton motive force.

of a [4Fe-4S] cluster. Thus, IscU provides a scaffold for assembly of iron and sulfide ions. The molecular details of these reactions are still under investigation. IscU interacts with two chaperones, the ATP-dependent Hsp70/DnaK homolog HscA and the Hsp70/DnaJ homolog HscB. The interaction strongly stimulates the ATPase activity of HscA, but the precise function of these two proteins remains to be elucidated. Fe-S cluster assembly depends on the transfer of electrons from the [2Fe-2S] ferredoxin Fdx. The reactions which need the input of electrons are not known yet, but may involve the release of sulfur from cysteine by IscS and the dissociation of the Fe-S cluster from IscU. IscA is thought to function as an alternative scaffold protein to facilitate the assembly of Fe-S clusters. A [2Fe-2S] cluster can be assembled on IscA and subsequently transferred to Fe-S proteins such as Fd.

The bacterial SUF machinery

The existence of a biogenesis system for Fe-S proteins in addition to the ISC assembly machinery has been noted only recently. When the *isc* operon is deleted in *Escherichia coli*, Fe-S proteins are assembled at an efficiency of only 2–10% of that in wild-type cells. The residual assembly activity is due to the *suf* (mobilization of sulfur) operon which contains the proteins SufA, SufB, SufC, SufD, SufS, and SufE. Many organisms contain only a subset of the *suf* genes. Overexpression of the *suf* gene cluster can restore Fe-S protein assembly in an *isc* deletion mutant. The *suf* operon is induced under iron-limiting conditions and repressed by the iron-sensing repressor protein Fur. Further, the *suf* operon is regulated by the transcription factor OxyR that senses oxidative stress conditions. This suggests that the SUF system is crucial under iron-limiting and oxidative stress conditions. The Suf proteins, homologs of Fdx and NifU, and an Fdx reductase are also present in apicoplasts, a plastid-like organelle of Apicomplexa such as *Plasmodium* and *Toxoplasma*.

SufS and SufA are homologs of IscS and IscA, respectively, and may serve similar functions in liberating sulfur from cysteine, in scaffolding, and in cluster assembly. SufC is an atypical ATPase which forms a complex with SufB and SufD. SufE accepts sulfur from SufS and acts synergistically with the SufBCD

complex to stimulate the cysteine desulfurase activity of SufS. This regulation of SufS activity may be important for limiting sulfur release during oxidative stress conditions. In addition, SufS has been assigned a function as a selenocysteine lyase in *E. coli*.

Assembly of superclusters in nitrogenase and hydrogenase

The iron–molybdenum cofactor (FeMo-co) in nitrogenase and the H cluster in FeFe-hydrogenase are each located in a cavity in the enzyme, bound to the protein by just one Cys. Each is constructed in a scaffold protein by a series of accessory proteins and chaperones before transfer to the finished enzyme. Central components in the biosynthesis of FeMo-co are NifN and NifE, which form a hetero-tetramer and serve as a scaffold for cluster assembly before transfer to NifKD apoproteins. Further components needed for FeMo-co assembly are the *nif* operon-encoded proteins NifQ (Mo processing), NifV (homocitrate synthase), NifB (sulfur and iron donor), as well as NifH and NifX, the functions of which are unknown. Studies on nitrogen-fixing bacteria have identified the *nif* (nitrogen fixation) operon as being responsible for maturation of the complex Fe–S protein nitrogenase. This enzyme consists of two oxygen-labile metalloproteins, the MoFe protein (encoded by NifKD) and the Fe protein (NifH). The latter contains a [4Fe–4S] cluster, bound between two subunits, whereas the MoFe protein is associated with two unique superclusters: the P cluster and the FeMo-co (Figure 2). The *nif* operon-encoded proteins NifS and NifU are involved in cluster biosynthesis by performing similar functions as their homologs IscS and IscU, respectively. The N terminus of NifU is homologous to IscU and provides a scaffold for assembly of transiently bound clusters, whereas the middle part of NifU contains a permanent [2Fe–2S] cluster that may carry out a redox function. Further, a homolog of IscA termed NifHscA is thought to be an alternative scaffold for cluster assembly. For FeFe-hydrogenase, the 2Fe subcluster of the H cluster is synthesized on HydF from a [2Fe–2S] cluster framework in a process requiring guanosine triphosphate (GTP) and two AdoMet radical proteins, HydE and HydG, before being transferred to the catalytic subunit HydA.

Biogenesis of Fe–S Proteins in Eukaryotes

The Mitochondrial Fe–S Cluster Assembly Machinery

In eukaryotes, mitochondria are central for biogenesis of cellular Fe–S proteins. Some eukaryotic protists, which lack functioning mitochondria, have mitosomes, intracellular organelles that contain the machinery for Fe–S cluster assembly. Loss of the ability to assemble Fe–S clusters is generally lethal to the cell. A number of inherited diseases are associated with defects in the system, notably the neurodegenerative disease Friedreich's ataxia, which is associated with a defect in frataxin. In keeping with their bacterial origin, these organelles contain homologs of the bacterial ISC and NIF machinery. All proteins of the mitochondrial assembly machinery are encoded by nuclear DNA and are imported after translation in the cytosol. The mitochondrial ISC assembly machinery comprises more proteins than in bacteria, though the underlying mechanistic principles are analogous. For instance, the cysteine desulfurase Nfs1 and the two highly homologous proteins, Isu1 and Isu2, execute similar

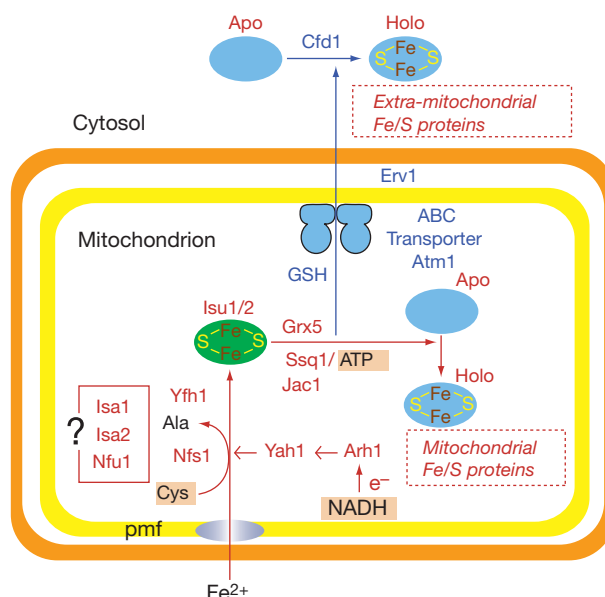


Figure 5 A working model for biosynthesis of Fe–S proteins in yeast. For details, see text. Abbreviation: pmf, proton motive force.

functions in elemental sulfur production and scaffolding as bacterial IscS and IscU, respectively (Figure 3). Formation of the transiently bound Fe–S cluster on the Isu proteins depends on the Fd reductase Arh1 and the Fd Yah1, which form an electron transfer chain using reduced nicotinamide-adenine dinucleotide (NADH) as a reducing agent. In addition, Yfh1, a homolog of frataxin, acts as an iron donor during Fe–S cluster synthesis by directly interacting with the Isu/Nfs1 complex. The dissociation of the transiently bound Fe–S cluster from the Isu proteins and/or its insertion into apoproteins requires the mitochondrial glutaredoxin Grx5 and the two chaperones Ssq1 and Jac1 (Figure 5). The IscA homologs Isa1 and Isa2 provide an Isu protein-independent scaffolding function for the assembly of Fe–S clusters. Nfu1 genetically interacts with Ssq1 and Isu1, but its biochemical role in mitochondrial Fe–S protein assembly is not understood to date.

Biogenesis of Fe–S Proteins in Chloroplasts

In plant cells, chloroplasts contain their own assembly machinery, independent of the machinery in mitochondria. Cyanobacteria, which are believed to be the evolutionary ancestors of chloroplasts, contain several homologs of the ISC assembly machinery and also the SUF system. In chloroplasts, a NifS homolog, CpNifS, and an Fe–S-binding P-loop NTPase, HCF101, have been identified as the cysteine desulfurase and scaffold proteins of this system.

The Cytosolic CIA System

The mitochondrial ISC assembly machinery not only is responsible for the assembly of mitochondrial Fe–S proteins but also plays an essential function in the maturation of Fe–S proteins outside mitochondria. To date, four components have been identified that are specifically needed for maturation of

extra-mitochondrial, but not of mitochondrial Fe–S proteins (Figure 5). The first-known protein of the so-called ISC export machinery is the mitochondrial ABC transporter Atm1. The other components are the sulfhydryl oxidase Erv1 of the intermembrane space and glutathione (GSH). Since virtually all components of the mitochondrial ISC assembly machinery are involved in the maturation of extra-mitochondrial Fe–S proteins, it is assumed that Atm1 exports a Fe–S cluster or a derivative thereof to the cytosol for transfer onto apoproteins. Mutation of the human homolog of Atm1, termed ABC7, gives rise to the iron storage disease X-linked sideroblastic anemia and cerebellar ataxia (XLSA/A). Finally, biogenesis of cytosolic Fe–S proteins requires Cfd1, a cytosolic P-loop ATPase that is conserved in all eukaryotes. The protein may assist in Fe–S cluster insertion into cytosolic apoproteins.

Mitochondria appear to be essential organelles because they perform crucial steps in the biosynthesis of cytosolic Fe–S proteins. Many components of the ISC machineries are essential for viability of yeast cells (e.g., Nfs1, Yah1, Arh1, Jac1, Erv1, and Cfd1) which demonstrates the importance of Fe–S protein biogenesis. Outside of the mitochondria, four highly conserved proteins (Cfd1, Nbp35, Nar1, and Cia1) have been identified as members of the cytosolic [Fe–S] protein assembly machinery. The structure of Nar1 is similar to that of the FeFe-hydrogenases, although it does not contain the H-cluster. The soluble P-loop NTPases Cfd1 and Nbp35 function as a scaffold for assembly of labile [Fe–S] clusters, which can be rapidly transferred and incorporated into target [Fe–S] apoproteins in a Nar1- and Cia1-dependent fashion.

See also: **Bioenergetics:** Ferredoxin; Ferredoxin-NADP⁺ Reductase; F_X, F_A, and F_B Iron–Sulfur Clusters in Type I Photosynthetic Reaction Centers; Hydrogenases, Structure and Function; Intercellular Ca²⁺ Waves: Mechanisms of Initiation and Propagation;

Protein/Enzyme Structure Function and Degradation: Disulfide Bond Formation.

Further Reading

- Balk J and Lobreaux S (2005) Biogenesis of iron–sulfur proteins in plants. *Trends in Plant Science* 10: 324–331.
- Beinert H (2000) Iron–sulfur proteins: Ancient structures, still full of surprises. *Journal of Biological Inorganic Chemistry* 5: 2–15.
- Beinert H and Kiley PJ (1999) Fe–S proteins in sensing and regulatory functions. *Current Opinion in Chemical Biology* 3: 152–157.
- Efremov RG, Baradaran R, and Sazanov LA (2010) The architecture of respiratory complex I. *Nature* 465: 441–461.
- Fontecilla-Camps JC, Amara P, Cavazza C, Nicolet Y, and Volbeda A (2009) Structure–function relationships of anaerobic gas-processing metalloenzymes. *Nature* 460: 814–822.
- Frey PA, Hegeman AD, and Ruzicka FJ (2008) The radical SAM superfamily. *Critical Reviews in Biochemistry and Molecular Biology* 43: 63–88.
- Johnson DC, Dean DR, Smith AD, and Johnson MK (2005) Structure, function, and formation of biological iron–sulfur clusters. *Annual Review of Biochemistry* 74: 247–281.
- Lill R (2009) Function and biogenesis of iron–sulphur proteins. *Nature* 460: 831–838.
- Lill R and Muhlenhoff U (2008) Maturation of iron–sulfur proteins in eukaryotes: Mechanisms, connected processes, and diseases. *Annual Review of Biochemistry* 77: 669–700.
- Meyer J (2008) Iron–sulfur protein folds, iron–sulfur chemistry, and evolution. *Journal of Biological Inorganic Chemistry* 13: 157–170.
- Palmer G and Reedijk J (1991) Nomenclature of electron-transfer proteins. *European Journal of Biochemistry* 200: 599–611.
- Rao PV and Holm RH (2004) Synthetic analogues of the active sites of iron–sulfur proteins. *Chemical Reviews* 104: 527–559.
- Rothery RA, Workun GJ, and Weiner JH (2008) The prokaryotic complex iron–sulfur molybdoenzyme family. *Biochimica et Biophysica Acta* 1778: 1897–1929.
- Sheftel A, Stehling O, and Lill R (2010) Iron–sulfur proteins in health and disease. *Trends in Endocrinology and Metabolism* 21: 302–314.
- Stephens PJ, Jollie DR, and Warshel A (1996) Protein control of redox potentials of iron–sulfur proteins. *Chemical Reviews* 96: 2491–2513.
- Sykes AG and Cammack R (eds.) (1999) *Iron–Sulfur Proteins: Advances in Inorganic Chemistry*, vol. 47. San Diego, CA: Academic Press.
- Vignais PM and Billoud B (2007) Occurrence, classification, and biological function of hydrogenases: An overview. *Chemical Reviews* 107: 4206–4272.

Keratins and the Skin

P A Coulombe, K M Bernot, and C-H Lee, Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Pierre A. Coulombe and Kelsie M. Bernot, volume 2, pp. 497–504, © 2004, Elsevier Inc.

Glossary

Apoptosis Orderly process of programmed cell death by which a cell commits suicide.

Coiled coil Tertiary structure of a protein motif mediating oligomerization in which two alpha helices wrap around each other forming a hydrophobic core with characteristic interdigitation of side chains between neighboring helices, known as ‘knob-in-hole packing’.

Complex epithelia One to several layers of cells in which nuclei are present on several planes covering either internal or external surfaces of the body.

Desmosome Junction formed between adjacent cells in order to form a tissue.

Hemidesmosome Junction that anchors cells to the basement membrane.

Intermediate filaments Nonpolar, 10-nm-wide filamentous proteins of the cytoskeleton that provide structural support to the cell.

Pilosebaceous unit Comprised of a hair follicle, sebaceous gland attached to it, and arrector pili muscle.

Simple epithelium A single layer of cells attached to a basal lamina, in which all nuclei are aligned in a single plane often covering an internal surface of the body.

Keratins are among the most abundant proteins in epithelial cells, in which they occur as a cytoplasmic network of 10-nm-wide intermediate filaments. Keratin proteins are encoded by a large multigene family in mammals, with 54 individual genes that can be readily partitioned into two major sequence types. A strict requirement for the heteropolymerization of types I and II keratin proteins during filament assembly underlies the pairwise transcriptional regulation of keratin genes. Modulation of keratin expression also depends upon both the type and stage of differentiation in epithelia. A major function fulfilled by keratin filaments is to act as a resilient yet pliable scaffold, which endows epithelial cells with the ability to sustain mechanical and nonmechanical stresses. Reflecting a partial or complete loss of this crucial scaffolding role, genetic mutations altering the coding sequence of keratins account for a large number of epithelial fragility disorders. Several additional roles for keratin proteins, affecting cell growth, migration, programmed cell death, and response to stresses, are manifested in a sequence- and context-dependent fashion and play key roles in normal physiology and in disease.

Keratin Classification

Keratins belong to the superfamily of intermediate filament (IF)-forming proteins. They are heterogeneous in size

(40–72 kDa) and charge ($pI \sim 4.7$ –8.4), and notoriously insoluble. The original keratin nomenclature, devised in 1982 by Roland Moll, Werner Franke, and their colleagues, was based upon protein separation by both charge and size via two-dimensional electrophoresis. A revised nomenclature was produced in 2006 to accommodate a large number of novel keratins (uncovered mainly through whole genome sequencing efforts) while complying with the human and mouse genome nomenclature committees. Type I keratins ($n=28$; [Figure 1](#)) tend to be smaller (40–64 kDa) and acidic ($pI \sim 4.7$ –6.1) compared to the larger (52–68 kDa) and basic–neutral ($pI \sim 5.4$ –8.4) type II keratins ($n=26$; [Figure 1](#)). Homology at both the protein (see [Figure 1\(a\)](#)) or DNA sequence levels, as well as gene substructure (number and position of introns), readily confirms the partitioning of keratin genes and proteins into type I and II IF sequences.

Keratin Gene Clusters Reflect Tissue Expression: Implications for Keratin Evolution

The coordinated regulation of two distinct types of keratin genes, along with the obligatory heteropolymerization of their products to form 10-nm-wide filaments, is present as far back as primitive chordates. In *Homo sapiens*, functional type I and type II keratin genes are clustered on the long arms of

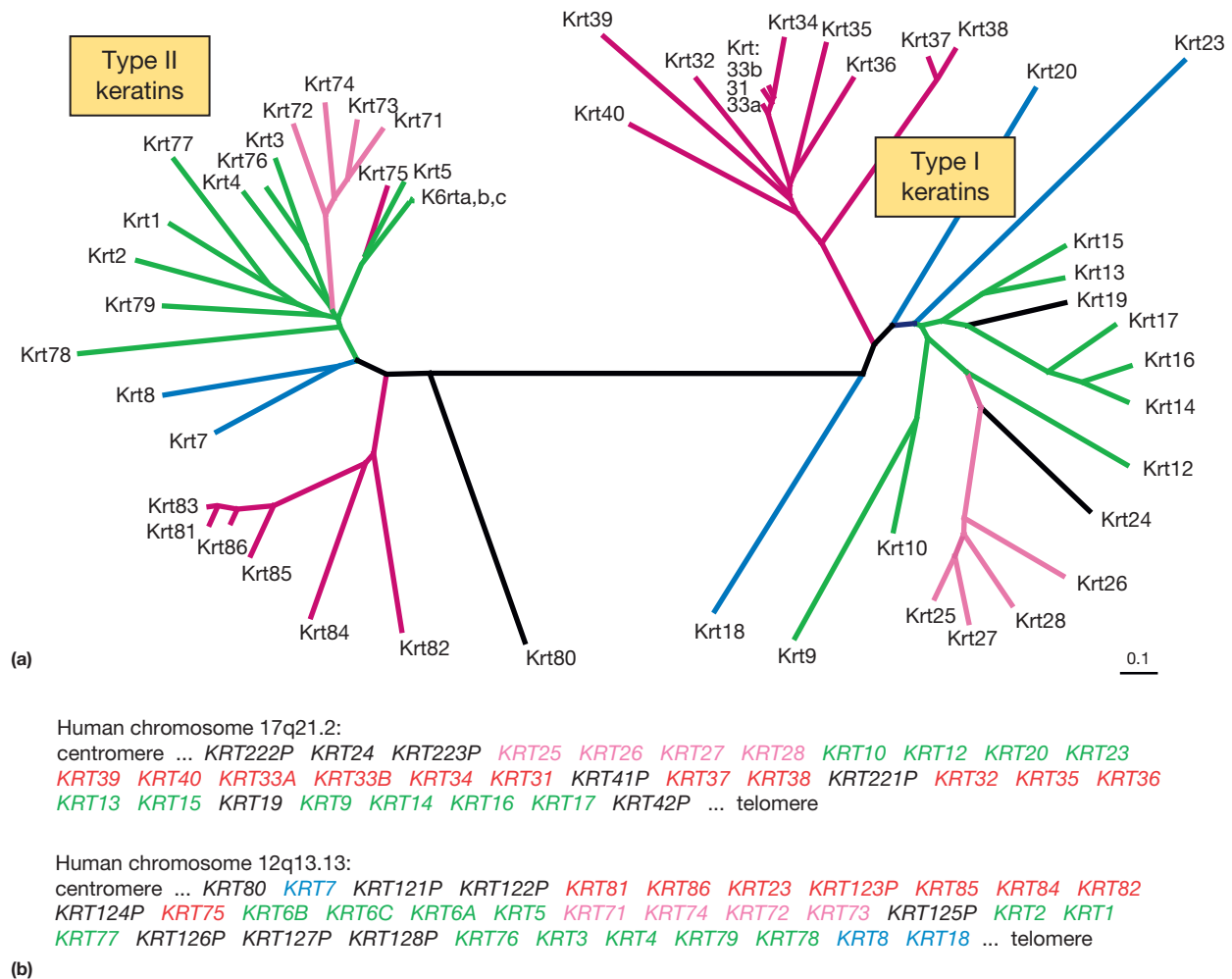


Figure 1 The human keratin family c. 2011. (a) Comparison of the primary structure of human keratins using the publicly available ClustalW and TreeView softwares. Sequence relatedness is inversely correlated with the length of the lines connecting the various sequences as well as the number and position of branch points. This comparison makes use of the sequences from the head and central rod domain for each keratin. Two major branches are seen in this tree, corresponding exactly to the known partitioning of keratin genes into type I and type II sequences. Beyond this dichotomy each subtype is further segregated into major subgroupings. (b) Location and organization of type I and type II keratin genes in the human genome. All functional type I keratin genes, except Krt18, are clustered on the long arm of human chromosome 17, while all functional type II keratin genes are located on the long arm of chromosome 12. Krt18, at type I gene, is located at the telomeric (Tel) boundary of the type II gene cluster. The suffix P identifies nonfunctional keratin pseudogenes. As highlighted by the color code used in frames (a) and (b), individual type I and type II keratin genes belonging to the same subgroup, based on the primary structure of their protein products, tend to be clustered in the genome. Moreover, highly homologous keratin proteins (e.g., Krt5 and Krt6 paralogs; also, Krt14, Krt16, and Krt17) are often encoded by neighboring genes, emphasizing the primary role of gene duplication in generating keratin diversity. These features of the keratin family are virtually identical in mouse (not shown).

chromosomes 17 and 12, respectively (Figures 1(b) and 1(c)). There is a notable exception to this principle, in that the type I keratin 18 (Krt18) locus is located next to Krt8 at the telomeric boundary of the type II cluster (Figure 1). Krt8 and Krt18 exhibit a tight coregulation in both embryonic and adult simple epithelia, and likely are direct descendants of the ancestral pair of keratin genes. The key molecular features of individual keratin genes (e.g., size and number of introns/exons) and those of the genomic clusters they form (e.g., transcriptional orientation and position relative to other family members) are highly conserved in human and mouse, and in many other species of mammals as well. This conservation has direct implications for the evolution of keratin genes and points to the

existence of locus control elements simultaneously affecting the regulation of several neighboring genes.

Comparing the primary structure of human keratin proteins also reveals the existence of distinct subfamilies within each of the two keratin types (Figure 1(a)), and provides additional insight into the mechanisms presiding over their evolution. Groupings of individual genes within each of the two major clusters tend to reflect expression in related types of epithelial cells or tissues. This situation applies for most genes expressed in simple epithelia (Krt7, Krt8, Krt18, and Krt20), soft complex epithelia (e.g., Krt1–Krt6; Krt9–Krt17), hard epithelia such as hair shaft and nail (Krt31–38; Krt81–86), and even the highly restricted inner root sheath (IRS)

compartment of hair follicles (Krt25–Krt28; Krt71–Krt74). A small subset of keratin genes exhibits a mixed regulation in two or more types of epithelia (e.g., Krt19 and Krt80). In addition to these intriguing features, the clustering of keratin genes is mirrored by a corresponding clustering of related protein products as revealed by phylogenetic analysis. This striking equivalence, highlighted by the color scheme employed in [Figure 1](#), implies a hierarchical pattern of gene duplication and specialization during evolution.

Keratin Proteins Form the Intermediate Filament Network of Epithelial Cells

Features of Keratin Proteins

Despite sequence differences, all keratins display the tripartite domain structure that is typical of IF-forming proteins ([Figure 2\(a\)](#)). The central domain consists of an extended α -helix featuring long-range heptad repeats that mediate coiled-coil dimerization. This rod domain is invariably 310 amino acids long and is flanked by highly variable sequences at the N-terminal head and C-terminal tail domains. Neither terminal domain exhibits known functional motifs other than the poorly understood glycine loops seen in epidermal keratins. The head and tail domains are protease- and kinase-accessible and, thus, are exposed at the surface of filaments, where they can foster interactions with neighboring filaments, other proteins, as well as serve as substrates for posttranslational modifications involved in their regulation. Given the marked heterogeneity observed in both their size and primary structure, the head and tail domains are poised to impact the regulation and function of individual keratin proteins *in vivo*.

Keratin Proteins Self-Assemble into 10–12-nm-Wide Filaments

The central rod domain of keratins is the main determinant of polymerization, with additional contributions provided by the head domain. By contrast, the tail domain appears to be largely dispensable for 10-nm filament assembly. Keratin polymerization begins with the formation of heterodimers in which the α -helical rod domains of type I and II keratins are aligned in parallel and perfect register. These heterodimers then interact along their lateral surfaces and then in an end-to-end fashion to give rise to the 10-nm-wide filaments. The extraordinary stability of keratin heterodimers (they resist 9 M urea!) and heterotetramers underscores the tightness of the interactions between type I and type II rod domains. Not surprisingly, most of the intracellular pool of keratin proteins (>95%) is in the polymerized form.

Organization and Regulation of Keratin Filaments *In Vivo*

The abundance and intracellular organization of keratin IFs *in vivo* varies according to the epithelium considered. Keratins are particularly abundant (from ~15% to ~80% of total cellular proteins) in surface-exposed stratified epithelia (e.g., epidermis, oral mucosa, cornea, etc.). In cells of such tissues, keratin IFs are organized in a pan-cytoplasmic network extending from the surface of the nucleus to the cytoplasmic periphery, where

they associate with membrane-spanning cell–cell or cell–matrix attachment complexes such as desmosomes ([Figure 2\(c\)](#)) and hemidesmosomes. In simple epithelia (e.g., liver, gut, pancreas, etc.), keratins remain a major fraction of proteins even though they are less abundant (from 1% to 5 % of total cellular proteins). Such tissues are made up of polarized epithelial cells that often contain asymmetrically organized keratin IFs, concentrated mostly at the cytoplasmic periphery and particularly at the apical pole. A number of associated proteins interact with and contribute to the organization of keratin IFs in these various settings. Some directly promote the bundling of keratin IFs (e.g., filaggrin and trichohyalin), their association with microtubules and microfilaments (e.g., plectin and bullous pemphigoid antigen (BPAG) isoforms), and/or with desmosomes or hemidesmosomes (desmoplakin, plakophilin, BPAG isoforms, etc.).

Keratin Posttranslational Modification

The organization of keratin filaments *in vivo* is also regulated through phosphorylation. All keratin proteins are phosphoproteins. Their phosphorylation by a variety of kinases occurs in a very dynamic and reversible fashion, usually on Ser/Thr residues, and involves multiple sites on the nonhelical head domain and, to a lesser extent, on the tail domain. Circumstances in which keratin IFs are reorganized in a rapid fashion, such as during mitosis or cellular stress, invariably correlates with increased phosphorylation of specific residues located in the end domains. The head domains of at least some keratins are also substrates for ubiquitination (which mediates their degradation) and O-glycosylation (whose significance has yet to be ascertained). As is the case for microtubules and actin microfilaments, the assembly and organization of keratin IFs is thus regulated through a variety of mechanisms in epithelial cells, and undoubtedly represents a key determinant of their function *in vivo*.

Keratin Gene Expression Mirrors Epithelial Differentiation: The Case of Skin

Skin provides a beautiful example of the tight relationship that exists between keratin gene regulation and epithelial differentiation and function. More than half of all keratin genes are expressed in mature mammalian skin tissue ([Figure 2](#) and [Table 1](#)). The architectural complexity of adult skin epithelia is achieved through a temporally and spatially regulated sequence of simple decisions, resulting in progressive restriction of fate determination and reflected by specific changes in keratin gene regulation. In the clinical setting, keratin typing is often exploited in diagnosing cancer type, its differentiation status (and therefore prognosis), as well as the origin of the cells forming metastatic foci. This strategy is also applied for diseases other than cancer.

Keratin Gene Expression in Thin Epidermis

In thin epidermis, mitotically active cells of the basal layer act as progenitors, and consistently express Krt5 and Krt14 as their main keratin pair ([Figure 2\(f\)](#)), along with lower levels of Krt15. Onset of differentiation coincides with the appearance

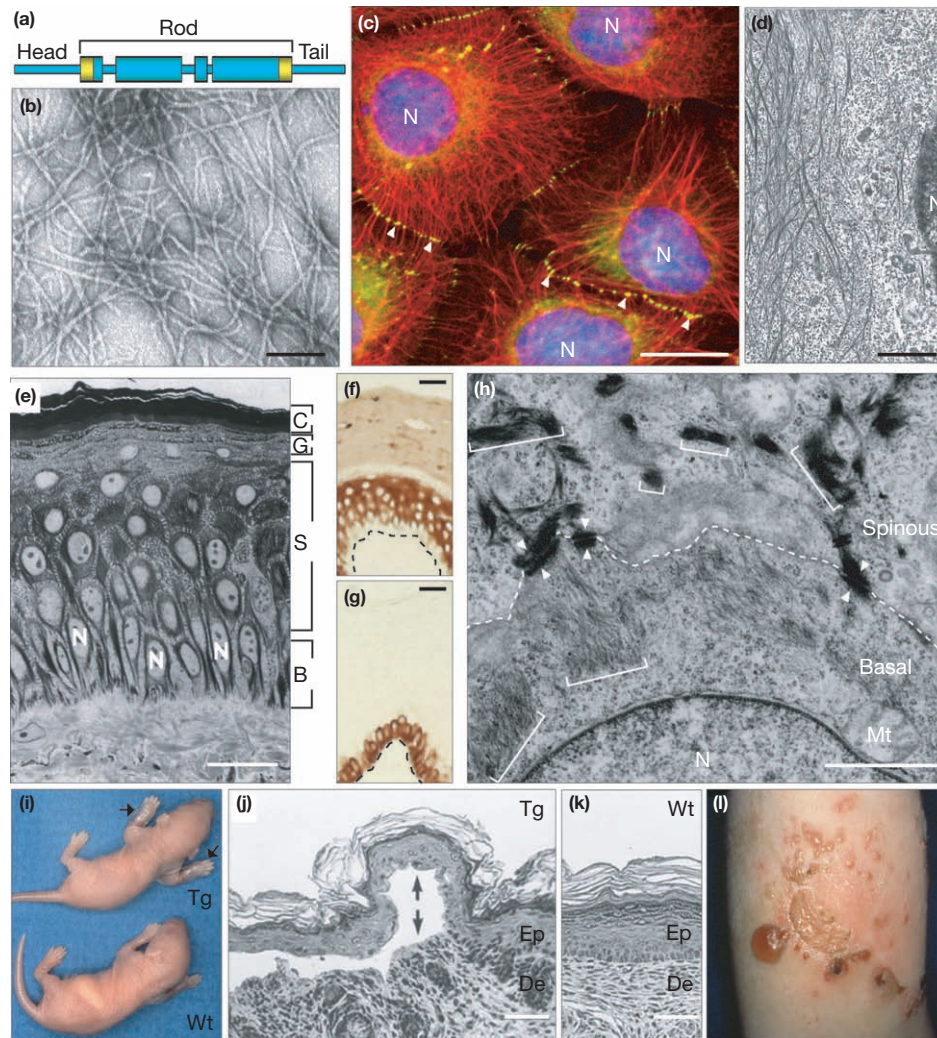


Figure 2 Attributes, differential regulation, and disease association of keratin. (a) Schematic representation of the tripartite domain structure shared by all keratin and other intermediate filament proteins. A central α -helical rod domain acts as the major determinant of self-assembly and is flanked by nonhelical head and tail domains at the N- and C-termini, respectively. The ends of the rod domain contain 15–20-amino-acid regions (yellow) highly conserved among all intermediate filaments. (b) Visualization of filaments, reconstituted *in vitro* from purified K5 and K14, by negative staining and electron microscopy. scale = 125 nm. (c) Triple-labeling for keratin (red chromophore) and desmoplakin, a desmosome component (green chromophore) and DNA (blue chromophore), by indirect immunofluorescence of epidermal cells in culture. Keratin filaments are organized in a network that spans the entire cytoplasm and are attached at points of desmosomal cell–cell contacts (arrowheads) between cells. Scale = $\sim 30 \mu\text{m}$. N, nucleus. (d) Ultrastructure of the cytoplasm of epidermal cells in primary culture, as visualized by transmission electron microscopy. Keratin filaments are abundant and tend to be organized in large bundles of loosely packed filaments in the cytoplasm. Scale = $2.5 \mu\text{m}$. (e) Histological cross-section of resin-embedded human trunk epidermis, revealing the basal (B), spinous (S), granular (G), and stratum corneum (C) compartments. Bar equals $\sim 50 \mu\text{m}$. N, nucleus. (f, g) Differential distribution of keratin epitopes on human skin tissue cross-sections (similar to frame (e)) as visualized by an antibody-based detection method. K14 occurs in the basal layer (g), where the epidermal progenitor cells reside. K10, on the other hand, is primarily concentrated in the differentiating, suprabasal layers of epidermis. Dashed line indicates the basal lamina. Scale = $\sim 100 \mu\text{m}$. (h) Ultrastructure of the boundary between the basal and suprabasal cells in mouse trunk epidermis, as seen by routine transmission electron microscopy. The sample, from which this micrograph was taken, is oriented in the same manner as frame (e). Organization of keratin filaments as loose bundles (see brackets in basal cell) correlates with the expression of K5–K14 in basal cells, whereas the formation of much denser and electron-dense filament bundles (see brackets in spinous cell) reflects the onset of K1–K10 expression in early differentiating cells. Arrowheads point to desmosomes connecting the two cells. Scale = $1 \mu\text{m}$. N, nucleus. (i) Newborn mouse littermates. The top mouse is transgenic (Tg) and expresses a mutated form of K14 in its epidermis. Unlike the control pup below (Wt), this transgenic newborn shows extensive blistering of its front paws (arrows). (j, k) Hematoxylin/eosin-stained histological cross-section through paraffin-embedded newborn mouse skin (similar to those shown in frame (i)). Compared to the intact skin of a control littermate ((k), Wt), the epidermis of the K14 mutant expressing transgenic pup ((j) Tg) shows intra-epidermal cleavage at the level of the basal layer, where the mutant keratin is expressed (see opposing arrows). Scale = $100 \mu\text{m}$. (l) Leg skin in a patient suffering from the Dowling–Meara (severe) form of epidermolysis bullosa simplex. Characteristic of this severe variant of this disease, several skin blisters are grouped in a herpetiform fashion.

Table 1 Distribution of keratins in the major compartments of skin epithelia

<i>Skin compartment</i>	<i>Cell type</i>	<i>Keratin expression (mRNA level)</i>
Interfollicular epidermis	Basal keratinocytes	Krt5, Krt14; also, Krt15
	Early differentiating keratinocytes	Krt1, Krt10
	Late differentiating keratinocytes	Krt2
	Merkel cells (in basal layer)	Krt8, Krt18, Krt20
Palmar/plantar epidermis	Basal keratinocytes	Krt5, Krt14; also Krt15, Krt17, Krt19
	Early differentiating keratinocytes	Krt1, Krt9, Krt10; Krt6, Krt16
	Late differentiating keratinocytes	Krt2
Hair follicles	Outer root sheath	Krt5, Krt14, Krt17; also, Krt15, Krt19
	Companion layer	Krt6a, Krt16 and Krt75, Krt17
	Inner root sheath	Krt25–28; Krt71–74; also, Krt7
	Hair shaft: cuticle and cortex	Primarily Krt31–38; Krt81–86
	Hair shaft: medulla	Krt17, Krt75; several others
	Glandular epithelium	Krt8, Krt18, Krt14; others?
Sebaceous glands (apocrine)	Cambial cells	Krt5, Krt14, Krt17
Sweat glands (eccrine)	Glandular epithelium	Krt8, Krt18; others?
	Myoepithelium	Krt5, Krt7, Krt14, Krt17
Nail	Duct	Krt5, Krt14; Krt7, Krt17; Krt6, Krt16
	Matrix	Krt5, Krt14, Krt17
	Nail plate	Primarily Krt31–38; Krt81–86
	Nail bed	Krt5, Krt14; Krt7, Krt17; Krt6, Krt16
	Proximal nail fold	Krt5, Krt14; Krt1, Krt10; Krt6, Krt16
	Hyponychium (upper layers)	Krt5, Krt14; Krt17; Krt6, Krt16

of the Krt1/Krt10 pair through a robust transcriptional induction that occurs at the expense of Krt5/Krt14. Accordingly, Krt1/Krt10 proteins are readily detectable in the lowermost suprabasal layer of epidermis (**Figure 2(g)**). The appearance of Krt1 and Krt10 correlates with a sudden and dramatic shift in the organization of cytoplasmic keratin filaments, which now exhibit significant bundling (**Figure 2(h)**). Another type II keratin gene, Krt2e, is expressed at a later stage of differentiation.

Keratin Gene Expression in Thick Epidermis

The epidermis of palm and sole skin is significantly thicker, owing to its specialization for resisting larger amounts of stress. This function is reflected in its architecture of alternating stripes of primary and secondary ridges, and specific changes in keratin expression. In the thick, stress-bearing, primary ridges, the major differentiation-specific type I Krt9 is presumed to foster a more resilient cytoskeleton. In the thinner secondary ridges, post-mitotic cells preferentially express the type II Krt6a and type I Krt16 and Krt17. Relative to Krt1, Krt9, and Krt10, the properties of Krt6a, Krt16, and Krt17 are believed to foster greater cellular pliability, thereby providing flexible hinge regions in-between the more rigid Krt1/Krt9-rich primary ridges. While it remains to be supported by experimentation, this attractive notion is consistent with the dramatic upregulation of K6a, K6b, K16, and K17 that occurs in keratinocytes recruited from wound margins to participate in the restoration of the epidermal barrier following injury.

Keratin Gene Expression in Hair Follicles

Consistent with its greater complexity, a larger number of keratin genes are transcribed in the pilosebaceous unit (**Table 1**).

A mature hair follicle houses eight distinct epithelial layers segregated into three histological compartments organized in concentric circles along its main axis. These compartments are, from inside out, the hair shaft proper; the IRS; and the outer root sheath (ORS), an epithelium that wraps around the follicle and is contiguous with the epidermis. The regulation of the >27 keratin genes expressed in hair follicles is detailed in **Table 1**. Key features are highlighted here. The Krt31–Krt38 and Krt81–Krt86 genes are unique to the hair shaft and related hard epithelia (e.g., nail; filiform papillae of the tongue). These keratin genes exhibit a stage-specific regulation during the differentiation of matrix progenitors in the bulb region to hair cortex keratinocytes. The cysteine-rich head and tail domains of the hard keratin proteins form disulfide bridges which, along with the γ -glutamyl cross-links catalyzed by transglutaminases, foster the organization of large bundles of densely packed keratin filaments. This organization endows hair (and nail) with its unique structural features, shape, and resilience. Krt25–28 and Krt71–74 are found in the IRS compartment. The ORS epithelium, on the other hand, expresses several of the keratins genes expressed in the epidermis, with the notable exception of the differentiation-specific Krt1, Krt2e, and Krt10 (**Table 1**). By the time an epidermal or hair cortex keratinocyte has completed differentiation, more than 80% of its protein content is keratin. This extreme situation is reminiscent of globin gene expression in erythrocytes, another highly specialized cell type, and underscores the fundamental importance of keratin proteins to surface epithelia.

Functions of Keratin Filaments

Structural Support

The properties typifying keratin filaments collectively point to an important structural role. Evidence for this function came from

Table 2 Representative skin disorders associated with mutations in keratin genes

<i>Disease</i>	<i>Main target tissue</i>	<i>Gene mutated</i>
Epidermolysis bullosa simplex	Basal cells/epidermis	Krt5, Krt14
Epidermolytic hyperkeratosis	Differ. cells/epidermis	Krt1, Krt10
Ichthyosis bullosa of Siemens	Differ. cells/epidermis	Krt2
Meesman's corneal dystrophy	Cornea	Krt3, Krt12
Monilethrix	Hair cortex	Krt81, Krt86
Oral white sponge nevus	Oral mucosa	Krt4, Krt13
Pachyonychia congenita		
Type 1 (Jadhasson–Lewandowsky)	Nail; other appendages	Krt6a, Krt16
Type 2 (Jackson–Lawler)	Nail; other appendages	Krt6b, Krt17
Palmoplantar keratoderma		
Epidermolytic	Palmar/plantar epidermis	Krt9
Nonepidermolytic	Palmar/plantar epidermis	Krt1, Krt16
Steatocystoma multiplex	Sebaceous glands	Krt17

transgenic mouse studies in which a mutated form of K14 capable of dominantly disrupting filament assembly and organization was targeted to the basal layer of epidermis *in vivo*. Unlike controls, mutant-expressing animals exhibited massive intracellular lysis of the basal cells in response to modest mechanical trauma to the skin (**Figures 2(i)–2(k)**). This mouse phenotype mimicked key aspects of a group of dominantly inherited epidermal disorders known as ‘epidermolysis bullosa simplex’ (EBS, **Figure 2(l)**), and it was shown shortly thereafter that the human condition arises from mutations in either Krt5 or Krt14. EBS turned out to be a very useful paradigm. We now know that keratin IFs fulfill a similar function in all complex epithelia, ranging from the surface of the eye (cornea) to hair tissue. *In vitro* studies showed that keratin IFs exhibit unique viscoelastic properties that support their involvement in a structural capacity, and that disease-causing mutations significantly decrease the strength of keratin assemblies subject to mechanical strain. Likewise, biophysical studies of live cells in culture showed that keratin IFs account for a substantial fraction of their viscoelastic properties, and that disrupting keratin cytoplasmic organization causes cells to soften considerably. Demonstration of a structural support role has been comparatively more difficult for simple epithelial tissues, but strong evidence to that effect now exists for liver hepatocytes as well as trophoblast giant cells.

Scaffolding Components of Signaling Pathways and Other Functions

Transgenic mouse studies that focused on adult liver and extra-embryonic tissue revealed that simple epithelial keratins 8/18 play an important cytoprotective role against chemical insults and in modulating the response to pro-apoptotic signals. In liver hepatocytes, the function of cytoprotection requires phosphorylation of K8/K18. In regard to apoptosis, studies showed that K8/K18 filaments can bind to, and either organize or sequester, key intracellular effectors of the signaling machinery downstream from engagement of the tumor necrosis factor- α (TNF- α) and fas receptors. Of note, the presence of K17 is also required for the survival of matrix epithelial cells in mature hair follicles. It appears likely that keratin IFs will be found to influence additional basic metabolic processes in the cell. In that regard, recent studies have shown that an intact K8/K18 filament network is required for the proper sorting of specific membrane proteins in polarized

simple epithelial cells. Determining how nonpolar assemblies are adapted to foster spatial organization in the cell promises to reveal additional keratin properties and functions.

Keratin Involvement in Disease

The role of keratin mutations in causing several epidermal, oral, ocular and hair-related diseases is well established (see **Table 2**, which emphasizes skin diseases). These diseases are genetically well defined and typically inherited in an autosomal-dominant fashion, although recessive inheritance is occasionally seen. The vast majority of genetic lesions consist of missense mutations, although premature stop codons, small deletions, and mutations affecting messenger RNA (mRNA) splicing occur with some frequency. In general, disease severity correlates with the location and nature of the mutation within the keratin backbone. A keratin mutation database is being maintained at the Centre for Molecular Medicine and the Bioinformatics Institute in Singapore and can be accessed online (Human Intermediate Filament Database, see section Relevant Website).

For most of the disorders listed in **Table 2**, mutations are causative for the disease and act through their ability to compromise the role of keratin IFs toward structural support. Mutations causing severe outcomes tend to affect amino acid residues that are highly conserved at either end of the central rod domain (**Figure 1(a)**), and often lead to the appearance of large aggregates of mispolymerized keratins in the cytoplasm of affected cells. In such instances, the aggregates appear to exacerbate the effect of the loss of a functional keratin IF network, and contribute to disease pathogenesis. Other types of mutations may also elicit alternative mechanisms of disease pathogenesis. Examples include mutations that could either affect keratin degradation during apoptosis or directly modulate keratin phosphorylation by introducing or removing a potential phosphorylation site. In contrast to other keratins, mutations affecting the simple epithelial keratins K8/K18 appear to predispose to, rather than cause disease *per se*. Again, this notion is supported by studies involving transgenic mice. Many of the diseases listed in **Table 2** are characterized by a clinical heterogeneity that cannot be entirely accounted for through the position and nature of the mutation along the keratin chain. For instance, the same missense allele of K17, Arg₉₄→Cys, can cause either type 2 Pachyonychia

Congenita (which affects primarily the nail) or Steatocystoma Multiplex (a glandular disorder with minimal if any nail alterations) in different kindreds. Together with studies involving gene manipulation in mice, this finding points to a sometimes dramatic impact of the genetic background on disease expression.

See also: [Cell Architecture and Function: Desmosomes and Hemidesmosomes](#); [Intermediate Filaments](#); [Intermediate Filament Linker Proteins: Plectin and BPAG1](#); [Neuronal Intermediate Filaments](#); [Nuclear Lamina](#).

Further Reading

- Coulombe PA, Kerns ML, and Fuchs E (2009) Epidermolysis bullosa simplex: A paradigm for disorders of tissue fragility. *Journal of Clinical Investigation* 119: 1784–1793.
- Coulombe PA and Omary MB (2002) 'Hard' and 'soft' principles defining the structure, function, and regulation of keratin intermediate filaments. *Current Opinion in Cell Biology* 14: 110–122.

- Erber A, Rierner D, Bovenschulte M, and Weber K (1998) Molecular phylogeny of metazoan intermediate filament proteins. *Journal of Molecular Evolution* 47: 751–762.
- Fuchs E (1995) Keratins and the skin. *Annual Review of Cell and Development Biology* 11: 123–153.
- Fuchs E and Cleveland DW (1998) A structural scaffolding of intermediate filaments in health and disease. *Science* 279: 514–519.
- Magin TM, Vijayaraj P, and Leube RE (2007) Structural and regulatory functions of keratins. *Experimental Cell Research* 313: 2021–2032.
- Moll R, Franke WW, Schiller DL, Geiger B, and Krepler R (1982) The catalog of human cytokeratins: Patterns of expression in normal epithelia, tumors, and cultured cells. *Cell* 31: 11–24.
- Omary MB, Coulombe PA, and McLean WHI (2004) Intermediate filament proteins and their associated diseases. *New England Journal of Medicine* 351: 2087–2100.
- Omary MB, Ku NO, Tao GZ, Toivola DM, and Liao J (2006) "Heads and tails" of intermediate filament phosphorylation: Multiple sites and functional insights. *Trends in Biochemical Sciences* 31: 383–394.
- Szeverenyi I, Cassidy AJ, Chung CW, et al. (2008) The human intermediate filament database: Comprehensive information on a gene family involved in many human diseases. *Human Mutations* 29: 351–360.
- Toivola DM, Strnad P, Habtezion A, and Omary MB (2010) Intermediate filaments take the heat as stress proteins. *Trends in Cell Biology* 20: 79–91.

Relevant Website

<http://www.interfil.org> — Human Intermediate Filament Database.

Kinesins as Microtubule Disassembly Enzymes

R Ohi, Vanderbilt University Medical Center, Nashville, TN, USA

L Wordeman, University of Washington, Seattle, WA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Susan L. Kline-Smith and Arshad Desai, volume 2, pp. 517–521, © 2004, Elsevier Inc.

Glossary

Centromere The primary constriction of a mitotic chromosome that connects two sister chromatids.

Kinetochores The proteinaceous structure that assembles onto each side of the centromere, allowing for chromosome

attachment to and movement on microtubules of the mitotic spindle.

Microtubule Polar, cylindrical polymers formed from ~13 laterally associated protofilaments of $\alpha\beta$ -tubulin.

Kinesins constitute a large evolutionarily conserved superfamily of enzymes that are fundamentally important for many aspects of unicellular and multicellular life. These molecular motors bind microtubules, which are polar cylindrical polymers of $\alpha\beta$ -tubulin that are important for spatial organization within the cell. Kinesin-1, the first kinesin to be isolated, was purified by taking advantage of its ability to couple ATP hydrolysis to translocation along microtubules. This activity underlies its ability to drive long-range transport of cellular cargos from one region of a cell to another. Sequences homologous to the catalytic domain responsible for the motor activity of Kinesin-1 have been discovered in a large number of proteins, and the complete genomic complement of kinesins is known for eukaryotes with fully sequenced genomes.

Although motility was the defining activity of kinesins, recent work has revealed that the catalytic domain is used in a subset of kinesins to drive microtubule disassembly rather than translocation. Physiologically, the regulation of microtubule dynamics by kinesins is likely to be of great significance in cellular organization, division, and differentiation. Here, we first briefly review the work that led to the recognition of specific kinesins acting as microtubule disassembly enzymes. We then discuss progress toward understanding the mechanism by which the catalytic domain is used to drive microtubule depolymerization and finally comment on the physiological relevance of this nonmotile activity.

Microtubule Dynamics Regulation in Cells – the Need for Depolymerases

In vitro, microtubules exhibit dynamic instability, a steady-state behavior wherein individual microtubule ends will interconvert between assembly and disassembly. The term catastrophe describes the transition from assembly to shrinkage, whereas rescue describes the transition from disassembly to assembly. In principle, dynamic instability permits the cell to rapidly reorganize its microtubule array by changing the number of microtubule nucleating sites or altering the frequency of the transitions between assembly and disassembly without changing the concentration of tubulin dimers via protein synthesis.

Early landmark studies of microtubules at the edges of live cells recorded similar stochastic transitions between assembly and disassembly which were independent of the behavior of neighboring microtubules. However, more recent studies, where the assembling tips of live microtubules are reported using '+TIP' proteins, enable microtubule polymerization to be scored at any position within the living cell. These studies reveal a distinctly different picture of microtubule turnover in cells. Consistent with the observation that microtubules assembled *in vitro* from tubulin purified from non-neuronal sources is considerably less dynamic than tubulin purified from brain, we now know that microtubules in cultured epithelial cells assemble persistently from the centrosome to the cell margin. Only upon reaching the cell margin are microtubule ends likely to transition between phases of shortening, assembly, and paused states. This behavior is accomplished by a host of assembly and disassembly factors that camouflage the intrinsic dynamic instability of microtubules.

Persistent assembly of microtubules is under the control of proteins such as the +TIP EB1 that, in addition to its own microtubule assembly promoting activity, can recruit microtubule regulatory proteins to the microtubule tip. Dynamic instability of microtubules in cells is suppressed by assembly factors and must, in turn, be reversed by disassembly factors. This service is performed, to a large extent, by two key kinesin protein families: the Kinesin-13 family of microtubule depolymerizers and, potentially, the Kinesin-8 family (Figure 1).

The most potent and well-characterized of the Kinesin-13s is MCAK/Kif2C, which removes tubulin dimers in an ATP-dependent fashion from the ends of artificially stabilized microtubules, such as those assembled with the GTP analog guanosine-5'-[(α , β)-methylene]triphosphate (GMP-CPP) (Figure 2(a)). For physiological microtubules that are intrinsically unstable, MCAK/Kif2C triggers transitions from assembly to disassembly (catastrophe) without altering either the rate of disassembly or assembly itself. Although MCAK/Kif2C strongly catalyzes catastrophes on microtubules in isolation; its activity is significantly enhanced both in cells and *in vitro* by recruitment to microtubule ends by EB-family members. Notably, of the three mammalian Kinesin-13 family members, only MCAK/Kif2C is directly recruited to microtubule ends by EB proteins.



Figure 1 Domain structure of the microtubule-depolymerizing kinesins. Proteins containing a kinesin-like catalytic domain (green) are defined as members of the kinesin superfamily. Motors are oriented with their N-termini positioned leftward. Kinesin-13s contain an internally located motor domain that is sandwiched by a globular N-terminus (ochre) and a C-terminal coiled-coil (black). Kinesin-8s are N-terminal kinesins, with a C-terminal tail (red) that is linked to the motor domain by a coiled coil. Kinesin-14s are C-terminal kinesins with a tail domain located at the N-terminus.

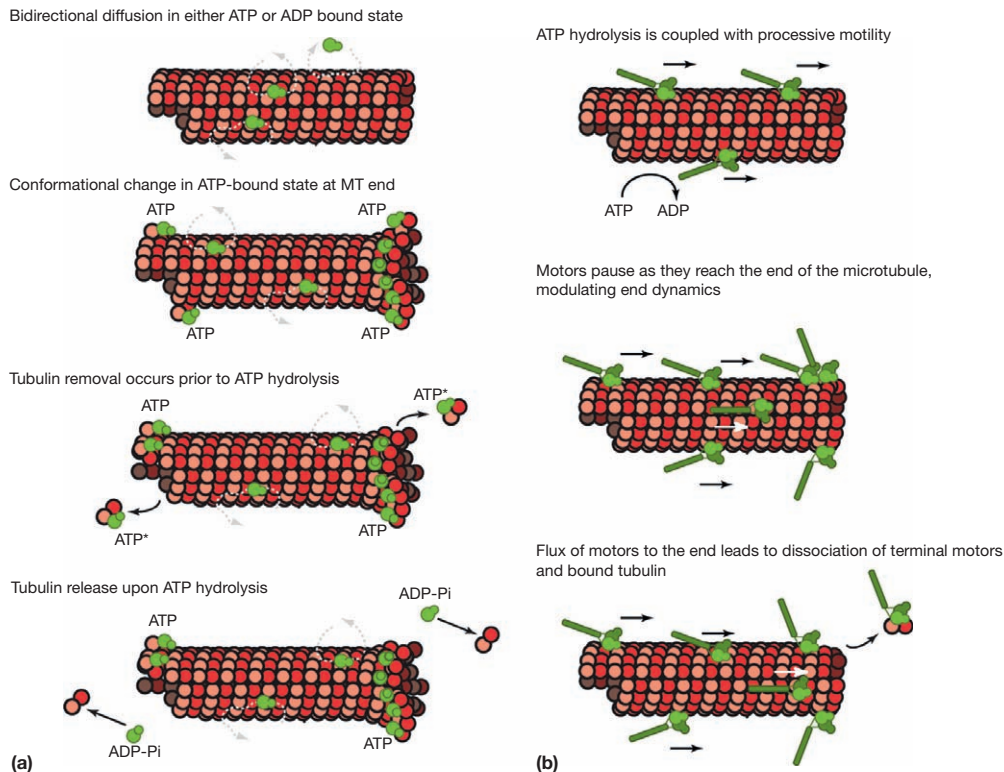


Figure 2 Enzymology of microtubule-depolymerizing kinesins. (a) Kinesin-13 family. MCAK/Kif2C (Kinesin-13) can diffuse bidirectionally on the microtubule lattice in the absence of nucleotide hydrolysis. Upon encountering either the plus or minus end of the microtubule, the motor can form a high-affinity complex with the bent protofilament in an ATP-bound state. The motor removes a tubulin dimer from the microtubule coordinately with the attainment of the transition state for ATP hydrolysis. Hydrolysis culminates in the release of the free tubulin dimer from the motor itself recycling the motor for another round of tubulin removal. Kinesin-13 motors can disassemble microtubules from either end. (b) Kinesin-8 family. Kinesin-8 motors utilize the energy of ATP hydrolysis to translocate many hundreds of steps toward the plus end of the microtubule without detaching from the polymer. Motors pause at the plus end and may suppress dynamics there via an unknown mechanism. In the case of Kip3 (*S. cerevisiae*), the arrival of another motor triggers the loss of the terminal motor, and its associated tubulin dimer, leading to preferential disassembly of the plus end. It is not known how ATP hydrolysis contributes to this flux-dependent tubulin removal.

The Kinesin-8s also significantly influence microtubule dynamics. On stabilized microtubules *in vitro*, Kinesin-8s will walk unidirectionally toward the plus end of the microtubule in an ATP-dependent fashion where they proceed to disassemble the microtubule from the plus end only (Figure 2(b)). In cells, the effects of these kinesins are more equivocal as they appear to suppress both assembly and disassembly of microtubules.

This leads to a general picture of microtubule turnover in cells that is dependent on Kinesin-8s, Kinesin-13s, and

numerous other regulatory factors. While dynamic instability is a key property that describes the fundamental behavior of microtubules, microtubules in cells are considerably less dynamic than previously appreciated. Microtubules in cells depend on accessory factors, such as the end binding (EB) proteins, to promote persistent assembly and disassembly factors, such as Kinesin-13s and Kinesin-8s, to antagonize assembly and promote turnover. Utilizing these regulatory factors, microtubules can effectively fill the cellular space available.

Microtubule-Depolymerizing Kinesins: Key Discoveries

In the mid-1980s, Ron Vale and Scott Brady independently reported the purification of a force-generating protein from the axoplasm of the squid giant axon that is capable of driving vesicle transport. This protein, termed conventional kinesin, is a homodimer that binds microtubules via one of two catalytic motor domains and uses energy derived from adenosine triphosphate (ATP) hydrolysis to translocate toward microtubule plus ends. Conventional kinesin, classified as a Kinesin-1, is one of 45 kinesin-like proteins encoded by the human genome. It was generally assumed that all kinesins would be transport motors that move along microtubules without influencing their dynamic properties.

In 1994, Endow and colleagues provided the first evidence that kinesins might also regulate the dynamic behavior of the tracks they translocate on. They observed that budding yeast Kar3 (Figure 1) is a slow minus-end-directed motor that has the capacity to depolymerize Paclitaxel-stabilized microtubules from minus ends. This finding was consistent with the observation that benomyl, a microtubule-destabilizing drug, could suppress the karyogamy defects exhibited by *kar3* mutants. In yeast, Kar3 controls microtubule plus-end dynamics during mating and is targeted to this location using the regulatory factor Cik1, a protein that heterodimerizes with Kar3. Kar3, a Kinesin-14, is evolutionarily conserved and its orthologs are characterized by having their catalytic domains at the C-terminus. Other well-studied Kinesin-14s such as *Drosophila* *ncd*, *Xenopus laevis* XCTK2, and human HSET also display minus-end-directed motility but thus far have not been reported to possess a microtubule destabilizing activity. Microtubule depolymerase activity may thus be restricted to fungal Kinesin-14s.

Members of the Kinesin-13 subfamily are the most potent of the microtubule disassembling kinesins. These motors are ancient, with orthologs present in divergent eukaryotes such as the intestinal parasite *Giardia intestinalis*. Kinesin-13s have an internally positioned motor domain (Figure 1) that is incapable of powering ATP-dependent motility. Kinesin-13s instead utilize ATP hydrolysis to catalyze the removal of tubulin subunits from both the plus and minus ends of microtubule protofilaments (Figure 2(a)). Humans express four Kinesin-13s: mitotic centromere-associated kinesin (MCAK; also called Kif2C), the first Kinesin-13 to be identified; Kif2A; Kif2B; and KIF24, a little studied motor thought to be the ancestral Kinesin-13.

Seminal studies of XKCM1, the MCAK/Kif2C ortholog in the African clawed frog *X. laevis*, in frog egg extracts were critical in demonstrating that Kinesin-13s are microtubule-destabilizing enzymes. Inactivation of XKCM1 led to the formation of long microtubules (Figure 5(a)), an effect that could be attributed to a reduction in the frequency at which microtubules transited from growth to shortening (catastrophe). Subsequent experiments with pure MCAK/Kif2C and Kif2A proteins revealed that Kinesin-13s could disassemble microtubules *in vitro* (Figure 3(a)), establishing a direct role for Kinesin-13s in the regulation of microtubule dynamics. Owing to their potency and biological importance, Kinesin-13s are the best understood and most intensively studied of the microtubule depolymerizing kinesins.

Kip3, a budding yeast Kinesin-8, is the most recent addition to the group of kinesins that possess a microtubule depolymerizing activity (Figure 1). Because yeast cells lacking *KIP3* contain long microtubules and because the motor domain of Kip3 is least divergent from the Kinesin-13s, it had long been postulated that members of the Kinesin-8 subfamily might be microtubule depolymerases. Recent studies with pure Kip3 motor have confirmed this and demonstrated that Kip3 is also a plus-end-directed motor. It is currently unclear whether the microtubule depolymerization activity of Kip3 has been retained throughout the eukaryotic kingdom. The heterodimeric Klp5/Klp6 Kinesin-8 from the evolutionarily distinct fission yeast *Schizosaccharomyces pombe* exhibits plus-end-directed motility but does not depolymerize microtubules *in vitro*. Similarly, Kif18A, one of three Kinesin-8s encoded by the human genome, does not appear to catalyze microtubule disassembly, but rather attenuates the dynamics of microtubule plus ends.

Biophysical Studies of Kip3 and Kinesin-13s

Considerable insight into Kinesin-13 and Kinesin-8 mechanisms has been obtained from biophysical and structural studies. The mysterious preference of MCAK/Kif2C for microtubule ends was explained by structural studies demonstrating that the high-affinity binding site for Kinesin-13 family members was a curved, rather than straight, protofilament. Curved protofilaments are most easily induced at the microtubule end. These motors stabilize the curved protofilament structure characteristic of depolymerizing microtubules. For unstabilized microtubules, such as those found in living cells, this has the effect of promoting catastrophes. Because Kinesin-13 family members can promote this activity only at the ends of microtubules, the motors must find the end of the microtubule without the benefit of directed motility. These motors can associate with the microtubule surface and diffuse in a directionally unbiased manner without any ATP hydrolysis until a microtubule end is encountered (Figure 3(c)). The rapidity with which the motor can associate with the microtubule lattice (on-rate) is the principal parameter governing the depolymerizing activity of the motor. This parameter is controlled by electrostatic interactions between the microtubule and positively charged regions outside of the motor domain. By changing the on-rate for the microtubule lattice, phosphorylation or posttranslational modifications of the microtubule itself can modulate the activity of these microtubule depolymerizers. Once associated with the microtubule lattice, the motor can use diffusive activity to encounter the microtubule end.

Once MCAK/Kif2C reaches the end of the microtubule, a high-affinity binding state to the curved protofilament end can be achieved (Figure 3(b)). Mutational analysis of switch II mutants of MCAK/Kif2C suggests that in the ATP-bound state, MCAK/Kif2C assumes a high-affinity state bound to the curved protofilament, breaks off one dimer as it reaches the ATP transition state, and then releases the dimer as hydrolysis is completed. Kinetic and biophysical studies have also predicted that one molecule of MCAK/Kif2C can remove approximately 20 tubulin dimers from GMP-CPP-stabilized microtubules before releasing from the microtubule. Whether

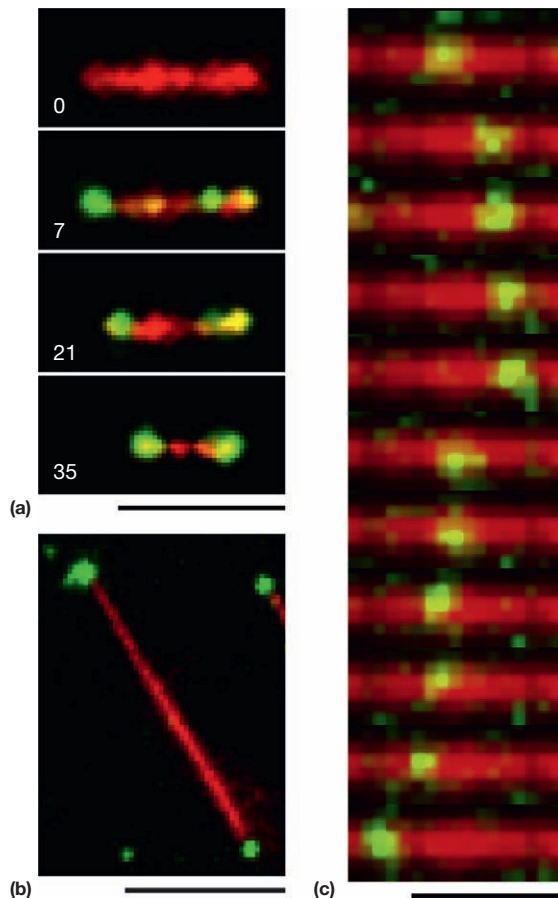


Figure 3 Biophysical properties of the Kinesin-13 MCAK/Kif2C. (a) MCAK/Kif2C depolymerizes microtubules from both ends. In the presence of ATP, GFP-MCAK/Kif2C targets the ends of a microtubule (red; here, stabilized with Paclitaxel) and catalyzes the removal of tubulin subunits, which is visualized as microtubule shortening. Time following exposure of the microtubule to GFP-MCAK/Kif2C is indicated in minutes. Scale = 1 μ m. (b) MCAK/Kif2C exhibits high affinity for microtubule ends. In the presence of the nonhydrolyzable ATP analog AMP-PNP, GFP-MCAK/Kif2C localizes to microtubule ends. Scale = 2 μ m. (c) MCAK/Kif2C diffuses along the microtubule lattice. In addition to direct microtubule end-binding, MCAK/Kif2C also targets microtubule ends by two-dimensional diffusion on the microtubule lattice. Shown are still images from a video depicting ATP-independent movement of GFP-MCAK/Kif2C along a microtubule. Time step is 200 ms. Scale = 2 μ m. (a) Image provided by Jeremy Cooper and Linda Wordeman. (b) Image provided by Ayana Moore and Linda Wordeman. (c) Image provided by Jeremy Cooper and Linda Wordeman.

MCAK/Kif2C would track with a disassembling dynamic microtubule once eliciting a catastrophe is not known, although a related *Drosophila* Kinesin-13 has been observed to track with the plus ends of disassembling microtubules. Finally, unlike the case with motile Kinesin-1s, it is still not known how or if the two heads of dimeric MCAK/Kif2C molecules are coordinated during tubulin removal.

The Kinesin-8 family members represent an interesting intersection between the microtubule depolymerizing kinesins described above and the well-studied motile kinesins of the Kinesin-1 family. Similar to Kinesin-1s, Kinesin-8s translocate to the plus end of the microtubule in an ATP-dependent

manner. The *S. cerevisiae* Kinesin-8, Kip3, can disassemble microtubules specifically from the plus end once a sufficient concentration of motors has accumulated there. Single-molecule studies reveal that Kinesin-8 motors exhibit strikingly higher processivity than the Kinesin-1s, taking many hundreds of steps before detaching from the microtubule. This interesting property suggests that Kinesin-8 motors are capable of translocating uninterrupted from one end of a cellular microtubule to the other without detaching. Furthermore, if the motors are capable of binding to the microtubule at any point along the lattice, the resulting motors will accumulate in a gradient extending from the plus end of the microtubule, a property that has been observed in mammalian Kinesin-8s. This has been proposed as a mechanism by which cells can set microtubule length or impart positional information along the length of the microtubule.

Upon reaching the plus end of the microtubule, the motor pauses until another motor approaches from behind. The arrival of another motor leads to the release of both the end-attached motor and its terminal tubulin dimer (Figure 2(b)) in a mechanism that is not understood but seems distinctly different from the Kinesin-13 mechanism of tubulin removal (Figure 2(a)). This behavior, in conjunction with observations from cellular and genetic studies, has led researchers to classify Kinesin-8s as modulators of microtubule-end dynamics similar to Kinesin-13s. While *in vitro* biophysical and kinetic studies utilizing dynamic microtubules have confirmed that Kinesin-13s promote microtubule catastrophes and lead to increased microtubule dynamics, similarly definitive studies using live microtubules remain ongoing for Kinesin-8s. However, preliminary studies in cellular systems suggest that Kinesin-8s regulate microtubule-end dynamics differently, possibly antagonistically, relative to Kinesin-13s.

Physiological Roles of Microtubule-Depolymerizing Kinesins

Microtubule-depolymerizing kinesins control microtubule polymerization dynamics to influence the architecture and functions of cellular microtubule-based assemblies. A key process in cell physiology that microtubule-depolymerizing kinesins participate in is cell division. During this period, the interphase microtubule array disassembles and re-organizes into the mitotic spindle, a highly dynamic apparatus that powers chromosome segregation. In part, this gross reorganization of the microtubule cytoskeleton is driven by a dramatic increase in microtubule dynamicity owing to a nearly five- to tenfold increase in the frequency of microtubule catastrophes. Elegant *in vitro* and *in vivo* studies have demonstrated that Kinesin-13s, particularly MCAK/Kif2C, are largely responsible for this change in microtubule dynamics. Without MCAK/Kif2C, spindle microtubules exhibit unbounded growth, an effect which impacts mitotic spindle assembly to varying degrees depending on the biological system. In embryonic *Xenopus* egg extracts, where the tubulin monomer pool is large, spindle formation is completely blocked whereas inhibition of MCAK/Kif2C in metazoan somatic cells leads to increases in spindle microtubule polymer, delayed centrosome separation, and delayed mitosis (Figures 5(a) and 5(b)).

In fungi, Kinesin-8 proteins also negatively regulate microtubule length, but the mechanism(s) and degree to which these proteins modify microtubule dynamics are unclear since most studies of these factors have been performed in live cells or on nonphysiological GMP-CPP microtubule substrates. Gene deletions of Kinesin-8s in both budding and fission yeasts (Kip3 and klp5/klp6, respectively) cause interphase microtubules to grow excessively long, and they are often observed to curl around cell tips. Catastrophe frequencies are reduced in cells lacking Kip3 and klp5/klp6 and, as discussed above, it is thought that these motors might actively catalyze microtubule shortening. In budding yeast, astral microtubule–cortex interactions specify the orientation and positioning of the mitotic spindle. Shortly after establishing a linkage, cortically attached microtubules depolymerize without dissociating from cortical attachment factors, causing nuclear translocation. Yeast cells lacking Kip3 misposition the mitotic spindle as a consequence of increased astral microtubule stability.

Kinesin-13s and Kinesin-8s also have critical functions at or near the plus ends of microtubules that interact with kinetochores, the chromosomal regions that constitute their primary interaction site with spindle microtubules. All Kinesin-13s and Kinesin-8s analyzed thus far exhibit prominent kinetochore localization (Figures 4(a)–4(c)). MCAK/Kif2C localizes to the inner kinetochore and centromeres (Figure 4(a)) and at this location polices kinetochore-microtubule attachments to ensure that chromosome segregation occurs without error. To segregate properly, chromosomes bi-orient or attach spindle microtubules such that each sister kinetochore is anchored to opposite spindle poles. MCAK/Kif2C depolymerizes flawed kinetochore-microtubule attachments, and, consistently, chromosomes are observed to mis-segregate in cells siRNA-depleted of MCAK/Kif2C (Figure 5(c)). How MCAK/Kif2C distinguishes between correct and improper attachments is not well understood. One model posits that MCAK/Kif2C continuously acts to depolymerize all kinetochore microtubules, and that

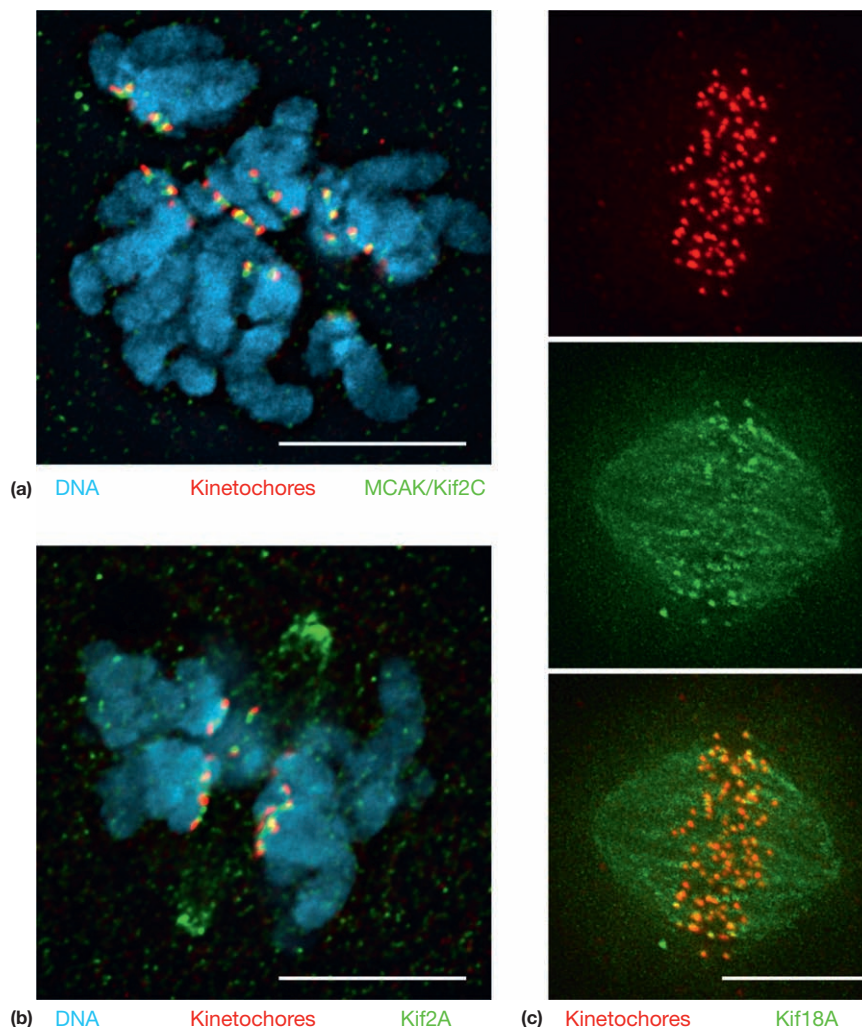


Figure 4 Localization of vertebrate microtubule-depolymerizing kinesins. (a) Localization of MCAK/Kif2C in a mitotic mammalian epithelial (PtK1) cell. The cell was fixed and immunostained for MCAK/Kif2C (green), kinetochores (red), and DNA (blue). (b) Localization of Kif2A in a PtK1 cell. The cell was fixed and immunostained for Kif2A (green), kinetochores (red), and DNA (blue). (c) Localization of Kif18A in mitotic human cervical carcinoma (HeLa) cell. The cell was fixed and immunostained for Kif18A (green) and kinetochores (red). Scale = 10 μm. Images provided by Ryoma Ohi.

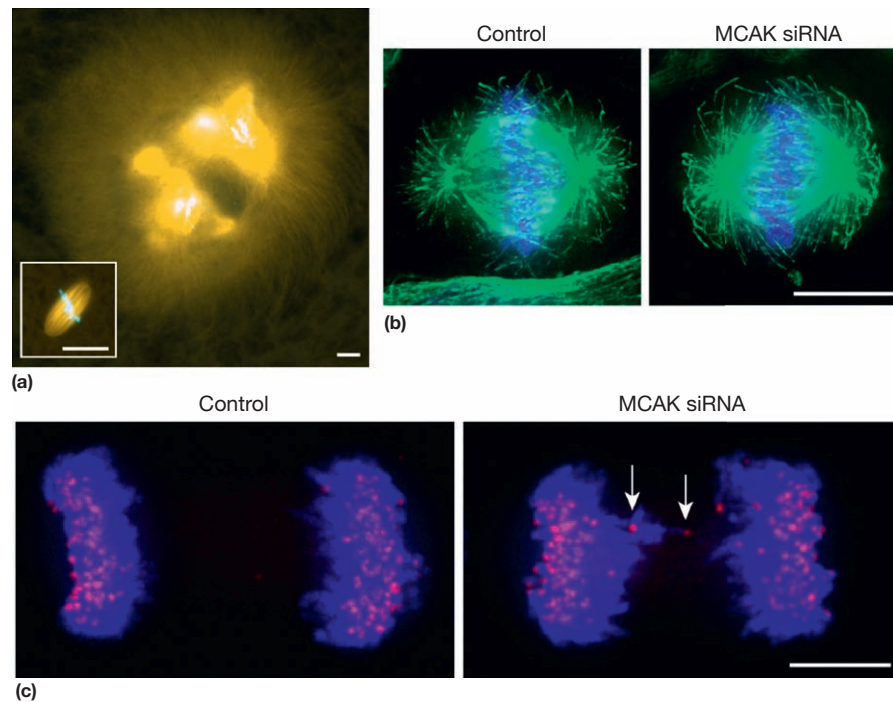


Figure 5 MCAK/Kif2C uses its microtubule-depolymerizing activity to limit the length of mitotic spindle microtubules and prevent erroneous kinetochore-microtubule attachments. (a) The immunodepletion of MCAK/Kif2C from *Xenopus* egg extracts causes microtubules (orange) to grow abnormally long as a result of a reduction in microtubule catastrophe frequency. Normal spindles (inset) fail to form. Scale = 20 μm . (b) Removal of MCAK/Kif2C from HeLa cells by RNAi causes lengthening of astral microtubules (green). Scale = 10 μm . (c) MCAK/Kif2C prevents errors in kinetochore-microtubule attachment. In anaphase cells, siRNA-depleted of MCAK/Kif2C, mis-attached kinetochores (red) are observed to lag at the spindle equator (arrows). Scale = 10 μm . (a) Image provided by Ryoma Ohi. (b) Image provided by Kathleen Rankin and Linda Wordeman. (c) Image provided by Sarah Domnitz and Linda Wordeman.

increased stability of proper kinetochore-microtubule attachments protects them from detaching. However, MCAK/Kif2C is known to be regulated by the Aurora B kinase, a molecule with a well-established role in correcting kinetochore-microtubule attachment errors, and it is also expected that the ability of MCAK/Kif2C to depolymerize improper attachments is actively controlled by phosphorylation.

Kinesin-8 proteins localize to the plus ends of kinetochore microtubules (Figure 4(c)), where they play a key role in positioning chromosomes at the spindle equator. Careful analysis of chromosome movements in tissue culture cells siRNA-depleted of the Kinesin-8 motor Kif18A demonstrated that this motor acts to slow kinetochore movements and increase the frequency at which chromosomes switch their direction of movement. Paradoxically, this effect opposes the expected consequence of removing a kinetochore-associated microtubule depolymerase. It is conceivable that Kinesin-8 proteins act differently on chemically distinct microtubule substrates. The microtubule depolymerase activity of Kip3 has been studied most intensively on GMP-CPP microtubules that mimic the stable GTP cap, but how Kip3 acts on dynamic microtubules assembled from GTP is not known. Human Kif18A has been reported to attenuate the polymerization dynamics of GTP microtubules in a manner which reflects the effects of the motor on kinetochore oscillations. It is likely that ongoing work will clarify the mechanism(s) by which Kinesin-8 proteins govern kinetochore movements.

Kinesin-13s, particularly Kif2A (Klp10A in the fruit fly *Drosophila melanogaster*), also localize to spindle poles (Figure 4(b)) where they are thought to drive the poleward translocation of kinetochore microtubules in the mitotic and meiotic spindles of higher eukaryotic cells. In this process, termed poleward microtubule flux, kinetochore-attached microtubules treadmill polewards by depolymerizing at spindle poles where their minus ends are anchored. Depending on cell type, poleward flux of kinetochore microtubules can produce 20–100% of the forces necessary for sister chromatid segregation during anaphase. Although data suggesting that Kinesin-13s drive polewards flux in somatic tissue cells is strong, it is unclear if Kinesin-13s are important for poleward flux in the *Xenopus* extract spindle. In this system, the poleward translocation of most spindle microtubules is driven by Eg5, a homotetrameric Kinesin-5 motor that slides anti-parallel microtubules apart, and is largely unaffected by disrupting MCAK/Kif2C and Kif2A functions. Unlike somatic cells, however, kinetochore microtubules represent a small fraction of spindle microtubules and it is possible that changes in the flux rates of kinetochore microtubules are particularly difficult to measure.

MCAK/Kif2C limits the length of astral microtubules during cytokinesis, the process that individualizes two proto-daughter cells. Abnormally long microtubules, such as those produced in the absence of MCAK/Kif2C (Figure 5(b)), cause the anaphase spindle to oscillate along the spindle axis, putting the

cell at risk of incorrectly segregating its replicated genome. Such spindle rocking is caused by the formation and resorption of abnormally large blebs, which are balloon-like protrusions on the cell surface that form as a result of positive hydrostatic forces that rupture membrane-actin bonds. Long microtubules promote bleb inflation by invading them and delaying their myosin II-dependent resorption. If blebs are too large, strong cytoplasmic currents are generated that are capable of displacing the entire mitotic spindle.

The organismal function of Kif2A has been analyzed in mice through gene knockout technology. *Kif2A*^{-/-} mice survive to gestation, but die within 1 day after birth as a result of feeding abnormalities. This behavioral defect is thought to arise from widespread mistakes in neuronal wiring. Cortical neurons from homozygous-mutant animals exhibit migratory defects that are caused by the formation of long collateral branches, projections that form along the length of the primary axon. Collateral branch extension is thought to be driven by protrusive forces generated by microtubule plus ends that polymerize against the neurite edge. Excess Kif2A appears to suppress collateral branch formation by depolymerizing such microtubule plus ends.

Lastly, an ancient function of the Kinesin-13 motors is to regulate the length of flagella, slender whiplike microtubule-based projections that beat with regular rhythmicity. Flagellar length is dictated by a balance of microtubule-plus-end assembly and disassembly at the distal tip. Flagellar beating

powers the motility of sperm and many unicellular organisms, including protozoa of the Trypanosomatida family. *Leishmania* and *Trypanosoma*, two members of this family, have been shown to encode five Kinesin-13s. One, LmjKIN13-2, localizes to flagella tips where it uses its microtubule depolymerizing activity to actively shorten flagella. The single Kinesin-13 homolog in *G. intestinalis* also regulates flagellar length using ATP-dependent depolymerase activity. Whether Kinesin-13s regulate flagellar or ciliary length in animal cells is currently unknown.

See also: **Cell Architecture and Function:** Centromeres; Kinesin Superfamily Proteins; Microtubule-Associated Proteins; Tubulin and Its Isoforms.

Further Reading

- Desai A and Mitchison TJ (1997) Microtubule polymerization dynamics. *Annual Review of Cell and Developmental Biology* 13: 83–117.
- Ems-McClung SC and Walczak CE (2010) Kinesin-13s in mitosis: Key players in the spatial and temporal organization of spindle microtubules. *Seminars in Cell and Developmental Biology* 21: 276–282.
- Howard J and Hyman AA (2007) Microtubule polymerases and depolymerases. *Current Opinion in Cell Biology* 19: 31–35.
- Moores CA and Milligan RA (2006) Lucky 13 – microtubule depolymerization by kinesin-13 motors. *Journal of Cell Science* 119: 3905–3913.
- Vale RD and Fletterick RJ (1997) The design plan of kinesin motors. *Annual Review of Cell and Developmental Biology* 13: 745–777.

Kinesin Superfamily Proteins

N Hirokawa, University of Tokyo, Tokyo, Japan

R Takemura, National College of Nursing, Tokyo, Japan

© 2013 Elsevier Inc. All rights reserved.

Glossary

Axonal transport Axons transmit the signal from the cell body to the synaptic terminal. Because axons lack ribosomes, that is, the protein synthetic machinery, most of the proteins needed in the axon and synaptic terminal need to be synthesized at the cell body and transported. The transport in the direction from the cell body to synapse is called the 'anterograde transport'. Some molecules are transported in the opposite direction ('retrograde transport'). The proximal part of the dendrite contains ribosomes, although the number reduces at the distal region.

Cilia and flagella These contain microtubules in the '9 + 2' axonemal configuration (nine doublets and central pair of microtubules). The microtubules in the doublets slide along one another by the stroke of axonemal dynein, which bridges neighboring doublets.

Dynein Minus-end-directed microtubule motor. It forms a massive multisubunit complex composed of two to three heavy chains (~530 kDa), which have motor domains, and variable numbers of associated intermediate and light chains. There are two classes, axonemal and cytoplasmic.

Axonemal dyneins are attached to microtubule doublets in cilia and flagellar, and slide against microtubules producing the beating movement. Cytoplasmic dyneins are expressed in most eukaryotic cells and important in vesicle trafficking and cell division.

Microtubule A filament of 25 nm diameter assembled from α - and β -tubulin subunits (55 kDa). Microtubules, microfilaments (actin filaments), and intermediate filaments are the major cytoskeletal filaments. In addition to the structural role, microtubules serve as rails for the long-distance transport, and microfilaments serve as rails for the short-distance transport. Microtubules are also reorganized as mitotic spindle during mitosis, and play an important role in cell division.

Mitosis Replicated chromosomes are separated into two daughter cells, by a complex cytoskeletal machine called 'mitotic spindles'. The process proceeds as prophase, prometaphase, metaphase, anaphase, and telophase. Several different classes of KIFs and the cytoplasmic dynein participate in the complex movement of mitotic spindles.

Kinesin was first identified biochemically as a microtubule-dependent motor protein responsible for transport of membranous organelles in the axon. It is now recognized that microtubule-dependent motor proteins form a large gene family, kinesin superfamily proteins (KIFs). The human genome contains 45 KIF genes. KIFs have high homology at the so-called 'motor domain', which is a globular domain responsible for moving along microtubules by hydrolysis of adenosine triphosphate (ATP). Outside the motor domain, the sequence is unique to each member. The motors bind to the 'cargoes', the molecule to be transported, at this domain. KIFs transport many different types of cargoes including membranous organelles, protein complexes, and messenger RNAs (mRNAs). They play important roles in a wide variety of intracellular transport, such as transport from Golgi to plasma membrane pathways involved in exocytosis and endocytosis, axonal transport, transport in dendrites, and special transport called intra-flagellar transport. In addition, recent molecular genetic experiments have uncovered unexpected roles for KIFs in the regulation of such physiologic processes as higher brain function, tumour suppression, and developmental patterning. They also play important roles in mitosis. Functions and/or dysfunctions of KIF motors underlie some diseases. The molecular mechanisms by which different KIFs recognize and bind to specific cargoes, and how their binding is regulated, have been identified for some KIFs.

Introduction

The study of axonal transport provided the first clue that there is a microtubule-dependent motor protein that is not related to dynein. It has been known for quite some time that various membranous organelles and proteins are transported within axon, but the rails and motors were not known. Experimental evidence then suggested that long-distance transport within the axon is microtubule dependent. Quick-freeze deep-etch electron microscopy revealed that short cross-bridges were present between membranous organelles and microtubules, and these were likely to be a candidate for motor proteins (Figure 1). In the mid-1980s, 'conventional kinesin (KIF5)' was first isolated biochemically as motor protein responsible for the transport of membranous organelles in the direction of cell body to synapse. Conventional kinesin has a rod-like structure composed of two globular heads (10 nm in diameter), a stalk, and a fanlike end, with a total length of 80 nm (Figure 2). The globular heads bind to microtubules and have an adenosine triphosphate (ATP)-binding sequence and a microtubule-binding sequence. A systematic molecular biological search of kinesin superfamily genes coding for proteins containing similar ATP-binding and microtubule-binding sequences led to the discovery of a group of motors, which is now called kinesin superfamily proteins, KIFs (Figures 2 and 3). Human genome contains 45 KIF genes. KIFs in the human

and mouse genomes are presented in a phylogenetic tree along with those in *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Caenorhabditis elegans* in **Figure 4**. KIFs identified in mouse are named as KIF1–KIF26, and KIFC1–KIFC3. KIFs identified in other species have various names. **Figure 4** also shows classes based mostly on murine and human KIFs, N-1 to N-11, M, C-1, and C-2. However, recently, a group of researchers in the field defined a new standardized nomenclature for classification of kinesins, based on 14 large families, Kinesin-1 to -14, to

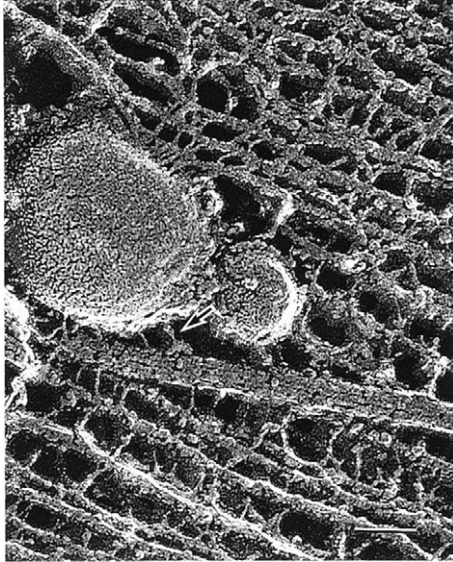


Figure 1 Quick-freeze deep-etch electron micrographs showing short cross-bridges between membranous organelles and microtubules that are structural candidates for KIFs. Scale = 50 nm. Reproduced from Hirokawa N (1998) Kinesin and dynein superfamily proteins and the mechanism of organelle transport. *Science* 279: 519–526, with permission.

facilitate understanding of evolutionary relatedness of genes identified in different names in various phylogenies (**Table 1**). To minimize confusion, it was decided that the names of individual motors will remain unchanged. The relationships of newly standardized family names and former classes are shown in **Table 1**. For clarity, the nomenclature in the mouse will be used as much as possible in this article. However, for some KIFs other names will be used.

Although all members of KIFs share a highly homologous domain having ~350 amino acids containing a site for ATP binding and one for microtubule binding, they are quite divergent outside the motor domain. This seems to allow the binding of KIF to various cargoes, including membrane-enclosed vesicles, macromolecular complexes, and messenger RNAs (mRNAs), and they play important roles in various transport processes including axonal transport, transport in dendrite, intracellular vesicular trafficking such as pathways involved in endocytosis and exocytosis, and special transport called ‘intraflagellar transport’ (**Table 1**). For most of the KIFs, the motor domain resides at the N-terminal portion of the molecule (N-kinesin), but the motor domain could be in the C-terminal (C-kinesin) or the middle of the molecule (M-kinesin) (**Figure 3**). The range of the size of a typical KIF molecule is from 600 to 1800 amino acids (**Figure 3**). Different KIFs take various forms (**Figure 2**). Some are homodimer with or without associated light chains; some are monomeric; some are heterodimer with associated protein; and some are homotetramer forming bipolar minifilaments. Some KIFs bind their cargo directly, and some do so via light chains or associated proteins, discussed individually in this article.

Microtubule has a polarity. It has a fast-growing plus-end, and opposite minus-end. In the nerve cell axons, the microtubules are unipolar, the plus-end pointing to the direction of synapse (**Figure 5(a)**). However, in dendrites microtubules have mixed polarity in proximal region, while they are oriented like in axon in the distal region. In interphase cells, microtubules are in a radial array with the minus-end pointing

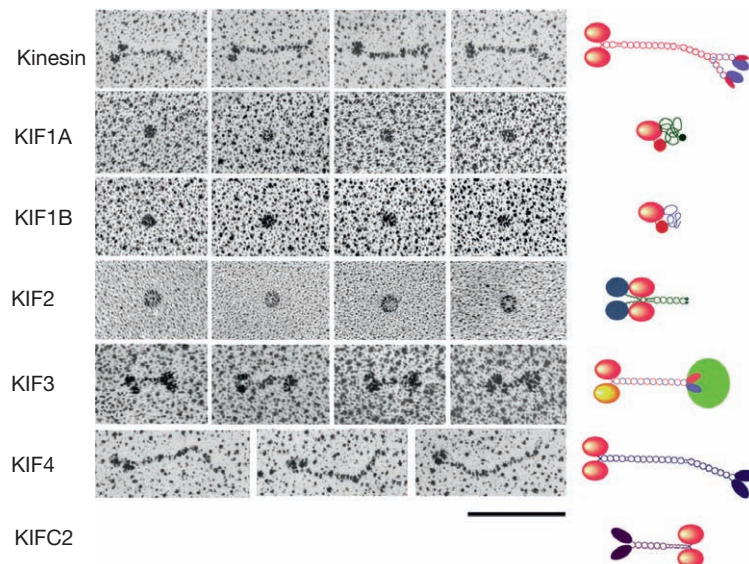


Figure 2 (Left) Panels showing main KIFs functioning in intracellular transport, as observed by low-angle rotary shadowing EM. Scale = 100 nm. (Right) Schematic illustration of the same KIFs based on EM studies or predicted from analysis of primary structures. Reproduced from Hirokawa N (1998) Kinesin and dynein superfamily proteins and the mechanism of organelle transport. *Science* 279: 519–526, with permission.

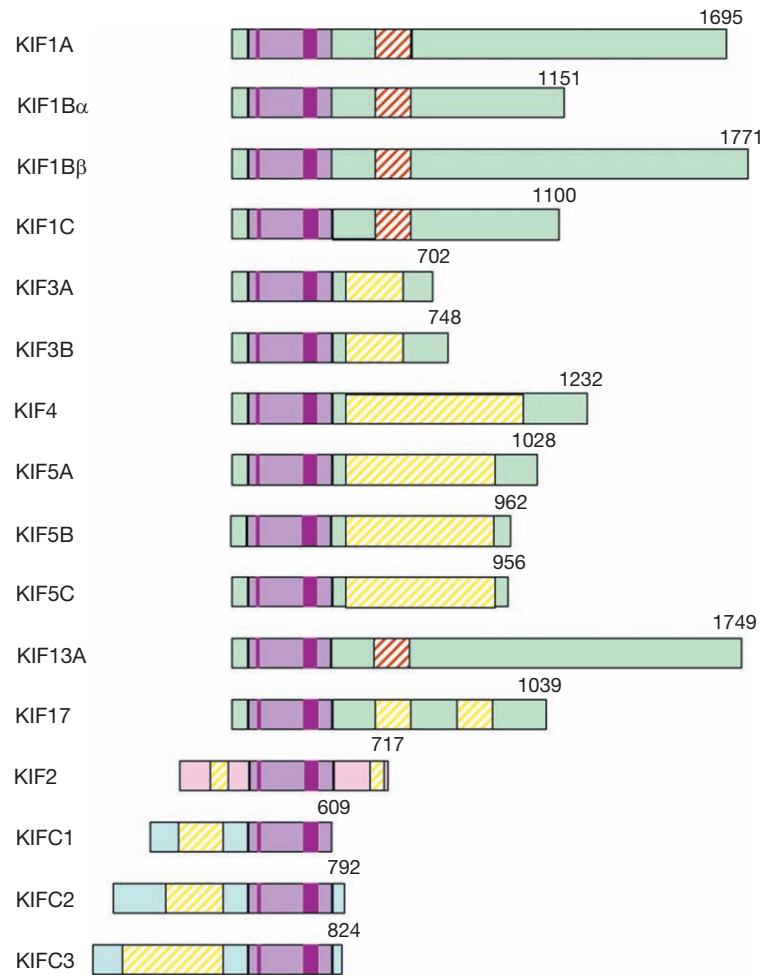


Figure 3 Structure of cDNAs from murine kinesin superfamily proteins (KIFs). Purple, motor domain; thin red line, ATP-binding consensus sequence; thick red line, microtubule-binding consensus sequence. Reproduced from Hirokawa N and Takemura R (2003) Kinesin superfamily proteins. In: Schliwa M (ed.) *Molecular Motors*, pp. 79–109. Weinheim: Wiley-VCH, with permission.

the cell center, and the plus-end pointing the cell periphery (Figure 5(b)). The direction of transport and velocity of each motor can be assessed in the so-called ‘*in vitro* motility assay’. Most of the KIFs are plus-end-directed motors, including conventional KIFs. Some KIFs are minus-end-directed motor, but the minus-end-directed transport is also carried out by cytoplasmic dynein (Figures 5(a) and 5(b)). In general, N-kinesin and M-kinesin are plus-end directed, and C-kinesin is minus-end directed. The typical velocity is within the range of $0.1\text{--}1.5\ \mu\text{m s}^{-1}$. In comparison, microtubule minus-end directed motors dyneins, which are one of the fastest motors, have a velocity of $\sim 14\ \mu\text{m s}^{-1}$. The direction obtained from *in vitro* assay generally agrees with that predicted from *in vivo* study. However, one may need to be somewhat cautious in interpreting the velocity of *in vitro* assay, because, for some KIFs, the condition of *in vitro* motility assay may not be optimal.

The characteristics of some KIFs, whose functions are relatively well characterized, are summarized in Table 1. Cargoes and physiological roles of some well-characterized KIF are described next.

N-Kinesin

N-kinesin holds by far the largest number of members and it consists of 12 classes, Kinesin-1 through Kinesin-12. Some classes have more than one family.

Kinesin-1 (N-1 Kinesins)

The KIF 5 family

Conventional kinesin corresponds to KIF5. KIF5 forms tetramers consisting of two kinesin heavy chains (120 kDa) and two light chains (64 kDa). Light chains are unrelated in sequence to kinesin. They are a rod-like structure composed of two globular heads, a stalk, and a fanlike end, in which the light chains reside (Figure 2). KIF5 is a plus-end-directed motor with a velocity of $\sim 0.5\ \mu\text{m s}^{-1}$. KIF5 is relatively abundantly expressed in both nervous tissue and other tissues, and transports an amazingly wide variety of cargoes including vesicles containing amyloid precursor protein (APP), GAP43, ApoER2, the receptor for the Reelin ligand, or the α -amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA)-type glutamate receptor GluR2,

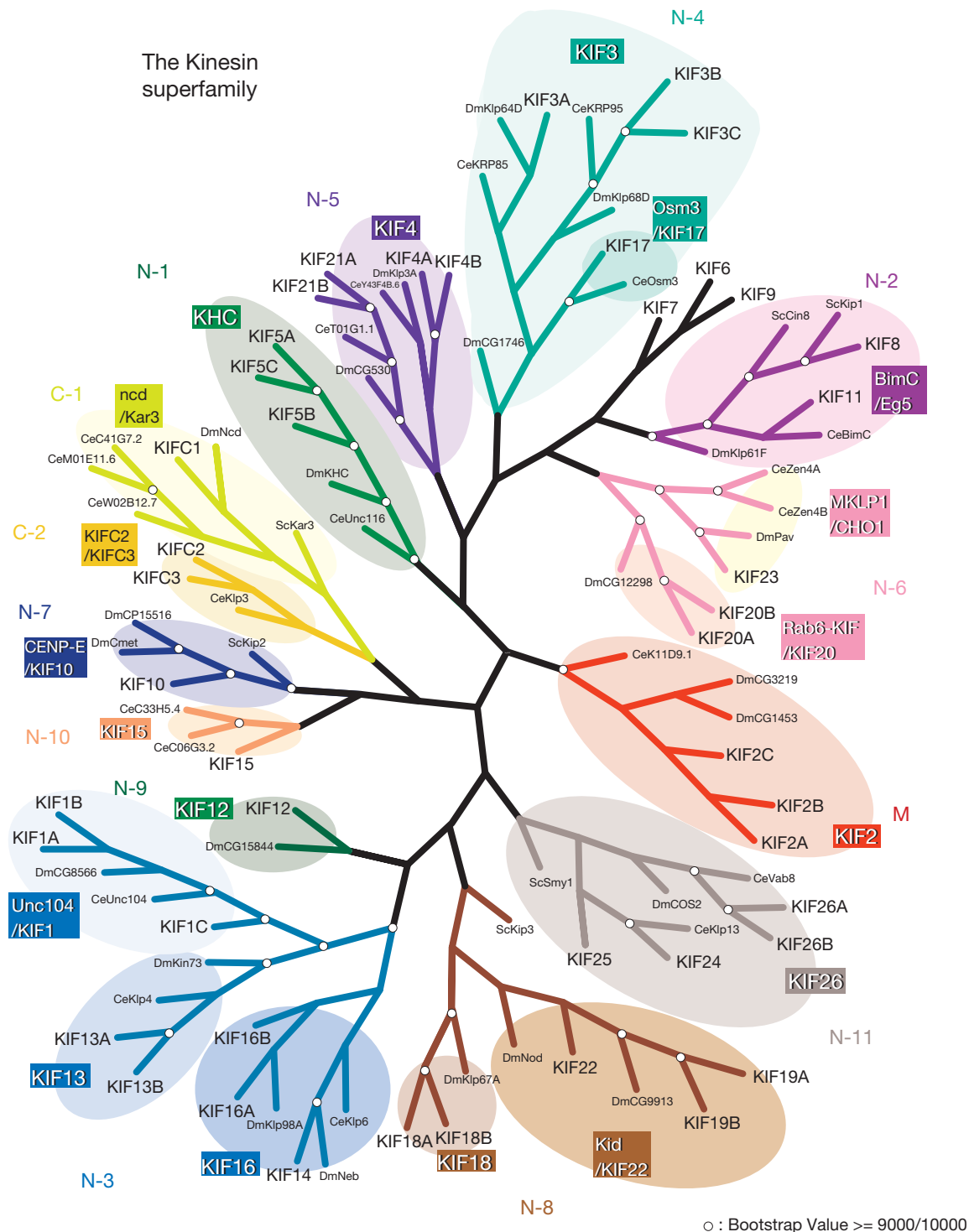


Figure 4 Phylogenetic analysis of all KIFs expressed in mouse/human, *D. melanogaster*, *C. elegans*, and *S. cerevisiae*. Amino acid sequences were aligned using maximum parsimony. Reproduced with permission from Miki H, Setou M, Kaneshiro K, and Hirokawa N (2001) All kinesin superfamily protein, KIF, genes in mouse and human. *Proceedings of the National Academy of Sciences of the United States of America* 98: 7004–7011, © 2001 National Academy of Sciences, USA.

mitochondria, and lysosomes (Figures 5(a) and 5(b)). KIF5 could also convey mRNA complexes, tubulin oligomers and intermediate filament proteins. Recently the direct interactions between KIF5 and some of the cargoes have been shown

(Figure 6). KIF5 may interact with certain cargoes via light chains, and yet with other cargoes via heavy chain. Moreover, whether KIF5 interacts via light chains or heavy chains may determine the destination. Glutamate receptor interacting

Table 1 Some members of KIF and their properties

Family name ^a	Class ^b	Name	Subunit structure	Direction of transport, velocity	Functions	Cargo
<i>N-kinesin</i>						
Kinesin-1	N-1	KIF5 (conventional kinesin)	Homodimer plus two light chains	Microtubule plus-end ~0.5 μm s ⁻¹	Axonal transport Dendritic transport Intracellular transport	Vesicles containing APP and GAP43, or ApoER2, tubulin oligomers Vesicles containing AMPA-type glutamate receptor Mitochondria, lysosomes, vimentin protein, mRNA complexes
Kinesin-2	N-4	KIF3	Heterodimer plus associated protein (KIF3A/3B-KAP3)	Microtubule plus-end ~0.3 μm s ⁻¹	Axonal transport Intracellular transport Intraflagellar transport (IFT) Dendritic transport	Vesicle containing fodrin, choline acetyltransferase Melanosomes IFT particles NMDA-type glutamate receptor
Kinesin-3	N-3	KIF17	Homodimer	Microtubule plus-end 0.7–1.2 μm s ⁻¹		
		KIF1A, KIF1Bβ	Monomer	Microtubule plus-end ~1.5 μm s ⁻¹	Axonal transport	Precursors of synaptic vesicles
		KIF1Bα	Monomer	Microtubule plus-end ~0.5 μm s ⁻¹	Axonal and intracellular transport	Mitochondria
Kinesin-5	N-2	KIF13A		Microtubule plus-end 0.1–0.3 μm s ⁻¹	Intracellular transport (endocytic, exocytic pathway)	Vesicles containing mannose-6-phosphate receptor
		<i>H. sapiens</i> Eg5, <i>C. elegans</i> BimC	Bipolar homotetramer	Microtubule plus-end	Mitosis	Microtubule (cross-linking and sliding of antiparallel interpolar microtubules at midzone, producing expanding force)
<i>M-kinesin</i>						
Kinesin-13	M	KIF2	Homodimer	Microtubule plus-end ~0.5 μm s ⁻¹	Axonal transport Intracellular transport Mitosis	Vesicles containing IGF-1 Lysosomes Microtubule (depolymerizing kinetochore microtubules at the kinetochore)
<i>C-kinesin</i>						
Kinesin-14	C-1	<i>Drosophila</i> Ncd		Microtubule minus-end	Mitosis	Microtubules (cross-linking and sliding of antiparallel interpolar microtubules at midzone, opposing the effect of N-2 kinesins)
	C-2	KIFC2 KIFC3	Homodimer	Microtubule minus-end ~4.5 μm s ⁻¹	Dendritic transport Intracellular transport	Multivesicular body-like membranous organelles Vesicles containing annexin XIIIb containing vesicles

Relationship of new family names and previous classes for other KIFs are Kinesin-4 (N-5), Kinesin-6 (N-6), Kinsin-7 (N-7), Kinesin-8 (N-8), Kinesin-9 to Kinesin-12 do not have counterparts in previous classification.

^aNew family names to be used for all phylogenies.

^bPrevious class names based mostly on murine and human KIFs.

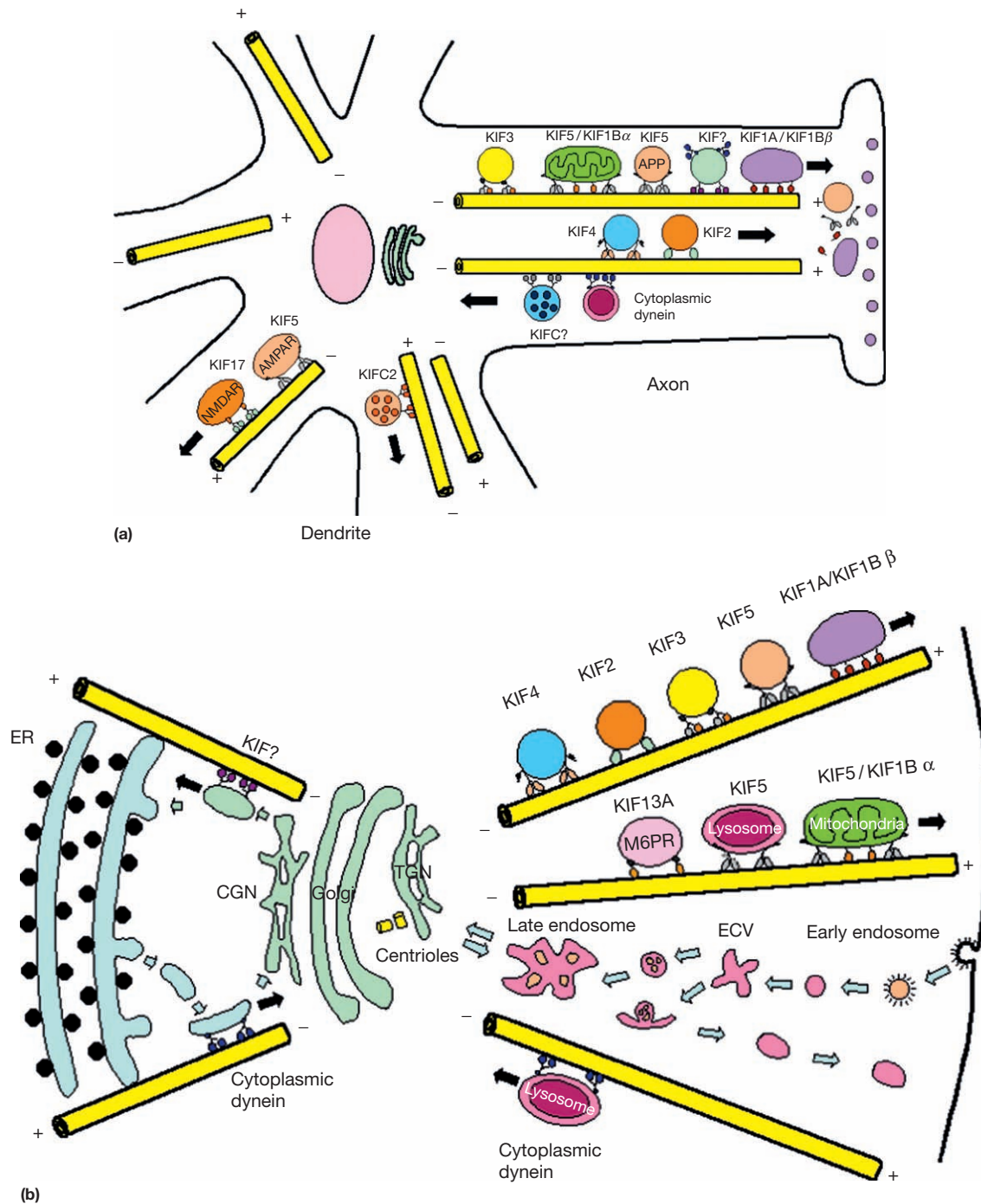


Figure 5 Scheme of KIFs and their cargo organelles in neurons (a) and in cells in general (b). In (b), neuron-specific KIFs and ubiquitous KIFs are illustrated in the same cell. Black arrows indicate the direction of transport. CGN, *cis*-Golgi network; TGN, *trans*-Golgi network; ECV, endosomal carrier vesicle. Reproduced from Hirokawa N and Takemura R (2003) Kinesin superfamily proteins. In: Schliwa M (ed.) *Molecular Motors*, pp. 79–109. Weinheim: Wiley-VCH, with permission.

protein 1 (GRIP1) binds to the tail domain of kinesin heavy chain and transports AMPA-type glutamate receptor (GluR2)-containing vesicles to dendrites, while JIP1,2 steers ApoER2-containing vesicles to axons through its interaction with light chains (Figure 6).

Kinesin-3 (N-3 Kinesins)

The *Unc104/KIF1* family

KIF1A and KIF1B β are plus-end-directed motors of ~190 kDa, which are relatively abundantly expressed in the axon and

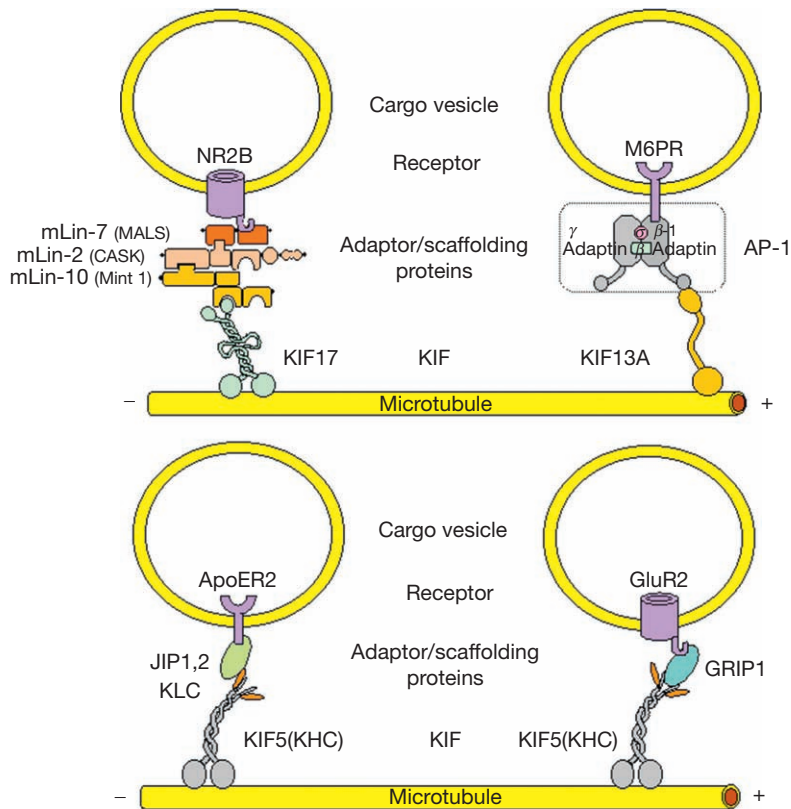


Figure 6 Scheme of how KIFs recognize and bind cargoes. Reproduced from Hirokawa N and Takemura R (2003) Kinesin superfamily proteins. In: Schliwa M (ed.) *Molecular Motors*, pp. 79–109. Weinheim: Wiley-VCH, with permission.

transport precursors of synaptic vesicles. They are monomeric motors, which is rather unique compared to other KIFs (Figure 2). They are also one of the fastest KIFs ($\sim 1.5 \mu\text{m s}^{-1}$) (Table 1). Unc104 is a homolog of mouse KIF1A. The knockout mice of KIF1A and KIF1B β show aberrant neuronal functions due to the defect in the transport of precursors of synaptic vesicles. Further study showed that patients with Charcot–Marie–Tooth disease type 2A carry a loss-of-function mutation in the motor domain of the *kif1b* gene. This was the first clear indication that impaired transport due to dysfunction of KIF motors underlies the cause of neurodegenerative disease. KIF1B α is a plus-end-directed motor ($\sim 0.5 \mu\text{m s}^{-1}$) expressed ubiquitously and transports mitochondria.

The KIF13 family

KIF13A is a plus-end-directed motor with the velocity of $0.1\text{--}0.3 \mu\text{m s}^{-1}$ and transports vesicles containing the mannose-6-phosphate receptor from the *trans*-Golgi network to the plasma membrane (Figure 5(b)). The interaction of KIF13A with the vesicle is mediated via direct interaction between KIF13A tail and β -1-adaptin, a subunit of the AP-1 adaptor complex (Figure 6).

Kinesin-2 (N-4 Kinesins)

The KIF3 family

The KIF3 forms a heterotrimeric complex, in which a heterodimer of KIF3A (80 kDa) and KIF3B (~ 95 kDa) interact at its

tail domain with a nonmotor protein, kinesin superfamily-associated protein 3 (KAP3, ~ 100 kDa) (Figure 2). The KIF3 complex is plus-end directed with a velocity of $\sim 0.3 \mu\text{m s}^{-1}$. KIF3 is relatively abundantly expressed in the axon and other tissues, and functions in axonal transport, intracellular transport, and in the intraflagellar transport (IFT). In the axon, the complex transports vesicles that carry fodrin and choline acetyltransferase, a soluble protein. In melanophore, it disperses pigment organelles called ‘melanosomes’. In cilia and flagella, the KIF3 complex transports macromolecular protein complexes, called ‘IFT particles’, from the bottom to the distal tip, which is important in the formation and maintenance of cilia and flagella. In retinal photoreceptor cells, the KIF3 complex is localized in the connecting cilium, which is located between the outer and inner segment. KIF3 complex participates in the transport of opsin from the inner to the outer segment at the connecting cilium; the impairment of this transport process could be one of the causes of retinitis pigmentosa.

An interesting aspect of KIF3 is its involvement in development. First, by its involvement in cilium formation, it plays a significant role in the determination of the left–right axis in early mouse development. During mouse development, there is a monocilium at the nodal cell that rotates to produce unidirectional leftward flow of extraembryonic fluid. This nodal flow could generate concentration gradients of putative secreted morphogen, which could switch on a gene cascade leading to left–right asymmetry. In *kif3A* (–/–) or *kif3B* (–/–) mouse, nodal cilia cannot be formed, nodal flow is absent, and the left–right axis formation is fundamentally impaired.

Second, KIF3 is involved in post-Golgi transport of N-cadherin/ β -catenin, which plays a role in adhesion of neuronal progenitor cells. In mice deficient in expression of KAP3 at post-midgestation stage, the subcellular localization of N-cadherin/ β -catenin was disrupted. β -Catenin tended to be localized in the nucleus to work as a transcriptional factor to enhance cell proliferation. The tumor-like abnormal hypertrophy of the cerebral cortex was observed suggesting that KIF3 may play roles in tumor suppression by transporting N-cadherin/ β -catenin from the Golgi to the plasma membrane.

The KIF17 family

KIF17 is localized to the dendrites and transports vesicles containing N-methyl-D-aspartate (NMDA)-type glutamate receptor 2B toward the microtubule plus end to the postsynaptic site, where the receptor plays an important role in synaptic plasticity (Figure 5(a)). The large protein complex containing mLin-10, mLin-2, mLin-7, and NR2B mediates the attachment of KIF17 to the vesicle (Figure 6). The tail domain of KIF17 interacts directly with the PDZ domain of mLin-10. Transgenic mice overexpressing KIF17 increase synthesis of KIF17 and NR2B and enhance spatial and working memories.

Other N-Kinesins

For the rest of the N-kinesin members, the physiological function has been elucidated to varying degrees. The members of Kinesin-5 (N-2 kinesins), *Homo sapiens* Eg5 or *Caenorhabditis elegans* BimC, form bipolar tetramer and participate in mitosis. They localize at the midzone of interpolar microtubules, where microtubules run in antiparallel orientation. They cross-link microtubules and allow their sliding against each other. Kinesin-6 (N-6 kinesins) is involved in both mitosis and vesicle transport. Kinesin-7 (N-7 kinesins) and Kinesin-8 (N-8 kinesins) are also involved primarily in mitosis.

KIFs are also involved in signaling transductions. The member of Kinesin-4 (N-5 kinesins), KIF4, regulates neuronal survival by binding to poly (ADP-ribose) polymerase-1 (PARP-1), a nuclear enzyme known to maintain cell homeostasis. The C-terminal domain of KIF4 is a module that suppresses the activity of PARP-1. When neurons are stimulated, KIF4 dissociates from PARP-1, resulting in upregulation of PARP-1 activity, which supports neuronal survival. Thus, KIF4 plays a significant role during brain development as a key molecule to control activity-dependent neuronal survival. The member of Kinesin-11 (N-11 kinesins), KIF26A, lacks ATPase activity, but suppresses glial cell-derived neurotrophic factor (GDNF) signaling pathway by binding to Grb2, an essential component of GDNF/Akt/ERK signaling. GDNF signaling regulates development of enteric nervous system. Mice with a homozygous deletion of KIF26A develop a megacolon with abnormal enteric nerve system. Kinesin-2 also includes the Costal 2, which is a part of the hedgehog-signaling cascade.

M-Kinesins

Kinesin-13 (M-Kinesins)

M-kinesins have a motor domain at the center of the molecule. There is only one family in this group, namely the KIF2 family,

which is composed of *kif2a*, *kif2b*, and *kif2c* genes. KIF2 family has functional roles in vesicle transport and mitosis. However, a unique feature of M-kinesins is regulation of microtubule dynamics, namely a microtubule-depolymerizing activity, which has been shown by X-ray crystallography to be due to the position of the motor domain within a molecule. *Xenopus* XKCM1 depolymerizes microtubules at the kinetochore and promotes chromosome segregation. In the brain of *kif2a* (–/–) mouse, neurons form abnormally elongated collateral branches. In contrast, in the wild-type mouse, shorter collaterals are formed. It was shown that in case of wild-type mouse, KIF2A suppresses microtubule polymerization at the tips of growth cones of developing neurons, which results in controlling the extension of collateral branches. This indicates that KIFs could play roles in brain wiring.

C-Kinesins

Kinesin-14A (C-1 Kinesins)

The KIFC1 family

Drosophila Ncd cross-links and slides antiparallel spindle microtubules in relation to another, pulling them apart, opposing the effect of bipolar plus-end-directed motors.

Kinesin-14B (C-2 Kinesins)

The KIFC2/KIFC3 family

KIFC2 transports multivesicular body-like membranous organelles in dendrites (Figure 5(a)). It has not been rigorously demonstrated, but it is presumed that KIFC2 acts as a minus-end-directed motor.

KIFC3 is a minus-end-directed microtubule motor protein that exists in kidney epithelial cells and other types of cells. In epithelial cells, microtubule organization is different from that in other types of cell; microtubule minus ends run toward the apical surfaces of the cell. KIFC3 accumulates on the apical surface of epithelial cells and transports vesicles associated with annexin XIIIb, a previously characterized membrane protein for apically transported vesicles.

Perspectives

Cargoes of KIFs: Specificity and Redundancy

As mentioned earlier, KIFs convey various types of cargoes. In some cases, a single KIF transports several distinct cargoes. How the same KIF can transport different cargoes is an interesting and important question that needs to be addressed.

At the same time, KIFs sometimes redundantly convey similar cargoes. KIF1B α transports mitochondria, while mitochondria have another redundant motor KIF5. KIF1A's cargo is a synaptic vesicle precursor, whereas KIF1B β also conveys synaptic vesicle precursors. Another example is lysosomes that have at least two microtubule plus-end-directed motors, KIF5 and KIF2A β . This redundancy could be a cause of subtle phenotypes of knockout mice in which some *kif* genes have been knocked out.

Recognition and Binding to Cargoes

To address the question of how specificity and redundancy are controlled, it is important to understand how KIFs recognize

and bind to their own cargoes. As discussed earlier, some of the interactions of KIFs and cargoes are beginning to be uncovered.

In terms of how KIFs recognize and bind cargoes, several forms can be categorized (Figure 6):

1. *KIF tail – scaffolding protein(s) or adaptor protein(s) – membrane protein.* This is applicable to KIF17-mLin-10-mLin-2-mLin-7-NR2B, KIF13A- β 1adaptin-AP-1adaptor complex-M6PR, KLC-JIP1/JIP2-ApoER2, the reelin receptor, and KIF5 (KHC)-GRIP1-GluR2.
2. *KIF tail – membrane protein.* Kinesin was reported to bind directly to membrane proteins through its interaction with KLC. Sunday Driver (also called JSAP1 and JIP3) and APP were proposed to be membrane proteins that bind directly to kinesin through KLC, while it needs to be studied further if Sunday Driver (JSAP1 and JIP3) is indeed a membrane protein.

It is only now being understood how KIFs recognize and bind to their specific cargoes, but thus far what was only known was that rather complex mechanisms are involved in these events. This may be related to how the association and dissociation between KIFs and cargoes are regulated. Another question is how KIFs recognize and bind to protein complexes such as tubulin and intermediate filament proteins, chromosomes, and mRNAs. These are obviously significant questions that need to be answered in the near future.

Regulation of Binding to Cargoes

For efficient cell function and morphogenesis, cargoes need to be delivered to their destinations by specific KIFs and unloaded from these KIFs at the appropriate time and place. Recent studies have started to reveal the regulatory mechanisms of the spatiotemporal delivery of cargoes. In some cases, the phosphorylation of KIFs seems to regulate their binding to cargoes. For example, phosphorylation of KIF17 by the Ca^{2+} /calmodulin-dependent protein kinase CaMKII disrupts the interaction of KIF17 and its interacting partner mLin-10. In other cases, members of Rab family of small guanosine triphosphatases seem to be involved in the regulation. For example, KIF1A and KIF1B β transport Rab3-carrying synaptic vesicle precursors through the adapter protein, mitogen-activated protein, kinase-activating death domain (MADD; also known as DENN/MADD). DENN/MADD recognizes the guanosine triphosphate-bound form of Rab3 on vesicles, but not the guanosine diphosphate-bound form.

How to Determine Directions of Transport

Another intriguing question is how a cell regulates direction of transports, for example, axon versus dendrites. In the case of KIF5, whether KIF5 binds to cargoes via heavy chain or light chain seems to determine the direction of transport, and JIP1.2

and GRIP1 are functioning as drivers (Figure 6). There are motors that transport cargoes mainly to dendrites. KIF17 conveys NMDA receptors mainly to dendrites and KIFC2 transports new multivesicular body-like organelles to dendrites. KIF21B was also proposed to be a motor for dendritic transport. How differential transports are performed is clearly the fundamental question that needs to be solved in the near future. The mechanism is probably not simple and could involve several distinct events.

Thus, cells use interestingly many KIFs and very precisely control the direction and velocity of transport of various distinct cargoes that are fundamental in cellular basic functions. This field is obviously very important and rapidly advancing. Further studies need to be carried out to understand the entire mechanism of intracellular transport.

See also: [Cell Architecture and Function: Dynein; Kinesins as Microtubule Disassembly Enzymes; Mitosis.](#)

Further Reading

- Alberts B, Johnson A, Lewis J, et al. (2008) *Molecular Biology of the Cell*, 5th edn. New York: Garland Science.
- Grafstein B and Forman DS (1980) Intracellular transport in neurons. *Physiological Reviews* 60: 1167–1283.
- Heald R (2000) Motor function in the mitotic spindle. *Cell* 102: 399–402.
- Hirokawa N (1998) Kinesin and dynein superfamily proteins and the mechanism of organelle transport. *Science* 279: 519–526.
- Hirokawa N (2000) Stirring up development with the heterotrimeric kinesin KIF3. *Traffic* 1: 29–34.
- Hirokawa N, Nitta R, and Okada Y (2009) The mechanisms of kinesin motor motility: Lessons from the monomeric motor KIF1A. *Nature Reviews Molecular Cell Biology* 10: 877–884.
- Hirokawa N and Noda Y (2008) Intracellular transport and kinesin superfamily proteins, KIFs: Structure, function, and dynamics. *Physiological Reviews* 88: 1089–1118.
- Hirokawa N, Noda Y, and Okada Y (1998) Kinesin and dynein superfamily proteins in organelle transport and cell division. *Current Opinion in Cell Biology* 10: 60–73.
- Hirokawa N, Noda Y, Tanaka Y, and Niwa S (2009) Kinesin superfamily motor proteins and intracellular transport. *Nature Reviews Molecular Cell Biology* 10: 682–696.
- Hirokawa N and Takemura R (2003) Kinesin superfamily proteins. In: Schliwa M (ed.) *Molecular Motors*, pp. 79–109. Weinheim: Wiley-VCH.
- Hirokawa N and Takemura R (2004) Biochemical and molecular characterization of diseases linked to motor proteins. *Trends in Biochemical Sciences* 28: 558–565.
- Hirokawa N and Takemura R (2005) Molecular motors and mechanisms of directional transport in neurons. *Nature Reviews Neuroscience* 6: 201–214.
- Lawrence CJ, Dawe RK, Christie KR, et al. (2004) A standardized kinesin nomenclature. *Journal of Cell Biology* 167: 19–22.
- Schliwa M (2003) *Molecular Motors*. Weinheim: Wiley-VCH.
- Tanaka Y and Hirokawa N (2002) Mouse models of Charcot-Marie-Tooth disease. *Trends in Genetics* 18: S39–S44.
- Terada S and Hirokawa N (2000) Moving on to the cargo problem of microtubule-dependent motors in neurons. *Current Opinion in Neurobiology* 10: 566–573.
- Verhey KJ and Rapoport TA (2001) Kinesin carries the signal. *Trends in Biochemical Sciences* 26: 545–550.

Kinetic Isotope Effects

J P Roth, Johns Hopkins University, Baltimore, MD, USA

J P Klinman, University of California, Berkeley, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Commitment to catalysis The rate constant for the conversion of ES to EP divided by the rate constant for the return of ES to E + S.

Intrinsic isotope effect The isotope effect when the bond cleavage or bond reorganization event completely limits the reaction rate.

Nuclear tunneling A process whereby an atomic nucleus, because of its wave-mechanical nature, is able to penetrate rather than surmount an energy barrier.

Zero-point energy The potential energy present in a bond at 0 K.

Isotopes are atoms that have identical numbers of protons and different numbers of neutrons in their nuclei. Because the electron distributions of molecules composed of isotopic atoms are identical, they exhibit identical chemical reactivity patterns. In reactions where bonds to isotopically labeled atoms are made and broken, rates are affected in proportion to the masses of the reacting bonds. Ratios of rates for such reactions are termed kinetic isotope effects (KIEs). KIEs are largest when bonds are made and broken at the site of isotopic substitution (a primary effect) but also are seen when significant reorganization occurs peripheral to a site that undergoes a change in bonding (a secondary effect). Both types of KIEs reflect the nature of the energy barrier to a fundamental chemical reaction. Primary hydrogen KIEs typically vary between two and eight, while the largest secondary hydrogen effects are <1.5. Thus, secondary KIEs often require the use of competitive methods and highly sensitive techniques to analyze product distributions, whereas primary KIEs can be detected using noncompetitive and less accurate techniques.

Physical Origin of KIEs

Primary KIEs have mostly been determined for reactions of carbon, nitrogen, and oxygen atoms bonded to the isotopes of hydrogen: protium (^1H), deuterium (^2H), and tritium (^3H). Such processes, which include proton, hydrogen atom, and hydride transfer, have been studied extensively in enzymes, solution phase and to a lesser extent in the gas phase. The observation of a primary KIE is conventionally understood using a semiclassical model. The model is semiclassical because it takes into account the quantization of energy in chemical bonds but does not explicitly deal with the dual wave/particle nature of the constituent atoms.

Chemical Bonds as Harmonic Oscillators

A chemical bond can be described as two massive bodies attached to a spring. The particles experience a restoring force that is proportional to their displacement (x) by the

force constant (F). Hooke's law (eqn [1]) relates the frequency of motion (ν) to the reduced mass (μ) of the atoms undergoing oscillation (eqn [2]) and the force constant characteristic of the bond. Force constants are the same for isotopic bonds and changes in vibration frequency result from changes in μ . The largest perturbation in ν is caused by changes in μ upon substituting deuterium or tritium for protium:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{F}{\mu}} \quad [1]$$

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad [2]$$

The oscillations of atoms from their equilibrium positions can be approximated by a harmonic function. The energy of a bound particle ($h\nu$) is defined by the harmonic-oscillator approximation according to eqn [3], where h is Planck's constant ($h = 6.626 \times 10^{-34} \text{ J s}$) and vibrational energy levels are designated n_{vib} . The zero-point energy ($n_{\text{vib}} = 0$) is the lowest vibrational energy level of a harmonic oscillator. These zero-point energies dominate the origin of KIEs within a semiclassical picture:

$$E_{\text{vib}} = (n_{\text{vib}} + 1/2)h\nu, \quad n_{\text{vib}} = 0, 1, 2, \dots \quad [3]$$

Wave/Particle Dualism

Particles also have wavelengths, with this property dependent on the mass (m) and velocity (v) of the particle. At constant energy, the wavelengths of unbound particles follow from the de Broglie equation (eqn [4]):

$$\lambda = h/mv \quad [4]$$

Defining the unbound particle's kinetic energy as $E_k = (\frac{1}{2}mv^2)$ allows de Broglie wavelengths to be calculated (Table 1). The arbitrarily assigned kinetic energies are characteristic of barrier heights encountered in many (C, N, or O)–H cleavage reactions. According to eqn [3], $E_k \sim 4 \text{ kcal mol}^{-1}$ approximates the potential energy present in (C, N or O)–H bonds ($\nu = 3000 \text{ cm}^{-1}$) at 0 K. When the electron is compared to the proton and then to an oxygen nucleus, there is an extremely

Table 1 de Broglie wavelengths as a function of mass and kinetic energy

Particle	e^-	^1H	^2H	^3H	^{12}C	^{16}O	^{80}Br
Mass ($\times 10^{27}$ kg)	0.000 911 0	1.661	3.321	4.982	19.93	26.57	132.8
λ (Å), $E_k = 4$ kcal mol $^{-1}$	29.40	0.69	0.49	0.40	0.20	0.17	0.077
λ (Å), $E_k = 10$ kcal mol $^{-1}$	18.60	0.44	0.31	0.25	0.13	0.11	0.049

large variation in the quantum wavelengths ranging from 29.40 Å to 0.69 Å to 0.17 Å, respectively. The values indicate that the de Broglie wavelengths for the electron and hydrogen are comparable to the distances traversed by these particles in chemical reactions, consistent with experimental findings that indicate a dominance of their reactivity by quantum mechanical behavior.

KIEs as Mechanistic Probes

KIE measurements are routinely used for the elucidation of reaction mechanisms. The reaction coordinate represents the three-dimensional motion of all nuclei along the lowest-energy path that interconverts reactants and products. The reaction coordinate may represent the displacement of atoms within a chemical bond and/or the reorientation of surrounding solvent molecules. Regardless of the nature of the reaction coordinate, isotope substitution does not change the electronic potential energy surface for the reaction. Reaction rates vary with isotopic perturbation due to changes in vibrations that directly or indirectly couple to the reaction coordinate and by a decrease in wavelength for heavier atoms.

Activation Energy and Transition State Theory

In the year 1889, Svante Arrhenius documented the observation that chemical reaction rates vary as a function of temperature, predating the discovery of the first isotope, deuterium, in 1932. The Arrhenius equation (eqn [5]) expresses a rate constant (k) as an exponential function of thermal activation energy (E_{act}). Together, the temperature (T) and the gas constant (R) define the population of states that, as the result of molecular collisions, have sufficient E_{act} for reaction. The A term, originally referred to as a frequency factor, reflects the rate at which an activated complex (one possessing E_{act}) decomposes to product:

$$k = A \exp(-E_{\text{act}}/RT) \quad [5]$$

The transition-state theory (TST) developed separately by Eyring and by Evans and Polanyi (c. 1935) is the preeminent model for interpretation of energy barriers to chemical reactions. TST gives the rate of reaction in terms of the molecular-level properties of an activated complex or transition state. Unlike the phenomenological Arrhenius equation, TST defines the energy difference between reactant and transition state using translational, rotational, vibrational, and electronic (gas phase) partition functions. The reaction coordinate is treated classically and separated from all of the other classical and

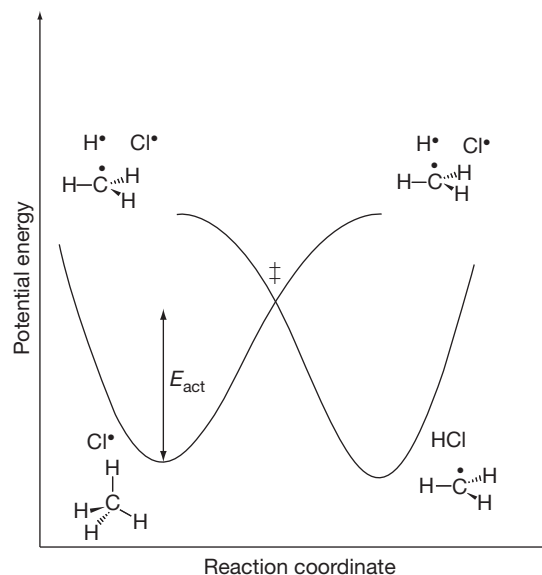


Figure 1 Reaction coordinates for $\text{H}^\bullet$ abstraction from methane by $\text{Cl}^\bullet$ depicting the transition state as the intersection of potential energy surfaces for stretching C–H and H–Cl bonds.

quantum modes. The transition state occupies the dividing surface separating reactants from products and contains all degrees of freedom orthogonal to the reaction coordinate. The fundamental assumption of classical TST is that reactants do not re-cross the dividing surface and products are formed with the frequency of a bond vibration or solvent fluctuation once the transition-state configuration is attained. This assumption is relaxed in the semi-classical expression given by eqn [6], where the rate constant is expressed in terms of an equilibrium constant for the formation of the transition state from the reactant state $\{K^\ddagger = \exp(-\Delta G^\ddagger/RT)\}$ and a frequency factor ($\kappa(k_B T/h)$). In the pre-exponential term, k_B is Boltzmann's constant and κ appears as a correction factor for the nonunit probability that products will form each time the reactants attain the energy of the transition state:

$$k = \kappa \frac{k_B T}{h} \exp(-\Delta G^\ddagger/RT) \quad [6]$$

Figure 1 shows a curve-crossing diagram for a molecule of methane reacting with atomic chlorine. The transition state is denoted by 'double dagger' and represents a configuration for H atom transfer where the C–H bond is stretched and there is an attractive force between the Cl and H atoms. Thus, the reaction coordinate is the stretching of the bond. This situation is depicted further in Figure 2, where reaction occurs when an energy barrier is surmounted that converts the stretching

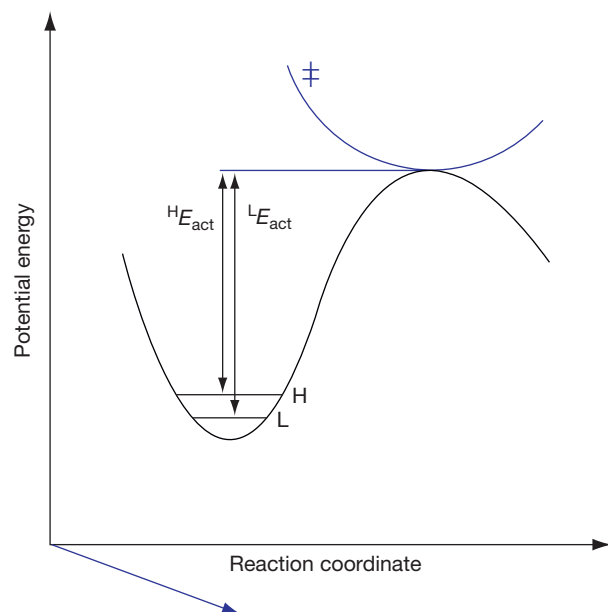


Figure 2 Energy diagram representing the origin of KIE on cleaving a C-H or C-L bond ($H = {}^1\text{H}$, $L = {}^2\text{H}$ or ${}^3\text{H}$). In this figure, a stretching mode in the reactant has been converted to a translational mode in the transition state.

motion of a C-H (or C-L, where $L = {}^2\text{H}$ or ${}^3\text{H}$) bond to a translational degree of freedom.

Although the nature of the reaction coordinate is not specified in the original derivation of TST, diagrams akin to **Figure 2** have been used extensively for pedagogical reasons. The transition state is not an intermediate and has no detectable lifetime. The curvature of the reaction coordinate in the transition state region is defined by an imaginary frequency for the C- ${}^1\text{H}$ or C-L vibration that is lost. Clearly, this formulation of a bond-stretching reaction coordinate is an oversimplification when frequencies are high. When $h\nu \gg k_B T$, the reacting bonds are essentially frozen and should not contribute much to the activation energy for reaction.

Isotope Effects from Zero-Point Energies

According to the Born–Oppenheimer approximation, the mass disparity between an electron and a nucleus allows the timescale for electronic motion to be separated from the much slower nuclear motion. It follows that reduced mass differences cause the zero-point energies to follow the trend $\text{C}-{}^1\text{H} > \text{C}-{}^2\text{H} > \text{C}-{}^3\text{H}$, as shown in **Figure 2**. Neglecting differences for reactions of isotopic molecules due to their quantum wavelengths, KIEs can be calculated from activation energies.

In the simplest case, the reaction coordinate involves conversion of a ground-state stretch to a translation and, hence, neglects isotopic zero-point energies in the transition state. The differences in the zero-point levels of the reactant state give the upper limit to the primary KIE through **eqn [7]**, where the subscript L designates either of the heavier isotopes of hydrogen. From the stretching frequencies of C- ${}^1\text{H}$

($\nu = 3000\text{ cm}^{-1}$), C- ${}^2\text{H}$ ($\nu = 2200\text{ cm}^{-1}$), and C- ${}^3\text{H}$ bonds ($\nu = 1800\text{ cm}^{-1}$), KIEs of 7 and 18 are estimated for $k_{1\text{H}}/k_{2\text{H}}$ and $k_{1\text{H}}/k_{3\text{H}}$ at 298 K, respectively:

$$\frac{k_{1\text{H}}}{k_L} = \exp\left[\frac{h(\nu_H) - h(\nu_L)}{2RT}\right] \quad [7]$$

The Melander–Westheimer principle (c. 1960) models the parabolic variation of primary KIEs in terms of a linear transition state for hydrogen transfer (cf. **Figure 1**). The symmetry of the transition state depends on the overall reaction thermodynamics. Asymmetry increases as the transition state becomes more reactant like (or earlier along the reaction coordinate) for exothermic reactions or more product like (later) in an endothermic reaction. For real molecules that contain multiple degrees of vibrational freedom, compensating vibrations that are isotopically sensitive will reduce the size of the isotope effect from its peak at a symmetric transition state.

Deviations from Predicted Values

In general, observed primary KIEs are less than the upper limit set by **eqn [7]**. Observations of values larger than those predicted by **eqn [7]** have been observed and tentatively attributed to an insufficient description of the reaction coordinate. For instance, KIEs may appear inflated when isotope labels in secondary positions contribute to differences in E_{act} . As discussed in the monograph by Melander and Saunders, KIEs may be up to threefold larger because of restricted transition-state geometries wherein bending vibrations present in the reactant state have been lost.

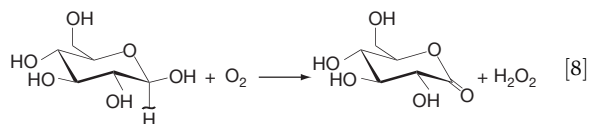
In contrast to effects ascribable to the ratio of reactant and transition-state partition functions, non-Boltzmann effects such as nuclear tunneling may also give rise to large isotope effects. Tunneling has historically been treated as a correction, with the dominant contribution to the rate constant viewed as the activation energy associated with distortion of the reacting bond. However, there has been a resurgence of attention to quantum phenomena in the chemical and biochemical literature. New models for proton, hydrogen atom, and hydride transfer reactions are being developed which presuppose tunneling mechanisms and formulate reaction coordinates in terms of the collective low-frequency motions of heavy atoms and the surrounding solvent molecules or protein. It is important to realize that a wide range of KIEs with values both larger and smaller than those predicted by **eqn [7]** are consistent with mechanisms of tunneling, where reactants do not need to have the activation energy required to reach the transition state and reaction may occur by passing through rather than over the potential energy barrier.

KIEs as Mechanistic Probes during Enzyme Catalysis

By virtue of being a catalyst, an enzyme accelerates a chemical reaction without changing the overall thermodynamics. The term catalysis frequently refers to a multistep reaction such as the oxidation of glucose to gluconolactone and reduction of O_2 to H_2O_2 as it occurs in the glucose oxidase reaction (**eqn**

[8]). Additionally, the term catalysis may refer specifically to a single chemical step that is accelerated by lowering or circumventing an energy barrier.

To understand the role of an enzyme during a catalytic transformation, the substrate-binding step and chemical step, where reactants are converted to products, must be analyzed separately. Assessment of barrier-lowering effects requires comparison of the enzymic reaction to a solution-phase reaction, both referenced to some standard state:



Experimental Considerations

Primary, secondary, and solvent KIEs can be used to elucidate the extent to which a change in bonding occurs in or before the rate-determining step. Primary effects are invariably normal, that is, the light isotope reacts faster than the heavy one. Secondary effects may show inverse behavior when the reaction is characterized by an increase in bonding and the difference in zero-point energies in the transition state is greater than in the reactant state. Primary effects are typically measured for reactions where protium is exchanged for deuterium and/or tritium, but isotope labeling of the heavy atoms may also give rise to measurable effects.

The largest effects are predicted upon isotopic substitution of protium because of the sizable perturbation to the zero-point energy levels as discussed earlier. Because reactions involving the transfer of protium and its isotopes give rise to the largest effects, experiments performed noncompetitively, using the pure labeled substrates in separate reaction solutions, require high isotopic enrichment in addition to chemical purity. Isotopic purity becomes crucial when KIEs are >10 and only 5% contamination of the heavy isotope by the light isotope will cause an apparent deflation of the KIE by a factor of >1.4 .

Secondary KIEs are typically much smaller and are, therefore, most often measured by competitive techniques. While isotope purity is not required, sensitive methods such as scintillation counting of radiolabeled products or mass-spectrometric analysis of chemically labeled products are necessary. Intra-molecular measurements can be more informative than inter-molecular measurements, as spurious isotope discrimination due to differential binding effects is absent. However, the determination of intra-molecular primary isotope effects on enzymatic reactions is, most generally, precluded by the inherent stereoselectivity imparted by the protein.

An Example from Glucose Oxidase Catalysis

Glucose oxidase mediates the oxidation of glucose and other simple six-carbon sugars. Enzymes isolated from fungi have been the subjects of extensive mechanistic studies due in large part to their robustness under a variety of temperature and pH

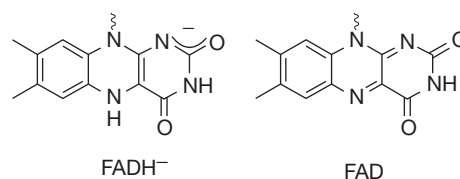


Figure 3 The isoalloxazine rings of oxidized (FAD) and reduced (FADH⁻) flavin adenine dinucleotide.

conditions. The active enzyme is a homodimer that contains one noncovalently bound flavin adenine dinucleotide cofactor (Figure 3) per subunit. The flavin adenine dinucleotide (FAD) is reduced by simple sugars in the reductive half-reaction to produce FADH⁻, which is oxidized by molecular O₂ in the oxidative half-reaction. A ping-pong-type kinetic mechanism has been demonstrated for glucose oxidase wherein the first substrate binds, gets converted to product, and is then released prior to the reaction of the second substrate (Figure 4).

The mechanisms advanced for the reactions of sugars with glucose oxidase involve cleavage of the anomeric C–H in the rate-determining step (eqn [8]). The main evidence is the significant primary KIEs observed upon isotope labeling of this position in glucose and in the alternative, slower-reacting substrate 2-deoxyglucose. Under steady-state conditions at pH 5 and 298 K, the second-order rate for the reaction of substrate (k_{cat}/K_m) is 3–4 times faster for [1-¹H]-glucose than for [1-²H]-glucose (Table 2). The k_{cat}/K_m values for [1-¹H] and [1-²H] 2-deoxyglucose are 860 and 120 M⁻¹ s⁻¹, respectively, revealing a primary KIE of 7. The rates are an order of magnitude slower than for the glucose [1-¹H] and [1-²H] for which k_{cat}/K_m is 8200 and 2600 M⁻¹ s⁻¹. The difference in the KIEs shows how C–H cleavage in the slower substrate has a greater contribution than substrate binding to its second-order rate constant. The reasoning is consistent with the observation that the unimolecular rate constant that controls turnover (k_{cat}) is more than order of magnitude faster for glucose (1200 s⁻¹) than for 2-deoxyglucose (50 s⁻¹) under the same conditions.

The degree to which bond cleavage limits the rate of sugar oxidation may vary with pH and with temperature depending on relative heights of energy barriers along the potential energy landscape. For oxidation of 2-deoxyglucose by glucose oxidase, the commitment to catalysis defined by Northrop as the ratio of the product-forming rate constant to the rate constant for substrate dissociation appears to be low and independent of pH. At pH 5, noncompetitive measurements with pure [1-¹H] and [1-²H] 2-deoxyglucose have allowed a direct comparison of k_{cat} to k_{cat}/K_m effects; these show that the isotope effect on k_{cat}/K_m approaches the intrinsic isotope effect on k_{cat} (¹H vs. ²H) 8. Thus, the rate-determining step in both the first- and second-order kinetic processes includes the cleavage of the bond to the anomeric carbon (the E·S to E·P step in Figure 4). Similar results at pH 9.0, where [1-¹H] 2-deoxyglucose reacts with $k_{\text{cat}}/K_m = 580 \text{ M}^{-1} \text{ s}^{-1}$ and exhibits ¹H versus ²H KIEs on both k_{cat}/K_m and k_{cat} of 7, indicate that bond cleavage represents the highest point on the energy landscape (Figure 4). This result has been confirmed by highly accurate KIEs measured using radiolabeled substrates under competitive conditions. Here the KIEs correspond to differential reactivity

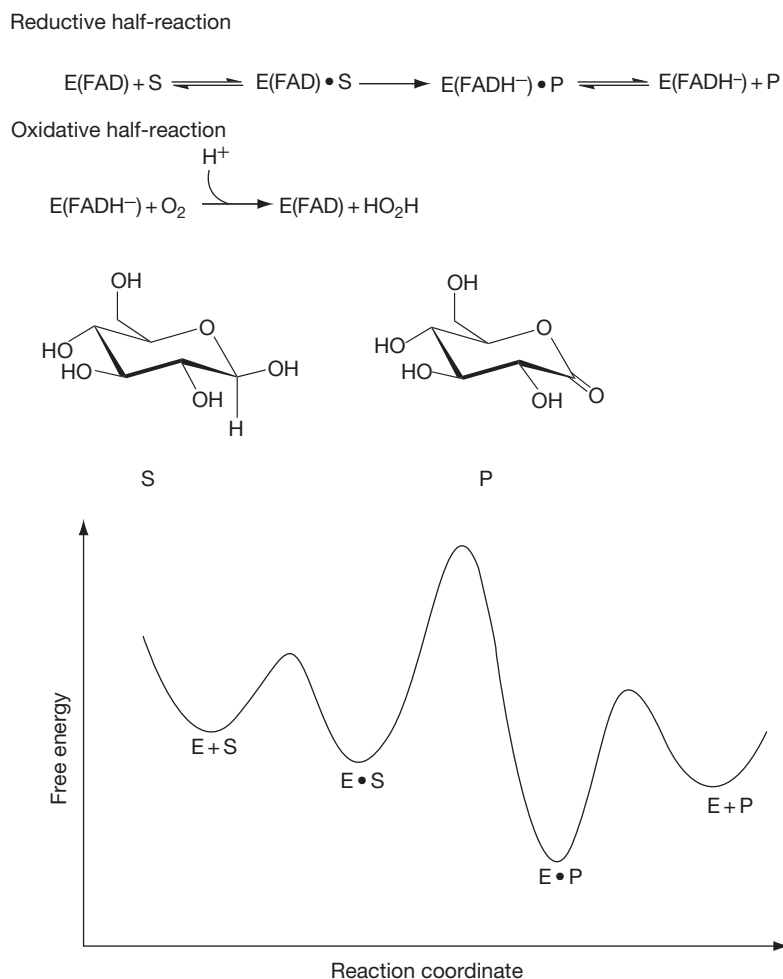


Figure 4 Ping-pong kinetic scheme and energy landscape for the reductive half-reaction of glucose oxidase. The E•S and E•P are arbitrarily defined as being more stable than separated reactants and products.

Table 2 Rate constants and isotope effects under air saturation at 298 K

Substrate	$k_{\text{cat}}/K_{\text{m}} (M^{-1} s^{-1})$	$^L (k_{\text{cat}}/K_{\text{m}})$	$k_{\text{cat}} (s^{-1})$	$^L (k_{\text{cat}})$
[1- ^1H] glucose (pH 5)	$8.2 (1.9) \times 10^3$	3–4	$1.2 (0.3) \times 10^3$	~6
[1- ^2H] glucose (pH 5)	$2.6 (0.6) \times 10^3$		$2.0 (0.5) \times 10^2$	
[1- ^1H] 2-deoxyglucose (pH 5)	$8.6 (0.8) \times 10^2$	~7	$4.4 (1.2) \times 10^1$	~8
[1- ^2H] 2-deoxyglucose (pH 5)	$1.2 (0.4) \times 10^2$		5.7 (0.1)	
[1- ^1H] 2-deoxyglucose (pH 9)	$5.8 (0.2) \times 10^2$	6.6	$2.6 (1.0) \times 10^1$	~7
[1- ^2H] 2-deoxyglucose (pH 9)	$8.8 (0.2) \times 10^1$		3.5 (0.6)	
[1- ^3H] 2-deoxyglucose (pH 9)	$3.6 (0.2) \times 10^1$	16	N/A	

L represents ^2H or ^3H .

of isotopically labeled substrates in the same reaction solution. Isotope effects on $k_{\text{cat}}/K_{\text{m}}$ for 2-deoxyglucose at pH 9.0 and 298 K reveal that the ^3H substrate reacts 15.9 (0.5) times slower than the ^1H substrate and 2.44 (0.04) times slower than the ^2H substrate. From the above data, the calculated ^1H versus ^2H isotope effect on $k_{\text{cat}}/K_{\text{m}}$ is found to be 6.5 (0.5) close to the value derived from the noncompetitive measurements. Based on observed KIEs under a variety of conditions, it can be concluded that the observed k_{cat} reflects the free-energy barrier associated with the chemical step wherein sugar is transformed to lactone.

The kinetics of FADH^- oxidation by O_2 has also been analyzed using isotope effects. Solvent KIEs and heavy-atom KIEs on $k_{\text{cat}}/K_{\text{m}}$ for O_2 have been determined for high- and low-pH forms of glucose oxidase. The latter measurements were determined competitively by using isotope ratio mass spectrometry to follow the enrichment of ^{18}O in natural abundance O_2 as a function of reaction. The ratio of $k_{\text{cat}}/K_{\text{m}}$ for $^{16,16}\text{O}_2$ to $^{16,18}\text{O}_2$ ranges from 1.027(3) to 1.028(4) for the two enzyme forms. The KIE is close to the equilibrium isotope effect estimated using the Bigeleisen approach, which employs

gas-phase partition functions for O_2 and superoxide anion, the reactant and product of the rate-determining step. Solvent isotope effects accompanying O_2 reduction were found to be close to unity, showing that electron transfer, and not proton transfer, limits $k_{cat}/K_m(O_2)$. In addition, the observed isotope effects aid a direct comparison of rate constant for the enzymatic reduction of O_2 and the nonenzymatic reaction which in turn allows the extent of barrier lowering (catalysis) in the chemical step to be addressed.

Various types of KIEs have been used to elucidate the chemical mechanisms of glucose oxidase catalysis. The observed KIEs indicate that the rate of the chemical step can be extracted from the steady-state measurements using the slowly reacting substrate 2-deoxyglucose. The observed KIEs demonstrate rate-determining C–H cleavage in the reductive half-reaction and a rate-determining change in oxygen bond order upon electron transfer in the oxidative half-reaction. In addition to exposing which steps limit catalysis, KIE probes have been invaluable in demonstrating the role of the protein environment in facilitating catalysis.

See also: [Protein/Enzyme Structure Function and Degradation: Flavins; Substrate Binding, Catalysis, and Product Release.](#)

Further Reading

- Bell RP (1980) *The Tunnel Effect in Chemistry*. New York, NY: Chapman and Hall.
- Bright HJ and Gibson QH (1967) The oxidation of 1-deuterated glucose by glucose oxidase. *Journal of Biological Chemistry* 242: 994–1003.
- Cook PF (ed.) (1991) *Enzyme Mechanism from Isotope Effects*. Boca Raton, FL: CRC Press.
- McQuarrie DE (1983) *Quantum Chemistry*. Mill Valley, CA: University Science Books.
- Melander L and Saunders WH, Jr. (1980) *Reaction Rates of Isotopic Molecules*. New York, NY: Wiley-Interscience.
- Northrop DB (1981) The expression of isotope effects on enzyme-catalyzed reactions. *Annual Review of Biochemistry* 50: 103–131.
- Roth JP and Klinman JP (2003) Catalysis of electron transfer during O_2 activation by the flavoprotein glucose oxidase. *Proceedings of the National Academy of Sciences of the United States of America* 100: 62–67.
- Steinfeldt JI, Francisco JS, and Hase WL (1989) *Chemical Kinetics and Dynamics*. Englewood Cliffs, NJ: Prentice-Hall.



***lac* Operon**

L Swint-Kruse, University of Kansas Medical Center, Kansas City, KS, USA

K S Matthews, Rice University, Houston, TX, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Activation Process that enhances transcription to boost cellular levels of regulated proteins.

Conformational change A significant rearrangement of atoms within a protein molecule in response to binding a specific small molecule (e.g., sugar binding by LacI).

Genome The entire DNA complement composing the genetic material of an organism.

Inducer A sugar that elicits LacI conformational rearrangement that releases operator DNA.

Induction The increase in *lac* enzyme production triggered by the presence of an inducer sugar.

Messenger RNA (mRNA) A copy of the DNA sequence that is used as a template for protein translation.

Operator The DNA sequence to which the repressor protein binds and thus inhibits RNA polymerase transcription.

Operon A region of bacterial DNA that is regulated as a unit, for example, the entire region encoding the *lac* repressor, metabolic enzymes, and regulatory regions.

Promoter The DNA sequence recognized by RNA polymerase for transcription initiation; in bacteria this region frequently contains a binding site for the catabolite repressor protein (CRP).

Repression The inhibition of transcription to maintain low levels of metabolic proteins/enzymes in the absence of appropriate conditions, such as lack of substrate for a metabolic pathway.

Repressor The protein that mediates inhibition of transcription by binding to a specific DNA sequence (see 'operator').

Transcription The act of copying the information encoded in DNA into a new mRNA molecule as a template for translation of the amino acid sequence of a protein.

Living organisms must sense a wide range of exogenous and endogenous stimuli in order to match gene expression to environmental state. From an energetic perspective, this process is most efficient at the point of DNA transcription into messenger RNA. The *lac* operon was the first system for which this mechanism was detailed and paved the way to discovering a variety of strategies for modulating RNA transcription. Regulation of *lac* operon expression allows bacteria to respond to changes in the availability of external lactose (milk sugar) and glucose concentrations by shifting production of enzymes used to metabolize lactose. In this manner, the bacteria are able to effectively utilize lactose when available and yet minimize energetic cost of producing metabolic proteins when lactose is unavailable. This article explores the historical context for the discovery of the prototypic lactose regulatory system and the structural features of the component proteins.

Historical Perspective

During the post-World War II years, at the Pasteur Institute in France, Jacques Monod and Francois Jacob were examining the

ability of bacteria to adapt to different sources of carbon as their energy source. They discovered that, when lactose (milk sugar) was absent from the growth medium, the bacterial enzymes that metabolize lactose were present at very low levels. When lactose was added, enzyme production increased ~1000-fold, a phenomenon termed 'induction'. (The three proteins encoded by the *lac* operon that import and metabolize lactose have historically been called 'structural' proteins and their coding regions 'structural genes'; however, we use the term 'metabolic' for these proteins and genes to distinguish them from later structural studies of the *lac* repressor and CRP proteins.)

Next, Jacob and Monod varied the molecular structure of lactose and the chemically related galactoside sugars and compared the consequent effects on growth and activity of *lac* metabolic proteins for *Escherichia coli* variants. They found that a subset of these sugars could induce but not be metabolized, and a different set could be metabolized but not serve as inducer. Thus, the function of inducing the *lac* metabolic proteins can be separated from the transport and metabolism of lactose. In the early 1960s, Jacob and Monod found that an inducer sugar functioned by relieving inhibition of

transcription for the metabolic genes. The inducible agent was finally identified as the lactose repressor protein, LacI. When the LacI protein was inactivated in the bacteria, the metabolic enzymes were always expressed at high levels, whether or not an inducer sugar was present. From these experiments, Jacob and Monod demonstrated that expression of RNA encoding an enzyme (or other protein) can be negatively controlled. This phenomenon of inhibiting transcription is termed 'repression'.

Further experimental observations indicated that, by itself, the presence of lactose does not always fully release repression for the *lac* operon. An additional factor contributing to the repression/induction balance was found to be glucose concentration. When glucose is abundantly available, full induction does not occur even in the presence of high lactose levels. From the cellular perspective, direct glucose utilization provides a more efficient energy source than converting lactose to glucose and galactose (and then galactose to glucose). Thus, in addition to repression by LacI, another regulatory mechanism must contribute to determining the level of *lac* enzyme expression. Two different mutations were discovered in the late 1960s/early 1970s that illuminated this issue, and a cell-free system was used to

demonstrate that one of these mutations disrupted the catabolite repressor protein (CRP) (cAMP receptor protein, also referred to as the catabolite activator protein, CAP). CRP binds the signaling molecule cAMP, which is, in turn, produced by the second protein identified in these mutagenesis experiments, adenylate cyclase. Adenylate cyclase activity is inhibited by high glucose concentrations. Together, CRP and adenylate cyclase provide a means to alter transcription by sensing glucose levels via detection of cellular cAMP concentrations. In contrast to the high specificity of the *lac* repressor (which only regulates the *lac* operon), the CRP/adenylate cyclase glucose sensor activates transcription of numerous systems in *E. coli* in its cAMP-bound form.

lac Operon

The region of the *E. coli* genome that is involved in lactose metabolism and control of *lac* gene expression is termed the *lac* operon. This region of DNA comprises sequences that contain several types of information – open reading frames

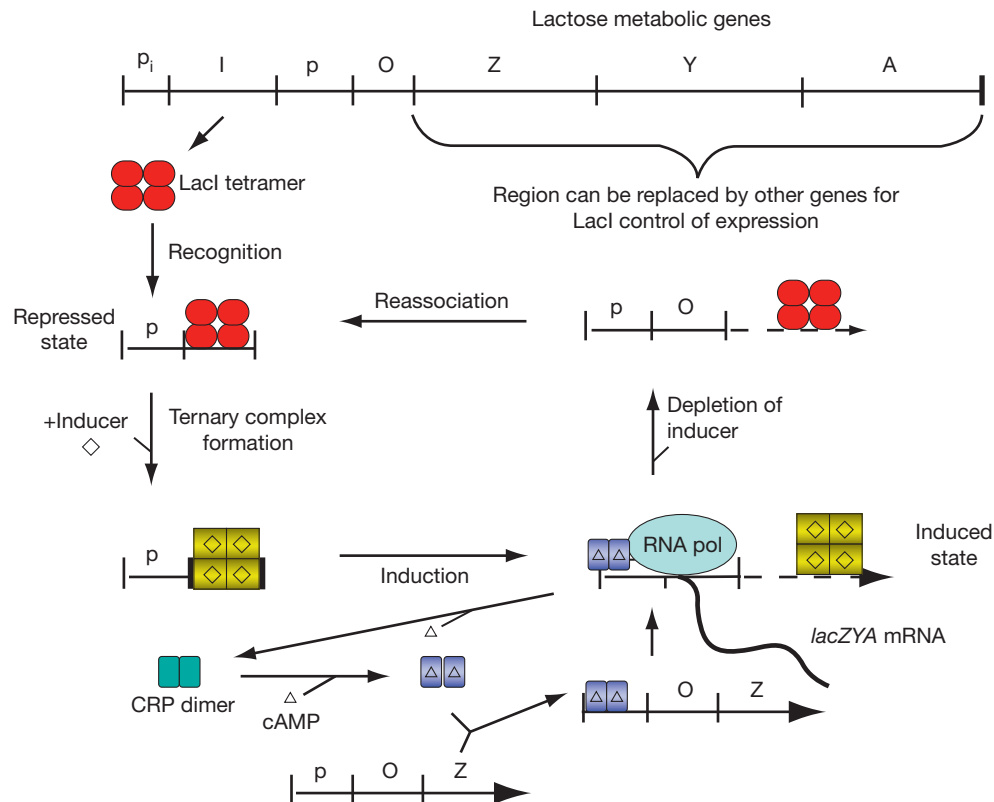


Figure 1 Schematic of the *lac* operon. The top horizontal bar depicts regions of the *lac* operon DNA: the promoter (P_i) for the tetrameric lactose repressor (*lacI* gene, '*I*'), the promoter (*p*) for the metabolic enzymes, the primary operator sequence LacO ('*O*', site for LacI binding), and the co-transcribed metabolic enzyme genes (*lacZYA*, '*Z*,' '*Y*,' and '*A*'). The cartoon of the functional cycle represents LacI tetramer (red ovals) binding to operator, LacI association with inducer (diamond) to form LacI-inducer complex (diamond inside yellow squares), LacI dissociation from operator (induction), RNA polymerase ('RNA pol') binding and transcription, and subsequent reassociation of LacI with LacO when sugar levels diminish. Shown at the bottom of the figure is the dimeric structure of CRP (teal rectangles), CRP association with cAMP (triangle) to generate CRP-cAMP complex (triangles inside purple rectangles), and CRP-cAMP binding within the promoter region to facilitate RNA polymerase binding and transcription. CRP dissociates from the promoter when glucose levels increase and cAMP levels fall. Changes in color/shape for protein symbols indicate conformational changes in response to binding the respective ligand (inducer or cAMP). Substitution of the *lac* metabolic genes by other protein coding sequences (prokaryotic or eukaryotic) allows effective control of specific protein expression using inducer/cAMP levels.

for two classes of proteins (metabolic and regulatory), 'promoter regions' that provide binding sites for RNA polymerase (the protein that transcribes DNA to mRNA), and binding sites for regulatory proteins. In order of their genomic appearance (Figure 1), the elements of the *lac* operon include:

1. the promoter for initiating transcription of *LacI*, followed by the *lacI* gene, which encodes the *lac* repressor protein (*LacI*);
2. the promoter for the *lac* metabolic genes involved in transporting and metabolizing lactose;
3. DNA-binding sites for *lac* repressor (*LacO*) and CRP that overlap this promoter; and
4. the genes for the *lac* metabolic proteins.

The latter encode proteins required for the transport and metabolism of lactose: β -galactosidase (*LacZ*), lactose permease (*LacY*), and thiogalactosidase transacetylase (*LacA*). Further discussion will focus on the functions and structures of the two regulatory proteins, *LacI* and CRP.

Lactose Repressor Protein

LacI Function

In the absence of available lactose, the role of *LacI* is to inhibit mRNA production for proteins encoded by the *lac*

operon. Transcription is not completely eliminated, but *lacZYA* mRNA is transcribed only at very low levels. This function is accomplished by *LacI* protein binding with high specificity and high affinity to the *lac* operator DNA sequence.

Transcription is subsequently inhibited via a variety of mechanisms. Since the *lac* operator (*LacO*) overlaps the promoter, *LacI* binding directly competes with RNA polymerase for binding this region. *LacI* can also impede transcription initiation and/or block elongation of mRNA to ensure effective transcription repression.

When lactose is available as a carbon source, the low levels of the metabolic protein *LacY* transport a small amount of lactose into the bacterium. Next, residual *LacZ* metabolizes lactose to glucose and galactose, which produces energy for the bacterium. Notably, this catalytic process also generates low levels of allolactose via a rearrangement of the β -1,4 linkage between glucose and galactose to a β -1,6 linkage (Figure 2). The by-product allolactose binds to *LacI* and elicits a conformational change in the repressor protein that results in release of the operator DNA sequence (induction). Consequently, RNA polymerase is freed to generate numerous copies of mRNA encoding the *lac* metabolic proteins. When translated into proteins, these metabolic proteins allow the bacterium to transport and metabolize large quantities of lactose as its carbon energy source, taking advantage of an environmental

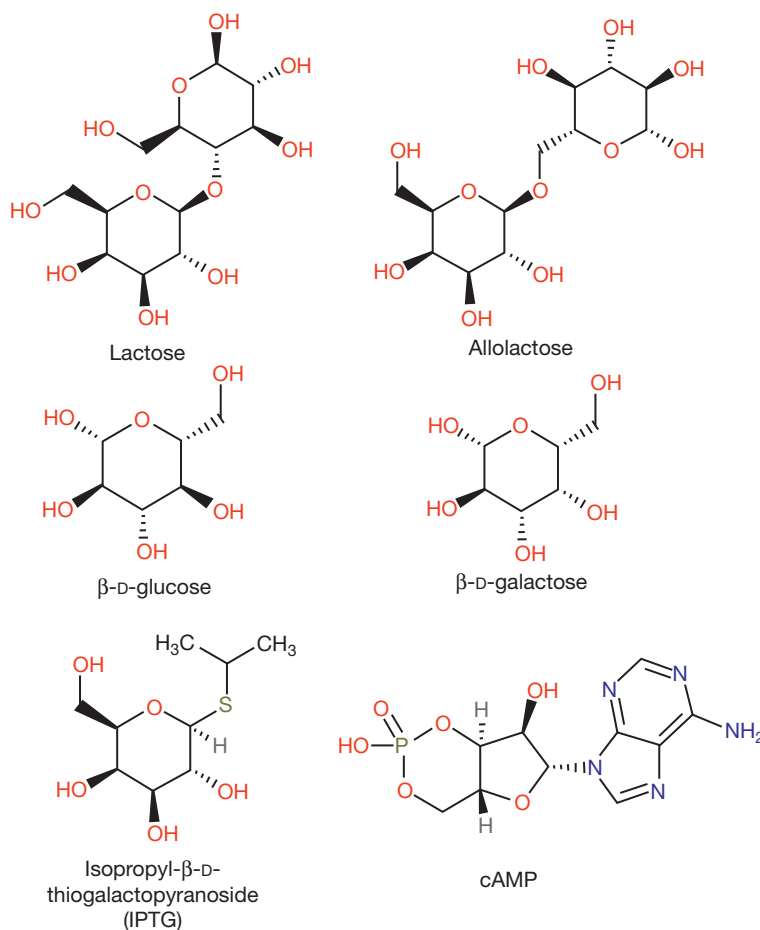


Figure 2 Chemical structures for lactose, allolactose (the natural inducer of *LacI*), the sugar components of lactose (glucose and galactose), the *LacI* gratuitous inducer IPTG, and cAMP. Structures were derived from ChemSpider (<http://www.chemspider.com>).

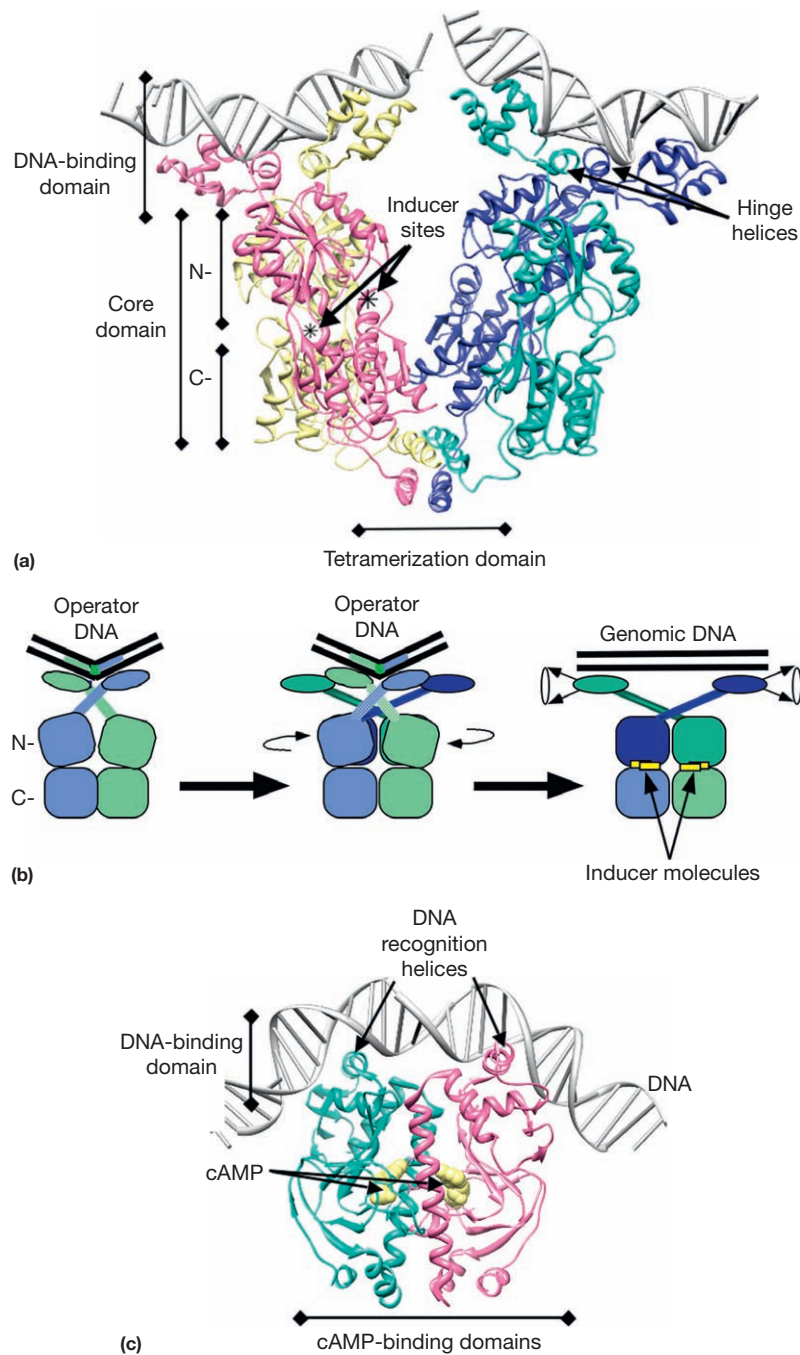


Figure 3 Structures of LacI and CRP. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081). (a) The tetrameric structure of the lactose repressor is shown in ribbons format; each monomer is a different color. This structure is from PDB file 1LBG. (This view does not show all of the atoms in the protein, since some are not resolved in the original structure.) DNA is shown at the top of the figure. Note that the dimer is the DNA-binding unit, whereas the monomer is the inducer-binding unit. The DNA-binding domain, hinge helices, core domain (with its N- and C-subdomains), and tetramerization domain are labeled. (b) Cartoon depicts the conformational change of a LacI dimer when it binds inducer. (c) The dimeric structure of CRP from PDB file 1RUN is shown with the sites for cAMP- and DNA-binding indicated. The DNA recognition helices (one per monomer) are also indicated. Two structures are known for apo-CRP (not bound to DNA or cAMP). Both structures show that the DNA recognition helices move significantly in the absence of cAMP binding, precluding effective contact with DNA. However, the two structures do not agree as to the final position of the helices.

opportunity. One result of the studies by Jacob and Monod was the discovery that a variety of non-natural galactoside sugars (e.g., IPTG; [Figure 2](#)) can induce LacI and relieve transcription repression of *lacZYA*. This discovery has been very important for biotechnology applications of this regulatory system.

LacI Structural Characteristics

Structural detail of the LacI protein has provided insight into the atomic-level mechanism by which this protein performs its functions. LacI is a multimeric, multidomain protein, an

arrangement now known to be typical for genetic regulatory proteins. LacI has four identical polypeptide chains (**Figure 3(a)**). Each polypeptide folds into distinct regions that have specific functions, and four monomers associate to form tetrameric LacI.

Folded protein domains can be isolated from the intact protein and maintain functions that are highly similar to the intact protein. For LacI, the N-terminal domain of a monomer, either with or without parts of the hinge-helix sequence, can be isolated by proteolysis. An isolated N-terminal domain can bind to one-half of LacO DNA with low affinity. If two of these domains plus the hinge-helix sequences are linked by a disulfide bond, this protein fragment binds to full-length LacO with very high affinity (but is not inducible by galactoside sugars). (Note that the LacO site is highly, but imperfectly, symmetric in sequence, therefore accommodating the two identical DNA-binding domains.) Only high-affinity binding by LacI represses mRNA transcription. Thus, the LacI dimer is the functional unit, and a LacI tetramer has two complete DNA-binding sites. Within the intact LacI–LacO complex, each N-terminal domain associates with bases in the DNA major groove, and the hinge helices insert into the minor groove to elicit a bend in the DNA (see **Figure 3(a)**).

The other product of proteolysis is the LacI core domain. This fragment is isolated as a tetramer, with each monomer comprised of two subdomains that flank the site for sugar binding. Thus, each monomer is able to bind an inducer molecule, creating four sugar binding sites per tetramer. (Note that the N- and C-subdomains within the core domain monomer are interconnected by multiple strands and cannot be separated by proteolysis.) When the inducer-binding site is occupied in full-length LacI, structural changes in the

N-subdomain affect the DNA-binding domain (**Figure 3(b)**). These alterations result in decreased operator-specific binding affinity; the inducer-bound repressor protein dissociates from the operator DNA site and is nonspecifically recruited to other regions of the genomic DNA.

The final domain of LacI is the C-terminal tetramerization domain, which connects two dimers via an antiparallel four-helix bundle (**Figure 3(a)**). When this domain is removed from the protein, the resulting LacI dimer is still capable of transcription repression. However, the amount of repression is decreased ~50-fold relative to that of tetramer LacI, for reasons described below.

Multiple Sites/Multiple Targets – DNA Looping

In addition to the primary operator, LacO, shown in **Figure 1**, two ‘pseudo-operator’ sequences are present within the *lac* operon sequence and contribute to repression. The DNA sequences of the pseudo-operators are very similar, but not identical, to LacO and are bound by LacI more weakly. The presence of two DNA-binding sites in LacI protein tetramer suggested a mechanism by which pseudo-operators could enhance repression – one LacI tetramer could bind two separate operators and generate a looped DNA structure. Experimental evidence for DNA looping has been obtained from a variety of laboratories. Recent evidence indicates that the angle between the two dimers must open for loop formation to occur, as illustrated in **Figure 4**. These looped structures are highly stabilized, accounting for the significant repression of *lacZYA* expression observed in bacterial cells. Indeed, DNA containing multiple operator sequences and with the supercoiling density characteristic of *E. coli* exhibits a half-life for the complex that exceeds 2 days. However, even these looped complexes respond rapidly (in less than 30 s) to the presence of inducer sugars, allowing quick adaptation to an external lactose source that may be transient.

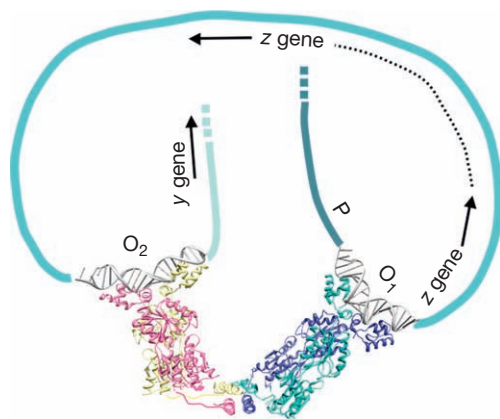


Figure 4 Looped DNA structure. The teal blue curved line depicts the *lac* operon DNA (with shading to indicate nearness to observer), which contains three possible LacI-binding sites (two of which, O_1 and O_2 , are shown bound to LacI). The pseudo-operator sequence O_2 is located within the *lacZ* gene, and the primary operator sequence O_1 overlaps the promoter sequence for the *lacZYA* metabolic genes (**Figure 1**). Tetrameric LacI is shown at the bottom of the figure as simultaneously interacting with O_1 and O_2 . This structure loops the DNA and generates a complex with very high stability. Note that the dimers within a LacI tetramer separate and adopt a larger angle between them when looping between O_1 and O_2 than in the absence of looping (i.e., the structure shown in **Figure 3(a)**). The need for flexibility between the dimers for looping to occur is supported by experimental evidence.

Catabolite Repressor Protein

CRP Function

Bacteria find it energetically beneficial to utilize lactose only when glucose is unavailable. The *lac* operon thus has two regulatory switches, one for each sugar. Glucose levels are measured through a protein designated CRP. This monitoring system is indirect and occurs by detecting levels of a signaling molecule called ‘cAMP’ (**Figure 2**). cAMP is created by another protein called adenylate cyclase, and adenylate cyclase is directly regulated (inhibited) by high glucose. Thus, levels of glucose and cAMP are inversely related. As glucose levels decrease, cAMP levels rise, and cAMP binds to CRP. The CRP–cAMP complex has high affinity for specific target sequences within the bacterial genome (**Figures 1** and **3**). One of the DNA target sequences for cAMP–CRP is found within the promoter for the *lac* metabolic proteins, upstream of the *lac* operator sequence (**Figure 1**). Note that cAMP binding enhances CRP binding to DNA, whereas with LacI, sugar binding decreases the DNA binding. Another difference is that DNA binding by the cAMP–CRP complex enhances RNA polymerase transcription of *lacZYA* mRNA, whereas LacI binding represses it. The end result is that the proteins for lactose metabolism are

expressed at high levels only when environmental glucose is low and lactose levels are high.

CRP Structural Characteristics

Like LacI, CRP is multimeric and has multiple domains (Figure 3(c)). In this case, the CRP structure is dimeric, and each monomer comprises two domains. The N-terminal domain facilitates dimer formation and binds at least one molecule of cAMP (two/dimer). In the absence of cAMP, the two DNA-binding domains, one from each monomer, are not aligned properly for sequence-specific DNA recognition. In the presence of cAMP, the structure rearranges to bring the two DNA-binding helix–turn–helix structures into the proper orientation for effective DNA binding. Similar to LacI, CRP–cAMP binding also elicits a bend in the target DNA-binding site. Each monomer only binds to one-half of the DNA-binding site, which is symmetric through the center of the sequence. Indeed, the symmetry seen in the *lac* operon regulatory sites is a common feature of DNA regulatory binding sites, going hand in hand with the generally multimeric natures of these proteins.

Three Common Errors in Understanding the *lac* Operon

Several errors are encountered frequently in presentations of the *lac* operon. The most common error, propagated in multiple venues, is that lactose is the inducer that relieves LacI repression. As indicated in this article, the molecule that serves as inducer *in vivo* is a derivative of lactose, allolactose, that is generated by β -galactosidase as a side reaction in the cleavage of lactose to glucose and galactose (see Figure 2). This arrangement ensures that the inducer is not produced unless the metabolic potential to utilize lactose is present in the cell.

The second common error is that *lacZYA* mRNA is not produced at all in the absence of lactose (and hence the inducer allolactose). Even when no inducer is available, normal dissociation events of LacI from LacO allow a small amount of read-through by RNA polymerase, resulting in transcription of a very low level of *lacZYA* mRNA. This low-level transcription ensures that small amounts of lactose permease and β -galactosidase proteins are available to transport lactose and to generate allolactose for induction. This leaky transcription can be a problem when the LacI/LacO system is used to control recombinant protein expression: if the recombinant protein is toxic to *E. coli*, even tiny quantities can kill the growing bacteria. Thus, LacI/LacO regulation has been combined with other control systems (such as bacteriophage T7 polymerase) in order to exert a high level of repression while bacterial cultures grow to high densities, after which protein production is induced.

The third error is the assertion that induced LacI (bound to allolactose or gratuitous inducers) can no longer bind to the LacO operator. In fact, LacI inducer can bind to LacO, but the strength of the binding is greatly diminished. With this decrease in binding affinity, the LacI-inducer complex can no longer effectively discriminate LacO from the rest of the DNA in the bacterial genome. Since the genome is vastly larger than LacO (which represents only a very, very small fraction of the *E. coli* DNA sequence), the repressor protein is competed

away by this excess of nonspecific DNA, and RNA polymerase is able to transcribe *lacZYA*.

Implications for Complex Organisms

This simple bacterial regulatory system illustrates the complexity that can be generated with only two on/off switches. In fact, the response of this system is more analogous to a rheostat. Since the two protein switches respond to different signals, graded response occurs at intermediate ratios of glucose/lactose. Gene regulation in response to specific signals – whether related to environmental or internal changes – has a twofold advantage. The organism conserves energy/resources under conditions where gene expression would be useless (and expensive), but is able to selectively respond to a variety of input information. This type of regulatory control, with binding modulated by interactions with small ligands or other proteins, underlies the incredibly complex pathways by which gene expression is fine-tuned to create specific temporal, spatial, and environmental patterns in higher organisms.

See also: **Metabolism Vitamins and Hormones:** Regulation of Gene Transcription by Hypoxia-Inducible Factor 1; **Molecular Biology:** DNA Sequence Recognition by Proteins; DNA Supercoiling; Gene Expression in Bacterial Systems: The *trp* Operon and Attenuation; LexA Regulatory System; Messenger RNA Degradation in Bacteria; Organization of the Bacterial Nucleoid; RNA Polymerase Reaction in Bacteria; RNA Polymerase Structure, Bacterial; Small RNAs in Bacteria; T7 RNA Polymerase; Transcription Termination; Translation Initiation in Bacteria: Factors and Mechanisms; **Protein/Enzyme Structure Function and Degradation:** Allosteric Regulation; Zinc Fingers.

Further Reading

- Busby S and Ebright RH (1999) Transcription activation by catabolite activator protein (CAP). *Journal of Molecular Biology* 293: 199–213.
- Emmer M, deCrombrughe B, Pastan I, and Perlman R (1970) Cyclic AMP receptor protein of *E. coli*: Its role in the synthesis of inducible enzymes. *Proceedings of the National Academy of Sciences of the United States of America* 66: 480–487.
- Harman JG (2001) Allosteric regulation of the cAMP receptor protein. *Biochimica et Biophysica Acta* 1547: 1–17.
- Jacob F and Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3: 318–356.
- Kolb A, Busby S, Buc H, Garges S, and Adhya S (1993) Transcriptional regulation by cAMP and its receptor protein. *Annual Review of Biochemistry* 62: 749–795.
- Lewis M (2005) The lac repressor. *C. R. Biology* 328: 521–548.
- Lewis M, Chang G, Horton NC, et al. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* 271: 1247–1254.
- Matthews KS and Nichols JC (1998) Lactose repressor protein: Functional properties and structure. *Progress in Nucleic Acid Research and Molecular Biology* 58: 127–164.
- Miller JH and Reznikoff WS (eds.) (1980) *The Operon*, 2nd edn. New York, NY: Cold Spring Harbor Laboratory.
- Müller-Hill B (1996) *The lac Operon: A Short History of a Genetic Paradigm*. Berlin, New York: Walter de Gruyter.

- Popovych N, Tzeng S-R, Tonelli M, Ebright RH, and Kalodimos CG (2009) Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. *Proceedings of the National Academy of Sciences of the United States of America* 106: 6927–6932.
- Sharma H, Yu S, Kong J, Wang J, and Steitz TA (2009) Structure of apo-CAP reveals that large conformational changes are necessary for DNA binding. *Proceedings of the National Academy of Sciences of the United States of America* 106: 16604–16609.
- Swint-Kruse L and Matthews KS (2009) Allostery in the LacI/GalR family: Variations on a theme. *Current Opinion in Microbiology* 12: 129–137.
- Wilson CJ, Zhan H, Swint-Kruse L, and Matthews KS (2007) Lactose repressor system: Paradigms for regulation, allosteric behavior and protein folding. *Cellular and Molecular Life Sciences* 64: 3–16.
- Zubay G, Schwartz D, and Beckwith J (1970) Mechanism of activation of catabolite-sensitive genes: A positive control system. *Proceedings of the National Academy of Sciences of the United States of America* 66: 104–110.

Lectins

N Sharon[†] and H Lis, Weizmann Institute of Science, Rehovot, Israel

Published by Elsevier Inc.

Glossary

Glycan Any compound consisting of a few or many monosaccharides in free form or attached to another substance; it is referred to as a carbohydrate or saccharide.

Glycoconjugate A substance that consists of one or more glycans covalently attached to a noncarbohydrate constituent, typically protein (to give a glycoprotein) or lipid (forming a glycolipid).

Innate immunity A form of defense of animals and other organisms that acts against infection in the first minutes or hours. It is different from acquired or adaptive immunity that takes days or weeks to develop.

Mitogen A substance that stimulates cell division, especially of lymphocytes.

Lectins are a class of proteins that bind carbohydrates specifically and reversibly. They occur widely in animals, plants, and microorganisms, where their main function is to serve as cell recognition molecules. Their main use is for the detection, isolation, and characterization of glycoconjugates in solution as well as in and on cells.

Lectins (from Latin, *legere*, to select or choose) are referred to as proteins that combine with mono- and oligosaccharides reversibly and with high specificity, but are devoid of catalytic activity and, in contrast to carbohydrate-binding antibodies, are not products of an immune response. They are found in all classes of organisms, from viruses and bacteria to higher animals, including humans (Table 1) and occur in different shapes and forms (Figure 1). Hundreds of lectins have been isolated by affinity chromatography on immobilized carbohydrates, as well as by recombinant DNA techniques. The richest source of these proteins are plants, from the seeds of which several of the long-known proteins of this class – such as concanavalin A, phytohemagglutinin (PHA), and soybean agglutinin (SBA) – have been obtained. Most lectins are bi- or oligovalent with respect to carbohydrate binding; therefore, when they combine with sugars on the surface of cells, such as erythrocytes, they cause cross-linking of the cells and their subsequent precipitation, a phenomenon referred to as cell agglutination. The erythrocyte-agglutinating, or hemagglutinating, activity of lectins is a major attribute of these proteins and is used routinely for their detection and characterization. Lectins also possess the ability to precipitate polysaccharides or glycoproteins from solution; both the cell agglutinating and precipitating activities of lectins are inhibited by the sugars for which the lectins are specific.

Carbohydrate Specificity

Lectins are classified primarily into five specificity groups, according to the monosaccharide for which they exhibit the highest affinity: mannose, galactose/*N*-acetylgalactosamine, *N*-acetylglucosamine, fucose, and *N*-acetylneuraminic acid (sugars are of the D-configuration except for fucose which

is L). Relevant for the biological functions of lectins is the fact that, of the numerous monosaccharides found in nature, these six are typical constituents of the glycans present on the surfaces of eukaryotic cells. Some lectins specific for mannose also bind glucose, which differs from mannose in the configuration of the 2-hydroxyl, but only in rare cases they react with galactose, which differs from glucose in the configuration of the 4-hydroxyl. Lectins of the same specificity group may differ markedly in affinity for oligosaccharides or cells, while a few of them recognize only oligosaccharide derivatives of the above monosaccharides. Specificity for other sugars has been rarely encountered. The affinity of lectins to monosaccharides is usually weak, with association constants in the millimolar range, whereas that to oligosaccharides is often much higher, up to three orders of magnitude.

Molecular Properties

In general, lectins are oligomeric proteins of diverse structure and molecular size. They differ in primary, secondary, and tertiary structure, number of subunits and subunit assembly (Figure 2), metal requirement, as well as in the constitution of carbohydrate-binding sites. The polypeptide chains often consist of several molecular domains, only one of which is the carbohydrate recognition domain (CRD). Lectins belong to distinct protein families or superfamilies that share sequence homologies and similar tertiary and quaternary structures. These families are often of the same phylogenetic origin (e.g., the cereal lectins), although they sometimes cross phylogenetic barriers (as in the case of the structurally similar legume lectins that belong to different specificity groups and the galectins, animal lectins that bind exclusively β -galactosides, such as lactose and *N*-acetyllactosamine [$\text{Gal}\beta(1\text{--}4)\text{GlcNAc}$]). The C-type (Ca^{2+} -requiring) lectins are members of a superfamily of diverse specificities that include endocytic lectins, collectins, selectins, and siglecs, many of which are membrane bound. In collectins, the CRD is linked to a collagen-like region, whereas in selectins – E-selectin, P-selectin, and L-selectin – it is attached to an epidermal growth-factor-like domain and several short repeating units related to complement-binding

[†]Deceased.

Table 1 Representative lectins

Source	Organism or lectin name	Specificity	Mol. wt. (kDa)	No. of subunits
Viruses	Influenza A and B	NeuAc	225	3
Bacteria	Cyanovirin-N	OligoMan	11	
	<i>Escherichia coli</i> type 1	Man		Multi
	<i>Escherichia coli</i> P	Gal α 4Gal		Multi
	<i>Pseudomonas aeruginosa</i>	Fuc	47	4
Protozoa	<i>Entamoeba histolytica</i>	Gal/GalNAc	260	2
Plants	<i>Arachis hypogaea</i> (Peanut) (PNA)	Gal	110	4
				4
	<i>Artocarpus integrifolia</i> (Jackfruit)	Gal	66	2
		Glc/Man		
	<i>Canavalia ensiformis</i> (Jack bean) (Con A)	Glc/Man	106	4
	<i>Erythrina corallodendron</i> (Coral tree)	Gal/GalNAc	57	2
	<i>Galanthus nivalis</i> (Snowdrop) (GNA)	Man	50	4
	<i>Glycine max</i> (Soybean) (SBA)	Gal/GalNAc	120	4
	<i>Lotus tetragonolobus</i> (Asparagus pea)	Fuc	110	4
	<i>Ricinus communis</i> (Castor bean)			
	Toxin	Gal/GalNAc	60	2
	Lectin	Gal/GalNAc	120	4
	<i>Sambucus nigra</i> (Elderberry) (SNL)	NeuAc α 2,6Gal	240	4
	<i>Triticum vulgare</i> (Wheat) (WGA)	GlcNAc	36	2
Invertebrates	<i>Anguilla anguilla</i> (Eel) (AAA)	Fuc	50	3
	<i>Helix pomatia</i> (Garden snail)	GalNAc	100	6
	<i>Limulus polyphemus</i> (Horseshoe crab)	NeuAc	400	
		Gal	28	2
Vertebrates	Mammalian hepatic binding protein (HBP)	Gal/GalNAc	130–190	3–4
	Macrophage mannose receptor (MMR)	Man	175	1
	Mannose-binding lectin A (MBL-A)	Man	200–600	6–18
	Selectins E, P and L	Sialyloligosach.	90–140	1
	Man-6-P receptors	Man6P	90–300	1–2
	Siglec-1	NeuAc α 2,3Gal	185	1
	Calnexin	Glc α 3Man	65	1
	Ficolin	GalNAc/GlcNAc	140	4

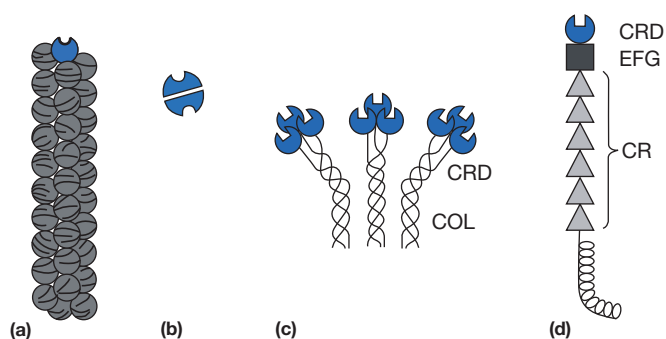


Figure 1 Schematic representation of the overall structure of lectins from different sources. (a) Part of fimbriae of *Escherichia coli*. The fimbriae are an assembly of different types of subunits, only one of which (blue) contains a carbohydrate-binding site. (b) A dimeric legume lectin (e.g., from lentils or peas). (c) Mammalian-soluble mannose-binding lectin, a collectin. (d) E-selectin; the coil represents the membrane-spanning and cytoplasmic domains. COL, collagenous region; CR, complement regulatory repeats; EGF, epidermal growth factor-like domain; CRD, carbohydrate recognition domain. The indentations represent binding sites. Reproduced from Sharon N and Lis H (1995) Lectins – proteins with a sweet tooth: Functions in cell recognition. *Essays in Biochemistry* 30: 59–75.

protein (Figure 1). The siglecs are a family of sialic-acid-specific cell-surface lectins with variable numbers of extracellular immunoglobulin-like domains and are thus members of the immunoglobulin superfamily. Bacterial lectins are typically in the form of fimbriae (or pili), hair-like submicroscopic appendages, consisting of an assembly of several different

types of subunits, only one of which has carbohydrate-binding activity, for example, for mannose (in type 1 fimbriae of *Escherichia coli* and *Klebsiella pneumoniae*) or galabiose, Gal α (1–4)Gal (in P fimbriae of *E. coli*).

Besides the surface lectins described above, a small number of soluble bacterial lectins have been described. These include

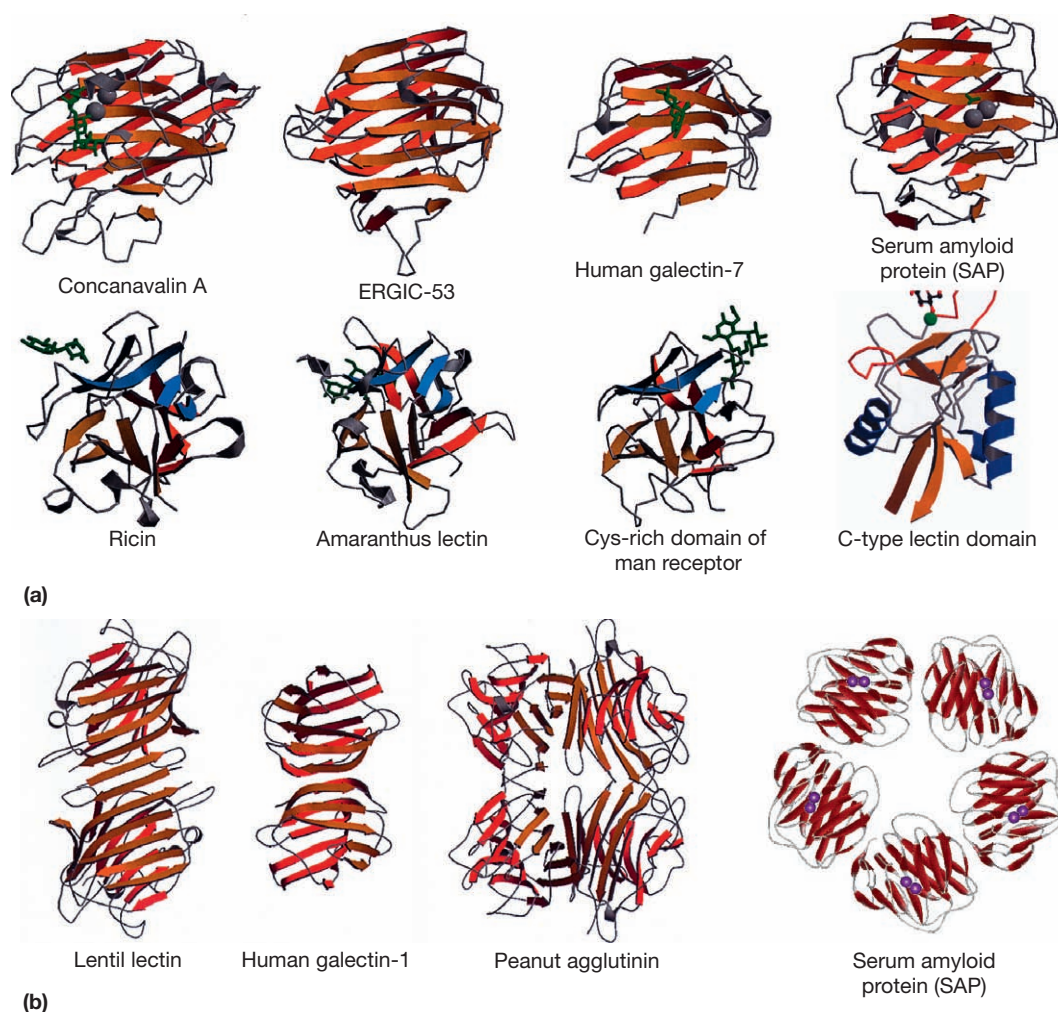


Figure 2 Structures of different lectins represented as ribbon diagrams. (a) Upper row shows monomers of lectins from different sources that share the jelly roll or lectin fold. First three lectins (from left) in the lower row all exhibit the β -trefoil fold. (b) Variations in quaternary lectins structures. The gray spheres represent metal ions; bound carbohydrate is shown in ball-and-stick representation. The diagram of SAP is from PDB entry sac. All diagrams, except for that of SAP: reprinted from Loris R (2002) Principles of structures of animal and plant lectins. *Biochimica et Biophysica Acta* 1572: 198–208.

cyanovirin-N, an oligomannose-specific cyanobacterial lectin that is a potent inactivator of all strains of human immunodeficiency virus and simian immunodeficiency virus and has a potential for clinical use as an antiviral agent.

High-resolution X-ray crystallography of lectins in complex with their ligands and site-directed mutagenesis allowed the identification of the chemical groups on the protein and on the carbohydrate that interact with each other and of the types of bonds formed. These studies have shown that, like other carbohydrate-binding proteins, lectins combine with their ligands primarily by a network of hydrogen bonds and hydrophobic interactions; in rare cases, electrostatic interactions (or ion pairing) and coordination with metal ions also play a role. One or more water molecules sometimes mediate the bonding. Although in a single lectin a limited set of residues contribute to the formation of the bonds with the ligand, overall almost all kinds of amino acids participate in ligand binding.

Functions

Lectins possess numerous biological activities, nearly all of which are based on their ability to recognize and bind carbohydrates specifically (Table 2). In other words, as a result of their unusual capacity to decode the information encoded in carbohydrates, they act as recognition determinants in diverse biological processes. Thus, they function in cell–cell, cell–molecule, or molecule–molecule recognition. Influenza virus hemagglutinins A and B, lectins specific for *N*-acetylneuraminic acid (or hemagglutinin C specific for 9-*O*-acetyl-*N*-acetylneuraminic acid), as well as bacterial surface lectins mediate the binding of the pathogens to host cells, a prerequisite for initiation of infection. Inhibitors of bacterial lectins protect animals against experimental infection by the lectin-carrying organisms, providing a basis for attempts to develop anti-adhesion therapy of microbial infections. In some cases the bacterial lectins are responsible for the attachment to sugar residues on phagocytic cells, permitting the

Table 2 Lectin functions

Functions	Examples of lectins involved
<i>Microorganisms</i>	
Infectivity	<i>Escherichia coli</i> fimbriae; <i>Entamoeba histolytica</i> Gal/GalNAc-specific lectin; influenza virus hemagglutinin
<i>Plants</i>	
Defence	Snowdrop lectin, jacalin, wheat germ agglutinin
Symbiosis with nitrogen-fixing bacteria	Pea lectin
<i>Animals</i>	
Clearance of glycoproteins from circulation	Asialoglycoprotein receptor
Control of lymphocyte migration	Selectins
Control of glycoproteins biosynthesis	Calnexin, calreticulin, ERGIC-53
Innate immunity	MBLs, MMR, SP-A, SP-D
Complement activation	MBLs
Lectinophagocytosis	MMR, SP-A, SP-D
Modulation of cell-cell and cell-substratum interactions	Galectins
Modulation of signal transduction by B lymphocytes	Siglec-2
Neuronal myelination and regeneration	Siglec-4a and -4b
Targeting of glycoproteins to lysosomes	Man6P receptors
Tumor metastases	Galectins-1, -3, and -8; selectins

latter cells to kill the bacteria in the absence of serum factors, in a process designated lectinophagocytosis.

In plants, lectins may serve as defense agents against different kinds of phytopathogenic fungi and insects, as well as predatory invertebrates and vertebrates. They may also be involved in the establishment of symbiosis between nitrogen-fixing bacteria, mainly rhizobia, and leguminous plants, a process of cardinal importance both in the nitrogen cycle of terrestrial life and in agriculture.

The galectins are believed to function in cell adhesion. Their elevated levels on the surfaces of metastatic cancer cells may be responsible for the adhesion of these cells to target organs, a step necessary for metastasis formation. The mannose-6-phosphate (Man-6-P) receptors that are located intracellularly target lysosomal enzymes into the lysosomes. A defect in the synthesis of the Man-6-P marker on these enzymes results in I-cell disease (mucopolipidosis II or MLI). This is an inherited lysosomal storage disease, characterized by a lack of enzymes in the lysosomes that normally carry the marker, resulting in accumulation in these subcellular organelles of undigested glycoconjugates.

The endocytic lectins are responsible for the clearance from the circulatory system of glycoproteins (e.g., ceruloplasmin and α_1 -acid glycoprotein) from which terminal sialic acid has been removed, and possibly of cells too (e.g., aged erythrocytes or bacteria). A prominent representative is the mammalian

asialoglycoprotein receptor (or hepatic-binding protein) found on hepatocytes, specific for galactose and *N*-acetylgalactosamine. The mannose-specific lectin present on the surface of macrophages (MMR) has been implicated in antimicrobial innate immunity. It binds infectious organisms that expose mannose-containing glycans on their surface, leading to their killing. The soluble mannose-binding lectins (MBLs) of mammalian serum and liver and the pulmonary surfactant proteins A and D (SP-A and SP-D), all members of the collectin family, also function in innate immunity. MBL deficiency syndrome is associated with recurrent, severe bacterial infections in infants; it is caused by a mutation of a single amino acid in the collagen-like domain of the lectin. The selectins mediate the adhesion of circulating leukocytes to endothelial cells, a prerequisite for their extravasation into other tissues. They thus control the migration (homing) of lymphocytes to specific lymphoid organs and of leukocytes to sites of inflammation, where they act to eliminate bacteria and other foreign intruders. Individuals unable to synthesize oligosaccharides recognized by the selectins, such as sialyl-Le^x [NeuAc α (2-3)Gal β (1-4)[Fuc α (1-3)]GlcNAc] (as seen, for example, in cases of leukocyte deficiency type II (LADII)), suffer from recurrent microbial infections. The mechanism that helps leukocytes to break the endothelial barrier as a step in the fulfillment of their infection-fighting duties, is also responsible for their undesirable accumulation in tissues where they do not belong (e.g., joints), thereby causing damage, swelling, and pain. In reperfusion injury, which may occur after heart infarct or cerebral stroke, the accumulated leukocytes destroy tissues that have been damaged by temporary lack of oxygen. As shown in experimental animals, sialyl-Le^x attenuates myocardial necrosis after myocardial ischemia and reperfusion, another example of the great potential of antiadhesion therapy. In addition to their involvement in inflammation, the selectins may play a role in the attachment of the human blastocyst to the uterine lining and in the spread of cancer cells from the main tumor throughout the body to form metastases. The singlens have been implicated in cell-cell interactions in the immune system and in the process of mediation. Intracellular lectins such as connexin, calreticulin, and ERGIC-53 participate in the control of glycoprotein biosynthesis.

Applications

Lectins, mainly those from plants, are invaluable tools for detection, isolation, and structural and functional investigation of complex carbohydrates, especially glycoproteins. In immobilized form, for example, covalently bound to Sepharose, they are indispensable for purification and isolation by affinity chromatography of glycoproteins, glycopeptides, and oligosaccharides. Lectins are also useful as histochemical and cytochemical reagents: derivatized with fluorescent dyes, gold particles, or enzymes they serve for the detection of glycoconjugates in tissue sections and on cells, as well as on and in subcellular organelles, mainly for examination of changes that occur on cell surfaces and in cells during physiological and pathological processes, from differentiation to cancer.

An important recent addition to the armament of lectin-based analytical tools is that of microarrays. These microarrays consist of discrete probes of tiny spots of lectins with different specificities immobilized on a solid support such as glass or nitrocellulose. When charged with a fluorescent glycoprotein (or similarly labeled cells) a pattern characteristic for the glycoprotein is observed providing information on the structure of its glycans.

A few lectins are employed in blood typing, for example, those of the seeds of *Lotus tetragonolobus* and *Ulex europaeus* for the identification of type O cells, and of secretors of blood group substances. PHA and concanavalin A are potent *in vitro* T-lymphocyte mitogens, while another lectin, pokeweed mitogen, stimulates both T and B cells. Mitogenic stimulation by lectins provides a simple means to assess the immunocompetence of patients and to monitor the effects of immunosuppressive and immunotherapeutic treatments, as well as to prepare chromosome maps for karyotyping, sex determination, and detection of chromosome defects.

Selective agglutination by PNA permits the facile separation of mouse and human cortical (immature) thymocytes from the medullary (mature) ones. Treatment with SBA removes mature T cells from human bone marrow and affords a fraction enriched in stem cells. Lectin-purged bone marrow of haplo-identical donors serves routinely for transplantation into children born with severe combined immune deficiency (bubble children) with close to 75% success. The method was also used experimentally in the 1980s for bone marrow transplantation of leukemia patients. Knowledge of the properties of lectins was essential for the development of enzyme replacement therapy for the treatment of Gaucher disease; there is hope for lectin replacement therapy for patients with defects in innate immunity, for example, MBL-deficient patients.

A major application of toxic lectins, such as ricin, is for the selection of resistant animal cell lines. These cell lines provide access to glycoproteins with modified glycan structures and to novel glycosyltransferases. They are useful tools for the investigation of the mechanism of protein glycosylation and for the large-scale production of therapeutic glycoproteins with desired carbohydrate structures.

See also: [Lipids Carbohydrates Membranes and Membrane Proteins: Siglecs.](#)

Further Reading

- Barrientos LG and Gronenborn AM (2005) The highly specific carbohydrate-binding protein cyanovirin-N: Structure, anti-HIV/Ebola activity and possibilities for therapy. *Mini-Reviews in Medicinal Chemistry* 5: 21–31.
- Dommett RM, Klein N, and Turner MW (2006) Mannose-binding lectin in innate immunity: Past, present and future. *Tissue Antigens* 68: 193–209.
- Fuster MM and Esko JD (2005) The sweet and sour of cancer: Glycans as novel therapeutic targets. *Nature Reviews of Cancer* 5: 526–542.
- Hsu K-L and Mahal LK (2009) Sweet tasting chips: Microarray-based analysis of glycans. *Current Opinion in Chemical Biology* 13: 427–432.
- Kilpatrick DC (2004) Animal lectins: A historical introduction. *Biochimica et Biophysica Acta* 1572: 187–197.
- Lannoo N and Van Damme EJM (2010) Nucleocytoplasmic plant lectins. *Biochimica et Biophysica Acta* 1800: 190–201.
- Liu FT and Rabinovich GA (2005) Galectins as modulators of tumour progression. *Nature Reviews of Cancer* 5: 29–41.
- Mitoma J, Bao XF, Petryanik B, et al. (2007) Critical functions of N-glycans in L-selectin-mediated lymphocyte homing and recruitment. *Nature Immunology* 8: 409–418.
- Patnaik SK and Stanley P (2006) Lectin-resistant CHO glycosylation mutants. *Methods in Enzymology* 416: 159–182.
- Sharon N (2006) Carbohydrates as future anti-adhesion drugs for infectious diseases. *Biochimica et Biophysica Acta* 1760: 527–537.
- Sharon N (2009) Lectins – past, present and future. *Biochemical Society Transactions* 36: 1457–1460.
- Sharon N and Lis H (1993) Carbohydrates in cell recognition. *Scientific American* 268 (1): 82–89.
- Sharon N and Lis H (1995) Lectins – proteins with a sweet tooth: Functions in cell recognition. *Essays in Biochemistry* 30: 59–75.
- Sharon N and Lis H (2004) History of lectins: From hemagglutinins to cell recognition molecules. *Glycobiology* 14: 54R–64R.
- Tateno H, Uchiyama N, Kuno A, et al. (2007) A novel strategy for mammalian cell surfaces glycome profiling using lectin microarray. *Glycobiology* 17: 1138–1146.
- Taylor ME and Drickamer K (2009) Structural insights into what glycan arrays tell us about how glycan-binding proteins interact with their ligands. *Glycobiology* 19: 1155–1162.
- van Kooyk Y and Rabinovich GA (2008) Protein–glycan interactions in the control of innate and adaptive immune responses. *Nature Immunology* 9: 593–601.
- Vasta GR and Ahmed H (2009) *Animal Lectins. A Functional View*, 538 pp. Boca Raton, FL: CRC Press/Taylor and Francis.
- Wellens A, Garofalo C, Nguyen H, et al. (2008) Intervening with urinary tract infections using anti-adhesives based on the crystal structure of the FimH–oligomannose-3 complex. *PLoS ONE* 3: e2040.

LexA Regulatory System

M Butala and D Žgur-Bertok, University of Ljubljana, Ljubljana, Slovenia
S J W Busby, The University of Birmingham, Birmingham, UK

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Veronica G. Godoy, Penny J. Beuning, and Graham C. Walker, volume 2, pp. 546–550, © 2004, Elsevier Inc.

Glossary

Autoregulation A gene product regulates expression of its own gene.

Chromatin immunoprecipitation Technique used to precipitate a protein antigen using specific antibody to identify protein–DNA interactions at the genome level.

DNA microarrays A surface carrying an array of probes, DNA oligonucleotides corresponding to genes of interest, which are hybridized with cDNA from RNA isolated from cells under a given condition.

Operator Specific DNA site where transcription factor binds and modulates initiation of gene transcription.

Promoter Sequence located upstream of a gene to which RNA polymerase binds to initiate transcription.

Regulon Group of genes whose expression is regulated by a common regulator(s).

Repressor Protein that inhibits gene expression by sterically interfering with binding of RNA polymerase or by binding to RNA.

Introduction

The *Escherichia coli* LexA regulon is a regulatory network, encompassing at least 57 genes whose products govern a coordinated bacterial response to DNA damage. The induced LexA regulatory system has also been designated the SOS response to emphasize its role in the cellular response to distress. The expressed SOS functions not only repair DNA damage but also enhance adaptation through mutagenesis and genetic exchange. The SOS response thus plays a broad role, modulating evolution and dissemination of drug resistance and virulence factor genes, as well as the synthesis and secretion of virulence factors. In addition, the SOS system controls persistence and multidrug tolerance in a subpopulation of bacterial cells. The SOS system is widespread among bacteria but exhibits considerable variation with regard to its components and regulation. This article outlines regulation by LexA in *E. coli*, which is the best-understood SOS system and has been studied most extensively.

The *E. coli* LexA Regulatory System

Control of gene expression in response to environmental assaults, and the maintenance of the structural and functional integrity of the genome are essential for cell survival. The bacterial SOS system is an inducible DNA repair and damage-tolerance response triggered either by extrinsic treatments that elicit DNA damage or by intrinsic events that disrupt DNA replication.

A comprehensive response to DNA lesions was first described in detail in *E. coli*. Evelyn M. Witkin postulated that cellular filamentation and phage induction are regulated by a common repressor, which is inactivated in response to DNA damage. In the 1970s, Miroslav Radman proposed that a coordinated cellular response controlled by the interplay of two key proteins, a repressor and an inducer, is mounted

upon DNA damage. The product of the *lexA* gene (locus for X-ray sensitivity A) is the repressor of the regulon while recombinase A (RecA) is involved in sensing DNA damage and induces inactivation of the LexA repressor. During normal bacterial growth, LexA downregulates expression of its own gene and, in *E. coli*, the expression of more than 50 unlinked genes. In response to DNA damage, RecA (bound to adenosine triphosphate (ATP)) polymerizes onto single-stranded DNA (ssDNA) exposed upon repair or replication of damaged DNA, creating a helical nucleoprotein filament. The active ssDNA–ATP–RecA filament (RecA*) interacts with LexA and activates its latent self-cleaving activity. Cleavage inactivates LexA, instigating repressor dissociation from its DNA targets (SOS boxes) and induction of the LexA regulon. Subsequently, as DNA damage is repaired or bypassed, the level of ssDNA, the SOS-inducing signal, decreases and the co-protease activity of RecA filaments disappears (note, RecA* does not participate directly in the proteolysis reaction but instead stimulates LexA cleavage and is thus termed a ‘co-protease’). Functional LexA rapidly re-accumulates, returning the system to its repressed state (Figure 1).

Defining the LexA Regulon

Genes of the SOS regulon are characterized by (1) basal-level expression during normal bacterial growth and induction following DNA damage; (2) absence of induction in the *lexA* (*ind*) mutant strain with noncleavable LexA protein; (3) constitutive induction in strains carrying the *lexA* (*def*) allele, due to impaired repressor dimerization and unstable DNA association; and (4) promoter regions that carry DNA targets that resemble the conserved LexA operator sequence.

The first investigations to show that the SOS response is a global genomic response to DNA damage were performed in Graham C Walker’s laboratory. Through random insertion of a *lacZ* reporter gene into the *E. coli* chromosome, they

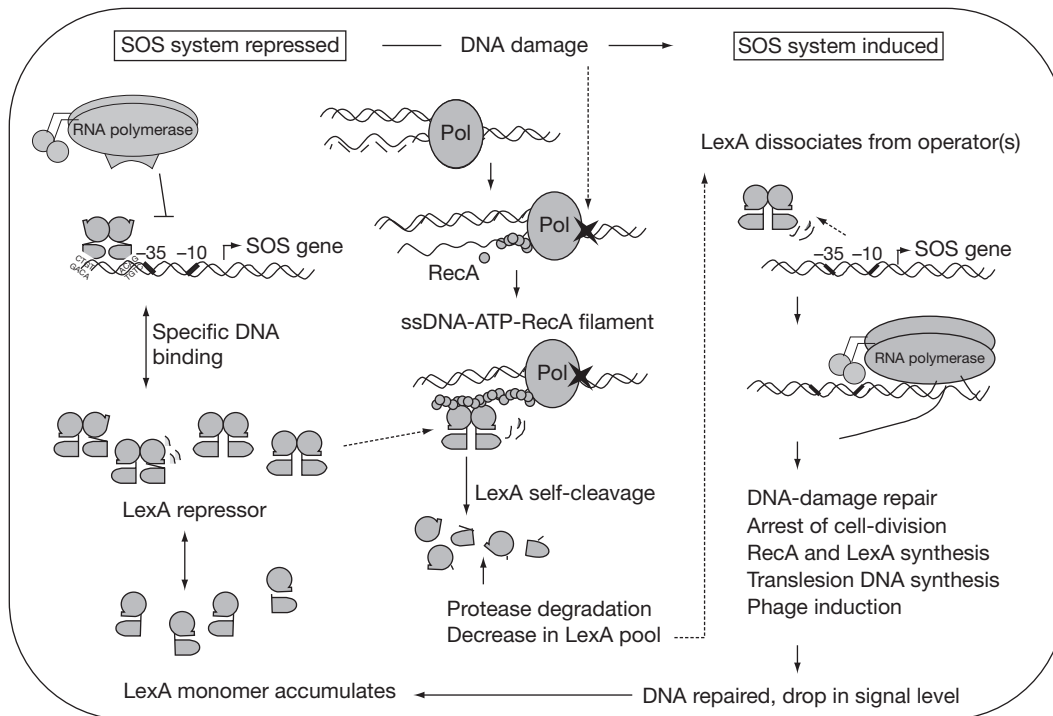


Figure 1 An overview of the SOS response in *E. coli*. In the uninduced state, LexA repressor binds to the promoter regions of SOS genes and sterically precludes their transcription. The polymerase III holoenzyme (Pol) carries out DNA replication. In the induced state, at the site of DNA damage PolIII arrests, ssDNA accumulates and active RecA filament is formed. Due to induced LexA self-cleavage, specifically bound LexA repressor dissociates from operators, leading to de-repression of SOS genes. Subsequently, as DNA damage is repaired, SOS induction is reversed. Adapted from Butala M, Žgur-Bertok D, and Busby SJW (2009) The bacterial LexA transcriptional repressor. *Cellular and Molecular Life Sciences* 66: 82–93, with permission from Springer Basel AG.

identified genes whose expression was induced following DNA damage. Characterization of genes upregulated in a *recA*/*lexA*-dependent manner revealed a 20-base-pair consensus LexA-binding site in promoter regions of SOS genes. Whole genome technologies that use microarrays to analyze transcriptome or chromatin immunoprecipitation experiments have now identified the full catalog of genes regulated by LexA. While the roles of most of the newly identified LexA-regulated genes are still unknown, unraveling their particular functions will yield insight into the molecular mechanisms underlying the SOS response. Several gene transcripts are decreased following DNA damage and some, while exhibiting a similar expression profile as genes of the LexA regulon, are not directly regulated by LexA. It thus seems that the SOS response is part of a larger, coordinated response network.

The LexA Regulatory System in the Repressed State

LexA exerts repression by binding to target sites located near promoters of SOS genes, blocking access of RNA polymerase. The C-terminal domain (CTD) of LexA is involved in dimerization and the N-terminal domain (NTD) in DNA binding. Intact LexA dimerizes by the CTD, and binds to DNA via a helix–turn–helix in its NTD (**Figure 2**).

LexA binding motifs are conserved in many Gram-negative bacteria. The consensus DNA target in *E. coli* is a palindromic dyad taCTGT-(at)₄-ACAGta and is designated the LexA box or

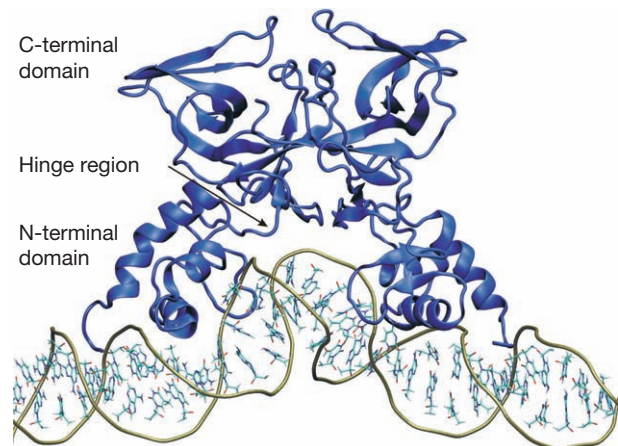


Figure 2 Model of the *E. coli* LexA repressor bound to the operator DNA site. LexA dimerises by the carboxy-terminal domain, and interacts with DNA by the amino-terminal domain. The two domains are linked by a flexible hinge region. Adapted from Butala M, Žgur-Bertok D, and Busby SJW (2009). The bacterial LexA transcriptional repressor. *Cellular and Molecular Life Sciences* 66: 82–93, with permission from Springer Basel AG.

SOS box. Functional LexA repressor is a homodimer while intracellular monomer levels are very low. Each of the two symmetrically inverted DNA-binding elements accommodates one LexA subunit. For stable and specific DNA binding, a

conformational change in LexA must occur. Binding to consensus targets with dyad symmetry requires LexA subunit–subunit interactions that enable high specificity and stabilizes interactions with both halves of the DNA duplex.

The LexA box exhibits considerable diversity; thus, no two sequences are alike and LexA binds with different affinities to the various variants enabling differential induction of the LexA regulon genes. The location of SOS boxes at promoters varies with respect to the transcription start site; some are positioned between the –35 and –10 elements, some overlap with the promoter elements, while others are adjacent to the target promoter. Although most *E. coli* LexA regulon genes possess a single LexA operator site, the number can range up to three SOS boxes. For example, the promoter region of the *lexA* gene carries separated tandem operators. LexA autoregulation sets a control of its own intracellular level via a feedback mechanism, enabling a rapid response to even small amounts of the inducing signal.

Triggers of the SOS Response

SOS genes can be induced by diverse exogenous treatments such as irradiation or chemicals, and can also be induced by DNA damage, caused by metabolic intermediates within the cell, by stalled replication forks, or by defects following recombination or chromosome segregation. Physical stress, such as high pressure that induces activity of the type IV restriction endonuclease, and even certain antibiotics, most notably fluoroquinolones such as ciprofloxacin, are also known to induce the SOS response. Note that the SOS-inducing signal is persistent regions of ssDNA that are generated when growing cells attempt to replicate damaged DNA. Depending upon the nature of the inducing signal, either the RecBCD or the RecFOR complex expose ssDNA to RecA.

The SOS response can also be triggered independently of RecA at low intracellular pH when LexA forms aggregates, which results in induction of LexA-repressed genes. Transient

failure of pH homeostasis occurs in *E. coli* upon shifts of extracellular pH or in mutants with improper intracellular pH regulation. Presumably, this is a bacterial survival strategy when crossing the gastric acid barrier.

Sensing the Signal and Inducing LexA Inactivation

The major SOS-inducing signal is the accumulation of ssDNA. During normal growth a limited amount of ssDNA is tolerated; however, above this threshold, the SOS system is induced in a LexA-dependent manner. Long-lived ssDNA is protected and stabilized by the ssDNA-binding (SSB) protein. Tetrameric SSB migrates along ssDNA, transiently melting short DNA hairpins and stimulating RecA filament elongation on DNA. Association of ATP-liganded RecA protomers constitutes an activated nucleoprotein filament (RecA*) (Figure 3). RecA-mediated SOS induction requires an extended filament conformation but no ATP hydrolysis (note that RecA protein besides working as a co-protease and activator of the DNA polymerase V plays a central role in recombination and is involved in a surprising range of other reactions in *E. coli*).

LexA is recognized by proteases only following self-cleavage, when otherwise latent protease recognition signals are exposed in the cleaved fragments. The self-cleavage of LexA generates LexA N- and C-terminal fragments of 83 and 118 amino acids, respectively. The fragments are rapidly degraded by the ClpXP protease and the degradation of the cleaved C-terminal fragment is facilitated by the Lon protease. Proteolysis ensures proper regulation of induction of the SOS response, since the LexA N-terminal fragment, that contains the DNA binding domain, still retains some repressor function.

Insights into the Key Step in the SOS Response

The LexA repressor is stable in normal growing cells, with a half-life of nearly 1 h. *E. coli* contains approximately 1300 LexA

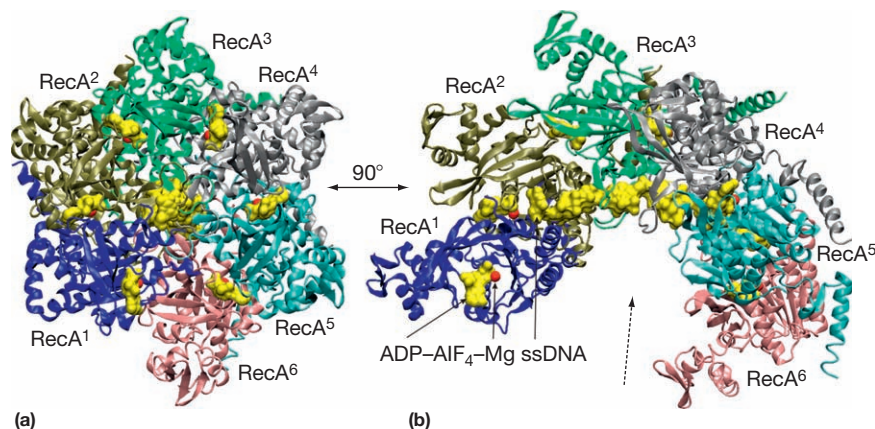


Figure 3 Crystal structure of the active *E. coli* RecA filament (pdb ID: 3CMU), the front (a) and the side view (b). The six RecA protomer monomers (numbered) form a filament on the 18 nt ssDNA (nucleotides are in yellow). ADP–aluminum fluoride–Mg (ADP–AF₄–Mg) is a nonhydrolyzable ATP analog. ADP–AlF₄–Mg is sandwiched between two adjacent RecA protomers (ADP in yellow, Mg in red). Dotted arrow indicates deep helical groove. Figure prepared with visual molecular dynamics (VMD). Reproduced from Humphrey W, Dalke A, and Schulten K (1996) VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 14: 33–38.

molecules. Repressor self-cleavage commences approximately 1 min after exposure to UV and, after 5 min, the level of LexA falls 10-fold. Self-cleavage takes place only after LexA has dissociated from its target, since dimers that are bound at specific operator targets cannot be inactivated.

Upon LexA interaction with the deep helical groove of RecA*, intramolecular cleavage of the repressor occurs. LexA is specifically cleaved at its Ala84–Gly85 bond. John W Little and colleagues proposed a Ser–Lys dyad mechanism for LexA autodigestion. The uncharged form of Lys156 helps remove a proton from the Ser119 hydroxyl group, which then acts as a nucleophile to attack the Ala84–Gly85 bond (Figure 4). *In vivo* cleavage requires RecA but, *in vitro*, it can proceed independently of RecA at alkaline pH (a reaction termed autocleavage).

Crystal structures of LexA mutants revealed that the cleavage site can adopt two conformations. In the cleavable state, the cleavage site is located adjacent to the catalytic center, the Ser119–Lys156 dyad, while in the noncleavable conformation it is ~20 Å away from the active site. It has been suggested that interaction with RecA* induces a conformational change in LexA and deprotonation of Lys156. It was also suggested that RecA* may preferentially interact with and stabilize the LexA cleavable state. However, recent evidence suggests that RecA* can bind to LexA in both the cleavable and noncleavable states. Residue Lys156 is solvent exposed and likely protonated in the LexA noncleavable conformation. The energetic cost of burying the charged group of Lys156, which is required for cleavage, provides another layer of regulation of LexA cleavage and helps to prevent autodigestion. Thus, by acting as a co-protease, RecA inactivates LexA, thereby inducing its expression, together with more than 50 other SOS gene products.

DNA Damage Repair

The level, timing, and duration of expression of each individual LexA regulon genes differ significantly. Most genes of the LexA regulon, including *recA*, are, in the absence of induction, expressed at a basal level. Specifically bound LexA molecules cannot be inactivated, which accounts for the precise timing of expression of the SOS genes following induction. Genes with high-affinity SOS boxes are expressed late in the SOS response due to a persistent decrease in the intracellular LexA pool. On the contrary, selective derepression of SOS genes with weaker operators occurs in response to minor inducing signals.

The SOS response is characterized by temporal control. Initially, SOS products (*recA*, *ssb*) sense DNA damage to protect and maintain the structural integrity of the replication fork. The LexA repressor is also induced immediately. Active RecA* initially signals the upregulation of SOS genes involved in high-fidelity DNA repair. Early induced genes include nucleotide excision repair genes *uvrA*, *uvrB*, *uvrD* that enable single-strand repair catalyzed by the UvrABCD proteins. To facilitate the resumption of processive replication, genes *recA*, *recN*, *ruvAB* of recombinational repair are induced. In order to circumvent lesions that inhibit DNA replication even after enhanced recombinational repair, low-fidelity DNA damage tolerance pathways are induced and DNA polymerases, PolIII (*polB*), PolIV (*dinB*), PolV (*umuC*, *umuD*) that operate in a poorly processive and error-prone manner are synthesized. Their ability to perform translesion DNA synthesis, allows a lethal event to be bypassed and replication to recover. These polymerases are the main contributors to SOS mutagenesis, which is an active process.

Precise temporal modulation of SOS gene expression is coordinated with DNA repair processes and influences many other

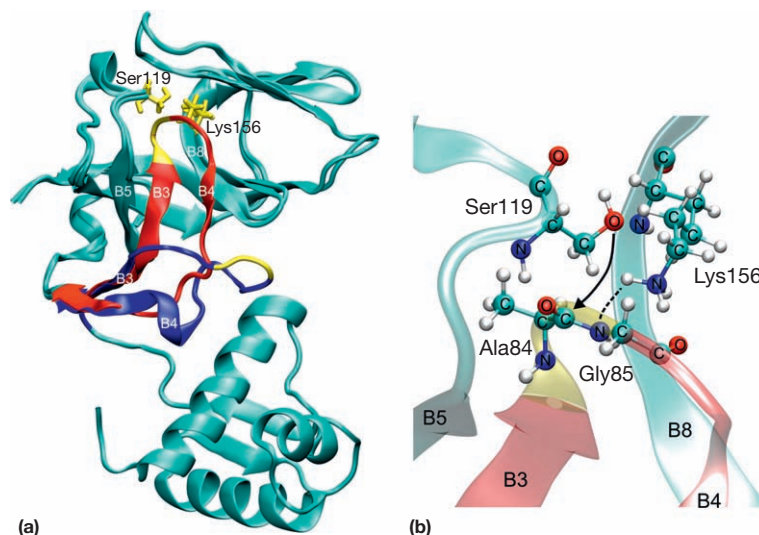


Figure 4 Two distinct conformations of the LexA cleavage site region and a detailed view of the active site. (a) Cleavage site region in the noncleavable state (pdb ID: 1jhh, chain A) is presented in blue and the CTD (pdb ID: 1jhe, chain A) in the cleavable state in red. The catalytic dyad, Ser119 and Lys156, is presented as a stick model and cleavage site Ala84–Gly85 as a ribbon presentation in yellow. (b) Model of the LexA self-cleavage mechanism. Neutral base Lys156 activates the nucleophile LexA119. Hydroxyl group of the activated nucleophile attacks the carbonyl carbon of the scissile peptide bond (arrow), followed by the transfer of the proton to the newly generated amino group (dotted line). The figure was generated by VMD and adapted from Butala M, Žgur-Bertok D, and Busby SJW (2009) The bacterial LexA transcriptional repressor. *Cellular and Molecular Life Sciences* 66: 82–93, with permission from Springer.

cellular processes. Damage inflicted on bacterial DNA leads to fast and massive intracellular coaggregation of RecA and DNA into a lateral macroscopic assembly. These intracellular assemblies are the functional target for DNA repair and are responsible for protection of the cell's DNA heritage.

Cell-Cycle Checkpoints

The expression of SOS genes is turned on in a pattern of discrete activation pulses; therefore, the system is not simply induced and turned off when DNA damage is repaired. To prevent the overlap of cell-cycle processes, the SOS system regulates DNA damage and cell division checkpoints.

E. coli cell-cycle checkpoints are regulated by the *umuDC* and *sulA* gene products. Uncleaved UmuD₂ in complex with UmuC activates a DNA damage replication checkpoint. UmuD₂C inhibits DNA synthesis directly by associating with the DNA replication complex. If high-fidelity repair is insufficient, the UmuD'₂C complex, PolV polymerase, is formed. Following SOS induction, dimeric UmuD is converted to functionally active UmuD' by RecA*-induced self-cleavage that is similar to inactivation of LexA. However, RecA*-mediated self-cleavage of UmuD is much slower than self-cleavage of LexA, providing time for accurate repair prior to recovery of replication by translesion DNA synthesis. The UmuD'₂C complex is activated by interacting with a single RecA-ATP transferred from the RecA* filament. Translesion DNA synthesis by the PolV polymerase enables replication over any remaining DNA lesions.

During the DNA repair process, cell division is inhibited which leads to the formation of cellular filaments. Notably, upon damage to the genome, the LexA-regulated *sulA* gene product is highly expressed and interacts with the FtsZ protein, involved in septum formation prior to cell division. Most likely, this checkpoint serves to delay cell division until DNA damage has been repaired. In addition, by inhibiting cell division the two daughter chromosomes are not separated enabling recombinational repair.

Turning Off the SOS Response

Once DNA damage is repaired and replication resumed, the co-protease activity of RecA disappears resulting in re-accumulation of LexA and repression of the SOS genes. Intracellular proteolysis of SOS gene products is also triggered to control and restrict their activity during the repair and recovery phases of the SOS response respectively.

Members of the LexA Super-Family

Jeffrey W Roberts and colleagues demonstrated that exposure of lysogens containing bacteriophage λ to DNA-damaging treatments results in RecA-mediated cleavage of the λ CI repressor. SOS regulation enables temperate λ -like bacteriophages to sense the physiological condition of the host cell and switch the phage from lysogenic to lytic growth. LexA, UmuD, and several λ CI-like repressors, exhibit CTD homology and undergo completely parallel cleavage reactions in helical

groove of the RecA* filament. Self-cleavage of LexA is intramolecular while UmuD is cleaved in an intermolecular reaction. Note that upon self-cleavage, dimeric UmuD is converted to the functionally active UmuD', in contrast to repressors that are inactivated by cleavage. Remarkably, compared to LexA, RecA* catalyzes slow self-cleavage of the CI repressor and UmuD; hence, prophage induction and mutagenesis are induced only when DNA is severely damaged.

Plasmid-Encoded Genes of the LexA Regulon

Some plasmid-encoded genes, with broader functions than defense against DNA damage and adaptation through mutagenesis, are also part of the LexA regulon. For example, colicins are plasmid-encoded bacteriocins, synthesized by and active against *E. coli* strains and its close relatives. Colicins are released into the environment only after lysis of the host cell. Expression of operons encoding colicin functions are always strongly repressed by LexA, and slow dissociation from the operators may account for the late induction of colicin genes during the SOS response. RecA-mediated production of bacteriocins thus resembles prophage induction, leading to cell lysis upon persistent, high level DNA damage. Many colicins can promote genetic diversity in *E. coli* populations pointing to a role in evolution.

The *qnr* genes, which encode fluoroquinolone-resistance determinants, provide another example of plasmid-borne LexA-repressed genes. These are widespread in Enterobacteriaceae and are all directly regulated by LexA. Since fluoroquinolones induce self-cleavage of LexA, this is the first example of SOS-dependent regulation of an antibiotic-resistance mechanism in response to the antibiotic itself.

Bacterial LexA Regulon Diversity

Although the SOS system is highly conserved among bacteria, the genes controlled by LexA, their regulation and consensus LexA-binding sites differ significantly. In *Bacillus subtilis* LexA regulates 26 operons encompassing 63 genes (note that the *B. subtilis* LexA protein is also designated DinR). In comparison, the *E. coli* LexA regulon comprises 57 genes and has only eight orthologs in *B. subtilis*. To further illustrate the diversity found in SOS networks, in both *Rhodobacter sphaeroides* and the cyanobacterium *Synechocystis* sp., the LexA paralogue can function both to repress and to activate transcription.

The Virulent Side of the SOS Response

Besides high-fidelity repair pathways, SOS genes encode low-fidelity translesion DNA polymerases (in *E. coli*, PolII [*polB*], PolIV [*dinB*], and PolV [*umuC*, *umuD*]) that enable bacteria to increase their mutation rate in times of stress. Studies employing therapeutic drugs showed that low or subinhibitory concentrations of certain antibiotics, that interfere with DNA replication as well as cell wall synthesis, can trigger the SOS response. Hence, antibiotics can accelerate evolution by, for example, the acquisition of point mutations that result in inactivation or efflux of the drug.

SOS-inducing antibiotics also affect virulence in several pathogenic bacteria. Antibiotics that activate RecA*-mediated inactivation of LexA also trigger self-cleavage of phage repressors of resident prophages in *E. coli*, *Vibrio cholerae*, and *Staphylococcus aureus*. Consequently, certain antibiotics promote the horizontal spread of temperate phage and associated pathogenicity islands. In addition, the lateral transfer of integrating conjugative elements, for example, the *V. cholerae* SXT element encoding antibiotic resistance, can be induced. Thus, SOS-induced mobilization and high-frequency horizontal transfer of DNA elements accelerate the spread of virulence factors and drug resistance genes. In *E. coli*, induction of the LexA regulon has been shown to be required for the acquisition of resistance to ciprofloxacin and rifampicin. In addition, recombination of integrons, genetic elements capable of incorporating and expressing promoterless genes, was shown to be controlled by the SOS response.

Cells in a bacterial population can survive antibiotic stress by forming dormant cells, designated as persisters that are highly tolerant to antibiotics. Persisters are not mutants but rather phenotypic variants of sensitive cells. Recently, a small membrane-acting peptide encoded by the LexA-regulated gene, *tisB*, was suggested to control persister formation.

Distinct from drug-induced mobilization of DNA elements, the SOS system also induces chromosomal virulence gene expression. For example, prophage encode the *E. coli* Shiga toxin. In enteropathogenic *E. coli*, SOS regulates a type III secretion system responsible for secretion of virulence-associated factors into host cells. Interestingly, in some *S. aureus* strains, a

LexA-regulated gene encodes the fibronectin binding protein (FnBB) that mediates tissue attachment and the establishment of infection.

See also: **Cell Architecture and Function:** Cell Cycle: DNA Damage Checkpoints; **Molecular Biology:** DNA Damage: Alkylation; DNA Mismatch Repair and the DNA Damage Response; Nucleotide Excision Repair, Bacterial: The UvrABCD System; UmuC D Lesion Bypass DNA Polymerase V.

Further Reading

- Butala M, Žgur-Bertok D, and Busby SJW (2009) The bacterial LexA transcriptional repressor. *Cellular and Molecular Life Sciences* 66: 82–93.
- Courcelle J, Khodursky A, Peter B, Brown PO, and Hanawalt PC (2001) Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. *Genetics* 158: 41–64.
- Erill I, Campoy S, and Barbe J (2007) Aeons of distress: An evolutionary perspective on the bacterial SOS response. *FEMS Microbiology Reviews* 31: 637–656.
- Foti JJ, Simmons LA, Beuning PJ, and Walker GC (2010) Signal transduction in the *Escherichia coli* SOS response. In: Bradshaw RA and Dennis EA (eds.) *Handbook of Cell Signaling*, vol. 3, pp. 2127–2136. New York, NY: Elsevier.
- Kelley WL (2006) Lex marks the spot: The virulent side of SOS and a closer look at the LexA regulon. *Molecular Microbiology* 62: 1228–1238.
- Little JW (1991) Mechanism of specific LexA cleavage: Autodigestion and the role of RecA coprotease. *Biochimie* 73: 411–421.
- Luo Y, Pfuetzner RA, Mosimann S, et al. (2001) Crystal structure of LexA: A conformational switch for regulation of self-cleavage. *Cell* 106: 585–594.
- Sassanfar M and Roberts JW (1990) Nature of the SOS-inducing signal in *Escherichia coli*. The involvement of DNA replication. *Journal of Molecular Biology* 212: 79–96.

Ligand-Operated Membrane Channels: Calcium (Glutamate)

E K Michaelis, University of Kansas, Lawrence, KS, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Ionotropic glutamate receptors This general class of glutamate receptors represents ligand-gated, ion-channel-forming receptor complexes in neuronal membranes. Channel activation occurs following the application of L-glutamate or one of its analogs. This class of receptors is further subdivided into three families of receptor-ion channels based on their sensitivity to select agonists.

Long-term depression (LTD) Synaptic plasticity that occurs following low-frequency stimulation of synapses and which manifests itself as a decreased response to a subsequent single stimulus applied to the same synapses.

Long-term potentiation (LTP) Synaptic plasticity that occurs following high-frequency stimulation of synapses and which manifests itself as an increased response to a subsequent single stimulus applied to the same synapses.

Metabotropic glutamate receptors A class of G-protein-coupled glutamate receptors in neuronal membranes. These receptors may either activate phospholipase C or inhibit adenylyl cyclases, and they do not directly activate ion channels. This class of receptors is further subdivided into three groups, referred to as groups I–III, based on their sensitivity to select agonists and antagonists.

Synaptic plasticity The adaptive changes in neurotransmission across synapses occurring as a result of repeated stimulation of a set of synapses. The changes in the electrical response of a set of synapses to a single stimulus applied after the delivery of a series of high-frequency repetitive stimuli can last for many hours or many days. These changes in synaptic efficacy are accompanied by alterations in gene transcription, protein synthesis and enzyme, ion channel, and receptor activity.

The mammalian central nervous system (CNS) has a great capacity to adapt to many stimuli and perturbations of normal physiological function and this adaptability or plasticity is crucial in coding, storing, and retrieving information. Over the past decade, it has become clear that transient elevations of intracellular calcium (Ca^{2+}) in neurons represents the key signals used by nerve cells to alter their molecular structure and function and to produce long-lasting changes in cell function. These Ca^{2+} -activated molecular events are crucial not only for phenomena such as learning and memory but also for determining neuronal survival or damage. Key regulators of Ca^{2+} entry into neurons are the receptors for the excitatory neurotransmitter, L-glutamic acid.

Calcium Signaling in Neurons

Adaptation of brain function to incoming stimuli or to altered behavioral states is a well-demonstrated phenomenon. Such adaptation is thought to involve long-lasting alterations in the sensitivity of select neuronal networks or select synapses of individual neurons. The phenomenon of long-lasting changes in synaptic activity in neurons is, in part, the result of altered gene transcription and protein synthesis. Changes in the concentration of free intracellular calcium (Ca^{2+}) brought about either through influx into nerve cells or release from intracellular stores are considered to be the major signals for such adaptive changes in neurons. Under normal physiological conditions in unstimulated neurons, the concentration of free intracellular Ca^{2+} ($[\text{Ca}^{2+}]_i$) is maintained at ~45–75 nM, that is, a concentration that is ~10 000-fold lower than that in the extracellular environment of neurons. Following the stimulation of neurons, through either synaptic

activation, primarily of glutamate receptors (but also of other receptors, such as nicotinic acetylcholine receptors (AChRs)), or direct electrical stimulation, the intracellular levels of free Ca^{2+} can change quite rapidly and substantially, as described below (Figure 1).

The low free $[\text{Ca}^{2+}]_i$ is maintained through buffering mechanisms that are dependent on the activity of both Ca^{2+} -transport systems and Ca^{2+} -binding proteins present in neurons. Both of these Ca-buffering processes are important in reestablishing low free $[\text{Ca}^{2+}]_i$ following perturbations of the concentrations after synaptic activation and membrane depolarization of neurons. Among the buffering mechanisms, the transport carriers either pump Ca^{2+} out of the intracellular environment, such as the plasma membrane ($\text{Ca}^{2+} + \text{Mg}^{2+}$)-ATPase (PMCA) and the $\text{Na}^+ - \text{Ca}^{2+}$ exchange carriers (NCX), or sequester Ca^{2+} in intracellular compartments, such as the endoplasmic reticulum, through the activity of the sarcoplasmic/endoplasmic reticulum Ca-ATPase (SERCA), and the mitochondria, through the activity of the mitochondrial Ca^{2+} uniporter (Figure 1). Calcium-binding proteins (Ca^{2+} BP) that have been shown to buffer intracellular Ca^{2+} include parvalbumin and calbindin as the primary Ca^{2+} -buffering proteins in neurons, as well as many other Ca^{2+} -binding proteins, such as calmodulin, calyphosine, neurocalcin, hippocalcin, and visinin-like protein 3.

Electrical stimulation of neurons with a stimulus of sufficient magnitude to depolarize the cell membrane from the normal electrical potential of ~−70 mV internal negative to a potential that is 10–20 mV or more, less negative internally, can trigger the initiation of a regenerative, self-propagating form of membrane depolarization that usually starts in the region where the axon process of a neuron emerges from the cell body and travels down to the fine terminal branching

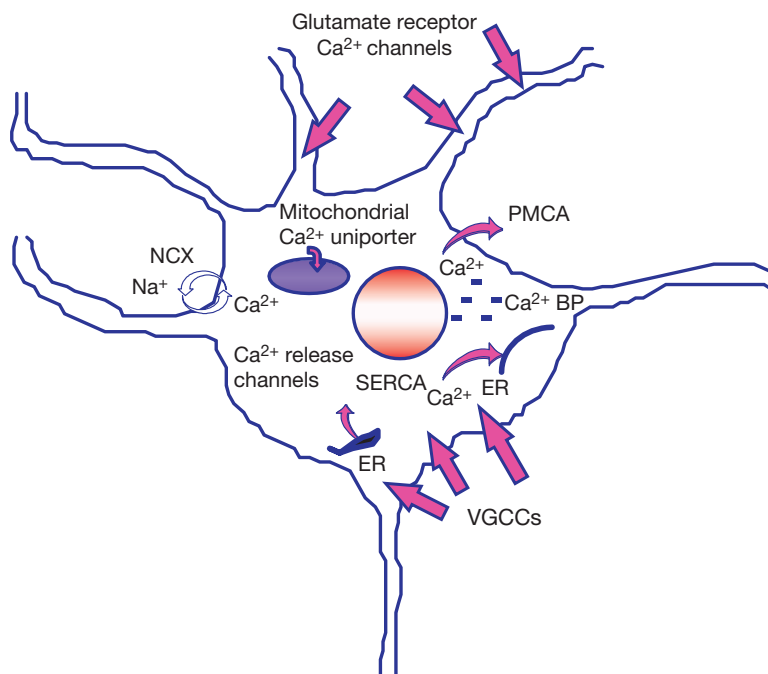


Figure 1 Pathways of Ca^{2+} entry, extrusion, and sequestration in neurons.

points of the axon (Figure 2). This self-propagating depolarization, known as an 'action potential', depends on the rapid opening of voltage-gated sodium channels (VGSCs).

Depolarization of the nerve terminal region of a neuronal axon by an action potential leads also to the opening of voltage-gated Ca^{2+} channels (VGCCs) in the axon terminal region. VGCCs allow for the rapid entry of Ca^{2+} from the extracellular environment, an event that is crucial for the first step in the transformation of the electrical event that has traveled down the axon to a chemical-signaling event, that is, the release of chemical transmitters. The molecular events that lead to Ca^{2+} -induced exocytosis of transmitter molecules stored within synaptic vesicles at the nerve terminal are progressively being defined through biochemical, biophysical, and genetic studies but are not discussed further here.

Action potentials initiated at the beginning of the axon as it exits the cell body spread in two directions, toward the axon terminal as described above, as well as toward the cell body and the other processes emanating from the cell body of a neuron, the dendrites (Figure 2). Dendrite processes are considered to be the primary site for the reception of chemical information transmitted by means of neurotransmitters released from hundreds or thousands of axon terminals impinging upon the dendrites or the soma of a single neuron. The contacts that nerve terminals make with either dendrites or the soma of a neuron are, of course, the synapses across which chemical neurotransmission takes place. Most synapses on dendrites of brain neurons occur through close approximation of nerve terminals with specialized, small processes protruding from the main shaft of a dendrite (Figure 2). These knob-like protrusions are referred to as 'dendritic spines' and represent a specialized compartment

in a neuronal dendrite. The signals initiated by Ca^{2+} entry through transmitter receptor-ion channels and back-propagating action potentials in a spine differ from those initiated in the main body of a dendrite. Therefore, integration of signals that lead to long-lasting changes in neuronal responses to incoming stimuli and to changes in gene transcription and protein synthesis is determined by signals initiated in spines as well as those in dendrites.

The narrowness of the neck of spines that connects them to the respective dendrite is thought to delay diffusion and equilibration of intra-spine Ca^{2+} with that in the dendrites. Thus, the levels of free $[\text{Ca}^{2+}]_i$ within the head of a spine are determined not only by the amount of Ca^{2+} entering the spine through receptor-activated and voltage-gated ion channels but also through the kinetics of diffusion out of the spine and the Ca-buffering capacity within the spine. Action potentials initiated in axons and invading the cell body and dendrites produce supralinear increases in free $[\text{Ca}^{2+}]_i$ in dendrites when they are coincident with synaptic activation of neurotransmitter receptor-associated Ca^{2+} channels. Supralinear increases in $[\text{Ca}^{2+}]_i$ are defined as increases that exceed the sum of net increases in free $[\text{Ca}^{2+}]_i$ brought about by the simultaneous activation of both neurotransmitter receptor channels and VGCCs. Free $[\text{Ca}^{2+}]_i$ in spines following the activation of back-propagating action potentials has been estimated to reach levels as high as $2\text{ }\mu\text{M}$. If back-propagating potentials are coupled to neurotransmitter activation of receptors that allow the influx of Ca^{2+} into neurons, such as glutamate receptors, then the levels are estimated to be more than 10-fold higher, that is, $26\text{ }\mu\text{M}$.

The coincidence of glutamate receptor-mediated and action potential-induced Ca^{2+} signals plays a crucial role in the phenomenon known as 'synaptic plasticity', that is, the

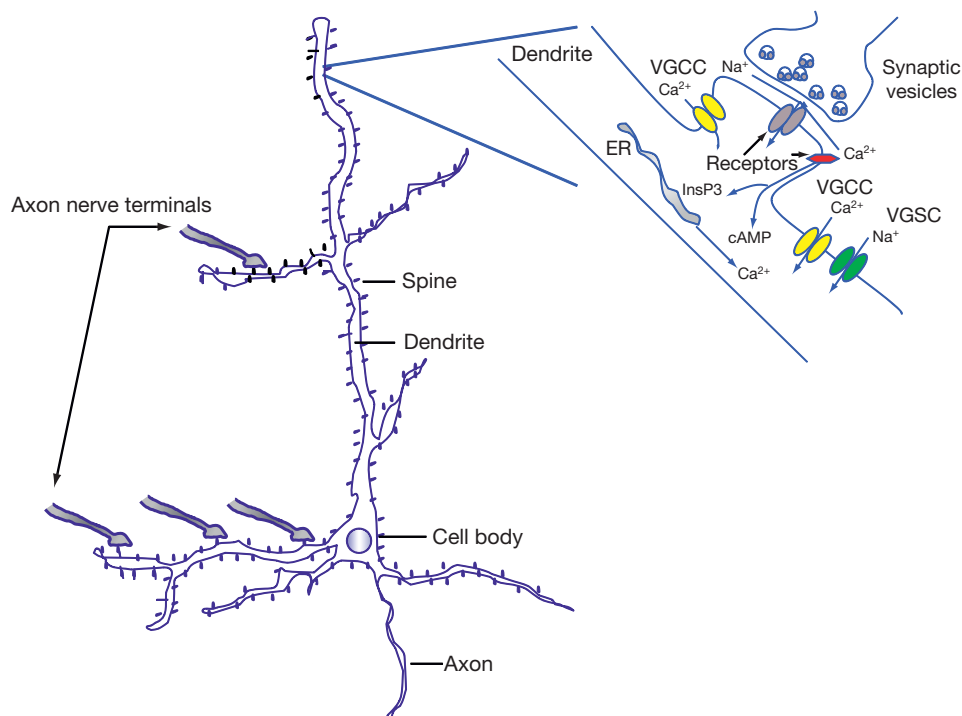


Figure 2 Diagrammatic representation of a large neuron that might be found in the neocortex or hippocampus of the brain. The structures of dendrites, dendritic spines, and axon terminals are highlighted.

long-term change in the magnitude of responses across a synapse following a series of activating stimuli delivered to a set of synapses in a neuron. Synaptic plasticity is fundamental to the processes of learning and memory and to the adaptive brain changes described above. Long-term increases in synaptic responses to a given stimulus are known as 'long-term potentiation' (LTP), and long-term decreases in such responses are known as 'long-term depression' (LTD). At the center of the molecular events that lead to synaptic plasticity are glutamate receptors, VGCCs, and intraneuronal Ca^{2+} signaling.

Glutamate Receptors in the Mammalian CNS: Structure and Function

Physiological and Pharmacological Properties of Glutamate Receptors

L-Glutamic acid is the most prevalent excitatory neurotransmitter in the vertebrate nervous system. Glutamate receptor activation is one of the stimuli for neuronal migration and synapse formation in the developing CNS, causes the elongation of neuritic processes, and enhances the survival of some neurons. However, excessive stimulation of glutamate receptors by moderately high concentrations of L-glutamate and several of its analogs, such as *N*-methyl-D-aspartate (NMDA), leads to delayed neuronal damage through the activation of apoptosis. Stimulation of glutamate receptors by even higher concentrations of L-glutamate or NMDA leads to acute toxicity through necrosis. Both the phenomena of synaptic plasticity and neuronal damage are dependent on the entry of Ca^{2+} into neurons

but they differ, of course, in the magnitude and duration of the changes in free $[\text{Ca}^{2+}]_i$.

Before proceeding with a discussion of the regulation of Ca^{2+} signaling in neurons by glutamate receptors, it is important to define the various types of glutamate receptors present in the mammalian CNS. The classification is based on pharmacological and physiological properties of neuronal glutamate receptors (Table 1). On the basis of functional characteristics, two general categories of glutamate receptors have been identified, the ion-channel-forming receptors or 'ionotropic receptors', and the receptors linked to the activation of phospholipase C (PLC) or the inhibition of adenylyl cyclases (ACs), known as the 'metabotropic receptors'. Metabotropic receptors that activate PLC also stimulate phospholipase D. Ionotropic receptors fall into three categories: those that are activated by NMDA, those that respond to kainic acid (KA), and those that are activated by α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) (Table 1). Metabotropic glutamate receptors are subdivided into groups I, II, and III based on their sensitivity to various agonists and antagonists, as well as their effects on either PLC or ACs.

KA and AMPA receptors (KARs and AMPARs) are the primary receptors for rapid excitatory transmission in the CNS. The receptor-associated ion channels activated by AMPA or KA are primarily permeable to Na^+ and K^+ . However, some AMPARs as well as some KARs form ion channels that are permeable to Ca^{2+} . AMPARs in brain neurons are activated by both AMPA and KA, but AMPA is the more potent agonist. AMPARs can be distinguished from KARs on the basis of the effects that selective allosteric modulators, the benzothiazides and AMPAkinases, and selective noncompetitive

Table 1 Classification of L-glutamate receptor subtypes found in the mammalian central nervous system

Receptor classes	Inotropic	Metabotropic		
Receptor subtype	NMDA	AMPA	Kainate	Quisqualate, L-AP4
Selective agonists (plus co-agonists)	NMDA (plus glycine or D-serine)	AMPA	Kainate	1S3RACPD, L-AP4LCCG1, LY354740, LY404039, DCG IV, ACPT III
Functional characteristics	Activation of channels for Na ⁺ , K ⁺ , and Ca ²⁺ ; slow excitation, slow desensitization	Activation of channels for Na ⁺ and K ⁺ , some Ca ²⁺ ; fast excitation, fast desensitization	Activation of channels for Na ⁺ and K ⁺ ; fast excitation, fast desensitization	Activation of PLC; activation of PLD; inhibition of AC
Other agonists	Ibotenate, L-HCA, quinolinate	(S)-ATPA, quisqualic acid, kainate, domoate, willardine	(2S, 4R)-4-methylglutamic acid, dysiherbaine, domoate, (S)-ATPA, acromelic acids A and B	Quisqualate, L-serine-O-phosphate, ibotenate, CCG
Allosteric modulators	Spermine, ifenprodil	Aniracetam, benzothiazides, CX164 (AMPAkine)	Concanavalin A (Con A)	
Competitive antagonists	2-AP5, 2-AP7, CPP, CPP-ene, CGP39653, CGS19755, biphenyl	CNQX, NBQX, DNQX	GAMS, γ-D-glutamyl-glycine	Phenylglycine analogs (3HPG, 4CPG, 4C3HPG, MCPG)
Antagonists at co-agonist or modulator sites	5,7-diCl-Kyn, HA-966, CNQX, L-689560, ifenprodil	2,3-Benzodiazepines (GYKI52466), SYM2206		
Ion channel inhibitors	PCP, MK-801, ketamine, memantine	JST, barbiturates		

AC, adenyl cyclase; ACPD, 1-aminocyclopentane-1,3-dicarboxylate; Aniracetam, 1-(4-methoxybenzoyl)-2-pyrrolidinone, L-AP4, L-2-amino-4-phosphonobutanoic acid; 2-AP5, D-2-amino-5-phosphonopentanoic acid; 2-AP7, D-2-amino-7-phosphonoheptanoic acid; (S)-ATPA, (S)-2-amino-3-(5-*tert*-butyl-3-hydroxy-4-isoxazolyl)propionic acid; CCG, 2-carboxycyclopropylglycine; CGP39653, [(±)-(E)-2-amino-7-amino-4-propyl-5-phosphonopentenoic acid; CGS19755, [(±)-2-carboxypiperidin-4-yl]methyl]-phosphonic acid; 4C3HPG, (S)-4-carboxy-3-hydroxy-phenylglycine; CNQX, 6-cyano-7-nitroquinoxaline-2,3-dione; 4CPG, (S)-4-carboxy-phenylglycine; CPP, [3-[(±)-carboxypiperazin-4-yl]prop-1-yl]-phosphonic acid; CPP-ene, [3-[(±)-2-carboxypiperazin-4-yl]-propen-1-yl]-phosphonic acid; CX614, 2H,3H,6aH-pyrrolidino[2'',1''_3',2']1,3-oxazino[6',5'-5,4]benzo[*a*]1,4-dioxan-10-one; 5,7-di-Cl-Kyn, 5,7-dichlorokynurenic acid; DNQX, 6,7-dinitroquinoxaline-2,3-dione; GAMS, γ-D-glutamylamino-methylsulfonate; HA-966, 3-amino-1-hydroxy-2-pyrrolidone; 3HPG, (S)-3-hydroxy-phenylglycine; JST, Joro spider toxin; L-689,650, *trans*-2-carboxy-5,7-dichloro-4-phenylaminocarbonylamino-1,2,3,4-tetrahydroquinoline MCPG, (+)-α-methyl-4-carboxy-5,7-dichloro-4-phenylaminocarbonylamino-1,2,3,4-tetrahydroquinoline; LY354740, (1S,2S,5R,6S)-2-aminobicyclo[3.1.0]hexane-2,6-dicarboxylate monohydrate; LY404039, (−)-(1*R*,4*S*,5*S*,6*S*)-4-amino-2-sulfonylbicyclo[3.1.0]hexane-4,6-dicarboxylic acid; MCPG, (+)-α-methyl-4-carboxy-phenylglycine; MK-801, (+)-5-methyl-10,11-dihydro-5H-dibenzo[*a,d*]cyclohepten-5,10-imine; NBQX, 2,3-dihydroxy-6-nitro-7-sulphamoyl-benzo-(F)-quinoxaline; PLC, phospholipase C; SYM 2206, (±)-4-(4-aminophenyl)-1,2-dihydro-1-methyl-2-propylcarbamoyl-6,7-methylenedioxyphthalazine.

inhibitors, the 2,3-benzodiazepines, have on AMPARs but not on KARs (Table 1). Some of the selective modulators of AMPARs, the AMPAkinases, were explored as agents that might enhance memory formation in diseases that produce decrements in memory, such as the dementia of Alzheimer's disease.

NMDA receptors (NMDARs) respond to glutamate in the presence of the co-agonist glycine and such responses are slower than those following activation of either KARs or AMPARs. The slow response of NMDARs is due to the fact that they are tonically inhibited by Mg²⁺. Mg²⁺ inhibition of NMDARs is a voltage-dependent phenomenon and membrane depolarization relieves such blockade. Thus, the rapid activation of KARs and AMPARs by glutamate and the subsequent depolarization of the membrane diminishes the inhibition of NMDARs by Mg²⁺ and leads to the activation of these receptors by glutamate. In neuronal synapses on dendritic spines, both AMPA- and NMDA-sensitive receptors are proximal to each other. Back-propagating action potentials in dendrites or action potentials generated within spines and dendrites also relieve the Mg²⁺-induced inhibition of NMDAR channels.

The opening of NMDAR ion channels increases the permeability of the membrane to Na⁺, K⁺, and Ca²⁺. A unique feature of NMDARs is the requirement for the presence of nanomolar to micromolar concentrations of the co-agonist glycine in order to be fully activated (Table 1). Because of the multiplicity of sites that activate or inhibit NMDAR ion channels, antagonists for NMDARs fall under four categories (Table 1): antagonists that compete for the recognition sites of L-glutamate and NMDA (e.g., 2-AP5, 2-AP7), those that block the binding sites for the coagonist glycine (e.g., 5,7-diCl-kynurenic acid), those that inhibit the allosteric enhancement of agonist activation by the polyamines (e.g., ifenprodil), and those that inhibit the receptor-associated ion channel (e.g., dizocilpine or MK801).

Metabotropic glutamate receptors (mGluRs) are localized primarily in what is known as the perisynaptic region of dendritic spines, that is, not within the synapse as AMPARs, KARs, or NMDARs do. These receptors are apparently not activated by the release of neurotransmitter glutamate following single-action potentials but rather following repeated, high-frequency stimulation of axons and their nerve endings. Of particular interest with regard to Ca²⁺ signaling in neurons is the fact

that group I mGluRs produce two types of neuronal depolarization, a rapid, transient depolarization related to the release of Ca^{2+} from intracellular stores and a prolonged and larger depolarization resulting from activation of transient receptor potential Ca^{2+} channels. The release of Ca^{2+} from intracellular stores is the result of inositol-1,4,5-triphosphate (InsP_3) formation and the activation of InsP_3 receptor-ion channels on endoplasmic reticulum membranes. This transient elevation in intracellular Ca^{2+} causes propagating increases in free $[\text{Ca}^{2+}]_i$ and elicits changes in the activity of CNS neurons that are dependent on protein synthesis, that is, it has some of the characteristics of long-term changes associated with synaptic plasticity. In addition, these Ca^{2+} transients can produce supralinear increases in free $[\text{Ca}^{2+}]_i$ in dendrites and spines

of CNS neurons if they occur at the same time as back-propagating action potentials.

Molecular Structure of Glutamate Receptors

The cloning of the complementary DNAs (cDNAs) for glutamate receptors has revealed a great complexity in the structure of these receptors (Table 2 and Figures 3 and 4). The size of all glutamate receptor proteins represented by the cDNAs for GluR1, NMDAR1, mGluR1, and their homologs (95–163 kDa) is larger than that for other neurotransmitter receptor proteins, both ion-channel-forming receptors and G-protein-coupled receptors (Table 2 and Figures 3 and 4). Based on the amino acid sequence of the proteins for the ionotropic receptors

Table 2 Characteristics of the cloned glutamate receptors, including some putative receptors with ligand-binding characteristics similar to glutamate receptors, from mammalian and amphibian brain

Subunits	Receptor type	Protein size (kDa)	Protein assembly	Functional properties
GluR1	AMPA (KA)	99.8	GluR1–4; most native receptors contain GluR 2	Glutamate, AMPA, KA-activated channels; low Ca^{2+} permeability
GluR2		96.4		
GluR3		98		
GluR4		101		
GluK1	KA	102.8	GluK1 with GluK2–GluK5	Glutamate and KA-activated channels; rapidly desensitizing
GluK2		93.9		
GluK3		100		
GluK4		105		
GluK5		109		
KBP-chick	KA binding	51.8	Single subunit	No channel function
KBP-frog		52.5		
$\delta 1$	Orphan receptor	110.4	Unknown	No channel function
$\delta 2$		113.2		
NMDAR1a	NMDA	103.5	With NMDAR2 or NMDAR3	NMDA-activated channels; glycine a co-agonist high Ca^{2+} permeability, MK-801 and ketamine block; voltage-dependent Mg^{2+} block
NMDAR2A		163.3	With NMDAR1 and other NMDAR2	
NMDAR2B		162.9	With NMDAR1 and other NMDAR2	
NMDAR2C		133.5	With NMDAR1 and other NMDAR2	
NMDAR2D		140.6	With NMDAR1 and other NMDAR2	
NMDAR3A		124.5	With NMDAR1	Glycine-activated excitatory responses; Na^+ , K^+ channels; not permeable to Ca^{2+} ; no Mg^{2+} blockade
NMDAR3B		109	With NMDAR1	Same as for NR-3A
GBP	NMDAR-like complex	57	With GlyBP, TCP-BP, CPP-BP,	Reconstituted complex forms NMDA-activated ion channels; glycine a co-agonist, MK-801 and ketamine inhibit; low sensitivity to Mg^{2+}
GlyBP	NMDAR-like complex	52.7	With GBP, TCP-BP, CPP-BP	
CPP-BP	NMDAR-like complex	78.3	With GBP, GlyBP, TCP-BP	
TCP-BP	NMDAR-like complex	8.9 & 9.5, Bicistronic	With GBP, GlyBP, CPP-BP	
mGluR1a	Metabotropic GluRs	133.2	Single protein receptor; coupled to G proteins	mGluR1, mGluR5: activation of PLC and PLD
mGluR2		95.8		mGluR2, 3, 4, 6, 7, 8: inhibition of adenylyl cyclases
mGluR3		99		
mGluR4a		101.8		
mGluR5a		128.3		
mGluR6		95.1		
mGluR7		102.2		
mGluR8		97.5		

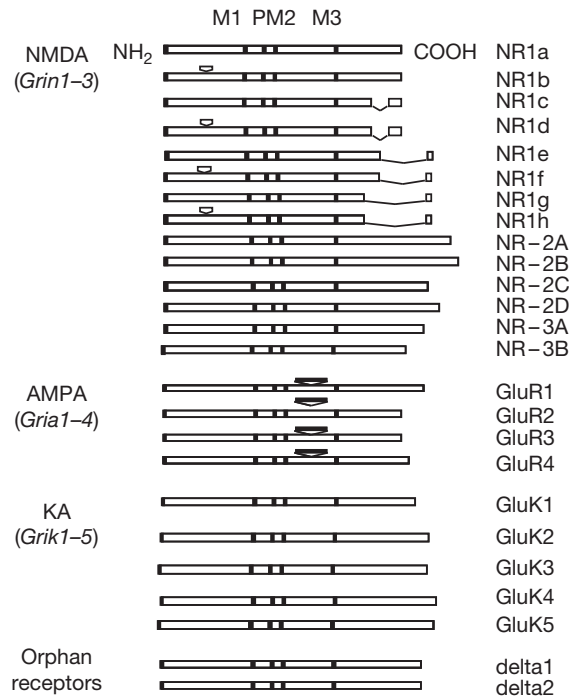


Figure 3 Linear representation of the structure of the cloned ionotropic glutamate-receptor proteins, including two orphan receptors. Shown are the NMDA, AMPA, and KA receptor proteins, as well as some of the orphan receptor proteins, and the variant forms of NMDAR1 that result from alternative RNA splicing. Such alternative splicing leads to the addition or deletion of amino acid residues at either the amino (NH₂) or carboxyl (COOH) terminal of the protein. All proteins are shown to contain three hydrophobic domains (designated M1–M3) and the P loop at the pore region (black boxes). For the proteins GluR1–4, the presence of the flip and flop splice variants in the region between M2 and M3 is indicated by boxes above the diagram of each receptor protein. Modified from Michaelis EK (1998) Molecular biology of glutamate receptors in the central nervous system and their role in excitotoxicity, oxidative stress and aging. *Progress in Neurobiology* 54: 369–415.



Figure 4 Linear representation of the structure of the cloned metabotropic glutamate receptor proteins mGluR1–8, including several splice variants. The seven hydrophobic domains considered as the likely transmembrane regions of the proteins are designated as M1–M7 and are shown as black boxes. Modified from Michaelis EK (1998) Molecular biology of glutamate receptors in the central nervous system and their role in excitotoxicity, oxidative stress and aging. *Progress in Neurobiology* 54: 369–415.

and the relative hydrophobicity of various domains in the proteins, it was predicted that each ionotropic receptor protein cloned had a large extracellular amino-terminal region and three transmembrane domains, M1–M3. Ionotropic glutamate receptors are macromolecular complexes composed of four subunits. Identification of the agonist-binding sites of the GluR and GluK subunits was initially achieved by analysis of chimeras made of GluR3, an AMPAR protein, and GluK2, a KAR protein. The last 150 amino acids before the M1 domain and the amino acids right after M2 in the extracellular loop between M2 and M3 are crucial in determining ligand-binding selectivity of these proteins. Site-specific mutagenesis and crystallographic studies of the extracellular domain have led to the characterization of the agonist-binding site as one that is formed in a bi-lobed manner by the S1 and S2 domains (Figure 5). The S1 and S2 domains are also the sites of binding of allosteric modulators that prevent receptor desensitization, for example, the binding site for cyclothiazide to AMPARs.

On the intracellular side of the plasma membrane, the domain between M1 and M2 forms a loop at the opening of the ion channel but does not traverse the membrane bilayer. This is the P domain (Figure 5). Structurally, the ion-channel-forming region of KARs, AMPARs, and NMDARs exhibits high homology to voltage-gated K⁺ channels, especially the P loop at the opening of the ion channel. Single amino acid substitutions in the P loop affect the permeability of the channel to Ca²⁺ dramatically.

AMPA Receptor Genes and Protein Structure

With regard to the structure of native AMPARs, homomers of any of the proteins GluR1–4 form ion channels that are activated by both AMPA and kainate and exhibit the characteristics of neuronal AMPARs. GluR1–4 receptor proteins (*Gria1–4* corresponding genes) are expressed broadly throughout the CNS. Forms of AMPARs that are derived from alternatively spliced exonic sequences in the respective messenger RNAs (mRNAs) are also expressed in the CNS (Figure 3). With regard to Ca²⁺ permeability of the ion channels associated with AMPARs, the key subunit is the GluR2 subunit. When this subunit is present in a receptor complex, then the AMPARs exhibit a linear relationship between voltage applied to the membrane and the current conducted through the receptor channels and have very low permeability to Ca²⁺, that is, they resemble most native AMPA receptors of CNS neurons. Homomeric complexes made of the GluR1, GluR3, or GluR4 proteins form channels that are permeable to Ca²⁺. The restriction of Ca²⁺ permeability imparted to receptor complexes by the presence of the GluR2 subunit is dependent on a crucial Arg586 (R586) in the P loop of this subunit. The corresponding residue in GluR1 is Gln582 (Q582) (Figure 5), and mutation of this Gln residue to Arg abolishes the Ca²⁺ permeability of channels formed by homomeric expression of the GluR1 protein. GluR3 and GluR4 also have Gln residues in the P loop.

The Arg residue in the P loop of GluR2 is the result of RNA editing. The gene codon for the amino acid at this position is cytosine–adenine–guanine (CAG) for Gln, not cytosine–guanine–guanine (CGG) for Arg. The pre-messenger

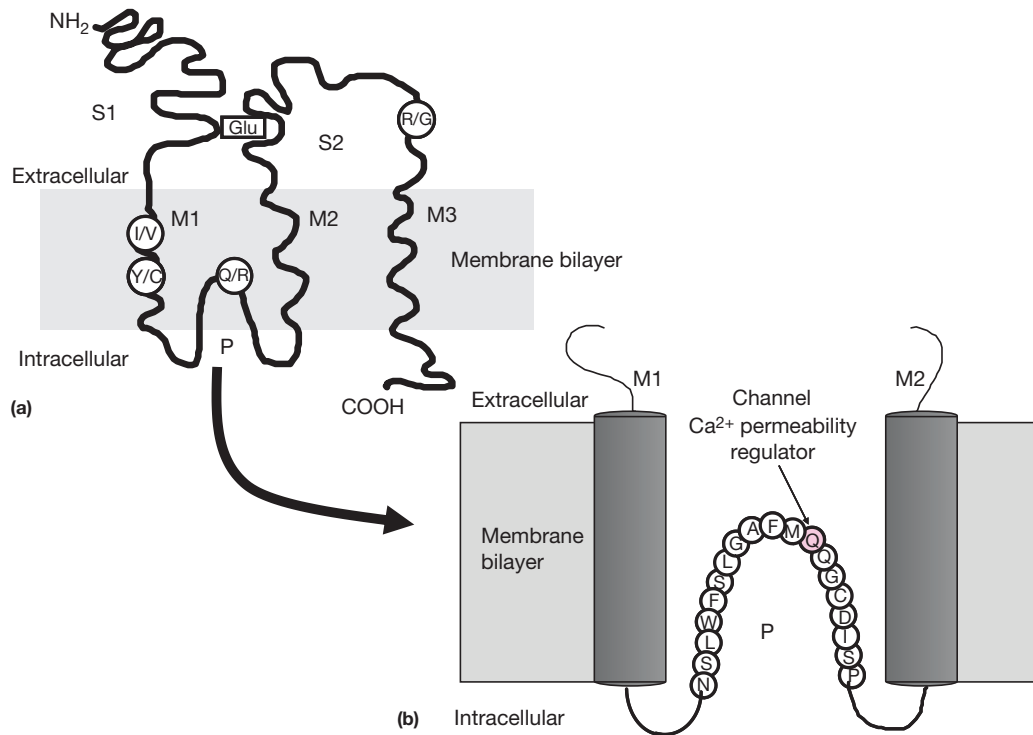


Figure 5 A representation of the distribution of different domains of the AMPA receptor (GluR proteins) in the neuronal membrane (a) and an enlarged view of the formation of a hairpin loop at the inner opening of the ion channel by the P loop. The formation of a glutamate-binding site by the two lobes of the protein projecting to the extracellular environment, domains S1 and S2, is also shown. Reproduced from Michaelis E (1998) Molecular biology of glutamate receptors in the central nervous system and their role in excitotoxicity, oxidative stress and aging. *Progress in Neurobiology* 54: 369–415.

RNAs (pre-mRNAs) are edited by an adenosine deaminase whose activity depends on the presence of the double-stranded RNA (dsRNA) structure between exon 11 and intron 11 of the GluR2 pre-mRNA. The adenosine in the CAG codon is deaminated to inosine, thus forming the cytosine-inosine-guanine (CIG) codon for Arg. This dsRNA deaminase is expressed in brain and other tissues and edits with very high efficiency (~90%) and selectivity GluR2 pre-mRNA. In the disease amyotrophic lateral sclerosis, editing of GluR2 is significantly less complete and leads to the expression of AMPARs in spinal neurons that are highly permeable to Ca^{2+} . This increased permeability to Ca^{2+} may cause neuronal death during the lifespan.

KA Receptor Genes and Protein Structure

Homomers or heteromers of GluK1, GluK2, or GluK3 (*Grik1–3* corresponding genes) form channels that resemble KARs in brain neurons. On the other hand, homomers of GluK4 or GluK5 (*Grik4–5* corresponding genes) fail to form channels unless they are co-expressed with GluK1–3. The channels formed by co-expression of GluK4 or GluK5 with one of the other KAR subunits leads to channel formation with differential pharmacological sensitivities and channel activation and desensitization kinetics. The expression of the KAR-like proteins $\delta 1$ and $\delta 2$, either as homomeric or heteromeric complexes, does not lead to ion channel formation. Therefore, these two genes are designated as orphan receptors. Based on these observations, GluK1–5 are considered to be the true

subunits of neuronal KARs. Different sets of neurons in the brain express these KAR receptor proteins and some of them, for example, those in the CA3 field of the hippocampus, are some of the most vulnerable brain neurons to KA-induced neurotoxicity (neurotoxic KARs) as well as to seizure-induced neuronal damage.

The RNA for GluK1 and GluK2 is also edited leading to a Gln to Arg change in the P loop. However, the efficiency of editing (35% and 75%, respectively) is lower than that observed for GluR2. For the GluK2 receptor protein, the permeability of the channel for Ca^{2+} is controlled not only by the editing of the Gln residue in the P loop but also by the editing of two other residues in the M1 domain of the protein.

NMDA Receptor Genes and Protein Structure

Expression of the NMDAR2A–D (*Grin2a–d* corresponding genes) subunits by themselves does not lead to the formation of functional NMDARs, but the co-expression of the NMDAR2 subunits together with NMDAR1 (*Grin1* corresponding gene) leads to the formation of ion channels that have conductances very similar to those of NMDA receptors in neurons. In the NMDAR1 protein, the amino acid present in the corresponding position as Arg586 in GluR2 is Asn598. The channels formed by these subunits of NMDARs are highly permeable to Ca^{2+} . Finally, each of the combinations of NMDAR1 and one of the NMDAR2 subunits exhibit different degrees of sensitivity to

the co-agonist glycine, the competitive antagonist 2-AP5, and the channel inhibitor Mg^{2+} .

The expression of NMDAR3A, B (*Grin3a,b* corresponding genes) together with NMDAR1 and NMDAR2 subunits leads to receptor-ion channels activated by glycine. These channels have lower conductance levels than those of NMDAR1/NMDAR2 receptor-ion channels, low Ca^{2+} permeability, and insensitivity to Mg^{2+} -induced channel block.

The glycine-binding site of NMDARs is in the NMDAR1 subunit and is formed by domains equivalent to the S1 and S2 domains of GluR1–4. The primary glutamate/NMDA-binding site of NMDARs is localized on the corresponding extracellular domains of the NMDAR2 subunits. Each NMDAR1 and NMDAR2 subunit pair forms a dimer through interactions at the S1–S2 domains. In addition to the seven isoforms of NMDARs, there are eight splice variant forms of NMDAR1 (NMDAR1a–h, [Figure 3](#)). These splice variant forms exhibit clearly distinguishable sensitivities to agonists, antagonists, Zn^{2+} , and polyamines. Thus, a wide array of receptor complexes with differing affinities for ligands may be expressed in neurons through different combinations of the NMDAR1/R2 subunits, as well as the NMDAR1 splice variants. Differential expression and localization of the NMDAR1 subunits in various brain regions or even within a single neuron, for example, cell body versus dendrites, could influence the responses of neurons to synaptically released L-glutamic acid.

Most synaptic NMDARs are formed by a combination of NMDAR1/R2A/R2B. Localization of the NMDAR proteins in synaptic regions is determined by the carboxyl-terminal region of the NMDAR1 and NMDAR2 subunits (NMDAR2A and 2B). Mice lacking one of the NMDAR2 subunits (NMDAR2A) have diminished NMDA receptor channel activity and a reduction in the process of induction of LTP in CNS neurons.

Neuronal toxicity produced by NMDA also depends on the specific combination of subunits that form the receptors in neurons. Cells of the human embryonic kidney cell line 293, for example, die when they are transiently transfected with the NMDAR1 and NMDAR2 subunits and such toxicity is greatest when NMDAR1 and NMDAR2A are co-expressed, less if NMDAR2B is co-expressed with NMDAR1, and it is not apparent when the NMDAR2C subunit is co-expressed with NMDAR1.

mGluR Genes and Proteins

Eight forms of mGluR genes have been identified ([Figure 4](#)). Group I are the mGluR1 and mGluR5, group II are the mGluR2 and mGluR3, and group III are mGluR4 and mGluR6–8. The activation of PLC by mGluR1 and 5 appears to involve transduction of the signal through coupling of the receptor proteins with either the Gi–Go family of guanosine triphosphate (GTP)-activated proteins (G proteins) or the Gq family. All of the group II and III mGluR proteins appear to be coupled to the Gi–Go family of G proteins. All mGluR proteins have a large extracellular amino-terminal domain, which contains a region that is homologous to bacterial periplasmic-binding proteins. It is likely that the glutamate-binding sites on mGluRs are within this extracellular domain. Insertions or deletions due to alternative mRNA splicing are confined to the carboxyl-terminal region and do not disrupt the amino-terminal domain of mGluRs.

Glutamate Receptors, Ca^{2+} -Signaling, and Synaptic Plasticity

Coincidence of action potentials in spines and dendrites with synaptic activation of glutamate receptors, primarily NMDARs, can enhance Ca^{2+} entry into neurons and produce long-lasting changes in synaptic activity. The timing of the two events, action potentials, and synaptic activation of receptor-induced Ca^{2+} influx, appears to be crucial. If receptors are activated after an action potential invades dendrites, it produces LTD. If it precedes or is coincident with the action potential it produces LTP. Activation of Ca^{2+} influx seems to be a key event in the expression of synaptic plasticity and neuronal adaptation but such increases in free $[Ca^{2+}]_i$ frequently work in a cooperative manner with other signaling events, such as the formation of cyclic adenosine monophosphate (AMP) (cAMP) and the activation of cAMP-dependent protein kinase A (PKA). For example, the establishment of LTP in some neurons is strengthened by the concurrent activation of neurotransmitter receptors that activate ACs, such as, dopamine, muscarinic acetylcholine, and β -adrenergic receptors.

The entry of Ca^{2+} into neurons following glutamate receptor activation, especially activation of NMDA receptors, or the opening of VGCCs leads to altered gene transcription, mRNA splicing, mRNA translation, and protein synthesis. These effects are, to a substantial degree, dependent on the stimulation of several kinases, including Ca^{2+} /calmodulin (CaM)-dependent kinases (CaMK), protein kinase C (PKC), extracellular signal-regulated kinase or mitogen-activated kinase (ERK or MAPK), MAPK-ERK kinase (MEK), stress-activated protein kinase, aurora kinase, and protein tyrosine kinases such as proline-rich tyrosine kinase and src kinase ([Figure 6](#)). However, not all types of Ca^{2+} influx into a cell lead to the same level of activation of gene transcription, even when the levels of $[Ca^{2+}]_i$ achieved are the same. There is clear compartmentalization of Ca^{2+} arriving at synaptic junctions as opposed to those received following generalized depolarization of a neuron. Ca^{2+} that enters following activation of either a receptor-ion channel or a VGCC apparently remains very close to the internal mouth of the channel (~ 100 nM radius). Even if all other free Ca^{2+} in dendrites and cell bodies of neurons is chelated, the localized Ca^{2+} near ion channels still function as an intracellular signal. Most likely readout of Ca^{2+} signaling is that of changes in the frequency of transient Ca^{2+} elevations and the most likely sensor of such frequency modulations is CaMKII. It is important to note that the mRNA for CaMKII α is selectively transported to the dendritic region of neurons and that activation of NMDARs leads to polyadenylation of CaMKII α mRNA and initiation of its translation through a process dependent on aurora kinase stimulation.

Stimulation of protein phosphorylation is pivotal in the activation of signal transduction cascades that leads to altered transcriptional regulation ([Figure 6](#)). A transcription factor that is sensitive to activation by several incoming stimuli and that regulates the expression of as many as 100 other genes, is the cAMP response element (CRE)-binding protein (CREB). A key gene whose transcription is regulated by CREB is that for brain-derived nerve growth factor (BDNF), a gene that is known to influence the establishment of LTP. CREB activation is dependent on phosphorylation of crucial Ser residues.

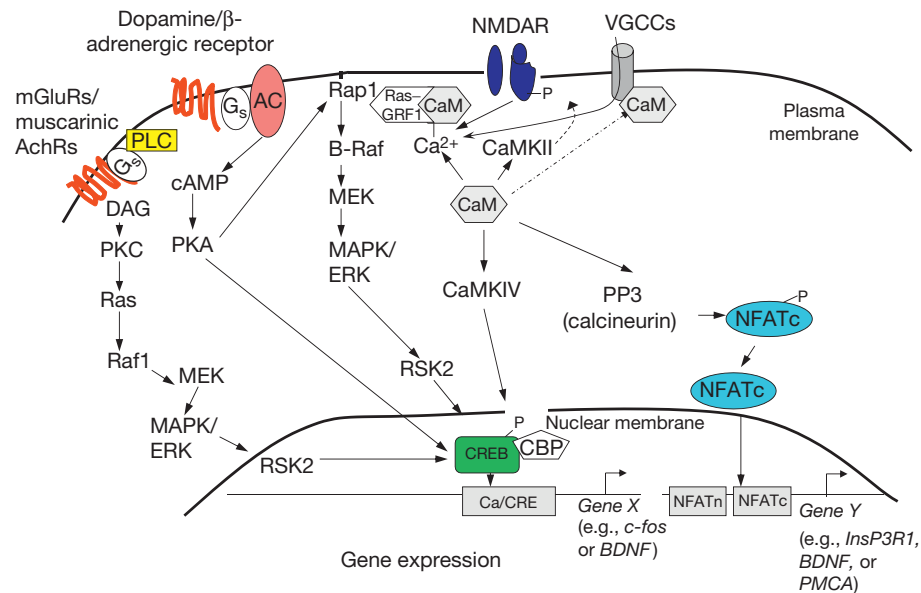


Figure 6 Regulation of transcription in neurons by Ca^{2+} following activation of NMDARs or VGCCs, and the interaction with pathways of gene transcription regulation following stimulation of G-protein-coupled receptors. Shown is the Ca^{2+} -induced and cAMP or diacylglycerol (DAG)-induced stimulation of kinases that phosphorylate CREB and cause activation of gene transcription. Also shown is the activation by Ca^{2+} and CaM of the protein phosphatase 3 (PP3) or calcineurin and the subsequent dephosphorylation of NFATc, the translocation of NFATc into the nucleus and the regulation of gene transcription in neurons. More details are presented in the text. CBP, CREB-binding protein.

CaMKIV, when directly activated by Ca^{2+} and CaM in the nucleus, phosphorylates CREB. But Ca^{2+} entering neurons through NMDA and VGCCs also leads to CREB phosphorylation through the activation of MAPK/ERK and RSK2 kinases (Figure 6). Activation of the small GTPase Rap 1 follows binding of the Ca^{2+} -CaM complex to a guanine-nucleotide exchange factor, Ras guanine-nucleotide-releasing factor (Ras-GRF1, Figure 6). The same cascade of kinases can also be activated by cAMP and PKA and by diacylglycerol (DAG) and PKC, thus leading to the potential strengthening of signal transduction by Ca^{2+} following coincident activation of receptors coupled to ACs or those coupled to PLC (Figure 6).

Ca^{2+} entry following NMDAR or VGCC activation also leads to the stimulation of the Ca^{2+} plus calmodulin-activated phosphatase PP3 (calcineurin) and, as a consequence of that, to the activation of the transcription factors nuclear factor of activated T cells (NFATc) and myocyte enhancement factor 2a. Activation of these two transcription factors is important in determining neuronal survival, axon and dendrite growth, synapse formation, and overall brain development. The transcription of some 34 genes is thought to be either up- or downregulated by NFATc in neurons. A prominent gene whose transcription is upregulated is the InsP_3 receptor-ion channel (Figure 6).

The activation of kinases following stimulation of glutamate receptors is thought to be a key step in the formation of LTP and memory formation, whereas the activation of calcineurin following weak stimulation of glutamate receptors is thought to cause primarily LTD. It is also noteworthy that phosphorylation of NMDARs by PKC, CaMKII α , or protein tyrosine kinases leads to increases in the response to agonists whereas activation of calcineurin or protein tyrosine phosphatases diminishes NMDAR responses. Given the importance of

NMDARs in synaptic plasticity and neuronal adaptation, it is not surprising to find such complex orchestration of protein modifications by kinases and phosphatases and of the consequent alterations in the activity of the receptors.

Glutamate Receptors, Ca^{2+} , and Neurodegeneration

Excessive activation of glutamate receptors leads to neuronal degeneration in the mammalian brain, such as that occurring during ischemia, anoxia, hypoglycemia, epileptic seizures, and chronic neurodegenerative diseases, for example, Alzheimer's disease. The entry of Ca^{2+} into neurons increases the activity of the calmodulin-activated form of nitric oxide synthase. Activation of this enzyme leads to the generation of intracellular nitric oxide radical ($\text{NO}^\bullet$) which, together with superoxide anion ($\text{O}_2^{\bullet-}$), forms peroxynitrite and leads to delayed neurotoxicity. Also, Ca^{2+} entry after activation of populations of nonsynaptic NMDA receptors is the primary contributor to alterations in mitochondrial Ca^{2+} handling and the initiation of neuronal apoptosis. The entry of Ca^{2+} through NMDA receptors also leads to the activation of proteases, such as Ca^{2+} -activated calpain, and enhanced proteolytic activity. Finally, activation of calcineurin and NFATc by Ca^{2+} entry into neurons has been shown to play an important role in Alzheimer's disease-related neuropathology.

See also: **Bioenergetics:** Calcium-Buffering Proteins: ER Luminal Proteins; Calcium in the Regulation of Gene Expression; Calcium Signaling: Calmodulin-Dependent Phosphatase; Calcium Signaling: NO Synthase; Calcium-Buffering Proteins: Calbindin; Calcium/Calmodulin-Dependent Protein Kinase II; ER/SR Calcium Pump;

Structure; Inositol Trisphosphate and Calcium Signaling; IP₃ Receptors; Membrane Transporters: Na⁺/Ca²⁺ Exchangers; Nuclear Calcium in Synapse-to-Nucleus Communication: Common Regulator of Persistent Adaptations in the Nervous System; Plasma-Membrane Calcium Pump: Structure and Function; Sodium–Calcium Exchanger: Structural Aspects; Voltage-Dependent K⁺ Channels; Voltage-Gated Ca²⁺ Channels; Voltage-Gated Sodium Channels: Structure, Function, and Pathophysiology; **Lipids Carbohydrates Membranes and Membrane Proteins**: Ion Channel Protein Superfamily; **Signaling**: Glutamate Receptors, Metabotropic; Mitogen-Activated Protein Kinase Family; Neurotransmitter Transporters.

Further Reading

- Cullen PJ and Lockyer PJ (2002) Integration of calcium and Ras signaling. *Nature Reviews* 3: 339–348.
- Furukawa H, Singh SK, Mancusso R, and Gouaux E (2005) Subunit arrangement and function in NMDA receptors. *Nature* 438: 185–192.

- Genazzani AA, Carafoli E, and Guerini D (1999) Calcineurin controls inositol 1,4,5-trisphosphate type I receptor expression in neurons. *Proceedings of the National Academy of Sciences of the United States of America* 96: 5797–5801.
- Hollmann M and Heinemann S (1994) Cloned glutamate receptors. *Annual Review of Neuroscience* 17: 31–108.
- Jin R and Gouaux E (2003) Probing the function, conformational plasticity, and dimer–dimer contacts of the GluR2 ligand-binding core: Studies of 5-substituted willardiines and GluR2 S1S2 in the crystal. *Biochemistry* 42: 5201–5213.
- Johansen TN, Greenwood JR, Frydenvang K, Madsen U, and Krosgaard-Larsen P (2003) Stereostructure-activity studies on agonists at the AMPA and kainite subtypes of ionotropic glutamate receptors. *Chirality* 15: 167–179.
- Michaelis EK (1998) Molecular biology of glutamate receptors in the central nervous system and their role in excitotoxicity, oxidative stress, and aging. *Progress in Neurobiology* 54: 369–415.
- Pin J-P and Duvoisin R (1995) Neurotransmitter receptors I: The metabotropic glutamate receptors: Structure and functions. *Neuropharmacology* 34: 1–26.
- West AE, Griffith EC, and Greenberg ME (2002) Regulation of transcription factors by neuronal activity. *Nature Reviews* 3: 921–931.
- Yasuda R, Sabatini BL, and Svoboda K (2003) Plasticity of calcium channels in dendritic spines. *Nature Neuroscience* 6: 948–955.

Ligand-Operated Membrane Channels: GABA

F Minier, François-Rabelais University, Tours, France

E Sigel, University of Bern, Bern, Switzerland

© 2013 Elsevier Inc. All rights reserved.

Glossary

Chloride equilibrium potential Membrane potential at which opening of chloride-ion-selective channel mediates no net ion flux due to the concentration gradient of the corresponding ion.

Conservative substitution Replacement of an amino acid residue by a residue with similar chemical properties.

Heterologous expression Genetic information in the form of complementary DNA (cDNA) or complementary RNA

(cRNA) is introduced into a foreign cell where the encoded protein is synthesized.

Hydrophobicity analysis The amino acid residue sequence predicted by the coding cDNA is scanned for the average hydrophobicity of a window of about 20 succeeding amino acid residues. A longer hydrophobic stretch is indicative of a transmembrane sequence in the protein.

Voltage clamp Imposition of a desired membrane potential on a cell using electrophysiological techniques.

Handling of the Neurotransmitter GABA: Biosynthesis, Transport, and Degradation

γ -Aminobutyric acid (GABA) is synthesized from glutamate in the cytosol of GABAergic neurons by two types of glutamic acid decarboxylase (GAD), namely, GAD65 and GAD67, according to their molecular weights (Figure 1). GAD65 is important for the local control of GABA synthesis at the synaptic sites, whereas GAD67, more widely distributed in the cell, is responsible for maintaining GABA baseline levels. After biosynthesis, GABA is loaded into synaptic vesicles by a vesicular transporter named 'vesicular inhibitory amino acid transporter (VIAAT)' or 'vesicular GABA transporter (VGAT)'. This transporter is predicted to be a 10-transmembrane domain protein and belonging to a superfamily of H^+ -neurotransmitter transporters. Presynaptic membrane depolarization triggers release of GABA into the synaptic cleft. After the activation of $GABA_A$ and $GABA_B$ receptors, GABA is taken up in surrounding neuronal and glial cells by high-affinity GABA transporters (GATs). This re-uptake contributes to the termination and modulation of GABA neuronal transmission. GATs belong to a superfamily of Na^+/Cl^- -dependent neurotransmitter transporters. Four GATs have been cloned: GAT-1, GAT-2, GAT-3, and betaine GABA transporter (BGT-1). They are 12-transmembrane domain proteins with both intracellular N and C termini, and an extracellular loop between transmembrane domains 3 and 4. Each of these transporters displays specific features: GAT-1 seems to be highly expressed throughout the brain at a neuronal presynaptic localization, while GAT-2 is mainly extra-parenchymal and found in the meninges. GAT-3 displays a similar macroscopic distribution as GAT-1. It is expressed in both presynaptic neurons and astrocytes. BGT-1 is weakly expressed throughout the brain in postsynaptic neuronal cells. The GABA degradation pathway, called 'GABA shunt', is one of the energy synthetic pathways in the brain. This pathway is composed of consecutive actions of three enzymes. The mitochondrial enzyme GABA-transaminase (GABA-T) catabolizes GABA to succinic semialdehyde (SSA). Then, SSA is converted either to succinic acid (a substrate for the Krebs cycle) by succinic semialdehyde dehydrogenase (SSADH)

or to γ -hydroxybutyrate (GHB) by succinic semialdehyde reductase (SSAR). All these pathways are summarized in Figure 1.

Early $GABA_A$ Receptor Biochemistry

Indirect evidence suggested a close relation of the $GABA_A$ receptor and the benzodiazepine-binding site. The benzodiazepine drugs are widely used for their sedative, anxiolytic, muscle relaxant, and anticonvulsant action. Affinity chromatography directed at the benzodiazepine-binding site led to the isolation of a pure protein complex with binding properties expected for the $GABA_A$ receptor and for its chloride pore. This demonstrated the presence of benzodiazepine-binding sites within the $GABA_A$ receptor. The subunits of the complex had a molecular mass of 49–58 kDa and the entire complex of ~250 kDa. Later, it became clear that five subunits surround the central ion pore.

Molecular Biology of the $GABA_A$ Receptor

Two $GABA_A$ receptor subunits were initially cloned. Hydrophobicity analysis indicated the presence of four membrane-spanning regions (Figure 2(b)). The second transmembrane segment of each of the five subunits contributes to the ion pore. Initial cloning was quickly followed by cloning of closely related subunits α_{1-6} , β_{1-3} , γ_{1-3} , δ , ϵ , θ , and π . At the genomic level, these subunits are organized on chromosomes in clusters, raising the possibility of a cluster-specific regulation of expression (Table 1). If the degree of amino acid sequence identity was >80%, subunits were grouped under the same Greek letter, if it was lower a new Greek letter was used. In both cases, subunits share predicted topology and there are many conservative substitutions of amino acid residues. With a sequence identity of ~20%, the subunits are also homologous to subunits of the other members of the same channel family including the nicotinic acetylcholine receptor, the glycine receptor, and the 5-hydroxytryptamine 3 (5HT₃) receptor.

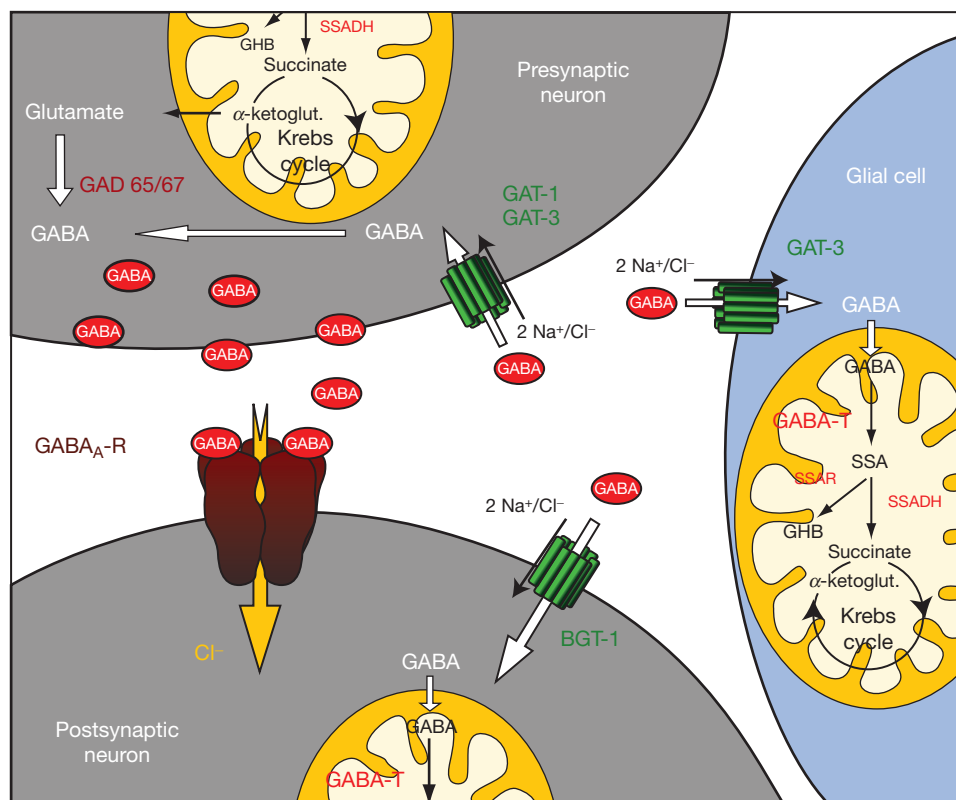


Figure 1 GABAergic synapse and main GABA metabolic pathways. GABA is synthesized from glutamate by the enzyme glutamate decarboxylase (GAD). Glutamate itself is synthesized from α -ketoglutarate (α -ketoglut.). The end of its action is determined by the GABA transporters of the GAT type or BGT-1 (for betaine/GABA transporter) and intracellular degradation by GABA transaminase (GABA-T) to succinic semialdehyde (SSA) and subsequently by succinic semialdehyde reductase (SSAR) to γ -hydroxybutyrate (GHB) or by succinic semialdehyde dehydrogenase (SSADH) to succinate, which is part of the Krebs cycle. Succinate in turn is converted to α -ketoglutarate.

Thus, 16 subunits of the GABA_A receptor are available and frequently many of them are coexpressed in the same cell. On the other hand, only five subunits can be accommodated in a single channel. Today, it is not known how many different GABA_A receptors exist and what their subunit composition is, but this diversity of subunits indicates that there may exist a large variety of GABA_A receptors. The major adult GABA_A receptor isoform is likely to be composed of $2\alpha_1$, $2\beta_2$, and $1\gamma_2$ subunits. Attempts to demonstrate coassembly of different subunit isoforms into a common pentamer have included immunoprecipitation and subsequent analysis of the precipitate, a type of approach heavily relying on excellent quality antibodies.

Physiological Function

In Situ Studies

GABA induces a conformational change in the receptor protein, such as to allow chloride ion flux. This ion flux may be measured as current flowing through the membrane by using electrophysiological techniques. Current amplitude depends on the concentration of GABA applied. While it is thought that the synaptic cleft is briefly flooded with GABA during activation of the synapse and, as a consequence, receptors on the postsynaptic membrane are maximally activated, a graded response may be obtained in receptors localized extrasynaptically and in

artificial systems. As a consequence of chloride channel opening, the membrane potential of the corresponding neuron will tend to move toward the chloride equilibrium potential which is about -75 mV in mature neurons. In case the GABA_A receptor has a postsynaptic localization, this opening of chloride ion channels following neurotransmitter release from the presynaptic neuron will be brief (~ 50 ms) and transient. These receptors mediate fast synaptic inhibition. GABA will then partially diffuse out of the synaptic cleft and trigger opening of extrasynaptic receptors. In this case, opening will be prolonged and a tonic inhibition will result.

Modulation

Benzodiazepines, neurosteroids, and barbiturates modulate the ion flux, while picrotoxinin inhibits it by blocking the pore and bicuculline, a plant alkaloid, acts at the GABA site as competitive inhibitor. Modulation by benzodiazepines will be briefly summarized here. On their own, classical benzodiazepines (such as diazepam) do not induce opening of the ion pore. If they are coapplied with the agonist GABA, they lead to an increase in the current (**Figure 3**) through an allosteric increase of the apparent affinity of the receptor for GABA to elicit channel opening. At the single channel level, this stimulation is evident as an increase in channel open frequency, at least when GABA is applied at submaximal

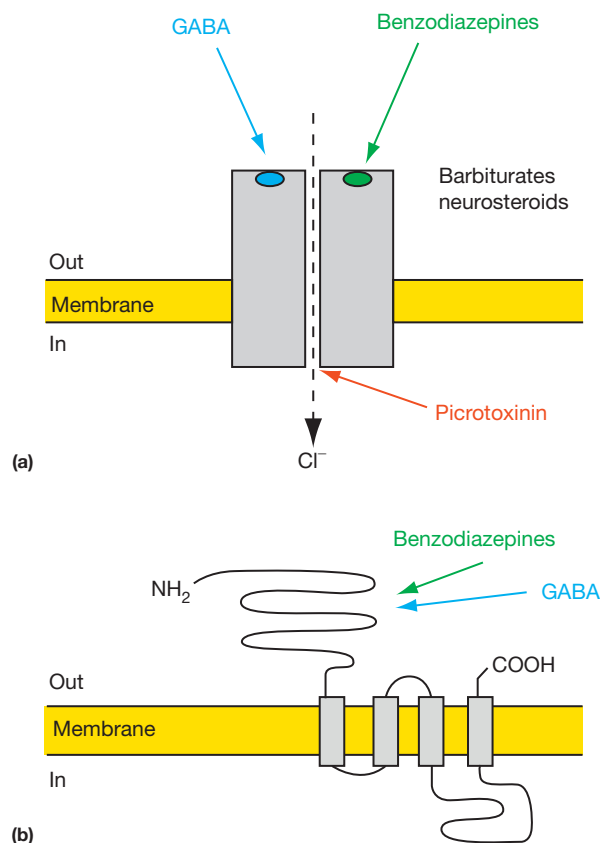


Figure 2 (a) The GABA_A receptor is built of five membrane protein subunits that surround a central chloride selective channel. Opening of this ion channel is caused by a conformational change in the protein occurring as a consequence of the binding of the channel agonist GABA. Various drugs (benzodiazepines and neurosteroids) modify channel opening and picROTOXIN blocks the pore. (b) Topology of a subunit characterized by a large N-terminal extracellular domain carrying the drug-binding sites. The second transmembrane domain lines the ion channel. The large loop between transmembrane segments 3 and 4 carries sites for posttranslational modification (phosphorylation).

Table 1 At the genomic level, GABA_A receptor subunits are organized into gene clusters whose chromosomal location is given for the human species

Subunits	Locus
δ	1p
$\alpha_2, \alpha_4, \beta_1, \gamma_1$	4p14–p12
$\alpha_1, \alpha_6, \beta_2, \gamma_2, \pi$	5q31–q35
$\alpha_5, \beta_3, \gamma_3$	15q11–q13
$\alpha_3, \epsilon, \theta$	Xq28

It is a major challenge for basic research to bring order in this subunit chaos, to establish the subunit composition of GABA_A receptors really existing *in vivo*, and, most importantly, to find drugs that selectively modulate their function.

concentrations. These benzodiazepines are therefore called ‘positive allosteric modulators’. There exist competitive antagonists (such as Ro 15–1788) and negative allosteric modulators as illustrated here with the β -carbolines (Figure 3). The picture

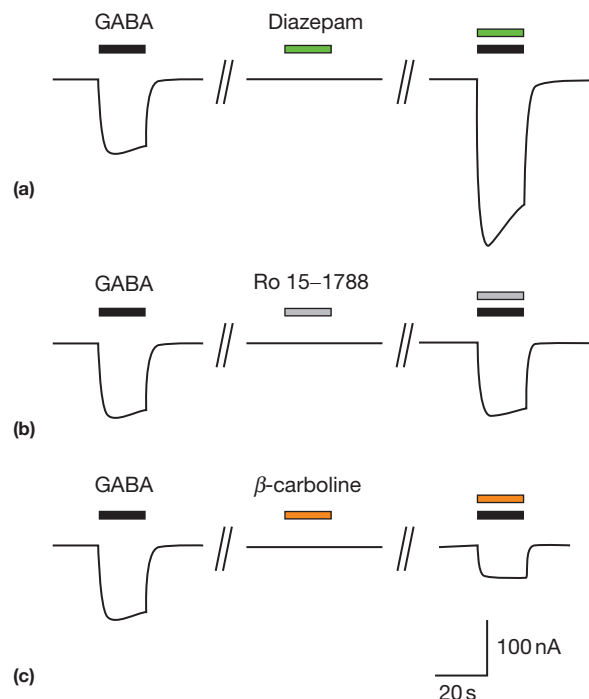


Figure 3 (a) Superfusion of a cell expressing GABA_A receptors with GABA induces an ion current that is measured using the two-electrode voltage-clamp technique. Superfusion of the same cell with the positive-allosteric modulator diazepam (prototype of benzodiazepines) does not elicit any current response while combined application of GABA and diazepam elicits a much bigger response than GABA alone. (b) Analogous experiments with the benzodiazepine antagonist Ro 15–1788, which binds with high affinity to the benzodiazepine-binding site but fails to alter the current responses. (c) Analogous experiment with a negative allosteric modulator of the β -carboline class, leading to a decrease in the current response.

is complicated by the fact that there exists a continuum of partial positive and negative allosteric modulators.

In Vitro Studies

Heterologous expression has been used to identify subunit isoforms that can form together a functional pentamer. This has been possible in cases where a certain subunit isoform confers to the receptor distinct functional properties. Such studies established, for example, that the combination of α - and β -subunits produces GABA-gated currents, but coexpression of a γ -subunit is required for benzodiazepine sensitivity of the expressed receptors. It has also been possible to exclude many theoretically possible subunit combinations.

Transgenic Mice

Transgenic mice either lacking a subunit isoform or carrying a diazepam-insensitive subunit isoform have been used to assign function or drug effects to an individual subunit isoform. Such studies have shown that α_1 contributes to the sedative effects, α_2 and α_3 to the anxiolysis induced by diazepam, and that α_5 is involved in memory functions.

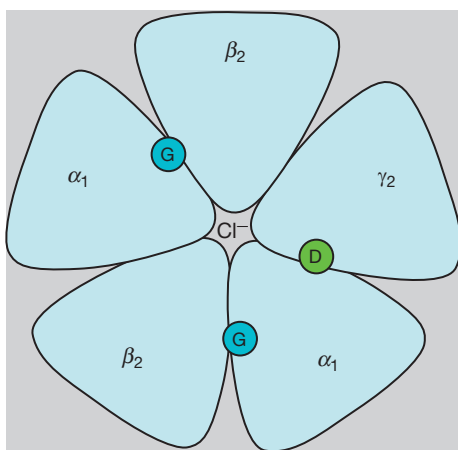


Figure 4 Subunit arrangement of the major isoform of the GABA_A receptor showing the arrangement of subunits around the central ion channel as seen from the synaptic cleft. The benzodiazepine-binding site (D) is located at the $\alpha_1\gamma_2$ -subunit interface in a homologous position to the two GABA (G)-binding sites located at $\beta_2\alpha_1$ -subunit interfaces.

Agonist- and Drug-Binding Sites

The following techniques have contributed to the description of amino acid residues involved in the interaction with the channel agonist GABA, its competitive antagonists, for ligands of the benzodiazepine-binding site, and for picrotoxinin: photoaffinity labeling, point mutation combined with either ligand-binding studies or functional characterization, cysteine point mutant ligand accessibility as determined by covalent modification by a chemically modified, cysteine-reactive ligand, and its protection by the corresponding ligand. The binding sites of other drugs such as barbiturates are still characterized to a lesser extent. Amino acid residues located in the N-terminal extracellular part of both the β_2 - and α_1 -subunits are taking part in the formation of the agonist-binding site, while residues on the α_1 - and γ_2 -subunits are taking part in the formation of the benzodiazepine-binding site in the major isoform of the GABA_A receptor. Both binding sites are thus located at subunit interfaces. Interestingly, both sites show a high degree of homology to each other, but only at the structural level and not at the functional level, as GABA can open the channel while benzodiazepines cannot. It should be noted that the benzodiazepine-binding site accepts not only benzodiazepines as allosteric modulators, but also some drugs with β -carboline, imidazopyridine, triazolopyridazine, and cyclopyrrolone structures.

Recently, an acetylcholine-binding protein homologous to the N-terminal extracellular portion of the nicotinic acetylcholine receptor has been crystallized. The GABA_A receptor shares structural and, to a small extent, sequence homology with the crystallized protein. The crystallized protein in its native state is able to bind acetylcholine, and a cavity binding this ligand has been identified. As the agonist-binding sites of the ligand-gated ion channel family have been conserved and as the binding pocket for benzodiazepines is homologous to the agonist site, this cavity must then be homologous to the binding pocket for benzodiazepines. The binding site is located in a cleft between

two subunits. This fits very well with the data obtained using classical methods, which indicated importance of the corresponding amino acid residues on α_1 and γ_2 that participate in the formation of the binding site. Thus, in $\alpha_1\beta_2\gamma_2$ GABA_A receptors, the binding pocket for benzodiazepines is located in a subunit cleft between α_1 - and γ_2 -subunits in a position homologous to the agonist-binding site for GABA that is located between β_2 - and α_1 -subunits (Figure 4).

Subtly Altered Receptor Subunits Cause Diseases

In several different cases, human point mutations in the γ_2 -subunit have been shown to occur in families with inherited forms of epilepsy. In both cases, not only the mutation has been tightly linked to the disease but also chloride ion flux is decreased, leading to a reduced state of neuronal inhibition. It can be anticipated for the future that many other diseases will be linked to isoforms of this important receptor. The GABA_A receptor has especially been implicated in pathological anxiety, insomnia, and epilepsy.

Architecture

The structure of the acetylcholine-binding protein has given an idea of how the extracellular domain of the GABA_A receptor could look like. The availability of a crystal structure of a protein homologous to the binding domain has also allowed establishment of the steric orientation of the subunits. From the study of covalently linked subunits, it has been concluded that the arrangement of subunits of the major GABA_A receptor isoform is $\beta\alpha\gamma\beta\alpha$. However, this still allows for a clockwise or counterclockwise arrangement. Figure 4 shows the correct arrangement as viewed from the synaptic cleft taking into account the information provided by the crystallization study.

See also: [Signaling: GABA_B Receptor; Neurotransmitter Transporters.](#)

Further Reading

- Barnard EA, Skolnick P, Olsen RW, et al. (1998) International union of pharmacology: XV. Subtypes of gamma-aminobutyric acid A receptors: Classification on the basis of subunit structure and receptor function. *Pharmacological Reviews* 50: 291–313.
- Borden LA (1996) GABA transporter heterogeneity: Pharmacology and cellular localization. *Neurochemistry International* 29: 335–356.
- Macdonald RL and Olsen RW (1994) GABA_A receptor channels. *Annual Review of Neuroscience* 17: 569–602.
- Moss SJ and Smart TG (2001) Constructing inhibitory synapses. *Nature Reviews Neuroscience* 2: 240–250.
- Rabow LE, Russek SJ, and Farb DH (1995) From ion currents to genomic analysis, recent advances in GABA_A receptor research. *Synapse* 21: 189–274.
- Rudolph U, Crestani F, and Möhler H (2001) GABA(A) receptor subtypes: Dissecting their pharmacological functions. *Trends in Pharmacological Sciences* 22: 188–194.
- Sieghart W, Fuchs K, and Tretter V (1999) Structure and subunit composition of GABA (A) receptors. *Neurochemistry International* 34: 379–385.
- Sigel E and Buhr A (1997) The benzodiazepine binding site of GABA_A receptors. *Trends in Pharmacological Sciences* 18: 425–429.

Light-Harvesting Complex I and II: Pigments and Proteins

S Jansson, University of Umeå, Umeå, Sweden

© 2013 Elsevier Inc. All rights reserved.

Glossary

Light-harvesting complex I (LHC I) The light-harvesting system containing proteins and pigments, associated with PS I.

Light-harvesting complex II (LHC II) The light-harvesting system containing proteins and pigments, associated with PS II.

Xanthophyll cycle The light-dependent conversions of the three xanthophylls – violaxanthin, antheraxanthin, and zeaxanthin; it plays a significant role in photoprotection.

Heterogeneity of Light-Harvesting Systems

Different taxa of photosynthetic organisms have organized light-harvesting systems in different ways, that is, have rather different light-harvesting proteins and pigments. In fact, phylogenetic grouping of photosynthetic organisms is often based on their light-harvesting pigments. Green sulfur bacteria get their color from bacteriochlorophyll *c*, *d*, or *e* in the chlorosome complex; red algae from phycoerythrin; and cyanobacteria from phycocyanin in their phycobilisomes. Green algae and higher plants have chlorophyll (chl) *a* and *b* in their light-harvesting complex I (LHC I) and LHC II. The antenna proteins from these phylogenetic group complexes share no sequence homology and light-harvesting systems are therefore likely to have evolved several times in contrast to photosynthetic reaction centers that seem to have only evolved once. The LHC systems of all higher plants (angiosperms and gymnosperms) as well as ferns, mosses, and green algae are, however, homologous structures and have a lot in common and these antenna systems will be described using examples from angiosperms, whose antennas are best studied.

The Higher Plant Light-Harvesting Antenna

The higher plant LHC proteins bind, via noncovalent bonds, the antenna pigment chl *a*, chl *b*, and a set of xanthophylls (carotenoids). The LHC proteins of most plants bind the same set of xanthophylls, lutein, violaxanthin (that can get photoconverted), and neoxanthin, although some plant species may contain unusual light-harvesting carotenoids. Since the excited state of chl *b* has a higher energy as compared to that of chl *a*, the preferred energy transfer path is from chl *b* to chl *a*. This energy transfer is supposed to take place through the Förster mechanisms, that is, excitation energy can jump between electrons in neighboring pigment molecules by a dipole–dipole mechanism. The chl molecules bound to the LHC proteins are, however, in different chemical environments and have thus different excitation energies, so some chl *a* molecules have slightly lower energy than the others. The differences in energy levels in general are typically small; so the energy transfer direction is not unidirectional and many jumps between antenna pigments could occur before it ends up in the reaction centers that have the lowest energy levels. Other light-harvesting systems are, however, organized

slightly differently. In phycobilisomes of cyanobacteria and red algae, peripheral pigments have considerably higher energy levels than those closer to the reaction center, so energy transfer there is unidirectional.

The Protein Components

The Lhca and Lhcb proteins

There are 10 major LHC proteins in angiosperms, whose genes are designated Lhca1–4 and Lhcb1–6. The Lhca genes code for LHC proteins associated with photosystem I, whereas the Lhcb3–6 proteins associate with PS II. Lhcb1 and Lhcb2 serve as antenna for both photosystems. Although all these proteins share a common secondary structure with three membrane-spanning regions, embedded in the thylakoid membrane of the chloroplasts, their tertiary structure differ. Lhcb1, Lhcb2 that typically are much more abundant than the others, form together with Lhcb3 trimeric complexes, the Lhcb4–6 (commonly named CP29, CP26, and CP24) proteins are monomeric, and the Lhca proteins seem to form dimers. This organization is not static, for example, Lhcb1 and Lhcb2 can monomerize and Lhcb5 can form trimers. The true molecular weight of each protein varies only slightly between different plants, although the apparent mobilities change considerably with both plant species and gel systems used, but they all fall within the range of 20–30 kDa. Since several of the proteins typically comigrate, identification solely based on molecular weight is often not possible.

Structure of the LHCs

In the monomeric structure, two homologous membrane-spanning regions form the core of the complex. Closely associated with these are two carotenoid molecules that in the case of LHC II are lutein. A third membrane-spanning region is located more peripheral in the complex, and a short C-terminal helix, not spanning the membrane, is located on the luminal side of the complex. Fourteen chlorophyll molecules (eight chl *a* and six chl *b*) are bound to each monomer, with the head groups in the interior of the complex and the phytol tails protruding toward the surrounding membrane.

Organization of LHC II

The organization of the LHC proteins around PS II in the so-called 'PS II supercomplex' is shown in Figure 1. The monomeric CP29 and CP26 proteins associate most closely with the

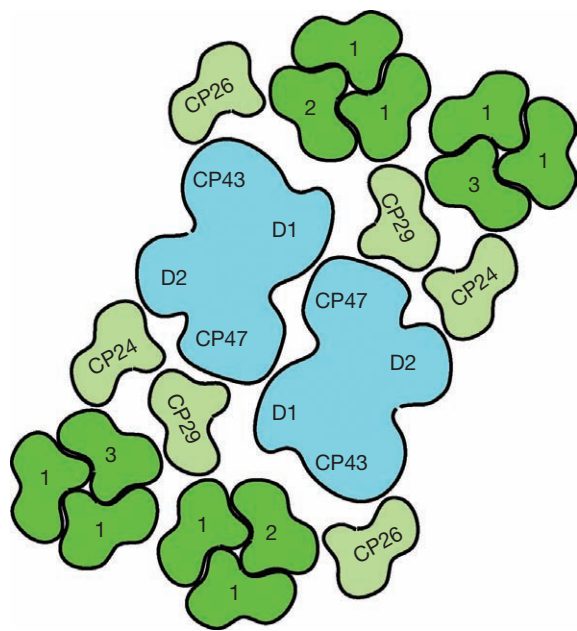


Figure 1 Schematic representation of the photosystem II dimer including the peripheral antenna. D1 and D2 are reaction center polypeptides, CP43 and CP47 core antenna proteins, CP24, CP26, and CP29 are monomeric LHC proteins, and 1, 2, and 3 denote tentative positions of Lhcb1, Lhcb2, and Lhcb3 proteins, respectively, in the peripheral LHC II trimers.

core complex, and CP24 sits next to CP29. LHC II trimers (consisting of trimers of Lhcb1, Lhcb2, and Lhcb3 in different combinations) attach to the monomeric proteins, which thus form a bridge between the trimeric proteins and the core. In the supercomplex, only three trimers could attach to each PS II center. In low light conditions, there are more trimers functionally attached to each PS II, although they appear not to be so tightly bound to the supercomplex. Larger complex of trimers has been observed, but it is not known how abundant they are *in vivo* and to what they are associated.

Organization of LHC I

The subunit LHC proteins seem to bind to only one side of the PS I complex. Lhca1 and Lhca4 form a heterodimer, named LHCI-730 after its characteristic low-temperature fluorescence emission peak, and Lhca2 and Lhca3 may perhaps also form heterodimers. The latter two proteins have also, based on the fluorescence emission of isolated complexes, been named LHCI-680, although it is likely that they, in the purified forms, have lost chlorophylls emitting at long wavelengths that *in vivo* associate with the complexes, so this name is misleading. The long-wavelength-emitting chlorophylls make up a small portion of the total pool of LHC I pigments. These are kept in a chemical environment that results in an absorption shifted to longer (red) wavelengths and their role seems to be to allow use of wavelengths just outside of the visible spectrum to excite PS I, despite the fact that they have lower energy than that of the PS I reaction center. In addition to the LHC I subunits, a fraction of the LHC II trimers associates with PS I. The attachment site for these seems to be opposite to that of

LHC I; so excitation energy presumably transfers directly from the trimers to the PS I core, not via LHC I.

The Pigment Components

The pigment/protein stoichiometry of the LHC proteins has been surprisingly hard to determine. The LHC II trimer is a pigment-dense complex and may bind in total 24 chl *a*, 18 chl *b*, 6 lutein, 3 neoxanthin, and 3 violaxanthin molecule. CP29 may bind 7 chl *a*, 2 chl *b*, and one each of the three carotenoids, and CP26 one additional chl *b*. The individual Lhca proteins, on the other hand, may also associate with about 14 chlorophylls, one or two luteins, a small amount of violaxanthin and perhaps also trace amounts of β -carotene, a nonxanthophyll carotenoid. The functions of the carotenoids are more complex than those of the chlorophylls. Carotenoids are indeed antenna pigments, also harvesting lights at wavelengths (~ 500 nm) where the chlorophylls have a low absorptivity, enabling plants to efficiently utilize daylight, although the chlorophylls mainly absorb in the red and blue regions of the spectrum. In addition, they have an equally important function in photoprotection against excess light. Photoprotective processes are necessary in the antenna systems if the excitation energy is not able to quickly enter into the reaction centers but stays in the antenna system for prolonged times. This could, for example, happen if the energy-consuming reactions downstream of the photosynthetic light reaction are not able to utilize the energy in the same speed as it is generated, for example, if the system is at low temperature when enzymatic reactions are slow. This creates a block in electron flow away from the reaction centers that will not be able to accept another energy package from the antenna. Excited chlorophyll molecules can, via a triplet state, under such circumstances, produce singlet oxygen molecules that are highly reactive and thus dangerous for the system.

Mechanisms of Photoprotection

The molecular details behind the roles of carotenoids in photoprotection are not fully understood, but it involves both direct effects (quenching of reactive oxygen species) and an indirect effect, namely, an increase in energy dissipation in the antenna resulting in loss of excitation energy. This process, named 'feedback de-excitation' or 'qE type of nonphotochemical quenching', is mediated by PsbS, a protein related to the LHC proteins. Feedback de-excitation is triggered (within seconds) when the luminal pH drops below a certain point, as a result of photosynthetic proton pumping over the thylakoid exceeding the backflow through the ATP synthase complex. The PsbS protein is then protonated but the low pH does also bring about a change in the photosynthetic xanthophylls (the xanthophyll cycle), violaxanthin is converted to zeaxanthin via antheraxanthin by the enzyme violaxanthin de-epoxidase, activated by the low luminal pH. Zeaxanthin accumulates in the antenna and is also found free in the thylakoid membrane. Zeaxanthin could protect directly against oxygen radicals, but when zeaxanthin is bound to protonated PsbS, a conformational change takes place in the antenna, resulting in the creation of a highly quenching species that relieves the excitation

pressure from the system into heat, and thus contributes to photoprotection. Only a small fraction of the zeaxanthin formed is, however, associated with PsbS; most of the other LHC proteins can also have zeaxanthin bound to them, but whether this also contributes to photoprotection is not clear. Zeaxanthin is backconverted to violaxanthin when the light level drops. It appears also as if a role of this type of quenching is not restricted to zeaxanthin, although it is likely that under *in vivo* conditions, zeaxanthin is a major regulator of quenching.

Other Regulatory Roles

LHC II is also involved in another regulatory process, referred to as 'state transitions', based upon the different states of chl fluorescence that can be obtained by treating the plant with light of different quality. If the plant is exposed to light preferentially absorbed by PS I (wavelengths >680 nm) or normal light, the fraction of LHC II trimers functionally associated with the photosystems can change in order to create an even excitation pressure of the photosystems. Overcapacity of PS II leads to over-reduction of the interphotosystem electron carriers that, subsequently, leads to the induction of an LHC II kinase. Phosphorylation of LHC II results in a reduction of the PS II antenna size, releasing the system from the imbalance. There is a connection between light quality and quantity, LHC II phosphorylation, LHC II migration between the different thylakoid regions and the amount of *grana* stacking, but the details are not yet fully understood. In any case, the reversible phosphorylation takes place on a Thr or Ser residue close to the N terminus of the protein, which protrudes into the stroma. The phosphorylation leads presumably to a conformational change that, in turn, changes the properties of the complex.

The amount of LHC proteins varies considerably in different light regimes. Higher light results in decreased amounts of

LHC and, as a consequence, since chl *b* is confined to the LHC proteins, an increased chl *a/b* ratio of the plant. Hence, this ratio is an indicator of the amount of antenna protein and, thus, antenna size of the plant. This is an acclimation to the conditions when light harvesting is not limiting for photosynthesis. However, the different LHC proteins do not decrease simultaneously: the quantity of Lhcb1 and Lhcb2 (making up the bulk of the trimers) is most variable; but at full sunlight, Lhcb1 and Lhcb2 quantities are low, and the other LHC proteins also decrease in abundance.

See also: **Bioenergetics:** Chlorophylls and Carotenoids; Dynamic Behavior of Photosystem II Light Harvesting; Photosystem I; Photosystem II: Assembly and Turnover of the D1 Protein; Photosystem II: Redox and Protein Components; Photosystem II: Water Oxidation, Overview; **Metabolism Vitamins and Hormones:** Photosynthesis.

Further Reading

- Bassi R, Giuffrè E, Croce R, Dainese P, and Bergantino E (1996) Biochemistry and molecular biology of pigment binding proteins. In: Jennings RC, Zucchelli G, Gheiti F, and Colombetti G (eds.) *Light as an Energy Source and Information Carrier in Plant Physiology*, pp. 41–63. New York: Plenum.
- Horton R, Ruban AV, and Walters RG (1996) Regulation of light harvesting in green plants. *Annual Review of Plant Physiology and Plant Molecular Biology* 47: 655–684.
- Jansson S (1994) The light-harvesting chlorophyll *a/b* binding-proteins. *Biochimica et Biophysica Acta – Bioenergetics* 1184: 1–19.
- Jansson S (1999) A guide to the Lhc genes and their relatives in *Arabidopsis*. *Trends in Plant Science* 4: 236–240.
- Niyogi KK (1999) Photoprotection revisited: Genetic and molecular approaches. *Annual Review of Plant Physiology and Plant Molecular Biology* 50: 333–359.
- Young AJ and Britton G (1993) *Carotenoids in Photosynthesis*. London: Chapman and Hall.

Lipases

H L Brockman, University of Minnesota, Austin, MN, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 571–575, © 2004, Elsevier Inc.

Glossary

Amphipath A molecule that contains both polar, hydrophilic (water-loving) and apolar, hydrophobic (water-avoiding) groups that are spatially distinct, conferring a tendency to localize at and stabilize interfaces between water and oil phases.

Designer fat Dietary fats or fat substitutes that have been *de novo* synthesized or modified, either chemically or enzymatically, to achieve desired intestinal absorption ability, taste, and thermal or rheological properties.

Interface A quasi-two-dimensional phase between two bulk phases in which the physical environment of the molecules that comprise the region is distinct from the bulk phases.

Interfacial activation The process by which an enzyme increases its activity in the presence of an interface. For lipases it results from the combination of enzyme and substrate being oriented and concentrated at interfaces as a consequence of their amphipathic nature (the substrate theory) and from the conformational transition of the lipase to an open conformation as it binds to the interface (the enzyme theory).

Lipases are water-soluble, ester hydrolases that are traditionally defined by their marked preference for apolar, water-insoluble ester substrates. This group of enzymes also includes species referred to as cholesterol esterases. Lipases and cholesterol esterases are distinguished from phospholipases that catalyze the hydrolysis of acyl ester bonds of highly amphipathic phospholipids having an *sn*-glycero-3-phospho-X moiety and from carboxylesterases that hydrolyze polar, water-soluble esters. These distinctions are relative, however, because some lipases exhibit activity toward phospholipids or soluble esters. Typical natural lipase substrates include, in order of amphipathicity, long aliphatic chain acyl esters of cholesterol (cholesteryl esters), triacyl esters of glycerol (triacylglycerols), acyl esters of long chain alcohols (wax esters), diacyl esters of glycerol (diacylglycerols), and monoacyl esters of glycerol. Because lipase substrates tend to be oily and only weakly amphipathic, they reside primarily in a bulk oil phase in preference to the aqueous phase or to the interface, that is, monomolecular surface phase that separates the bulk oil and aqueous phases. It follows, because lipases are water-soluble enzymes, that the site of lipolysis is the quasi-two-dimensional interface. The focus of basic research on lipases has been to understand how a reaction involving such a change in dimensionality can occur and how it is regulated. Medically, lipases are targets for therapeutic intervention in the treatment of obesity. The focus of applied research with lipases has been to exploit the unusual properties of lipolytic systems for the production of chiral pharmaceuticals, improved detergents, and designer fats.

Roles of Lipases

In times of abundant energy supply cells accumulate triacylglycerols or wax esters as energy-storage depots. This is exemplified by adipose tissue in animals and by the formation of seed oils in plants. Cholesteryl ester storage provides a potential reservoir of cholesterol for steroidogenic tissues. Triacylglycerol

and cholesteryl ester accumulation also provides a mechanism of cellular defense by ameliorating the toxic effects of excess fatty acids and cholesterol. Most notable in this regard is the accumulation of cholesteryl esters in the arterial endothelium during atherogenesis. The mobilization of these bulk lipid inclusions intra- and extracellularly is catalyzed by lipases and cholesterol esterases. Once the ester bond is split, the more amphipathic hydrolysis products can be used metabolically or mobilized, in higher organisms, for transport among tissues. Because of the requirement, for hydrolysis for transport, lipases are key enzymes in maintaining lipid homeostasis in multiorgan animals and plants. For example, in mammals there are secreted gastric and pancreatic lipases for digesting dietary lipids. From the resulting hydrolysis products, long-chain fatty acids, and monoacylglycerols, triacylglycerols are resynthesized in the intestinal wall. These circulate in the form of lipoproteins that are then subjected to hydrolysis by lipoprotein lipase for entry into peripheral tissues and by hepatic lipase for entry into the liver, respectively. Within peripheral tissues (e.g., adipose tissue), triacylglycerols are resynthesized and, when needed, are mobilized by hormone-sensitive lipase/cholesterol esterase. The resulting fatty acids either re-enter the circulation for transport back to the liver for re-esterification or to organs such as muscle and heart for energy generation.

Hydrolysis Mechanism and Specificity

The catalytic domain of lipases contains both lipid-binding determinants and the active site. The core structure of the catalytic domain is an α/β -hydrolase fold, a motif common to other classes of hydrolases, that consists of β -sheets connected by α -helices. This supports a Ser-His-Glu/Asp catalytic triad resembling that of serine proteases and other hydrolytic enzymes. The active site serine is contained in a consensus sequence of Gly-X-Ser-X-Gly that forms a tight turn exposing the serine hydroxyl group to the scissile ester bond of the

substrate. The general mechanism of the reaction catalyzed by lipases is shown in **Figure 1**.

Substrate specificity among various lipases is variable because of the long flexible hydrocarbon-like groups that must be accommodated by the active site. This is in contrast to carboxylesterases that act on more soluble substrates and tend to have much more restrictive sites to accommodate a particular substrate. Lipases (e.g., pancreatic triacylglycerol lipase) that have relatively restricted substrate access to the active site tend to be specific for esters of primary alcohols, whereas others (e.g., *Candida rugosa* lipase) readily catalyze hydrolysis of esters of secondary alcohols as well. These latter enzymes are also more likely to exhibit activity against phospholipids and other nonconventional substrates. The list of atypical substrates includes peptides, simple amides, Schiff bases, (lyso) phospholipids, percarboxylic acids, and polyesters.

In addition to differing specificities for acyl and alkyl size, some lipases (e.g., *Rhizomucor miehei* lipase) show pronounced stereospecificity, whereas others (e.g., porcine pancreatic triacylglycerol lipase) do not. Interestingly, stereospecificity extends to prochiral molecules (e.g., trioleoylglycerol), in which the lipid substrate is symmetrical but the product of the reaction following removal of an acyl group from position 1 or 3 is chiral. Overall, there is interplay among the various types of specificity in lipases so that generalizations are difficult and specificity must be determined on a substrate-by-substrate basis. Also, the presentation of the substrate to the lipase can affect measured specificity. For pharmaceutical and other fine chemical applications, there have been extensive efforts to engineer lipases to carry out particular reactions with high stereospecificity.

Although lipases are defined and function *in vivo* as hydrolytic enzymes, they are increasingly employed industrially for *de novo* ester synthesis and acyl exchange. Most globular proteins, including lipases, will irreversibly denature if confronted with an interface between an aqueous phase and a highly

apolar phase (e.g., hydrocarbon-like). However, if water is removed from the enzyme to the extent that its chemical activity is reduced to very low levels, for example, by freeze-drying, the enzymes are stable and potentially active in an apolar medium. While this is true of many enzymes, it is particularly advantageous with lipases. This is because water is a reactant in the system (**Figure 1**) and because the substrates are nonpolar. The absence of water shifts the equilibrium of the reaction from ester hydrolysis to ester synthesis and the reactants are compatible with or can even serve as their own apolar solvent. This enables the synthesis of esters in high yield or the exchange of one ester acyl group for another (**Figure 1**).

Lipolysis Regulation

Physiochemical Importance of Interfaces

Lipases are classically defined as enzymes, generally monomeric and water soluble, that catalyze the hydrolysis of ester substrates that are apolar and, therefore, generally water insoluble. Although this definition indicates what lipases do, it does not reveal how they do it. The key to understanding how lipases work is to appreciate the role of the interface between the oil and water phases as an enabler of catalysis. This is shown schematically in **Figure 2**. Having ester functional groups, lipase substrates are more polar than simple hydrocarbons and have some tendency to partition between the bulk oil phase and the interface. Likewise, lipases will also partition to an oil–water interface from the aqueous phase. The consequence of these partitioning reactions is to orient and concentrate the substrate and enzyme together in a quasi-two-dimensional phase. This phenomenon is known as the substrate theory of why lipases are more active on insoluble substrates than soluble ones. In this view of lipolysis, the interface assumes part of the role played by classical enzyme–substrate binding in solution, but in a nonspecific way.

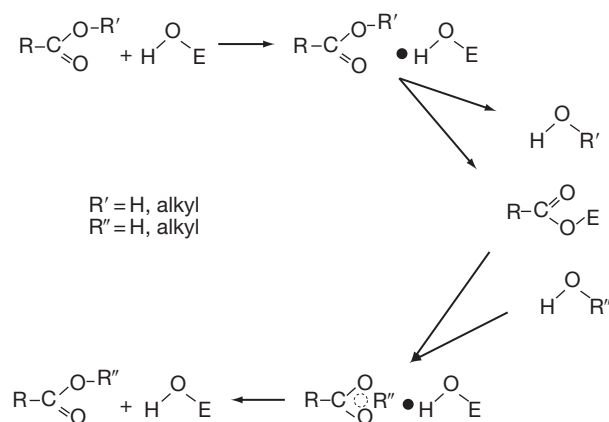


Figure 1 Lipolysis mechanism. E–O–H represents the active site serine which forms the acyl-enzyme intermediate in catalysis. The interchangeability of H, R', and R'' shows that lipases readily catalyze ester synthesis, ester hydrolysis, and acyl exchange, depending on the activity of water in the system, and the circular bond in the lower right structures indicates that the carboxyl oxygen atoms become randomized during the catalytic cycle.

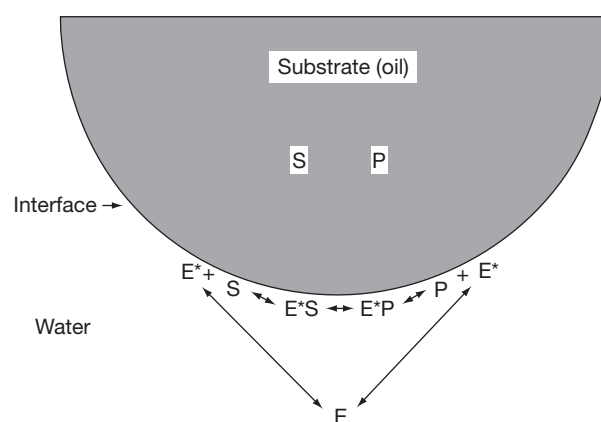


Figure 2 The interfacial nature of lipolysis. The substrate (S) is initially present in or as an oil droplet (shaded) whereas the lipase (E) is in the aqueous phase. Catalysis requires that both E and S partition to the interface. Partitioning of E involves a conformational change with the bound species noted as E*. Reaction products (P) will remain in the interface or partition to the oil phase, depending on their amphipathicity.

The most important consequence of enzyme and reactant partitioning is that the instantaneous rate of lipolysis depends on the two-dimensional concentrations of enzyme and reactants in the interface rather than their concentrations in either of the bulk phases.

This sequential scheme for lipolysis has other consequences as well. For a substrate, it means that its amphipathicity, together with its bulk concentration in the oil phase, will contribute to its two-dimensional concentration in the interface. Simplistically, if one had an equimolar mixture of glycerides in an oil phase, their interfacial concentrations would be in the relative order monoacylglycerols > diacylglycerol >> triacylglycerol. Measuring initial rates of hydrolysis for this mixture using a hypothetical lipase with no positional specificity, and normalizing for the number of ester bonds available would show this same apparent order leading, incorrectly, to the conclusion that the lipase is specific for monoacylglycerols. Another consequence of the interfacial nature of lipolysis is that the reaction products will accumulate in the interface, unless they are removed by partitioning into the oil phase or are solubilized in the aqueous phase. With respect to the latter, albumin or cyclodextrins are often used *in vitro* to trap fatty acids and monoacylglycerols in the aqueous phase during the hydrolysis of tri- or diacylglycerols. When it occurs, the accumulation of reaction products in the interface enriches the interface in more amphipathic molecules, greatly complicating the measurement of specificity and interpretation of kinetic data. Product accumulation can also have the effect of slowing the reaction due to fatty acid accumulation while at the same time causing the area of the interface to expand. Area expansion will have the opposite effect by favoring the partitioning of more lipase to the interface. Thus, once lipolysis is initiated in a system having a bulk oil phase as well as an aqueous phase, the regulation of lipolysis rate and extent becomes extremely complex. It is for this reason that researchers have utilized simpler systems, like mono-molecular lipid films, in efforts to understand lipase regulation.

That lipolysis is heterogeneous, unequivocally demonstrated by the P. Desnuelle group in 1965. They showed that if the same insoluble substrate was presented to a lipase as coarse and fine dispersions in buffer, the apparent Michaelis constant for the reaction (i.e., the three-dimensional concentration of substrate at which the hydrolysis rate was half maximal) was lower for the fine dispersion. However, if the rate data were analyzed as a function of total substrate particle surface area rather than bulk substrate concentration, the constant was the same for both dispersions. This occurred because the two-dimensional concentration of substrate at the particle interface was actually the same for the coarse and fine dispersions and the observed saturation of reaction rate with increasing substrate dispersion reflected saturation of lipase partitioning to that surface with increasing interfacial area. Despite the complexity of lipolysis kinetics arising from the interfacial nature of the process, it has been possible to develop kinetic schemes for analyzing lipase kinetics, at least for simple experimental systems, for example, where reaction products are removed from the interface when formed. These are valuable for screening potential lipase inhibitors to determine the efficiency at which each step of the reaction (i.e., partitioning or interfacial catalysis) contributes to the observed inhibition.

Structural Adaptation of Lipases at Interfaces

It might be presumed that lipases would possess a high affinity for lipid–water interfaces, that is, that they would be highly amphipathic. However, their affinities toward interfaces dominated by amphipathic lipids, such as phospholipids, is often no higher than that of other water-soluble globular proteins. Indeed, many lipases can be isolated as soluble monomeric proteins that do not show a propensity to aggregate. However, if exposed to a surface of a pure substrate, such as triacylglycerol, lipases bind with high affinity. Related to this observation is their ability to efficiently catalyze the hydrolysis of insoluble substrates, but not soluble ones. To explain this observation, it was proposed early on that lipases somehow became activated at interfaces. In the 1990s, the determination of lipase structures under varying conditions revealed the physical basis of this phenomenon (Figure 3). In aqueous solution the active site of most lipases is not exposed to the medium but lies in the interior of the protein under a lid or flap structure. This keeps soluble ester substrates away from the catalytic triad. As implied by its name, the lid can open to reveal the active site and, simultaneously, to create and expose a large area comprised of hydrophobic amino acids. Whereas this open conformation is energetically unfavorable in the aqueous phase, it may be induced or trapped by the presence of an apolar surface. The lipase then binds with high affinity to the interface; the active site is accessible for diffusion of substrate molecules into it and catalysis can proceed. It is the conformation of this open form of the enzyme that determines specificity for

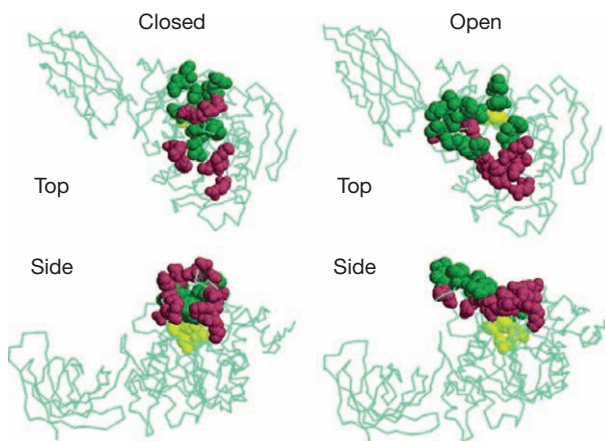


Figure 3 Closed and open conformations of pancreatic triacylglycerol lipase. The backbone structure of the lipase is shown in light blue-green with the C2 regulatory domain on the left and the catalytic domain on the right. Nonpolar (green) and polar/charged (red) residues in the vicinity of the catalytic triad (yellow) are shown space filled. Note that in the closed conformation, presumed to exist in solution (left) and designated E in Figure 2, mostly polar/charged residues are exposed to water and the catalytic triad is shielded from water. In the closed conformation, presumed to exist at the interface and designated E' in Figure 2, a large, apolar surface is exposed on the top of the molecule as shown and the catalytic triad becomes accessible from the top. This surface presumably interacts with the substrate-containing oil droplet. The closed structures were drawn from coordinates kindly provided by Dr. C. Cambillau and the open structure from the Protein Data Base, file 1LPA.pdb.

particular substrates. The changing of conformation by lipases in the vicinity of an interface is known as the enzyme theory of why lipases are more active on insoluble substrates than on soluble ones. It should be noted that the enzyme theory of lipolysis and the substrate theory of lipolysis described above are not mutually exclusive.

A still unresolved question is what kind of a surface a lipase needs for surface binding in the open conformation. Natural interfaces tend to be dominated by phospholipids and proteins. These are necessary to emulsify the apolar lipid substrate, that is, to stabilize it in a dispersed state with high surface area, but tend to inhibit lipase binding to the interface. Research suggests that lipases need a defect- or a patch-free binding inhibitor in the interface to initiate lipolysis. Once bound, the generation of lipolysis proceeds and its reaction products increase the number of binding sites and more lipase molecules can bind. Thus, lipolysis often appears to be autocatalytic in the presence of inhibitory phospholipids or proteins, that is, there is an initial lag in lipolysis followed by a burst. Overcoming the inhibition of lipases by amphipathic lipids and proteins is important to understand not only their roles and regulation *in vivo*, but also their applications, for example, improve the efficiency of detergents.

Lipase Regulatory Domains

Many lipases, such as gastric lipase, consist of a single protein domain but others have a second domain connected to the catalytic domain by a flexible linker. The most studied of these are members of the lipase gene family, exemplified by pancreatic triacylglycerol lipase (Figure 3), lipoprotein lipase, and hepatic lipase. These have a C2 domain, a motif common to other peripheral (also called amphitrophic) proteins, that is, proteins that reside in solution but translocate to and function at a lipid–water interface in response to a signal. The C2 domain framework is an eight-stranded, antiparallel β -sandwich with connecting loops. In lipases having C2 domains one of these loops exposes hydrophobic amino acid side chains, such as leucine, tyrosine, and tryptophan. These confer an amphipathic character to the domain and, depending on assay conditions, can be essential for lipase activity to be expressed. Given the large hydrophobic surface created near the active site of the catalytic domain by the opening of the lid, such dependence on a C2 domain for initiating lipolysis suggests that it is needed to associate the enzyme with a substrate-containing interface long enough for catalytic domain to open and lock the enzyme on the interface.

Lipase Cofactors

Two members of the lipase gene family, pancreatic triacylglycerol lipase and lipoprotein lipase, utilize cofactor proteins for

carrying out efficient lipolysis. As noted above, lipase interface binding tends to be inhibited by those molecules that help create and stabilize interfaces, such as phospholipids. Cofactor proteins facilitate the initiation of lipolysis at such interfaces. The traditional view of cofactors has been that they are amphipathic proteins that simply bind both the interface and the lipase and thereby anchor the lipase to the interface. However, the comparable affinities of lipases and their cofactors for protected interfaces suggest that they do more. Structural studies of the complex between pancreatic lipase and its protein cofactor, colipase, show that the cofactor may facilitate the opening of the lid through protein–protein interactions with the open conformation of the catalytic domain of the enzyme. Another role has been suggested by comparison of the ability of the cofactor to bind to interfaces with different compositions. Specifically, if a phospholipid-rich interface also contains the substrates and products of lipolysis, colipase binding is strengthened. This occurs because the colipase molecules tend to laterally concentrate the nonphospholipid molecules in their vicinity. In this manner, they help to create a nanometer-sized region from which phospholipid tends to be excluded. This facilitates lipase binding by a combination of lipid–protein and protein–protein interactions and thereby helps initiate lipolysis. Another speculative role for lipase cofactors is that their preference for binding to substrate-containing surfaces acts to target lipase to those surfaces in preference to interfaces that do not contain substrate.

See also: **Metabolism Vitamins and Hormones:** Cholesterol Synthesis and Regulation; Fatty Acid Oxidation; Fatty Acid Structure and Synthesis; **Protein/Enzyme Structure Function and Degradation:** Lipid Modification of Proteins: Targeting to Membranes; **Lipids Carbohydrates Membranes and Membrane Proteins:** Lipid Bilayer Structure.

Further Reading

- Brockman HL (2000) Kinetic behavior of the pancreatic lipase–colipase–lipid system. *Biochimie* 82: 987–995.
- Fojan P, Jonson PH, Petersen MTN, and Petersen SB (2000) What distinguishes an esterase from a lipase: A novel structural approach. *Biochimie* 82: 1033–1041.
- Kraemer FB and Shen WJ (2002) Hormone-sensitive lipase: Control of intracellular tri-(di-)acylglycerol and cholesteryl ester hydrolysis. *Journal of Lipid Research* 43: 1585–1594.
- Lombardo D (2001) Bile salt-dependent lipase: Its pathophysiological implications. *Biochimica et Biophysica Acta* 1533: 1–28.
- Miled N, Canaan S, Dupuis L, et al. (2000) Digestive lipases: From three-dimensional structure to physiology. *Biochimie* 82: 973–986.
- Panaïotov I and Verger R (2000) Enzymatic reactions at interfaces: Interfacial and temporal organization of enzymatic lipolysis. In: Baszkin A and Norde W (eds.) *Physical Chemistry of Biological Interfaces* pp. 359–400. New York: Dekker.

Lipid Bilayer Structure

E London, Stony Brook University, Stony Brook, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Biological membranes A thin shell surrounding either internal compartments in cells or the entire cell. Composed of special proteins and lipids.

Lipids Small biomolecules with hydrophobic characteristics.

Polar headgroup The portion of a lipid molecule that interacts with aqueous solution; it usually contains several chemical groupings that can hydrogen-bond with water.

Chemical Structure of Bilayer-Forming Lipids

Lipid bilayers are composed primarily of polar lipids. Glycerophospholipids and sphingolipids are usually the predominant polar lipids in natural membranes. Unlike dietary lipids (fats), these molecules possess a highly polar headgroup in addition to (two) nonpolar (i.e., hydrophobic) tails ([Figure 1](#)).

The hydrophobic tails consist of the hydrocarbon chain portions of fatty acids and/or the base sphingosine. Hydrocarbon chains are linear in eukaryotic organisms, but sometimes have branching methyl groups in bacteria.

The polar headgroups of the lipids can have various chemical structures. Often they contain phosphate to which a polar or charged alcohol derivative is attached. In other cases they contain carbohydrates. The polar headgroup can be electrically neutral, zwitterionic, or negatively charged.

In eukaryotes, sterols such as cholesterol are also major components of the bilayer. Sterols differ from other polar lipids in having a simple, small hydroxyl group as their polar portion. This hydroxyl group is attached to four fused aliphatic hydrocarbon rings. In addition, they contain a branched hydrocarbon chain that is attached to the end of the rings opposite the hydroxyl.

Organization of Lipids in Bilayers

In bilayers lipids orient such that their polar groups are in contact with the aqueous environment and their hydrocarbon segments face each other ([Figure 1](#)). As a result, the polar groups in each half of the bilayer (called a monolayer or leaflet) face away from one another. As one moves from the aqueous solution to the bilayer center, a somewhat ill-defined gradient of decreasing polarity is encountered.

The formation of the bilayer is driven by the hydrophobic effect, which is conventionally thought of as primarily a decrease in water-free energy that occurs when the clustering of hydrophobic groups allows them to avoid contact with water molecules. Because of the shape of polar lipids, they usually form topologically closed bilayers, for example, a spherical or elliptical shell. Such closed structures eliminate direct contact between lipid hydrocarbon and water. They contain an internal aqueous lumen separated from the external aqueous solution by the bilayer. Artificially formed bilayers of this type are called liposomes. The bilayer plus the internally trapped aqueous solution

is known as a vesicle ([Figure 2](#)). Vesicles can contain one lipid bilayer, in which case they are said to be unilamellar vesicles, or multiple bilayers each separated by thin layers of aqueous solution, in which case they are called multilamellar vesicles. Flattened stacks of bilayers with the internal aqueous solution largely squeezed out can also be prepared.

Membrane Proteins and Bilayers

Natural membranes are associated with numerous proteins. These proteins that span the lipid bilayer are said to be transmembraneous ([Figure 2](#)). Non-transmembraneous proteins associated with the polar surface of the bilayer are also present ([Figure 2](#)). Proteins can have a significant influence on membrane structure. For example, they are likely to be responsible for the distortion of natural membranes into more complex shapes than those observed for pure lipid bilayers. Familiar examples are the membranes of red blood cells and the variable-shape membranes surrounding amoeba. On the other hand, it is clear that natural membranes contain bilayers that have physical properties similar to those of pure lipid bilayers. The fact that the lipid-facing amino acid residues of membrane proteins form hydrophobic surfaces that match lipid dimensions is no doubt an important factor explaining this behavior. Nevertheless, the physical properties of the hydrophobic segments of membrane proteins are very different than those of lipids and the consequences of these differences remain to be fully explored.

Bilayer Dimensions

Experiments on liposomes show that the diameter of spherical lipid bilayers can be any value upward from 250 Å. Bilayer width (thickness) is determined by the length of the nonpolar tails of the polar lipids and sterols. This usually ranges from 16 to 24 carbon atoms for the hydrocarbon chains of phospholipids and sphingolipids. Combined with the effects of lipid state and double bonds (see below), this results in a hydrocarbon core of the bilayer that is ~30 Å in width. Total width including the lipid headgroups is on the order of 45–50 Å, but the exact value depends on the structure of the lipid polar headgroups, which can be very large in some cases.

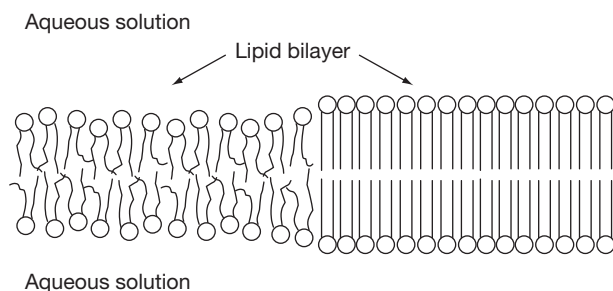


Figure 1 Schematic representation of cross section through a lipid bilayer composed of a single type of phospholipid or sphingolipid. Polar headgroups are illustrated as circles and hydrocarbon tails are shown as lines. Notice each lipid has one headgroup and two tails. A bilayer composed of both a disordered liquid domain (left) and an ordered gel domain (right) is illustrated.

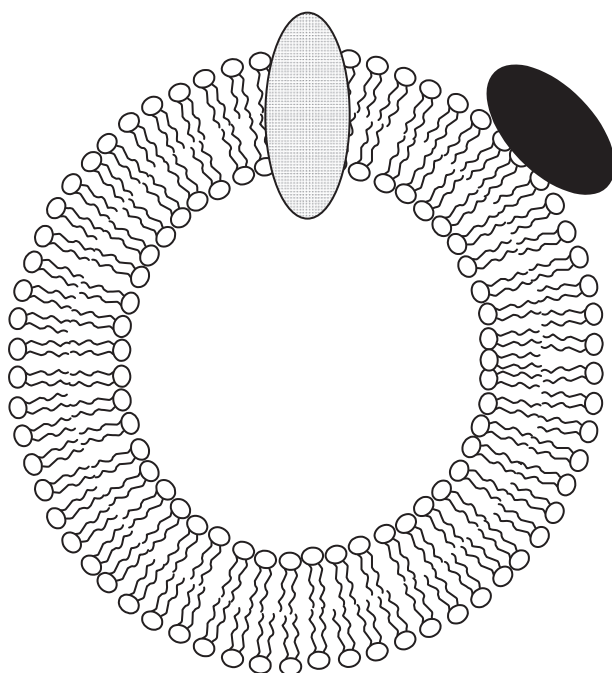


Figure 2 A unilamellar lipid vesicle formed by a lipid bilayer and containing one transmembrane protein (striped) and one protein associated with the lipid polar headgroups (filled).

Relationship between Bilayer Form and Bilayer Function

The properties of bilayers are admirably suited to their biological function. The closed structure of the bilayer allows biological membranes to completely encase cellular contents (such as nucleic acids, proteins, sugars, and amino acids) in a hydrophobic barrier that prevents them from escaping into the surrounding solution.

The width of the bilayer is such that proteins of ordinary size are able to span the lipid bilayer. This allows membrane proteins to maintain contact with both the internal and external aqueous milieu. As a result, membrane proteins

can carry out functions such as the transport of molecules across bilayers. Such abilities are central to membrane function. A bilayer that was much wider than that formed by natural lipids (i.e., one formed by much longer fatty acid chains) would require the metabolically wasteful synthesis of extra-large transmembrane proteins as well as that of longer lipids. A bilayer that was significantly thinner than that formed by natural lipids would not be stable.

Physical State of Bilayer Lipids

Another important property of lipid bilayers is their ability to exist in different physical states (or phases). At low temperatures bilayers often exist in a relatively solid-like state called the gel or $L\beta$ state or phase. At higher temperatures gel-state bilayers undergo a melting transition and the gel state is replaced by a more disordered, liquid-like state, variously designated as the liquid-crystalline, liquid-disordered, or $L\alpha$ state. Melting temperatures (or if temperature is decreased, freezing temperatures) depend on the structure of both the hydrophobic tails and the polar headgroups of the individual lipid molecules. The melting temperature of lipids with linear saturated tails, which contain only single bonds connecting carbon atoms to each other, is much higher than that of lipids with unsaturated tails, which have *cis* double bonds, as well as single bonds. Because sphingolipids tend to contain hydrocarbon chains lacking *cis* double bonds, they tend to have much higher melting temperatures than glycerophospholipids, which tend to have more highly unsaturated hydrocarbon chains.

Bilayers containing sterols can exist in a third lipid state called the liquid ordered or L_o state. This state is observed most frequently in mixtures of sterols with relatively saturated chain polar lipids. The L_o state has properties that are intermediate between the disordered liquid and gel states.

Studies in synthetic lipid bilayers show that in a single bilayer, different regions (known as domains) can exist in different physical states (Figure 1). It is believed that the disordered liquid state is by far the most common state present in biological membranes. However, in eukaryotic cells it is now thought that ordered liquid-state domains can coexist with disordered liquid domains. L_o state domains in cell membrane, called lipid rafts, are rich not only in sphingolipids and sterols, but also in special subsets of membrane proteins. Rafts appear to have a specialized role in a number of important biological processes.

The Effect of Lipid State upon Lateral Organization of Lipid Molecules

Lipid bilayers in different physical states differ in a number of properties. One is the degree of lateral interaction between various bilayer lipids. In the disordered liquid-like state, bilayers are believed to exist as a more-or-less random mixture of various lipid species. In other words, there are few or no cases in which tight associations between two different lipid molecules occur. As special constraints tend to be much more restrictive in solids, a much higher degree of lateral organization is likely to exist in a gel-state bilayer containing a mixture

of different lipids. This is little studied except to the extent that it appears that separate gel-phase domains with different lipid compositions can coexist in a single bilayer. The nature of lateral interactions between different lipids in the liquid-ordered state is being actively studied, but remains unclear. There may be specific arrangements of sterol molecules relative to the other polar lipids present in the bilayer, and various liquid-ordered domains with different compositions may be able to coexist in a single bilayer.

Interactions between each of the monolayers within a single bilayer are also poorly understood in most cases. Under some conditions certain combinations of hydrocarbon chain lengths allow interdigitation of hydrocarbon chains from opposite monolayers.

Influence of Physical State on Lipid Conformation and Lateral Diffusion

The conformation of individual lipid molecules within a bilayer is also dependent upon lipid state. In all states, the hydrocarbon chains of the lipids tend to align parallel to one another and perpendicular to the surface. A high degree of parallel alignment characterized by the presence of tightly packed lipids occurs in the gel- and liquid-ordered states. In contrast, in the disordered liquid state, lipids are more loosely packed and the degree of alignment is reduced because carbon-carbon single bonds can twist a manner that results in bending of the hydrocarbon chains. This also results in the bilayer being thinner than in other states (Figure 1). The relatively random location of such bends within a hydrocarbon chain results in increased disorder as one moves from the bilayer surface toward the bilayer center. There is also a contribution to disorder from the presence of double bonds. These are usually in the form of *cis* double bonds, which form permanent bends in hydrocarbon chains. Headgroup conformation, and its dependence on lipid state, are more poorly characterized in most cases.

Another property of bilayers that depends on lipid state is lateral diffusion. This refers to random sliding motions of lipids within the plane perpendicular to the average direction of the hydrocarbon tails. Lateral motions are rapid in both the liquid-disordered and liquid-ordered state, but much reduced in the gel state. It should be emphasized that the ability to undergo relatively rapid lateral motions is an important feature of biological membranes. It allows membrane components to rearrange within the lipid bilayer such that they can form and break functionally important complexes.

Transverse Lipid Movements

Lipids can also undergo transverse movement, in which they move (or flip) from one monolayer of the bilayer to the opposite one. For good reason this motion is many orders of magnitude slower than lateral diffusion. Transverse flipping requires that the polar/charged portions of lipids transiently dissolve within the hydrophobic core of the bilayer. This is very unfavorable energetically. However, it should be noted that in natural membranes specific flippase proteins are believed to catalyze rapid flipping events in some cases. Furthermore, it appears that spontaneous flipping is more rapid for sterol molecules, which have a headgroup that is very small and not excessively polar, than it is for most phospholipids and sphingolipids, although this remains an area of some controversy.

Effect of Lipid Bilayers on Local Aqueous Environment

Finally, it is important to remember that lipid bilayers can also affect the structure of their immediate aqueous environment. When a significant fraction of lipids in a bilayer is anionic, they tend to attract significant numbers of cations while repelling anions. As a result a special double layer of solution depleted in anions and enriched in cations surrounds the bilayer. This layer can alter the structure and function of membrane proteins.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Flippases; Glycerolipid Structure, Function, and Synthesis in Eukaryotes; Lipid Rafts; Phospholipid Synthesis in Yeast; Sphingolipid Biosynthesis; Sphingolipid Catabolism; **Metabolism Vitamins and Hormones:** Fatty Acid Structure and Synthesis; **Protein/Enzyme Structure Function and Degradation:** Lipid Modification of Proteins: Targeting to Membranes.

Further Reading

- Brown DA and London E (2000) Structure and function of sphingolipid- and cholesterol-rich rafts. *Journal of Biological Chemistry* 275: 17221–17224.
- Gennis RB (1989) *Biomembranes: Molecular Structure and Function*. New York, NY: Springer.
- Jain MK (1988) *Introduction to Biological Membranes*. New York, NY: Wiley.
- Koynova R and Caffrey MB (1998) Phases and phase transitions of the phosphatidylcholines. *Biochimica et Biophysica Acta* 1376: 91–145.
- Lee AG (1977) Lipid phase transitions and phase diagrams. *Biochimica et Biophysica Acta* 472: 237–281.

Lipid Modification of Proteins: Targeting to Membranes

M D Resh, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Farnesyl A 15-carbon isoprenoid unit attached via thioether linkage to proteins.

Fatty acylation The process whereby fatty acids are covalently attached to proteins.

Geranylgeranyl A 20-carbon isoprenoid unit attached via thioether linkage to proteins.

GPI anchor A complex structure consisting of ethanolamine phosphate, 3 mannoses, *N*-acetylglucosamine, and phosphatidylinositol.

Myristoylation The process whereby myristate, a 14-carbon fatty acid, is covalently attached to the N-terminal glycine of a protein.

Palmitoyl acyltransferase (PAT) An enzyme that catalyzes attachment of palmitate to proteins.

Palmitoylation The process whereby palmitate, a 16-carbon fatty acid, is covalently attached to a residue (usually cysteine) within a protein.

Prenylation The process whereby isoprenoid groups are covalently attached to a cysteine residue in a protein.

Rafts Membrane regions that are enriched in cholesterol and sphingolipids.

S-acylation Attachment of a fatty acid, typically palmitate, via thioester linkage to a cysteine residue within a protein.

Lipid Modification of Proteins

Three different types of lipid moieties are typically found attached to proteins: fatty acids, isoprenoids, and glycosylphosphatidyl inositol (GPI) anchors (Figure 1 and Table 1). Linkage of each of these groups occurs within different protein sequences and is catalyzed by different enzymes.

Fatty Acylation

The two major types of fatty acids that are attached to fatty acylated proteins are myristate and palmitate. Myristoylation occurs exclusively at the N-terminus, so only one myristate can be attached to a given protein. Palmitate is usually attached to proteins at internal cysteine residues. Proteins can be fatty acylated with one or several palmitates or with both myristate and palmitate.

Myristoylation

Myristate is a 14-carbon saturated fatty acid that is directly linked to the N-terminus of select proteins in a process known as N-myristoylation. Approximately 0.5% of all proteins in eukaryotes are modified in this manner; in humans, this corresponds to about 150 proteins. Nearly all N-myristoylated proteins begin with the sequence Met-Gly. During protein translation, the initiating methionine is removed by methionine aminopeptidase, exposing glycine at the N-terminus. The fatty acid is then linked to the protein via an amide bond between myristate and the N-terminal glycine residue. Additional amino acids near the N-terminus are important for determining whether a protein becomes N-myristoylated. Algorithms that predict with high accuracy whether a particular protein sequence is N-myristoylated are available on the Internet. The high stability of the amide linkage makes N-myristoylation an essentially irreversible protein modification.

N-myristoylation is catalyzed by the enzyme N-myristoyl transferase (NMT). Most organisms have one or two NMT genes, and NMT has been shown to be an essential gene product that is required for the growth of fungi, flies, and plants. The N-myristoylation reaction occurs in a cotranslational manner, that is, while the newly translated polypeptide is still attached to the ribosome. The preferred fatty acid substrate for NMT is myristoyl-CoA, an activated form of myristate. Fatty acids that are shorter or longer than myristate are generally not transferred from their respective fatty acyl-CoAs to proteins. However, NMT will transfer unsaturated 14 carbon fatty acids to proteins, a process that occurs primarily in the retina.

The requirement for a free N-terminal glycine prevents N-myristoylation from occurring on internal glycines within the polypeptide chain. However, during apoptosis, caspase-mediated protein cleavage generates newly exposed N-termini which, if they contain Gly within a consensus sequence, can serve as posttranslational targets for N-myristoylation. Proteins modified in this manner, termed 'morbid myristoylation' include BID, actin, and gelsolin.

Palmitoylation

Several hundred eukaryotic proteins are posttranslationally modified by covalent attachment of the 16-carbon fatty acid palmitate. The fatty acid is generally attached to one or more cysteine residues within the protein through a thioester linkage, and this process is referred to as S-palmitoylation. Palmitoylation can occur within a variety of types of protein sequences: (1) at or near the transmembrane domain of a membrane protein; (2) near a C-terminal prenylated cysteine (see later); (3) near the N- or C-terminus; (4) adjacent to or near an N-terminal myristoylated glycine; or (5) at an N-terminal cysteine via an amide bond. To date, there is no definitive consensus sequence for protein palmitoylation, but Web-based algorithms can be used to make predictions of potential palmitoylation sites.

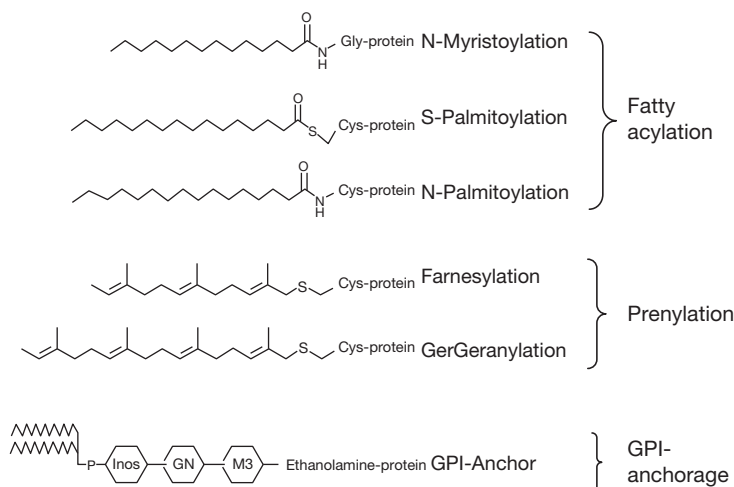


Figure 1 Structures of lipid-modifying groups on proteins. Structures of the different lipid modifying groups and their linkages to proteins are depicted. For the GPI anchor, Inos = inositol, GN = *N*-acetylglucosamine, and M3 = 3 mannoses.

Table 1 Examples of lipid-modified proteins

Fatty acylated

Myristate

Src family kinases, G α subunits, Arf, Recoverin, MARCKS, mammalian retroviral Gag proteins, calcineurin B, eNOS, cytochrome b5 reductase, AKAP

Palmitate

Transferrin receptor, CD4, caveolin-1, G-protein-coupled receptors (GPCRs), H-Ras, N-Ras, G α subunits, GAP43, PSD-95, Src family kinases, Huntingtin, eNOS, SNARE proteins

Prenylated

Farnesyl

H-Ras, K-Ras, N-Ras, prelamin A, Rheb, CENP-E, CENP-F

GeranylGeranyl

G γ subunits, Rho family proteins, Rab proteins, RalA, RalB

GPI-anchored

Alkaline phosphatase, acetylcholinesterase, CD55 (DAF), CD59, 5'-nucleotidase, Thy-1, PrP^C prion protein, trypanosomal VSG, NCAM

The fatty acid donor for protein palmitoylation is palmitoyl-CoA. In eukaryotic cells, two main families of palmitoyl acyl transferases (PATs) catalyze attachment of palmitate to proteins: DHHC PATs and MBOAT PATs. DHHC PATs are multipass membrane proteins with a cysteine-rich domain that contains a conserved sequence, Asp-His-His-Cys. Seven DHHC PATs are expressed in *S. cerevisiae* and these are responsible for nearly all the protein palmitoylation in the yeast cell. At least 23 DHHC enzymes are present in the human genome, many of which have recently shown to exhibit distinct substrate specificity. The MBOAT (membrane-bound O-acyl transferase) family contains two members, Hhat and Porc, which mediate palmitoylation of secreted proteins. Both Hhat and Porc are transmembrane proteins that are predicted to traverse the bilayer 8–12 times. Hhat palmitoylates the hedgehog family of proteins on an N-terminal cysteine that is exposed after signal-sequence cleavage. Hedgehog protein modification is unusual in two respects. First, unlike nearly all other palmitoylated proteins, the link between palmitate and hedgehog

is an amide bond. Second, hedgehog proteins undergo an additional lipid modification, covalent attachment of cholesterol to the C-terminus of the mature, processed protein product. Both of these modifications are required for hedgehog signaling function. A different MBOAT PAT, Porcupine (Porc) has been implicated in mediating palmitoylation of Wnt and Wingless proteins. At least two fatty acylation reactions have been ascribed to Porc: attachment of palmitate to a cysteine residue (Cys77 in Wnt3a) via thioester linkage, and attachment of palmitoleic acid (16:1) to a downstream serine residue (Ser209 in Wnt3a). Palmitoylation at both sites regulates Wnt secretion and signaling.

Acylation with other fatty acids

In addition to palmitate, lipidated proteins can contain other longer- or shorter-chain fatty acids. These include saturated (stearate), and unsaturated (arachidonate and palmitoleic acid) fatty acids. The alternative fatty acids may be attached at the same position as the palmitoylation site or at a different site. At least one protein, the appetite-stimulating hormone ghrelin, has been reported to be modified by attachment of the 8-carbon fatty acid oleate to a serine residue at position 3. Acylation of ghrelin is catalyzed by GOAT (Ghrelin O-acyltransferase), an MBOAT family member that shares homology with Hhat and Porc.

Prenylation

Protein prenylation is a posttranslational modification that occurs in the cytosol. In the human genome, over 200 proteins are predicted to be modified by prenylation. Prenyl groups are built from 5-carbon building blocks known as isoprene. Prenylation involves the attachment of two types of isoprenoid groups, 15-carbon farnesyl or 20-carbon geranylgeranyl, via thioether linkage to a cysteine residue at or near the C-terminus. The prenyl group donors for these reactions are farnesyl diphosphate and geranylgeranyl diphosphate, respectively. Most prenylated proteins contain a CAAX box sequence at their C-terminus (C = Cys, A = aliphatic amino acid, X = any amino acid). The identity of the C-terminal amino acid

determines whether the protein is recognized by farnesyl transferase (FTase) (X = Met, Ser, Ala, or Gln) or geranylgeranyl transferase I (GGTase I) (X = Leu or Phe). PrePS is a Web-based algorithm that can be used to predict whether a protein will be farnesylated or geranylgeranylated. FTase and GGTase I are heterodimers consisting of a common α subunit and a distinct β subunit. Following prenylation of the CAAX box, two additional modifications occur in the ER. The three C-terminal amino acids are cleaved by the endoprotease Rce1, and the prenylated C-terminal cysteine is carboxymethylated by the carboxymethyltransferase pcCMT/Icmt. A third prenyl transferase, Rab GGTase or GGTase II, attaches geranylgeranyl to Rab proteins that terminate with -CC, CXC, CCX, or CCXX sequences. Rab proteins must be bound to Rab Escort Protein (REP) in order to be prenylated by GGTase. Genetic mutations in REP cause choroideremia, a disease characterized in humans by retinal degeneration and blindness.

GPI Anchors

GPI anchors are attached to the C-termini of approximately 1% of all eukaryotic proteins. These complex structures consist, in general, of ethanolamine phosphate, three mannoses, *N*-acetylglucosamine, and phosphatidylinositol (PI). At least 20 enzymes are involved in the synthesis and attachment of GPI anchors, a process that occurs in the endoplasmic reticulum (ER) membrane. A C-terminal hydrophobic region in the target protein serves as a signal for GPI anchor attachment. This sequence is cleaved and replaced with the GPI anchor at the C-terminus in a reaction catalyzed by GPI transamidase, a complex consisting of five membrane-bound proteins. The modified protein is transported through the Golgi apparatus and ultimately to the extracellular leaflet of the plasma membrane. Putative GPI-modified proteins can be predicted using the algorithm mentioned in the 'Relevant Websites' section.

GPI anchors can undergo remodeling reactions that change the structure of the anchor. An acyl chain attached to the diacylglycerol moiety of inositol is removed and replaced with a longer-chain fatty acid. This lipid remodeling event promotes transport of the GPI-modified protein from the ER to the Golgi and is required for interaction of the protein with lipid rafts. Many GPI-anchored proteins of fungi undergo an additional processing event after delivery to the outer leaflet of the plasma membrane. The GPI anchor is cleaved between a mannose and the *N*-acetylglucosamine moiety, removing the inositol lipid portion. The remainder of the modified protein is then cross-linked to sugar residues in the cell wall.

Detection of Lipid-Modified Proteins

Mass spectroscopy analysis provides definitive identification of the modifying group on proteins. Fatty acylated and prenylated proteins can be detected by radiolabeling cells with ^3H - or ^{125}I -labeled fatty acids or isoprenoid precursors, followed by isolation of the protein with a specific antibody and autoradiographic analysis. Alternatively, chemical labeling using acyl-biotinyl exchange, azido-fatty acids, or click chemistry has recently been developed. A hallmark of many GPI-anchored proteins is that they can be released from the cell surface by *in vitro* cleavage with PI-specific phospholipase

C (PI-PLC); this process occurs *in vivo* for certain GPI-anchored proteins (e.g., alkaline phosphatase). However, other GPI-anchored proteins have palmitate attached to the inositol ring, making them resistant to PI-PLC cleavage.

Membrane Targeting of Lipid-Modified Proteins

Lipid modifications play important roles in promoting interaction of proteins with membranes. The nature of the lipid group determines how the modified protein will traffic through the cell, how tightly the protein will bind to a membrane, and which membrane it will bind to. Membrane targeting is essential for the function of many types of lipid-modified proteins. Localization of proteins to a membrane increases their local concentration and amplifies the efficiency of interaction with other membrane-bound substrates and targets. When lipid modification is inhibited, the affected proteins are mislocalized within the cell and no longer function effectively.

Membrane Binding

The presence of a molecule of palmitate, geranylgeranyl, or a GPI-anchor is sufficient for membrane binding of the modified protein. By contrast, the hydrophobicity of myristate or farnesyl is not sufficient to stably anchor a protein to the membrane. *N*-myristoylated and farnesylated proteins that do not contain a transmembrane domain need to use a second signal to promote membrane binding. The second signal is either a cluster of basic amino acids (lysine or arginine) or covalently bound palmitate contained within the protein. The positively charged basic cluster provides electrostatic interaction with negatively charged lipids, such as phosphatidylserine and phosphatidylinositols, which are enriched on the cytoplasmic surface of the plasma membrane. Synergy is achieved by combining the two forces, hydrophobic and electrostatic, resulting in enhancement of membrane binding by several orders of magnitude. Myristoylated proteins that utilize a myristate + basic motif for membrane binding include Src, MARCKS, and HIV-1 Gag, whereas K-Ras4B is the prototype for proteins with a farnesyl + basic membrane-binding mechanism.

Attachment of palmitate to a myristoylated or prenylated protein can also be used as a second signal to enhance hydrophobic interactions with the phospholipid bilayer. Examples include members of the Src family of tyrosine protein kinases, which are dually fatty acylated with myristate and palmitate, and H-Ras and N-Ras, farnesylated proteins that contain one (N-Ras) or two (H-Ras) additional palmitate moieties attached to cysteine residues just upstream from the prenylated C-terminus. Palmitoylation of these proteins is typically dependent on prior lipid modification, most likely because myristoylation or farnesylation enhances the ability of the protein to interact with a membrane-bound PAT. The combination of two lipid groups provides sufficient binding energy to promote binding of the modified protein to a membrane.

Reversible Membrane Binding and Switch Mechanisms

Several mechanisms exist *in vivo* to regulate membrane binding of lipid-modified proteins. The first is removal of the

modifying group. As indicated above, some GPI-anchored proteins can be released from the membrane after cleavage of the GPI-anchor by PI-phospholipase. S-Palmitoylation is a reversible reaction; palmitate is removed from proteins by palmitoyl protein thioesterases. The reciprocal action of PATs and thioesterases sets up a dynamic balance of palmitoylation/depalmitoylation reactions in the cell. As a consequence, proteins that are S-palmitoylated undergo reversible cycles of membrane association and dissociation.

Myristoylated and prenylated proteins contain stable lipid-protein linkages and are thus not subject to this mode of regulation. Instead, exposure of the modifying group can be regulated by a switch mechanism. For example, recoverin and Arf are N-myristoylated proteins whose membrane binding is regulated by an intramolecular myristoyl switch. These proteins exist in two conformations. In one conformation, myristate is hidden within a hydrophobic pocket inside the protein, and the protein is cytosolic. Binding of ligand triggers the myristoyl switch: a conformational change results in exposure of myristate and promotion of membrane binding. An electrostatic switch is another mechanism to regulate membrane-binding activity of either myristoylated or farnesylated proteins that contain a polybasic cluster. Phosphorylation of serine or threonine residues within the polybasic motif introduces negative charge. As a result, the electrostatic forces that were contributing to membrane binding are diminished, and the phosphorylated protein is released from the membrane. MARCKS and K-Ras4B are examples of myristoylated and farnesylated proteins, respectively, that undergo electrostatic switch-dependent release from the membrane upon phosphorylation by protein kinase C.

Switches that regulate the exposure of prenyl groups have been described, but these mechanisms involve intermolecular interactions. Rho and Rab are membrane-bound proteins that are geranylgeranylated at their C-termini. Upon interaction with specific guanine nucleotide dissociation inhibitor (GDI) proteins, the isoprene units become sequestered within a hydrophobic cleft of the GDI, inducing extraction of the Rho:GDI or Rab:GDI complex from the membrane into the cytosol.

Trafficking and Targeting to a Specific Membrane

Proteins that contain one or more transmembrane domains obviously do not require lipid modification to promote membrane binding. Rather, attachment of lipophilic groups has been shown to regulate their intracellular trafficking. Palmitoylation of G-protein-coupled receptors (GPCRs) is required for efficient delivery of these seven-transmembrane proteins to the cell surface. Nonpalmitoylated GPCRs accumulate in intracellular membranes resulting in decreased localization at the plasma membrane. For other types of receptors, for example, the AMPA receptor, palmitoylation has the opposite effect. When coexpressed with its cognate PAT (DHHC3/GODZ), the GluR subunits of the AMPA receptor are palmitoylated and retained in the Golgi, and cell-surface expression is reduced.

Proteins that are modified with either myristate or farnesyl alone have weak affinities for membranes and are randomly distributed throughout the cell. These mono-lipidated proteins are found in the cytosol as well as nonspecifically associated

with intracellular membranes. Further modification with palmitate can have a dramatic effect on protein targeting and subcellular localization. This concept is best illustrated by pioneering studies from the laboratories of Bastiens and Kenworthy on H-Ras and N-Ras proteins. These Ras proteins transit from their site of synthesis in the cytoplasm to the plasma membrane via vesicular trafficking mediated by the secretory pathway. H-Ras and N-Ras are first farnesylated in the cytosol, and then bind to the ER, where the CAAX box is further modified. Vesicular trafficking brings the proteins to the Golgi, where they are palmitoylated by a complex containing the DHHC9 PAT, and then delivers farnesylated, palmitoylated Ras proteins to the plasma membrane. Depalmitoylation at the plasma membrane occurs rapidly and results in release of Ras into the cytosol. Farnesylated (nonpalmitoylated) Ras is initially distributed randomly throughout the cell, but ultimately is concentrated or trapped at the Golgi as it becomes palmitoylated by its Golgi-associated PAT. Thus, the membrane containing the modifying PAT specifies the localization of the newly palmitoylated protein, a phenomenon known as 'kinetic bilayer trapping'. If palmitoylation is inhibited by addition of 2-bromopalmitate, or if depalmitoylation is blocked by attachment of a noncleavage palmitate analog to Ras, then Ras becomes mislocalized to all internal membranes.

There are now multiple examples of proteins that undergo reversible shuttling between the plasma membrane and the nucleus in a palmitoylation-dependent manner. RGS7 is a plasma-membrane-associated protein that regulates G-protein signaling. RGS7 location is regulated by binding to R7BP, a palmitoylated protein. Upon depalmitoylation of R7BP, the RGS7:R7BP complex is released from the plasma membrane and translocates into the nucleus. A polybasic sequence within R7BP functions as a nuclear localization signal (NLS). Likewise, Wrch-1, a Cdc42-like GTPase, contains a C-terminal polybasic region that also contains palmitoylation sites. Translocation of Wrch-1 between plasma membrane and nucleus is regulated by palmitoylation and depalmitoylation reactions.

The ultimate location and fate of a protein may be determined by competition between lipid modification and another signal or posttranslational modification. Here are three examples of reactions that compete with palmitoylation: (1) as described above, when a palmitoylation site lies in close proximity to an NLS, attachment of palmitate may mask the NLS and prevent access to nuclear import receptors; (2) for the yeast SNARE protein Tlg 1 and anthrax toxin receptors, palmitoylation protects the protein from ubiquitination and degradation; and (3) GPCRs have palmitoylation sites close to the phosphorylation sites that regulate receptor desensitization. Palmitoylation reduces receptor phosphorylation. In all three instances, depalmitoylation exposes an additional signal (NLS) or site (ubiquitination or phosphorylation), thereby mediating an alternate fate for the delipidated protein.

Targeting to Membrane Rafts and Other Specialized Membrane Regions

Membrane rafts are small subdomains of the plasma membrane that are enriched in cholesterol and sphingolipids. The lipids in rafts predominantly contain saturated fatty acid chains, which promote tight packing and formation of a

specialized region of the membrane bilayer known as a 'liquid-ordered' domain. Proteins modified by GPI-anchors or saturated fatty acids preferentially localize to rafts. The saturated nature of the fatty acid chain is an important factor in promoting raft association; replacement with a *cis*-unsaturated fatty acid results in displacement of the modified protein from rafts. Myristoylated and palmitoylated proteins localize to the cytoplasmic side of plasma membrane rafts, while GPI-anchored proteins are sorted into raft domains in the *trans*-Golgi network and then trafficked to the exoplasmic leaflet of the plasma membrane. Raft localization of these proteins enhances protein-protein interactions and is important for propagating efficient signal transduction by growth factor and immune cell receptors.

Lipid modification can also serve to direct proteins to other types of distinct membrane subdomains. In polarized cells such as epithelial cells, the plasma membrane is divided into two distinct domains – apical and basolateral – that differ in protein and lipid composition. The GPI-anchor provides a sorting signal that directs the modified protein to the apical surface in most epithelial cells. Palmitoylation has also been shown to regulate protein sorting within membranes. For example, attachment of palmitate dictates targeting to the axons or to the postsynaptic density in neurons, to the myelin membrane in oligodendrocytes, or to tight junctions in epithelial cells.

Inhibition of Lipid Modification: Clinical Implications

The importance of lipid modification for correct protein localization and function has led to development of inhibitors that specifically block attachment of individual hydrophobic groups. Compounds that interfere with N-myristoylation (2-hydroxymyristate) or palmitoylation (2-bromopalmitate) have been widely used in the laboratory to inhibit signal-transduction pathways dependent on fatty acylated proteins. A number of disease states are regulated by lipid-modified proteins, and drugs that inhibit lipid modification are currently in use in the clinic or are in the pipeline. For example, fungal infections are being combated with NMT inhibitors. The substrate-binding properties of fungal NMT are sufficiently different from human NMT to allow development of inhibitors that function as antifungal agents. There is great clinical interest in prenylation inhibitors (farnesyl transferase inhibitors (FTIs) and geranylgeranyl transferase inhibitors (GGTIs)), which are designed to inhibit prenylation of Ras and oncogenic forms of Rho and reduce the growth of Ras-induced human tumors. FTIs have also been shown to prevent cardiovascular disease in a mouse model of progeria, a premature aging syndrome that is driven by abnormal processing of farnesylated lamin A.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Lipid Bilayer Structure; **Protein/Enzyme Structure Function and Degradation:** Protein N-Myristoylation; Protein Palmitoylation.

Further Reading

- Bijlmakers MJ and Marsh M (2003) The on–off story of protein palmitoylation. *Trends in Cell Biology* 13: 32–42.
- Escribá PV, Wedegaertner PB, Goñi FM, and Vögler O (2007) Lipid–protein interactions in GPCR-associated signaling. *Biochimica et Biophysica Acta* 1768: 836–852.
- Farazi TA, Waksman G, and Gordon JL (2001) The biology and enzymology of protein N-myristoylation. *Journal of Biological Chemistry* 276: 39501–39504.
- Goodwin JS, Drake KR, Rogers C, et al. (2005) Depalmitoylated Ras traffics to and from the Golgi complex via a nonvesicular pathway. *Journal of Cell Biology* 170: 261–272.
- Liang X, Nazarian A, Erdjument-Bromage H, et al. (2001) Heterogeneous fatty acylation of Src family kinases with polyunsaturated fatty acids regulates raft localization and signal transduction. *Journal of Biological Chemistry* 276: 30987–30994.
- Linder ME and Deschenes RJ (2007) Palmitoylation: Policing protein stability and traffic. *Nature Reviews in Molecular Cell Biology* 8: 74–84.
- Orlean P and Menon AK (2007) Lipid posttranslational modifications. GPI anchoring of protein in yeast and mammalian cells, or how we learned to stop worrying and love glycosphospholipids. *Journal of Lipid Research* 48: 993–1011.
- Paulick MG and Bertozzi CR (2008) The glycosylphosphatidylinositol anchor: A complex membrane-anchoring structure for proteins. *Biochemistry* 47: 6991–7000.
- Pechlivanis M and Kuhlmann J (2006) Hydrophobic modifications of Ras proteins by isoprenoid groups and fatty acids – more than just membrane anchoring. *Biochimica et Biophysica Acta* 1764: 1914–1931.
- Peitzsch RM and McLaughlin S (1993) Binding of acylated peptides and fatty acids to phospholipid vesicles: Pertinence to myristoylated proteins. *Biochemistry* 32: 10436–10443.
- Planey SL and Zacharias DA (2009) Palmitoyl acyltransferases, their substrates, and novel assays to connect them. *Molecular Membrane Biology* 26: 14–31.
- Resh MD (2006a) Palmitoylation of ligands, receptors, and intracellular signaling molecules. *Science Signaling: The Signal Transduction Knowledge Environment* 359: re14.
- Resh MD (2006b) Trafficking and signaling by fatty acylated and prenylated proteins. *Nature Chemical Biology* 2: 584–590.
- Rocks O, Peyker A, Kahms M, et al. (2005) An acylation cycle regulates localization and activity of palmitoylated Ras isoforms. *Science* 307: 1746–1752.
- Webb Y, Hermida-Matsumoto L, and Resh MD (2000) Inhibition of protein palmitoylation, raft localization, and T cell signaling by 2-bromopalmitate and polyunsaturated fatty acids. *Journal of Biological Chemistry* 275: 261–270.

Relevant Websites

- <http://plantsp.genomics.purdue.edu> – A plant-specific myristoylation predictor.
- <http://csspalm.biocuckoo.org> – CSS Palm – Prediction of Palmitoylation Sites.
- <http://mendel.imp.ac.at> – GPI Lipid Anchor Project.
- <http://www.bioinfo.tsinghua.edu.cn> – NBA-Palm: Prediction of Palmitoylation Sites.
- <http://mendel.imp.ac.at> – NMT: The MYR Predictor.
- <http://mendel.imp.ac.at> – PrePS: Prenylation Prediction Suite.

Lipid Rafts

D A Brown, State University of New York, Stony Brook, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Caveolae 50–100 nm plasma membrane pits coated with tightly packed oligomers of caveolin-1 and/or other caveolin proteins.

Ganglioside Acidic glycosphingolipid containing one or more sialic acid moieties.

Glycosphingolipid Sphingolipid with one or more sugars as a head group.

GPI-anchored protein Protein anchored on the outside of the plasma membrane by covalent linkage to glycosyl phosphatidylinositol.

Lipid raft Membrane microdomain in the liquid-ordered (l_o) phase; in cell membranes, rich in sterol and sphingolipids.

Sphingolipid Membrane lipid with a ceramide backbone consisting of a sphingoid base backbone and a long saturated acyl chain, and a polar head group.

Introduction

Lipids do not always mix uniformly in membranes, but can cluster to form microdomains. A certain class of these microdomains has been termed 'lipid rafts'. These are enriched in cholesterol and sphingolipids. Rafts probably exist in membranes in the liquid-ordered phase or a phase with similar properties. Increasing evidence suggests that phospholipid-rich liquid-crystalline phase domains and sphingolipid-rich liquid-ordered phase domains (rafts) can exist in equilibrium in biological membranes, especially the plasma membrane. Preferential partitioning of membrane proteins into rafts can affect function. Among the proteins that are targeted to rafts are those anchored in the outer leaflet of the membrane through covalent attachment to a special glycolipid, glycosyl phosphatidylinositol (GPI). Other proteins that are linked to saturated acyl chains, such as those that are directly acylated with two or more palmitate chains, or a palmitate and a myristate chain, are also targeted to rafts. Targeting of GPI-anchored proteins and other proteins to rafts plays a role in signal transduction, especially in hematopoietic cells, and possibly also in sorting in intracellular membranes and regulation of cell-surface proteolysis in other mammalian cells.

Physical Properties

The fact that membrane lipids have such heterogeneous structures suggests that they may mix nonrandomly in the bilayer to form lipid microdomains. However, despite much interest in the subject, good evidence that lipids cluster in cell membranes has been slow to emerge. We will discuss one type of microdomain that has recently been described in cell membranes. These are rich in cholesterol and sphingolipids, and probably exist in membranes in the liquid-ordered (l_o) phase or a phase with similar properties. These domains have also been called rafts. The fact that rafts are in the l_o phase gives them a surprising but technically useful property; they are resistant to solubilization by nonionic detergents such as Triton X-100 in the cold. Thus, after extraction of cells in detergent, insoluble rafts can be separated from fully solubilized nonraft proteins and lipids, and isolated for further analysis. We will use the

term 'raft' to refer to the domain in the intact membrane, and the term 'detergent-resistant membrane' (DRM) to refer to the structure isolated by detergent insolubility.

Certain proteins, including those anchored in the membrane by GPI and those linked to dual saturated acyl chains such as palmitate and myristate, are targeted to rafts. We discuss GPI-anchorage, an unusual form of membrane anchorage, in detail, covering synthesis and the types of proteins that are GPI-anchored. We also discuss the functions of GPI-anchorage, focusing on the ability of this modification to target proteins to lipid rafts (Figure 1).

Lipid-Phase Behavior

Since rafts are probably in a separate phase from the rest of the bilayer, we briefly review membrane lipid-phase behavior. Artificial bilayers consisting of one pure phospholipid species exist in a frozen, ordered, gel phase at low temperatures. Acyl chains of gel-phase lipids are extended and packed tightly together, and lipids have low rates of rotational mobility and diffusion in the bilayer. These frozen acyl chains have a melting temperature (T_m) that is characteristic of each phospholipid species. Above this temperature, the bilayer is present in a phase, termed 'liquid-crystalline (l_c)' or 'liquid-disordered (l_d)', in which the lipid acyl chains are fluid and disordered. In the l_c phase, the acyl chains are disordered and do not pack tightly together. They also have high rates of rotational and lateral mobility in the bilayer. Binary mixtures of phospholipids with different T_m can be examined at temperatures between the T_m of the two lipids. When one component is present at low levels, the mixture is uniform, and generally remains in the phase favored by the major component. Above a threshold concentration of one component, cooperative phase separation occurs, and gel and l_c phase domains can coexist. In these two-phase systems, the gel phase is enriched in the high- T_m lipid, and the l_c phase is enriched in the low- T_m lipid. Phase separation in these mixtures is affected by temperature; high temperatures favor the l_c phase, while low temperatures favor the gel phase. Whenever two phases are present, the lipid composition of each phase remains constant. Thus, as the lipid composition and the temperature are changed, only the

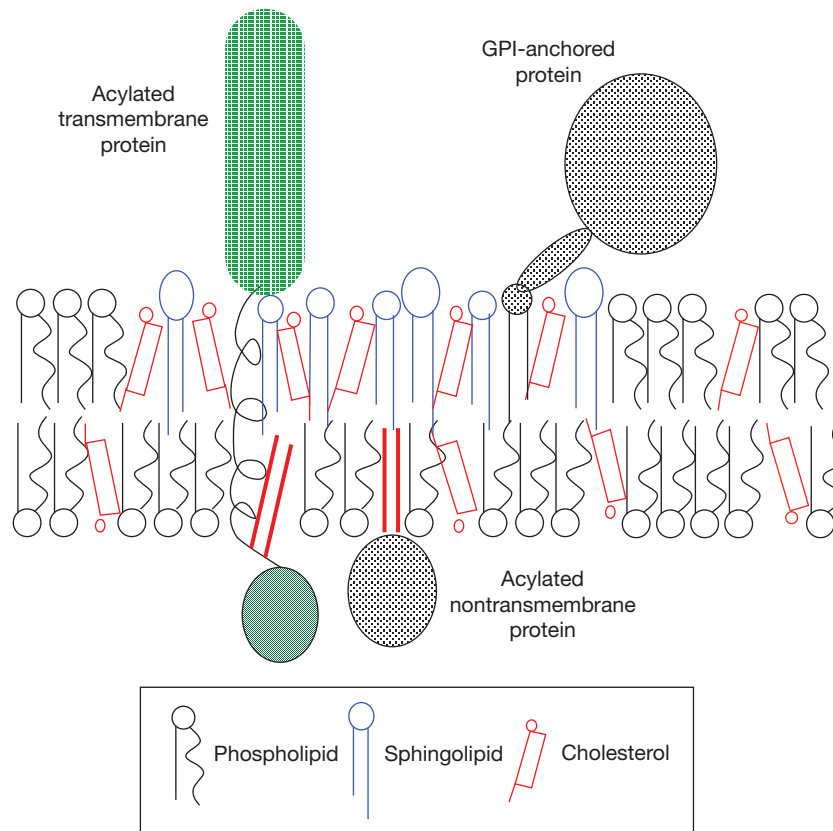


Figure 1 Diagram of a membrane raft. Rafts form in membranes rich in cholesterol or other sterols (red) and sphingolipids (blue). Rafts are enriched in these lipids, and relatively depleted in phospholipids (black). Most phospholipids in biological membranes contain one unsaturated acyl chain (curvy line). Lipids can form rafts spontaneously in model membranes, though proteins may affect raft formation in cell membranes. Most sphingolipids are present in the outer leaflet of the plasma membrane. The lipid composition of inner leaflet rafts is not known. Proteins with a high affinity for the ordered lipid environment present in rafts can concentrate there. These include GPI-anchored proteins (whose phosphatidyl inositol moieties generally contain two saturated acyl chains), acylated cytoplasmic proteins such as Src-family tyrosine kinases and α subunits of heterotrimeric G proteins, and some acylated transmembrane proteins (which are generally modified with palmitate chains). Saturated acyl chain lipid modifications on proteins are represented by thick red lines.

relative amount of the two different phases in the membrane changes.

Eukaryotic cell membranes contain mixtures of glycerolipids (in mammalian cells, all phospholipids except sphingomyelin), sphingolipids, and sterols. Biological glycerolipids generally have very low T_m , while sphingolipids (especially glycosphingolipids) have much higher T_m . This disparity has suggested that phase separation between glycerolipid- and sphingolipid-rich domains (possibly in the l_c and gel phases, respectively) might occur in membranes in cells. Because lipids in the gel phase are insoluble in detergent, DRMs might be gel-phase domains. However, although phase separation between l_c and gel phases has been well characterized in model membranes, the gel phase does not appear to exist in biological membranes, except in unusual cases. In addition, DRMs are rich in cholesterol, making it unlikely that they are present in the gel phase. However, another type of phase separation has also been described in model membranes. This is phase separation between two fluid phases in binary mixtures of a saturated-chain phospholipid (which has a high T_m) and cholesterol. In this case, above the T_m of the phospholipid, a liquid-ordered (l_o) phase can separate from the l_d phase as

the amount of cholesterol is increased. Acyl chains of lipids in the l_o phase have properties that are intermediate between those of the gel and l_d phases. They are extended, ordered, and tightly packed together, as in the gel phase, but have high lateral mobility in the bilayer, as in the l_d phase. Phase separation between the l_d and l_o phases also occurs in model membranes containing ternary mixtures of high- and low- T_m lipids and cholesterol, although the process is more complex. In particular, l_c/l_d phase separation occurs at 37 °C in mixtures of phospholipids, sphingolipids, and cholesterol in proportions similar to those found in the plasma membrane. This result is important because it shows that the formation of l_o phase domains in the plasma membrane is plausible from a biophysical standpoint.

DRMs Are in the l_o Phase

Several studies suggest that DRMs isolated from cells exist in intact membranes as domains in the l_o phase. First, DRMs are in the l_o phase when they are isolated. Second, controlled experiments have addressed the possibility of detergents

introducing artifacts. One such artifact is that lipids in membranes might be in a uniform phase, with a concentration of sphingolipids (high- T_m lipids) too low to induce phase separation. However, detergent might preferentially extract low- T_m phospholipids first. As phospholipids are gradually removed, the concentration of sphingolipids in the membrane might gradually increase. This process might eventually lead to phase separation when a threshold concentration of sphingolipids is reached. Detergents do not act in this manner, as shown by model membrane experiments. A series of liposomes whose phase behavior is known from methods independent of detergent extraction is made. Some of these liposomes contain sphingolipids and cholesterol, but at levels too low for phase separation. When these liposomes are extracted with detergent, no DRMs are produced. This shows that detergent does not rearrange membrane lipids to induce DRM formation.

Two different models of raft formation have been proposed. The first model, supported by most data, suggests that rafts form by phase separation. Thus, structural features of lipids that affect T_m and promote phase separation will also promote raft formation. One key difference between phospholipids and sphingolipids is the length and saturation of their acyl chains. Lipid head-groups can also contribute to T_m by altering the way lipids pack together. According to our model, acyl chain packing plays a key role in raft formation. Head group interactions might also be important, although these interactions may affect raft formation indirectly by affecting acyl chain packing. Cholesterol is recruited to rafts due to its ability to pack tightly with lipids of high T_m when the l_o phase is formed. Another group has proposed a different model for raft structure. According to the second model, rafts are primarily clusters of glycosphingolipids held together through hydrogen bonding or other interactions between glycosphingolipid head groups. Because cholesterol has a small head group, it is recruited to rafts because it fits well in the gaps between bulky headed glycosphingolipids.

Proteins in DRMs

The physiological significance of raft formation is likely to stem from the preferential partitioning of certain proteins into rafts. It has been known for several years that proteins anchored in membranes by GPI are detergent-insoluble and are present in DRMs. Acylated proteins, which are covalently linked to fatty acids, can also be targeted to DRMs. In both cases, this is likely to occur through partitioning of order-preferring lipid modifications into rafts. Thus, proteins that have an affinity for an ordered lipid phase will partition into rafts.

GPI-anchorage can target proteins to rafts. It has become clear in the last few years that targeting to rafts is an essentially universal property of GPI-anchored proteins in mammalian cells. Proposed function(s) of this targeting, which is still being elucidated, includes sorting in polarized epithelial cells, signal transduction, and regulation of cell-surface hydrolytic activity. GPI-anchored proteins have long been known to be insoluble in nonionic detergents such as Triton X-100 in the cold. The basis of this property was not understood until recently. Integral membrane proteins are generally detergent soluble

unless they are linked to cytoskeleton. However, GPI-anchored proteins do not have access to cytoskeleton, as they are located in the outer leaflet of the plasma membrane. Detergent-insoluble GPI-anchored proteins are enriched in DRMs. This finding led us to postulate that partitioning of the saturated acyl chains on GPI-anchored proteins into the l_o phase targets the proteins to rafts. Alternatively, protein-protein contacts might be required for detergent insolubility of GPI-anchored proteins. To test these possibilities, we purified a GPI-anchored protein, placental alkaline phosphatase (PLAP), and incorporated it into sphingolipid/cholesterol-rich liposomes (SCRL) that (as discussed above) probably contain both detergent-soluble l_d and detergent-resistant l_o phase domains. When we incorporated PLAP into these liposomes, subjected them to detergent extraction, and separated DRMs and the Triton-soluble fractions, we found that PLAP was associated with the DRMs. This result shows that the protein interacts directly with lipids in the l_o -phase. Other GPI-anchored proteins behave the same way, while a control transmembrane protein is fully solubilized.

Functions of Rafts

Rafts have been proposed to form in the *trans*-Golgi network, and to function during sorting of apical and basolateral proteins into separate transport vesicles in polarized epithelial cells. In what appears to be an analogous process, rafts have also been implicated in axonal versus dendritic sorting of proteins in neurons. Rafts are required for infection of mammalian cells by a number of viruses, including HIV, measles, and Ebola virus. They are also required for both entry of viruses into target cells and budding of mature virions from infected cells. However, the best-characterized functions of rafts are in signaling, especially in hematopoietic cells, and in membrane pits called caveolae, as described next.

Signaling in Hematopoietic Cells

Signaling through the antigen receptors in both T cells and basophils also occurs in rafts. For instance, in T cells, the T-cell receptor (TCR), associated kinases Lck and Fyn, the downstream adaptor protein LAT, and the costimulatory protein CD28 are present in rafts during signaling. Mutations in Lck and LAT that prevent association of these proteins with rafts affect signaling through the TCR, as do agents that disrupt rafts. These results show that rafts play an important role in signaling in T cells. This novel mechanism of regulating interactions between proteins by selective partitioning into membrane domains is unlikely to be restricted to one cell type. Further work is likely to demonstrate the importance of rafts in signaling in other cell types as well.

Caveolae

Caveolae are 50–100 nm pits in the plasma membrane of most mammalian cells. The membrane protein caveolin-1 plays an important structural role in caveolae, and can induce their formation when expressed in cells that lack caveolae. Unlike

clathrin-coated pits, whose formation is induced by transient recruitment of soluble clathrin and adaptor complex coat to membranes, caveolin-1 is an integral membrane protein, and remains embedded in the membrane. Caveolin-1 (22 kDa) forms high-molecular-weight oligomers (400–600 kDa). These oligomers associate with each other in the plane of the membrane in caveolae. Several pieces of evidence show that lipid rafts are concentrated in caveolae. For instance, GPI-anchored proteins have a high affinity for caveolae and will cluster there when cross-linked by antibodies. Gangliosides, acidic glycosphingolipids that are important components of lipid rafts, show similar behavior. Caveolin-1 itself has a high affinity for rafts, and is enriched in DRMs.

Several functions have been ascribed to caveolae. They appear to be signaling centers, where a number of signaling proteins are concentrated. A role for caveolae in intracellular cholesterol transport has also been proposed. Caveolin-1 is a cholesterol-binding protein, which may contribute to a role in cholesterol homeostasis. Caveolae can also internalize from the cell surface. In endothelial cells, caveolae may participate in transport of albumin across the cell monolayer. At least some pathogens have co-opted this pathway. SV40 virus and some pathogenic *Escherichia coli* bacteria enter mammalian cells through caveolae during infection.

See also: **Bioenergetics:** V-ATPases; **Lipids Carbohydrates Membranes and Membrane Proteins:** Glycosylphosphatidylinositol Anchors; Lipid Bilayer Structure; **Metabolism Vitamins and Hormones:** Adipocyte Fat Mobilization: The Regulatory Roles of Perilipin 1, Adipose Triglyceride Lipase, and Hormone-Sensitive Lipase; Vitamin D; **Signaling:** Src Family of Protein Tyrosine Kinases.

Further Reading

- Brown DA and London E (1998) Structure and origin of ordered lipid domains in biological membranes. *Journal of Membrane Biology* 164: 103–114.
- Harder T (2001) Raft membrane domains and immunoreceptor functions. *Advanced Immunology* 77: 45–92.
- Jacobson K and Dietrich C (1999) Looking at lipid rafts? *Trends in Cell Biology* 9: 87–91.
- Mukherjee S and Maxfield FR (2000) Role of membrane organization and membrane domains in endocytic lipid trafficking. *Traffic* 1: 203–211.
- Simons K and Ikonen E (1997) Functional rafts in cell membranes. *Nature* 387: 569–572.
- Smart EJ, Graf GA, McNiven MA, et al. (1999) Caveolins, liquid-ordered domains, and signal transduction. *Molecular and Cellular Biology* 19: 7289–7304.
- van der Goot FG and Harder T (2001) Raft membrane domains: From a liquid-ordered membrane phase to a site of pathogen attack. *Seminars in Immunology* 13: 89–97.

Lipid Signaling and Ion Channels

J B Jensen, B H Falkenburger, and B Hille, University of Washington School of Medicine, Seattle, WA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Affinity A measure of the strength of binding between two molecules, measured, for example, by radioligand binding. Distinguished from apparent affinity, which is based on measurements of effects rather than chemical binding.

Excised patch A patch clamp configuration for measuring currents through ion channels wherein a glass pipette is used to pull a piece of plasma membrane off a cell, exposing the inner surface of the membrane to the bathing medium.

Förster resonance energy transfer (FRET) The radiationless transfer of energy between two fluorophores that occurs when: (1) the emission spectrum of the donor fluorophore overlaps the excitation spectrum of the acceptor fluorophore and (2) the two fluorophores are in sufficient molecular proximity. Used to assess proximity between two molecules, such as fluorophore-labeled pleckstrin homology domains. Sometimes called fluorescence resonance energy transfer.

Homotetramer A protein complex (such as an ion channel) consisting of four identical subunits. A heterotetramer would consist of four nonidentical subunits.

Inward rectifier A potassium channel type that conducts K^+ ions more readily in the inward direction than in the outward direction.

Ion channel An integral membrane protein (or protein complex) that forms a pore permeable to ions. Ions permeate through ion channels following their electrochemical gradient.

Pleckstrin homology (PH) domain A fold occurring in pleckstrin and several other proteins that, in many cases, binds phosphoinositides such as PIP_2 . In the context of this article, we refer to the PIP_2 -binding PH domain from phospholipase C δ_1 .

Transporter An integral membrane protein that passes ions against their electrochemical gradient by using energy from adenosine triphosphate (ATP) (primary active transport) or from the electrochemical gradient of another ion (secondary active transport). To be distinguished from ion channels, which permit the passive flow of ions along their concentration gradients.

Voltage-sensitive phosphatase A recently discovered integral membrane protein consisting of a voltage sensor linked to a phosphoinositide 5-phosphatase. The phosphatase is activated by membrane depolarization.

Introduction

Membrane lipids have a structural role, forming the fluid-insulating bilayer that supports membrane proteins. In addition, certain minority lipids have informational roles. They turn over constitutively, and in response to stimuli, they regulate activities of many membrane proteins, and they recruit cytoplasmic proteins to the membrane interface. Prominent among the informational lipids are the phosphoinositides.

Diversity and Localization of Phosphoinositides

Phosphoinositides are phospholipids that reside in the cytoplasmic leaflet of the plasma membrane and of organellar membranes. They interact with membrane proteins, including ion channels. Phosphoinositides have the cyclic six-carbon inositol head group linked through phosphoglycerol to two fatty acid tails that extend into the membrane, typically arachidonic acid at the 2-position of glycerol and stearic acid at the 1-position. The inositol ring can be phosphorylated at positions D-3, D-4, and D-5, generating seven phosphorylated phosphoinositide species from bare phosphatidylinositol (PI): singly phosphorylated $PI(3)P$, $PI(4)P$, and $PI(5)P$; doubly phosphorylated $PI(3,4)P_2$, $PI(3,5)P_2$, and $PI(4,5)P_2$; and triply phosphorylated $PI(3,4,5)P_3$. Each phosphoinositide species

is recognized by different proteins. Phosphoinositides are converted from one species to another by lipid kinases and lipid phosphatases. They move between cellular membranes during membrane trafficking and some are transported by phospholipid transfer proteins. Compared to PI, and especially the other kinds of phospholipids of the cell, all polyphosphoinositides have low abundance and relatively fast turnover.

Different cellular membranes have distinct characteristic phosphoinositide species. PI is synthesized in the endoplasmic reticulum; $PI(4)P$ is present in the Golgi complex, secretory vesicles, and probably on the plasma membrane; $PI(4,5)P_2$ and PIP_3 populate the plasma membrane; $PI(3)P$, $PI(3,4)P_2$, and $PI(3,5)P_2$ are present in the endosomal pathway; and several species may be present in the nucleus. These phosphoinositide 'signatures' are maintained by enzyme compartmentalization. The pool of $PI(4,5)P_2$ is tonically present in the plasma membrane, whereas PIP_3 is produced only transiently when PI 3-kinases are activated by receptors.

Phosphoinositides have diverse functions – they are involved in membrane budding and fusion, cytoskeletal dynamics, control of nuclear function, and assembly of signaling proteins at membranes. They also regulate the activity of plasma membrane ion channels and transporters. Phosphoinositides gain all these functions through specific lipid–protein interactions. As $PI(4,5)P_2$ is the best-studied plasma membrane phosphoinositide,

Table 1 Tools and approaches to probe phosphoinositide action on ion channels

Where and how abundant are phosphoinositides?

- Phosphoinositide-binding domains fused to fluorescent proteins can report location and spatiotemporal changes within living cells.
- Biochemical methods (membrane extraction and chromatography) in populations of cells can give absolute amounts of phosphoinositide species (normalized to anionic lipids or protein).
- Ionic current through known PIP₂-sensitive ion channels can report relative changes in PIP₂.

Do changes in phosphoinositide levels affect ion channel activity?

- Add phosphoinositides:
 - Soluble analogs (e.g., short-chain PIP₂) to inside-out membrane patches.
 - Carrier shuttle protein to escort lipid into cell.
- Chelate/sequester phosphoinositides from cytoplasmic side:
 - Polycations (neomycin, polymycin, spermine, and Mg²⁺).
 - Antibodies.
 - Binding proteins (palmitoylated peptides, pleckstrin homology (PH) domains, and MARCKS proteins).
- Stimulate lipid enzymes:
 - Overexpress phosphoinositide kinases or phosphatases.
 - Stimulate upstream receptors or raise Ca²⁺ to activate PLC.
 - Induce membrane targeting of enzymes using dimerization of FKBP–FRB by rapamycin.
 - Express voltage-activated phosphatase (VSP, a PI 5-phosphatase).
- Inhibit lipid enzymes:
 - RNA interference (established for several PI 5-kinase isoforms).
 - Nonhydrolyzable ATP analogs (stops kinases only).
 - Pharmacological blockers (10 μ M phenylarsine oxide for PI4KIII α , 10 μ M wortmannin for PI4KIII α , and β ; several PI3-kinase inhibitors).

What is the affinity of channels for phosphoinositides?

- Concentration–response relationship applying soluble PIP₂ analogs on excised membrane patches.
- Biochemical methods (PIP strips, pull-down assays).

What regions of the channel are important for interaction with phosphoinositides?

- Mutate individual amino acids and test responsiveness or affinity to phosphoinositides.
- Solve ion-channel crystal structure; model interaction with phosphoinositides.
- Determine crystal structure of channel with the lipid.

the further discussion emphasizes this molecule, referred to hereafter as PIP₂. Levels of PIP₂ can be downregulated dynamically by physiological stimulation of receptors that activate PIP₂-selective phospholipase C (PLC).

Ion Channels and Transporters Modulated by Phosphoinositides

Phosphoinositide modulation of membrane proteins has been explored by the methods outlined in Table 1. The experiments have shown that several classes of plasma membrane ion channels are regulated by PIP₂. Most of these channels conduct less current when PIP₂ is depleted, but a few conduct more. So far, the list includes: inward rectifier potassium channels, Kir1 (ROMK), Kir2 (IRK), Kir3 (GIRK), and Kir6 (K_{ATP}); voltage-gated potassium channels, especially Kv7 (KCNQ); voltage-gated calcium channels (Ca_v); most transient receptor potential channels (TRP); the epithelial sodium channel (ENaC); possibly the cystic fibrosis transmembrane conductance regulator; and connexin43. To illustrate with one example: KCNQ K⁺ channels can be turned off more than 90% by activating the enzymes PLC or PIP₂ 5-phosphatase (Figure 1) and by chelating PIP₂ with polycations; the recovery of current after PLC activation is prevented when there is no hydrolyzable adenosine triphosphate (ATP) or when PI 4-kinase is blocked; the recovery is speeded by overexpression of PI 4-kinase or PIP 5-kinase, or by applying soluble forms of PIP₂ to the cytoplasmic face of the membrane.

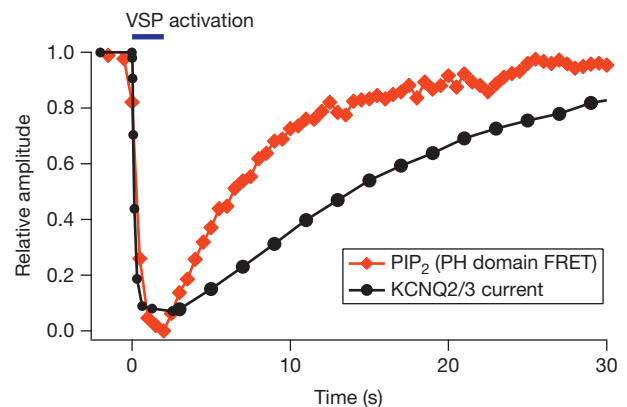


Figure 1 PIP₂ is required for KCNQ2/3 channel activity. This experiment uses expression of a unique voltage-sensitive phosphatase which is activated by large membrane depolarizations. Activation of the voltage-sensitive 5-phosphatase depletes PIP₂ (as reported by loss of PH domain FRET) and suppresses whole-cell KCNQ current recorded by patch clamp. Current records averaging data from several cells were obtained by the methods of Falkenburger BH, Jensen JB, and Hille B (2010) Kinetics of PIP₂ metabolism and KCNQ2/3 channel regulation studied with a voltage-sensitive phosphatase in living cells. *Journal of General Physiology* 135: 99–114.

Phosphoinositides also regulate transporters and pumps: the sodium/calcium exchanger (NCX); the plasma membrane calcium adenosine triphosphatase (ATPase); and the sodium/proton exchanger. These transporters slow or stop if PIP₂

is depleted. NCX was one of the first membrane proteins for which slowing by PIP₂ depletion was shown. Here, we consider transporters collectively with ion channels for simplicity.

Features of Phosphoinositide Modulation of Ion Channels

PIP₂ Sensitivity and Apparent Affinity

Ion channels differ in their sensitivity to PIP₂. In several channels, the open probability can be increased from virtually zero to almost unity by exposing PIP₂-depleted membranes to high concentrations of soluble PIP₂. The midpoint of this concentration–response curve defines the apparent affinity of the channel for PIP₂. Differences in PIP₂ affinity among channel subtypes are physiologically relevant. At resting levels of PIP₂ in a cell, high-affinity channels will be almost saturated with PIP₂, whereas low-affinity channels will have fewer PIP₂-binding sites occupied. Given similar channel densities, current through the high-affinity channels will be larger. For example, among inward rectifier potassium channels, IRK and ROMK channels are distinguished by their constitutive activity resulting from high affinity for PIP₂. Other inward rectifier channels, such as GIRK, have low resting PIP₂ affinity and low constitutive activity. Activation of these channels requires binding of G-protein $\beta\gamma$ subunits, which increase channel affinity for PIP₂.

High- and low-affinity channels also differ in their response to changes in PIP₂ levels. Low-affinity channels let go of their PIP₂ more easily than high-affinity channels. PIP₂ levels can fall quickly during activation of PLC by receptors and are changed by other physiological stimuli as well. With falling PIP₂ levels, currents through low-affinity channels will drop more quickly and more thoroughly than currents through high-affinity channels. Thus, less PIP₂ depletion is required to inhibit 50% of resting current for low-affinity channels than for high-affinity channels. For example, heterotetrameric KCNQ2 or KCNQ4 channels have a 100-fold lower apparent affinity for PIP₂ than do heterotetrameric KCNQ3 channels, making them much more responsive to PLC-induced PIP₂ depletion. Choosing high-affinity versus low-affinity channels is one way by which cells can tune their responsiveness to PLC-coupled receptors.

With most means of PIP₂ depletion, physiological or experimental, PIP₂ levels do not fall to zero, but to a new lower steady state that balances PIP₂ synthesis and depletion. At this steady state, a substantial proportion of high-affinity channels still might have some PIP₂ bound, so that current inhibition is incomplete. Experimentally, it can be difficult to determine whether residual current in PIP₂-sensitive channels is a function of incomplete PIP₂ depletion or of partial channel function in the absence of PIP₂. Other acidic phospholipids might support partial channel activity.

Interaction with Other Channel Regulators

In many cases, PIP₂ apparently binds directly to the ion channel to activate it, and channel inhibition upon receptor activation is mediated by PIP₂ depletion. But there are examples where the situation is more complex:

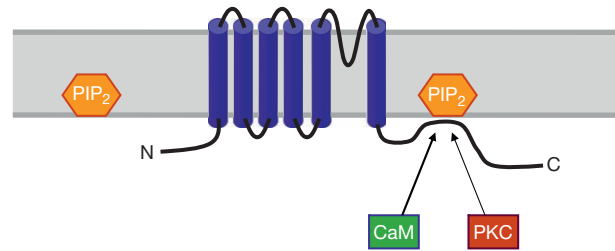


Figure 2 Ion-channel-binding sites for PIP₂ and intracellular regulators overlap. For several classes of channels, including KCNQ and TRP, binding sites for intracellular signals such as calmodulin and protein kinase C overlap with putative PIP₂-binding sites in the proximal C terminus. For KCNQ channels, this domain is located between C-terminal helices called A and B (which also coincides the binding site for scaffolding protein AKAP79/150), and for TRP channels it is in the vicinity of the consensus C-terminal TRP box.

1. PIP₂ does not need to bind to the channel itself. For NCX-1, PIP₂ appears to have indirect effects in addition to a known direct interaction with the exchanger. These include regulation via cytoskeletal remodeling and membrane trafficking.
2. PIP₂ changes the affinity of the channel for other regulatory elements. In the cardiac NCX exchanger, PIP₂ increases the affinity of a regulatory binding site for Ca²⁺, thus activating the exchanger. PIP₂ decreases the apparent affinity of K_{ATP} channels for the negative regulator ATP, thereby increasing current. For TRP and KCNQ1 channels, PIP₂ also reduces calmodulin binding, thereby relieving inhibition of current (Figure 2).
3. Other regulatory elements change PIP₂ affinity. Phosphorylation by PKA increases the affinity of ROMK channels for PIP₂, increasing channel activity. Phosphorylation of GIRK channels by PKC lowers their affinity for PIP₂, whereas G-protein $\beta\gamma$ subunits and intracellular Na⁺ increase GIRK channel affinity for PIP₂.

Based on allosteric considerations, examples (2) and (3) should be synonymous. They are described separately to reflect the different experimental approaches used. The interaction between PIP₂ and other channel modulators is not surprising given that channel structures would change upon binding of phosphoinositides and that binding sites for phosphoinositides and other regulators can overlap. In many channels, including KCNQ and TRPV1, this ‘hot spot’ for binding of channel modulators is located in the proximal C-terminus (Figure 2).

Phosphoinositide Specificity in Ion-Channel Regulation

Whether an ion channel is regulated by a particular phosphoinositide depends on several factors. (1) Phosphoinositide availability: PI(4,5)P₂ and PI(4)P each make up 1% of the total anionic phospholipids of the cell. PI(3,4)P₂, PI(3,5)P₂, and PI(3,4,5)P₃ are much less abundant. PI(4,5)P₂ is unique in its reliable presence at the plasma membrane, but many enzymes transiently change the density of PI(4,5)P₂, making it a dynamic signaling molecule. These features make PI(4,5)P₂ the most physiologically relevant phosphoinositide regulator of ion channels at the plasma membrane. (2) Charge: In channels as in other phosphoinositide-binding proteins, several

(three or more) basic amino acid residues close to the plasma membrane interface are thought to interact electrostatically with the acidic phosphate groups of phosphoinositide head groups. A doubly phosphorylated phosphoinositide will bind more strongly than a monophosphorylated one. (The phosphoinositide-interacting residues need not be adjacent in the sequence of the protein.) (3) Structure of the ion-channel phosphoinositide-binding site: The specificity of phosphoinositide-binding domains in various proteins ranges from highly selective (PH, PX, and FYVE domains) to promiscuous (MARCKS proteins). Proteins with a structured binding pocket even can discriminate between isomers, for example, PI(4,5)P₂ versus PI(3,4)P₂; others are less selective. ENaC channels are stimulated by PI(4,5)P₂, PI(3,4)P₂, and PI(3,4,5)P₃, but appear to bind PIP₂ and PIP₃ at distinct sites. IP₃ receptors are also sensitive to PIP₂, and the activity of TRPV1 can be supported by PIP, PIP₂, and PIP₃. For proteins with a less-structured interaction, some functional specificity may be achieved by co-localization with kinases and/or phosphatases that generate a microdomain with a specific phosphoinositide population (specificity by abundance, see above).

Some ion channels present a bimodal sensitivity to phosphoinositides. In a heterologous system, current through some Ca_v channels is stimulated by low PIP₂ and depressed by high PIP₂ levels. The effects of low PIP₂ do not change voltage dependence of gating, whereas those of high PIP₂ shift the voltage dependence of opening to more positive potentials. The latter shift is similar to the willing/reluctant paradigm of modulation of Ca_v by G-protein βγ subunits. It is postulated that these channels contain two binding sites for PIP₂, one of lower and one of higher affinity, a mechanism also suggested for TRPV1.

Structural View of Ion-Channel Modulation

Voltage-gated potassium and TRP channels are formed by four homologous subunits. Each subunit has an N-terminal voltage-sensor domain, a pore domain, and a C-terminal tail. Most evidence suggests that PIP₂ interacts with residues in the proximal C-terminus (Figure 2). The phosphoinositide specificity of ion channels provides a clue to the structure of their phosphoinositide-sensitive domains, with promiscuous channels employing a cloud of charges and more discriminating channels using a better defined binding pocket analogous to PH domains. However, some members of the TRP family of ion

channels (e.g., TRPM4) may employ both motifs. In no case is the structure of a putative-binding pocket actually known.

Conclusions

An accumulating body of work demonstrates that many plasma membrane proteins – ion channels and transporters – require PIP₂ to function. Changing PIP₂ levels over time allows for temporal modulation of these channels, for instance by G_q-coupled receptors. Different phosphoinositide compositions of cellular membrane compartments potentially allow for spatial regulation of channel function, determining where ion channels work, and where they do not. A similar phosphoinositide regulation may be found in the future for other classes of membrane proteins and for intracellular membranes in addition to the plasma membrane.

See also: **Bioenergetics:** Membrane Transport, General Concepts; Transient Receptor Potential Channels; Voltage-Dependent K⁺ Channels; Voltage-Gated Ca²⁺ Channels; **Lipids Carbohydrates Membranes and Membrane Proteins:** Glycerolipid Structure, Function, and Synthesis in Eukaryotes; Ion Channel Protein Superfamily; **Signaling:** Phosphatidylinositol Bisphosphate and Trisphosphate.

Further Reading

- Balla T, Szentpetery Z, and Kim YJ (2009) Phosphoinositide signaling: New tools and insights. *Physiology* 24: 231–244.
- Delmas P and Brown DA (2005) Pathways modulating neural KCNQ/M (Kv7) potassium channels. *Nature Reviews in Neuroscience* 6: 850–862.
- Falkenburger BH, Jensen JB, Dickson EJ, Suh B-C, and Hille B (2010) Phosphoinositides: Lipid regulators of membrane proteins. *Journal of Physiology* 558: 3179–3185.
- Gamper N and Shapiro MS (2007) Regulation of ion transport proteins by membrane phosphoinositides. *Nature Reviews Neuroscience* 8: 921–934.
- Hilgemann DW, Feng S, and Nasuhoglu C (2001) The complex and intriguing lives of PIP₂ with ion channels and transporters. *Science STKE* 111: re19.
- Rohacs T and Nilius B (2007) Regulation of transient receptor potential (TRP) channels by phosphoinositides. *Pflügers Archiv – European Journal of Physiology* 455: 157–168.
- Suh BC and Hille B (2005) Regulation of ion channels by phosphatidylinositol 4,5-bisphosphate. *Current Opinion in Neurobiology* 15: 370–378.
- Suh BC and Hille B (2009) PIP₂ is a necessary cofactor for ion channel function: How and why? *Annual Reviews of Biophysics* 37: 175–195.

Live Cell Imaging of Nuclear Dynamics

J C Harr and K L Reddy, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

The search for the mechanisms underlying nonrandom chromosome and gene organization within the nuclear volume and the precise functional consequences of such a configuration lies at the heart of research into nuclear dynamics. It has been demonstrated that chromatin is likely organized into discrete domains by association with distinct nuclear compartments enriched in structural and regulatory proteins. Growing evidence suggests that gene activity is modulated by interactions with these subnuclear compartments and that such interactions can be cell type specific. Analyzing how nuclear architecture controls genome activity and the mechanisms that drive such organization is necessary to fully understand complex biological processes such as development and disease. The most widely used approaches in understanding and illuminating nonrandom organization of the genome have relied on high-resolution two- or three-dimensional fluorescent *in situ* hybridization (FISH) techniques at the single cell level and on bulk cell population analyses using molecular methods such as chromosome conformation capture (3C and its derivatives such as Hi-C) and DNA adenine methylation identification (Dam-ID) which measure probabilities of chromosomal regions interacting with each other or with a nuclear domain, respectively. While these powerful molecular and high-resolution microscopy techniques have given us a wealth of understanding about nuclear architecture, they reflect an average of a bulk population and are necessarily static. Complementary live cell imaging techniques using the living cell as both a visual and molecular test tube are vital to our understanding of the dynamics and mechanisms underlying nuclear organization. The most insight will be gained by studies using genome wide studies of nuclear architecture, such as Dam-ID and Hi-C, that illuminate frequencies or averages of nuclear structure in a cell population, coupled with live cell studies that capture real-time spatial dynamics of nuclear positioning and molecular consequences of nuclear compartmentalization.

The Dynamically Organized Nucleus

The mammalian nucleus has a complex structural organization that dynamically interacts with the genome. Chromatin is organized into discrete domains by association with distinct nuclear compartments that are defined by their enrichment in structural and regulatory proteins. Growing evidence suggests that gene activity is modulated by interactions with these subnuclear compartments. Such compartmentalization has been proposed to bring together genes whose transcription or repression is coordinately regulated. Recent evidence suggests that the nonrandom organization of chromosomes in the nucleus is a contributing factor in facilitating specific translocations and that disruption of normal nuclear architecture leads to a variety of diseases including premature aging, Burkitt's lymphoma, some types of muscular dystrophy, and other so-called laminopathies.

However, while mounting evidence suggests that gene activity is regulated by nuclear compartmentalization, relatively little is known about how specific loci are directed to or associated with different nuclear domains. Additionally, the molecular consequences of such organization are unknown. In fact, most of the studies to date have relied on cytological analysis of fixed specimens and are, thus, correlative. By contrast, live cell methodologies that direct compartmentalization allow one to discriminate between positioning and molecular consequence. The consequences of such directed re-positioning or compartment formation on gene activity or other DNA transactions (e.g., replication timing) can then be analyzed by molecular or fixed cell approaches. In addition, other live cell technologies that detect protein–protein interactions, such as bi-molecular fluorescence complementation (BiFC), and those that detect nascent RNA production (e.g., through MS2-hairpin recognition) in combination with the directed compartmentalization techniques allow the living cell to give both 3D/4D position and quantifiable molecular data.

3D-FISH and Genome Wide Molecular Approaches: Insights into Nuclear Architecture

Three different technologies have been widely used to analyze and study nuclear architecture and dynamics: 3D-FISH, chromosome conformation capture or 3C (more recently its derivatives, ChIA-PET and Hi-C), and Dam-ID. In all of these techniques, the assays are designed to reflect and probe for nuclear architecture and therefore are either performed on fixed (FISH and Hi-C) or initiated in live cells (Dam-ID). As such, the Dam-ID and Hi-C techniques represent a great step forward from canonical molecular assays which disregard three-dimensional context. 3D-FISH analyses give greater insight into position and colocalization with protein compartments than its ancestors 2D-FISH and simple immunofluorescence detection of proteins.

Fluorescence *in situ* hybridization on three-dimensionally preserved nuclei (3D-FISH) has become a major tool to study higher-order chromatin architecture in the interphase nucleus. Even with the development of live cell imaging assays, this technique will remain a powerful tool since it can be used on virtually any cell or tissue type. The 3D-FISH technique involves hybridization of labeled DNA probes to three-dimensionally preserved nuclei to detect single genes, chromosomal domains, or entire chromosome territories. Unfortunately, 3D-FISH requires harsh pretreatments, including a denaturation step, for the best accessibility and hybridization of probes to nuclear target DNA. Light microscopy observations of individual ~1-Mb chromatin domains in the same cells *in vivo*, after fixation, and after 3D-FISH demonstrate that number, size, and spatial distribution of ~1-Mb chromatin domains are generally well preserved throughout the 3D-FISH procedure. In addition,

other studies indicate that both the position of DNA segments and the localization of specific nuclear structures (e.g., promyelocytic leukemia (PML) bodies, nucleoli, and centromere regions of interphase chromosomes) do not change significantly at the light microscopy level even after 3D-FISH treatments. Importantly, regardless of minor perturbations in nuclear structure due to technical issues, 3D-FISH analyses have demonstrated that repositioning of genes and chromosomal domains does occur and that this rearrangement coincides with molecular events, such as transcription of the gene being studied. Such studies have been done on regions to demonstrate proximity to either other chromosomal domains (such as chromosome territories or coordinately regulated loci) or nuclear compartments (such as the nuclear lamina, the nucleolus, transcription factories, PML bodies, splicing factories, etc.). For example, it has been shown that germ-line immunoglobulin heavy chain loci are preferentially localized to the nuclear lamina in hematopoietic progenitors, T lineage cells, and non-B cells (such as fibroblasts) but centrally positioned in the nucleus in pro-B cells, where they are active. Similar repositioning has been shown for the IgK locus, the TCR loci, Ikaros and the cystic fibrosis gene, CFTR, among others. This cytological demonstration of repositioning, while correlative, indicates the dynamics of nuclear architecture in a cell-type-specific manner and hints at a role for nuclear structure in regulation of gene activity. Lacking in these types of assays are real time associations and demonstration of any molecular interactions with, or functional consequences of, association with a given nuclear compartment, or with other loci.

In contrast to 3D-FISH, Dam-ID uses the living cell to mark regions that are in molecular contact with protein components. In this protocol, DNA adenine methyltransferase (Dam) from *Escherichia coli* is fused to a DNA-interacting protein. Dam adds methyl groups to adenines in the vicinity of the fusion protein, thereby marking the interaction site on the underlying DNA segment. This fusion protein is expressed at extremely low levels to ensure no 'off-target' effects. Since higher eukaryotes do not normally have methylated adenines, any marked adenine can be presumed to be a consequence of the fusion protein interacting at the marked site. Moreover, mammalian nuclei appear to be blind to this mark as there is no mechanism for 'reading' adenine methylation. This technique is uniquely suitable for those proteins which are difficult to subject to chromatin immunoprecipitation (ChIP), notably proteins of the inner nuclear membrane INM/lamina or other nuclear compartments. Once the protein is introduced into cells via viral transduction, DNA is isolated and subjected to restriction digestion followed by ligation-mediated polymerase chain reaction (PCR) to amplify regions demarcated by the fusion protein. This DNA can then be analyzed by qPCR to specific genic regions, subjected to hybridization to tiled arrays or, theoretically, high-throughput sequencing. Dam-ID has emerged as a powerful tool to map the interaction of chromosomal segments with INM/lamina proteins, the so-called lamin associated domains or LADs. It should be noted that 3D-FISH analyses hinted at the importance and widespread interaction of some chromosomal domains with the INM/lamina, but the FISH technique is ill-suited to give genome-wide high-resolution maps of association and incapable of indicating

true molecular interactions between chromatin and protein components of a subnuclear compartment. In human fibroblasts, the Dam-ID technique has revealed large (0.1–10 Mb) delimited chromosomal regions which contain mostly inactive chromatin. Many of these domains contain developmentally regulated genes that change their association with this compartment in a developmentally regulated manner. However, it remains unclear how this architecture might functionally contribute to the repression of lineage specific genes and how nuclear organization might vary in other lineages. Thus, Dam-ID is capable of marking chromatin that is associated with a specific nuclear compartment in living cells to give high-resolution maps of chromatin domains. The limitation of this technique is, of course, that the demarcated domains represent an analysis of a bulk population of cells and the dynamics of association are completely obscured since the methyl mark deposited by the methyltransferase is permanent. In addition, since the Dam-ID technique reflects frequencies of association, 3D-FISH analyses are routinely used to verify that a given region is indeed associated with the domain being assayed in a majority of cells.

Like 3D-FISH, the chromosome conformation capture (3C) technique and its derivatives depend upon preserving nuclear architecture by fixation with formaldehyde. In 3C this fixation step is followed by restriction enzyme digestion to leave cross-linked sequences attached. The DNA is then ligated under dilute conditions to favor self-ligation of cross-linked DNA fragments. The ligated fragments, that should contain both pieces of interacting DNA, are then analyzed by PCR with specific primers to the two loci of interest. In this case, one must have an idea of the two interacting regions. While useful for specific loci of interest, 3C has very limited throughput. There have been many variations on this technique, but two have proved to be especially useful in looking at nuclear structure: ChIA-PET (chromatin interaction analysis by paired-end tag sequencing) and Hi-C. The ChIA-PET method combines ChIP-based methods and 3C. Specifically, after fixation, chromatin is sonicated and an immunoprecipitation is performed to a specific protein to enrich for chromatin associated with this protein of interest. After precipitation, two differentially bar-coded linkers are ligated to the sheared ends of the chromatin and then ligated under dilute conditions (as in 3C). After digestion with a restriction endonuclease which cuts outside of the linkers, the resulting complexes are subjected to reverse cross-linking, purification, and high-throughput sequencing. The results give a genome-wide interaction map of *cis*- and *trans*-DNA sequences that are associated with one another through their association with the protein of interest. Hi-C which was designed to detect all pairwise physical associations of DNA in the genome is perhaps the ultimate extension of the chromosome conformation capture (3C) assay. As in 3C and its variants, the cross-linked genome is enzyme-digested and the resulting pieces of genomic DNA are ligated to each other based on proximity. Thus, linear linkages along the chromosomes are virtually erased, and new connections are generated based on three-dimensional proximity within the nucleus. In this case, isolation of pairwise interactions is achieved by incorporating biotin into the ends of the digested DNA before ligation, and then carrying out physical selection of these fragments on streptavidin beads followed by

paired-end sequencing. Like ChIA-PET, this technique gives both *cis*- and *trans*-DNA sequence interactions. However, unlike ChIA-PET, the technique is not constrained (or enhanced, depending upon your point of view) by choosing a particular protein of interest. Both ChIA-PET and Hi-C generate massive amounts of genome-wide data and require very stringent controls to ensure that meaningful interactions are uncovered. Intriguingly, even though these techniques are relatively new, both Hi-C and ChIA-PET techniques have pointed to CTCF-binding sites as regions important for genome organization. Hi-C has demonstrated that active regions and inactive regions of the genome segregate from one another. Again, the pitfalls of these expensive technologies are that they represent a bulk analysis of the chromosomal interactome and are also unable to uncover dynamic molecular processes involved in the organization they uncover.

Insights into Chromosome Dynamics Using Live Cells: The *LacO* System

The LacI/*lacO* system has been widely used in mammalian cells to study chromosomal dynamics and chromatin decondensation. These arrays can and have been used to visualize a chromosomal domain and its movements in living cells under many conditions (Figure 1). Previous studies have demonstrated that LacI binds to *lacO* sites with high affinity ($K_d = 10^{-10}$) and DNA binding is reversed upon incubation with the allosteric inhibitor isopropyl β -D-1-thiogalactopyranoside (IPTG). Importantly, in mammalian cells a fluorescently tagged LacI protein is able to bind to integrated *LacO* sites (256 copies) specifically and can be visualized using simple live cell microscopy techniques. Thus, this system provides for highly specific and stable DNA binding that is reversible using a relatively inexpensive reagent (IPTG). One could also use iterated *tetO*

sequences and tetR fusion proteins in a similar manner with doxycycline to control binding. These artificially designed chromosomal domains have been used to demonstrate *in vivo* dynamics of large amplified arrays of *lacO* binding sites occupying megabases of DNA as well as more discrete regions comprising only one or two sites of integration (10 kb). These integrated arrays have been demonstrated to exhibit simple Brownian motion in interphase nuclei and the larger (Mb) sized integrations, are quite often condensed and heterochromatic. Upon binding of a transcriptional activator to these large designer chromosomal domains, the decondensation of the domain is dramatic and easily seen by light microscopy. In addition, a random integration positioned reproducibly at the nuclear periphery shows dramatic relocalization to the nuclear interior upon trans-activation along with rapid and apparently directed movements. Finally, the extent of movement of individual tagged chromosomal domains varies depending upon their nuclear environment; specifically, chromosomal domains located at the nuclear periphery or chromocenters move in a smaller diameter than those regions located in the nuclear interior. Taken together, such studies demonstrate the utility of using tagged chromosomal domains in living cells to study the interplay of chromosome dynamics and gene regulation in space and time.

Forced Repositioning of Chromosomal Domains Using Molecular Tethers in Living Cells: Insights into Consequences of Repositioning

More recently, the *lacO* system has been used to position integrated test genes to the inner nuclear membrane and to test the consequences of such repositioning on gene activity (Figure 2). Three studies took advantage of targeting strategies

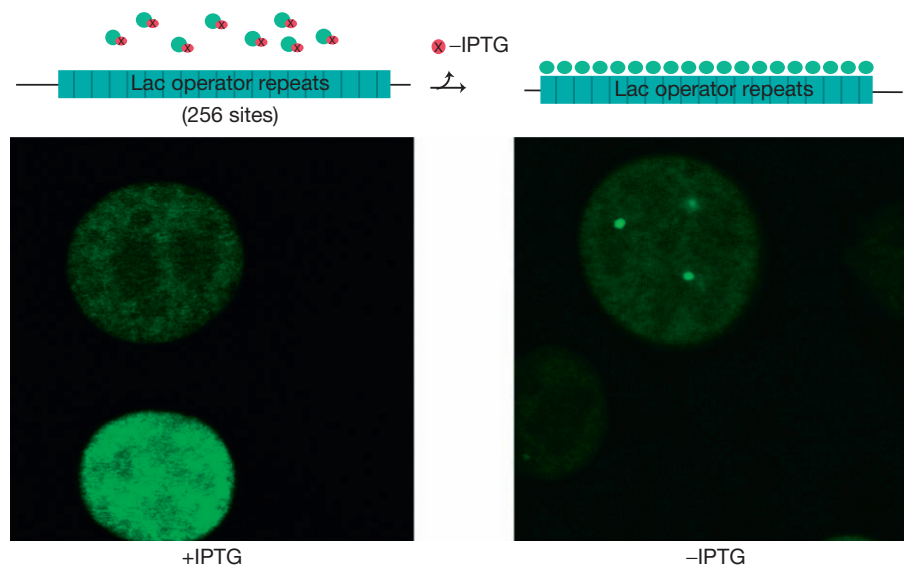


Figure 1 Stably integrated arrays of *lacO* binding sites are visible upon binding of GFP-LacI in living cells. Shown are cells containing the *lacO* repeats that are also expressing GFP-LacI. This interaction is controlled by the allosteric inhibitor IPTG. IPTG can be used in the culture medium to prevent GFP-LacI from binding (+IPTG). Upon IPTG withdrawal (–IPTG), accumulation of the GFP-LacI protein on the *lacO* arrays occurs in a matter of hours. This system has been widely used to look at chromosomal domain dynamics (several megabases) with large amplified regions of *lacO* repeats or dynamics of smaller regions (~10 kb) using single integrations.

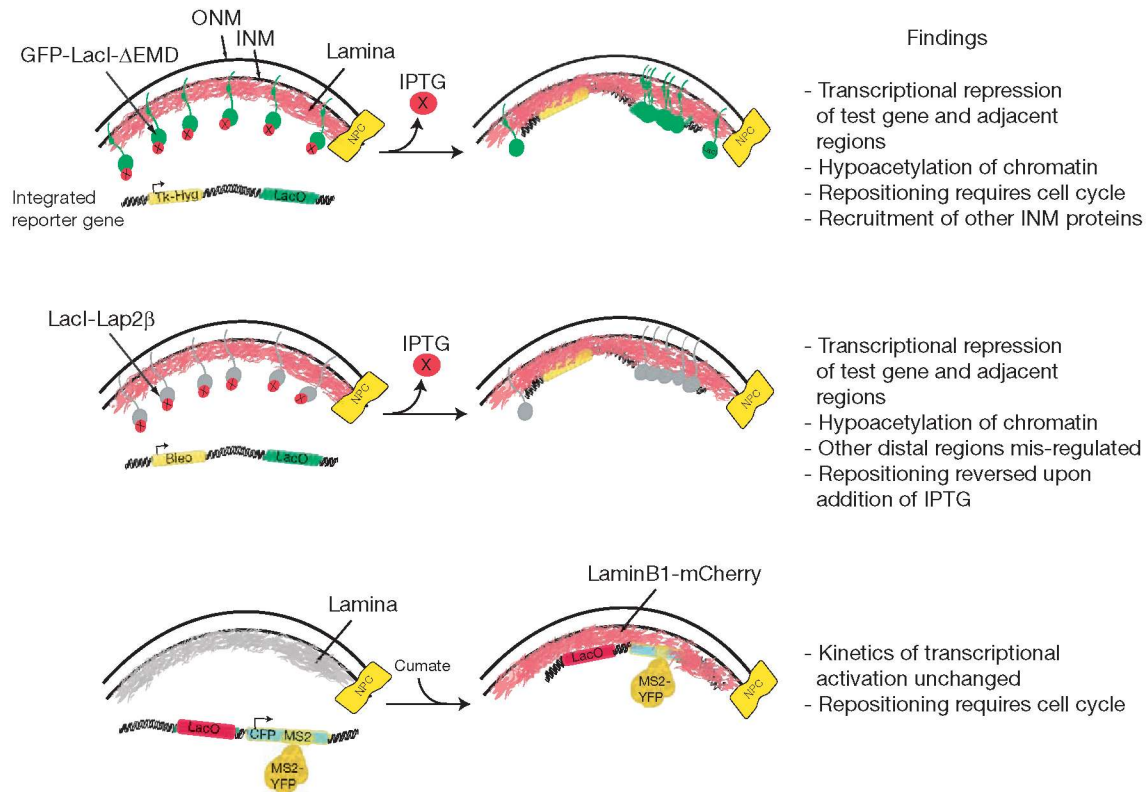


Figure 2 Three different strategies expanded upon the original *lacO*/LacI system to tether chromosomal domains to the INM/Lamina. The top two panels demonstrate similar strategies using an INM protein. In the topmost panel, a truncated version of Emerin, Δ EMD, fused to LacI was used to reversibly interact with chromosomal domains. In the middle panel, the INM protein Lap2 β was used as the molecular tether to recruit *lacO* sites. In both cases binding was able to be controlled by introducing the allosteric inhibitor IPTG. In the bottom panel, a cumate inducible LaminB1-LacI fusion was used to recruit *lacO* insertions to the nuclear lamina. The findings from each of the studies are outlined to the right.

using *lacO* sites positioned either well upstream or downstream of a test gene to avoid inhibition of transcription due to binding of LacI fusion proteins. These target constructs (containing *lacO* repeats) were then tethered to the INM/lamina using either a portion of the resident INM proteins (Emerin or Lap2 β) or the nuclear lamina protein LaminB1. These studies focused on recruitment to the nuclear periphery because it is one of the few stable structures in the nucleus (i.e., it has defined location, molecular make-up, and boundaries) and it has been directly implicated in regulating gene activity, specifically gene repression. Other compartments within the nucleus, such as splicing factories, cajal bodies, can occupy any position within the nuclear volume making it more difficult to tease apart recruitment or *de novo* compartment formation that may occur by simply recruiting a high local concentration of regulatory proteins. In both studies using INM protein domains, the chimeric proteins were stably expressed in cells and they resulted in the recruitment of the target *lacO* containing chromosomal domains to the INM upon removal of IPTG. In the case of the LaminB1-LacI chimeric fusion, the protein itself was inducibly expressed and was not regulated by IPTG. In all cases, the *lacO* insertions were reproducibly recruited to the nuclear periphery, allowing, for the first time, a study of the consequences of forced compartmentalization. The two studies utilizing INM proteins (Emerin and Lap2 β) as tethers resulted in transcriptional repression accompanied by changes in chromatin status

around the recruited domain. Intriguingly, the study using LaminB1 focused on the kinetics of gene induction of a reporter gene tethered to the nuclear lamina. The activity of this reporter gene, encoding the cyan fluorescent protein (CFP) and containing MS2-RNA repeats in the intron of the nascent RNA which can be recognized by MS2-YFP protein, allows for visualization of gene activity in individual live nuclei. For this reporter system activity levels were measured by either CFP fluorescence or MS2-YFP accumulation at sites of nascent transcription. Intriguingly, these forced tethering studies also uncovered a potential link between progression through the cell cycle and recruitment to the nuclear lamina. These observations would never have been made using molecular or fixed cell approaches. These adaptations of the original LacI/*lacO* system demonstrate the versatility of using live cell systems to discover not only dynamics of nuclear architecture but also consequences of nonrandom association with a nuclear compartment.

The Future: Induced Repositioning of Chromosomal Domains Using 'Nuclear Zipcodes' or Specific *trans*-Factors Identified by Genome-Wide/Molecular Analyses

How is the architecture of a nucleus set up? This is one of the big questions in the field of nuclear dynamics. The studies

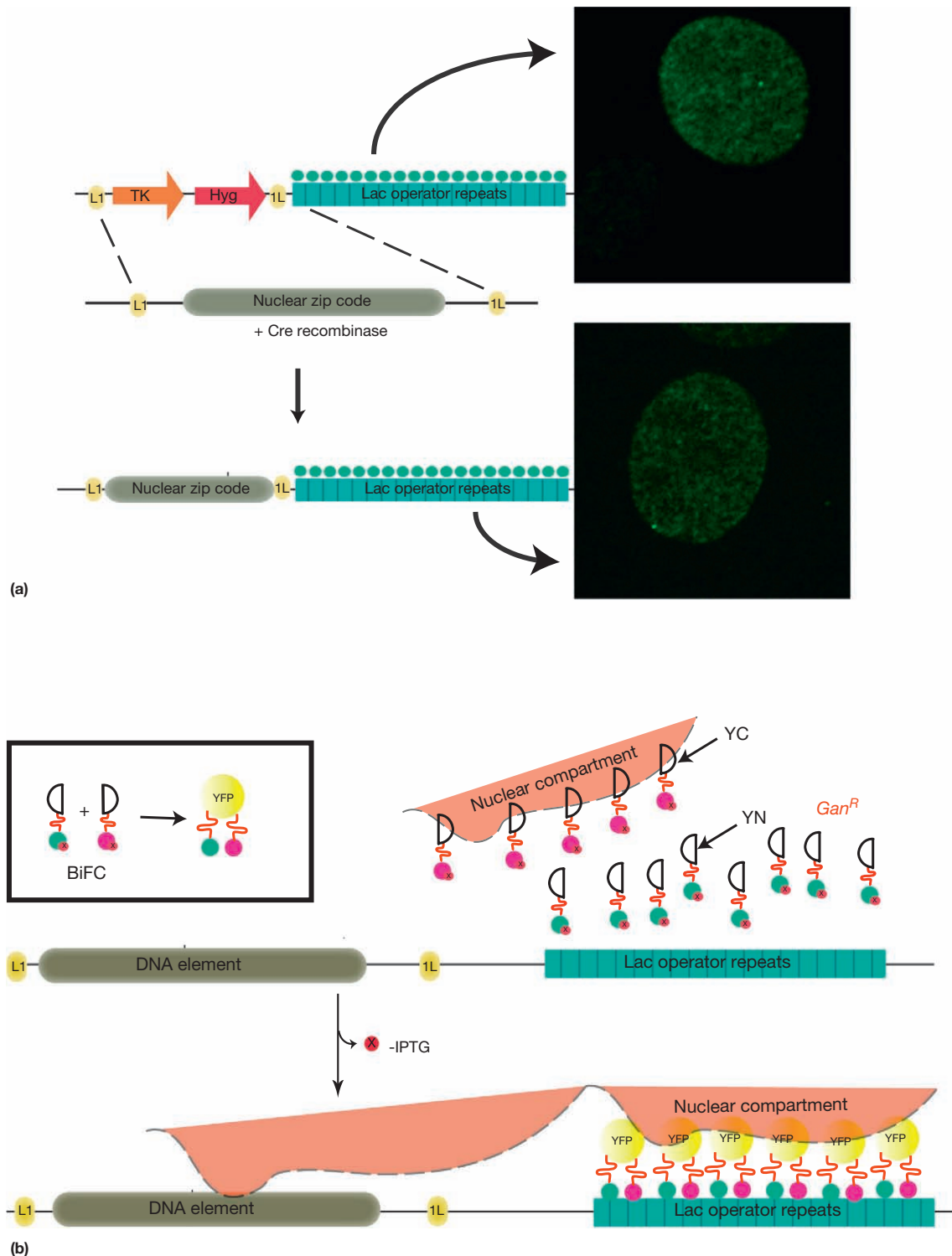


Figure 3 Newer uses of the lacO/LacI system. (a) A new system has been designed to allow for integration of 'nuclear zip codes' into a specific genomic site. This is accomplished by flipping the sequence of interest into an engineered chromosome region containing inverted loxP sites adjacent to *lacO* sequences. In the absence of a sequence that is able to direct compartmentalization, the array is located in the nuclear interior (nucleus at top). After introducing a targeting sequence containing a putative nuclear zip code, the array is now preferentially located at the nuclear periphery (nucleus at the bottom of panel (a)). Analyses using BiFC (see below) or expression from a test gene can be incorporated into this system to garner positional, molecular, and regulatory information. (b) Design of a system to detect molecular interaction of a genomic region with a nuclear compartment. The basic concept of BiFC is illustrated in the box in panel (b). LacI is fused to one-half of fluorescent YFP and a protein located in a nuclear subcompartment is fused to the cognate half of the YFP molecule. If the two proteins are in molecular proximity, a complete fluorescent YFP will be formed. A nuclear zip code is used (as in (a)) to direct a sequence to a subnuclear domain. The LacI fusion protein bound to the *lacO* repeats will be directed to the compartment and will be brought into proximity to allow for formation of the fluorescent protein. This type of analysis allows one to get specific real time information about molecular interactions.

detailed above describe artificial systems where genic regions are ‘forced’ to associate with a nuclear domain by molecular tethers, specifically the INM/nuclear lamina and, thus, did not address this basic question. As mentioned above, the nuclear periphery is one of the few stable structures in the nucleus. In addition, Dam-ID studies have indicated sequences that are in high-frequency association with this domain. Thus, the preferential association of a genomic region with this structure affords the unique opportunity to uncover the role that specific *cis*-elements have in determining genome organization. By definition, such *cis*-elements or nuclear ‘zipcodes’, would direct or facilitate association of chromosomal domains with a specific nuclear compartment. Of course, this concept of nuclear addressing also requires the action of specific *trans*-acting factors that interact with the *cis*-elements to mediate their association within a specific nuclear compartment. For example, the nuclear periphery of mammalian cells is made up by the nuclear lamina (such as Lamin A and B) and the INM, which contains many specific proteins implicated in gene regulation, such as emerin, and Lap2 β .

Using data generated by Dam-ID or ChIA-PET/Hi-C, potential *cis*-elements can be tested for their ability to direct nuclear compartmentalization. The advantage of using such a system is that the *cis*-elements tested would be subject to the normal cell cycle and other cellular controls involved in establishing nuclear architecture – a feature missing from the forced association studies using molecular tethers described above. One can imagine two ways to test the ability of *cis*-elements to reposition an ectopic site to a new position in the nuclear volume in living cells. First, one could simply randomly integrate these elements with *lacO* sites and assay the frequency of association with a given compartment (e.g., the nuclear lamina) by looking at GFP-LacI foci by light microscopy. Unfortunately, such an approach yields multiple integrations and different sites for each different sequence tested, making any conclusions about repositioning suspectable. Moreover, since there is no control over where the insertions take place, analyzing that the consequences of repositioning would not be possible. A better strategy would employ a system that allowed for directed insertion of the *cis* sequences to be tested into the same chromosomal location. An example of a modification of such a system that has been demonstrated to work in this way is shown in [Figure 3](#). This tagged chromosomal insertion site (TCIS) system incorporates the ability to tag a chromosomal region via the *lacO*/LacI system and to recombine specific sequences into the same location using recombination-mediated cassette exchange (RMCE), a well-documented method for site-specific integration of DNA sequences. The engineered chromosomal insertion is comprised of 256 *lacO* sites with an adjacent pair of inverted loxP sites which support recombination of specific sequences into this site. This engineered locus can be followed by introduction of GFP-LacI, which binds to the iterated *lacO* sites and accumulates to generate a discrete focus marking the insertion site. Specific sequences, in a switch cassette, can then be introduced via Cre recombinase to this locus and the disposition of the chromosomal domain within the nucleus can be followed in living cells. Thus, the TCIS system represents a rapid way to assay *cis*-elements for nuclear disposition and activity. Moreover, since the multiple sequences are tested

within the same nuclear environment, artifactual results due to position effects are avoided.

In addition to using live cells to study the mechanism by which *cis*- or *trans*-elements are able or sufficient to cause a chromosomal domain to reposition to a new nuclear domain, live cells can also be used to detect the molecular association of engineered chromosomes with a protein present in a nuclear compartment. BiFC allows the direct visualization of protein complexes in their native environment. BiFC is based on the formation of a fluorescent protein from two nonfluorescent constituents. Specifically, BiFC employs a strategy that splits either GFP or YFP into two domains (an N- and a C- terminal portion) that can complement one another to form a fluorescent protein, but individually have no fluorescence. The two domains only fluoresce when they are brought into close molecular proximity to one another, for example, when they are fused to two proteins that interact with one another ([Figure 3](#)). Coupling this technology with the *lacO*/LacI system should enable detection of such nuclear events as chromosome looping and compartmentalization by fusing one-half of the BiFC to LacI and the other half to either a chromatin interacting protein or a protein resident in a specific nuclear compartment, for example, the nuclear lamina ([Figure 3](#)). BiFC has been used quite successfully to demonstrate novel dynamics of the protein CBX binding to chromatin. CBX recruits the polycomb repressive complex 1 (PRC1) and is very important in establishing heterochromatin domains. Numerous previous studies relied on immunofluorescence to detect CBX proteins; however, such a strategy is unable to detect whether the protein is actually interacting with chromatin. Of course, ChIP can detect specific chromatin interactions, but those studies lack spatial information. Using BiFC to CBX and histones, the patterns of chromatin interacting versus total CBX were quite distinct. Thus, BiFC allows measurement of spatial, contextual, and temporal changes in protein complexes and will likely be an important tool in detecting the dynamics of nuclear compartmentalization.

In summary, live cell imaging techniques are poised to take advantage of the many genome-wide molecular assays that are currently being employed to describe genome-wide nuclear associations to uncover the mechanisms and consequences of nuclear architecture and organization on gene activity. Using the well-established *lacO*/LacI system and other techniques such as BiFC and nascent RNA detection methods in living cells will enable spatial, temporal, and molecular analysis to be carried out in a single cell giving an unprecedented view of the form and function of nuclear dynamics.

See also: [Cell Architecture and Function: Nuclear Lamina;](#)
[Molecular Biology: lac Operon.](#)

Further Reading

- Belmont AS, Li G, Sudlow G, and Robinett C (1999a) Visualization of large-scale chromatin structure and dynamics using the lac operator/lac repressor reporter system. *Methods in Cell Biology* 58: 203–222.
- Belmont AS, Stephen D, Nye AC, Strukov YG, and Tudorita T (1999b) Large-scale chromatin structure and function. *Current Opinion in Cell Biology* 11(3): 307–311.

- Botta M, Haider S, Leung I, Lio P, and Mozziconacci J (2010) Intra- and inter-chromosomal interactions correlate with CTCF binding genome wide. *Molecular Systems Biology* 6: 426.
- Finlan LE, Sproul D, Thomson I, et al. (2008) Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genetics* 4(3): e1000039.
- Handoko L, Xu H, and Li G (2011) CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature Genetics* 43(7): 630–638.
- Kerppola TK (2008) Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annual Review of Biophysics* 37: 465–487.
- Kumaran RI and Spector DL (2008) A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *Journal of Cell Biology* 180(1): 51–65.
- Li G, Fullwood MJ, Xu H, et al. (2010) ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biology* 11(2): R22.
- Misteli T and Spector DL (eds.) (2010) *The Nucleus (Perspectives in Biology)*. Plainview, NY: Cold Spring Harbor Laboratory Press.
- Reddy KL and Singh H (2008) Using molecular tethering to analyze the role of nuclear compartmentalization in the regulation of mammalian gene activity. *Methods (San Diego, Calif.)* 45(3): 242–251.
- Reddy KL, Zullo JM, Bertolino E, and Singh H (2008) Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* 452(7184): 243–247.
- Solovei I, Cavallo A, Schermelleh L, et al. (2002) Spatial preservation of nuclear chromatin architecture during three-dimensional fluorescence *in situ* hybridization (3D-FISH). *Experimental Cell Research* 276(1): 10–23.
- Tumbar T, Sudlow G, and Belmont AS (1999) Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *Journal of Cell Biology* 145(7): 1341–1354.
- Vincenz C and Kerppola TK (2008) Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. *Proceedings of the National Academy of Sciences of the United States of America* 105(43): 16572–16577.
- Vogel MJ, Peric-Hupkes D, and van Steensel B (2007) Detection of *in vivo* protein-DNA interactions using DamID in mammalian cells. *Nature Protocols* 2(6): 1467–1478.

Low-Barrier Hydrogen Bonds

P A Frey, University of Wisconsin, Madison, WI, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 594–598, © 2004, Elsevier Inc.

Glossary

Deuterium ractionation factor The equilibrium constant for exchange of a proton with the deuterons in D_2O .

Isotope shift The upfield perturbation of a nuclear magnetic resonance signal of an atom when a heavy isotope is substituted for it.

Transition state analog A stable molecule that has nearly all of the structural features of a transition state.

Proteins contain mainly weak, conventional hydrogen bonds; however, a few enzymes have low-barrier hydrogen bonds (LBHBs) in transition state analog complexes. Hydrogen bonds display variations in physicochemical properties including length, spectroscopic characteristics, and strength. Three classes of hydrogen bonds have been defined – weak ($2\text{--}8\text{ kcal mol}^{-1}$), strong ($10\text{--}20\text{ kcal mol}^{-1}$), and very strong ($24\text{--}40\text{ kcal mol}^{-1}$). In a weak hydrogen bond, the proton is bonded to one heteroatom by a dipolar covalent bond and engages in a weak electrostatic attraction with another heteroatom. In a strong hydrogen bond, or LBHB, the heteroatoms are closer than a van der Waals contact, the covalent bond to the proton is elongated, and the contact between the proton and the second heteroatom is significantly shorter than in a weak hydrogen bond. In a very strong, or single-well hydrogen bond, the heteroatoms are much closer than a van der Waals contact, and the proton is nearly centered between them. LBHBs in proteins and small molecules are characterized by their spectroscopic and thermodynamic properties, deuterium fractionation factors, and crystallographic structures. LBHBs have been identified in transition state analog complexes of a few enzymes and are postulated to stabilize the transition states in catalysis.

The Nature of Hydrogen Bonds

Hydrogen bonds were controversial throughout the twentieth century. By mid-twentieth century, the weak conventional hydrogen bonds were reasonably well understood and widely accepted. Unlike covalent bonds, which vary in strength within a factor of ~ 4 ($30\text{--}120\text{ kcal mol}^{-1}$), hydrogen bonds are much less constrained in their geometry and physical properties, and they vary in strength by a factor of at least 20-fold ($2\text{--}40\text{ kcal mol}^{-1}$). Very few strong hydrogen bonds had been documented before the mid-twentieth century, principally the very strong hydrogen bond in hydrogen difluoride $[F\cdots H\cdots F]^-$ (40 kcal mol^{-1}). Such strong hydrogen bonds were thought to be limited to crystalline states. In the latter half of the century, strong hydrogen bonds were discovered in several classes of organic molecules in crystals and aprotic liquid phases and characterized by crystallographic, spectroscopic, and chemical methods.

Weak, Strong, and Very Strong Hydrogen Bonds

The distinctions among hydrogen bonds can be illustrated by the differences among their energy profiles, as shown in Figure 1. Figure 1(a) is the energy profile for a weak hydrogen bond, that in Figure 1(b) is for a strong or LBHB, and that in Figure 1(c) is for a single-well or very strong hydrogen bond. The three classes are sometimes referred to as weak, moderate, and strong hydrogen bonds. Most hydrogen bonds in proteins and nucleic acids involve oxygen or nitrogen as the heteroatoms. For that reason, the hydrogen bonding properties delineated below refer to N or O as heteroatoms. Larger or smaller heteroatoms such as sulfur or fluorine display analogous properties with physical parameters outside the ranges for N and O.

Weak Hydrogen Bonds

As illustrated in Figure 1(a), the hydrogen (or deuterium) is confined to covalent bonding with one heteroatom by the fact that their zero-point vibrational energies are much lower than the barrier separating the heteroatoms. The covalent bond linking the electronegative atom to a hydrogen is dipolar, with a partial positive charge on the proton and a partial negative charge on the heteroatom. The electrostatic attraction between the proton and another partially negative heteroatom is a weak hydrogen bond. The strengths of weak hydrogen bonds range between 2 and 8 kcal mol^{-1} , depending on the type of heteroatoms, their electrostatic charge, the distance separating them, and the dielectric medium. Because they are weak, conventional hydrogen bonds easily undergo exchange with protons of water, with enthalpies of activation on the order of $2\text{--}3\text{ kcal mol}^{-1}$.

LBHBs – Strong Hydrogen Bonds

In an LBHB (Figure 1(b)) the heteroatoms are closer together and separated by significantly less than the van der Waals contact distance. As a consequence, the energy barrier between them is lowered to the range of the zero point vibrational energy for hydrogen, so that the hydrogen feels increased attraction to the other heteroatom and is drawn away from the first heteroatom, $A\cdots H$ in Figure 1(b). Deuterium feels less

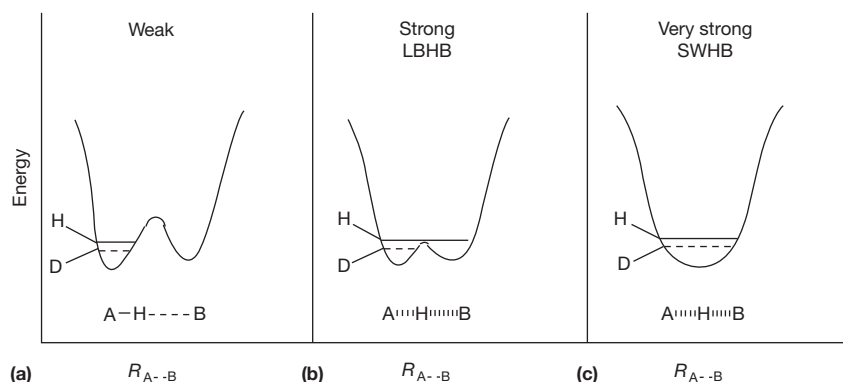


Figure 1 In (a), the double well potential for a proton covalently bonded to heteroatom A and weakly hydrogen-bonded to heteroatom B is illustrated. The zero point energies for H and D are both well below the barrier, and both are strictly covalently bonded to A. In (b) for an LBHB, the heteroatoms are drawn close together in violation of van der Waals contact distance, the proton is moved toward B and away from A, and the barrier is lowered to near the zero point energy of H. The zero point energy of D is lower, so that deuterium is less attracted to B. In (c) for a single-well hydrogen bond, the heteroatoms are even closer together, and the proton is nearly centered between them. In the LBHB and SWHB, the bonding of the proton to heteroatom B is partially covalent.

attraction to the second heteroatom, owing to its lower zero point energy. Because of this difference, an LBHB displays deuterium and tritium isotope effects. LBHBs are stronger ($10\text{--}20\text{ kcal mol}^{-1}$) than weak hydrogen bonds because of the partial covalency between hydrogen and the second heteroatom, $\text{H}\cdots\text{B}$ in **Figure 1(a)**. As they are strong, LBHBs undergo exchange with protons of water less readily than weak hydrogen bonds. In an LBHB, the proton is strongly held, and the activation enthalpy for exchange with the protons of water is much higher than the $\sim 3\text{ kcal mol}^{-1}$ for a weakly hydrogen bonded proton. For LBHBs in proteins, the activation enthalpies for exchange are typically $10\text{--}19\text{ kcal mol}^{-1}$.

Single-Well or Very Strong Hydrogen Bonds

In a single-well hydrogen bond the heteroatoms are so close together that the zero point energies of both hydrogen and deuterium lie above the barrier, as in $\text{A}\cdots\text{H}\cdots\text{B}$ of **Figure 1(c)**. The proton is almost equally bonded between the heteroatoms and is nearly centered between them. Perfectly symmetrical, single-well hydrogen bonds are the strongest, $24\text{--}40\text{ kcal mol}^{-1}$, and are observed in very few small molecules. Such hydrogen bonds have not been detected in proteins.

Physicochemical Properties

Bond Lengths

Weak hydrogen bonds

The covalent bond length ($\text{O}, \text{N}\text{--}\text{H}$) is about $0.95\text{ \AA}$, the length of the electrostatic contact with the second heteroatom is $\geq 1.7\text{ \AA}$, and the separation between the heteroatoms is $\geq 2.6\text{ \AA}$ for $\text{O}\cdots\text{O}$, $\geq 2.7\text{ \AA}$ for $\text{O}\cdots\text{N}$, and $\geq 2.8\text{ \AA}$ for $\text{N}\cdots\text{N}$. Distances separating the heteroatoms can vary from the minima quoted above to as much as $3.5\text{ \AA}$.

LBHBs

In an LBHB (**Figure 1(b)**) the covalent bond is elongated to $1.05\text{--}1.1\text{ \AA}$, and the distance to the second heteroatom is

shortened, for example, to $1.4\text{--}1.5\text{ \AA}$. The separation between heteroatoms is shortened to $\leq 2.6\text{ \AA}$ for $\text{O}\cdots\text{O}$, $\leq 2.7\text{ \AA}$ for $\text{O}\cdots\text{N}$, and $\leq 2.8\text{ \AA}$ for $\text{N}\cdots\text{N}$. A short contact distance between heteroatoms is necessary but not sufficient for an LBHB. To be an LBHB, the hydrogen must be positioned in alignment with the heteroatoms. If the hydrogen is out of alignment, the hydrogen bond will be weak. In the strongest LBHBs, the proton affinities of the heteroatoms are similar or identical. The pK_a of the conjugate acid of a heteroatom is a measure of proton affinity, and heteroatoms in functional groups with similar or identical values of pK_a form the strongest hydrogen bonds.

Single-well hydrogen bonds

The hydrogen lies nearly equidistant from the heteroatoms and is approximately equally shared. The hydrogen is $1.2\text{--}1.3\text{ \AA}$ from each heteroatom, depending on whether it is O or N. Perfectly symmetrical hydrogen bonds are rare and so far observed only in small molecules incorporating special structural constraints that favor single-well hydrogen bonding. Single-well hydrogen bonds range in strength from 24 to 40 kcal mol^{-1} and are the strongest hydrogen bonds. They are formed between heteroatoms of functional groups with identical values of pK_a , in molecules or lattices in which the groups are held in close proximity.

Spectroscopic Properties

Weak hydrogen bonds

The nuclear magnetic resonance (NMR) chemical shift of the proton ranges between 9 and 11 ppm and does not display an isotope effect (isotope shift) upon replacement with deuterium or tritium.

LBHBs

Due to the elongated bond to the hydrogen, $\text{A}\cdots\text{H}$ in **Figure 1(b)**, the proton is substantially deshielded from the nonbonding electrons of the heteroatom. Consequently, the NMR signal for an LBHB involving O or N is downfield in the range of $18\text{--}20\text{ ppm}$ on the chemical shift scale. In the

absence of an internal magnetic effect, such as complexation with a paramagnetic metal ion, a downfield chemical shift is a strong indicator of an LBHB or single-well hydrogen bond. In an LBHB as in **Figure 1(b)**, the zero point energy difference between hydrogen and deuterium leads to deuterium and tritium isotope effects on the NMR signal, with that for deuterium falling upfield by up to 0.6 ppm from that for hydrogen.

As the proton in an LBHB interacts with both heteroatoms to a greater degree than a deuterium, there is an isotope effect on the infrared stretching frequency. The ratio of stretching frequencies $\nu_{\text{OH}}/\nu_{\text{OD}}$ for O–H and O–D engaged in an LBHB differs from the ratio typical of weak hydrogen bonds. For weak hydrogen bonds $\nu_{\text{OH}}/\nu_{\text{OD}} = 1.4$ and represents the mass difference between hydrogen and deuterium. For an LBHB, $1.0 \leq \nu_{\text{OH}}/\nu_{\text{OD}} < 1.4$ in small molecules. This is because of the greater strength of the LBHB with H than with D, and the difference in strength arises because of the differential interactions of H and D with the barrier in **Figure 1(b)**. This property of LBHBs in small molecules has been extensively documented. Among the physicochemical properties that have been used in characterizing LBHBs, the deuterium isotope effect on the infrared stretching frequencies is the only one that has not been applied to LBHBs in proteins.

Single-well hydrogen bonds

Single-well hydrogen bonds display downfield NMR signals at 21–22 ppm when the heteroatoms are O or N. Because the zero point energies of both H and D lie above the barrier in **Figure 1(c)**, single-well hydrogen bonds do not display deuterium or tritium isotope effects on NMR signals. They also do not display deuterium isotope effects on infrared stretching frequencies, that is, $\nu_{\text{OH}}/\nu_{\text{OD}} = 1.4$.

Deuterium Fractionation Factors

Weak hydrogen bonds

Weakly hydrogen bonded protons display deuterium fractionation factors Ψ of about 1.0–1.2. The fractionation factor is the equilibrium constant for the exchange of hydrogen-bonded protons with deuterons in deuterium oxide. A fractionation factor of 1.0 means that deuterons and protons are bound with equal affinities.

LBHBs

In an LBHB, hydrogen is more strongly bonded than deuterium because of the effect illustrated in **Figure 1(b)**. Therefore, the deuterium fractionation factor Ψ for an LBHB is less than 1.0, typically 0.3–0.7.

Single-well hydrogen bonds

In a single-well hydrogen bond, the zero point energies of both H and D lie above the barrier, so that they interact similarly with both heteroatoms. Therefore, the fractionation factor is generally ~ 1.0 .

LBHBs in Enzymes

Serine Proteases

An LBHB was first assigned to the proton bridging His57 and Asp102 in chymotrypsin at low pH and in transition state

analog complexes of chymotrypsin. The assignment was based initially on the downfield NMR chemical shift for this proton of 18.0 ppm for chymotrypsin at low pH and 18.9 ppm for chymotrypsin a transition state-analog complex, and on the close contact between His57 and Asp102 ($< 2.6 \text{ \AA}$) in crystal structures of transition state analog complexes. The LBHB was postulated to stabilize the transition state and facilitate its formation. It was shown in this connection to increase the basicity of His57. The characterization of this LBHB in several transition state analog complexes was completed by measurements of its fractionation factor of 0.3–0.4, its enthalpy of activation for exchange of 14–19 kcal mol^{−1}, and the deuterium and tritium isotope effects on its NMR chemical shifts. The role of the LBHB in catalyzing the formation of the acyl-chymotrypsin intermediate in the action of chymotrypsin is illustrated in **Figure 2**. In resting chymotrypsin, His57 is not protonated and is engaged in weak hydrogen bonding with Asp102. His57 abstracts the proton from Ser195 as it attacks the acyl carbonyl group of a substrate to form the tetrahedral

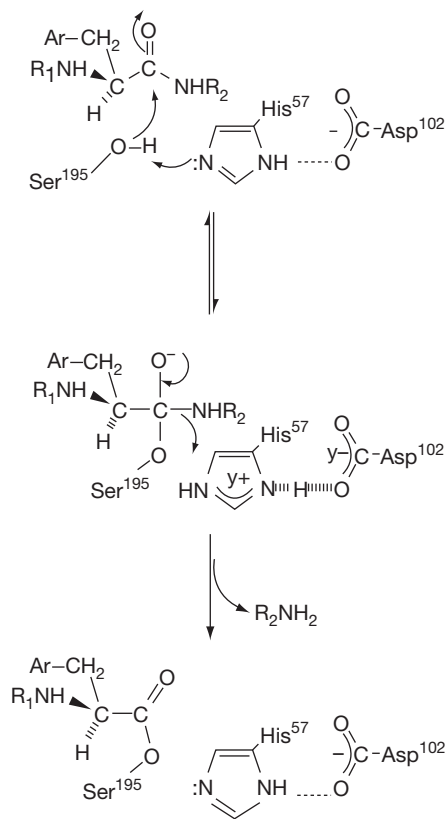


Figure 2 The role of the LBHB in the mechanism of acylation of serine proteases. In the Michaelis complex, the active site histidine (His57 in chymotrypsin numbering) is neutral and weakly hydrogen bonded to Asp201. In the first step, His57 abstracts a proton from Ser195 as it undergoes nucleophilic addition to the substrate carbonyl group. As His57-Nε2 becomes protonated, the hydrogen bond between His57-Nδ1 and Asp-Oδ2 becomes a strong LBHB, and this increases the basicity of His57 as it enters the transition state. The LBHB stabilizes the transition state and transient tetrahedral intermediate. In the second step, His57-Nε2 donates the proton to the N-terminal amino group of the leaving peptide.

intermediate. In the transition state for proton abstraction, His57 becomes protonated on His57-N ϵ 2, and the LBHB is formed between His57-N δ 1 and Asp102. The resulting LBHB stabilizes the transition state and tetrahedral adduct and thereby facilitates the formation of the metastable intermediate. The LBHB maintains the basicity of His57 with a pK_a of 10.6 to 12, intermediate between that of Ser195 (13.4) and that of the amino terminus (9.2) of the peptide leaving group. Thus, the LBHB makes His57 an optimal acid/base catalyst for the acylation of Ser195 by a peptide substrate.

Evidence, similar to but less extensive than that above, for chymotrypsin has been obtained for other serine proteases, especially trypsin and subtilisin. The crystal structure of subtilisin at 0.78 Å resolution shows electron density for hydrogen in the LBHB linking His64 and Asp32. The electron density for the proton places it 1.2 Å from His57-N δ 1 and 1.5 Å from Asp32-O δ 2. ¹⁵N NMR analysis of the LBHB in a transition state analog complex of subtilisin indicates that the proton, 20–30%, is transferred toward Asp32. Both the crystallography and NMR spectroscopy verify that this is an LBHB and support its catalytic role as originally postulated in the action of chymotrypsin.

Serine Esterases

Serine esterases have catalytic triads similar to serine proteases and function by a similar mechanism. Extensive NMR analysis of the hydrogen bond, linking the active site histidine with the active site glutamate, documents the presence of an LBHB.

Δ^5 -3-Ketosteroid Isomerase

The isomerization requires the enolization of the 3-keto group of the substrate to form a homoenolate anion as a metastable intermediate. The homoenolate is within hydrogen bonding distance of Tyr14, and the pK_a of the homoenol is similar to that of tyrosine or phenol. The phenolic inhibitor dihydroequilinen forms an LBHB with Tyr14 in analogy with a putative LBHB in the complex of the metastable homoenolate with Tyr14. The LBHB displays the properties including a downfield NMR signal, low deuterium fractionation factor, and close contact in the crystal structure. The LBHB is postulated to stabilize the homoenolate and thereby facilitate its formation in the catalytic mechanism.

See also: Protein/Enzyme Structure Function and Degradation: Kinetic Isotope Effects; Substrate Binding, Catalysis, and Product Release.

Further Reading

- Choi G, Ha NC, Kim SW, et al. (2000) Asp-99 donates a hydrogen bond not to Tyr-14 but to the steroid directly in the catalytic mechanism of Delta 5-3-ketosteroid isomerase from *Pseudomonas putida* biotype B. *Biochemistry* 39: 903–909.
- Cleland WW (1992) Low-barrier hydrogen bonds and low fractionation factor bases in enzymatic reactions. *Biochemistry* 31: 317–319.
- Cleland WW and Kreevoy MM (1994) Low-barrier hydrogen bonds and enzymic catalysis. *Science* 264: 1887–1890.
- Frey PA, Whitt SA, and Tobin JB (1994) A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science* 264: 1927–1930.
- Gerlt JA and Gassman PM (1993) Understanding the rates of certain enzyme-catalyzed reactions: Proton abstraction from carbon acids, acyl-transfer reactions, and displacement reactions of phosphodiesteres. *Biochemistry* 32: 11943–11952.
- Halkides CJ, Wu YQ, and Murray CJ (1996) A low-barrier hydrogen bond in subtilisin: 1H and 15N NMR studies with peptidyl trifluoromethyl ketones. *Biochemistry* 35: 15941–15948.
- Hibbert F and Emsley J (1990) Hydrogen bonding and chemical reactivity. *Advances in Physical Organic Chemistry* 26: 255–379.
- Jeffrey GA (1997) *An Introduction to Hydrogen Bonding*. New York, NY: Oxford University Press.
- Kuhn P, Knapp M, Soltis SM, et al. (1998) The 0.78 Å structure of a serine protease: *Bacillus lentus* subtilisin. *Biochemistry* 37: 13446–13452.
- Lin J, Cassidy CS, and Frey PA (1998) Correlations of the basicity of His 57 with transition state analogue binding, substrate reactivity, and the strength of the low-barrier hydrogen bond in chymotrypsin. *Biochemistry* 37: 11940–11948.
- Lin J, William M, Westler W, et al. (1998) Fractionation factors and activation energies for exchange of the low barrier hydrogen bonding proton in peptidyl trifluoromethyl ketone complexes of chymotrypsin. *Proceedings of the National Academy of Sciences* 95: 14664–14668.
- Massiah MA, Viragh C, Reddy PM, et al. (2001) Short, strong hydrogen bonds at the active site of human acetylcholinesterase: Proton NMR studies. *Biochemistry* 40: 5682–5690.
- Mildvan AS, Harris TK, and Abeygunawardana C (1999) Nuclear magnetic resonance methods for the detection and study of low-barrier hydrogen bonds on enzymes. *Methods in Enzymology* 308: 219–245.
- Neidhart D, Wei Y, Cassidy C, et al. (2001) Correlation of low-barrier hydrogen bonding and oxyanion binding in transition state analogue complexes of chymotrypsin. *Biochemistry* 40: 2439–2447.
- Westler WM, Frey PA, Lin J, et al. (2002) Evidence for a strong hydrogen bond in the catalytic dyad of transition-state analogue inhibitor complexes of chymotrypsin from proton-triton NMR isotope shifts. *Journal of the American Chemical Society* 124: 4196–4197.
- Zhao Q, Abeygunawardana C, Talalay P, and Mildvan AS (1996) NMR evidence for the participation of a low-barrier hydrogen bond in the mechanism of 5-3-ketosteroid isomerase. *Proceedings of the National Academy of Sciences* 93: 8220–8224.

Luft's Disease

S DiMauro, Columbia University, New York, NY, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 599–601, © 2004, Elsevier Inc.

Glossary

Erythema Redness of the skin due to capillary dilatation or proliferation.

Hypermetabolism A condition of enhanced metabolic activity in one or more tissues, resulting in increased heat production by the body, typically associated with hyperthyroidism.

Polydipsia Prolonged excessive thirst and excessive intake of liquids.

Polyphagia Excessive eating.

Tachycardia Rapid beating of the heart, usually above 90 beats min^{-1} .

Luft's disease is an extremely rare human disorder characterized by loose coupling of mitochondrial oxidative phosphorylation in skeletal muscle, that is, loss of the control normally exerted by adenosine triphosphate on the rate of respiration of muscle mitochondria. Although described only in two patients, Luft's disease is of great historical importance because it was the first example of organellar medicine and it opened the field of mitochondrial diseases, which has expanded into a large and complex area involving all subspecialties of medicine.

Clinical Considerations

Both patients were women, one from Sweden and the other from Jordan, without any family history of similar disorders. The clinical hallmark was hypermetabolism, with fever, heat intolerance, profuse perspiration, polyphagia, polydipsia, and tachycardia at rest. Although both patients suffered from exercise intolerance, weakness was mild. The Jordanian patient also had diffuse erythema of the legs due to capillary proliferation, maybe an attempt by the body to create a radiator to facilitate heat dispersion. The initial diagnosis was hyperthyroidism, but Rolf Luft, an endocrinologist at the Karolinska Institute in Stockholm, ruled out this possibility and astutely suggested that mitochondrial dysfunction of skeletal muscle, the most abundant tissue in the body, could underlie this condition. A muscle biopsy confirmed that mitochondria were excessively numerous and morphologically abnormal, and polarographic studies of isolated mitochondria showed loose coupling. The paper describing these findings is a classic of clinical investigation, which amply justifies the eponym.

Onset of symptoms had been in childhood in both patients, and the course was slowly but relentlessly progressive: both women died, in their middle ages, of respiratory failure. Autopsies were not performed, but clinical evidence strongly suggests that the disease is confined to skeletal muscle, as heart function was normal and there were no symptoms of central nervous system involvement. Electromyography was compatible with myopathy.

Therapy

In 1971, Haydar employed a naive but pioneer approach to gene therapy in the second patient with Luft's disease, who was given chloramphenicol to inhibit mitochondrial protein synthesis and reduce her basal metabolic rate. During the first trial, some reduction of the basal metabolic rate was achieved and was accompanied by mild subjective improvement, but the second trial had to be interrupted because of drug toxicity.

Muscle Biopsy

Morphology

In both women, there was massive proliferation of mitochondria, documented with the modified Gomori trichrome histochemical stain by the presence of ragged-red fibers in the second patient, and by electron microscopy in both patients. Ultrastructurally, there were large pools of mitochondria, predominantly under the sarcolemma. Many of the mitochondria were greatly enlarged and contained tightly packed cristae, some also contained abnormal osmiophilic inclusions (Figure 1). Intramuscular capillaries were overabundant.

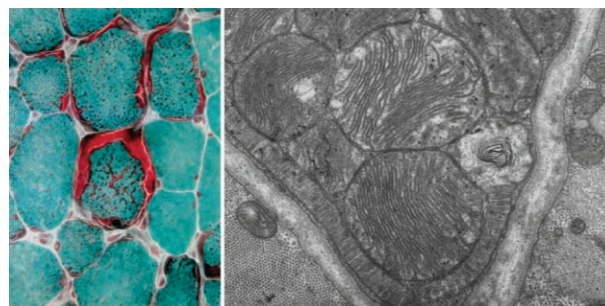
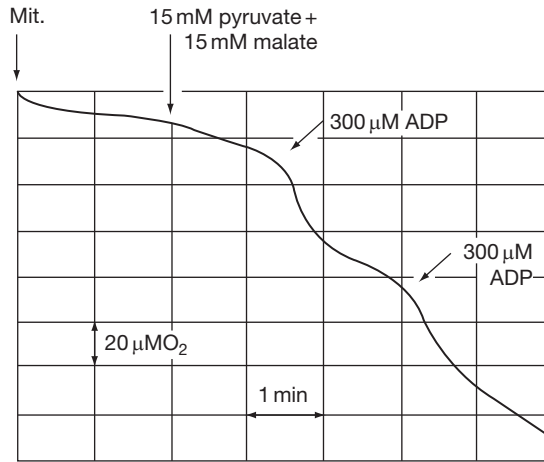
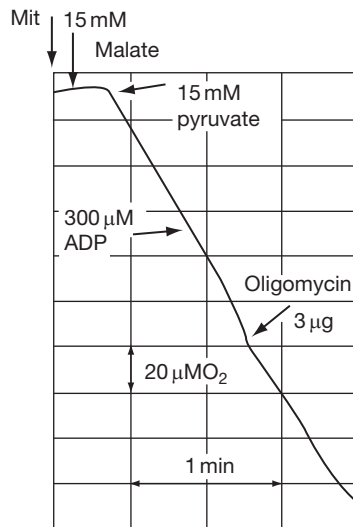


Figure 1 Morphology of a muscle biopsy from the second patient with Luft's disease. The left panel shows the histochemical stain of a frozen section with the modified Gomori trichrome, which reveals abnormal accumulations of mitochondria as reddish blotches (ragged-red fibers). The right panel is an electron micrograph showing a portion of one fiber (bottom) with normal morphology. The adjacent fiber (top) shows massive accumulation of mitochondria under the cell membrane. The accumulated mitochondria are greatly enlarged and have tightly packed cristae.



(a)



(b)

Figure 2 Polarographic tracings of isolated muscle mitochondria from the second patient with Luft's disease (b) and from a normal individual (a). Note how addition of micromolar amounts of adenosine diphosphate (ADP) stimulate oxygen uptake in normal mitochondria, but not in the patient's mitochondria, whose rate of respiration is maximal after the addition of substrates (pyruvate + malate). This high rate of respiration is neither stimulated by ADP nor inhibited by oligomycin. This condition, in which adenosine triphosphate is formed normally but does not control the rate of respiration, is defined 'loose coupling' of oxidation and phosphorylation.

A skin biopsy in the second patient was normal, except for capillary proliferation.

Biochemistry

In polarographic studies of isolated skeletal muscle, mitochondria showed maximal oxygen consumption (respiratory rate) even in the absence of adenosine diphosphate, indicating that respiratory control was lost, although phosphorylation capacity was normal (Figure 2). Spectra and content of cytochromes were normal, and the activities of individual enzymes of the respiratory chain were also normal. However, basal adenosine triphosphatase activity was greatly increased and poorly stimulated by the uncoupler 2,4-dinitrophenol.

In 1976, DiMauro et al. found in the second patient that the rate of energy-dependent calcium uptake by isolated mitochondria was normal, but the amount of calcium accumulated decreased, and intramitochondrial calcium could not be retained and was quickly released into the medium. A vicious cycling of calcium uptake and release was proposed as a mechanism to explain the waste of electrochemical energy as heat.

Genetics

The lack of maternal inheritance, the apparent muscle specificity of the disease, and the lack of large-scale rearrangements of mitochondrial DNA in the second patient are circumstantial evidence that Luft's disease may be due to a nuclear DNA defect. Thus, the molecular basis of the prototypical mitochondrial myopathy remains obscure 40 years after Luft described the first patient, a situation made more difficult by the fact that only one other patient with Luft's disease has been identified; both patients are deceased, and no tissues are available, except fibroblasts from the second patient.

See also: [Bioenergetics: Mitochondrial DNA; Nuclear Genes Involved in Mitochondrial Function and Biogenesis.](#)

Further Reading

- DiMauro S, Bonilla E, Lee CP, et al. (1976) Luft's disease. Further biochemical and ultrastructural studies of skeletal muscle in the second case. *Journal of the Neurological Sciences* 27: 217–232.
- DiMauro S and Schon EA (2003) Mitochondrial respiratory-chain diseases. *New England Journal of Medicine* 348: 2656–2668.
- Haydar NA, Conn HL, Afifi A, et al. (1971) Severe hypermetabolism with primary abnormality of skeletal muscle mitochondria. *Annals of Internal Medicine* 74: 548–558.
- Luft R, Ikkos D, Palmieri G, et al. (1962) A case of severe hypermetabolism of nonthyroid origin with a defect in the maintenance of mitochondrial respiratory control: A correlated clinical, biochemical, and morphological study. *Journal of Clinical Investigation* 41: 1176–1804.
- Pon LA and Schon EA (eds.) (2001) *Mitochondria*. San Diego, CA: Academic Press.

ENCYCLOPEDIA OF **BIOLOGICAL CHEMISTRY**

SECOND EDITION

Volume 3

M–Q

Major Sperm Protein and Sperm Locomotion

K Shimabukuro and T M Roberts, Florida State University, Tallahassee, FL, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Actin Cytoskeletal protein found in nearly all eukaryotic cells. Self-assembly of actin monomers form filaments that impart organizational and mechanical properties in cells and participate in activities ranging from whole cell migration to cytokinesis.

Cell body The trailing portion of a crawling cell, usually packed with organelles, including the nucleus and mitochondria.

Cytoskeleton A network of protein polymers involved in cell structure, organization, and motility.

Lamellipod The flattened forward region of a crawling cell containing the cytoskeletal components responsible for powering motility.

Major sperm protein (MSP) It forms the filaments that comprise the cytoskeleton of nematode sperm, also functions as a signaling molecule.

Protrusion The forward extension of the leading margin of a crawling cell.

Retraction The process that pulls the cell body of a crawling cell forward behind the motile lamellipod.

Nematode Sperm Motility

Nematode sperm lack a flagellum, and, instead, locate eggs by amoeboid locomotion (Figure 1(a)). The movement of nematode sperm is practically indistinguishable from that of other amoeboid cells but they have discarded the actin-based motility apparatus that normally powers crawling movement. These cells have replaced actin with a dynamic cytoskeleton constructed from major sperm protein (MSP). Like actin, MSP polymerizes to form filaments that pack the lamellipod responsible for pulling the cell over its substratum. The organization and dynamics of the MSP cytoskeleton are displayed in a particularly prominent fashion in sperm from the large parasitic nematode, *Ascaris suum*, where the MSP filaments are arranged into several long, branched meshworks, called 'fiber complexes' that span the length of the lamellipod (Figure 1(b)). The fiber complexes are assembled at the leading margin of the lamellipod and taken apart at its base and, thus, treadmill rearward as the sperm moves forward. The rates of cytoskeletal assembly and disassembly are balanced and coupled to the rate of sperm locomotion. Disruption of this finely tuned cytoskeletal choreography, for example, via an uncoupling of fiber complex assembly from its disassembly, has shown that both processes are required for sperm movement. Assembly of MSP produces a force required for protrusion of the leading edge of the lamellipod while disassembly

is involved in creating a retraction force that pulls the cell body forward.

MSP as a Signaling Molecule

MSP had been known as a cytoskeleton protein for several years when Greenstein and colleagues made the surprising discovery that MSP leads a double life as an extracellular signaling molecule. In *Caenorhabditis elegans*, MSP derived from sperm is a bipartite signaling factor that stimulates both oocyte meiotic maturation and contraction of a smooth-muscle-like gonadal sheath cell that surrounds the proximal oocyte, resulting in the release of an egg from the ovary to the spermatheca, the sperm-storage compartment where fertilization takes place. The N-terminal portion of MSP promotes oocyte maturation while the C-terminal part of the molecule stimulates sheath cell contraction. Several aspects of this part of the double life of MSP remain to be unraveled. For example, in order to carry out this signaling function, MSP must be secreted from the lamellipod into the extracellular space. However, sperm are cells ultimately specialized for motility that discard the standard components for secretion very early in their development. Further, MSP lacks an N-terminal signal sequence. Thus, secretion of the protein occurs through unconventional mechanism that appears to involve the budding of small MSP-packed

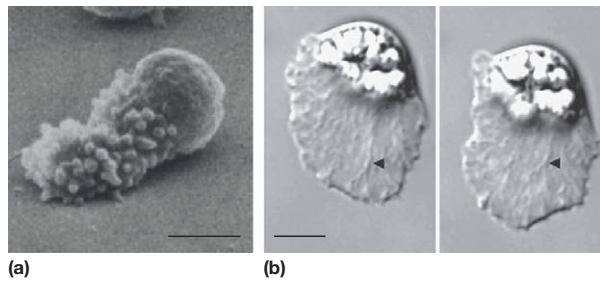


Figure 1 (a) A scanning electron micrograph of a crawling *C. elegans* sperm. These cells crawl by extending a motile lamellipod to grip the substratum and pull the hemispherical cell body forward. (b) Two images of a crawling *Ascaris* sperm illustrating the dynamics of fiber complexes within lamellipod. The time interval between the two images is 10 s. The MSP fiber complexes are readily visible by light microscopy, so their dynamics can be followed without using invasive methods. The fiber complexes treadmill through the lamellipod by assembling at the front and disassembling at the rear. The arrowhead indicates a branch in a fiber complex that remains stationary relative to the substratum as the cell progresses. Scale = 2 μm in (a) and 5 μm in (b). (a) From Nelson GA, Roberts TM, and Ward S (1982) *Caenorhabditis elegans* spermatozoan locomotion: Amoeboid movement with almost no actin. *Journal of Cell Biology* 92: 121–131. (b) From Italiano JE, Stewart M, and Roberts TM (1999) Localized depolymerization of the major sperm protein cytoskeleton correlates with the forward movement of the cell body in the amoeboid movement of nematode sperm. *Journal of Cell Biology* 146: 1087–1095.

vesicles from the sperm lamellipod. The released vesicles then break down allowing MSP to spread into the oviduct.

Molecular Biology and Biochemistry of MSP

MSP, which was first reported in sperm from *C. elegans*, is a 14-kDa protein found exclusively in nematode sperm. As implied by its name, MSP is the most abundant protein in sperm and comprises about 40% of soluble protein in the cell. Although MSP forms the cytoskeleton within the lamellipod of sperm, it does not have any sequence homology to other cytoskeletal proteins such as actin, tubulin, or intermediate filament proteins.

All nematode species that have been examined, including free-living and parasitic forms, have MSP genes. The sequences are well conserved and exhibit at least 80% amino acid identity between species. The numbers of copies of MSP genes differ considerably among species. For example, *C. elegans* has as many as 50 MSP genes and pseudogenes whereas *Ascaris suum* contains only two genes.

MSP from *A. suum* has been studied more extensively than that from other species because the abundance of sperm from these large parasites made it relatively easy to obtain large quantities of the protein. There are two isoforms of MSP in *Ascaris*, named α and β , each composed of 126 amino acids and differing at only four residues. Both isoforms contain an acetylated N-terminal alanine, but no other posttranslational modifications have been detected. β -MSP is slightly less basic (pI 8.3 vs. 8.9) and fivefold more abundant than α -MSP. Both isoforms form homodimers in solution. Monomers are tightly associated ($K_d < 50$ nM) in the dimer, implying that almost all of the unpolymerized MSP in sperm, where the free MSP concentration is ~ 2 mM, is dimeric.

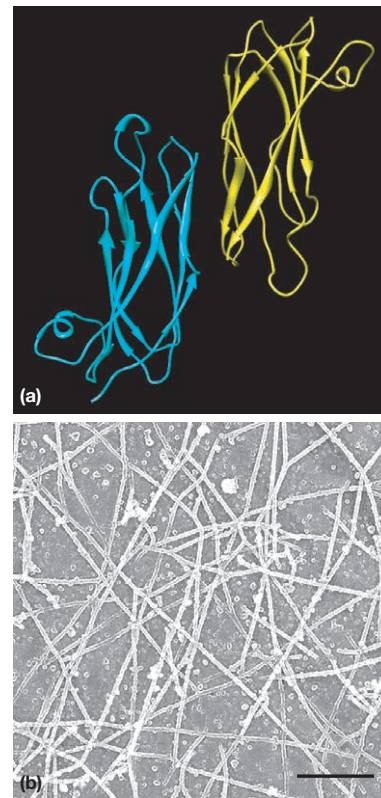


Figure 2 The structure of MSP and the filaments it forms. (a) A ribbon diagram showing the structure of a dimer of α -MSP from *Ascaris* based on X-ray crystallographic analysis. MSP contains a seven-stranded β sandwich with opposing three-stranded and four-stranded β sheets. (b) MSP filaments assembled in a cell-free extract of *Ascaris* sperm and prepared as a platinum replica. Scale = 0.5 μm . (a) From Roberts TM and Stewart M (2000) Acting like actin. The dynamics of the nematode major sperm protein (MSP) cytoskeleton indicate a push-pull mechanism for amoeboid cell motility. *Journal of Cell Biology* 149: 7–12.

Structure of MSP

Structural analysis of *Ascaris* α -MSP by X-ray crystallography and NMR spectroscopy showed that the MSP polypeptide chain has an immunoglobulin-like fold based on a seven-stranded β sandwich with opposing three-stranded and four-stranded β sheets (Figure 2(a)), a feature shared with the N-terminal domain of PapD, a bacterial chaperone involved in assembly of the flagellar rod. However, PapD and MSP do not appear to share a common function. The MSP dimer is a symmetric molecule in which the two subunits are related by twofold rotational symmetry, a property that has important consequences for the structure and function of MSP filaments.

Purified MSP polymerizes into filaments when incubated in a solution containing 30% ethanol. Although this is not physiological, the filaments formed in ethanol are indistinguishable from native filaments obtained from lysed sperm. Each filament is 11 nm in diameter and composed of two helical subfilaments that are wrapped around one another. The polymerizing unit is the symmetric dimer, and the way that these subunits are arranged to form subfilaments produces an MSP

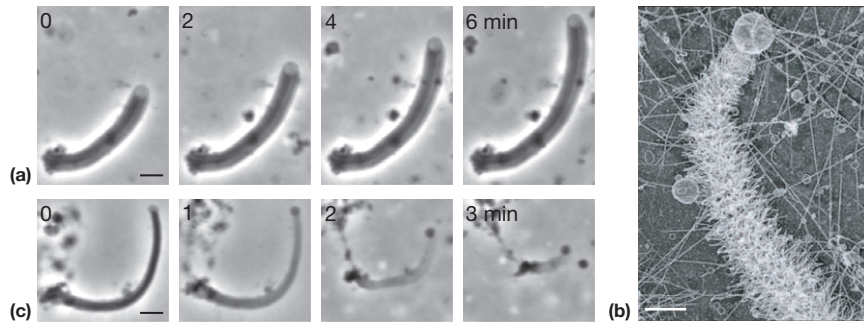


Figure 3 *In vitro* reconstitution of MSP-based motility. (a) Time lapse sequence showing the elongation of a fiber *in vitro*. Addition of ATP to cell-free sperm extract triggers MSP assembly on the surface of spherical vesicle derived from the leading edge membrane of lysed sperm. Polymerizing MSP forms a cylindrical meshwork of filaments, called a ‘fiber’, behind the vesicle. Growing fibers push their vesicle forward in the same manner as the fiber complexes in crawling sperm push the leading edge membrane forward (see [Figure 1\(b\)](#)). (b) An image of a platinum replica of a fiber showing the spherical vesicle at the head of a column of tightly packed MSP filaments. (c) Time lapse sequence of images illustrating retraction of a fiber that has been treated with sperm extract supplemented with PTPase. Note the pronounced shrinkage of the fiber over time accompanied by the decrease in its optical density indicative of the gradual disassembly of the constituent filaments. Scale = 5 μm in (a) and (c), 1 μm in (b).

filament with no structural polarity. This property sets MSP filaments apart from actin filaments and microtubules where structural polarity is critical to important such processes as polymerization kinetics, operation of motor proteins, intracellular organization, and interaction with accessory proteins.

In Vitro Reconstitution of the MSP Motility Apparatus

The MSP motility apparatus from *Ascaris* sperm has been reconstituted *in vitro* by the addition of adenosine triphosphate (ATP) to cell-free sperm extracts. Under these conditions, MSP filaments assemble at the surface of small vesicles, derived from the lamellipodial plasma membrane, which contain the MSP nucleation machinery. This site-directed polymerization produces dense meshworks of filaments behind the vesicle, called ‘fibers’ ([Figures 3\(a\)](#) and [3\(b\)](#)). As the fibers grow they push their vesicles forward in the same way that MSP polymerization into fiber complexes pushes the plasma membrane at the leading edge of the lamellipod forward in crawling sperm.

The *in vitro* motility system has also been used to reconstitute the cytoskeletal dynamics underlying the retraction of the cell body that occurs in crawling sperm. Treatment of fibers with cell-free sperm extract supplemented with a protein tyrosine phosphatase (PTPase) caused the filaments within the fiber to disassemble and, surprisingly, also led to a dramatic shrinkage of the fiber along its length ([Figure 3\(c\)](#)) generating sufficient force to move plastic beads attached to the shortening fiber. By contrast, fibers treated with buffer alone disassemble slowly without appreciable shrinkage. Subsequent assays showed that the PTPase that induces fiber retraction works by activating a serine/threonine phosphatase in the sperm extract. This activity was eventually traced to a protein phosphatase 2A (PP2A) present in the sperm extract. The disassembly associated shrinkage of fibers in response to phosphatase treatment is similar to the disassembly mediated shortening of the fiber complexes at the base of the sperm lamellipod that generates the retraction force that pulls the cell body forward.

Table 1 Known MSP cytoskeletal proteins

Protein	Isoforms	Localization	Function
MSP	α, β	Lamellipod, cytosolic	Subunit of filaments
MFP1	α, β, γ	Lamellipod, cytosolic	Fiber-binding protein, inhibits fiber growth, capping protein?
MFP2	a, b, c	Lamellipod, cytosolic	Fiber-binding protein, accelerates polymerization, threonine-phosphorylated protein
MFP3		Lamellipod, cytosolic	Fiber-binding protein, stabilizes filaments? Threonine-phosphorylated protein
MPOP		Leading edge membrane	Membrane-embedded protein forms nucleation complex with MPAK, tyrosine-phosphorylated protein
MPAK		Leading edge membrane	Ser/Thr kinase, forms nucleation complex, phosphorylates MFP2
PP2A		Lamellipod–cell body interface	Phosphatase, triggers retraction Dephosphorylates MFP3

MSP Accessory Proteins

The versatility of the reconstituted MSP motility system suggests that it contains all of the components needed for both protrusion and retraction. Therefore, fibers became a valuable starting point for the isolation and identification of the accessory proteins involved in sperm locomotion. Stepwise dissection of fibers led, thus far, to the discovery of six such proteins ([Table 1](#) and [Figure 4](#)).

MSP-Organizing Protein

MSP-organizing protein (MPOP) is a 48-kDa integral membrane phosphotyrosine protein and the only membrane

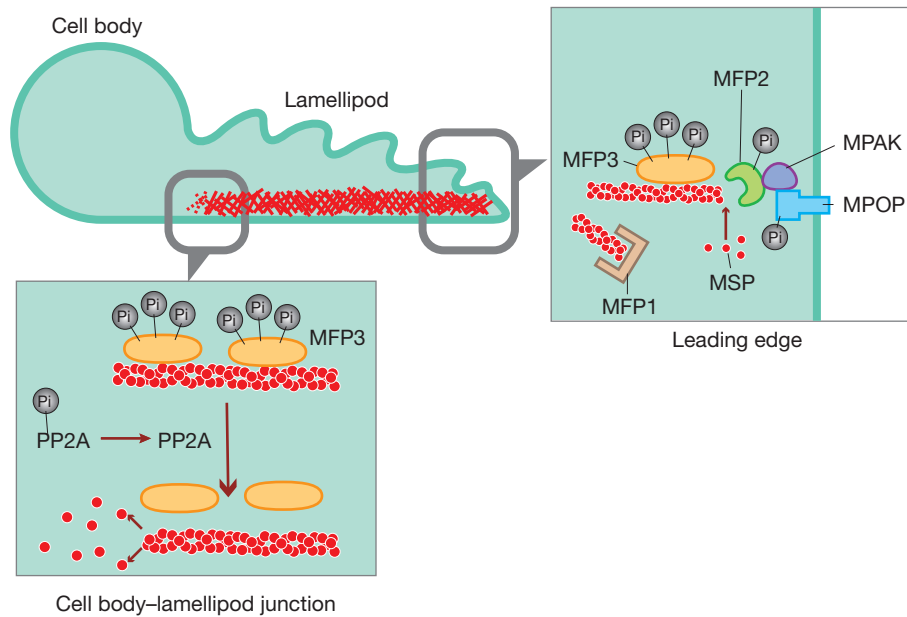


Figure 4 Model of the biochemical basis of the MSP assembly-disassembly cycle associated with motility in sperm. In the schematic drawing of the whole sperm, the MSP filaments that make up a fiber complex are shown in red. At the leading edge of sperm (right inset), tyrosine-phosphorylated MPOP recruits MPAK to the membrane and forms a nucleation complex. MPAK, then, activates MFP2 by phosphorylation which, in turn, stimulates MSP assembly by an unknown mechanism. MSP polymerization may be terminated by MFP1. Phosphorylated MFP3 binds to and may, in fact, stabilize, the newly polymerized filaments. The filaments then treadmill to the base of the lamellipod where they disassemble (left inset). PP2A concentrated at the cell body-lamellipod junction is activated by a PTPase and dephosphorylates MFP3, causing release of MFP3 and depolymerization of the filaments.

protein required for MSP polymerization. Phosphorylated MPOP localizes to the leading edge of crawling sperm where MSP assembles into filaments and forms fiber complexes and is also found in the membrane of the vesicles that assemble fibers *in vitro*. The phosphorylation of MPOP is pH sensitive and its dephosphorylation at low pH results in the loss of polymerization-inducing activity. Only a partial amino acid sequence of MPOP is available and there is, as yet, no significant homology to known proteins from other types of organisms.

MSP Polymerization-Activating Kinase

MSP polymerization-activating kinase (MPAK) is a 34-kDa putative Ser/Thr kinase that exhibits sequence homology to several members of the casein kinase 1 family. MPAK is a cytosolic component but is recruited to the membrane surface by binding to phosphorylated MPOP. Thus, MPAK is located at sites of cytoskeletal assembly *in vivo* and *in vitro* where it phosphorylates a protein required for MSP polymerization at the membrane (see below).

MSP Fiber Protein 2

MSP fiber protein 2 (MFP2) is a cytosolic component in the MSP cytoskeleton that is required for fiber assembly *in vitro*. There are three MFP2 isoforms: a, b, and c, that range in molecular weight from 38 to 41 kDa. MFP2 is phosphorylated on a threonine residue by MPAK, and the phosphorylated form

of the protein incorporate into the fiber. The rate of fiber elongation *in vitro* is directly related to the concentration of MFP2 in the sperm extract but the exact role of MFP2 in MSP polymerization is not known. Crystal structures of both MFP2a and MFP2b revealed that both isoforms consist of two similar domains connected by a linker.

MSP Fiber Protein 1

MSP fiber protein 1 (MFP1) is also a cytosolic component that is incorporated into the MSP cytoskeleton as it is assembled. Although it binds to fibers, MFP1 is not essential for MSP polymerization and also has no known role in retraction. MFP1 causes a concentration-dependent decrease in the rate of fiber growth *in vitro*. There are three MFP1 isomers: α , β , and γ , all of which are acidic proteins ($pI=4.5$) with molecular weights of about 16 kDa. MFP1 contains an MSP domain, which has led to speculation that it may function as a capping protein by binding to the ends of filaments thereby blocking or hindering the addition of MSP dimers to the polymer.

MSP Fiber Protein 3

MSP fiber protein 3 (MFP3) is a 43-kDa cytosolic threonine-phosphorylated protein that contains 19 tandem repeats of a 15-residue motif in the middle of its sequence. Each repeat has at least one threonine residue that can be phosphorylated. The phosphorylated form of the protein binds to MSP filaments but the number of phosphorylated residues required for binding has

not been defined. Upon dephosphorylation, MFP3 is released from the filament and the loss of MFP3 leads to the depolymerization associated with fiber retraction. Therefore, multiple phosphorylation of MFP3 could, in principle, generate a molecule with a large negative charge that could bind to and stabilize filaments constructed from positively charged MSP dimers.

Protein Phosphatase 2A

The discovery that protein phosphatases trigger retraction of fibers led to the identification of an endogenous PP2A in *Ascaris* sperm. This phosphatase is concentrated at the cell body–lamellipod junction where fiber complexes are taken apart. Purified PP2A from sperm can induce retraction *in vitro* in absence of any other soluble components in cell-free extract. Moreover, the activity of this PP2A itself is regulated by phosphorylation. Thus, the induction of retraction by the sperm PP2A is activated by tyrosine dephosphorylation, a property consistent with the observation that sperm extract supplemented with PTPase triggers fiber shrinkage.

See also: [Cell Architecture and Function: Kinesin Superfamily Proteins](#); [Live Cell Imaging of Nuclear Dynamics](#); [Microtubule-Associated Proteins](#).

Further Reading

- Greenstein D (2005) Control of oocyte meiotic maturation and fertilization. *WormBook* 28: 1–12.
- Italiano JE, Jr., Roberts TM, Stewart M, and Fontana CA (1996) Reconstitution *in vitro* of the motile apparatus from the amoeboid sperm of *Ascaris* shows that filament assembly and bundling move membranes. *Cell* 84: 105–114.
- L'Hernault SW (2006) Spermatogenesis. *WormBook* 20: 1–14.
- Miao L, Vanderlinde O, Stewart M, and Roberts TM (2003) Retraction in amoeboid cell motility powered by cytoskeletal dynamics. *Science* 302: 1405–1407.
- Roberts TM and Stewart M (2000) Acting like actin. The dynamics of the nematode major sperm protein (MSP) cytoskeleton indicate a push–pull mechanism for amoeboid cell motility. *Journal of Cell Biology* 149: 7–12.
- Sepsenwol S, Ris H, and Roberts TM (1989) A unique cytoskeleton associated with crawling in the amoeboid sperm of the nematode, *Ascaris suum*. *Journal of Cell Biology* 108: 55–66.
- Tarr DE and Scott AL (2005) MSP domain proteins. *Trends in Parasitology* 21: 224–231.

Mammalian DNA Methyltransferase Structural Themes

X Cheng and H Hashimoto, Emory University School of Medicine, Atlanta, GA, USA

© 2013 Elsevier Inc. All rights reserved.

Introduction

In mammals and other vertebrates, DNA methylation occurs at the C5 position of cytosine (5mC), mostly within CpG dinucleotides (Figure 1(a)), with the Dnmt enzymes using a conserved mechanism that has been studied best in the bacterial 5mC methyltransferase M.HhaI. Briefly, this mechanism involves methyltransferase binding to the DNA, eversion of the target nucleotide so that it projects out of the double helix (base flipping), covalent attack of a conserved Cys nucleophile on cytosine C6, transfer of the methyl group from S-adenosyl-L-methionine (AdoMet) to the activated cytosine C5, and the various release steps. This methylation, together with histone modifications, plays an important role in modulating chromatin structure, thus controlling gene expression and many other chromatin-dependent processes. The resulting epigenetic effects maintain the various patterns of gene expression in different cell types. In mammals, Dnmts include three members, in two families, that are structurally and functionally distinct (see below). The Dnmt3a and Dnmt3b establish the initial CpG methylation pattern *de novo*, while Dnmt1 maintains this pattern during chromosome replication and repair (Figure 1(b)).

Dnmt3 Family

The Dnmt3 family includes two active *de novo* Dnmts, Dnmt3a and Dnmt3b, and one regulatory factor, Dnmt3-like protein (Dnmt3L) (Figure 2(a)). Dnmt3a and Dnmt3b have similar domain arrangements: both contain a variable region at the N-terminus, followed by a PWWP (Pro-Trp-Trp-Pro) domain that may be involved in nonspecific DNA binding or binding of histone H3 trimethylated at lysine 36, a Cys-rich 3-Zn binding ADD (ATRX-DNMT3-DNMT3L) domain, and a C-terminal catalytic domain. The amino acid sequence of Dnmt3L is very similar to that of Dnmt3a and Dnmt3b in the ADD domain, but it lacks the conserved residues required for DNA methyltransferase (MTase) activity in the C-terminal domain.

Dnmt3L is a Regulatory Factor for *De Novo* DNA Methylation

The phenotype of Dnmt3L knockout mice is indistinguishable from that of a Dnmt3a germ cell-specific conditional knockout, with both having altered sex-specific *de novo* methylation of DNA sequences in germ cells and dispersed retrotransposons. These results indicate that Dnmt3a and Dnmt3L are both required for the methylation of most imprinted loci in germ cells. The minimal regions required for interaction between Dnmt3L and Dnmt3a, and for stimulated activity, are in the C-terminal domains of both proteins. The overall Dnmt3a/Dnmt3L

C-terminal complex is ~16 nm long, which is greater than the diameter of a 11-nm core nucleosome (Figure 2(b)). This complex contains two monomers of Dnmt3a and two of Dnmt3L, forming a tetramer with two 3L–3a interfaces and one 3a–3a interface (3L–3a–3a–3L). Substituting key noncatalytic residues at the Dnmt3a–3L or Dnmt3a–3a interfaces eliminates enzymatic activity, indicating that both interfaces are essential for catalysis.

Dimeric Dnmt3a Suggests *De Novo* DNA Methylation Depends on CpG Spacing

Among known active DNA MTases, Dnmt3a and Dnmt3b have the smallest DNA-binding domain (though it is absent altogether in Dnmt3L). However, dimerization via the 3a–3a interface yielded a model such that the two active sites are located in the DNA major groove and dimeric Dnmt3a could methylate two CpGs separated by one helical turn in one binding event (Figure 2(c)). A periodicity in the activity of Dnmt3a on long DNA substrates revealed a correlation of methylated CpG sites at distances of ~10 base pairs, and the structural model of oligomeric Dnmt3a docked to DNA may explain this pattern. Similar periodicity is observed for the frequency of CpG sites in the differentially methylated regions of 12 maternally imprinted mouse genes. These results suggest a basis for the recognition and methylation of differentially methylated regions in imprinted genes, involving detection of both CpG spacing and histone modification (see the next section). CpG DNA methylation patterns in human chromosome 21 are correlated with the CpG periodicity of ~10 base pairs. An ~10 base-pair periodicity has also been evident for non-CpG methylation in embryonic stem cells. Non-CG methylation disappeared upon induced differentiation of the embryonic stem cells and was restored in induced pluripotent stem cells. Similarly, a 10-bp correlation of non-CpG DNA methylation by *Arabidopsis thaliana* DRM2 (which is related to mammalian Dnmt3a) has been observed.

Dnmt3L Connects Unmethylated Histone H3 Lysine 4 to *De Novo* DNA Methylation

Genome-scale DNA methylation profiles suggest that DNA methylation is better correlated with histone methylation patterns. Specifically, DNA methylation is correlated with the absence of histone H3 lysine 4 (H3K4) methylation and the presence of histone H3 lysine 9 (H3K9) methylation. Considering the inverse relationship between H3K4 methylation and DNA methylation, it is important to note that mammalian LSD1 and LSD2, two related lysine-specific demethylases whose substrates include di- and mono-methylated H3K4 (H3K4me2/me1), are absolutely essential in maintaining

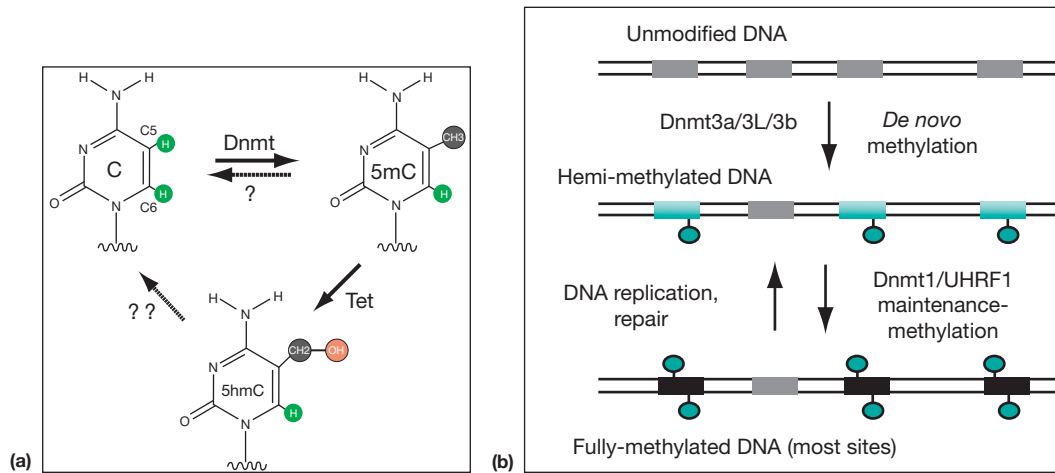


Figure 1 DNA cytosine methylation, hydroxylation, and demethylation. (a) The question mark indicates possible activity of DNA demethylases. Tet indicates conversion of 5mC to 5hmC in mammalian DNA, by the MLL fusion partner TET1. The double question mark indicates that it is currently unknown whether 5hmC is an end product or an intermediate in active DNA demethylation. (b) Maintenance versus *de novo* methylation. The pale-blue rectangular segments are substrate sequences (usually CpG), and the turquoise shapes represent methyl groups on the cytosines. After replication or repair, the duplex is methylated on one strand only (hemi-methylated).

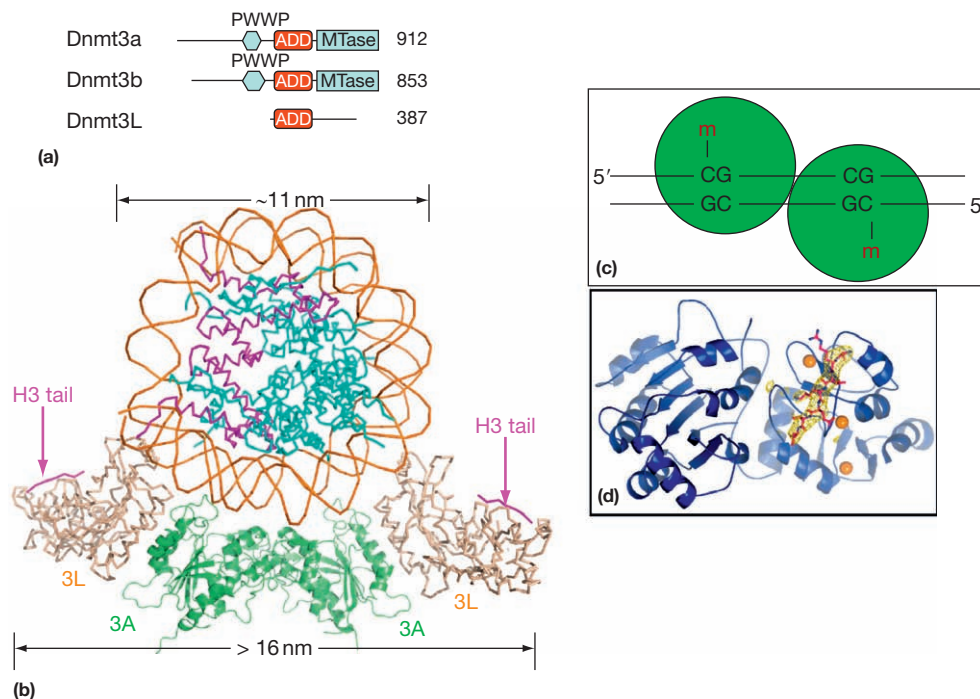


Figure 2 Dnmt3a. (a) Schematic representation of Dnmt3 family. (b) A model of interactions between Dnmt3a-3L tetramer and a nucleosome. A nucleosome is shown, docked to a Dnmt3L-3a-3a-3L tetramer (3a-C in green; 3L full length in gray). The position of a peptide derived from the sequence of the histone H3 amino terminus (purple) is shown, taken from a cocrystal structure with this peptide bound to Dnmt3L. By wrapping the tetramer around nucleosome, the two Dnmt3L molecules could bind both histone tails from one nucleosome. (c) The Dnmt3a dimer could, in theory, methylate two CpGs separated by one helical turn in one binding event. (d) Structure of Dnmt3L with a bound histone H3 N-terminal tail (orange).

global DNA methylation or establishing maternal DNA genomic imprints, respectively. Indeed, disruption of LSD1 results in earlier embryonic lethality and a more severe hypomethylation defect than disruption of the Dnmt's themselves.

The mammalian *de novo* DNA methylation Dnmt3L-Dnmt3a machinery could translate patterns of H3K4 methylation into

heritable patterns of DNA methylation that mediate transcriptional silencing of the affected sequences. Dnmt3a and Dnmt3b bind nucleosomal DNA. Dnmt3a2 is a shorter isoform of Dnmt3a that is also required for genomic imprinting. Dnmt3a2 and Dnmt3b, along with the four core histones, were identified as the main *in vivo* interaction partners of

epitope-tagged Dnmt3L. Peptide interaction assays showed that Dnmt3L specifically interacts with the extreme amino terminus of histone H3; this interaction was strongly inhibited by H3K4 methylation, but was insensitive to modifications at other positions. Co-crystallization of Dnmt3L with the amino tail of H3 showed this tail bound to the ADD domain of Dnmt3L (Figure 2(d)), and substitution of key residues in the binding site eliminated the H3–Dnmt3L interaction. These data suggest that Dnmt3L is a probe of H3K4 methylation, and if the methylation is absent then Dnmt3L induces *de novo* DNA methylation by docking activated Dnmt3a to the nucleosome.

Histone–Dnmt3L–Dnmt3a–DNA interactions have been studied in the budding yeast *Saccharomyces cerevisiae*, which has no detectible DNA methylation and lack Dnmt orthologs. Introduction of the murine methyltransferase Dnmt1 or Dnmt3a in yeast leads to detectible but extremely low levels of DNA methylation. By contrast, a substantially higher level of *de novo* methylation could be achieved in yeast by coexpressing murine Dnmt3a and Dnmt3L. This induced DNA methylation was found preferentially in heterochromatic regions where H3K4 methylation is rare. When genes for components of the H3K4-methylating complex COMPASS/Set1 were disrupted in the context of Dnmt3a/3L overexpression, a greater level of genomic DNA methylation was observed. Deletions or targeted mutations in the ADD domain of Dnmt3L inhibited both global DNA methylation and the ability of Dnmt3L to associate with an H3K4me0 peptide (Figure 2(d)). These same Dnmt3L mutants failed to restore normal DNA methylation to a specific promoter when introduced into embryonic stem cells from Dnmt3L^{−/−} mice.

Structure of Dnmt1

The large Dnmt1 protein contains multiple functional domains (1619 residues for mouse Dnmt1 and 1616 residues for human DNMT1) (Figure 3(a)): an N-terminal region that interacts with Dnmt1 associated protein(s) (DMAP1), followed by a methylation and phosphorylation switch between an adjacent lysine and serine that determines Dnmt1 stability, a proliferating cell nuclear antigen (PCNA) interacting sequence, a replication foci targeting sequence (RFTS) that interacts with the SET- and RING-associated (SRA) domain of UHRF1, a CpG-interacting CXXC domain, a tandem bromo-adjacent homology (BAH) domain, and a DNA methyltransferase domain including the target-recognizing domain (TRD) at the C-terminus.

Two structures are currently available for mouse Dnmt1: a large fragment (residues 357–1612) of Dnmt1 in complex with AdoMet or its product S-adenosyl-L-homocystein (AdoHcy) and a shorter fragment (residues 651–1600) bound to DNA-containing unmethylated CpG sites. In the absence of DNA, the N-terminal RFTS domain is inserted into the DNA-binding surface cleft (Figure 3(b)), indicating that this domain must be removed for methylation to occur. When the isolated RFTS domain added in trans to the RFTS-lacking Dnmt1 protein, the RFTS domain is a DNA-competitive inhibitor of Dnmt1. In the structure of the shorter Dnmt1 fragment, in the absence of the linked RFTS domain, the DNA-bound CXXC domain positions itself in the catalytic domain and prevents aberrant *de novo* methylation of CpG sequence (Figure 3(c)). Thus, a multistep process accompanied by structural changes must occur during maintenance methylation

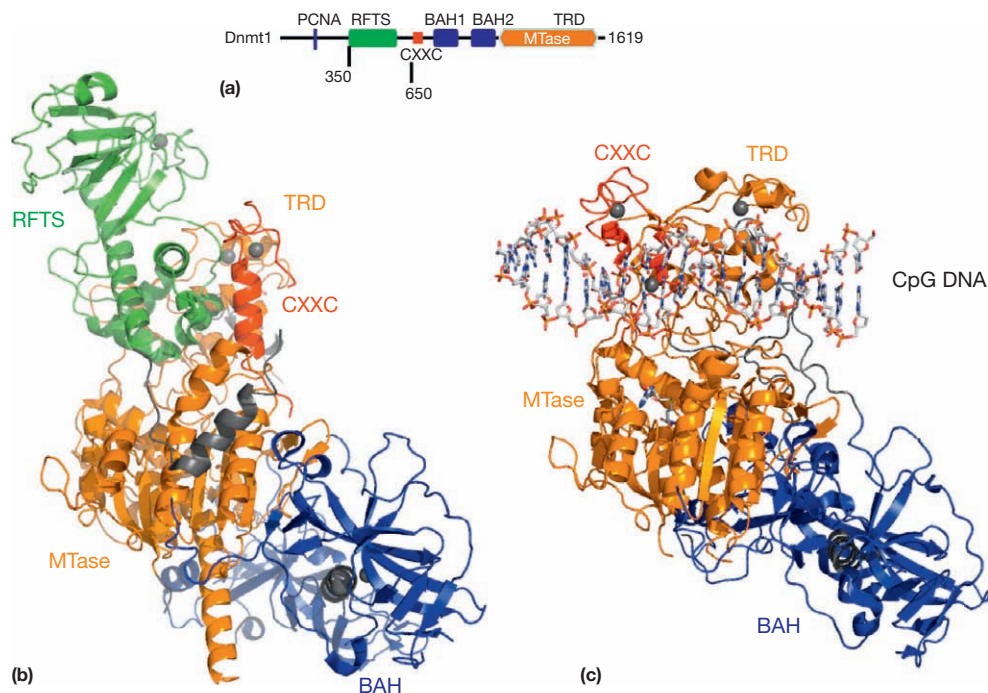


Figure 3 Dnmt1. (a) Schematic representation of Dnmt1. (b) Structure of mouse Dnmt1 (residues 357–1612) containing RFTS (green), CXXC (red), BAH1 (front), and BAH2 (back) in dark blue, and C-terminal MTase domain with TRD in orange (PDB 3AV4). (c) Structure of mouse Dnmt1 (residues 651–1600) with the CXXC domain specifically bound to DNA containing an unmethylated CpG site (PDB 3PT6).

of hemi-methylated CpG DNA by removing both RFTS and CXXC domains away from the catalytic center.

The SRA Domain of UHRF1 Flips 5-Methylcytosine Out of the DNA Helix

An accessory protein UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1) (**Figure 4(a)**) targets Dnmt1 to hemimethylated replication forks (and presumably repair sites). The crystal structure of the SET and RING associated (SRA) domain of UHRF1 in complex with DNA containing a hemimethylated CpG site was determined. They reveal that the SRA domain flips the 5-methylcytosine (5mC) completely out of the DNA helix (**Figure 4(b)**) and is positioned in a binding pocket. The structure also suggests an explanation for the preference for hemimethylated sites. In the major groove side, a backbone carbonyl oxygen is close to the C5 ring carbon of the unmethylated cytosine, forming a $C=O \cdots H-C$ hydrogen bond. The addition of a methyl group to C5 of the unmethylated cytosine would cause a steric clash between the methyl group and SRA.

DNA methyltransferases or DNA repair enzymes use base flipping to gain access to a DNA base to perform chemistry on it, but the SRA domain probably uses base flipping to increase its protein–DNA interface and to prevent the SRA domain from linear diffusion away from the site on the DNA. This may be particularly important for the SRA domain, as its recognition sequence is only two base pairs. We suggest

that the SRA–DNA interaction (through recognition and flipping of the 5mC) serves as an anchor to keep UHRF1 at hemimethylated CpG site where it recruits Dnmt1 for maintenance methylation, and perhaps other proteins such as DNA repair enzymes for mismatch repair. The 5mC base flipping by the SRA domain might also provide a more general mechanism to distinguish the methylated parental strand from the unmethylated daughter strand, an ability particularly important for mismatch repair if an error occurs during DNA replication. Supporting this hypothesis, the expression of UHRF1 is deregulated in cancer cells, and mouse UHRF1-null cells are more sensitive to DNA-damaging agents and DNA replication arrest.

DNA Demethylation via Hydroxylation?

5-Hydroxymethylcytosine (5hmC) has long been noted in bacterial phage, and its presence in mammalian cells was believed to be a by-product of oxidative DNA damage. Recently, using isolated relatively homogenous populations of Purkinje and granule neuronal nuclei of adult mouse brains, Kriaucionis and Heintz found that significant fractions (~40%) of cytosine nucleotides correspond to 5hmCs, the amount of which inversely correlates with 5mC and nuclear heterochromatin in neurons (Kriaucionis and Heintz, 2009). Even more fascinating, a conserved mammalian-specific family of ten-eleven translocation (TET) proteins that converts 5mC to 5hmC (**Figure 1(a)**) was identified. Overproduction of TET1 in human cells led to the appearance of 5hmC. A concomitant reduction in DNA 5mC indicated that 5hmC is an oxidation product of 5mC. 5hmC was detected in embryonic stem cells, especially enriched at the start sites of genes whose promoters bear ‘bivalent’ marks – dual H3K27me3 and H3K4me3 marks. The surprising finding of a 5mC oxidation pathway raises numerous questions, such as whether oxidation of 5mC is an important epigenetic modification, either as an end product or as an intermediate in active DNA demethylation, as supported by the presence of 5hmC DNA excision repair glycosylase. New lines of research will likely be catalyzed by the presence of 5hmC in mammalian DNA.

DNA Demethylation via Glycosylation?

A recent study reported that MBD4, a protein containing an N-terminal methyl-CpG-binding domain (MBD) and a C-terminal glycosylase domain, is phosphorylated via protein kinase C (PKC) by parathyroid hormone stimulation (**Figure (5)**). Phosphorylated MBD4 promotes incision of methylated DNA through glycosylase activity, and a base-excision repair process seems to complete DNA demethylation in the MBD4-bound promoter. Such parathyroid-hormone-induced MBD4 phosphorylation and subsequent DNA demethylation and transcriptional derepression are impaired in Mbd4^{-/-} mice.

Summary

The experimental characterizations of Dnmts and their associated protein factors provide a rapid and convergent picture

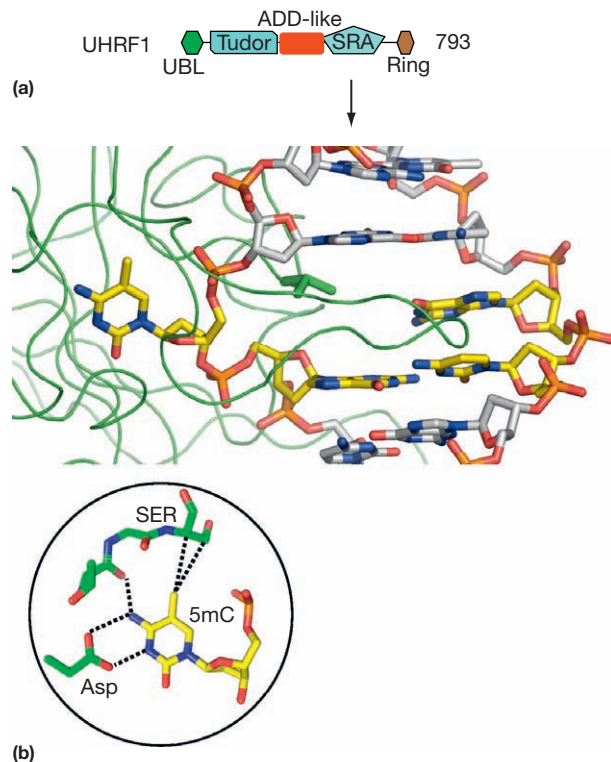


Figure 4 UHRF1. (a) Schematic representation of UHRF1 – a multi-domain protein. (b) Structure of SRA–DNA complex. The 5mC flips out and is bound in a cage-like pocket involving hydrogen bonding and van der Waals interactions (inserted).

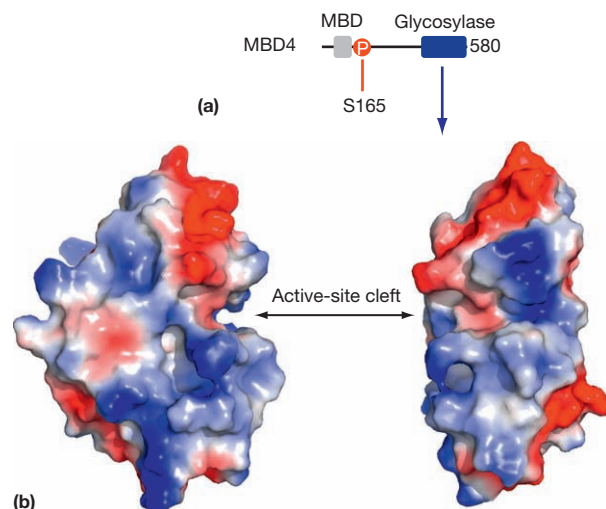


Figure 5 MBD4. (a) Schematic representation of MBD4, with letter P indicating the site of phosphorylation. Serine 165 is conserved among human, mouse, and rat MBD4. (b) Structure of MBD4 glycosylase domain (PDB 1NGN), showing the active-site cleft.

of the kinetic mechanisms (autoinhibition of Dnmt1 by RFTS domain), binding partners such as UHRF1-Dnmt1 and Dnmt3L-Dnmt3a, chromatin recognition such as binding of H3K4me0 by Dnmt3L, RNA-directed DNA methylation, methylation–phosphorylation-dependent regulation of Dnmt1 stability, phosphorylation-induced MBD4 activity, and the discovery of 5hmC in mammalian genome. However, understanding the basis for establishing, maintaining, and disturbing DNA methylation patterns will require a much better understanding of the union between structure and function in the Dnmts and their associated protein factors than we currently possess. Without understanding the interactions and spatial relationships between their modular domains, or whether inter-domain interactions contribute to target specificity, it is not possible to construct a temporal sequence of events or causal relationships in gene silencing.

See also: **Molecular Biology:** DNA Methyltransferases: Eubacterial GATC; DNA Mismatch Repair and Homologous Recombination; DNA Mismatch Repair and the DNA Damage Response.

Further Reading

- Arita K, Ariyoshi M, Tochio H, Nakamura Y, and Shirakawa M (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* 455: 818–821.
- Avvakumov GV, Walker JR, Xue S, et al. (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* 455: 822–825.
- Bostick M, Kim JK, Esteve PO, et al. (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317: 1760–1764.
- Ciccone DN, Su H, Hevi S, et al. (2009) KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* 461: 415–418.
- Cokus SJ, Feng S, Zhang X, et al. (2008) Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452: 215–219.
- De Carvalho DD, You JS, and Jones PA (2010) DNA methylation and cellular reprogramming. *Trends in Cell Biology* 20: 609–617.
- Esteve PO, Chang Y, Samaranyake M, et al. (2011) A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. *Nature Structure and Molecular Biology* 18: 42–48.
- Hashimoto H, Horton JR, Zhang X, et al. (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* 455: 826–829.
- Hu JL, Zhou BO, Zhang RR, et al. (2009) The N-terminus of histone H3 is required for *de novo* DNA methylation in chromatin. *Proceedings of the National Academy of Sciences of the United States of America* 106: 22187–22192.
- Jia D, Jurkowska RZ, Zhang X, Jeltsch A, and Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation. *Nature* 449: 248–251.
- Kim MS, Kondo T, Takada I, et al. (2009) DNA demethylation in hormone-induced transcriptional derepression. *Nature* 461: 1007–1012.
- Kriaucionis S and Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324: 929–930.
- Meissner A, Mikkelsen TS, Gu H, et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454: 766–770.
- Ooi SKT, Qiu C, Bernstein E, et al. (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* 448: 714–717.
- Pastor WA, Pape UJ, Huang Y, et al. (2011) Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 473: 394–397.
- Sharif J, Muto M, Takebayashi S, et al. (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450: 908–912.
- Song J, Rechkooblit O, Bestor TH, and Patel DJ (2011) Structure of DNMT1–DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* 331: 1036–1040.
- Tahiliani M, Koh KP, Shen Y, et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324: 930–935.
- Takeshita K, Suetake I, Yamashita E, et al. (2011) Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). *Proceedings of the National Academy of Sciences of the United States of America* 108: 9055–9059.
- Wang J, Hevi S, Kurash JK, et al. (2009) The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nature Genetics* 41: 125–129.

Mass Spectrometry of Native Complexes

R J Rose and A J R Heck, Biomolecular Mass Spectrometry and Proteomics, Utrecht University, Utrecht, The Netherlands; Netherlands Proteomics Center, Utrecht, The Netherlands

© 2013 Elsevier Inc. All rights reserved.

Glossary

Collisional cross-section The cross-section of a molecular structure that can undergo collisions with a buffer gas, averaged over all orientations.

Electrospray ionization Method of producing gas phase molecular ions by the application of voltage to a solution; solvent evaporation and charge-induced fission result in multiply charged protein ions.

Ion mobility spectrometry Measurement of mobility of ions through a buffer gas under the influence of an electric

field; mobility of an ion is determined by its mass, charge, and physical size.

Native mass spectrometry Mass analysis of molecular structures, whereby the analyte retains as much as possible its ternary and quaternary solution phase structure.

Tandem mass spectrometry The mass analysis of molecular structures prior to and subsequent to a gas phase reaction, typically an activation/dissociation event.

Introduction

The association of proteins into large and functional assemblies is essential for life; indeed, the average number of interacting partners of a protein during its lifetime is thought to be as high as eight. The resulting complexes can be large and heterogeneous, as well as dynamic, which can make the study of their structural properties challenging. Mass spectrometry (MS) offers a unique possibility to analyze intact protein complexes in a native-like state, over a theoretically unlimited mass range. This allows the deduction of information regarding the structure and dynamics of these assemblies. Native MS is a comparatively new and increasingly popular technology for such applications, and provides important information highly complementary to other techniques commonly used in structural biology and molecular biophysics.

Structural Biology Analysis of Protein Complexes

There is increasing demand for structural biological studies of large protein complexes and assemblies, and a concomitant requirement for applicable techniques. Native MS is a unique method, allowing determination of and insight into overall mass of a protein complex, stoichiometry, mass of composite subunits, subunit arrangement, complex stability, and dynamics. Furthermore, recent coupling of ion-mobility spectrometry (IMS) with native MS has added a new dimension to these analyses, whereby the overall size or shape of the analyte can additionally be measured. Thus, the data obtained from native MS can bridge the gap between high-resolution nuclear magnetic resonance (NMR) and X-ray crystallography structures, usually limited to relatively small macromolecules, and high-throughput protein identification and interactomics studies.

Native MS Methodology and Instrumentation

In order to analyze proteins and protein complexes by MS, they must first be transferred from solution phase to gas phase, while maintaining native contacts and native (or native-like) structure. The most effective ionization technique for achieving this is nano-electrospray ionization (nanoESI), a miniaturized version of ESI, a method widely exploited for the MS analysis of biomolecules. Importantly, for nanoESI, protein complexes can be in aqueous solutions at neutral pH, in which they will generally populate native states. Due to the vaporization step necessary prior to MS, the presence of involatile salts and buffers is problematic; a typical buffer-like salt adopted for native MS analyses, to circumvent this problem, is ammonium acetate. During nanoESI, solvent and buffer evaporate, leaving multiply charged macromolecular ions (**Figure 1**). With this soft ionization technique, native structure can to a certain degree be retained; for protein complexes, this means the retention of both tertiary and quaternary structure.

The production of protein ions with multiple charges by ESI results in numerous peaks on a mass-to-charge (m/z) scale, with each peak relating to a different charge state of the same protein (**Figure 1**). The molecular mass of the protein (or complex) can be determined from these charge-state peaks. Furthermore, the charge state of a protein, that is, the number of protons which have associated during the ESI process, is related to the conformational state of the protein. Compact, folded conformations obtain fewer charges than extended, unfolded conformations, due to the decreased number of solvent-accessible basic sites, and thus the two different states will result in discrete charge-state distributions within a mass spectrum.

In order to analyze the protein complex ions produced by ESI, a mass spectrometer with a large m/z range accessible is typically necessary. For these purposes, time-of-flight (ToF) or hybrid ToF (e.g., quadrupole time-of-flight, Q-ToF) mass

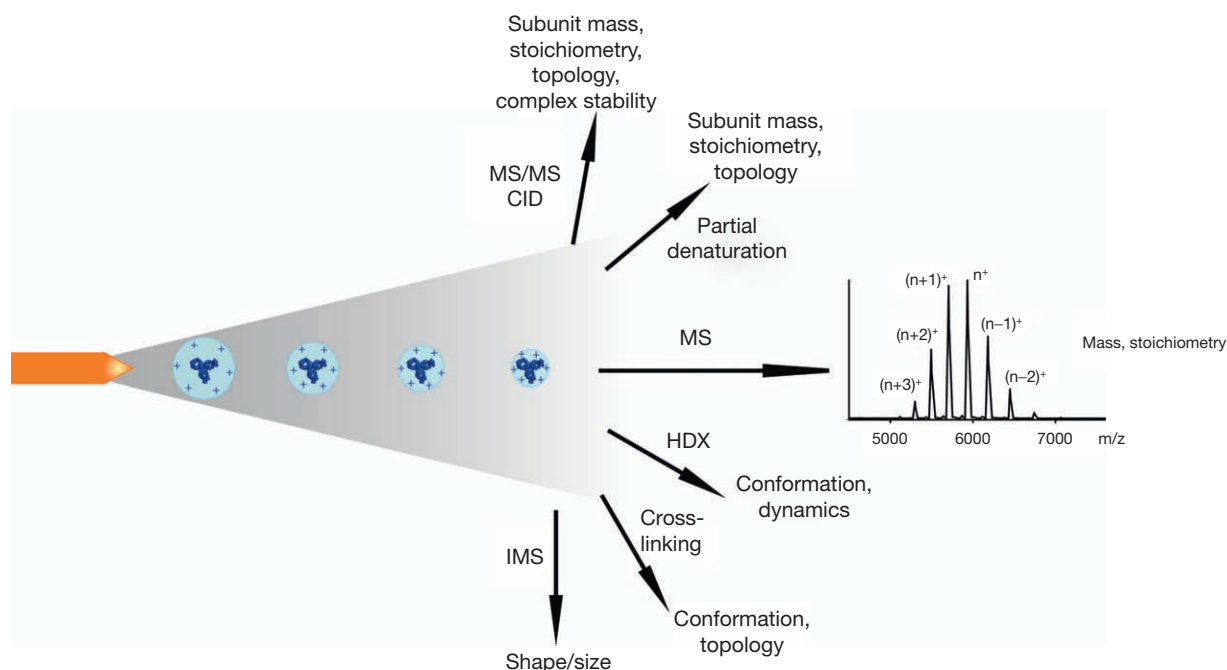


Figure 1 (Left) Schematic of nanoESI, whereby multiply charged protein ions are produced after a voltage is applied to a capillary. (Right) Typical mass spectrum of a multiply charged protein ionized by ESI. Different types of MS experiments, and the information gained from them, are listed (CID=collision-induced dissociation, HDX=hydrogen-deuterium exchange, IMS=ion-mobility spectrometry).

analyzers are usually employed. ToF instruments have theoretically unlimited m/z ranges, although some instrumental challenges currently impose a maximum on the achievable values. Modifications have been made to Q-ToF instruments to improve the transmission and detection efficiency of large macromolecular ions, and ions can now be observed up to approximately m/z 100 000. An important adaptation is to increase the pressure in certain regions within the mass spectrometer, which results in improved transmission of large noncovalent ions. Other modifications include the use of low-frequency quadrupoles, as well as low-frequency ToF pusher voltages, to allow ions with high m/z values to traverse these sections of the mass spectrometer.

In addition to the mass spectral information typically obtained from these instruments, performing MS in combination with other experimental techniques (summarized in [Figure 1](#)), such as hydrogen-exchange, cross-linking, and IMS, can provide many data regarding biophysical properties of macromolecules. Tandem MS (MS/MS) experiments are commonly used to glean more information regarding the structural properties of the complex. In Q-ToF instruments, a collision cell between the quadrupole and ToF analyzers is filled with an inert gas, such as argon, and acceleration of the ions through this cell causes energetic collisions. This activation can result in dissociation of a protein complex into composite parts (collision-induced dissociation, CID), which are then identified by the ToF MS.

Recent instrumental developments have allowed the integration of IMS into MS experiments on large protein complexes. This technique can be seen as a gas phase equivalent of gel filtration chromatography, and adds another dimension of analysis or separation, based on the mobility of ions

traveling through a gas-filled cell. This provides a measurement of the ion's collisional cross-section, an indication of the overall size and shape of the ion. The ability to gain such additional information is enhancing further the application of native MS within the structural biology field.

Protein Complex Composition

The measurement of molecular weight of a native protein complex can provide direct evidence for or confirmation of the subunit stoichiometry. By denaturing the complex, for example, using organic or acidic solution modifiers (a mixture of acetonitrile or methanol with formic acid or trifluoroacetic acid is often utilized in MS), the masses of the individual composite subunits can be measured, to aid in interpretation of the mass of the intact complex.

An example of the use of native MS in confirming the stoichiometry of a large, homo-oligomeric complex is in the study of the hepatitis-B virus (HBV) capsid. This particle was previously known to populate two different morphologies, with the number of subunits assumed from their symmetry. The two particles were confirmed by MS experiments to contain 180 or 240 copies of the subunit monomer, as determined from the measured masses of 3.012 and 4.014 MDa ([Figure 2](#)). MS has also been used to determine for the first time the composition of the yeast RNA polymerase III. It was shown that this complex contained a single copy of each of the 17 subunits previously identified. Furthermore, the capability of MS to observe simultaneously populated species also revealed the presence of a complex lacking a heterodimer of two of the subunits. The tryptophan RNA-binding

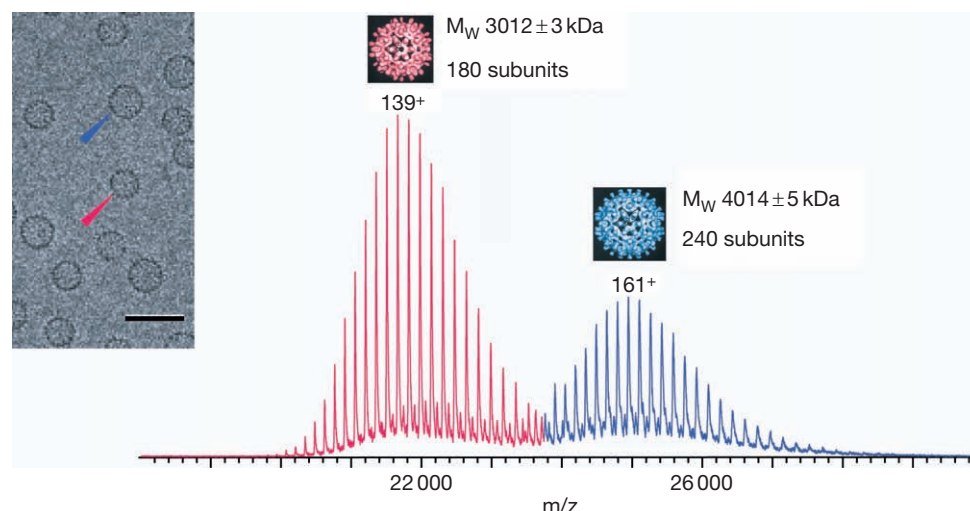


Figure 2 NanoESI mass spectrum of hepatitis-B virus capsids. Two distinct charge-state distributions are observed, corresponding to two stoichiometries of subunit assembly: 180 subunits (measured mass 3012 kDa, pink) and 240 subunits (measured mass 4014 kDa, blue). Inset: capsid particles as visualized by cryo-electron microscopy, showing two distinct geometries (indicated by arrowheads). Scale = 50 nm.

attenuation protein (TRAP) was known to populate a homo-oligomeric 11-mer state, and had been proposed also to form higher order structures. MS was used to analyze TRAP in the presence of trp ligand to identify both the oligomeric composition of the complex, and its binding behavior. In addition to the 11-mer, a 12-mer TRAP complex was identified, which was found to bind 11 trp ligands, and also a TRAP24-trp22 complex. These findings not only confirmed the composition of TRAP complexes, but additionally gave insights into how the oligomers assemble, and differences in trp binding.

Dynamic Protein Complexes

MS analyses of protein complexes can be performed on a relatively fast timescale (data can be acquired within a few minutes). This, plus the advantage of monitoring copopulated species, allows dynamic processes to be followed in real time. Reactions between different protein complexes can be studied, with the relative rates of reactant decrease and product formation being measured, as well as the possibility to observe and identify on-pathway intermediates. This can reveal important information about kinetic processes and reaction mechanisms. For example, the ordered assembly of specific subunits into bacterial pilus fibers was shown to be kinetically controlled using time-resolved MS studies. Monitoring the transthyretin tetramer as a function of time by MS also revealed very interesting behavior regarding the exchange of subunits in and out of this oligomer, which is involved in systemic amyloid disease. Using stable isotope labeling to increase the molecular weight of a fraction of the proteins, the rate of exchange could be measured. Mixing the wild-type with a variant which is particularly associated with amyloidosis, L55P, showed the formation of mixed tetramers. Interestingly, it was also found that the L55P affected the rate of wild-type transthyretin dissociation. In this way, information about the formation of hybrid tetramers, with important biological implications, was obtained from native MS experiments.

A great advantage of MS is that it can be utilized to analyze systems of a very complex and dynamic nature. An example of this is in the analysis of an intricate circadian clock system from cyanobacteria, which can be reproduced *in vitro* with three proteins. Native MS has been used to identify the many different complexes these proteins form, as well as monitoring their appearance and disappearance throughout this cyclic reaction.

Topology of Complexes and Assembly Pathways Determined by Dissociation

The use of CID can provide information about the quaternary structure and overall architecture of protein complexes. This process is analogous to CID of peptides or small molecules, where the activation energy eventually causes fragmentation of covalent bonds, but here, the activated complex generally loses a single, noncovalently bound subunit. The dissociating subunit is typically highly charged compared to the remaining subcomplex; indeed, it is thought that the dissociation occurs concomitantly with subunit unfolding, explaining the ability of this protein to sequester a high proportion of the protons. Which subunit dissociates first from a complex is apparently dependent on two factors: first, how strong or extensive interactions are with other subunits; second, whether the subunit is located at the periphery or in the core of the complex. This is well demonstrated by CID experiments on the wild-type and a variant of the 20S proteasome complex from *Rhodococcus*. The K151A/K153A variant forms a $\alpha_7\beta_7$ structure, where a ring of α subunits is bound face-to-face with a ring of β subunits. When subjected to collisional activation, a β subunit is released (**Figure 3(a)**); this is likely due to the lower area of surface contact between β subunits (800 Å²) compared to between α subunits (1440 Å²). However, in the wild-type complex, a $\alpha_7\beta_7\beta_7\alpha_7$ structure is formed, whereby two β rings are sandwiched between two α rings. In this case, collision activation induces the loss of an α subunit, since the topology of the complex prevents the

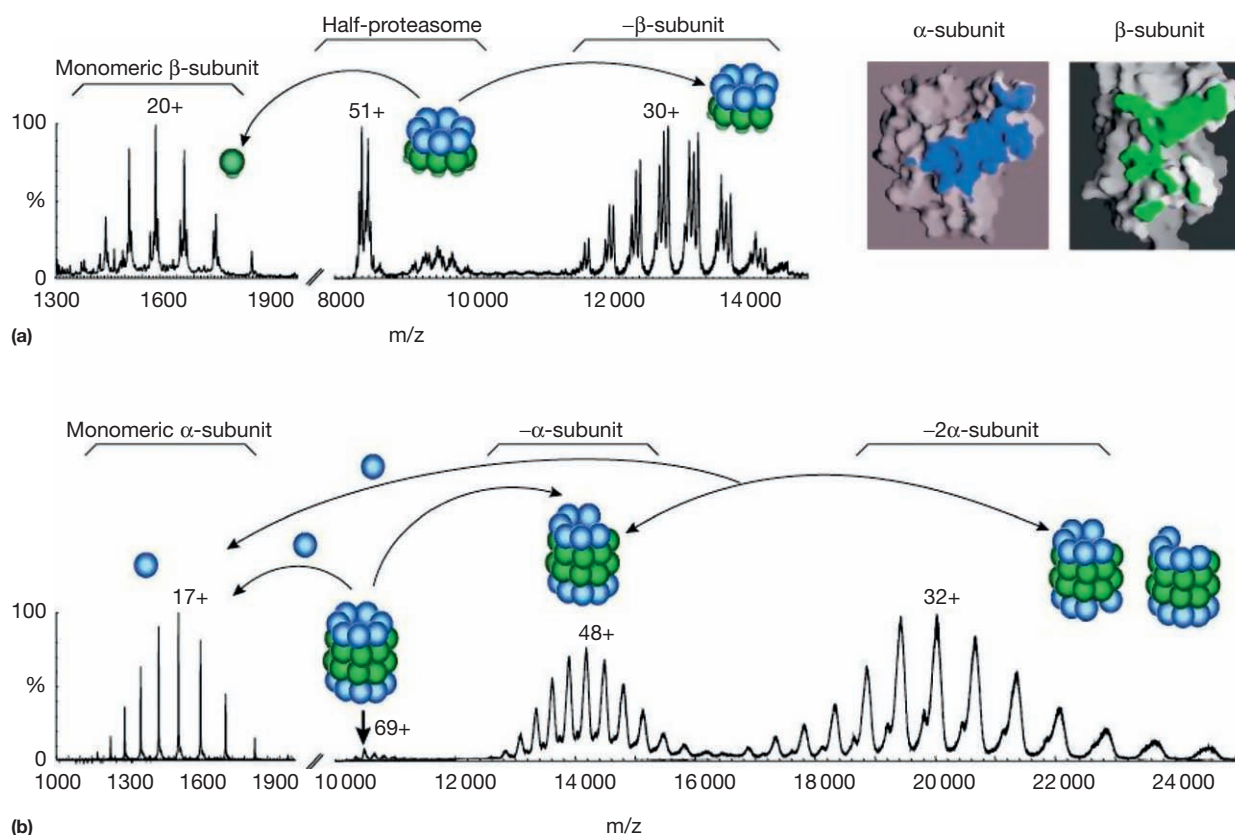


Figure 3 Tandem MS (CID) spectra of (a) K151A/K153A and (b) wild-type 20S proteasome, showing dissociation of beta (green) or alpha (blue) subunits. Inset (a) shows contact surface area of alpha and beta subunits.

dissociation of a β monomer (Figure 3(b)). Thus, it is clear that CID MS can provide information regarding the overall topology of protein complexes.

An alternative angle from which to approach the deduction of the topology of a protein complex is to use partial denaturation in solution to produce subcomplexes of the overall assembly. In this way, local protein interactions can be mapped, and the location of different subunits in relation to each other determined. Indeed, using different dissociation methods allows a wide range of different subcomplexes to be identified, and thus extensive maps of the topology of complexes to be determined. For example, in combination with computational modeling, MS data has provided a detailed picture of the quaternary structure of the yeast RNA exosome, a complex containing 10 distinct subunits. Thus, the use of either gas-phase (CID) or solution-phase dissociation evidently allows important structural information of large, homo- or heterogeneous protein complexes to be obtained.

Maintaining Native Structure/Function in the Gas Phase

An important question to address when using MS in the analysis of protein complexes is to what extent the native structure is maintained during the course of the experiment. Whether or not the structures of gas phase complexes bear close resemblance to those they adopt in solution has been the subject of

wide debate. Obviously, the desolvation of such a molecule is likely to significantly affect the native interactions it forms, altering the viability and/or strength of hydrophobic and electrostatic interactions, and hydrogen bonds. However, many factors can affect whether structural features are maintained or not, including the identity of the complex of interest, the retention of residual interfacial or surface-bound solvent molecules, and the number of charges obtained during the ESI process. Much evidence is being accrued for the validity of gas phase analyses in determining solution phase properties. For example, the tobacco mosaic virus capsid was subjected to ESI-MS, after which it was collected and analyzed by transmission electron microscopy, revealing that it retained its native three-dimensional structure. Moreover, it was proven still to be infective, implying that the particle either remains unchanged throughout the MS process, or undergoes only changes that are reversible upon rehydration.

The recent advent of IMS hyphenated with MS for the analysis of large proteins and complexes has allowed more proof of the retention of native structures to be gathered. In a seminal IMS-MS study of protein complexes, the native ring structure of the oligomeric TRAP complex was seen to be preserved in the gas phase. A measurement of the size of an ion, the collisional cross-section, can be extracted from IMS data, and compared to structural models of a molecule, for example, from X-ray crystallography or computational methods. In the case of TRAP, and many other proteins and complexes, reasonable agreement is seen between the IMS-MS data and the solved structures.

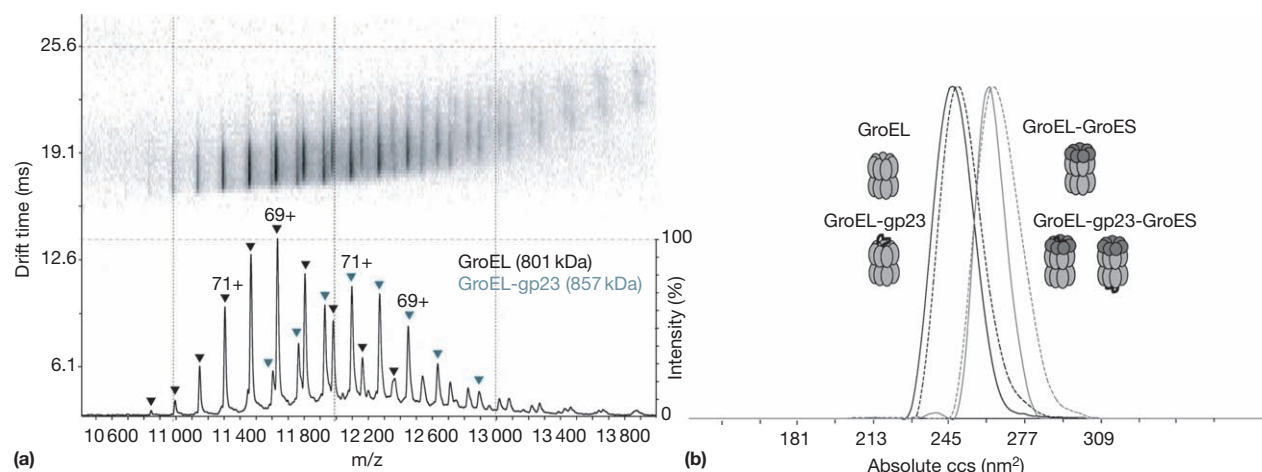


Figure 4 (a) Mass spectrum and associated ion mobility drift time data for GroEL alone and bound to a substrate, gp23. (b) Collisional cross-sections for complexes of unbound GroEL, GroEL–GroES, GroEL–gp23, and GroEL–GroES–gp23.

However, the effect of charge can be apparent on gas phase structures, in that a high number of charges can give rise to an extended or even a collapsed protein structure. Thus, although native or native-like structures evidently can be studied by ESI–MS, it is always wise to interpret such data with care.

IMS–MS is increasingly being applied as an analytical tool to study biological systems. One such study focused on the GroEL–GroES chaperonin system. Substrate polypeptides were seen to remain bound to the chaperonin in the gas phase (Figure 4(a)). Furthermore, the dimensions of the GroEL or GroEL–GroES complexes were found not to change upon substrate binding, implying that the substrate is trapped inside the hollow structure (Figure 4(b)). These results are part of the cumulating evidence that protein assemblies can largely retain their solution phase structural properties in the gas phase.

Pushing the Boundaries of Native MS

Native MS is increasingly being used for the study of native complexes and their structural and biophysical properties. Complexes can now be analyzed into the megaDalton range, and thus with particle sizes of tens of nanometers. Although the theoretical size of particles measurable by MS is unlimited, practicalities currently prevent the analysis of much larger complexes. Some problems also arise with charge-state resolution and complete desolvation, as well as assignment of charge states for accurate mass assignment. Continuing instrument and software development should aid with the improvement of such factors. Furthermore, parallel improvements in the resolution and calibration of IMS, as well as the continuing advance in computational modeling methods, should further the level of detail possible from IMS–MS experiments.

One field of structural biology very challenging for MS is the analysis of membrane proteins and complexes thereof. Recent developments, however, have pushed this field forward. It has been shown that ESI of a micelle containing a protein complex, followed by CID of the detergent molecules

within the micelle, allow intact complexes of membrane proteins to be analyzed by MS. This opens up an avenue for extending the application of native MS to samples not previously possible.

Another area in which native MS of protein complexes is likely to be increasingly exploited is in biopharmaceutical research. Here, it can be applied to numerous analytical problems, including protein ligand screening, measuring antibody–antigen interactions, epitope mapping, and monitoring conformational changes. The ability to obtain unique information, with a fast data acquisition time and low sample consumption, means that this technique fits ideally into the niche of biopharmaceutical analysis. As native MS pushes the boundaries of the analysis of protein complexes, its application range broadens, and it is proving to be an invaluable tool in many areas of structural biology.

See also: **Protein/Enzyme Structure Function and Degradation: Chaperonins.**

Further Reading

- Ashcroft AE (2005) Recent developments in electrospray ionization mass spectrometry: Noncovalently bound protein complexes. *Natural Product Reports* 22: 452–464.
- Atmanene C, Wagner-Rousset E, Malissard M, et al. (2009) Extending mass spectrometry contribution to therapeutic monoclonal antibody lead optimization: Characterization of immune complexes using noncovalent ESI–MS. *Analytical Chemistry* 81: 6364–6373.
- Barrera NP, Di Bartolo N, Booth PJ, and Robinson CV (2008) Micelles protect membrane complexes from solution to vacuum. *Science* 321: 243–246.
- Brettschneider C, Rose RJ, Hertel S, et al. (2010) A sequestration feedback determines dynamics and temperature entrainment of the KaiABC circadian clock. *Molecular Systems Biology* 6: 389.
- Ganem B, Li YT, and Henion JD (1991) Detection of noncovalent receptor–ligand complexes by mass spectrometry. *Journal of the American Chemical Society* 113: 6294–6296.
- Heck AJ (2008) Native mass spectrometry: A bridge between interactomics and structural biology. *Nature Methods* 5: 927–933.
- Heck AJ and van den Heuvel RH (2004) Investigation of intact protein complexes by mass spectrometry. *Mass Spectrometry Reviews* 23: 368–389.

- Hernandez H and Robinson CV (2007) Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nature Protocols* 2: 715–726.
- Karas M, Bahr U, and Dulcks T (2000) Nano-electrospray ionization mass spectrometry: Addressing analytical problems beyond routine. *Fresenius' Journal of Analytical Chemistry* 366: 669–676.
- Lorenzen K, Vannini A, Cramer P, and Heck AJ (2007) Structural biology of RNA polymerase III: Mass spectrometry elucidates subcomplex architecture. *Structure* 15: 1237–1245.
- Rose RJ, Verger D, Daviter T, et al. (2008) Unraveling the molecular basis of subunit specificity in P pilus assembly by mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 105: 12873–12878.
- Ruotolo BT, Giles K, Campuzano I, et al. (2005) Evidence for macromolecular protein rings in the absence of bulk water. *Science* 310: 1658–1661.
- Sharon M and Robinson CV (2007) The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annual Review of Biochemistry* 76: 167–193.
- Siuzdak G, Bothner B, Yeager M, et al. (1996) Mass spectrometry and viral analysis. *Chemistry and Biology* 3: 45–48.
- Synowsky SA, van Wijk M, Raijmakers R, and Heck AJ (2009) Comparative multiplexed mass spectrometric analyses of endogenously expressed yeast nuclear and cytoplasmic exosomes. *Journal of Molecular Biology* 385: 1300–1313.
- Taverner T, Hernandez H, Sharon M, et al. (2008) Subunit architecture of intact protein complexes from mass spectrometry and homology modeling. *Accounts of Chemical Research* 41: 617–627.
- Uetrecht C, Rose RJ, van Duijn E, Lorenzen K, and Heck AJ (2010) Ion mobility mass spectrometry of proteins and protein assemblies. *Chemical Society Reviews* <http://dx.doi.org/10.1039/b914002f>.
- Uetrecht C, Versluis C, Watts NR, et al. (2008) High-resolution mass spectrometry of viral assemblies: Molecular composition and stability of dimorphic hepatitis B virus capsids. *Proceedings of the National Academy of Sciences of the United States of America* 105: 9216–9220.
- van den Heuvel RH, van Duijn E, Mazon H, et al. (2006) Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Analytical Chemistry* 78: 7473–7483.

Meiosis

N Hunter, Howard Hughes Medical Institute, University of California Davis, Davis, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Chiasma The point where homologous chromosomes exchange arms. Chiasmata (pl.) are the cytological manifestations of crossing-over.

Crossing-over A reciprocal exchange between two DNA molecules. This occurs via the DNA-repair process called homologous recombination.

Homologous recombination A DNA repair process in which a broken chromosome uses a homologous DNA sequence (generally located on a different chromosome) as a template for its repair.

Ploidy The number of copies of the genome or the number of chromosome sets in a cell. Haploid, one genome and

one set of chromosomes; diploid, two copies of the genome and two sets of chromosomes; and aneuploid cells have unbalanced sets of chromosomes, that is, at least one extra or one missing chromosome.

Sister-chromatid cohesion The connections that hold newly replicated chromosome pairs together until they are ready to be distributed into daughter cells.

Synaptonemal complexes Proteinaceous structures, resembling zippers or railroad tracks, which assemble between homologous chromosomes during meiosis. Synaptonemal complexes hold chromosomes together during the pachytene stage and promote the formation of chiasmata.

Meiosis is the specialized type of cell division by which gametes are produced. This process achieves the remarkable feat of halving the chromosome complement of a cell, typically from diploid to haploid. Meiosis thereby allows the diploid state to be restored when two gametes fuse during fertilization to form a zygote. Reduction of ploidy during meiosis occurs because two successive rounds of chromosome segregation follow a single round of DNA replication. The first round of chromosome segregation is unique in that the sister chromatids remain associated while parental homologs (pairs of sisters) are segregated. As prerequisites for their accurate segregation, homologs become closely paired along their lengths, by a structure called the synaptonemal complex (SC), and then undergo crossing-over – a reciprocal exchange of chromosome arms between two nonsister chromatids. Homolog pairing and crossing-over both involve homologous recombination, which occurs by the programmed formation and repair of DNA double-strand breaks (DSBs). Crossing-over is an essential feature of meiosis that underlies the fundamental laws of heredity: first, crossovers are essential for genetic transmission because they create connections between homologs that facilitate their bipolar orientation on the segregation apparatus (the spindle) and thereby facilitate orderly segregation. Second, crossovers (together with the independent assortment of parental chromosomes) produce diverse combinations of alleles on which genetic selection can act. Defects in meiotic chromosome segregation are linked to infertility, miscarriage, and genetic disease in humans.

The Sexual Life Cycle

In sexually reproducing organisms, successive generations maintain a constant number of chromosomes. This is because syngamy, the fusion of two gametes to form a zygote

(that subsequently gives rise to a new individual), is alternated with meiosis, which halves the cellular chromosome number. Without meiosis, syngamy would result in an unsustainable doubling of the chromosome content with each successive generation. Thus, the alternation of diploid (two copies of each chromosome) and haploid (one copy of each chromosome) generations of cells is a fundamental feature of most sexually reproducing organisms. In metazoans, syngamy is typically followed by extensive cell proliferation, by mitosis, and the development of a new individual. Meiosis in metazoans occurs exclusively in the germ-line cells of the gonads.

Mitotic and Meiotic Chromosome Cycles

Understanding how meiosis achieves the remarkable feat of halving the cellular chromosome number first requires an appreciation of how cells segregate chromosomes during mitotic cell division ([Figure 1](#)).

Mitosis

During mitosis, copies of both maternal and paternal chromosomes are distributed to two daughter cells. Replication first produces two identical copies of each chromosome called sister chromatids ([Figure 1](#), top row, panel ii.). Microtubules of the spindle apparatus attach to the chromatids via complex protein structures called kinetochores, which assemble at a unique location on each chromosome called a centromere ([Figure 1](#), top row, panel iii.). The chromatid pairs are then pulled apart to opposite poles of the cell, which then divides ([Figure 1](#), top row, panel iv.). The result is two identical daughter cells each containing two complete (maternal and paternal) sets of chromosomes. Two interdependent features ensure that chromosome segregation is an accurate process:

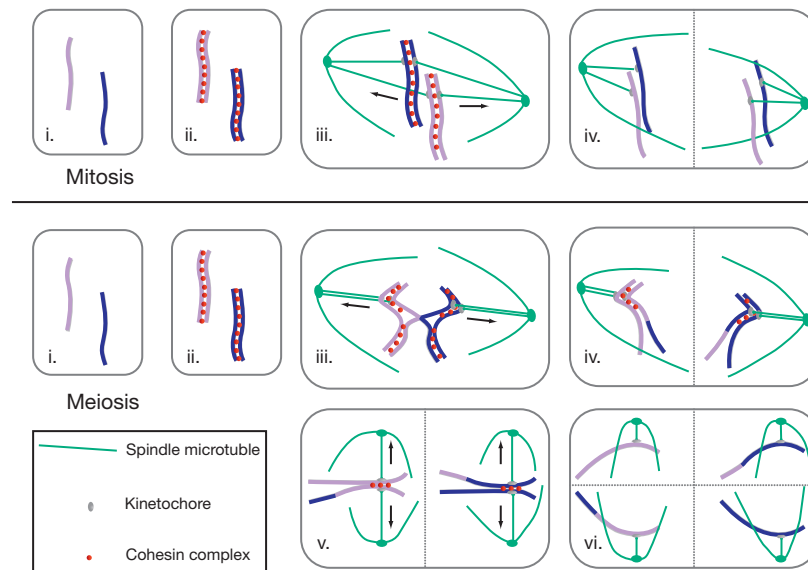


Figure 1 Comparison of Mitotic and Meiotic chromosome cycles (see text for details). Arrows indicate directions of the pulling forces generated by microtubules. Dashed lines indicate the planes of cell division.

(1) connections between sister chromatids and (2) a checkpoint mechanism that senses when each pair of chromatids has become attached to microtubules emanating from both poles of the spindle. Connections between sister chromatids, called sister-chromatid cohesion, form during replication and hold the sister chromatids together until they are ready to be segregated. When the two kinetochores of a sister-chromatid pair are attached to microtubules emanating from opposite poles of the cell, the pulling forces of the spindle are resisted by the cohesion between sister centromeres. The resulting tension stabilizes the microtubule-kinetochore attachments. When all chromatid pairs have achieved this biorientation on the spindle, sister-chromatid cohesion is destroyed and chromatids are pulled to the cell poles.

Meiosis

Meiosis, like mitosis, begins with replication to produce pairs of sister chromatids connected by cohesion (Figure 1, bottom row, i and ii). Then, in contrast to mitosis, a single copy of every chromosome must be accurately distributed to four different nuclei. Meiosis achieves this in the only logical way: via two successive rounds of nuclear division, first segregating homologs (the maternal and paternal chromatid pairs) and then segregating sister chromatids, as in mitosis. Regular homolog segregation is unique to meiosis, but utilizes the same mechanical principles as mitotic chromosome segregation. Thus, four modifications to the mitotic chromosome cycle are required:

1. Maternal and paternal homologs associate and become connected by structures known as chiasmata (Figure 1, bottom row, panel iii. and Figure 2). A chiasma results from crossing-over: the reciprocal breakage and joining of a maternal with a paternal chromatid. A chiasma holds the homologs together by virtue of the sister-chromatid

cohesion that was established as the chromosomes were replicated. Chiasma formation is the physical basis of genetic crossing-over.

2. The kinetochores of sister chromatids are modified such that they are unable to attach to microtubules emanating from different poles. Thus, the two sister kinetochores of a homolog behave like the kinetochore of a single sister chromatid in mitosis. Together with chiasmata, this monopolar property allows the homolog pair to biorient on the spindle. In this case, spindle forces are resisted by the combination of the connections between homologs and sisters (chiasmata and sister-chromatid cohesion; Figure 1, bottom row, panel iii.). The monopolar modification of sister kinetochores is reversed in preparation for the second meiotic division in order for sister chromatids to be segregated (Figure 1, bottom row, panel v).
3. The cohesion between the chromosome arms is destroyed prior to the first division, whereas cohesion between sister centromeres is protected until the second round of segregation ensues. This modification allows homologs to be segregated at the first meiotic division, while leaving sufficient cohesion to facilitate sister segregation at the second division (Figure 1, bottom row, panels iii–v).
4. The two meiotic divisions occur without an intervening round of replication by modulating the activities of cell-cycle kinases.

Meiotic Prophase I

The first meiotic division is preceded by a protracted G2 phase during which the chromosomes are paired and chiasmata are formed. This period is divided into five stages, defined by the appearance of the chromosomes under the light microscope (Figure 3):

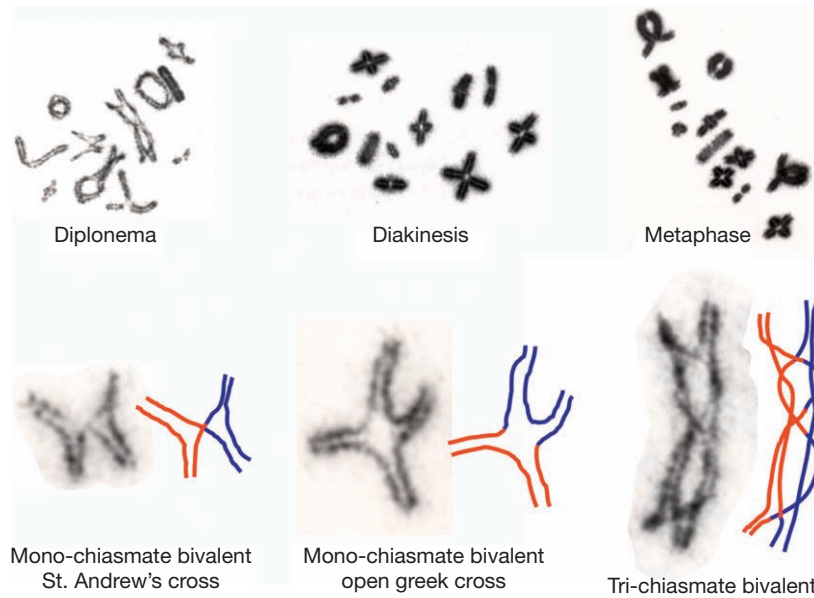


Figure 2 Chiasmata in the migratory locust, *Locusta migratoria*. (Top row) Chiasmata are observed during three distinct stages of meiotic prophase I, after which the homologs are segregated. (Bottom row) Enlargements of individual homolog pairs; cartoons show the topological relationships between chromosomes and the points of exchange. Images kindly provided by Professor Gareth Jones, Birmingham University.

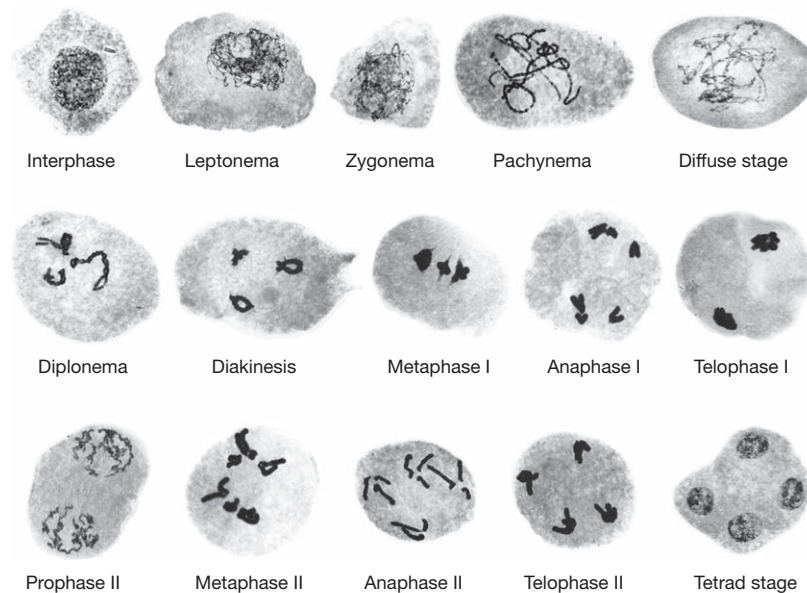


Figure 3 Classical stages of meiosis in *Crepis capillaris*, a flowering plant of the family *Compositae*. The cells of the tetrad eventually differentiate into pollen. Images kindly provided by Professor Gareth Jones, Birmingham University. Reproduced from Funcham JRS (1983) *Genetics*. Sudbury, MA: Jones & Bartlett, with permission.

1. *Leptonema* (adjective = leptotene; from the Greek *lepto*, meaning 'thin'). During leptonema, which follows chromosome replication, chromosomes are first visible as long thin strands.
2. *Zygonema* (adjective = zygotene; from the Greek *zygo*, meaning 'yoke', Old English for couple or unite). Homologs begin to closely associate during zygonema. The extent of association increases throughout this stage.
3. *Pachynema* (adjective = pachytene from the Greek *pachy*, meaning 'thick'). Homologs are associated along their entire lengths and thus appear as thick, morphologically indistinguishable entities.
4. *Diplonema* (adj = diplotene; from the Greek *diplo*, meaning 'double'). Diplonema is usually preceded by the diffuse stage during which chromosomes appear fuzzy and indistinct. When they reappear at diplonema, homologs are no

longer associated along their lengths, but are visibly connected by chiasmata (Figure 2). In female animals, meiosis typically arrests after diplonema and oocytes move into a quiescent dictyate stage in which the chromosomes are less condensed. In most vertebrates, the first meiotic division is only completed in response to hormonal stimulation during ovulation.

5. *Diakinesis* (from the Greek *dia*, meaning 'across' and *kinesis*, meaning 'movement'). During diakinesis, the nuclear membrane begins to break down and spindle microtubules develop and connect to the monopolar kinetochores of the homologs. The homolog pairs condense dramatically, becoming progressively shorter and thicker and they move toward the equator of the cell (Figures 2 and 3).

After Prophase I

Following the events of prophase I, the cell prepares the chromosomes for segregation and then segregates them via two additional stages, which are the meiotic counterparts of the final stages of mitosis (Figure 3). A second round of chromosome segregation and differentiation of the gametes then ensues:

1. *Metaphase I* (from Greek *meta* = between, with, beside, after + *phasis* = appearance). At metaphase I, the condensed chromosomes are aligned along the equator of the spindle, typically at the central plane of the cell. This configuration is the consequence of stable biorientation of the homolog pairs on the spindle (Figures 2 and 3).
2. *Anaphase I* (from Greek *ana* = up, throughout, according to + *phasis* = appearance). At anaphase I, cohesion between chromosome arms is destroyed and the homologs are pulled toward the poles of the spindle.
3. *Telophase I and Interphase I* (from Greek *telos* = end; Latin *inter* = between, among; + Greek *phasis* = appearance). Telophase is the last phase of mitosis, in which the segregated chromosomes decondense and are enclosed in a new nuclear membrane. Interphase is any period or stage between two successive mitotic divisions. These stages are not a general feature of meiosis and in many cases are skipped as the cell proceeds directly to the second meiotic division. In metazoan oocytes, cell division is asymmetric, resulting in the production of a small polar body plus a large oocyte cell.
4. *Meiosis II*. During prophase II, the chromosomes recondense and spindle microtubules develop. The sister-chromatid pairs then attach to the spindle in a bipolar fashion and congress to the spindle equator at metaphase II. The oocytes of vertebrates arrest for a second time at metaphase II, where they remain until fertilization. Typically, the plane of the meiosis II spindles is perpendicular to that of the meiosis I spindle. Sister chromatids are pulled apart at anaphase II and new nuclear envelopes form during telophase II. In oocytes, a second polar body is formed and only one haploid gamete is produced from the four sets of chromosomes that were present in the original precursor cell.
5. *Gametogenesis*. At this stage, the products of meiosis are known as meiocytes. The differentiation of meiocytes then

produces gametes that are characteristic of the specific organism, for example, sperm, eggs, pollen, and spores.

Meiotic Chromosome Structure and the SC

Loops and Axes

Meiotic chromosomes have a well-defined loop-axis structure (Figures 4(a)–4(c)), which is revealed by ultrastructural studies using the electron microscope. The loops are loops of chromatin that vary in size from ~20 kb in simple protists like yeast to ~2500 kb in grasshopper, an organism with particularly large chromosomes. Thus, the large differences in genome size between different organisms are accommodated in large part by changes in chromatin loop size, such that meiotic chromosome length does not vary linearly with genome size. The loops of every pair of chromatids are connected at their bases, in a linear arrangement, along a rod-like axis or core, that is, one axis = two chromatids = one homolog. Protein constituents of the homolog axes include those involved in sister-chromatid cohesion. The basic loop/axis structure is thought to be organized by the cohesion proteins; this then serves as a platform for the binding of additional proteins whose binding gives rise to the distinct axes detected by electron microscopy (Figures 4(c)–4(f)).

Synaptonemal Complexes

Meiotic chromosome pairing culminates with the formation of SCs, prominent proteinaceous structures that form between homologs, along their entire lengths. SC morphology, as defined by ultrastructural studies, is remarkably conserved between different organisms (Figure 4(f)): two dense lateral elements flank a central region, which contains a less dense central element. The lateral elements correspond to the homolog axes, described above. Transverse filaments lie across the central region to create a striated, zipper-like appearance. Central region proteins (transverse filaments) share a broadly similar secondary structure; a long coiled-coil flanked by globular C- and N-termini. Coiled-coil lengths of ~70–90 nm fit the prediction that these proteins span the 100-nm central region in an overlapping head-to-head configuration. SCs are thought to assemble by a two-step process: nucleation, by installing central region proteins at sites where homolog axes are very closely paired, followed by polymerization between and along the homolog axes. SCs are important for the normal formation of chiasmata and may also help minimize chromosome entanglement and nonhomologous recombination.

Prophase I Stages at the Ultrastructural Level

The homolog axes develop and become distinct during leptotema. The homologs progressively pair, eventually becoming coaligned along their lengths at a distance of ≥ 400 nm. During this presynaptic alignment, axes closely associate at several sites along each pair of homologs. These axial associations are the sites where DNA molecules are interacting via homologous recombination and are often associated with densely staining nodular structures that contain recombination proteins (see below and Figures 4(d) and 4(e)). During zygonema,

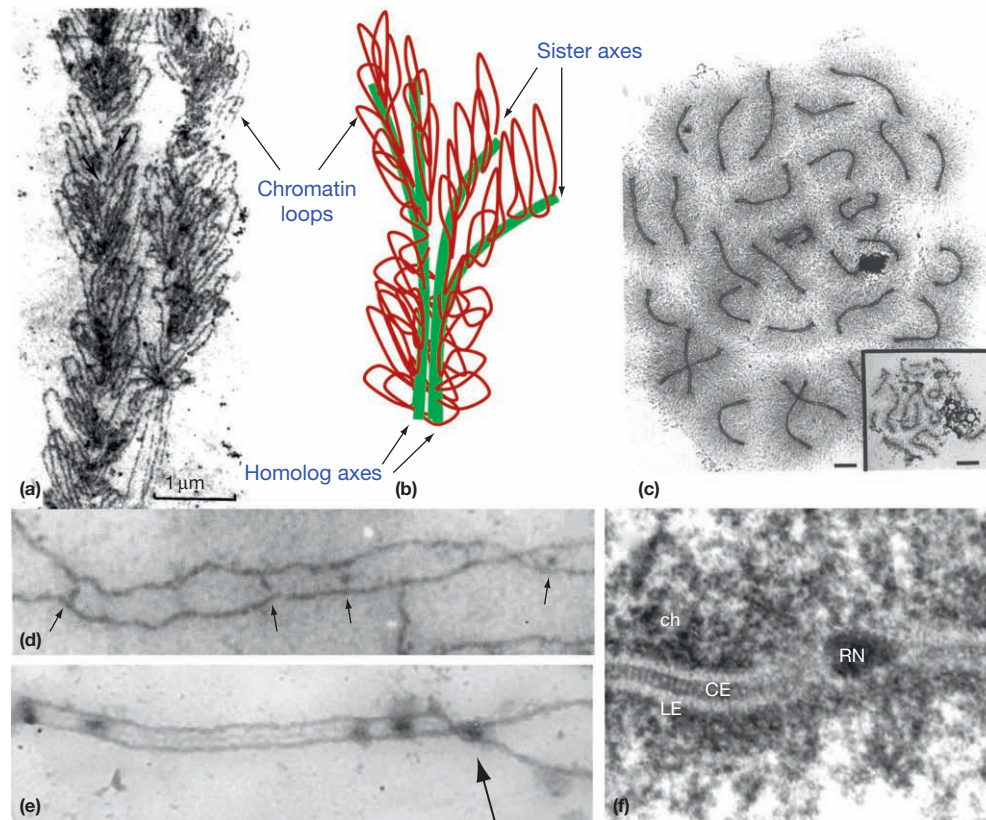


Figure 4 Chromosome structure in meiotic prophase. (a) Late pachytene stage chromosomes from the midge, *Chironomus*. (b) Cartoon rendering of loop-axis structure showing that each distinct homologue axis comprises two conjoined sister axes. (c) Spread chromosomes from the moth, *Hyalophora columbia*, and baker's yeast, *Saccharomyces cerevisiae* (inset; bars = 2 μ m); note the large difference in chromatin loop size. (d) Part of a spread chromosome pair in early zygonema from the onion, *Allium cepa*, showing early recombination nodules at the sites of axial association (indicated by arrows). (e) Part of a spread chromosome pair in mid-late zygonema from the tree tomato, *Cyphomandra betacea*, showing SC-associated early recombination nodules. Arrow highlights an early nodule at the junction of synapsed and unsynapsed axes. (f) SC longitudinal section from the beetle, *Blaps cribrosa*, with associated late recombination nodule. LE, lateral element; CE, central element; RN, recombination nodule; ch, chromatin. (a) Reproduced from Keyl HG (1975) Lampbrush chromosomes in spermatocytes of *Chironomus*. *Chromosoma* 51: 75–91, with permission from Springer-Verlag. (c) Reproduced from Moens PB and Pearlman RE (1988) Chromatin organization at meiosis. *BioEssays* 9: 151–153, with permission from John Wiley and Sons, Inc. (d) Reproduced from Albini SM and Jones GH (1987) Synaptonemal complex spreading in *Allium cepa* and *A. fistulosum*. I. The initiation of pairing. *Chromosoma* 95: 324–338, with permission from Springer-Verlag. (e) Reproduced from Anderson LK, Hooker KD, and Stack SM (2001) The distribution of early recombination nodules on zygotene bivalents from plants. *Genetics* 159: 1259–1269, with permission. (f) Reproduced from Shmekel K and Daneholt B (1988) Evidence for close contact between recombination nodules and the central element of the synaptonemal complex. *Chromosome Research* 6: 155–159, with permission from Springer-Verlag.

the polymerization of SCs is nucleated at a subset of these sites. Cells enter pachynema when SCs extend along the lengths of every pair of homologs and a second type of nodular structure develops which is typically larger, denser, and rounder, and much less numerous than those observed during zygonema (Figure 4(f)). These late recombination nodules mark the sites where chiasmata will appear following the breakdown of SCs and entry into diplonema.

Homologous Recombination during Meiosis

Homologous recombination is a DNA repair process in which a broken chromosome uses a second homologous chromosome as a template for its repair. The central reaction of recombination involves the pairing and strand exchange between a DNA strand, from an end of a broken chromosome, and a

homologous template duplex. The resulting joint molecule (JM) intermediate, called a displacement loop (D-loop), provides a primer-template substrate for the new DNA synthesis required to repair the damaged chromosome. Each recombination event has one of two potential outcomes: a crossover (the reciprocal exchange of chromosome arms) or a noncrossover (a local transfer of genetic information without exchange of chromosome arms). Homologous recombination typically plays two successive roles during meiosis: (1) to mediate the close pairing of homologs during leptonema/zygonema and (2) to connect homologs via chiasmata from diplonema through to anaphase I.

Mechanism and Regulation of Meiotic Recombination

Meiotic recombination is initiated by the programmed formation of numerous DNA DSBs during leptonema (Figure 5).

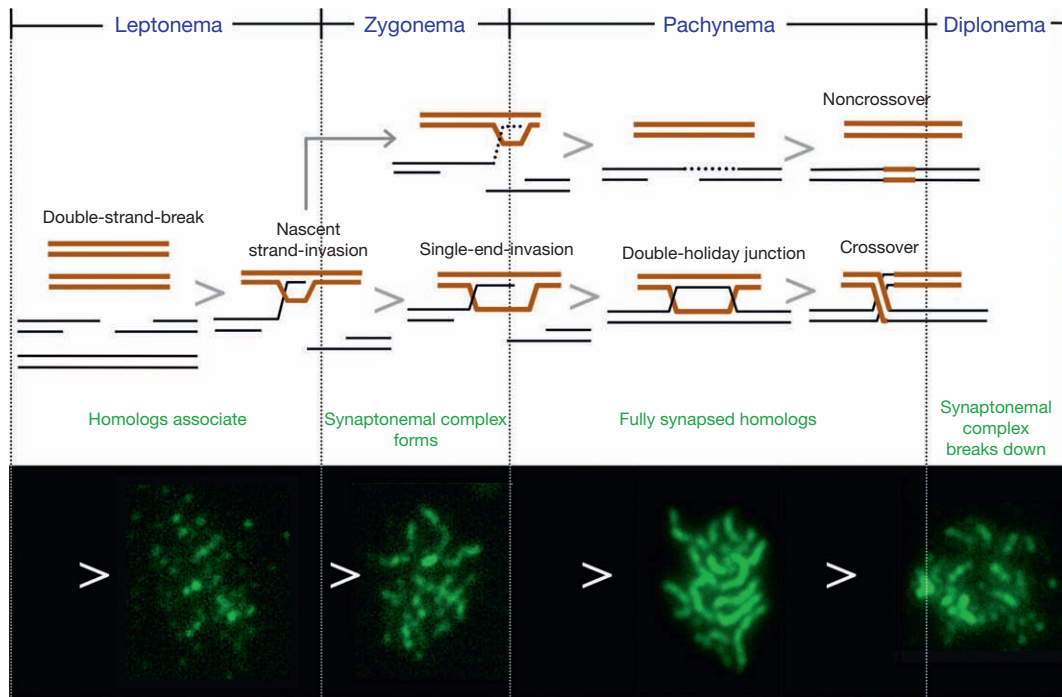


Figure 5 Relative time line of the DNA events of recombination and the cytological stages of meiosis. (Top row) The stages of meiotic prophase I. In mammals, this period occupies about 3 weeks; in a simple protist like yeast, it takes just 5 h. (Middle row) Relative timing and molecular details of meiotic recombination. Each pair of lines represents a DNA duplex. All four chromatids are shown at the DSB stage, but for clarity only the two chromatids involved in the recombination event are shown in subsequent stages (see text for details). (Bottom row) Synaptonemal complex formation in baker's yeast, *Saccharomyces cerevisiae*. Spread meiotic chromosomes were stained with fluorescent antibodies to a synaptonemal complex protein.

In yeast and similarly in mouse and human, it can be inferred that ≥ 200 DSBs are formed per meiotic cell. Such a large number of recombination events are required to bring homologs into close register and facilitate the formation of SCs. Only a subset of DSBs ($\sim 5\text{--}30\%$ depending on the organism), go on to produce crossovers. The ends of DSBs are processed to give long single-stranded tails, which load proteins related to the bacterial recombinase, RecA. These RecA homologs have the amazing ability to coat a single strand of DNA and then rapidly locate and pair with an identical DNA sequence anywhere in the genome. Meiotic chromosomes use this homologous pairing activity to recognize and pair with their homologous partner during leptonema.

A key aspect of meiotic recombination is its regulation. To ensure that each pair of homologs becomes connected by chiasmata, meiotic recombination must be regulated at multiple levels: (1) DSB formation, to ensure that recombination is initiated on all homologs; (2) template choice, to favor inter-homolog interactions (as opposed to inter-sister recombination); (3) the crossover or noncrossover outcome, to produce at least one crossover per homolog pair; and (4) spatial-temporal integration with the other events of meiotic prophase, that is, homolog pairing, synapsis, and segregation.

Our current understanding of these regulatory processes is shown in Figure 5. The crossover or noncrossover outcome is thought to be assigned at an early step in the recombination pathway, at the transition into zygonema. At this stage, DSBs

that have been designated to become crossovers exchange DNA strands with the homologous chromosome to form first a single-end invasion (SEI) and then a double-Holliday Junction (dHJ; named after Robin Holliday who proposed the existence of these strand-exchange structures during homologous recombination; Figure 5). dHJs form during pachynema and are resolved toward the end of this stage, by a nick and ligation mechanism, to give crossovers. DSBs that are not designated as crossovers are repaired by a process called synthesis-dependent strand-annealing (SDSA), which involves first copying DNA sequences from the homologous chromosome and then annealing the two DSB ends to seal the break. SDSA does not appear to involve stable, long-lived strand-exchange intermediates like SEIs and dHJs.

Recombination-Independent Mechanisms of Homolog Pairing and Segregation

In most organisms, recombination-independent mechanisms make important contributions to homolog pairing and synapsis. In fission yeast, microtubules attach to the chromosome ends and vigorously drag the chromosomes back and forth, bringing them into alignment. Actin-based chromosome movements facilitate pairing and synapsis in other organisms. Some species, such as nematode worms and fruit flies, have evolved special chromosomal sites that allow homologs to pair

independently of recombination. In fact, in fruit flies, homologs are already paired in somatic cells prior to meiosis. A completely recombination-independent meiotic process is apparent in male fruit flies and female moths, in which homologs are segregated without the benefit of chiasmata. In the case of female moths, SCs are maintained until anaphase I and thereby appear to negate the need for chiasmata.

Meiosis and Human Disease

Defects in meiosis can lead to programmed cell death, reduced numbers of gametes, and reduced fertility. More subtle defects in meiotic chromosome segregation result in aneuploid gametes that either lack or contain an extra copy of a chromosome. The monosomic or trisomic zygotes that result from aneuploid gametes are generally inviable due to gene imbalances that are incompatible with development. Well-known exceptions are Down syndrome (trisomic for chromosome 21), Turner syndrome (monosomic chromosome X), and Klinefelter's syndrome (males with two X chromosomes). In humans, ~30% of pregnancy miscarriages are the result of meiotic segregation errors that arise, primarily, in the female germline. The frequency of aneuploid oocytes (and the ensuing risk of miscarriage and aneuploid offspring) increases dramatically with advancing maternal age. The so-called 'maternal-age effect' is thought to result from the deterioration of oocytes that are arrested in the dictyate stage, rendering them less capable of efficiently segregating chromosomes when meiosis resumes at the time of ovulation.

See also: **Cell Architecture and Function:** Cell Cycle: Mitotic Checkpoint; Chromosome Organization and Structure, Overview; Mitosis; **Molecular Biology:** DNA Mismatch Repair and Homologous Recombination; Homologous Recombination in Meiosis; Recombination: DNA Helicases and Nucleases; Recombination: DNA-Strand Transferases.

Further Reading

- Bhalla N and Dernburg AF (2008) Prelude to a division. *Annual Review of Cell and Developmental Biology* 24: 397–424.
- Gerton JL and Hawley RS (2005) Homologous chromosome interactions in meiosis: Diversity amidst conservation. *Nature Reviews Genetics* 6: 477–487.
- Handel MA and Schimenti JC (2010) Genetics of mammalian meiosis: Regulation, dynamics and impact on fertility. *Nature Reviews Genetics* 11: 124–136.
- Hassold T, Hall H, and Hunt P (2007) The origin of human aneuploidy: Where we have been, where we are going. *Human Molecular Genetics* 16(Spec No. 2): R203–R208.
- Hunter N (2006) Meiotic recombination. In: Aguilera A and Rothstein R (eds.) *Molecular Genetics of Recombination*, pp. 381–442. Heidelberg: Springer.
- Jones GH (1984) The control of chiasma distribution. *Symposium of the Society for Experimental Biology* 38: 293–320.
- Marston AL and Amon A (2004) Meiosis: Cell-cycle controls shuffle and deal. *Nature Reviews Molecular Cell Biology* 5: 983–997.
- Page SL and Hawley RS (2003) Chromosome choreography: The meiotic ballet. *Science* 301: 785–789.
- Petronczki M, Siomos MF, and Nasmyth K (2003) Un menage a quatre: The molecular biology of chromosome segregation in meiosis. *Cell* 112: 423–440.
- Szekvolgyi L and Nicolas A (2010) From meiosis to postmeiotic events: Homologous recombination is obligatory but flexible. *FEBS Journal* 277: 571–589.
- Zickler D (2006) From early homologue recognition to synaptonemal complex formation. *Chromosoma* 115: 158–174.
- Zickler D and Kleckner N (1999) Meiotic chromosomes: Integrating structure and function. *Annual Review of Genetics* 33: 603–754.

Melanocortin System

J Pantel, Centre de Psychiatrie et Neurosciences, Paris, France

R D Cone, Vanderbilt University Medical Center, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Roger D. Cone, volume 2, pp. 617–620, © 2004, Elsevier Inc.

Glossary

Adrenocorticotropin ACTH, a 39-amino-acid peptide cleaved from proopiomelanocortin (POMC) in the anterior pituitary, that is secreted into the bloodstream and regulates the synthesis and release of the glucocorticoid class of steroid hormones.

Agouti proteins Along with agouti-related protein (AGRP), a family of small secreted proteins that bind to and antagonize melanocortin receptors.

Cachexia A condition associating anorexia, increased energy expenditure and lean body mass loss that accompanies chronic diseases.

Energy homeostasis The process by which the brain detects long-term energy reserves, stored in adipocytes in the form of fat, and regulates food intake and metabolic processes so as to keep reserves constant.

Melanocortin A peptide derived from the proopiomelanocortin gene containing a –His–Phe–Arg–Trp– motif within its sequence that binds to and activates one or more of a family of five melanocortin receptors.

Melanocyte A cell type populating skin and hair follicles that pigments hair and skin by secreting a variety of forms of melanin.

Tachyphylaxis A phenomenon characterizing a rapid decrease in drug response upon repeated doses.

The melanocortins are a family of peptides defined by their ability to modulate the activity of a family of five related G-protein-coupled receptors (GPCRs), called the ‘melanocortin receptors’ (MC1-R, also called the ‘melanocyte stimulating hormone receptor’; MC2-R, also called the ‘adrenocorticotropin receptor’; MC3-R; MC4-R; and MC5-R). The name is derived from the two original physiological activities identified for the family of peptides, stimulation of melanin synthesis and deposition, and stimulation of glucocorticoid synthesis and release by the adrenals. Four major melanocortin agonist peptides (α -melanocyte-stimulating hormone (MSH), β -MSH, γ -MSH, and adrenocorticotrophic hormone (ACTH)) are encoded by a single gene, the proopiomelanocortin (POMC) gene. Major sites of POMC gene expression include the pituitary, hypothalamus, brainstem, and skin. The melanocortin system is quite unique in that specific receptor antagonists, called ‘agouti’ and ‘agouti-related protein (AGRP)’, also exist. Together, these melanocortin agonists and antagonists regulate a remarkably diverse array of physiological processes, including the pigmentation of hair and skin, production of glucocorticoids, energy homeostasis, natriuresis, secretion of diverse exocrine gland products, and erectile function. A third agouti protein, AgRP2, has been discovered in teleosts, and appears to play a role in background color adaptation.

Melanocortin Ligands and Receptors

Elements of the mammalian melanocortin system are encoded by eight genes, the POMC gene, encoding the melanocortin peptide agonists, the five melanocortin receptor genes, and two genes, agouti and AGRP, that encode secreted melanocortin receptor antagonists. The melanocortin system appears to have developed with the evolution of vertebrates. While melanocortin-like

peptides can be found in invertebrates, the receptors appear first in vertebrate species. Most fish species contain all five melanocortin receptors. The lamprey genome, on the other hand, contains two melanocortin receptors, MCA, the likely precursor of the MC1 and MC2 receptor, and MCB, the likely precursor of the MC3, MC4, and MC5 receptors. The agouti and AgRP genes are not found in lampreys, and thus appear to be a later addition to the melanocortin system.

Peptide agonists are encoded within three different regions of the POMC preprohormone precursor. The POMC gene is expressed primarily in the anterior and intermediate lobes of the pituitary, the arcuate nucleus of the hypothalamus and the nucleus of the solitary tract of the brainstem, and in skin and a handful of peripheral sites where the peptides are thought to act in a paracrine or autocrine fashion. Melanocortin peptides are cleaved from three different regions of POMC to yield γ -MSH (amino terminal), ACTH and α -MSH (overlapping peptides from the central portion, and β -MSH (carboxy-terminal end) peptides. Specific patterns of cleavage occur in different tissues, so, for example, the 39-amino-acid peptide encoding ACTH is produced in the anterior pituitary by the peptidase PC1. A peptidase expressed in the intermediate lobe, PC2, makes a second cleavage to yield α -MSH, derived from the first 13 amino acids of ACTH. Melanocortin peptides derived from any of the three regions of POMC contain a His–Phe–Arg–Trp amino acid motif; this peptide motif is required for high affinity binding of these peptides to the melanocortin receptors.

In addition to peptide melanocortin agonists derived from POMC, two genes, agouti and AGRP, encode small secreted antagonists of the melanocortin receptors. Agouti is a 131-amino-acid protein with 10 disulfide-bonded cysteine residues related in structure to small peptide toxins like the conotoxins and plectotoxins, produced by cone snails and spiders,

respectively. Agouti is produced by hair follicle cells, where it acts on adjacent follicular melanocytes to antagonize the MC1-R and switch the cell from eumelanin (brown–black) to pheomelanin (yellow–red) pigment synthesis. AGRP, also 131 amino acids in length, shares sequence similarity with agouti in the cysteine-rich carboxyterminal domain, and is an endogenous antagonist of the MC3-R and MC4-R. AGRP is expressed in the adrenal gland and in the arcuate nucleus of the hypothalamus; in this latter site, AGRP is co-expressed with neuropeptide Y in a set of neurons adjacent but distinct from POMC arcuate neurons. In the brain, the AGRP gene is dramatically upregulated by fasting. The arcuate AgRP neurons are required for food intake, as specific ablation of these neurons results in cessation of food intake and, ultimately, death.

The melanocortin receptors, MC1-R through MC5-R, are encoded by a family of five independent genes; each receptor amino acid sequence is found in a single coding exon. The receptors are in the seven-membrane-spanning GPCR class, and all five receptors couple, via Gs, to activation of adenylyl cyclase. With the exception of the MC2-R, all of the melanocortin receptors bind multiple melanocortin peptide species. The MC2-R binds only the ACTH peptide, and the MC3-R is the only receptor that exhibits high affinity γ -MSH binding. Much of the physiological specificity of melanocortin peptide action derives from the unique tissue distribution of the five receptor subtypes (Table 1).

Pigmentation

The genetic study of coat colors led to the identification of over 60 genes involved in pigmentation. These genes encode proteins involved in all stages and aspects of the process by which melanocytes migrate from the neural crest, populate skin and hair follicles, and synthesize and secrete melanin or pigment granules. The biochemically complex melanin polymers can be divided into two classes, the yellow–red pheomelanins and brown–black eumelanins. The regulation of enzyme activities leading to these two pigment classes is termed the ‘eumelanin–pheomelanin switch’. Genetic studies of pigmentation in the mouse and a variety of other mammalian species led to the identification of two gene loci primarily involved in the regulation of the eumelanin/pheomelanin switch, *agouti* and

extension. The *extension* locus encodes the MSH receptor, or MC1-R, while *agouti* encodes an endogenous MC1-R antagonist expressed in skin and hair follicle. Mutations inactivating the MC1-R lead to yellow or red coat color, while mutations inactivating agouti lead to dark black coat colors. Loss-of-function mutations in the human MC1-R are responsible for approximately 80% of red hair color and type I skin pigmentation in humans. In many mammals in the wild, upregulation of agouti expression during a brief period of synthesis of a hair shaft leads to blockade of MC1-R activity, and results in deposition of a pheomelanin band along the otherwise darkly pigmented hair, known as the ‘wild-type agouti pigmentation pattern’ (Figure 1). In dogs, the *K* locus, encoding one of the endogenous anti-microbial β -defensin peptides, yields black coat color by blocking agouti action via a mechanism that has not yet been fully elucidated.

Adrenocortical Steroidogenesis

The ACTH receptor, or MC2-R, is expressed almost exclusively in the cortex of the adrenal glands, where it regulates synthesis and release of glucocorticoids in response to release of ACTH by the pituitary gland. ACTH also has long-term effects on the growth and differentiation of the cells of the adrenal cortex. ACTH is a 39-amino-acid melanocortin peptide cleaved from POMC in the anterior lobe of the pituitary. ACTH release, in turn, is regulated by production of corticotropin-releasing hormone, and hypothalamic releasing factor produced in response to a wide variety of stressors. These factors, in concert, constitute what is widely known as the ‘hypothalamic–pituitary–adrenal (HPA)’, or stress axis, which functions to allow organisms to adapt to internal and external challenges to homeostasis. Glucocorticoids serve many functions, but in general serve an adaptive role in coping with challenges to the organism by mobilizing carbohydrate energy stores and suppressing the immune system. Glucocorticoids are also essential for development. A rare syndrome of ACTH resistance resulting from loss-of-function mutations in the MC2-R, known as familial ACTH resistance or familial glucocorticoid deficiency (FGD), often presents with neonatal hypoglycemia, hepatitis, hyperpigmentation of skin due to elevated ACTH and, in some cases, severe septic events. Approximately 20% of FGD cases result

Table 1 The melanocortin receptors and their known functions

Receptor subtype	Site of action	Ligand	Biological function
MC1-R	Melanocyte	α -MSH = β -MSH = ACTH > γ -MSH > ACTH _{4–10} Antagonists – Agouti (IC ₅₀ ~ 1 nM)	Pigmentation
MC2-R	Adrenal cortex	ACTH	Steroidogenesis
MC3-R	CNS, stomach, duodenum, placenta, and pancreas	γ ₁ -MSH = γ ₂ -MSH = α -MSH = β -MSH = ACTH Antagonists – AGRP (~2 nM)	Energy homeostasis; natriuresis
MC4-R	Adult CNS, spinal cord, fetal brain, spinal cord, and autonomic nervous system	α -MSH (EC ₅₀ = 0.2–1.5 nM) = β -MSH = ACTH > ACTH _{4–10} = γ -MSH > ACTH _{4–10} Antagonists – AGRP (IC ₅₀ ~ 2 nM)	Energy homeostasis
MC5-R	Exocrine glands	α -MSH > β -MSH = ACTH > γ -MSH > ACTH _{4–10}	Synthesis and secretion of exocrine gland products

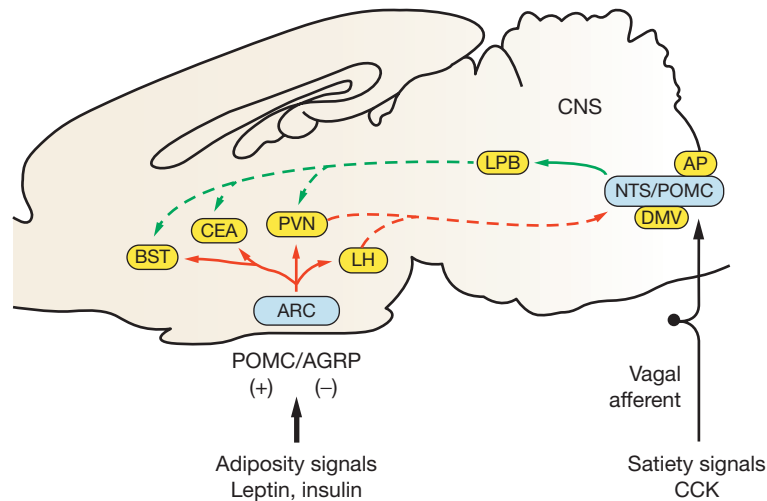


Figure 1 Integration of long-term adipostatic signals and acute satiety signals by the central melanocortin system. Blue: nuclei containing POMC neurons and AGRP neurons; yellow: nuclei containing MC4-R neurons that may serve to integrate adipostatic and satiety signals; red arrows: adipostatic signaling; green arrows: satiety signaling; BST, bed nucleus of the stria terminalis; CEA, central nucleus of the amygdala; PVN, paraventricular nucleus of the hypothalamus; LH, lateral hypothalamic area; LPB, lateral parabrachial nucleus; AP, area postrema; DMV, dorsal motor nucleus of the vagus. Reprinted with permission from Nature Neuroscience.

from mutations in melanocortin receptor accessory protein (MRAP), a transmembrane protein that is essential for both transport of the MC2-R to the plasma membrane and binding of ACTH by the receptor.

Energy Homeostasis and Other Central Nervous System Actions

The cloning of the MC1-R (MSH-R) and MC2-R (ACTH-R) genes led to the identification of three additional GPCRs that were highly related in primary amino acid sequence and also were high-affinity receptors for melanocortin peptides. Since these receptors had, at the time, unknown physiological roles, they were termed 'MC3-R', 'MC4-R', and 'MC5-R'. Remarkably, a novel mutation in the agouti gene that caused not only yellow coat color but also an obesity syndrome in the mouse ultimately led to the discovery that the MC3-R and MC4-R play a role in the central control of energy homeostasis. The lethal yellow allele of agouti results from a promoter rearrangement that causes ubiquitous expression of the agouti gene, in contrast to its normal expression limited to the hair follicle, where it normally serves to block eumelanin synthesis. Aberrant expression of agouti in the brain results in blockade of the brain MC4-R, resulting in the agouti obesity syndrome. Discovery of AGRP identified an endogenous hypothalamic antagonist of the central MC3-R and MC4-R that has a physiological role in the regulation of the central melanocortin system. Pharmacological blockade of the MC4-R, deletion of the gene, or deletion of the ligand-encoding gene, POMC, also reproduce the agouti obesity syndrome. In humans, haploinsufficiency of the MC4-R, that is loss-of-function or hypomorphic mutations in one allele, appear to be responsible for as much as 5% of cases of severe early onset obesity, making the MC4-R the most common genetic cause of severe obesity known. Less is known regarding the MC3-R; however, it is expressed in a majority of arcuate POMC neurons where it

appears to function as an inhibitory autoreceptor. Deletion of the MC3-R also causes a mild obesity syndrome associated with Cushing's syndrome-like symptoms.

The central melanocortin system is defined as the hypothalamic POMC and AGRP neurons that originate in the arcuate nucleus, brainstem POMC neurons originating in the nucleus of the solitary tract, and their target MC3-R- and MC4-R-expressing neurons, found throughout the central nervous system (CNS), in approximately 200 discrete locations (Figure 1). Central administration of melanocortin agonists both inhibits food intake and increases energy expenditure. These data, and the fact that blockade of MC4-R signaling causes obesity, argue that the POMC neurons tonically inhibit food intake and energy storage. The melanocortin system does this through effects on both food intake and on autonomic outflow to metabolically important tissues in the periphery. POMC and AGRP neurons express receptors for the adipostatic hormone leptin, and are regulated as well by a network of leptin-receptor-expressing GABAergic neurons, and play an important role in sensing and responding to the leptin signal from adipose tissue that informs the brain of long-term energy stores. POMC neurons also appear to be important for sensing and responding to acute satiety and hunger signals. The central melanocortin system affects autonomic outflow to a wide variety of tissues; melanocortins also have effects on glucose homeostasis, blood pressure, heart rate, natriuresis, inflammatory response, and the male erectile response. Therapeutic applications of melanocortins may include obesity, diabetes, cachexia, and erectile dysfunction.

Exocrine Gland Function and Other Peripheral Actions

MC5-R is expressed at high levels in a diverse array of exocrine glands, such as the sebaceous, lacrimal, preputial, and Harderian glands of the mouse. The receptor appears to stimulate the

production and release of exocrine gland products from these tissues. For example, a knockout mouse specifically lacking the MC5-R gene was found to produce reduced amounts of certain lipids normally secreted by sebaceous glands onto the coat of the animal, thus reducing the ability of the coat to repel water. Male MC5-R knockout mice also exhibit a defect in initiating aggressive behavior relative to their wild-type counterparts, which has been attributed to defects in the secretion of pheromones by glands like the preputial, known to express high levels of the receptor. Little is known regarding the actual physiological roles of the MC5-R in the various exocrine tissues. For example, it is not clear whether the source of ligand for the MC5-R is the predominant serum melanocortin, ACTH, or whether this receptor is regulated by yet uncharacterized paracrine sources of POMC. If the former is the case, the system would provide a mechanism for stress-induced stimulation of exocrine gland function. Suppression of sebaceous gland function in acne by MC5-R antagonism is a potential therapeutic utility of this receptor.

The MC3-R is also expressed in a handful of peripheral tissues, although little is known regarding its actions. The MC3-R and γ -MSH appear to be involved in the regulation of natriuresis; the reflex natriuresis that occurs after unilateral nephrectomy requires the MC3-R, and MC3-R knockout mice exhibit salt-sensitive hypertension.

Synthetic Ligands and Drug Discovery Efforts

The MC4-R has attracted the attention of the pharmaceutical industry, given possible applications of agonist drugs as potential obesity and sexual dysfunction treatments, and the application of antagonists for the treatment of cachexia, a disease condition commonly associated with chronic diseases such as cancer, AIDS, and cardiac or renal failure.

The first compounds to be discovered were modified peptides based on the -His-Phe-Arg-Trp- motif of the endogenous melanocortin ligands such as α -MSH. The first important agonist to be discovered, melanotan-II (MT-II), is a heptapeptide that was cyclized for increased half-life. This compound is around 100-fold more potent than α -MSH and is equally potent for MC3R and MC4R. This molecule was successfully shown to be erectogenic and to decrease food intake in both animals and humans. Based on a similar cyclic structure, SHU9119, a potent competitive antagonist was developed. This antagonist was shown to be able to restore food intake in cachexia models in the animal. Similar results could be obtained with the physiological inverse agonist AgRP.

Bremelanotide, another cyclic agonist, entered clinical trials to treat sexual dysfunction. In the recent years, the first non-peptidic small melanocortin molecules were discovered on the basis of structure-activity relationships focusing on the MSH pharmacophore or through large pharmacological drug screens. From these studies, tetrahydroisoquinoline (THIQ), a potent selective MC4R agonist, was the first nonpeptidic compound to be discovered. Overall, these nonpeptidic compounds present advantages regarding their pharmacokinetic properties as well as being orally available compared to their peptidic counterparts.

Therapeutic Potential

Antiobesity

The development of melanocortin agonists looked initially very promising as antiobesity drugs. However, the long-term administration of the agonist such as MT-II only induces a transient decrease in food intake that resolves within days. The exact explanation of this phenomenon referred to as 'tachyphylaxis' remains unknown. In this regard, the benefits of a long-term treatment based on melanocortin agonist drugs might potentially be affected. The very first clinical trials reported so far failed due to efficacy problems or side effects, such as an increase in blood pressure.

Sexual dysfunction

The implication of the melanocortin system in penile erection was first discovered with the agonist MT-II. These effects are believed to be centrally mediated. Bremelanotide is considered as a promising initiator of erection tested independently or together with the phosphodiesterase type 5 (PDE5) inhibitor sildenafil, and also exhibited unacceptable pressor activity.

Cachexia

Many pharmacological approaches based on the melanocortin antagonist SHU9119 or the endogenous inverse agonist AgRP successfully improved the phenotype of different acute (lipopolysaccharide injection) or chronic cachexigenic murine models (cancer induced or nephrectomy). The implication of the melanocortin system in the pathophysiology of cachexia was further confirmed by means of the genetic Mc4r knockout mice. These observations provide a strong rationale to develop small antagonist molecules selective for MC4R. So far, these molecules still remain at early preclinical stages.

See also: **Metabolism Vitamins and Hormones:** Anorexigenic and Orexigenic Gut Peptides; **Metabolic Roles of the Orexigenic and Anorexigenic Neuropeptides;** **Signaling:** Evolving Concepts of Leptin.

Further Reading

- Butler AA, Kesterson RA, Khong K, et al. (2000) A unique metabolic syndrome causes obesity in the melanocortin-3 receptor-deficient mouse. *Endocrinology* 141: 3518–3521.
- Chen W, Opitz-Arraya, and Cone RD (1997) Exocrine gland dysfunction in MC5-R-deficient mice: Evidence for coordinated regulation of exocrine gland function. *Cell* 91: 789–798.
- Cone RD (ed.) (2000) *The Melanocortin Receptors*. Totowa, NJ: Humana Press.
- Cone RD (ed.) (2003) The melanocortin system [Themed issue]. *Annals of New York Academy of Sciences*, 994.
- Cone RD (2005) Anatomy and regulation of the central melanocortin system. *Nature Neuroscience* 8: 571–578.
- Farooqi IS and O'Rahilly S (2008) Mutations in ligands and receptors of the leptin-melanocortin pathway that lead to obesity. *Nature Clinical Practice Endocrinology and Metabolism* 4: 569–577.
- MacNeil DJ, Howard AD, Guan X, et al. (2002) The role of melanocortins in body weight regulation: Opportunities for the treatment of obesity. *European Journal of Pharmacology* 440: 141–157.
- Rees JL (2000) The melanocortin 1 receptor (MC1R): More than just red hair. *Pigment Cell Research* 13: 135–140.
- Webb TR and Clark AJ (2010) Minireview: The melanocortin-2 receptor accessory proteins. *Molecular Endocrinology* 24: 475–484.
- Wikberg JE and Mutulis F (2008) Targeting melanocortin receptors: An approach to treat weight disorders and sexual dysfunction. *Nature Reviews Drug Discovery* 7: 307–323.

Membrane-Associated Energy Transduction in Bacteria and Archaea

G Schäfer, University of Lübeck, Lübeck, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

Archaea The prokaryotic domain consists of two branches of the phylogenetic tree of life, one occupied by the bacterial domain and the other by the archaeal domain. The latter splits into the euryarchaeota and the crenarchaeota. Both archaeal sub-branches are host to many extremophilic genera. The methanogens belong to the euryarchaeota, the thermoacidophiles with a few exceptions to the crenarchaeota.

Chemolithotrophs Prokaryotes assimilating carbon dioxide as sole carbon source and extracting energy from inorganic redoxreactions (in contrast to organotrophs that use reduced organic substrates as their carbon and/or nitrogen source).

Chromophore Usually, a colored organic molecule bound to a protein and functioning as a cofactor in catalysis. Hemes are chromophores of cytochromes and host Fe ions in the core of a porphyrin ring; other tetrapyrrol derivatives such as chlorophyll host Mg or Co. Chlorophylls and carotenoids in reaction centers or

antenna systems, and retinal in rhodopsins, are light-energy transducers.

Endosymbiosis Uptake of an originally independent free-living organism into cells of a host organism, where it can reproduce under the control of the host organism and share metabolic functions (the energy-transducing organelles of eukaryotes originated by permanent transfer of genetic material from the endosymbiont to the host genome).

Prokaryotes Unicellular microbes without a cell nucleus (in contrast to the eukaryotes that are located on a distinct branch of the universal phylogenetic tree of life, the eukaryotic domain).

Thermoacidophiles Prokaryotes with a lifestyle adapted to hyperthermophilic conditions (75–110 °C) as well as to extremely low environmental pH (3.5–0.5). Most of them belong to the archaea.

Vacuoles Membrane-encapsulated compartments within eukaryotic cells, in most cases, filled with liquid of low pH; their membrane contains an adenosine triphosphate (ATP)-driven proton pump, the V-type ATPase.

Archaea and bacteria represent the two prokaryotic branches of the universal tree of life. Their plasma membranes host protein complexes specialized in primary and secondary energy-transducing mechanisms. Common to bacteria and archaea is the underlying principle of primary energy conservation, the chemiosmotic mechanism generating an electrochemical potential of ions across the membrane on expense of the energy derived from membrane-associated fermentative, redox- or light-driven processes. In evolution, two of these, oxygen respiration and oxygenic photosynthesis, were transferred to eukaryotes by endosymbiotic uptake of putative ancient bacteria, the precursors of mitochondria and chloroplasts. Archaea differ from bacteria not only by genetic and taxonomic markers, membrane structure, specialized metabolic functions, and archaetypical coenzymes, but also by their extremophilic lifestyles regarding thermo-, pH-, and salinity-tolerance. Several energy-transducing processes are confined to the archaeal domain such as methanogenesis and rhodopsin-dependent photosynthesis. Chlorophyll-dependent photosynthesis does not exist in the archaeal domain. Secondary membrane-associated energy transduction such as the synthesis of adenosine triphosphate (ATP) from adenosine diphosphate (ADP) and various active transport systems are consumers of the primarily generated electrochemical ion gradient. The ions that couple primary and secondary energy transduction are predominantly hydrogen ions; in several bacteria and archaea sodium ions are also used.

Membranes

The matrix of polar lipids in plasma membranes serves as the environment for associated and membrane-spanning proteins. In bacteria, glycerol diester phospho- and glycolipids form a liquid-crystalline bilayer into which the energy-transducing membrane-protein complexes are inserted. The hydrophobic properties and fluidity of the core are determined by the fatty acid hydrocarbon chains; the permeability properties are additionally determined by the polar head groups. The lipid profile can serve as a taxonomic marker. Archaea, in contrast, contain unique lipids consisting of derivatives of diphytanylglycerol diethers and of di-biphytanyl-diglycerol-tetraethers. Consistently, besides bilayer structures, these membranes are composed of membrane-spanning lipid monolayers conferring extremely high stability, especially to the hyperthermophilic genera of the archaeal domain. The diether- and tetraether-derived lipids include phospholipids, glycolipids, and sulfoglycolipids. Generally, these lipids are glycosylated with the carbohydrate groups facing the plasma membrane outside. The unique lipid composition clearly delineates the archaea from all other organisms. Also, the cell envelopes of archaea are distinct from those of bacteria. Murein and lipopolysaccharide-containing outer membranes are not found in archaea. The quasi-crystalline surface layers of archaea are high-molecular-weight glycoproteins and have mainly form-stabilizing function and, in most cases, are the only envelope layer outside the plasma membrane.

Energy-Transducing Functions

Membrane-associated energy transduction proceeds by linking an exergonic process catalyzed by a membrane-residing enzyme (or protein complex) to the simultaneous vectorial translocation of charge (electrons or ions) across this membrane by the same enzyme. This principle is common to bacteria and archaea (Figure 1). Typical processes can be metabolic substrate conversions such as redox reactions or decarboxylation, respiratory electron transport from a low-potential donor to a high-potential acceptor (respiratory chains), or the decay from an energized state of a molecule generated by absorption of light to the low-energy ground state.

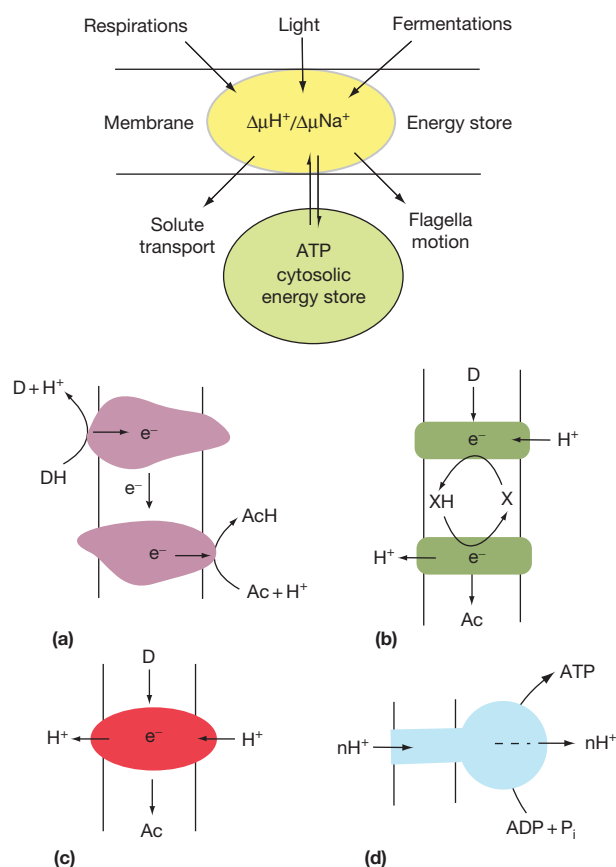


Figure 1 Schematic illustration of energy-conserving principles used by bacteria and archaea for the generation of an electrochemical potential of ions ($\Delta\mu\text{Na}^+$ or $\Delta\mu\text{H}^+$) across their plasma membrane. (a) Symbolizes $\Delta\mu\text{H}^+$ generation by scalar or so-called chemical protons consumed or liberated on opposite sides of the membrane; only electrons are transduced through the membrane. (b,c) Illustrate pumps: (b) functions via the intermediary quinol pool (x), (c) integrates pumping and electron transport within the same protein module like cytochrome-*c* oxidases; (d) symbolizes the universal ATP synthetase; its vectorial action and orientation is consistently opposite to that of the primary pumps.

Fermentation

Fermentations are anaerobic redox processes in which ATP is usually generated by substrate-level phosphorylation. In some special cases, partial reactions of fermentative pathways are catalyzed by membrane-residing enzymes, and the free-energy change of the reaction is coupled to the generation of an electrochemical-ion gradient. Examples are the biotin-dependent methylmalonyl-coenzyme A (CoA) decarboxylase of *Propionigenium modestum*, the oxaloacetate decarboxylase involved in citrate fermentation, or the glutamyl-CoA decarboxylase involved in glutamate fermentation – they generate a $\Delta\mu\text{Na}^+$. By means of H^+/Na^+ exchange transporters, the sodium gradient can be coupled to a proton gradient, or the $\Delta\mu\text{Na}^+$ can directly drive ATP synthesis by a Na^+ -translocating ATP synthetase. In methanogenic archaea, methyl-group transfer from an N atom to an SH group drives a sodium pump and also generates a $\Delta\mu\text{Na}^+$. The above examples demonstrate that in anaerobic fermentations also nonredox reactions can play an important role in membrane-associated energy transduction.

Anaerobic Respiration

Many organotrophic and chemolithotrophic bacteria and archaea can conserve energy by means of membrane-residing electron transport chains employing terminal acceptors other than oxygen. This group of obligately or facultatively anaerobic microbes uses electron acceptors such as nitrate, nitrite, fumarate, sulfur, ferric ion, carbon dioxide, sulfate, or even organic chlorocompounds for oxidation of organic molecules or hydrogen. Energy is conserved as a $\Delta\mu\text{H}^+$. Example reactions are given in Table 1.

The block on top gives examples for electron-accepting reactions in anaerobic respirations, the lower block exemplifies electron-donor reactions used by organotrophic and chemolithotrophic oxygen respiration. $[\text{H}]_2\text{-X}$ symbolizes an organic hydrogen-donating molecule, for example, reduced nicotinamide-adenine-dinucleotide (NADH) or succinate.

Denitrification

Denitrification is a typical example of anaerobic respiration and is found in over 40 genera; in archaea only in some

Table 1 Standard free energy changes of anaerobic and aerobic respiratory reactions

Chemical reaction	ΔG (kJ mol ⁻¹)
$\text{S}^0 + 2[\text{H}] \Rightarrow \text{H}_2\text{S}$	-33.5
$\text{SO}_4^{2-} + 8[\text{H}] + 2\text{H}^+ \Rightarrow \text{H}_2\text{S} + 4\text{H}_2\text{O}$	-151.7
$2\text{NO}_3^- + 5\text{H}_2 + 2\text{H}^+ \Rightarrow 6\text{H}_2\text{O} + \text{N}_2$	-959.8
$2\text{NO} + 2\text{H}^+ + 2\text{e}^- \Rightarrow \text{N}_2\text{O} + \text{H}_2\text{O}$	-305.9
$\text{CO} + \text{H}_2\text{O} \Rightarrow \text{CO}_2 + 2\text{H}^+ + 2\text{e}^-$	-492
$[\text{H}]_2 - \text{X} + 1/2\text{O}_2 \Rightarrow \text{H}_2\text{O} + \text{X}$	-219.0
$2\text{H}_2 + \text{O}_2 \Rightarrow 2\text{H}_2\text{O}$	-472.5
$2\text{S}^0 + 2\text{H}_2\text{O} + 3\text{O}_2 \Rightarrow 2\text{H}_2\text{SO}_4$	-1014.0
$2\text{FeS}_2 + \text{H}_2\text{O} + 7\text{O}_2 \Rightarrow 2\text{FeSO}_4 + 2\text{H}_2\text{SO}_4$	-1498.4
$2\text{CO} + \text{O}_2 \Rightarrow 2\text{CO}_2$	-504.9
$4\text{Fe}^{2+} + \text{O}_2 + 4\text{H}^+ \Rightarrow 4\text{Fe}^{3+} + 2\text{H}_2\text{O}$	-17.7

extreme halophiles. However, nitrate is reduced in four steps to $\text{N}_2 \rightarrow \text{NO}_2^- \rightarrow \text{NO} \rightarrow \text{N}_2\text{O} \rightarrow \text{N}_2$. The enzymes are integrated into the plasma membrane or located in the periplasm. Oxygen represses their synthesis. Nitrate reductase is a molybdopterin dinucleotide-containing enzyme associated with an iron-sulfur protein (four FeS clusters) and a membrane spanning *b*-type cytochrome. Energy from NO_3^- reduction is conserved by release of protons from the reductant, ubiquinone, or menaquinone, at the membrane outside, whereas protons for water formation are taken up from the cytosol. NO reductase displays analogy to heme/Cu oxidases but hosts Fe instead of Cu in the binuclear reaction center; in contrast to the latter, it does not pump protons. Denitrifying pathways are widespread in Gram-negative proteobacteria (e.g., *Paracoccus*, *Alcaligenes*, *Pseudomonas*, and *Thiobacillus*). Associated with denitrification is ammonification, which saves NH_4^+ by reduction of NO_2^- with NAD(P)H. Ammonification is a cytosolic process catalyzed by a siroheme/[4Fe4S]-containing enzyme.

Fumarate Respiration

In *Wolinella*, nitrate reduction can be achieved with molecular hydrogen or with formate as electron donors. It involves fumarate respiration (Figure 2), as an alternative mechanism to generate a proton-motive force. H^+ is released at the outer side of the membrane by oxidation of hydrogen or formate, whereas proton uptake and reduction of fumarate occur at the inner side. The dehydrogenases as electron donors and the fumarate reductase as acceptor are interacting via *b*-type cytochromes and the menaquinone pool in the membrane ($E'_0 = -100$ mV). The reduction of fumarate to succinate is catalyzed by an enzyme similar to nitrate reductase, but molybdopterin is replaced by flavine-adenine-dinucleotide (FAD). The enzyme is structurally and evolutionarily related closely to succinate dehydrogenases of aerobic organisms which catalyze the reverse reaction using ubiquinone

($E'_0 = +60$ mV), or in thermoacidophilic archaea *Caldariella* quinone ($E'_0 = +106$ mV), as primary electron acceptor.

Sulfur Respiration

Some bacteria and archaea perform sulfur respiration – *Wolinella succinogenes* is an example. Instead of insoluble elemental sulfur polysulfide (S-S_n^-) serves as electron acceptor. The electron transport chain involves three membrane-residing enzymes with their substrate-binding sites facing the periplasm: formate dehydrogenase or hydrogenase, and polysulfide reductase. In contrast to fumarate respiration, a quinone is not involved in electron transfer. The reduction of sulfur to H_2S is coupled to generation of $\Delta\mu\text{H}^+$ at a low H^+/e^- ratio by a hitherto unknown mechanism.

Ferric Ion Reduction

Members of the genera *Clostridium*, *Escherichia*, and *Bacillus*, among others, are able to reduce ferric to ferrous ions. It has not been shown unequivocally that the reaction is coupled to energy conservation. Although the couple $\text{Fe}^{3+/2+}$ has a high redox potential ($E'_0 = +770$ mV) similar to oxygen, organisms using Fe^{3+} as terminal electron acceptor have to cope with its extreme insolubility, except in a strongly acidic environment.

Oxygen Respiration

Aerobic Organotrophs

Oxygen is a superior electron acceptor for aerobic microbes that use reduced substrates or even molecular hydrogen as donors of reducing equivalents. Organotrophs deliver electrons into a respiratory chain (Figure 3) via ubi- or menaquinone-reducing membrane-anchored or integral membrane protein complexes. A major complex is the

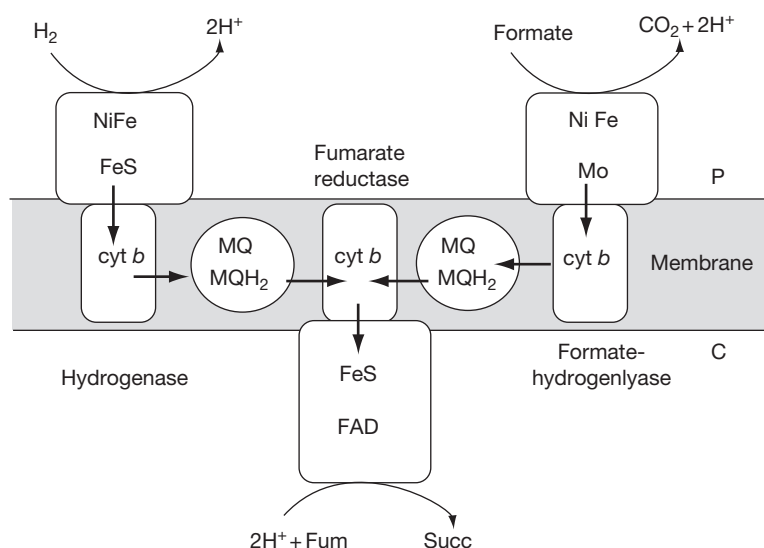


Figure 2 Scheme of an anaerobic respiration. Fumarate respiration is taken as an example with two possible electron-donating systems. The type of energy conservation corresponds to principle A in Figure 1. MQ, menaquinone; P, periplasm; C, cytosol.

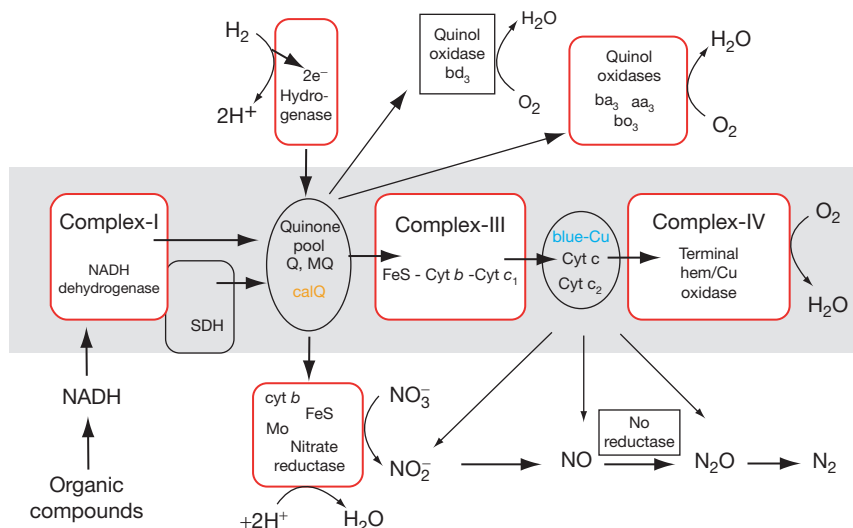


Figure 3 Illustration of the diversity of energy-transducing membrane systems in bacteria and archaea. All boxed or encircled components are membrane-integral or membrane-associated. All red boxes are proton pumps acting according to principles (b) and (c) of Figure 1. The scheme depicts the situation in *P. denitrificans*. The shaded block from left to right symbolizes a regular and complete aerobic respiratory chain as present, for example, in purple bacteria and also in eukaryotic mitochondria. The linear arrows indicate the direction of electron flux between donor and acceptor complexes or compounds. The symbols aa₃, ba₃, and bo₃ in the box quinoloxidases indicate the species-dependent variability of observed heme-compositions; calQ (caldariella quinone) applies to some archaea, as does the blue-Cu protein instead of c-type cytochromes. SDH, succinate dehydrogenase, complex II.

FMN-dependent type-I NADH dehydrogenase (complex I), composed of 14 different polypeptides (e.g., *Paracoccus* and *Escherichia coli*). It couples the translocation of (probably four) protons to the reduction of quinone and consists of a peripheral substrate oxidizing subcomplex facing the cytosol, a membrane integral transducer subcomplex with several FeS centers, and a receiver subcomplex transferring electrons to the quinone. Alternatively, in many bacteria and in archaea, type-II NADH dehydrogenases are found; these have less complex composition and cannot couple ion translocation to substrate oxidation. Several alkaline-adapted bacteria couple the translocation of Na⁺ instead of protons to NADH oxidation. As another quinone-reductase succinate dehydrogenase (complex-II) is found in the plasma membranes of all aerobic microbes. The small change in free energy of this reaction is insufficient to drive energy-conserving ion translocation. The reoxidation of reduced quinones is either catalyzed directly by heme/Cu-type quinol oxidases as, for example, in *E. coli*, *Paracoccus denitrificans*, and *Rhodobacter sphaeroides*, or by a second energy-conserving module of respiratory chains, the bc₁ complex (complex-III). The essential redox catalysts of this module are an [2Fe₂S] iron-sulfur protein (Rieske protein), a diheme b-type cytochrome, and a c-type cytochrome c₁. Complex-III couples the stepwise reoxidation of QH₂ to the transport of 2H⁺ across the membrane for each electron transduced to cytochrome c₁. The complex reaction mechanism is known as Q-cycle. The mobile redox carrier cytochrome c associated with the outer surface of the plasma membrane accepts electrons from cytochrome c₁ and serves as the connector to the terminal oxidase module, the heme/Cu-type cytochrome c oxidase (⇒). All oxygen-reducing heme/Cu oxidases are proton pumps. The path of electrons from NADH to oxygen can thus conserve the energy equivalent to 12H⁺ pumped across the membrane

per [H₂]. Depending on the genetic disposition, either one of the energy-transducing complexes can be missing, except a terminal oxidase. The modular structure of microbial respiratory chains allows a variety of special adaptations. Besides an aerobic respiratory chain, all modules for denitrification and even several terminal oxidases can be present in parallel (e.g., Figure 3); their individual contribution to energy transduction is usually regulated via the expression of the respective genes in response to nutritional and environmental conditions.

Aerobic Chemolithotrophs

A large variety of bacteria, the chemolithotrophs, can derive energy from oxidation of inorganic electron donors such as hydrogen, carbon monoxide, sulfur and nitrogen compounds, or divalent cations (e.g., Fe²⁺ and Mn²⁺). Many of these use molecular oxygen as oxidant. The mechanisms of energy transduction in chemolithotrophs are essentially the same as in organotrophs, that is, the electrons are channeled into a cytochrome chain and flow down to the terminal oxidase enabling a number of proton-translocating redox loops. Examples are given in Table 1. The acidophile *Thiobacillus ferrooxidans* uses a periplasmic Fe²⁺ oxidase transferring electrons via cytochrome c and the Cu-protein rusticyanin to a heme/Cu-terminal oxidase, which reduces oxygen on the plasma membrane inside. The reaction generates a proton-motive force by scalar consumption of H⁺ on the inside in addition to proton pumping by the aa₃-type terminal oxidase. So-called Knallgas bacteria comprising a large group of Gram-negative and Gram-positive bacteria, which can oxidize molecular hydrogen. Essential for the mechanism is a Ni-Fe hydrogenase dimer (e.g., in *Alcaligenes eutrophus*) in the cytoplasmic

membrane. It transfers electrons from hydrogen via the bimetallic reaction center, several Fe–S clusters, and a *b*-type cytochrome to the respiratory electron transport chain. By this process, scalar protons are produced outside in addition to protons pumped by the respiratory chain. Several Knallgas bacteria contain a second, cytoplasmic hydrogenase which is used for the reduction of NAD^+ via FMN by an enzyme with homology to NADH:ubiquinone oxidoreductase (complex-I).

Phototrophic Bacteria

The utilization of light by cyanobacteria, phototrophic purple, and green bacteria follows the same principles as in algae and higher plants. Also the light-absorbing pigments, other cofactors, and their hosting membrane proteins are essentially similar. Both, oxygenic and anoxygenic photosynthesis, are found; the latter is restricted to purple, green, and heliobacteria.

Pigments of Energy Transducers

Absorption of light energy and transduction of energized states is mediated by antenna complexes. These contain the cyclic tetrapyrrol derivatives chlorophyll and bacteriochlorophyll, carotenoids, and in cyanobacteria open-chain tetrapyrrols (phycobillins). The conjugated polyene compounds absorb light from the near-ultraviolet (UV) to the near-infrared (IR) and are noncovalently bound to membrane-integral and membrane-associated proteins in arrays. They transduce the excited state to the reaction centers (RCs) proper. The ratio is about 25–30 mol antenna chlorophyll per mol RC. In purple bacteria and heliobacteria (e.g., *Rhodobacter*, *Rhodospseudomonas*, *Chromatium*, and *Heliobacillus*), the antenna complexes are integrated within the cytoplasmic membrane. In green bacteria (e.g., *Chlorobium* and *Chloroflexus*), the antenna complexes are organized as chlorosomes, that is, vesicles attached to the inner side of the plasma membrane.

Anoxygenic Photosynthesis

Anoxygenic photosynthesis produces an electrochemical proton gradient across the membrane which drives ATP synthesis and other energy-consuming processes. Excitation energy is transferred from antennas to the RC within ~ 50 ps. In the RC, the energy is utilized to generate a charge separation ($\Delta\psi$) across the membrane. It occurs by electron transfer between the primary donor, a special pair of chlorophylls, localized toward the cytosolic side of the membrane and secondary acceptors at the opposite side. These can be quinones (type-II RC) or Fe–S clusters (type-I RC). From there the electrons flow back to cytochrome c_2 and the special pair via a quinone pool and through the bc_1 complex, as it is also used for respiratory energy conservation; a $\Delta\mu\text{H}^+$ is produced. Thus, anoxygenic photosynthesis can occur as a cyclic process. Production of reducing equivalents is not coupled to anoxygenic photosynthesis. In facultatively phototrophic/chemotrophic bacteria (*Rhodobacter* and *Rhodospirillum*), photosynthesis and respiration are competing for the bc_1 complex depending on the environmental conditions (oxygen tension and light intensity).

Oxygenic Photosynthesis

Cyanobacteria (*Synechococcus*, *Synechocystis*, and *Anacystis*) perform oxygenic photosynthesis. It employs two photosystems in sequence in its membrane (PS-I, PS-II). PS-I is essentially equivalent to the RC of anoxygenic phototrophic bacteria but channels the electrons to ferredoxin as terminal acceptor. Ferredoxin serves as reductant for NADP^+ , thus generating reducing equivalents for the cell. The reaction center PS-II is an H_2O :plastoquinone oxidoreductase and couples the reduction of the primary e^- -donor, the special pair of chlorophylls, to the oxidation of water by a Mn-enzyme associated with PS-II at the periplasmic surface. The water-splitting Mn-enzyme liberates molecular oxygen and H^+ . The reducing equivalents are passed on to PS-I via the plastoquinone quinone pool, the b_6f -complex, and the Cu-protein plastocyanin. The b_6f -complex is functionally and structurally analogous to the bc_1 complex. It is composed of a diheme cytochrome *b*, a Rieske Fe–S protein, and the mono-heme cytochrome *f*. b_6f acts as a proton pump performing a Q-cycle ($\Rightarrow$). The photosynthetic apparatus of cyanobacteria is localized in thylacoids; the antenna systems are organized as phycobilisomes attached to the thylacoid membranes. These structures largely resemble the structure of eukaryotic organelles. Also, respiratory enzymes (NADH dehydrogenase and cytochrome *c* oxidase) are localized in the thylacoid and partially also in the plasma membrane of cyanobacteria.

Special Mechanisms in Archaeal Energy Transduction

Methanogenesis from one- or two-carbon atom compounds and light-driven energy conservation by bacteriorhodopsins (BR) are mechanisms restricted to archaea. Methanogens comprise the genera *Methanosarcina*, *Methanobolus*, *Methanobacterium*, *Methanotrix*, *Methanothermus*, *Methanococcus*, and others among about 50 known species. Examples for the best-investigated genera of extreme halophilic archaea are *Halobacterium*, *Haloferax*, *Haloarcula*, *Halococcus*, and *Natronomonas*.

Methanogenesis

Methanogenesis is a reductive process under strictly anaerobic conditions. It involves reduction of C_1 compounds (CO_2 , CO , CH_3OH , and CH_3NH_2) mainly by H_2 , and also the disproportionation of acetate, formate, or methanol. Many of the methanogenic archaea are extremely thermophilic (growth at $\approx 100^\circ\text{C}$). For example, the change in free energy of the reaction $\text{CO}_2 + 4\text{H}_2 \Rightarrow \text{CH}_4 + 2\text{H}_2\text{O}$ is $\Delta G = -131 \text{ kJ mol}^{-1}$ and is comparable to CH_4 formation from other sources; that of acetate disproportionation is only -36 kJ mol^{-1} . Both reactions can drive chemiosmotic ATP synthesis by linking membrane-associated partial reactions to generation of ion gradients ($\Delta\mu\text{H}^+$ and $\Delta\mu\text{Na}^+$). The reactions involve unusual coenzymes, not found in bacteria or eukaryotic organisms. These are the fluorescent 5-deazaflavin F_{420} , methanofuran, methanopterin (H_4MPT), coenzyme-M (mercaptoethanesulfonate), a Ni-tetrapyrrole F_{430} , 5-hydroxybenzimidazolyl-cobamide, and methanophenazine as a substitute for membrane-residing quinones. For details of pathways the reader is referred to

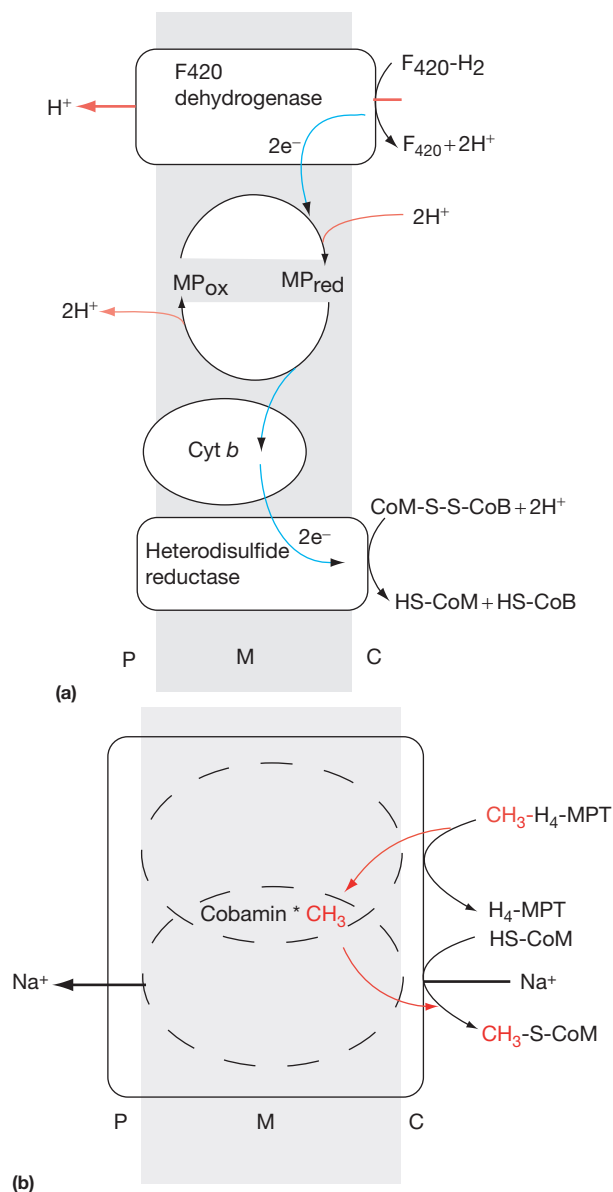


Figure 4 (a) Illustration of chemiosmotic energy conservation in methanogenic archaea. The example combines mechanisms (b) and (c) of Figure 1. P, periplasmic space; M, membrane; C, cytosol; MP, methanophenazine; CoM, coenzyme M. (b) Energy transduction by a non-edox reaction in archaeal methanogenesis. The sodium pump is driven by the energy change of the transmethylation reaction. H_4 -MPT, tetrahydro-methanopterin. The mechanism is thought to involve conformational coupling between the group-transferring and the ion-transferring modules; similar mechanisms apply to the sodium pumps driven by decarboxylation reactions.

specialized publications. Important for membrane-associated energy transduction are electron transport chains involved in methanogenesis, as illustrated in Figure 4. The ion-translocating redox chains involve hetero-disulfide reductase, two *b*-type cytochromes, methanophenazine, and Ni-hydrogenase or another donor of reducing equivalents. Redox cycling of methanophenazine with asymmetric uptake and release of protons can act as a proton pump with $H^+/e^- = 1$. Hydrogen can

also be provided from ferredoxin:methanophenazine oxidoreductase, or from F_{420} dehydrogenase, depending on the individual organisms and the C_1 carbon source used for methanogenesis. Whether the latter functions as proton pump is under debate; the observed total H^+/e^- ratios of 3–4, are higher than calculated for the methanophenazine cycle alone. Hydrogenases and F_{420} dehydrogenase display structural similarities to proton-translocating respiratory NADH dehydrogenases and are bearing subunits with several $[Fe-S]$ clusters. Also, heterodisulfide reductase contains two $[4Fe-4S]$ clusters. Apart from proton-motive electron transport chains, all methanogens have a primary Na^+ pump, the membrane integral methyl- H_4 MPT:CoM-methyltransferase. This enzyme contains the above-mentioned cobamide cofactor and a $[4Fe-4S]$ cluster. The reaction proceeds in two steps: first, the transfer of the methyl group to the corrin cofactor, second, methyltransfer to the cytosolic cofactor HS-Coenzyme M. Sodium pumping is associated with the second partial reaction producing a Na^+ gradient equivalent to a -60 mV membrane potential. The sodium pump consists of six to eight polypeptides (species dependent). The central pathway of methanogenesis is reversible. Accordingly, the enzyme acts as a sodium pump during methanogenesis from CO_2 or acetate but drives the reverse reaction by means of a sodium potential in methanogenesis from methyl group containing C_1 compounds.

Utilization of Light

Photosynthetic apparatuses as found in phototrophic bacteria and eukaryotes do not exist in archaea. However, the aerobic extreme halophiles like *H. salinarum* can synthesize a light-driven proton pump under conditions of low-oxygen tension composed of a single integral-membrane protein BR. BR spans the membrane with seven α -helical domains enclosing the chromophore retinal, which is bound as protonated Schiff's base to a lysine in the hydrophobic core of the protein. Absorption of light induces an all-*trans* to 13-*cis* isomerization of retinal accompanied by a deprotonation of the Schiff's base with concomitant release of the proton on the periplasmic membrane surface via a channel delineated by a series of protonable amino acid residues and water molecules. The light-triggered reaction is accompanied by conformational changes of the protein and initiates a photocycle of the chromophore that relaxes through five to six consecutive steps to the all-*trans* ground state. In that phase reprotonation of the Schiff's base from the cytosolic side of the plasma membrane is attained. The proton pumping power stroke corresponds to about -50 kJ mol $^{-1}$ and can generate a proton-motive potential of about 300 mV. The light-induced proton gradient $\Delta\mu H^+$ can drive ATP synthesis, the flagellar motor, or a Na^+/H^+ antiporter to drive sodium out of the cell. Besides BR, three other archaeal rhodopsins were found. Halorhodopsin (HR) has a similar structure but acts as an inwardly directed chloride pump. The two other so-called sensory rhodopsins (SR 1 + 2) are photoreceptors that trigger the flagellar motor (phototactic/photophobic) via conformational coupling of membrane-associated transducer units (Figure 5). As such they are basic models for the large variety of seven-helix membrane receptors. All four archaeal rhodopsins have striking similarity regarding

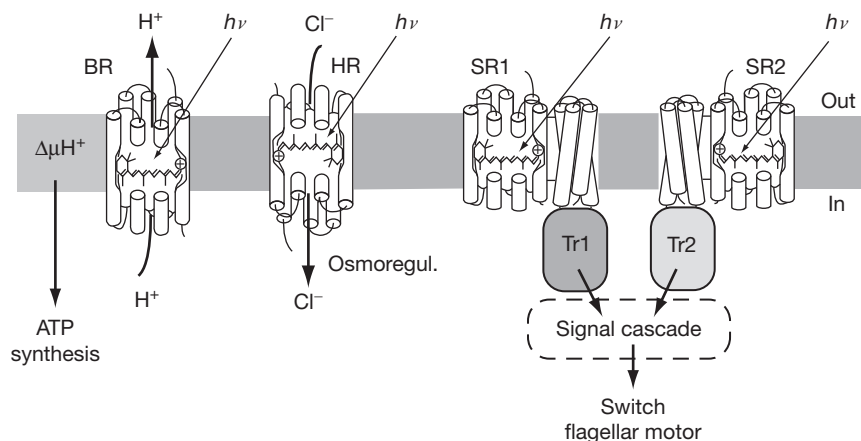


Figure 5 Illustration of energy transduction and sensory functions of rhodopsins in the extreme halophilic archaea. BR, bacteriorhodopsin; HR, halorhodopsin; SR1 and SR2, sensory rhodopsins. Only BR and HR are ion pumps. The homodimeric transducer complexes Tr1, 2 are conformationally coupled to SR1, SR2, and transfer the conformational signal via their cytoplasmatic domains to a signaling cascade controlling the rotational direction of the flagella motor. The orientation of vectorial reactions is indicated with 'in' symbolizing the cytosolic side of the plasma membrane.

protein structure, the retinal chromophore and its photocycle. However, the ion-transducing BR and HR perform fast photocycles in the millisecond range, whereas SR 1 + 2 are comparatively slow (~100 ms).

Respiratory Complexes

The euryarchaeota *Halobacteria*, *Thermoplasma*, and many species of the crenarchaeota comprising genera as *Sulfolobus*, *Acidianus*, *Metallosphaera*, or *Pyrobaculum* are obligate or facultative aerobes. Their respiratory systems essentially resemble modular components of respiratory chains as found in oxygen-respiring bacteria. A significant difference is the lack of a proton translocating NADH:quinol oxidoreductase. Instead, type-II NADH dehydrogenases were found, whereas complex-II analogous succinate dehydrogenases are present in all aerobic archaea. Two groups of enzymes can be distinguished: one group resembles the properties of SDHs from bacteria and mitochondria, and the other represents a novel class with unusual iron-sulfur clusters, as well as additional ones in a subunit with homology to methanobacterial heterodisulfide reductase, suggesting a novel electron pathway to the quinone pool. In *Halobacteria*, menaquinone and ubiquinone function as membrane-integral electron acceptors; in *Thermoplasma* and thermoacidophilic crenarchaeota like *Sulfolobales*, these are replaced by caldariella quinone and a variety of similar sulfur-containing thiopheno-benzoquinones. Several archaea such as *A. ambivalens* contain only fragmentary respiratory chains established from NADH- and succinate-quinone reductases and a heme/Cu-type quinol oxidase as terminal electron acceptor; the latter serves as the only energy-conserving proton pump. Rieske Fe-S proteins are present in *Halobacteria*, *Sulfolobales*, and *Pyrobaculum* for example, but cytochrome *c* and regular quinol:cytochrome *c* reductases are absent in many species. Instead, analogous functions are replaced by alternate membrane protein assemblies, for example, the SoxLN complex of *S. acidocaldarius*, using different electron acceptors as, for example, a mono-heme *b*-type cytochrome and/or blue copper proteins like sulfocyanine from various *Sulfolobus* species, or

halocyanines from *Natronomonas* or *Halobacterium salinarum*. The proton pumping terminal heme/Cu-type oxidases are organized as supercomplexes in some thermoacidophilic crenarchaeota that combine features of quinol- and cytochrome *c* oxidases. The best-investigated examples are the SoxM complex and the SoxABCD complex from *Sulfolobus*.

Secondary Energy Transducers

ATP synthetases ($\Rightarrow$) are the most important universal secondary energy transducers integrated into plasma membranes. Bacterial ATP synthetases and those of mitochondria and chloroplasts are of the F_0F_1 -type ($\Rightarrow$). Functionally, the ATP synthetases are reversible ion-transporting ATPases driving the reaction $ADP + P_i \rightarrow ATP + H_2O$ by dissipation of a $\Delta\mu H^+$ or $\Delta\mu Na^+$ across the membrane. F_1 extends to the cytosol and bears the substrate-binding sites; F_0 is the membrane integral ion-conducting subcomplex. Archaea contain A_0A_1 -type ATP synthetases that structurally resemble the eukaryotic vacuolar V-type ATPases ($\Rightarrow$) but, in contrast to the latter, function as ATP synthetases driven by an ion-motive $\Delta\mu H^+$ or $\Delta\mu Na^+$ according to the rotational reaction mechanism of F_0F_1 . The peripheral substrate-binding A_1 moiety has been identified from many *Halobacteria*, *Sulfolobales*, and *Methanobacteria*. The only integrated A_0A_1 -complex was isolated from the methanogen *Methanococcus jannashii*.

Evolutionary Links

In the evolution of higher organisms, cyanobacteria are the most likely precursor of chloroplasts which originated from an endosymbiotic uptake of such a prokaryotic phototroph into a precursor form of a eukaryotic cell. Likewise, the mitochondria have been derived from aerobic prokaryotes similar to the recent purple bacteria by endosymbiosis. The recipient might have been an archaeobacterium.

See also: **Bioenergetics:** Bioenergetics: General Definition of Principles; Chemolithotrophy; Cytochrome *bc₁* Complex (Respiratory Chain Complex III); Cytochrome Oxidases, Bacterial; Energy Transduction in Anaerobic Bacteria; F-ATPases; Green Bacteria: Chlorophyll Biosynthesis, Light-Harvesting, Reaction Centers, and Electron Transport; Purple Bacteria: Photosynthetic Reaction Centers; Respiratory Chain Complex I; Respiratory Chain Complex II and Succinate: Quinone Oxidoreductases; Respiratory Chain Complex IV.

Further Reading

Anthony C (ed.) (1988) *Bacterial Energy Transduction*. New York: Academic Press.
Driessen AJM, Rosen BP, and Konings WN (2000) Diversity of transport mechanisms: Common structural principles. *Trends in Biochemical Sciences* 25: 397–401.

Harold FM (1986) *The Vital Force: A Study of Bioenergetics*. New York: Freeman.
Kates M, Kushner DJ, and Matheson AT (1993) *The Biochemistry of Archaea*. Amsterdam: Elsevier.
Lengeler JW, Drews G, and Schlegel Hans G (eds.) (1999) *Biology of the Prokaryotes*. Stuttgart: Thieme.
Lolkema JS, Speelmans G, and Konings WN (1994) Na⁺-coupled versus H⁺-coupled energy transduction in bacteria. *Biochimica et Biophysica Acta* 1187: 211–215.
Schäfer G (2004) Respiratory systems of the Archaea: From minimal structures to supercomplexes. In: Zannoni D (ed.) *Progress in Photosynthesis and Respiration*, vol. 16, pp. 1–33. Dordrecht: Springer.
Schäfer G (2004) Extremophilic archaea and bacteria: New insights into their roles in membrane transport and other bioenergetic functions (Minireview Series). *Journal of Bioenergetics and Biomembranes* 36: 3–159.
Schäfer G, Engelhard M, and Müller V (1999) Bioenergetics of the archaea. *Microbiology and Molecular Biology Reviews* 63: 570–620.
Schäfer G and Penefsky H (eds.) (2008) *Bioenergetics, Energy Conversion and Conservation*. Berlin, New York: Springer.

Membrane Fusion

R K Plumper and G B Melikyan, Emory University School of Medicine, Atlanta, GA, USA; Children's Healthcare of Atlanta, Atlanta, GA, USA

Published by Elsevier Inc.

This article is a revision of the previous edition article by Joshua Zimmerberg, Leonid V. Chernomordik, volume 2, pp. 621–626, © 2004, Elsevier Inc.

Glossary

4-HB and 6-HB Highly stable four- and six-helix bundle structures composed of HR domains.

Fusion peptide and fusion loop Hydrophobic attack domains of viral fusion proteins, inserted into target membranes for viral entry.

Fusion pore Resolving of the hemifusion stage through merger of the inner layer, resulting in completion of the fusion process.

Hemifusion Intermediate fusion state in which the outer layers of two approaching lipid bilayers have formed a continuum while the integrity of the inner layers remains uninterrupted.

Heptad repeat (HR) The 4-3 α -helical HR domain, harboring hydrophobic residues in every fourth and third position.

Pseudoatomic structures Overlay of high-resolution X-ray data on electron microscopy-derived electron density maps to reveal the structure of protein complexes *in situ*.

Single-molecule reconstitution Analysis of the activity of individual fusion protein complexes in artificial lipid environments.

Soluble N-ethylmaleimide-sensitive factor association protein receptor (SNARE) A key mediator of vesicular fusion.

Introduction

Membrane fusion is essential for the maintenance of basic functionality of eukaryotic cells and the formation of multicellular organisms. Extracellular fusion, that is, merger of two neighboring cells into a multinucleated syncytium, occurs during embryogenesis and, for instance, during muscle tissue development. Intracellular fusion, that is, merger of internal cellular membranes with each other or the plasma membrane, is, among others, essential for functionality of the eukaryotic secretory system through vesicular transport, activity of the endocytotic system and exocytosis of synaptic vesicles. In a variation of extracellular fusion, lipid-wrapped (enveloped) viral pathogens are required to merge their envelope with cellular membranes for the transfer of their genetic information into a target cell and for the initiation of a new replication cycle.

Despite the fact that every second numerous membrane fusion events take place in a single cell, neither biological nor artificial membranes fuse spontaneously. Rather, efficient membrane fusion depends on the activity of highly specified fusogenic transmembrane proteins. The article commences with an overview of protein mediators of cellular membrane fusion, then proceeds to a discussion of distinct structural classes of fusogenic viral envelope glycoproteins, and concludes with a summary of experimental methods applied to gain structural and mechanistic insight into the membrane-fusion process.

General Principles of Membrane Fusion

The lipid bilayer core represents the energetically most favorable structure of lipids in an aqueous environment. Summarized in the stalk–pore hypothesis ([Figure 1](#)), membrane merger is viewed to proceed through a local bending and

merger of two membranes. A close contact between planar bilayers is disfavored by powerful hydration forces, but highly localized contacts between the tips of protruding dimples appear feasible. Originating at the apex of these dimples, a hemifusion stalk is considered to emerge in which the outer layers of donor and target membrane form a continuum, while the integrity of the inner layers remains intact. Lateral expansion of a stalk results in creation of a hemifusion diaphragm, which is a new bilayer formed by inner monolayers of fusing membranes. Disruption of the hemifusion diaphragm then creates a fusion pore, and subsequent pore enlargement completes the fusion process.

All lipid intermediates described are energetically unfavorable due to the large elastic energy generated upon bilayer bending. In spite of these energetic penalties, membrane merger is effectively induced by cellular and viral fusion proteins. These considerations provide a framework for thinking about the necessary requirements placed on fusion proteins, which include: (1) bringing two membranes into close proximity; (2) promoting bilayer bending and destabilization; and (3) mediating their merger and fusion pore dilation. Additional clues to the mode of action of fusogens are once again provided by lipids themselves. It is well established that lipids with effectively smaller polar headgroup and larger hydrophobic moiety are prone to form curved inverted cylindrical structures and, thus, favor membrane hemifusion that progresses through a stalk intermediate of a similar net curvature ([Figure 1](#)). The membrane-interacting domains of fusion proteins are thought to alter the lipid environment in a way that favors the formation of highly curved nonbilayer intermediates. Recent structural, biochemical, and functional studies revealed that disparate fusion proteins have emerged during evolution. However, they promote membrane fusion through a universal mechanism, perhaps dictated by the common physical properties of lipid bilayers.

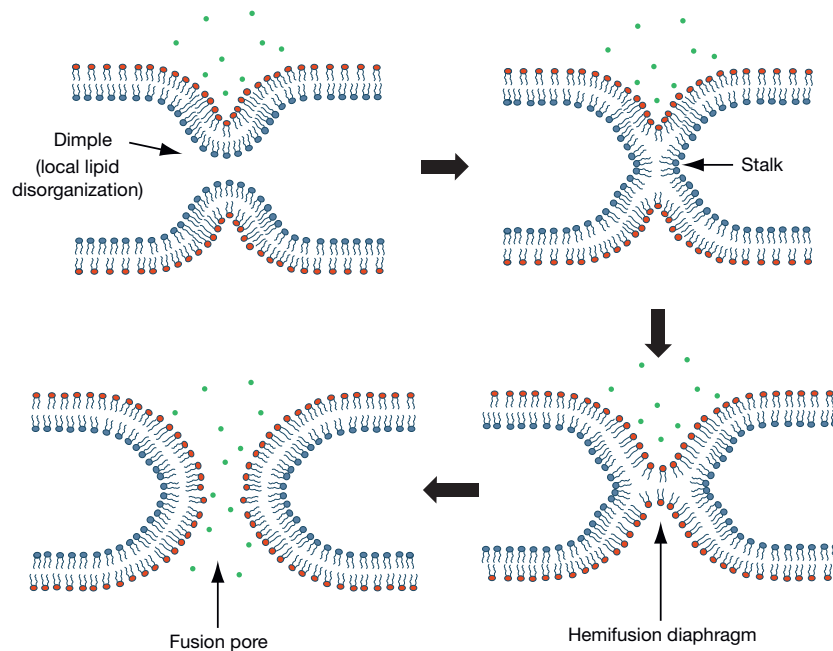


Figure 1 Fusion of pure lipid bilayers according to the stalk-pore hypothesis. Key intermediate steps of fusion are the local approach of two bilayers through dimpling, introduction of local high membrane curvature at the apex of protruding dimples, formation of a monolayer stalk, local hemifusion characterized by abutment of distal leaflets, and, finally, creation and enlargement of a fusion pore that completes the fusion process. Lipid headgroups of contacting and distal monolayers are colored blue and red, respectively.

Cellular Fusion Proteins

Although the nature of most *bona fide* mediators of extracellular fusion remains obscure at present, cellular fusogens involved in intracellular and synaptic fusion have been well characterized structurally and mechanistically. We, therefore, take a more detailed view on vesicle budding and fusion to exemplify the intracellular fusion process.

In striking similarity to the principal steps of virion formation and cell entry of enveloped viruses, vesicular transport involves concentration of cargo into the area of a budding vesicle at the donor membrane, scission of the lipid connector between the fully budded vesicle and the donor compartment, specific tethering of the free vesicle to an acceptor membrane, and protein-mediated membrane fusion, releasing the cargo into the acceptor compartment. Unlike the single fusion cycle of viral fusion proteins, however, cellular fusion proteins must be present on both donor and acceptor membrane and undergo an energy-dependent recycling step subsequent to fusion, rendering them capable of mediating multiple fusion events.

To accomplish membrane merger, it is essential that fusion protein rearrangement, which releases energy, is properly coupled to lipid membranes. The fusion process must, thus, be viewed in general as interplay between a complex system of proteins and lipids. Central to the cellular vesicular fusion and recycling machinery is the *N*-ethylmaleimide-sensitive factor (NSF) adenosine triphosphatase, a soluble NSF association protein (SNAP), and, constituting the *bona fide* fusogens, members of the superfamily of transmembrane SNAP receptors (SNAREs). It should be noted that in addition to this core fusion machinery, highly diverse families of tethering factors

are required for fusion activity. Organizing proper vesicular trafficking, these tethering factors form large complexes that, in many cases, directly associate with particular SNAREs and are dedicated to distinct trafficking pathways.

Although SNAREs vary widely in size and structure, they typically feature a short carboxy-terminal luminal tail, a trans-membrane domain (TMD), and a larger cytosolic domain harboring the defining 60–70-amino-acid SNARE motif. Located adjacent to the TMD, the latter includes eight heptad repeat (HR) domains characteristic for α -helical coiled coils. For fusion, SNAREs must be present on both the vesicular (v-SNARE) and the target (t-SNARE) membrane. Binding of a v-SNARE to a cognate t-SNARE complex in either 1:3- or 2:2-stoichiometry then results in the formation of a tetrameric *trans*-SNARE complex, in which the HR domains wrap into a thermodynamically highly stable coiled structure, bringing the v-SNARE and t-SNARE TMDs, and, thus, donor and target membrane into close proximity (Figure 2). Energetically favorable zippering of v- and t-SNAREs into this four-helix bundle structure is considered to be directly coupled to the introduction of local membrane curvature, leading to the formation of a hemifusion stalk. Opening of a fusion pore and, thus, completion of the fusion process coincide with the transition to the *cis*-SNARE complex that unites the TMDs of v- and t-SNAREs in a single lipid bilayer (Figure 2).

Docking of SNAP to the *cis*-SNARE complex and subsequent recruitment of NSF by SNAP initiates SNARE recycling. NSF is considered to catalyze adenosine triphosphate (ATP)-hydrolysis-dependent unwinding of the four-helix bundle, allowing dissociation of the complex and packing of the v-SNARE into a recycling vesicle for transport back to the donor compartment.

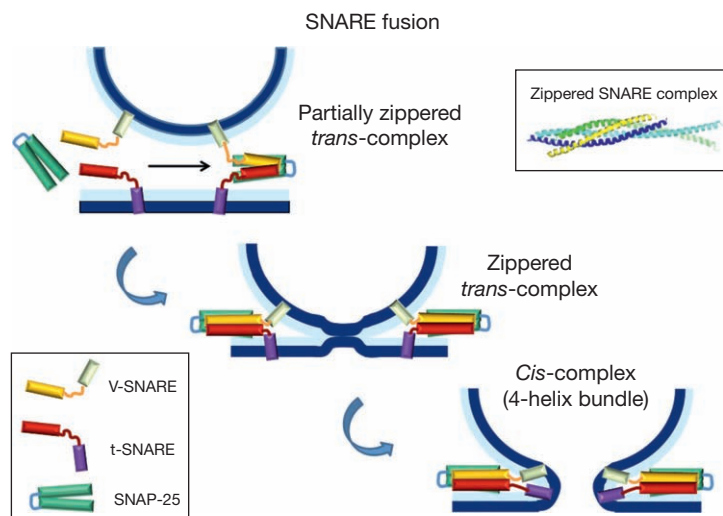


Figure 2 Cellular SNARE-mediated vesicle fusion with the plasma membrane. Inset: crystal structure of the fully zippered SNARE four-helix bundle (Protein Data Bank coordinates 3HD7). Top panel shows the vesicle tethering through the initiation of the *trans*-SNARE complex. Close contact and hemifusion are thought to occur through the formation and zippering of the *trans*-SNARE complexes (middle panel). Fusion pores (bottom panel) are likely directly coupled to *cis*-SNARE complex formation.

Viral Fusion Proteins

A unique feature of viral fusion proteins compared to the cell fusion machinery is that the former act in the absence of an external energy source. Here, a key strategy is to generate metastable, kinetically trapped prefusion conformations characterized by a relatively high free energy. An energized conformation of viral glycoproteins is created through a post-translational proteolytic cleavage or through other means. These structures are, thus, ready to refold into a more energetically favorable state while releasing the energy that drives membrane fusion. Both strengths and weakness of viral fusion stem from this principle. On the one hand, the fusion step cannot be usually inhibited by metabolic interventions. On the other hand, the requirements for successful virus entry are quite stringent, so that even under optimal conditions the probability of fusion is usually low. Perhaps as a result of a limited conformational energy, viral fusion is sensitive to the lipid composition of fusing membranes and requires a high density of activated fusion protein oligomers (>1 , probably 5–15 oligomers). The fact that enveloped viruses are devoid of active repair mechanisms has been recently exploited for designing a new class of antivirals that irreversibly diminish the propensity of the viral membrane to fuse, while exerting virtually no adverse effect on the cell membrane that quickly repairs itself.

When triggered by various cues (cellular receptors and/or low pH), viral proteins are thought to undergo the following conformational changes.

First, conserved hydrophobic segments, referred to as fusion peptides or fusion loops, are exposed and inserted into a target membrane (Figures 3–5). This step is critical for coupling the subsequent protein refolding to membrane merger. The insertion of hydrophobic domains also destabilizes the bilayer structure, facilitating the formation of curved lipid intermediates.

Second, viral proteins fold back from an extended conformation, positioning their fusion peptides/loops and TMDs at the same end of an elongated hairpin structure. As the N- and C-terminal ends of viral proteins are anchored to opposing membranes, the hairpins formation is expected to bring membranes together and force their fusion. There is strong evidence that an intermediate referred to as hemifusion (merger of contacting leaflets without the formation of a fusion pore) precedes full fusion. Hemifusion is likely induced by partially refolded protein conformations referred to as prehairpins (Figures 3–5). Progression of fusion through a hemifusion step ensures a nonleaky transition from two bilayers to one continuous membrane, as the inner and outer leaflets are not disrupted at the same time.

Third, in order to release the viral nucleocapsid and initiate infection, nascent fusion pores must enlarge to sizes comparable to the size of a virion. Forces driving the pore enlargement are poorly understood. This step could be driven by recruitment of new fusion proteins into a fusion pore and/or by cellular processes that promote pore dilation.

In summary, structurally diverse unrelated viral proteins appear to undergo a conserved ‘cast-and-fold’ type rearrangement *en route* to fusion. This refolding brings two membranes in contact and promotes their merger by favoring the formation of nonbilayer structures, such as a stalk and a hemifusion diaphragm. These proteins then destabilize the hemifusion diaphragm, likely by pushing the TMDs through the nonbilayer lipid structure. It must be stressed that viral fusion is not a deterministic process. Depending on the experimental conditions, such as the receptor density, pH, and other factors, a varied, but usually small, fraction of virions fuse with a cell membrane. The minimal viral fusion machinery, which relies solely on energy released upon the viral glycoprotein refolding, dictates the need to merge membranes through the least energy-intensive pathway.

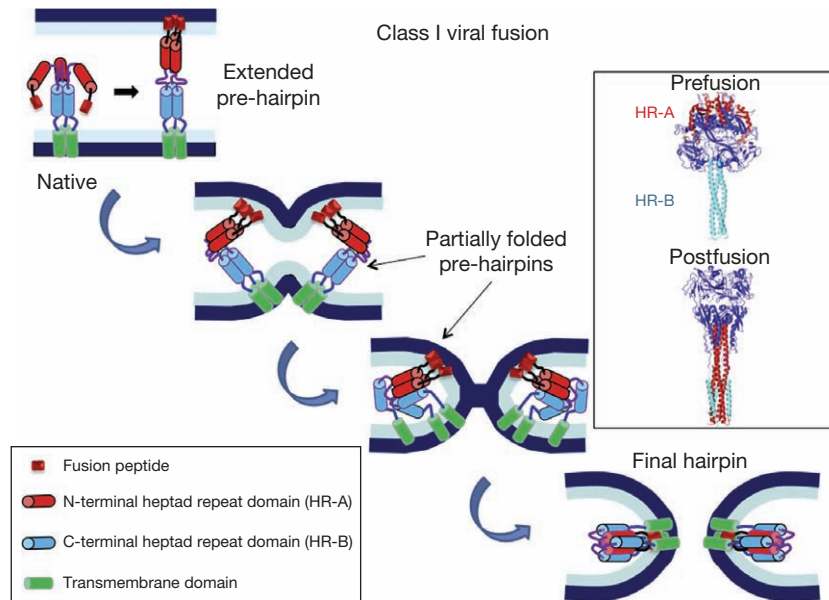


Figure 3 Membrane fusion driven by viral class I proteins. Inset: crystal structure of a class I viral fusion protein (paramyxovirus F protein) in pre- and postfusion conformations (Protein Data Bank coordinates 2B9B and 1ZTM). HR domains A and B are colored red and cyan, respectively, in the structure and in the schematic drawing of the F-protein refolding. The model of class I protein-mediated hemifusion and fusion depicts the progression through an extended prehairpin followed by breaking the threefold symmetry and dissociation of the HR-B domains. The final 6-helix-bundle structure is formed by a hairpin arrangement of three monomers with the peripheral positioning of the HR-B domains against the antiparallel central coiled-coil stem formed by HR-A.

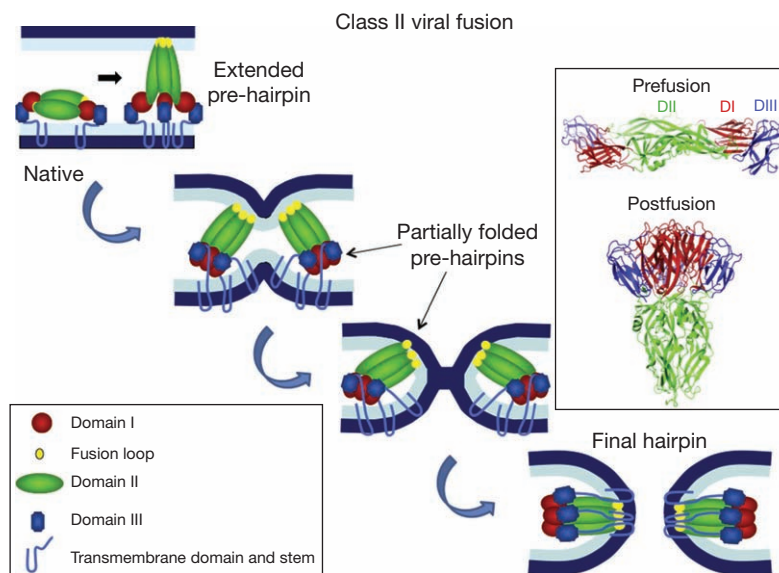


Figure 4 Class II viral protein-mediated fusion. Inset: crystal structure of a class II viral fusion protein (Dengue virus E protein) in its prefusion (antiparallel homodimer) and postfusion (homotrimer) conformations (Protein Data Bank coordinates 10AN and 10K8). The color coding of domains I–III is preserved throughout the figure. The schematic of class II protein-induced membrane fusion shows dissociation of prefusion dimers oriented nearly parallel to the viral membrane, engagement of the target membrane through insertion of the fusion loops (yellow) and reorganization into homotrimers. Subsequent refolding of the extended trimeric conformation into a hairpin structure promotes hemifusion and fusion pore formation.

Class I Fusion Proteins

Viral class I fusion proteins are found on major human and also on animal pathogens such as human immunodeficiency virus, influenza virus, Ebola virus, and measles virus, among

many others. Of all fusogenic viral envelope glycoproteins, class I representatives were the first to be described and crystallized in both the pre- and postfusion conformations. Class-defining features are assembly into noncovalently linked homotrimers, synthesis as inactive precursor molecules that

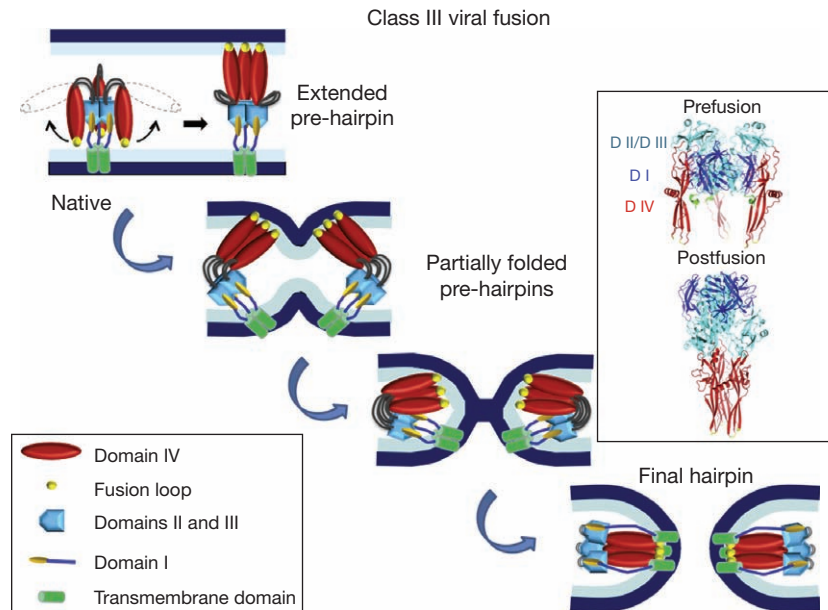


Figure 5 Membrane merger induced by class III viral fusogens. Inset: crystal structure of a rhabdovirus fusion protein (VSV G protein) in its pre- and postfusion conformations (Protein Data Bank coordinates 2J6J and 2CMZ). The coloring of domains I–IV is preserved in the model drawing. The schematic shows that, in reminiscence of the other two classes of viral fusion proteins, also class III proteins engage a target membrane by inserting their fusion loops in an extended intermediate conformation, followed by folding back onto themselves through hairpin formation, thus juxtaposing the fusion loops and the TMDs and forcing membrane merger.

require proteolytic maturation to gain fusion competence, and formation of a thermodynamically highly stable six-helix bundle fusion core structure.

In contrast to host cell fusion complexes, activity of viral fusion proteins is ATP independent. Upon synthesis, class I proteins are rather thought to fold into a metastable, spring-loaded conformation. Proteolysis then liberates the N-terminus of the fusion peptide domain, setting the stage for irreversible refolding into a thermodynamically stable postfusion conformation when a target membrane is in reach. In many cases, the cleavage step is mediated by ubiquitous host furin endoproteases located in the secretory system of the host cell, ensuring that proteolytically matured, fusion-competent complexes reach the cell surface and are incorporated into progeny viral particles. Alternatively, some viruses rely on extracellular proteases for proteolytic maturation subsequent to particle formation. For instance, the haemagglutinin protein of many influenza virus strains is recognized and specifically cleaved by trypsin-like extracellular proteases of the respiratory tract for maturation.

Considering the irreversible nature of class I fusion protein conformational rearrangements, proper timing of the initiation of refolding is essential for productive infection: premature triggering in the absence of a target membrane results in self-inactivation of a viral particle, while delayed fusion activation may miss the narrow window of opportunity available for successful infection. Depending on the viral family, conformational rearrangements can be triggered either at neutral pH through receptor binding (i.e., members of the paramyxovirus family such as measles virus and mumps virus) or through protonation in the acidic environment of the endosome after endocytotic uptake of viral particles (i.e., influenza virus).

Independent of the mechanism of fusion triggering, characteristic for all class I fusion proteins are two HR domains, positioned adjacent to the fusion peptide and TMD, respectively. **Figure 3** illustrates major structural features of the class I fusion process by example of the paramyxovirus fusion (F) protein. In the prefusion conformation, the F protein assumes a lollipop-like structure consisting of a globular head domain that is linked to the TMDs and short luminal tails through the TMD-proximal HR sequences. In this metastable conformation, the fusion peptide-proximal HR domains are broken up into 11 distinct segments forming part of the head structure. Upon fusion activation, major structural rearrangements lead to assembly of the head-domain HR segments into a central, trimeric α -helical coiled-coil structure, propelling the fusion peptides toward the target membrane. Subsequent hairpin-like refolding then positions the individual TMD-proximal HR domains in the grooves of the central triple helix, resulting in the formation of the stable six-helix bundle (6HB) postfusion structure. This final arrangement of a fusion peptide-proximal core coiled-coil structure surrounded by the three TMD-proximal HR helices is a defining feature of class I fusion proteins.

In striking similarity to the coupling of four-helix bundle formation to membrane merger in SNARE complex-mediated vesicular fusion, 6HB formation appears closely linked to the opening of a fusion pore. In both cases, zippering of thermodynamically highly stable α -helical structures is considered to provide the major driving force for pore formation and expansion. Consequently, peptides or small molecules that prevent final 6HB formation of viral class I proteins by interfering with refolding of the fusion protein function as potent viral entry inhibitors and are, as in the case of the HR-derived peptidic HIV entry inhibitor Fuzeon, in clinical use as antiviral therapeutics.

It becomes immediately evident from the schematic shown in **Figure 3** that refolding of the fusion protein trimer cannot proceed symmetrically, as this would wrap the protein into the lipid of the viral envelope. Instead of an umbrella-like folding of the protein through itself, an asymmetric process must, therefore, take place in which individual monomers of the trimer progress through the conformational rearrangements at different rates and individual TMD-proximal HR domains possibly engage the grooves in the central triple helix coiled-coil consecutively rather than in a concerted action.

Class II Fusion Proteins

The envelope glycoproteins of togaviruses and flaviviruses are grouped into class II fusion proteins. Unlike class I viral proteins, these proteins exist as homo- or heterodimers in their native state. With high content of β -sheets, they are structurally distinct from other classes of viral fusogens (**Figure 4**). Homodimer arrangements are typical for the flavivirus E protein exemplified by the tick-borne encephalitis and dengue viruses, whereas E1 proteins of togaviruses (e.g., Semliki Forest virus) are associated with a chaperone protein in their mature form. Furthermore, in contrast to class I proteins, class II proteins are not oriented perpendicular to the viral membrane but run parallel or nearly parallel to the envelope, covering the surface of small icosahedral viral particles.

Class II proteins consist of three major structural domains (DI, DII, and DIII), of which the C-terminal DIII domain is connected to the TMD through a stem region. Another important difference between these and class I fusogens is that, instead of the N-terminal or N-proximal fusion peptides, each class II protein contains an internal fusion loop, which is a part of the DII domain. In the native structure, these loops are sequestered at the dimer interface and are positioned close to the viral membrane.

Although the receptors for class II viruses are generally not identified, these proteins mediate fusion with nearly all cell types by entering through endocytic routes. The ability to fuse at mildly acidic pH extends to fusion with artificial liposomes. The low pH-induced conformational changes proceed through dissociation of homo- or heterodimers and formation of E/E1 homotrimers that engage a target membrane by inserting their fusion loops (**Figure 4**). Subsequent conformational changes involve the reorientation of the DIII domain around a hinge region that positions the transmembrane region and the fusion near each other and presumably within the same membrane. In other words, in spite of the marked differences in their structures, class I and class II proteins appear to progress through a similar refolding pathway: formation of extended conformations followed by the folding into a collapsed hairpin structure, characterized by antiparallel arrangements of the major structural subdomains. As in the final postfusion structure of paramyxovirus F protein discussed above, tight interactions between the inverted DIII domain and the central core stabilize the hairpin structure and thereby promote membrane merger. Functional evidence supporting this notion includes the ability of a recombinant DIII protein to bind to the extended conformation of a class II fusion protein and to block its subsequent refolding and membrane merger.

An interesting feature unique to some class II proteins is that fusion loop insertion into the target membrane and fusion is dependent on the lipid composition. For instance, the low pH-triggered insertion of the Semliki Forest virus E1 into the target membrane is cholesterol dependent, while its ability to fuse membranes is dependent on the presence of small amounts of sphingolipids. In addition, accumulating evidence supports that the late steps of fusion mediated by Dengue virus of the flavivirus family require acidic lipids.

Class III Fusion Proteins

The group of viral class III fusion proteins currently comprises rhabdovirus G protein, herpesvirus gB, and baculovirus gp64. In addition to postfusion conformations of gB and gp64, the structural understanding of this class was greatly advanced by crystallization of the vesicular stomatitis virus (VSV) G-protein trimer in the pre- and postfusion conformations (**Figure 5**). The G protein mediates both receptor binding and, triggered by the low pH of the endosome, membrane merger. Resembling central features of class I and II fusion proteins, refolding from the prefusion conformation assumed at neutral pH involves a prehairpin intermediate that engages both viral envelope and target membranes, and a hairpin-like postfusion conformation in which fusion loop and TMDs rest in close proximity. In striking contrast to the other fusion protein classes, however, G refolding is reversible, adhering to a pH-driven equilibrium. Consequently, the native prefusion conformation is not metastable, enabling the trimer to reassume its native configuration subsequent to transport through the low-pH environment of the Golgi apparatus.

Native G trimers fold into a tripod-like arrangement consisting of four defined sections, a lateral domain with high β -sheet content, a trimerization domain, a PH domain located toward the top of the tripod and, forming the tripod legs, a fusion domain (**Figure 5**). The fusion loops are positioned at the tip of each leg, pointing toward the viral envelope. Of these, the trimerization domain becomes subject to major refolding, while the other domains largely maintain their fold during transition from pre- to postfusion conformation. Among others, exposure to low pH results in protonation of a specific cluster of histidine residues that is located at the intersection between domains and is considered to exude a major destabilizing effect on the prefusion structure. Once triggered, the tripod legs are proposed to swing upward, driving the fusion loops toward the target membrane. Repositioning of the domains with conserved tertiary structure relative to each other through secondary structure reorganization in hinge regions and major changes of the trimerization domain then ultimately result in a classic hairpin postfusion conformation, in which fusion loops and TMDs are positioned in proximity in the same lipid space.

Despite the absence of HR domains in VSV G, two helices of each postfusion domain II engage in formation of an antiparallel 6HB in the postfusion trimer in strong reminiscence of the class I protein fusion core structure. A set of acidic amino acids, brought into close proximity in the central 6HB in the postfusion conformation, appear to serve as a second

major molecular switch to initiate transition back to the native state upon deprotonation at elevated pH.

Methods to Study Membrane Fusion

As discussed above, membrane fusion is a transient and local event. An important confounding factor for mechanistic studies is that fusion occurs in parallel with an excessive number of dead-end protein–lipid interactions that do not lead to fusion. These nonproductive events mask the actual fusion reaction, thereby limiting the use of standard experimental approaches and necessitating the development of new strategies to study cellular and viral fusion.

Fusion Protein Structure

Although membrane proteins in general and viral fusion proteins in particular (by virtue of their metastable nature) are difficult to crystallize, high-resolution structures have been obtained of the ectodomains of a number of SNARE complexes and viral fusion proteins. These are typically based on expression of soluble ectodomain fragments lacking the TMDs. For viral fusion proteins, these ectodomain structures in pre- and postfusion conformations revealed stunning conservation of the refolding pathways. However, less-stable envelope glycoproteins, such as the HIV-1 gp120/gp41, have not been crystallized yet in their native oligomeric form. Recent advances in cryo-electron tomography and image reconstruction techniques bridged the gap between *in situ* light microscopy and crystallography by achieving a resolution of as low as five nanometers for morphologically uniform symmetric viruses. These techniques provide new insights into the architecture of proteins in the viral membrane, their oligomeric status, and lateral distribution. Overlaying cryo-electron tomography-derived electron density maps with available X-ray data allow the generation of pseudoatomic structures that reveal the spatial organization of the native fusion protein complexes at high-resolution *in situ*. Other methods, such as nuclear magnetic resonance, have been successfully applied to solve the structure of smaller protein fragments in solution, including trimeric hairpin subdomains and fusion peptide domains.

Conformational Changes

Major structural changes in fusion proteins have been documented by crystallography, but the available structures are usually limited to native and postfusion conformations of their ectodomains. Less-direct approaches include changes in reactivity to conformation-specific monoclonal antibodies or virus-derived peptides that specifically bind to transiently emerging complementary regions on a refolding fusion protein. Biophysical techniques, such as circular dichroism, calorimetry, and spectroscopy, have been used to study the secondary structure of soluble fragments and changes in the protein structure under fusion-permissive conditions. Exposure of hydrophobic domains (fusion peptides or loops) of viral proteins is often detected based on their association with liposomes or the enhanced fluorescence of dyes binding to hydrophobic domains of these proteins.

Protein–Protein Interactions

Whether or not SNARE-mediated cellular and viral fusions are cooperative processes that require a concerted action of several fusogenic proteins is debated. At least in the case of SNARE-mediated fusion, single-molecule *in vitro* fusion studies support the notion that a lone SNARE complex may in fact be sufficient for fusion. However, efficient, rapidly progressing SNARE-mediated fusion and viral entry require the coordinated activity of multiple fusion complexes. A prevailing view is thus that fusion proteins assemble into multimeric pre-fusion complexes that act cooperatively to promote membrane merger. Importantly, fusion proteins of viruses that rely on auxiliary receptor-binding or regulatory proteins (e.g., herpesviruses and paramyxoviruses such as measles virus and mumps virus) are triggered through lateral interactions with these proteins. Protein–protein interactions and the stoichiometry of pre-fusion complexes have been investigated by biochemical and fluorescence techniques, including co-immunoprecipitation, Foster resonance energy transfer (FRET), and bimolecular complementation. The latter fluorescence techniques report the molecular proximity of two proteins based on either quenching of donor fluorescence by fluorescence acceptor molecules attached to viral proteins or SNARE counterparts (FRET) or assembly of a functional green fluorescent protein (GFP) molecule from its two nonfluorescent subdomains as a result of fusion protein dimerization or oligomerization.

Functional Studies

Functional approaches are widely employed in studies of membrane fusion, permitting detection of lipid merger and monitoring the dynamics of this process. Because membrane fusion involves the establishment of membrane continuity and merger of aqueous compartments delimited by two membranes, fusion is manifested in redistribution of membrane probes and in transfer of aqueous content. Sensitive lipid-mixing assays are also used to detect and characterize a hemifusion intermediate (lipid mixing without content transfer) formed during intracellular, cell–cell and virus–cell fusion. Traditionally, functional studies of viral fusion have relied on the cell–cell fusion model, which provided a convenient means to study the activity of transiently expressed envelope glycoproteins without interference by other viral proteins. This model system is amenable for mechanistic studies employing quantitative fluorescence microscopy, electrophysiology, and flow cytometry assays that detect early steps of fusion on a single-cell level, as manifested by the redistribution of membrane or cytoplasmic fluorescent markers. Somewhat less sensitive, albeit high-throughput-capable, assays rely on protein transfer (enzymes, autofluorescent proteins, or promoters of reporter gene expression) between fusing cells. The end-point measurements of syncytia formation (seen as multinucleated cells) correspond to an ultimate dilation of a fusion pore, which is contingent on a large-scale cytoskeleton rearrangement. Because this assay detects remote consequences of fusion, a lack of syncytia formation may not always indicate absence of function, as small nonenlarging fusion pores can still be present under these conditions.

Single Virion Entry Assays

Direct techniques for monitoring virus–cell fusion (not counting the traditional infectivity assays) have been developed only recently. These assays are based on transfer of virus-incorporated reporter enzymes into the cytosol of a target cell. Advanced fluorescence microscopy techniques permit visualizing entry and fusion of single viral particles. Viral fusion has also been examined using artificial lipid membranes (liposomes and supported bilayers) as targets. Fusion in these minimal model systems allows studies of lipid-dependence and virus–lipid interactions.

In Vitro Reconstitution and Single-Molecule Analysis

Reconstitution of SNARE-mediated fusion of artificial liposomes under *in vitro* conditions has set the stage for the analysis of individual fusion events, a prerequisite for gaining detailed molecular insight into the mechanism of membrane merger as discussed above. Assessment of single molecules in a reconstituted system has been a major breakthrough in our insight into the organization of fusion complexes and the dynamics of the process, because it enables the separation of individual complexes engaged in fusion from the bulk of all molecules in diverse conformational stages that are always present in complex biological systems. Although bulk analysis in the absence of synchronization, thus, returns only the average behavior of the system, the isolated assessment of specific molecules provides a view of the individual molecular fate of a complex engaged in membrane fusion. Frequently coupled with fluorescence methodology, for instance, in single-molecule-FRET studies, protein–protein interactions and protein complex formation can be monitored through this approach while membrane fusion proceeds. These techniques, thus, allow to track the dynamic protein–protein and protein–lipid interactions required for fusion in nearly real time.

Concluding Remarks

The basic principles of viral envelope glycoprotein-mediated membrane fusion, engagement of donor and target membranes followed a juxtaposing membrane-associated domains in the same lipid environment, and local disturbance of the lipid bilayer through introduction of a high degree of membrane curvature have emerged as universally conserved. All viral fusion proteins harbor fusion peptides or loops in addition to the TMDs. Anchoring these in the target membrane in the extended intermediate conformation engages both membranes simultaneously while bypassing the need for a t-SNARE equivalent on the target membrane. By contrast, cellular SNARE proteins must be present in both donor and target membrane, forming a thermodynamically highly stable four-helix bundle that brings the TMDs into proximity. Either four-helix bundle formation or refolding into a hairpin-like structure ultimately brings then the TMDs or TMDs and fusion peptides/loops in close proximity in a continuous lipid environment. This step is directly linked to opening of a fusion pore.

Although X-ray structures of SNARE complexes and viral fusion protein ectodomains in pre- and postfusion

conformations have resulted in a breakthrough in our general structural understanding of membrane fusion, much less is known about the molecular details of lipid merger, the nature of short-lived intermediate conformations of the fusion complexes, and, in the case of viral fusion proteins, their spatial arrangement on the surface of infectious particles. Addressing these questions will likely require interfacing novel-imaging approaches such as cryo-electron tomography reconstructions of emerging fusion pores and refolding complexes with single-molecule analyses of the fusion dynamics in reconstituted systems. Recent advances toward the analysis of SNARE-mediated vesicle fusion have opened a path toward experimentally illuminating the dynamic changes of individual complexes. Application of this approach to viral fusion proteins is anticipated to further boost our understanding of conserved and unique features of the different fusion pathways. Ultimately, detailed, high-resolution structural and mechanistic insight into the viral fusion process has high potential to aid in the development of novel therapeutics through structure-guided drug design.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Detergent Properties; **Protein/Enzyme Structure Function and Degradation:** Lipid Modification of Proteins: Targeting to Membranes.

Further Reading

- Boninfacino JS and Glick BS (2004) The mechanisms of vesicle budding and fusion. *Cell* 116: 153–166.
- Brunger AT, Weninger K, Bowen M, and Chu S (2009) Single-molecule studies of the neuronal SNARE fusion machinery. *Annual Reviews of Biochemistry* 78: 903–928.
- Chernomordik LV and Kozlov MM (2003) Protein–lipid Interplay in fusion and fission of biological membranes. *Annual Reviews of Biochemistry* 72: 175–207.
- Harrison SC (2007) Principles of virus structure. In: Knipe DM and Howley PM (eds.) *Fields Virology*, 5th edn., pp. 59–98. Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins.
- Harrison SC (2008) Viral membrane fusion. *Nature Structural and Molecular Biology* 15: 690–698.
- Helenius A (2007) Virus entry and uncoating. In: Knipe DM and Howley PM (eds.) *Fields Virology*, 5th edn., pp. 99–118. Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins.
- Jahn R, Lang T, and Suedhof TC (2003) Membrane fusion. *Cell* 112: 519–533.
- Kielian M and Rey FA (2006) Virus membrane-fusion proteins: More than one way to make a hairpin. *Nature Reviews Microbiology* 4: 67–76.
- Lamb RA and Parks GD (2007) *Paramyxoviridae*: The viruses and their replication. In: Knipe DM and Howley PM (eds.) *Fields Virology*, 5th edn., pp. 1449–1498. Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins.
- Lindenbach BD, Thiel HJ, and Rice CM (2007) *Flaviviridae*: The viruses and their replication. In: Knipe DM and Howley PM (eds.) *Fields Virology*, 5th edn., pp. 1101–1152. Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins.
- Lyles DS and Rupprecht CE (2007) *Rhabdoviridae*. In: Knipe DM and Howley PM (eds.) *Fields Virology*, 5th edn., pp. 1363–1408. Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins.
- Roche S, Albertini AA, Lepault J, Bressanelli S, and Gaudin Y (2008) Structures of vesicular stomatitis virus glycoprotein: Membrane fusion revisited. *Cellular and Molecular Life Sciences* 65: 1716–1728.
- White JM, Delos SE, Brecher M, and Schornberg K (2008) Structures and mechanisms of viral membrane fusion proteins: Multiple variations on a common theme. *Critical Reviews in Biochemistry and Molecular Biology* 43: 189–219.
- Wickner W (2010) Membrane fusion: Five lipids, four SNAREs, three chaperones, two nucleotides, and a Rab, all dancing in a ring on yeast vacuoles. *Annual Review of Cell and Developmental Biology* 26: 115–136.

Membrane Transporters: Na⁺/Ca²⁺ Exchangers

J Lytton, University of Calgary, Calgary, AB, Canada

© 2013 Elsevier Inc. All rights reserved.

Glossary

Alternative splicing A process in which the individual expressed regions of a gene (exons) are arranged in different patterns to produce different messenger RNA (mRNA) molecules that often encode different proteins.

Excitation–contraction coupling The process by which an electrical signal generated in the plasma membrane of a muscle cell results in movement of the muscle fiber proteins and contraction of the cell.

Na⁺/Ca²⁺ exchanger A plasma membrane protein that couples the movement of Na⁺ ions in one direction with Ca²⁺ ions in the other.

Plasma membrane The membrane that surrounds every cell, separating inside from outside.

Retinal rod photoreceptors Long, cylindrically shaped, cells in the retina of the eye that respond to changes in low levels of light.

α-Repeat A short stretch of amino acids identified within the Na⁺/Ca²⁺ exchanger that is conserved among a large superfamily of membrane proteins. Two copies of this conserved sequence are present in each Na⁺/Ca²⁺ exchanger, which are modeled to interact to form the ion-binding and translocation pocket.

Cellular Ca²⁺ Homeostasis

Calcium ion (Ca²⁺) is one of the most ubiquitous intracellular second messengers. Changes in the concentration of Ca²⁺ control a variety of cellular processes including fertilization, proliferation, differentiation, muscle contraction, neuronal signaling, and even programmed cell death. The coordination of such a diverse set of functions is thought to be achieved through Ca²⁺-binding events that are sensitive to complex spatial and temporal patterns of intracellular Ca²⁺ concentration. Unlike other second messenger systems, Ca²⁺ cannot be metabolized, and its concentration is controlled primarily by transport across membranes that bound intracellular compartments (such as the sarco- or endoplasmic reticulum and the mitochondrion) and the cell periphery (the plasma membrane). Cells possess a complex array of different Ca²⁺-transporting mechanisms, and the process of Na⁺/Ca²⁺ exchange is a major contributor to Ca²⁺ efflux across the plasma membrane. The unique and precise nature and arrangement of different Ca²⁺ transport protein molecules in different cells are generally thought to provide a major component of control over complex Ca²⁺ dynamics. When this carefully orchestrated homeostasis goes awry and Ca²⁺ concentrations are dysregulated, the repercussions are usually severe, resulting in damage and cell death which then cause the pathological consequences of diseases such as stroke, neurodegeneration, and heart disease.

Discovery of the Na⁺/Ca²⁺ Exchanger

A competitive relation between the effects of extracellular Na⁺ and Ca²⁺ on intracellular Ca²⁺ and contraction in cardiac, skeletal, and smooth muscle was first noted over 50 years ago. Then, in the late 1960s, three different groups (Reuter, Baker, and DeLuca), working on three different tissue preparations (mammalian cardiac muscle, squid axon, and mammalian

intestinal epithelium) concluded that the coupled counter-transport of Na⁺ for Ca²⁺ across the plasma membrane – so-called Na/Ca exchange – could explain the apparent competitive relation between Na⁺ and Ca²⁺. This finding established the physiological importance of Na⁺/Ca²⁺ exchanger as a mechanism contributing to the control of Ca²⁺ homeostasis, and launched the search for a protein molecule underlying the activity.

The development of biochemical measurements of Na⁺/Ca²⁺ exchanger function in membrane preparations and effective solubilization and reconstitution procedures allowed several groups to partially purify Na⁺/Ca²⁺-exchange activity. Aided by immunological approaches, Philipson's group provided strong evidence that a 120-kDa protein of canine cardiac sarcolemmal membranes, and its 70-kDa proteolytic product, corresponded to the Na⁺/Ca²⁺ exchanger. The use of recombinant expression cloning techniques then led to the seminal molecular cloning work from Philipson's laboratory, published in *Science* in 1990, which confirmed the identity of the canine cardiac Na⁺/Ca²⁺ exchanger.

During the 1980s, studies on visual adaptation in mammalian retinal rod photoreceptors also provided evidence for a Na⁺/Ca²⁺-exchange mechanism responsible for controlling cytosolic Ca²⁺ in that environment. Further biochemical studies of rod Na⁺/Ca²⁺ exchange revealed an obligate dependence on K⁺, and demonstrated that the rod Na⁺/Ca²⁺ exchanger catalyzed the exchange of Na⁺ for both Ca²⁺ and K⁺. As with the cardiac exchanger, solubilization and reconstitution studies allowed purification of the bovine rod photoreceptor Na⁺/Ca²⁺+K⁺ exchanger, which was identified as a ~220-kDa protein. Immunological tools were then used to clone the molecule in 1992 by Cook's laboratory. Surprisingly, despite the mechanistic similarity and a broadly similar topology pattern (Figure 1), the cardiac and rod exchangers share sequence similarity in only two short stretches, which were subsequently termed the 'α-repeats' (see below).

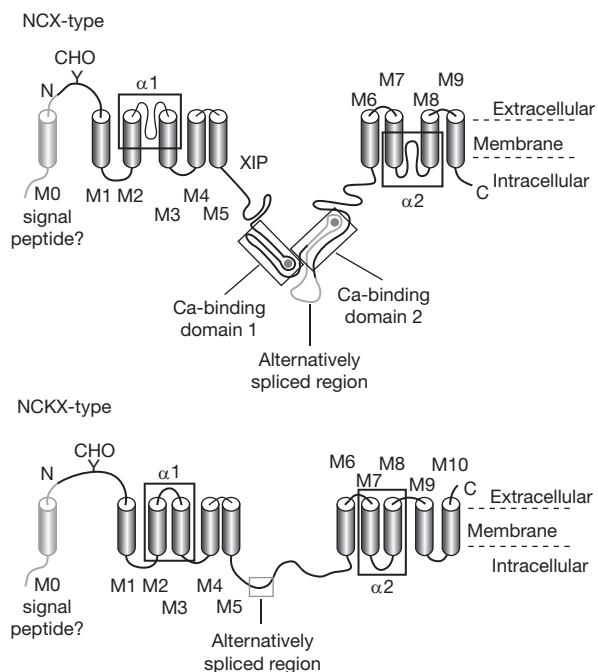


Figure 1 Topology models for $\text{Na}^+/\text{Ca}^{2+}$ exchangers (NCX-type) and $\text{Na}^+/\text{Ca}^{2+}+\text{K}^+$ exchangers (NCKX-type). The predicted folding of each class of protein with respect to the membrane, based on hydropathy analysis and experimental data, is shown. The cylinders labeled M0 (putative signal peptide, removed co- or posttranslationally) to M10 indicate predicted transmembrane helices. The conserved α -repeat regions (α -1 and α -2) are boxed and labeled, as are the two calcium-binding domains of NCX. The location at which Ca^{2+} binds within each domain is illustrated with a gray sphere (four Ca^{2+} bind at this site in domain 1, while two Ca^{2+} bind in domain two). The region subject to alternative splicing is indicated, and shown in gray for NCX and boxed for NCKX. The N-terminus (N) and C-terminus (C) of the mature protein, as well as sites of glycosylation (CHO) are also indicated.

An Expanding Gene Family

Cloning of the canine cardiac $\text{Na}^+/\text{Ca}^{2+}$ exchanger (NCX1 – gene *SLC8A1*) and the bovine rod photoreceptor $\text{Na}^+/\text{Ca}^{2+}+\text{K}^+$ exchange (NCKX1 – gene *SLC24A1*) opened the exchanger field to molecular analysis. NCX1 was subsequently cloned from heart, brain, and kidney of several different animal species. In the course of these studies, it became evident that the NCX1 gene gives rise to a complex set of alternatively spliced messenger RNA transcripts whose protein products differ in one relatively small region of the protein (Figure 1). The power of molecular cloning was also harnessed to uncover the existence of structurally related gene products, revealing that NCX1 is a member of a gene family that includes two other members, NCX2 (*SLC8A2*) and NCX3 (*SLC8A3*), whose products display about 70% amino acid identity with one another. NCX3 is also subject to alternative splicing, in a similar location as NCX1.

Advances in the molecular analysis of the NCKX family were not so rapid, but by the end of the 1990s a second member, NCKX2 (*SLC24A2*), had been identified. At the turn of the millennium, information contributed to the databases from random-sequencing efforts and from the various animal genome projects had turned the identification of homologous genes

from a bench experiment into a computer search. The discovery of NCKX3 (*SLC24A3*), NCKX4 (*SLC24A4*), and, most recently, NCKX5 (*SLC24A5*) followed rapidly. The NCKX protein family shares lower amino acid similarity than NCX proteins, and is subdivided, with the members of the pairs NCKX1 and 2, and NCKX3 and 4, each sharing about 60% identity, while the identity between pairs is much lower (Figure 2).

During the early phase of homologous gene discovery, it was recognized that NCX1 contains two regions of paired internal similarity, suggesting an ancient gene-duplication event. One pair is found in the cytoplasmic loop of the protein (originally noted as the ‘ β -repeat’) and the other pair resides, one member within each hydrophobic membrane-spanning domain (the ‘ α -repeat’; see Figures 1 and 2). As noted, the α -repeats contain the only regions of similarity between NCX and NCKX exchangers. Moreover, a database search with the α -repeat motif identifies a large number of genes in species ranging from higher vertebrates to flies, worms, plants, yeast, and bacteria (the Ca^{2+} -cation antiporter (CaCA) superfamily). The deduced proteins of these database hits contain two sequences homologous to the α -repeat, each within a relatively hydrophobic region, suggesting that these proteins are all membrane transporters, although, in most cases, the function of the gene products has not been determined experimentally. In some cases, the degree of sequence identity extends beyond the α -repeats, and these genes are likely to be orthologs of known mammalian NCX or NCKX exchanger genes. In other cases, the sequences are quite divergent, and the function of the encoded protein is not clear, but is presumed to involve cation exchange, for example, $\text{Ca}^{2+}/\text{H}^+$ or $\text{Mg}^{2+}/\text{H}^+$ exchange.

Expression of NCX and NCKX in Mammals

NCX1 is ubiquitously expressed in almost all mammalian tissues, with the highest level of expression found in heart, brain, and in kidney. The predominant NCX1 transcript in each of these three tissues corresponds to a different alternatively spliced species. The expression of NCX2 and NCX3 is highly restricted, with NCX2 present at high levels in the brain, but not at significant levels elsewhere, while NCX3 is present in skeletal muscle and at lower levels in selected brain regions, but is not significantly expressed in other tissues. The NCX1 protein has been localized mainly to the T-tubules in cardiomyocytes, to axonal and dendritic processes of neurons, and to the basolateral membrane of distal-convoluted tubule connecting segment cells in the kidney. Studies on localization of the NCX2 and NCX3 proteins are less well developed.

NCKX1, originally defined in rod photoreceptors, is expressed almost exclusively in the eye. NCKX2 is expressed at high abundance in neurons throughout the brain and in the cone photoreceptors, but only at much lower levels elsewhere. Both NCKX3 and NCKX4 are abundant in brain neurons but, in contrast to NCKX1 and NCKX2, are also broadly distributed in a number of other tissues and cell types. NCKX5 is abundant in skin, eye, and selected brain regions. The NCKX1 protein is present in the outer segments of rod photoreceptors. NCKX2 is enriched in neuronal processes particularly in the hippocampus. Little is known about endogenous NCKX3 and NCKX4 protein expression, but when expressed

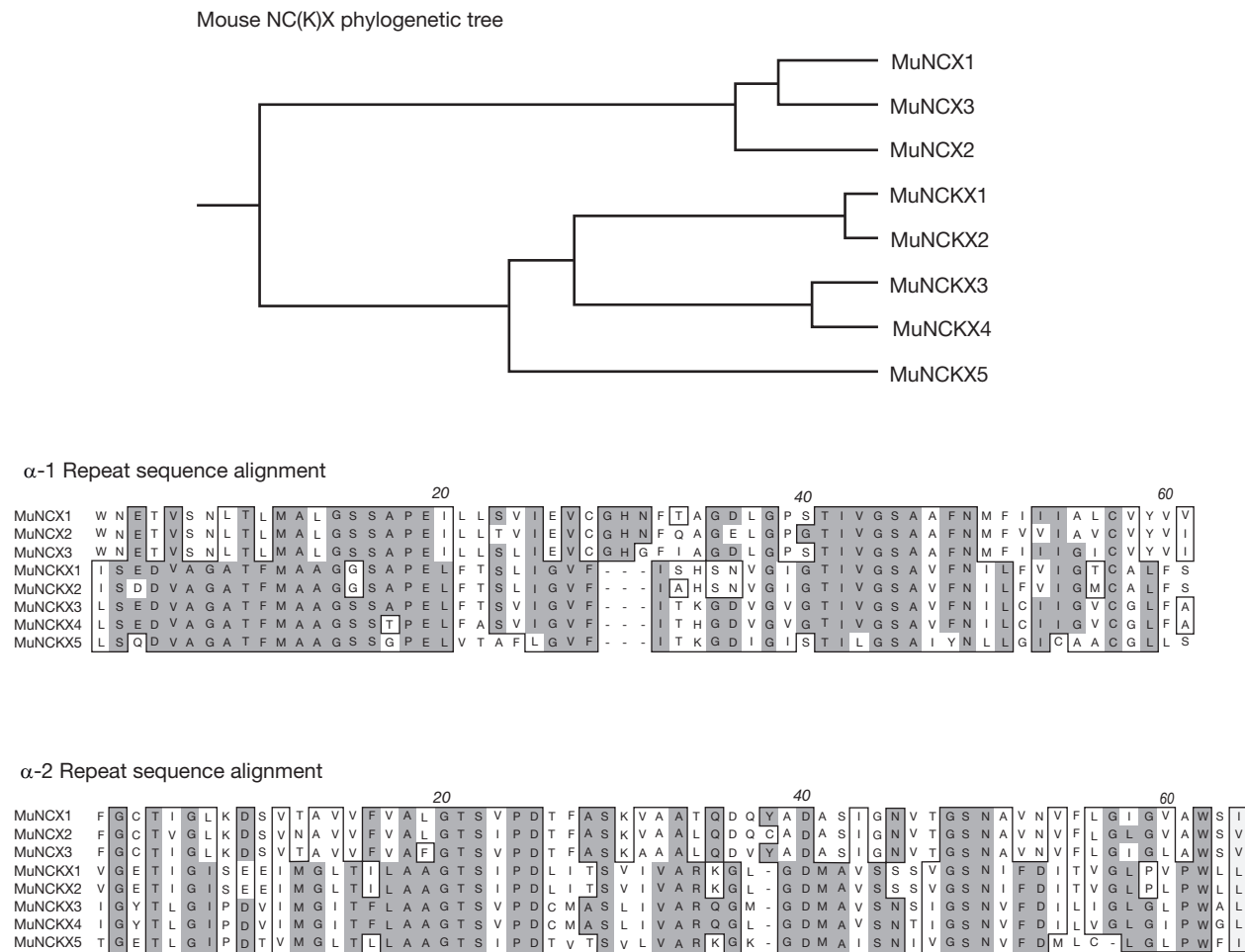


Figure 2 Phylogenetic relation and sequence alignment. The evolutionary relation between the different murine NC(K)X molecules is shown as a phylogenetic tree based on the combined similarities of the two α -repeat regions. A qualitatively similar pattern is obtained if the entire protein sequence from each isoform is used for the analysis. The sequences corresponding to the two α -repeats are shown aligned. Amino acids that are identical in five or more of the sequences shown boxed, dark shaded and highlighted in bold; chemically similar amino acids are boxed and light shaded.

in recombinant systems, these proteins are clearly found at the plasma membrane like NCKX1 and NCKX2. In distinct contrast, NCKX5 expressed in melanocytes is found on intracellular membranes associated with the *trans*-Golgi complex and melanosome particles. These data define NCKX5 as the only member of the NCX and NCKX families that appears not to reside on the plasma membrane.

Very recently, the only known member of a distinct clade of the CaCA superfamily (the CCX branch), previously characterized as a Na⁺/Ca²⁺ exchanger capable of using Li⁺ to drive Ca²⁺ flux and variously named NCLX or NCKX6 (though not actually a true member of the NCKX branch), was implicated as a component of the long-sought-after mitochondrial Na⁺/Ca²⁺ exchanger. Further analysis and confirmation of this important finding are eagerly awaited.

Structural Properties

Detailed discussion of the structural features of Na⁺/Ca²⁺ exchangers is covered in another chapter in this volume, and

is mentioned here only briefly. Broadly speaking, all the Na⁺/Ca²⁺-exchanger protein family members are predicted to have a similar overall architecture (Figure 1), comprising two hydrophobic membrane-embedded regions separated by a long central cytoplasmic loop. Within each branch of the family (i.e., NCX and NCKX), the length and sequence of each hydrophobic segment are quite well conserved, especially in the α -repeat regions (Figure 2). In contrast to the hydrophobic regions, the sequence of the exposed hydrophilic loops is quite variable between different family members. Of note, although NCKX-family members are predicted to have 10 transmembrane spans and an extracellular C-terminus, NCX1 is predicted to have 9 transmembrane spans and an intracellular C-terminus (Figure 1).

Theoretical models and experimental data indicate that within the transmembrane domains, the two α -repeat regions are oriented in opposite directions (Figure 1). In the case of NCX1, both α -repeats are suggested to form reentrant loop structures that, in the intact protein, come into close physical proximity with one another and form part of the ion-binding and translocation pathway. Studies on NCKX2 have confirmed

a similar structural arrangement, although there is no clear evidence for reentrant loops in either α -repeat in the NCKX proteins.

Recent structural studies on the NCX1 cytoplasmic loop have defined two homologous domains (corresponding to the two β -repeat regions) each comprising an immunoglobulin-like fold with Ca^{2+} bound by the loops at one end of the β -sandwich in a manner reminiscent of C2 domains found in some membrane-associated Ca^{2+} -binding proteins. These two Ca^{2+} -binding domains, together with other regions of the cytoplasmic loop, are responsible for mediating regulation of NCX1 activity (see below). Alternative splicing changes the sequence and properties of the second of the Ca^{2+} -binding domains, resulting in NCX1 proteins with different regulatory properties.

In contrast to NCX1, the large cytoplasmic loop of the NCKX family has not been investigated in any detail. NCKX1, but not the other family members, has an unusually long loop as a consequence of a simple tandem repeat structure. Though this motif is present in NCKX1 molecules from various animal species, the sequence and extent of the repeat are different in each. Several of the NCKX family members are subject to alternative splicing that changes a short region toward the N-terminal end of the cytoplasmic loop. The consequences of these differences on transport function or its regulation are currently not known.

Functional Properties

As their name implies, all $\text{Na}^+/\text{Ca}^{2+}$ exchangers couple the movement of Na^+ ions in one direction to Ca^{2+} ions in the other. An extensive collection of experimental data supports a simple model where the transporter has a single set of ion-binding and transport sites that can be occupied by either Na^+ ions or a Ca^{2+} ion (and in the case of NCKX exchangers, a K^+ ion as well), but cannot be occupied by both Na^+ and Ca^{2+} ions simultaneously. The ion-binding sites on the exchanger must be fully occupied in order for their aqueous accessibility to shift across the membrane barrier (Figure 3). The direction of net ion flux is determined by the relative electrochemical ion gradients across the membrane, and the number of ions that bind to the transporter. The exchanger can, thus, catalyze half cycles of Na^+/Na^+ exchange or $\text{Ca}^{2+}/\text{Ca}^{2+}$ (or in the case of NCKX, $(\text{Ca}^{2+}+\text{K}^+)/(\text{Ca}^{2+}+\text{K}^+)$) exchange, as well as the full cycle of $\text{Na}^+/\text{Ca}^{2+}$ exchange. For NCX1, the ion-binding stoichiometry has been a subject of intense and controversial investigation, but most studies suggest 3 Na^+ :1 Ca^{2+} . For NCKX1 and NCKX2, the stoichiometry has been established as 4 Na^+ :1 Ca^{2+} + 1 K^+ . The stoichiometry of the other NCX and NCKX proteins has not been determined, but is likely to be similar to their paralogous family members.

Although the direction of flux is dictated by thermodynamic considerations, the magnitude of the flux is determined by kinetic and regulatory factors. For NCX1, which has a very high turnover rate, the enzyme is normally limited by the concentration of cytosolic Ca^{2+} , which at rest is well below the concentration for half-maximal occupancy of the transport sites. NCX1 is, thus, activated only after cytosolic $[\text{Ca}^{2+}]$ is elevated by a signaling event. NCX1 activity is also regulated in a complex fashion through various domains of the central

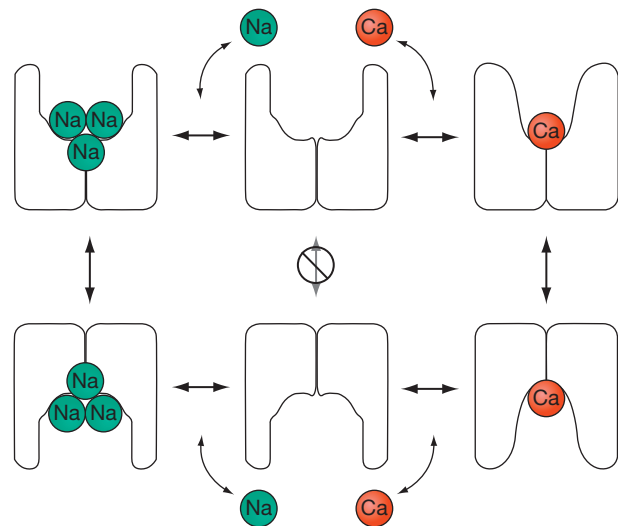


Figure 3 The $\text{Na}^+/\text{Ca}^{2+}$ exchanger reaction cycle. A simple six-state model is shown, where either three Na^+ ions or a single Ca^{2+} ion can bind to the transport site from either side of the membrane in a competitive fashion. Only fully loaded transporter can reorient the aqueous accessibility of the ion-binding sites across the membrane. Note that, in the case of $\text{Na}^+/\text{Ca}^{2+}+\text{K}^+$ exchanger, four Na^+ ions bind instead of three, and K^+ binds and is transported together with Ca^{2+} .

cytosolic loop. Intracellular Ca^{2+} activates the exchanger, while a high concentration of Na^+ causes inactivation. Acidic phospholipids, and adenosine triphosphate through phosphorylation of phosphatidylinositol, also activate the exchanger. Modulation of various kinase pathways has a modest effect on $\text{Na}^+/\text{Ca}^{2+}$ -exchange activity, but there is little convincing evidence for direct phosphorylation of NCX1. In addition, a variety of interaction partners have been proposed to regulate NCX1 activity. Though not studied in nearly as much detail, it appears that the other NCX family members are subject to a similar range of regulatory phenomena.

The rod photoreceptor exchanger, NCKX1, appears to turn-over more slowly than NCX1. On the other hand, NCKX1 has a higher apparent affinity for cytosolic Ca^{2+} and should theoretically be able to drive Ca^{2+} down to very low levels. That this does not happen in rod photoreceptors, suggests that other kinetic or regulatory factors may influence the activity of the molecule. One good candidate for such regulation is the rod-cyclic nucleotide-gated channel, which interacts functionally via Ca^{2+} cycling and also physically associates with NCKX1. Analysis of NCKX function in brain neurons confirms expectations from recombinant systems and reveals a high-capacity transport system that contributes to Ca^{2+} homeostasis when intracellular $[\text{Ca}^{2+}]$ levels are elevated above baseline levels. Recent studies have suggested that, analogous to NCX1, NCKX2 is subject to a Na^+ -dependent inactivation process involving occupancy of cytoplasmically disposed Na^+ -transport sites. NCKX2, but not NCKX3 or NCKX4, is stimulated by protein kinase-C in a complex manner that appears to involve phosphorylation of three serine/threonine residues residing in separate cytoplasmic locations. The mechanism that leads from phosphorylation to activation of transport in NCKX2 has not yet been delineated.

Physiological Relevance

In the heart, it is clearly established that NCX1 plays a critical role in relaxation by lowering Ca²⁺ during diastole. Additionally, during the early phase of the cardiac action potential, as the membrane depolarizes and ion gradients change, NCX1 probably contributes to Ca²⁺ entry and excitation–contraction coupling. Furthermore, as NCX1 switches to Ca²⁺ extrusion later in systole, it creates a depolarizing current that can significantly modulate the shape of the action potential. Normally, these contributions of NCX1 to cardiac excitation–contraction coupling are integrated into overall Ca²⁺ homeostasis. However, during pathological circumstance, the expression pattern of Ca²⁺ channels and transporters changes, and ion concentrations and gradients are altered dramatically. Under these conditions, changes in NCX1 activity may be maladaptive, and so may contribute to, rather than protect against, deleterious consequences. Recent studies using mice harboring a cardiac-specific knockout of NCX1 have shed light on several of these intriguing issues.

In the kidney, very high levels of NCX1 are found on the basolateral aspect of a selected population of distal nephron cells. In this environment, NCX1 contributes to the control of extracellular, systemic, Ca²⁺ homeostasis by driving the active transcellular reabsorption of Ca²⁺. NCX1 is also present in brain neurons and astrocytes, where it presumably plays a role in Ca²⁺ extrusion under resting conditions, but may participate in Ca²⁺ entry under certain circumstances. Recent evidence has suggested that expression of NCX1 in vascular smooth muscle cells plays a central role in regulating vessel tone and may contribute to susceptibility to salt-sensitive hypertension.

The unique and specialized role that NCX1 plays in controlling Ca²⁺ flux and homeostasis in heart and kidney is also reflected in the fact that this gene is expressed under the control of three different, and tissue-specific, promoters. One promoter is used essentially exclusively in heart, one exclusively in kidney, while the third is used widely in various other tissues. This arrangement provides the opportunity for expression of the heart and kidney NCX1 molecules to be regulated independently and to be responsive to the specialized Ca²⁺ homeostasis requirements of these tissues.

The distinct contributions provided by the NCX2 and NCX3 proteins have recently been investigated using gene-targeted knockout models. NCX2 was demonstrated to have an important stabilizing contribution to Ca²⁺ homeostasis in neuronal processes at both pre- and postsynaptic sites. Inactivation of NCX2 resulted in increased hippocampal plasticity and enhancements of memory and learning. NCX3 knockout animals, by contrast, displayed defects in neuromuscular transmission and increased muscle necrosis compared to wild-type counterparts.

In the rod photoreceptors, NCKX1 is thought to play a critical role in maintaining Ca²⁺ homeostasis, especially in

the dark when cyclic-nucleotide-gated channels are open and so the membrane is depolarized, the Na⁺ gradient is low, and Ca²⁺ flux is high. Under these circumstances, the thermodynamic power derived from the 4 Na⁺:1 Ca²⁺ + K⁺ coupling ratio is necessary to allow continued Ca²⁺ efflux. It would seem reasonably likely that similar requirements for continued Ca²⁺ efflux in the face of depleted ion gradients, reduced membrane potential, and elevated cytoplasmic [Ca²⁺] – for example, during high-frequency action potential firing in a neuron – underlie the expression of NCKX isoforms at various cellular and tissue sites. The specific roles for the different NCKX isoforms are now beginning to be resolved through analysis of mice models with specific gene-targeted knockouts. Recent data demonstrate that NCKX2 has a nonredundant role in supporting normal hippocampal synaptic plasticity underlying a set of experience-dependent motor learning and memory behaviors.

In a recent exciting discovery, NCKX5 was demonstrated to be essential for normal pigmentation in zebrafish. Importantly, a nonsynonymous single-nucleotide polymorphism within a conserved part of the NCKX5 protein-coding region is associated with a major quantitative contribution to skin pigmentation in human populations. The resulting nonconservative substitution of Thr for Ala at position 111 is associated with lighter skin color in people of Northern-European ancestry. Intriguingly, NCKX5 function has not yet been measured, and how the A111T polymorphism leads to changes in pigmentation still remains to be elucidated.

The distinct individual physiological roles, as well as the underlying molecular mechanisms, for the expanding family of NCX and NCKX protein molecules continues to be an active area of research that is likely to provide exciting novel insights into Ca²⁺ homeostasis and its dysregulation in disease.

See also: **Bioenergetics: Sodium–Calcium Exchanger: Structural Aspects.**

Further Reading

- Berridge MJ, Bootman MD, and Roderick HL (2003) Calcium signalling: Dynamics, homeostasis and remodelling. *Nature Reviews. Molecular Cell Biology* 4: 517–529.
- Bers DM (2002) Cardiac excitation–contraction coupling. *Nature* 415: 198–205.
- Blaustein MP and Lederer WJ (1999) Sodium/calcium exchange: Its physiological implications. *Physiological Reviews* 79: 763–854.
- Nicoll DA, Longoni S, and Philipson KD (1990) Molecular cloning and functional expression of the cardiac Na, Ca-exchanger. *Science* 250: 562–565.
- Pott C, Henderson SA, Goldhaber JL, and Philipson KD (2007) Na⁺/Ca²⁺ exchanger knockout mice: Plasticity of cardiac excitation–contraction coupling. *Annals of the New York Academy of Science* 1099: 270–275.
- Quednau BD, Nicoll DA, and Philipson KD (2004) The sodium/calcium exchanger family-SLC8. *Pflügers Archiv European Journal of Physiology* 447: 543–548.
- Reiländer H, Achilles A, Friedel U, et al. (1992) Primary structure and functional expression of the Na/Ca, K-exchanger from bovine rod photoreceptors. *EMBO Journal* 11: 1689–1695.
- Visser F and Lytton J (2007) K⁺-dependent Na⁺/Ca²⁺ exchangers: Key contributors to Ca²⁺ signaling. *Physiology (Bethesda)* 22: 185–192.

Membrane Transport, General Concepts

S G Schultz, University of Texas Medical School, Houston, TX, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 627–630, © 2004, Elsevier Inc.

Glossary

Active transport Transport of a solute against a chemical (concentration) and/or electrical potential difference that is directly (primary) or indirectly (secondary) driven by metabolic energy.

Carrier Integral proteins that span the membrane phospholipid bilayer and are capable of binding a solute(s) and catalyzing translocation from one side of the membrane to the other.

Channel Integral proteins that span the membrane phospholipid bilayer and provide water-filled passageways (pores) for the diffusion of solutes.

Diffusion Flow of a solute from a region of higher to a region of lower chemical (concentration) and/or electrical potential driven solely by thermal random motion of solute particles.

Facilitated diffusion Carrier-mediated transport processes that are not coupled to a supply of metabolic energy and, thus, can only result in the equilibration of a transport solute across the membrane.

The term 'membrane transport' refers to processes that bring about the movement of solutes and water across the barrier, or envelope that surrounds all living cells in the animal and plant kingdoms. These processes are responsible for two ubiquitous characteristics of living cells, namely: (1) the ability to maintain constant, or near constant, intracellular compositions that differ from the extracellular environment, containing the highly specialized molecules essential for metabolism and replication and, at the same time, (2) the ability to selectively exchange matter and energy with that environment.

All cell membranes consist of a double layer of phospholipids oriented so that their electrically charged, hydrophilic (or water-loving) head projects into the interior and exterior of the cells, respectively, and their lipid tails, which are hydrophobic (or water-fearing), project inward and form an oily core. Floating in this oily structure are surface or peripheral proteins that do not extend through the thickness of the bilayer, and integral proteins that span the bilayer and provide pathways for direct communication between the intracellular and extracellular compartments. This lipoprotein bilayer model, which consists of large expanses of lipid studded here and there with protein, is illustrated in [Figure 1](#).

Exchange of solutes and water across the lipoprotein barrier takes place through the lipid as well as through integral proteins and can be classified as (1) simple diffusion; (2) diffusion through pores or channels; and (3) carrier-mediated transport. The latter may be subdivided into facilitated diffusion (or facilitated transport), primary active transport, and secondary active transport.

Simple Diffusion

The simplest form of membrane transport was recognized more than a century ago by Overton, who noted that the ease with which a large variety of molecules cross plant cell membranes is proportional to their solubility in lipids; indeed, it is the observation that first suggested that the membranes

were composed of lipids. In essence, he suggests that the molecule simply dissolves in the lipid across one surface, passes through this oily barrier, and exits at the other side. The rate of this process is given by a modification of a law derived by Fick in 1855:

$$J_i = k_i D_i \Delta C_i$$

where J_i is the rate of diffusion of solute i across the membrane (in amount of solute per unit time), k_i a measure of the solubility of the solute i in oil compared to water (oil:water partition coefficient), D_i the diffusion coefficient of i or the mobility of i in the lipid, and ΔC_i the concentration difference of i across the membrane. The product $k_i D_i$ is often abbreviated as the permeability coefficient P_i .

Overton demonstrated that a linear relation between J_i and k_i in plant cells and this relation, now referred to as Overton's law, has been confirmed repeatedly.

Diffusion through Pores or Channels (Restricted Diffusion)

Lipid membranes are essentially impermeable to ions and are only slightly permeable to water. It is now clear that these highly hydrophilic molecules traverse cell membranes through integral proteins that form water-filled pores or channels that span the bilayer. Although the presence of holes in membranes had been speculated as early as the 1850s and were strongly suggested by the pioneering work of Hodgkin, Huxley, and Keynes on the squid axon in the 1950s, they were established beyond doubt, more recently, by Hille, Armstrong, and MacKinnon.

Although some use the terms pore and channel interchangeably, others prefer to employ the word pore for integral proteins that are continually open to the two aqueous compartments on both sides of the membrane (using the analogy of an open pipe) and reserve the word channel for a pore that has a gate that determines how long it will be in the fully open

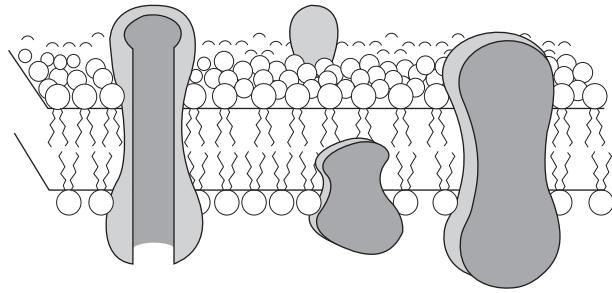


Figure 1 Illustration of phospholipid bilayer with peripheral proteins adhering to the surfaces and integral proteins spanning the thickness of the membrane. One integral protein is portrayed as a channel. The charged head groups of the phospholipids are represented as circles and the lipophilic fatty tail groups as the wavy lines forming the core of the bilayer.

state and how long it will be closed (the analogy being a pipe with shutter).

In general, then, channels have gates that determine the probability that they will be in the open state at any time (open-time probability) and how long they will stay open (mean open time). The properties of these gates may be influenced by the electrical potential difference across the membrane (voltage-gated channels), the binding of one or more ligands (ligand-gated channels) or both. In addition, they possess 'filters' that confer selectivity on the channel; the structural characteristics of ion channels that result in a high preference for one ion over the other have recently been unraveled for potassium channels.

Movement of an ion through an open channel obeys a form of Ohm's law for current flow caused by a conductance driven by an electrochemical potential difference. Thus,

$$i_i = g_i(V_m - E_i)$$

where i_i is the rate of flow of the ion through a single open channel given in electrical terms of amperes (this is analogous to J_i) and g_i the conductance of the channel (analogous to P_i). V_m is the electrical potential difference across the membrane and E_i the Nernst equilibrium potential for the ion so that the difference ($V_m - E_i$) is the displacement of the ion from equilibrium in electrical terms, and is thus the driving force for flow. More than one million ions may traverse a single channel per second so that single channel currents may be readily recorded using a variety of techniques. The ionic current through an ensemble of such channels is $I_i = i_i N_i p_o$, where N_i is the total number of channels and p_o the open probability.

Carrier-Mediated Transport

Carrier-mediated transport differs from diffusion through lipid phase, pores, or channels in that after entering the integral protein the transported solute(s) bind to a specific site(s) on this protein which then undergoes some translocation that exposes and/or releases them to the other side. Unlike pores or channels, carriers are not open to both sides of the

membrane simultaneously. Because conformational changes are necessary for carriers to function, the number of molecules transported per second is orders of magnitude less than pores. In this sense, carriers may be viewed as enzymes that do not chemically alter the transported species (substrate) but, instead, catalyze a change in their location. For this reason, the terms permeases and translocases are used in some instances as synonymous with carriers. Indeed, one characteristic of carrier-mediated transport is that it displays Michaelis–Menten saturation kinetics, marked by a maximum transport velocity (J_m) and a solute concentration needed to achieve a half-maximal transport velocity (K_t), and often obeys the simple relation:

$$J_s = \{[S]J_m/K_m + [S]\}$$

where J_s is the rate of transport and $[S]$ the concentration of the transported solute. (Note: this relation describes the initial rate of transport when there is solute present on one side of the membrane only.)

Facilitated Diffusion

The earliest recognized and simplest form of carrier-mediated transport is facilitated diffusion, often called facilitated transport, in which an otherwise impermeant solute binds to a site on an integral protein (carrier) from one side of the membrane and then undergoes a translocation that provides the solute access to the other side. The classic example of facilitated diffusion is glucose transport across the membranes of cells such as erythrocytes, muscle, and adipocytes. The carriers that mediate this transport have been cloned and sequenced and fall into a group of proteins that have 12 membrane-spanning segments called glucose transporter (GLUT). Urea is transported across the membranes of many cells by a facilitated transporter called UT.

It must be emphasized that because these carriers are not directly or indirectly coupled to a supply of energy, they cannot perform osmotic work, that is, they cannot transport a neutral solute against a concentration difference (e.g., from lower to higher concentration) or a charged solute against a combined electrochemical potential difference. Thus, like diffusion, facilitated diffusion carriers bring about equilibration; transport ceases when intrinsic thermodynamic driving forces are abolished.

Active Transport

Active transport is the term reserved for transport processes that result in the movement of a solute uphill or against its natural direction. For the case of a neutral solute (at constant temperature and pressure), this resolves into movement against a concentration difference; for a charged solute, it is movement against the combination of concentration and electrical potential differences. This is work (e.g., charging a battery) and it therefore requires the investment of energy.

There are two groups of carrier-mediated active transport processes: those where the coupling to energy is direct and those where the coupling to energy is indirect.

Primary Active Transport

Primary active transport processes are capable of coupling energy directly to the uphill movement of the transported species. By far the most abundant, and well-understood, primary active transporters are the retinylidene proteins or, more commonly called, rhodopsins, capable of coupling light energy to ion transport. Bacteriorhodopsins found in *Archaea*, *Cirrhacea*, and some eukaryotic and prokaryotic cells pump protons out of these cells, thereby establishing a 'proton-motive force' across the membrane. Halophilic bacteria possess a form referred to as halorhodopsin that pumps Cl^- into cells. These transporters have been studied extensively using high-resolution crystallography and all possess seven membrane-spanning segments attached to a chromophore.

In higher animals, all primary active transporters are ATPases, that derive the energy for moving the transported solute against a thermodynamic driving force (i.e., uphill) from adenosine triphosphate (ATP) hydrolysis. The most prevalent is the Na-K pump found in virtually all cells from higher animals and responsible for pumping three Na^+ ions out of the cell and two K^+ ions into the cell for every ATP hydrolyzed. This pump is responsible for the fact that these cells have a lower intracellular Na^+ concentration and higher intracellular K^+ concentration than the concentrations of these ions in the surrounding extracellular fluid. As is discussed below, the extrusion of Na^+ from the cell establishes a 'natrio-motive' force across the membrane analogous to the proton-motive force established by light-driven proton pumps in lower life forms. The detailed mechanism of action of the Na-K pump is not well established except for the fact that it is a member of a superfamily of ATPases referred to as $\text{E}_1\text{-E}_2$ ATPases or P-type ATPases, because the operation of the pump involves cycling between two configurations of the protein driven by phosphorylation and dephosphorylation reactions.

Other ATPases that perform primary active transport include the Ca^{2+} ATPase found in muscle and responsible for regulation of the Ca^{2+} activity of the cytoplasm, the $\text{H}^+\text{-K}^+$ ATPase of gastric parietal cells responsible for secreting protons into the gastric juice in response to appropriate stimuli, and the H^+ ATPase that is responsible for acidifying intracellular organelles such as lysosomes and endosomes.

Secondary Active Transport

Secondary active transport processes bring about the uphill transport of one or more solutes but do not derive the necessary energy from direct coupling to a source of metabolic energy. Instead, these transport processes are energized by coupling those movements to the downhill movement of another solute.

Co-Transport (Symport)

The uphill or secondary active transport of a large number of solutes into cells of higher animals is coupled to the downhill

movement of Na^+ . Thus, the carrier protein binds both the transported solute and Na^+ from the extracellular solution. Inasmuch as the Na^+ concentration in the intracellular solution is lower than that in the extracellular solution, because of the action of the $\text{Na}^+\text{-K}^+$ pump as discussed, this coupling permits the downhill movement of Na^+ and provides the energy for the uphill movement of the coupled solute. In essence, the energy invested into the $\text{Na}^+\text{-K}^+$ pump by the hydrolysis of ATP has set up a Na^+ -gradient or natrio-motive force across the membrane that can be deployed to energize the secondary active transport of solutes whose translocation across the membrane is coupled to that of Na^+ .

In many lower forms of life, secondary active co-transport is energized by the H^+ gradient or proton-motive force across the membrane established and maintained by light-driven primary active extrusion of protons, as discussed above.

Counter-Transport (Antiport)

Coupling to the downhill movement of Na^+ into cells can also serve to energize the uphill or secondary active transport of solutes out of cells. Important examples of counter-transporters are the $\text{Na}^+\text{-H}^+$ exchanger responsible for extruding H^+ from cells and the regulation of cell pH and the $\text{Na}^+\text{-Ca}^{2+}$ exchanger responsible for extruding Ca^{2+} from many cells.

Na^+ -coupled co- and counter-transport processes are very diverse and are beautiful examples of bioenergetic economy where a single primary active transport process directly coupled to the hydrolysis of ATP can establish a transmembrane ion (Na^+) gradient that can, in turn, be utilized to energize a wide variety of secondary active transport processes.

See also: **Bioenergetics:** ER/SR Calcium Pump: Function; ER/SR Calcium Pump: Structure; Membrane Transporters: $\text{Na}^+/\text{Ca}^{2+}$ Exchangers; Plasma-Membrane Calcium Pump: Structure and Function; P-Type Pumps: Copper Pump; P-Type Pumps: H^+/K^+ Pump; P-Type Pumps: Na^+/K^+ -ATPase; P-Type Pumps: Plasma-Membrane H^+ Pumps.

Further Reading

- Bretscher MS (1985) The molecules of the cell membrane. *Scientific American* 253(4): 100–108.
- Doyle DA, Cabral JM, Pfuetzner RA, et al. (1998) The structure of the potassium channel: Molecular basis of K^+ conductance and selectivity. *Science* 280: 69–77.
- Hille B, Armstrong CM, and MacKinnon R (1999) Ion channels: From idea to reality. *Nature Medicine* 5: 1105–1109.
- Schultz SG (1982) *Basic Principles of Membrane Transport*. Cambridge: Cambridge University Press.
- Schultz SG (2003) Membrane transport. In: Johnson LR (ed.) *Essential Medical Physiology*, 3rd edn., pp. 37–70. San Diego, CA: Academic Press.
- Spudich JL, Yang C-S, Jung K-H, and Spudich EN (2000) Retinylidene proteins: Structures and functions from Archaea to humans. *Annual Review of Cell and Developmental Biology* 16: 365–392.

Messenger RNA Degradation in Bacteria

D H Bechhofer, Mount Sinai School of Medicine, New York, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Endonuclease Enzyme that cleaves an internal phosphodiester bond.

Exonuclease Enzyme that cleaves a terminal phosphodiester bond, at a 5' end or a 3' end.

Oligoribonuclease Enzyme that degrades short oligonucleotides to mononucleotides.

Phosphorolysis Cleavage of a nucleotide bond by addition of inorganic phosphate, rather than the more common hydrolysis, which adds water.

Poly(A) polymerase Template-independent enzyme that adds A nucleotides at an RNA 3' end.

Transcription termination The process by which RNA polymerase is removed from its DNA template and RNA synthesis ceases.

Transfer messenger RNA (tmRNA) The RNA that binds to a stalled ribosome and transfers it to its own coding sequence for renewed translation. The peptide that is generated is marked at its C-terminal end for degradation by proteases.

General Aspects of Bacterial Messenger RNA Decay

Regulation of gene expression occurs at many levels, including the fate of a messenger RNA (mRNA) after its synthesis. In eukaryotes, regulation of gene expression at this posttranscriptional level includes all activities that affect the ability of an initial transcription product to function in translation, such as capping, splicing, polyadenylation, export to the cytoplasm, and degradation. In prokaryotes, self-splicing occurs rarely, transcription and translation are not separated by a membrane barrier, and polyadenylation occurs but to a much lesser extent. The study of posttranscriptional gene regulation in prokaryotes is therefore concerned mainly with mRNA degradation.

Bacterial mRNA half-lives are measured by adding an inhibitor of bacterial RNA polymerase, most often rifampin, and following the decline in the level of a specific mRNA over time by Northern blot analysis or reverse transcription-polymerase chain reaction, or following decay of mRNA globally in a microarray experiment. The half-lives of bacterial mRNAs are generally only a few minutes, unlike the much longer-lived eukaryotic mRNA. Rapid turnover of bacterial mRNA is likely an important factor in the ability of bacteria to respond to changes in their environment and nutritional sources. Thus, the degradation of an mRNA will free up ribosomes to translate newly synthesized mRNAs that are induced by such changes. In addition, mRNA degradation releases free nucleotides to the pool from which new nucleoside triphosphates are made and then used in RNA synthesis. As such, decay of mRNA is an essential process, required for life. Nevertheless, the range of bacterial mRNA half-lives varies from <1 min to >30 min. The half-life of a particular mRNA will be a function of susceptibility to ribonuclease attack, which is determined by its specific sequence/structure and its interaction with elements of the translational apparatus and other proteins or small RNAs.

The process of mRNA decay is divided into two stages: (1) rate-determining initiation of decay and (2) complete turnover of the mRNA to nucleotides. The most efficient way to remove an mRNA from the set of translated messages would be to make the mRNA unrecognizable to ribosomes by degrading the sequence that constitutes the 5'-proximal

ribosome-binding site (RBS). This could be accomplished by exoribonucleolytic decay from the 5' end. However, a 5'-to-3' exoribonuclease is not known to exist in *Escherichia coli*, and, until very recently, it was thought that the absence of such an enzyme in *E. coli* meant that bacteria generally do not contain a 5' exoribonuclease. In eukaryotes, by contrast, decapping enzymes and 5' exoribonucleases are widespread. Because of the absence of a 5'-to-3' exonuclease, the hypothesis developed that initiation of bacterial mRNA decay occurs by an internal endonucleolytic cleavage, followed by rapid decay of the upstream and downstream products of such cleavage. This mechanism is reflected by the fact that Northern blot analysis of an mRNA in a wild-type bacterial strain will usually detect only the full-length mRNA and no decay intermediates. Strains that are deficient in one or more ribonucleases are required to detect intermediates in the decay process.

Surprisingly, over a decade after complete genome sequences of well-studied organisms began appearing, new enzymes that participate in mRNA decay are still being discovered. We focus on the ribonucleases of two representative eubacteria: *E. coli*, the model Gram-negative organism, and *Bacillus subtilis*, the model Gram-positive organism. The overwhelming majority of studies on bacterial mRNA decay have been done in these two organisms, and the complement of ribonucleases (Table 1) and the mechanism of mRNA decay differ significantly between them.

Ribonucleases Activities

RNase E

The essential endoribonuclease of *E. coli* is named RNase E. RNase E-related enzymes are present in many bacterial species and are thought to be the primary decay-initiating enzyme in these organisms. RNase E is a 5'-end-dependent endonuclease, that is, it binds to the 5' end of an mRNA and then locates a target site at which it cleaves endonucleolytically. The activity of RNase E is at least an order of magnitude more active on an mRNA with a 5'-monophosphate end than the 5'-triphosphate end that is present in the initial transcript. This feature ensures that an mRNA is not attacked immediately as RNA polymerase

Table 1 Ribonucleases involved in mRNA decay in *E. coli* and *B. subtilis*

Ribonuclease	Gene name	Cleavage specificity	<i>E. coli</i>	<i>B. subtilis</i>	Characteristics
RNase E	<i>rne</i>	Endonuclease	Yes	No	5'-End-dependent, inhibited by 5' triphosphate end
RNase G	<i>rng</i>	Endonuclease	Yes	No	Some functional overlap with RNase E
RNase II	<i>rnb</i>	3' Exonuclease	Yes	No	Major turnover enzyme in <i>E. coli</i>
RNase III	<i>rnc</i>	Endonuclease	Yes	Yes	Recognizes double-stranded structure
RNase J1	<i>rnjA</i>	Endonuclease 5' Exonuclease	No	Yes	Endonuclease activity insensitive to 5' phosphorylation state; exonuclease activity requires 5' monophosphate end
RNase J2	<i>rnjB</i>	Endonuclease	No	Yes	Forms a complex with RNase J1; poor 5' exonuclease activity
RNase R	<i>rnr</i>	3' Exonuclease	Yes	Yes	Degrades through secondary structure
RNase Y	<i>rny</i>	Endonuclease	No	Yes	Prefers 5'-monophosphate end
PNPase	<i>pnp</i>	3' Exonuclease	Yes	Yes	Phosphorolytic; major turnover enzyme in <i>B. subtilis</i>

begins synthesizing it. Indeed, recent studies have shown that an early step in the decay of a number of *E. coli* mRNAs is the conversion of the 5'-triphosphate nucleoside to a 5'-monophosphate nucleoside by the action of a phosphohydrolase enzyme. Thus, accessibility of the 5' end is a major determinant of mRNA half-life. For example, the formation of strong secondary structure at the 5' end by base-pairing of 5'-proximal nucleotides will stabilize diverse downstream sequences. The internal target site for RNase E cleavage is in a single-stranded segment that is AU rich, often located near a region of RNA secondary structure. How a 5'-binding enzyme finds its internal target site is a matter of intense speculation. The mechanism could involve looping out of uncleaved mRNA and positioning of the target sequence in the enzyme catalytic site. As RNase E is thought to exist mostly as a tetramer, 5'-end binding could coordinate structural changes of monomers in the complex to position the target site appropriately. In some cases, RNase E can locate and cleave at an internal target site independent of binding at the 5' end (direct access).

Starting mRNA decay process by RNase E cleavage could be a dangerous strategy, if the resulting upstream fragment was not degraded rapidly. A ribosome could still bind and initiate translation of this fragment, leading not only to synthesis of a nonfunctional peptide but also to a ribosome stuck at the 3' end of the fragment with no stop codon to signal its release. Although bacteria do have the specialized transfer messenger RNA (tmRNA) system (see below) that allows off-loading of ribosomes from nonstop mRNA fragments (i.e., mRNA fragments that contain an RBS and an open reading frame but no stop codon), the capacity of this system is limited. In fact, the products of RNase E cleavage do not persist due to the nature of their ends (Figure 1(a)). The upstream fragment that is generated by RNase E has a 3'-hydroxyl end, and is rapidly degraded by the action of 3' exonucleases: RNase II, polynucleotide phosphorylase (PNPase), and RNase R (Table 1). There are several other *E. coli* 3' exonucleases, but they are involved in processing of stable RNAs. Thus, as soon as RNase E makes an endonucleolytic cut, the upstream fragment bearing the RBS is on its way to being converted to mononucleotides. The downstream mRNA fragment generated by RNase E cleavage bears a 5'-monophosphate end, and this is an excellent substrate for further RNase E binding and cleavage. Accordingly, the initial cleavage by RNase E is the trigger for a rapid process that leaves no long-lasting intermediates (Figure 1(a)).

Degradosome

An interesting aspect of RNase E is that it nucleates a protein complex known as the RNA degradosome. The N-terminal half of the RNase E protein, which is strongly conserved in other species, contains the catalytic domain. Many bacteria also contain a smaller ribonuclease known as RNase G, which resembles the N-terminal half of RNase E and has similar endonucleolytic activity. The fact that RNase E is essential suggests that RNase G cannot substitute for it *in vivo*. The C-terminal half of RNase E is less well conserved and contains an arginine-rich RNA-binding domain, as well as several protein sites that are docked by three other proteins: (1) RhlB, an RNA helicase that unwinds double-stranded RNA; (2) the aforementioned 3'-exonuclease PNPase; and (3) enolase, an enzyme involved in glycolysis. The C-terminal region of RNase E thus serves as a scaffold for the protein complex that constitutes the degradosome. Clearly, RNase E, RhlB, and PNPase could function in concert to accomplish mRNA decay, with RNase E making an initial cut, PNPase degrading the upstream fragment in the 3'-to-5' direction, and RhlB unwinding helices ahead of PNPase (see below). The presence of a degradosome complex, however, is not essential, as *E. coli* cells that carry only the N-terminal half of RNase E are viable. The existence of a degradosome complex has been established for a number of bacterial species, but remains controversial in others, including *B. subtilis*.

3'-to-5' Exoribonucleases

We mentioned the activity of 3'-to-5' exoribonucleases in degrading the upstream product of endonucleolytic initiation of decay. Most bacterial species contain more than one 3'-to-5' exonuclease that degrades mRNA fragments. In *E. coli*, the major exoribonucleases that perform this function are RNase II and PNPase, and a strain that is missing both of these activities is inviable. While RNase II and most exoribonucleases are hydrolytic enzymes, which add water across a phosphodiester bond, PNPase is unusual in that it is a phosphate-dependent enzyme that carries out phosphorolysis, which releases a nucleoside 5' diphosphate (Figure 2). It has been argued that bacteria that live in energy-poor environments may rely more heavily on PNPase activity for RNA decay, as this preserves the phosphate bond energy in the decay products. Interestingly, *B. subtilis*, which lives in energy-poor

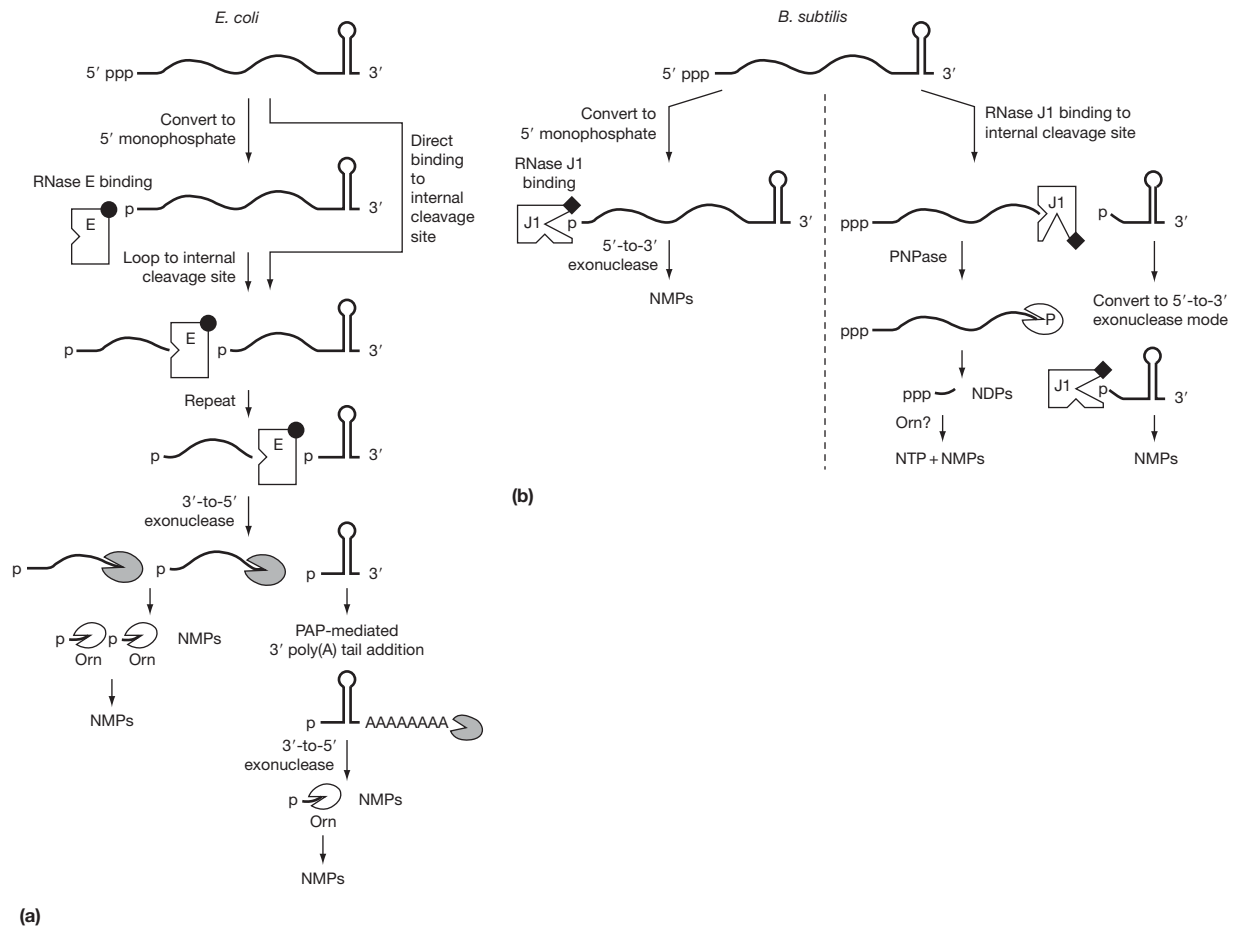


Figure 1 Models for mRNA decay in (a) *E. coli* and (b) *B. subtilis*. Alternative pathways are shown for *B. subtilis*. Shaded 3'-to-5' exonucleases in (a) represent RNase II or PNPase. The major 3'-to-5' exonuclease in (b) is PNPase (indicated by the P). Orn, oligoribonuclease; NMPs, nucleoside monophosphates; NDPs, nucleoside diphosphates.

soil (unlike *E. coli*, which lives in the energy-rich gut), contains PNPase but not RNase R. Another 3'-to-5' exonuclease that has been shown to be involved in mRNA decay is RNase R, which is unusual in being able to degrade through strong secondary structure, provided it is given a long-enough single-stranded tail for binding (optimally 10 nucleotides). According to recent models, the choice of which 3' exonuclease will degrade the mRNA fragments generated by endonuclease cleavage will depend on the intrinsic RNA structure. Unstructured RNAs will be substrates for any of the three enzymes. Structured RNAs will be substrates for PNPase in conjunction with RhlB, or for RNase R.

Because most bacteria contain multiple 3' exoribonucleases, the effect of knocking out a gene encoding a particular 3' exoribonuclease may be difficult to detect, and multiply mutant strains are often required to see the effects of a missing ribonuclease. Such is not the case for *B. subtilis*, in which PNPase is the dominant ribonuclease activity of its four known 3' exoribonucleases. This is seen in experiments that measure the steady-state level of decay intermediates in wild-type versus ribonuclease mutant strains (an example is shown in Figure 3). In strains that contain all four of the known 3' exonucleases or only PNPase, only the full-length RNA is

detected. In strains that do not contain PNPase, decay intermediates are detected (although much less so in the strain containing RNase R alone) at levels that exceed the amount of full-length RNA.

Despite the presence of multiple 3' exonucleases, initiation of decay from the native 3' end of mRNAs contributes little to the overall decay mechanism. This is because of the nature of the 3' end of mRNAs that are terminated in a factor-independent manner. Factor-independent termination of transcription relies on the formation of a structure by complementary GC-rich RNA sequences, which is followed by several U residues. Synthesis of such a sequence causes RNA polymerase to fall off the DNA template, leaving an RNA with a strong secondary structure very near the 3' end and only a short single-stranded tail. Such structures are intrinsically resistant to degradation by 3'-to-5' exoribonucleases, such as RNase II and PNPase, and the 3' single-stranded tail is too short for RNase R binding. There is another mechanism of transcription termination that relies on the RNA-binding protein Rho, which leaves unstructured 3' ends. However, termination that is not Rho dependent is much more prevalent, and most studies of bacterial mRNA decay deal with mRNAs that are terminated in a factor-independent manner.

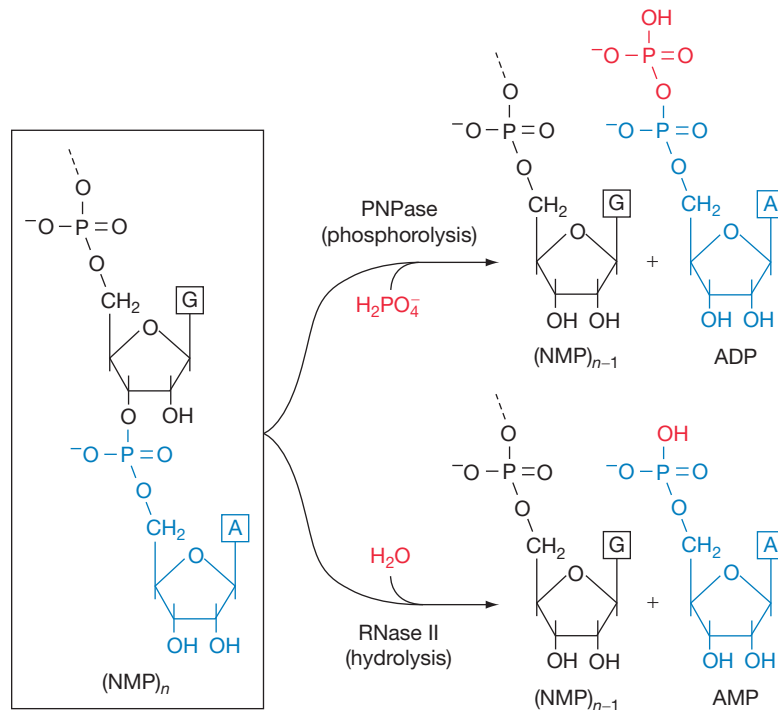


Figure 2 Phosphorolysis vs. hydrolysis. The boxed structure is the 3' end of an RNA chain of n residues, ending in guanosine and adenosine residues. The end nucleotide can be cleaved by the addition of inorganic phosphate (top) or water (bottom), generating a chain of $n-1$ residues and a nucleoside diphosphate (top) or a nucleoside monophosphate (bottom).

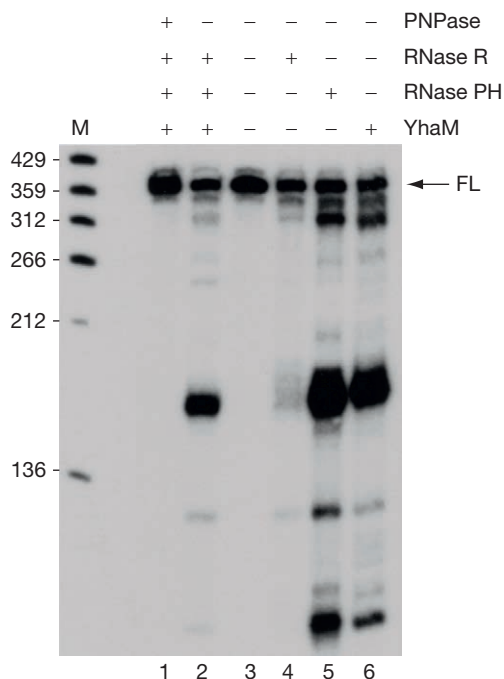


Figure 3 Northern blot analysis of accumulation of mRNA decay intermediates in *B. subtilis* ribonuclease mutants. Migration of the full-length (FL) mRNA is indicated by the arrow at right. Decay intermediates are observed neither in the wild-type strain (lane 1) nor in the strain containing only PNPase (lane 3). Marker lane (M) contains labeled DNA fragments of the indicated sizes.

PAP-mediated decay of 3'-end-containing fragments

The mechanism of *E. coli* mRNA decay described so far does not have a way of dealing with mRNA fragments that contain the 3'-terminal transcription termination structure, which would not be substrates for the single-strand-specific activity of RNase E. Rather, these fragments are degraded from their 3' end with the assistance of poly(A) polymerase (PAP), a DNA-independent RNA polymerase that adds A residues to a 3' end. Extension of the 3' end by poly(A) addition provides a toehold for 3' exonucleases to bind and begin degradation in the 3'-to-5' direction (Figure 1(a)). When the exonuclease proceeds a short distance but is blocked by secondary structure and falls off, PAP again adds a poly(A) extension and gives a 3' exonuclease another crack at degrading from the 3' end. This process occurs iteratively, until the 3'-proximal structure can no longer form, and the fragment is degraded.

RNase J Family

When completed bacterial genome sequences began accumulating in the late 1990s, it soon became clear that many species, including *B. subtilis*, had no sequence homolog of RNase E/G. Nevertheless, many studies showed that initiation of mRNA decay in these organisms was also 5'-end dependent. Uncertainty about decay initiation in *B. subtilis* began to resolve in 2005, when the first RNase J enzymes were identified. *B. subtilis* contains two RNase J enzymes, the essential RNase J1 and the nonessential RNase J2, which are similar in size and share 49% identical residues. In a strain with no RNase

J2 and a fivefold lower RNase J1 activity than wild type, the level of hundreds of mRNAs is increased or decreased significantly. Most bacterial species have either an RNase E or an RNase J family enzyme, but not both. RNase J1 of *B. subtilis* is, so far, the most well-studied RNase J family member. Like RNase E, RNase J1 is believed to be a 5'-end binding enzyme. However, unlike RNase E, the endonuclease activity of RNase J1 is effective on substrates with either 5'-monophosphate or 5'-triphosphate ends.

RNase J1 is also a 5'-to-3' exoribonuclease

The finding of RNase J1 endonuclease activity, which may be responsible for initiation of decay in *B. subtilis*, was followed (in 2007) by the remarkable discovery that RNase J1 is not only an endonuclease but also a 5' exonuclease, the very activity that was thought not to exist in bacteria. Thus, RNase J1 is a dual-acting ribonuclease, with both endonuclease and exonuclease activities – a rare enzyme in biology. Unlike the endonuclease activity of RNase J1, which is insensitive to the 5'-end phosphorylation state, the exonuclease activity of RNase J1 absolutely requires a 5'-monophosphate end. The reason for this restriction is clear from the crystal structure of an RNase J enzyme, which shows a 5'-end binding site that can accommodate a monophosphate (but not a triphosphate) nucleoside and that sits a one-nucleotide distance away from the catalytic site. This arrangement perfectly explains the 5'-exonuclease activity of RNase J1 and its sensitivity to 5'-end phosphorylation. However, at this time, it is unclear how RNase J1 endonuclease activity is achieved, as biochemical experiments indicate that the identified catalytic site is used for both exonucleolytic and endonucleolytic cleavage.

Other Endoribonucleases

RNase III is an endonuclease that recognizes specific double-stranded RNA structures. RNase III family enzymes are ubiquitously expressed in bacteria and even in higher eukaryotes, where the RNA_i enzyme known as dicer is an RNase III family member. RNase III's major function is in processing of stable RNAs, and it does not appear to play a significant role in the initiation of mRNA decay, other than a single known example of a particular operon mRNA and the examples of ribonuclease gene regulation cited below.

A very recent discovery in *B. subtilis* is the essential endonuclease RNase Y. A mutant *B. subtilis* strain, which contains lower than normal levels of RNase Y, shows a twofold increase in global mRNA half-life, indicating a significant impact of this enzyme on mRNA decay. It is expected that future studies will reveal an important role for this new enzyme in mRNA decay, in the diverse bacterial species that carry it.

Oligoribonuclease

3' Exoribonucleases can digest long polyribonucleotides in a processive manner, that is, the enzyme binds to the 3' end of the RNA and catalyzes phosphodiester bond cleavage to liberate successive end-nucleotides without releasing the substrate. As the length of the substrate becomes shorter than approximately 12 nucleotides, the reaction becomes distributive, that is, the enzyme releases the substrate and randomly binds a

3' end. Chains that are only 2–4 nucleotides long are no longer substrates for 3'-exoribonuclease activity, and they accumulate in *in vitro* reactions. Such accumulation could not occur *in vivo*, however, where RNA substrates need to be turned over completely to mononucleotides in order to maintain the ribonucleotide pool that is used for new RNA synthesis. Thus, bacteria have a specialized, essential 3' exonuclease called oligoribonuclease, which is able to digest very short oligonucleotides to completion (see Figure 1).

Comparison of Models for mRNA Decay

The existence of a 5'-exonuclease activity in *B. subtilis* has stimulated new models for initiation and complete turnover of an mRNA. Figure 1 shows a comparison of models for mRNA decay in *E. coli* (discussed in detail earlier) versus *B. subtilis* and other RNase J-containing organisms. In the latter case, decay can initiate either by an endonuclease cleavage (Figure 1(b)), right), or by attack from the 5' end (Figure 1(b)), left) if the 5' phosphate has been converted from a triphosphate to a monophosphate end. In addition, endonucleolytic cleavage can be followed by 5'-exonuclease activity that degrades the downstream fragment, including the 3' terminus. It is known that the 5' exonuclease activity of RNase J1 can degrade through secondary structure. It has been hypothesized that the exonuclease activity of RNase J1 is stimulated by prior endonuclease activity. Thus, RNase J1 endonuclease cleavage may be followed by rapid degradation of the downstream product without the release of the substrate. The near future will see more studies in *B. subtilis* aimed at clarifying whether the 5'-exonuclease activity or the endonuclease activity of RNase J1 is essential, and whether RNase Y plays a major role, perhaps even more important than RNase J1.

Regulation of RNase Expression and Activity

As control of mRNA stability is an important element in gene expression, it stands to reason that the activity of enzymes involved in mRNA decay may be regulated. In fact, the key *E. coli* endonuclease, RNase E, autoregulates its expression quite tightly. When the cellular concentration of RNase E increases, the enzyme cleaves *rne* mRNA, resulting in unstable *rne* mRNA and less RNase E translation. Similarly, the narrow-specificity RNase III enzyme autoregulates its own expression by cleaving the 5'-leader region of *rnc* mRNA, leading to mRNA instability and less RNase III expression. Interestingly, cross-regulation of ribonucleases is also known. In *E. coli*, RNase III cleaves in the 5'-proximal region of the *pnp* mRNA, resulting in PNPase-mediated degradation of the 5'-terminal fragment that is bound by complementary base-pairing to the sequence upstream of the PNPase translation initiation region (see Figure 4). This results in access of the 5' end of the downstream fragment to RNase E-initiated decay. Thus, the combined action of three ribonucleases participates in the regulated expression of one of them. Regulation of ribonuclease gene expression is not only at the level of mRNA decay. Another example is the effect of the *gmr* gene product on RNase II activity (*gmr* = gene modulating RNase II). The Gmr protein

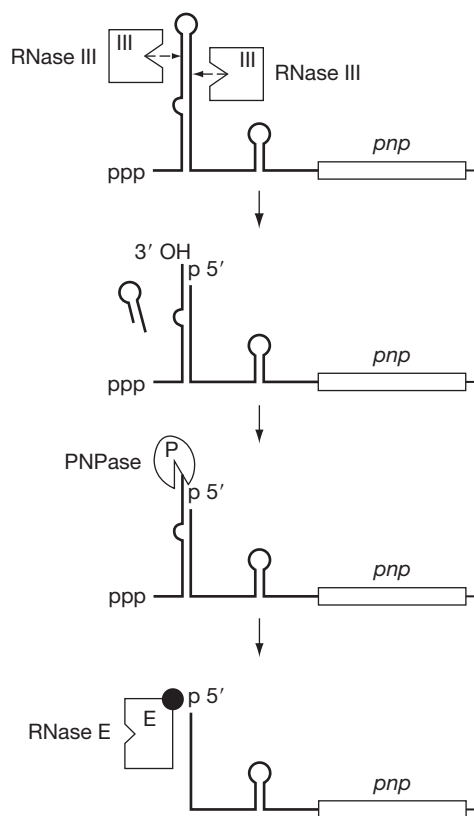


Figure 4 Regulation of *pnp* mRNA stability by RNase III, PNPase, and RNase E. The mRNA schematic shows the 5'-proximal secondary structure and the *pnp* coding sequence (open rectangle).

does not affect RNase II mRNA levels but, rather, controls the stability of the RNase II protein itself.

In addition to the regulation of ribonuclease levels, auxiliary proteins are known that bind a ribonuclease and regulate its activity. Two regulators of *E. coli* RNase E have been identified and named RraA and RraB (regulator of ribonuclease activity). These regulators bind at different sites in the C-terminal scaffold region of RNase E and inhibit its activity in specific ways. It will be interesting to determine how the ribonuclease regulators are themselves regulated.

Small Regulatory RNAs

A recent development in RNA biology has been an appreciation of the large number of small RNAs (sRNAs) that function as regulators of gene expression. Dozens of such sRNAs have been found in intergenic regions using bioinformatics approaches, microarray experiments, and deep sequencing. The sRNAs function in several ways to affect gene expression, one of which is to direct decay of one or more target mRNAs, similar to the function of microRNAs in eukaryotic systems. The action of sRNAs is directed by limited or extensive base complementarity between a specific mRNA and an sRNA, and is mediated by Hfq, a small RNA-binding protein that is an obligatory partner in sRNA-mediated decay. Hfq binds to the

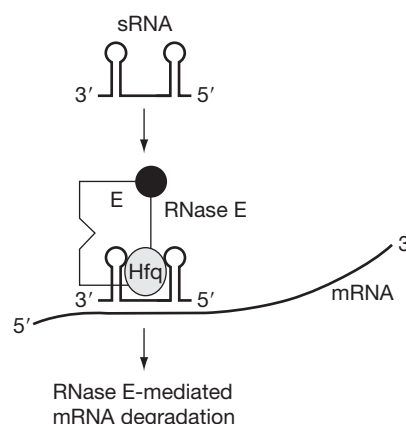


Figure 5 sRNA-mediated decay of a specific mRNA via pairing of sRNA to mRNA, binding of Hfq, and recruitment of RNase E.

sRNA and to RNase E to form a complex that is targeted to a specific mRNA (Figure 5). By virtue of its indirect interaction with RNase E (via Hfq), the binding of sRNA causes an increase in the local concentration of RNase E, resulting in a more rapid initiation of decay.

Translation and mRNA Decay

The intimate association of ribosomes and mRNA portends an effect of translation on mRNA half-life. An efficiently translated mRNA will have a high degree of ribosome occupancy, which should provide protection against decay-initiating endonuclease cleavage. However, we know very little about the parameters for steric hindrance to endonucleases that a ribosome flow might provide, and the relationship between translation and mRNA decay is not straightforward. What has been observed in many cases in *E. coli* and *B. subtilis* is that the strength of the ribosome:mRNA interaction at a 5'-proximal RBS can affect mRNA half-life. Thus, for an RBS that is located close to the 5' end, mutations that weaken the rRNA:RBS interaction most often result in destabilization. Similarly, mutation of the start codon to a codon that does not function in translation initiation results in a decrease in mRNA stability. It is thought that efficient and tight binding of the ribosome to the RBS restricts access to the 5' end by a decay-initiating endonuclease. In accordance with this hypothesis, the stabilizing effect of a favorable rRNA:RBS interaction is abolished when the RBS is located some distance away from the 5' end. What is less clear is the effect of a premature stop codon. In some cases in *E. coli* (but not all), insertion of a premature stop codon causes destabilization. Thus, in addition to ribosome binding near the 5' end, a flow of ribosomes down the body of the message seems to be required for optimal stability. On the other hand, experiments done on a limited number of *B. subtilis* mRNAs have shown that inserting an early stop codon, even immediately after the start codon, has no effect on mRNA half-life. Whether or not ribosome flow is relevant to stability could depend on where the initial endonuclease cleavage occurs; that is, if this cleavage occurs upstream of the

coding sequence, then ribosome flow in the coding sequence may not affect stability, as the initial cleavage will generate a new 5' end that promotes efficient decay, even in the presence of ribosome flow. However, if the initial cleavage occurs in the coding sequence itself, the presence of closely spaced ribosomes could inhibit initiation of decay.

We mentioned earlier the fascinating tmRNA system, which is used by cells to offload ribosomes that have reached the end of a nonstop mRNA. The tmRNA–protein complex includes RNase R, which degrades the nonstop mRNA in the 3'-to-5' direction. A nonstop mRNA can be generated by a decay-initiating endonuclease cleavage, or as a consequence of ribosome pausing during translation, which occurs at particular sequences/structures and under amino acid starvation conditions. Ribosome pausing activates a ribosome-bound mRNA interferase, which cleaves the codon that occupies the ribosomal A site, generating a nonstop mRNA from which the ribosome is offloaded by tmRNA and which is degraded by RNase R. There is also a second class of mRNA interferases, which function in a ribosome-independent manner. These mRNA interferases are activated under stress conditions, and they cleave mRNA endonucleolytically in a site-specific manner (e.g., at every ACA triplet sequence). The degradation

of mRNAs by mRNA interferases results in bacteriostasis, allowing the cell to survive the stress condition.

See also: [Molecular Biology: Pre-tRNA and Pre-rRNA Processing in Bacteria](#); [Small RNAs in Bacteria](#).

Further Reading

- Aiba H (2007) Mechanism of mRNA silencing by Hfq-binding small RNAs. *Current Opinion in Microbiology* 10: 134–139.
- Bechhofer DH (2009) Messenger RNA decay and maturation in *Bacillus subtilis*. *Progress in Molecular Biology and Translational Science* 85: 213–273.
- Carpousis AJ (2007) The RNA degradosome of *Escherichia coli*: An mRNA-degrading machine assembled on RNase E. *Annual Review of Microbiology* 61: 71–87.
- Carpousis AJ, Luisi BF, and McDowall KJ (2009) Endonucleolytic initiation of mRNA decay in *Escherichia coli*. *Progress in Molecular Biology and Translational Science* 85: 91–135.
- Condon C and Putzer H (2002) The phylogenetic distribution of bacterial ribonucleases. *Nucleic Acids Research* 30: 5339–5346.
- Deana A and Belasco JG (2005) Lost in translation: The influence of ribosomes on bacterial mRNA decay. *Genes and Development* 19: 2526–2533.
- Hayes CS and Keiler KC (2010) Beyond ribosome rescue: tmRNA and co-translational processes. *FEBS Letters* 584: 413–419.

Messenger RNA Processing in Eukaryotes

K Potter, N Cremona, and J A Wise, Case Western Reserve University School of Medicine, Cleveland, OH, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Alternative splicing A process through which a given primary transcript undergoes differential exon joining events to produce mature messenger RNAs (mRNAs), in which some segments differ; in most cases, these mRNAs are translated to yield proteins that may have different properties.

Chromatin The complex of genomic DNA and its associated proteins present in the nuclei of eukaryotic cells.

Exon A block of sequence within the pre-mRNA that is retained in the mature mRNA following splicing.

Intron A block of sequence within the pre-mRNA that is removed via splicing in the cell nucleus.

5' mRNA capping Addition of 7-methylguanosine to the 5' phosphorylated end of a nascent RNA polymerase II transcript through an enzymatic mechanism that produces a 5'-5' triphosphate linkage.

mRNA The class of RNA that carries the coding information from the DNA in the nucleus to the cytoplasm and serves as the blueprint for protein synthesis by cytoplasmic ribosomes.

Nucleosome The repeating unit of chromatin, which consists of DNA wrapped around a histone octamer consisting of two copies each of the H2A, H2B, H3, and H4 proteins.

ORF An open reading frame, defined as a continuous segment of nucleic acid that contains no nonsense codons and, therefore, can be translated into a full-length polypeptide.

Pre-mRNA The primary transcript of a gene synthesized by RNA polymerase II, which serves as the substrate for

RNA-processing reactions. Primary transcripts, also known as precursor mRNAs, extend from the site of transcription initiation to beyond the 3' end of the mature mRNA.

Ribonucleoprotein particle A general term used to describe any complex of RNA and protein.

RNA polymerase II The eukaryotic enzyme that catalyzes synthesis of mRNA precursors as well as a subset of small stable RNAs.

RNA splicing A process by which primary transcripts are converted to mature mRNAs via the removal of internal segments known as introns accompanied by precise joining of exons.

snRNP An acronym for small nuclear ribonucleoprotein particle, generally used to describe complexes between one or two small stable RNAs and a variable number of proteins; these reside in the cell nucleus and function in RNA-processing reactions.

Spliceosome A macromolecular complex composed of snRNPs and proteins that assembles on pre-mRNAs and catalyzes the splicing reaction.

3' mRNA polyadenylation Addition of multiple adenosine monophosphate (AMP) units derived from adenosine triphosphate (ATP) to the 3' end of a nascent RNA by poly(A) polymerase at a site created by cleavage of the precursor RNA. The cleavage/polyadenylation site is defined by specific 3' signals recognized by components of the polyadenylation machinery other than the polymerase.

Transcription The process through which RNA is synthesized in the cell nucleus by enzymes known as RNA polymerases, which copy the DNA template.

Historical Overview

In prokaryotes, all steps in gene expression take place in the same cellular compartment and, thus, transcription and translation of a messenger RNA (mRNA) are intimately linked, whereas in eukaryotes, the nuclear membrane provides a physical barrier between the compartments in which an mRNA is synthesized and translated. During its sojourn in the nucleus, the primary RNA transcript undergoes many physical transformations including 5' capping, removal of internal, noncoding sequences via splicing, and addition of a 3' poly(A) tail. Capping, the enzymatic addition of a 7-methyl guanosine not encoded in the DNA to the 5' end of the transcript, occurs very early in the biosynthesis of an mRNA; this modification confers resistance to nucleases that degrade RNA in a 5'-3' direction and also facilitates translation. The poly(A) tail, which can reach lengths of several hundred nucleotides in mammals, is also not encoded in the DNA. Like the cap

structure, the poly(A) tail enhances RNA stability and promotes translation in the cytoplasm; these functions require binding to the major poly(A)-binding protein (designated PAB1 in yeast and PABPC in mammals). The third major nuclear event, pre-mRNA splicing, is the process by which internal noncoding segments (introns) are removed from the primary transcript through a concerted mechanism that concomitantly joins the segments that form the mature mRNA (exons). Finally, some pre-mRNAs are subject to editing, a process through which the sequence encoded in the DNA is altered in the RNA. The ultimate fate of a eukaryotic mRNA is to be degraded (turned over), which generally occurs in the cytoplasm after it undergoes multiple rounds of translation. Throughout its life cycle, a given mRNA is subject to quality-control processes known as RNA surveillance, which detect aberrant processing or damage and target nonfunctional RNA molecules for elimination, in some cases before they can be exported from the nucleus.

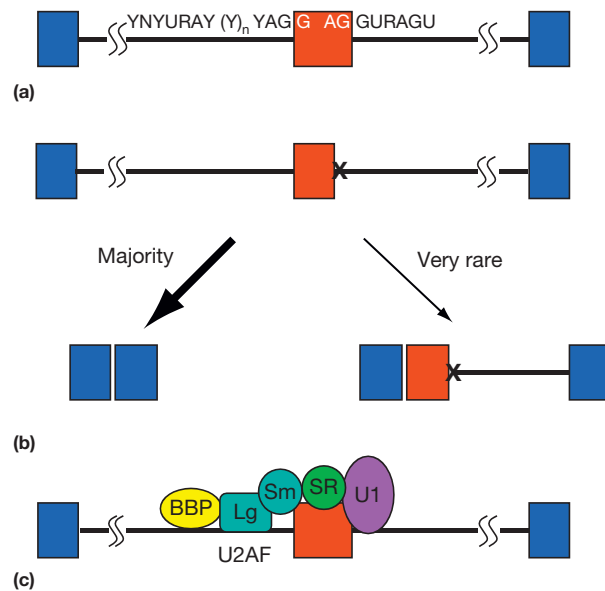


Figure 1 (a) Degeneracy of mammalian splicing signals. An internal portion of a pre-mRNA is shown with the consensus sequences flanking the middle exon indicated. (b) Consequences of mutating the 5' splice site in a mammalian pre-mRNA. In the vast majority of cases, the outcome is skipping of the preceding exon (left side), consistent with splice site pairing via exon definition. In rare cases, generally involving small introns that resemble those found in simpler eukaryotes, the intron is retained. (c) Composition of an exon definition complex. The functions of individual proteins are discussed under the section The Roles of RNA-Binding Proteins in the Life Cycle of an mRNA. BBP, branchpoint-binding protein; U2AF, U2 snRNP auxiliary factor, composed of a large and a small subunit; SR, serine-arginine-rich protein; U1, U1 snRNP. The drawing is not to scale.

For a number of years after the cap structure and poly(A) tail were identified and characterized, it remained unclear why cytoplasmic mRNAs were so much smaller than their nuclear precursors, especially in mammalian cells. This mystery was solved by the discovery of premessenger RNA (pre-mRNA) splicing in the late 1970s. The presence of introns was initially revealed by an electron microscopic technique that demonstrated covalent joining at the RNA level of discontinuous gene segments. Soon after the discovery of pre-mRNA splicing, conserved RNA sequence motifs were noted at the exon/intron boundaries and just upstream from the 3' splice site, although their significance did not become clear until elucidation of the splicing mechanism in the early 1980s. Despite the relatively simple chemistry of splicing, the reactions take place in a very large complex dubbed the spliceosome. At the heart of this macromolecular machine are four small nuclear ribonucleoprotein particles (abbreviated snRNPs and pronounced 'snuRPs'), each of these contains one or two snRNA molecules and a variable number of proteins. During the course of spliceosome assembly, the snRNA components recognize splicing signals via base-pairing interactions. At the 5' exon-intron boundary, a hexanucleotide (GUAAGU in yeast) is recognized early in the pathway by U1 snRNA, which is subsequently replaced by U6 snRNA. At the branch site, so named because it is joined to

the 5' splice junction during the first step of splicing to form a lariat structure, is a heptanucleotide (UACUAAAC in yeast) that is complementary to U2 snRNA. The 3' intron-exon boundary, which is joined to the 5' splice junction during the second step of splicing, is demarcated by the trinucleotide YAG preceded by a more variable element (U-rich in yeast). Finally, the moderately conserved nucleotides immediately surrounding the intron form base pairs with U5 snRNA. Although the yeast consensus splicing signals function optimally in mammalian cells, the actual 5' and 3' splice sites are often highly degenerate, with only the terminal GU and AG serving as invariant anchors and the branchpoint motif sometimes barely recognizable (Figure 1(a)). Beyond their roles in splice site recognition, a growing body of evidence indicates that the RNA components of two spliceosomal snRNPs, U2 and U6, are responsible for catalyzing the sequential transesterification reactions through which introns are removed and exons are ligated.

The Relationship between RNA Processing and Organismal Complexity

When the RNA-processing field was in its infancy, splicing was widely believed to make only a minor contribution to either regulated or constitutive gene expression. However, its status was greatly elevated upon completion of the human genome sequence, which revealed far fewer protein-coding genes than expected. The number of distinct proteins necessary to make up a human being is estimated at >100 000; remarkably, these are produced from only ~25 000 genes. Instead of creating entirely new open reading frames (ORFs), the necessary complexity was achieved largely through alternative splicing, a process through which the same genetic information is used to produce multiple functional mRNAs and, therefore, multiple distinct proteins. Although pre-mRNA splicing plays a pivotal role in both regulated and constitutive expression of genes in other eukaryotes, the discrepancy between genome size and the number of protein-coding genes is not nearly so great as in humans. For example, the fruit fly haploid genome contains only ~1/20th the amount of DNA as the human haploid genome but the number of genes is ~13 000, and even the lowly nematode worm has ~18 000 distinct ORFs. Similar to humans and other vertebrates, these invertebrates use alternative splicing to increase the coding capacity of their genomes; in fact, the most complex gene yet to be identified is fruit fly Dscam (Down syndrome cell adhesion molecule), which has the capacity to produce tens of thousands of alternative mRNAs. In contrast to metazoans, unicellular eukaryotes do not make wide use of alternative splicing, and the number of protein-coding genes is similar to the number of proteins, ~6000 in budding yeast and just over 5000 in fission yeast.

Although alternative splicing of human genes is believed to be the primary factor that accounts for the discrepancy between the number of genes and the number of proteins, other processes also contribute. First, alternative polyadenylation or a competition between splicing and polyadenylation within the terminal exon can produce different protein isoforms, and another source of proteomic diversity is editing, a

process in which the sequence of the gene is altered at the RNA level to either change the encoded amino acid or alter the length of the ORF by the introduction or elimination of a nonsense codon. Like splicing, RNA editing often occurs in a tissue-specific manner to produce different forms of the gene product. Finally, the use of alternative promoters can sometimes result in the presence or absence of new exons, but more often alters the length of the 5' untranslated region (UTR) of an mRNA; this, in turn, can affect translation through protein binding or the presence of RNA secondary structure. Similarly, alternative polyadenylation most often changes the length of the 3' UTR, which frequently contains elements that regulate translation and/or stability of the mRNA. By contrast, alternative splicing most commonly results in the inclusion or exclusion of a segment within the protein-coding region, though even in simple eukaryotes introns are sometimes present in the UTRs. A recent estimate suggests that only ~2% of the human genome is utilized to produce protein-coding mRNAs. Of this fraction, the ORFs constitute only ~1/2, with the remainder devoted to the 5' and 3' UTR's.

A Daunting Challenge: Finding Exonic Islands in a Sea of Intronic RNA

Exon Definition

Although the human genome does contain vast regions that largely or completely lack genes, a substantial fraction of the excess DNA is actually copied to produce the primary transcripts that are processed into mature mRNAs for export to the cytoplasm. Given that virtually every mammalian gene contains multiple introns, and that in general these are much larger than the exons that make up mRNAs (Figure 2(a)), a key question was how the small islands of coding RNA are accurately identified and precisely joined. An important clue came from the consequences of splice site mutations: if either the 5' splice site preceding an exon or the 3' splice site immediately downstream sustains a mutation that renders it nonfunctional, the result in the vast majority of cases is skipping of the exon (Figure 1(b), left side) rather than retention of the intron (Figure 1(b), right side). This surprising observation led Susan Berget to propose the 'exon definition' model, which

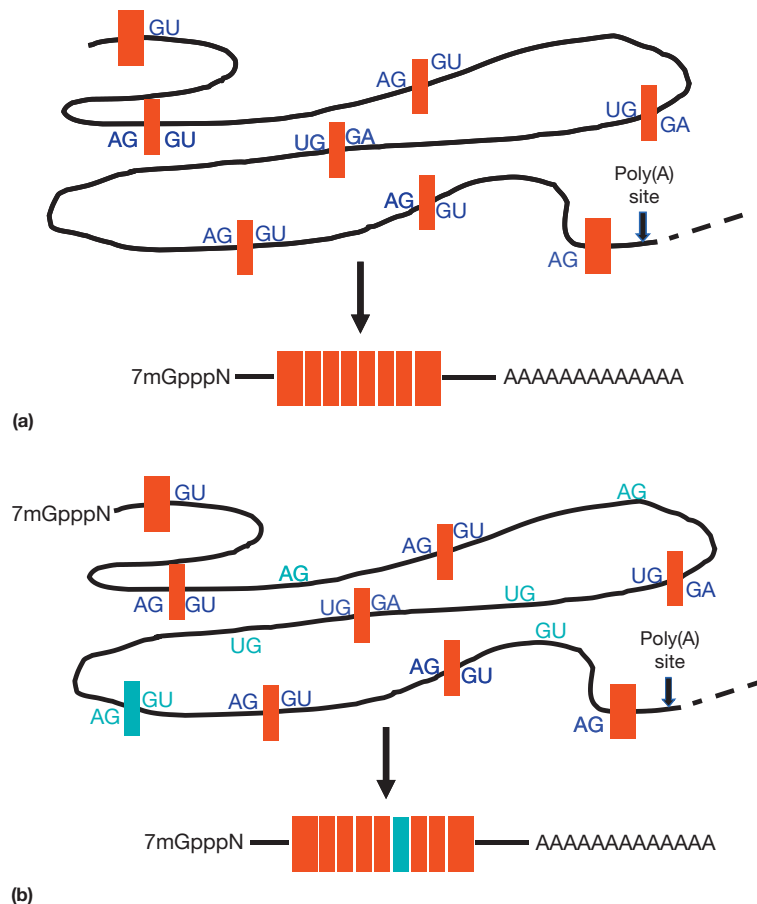


Figure 2 (a) (Top) Architecture of a typical mammalian primary transcript. On average, pre-mRNAs contain eight exons that are quite uniform in size, clustering around 300 nucleotides, interrupted by seven introns that are generally 1.5–3 kilobases in length but may be as long as half a megabase. Also indicated is the polyadenylation site located near the 3' end of the transcript. (Bottom) The fully processed mRNA after intron removal by splicing and addition of the 5' cap structure and 3' poly(A) tail. (b) Evolution of an alternatively spliced exon from an intronic region. The locations of sequences resembling canonical splicing signals are indicated by AG and GU dinucleotides; acquisition through mutation of a pair of splice sites at a distance that allows recognition by the splicing machinery can lead to exonization of a segment in the pre-mRNA that is part of an intron in the otherwise identical transcript depicted in (a). Inclusion of the alternative exon can be increased by improving the match between its splicing signals or acquisition of a nearby exonic splicing enhancer.

postulates that splice site pairing in mammalian pre-mRNAs initially occurs across the exons. Direct experimental support for this model came from the finding that at the earliest stage of spliceosome assembly, a serine-arginine-rich (SR) protein bridge links the factors bound to the 3' and 5' splice sites on opposite sides of an exon (Figure 1(c)). In simpler eukaryotes, introns are generally much smaller than exons and the splice sites are paired via intron definition, consistent with the finding that splice site mutations generally lead to intron retention. The unifying feature is that in both cases proximity is the driving force for splice site pairing. The exon definition model also helped to explain why, in humans and other higher eukaryotes, exon length is much more uniform than intron length, as well as how exon skipping is generally avoided in spite of the extremely large sizes of many mammalian introns.

Ancillary RNA-Processing Signals

Because unambiguous specification of splice sites is essential to avoid producing mRNAs with frameshifts that would preclude translation into functional proteins, the low information content of mammalian splicing signals was initially puzzling, particularly in light of the frequent occurrence of sequences related to authentic splicing signals within the very large introns. The mystery of the missing splicing information was solved by the discovery of ancillary splicing signals that function in collaboration with the degenerate 'classical' splicing signals (Figure 3(a)). The first group of elements to be discovered, the so-called exonic splicing enhancers, activates weak (i.e., nonconsensus) splice sites via binding to positively acting factors (discussed further below). Subsequently, different classes of splicing enhancer elements were found within introns, and multiple families of elements that repress splicing, known as silencers, can also affect the composition of the mRNA from either exonic or intronic locations. Although enhancers were initially uncovered based on their roles in promoting tissue- or

developmental stage-specific alternative splicing, they also function during constitutive splicing to ensure that exons surrounded by suboptimal splicing signals are included in the mature mRNA. Indeed, the mode of exon recognition in constitutively spliced pre-mRNAs provides the mechanistic basis for positive and negative regulation of alternative splicing. On an evolutionary scale, the degenerate nature of mammalian splice site sequences provides a source for the creation of new exons from sequences that were previously intronic (Figure 2(b)), prompting the designation of introns as the 'playground of evolution'. Conversely, mutations at ancillary splicing signals (particularly splicing enhancers) can lead to exon skipping and, if fixed in the gene pool, permanent loss. Recent estimates suggest that 60% of mutations that cause human disease do so by affecting splicing, via changes either in ancillary or in classical splicing signals.

Despite dramatic progress in understanding the role of splicing on a genome-wide basis, mechanistic questions still remain. For example, regardless of the initial mode of recognition, splice sites must ultimately be paired across introns (Figure 3(b)) to allow the U2 and U6 snRNAs to carry out their catalytic functions. An important challenge for the future is to learn how the necessary rearrangements occur.

The Roles of RNA-Binding Proteins in the Life Cycle of an mRNA

In addition to the snRNPs discussed above, purification and proteomic analyses of spliceosomes at different stages of the splicing cycle revealed a remarkably large number of protein constituents. Among the >300 polypeptides that copurify with the mammalian spliceosome, many turned out to be factors previously identified through biochemical fractionation of cell-free splicing extracts. These include a heterodimeric protein, which was designated U2 auxiliary factor (U2AF) based

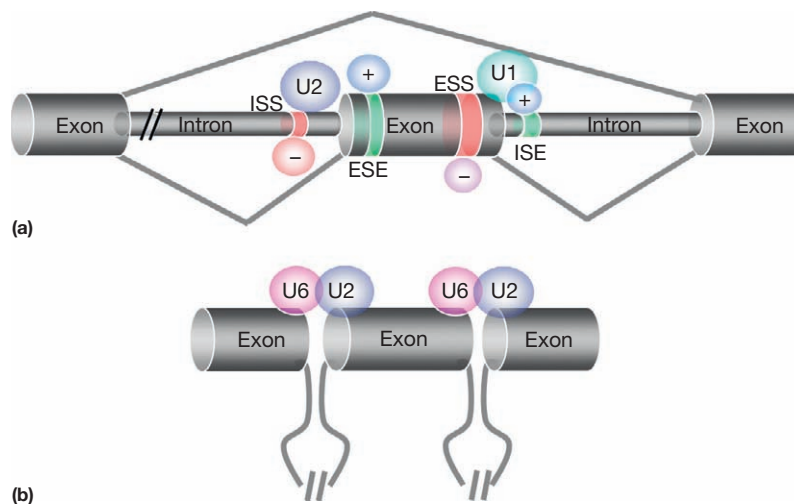


Figure 3 (a) Ancillary splicing signals: ESE, exonic splicing enhancer; ISE, intrinsic splicing enhancer; ESS, exon splice site; ISS, intron splice site. The U1 and U2 snRNPs are indicated by cyan and purple balls, respectively, and the effect of each auxiliary splicing element is indicated by a positive or negative symbol. (b) Exon juxtaposition. Before splicing can occur, several poorly understood rearrangements must occur such that the exons are juxtaposed and the catalytic components of the spliceosome (believed to reside within the U2 and U6 snRNP; see text) are positioned across the intron rather than the exon.

on its role in promoting addition of the U2 snRNP to the growing spliceosome. During the earliest phase of spliceosome assembly, the large and small subunits of U2AF recognize the polypyrimidine tract and 3' AG dinucleotide, respectively (Figure 1(c)). The large subunit makes contacts with the branchpoint-binding protein, which is subsequently displaced by the U2 snRNP, while the small subunit forms the 5' anchor of a bridge to the U1 snRNP bound to the downstream 5' splice site. The final constituents of the exon definition complex are the SR family of splicing factors, which form a bridge across the exon. In general, these factors associate with splicing enhancer elements and, in at least some cases, function to promote binding of U2AF, and ultimately U2 snRNP, to the upstream 3' splice site. The roles of splicing silencers are less well understood, but splicing repression has often been associated with members of the heterogeneous ribonucleoprotein family of proteins, members of which also copurify with spliceosomes. Finally, it is important to note that, although the chemical reactions are almost certainly RNA-catalyzed, many proteins also play critical roles in splicing. The most highly conserved of these, a very large protein known as PRP8, makes contact with both the 5' and 3' ends of the introns and remains associated throughout the catalytic cycle.

Many of the proteins implicated in splicing as well as other aspects of RNA metabolism contain domains that recognize short, moderately conserved motifs. The largest group of sequence-specific RNA-binding proteins, designated the RNA recognition motif family, is encoded by over 500 different genes in mammalian cells. Other common RNA-binding motifs include the KH domain, the double-stranded RNA-binding domain, zinc fingers, RGG boxes, and the Pumilio domain. The proliferation of RNA-binding proteins, even in simple eukaryotes, reflects the fact that, throughout its life cycle, an mRNA is associated with a diverse array of proteins, some of them stably bound and others subject to dynamic exchange. In many cases, the protein components of messenger RNPs (mRNPs) serve as adaptors to facilitate association with the macromolecular machines that process it in the nucleus, use it as a blueprint for protein synthesis in the

cytoplasm, or cause its demise via RNA turnover once its useful lifetime has ended.

Mechanistic Coupling of Different Steps in Gene Expression

Until recently, the multiple events required for the production of a translation-competent mRNA from a primary transcript in the nucleus were generally studied in isolation, often in cell-free extracts. Nevertheless, there were early indications that splicing occurs cotranscriptionally *in vivo*, for example, the Beyer laboratory's pioneering electron micrographic studies. Multiple instances in which the same protein was isolated as a splicing factor in one laboratory and a transcription factor in another provided further circumstantial evidence for coupling, but it was not until the late 1990s that the underlying mechanisms began to emerge (Figure 4). A landmark study from the Bentley laboratory used a drug to inactivate endogenous RNA polymerase II (RNAP II) in cells carrying a drug-resistant mutant lacking the C-terminal domain (CTD). Analysis of transcripts made by the mutant polymerase revealed that the CTD is required for both splicing and polyadenylation. The extended structure of this domain, in conjunction with the changes in association of RNA-processing factors due to altered CTD phosphorylation, led to a widely accepted model in which it serves as a recruiting platform at the promoter for delivery of bound factors at appropriate times during the transcription cycle (Figure 4). Remarkably, in addition to capping and splicing factors, factors required for cleavage and polyadenylation to form the mature 3' end of the mRNA are also present at the promoter. Although it has long been known that 3' processing is tightly coupled to transcription termination by RNAP II, the mechanism has been controversial. Recent studies provide support for the so-called 'Torpedo' model, in which degradation of the RNA downstream from the polyadenylation site helps to displace the polymerase from the DNA.

In addition to coupling between RNA processing and transcription, many links have been uncovered between different

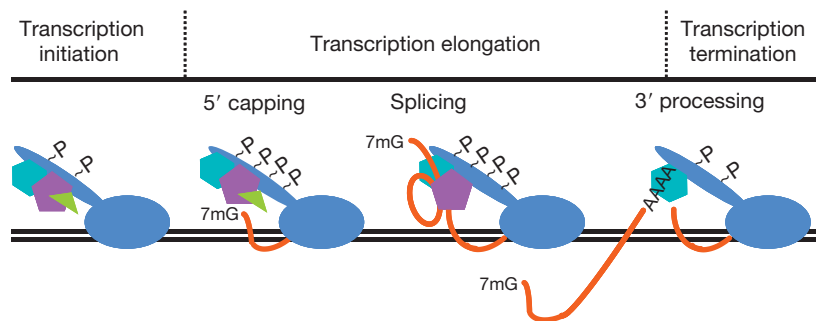


Figure 4 'Assembly line' view of cotranscriptional RNA processing. Each stage of RNA polymerase II transcription is indicated at the top, while the next line shows the coupled RNA-processing events that occur during maturation of the primary transcript. RNAP II is indicated by the blue oval with the C-terminal domain (CTD) of its largest subunit indicated by a tail. The capping enzymes are indicated by a green triangle, the spliceosome by a purple pentagon, and the polyadenylation complex by a turquoise hexagon. For simplicity, the complete machinery involved in each step of RNA processing is shown assembled on the CTD at the promoter; in reality, only a subset of the factors, as yet not fully defined, associate with the promoter. Soon after transcription initiation, the 5' end of the nascent RNA acquires a cap, which associates with a heterodimeric protein that stimulates splicing. Spliceosome assembly on the nascent RNA occurs during the transcription elongation phase and, for longer transcripts, introns near the 5' end are spliced out before synthesis of the primary transcript is completed. Processing at the 3' end by cleavage and polyadenylation is coupled to transcription termination.

RNA-processing reactions. For example, both capping and polyadenylation stimulate splicing, and splicing factors have been implicated in RNA export, as has polyadenylation of the newly synthesized RNA. In some cases, the crosstalk is bidirectional; for example, increased transcription stimulates splicing and the presence of a functional 5' splice site promotes RNA synthesis. Perhaps most remarkable is the ability of the exon junction complex, which is deposited on the RNA during splicing in the nucleus, to influence turnover of the RNA in the cytoplasm.

Future Directions

An area that has recently come under intense scrutiny is how the structure of chromatin, the actual template for transcription, affects RNA-processing events. Although it is now clear from both genome-wide studies and investigation of individual genes that histone modifications and chromatin-remodeling factors influence splicing, elucidating the underlying mechanisms will undoubtedly be a fertile area of investigation in the years to come.

See also: **Cell Architecture and Function:** Nuclear Pores and Nuclear Import/Export; **Molecular Biology:** Alternative Splicing; Regulation of *Drosophila melanogaster* Sex Determination; mRNA Polyadenylation in Eukaryotes; Regulation of Chromatin Dynamics.

Further Reading

- Gott JM and Emeson RB (2000) Functions and mechanisms of RNA editing. *Annual Review of Genetics* 34: 499–531.
- Komili S and Silver PA (2008) Coupling and coordination in gene expression processes: A systems biology view. *Nature Reviews Genetics* 9: 38–48.
- Kornblihtt AR (2005) Promoter usage and alternative splicing. *Current Opinion in Cell Biology* 17: 262–268.
- Moore MJ (2005) From birth to death: The complex lives of eukaryotic mRNAs. *Science* 309: 1514–1518.
- Nilsen TW and Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463: 457–463.
- Perales R and Bentley D (2009) "Cotranscriptionality": The transcription elongation complex as a nexus for nuclear transactions. *Molecular Cell* 36: 178–191.
- Richard P and Manley JL (2009) Transcription termination by nuclear RNA polymerases. *Genes and Development* 23: 1247–1269.
- Wang GS and Cooper TA (2007) Splicing in disease: Disruption of the splicing code and the decoding machinery. *Nature Reviews Genetics* 8: 749–761.

Metabolic Control during Ischemia of the Heart

J R Ussher and G D Lopaschuk, University of Alberta, Edmonton, AB, Canada

© 2013 Elsevier Inc. All rights reserved.

Glossary

Cardiac efficiency The ratio of external cardiac power to energy expenditure by the left ventricle.

Ischemia A restriction in blood supply.

Reperfusion Restoration of blood flow/supply.

Cardiovascular disease is now the primary cause of death worldwide. For the majority of people living with cardiovascular disease, the underlying cause is a condition coined ischemic heart disease. Ischemia of the heart occurs when coronary blood flow is inadequate, and hence the oxygen supply to the myocardium is not sufficient to meet the oxygen demand of the heart. Due to the heart's high demand for energy to drive cardiac contraction, the heart is extremely susceptible to injury if an interruption to its oxygen supply occurs. The manifestations of myocardial ischemia depend upon the nature and severity of the ischemic episode, and the subsequent re-establishment of flow (reperfusion). Clinically, ischemic heart diseases can manifest as angina pectoris, acute myocardial infarction, and heart failure.

In the aerobic heart, energy production and cardiac function are closely matched. However, ischemia disturbs the relationship between fatty acid and glucose metabolism. In isolated working heart studies, metabolic disturbances during ischemia include an acceleration of anaerobic glycolysis, and of what residual oxidative metabolism remains, a greater reliance of the heart on fatty acids as a fuel source, which occurs at the expense of glucose. This results in an uncoupling between glycolysis and glucose oxidation, which leads to the production of protons and development of intracellular acidosis. As a result, adenosine triphosphate (ATP) must be diverted away from contractile function in order to re-establish ionic homeostasis, which decreases cardiac efficiency.

During successful reperfusion of the ischemic heart, glycolytic rates remain high, and the oxidation of fatty acids quickly recovers, once again at the expense of glucose. Thus, ATP must continue to be used to deal with the intracellular acidosis arising from uncoupled glucose metabolism, and cardiac efficiency remains impaired.

The consequences of myocardial ischemia are numerous, and metabolic perturbations with regards to the availability of circulating energy substrates, as well as the subcellular regulation of energy substrate metabolism are important factors underlying some of these consequences. The purpose of this article is to outline the pathways of metabolism in the healthy heart, the metabolic changes that take place in the heart during ischemia, and how they adversely affect cardiac function. In addition, we also highlight how alterations in certain signaling enzymes contribute to these ischemic-induced changes in metabolism. Finally, we will touch upon studies in both rodents and humans showing that modulating these metabolic disturbances is a novel and promising therapeutic approach for ischemic heart disease.

Carbohydrate Metabolism in the Heart

The primary carbohydrate used as an energy source by the heart is glucose, and the majority of glucose that the heart metabolizes is obtained from the blood with its uptake being facilitated by glucose transporters (GLUTs). GLUT1 is responsible for maintaining basal glucose uptake, whereas GLUT4 translocates from an intracellular pool to the sarcolemmal membrane in response to insulin, increased work demand, or ischemia.

The metabolism of glucose (Figure 1) can be separated into two major components, glycolysis and glucose oxidation. Glycolysis results in the production of pyruvate and accounts for less than 10% of the total ATP produced by the aerobic heart. Flux through glycolysis is primarily regulated at the level of phosphofructokinase-1 (PFK1), which is responsible for catalyzing the rate-limiting, irreversible conversion of fructose-6-phosphate into fructose-1,6-bisphosphate. PFK1 is allosterically regulated by the energy status of the cell, with metabolites of low cellular energy levels such as adenosine monophosphate, adenosine diphosphate, and inorganic phosphate increasing its activity, and metabolites of high cellular energy levels such as ATP and citrate decreasing its activity. If glycolysis is coupled to glucose oxidation, the pyruvate generated from glycolysis is converted to acetyl coenzyme A (CoA) (which is subsequently oxidized in the tricarboxylic acid (TCA) cycle) by the enzymatic action of the multienzyme complex, pyruvate dehydrogenase (PDH) (Figure 1). The majority of pyruvate that is not converted into acetyl-CoA will be converted into lactate via the enzymatic activity of lactate dehydrogenase (LDH), but a small fraction (~2–6% in the aerobic heart) will be carboxylated into malate or oxaloacetate and enter the TCA cycle via anaplerosis.

The PDH complex is under tight regulation by an upstream kinase, PDH kinase (PDK), which phosphorylates and inhibits its activity, whereas PDH phosphatase dephosphorylates and increases its activity (Figure 1). There are four PDK isoforms, with PDK4 being the primary isoform in the heart. PDK is positively regulated by acetyl-CoA and nicotinamide adenine dinucleotide (NADH). As mitochondrial acetyl-CoA/CoA and NADH/NAD⁺ ratios are increased by elevated rates of fatty acid β -oxidation, it leads to an increase in PDK activity and a potent inhibition of PDH and glucose oxidation. This reciprocal relationship between glucose and fatty acids for oxidative metabolism was first described by Sir Philip Randle and colleagues in the 1960s and has been termed the Randle cycle. It will be described in later sections how

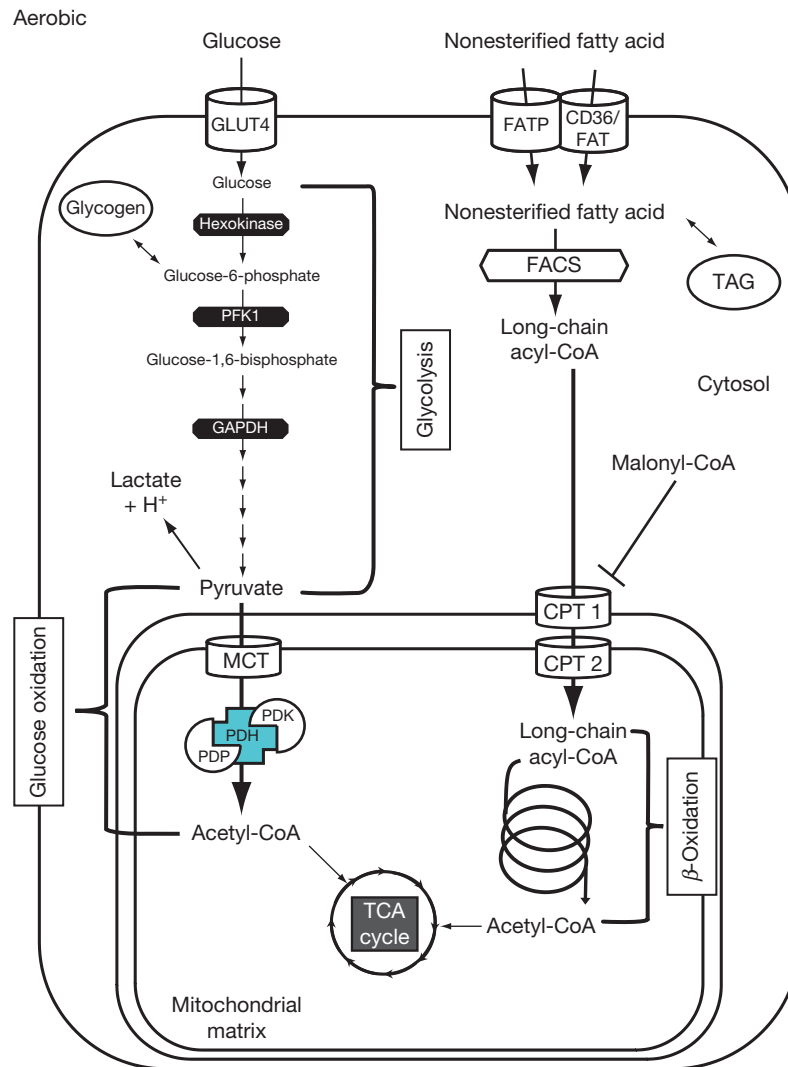


Figure 1 Energy metabolism in the aerobic heart. In the aerobic heart, energy metabolism is primarily derived from either plasma glucose or nonesterified fatty acids. Glucose is taken up via glucose transporters (GLUTs) such as GLUT4 and undergoes glycolysis to become pyruvate, which is then converted either into lactate and a H⁺, or taken up into the mitochondria and oxidized via PDH to produce acetyl-CoA for the TCA cycle. Fatty acids are primarily taken up via protein-mediated transporters, such as CD36/FAT, and then activated by esterification to CoA via the enzymatic activity of fatty acyl-CoA synthetase (FACS). The activated fatty acyl-CoA is taken up into the mitochondria via CPT 1, whereby it undergoes β -oxidation to produce acetyl-CoA for the TCA cycle. In the aerobic heart, there is a steady balance between the contribution of glucose and fatty acids for oxidative energy production.

this reciprocal substrate cycle is affected by ischemia and contributes to the pathology of ischemic heart disease.

Alternative sources for glucose metabolism outside of plasma glucose include endogenous glycogen stores and lactate. The mobilization of endogenous glycogen stores can generate glucose-6-phosphate, which bypasses the hexokinase reaction of glycolysis. Thus, glycolysis of glycogen-derived glucose-6-phosphate generates three molecules of ATP, as opposed to the two normally seen from glycolysis of glucose taken up from the plasma.

While being a primary end product of glycolysis, lactate from other tissue sources such as the skeletal muscle is also a major energy source for the aerobic heart, and is converted into pyruvate via the enzymatic activity of LDH. The pyruvate

molecule is then oxidized in the same manner that pyruvate generated from glycolysis is, via the enzymatic activity of PDH. It is important to note that although the heart is normally a lactate consumer, during times of metabolic stress (such as myocardial ischemia) the heart becomes a lactate producer.

Fatty Acid Metabolism in the Heart

Fatty acids enter the cardiac myocyte by either passive diffusion or protein-mediated transport across the sarcolemmal membrane (Figure 1). The rate of fatty acid uptake by the heart is primarily determined by the concentration of free fatty acids (FFAs) in the plasma, which can vary over a fourfold range in

healthy humans during the course of the day (~ 0.2 – 0.8 mM). Once transported across the sarcolemmal membrane, fatty acids are subsequently activated by esterification to CoA by fatty acyl-CoA synthetase. This fatty acyl-CoA can either be partitioned into intracellular lipids, such as triacylglycerol, or converted to long-chain fatty acylcarnitine by carnitine palmitoyltransferase 1 (CPT-1) for subsequent oxidation in the mitochondria. The primary fate of fatty acids taken up into the heart is β -oxidation, as studies in humans have shown that 80% of long-chain fatty acids rapidly appear as CO_2 in coronary venous blood.

Fatty acid β -oxidation occurs predominantly in the mitochondria and to a smaller extent in the peroxisomes. As mentioned, for mitochondrial fatty acid β -oxidation to begin, the cytoplasmic long-chain fatty acyl-CoA must first be transported into the mitochondrial matrix through a carnitine-dependent transport system. This carnitine transport system involves three enzymes: CPT-1, carnitine acyltransferase, and CPT-2. Of these three enzymes, CPT-1 is rate limiting in regards to mitochondrial fatty acid uptake and is subject to potent inhibition via malonyl-CoA. Changes in intracellular malonyl-CoA concentrations during ischemia and reperfusion are key determinants regulating flux through both the fatty acid and glucose oxidative pathways.

Once in the mitochondria, β -oxidation repeatedly cleaves off two carbon acetyl-CoA units from fatty acyl-CoA, generating NADH and reduced flavin adenine dinucleotide (FADH_2) in the process. The β -oxidation process involves four enzymatically catalyzed reactions (acyl-CoA dehydrogenase, 2-enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase, and 3-ketoacyl-CoA thiolase), with the last step regenerating acyl-CoA for another round of β -oxidation and releasing acetyl-CoA for the TCA cycle. As mentioned earlier, β -oxidation of fatty acids increases the acetyl-CoA/CoA ratio, which inhibits PDH and subsequent glucose oxidation via the Randle cycle, and is a critical component of how ischemic-induced alterations in metabolism contribute toward the pathology of ischemic heart disease.

Energy Metabolism in the Ischemic Heart

Cardiac energy metabolism is drastically altered during ischemia (Figure 2), with the primary effect of ischemia being a lack of oxygen and nutrient supply and decreased clearance of metabolic by-products from the affected region(s) of the heart. During ischemia, there is a rapid acceleration in the conversion of pyruvate to lactate via LDH, which accelerates in an attempt to provide an anaerobic source of ATP to make up for the reduction in oxidative ATP production. As previously alluded to, PFK1 is the rate-limiting enzyme determining flux through the glycolytic pathway, and is allosterically activated by metabolites such as AMP and P_i , both of which are elevated during ischemia. However, as ischemia progresses and becomes more severe, glyceraldehyde-3-phosphate dehydrogenase can become the rate-limiting enzyme of glycolysis, due to the continual buildup of lactate, protons (H^+ s), and NADH in the cytosol, which inhibit its activity. Further progression of ischemia eventually results in reduced glycolytic rates as glycogen content depletes in combination with the reduction

in coronary flow. Moreover, just before irreversible damage takes place as a full-fledged myocardial infarction, PFK1 translocates from the cytosol to the sarcolemmal membrane and is subsequently inactivated.

Although glycolysis can provide a limited amount of ATP during ischemia, the hydrolysis of glycolytically derived ATP in the absence of subsequent pyruvate oxidation leads to an accumulation of lactate and H^+ s, which cause intracellular acidosis and further aggravate the ionic disturbances brought about by ischemia. Thus during ischemia, when glycolysis is accelerated, a greater proportion of ATP hydrolysis must be diverted toward performing chemical work (re-establishing ionic homeostasis) rather than maintaining contractile function, which is already jeopardized. It should be noted that the accumulation of lactate and H^+ s is less of an issue if the cardiac myocyte is exposed to hypoxia or a very mild ischemia, as opposed to a severe ischemia, since these by-products of glycolysis are rapidly removed from the cardiac myocyte.

When considering the effects of fatty acids on the heart during and following ischemia, it is important to recognize that fatty acids can have differing effects depending on whether the myocardium is hypoxic or ischemic. The obligate requirement for oxygen in the process of oxidative phosphorylation results in a rapid decrease in ATP production from the catabolism of fatty acids and pyruvate in proportion to the degree of ischemia. However, fatty acid β -oxidation dominates as the major source of residual oxidative metabolism with no increase in the relative contribution of carbohydrate oxidation. This contributes to the uncoupling between glycolysis and glucose oxidation, further aggravating intracellular acidosis and dysregulated ionic homeostasis.

With total ischemia there is an accumulation of reducing equivalents in the form of NADH and FADH_2 . Both the acyl-CoA dehydrogenase and 3-hydroxyacyl-CoA dehydrogenase enzyme reactions of fatty acid β -oxidation are sensitive to the redox state of the matrix (NAD^+/NADH and FAD/FADH_2 ratio). The inhibition of fatty acid β -oxidation secondary to the accumulation of reducing equivalents can result in the accumulation of fatty acid intermediates in distinct cellular compartments. Fatty acylcarnitine species can accumulate in both the mitochondrial matrix and cytosol, whereas fatty acyl-CoA species accumulate primarily in the mitochondrial matrix as this pool of CoA does not exchange with the relatively small cytosolic CoA pool. The accumulation of these acylcarnitine and acyl-CoA esters promotes disruption of mitochondrial cristae and the formation of amorphous intramitochondrial densities, changes that may ultimately disrupt mitochondrial function.

Ischemic-Induced Alterations in Metabolic Signaling

The aforementioned changes in metabolism that take place during ischemia are influenced by a number of signaling enzymes (Figure 3). One of the most important signaling enzymes that contributes toward the ischemic-induced elevation of glycolytic rates in the heart is 5'AMP-activated protein kinase (AMPK). AMPK is generally considered to be a cellular energy gauge whose activity reflects the energy status of the cell. During ischemia, the rapid increase in intracellular AMP

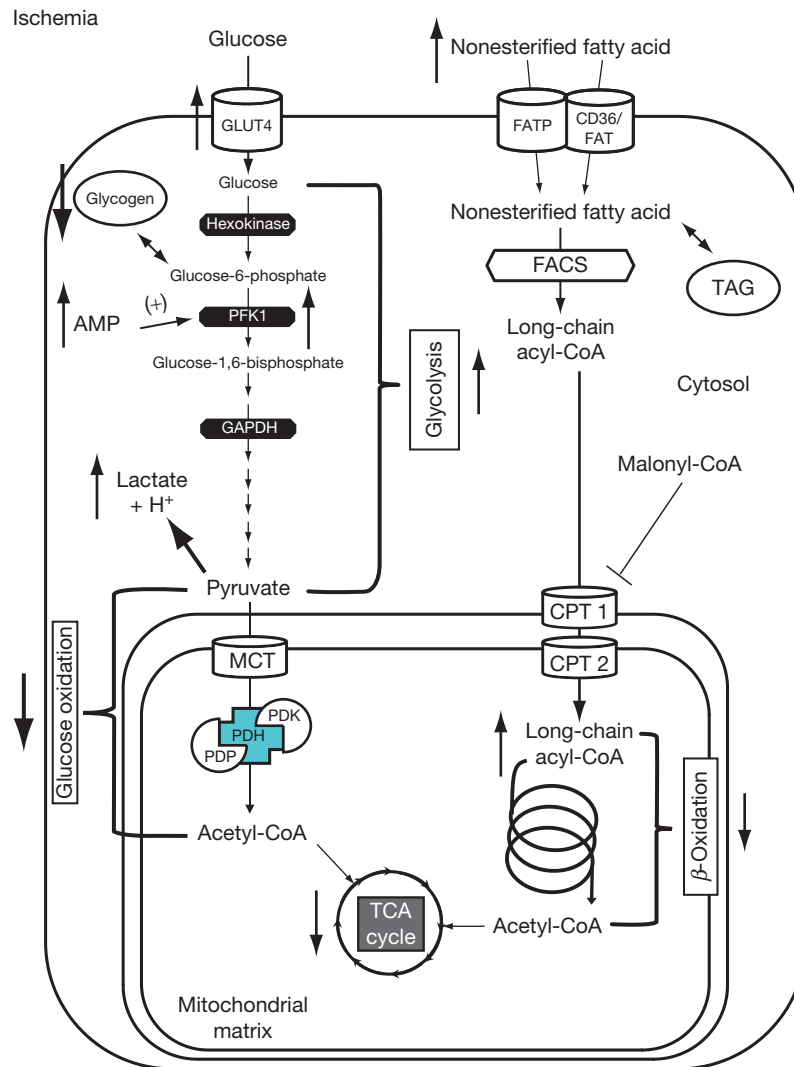


Figure 2 Energy metabolism in the ischemic heart. During ischemia, energy metabolism is drastically altered, as lack of oxygen leads to a dramatic reduction in oxidative energy production. However, fatty acids remain the primary source of what residual oxidative metabolism remains. To compensate for the reduction in oxidative energy production, rates of glycolysis are enhanced to provide an anaerobic source of energy. However, because glucose oxidation is impaired, one has an increased conversion of pyruvate into lactate and H^+ s, which results in the development of intracellular acidosis, reducing cardiac efficiency.

concentrations results in a rapid activation of AMPK, which then contributes to the ischemic-induced increase in glycolysis by phosphorylating PFK2 at serine residue 466, increasing PFK2 activity. PFK2 converts fructose-6-phosphate into fructose-2,6-bisphosphate, which is a potent allosteric stimulator of PFK1 activity. Furthermore, increased AMPK activity has been linked to GLUT4 translocation and increased glucose uptake, which would also contribute to the elevated rates of glycolysis observed during glycolysis.

The activation of AMPK during ischemia also results in the phosphorylation and subsequent inactivation of acetyl-CoA carboxylase (ACC), which is the enzyme responsible for the synthesis of malonyl-CoA. This reduction in malonyl-CoA content likely contributes to fatty acid β -oxidation rates accounting for the majority of residual oxidative metabolism observed during ischemia.

Hypoxia-inducible factor-1 α (HIF-1 α) is a transcription factor whose expression is rapidly induced during ischemia in the heart, and its target transcripts include a number of genes involved in the glycolytic pathway such as hexokinase and PFK1/2. HIF-1 α has also recently been shown to increase the expression of PDK1, which would inactivate PDH and further contribute to fatty acid β -oxidation dominating as the major source of residual oxidative metabolism.

Signaling changes in other tissues may also affect metabolism in the heart during ischemia. In particular, ischemia rapidly increases catecholamine discharge, and plasma norepinephrine concentrations increase within minutes and can remain elevated for periods of up to 20 h, depending on the severity of the ischemic insult and ensuing stress response. This activates hormone-sensitive lipase in the adipose tissue and elevates circulating FFA concentrations by promoting

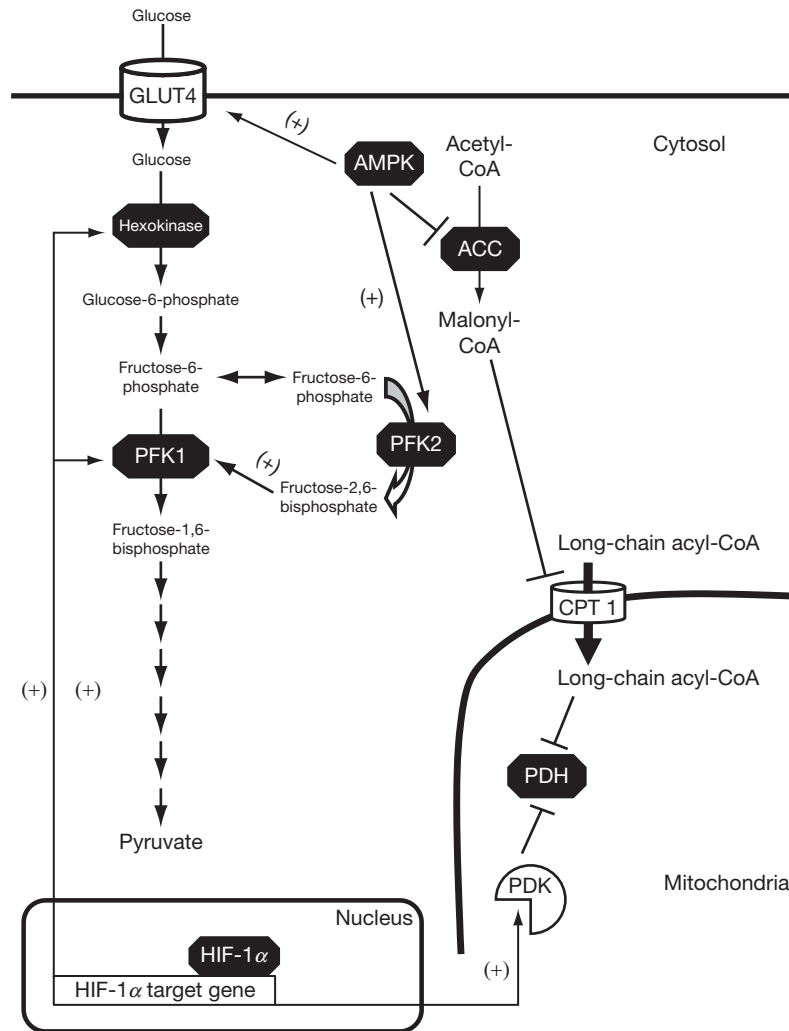


Figure 3 Signaling changes during ischemia that affect energy metabolism in the heart. During ischemia, AMPK is rapidly activated, which increases GLUT4 translocation and activates PFK2. Changes in both of these contribute to the elevation of glycolytic rates observed during ischemia. Furthermore, AMPK activation results in the inhibition of ACC which decreases malonyl-CoA content and contributes to fatty acid oxidation rates dominating as the primary source of residual oxidative metabolism during ischemia. HIF-1 α is also rapidly activated during ischemia, and this transcription factor increases the transcription of a number of glycolytic enzymes such as hexokinase and PFK1/2. In addition, HIF-1 α also increases the transcription of PDK1, which like AMPK's effect on ACC, contributes to fatty acid oxidation rates dominating as the primary source of residual oxidative metabolism during ischemia.

lipolysis. Furthermore, elevated plasma concentrations of hydrocortisone accompany the stress of ischemia, having permissive effects on adipose tissue lipolysis. An increase in circulating plasma FFAs during and following ischemia thus increases the delivery of fatty acids to the heart and can alter fatty acid utilization during both the ischemic and postischemic period.

In total, all these signaling changes collectively result in a dramatic increase in glycolysis to compensate for the reduction in oxidative energy metabolism, and due in part to the Randle cycle, a reliance on fatty acids for what residual oxidative metabolism remains. This leads to the uncoupling of glycolysis from glucose oxidation, which results in the production of H^+ s and induces intracellular acidosis, reducing the efficiency of converting metabolism into contractile function.

Energy Metabolism during Reperfusion of the Ischemic Heart

During reperfusion of the ischemic heart, glycolytic rates remain high, while fatty acid β -oxidation rates recover to a greater extent than that of carbohydrates such as glucose (Figure 4). As mentioned in the previous section, this likely arises from elevated circulating FFA concentrations due to sympathetic nervous system activation and catecholamine discharge from the stress of ischemia, and a continued reduction in intracellular malonyl-CoA content. In the rat heart this is accompanied via continued activation of AMPK and subsequent inhibition of ACC, although in the mouse heart AMPK activity and malonyl-CoA content return to preischemic values. Regardless of these species' differences, glycolysis remains dramatically

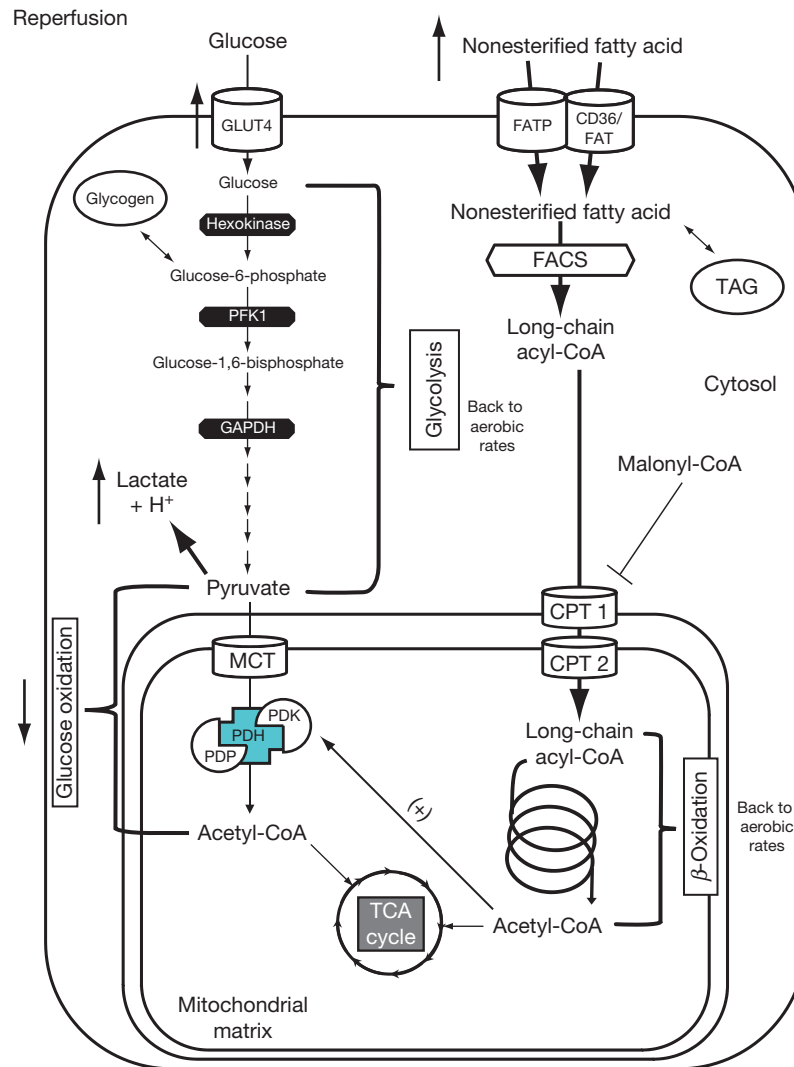


Figure 4 Energy metabolism in the reperfused heart. In the critical period immediately after successful reperfusion of the ischemic heart, glycolytic rates remain high, while fatty acid oxidation rates rapidly recover. Increased fatty acid oxidation is due in part to decreased malonyl-CoA content. Increased by-products of fatty acid oxidation, such as acetyl-CoA and NADH, activate PDK isoforms such as PDK4, which inhibit PDH and impair the recovery of glucose oxidation rates. Thus, pyruvate from glycolysis is not oxidized via PDH, and once more goes into lactate and H^+ s, intracellular acidosis continues, cardiac efficiency remains impaired, and this limits the recovery of cardiac function.

uncoupled from glucose oxidation in both the rat and mouse heart, H^+ production, and intracellular acidosis ensue, which limits the recovery of cardiac efficiency and function. Recent data in humans support the findings reported in rodents, as use of a comprehensive metabolomics approach revealed that fatty acid extraction persisted as the major fuel substrate contributing to myocardial ATP requirements in patients with coronary artery disease after reperfusion following cardiac surgery versus that in control patients.

Intracellular Acidosis and Cardiac Efficiency during Ischemia/Reperfusion

Cardiac efficiency is defined as the ratio of external cardiac power to energy expenditure by the left ventricle, and as

alluded to throughout this article, the metabolic alterations that take place in the heart during ischemia and reperfusion result in an intracellular acidosis that contributes to a reduction in cardiac efficiency.

The marked decrease in ATP production during ischemia leads to the inhibition of the Na^+/K^+ ATPase, which is responsible for the extrusion of three Na^+ ions in exchange for two K^+ ions and is crucial in regulating resting membrane potential. Impaired function of the Na^+/K^+ ATPase thus results in intracellular Na^+ overload. Impaired activity of the sarcoplasmic Ca^{2+} ATPase, which is responsible for the reuptake of Ca^{2+} following myocyte contraction, contributes to Ca^{2+} overload. As Ca^{2+} is required for cardiac muscle contraction, it is a major determinant of the pathophysiology of ischemic and postischemic contractile dysfunction, via mechanisms involving decreased responsiveness of the contractile proteins to

activator Ca^{2+} . Intracellular acidosis itself impairs the response of the contractile filaments to Ca^{2+} , thereby contributing to the impaired recovery of function during reperfusion. The increased contribution of fatty acid β -oxidation to myocardial energy requirements occurs again at the expense of glucose oxidation during reperfusion, and in conjunction with the alterations in ionic homeostasis occurring during ischemia, primes the myocardium to further injury. The uncoupling of glycolysis from glucose oxidation during the postischemic period results in an increase in H^+ production and produces a large pH gradient across the sarcolemmal membrane, which promotes Na^+/H^+ exchange, further aggravating intracellular Na^+ overload. This in turn promotes reverse mode $\text{Na}^+/\text{Ca}^{2+}$ exchange, and the sequence of events associated with intracellular Ca^{2+} overload, including contracture, mitochondrial dysfunction, the activation of Ca^{2+} -dependent proteases, and cardiac myocyte cell death culminating in the impaired recovery of cardiac function and efficiency.

Optimization of Energy Metabolism to Treat Ischemic Heart Disease

Because the metabolic alterations that take place during ischemia/reperfusion have a major role in determining outcome, a number of studies have attempted to optimize metabolism in the heart as a novel approach to treat ischemic heart disease (see Table 1 for an overview of current metabolic agents and their effects). Via the Randle cycle, a decrease in fatty acid β -oxidation rates leads to a subsequent increase in glucose oxidation rates, which can improve the coupling of glucose metabolism, decrease H^+ production, and therefore increase cardiac efficiency and functional outcome. In addition, the oxidation of glucose produces more ATP per consumption of one oxygen molecule versus the oxidation of a fatty acid such as palmitate, which also contributes to the increase in cardiac efficiency.

Studies in humans have shown that inhibiting fatty acid β -oxidation with agents such as trimetazidine or perhexiline has beneficial effects in the treatment of angina, acute myocardial infarction, and even heart failure. Furthermore, direct stimulation of glucose oxidation with an agent such as dichloroacetate has also shown some benefit in the treatment of heart failure, but it is limited by a very short half-life.

In animal studies, a new approach to optimizing myocardial metabolism is to increase malonyl-CoA content in the heart, which would inhibit fatty acid β -oxidation rates by decreasing mitochondrial fatty acid uptake. This can be achieved by inhibiting malonyl-CoA decarboxylase (MCD), which is the primary enzyme responsible for the degradation of malonyl-CoA. A number of recent studies have shown that MCD inhibition increases malonyl-CoA levels and improves functional outcome during both angina and ischemia/reperfusion injury, which is associated with a significant increase in glucose oxidation rates and a decrease in H^+ production.

Summary

Overall, ischemia of the heart is associated with dramatic alterations in energy fuel metabolism. These metabolic alterations primarily involve a robust increase in glycolysis and a

Table 1 Metabolic agents used to treat ischemic heart disease

Metabolic agent	Metabolic effect	Clinical use
Perhexiline	CPT 1 inhibitor ↓ Fatty acid oxidation rates	Anti-anginal agent since the 1970s – limited clinical use
Trimetazidine	3-Ketoacyl-CoA thiolase inhibitor ↓ Fatty acid oxidation rates	Approved anti-anginal agent in over 80 countries
CBM-301106	MCD inhibitor ↓ Fatty acid oxidation rates	Improves functional recovery against ischemia/reperfusion in rodents No clinical data available
Dichloroacetate	PDK inhibitor ↑ Glucose oxidation rates	Currently under clinical trial Limited to acute use due to short half-life

decrease in oxidative metabolism that depends on the severity of the ischemia itself. However, fatty acids are the preferred substrate for what residual oxidative metabolism remains. These metabolic alterations persist during reperfusion of the ischemic heart, and thus in both scenarios glycolysis is severely uncoupled from glucose oxidation, which culminates in the production of H^+ s and the development of intracellular acidosis. Thus, during both ischemia and reperfusion, ATP is diverted away from the contractile function toward maintaining pH and other ionic disturbances, which impairs the heart's efficiency of converting energy metabolism into mechanical work. This is detrimental to the recovery of the heart during ischemia/reperfusion, and as such, novel therapies for ischemic heart disease are now focusing on optimizing heart metabolism in order to improve cardiac efficiency and function.

See also: **Bioenergetics:** Mitochondria in Myocardial Ischemia; **Metabolism Vitamins and Hormones:** Fatty Acid Oxidation; Structure and Regulation of Pyruvate Dehydrogenases.

Further Reading

- Dennis SC, Gevers W, and Opie LH (1991) Protons in ischemia: Where do they come from; where do they go to? *Journal of Molecular and Cellular Cardiology* 23: 1077–1086.
- Jaswal JS, Ussher JR, and Lopaschuk GD (2009) Myocardial fatty acid utilization as a determinant of cardiac efficiency and function. *Clinical Lipidology* 4: 379–389.
- Lopaschuk GD, Ussher JR, Folmes CDL, Jaswal JS, and Stanley WC (2010) Myocardial fatty acid metabolism in health and disease. *Physiological Reviews* 90: 207–258.
- Neely JR and Morgan HE (1974) Relationship between carbohydrate metabolism and energy balance of heart muscle. *Annual Review of Physiology* 36: 413–459.
- Randle PJ (1998) Regulatory interactions between lipids and carbohydrates: The glucose fatty acid cycle after 35 years. *Diabetes/Metabolism Reviews* 14: 263–283.
- Stanley WC (2002) Partial fatty acid oxidation inhibitors for stable angina. *Expert Opinion on Investigational Drugs* 11: 615–629.
- Young LH (2008) AMP-activated protein kinase conducts the ischemic stress response orchestra. *Circulation* 117: 832–840.

Metabolic Roles of Adiponectin

K Sun and P E Scherer, University of Texas Southwestern Medical Center, Dallas, TX, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Adipokine Adipocyte-derived secretory protein. They are important factors for regulating metabolism, inflammation and many other cellular processes.

AdipoR1, AdipoR2 AdipoR1 and AdipoR2 serve as receptors for adiponectin and mediate some of the biological effects of adiponectin.

Lipotoxicity Pathological accumulation of lipids in non-adipose tissues.

LMW, HMW Low- and high-molecular weight complexes of adiponectin.

Thiazolidinediones (TZDs) Synthetic ligands for activation of the nuclear receptor PPAR γ .

Characterization and Molecular Structure

Adiponectin is a circulating hormone secreted predominantly by adipocyte cells. It was originally referred to as adipocyte complement-related protein of 30 kDa (ACRP30), but was later also described as apM1, adipoQ, or GBP28. The molecular weight of the basic monomer is about 30 kDa. It has an amino-terminal signal sequence, a hypervariable region that differs from species to species, a collagenous domain, and a carboxy-terminal globular domain. Many published studies use a recombinant form of the protein that consists of only the globular domain of adiponectin (Figure 1).

In humans and rodents, adiponectin circulates in plasma at approximately $10 \mu\text{g ml}^{-1}$ and therefore contributes approximately 0.01% of total plasma protein. Adiponectin has been shown to play many roles in a plethora of physiological processes. Adiponectin levels in circulation are tightly linked to insulin resistance, obesity, diabetes, and cardiovascular disease. Adiponectin has a complex structure and forms basic homotrimeric building blocks. These trimers associate through disulfide bonds to form higher-order structures, which include low-molecular-weight (LMW) hexamers and high-molecular-weight (HMW) complexes of more than 540 kDa (Figure 2). All three forms are found in circulation in humans, while HMW and LMW forms are the more predominant forms in rodents. Several thousand papers report interesting correlations of adiponectin with a number of parameters reflecting metabolic health. These correlations generally hold up by measuring total plasma adiponectin. However, a number of papers point out stronger correlations of metabolic parameters with specific levels of adiponectin subcomplexes. Particularly, the HMW form of adiponectin has found widespread acceptance as a better correlative factor for insulin sensitivity. Alternatively, the ratio of HMW adiponectin/total adiponectin can also be correlated, in an excellent manner, with various aspects of the metabolic syndrome.

Regulation of Adiponectin Levels

Unlike other hormones such as insulin, the circulating adiponectin levels are tightly controlled and remain relatively stable within a narrow range, suggesting that there is very little tolerance

for fluctuations in plasma adiponectin levels. Normal circulating adiponectin levels range from 5 to $30 \mu\text{g ml}^{-1}$. This is particularly surprising in the light of a relatively rapid turnover of the protein in circulation, with plasma half-lives of 60 min or less. Even though adiponectin is primarily a product of adipocytes, there is a paradoxical decrease of adiponectin in the plasma of obese subjects with higher fat mass. In addition to correlations with total fat mass, adiponectin levels are also affected by body fat distribution. Plasma levels of adiponectin exhibit strong negative correlations with intra-abdominal fat mass. However, *ex vivo* analysis of adipose tissue in lean and obese subjects suggests lower adiponectin protein and mRNA levels in omental versus subcutaneous fat pads. Importantly though, it is unlikely that adiponectin levels in plasma are regulated transcriptionally in adipocytes. In fact, the main regulatory nodes for adiponectin release are likely to be found at the posttranslational level in the secretory pathway. Therefore, it is the rate of adiponectin release from the cells rather than a static mRNA or protein measurement associated with the tissue that is of relevance. Additional clinical studies have been carried out to investigate the effect of lipodystrophy, characterized by adipose tissue loss or adipose tissue misdistribution. The results show that both congenital and human immunodeficiency virus (HIV)-related lipodystrophies are associated with lower adiponectin levels. Treatments that reconstitute normal fat mass also normalize adiponectin levels.

Interestingly, circulating adiponectin and relative levels of the HMW complex are higher in females who have higher levels of body fat mass compared with males. This hints at the complex hormonal regulation of adiponectin that includes effects of estrogen and androgen that exert profound effects on adiponectin levels at various stages of development. There is also a nutritional component involved in the regulation of adiponectin levels, such that diets associated with improvements in insulin sensitivity, such as foods enriched in $\Omega 3$ fatty acids and soy-based diets, lead to higher adiponectin levels. Table 1 summarizes some of the critical factors known to affect adiponectin concentrations in plasma.

Adiponectin Receptors

Several putative receptors have been identified for adiponectin. The biological functions of adiponectin have been postulated

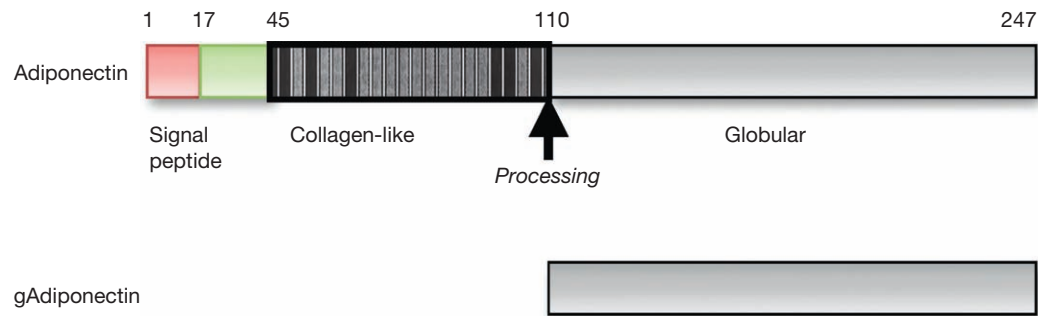


Figure 1 Schematic structure of adiponectin. Adiponectin is composed of an N-terminal collagen-like domain and a C-terminal globular domain.

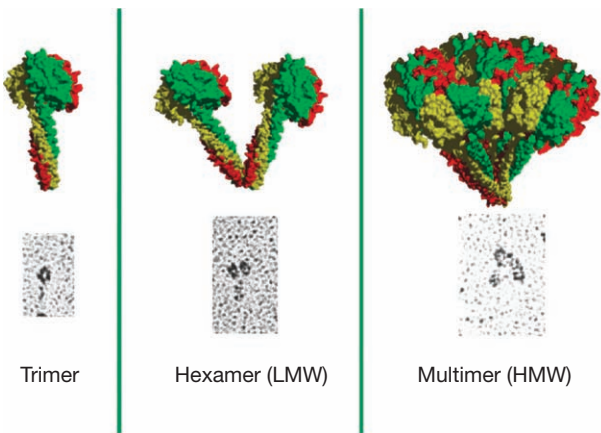


Figure 2 Schematic structures of the three most abundant complexes of adiponectin in circulation. The different colors represent individual adiponectin subunits in the basic trimeric blocks. The electron micrographs of purified complexes are also shown next to the schematic models.

Table 1 Regulation of adiponectin levels by different factors

<i>Upregulation</i>
Exercise
Mediterranean diet
Female
Lean subjects
Type 1 diabetes
PPAR γ agonists
Angiotension II receptor blockers
Heart failure
Kidney failure
<i>Downregulation</i>
Obesity
Insulin resistance
Oxidative stress
Sympathetic nervous system activity
Pro-inflammatory cytokines
Dyslipidemia
Carbohydrate-rich diet
Male
Smoking
Type 2 diabetes

to be exerted mainly through the two main adiponectin receptors referred to as AdipoR1 and AdipoR2. Both AdipoR1 and AdipoR2 are surface membrane proteins with seven transmembrane domains, have similar overall topology, and are both widely expressed in many tissues, including liver, muscle, and fat tissue, with most cell types preferentially expressing one of the receptors more prominently than the other. AdipoR1 has been shown to have a higher affinity for globular adiponectin and is more prominently expressed in skeletal muscle. By contrast, AdipoR2 is a medium-affinity receptor for both globular and full-length HMW adiponectin, more highly enriched in liver. Both AdipoR1 and AdipoR2 are thought to mediate increased AMP-activated protein kinase (AMPK) and peroxisome proliferator-activated receptor alpha (PPAR α) activities, leading to increased fatty acid oxidation and glucose uptake, but much more work needs to be done to better understand the role these receptors play in mediating improvements in insulin sensitivity. The expression levels of AdipoR1 and -R2 are decreased in insulin resistant *ob/ob* mice, as well as in the skeletal muscle of type 2 diabetic patients.

Adaptor protein containing PH domain, PTB domain, and Leucine zipper motif (APPL)1, an adaptor molecule with a pleckstrin homology domain and a phosphotyrosine-binding domain, may be a downstream mediator of activated adiponectin receptors. It has been shown to interact with the intracellular region of AdipoRs, thereby creating a link to AMPK upregulation, glucose uptake, and β -oxidation.

A third receptor has been cloned and shown to be T-cadherin that can mediate adiponectin binding to endothelial cells. It is primarily thought of as a tethering receptor for adiponectin, since T-cadherin lacks an intracellular signaling domain. The more detailed involvement of T-cadherin in adiponectin signaling has yet to be described.

Physiological Functions of Adiponectin

There are a remarkable number of physiological functions that have been attributed to adiponectin. While there is no doubt about the fact that modest changes in adiponectin levels yield remarkable physiological changes, a unifying mechanism of action has yet to be proposed to account for the anti-inflammatory, insulin-sensitizing, anti-apoptotic, and pro-angiogenic actions.

Insulin Sensitivity Effect

One of the most important aspects of adiponectin is its association with improvements in insulin sensitivity. During the past decade, a large number of clinical studies have associated adiponectin as an important factor regulating whole-body insulin sensitivity. Decreased circulating adiponectin is strongly associated with decreased insulin sensitivity under a vast number of pathophysiological conditions. Beyond its usefulness as a mere marker for insulin sensitivity, studies in rodents have demonstrated that the manipulation of adiponectin levels have a direct causative relationship to insulin sensitivity. Adiponectin null mice show decreased hepatic insulin sensitivity, whereas transgenic overexpression of adiponectin or supplying the molecule in a recombinant form enhances insulin sensitivity. These effects are further exacerbated when the mice are exposed to a high-fat diet. The correlations with insulin sensitivity can be further refined both in rodents' and humans' samples by focusing on the relative levels of the three different complexes in circulation. Particularly, the levels of the HMW adiponectin form show improved correlations compared to total adiponectin levels, at least under some physiological conditions or pharmacological interventions. The underlying mechanisms for the improvements in insulin sensitivity remain to be further established, but adiponectin has potent anti-lipotoxic activities on the liver by redistributing hepatic lipids to adipose tissue.

Vasculature and Anti-Atherogenic Effects

Adiponectin has beneficial effects on vascular endothelium. Adiponectin directly stimulates the production of nitric oxide (NO) by endothelial cells, via PI3-dependent pathways, which enhance endothelial NO synthase (eNOS) activity. Apart from the beneficial effects on eNOS, adiponectin improves endothelial redox state by suppressing nicotinamide adenine dinucleotide phosphate oxidase-derived superoxide generation. The HMW adiponectin also suppresses endothelial cell apoptosis and promotes vascular healing and angiogenesis.

Adiponectin may also be protective against the development of atherosclerosis. Adiponectin strongly inhibits the expression of adhesion molecules, including intracellular adhesion molecule-1, vascular cellular adhesion molecule-1, and E-selectin. The inhibition attributes directly to protection effect. In addition to the direct effects of adiponectin on the vasculature system, adiponectin may exert additional anti-atherogenic effects through the modification of lipid metabolism. Adiponectin levels are negatively correlated with serum triglycerides and small dense low-density lipoprotein, and positively correlated with high-density lipoprotein. This leads to the clinically widely studied inverse correlations with cardiovascular disease. High circulating levels of adiponectin are associated with a decreased risk of myocardial infarction in men. Major questions remain in this area, since end-stage cardiovascular disease is associated with a surprising increase in circulating levels of adiponectin. In addition, studies on the anti-atherogenic properties of adiponectin yielded mixed results.

Anti-Inflammatory Effect

Adiponectin reduces expression of pro-inflammatory cytokines, such as tumor necrosis factor (TNF α) and C-reactive protein. It also downregulates the expression of macrophage-attracting adhesion molecules, inhibits the activation of inflammatory signaling cascades, and prevents the production of excessive mitochondrial reactive oxygen species that lead to oxidative stress damage. Clinical observations have demonstrated a strong inverse relationship between serum adiponectin levels and inflammatory markers in a large number of studies. Anti-inflammatory drugs, such as TNF α neutralizing compounds and PPAR γ agonists, increase circulating adiponectin levels.

Adiponectin and Adiponectin Receptors as Therapeutic Targets?

Based on their beneficial effects on many aspects of human physiology, adiponectin and adiponectin receptors may lend themselves as excellent therapeutic targets for the treatment of insulin resistance, inflammation, and cardiovascular disease. Recombinant adiponectin treatment has only been done in preclinical models. Due to its short half-life, high serum concentrations, and complex tertiary and quaternary structure, adiponectin is not an attractive recombinant therapeutic protein. An alternative strategy is to develop agents that may induce adiponectin expression and secretion. To date, the most effective pharmacological intervention resulting in an increase in circulating levels are the PPAR γ agonists (thiazolidinediones (TZDs)). These PPAR γ agonists have been widely used to treat type 2 diabetes. Plasma adiponectin levels, especially HMW adiponectin, dramatically increase in response to TZD exposure. In fact, the full therapeutic potential of these PPAR γ agonists critically depends on their ability to induce adiponectin.

Since AdipoR1 and -R2 are downregulated in the obese and insulin-resistant state, strategies aimed at restoring normal levels of the receptors seem worthwhile. More importantly, small molecule activators of these receptors promise to be a powerful strategy to induce the full activity for these receptors. The screening for small compounds that enhance or mimic adiponectin action through modulation of expression and/or function of either adiponectin or its receptors may therefore represent powerful therapeutic strategies for targeting insulin resistance, type 2 diabetes, metabolic syndrome, and cardiovascular disease.

See also: [Metabolism Vitamins and Hormones: Adipogenesis; Insulin Mechanisms/Metabolic Actions.](#)

Further Reading

- Asterholm IW and Scherer PE (2010) Enhanced metabolic flexibility associated with elevated adiponectin levels. *American Journal of Pathology* 176(3): 1364–1376.
- Berg AH, Combs T, Du XL, Brownlee M, and Scherer PE (2001) The adipocyte-secreted protein Acrp30 enhances hepatic insulin action. *Nature Medicine* 7(8): 947–953.

- Kawano J and Arora R (2008) The role of adiponectin in obesity, diabetes, and cardiovascular disease. *Journal of the CardioMetabolic Syndrome* 4: 44–49.
- Kusminski CM and Scherer PE (2009) The road from discovery to clinic: Adiponectin as a biomarker of metabolic status. *Clinical Pharmacology and Therapeutics* 86(6): 592–595.
- Scherer PE (2006) Adipose tissue: From lipid storage compartment to endocrine organ. Outstanding Scientific Achievement Award Lecture. *Diabetes* 55: 1537–1545.
- Trujillo ME and Scherer PE (2005) Adiponectin – journey from an adipocyte secretory protein to biomarker of the metabolic syndrome. *Journal of Internal Medicine* 257: 167–175.
- Yamauchi T and Kadowaki T (2008) Physiological and pathophysiological roles of adiponectin and adiponectin receptors in the integrated regulation of metabolic and cardiovascular diseases. *International Journal of Obesity* 32: s13–s18.
- Ziemke F and Mantzoros CS (2010) Adiponectin in insulin resistance: Lessons from translational research. *American Journal of Clinical Nutrition* 91: 258–261.

Metabolic Roles of the Orexigenic and Anorexigenic Neuropeptides

G J Morton and M W Schwartz, University of Washington, Seattle, WA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Blood–brain barrier A barrier between circulating blood and the brain that only allows certain substances to pass through and enter the brain.

Central nervous system (CNS) Part of the nervous system that consists of the brain and spinal cord.

Energy homeostasis The biological process whereby energy intake is matched to its expenditure to promote the stability of body fuel stored as fat.

Hyperphagia An increase in food intake.

Insulin resistance A condition in which the normal biological response to insulin is impaired.

Satiety The state of being satisfactorily full.

Introduction

Each of us makes decisions several times per day about whether and when to eat, and how much to eat in any given meal. Although such decisions are often made with little conscious thought, they are nonetheless shaped by a myriad of biological, environmental, and lifestyle factors. Among these are social variables, learned preferences, time of day, emotional state, stress, nutritional status, and the smell, taste, and sight of food. Consequently, both the number of calories consumed in any given meal, and the total calories consumed on any given day can vary considerably. Despite this inherent variability, energy intake is matched to energy expenditure over long time intervals so as to promote stability in the amount of body fuel stored as fat, through a process termed as ‘energy homeostasis’. The central nervous system (CNS) plays a key role in this process through its ability to sense and integrate information from afferent signals regarding energy stores and nutrient availability, and then to orchestrate appropriate adjustments of energy balance. During times of negative energy balance, for example, the brain initiates adaptive metabolic responses to stimulate food intake and reduce energy expenditure that ultimately enable the recovery of lost weight. Presumably arising over millions of years of mammalian evolution, energy homeostasis offers a straightforward and compelling explanation for the frustration experienced by dieters seeking to maintain weight loss achieved through voluntary caloric restriction. Conversely, during times of overnutrition and positive energy balance, the brain initiates biological responses that include reduced energy intake and increased energy expenditure that protect against pathological weight gain. In this overview, we focus on CNS mechanisms that respond to hormonal- and nutrient-related signals involved in energy homeostasis.

Energy Homeostasis

More than 50 years ago, Kennedy first hypothesized the existence of signals that circulate in direct proportion to body fat and act in the CNS to inhibit food intake, thus creating a negative feedback loop to promote stability of body fat mass. Two hormones postulated to act as adiposity signals are the pancreatic β -cell hormone, insulin, and the adipocyte

hormone, leptin. Both hormones circulate at levels proportionate to body adiposity and enter the CNS in proportion to their plasma level. Leptin receptors and insulin receptors are expressed in an overlapping distribution within key brain areas, such as the hypothalamus, which are responsible for the control of food intake and energy expenditure, and the administration of peptide directly into the brain reduces food intake, while promoting increased energy expenditure, thus reducing body weight. Conversely, deficiency of either hormone or its CNS receptor causes hyperphagia and obesity. [Figure 1](#) describes a model in which body adiposity is regulated by a negative feedback loop involving signals that circulate in proportion to body fat, and act in the brain to inhibit food intake and stimulate energy expenditure.

The discovery of leptin and, subsequently, that leptin deficiency causes obesity both in humans as well as mice raised hopes for a new approach to the treatment of human obesity. Most forms of obesity are associated with at least some degree of leptin resistance. Consequently, clinical trials of leptin as a weight-loss agent in obese humans have yielded largely disappointing results. Although several mechanisms have been invoked to explain obesity-associated leptin resistance, including an impaired ability of circulating leptin to either cross the blood–brain barrier or activate its signal transduction pathway, a clear understanding of both the mechanisms underlying this phenomenon and their relevance to obesity pathogenesis is still awaited.

Hypothalamic Neuronal Systems Controlling Energy Homeostasis

Arcuate Nucleus

The hypothalamic arcuate nucleus (ARC) emerged early on as an area of particular importance to energy homeostasis. Situated adjacent to the floor of the third ventricle and the median eminence (a circumventricular organ that lacks a functional blood–brain barrier), ARC neurons can sense and respond to both circulating signals and input from neurons located in other brain areas. Within the ARC are two distinct neuronal populations that exert potent, opposing effects on food intake and energy expenditure in response to a wide assortment of hormonal and nutrient-related inputs.

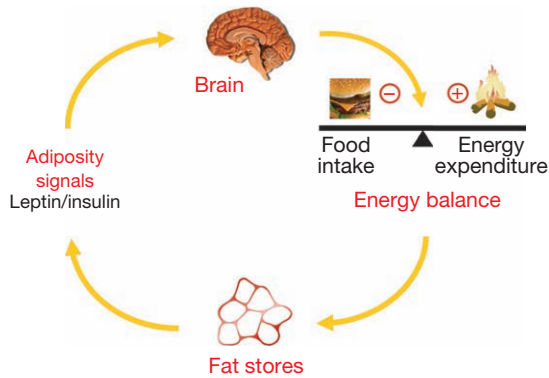


Figure 1 Negative feedback regulation of body fat mass.

Melanocortin Neurons

Among leptin-sensitive ARC neuronal populations involved in energy homeostasis are those that express proopiomelanocortin (POMC), a precursor polypeptide that is cleaved to generate α -melanocyte stimulating hormone (α -MSH), an agonist of neuronal melanocortin receptors (Mc3r/Mc4r). Activation of these receptors in hypothalamic areas, including the ventromedial and paraventricular nuclei, as well as in brainstem areas such as the nucleus of the solitary tract (NTS), inhibits food intake, increases energy expenditure, and promotes weight loss. A role for melanocortin neurons in energy homeostasis is supported by the observations that (1) these cells are potently stimulated by leptin, (2) leptin-induced weight loss requires melanocortin receptor signaling, and, most importantly, (3) genetic or pharmacological inhibition of POMC neurons or signaling via melanocortin receptors causes obesity in humans as well as in rodent models.

Neuropeptide Y/Agouti-Related Protein Neurons

Adjacent to POMC cells in the hypothalamic ARC is a distinct neuronal subset that expresses two peptides: neuropeptide Y (NPY) and agouti-related protein (AgRP), which promote hyperphagia, positive energy balance, and weight gain. NPY is among the most potent of orexigenic agents and acts via NPY Y1 and Y5 receptors to stimulate food intake. With repeated intracerebroventricular (icv) administration, NPY readily induces obesity and insulin resistance. Also synthesized within these neurons is AgRP, an inverse agonist of Mc3r/Mc4r that reduces endogenous melanocortin signaling. Similar to NPY, central administration of AgRP or pharmacological melanocortin receptor antagonists promotes hyperphagia, weight gain, and insulin resistance. Unlike POMC neurons, NPY/AgRP neurons are inhibited by leptin and insulin (Figure 2). Therefore, in conditions of reduced leptin signaling (e.g., during fasting or diet-induced weight loss), NPY/AgRP neurons are activated, whereas POMC neurons are inhibited, a combination of responses that reduce neuronal melanocortin signaling via multiple mechanisms, thus promoting a state of positive energy balance until lost weight has been recovered. In addition, NPY/AgRP neurons inhibit POMC neurons through release of the inhibitory neurotransmitter γ -aminobutyric acid (GABA), and loss of GABA tone is implicated in the potent

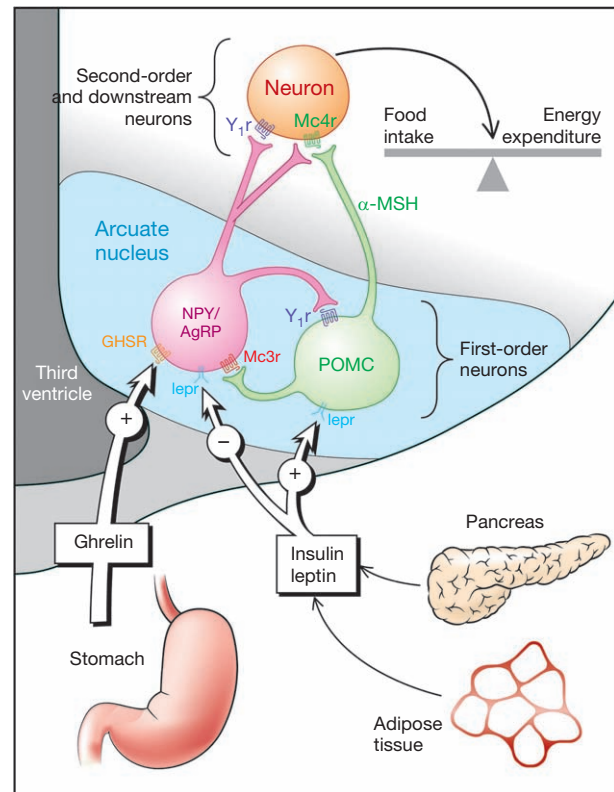


Figure 2 Model for hypothalamic regulation of energy homeostasis. Adapted from Barsh GS and Schwartz MW (2002) Genetic approaches to studying energy balance: Perception and integration. *Nature Reviews Genetics* 3: 589–600, with permission.

anorexia observed following acute, selective destruction of NPY/AgRP neurons in mouse models. These neurons are also targets of several gut-derived peptides, including ghrelin and PYY_{3-36} , and respond to nutrient-related signals, including oleic acid and glucose.

Several observations support a key role for these ARC neurons in energy homeostasis. First, leptin-deficient *ob/ob* mice exhibit sustained hyperphagia associated with reduced hypothalamic expression of POMC messenger RNA (mRNA) and increased expression of NPY and AgRP mRNA, and each of these responses is reversed upon leptin replacement. Second, Mc4r-deficient mice develop a hyperphagic obesity syndrome accompanied by leptin resistance, and the catabolic action of leptin is similarly blocked in normal animals by melanocortin receptor antagonists. Selective leptin receptor deletion from NPY/AgRP neurons also causes excess weight gain (although mild), and selective deletion of insulin receptors from these neurons causes peripheral insulin resistance (mainly at the level of the liver).

Ventromedial Hypothalamus

Brain lesioning and electrical stimulation studies conducted more than half a century ago implicated neuronal populations in the ventromedial hypothalamus (VMH), the lateral hypothalamic area (LHA), and the paraventricular nucleus (PVN) as key areas for food intake control (Figure 3). These classic

studies identified the VMH as a satiety center, as lesions to the VMH resulted in hyperphagia and obesity, while electrical stimulation of the VMH reduced food intake. As leptin receptors are expressed in the VMH and leptin activates neurons in this brain area, the VMH has emerged as a key brain area mediating leptin's anorexic effects. Microinjection of leptin directly into the VMH suppresses food intake, while conversely, selective deletion of leptin receptors from neurons expressing steroidogenic factor-1 (which is expressed selectively by VMH neurons) results in mice with a phenotype characterized by hyperphagia, obesity, glucose intolerance, and insulin resistance. Second, the VMH plays a role in the regulation of melanocortin signaling. Recent evidence suggests that brain-derived neurotrophic factor (BDNF) neurons in the VMH are regulated by both nutritional state, as well as by input from leptin and melanocortin s, and mice with reduced expression of the BDNF receptor, TrkB tyrosine kinase (TrkB), exhibit a phenotype characterized by obesity, hyperphagia, and increased body length, similar to *Mc4r*-deficient mice. In addition, a subset of VMH neurons sends strong excitatory inputs to POMC neurons in the ARC. Taken together, these data suggest that neurons in the VMH are integral components of energy homeostasis that act in part to transduce input from leptin and melanocortins into changes of energy balance.

Lateral Hypothalamic Area

In contrast to the VMH, the LHA is traditionally referred to as a hunger center, as stimulation of the LHA induces food intake, whereas its disruption causes hypophagia and weight loss. Within this brain area are two distinct neuronal populations that are functionally linked to ARC – POMC and NPY neurons – one that expresses melanin concentrating hormone (MCH) and another that produces orexin/hypocretin. Both

neuronal populations are orexigenic, as icv administration of either MCH or orexin stimulates feeding. Similarly, expression of genes encoding both MCH and orexin is increased by fasting and mice deficient in MCH are lean and hypophagic, while targeted deletion of orexin results in narcolepsy, a condition characterized by sudden daytime sleeping, suggesting that orexin signaling also plays a role in sleep, as well as feeding.

Paraventricular Nucleus

In addition to their projections to the VMH and the LHA, NPY/AgRP neurons and POMC neurons in the ARC also project extensively to the PVN, and bilateral PVN lesions cause hyperphagia and obesity, while electrical stimulation of this brain area inhibits food intake. These observations imply the existence of PVN neurons that play a physiological role to inhibit food intake and constrain weight gain. Consistent with this hypothesis, PVN neuronal subsets express both *Mc3r* and *Mc4r*, and several neuropeptides that reduce food intake and body weight, including thyrotropin-releasing hormone (TRH), corticotrophin-releasing factor (CRF), and oxytocin (OXY), are also synthesized in this key brain area. Administration of CRF and OXY into the brain reduces food intake, while leptin-induced anorexia is blocked by administration of either a CRF- or oxytocin-receptor antagonist. Thus, coordinate regulation of the activity of discrete neuronal cell types in multiple hypothalamic areas allows the brain to balance the activity of anabolic and catabolic neurocircuits in order to maintain energy homeostasis.

Integration of Adiposity and Satiety Signals

In comparison to our rapidly increasing understanding of hypothalamic neurocircuits involved in energy homeostasis, relatively less is known about how changes in energy intake are adjusted to meet changing energy needs on a meal-to-meal basis. Since food in general is consumed in discrete bouts or meals, hypothalamic mechanisms involved in energy homeostasis must ultimately impinge upon systems responsible for initiation and/or termination of individual meals. Indeed, recent evidence suggests that long-term adiposity signals are integrated with input from short-term, meal-related signals to influence meal size, and that changes of meal size are a key mechanism for adaptive changes of food intake in the service of energy homeostasis (Figure 4).

Cholecystokinin

Cholecystokinin (CCK) is a gut-derived peptide that is released from I-cells following nutrient entry into the duodenum, and it induces satiety by activating vagal afferent fibers that terminate in the NTS. In addition to promoting the termination of a meal, CCK also slows gastric emptying, reduces gastric acid secretion and promotes contraction of the gall bladder. Although exogenous CCK reliably reduces the size of individual meals, repeated CCK administration does not effectively lower body weight. This is because as weight is lost and plasma levels of insulin and leptin decline, an increase in the number of meals occurs that negates the effect of CCK to reduce meal size. This

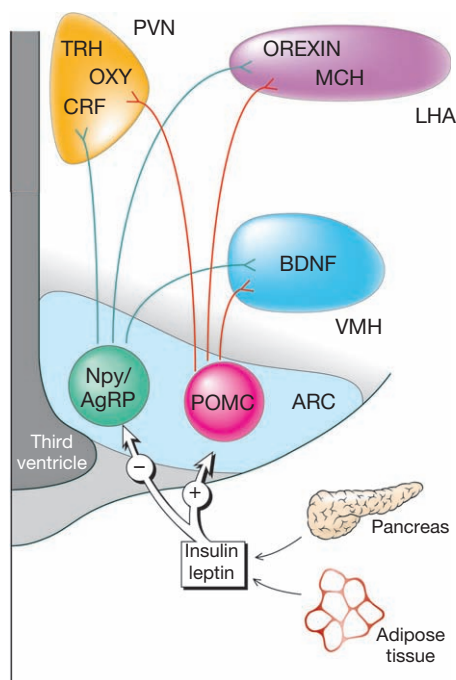


Figure 3 Hypothalamic neurocircuits involved in energy homeostasis.

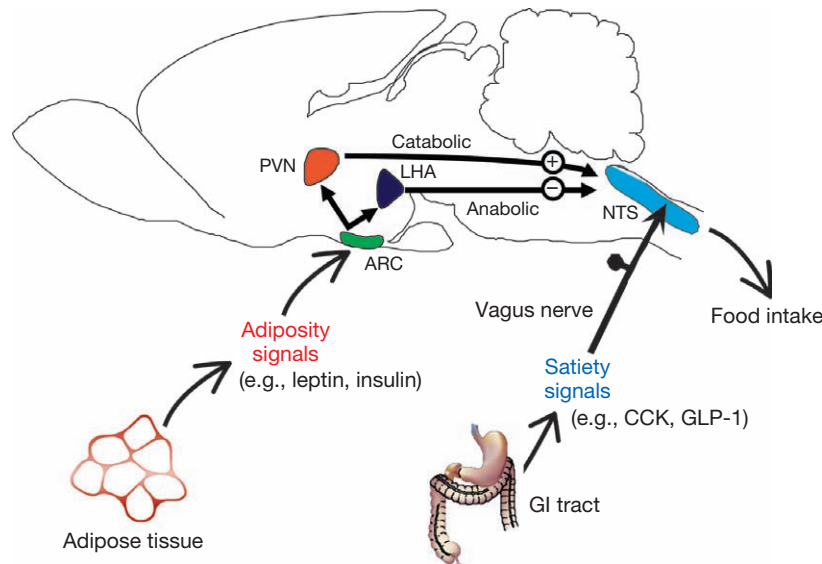


Figure 4 Integration of adiposity signals and satiety signals.

finding highlights the inherently limited potential of meal-related satiety peptides to promote therapeutic weight loss.

Several lines of evidence suggest that leptin reduces food intake, at least in part, by increasing sensitivity to meal-related signals such as CCK. Leptin-mediated food intake suppression is characterized by reduced meal size and exogenous leptin potentiates the satiety effect of CCK. Conversely, in conditions of reduced or deficient leptin signaling, increased energy intake is due, in part, to an increase in meal size associated with reduced sensitivity to satiety signals. Therefore, if an individual goes on a calorically restricted diet and loses weight, the decline in circulating levels of both insulin and leptin ultimately reduces sensitivity to endogenous satiety signals such as CCK. Consequently, the amount of food consumed during each meal tends to increase, amounting to a major contributor to positive energy balance and recovery of lost weight. In addition to CCK, there are a host of other meal-related signals released from the gastrointestinal tract, several of which are already known to interact with adiposity signals including glucagon-like peptide-1 (GLP-1) and amylin.

Ghrelin

Whereas CCK and other meal-related peptides promote meal termination, the gastric hormone ghrelin, stimulates feeding. Plasma ghrelin levels are increased with fasting and fall rapidly following a meal. Moreover, exogenous administration of ghrelin when given systemically or centrally stimulates feeding and activates NPY/AgRP neurons in the ARC. Thus, during weight loss, elevated plasma ghrelin levels, in combination with reduced insulin and leptin signaling, contribute to the regain of lost weight. One caveat to this widely accepted model of ghrelin biology is that for biological activity, ghrelin must be acylated (specifically, it is octanoylated), and deacylation occurs rapidly in plasma, such that most immunoreactive ghrelin in circulation may be devoid of bioactivity. Whether circulating, biologically active ghrelin is affected by changes in

energy balance, and whether such changes play a vital role in energy homeostasis, await further study.

Summary

In conclusion, afferent input from adiposity signals, such as insulin and leptin, convey information regarding long-term energy stores to key brain areas that integrate this input with information regarding short-term energy availability from nutrient-related signals. Thus, when the brain perceives that both short- and long-term energy needs are met, anabolic neurocircuits that stimulate food intake and reduce energy expenditure are inhibited, whereas catabolic pathways that limit food intake and increase energy expenditure are activated so as to maintain energy homeostasis. Adaptive changes in the activity of this reciprocally regulated, opposing, dual-effector control system lie at the core of the energy homeostasis system.

See also: **Metabolism Vitamins and Hormones: Anorexigenic and Orexigenic Gut Peptides; Regulation of Body Weight by Malonyl-CoA in the CNS; Signaling: Evolving Concepts of Leptin; Melanocortin System.**

Further Reading

- Barsh GS and Schwartz MW (2002) Genetic approaches to studying energy balance: Perception and integration. *Nature Reviews Genetics* 3: 589–600.
- Lenard NR and Berthoud HR (2008) Central and peripheral regulation of food intake and physical activity. Pathways and genes. *Obesity (Silver Spring)* 16(supplement 3): S11–S22.
- Morton GJ, Cummings DE, Baskin DG, Barsh GS, and Schwartz MW (2006) Central nervous system control of food intake and body weight. *Nature* 443: 289–295.
- Schwartz MW, Woods SC, Porte D, Jr., Seeley RJ, and Baskin DG (2000) Central nervous system control of food intake. *Nature* 404: 661–671.
- Woods SC (2009) The control of food intake: Behavioral versus molecular perspectives. *Cell Metabolism* 9: 489–498.

Metabolite Channeling: Creatine Kinase Microcompartments

U Schlattner and M Tokarska-Schlattner, University Joseph Fourier and Inserm, Grenoble, France
T Wallimann, Institute of Cell Biology, ETH Zürich, Switzerland

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Uwe Schlattner and Theo Wallimann, volume 2, pp. 646–651, © 2004, Elsevier Inc.

Glossary

Metabolic compartmentation Segregation of intermediates and enzymes of a metabolic pathway by membranes, binding to a specific surface or direct interaction in protein complexes allowing metabolite channeling.

Metabolite channeling Local transfer of metabolic intermediates between sequential enzymes or transport reactions without equilibration with bulk solution.

Microcompartment Structural unit allowing metabolic compartmentation, also called metabolon.

Mitochondrial contact sites Close adhesions of inner and outer mitochondrial membrane that can be observed by electron microscopy and can be isolated as a

separate microcompartment. Contact sites consist of multilipid/protein complexes with variable composition and are involved in energy transduction (e.g., containing adenylate translocator (ANT)/voltage-dependent anion channel (VDAC) or ANT/mitochondrial creatine kinase (MtCK)/VDAC) or in protein import.

Mitochondrial permeability transition

pore A multienzyme complex, probably composed of VDAC, ANT, Bax, cyclophilin, MtCK, and others, that is crucially involved in early events that trigger apoptosis-like release of cytochrome *c* and other apoptosis-inducing factors into the cytosol.

Subcellular microcompartments, consisting of multienzyme complexes embedded within the cellular, highly viscous matrix, associated with the cytoskeleton, or situated along membranes, are operating according to exclusion principles and favor preferred pathways of intermediates. This process, called metabolite or substrate channeling, is defined as transfer of intermediates between sequential enzymes without equilibration of these metabolites with the surrounding bulk solution. Such an association between two or more sequential enzyme or transport reactions in a microcompartment, forming a distinct functional pool of intermediates, is also called functional coupling. It can be considered as a general mechanism to increase efficiency of sequential reactions in a metabolic pathway. As metabolite channeling leads to segregation of a metabolic pathway from other cellular reactions, it represents a specific kind of metabolic compartmentation similar to that operating within membrane-separated organelles or by restricted two-dimensional diffusion at surface boundary layers. Here, metabolite channeling is described with special emphasis on high-energy phosphate channeling by creatine kinase (CK), the phosphocreatine circuit or shuttle, and the mitochondrial CK (MtCK) isoenzyme.

Subcellular Microcompartments and Mechanisms of Metabolite Channeling

Life most likely originated autotrophically *de novo* in metabolic complexes organized on FeS₂ (pyrite) mineral surfaces, the earliest form of microcompartments. Likewise, a cell is by no means represented best by a well-mixed bag of enzymes, behaving in complete equilibrium according to solution kinetics. Because of the intricate structural and functional organization of living cells, enzymes and metabolites do not behave as if they were freely diffusible in solution. Instead, they may form

structurally, functionally, and temporally defined subcellular microcompartments, either via strong static or via fickle, dynamic interactions with other enzymes, proteins, or subcellular structures. Such a structural organization of pathway components is a general prerequisite for metabolite channeling. It may involve (1) huge covalently linked enzyme complexes (or multifunctional enzymes) such as fatty acid synthase (FAS), (2) kinetically stable multienzyme complexes such as pyruvate dehydrogenase (PDH) or bacterial and plant tryptophan synthase (TS), (3) more dynamic, reversibly associating enzymes such as glycolytic complexes containing glyceraldehyde phosphate dehydrogenase (GAPDH) or glycerol phosphate dehydrogenase (GPDH), or (4) colocalization on subcellular particles or biological membranes. These associations allow the transfer of intermediates between the channeling components by different mechanisms: (1) physical hindrance or electrostatic effects prevent mixing with bulk solution and drive a directed diffusion (e.g., TS and FAS), (2) sequential covalent binding to very close active sites in the reaction sequence (e.g., PDH), (3) transfer of noncovalently bound intermediates between active sites (e.g., nicotinamide adenine dinucleotide dehydrogenase), and (4) transfer in dynamic multienzyme complexes (GAPDH and GPDH) or in an unstirred membrane surface layer (MtCK). These mechanisms can be fulfilled in both, static and dynamic enzyme associations. However, although static associations allow for an almost perfect or tight channeling of metabolites, dynamic channeling is often only partial or leaky.

Advantages of Metabolite Channeling

Sequestering of intermediates in a microcompartment through channeling provides kinetic and regulatory advantages not only for the reaction sequence itself (see [Figure 1](#)) but also for cellular metabolism. In general: (1) enzyme reaction rates

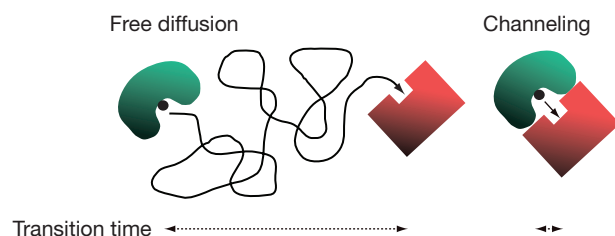


Figure 1 Free diffusion vs. metabolite channeling. Compartmentation of a reaction sequence without equilibration with bulk solution leads to shorter transition times and further advantages (see text).

and equilibria are controlled by local and enzyme-bound substrates, rather than bulk-phase substrate concentrations; (2) for a readily reversible reaction, local supply of substrate and removal of product may drive the reaction in the desired direction; (3) sequestered intermediates are present at high local concentrations and an apparently low K_m for these intermediates can be observed with the channeling complex compared to the nonchanneling situation measured with isolated components; (4) metabolites are isolated from competing reactions, for example, between anabolic and catabolic pathways; (5) the lifetime of the intermediate in the solvent phase is shortened relative to free diffusion, which may be essential in the case of unstable intermediates; (6) in certain cases, the unfavorable energetics of desolvating the substrate that precedes binding to the enzyme is avoided; (7) channeling components can be regulated by modulators that affect enzyme associations; and (8) a larger degree of metabolic control of the overall flux of the reactions can be achieved, for example, via feedback regulatory mechanisms such as substrate activation, product inhibition, and cooperativity.

Subcellular Targeting of Glycolytic Multienzyme Complexes

In muscle, glycolytic enzymes are targeted to the actin-containing thin filaments at the sarcomeric I-band region where they form highly complex metabolons. The I-band in *Drosophila* flight muscle contains a multienzyme complex consisting of GDPH-1, aldolase, and GAPDH. By elegant experiments with transgenic *Drosophila* expressing GDPH-3 instead of GDPH-1, it could be shown that all three glycolytic enzymes no longer colocalize in the I-band to form a microcompartment. Even though the full complement of active glycolytic enzymes was still present, their failure to colocalize in the sarcomer resulted in the inability to fly. Thus, correct targeting and formation of multienzyme complexes that lead to functionally coupled microcompartments and substrate/product channeling seem to be a prerequisite for proper function of glycolysis and ultimately for correct muscle function. In mammalian cells, CK is also participating in the glycolytic metabolon.

More recently, metabolomic studies have envisaged a comprehensive analysis of cellular metabolites, their changes in time, as well as the modeling of these processes. Such systems biology approaches have the potential to reveal many more cases of dynamic and time-resolved metabolite-channeling events within the cellular metabolic network.

Compartmentation of CK Isoenzymes and Channeling of High-Energy Phosphates

The CK/Phosphocreatine Circuit or Shuttle

One fundamental requirement of life is energy supply. Cellular energy demand and supply are balanced, and tightly regulated for economy and efficiency of energy use. CK is a major enzyme of higher eukaryotes that copes with high and fluctuating energy demands to maintain cellular-energy homeostasis, in general, and to guarantee stable, locally buffered adenosine triphosphate (ATP)/adenosine diphosphate (ADP) ratios, in particular.

The enzyme catalyzes the reversible phosphoryl transfer from ATP to creatine (Cr) to generate ADP and phosphocreatine (PCr). Thus, CK is able to conserve energy in the form of metabolically inert PCr and vice versa, to use PCr to replenish global as well as local cellular ATP pools. As PCr can accumulate to much higher cellular concentrations than ATP, the CK/PCr system constitutes an efficient and immediately available cellular energy buffer. In addition, tissue-specific CK isoenzymes are located in the cytosol (dimeric muscle-type MM-CK and brain-type BB-CK) and within the mitochondrial intermembrane space (sarcomeric MtCK and ubiquitous MtCK, both forming octamers and dimers). CK isoenzymes are often associated with sites of ATP supply, where they generate PCr, or with sites of ATP consumption, where they regenerate ATP by using PCr. Thus, together with the faster diffusion rate of PCr as compared to ATP, the CK/PCr system also supports an intrinsic energy-transfer system (CK/PCr circuit or shuttle), coupling sites of energy generation (oxidative phosphorylation or glycolysis) with sites of energy consumption (Figure 2). This circuit is particularly important in large and/or polar cells, such as spermatozoa where diffusional limitations of adenine nucleotides, especially ADP, along the sperm tail become especially apparent.

Channeling with Cytosolic CK

Cytosolic CK is only partially soluble. A significant fraction is structurally and functionally associated or colocalized with different, structurally bound ATPases or ATP-regulated processes. These include (1) different ion pumps in the plasma membrane, (2) the sarcomeric M-band of the myofibrils in muscle, (3) the calcium pump of the muscular sarcoplasmic reticulum, and (4) the ATP-gated K^+ -channel. In all these cases, PCr is used for the local regeneration of ATP, which is directly channeled from CK to the ATP-consuming ATPase without major dilution by the surrounding bulk solution. Only in some cases, the physical basis for the metabolite channeling is known. For example, MM-CK uses a charge clamp consisting of four lysine residues to specifically bind to partner molecules, myomesin and M-protein in the M-band. This allows for an isoenzyme-specific association of MM-CK with the sarcomeric M-band of the myofibrillar apparatus. There, MM-CK is ideally positioned in the middle of the acto-myosin overlap zones, situated symmetrically on both sides of the M-band, to regenerate *in situ* the ATP that is hydrolyzed during muscle contraction.

Cytosolic CK is structurally associated with the key regulatory enzyme of glycolysis, phosphofructokinase, which itself

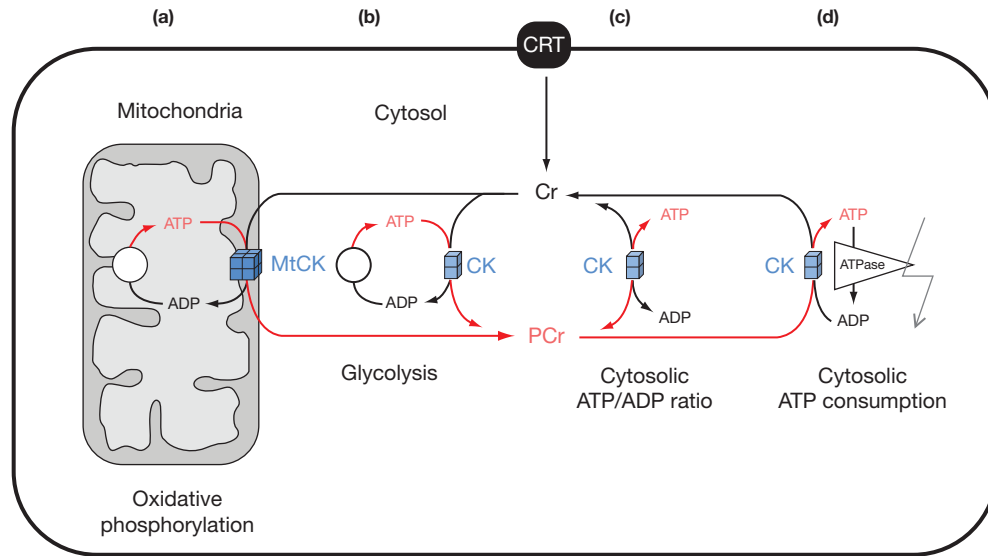


Figure 2 The creatine kinase (CK)/phosphocreatine (PCr) system. Isoenzymes of CK (blue) are found in different compartments such as mitochondria (octameric MtCK, (a)) and cytosol (dimeric CK, (b)–(d)) in soluble form (c) or associated to a different degree to ATP-delivering ((a) and (b)) or -consuming processes (d). A large cytosolic PCr pool up to 30 mM is built up by CK using ATP from oxidative phosphorylation like in heart (a) or glycolysis like in fast-twitch glycolytic muscle (b). PCr is then used to buffer global (c) and local (d) ATP/ADP ratios. In cells that are polarized and/or have very high or localized ATP consumption, these CK isoenzymes, together with easily diffusible PCr, also maintain an energy shuttle between ATP-providing or -consuming processes ((a) and (d)). Metabolite channeling occurs where CK is associated with ATP-providing or -consuming transporters, pumps, or enzymes ((a), (b), and (d)). Creatine (Cr) is synthesized in only few cell types (e.g., liver and kidney) and has to be taken up from the blood stream by a specific Cr transporter (CRT) that is highly expressed by Cr-containing target cells.

is regulated by ATP. Likewise, structural and functional interactions of cytosolic CK with the PAR-1 receptor of the thrombin-signaling pathway and with the ATP-gated K^+ -channel, respectively, have been demonstrated. A tight functional coupling of CK to ATPases, for example, actomyosin ATPase and ion pumps, such as the K^+ /Na $^+$ -ATPase or the Ca $^{2+}$ -ATPase, has the advantage (1) that product inhibition of the ATPase by ADP and H $^+$ is avoided, because the latter are both substrates of the CK reaction ($PCr + ADP + H^+ \leftrightarrow Cr + ATP$) and (2) that the high free energy of ATP hydrolysis (ΔG_{ATP}) at sites of ATP hydrolysis is preserved by keeping locally very high ATP/ADP ratios due to coupling of CK with those ATPases *in situ* and, thus, preventing energy dissipation that would otherwise be caused by transport of ATP and mixing it with the bulk surrounding. Interestingly, the strongest phenotype of CK double knockout mice, which no longer express cytosolic MM-CK and mitochondrial MtCK in muscle, is characterized by significant difficulties with intracellular calcium handling and muscle relaxation, emphasizing the physiological importance of the CK system for the energetics of intracellular calcium homeostasis, in general, and the delivery of ATP to the energetically demanding Ca $^{2+}$ -ATPase, in particular.

Channeling in Energy-Transducing Mitochondrial Microcompartments

MtCK forms mainly large, cuboidal octamers (Figure 3) that are present (1) between the outer and inner mitochondrial membrane (the so-called intermembrane space of

mitochondria) and preferentially localized at the so-called mitochondrial contact sites between outer and inner mitochondrial membrane, as well as (2) in the cristae space (see Figure 3 inset). The kinase catalyzes the direct transphosphorylation of intramitochondrial-produced ATP and Cr from the cytosol into ADP and PCr. ADP then enters the matrix space to stimulate oxidative phosphorylation, giving rise to mitochondrial recycling of a specific pool of ATP and ADP, while PCr is the primary high-energy phosphoryl compound that leaves mitochondria into the cytosol. The molecular basis for such directed metabolite flux is channeling between the large, cuboidal MtCK octamer and two transmembrane proteins, adenylate translocator (ANT) and mitochondrial porin or voltage-dependent anion channel (VDAC). ANT is an obligatory antiporter for ATP/ADP exchange across the inner mitochondrial membrane, while VDAC is a nonspecific, potential-dependent pore in the outer mitochondrial membrane. The MtCK-linked metabolite channeling is based on (1) colocalization, (2) direct interactions, and (3) diffusion barriers. MtCK tightly binds to cardiolipin, an acidic phospholipid that is specific for the mitochondrial inner membrane. As ANT molecules are situated in a cardiolipin patches, this leads to colocalization and metabolite channeling between both proteins, MtCK and ANT, in the cristae as well as in the intermembrane space (Figure 3). MtCK in the intermembrane space further interacts with outer membrane phospholipids and VDAC, thus virtually cross-linking inner and outer membrane and contributing to the mitochondrial contact sites. Increasing the external calcium concentrations strengthens the interaction of MtCK with VDAC, which may improve channeling under cytosolic calcium overload as occurring at

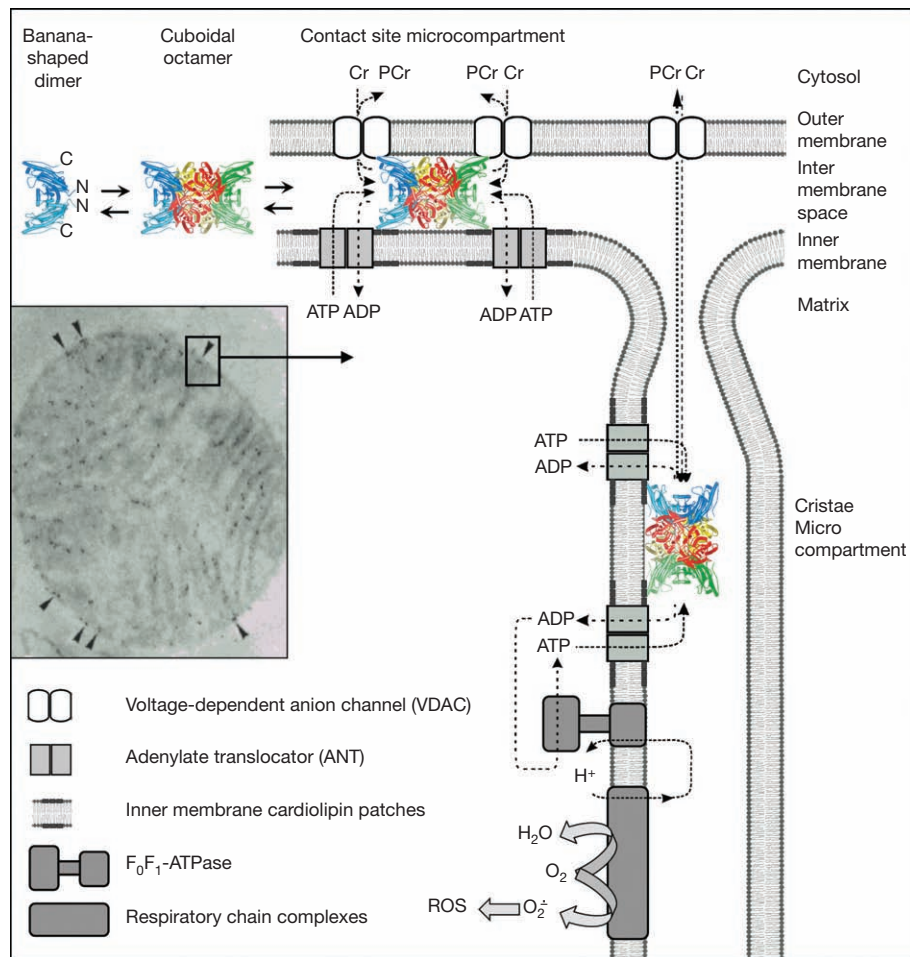


Figure 3 (Inset) Dual localization of mitochondrial CK by electron microscopy. Post-embedding immuno-gold labeling of MtCK in a mitochondrion from photoreceptor cells of chicken retina, showing localization of MtCK in the peripheral intermembrane space (arrows) and along the cristae membranes. Scheme: microcompartments and metabolite channeling of mitochondrial CK. After the import of nascent MtCK over the mitochondrial outer membrane and cleavage of the targeting sequence, MtCK first assembles into dimers. Dimers rapidly associate into octamers (top left), and although this reaction is reversible, octamer formation is strongly favored by high MtCK concentrations and MtCK binding to acidic phospholipids. MtCK is then found in two locations: (1) in the so-called mitochondrial contact sites associated with ANT and VDAC (top center) and (2) in the cristae associated with ANT only (at right). In contact sites, MtCK simultaneously binds to inner and outer membrane due to the identical top and bottom faces of the octamer. Binding partner in the inner membrane is the twofold negatively charged cardiolipin, which allows a functional interaction with ANT. This carrier is situated in cardiolipin membrane patches that can be induced by MtCK itself. In the outer membrane, MtCK interacts with other acidic phospholipids and, in a calcium-dependent manner, directly with VDAC. Substrate and product fluxes between MtCK and the associated proteins are depicted as arrows. In contact sites, this substrate channeling allows for a constant supply of substrates and removal of products at the active sites of MtCK. In cristae, only ATP/ADP exchange is facilitated through direct channeling to the MtCK active site, while Cr and PCr have to diffuse along the cristae space to reach VDAC. Further channeling may take place within mitochondrial supercomplexes that include carriers, ATPase and respiratory chain complexes (lower part). The active ATP/ADP exchange maintained by coupled MtCK favors ATP generation by the F₀F₁-ATPases and, thus, proper functioning of the respiratory chain, which could otherwise generate elevated levels of superoxide and reactive oxygen species (ROS).

low-cellular-energy state. Some studies suggest that only the membrane-bound, octameric form of MtCK is able to maintain the metabolite channeling described above. Finally, the limited permeability of VDAC and, thus, of the entire outer membrane creates a dynamic microcompartmentation of metabolites in the intermembrane space that contributes to MtCK-linked channeling and separate mitochondrial ATP and ADP pools. Similar to MtCK, hexokinase is able to use intramitochondrially produced ATP by binding to VDAC from the cytosolic mitochondrial surface at contact sites containing only ANT and VDAC. The direct functional coupling

of MtCK to oxidative phosphorylation can be demonstrated with oxygraph respirometry on skinned muscle fibers from normal and transgenic mice lacking MtCK. Although in normal muscle fibers Cr stimulates mitochondrial respiration, this phenomenon is missing in fibers of MtCK knockout mice.

By favoring ATP/ADP exchange through ANT in the mitochondrial inner membrane, cytosolic Cr and the MtCK reaction not only stimulate the rate of mitochondrial respiration, but also lead to improved coupling between respiration, ATP synthesis, and mitochondrial energy channeling, which finally reduces formation of potentially deleterious reactive

oxygen species (ROS; [Figure 3](#), lower part). Both effects of MtCK-related metabolite channeling, stimulation of mitochondrial respiration and lowering of ROS production, may contribute to the remarkable cell- and specifically neuroprotective action of Cr supplementation *in vitro* and *in vivo*. This is an instructive example of how efficient, multiple metabolite channeling events can be beneficial for human health.

Oxidative damage of MtCK, induced by reactive oxygen and nitrogen species, generated under cellular stress situations, for example, in infarcted heart or under chemotherapeutic intervention by anthracyclines, leads to inactivation and dimerization of MtCK, as well as to dissociation of the enzyme from the mitochondrial inner membrane. Thus, important prerequisites for efficient channeling of high-energy phosphates by MtCK are negatively affected. These events contribute to cardiac energy failure and specific anthracycline cardiotoxicity.

Mt CK, Intramitochondrial Inclusions, and Low-Cellular-Energy State

MtCK is an indicator for cellular low-energy stress, that is, the expression of this enzyme is highly upregulated in patients with mitochondrial myopathies in the so-called ragged-red



Figure 4 Intramitochondrial inclusions in patients with mitochondrial myopathies consist of mitochondrial MtCK. Immunoelectron histochemistry of human intramitochondrial crystalline ‘railway-track’ inclusions seen in patients with mitochondrial myopathies showing ‘ragged-red’ skeletal fibers as a hallmark of the disease. Note the grossly enlarged mitochondrion, here from a patient with Kearns–Sayre syndrome, displaying the typical regularly spaced intracristae inclusions (dark) surrounded by mitochondrial inner cristae membranes. The section has been stained by rabbit anti-MtCK antibodies, followed by colloidal gold-conjugated second antibody (in collaboration with Dr. AM Stadhouders, Nijmegen, The Netherlands). Note that the dark inclusion bodies are heavily and specifically stained by the small 10-nm gold particles, indicating a high propensity of MtCK at these locations. Isolation of such inclusions plus image processing of sections through them revealed that they consist mainly of crystalline MtCK octamers, which crystallize in sheet-like structures. Modified from Stadhouders AM, Jap PHK, Winkler H-P, Eppenberger HM, and Wallimann T (1994) Mitochondrial creatine kinase: A major constituent of pathological inclusions seen in mitochondrial myopathies. *Proceedings of the National Academy of Sciences of the United States of America* 91(11): 5089–5093.

skeletal muscle fibers, where mitochondrial volume and size are markedly increased and where characteristic intramitochondrial railway-track inclusions are observed as a hallmark of pathology. The latter were shown to consist of crystalline sheets of MtCK (see [Figure 4](#)). Similar MtCK inclusions can also be induced in animals by chronic Cr depletion, leading to cellular low-energy state. Even more generally, cellular low-energy stress, be it induced by chronic endurance training, fasting, Cr depletion, or pathologies in ATP generation, such as mitochondrial dysfunction seen in patients with mitochondrial cytopathies, induces a coordinated induction of energy gene expression to compensate for impaired energy supply and transport. In the case of MtCK, one of the most prominent among these genes, overexpression is such that it leads to crystallization of the enzyme in a pathological state as described above.

Mt CK and the Mitochondrial Permeability Transition Pore

MtCK, together with its substrate Cr, has been implicated in regulation of the mitochondrial permeability transition pore that is crucially involved in triggering apoptosis. This seems to be due to metabolite channeling in the MtCK/ANT microcompartment also, which maintains high ADP concentrations in the matrix space that are inhibitory for permeability transition. Thus, the CK system and its substrates seem to exert additional effects that are not necessarily directly coupled to improving cellular energetics. This may explain the remarkable cell- and neuroprotective effects of Cr supplementation that have been reported.

See also: **Bioenergetics:** Mitochondria: Source and Target of Free Radicals; Mitochondrial Channels; Mitochondrial Membranes, Structural Organization; Mitochondrial Metabolite Carrier Protein Family; Mitochondrial Outer Membrane and the VDAC Channel; P-Type Pumps: Na⁺,K⁺-ATPase.

Further Reading

- Agius L and Sherratt HSA (eds.) (1997) *Channelling in Intermediary Metabolism Research Monograph IX*. London: Portland Press.
- Dolder M, Walzel B, Speer O, Schlattner U, and Wallimann T (2003) Inhibition of the mitochondrial permeability transition by creatine kinase substrates. Requirement for microcompartmentation. *Journal of Biological Chemistry* 278: 17760–17766.
- Kay L, Nicolay K, Wieringa B, Saks V, and Wallimann T (2000) Direct evidence for the control of mitochondrial respiration by mitochondrial creatine kinase in oxidative muscle cells *in situ*. *Journal of Biological Chemistry* 275: 6937–6944.
- Meyer LE, Machado LB, Santiago AP, et al. (2006) Mitochondrial creatine kinase activity prevents reactive oxygen species generation: Antioxidant role of mitochondrial kinase-dependent ADP recycling activity. *Journal of Biological Chemistry* 281: 37361–37371.
- Ovádi J (1995) *Cell Architecture and Metabolic Channeling*. Heidelberg: Springer.
- Ovádi J and Srere PA (2000) Macromolecular compartmentation and channeling. *International Review of Cytology* 192: 255–280.
- Schlattner U, Forstner M, Eder M, et al. (1998) Functional aspects of the X-ray structure of mitochondrial creatine kinase: A molecular physiology approach. *Molecular and Cellular Biochemistry* 184: 125–140.

- Schlattner U, Tokarska-Schlattner M, Ramirez S, et al. (2009) Mitochondrial kinases and their molecular interaction with cardiolipin. *Biochimica et Biophysica Acta* 1788: 2032–2047.
- Schlattner U, Tokarska-Schlattner M, and Wallimann T (2006) Mitochondrial creatine kinase in human health and disease. *Biochimica et Biophysica Acta* 1762: 164–180.
- Srere PA and Knull HR (1998) Location–location–location. *Trends in Biochemical Sciences* 23: 319–320.
- Srivastava DK and Bernhard SA (1986) Metabolite transfer via enzyme–enzyme complexes. *Science* 234: 1081–1086.
- Stadhouders AM, Jap PHK, Winkler H-P, Eppenberger HM, and Wallimann T (1994) Mitochondrial creatine kinase: A major constituent of pathological inclusions seen in mitochondrial myopathies. *Proceedings of the National Academy of Sciences of the United States of America* 91(11): 5089–5093.
- Wallimann T, Tokarska-Schlattner M, Neumann D, et al. (2007) The phospho-creatine circuit: Molecular and cellular physiology of creatine kinases, sensitivity to free radicals and enhancement by creatine supplementation. In: Saks VA (ed.) *Molecular Systems Bioenergetics: Energy for Life, Basic Principles, Organization and Dynamics of Cellular Energetics*, pp. 195–264. Weinheim: Wiley-VCH.
- Wallimann T, Wyss M, Brdiczka D, Nicolay K, and Eppenberger HM (1992) Intracellular compartmentation, structure and function of creatine kinase isoenzymes in tissues with high and fluctuating energy demands: The 'phosphocreatine circuit' for cellular energy homeostasis. *Biochemical Journal* 281: 21–40.
- Winkel BS (2004) Metabolic channeling in plants. *Annual Review of Plant Biology* 55: 85–107.

Metalloproteases

D S Auld, Harvard Medical School, Boston, MA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Aminopeptidase Protease catalyzing the cleavage of amino acids from the N terminus of polypeptides.

Carboxypeptidase Protease catalyzing the cleavage of amino acids from the C terminus of polypeptides.

Dipeptidyl peptidase Protease that removes two amino acids at a time from the N or C terminus of the polypeptide.

Endoprotease Protease that acts internally on a polypeptide chain.

Protease or proteinase Enzyme catalyzing the hydrolysis of peptide bonds in proteins.

Metalloproteases are a class of enzymes that use protein-bound metal ion(s) and a coordinated water molecule to catalyze the addition of water to a peptide bond. Proteases are necessary for the survival of all living creatures, and are encoded by ~2% of the genes in all kinds of organisms. They are an exceptionally important group of enzymes in biology, medical research, and biotechnology.

Functional Classification

Scientists who work in the general area of proteolytic enzymes have given much latitude to the naming of such enzymes. At present, they are frequently referred to as proteases, proteinases, and peptidases, in an interchangeable manner. Some of the terms used are more illustrative of their function. Exopeptidases cleave one or a few amino acids from the N or C terminus of a polypeptide, whereas endoproteases act internally on the polypeptide chain. Aminopeptidases and carboxypeptidases act on the N and C terminus of the polypeptide, respectively, usually cleaving off one amino acid at a time, but some are capable of removing two (dipeptidyl peptidases) or three (tripeptidyl peptidases) amino acids at a time. These enzymes are often classified according to their general mechanism type as structural/functional information becomes available. This classification results in four major types: aspartic, cysteine, metallo-, and serine, each named for a critical component of the active site. In the case of metalloproteases, the metal that is most frequently found is zinc. However, manganese, cobalt, and iron are also observed for some specific cases. Proteases are often named according to a specific physiological substrate that they hydrolyze. Angiotensin I converting enzyme (ACE I) is a zinc metalloprotease that cleaves the C-terminal dipeptide from angiotensin I to produce the potent vasopressor, octapeptide, angiotensin II and inactivates bradykinin by the sequential removal of two dipeptides from its C terminus. The role of this enzyme in blood pressure control and water and salt metabolism was identified through the use of potent inhibitors of its action. It is also defined as a zinc dipeptidyl carboxypeptidase. Tumor necrosis factor- α (TNF- α) is a cytokine that induces protective inflammatory reactions and kills tumor cells. Soluble TNF- α is released from its membrane-bound precursor by a membrane-anchored proteinase, identified as a multidomain metalloproteinase

called TNF- α -converting enzyme (TACE). This enzyme is also named as a disintegrin-like and metalloproteinase site (ADAMS) 17 endopeptidase for a classification based on structural considerations. Frequently, such a substrate-based classification requires a very general term when the substrate specificity is only known in general terms, such as matrix metalloproteinase (MMP).

Structural Classification

During the past decade, the advent of sequencing the protein through its gene and the ability to obtain three-dimensional (3D) structures on proteins allowed the recognition of metalloprotease families. These studies permit several different classifications based, for example, on the type of metal site present and/or similarity in overall protein fold and/or amino acid sequence similarities.

Metal Sites

The metal that is found in most metalloproteases is zinc. The function of these enzymes is therefore related to the chemistry of zinc, which is quite versatile. It has a remarkably adaptable coordination sphere that allows it to accommodate a broad range of coordination numbers and geometries when forming complexes. If the zinc retains a positive charge after ligating to protein side chains, it can act as a Lewis acid in catalysis. If a zinc-bound water is converted to hydroxide, the zinc site can act as a base or nucleophile in catalysis. In this sense, the catalytic zinc sites are amphoteric. The oxidation/reduction properties of its neighboring transition metals are a major cause of their ligand exchange rates, amphoteric properties, and coordination geometries. As zinc has a filled *d*-shell, it does not have oxidation/reduction properties, thus providing a stable metal ion species in a biological medium whose redox potential is in constant flux. There are now about six dozen X-ray diffraction and nuclear magnetic resonance (NMR) structures of individual metalloproteases. These structures provide standards of reference for the identity and nature of zinc ligands in other proteins, for which only the primary structure is known, and thus allow the formation of metalloprotease families. Three types of distinct zinc-binding motifs emerge

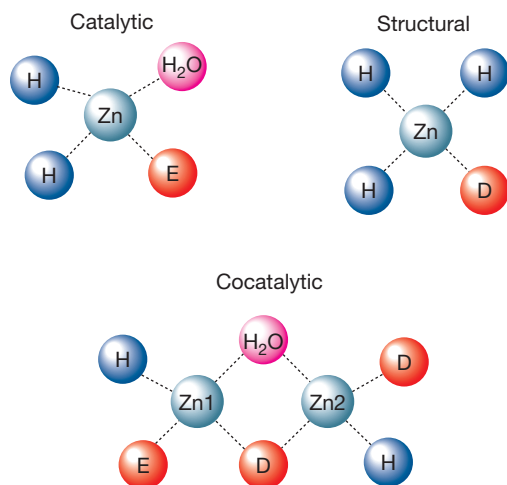


Figure 1 Zinc-binding sites in metalloproteases. Schematic shown is for a 'catalytic' site, represented by carboxypeptidase A, the 'structural' site for the matrix metalloproteases, and the 'cocatalytic' site of *Aeromonas* aminopeptidase.

from these analyses: 'catalytic', 'cocatalytic', and 'structural' sites (Figure 1). In zinc metalloenzymes, the most common amino acids that supply ligands to these sites are His, Glu, Asp, and Cys. Thus far, Cys has been found as a ligand in only one metalloprotease.

Catalytic Sites

3D structures of the catalytic sites of about five dozen metalloproteases show that zinc forms complexes with any three nitrogen (His) and oxygen (Glu or Asp) donors, with His being the predominant amino acid chosen by a ratio of ~2:1 over Glu plus Asp. Histidine (usually the N ϵ 2 nitrogen) may be chosen because of its capacity to disperse charge through H-bonding of its nonligating nitrogen. The carboxylate anion of Glu and Asp ligands will reduce the charge on the metal, making it more difficult for the metal-bound water to ionize and for the metal to act as a Lewis acid catalyst. The first two metal ligands are separated by a short amino acid spacer with the third ligand being supplied by a longer spacer. In catalytic zinc sites of proteases, the ligands are generally supplied from within an α -helix or a β -sheet that places restraints on the coordination of the metal by the side chains of the amino acids. The overall length of such sites can be as small as 11 amino acids, as is observed in the astacin super family of zinc proteases.

Catalytic zinc sites are usually four or five coordinates in metalloproteases and the geometry in the free state is frequently distorted tetrahedral or trigonal bipyramidal. Water is always a ligand to these sites. In principle, the zinc and its bound water molecule can be involved in catalysis by a number of different means (Figure 2). Metalloproteases generally have a Glu or His residue within H-bonding distance of the zinc-bound water that could play the role of a general acid/base catalyst or may stabilize the transition state of a tetrahedral intermediate. Alternatively, the direct ionization of the metal-bound water can lead to nucleophilic catalysis by zinc hydroxide, while displacement of water

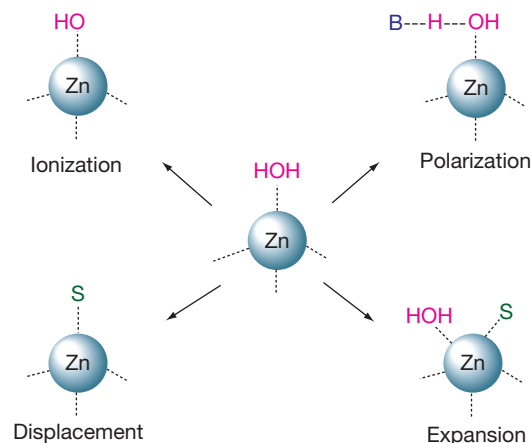


Figure 2 The role of the metal and bound water in catalysis. The water can ionize either (1) without or (2) with the help of the adjacent base supplied by an amino acid residue. The zinc-bound water can also (3) be displaced by substrate or (4) the coordination sphere be expanded upon interaction with substrate.

or expansion of the coordination sphere can result in Lewis acid catalysis by the catalytic zinc atom. The three protein ligands, their spacing, and secondary interactions with neighboring amino acids, in conjunction with the vicinal properties of the active center created by protein folding, are all critical to the mechanisms through which zinc is involved in catalysis.

Structural Sites

Metalloproteases frequently use either disulfides or calcium ions to aid in stabilizing the structure of the enzyme. Structural metal sites have four protein ligands and no metal-bound water molecule. Cys ligands, followed by His, are the preferred ligands in such sites in zinc metalloenzymes. However, only one Cys-containing zinc site is observed in the class of metalloproteases. Human picornavirus endoprotease 2A binds one zinc to three Cys and one His in the sequence Cx₁Cx₅₇Hx₁C. One other metal site that fits the criteria for a structural zinc site has been observed in the matrix metalloproteases and has a signature of Hx₁Dx₁₂Hx₁₂H (Figure 1). This site may indirectly affect the activity of the enzyme. The amino acids adjacent to the third and fourth His ligands to the structural zinc site provide a number of hydrophobic residues that border a catalytic Glu residue. These residues could provide a hydrophobic environment for the Glu carboxylate group and thus raise its pK_a and allow it to play a role in stabilizing the transition state in catalysis. The zymogen form of the MMPs has a Cys ligand in place of the zinc-bound water in the catalytic zinc site, thus having the properties of a structural zinc site.

Cocatalytic Sites

These sites in metalloenzymes contain two or three metals in close proximity with two of the metals bridged by a side-chain moiety of a single amino acid residue, such as Asp,

Glu, His, or a carboxylated lysine, LysCO, and sometimes a water/hydroxide molecule (Figure 1). Thus far, structures for amino-, carboxy-, and a tripeptidase have been reported. Asp and Glu and, sometimes, an additional water molecule are the preferred bridging ligands for these sites. The distance between the metals is dictated by the type of bridging ligand, 2.8–3.6 Å, for the carboxylate. No Cys ligands have been reported for these sites.

The bridging amino acids and H₂O could have critical roles in catalysis. Thus, their dissociation from either or both metal atoms during catalysis could change the charge on the metal promoting its action as a Lewis acid or allowing interaction with an electronegative atom of the substrate. Alternatively, the bridging ligand might participate transiently in the reaction as a nucleophile or general acid/base catalyst. In this manner, the metal atoms and their associated ligands would play specific roles in each step of the reaction that works in concert to bring about catalysis.

The ligands to these sites often come from nearly the entire length of the protein. The metals may therefore be important to the overall fold and stability of the protein as well as catalytic function. The secondary structural elements, α -helices and β -sheets, again play a strong role in supplying the ligands. However, in contrast to the catalytic zinc sites, the ligands often reside one or two amino acids from the N or C terminus of an α -helix or a β -sheet. These positions should allow more flexibility in forming the cocatalytic site.

Some of the cocatalytic sites contain metal ions other than zinc ions. The *Escherichia coli* methionine aminopeptidase-1, MetAP-1, the hyperthermophile *Pyrococcus furiosus* methionine aminopeptidase-2, and human methionine aminopeptidase-2 have been isolated as dicobalt enzymes and the *E. coli* proline aminopeptidase as the dimanganese enzyme. The physiological metal for these enzymes is still not certain. Zinc works as well as cobalt in the yeast aminopeptidase-1, and recent studies of the *E. coli* MetAP-1 indicate that it functions as an Fe(II) enzyme.

Genetic Origin and Ligand Nature

The identification of the ligands to these metal sites and the spacing characteristics in combination with modern sequencing techniques have led to further classifications of the metalloproteases. An excellent source of information on proteolytic enzymes is the *Handbook of Proteolytic Enzymes* and its associated website of MEROPS. This protease database groups peptidases into three levels – clans, families, and an individual peptidase. A clan contains all the peptidases that are believed to have diverged from a single evolutionary origin. In terms of the metalloproteases, they may have similar catalytic metal sites and overall folds. The clans can be further divided into families of proteases that show significant similarities in amino acid sequence. Finally, a member of a family is an individual peptidase. The Release 9.3 (December, 2010) of MEROPS gives information on 60 846 sequences of metallopeptidases and groups them into 21 clans, 58 families, and 724 identified metallopeptidases with links to 165 3D structures.

Physiological Function

These enzymes are important to a wide range of physiological processes that are too numerous to cover completely herein. Carboxypeptidase A (CPD A), was the first zinc protease and second zinc enzyme discovered. The zinc exopeptidases, CPD A and B, compliment the action of the serine endopeptases, trypsin, and pepsin, in the degradation of food proteins. The serine proteases cleave proteins internally either at positions of basic amino acids such as Lys or Arg (trypsin) or in hydrophobic aromatic amino acids such as Phe or Trp (pepsin). The zinc exopeptidases then liberate the essential amino acids Lys and Phe from the C terminus of the peptides generated by the endopeptases. The combination of structural and functional studies on these metalloproteases has revealed the presence of other members of this family that are important to hormone processing, immune/inflammatory activity, and blood-clotting functions. Information on the active sites of CPD A led to the development of specific inhibitors of ACE decades before a crystal structure was known for it, based on the proposition that both enzymes contained a zinc-bound water. Drugs for control of hypertensin were identified by these means. Methionyl aminopeptidases, MetAP, remove the N-terminal methionine residue from newly transcribed polypeptides. The physiological importance of these enzymes is underscored by the lethal nature of the removal of the MetAP gene from a number of bacterial sources. The *Handbook of Proteolytic Enzymes* along with its associated MEROPS website is an excellent source for detailed information on metalloproteases.

Metalloprotease Inhibition

The crucial nature of these enzymes to physiological processes has led to considerable research efforts on discovering their natural inhibitors that could be involved in their regulation. The MEROPS website contains a good compilation of such inhibitors. The importance of the metal and its bound water is the key to inhibiting the metalloproteases (Figure 3). Inhibition can be accomplished by one of the three ways: (1) displace the water by an inhibitor binding to the metal,

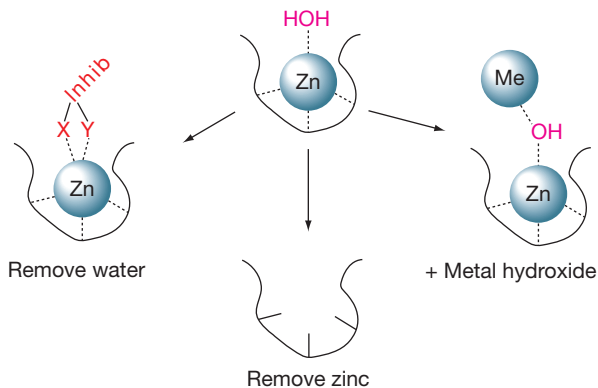


Figure 3 Inhibition of zinc proteases is accomplished by either (1) removing the water, (2) removing the metal, or (3) binding a metal hydroxide complex.

(2) remove the metal, or (3) bind a metal hydroxide complex to the catalytic metal and an active site residue.

Displacement of the Metal-Bound Water

The importance of zinc proteases in both normal and pathological processes makes them targets for drug design. The fact that zinc can readily form four-, five-, and six-coordinate complexes has led to the design of a number of synthetic inhibitors that act by displacing the metal-bound water. These studies have shown that the strength of the chelator–metal interaction increases the potency of the agent as a drug. The inhibitors are designed to have both structural specificity for the active site and a war head that will bind to the catalytic metal. Specificity is incorporated on either side of the metal-binding ligand. Popular warheads are hydroxamates, thiolates, and phosphoryl groups as chelating ligands. Carboxyalkyl and phosphinic acid groups are used as transition state or tetrahedral intermediate mimics.

Catalytic Chelators

It should also be possible to inhibit zinc proteases by removing the metal. This is often not very specific and many common chelators, such as ethylenediamine tetra-acetic acid (EDTA), are ineffective. However, D-penicillamine, a drug used to treat Wilson's disease and rheumatoid arthritis, has been shown to catalyze the release of zinc from CPD A. In the presence of a second scavenger chelator, such as EDTA or the physiological chelator, thionein (apo-metallothionein), the inhibition is quite potent. From a physiological perspective, such a mechanism of inhibition would be essentially irreversible, given the low levels of free zinc believed to be present in the body.

Inhibition by Metal Hydroxide Complexes

Many zinc proteases are inhibited by adding zinc to the assays. In the case of the exopeptidase, CPD A, and the endoprotease, thermolysin, a combination of kinetic and structural studies have shown this inhibition is due to zinc hydroxide binding the ionized carboxylate of an active site glutamate residue

which then bridges to the catalytic zinc by displacing its bound water. In the case of CPD A, both zinc and lead inhibit in the form of $M(OH)Cl$ with submicromolar-inhibition constants. If multidentate ligands interact with the inhibitory metal and the enzyme-active site, significantly stronger inhibition is anticipated. Inhibition by such metal hydroxide complexes could be important to regulatory and/or toxicological processes of zinc proteases.

See also: **Protein/Enzyme Structure Function and Degradation:** Aminopeptidases; Aspartic Proteases; Cysteine Proteases; Proteases in Blood Clotting; **Signaling:** Angiotensin Receptors.

Further Reading

- Auld DS (2001) Zinc coordination sphere in biochemical zinc sites. *BioMetals* 14: 271–313.
- Auld DS (2001) Zinc sites in metalloenzymes and related proteins. In: Bertini I, Sigel A, and Sigel H (eds.) *Handbook on Metalloproteins*, pp. 881–959. New York, NY: Dekker.
- Auld DS (2004) Cocatalytic zinc sites. In: Messerschmidt A, Bode W, and Cygler M (eds.) *The Handbook of Metalloproteins*, vol. 3, pp. 416–431. Chichester: Wiley.
- Barrett AJ and Rawlings ND (2004) MEROPS the peptidase database. <http://merops.sanger.ac.uk/index.htm> (accessed November 2010).
- Chong CR and Auld DS (2000) Inhibition of carboxypeptidase A by D–penicillamine: Mechanism and implications for drug design. *Biochemistry* 39: 7580–7588.
- Lowther WT and Matthews BW (2000) Structure and function of the methionine aminopeptidases. *Biochimica et Biophysica Acta* 1477: 157–167.
- (1998) Metallopeptidases. In: Barrett AJ, Rawlings ND, and Woessner JF (eds.) *Handbook of Proteolytic Enzymes*, 2nd edn., pp. 268–288. London: Academic Press.
- Vendrell J, Querol E, and Aviles FX (2000) Metallo-carboxypeptidases and their protein inhibitors – structure, function and biomedical properties. *Biochimica et Biophysica Acta* 1477: 284–298.
- Woessner JF and Nagase H (2000) *Matrix Metalloproteinases and TIMPS*. Oxford: Oxford University Press.

Relevant Website

<http://merops.sanger.ac.uk> – MEROPS the peptidase database.

Metalloproteinases, Matrix

H Nagase, University of Oxford, Oxford, UK

G Murphy, University of Cambridge, Cambridge, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

α_2 -Macroglobulin A glycoprotein in blood of 720 kDa consisting of four identical subunits. It inhibits most endopeptidases regardless of their catalytic mechanism by physical entrapment of the enzyme molecule upon proteolytic attack on the macroglobulin.

ADAM Abbreviation for type I transmembrane proteins with a disintegrin and metalloprotease domain, some of which are involved in shedding of cell-surface molecules and have some overlapping properties shared with MMPs. While the primary structures of the metalloprotease domain are homologous to each other, about half of the members of this family do not contain the strictly conserved zinc-ion-binding motif, and are probably proteolytically inactive.

ADAMTS Abbreviation for proteins with a disintegrin and metalloprotease domain with type I thrombospondin motifs. The metalloproteinase domains are homologous to ADAMs, but they do not possess a transmembrane domain.

Extracellular matrix A complex network of secreted extracellular macromolecules (e.g., collagens,

proteoglycans, glycoproteins, fibronectin, and complex carbohydrates) which provides an organized lattice and helps to hold cells and tissues together in multicellular organisms.

Metzincin Metalloproteinases that contain the zinc-ion-binding motif HEXXHXXGXXH and a conserved methionine that forms a unique Met-turn structure, e.g., matrix metalloproteinases, astacins, repolysins, ADAMs, ADAMTSs, pappalysins, serralsins, and leishmanolysin.

MMP Abbreviation for matrix metalloproteinases that are secreted from cells or bound to the cell surface and degrade extracellular matrix molecules.

Proteinase Proteolytic enzymes, also known as proteases or endopeptidases, that hydrolyze internal peptide bonds of proteins and polypeptide chains (cf. exopeptidase, enzymes that hydrolyze one or a few amino acids from the N- or C-terminus of polypeptides).

TIMP An acronym for tissue inhibitors of metalloproteinases with molecular masses of 22–30 kDa that inhibit MMPs.

Matrix metalloproteinases (MMPs), also called matrixins, are a family of structurally related proteolytic enzymes in the Metzincin clan, containing a zinc ion liganded to three histidines in the active site. They are secreted from cells or are bound to the plasma membrane and hydrolyze extracellular matrix (ECM) and cell-surface molecules. As exemplified by the discovery of collagenase in the tadpole tail undergoing metamorphosis, MMPs play key roles in embryonic development, organ morphogenesis, ovulation, embryo implantation, nerve growth, bone remodeling, wound healing, angiogenesis, and apoptosis. Aberrant activities of MMPs are often associated with the progression of diseases such as cancer, arthritis, cardiovascular disease, nephritis, neuro-degenerative diseases, periodontal disease, skin ulceration, gastric ulcer, corneal ulceration, liver fibrosis, emphysema, fibrotic lung, and many others. Many cell types have the ability to synthesize MMPs but their production and turnover (transcription, translation, and cellular trafficking) are tightly controlled by growth factors, cytokines, physical stress to the cell, oncogenic transformation, hormones, and cell–cell and cell–ECM interactions. MMP activities are also regulated at the level of proteolytic and oxidative activation of the precursor zymogens, interaction with ECM and cell-surface molecules, and inhibition by endogenous inhibitors.

MMP Family

MMPs are found in vertebrates, fruit flies (*Drosophila melanogaster*), nematodes (*Caenorhabditis elegans*), sea urchins (*Paracentrotus lividus*), hydras (*Hydra vulgaris*), and plants (e.g., *Arabidopsis thaliana* and *Glycine max*). The human genome has 24 MMP genes (one gene duplication). There are two genes in *Drosophila*, six genes in *C. elegans*, and five genes in *Arabidopsis*. They are classified as the matrixin subfamily of metalloprotease family M10 in the MEROPS database. Vertebrate matrixins are assigned with MMP numbers and trivial names in many cases (Table 1). MMP numbers missing in the list (i.e., MMP-4, -5, -6, and -22) were found to be identical to one of the other MMPs. Expression of individual MMPs varies depending on cell types (Figure 1).

Most MMPs consist of a prodomain, a catalytic domain, a hinge (or linker), and a hemopexin domain. The signatures to assign proteinases to the MMP family are the cysteine switch motif with amino acid sequence PRCGXPD in the propeptide and the zinc-binding motif HEXGHXXGXXH in the catalytic domain, where the three histidines coordinate the zinc ion (Zn^{2+}). The primary structure of the catalytic domain is homologous to collagenase 1 (MMP-1). On the basis of substrate specificity sequence similarity and domain organization, MMPs are divided into six subgroups (Table 1 and Figure 2).

Table 1 Examples of substrates of mammalian MMPs

Enzyme	MMP	Human chromosome	Substrates
Soluble types			
<i>Collagenases</i>			
Interstitial collagenase (collagenase 1)	MMP-1	11q22–q23	Collagens I, II, III, VII, VIII, X, and XI, gelatin, entactin, tenascin, aggrecan, link protein, fibronectin, perlecan, IGFBP-2,-3,-5, α_1 PI, proIL-1 β , CTGF
Neutrophil collagenase (collagenase 2)	MMP-8	11q21–q22	Collagens I, II, and III, aggrecan, link protein, α_1 PI, substrate P
Collagenase 3	MMP-13	11q22.3	Collagens I, II, III, IV, IX, X, and XIV, gelatin, collagen telopeptides, fibronectin, aggrecan, tenascin, osteonectin, laminin, perlecan, CTGF, pro TGF β , MCP-3
Collagenase 4 (<i>Xenopus</i>)	MMP-18	–	Collagen I
<i>Gelatinases</i>			
Gelatinase A	MMP-2	16q13	Gelatin, collagens I, II, III, IV, V, VII, and X, fibronectin, laminin, aggrecan, elastin, tenascin, decorin, fibrillin, fibulin, galectin-3, plasminogen, α_1 PI, IL-1 β , IGFBP-3, FGFR1, MCP-3, pro TGF β , SDF-1, S100A8/9
Gelatinase B	MMP-9	20q12.2–q13.1	Gelatin, collagens IV, V, XI, XIV, XVIII, elastin, aggrecan, decorin, laminin, entactin, pro TGF β , α_1 PI, proIL-1 β , ICAM-1, galectin-3, IGFBP-3, IL-2 receptor α
<i>Stromelysins</i>			
Stromelysin 1	MMP-3	11q23	Collagens III, IV, V, IX, X, XI, XVIII, telopeptides of collagens I and II, gelatin, aggrecan, fibronectin, laminin, entactin, tenascin, decorin, myelin basic protein, α_1 PI, proIL-1 β , HB-EGF, CTGF, E-cadherin, IGFBP-3, activation of proMMP-1, proMMP-3, proMMP-7, proMMP-8, proMMP-9, proMMP-13, plasminogen, uPA
Stromelysin 2	MMP-10	11q22.3–q23	Collagen III, IV, and V, gelatin, fibronectin, elastin, aggrecan, activation of proMMP-1, proMMP-7, proMMP-8, proMMP-9
Stromelysin 3	MMP-11	22q11.2	α_1 PI (mouse MMP-11 has weak activities on collagen IV, gelatin, fibronectin, laminin, aggrecan), IGFBP-1
<i>Matrilysins</i>			
Matrilysin 1	MMP-7	11q21–q22	Collagen IV, XVIII, gelatin, aggrecan, link protein, elastin, fibronectin, vitronectin, laminin, SPARC, entactin, decorin, myelin basic protein, tenascin, fibulin, α_1 PI, proTNF α , E-cadherin, syndecan-1, pro α -defensin, activation of proMMP-1, proMMP-2, proMMP-9
Matrilysin 2	MMP-26	11p15	Collagen IV, fibronectin, fibrinogen, vitronectin, gelatin, α_1 PI, proMMP-9
<i>Others</i>			
Metalloelastase	MMP-12	11q22.2–q22.3	Elastin, collagen IV, gelatin, fibronectin, vitronectin, laminin, entactin, aggrecan, nidogen, myelin basic protein, osteopontin, α_1 PI, proTNF α , plasminogen, apolipoprotein a
(No name)	MMP-19	12q14	Collagen IV, laminin, entactin, tenascin, fibronectin, gelatin, aggrecan, COMP, IGFBP-3
Enamelysin	MMP-20	11q22.3	Ameloblastin, amelogenin, collagen XVIII, aggrecan, COMP
(No name)	MMP-21	10	Gelatin, α_1 PI
CA-MMP	MMP-23*	1p36.3	Gelatin
(No name)	MMP-27	11q24	–
Epilysin	MMP-28	17q21.1	Casein, Nogo-A, NCAM-1
Membrane types			
<i>Transmembrane</i>			
MT1-MMP	MMP-14	14q11–q12	Collagens I, II, and III, gelatin, fibronectin, vitronectin, laminins 1 and 5, syndecan-1, entactin, aggrecan, fibrin, α_1 PI, decorin, proTNF α , CD44, tissue transglutaminase, activation of proMMP-2 and proMMP-13
MT2-MMP	MMP-15	8q21	ProMMP-2, collagen I, collagen IV NCI domain, laminin, fibronectin, tenascin, entactin, aggrecan, perlecan, proTNF α , proMMP-2, tissue transglutaminase
MT3-MMP	MMP-16	15q13q21	ProMMP-2, collagen III, fibronectin, tissue transglutaminase
MT5-MMP	MMP-24	20q11.2	ProMMP-2, proteoglycans, gelatin
<i>GPI-anchored</i>			
MT4-MMP	MMP-17	12q24.3	Fibrinogen, fibrin, gelatin, proTNF α
MT6-MMP	MMP-25	16q13.3	Collagen IV, gelatin, fibrinogen, fibrin, fibronectin, proMMP-2

Further details are available in <http://www.merops.co.uk/merops/merops.htm> and references therein.

Collagenases

Collagenase 1 (MMP-1), collagenase 2 (MMP-8), collagenase 3 (MMP-13), and collagenase 4 in *Xenopus* (MMP-18) belong to this subgroup. The characteristic feature of these enzymes

is their ability to cleave native fibrillar collagen types I, II, and III into $\frac{3}{4}$ and $\frac{1}{4}$ fragments. Some members of other MMP subgroups also degrade fibrillar collagen (e.g., MMP-2 and MMP-14). Collagenases also digest other ECM and non-ECM proteins (Table 1).

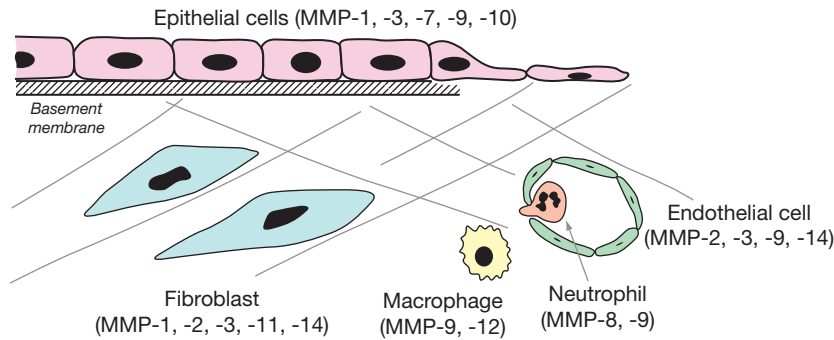
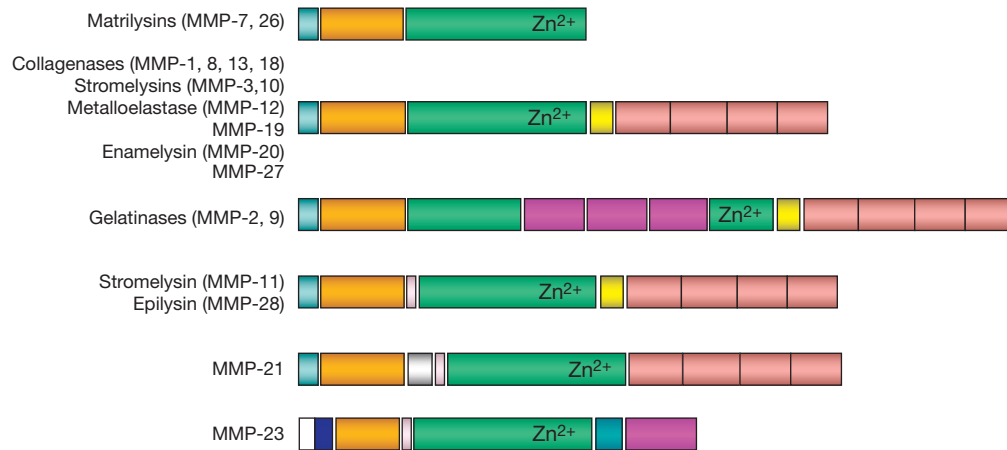


Figure 1 Tissue expression of MMPs. Representative MMPs produced by various cell types are shown. The production of these MMPs is usually transcriptionally controlled by stimulatory factors.

Soluble MMPs



Membrane-type MMPs

Transmembrane



GPI-anchored

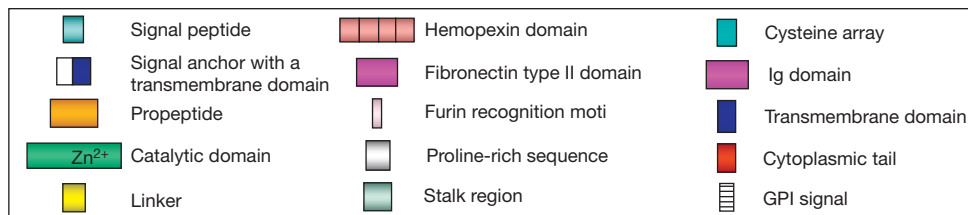


Figure 2 Domain arrangements of vertebrate MMPs.

Gelatinases

Gelatinase A (MMP-2) and gelatinase B (MMP-9) are in this subgroup. They have three repeats of fibronectin type II motifs inserted within the catalytic domain, which distinguishes them

from other MMPs. These repeats interact with collagens, gelatins, and laminin. Both MMPs readily digest denatured collagens, that is, gelatins, and a number of ECM molecules (Table 1). MMP-2, but not MMP-9, cleaves collagen I, II, and III into $\frac{3}{4}$ and $\frac{1}{4}$ fragments. MMP-2 is an important enzyme for

the release of matrix sequestered growth factors and cytokines and the processing of chemokines.

Stromelysins

Stromelysin 1 (MMP-3), stromelysin 2 (MMP-10), and stromelysin 3 (MMP-11) are assigned to this subgroup. While MMP-3 and MMP-10 have similar enzymatic properties, including their abilities to cleave various ECM molecules and activate several other MMPs (Table 1), MMP-11 has a more divergent amino acid sequence and substrate specificity. Human MMP-11 does not cleave ECM components. Alternative splicing and promoter usage generate an intracellular form of MMP-11.

Matrilysins

Matrilysin 1 (MMP-7) and matrilysin 2 (MMP-26) are assigned to this subgroup. They consist of a pro-domain and the catalytic domain only (Figure 2). MMP-7 is produced in epithelial cells and secreted apically. Besides digesting ECM molecules, it processes pro- α -defensin, Fas ligand, protumor necrosis factor α , and E-cadherin from the cell surface. MMP-26 digests several ECM molecules (Table 1).

Membrane-Type MMPs (MT-MMPs)

There are six MT-MMPs in humans. They are numerically assigned from MT1-MMP to MT6-MMP; four are type I trans-membrane proteins (MMP-14, MMP-15, MMP-16, and MMP-24) and two are glycosylphosphatidylinositol (GPI)-anchored proteins (MMP-17 and MMP-25). The catalytic domain and the hemopexin domain are exposed on the cell surface. With the exception of MT4-MMP (MMP-17), MT-MMPs activate proMMP-2 as well as digest ECM components. MT1-MMP, MT2-MMP, and MT3-MMP digest fibrillar collagens to varying degrees, and MT1-MMP activates proMMP-13. While MT1-MMP is expressed in many cell-types, MT5-MMP is expressed in brain, mainly in the cerebellum, and MT6-MMP in leukocytes.

Other MMPs

Seven MMPs do not belong to the above subgroups. Metalloelastase (MMP-12) is primarily expressed in macrophages. It digests elastin and other proteins and is essential for macrophage migration. The hemopexin domain of MMP-12 has bactericidal activity. MMP-19 is widely expressed in human tissues and implicated in tissue remodeling, wound healing, keratinocyte proliferation and migration, T-cell development, and T-cell-mediated immune responses. Enamelysin (MMP-20) is mainly expressed in newly formed tooth enamel. It processes amelogenin, and resulting products are essential for tooth enamel formation. MMP-21 is expressed in various fetal and adult tissues and in several invasive cancer cells. It digests gelatin and α_1 -proteinase inhibitor. MMP-23 is mainly

expressed in reproductive tissues. Two identical copies of the MMP-23 gene are located in human chromosome 1p36 loci in a head-to-head arrangement. The enzyme is a type II trans-membrane protein with a membrane-anchoring sequence located in the N-terminal end of propeptide. The propeptide is removed by a proprotein-converting proteinase; therefore, it is secreted from the cell as an active enzyme. The enzyme lacks both the cysteine switch motif in the propeptide domain and a hemopexin domain. The latter is replaced by a cysteine array and an immunoglobulin domain. MMP-27 is an ortholog of chicken MMP (originally assigned as MMP-22) and is expressed in B lymphocytes. The chicken enzyme digests gelatine and casein, but physiological substrates are not known. The enzymatic properties of MMP-27 have not been characterized. Epilysin (MMP-28) is expressed in keratinocytes, and its expression increases during wound healing. Its expression is also elevated in cartilage of patients with osteoarthritis and rheumatoid arthritis. Overexpression of MMP-28 in adenocarcinoma cells results in epithelial to mesenchymal transition, accompanied by loss of E-cadherin and activation of TGF β .

Structural Chemistry

The propeptide domain plays a key role in maintaining the MMPs in their zymogen forms (proMMPs) by blocking the active site. It has ~80 amino acids and consists of three α -helices and extended peptides including the cysteine switch motif (Figure 3(a)). The residues in the cysteine switch interact with the substrate-binding cleft in the catalytic domain by forming β -structure-like hydrogen bonds in a manner similar to a substrate, but in the opposite direction. The cysteine residue coordinates with the catalytic Zn²⁺ to maintain the latency of the enzyme.

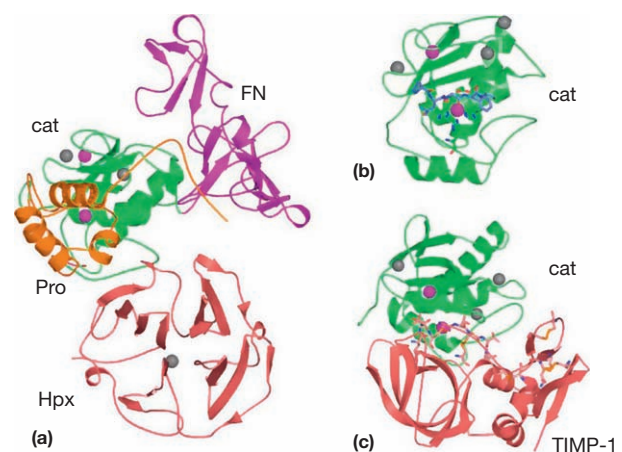


Figure 3 Three-dimensional structure of MMPs and TIMPs. (a) Ribbon diagram of proMMP-2. The propeptide (pro) is shown in orange; the catalytic (cat) domain in green; the fibronectin (FN) domain in pink; and the hemopexin (Hpx) domain in red. Zinc ions and calcium ions are in pink and gray, respectively. (b) The catalytic domain of MMP-1 with a peptide substrate (stick representation) bound to the active site. (c) The complex of TIMP-1 and the catalytic domain of MMP-3.

The catalytic domain has ~170 amino acids including the zinc-binding motif (HEXGHXXGXXH) and a conserved methionine that forms a unique Met-turn structure. The domain consists of three α -helices, a five-stranded β sheet, and bridging loops (Figure 3(b)). It contains two zinc ions (Zn^{2+}): one is essential for catalysis and the other together with three calcium ions stabilizes the structure. The overall polypeptide folds, including the Met-turn, are similar to those of astacins, reprolysins, ADAMs, and bacterial proteinases serralsins, which together constitute the metzincins.

The hemopexin domain consists of ~190 amino acids and forms a four-bladed β -propeller structure arranged almost symmetrically around a central axis, and each blade is comprised of a four-stranded anti-parallel β sheet (Figure 3(a)). The domain is stabilized by a calcium ion and a disulfide bond between blades I and IV.

The catalytic domain and the hemopexin domain are connected by a hinge (also called linker) region of variable length, which does not have a specific structure. The fibronectin type II motif found in MMP-2 and MMP-9 consists of two double-stranded anti-parallel β -sheets and two large loops (Figure 3(a)). Each β -sheet has one disulfide bond. Several MMPs are glycoproteins. In particular, proMMP-8, proMMP-9, and proMMP-23 are highly glycosylated, with ~20–45% of their molecular mass being carbohydrate.

Mechanism of Peptide Bond Hydrolysis

The active site of MMPs consists of the catalytic Zn^{2+} bound to three imidazoles of histidines in the HEXGHXXGXXH motif and the glutamate in the motif. A proposed mechanism of peptide bond hydrolysis by MMPs is shown in Figure 4. In an active MMP, one water molecule is bound to the catalytic Zn^{2+} . Upon binding of a peptide substrate to the active site of the enzyme, the carbonyl group of the P1 residue interacts with the catalytic Zn^{2+} and the water molecule is displaced toward the carboxyl group of the glutamate. This polarizes the water molecule and promotes its nucleophilic attack on the carbonyl carbon of the scissile bond, resulting in formation of a hemiketal intermediate and the subsequent hydrolysis of the peptide bond. The glutamate functions as a general acid/base during catalysis. Mutation of the glutamate to aspartate reduces the enzymatic activity to less than 1% and mutation to alanine reduces activity to 0.01% of the wild type.

Activation of proMMPs

MMPs are synthesized as pre-pro-MMPs and many are secreted from cells as zymogens (proMMPs) which need to be activated outside the cell. The major activation pathways are proteolytic processes, which can include auto-cleavages.

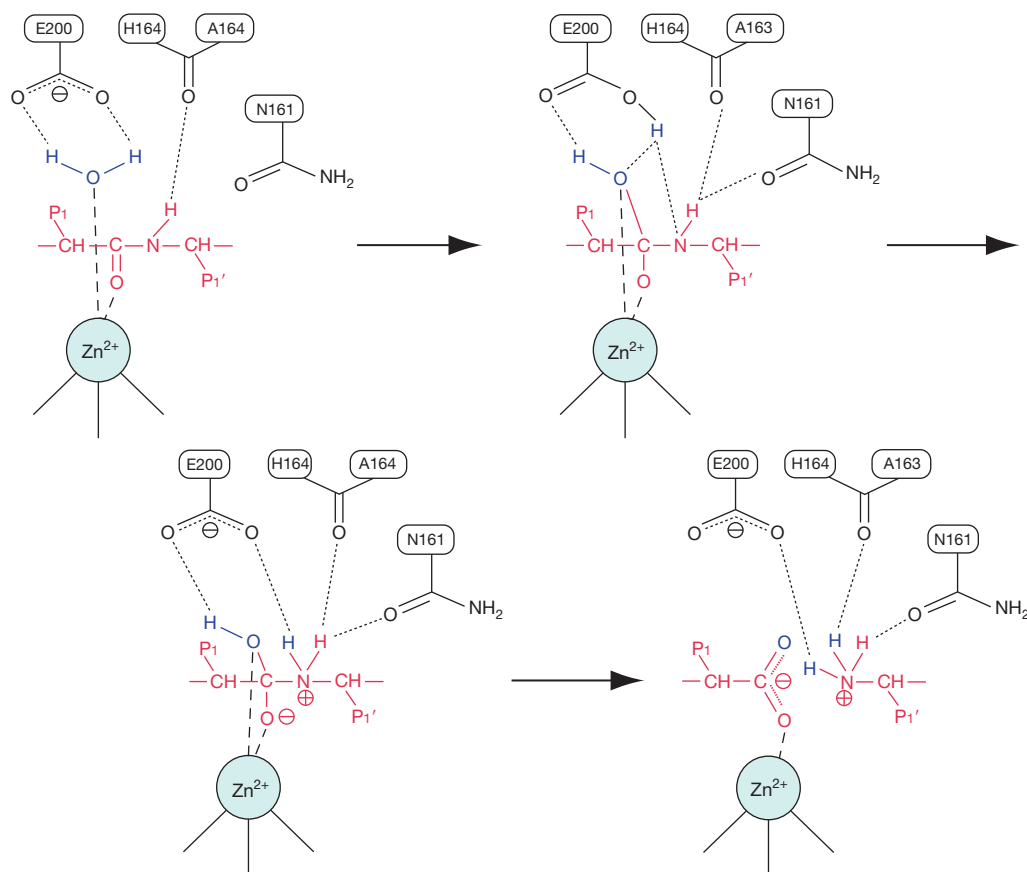


Figure 4 A proposed mechanism of action of MMPs. A peptide substrate and a water molecule that are involved in hydrolysis are shown red and blue, respectively. The residues shown are in MMP-1. See text for details.

Activation by Proteinases

The six MT-MMPs and MMP-11, -21, -23, and -28 contain a furin-like proprotein converting proteinase recognition sequence at the C-terminal end of the propeptide domain. Most of them are therefore activated within the cell and secreted or bound to the membrane as the active form.

Other MMPs are secreted from cells as proMMPs in the tissue. ProMMPs are likely to be activated by tissue or plasma proteinases or opportunistic bacterial proteinases. Figure 5 illustrates activation pathways that are involved in pericellular ECM hydrolysis. The urokinase-type plasminogen activator/plasmin system initiates the activation of many proMMPs near the cell surface. Unlike many other MMPs, proMMP-2 is not readily activated by plasmin or other proteinases, but activated on the cell surface by MT1-MMPs. This activation process is assisted by binding to a TIMP-2-MT1-MMP complex which presents proMMP-2 to an adjacent active MT1-MMP. ProMMP-13 is also activated by MT1-MMP, but this does not require TIMP-2 (Figure 5).

Activation by Chemical and Physical Means

ProMMPs can be activated *in vitro* by SH-reactive agents (e.g., 4-aminophenylmercuric acetate, iodoacetamide, and oxidized glutathione), chaotropic agents (e.g., urea, SCN^- , I^-), sodium dodecylsulfate, reactive oxygen, low pH, and heat treatment. This is due, in part, to the disruption of the cysteine–zinc interaction formed between the propeptide and catalytic domains and subsequent removal of the propeptide by autoproteolysis. Oxidation may activate some proMMPs, but it may also destroy the enzyme.

Activity and Substrate Specificity

MMPs exhibit optimal proteolytic activities at around pH 7.0–8.0 for most substrates, but, in some cases, they exhibit a broader pH range. An exception is human stromelysin 1 which has optimal activity at around pH 5.5–6.0 with a shoulder of activity at pH 7.5–8.0.

MMPs degrade ECM and non-ECM proteins (see Table 1 for representative substrates), indicating roles in many cellular processes. The potential substrates of more recently discovered MMPs (higher numbers) have not been well characterized. In general, the substrate needs to be six amino acids or longer, and the enzymes have a strong preference for hydrophobic residues at the P1' site (the residue located on the right of the cleaved peptide bond) such as Leu, Ile, Met, Phe, or Tyr, but the amino acid at the P1 site (the residue on the left of the cleaved peptide bond) is not important. Nuncatalytic domains greatly influence substrate specificity for some MMPs when substrates are extracellular macromolecules; for example, fibronectin domain repeats of MMP-2 and MMP-9 play an important role in expressing the activity on type IV collagen, gelatin, and elastin; and the hemopexin domain of collagenases is essential for the collagenolytic activity.

MMPs are inhibited by chelating agents such as EDTA, 1,10-phenanthroline, cysteine, and dithiothreitol. Numerous synthetic inhibitors of MMPs have been designed. They commonly contain a zinc-chelating moiety such as a hydroxamic acid, a carboxyl, a thiol, or a phosphinic group, although more recent developments have tended to avoid zinc chelators. Some highly selective inhibitors for specific MMPs have been generated, notably focusing on the S1' pocket, but many still broadly inhibit MMPs and other metalloproteinases. Phosphoramidon is not inhibitory. Tissue inhibitors of metalloproteinases

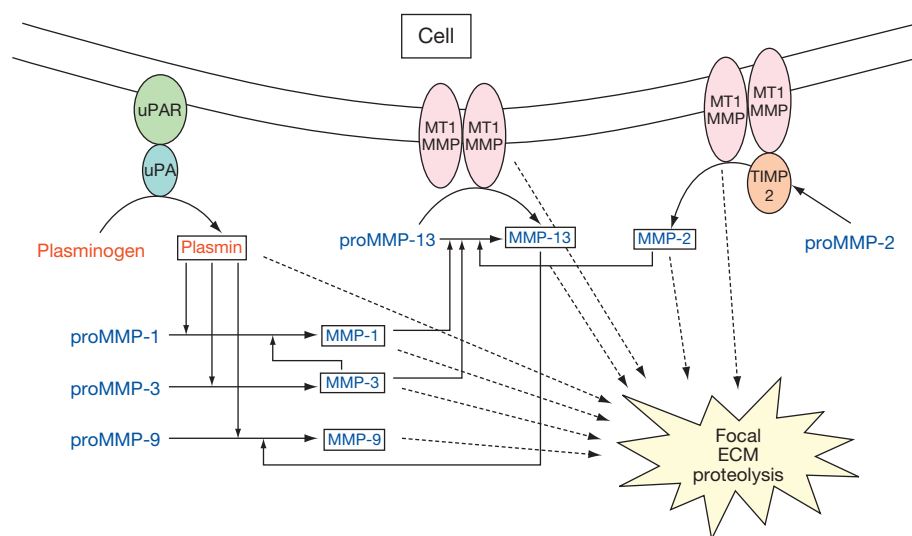


Figure 5 Activation pathways of proMMPs involved in pericellular ECM proteolysis. Urokinase-like plasminogen activator (uPA) bound to uPA receptor (uPAR) activates plasminogen to plasmin, which in turn activates a number of proMMPs. MT1-MMP activates proMMP-2; this process is assisted by binding to the TIMP-2-MT1-MMP complex which presents proMMP-2 to an adjacent active MT1-MMP. Activation of proMMP-13 by MT1-MMP does not require TIMP-2. Activated MMPs also participate in the activation of other proMMPs. Active enzymes are boxed.

(TIMPs), $\alpha 2$ -macroglobulins, and ovomacroglobulins (ovostatins) are the known natural protein inhibitors.

Endogenous Inhibitors of MMPs

$\alpha 2$ -Macroglobulins

Human $\alpha 2$ -macroglobulin is a plasma glycoprotein of 725 kDa consisting of four identical subunits of 180 kDa. It inhibits most endopeptidases by a trap mechanism: the interaction with a proteinase is initiated by the proteolytic cleavage in the middle of the subunits, causing a conformational change in the macroglobulin, hence entrapping the proteinase molecule. The active site of the trapped proteinase is not blocked, but its access to large protein substrates is sterically hindered. Related macroglobulins in other species, including ovomacroglobulins (ovostatins), have a similar inhibition mechanism.

TIMPs

TIMPs are structurally related proteins which inhibit matrixins with a 1:1 molar stoichiometry. The human genome has four TIMP genes and there are two genes in *C. elegans* and one in *Drosophila*. The four human isoforms (TIMP-1 to -4) consist of 184–194 amino acids, subdivided into N-terminal and C-terminal domains, each having three disulfide bonds. The N-terminal domain is the inhibitory domain and it binds to the active site of the MMPs. ProMMP-2 and proMMP-9 zymogens of MMPs bind to TIMPs via interaction between the C-terminal noninhibitory domain of TIMP and the C-terminal hemopexin domain of proMMP. These complexes act as MMP inhibitors since the N-terminal inhibitory domain of TIMPs is not blocked. TIMP-3 is unique among four TIMPs as it also inhibits ADAMs and ADAMTSs, metalloproteinases containing a disintegrin domain.

Inhibition Mechanism

The inhibitory domain of TIMPs has an OB-fold structure with a reactive ridge formed by the N-terminal segment Cys¹-Thr-Cys-Val⁴ and the Ser⁶⁸-Val-Cys⁷⁰ (residues are in TIMP-1) where Cys¹ and Cys⁷⁰ are disulfide linked. This ridge structure slots into the active site cleft of the MMPs (Figure 3(c)). The nitrogen of the N-terminal NH₂-group and the carbonyl oxygen of Cys¹ bidentately chelate the Zn²⁺ in the active site of the enzyme. MMPs do not form covalent bonds with TIMPs nor cleave TIMPs, but form tight complexes with inhibition constants in the sub nanomolar range.

Biological Activities of MMPs

The main function of MMPs is considered to be removal of ECM macromolecules from the tissues; they also alter the function of ECM molecules by specific proteolysis and express new biological functions of these molecules. Some secreted MMPs bind to specific cell-surface receptors, membrane-anchored proteins, or cell-associated ECM. This allows them to function in a focused way to modulate the pericellular environment. MMPs can also release growth factors bound to

ECM and from carrier proteins, or inactivate inflammatory cytokines and chemokines, which then influence cellular behavior. For example, the cleavage of type I collagen is associated with keratinocyte migration during wound healing, activation of osteoclasts during bone remodeling, and induction of apoptosis of amnion epithelial cells before the onset of labor. Cleavage of laminin 5 and CD44 by MMPs promotes cell migration. The expression of MT1-MMP and activation of proMMP-2 occur in invasive tumor cells and endothelial cells at the migratory front and promotes cells to migrate and invade by cleaving and shedding these ECM and cell-surface molecules. The release of TGF β and VEGF from the matrix by MMPs modulates ECM synthesis and angiogenesis, respectively, which are key events in development, morphogenesis, reproduction, tissue remodeling, and repair. Inactivation of monocyte chemoattractant proteins or interleukin 1 β by MMP reduces inflammation. MMPs are now acknowledged as modulators of many extracellular signaling pathways, with individual specificities for associated proteins, including cytokines, growth factors, cell adhesion molecules, and others.

Human Genetic Disorders and DNA Polymorphisms

Unregulated MMP activities are implicated in the progression of a variety of diseases characterized by aberrant breakdown of ECM molecules. In many cases, their activities are controlled at the cellular level by various stimulatory factors. In addition, a few inherited human disorders associated with the MMP and TIMP genes have been described.

Mutations in human MMP-2 resulting in the absence of active MMP-2 are linked to a rare autosomal recessive genetic disorder, multicentric osteolysis (Torg syndrome, nodulosis–arthropathy–osteolysis syndrome, and Winchester syndrome), which causes destruction of the affected bones and joints. A mutation in the coding region for MMP-13 is associated with spondyloepimetaphyseal dysplasia, a dwarfism resulted from abnormal development of the vertebrae and extremities. Amelogenin imperfecta, a genetic disorder caused by defective enamel formation, is due to a mutation at MMP-20 cleavage sites.

Mutations in the coding region of the TIMP3 gene leading to the presence of an extra Cys in the C-terminal region of the expressed protein cause the rare autosomal dominant Sorsby's fundus dystrophy, a form of blindness characterized by death of the retinal pigment epithelium.

Various DNA polymorphisms in MMP genes are associated with a number of diseases. In many cases, polymorphic sites are found in the promoter regions of MMP genes: polymorphisms in the *MMP1* gene are associated with coronary heart disease, rheumatoid arthritis, periodontitis, and cancers (e.g., cervix, ovary, colon, kidney, lung, and skin); in the *MMP2* gene with the risk of developing lung cancer; in the *MMP3* gene with coronary heart disease, coronary aneurysms, and cancers (colon and breast); in the *MMP9* gene with coronary heart disease, aneurysms, nephropathy, and preterm prenatal rupture of the fetal membrane; and in the *MMP12* gene with coronary atherosclerosis.

See also: **Protein/Enzyme Structure Function and Degradation:** Collagenases; Collagens.

Further Reading

- Barrett AJ, Rawlings ND, and Woessner JF (2004) *Handbook of Proteolytic Enzymes*, 2nd edn. London: Academic Press/Elsevier Science.
- Murphy G and Nagase H (2011) Matrix metalloproteinases in health and disease: Localising MMP activities in the pericellular environment. *FEBS Journal* 278: 2–15.
- Parks WC (1998) *Matrix Metalloproteinases*. San Diego, CA: Academic Press.
- Rawlings ND, O'Brien EA, and Barrett AJ (2002) MEROPS: The protease database. *Nucleic Acids Research* 30: 343–346.

- Sternlicht MD and Werb Z (2001) How matrix metalloproteinases regulate cell behaviour. *Annual Review of Cell and Developmental Biology* 17: 463–516.
- Woessner JF Jr. and Nagase H (2000) *Matrix Metalloproteinases and TIMPs*. Oxford: Oxford University Press.

Relevant Website

<http://www.merops.co.uk> – bluehost, Affordable, Reliable Web Hosting Solution.

MicroRNAs in Eukaryotes

R N Douglas and J A Birchler, University of Missouri, Columbia, MO, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Argonaute A family of proteins that bind micro-RNAs to form a functional RNA-induced silencing complex. Some members of the Argonaute family possess RNase H-like activity and are able to cleave target messenger RNA (mRNA) transcripts, while other members of the Argonaute family lack RNase-H like activity and inhibit translation of mRNA by other means.

Micro-RNA A small RNA molecule found in eukaryotes, typically ~22 nucleotides in length, that is excised from an endogenous double-stranded RNA precursor containing a stem-loop structure.

Mirtron A micro-RNA precursor that is spliced out of an intron.

RNA-induced silencing complex A multiprotein complex that binds a mature micro-RNA and then targets a complementary RNA transcript to inhibit translation.

Small RNA Eukaryotic noncoding RNAs ~18–24 nucleotides in length that often have regulatory roles in cells. Classes of small RNAs include: micro-RNA, small nuclear RNA, small nucleolar RNA, piwi-interacting RNA, small interfering RNA and trans-acting small interfering RNA.

MicroRNAs in Eukaryotes

Several classes of noncoding small RNAs, RNA molecules consisting of ~18–24 nucleotides (nt), are found in eukaryotes. MicroRNAs (miRNAs) are a distinct subclass of small RNA molecules that are ~21–23 nt long. miRNAs arise from endogenous, double-stranded hairpin-shaped RNAs and they function to silence endogenous genes at transcriptional and/or posttranscriptional levels in eukaryotes. miRNAs inhibit transcription by guiding an RNA-induced silencing complex (RISC) to cleave a complementary messenger RNA (mRNA) transcript. Alternatively, miRNAs may prevent translation of an mRNA by targeting a RISC to an mRNA to impede the action of ribosomes by one of several possible mechanisms including deadenylation, ribosome inhibition, or repression of translation initiation. The importance of miRNAs is underscored by the discovery that many miRNA families are conserved across distant taxa and that upward of 50% of all protein-coding mammalian genes are miRNA targets. Although small in size, up to 10 000 individual miRNAs may be present in a single cell and they function in myriad biological processes including plant and animal development, epigenetics, and cancer growth.

miRNAs are very similar to another class of small RNAs termed 'short interfering RNAs' (siRNAs) that require the same families of proteins, Dicers and Argonautes (see below), to silence their targets. siRNAs were once thought of as arising only when a virus, transgene, or transposon was attacking a genome, whereas miRNAs were endogenously encoded. However, recent research has shown that siRNAs are not always exogenous in origin, and the distinction between miRNAs and siRNAs is beginning to become more obscured. Today, the major difference between siRNAs and miRNAs is that siRNAs originate from within fully complementary double-stranded RNAs (dsRNAs), while miRNAs are derived from a dsRNA that takes on a stem-loop conformation with imperfect complementarity.

Discovery of miRNAs

The first clues to the existence of small RNAs as a mediator of gene silencing arose in the mid-1980s, when researchers injected anti-sense RNA into a murine (*Mus musculus*) cell line and observed a decrease in the accumulation of the corresponding sense RNA transcript. It was soon discovered that similar experiments with frog (*Xenopus* sp.) oocytes, fruit fly (*Drosophila melanogaster*) embryos, and wild carrot (*Daucus carota*) protoplasts also reduced the accumulation of complementary endogenous mRNAs, which indicated that this phenomenon was widespread in eukaryotes.

A major breakthrough in small RNA research occurred in 1993, when the first miRNA was discovered in the nematode *Caenorhabditis elegans*. The *lin-4* locus encodes a product required for heterochronic switching in post-embryonic development by repressing the accumulation of the LIN-14 protein and allowing for properly timed switches between larval developmental stages. The *lin-4* loss-of-function mutants revert to early developmental stages later in development, which can lead to the absence of typical adult structures. Intriguingly, the *lin-4* locus was not found to encode a protein. The *lin-4* locus produces 22- and 61-nt RNA transcripts that are complementary to seven repeats in the 3' untranslated region (UTR) of *lin-14* mRNA in *C. elegans*. The discovery that *lin-4* RNA transcripts could bind *lin-14* mRNA strongly suggested that *lin-4* inhibits *lin-14* via an antisense RNA–RNA interaction. However, *lin-14* mRNA levels are unaltered in *lin-4* loss-of-function mutants, so it was believed that *lin-4:lin-14* RNA–RNA interactions inhibited *lin-14* mRNA translation rather than transcription. Several years later, the second miRNA locus, *let-7*, was also discovered in *C. elegans*. Researchers showed that *let-7*, another heterochronic switching factor, produces a 21-nt miRNA that has complementarity to several additional heterochronic switching factors including *lin-14*. Early work in *C. elegans* demonstrated that miRNAs were intricately linked to developmental timing,

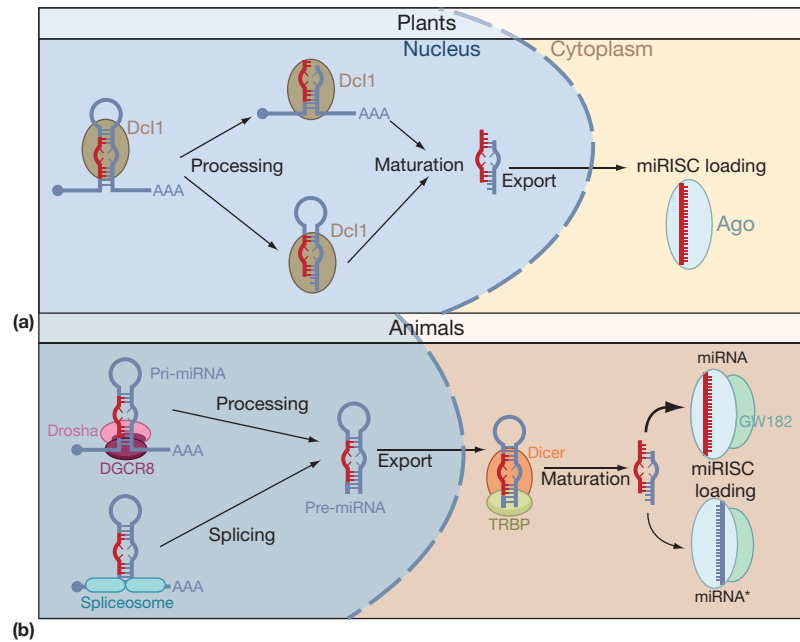


Figure 1 MicroRNA biogenesis in plant and animal cells. (a) In plant cells DICER-LIKE1 (Dcl1) processes the primary microRNA (pri-miRNA) into a precursor miRNA (pre-miRNA) and then to a ~22-nt-long miRNA/miRNA* duplex in the nucleus. The miRNA/miRNA* duplex is exported from the nucleus to the cytoplasm where it is loaded into an ARGONAUTE (AGO) protein to form a functional microRNA-induced silencing complex (miRISC). (b) In animal cells Drosha, a protein similar to DCL1, processes pri-miRNAs into pre-miRNAs with DGCR8, a double-stranded RNA binding protein. Alternatively, the spliceosome can excise pre-miRNAs from mRNA introns. Unlike in plants, the pre-miRNA is exported from the nucleus to the cytoplasm before being processed into a mature miRNA/miRNA* duplex. The pre-miRNA is processed in the cytoplasm by Dicer and a transactivation-responsive RNA-binding protein (TRBP). From there the miRNA and miRNA* may be incorporated into an AGO miRISC with the aid of a glycine-tryptophan protein of 182 kilodaltons (GW182). Reproduced from Carthew RW and Sontheimer EJ (2009) Origins and mechanisms of miRNAs and siRNAs. *Cell* 136: 642–655, with permission from Elsevier.

and subsequent work soon showed that miRNAs are abundant in many invertebrate and vertebrate species. Around the same time, researchers began to identify miRNAs in plants as well, indicating that miRNAs likely originated early in eukaryote evolution prior to the split between animals and plants.

Biogenesis of miRNAs

Upward of 1000 discrete miRNAs may be encoded in a single genome and they are all inexorably linked to a small group of proteins required to process the miRNAs in the nucleus, export them from the nucleus to the cytoplasm, and act on the target gene to inhibit transcription or translation. While most plant miRNA loci are found singly in intergenic regions and individually transcribed by their own promoters, animal miRNAs are often found in clusters, within introns of protein-coding genes, and/or transcribed in polycistronic units. miRNA loci in plants and animals are typically transcribed by RNA polymerase II (Pol II), and these primary miRNA (pri-miRNA) transcripts are capped with 7-methylguanylate on the 5' end and polyadenylated on the 3' end. The pri-miRNA folds back on itself in an imperfect manner to form a stem-loop structure and may be several thousand nucleotides long, although the stem is usually only ~33 nt long. Biogenesis of miRNAs is similar in plants

(Figure 1(a)) and animals (Figure 1(b)), but there are several subtle differences between the two kingdoms.

An RNA-binding protein, DAWDLE in plants or the orthologous SNIP1 in animals, stabilizes pri-miRNAs as they are readied for processing into precursor miRNAs (pre-miRNAs). In plants, pri-miRNAs are moved to a nuclear processing region called a D-body where it interacts with a protein complex that includes HYPONASTIC LEAVES1, a dsRNA-binding protein, SERRATE, a C2H2-zinc finger protein, and DICER-LIKE1 (DCL1), which transforms the pri-miRNA into a ~60–70-nt-long hairpin-shaped pre-miRNA. DCL1 is an RNase III family member and consists of several domains, including a DEXD/H ATPase domain, a domain of unknown function (DUF) 283 domain, an RNA helicase motif, a Piwi/Argonaute/Zwille (PAZ) domain to hold the 3' end of small RNA molecules, two RNase III domains, and one or more dsRNA-binding domains. DCL1 is the enzyme that physically cleaves the pri-miRNA into the pre-miRNA in plants. Following a second round of processing by DCL1, in which the pre-miRNA is reduced to a ~22-nt dsRNA consisting of the mature miRNA and its antisense complement, or passenger strand (miRNA*), the resulting miRNA/miRNA* duplex is methylated and stabilized by HUA ENHANCER1, an S-adenosyl methionine-dependent methyltransferase. Methylation prevents the miRNA from being degraded by a family of small RNA-degrading nucleases in

the cytoplasm. The miRNA/miRNA* duplex is exported from the nucleus to the cytoplasm via the animal Exportin-5 homolog HASTY, although there are likely additional miRNA/miRNA* nuclear export pathways in plants as some miRNAs are still exported from the nucleus in *hasty* loss-of-function mutants. It is also not entirely clear where or when the miRNA and miRNA* separate in plants, but once in the cytoplasm the miRNA associates with an ARGONAUTE (AGO) protein to form a RISC, which may then target transcripts with complementarity to the miRNA. AGO proteins consist of an N-terminal domain, a PAZ domain, which is shared with DCL and Dicer (see below) proteins, a middle (MID) domain to hold the 5' end of a small RNA molecule and a P-element-induced Wimpy testis (PIWI) domain, which has RNase H-like activity to cleave target transcripts.

In animal nuclei, the pri-miRNA interacts with Drosha, an RNase III family member similar to, but not homologous to, DCL1, and DGCR8/Pasha, a dsRNA-binding protein that helps to position the catalytic residues of Drosha at the correct position on the pri-miRNA. Drosha is responsible for cleaving the pri-miRNA into the much smaller pre-miRNA. The pre-miRNA is immediately processed into the miRNA/miRNA* duplex in plants, but in animals the pre-miRNA is exported from the nucleus to the cytoplasm via Exportin-5. Once it reaches the cytoplasm, the pre-miRNA is loaded into a complex including Dicer, AGO2, and a transactivation-responsive RNA-binding protein. Although plant genomes often encode multiple DCL family members that are able to specialize to only process either miRNAs or siRNAs, the genomes of mammals and *C. elegans* encode only one Dicer protein, which processes both miRNAs and siRNAs. The pre-miRNA is processed into an miRNA/miRNA* duplex by Dicer. The duplex has a short life span, and either strand may be incorporated into a functional RISC, although the more stable strand is termed the 'miRNA' and is transferred to AGO more often than miRNA*, which is often degraded. However, miRNA* has occasionally been found to be able to form a functional RISC and target mRNA transcripts, adding a new wrinkle to miRNA biogenesis.

Animals also have many miRNAs transcribed from within introns rather than from intergenic regions, and these loci are called 'mirtrons'. Mirtron expression is regulated by the promoter of its host transcript and they are transcribed by Pol II as a canonical gene transcript, but when the spliceosome removes the intron containing the miRNA, it is spliced out as a pre-miRNA. This pre-miRNA then undergoes normal processing on its way to becoming a mature miRNA.

Function of miRNAs

Once an miRNA has been generated and joins a functional RISC, it will guide the RISC to a complementary sequence on an mRNA transcript. Animal miRNAs often have imperfect complementarity with their targets and they usually target sequences in the 3' UTR of the mRNA transcript. Recent work suggests that ~60% of murine miRNA target sequences are in the 3' UTR, while 25% are in exons and 12% are within introns. In contrast to animals, plant miRNAs typically have perfect complementarity with their target transcripts and the targeted base pairs are generally within an exon.

The miRNA-guided RISC can regulate target transcripts in both plants and animals either by cleaving the transcript or by inhibiting translation of the transcript into a protein. Not all AGO family members have the ability to cleave target transcripts, and those that do have this ability are often referred to as 'slicers'. The miRNA itself also is an important determinant in regard to whether or not a target mRNA can be cleaved. If the miRNA does not have perfect complementarity with the target between nucleotides 9 and 11, a bulge in the miRNA-mRNA duplex will form and the active site residues of AGO will not be properly positioned to cleave the target mRNA, which will likely lead to translational inhibition. However, if there is perfect complementarity between the miRNA and the targeted mRNA, then endonucleolytic cleavage of the mRNA is possible. The discovery of *miRNA-action deficient* mutants in *Arabidopsis* showed that mRNA cleavage and translational inhibition are equally likely to occur in plants, genetically separable, and that translational inhibition is the default method for miRNA-mediated gene silencing in plants and animals.

There are several proposed models for posttranscriptional silencing via RISC activity in animals. It is not yet known if an miRNA-guided RISC prevents translation before or after initiation, although experimental data suggest that inhibition may occur at both stages. In order to begin mRNA translation in mammals, an eIF4F initiation complex recognizes the 5' cap on an mRNA and recruits other initiation factors and ribosomal subunits to the 5' end of the mRNA (Figure 2(a)). The recruitment process also causes the mRNA to circularize, which improves translation efficiency. One model of posttranscriptional silencing simply suggests that the presence of a RISC on an mRNA transcript promotes ribosomal drop-off, which would prevent translation from occurring (Figure 2(b)). A second model suggests that the MID domain of human AGO2, which resembles the structure of eIF4F, can compete with the initiation complex to bind the 5' cap on the mRNA (Figure 2(c)). If eIF4F is outcompeted by AGO2, then eIF4F would be unable to bring ribosomes to the translation start site and translation would not occur. A third proposed model suggests that a RISC may inhibit 60S ribosomal subunits from joining the translation initiation complex. *In vitro* experiments have shown that human AGO2 can bind the initiation factor eIF6 and 60S ribosomal subunits. eIF6 acts to keep 60S ribosomal subunits from prematurely interacting with 40S ribosomal subunits, so if eIF6 and 60S subunits were to bind AGO2, then the 60S subunits would be unable to bind 40S subunits and begin mRNA translation (Figure 2(d)). Finally, the fourth model proposes that the presence of a RISC on an mRNA promotes deadenylation of the poly(A) tail. The removal of adenosines is purported to inhibit circularization of the mRNA molecule, which would greatly decrease translational efficiency and effectively silence the mRNA transcript.

The *Arabidopsis* gene *VARICOSE* (VCS) is homologous to the mammalian gene *Ge-1*, a decapping enzyme required for miRNA-guided posttranslational inhibition. Loss-of-function *vcs* mutants show no increase in mRNA accumulation levels for miRNA targets, but protein levels do increase, which implies that VCS, like *Ge-1*, is required for posttranslational inhibition. This discovery shows that plants and animals likely use similar mechanisms to regulate mRNA at the posttranslational level with miRNA-guided RISCs.

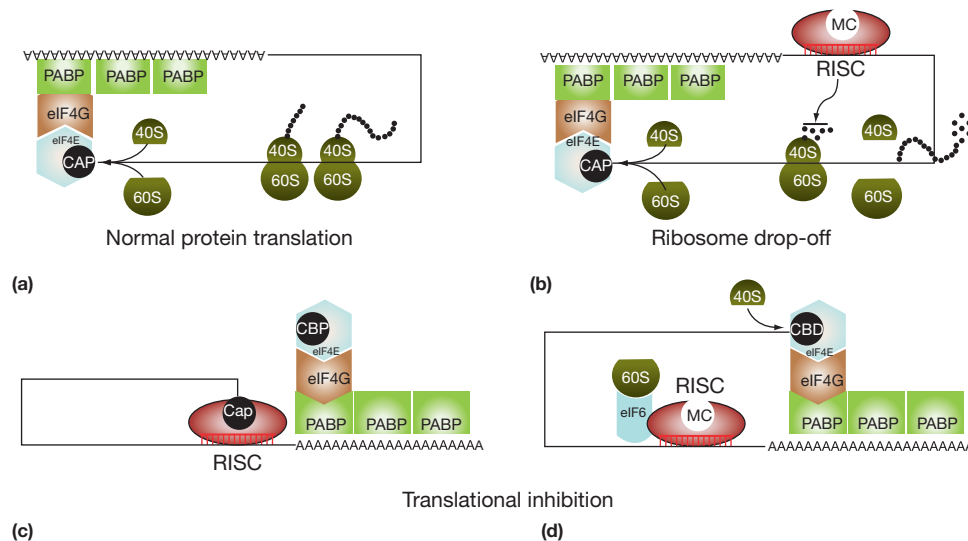


Figure 2 Modes of microRNA silencing. Several possible mechanisms have been proposed for the regulation of mRNA transcripts by miRNA-guided RISC activity. (a) Efficient mRNA translation requires circularization of the mRNA. Eukaryotic translation initiation factors (eIF4E and eIF4G) bind to the 5' cap (CAP) of the mRNA and to poly(A)-binding proteins (PABPs) to circularize the mRNA and allow translation initiation by 40S and 60S ribosomal subunits. (b) eIF4E can begin to circularize mRNA transcripts, but miRNA-guided RNA-induced silencing complexes (RISCs) cause ribosomes to drop off the mRNA transcript and inhibit full mRNA translation. (c) The MID-domain (MC) of an Argonaute protein in a RISC may outcompete eIF4E to bind the 5' cap on miRNA-targeted transcripts and inhibit the start of mRNA translation. (d) The AGO protein in the RISC may bind eIF6, which binds the 60S ribosomal subunit. This binding could prevent the 60S and 40S ribosomal subunits from forming a translation complex. Reproduced from Tang G, Tang X, Mendu V, et al. (2008) The art of microRNA: Various strategies leading to gene silencing via an ancient pathway. *Biochimica et Biophysica Acta* 1779: 655–662.

Biological Roles of miRNAs

Any one miRNA may target several dozen distinct mRNA transcripts, and an mRNA transcript may be under the regulation of several miRNA species. miRNAs are prevalent in essentially all eukaryotic cells, so it is not surprising that they have adopted functions in diverse biological processes including, but not limited to, plant development, animal development, epigenetics, and cancer growth.

One of the most controversial aspects of current miRNA research in plants is whether or not miRNAs can act noncell autonomously, that is, whether or not they are mobile. Unlike in animals, plant cell fate is not determined by cell lineage; rather, it is established by positional clues from neighboring cells. Plant cells are known to rely on mobile signals to establish their fate, so if regulatory molecules such as miRNAs are mobile, it could have major implications in understanding how plants develop. Early research using artificial miRNAs and/or miRNAs with non-native promoters in plants suggested that miRNAs could only act in a cell-autonomous manner. However, recent research strongly suggests that plant miRNAs are in fact mobile and capable of acting some distance away from their origin via transport through phloem and/or plasmodesmata. Several lines of research support the idea of mobile miRNAs: several miRNAs known to respond to physiological stresses are readily found in phloem sap; in maize (*Zea mays*), *MIR390* precursor loci are expressed only in the outermost cell layer of the shoot apical meristem, but *in situ* hybridization shows that mature miR390 accumulates in inner cell layers as well as the outermost layer; lastly, it has been shown

that a mobile miRNA acts in a reciprocal manner with a mobile transcription factor to pattern the *Arabidopsis* root.

miRNAs also play important roles in animal development beginning with embryogenesis. The vital role of the miRNA biogenesis pathway in embryo development is underscored by Dicer knockouts in mice; they are embryonic lethal. However, stem cell populations that lack functional Dicer proteins have been recovered. Since these stem cells lack differentiation markers even after differentiation has been induced, and often show cell-cycle arrest, it can be inferred that miRNAs play an important role in the establishment of different tissue types during embryogenesis. Indeed, it has been revealed that distinct miRNA species play important roles in the formation of a diverse array of mammalian cell types.

Many studies have implicated siRNAs in directing epigenetic changes to an organism's genome, but miRNAs have also been shown to be vital to this process. Epigenetics encompasses changes to an organism's genome without changing the DNA sequence. Epigenetic changes may modify gene expression by causing methylation of individual DNA nucleotides and/or by methylating or acetylating histones, which can change the chromatin state back and forth from easily accessible euchromatin to inaccessible heterochromatin. miRNAs manipulate epigenetic changes by regulating the accumulation of DNA methyltransferases, which methylate DNA, and histone deacetylases and histone methyltransferases. Interestingly, many miRNA genes, perhaps up to 15% in humans, have been found to be targets for silencing by the very epigenetic machinery they target, indicating that a complex feedback loop likely regulates epigenetic changes. Similarly, in

Arabidopsis, miR168 has been shown to negatively regulate the accumulation of AGO1 mRNA, which indicates that miRNA genes and RISC-related genes are under tight regulation. Many of the epigenetic influences caused by miRNAs have been uncovered during research of various types of cancer. In most cancers, the overall expression of miRNAs is reduced, but individual miRNAs may increase their accumulation. Typically, oncogenes have increased accumulation due to a decrease in their regulatory miRNAs, while tumor-suppressor genes show decreased accumulation due to an increase in their regulatory miRNAs. Interestingly, it seems that each type of cancer has a distinct collection of miRNAs that could be used as biomarkers in the future to identify cancers in humans at the earliest stages of tumor development.

Conclusion

This article discusses the mechanisms of miRNA biogenesis. The exact mechanism of miRNA-guided RISC translational inhibition is currently unknown, but it likely involves several, if not all, of the proposed models covered herein. Since their discovery in 1993, the study of miRNAs has progressed rapidly and more will surely be uncovered in the years to come.

See also: **Molecular Biology:** Messenger RNA Degradation in Bacteria; Messenger RNA Processing in Eukaryotes; mRNA Polyadenylation in Eukaryotes; Small RNAs in Bacteria.

Further Reading

- Carthew RW and Sontheimer EJ (2009) Origins and mechanisms of miRNAs and siRNAs. *Cell* 136: 642–655.
- Chitwood DH and Timmermans MCP (2010) Small RNAs are on the move. *Nature* 467: 415–419.
- Kim VN (2005) MicroRNA biogenesis: Coordinated cropping and dicing. *Nature Reviews Genetics* 6: 376–385.
- Krol J, Loedige I, and Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nature Reviews Genetics* 11: 597–610.
- Lee RC, Feinbaum RL, and Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75: 843–854.
- Tang G, Tang X, Mendu V, et al. (2008) The art of microRNA: Various strategies leading to gene silencing via an ancient pathway. *Biochimica et Biophysica Acta* 1779: 655–662.
- Voinnet O (2009) Origin, biogenesis and activity of plant microRNAs. *Cell* 136: 669–687.

Relevant Website

<http://www.mirbase.org> – miRBase.

Microtubule-Associated Proteins

N Hirokawa, University of Tokyo, Tokyo, Japan

R Takemura, National College of Nursing, Tokyo, Japan

© 2013 Elsevier Inc. All rights reserved.

Glossary

Axon Structure that transmits the signal from the cell body to the synaptic terminal. Single axon extends from the cell body. Microtubules and neurofilaments are major components of the axonal cytoskeleton. Microtubules in the axons are organized more densely than in the dendrites.

Dendrite Structures that receive signals and transmit them to the direction of the cell body. Usually, multiple dendrites emanate from the cell body in 'dendritic' shape.

Microtubules are the major components of the dendritic cytoskeleton, and microtubules in the dendrites are organized more loosely than in the axon.

Microtubule Filament of 25 nm diameter.

Assembled from α - and β -tubulin subunits (55 kDa), it is actually a hollow cylinder (hence tubule) composed of 13 protofilaments. During mitosis, it forms a mitotic spindle. Compared to intermediate filaments (10 nm diameter), microtubules turn over much more dynamically. New subunits are added at the 'plus' end, while subunits dissociate from the 'minus' end (treadmilling). In addition, microtubules show the property called 'dynamic instability', in which individual microtubules are either growing or shrinking and they stochastically switch between the two states.

Classical Microtubule-Associated Proteins

Microtubules are formed by polymerization of α - and β -tubulin. Microtubules play important roles in organizing and maintaining the shape of the cell, and also act as tracks for the transport of membranous organelles and macromolecules. The so-called motor proteins (kinesin superfamily proteins and dynein superfamily proteins) are involved in the transport processes, and they use the energy derived from the hydrolysis of ATP. Microtubule-associated proteins (MAPs) bind to microtubules without a requirement for nucleotide involvement, and organize the spatial arrangement of microtubules within the cell; they also regulate the polymerization of microtubules. Historically, MAPs have been isolated from the mammalian brain by their property that they remain bound to microtubules through repeated cycle of polymerization and depolymerization. The quick-freeze, deep-etch microscopy reveals that in neurons, there are abundant filamentous structures of various length associated with microtubules. These filamentous structures were proved to be mainly composed of MAPs (Figures 1 and 2). These MAPs have been studied extensively, and we term these well-characterized MAPs as 'classical MAPs'. The major part of this article will be spent on these classical MAPs. Some other MAPs are known and some new MAPs have been identified recently. These will be dealt with in relevant places. Classical MAPs include MAP1A, MAP1B, MAP2A, MAP2B, MAP2C, and tau (from mammalian brains) and MAP4 (found ubiquitously) (Table 1). All these MAPs are fibrous molecules of different lengths, ranging from 50 to 185 nm, and form filamentous projections from the microtubule surface. First, the basic biochemical properties and their expression patterns will be described for each of the classical MAPs.

MAP1A and MAP1B

MAP1A and MAP1B are high-molecular-weight, heat-labile, filamentous proteins of ~350 and 320 kDa on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE). MAP1A and MAP1B are transcribed from the separate genes. MAP1B has also been called MAP1x or MAP5. The protein originally called MAP1C turned out to be cytoplasmic dynein. Therefore, it is not usually included in MAPs. MAP1A encodes protein of 2774 amino acid residues, and MAP1B encodes protein of 2464 amino acid residues. Rotary-shadowing electron microscopy reveals that MAP1A and MAP1B are highly elongated, flexible molecules of ~150–185 nm in length. Both MAP1A and MAP1B have associated light chains. Interestingly, the light chains, LC2 and LC1, are encoded at the C-terminal portion of MAP1A and MAP1B molecules, respectively. At the N-terminal portion of both MAP1A and MAP1B molecules, there is a domain enriched in the positively charged amino acids and contains the short motif (KKEX), repeated 11 and 21 times in MAP1A and MAP1B (Table 1 and Figure 3). This domain serves as microtubule-binding domain, and interacts with C terminus of tubulin, which has a large number of acidic residues. Both MAP1A and MAP1B form filamentous structures associated with microtubules in neurons.

MAP1A and MAP1B are predominantly expressed in neurons, although a smaller amount may be expressed in other tissues as well. There is a developmental regulation in the expression of MAP1. MAP1A is more abundantly expressed in the adult, and MAP1B is more abundantly expressed in the developing nervous system. In fact, the expression of these two proteins appears to be complementary to each other. In the rat brain development, the level of MAP1B expression is highest at the first few postnatal days, when many neurons are extending their axons. Thereafter, the expression of MAP1B declines.

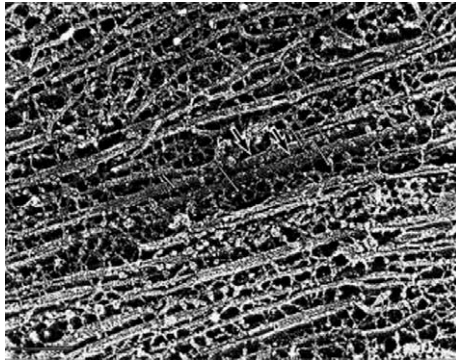


Figure 1 Axonal cytoskeleton observed by quick-freeze, deep-etch electron microscopy. Between microtubules (thick black arrows), short cross-linkers (thin arrows) and longer cross-linkers are found. Scale = 100 nm. Reproduced from Hirokawa N (1991) Molecular architecture and dynamics of the neuronal cytoskeleton. In: Burgoyne RD (ed.) *Neuronal Cytoskeleton*, pp. 5–74., New York, NY: Wiley-Liss, with permission of Wiley-Liss, Inc.

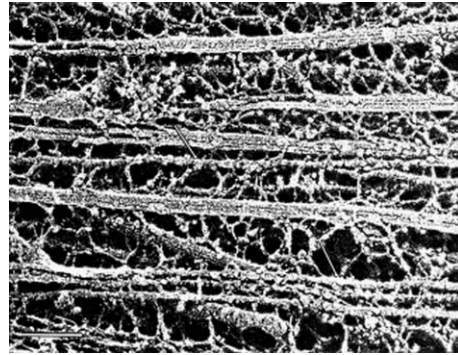


Figure 2 Dendritic cytoskeleton observed by quick-freeze, deep-etch electron microscopy. Between microtubules, and between microtubules and neurofilaments (arrows), long cross-linkers are found. Scale = 100 nm. Reproduced from Hirokawa N (1991) Molecular architecture and dynamics of the neuronal cytoskeleton. In: Burgoyne RD (ed.) *Neuronal Cytoskeleton*, pp. 5–74., New York, NY: Wiley-Liss, with permission of Wiley-Liss, Inc.

Table 1 Characteristics of classical MAPs

MAPs	App. mol. weight on SDS-PAGE (kDa)	Number of amino acids	Length of isolated molecule (nm)	Distance in microtubule bundles (nm)	Microtubule-binding domain	Developmental and topological expression pattern
MAP1A ^a	350	2774	150–185		N-terminal KKEX repeats	Abundant in mature dendrite
MAP1B ^a	320	2464	150–185		Same as above	Abundant in developing axon
MAP2A ^b	280	1911	~100	~65	C-terminal tandem repeats	Specific to developing dendrite
MAP2B ^b	270	1828	~100	~65	Same as above	Specific to developing and mature dendrite
MAP2C ^b	70	467		~20	Same as above	Abundant in developing axon and dendrite
tau ^c	55–65	352–441	~50	~20	Same as above	Abundant in developing and mature axon
MAP4	210	1072			Same as above	Ubiquitously expressed

^aMAP1A and MAP1B are transcribed from separate genes.

^bMAP2A, MAP2B, and MAP2C are splice variants.

^cSix isoforms of tau are splice variants.

By contrast, the expression of MAP1A is very low at this period, but the level of its expression increases thereafter, and reaches its maximal level in adults. There is a spatial difference in the expression of MAP1B and MAP1A. MAP1B is predominantly expressed in the developing axon, although smaller amount is found in the dendrite of juvenile and mature neurons. By contrast, MAP1A is abundantly expressed in the mature dendrites, with the smaller amount of expression at the axons.

MAP2A, MAP2B, and MAP2C

MAP2A and MAP2B were identified as high-molecular-weight, filamentous protein of ~280 and 270 kDa on SDS-PAGE. MAP2A and MAP2B have been known to be heat stable, which is rather unique considering the high molecular weight of these proteins. MAP2B encodes proteins of 1828 amino acid residues;

MAP2A is a splice variant having additional exon of 83 amino acid residues in the middle of the molecule. Rotary-shadowing electron microscopy revealed that MAP2A and MAP2B are highly elongated, flexible molecules of ~100 nm in length. Subsequent study revealed that there is an additional low-molecular-weight form, MAP2C, which is ~70 kDa on SDS-PAGE. In MAP2C, the N- and C-terminal ends are joined, resulting in the protein of 467 amino acid residues. The unique characteristic of MAP2 is that they have binding domain for the regulatory subunit of type II cyclic AMP-dependent kinase on the N terminus. The C-terminal portion of the MAP2 forms the microtubule-binding domain (Table 1 and Figure 3). The sequence of this region does not have homology to MAP1, but it has unique imperfect tandem repeats of 31–32 amino acids; similar repeats of 18 amino acids joined by less-conserved, variable ‘spacers’ of 13–14 amino acids. These repeats bind to the C-terminal ends of tubulin. These imperfect tandem repeats are also used in the microtubule-

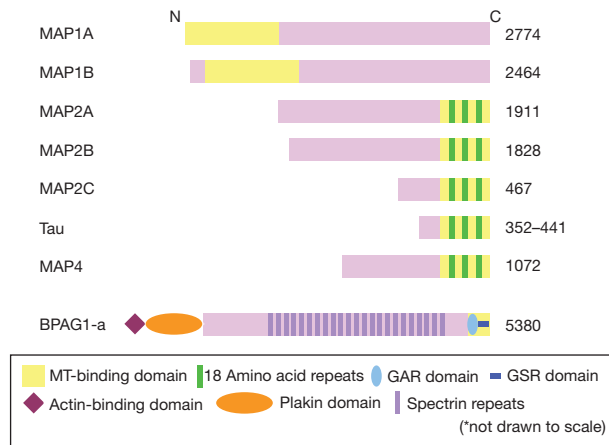


Figure 3 Schematic diagram of representative MAPs. Domain structures of MAP1A, MAP1B, MAP2A, MAP2B, MAP2C, tau, MAP4, and BPAG1-a are shown. Numbers of amino acids of representative molecules are shown on the right. MAP1A and MAP1B have a microtubule-binding domain on the N-terminal portion, which is domain enriched in the positively charged amino acids and contains the short motif (KKEX), repeated 11 and 21 times in MAP1A and MAP1B, respectively. C-terminal portion serves as a projection domain. MAP2A, MAP2B, MAP2C, tau, and MAP4 have a microtubule-binding domain on the C-terminal portion, which is a unique imperfect tandem of three or four basic repeats of 31–32 amino acids, similar repeats of 18 amino acids joined by less conserved, variable ‘spacer’ of 13–14 amino acids, and is quite well preserved among these MAPs. Molecules with three repeats are depicted in the drawing. BPAG1-a, one of the new MAPs, has actin-binding domain, plakin domain, spectrin-repeat-containing rod domain, GAR domain, and GSR domain. The GAR and the GSR domain together constitute the microtubule-binding domain in BPAG1-a as well as in MACF. In plakin, which is not depicted here, GAR domain is missing and GSR domain at the C-terminal portion can associate with microtubules. Not drawn to scale.

binding domain of tau and MAP4, which also reside in the C-terminal part of the molecule. Each molecule has three or four repeats. The central portion of the MAP2A and MAP2B molecule corresponds to the extended arm. Lacking this arm portion, MAP2C is revealed as filamentous molecule ~50 nm in length by the rotary-shadowing electron microscopy.

MAP2 is almost exclusively expressed in neurons. Importantly, MAP2A and MAP2B are specifically localized to the dendrites. It is often used as markers for the dendrites. There is a developmental regulation of MAP2 expression. MAP2C is expressed in juvenile neuron; it is found in cell bodies, dendrites, and axons. MAP2A is expressed in the adult, and MAP2B is expressed throughout the developmental stage.

Tau

Tau was the first MAP purified from the brain. It is heat stable, consists of several bands of 55–65 kDa on SDS-PAGE. It is now known that tau consists of six isoforms (ranging from 352 to 441 amino acid residues), which are splice variants of single tau gene. In the C-terminal half of the molecule, tau has three or four tandem repeats of 31 or 32 amino acids, as described above, which serves as the microtubule-binding domain (Table 1 and Figure 3). In rotary-shadowing electron microscopy, tau forms filaments of ~50 nm in length. Tau

forms ~20 nm short filamentous structures associated with microtubules. It is known that in the peripheral nervous system, additional isoform of 110 kDa is expressed. However, this isoform is less well characterized.

There is a developmental regulation of expression of tau isoforms. In the juvenile brain, only the three repeat isoforms are expressed, and in the adult brain, tau isoforms with both three and four repeats are expressed. Tau is posttranslationally modified by phosphorylation. A part of the heterogeneity of the tau bands on the SDS-PAGE results from phosphorylation. When antibodies that recognize all tau isoforms are used, tau is predominantly localized in the axon, with some expression in cell bodies and dendrites.

MAP4

Although MAP1, MAP2, and tau are rather abundantly expressed in the nervous system, their expression outside the nervous system is limited. MAP4 is a major MAP found outside the nervous system. It is a protein of 210 kDa on SDS-PAGE. MAP4 encodes protein of 1072 amino acid residues. It is composed of an N-terminal projection domain and a C-terminal microtubule-binding region, which has homology to the microtubule-binding domain of tau and MAP2 (Table 1 and Figure 3).

Control of Microtubule Dynamics

Microtubules are dynamic and unstable. By binding along the sides of the microtubules, MAPs control microtubule dynamics *in vitro* and *in vivo*. Many MAPs, including classical MAPs, stabilize microtubules. Classical MAPs were usually expressed predominantly in the nervous system. Microtubules in the neuron are usually much more stable than those in most interphase cells. Therefore, it is not surprising that one important function of classical MAPs is the stabilization of microtubules against disassembly. As described above, tau, MAP2, and MAP4 share conserved C-terminal three or four pseudorepeats, which serves as microtubule-binding domain. In tau, it is shown that one of these repeats bind to the tubulin C-terminal end, and other repeats bind to another internal site of tubulin. Therefore, these tandem repeats are likely to bind to microtubule in such a way to cross-link adjacent tubulin subunits and stabilize microtubules against disassembly. These bindings are controlled by phosphorylation. Both tau and MAP2 have many residues, which can be phosphorylated.

There are other modes of regulation of microtubule dynamics by MAPs. Microtubule is usually disassembled in cold temperatures. Stable tubule-only polypeptide (STOP) is a MAP known to stabilize microtubules against cold temperatures. A few MAPs are also known to destabilize microtubules, including katanin, Op18, and stathmin. In addition to structural MAPs, some of the kinesin superfamily proteins belonging to KIF2 class also destabilize microtubules.

Bundling of Microtubules: MAPs and Neurite Outgrowth

Neuron extends axon and dendrites from the cell body. In the axon and dendrites, microtubules run in parallel, in

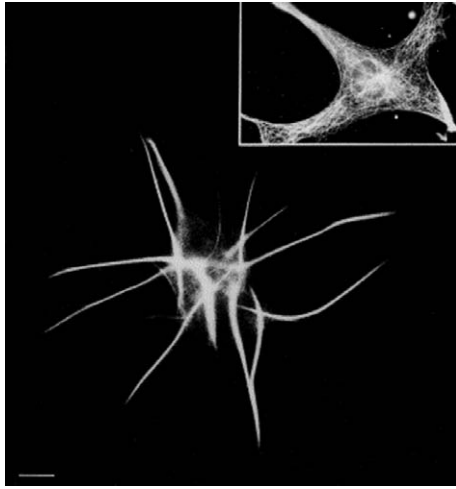


Figure 4 Microtubule bundle formation in fibroblasts transfected with tau cDNA. Immunofluorescence using anti-tubulin antibody. Long processes are formed and microtubule bundles extend into the processes. Upper right corner is the immunofluorescence of nontransfected control fibroblasts stained with antitubulin antibody. Microtubules radiate in various direction from the microtubule-organizing center near the nucleus. Scale = 10 μ m. Reproduced from Kanai Y, Takemura R, Oshima T, et al. (1989) Expression of multiple tau isoforms and microtubule bundle formation in fibroblasts transfected with a single tau cDNA. *Journal of Cell Biology* 109: 1173–1184, with permission of the Rockefeller University Press.

longitudinal array. In the quick-freeze, deep-etch electron microscopy, it can be shown that microtubules are highly cross-linked by meshwork of fine filamentous structures in axons and dendrites (Figures 1 and 2). MAPs form these cross-linkers. Moreover, MAPs induce microtubule bundle formation and outgrowth of processes from the cell. When classical MAPs are transfected and overexpressed in the cells, such as fibroblast or Sf9 cells, unusual microtubule bundles are formed, and long processes very much like neurites can be induced (Figures 4 and 5).

In addition, MAPs determine the spacing between neighboring microtubules to be axon like or dendrite like. Namely, MAPs expressed in the axon, such as tau and MAP2C, produce microtubule bundles similar to those seen in the axon (wall-to-wall distance of neighboring microtubules, ~ 20 nm), and MAPs expressed in the dendrites, such as MAP2A and MAP2B, produce microtubule bundles similar to those seen in the dendrites (wall-to-wall distance of neighboring microtubules, ~ 65 nm) (Figure 5). Therefore, in addition to the stabilization of microtubules, projection domains of MAPs induce microtubule bundle formation, induce process outgrowth, and determine the spacing between microtubules.

Analysis of Transgenic and Knockout Mice: Relation of MAPs to Neurodegenerative Diseases

Recently, genetically altered mice gave some insights into the role of individual MAPs *in vivo*. In general, lack of MAP

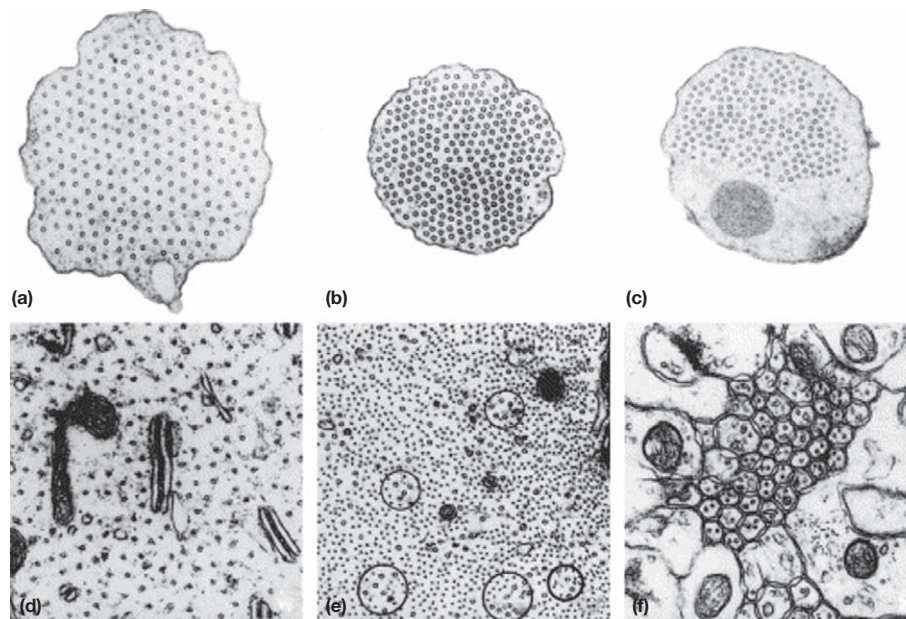


Figure 5 Cross-section electron micrographs of processes induced by transfection of MAP2 (a), MAP2C (b), and tau (c) into Sf9 cells. Each MAP2 forms microtubule bundles, but the organization is different. When MAP2 is transfected, wall-to-wall distance between nearest adjacent microtubules is ~ 65 nm, which is similar to the microtubule organization of Purkinje cell dendrite in the cerebellum (d). By contrast, when MAP2C or tau are transfected, the distance between microtubules is ~ 20 nm, which is similar to the microtubule organization of large myelinated axons in the spinal cord (inside the circle in (e)) or parallel fiber axons in the cerebellum (f). Reproduced from Chen J, Kanai Y, Cowan NJ, and Hirokawa N (1992) Projection domains of MAP2 and tau determine spacings between microtubules in dendrites and axons. *Nature* 360: 674–677, with permission of Nature Publishing Group.

produces the expected effect, but the overall effect is rather limited. This implies that there is certain overlap in the function of MAPs. In fact, when the double knockout mice were formed in the combination of two MAPs, the expected effect is much more severely observed. For example, in *tau*^{-/-} mice, an effect on the microtubule stabilization and bundle formation in the axon was observed, but the effect is evident mainly in the small caliber axon. Likewise, in *MAP2*^{-/-} mice, an effect on microtubule organization in dendrite is evident, but the effect is somewhat limited. However, when both tau and MAP1B, or both MAP1B and MAP2 are deleted, impaired microtubule bundle formation and outgrowth of axon or dendrite was markedly observed.

Tau has been implicated in the pathogenesis of certain cognitive neurodegenerative diseases. The diseases are sometimes grouped as tauopathies, which comprise a heterogeneous group of age-dependent cognitive disorders, including Alzheimer's disease. In these diseases, abnormally hyperphosphorylated tau forms filamentous aggregates within the cell. The aggregates do not contain microtubules, and the formation of aggregates is independent of microtubules. There is a loss of neurons in the affected brain area. In 1998, tau gene mutations were identified in patients with dementia and parkinsonism linked to chromosome 17 (FTDP-17). Transgenic mice with these mutations have been developed for an analysis.

New MAPs

Recently, a family of proteins of a very high molecular weight, plakin family, has been identified as cross-linker proteins that connect different cytoskeletal filaments. Some of the members, such as plectin, Bullous pemphigoid antigen 1 (BPAG1), and microtubule-actin cross-linking factor (MACF), have microtubule-binding domain, and they are likely to cross-link microtubule to intermediate filaments, or actin filaments. Microtubule-binding domain used in the plakin family proteins are of two kinds: Gas2-related (GAR) domain in BPAG1 and MACF; and glycine-serine-arginine (GSR) domain in plakin,

BPAG1, and MACF (Figure 3). That these new cross-linkers play an important physiological role is suggested by the fact that mutations in these proteins cause neurological defects *in vivo*. For example, mutations in BPAG1 cause degeneration of sensory neurons, and mutations in MACF cause deficits in axon outgrowth and dendritic sprouting.

See also: [Cell Architecture and Function: Kinesin Superfamily Proteins; Kinesins as Microtubule Disassembly Enzymes.](#)

Further Reading

- Alberts B, Johnson A, Lewis J, et al. (2002) *Molecular Biology of the Cell*. New York, NY: Garland Publishing.
- Chen J, Kanai Y, Cowan NJ, and Hirokawa N (1992) Projection domains of MAP2 and tau determine spacings between microtubules in dendrites and axons. *Nature* 360: 674–677.
- Desai A and Mitchison TJ (1997) Microtubule polymerization dynamics. *Annual Review of Cell and Developmental Biology* 13: 83–117.
- Garcia ML and Cleveland DW (2001) Going new places using an old MAP: Tau, microtubules and human neurodegenerative disease. *Current Opinion in Cell Biology* 13: 41–48.
- Hirokawa N (1991) Molecular architecture and dynamics of the neuronal cytoskeleton. In: Burgoyne RD (ed.) *Neuronal Cytoskeleton*, pp. 5–74. New York, NY: Wiley-Liss.
- Hirokawa N (1994) Microtubule organization and dynamics dependent on microtubule-associated proteins. *Current Opinion in Cell Biology* 6: 74–81.
- Homma N, Takei Y, Tanaka Y, et al. (2003) Kinesin superfamily protein 2A (KIF2A) functions in suppression of collateral branch extension. *Cell* 114: 229–239.
- Houseweart MK and Cleveland DW (1999) Cytoskeletal linkers: New MAPs for old destinations. *Current Biology* 9: R864–R866.
- Hyams JS and Lloyd CW (1994) *Microtubules*. New York, NY: Wiley-Liss.
- Kanai Y, Takemura R, Oshima T, et al. (1989) Expression of multiple tau isoforms and microtubule bundle formation in fibroblasts transfected with a single tau cDNA. *Journal of Cell Biology* 109: 1173–1184.
- Leung CL, Green KJ, and Liem RKH (2002) Plakins: A family of versatile cytolinker proteins. *Trends in Cell Biology* 12: 37–45.
- Takei Y, Teng J, Harada A, and Hirokawa N (2000) Defects in axonal elongation and neuronal migration in mice with disrupted tau and map1b genes. *Journal of Cell Biology* 150: 989–1000.
- Teng J, Takei Y, Harada A, et al. (2001) Synergistic effects of MAP2 and MAP1B knockout in neuronal migration, dendritic outgrowth, and microtubule organization. *Journal of Cell Biology* 155: 65–76.

Mitochondria in Myocardial Ischemia

F Di Lisa, University of Padua, Padua, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Diastole Period of relaxation of the heart muscle, accompanied by the filling of the chambers with blood that is followed by systole.

Heart failure A condition resulting from a variety of cardiac diseases associated with an inadequate pumping function of the heart.

Infarction Cellular death caused by a prolonged obstruction of the inflow of arterial blood that is generally caused by the thrombotic occlusion of an atheromatous lesion.

Ischemia Deficient supply of blood to a body part (as the heart or brain) that is due to obstruction of the inflow of arterial blood (i.e., coronary arteries in the case of myocardial ischemia).

Ischemic postconditioning An intrinsic process whereby brief periods of ischemia alternating with brief periods

of re-flow applied at the onset of reperfusion following sustained ischemia reduces myocardial infarct size.

Ischemic preconditioning An intrinsic process whereby repeated short episodes of ischemia alternating with brief periods of re-flow protect the myocardium against a subsequent prolonged ischemic insult that would be otherwise lethal.

Myocyte The type of cell found in muscles.

Reperfusion Restoration of the flow of blood to a previously ischemic tissue or organ.

Sarcomere The basic unit of a myocyte composed of long, fibrous proteins, mostly actin and myosin, that slide past each other when the muscles contract and relax.

Systole Period of contraction of the ventricles of the heart that causes the ejection of blood into the aorta and pulmonary trunk.

In the heart, oxygen supply is generally reduced by coronary obstruction resulting in a critical reduction of flow (ischemia). In the affected region deprived of oxygen, drastic changes of metabolism and lack of adequate washout determine an abnormal accumulation of ions and metabolites. The readmission of coronary flow (post-ischemic reperfusion), which is obviously necessary for the maintenance of cardiomyocyte viability, is beneficial only after short periods of ischemia. When the duration of ischemia is prolonged, reperfusion exacerbates functional and structural alterations induced by ischemia creating conditions that are hardly compatible with cell survival. The following paragraphs are focused on the alterations in energy-linked functions of mitochondria that are involved in cardiomyocyte death during ischemia and reperfusion.

Ischemia and Energy Metabolism

At a cellular level, the onset of ischemia is determined by an insufficient availability of oxygen for mitochondrial oxidations. As a consequence of the reduced or absent oxidative phosphorylation, intracellular creatine phosphate is rapidly depleted with a concomitant rise in P_i , both factors stimulating glycolysis and lactate formation. The accumulation of lactate and the hydrolysis of adenosine triphosphate (ATP) decrease intracellular pH. During the early phase of ischemia, the failure of contraction and the rigor contracture, which are the most relevant changes in myocardial function, are caused by intracellular acidosis and severe reduction in ATP content respectively (Figure 1). Besides these relevant functional changes, several cellular activities are modified by the fall in pH and ATP content, including the inhibition of glycolysis

which is observed under conditions of severe ischemia. A low pH and the accumulation of lactate decrease the activity of both phosphofructokinase (PFK) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH). In addition, the fall in ATP is associated with membrane binding and inactivation of PFK. Thus, the failure in mitochondrial ATP production, which initially stimulates glycogen breakdown and glycolysis, is followed by the inhibition of glycolysis that is the only alternative pathway for ATP synthesis in anaerobic cells.

Despite the arrest of the respiratory chain, the mitochondrial membrane potential ($\Delta\psi_m$) is not immediately collapsed at the onset of ischemia. In fact, $\Delta\psi_m$ is maintained by the inverse operation of the F₀F₁ adenosine triphosphatase (ATPase) that uses ATP generated by glycolysis. Therefore, during ischemia not only do mitochondria cease to be the major ATP producers of the cell, but they can also become its major consumers.

Information from Isolated Cardiomyocytes

The multifaceted relationship between ischemia-induced mitochondrial dysfunction and alterations of cardiac structure and function have been elucidated by studies in isolated cardiomyocytes. These cells show a rod-shaped morphology and do not display spontaneous contractions. Rhythmic contractions can be evoked by electrical stimulation. Under de-energizing conditions, rod-shaped myocytes can either change into square forms (rigor state or contracture) or hypercontract into rounded dysfunctional forms in which the typical sarcomeric striature is no longer distinguishable (hypercontracture).

At the single cell level, hypercontracture is the paradigm of irreversible damage and is considered equivalent to the abrupt rise in diastolic pressure in the absence of contractile activity

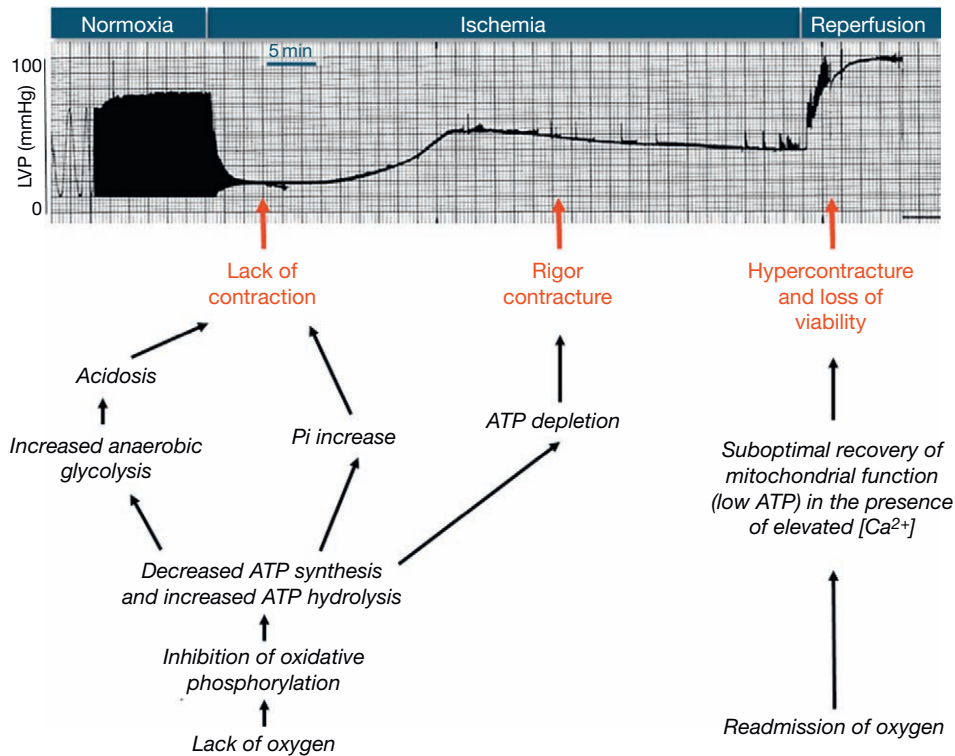


Figure 1 Relationships between mitochondrial dysfunction and the characteristic sequence of cardiac alterations occurring during ischemia and reperfusion. The upper panel is a typical example of the changes in left ventricular pressure (LVP) that are observed in isolated and perfused rodent hearts subjected to prolonged ischemia (60 min) followed by reperfusion. Pi, inorganic phosphate.

which is observed in the intact heart when reperfusion takes place after a prolonged ischemia. However, in the intact heart, irreversible damage is characterized by the loss of cell membrane integrity, easily demonstrated by the release of intracellular proteins in the coronary effluent (Figure 1). Conversely, hypercontracted myocytes often retain the integrity of sarcolemma, as demonstrated by the retention of small molecules such as intracellularly trapped fluorescent probes. The difference between single cells and intact hearts is likely explained by the lack of cell attachments. The same morphological distortion which determines the hypercontracture in isolated myocytes could probably cause the rupture of the sarcolemma in the intact heart by tearing off adjacent myocytes.

Mitochondria in the Transition from Reversible to Irreversible Injury

Prolongation of ischemia (i.e., more than 20 min of no-flow ischemia in perfused hearts) leads myocytes to a point of no return beyond which even readmission of oxygen fails to rescue cell viability.

Myocardial tissue exhibits a peculiar transition from reversible to irreversible damage. In every cell type, the lack of ATP production and a severe decrease of its content result in a slow degenerative process which eventuates in the loss of plasma membrane integrity with release of intracellular components. Thus, if the initial pathological stimulus (i.e., lack of oxygen) is not removed, the inhibition or the alteration of mitochondrial

function results in cell death. In addition to this common mode of cell death, which occurs gradually, the myocardium exhibits a peculiar transition to irreversible injury, which occurs suddenly when oxygen supply is re-established after a prolonged period of ischemia. In this case, the uncontrolled hypercontracture of myofilaments is associated with a rapid increase in cell membrane permeability followed by the release of intracellular constituents.

The sudden changes produced by reperfusion on myocardial viability are concomitant with massive accumulation of calcium within the mitochondrial matrix. This seminal finding that highlighted the role of intracellular Ca²⁺ overload in myocardial pathology indicates that the rapid transition toward cell death requires a coupled mitochondrial respiration. In fact, respiratory chain inhibition or uncoupling of oxidative phosphorylation blunts enzyme release occurring at the onset of post-ischemic reperfusion. Therefore, oxidative phosphorylation not only is essential for cell recovery, but can also contribute to the processes causing cell necrosis.

Two major questions arise from these seminal findings: (1) Why is mitochondrial function required to cause hypercontracture or tissue injury upon reperfusion? (2) Which are the causes of mitochondrial dysfunction?

The apparent paradox linking mitochondrial function with cell death is related to the different ATP requirements of contraction and relaxation at the myofilament level. A suboptimal mitochondrial function could produce low levels of ATP which, in the presence of even a modest rise in [Ca²⁺]_i, might be sufficient for the formation of rigor bonds but not for the

relaxation process, resulting in myofilament hypercontracture. Supporting this concept, at reperfusion hypercontracture is abrogated not only by preventing oxidative phosphorylation, but also by inhibiting myosin ATPase.

Regarding the causes of mitochondrial dysfunction, upon reperfusion mitochondria are exposed to high levels of Ca^{2+} that is increased in the cytosol due to both the failure of Ca^{2+} ATPases and uptake through the sarcolemma $\text{Na}^+/\text{Ca}^{2+}$ exchanger. Since mitochondrial Ca^{2+} uptake and ATP synthesis utilize the same driving force, namely the $\Delta\psi_{\text{m}}$, mitochondrial Ca^{2+} accumulation results inevitably in a significant decrease in ATP formation. In addition and importantly, as detailed in the following paragraphs, the rise in intramitochondrial $[\text{Ca}^{2+}]$ ($[\text{Ca}^{2+}]_{\text{m}}$) promotes opening of the mitochondrial permeability transition pore (PTP) and is associated with increased formation of reactive oxygen species (ROS). Eventually a vicious cycle is generated whereby high $[\text{Ca}^{2+}]_{\text{m}}$, PTP opening and ROS formation hamper mitochondrial function jeopardizing the maintenance of cell viability.

Although ROS can be generated at several intracellular and extracellular sites, in cardiac myocytes the major source is represented by mitochondria that are on the other hand a prominent target of oxidative stress. A slight generation of ROS has been observed in ischemic or hypoxic cardiomyocytes, especially in isolated cells. However, a major burst in ROS formation occurs at the onset of reperfusion providing a significant contribution to the rapid evolution toward cell death. Alterations of respiratory chain induced by ischemia are likely to contribute largely to mitochondrial ROS generation during reperfusion. However, it must be pointed out that besides respiratory chain complexes, especially I and III, three other systems have been described for ROS generation in mitochondria, namely 'monoamine oxidases' (MAOs), p66^{Shc} , and NADPH oxidase 4 (NOX4). While MAO are located to the outer mitochondrial membrane (OMM), p66^{Shc} is translocated from cytosol into the intermembrane space where it catalyzes electron transfer from cytochrome *c* to oxygen-generating H_2O_2 . The recent addition of NOX 4 to the list of mitochondrial systems generating ROS is more controversial, since also a localization to endo/sarcoplasmic reticulum has been reported. However, due to close vicinity, ROS formed at endo/sarcoplasmic reticulum would easily attack mitochondrial components. Interestingly, pharmacological inhibition or genetic deletion of these ROS-generating systems results in a significant decrease of reperfusion-induced loss of cardiomyocyte viability. Additional studies are required to elucidate the relationships among the mitochondrial systems that generate ROS especially during post-ischemic reperfusion.

The Mitochondrial PTP as a Target of Ischemic Injury and Pharmacological Interventions

A general consensus exists that the PTP is a major factor in determining cardiomyocyte death, and that PTP inhibition affords significant cardioprotection against the ischemia/reperfusion injury, a concept that has been successfully translated to clinical settings.

As detailed elsewhere in the encyclopedia, major functional and structural changes are induced in mitochondria, and

consequently to the whole cell, by PTP opening that causes a sudden increase in the permeability (i.e., permeability transition) of the inner mitochondrial membrane (IMM) to solutes with molecular masses up to 1500 Da. The PTP can be defined as a high-conductance channel located in the IMM. Its opening is favored by a decrease in mitochondrial membrane potential, elevation in matrix $[\text{Ca}^{2+}]$, and oxidative stress.

Among the many proteins described in PTP regulation, major attention has focused on the matrix protein cyclophilin (CyP) D due to its importance in pharmacological control of PTP opening and to the pivotal information provided by its genetic deletion. PTP opening is facilitated by binding of CyPD to the IMM in a process modulated by Ca^{2+} , Pi, and ROS. Notably, CyPD binding to the IMM is prevented by cyclosporine A (CsA) and by other molecules interacting with CyPD that are usually described as PTP inhibitors. However, since increasing $[\text{Ca}^{2+}]$ results in PTP opening even in the presence of these molecules, PTP desensitization describes better than PTP inhibition the effect of this family of compounds, a concept that also applies to mitochondria devoid of CyPD.

The relationship between CyPD and PTP has initially been demonstrated by pharmacology through the use of CsA, and then corroborated by using other CyPD ligands. However, conclusive evidence for a regulatory role of CyPD was provided by genetic studies demonstrating that the propensity of the PTP to open is reduced in mitochondria devoid of CyPD, while susceptibility to PTP opening and to cell death is increased when CyPD is overexpressed.

Although PTP structure is still elusive and its modulation is far from having been conclusively elucidated, the relationship between PTP opening and cell death is supported by a large body of evidence, especially in the field of acute myocardial ischemia and reperfusion. PTP opening is causally linked to the loss of cell viability as demonstrated by the reduction in infarct size obtained when PTP opening is counteracted by pharmacological inhibitors or by genetic ablation of CyPD.

PTP opening favors ATP depletion, deregulation of cellular Ca^{2+} homeostasis, and oxidative stress, all of which synergize in jeopardizing cell survival. The sequence of events can be summarized as follows. PTP opening determines an immediate collapse of $\Delta\psi_{\text{m}}$ that is followed by ATP depletion. When opening is prolonged, the initial uncoupling-like effect is rapidly followed by respiratory inhibition caused by loss of pyridine nucleotides and of cytochrome *c*. This latter event can be contributed by the rupture of the OMM that follows PTP-dependent matrix swelling, or by making more cytochrome *c* available for release through cristae junction widening followed by release through Bax/Bak channels through an otherwise intact OMM. The resulting inhibition of electron flow might explain the increased ROS formation induced by PTP opening. Since the latter event is favored by ROS, a vicious cycle of injury amplification is likely to be established, especially at the onset of reperfusion.

Regarding ischemia/reperfusion injury, during ischemia the build-up of factors favoring PTP opening (e.g., depolarization and increased Ca^{2+} and Pi matrix levels) is balanced by the presence of PTP antagonists (i.e., intracellular acidosis, high levels of Mg^{2+} , and adenosine diphosphate (ADP)), which may prevail and prevent PTP opening. Conversely, upon reperfusion, the recovery of pH together with a burst in ROS

formation in the presence of high-matrix concentrations of Ca^{2+} and Pi creates optimal conditions for PTP opening in spite of the antagonizing effect of $\Delta\psi_m$ recovery. The concept that PTP is more likely to open upon reperfusion is supported by findings obtained in intact hearts using direct methods for PTP assessment. Notably, the promoting effect of Ca^{2+} is also due to indirect effects that add to its direct action on mitochondria. For instance, arachidonic acid released from phospholipids through the action of Ca^{2+} -dependent phospholipase A_2 and the activation of the Ca^{2+} -dependent proteinase calpain are likely to further facilitate PTP opening.

CsA-dependent cardioprotection is likely to involve also mechanisms other than CyPD inhibition. For instance, the CsA/CyPA complex binds to and inhibits the cytosolic phosphatase calcineurin preventing mitochondrial fragmentation. In fact, due to its Ca^{2+} -dependent phosphatase activity, calcineurin promotes the intracellular redistribution of dynamin-related protein 1 (Drp1) that only in its dephosphorylated form translocates to mitochondria causing their fragmentation. The relevance of this process in myocardial ischemia is highlighted by recent reports showing that genetic or pharmacological inhibition of Drp1 affords a significant protection against various injury conditions including post-ischemic reperfusion. Therefore, PTP opening might be the consequence and not only the cause of derangements in mitochondrial structure and function.

Mitochondria and Endogenous Self-Defenses Triggered by Conditioning Stimuli

Endogenous defense mechanisms are triggered by conditioning stimuli to maintain cardiomyocyte viability during ischemia or reperfusion. The term 'ischemic preconditioning' (IPC) describes the process whereby repeated short episodes of ischemia and reperfusion protect from a subsequent prolonged ischemia. The deleterious effects of post-ischemic reperfusion have been shown to be limited also by performing a series of very short periods of ischemia and reperfusion at the time of reperfusion. This maneuver is termed 'ischemic post-conditioning' (IPoC) and has received great attention as a promising protective strategy in clinical settings.

Both IPC and IPoC have been related to PTP based upon the following major lines of evidence: (1) IPC results in a decreased occurrence of PTP opening during post-ischemic reperfusion; (2) IPC and IPoC reduce PTP susceptibility in mitochondria isolated during reperfusion; and (3) cardioprotection by IPC and IPoC is similar to that obtained using compounds or interventions that decrease PTP opening.

The notion that the PTP contributes to mediate IPC has been corroborated by observations suggesting that CyPD inhibition or deletion abolishes the cardioprotective effects of IPC itself. This apparent paradox reminds of, and fits well with, the concept that a slight increase in ROS formation during the preconditioning phase prevents the large accumulation in ROS induced by reperfusion, and thus results in maintenance of myocardial viability.

The mechanisms by which IPC and/or IPoC affect mitochondrial function and structure, as well as PTP opening, have not yet been elucidated. Evidence has been provided that IPC/

IPoC-induced changes in signaling pathways are related to the increased resistance of mitochondria to ischemia/reperfusion (I/R) injury. For instance, IPC-induced inactivation of glycogen synthase kinase 3β (GSK- 3β) results in PTP inhibition. The constitutively active GSK- 3β would be inactivated by phosphorylation at Ser⁹ catalyzed by upstream protein kinases. These have been grouped under the acronyms of RISK (reperfusion injury salvage kinase), including phosphatidylinositol-3-OH kinase-Akt and ERK 1/2, and SAFE (survivor-activating factor enhancement) describing the activation of the cytokine tumor necrosis factor α and signal transducer and activator of transcription-3 (STAT-3). In addition, activation of cyclic guanosine monophosphate (GMP)-dependent protein kinase (PKG), cyclic adenosine monophosphate (AMP)-dependent protein kinase (PKA), and Ca^{2+} -dependent protein kinase ϵ (PKC ϵ) have been reported to contribute to the maintenance of mitochondrial function in hearts undergoing I/R, whereas PKC δ (together with GSK- 3β) appears to facilitate mitochondrial dysfunction.

Despite the large amount of data supporting the relevance of (de)phosphorylation processes in IPC-mediated mitochondrial and cardiomyocytes protection, the field is still controversial. First, the mechanisms linking signaling pathways activated in the cytosol with energy-linked process occurring within the IMM are far from being elucidated. Candidate (de) phosphorylation reactions occurring at the level of the OMM might involve the voltage-dependent anion channel (VDAC), thus affecting adenine nucleotide traffic between mitochondria and cytosol, and/or binding of the anti-apoptotic protein Bcl-2. A PTP-regulatory role has been proposed also for IPC-induced opening of the mitochondrial K_{ATP} channel by PKC ϵ , which may cause a slight increase in H_2O_2 formation eventually causing PTP inhibition. Second, in some experimental models activation of the RISK pathway appears to be independent of any protective effects. It has been suggested that rather than causing direct and immediate modifications in protein (de)phosphorylation, IPC reduces the extent of oxidative stress during ischemia and reperfusion. Although the burst in ROS formation occurring during reperfusion is clearly reduced in IPC-treated cardiomyocytes, the mechanism by which IPC or IPoC might affect oxidative stress is yet ill defined.

See also: [Bioenergetics: Calcium Transport in Mitochondria](#); [Mitochondria: Source and Target of Free Radicals](#); [Mitochondrial Dynamics](#); [Mitochondrial Permeability Transition Pore](#); [Monoamine Oxidase](#); [Reactive Oxygen and Nitrogen Species: Interactions with Mitochondria and Pathophysiology](#).

Further Reading

- Brookes PS, Yoon Y, Robotham JL, Anders MW, and Sheu SS (2004) Calcium, ATP, and ROS: A mitochondrial love-hate triangle. *American Journal of Physiology Cell Physiology* 287: C817–C833.
- Di Lisa F, Carpi A, Giorgio V, and Bernardi P (2011) The mitochondrial permeability transition pore and cyclophilin D in cardioprotection. *Biochimica et Biophysica Acta* 1813: 1316–1322.
- Di Lisa F, Kaludercic N, Carpi A, Menabò R, and Giorgio M (2009) Mitochondrial pathways for ROS formation and myocardial injury: The relevance of p66(Shc) and monoamine oxidase. *Basic Research in Cardiology* 104: 131–139.

- Di Lisa F, Menabò R, Canton M, and Petronilli V (1998) The role of mitochondria in the salvage and the injury of the ischemic myocardium. *Biochimica et Biophysica Acta* 1366: 69–78.
- Dorn GW and Scorrano L (2010) Two close, too close: Sarcoplasmic reticulum–mitochondrial crosstalk and cardiomyocyte fate. *Circulation Research* 107: 689–699.
- Halestrap AP, Clarke SJ, and Javadov SA (2004) Mitochondrial permeability transition pore opening during myocardial reperfusion – a target for cardioprotection. *Cardiovascular Research* 61: 372–385.
- Lee JA and Allen DG (1991) Mechanisms of acute ischemic contractile failure of the heart. Role of intracellular calcium. *Journal of Clinical Investigation* 88: 361–367.
- Murphy E and Steenbergen C (2007) Preconditioning: The mitochondrial connection. *Annual Review of Physiology* 69: 51–67.
- Neely JR and Feuvray D (1981) Metabolic products and myocardial ischemia. *American Journal of Pathology* 102: 282–291.
- Ovize M, Baxter GF, Di Lisa F, et al. (2010) Postconditioning and protection from reperfusion injury: Where do we stand? Position paper from the Working Group of Cellular Biology of the Heart of the European Society of Cardiology. *Cardiovascular Research* 87: 406–423.
- Reimer KA and Jennings RB (1992) Myocardial ischemia, hypoxia and infarction. In: Fozzard H, Haber E, Jennings RB, Katz A, and Morgan H (eds.) *The Heart and Cardiovascular System*, 2nd edn., pp. 1875–1973. New York: Raven Press.
- Silverman HS and Stern MD (1994) Ionic basis of ischaemic cardiac injury: Insights from cellular studies. *Cardiovascular Research* 28: 581–597.
- Yellon DM and Hausenloy DJ (2007) Myocardial reperfusion injury. *New England Journal of Medicine* 357: 1121–1135.

Mitochondrial Auto-Antibodies

H Baum, King's College London, London, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

CD4⁺-T cell T-lymphocyte, with a receptor recognizing a specific short peptide, derived from an antigen and presented by an antigen-presenting cell and, thereafter, primed to help B-cells that also present that peptide to secrete antibodies against the antigen.

Epitope The specific target recognized by a T-cell or antibody.

Immunofluorescence A technique for detecting antibodies that bind to specific structures on tissue sections, by

treatment with a fluorescently labeled, anti-immunoglobulin antibody.

Lipoyl residue The amide of 6,8-dithio-octanoic acid, the oxidized, disulfide form of which is reductively acylated in the catalytic cycle of the 2-oxoacid dehydrogenase complexes.

Molecular mimicry Similarity in structure, either accidental or evolutionarily selected, between two antigens, leading to immunological cross-reactivity.

Detection of Mitochondrial Auto-Antibodies

Anti-mitochondrial antibodies (AMAs) have been detected by immunofluorescence (IFL) (where they are visualized as bright granular staining in cell types with high mitochondrial content), or by applying complement fixation tests (CFTs), enzyme-linked immunosorbent assays (ELISAs), Line or Dot-blot immunoassays, or Western immunoblotting (IB) to tissue extracts, purified subcellular constituents, or specific, recombinant mitochondrial species. Proof that mitochondria were the target of antibodies detected by such methods was provided in early studies, such as the landmark paper of Walker et al. by the specific absorption of serum reactivity using preparations of mitochondria. It was subsequently recognized that AMAs recognizing different mitochondrial antigens might characterize different clinical conditions. The molecular identity of many, but not all, of these antigens is now established. Not all of these AMAs are detectable by all of the above methodologies. The preferred test in some general diagnostic laboratories is IFL on a composite substrate of rodent tissues, complemented by ELISAs when appropriate. In other laboratories, particularly when primary biliary cirrhosis (PBC) is the concern, the preferred test is ELISA. In the research laboratory, and where specific AMAs are under investigation, the preferred assays are ELISAs and IB, using purified or recombinant antigens, or synthetic peptides.

The M-Classification of AMAs

In the early 1980s, an arbitrary classification of AMAs, as anti-M1-9, was developed, mainly by Berg's group in Tübingen, on the basis of patterns of IFL and/or properties of target antigens as determined by various assay procedures. It is noteworthy that in all these cases, the AMAs are not species specific. Although this M-classification is now more or less obsolete, and does not include all known mitochondrial targets for AMAs, it is still a helpful convention for listing reactivities (Table 1).

Anti-M1 AMAs and Cardiolipin

The M1 pattern of IFL is exhibited exclusively with sera of patients with untreated secondary syphilis. The reaction is abolished completely by absorption of sera with cardiolipin, the characteristic phospholipid of the mitochondrial inner membrane. However, the simple conclusion that anti-M1 antibodies are specific for cardiolipin is complicated by another observation. The earlier diagnostic test for syphilis, the Wasserman reaction, is based on CFTs against cardiolipin. However, not all sera positive in this test show an M1 pattern of IFL. Moreover, Wasserman-positive sera that are negative for M1 give a strongly positive ELISA reaction against purified cardiolipin, whereas anti-M1 AMAs are scarcely detectable by ELISAs using purified cardiolipin, while they react strongly in ELISAs against submitochondrial particles. Interestingly, M1-negative, Wasserman-positive sera are frequently false positives for syphilis and may instead reflect collagen disorders such as systemic lupus erythematosus (SLE). It seems likely that these anomalies reflect antibodies to different epitopes in cardiolipin, the different assay systems exposing different conformations of the same molecule. (Indeed some of the anti-cardiolipin antibodies in M1-negative sera of SLE patients may be detected as antibodies against nucleic acids, cross-reactivity arising by virtue of the similarity between the glycerol phosphate arrays in cardiolipin arrays and the ribose phosphate backbone of nucleic acids.) Anti-M1 AMAs probably recognize a more native form of the lipid, as it exists in the intact mitochondrial membrane (and may arise from immunological cross-reactivity with a membrane constituent of the agent of syphilis, *Treponema pallidum*).

Anti-M2 AMAs

The most important mitochondrial auto-antibodies, anti-M2 AMAs, exhibit a characteristic pattern of IFL and were initially defined both by their disease specificity and by properties of the antigen that they recognized. This was shown to be trypsin sensitive and associated with (but detachable from) the inner surface of the inner membrane of all mitochondria tested, from

Table 1 The M-classification of specificities of mitochondrial auto-antibodies

Antigen	Occurrence of auto-antibodies
M1	Secondary syphilis
M2	PBC
M3	Phenopyrazone-induced pseudolupus syndrome
M4	PBC (prognostic indicator?); other chronic liver diseases
M5	Subgroup of systemic lupus; anti-phospholipid syndrome
M6	Iproniazid-induced hepatitis
M7	Cardiomyopathies
M8	PBC (prognostic indicator?)
M9 ^a	PBC (prognostic indicator?)

^aIn fact this antigenic specificity is not mitochondrial (see text).

human to fungal. Anti-M2 antibodies were also shown to cross-react with several bacteria, anti-mitochondrial activity, for example, being absorbed out of PBC sera by *Escherichia coli*.

Association with primary biliary cirrhosis

PBC is a chronic, cholestatic liver disease, predominantly affecting middle-aged women. It is characterized by spontaneous inflammatory damage to the epithelial cells of the bile duct (BEC). While the clinical presentation and disease progression are variable, and the disease may be virtually asymptomatic for many years, the end stage is cirrhotic liver failure, and the disease is the second most frequent reason for liver transplantation in the developed world. Although the most clinically significant pathology relates to the liver, PBC is actually a multi-system disorder and may feature lachrymal salivary and pancreatic hyposecretion. The paper by Walker, referred to earlier, and many subsequent studies, demonstrated clearly that PBC is characterized by high and persistent titers of AMAs of a kind not seen in normal subjects or other diseases. Since the description of this close association, the detection of this class of AMAs, designated as anti-M2, has acted as a powerful diagnostic tool. Anti-M2 antibodies (both IgG and IgM) are found in up to 98% of PBC patients, the precise figure depending on the sensitivity of the assay used. Interestingly, these AMAs are not restricted to serum, being found in the bile, saliva, and urine of patients with PBC. Moreover, anti-M2 AMAs have turned out to be predictors for the disease; asymptomatic subjects, accidentally having been identified as being positive for anti-M2 AMAs, are frequently found to develop the disease subsequently.

Identity of M2 antigens

Western IB soon revealed that M2 actually consists of a family of up to seven co-purifying peptides of differing molecular weight, encompassing a spectrum from around 75 to 36 kDa. Different PBC sera demonstrated different patterns of reactivity against these (see Table 2), but virtually all recognized a predominant peptide that was eventually identified as being the E2 component (dihydrolipoamide acetyltransferase) of the pyruvate dehydrogenase complex. Thereafter, it soon became clear that the other M2 antigens were also constituents of one or other of the 2-oxo-acid dehydrogenase multienzyme complexes (OADCs): pyruvate, 2-oxoglutarate, and branched-chain 2-oxo acid dehydrogenase complexes (PDC, OGDC, and BCOADC, respectively). Each of the three

Table 2 Frequency of occurrence of anti-M2 AMA to specific antigens in PBC

Complex	Subunit	Apparent M_r (kDa)	AMA-positive (%) ^a
PDC	E1 α	41	53
PDC	E1 β	34–36	4
PDC	E2	70–74	96
PDC	E3BP (X)	50–56	96
OGDC	E2	48–50	74
BCOADC	E1 α	46	50
BCOADC	E2	50–52	56
All	E2's		50
At least one	E2		98

^aApproximate average of published data using defined antigens.

complexes occupies a key position in energy metabolism in aerobic cells. Each consists of multiple copies of at least three enzymes (E1, E2, and E3), which are encoded by genes in the nucleus and separately imported into mitochondria for assembly into high-molecular-weight multimers localized to the matrix aspect of the inner membrane. The structural core of these complexes is provided by E2 (the acyltransferase), to which multiple copies of E1 (the decarboxylating dehydrogenase) and E3 (dihydrolipoamide dehydrogenase) are noncovalently bound. The E2 enzymes are notably homologous in structure as between different complexes and different species (e.g., there is 52% similarity in sequence as between rat PDC-E2 and OGDC-E2 of *E. coli*). E3 is common to all three complexes, whereas E1 and E2 are unique to each complex. PDC contains a fourth polypeptide with a structural role, the E3-binding protein (E3BP), once termed 'protein X'. This contains a lysine-linked lipate moiety, located in a domain that is strongly homologous to corresponding domains in PDC-E2.

In view of this latter homology, it is not surprising that all sera recognizing PDC-E2 also bind E3BP and that the antibodies to the two peptides are cross-reactive. The other members of the M2 family were shown to include the E2's of OGDC and BCOADC. However, notwithstanding the general similarity between the respective E2's, there seems only to be a limited degree of antibody cross-reactivity between them and only about 50% of PBC sera recognize all three. (On the other hand, AMAs against a particular E2 recognize that peptide in mitochondria of all species is being tested.) About half of all PBC sera also recognize the α -subunit of PDH-E1 (with the β -subunit being rarely recognized). Additionally, human auto-antibodies, specific for BCOADC-E1 α , were found, in a more recent study, in the sera of almost half of the PBC patients tested. A summary of frequency of occurrence of AMAs to individual members of the M2 family is given in Table 2. The E3 component, a common subunit of all three complexes, is not listed in the table. This is because, although specifically anti-mammalian E3 antibodies are found in PBC, these are in relatively low amount and are not totally disease specific.

Enzyme inhibition by anti-M2 AMAs

A notable characteristic of AMAs in PBC sera is their capacity to block the catalytic function of the 2-OAD multienzyme complexes *in vitro*. For example, sera with AMAs recognizing PDC-E2 rapidly inhibit PDC at dilutions of 1/500–1/5000, much of this inhibitory activity being removed by pre-

absorption of the sera with recombinant PDC-E2. Similarly, affinity-purified antibody against PDC-E1 α also specifically inhibits PDC reactivity.

Possible target for anti-M2 AMAs on bile duct cells

What may be a key fact relating to the occurrence and disease association of AMAs in PBC is the reactivity that has been observed of anti-M2 AMAs (and of selected monoclonal antibodies against PDC-E2) with some component at the surface of a subset of BEC in livers of PBC patients. Conflicting views have been that this antigen is an inappropriately localized component of PDC, or that it is a different molecule with a sequence similarity to a dominant B-cell epitope. What does seem clear is that susceptibility of BEC is not host specific, since PBC may recur in allogeneic-transplanted livers. What is also very significant is that PDC-E2 has more recently been shown to be uniquely conserved in BEC following apoptosis, with evidence that it remains accessible to the immune system within apoptotic blebs. However, it has yet to be unambiguously demonstrated that the pathogenetic focus of the disease is reactivity of AMA with PDC-E2 at the surface of BEC.

Cross-reactivity of anti-M2 AMAs against bacterial E2's

Early reports of cross-reactivity of PBC-specific AMAs against a variety of bacterial species were confirmed in studies of reactivity against individual microbial E2's of the three OADC complexes. (There is no homology between microbial and corresponding mitochondrial E1's; neither is there reactivity of E1-specific AMAs against bacteria.) Further evidence of cross-reactivity is that antibodies raised in rabbits, against bacterial mutants having fragile cell walls, were reported to react with a typical M2 pattern on IB against mitochondria. However, detailed absorption/elution and other studies have led to a current view that, while there is indeed some cross-reactivity between corresponding enzymes, PBC sera contain distinct sets of antibodies against mitochondrial and microbial OADCs, respectively.

Anti-M2 AMAs and chronic bacterial infection

In a study of women with asymptomatic recurrent bacteriuria, a majority of the patients were found to be weakly positive for antibodies to PDC-E2 by immunoblot, in the absence of evidence of liver disease. AMAs were also detected by ELISAs and IB in 43% of patients with active pulmonary tuberculosis and in patients with leprosy. However, it is not clear whether what is being detected in these cases are true AMAs or antibodies against corresponding microbial enzymes that are partially cross-reactive against mitochondrial peptides.

Epitopes recognized by anti-M2 AMAs

Most studies on B-cell epitopes in PBC have focused on the E2 antigens. Here, it is clear that the main immunogenic regions are the lipoyl domains. Short peptides (e.g., of around 20 residues in length) encompassing the lipoyl-binding site are rather weakly antigenic, and full immunological reactivity is only seen for much longer peptides. Thus, for PDC-E2, the 93-residue sequence, 128–221, of the inner lipoyl domain is required for maximal antibody binding. The outer lipoyl domain, although containing a highly homologous region around the lysine residue that is bound to lipoate, is only weakly antigenic. These and other findings suggest that the

overall epitopes are conformational ones, but centered on the lipoyl-binding sites. The strong homologies between these domains among the three OADCs, and between mitochondrial and microbial enzymes, may account for partial cross-reactivities, with extended flanking regions being responsible for specificity and full reactivity. There is some disagreement as to whether lipoyl residues themselves are necessary for full reactivity, but corresponding peptides not containing lipoate are undoubtedly antigenic.

Stimulus for anti-M2 AMAs production in PBC

There is good evidence that B-cell clones producing AMAs are maximally stimulated, and it is noteworthy that the dominant epitopes of the CD4⁺ T-cells, that provide the help for antibody production, are short peptides of the lipoyl-binding domain. (Similar peptides are also the dominant epitopes of the cytotoxic CD8⁺ T-cells that infiltrate the bile ducts in this disease.) What then is the antigenic drive in PBC? It cannot be a secondary response to tissue damage, since M2 AMAs are not detectable in many other destructive diseases of the liver.

As mentioned earlier, chronic microbial infection may be characterized by low titers of anti-M2 AMAs. There is also epidemiological evidence of association between recurrent urinary tract infection (UTI) and PBC. Moreover, UTI is particularly common in middle-aged women, the group in which PBC most commonly occurs. Therefore, it is suggested that the initial stimulus for an immune reaction directed specifically at the E2's of the OADCs (and E3BP of PDC) is chronic exposure, in the context of co-stimulatory signals (and probably a particular genetic background), to microbial peptides that share strong sequence similarities with the lipoyl-binding domains at the core of the dominant epitopes. Expansion of reactivity and specificity to incorporate the entire domain, and then also to recognize E1 subunits, would be by epitope spreading. Such microbial mimics might not necessarily be the homologous OADCs. A number of other bacterial and viral peptides have been identified with strong sequence similarities to the core peptidyl epitope of PDC-E2, and have been shown, by ELISAs, to react specifically with PBC sera, the antibodies being cross-reactive with that core peptide. In particular, it has recently been shown that the amino acid sequences of two proteins of the ubiquitous, free-living bacterium, *Novosphingobium aromaticivorans*, present high degrees of homology with the main human PDC-E2 autoepitope, and that reactivity of these proteins against all anti-PDC-E2 positive sera was exceptionally high.

A parallel scenario to microbial infection has recently been suggested as initiating the autoimmune process, and this is modification of lipoyl residues of E2's by environmental xenobiotics. The central role of the liver in xenobiotic metabolism would then be a rationale for the location of the disease.

One other important fact needs bearing in mind in any speculation about antigenic drive. Once it is established, it persists, even though there might no longer be any relic of the triggering process. Thus, even after liver transplantation, high titers of anti-M2 AMAs are still found.

Role of AMAs in the pathogenesis of PBC

Anti-M2 AMAs are so focused in specificity and so disease specific that it has to be assumed that they are relevant to

disease pathogenesis, but most workers have concluded that it is T-cells rather than the antibodies themselves that are the direct cause of tissue damage. For example, high titers of AMAs can be raised in experimental animals, for example, by immunization with recombinant human PDC-E2, with no evidence of liver damage even after a prolonged period, and only a minority of patients exhibit unambiguous recurrence of PBC following liver transplantation, despite the persistence of AMAs. However, in the former case, fine specificities of the AMAs may not be identical to those of the PBC-specific antibodies, and in the latter, immunosuppressive drugs may be protective. If the AMAs were to be directly cytotoxic, it might be by inhibition of mitochondrial function following uptake into BEC. Such uptake might be mediated by binding to the elusive antigen that might be the primary recognition target at the surface of BEC.

Anti-M3 AMAs and Pseudolupus Syndrome

Some time ago, it was recognized that there was an association between the taking of the drug phenopyrazone, for cardiovascular disease, and a disorder designated as pseudolupus syndrome (PLE). This syndrome was characterized by AMAs that could be demonstrated by IFL, CFTs, and ELISAs, but not by IB. The antigen, classified as M3, was shown to be associated with the outer mitochondrial membrane and was not organ specific. High titers of anti-M3 were detectable during the acute phase of the disease, but they declined on cessation of treatment with phenopyrazone, a drug that has since been withdrawn. No further cases of PLE have been reported, and the M3 antigen has not been further characterized. However, it has been speculated that it might be a drug-metabolizing enzyme such as monoamine oxidase (MAO), rendered immunogenic through tight binding to the drug.

Anti-M4 AMAs and Sulfite Oxidase

In addition to anti-M2, another complement-fixing antibody has been shown to be associated with PBC. The target antigen in this case, designated as M4, was found to co-purify with the outer mitochondrial membrane and be insensitive to trypsin. It was identified as being sulfite oxidase, an enzyme of the inter-membrane space, a precursor of which binds to the outer membrane. Sulfite oxidase in an evolutionarily highly conserved enzyme, also found in bacteria, but enzyme activity does not seem to be inhibited by M4-positive PBC sera. Nevertheless, recombinant sulfite oxidase was shown by ELISA and IB to react with around 5% of sera from patients with PBC.

The clinical significance of the detection of anti-M4 in addition to anti-M2 AMA is controversial. There have been claims that anti-M4 AMAs reflect a mixed-form pathology of PBC with histological features of chronic active hepatitis, and/or that their detection is an indicator of poorer prognosis toward the end stage of the disease. However, it is now clear that this reactivity is not specific for PBC, and is indeed characteristic of chronic liver disorders in general, with the highest reactivity being seen in primary sclerosing cholangitis.

Anti-M5 AMAs and Phospholipid

Mention has been made that false Wasserman-positive tests may reflect a different conformation of cardiolipin than that

recognized by the M1-positive, syphilis-specific pattern of IFL, and are frequently associated with SLE. Such false-positive reactions often correlate with a characteristic IFL pattern, designated as M5. This pattern is seen with both a subset of SLE and patients with antiphospholipid syndrome, characterized by thrombocytopenia and recurrent fetal loss. While anti-M5 reactivity may be absorbed out by inner mitochondrial membranes, no specific reactivities have been detected against any mitochondrial proteins, and it is concluded that M5 is some form of phospholipid, although the reactivity is not absorbed out by cardiolipin liposomes or cardiolipin-glycoprotein complexes.

Anti-M6 AMAs and Immuno-Allergic Hepatitis

Anti-M6 AMAs have been detected, by a characteristic pattern of IFL, in sera of patients who had developed acute hepatitis associated with taking the anti-depressive drug iproniazid. The M6 antigen could be detected by CFTs and ELISAs, was partially tissue specific (liver-kidney), and was associated with the inner membrane. The antigen has now been identified as monoamine oxidase B (MAO-B) and the activity of this enzyme against substrates such as tyramine is inhibited by these AMAs. As MAO-B is irreversibly inhibited by iproniazid, the stimulus for AMA production may arise from metabolite/enzyme haptenization as also suggested, above, in the case of anti-M3.

Anti-M7 AMAs and Flavoenzymes

Around 60% of patients with acute myocarditis and 30% of patients with cardiopathy of unknown etiology exhibit a class of AMAs designated as anti-M7. The antibodies in acute myocarditis are predominantly of the IgM class and disappear on recovery, while in chronic disease they are IgG and persist. Such antibodies are not seen in coronary artery conditions and have been considered as reflecting infective disease. The target antigen was originally reported as being heart specific and tightly associated with the inner mitochondrial membrane, but it is now recognized that this class of AMA is directed generally against flavoenzymes. Hence, in heart mitochondria a dominant target is the flavoprotein of succinate dehydrogenase, while in liver it includes sarcosine dehydrogenase and dimethylglycine dehydrogenase.

Anti-M8 AMAs and Anti-M9 AMAs in PBC

Two other sets of AMAs have been claimed to be specific to PBC and, like anti-M4, important in prognosis of disease progression. One of these recognizes a mitochondrial target categorized as M8. This is an outer membrane, nonorgan-specific, trypsin-sensitive antigen, detectable by CFTs and ELISAs but not by IB. To date, M8 is defined only by the protocol for its preparation. The other specificity, categorized as M9, was defined in terms of a trypsin-insensitive constituent of the inner membrane of liver mitochondria. However, it has since been identified as the nonmitochondrial, cytoplasmic enzyme, glycogen phosphorylase, which presumably co-purifies with inner-membrane fragments on ultra-centrifugation. Hence, anti-M9 antibodies can no longer be considered as AMAs.

Other AMAs

Naturally Occurring Anti-Mitochondrial Antibodies

Sensitive detection methods reveal that sera of healthy individuals exhibit a spectrum of naturally occurring anti-mitochondrial antibodies (NOMAs), perhaps reflecting an arm of the normal immunoregulatory network, and/or a history of exposure to microbial antigens with structures that mimic those of mitochondrial constituents. Where NOMA reactivities against particular antigens have been studied, their fine specificities have been shown to differ from those of disease-associated AMAs recognizing the same molecular species. Interestingly, it has been reported that there is a relative lack of NOMA in patients with PBC, which might be a clue to their normal role in immune regulation.

AMAs Arising from Tissue Damage

Following acute myocardial infarction, many patients demonstrate the short-term production of AMAs against human heart mitochondria. These cross-react against mitochondria from human skeletal muscle, but not with heart mitochondria of other species. Similarly, synovial B-cells in rheumatoid arthritis have been reported as producing AMAs, the finding being interpreted as reflecting local tissue destruction. Such an interpretation may also apply to the demonstration, in insulin-dependent diabetes, of AMAs against the mitochondrial enzyme of pancreatic β -islet cells, glycerophosphate dehydrogenase.

Pathogenetic AMAs against the Adenine Nucleotide Transporter

Myocarditis and dilated myocarditis are diseases related to viral infection, and (independently of the anti-M7 AMAs described above) are characterized by the production of organ-specific

AMAs directed at the mitochondrial adenosine diphosphate/adenosine triphosphate (ADP/ATP) carrier. These antibodies are cytotoxic when tested against guinea pig ventricular myocytes, and in mouse heart, where the transport activity of the carrier has been shown to decline in a manner related to the decrease of cardiac function. The production of these AMAs declines over a period of 2–3 months, and it has been speculated that the underlying mechanism might involve molecular mimicry with a viral constituent. However, they may arise as a secondary response to virally induced tissue injury.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Glycerolipid Structure, Function, and Synthesis in Eukaryotes; **Metabolism Vitamins and Hormones:** Primary Biliary Cirrhosis.

Further Reading

- Bogdanos DP, Baum H, Sharma UC, et al. (2001) Antibodies against homologous microbial caseinolytic proteases P characterize primary biliary cirrhosis. *Journal of Hepatology* 36: 14–21.
- Gershwin ME, Ansari AA, Mackay IR, et al. (2000) Primary biliary cirrhosis: An orchestrated immune response against epithelial cells. *Immunological Reviews* 174: 210–225.
- Haydon GH and Neuberger J (2000) PBC: An infectious disease? *Gut* 47: 586–588.
- Mackay IR, Whittingham S, Fida S, et al. (2000) The peculiar autoimmunity of primary biliary cirrhosis. *Immunological Reviews* 174: 226–237.
- Selmi and Gershwin ME (2009) The role of environmental factors in primary biliary cirrhosis. *Trends in Immunology* 30: 415–420.
- Van de Water J, Ishibashi H, Coppel RL, et al. (2001) Molecular mimicry and primary biliary cirrhosis: Premises not promises. *Hepatology* 33: 771–775.
- Walker JG, Daniach D, Roitt IM, and Sherlock S (1965) Serological tests in diagnosis of primary biliary cirrhosis. *Lancet* 39: 827–831.
- Yeaman SJ, Kirby JA, and Jones DE (2000) Autoreactive responses to pyruvate dehydrogenase complex in the pathogenesis of primary biliary cirrhosis. *Immunological Reviews* 174: 238–249.

Mitochondrial Calcium Transport: Historical Aspects

E Carafoli, University of Padua, Padua, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Ca²⁺ micropools Areas in the cytoplasm created by the opening of channels in the endoplasmic reticulum, in which the concentration of Ca²⁺ may transiently exceed 20 μ M. The channel in the endoplasmic reticulum is gated by 1,4,5-inositol-*tris*-phosphate (InsP3), produced as a consequence of the interaction of specific agonists with receptors in the plasma membrane.

Ca²⁺/O ratio The stoichiometry between Ca²⁺ accumulated and oxygen consumed during the uptake of Ca²⁺, driven by the activity of the respiratory chain.

Ca-phosphate deposits in the mitochondrial

matrix Precipitates in the matrix of mitochondria that have accumulated Ca²⁺ and phosphate. Composed mostly of amorphous hydroxyapatite and also organic components, for example, adenosine triphosphate (ATP) and adenosine diphosphate (ADP).

Chemical theory of mitochondrial energy conservation

A theory claiming that energy from the activity of the respiratory chain is conserved in the form of a hypothetical

high-energy intermediate formed on one of the components of the respiratory chain.

Chemiosmotic theory of energy conservation A theory claiming that energy in mitochondria is conserved in the form of a separation of charges across the inner membrane.

Electrophoretic uniporter A still unidentified proteinaceous system that mediates the charge-uncompensated transfer of Ca²⁺ across the inner mitochondrial membrane. Now characterized as a Ca²⁺-selective channel.

Limited accumulation of Ca²⁺ Accumulation of limited amounts of Ca²⁺ in the absence of simultaneous accumulation of inorganic phosphate.

Massive accumulation of Ca²⁺ Accumulation of large amounts of Ca²⁺ and inorganic phosphate as insoluble deposits in the mitochondrial matrix.

Na⁺/Ca²⁺ exchanger A system that promotes the release of mitochondrial Ca²⁺ in exchange for cytosolic Na⁺. Now identified with NCLX, a member of the Na⁺/Ca²⁺ exchanger protein superfamily.

The story of mitochondrial Ca²⁺ transport, now about 50 years old, has had ups and downs: the most important properties of the process were established in a feverish period of activity that followed its discovery. The uptake process was recognized to be mediated by an electrophoretic uniporter, and to be coupled to a release route that exchanges Na⁺ for Ca²⁺ with a probable 3:1 stoichiometry (a H⁺/Ca²⁺ exchanger was also found to release Ca²⁺ in some mitochondrial types). It was soon recognized that the Ca²⁺ cycle regulates two tricarboxylic (TCA) cycle dehydrogenases and pyruvate dehydrogenase-phosphate phosphatase in the mitochondrial matrix, thus controlling the delivery of reducing equivalents to the respiratory chain and, ultimately, the synthesis of adenosine triphosphate (ATP). The uptake of Ca²⁺ may also be accompanied by that of inorganic phosphate, resulting in the precipitation of insoluble Ca²⁺-phosphate deposits (amorphous hydroxyapatite) in the matrix. The precipitation of Ca²⁺ and phosphate was proposed to be a mechanism to enable cells to survive (brief) periods of cytosolic Ca²⁺ overload. It was also suggested to have a potential role in the physiological calcification process. Disappointingly, the affinity of the uptake system for Ca²⁺ was soon found to be at least one order of magnitude lower than what would have been necessary to efficiently handle Ca²⁺ in the concentration range of the bulk cytosol. This dampened the initial excitement very significantly, and initiated a protracted period of lower interest in the process. The key problem was that of reconciling the poor affinity of the uptake system, which would negate its involvement in the regulation of cell Ca²⁺, with evidence showing that the system was nevertheless active *in situ*. The paradox was eventually

solved by the demonstration that the concentration of Ca²⁺ in the bulk cytosol did not reflect the concentration in the vicinity of mitochondria, where it could transiently reach values even higher than 20 μ M, thanks to the agonist-mediated release of large amounts of Ca²⁺ from vicinal endoplasmic reticulum. The finding thus showed that mitochondria could, indeed, efficiently regulate cytosolic Ca²⁺ in spite of their low affinity for it. It initiated a robust revival of the process, which has now become again a hot topic. As a result of the renewed activity in the area, results have also appeared on the (molecular) characterization of the components of the transport process: the electrophoretic uniporter, which has now been characterized as a Ca²⁺-selective channel, and the Na⁺/Ca²⁺ exchanger, which has been identified as NCLX, a member of the Na⁺/Ca²⁺ exchanger protein superfamily.

The interplay of mitochondria with Ca²⁺ is of central importance to cell biology. Following its discovery 50 years ago, the standing of the process in the panorama of cell science has had an oscillatory pattern: immediately after its discovery, it enjoyed a period of great interest and popularity. For reasons that will be discussed, popularity and interest did not last long, and were followed by a protracted phase of essential neglect. Then, in the 1990s, the process suddenly initiated a vigorous revival that is still in full swing today. This article will offer a detailed coverage of the origins and coming of age of the topic, which, as is evident from the literature, are poorly known. It will only briefly discuss the more recent developments, as they have been handled in a number of comprehensive reviews. Historically, the critical year was 1961, when it was discovered that isolated mitochondria took up large amounts

of Ca^{2+} from the medium in a reaction that depended on respiratory energy. This landmark finding was the first direct demonstration of the ability of mitochondria to transfer Ca^{2+} to their internal space; however, numerous indirect hints had appeared in the preceding decade, suggesting that the exposure of mitochondria to Ca^{2+} resulted in its penetration into the organelle. Most of this indirect evidence came from experiments on mitochondrial swelling, which was a very popular experimental topic in the 1950s. For instance, it had been found that Ca^{2+} caused a large-amplitude swelling of isolated mitochondria, a finding that was not related – it must be remembered that those were the prechemiosmotic days – to osmotic causes, but to the stimulation of the synthesis of an endogenous uncoupling agent, or to the activation of hypothetical mitochondrial contractile elements. More direct evidence for the penetration of Ca^{2+} into mitochondria was actually provided by a 1953 study that had found that most of the Ca^{2+} of heart homogenates was recovered in the mitochondrial fraction. The study had directly corroborated the observation with the finding that mitochondria incubated with Ca^{2+} , indeed accumulated it (Table 1). Unfortunately, however, the authors (EC Slater and KW Cleland) had discounted the result as an artifact, because the work had been performed at 0° , that is, under conditions that the authors presumed would not support metabolic activities by mitochondria. Two years later, another study had come closer to demonstrating that respiring mitochondria could indeed take up Ca^{2+} . The study had found that Ca^{2+} , well known at that time as an uncoupler of oxidative phosphorylation, behaved peculiarly, that is, it activated the resting respiration of isolated mitochondria reversibly, instead of irreversibly as orthodox uncouplers do (Figure 1). Thus,

Ca^{2+} behaved like adenosine diphosphate (ADP), and the study even determined that the amount of Ca^{2+} necessary to elicit the same amount of extra oxygen consumption by mitochondria was two-third times that of ADP. The concept of a Ca:O stoichiometry (later due to become an important parameter of the coupling efficiency of mitochondria) was, thus, established *ante litteram*. Clearly, the author of the article, B Chance, had found that Ca^{2+} had somehow become expended as it activated respiration, which was another way of saying that it had become bound to mitochondria. Remarkably, this study had appeared 6 years before the evidence for Ca^{2+} uptake by mitochondria would become direct and unambiguous. Something else, which is worth mentioning, had also happened before the historical 1961 date. NE Saris, a postdoctoral fellow in the laboratory of Chance in the summer of

Table 1 Ca^{2+} uptake by heart mitochondria

<i>Ca²⁺ content of mitochondria (μmoles per mg protein)</i>		
<i>Uptake medium</i>	<i>EDTA-washed preparation (%Δ)</i>	<i>Saline preparation (%)</i>
no Ca^{2+}	0.0015	0.083
0.1 mM Ca^{2+}	0.068+~450	0.127~50

Heart mitochondria were prepared either in saline, or in saline plus 10-mM EDTA. The reaction medium contained 0.135-M KCl, 20-mM phosphate, 0.6-mM Ca^{2+} . Incubation was at 0° . After 30 min mitochondria were sedimented by centrifugation, and Ca^{2+} was determined in the supernatant. Adapted from Slater EC and Cleland KW (1953) The effect of calcium on the respiration and phosphorylative activities of heart muscle sarcosomes. *Biochemical Journal* 55: 566–580.

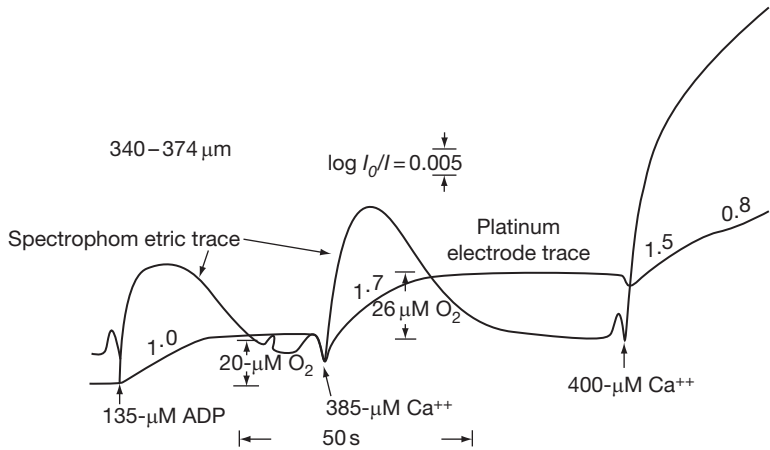


Figure 1 Simultaneous recordings of respiration (black trace) and NADH (diphosphopyridine nucleotide, called DPNH in the 1950s) oxidation (gray trace) in isolated guinea pig liver mitochondria. Oxygen (O_2) consumption was recorded as an upward of the platinum microelectrode trace: the medium (K^+ -free, isotonic, temperature not specified) contained a respiratory substrate (glutamate), but the respiration was maintained at a low level (state 4 rate) owing to the absence of the phosphate acceptor ADP. The legend does not state it explicitly but 4.5-mM phosphate was also present, judging from the legends to other figures in the article. Addition of ADP caused a burst of O_2 consumption (state 3 rate) that ceased when all ADP had been transformed into ATP. The addition of a pulse of Ca^{2+} induced a similar reversible increase of the respiratory rate. A second addition of Ca^{2+} -activated respiration indefinitely. NADH (DPNH) oxidation (not discussed in the text) was recorded as an upward deflection of the spectrophotometric trace, which reflects the decrease of optical density of the mitochondrial suspension at 340 mμ. The oxidation of NADH (DPNH), which accepted the reducing hydrogen from the substrate at the proximal end of the respiratory chain, was an alternative way of expressing the changes in respiratory rate induced by ADP and by Ca^{2+} . Reproduced from Chance B (1956) On possible mechanism for the control of electron transport in the respiratory chain. In: Liebecq C (ed.) *Proceedings of the 3rd International Congress of Biochemistry, Brussels 1955*, pp. 300–304. New York, NY: Academic Press.

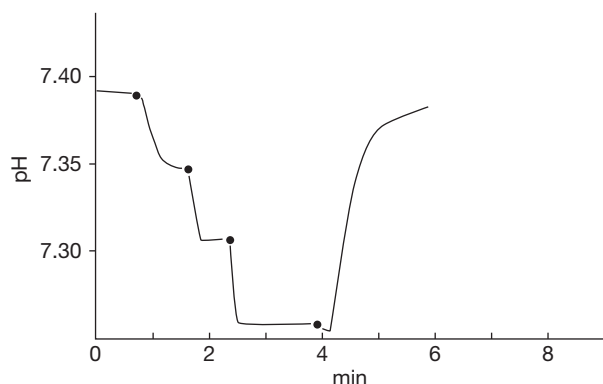


Figure 2 pH changes induced by the addition of Ca^{2+} to a mitochondrial suspension. Rat liver mitochondria were prepared by a conventional cell fractionation procedure. The medium contained 50-mM sucrose, 40-mM KCl, 4.5-mM succinate, 1.9 mg of mitochondrial protein per ml. The starting pH was 7.4, and the temperature 20.23 °C. The additions of Ca^{2+} were made at the points of deflection of the trace. The final concentration of Ca^{2+} added was 91 mM. Adapted from Saris NE (1963) The calcium pump of mitochondria *Societas Scientiarum Fennica. Commentationes Physico-Mathematicae* 28: 3–59.

1959, had discovered that the addition of Ca^{2+} to liver mitochondria induced a pH drop in the medium. He had explained the observation on the basis of the formation of a $\text{H}^+/\text{Ca}^{2+}$ gradient resulting from the transport of the latter ion into mitochondria. Those were important observations and explanations (Figure 2), because they were compatible with the mechanism for the uptake of Ca^{2+} that would become recognized much later with the triumph of the chemiosmotic concept. Saris published his first observations in 1959 in an obscure Finnish journal (where they went predictably unnoticed), and continued his experiments demonstrating that the Ca^{2+} -induced pH cycle was actually linked to the uptake of Ca^{2+} by isolated mitochondria. He summarized the results in his PhD Thesis, which, however, only appeared in 1963.

Discovery of the Ca^{2+} -Uptake Process

Nothing significant then happened in this area until 1961, when two groups, working at Johns Hopkins and at the University of Wisconsin, at last succeeded in unambiguously demonstrating that isolated kidney mitochondria took up large amounts of Ca^{2+} in a process that depended on the activity of the respiratory chain (Figure 3); the uptake was inhibited by uncouplers of oxidative phosphorylation, showing that the respiration had to be coupled. Actually, only the study of the Johns Hopkins group showed the orthodox response of the process to all inhibitors, fully justifying the conclusion that the uptake of Ca^{2+} was a means to harvest the energy made available by the respiratory chain. The study of the University of Wisconsin group had some inconsistencies, for example, it found that the uptake process was inhibited by some uncouplers but not by all, and by antimycin A but not by cyanide. The conclusion of the Wisconsin group was, thus, that “the process of Ca^{2+} uptake was not directly dependent on

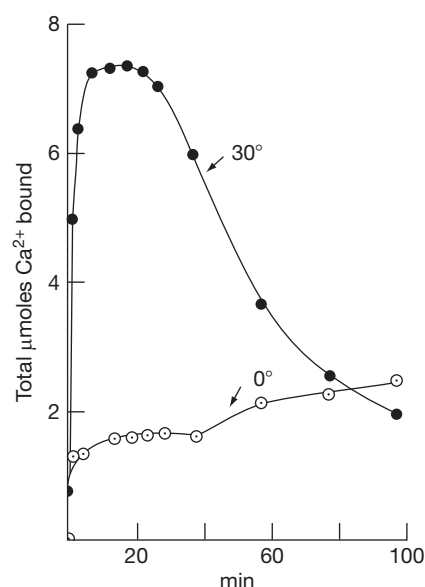


Figure 3 Uptake of Ca^{2+} by isolated rat kidney mitochondria. Mitochondria were isolated by a conventional cell fractionation procedure. The reaction medium (isotonic NaCl, pH 7.0 at 30 °C), contained a respiratory substrate (succinate), ATP, Mg^{2+} , and 2.5-mM radioactive Ca^{2+} . At the end of the incubation, mitochondria were sedimented by centrifugation, and radioactivity was measured in the supernatant in a Geiger counter. Reproduced from Vasington FD and Murphy JV (1962) Active binding of Ca^{2+} by mitochondria. *Journal of Biological Chemistry* 237: 2670–2677.

oxidative phosphorylation, nor on the activity of the entire respiratory chain, but only on the function of its mid-portion.” Remarkably, however, both studies had found that the uptake of Ca^{2+} , even if energized by the respiratory chain, also required the presence of ATP (Figure 4). As the uptake was not inhibited by oligomycin, ATP was evidently not used as a source of energy. The reason for the necessity of ATP was only rationalized a couple of years later, when it was shown by E. Carafoli that ATP was taken up by mitochondria alongside with Ca^{2+} . ATP, however, could also be used as a direct energy source for the uptake process, as it was soon found that it could support the uptake of Ca^{2+} in the absence of respiratory chain activity (i.e., in the presence of respiratory inhibitors), in which case the process was inhibited by oligomycin.

The two first studies, and other that followed immediately afterwards, had established that the amounts of Ca^{2+} that could be accumulated were hundreds of times greater than the original mitochondrial content. However, these large amounts could only be accumulated in the presence of inorganic phosphate. In its absence, the accumulated amounts were much smaller: two uptake conditions were, thus, soon defined, the limited and the massive uptake. The requirement for phosphate was understood when it was shown that phosphate accompanied Ca^{2+} on its way to the mitochondrial matrix, to precipitate it there as an insoluble salt. Soon, the Ca-phosphate deposits were visualized in the electron microscope within the mitochondrial profiles as electron dense granules (Figure 5). They were even isolated and shown to be mostly composed of hydroxyapatite, the principal mineral

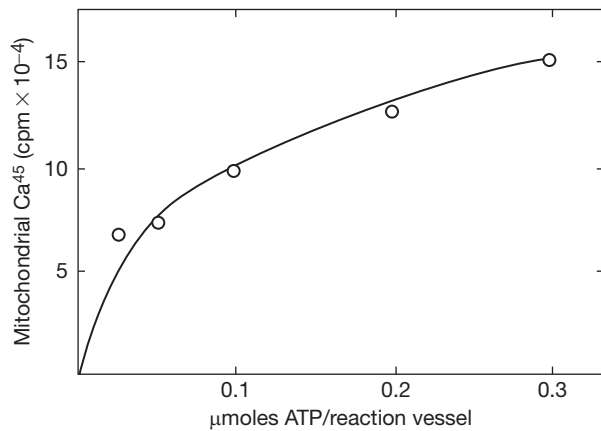


Figure 4 Dependence on ATP for Ca^{2+} uptake by kidney mitochondria. Mitochondria were isolated from the kidney of a vitamin D-deficient rat by a conventional cell fractionation procedure and incubated for 5 min at 30 °C in a medium (isotonic sucrose-KCl) containing a respiratory substrate, Mg^{2+} , various amounts of ATP, and about 300-mM radioactive Ca^{2+} . At the end of the incubation, mitochondria were sedimented by centrifugation, and radioactivity measured in the supernatant in a Geiger counter. Reproduced from De Luca HF and Engstrom GW (1961) Calcium uptake by rat kidney mitochondria. *Proceedings of the National Academy of Sciences of the United States of America* 47: 1744–1750.

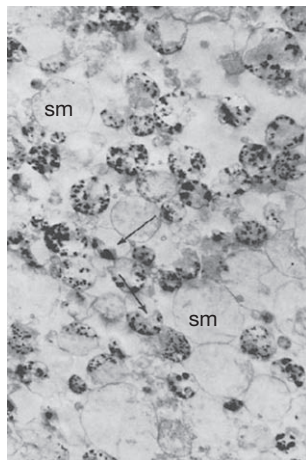


Figure 5 Electron-dense calcium-phosphate precipitates in isolated rat liver mitochondria after accumulation of Ca^{2+} and phosphate. Mitochondria were isolated by a conventional cell fractionation procedure. The composition of the medium for the uptake reaction is not specified in the legend but, according to the reference quoted, it contained isotonic NaCl at pH 7.0, a respiratory substrate (succinate), 4.0-mM phosphate, 3.0-mM ATP, and 4.0-mM Ca^{2+} . The incubation lasted 20 min at 30°. Mitochondria were then sedimented by centrifugation, fixed in OsO_4 , and observed in an electron microscope, $\times 8,000$. Reproduced from Greenawalt JW, Rossi CS, and Lehninger AL (1964) Effect of active accumulation of calcium and phosphate ions on the structure of rat liver mitochondria. *Journal of Cell Biology* 23: 21–38.

component of bones. They also contained organic components, among them ATP and ADP. The presence of ATP and ADP in the granules explained the requirement of ATP for the Ca^{2+} -uptake process even when it was energized by the respiratory chain, no granule formation occurred, and no granules

could be isolated, in the absence of ATP. As mentioned, ATP (or ADP) was accumulated together with Ca^{2+} and phosphate; the uptake of adenine nucleotides only occurred when phosphate was also taken up, and when the amounts of Ca^{2+} and phosphate accumulated were massive. In all probability, adenine nucleotides stabilized the Ca-phosphate deposits, or perhaps primed their precipitation. The uptake of ATP was inhibited by the classical blocker of the ATP/ADP carrier, but the inhibition was only partial when the amounts of accumulated Ca^{2+} and phosphate were massive, a finding that was attributed to the damage to the mitochondrial membrane by the excess accumulation of Ca^{2+} and phosphate. The precipitation of the insoluble deposits initiated on the cristae, and then, as the precipitates grew in size, the deposits moved to freer spaces in the matrix. The simultaneous accumulation of phosphate was exploited as a convenient laboratory device to increase the amount of Ca^{2+} than could be stored inside mitochondria, and also gradually became recognized as potentially important in physiology. Early after the discovery of the Ca^{2+} - and phosphate-uptake process, interest developed in the possibility that the storage of insoluble Ca-phosphate salts within mitochondria could have a role in the process of physiological calcification. As the biological potential to undergo calcification is ubiquitously present in animal tissues, it was appealing to relate the transport of Ca^{2+} by mitochondria, and, in particular, the deposition of Ca^{2+} -phosphate granules, to the process. A.L. Lehninger, thus, proposed that mitochondria could concentrate Ca^{2+} and phosphate to a level that could not be attained spontaneously in the extracellular fluids. Micropackets of colloidal dimensions would then dissociate from the large Ca^{2+} -phosphate granules and would somehow pass through the mitochondrial membranes to be exported from the cell and to the calcification sites; the finding that the hydroxyapatite formed inside mitochondria remains indefinitely amorphous would facilitate the process. Clearly, the transfer of the Ca^{2+} deposits through the mitochondrial membrane and their diffusion to the calcification sites is a major problem. Some indirect evidence in favor of the proposal has appeared over the years, but convincing experimental validation has yet to appear.

The storage of massive amounts of Ca^{2+} and phosphate inside mitochondria was also proposed to be a means to clear excess Ca^{2+} from the cytosol whenever pathological events would permit its abnormal entry from the extracellular spaces. Mitochondria, thus, have come to be regarded as devices that would permit cells to survive a transient period of cytosolic Ca^{2+} overload. The concept was reinforced by the finding that the hydroxyapatite deposits in the matrix remained indefinitely amorphous, that is, in a state that would facilitate their dissolution, and the gradual release of Ca^{2+} from mitochondria (once the Ca^{2+} emergency had passed) at a rate compatible with the exporting capacity of the plasma-membrane systems.

Early Findings on the Mechanism of the Uptake Process

In the wake of the early findings, mitochondrial Ca^{2+} transport rapidly became very popular, and became recognized as an alternative to ADP phosphorylation in the harvesting of the energy made available by the respiratory chain. For a long time,

the findings on the process were interpreted within the framework of the chemical theory of energy transduction; in hindsight, it may seem surprising that the process was not immediately related to the then emerging chemiosmotic theory of energy transduction, for which it evidently provided essential support. Chemiosmosis, however, only crept into the field laboriously, and did not take it over until well into the 1960s. Meanwhile, numerous studies had established the essential phenomenology of the process; thus, the stoichiometry of Ca^{2+} uptake to oxygen consumption was measured, finding that two Ca^{2+} ions were required to elicit the same extra consumption of oxygen elicited by one molecule of ADP, a finding that clearly confirmed the earlier conclusion by B Chance that the Ca^{2+}/O stoichiometry in the transition from resting (state 4) to activated (state 3) respiration was 2–3 times the ADP/O stoichiometry. It also became clear that in the absence of phosphate the small amounts of Ca^{2+} that were taken up were maintained within mitochondria in a dynamic steady state, in which the leak of Ca^{2+} was balanced by its reuptake during the phase of resting respiration that followed the activated phase during which the pulse of Ca^{2+} had been taken up. It was an important finding, as it established the concept of a continuous cycling of Ca^{2+} across the mitochondrial membrane (the mitochondrial Ca^{2+} cycle), its long-term storage in the matrix only occurring when phosphate was also present and taken up. The finding that mitochondria could dynamically take up and release Ca^{2+} had obvious interest in the light of the emerging importance of Ca^{2+} in the regulation of cell activities, and placed mitochondria in a central position as potential regulators of cell function. At that time, however, an element in the scenario was still missing: while the fourth portion of the cycling process was defined, the back portion, that is, the route (mechanism) for the release of Ca^{2+} was not, and was only discovered some years later.

The limited uptake of Ca^{2+} in the absence of phosphate was eventually used to test the proposal of an electrophoretic mechanism for the uptake of Ca^{2+} , which would have been consistent with the increasingly popular chemiosmotic principles. Diffusion potentials were artificially imposed across the mitochondrial membrane, showing that Ca^{2+} indeed traversed the membrane in response to an electrochemical gradient. Uncertainties of the various types complicated experiments aimed at establishing whether the mechanism of the uptake process was purely electrophoretic or partially charge compensated, but the presence of an electrophoretic component in the uptake process became generally accepted. As a result, the process of Ca^{2+} transport, which had been initially described as an active uptake followed by a passive release, became an energetically downhill uptake process, followed by an energetically uphill release leg. The uptake leg of the process was postulated to occur on an electrophoretic uniporter, its study being aided greatly by the discovery of specific inhibitors, the most popular being the histochemical stain Ruthenium Red (about 20 years later a derivative, RU360, was introduced, which specifically inhibited the uptake of Ca^{2+} with an IC_{50} of 0.2–2.0 nM).

The Release Leg of the Mitochondrial Ca^{2+} -Transport Cycle

The release leg of the Ca^{2+} -transport process was eventually identified in 1974, when E Carafoli and co-workers found that Na^+ could specifically release Ca^{2+} from mitochondria (Figure 6). The process was later characterized by the group as a carrier-mediated antiport that exchanged Na^+ for Ca^{2+} . The curve relating the rate of Na^+ -induced Ca^{2+} release to the concentration of extramitochondrial Na^+ had a pronounced

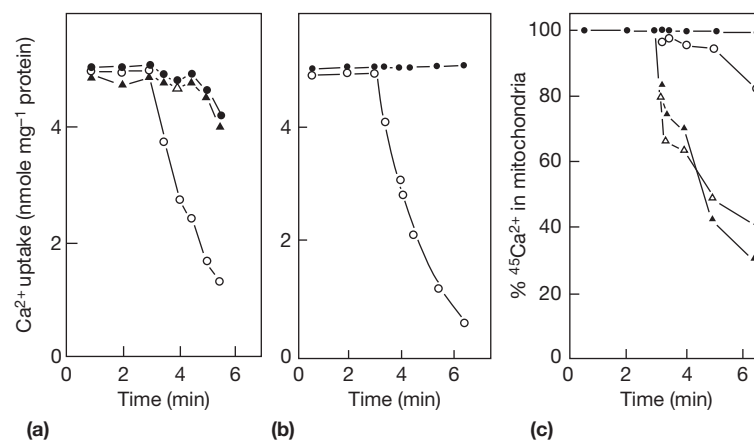


Figure 6 Na^+ -induced Ca^{2+} release from rat heart mitochondria. Mitochondria were isolated by a conventional cell-fractionation procedure. The release of Ca^{2+} was followed isotopically by rapid Millipore filtration after mitochondria had been loaded with Ca^{2+} , labeled with $^{45}\text{Ca}^{2+}$. The medium (pH 7.4, 25°) contained 210-mM mannitol, 70-mM sucrose, and a respiratory substrate (succinate). No substrate was present in (a), in which mitochondria had also been preincubated with rotenone to block the reuptake of the lost Ca^{2+} driven by endogenous respiratory chain activity. In (b), the reuptake of the lost Ca^{2+} had instead been blocked by 10-nM ruthenium red, added 30 s before the addition of Na^+ . The amounts of Ca^{2+} added to the medium were extremely small (between 5 and 10 nmol per mg of mitochondrial protein). In (c), no Ca^{2+} was added, as the rats had been injected intraperitoneally with 180 μCi of $^{45}\text{Ca}^{2+}$ 5 min before being killed. (a) Open circles, 100-mM Na^+ ; triangles, 100-mM K^+ ; closed circles, equal concentration of reaction medium. (b) Open circles, 50-mM Na^+ ; closed circles, equal concentration of reaction medium. (c) Open circles, 5-mM Na^+ , closed triangles, 20-mM Na^+ , open triangles, 50-mM Na^+ ; closed circles, 100-mM reaction medium. Adapted from Carafoli E, Tiozzo R, Lugli G, Crovetto F, and Kratzing C (1974) The release of calcium from heart mitochondria by sodium. *Journal of Molecular and Cellular Cardiology* 6: 361–371.

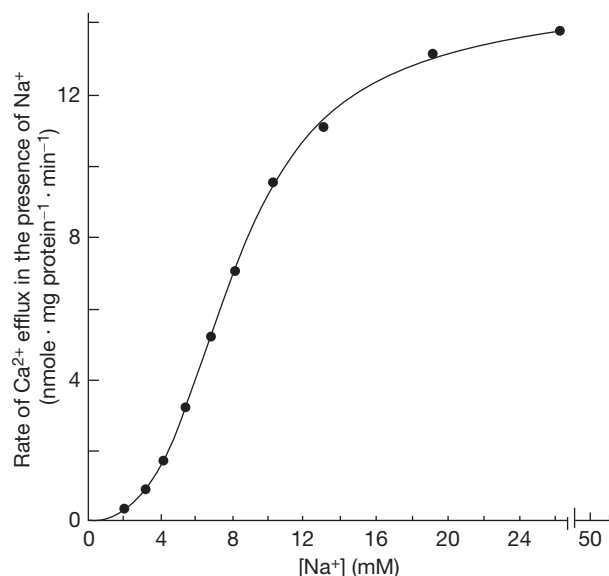


Figure 7 Dependence of the initial rate of Ca^{2+} release from mitochondria on the concentration of external Na^+ . Rat heart mitochondria were prepared by a conventional cell fractionation procedure and incubated in a medium containing 120-mM KCl, 10-mM HEPES pH 7.0, 1-mM succinate, 60 nmol of Ca^{2+} , and 2.9 mg of mitochondrial protein in a final volume of 4.0 ml at 25 °C. After 4 min, 0.15 μM ruthenium red was added to prevent the reuptake of the released Ca^{2+} . Then, Na^+ was added to induce the efflux of Ca^{2+} and the rate of Ca^{2+} release, which was constant for at least 1 min, was measured with a Ca^{2+} -specific electrode. Modified from Crompton M, Capano M, and Carafoli E (1976) The sodium-induced efflux of calcium from heart mitochondria. A possible mechanism for the regulation of mitochondrial calcium. *European Journal of Biochemistry* 69: 453–462.

sigmoidal character, with an EC_{50} of about 8 mM (Figure 7); the figure was interesting, as it was within the range of Na^+ concentration in the normal cytoplasm, that is, the rate of Ca^{2+} release from mitochondria changed significantly by relatively small changes of Na^+ in the 4–10-mM range, implying that the efflux of Ca^{2+} from mitochondria could be regulated by physiological changes of the cytosolic Na^+ concentration. The maximal velocity of the Na^+ -mediated release process (about 18 nmol per mg of mitochondrial protein per minute) was about 10 times lower than the V_{max} of the uptake process. The exchanger was found to be particularly active in mitochondria of excitable tissues, for example, heart (a $\text{H}^+/\text{Ca}^{2+}$ -exchange process was later found to preferentially mediate the release of Ca^{2+} from mitochondria that were less Na^+ responsive, e.g., those of liver). The original work on the Na^+ -mediated Ca^{2+} release found that it was inhibited by lanthanides, which also inhibited the Ca^{2+} -uptake process, albeit with a different response to the lanthanide series. Later on, the release process was found to be inhibited much more effectively by benzothiazepines. The most effective was CGP-37157, which has become the inhibitor of choice in the studies of the exchanger.

The $\text{Na}^+/\text{Ca}^{2+}$ exchanger operated in the presence of the large negative membrane potential maintained inside mitochondria by the vectorial operation of the respiratory chain; therefore, it had to operate either electroneutrally or

electrogenically, exchanging more than two Na^+ per one Ca^{2+} . The stoichiometry of the exchange reaction has been debated, but now appears to be settled on three Na^+ per one Ca^{2+} .

The Affinity of Mitochondria for Ca^{2+} Is Too Low for the Regulation of Cell Ca^{2+}

The question of the affinity of the mitochondrial uptake system for Ca^{2+} was obviously important, if mitochondria were to play any role in the regulation of cytosolic Ca^{2+} , they had to be able to interact with it in the concentrations prevailing in the cytosol, that is, in the nM range. At this point, problems soon emerged: numerous measurements with different methods, and on mitochondria from different tissues, invariably yielded K_m values for the uptake process higher than 1 μM , most frequently in the 10- μM range. The problem was further exacerbated by Mg^{2+} , which decreased the Ca^{2+} affinity of the uptake system in the concentrations known to exist in the cytosol. Factors were identified that increased the affinity, for example, spermine, which made mitochondria responsive to physiological concentrations of Ca^{2+} , but the idea gradually gained ground that mitochondria were essentially inactive in the Ca^{2+} concentrations of the cytosol. The rationale for the existence of the sophisticated Ca^{2+} uptake and release system, thus, shifted from the regulation of Ca^{2+} in the cytosol to the regulation of Ca^{2+} within mitochondria themselves, which was an important task, because RM Denton, JG McCormack and co-workers had discovered that two TCA cycle dehydrogenases (α -ketoglutaric dehydrogenase and NAD^+ -isocitric dehydrogenase) and the phosphatase that dephosphorylates pyruvic dehydrogenase were allosterically regulated by Ca^{2+} in the μM range, the production of reducing equivalents to be delivered to the respiratory chain and, ultimately, the synthesis of ATP, thus depended on the precise regulation of Ca^{2+} in the mitochondrial matrix. Therefore, in the physiological conditions of most cytosols, the mitochondrial Ca^{2+} -transporting system would only operate at a minor fraction of its maximal potential, which, however, would still be adequate to translate Ca^{2+} signals inside mitochondria for the purpose of modulating the production of energy.

Considering that in the 1960s and 1970s, it was generally assumed that the concentration of Ca^{2+} in the cytosol oscillated uniformly in the 100–200-nM range, the skepticism on the role of mitochondria in the regulation of cytosolic Ca^{2+} seemed justified, not entirely; however, as a number of findings had shown that mitochondria, regardless of their documented poor affinity for Ca^{2+} , were somehow still able to efficiently take it up Ca^{2+} *in situ*. The original 1953 work by Slater and Cleland had already shown that mitochondria sequestered most of the Ca^{2+} of heart homogenates, and similar work had shown that this was true also of Mn^{2+} , a cation that is transported by mitochondria with the same rules as Ca^{2+} . However, more recent work on the distribution of injected radioactive Ca^{2+} in the tissues of rats had shown that the largest fraction of the radioactivity was recovered in the mitochondrial fraction. Importantly, the fraction of radioactivity recovered in mitochondria decreased very significantly if rats had been previously injected with an uncoupler; evidently, efficient energy-linked uptake of Ca^{2+} by mitochondria had

managed to occur *in situ*, somehow overcoming the poor affinity of the system.

However, the mediocre Ca^{2+} affinity of the uptake system was difficult to circumvent. Actually, the finding that the amount of Ca^{2+} which could be presumably mobilized from mitochondria *in situ* in response to physiological demands was insignificantly low (1–2 nm per mg of mitochondrial protein) added significantly to the general aura of skepticism; by the late mid-1970s, mitochondrial Ca^{2+} transport had lost a large portion of the appeal it had once enjoyed.

New Findings Overcome the Problem of the Low Affinity of the Uptake System

Important findings continued to be made during the period of dampened interest on mitochondrial Ca^{2+} transport, for instance, showing that the uniporter had two Ca^{2+} -conductivity states, and identifying a new path for the release of Ca^{2+} from mitochondria, the permeability transition pore (PTP). Findings also occasionally appeared showing that mitochondria efficiently buffered physiological Ca^{2+} loads *in situ*, for example, in neurons. The large and sophisticated body of information that had become available since the 1960s (summarized in the cartoon in Figure 8) was impressive, but the poor Ca^{2+} affinity of the system remained a formidable barrier that confronted all serious attempts to involve mitochondria in the regulation of cell Ca^{2+} . The barrier was finally removed at the beginning of the 1990s, when S Rizzuto and his co-workers showed that the low nanomolar concentrations of Ca^{2+} routinely measured in the bulk cytosol did not

necessarily reflect those in the vicinity of mitochondria. The conclusion was made possible by the targeting of the Ca^{2+} -responsive photoprotein, aequorin, to the matrix of mitochondria, where it directly monitored the changes of the concentration of Ca^{2+} ; the original experiment, for which bovine endothelial cells were used, showed that plasma-membrane agonists that transiently elevated cytosolic Ca^{2+} surprisingly elicited the rapid and transient increase of mitochondrial Ca^{2+} (Figure 9). The experiment took advantage of the ability of the newly discovered second messenger 1,4,5-inositol-*tris*-phosphate (InsP3) to open a specific Ca^{2+} -release channel in the endoplasmic reticulum. It was a direct demonstration that the mitochondrial Ca^{2+} -transport system responded rapidly to the increase in cytosolic Ca^{2+} induced by physiological stimuli, which in this case were plasma-membrane agonists that generated InsP3. Importantly, no mitochondrial Ca^{2+} response was elicited by adding Ca^{2+} to permeabilized cells to artificially modulate its concentration throughout the cytosol; evidently, the opening of the InsP3-gated channels in the endoplasmic reticulum had generated micropools of Ca^{2+} in the vicinity of mitochondria, in which the concentration of the ion was much higher than in the cytosol at large, and evidently adequate to activate the low-affinity uniporter. Later measurement actually showed that the concentration of Ca^{2+} in these hot spots could reach values in excess of 20 μM . The findings were reinforced by the demonstration of the intimate physical interplay between mitochondria and the endoplasmic reticulum in a number of cell types. They were followed by numerous studies correlating the time response of mitochondrial Ca^{2+} with that of ATP production, which depended on the activation of the Ca^{2+} -responsive matrix

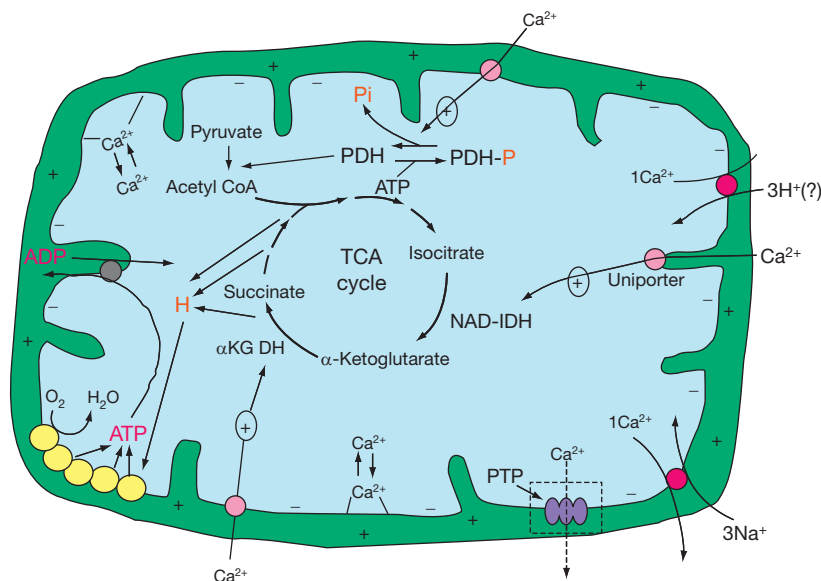


Figure 8 The systems for the control of Ca^{2+} homeostasis in mitochondria. The cartoon shows the three matrix enzymes that are controlled by Ca^{2+} : α -ketoglutaric dehydrogenase (α -KGDH), NAD-dependent isocitric dehydrogenase (NAD-IDH), and pyruvate dehydrogenase phosphate phosphatase, that acts on the substrate PDH-P. The (–) and (+) symbols indicate the negative inside membrane potential maintained by the asymmetric orientation of the respiratory chain carriers in the inner membrane. The (?) symbol next to the $\text{H}^+/\text{Ca}^{2+}$ exchanger indicates that the stoichiometry of the system has not yet been conclusively established. The permeability transition pore (PTP) is engaged in a dashed box to signify that, even if it lets Ca^{2+} through, it is not a *bona fide* Ca^{2+} channel. The uniporter is indicated as a Ca^{2+} -selective channel, as shown by recent findings (see text). The cartoon also shows the reversible (low affinity) binding of Ca^{2+} to negatively charged membrane components (phospholipids?)

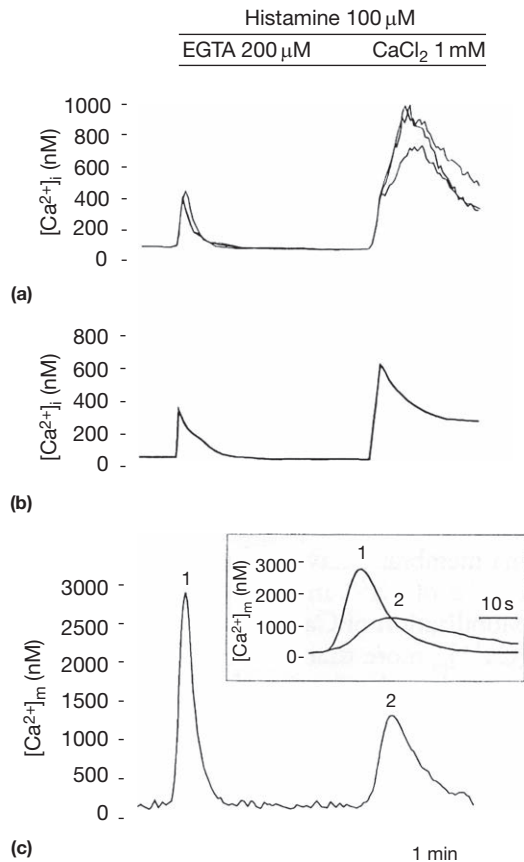


Figure 9 Changes of cytoplasmic Ca^{2+} (Ca^{2+}_c) and mitochondrial Ca^{2+} (Ca^{2+}_m) in a HeLa cell clone that stably expressed mitochondrial aequorin or cytoplasmic aequorin. The cells were trypsinized, plated on glass cover slips, left in culture for 2 days, and loaded with fura 2 and coelenterazine (to reconstitute aequorin). Full experimental conditions, including observation of cells, are described in Rizzuto R, Simpson AWM, Brini M, and Pozzan T (1992) Rapid changes of mitochondrial Ca^{2+} revealed by specifically targeted recombinant aequorin. *Nature* 358: 325–327. (a) Measurement of Ca^{2+}_i in three representative single cells. (b) Measurement of Ca^{2+}_i in a monolayer of cells. (c) Measurement of Ca^{2+}_i in a monolayer of cells. In the inset, the (Ca^{2+}_m) peaks (1, mobilization; 2, influx), are compared on an expanded timescale. Reproduced from Rizzuto R, Brini M, Murgia M, and Pozzan T (1993) Microdomains with high calcium close to IP_3 sensitive channels that are sensed by neighbouring mitochondria. *Science* 93: 744–747.

dehydrogenases, and of NAD(P)H generation. They thus provided the long-sought explanation for the paradoxical ability of the uniporter to efficiently operate *in situ* regardless of its very poor affinity for Ca^{2+} , and revived interest in mitochondrial Ca^{2+} transport.

The Molecular Nature of the Mitochondrial Ca^{2+} Transporters

Research on the molecular nature of the mitochondrial Ca^{2+} transporters had been going on since the early days. However, in spite of much work, for a long while the results were consistently meager: the uniporter and the two exchangers defeated

all attempts of molecular characterization, remaining elusive entities. The renaissance of the topic in the 1990s brought new impetus to the search, which has recently seen very substantial progress. Patch clamping work on mitochondria of COS-7 cells deprived of the outer membrane permitted D Clapham and co-workers to characterize the uniporter as a novel, highly selective Ca^{2+} channel that was sensitive to ruthenium red and to RU360, and which had the same divalent cation conductance selectivity of the Ca^{2+} uptake reaction in isolated mitochondria. More recently, the same laboratory performed a genome-wide *Drosophila* RNAi screening that led to the identification of a protein (Letm1) with no appreciable homology to bacterial and plant Ca^{2+}/H^{+} and Ca^{2+}/Na^{+} exchangers that mediated Ca^{2+} uptake coupled to H^{+} extrusion in permeabilized S2 cells. The expressed protein was purified from bacteria, and incorporated into liposomes where it supported the rapid accumulation of Ca^{2+} coupled to the efflux of H^{+} . The authors concluded that Letm 1 was the mitochondrial Ca^{2+}/H^{+} exchanger. However, some of the findings were difficult to reconcile with the conclusion, for example, the exchanger in reconstituted liposomes was blocked by RU 360, which completely blocks the uniporter but has no effect on the Ca^{2+}/H^{+} exchange.

The most recent finding on the molecular identity of the mitochondrial Ca^{2+} transporters is the identification by a composite group of researchers working in Israel, continental Europe, and the United States, of the Ca^{2+}/Na^{+} exchanger as a novel member of the family of Na^{+}/Ca^{2+} exchangers protein superfamily, NCLX. NCLX is the single mammalian member of a phylogenetically ancestral branch of the Na^{+}/Ca^{2+} exchanger superfamily. It performs Na^{+} - or Li^{+} -dependent transport of Ca^{2+} , a feature only shared by the mitochondrial exchanger, and is expressed with particular prominence in skeletal muscle, stomach (smooth muscle), and pancreas, three tissues in which the mitochondrial transport of Ca^{2+} plays a critical role. The authors convincingly corroborated this indirect evidence by showing that NCLX was enriched in mitochondria, where it located to the inner membrane criste. Its overexpression in model cell increased the Na^{+} -promoted mitochondrial Ca^{2+} efflux, whereas its silencing reduced it. The exchange activity in silenced mitochondria was recovered by the expression of heterologous NXCL, and, finally, the NCLX-mediated mitochondrial Ca^{2+} transport was inhibited by the classical inhibitor of the mitochondrial Na^{+}/Ca^{2+} exchanger CGP-37157; the long-sought mitochondrial Na^{+}/Ca^{2+} exchanger, thus, appears to have been identified. Still more recently comprehensive efforts based on a number of predictive *in silico* approaches have led to the identification of a uniporter regulatory protein, and of a two-transmembrane domain, 40-kDa channel-forming protein that fulfills the requirements for being the uptake uniporter.

Mitochondrial Ca^{2+} Transport and the Disease Process

The robust renaissance of mitochondrial Ca^{2+} transport parallels the vigorous expansion of the Ca^{2+} -signaling field. However, Ca^{2+} is an ambivalent messenger; it is essential to the correct functioning of cells as soon as its concentration and

movements are under careful control, but easily becomes a conveyor of death if the control somehow fails. Considering the all-encompassing role of Ca^{2+} in the regulation of cell functions, it is hardly surprising that dysfunctions of the ability of mitochondria to control Ca^{2+} should be harmful to cells. Disease conditions have, for instance, been described in which defects in the respiratory chain prevented mitochondria from efficiently taking up Ca^{2+} , thus depressing the activity of the Ca^{2+} -dependent matrix dehydrogenases and the synthesis of ATP. Significantly, the defects were corrected by limiting the release of Ca^{2+} through the $\text{Na}^+/\text{Ca}^{2+}$ exchange route with the inhibitor CGP 37157. The cellular dyshomeostasis of Ca^{2+} appears to be a strong causative element in a number of neurodegenerative diseases, paramount among them Huntington's disease. Neuronal mitochondria in these diseases experience a condition of stress that compromises their ability to correctly handle Ca^{2+} . Evidence is now accumulating that the abnormal opening of the PTP would promote the release of proapoptotic factors, and of Ca^{2+} itself, from mitochondria to the cytoplasm. The $\text{Na}^+/\text{Ca}^{2+}$ exchanger is also becoming particularly popular as a possible actor in the mitochondrial Ca^{2+} dysfunction in the disease process, especially in tissues such as heart and brain, where the $\text{Na}^+/\text{Ca}^{2+}$ exchange plays a particularly important role in the control of cellular Ca^{2+} homeostasis by mitochondria. Recent evidence has involved it in the

molecular pathogenesis of diseases such as heart failure and neurodegenerative conditions such as Parkinson's and Alzheimer's diseases.

See also: [Bioenergetics: Calcium Transport in Mitochondria](#).

Further Reading

- Carafoli E (2003) Mitochondria and calcium: Ups and downs of an unusual relationship (Review). *Trends in Biochemical Sciences* 28: 175–181.
- Carafoli E (2010) The fateful encounter of mitochondria with calcium: How did it happen? (review). *Biochimica et Biophysica Acta* 1797: 5595–5606.
- Kirichok Y, Krapivinsky G, and Clapham DE (2004) The mitochondrial Ca^{2+} uniporter is a highly selective ion channel. *Nature* 427: 360–364.
- Lehninger AL (1970) Mitochondria and calcium ion transport. *Biochemical Journal* 119: 129–138.
- McCormack JG, Halestrap AP, and Denton RM (1990) Role of calcium ions in the regulation of mitochondrial metabolism (review). *Physiological Reviews* 70: 391–425.
- Palty R, Silverman WF, Hershfinkel M, et al. (2010) NCLX is an essential component of mitochondrial $\text{Na}^+/\text{Ca}^{2+}$ exchange. *Proceedings of the National Academy of Sciences of the United States of America* 107: 436–441.
- Rizzuto R, Brini M, Murgia M, and Pozzan T (1993) Microdomains of high Ca^{2+} close to the IP3-sensitive Ca^{2+} channels that are sensed by neighbouring mitochondria. *Science* 262: 744–747.
- Rossi CS and Lehninger AL (1963) Stoichiometric relationships between accumulation of ions by mitochondria and the energy coupling sites in the respiratory chain. *Biochemische Zeitschrift* 338: 698–713.

Mitochondrial Channels

M C Sorgato and A Bertoli, University of Padua, Padua, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Conductance G Expressed in siemens ($S = \text{ampere/volt}$ (A/V)), conductance G is the ratio between the electric current and the applied voltage. At 100 mV (10^{-1} V), a G of 10 pS (10^{-11} S) corresponds to a current of 1 pA (10^{-12} A) generated by about six million monovalent ions per second.

Electrochemical proton gradient Generated by the H^+ -pump activity of mitochondrial inner membrane enzyme complexes, it is commonly expressed in millivolts (mV) as proton motive force ($\Delta p = \Delta\mu_{H^+} / F = \Delta\psi - 60\Delta pH$). $\Delta\psi$ defines the difference in charges, and ΔpH that in H^+ concentrations, between the matrix and the aqueous phase external to the inner membrane. A Δp of 180–200 mV (matrix negative) is normally found, with $\Delta\psi$ contributing for 150–180 mV.

Lipid bilayer It allows the recording of electric activities from artificial planar membranes fused with vesicles obtained from native membranes, or reconstituted with purified proteins.

Mitochondria Organelles composed of an outer membrane, an inter-membrane space, and an invaginated inner membrane delimiting a matrix space that houses crucial metabolic routes.

Patch clamp By means of fine electrodes, it records ionic currents directly from the membrane of isolated cells or organelles, as well as from large lipid vesicles reconstituted with channel-containing membrane fractions (proteoliposomes).

Mitochondria are organelles composed of two membranes, in either of which channels have been detected. However, while channels in the outer membrane (OM) have been rationalized by the overall high permeability of the membrane, the presence of channels in the inner membrane (IM) was unexpected in view of the known implication of the IM in the process of oxidative phosphorylation requiring tightly controlled ion fluxes. After the initial phenomenological description, and despite the largely unknown molecular identity, substantial advances in the functional identification of mitochondrial channels have now disclosed the possibility that these channels may take part in crucial mitochondrial functions, and that their dysregulation triggers dramatic cell events.

Channels in Mitochondrial Bioenergetics

The energy-requiring synthesis of ATP is indispensable for the life of the cell, and is catalyzed by the IM-spanning adenosine triphosphate (ATP)-synthase complex. Following the chemiosmotic principles, ATP-synthases utilize the electrochemical proton gradient that is built up by the active proton translocation from the inner compartment (matrix) of mitochondria outwardly, accomplished by IM enzyme complexes during substrate oxidation. As intuitively expected, mitochondria need to avoid dissipation of the proton gradient by uncontrolled fluxes of protons, and ions in general, toward the negatively charged matrix. Accordingly, it has long been established that across the IM there are tightly regulated ion movements that proceed at a much lower rate than ions through channels, and which are catalyzed by proteins such as: the ATP-synthase; the H^+ -carrier that in brown fat dissipates energy as heat; carriers that accumulate anionic substrates in the matrix, or counterbalance the influx of positive charges carried, for

example, by Ca^{2+} and K^+ . A different reasoning applies to the OM, which is devoid of energy-transducing apparatus, and cannot, therefore, develop a proton gradient. This fact excludes conceptual barriers for the OM permeability properties, and indeed the electric features of OM channels, the first to be identified in mitochondria, have led to assume that, rather than classical ion conducting elements, they act as principal pathways regulating the intense exchange of high-molecular-mass molecules between mitochondria and the rest of the cell.

Functional and Molecular Aspects of Mitochondrial Channels

Outer Membrane Channels

The voltage-dependent anion channel (VDAC, or mitochondrial porin) is the most abundant OM protein, and one of the few mitochondrial channels with known molecular identity. Its presence in all eukaryotes examined so far speaks in favor of a key role for this protein. Extensive electrophysiologic studies – carried out by incorporating purified OM fractions in artificial membranes – have revealed that VDACs from different species have amazingly conserved biophysical behaviors. Paradigmatic is the symmetrical closure upon raising the transmembrane electric field, in that the large conductive (maximally open) state adopted by VDAC at low applied voltages (roughly between ± 10 mV) (Table 1) is drastically reduced (albeit not nullified) as voltage is increased on both directions (positive and negative). Attainment of the lower conducting states, nicknamed ‘closed’, parallels the change of the ionic selectivity from slightly anionic to cationic. Such features, in particular the voltage-dependent permeability, and other biochemical and genetically based findings, have thus frequently suggested that the protein acts as a molecular

Table 1 Main mitochondrial conductances

Location	Name	Maximal conductance (pS)	Major putative role(s)
Outer membrane	VDAC	600–750	Transport of metabolites, MOMP/apoptosis, PTP complex
	Tom40	1000	Protein import
	MAC	2500	Apoptosis
Inner membrane	Miscellaneous	~300	Transport of metabolites
	K _{ATP}	~10	Volume regulation, ischemic preconditioning
	mtBK _{Ca} (K _{Ca} 1.1)	295	Ischemic preconditioning, protection against Ca ²⁺ overload
	mtIK _{Ca} (K _{Ca} 3.1)	90	?
	Kv1.3	17	Apoptosis
	MiCa/MCU	~6	Ca ²⁺ import
	mCS (IMAC)	107	Volume regulation.
	TIM (MMC/MCC)	1000	Transport of metabolites
	PTP	1500	Protein import
			Cell death

Summary of the channel activities present in the OM and IM of mitochondria, and of the putative role attributed to them. For details, see text.

sieve, thereby regulating the process of oxidative phosphorylation, if not the entire cell metabolism.

Three distinct VDAC genes (Vdac1-3) have been identified in mammals. Mice deficient for either Vdac1 or Vdac3 are viable, although showing alterations of mitochondrial functionality and morphology. Conversely, deletion of Vdac2 results in embryonic lethality, which is possibly linked to enhanced activities of mitochondria-dependent death pathways. VDACS from different organisms show little conservation of the amino acid sequence, thus suggesting that the almost overlapping electric picture of all VDACS could arise from similar secondary structures. Indeed, prediction algorithms based on sequence analysis and functional properties have proposed a general feature of the pore's wall formed by antiparallel β sheets, with the hydrophilic amino acids expectedly facing the internal aqueous ion-conducting space. The recent availability of high resolution three-dimensional (3D) structures of mammalian VDAC1 has partly confirmed these predictions, showing that the protein is a 19-stranded β barrel, with an α helix located within the aqueous pore. The α helix might serve as a voltage sensor, or as a filter for ion selectivity. Intriguingly, however, the functional properties deduced from the 3D structures did not correlate with the long-standing electrophysiologic picture of VDAC in artificial membranes. In addition, the same recombinant protein utilized for the structural studies showed substantially different electric behaviors.

Within this issue, it is important to mention that the conserved electric pattern of VDAC was never observed *in situ*. Instead of the typical bell-shaped curve, application of a patch clamp electrode to intact mitochondria has shown that the macroscopic current of the entire OM responded asymmetrically to positive and negative voltages, and, likewise, that single-channel behaviors were never reminiscent of reconstituted VDAC properties. Clearly, these data suggest that VDAC in native membranes is under the control of either OM or inter-membrane regulators that are lost during reconstitution of OM fractions in artificial membranes; or that, despite VDAC abundance, channels different from VDAC are the major contributors to the OM electric behavior. As yet, however, the

physiology and molecular structure of most of such channels are unknown, although – consistent with the present notion of OM functionality – they likely transport metabolites rather than only small ions. Finally, the only attempt to analyze mitochondria in its cellular context (at the presynaptic terminals of large neurons) has demonstrated that the OM displays intense electrical activities especially during synapse activation, which, however, do not resemble those reported for VDAC in artificial membranes.

Of relevance, the electrophysiologic analysis of OM- (or IM-) containing proteoliposomes, and of some membrane components in artificial membranes, has eventually proven the long-lasting hypothesis that most (almost 99%) mitochondrial proteins – encoded by the nuclear genome and synthesized in the cytosol – reach their final destination by passing through large permeation pathways spanning the OM and IM. Specifically, it has been found that both the purified multisubunit translocase of the OM (TOM) (which does not include VDAC) and the TOM component putatively involved in translocation, Tom40, form a pore with a large full open state of around 1000 pS, and with a slight preference for cations (Table 1). These features, which are *per se* consistent with a role of the TOM machinery in allowing permeation of unfolded polypeptides, have been further substantiated by the disturbance to ionic currents provoked by cationic peptides mimicking the mitochondrial targeting domain of several proteins. This perturbation likely reflects transient occlusions of the pore during the passage of the peptides.

Inner Membrane Channels

The *in situ* electric recording of the IM of isolated mitochondria demands that the OM be removed first; this allows expansion of the IM invaginations and formation of a vesicle (mitoplast) sufficiently large to be patch clamped. Except for one type of conductance (called mitochondrial mega- (MMC), or multi-conductance channel (MCC)) showing huge and unselective fluxes of ions, this strategy has been instrumental to the

description of two main kinds of channels, distinguishable by several criteria, not the least the ion selectivity.

Cation selective channels

Following the patch clamping of intact mitoplasts, it has been proposed that the IM harbors more than one type of K^+ channels, which have been discriminated upon the mode of their regulation, the single channel conductance, and, when available, the molecular identity. One of these is the K_{ATP} channel, with a conductance of around 10 pS and no clear voltage dependence, which was found to be active only at low concentrations of matrix ATP (Table 1). The similar action of ATP and drug molecules has then allowed to functionally identify the K_{ATP} channel with the K^+ uniport that, together with anion-conducting systems (see below), is known to guarantee mitochondrial osmotic stability. Indeed, enhancement of K^+ influx under the stressed condition of low matrix ATP renders the channel suitable at serving in the K^+ cycle-based volume regulatory mechanism, especially within the proposed control of mitochondrial oxidation rate by matrix volume changes.

It has also been shown in isolated hearts that the same drugs opening (cromakalin and diazoxide), or closing (glibenclamide, 5-hydroxydecanoate), mitochondrial K_{ATP} channels, enhance or abolish, respectively, the beneficial effects that arise by imposing to the myocardium brief ischemic episodes (named 'ischemic preconditioning'). During these episodes, it has been hypothesized that mitochondria produce moderate amounts of reactive oxygen species (ROS) that, by triggering ROS-quenching mechanisms, or by stimulating cell-survival pathways, protect hearts from the ROS-mediated irreversible injury caused by subsequent prolonged ischemic periods. Thus, the cardioprotective role of K_{ATP} could be rationalized in the mitochondrial depolarization provoked by the channel-mediated K^+ entry, which enhances the rate of substrates oxidation, and hence, stimulates the production of ROS.

The second K^+ channel described in the IM has a larger conductance (of around 300 pS), and is activated by matrix Ca^{2+} (mtBK_{Ca}) (Table 1). Because the channel-mediated K^+ influx causes a mitochondrial depolarization, this channel could serve in ischemic preconditioning as hypothesized for K_{ATP} . However, its opening by matrix Ca^{2+} also entails that mtBK_{Ca} could guard mitochondria against dangerous Ca^{2+} overloads, given that the K^+ -induced depolarization reduces the voltage-dependent entry of Ca^{2+} into the mitochondrial matrix. As for its molecular identity, the similar electrophysiological and pharmacologic features have suggested that mtBK_{Ca} coincides with the BK type K_{Ca} channel ($K_{Ca1.1}$) present in the plasma membrane. However, despite the support to this hypothesis by immunological approaches, it is still debatable whether mitochondrial mtBK_{Ca} coincides with that of the plasma membrane, or, conversely, whether the two molecules share only an antigenic similarity.

Intriguingly, the IM of mitochondria could harbor two other channels that have been previously described in the plasma membrane. One of these is the Ca^{2+} -activated K^+ channel $K_{Ca3.1}$, immunologically and electrophysiologically characterized in mitoplasts from human colon cancer cells (mtIK_{Ca}) (Table 1). The presence of $K_{Ca3.1}$ in mitochondria of other cell types, and the physiologic relevance of its IM localisation, remain, however, to be determined. The other

channel, identical to the plasma membrane voltage-dependent Shaker-type K^+ channel, Kv1.3 (Table 1), has been observed in mitoplasts from lymphocytes of wild-type, but not Kv1.3-null, mice. As proposed, mitochondrial Kv1.3 channels could act in mitochondria-mediated apoptotic cell death (see below).

Finally, it is important to mention the Ca^{2+} transport across the IM, in light of the many reactions controlled by the ion, and the dangerous consequences of mitochondrial Ca^{2+} overloads for the life of the cell. The protein responsible for this passage, the so-called Ca^{2+} uniport, has been recently identified following a combination of genetic strategies, patch clamping of mitoplasts, and other electrophysiological approaches. It has a molecular mass of around 40 kDa, and acts as a Ca^{2+} -selective channel of small conductance (MiCa channel/MCU) (Table 1).

Anion selective channels

The first channel identified in the IM of mitochondria has a conductance of around 100 pS (named mitochondrial Centum pico-Siemens, mCS, or mCtS) (Table 1), and is preferentially permeated by anions. Although several molecules could modulate the channel (e.g., nucleotides, Ca^{2+}), its most prominent feature is the voltage dependence, which opens the channel at nonphysiologic (positive) matrix voltages and closes it at the high negative potentials normally sustained by mitochondria.

To search for the role of this channel, much experimentation has followed the concept that the IM transport of small ions may guarantee fundamental aspects of mitochondrial physiology, such as osmotic stability. Detailed insight into such routes has been provided by monitoring the osmotic swelling of isolated mitochondria that, under the used conditions, is the direct consequence of the ability of a specific solute to cross the IM. Several carrier-mediated pathways have been described, two of which are of interest in the context of the ion channels' theme: the above-mentioned K^+ uniport (which is functionally identical to the K_{ATP} channel) and the inner membrane anion channel (IMAC) that allows the permeation of anions different from phosphate, adenine nucleotides, and other metabolites. The initial classification as channel of IMAC derived from its high translocation rate, not because of the, as yet unknown, molecular nature. Accurate electrophysiological reinvestigations of mCS single channels, and of the entire IM, have provided substantial evidence for the functional homology of IMAC and mCS channels; both conduct anions of different valence, are active at alkaline matrix pHs (i.e., alkaline pHs render mCS channels operative also at negative matrix potentials), and are equally blocked by specific drugs. The broad biochemical and pharmacological characterization of K_{ATP} and IMAC has led to propose that both transport systems are implicated in maintaining mitochondrial volume homeostasis. The particular conditions of IMAC activation in intact mitochondria have suggested that the channel mediates efflux of matrix anions as part of the mechanism designed to protect against excessive swelling.

More recently, by relying on pharmacological tests on isolated cardiomyocytes, it has been proposed that the opening of mCS/IMAC may trigger the metabolic oscillations observed in cells deprived of substrates, which, at the mitochondrial level, give rise to periodic, redox-state transitions, and a

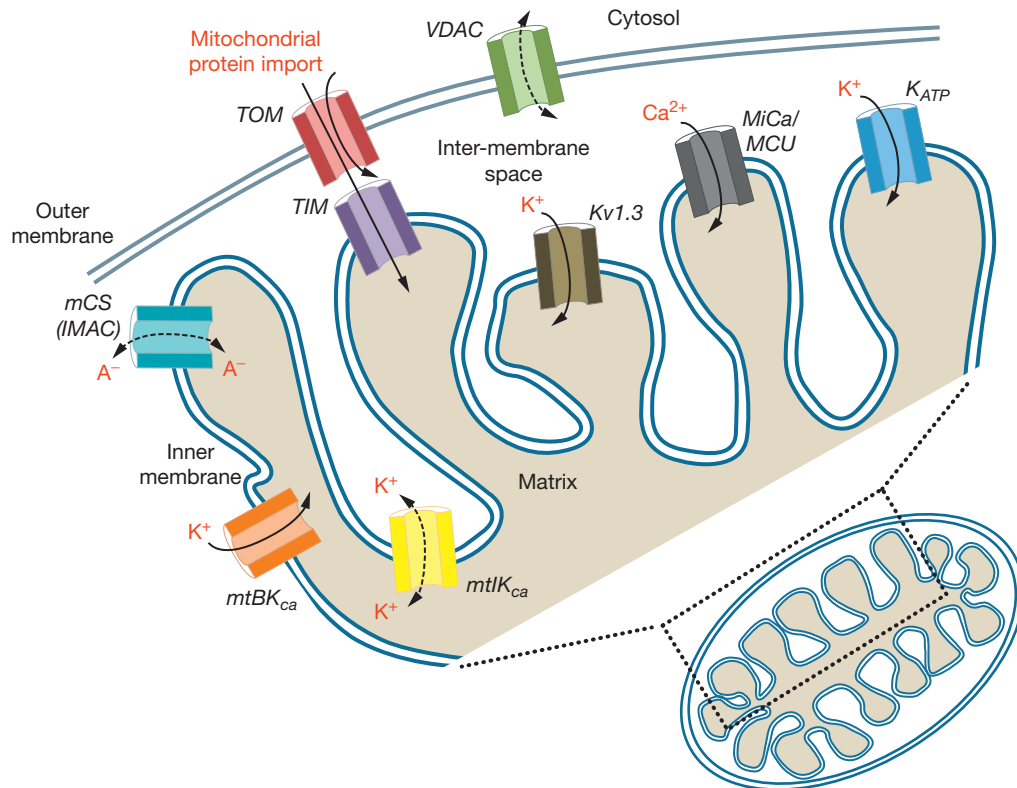


Figure 1 Major putative channels in mitochondrial membranes. The cartoon shows the channels that could be responsible for some of the conductances observed in mitochondrial membranes (listed in **Table 1**). Solid arrows specify the experimentally determined direction of the ion flow through a channel. Conversely, dashed arrows indicate that the direction of the flow is still uncertain. A^- designates anions. No ionic species is indicated for VDAC in view of the largely unselective conductance of the channel. In regions in which the outer and the inner membranes are closely juxtaposed, TOM and TIM together may form a permeability pathway that allows translocation of proteins from the cytosol to the mitochondrial matrix.

continuous switching between dissipation and re-establishment of the proton gradient. As this phenomenon – if uncontrolled – is likely to provoke a whole set of excitability disturbances in the heart, including arrhythmias, mCS channels may thus become important targets of pharmacological heart treatments.

Similar to the TOM machinery, the reconstituted translocase of the IM (TIM) forms permeation pathways with the large conductance (of around 1000 pS) expected for protein-transporting pores (**Table 1**). Complemented by immunologic approaches, this study has also suggested the functional (and molecular) identity of the TIM pore with the MMC/MCC originally described by patch clamping mitoplasts.

A cartoon illustrating the major channels of the OM and IM is present in **Figure 1**.

Involvement of Mitochondrial Channels in Cell Death Mechanisms

Increasingly, it has been reported that mitochondria are implicated in the cell fate through the regulation of the apoptotic process. Apoptosis is a naturally occurring cell death program, which is essential for normal embryonic development and

tissue homeostasis in adults. The process is controlled by a variety of factors, including members of the Bcl-2 family (e.g., the pro- and anti-apoptotic Bax/Bak and Bcl-2, respectively) and mitochondrial proteins of the inter-membrane space (e.g., cytochrome *c*) that, upon release into the cytosol, activate downstream death events. By necessity, passage of these high-mass molecules across the OM calls for a further increase of the already large OM permeability, a process now referred to as mitochondrial outer membrane permeabilization (MOMP). MOMP has been associated with an extremely large single channel conductance, named mitochondrial apoptosis-induced channel (MAC) (**Table 1**), which was observed by patch-clamping proteoliposomes containing OM fractions of cells subjected to apoptotic stimuli. MAC currents are dependent on the presence of Bax, and are abrogated by Bcl-2 overexpression. This has allowed hypothesizing that MAC activity could be related to the capacity of pro-death Bcl-2 family members to insert into the OM and to generate huge pores able to translocate apoptosis-triggering proteins. An alternative explanation entails that the translocation of Bax/Bak to mitochondria could enhance VDAC permeability, but the direct involvement of VDAC in MOMP is still strongly debated.

On the other hand, several other observations suggest that the formation of high conductance pathways in the IM could

play a key role in apoptosis, following the dramatic increase of the IM permeability known as mitochondrial permeability transition (MPT). MPT leads to excessive ion intake by, and swelling of, the mitochondrial matrix, which, in turn, disrupts the OM and releases inter-membrane proteins. Such a process, reported in isolated organelles and in cells challenged with apoptotic stimuli, is caused by the opening of a large, unselective IM pathway, the permeability transition pore (PTP) (Table 1), tentatively identified with the MMC/MCC electric activity. From a molecular point of view, the PTP would be formed by a multisubunit complex that includes proteins located in the matrix (cyclophilin D), IM (adenine nucleotide translocator), and OM (VDAC) proteins. Recently, however, this view has been questioned by studies using mitochondria carrying the ablation of each of these proteins, which have suggested their modulatory, rather than obligatory, role in the PTP system.

See also: **Bioenergetics:** Calcium Transport in Mitochondria; Mitochondrial Membranes, Structural Organization; Mitochondrial

Outer Membrane and the VDAC Channel; Mitochondrial Permeability Transition Pore; Protein Import into Mitochondria; **Cell Architecture and Function:** Potassium Channels in the Inner Membrane of Mitochondria in Various Organisms: From Unicellular Eukaryotes to Higher Plants and Mammals.

Further Reading

- Colombini M (2009) The published 3D structure of the VDAC channel: Native or not? *Trends in Biochemical Sciences* 34: 382–389.
- Kroemer G, Galluzzi L, and Brenner C (2007) Mitochondrial membrane permeabilization in cell death. *Physiological Reviews* 87: 99–163.
- Neupert W and Herrmann JM (2007) Translocation of proteins into mitochondria. *Annual Review of Biochemistry* 76: 723–749.
- O'Rourke B (2007) Mitochondrial ion channels. *Annual Review of Physiology* 69: 19–49.
- Sakmann B and Neher E (eds.) (1995) *Single Channel Recording*. New York: Plenum Press.
- Sorgato MC and Moran O (1993) Channels in mitochondrial membranes: Knowns, unknowns, and prospects for the future. *Critical Reviews in Biochemistry and Molecular Biology* 18: 127–171.

Mitochondrial DNA

G Schatz, Swiss Science and Technology Council, Bern, Switzerland

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 2, pp. 693–696, © 2004, Elsevier Inc.

Glossary

Cytochrome *bc*₁ complex A multi-subunit complex of the respiratory chain that transfers electrons from ubiquinone to cytochrome *c*.

Cytochrome oxidase A multi-subunit green heme protein of the respiratory chain reducing oxygen gas to water.

Endosymbiont Organism living symbiotically inside another organism.

Kinetoplast The single, giant mitochondrion of a trypanosome cell.

Nebenkern Spiral-like single mitochondrion wrapped around the flagellum of certain spermatozoa.

Mitochondria, the eukaryotic organelles of oxidative phosphorylation, contain their own DNA genome. This genome is much smaller than that in the nucleus and encodes only ~1% of the mitochondrial proteins. Yet it is essential for the formation of fully functional mitochondria.

Discovery of Mitochondrial DNA

Some concepts take the scientific world by storm, but others conquer it only after many skirmishes. The discovery of mitochondrial DNA (mtDNA) belongs to this second category. Biochemists, histologists, and electron microscopists had seen DNA in mitochondria for years, but most of them were not ready for the idea that the DNA really belonged there. This may explain why textbook accounts of mtDNA almost never tell how this DNA was discovered.

After the basic building plan of the eukaryotic cell had been revealed in the early 1950s by the electron micrographs of Palade, Sjöstrand, and others, biochemists embraced de Duve's dogma that every macromolecule had one, and only one, intracellular location. In the analysis of cell fractions, cytochrome oxidase was taken as a marker for the mitochondria, nicotianamide adenine dinucleotide phosphate (NADPH)-cytochrome *c* reductase for the endoplasmic reticulum, and DNA for the nucleus. Given this frame of mind, it is easy to understand why the presence of DNA in mitochondrial fractions was generally attributed to contamination by nuclear fragments. Histochemical DNA stains, such as the Feulgen reaction, also stained the kinetoplasts of Trypanosomes and the 'Nebenkern' of insect spermatozoa, but at that time it was not yet recognized that these structures were, in fact, unusual mitochondria. Massive amounts of extranuclear DNA were also detected in the cytoplasm of amphibian oocytes, but it took many years to realize that this DNA was, in fact, mtDNA whose abundance reflected the enormous amount of mitochondria in these large cells. In 1961, at the Fifth Annual Meeting of the American Society of Cell Biology in Chicago, Hans Ris showed electron micrographs of mitochondria with inclusions resembling the DNA-containing nucleoids of bacteria and made the heretical proposal that mitochondria (and also chloroplasts) contain their own DNA. In a paper that appeared in the following

year, Ris and Walter S Plaut further documented and expanded these observations. Soon thereafter, biochemical and morphological evidence from several groups confirmed the presence of DNA in chloroplasts.

The discovery of chloroplast DNA made biochemists take a fresh look at early findings by Margaret Mitchell and Boris Ephrussi that certain mutations affecting mitochondrial function in the mold *Neurospora crassa* and the yeast *Saccharomyces cerevisiae* were not inherited according to Mendel's laws. It seemed tempting to speculate that the unknown 'extrachromosomal factors' implicated in these mutations were, in fact, mtDNA.

By 1962, the ground for the concept of mtDNA was thus well prepared, but the concept itself was not generally accepted. In retrospect, it seems that the scientific community was waiting for convincing studies that documented the existence of mtDNA by several diverse methods.

One of these studies came from the electron microscopists Margit MK Nass and Sylvan Nass, who were then working at the Wenner Gren Institute of the University of Stockholm. They showed that the matrix of osmium-fixed chick embryo mitochondria contained thread-like inclusions whose appearance after different fixation procedures closely paralleled that of the histone-free DNA nucleoid of bacteria: after fixation with osmium tetroxide, the inclusions appeared clumped and as bars with a diameter of ~400 Å; fixation of the tissues with osmium tetroxide followed by treatment with uranyl acetate before dehydration made them appear as 15–30-Å thin fibers. Even more convincing evidence for the presence of DNA in these inclusions was the observation that the inclusions could be removed by treating the lightly fixed embryonic tissue with DNase. Treatment with pepsin, with RNase, or with DNase-free buffer controls was ineffective. The clarity of these electron micrographs and the careful controls that were included had a compelling impact on cell biologists. MMK Nass and S Nass published their work in two back-to-back papers in a 1963 issue of the *Journal of Cell Biology*. At that time, however, cell biology and biochemistry were still rather different disciplines and most biochemists did not peruse journals devoted to cell biology. It therefore took a while before the findings by MMK Nass and S Nass entered the consciousness of the biochemical community.

At about the same time, Ellen Haslbrunner, Hans Tuppy, and Schatz at the Biochemistry Institute of the University of Vienna were trying to find a biochemical basis for the extra-chromosomal mutations that abolished respiratory function in the yeast *S. cerevisiae*. In the early 1960s, many biochemists were still reluctant to consider the 'respiratory granules' of yeast as bona fide mitochondria, which placed the research by Haslbrunner et al. well outside the mainstream of mitochondrial biochemistry in the United States and elsewhere.

To look for DNA in mitochondria, a biochemical approach was chosen. Yeast mitochondria were purified by the best methods available, and their DNA content was measured by the time-honored 'Diesche' color reaction. A few years before, de Duve and co-workers had shown that centrifuging subcellular fractions to equilibrium in a density gradient often gave a clean separation of different organelles. Surprisingly, the usual sucrose gradients did not separate yeast mitochondria from nuclear fragments, but when sucrose was replaced with the X-ray contrasting agent 'Urografin', the mitochondria formed an extremely sharp band, and DNA was present in only two fractions: most was at the bottom of the centrifuge tube, and a very small amount, but discrete peak coincided exactly with that of the mitochondria. The DNA in the bottom fraction was easily digested by DNase and apparently represented nuclear DNA. The DNA in the mitochondrial fraction was not readily digested by DNase unless the organelles were first disrupted with trichloroacetic acid; presumably, it represented DNA enclosed by the mitochondrial membranes. Its concentration was very constant between different experiments – between 1 and 4 $\mu\text{g mg}^{-1}$ mitochondrial protein. Urografin – purified mitochondria from rat liver, rat kidney, and bovine heart – contained almost 10 times less DNA, between 0.2 and 0.6 μg DNA per mg protein. The typical mammalian mitochondrion was calculated to contain 3×10^{-17} g DNA. Assuming that the DNA was double-stranded, it could encode no more than 1.2 MDa of polypeptide chains. This result was considered important, because it firmly excluded the possibility that mtDNA encoded all of the mitochondrial proteins. Today, this early calculation by Haslbrunner et al. could be challenged on several grounds, yet it came remarkably close to reality: the 13 polypeptides encoded by mammalian mtDNA have a total mass of 0.423 MDa, and the remainder of the coding potential is largely accounted for by genes for ribosomal and transfer RNAs, as well as by the fact that mitochondria usually have more than one copy of their DNA genome.

Properties of mtDNA

Soon thereafter, several laboratories showed that mtDNA from several organisms had a different base composition and sometimes also a different buoyant density in cesium chloride gradients than the corresponding nuclear DNA. Particularly, striking was the observation that mtDNA from mammals was a small double-stranded circle. Late in 1964, the existence of DNA in mitochondria was generally accepted, and efforts focused on its function. Its identity with the 'extrachromosomal factors' affecting respiratory function in *S. cerevisiae* was established by the observations that mtDNA was either altered or missing in the corresponding mutants. Several

laboratories then showed that mitochondria from yeast and *N. crassa* synthesized the three large subunits of cytochrome oxidase, the heme-carrying subunit of the cytochrome *bc*₁ complex, and at least one membrane-bound subunit of the adenosine triphosphate synthase complex.

Sequence and Gene Content

In 1981, Fred Sanger and co-workers published the complete sequence of the 16 569 nucleotides of human mtDNA. This seminal paper provided the first complete sequence of a eukaryotic genome and identified the mitochondrial genes. Subsequent work by others then showed that the size of mtDNA can vary dramatically between different species (Table 1). There are also considerable differences in gene content, whereas mtDNA from humans and most mammals encodes 13 proteins and 24 RNAs (Figure 1 and Table 2),

Table 1 Sizes of mtDNAs from different species

Species	Kilo-base pairs
Mammals	16.5
Fungi	
<i>Saccharomyces cerevisiae</i>	78
<i>Schizosaccharomyces pombe</i>	17
<i>Aspergillus nidulans</i>	32
Plants	

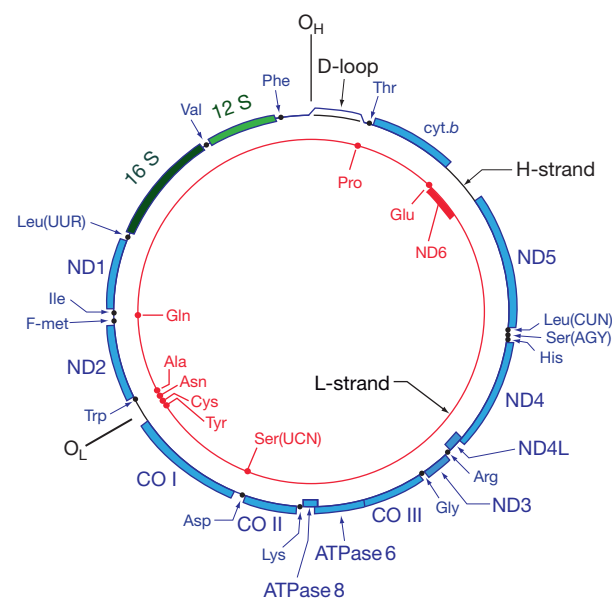


Figure 1 Genes on human mtDNA. 12S and 16S, rRNA of small and large ribosomal subunit; ND1–6 and ND4L, subunits of NADH dehydrogenase; CO I–III, subunits 1–3 of cytochrome oxidase; cyt. *b*, cytochrome *b* subunit of cytochrome *bc*₁ complex; ATPase 6 and 8, subunits of ATP synthase complex; amino acids designate the corresponding tRNAs; D-loop, highly variable gene-free region; O(H) and O(L), sites previously proposed by some as strand-specific replication origins. Reproduced from Attardi G (1986) The elucidation of the human mitochondrial genome: A historical perspective. *Bioassays* 5: 34–39.

Table 2 Genes on human mtDNA

Function	Molecule(s)	Number
Protein synthesis	rRNAs	2
	tRNAs	22
Electron transport	Cytochrome oxidase	3
	NADH dehydrogenase	7
	Cytochrome <i>bc</i> ₁ complex	1

Table 3 Unusual codon usage in mtDNAs

Codon	Universal usage	Mitochondrial usage	Species
CUA	Leucine	Threonine	Yeast
UGA	Stop	Tryptophan	Most species, not plants
AUA	Isoleucine	Methionine	Most species, not plants
AG(A,G)	Arginine	Stop	Mammals
		Serine	Insects
		Arginine	Yeast

that from the freshwater protozoan *Reclinomonas americana* carries no fewer than 97 genes. Sequence comparisons between some of the genes and the corresponding polypeptide products also uncovered the astounding fact that the genetic code of mtDNA differs slightly but significantly from that of nuclear DNA (Table 3). For example, the codon UGA defines 'stop' in the nucleus, but 'tryptophan' in mtDNA from mammals, yeast, and many other species. This unexpected finding explained why so many earlier attempts to identify the protein products of mtDNA by coupled transcription/translation in *Escherichia coli* or reticulocyte lysates had failed.

Outlook

There is no end to the fascinating properties of the mitochondrial genome, and many of these properties still beg an explanation. Why is the mitochondrial genome of mammals one of the most compact genomes known, whereas that of plants and yeast spreads its genes luxuriously over 5, 10, or even 100 of times more DNA? Why has the genetic code diverged from the 'standard' one that is used in the nucleus? Why does mtDNA mutate ~10 times faster than the corresponding nuclear DNA? Why some mitochondria have dozens of copies of their genome? Most importantly, why does mtDNA exist at all? Maintaining its few protein-coding genes from one generation to the next keeps hundreds of nuclear genes

occupied, which seems highly uneconomical. Evolution provides the answer: it is now generally accepted that today's mitochondria are descendants of free-living prokaryotes that entered into symbiosis with other cells and transferred most of their genome to the nucleus of their host. The mtDNA is the small remnant of the genome of these endosymbionts.

As a typical somatic cell has only two copies of a nuclear gene, but as many as a thousand copies of each mitochondrial gene, mtDNA has become a favorite tool for genetic fingerprinting and evolutionary studies. Also, it is now well established that mutations in mtDNA can cause a host of devastating maternally inherited diseases, such as blindness, deafness, muscular degeneration, diabetes, and severe neurodegenerative syndromes. A possible involvement of mtDNA in aging and cancer is under intense investigation. There is also mounting evidence that mutations of mtDNA may impair mammalian fertility and the success of *in vitro* fertilization experiments. Indeed, many of the most exciting functions of mtDNA may yet await discovery.

See also: [Bioenergetics: Cytochrome *bc*₁ Complex \(Respiratory Chain Complex III\); Cytochrome Oxidases, Bacterial; Mitochondrial Genes and Their Expression: Yeast.](#)

Further Reading

- Andersen S, Bankier AT, Barrell BG, et al. (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290: 457–465.
- Attardi G (1986) The elucidation of the human mitochondrial genome: A historical perspective. *Bioassays* 5: 34–39.
- Attardi G and Schatz G (1988) Biogenesis of mitochondria. *Annual Review of Cell and Developmental Biology* 4: 289–333.
- Attardi G, Yoneda M, and Chomyn A (1995) Complementation and segregation behavior of disease-causing mitochondrial DNA mutations in cellular model systems. *Biochimica et Biophysica Acta* 1271: 241–248.
- Chomyn A and Attardi G (2003) MtDNA mutations in aging and apoptosis. *Biochemical and Biophysical Research Communications* 304: 519–529.
- Clayton DA (1991) Replication and transcription of vertebrate mitochondrial DNA. *Annual Review of Cell and Developmental Biology* 7: 453–478.
- Grivell LA (1983) Mitochondrial DNA. *Scientific American* 248: 78–89.
- Nass MMK and Nass S (1963) Intramitochondrial fibers with DNA characteristics II. *Journal of Cell Biology* 19: 613–629.
- Nass S and Nass MMK (1963) Intramitochondrial fibers with DNA characteristics I. *Journal of Cell Biology* 19: 593–612.
- Schatz G, Haslbrunner E, and Tuppy H (1964) Deoxyribonucleic acid associated with yeast mitochondria. *Biochemical and Biophysical Research Communications* 15: 127–132.
- Schatz G, Haslbrunner E, and Tuppy H (1964) Intramitochondriale Deoxyribonukleinsäure in Säugetierzellen. *Monatshette für Chemie* 95: 1135–1140.
- Scheffler IE (1999) *Mitochondria*, pp. 364. New York, NY: Wiley.
- Scheffler IE (2000) A century of mitochondrial research: Achievements and perspectives. *Mitochondrion* 1: 3–31.
- Wallace DC (1977) DNA mitochondrial, in aging and disease. *Scientific American* 277: 40–47.

Mitochondrial Dynamics

L Scorrano, University of Geneva, Geneva, Switzerland

Published by Elsevier Inc.

Glossary

Cristae The folds of the inner mitochondrial membrane; these pleiotropic structures are the site of oxidative phosphorylation.

Dynamins Large ubiquitous guanosine triphosphate (GTP)-hydrolyzing enzymes that participate in fusion, fission, and tubulation of biological membranes.

Microdomain A locale of high concentration of an ion or a second messenger, often limited in space and time.

Tethering The close apposition between the outer mitochondrial membrane and the endoplasmic reticulum.

Introduction

Mitochondria were once regarded as the organelles merely devoted to energy conversion until it was discovered that they regulate apoptosis by releasing cytochrome *c*, the only soluble components of the respiratory chain, as well as other proteins that activate the machinery of cell self-destruction. However, the role of these organelles is not limited to the cell life and death decision, but they are also key decision makers in a variety of cellular cascades that range from Ca^{2+} signaling to cellular movement, to establishment of synapse, even to epigenetics. This functional versatility goes hand in hand with the morphological complexity of this organelle. Mitochondria possess two membranes and their inner membrane can be functionally further subdivided in a boundary membrane and in the cristae membrane, pleiomorphic infolded structures of the inner membrane. Furthermore, in the cytoplasm the organelle can live a solitary life, or be interconnected with other mitochondria to form a network. Shape and morphology are continuously regulated by the balance between fusion and fission events. In the last few years, we learned that mitochondrial fusion and fission are essential for embryonic development, participate in apoptosis, maintenance of dendritic spines, migration of white blood cells, Ca^{2+} signaling, determination of life span, and are affected in several human genetic diseases. This article wishes to offer to the reader an account of the surging field of research on how, why, and when mitochondria fuse and divide, and of the functional and pathophysiological consequences of changes in mitochondrial morphology.

Mitochondrial Morphologies

Mitochondria are the metabolic hub of the cell, ensuring the conversion of the dietary intake into adenosine triphosphate (ATP), the usable energy molecule that is required for most of the endergonic processes. This highly efficient process is carried on by the oxidative phosphorylation, where electrons are extracted from reducing equivalents generated and passed in a stepwise process to molecular oxygen. The energy liberated by the reducing steps is coupled to the pumping of protons and therefore to the generation of an electrochemical gradient

across the inner mitochondrial membrane. According to Mitchell's chemiosmotic theory, the free energy accumulated across the inner membrane is used by the ATP synthase to generate ATP from adenosine diphosphate (ADP) and phosphate. This well-organized system takes advantage of the specialized morphology of the organelle. As mentioned above, mitochondria possess an outer membrane that separates them from the cytosol, and an inner one that constitutes the active site of the oxidative phosphorylation, where the four complexes of the respiratory chain as well as the ATPase reside. The two membranes enclose two aqueous spaces, the matrix, where Krebs cycle is located, and the intermembrane space, where many proteins are stored, including the one and only soluble component of the mitochondrial respiratory chain, cytochrome *c*.

Recent advances in the techniques used to visualize the ultrastructure of the mitochondria unveiled that the inner membrane can be further subdivided into two different compartments, the inner boundary membrane and the cristae (Figure 1). These latter, according to the classic model elaborated in the 1950s by G. Palade, were invaginations of the inner membrane widely opened to the intermembrane space. Nowadays, thanks to the pioneering electron tomography (ET) studies of Mannella and Frey, cristae are identified as a separated compartment, resembling pleomorphic bags connected by a bottleneck called 'cristae junction' to the inner boundary membrane. The bag-like appearance of the cristae has a clear functional consequence on the efficiency of the energy conversion system of mitochondria. Integral proteins of the inner membrane are not distributed in a uniform manner between inner boundary and cristae membrane. Accurate immunoelectron microscopy studies revealed that the complexes of the respiratory chain accumulate in the membrane of the cristae, whereas the components of the machinery for mitochondrial protein import are mostly in the inner boundary membrane. This organization takes advantage of the increased surface of the cristae compared to that of the inner boundary membrane, thus maximizing the condenser capacity of the membrane and increasing the efficiency of the oxidative phosphorylation. On the other hand, it requires specialized machineries that regulate the shape of the organelle and ensure the subcompartmentalization of the inner membrane of the organelle.

A further layer of complexity in the regulation of mitochondrial morphology is introduced by the fact that mitochondria

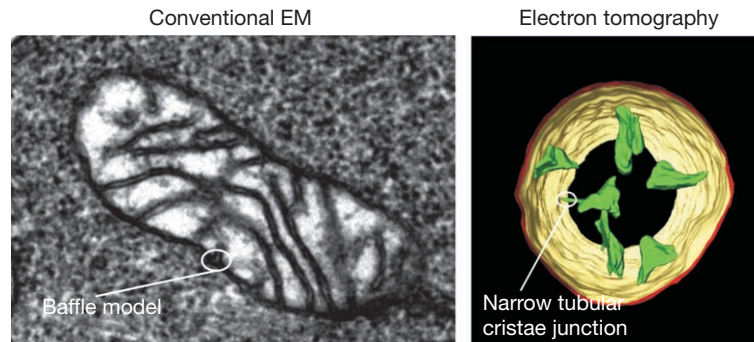


Figure 1 Mitochondrial ultrastructure: from electron microscopy to electron tomography. A conventional electron microscopy and an electron tomogram of mitochondria from mouse liver hepatocytes are shown. In the electron tomography, the outer membrane is in red, the inner boundary membrane in beige, and the cristae in green. The circled area points to the different appearance of the cristae junctions with the inner membrane.

are often organized in the cytoplasm in a network, a reticulum of interconnected organelles continuously shaped by opposing fusion and fission processes. Interestingly, in different cell types morphology of the mitochondria differs: for example, solitary organelles are retrieved in hepatocytes, whereas in many epithelial cells they form an intricate network of interconnected ones. The case of myocytes is even more interesting, with a zonal distinction in mitochondrial morphology: perinuclear mitochondria are mostly globular, while rod-shaped mitochondria are predominantly retrieved in the subsarcolemmal space. Finally, intermyofibrillar mitochondria have exactly the same size of a sarcomere. Thus, different organizations even in the context of the same cell highlight the requirement for a refined machinery that regulates the processes of fusion and fission which are the basis for the observed steady-state morphology of the organelle.

Mitochondrial Fusion and Fission

The final shape of mitochondria results from the equilibrium between two antagonistic processes, fusion, and fission, that constantly take place. When either of them is blocked, the final morphology of the mitochondria is the consequence of the unopposed progression toward the other side of the equilibrium. Members of the machinery regulating mitochondrial fusion have been identified by the analysis of organellar morphology in collections of mutants of *Saccharomyces cerevisiae*. The shape of mitochondria depends on several guanosine triphosphatases (GTPases) with structural homology to dynamins, large ubiquitous guanosine triphosphate (GTP)-hydrolyzing enzymes that participate in fusion, fission, and tubulation of biological membranes. Proteins that regulate mitochondrial shape share with the prototypical dynamins at least the GTPase domain and a C-terminal coiled-coil domain that can function as GTPase effector domain or to mediate protein-protein interaction. On the other hand, the dynamin-related proteins that impinge on mitochondrial fusion are peculiar in that most of them are integral membrane proteins. We now discuss the properties of mitochondria-shaping proteins of mammals.

Mitochondrial Fission

In mammalian cells, mitochondrial fission depends on dynamin related protein 1 (Drp1), a cytoplasmic large GTPase similar to dynamin that mediates the fragmentation of mitochondria, peroxisomes, and endoplasmic reticulum (ER). Drp1 translocates to mitochondria in response to cellular and mitochondrial cues. Following mitochondrial dysfunction, cytoplasmic Ca^{2+} rises, leading to activation of calcineurin and dephosphorylation of the conserved Ser 637 of Drp1. This residue is phosphorylated by protein kinase A, linking mitochondrial morphology to another crucial second messenger, cyclic adenosine monophosphate (cAMP). The phosphorylation status of this site is dominant over that of Ser 616, controlled by cyclin-dependent kinase 1 to drive mitochondrial fission during mitosis. Mitochondrial Drp1 can then be stabilized on the surface of the organelle by SUMOylation, a process known to subtract molecules to degradation by the ubiquitin-proteasome system.

Fis1 is a membrane protein homogeneously distributed in the outer mitochondrial membrane (OMM) via a transmembrane domain located in the C-terminal region, with only a small portion of the molecule protruding into the intermembrane space (IMS). The cytoplasmic region contains six α -helices, four of which $\alpha 2$ – $\alpha 5$ form two tetratricopeptide repeat (TPR)-like domains, predicted to allow protein-protein interactions. Overexpression of Fis1 induces mitochondrial fragmentation, but since it does not possess enzymatic activity, its role is probably restricted to anchoring effector proteins to mitochondria. Accordingly, fragmentation of mitochondria caused by Fis1 overexpression can be blocked by expression of dominant negative mutant of Drp1. Moreover, Drp1 and Fis1 appear to interact, as judged by cross-linking and coimmunoprecipitation. The Drp1-Fis interaction is a transient event and efficient fission requires dissociation of the complex. Accordingly, genetic modifications of TPR-like domain, which stabilize this complex, abolish fragmentation of mitochondria. Binding to Drp1 is executed by $\alpha 2$ – $\alpha 4$ helices of TPR-like domain, while $\alpha 1$ has a negative regulatory function. However, the absolute requirement of Fis1 for binding of Drp1 to mitochondria is somewhat weakened by the observation that downregulation of Fis1 diminishes Drp1 recruitment to mitochondria only partially, and some

novel receptors, called mitochondrial fission factor (Mff) and mitochondrial division (MiD) 49 and 51 have been recently discovered. Mff, previously identified in a *Drosophila* RNA interference (RNAi) screening for mitochondrial morphology alterations, is a C-tail anchored protein of the outer membrane of the organelle, exposed to the cytosol. Ablation of Mff causes mitochondrial elongation and results in the retention of Drp1 in the cytoplasm, whereas its overexpression induces Drp1 recruitment and mitochondrial fission. Indeed, Mff transiently interacts with Drp1 via its cytoplasmic region. The current model indicates that Mff may facilitate membrane fission by promoting the self-assembly and the GTPase activity of Drp1, probably in conjunction with some yet unknown factors. MiD49 and MiD51 are N-terminal anchored MOM proteins, forming foci or ring-like structures on mitochondria and similarly involved in the recruitment of Drp1 onto mitochondria. The relationship between MiDs and Mff remains unclear and is to be investigated.

Mitochondrial Fusion

Fusion of the OUM is governed by two dynamin-related GTPases – mitofusins 1 and 2 (Mfn1 and Mfn2). They display a very high degree of homology and accordingly a similar structure composed of a terminal GTPase domain, two hydrophobic heptads repeats (HRs) and two transmembrane domains that insert them in the OMM. Nevertheless, they are not functionally equal. GTPase activity of Mfn1 is much higher but affinity toward GTP is lower, seemingly playing slightly different roles in mitochondrial fusion. Mfn1 is responsible for mitochondria tethering *in trans*. The role of Mfn2 is somewhat elusive, but it can be retrieved in heterooligomers with Mfn1 and is suggested to participate in later steps of mitochondrial fusion. In addition, levels of Mfn2 correlate with the oxidative metabolism of skeletal muscle and the proliferation ability of vascular smooth muscle cells, by sequestering the proto-oncogene Ras. Mfn2 also controls the shape of ER and tethers it to mitochondria. On the other hand, at least in

fibroblasts, Mfn1, but not Mfn2, is required for fusion triggered by the inner membrane dynamin-related protein Optic atrophy 1 (Opa1), further complicating the picture.

Opa1 is anchored to the IMM by a transmembrane domain located close to the N-terminus, and most of the protein is exposed to the IMS. In humans, there are eight splice variants of Opa1, while in mice only four. All alternatively spliced exons are located in the GTPase domain on N-terminus. Isoforms of Opa1 are then subject to a complex posttranslational cleavage. A model of sequential cleavage has been proposed in which the *m*-AAA protease paraplegin or AFG3L2, or the *i*-AAA protease Yme1L, produces the lower MW forms of Opa1, on which then the rhomboid protease Parl acts to release soluble Opa1 in the IMS space. So far two functions of Opa1 have been defined: as mentioned above, Opa1 drives Mfn1-dependent fusion of mitochondria. Fusion requires the presence of both long and short forms of Opa1, suggesting that different forms of Opa1 may complex to govern mitochondrial fusion. Opa1 is also crucial for maintenance of the structure of IMM. In contrast to the control over mitochondrial fusion, dynamics of IMM seem to depend exclusively on Opa1. Its downregulation causes the appearance of vacuolar cristae and the widening of the cristae junctions, as a consequence of the disruption of a complex comprising at least IMM and IMS forms of the protein. A cartoon depicting our current knowledge of the players involved in mammalian mitochondrial fusion is presented in Figure 2.

Mitochondrial Tethering to the ER

Mitochondria are often found in close association with other cellular structures, ranging from cytoskeletal components to organelles, to the plasma membrane. Probably the best characterized interaction, and the one for which mitochondrial dynamics has been proposed to play a role, is that with the ER, which is not only structural, but also functional, impacting on transport of proteins, Ca^{2+} signaling, and cell death. Since

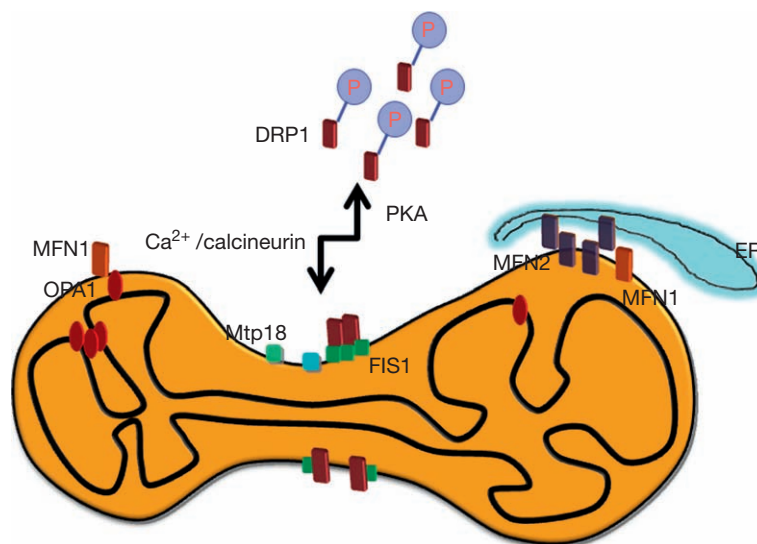


Figure 2 A cartoon of the key players in mitochondrial fusion and fission and tethering to ER. Mitochondria-shaping proteins and other proteins involved in the fission and fusion machinery are depicted. For a detailed description, the reader is referred to the main text.

the 1970s ultrastructural studies proposed the existence of a close physical relationship between ER and mitochondria in animal and plant cells. This was confirmed more recently by fluorescence microscopy and ET. Up to 20% of the mitochondrial surface is in close apposition with the ER membrane, and the two organelles move in synchrony and exchange several metabolites. This suggests that the machineries of mitochondrial morphology and tethering are somehow connected, as testified by several emerging studies.

A complex formed by some mitochondrial morphology proteins was found to link ER to mitochondria in yeast. The ER-mitochondria encounter structure (ERMES) is formed by Mmm1, localized on the ER membrane, and Mdm10, on the OMM. It comprises also Mdm12 and Mdm34, which are essential for the Mmm1–Mdm10 binding. Mutations associated with one of the ERMES components cause lipid reduction and defects in ER–mitochondria lipid transfer.

Mfn2 is the first mammalian protein identified that directly acts in maintaining ER–mitochondria linkage. It is present in both organelles and regulates their morphology. Mfn2 is enriched in MAMs and connects ER to mitochondria via direct interactions between the protein localized in the ER and Mfn1 or the same Mfn2 present in the OMM. Ablation of Mfn2 causes the destruction of the typical tubular ER structure, forming large vesicles. Moreover, it slows down mitochondrial Ca^{2+} uptake, as a consequence of the increased distance between the two organelles that ultimately disrupts the Ca^{2+} microdomain at the interface between the two.

The Role of Mitochondrial Shape in Cell Function

In the last few years, we learned that changes in mitochondrial shape influence crucial cellular functions, from Ca^{2+} signaling to generation of reactive oxygen species, to neuronal plasticity, to muscle atrophy, to lymphocyte migration, and even to life span. Why is mitochondrial morphology so important for cellular physiology? We think that a few aspects could account for this: compartmentalization achieved by mitochondrial localization, the mitochondrial shape–function relationship, and effects of mitochondrial shape on viability. Mitochondria localize where high ATP amounts are required, or where Ca^{2+} signaling needs to be regulated, as if they are movable tuners of signaling events. Not surprisingly, mitochondrial movement is coordinated with shape changes, in order to produce organelles whose size is compatible with their movement. For example, the expression of pro-fusion shaping proteins decreases mitochondrial movement along axons and dendrites and consequently the amount of dendritic spines and synapses, whose formation or maintenance probably requires locales of high mitochondrial ATP production. Mitochondria-shaping proteins can also influence organellar function. Back in the 1960s, Hackebrock described the classical transition from the orthodox morphology of quiescent ‘state 4’ to the condensed one of respiring, ‘state 3’ mitochondria, showing a clear structure–function relationship for mitochondria. Recently, an association between different phosphorylation abilities, nutrient availability, induction of catabolic processes such as autophagy and mitochondrial morphology has been unveiled; however, an unequivocal relationship between P/O ratios (the number of

ATP molecules produced per oxygen molecule oxidized) and mitochondrial morphology is lacking, raising the question of whether the two are directly linked. Finally, mitochondria-shaping proteins have been reported to play a crucial role in apoptosis, as will be described extensively in the next section.

The Role of Mitochondrial Shape in Cell Death

Several reports indicate that mitochondrial shape and ultrastructure vary during apoptosis. We will review the evidence implicating the mitochondria-shaping proteins in cell death and the proposed mechanisms by which they might influence it.

Mitochondrial Fragmentation

The group of J.C. Martinou originally reported that during neuronal apoptosis mitochondria fragment. Subsequent studies by the group of R. Youle discovered that this apoptotic fission occurs in most cell types and can be blocked by a dominant negative mutant of Drp1, limiting cell death. In the membrane, Drp1 does not distribute homogeneously but forms foci, which usually colocalize later with scission sites. Interestingly, upon apoptotic stimulation at least two other proteins colocalize in foci: Mfn2 and the pro-apoptotic member of the Bcl-2 family Bax, which is required for outer membrane permeabilization. Our understanding of these foci remains elusive, but several lines of evidence suggest that foci may serve as platforms for cross-talk between pro- and anti-apoptotic as well as fusion and fission molecules. Interestingly, Drp1 seems to control the release of cytochrome *c* but not of the other pro-apoptotic factor second mitochondrial activator of apoptosis/direct IAP binding protein with low pI (Smac/DIABLO) from mitochondria, calling for a deeper understanding of the role of Drp1 in outer membrane permeabilization and in the cross-talk with the members of the Bcl-2 family.

Cristae Remodeling

Under basal conditions, the majority of cytochrome *c* is located in the cristae, where the complexes of the respiratory chain reside. To enable its complete release during apoptosis, cytochrome *c* is redistributed to the peripheral IMS in a process discovered by our laboratory and christened ‘cristae remodeling’, characterized by fusion of individual cristae and widening of the narrow cristae junctions.

Remodeling of the cristae occurs in response to a variety of apoptotic stimuli, including pro-apoptotic members of the Bcl-2 family such as Bid, Bim, Bnip3, and Bik. In the latter case, remodeling seems to be mediated by Drp1-dependent mitochondrial fission, thereby linking the outer to the inner side of mitochondrial shape changes during cell death. In accordance with its proposed function in cristae biogenesis, Opa1 is the master regulator of the shape of the cristae and therefore of the rate and extent of cytochrome *c* during apoptosis. This function depends on its GTPase activity, on its oligomerization, and on the efficient production of the soluble forms of the protein controlled by the rhomboid protease Parl. During apoptosis, this Opa1-containing oligomer is targeted and disrupted, allowing the complete release of cytochrome *c* and the progression of the apoptotic cascade.

Conclusions and Perspectives

The field of mitochondrial dynamics is very young and expanding. The recent discoveries that the shape of the organelle is linked to many pathophysiological processes, the availability of molecular tools to investigate if and how shape and function go hand in hand, and the cross-field interest are all good omens for intense research that will further elucidate the structure–function relationship of mitochondria.

See also: **Bioenergetics:** Cell Death by Apoptosis and Necrosis; **Cytochrome *c*;** **Cell Architecture and Function:** Endoplasmic Reticulum-Associated Protein Degradation.

Further Reading

- de Brito OM and Scorrano L (2008) Mitofusin 2 tethers endoplasmic reticulum to mitochondria. *Nature* 456: 605–610.
- de Brito OM and Scorrano L (2010) An intimate liaison: Spatial organization of the endoplasmic reticulum–mitochondria relationship. *EMBO Journal* 29: 2715–2723.
- Frank S, Gaume B, Bergmann-Leitner ES, et al. (2001) The role of dynamin-related protein 1, a mediator of mitochondrial fission, in apoptosis. *Developmental Cell* 1: 515–525.
- Frey TG and Mannella CA (2000) The internal structure of mitochondria. *Trends in Biochemical Sciences* 25: 319–324.
- Frezza C, Cipolat S, Martins de Brito O, et al. (2006) OPA1 controls apoptotic cristae remodeling independently from mitochondrial fusion. *Cell* 126: 177–189.
- Gomes LC, Di Benedetto G, and Scorrano L (2011) During autophagy mitochondria elongate, are spared from degradation and sustain cell viability. *Nature Cell Biology* 13: 589–598.
- Hackenbrock CR (1966) Ultrastructural bases for metabolically linked mechanical activity in mitochondria. I. Reversible ultrastructural changes with change in metabolic steady state in isolated liver mitochondria. *Journal of Cell Biology* 30: 269–297.
- Li Z, Okamoto K, Hayashi Y, and Sheng M (2004) The importance of dendritic mitochondria in the morphogenesis and plasticity of spines and synapses. *Cell* 119: 873–887.
- Liesa M, Palacin M, and Zorzano A (2009) Mitochondrial dynamics in mammalian health and disease. *Physiological Reviews* 89: 799–845.
- Martinou I, Desagher S, Eskes R, et al. (1999) The release of cytochrome *c* from mitochondria during apoptosis of NGF-deprived sympathetic neurons is a reversible event. *Journal of Cell Biology* 144: 883–889.
- Palade GE (1952) The fine structure of mitochondria. *Anatomical Record* 114: 427–451.
- Romanello V, Guadagnin E, Gomes L, et al. (2010) Mitochondrial fission and remodelling contributes to muscle atrophy. *EMBO Journal* 29: 1774–1785.
- Scheckhuber CQ, Erjavec N, Tinazli A, et al. (2007) Reducing mitochondrial fission results in increased life span and fitness of two fungal ageing models. *Nature Cell Biology* 9: 99–105.
- Scorrano L, Ashiya M, Buttle K, et al. (2002) A distinct pathway remodels mitochondrial cristae and mobilizes cytochrome *c* during apoptosis. *Developmental Cell* 2: 55–67.
- Scorrano L and Korsmeyer SJ (2003) Mechanisms of cytochrome *c* release by proapoptotic BCL-2 family members. *Biochemical and Biophysical Research Communications* 304: 437–444.

Mitochondrial Genes and Their Expression: Yeast

P P Slonimski[†], University Pierre et Marie Curie, Paris, France
G Carignani, University of Padova, Padua, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Self-splicing The intrinsic ability of some introns to remove themselves and link together the two adjacent RNA exons, by transesterification reactions.

Spliceosome The large RNA–protein body upon which splicing of nuclear messenger RNA precursors occurs.

Suppressor gene A gene which, by mutation, is able to eliminate the effects of a mutation in another gene.

Transporter outer membrane–transporter inner membrane (TOM–TIM) proteins Proteins forming receptor complexes that are needed for translocation of proteins across the mitochondrial outer (TOM) or inner (TIM) membrane.

Mitochondria possess their own genome, which encodes a small part of the proteins that make up the organelle. The remaining proteins, that is, the majority, are encoded by the nuclear genome. Mitochondrial life thus depends on the coordinated expression and interaction of nuclear genes and genes present on the organellar genome (mitochondrial DNA (mtDNA)) itself. Although mtDNA encodes very few subunits of the energy-transducing complexes of the inner mitochondrial membrane, these subunits constitute the key elements of the overall process: cytochrome *b* for the reductase activity, cytochrome *c* oxidase subunit 1 for the oxidative activity, and Atp6 for the synthetase activity. The genes coding for these three subunits are always located in the mtDNA in all eukaryotes, whether fungi, protists, animals, or plants. The mtDNA is supposed to be a remnant of a prokaryotic genome that would have originated from a symbiotic event between different cellular species. As the majority of the prokaryotic genes would, however, in the course of evolution, have migrated to the nucleus of the new cell, it is surprising that even today a mitochondrial genome exists. A plausible reason might reside in the compulsory evolution of the formation mechanism of the enzymatic complexes of the inner mitochondrial membrane. This might depend on a topological constraint of the expression and assembly of the different subunits, some of which are very hydrophobic and have to be transcribed, translated, inserted, and assembled from inside the organelle. It follows then that these key genes are present inside the mitochondrion. This would also explain the persistence of mtDNA and its complex machinery of expression for a billion years since the original symbiotic event.

The knowledge of the conjoint expression of both the mitochondrial and nuclear genes involved in energy transduction is best advanced in the study of baker's yeast, essentially due to the extraordinary power of molecular genetics both *in vivo* and *in vitro*.

Yeast as a Model Organism for the Study of Mitochondrial Biogenesis and Function

The facultative anaerobic yeast *Saccharomyces cerevisiae* has provided the main information about mitochondrial biogenesis,

from the discovery of petite mutants by Boris Ephrussi and Piotr Slonimski and their collaborators in the 1950s. This was followed by the exhaustive analysis of mitochondrial genetics in the 1970s in the laboratory of Gif-sur-Yvette and by the identification of mtDNA by Gottfried Schatz, and still remains an excellent experimental organism, owing to its capacity of surviving in the absence of competent mitochondria, to the facility of isolation or construction of mutants (several thousands have already been characterized) and to the possibility of mitochondrial transformation. The search for nuclear genes involved in mitochondrial biogenesis and function is constantly carried out in this organism and the results continually lead to the isolation of functionally homologous human genes. In particular, many mitochondrial disease genes have been identified or studied in this way.

Mitochondrial life depends on the coordinated expression and interaction of nuclear genes and genes present on the organellar genome (mtDNA) itself. The mitochondrial genetic system is required for the synthesis of a limited number of proteins, in particular in yeast of seven highly hydrophobic subunits of the energy-transducing complexes of the inner membrane (Figure 1).

All the processes necessary for the expression of these genes (i.e., transcription, RNA processing and translation), as well as those required for mtDNA maintenance and integrity, take place in the same compartment, the mitochondrion itself, as they occur in a bacterium, the hypothetical mitochondrial ancestor. Also, the different processes are functionally linked, frequently through the activity of proteins exhibiting multiple roles.

Expression of Mitochondrial Genes

Mitochondrial gene expression requires the coordinated interaction of mitochondrial and nuclear-encoded factors: the mitochondrial genome (Figure 1) participates in the process with the genes encoding the 9 S RNA-subunit (*RPM1*) of mitochondrial ribonuclease P (RNase P), the 21S and 15S ribosomal RNAs (rRNAs), a complete set of transfer RNAs (tRNAs), and one protein of the small mitoribosomal subunit, Var1p; in addition, some introns encode proteins necessary for RNA splicing (RNA maturases). The nuclear genome encodes all other factors necessary for mitochondrial genome maintenance and

[†]Deceased.

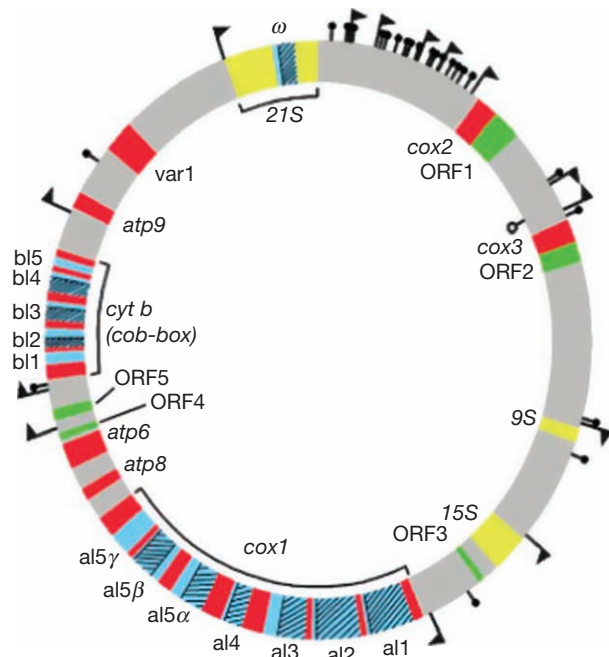


Figure 1 The mitochondrial genome of *Saccharomyces cerevisiae*. The circular map is in agreement with the genetic results and with the observation that mtDNA replicates by the rolling-circle mode. Different yeast strains contain mtDNA molecules of different lengths (comprised between 70 and 85 bp). These differences are essentially due to the presence/absence of introns and hypothetical open reading frames (ORFs 1–5 in this representation). Yeast mtDNA encodes seven highly hydrophobic subunits of the energy-transducing complexes of the inner membrane, that is, apocytochrome *b* (encoded by the *cyt b*, or *cob-box*, gene), three subunits of cytochrome *c* oxidase (encoded by the *cox1*, *cox2*, and *cox3* genes) and three subunits of ATP synthase (encoded by the *atp6*, *atp8*, and *atp9* genes). The *cytb*, *cox1*, and *21S rRNA* genes contain introns, some of which are translated, independently or in frame with their upstream exons, to produce maturases (e.g., the *bl2* intron of the *cytb* gene, see [Figure 3](#)) or site-specific endonucleases (the ω intron of the *21S rRNA* gene); others encode complex proteins with multiple functions, that is, maturases, reverse transcriptases, and endonucleases (e.g., the *al1* intron of the *cox1* gene). Red: exons of protein-coding genes; blue: introns (*al1*–*al5*; *bl1*–*bl5*; ω); hatched blue: intron-encoded ORFs; green: hypothetical ORFs; yellow: 9S, 15S, and 21S rRNAs; solid circles: tRNAs. Flags indicate the transcription initiation sites and their orientation. The orientation of all the transcription units is clockwise, with the exception of the *trnA^{thr}* gene (open circle), which is transcribed from the opposite strand.

expression, as well as all other mitochondrial proteins (about 800 yeast proteins located in the organelle are currently known). These are translated in the cytoplasm and then imported into different compartments of the organelles, using protein-specific import mechanisms. During the last decade, an in-depth analysis of the messenger RNAs (mRNAs) transcribed by these nuclear genes was performed in the laboratory of Claude Jacq in Paris. They have shown that about half of these mRNAs are positioned and translated close to mitochondria, while the other half are translated on free cytosolic polyosomes. The mitochondrially localized mRNAs (MLRs) are addressed there by a targeting sequence present at the amino terminus of translated precursors (mitochondrial targeting sequence (MTS)) and by elements present in the mRNA coding sequence. As MTS is recognized by the mitochondrial

receptor–translocator TOM–TIM, which is responsible for importing protein into mitochondria, it ensues that, for these proteins, the two events, translation and import, are closely connected (co-translational import). Interestingly, MLR proteins are prevalently of prokaryotic origin and are key elements of the core construction of the molecular complexes. The peripheral proteins are instead translated on free cytosolic polyosomes and then imported into mitochondria. The fraction of MLR–mRNA encoding proteins needed for mitochondrial translation and assembly of protein complexes is expressed early in mitochondrial biogenesis and depends on the translational regulator Puf3p. Instead, Puf3p-independent MLR–mRNAs, which encode structural proteins of protein complexes, are expressed successively.

General Features of Simple and Mosaic Genes

To describe the expression of yeast mitochondrial genes, we distinguish between simple genes and mosaic genes, i.e., genes containing introns and thus subject to a more complicated expression pathway.

Transcription

The yeast mitochondrial genome is ~4 times larger than mammalian mitochondrial genomes, although they contain approximately the same number of genes. The difference is due to the lengths of intergenic regions and to the presence of introns in some of the yeast genes. Whereas the 16 000-base-pair mammalian mitochondrial genome is transcribed by two opposite promoters (one for each strand), genes on the 80 000-base-pair yeast mtDNA are transcribed, singly or in clusters, from several promoters (14 active promoters have been identified, see [Figure 1](#)). The consensus sequence of mitochondrial promoters is a nonanucleotide 5'ATATAAGTA 3', whose last A is the +1 position of the transcript, and termination of transcription is indicated by the dodecanucleotide consensus 5'AAUAAUUAUCUU 3' (see [Figure 2](#)).

As in the case of bacterial transcription, all mitochondrial genes are transcribed by the same RNA polymerase. This is composed of only two nuclear-encoded subunits, a core RNA polymerase, Rpo41p, which is similar to the very simple T-odd phages RNA polymerases, and the mitochondrial transcription factor, Mtf1p, which is required for the recognition of the promoter by the holoenzyme and is released when Rpo41p enters into the elongation mode, as occurs for sigma factors of bacterial RNA polymerases. Mtf1p, in fact, shares sequence similarity with regions 2 and 3 of bacterial sigma factors, but, unexpectedly, crystal structure of Mtf1p has revealed clear similarity to the family of dimethyladenosine methyltransferases, and one of the two human homologs of Mtf1p has been proved to possess RNA methyltransferase activity. This is one of the several evidences where the structure/function relationship of a present-day protein indicates that it might have evolved from a multifunctional ancestor or from a protein with a different function.

At least two additional factors, Nam1p (Mtf2p) and Sls1p, interact with the amino-terminal domain of mitochondrial RNA polymerase: these are also involved in posttranscriptional processing and in translation, thus allowing the coordination of transcription with the subsequent events.

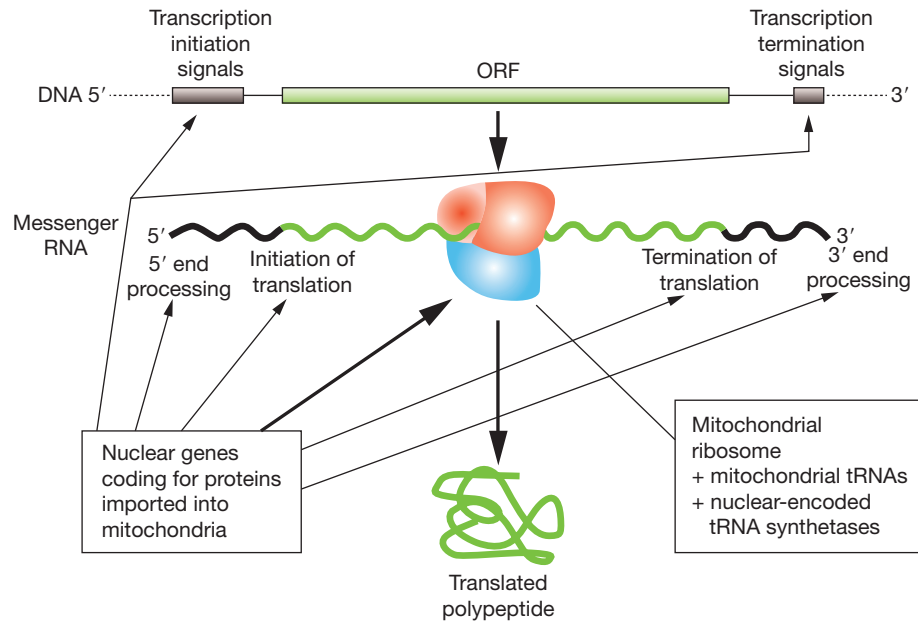


Figure 2 Main features of the expression of a simple mitochondrial protein-coding gene. The open reading frame (ORF) of the gene located in the mtDNA is transcribed into messenger RNA and translated into a polypeptide chain according to the general principles of molecular biology. However, most of the elements of these two machines derive their information from a different cellular compartment than the one where these processes take place. The genes coding for the transcriptional machinery as well as those involved in the processing or stability of mRNA are located in the nucleus, the proteins synthesized in the cytoplasm and imported inside the mitochondrion where they recognize specific signals, that is, mtDNA or mtRNA sequences or structures indicating the beginning and the end of DNA transcription or RNA processing (e.g., Mtf1 and Nam1). The next step of gene expression is even more complex, since in addition to numerous nuclear genes coding for proteins involved in initiation/termination of translation, as well as the translation process itself (more than 70 ribosomal proteins, e.g., Nam9, all translation factors, e.g., Cbp1, and all specific mitochondrially located tRNA synthetases), a number of genes located in the mitochondrial genome participate in the process. They encode essentially the catalytic RNAs: the two, large and small, rRNAs, the complete set of tRNAs, and a subunit of RNase P. As a result of this double inheritance scenario, a mutational lesion in any one of the elements (e.g., the nuclear-encoded leucine tRNA synthetase by gene NAM2 or a mtDNA-encoded tRNA) will ineluctably and irreversibly abolish the expression of all the genes located in mtDNA.

RNA Processing: 5' and 3' Processing

Mitochondrial transcription units are unusual as they are characterized by the presence of mRNAs interspersed with tRNA and rRNAs. In human mtDNA tRNAs genes 'punctuate' the very long principal transcription unit and their processing at both ends 'releases' the majority of other RNAs. In yeast, the mitochondrial transcription units not only are shorter but also contain various combinations of RNAs. tRNA precursors are processed at 5' end by mitochondrial RNase P and at 3' end by a specific endonuclease (successively tRNAs are matured by CCA addition and nucleotide modifications). The processing of mRNAs at their 5' end consists of a cleavage to generate mature mRNA at a position specific for each transcript. In some transcription units, the transcript is first released by the 3' end processing of a preceding tRNA or mRNA and then processed by nuclear-encoded factors (e.g., Pet127p and Cbp1p for the cytochrome *b* transcript). Processing at 3' end of mRNAs is done near the conserved dodecamer sequence 5'AAUAAUUAUUCUU3', which is protected by a complex of three proteins before cleavage by a probably specific endonuclease. All these processes are assisted by several nuclear-encoded proteins with a gene-specific or a more general role (see Figure 2).

The correct expression of the yeast mitochondrial genome depends on the balance between transcription and RNA

degradation. The mitochondrial degradosome complex (mtEXO) is composed of a helicase (Suv3p) and an exoribonuclease (Dss1p) that together create an adenosine triphosphate (ATP)-dependent 3'-to-5' exoribonuclease activity.

Processing of Mosaic Genes Is More Complex: RNA Splicing

A typical feature of yeast mtDNA, which is not shared by human mtDNA (but is common to many mtDNA from plants, fungi, and protists), is that some of its genes are split by introns. In *S. cerevisiae* these are the 21S rRNA, *cytb* (*cob-box*), and *cox1* genes. Splicing of these introns is catalyzed by the intron itself (catalytic RNA or ribozyme), and depends on the formation of a conserved RNA three-dimensional structure, which enables them to undergo self-splicing *in vitro*, albeit inefficiently and only under nonphysiological conditions. *In vivo*, however, splicing requires the presence of specific proteins that function to promote a stable and active RNA conformation. Furthermore, some yeast mitochondrial introns behave as mobile genetic elements, as they can be inserted into an intronless version of the gene at the same position it already occupies in the intron-containing version (intron homing) or to a novel location (intron transposition). Mitochondrial introns belong to two classes, group I and II, based on the secondary and tertiary structure of the intron transcript and on the mechanism of splicing. The mechanism of splicing

of group II introns has suggested that they might be progenitors of nuclear pre-messenger introns and of the spliceosome-catalyzed splicing process, with group II intron domains having evolved into small nuclear RNAs (snRNAs).

Several features of mitochondrial introns have been uncovered in the laboratory at Gif-sur-Yvette in the late 1970s; among these, some encode proteins, that is, RNA maturases, which assist the splicing of the intron itself and sometimes also of a different intron (see Figure 3), besides promoting intron mobility. Intron-encoded proteins also belong to two groups. Group I proteins are members of the LAGLI-DADG family of DNA endonucleases (characterized by the presence of the conserved amino acid motif LAGLIDADG) and, besides the endonuclease activity required for intron mobility via DNA (as occurs for prokaryotic transposons), some of them have high-affinity RNA-binding properties necessary for their role in splicing. Mobile group II introns (e.g., intron aI1 of the *cox1* gene) not only encode multifunctional proteins, endowed with maturase and DNA endonuclease activity, but also encode reverse transcriptases that bind specifically to the intron RNAs to promote both intron mobility and RNA splicing. In the case of group II introns, transposition occurs via RNA, the procedure adopted by eukaryotic nuclear retrotransposons and retroviruses. Some of the mitochondrial intron-encoded proteins are translated in-frame with the upstream exon and

subsequently processed to mature proteins by various proteases. This process shows how the different events of mitochondrial gene expression (in this case, splicing and translation) are strictly interconnected and also enlightens a peculiarity of mitochondrial translation, that it starts immediately after transcription, as in prokaryotes, however before RNA processing, an occurrence unique to this system.

In vivo self-splicing of yeast mitochondrial introns is thus assisted by proteins that, in general, have the role to promote the folding of the catalytic core of the intron, inducing the formation of nucleotide interactions required for catalytic activity. These proteins are the maturases encoded by the intron itself and other proteins, encoded in the nuclear genome, that concur in assisting the splicing event. Many of these proteins have been identified as suppressors of splicing-defective mitochondrial mutations and most of them are proteins that, besides their role in mitochondrial splicing, have also other functions. An example is the mitochondrial enzyme leucyl-tRNA synthetase, Nam2, which is also necessary for the splicing of several group I introns, as it recognizes conserved tRNA-like structural features of the intron RNA. Another example is the Mss116p DEAD box protein, an RNA chaperone with helicase activity, which is required both for efficient splicing of all group I and II introns and for translation activation. The Mrs2 protein, which assists splicing of group II introns, has

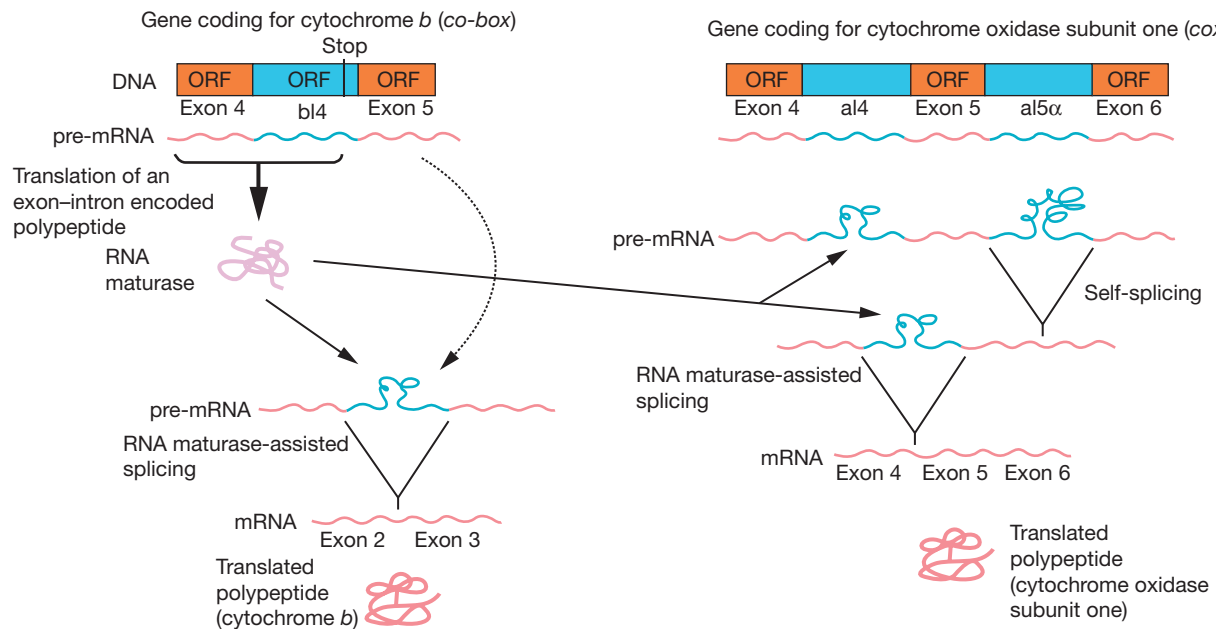


Figure 3 Specific features of the expression of a mosaic mitochondrial protein-coding gene. Several yeast mitochondrial genes contain introns and some of these introns contain ORFs (see Figure 1). The expression of such mosaic genes occurs in two successive steps supplementary to the general features of expression of simple genes (see Figure 2). The primary, unspliced RNA transcript (pre-mRNA) for apocytochrome *b* acts as a messenger for a protein translated from the upstream exon ORF of cytochrome *b* and the ORF of the successive intron. This fusion protein, mRNA maturase (in the example shown referred to as the bI4 RNA maturase), promotes catalytically the excision of the intervening sequence from the pre-mRNA and ligation of cytochrome *b* exonic RNA leading to the formation of cytochrome *b* mRNA. The maturase recognizes specific structures of the intervening sequence, helps the productive folding, and facilitates the activity of the ribozyme. At the same time, the maturase activity destroys the RNA which codes for itself and thus exerts a negative feedback for its own biosynthesis. As a consequence, the amount of maturase in the organelle is very low, and the protein can be detected only when its splicing function is impaired. Generally, the maturases act selectively on the introns which encode them. A notable exception is the maturase shown in the example, that is, the maturase encoded by the fourth intron of the cytochrome *b* gene (bI4): its activity is essential also for the splicing of the fourth intron (aI4) of a different gene, *cox1*, which encodes the subunit one of cytochrome *c* oxidase. In this manner, the expression of the key catalysts of the electron transfer cascade may be coordinated.

actually a more general function, as a mitochondrial inner membrane Mg^{2+} channel. A particularly interesting protein is Nam1, which is involved in the removal of introns from the *cytb* and *cox1* transcripts and in overall translation capacity, and is also responsible for coupling these processes to transcription, through direct interaction with the amino terminus of mitochondrial RNA polymerase, Rpo41.

Translation

Translation of mitochondrial mRNAs resembles more closely its prokaryotic than eukaryotic counterpart with respect to the spectrum of antibiotics inhibiting mitochondrial translation. Also, mitochondrial mRNAs are uncapped and lack the poly(A) tail, as prokaryotic mRNAs. However, mitochondrial translation has unique features, for example, the use of an alternative genetic code and the composition of mitochondrial ribosomes. These are made up of nuclear-encoded proteins (with the exception of Var1), the majority of which have no recognizable homology to known proteins, that is, are unique. Some of these proteins have more specialized functions connected with the coupling of the translation process to the inner mitochondrial membrane, where the seven hydrophobic mtDNA-encoded proteins have to be assembled.

The 5'-untranslated leader sequences (UTLs) of mitochondrial mRNAs contain sequence or structural signals that indicate the translation initiation site and that are recognized by mRNA-specific activator proteins (e.g., Cbp1 for *cytb*, Pet309 for *cox1*, Pet 111 for *cox2*, and Pet494p for *Cox3* mRNA). Additional general factors are involved in translation initiation, elongation, and termination. Some of the translational activators, which mediate mRNA interactions with mitochondrial ribosomes, are organized on the surface of the inner mitochondrial membrane in a way that facilitates assembly of the inner membrane complexes. Assembly of these complexes necessitates also other proteins, for example, Oxa1p ensures co-translational insertion of the nascent mitochondrial-encoded polypeptides into the membrane.

In this way, synthesis of the mitochondrially encoded proteins is coordinated with their assembly into multimeric respiratory complexes. As the latter are made up also of nuclear-encoded subunits, translation of polypeptides from both origins and their assembly into complexes must be topologically coordinated. This suggests that a process of co-translation occurs where the inner and outer membranes

are in intimate association (contact sites): actually, it has been observed that ribosomes interact with yeast mitochondria at outer-inner membrane junctions.

See also: **Bioenergetics:** Bioenergetics and the Mitochondrial Genome; Mitochondrial DNA; **Molecular Biology:** Messenger RNA Degradation in Bacteria; mRNA Polyadenylation in Eukaryotes; Pre-tRNA and Pre-rRNA Processing in Bacteria; **Signaling:** Adrenergic Receptors.

Further Reading

- Barrientos A, Gouget K, Horn D, Soto IC, and Fontanesi F (2009) Suppression mechanisms of COX assembly defects in yeast and human: Insights into the COX assembly process. *Biochimica et Biophysica Acta* 1793: 97–107.
- Belfort M (2003) Two for the price of one: A bifunctional intron-encoded DNA endonuclease–RNA maturase. *Genes and Development* 17: 2860–2863.
- Bonnefoy N, Fiumera HL, Dujardin G, and Fox TD (2009) Roles of Oxa1-related inner-membrane translocases in assembly of respiratory chain complexes. *Biochimica et Biophysica Acta* 1793: 60–70.
- Chen XJ and Clark-Walker GD (2000) The petite mutation in yeasts: 50 years on. *International Review of Cytology* 194: 197–238.
- Contamine V and Picard M (2000) Maintenance and integrity of the mitochondrial genome: A plethora of nuclear genes in the budding yeast. *Microbiology and Molecular Biology Reviews* 64: 281–315.
- Dieckmann CL and Staples RR (1994) Regulation of mitochondrial gene expression in *Saccharomyces cerevisiae*. *International Review of Cytology* 152: 145–181.
- Dujon B, Colleaux L, Jacquier A, Michel F, and Montellier C (1986) Mitochondrial introns as mobile genetic elements: The role of intron-encoded proteins. *Basic Life Sciences* 40: 5–27.
- Gelly JC, Orgeur M, Jacq C, and Lelandaïs G (2011) MitoGenesisDB: An expression data mining tool to explore spatio-temporal dynamics of mitochondrial biogenesis. *Nucleic Acids Research* 39: D1079–D1084.
- Lazowska J, Jacq C, and Slonimski PP (1980) Sequence of introns and flanking exons in wild-type and box3 mutants of cytochrome *b* reveals an interlaced splicing protein coded by an intron. *Cell* 22: 333–348.
- Lipinski KA, Kaniak-Golik A, and Golik P (2010) Maintenance and expression of the *S. cerevisiae* mitochondrial genome – from genetics to evolution and systems biology. *Biochimica et Biophysica Acta* 1797: 1086–1098.
- Michel F, Costa M, and Westhof E (2009) The ribozyme core of group II introns: A structure in want of partners. *Trends in Biochemical Sciences* 34: 189–199.
- Michel F and Ferat JL (1995) Structure and activities of group II introns. *Annual Review of Biochemistry* 64: 435–461.
- Neupert W and Herrmann JM (2007) Translocation of proteins into mitochondria. *Annual Review of Biochemistry* 76: 723–749.
- Saint-Georges Y, Garcia M, Delaveau T, et al. (2008) Yeast mitochondrial biogenesis: A role for the PUF RNA-binding protein Puf3p in mRNA localization. *PLoS ONE* 3: e2293.
- Zee JM and Glerum DM (2006) Defects in cytochrome oxidase assembly in humans: Lessons from yeast. *Biochemistry and Cell Biology* 84: 859–869.

Mitochondrial Membranes, Structural Organization

C A Mannella, New York State Department of Health, Albany, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Cristae The complex internal compartments of mitochondria that are formed by invagination of the inner membrane.

Crista junctions Tubular regions of the inner membrane that connect the internal cristae compartments to the region of the membrane along the mitochondrial periphery.

Electron tomography A technique involving computation of a three-dimensional density map from scores of images recorded in an electron microscope from a serially tilted specimen.

Mitochondria The organelle in eukaryotic cells, composed of an outer and inner membrane, that uses the free energy of substrate oxidation for adenosine triphosphate (ATP) production.

Mitochondrial Structure

The mitochondrion is the organelle in eukaryotic cells that is the primary generator of adenosine triphosphate (ATP), the molecule that fuels most of the cell's machinery. The chemiosmotic mechanism of ATP production involves establishment of an electrochemical gradient across the mitochondrial inner membrane by the respiratory chain complexes. The free energy in this gradient is coupled to phosphorylation of adenosine diphosphate (ADP) by the F_1F_0 ATP synthase on the inner membrane. Since chemiosmosis entails very rapid fluxes of ions and metabolites between the mitochondrial interior and the cytosol, the shape as well as the permeability properties of the energy-transducing inner membrane can be expected to affect mitochondrial function.

Electron Microscopy and the Structure of Mitochondria

Pioneering work of G Palade, F Sjostrand, and many others in the 1950s led to the first high-quality electron microscopic images of mitochondria. Controversies ensued, partly due to the inherent difficulty of interpreting images of plastic sections that were too thin (80 nm) to convey three-dimensional (3D) structure. Conversely, the section thickness imposed a lower limit on resolution in the axial direction (parallel to the electron beam) in these images. Thus, despite the general acceptance of a model by the early 1960s (Figure 1), determining the true morphology of mitochondrial membranes in 3D detail was a work in progress for several decades. An important advance was the development and maturation in the 1990s of electron tomography, which uses numerous micrographs of a serially tilted specimen to generate a 3D image (tomogram). Using high-voltage electron microscopes, this technique has been used to provide actual 3D images of mitochondria in thick plastic sections (200–1000 nm) with axial resolution on the order of 5–10 nm (Figure 2). Moreover, with computer-controlled cryo-electron microscopes, it has been possible to obtain 3D images of mitochondria in near-native, vitreously frozen state, without the chemical fixatives and metal stains employed with plastic sections. Cryo-tomograms have been obtained for intact, isolated mitochondria (Figure 3),

mitochondria in thin regions of culture cells, and mitochondria in mammalian tissue.

Mitochondrial Membranes and Compartments

The mitochondrion is composed of two nested membranes, which define three classes of compartments.

Outer membrane

The mitochondrial outer membrane is the interface between the organelle and the cytosol. It excludes proteins and other macromolecules, but is normally permeable to solutes below a molecular weight of ~2000 since it carries many copies of a pore-forming protein called 'voltage-dependent, anion selective channel (VDAC)' or 'mitochondrial porin'.

Inner membrane

The inner membrane is the site of the respiratory chain, the ATP synthase, and numerous carrier proteins such as the adenine nucleotide transporter (ANT) that exchanges ADP and ATP.

Intermembrane space

The space between the outer and inner membranes is called the 'intermembrane (or peripheral) space'. Its protein content, which is much lower than that of the matrix, includes kinases (for interconversion of phosphorylated metabolites) and cytochrome *c* (involved in both electron transfer and apoptosis).

Matrix space

The space enclosed by the inner membrane is occupied by a very dense, largely protein matrix. This compartment contains a variety of enzymes, particularly those of the Krebs cycle, along with several copies of the circular mitochondrial genome and numerous mitochondrial ribosomes. (Mitochondria synthesize a small number of respiratory chain proteins; 90% of the mitochondrial protein mass is imported from the cytosol.)

Cristae

The mitochondrial inner membrane has a much larger surface area than the outer membrane, typically by a factor of 2.5–10. There is a rough correlation between the energy demand in a

particular tissue and the inner membrane surface, for example, heart mitochondria have more inner membrane (and so more ATP-generating capacity) than liver mitochondria. The inner membrane is contained by the smaller outer membrane because the inner membrane is invaginated or involuted. The internal compartments defined by the infoldings of the inner membrane are called 'cristae' and the space they enclose is the intracristal space. The intracristal space connects to the intermembrane space through the openings of the cristae in the peripheral region of the inner membrane.

Mitochondrial Inner Membrane Topology

The early model of mitochondrial structure portrayed the cristae as baffle-like folds in the inner membrane (**Figure 1**). 3D images provided by electron tomography of a wide variety of mitochondria, isolated and *in situ*, indicate a fundamentally different internal architecture. The cristae, in fact, are pleiomorphic compartments that connect to each other and to the peripheral (boundary) region of the inner membrane via narrow tubular segments (**Figure 2**). These connecting regions (crista junctions) can vary in length from a few tens to several

hundred nanometers. The shape of the connecting segments can vary from narrow cylinders with openings ~ 20 nm in diameter, to slot-like with openings of 40–100 nm or wider.

Functional Implications of the Crista Junctions

The fact that cristae open into the intermembrane space through narrow, sometimes very long tubular segments suggests that diffusion of solutes between the internal compartments of mitochondria might be restricted. This was not a concern with the early structural model, which depicted cristae as baffles with very wide openings into the intermembrane compartment. Computer simulations indicate that diffusion of ADP between the intermembrane and intracristal compartments via long, narrow tubes can cause depletion of ADP in the intracristal space and locally diminished ADP phosphorylation. Thus, the topology of the mitochondrial inner membrane may directly influence the efficiency of ATP generation by the organelle.

Cristae Junctions Form Spontaneously

Tubular crista junctions appear to be a highly conserved feature of the mitochondrial inner membrane. If they are functionally important, it is expected that their formation would be spontaneous or energetically favored. This has been demonstrated with yeast mitochondria. Starting with inner membranes unfolded by osmotic swelling, tubular junctions between cristae and the peripheral region of the inner membrane re-form over a matter of minutes following addition of osmoticant. Interestingly, the inner membranes of relic mitochondria in protozoa that lack oxidative phosphorylation fold randomly and lack tubular connecting segments.

Dynamics of the Mitochondrial Inner Membrane

Although the mitochondrial inner membrane has a high protein-to-lipid ratio and encases a dense protein matrix, it is

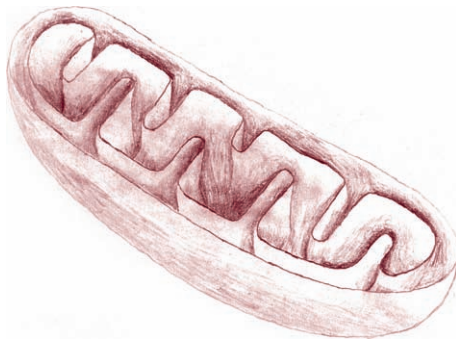


Figure 1 The early textbook model for mitochondrial structure. From Philomena Sculco, Wadsworth Center.

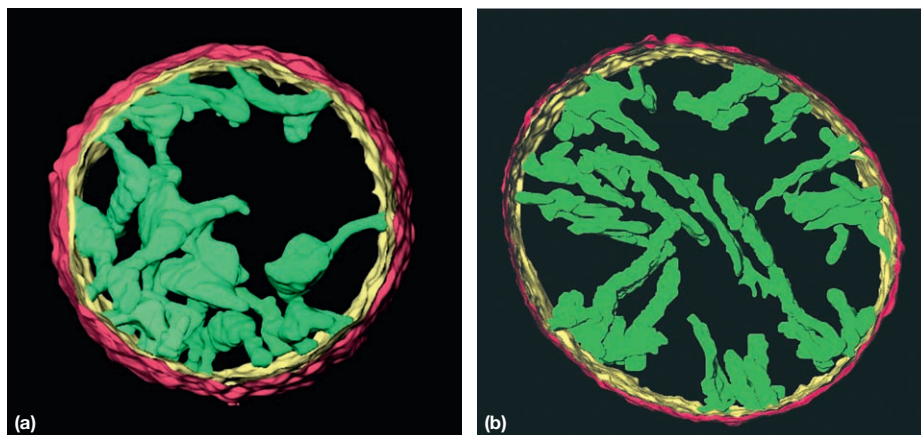


Figure 2 Tomographic reconstruction of membranes in isolated rat liver mitochondria in (a) condensed and (b) orthodox conformations. Mitochondria were chemically fixed, plastic embedded and stained by conventional procedures. Approximate dimensions: (a) 1500 nm diameter, 500 nm thickness, and (b) 500 nm diameter, 200 nm thickness. (a) Adapted from **figure 1(a)** in Mannella CA, Pfeiffer DR, Bradshaw PC, et al. (2001) Topology of the mitochondrial inner membrane: Dynamics and bioenergetic implications. *IUBMB Life* 52: 93–100.



Figure 3 Tomographic reconstruction of membranes in an intact, frozen-hydrated rat liver mitochondrion, prepared by rapid plunging into cryogen with no chemical fixatives or stains. Approximate dimensions: 700 nm diameter, 450 nm thickness. Adapted from figure 4(b) in Mannella CA, Pfeiffer DR, Bradshaw PC, et al. (2001) Topology of the mitochondrial inner membrane: Dynamics and bioenergetic implications. *IUBMB Life* 52: 93–100.

very plastic. The morphology of this membrane in isolated mitochondria changes readily in response to osmotic and metabolic conditions.

Osmotic Adjustments in Inner Membrane Topology

H Tedeschi and D Harris showed in the 1950s that isolated mitochondria behave as perfect osmometers over a range of osmotic conditions. The matrix volume adjusts to changes in osmotic conditions (external solute concentration) by expanding (water influx) or shrinking (water efflux). Isolated mitochondria with a shrunken matrix (caused by water efflux in response to hyper-osmotic external conditions) are said to be in the condensed conformational state (Figure 2(a)). Mitochondria with a more expanded matrix (exposed to iso-osmotic or slightly hypo-osmotic conditions) are said to be in the orthodox (cell-like) state (Figure 2(b)). Exposure of mitochondria to very hypo-osmotic media causes extreme matrix swelling and eventual rupture of the outer membrane. (The latter is considered osmotically inactive due to its permeability to small solutes.)

Changes in Inner Membrane Topology Associated with Metabolic State

In the 1960s, C Hackenbrock demonstrated that isolated mitochondria could reversibly switch between orthodox and condensed conformations during respiration. When ADP is added, respiring mitochondria change from orthodox to condensed conformation, then back to orthodox as ADP runs out. Condensed mitochondria tend to have larger internal

compartments with longer cristae junctions, which can limit ADP diffusion into the cristae. Thus, the switch to orthodox conformation (smaller internal compartments, shorter tubular junctions) as ADP levels drop may prevent intracristal depletion of ADP and so maintain more efficient ATP production. Note that cellular ADP levels are normally low and mitochondria *in situ* are usually in the orthodox conformation.

Inner Membrane Fusion and Fission

Comparison of 3D images of condensed and orthodox mitochondria suggests an important role for membrane fusion and fission in regulation of inner membrane topology. Condensed mitochondria have large cristae with multiple tubular connections (crista junctions) at the periphery of the inner membrane (Figure 2(a)). Conversely, orthodox mitochondria have predominantly smaller cristae with fewer peripheral connections (Figure 2(b)). Interconversion of these two membrane topologies cannot be achieved by simple unfolding and refolding of the inner membrane. The topologies can be reconciled by postulating that large cristae with multiple tubular segments form by fusion of individual smaller cristae that make contact during matrix contraction. Likewise, matrix expansion and associated lateral stress on the inner membrane may trigger fission of larger cristae into individual tubular membranes. Note that, in tomographic images of frozen-hydrated mitochondria, large cristae clearly appear to be composed of fused tubular membranes (Figure 3). Mitochondria contain several classes of proteins that regulate outer and inner membrane dynamics, curvature and tethering during fusion and fission of the organelles. One is the dynamin-like protein Opa-1 (called ‘Mgm-1’ in yeast) which has been shown to be required not only for mitochondrial fusion but also for maintaining normal crista structure. Thus, crista morphology likely reflects a balance between inner membrane fusion and fission regulated by Opa-1 and similar proteins.

Role of ATP Synthase in Shaping the Inner Membrane

ATP synthase is a multimeric inner-membrane motor that uses the free energy stored in the proton gradient across the inner membrane to drive phosphorylation of ADP. ATP synthase has the property of forming dimers, which can associate laterally to form dimer ribbons. Mutation or knockdown of proteins involved in ATP synthase dimerization result in mitochondria that lack normal cristae with tubular junctions. High-resolution electron microscopy of isolated ATP synthase dimers and inner membrane fragments suggests that the dimer ribbons impart the curvature in the inner membrane responsible for forming narrow tubular regions, first proposed by R Allen using scanning electron microscopy.

Functional Changes in Inner Membrane Topology

In functionally intact mitochondria, inner membrane shape is not random. It is regulated by proteins, such as Opa-1 and ATP synthase subunits, and it responds to metabolic state. Several other examples of function-linked changes in the shape

of the mitochondrial inner membrane are described below. The emerging hypothesis is that mitochondrial inner membrane topology is a parameter regulated by the cell to optimize mitochondrial function in response to metabolic signals and stresses.

Crista Dilation and Oxidative Stress

Inner membranes of mitochondria in steroidogenic tissues like adrenal glands undergo a rapid transition from lamellar compartments (with typical crista junctions) to parallel arrays of dilated tubular membranes (2–3 times wider than the junctions) upon activation of the P450 hydroxylation system. A similar crista transition to extended dilated tubes is observed in flight muscle mitochondria of *Drosophila* after exposure to hyperoxic conditions. An even more drastic transition is induced by fasting in the amoeba, *Chaos carolinensis*, with the mitochondrial inner membrane converting from narrow tubes to an interconnected labyrinth of wide internal compartments displaying cubic crystalline order. These three very disparate systems have two things in common. In each case, membrane remodeling is associated with increased free radical production (oxidative stress), and the resulting membrane topology would ameliorate potential damage to the mitochondria. Wider intracristal compartments enhance internal diffusion and efflux of toxic radicals from the mitochondrial interior, while increasing the radius of curvature of membranes inhibits oxidation of lipid acyl chains by oxygen radicals.

Cytochrome *c* Release during Apoptosis

The process of apoptosis, or programmed cell death, in many cell types requires release of mitochondrial factors, including cytochrome *c*, to activate caspases involved in cellular degradative reactions. This release of mitochondrial proteins involves members of the Bcl-2 family of proteins known to regulate apoptosis, as described by S Korsmeyer and others. Truncated Bid (tBid) triggers migration of Bax from the cytosol to the mitochondrial surface, where it oligomerizes and creates large pores (by an as yet undefined mechanism) through which the proteins diffuse. Mitochondrial fragmentation (fission) occurs later in apoptosis, associated with loss of Opa-1. Not only does tBid trigger outer membrane permeabilization, but it also induces (in isolated liver mitochondria) a dramatic remodeling of the topology of the inner membrane that correlates with increased accessibility of cytochrome *c* to the intermembrane space and its more complete release. Inner-membrane curvature is locally reversed and crista junctions widen into long slots, which could expedite release of cytochrome *c* from the cristae. The topological change reflects a fundamental alteration in molecular organization of the inner membrane induced by interaction of tBid with cardiolipin. This interaction also reduces low-affinity cytochrome *c* binding to the

inner membrane surface, which would also enhance cytochrome *c* mobilization and release.

See also: **Bioenergetics:** ATP in Plant Mitochondria: Substrates, Inhibitors, and Uncouplers; ATP Synthesis: Mitochondrial Cyanide-Resistant Terminal Oxidases; Bioenergetics and the Mitochondrial Genome; Cell Death by Apoptosis and Necrosis; Cytochrome *c*; Mitochondrial DNA; Mitochondrial Outer Membrane and the VDAC Channel; Nuclear Genes Involved in Mitochondrial Function and Biogenesis; Protein Import into Mitochondria; **Molecular Biology:** DNA Replication, Mitochondrial.

Further Reading

- Davies KM, Strauss M, Daum B, et al. (2011) Macromolecular organization of ATP synthase and complex 1 in whole mitochondria. *Proceedings of the National Academy of Sciences of the United States of America* 108: 14121–14126.
- Frey TG and Mannella CA (2000) The internal structure of mitochondria. *Trends in Biochemical Sciences* 25: 319–324.
- Frey TG, Perkins GA, and Ellisman MH (2006) Electron tomography of membrane-bound cellular organelles. *Annual Reviews of Biophysics and Biomolecular Structure* 35: 199–224.
- Heath-Engel HM and Shore GC (2006) Mitochondrial membrane dynamics, crista remodelling and apoptosis. *Biochimica et Biophysica Acta* 1763: 549–560.
- Hsieh CE, Marko M, Leith A, Mannella CA, and Frank J (2006) Towards high-resolution three-dimensional imaging of native mammalian tissue: Electron tomography of frozen-hydrated rat liver sections. *Journal of Structural Biology* 153: 1–13.
- Mannella CA (2008) Structural diversity of mitochondria: Functional implications. *Annals of the New York Academy of Sciences* 1147: 171–179.
- Mannella CA, Marko M, Penczek P, Barnard D, and Frank J (1994) Internal compartmentation of rat-liver mitochondria: Tomographic study using the high-voltage transmission electron microscope. *Microscopy Research and Technique* 27: 278–283.
- Mannella CA, Pfeiffer DR, Bradshaw PC, et al. (2001) Topology of the mitochondrial inner membrane: Dynamics and bioenergetic implications. *IUBMB Life* 52: 93–100.
- Meeusen S, DeVay R, Block J, et al. (2006) Mitochondrial inner-membrane fusion and crista maintenance requires the dynamin-related GTPase Mgm1. *Cell* 127: 383–395.
- Munn EA (1974) *The Structure of Mitochondria*. London: Academic Press.
- Nicasro D, Frangakis AS, Typke D, and Baumeister W (2000) Cryo-electron tomography of *Neurospora* mitochondria. *Journal of Structural Biology* 129: 48–56.
- Paumard P, Vallier J, Couly B, et al. (2002) The ATP synthase is involved in generating mitochondrial cristae morphology. *EMBO Journal* 21: 221–230.
- Rasmussen N (1995) Mitochondrial structure and the practice of cell biology in the 1950s. *Journal of the History of Biology* 28: 381–429.
- Scorrano L, Ashiya M, Buttle K, et al. (2002) A distinct pathway remodels mitochondrial cristae and mobilizes cytochrome *c* during apoptosis. *Developmental Cell* 2: 55–67.
- van Driel LF, Valentijn JA, Valentijn KM, Koning RI, and Koster AJ (2009) Tools for correlative cryo-fluorescence microscopy and cryo-electron tomography applied to whole mitochondria in human endothelial cells. *European Journal of Cell Biology* 88: 669–684.

Relevant Websites

- <http://www.wadsworth.org> – EM Animation Gallery: Frozen-Hydrated Rat Liver Mitochondrion.
- <http://www.sci.sdsu.edu> – Mitochondria Reconstructed by Electron Tomography.

Mitochondrial Metabolite Carrier Protein Family

F Palmieri, University of Bari, Bari, Italy

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Ferdinando Palmieri and Martin Klingenberg, volume 2, pp. 725–732, © 2004, Elsevier Inc.

Glossary

Electrogenic, electrophoretic transport The movement of a net electrical charge across the membrane coupled to the transport of metabolites. It is driven by the membrane potential.

Exchange carriers Carriers that catalyze transport only by counter-exchange of solutes across the membrane.

Mitochondrial carriers Proteins mostly located in the inner mitochondrial membrane which catalyze transport of metabolites between cytosol and mitochondrial matrix space.

Introduction

Mitochondrial carriers (MCs) are a large family of membrane-embedded proteins that are mostly localized in the inner membrane of mitochondria. Given that this membrane is highly impermeable and only some uncharged molecules, such as O₂ and CO₂, pass through in a non-protein-mediated manner, MCs provide a link between metabolic reactions occurring in the cytosol and the mitochondrial matrix by catalyzing the translocation of numerous solutes across the membrane. Because mitochondria and cytosol cooperate in a large number of metabolic cellular processes due to enzyme compartmentalization, a continuous and highly diversified flux of solutes into and out of mitochondria takes place across the inner membrane. Indeed, MCs are involved in many important metabolic pathways such as oxidative phosphorylation, the citric acid cycle, fatty acid oxidation, amino acid degradation, gluconeogenesis, lipogenesis, transfer of reducing equivalents, urea synthesis, replication and repair of mitochondrial DNA, mitochondrial RNA and protein synthesis, heme biosynthesis, and ketone body production and utilization (Figure 1).

All mitochondrial proteins transporting metabolites (i.e., MCs) share common structural features and belong to a protein family known as the mitochondrial carrier family (MCF). This family, which is probably the largest of the solute carrier families, can be subdivided into groups according to their substrate specificity: carriers for nucleotides, carboxylates and keto acids, amino acids and other substrates; each group can be subdivided into several subfamilies. With a few exceptions, these subfamilies are widespread in the eukaryotic kingdom indicating that they play a role that has been conserved throughout the evolutionary process. The substrates transported by the MCF vary widely not only in their structure but also in their size from the smallest, H⁺, to the largest and most highly charged solutes translocated across membranes, such as nicotinamide adenine dinucleotide (NAD⁺) or coenzyme A. Most of the substrates are negatively charged, but some are positively charged or zwitterions at physiological pH values. Several MCs have isoforms encoded by different genes and others, like the human phosphate carrier, have isoforms generated by alternative splicing. The level of MC expression is greatly variable in unicellular organisms depending on the metabolic environmental conditions. In animalia and plants some MCs,

or one of their isoforms, if any, are widely distributed in all or nearly all tissues, whereas others are tissue specific and their limited distribution reflects their importance in special functions. Furthermore, given that MCs are nuclear-coded proteins they must be imported into the inner mitochondrial membrane. However, unlike most nuclear-coded mitochondrial proteins, MCF members contain targeting information in their mature sequence. Indeed, only a minority of MCs possess cleavable presequences which are not essential for targeting and probably fulfill other functions, that is, provide a binding site for chaperone heat-shock cognate Hsc70 (the phosphate carrier) or act as an internal chaperone (citrate carrier). Finally, the MCF is not restricted to mitochondria. Some MCs are localized in other cell organelles, such as peroxisomes and chloroplasts.

The aim of this concise review is to provide a comprehensive reference source of current knowledge on the structure, function, and evolution of the MCs as well as on their associated diseases.

Structure

All MCF members of known function have common structural characteristics which are different from those of any other transporter family. A unique feature of all MCs is their tripartite structure, that is, they consist of three tandemly repeated homologous domains about 100 amino acids in length. Each repeat is comprised of two transmembrane α -helices connected by an extensive polar loop exposed on the matrix side of the inner mitochondrial membrane, and the three repeats are linked by two shorter loops exposed on the cytosolic side. Therefore, a typical MC has six α -helices traversing the membrane linked by hydrophilic loops with both the N- and C-termini on the cytosolic side. Furthermore, each repeat contains a signature sequence motif PX[D/E]XX[K/R]X[K/R] (20–30 residues) [D/E]GXXXX[W/Y/F][K/R]G. Unlike the majority of the MCF members, some MCs have long extensions at the N-terminus. In particular, the Ca²⁺-binding aspartate/glutamate and adenosine triphosphate (ATP)-Mg/Pi subfamilies have an extensive N-terminal domain (more than 150 amino acids), which usually contains three or four EF-hand Ca²⁺-binding motifs and fulfills a regulatory function.

Until now, the only three-dimensional (3D) MC structure is that of the adenosine diphosphate (ADP)/ATP carrier in

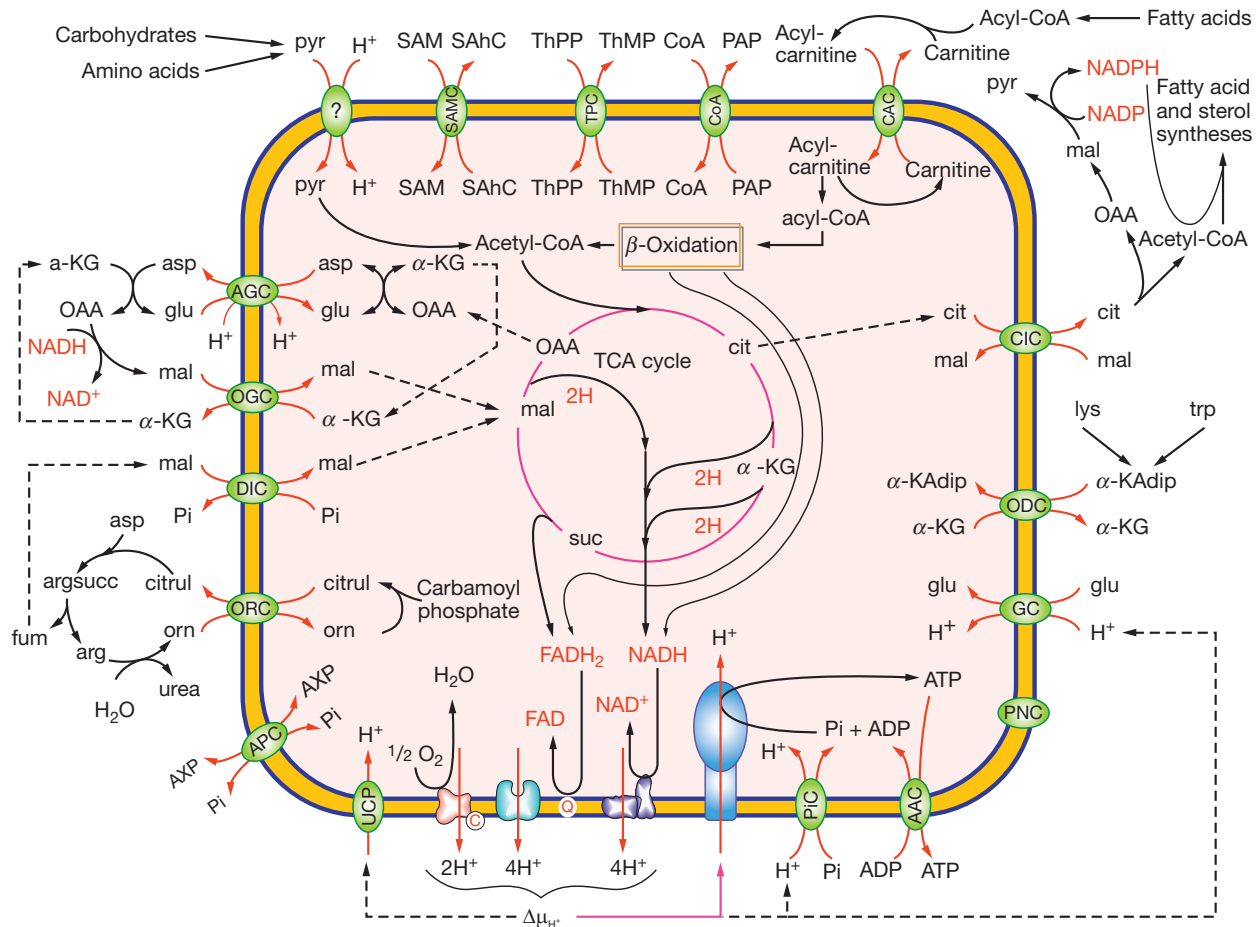


Figure 1 Metabolic roles of mitochondrial carriers. The scheme does not include all of the carriers listed in Table 1 and does not show all the metabolic pathways in which individual carriers are involved (see Table 1). Abbreviations as in Table 1.

complex with its powerful inhibitor carboxyatractyloside. Basically, this structure is composed of six transmembrane α -helices (H1–H6) and three short α -helices (h12, h34, and h56) located on the matrix side and parallel to the membrane plane. H1–H6 line a funnel-shaped cavity (occupied by the inhibitor) which is open toward the cytosol and closed on the matrix side by a salt-bridge network formed by the charged residues of the first part of the three signature motifs, PX[D/E]XX[R/K]. This structure (1) exhibits a threefold pseudo-symmetry in line with the threefold sequence repeats; (2) roughly corresponds to the 'c' (cytosolic)-state of the ADP/ATP carrier (with the cavity open toward the cytosol), as carboxyatractyloside is a high-affinity ligand that binds the carrier from the cytosolic side of the membrane; and (3) based on the significant sequence conservation in the MCF, most likely represents the common folding of all the other MCF members and has therefore been used as a template for building homology models of various carriers. Furthermore, the ion-pair network mentioned above constitutes the gate of MCs on the matrix side (Figure 2(a)).

Recently the MC threefold pseudo-symmetry has been further exploited principally by Dr. E. Kunji and co-workers. The analysis of inter-repeat multiple sequence alignment of MCs of known function (substrate specificity) highlighted

symmetry-related triplets formed by the aligned residues of each carrier (see Figure 3). These triplets can be either symmetric, consisting of three identical amino acids, or asymmetric, consisting of different amino acids. Symmetric triplets are important for structure and/or the transport mechanism, while asymmetric triplets are important for recognition and/or binding of the substrate (which usually has an asymmetric distribution of functional groups). Typical symmetric triplets (30 and 33 in Figure 3) include the negatively and positively charged residues involved in the matrix salt-bridge network or matrix gate (formed by the charged residues of the sequence motif PX[D/E]XX[K/R] localized at the C-terminus of the odd-numbered transmembrane α -helices). Other symmetric triplets (93 and 96 in Figure 3), though less conserved, are the charged residues of the [F/Y][D/E]XX[R/K] motif localized at the C-terminus of the even-numbered transmembrane α -helices. These residues would form a salt-bridge network on the cytosolic side which would close the carrier in the 'm' (matrix)-state (with the cavity open toward the matrix) constituting the gate of MCs on the cytosolic side (Figure 2(b)). Typical asymmetric triplets (80 and 81 in Figure 3) are the two residues of the three even-numbered transmembrane α -helices that have been proposed to form the contact points I, II, and III of MCs for their substrates. Based on the 3D structure of the ADP/ATP

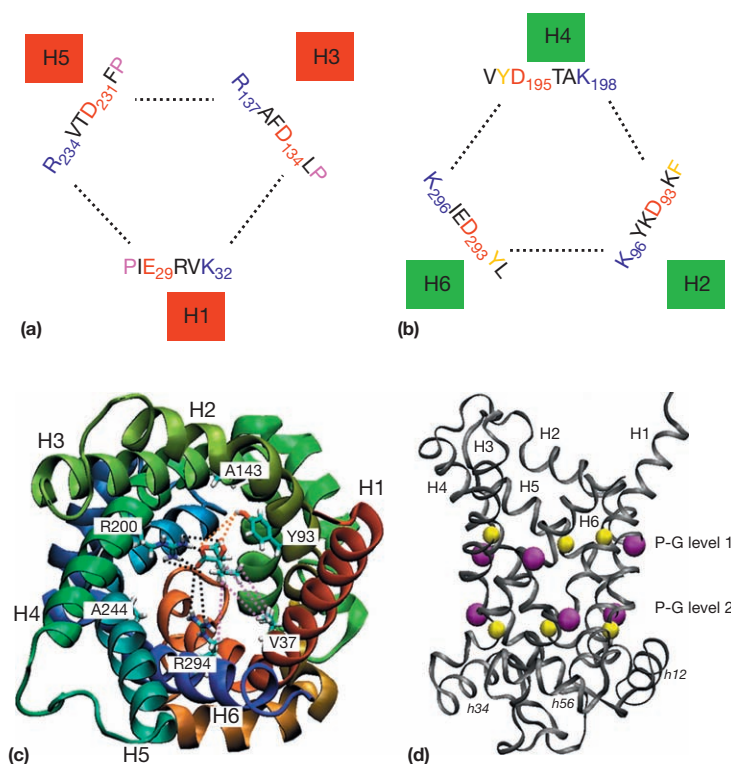


Figure 2 Structural properties of MCs. (a) Matrix salt-bridge network constituting the matrix gate of MCs. The salt bridges are formed by the positive and negative residues of the signature motifs PX(DE)XX(RK) located at the C-terminus of the three odd transmembrane α -helices. This gate closing the carrier cavity on the matrix side is documented by the 3D structure of the carboxyatractyloside-inhibited ADP/ATP carrier. (b) Cytosolic salt-bridge network constituting the cytosolic gate of MCs. The salt bridges are formed by the charged residues of the motifs [F/Y][D/E]XX[R/K] located at the C-terminus of the three even transmembrane α -helices. This network closing the carrier cavity on the cytosolic side has been hypothesized by Dr. E.R.S. Kunji and co-workers. (c) Top view from the cytosolic side of the structural homology model of the yeast oxaloacetate/sulfate carrier (Oac1p) highlighting the carrier substrate-binding site. Transmembrane α -helices are colored as follows: H1, red; H2, tan; H3, green; H4, emerald-green; H5, cyan; and H6, blue. The residues (V37, Y93, A143, R200, A244, and R294) of the six transmembrane α -helices protruding into the cavity at the height of the binding site and the substrate α -isopropylmalate are shown in stick representation using the VMD color code: cyan, C-atoms; red, O-atoms; white, H-atoms; and blue, N-atoms. Possible interactions between the substrate and residues (V37, Y93, R200, and R294) are highlighted as follows: electrostatic/polar (black dashed lines), hydrophobic (purple dashed lines) and via bridging water (orange dashed lines). (d) Lateral view of the structural homology model of human aspartate/glutamate carrier (isoform 1) highlighting the position of the conserved prolines and glycines in the even and odd transmembrane α -helices. The triplet PPP of the odd helices (no. 28 in Figure 3) and the triplet PPP of the even helices (no. 83 in Figure 3) are in purple; the triplet GGG of the odd helices (no. 19 in Figure 3) and the triplet GGG of the even helices (no. 73 in Figure 3) are in yellow. (c) Reproduced from Marobbio CMT, Giannuzzi G, Paradies E, Pierri CL, and Palmieri F (2008) α -Isopropylmalate, a leucine biosynthesis intermediate in yeast, is transported by the mitochondrial oxaloacetate carrier. *Journal of Biological Chemistry* 283: 28445–28453, with permission from The American Society for Biochemistry and Molecular Biology.

carrier, these residues protrude into the cavity of the carrier at the midpoint of the membrane one-and-a-half helix turns above the matrix salt-bridge network that closes the bottom of the cavity (Figure 2(c)). Among the three proposed contact points, point II on helix 4 is the most important because it alone can discriminate between amino acids, carboxylic acids (including keto acids), and nucleotides. In fact, point II is defined by G[IVLMT] in nucleotide carriers, by R[QHNTAV] in carboxylic acid carriers and by R[D/E] in amino acid carriers. The above-mentioned contact points constitute the common and similarly located substrate-binding site of MCs or part of it when, depending on the size, shape, and chemistry of the substrate, residues of the odd-numbered transmembrane α -helices located in the cavity at the same level and/or other residues above and below are also involved in substrate binding (Figure 2(c)). Other asymmetric triplets are, for example,

84 and 88 in the ATP-Mg/Pi carrier subfamily and 84, 85, and 88 in the aspartate/glutamate and glutamate carrier subfamilies (Figure 3).

Finally, the analysis of inter-helix multiple sequence alignment revealed first the presence of a well-conserved glycine nine residues before the prolines of the PX[D/E]XX[K/R][X/K/R] sequence motif in the odd transmembrane helices; and second, the presence of a conserved proline in the even transmembrane helices 10 residues after the glycine corresponding to the last residue of the second part ([D/E]GXXXX[W/Y/F][R/K]G) of the sequence motif (Figure 2(d)). Interestingly, the above-mentioned Gly and Pro of the odd helices are aligned with the Gly and Pro of the even helices in an antiparallel fashion. Moreover, these Pro and Gly residues are located strategically between the substrate-binding site and the gates on both sides. As documented for the Pro of the odd helices by the 3D

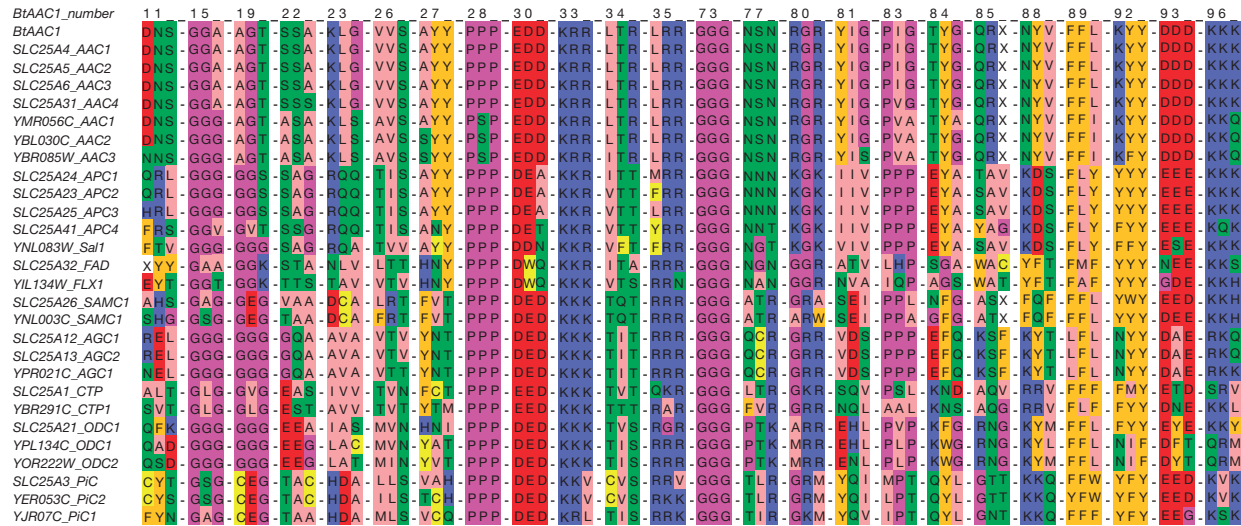


Figure 3 Alignment of symmetry-related amino acid triplets of representative carriers for nucleotides, amino acids and acids in *Homo sapiens*, *Saccharomyces cerevisiae*, and *Arabidopsis thaliana*. Each triplet of a single carrier is formed by the three aligned residues derived from the inter-repeat multiple sequence alignment. The triplets are ordered horizontally according to the number of the first-repeat amino acids of the bovine ADP/ATP carrier sequence (NP_777083). Only symmetry-related triplets of residues protruding into the carrier cavity are shown with the exception of triplets 19, 28, 73, and 83. ADP/ATP carriers (AAC), ATP-Mg/Pi carriers (APC), *S. cerevisiae* ATP-Mg/Pi carrier (Sal1), human FAD/folate carrier (FAD), *S. cerevisiae* FAD carrier (FLX1), *S*-adenosylmethionine carriers (SAMC), aspartate/glutamate carriers (AGC), citrate transporter proteins (CTP), oxodicarboxylate carriers (ODC), and phosphate carriers (PIC) are reported. Amino acids are colored according to the default Jalview-Zappo style.

structure of the ADP/ATP carrier, it has been proposed that the Gly of the odd helices and the Gly and Pro of the even helices can also act as hinges to open or close the carrier gate on the matrix or cytosolic side (see section **Mechanism of Transport**).

Extension of the MCF

The MCF is typified by the sequence features of its members: a tripartite structure, a threefold repeated signature motif, and six transmembrane α -helices (two in each of the three repeats) separated by hydrophilic loops. The term MCF was established after identifying its first six members, which had been purified from mitochondria. Indeed, the ADP/ATP carrier, the uncoupling protein and the carriers for phosphate, citrate, oxoglutarate/malate, and carnitine/acylcarnitines were sequenced by Edman degradation or by molecular biology techniques based on minimal protein sequence information. In the genomic era, searching genome databases using the sequence features of the MCF has revealed a large number of MCF members. For example, the genome of *Saccharomyces cerevisiae* encodes 35 MCs, that of *Arabidopsis thaliana* 58, and the *Homo sapiens* genome 53.

Approximately half of the *H. sapiens* and *S. cerevisiae* MCs have been characterized in terms of substrate specificity and kinetic parameters, in particular by transport assays upon heterologous gene expression, purification, and reconstitution into liposomes. This gene-to-function strategy was first employed for the bacterial overproduction and functional reconstitution of the bovine oxoglutarate carrier, which was indeed the first eukaryotic membrane protein to be expressed in *Escherichia coli* and renatured. In addition, for some carriers, subcellular localization, tissue distribution, and their physiological role in

cell metabolism and specialized cell functions have also been determined. Until now, 23 subfamilies of MCs have been disclosed based on substrate specificity (Table 1). Among these, the uncoupling protein subfamily has been included in Table 1 for its relevance, although the substrate specificity of several of its members is yet to be determined. Furthermore, MC subfamilies are characterized by specific sets of symmetry-related triplets (Table 1 and Figure 3). For example, all the members of the ADP/ATP carrier subfamily exhibit triplets 11 (DNS), 19 (AGT), 23 (KL[G/S]), 84 (TYG), 85 (QRX), and 88 (NYV), as well as all the members of the aspartate/glutamate carrier subfamily triplets 22 (GQA), 77 (QCR), 84 (EFQ), 85 (KSF), and 88 (KYT). Of note, related subfamilies transporting structurally similar substrates may share some characterizing triplets as well as some substrates. For example, the glutamate and aspartate/glutamate carrier subfamilies share triplets 22 and 85. Finally, the existence of subfamily-specific sets of symmetry-related triplets (Table 1) offers the possibility of predicting the function (substrate specificity) of yet unidentified carriers in different organisms.

Evolution of MCs

In agreement with the large number of MC subfamilies already identified (see above), the phylogenetic tree of all the MCs of *A. thaliana*, *H. sapiens*, and *S. cerevisiae* presented in Figure 4 shows that the MCs of these three evolutionary-distant species (146 carriers altogether) cluster into many different clades, suggesting a large variety of specialized functions. Furthermore, with a few exceptions (e.g., the guanosine triphosphate/guanosine diphosphate (GTP/GDP) carrier subfamily present only in fungi, the di-/tri-carboxylate carrier subfamily only in plants and the aspartate/glutamate carrier subfamily in fungi

Table 1 Mitochondrial carrier subfamilies defined by substrate specificity and symmetry-related amino acid triplets

Mitochondrial carrier subfamilies	Main substrates	Main metabolic roles	Characterizing triplets
1. <i>For nucleotides/dinucleotides</i>			
ADP/ATP (AAC)	ADP, ATP	Oxidative phosphorylation, energy transfer, phosphate transfer	11 (DNS), 19 (AGT), 23 (KL[G/S]), 84 (TYG), 85 (QRX), 88 (NYV)
Coenzyme A/PAP (CoC)	CoA, PAP, dephospho-CoA, ADP, ATP	Krebs cycle, fatty-acid β -oxidation, heme biosynthesis, branched-chain amino acid catabolism, urea cycle (acetylglutamate synthesis), protein acetylation, mitochondrial fatty-acid synthase	23 (K[V/A]Q), 34 (IVR), 88 triplets ([K/Q]SS)
ATP-Mg/Pi (APC)	ATP-Mg, Pi, ADP, AMP	Modulation of the matrix adenine nucleotide content, regulation of mitochondrial nucleotide-dependent enzymes	23 (RQ[Q/A]), 30 (DE[A/T/N]), 84 (EYA), 88 (KDS)
Thiamine pyrophosphate (TPC)	ThPP, ThMP, dNDP, dNTP, ADP, ATP	Regulation of mitochondrial ThPP-dependent enzymes, mitochondrial DNA synthesis	23 (R[T/S]K), 34 (IT[K/R]), 80 (L[A/T]K), 85 (GAT)
Pyrimidine nucleotides (PNC)	Pyrimidine (deoxy)nucleotides	Mitochondrial DNA and RNA synthesis and catabolism	19 (G[G/A]K), 27 (CNY), 30 ([D/E]WE), 37 (QQR), 83 ([PEP], 85 (R[I/V][S/T]))
FAD/folate	FAD, folates	Mono-carbon unit donor, redox reactions	19 (GGK), 27 (HNY), 30 (DWQ)
Adenine nucleotides in peroxisomes (ANT)	ATP, ADP, AMP	Peroxisomal fatty-acid β -oxidation	19 (SAK), 30 (DAI), 33 (KAK), 37 (QKR)
NAD ⁺ (NDT)	NAD ⁺ , (d)AMP, (d)GMP	Redox reactions, redox balance of mitochondria and chloroplasts, regulation of NAD ⁺ -dependent metabolic pathways	19 (GGK), 27 (CNY), 30 (DWE), 89 (FP[L/F])
GTP/GDP (GGC)	GTP, GDP, dGTP, dGDP	Mitochondrial protein and RNA synthesis, DNA replication and repair	22 (EGS), 23 (IEL), 84 (Q GK), 85 (RSL), 88 (KLS)
2. <i>For di-/tri-carboxylates and keto acids</i>			
Citrate (CIC)	Citrate, malate, isocitrate, <i>cis</i> -aconitate, phosphoenolpyruvate	Fatty-acid and sterol biosynthesis, histone acetylation, gluconeogenesis from lactate, isocitrate (citrate)/oxoglutarate shuttle	22 (E[A/S][S/T]), 84 (KN[S/D]), 88 (RRV)
Dicarboxylates (DIC)	Malate, succinate, Pi, sulfate, thiosulfate	Krebs cycle, gluconeogenesis from pyruvate, urea synthesis, sulfur metabolism	26 (TG[C/S]), 27 (H[N/T][S/Q/N]), 33 (K[N/M]K), 88 (RQ[I/L/T])
Succinate/fumarate (SFC)	Succinate, fumarate	Gluconeogenesis	22 (EAG), 84 (KNG), 88 (RNT)
Oxoglutarate (OGC)	Oxoglutarate, malate	Malate/aspartate shuttle, isocitrate (citrate)/oxoglutarate shuttle, nitrogen metabolism, glucogenesis from lactate	26 (VGS), 27 (QTM), 33 (KLK), 35 (RRR), 77 (GTY), 84 (YVH), 88 (RQT), 93 (TSE)
Oxodicarboxylates (ODC)	Oxoadipate, oxoglutarate	Lysine, hydroxylysine and tryptophan catabolism (and synthesis in yeast)	26 (IGS), 27 (QSL), 33 (KLK), 77 (GTY), 84 (YLH), 88 (RMT)
Oxaloacetate/sulfate (OAC)	Oxaloacetate, sulfate, thiosulfate, α -isopropylmalate	Krebs cycle, sulfur metabolism, transfer of reducing equivalents, leucine synthesis in yeast	23 (VAA), 26 (TGM), 30 (E[F/Y]D), 80 (YRR), 84 ([L/M]GH), 88 (RQ[C/S])
Di-/tri-carboxylates (DTC)	Oxoglutarate, citrate, malate, oxaloacetate, isocitrate	Fatty acid synthesis, isoprenoid synthesis, transport of redox equivalents, nitrogen assimilation	26 (IGS), 27 (QSL), 33 (KLK), 35 (RRQ), 77 (GTY), 84 (YLH), 88 (RMT), 93 ([K/R]DN)
3. <i>For amino acids</i>			
Glutamate (GC)	L-Glutamate	Urea synthesis, amino acid degradation (metabolism)	22 (GQA), 77 (NTR), 80 (LRV), 84 (EFL), 85 (KSF), 88 (KYA)
Aspartate/glutamate (AGC)	L-Aspartate, L-glutamate, L-cysteinesulfinate	Malate/aspartate shuttle, urea synthesis, gluconeogenesis, cysteine degradation	22 (GQA), 77 (QCR), 84 (EFQ), 85 (KSF), 88 (KYT)
Ornithine (ORC)	Ornithine, citrulline, lysine, arginine (histidine)	Urea synthesis, basic amino acid metabolism, polyamine biosynthesis	23 ([V/I][A/S]W) but (KSN) in <i>S. cerevisiae</i> , 26 (GL[V/C]) but (ELI) in <i>S. cerevisiae</i> , 84 (EGA), but (QAV) in AtBAC2
Carnitine (CAC)	Carnitine, acylcarnitines	Fatty-acid β -oxidation	23 (VTW), 85 (FSN)
S-Adenosylmethionine (SAMC)	SAM, S-adenosylhomocysteine	Mitochondrial DNA, RNA, and protein methylation	19 (G[E/G]G), 23 ([D/E][C/S][A/G]), 26 ([L/F]RT), 80 ([G/A]RW), 85 ([A/S][S/T/D]X), 88 (FQF)

(Continued)

	Reaction mechanism	Driving force	Mitochondrial carrier
Uniport		$2.3 \frac{RT}{F} \log \frac{[S_c]}{[S_m]}$	Carnitine/acylcarnitine
Symport		$2.3 \frac{RT}{F} \log \frac{[S_c]}{[S_m]} + 2.3 \frac{RT}{F} \Delta pH$	Glutamate, phosphate
Antiport		$2.3 \frac{RT}{F} \log \frac{[S_m^+]}{[S_c^+]} + 2.3 \frac{RT}{F} \Delta pH$	Ornithine (in yeast)
Antiport		$2.3 \frac{RT}{F} \log \frac{[S_{1c}^{3-}][S_{2m}^{4-}]}{[S_{1m}^{3-}][S_{2c}^{4-}]} + \Delta \Psi$	ADP/ATP
Antiport		$2.3 \frac{RT}{F} \log \frac{[S_{1c}^-][S_{2m}^-]}{[S_{1m}^-][S_{2c}^-]} + 2.3 \frac{RT}{F} \Delta pH + \Delta \Psi$	Aspartate/glutamate
Antiport		$2.3 \frac{RT}{F} \log \frac{[S_{1c}^{4-}][S_{2m}^{3-}]}{[S_{1m}^{4-}][S_{2c}^{3-}]} + 2.3 \frac{RT}{F} \Delta pH$	GTP/GDP

Figure 5 Modes of transport catalyzed by mitochondrial carriers and their driving forces. Typical examples of reaction mechanisms and driving forces are shown. Reproduced from Palmieri F and Pierri CL (2010) Mitochondrial metabolite transport. *Essays in Biochemistry* 47: 37–52, with permission from The Biochemical Society.

of species-specific carriers in the most recent branches of the tree. These later duplications concern *A. thaliana* (and other plants) more than *H. sapiens* and *S. cerevisiae*, suggesting a greater number of specific functional requirements in plants as compared to fungi and animalia.

Concerning the early evolution of MCs, their structural features clearly indicate that they result from the tandem triplication of a primordial 100-amino-acid, two-helix domain. Moreover, given that a low-grade sequence similarity is also detectable in the two helices of each repeat, it has been hypothesized that the primordial repeat may itself have evolved by the duplication of a DNA sequence encoding a single transmembrane segment.

Modes of Transport and Driving Forces

The majority of MCs catalyze strict solute antiport (exchange) reactions (Figure 5). Some mediate uniport (unidirectional substrate transport) besides antiport, whereas the uncoupling protein catalyzes uniport as the exclusive transport mode.

Based on the electrical nature of their transport reactions, MCs can be divided into electrogenic (or electrophoretic) and electroneutral transporters (Figure 5). So far, three well-characterized carriers have been shown to be electrophoretic. The ADP/ATP and aspartate/glutamate carriers catalyze exchanges that result in a charge imbalance because they transport ADP^{3-} for ATP^{4-} and aspartate⁻ for glutamate⁻ + H^+ , respectively, across the mitochondrial membrane. The third, the

uncoupling protein, catalyzes the unidirectional transport of H^+ . Electroneutral balance can be achieved by cotransport (symport) and counter-transport (antiport or exchange) of solutes, and by uniport of electroneutral metabolites (Figure 5). In some cases, electroneutrality is imposed by simultaneous carrier-mediated H^+ movement. The carrier subfamilies for phosphate and glutamate, and in yeast for oxaloacetate/sulfate, mediate the transport of anions together with an equivalent amount of H^+ (anion/ H^+ symport), and the yeast GTP/GDP carrier exchanges GTP^{4-} plus a H^+ against GDP^{3-} . Other subfamilies catalyze an exchange of anions or cations. The oxoglutarate carrier, for instance, exchanges oxoglutarate²⁻ for malate²⁻ and the ornithine carrier ornithine⁺ for lysine⁺, arginine⁺, or an H^+ .

As their driving force, MCs utilize the concentration gradient of the transported solutes and in most cases the H^+ electrochemical potential generated across the inner mitochondrial membrane by the functioning of the respiratory chain (Figure 5). Because the electrical component of $\Delta \mu H^+$ is rather high, the electrical nature of the ADP/ATP carrier and the aspartate/glutamate carrier provides a powerful means of ejecting ATP and aspartate against the concentration gradient from the mitochondrial matrix to the cytosol. In cases of H^+ symport or H^+ exchange, the transmembrane pH gradient regulates the distribution of anionic and cationic solutes across the membrane. For example, with a higher pH value inside the mitochondria, the carrier-mediated, H^+ -compensated uptake of phosphate or glutamate is stimulated, as well as the efflux of cationic solutes such as the export of ornithine in yeast.

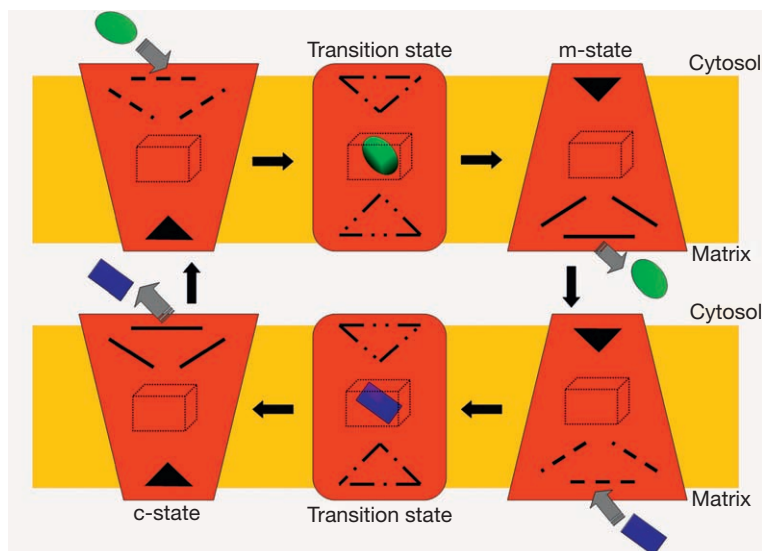


Figure 6 Schematic representation of the transition from the c-state to the m-state, and vice versa, which mitochondrial carriers undergo in the catalytic exchange transport cycle. Trapezoids (left) illustrate the c-state after the release of the substrate toward the cytosol (bottom) and immediately after the entry of the substrate from the cytosolic side (top); trapezoids (right) show the m-state after the release of the substrate into the matrix (top) and immediately after the entry of the substrate from the matrix side (bottom); and the two central rounded rectangles depict the transition states of the carrier with the bound substrate entered from the cytosol (top) and from the matrix (bottom). The box in the center of the carrier indicates the presence of an internal cavity without reference to its shape (which varies in the different states depicted). Green ovals and blue rectangles represent the substrate entering from the cytosol and from the matrix, respectively; solid triangles indicate closed gates, and dotted triangles open/partially closed gates. Reproduced from Palmieri F and Pierri CL (2010) Mitochondrial metabolite transport. *Essays in Biochemistry* 47: 37–52, with permission from The Biochemical Society.

Mechanism of Transport

As predicted by the single-binding center-gating pore hypothesis, MCs have only one binding site in the center of the carrier which is open alternately to the two opposite sides of the membrane, that is, either toward the cytosol in the c-conformation or toward the mitochondrial matrix in the m-conformation. In the c-state, the substrate enters the carrier from the cytosolic side and binds to the carrier (see Figure 6). As binding occurs, the protein rearranges, according to the induced transition fit theory of carrier catalysis, to attain an optimum fit between the protein and the substrate corresponding to the transition conformation. In the transition state, the substrate is bound in the cavity approximately at the center of the carrier and the carrier is compactly structured around the substrate almost entirely closed on either side of the membrane (Figure 6). The total binding energy of the optimum-fit interactions between the carrier and the substrate in the transition state triggers additional structural changes leading to the m-conformation, in which the c-gate is closed and the m-gate is opened. At this stage, the substrate which entered the carrier from the cytosolic side exits into the matrix, and the catalytic cycle continues with the entry of another substrate from the matrix (see Figure 6). This mechanism was first suggested for the ADP/ATP carrier because of the existence of high-affinity inhibitor ligands that bind to the carrier either from the cytosolic (atractyloside) or from the matrix side (bongkrekate). These ligands remove ADP or ATP and their binding is mutually exclusive, indicating a single reorienting binding site. The existence of a c-state has been demonstrated

by the crystal structure of the carboxyatractyloside-ADP/ATP carrier complex. Furthermore, the reorientation of the single binding site is supported by the common and similarly located substrate-binding site and by the kinetic ping-pong mechanism of the carnitine/acylcarnitine carrier (CAC)-catalyzed transport. Unfortunately, a 3D structure of MCs in the m-state is not yet available. Similarly, the structural changes occurring during the transition from the c- to the m-state, and vice versa, are still unknown. However, it has been proposed that they are largely due to the movements of the even and odd transmembrane α -helices (flexible hinged helix movements) consisting of a tilt of the helical segments and a kink/swivel of their termini at the level of the conservative Pro and Gly. Viewing the carrier from the cytosol, the even and odd helices would be seen rotating clockwise during the c- to m-state transition, counterclockwise during the m- to c-state transition, and vice versa, viewing the carrier from the matrix side. Moreover, during the transition from the c- to the m-state the helix termini, hinging on the Pro and Gly residues, move apart on the matrix side (opening the m-gate) and come together on the cytosolic side (closing the c-gate). The opposite movements take place during the transition from the m- to the c-state.

The above-reported mechanism describes the MC-mediated antiport mode of transport. It is thought that MCs catalyzing uniport, besides antiport, are able to undergo a reversible transition between the c-state and the m-state in the absence of the substrate. This transition requires that the activation energy barrier needed for the conformational change of these carriers be much lower than that of the obligatory 1:1 exchange carriers.

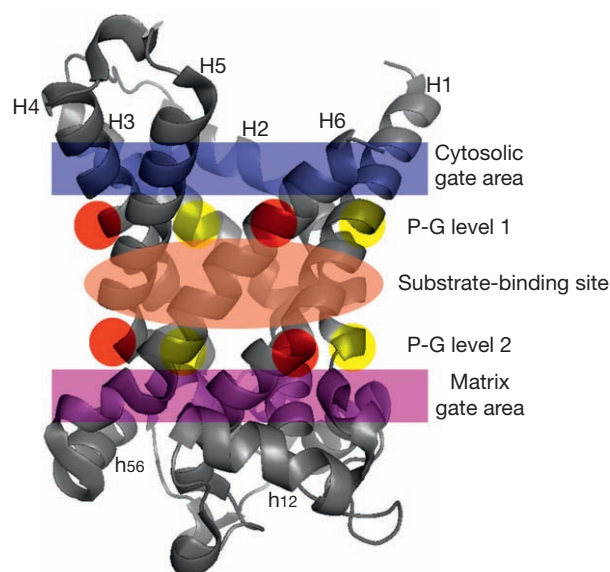


Figure 7 MC structural areas in which most transport-inactivating experimentally or naturally produced mutations are located. The 3D crystal structure of the carboxyatractyloside-ADP/ATP carrier complex (devoid of the inhibitor) is shown with the following regions in color: cytosolic gate (blue); substrate-binding site (orange); and matrix gate (purple). In P-G level 1 and P-G level 2, prolines are shown in red and glycines in yellow. Reproduced from Palmieri F and Pierri CL (2010) Structure and function of mitochondrial carriers – role of the transmembrane helix P and G residues in the gating and transport mechanism. *FEBS Letters* 584: 1931–1939.

Structure–Function Relationships

Most transport-inactivating residue replacements (with cysteine or alanine) in the oxoglutarate carrier and other MCs, as well as most missense mutations found in patients affected by MC-associated diseases (see below), are located in areas of carrier structure that are considered crucial for their function, that is, substrate-binding site, matrix gate, Pro–Gly levels 1 and 2, and cytosolic gate (Figure 7). Notably, nearly all the residues that interact with the lipid bilayer as well as the large majority of those participating in inter-helical interactions tolerate replacement with cysteine or alanine.

Diseases

The ongoing identification of MC gene function has favored the disclosure of several diseases associated with defective carriers. Until now, 11 MC-related diseases have been well characterized biochemically and genetically (Table 2). These disorders are rare errors of metabolism caused by alterations of the genes encoding MCs. The genes are essential for either oxidative phosphorylation (e.g., ADP/ATP and phosphate carriers) or other cell functions (e.g., carriers for glutamate, aspartate/glutamate, thiamine pyrophosphate, and carnitine/acylcarnitines). Therefore, the symptomatology of these diseases depends on the specific metabolism affected and its relevance in specific tissues. For example, CAC deficiency (OMIM 212138) is a fatty-acid β -oxidation disorder that

Table 2 Diseases caused by mutations in genes encoding MCs

Disease	Gene	Carrier
CAC deficiency	SLC25A20	Carnitine/acylcarnitine
AGC2 deficiency (CTLN2/NICCD)	SLC25A13	Aspartate/glutamate 2
HHH syndrome	SLC25A15	Ornithine/citrulline
adPEO	SLC25A4	ADP/ATP
Congenital Amish microcephaly	SLC25A19	Thiamine pyrophosphate
Neuropathy with bilateral striatal necrosis	SLC25A19	Thiamine pyrophosphate
AAC1 deficiency	SLC25A4	ADP/ATP
Early epileptic encephalopathy	SLC25A22	Glutamate
PiC (isoform A) deficiency	SLC25A3	Phosphate
AGC1 deficiency	SLC25A12	Aspartate/glutamate 1
Congenital sideroblastic anemia	SLC25A38	? (glycine/ALA)

adPEO, autosomal dominant progressive external ophthalmoplegia; ALA, aminolevulinic acid.

Other abbreviations as in Table 1.

manifests with life-threatening episodes of coma upon fasting, cardiomyopathy, muscle weakness, and abnormal liver function, and HHH syndrome (OMIM 238970) is a disorder of the urea cycle characterized by hyperammonemia, hyperornithinemia, and homocitrullinemia. With the exception of autosomal dominant progressive external ophthalmoplegia (adPEO, OMIM 157640), all the other diseases are inherited in an autosomal recessive manner. In addition, although the molecular defect(s) responsible for 2-oxoadipate acidemia have not yet been characterized, this disease has been suggested to be due to defective oxodicarboxylate carrier, that transports 2-oxoadipate into mitochondria, or oxoadipate dehydrogenase that oxidizes 2-oxoadipate in these organelles.

See also: [Bioenergetics: Mitochondrial Calcium Transport: Historical Aspects; Mitochondrial Outer Membrane and the VDAC Channel; Mitochondrial Permeability Transition Pore.](#)

Further Reading

- Cappello AR, Miniero DV, Curcio R, et al. (2007) Functional and structural role of amino acid residues in the odd-numbered transmembrane α -helices of the bovine mitochondrial oxoglutarate carrier. *Journal Molecular Biology* 369: 400–412.
- Fiermonte G, Walker JE, and Palmieri F (1993) Abundant bacterial expression and reconstitution of an intrinsic membrane transport protein from bovine mitochondria. *Biochemical Journal* 294: 293–299.
- Klingenberg M (2008) The ADP and ATP transport in mitochondria and its carrier. *Biochimica et Biophysica Acta* 1778: 1978–2021.
- Kunji E and Robinson A (2010) Coupling of proton and substrate translocation in the transport cycle of mitochondrial carriers. *Current Opinion in Structural Biology* 20: 440–447.
- Palmieri F (2004) The mitochondrial transporter family (SLC25): Physiological and pathological implications. *Pflügers Archiv – European Journal of Physiology* 447: 689–709.
- Palmieri F (2008) Diseases caused by defects of mitochondrial carriers: A review with new insights. *Biochimica et Biophysica Acta* 1777: 564–578.
- Palmieri F, Agrimi G, Blanco E, et al. (2006) Identification of mitochondrial carriers in *Saccharomyces cerevisiae* by transport assay of reconstituted recombinant proteins. *Biochimica et Biophysica Acta* 1757: 1249–1262.

- Palmieri F and Pierri CL (2010) Mitochondrial metabolite transport. *Essays in Biochemistry* 47: 37–52.
- Palmieri F and Pierri CL (2010) Structure and function of mitochondrial carriers – role of the transmembrane helix P and G residues in the gating and transport mechanism. *FEBS Letters* 584: 1931–1939.
- Palmieri F, Pierri CL, De Grassi A, et al. (2011) Evolution, structure and function of mitochondrial carriers: A review with new insights. *Plant Journal* 66: 161–181.
- Pebay-Peyroula E, Dahout-Gonzalez C, Kahn R, et al. (2003) Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. *Nature* 426: 39–44.
- Picault N, Hodges M, Palmieri L, and Palmieri F (2004) The growing family of mitochondrial carriers in *Arabidopsis*. *Trends in Plant Science* 9: 138–146.
- Robinson A and Kunji E (2006) Mitochondrial carriers in the cytoplasmic state have a common substrate-binding site. *Proceedings of the National Academy of Sciences of the United States of America* 103: 2617–2622.
- Robinson A, Overy C, and Kunji E (2008) The mechanism of transport by mitochondrial carriers based on analysis of symmetry. *Proceedings of the National Academy of Sciences of the United States of America* 105: 17766–17771.
- Traba J and Satrustegui J (2011) Adenine nucleotide transporters in organelles: Novel genes and functions. *Cellular and Molecular Life Sciences* 68: 1183–1206.
- Wohlrab H (2009) Transport proteins (carriers) of mitochondria. *IUBMB Life* 61: 40–46.

Mitochondrial Outer Membrane and the VDAC Channel

M Colombini, University of Maryland, College Park, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Apoptosis A self-destruct process present in cells of multicellular organisms that eliminates cells that are damaged or no longer needed.

Mitochondrion An organelle occurring in the vast majority of eukaryotic cells that is critical for energy transduction and

metabolism. It is also semi-autonomous with its own DNA and independent reproductive rate.

Voltage-dependent anion-selective channel (VDAC)

A protein found in the outer membrane of all mitochondria. It forms large channels in this membrane allowing metabolites to cross.

Origin and Structure of the Outer Membrane

Mitochondria have most likely evolved from endosymbiotic bacteria. The outer membrane may have originated as an endosomal membrane that engulfed the original endosymbiotic bacterium or may be related to the bacterial outer membrane. Whatever its origin, structurally it is a relatively simple limiting membrane forming a continuous boundary (Figure 1). It seems to limit the boundary of the mitochondrial space and, unlike the inner membrane, maintains a fairly constant total mitochondrial volume.

Further aspects of mitochondrial structure and function vary depending on the source and, thus, most of what will be presented applies to mammalian sources.

The outer membrane interacts intimately with neighboring membranes. Rivet-like connections between the outer and inner membranes often exist and are referred to as contact sites. These are mechanically sturdy as they interfere with the experimental separation of the outer and inner membranes. These contacts are proposed to serve a number of functions including the formation of macromolecular complexes between transporters and enzymes to expedite metabolic flow. The outer membrane also interacts closely with the endoplasmic reticulum and this close apposition may be the reason for preferential transfer of Ca^{2+} from the endoplasmic reticulum to the mitochondrial matrix compartment. This has been especially well studied in the heart where changes in Ca^{2+} level on a beat-by-beat basis alter dehydrogenase activity matching cellular energy consumption. If specialized docking structures are involved in this interaction with the endoplasmic reticulum, they must be more labile than the contact sites because isolation of mitochondria from endoplasmic reticulum is quite easy.

Functions of the Outer Membrane

The outer membrane of isolated mitochondria is highly permeable to metabolites due to the presence in this membrane of large aqueous channels formed by the protein called voltage-dependent anion-selective channel (VDAC). Their similarity to channel-forming proteins in bacterial outer membranes collectively referred to as porins has led some to refer to VDAC as

mitochondrial porin. Although the permeability characteristics conferred to the outer membrane by VDAC are rather complex, physical size is a major factor. Thus, under normal conditions, the outer membrane is essentially impermeable to molecules larger than 5 kDa. This allows proteins to be sequestered in the intermembrane space, the compartment between the outer and inner membranes. This sequestration serves at least two functions. First, it localizes proteins that could diffuse away to the site where they are needed. For example, the soluble protein, cytochrome *c*, is a member of the electron transport chain and must transmit electrons between complex III and IV so that respiration may proceed. Second, the sequestration prevents proapoptotic proteins from being released into the cytosol as this would initiate programmed cell death (apoptosis). Cytochrome *c* is also an example of a sequestered proapoptotic protein. The permeability barrier to proapoptotic proteins and others of similar size is lifted under some conditions that lead to cellular apoptosis.

The permeability of the outer membrane to small molecules can be regulated resulting in the outer membrane limiting metabolite flux and, consequently, the level of metabolic activity. This regulation is probably critical to changes in cell function.

Most mitochondrial proteins and RNA molecules are synthesized in the cytosol and imported into mitochondria. The Tom complex is responsible for much of the protein import through the outer membrane. Leader sequences on the proteins are recognized by Tom and the protein is treaded through a pore in the complex and delivered to the intermembrane space where it is picked up by the Tim complex in the inner membrane for translocation through that membrane. The processing of outer membrane proteins by this complex is less clear.

Mitochondria are usually found in the portion of cells where they are needed. For example, to minimize diffusion limitations they are located close to sites of high-energy dissipation. Sometimes they need to be delivered to distal sites in a cell (e.g., from cell body to nerve terminal of a neuron or to the reproductive bud of a yeast cell). Motors located on either microtubules or actin filaments move mitochondria from one location in the cell to another. The outer membrane must provide sites of attachment for the motors that perform the translocation.

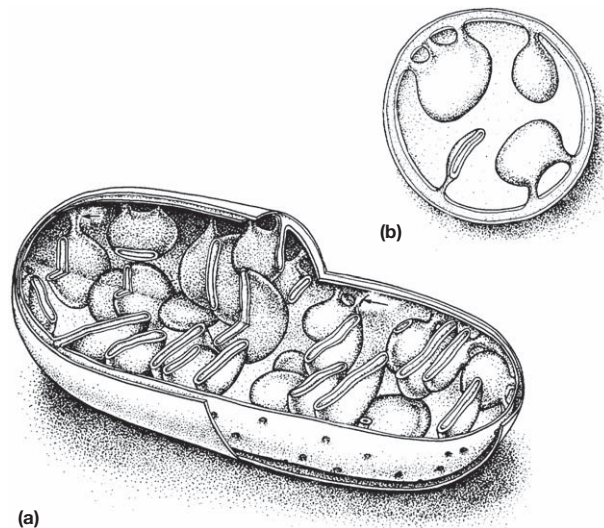


Figure 1 A drawing of the membranes and compartments of a typical small mitochondrion. (a) The mitochondrion with portions of the membranes cut out so as to reveal inner structure and compartmentation. (b) Illustrates a cross section. Reproduced from Lea PJ and Hollenberg MJ (1989) Mitochondrial structure revealed by high-resolution scanning electron microscopy. *The American Journal of Anatomy* 184: 245–257. Drawings made by Karen Visser. This material is used by permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.

Composition

The composition of the outer membrane varies somewhat depending not only on the species but also on the tissue from which it was isolated. With that caveat, it can be said that there are a number of similarities between the constituents of the outer membrane and the endoplasmic reticulum, for example, the presence of cytochromes *b* and *b₅*, antimycin-insensitive reduced nicotinamide adenine dinucleotide (NADH)-cytochrome *c* reductase, and ceramide synthase. Yet there are also many differences and the enzymes present in the two locations are usually different isoforms. Thus, the similarity may reflect similar origins but certainly these two membranous compartments are distinct.

The outer membrane also contains enzymes involved in fatty acid and phospholipid metabolism and an amine oxidase. The liver mitochondria contain a peripheral benzodiazepine receptor whose role is unclear. The membrane can also bind some kinases that are typical soluble enzymes and the binding is thought to accelerate energy transduction.

The outer membrane has a higher lipid content than the inner membrane, one more typical of other cell membranes. The lipid composition is also drastically different from that of the inner membrane, the outer membrane being rich in sterol, while the inner has higher levels of cardiolipin.

VDAC Channels in the Outer Membrane

VDAC Structure

Mitochondria from all eukaryotic kingdoms contain VDAC channels in their outer membrane. These are formed by a

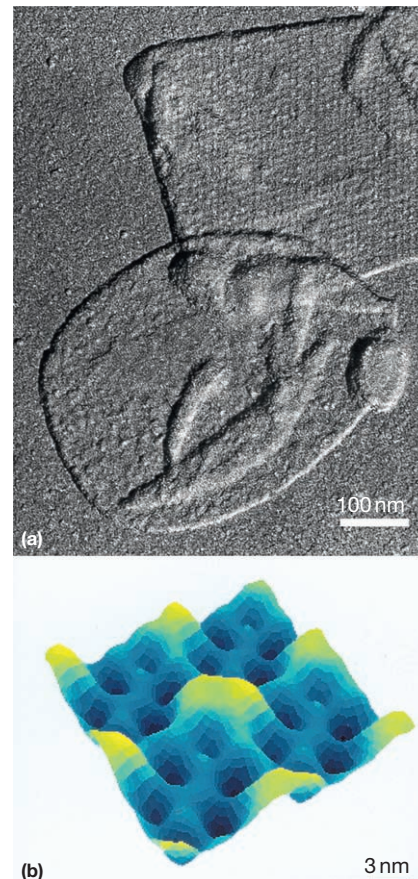


Figure 2 Views of mitochondrial outer membranes and the ubiquitous VDAC channels in these membranes. (a) Electron micrographs of freeze-dried mitochondrial outer membranes with two-dimensional crystals of VDAC channels (membrane with squared edges and ordered array of particles) and without such crystals (smooth membrane with rounded edges). Outer membrane vesicles isolated from mitochondria are dried down on a surface and they form collapsed sacs. (b) A surface view of the VDAC channels obtained after computer filtration and averaging of multiple micrographs containing images of two-dimensional crystals. The yellow shows regions of elevation and the dark blue areas are the openings of the pores (holes) that are formed through the membrane. These allow the metabolites to cross the outer membrane. This figure was assembled from the work of Dr. Lorie Thomas. Reproduced from Thomas L, Kocsis E, Colombini M, et al. (1991) Surface topography and molecular stoichiometry of the mitochondrial channel, VDAC, in crystalline arrays. *Journal of Structural Biology* 106: 161–171.

single polypeptide of approximately 30 kDa and form a single large aqueous pore. In some plant and fungal mitochondria, VDAC channels form two-dimensional crystals (Figure 2), which upon further analysis reveal fundamentally cylindrical proteins forming water-filled transmembrane pathways. The wall of the channel is almost certainly a single layer of protein mainly consisting of a beta structure. Experimental evidence leads to the conclusion that the transmembrane portion consists of 1 α -helix and 13 β -strands. The surface domains (yellow regions in Figure 2) likely act as binding sites for a variety of proteins that bind or interact with the channel. Three-dimensional structures of human VDAC1 refolded from

denatured protein and solved by nuclear magnetic resonance and X-ray crystallography show 19 transmembrane beta strands and virtually no surface domains. These structures reveal a nonnative conformation as they are incompatible with numerous functional studies.

VDAC Function and Dynamics

The primary role of mitochondria is metabolism that is closely integrated with the metabolism occurring in the cytosol. This requires the continuous and rapid exchange of metabolites between these two major compartments. VDAC channels in the outer membrane can exist in conformational states that either favor or hinder metabolite flux. The state that favors metabolite flux, the open state, does so by having an electrostatic environment that favors anions. In addition, the charge distribution in the lining of the channel is such that it favors the translocation of adenosine triphosphate (ATP) over that of anions of similar size and charge. In the closed states, the water-filled channel is somewhat narrower (the pore size decreases from 2.5 nm in the open state to 1.8 nm in the larger of the population of closed states) but, most importantly, the electrostatic nature is such that it hinders the flow of anions. A positively charged region lining the channel in the open state is translocated out of the channel resulting in a pathway with a net negative charge. This excludes ATP and drastically reduces the permeability of other anionic metabolites. At the same time, this change favors the flux of cations.

This gating of VDAC is voltage dependent and, thus, the channels can respond to changes in the voltage across the outer membrane. An electrical potential across this membrane may arise from the presence of different charged macromolecules in the aqueous compartments (Donnan potential) and/or differential metabolite flux responding to metabolic rates. In addition, numerous factors influence this gating process including cytosolic NADH levels and specific proteins in the cytosol and intermembrane space. The importance of this regulation is indicated by its remarkable conservation of VDAC from all species tested (including all eukaryotic kingdoms). The diversity of influences on the gating of VDAC indicates that VDAC may integrate the influence of a variety of cellular factors and, thus, tune mitochondrial activity appropriately.

VDAC's ability to bind with cytosolic hexokinase and intermembrane space creatine kinase has been proposed to result in organized structures that increase the efficiency of energy transduction.

The Outer Membrane and Apoptosis

Apoptosis, programmed cell death, is a process by which a cell repackages itself to be ingested by other cells, generally macrophages. The outer membrane separates proteins that once combined initiate an amplification cascade leading to 'full blown' apoptosis. The leakage of proteins from the intermembrane space into the cytosol is sufficient to cause apoptosis. Even injecting just one of these proteins, cytochrome *c*, into the cytosol is sufficient.

How the outer membrane becomes permeable to proteins is an area of intense investigation. The proteins that cross this membrane can be as large as 50–100 kDa, proteins of the Bcl-2 family partition between the cytosol and the outer membrane. They can exist as both soluble proteins and intrinsic membrane proteins capable of generating ion channels. Some of these are proapoptotic and others are antiapoptotic. Some antiapoptotic proteins are often associated with the outer membrane under normal conditions, that is, cells functioning normally, and may favor normal mitochondrial function. Others translocate to the outer membrane at the early stages of apoptosis, perhaps forming the protein efflux pathways. The sphingolipid, ceramide, can also be produced in mitochondria early in the apoptotic process and has been shown to be able to form channels large enough for protein flux.

Under some conditions of apoptosis induction, exchange of nucleotides between mitochondria and the cytosol is greatly inhibited. The outer membrane becomes poorly permeable to ATP and adenosine diphosphate and other anionic metabolites. This phase precedes the release of intermembrane space proteins and in this phase cells can be rescued. This gives the outer membrane the role of deciding between life and death.

Emerging Importance of the Outer Membrane

Once considered relatively unimportant, the mitochondrial outer membrane is emerging as a site for control of overall mitochondrial function and a location where the fate of the entire cell is decided. Despite its deceptively simple appearance, this membrane performs a variety of vital functions. By hindsight, its location at the frontier between the cytosol and the mitochondrial spaces makes it an ideal site for control of the flux of matter and transduction of information.

See also: [Bioenergetics: Cell Death by Apoptosis and Necrosis; Cytochrome *c*; Mitochondrial Membranes, Structural Organization.](#)

Further Reading

- Colombini M (2009) The published 3-D structure of the VDAC channel: Native or not? *Trends in Biochemical Sciences* 34: 382–389.
- Colombini M, Blachly-Dyson E, and Forte M (1996) VDAC, a channel in the outer mitochondrial membrane. In: Narahashi T (ed.) *Ion Channels*, pp. 169–202. New York, NY: Plenum.
- Gordon DM, Dancis A, and Pain D (2000) Mechanisms of mitochondrial protein import. *Essays in Biochemistry* 36: 61–73.
- Lea PJ and Hollenberg MJ (1989) Mitochondrial structure revealed by high-resolution scanning electron microscopy. *American Journal of Anatomy* 184: 245–257.
- Thomas L, Kocsis E, Colombini M, et al. (1991) Surface topography and molecular stoichiometry of the mitochondrial channel, VDAC, in crystalline arrays. *Journal of Structural Biology* 106: 161–171.
- Tyler DD (1992) *The Mitochondrion in Health and Disease*. Somerset, NJ: VCH Publishers (currently Wiley-VCH Verlag GmbH and Co., Wiley).
- Waterhouse NJ, Ricci J, and Green DR (2002) And all of a sudden it's over: Mitochondrial outer-membrane permeabilization in apoptosis. *Biochimie* 84: 113–121.

Mitochondrial Permeability Transition Pore

P Bernardi, University of Padua, Padua, Italy

© 2013 Elsevier Inc. All rights reserved.

Introduction

The mitochondrial permeability transition (PT) can be defined as a Ca^{2+} -dependent increase of mitochondrial inner membrane permeability to solutes with molecular masses up to about 1500 Da, which has been studied mostly in mammals. *In vitro* at least, it is accompanied by depolarization, matrix swelling, depletion of matrix pyridine nucleotides (PNs), outer membrane rupture, and release of intermembrane proteins including cytochrome *c*. The PT is believed to be due to opening of a high-conductance, voltage-dependent channel in the inner mitochondrial membrane, the PT pore (PTP); in spite of many efforts, its molecular identity remains unknown. The occurrence of swelling in isolated mitochondria and its basic features (stimulation by Ca^{2+} and Pi uptake and inhibition by Mg^{2+} , adenine nucleotides, and acidic pH) had been recognized very early in studies on isolated mitochondria, as were its detrimental effects on energy conservation. As a result, the occurrence of a PT has long been considered an *in vitro* artifact of little or no physiological interest. This state of affairs was paradoxically worsened by the clarification of the basic mechanism of mitochondrial energy conservation, which took place in the 1960s and 1970s, and culminated with the award of the 1978 Nobel prize in chemistry to the late Peter Mitchell. He understood that the initial event in energy conservation is charge separation at the inner mitochondrial membrane, for which he coined the term 'chemiosmosis'.

Chemiosmosis, Mitochondrial Channels, and the PT

The four basic postulates of chemiosmosis are (1) that the membrane-located adenosine triphosphatase (ATPase) reversibly couples the translocation of protons across the membrane to the flow of anhydro-bond equivalents between water and the couple adenosine triphosphate (ATP)/(adenosine diphosphate (ADP) + Pi); (2) that the membrane-located respiratory chain catalyzes the flow of reducing equivalents, coupling reversibly the translocation of protons across the membrane to the flow of reducing equivalents during oxido-reduction; (3) that the mitochondrial inner membrane has specific carriers allowing anion- OH^- and cation- H^+ exchanges that regulate the pH and osmotic differential across the membrane, and permit the flux of essential metabolites without collapse of the membrane potential; and (4) that the systems of the first three postulates are located in a specialized coupling membrane which has a low permeability to protons and to anions and cations generally. The fourth postulate was widely (and erroneously) implied to mean that the inner membrane did not possess cation channels. Given this view, prevailing as late as the 1990s, it comes as no surprise that the PTP, at an estimated diameter of 3 nm, did not attract too much interest in bioenergetics. Things have changed dramatically over the last 15 years,

following the demonstration that the inner mitochondrial membrane does possess cation channels that are expectedly tightly regulated, and that a PT can occur in cells, tissues, and organisms where it plays a role in cell death, and possibly in Ca^{2+} homeostasis.

Molecular Nature of the PTP

Two main theories were proposed to explain the PT: the membrane theory, postulating that the permeability increase occurred mainly or exclusively in the lipid bilayer following activation of a Ca^{2+} -dependent phospholipase, and the pore theory, postulating that the PT was due to opening of a regulated proteinaceous channel. The discovery that the PTP can be desensitized by nanomolar concentrations of cyclosporin A (CsA) without any effects on phospholipase activity shifted the balance in favor of the pore theory. The effects of CsA on the PTP are mediated by its high-affinity binding to cyclophilin (CyP)-D, a matrix peptidyl prolyl *cis-trans* isomerase that assists in protein folding, and is believed to act as a promoter of the PT. CyP-D is not a structural part of the channel but rather a modulator, and the only PTP-related molecule identified so far.

CyP-D

CyP-D is part of the CyP family of proteins, which in mammals has 16 members, among which cytosolic CyP-A is the most studied. The binding of CsA to CyP-A creates a complex that acquires the ability to bind to, and inhibit, cytosolic calcineurin, a Ca^{2+} -calmodulin-dependent phosphatase that catalyzes dephosphorylation (hence nuclear translocation) of nuclear factor of activated T-cells (NFAT), an essential element for the activation of the immune response against transplants. As a result of calcineurin inhibition, CsA prevents transplant rejection. It has been established that calcineurin inhibition instead is not required for PTP inhibition by CsA.

The mechanism through which CyP-D modulates the PTP was clarified quite recently. We found that CyP-D hinders a regulatory site for matrix inorganic phosphate (Pi), whose occupancy results in PTP inhibition (rather than the activation normally caused by Pi). When CyP-D is genetically ablated, or when CsA is added to displace CyP-D from the inner membrane, the PTP is desensitized by matrix Pi resulting in a decreased probability of pore opening. It must be stressed that mitochondria from lower eukaryotes such as yeast and *Drosophila* can undergo a permeability increase that has some points in common with the PT. This PT is insensitive to CsA, but it is inhibited by Pi; thus, the finding that Pi can be a desensitizer in higher eukaryotes as well may bridge the gap between yeast and mammals.

Recent studies have addressed the regulatory interactions of CyP-D with the F_1F_0 -ATP synthase, whose catalytic part (F_1)

and lateral stalk protrude into the matrix where CyP-D is located. CyP-D interacts with the ATP synthase under mild extraction conditions, and can be cross-linked with the oligomycin sensitivity conferring protein (OSCP), b, and d subunits (i.e., with the extrinsic part of the lateral stalk). These interactions are mentioned here because binding of CyP-D is favored by Pi, while CsA can displace CyP-D from the ATP synthase. These binding changes are matched by functional changes, with a Pi-dependent decrease of specific activity that is reversed by CsA. We presently do not know whether this interesting analogy with the PTP is a coincidence, or rather reflects an involvement of the F_1F_0 ATP synthase in pore formation and/or regulation, an issue that deserves further study.

Adenine Nucleotide Translocator and Voltage-Dependent Anion Channel

The adenine nucleotide translocator (ANT) has long been considered to be the main component of the PTP, possibly in association with the outer membrane voltage-dependent anion channel (VDAC). The PTP is modulated by ligands of the ANT. Atractylate, which inhibits the ANT and stabilizes it in the so-called 'c' conformation (nucleotide-binding sites facing the cytosolic side of the inner membrane), favors PTP opening, while bongkrekate, which also inhibits the ANT but stabilizes it in the 'm' conformation (nucleotide-binding sites facing the matrix side of the inner membrane), favors PTP closure. These findings are the main basis for the suggestion that the PTP may be formed by the ANT. An alternative explanation, which is based on PTP modulation by the membrane potential (see below), is that the transitions of the ANT from the 'm' to the 'c' conformation cause large changes of the surface potential that may explain PTP opening by atractylate and its closure by bongkrekate as an indirect effect on the membrane potential. Conclusive evidence that the ANT is not essential for PTP formation was obtained in mitochondria lacking all ANT isoforms, which still displayed a Ca^{2+} -dependent and CsA-sensitive PT (indicating that the ANT is not the binding partner of CyP-D either).

The PT of mammals can be induced by relatively high concentrations of substituted maleimides (see below), but this is not observed in mitoplasts (mitochondria devoid of the outer membrane), suggesting that inner membrane permeability changes also require an intact outer membrane. The suggestion that the outer membrane protein involved in PTP formation could be VDAC was based on some striking analogies with the PTP. For example, purified VDAC incorporated into planar phospholipid bilayers forms channels with a pore diameter of 2.5–3.0 nm whose electrophysiological properties strikingly resemble those of the PTP; VDAC channel properties are modulated by many conditions that also modulate the PTP; and chromatography of mitochondrial extracts on a CyP-D affinity matrix allowed purification of VDAC and the ANT, which in the presence of CyP-D catalyzed a CsA-sensitive permeabilization of liposomes to solutes. However, recent results obtained in mitochondria from mice lacking the major VDAC isoform (VDAC1), as well as from mammalian cells where all three VDAC isoforms had been deleted, indicate that each VDAC isoform is dispensable for occurrence of the PT,

demonstrating that VDACS cannot be essential components of the PTP.

Regulation of the PTP

Divalent Cations

Matrix Ca^{2+} is perhaps the single most important factor required for PTP opening. Ca^{2+} is best defined as a permissive factor, in the sense that although a PT is not observed in the absence of matrix Ca^{2+} , Ca^{2+} alone is not sufficient to induce PTP opening. Ca^{2+} is unique because all the other Me^{2+} ions that are transported by the mitochondrial Ca^{2+} uniporter (and hence can be accumulated in the matrix, such as Sr^{2+} and Mn^{2+}) are inhibitors of the PT, perhaps by competing with Ca^{2+} for the same binding sites. It has always been difficult to separate the PTP-inducing effects of Ca^{2+} from those of Pi (and possibly also of polyphosphate generated in the matrix) because the uptake of Pi (or of other species able to buffer matrix pH such as acetate or bicarbonate) is required for the uptake of substantial amounts of Ca^{2+} . The inducing effect of Pi is surprising because it should decrease free matrix $[Ca^{2+}]$; however, the effect could rather be due to the Pi-dependent decrease of free matrix $[Mg^{2+}]$, which at variance from Ca^{2+} inhibits the PTP (although Mg^{2+} is not transported rapidly across the inner membrane, it is present in substantial amounts in the matrix space). An external Me^{2+} -binding site has also been found. All tested Me^{2+} ions, including Ca^{2+} itself, cause a decreased probability of PTP opening through this site.

Proton Electrochemical Gradient

The PTP behaves as a voltage-gated channel, with an increase of the open probability as the membrane potential difference (negative inside) becomes less negative, that is, as the inner membrane depolarizes. The threshold voltage for PTP opening is affected by a variety of factors such as Ca^{2+} , Mg^{2+} , and adenine nucleotides; redox effectors including quinones; and amphiphilic molecules such as fatty acids. The existence of a voltage sensor has been postulated, which would detect the membrane potential at which opening occurs. Thus, PTP opening can be the consequence of depolarization (when the membrane potential drops below the threshold) or its cause (when effectors move the threshold toward the resting potential). The effects of the ANT inhibitors atractylate (PTP opening) and bongkrekate (PTP closure) can be easily explained within the framework of the voltage dependence. Indeed, the ANT 'c' conformation (stabilized by atractylate) neutralizes surface positive charges on the outer surface of the inner membrane, thus simulating a depolarization affecting the putative voltage sensor as a signal for PTP opening; conversely, the ANT 'm' conformation (stabilized by bongkrekate) increases the surface negative charges on the inner side of the inner membrane, thus simulating a hyperpolarization affecting the sensor as a signal for PTP closure.

The PTP is also profoundly affected by changes of matrix pH, with an optimum at $pH \approx 7.3$. Lower pH values decrease the PTP open probability via reversible protonation of histidyl residues, and higher pH values decrease pore opening via an undefined mechanism. It appears to be possible that the

inducing effect of Pi may partly depend on matrix pH buffering, as the pK_a of the $H_2PO_4^- \leftrightarrow H^+ + HPO_4^{2-}$ reaction is 7.2.

Redox Sensitivity

An important feature of the PTP is its modulation by redox effectors, a more oxidized state favoring PTP opening. This effect is conferred by at least three redox-sensitive sites that could be distinguished by a careful dissection of their regulatory factors and inhibitor sensitivity. The first is a sulfhydryl (SH) site accessible from the matrix, which is in apparent redox equilibrium with glutathione and can be blocked by low micromolar concentrations of *N*-ethylmaleimide (NEM) and monobromobimane but not by impermeant bimanes or by β -hydroxybutyrate. The second is a site of unknown chemical nature also accessible from the matrix and in apparent redox equilibrium with pyridine nucleotides, which can be blocked by NEM or β -hydroxybutyrate but not by monobromobimane. The third site is an external thiol (accessible from the intermembrane space) that undergoes oxidation and triggers PTP opening after a separate site has reacted with millimolar concentrations of NEM or low micromolar concentrations of copper-*o*-phenantroline. This complex array of PTP redox-sensitive sites has recently been confirmed by careful studies based on PTP inactivation/reactivation by photoradiation after loading mitochondria with hematoporphyrin, which induces a photodynamic effect in close proximity to the hematoporphyrin-binding sites.

Consequences of PTP Opening

The primary consequence of PTP opening is depolarization, which is followed by equilibration of ions and solutes with molecular mass below the pore cutoff of about 1500 Da. PTP opening *in vitro* is followed by release of matrix Mg^{2+} and of adenine and pyridine nucleotides, which will tend to stabilize the pore in the open conformation. PTP closure (e.g., by the addition of Ca^{2+} chelators) is followed by mitochondrial resealing and repolarization, indicating that no permanent changes of inner membrane permeability occurs as a consequence of a PT.

Matrix swelling follows stable PTP opening *in vitro*, and this can lead to damage the outer membrane with release of cytochrome *c* and other pro-apoptotic factors. Swelling results from solute and water flux to the matrix when there is an osmotic imbalance, and it can be prevented if the incubation medium contains low concentrations of pore-impermeant solutes such as polyethylene glycol 3400. Partial swelling may occur for very short PTP open times, which causes widening of the intracristal junctions that otherwise limit the diffusion of the bulk of cytochrome *c* to the intermembrane space. This event may contribute to release of cytochrome *c* through channels formed by activated Bak and Bax even when the outer membrane is intact. Mitochondria are able to shrink and fully recover energy-linked functions even after PT-dependent large-amplitude swelling with outer membrane rupture in saline media, provided cytochrome *c* is added back.

Occurrence of a PT causes secondary consequences on respiration. The initial uncoupling (due to the loss of membrane permeability) is followed by respiratory inhibition because

matrix pyridine nucleotides are readily lost at the onset of the PT. The extent of respiratory inhibition will then depend on the extent of matrix pyridine nucleotide depletion and subsequent hydrolysis by intermembrane glycohydrolase. Finally, the collapse of Δp caused by the PT will curtail ATP synthesis as long as the pore is open. Mitochondria will actually become consumers of any available glycolytic ATP, because in the absence of a proton gradient the ATP synthase will work in reverse as an ATP hydrolase. Imbalanced ion homeostasis (Ca^{2+} in particular), volume dysregulation, and cytochrome *c* release are all factors that favor cell death, as we will further discuss below.

A Role for the PTP in Cellular Ca^{2+} Homeostasis?

In respiring mitochondria, Ca^{2+} is taken up through a channel (the mitochondrial Ca^{2+} uniporter, MCU) which equilibrates the Ca^{2+} gradient with the membrane potential. The driving force is enormous, as for a voltage of -180 mV the predicted accumulation ratio at equilibrium would be 10^6 , with a matrix concentration of 0.1 M for a resting cytosolic $[Ca^{2+}]$ of 0.1 μ M. Equilibrium is never achieved because of the operation of 2 Ca^{2+} efflux pathways, the Na^+ - Ca^{2+} and H^+ - Ca^{2+} antiporters. Coupling of electrophoretic Ca^{2+} uptake via the MCU with Ca^{2+} efflux on the antiporters allows fine regulation of matrix and cytosolic Ca^{2+} (Figure 1). It should be noted that energy is required both for Ca^{2+} uptake and for Ca^{2+} release, owing to the electrophoretic nature of transport via MCU and to the fact that the stoichiometry of the antiporters is most likely $3Na^+$ (H^+) per Ca^{2+} . Operation of the antiporters is therefore favored by the membrane potential (and by the Na^+ , H^+ , and Ca^{2+} concentration gradients). As a result of this complex arrangement steady-state extramitochondrial $[Ca^{2+}]$ can be increased (and matrix $[Ca^{2+}]$ decreased) by increasing the activity of the antiporters, by decreasing the activity of the MCU, or both. Yet, the rate of uptake increases very steeply with extramitochondrial $[Ca^{2+}]$, reaching a V_{max} as high as 1400 nmol $Ca^{2+} \times mg$ protein $^{-1} \times min^{-1}$; while the maximal rate of operation of the combined antiporters is about 20 nmol $Ca^{2+} \times mg$ protein $^{-1} \times min^{-1}$. This situation exposes mitochondria to the deadly risk of Ca^{2+} overload and calcification, which can be prevented by PTP openings of short duration that have been documented in intact cells and organs. But why should have nature designed a high-conductance, apparently unselective channel for mitochondrial Ca^{2+} release? Can fast Ca^{2+} release rather occur via the MCU after depolarization?

Even after complete mitochondrial depolarization, Ca^{2+} efflux down its concentration gradient via a selective channel (like the MCU) is opposed by the buildup of a Ca^{2+} diffusion potential which, according to the Nernst equation, is -30 mV per decade. Since the inner membrane has a very low permeability to charged species (Mitchell's fourth postulate), the rate of Ca^{2+} efflux would be extremely slow and could only be increased by independent modulation of membrane ion permeability. A pore of large size like the PTP confers the obvious advantage of charge compensation within the channel itself, allowing maximal Ca^{2+} flux (i.e., at zero potential). Thus, a pore of large size would permit fast Ca^{2+} release even for very small $[Ca^{2+}]$ gradients. Since no K^+ and Na^+ concentration

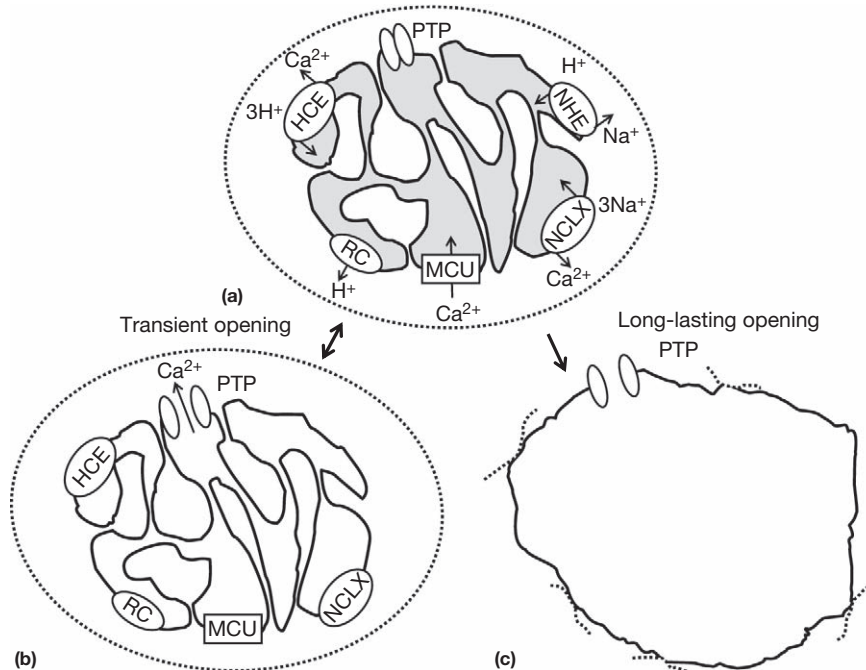


Figure 1 Possible role of the PTP in cellular Ca^{2+} homeostasis. (a) Owing to the low permeability of the inner membrane to charged species and solutes, under basal conditions (when the permeability transition pore, PTP, is closed) the proton-pumping activity of the respiratory chain (RC) creates a proton electrochemical gradient whose membrane potential component (depicted as a darker matrix space) represents a powerful driving force for Ca^{2+} accumulation through the Ca^{2+} uniporter (MCU). In spite of the enormous capacity of MCU, Ca^{2+} flux is slow because the resting Ca^{2+} in the cytoplasm and intermembrane space is well below the K_m of the MCU. In spite of their low combined $V_{\max}$ (two orders of magnitude lower than that of the MCU) the Na^{+} - Ca^{2+} exchanger (NCLX) and the putative H^{+} - Ca^{2+} exchanger (HCE; the depicted stoichiometry is tentative, but operation of this efflux system is favored at high membrane potential) are adequate to prevent matrix Ca^{2+} overload, and Na^{+} accumulation is prevented by the Na^{+} - H^{+} exchanger (NHE). (b) When efflux on the NCLX and HCE is saturated (e.g., following a large Ca^{2+} discharge from the sarcoplasmic reticulum) the matrix Ca^{2+} concentration rises and causes transient opening of the PTP with depolarization and Ca^{2+} release, which is followed by PTP closure and repolarization (back to (a)). (c) For long-lasting PTP openings depolarization is no longer reversible, and solute flux through the pore leads to matrix swelling and to release of intermembrane pro-apoptotic factors which can lead to cell death. For further explanation, see text. The solid lines depict the inner membrane, while the dashed lines represent the outer mitochondrial membrane, which has been drawn at a distance from the inner membrane only to accommodate the lettering.

gradients exist across the inner membrane, PTP opening can lead to selective Ca^{2+} release without major consequences on K^{+} and Na^{+} homeostasis. In this respect, the PTP would be similar to the Ca^{2+} release channel of the sarcoplasmic reticulum, which operates as a selective Ca^{2+} channel in spite of its large size ($>38 \text{ \AA}$), high conductance for monovalent cations ($\approx 1 \text{ nS}$ at saturating K^{+}), permeability to solutes such as glucose, and low permeability ratio when both K^{+} and Ca^{2+} are present. This hypothesis could not yet be tested because of the lack of true PTP blockers, and of the uncertainties about the molecular nature of the PTP which prevent experiments based on genetic elimination of specific pore components.

Signaling to the PTP

PTP regulation by kinase signaling has been widely investigated. The Reperfusion Injury Salvage Kinase (RISK) group confers cardioprotection when activated during post-ischemic myocardial reperfusion or ischemic pre- and postconditioning. RISKS include the survival kinases AKT and extracellular regulated kinase 1 and 2 (ERK1/2), protein kinase C epsilon (PKC ϵ), protein kinase G (PKG), and p70s6K, and protect from

cell death by both phosphorylation of Bcl-2 family proteins and inhibition of the mitochondrial PTP, events that may well be linked. A substrate of the RISK pathway is GSK3 β which, unlike most kinases, is constitutively active in unstimulated cells where it exerts a tonic inhibitory effect on its downstream targets. GSK3 β is inactivated by serine phosphorylation induced by a variety of signaling pathways, including the one involving ERK. We have recently found that a small fraction of ERK associates to mitochondria, where it can be activated to phosphorylate (hence inactivate) GSK3 β resulting in PTP desensitization. CyP-D was associated with ERK2 and GSK3 β , and it was Ser/Thr phosphorylated, phosphorylation correlating with binding to GSK3 β and with PTP opening.

Another key kinase is hexokinase II, whose binding to mitochondria is promoted by the survival kinase AKT, which is activated by growth factors and in most tumor cells. AKT-dependent phosphorylation of mitochondrial hexokinase II inhibits Ca^{2+} -induced cytochrome *c* release, and hexokinase II association to the outer mitochondrial membrane is favored when GSK3 β is inactivated by AKT phosphorylation. Accordingly, activation of GSK3 β induces release of hexokinase II and enhances susceptibility to cell death through increased sensitivity of mitochondria to PTP induction.

Further regulatory interactions involving the PTP are emerging in the literature. A recent study demonstrated a complex mitochondrial chaperone network that involves Hsp90, its related molecule TNF receptor associated protein 1 (TRAP-1), and CyP-D. CyP-D bound to the complex was no longer available to favor PTP opening, and its displacement from the complex by selective Hsp90 antagonists reactivated the PTP with the onset of mitochondrial depolarization, release of cytochrome *c*, and massive cell death. Since the Hsp90-TRAP-1-CyPD interactions are particularly prominent in tumor cells, this strategy has been successfully used for their selective killing.

The PTP in Disease

Largely through the use of CsA, and more recently through the study of *Ppif*^{-/-} mice lacking CyP-D, key advances have been made in understanding the role of CyP-D, and to some extent of the PTP, in diseases. Here we focus on systems where an involvement of the PTP has been investigated *in vivo*.

Myocardial Ischemia-Reperfusion

The role of mitochondria in cell death extends beyond the shortage of ATP supply, and involves generation of reactive oxygen species, impairment of ion homeostasis, accumulation, and release of potentially harmful metabolites, and release of pro-apoptotic proteins such as cytochrome *c*. In the heart, mitochondrial function can paradoxically precipitate the sudden cell death that occurs when coronary flow is reestablished after prolonged ischemia (post-ischemic reperfusion). The partial recovery of mitochondrial function generates enough ATP for contraction but not for relaxation, resulting in hypercontraction and sarcolemma rupture, and a key to cardioprotection is therefore preventing mitochondrial dysfunction.

It is now well documented that the PTP represents an ideal target for cardioprotection. While ischemia does not cause PTP opening, probably because of the protective effects of intracellular acidosis, it creates the conditions for PTP opening at reperfusion because recovery of respiration in the presence of elevated cytosolic [Ca²⁺] and [Pi] provides the ideal conditions to promote PTP opening, which would be further favored by overproduction of reactive oxygen species and recovery of neutral pH. Evidence that the PTP plays a role in reperfusion injury has been obtained in isolated cardiomyocytes, perfused hearts, experimental infarction in living animals, and in a pilot study in a small number of patients. In all these cases, acute CsA administration reduced the occurrence of contractile impairment and irreversible damage. Recent results in mice where CyP-D had been genetically ablated have confirmed that lack of CyP-D confers a remarkable protection against acute infarction.

Although these issues are still controversial, PTP inhibition may also be involved in ischemic preconditioning, that is, the ability of short, repeated ischemic episodes to confer protection against a prolonged and otherwise lethal ischemia, and in ischemic postconditioning, that is, the ability to confer protection by a brief ischemic episode applied at the onset of reperfusion. In general, it appears that development of PTP blockers will be of great interest in the cardiovascular disease area.

Liver Diseases

CsA has proved to be hepatoprotective in several disease models (treatment with ethanol, with lipopolysaccharide of *E. coli* after liver sensitization with heat-inactivated *Propionibacterium* acnes or D-galactosamine, or with lipopolysaccharide alone; treatment with anti-Fas antibody, diclofenac or acetaminophen; in liver ischemia-reperfusion; in a model of α -1 antitrypsin deficiency with liver injury and mitochondrial autophagy; and in one cohort of cases of human fulminant viral hepatitis). It is difficult to assess whether the protective effects should be ascribed to PTP desensitization or rather to inhibition of cytokine production and lymphocyte activation. We systematically investigated this problem in the rat, and found that under conditions of maximal PTP desensitization by CsA *in vivo* rats were protected from the otherwise hepatotoxic effects of lipopolysaccharide plus D-galactosamine, which act through liver sensitization to the pro-apoptotic effects of tumor necrosis factor- α (TNF α).

The PTP is also involved in at least one model of hepatocarcinogenesis in rodents, that is, feeding with the arylamine 2-acetylaminofluorene (AAF), which causes onset of liver tumors within 30–50 weeks after a sequence of alterations that resembles the clinical course of chronic hepatitis, with hepatocyte damage and fibrosis gradually evolving into cirrhosis. Hepatic metabolism of AAF generates compounds that are responsible for the toxic and mutagenic effects of AAF including 2-nitrosofluorene, a redox cyler that drains electrons from the respiratory chain and causes opening of the PTP in isolated mitochondria. Very early into AAF feeding an adaptive response takes place, which desensitizes the PTP and the liver mitochondrial apoptotic pathway to TNF α *in vivo*. The adaptive response is of epigenetic nature, and represents a mitochondrial tumor-promoting event centered on the PTP that may contribute to the selection of resistant hepatocytes in the population of chemically transformed cells.

Neurological Diseases

Two convincing sets of results link the PTP to neuropathology in experimental models of neuronal injury. The first is based on the protective effects of CsA in hyperglycemia, hypoglycemia, ischemia, and trauma, in cell death following facial motoneuron axotomy in neonatal rodents, and in photoreceptor apoptosis. The second is based on mice with genetic inactivation of the *Ppif*^{-/-} gene encoding for CyP-D, which displayed a decrease of the emispheric damage that follows middle cerebral artery occlusion; a very striking resistance to axonal damage in a mouse model of multiple sclerosis; and prolonged survival after crossing into a mouse strain mimicking the familial variant of amyotrophic lateral sclerosis. This set of experiments is important because it demonstrates that the neuroprotective effects cannot be due to inhibition of calcineurin. Lack of involvement of calcineurin had also been suggested by the findings that the immunosuppressant FK506 (which inhibits calcineurin but not the PTP) failed to protect mitochondria and neurons of the dentate gyrus against hypoglycemic damage, and that *N*-methyl-Val-4-CsA, which inhibits the PTP but not calcineurin, afforded significant neuroprotection in a rat model of transient focal ischemia.

Muscle Diseases

The hypothesis that Ca^{2+} -dependent mitochondrial dysfunction plays a role in muscular dystrophies has been put forward 30 years ago. An important development in this field has been the demonstration that the PTP plays a key role in the pathogenesis of muscular dystrophy caused by collagen VI deficiency. Inherited mutations of collagen VI genes cause two muscle diseases in humans, Bethlem myopathy and Ullrich congenital muscular dystrophy. Collagen VI-deficient (*Col6a1*^{-/-}) mice display a muscle phenotype strongly resembling Bethlem myopathy, with loss of contractile strength associated with ultrastructural alterations of sarcoplasmic reticulum and mitochondria, and spontaneous apoptosis. These defects are unexpectedly due to inappropriate PTP opening because they could be normalized by treatment with CsA or with Debio 025, a cyclophilin inhibitor that does not affect calcineurin, or by crossing with *Ppif*^{-/-} mice devoid of CyP-D. Pharmacological treatment and the genetic cure rescued mitochondrial function, muscle ultrastructural defects, and dramatically decreased the number of apoptotic nuclei *in vivo* in what represents the first successful pharmacological treatment of an animal model of a genetic muscle disease.

The same mitochondrial defect was found in myoblast cultures from patients with Ullrich congenital muscular dystrophy and Bethlem myopathy, findings that paved the way for a pilot clinical trial where a dramatic short-term improvement of mitochondrial function, decreased apoptotic rates, and onset of muscle regeneration could be demonstrated in five patients. Assessing whether long-term treatment also leads to a clinical improvement will be the next step, and one that raises some hope for a therapy of this and possibly other congenital muscular dystrophies.

See also: [Bioenergetics: Calcium, Biological Fitness of; Calcium Oscillations; Calcium Transport in Mitochondria; Cell Death by Apoptosis and Necrosis; Giant Mitochondria \(Megamitochondria\); Intracellular Calcium Waves; Membrane Transport, General Concepts; Mitochondrial Calcium Transport: Historical Aspects; Nuclear Calcium in Synapse-to-Nucleus Communication: Common Regulator of](#)

[Persistent Adaptations in the Nervous System; Ryanodine Receptor Calcium Ion Channels.](#)

Further Reading

- Basso E, Petronilli V, Forte MA, and Bernardi P (2008) Phosphate is essential for inhibition of the mitochondrial permeability transition pore by cyclosporin A and by cyclophilin D ablation. *Journal of Biological Chemistry* 283(39): 26307–26311.
- Bernardi P (1999) Mitochondrial transport of cations: Channels, exchangers and permeability transition. *Physiological Reviews* 79: 1127–1155.
- Bernardi P, Krauskopf A, Basso E, et al. (2006) The mitochondrial permeability transition from *in vitro* artifact to disease target. *FEBS Journal* 273(10): 2077–2099.
- Broekemeier KM, Dempsey ME, and Pfeiffer DR (1989) Cyclosporin A is a potent inhibitor of the inner membrane permeability transition in liver mitochondria. *Journal of Biological Chemistry* 264(14): 7826–7830.
- Crompton M, Ellinger H, and Costi A (1988) Inhibition by cyclosporin A of a Ca^{2+} -dependent pore in heart mitochondria activated by inorganic phosphate and oxidative stress. *Biochemical Journal* 255(1): 357–360.
- Giorgio V, Bisetto E, Soriano ME, et al. (2009) Cyclophilin D modulates mitochondrial F₀F₁-ATP synthase by interacting with the lateral stalk of the complex. *Journal of Biological Chemistry* 284(49): 33982–33988.
- Giorgio V, Soriano ME, Basso E, et al. (2010) Cyclophilin D in mitochondrial pathophysiology. *Biochimica et Biophysica Acta* 1797(6–7): 1113–1118.
- Gunter TE and Pfeiffer DR (1990) Mechanisms by which mitochondria transport calcium. *American Journal of Physiology* 258(5 Pt 1): C755–C786.
- Halestrap AP and Davidson AM (1990) Inhibition of Ca^{2+} -induced large-amplitude swelling of liver and heart mitochondria by cyclosporin is probably caused by the inhibitor binding to mitochondrial-matrix peptidyl-prolyl *cis-trans* isomerase and preventing it interacting with the adenine nucleotide translocase. *Biochemical Journal* 268(1): 153–160.
- Haworth RA and Hunter DR (1979) The Ca^{2+} -induced membrane transition of rat liver mitochondria. II. Nature of the Ca^{2+} trigger site. *Archives of Biochemistry and Biophysics* 195(2): 460–467.
- Hunter DR and Haworth RA (1979) The Ca^{2+} -induced membrane transition in mitochondria. I. The protective mechanisms. *Archives of Biochemistry and Biophysics* 195(2): 453–459.
- Hunter DR and Haworth RA (1979) The Ca^{2+} -induced membrane transition in mitochondria. III. Transitional Ca^{2+} release. *Archives of Biochemistry and Biophysics* 195(2): 468–477.
- Kang BH, Plescia J, Dohi T, et al. (2007) Regulation of tumor cell mitochondrial homeostasis by an organelle-specific Hsp90 chaperone network. *Cell* 131(2): 257–270.
- Kokoszka JE, Waymire KG, Levy SE, et al. (2004) The ADP/ATP translocator is not essential for the mitochondrial permeability transition pore. *Nature* 427(6973): 461–465.
- Szabó I and Zoratti M (1991) The giant channel of the inner mitochondrial membrane is inhibited by cyclosporin A. *Journal of Biological Chemistry* 266(6): 3376–3379.

Mitochondria: Source and Target of Free Radicals

A Boveris, University of Buenos Aires, Buenos Aires, Argentina

E Cadenas, University of Southern California, Los Angeles, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Apoptosis The active process by which cells execute a series of processes leading to cell death in a way that is not harmful for the tissue and the organism. It is opposite to inflammation and necrosis, processes in which toxic products reach the neighbor cells.

Free radicals Molecules, better understood as chemical species, having an unpaired electron in their external orbitals. The feature makes the chemical species unstable and highly reactive with low specificity. However, some biologically produced free radicals, such as superoxide radicals (O_2^-) and nitric oxide (NO), are relatively stable and selective in their reactions. Other biologically produced radicals, such as hydroxyl and peroxy ($HO^\cdot$, $ROO^\cdot$) radicals,

are highly reactive and unspecific. The point near the chemical symbols indicates the unpaired electron.

Hydrogen peroxide The product of the bivalent reduction of the oxygen molecule ($HOOH$; H_2O_2); it is not a free radical, but originates $HO^\cdot$ radical upon interaction with the omnipresent Fe^{2+} .

Mitochondrial electron transfer chain A series of biological and organic molecules embedded in the inner mitochondrial membrane having chemical groups able to undergo reversible cycles of reduction/oxidation. The mitochondrial respiratory chain channels the reducing equivalents from food components and derived products to the oxygen molecule and constitutes the biochemical pathway supporting respiration.

Introduction

More than two centuries after Lavoisier's description of animal respiration as a biological oxidation, there is still a fresh and active interest in the mechanisms and the regulation of the biochemical processes that link the oxidation of food components to energy production. There is a current recognition that the regulation of biological oxidations and of energy production and expenditure underlies the whole process of life and death in cells and organisms. At the foundation of biochemistry, in the 1920s, there was a controversy on how the hydrogen atoms of organic molecules end in the H_2O molecule, with two views: hydrogen activation and oxygen activation, championed by Wieland and Warburg, respectively. The first view had H_2O_2 as an important intermediate and the second one supported the idea of H_2O as the direct product of O_2 reduction. The monumental contribution of Warburg, identifying cytochrome oxidase as the enzyme-catalyzing O_2 uptake in biological systems, ended the dispute. However, the concept of the partial reduction of O_2 in biological systems emerged again in 1946 with Michaelis' ideas of univalent electron transfer. The four univalent steps of electron transfer to the O_2 molecule would produce, according to Michaelis, hydroperoxyl radical ($HO_2^\cdot$, the protonated base of superoxide radical, O_2^-), hydrogen peroxide (H_2O_2), hydroxyl radical ($HO^\cdot$), and water (H_2O). These concepts were strengthened when Gerschman postulated in 1954 that oxygen-free radicals were the common biochemical mechanisms of oxygen and radiation toxicity. At that time, $HO^\cdot$ and $HO_2^\cdot$ were known as toxic chemical species produced in radiation chemistry and biology, but by no means it was admitted that these highly reactive free radicals could be produced in physiological conditions in animals and humans. Gerschman exposed mice to hyperbaric O_2 to increase the tissue levels of oxygen-free radicals and found synergism

between X-radiation and hyperbaric O_2 in decreasing the survival time of exposed animals.

The transcendental and landmark discovery of superoxide dismutase (SOD) by McCord and Fridovich in 1969 introduced a new dynamism to the understanding of free radical reactions in biological systems. The existence of the enzyme, first described in erythrocytes and later in bacteria and in mammalian organs, implied that O_2^- , its substrate, was a normal metabolite. In a few years, Fridovich advanced the dogma that O_2^- and H_2O_2 interact in biological systems and produce the highly toxic $HO^\cdot$, being the three chemical species responsible for oxygen toxicity and biological damage. The elegant classification of bacteria as aerobes (protected by SOD and catalase), aerotolerant anaerobes (protected by SOD but not by catalase), and anaerobes (without protection by SOD and catalase) paved the way for the extension of the 'Fridovich dogma' to mammalian biology and human pathology. Simultaneously, mitochondria were recognized in 1971–74 as an active source of H_2O_2 that diffuses to the cytosol. A series of researchers, that is, Loschen, Oshino, Flohé, Boveris, Azzi, and Richter, working under the leadership of Britton Chance at or associated with the Johnson Research Foundation (Philadelphia, USA) participated in the development of the concept. The idea collided with de Duve's 'peroxisome concept', in which H_2O_2 was produced and utilized in the intraperoxisomal space and avoided as a cellular metabolite. Shortly after Boveris and Cadenas and Azzi and coworkers simultaneously described O_2^- as the stoichiometric precursor of mitochondrial H_2O_2 , Boveris and Cadenas identified the autooxidation of ubiquinone (UQH) as an active source of O_2^- . By 1980, mitochondria were recognized as the more generalized and quantitatively more important source of O_2^- in mammalian cells and organs. Other sources of O_2^- , such as liver endoplasmic reticulum, activated neutrophils, and xanthine

oxidase, were described but their physiological importance remained restricted to specific conditions.

The existence of the family of SODs, with the Mn-SOD isoenzyme specific for the mitochondrial matrix, and the mitochondrial production of $O_2^{\cdot-}$ in all aerobic cells, were the two concepts that gave the strongest support for the participation of oxygen-free radicals in mammalian and human physiology, pathology, and aging.

The mechanistic concept that was and is currently accepted for the deleterious action of oxygen-free radicals in biological systems is called the Fenton/Haber-Weiss pathway, which involves a first step with two $O_2^{\cdot-}$ -dependent reactions: the Mn-SOD-catalyzed reaction with production of H_2O_2 and the reduction of cellular or ferritin- Fe^{3+} with production of Fe^{2+} . The second step is the reaction between H_2O_2 and Fe^{2+} that involves the homolytic scission of the peroxide bond and the generation of the highly toxic $HO^{\cdot}$ radical.

The discovery of the generation of nitric oxide (NO) in mammals as a signaling molecule with roles as intercellular messenger in the cardiovascular system, as neurotransmitter and as component of the signaling and cytotoxic mechanisms in the immunological system, along with the identification of NO as a free radical, produced a revolution in the concepts of free radical metabolism and signaling in mammals. The linear Fenton/Haber-Weiss pathway evolved to a cycles-shaped scheme for the pathways of free radical reactions in mammalian cells and organs (Figure 1).

The Mitochondrial Production of $O_2^{\cdot-}$ and H_2O_2

Mitochondria were recognized as subcellular organelles by application of supravital stainings by the end of nineteenth century, and as the site of the biochemical process of oxidative phosphorylation in the 1950s, after the availability of electron microscopy and the modern methods of cell fractionation. Mitochondrial isolation permitted the identification of complexes I–V as constitutive proteins of the inner

mitochondrial membrane and the role of the inner membrane as the coupling place for the transduction of the redox chemical energy of reduced nicotinamide adenine dinucleotide (NADH) oxidation into adenosine triphosphate (ATP) high-energy phosphate bonds. By 1955, Chance and Williams produced a complete description of the redox states of the components of the mitochondrial respiratory chain, and defined the operational concepts of mitochondrial metabolic state and of respiratory control. The latter concept elegantly describes, as the state 4 to state 3 transition, the phenomenon by which adenosine diphosphate (ADP) regulates mitochondrial electron transfer and respiration. State 4, with availability of respiratory substrate but not of ADP, is described as ‘controlled or resting respiration’, whereas state 3, with ample respiratory substrate and ADP availability, is defined as the ‘active respiration’, that is, the maximal physiological rate of ATP production and O_2 consumption. By determination of the rates of mitochondrial O_2 uptake in both state 4 and state 3, and of the O_2 uptake in tissue slices and in perfused organs, it is estimated that mammalian tissue mitochondria under basal physiological conditions are mostly (60–70%) in state 4, whereas the rest (30–40%) are in state 3, thus meaning that in the physiological setting of basal metabolism there is only a 30–40% utilization of the ATP-producing capacity.

Heart and liver mitochondria were identified as sources of H_2O_2 , in a process modulated by the mitochondrial state 4/state 3 transition; being the rate of H_2O_2 generation in state 4 markedly higher (4–6 times) than in state 3. This fact indicates that a component of the respiratory chain, markedly changing its reduction level in the state 4/state 3 transition is the H_2O_2 generator. The rates of H_2O_2 production of mitochondria isolated from mammalian organs are in the range of $0.6\text{--}0.8\text{-nmol } H_2O_2 \times \text{min}^{-1} \times \text{mg protein}^{-1}$ for state 4 and $0.05\text{--}0.15\text{-nmol } H_2O_2 \times \text{min}^{-1} \times \text{mg protein}^{-1}$ for state 3. In the case of rat heart and liver, mitochondrial H_2O_2 production accounts for about 0.5% of the tissue O_2 uptake.

The mechanistic option for mitochondrial O_2 reduction and H_2O_2 production, either one 2-electrons step to yield H_2O_2 or two 1-electron steps to yield $O_2^{\cdot-}$ as intermediate followed by dismutation, was solved by the fact that submitochondrial particles devoid of Mn-SOD show a ratio of about 2:1 in the rates of $O_2^{\cdot-}$ and H_2O_2 productions, respectively. This ratio and the intramitochondrial localization of Mn-SOD established $O_2^{\cdot-}$ as the precursor of mitochondrial H_2O_2 . The main part of mitochondrial $O_2^{\cdot-}$, about 55%, is vectorially released into the mitochondrial matrix, where it encounters Mn-SOD. Steady-state concentrations of 0.08–0.2-nM $O_2^{\cdot-}$ and 5–50-nM H_2O_2 are estimated for the mitochondrial matrix, with a content of 3–10- μM Mn-SOD. The release of $O_2^{\cdot-}$ to the mitochondrial intermembrane space, accounting for 30% of the mitochondrial $O_2^{\cdot-}$ production, was also observed in functional relationship with the Cu, Zn-SOD of this compartment. The $O_2^{\cdot-}$ released into the intermembrane space could reach the cytosol through the voltage-dependent anion channel (VDAC). Hence, mitochondria may be considered as an eventual effective source of cytosolic $O_2^{\cdot-}$, with this species contributing to the univalent redox regulation of cell signaling.

Two main $O_2^{\cdot-}$ -generating reactions have been described for the mitochondrial respiratory chain. In both cases,

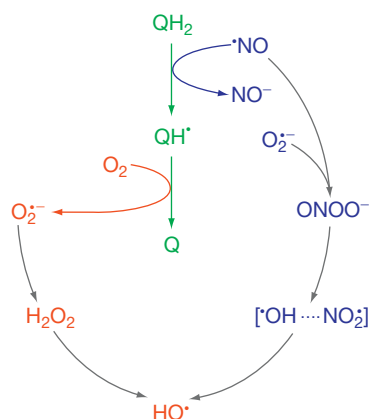


Figure 1 The biochemical free radical chain reactions encompassed by the redox transitions of $O_2^{\cdot-}$ (red), NO (blue), and ubiquinone (green). $O_2^{\cdot-}$, superoxide anion; H_2O_2 , hydrogen peroxide; $HO^{\cdot}$, hydroxyl radical; NO , nitric oxide; $ONOO^{\cdot}$, peroxynitrite; UQH_2 , ubiquinol; $UQH^{\cdot}$, ubisemiquinone; UQ , ubiquinone.

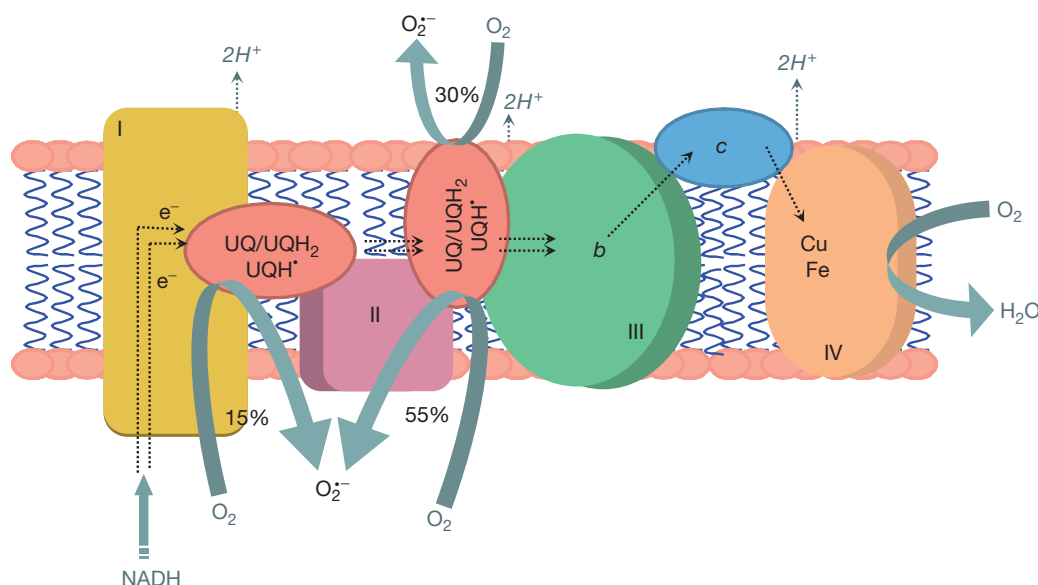


Figure 2 Sites of O_2^- production in the mitochondrial respiratory chain. UQH₂, ubiquinol; UQ, ubiquinone; and UQH[•], ubisemiquinone. The percentages close to O_2^- indicate the relative contribution of each chemical pathway to total mitochondrial O_2^- production.

the intermediate are semiquinones of redox pairs of the respiratory chain: the pair ubiquinol/ubiquinone and the pair reduced flavin mononucleotide (FMNH₂)/FMN component of NADH dehydrogenase. The two semiquinones, UQH[•] and FMNH[•], respectively, are collisionally and nonenzymatically autooxidized by molecular O_2 yielding O_2^- as product. In the UQH[•] autooxidation, the chemical process is called the Boveris–Cadenas reaction ([reaction \[1\]](#)), with a second-order reaction constant of $8 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$, and with the reaction considered the main mitochondrial O_2^- -producing pathway and a pacemaker of mitochondrial turnover and of the aging process.

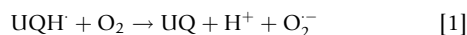


Figure 2 illustrates the mitochondrial reactions that produce O_2^- as well as their vectorial release to the cytosol and to the intermembrane space. Autooxidation of other components of the mitochondrial respiratory chain are thermodynamically spontaneous, but these reactions are not kinetically fast enough to be important on a quantitative consideration.

The spontaneous surface chemiluminescence of *in situ* organs supported the idea of mitochondria as the main physiological source of oxidizing free radicals. The assay is based in the production of the chemiluminescent species singlet oxygen (1O_2), a downstream byproduct of lipid peroxidation generated in the reaction of $ROO^{\bullet}$ radical recombination ([Figure 1](#)). *In situ* liver and brain chemiluminescence increased under conditions, such as hyperbaric O_2 , ischemia–reperfusion, and pharmacological treatments, following the quantitative effect of these conditions on mitochondrial O_2^- generation.

The Mitochondrial Production of 'NO

The extraordinary research activity on biological 'NO production at the beginning of the 1990s identified three different

nitric oxide synthases (NOSs): neuronal NOS (NOS-I or nNOS), inducible or macrophage NOS (NOS-II or iNOS), and endothelial NOS (NOS-III or eNOS). Immunocytochemical assays with antibodies against eNOS gave evidence of an NOS located in mitochondria, which was named mtNOS and advanced as a regulator of mitochondrial and cellular respiration. The determination of mtNOS enzymatic activity with 'NO as product was elusive for a few years, until two independent groups, Ghafourifar and Richter in Zurich, and Giulivi, Poderoso, and Boveris in Los Angeles and Buenos Aires, succeeded in 1997–98 in measuring a 'NO production by rat liver mitochondria and mitochondrial fragments. The original observations in liver mitochondria were later extended to mitochondria from brain, heart, thymus, kidney, and skeletal muscle, thus dissipating the skepticism on mtNOS that stemmed from the supposition of contamination during cell fractionation. Mitochondrial 'NO production is carried out by mtNOS, a classic NOS, that requires NADPH ($K_M = 12\text{--}20 \mu\text{M}$), arginine ($K_M = 5\text{--}50 \mu\text{M}$), O_2 ($K_M = 35\text{--}40 \mu\text{M}$), and Ca^{2+} for enzyme activity. Calmodulin, tetrahydrobiopterine, and thiols increase the enzymatic activity. Intramitochondrial concentrations of NADPH, arginine, O_2 , and Ca^{2+} are in excess or in the range needed for enzymatic activity. The methodology currently used to measure mtNOS activity determines 'NO production rates of $0.80\text{--}1.60\text{-nmol 'NO}_x \text{ min}^{-1} \times \text{mg protein}^{-1}$ in mitochondria and submitochondrial preparations.

In a pivotal contribution in 2002, Giulivi and coworkers sequenced the 1429 amino acids of rat liver mtNOS and found it identical to nNOS splice variant α , mirystoylated and phosphorylated in posttranslational processes. Transcripts corresponding to nNOS- α were found in liver, brain, heart, kidney, skeletal muscle, lung, testis, and spleen. Hence, it is now clear that mtNOS is a constitutive protein of the inner mitochondrial membrane with the reaction center facing the matrix side and with its substrates present in the mitochondrial matrix. Interestingly, a series of physiological situations were

reported to regulate the expression and activity of mtNOS that was reported upregulated by acute and chronic hypoxia, by insulin, and by cold exposure, and downregulated by thyroxine and angiotensin II. Pharmacological administration of enalapril increased heart and kidney mtNOS activity and treatments with haloperidol and chlorpromazine decreased brain mtNOS activity.

Intramitochondrial steady-state concentrations of $\cdot\text{NO}$ are calculated as 200–370 nM (200 nM in state 3 and 370 nM in state 4) and a release of 29 nM $\cdot\text{NO}$ was electrochemically measured after supplementation of a single mitochondrion with Ca^{2+} . In the tissues under physiological conditions at 20- μM O_2 , the $[\text{O}_2]/[\cdot\text{NO}]$ ratio is 100–55, which would inhibit cytochrome oxidase by 30–45 %. In the mitochondrial matrix, $\cdot\text{NO}$ and $\text{O}_2^{\cdot-}$ metabolism are linked by the very fast reaction between the two free radicals to produce ONOO^- . This oxidative utilization of $\cdot\text{NO}$ is the main (60–70%) pathway of $\cdot\text{NO}$ catabolism, but at the same time it is only a minor part (15%) of intramitochondrial $\text{O}_2^{\cdot-}$ utilization. The reductive utilization of $\cdot\text{NO}$ by ubiquinol and cytochrome oxidase provides a minor (20%) pathway of $\cdot\text{NO}$ catabolism. The cellular conditions in which $\cdot\text{NO}$ diffuses from mitochondria to cytosol, as well as the conditions in which $\cdot\text{NO}$ diffuses from cytosol to mitochondria are unsolved questions for the complex process of intracellular signaling. The available data support a bipolar distribution of NOS in the cells, with an isoenzyme in mitochondria (mtNOS) and the other in the cytosol (endoplasmic reticulum or synaptic vesicles). **Figure 3** shows a view of the integrated mitochondrial metabolism of $\text{O}_2^{\cdot-}$ and $\cdot\text{NO}$.

The main effect of $\cdot\text{NO}$ on mitochondrial electron transfer was reported in 1994 by two British research groups, Brown and Cooper and Cleeter et al., that described the inhibition of brain and muscle cytochrome oxidase (complex IV) activity by

low- $\cdot\text{NO}$ concentrations in a reversible and O_2 -competitive biochemical process. The observation was rapidly confirmed by other groups using $\cdot\text{NO}$ donors and pure $\cdot\text{NO}$ in liver, heart, and brown fat mitochondria. Nitric oxide at 0.1–0.2 μM decreases cytochrome oxidase activity and mitochondrial respiration to one-half. The inhibition is reversible by dilution, hemoglobin supplementation, or exposure to $\text{O}_2^{\cdot-}$. The $\text{O}_2^{\cdot-}$ competitive inhibition of cytochrome oxidase renders the effect more marked at low O_2 concentrations. A mathematical model by Antunes et al. explains the main features of the $\cdot\text{NO}$ effects on mitochondrial respiration: dependence with $\cdot\text{NO}$ concentration, dependence with O_2 concentrations (as the $\cdot\text{NO}/\text{O}_2$ ratio), and a more marked effect of $\cdot\text{NO}$ in state 3 than in state 4. A second sensitive point in the mitochondrial respiratory chain is complex III (ubiquinol–cytochrome *c* reductase); half inhibition of electron transfer between cytochromes *b* and *c* occurs at 0.2–0.4- μM $\cdot\text{NO}$ with the interesting property of enhancing the production of $\text{O}_2^{\cdot-}$ and H_2O_2 in submitochondrial particles and in mitochondria. A third sensitive respiratory component is NADH-dehydrogenase (complex I) and in this case, ONOO^- appears as the effective inhibitory agent in an irreversible process. Complex I inactivation is of great interest in pathological conditions as it has been reported in a series of neurodegenerative diseases, originally in Parkinson's disease, and referred as 'complex I syndrome'.

Figure 4 illustrates the direct and indirect effects of $\cdot\text{NO}$, in the latter case through ONOO^- formation, on the mitochondrial respiratory chain. The reversible effects of $\cdot\text{NO}$ on cytochrome oxidase and complex III are two aspects of the regulation of mitochondrial respiration by $\cdot\text{NO}$: the direct inhibition of cytochrome oxidase, and the enhancement of $\text{O}_2^{\cdot-}$ production to remove the respiratory inhibition, as it may become a physiological need in the case of high steady-state concentrations of $\cdot\text{NO}$, such as in inflammation and in

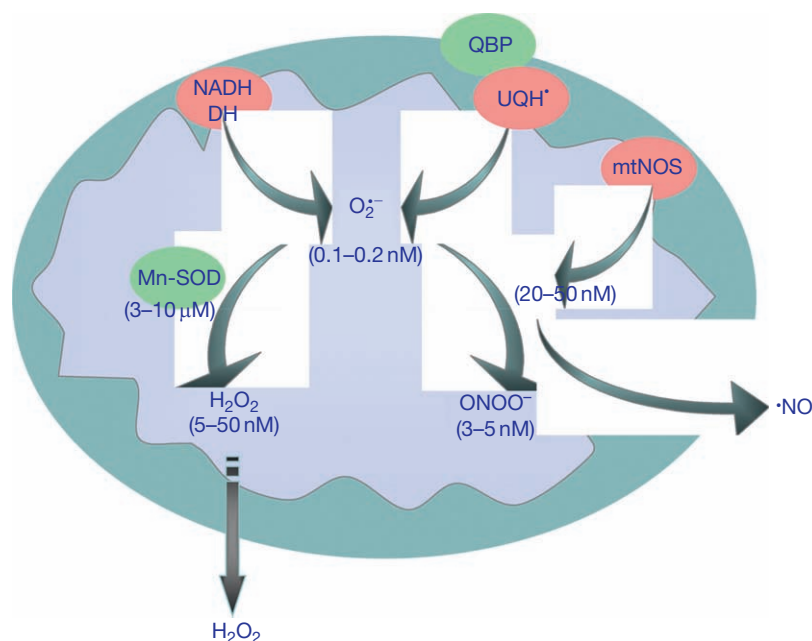


Figure 3 Metabolism of $\text{O}_2^{\cdot-}$ and $\cdot\text{NO}$ in the mitochondrial matrix. The numbers below the symbols indicate physiological steady-state concentrations. The arrows reaching outside the mitochondrion indicate diffusion of H_2O_2 and $\cdot\text{NO}$ to the cytosol in the indicated conditions.

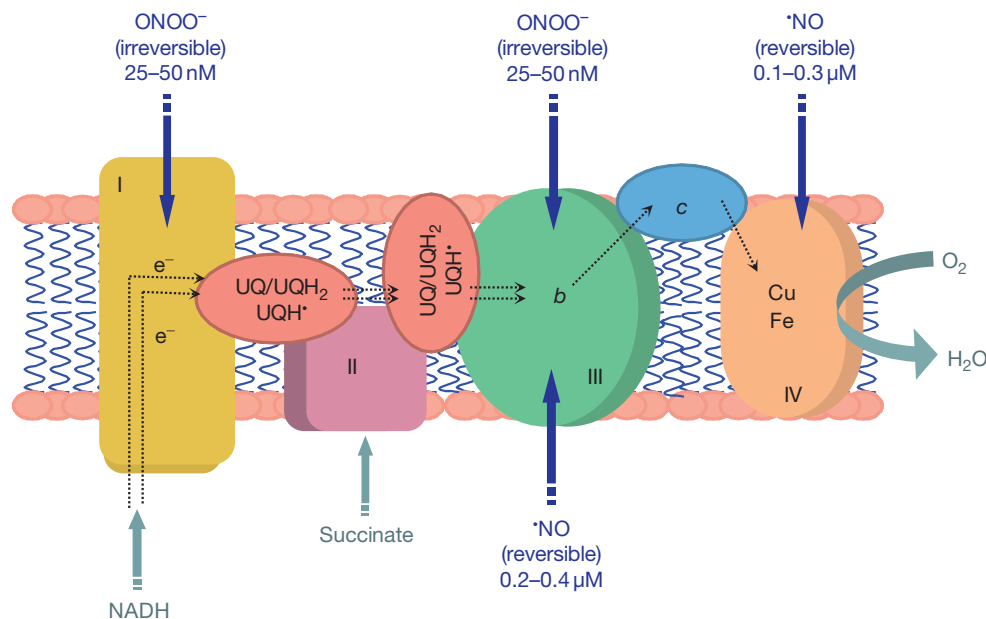


Figure 4 Sites of action of $\cdot\text{NO}$ and ONOO^- on the mitochondrial respiratory chain. Reversible and irreversible indicate the types of inhibition that are understood as regulatory or toxic processes, respectively. The indicated steady-state concentrations were determined or estimated to reduce enzymatic activities to one half, at a concentration of 100–150 μM O_2 .

ischemia–reperfusion. The irreversible effects of ONOO^- on complexes I and III are related to situations in which sustained high levels of ONOO^- lead to mitochondrial dysfunction and apoptosis.

Concerning the important question of whether or not mtNOS enzymatic activity regulates mitochondrial functions under physiological conditions, several reports described the modulation of mitochondrial O_2 uptake and H_2O_2 production by the activity of mtNOS in isolated mitochondria. The current view is that mtNOS does regulate mitochondrial respiration. Hence, mtNOS, cytochrome oxidase, and $\text{F}_1\text{-ATPase}$ are the three regulatory proteins of cellular O_2 uptake and energy production. The biochemical activity of mtNOS (expressed as $\text{nmol NO} \times \text{min}^{-1} \times \text{mg protein}^{-1}$) is usually measured as the mitochondrial $\cdot\text{NO}$ production (sensitive to a competitive inhibitor of NOS, such as NG-monomethyl-arginine (NMMA) or nitroarginine (NNA)). The regulatory activity of mtNOS, named the functional mtNOS activity, is best expressed by the difference in the rates of O_2 uptake or H_2O_2 production (in $\text{nmol of O}_2 \text{ or H}_2\text{O}_2 \times \text{min}^{-1} \times \text{mg protein}^{-1}$) between mitochondria supplemented with arginine and SOD, condition that provides the highest $\cdot\text{NO}$ steady state, and mitochondria in the presence of HbO_2 and of a NOS competitive inhibitor (NMMA or NNA), that gives the lowest $\cdot\text{NO}$ steady-state level. The regulatory capacity of $\cdot\text{NO}$ is estimated as the ratio of the rates of regulated O_2 uptake or H_2O_2 production over the rate of $\cdot\text{NO}$ production; the ratio is ~ 70 for O_2 uptake and ~ 0.4 for H_2O_2 production. The activity of mtNOS, that is, the steady-state level of $\cdot\text{NO}$ in isolated mitochondria and cells, markedly increases the $[\text{O}_2]_{0.5}$ of mitochondrial O_2 uptake; for mitochondria, the values changed from 1.1- μM O_2 in the case of inactive mtNOS to 2.2- μM O_2 in the case of an active mtNOS. However, it remains to be established to what extent these effects, observed with isolated mitochondria occur under physiological

conditions in myoglobin-containing tissues and in blood-perfused organs, due to the high affinity of the two mentioned hemoproteins for $\cdot\text{NO}$.

The Physiological Role of the Mitochondrial Production of $\text{O}_2^{\cdot-}$, H_2O_2 , and $\cdot\text{NO}$

The early recognition that $\text{O}_2^{\cdot-}$ and H_2O_2 are both able to initiate reactions harmful to cell and tissues is now complemented by the concept that $\text{O}_2^{\cdot-}$, H_2O_2 , and $\cdot\text{NO}$ are carefully regulated metabolites capable of signaling to the regulatory receptors of the biochemical and genetic systems of the cell. Nitric oxide was initially recognized as an intercellular messenger and later as an intracellular regulator. At present, the three chemical species, $\text{O}_2^{\cdot-}$, H_2O_2 , and $\cdot\text{NO}$, are considered to participate in integrated processes of intracellular regulation and in intercellular signaling, communication, and cytotoxicity. Although $\text{O}_2^{\cdot-}$ is considered a metabolite confined to the intramitochondrial space, with minor release to cytosol through VDAC, both H_2O_2 and $\cdot\text{NO}$ are permeable through the double mitochondrial membrane and plasma membrane and are likely regulators of cytosolic and neighbor cell functions. Hydrogen peroxide and $\cdot\text{NO}$ share the chemical property of being uncharged and the biological property of being highly diffusible through biological membranes. At variance, $\text{O}_2^{\cdot-}$ and ONOO^- are charged and nondiffusible molecules. Consequently, the first two species are suitable for cellular and intercellular signaling and the two latter for regulation and cytotoxicity in confined spaces. The regulation of gene expression and intercellular communication is just starting to be understood as a vital mechanism in health and disease.

The fine regulation by H_2O_2 of cell function was advanced by Antunes and Cadenas in 2001 when they treated Jurkat

T-cells with controlled H_2O_2 concentrations provided by the glucose–glucose oxidase system. Steady-state concentrations below $0.7\ \mu\text{M}$, H_2O_2 place cells in a proliferative state, whereas at $1.0\text{--}3.0\ \mu\text{M}$, H_2O_2 cells develop programmed cell death, and at levels higher than $3.0\ \mu\text{M}$ H_2O_2 cells undergo necrosis.

There is evidence that H_2O_2 and NO are able to upregulate mitogen-activated protein kinases (MAPKs), the widespread components of intracellular phosphorylation and dephosphorylation signaling cascades involved in cell survival, proliferation, differentiation, and death. The interactions are complex and involve intracellular glutathione and ONOO^- . Both, H_2O_2 and NO diffusing from mitochondria to the cytosol are considered a pleiotropic signal involved in the regulation of c-Jun-N-terminal kinase (JNK)-dependent signaling, with H_2O_2 upregulating JNK1 and NO inactivating JNK1, likely by S-nitrosylation). The concept of narrow ranges of intracellular messenger molecules for different or opposite biological actions, as observed for H_2O_2 levels in the proliferation/apoptosis transition, is now applied to effector systems with two likely regulators, H_2O_2 and NO , and with the consequence of different responses for the integration of the two signals in one response (Figure 5).

It has been understood for the last 40 years that excess generation of O_2^- and H_2O_2 leads to oxidative stress and harmful biological situations. Nitric oxide has been added to these two species in the last two decades. The classical concept of oxidative stress, as brought forward by Sies, implies an imbalance between oxidants and antioxidants, in favor of the former. The situation is understood as reversible, especially by supplementation with antioxidants, and also as leading, sooner or later, to biological dysfunction. The concept is applicable to subcellular organelles, cells, organs, systems, and whole organisms. From a biochemical perspective, oxidative stress is determined by an increased steady-state level of at least one of the following chemical species, O_2^- , H_2O_2 , NO ,

ONOO^- , $\text{HO}^\cdot$, $\text{ROO}^\cdot$, and $^1\text{O}_2$, leading, according to the mass action law, to increased production and levels of the other oxidizing species downstream of the free radical chain reaction pathway (Figure 1) and to biochemical damage.

Oxidative Stress and Dysfunctional Mitochondria: Mitochondria as Targets of Free Radicals

In the last few years, the concept of dysfunctional mitochondria was applied to mitochondria isolated from tissues showing oxidative stress, biological damage, physiological dysfunction or pathological conditions. Mitochondria are considered dysfunctional by comparison of their functions and properties with normal control mitochondria. The classic mitochondrial properties, that reflect the energy-producing physiological function, are: the state 3 rate of O_2 uptake, the respiratory control, the membrane potential, and the ADP:O ratio. These parameters, when decreased, support the consideration of dysfunctional mitochondria. In addition, dysfunctional mitochondria frequently show increased rates of O_2^- and H_2O_2 production, a condition that contributes to the perpetuation of the dysfunctional condition.

Mitochondrial dysfunction has been observed in ischemia-reperfusion, endotoxic and septic shock, a series of pathological situations, aging, and neurodegenerative diseases. The reported characteristics are: decreases in the rates of state 3 respiration and ATP synthesis, in mitochondrial membrane potential, and in respiratory control, and increases in state 4 respiration, in mitochondrial size and fragility, and in the content of free-radical-mediated oxidation products of proteins, phospholipids, and DNA. The reported decreases in the rates of state 3 respiration and of mitochondrial electron transfer in the isolated respiratory complexes are in the range of 30–40%; it seems that more damaged mitochondria are not compatible with living cells.

The continuous production of oxidizing chemical species is the concept that explains the molecular mechanisms involved in the evolution from normal to dysfunctional mitochondria. The free radical O_2^- is continuously produced in the autooxidation of $\text{UQH}^\cdot$ and $\text{FMNH}^\cdot$ and mainly dismutates to H_2O_2 ; also, the free radical NO is continuously produced by mtNOS as regulator of respiration. The species ONOO^- is produced by the collision of O_2^- and NO in the fastest reaction known in biological systems ($k = 1.9 \times 10^{10}\ \text{M}^{-1}\ \text{s}^{-1}$), except for the neutralization reaction of H^+ and $\text{HO}^\cdot$ ($k = 14 \times 10^{10}\ \text{M}^{-1}\ \text{s}^{-1}$). Peroxynitrite is a nitrating and oxidizing species; when produced in excess, as in ischemia-reperfusion and inflammation, leads to tyrosine nitration of mitochondrial proteins. Intramitochondrial levels of 3–5-nM ONOO^- have been estimated for physiological conditions, and levels above 20–30 nM are considered cytotoxic. The free radicals $\text{HO}^\cdot$ and $\text{ROO}^\cdot$ are the products of the Fenton/Haber-Weiss reaction and of the process of lipid peroxidation. The oxidation products of phospholipids and proteins (thiobarbituric acid reactive substance (TBARS), 4-HO-nonenal and protein carbonyls) are constantly detected in dysfunctional mitochondria. Dysfunctional mitochondria appears almost equally driven by excessive NO and ONOO^- production and by excessive lipid peroxidation.

Selective inhibition of NADH-dehydrogenase (complex I) activity associated to increased lipid peroxidation and nitration

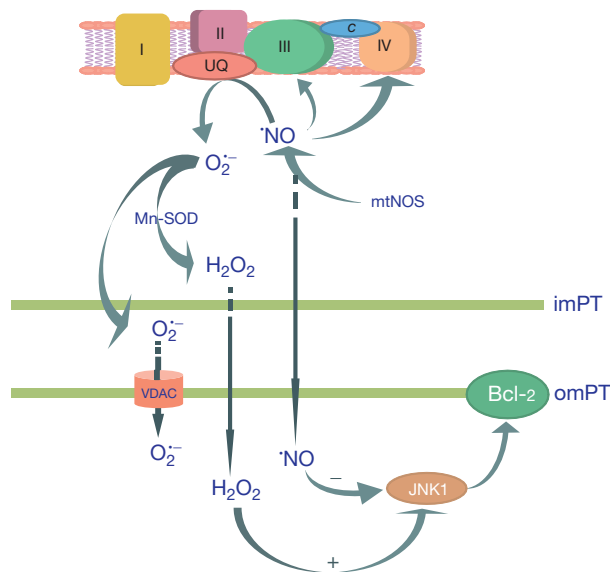


Figure 5 Mitochondrial production of oxygen- and nitrogen-centered radicals and cell signaling. The scheme shows the sites of action of NO in the respiratory chain, the ensuing production of O_2^- and of H_2O_2 , the release of these reactive species from the mitochondrion (controlled by VDAC in the case of O_2^-) and the effect on JNK signaling.

has been reported in mitochondria and submitochondrial preparations from dysfunctional organs, especially from experimental animals and human neurodegenerative diseases. This selective mitochondrial defect is understood as secondary to a sustained oxidative stress and was reported in aged humans and in patients with neurodegenerative disorders, among them Parkinson's and Alzheimer's diseases. The near-target production of $O_2^{\cdot-}$ (by NADH-dehydrogenase) and of $ONOO^-$ are considered determinant in the selective impairment of complex I in neurodegenerative diseases. The concept of complex I syndrome implies a decreased rate of state 3 mitochondrial respiration with NAD-linked substrates, a decreased complex I activity, and an increased content of proteins and phospholipids oxidation products. These characteristics describe the findings in animal models and in human patients of Parkinson's and Alzheimer's diseases and of dementia with Lewy bodies, among other neurodegenerative diseases.

The continuous mitochondrial production of free radicals and oxidizing species makes mitochondria the quantitatively most important source of these harmful species in the cell, and at the same time, the sensitivity of mitochondrial functions to increased levels of the oxidizing free radicals and related chemical species renders mitochondria the most sensitive cellular target for the deleterious action of free radicals. The rapid mitochondrial turnover ($t_{1/2}$ of 7–10 days in liver and of 30–40 days in heart and brain) clearly faster than cell turnover (40–50 days in liver and almost infinite in brain neurons) illustrates the double role of mitochondria as source and target of biologically harmful free radicals.

Mitochondrion-Dependent Apoptosis

The role of mitochondria in apoptosis is now well established; a specific mitochondrion-dependent pathway of cell death is differentiated from the death-receptor-mediated apoptosis. Mitochondrion-dependent apoptosis entails oxidative stress and is modulated by MAPK signaling through JNK activation associated with Bcl-2 downregulation, cytochrome *c* release, and the downstream activation of caspases. A very early activation of mtNOS has been reported in thymocyte-induced apoptosis. Thapsigargin supplementation of isolated thymocytes sequentially increases cytosolic Ca^{2+} and the activities of mtNOS and of the NOS of endoplasmic reticulum. A crosstalk between mitochondria and endoplasmic reticulum, in which Ca^{2+} and 1NO are the signals, is established in the initial phase of thymocyte apoptosis followed by the well-established mitochondrial changes that are characteristic of mitochondrion-mediated apoptosis: fall in mitochondrial membrane potential, mitochondrial dysfunction, opening of the membrane transition pores (MTPs), and cytochrome *c* release.

The central role of mitochondria in the cell death program is exerted during the first and reversible phase of early apoptosis through the markedly increased release of H_2O_2 and 1NO to either activate cytosolic JNK which attached to mitochondria catalyzes phosphorylation of Bcl-2 and Bcl- x_L and other not yet identified mitochondrial proteins. The JNK-mediated phosphorylations initiate the steps of the second and irreversible phase of apoptosis by opening the inner MTP and by releasing

Ca^{2+} and $O_2^{\cdot-}$ to the intermembrane space, followed by the opening of the outer MTP and release of cytochrome *c* to the cytosol. The next steps in the execution of the cell death program, such as apoptosome assembly and caspase activation, are independent of mitochondrial intervention.

The Mitochondrial Hypothesis of Aging – Aging and Neurodegenerative Diseases

In the last few years, a strong current of opinion has linked the processes of mitochondrial production of $O_2^{\cdot-}$ and H_2O_2 , with continuous oxygen toxicity, aging, life span, and neurodegenerative diseases. In the mitochondrial hypothesis of aging, mitochondria are considered as a biological clock for mitochondrial turnover, cell aging, and apoptosis. The free radical theory of aging was proposed by Harman in 1956, who after observing that supplementation of animal food with organic antioxidants increased the median life span in rats and mice, considered that aging is caused by the cumulative and random effects of free radicals that were cancelled by antioxidants. The antioxidant effect is now understood in terms of the continuous mitochondrial production of $O_2^{\cdot-}$ and 1NO as primary free radicals, with a whole series of derived and oxidizing reactive species, such as H_2O_2 , $ONOO^-$, $HO^{\cdot}$, $ROO^{\cdot}$, and 1O_2 . This continuous production supports the notion of mitochondria as pacemakers of cell and organ aging. As mentioned above, increased intramitochondrial steady-state concentrations of reactive oxygen and nitrogen species have been associated to the appearance of dysfunctional mitochondria in situations of acute oxidative stress, such as ischemia–reperfusion, and endotoxic or septic shocks and of chronic oxidative stress such as aging and neurodegenerative diseases. Increased levels of oxidative stress markers, such as 8-hydroxy-deoxyguanosine, protein carbonyls, hydroperoxides, and TBARS, have been consistently reported upon aging. Dysfunctional mitochondria with decreased state 3 respiration, membrane potential and respiratory control, and with increased size have been reported upon aging. Interestingly, some of the altered parameters were partially or near totally normalized by treatment with antioxidants, such as high doses of vitamin E and plant extracts.

It is noteworthy that some mitochondrial macromolecules appear selectively damaged upon aging. In this way, mitochondria are the pacemaker of aging, both as the main continuous source and as the main target of oxidizing free radicals. Ames and coworkers pioneered the field with the observation of increased mitochondrial DNA (mtDNA) damage, expressed as 8-hydroxy-deoxyguanosine in tissues and in urine, in old animals as compared with young ones. Other mitochondrial macromolecules, such as mitochondrial RNA (mtRNA) and the regulatory factor Bcl-2, have been reported damaged upon aging. Mitochondrial enzymes that are crucial for electron transfer and energy production, such as complexes I and IV, adenine nucleotide translocase, and carnitine-acyl transferase, are selectively decreased upon aging. Two hypothesis or lines of thought, link these observations. One of them considers that a decreased mitochondrial biogenesis is the driving phenomenon. The normal process of mitochondrial biogenesis, regulated by cellular 1NO , is slowed down by a lower level of 1NO steady-state concentration. Aged mitochondria

accumulate oxidation products and evolve to dysfunctional mitochondria, signaling for apoptosis. It is considered that mitochondria in simultaneous processes “respire (take up O_2 and release CO_2), generate $O_2^{\cdot -}$ and H_2O_2 , and undergo phospholipid peroxidation and enzyme inactivation.” The second hypothesis is that damaged mtDNA and mtRNA will produce, through defective mitochondrial protein synthesis, faulty polypeptides that would contribute along with protein carbonyls, hydroperoxides, and TBARS, to suboptimal enzyme function and the appearance of dysfunctional mitochondria. These dysfunctional mitochondria signal for lysosomal digestion, autophagy, or apoptosis. In all cases, the decrease in energy availability and in tissue-active cells, consequence of the decreased amount of functional mitochondria due to enhanced cell death, leads to the reduction of tissue physiological functions that is characteristic of aging.

See also: **Bioenergetics:** Cell Death by Apoptosis and Necrosis; Cytochrome c; Superoxide Dismutase; **Signaling:** Mitogen-Activated Protein Kinase Family; Nitric Oxide Signaling.

Further Reading

- Beckman KB and Ames BN (1998) The free radical theory of aging matures. *Physiological Reviews* 78: 547–581.
- Boveris A and Cadenas E (1982) Production of superoxide radicals and hydrogen peroxide in mitochondria. In: Oberley LW (ed.) *Superoxide Dismutase*, vol. II, pp. 15–30. Boca Raton, FL: CRC Press.
- Boveris A and Cadenas E (2000) Mitochondrial production of hydrogen peroxide, regulation by nitric oxide and the role of ubiquinone. *IUBMB Life* 50: 1–6.
- Cadenas E (1989) Biochemistry of oxygen toxicity. *Annual Review of Biochemistry* 58: 79–110.
- Chance B, Sies H, and Boveris A (1979) Hydroperoxide metabolism in mammalian organs. *Physiological Reviews* 59: 527–605.
- Elfering SL, Sarkela TM, and Giulivi C (2002) *Journal of Biological Chemistry* 277: 38079–38086.
- Fridovich I (1976) Oxygen radicals, hydrogen peroxide and oxygen toxicity. In: Pryor WA (ed.) *Free Radicals in Biology*, vol. 1, pp. 239–277. New York, NY: Academic Press.
- Gerschman R, Gilbert D, Nye SW, Dwyer P, and Fenn WO (1954) Oxygen radicals: A mechanism in common for oxygen and X-radiation toxicity. *Science* 119: 623–626.
- Ignarro LJ (2000) Introduction and overview. In: Ignarro LJ (ed.) *Nitric Oxide: Biology and Pathobiology*, pp. 3–19. New York, NY: Academic Press.
- Navarro A and Boveris A (2007) The mitochondrial energy transduction system in the aging process. *American Journal of Physiology – Cell Physiology* 292: C670–C686.

Mitogen-Activated Protein Kinase Family

H Teramoto and J S Gutkind, National Institute of Dental and Craniofacial Research, Bethesda, MD, USA

Published by Elsevier Inc.

This article is reproduced from the previous edition, volume 2, pp. 737–742, © 2004, Elsevier Inc.

Glossary

Extracellular signal-regulated protein kinase (ERK) One class of mitogen-activated protein kinases (MAPKs).

Guanosine triphosphatases (GTPases) Molecular switches that control cellular responses by exchanging guanosine diphosphate (GDP) (inactive form) for guanosine triphosphate (GTP) (active form) in response to a variety of stimuli. In the GTP-bound form they interact with effector molecules. Subsequently, they hydrolyze GTP to GDP, thereby returning to their inactive state.

Mitogen-activated protein kinases (MAPKs) A group of serine/threonine protein kinases that play an essential role

in signal transduction in response to changes in the cellular environment.

Protein kinases Phosphotransferases that catalyze the transfer of a phosphate group from adenosine triphosphate (ATP), or in some cases from GTP, to a protein substrate. Most protein kinases phosphorylate serine or threonine residues (serine–threonine protein kinases) or tyrosine residues (protein tyrosine kinases).

Ras A small GTPase that controls the activation of the ERK kinase cascade, as well as other downstream targets.

Mitogen-activated protein kinases (MAPKs) are a group of protein kinases that play an essential role in signal transduction by modulating gene transcription in the nucleus in response to changes in the cellular environment. MAPKs also play a key role in intracellular communication, and their activating pathways have been conserved throughout evolution, from plants, fungi, nematodes, insects, to mammals. In humans, there are at least 11 members of the MAPK superfamily, which can be divided into six groups: the extracellular signal-regulated protein kinases (ERK1 and ERK2); c-Jun N-terminal kinases (JNK1, JNK2, and JNK3); p38s (p38 α , p38 β , p38 γ , and p38 δ); ERK5 (ERK5); ERK3s (ERK3, p97 MAPK, and ERK4); and ERK7s (ERK7 and ERK8). Each group of MAPKs can be stimulated by a separate signal transduction pathway in response to different extracellular stimuli. In turn, MAPKs exhibit distinct substrate specificities. Thus, individual extracellular stimuli can lead to the differential activation of each MAPK and the consequent phosphorylation of a distinct set of MAPK substrates. Whereas components of some of these biochemical routes are now well established, others are still poorly understood. The specificity of the molecular mechanisms by which each MAPK can be activated ensures that cells mount an appropriate response to extracellular stimulation, as the resulting pattern of gene expression will depend on the integration of the combinatorial signals provided by the temporal activation of each of these groups of MAPKs.

ERK Family

ERK1 and ERK2

The extracellular signal-regulated protein kinase (ERK) group of MAPK includes ERK1 and ERK2. In certain cells their activation contributes to normal and aberrant growth, including malignant transformation, while in other cells this pathway can promote cell survival or initiate differentiation processes in others. The enzymatic activity of ERK1 and ERK2 is enhanced

by dual phosphorylation on Thr and Tyr residues within a short sequence that forms the activation loop, located close to the kinase active site, by a group of dual-specificity protein kinases (MAPK kinases; MAPKKs), represented by MEK1 and MEK2. Each MAPK family exhibits a distinct phosphor-acceptor motif in their activation loop (**Figure 1**), which provides specificity for the MAPKK–MAPK interactions. This phosphorylation leads to conformational changes in ERKs, thereby promoting their activation and nuclear localization. Subsequently, ERK activity is terminated by dephosphorylation on either Thr or Tyr by a Ser/Thr or Tyr phosphatase. This dephosphorylation can be mediated also by a dual specificity MAPK phosphatase. Similarly, the mammalian MAPKKs, such as MEK1 and MEK2, are activated by phosphorylation mediated by MAPK kinase kinases (MAPKKKs), which includes the Ras effector, Raf. Thus, the Raf group of MAPKKK that includes A-Raf, B-Raf, and c-Raf-1 plays an important role in the coupling of ERK MAPK activation to the Ras signaling pathway (see **Figure 2**).

The discovery of the Ras–Raf–MEK–ERK signaling route represented a breakthrough in cancer biology, and emerged as the product of the conversion of a large number of research efforts in a variety of otherwise poorly connected fields. Indeed, a large body of information was derived from genetic analyses of model organisms including *Drosophila*, *Saccharomyces cerevisiae*, and *Caenorhabditis elegans*, in addition to biochemical and biological studies in mammalian cells. They, collectively, led to the identification of Ras as a key downstream component of the pathway initiated by polypeptide growth factor receptors, and to the realization that this small GTPase acts subsequently as a master switch initiating the activity of a cascade of cytoplasmic kinases that culminate with the activation of the ERK1 and ERK2. ERKs, in turn, direct the activities of many nuclear transcription factors, thus regulating gene expression. These findings provided a better understanding of the molecular basis for the transduction of proliferative signals, as well as the first clue as to how the aberrant function of Ras proteins may contribute to the malignant conversion of cancer cells.

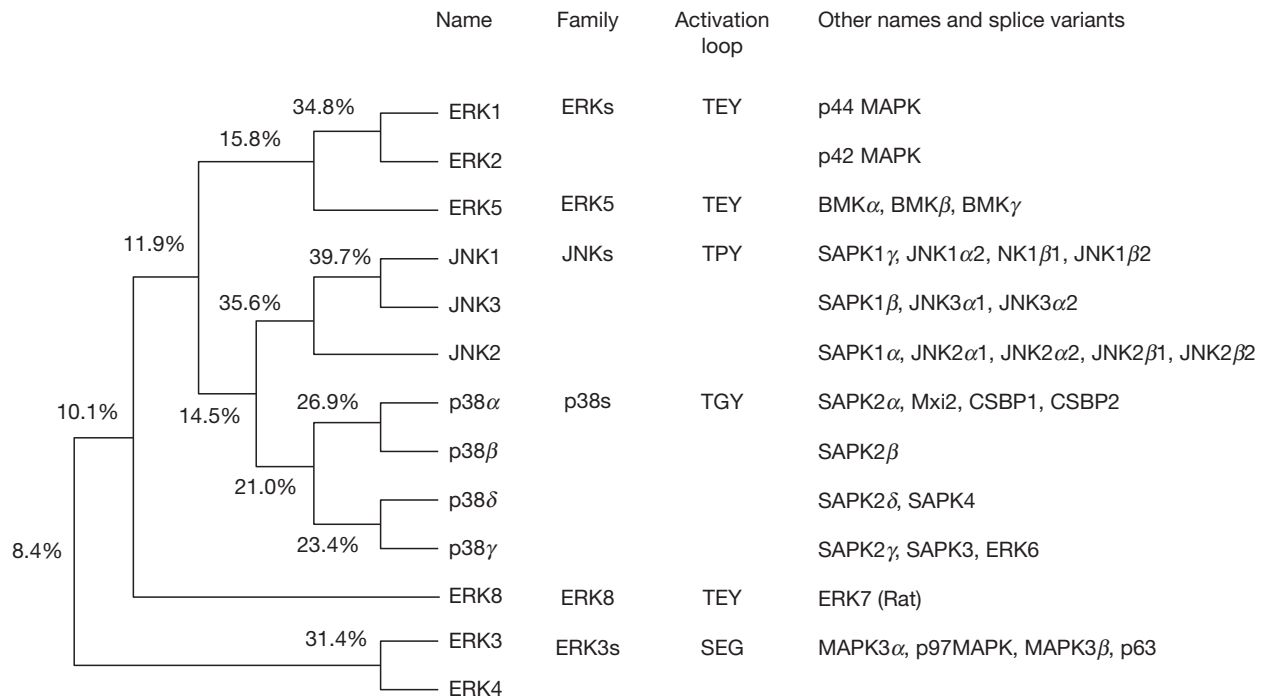


Figure 1 Phylogenetic tree of human MAPKs. A computer-generated phylogenetic tree of the MAPKs. The GenBank accession numbers are human ERK1 (X60188), human ERK2 (M84489), human ERK3 (S23429), human p97 (X80692), human ERK4 (P31152), human ERK5 (U25278), human BMK (U29725), Rat ERK7 (AF078798), human ERK8 (AY065978), human JNK1 (U34822), human JNK2 (U34821), human JNK3 (U34820), human Mxi2 (U19775), CSBP1 (S52419), CSBP2 (S52420), human p38 β (U53442), and human p38 γ (U66243).

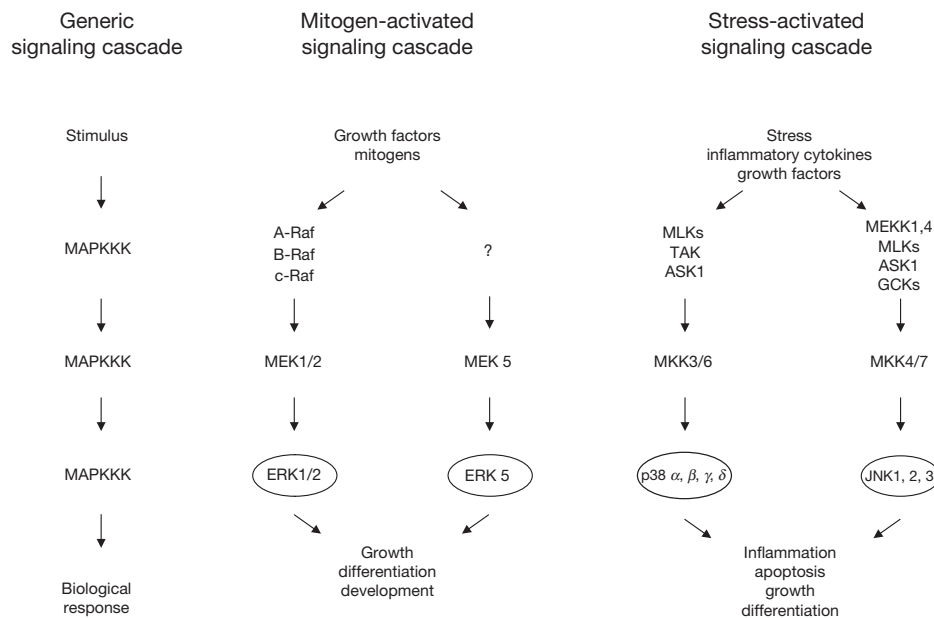


Figure 2 MAPK signaling cascades. Schematic representation of generic (left), mitogen-activated (center), and stress-activated (right) signaling cascades. Separate signal transduction pathways initiated by different extracellular stimuli activate each group of MAPKs that, in turn, exhibit distinct substrate specificities.

c-Jun N-Terminal Kinases (JNK1, JNK2, and JNK3)

JNK was initially discovered by its ability to phosphorylate the N-terminal transactivating domain of the transcription factor c-Jun. As its activity was stimulated primarily by cellular stress

rather than by mitogens, contrary to ERKs, it was also termed 'stress-activated protein kinase (SAPK)'. It is now known that there are three members of the JNK family, JNK1–3, and that they play numerous biological roles in response to a large variety of cell surface receptors as well as physical and chemical

stresses. Whereas the JNK1 and JNK2 genes and their alternatively spliced variants are expressed ubiquitously, generating a total of eight JNK isoforms, the JNK3 gene has a more restricted expression pattern, being its two alternative spliced transcripts expressed the highest in brain and testis. Studies of mammalian cells demonstrated that JNK is required for neuroexcitatory stress-induced neuronal apoptosis, as well as for certain inflammatory responses and for the survival and malignant transformation of specific cell types (e.g., B cells).

Following the distinct pattern of activation of ERK and JNK by cellular receptors, an unexpected discovery was that whereas Ras regulates ERKs, two members of the Rho family of GTPases, Rac1 and Cdc42, initiate an independent kinase cascade regulating JNK activity. These findings provided direct evidence of a role for Rho GTPases in transcriptional regulation, thus expanding previous efforts addressing the function of these Rho-related proteins, which were primarily focused on the molecular dissection of the mechanism(s) involved in their ability to control the dynamic assembly and remodeling of actin-containing cytostructures. Components of this pathway include several MAPKKKs: MEKK1, MEKK2, MEKK3, and MEKK4, apoptosis-stimulated kinase 1 (ASK1), mixed-lineage kinase 3 (MLK3), tumor progression locus 2, also known as Cot (TPL2), TGF- β -activated kinase (TAK), MAPK upstream kinase (MUK), germinal center kinase (GCK), and PAK, which can all contribute to activation of JNKs, apparently through two MAPKKs, MKK4 (also called Sek or JNKK1), and MKK7. Substrates for JNK include transcription factors of the c-Jun family, as well as other transcription factors, including Elk-1 and Elk-2, serum response factor accessory protein-1 (Sap-1), nuclear factor of activated T cells 4 (NFAT4), and activating transcription factor 2 (ATF2).

p38s (p38 α , p38 β , p38 γ , and p38 δ)

The family of p38 kinases has expanded over the past few years, including now four members known as p38 α (or CSBP-1), p38 β , p38 γ (also known as SAPK3 or ERK6), and p38 δ (or SAPK4), and in some cases their corresponding splice variants. These MAPKs share a Thr–Gly–Tyr phosphorylation motif in their activation loop, and are stimulated by receptors and environmental stresses similar to those provoking the activation of JNKs. These enzymes are particularly sensitive to their stimulation by exposure of cells to endotoxins, including bacterial lipopolysaccharides. Evidence of their role in transducing stress signals can be found as far phylogenetically as in yeast, where genetic analysis indicates that the Hog1p MAPK, the yeast homolog of p38, is required for the response to osmotic stress that involves the enhanced expressions of genes participating in the biosynthetic pathway of the osmotic stabilizer glycerol.

As for the ERKs, p38s are activated by dual phosphorylation on Thr and Tyr residues by MKK3 and MKK6, the latter being a general stimulator for p38s and the former exhibiting certain preference for p38 α . The immediate upstream activators for these kinases are not yet well defined and include the potential p38 MAPKKKs known as ASK1 and TAK1. The specific transcription factors that can be regulated by p38s include cAMP-responsive element-binding protein (CREB), activating transcription factor 1 (ATF1), ATF2, Max, C/EBP homologous

protein (CHOP), and MEF2C. In addition, these MAPKs can also trigger the activation of other serine–threonine kinases such as Mnk1 (MAPK-interacting kinase 1) and Mnk2 and MAPK-activated protein kinases (MAPKAPKs).

ERK5

ERK5, also known as big mitogen-activated protein kinase (BMK1), is distantly related to ERK1 and ERK2. Although it contains a Thr–Glu–Tyr motif in its activation loop, similar to that of ERK1 and ERK2, ERK5 is larger than any other known MAPK (~80 kDa) and is selectively activated by MEK5 but not by MEK1 or MEK2. This kinase can be stimulated by oxidative stress, and can play a role in early gene expression triggered by epidermal growth factor and serum by the direct phosphorylation of the transcription factors MEF2C or c-Myc, thereby participating in cell proliferation and progression through the cell cycle. Although information on this MAPK is still limited, evidence supporting its role in cell growth is also provided by its ability to cooperate with the activated Raf or MEK, which act on ERK1 and ERK2, to promote neoplastic transformation. The physiological function of ERK5 is nevertheless still unclear, and further studies are required to establish the biological role of this MAPK signaling pathway.

Others (ERK3, ERK4, ERK7, and ERK8)

Novel members of the MAPK family, ERK 7 (61 kDa) and ERK8, also contain a Thr–Glu–Tyr activation motif. ERK8 was cloned using the rat ERK7 complementary DNA (cDNA) to screen a human multiple tissue cDNA library and share an overall amino acid sequence identity of 69%. In contrast to rat ERK7, ERK8 does not display significant constitutive kinase activity, does not phosphorylate c-Fos, and is activated in response to stimulation by Src or serum. ERK3 contains a Ser–Glu–Gly activation motif, with gly in place of tyr. The ERK3 subfamily of MAPKs is composed of two functional genes, MAPK6 (p97 MAPK) and MAPK4 (ERK4), and several pseudogenes. Upstream activators of ERK3 are still poorly defined, and this kinase is not present in yeast or *C. elegans*. None of the currently known MEK family members is able to phosphorylate and activate ERK3 and no physiological substrates of ERK3 have been found. ERK3 is localized constitutively to the nucleus without conventional nuclear localization sequences.

Specificity and Fidelity in Nuclear Signal Transmission by MAPK Signaling Pathways

Protein kinases typically interact with their substrates by recognizing primary sequence determinants that surround the phosphorylation site(s). In the case of MAPK, the minimum general primary sequence required to be a substrate for protein phosphorylation is the presence of a Ser or Thr residue followed by Pro. Additionally, most MAPKs also appear to interact with their substrates by binding to a second site. This docking interaction is required for substrate phosphorylation by MAPK *in vitro* and is likely to be a relevant mechanism for the

selection of high affinity substrates by these kinases *in vivo*. The current view is that for substrate recognition, MAPKs first dock to the substrate through a binding interaction at one site leading to phosphorylation of the substrate at a second site at one or more Ser/Thr-Pro motifs. Recent studies demonstrate that the selective interaction of MAPK with substrates may influence the physiological regulation of MAPK targets *in vivo*. Examples of such docking interactions include the interaction of ERK or JNK with the transcription factor Elk-1. Furthermore, there is substantial evidence that the normal function of MAPK signaling modules requires the proper organization of their components within the cell.

Of interest, in the past decade, scaffolding proteins have emerged as general regulatory mechanisms ensuring the fidelity of intracellular pathways. These proteins bind components of signaling routes, physically organizing them to enable rapid physiological responses. The prototypical signaling scaffold is the yeast protein Ste5p which binds the components of the yeast MAPK cascade leading to the mating response after activation of the pheromone receptor, thus maintaining the signaling fidelity of this cascade. Several scaffold-binding components of MAPK cascades have now been described in multicellular organisms, including kinase suppressor of RAS (KSR) and c-Jun terminal kinase interacting protein family (JIP). Their role is to maintain the integrity of biochemical routes connecting cell surface receptors to the nucleus. For example, the JIP family of scaffolding proteins consists of three mammalian homologs JIP/IB-1, -1b, and -2. These molecules bind members of the mixed-lineage group of MAPKKK's (MLK) that have been reported to activate both JNK and p38. In addition, both IB1/JIP1 and IB2/JIP-2 have been found to bind the JNK MAPKK, MKK7, while IB2/JIP-2 has also been reported to bind MLK3 and p38 γ . An emerging concept derived from recent studies is that these scaffold proteins are organizers and keepers of signal specificity and fidelity. Indeed, it may be as important to keep signals separate between closely related MAPK cascades, as it is to integrate them in a coordinated response, particularly in gene expression control.

The activation of MAPK-mediated signaling pathways is often initiated by stimuli acting at the cell surface. However, many of the transcription factors that act as MAPK substrates are located in the nucleus, thus suggesting the existence of mechanisms by which the signals can be transmitted from the cytoplasm to the nucleus. This has been extensively investigated for the ERK1/2 pathway. For example, the activated forms of Raf (MAPKKK) are present at the plasma membrane, while activated MEK1 and MEK2 (MAPKK) are located in the cytoplasm, due to the presence of a nuclear export signal in the protein. Thus, the final step in signal transmission from the cytoplasm to the nucleus appears to be directly mediated by ERK1/2. Indeed, these ERKs are located in the cytoplasm of quiescent cells but in the nucleus of activated cells. The mechanism by which activation promotes the translocation of ERKs to the nucleus is still unclear. It has been suggested that ERKs may be actively sequestered in the cytoplasm by inactive MEK1/2, and upon their activation by Raf, MEK1 and MEK2 may dissociate from ERK1 and ERK2, which may then be able to migrate to the nucleus as homodimers. Whereas available evidence supports this model, it is still unclear whether the release of ERKs from their MAPKK,

MEK1, and MEK2 is sufficient to promote the nuclear accumulation of ERKs or whether additional processes are involved. Furthermore, whether other MAPKs use similar or distinct mechanisms to ultimately phosphorylate nuclear proteins is also poorly understood, and under current intense investigation.

Regulation of Gene Expression through MAPK Signaling Pathways

Although MAPKs can phosphorylate cytoplasmic substrates, most MAPK signaling pathways ultimately converge in the nucleus to regulate gene expression, either by acting at the pre-transcriptional, transcriptional, or posttranscriptional level. The most widely used mechanism by which MAPKs regulate gene expression is the phosphorylation of transcription factors, thereby enhancing their activity. For example, phosphorylation of the transcriptional activation domains of ATF2, CHOP, Elk-1, and c-Jun causes increased transcription. As studied in detail for the transcription factor CREB, transcription factor phosphorylation can cause changes in the interaction with co-activator molecules, including CBP and p300, which function, in part, through the regulation of the chromatin structure by acetylating histones.

Less frequent cases are found where MAPKs negatively regulate transcription factors by promoting their retention in the cytoplasm upon phosphorylation, and active dephosphorylation of these factors is then needed for their migration to the nucleus to stimulate gene transcription. As examples, phosphorylation of the transcription factors SMAD1 and NFAT4 by ERK1/2 and JNK, respectively, causes their cytoplasmic retention and therefore the inhibition of their transcriptional activity. Whether instead the phosphorylation by MAPKs of cytoplasmic transcription factors may cause their nuclear translocation is still unclear. Regulation of the DNA-binding activity is also a mechanism employed by MAPKs to control transcription factor activity. An example is the transcription factor Elk-1, which is phosphorylated by ERK, JNK, and p38 MAPKs leading to increased DNA binding.

Regulation of translation represents a fourth mechanism of regulation of gene expression by MAPKs. The best example is the translational regulation of cytokine expression observed in macrophages exposed to endotoxin. In this case, p38 MAPK acts on a *cis*-acting element in the tumor necrosis factor α (TNF α) messenger RNA (mRNA) to regulate its translation. Modulation of mRNA processing and nuclear export has not been described, although these biosynthetic steps also represent potential sites for intervention by MAPK.

The final mechanism employed by MAPK to regulate gene expression is the regulation of protein degradation. This is exemplified by the short-lived transcription factor c-Jun, which is rapidly degraded by the ubiquitin-proteasome pathway. Phosphorylation of c-Jun by JNK inhibits the ubiquitination of c-Jun and, therefore, its rapid degradation. Consequently, JNK activation prolongs the half-life of c-Jun, leading to the accumulation of the c-Jun protein. The regulation of protein stability by MAPK therefore contributes to the regulation of gene expression mediated by these signaling pathways.

See also: **Signaling:** Muscarinic Acetylcholine Receptors; Ras Family; Src Family of Protein Tyrosine Kinases.

Further Reading

- Cano E and Mahadevan LC (1995) Parallel signal processing among mammalian MAPKs. *Trends in Biochemical Sciences* 20(3): 117–122.
- Chang L and Karin M (2001) Mammalian MAP kinase signaling cascades. *Nature* 410(6827): 37–40.
- Gutkind JS (2000) Regulation of mitogen-activated protein kinase signaling networks by G-protein-coupled receptors. *Science's STKE* 40: RE1.
- Johnson G and Lapadat R (2002) Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science* 298: 1911–1912.
- Kultz D (1998) Phylogenetic and functional classification of mitogen- and stress-activated protein kinases. *Journal of Molecular Evolution* 46: 571–588.
- Ono K and Han J (2000) The p38 signal transduction pathway: Activation and function. *Cell Signal* 12(1): 1–13.
- Pearson G, Robinson F, Beers T, et al. (2001) Mitogen-activated protein (MAP) kinase pathways: Regulation and physiological functions. *Endocrine Reviews* 22(2): 153–183.
- Weston CR and Davis RJ (2002) The JNK signal transduction pathway. *Current Opinion in Genetics and Development* 12(1): 14–21.

Mitosis

P Wadsworth and J Titus, University of Massachusetts, Amherst, MA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Patricia Wadsworth and Nasser M. Rusan, volume 2, pp. 743–747, © 2004, Elsevier Inc.

Glossary

Centrosome Structure present at the spindle poles in many animal cells that is composed of a pair of centrioles and associated pericentriolar material.

Congression The motion of chromosomes to the midplane, or equator, of the spindle during prometaphase.

Kinetochores Structure located at the centromere of each chromatid that functions to link the chromosome to

the spindle. The kinetochore is also the location of components of the spindle assembly checkpoint.

Microtubule Polar, cytoskeletal polymer composed of tubulin dimers that are a major structural element of the mitotic spindle and are required for chromosome motion.

Spindle flux Poleward motion of microtubule polymers that occurs during metaphase and anaphase of mitosis.

Spindle Structure

The mitotic spindle at metaphase is a bilaterally symmetric, microtubule-based structure with chromosomes aligned in the middle. The spindle is bipolar: two spindle poles define the spindle axis and are the sites to which the chromatids are moved during anaphase (Figure 1). In many animal cells, a centrosome composed of a pair of centrioles and associated pericentriolar microtubule-nucleating material is present at the spindle poles (Figure 1). The major structural elements of the spindle are microtubules, dynamic cytoskeletal filaments that radiate from the two spindle poles. Microtubules attach chromosomes to the spindle and are essential for chromosome motion. Spindle microtubules are classified by their location in the spindle (Figure 1). Interpolar microtubules extend from each spindle pole and overlap in the middle of the spindle. Kinetochore microtubules have one end embedded in a specialized attachment site on the chromosome, the kinetochore. Astral microtubules extend away from the centrosomes toward the cell cortex and contribute to spindle positioning and in defining the position of the contractile ring at cytokinesis.

Considerable variation in spindle structure exists among diverse cells. In cells of higher plants, the spindle pole lacks centrioles and is broad, not focused. Plant spindles are also anastral, lacking astral microtubules. In yeasts, which undergo closed mitosis inside of an intact nuclear envelope, a spindle pole body is present at the two ends of the spindle. The spindle pole body, a trilaminar plaque-like structure embedded in the nuclear envelope, nucleates microtubules that radiate into the nucleus and cytoplasm. The number of spindle microtubules also varies among cell types. Yeasts have a simple spindle, and a single microtubule links each kinetochore to the spindle pole body. By contrast, typical animal cells have approximately 20 microtubules associated with each kinetochore and hundreds of microtubules present in each half-spindle. Despite these differences in spindle structure, in all cases mitosis accomplishes the accurate segregation of the replicated genetic material into the two daughter cells.

Phases of Mitosis

Prophase

Chromosome condensation

During the S (synthetic) phase of interphase, each chromosome is replicated resulting in two identical sister chromatids that are bound together by protein complexes called cohesions. Sister chromatids are very long strands of DNA that must be compacted for division. This process, called chromosome condensation, occurs by specialized proteins called condensins that coil the DNA into loops. As prophase progresses, the chromosomes become increasingly visible in the nucleus, first as threads or strands and later as highly condensed rod-like structures called chromosomes, each of which is composed of the two sister chromatids (Figure 2). The term mitosis comes from the Greek word *mito* meaning thread, referring to the appearance of thread-like structures in the prophase nucleus.

Kinetochore assembly

In order for the chromatids to be accurately segregated into the two daughter nuclei, they must be attached to, and later moved by, the spindle machinery. Mitotic chromosomes are characterized by an indentation, or primary constriction, called the centromere (Figure 1). The centromere is a region of highly repeated DNA sequences that serves as the assembly site for a plate-like structure called the kinetochore. The kinetochore functions to link each chromatid to the spindle via kinetochore microtubules and also functions as a signaling device that regulates progression through the cell cycle. Kinetochores of sister chromatids are arranged back to back so that the binding sites for spindle microtubules face in opposite directions to facilitate the association of each chromatid with microtubules from opposite spindle poles.

Centrosome separation

In animal cells, the centrosome is also duplicated during S phase of interphase, and the two daughter centrosomes remain adjacent to one another. Centrosome duplication ensures that two centrosomes are present at the beginning of

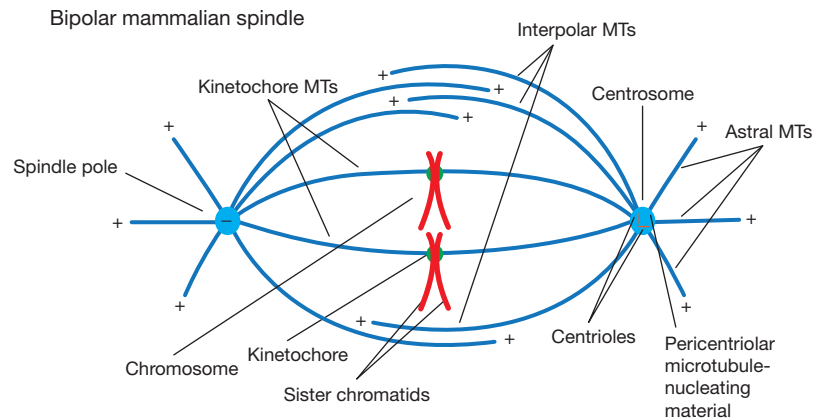


Figure 1 Highly schematic diagram of a bipolar mammalian mitotic spindle at the metaphase stage of mitosis. Microtubules are depicted as dark blue lines; interpolar, kinetochore, and astral microtubules (MTs) are labeled. The plus and minus ends of the microtubules are marked. Each pole has microtubules radiating from a centrosome (light blue). Contained within the pericentriolar microtubule-nucleating material (light blue) are two centrioles (orange). Examples of chromosomes, with sister chromatids in red and kinetochores in green, are shown aligned at the metaphase plate.

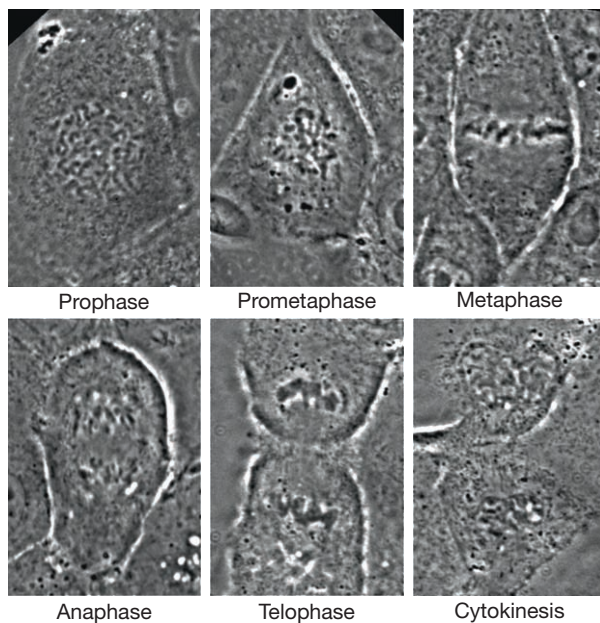


Figure 2 Phase-contrast light micrographs of LLC-PK1 mammalian cells in mitosis. Each panel shows a different cell at the indicated stage of mitosis.

mitosis to form the spindle poles and guarantees that each daughter cell inherits one centrosome. In prophase, the duplicated centrosomes separate and begin moving to opposite sides of the nucleus to establish a bipolar spindle. The microtubule-nucleating capacity of each centrosome also increases dramatically at prophase.

Prometaphase

Nuclear envelope breakdown

In cells that undergo open mitosis, the breakdown of the nuclear envelope marks the beginning of prometaphase. Microtubules assist nuclear envelope breakdown resulting in

the fragmentation of the nuclear membrane into small vesicles and consequently the release of the chromosomes into the cytoplasm.

Spindle assembly

Spindle assembly involves a dramatic reorganization of microtubules into a functional bipolar array. This process requires dynamic microtubules. Because microtubules are polar polymers, their two ends have different dynamic properties. The less dynamic end, called the minus end, is associated with the spindle pole material; the more dynamic end, called the plus end, extends outward toward the cell periphery (**Figure 1**). Microtubule plus ends display a behavior called dynamic instability, in which they switch stochastically between phases of assembly and disassembly. The result of dynamic instability behavior is that microtubule plus ends continually probe the cytoplasm as they grow and shorten. Microtubules in mitotic cells are more dynamic than microtubules in interphase cells; they switch from growing to shortening more frequently and shorten further before resuming growth. The result is a population of relatively short, highly dynamic microtubules that radiate from each spindle pole. At nuclear envelope breakdown, when the chromosomes are released into the cytoplasm, dynamic microtubules can interact with each kinetochore. Once a connection between a microtubule and kinetochore is established, the microtubule becomes less dynamic in order to maintain the connection. Cells monitor the attachment of chromosomes to the mitotic spindle and mitosis is delayed if even a single kinetochore is unattached. This regulatory system is called the spindle assembly checkpoint, or metaphase checkpoint.

Microtubule motor proteins, which utilize the energy of adenosine triphosphate hydrolysis to perform work in cells, are essential for assembling the bipolar spindle. The type and properties of the motor protein govern the direction of motion, either toward the minus end or the plus end of the microtubule. Multifunctional motors or motor protein complexes can exert force on microtubules by cross-linking neighboring microtubules and sliding them relative to one another or by

binding to one microtubule while walking along an adjacent one. Alternatively, the nonmotor region of the motor protein can bind to a tethered structure, the cell cortex for example, while the motor region of the protein applies a pulling force on the microtubule. Motor proteins are present at the kinetochores, between interpolar microtubules, along kinetochore microtubules and at the cell cortex.

Motor proteins contribute to various motions during mitosis. During centrosome separation in prophase, motors located between overlapping microtubules contribute to the separation of the two spindle poles. At the kinetochores, motors contribute to the initial attachment of chromosomes to spindle microtubules. Motors also contribute to congression, the movement of the chromosomes to the spindle equator following initial attachment. In the assembled spindle, a balance of the activities of oppositely directed motor proteins maintains spindle length; some motors tend to pull the poles together, and this force is resisted by motors that act to push the poles apart. Motor proteins also contribute to spindle positioning in mitotic cells, and to the dynamic turnover of mitotic microtubules.

Spindle assembly can also occur in the absence of centrosomes. In this situation, microtubules assemble in the vicinity of chromosomes and are subsequently sorted into a bipolar array by the activity of motor proteins. In both the presence and absence of centrosomes, dynamic microtubules and motor proteins contribute to spindle formation.

Metaphase

The cell is in metaphase when all of the duplicated chromosomes are aligned midway between the two spindle poles (Figure 2 and 3), a location referred to as the metaphase plate or spindle equator. Spindle length remains relatively constant at metaphase, but the spindle microtubules are not static. Rather, they undergo a behavior called spindle flux, in which spindle microtubules continually assemble at their plus ends and disassemble at the minus ends, coupled to poleward translocation of the polymer. As a result of flux, marks placed on spindle microtubules are observed to move slowly toward the poles during metaphase and anaphase of mitosis.

Not only are microtubules dynamic at metaphase, but the chromosomes are active as well, oscillating toward and away from each spindle pole. Chromosomes are also under tension, ready for subsequent movement in anaphase. If the connection between one sister chromatid and the pole is severed at metaphase, the chromosome moves to the pole to which it remains attached, demonstrating that the metaphase position results from a balance of pole-directed forces acting on the chromosomes.

Anaphase

During anaphase, sister chromatids separate and move to the spindle poles (Figures 2 and 3). Anaphase consists of two phases, anaphase A and B. During anaphase A, the chromosomes move to the poles and kinetochore fiber microtubules shorten; during anaphase B, the spindle poles move apart as interpolar microtubules elongate and slide past one another. Many cells undergo both anaphase A and B motions, but, in some cases, one or the other motion dominates.

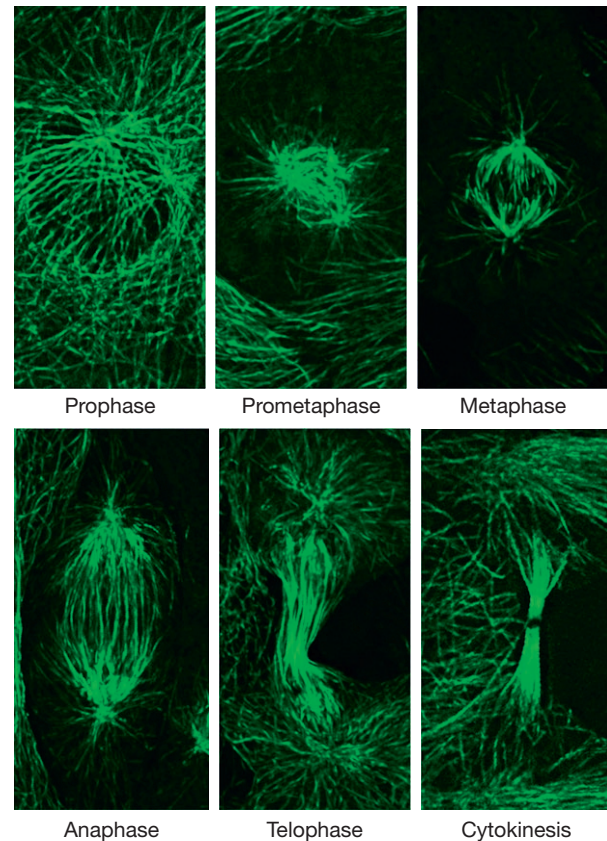


Figure 3 Distribution of microtubules throughout mitosis. Fluorescence light micrographs of cells that have been fixed and stained with antibodies to tubulin, to label microtubules, shown in green. Images are maximum projections of deconvolved image stacks taken with a spinning disk confocal microscope.

Separation of the paired sister chromatids is required for poleward motion in anaphase. Chromatid separation results from the proteolytic degradation of components that link the chromatids at the centromere. Degradation is triggered by the activity of the anaphase-promoting complex, which regulates cell-cycle progression. Chromatid separation is not the result of tugging by microtubules and motor proteins, and can be observed even in the absence of microtubules.

Although the motion of the chromosomes to the spindle poles in anaphase has fascinated biologists for many years, the molecular basis for this motion remains controversial and incompletely understood. During anaphase A, kinetochore microtubules must shorten as the chromosomes move poleward. Measurements of spindle flux show that subunit loss from microtubules occurs at the spindle poles during anaphase. In many cells, however, the rate that chromosomes move exceeds the rate of subunit loss at the pole, and, thus, subunit loss must also occur at the kinetochore.

Pioneering studies of mitosis in living embryonic cells demonstrated that assembly and disassembly of microtubule polymers result in chromosome motion. This work led to the hypothesis that microtubule disassembly drives chromosome motion. Later work identified molecular motors at the kinetochore, leading to the alternative hypothesis that forces generated

by molecular motors drive chromosome motion. One possibility is that molecular motors power chromosome motion, but kinetochore microtubule disassembly limits the rate of chromosome motion. Alternatively, disassembly may be responsible for chromosome motion, and motors may tether the chromosomes to the shortening fiber. The presence of potentially redundant mechanisms for chromosome motion may reflect the fact that mitotic fidelity is of utmost importance.

Telophase

In telophase, the chromosomes decondense, and the nuclear envelope begins to reform around them (Figures 2 and 3), returning the cell to the interphase condition. The short dynamic microtubules of the spindle are replaced with the less-dynamic microtubules of the interphase array that extend to the cell periphery. During telophase, the contractile ring assembles midway between the spindle poles in preparation for cytokinesis.

Cytokinesis

Cytokinesis is the process of constricting the cytoplasm between the two forming daughter nuclei resulting in the formation of two cells (Figures 2 and 3). Cytokinesis is achieved by a contractile ring, or belt, composed of filaments of the cytoskeletal polymer, actin, and the molecular motor protein myosin II, which provides the force for cytokinesis. The contractile ring is tethered to the cell membrane, and its contraction creates a cleavage furrow, or indentation of the plasma membrane, during cytokinesis. The contractile ring assembles midway between the two spindle poles in telophase, and disassembles when cytokinesis is completed. The signals that determine the location of contractile ring formation in the cell cortex are incompletely understood, but it is

well-established that microtubules are involved in their delivery to the cortex. In some cells, radiating astral microtubules that overlap midway between the spindle poles dictate the site of contractile ring formation. In other cases, interzonal microtubules located between the separating chromosomes play an important role in specifying the location of furrow formation. The completion of cytokinesis marks the end of mitosis and the birth of two completely independent daughter cells in interphase of their cell cycle.

See also: **Cell Architecture and Function: Cytoskeletal Motors: General Principles; Rho GTPases and Actin Cytoskeleton Dynamics.**

Further Reading

- Compton DA (2000) Spindle assembly in animal cells. *Annual Review of Biochemistry* 69: 95–114.
- Inoue S and Salmon ED (1995) Force generation by microtubule assembly/disassembly in mitosis and related movements. *Molecular Biology of the Cell* 6: 1619–1640.
- Karsenti E and Vernos I (2001) The mitotic spindle: a self-made machine. *Science* 294: 543–547.
- McIntosh JR and McDonald KL (1989) The mitotic spindle. *Scientific American* 261(4): 48–56.
- Nicklas RB (1997) How cells get the right chromosomes. *Science* 275: 632–637.
- Rieder CL and Khodjakov A (2003) Mitosis through the microscope: Advances in seeing inside live dividing cells. *Science* 300: 91–96.
- Scholey JM, Brust-Mascher I, and Mogilner A (2003) Cell Division. *Nature* 422: 746–752.
- Wittman T, Hyman A, and Desai A (2001) The Spindle: A dynamic assembly of microtubules and motors. *Nature Cell Biology* 3: E28–E34.

Relevant Websites

- <http://dms.dartmouth.edu> – Compton Laboratory.
- <http://www.bio.unc.edu> – The Salmon Laboratory.
- <http://www.bio.umass.edu> – Wadsworth Laboratory.

Monoamine Oxidase

A Mattevi, University of Pavia, Pavia, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Catalytic intermediate Chemical species that transiently forms during an enzymatic reaction.

Introduction

The human genome encode for two enzymes, monoamine oxidase A and B (MAO A and MAO B), that catalyze the oxidation of a number of amine-containing molecules. The two proteins are encoded by different genes although they have significant (about 70%) sequence identity. This similarity is matched by partly overlapping substrate specificities. In very general terms, MAO A accepts bulkier substrates than MAO B. This distinction can be best illustrated by the two neurotransmitters dopamine and serotonin; dopamine is a good substrate for MAO B whereas serotonin is the preferred substrate for MAO A. The main reason for the huge number of studies on MAOs resides in their pharmacological relevance. It was found already in the fifties that these enzymes represent valid targets for the treatment of neurological disorders; more specifically, depression and Parkinson's disease. In this article, we first describe the reaction catalyzed by MAOs. Next we discuss the inhibitors and drugs targeting these proteins in light of the more recent biochemical and structural studies.

Reaction Catalyzed

Human MAOs are typical flavin-dependent oxidases. They use a flavin-adenine dinucleotide (FAD) cofactor to catalyze the oxidation of the amine group of the substrate. FAD is covalently bound to a protein Cys residue and the formation of the covalent bond, as in most covalent flavoenzymes, is self-catalyzed – that is, it does not require the action of an external protein or cofactor. The enzymes take advantage of the redox properties of the flavin to perform the oxidation of the substrate $-CH_2-NH_2-$ to convert it into the corresponding imine (Figure 1). This compound spontaneously hydrolyzes to the aldehyde, which is, therefore, the final reaction product. Substrate oxidation leads to the formation of the two-electron-reduced flavin which needs to be re-oxidized. As it ought to be in oxidases, the electron acceptor is oxygen which is converted to hydrogen peroxide, the other reaction product. Normally, it is thought that the enzyme functions according to a ternary complex mechanism – that is flavin re-oxidation takes place when the imine product is not yet released by the enzyme. Indeed, the imine, at least for certain substrates, is known to enhance the reactivity of the reduced enzyme with oxygen. It is beyond the scope of this article to discuss the intricacies and the huge body of literature concerning the mechanism of substrate oxidation/flavin reduction. The

interested reader is referred to the reference list at the end of the article. Here, we simply summarize the state of the art about this hotly debated subject in the enzymology community. In essence, there are three proposals: the substrate is oxidized through a transient covalent intermediate with flavin; substrate oxidation takes place through direct transfer of a hydride anion from the carbon atom of the $-CH_2-NH_2-$ unit to the flavin; oxidation occurs through stepwise transfer of electrons and/or hydrogen atoms with transient formation of a radical intermediate. Future work is required to determine which of these mechanistic proposals will finally prove to be the correct one.

Cellular Localization

An important feature of MAOs is that they are membrane bound. Indeed, both MAO A and MAO B feature a C-terminal hydrophobic tail that, as initially predicted by sequence analysis and later confirmed by the crystal structures, forms a trans-membrane helix that anchors these dimeric enzymes to the outer mitochondrial membranes. The physiological significance of this localization remains unclear. At present, it is also unknown how the proteins are oriented in the membrane. Two possibilities exist; the globular domain region of the protein (Figure 2) is located in the space between the inner and outer membranes of the mitochondrion or the globular domain is exposed on the mitochondrial outer surface. Biochemical studies have shown that removal by protein engineering of the membrane-binding regions lead to a more water-soluble protein that retains catalytic activity but it is far less stable. A fascinating speculation is that the membrane localization might favor substrate binding. The idea is that the negatively charged membrane surface would attract the positively charged amine substrates which, as gathered from the crystal structures, enter the active site through a gating loop located at the protein–membrane interface (Figure 2).

Enzyme Substrates

Both MAO A and MAO B are rather promiscuous and known to accept a variety of substrates. The two enzymes feature internal cavities located in front of the flavin cofactor. These cavities exhibit different shapes on their rear side as consequence of a few critical amino acid differences which result in only partly similar substrate specificities. Both proteins are able to act on small aromatic amines such as benzylamine (Figure 1 and

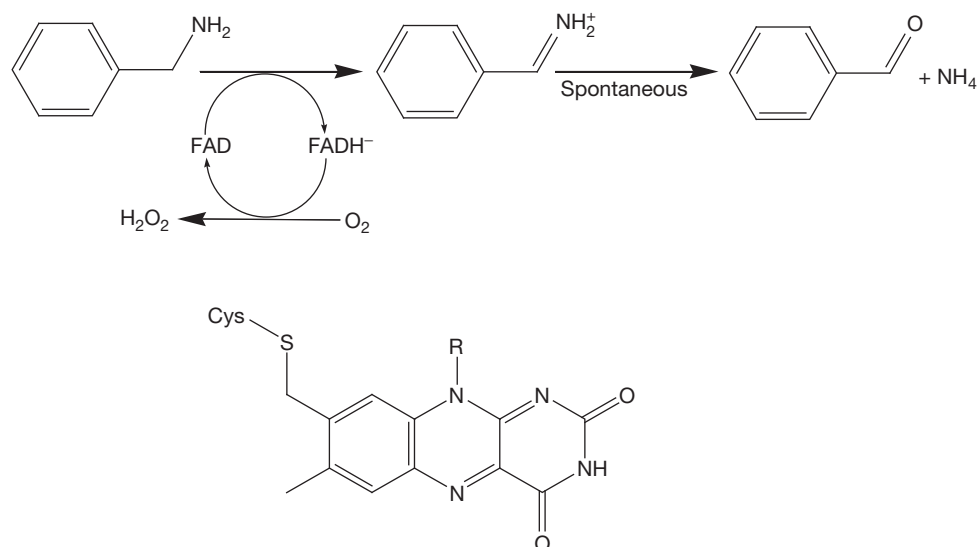


Figure 1 Scheme of the reaction catalyzed by MAOs with reference to the benzylamine substrate and of the covalently protein-bound flavin.

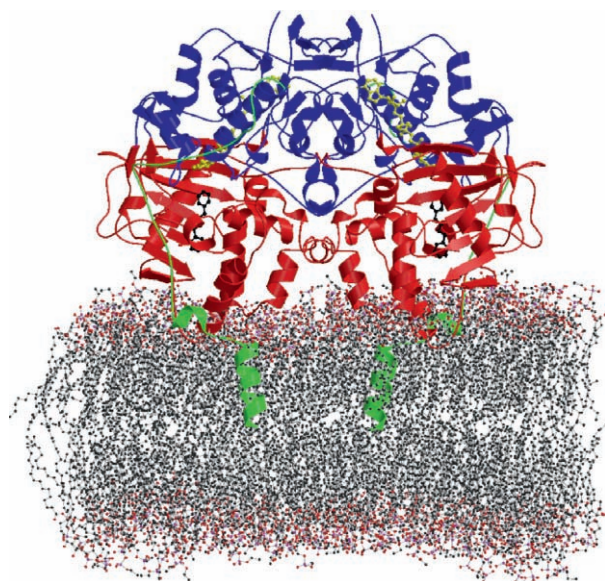


Figure 2 Schematic illustration derived from the crystal structure analysis of the human MAO B dimer bound to the outer mitochondrial membrane. The substrate-binding domain is in blue, the flavin-binding domain in red. The membrane-bound helices are in green. At the membrane–protein interface is located a loop that represents the entry point for the substrate into the active site, located in front of the flavin cofactor (yellow). In black is shown diphenylbutene, a MAO B inhibitor that binds in the enzyme active site.

Figure 3) but differences arise when larger molecules are considered. Thus, serotonin is a preferential substrate for MAO A whereas dopamine is a better (but not exclusive) substrate for MAO B. The enzymes are also able to act on various xenobiotics. A well-known example is 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP). This is a horribly potent neurotoxin that was initially discovered as contaminant in certain synthetic preparations of illicit drugs. MPTP is oxidized by MAO B becoming an incredibly effective toxin that in matters of hours causes a syndrome very similar to the Parkinson's

disease. Consistently, MPTP has become a widely used chemical tool to generate the disease in animal models. Another example of a peculiar substrate is the amine analog of the hydroxyl group of farnesol, an aliphatic compound that can be oxidized (though not very effectively) by MAO B.

Inhibition and Pharmacology

Because of their key role in the regulation of neurotransmitter levels, MAOs were soon discovered to be valuable targets for the treatment of neurological disorders. The existence of a tyramine oxidase activity was reported in a groundbreaking paper by Hare in 1928. Less than 30 years later, the first MAO inhibitors were reported. It is interesting to remember how these inhibitors were discovered by Zeller and coworkers. The discovery came from the observation that tuberculosis patients receiving isoniazid (isonicotinylhydrazine) therapy experienced mood elevation and, thus, phenylethylhydrazine was developed for use as a MAO inhibitor (MAO-I) for the treatment of depression. This finding initiated an enormous research effort, which lasted for more than 40 years and led to the publication of no less than 15 000 scientific articles on MAO inhibition. These complex efforts culminated in the development (in the late fifties) of tranylcypromine (parnate), an antidepressant still in use. Further important milestones were the development of inhibitors of the so-called propargyl class (they feature a propargylamine group). Among them, deprenyl (selegiline) and, more recently, rasagiline are proving to be an effective drug for the therapy of Parkinson's disease. The rationale for MAO B inhibitors as anti-Parkinson agents is that they increase dopamine levels by blocking MAO-mediated neurotransmitter degradation. This effect should counteract the reduced dopamine levels caused by the progressive degeneration of dopaminergic neurons. An additional effect is thought to be caused by the antioxidant effect of MAO-Is which block key and very active oxidative enzymes such as MAOs. An important point to be made is that these anti-MAO drugs are all covalent inhibitors that, upon binding,

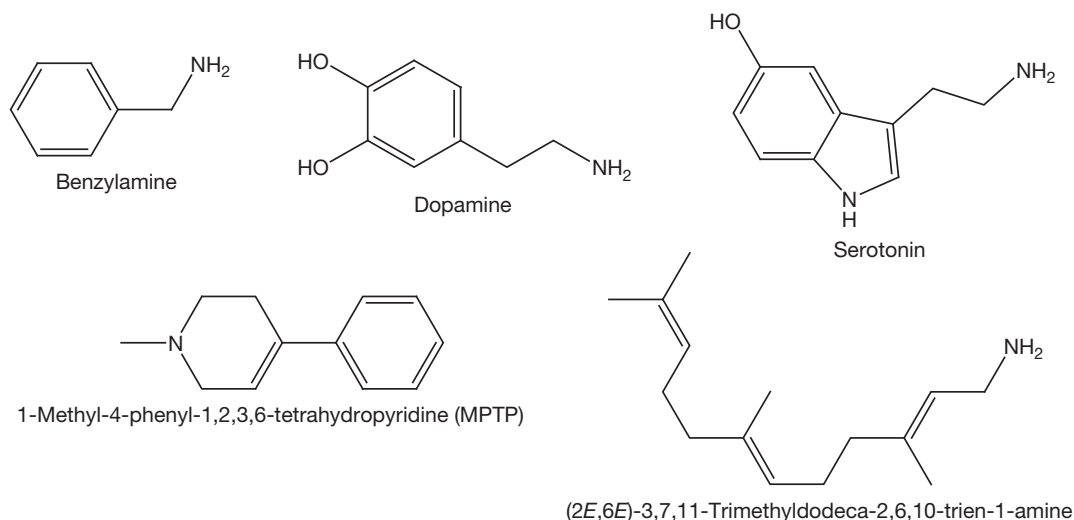


Figure 3 Chemical formulas of MAO A and/or MAO B substrates.

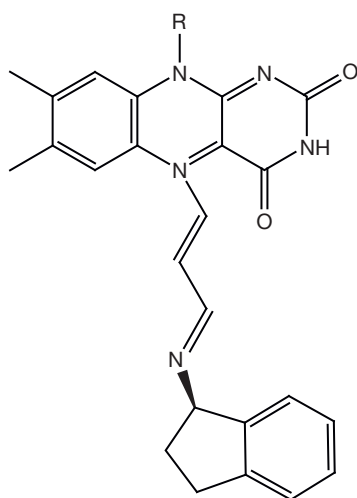


Figure 4 Chemical formulas of the adduct formed by the MAO B flavin with rasagiline, a recently approved anti-Parkinson drug.

react with the flavin forming a covalent adduct (Figure 4). Furthermore, most of these drugs are not very selective because they inhibit (though with different efficacies) both MAO A and B, and, as reported more recently, certain histone demethylases (LSD1 and LSD2) which share weak similarities with MAOs. In pharmacological terms, this lack of selectivity and mode of inhibition is reflected in significant side effects often exhibited by MAO-Is. Patients taking these drugs must monitor their blood pressure and pay attention to their diet, avoiding food that are rich in amines (the so-called cheese effect). At present, a handful of noncovalent highly selective MAO inhibitors are in advanced clinical trials and promise to be very effective for the solution of at least some of these problems.

MAO and Behavior

In recent years, MAOs have been making headlines in newspapers. The reason is that a few studies have linked genetic abnormalities in MAO A expression and activities to aggressive behavior – a MAO A gene has been described as the ‘warrior

gene’. The rationale is that low levels of MAO A activity can lead to increased levels of serotonin, enhancing the propensity to an aggressive and violent behavior. This idea originated from the genetic analysis of a Dutch family with a history of violence. It is of interest to note that this problem is expected to affect only males because both MAO genes are located on the X chromosome. Obviously, this is a hotly debated and highly controversial topic. Although it is hard to imagine that behavior can be simply linked to a single gene alteration, it remains possible that deficiencies in MAO function might affect central nervous system (CNS) activity by modifying neurotransmitter levels. However, this area needs many more studies. It is clear that after 80 years since their discovery, MAOs are likely to remain under the spotlight.

See also: **Bioenergetics:** Amine Oxidases; Reactive Oxygen and Nitrogen Species: Interactions with Mitochondria and Pathophysiology; Respiratory Chain Complex I; **Protein/Enzyme Structure Function and Degradation:** Enzyme Inhibitors; Flavins; **Signaling:** Neurotransmitter Transporters.

Further Reading

- Edmondson DE, Binda C, Wang J, Upadhyay AK, and Mattevi A (2009) Molecular and mechanistic properties of the membrane-bound mitochondrial monoamine oxidases. *Biochemistry* 48: 4220–4230.
- Gibbons A (2004) American association of physical anthropologists meeting. Tracking the evolutionary history of a “warrior” gene. *Science* 304: 818.
- Hare ML (1928) Tyramine oxidase: A new enzyme system in liver. *Biochemical Journal* 22: 968–979.
- Krishnan KR (2007) Revisiting monoamine oxidase inhibitors. *Journal of Clinical Psychiatry* 68(supplement 8): 35–41.
- Langston WJ and Palframan J (1995) *The Case of the Frozen Addicts*. New York: Pantheon Books.
- Wensley D and King M (2008) Scientific responsibility for the dissemination and interpretation of genetic research: Lessons from the “warrior gene” controversy. *Journal of Medical Ethics* 34: 507–509.
- Youdim MB, Edmondson D, and Tipton KF (2006) The therapeutic potential of monoamine oxidase inhibitors. *Nature Review Neuroscience* 7: 295–309.
- Zeller EA and Barsky J (1952) *In vivo* inhibition of liver and brain monoamine oxidase by 1-isonicotinyl-2-isopropyl hydrazine. *Proceedings Society Experimental Biology Medicine* 81: 459–461.

mRNA Polyadenylation in Eukaryotes

J L Manley and D C Di Giammartino, Columbia University, New York, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Capping The process by which a 7-methyl guanosine is added to the 5' end of a precursor-messenger RNA (pre-mRNA) by a 5' to 5' triphosphate linkage. 5' Capping is important for regulation of mRNA nuclear export, prevention of degradation by exonucleases, promotion of translation, and 5' proximal intron excision.

Messenger ribonucleic acid (mRNA) A molecule transcribed from DNA by an RNA polymerase. Mature mRNAs contain 5' and 3' untranslated regions (UTRs), a modified base at the 5' end called 'cap', a polyA tail at the 3' end, and a protein-coding region in between. mRNAs are exported to the cytoplasm where they will be translated into a polymer of amino acids which constitutes a protein.

Polyadenylation The process by which poly(A) polymerase (PAP) adds an unencoded poly(A) tail at the 3' end of a maturing mRNA. The site for poly(A) addition is created by a large protein complex that cleaves the pre-mRNA at the polyadenylation site in the 3' UTR. More than one

polyadenylation site might be present allowing a single gene to encode multiple transcripts with distinct 3' ends, an event called 'alternative polyadenylation'.

Splicing It is a two-step reaction that leads to the removal of an intron and the joining of two adjacent exons. The spliceosome is the large complex that directs this step of pre-mRNA maturation and is made of approximately 200 proteins and five small RNAs.

Transcription The process by which an RNA polymerase synthesizes an RNA molecule on a DNA template. In eukaryotes, there are three RNA polymerases, but only RNA polymerase II is responsible for transcription of protein-coding genes into mRNA.

3' UTR The untranslated region at the 3' end of an mRNA, after the stop codon. It includes regulatory regions such as those that direct cleavage and polyadenylation as well as sequences that regulate mRNA translation and stability by serving as binding sites for microRNAs or regulatory proteins.

Introduction

Eukaryotic precursor-messenger RNAs (pre-mRNAs) must undergo several processing events before the mature mRNA can be transported out of the nucleus and translated into proteins. Transcription, capping at the 5' end, splicing of introns, and polyadenylation of the 3' end are all complex reactions that require numerous proteins (and in the case of splicing, RNAs). Interestingly, these reactions have all been found to be interrelated, such that proteins involved in one step often also participate in the other/s. It is the accurate sequence of events of this highly regulated process that leads to the formation of an mRNA.

All eukaryotic pre-mRNAs, with the exception of replication-dependent histone transcripts, acquire a polyadenylate tail at their 3' end. The maturation of the 3' end of the pre-mRNA is achieved by a two-step reaction that involves an endonucleolytic cleavage of a phosphodiester bond in the pre-mRNA transcript, followed by addition of a polyadenylate tail on the upstream cleavage product, which corresponds to the mature mRNA. In order for these events to occur, four multi-subunit protein complexes, with additional accessory factors, have to bind specific RNA sequences that are found in the primary transcript, and this dictates the precise site where cleavage will occur. The importance of this step is emphasized by the absolute requirement of genes encoding 3' processing factors for cell viability in yeast and higher eukaryotes, and by the association of some human diseases with aberrant polyadenylation events. Such diseases can be caused either by mutations in the *cis*-acting RNA sequences or by mutations/deregulation of the *trans*-acting protein factors involved.

Polyadenylation is a fundamental step of gene expression and its success will determine many aspects of mRNA metabolism: transcription termination by RNAP II (which occurs past the polyadenylation site and has been shown to be dependent on the 3' processing), mRNA stability, mRNA export to the cytoplasm, and the efficiency of translation of the mRNAs into proteins are all dependent on 3' processing. Therefore, the effectiveness by which the 3' protein complex assembles onto defined sequence elements in the 3' untranslated region (UTR) will determine the final level of protein expression.

Recent studies indicate that more than 50% of human genes encode transcripts that contain multiple polyadenylation sites. These alternative sites either can be found in internal introns and therefore alternative polyadenylation will be coupled to alternative splicing events, producing different protein isoforms, or can all be in the 3' UTR, resulting in transcripts encoding the same protein but with 3' UTRs of different length. The length of the 3' UTR can, in turn, affect the stability, localization, transport, and translational properties of the mRNA. Interestingly, differential processing at multiple poly(A) sites can be influenced by physiological conditions such as differentiation and development, or by pathological events such as cancer and viral infection.

RNA Sequences That Direct Cleavage and Polyadenylation

In mammalian cells, three *cis*-elements define the cleavage site. The highly conserved hexanucleotide AAUAAA is found 10–30 nucleotides (nt) upstream of the cleavage site. Extensive

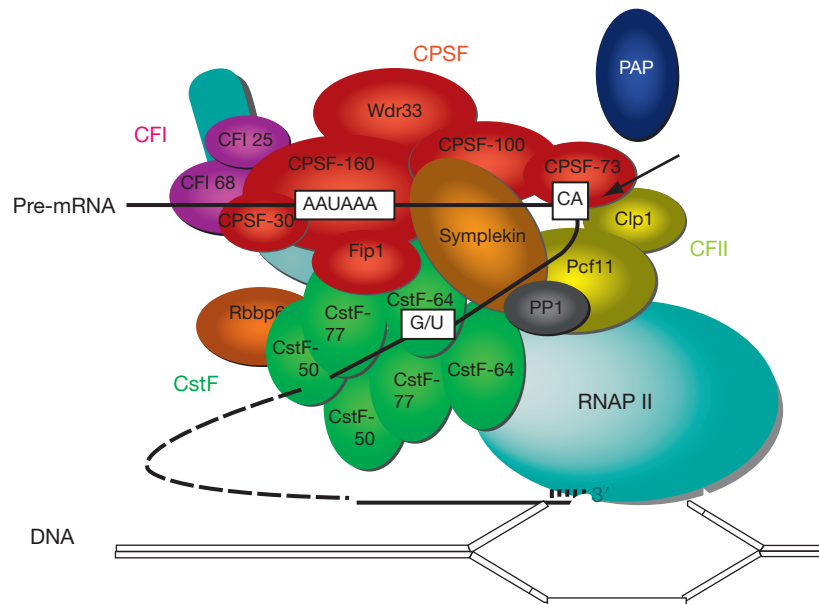


Figure 1 A schematic model of the mammalian pre-mRNA 3'-end processing machinery.

mutagenesis analysis and naturally occurring mutations have proved that this sequence is essential for both cleavage and specific polyadenylation. Variants of the AAUAAA motif are usually associated with alternatively used poly(A) sites.

The second motif of the core polyadenylation signal is the downstream element (DSE), which is located within 30-nt downstream of the cleavage site and consists of a U-rich or/and a GU-rich sequence in the form of YGUGUYY (Y = pyrimidine). This sequence is poorly conserved and mutations or small deletions have only a weak effect on the efficiency of 3' processing.

The cleavage site will be determined by the distance between the AAUAAA and the DSE. The sequence where cleavage will occur is not conserved, but, in 70% of the cases, is located immediately after an adenosine residue that is very often preceded by a cytosine residue. In addition to these three core polyadenylation sequences, an upstream (located upstream of the AAUAAA) and/or downstream (located after the DSE) auxiliary element are sometimes present: the first is a U-rich sequence while the second is a G-rich sequence, but these are very poorly conserved.

In yeast, the RNA signals that direct polyadenylation are different from those used in higher eukaryotes in both sequence and organization and appear to be much less conserved. The yeast poly(A) site is defined by four elements: the AU-rich efficiency element, the A-rich positioning element, the actual cleavage site, and the U-rich elements positioned upstream and downstream of the cleavage site.

Protein Complexes Involved in 3' Processing

The development of appropriate techniques to reproduce the cleavage and polyadenylation reactions *in vitro*, by using cell extracts and *in vitro*-transcribed pre-mRNAs, allowed the

biochemical characterization of the proteins involved in 3'-end maturation of pre-mRNAs.

Two processes as simple as cleavage and polyadenylation, which could be catalyzed by the action of only two enzymes (an endonuclease and a poly(A) polymerase (PAP)), are in fact supported by a vast number of protein factors (**Figure 1**) emphasizing the necessity to finely regulate the process and to coordinate it with other nuclear events.

The core molecular machinery responsible for 3' end formation in mammals includes four multi-subunit protein complexes: cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor I (CFI), and cleavage factor II (CFII). In addition, PAP, symplekin, poly(A)-binding protein II (PABPII), and the C-terminal domain (CTD) of the RNA polymerase II largest subunit are also part of the core machinery. Other proteins have been discovered in a recent proteomic study to associate with the core factors and are very likely to be part of the multi-subunit complexes mentioned above; although they are annotated in **Table 1**, they will not be discussed here, due to the lack of enough published data on their function.

Mammalian CPSF contains five subunits (CPSF-160, CPSF-100, CPSF-73, CPSF-30, and Fip1), which are all essential for efficient cleavage and polyadenylation. The CPSF complex binds to the AAUAAA sequence, mainly through CPSF-160, but the interaction with the RNA is greatly enhanced by cooperative binding between CPSF and CstF. CPSF-160 is known to bind CstF-77 and PAP, and it also associates with proteins involved in transcription initiation (TFIID) and elongation (PolII CTD). CPSF-73 has recently been reported to be the actual endoribonuclease, responsible for the cleavage at the poly(A) site, but it is the entire 3' processing complex that defines the exact cleavage site and confers sequence specificity. CPSF-73 contains a metallo- β -lactamase domain and is the founding member of the β -CASP protein superfamily.

Table 1 Characteristics of the protein factors of the mammalian pre-mRNA 3'-end processing complex

<i>Protein factor (processing step)</i>	<i>Subunits</i>	<i>Yeast homologue (sub-complex)</i>	<i>Sequence characteristics</i>	<i>Protein function</i>	<i>Interacting proteins</i>
CPSF (cleavage and polyadenylation)	CPSF-160	Cft 1p/Yhh1p (CPF)	Three possible β -propellers	Binds the AAUAAA sequence	CstF-77, Pol II CTD, PAP, Fip1, TFIID CPSF-73 CstF-64 symplekin CPSF-100, CstF64, symplekin Fip1 PAP, CPSF-160, CPSF-70, CstF-77
	CPSF-100	Cft2p/Ydh1p (CPF)	Non-metal binding β -lactamase domain		
	CPSF-73	Brr5p/Ysh1p (CPF)	Metallo β -lactamase domain	Endonuclease	
	CPSF-30 Fip 1	Yth1p (CPF) Fip1p (CPF)	Five zinc fingers and one zinc knuckle Pro-rich sequence. RD-rich sequence. Arg-rich sequence	Binds U-rich RNA sequences	
CstF (cleavage)	WDR33	Pfs2 (PFI)	WD repeats	Scaffolding protein, links CstF and CPSF	CPSF-160, CstF-64, CstF-50 Fip1 CstF-77, Symplekin
	CstF-77	Rna14p (CFIA)	HAT domain, Proline rich sequence		
	CstF-64	Rna15p (CFIA)	RRM, pro/gly-rich sequence. MEARA/G pentapeptide motif	Binds to G/U rich sequences	
CFIm (cleavage)	CstF-50		Seven WD40 repeats	Regulatory role during DNA damage	CstF-77, Pol II CTD, BARD1 PAP, CFI-68, PABPII CFI-25
	CFI-25		NUDIX domain, PAP interaction domain	Helps binding AAUAAA	
CFIIm (cleavage)	CFI-68	Clp1p (CFIA)	RBD, SR protein homology in C-terminus	Helps binding AAUAAA	CFIm, CPSF-100, CPSF-73, CstF-64, symplekin Pol II CTD
	hClp1		Walker A and B motifs for ATP binding	Tethers CPSF with CF Im	
Symplekin (cleavage)	Pcf11	Pcf11p (CFIA)	polII CTD interacting motif, two zinc fingers	Mediates interaction between CPSF and CstF Catalyzes the addition of the poly(A) tail to cleaved mRNA, non specific activity by itself	CstF, CPSF. Ssu72, pol II CTD CPSF-160, Fip1, CFIm
PAP (cleavage and polyadenylation)		Pta1p (CPF)	HEAT fold		
		Pap1p	Catalytic core at N-terminus, C-terminus contains RBS, bipartite NLS, ST-rich region		
PABPII (polyadenylation)		Pab1p	Two RRM domains	Responsible for processive elongation and control of poly(A) tail length, stabilizes the tail by binding	CPSF-30
Pol II CTD (cleavage)		Pol II CTD	YSPTSPS repeats (52 in humans)	Essential for co-transcriptional recruitment of CPSF and CstF and for cleavage	CPSF-160, CstF-77, CstF-50, Pcf11
PP1 (polyadenylation)		Glc7p	RS domain, RING finger, zinc knuckle, DWNN, pro-rich	Type 1 protein phosphatase	p53, Rb
RBBP6		Mpe1p (CPF)			

CstF consists of three subunits (CstF-77, CstF-64, and CstF-50), which are necessary for cleavage but not for polyadenylation. CstF-64 contains a typical RNP-type RNA-binding domain (RBD) that binds to the G/U-rich element in the mRNA precursor. In addition, it contains 12 repeats of a pentapeptide motif, MEAR(A/G), the function of which is unknown. In humans, a second isoform, known as tCstF-64, exists. CstF-77 contains a half A TPR (HAT) motif, which mediates its homodimerization. Both CstF-77 and CstF-50 bind specifically to the CTD of Pol II but the latter does so with higher efficiency. The WD-40 repeats in CstF-50 are involved in its interaction with CstF-77. CstF-50 was found to interact by yeast two-hybrid screen with breast cancer 1 (BRCA1) associated ring domain (BARD1), a protein associated with the tumor-suppressor BRCA1. Interestingly, the CstF/BARD1/BRCA1 complex is stabilized under conditions of DNA damage, leading to inhibition of 3' processing (see below for more details). CFI and CFII are required only for the cleavage reaction. The former is a heterodimer made up of CF-25 and one of three related subunits of 59-, 68-, or 72-kDa subunits; its primary function is to provide additional recognition of the pre-mRNA and aid the definition of the proper polyadenylation site. In addition, CFI-68 has recently been reported to be involved in mRNA export to the cytoplasm. CFII also consists of two subunits, Pcf11 and Clp1. Interestingly, Pcf11 has a PolII CTD-interacting domain (CID), and mutations in this domain cause incorrect transcription termination.

PAP is the polymerase responsible for the addition of the poly(A) tail to the cleaved mRNA. Several isoforms of this protein have been reported to arise from the same gene, with PAPII (referred here as PAP) being the predominant nuclear species in most cells. Two additional nuclear poly(A) polymerases transcribed from different genes are of particular interest: neo-PAP, which is overexpressed in human tumors and star-PAP, which is required for the polyadenylation of a group of genes encoding proteins involved in detoxification and/or oxidative stress response. PAP interacts with CPSF160, Fip1, and CF-25. Its N-terminal domain coordinates two metal ions (Mg or Mn) that are required for catalysis, while the CTD is subject to extensive posttranslational modifications that modulate PAP's activity (see below). CPSF and PAP are sufficient for poly(A) addition to a pre-cleaved RNA substrate but rapid elongation and control of poly(A) length (up to 200–300 bases in mammals) requires PABPII. In mammalian cells, PABPII binds nascent tracts of adenylate residues and, along with CPSF, stimulates PAP to switch from distributive synthesis to processive synthesis.

Symplekin is thought to function as a scaffold in the 3' processing complex. Recently, it has also been found to associate with a CTD phosphatase, Ssu72, and to stimulate its catalytic activity. Significantly, this interaction appears to be required for transcription-coupled polyadenylation, but not for polyadenylation of a presynthesized substrate.

The CTD is also required for efficient cleavage *in vitro*. In humans, the heptapeptide YSPTSPS is repeated 52 times and is subject to extensive phosphorylation events that are specific to transcriptional stages. One of these, phosphorylation of the ser 2 position, can enhance 3' processing *in vivo*. As mentioned earlier, several 3' processing factors can associate with

Pol II CTD at the promoter and remain associated during transcription.

The yeast 3' processing complex shares some similarities with the mammalian complex, reflected in a high number of homologous proteins, but the overall complex composition is different. The yeast sub-complexes include the cleavage factor IA (CFIA), cleavage factor IB (CFIB), and cleavage and polyadenylation factor (CPF). The latter, in turn, is composed of CFII and polyadenylation factor I (PFI). CFII and PFI contain subunits that are homologous to the mammalian CPSF, while the CFIA complex shares homologous subunits with mammalian CFII and CstF.

Regulation of mRNA 3' End Formation

One way to regulate cleavage and polyadenylation of pre-mRNA is through posttranslational modification of the 3' processing factors. Importantly, all types of modifications that will be mentioned are covalent but reversible, such that they provide an efficient but temporary way to control gene expression. Some of these modifications are cell-cycle dependent, such as PAP's phosphorylation, while others are triggered by environmental conditions such as stress and DNA damage. Phosphorylated serine, threonine, and, to a lesser extent, tyrosine residues have been detected within the core 3' processing proteins, either by phosphoproteomic screens or by specific biochemical studies. The most explicative example is the cell-cycle-dependent phosphorylation of PAP. During mitosis, PAP is hyperphosphorylated by Cdc2/Cyclin B, which reduces its enzymatic activity, contributing to a general repression of mRNA and protein production during mitosis. Another important posttranslational modification is sumoylation of lysine residues. Proteins modified by small ubiquitin-like modifier (SUMO) are CPSF-73, symplekin, and PAP. This modification was shown to enhance formation of the 3' processing complex as well as the nuclear localization of PAP. Other modifications include lysine acetylation in PAP and CFI25 and methylation of arginine residues in CFI68, although the functional significance of these is unclear.

Posttranslational modifications can regulate the formation of the 3' end of pre-mRNAs by different mechanisms: modification of a protein can interfere with its binding to the rest of the complex, and the absence of that particular protein may, in turn, lead to destabilization of the whole complex, as is the case for sumoylation. Conversely, a posttranslational modification may also lead to recruitment of a binding partner that would not otherwise be associated with the complex. An additional possibility is that the modification leads to a conformational change or a change in the charge of the protein that inhibits the binding of the protein to the mRNA. Some types of modifications, such as sumoylation, are well known to regulate the localization of the modified proteins and affect the nucleo-cytoplasmic shuttling, while others can lead to degradation of the target, for example, in the presence of DNA damage, for example, after UV treatment, the CTD will be ubiquitinated by a pathway involving BRCA1/BARD1, which are recruited to the site of RNA processing by CstF-50. Ubiquitination of CTD causes its proteosomal

degradation, leading to inhibition of transcription. These events would also inhibit the cleavage of pre-mRNA, since the 3' processing factors cannot be recruited properly without the CTD.

In the past 10 years, it has become increasingly evident that usage of alternative poly(A) sites is a common mechanism to regulate gene expression. Fifty percent of human genes encode multiple transcripts derived from alternative polyadenylation. In many mRNAs, the choice of the site changes in response to physiological conditions such as cell growth, development, and differentiation, while in other cases it is triggered by external conditions such as cancer and viral infection. Alternative polyadenylation may or may not be affected by splicing, depending on if the alternative site is present in an intron or in the 3' UTR. A classical example of an alternative polyadenylation is the IgM heavy chain expression, in which increased accumulation of CstF in plasma cells is sufficient to switch the heavy chain from the membrane-bound form to the secreted form during differentiation of B lymphocytes.

Several recent reports have examined alternative polyadenylation using microarray-based techniques and found a general correlation between 3' UTR shortening and cell proliferation and transformation, while there is a preferred usage of distal poly(A) sites during differentiation. 3' UTR shortening can have striking functional consequences, for example, a shorter transcript produces more protein, an effect that can be explained by the absence of inhibitory sequences, such as miRNA-binding sites or other destabilizing elements that would otherwise be present in the longer transcript. The length of the 3' UTR might affect not only the stability but also the localization, transport, and translational properties of the mRNA. The mechanisms that regulate such global events are mostly unknown and intense research is currently being carried out in order to better understand this phenomenon at the molecular level.

Coupling of 3' End Processing with Other Steps of Gene Expression

The maturation of the 3' end of pre-mRNA is closely connected to most steps of gene expression. Starting from transcription initiation, CPSF was first found to associate with TFIID and later, both CPSF and CstF were shown to be located at the transcription start site by chromatin immunoprecipitation experiments. These factors are believed to travel along the gene, during transcription elongation, through their interaction with the CTD. Once the polyadenylation signals are transcribed, the 3' processing factors associate with the signal sequences in the newly synthesized pre-mRNA. The connection with transcription does not end here. Indeed, transcription termination is dependent on the molecular machinery responsible for 3' end formation, and an intact polyadenylation signal has long been known to be necessary for transcription termination of protein-coding genes in human and yeast cells. In the last 10 years, it became increasingly evident that not only transcription and mRNA maturation, but virtually all steps of gene expression are coupled processes that can regulate one another.

A link between splicing and 3' processing has been established by several studies. In one example, the splicing factor

U2AF65, by binding to the polypyrimidine tract at the last intron 3' splice site, stimulates both cleavage and polyadenylation by recruiting the CFI complex to the poly(A) site. The coupling is bidirectional, as evidenced by the fact that U2AF interaction with PAP stimulates splicing of the last intron. Another example of interconnection with the splicing machinery is through the splicing repressor polypyrimidine tract-binding protein (PTB). PTB can play a repressive role by competing with CstF binding to the DSE, but it can also have a stimulating function when associated to hnRNP H. In this case, PTB increases binding of hnRNP H to the G-rich auxiliary element, which in turn stimulates cleavage by recruiting CstF and PAP.

A poly(A) tail has also been proved to be important for an efficient translation of the mRNA into protein; in the cytoplasm the PABPs will facilitate the formation of a loop structure by interacting on one side with the poly(A) tail and on the other with the elongation initiation factor eIF4G, which in turn binds to the 5' cap of the mRNA, thereby forming a closed circle which enhances translation. Other essential functions of the poly(A) tail are to promote transport of mRNA from the nucleus to the cytoplasm and to enhance mRNA stability, by preventing the action of the exosome complex of 3'→5' exonucleases.

Studies over the past 20 years have identified numerous factors involved in pre-mRNA 3' end processing, but more research is needed to elucidate the functional properties of many of the new protein factors found to be associated with this complex, especially those that might connect cleavage and polyadenylation with other nuclear events. An extensive characterization of the cross talk between 3' processing and other cellular processes will be important for better understanding gene regulation on a global level.

See also: [Molecular Biology: Messenger RNA Processing in Eukaryotes](#); [RNA Polymerase II and Its General Transcription Factors](#); [RNA Polymerase II Elongation Control in Eukaryotes](#); [Transcription Termination](#).

Further Reading

- Ahn SH, Kim M, and Buratowski S (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Molecular Cell* 13: 67–76.
- Calvo O and Manley J (2003) Strange bedfellows: Polyadenylation factors at the promoter. *Genes and Development* 17: 1321–1327.
- Colgan DF, Murthy KG, Prives C, and Manley JL (1996) Cell-cycle related regulation of poly(A) polymerase by phosphorylation. *Nature* 384: 282–285.
- Danckwardt S, Hentze MW, and Kulozik AE (2008) 3' end mRNA processing: Molecular mechanisms and implications for health and disease. *EMBO Journal* 27: 482–498.
- Glover-Cutter K, Kim S, Espinosa J, and Bentley DL (2008) RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nature Structural and Molecular Biology* 15: 71–78.
- Hirose Y and Manley JL (1998) RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* 395: 93–96.
- Mandel CR, Bai Y, and Tong L (2008) Protein factors in pre-mRNA 3'-end processing. *Cellular and Molecular Life Sciences* 65: 1099–1122.
- Mayr C and Bartel DP (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138: 673–684.

- Moore MJ and Proudfoot NJ (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 136: 688–700.
- Proudfoot NJ, Furger A, and Dye MJ (2002) Integrating mRNA processing with transcription. *Cell* 108: 501–512.
- Ryan K and Bauer DLV (2008) Finishing touches: Post-translational modification of protein factors involved in mammalian pre-mRNA 3' end formation. *International Journal of Biochemistry and Cell Biology* 40: 2384–2396.
- Shatkin AJ and Manley JL (2000) The ends of the affair: Capping and polyadenylation. *Nature Structural Biology* 7: 838–842.
- Shi Y, Campigli Di Giammartino D, Taylor D, et al. (2009) Molecular architecture of the human pre-mRNA 3' processing complex. *Molecular Cell* 33: 365–376.
- Zhao J, Hyman L, and Moore C (1999) Formation of mRNA 3' ends in eukaryotes: Mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiology and Molecular Biology Reviews* 63: 405–445.

mTOR and its Downstream Targets

C G Proud, University of Southampton, Southampton, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Autophagy A highly conserved process in which proteins and other macromolecules, or even whole organelles, can be degraded. Their components may be recycled to provide building blocks for anabolic processes and/or metabolic energy. Autophagy may also remove damaged cellular components. In autophagy, regions of the cytoplasm are first engulfed by a membrane to form autophagosomes, which then fuse with endosomes or lysosomes, leading to the degradation of the contents. Autophagy also provides a way by which damaged cells can be eliminated.

GTPases Typically small monomeric proteins which bind guanosine triphosphate (GTP) and can hydrolyze it to guanosine diphosphate (GDP). In almost cases, the GTP-bound form of the protein is active, so GTP hydrolysis switches it off. Often, the GTPase activity of the protein is low and is enhanced by a GTPase-activator protein (GAP). Regenerating the active GTP-bound form may require a guanine nucleotide-exchange factor (GEF). Both the GAP and GEF can provide regulatory inputs to control the proportion of the GTPase that is in its active form.

Protein kinases Enzymes that transfer phosphate groups from adenosine triphosphate (ATP) onto specific amino acid residues in proteins. This modification often alters the properties of the protein and is a very widespread physiological mechanism for controlling protein function.

Rapamycin A chemical compound first isolated from the bacterium *Streptomyces hygroscopicus* found on Easter Island. The local name for this island is Rapa Nui, hence rapamycin. Rapamycin was first studied as an antifungal agent and was later shown to inhibit the proliferation of cells of the immune system.

Scaffold proteins Bind to other proteins and thereby mediate the assembly of protein:protein complexes, thereby acting to enhance the efficiency and specificity of intracellular signaling pathways.

Tuberous sclerosis Also termed tuberous sclerosis complex (TSC) is a rare genetic disease caused by mutations in the genes for the proteins TSC1 (hamartin) or TSC2 (tuberin). TSC patients develop benign tumors in the brain or other organs such as the kidneys, heart, lungs, or skin. Symptoms may include seizures, behavioral problems, skin problems, lung, and kidney disease; mTOR.

Introduction

Recent years have witnessed a tremendous increase in interest in the mammalian target of rapamycin (mTOR) and its roles in normal cell regulation and in disease. This reflects improved knowledge about this protein kinase and its cellular functions, in particular the control of its function; the cellular processes that it regulates; and its importance in disease. In several areas, including oncology, there is a high level of interest in devising ways to inhibit mTOR function for therapeutic reasons.

mTOR and mTOR Complexes

mTOR is a protein which is composed of several distinct domains (Figure 1). These include the so-called HEAT repeats, which generally mediate protein:protein interactions; an FRB (FK506 and rapamycin-binding domain); and a kinase domain. The kinase domain is similar to those found in lipid kinases of the phosphatidylinositol 3-kinase (PI 3-kinase) family, although mTOR phosphorylates proteins on serine or threonine residues, not lipids. The FRB domain interacts with a complex between a drug termed rapamycin and a small protein (an 'immunophilin') called FKBP12. The rapamycin-FKBP12 complex inhibits some of the actions of mTOR, but it is not clear how. mTOR also contains the so-called FAT and FATC domains

(Figure 1) which may mediate protein:protein interactions and confer control of mTOR function by cellular redox status, although this requires further study.

mTOR binds several other proteins and forms two distinct kinds of multiprotein complex in the cell. mTOR complex 1 (mTORC1) contains mTOR, a large scaffold protein termed raptor (regulatory associated protein of mTOR), and mLst8 or GβL. Raptor plays a key role on the function of mTORC1 by interacting with substrates and recruits them to be phosphorylated by this complex. Such substrates include the p70 S6 kinases, which contain a 5 amino acid TOS (for TOR signaling) motif (see Figures 1 and 2).

mTOR complex 2 (mTORC2) also contains mTOR and mLst8, but not raptor. Instead, it contains other proteins, including rictor (rapamycin-insensitive companion of mTOR), protor, and mSin1 (Figure 1). mTORC2 phosphorylates different protein substrates from mTORC1; rictor may act as a scaffold to recruit mTORC2's substrates as raptor does for mTORC1.

Studies in genetically modified mice have revealed the key importance of mTOR and raptor: knocking out mTOR in mice causes a very early (pre-implantation) lethality, while raptor null blastocysts stop growing. Rictor knockout embryos show a developmental delay and die about e10.5. These data attest to the central role that mTOR complexes play in cellular function and in development.

Rapamycin blocks the phosphorylation of certain substrates of mTORC1, such as p70 S6 kinase, but not all of them. In the

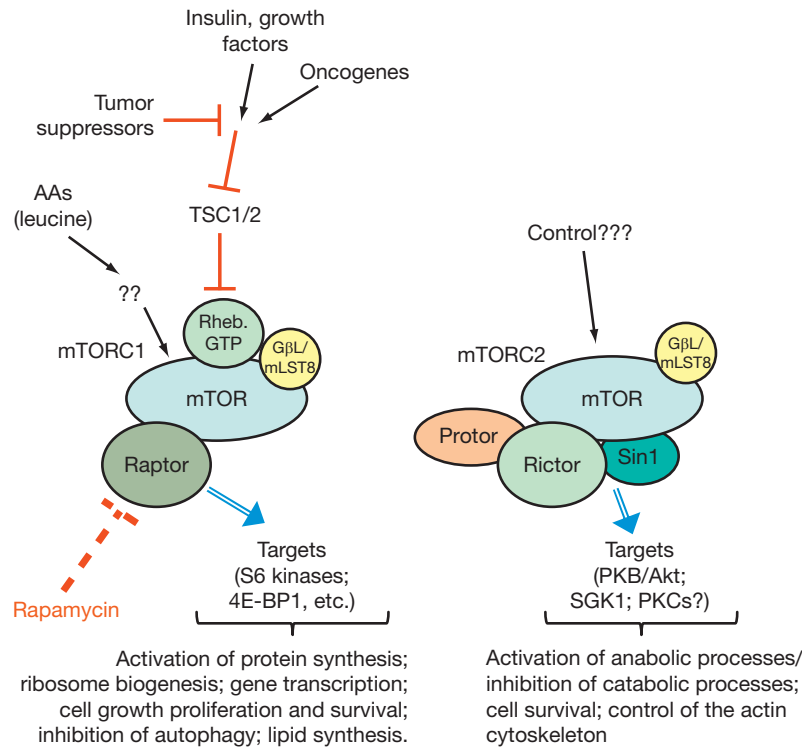


Figure 1 The composition of mTORC1 and mTORC2 and their best-characterized targets are depicted, together with an indication of the known mechanisms by which mTORC1 activity can be regulated. Also indicated are the main processes known to be controlled via these complexes.

short term, it does not affect the function of mTORC2, although it can do over longer time periods (24–72 h) in certain cell types.

Use of rapamycin has revealed that mTORC1 plays important and diverse cellular roles, as described in detail below. An early very important finding was that rapamycin blocks cell-cycle progression late in G1. It thus inhibits, for example, the activation of T lymphocytes, a feature that led to the application of rapamycin as an immunosuppressant to prevent (kidney) graft rejection.

Substrates for mTORCs

The S6 kinases were the first proteins discovered to be controlled by mTORC1 by virtue of the ability of rapamycin to inhibit their activation. S6Ks contain a TOS motif near their N-termini which allows them to interact with raptor and be recruited to mTORC1 to be phosphorylated by mTOR (Figure 2). The regulation of the S6 kinases shows the classical hallmarks of mTORC1 signaling, that is, they are activated by hormones and growth factors, and this activation (in many types of cells) is abrogated if cells are starved of amino acids and is also completely blocked by rapamycin.

The second well-understood group of mTORC1 targets are the eukaryotic initiation factor (eIF) 4E-binding proteins (4E-BPs; Figure 2). 4E-BPs bind to eIF4E, the translation initiation factor that interacts with the 5'-terminal 'cap' structure of eukaryotic mRNAs. When bound to a 4E-BP, eIF4E cannot interact with the scaffold protein eIF4G. mTORC1 directly

phosphorylates 4E-BP1, the best-understood 4E-BP, triggering its release from eIF4E, thereby activating eIF4E.

The functions of these mTORC1 substrates are discussed in detail below.

Other direct substrates for mTORC1 include the proline-rich Akt substrate of 40kDa, PRAS40, and ULK1, a protein kinase involved in controlling autophagy. PRAS40 has been described as both an upstream regulator and a downstream substrate of mTORC1; its full biological function remains obscure.

Regulation of mTOR Complexes

Signaling through mTORC1 is regulated by a wide variety of stimuli and conditions. It is activated by growth factors, mitogens and hormones, such as insulin. Key to the control of mTORC1 are Rheb, a small GTPase protein, and the proteins TSC1 and TSC2 (tuberous sclerosis 1 and 2). When bound to GTP, Rheb can stimulate mTORC1 signaling. TSC1/TSC2 act to hydrolyze Rheb-bound GTP to GDP, thus inactivating Rheb. TSC2 acts as a GTPase-activator protein (GAP) for Rheb. Agents that activate PI 3-kinase signaling switch on protein kinase B (also called Akt) which phosphorylates and apparently inactivates TSC2. This allows Rheb.GTP to build up, activating mTORC1. Stimuli that activate the classical mitogen-activated protein (MAP) kinase (extracellular ligand-regulated kinase (ERK)) pathway also appear to inactivate TSC2, activating mTORC1. As both the PI 3-kinase and ERK pathways contain proto-oncogenes (PI 3-k, Akt; Ras, Raf) and are inhibited by tumor suppressors (phosphatase and tensin homolog (PTEN);

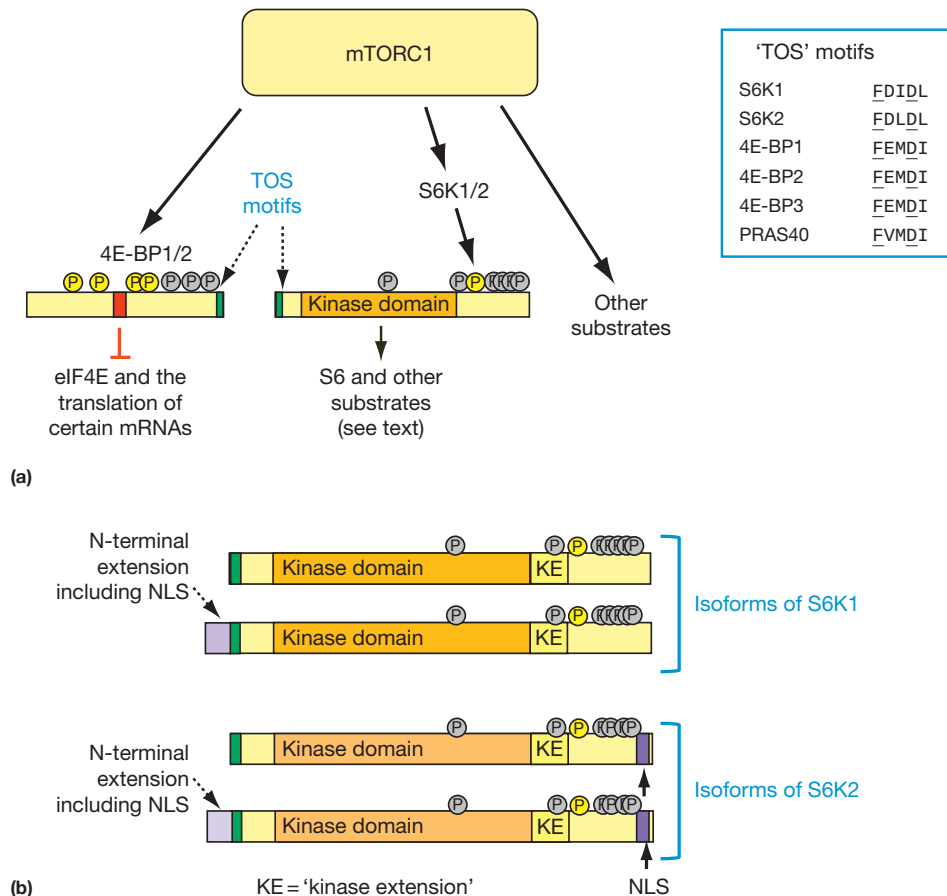


Figure 2 (a) The best-known substrates for mTORC1 are shown schematically; TOS motifs are indicated. Phosphorylation events catalyzed by mTORC1 are shown in yellow and others in gray. The box shows the TOS motifs in proteins in known mTORC1 substrates; completely conserved residues present are underlined. (b) Cartoon of the layout of S6K isoforms, showing the shorter and longer splice variants of S6K1 and S6K2 including their nuclear localization sequences (NLS). Site shaded yellow is the one phosphorylated by mTORC1.

neurofibromin 1 (NF1)), mTORC1 signaling is very frequently activated in tumor cells. p53, another important tumor suppressor, also inhibits mTORC1 signaling.

mTORC1 is also sensitive to the energy and nutrient status of cells. Starvation of cells for amino acids results in profound inhibition of mTORC1 signaling, leucine generally being the most potent regulator of mTORC1. The mechanism by which amino acids activate mTORC1 remains to be fully identified; it appears to be independent of TSC1/TSC2 and to involve the Rag family of small GTPases.

Depleting cellular energy (ATP, adenosine triphosphate) (e.g., during hypoxia) also results in the impairment of mTORC1 signaling. This effect involves the AMP-activated protein kinase, a key sensor of cellular energy status, which can phosphorylate TSC2. The control of mTORC1 signaling by nutrients and energy status make excellent physiological sense given that mTORC1 promotes energy-requiring anabolic processes such as protein, ribosome, and lipid synthesis.

The control of mTORC2 is not understood; since mTORC1 phosphorylates proteins such as Akt and SGK1 which lie downstream of PI 3-kinase, it may be that mTORC2 is activated in some way by PI 3-kinase signaling, but there is currently no information on this.

Processes that are Controlled by mTORCs

Rapamycin has proved to be a very useful tool for identifying the cellular roles of mTORC1. It should be noted that recent work has demonstrated that not all mTORC1's functions are inhibited by this drug and further functions likely await identification. Because there is no specific inhibitor of mTORC2, far less is known about its cellular roles.

mTORC1

As judged by sensitivity to rapamycin, mTORC1 controls several diverse cell processes, including protein synthesis, ribosome biogenesis, the cell cycle, gene transcription, mitochondrial biogenesis, lipid metabolism, and autophagy. Of these, protein synthesis is the one whose control by mTORC1 is best understood. Regulation of the translation of specific mRNAs contributes to mTORC1's control of cell proliferation and cell survival.

Protein synthesis – initiation and elongation

As already mentioned, mTORC1 phosphorylates the translational repressor 4E-BP1, thus eliciting 4E-BP1's release from

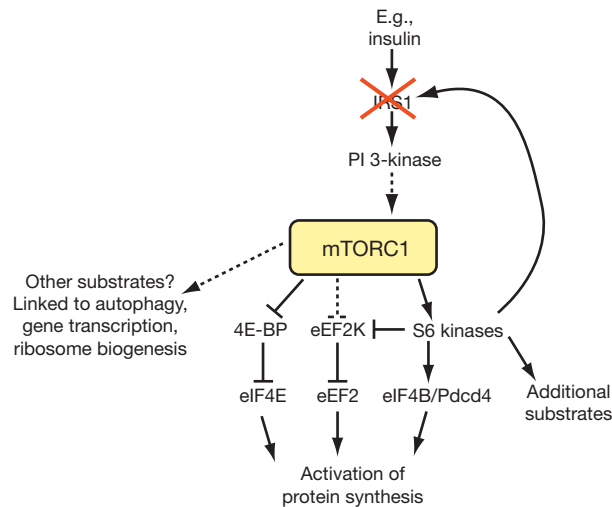


Figure 3 mTORC1 controls several proteins involved in mRNA translation. S6ks phosphorylate IRS1 and inhibit its function (red cross). They may also control other processes. The links between mTORC1 and the control of other processes remain to be elucidated. Dotted arrows indicate indirect links.

eIF4E, allowing eIF4E to bind the scaffold protein eIF4G. eIF4G binds to several other proteins which are involved in the initiation of mRNA translation, that is, the binding of the small ribosome subunit to the mRNA and the identification of the translation start codon. 4E-BPs can thus hinder efficient initiation; this is probably especially important for mRNAs whose 5'-untranslated regions (5'-UTRs) contain secondary structure (stem loops). Such stem loops, between the 5'-cap and the start codon, can hinder the scanning ribosome's movement along the mRNA to find the start codon but can be unwound by eIF4A, an RNA helicase, which is recruited to mRNAs via its binding to eIF4G (Figure 3). Thus, 4E-BPs likely inhibit, in particular, the translation of specific, structured mRNAs. The subset of mRNAs whose translation is thought to be enhanced by mTORC1 via 4E-BP1/eIF4E includes several involved in driving cell-cycle progression (cyclin D1), angiogenesis (VEGF), and other proteins involved in the cell cycle and related events (such as c-myc).

As mentioned above, oncogenic signaling activates mTORC1, while certain tumor suppressors restrain it. A good deal of work has focused upon the importance of mTORC1 signaling in tumorigenesis; a key finding was that rapamycin strongly inhibits the proliferation of certain tumor cells and can inhibit tumor growth in animal models. Several lines of evidence point to a key role of eIF4E, downstream of mTORC1, in tumor biology. Indeed, increasing the levels of eIF4E can transform cells in culture and induce tumors *in vivo*. There is currently high interest in targeting mTORC1 signaling and eIF4E for cancer therapy.

S6 kinases

There are two S6K genes in mice and humans, S6K1 and S6K2, each of which gives rise to two mRNAs by alternative splicing and thus to two distinct proteins (which differ at their N-termini; Figure 2). The longer isoforms possess nuclear localization signals that target them constitutively

to the nucleus where they may regulate ribosome biogenesis and gene transcription.

Ribosomal protein S6, a component of the small ribosomal subunit, was the first substrate for the S6 kinases to be identified; however, the cellular role of S6 phosphorylation still remains unclear. It had been thought that phosphorylation of S6 promoted the translation of a set of mRNAs called 5'-TOP mRNAs (see below for further details). However, data from cells in which both the S6K1 and S6K2 genes have been deleted showed that control of the translation of 5'-TOP mRNAs does not require the S6 kinase. Subsequently, mice were created in which the 5 phosphorylation sites in S6 were mutated to residues which cannot be phosphorylated; data from cells from these mice showed that phosphorylation of S6 is also dispensable for control of 5'-TOP mRNAs.

Meantime, several other substrates for the S6 kinases have been identified, mainly proteins linked to mRNA translation. These include eIF4B and Pdcd4, which are respectively (Figure 3), activators and inhibitors of the helicase eIF4A. This likely assists the translation of mRNAs with structured 5'-UTRs. S6Ks also phosphorylate and inactivate eukaryotic elongation factor kinase (eEF2K) which itself phosphorylates and inhibits eEF2, which mediates the movement of the ribosome along the ribosome during the main, elongation, stage of polypeptide synthesis. Agents that activate protein synthesis elicit the dephosphorylation and activation of eEF2, helping to stimulate protein synthesis. This effect is blocked by rapamycin showing it is mediated by mTORC1. The phosphorylation of eEF2K by S6Ks provides one way that mTORC1 can switch off eEF2K (Figure 3). S6Ks provide a feedback loop from mTORC1 to inhibit signaling through insulin receptor substrate 1 (IRS1), which normally couples the insulin receptor to activation of PI 3-kinase signaling. S6ks can phosphorylate IRS1, thereby impairing its ability to couple the insulin receptor to PI 3-kinase, whereas S6K1 knock-out mice remain sensitive to insulin (Figure 3). Interestingly, wild-type mice fed a high-fat diet become obese and develop insulin resistance, whereas S6K1 knockout mice do not. This feedback loop may be important in obesity and type II diabetes in humans too.

Other substrates for the S6Ks have been reported, including SKAR, a protein involved in pre-mRNA splicing, and CREM, a transcription factor. The physiological role of S6K-mediated phosphorylation of these proteins is not fully understood.

Probably, the most striking phenotype observed in S6K knockout animals is growth restriction. This was first noted for fruit flies when the only S6K gene was inactivated, and later also reported for mice lacking S6K1. In flies, this seems to result from a defect in cell size (growth) rather than cell number (proliferation). The molecular basis of this effect is not yet understood.

Ribosome biogenesis

mTORC1 signaling positively regulates ribosome biogenesis, which is especially important for growing or dividing cells in order to maintain their capacity for protein synthesis. Ribosome biogenesis is a complicated process requiring the synthesis of four different ribosomal RNAs (rRNAs) and around 80 proteins. mTORC1 promotes the transcription of the rRNAs by RNA polymerases I and III, and the translation of the mRNAs that encode ribosomal proteins. The translation

of these mRNAs is controlled by a short pyrimidine-rich sequence at their 5'-end (5'-tract of pyrimidines; 5'-TOP). The 5'-TOP represses their translation in serum-starved cells, this inhibition being overcome by addition of serum in an mTORC1-dependent manner.

Cell cycle and cell survival – via mRNA translation?

As mTORC1 regulates protein synthesis, including the translation of specific mRNAs, it may influence many cellular processes (which are affected by the levels of proteins encoded by mRNAs whose translation is controlled, e.g., by eIF4E/4E-BP1). Two examples are the cell cycle and cell survival. As mentioned above, rapamycin blocks G1-S progression. This is, at least in part, due to the upregulation by mTORC1 of the translation of the mRNA for cyclin D1, which is required for progression into S-phase. Other components are likely also involved in the regulation by mTORC1 of this stage of the cell cycle. mTORC1 may also regulate G2-M progression by controlling the activity of cdc2-cyclin B.

The influence of mTORC1 signaling on cell survival appears to vary, depending on the cell line and perhaps the circumstances. The pro-survival effects of mTORC1 likely involve its ability to promote the translation of the mRNA for Mcl-1, an anti-apoptotic protein.

The abilities of mTORC1 signaling to promote cell proliferation (cell-cycle progression) and cell survival are likely to be of major importance in the roles of mTORC1 in cell transformation and oncogenesis (see below).

Autophagy

Autophagy is inhibited by signaling through mTORC1 but it is only recently that light has begun to be shed upon the mechanisms underlying this. In mammalian cells, mTOR interacts with a complex of proteins which includes a protein kinase termed ULK1 that plays an important role in the induction of autophagy. ULK1 is phosphorylated by mTORC1. It is also suggested that eEF2 kinase plays a role in inducing autophagy. Autophagy may play a role in tumor cells although it remains unclear whether it aids tumor survival or promotes tumor cell death. This likely depends on the tumor and its nutritional status, for example.

Gene transcription

In addition to its role in regulating rRNA synthesis by Pol I/III, mTORC1 can also regulate Pol II-mediated transcription of protein-coding genes. For example, mTOR is required to maintain the oxidative function of mitochondria. Conversely, rapamycin decreases the expression of a mitochondrial transcription factor, peroxisome proliferator-activated receptor- γ coactivator (PGC)-1 α , and of mRNAs for mitochondrial proteins involved in oxidative phosphorylation and the tricarboxylic acid cycle. mTORC1 interacts with the transcription factor YY1, which promotes mitochondrial gene expression. Positive control of mitochondrial function by mTORC1 makes sense given that mTORC1 promotes processes that need energy.

Lipid metabolism can also be controlled by mTORC1. For example, mTORC1 can activate a transcription factor called SREBP1 (sterol-response element-binding protein 1). SREBP1 activates genes for enzymes involved in fatty acid biosynthesis.

mTORC1 may also repress genes involved in lipid breakdown (lipolysis). Thus, mTORC1 signaling can promote the accumulation of lipids. This is likely important, among other things, for growing or proliferating cells which need to create additional membranes.

Protein phosphatases

Signaling through mTORC1 may regulate the activity of protein phosphatases which would contribute to control of the states of phosphorylation of specific proteins in the cell. It is well established that yeast TORC1 controls protein phosphatases PP2, PP4, and PP6. However, much less is currently known about the control of protein phosphatases by mTORC1.

mTORC2

In contrast to the rather extensive knowledge about the functions of mTORC1 and its substrates, much less is known about mTORC2. This largely reflects the absence of a specific inhibitor of mTORC2 with which to define mTORC2's functions analogous to rapamycin for mTORC1. The only substrates for mTORC2 that have so far been identified are members of the AGC group of serine/threonine kinases. AGC family kinases, which also include the S6 kinases, are activated by phosphorylation in their activation loop (within the kinase domain) and at a second site located in a hydrophobic motif. mTORC1 and mTORC2 phosphorylate, respectively, the hydrophobic sites in the S6Ks and in PKB (Akt) and SGK1 (the serum-and glucocorticoid-induced kinase).

The significance of phosphorylation of the hydrophobic site in Akt for its activity against different substrates may vary; its importance for SGK activity seems more clear-cut. mTORC2 may also phosphorylate some other AGC family kinases, for example, members of the protein kinase C (PKC) group.

Like its yeast counterpart, mTORC2 also controls the actin cytoskeleton, perhaps via Rho GTPases.

mTOR in Human Disease

There is currently intense interest in the potential value of inhibiting mTOR signaling to treat human diseases and other conditions. These include sirolimus, ridaforolimus, temsirolimus, and everolimus. As discussed above, rapamycin has been used since the 1990s to tackle kidney graft rejection. Rapamycin, delivered via metal stents, is frequently used to treat restenosis, a complication of cardiac surgery caused by proliferation of smooth muscle cells. Rapamycin itself is a poor drug on account of its pharmacokinetic properties, leading the development of modified versions of rapamycin known as 'rapalogs'. Most recently, specific inhibitors of the kinase activity of mTOR have been developed; these inhibit mTORC1 and mTORC2, and have revealed that not all the functions of mTORC1 are inhibited by rapamycin.

A major area of therapeutic interest is cancer, based on observations such as rapamycin's potent ability to inhibit tumor cell proliferation *in vitro* or tumor growth in animals *in vivo*. A good deal of effort concerns the role of mTORC1-regulated translation in cell proliferation and survival, with eIF4E being considered a potentially valuable therapeutic

target. Autophagy may also be a relevant target for anti-tumor therapies linked to the inhibition of mTOR signaling. Other situations where inhibiting mTOR may be of therapeutic value are neurodegenerative conditions such as Huntington's and Parkinson's diseases; tuberous sclerosis, where initial data indicate benefit even in terms of patients' cognitive functions; and possibly type II diabetes and related disorders.

Concluding Remarks

Recent studies have provided a wealth of information about the cellular roles of mTOR and especially mTORC1 revealing the central role this pathway plays in cell physiology. These discoveries raise further important questions, for example, about which molecules provide the links between mTORC1 and the control of cellular processes such as ribosome production and gene transcription. The answers will be valuable for understanding both the normal roles of mTOR and how its defective control contributes to human disease.

See also: [Cell Architecture and Function: Cell-Cycle Controls in G₁ and G₀](#); [Metabolism Vitamins and Hormones: Insulin Mechanisms/Metabolic Actions](#); [Molecular Biology: Eukaryotic Protein Biosynthesis: The Elongation Cycle; Nucleolus, Overview; Ribosome Assembly; Translation Initiation in Eukaryotes: Factors and Mechanisms](#).

Further Reading

- Avruch J, Long X, Ortiz-Vega S, Rapley J, Papageorgiou A, and Dai N (2009) Amino acid regulation of TOR complex 1. *American Journal of Physiology Endocrinology and Metabolism* 296: E592–602.
- Dann SG, Selvaraj A, and Thomas G (2007) mTOR complex1-S6K1 signaling: At the crossroads of obesity, diabetes and cancer. *Trends in Molecular Medicine* 13: 252–259.
- Drygin D, Rice WG, and Grummt I (2010) The RNA polymerase I transcription machinery: An emerging target for the treatment of cancer. *Annual Reviews of Pharmacology and Toxicology* 50: 131–156.
- Gibbons JJ, Abraham RT, and Yu K (2009) Mammalian target of rapamycin: Discovery of rapamycin reveals a signaling pathway important for normal and cancer cell growth. *Seminars in Oncology* 36(Supplement 3): S3–S17.
- Hoeffer CA and Klann E (2010) mTOR signaling: At the crossroads of plasticity, memory and disease. *Trends in Neuroscience* 33: 67–75.
- Huang J and Manning BD (2008) The TSC1–TSC2 complex: A molecular switchboard controlling cell growth. *Biochemical Journal* 412: 179–190.
- Inoki K and Guan KL (2009) Tuberous sclerosis complex, implication from a rare genetic disease to common cancer treatment. *Human Molecular Genetics* 18: R94–R100.
- Jung CH, Ro SH, Cao J, Otto NM, and Kim DH (2010) mTOR regulation of autophagy. *FEBS Letters* 584: 1287–1295.
- Ma XM and Blenis J (2009) Molecular mechanisms of mTOR-mediated translational control. *Nature Reviews in Molecular Cell Biology* 10: 307–318.
- Polak P and Hall MN (2009) mTOR and the control of whole body metabolism. *Current Opinion in Cell Biology* 21: 209–218.
- Wang X and Proud CG (2006) The mTOR pathway in the control of protein synthesis. *Physiology (Bethesda)* 21: 362–369.
- Wang X and Proud CG (2009) Nutrient control of mTOR, a cell cycle regulator. *Trends in Cell Biology* 19: 260–267.

Mucin Family of Glycoproteins

R B Brown and M A Hollingsworth, University of Nebraska Medical Center, Omaha, NE, USA

Published by Elsevier Inc.

Glossary

Glycoproteins Proteins with oligosaccharide (carbohydrate) chains covalently attached to their polypeptide backbones.

Mucus Viscous aqueous secretions produced by epithelial cells mainly composed by mucins, ions, and water.

N- and O-linked oligosaccharides Oligosaccharide chains linked to serine or threonine and asparagine residues in glycoproteins, respectively.

Peptide motif Peptide sequence that has a critical function during the biosynthesis, processing, or assembly of a protein

(e.g., the CXCC motif is required for dimerization via the mucin CK-like domains), or for the role of a protein.

Protein domain A region in the polypeptide chain of a protein that has a specific function (e.g., the CK-like domain in gel-forming mucins is a dimerization domain).

Secretory pathway Biosynthetic pathway taken by mucins and other glycoproteins that involves different organelles, for example, endoplasmic reticulum and Golgi complex, and many consecutive co-/posttranslational modifications, including folding, N-/O-glycosylation, disulfide bond formation, sulfation, etc.

Mucin Glycoproteins

Mucins are a diverse group of high-molecular-weight glycoproteins expressed by secretory epithelia and other cell types. As components of the cell surface and of secretions, mucins form high-order structures through oligomerization and association with other molecules. Molecular features of mucins create specialized biochemical microenvironments at cellular surfaces and interfaces, specific to each cell type, that enable specific functions associated with those cell types. Many of the molecular features of mucins that are important to their function result from posttranslational processing that creates higher-order structures. For example, most mucins are comprised of greater than 90% carbohydrate that results from posttranslational modification by dozens of glycosyltransferases. Higher-order structures of mucins are created through oligomerization that is mediated by formation of disulfide bonds and other covalent and noncovalent interactions.

One biochemical feature that defines mucins is the presence of tandemly repeated domains rich in serine, threonine, and proline, that are heavily O-glycosylated (mucin-type O-glycosylation of serine and threonine by *N*-acetyl-galactosamine that is extended with linear and branched structures attached by hundreds of different glycosyltransferases). Tandem repeat domains are found in all mucins, but they are highly variable in length, number of repeats, and specific sequence. One function of the repeat domain is to create multivalent structures that serve to create locally high concentrations of specific molecular structures (protein, carbohydrate, or a combination of these, which is termed 'glycopeptide'). These structures function in different ways by serving as ligands for other molecules that are intercalated into a mucin layer or by creating a dense hygroscopic and possibly charged molecular environment that may serve as a barrier or to selectively facilitate transport of molecules across a protective barrier at epithelial surfaces.

Glycosylation

Glycosylation is a posttranslational modification of proteins processed in the secretion pathway of eukaryotes. The reaction is catalyzed by hundreds of glycosyltransferases, each catalyzing a specific glycosidic linkage.

Mucin-Type O-Glycosylation

Mucins are structurally diverse and can range in size from a few kDa to 50 MDa. Most mucins contain a tandem repeat domain that is comprised of conserved sequences of amino acids repeated many times. Some tandem repeats exhibit a high degree of polymorphism in length and/or sequence. The tandem repeat domain is often rich in Pro, and the hydroxy-amino acids Ser and Thr that are modified initially by covalent O-linked attachment of an *N*-acetyl-galactosamine followed by addition of a diverse array of oligosaccharide chains that include fucose, galactose, *N*-acetylneuraminic acid (sialic acid), *N*-acetylglucosamine, and *N*-acetylgalactosamine (GalNAc). O-linked oligosaccharides in the mucin domain comprise 50–90% of the physical mass of the protein and are largely responsible for the structural and biological properties of these glycoproteins, including protease resistance, pathogen sequestration, and binding of ions and water, which enables a high degree of hydration and affects their visco-elastic properties. Several saccharides present in O-linked oligosaccharide chains can be sulfated, including galactose, *N*-acetyl-galactosamine, and *N*-acetyl-glucosamine. These sulfated saccharides protect the oligosaccharide chains from bacterial glycosidases and establish a significant negative charge that may establish ionic repulsive forces between carbohydrate chains, or adhesive properties to positively charged molecules. The repulsive forces stiffen and elongate the protein while preventing rotation around the polypeptide backbone, permitting the molecules to occupy a large volume of the luminal space. The relative abundance of

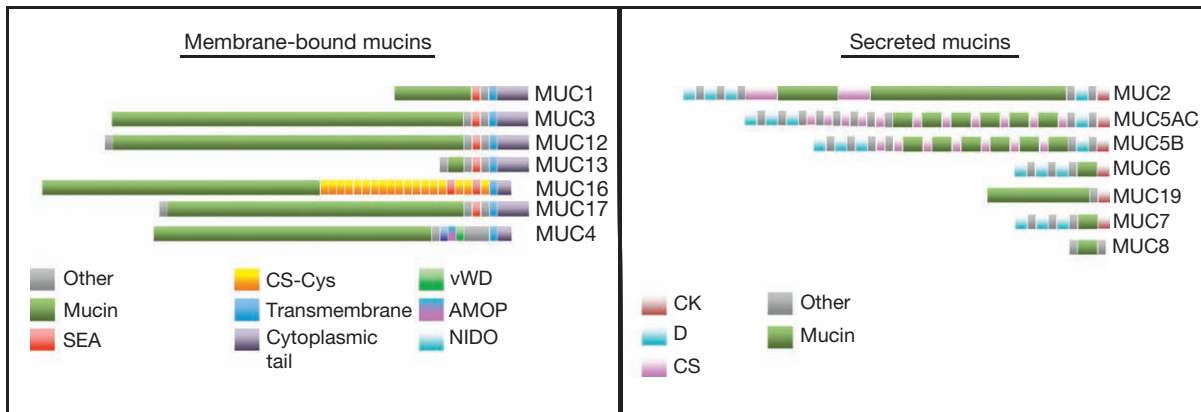


Figure 1 The VNTR consists of repetitive tandem repeats. The theme of a repeat is common to all members of the mucin family but there is significant diversity in sequence and length. This region may exist as a continuous, repeating domain, or the secreted mucins MUC2, MUC5AC, and MUC5B contain cysteine-rich CS domains interspersed throughout the VNTR. The transmembrane and cytoplasmic tail of membrane-bound mucins is found at the far C-terminus. The SEA domain depicted in red is present in all membrane-bound mucins, except MUC4. The posttranslational autoproteolytic cleavage event of these membrane-bound proteins yields two functional protein fragments that remain tightly associated through noncovalent bonds. The CS-Cys domains of MUC16 contributes to formation of tight higher-order oligomerized structures. These cysteine-rich regions are most typically found in the CK, D, and CS domains of the secreted mucins. The oligomerization of these proteins into high-order structures, with high degrees of hydration, give the mucus gel adhesive and viscoelastic properties.

sulfated saccharides available for O-glycan synthesis varies depending on the physiological state of the cell or tissue type. Differences in number and type of O-glycans added to mucins create a diverse range of mucin isoforms in the body.

N-Glycosylation

N-linked glycosylation of mucins typically occurs in cysteine-rich regions outside of the mucin domain, and at a minimum contributes to proper protein folding, and transport to the Golgi. N-linked oligosaccharides are attached via a GlcNAc linkage to the side chain nitrogen of Asn in the consensus sequence NXT/S ($X \neq P$) known as the 'glycosylation sequon'. The precursor branched carbohydrate chain contains three glucose, nine mannose, and two N-acetylglucosamine sugars that are attached by oligosaccharyltransferase to Asn. The removal of the three glucose sugars signals completion of N-linked glycosylation and transport to the Golgi where mannose is removed yielding a carbohydrate chain containing five-nine mannose sugars. Further removal of mannose residues can lead to the core structure containing three mannose and two N-acetylglucosamine residues, which may then be elongated with a variety of different monosaccharides including galactose, N-acetylglucosamine, N-acetylgalactosamine, fucose, and sialic acid, many of which can also exist in sulfated form.

Mucin Protein Family Members

To date there are 17 proteins classified as mucins. There are 10 genes that encode cell-tethered mucins (MUC1, MUC3A, MUC3B, MUC4, MUC12, MUC13, MUC15, MUC16, MUC17, and MUC20). The other seven genes encode for secreted mucins, classified as either secreted polymeric (MUC2, MUC5AC, MUC5B, MUC6, and MUC19) or secreted nonpolymeric

(MUC7 and MUC8). There is a large and diverse group of cellular proteins that contain a mucin tandem repeat domain in their structure; however, some of these proteins have distinct, well-characterized biological functions that do not classify them as typical mucins.

Membrane or Tethered Mucins

The tethered mucins are single-pass integral transmembrane proteins that contain a cytoplasmic tail that associates with cytoskeletal elements, cytosolic adapter proteins, and in some cases facilitates signal transduction. Some mucins exhibit the posttranslational autoproteolytic cleavage event that creates two tightly associated subunits. For example, an autoproteolytic cleavage event occurs at G↓SVVV in the SEA domain of MUC1. Other mucins contain these domains but whether the autoproteolytic cleavage occurs is not yet established. MUC4 does not contain a SEA domain but undergoes autoproteolytic cleavage at a G↓DPH. The large N-terminal subunit containing the tandem repeat is extracellular, while a smaller C-terminal subunit contains short extracellular, transmembrane, and cytoplasmic tail domains. Following autoproteolysis, the two subunits of membrane-associated mucins remain associated by noncovalent interactions. Other biochemical events result in shedding of membrane-associated mucins from the cell surfaces. For example, the amino-terminal subunit of MUC1 can be shed into the extracellular space when cleaved by the proteases such as TACE/ADAM17 and ADAM9. Recent characterization of MUC1 and MUC4 splice variants that lack transmembrane domains provides examples of soluble isoforms that lack the transmembrane domain and are directly secreted from the cell. The lactating mammary gland produces high levels of soluble MUC1 and MUC4 that has been found in breast milk. The ciliated cells comprising the airway epithelium also synthesize soluble forms of MUC1, MUC4, and MUC16 that contributes to 10% of the mucin raft.

Major Domains in Membrane Mucins

Epidermal Growth Factor-Like Domains

The epidermal growth factor (EGF)-like domain is a stretch of 30–45 amino acids containing six conserved cysteine residues, three of which form evenly spaced disulfide bonds that are similar among other EGF-like domain containing proteins. The domain is named due to similarity to the growth factor EGF. The membrane-bound mucins MUC3A, MUC3B, MUC4, MUC12, MUC13, and MUC17 contain these domains between the variable number tandem repeat (VNTR) and transmembrane domain. The functions of these domains are not fully understood, but they may contribute to ligand–receptor interactions between mucins and receptors. For example, there is evidence that MUC4 binds to the EGF-like receptor ERB-B2.

SEA-Like Domains

The SEA-like domain is a stretch of 85–110 amino acids located between the tandem repeat and the transmembrane domain of MUC1, MUC3A, MUC3B, MUC12, MUC13, and MUC17. SEA-like domain contains the GSVVV motif that contributes to autoproteolytic cleavage of the tethered mucins.

Other Domains

MUC4 contains an other protein (AMOP), *nidogen* (NIDO)-like, and adhesion-associated domains. The presence of similar domains on adhesion proteins suggests that they likely serve a function in cellular adhesion.

The Transmembrane Domain

Membrane-associated mucins are anchored in the membrane by a type I transmembrane domain, comprised of 20–35 hydrophobic amino acids near the C-terminus that transverses the membrane once, anchoring the protein to the cell.

The Cytoplasmic Domain

Many membrane-associated mucins contain a short stretch of <100 amino acids at the C-terminus, known as the ‘cytoplasmic domain’, or ‘cytoplasmic tail’. While each cytoplasmic tail varies in length and composition, individual mucin sequences are usually highly conserved among species, suggesting that they mediate fundamental biological processes. One major function of this domain is to engage in signal transduction within the cell in response to alterations in stimuli or the condition of the external environment or cell surface. The best studied of these is MUC1, which can be phosphorylated by a number of receptor tyrosine kinases at the cell surface. The phosphorylation of MUC1 by Met, epidermal growth factor receptor, and other kinases, has been shown to initiate MUC1 internalization and transport into the nucleus where it binds transcription factors at distinct and multiple promoters to control patterns of gene expression. While most studies have focused on Tyr phosphorylation, it is also known that a majority of the 21 Ser, Thr, and Tyr in the MUC1 cytoplasmic tail can be phosphorylated in complex patterns that dictate subsequent

binding to signaling proteins or transcription factors. Thus, signaling through membrane-associated mucins invokes alteration of transcriptional programs in response to external or cell surface stimuli.

Other examples include recent studies into MUC1 signaling that have shown that flagella present in *Pseudomonas aeruginosa* can bind oligosaccharides within the VNTR and initiate a signaling cascade that ultimately leads to ERK1/2 phosphorylation and activation. The pathway is not understood but involves a phosphorylation event that recruits Grb2-Sos-Ras-MEK1/2-ERK1/2.

AP-2 Domain

Some membrane mucins such as MUC1 are internalized into the cell through the formation of clathrin-coated pits via endocytosis. The cytoplasmic tail of MUC1 contains the AP-2 binding motif (YXXO) just C-terminal to the transmembrane domain. The binding of AP-2 to MUC1 is responsible for the recruitment of clathrin and formation of endocytic vesicles.

Secreted Mucins

Secreted mucins are large glycoproteins that lack a transmembrane domain and are usually produced by specialized mucus-producing cells (goblet cells or other specialized cell types in submucosal glands). Secreted mucins have gel-forming properties that are the result of extensive intermolecular cross-linking through disulfide and other covalent and noncovalent molecular bonds. These mucins are major components of the thin or thick mucus that coats many of the epithelial surfaces of the eyes, respiratory tract, gastrointestinal tract, urinary tract, breast, and reproductive tracts. Mucus gels have specialized adhesive and viscoelastic properties that are specific for the biological needs of the different epithelial surfaces on which they are expressed. Different combinations of MUC2, MUC6, MUC5AC, and/or MUC5B together constitute 90% of the mucin in mucus gels. These proteins are packaged and stored in large secretory vesicles known as ‘mucin granules’ that are secreted via exocytosis. The regulated secretion is activated by differentiation, cytokines, chemokines, neurotransmitters, and/or proteases.

Protein Domains in Secreted Mucins

D-Domains

D-Domains are named for their similarity to the D or dimerization domains of von Willebrand factor. The secreted mucins contain three homologous N-terminal D-domains (D1, D2, and D3) and, all but MUC6, a fourth domain, D4, at the C-terminus. A partial D-domain, D', is between D2 and D3 in all secreted mucins. The D-like domain is a cysteine-rich stretch of 325–400 amino acids that contains several conserved N-glycosylation acceptor motifs. The functions of D-like domains include the ability to mediate dimerization, trimerization, and oligomerization of secreted mucins and they may serve to cross-link mucins to other D-like domain containing glycoproteins. There is also evidence that these domains bind other smaller proteins found in secreted mucus, including

trefoil factors, cytokines, chemokines, and other secreted components.

Cystine Knot Domains

The cystine knot (CK)-like domain is a cysteine-rich domain at the C-terminus of secreted mucins, consisting of 90–100 amino acid residues. The domain is extensively N-glycosylated and is believed to contribute to mucin oligomerization through disulfide linkages.

Cys Subdomain Domains

The Cys subdomain (CS) domain is comprised of a cysteine-rich stretch of 100 amino acids. The domain is distributed throughout the VNTR of MUC2, MUC5AC, and MUC5B, and contains a C-mannosylation acceptor motif (WXXW), that is C-mannosylated on Trp, via covalent C–C link attachment of an α -mannopyranosyl residue to the indole C2 carbon of Trp. This form of glycosylation is fundamentally different from N- and O-linked glycosylation and has not been well studied.

Roles in Disease

Pulmonary Disorders

The heavily glycosylated tandem repeat of mucins is key to providing protection of epithelial cells in airways. The mucus layer in the airways requires maintenance of a protected epithelial surface at the air–water interface, laden with bacteria and physical debris, that must also undertake exchange of oxygen and carbon dioxide. The sialylated and sulfated O-glycans offer high degrees of hydration, lining the epithelium, such as the airways, with a layer of mucus. The diversity in glycosylation is likely to allow for specific binding to most, if not all, pathogens that come in contact with airway surfaces. The peristaltic motion provided by cilia move the pathogens attached to secreted mucins for secretion out of the body; however, these hypersecretory responses are also the major contributing factor to the three chronic airway diseases asthma, chronic obstructive pulmonary disorder, and cystic fibrosis. In some circumstances, mucus has been observed to be tethered to the epithelium. These thick layers of mucus typically have altered transport properties and can build in the airways, providing a warm moist environment where bacteria can flourish. These pathogens often activate the innate and adaptive immune response causing inflammation of the airway.

Gastrointestinal Disorders

Mucins are critical components of every epithelial surface of the gastrointestinal tract. The molecular features of the mucins produced by normal epithelial surfaces are specific to the biological requirements of the different organ sites. For example, the needs of the mouth and esophagus require lubrication and protection of surfaces associated with chewing and swallowing of relatively large items of food mixed with liquid. The stomach is lined by a mucus layer that separates the epithelial surface from masticated food in acid at pH 2. Different sections of the intestines deal with digestion and absorption of complex

foods, exchange of waste products, and peristalsis. Consequently, there are different mucin core proteins that are glycosylated in different ways at these different locations. MUC5b and MUC7, produced by the salivary glands, protect the oral cavity and esophagus. The mucus glands of the stomach produce MUC5ac, MUC6, and MUC1. The lower GI tract produces MUC2 and MUC5ac. Disruptions in the production of specific mucin proteins at these sites alters the surface properties of the gastrointestinal epithelia and can lead to alterations in the colonization of these areas by microbial flora, with a change from probiotic bacteria to pathogenic strains. This can also affect the protection of the surface of the epithelia and lead to inflammatory conditions of the intestine and secondary effects including cancer.

Ocular Surfaces

Mucins are an important component of the tear film, providing protection and lubrication of the ocular surface. MUC1, MUC4, and MUC16 are associated with tears, and both MUC2 and MUC5ac have been found in ocular secretions. Differences in the production and glycosylation of these mucins have been reported in conditions of dry eye syndrome and Sjogrens syndrome.

Cancer

Most human adenocarcinomas express mucins, in part, because they arise from the tissues and cell types that normally express mucins. Human tumors typically express molecular forms of mucins that are distinct from the corresponding normal cell type in both types of proteins expressed and in their glycosylation and other posttranslational modification. Studies of human tumors have shown overexpression of MUC1, MUC4, and MUC16 in a variety of cancers. MUC1 overexpression has been observed in greater than 90% of epithelial cancers, including lung, colon, ovarian, breast, and pancreas. For example, extensive studies of the pancreas show low-level expression of MUC1 and little to no MUC4 in normal epithelial cells. MUC4 becomes expressed in precursor lesions known as ‘pancreatic intraductal neoplasms’, or ‘PanINs’. Following malignant transformation, there are significantly higher levels of MUC1 and MUC4 that further increases with tumor progression. The increased level of MUC1 in some breast cancers correlates with poor prognosis and so is used as a prognostic serum marker known as ‘CA15-3’. Similarly, MUC16 (CA125) has longstanding use as a serum marker for ovarian cancer, and, more recently, MUC4 is being studied as a potential marker of aggressive pancreatic cancers.

There are other functional differences observed with tumor progression. MUC1 has been well studied for its ability to act as a cell-signaling protein through the cytoplasmic tail. On normal epithelial cells, MUC1 is found almost exclusively on the apical membrane, whereas following transformation cellular polarity is lost and MUC1 is expressed throughout the plasma membrane. The interaction of MUC1 and growth factor receptors alters cellular signaling pathways and typically leads to high levels of the MUC1 cytoplasmic tail throughout the cytoplasm and nucleus. The VNTR of the extracellular N-terminus of MUC1 undergoes significant alterations in its pattern of

glycosylation. The tumor-derived MUC1 is often aberrantly glycosylated, where the carbohydrate chains tend to be shorter and contain moieties not present on normal MUC1.

The exact roles that mucins play in cancer have not been completely elucidated. There have been many earlier reports that mucins contribute to tumor progression and metastasis through actions at the cell surface that involve the adhesive and anti-adhesive properties of the tandem repeat moiety. The uniform expression of mucin over the entire surface of a depolarized tumor cell would allow for a reconfiguration of the adhesive/antiadhesive properties, allowing the tumor cell flexibility during the process of invasion and metastasis. The steric hindrance provided by these structures could provide a protective boundary from attack by immune cells. The propensity for mucins to interact with adhesion molecules may also play an important role for invasion and metastasis; such is the case with MUC1 binding the endothelial protein ICAM-1, which is expected to play a critical role in invasion. Signal transduction by membrane-associated mucins is undoubtedly important to malignant progression of tumors to a metastatic condition. For example, specific phosphorylated forms of the cytoplasmic tail of MUC1 associate with transcription factors that are involved in cancer progression, such as p53. MUC1/p53 complexes are localized to promoter regions of groups of genes whose expression is altered as a consequence of signaling through the mucin cytoplasmic tail. This results in alterations in transcriptional programs in the cell that affect cell division, motility, apoptosis, differentiation status, or secretion of cytokines and growth factors. It is believed that tumors have appropriated these normal functions in order to survive under different conditions encountered during the phases of uncontrolled growth and metastasis.

See also: **Cell Architecture and Function:** Golgi Complex; **Lipids Carbohydrates Membranes and Membrane Proteins:** Free Oligosaccharides: Formation and Degradation (Free Oligosaccharides Related to N-Linked Glycans); Glycoprotein Folding and Processing Reactions; Mucins in Embryo Implantation; N-Linked Glycan-Processing Glucosidases and Mannosidases; Secretory Pathway; **Protein/Enzyme Structure Function and Degradation:** Disulfide Bond Formation.

Further Reading

- Bafna S, Kaur S, and Batra SK (2010) Membrane-bound mucins: The mechanistic basis for alterations in the growth and survival of cancer cells. *Oncogene* 29(20): 2893–2904.
- Hollingsworth MA and Swanson BJ (2004) Mucins in cancer: Protection and control of the cell surface. *Nature Reviews Cancer* 4(1): 45–60.
- Kufe DW (2009) Mucins in cancer: Function, prognosis, and therapy. *Nature Reviews Cancer* 9(12): 874–885.
- McGuckin MA, Linden SK, Sutton P, and Florin TH (2011) Mucin dynamics and enteric pathogens. *Nature Reviews Microbiology* 9(4): 265–278.
- Sheng YH, Hasnain SZ, Png CW, et al. (2011) Mucins in IBD and colorectal cancer. *Journal of Gastroenterology and Hepatology* 16: 377–383.
- Singh PK and Hollingsworth MA (2006) Cell surface-associated mucins in signal transduction. *Trends in Cell Biology* 16(9): 467–476.
- Tabak LA (2010) The role of mucin-type O-glycans in eukaryotic development. *Seminars in Cell and Developmental Biology* 21(6): 616–621.
- Tarp MA and Clausen H (2008) Mucin-type O-glycosylation and its potential use in drug and vaccine development. *Biochimica et Biophysica Acta* 1780(3): 546–563.
- Thornton DJ, Rousseau K, and McGuckin MA (2008) Structure and function of the polymeric mucins in airways mucus. *Annual Review of Physiology* 70: 459–486.
- Voynow JA and Rubin BK (2009) Mucins, mucus, and sputum. *Chest* 135(2): 505–512.

Mucins in Embryo Implantation

N Dharmaraj and D D Carson, Rice University, Houston, TX, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Daniel D. Carson, volume 2, pp. 765–769, © 2004, Elsevier Inc.

Glossary

Embryo implantation The process by which the mammalian embryo binds to and implants into the uterine wall.

MUC1 The first transmembrane mucin glycoprotein to be completely molecularly cloned.

Mucin Any member of a class of glycoproteins that contain a large number of oligosaccharides linked

through the hydroxyl groups of serine or threonine via *N*-acetylgalactosamine.

Uterus The upper portion of the female reproductive tract into which the mammalian embryo normally implants and in which the fetus develops.

Mucin glycoproteins are defined as proteins heavily substituted with oligosaccharides in O-glycosidic linkage between *N*-acetylgalactosamine and S/T residues. Carbohydrates can account for 50–90% of the mucin molecular weight and greatly influence the physical properties of these molecules. Mucins are also predominant constituents of the apical surfaces of reproductive tract epithelia. In many species, uterine mucin expression changes dynamically during the reproductive cycle and in preparation for embryo attachment. This article summarizes current knowledge of factors controlling the expression and function of uterine mucins with particular emphasis on MUC1 and the process of embryo implantation.

The Implantation Process

The process of embryo implantation into the uterine wall involves complex interactions between embryonic- and uterine-derived factors. These factors include secreted signals as well as cell-surface recognition events. With regard to the latter, a number of cell-surface components have the ability to support embryo attachment and are located at sites of embryo uterine attachment. These include integrins, proteoglycans, and cadherins. The failure of gene knockouts of any of these receptors to prevent the initial events in implantation indicates that there is considerable redundancy in function. This is consistent with *in vitro* studies demonstrating the ability of embryos to attach to a wide variety of cell types and substrates via different cell-surface receptor systems as well as the ability of trophoblast to modulate their cell-surface expression of such receptors in response to the extracellular matrix they encounter (Figure 1).

In spite of the highly invasive nature of blastocysts and trophoblast, the uterus displays the remarkable ability to limit both embryo and tumor cell attachment and invasion. Only during a well-defined window in time does the uterus permit attachment to and penetration of the endometrium. This control appears to be primarily exerted at the first point of encounter, namely the apical surface of the luminal epithelium. In response to ovarian steroid influences, the uterus develops through a series of stages that may be defined as

prereceptive, receptive, and postreceptive with regard to the ability to support embryo attachment. The end of the ovarian cycle is marked by a fall in steroid hormone levels and resetting of the uterine clock to begin another cycle of receptivity. In menstruating species, the endometrium is largely sloughed and new tissue is formed; however, in most species, menstruation does not occur and the endometrium is restructured via processes involving controlled apoptosis and cell proliferation as well as extracellular matrix remodeling. If implantation occurs, steroid hormone levels are maintained and the endometrium not only remains, but also differentiates further in many species in a process called the 'decidual response'.

During the receptive phase, the uterine epithelium also displays remodeling, which is evident by electron microscopy as well as by light microscopy if staining for specific markers is used. While there are many species-specific variations on the theme, there are features of this process that are well conserved across species. These include retraction of apical microvilli and reorganization of the cortical actin filaments, the transient appearance of large membrane protrusions (pinopods or uterodomes), and alterations in the apical glycocalyx. Much of the apical glycocalyx is composed of large, heavily glycosylated glycoproteins identified as mucins. This article briefly summarizes our current knowledge of the function and expression of mucins in the context of the embryo implantation process.

Mucins

Functions

As a general principle, mucins heavily coat the luminal surfaces of epithelia where they provide the first line of defense against infectious agents as well as degradative enzymes. Consistent with this role, mucin glycoproteins are highly resistant to proteolytic attack and require the actions of multiple glycosidases to remove their constituent carbohydrates. While mucins provide a barrier to infection, it is clear that many bacteria and viruses bind well to mucins. Thus, the concept emerges that these molecules serve as traps for the infectious agents that interact with them before they can establish contact with the plasma membrane of the host epithelium. A number

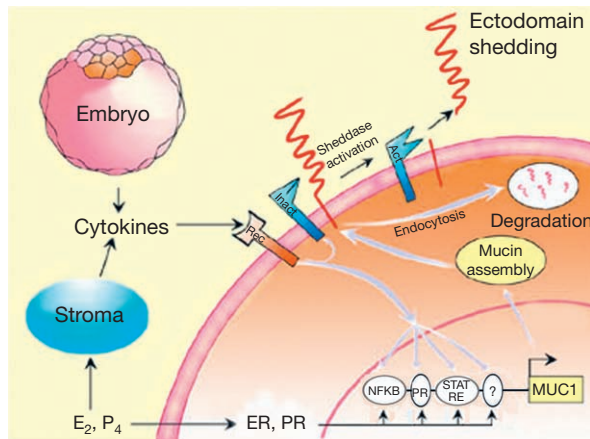


Figure 1 Mucin dynamics during embryo implantation. Uterine epithelial cells produce mucins, including transmembrane mucins such as MUC1 (represented by the squiggly lines in the figure). In general, these mucins are believed to be a barrier to embryo attachment that must be removed to create access to the apical surface of the uterine epithelium. In some cases, mucin gene expression can be regulated by ovarian steroids, estrogen (E_2) and progesterone (P_4) that may act through their nuclear receptors (ER and PR) directly or indirectly on the MUC1 promoter. Cytokines, produced by neighboring stromal cells or the embryo, which bind to cell-surface receptors (Rec) on uterine epithelia can also regulate MUC1 gene expression via NF κ B and STAT binding and transcriptional co-regulators. These same elements also may be downstream of steroid hormone actions. Cytokines can also trigger loss of mucin protein by causing increased expression or conversion of latent (Inact) cell-surface proteases or sheddases to active forms (Act) that catalyze mucin ectodomain release. Mucins may also be removed by endocytosis and intracellular degradation pathways.

of mucins occur as transmembrane proteins (Table 1). MUC1 is an example of a transmembrane mucin glycoprotein and experiments with MUC1-null mice demonstrate that these animals are much more prone to bacterial infection and inflammation than their wild-type counterparts. Thus, a paradox arises over how a glycoprotein that can promote attachment of an infectious agent simultaneously serves to protect against the microbe. As discussed subsequently, the host may be able to recognize attachment and shed the ectodomains of transmembrane mucins in response. This process then would convert the shed mucins into a soluble mucin trap, removing the potential pathogen from the cell surface.

In the context of embryo implantation, mucins also appear to be anti-adhesive. Polarized uterine epithelia of rodents are constitutively nonreceptive *in vitro* and abundantly express mucins. Selective, enzymatic removal of mucins or genetic ablation of Muc1 converts these cells into a functionally receptive state. Early pregnancy studies in the human MUC1 mouse model indicate a reduced, yet persistent level of human MUC1 expression in the peri-implantation stages. Moreover, transfection and overexpression of MUC1 into cells that normally adhere well to blastocysts markedly reduce their adhesive activity. In contrast, under some conditions, mucins can carry selectin ligands and, thereby, potentially support cell attachment. Recent studies indicate that this occurs at the fetal-maternal interface and may contribute to aspects of trophoblast binding.

In addition to being abundantly expressed at the apical surface of uterine epithelia, mucins are physically very large with individual molecules being hundreds of nanometers in length. By contrast, typical cell-surface receptors extend 10–30 nm from the cell surface. Due to their heavily glycosylated nature, individual mucins have large hydration spheres. In addition, mucins often form complex aggregates or gels. Consequently, the mucin coat presents a large, formidable matrix that must be somehow traversed or removed in order for molecules or cells to establish contact with the plasma membrane. The size of the ectodomains is an important determinant of transmembrane mucin function as an anti-adhesive molecule. Truncation mutations of the ectodomains of both MUC1 and MUC4 demonstrate that these mucins must extend more than 50 nm from the cell surface to be effective. This would extend well beyond the distance spanned by most cell-surface receptors.

Other interesting functions associated with mucins include immunosuppression and association with intracellular signal-transducing molecules. In the former case, the mechanism of action is not clear, although effects on several cell types involved in cell-mediated immunity have been reported. In the case of signal transduction, MUC1 has been shown to be associated with β -catenin, Grb2, epidermal growth factor (EGF) receptor, and participates in numerous other intracellular signal transduction pathways. Furthermore, tyrosine phosphorylation of the cytoplasmic tail is stimulated when ligand is added to cells expressing a fusion protein consisting of a ligand binding extracellular domain and the MUC1 cytoplasmic tail. The physiological significance of these interactions has not been demonstrated, although EGF treatment appears to both increase MUC1 association with EGF receptor and activated EGF receptor can tyrosine-phosphorylate the cytoplasmic tail of MUC1, an event suggested to increase MUC1: β -catenin association. Since most of these interactions have only been observed in cancer cells or cancer cell lines, it is possible that these associations only occur in contexts where MUC1 is overexpressed or aberrantly localized. An exception is the association with epidermal growth factor receptor (ErbB) family receptors, which have been shown in both normal and cancer cells.

Expression and Transcriptional Control

To date, 21 mammalian mucin genes have been identified. The predicted protein structures indicate that the protein cores fall into two classes, soluble and transmembrane. As discussed later, even transmembrane mucins may be shed from the cell surface and so may contribute to the pool of soluble mucins. As a general principle, much of the core protein structure is composed of tandemly repeated sequences rich in serine, threonine, and proline, thereby capable of carrying many O-linked oligosaccharide chains. Secreted mucins also usually contain multiple cysteine-rich regions that participate in disulfide bond formation and the assembly of mucin gels. A hallmark feature is the presence of a large number of O-linked oligosaccharides that can constitute 50–90% of the molecular weight of the glycoprotein. A number of mucin oligosaccharide structures have been determined and considerable information on the glycosyltransferases that assemble these structures is available. The oligosaccharides are branched, can be reasonably complex,

Table 1 Uterine mucin expression

Mucin	Type ^a	Found in endometrium?	Species ^b	References
MUC1	TM,Sec	Yes	Hu, Pri, Sh, Bov, Por, Rab, Rat, Mu	Reviewed in Thathiah and Carson (2002)
MUC2	Sec	Yes	Hu	Hebbar et al. (2005), Alameda et al. (2007)
MUC3	TM, Sec	No	Hu	Gipson et al. (1997), Ho et al. (1993)
MUC4	TM, Sec	Yes	Hu, Pig, Rat	Koscinski et al. (2006), Ferrel et al. (2003), Carraway and Iddris (2001)
MUC5AC	Sec	Yes (endometrial hyperplasia)	Hu	Hebbar et al. (2005), Alameda et al. (2007)
MUC5B	Sec	Yes	Hu	Hebbar et al. (2005)
MUC6	Sec	Yes	Hu	Gipson et al. (1997), Alameda et al. (2007)
MUC7	Sec	No	Hu	Gipson et al. (1997)
MUC8	Sec	Yes	Hu	Hebbar et al. (2005) D'Cruz et al. (1996)
MUC9	Sec	Yes	Ham	Martoglio and Kan (1996)
		No	Rab	Yong et al. (2002)
MUC10		ND ^c	-	
MUC11	Sec	ND	-	
MUC12	TM	ND	-	
MUC13	TM	ND	-	
MUC15	TM	ND	-	
MUC16	TM	Yes	Hu, Mu	Gipson et al. (2008), Wang et al. (2008)
MUC17	TM	ND	-	-
MUC18	TM	ND	-	-
MUC19	TM	ND		
MUC20	Sec	ND		
MUC21	TM	ND		

^aTransmembrane (TM) or secreted (Sec); some can be either depending on alternative mRNA splicing.

^bHuman (HU), primates (Pri), sheep (Sh), cattle (Bov), swine (Por), rabbits (Rab), rats (Rat), mice (Mu), hamster (Ham).

^cNot determined (ND).

and carry a net negative charge due to the presence of sialic acid and/or sulfate residues.

The structures of reproductive tract mucin oligosaccharides change during the cycle under the influence of ovarian steroids. Consistent with this are studies showing that glycosyltransferase activities in endometrium also change in parallel. Expression of messenger RNA (mRNA) encoding certain mucin core proteins also changes during the cycle; however, only changes in the expression of MUC1 have been extensively characterized. MUC4 and MUC16 are other mucins that have been studied in any detail, in this regard (see Table 1). In rodents, estrogen strongly stimulates Muc1 expression. Progesterone alone has no effect on Muc1 expression, but antagonizes this estrogen action. Both the actions of estrogen and progesterone on Muc1 expression are mediated by their corresponding nuclear receptors. The murine Muc1 promoter contains a number of sequences that are good candidates for estrogen receptor as well as for progesterone receptor (PR)-binding sites; however, none of these sites appears to be functional with regard to estrogen receptor- α or - β . Thus, estrogen actions are likely to be indirect and may be mediated in a paracrine fashion by estrogen-regulated growth factors or cytokines. Consistent with this is the conservation of a functional STAT and nuclear factor kappa B (NF κ B)-binding site in the mouse and human MUC1 promoter. Proinflammatory cytokines strongly stimulate MUC1 expression in various cellular contexts. In human mammary and uterine epithelial cells, STATs stimulate MUC1 expression via the STAT-binding site. Synergy with other cytokines has been observed, notably tumor necrosis factor- α (TNF- α). In this case, cooperativity with p65 and a corresponding NF κ B-

binding site appear to further enhance MUC1 transcription. It is noteworthy that these cytokines are highly expressed in the mouse uterus at times when MUC1 levels are high as well, for example, estrus and day 1–2 *post coitum*, most likely in response to the microbial challenge associated with coitus. Thus, this response may help protect the reproductive tract from infections that would trigger implantation failure or abortion.

Hormonal regulation of reproductive tract mucin expression is observed in other species as well, although in many cases progesterone stimulates mucin expression. PR regulates MUC1 gene expression in a PR isoform-specific fashion with PRB being stimulatory and PRA antagonizing PRB actions in both, human uterine epithelial cell line and mouse uterine epithelia *in vivo*. The progesterone responsiveness is mapped to a region and not specific to a PR-binding site in the human MUC1 promoter. Progesterone and changes in the levels of proinflammatory cytokines contribute toward a successful implantation. Progesterone and cytokines, in concert with each other, control MUC1 gene expression, regulated by transcriptional cofactors (Figure 2). Steroid hormone receptor coregulators regulate gene transcription by bridging the transcriptional machinery. Furthermore, steroid regulated co-activators (SRCs) are expressed in a cyclical manner in the human endometrium. SRC-1, SRC-3, and nuclear receptor corepressor (NCoR) bind to the MUC1 promoter and regulate MUC1 gene expression. Other cofactors such as p300 are also important in the stimulation of MUC1 gene expression.

In rats, MUC4 expression is also strongly regulated by progesterone and estrogen. In addition, transforming growth factor beta (TGF- β) downregulates MUC4 expression in both

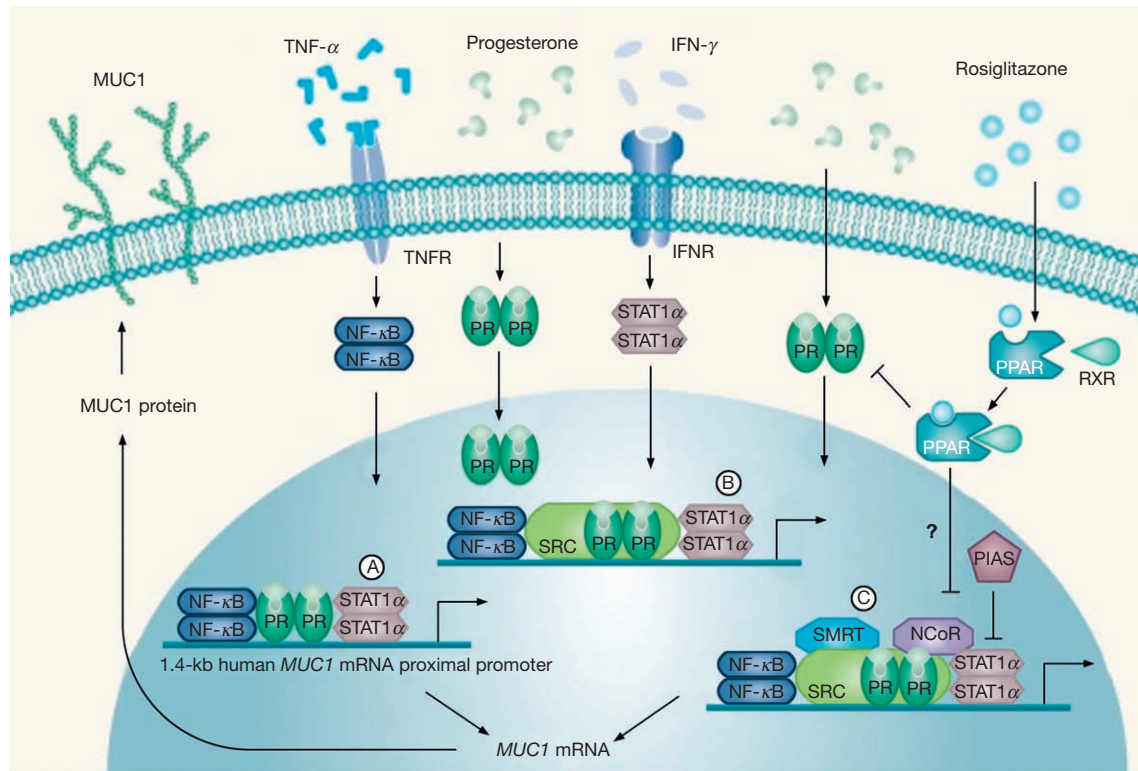


Figure 2 Proposed model for transcriptional regulation of MUC1. This illustration depicts three proposed pathways for regulation of MUC1 expression. Pathway (1) illustrates PR isoforms interacting directly with NF- κ B and STAT1 α to regulate MUC1 expression; (2) illustrates MUC1 expression by interaction between NF- κ B, STAT1 α , nuclear steroid receptor coactivators (SRC), and PR isoforms; pathway (3) illustrates antagonism by PIAS family members and PPAR γ , possibly by recruiting corepressors (NCoR and SMRT). Both PIAS family members and PPAR γ repress MUC1 expression, possibly by altering the transcriptional initiation complex instead of binding to MUC1 promoter directly. Rosiglitazone activates PPAR γ and antagonizes P-stimulated MUC1 expression by interfering with PR-dependent gene activation by as-yet undefined mechanisms and/or causing a decrease in progesterone receptor (PR) protein levels. PIAS represses both basal and P-stimulated MUC1 expression, possibly by reducing the recruitment of a cofactor, NCoR, that appears to have a nonconventional association with the activated, rather than repressed, state of the MUC1 promoter. Reproduced from Carson DD, Dharmaraj N, and Wang P (2008) Transcriptional control of the expression of MUC1. *Expert Review of Endocrinology and Metabolism* 3: 463–471, with permission.

mammary and uterine epithelium. Another transmembrane mucin, MUC16, regulated by steroid hormones, is reduced during receptive phase in humans. Few negatively acting factors on MUC1 expression have been recently reported. Members of the PIAS family were found to inhibit both progesterone- and cytokine-stimulated MUC1 expression. A peroxisome proliferator-activated receptor gamma (PPAR γ) agonist, rosiglitazone, antagonizes progesterone-stimulated MUC1 expression, in contrast to the murine Muc1 expression by activated PPAR γ . Nonetheless, mucin expression changes differently in different parts of the female reproductive tract in response to steroid hormone influences. Thus, regulation of these genes is tissue specific and complex. MUC1 appears to be differentially spliced as well, since forms lacking the extracellular tandem repeats or the cytoplasmic tail can also be detected in certain cell types. It is not clear what fraction of the total MUC1 pool is contributed by these splice variants or exactly what functions these forms play in the implantation process. Other mucins, including MUC3 and MUC4, also undergo alternative splicing to generate transmembrane or soluble variants.

Transport and Metabolism

Mucin assembly involves synthesis and transit of the core protein from the rough endoplasmic reticulum through the Golgi apparatus where most of the glycosylation occurs. Subsequently, the fully assembled mucin is transported to the cell surface. The entire process described above takes 60–90 min for the average mucin molecule in several cell types. Both MUC1 and MUC4 are synthesized from a single mRNA, but are proteolytically cleaved during their biosynthesis, resulting in the formation of stable heterodimers of their respective ectodomains and transmembrane/cytoplasmic tail domains. MUC1 undergoes an autoproteolytic cleavage within the sea-urchin sperm protein, enterokinase, and agrin (SEA) module in the ectodomain. While the exact nature of the association of the heterodimers is not clear, in the case of MUC1 the complex is not dissociable by a number of agents, including urea, heating, high salt, or reducing agents; however, it is readily dissociable by sodium dodecyl sulfate in a variety of normal cells and tumor-derived cell lines.

After arriving at the cell surface, MUC1 undergoes one of the following three fates:

1. it can be recycled and further sialylated intracellularly and returned to the cell surface, or
2. it may be endocytosed and degraded in acidic compartments, presumably lysosomes, or
3. the ectodomains may be released or shed from the cell surface.

Although the process of mucin shedding has been recognized for many years, the identity of the sheddases mediating this release has remained elusive. Recent evidence indicates that tumor necrosis factor- α -converting enzyme/ADAM metallo-peptidase domain 17 (TACE/ADAM17) and membrane type 1 matrix metalloproteinase (MT1-MMP) as such sheddases. TACE/ADAM17-mediated shedding of MUC1 results in a cytoplasmic tail fragment that serves as a substrate for γ -secretase and can undergo rapid intramembranous proteolysis. Mucin shedding is a property of both normal and tumor cells and, indeed, serum levels of shed ectodomains are used as markers of tumor burden in certain cancers. Shedding also may be stimulated by physiologically relevant agents, for example, cytokines, as well as other compounds, such as phorbol esters. In the context of embryo implantation, stimulation of mucin shedding may be the key method to create cell-surface access to embryo receptors in certain species, for example, rabbits and humans. In other species, shedding may be coupled with decreased mucin expression to clear the apical cell surface in preparation for embryo attachment.

Summary and Future Directions

The process of embryo implantation is under the control of the actions of ovarian steroids that coordinate the development of the embryo with the maturation of the endometrium. A panorama of receptors and binding factors may participate in a redundant fashion to promote embryo–uterine interactions during implantation. Nonetheless, under most conditions, the thick mucin coat that covers the apical surface of the uterine epithelium serves not only as a protective barrier to infection and enzymatic digestion, but also as a barrier to embryo attachment. This barrier must be removed in order for embryo attachment to occur. In many species, the same ovarian steroids that coordinate the implantation process also control mucin expression, leading to the loss of the mucin coat at the time of implantation. In other species, including humans, ovarian steroids do not downregulate mucin expression. Rather, localized activation of mucin shedding at the cell surface has been proposed as a mechanism to create access to the apical cell surface of the uterine epithelium. Identification of factors responsible for regulation of mucin gene expression as well as shedding may provide novel avenues to promote fertility as well as improve protection of the reproductive tract from infectious agents. In addition, the findings that certain transmembrane mucins interact with intracellular signal trans-

ducing molecules suggest novel roles for these complex glycoproteins in sensing and triggering responses to the extracellular environment and indicate opportunities for novel therapeutic interventions to modulate mucin expression in normal and pathological states.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Glycoprotein-Mediated Cell Interactions, Nonmucin-Type O-Linked; Mucin Family of Glycoproteins.

Further Reading

- Alameda F, Mejias-Luque R, Garrido M, and De Bolos C (2007) Mucin genes (MUC2, MUC4, MUC5AC and MUC6) detection in normal and pathological endometrial tissues. *International Journal of Gynecological Pathology* 26: 61–65.
- Brayman MJ, Julian J, Mulac-Jericevic B, et al. (2006) Progesterone receptor isoforms A and B differentially regulate MUC1 expression in uterine epithelial cells. *Molecular Endocrinology* 20: 2278–2291.
- Carson DD (2008) The cytoplasmic tail of MUC1: A very busy place. *Science Signaling* 1: Pe35.
- Carson DD, Bagchi I, Dey SK, et al. (2000) Embryo implantation. *Developmental Biology* 223: 217–237.
- Carson DD, Dharmaraj N, and Wang P (2008) Transcriptional control of the expression of MUC1. *Expert Review of Endocrinology and Metabolism* 3: 463–471.
- D'Cruz OJ, Dunn TS, Pichan P, Hass GG, Jr, and Sachdev GP (1996) Antigenic cross-reactivity of human tracheal mucin with human sperm and trophoblasts correlates with the expression of mucin 8 gene messenger ribonucleic acid in reproductive tract tissues. *Fertility and Sterility* 66: 316–326.
- DeSouza MM, Surveyor GA, Price RE, et al. (1999) MUC1/episialin: A critical barrier in the female reproductive tract. *Journal of Reproductive Immunology* 45: 127–158.
- Dharmaraj N, Gendler SJ, and Carson DD (2009) MUC1 during early pregnancy in the human MUC1 transgenic mouse model. *Biology of Reproduction* 81: 1182–1188.
- Gendler SJ (2001) MUC1, the renaissance molecule. *Journal of Mammary Gland Biology and Neoplasia* 6: 339–353.
- Gipson IK, Blalock T, Tisdale A, et al. (2008) MUC16 is lost from the uterodome (pinopode) surface of the receptive human endometrium: *In vitro* evidence that MUC16 is a barrier to trophoblast adherence. *Biology of Reproduction* 78: 134–142.
- Gipson IK, Ho SB, Spurr-Michaud SJ, et al. (1997) Mucin genes expressed by human female reproductive tract epithelia. *Biology of Reproduction* 56: 999–1011.
- Hebbbar V, Damera G, and Sachdev GP (2005) Differential expression of MUC genes in endometrial and cervical tissues and tumors. *BMC Cancer* 5: 124.
- Julian J, Dharmaraj N, and Carson DD (2009) MUC1 is a substrate for gamma-secretase. *Journal of Cellular Biochemistry* 108: 802–815.
- Koscinski I, Viville S, Porchet N, et al. (2006) MUC4 gene polymorphism and expression in women with implantation failure. *Human Reproduction* 21: 2238–2245.
- Martoglio AM and Kan FW (1996) Immunohistochemical localization of oviductin in the endometrial lining of the golden hamster (*Mesocricetus auratus*) during the estrous cycle and early gestation. *Histochemical Journal* 28: 449–459.
- Pimental RA, Julian J, Gendler SJ, and Carson DD (1996) Synthesis and intracellular trafficking of Muc-1 and mucins by polarized mouse uterine epithelial cells. *Journal of Biological Chemistry* 271: 28128–28137.
- Surveyor GA, Gendler SJ, Pemberton L, et al. (1995) Expression and steroid hormonal control of Muc-1 in the mouse uterus. *Endocrinology* 136: 3639–3647.
- Thathiah A, Blobel CP, and Carson DD (2003) Tumor necrosis factor- α converting enzyme/ADAM 17 mediates MUC1 shedding. *Journal of Biological Chemistry* 278: 3386–3394.
- Thathiah A and Carson DD (2002) Mucins and blastocyst attachment. *Reviews in Endocrine and Metabolic Disorders* 3: 87–96.
- Yong P, Gu Z, Luo JP, Wang JR, and Tso JK (2002) Antibodies against the C-terminal peptide of rabbit oviductin inhibit mouse early embryo development to pass 2-cell stage. *Cell Research* 12: 69–78.

Multidrug Resistance Membrane Proteins

N C Rockwell, University of California at Davis, Davis, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

ATP-binding cassette (ABC) transporter A member of a family of membrane transport proteins distinguished by a conserved sequence (the ABC). Most ABC transporters directly couple ATP hydrolysis to vectorial transport across biological membranes.

Cross-resistance The general phenomenon whereby a cell selected for resistance to one toxin simultaneously acquires resistance to other toxins not presented to the cell; multidrug resistance is a special case of cross-resistance.

Homolog A protein with significant sequence similarity to another protein, either across its entire length or in certain regions or domains.

Nucleotide-binding domain (NBD) A conserved domain, or discretely folded protein structure, which is part of an ABC transporter. The ABC signature sequences reside in the NBD, and NBDs power transport by binding and hydrolyzing ATP.

Polyspecific Having very broad specificity, with minimal overlap in chemical structure.

MDR1 and the Clinical MDR Phenotype

Chemotherapy remains one of the main approaches available for the treatment of cancer. In chemotherapy, chemical compounds preferentially more toxic to cancer cells are administered to the patient. However, cancer cells are genetically and biochemically quite similar to the surrounding tissue, so a fairly narrow range of concentrations is available for the administration of chemotherapeutic drugs, and factors that reduce the accumulation of such drugs within the actual cancer cell will confound treatment. In practice, chemotherapy often fails or offers only a brief respite, and there are a number of mechanisms for such failure. In one such mechanism, the cancer cell may express one or more proteins that directly export the drug, preventing its accumulation to effective concentrations. The genetic instability of cancer cells permits them to alter expression of proteins or to develop mutated forms of proteins during the course of the disease, and the use of chemotherapy is in essence a selection for those cancer cells which are able to resist the drug or drugs used. Once such resistant cells arise, continued treatment will cause the resistant cells to dominate the population of cancer cells over time. Often, the appearance of resistance to one chemotherapeutic drug is accompanied by simultaneous appearance of resistance to a range of other drugs that are not structurally related, leading to the name multidrug resistance (MDR). The appearance of MDR is a major complication in chemotherapy.

The Discovery of MDR1

One can study MDR by selecting for resistance to a given drug in tissue culture. In the laboratory of Victor Ling, cells selected for resistance to the chemotherapeutic drug colchicine were shown to have simultaneously gained cross-resistance to a number of other compounds with no obvious overlap in chemical structure. Cross-resistance correlated with the expression of an integral membrane protein named P-glycoprotein or Mdr1, which was subsequently shown to be encoded by the *MDR1* gene (also known as *ABCB1*) through the work of Michael Gottesman and others. The amino acid

sequence encoded by *MDR1* has the hallmarks of an adenosine triphosphate (ATP)-binding cassette (ABC) transporter: the sequence contains a membrane-spanning region with six transmembrane segments followed by a nucleotide-binding domain (NBD) containing the ABC followed by another membrane-spanning region and a second NBD. *MDR1* is thus predicted to be a fully functional transporter in the absence of other proteins. Later studies have confirmed that Mdr1 is an ATP-dependent pump that extrudes drug molecules from the cell.

Clinical Relevance of MDR1

In the years since the discovery of *MDR1*, much research has focused on its role in the clinical outcomes of patients. The idea behind this work can be summarized in two questions: first, does expression of *MDR1* actually correlate with the results of treatment, and second, does blocking *MDR1* function alter the results of treatment for the better? The answer to the first question has proved to be complex. For certain types of cancer, there is clear evidence that *MDR1* is a significant factor in the outcome of treatment. However, evidence for a role for *MDR1* in other types of cancer is much weaker, even in more recent studies. Complications arise in answering this question, such as the role of other transporters in MDR, individual variation in the level of *MDR1* present, and similar individual variation in the presence of any mutations in the *MDR1* sequence. Nonetheless, at this point it seems clear that *MDR1* is a clinically relevant factor in treating certain cancers, and new genetic screening approaches are being developed to permit a finer diagnosis of whether *MDR1* might be a complication in treating other cancers in particular individuals. The answer to the second question has proved to be even more complex. Some studies have reported that blocking MDR function can improve the results of treatment, but others have argued that such inhibition has no effect. Factors which lead to these disparate results include the frequent lack of surrogate assays ensuring that *MDR1* is truly being inhibited, the lack of examination of other transporters in the patient sample populations, and the observation that blocking *MDR1* can increase

the toxicity of chemotherapeutic drugs to the patient and thereby alter the outcome of chemotherapy indirectly.

Structure and Function of MDR1

MDR1 is a member of subfamily B of the ABC transporter superfamily and functions at the plasma membrane. Like most ABC transporters, MDR1 hydrolyzes ATP at the NBDs, and this hydrolysis is coupled to transport of drugs out of the cell. Thus, MDR1 is a direct active transporter, able to work against a concentration gradient for the molecule being exported without the aid of other gradients by coupling ATP hydrolysis to vectorial transport. Due to the clinical relevance of the MDR phenotype, MDR1 has been the focus of intense study. This work has made

good progress toward understanding how MDR1 actually works to extrude drug substrates from cells. As expected, it has been confirmed that ATP binds to the NBDs of MDR1 and other ABC transporters, while transport substrates such as drugs instead bind to the membrane-spanning regions. For MDR1, this work has now culminated in the determination of structures for mouse MDR1 with and without bound inhibitors in 2009.

The Structure and Transport Cycle of MDR1

The overall structure of the protein exhibits two approximately symmetrical halves (Figure 1(a)) and is not substantially changed upon inhibitor binding. The distribution of charged residues on the protein permits assignment of an approximate plane for the hydrophobic part of the lipid bilayer in which

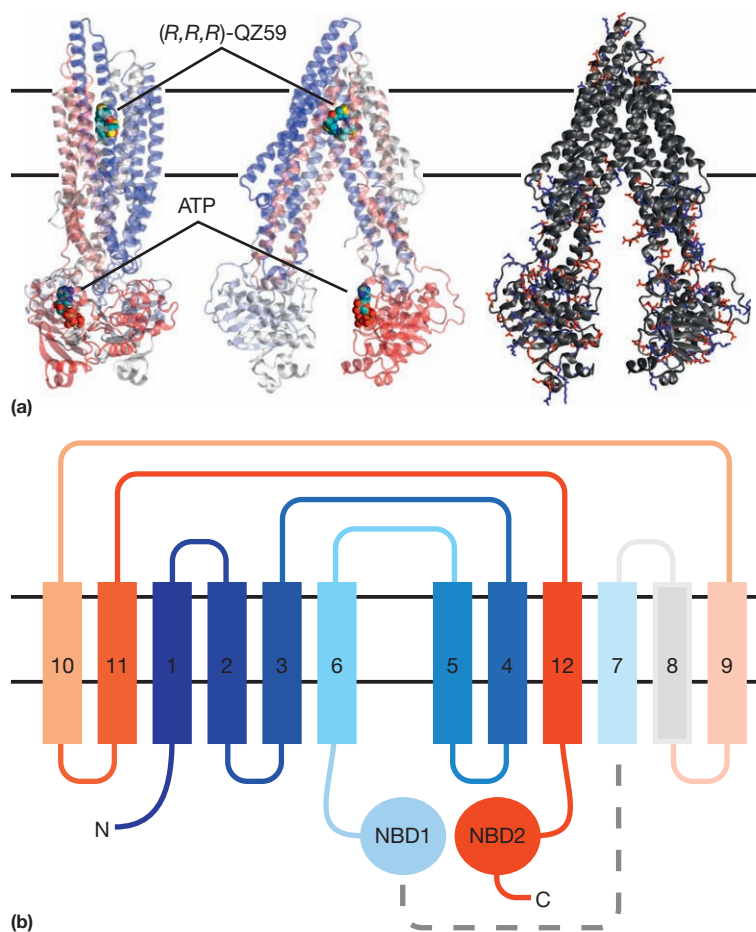


Figure 1 Structure of MDR1. (a) The crystal structure of MDR1 (PDB accession 3G60) is shown with the approximate hydrophobic membrane bilayer indicated by horizontal black lines. The left and center views are colored from the N-terminus (blue) to the C-terminus (red). The bound inhibitor (*R, R, R*)-QZ59 is shown as a space-filling model within the drug-binding pocket. This pocket is not accessible to the extracellular medium or outer leaflet of the bilayer, but it can readily be accessed from the inner leaflet. The approximate position of ATP within one of the two NBDs was determined by superimposing the structure of the HisP NBD (PDB accession 1B0U) onto the MDR1 NBD, and is shown as a space-filling model. The right view shows the backbone in gray, with acidic (Asp, Glu: red) and basic (Arg, Lys: blue) residues shown in stick representation. Charged residues are absent from the region presumed to interact with the hydrophobic portion of the membrane. (b) A cartoon schematic of the MDR1 structure is shown to illustrate the overlapping architecture of the transmembrane segments, in which segments from both halves of the primary sequence are incorporated into each structural half. The protein is color-coded from the N-terminus (blue) to the C-terminus (red), as in (a), and the inter-domain linker connecting the two halves (not resolved in the structure) is shown as a dashed line. (a) PDB accession 3G60: From Aller SG, Yu J, Ward A, et al. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* 323: 1718–1722. PDB accession 1B0U: From Hung LW, Wang IX, Nikaido K, et al. (1998) Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature* 396: 703–707.

MDR1 resides. Interestingly, while the primary amino acid sequence of MDR1 suggests such a symmetrical arrangement, the two membrane-spanning regions observed in the structure are composed of individual transmembrane helices from both halves of the primary sequence (**Figure 1(b)**). Inhibitors bind in a large hydrophobic cavity well situated for lateral access from the inner leaflet of the membrane but not accessible to the outer leaflet. ATP would bind to NBDs, which can transiently dimerize during the transport cycle. Comparison of this structure with the structures of other ABC transporters permits the development of a conceptual transport cycle (**Figure 2(a)**). Drugs partition into the plasma membrane from the extracellular medium due to their hydrophobic character. MDR1 is resting in an inward-facing orientation, such

that drug molecules diffuse laterally in the inner leaflet and enter the MDR1 substrate pocket. In the inward-facing conformation, the drug-binding pocket is not accessible to the outer leaflet or the extracellular medium. Drug binding apparently stimulates ATP binding, implying a cross-talk between the membrane-spanning domains and the NBDs. Structural studies of the MDR1 catalytic cycle from the laboratory of Chris Higgins indicate that major conformational changes take place upon ATP binding, and subsequent crystallographic characterization of other ABC transporters has confirmed that bound ATP switches the transporter to an outward-facing conformation. From this outward-facing conformation, drugs are able to diffuse out of the pocket into either the outer leaflet or the extracellular medium, but they

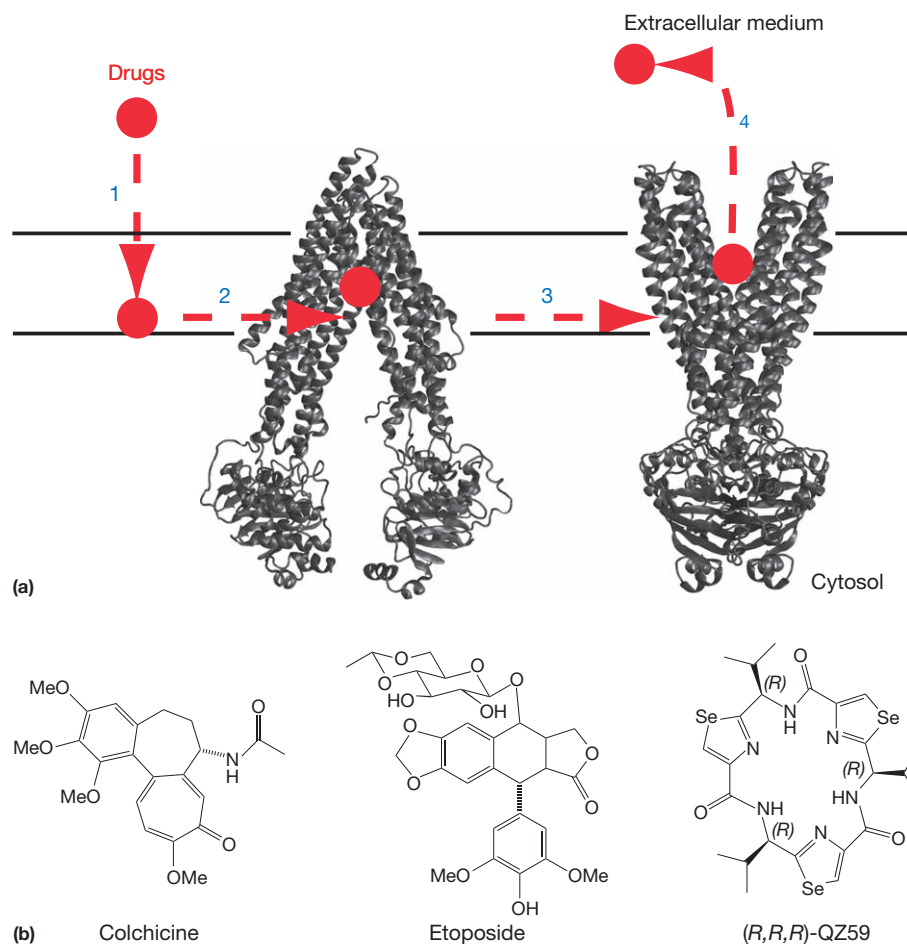


Figure 2 Polyspecific transport of hydrophobic substrates by MDR1. (a) The inhibitor-bound inward-facing structure of MDR1 (left, PDB accession 3G5U) and the nucleotide-bound outward-facing structure of the bacterial ABC transporter MsbA (right, PDB accession 3B60) are shown in the plane of the plasma membrane (black lines) in a hypothetical transport scheme. Drug molecules (red circles) are exported from the cell in four conceptual steps: (1) the hydrophobic character of the drug permits it to cross to the inner leaflet of the plasma membrane; (2) it then enters the substrate-binding pocket of MDR1 via lateral diffusion; (3) ATP then binds and triggers a change to the outward-facing conformation; and (4) the drug then diffuses off MDR1 into either the outer leaflet or the extracellular medium. To have a complete cycle, ATP hydrolysis would then occur and reset MDR1 to the inward-facing state. (b) Structures are shown for three molecules which interact with MDR1: colchicine, originally used in selecting for resistant cell lines that overexpressed MDR1; etoposide, whose pharmacokinetics are regulated by MDR1, MRP2, and MRP3; and (*R,R,R*)-QZ59, one of two inhibitors which has been co-crystallized with MDR1 to permit examination of the MDR1 substrate-binding pocket. (a) Left, PDB accession 3G5U: Aller SG, Yu J, Ward A, et al. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* 323: 1718–1722. Right, PDB accession 3B60: Ward A, Reyes CL, Yu J, Roth CB, and Chang G (2007) Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proceedings of the National Academy of Sciences of the United States of America* 104: 19005–19010.

are no longer able to access the inner leaflet. In MDR1, it is thought that ATP hydrolysis may subsequently serve to reset the transporter and restore the inward-facing conformation. This is in marked contrast to the roles reported for ATP hydrolysis in the distantly related cystic fibrosis transmembrane conductance regulator, an ABC protein which functions as an ATP-gated ion channel rather than as an active transporter and which is mutated in the genetic disease cystic fibrosis.

The Conundrum of MDR1: Broad Specificity with High Affinity

A very wide range of hydrophobic compounds can serve as substrates for MDR1, without obvious structural motifs that distinguish substrates from nonsubstrates (Figure 2(b)). In spite of this, MDR1 is able to transport these compounds with seemingly high affinity (low K_M). Paradoxically, conventional enzymology dictates that high affinity is achieved via extensive interactions between enzyme and substrate, implying conserved structural features in the substrate. However, hydrophobic substrates accumulate in the membrane at much higher concentrations than in the total volume, and such substrates can bind to the inward-facing conformation of MDR1 from the inner leaflet of the plasma membrane. This means that the apparently high affinity of MDR1 for its substrates is misleading and allows a suitably hydrophobic binding pocket to accommodate many substrates in slightly different binding modes with only modest affinity and without making extensive enzyme–substrate contacts. MDR1 is thus a polyspecific transporter, meaning that it has very broad specificity but does actually recognize specific features of its substrates.

Other MDR Membrane Proteins

In addition to MDR1, other ABC transporters have been identified as clinically relevant MDR membrane proteins. Moreover, with the availability of an ever-increasing number of sequenced genomes, it is possible to identify all the ABC transporters within a given organism. This permits the experimentalist to ask whether a given protein can function as an MDR transporter by expressing it in tissue culture or some other recombinant system.

MDR Protein 1

MDR protein 1 (MRP1 or *ABCC1*) was isolated by Susan Cole and Roger Deeley as a protein implicated in MDR in lung cancer cells. MRP1 belongs to subfamily C of the ABC transporters. Whereas MDR1 transports drugs without the need for any chemical modification, MRP1 and other ABC transporters of subfamily C frequently transport drugs that have first been chemically conjugated to glutathione or glucuronic acid. Exceptions to this rule are sometimes co-transported with unlinked glutathione molecules. As with MDR1, the presence of MRP1 correlates with clinical outcome for certain cancers.

ABCG2

ABCG2 (also called MXR or BCRP) was isolated by several labs as a transporter implicated in drug resistance in breast

cancer cells. It belongs to the G subfamily of ABC transporters. The specificity of ABCG2 seems more narrow than that of MDR1 or MRP1, but chemical conjugation of drugs is not required. There are thus at least three transporters thought to be important in the presentation of clinical MDR, any or all of which may be upregulated or mutated in individual tumors.

Homologous Transporters Not Associated with Drug Resistance

Each transporter associated with clinical drug resistance has close relatives not associated with such resistance. For instance, the two closest relatives of MDR1 are MDR3 (encoded by the *ABCB4* gene) and BSEP (encoded by *ABCB11*). Although both proteins exhibit weak drug-efflux activity in tissue culture or *in vitro*, not even one has been clinically implicated in drug resistance. The closest relative to MRP1 is MRP3, which can also confer weak drug resistance when overexpressed in tissue culture. Another relative, MRP2, can be amplified in cell lines selected for resistance to cisplatin and has been detected in clinical samples, but its importance in clinical MDR is not yet clear. Similarly, at least two relatives of ABCG2, ABCG5 and ABCG8, are not implicated in drug resistance.

Physiological Functions of MDR Transporters

While the existence of MDR membrane proteins complicates chemotherapy, such proteins are actually advantageous for organisms in the absence of cancer. As we eat, a number of molecules are ingested that are not part of the normal human metabolism, and such polyspecific transporters can play a key role both in facilitating their excretion and in creating safe havens from such compounds within the organism. We are also beginning to understand the biological functions of homologous transporters that do not contribute to clinical MDR. Examination of the tissue distribution of different transporters provides valuable information about transporter function. The development of knockout mouse models has also been invaluable in such studies, because such models allow phenotypic analysis of null alleles in a genetically tractable organism that retains much of the complexity of human biochemistry and physiology.

The Physiological Role of MDR1

Homozygous *mdr1*^{−/−} knockout mice have altered ability to withstand and excrete toxic compounds known to be MDR1 substrates. This is consistent with the idea that MDR1 is normally responsible for extrusion of toxic metabolites or xenotoxins from cells to allow facile excretion from the body. MDR1 is found at high levels in the blood–brain barrier, presumably forming a safe haven within the organism for the metabolically reduced neurons of the central nervous system. This population of MDR1 has important consequences for drug delivery across the blood–brain barrier, and the use of MDR1 inhibitors to improve bioavailability within the central nervous system is attracting increasing attention.

The Physiological Role of MRP1

As with MDR1, at least one physiological role of MRP1 is the removal of various toxins from the body. MRP1 expression is quite widespread. This protein has a physiological role in the secretion of leukotrienes, and this has been shown to have implications for response to bacterial lung infections. Recently, MRP1 has also been implicated in determining the tissue distribution of cobalamin (vitamin B12).

The Physiological Role of ABCG2

ABCG2 is expressed in intestine and placenta as well as in a side population of hematopoietic stem cells. It is likely to play a role in detoxification, and it has also been recently shown to act as a glutathione transporter, implicating it in maintaining redox balance.

The Physiological Functions of Homologous Transporters

MDR3 and BSEP (the two closest homologs of MDR1) are both implicated in proper function of the liver: MDR3 is an outward-directed transporter for the membrane phospholipid phosphatidylcholine, and BSEP is a bile salt exporter. Mutations in the genes encoding these two proteins interfere with proper secretion of bile, and such mutations are associated with different subtypes of progressive familial intrahepatic cholestasis. MRP2 is the pump responsible for exporting a wide range of anionic conjugates across the apical membrane of polarized cells, and both MRP2 and MRP3 have been proposed to play important roles in the pharmacokinetics of the chemotherapeutic drug etoposide even though they are not associated with clinical MDR. The ABCG2 homologs ABCG5 and ABCG8 form a functional heterodimer which is responsible for the export of plant sterols from the cells of the intestine into the intestinal lumen.

MDR as Part of a Broader Picture

The phenomenon of MDR is by no means unique to mammalian cells. The ABC transporter family is ubiquitous, and every eukaryotic organism examined to date has its own cohort of these proteins. Examples of such ABC transporters from other organisms implicated in drug resistance or detoxification include AtMRP proteins from *Arabidopsis thaliana*, Pdr5 from *Saccharomyces cerevisiae*, and PGPA from *Leishmania*. This

phenomenon is not unique to eukaryotes; the bacterium *Lactococcus lactis* expresses an MDR1 homolog, LmrA, which is able to complement the loss of MDR1 function in mammalian cells. There are also similarly polyspecific drug exporters belonging to other families of transport proteins, such as AcrB from *Escherichia coli*. Thus, mammalian MDR membrane proteins are a special case of a general theme in biology: removal of a wide variety of toxic compounds from the cell via a small number of polyspecific transport proteins. This principle also casts new light on antibiotic resistance, because overexpression or mutation of a transport protein to acquire resistance to an unnecessary or ill-chosen antibiotic may also imply acquisition of cross-resistance to others.

See also: [Bioenergetics: ABC Transporters; Membrane Transport, General Concepts.](#)

Further Reading

- Aller SG, Yu J, Ward A, et al. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* 323: 1718–1722.
- Bates SE (2003) Solving the problem of multidrug resistance: ABC transporters in clinical oncology. In: Holland IB, Cole SPC, Kuchler K, and Higgins CF (eds.) *ABC Proteins: From Bacteria to Man*, pp. 359–391. London: Academic Press/Elsevier.
- Crowley E, McDevitt CA, and Callaghan R (2010) Generating inhibitors of P-glycoprotein: Where to, now? In: Zhou J (ed.) *Special Issue: Multi-Drug Resistance in Cancer. Methods in Molecular Biology*, vol. 596, pp. 405–432. Totowa, NJ: Humana Press.
- Dawson RJ and Locher KP (2006) Structure of a bacterial multidrug ABC transporter. *Nature* 443: 180–185.
- Deeley RG and Cole SPC (2003) Multidrug resistance protein 1 (ABCC1). In: Holland IB, Cole SPC, Kuchler K, and Higgins CF (eds.) *ABC Proteins: From Bacteria to Man*, pp. 393–422. London: Academic Press/Elsevier.
- Ernst R, Kueppers P, Stindt J, Kuchler K, and Schmitt L (2010) Multidrug efflux pumps: Substrate selection in ATP-binding cassette multidrug efflux pumps – first come, first served? *FEBS Journal* 277: 540–549.
- Gillet JP and Gottesman MM (2010) Mechanisms of multidrug resistance in cancer. In: Zhou J (ed.) *Special Issue: Multi-Drug Resistance in Cancer. Methods in Molecular Biology*, vol. 596, pp. 47–76. Totowa, NJ: Humana Press.
- Higgins CF (2007) Multiple molecular mechanisms for multidrug resistance transporters. *Nature* 446: 749–757.
- Hung LW, Wang IX, Nikaido K, et al. (1998) Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature* 396: 703–707.
- Linton KJ, Rosenberg MF, Kerr ID, and Higgins CF (2003) Structure of ABC transporters. In: Holland IB, Cole SPC, Kuchler K, and Higgins CF (eds.) *ABC Proteins: From Bacteria to Man*, pp. 65–80. London: Academic Press/Elsevier.
- Seeger MA and van Veen HW (2009) Molecular basis of multidrug transport by ABC transporters. *Biochimica Biophysica Acta* 1794: 725–737.
- Ward A, Reyes CL, Yu J, Roth CB, and Chang G (2007) Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proceedings of the National Academy of Sciences of the United States of America* 104: 19005–19010.

Muscarinic Acetylcholine Receptors

N M Nathanson, University of Washington, Seattle, WA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Agonist A drug which binds to a receptor and mimics the actions of a natural transmitter.

Antagonist A drug which binds to a receptor and blocks the action of an agonist.

G proteins A family of proteins which bind guanosine triphosphate and which couple receptors to signaling proteins such as ion channels, adenylyl cyclase, and phospholipase C.

Neurotransmitter Chemical signal released by a neuron.

Introduction

Acetylcholine (ACh) is an important neurotransmitter in both the central and peripheral nervous systems. ACh is synthesized in the cytoplasm of nerve terminals by the enzyme choline acetyltransferase, and is then transported into synaptic vesicles. Depolarization of the nerve terminal causes an influx of calcium into the nerve terminal and evokes the release of ACh into the synaptic cleft; the release of ACh can be blocked by botulinum toxin. The actions of ACh are terminated by the enzyme acetylcholinesterase, which hydrolyzes ACh. The activity of acetylcholinesterase can be inhibited by drugs such as neostigmine and the nerve gas agent sarin.

ACh is released from a nerve terminal and binds to a receptor on the cell surface of a target cell to allow transfer of information across a chemical synapse. Sir Henry Dale, in 1914, divided receptors for ACh into two classes based on their distinct pharmacological properties, nicotinic and muscarinic. Muscarinic ACh receptors (mAChRs) are present on central and peripheral neurons, and in such target organs of the parasympathetic nervous systems as cardiac and smooth muscle, and many exocrine glands. The binding of ACh to mAChR can be blocked by antagonists such as atropine.

The mAChR Gene Family

While pharmacological and biochemical analyses suggested, for many decades after Dale's classification, that there was only a single type of muscarinic receptor, drugs were eventually developed that distinguished between muscarinic receptors in different tissues. For example, gallamine was an antagonist at nicotinic receptors that selectively blocked muscarinic receptors in the heart compared to other tissues, while pirenzepine blocked mAChR in certain neurons with higher affinity than in the heart. These studies indicated that there were multiple pharmacologically distinct muscarinic receptors. In the 1980s, mAChR were purified from both heart and brain, which allowed the isolation of complementary DNA (cDNA) and genomic clones encoding the mAChR. There are five subtypes of mAChR, termed M_1 , M_2 , M_3 , M_4 , and M_5 , which are encoded by distinct genes and differentially expressed in various organs and parts of the nervous system (Table 1).

Muscarinic Receptor-Mediated Signal Transduction

Muscarinic Receptors Signaling Utilizes G Proteins

While nicotinic ACh receptors are ligand-gated ion channels (i.e., the same protein both binds ACh and acts as an ion channel), mAChRs are members of the G-protein-coupled receptor (GPCR) superfamily. G proteins are a family of guanosine triphosphate (GTP)-binding heterotrimeric proteins (they consist of three nonidentical subunits) and they couple receptors to various effector proteins, such as adenylyl cyclase, phospholipase C, and ion channels. In the inactive state, guanosine diphosphate (GDP) is bound to the α -subunit. The binding of ACh to mAChR promotes the dissociation of GDP and the binding of GTP to the α -subunit. This, in turn, causes the dissociation of the α -subunit from the $\beta\gamma$ -subunits; both the α - and $\beta\gamma$ -subunits can interact with effector proteins to regulate their activity. The α -subunits have GTPase activity which can be increased by the action of regulator of G-protein signaling (RGS) proteins. This hydrolysis of GTP to GDP promotes the reassociation of the α - and $\beta\gamma$ -subunits and terminates signaling unless ACh is still present to reinitiate the signaling cycle.

Muscarinic Receptor-Mediated Regulation of Second-Messenger Pathways

The mAChR can be divided into two groups on the basis of both sequence similarities and functional responses. The M_1 , M_3 , and M_5 receptors preferentially couple to the G_q family of G proteins, whose α -subunits can activate phospholipase C (PLC). PLC activity produces two intracellular second messengers: inositol trisphosphate, which causes release of calcium from intracellular stores, and diacylglycerol, which activates protein kinase C. The increased intracellular calcium, in turn, can produce many potential effects, such as activation of protein kinases, activation or inhibition of adenylyl cyclase, activation of phosphodiesterases, and activation of nitric oxide synthetase, which can result in increased intracellular cyclic guanosine monophosphate (cGMP) levels. The M_2 and M_4 receptors preferentially couple to the G_i family of G proteins to inhibit adenylyl cyclase activity, thus lowering intracellular cyclic adenosine monophosphate (cAMP) levels. The types of signals produced by any mAChR can depend on both the level

Table 1 Subtypes of muscarinic receptors

Subtype	Number of amino acids	G-protein family	Some tissues with high expression
M ₁	460	G _q	Cortex, hippocampus, striatum
M ₂	466	G _i /G _o	Heart, cerebellum
M ₃	590	G _q	Exocrine glands, smooth muscles cortex, thalamus
M ₄	479	G _i /G _o	Striatum
M ₅	532	G _q	Hippocampus, ventral tegmentum

The table shows the number of amino acids in the five human mAChR subtypes, the G-protein family to which each receptor preferentially couples, and representative tissues which express the various subtypes.

of receptor expression and the particular complement of signaling proteins expressed in a given cell.

Ion Channels

Many of the physiological responses mediated by the activation of mAChR on excitable cells are due to the regulation of the activity of ion channels. Regulation of ion channel activity can occur by many different mechanisms. Modulation of ion channel activity can occur due to protein phosphorylation. For example, activation of protein kinase C or decreased activity of cAMP-dependent protein kinase (due to inhibition of adenylyl cyclase activity) can lead to decreased activity of sodium and calcium channels, respectively. Activation of tyrosine kinases can regulate potassium channel activity.

Intracellular second messengers can directly regulate ion channel activity. Increased intracellular calcium can increase the activity of calcium-activated potassium channels, and increased cGMP can activate cyclic nucleotide-gated ion channels. In addition, $\beta\gamma$ -subunits released following activation of G_i-family G protein can interact with and regulate ion channel activity. For example, inwardly rectifying potassium channels in the heart are activated in this manner following M₂ mAChR activation.

Muscarinic receptor activation can also regulate ion channel activity by regulating the trafficking of channels to or from the cell membrane. In addition, the activity of some ion channels is regulated by PIP₂, and activation of G_q-coupled mAChR can regulate the activity of these channels by the hydrolysis of PIP₂ due to activation of PLC.

Mitogen-Activated Protein Kinase and Related Pathways

Muscarinic receptors can mediate regulation of a series of protein kinase cascades implicated in a wide variety of processes, including cell growth, differentiation, and apoptosis. These protein kinase cascades lead to activation of the mitogen-activated protein kinases (MAPKs) Erk1 and Erk2, p38 kinase, and c-Jun terminal kinase. The particular cascades which are regulated, the mechanisms responsible for this regulation, and whether there is activation or inhibition of a particular kinase pathway depend on both the specific subtype of mAChR and the cell type the receptor is expressed in.

Structural Studies

Transmembrane Topology

Like other members of the GPCR superfamily, the muscarinic receptors are heptahelical, that is, they have seven helical transmembrane domains, with the amino-terminal extracellular and the carboxyl-terminal intracellular. The amino-terminal domains of the various mAChR subtypes contain asparagine residues which are sites for N-linked glycosylation. Based on the crystal structures determined for rhodopsin and other GPCR and on mutagenesis and structural analyses of the mAChR, it is likely that the transmembrane helices are arranged in a bundle.

Ligand Binding to the mAChR

ACh binds to the mAChR by interacting in a crevice with multiple transmembrane domains in the plane of the membrane. A negatively charged aspartic acid residue in the third transmembrane domain interacts with the positively charged ammonium group of ACh. Mutagenesis experiments have also implicated specific residues in the second, fifth, sixth, and seventh transmembrane domains in the binding of cholinergic ligands. Although classical agonists and antagonists bind to the ACh-binding site (the orthostatic site), drugs which act at a distinct (allosteric) site which either positively or negatively affect ACh binding have more recently been identified.

Regions of the mAChR Involved in Coupling to G Proteins

The binding of ACh to the mAChR induces a conformational change in the receptor which allows it to activate the appropriate G protein(s). The regions responsible for determining the specificity of coupling of the mAChR to G proteins has been determined primarily by construction of chimeric receptors between two mAChR, such as M₂ and M₃, which couple to different G proteins. The membrane proximal portions of the third cytoplasmic loop are the main determinants of the specificity of G-protein coupling, although other regions such as the second intracellular domain may also play a role in mAChR–G-protein interactions. A variety of studies indicate that multiple transmembrane domains of the mAChR and other GPCR undergo conformational changes due to agonist activation.

Muscarinic Receptor Actions In Vivo

Muscarinic Receptor Actions in the Periphery

Muscarinic receptors are present in many target organs of the autonomic nervous system. A generalization based on pharmacological and molecular genetic approaches is that the M₂ receptors are present in heart, while the M₃ receptor is present in smooth muscles and exocrine glands; however, this generalization must be tempered by the presence of additional subtypes of mAChR present in low abundance in both cardiac and smooth muscles and by the possibility of differences in subtype expression between different animal species. Muscarinic receptors decrease the rate and force of contraction of the heart, and increase the tone and motility of the gastro-intestinal and urinary tracts. Muscarinic receptors increase secretions from

salivary and sweat glands, and cause bronchoconstriction and increased bronchial secretions. Intravenous administration of ACh causes vasodilation. This is due not to activation of mAChR on the smooth muscle cells but to activation of the mAChR on the endothelial cells in the blood vessel. mAChR activation results in the production of NO, which then diffuses out of the endothelial cells and into the smooth muscle cells, resulting in relaxation of the blood vessel. Muscarinic receptors in pancreatic beta cells regulate insulin release and glucose homeostasis.

Muscarinic Receptor Actions in the Central Nervous System

Muscarinic receptors play an important role in the central nervous system. Many studies have implicated mAChR in learning and memory. Administration of muscarinic antagonists interfere with performance in many learning and memory tests, and the decreased cognition in patients with Alzheimer's disease is thought to be due, in part, to impaired muscarinic transmission. Both pharmacological and genetic experiments have implicated the M₁ mAChR in memory and learning, although it is likely that multiple muscarinic receptor subtypes play a role in these processes.

Muscarinic receptors in the spinal cord and at supraspinal sites (i.e., the brain) can mediate an analgesic response. Genetic studies in mice show that both the M₂ and M₄ receptors are involved in this anti-nociceptive effect.

Muscarinic receptors have been implicated in many other processes as well. For example, muscarinic receptors regulate the function of the basal ganglia, where multiple subtypes of muscarinic receptors modulate dopaminergic signaling. M₅ receptors are present in the ventral tegmentum and are involved in drug-reward mechanisms. M₃ receptors in hypothalamic neurons regulate growth by stimulating the release of growth-hormone-releasing hormone.

Pharmacological analyses carried out over many years have led to the development of a number of drugs which produce their therapeutic actions by either activating or blocking muscarinic receptors. The increased understanding of the actions of individual mAChR subtypes should be a great aid in the

development of drugs with increased therapeutic actions while decreasing effects at undesired mAChR subtypes.

See also: [Signaling: Adenylyl Cyclases; Mitogen-Activated Protein Kinase Family; Phospholipase C.](#)

Further Reading

- Anagnostaras SG, Murphy GG, Hamilton SE, et al. (2003) Selective cognitive dysfunction in acetylcholine M₁ muscarinic receptor mutant mice. *Nature Neuroscience* 6: 51–58.
- Birdsall NJM, Nathanson NM, and Schwarz RD (2001) Muscarinic receptors: It's a knockout. *Trends in Pharmacological Sciences* 22: 215–219.
- Caulfield MP and Birdsall NJM (1998) International union of pharmacology: XVII. Classification of muscarinic receptors. *Pharmacological Reviews* 50: 279–290.
- Conn PJ, Jones CK, and Lindsley CW (2009) Subtype-selective allosteric modulators of muscarinic receptors for the treatment of CNS disorders. *Trends in Pharmacological Sciences* 30: 148–155.
- Duttaroy A, Gomez J, Gan JW, et al. (2002) Evaluation of muscarinic agonist-induced analgesia in muscarinic acetylcholine receptor knockout mice. *Molecular Pharmacology* 62: 1084–1093.
- Gautam Jeon J, Starost MF, et al. (2009) Neuronal M₃ muscarinic acetylcholine receptors are essential for somatotroph proliferation and normal somatic growth. *Proceedings of the National Academy of Sciences of the United States of America* 106: 6398–6403.
- Gutkin JS (1998) The pathways connecting G-protein-coupled receptors to the nucleus through divergent mitogen-activated protein kinase cascades. *Journal of Biological Chemistry* 273: 1839–1842.
- Hulme EC, Bee MS, and Goodwin JA (2007) Phenotypic classification of mutants: A tool for understanding ligand binding and activation of muscarinic acetylcholine receptors. *Biochemical Society Transactions* 35: 742–745.
- Levey AI (1996) Muscarinic ACh receptor expression in memory circuits: Implications for treatment of Alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America* 93: 13541–13546.
- Nathanson NM (1987) Molecular properties of the muscarinic acetylcholine receptor. *Annual Review of Neuroscience* 10: 195–236.
- Nathanson NM (2000) A multiplicity of muscarinic mechanisms: Enough signaling pathways to take your breath away. *Proceedings of the National Academy of Sciences of the United States of America* 97: 6245–6247.
- Wess J (2004) Muscarinic acetylcholine receptor knockout mice: Novel phenotypes and clinical implications. *Annual Review of Pharmacology and Toxicology* 44: 423–450.
- Wess J, Han SJ, Kim SK, Jacobson KA, and Li JH (2008) Conformational changes involved in G-protein-coupled-receptor activation. *Trends in Pharmacological Sciences* 29: 616–625.

Myosin Motors

M A Titus, University of Minnesota, Minneapolis, MN, USA

R E Larson, University of Sao Paulo, Sao Paulo, Brazil

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Roy E. Larson, volume 2, pp. 778–781, © 2004, Elsevier Inc.

Glossary

α -Helix The specific helical arrangement of a polypeptide chain formed by repetitive hydrogen bonding between atoms of peptide bonds separated by four amino acids along the chain.

Calmodulin A calcium-binding protein, found ubiquitously in eukaryotic cells, which acts as a calcium-sensitive regulator of many enzymes via its physical binding to the target protein.

Focal adhesions Sites of attachment of cells to the underlying substratum involving specialized protein complexes linking the intracellular actin cytoskeleton to the extracellular matrix.

Heptad repeat A seven-amino-acid sequence, where the third and seventh amino acids are hydrophobic, which tends to form an amphipathic α -helix able to dimerize with another of its kind via the hydrophobic residues to form a double α -helix referred to as a coiled coil.

Protein conformation The three-dimensional form of the protein considering the specific spatial localization of all the atoms in the molecule.

Quaternary structure The three-dimensional conformation of a protein formed by two or more polypeptides.

Sarcomere The functional contractile unit of skeletal and cardiac muscle cells consisting of interdigitating actin (thin) and myosin (thick) filaments.

Overall Molecular Architecture

Although structural variations among the myosins reflect their specific cellular type and function, and also their phylogeny, there is a basic architectural plan followed by these motors (Figures 1 and 2). The protein structure of most myosins can be divided into three principal domains with generic functional homology, constructed from a long polypeptide – the heavy chain – and several small subunits – the light chains. The heavy chain encodes the highly conserved, mechano-chemical generator, usually referred to as the ‘head’ or motor domain, which is a globular domain that contains the actin binding and adenosine triphosphate (ATP) hydrolysis sites, almost universally positioned at the amino terminal portion of the protein. Immediately following the head domain is an elongated, α -helical region in the heavy chain to which calmodulin and/or calmodulin-like light chains bind, forming the ‘neck’ domain. This domain acts as a lever arm to amplify small conformational changes in the motor domain that result in movement of the motor along actin filaments. Myosin light chains are associated with the neck domain and impose an inhibition on the mechano-chemical cycle. In this regard, the neck domain of the heavy chain acts as the regulatory element that triggers the mechano-chemical events when this inhibition is released by either calcium binding to or phosphorylation of the bound light chains. Crystal structures of the head and neck domains from widely distant myosins (vertebrate myosins II and V, *Dictyostelium* myosins I and II, scallop muscle myosin II, and chicken and fruit fly myosins VI) have revealed highly similar atomic structures among them, indicating that the basic molecular machinery for movement is highly conserved.

The region that follows the neck domain is generally called the ‘tail’ and may, or may not, contain a tandem series of a

linear sequence called the heptad repeat, which tends to form a quaternary structure composed of two intertwining α -helices, referred to as coiled coils. Many myosins have dimerized heavy chains due to these coiled coils. The prime example is skeletal muscle myosin II whose tail domain is entirely composed of a coiled-coil motif, which aligns and joins together with other tail domains to form the bipolar, thick filaments of the sarcomere. Other myosins, however, such as myosin I, are without heptad repeats and thus are single headed. The rest of the molecule is referred to as the ‘globular tail’ or targeting domain and is the most structurally diverse part, varying extensively between classes (Figure 3), yet maintaining significant homology within a class. This is a region of interaction with lipid membranes, adaptor proteins, and signaling molecules, important for site recognition, receptor binding, and other specific functions.

Evolutionary Diversity and Phylogenetic Classification

The advent of whole genome sequencing led to the discovery that the myosin family is remarkably diverse. Comprehensive phylogenetic analyses of the conserved myosin motor domain established that there are at least 37 distinct classes of eukaryotic myosin (a class is defined as a group of myosins from several different taxa having highly similar motor domain sequences and tail structures). The classes are each designated by a Roman (or Arabic) numeral, assigned based on the order in which they were identified. The tail regions are unique to each class and are characterized by the specific domain or combination of domains present in this region (e.g., see Figure 3). Some myosins also have N-terminal extensions (e.g., class III) and a few have extremely short tail regions without any notable domains (e.g., class XIV). A large number of myosins with unique tail domains remain unclassified as they are

found only in a small number of closely related species – these are referred to as ‘orphan myosins’. No single myosin class is found in all organisms and there are a number of myosins that are specific to particular kingdoms (such as the class VIII, XI, and XIII myosins found only in plants), suggesting that these motors evolved early and have become highly specialized over time. Class I, II, and V myosins are found in the Fungi, Amoeba, and Metazoans, indicating that these are the ancestral myosins. Remarkably, Prokaryotes and Archae do not appear to possess a myosin or myosin-like motor and it is tempting to speculate that the emergence of myosin motors was an early evolutionary innovation that provided the impetus for eukaryotic cells to gain access to new environmental niches.

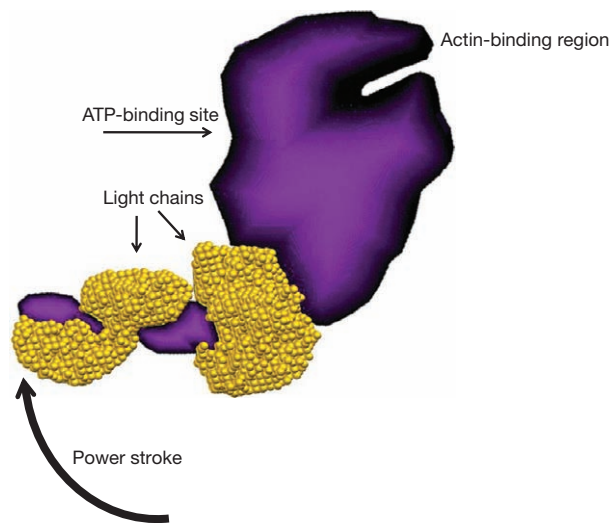


Figure 1 Schematic diagram of the head and neck domains of myosin. The relative positions of the actin-binding region to the ATP site in the head and to the light chains in the neck domain are illustrated. The rotation of the neck domain relative to the head domain, which represents the power stroke, is indicated as a curved arrow. The purple structure represents the contribution of the myosin heavy chain to the conformation of the head and neck domains. The yellow structures represent two light-chain proteins that together with the heavy chain extension form the neck domain. Drawn by Claudia Tavares.

It should be noted that the myosin II class is often referred to as ‘conventional’ myosin due to its historical identification and characterization. All other myosins were then collectively designated as ‘unconventional’. However, myosin II is now viewed as being quite unique in that no other class within the myosin superfamily is known to form bipolar filaments. Clearly, the class II myosins have been refined by evolution, culminating in their role in the generation of macroscopic force and velocity of muscle contraction, whereas all other myosins have globular tail regions and are without the propensity for filament formation.

The early appearance of myosins in evolution is consistent with these motors having fundamentally important cellular functions. As the myosin family expanded, organisms often acquired second or third copies of most myosin genes resulting in the expression of two or more genes encoding closely related, but functionally different myosins of the same class (designated by a letter, e.g., myosin IIA and myosin IIB). Each of these myosins has quite similar motor and tail domains but carries small changes in the motor and tail domain that confer slightly different properties to each myosin. Thus, in addition to the large number of myosin classes found in eukaryotes, there are often multiple forms of each myosin class in organisms with their own specific activities and functions.

Mechano-Chemical Features

A description of the molecular mechanism of energy transduction by mechano-enzymes has been a challenge for workers in the field for half a century. The determination of the atomic structures of myosins in several conformational states with nucleotide derivatives bound, in combination with detailed kinetic studies, has brought this goal close to realization. An overall description of the molecular events taking place during the mechano-chemical cycle, which lead to movement and force generation, is given here (Figure 4). There are three elements to keep in mind: first, ATP binds to the myosin motor domain at an active site capable of hydrolyzing it to adenosine diphosphate (ADP) and inorganic phosphate (Pi); second, conformational changes in the myosin structure occur depending on whether

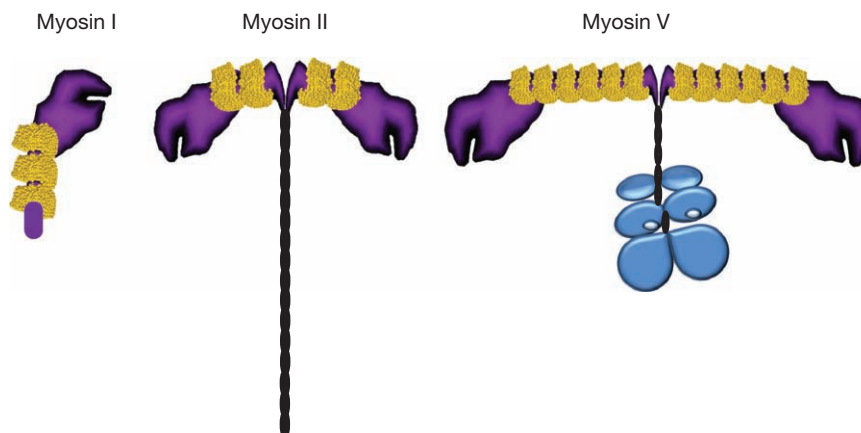


Figure 2 Schematic diagram of three typical myosins representing classes I, II, and V. The purple and yellow structures represent the head and neck domains, respectively, as in Figure 1. The coiled coil is represented by intertwined lines shown in both myosins II and V. The light blue structures represent globular regions in the tail domain of myosin V. Drawn by Claudia Tavares.

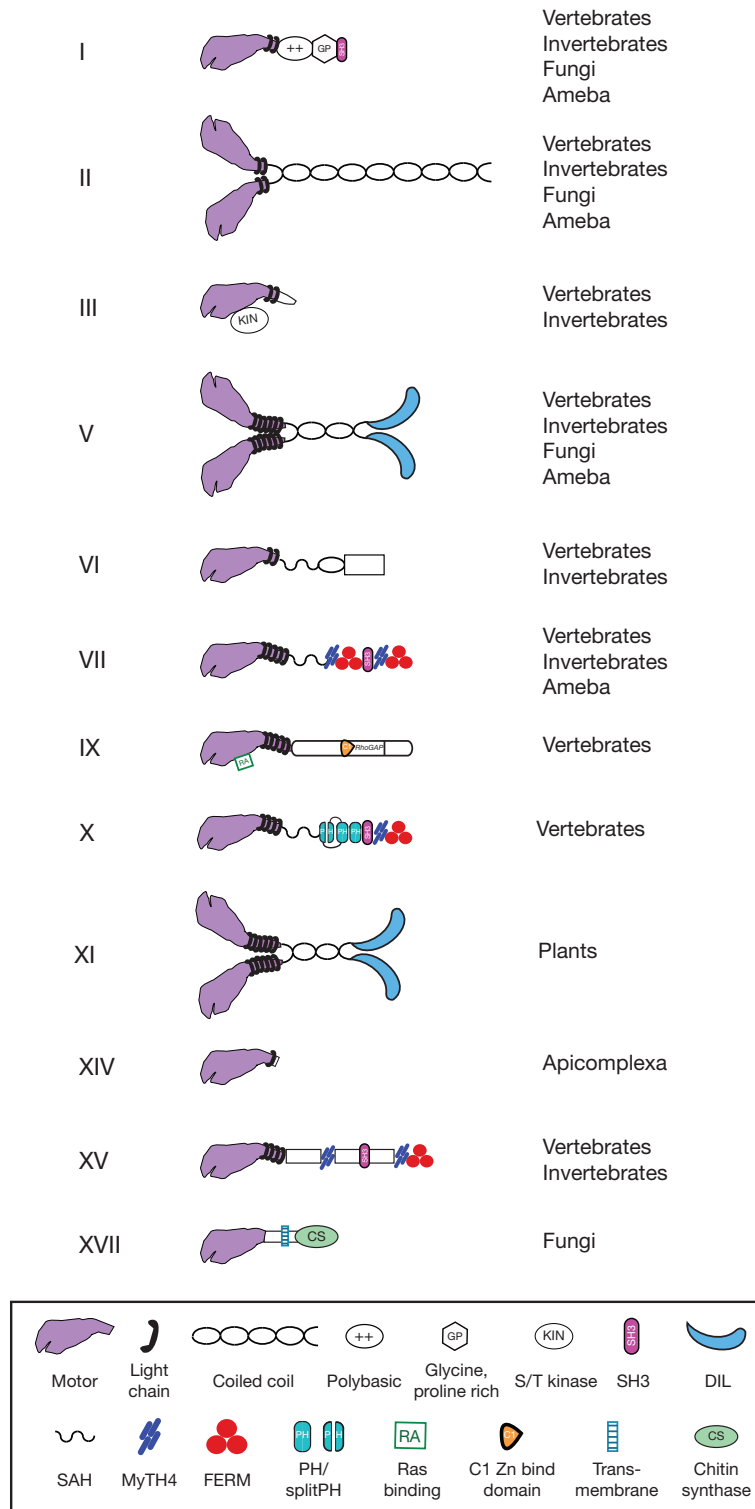


Figure 3 Myosin diversity. A selection of different myosin classes (indicated by the Roman numeral in the left column) and where they are found in nature (column to the right) highlighting the diversity of myosins found in nature. The protein domains present at the N-terminus or in the tail region are indicated and regions unique to each myosin are shown as open rectangles or ovals (not drawn to scale). DIL, dilute domain; FERM, band 4.1, ezrin, radixin, moesin domain; MyTH4, myosin tail homology 4; SAH, single alpha-helix; PH, pleckstrin homology domain; SH3, src-homology 3 domain; S/T kinase, serine/threonine kinase domain.

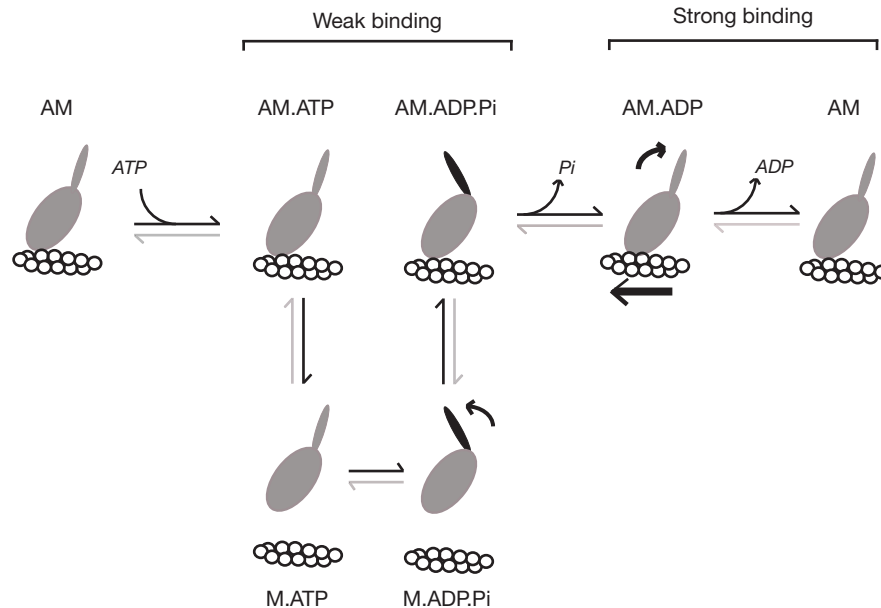


Figure 4 Summary of the kinetic cycle of myosins. The myosin head and neck region are represented by the large and small gray ovals and a segment of a helical actin filament by a series of small circles. The change in the position of the lever arm following ATP hydrolysis is indicated by the black color of the small oval. The biochemical states are indicated above (M, myosin; A, actin; ATP, adenosine triphosphate; ADP, adenosine diphosphate; Pi, inorganic phosphate). While all of the reactions in the cycle are reversible, the favored direction is indicated by the dark line.

ATP or ADP/Pi or no nucleotide is occupying the active site; and, third, myosin binds to filaments of actin (F-actin) either strongly or weakly depending on the molecular conformation of the myosin. Beginning with the binding of ATP to the nucleotide-binding site in the motor domain of myosin, this event puts the myosin into an extended, fully cocked, high-energy conformation, but with low affinity for F-actin, that is, most of the myosins are not bound to F-actin. Myosin, however, quickly hydrolyzes the ATP, which converts myosin into a high-affinity conformation that now strongly binds to F-actin. Triggered by this binding, the reaction products Pi and ADP are sequentially released from the active site accompanied by a major conformational change in the myosin head, which results in the rotation of the lever arm relative to the longitudinal axis of the actin filament. This movement of the lever arm is known as the 'power stroke'. Thus, the conformational change in the myosin molecule, representing a transition from the high-energy to the low-energy state, associated with product release, occurs simultaneously with tight coupling to F-actin and thus converts protein conformational energy into tension development and movement. A return to the weak binding state and disassociation of myosin from F-actin occurs with the rebinding of ATP to the nucleotide site, thus initiating a new cycle. In summary, the mechano-chemical cycle results from the tight coupling of protein conformational energy in the myosin head and neck domains with chemical energy (ATP binding and hydrolysis) and cyclic transitions between strong and weak binding between myosin and F-actin. As presented, this cycle will continue as long as myosin can contact actin and a steady supply of ATP is maintained. (Note that *rigor mortis* results from the interruption of the metabolic supply of ATP that comes with death and represents the strong binding, low-conformational energy state of the actomyosin in muscle at the end of the power stroke.)

Surprisingly, in spite of their highly conserved structure and mechanism of action, the myosin heads differ in a fundamental property related to their mechano-chemical cycle, which affects their cellular function. For example, individual myosin II molecules bind to actin for only a small fraction of the cycle duration, spending a large proportion of their time in the weak binding state. This property is essential for the smooth sliding of myosin filaments in muscle with their arrays of many heads interacting with the organized actin filaments of the sarcomere. On the other hand, at least two myosins (a class V and a class VI myosin) have the property of processivity, that is, they are able to move as individual molecules along actin filaments for some distance without losing contact. Present evidence indicates that these two-headed myosins walk in a hand-over-hand manner along the actin filament, with each head binding to actin for more than 50% of the cycle duration. This property, therefore, allows these myosins, acting alone, to carry cargo without losing a grip on the actin track. All myosins move unidirectionally along polarized actin filaments, with the large majority of myosins moving toward the fast-growing or plus end of actin filaments. However, myosin VI is unique (so far) among all myosins in that it moves in the opposite direction (toward the slow-growing or minus end) along actin filaments. A 50-amino-acid linker between the head and neck domains of myosin VI that reverses the direction of the lever arm seems to be the structural element responsible for this property.

Regulation

The mechano-chemical cycle of myosin must be turned on and off as needed, as is true for all motor proteins. There are several mechanisms to do this, although calcium is the most common

triggering element. For example, in striated muscles a calcium-sensitive system of regulator proteins (tropomyosin and troponins) decorates F-actin and sterically blocks access of the myosin head to the actin sites. Calcium relieves this blockage, such that an increase in its intracellular concentration determines actomyosin activation. For molluscan myosin and vertebrate smooth muscle myosin, regulation is myosin based, in that an inhibitory effect of the neck domain can be relieved by calcium binding to a regulatory light chain or by calcium-activated phosphorylation. Calmodulin has been identified as the light chain for a large number of myosins, and calcium binding to it can trigger the mechano-chemical cycle of myosin and activate motility. However, it should be noted that the role of calcium-calmodulin in myosin regulation is complex and unique to each type of myosin. It has also been demonstrated that conformational folding, that is, the physical approximation of head and tail regions or the doubling over of extended tail structures can control the mechano-chemical cycle, typically inhibiting myosin activity. Again, calcium and/or phosphorylation are also key factors in this regulation.

A Multitude of Cellular Functions

Simplistically, myosins have two functional ends: the motor end, which generates force and movement on actin filaments, and the tail end, which links it to its cargo or to itself to form dimers. Thus, the function of any given myosin is dictated by a combination of the kinetic properties of its motor and the interaction of the tail with its binding partner. The highly diverse nature of the tail domain largely determines the enormous variety of function within the myosin superfamily. For example, the coiled-coil tail of the class II myosins directs both dimerization of the heavy chains and formation of bipolar filaments. It is the tail that specifies the intracellular localization of myosin II filaments where they drive the contraction of skeletal and cardiac muscle or the constriction of the assembled cleavage furrow during cell division. Sorting out the specific functions of individual myosins and myosin classes has been a formidable challenge to researchers in the field. To exemplify the extent of the problem, a single vertebrate cell type – the human intestinal epithelial cell – expresses at least 12 myosins, representing six classes.

Myosins have a broad spectrum of functions, including amoeboid movements, pseudopodia extensions, phagocytosis, exocytosis, cytokinesis, cell gliding over a surface, and other manifestations of plasma membrane dynamics. While the list of known myosin functions continues to grow, myosins appear to have three general roles: intracellular transport, powering large cellular movements, and building or maintaining cellular structures. The most widely studied of myosin functions is organelle transport and many examples of intracellular vesicle and organelle transport that depend on myosin motors have been discovered. The dramatic cytoplasmic streaming in plants is a good example. This process involves class XI myosins, one of only three classes found in plants (the others are class VIII and XIII). The role of class V myosins in the distribution of pigment granules in vertebrate melanocytes and of membrane vesicles in dividing yeast is another example. Several linker proteins have been described that connect myosin motors to

their cargo, such as melanophilin and Rab27a, which link pigment granules to myosin V, and Vac17p, a key component of the myosin receptor on yeast vacuoles. The demonstration in squid axoplasm of translocation of moving vesicles from microtubule tracks to actin filaments suggested that myosins can interact cooperatively with microtubule-based movement for coordinated vesicle transport. These examples serve to illustrate that myosin motors have important roles in the transport of membrane-delimited organelles. In addition, recent evidence shows that myosins also transport macromolecular complexes. Myo4p, a class V myosin in yeast, is directly involved, together with at least two other proteins, in the transport of a specific messenger RNA (mRNA) encoding a transcription factor, to the daughter cell of budding yeast, thus playing an essential role in regulating differential gene expression in this organism.

Myosins contribute to various aspects of cell migration. In mammalian cells, the cytoplasmic class II myosins (IIA and IIB) act in two separate steps of migration – myosin IIB provides contractile force at the rear of the cell and myosin IIA regulates pseudopod extension. The single-headed, class I myosins also contribute to pseudopod extension by binding directly to membranes via the phospholipid-binding regions in their tail domains, thus giving them the ability to orient force generation directly toward membranes. Some of the class I myosins may also contribute to pseudopod extension by recruiting components of the actin polymerization machinery to the leading edge. On the other hand, gliding of Apicomplexa parasites, such as *Toxoplasma gondii*, is powered by the class XIV myosins that are embedded in the space between the plasma membrane and an inner membrane complex. Myosin XIV appears to walk along short, membrane-associated actin filaments, resulting in movement of the membrane against the substrate (similar to a tank tread).

Myosins are required for the formation of specialized actin-filled structures such as filopodia and the maintenance of a related class of structures in the sensory cells of the cochlea called ‘stereocilia’. This may be due to a role for several myosins to localize critical regulators of actin polymerization. For example, filopodia and stereocilia are composed of bundled, parallel actin filaments that grow out from the plasma membrane. Myosin VII and X are required for filopodia formation in several different organisms. Myosin X has been shown to transport vasodilator-stimulated phosphoprotein (VASP), an activator of actin polymerization that binds to the tail of myosin X, to the tips of filopodia and it appears that loss of VASP from the tip abrogates filopodia growth. The class VII and XV myosins play roles in the regulation of stereocilia length. Mice lacking myosin VII have slightly elongated stereocilia and those lacking myosin XV have significantly shorter stereocilia than wild-type mice. The shorter stereocilia found in the myosin XV mutants results from the loss of whirlin, a protein required for full elongation of stereocilia that binds to the tail of myosin XV, from the tips of stereocilia. The ability of these myosins to move along actin filaments toward the membrane allows them to maintain their localization at the tip and aid in the delivery or localization of molecules critical for the growth of the filopodia or stereocilia to the correct place.

One may expect that myosins are end targets for cell signaling. As previously discussed, myosins bind to calmodulin and are indeed activated or modulated by ionic calcium and

protein kinases, all of which are important intracellular messengers for signal transduction pathways. However, it is noteworthy that myosins also seem to participate in signaling pathways themselves, as a kind of motorized signaling device. Evidence for this first appeared upon analysis of tail sequences, where sequence motifs related to cell signaling pathways, such as SH3, PDZ, and PH motifs, were recognized in several distinct myosin classes. These motifs are signals for protein–protein and protein–phosphoinositide binding that participate in the assembly of signaling complexes, although the functional role of myosins containing these signal motifs is still not clear. The tail domain of myosins IX has a region with homology to Rho GTPase, activating proteins (RhoGAPs), which act as molecular switches in the organization of the actin cytoskeleton, particularly in the formation of cell–matrix focal adhesions. Cells lacking myosin IXB fail to spread and polarize, a phenotype that is strikingly similar to cells that have active Rho, suggesting that myosin IX is necessary to turn off Rho signaling. This example illustrates that myosins are not only final targets for cellular signaling but also potential intermediate players in the signaling cascades. Myosins are involved in other cell processes in ways not yet well understood, such as mitotic spindle orientation in yeast, phototransduction in

Drosophila retina, and auditory function of the vertebrate inner ear. Indeed, there is much to be learned about the functional roles of myosin motors throughout the biosphere.

Myosins in Disease

Myosins have fundamentally important cellular roles and it is thus not surprising to find that they underlie the dysfunction observed in a number of human diseases (see [Table 1](#)). The human genome harbors 40 myosin genes, encoding myosins from 12 different classes and mutations in over 10 of these myosin heavy chains have been identified as the cause of hereditary disease. The majority of mutations have been found in the motor domain of a given myosin, although deletions or mutations in tail regions also result in myosin dysfunction and disease. A large number of myosin genes encode class II myosins (12 muscle myosins and three cytoplasmic myosins) and are required for normal muscle function. A major cause of familial hypertrophic cardiomyopathy lies in mutations of the head domain of either of the two cardiac muscle myosin IIs, and mutations in other skeletal myosin genes cause defects in skeletal muscle such as muscle weakness.

Table 1 Human diseases caused by mutations in myosin heavy chains

<i>Gene</i>	<i>Myosin</i>	<i>Disease</i>	<i>Phenotype</i>
Myh2	Adult skeletal myosin IIA	Inclusion body myopathy 3	Dystrophic muscle
Myh3	Embryonic skeletal myosin II	Freeman–Sheldon syndrome (also known as distal arthrogryposis type 2A)	Joint contraction, muscle weakness, fibrosis
Myh6	Cardiac myosin IIA	Cardiomyopathy (familial hypertrophic cardiomyopathy 14, dilated cardiomyopathy 1EE)	Heart failure
		Atrial septal defect 3	
Myh7	Cardiac myosin IIB	Cardiomyopathy (hypertrophic cardiomyopathy 1, dilated cardiomyopathy 1 S)	Premature death due to sudden heart attack
		Myosin storage myopathy	Muscle weakness
Myh8	Perinatal skeletal myosin II	Carney complex	Persistent joint flexing/contracture
Myh9	Cytoplasmic myosin IIA	May–Hegglin anomaly	Macrothrombocytopenias (abnormal platelets)
		Sebastian syndrome	
		Fechtner syndrome	
		Epstein syndrome (MHY9-related diseases)	
		DFNA 17	Sensorineural deafness
Myh11	Smooth muscle myosin II	Thoracic aortic aneurysm/aortic dissection 4	Weakened aortic wall, aneurysm
		Intestinal neoplasia	Colorectal cancer
Myh14	Cytoplasmic myosin IIC	DFNA4	Sensorineural deafness
Myo3A	Myosin IIIA	DFNA30	Progressive hearing loss
Myo5A	Myosin VA	Griscelli syndrome	Pigment dilution, immunologic and nervous system defects
Myo5B	Myosin VB	Microvillar inclusion disease	Diarrhea in infants
Myo6	Myosin VI	DFNA22	Progressive hearing loss
		DFNB37	Congenital hearing loss
		Deafness with hypertrophic cardiomyopathy	Progressive, late-onset deafness accompanied by disease
Myo7A	Myosin VIIA	Usher's syndrome type IB	Deafness, early-onset retinitis pigmentosa
Myo15A	Myosin XVA	DFNB3	Congenital deafness

DFNA, deafness, autosomal dominant; DFNB, deafness, autosomal recessive.

Summary of known human diseases caused by mutations in myosin heavy chains.

Detailed information about alleles and phenotypes can be obtained at OMIM (<http://www.ncbi.nlm.nih.gov/omim/>).

Mutations in several different myosin classes result in deafness, reflecting both the complexity of the sensory cells in the ear and the specialization of myosin function. Usher's syndrome type IB is the most common cause of inherited deaf-blindness in patients and it is caused by mutations in the MyoVIIA gene. Other forms of Usher's syndrome type I result from mutations in proteins that interact or work in concert with MyoVIIA. Mutations in other classes of myosin result in nonsyndromic deafness; these include myosin IIA, IIC, IIIA, and XVA. Mutations in myosin VI were identified in the Snell's Waltzer mouse, which suffers from deafness due to degeneration of the sensory neuroepithelia in the inner ear, and subsequently in human patients. The availability of mouse models for many of these human deafness diseases has provided critical insights into how each myosin contributes to hearing and the identity of their binding partners. Mutations in the two different class V myosins, MyoVA and MyoVB, result in distinct human diseases. Mutations in MyoVA result in Griscelli syndrome, where children exhibit hair hypopigmentation, linked to severe immunological and neurological deficits that frequently lead to early death. As might be expected, mutations in Rab27A, a myosinVA binding partner that links myosinVA to organelles, also cause Griscelli syndrome. In contrast, mutations in MyoVB cause microvillar inclusion disease due to a loss of epithelial cell polarity. Rapid progress in mapping human disease loci in combination with the generation of mouse models suggests that many more connections will be made between human disease and myosin function in the near future.

See also: **Bioenergetics:** Calcium/Calmodulin-Dependent Protein Kinase II; **Cell Architecture and Function:** Actin Assembly/Disassembly; Cytoskeletal Motors: General Principles; **Molecular Biology:** DNA Sequence Recognition by Proteins.

Further Reading

- Berg JS, Powell BC, and Cheney RE (2001) A millennium myosin census. *Molecular Biology of the Cell* 12(4): 780–794.
- Coluccio LM (2008) *Myosins. A Superfamily of Molecular Motors*, p. 475. Dordrecht: Springer.
- Foth BJ, Goedecke MC, and Soldati D (2006) New insights into myosin evolution and classification. *Proceedings of the National Academy of Sciences of the United States of America* 103(10): 3681–3686.
- Lowey S and Trybus KM (2010) Common structural motifs for the regulation of divergent class II myosins. *Journal of Biological Chemistry* 285: 16403–16407.
- O'Connell CB, Tyska MJ, and Mooseker MS (2007) Myosin at work: Motor adaptations for a variety of cellular functions. *Biochimica et Biophysica Acta* 1773(5): 615–630.
- Odrionitz F and Kollmar M (2007) Drawing the tree of eukaryotic life based on the analysis of 2,269 manually annotated myosins from 328 species. *Genome Biology* 8: R196.
- Pollard T and Earnshaw W (2008) *Cell Biology*, 2nd edn., p. 905. Philadelphia, PA: Saunders.
- Richards TA and Cavalier-Smith T (2005) Myosin domain evolution and the primary divergence of eukaryotes. *Nature* 436: 1113–1118.
- Sellers JR (2004) Fifty years of contractility research post sliding filament hypothesis. *Journal of Muscle Research and Cell Motility* 25: 475–482.
- Sellers JR and Knight PJ (2007) Folding and regulation in myosin II and V. *Journal of Muscle Research and Cell Motility* 28: 363–370.
- Sweeney HL and Houdusse A (2010) Structural and functional insights into the myosin motor mechanism. *Annual Review of Biophysics* 39: 539–557.
- Taylor KA (2007) Regulation and recycling of myosin V. *Current Opinion in Cell Biology* 19(1): 67–74.
- Woolner S and Bement WM (2009) Unconventional myosins acting unconventionally. *Trends in Cell Biology* 19(6): 245–251.

Relevant Website

<http://www.cymobase.org> – A database for cytoskeletal and motor protein sequence information.

Natriuretic Peptides, Their Receptors and Therapeutic Applications

K Barbieri and L R Potter, University of Minnesota, Minneapolis, MN, USA

Published by Elsevier Inc.

This article is a revision of the previous edition article by Lincoln R. Potter, volume 3, pp. 1–5, © 2004, Elsevier Inc.

Glossary

Cyclic GMP A cyclic mononucleotide of guanosine that acts as an intracellular second messenger for hormones, such as natriuretic peptides, guanylin, and nitric oxide.

Guanylyl cyclase An enzyme that catalyzes the synthesis of cyclic GMP and pyrophosphate from guanosine triphosphate.

Natriuresis Urinary sodium excretion.

Natriuretic peptides are an ancient family of hormones/paracrine factors that regulate blood pressure, cardiovascular homeostasis, and long bone growth. The mammalian family consists of atrial natriuretic peptide (ANP), B-type natriuretic peptide (BNP), and C-type natriuretic peptide (CNP). All three peptides are derived from separate genes, but share a common 17-amino-acid disulfide ring that is required for biological activity. A family of three cell-surface receptors binds natriuretic peptides. Two are receptor guanylyl cyclases, enzymes that catalyze the synthesis of the intracellular second messenger, cyclic guanosine mononucleotide (cGMP), whereas one is a clearance receptor that controls local natriuretic peptide concentrations via constitutive receptor-mediated endocytosis. In this article, the structure, function, and regulation of each hormone and receptor are discussed.

Natriuretic Peptides

Atrial Natriuretic Peptide

ANP was the first natriuretic peptide to be identified. It is also called atrial natriuretic factor or atriopeptin. Pioneering studies by de Bold and co-workers found that intravenous infusion of atrial, but not ventricular, homogenates into rats caused a rapid and dramatic increase in renal sodium and water excretion that was accompanied by reduced blood pressure and increased hematocrits. The peptide responsible for these activities was purified and sequenced from multiple species by several laboratories. The sequence of polypeptide precursors was derived from complementary DNA sequences from several species.

All natriuretic peptides are synthesized as preprohormones. Human preproANP is 151 amino acids in length. Proteolytic removal of the amino terminal signal sequence yields a 126-residue proANP peptide, which is the predominant form

found concentrated in dense atrial granules. Upon secretion, proANP is cleaved by corin, a transmembrane serine protease, to a nonfunctional 98-residue amino-terminal peptide and a 28-residue carboxyl-terminal bioactive peptide. Both fragments circulate in the plasma and are elevated under conditions where vascular volume is increased, such as congestive heart failure. The carboxyl-terminal peptide is the mature form of ANP that mediates the known biological effects associated with the ANP gene ([Figure 1](#)); this peptide contains 28 residues, with a six-residue amino tail and a five-residue carboxyl tail. The majority of ANP gene expressions occur in the cardiac atria, but low-level expressions take place in the kidney. Atrial-wall stretch, reflecting increased intravascular volume, is the dominant stimulus for ANP release. However, several hormones and neurotransmitters such as endothelin, arginine vasopressin, and catecholamines, also stimulate ANP secretion.

In addition to stimulating natriuresis, diuresis, and vasorelaxation, ANP also inhibits renin, vasopressin, and aldosterone release, as well as cell proliferation and hypertrophy. Hence, ANP's short-term (natriuresis, diuresis, vasorelaxation) and long-term (antihypertrophy) effects are generally considered beneficial.

Transgenic mice exhibiting lifelong elevated plasma ANP levels display reduced blood pressure and heart size. In contrast, mice lacking ANP are hypertensive. The initial report describing these animals suggested that ANP regulates blood pressure in a salt-sensitive manner, but a subsequent communication indicated that these same mice were also hypertensive when fed a low-salt diet, consistent with ANP mediating its effects independently of body sodium levels.

ANP is degraded by extracellular proteases and through natriuretic peptide clearance receptor (NPR-C)-dependent internalization and degradation. Neutral endopeptidase (NEP) and insulin-degrading enzyme have been shown to proteolyze ANP.

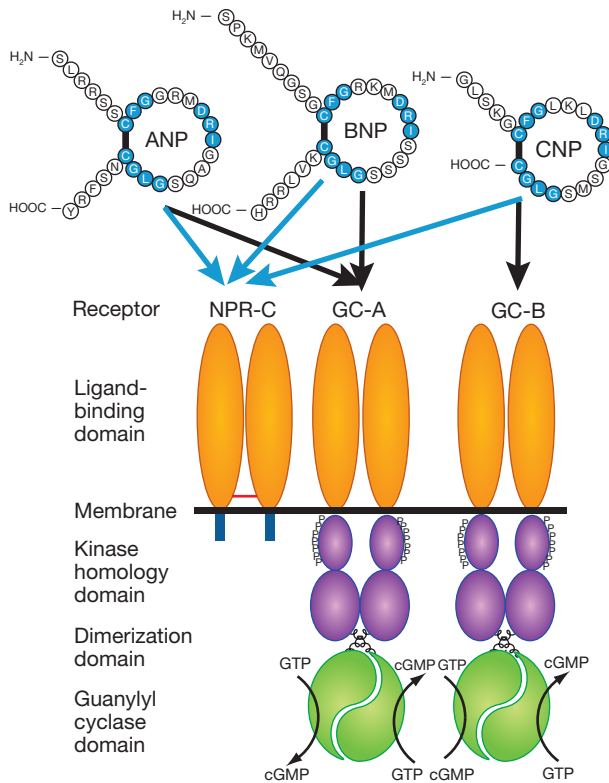


Figure 1 Schematic representation of natriuretic peptides and their receptors. Black arrows indicate peptide binding to signaling receptors. The blue arrows indicate that they bind the NPR-C, which leads to degradation. Blue circles in natriuretic peptide structures indicate conserved residues. In the kinase homology domain, the black P's indicate phosphorylation sites. Abbreviations: ANP, atrial natriuretic peptide; BNP, brain natriuretic peptide; CNP, C-type natriuretic peptide; cGMP, cyclic guanosine monophosphate; GTP, guanosine triphosphate; GC-A, natriuretic peptide receptor A; GC-B, natriuretic peptide receptor B; NPR-C, natriuretic peptide receptor C.

B-Type Natriuretic Peptide

B-type natriuretic peptide was first purified from porcine brain extracts and, therefore, named brain natriuretic peptide. However, subsequent experiments found that the peptide exists in higher concentrations in stressed cardiac tissues and, for this reason, is most often called BNP.

BNP exhibits many of the same characteristics as ANP and is also encoded on chromosome 1. The resulting prohormone contains 134 residues but only the 108-residue prohormone is found in cells because the signal sequence is cleaved during translation. The protease responsible for processing of proBNP to BNP is not definitively known but corin is a reasonable candidate. The bioactive form of human BNP contains 32 residues, with a nine-residue amino tail and a six-residue carboxyl tail.

BNP is expressed along with ANP in the atrial granules, but higher levels of expression occur in cardiac ventricles during periods of cardiovascular stress such as heart failure. In the ventricles, BNP is transcriptionally regulated and constitutively released once it is synthesized. It is not stored in ventricular granules.

Plasma concentrations of BNP in normotensive patients are much lower than those of ANP (about 1 fmol ml^{-1} vs. 10 fmol ml^{-1} , respectively). However, BNP production is elevated to a greater degree in patients with congestive heart failure, and, in extreme cases, plasma concentrations can increase more than 300-fold. This makes BNP a better serum marker of heart failure than ANP.

BNP is degraded by NPR-C-mediated internalization and also by cell-surface proteases. However, BNP is not a good substrate for NEP. Instead, a protease that is inhibited by leupeptin degrades BNP. This reduced degradation by NEP may contribute to the 20-min as opposed to 2-min half-lives for BNP and ANP, respectively.

C-Type Natriuretic Peptide

CNP was initially purified from porcine brain extracts and is the most highly expressed natriuretic peptide in this tissue. Human preproCNP is 126 residues in length. The 103-amino-acid CNP propeptide is processed to either a 53- or a 22-amino-acid form. The bioactive forms of CNP are structurally similar to ANP and BNP, but they lack a carboxyl tail (Figure 1). Furin has been suggested to mediate this final cleavage step. Both the 53- and 22-amino-acid forms have been identified *in vivo*. The 22-amino-acid peptide predominates in fluids such as blood, and the 53-amino-acid version is found mainly in tissues. The bioactivity of the two CNP peptides is similar. CNP relaxes vascular smooth muscle and is a more potent dilator of veins than ANP or BNP. CNP is not natriuretic despite its name.

Homozygous deletion of the gene encoding CNP in mice caused more than half of the offspring to die during postnatal development. Surviving animals exhibited shortened body length and defects in endochondral bone development, which is analogous to human dwarfism. cGMP-dependent protein kinase G II (PKGII) knockout mice also develop dwarfism as a result of impaired endochondral ossification, suggesting that PKGII mediates the CNP-dependent skeletal effects. Consistent with a linear CNP/cGMP/PKGII pathway, the targeted expression of CNP in growth plate chondrocytes cannot increase the longitudinal bone growth or chondrocytic proliferation and hypertrophy in PKGII knockout mice. Chromosomal translocations that increase CNP concentrations in humans cause marfanoid-like skeletal overgrowth.

CNP is the only natriuretic peptide that does not have a C-terminal tail. It does, however, have a 36-residue amino-terminal tail in CNP-55 and a five-residue amino-terminal tail in CNP-22. CNP is degraded by NPR-C-mediated endocytosis and by NEP and insulin-degrading enzyme.

Natriuretic Peptide Receptors

There are three known natriuretic peptide receptors. Guanylyl cyclase A (GC-A) and guanylyl cyclase B (GC-B) are transmembrane guanylyl cyclases, whereas the NPR-C does not possess any known enzymatic activity. Instead, NPR-C controls the local concentration of natriuretic peptides that are available

to bind GC-A and GC-B through receptor-mediated internalization and degradation.

Guanylyl-Cyclase-Linked Natriuretic Peptide Receptors

ANP and BNP elicit their effects through GC-A. CNP activates GC-B. Both guanylyl-cyclase-linked receptors have five homologous domains: the extracellular ligand-binding domain, a membrane-spanning region, a kinase homology domain (KHD), a hinge dimerization region, and C-terminal guanylyl cyclase catalytic domain (Figure 1).

The ligand selectivity for GC-A is $\text{ANP} \geq \text{BNP} \gg \text{CNP}$, whereas the ligand selectivity for GC-B is $\text{CNP} \gg \text{ANP} = \text{BNP}$ (Figure 1). Mice lacking GC-A are hypertensive and develop cardiac hypertrophy and ventricular fibrosis. In addition, they are completely unresponsive to the renal and vasorelaxing effects of ANP and BNP, consistent with GC-A being the sole signaling receptor for these peptides in mice. GC-B mediates the endochondrial ossification effects of CNP. Homozygous inactivating mutations in GC-B cause acromesomelic dysplasia-type maroteaux dwarfism in humans.

Both GC-A and GC-B consist of an approximately 450-amino-acid extracellular ligand-binding domain, a 21-residue hydrophobic membrane-spanning region, and a 566- or 568-amino-acid intracellular domain, respectively (Figure 1). The latter can be divided into a ~250-amino-acid KHD, a 41-amino-acid hinge-dimerization region, and a 250-amino-acid C-terminal guanylyl cyclase catalytic domain.

The GC-A extracellular domain contains three intramolecular but no intermolecular disulfide bonds. Similar cysteine residues are conserved in the extracellular domain of GC-B. Both GC-A and GC-B are highly glycosylated, and several studies suggest that glycosylation is required for proper receptor folding and/or transport of the receptor to the cell surface, but is not required for hormone binding.

Activation of the guanylyl-cyclase-linked natriuretic peptide receptors is incompletely understood. Both receptors are homooligomers in the absence and presence of their respective ligands, indicating that receptor activation does not simply result from ligand-dependent dimerization. However, ANP binding does cause a conformational change in GC-A which brings the extracellular juxtamembrane regions of each monomer closer together. In addition to natriuretic peptides, adenosine triphosphate (ATP) is also required to maximally activate the guanylyl-cyclase-linked receptors in broken cell preparations. In the resting state, both GC-A and GC-B are highly phosphorylated on multiple serines and threonines within their KHD domains (Figure 1). Mutation of any of these phosphorylated residues to alanine in order to mimic a dephosphorylated serine or threonine results in decreased ANP- or CNP-dependent guanylyl cyclase activity. Furthermore, receptors containing alanine substitutions at four or more phosphorylation sites are completely unresponsive to hormone. Both GC-A and GC-B are desensitized by dephosphorylation.

The Natriuretic Peptide Clearance Receptor

In addition to GC-A or GC-B, all three natriuretic peptides bind NPR-C, which is found as a disulfide-linked homodimer

in cells. NPR-C has a large amino-terminal extracellular ligand-binding domain, a single transmembrane domain, and an intracellular domain that contains only 37 amino acids (Figure 1). It not only binds ANP, BNP, and CNP with similar affinities, but can also bind tightly to several smaller carboxyl-terminal and ring-deleted peptide analogs that do not bind well to GC-A or GC-B. C-ANF (atrial natriuretic factor), the prototypical form of these peptides, is sometimes used to distinguish NPR-C from GC-A in both binding and cGMP stimulation assays.

NPR-C is the most abundant natriuretic peptide receptor on the surface of most cells, and its primary function is to remove natriuretic peptides from circulation via receptor-mediated internalization and degradation, thereby controlling the local concentrations of the natriuretic peptides that are available to bind GC-A and GC-B. Like many nutrient receptors, NPR-C internalization is constitutive, that is, it is not increased by natriuretic peptide binding. In addition to its clearance role, NPR-C has also been implicated in signaling, possibly in a G-protein-dependent manner. However, mice lacking NPR-C displayed a reduced ability to clear ^{125}I -ANP from the circulation, but no reduction in known signaling functions. In fact, pathways activated by GC-A and GC-B were increased in these animals.

Natriuretic Peptide Therapies

The cardiac natriuretic peptides, ANP and BNP, trigger natriuresis and diuresis, which decreases vascular volume and allows the circulatory system to compensate for the failing heart. In addition, they also exert local inhibitory effects on the heart by preventing cardiac hypertrophy and fibrosis. Because of these natriuretic and diuretic effects, synthetic forms of ANP and BNP have been approved for the treatment of acute decompensated heart failure.

Synthetic ANP, known as Carperitide, was approved for use in patients in Japan in 1995. Human BNP, synthesized recombinantly in *Escherichia coli*, is called Nesiritide and was shown to be a better renal activator than Carperitide in heart-failed dogs. Based on positive results from the Vasodilation in the Management of Acute Congestive Heart Failure (VMAC) trial, Nesiritide was approved for the treatment of acute decompensated heart failure in the United States in 2001. Subsequently, the safety of Nesiritide was questioned. Complicated statistical manipulation of clinical trial data from multiple studies, called meta-analysis, led researchers to suggest that Nesiritide worsened patients' renal function and increased the likelihood of death. However, the Follow-Up Serial Infusion of Nesiritide (FUSION) trials, specifically designed to investigate the potential adverse renal effects of Nesiritide in patients with advanced heart failure, found no evidence of decreased renal function. Unfortunately, they also found little clinical benefit of the drug.

One possibility for the negative results is that congestive heart-failure patients may have developed resistance to natriuretic peptides. Consistent with this idea, mice with heart failure were shown to have reduced levels of GC-A but not GC-B. New natriuretic peptide-based therapies are focusing on chimeric natriuretic peptides that activate both receptors.

See also: **Bioenergetics:** Calcium/Calmodulin-Dependent Protein Kinase II; **Protein/Enzyme Structure Function and Degradation:** Aminopeptidases; Metalloproteases; **Signaling:** Protein Kinase C Family; Vasopressin/Oxytocin Receptor Family.

Further Reading

Potter LR (2011) Guanylyl cyclase structure, function and regulation. *Cellular Signalling* 23: 1921–1926.

Potter LR (2011) Natriuretic peptide metabolism, clearance and degradation. *FEBS Journal* 278: 1808–1817.

Potter LR (2011) Regulation and therapeutic targeting of peptide-activated receptorguanylyl cyclases. *Pharmacology and Therapeutics* 130: 71–82.

Potter LR, Abbey-Hosch S, and Dickey DM (2006) Natriuretic peptides, their receptors, and cyclic guanosine monophosphate-dependent signaling functions. *Endocrine Reviews* 27(1): 47–72.

Neoglycoproteins

Y C Lee and R T Lee, Johns Hopkins University, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Glycoside clustering effect When glycosides are bundled together at proper distances, the total effect can be much greater ($>10^6$ -fold) than the simple sum of individual glycoside.

Neoglycoconjugates The original term of neoglycoproteins has now been expanded, and the term neoglycoconjugates has been introduced to encompass all the glyco-modified materials.

Preparation

Direct Coupling of Oligosaccharides to Proteins

The easiest method is to attach naturally available oligosaccharides to the amino groups of proteins. This process does not require any chemical manipulations other than the conjugation process by reductive amination. Some of the readily available di- and oligosaccharides include: lactose, maltose, cellobiose, melibiose, *N*-acetyl-lactosamine, di-*N*-acetyl-chitobiose, and a number of milk oligosaccharides. An example of coupling of lactose by reductive amination to protein using sodium cyanoborohydride or pyridine borane as reducing agent, which preferentially reduces aldimine (Schiff's base) over aldehyde, is shown in [Figure 1](#). This method relies on the fact that reducing sugars in acyclic aldehyde form, although present in a very minor proportion, are in equilibrium with the cyclic forms. As the aldehyde form of oligosaccharide is consumed, more will become available by conversion from the cyclic forms. However, the conjugation by this simple method is a very slow process, requiring several days or longer, because of the low concentration of the aldehyde form of oligosaccharide in solution.

The reductive amination converts the primary amino groups of N-terminus and lysine side chains to secondary amines initially, and eventually to tertiary amines (with two sugar groups attached to a single nitrogen), if both oligosaccharide and reducing agent are present in excess. It is important to note that after such a reaction, the reducing terminal sugar of the oligosaccharide will become acyclic, and for this reason, conjugation with monosaccharide is meaningless. The shortcoming of this approach, in addition to long reaction time mentioned above, is the limited availability of suitable oligosaccharides. However, the number of commercially available oligosaccharides that are naturally derived or synthetically prepared has been steadily increasing.

Using Synthetic Carbohydrate Derivatives

Neoglycoproteins prepared by direct conjugation of oligosaccharides sometimes can be disadvantageous, due to sugar residues that are recognition target being too close to the protein surface. Synthetic carbohydrate derivatives would expand the scope of neoglycoprotein considerably both in terms of sugar

structure and its disposition on the protein surface. The procedures can be totally synthetic or can be chemical modifications of naturally available glycans either in a free sugar form or linked to a group such as peptide. A few of the popular procedures are presented below.

Aromatic glycosides

Commercially available *p*-aminophenyl glycosides are convenient derivatives for making neoglycoproteins either by diazo-coupling to tyrosine side chains or via transformation into isothiocyanate (reaction with thiophosgen), which then reacts with amino groups. *p*-Nitrophenyl glycosides, commonly used chromogenic glycosides, are available in even wider variety, and can be easily reduced to the corresponding *p*-aminophenyl glycosides. A shortcoming of these derivatives is that they can increase the hydrophobicity of product neoglycoprotein considerably and consequently may alter the nature of protein and give anomalous results in the probing of carbohydrate function.

Nonaromatic glycosides

The problem of excessive hydrophobicity can be alleviated by the use of derivatives which do not contain aromatic rings and are not as hydrophobic as aromatic glycosides (e.g., [Figure 2](#)). Some of these derivatives are thioglycosides, which have the advantage of being more resistant to most exoglycosidases and also specifically cleavable with mercuric salts. Examples shown in [Figure 2](#) are: (a) sugar imidate reacting with amino groups of protein to produce amidino-type neoglycoproteins; (b) an ω -aldehyde glycoside being coupled to protein via reductive amination; and (c) an ω -aminoalkyl glycoside converted into a squaric acid derivative which can react with amino groups of protein. In both (b) and (c), the length of a glycon can be varied.

Reduced and periodate-oxidized natural oligosaccharides

Synthesis of oligosaccharides having complex structure is still a tedious and difficult task. Therefore, it is prudent sometimes to utilize naturally isolable oligosaccharides, if the quantity of neoglycoprotein needed is not too large. Natural oligosaccharides, pre-existing or obtained from natural glycoconjugates by enzymatic cleavage, can be made more reactive by first reducing them with sodium borohydride (to convert the reducing

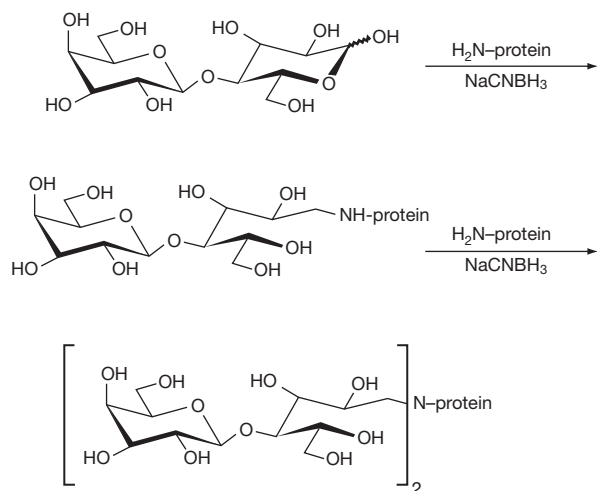


Figure 1 Conjugation of lactose via reductive amination.

sugar residue into acyclic alditol) followed by mild periodate oxidation (e.g., 10 mM NaIO₄ for 10 min at room temperature), which will preferentially oxidize acyclic glycols to generate aldehyde group(s) without affecting the rest of the cyclic sugar residues. The newly generated aldehyde can be used in the reductive amination reaction as described above.

Using Other Functional Groups of Proteins

Most of the conjugation methods described above utilize amino groups of proteins for conjugation, which are often abundant and reactive because of the distance from the protein backbone. However, other functional groups such as phenolic, sulfhydryl, and carboxyl groups can also be used for conjugation. Diazo coupling of aryl glycosides to tyrosine side chain has been mentioned earlier. Thiol groups of protein are easy target for oligosaccharide attachment, since chemically mild, thiol-specific reactions are readily available. For example, a heterobifunctional linker with an amino-reacting group on

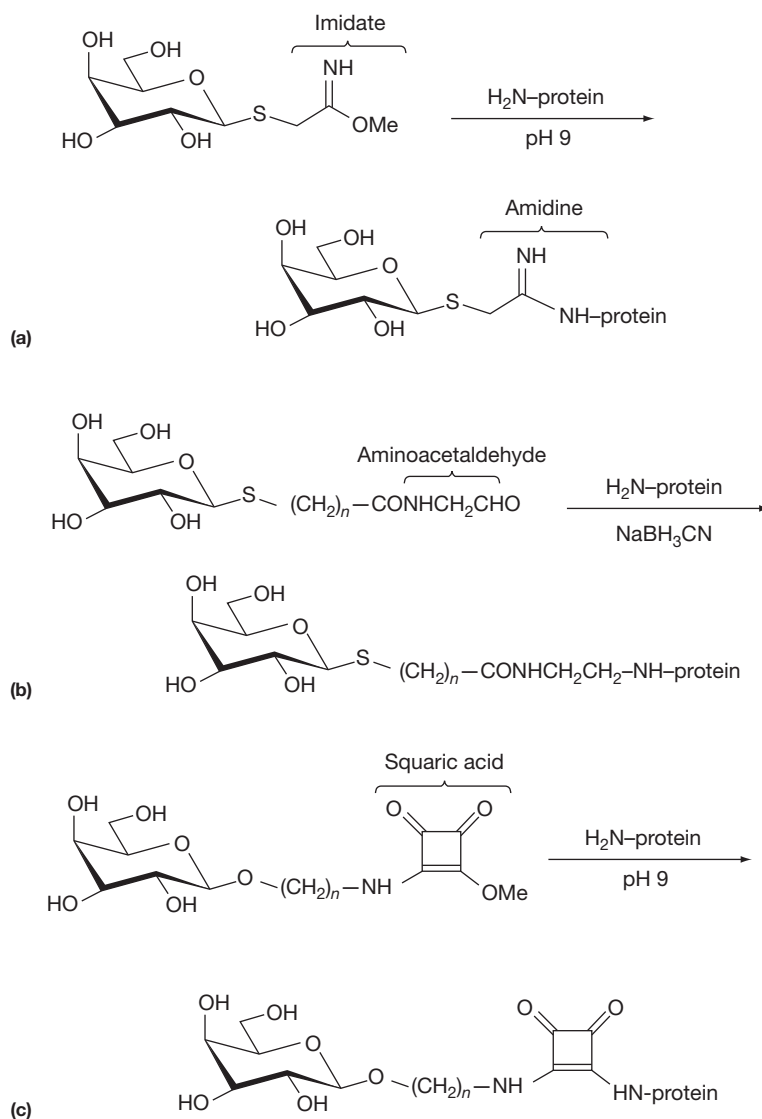


Figure 2 (a–c) Example of synthetic methods for neoglycoprotein preparations.

one end and a maleimido or iodoacetamido group on the other end is reacted first to an Asn-oligosaccharide (or other glycopeptide) via its amino terminus, which is then reacted with protein thiol groups to produce neoglycoprotein in high yield. Bovine serum albumin (BSA) has a single sulfhydryl group which can be utilized in this way to produce a neoglycoprotein carrying a single oligosaccharide chain of defined structure. An interesting approach is to react glycosylthiol with protein thiol group to form a disulfide bond. This will result in a product quite similar in structure to the natural *N*-glycosylated protein in terms of linkage length.

Use of Glycosyl Transferases

As the number of glycosyl transferases available commercially increases steadily, these enzymes have become a powerful synthetic tool in neoglycoprotein synthesis. The advantage of glycosyl transferases is their exquisite specificity for anomeric configuration as well as the position of attachment. Thus, these enzymes are especially suited for modifying a pre-existing glycan structure (even a monosaccharide) at a specific position. Attachment of terminal α -L-fucosyl group or sialic acid group is a prime example. The enzymatic reaction can be carried out on the oligosaccharide before conjugation or after the conjugation to proteins.

Conjugation of Glycopeptides

For the conjugation of glycan structures from glycoprotein, sometimes it is easier to generate glycopeptide than to generate reducing oligosaccharide. A heterobifunctional crosslinker with a hydrazide and an acetal group at two ends of molecule was devised to react first with amino group of glycopeptide (or any other amino-terminated carbohydrate derivatives) via azide reaction, which is followed by the generation of aldehyde function and then coupling to protein via reductive amination.

In all of these reactions, the level of sugars on neoglycoproteins can be controlled either by the ratio of reagent to protein or by the length of reaction. It should be borne in mind that these modifications generally produce a distribution of different sugar substitution in terms of both position and number. However, the ability to vary the degree of substitution turned out to be a great advantage for probing of carbohydrate function in many biological systems that is not readily available in natural glycoproteins. This is because recognition of carbohydrates often requires multivalency (glycoside clustering effect) in order to manifest its function unequivocally. Neoglycoproteins can readily provide multivalency and thus are powerful ligands for carbohydrate binding.

In addition, neoglycoproteins can be radioiodinated (using the tyrosine side chain) or modified with fluorescent groups, such as fluorescein (using available ϵ -amino group) for easy monitoring.

Applications

Carbohydrate Receptor Analyses

Using neoglycoproteins, one can easily survey sugar-binding receptors (or lectins) in a biological specimen. For example, a series of neoglycoproteins have been used to determine that

alligator hepatic receptor has binding specificity for mannose and L-fucose. The binding of tagged neoglycoproteins (e.g., ^{125}I -labeled) to receptors can be probed both in solution phase and solid phase. Properly labeled neoglycoproteins can be used as cytochemical markers.

Affinant for Isolation

Neoglycoproteins can serve as effective affinants for the isolation of lectins and receptors. For such a purpose, a neoglycoprotein can be conveniently immobilized to a suitable solid phase (e.g., sepharose). In an alternate case, a protein (such as BSA) is immobilized first and then modified with one of the reagents mentioned above to form neoglycoproteins on solid phase.

Substrates for Enzymes

Neoglycoproteins can be used as convenient substrate for enzymes. For example, GlcNAc-BSA adsorbed on microtiter wells was used conveniently as substrate to assay β -galactosyl transferase with great sensitivity, which in turn can be used to assay α -fucosyl transferase.

Targeting Device

Neoglycoproteins can be used to target drugs and other materials to specific organs. For example, Gal- or GalNAc-containing neoglycoproteins, when properly conjugated to drug or anti-sense materials, can be used to target such materials to mammalian liver.

Examination of Glycoprotein Structure/Function Relationship

Carbohydrate chains of glycoprotein can have biological and biochemical function unrelated to specific sugar-lectin interaction. One of the better documented cases is the glycan in the Fc domain of IgG, which is required for interaction with Fc receptor. Neoglycoproteins were prepared by disulfide formation between Cys (in lieu of Asn) in the Fc domain and glycosylthiols of various structures to examine the effect of glycan structure on Fc receptor binding activity. Moreover, if the amino acid sequence and X-ray structure are known, one can place a Cys at different surface locations by genetic engineering, which would provide more latitude in examination of the role of carbohydrates on glycoproteins.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Glycosylation, Congenital Disorders of; Glycoprotein-Mediated Cell Interactions, Nonmucin-Type O-Linked; Glycoprotein Folding and Processing Reactions; Glycoproteins, N-Linked; **Molecular Biology:** DNA Mismatch Repair and the DNA Damage Response; DNA Ligases: Mechanism and Functions; DNA Ligases: Structures; Nonhomologous Recombination: Retrotransposons; Nucleotide Excision Repair, Bacterial: The UvrABC System; **Signaling:** Protein Kinase C Family; Immunoglobulin (Fc) Receptors.

Further Reading

Lee YC and Lee RT (1994) *Methods in Enzymology, Vol. 242. Neoglycoconjugates: Part B. Biomedical Applications*. San Diego, CA: Academic Press.

- Lee YC and Lee RT (1994) *Methods in Enzymology, Vol. 247. Neoglycoconjugates: Part A. Synthesis*. San Diego, CA: Academic Press.
- Lee YC and Lee RT (1994) *Neoglycoconjugates: Preparation and Applications*. San Diego, CA: Academic Press.
- Lee RT and Lee YC (1997) Neoglycoproteins. In: Montreuil J, Vliegenthart JFG, and Schachter H (eds.) *Glycoproteins II*, pp. 601–620. Amsterdam: Elsevier.
- Macindoe WM, Van Oijen AH, and Boons GJ (1998) A unique and highly facile method for synthesizing disulfide linked neoglycoconjugates: A new approach for remodelling of peptides and proteins. *Chemical Communications* 7: 847–848.
- Monsigny M, Mayer R, and Roche AC (2000) Sugar–lectin interactions: Sugar clusters, lectin multivalency and avidity. *Carbohydrate Letters* 4: 35–52.
- Monsigny M, Midoux P, Mayer R, and Roche AC (1999) Glycotargeting: Influence of the sugar moiety on both the uptake and the intracellular trafficking of nucleic acid carried by glycosylated polymers. *Bioscience Reports* 19: 125–132.

Neuronal Intermediate Filaments

R K H Liem, Columbia University College of Physicians and Surgeons, New York, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Axonal spheroids Accumulations of protein aggregates in axons, usually containing neurofilamentous structures.

Axonal transport The movements of vesicles, mitochondria, and other cytoplasmic structures down the axon, necessitated by the fact that most protein synthesis occurs in the axon and the proteins must sometimes travel long distances to reach the end of the axon. This transport generally occurs by way of motor proteins along microtubules and microfilaments.

Charcot-Marie-Tooth (CMT) disease An inherited neurological disorder that affects the peripheral nerves.

It is named after the three physicians who described it. CMT patients slowly lose the normal use of their limbs as the nerves to the extremities degenerate. There are a large number of different causes for the disease, including the recently identified neurofilament mutations.

Cytoskeleton A network of interconnected cytoplasmic filaments, which consists of microtubules, microfilaments, and IFs.

Transgenic mouse A mouse that has been genetically altered to express a gene that is not normally expressed or to have a gene deleted (also called a knockout mouse).

Introduction

The cytoskeleton of most eukaryotic cells is composed of microtubules, microfilaments, and intermediate filaments (IFs). Microtubules are 25 nm in diameter and are composed of two subunit proteins called α - and β -tubulin, as well as a number of associated proteins, microfilaments are 6–8 nm in diameter and made up of a protein called actin. IFs are intermediate in size between the microtubules and microfilaments. All IF proteins contain a central α -helical rod domain flanked by nonhelical N-terminal head and C-terminal tail domains. The rod domain contains hydrophobic repeats (called heptad repeats) that mediate dimerization by allowing two α -helices to wrap around each other forming a 'coiled coil'. Subsequent polymerization into filaments involves the formation of tetramers and higher-order structures for which the N-terminal head domain is particularly important. Based on their protein sequence similarities and their gene structure, the IF protein family is separated into six or seven types. In the nervous system, IFs are expressed in both neurons and glial cells. In the central nervous system (CNS), astroglial cells express the glial fibrillary acidic protein (GFAP) and sometimes vimentin, whereas oligodendrocytes appear to be one of the few cell types that express no cytoplasmic IFs. In the peripheral nervous system (PNS), Schwann cells express the more ubiquitous IF protein vimentin. Vimentin and GFAP are quite similar in sequence and gene structure and are part of a subfamily of IFs called type III IFs.

Neuronal IFs are the major and most visible filamentous structures in the axons. The number of neuronal IFs is correlated with axon diameter and they are believed to be primary determinants of axon diameter. The neurofilament triplet proteins (NFTPs) are the most ubiquitous IF proteins in neurons; they are present in both central and peripheral nerves. The NFTPs and α -internexin are more closely related to each other than to other IF proteins and are classified together as a

subfamily called type IV IFs. The expression of peripherin predominates in the PNS. Although peripherin is expressed exclusively in neurons, based on its protein sequence homologies and its gene structure, peripherin is more similar to vimentin and GFAP and is, therefore, considered a type III IF. Nestin is the IF protein found in neuroepithelial stem cells, as well as muscle progenitor cells. The protein sequence of nestin is quite unique and nestin is sometimes considered to be a separate type VI IF. However, on the basis of its gene structure, it resembles the other type IV neuronal IFs.

One of the hallmarks of a number of neurodegenerative diseases is the accumulation of neuronal IFs in the form of axonal spheroids. These axonal spheroids have been observed in amyotrophic lateral sclerosis (ALS, also called Lou Gehrig's disease) as well as other neuropathies. A number of recent studies from transgenic mice have indicated that the disorganization of neuronal IFs, especially in the forms of axonal spheroids, can lead to neurodegeneration. The relationship between the various neuronal IF proteins and neurodegeneration is discussed later.

The Neurofilament Triplet Proteins

The NFTPs are composed of three proteins called NFL, NFM, and NFH for neurofilament light, medium, and heavy chain, respectively. NFTPs have been identified in all mammals and birds. The identification of the NFTPs was first done on the basis of studies on the transport of proteins in the axon. To study the transport of particular proteins, early studies used radioactive amino acids to label the newly synthesized proteins, and the nerves were then analyzed to determine the rates at which different proteins move. The slowest moving fraction moved at 0.1–1-mm per day and was believed to be the structural components. There were five major proteins, α - and β -tubulin and three proteins with apparent molecular

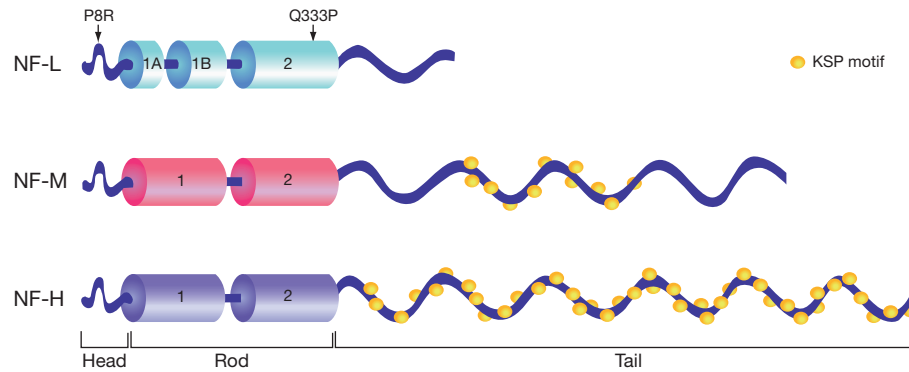


Figure 1 Diagrammatic representation of the NTFs. Note that the size of the proteins is dependent on the lengths of the C-terminal tails. The α -helical rod regions are shown as cylinders. The KSP phosphorylation sites in the tail regions are shown as yellow circles. The first two mutations linked to CMT disease in NFL are indicated.

weights of 68, 145, and 200 kDa. As actin, the component of the microfilaments, moved somewhat faster than these five proteins, the three unidentified proteins were identified as the subunits of the other structural element, the neurofilaments. Biochemical verification came when neurofilaments were purified and isolated and the contaminating astroglial filaments were removed. Each subunit was purified and the ability of the purified proteins to assemble *in vitro* was determined and assayed by examining the assembled filaments under the electron microscope. These *in vitro* assembly studies showed that the 68 kDa protein, later called NFL, was necessary and sufficient for IF formation. The 145 kDa (NFM) and 200 kDa (NFH) proteins coassembled with NFL and their long C-terminal tails appeared to form the links that can be observed between filaments in the electron microscope. The actual sequences of each of these proteins were determined initially by protein sequencing and later by complementary DNA (cDNA) cloning. They are shown schematically in **Figure 1**. Like all other IFs they contain a central α -helical rod domain flanked by nonhelical N-terminal head and C-terminal tail. The rod domain contains α -helical segments that are represented as cylinders. The rod domain of NFL can be divided into three separate segments (1A, 1B, and 2), whereas those of NFM and NFH are separated into two segments. The NF proteins have variable size C-terminal tails. The lengths of the C-terminal tails determine the ultimate sizes of the three proteins.

The tails of NFM and NFH contain multiple lysine-serine protein (KSP) repeats (shown as yellow circles), which are sites for phosphorylation as described below. As mentioned above, NFL is necessary for filament formation. Using various assays to study protein-protein interactions, it has been demonstrated that NFL can form homodimers and heterodimers with NFM and NFH. These results are consistent with the earlier observations from *in vitro* assembly studies that NFL can form a filament, whereas the other two proteins require NFL to assemble into filaments. Interestingly, transfection studies have shown that rat or mouse NFL needed either NFM or NFH to form a filamentous network *in vivo*, indicating that rodent neurofilaments are obligate heteropolymers. Human NFL can self-assemble in the transfected cells, but this assembly is not very efficient. The other neuronal IF

proteins, peripherin, and α -internexin are able to self-assemble into homopolymers *in vitro* and *in vivo*, although NFM appears to have an inhibitory effect on the ability of peripherin to assemble into a filamentous network.

Neurofilaments have been suggested to play a role in regulating axonal caliber and thereby influencing axonal conduction velocity. This role has been confirmed by studies of the Japanese *quiver* quail, which has a premature termination codon in its NFL gene and consequently no neurofilaments in its axons. Normal radial growth is severely inhibited in the large caliber fibers. When the NFL gene was deleted in NFL knockout mice, no visible IFs were observed and there was axonal hypotrophy. There was also a substantial decrease in NFM and NFH in the NFL knockout mice, presumably because they were unstable, because they could not polymerize into filaments. Curiously, in spite of this hypotrophy, the NFL-deficient mice developed normally and showed no obvious neurological problems. However, loss of function of NFL in humans causes a severe neuropathy. NFM $^{-/-}$ and NFH $^{-/-}$ mice have even less severe phenotypes. Loss of NFM also resulted in some reduction of the caliber of large axons and some axonal loss, whereas the absence of NFH in transgenic mice resulted in an even more modest reduction in axonal caliber. Overexpression of NFL and either NFH or NFM in transgenic mice resulted in an increase in the number of neurofilaments as well as axonal caliber, also consistent with the role of neurofilaments in regulating axonal calibers.

NFs are characterized by their phosphorylation, a posttranslational mechanism that is thought to regulate their functions. NFM and NFH are phosphorylated mainly in the KSP repeats located in their C-terminal tail domains. These phosphorylation sites are depicted in the tail domains seen in **Figure 1**. NFH has 44–51 KSP residues, whereas NFM has 5–14 of these residues, depending on species. Phosphorylation of NFs induces extension of their side arms from the filament backbone, apparently increasing the spacing between individual filaments and ultimately the axonal caliber. Phosphorylated NFs are found in axons, while the nonphosphorylated forms are found in neuronal cell bodies and dendrites. A number of different kinases have been shown to phosphorylate the KSP residues of NFM and NFH. Phosphorylation of the KSP

residues has been proposed to be important for the increase of the radial growth of axons by increasing the nearest neighbor distances between the neurofilaments, possibly through electrostatic repulsion between the neurofilament side arms. The evidence for this proposal came from electron microscopy studies that showed that dephosphorylated neurofilaments were more closely spaced together than phosphorylated neurofilaments. In the dysmyelinated *Trembler* mouse mutant, there is a decrease in neurofilament phosphorylation and smaller axonal diameter. When segments of *Trembler* sciatic nerves were grafted onto control nerves, the *Trembler* segments still maintained their smaller diameters and dephosphorylated neurofilaments, whereas adjacent regions showed normal diameters and neurofilament phosphorylation. These studies were considered as strong evidence that phosphorylation of neurofilaments (especially of NFH) is a determinant of axonal diameter. It, therefore, came as a surprise that the NFH knockout mice showed only a slight decrease in nearest neighbor spacing of neurofilaments and did not change the axon caliber, indicating that NFH phosphorylation may be less important for the determination of axonal caliber than was predicted. To test this particular property of NFH more rigorously, the NFH gene was replaced in transgenic mice with one deleted in the NFH tail. The results showed that the loss of the NFH tail and all of its phosphorylation sites did not affect the interfilament spacing of the neurofilaments. By contrast, the NFM tail and its state of phosphorylation may be more important for radial growth, even though it contains considerably fewer phosphorylation sites.

As described before, neurofilament proteins are synthesized in the cell body and have been found to move down the axon as part of the slow component of axonal transport with a velocity of $\sim 0.1\text{--}1\text{ mm d}^{-1}$. The initial hypothesis was that the entire axoplasm moved as a coherent block, but more recent work has shown that the entire axoplasm in its totality does not move. Neuronal IFs are now thought to move as monomeric and polymeric units. Studies using proteins tagged with green fluorescent protein and live-video microscopy have demonstrated that neuronal IFs undergo rapid, intermittent, and nonsynchronous transport in both directions, with predominance of the anterograde transport and long pauses in between movements along the axons. These studies indicate that slow transport is actually the result of fast motions interrupted by long pauses. This fast transport is likely to utilize one of the many motors that have been identified for fast transport. These motor proteins could also be responsible for some of the cross-links observed between IFs and microtubules. Other potential cross-linkers are members of the plakin family of proteins that were first identified in junctions in epithelial cells. More recent studies have shown that some plakins are also expressed in the nervous system. Plakins have distinct domains that include microtubule, microfilament, and IF-binding domains and could, therefore, be important in the formation of cross-links between these cytoskeletal elements. The phosphorylation of the side chains of the neurofilaments described above has also been considered to be important for the regulation of transport of neurofilaments through axons. Two studies have provided different conclusions regarding the role of NFH side arm

phosphorylation in neurofilament transport. In one study, all the potential phosphorylation sites of NFH were mutated either to preclude phosphorylation or mimic permanent phosphorylation of the NFH side arms. These mutants displayed altered rates of transport in a bulk transport assay. The mutant mimicking permanent phosphorylation spent a higher proportion of time pausing than one that could not be phosphorylated, indicating that NFH phosphorylation slows neurofilament transport. By contrast, in transgenic mice, where the endogenous NFH was replaced by a tailless NFH, there were no changes in the rate of bulk transport. These studies indicate that the mouse studies are not always consistent with the studies in cultured cells, in part, due to additional compensatory mechanisms that may exist in the mouse. In any case, some of these aspects of neurofilament transport will become clearer when the specific motors that move neurofilaments are identified.

α -Internexin

α -Internexin was first isolated from optic nerve and spinal cord and found to be particularly abundant in the developing nervous system. α -Internexin has an apparent molecular weight of 66 kDa and is, therefore, a little smaller than the NTFPs. Although it is present in both the CNS and PNS early in development, it is mainly CNS specific in the mature mammalian nervous system. Consistent with its expression early in development in the absence of other IFs, α -internexin has been shown to be able to self-assemble both *in vitro* and in transfected cells. α -Internexin is closely related to the NTFPs and is, therefore, also classified as a type IV IF protein. α -Internexin is the first neuronal IF expressed in neurons after they are committed to the neuronal lineage. The expression of α -internexin is high early in development and then decreases after the axon matures. In the PNS, the level of α -internexin decreases significantly in adulthood, although it persists especially in thinner unmyelinated axons. In the CNS, α -internexin expression is also lower after the axon matures. Of particular interest is the fact that α -internexin is the only neuronal IF expressed in some neurons, for example, the granule cell neurons in the cerebellum. The axons of these neurons are thinner in diameter than the ones that express the NTFPs suggesting that α -internexin is not important for maintaining axonal diameter. The early expression of α -internexin led to the suggestion that it might be important for axon outgrowth. Although this idea was supported by antisense inhibition experiments in cultured cells, the α -internexin knockout mice have no obvious phenotype. No compensatory changes in the levels of other neuronal IFs were observed and it appears that α -internexin is not necessary for axon outgrowth in these mice. Unfortunately, these knockout studies tell us little about other potential functions of this protein. Overexpression of α -internexin in transgenic mice results in the formation of cerebellar torpedoes. These torpedoes are swellings in Purkinje cell axons. The presence of these torpedoes results in the degeneration and ultimate death of the Purkinje cell. There is a dysfunction of the cell as determined behaviorally from the transgenic mice before the cells die.

Peripherin

Peripherin was first identified as an IF protein from neuroblastoma and its cDNA was cloned from a cell line that can be induced to grow neuronal processes by the addition of nerve growth factor, called PC12 cells. Peripherin has a molecular weight of 57 kDa and belongs to the type III IF proteins, which also consists of vimentin, the glial-specific GFAP, and the muscle-specific IF protein called desmin. Peripherin is, therefore, the only type III neuronal IF protein, because all the other neuronal IFs are in the type IV category. The significance of this difference is not known. Like all the other type III IFs, peripherin can self-assemble into a filamentous network. Peripherin can also coassemble with other neuronal IFs. Peripherin is predominantly found in the PNS (hence its name), but some CNS neurons also express peripherin. As described above, peripherin was identified as a gene that was increased in PC12 cells upon stimulation by nerve growth factor. Peripherin has also been shown to be upregulated following axonal injury. These studies suggested that peripherin may play a role in axon outgrowth and elongation. Peripherin depletion experiments with antisense technologies yielded conflicting results: initial studies showed that antisense oligonucleotides did not inhibit neurite outgrowth, whereas more recent studies using small interfering RNAs (siRNAs) showed that peripherin siRNA inhibits the initiation, extension, and maintenance of neurites in PC12 cells. The reason for this discrepancy is not clear. Peripherin knockout mice have normal caliber axons and show no axonal loss. As described in more detail below, some interesting observations have led to the possibility that misexpression of peripherin can lead to neurodegenerative disease.

Nestin

Nestin was first described as a marker of neuroepithelial stem cells. Subsequent cDNA cloning showed that a large IF protein encoded this marker. Nestin has the basic hallmarks of an IF protein with an α -helical rod, capable of forming coiled-coil dimers, a very short head domain and a very long tail domain. Nestin has little homology to the other mammalian IF proteins, but its genomic structure is similar to the NTFs and α -internexin. As a result, nestin is placed either as a separate type of IF (type VI), or as a type IV IF along with the NTFs and α -internexin. During early development, nestin is expressed in neuroepithelial cells and muscle progenitor cells. Transgenic mouse studies have indicated that different regulatory elements in the nestin gene are responsible for directing gene expression in developing muscle and neural precursor cells, respectively. Nestin is not able to self-polymerize, but requires vimentin to form a filamentous network. There have been no published reports of a nestin knockout mouse.

Neuronal Intermediate Filaments and Disease

Overexpression of neuronal IFs in transgenic mouse models results in the formation of mislocalized protein aggregates,

which affect neuronal function and survival. This has been shown in the case of human NFH, mouse NFL, peripherin, and rat α -internexin. The relationship between misexpression, disrupted transport, and mislocalization of neuronal IFs to neurodegeneration is also apparent from the appearance of axonal neurofilamentous accumulations in a number of neurodegenerative diseases. In particular, aggregation of IFs in motor neurons and alteration in their axonal transport has been reported in ALS. It appears likely that mutations in neurofilament proteins would also lead to neurodegeneration. When an assembly disrupting NFL mutant was produced in a transgenic mouse model, the mutant NFL caused massive degeneration of motor nerves. Several mutations have now been identified in neurofilament proteins that lead to neurodegenerative diseases. Deletions and insertions in the tail domain of NFH have been reported in 1–2% of patients with sporadic ALS; however, no familial cases of NFH mutations have been identified, making this connection a little bit more tenuous. There is a better-documented relationship between peripherin and ALS. Peripherin has been found in axonal spheroids in patients with ALS. Furthermore, overexpression of peripherin by four- to sevenfold in transgenic mice leads to late onset progressive motor neuron disease, which is accelerated in mice that lack NF. A number of mutations in peripherin have also been described in sporadic ALS cases. A number of mutations in the NFL gene have now been linked to Charcot–Marie–Tooth neuropathy type 2E. Two of these mutations (P8R and Q333P) are shown in [Figure 1](#). When these same mutations are introduced in an NFL cDNA and the mutant NFL constructs are expressed in cultured cells, defects in assembly and transport of NFL are observed. It appears, therefore, that these mutations in NFL might alter axonal transport, which, in turn, would cause neurodegeneration. This notion is consistent with the hypothesis that alterations in axonal transport (in which neuronal IFs could be directly or indirectly involved) play a role in the pathogenesis of other neurodegenerative diseases. A number of additional mutations have been described in both sporadic and familial CMT with varying degrees of severity depending on the site of the mutation. The mutations that have been described have all been in the head and the rod domain of NFL. A mouse model with one of these mutations (P22S) causes neuropathy in transgenic mice, which can be reversed by turning off the transgene. It is also of interest that in humans, a recessive *NEFL* mutation that appears to cause a simple loss of function results in a severe, early-onset axonal neuropathy.

See also: [Cell Architecture and Function: Actin Assembly/Disassembly; Actin-Capping and -Severing Proteins; Actin-Related Proteins; Intermediate Filament Linker Proteins: Plectin and BPAG1; Intermediate Filaments; Microtubule-Associated Proteins; Rho GTPases and Actin Cytoskeleton Dynamics; Tubulin and Its Isoforms.](#)

Further Reading

Coulombe PA, Ma L, Yamada S, and Wawersik M (2001) Intermediate filaments at a glance. *Journal of Cell Science* 114: 4345–4347.

- Helfand BT, Chang L, and Goldman RD (2003) The dynamic and motile properties of intermediate filaments. *Annual Review of Cell and Developmental Biology* 19: 445–467.
- Lariviere RC and Julien JP (2004) Functions of intermediate filaments in neuronal development and disease. *Journal of Neurobiology* 58: 131–148.
- Leung CL, Flores RG, and Liem RKH (1998) The complexity of intermediate filaments in the nervous system. In: Herrmann H and Warton R (eds.) *Intermediate Filaments*, pp. 497–526. New York, NY: Plenum.
- Leung CL, Green KJ, and Liem RKH (2002) Plakins: A family of versatile cytolinker proteins. *Trends in Cell Biology* 12: 37–45.
- Leung CL and Liem RKH (2003) Neuronal intermediate filament overexpression and neurodegeneration in transgenic mice. *Experimental Neurology* 184: 3–8.
- Liem RKH and Messing A (2009) Dysfunctions of neuronal intermediate filaments in disease. *Journal of Clinical Investigation* 119: 1814–1824.
- Shah JV and Cleveland DW (2002) Slow axonal transport: Fast motors in the slow lane. *Current Opinion in Cell Biology* 14: 58–62.
- Xiao S, McLean J, and Robertson J (2006) Neuronal intermediate filaments and ALS: A new look at an old question. *Biochimica et Biophysica Acta* 1762: 1001–1012.

Neurotransmitter Transporters

K Erreger, H J G Matthies, A Galli, and C Saunders, Vanderbilt University, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Aurelio Galli, Randy D. Blakely, and Louis J. DeFelice, volume 3, pp. 37–40, © 2004, Elsevier Inc.

Glossary

Neurotransmitter Message chemicals released from one neuron at the presynaptic nerve terminal.

Polymorphism A genetic variant in DNA sequence.

Synaptic cleft The gap between the pre- and postsynaptic neurons.

Neurotransmitters are chemical messengers by which neurons communicate with each other. Timely removal of neurotransmitters from the synaptic cleft is crucial for synaptic neurotransmission. If neurotransmitter molecules released during synaptic events were allowed to remain in the synaptic cleft, they would prevent new neuronal signaling mainly because the synapse would become refractory due to continued exposure of neurotransmitter. Removal of neurotransmitter from the synaptic cleft occurs by three mechanisms: diffusion, enzymatic degradation, and reuptake. High-affinity reuptake for the released neurotransmitters is mediated by transporter proteins localized at the membrane of nerve terminals and glial cells, and is the most common mechanism for the termination of synaptic events. This mechanism serves not only to clear the extracellular component of the synapses of neurotransmitter but also to recapture transmitter molecules for later reuse. In addition to their obvious and crucial role in neurotransmitter homeostasis, transporters are also important targets for therapeutic drugs as well as drugs of abuse and known neurotoxins.

Neurotransmitter transporters utilize the electrochemical energy derived from the movement of ions to drive the intracellular accumulation of neurotransmitter. In the early 1990s, several of these membrane proteins were cloned (the genetic sequence of DNA was identified) and two wide families have been identified. The first (SLC6 family), which includes the norepinephrine (NE), dopamine (DA), γ -aminobutyric acid (GABA), and serotonin (5HT) transporters, shares no significant sequence homology with the second (SLC1 family), which includes transporters for glutamate. After the cloning of these membrane proteins, there were several intriguing and exciting questions: What is the mechanism of transport? How does the protein structure define the functional properties of the transporter? What are the cellular signals and determinants that regulate transporter expression and function? Are there disease states related to neurotransmitter transporter dysfunction?

Mechanisms of Transport

Reuptake and Electrogenicity

Neurotransmitter transporters operate by coupling the transmembrane translocation of substrate to the movement of driving ions down pre-established electrochemical gradients (co-transport). Members of the SLC6 family include the DA

transporter (DAT), 5HT transporter (SERT), NE transporter (NET), and GABA transporter (e.g., GAT1) that are dependent on both extracellular sodium (Na^+) and chloride (Cl^-) to function, whereas the glutamate transporters (EAATs) depend on extracellular Na^+ and intracellular potassium (K^+). Both families of transporters use the Na^+ gradient generated by high extracellular Na^+ and low intracellular Na^+ to drive uptake of neurotransmitter into cells. The most widely held concept of how co-transporters function is founded on the alternating access model, in which the binding sites for substrates and ions are alternately exposed to extracellular or cytoplasmic environments via conformational changes in the transporter protein. The thermodynamic work of the transporter is accomplished by coupling the flux of substrate to the movement of co-transported ions down their electrochemical gradients. The exact stoichiometry of ions depends on the transporter and can even be modulated by regulatory proteins. The neurotransmitter transporter process is electrogenic, meaning that the transport process moves net charges across the plasma membrane due to the movement of ions accompanying the movement of the neurotransmitter. Indeed, in both heterologous cell lines expressing transporters and in native neurons in cell culture, electrical currents generated by transporters have been reported. In addition to the stoichiometric current due to movement of ions coupled to neurotransmitter transporters, electrical current that is independent of, or uncoupled from, neurotransmitter movement has been reported through transporters. Particularly notable are the relatively large Cl^- currents through the glutamate transporters, although uncoupled currents are also known for the monoamine transporters including DAT. These uncoupled currents through the transporters are large enough in magnitude to alter neuronal membrane potential and excitability.

Recent electrophysiological studies of both native and cloned neurotransmitter transporters revealed that they have brief and rare channel-like openings comparable to those generated by ligand-gated ion channels. Models for this exiting mode of the neurotransmitter transporters have been established, in which substrate and ions induce the carrier to transport in the alternating access mode (Figures 1(a) and 1(b)), and rarely switch to a channel-like mode (Figure 1(c)). The alternating access model assumes that the substrate permeation occurs through a state transition ($A \leftrightarrow B$) and results in the transport of a single neurotransmitter molecule (Figures 1(a) and 1(b)). The channel-like mode (Figure 1(c))

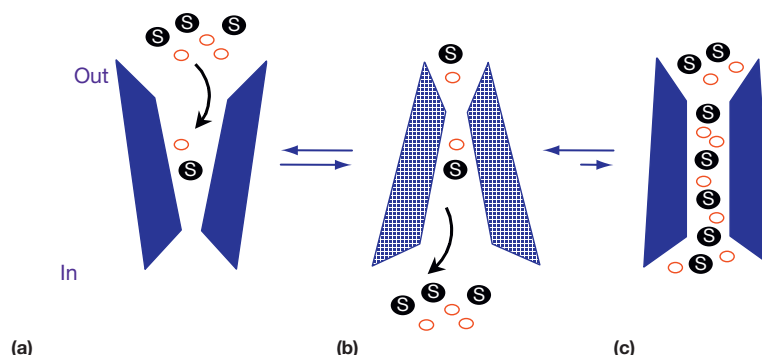


Figure 1 Model for neurotransmitter transporters function. S indicates a substrate molecule and the red circles depict a co-transported ion. The conformational switch between A and B is the classical alternating access model of co-transport. C represents the channel-like conformation that allows rapid bursts of ions and/or substrates to move across the membrane.

is a low-probability event that consists of hundreds of ions crossing the membrane (using the transporter aqueous pore) down their electrochemical gradients.

Reverse Transport

In addition to mediating the reuptake of neurotransmitter into nerve terminals, transporters can also cause nonvesicular release (efflux) of the neurotransmitter from the nerve terminal due to the transporter operating in reverse moving the substrate from the inside to the outside of the cell. This process can be induced by conditions that change the gradient of neurotransmitters and/or ions. For example, the abused psychostimulant amphetamine increases the concentration of dopamine on the inside of the cell as well as alters the gradient for Na^+ , resulting in a reverse transport of dopamine from the inside to the outside of the cell. Nonvesicular neurotransmitter release is thought to play a role under certain pathophysiological conditions. Ischemia during stroke impairs the normal Na^+ and K^+ gradients due to the inability of the Na^+/K^+ adenosine triphosphatase (ATPase) pump to function properly. The alteration of these ion gradients leads to reverse transport of glutamate through GluTs and subsequent toxicity from excess glutamate receptor activation. The direction of transport is also voltage dependent and there is evidence that physiological depolarization may induce some reverse transport in response to neuronal depolarization for both dopamine transporters and glutamate transporters.

Structure and Function

A major milestone for understanding the relationship between the structure and function of neurotransmitter transporters came in the form of the determination of the crystal structure for bacterial homologs for both the glutamate (SLC1) family and the monoamine (SLC6) family. Both structures generally confirmed earlier models of transmembrane domain topology based on hydrophathy predictions, site-directed mutagenesis, accessibility of individual residues to domain-specific antibodies, and accessibility of individual residues to chemical modification of engineered cysteine residues. Nonetheless, some details were surprising such as the broken conformation of helices surrounding the substrate-binding sites. The crystal structures have become

valuable tools both for the rational design and the interpretation of structural and function studies examining the interaction of substrates and drugs with the transporter, the conformational changes underlying transporter function, and the effects of altering specific amino acid structures on the transporters.

Regulation

It is not surprising that molecules regulating the spatial and temporal dimension of synaptic events are tightly regulated. Several reports have shown that many of the neurotransmitter transporters can be rapidly regulated by various signaling molecules. Regulators of neurotransmitter transporter activity include G-protein-coupled receptors such as the D_2 and mGluR5 receptors, kinases such as protein kinase C (PKC), phosphatidylinositol 3-kinase, tyrosine kinase, Ca^{2+} /calmodulin kinase, protein kinase B, neurotrophic factors, and hormones including insulin. In addition to uptake rates, some studies suggest that the regulation of neurotransmitter transport capacity might originate from a change in transporter cell-surface expression. In this context, it is important to point out that plasma membrane expression level of neurotransmitter transporters is finely controlled by transporter interacting proteins. For example, in addition to PKC and other signaling pathways, cell-surface expression and/or activity of GAT1, SERT, NET, and DAT are regulated by syntaxin 1A, a protein involved in synaptic vesicle fusion to the plasma membrane. Other regulators of DAT, NET, and SERT cell-surface expression include psychostimulants such as amphetamine and cocaine.

The general process of transporter trafficking, that involves endocytosis, intracellular sorting, and exocytosis of these membrane proteins, is acquiring great interest for neuroscientists. This is because researchers studying transporter cell-surface redistribution are envisioning the possibility of discovering new cellular targets for the cure of neurological disorders.

Relationships between Disease States and Neurotransmitter Transporters

The ability of the neurotransmitter transporters to regulate normal synaptic signaling implies that functional modification

of transporter activity might contribute to the etiology of multiple neurobiological diseases. Indeed, early studies have shown that the monoamine transporters NET, DAT, and SERT play an important role in regulating mood, learning, and motor activity, while GABA transporters have been implicated in neuronal excitability dysfunction such as epilepsy. While many years of innovative pharmacology resulted in the development of compounds that, by targeting neurotransmitter transporters, alleviate the symptoms of neurological diseases such as drug abuse and attention-deficit hyperactivity disorder (ADHD), new hopes for the cure of these conditions are coming from studying the gene organization and polymorphisms of these membrane proteins. Indeed, it has been shown that genetic variations in the human neurotransmitter transporters sequences known as 'polymorphisms' may alter transporter-expression level, activity, or regulation that ultimately may influence the levels of extracellular neurotransmitter.

Recently, genetic studies have associated specific polymorphisms in neurotransmitter transporters with human disease, and especially polymorphisms that lead to changes in the coding sequences have facilitated studies to determine the molecular mechanisms for these diseases. An example of a specific transporter genetic variant associated with human disease is the DAT variant A559V found in some children with ADHD. Molecular studies of this transporter suggest that anomalous reverse transport of dopamine may contribute to ADHD and that variation in the DAT gene may explain how psychostimulants are effective treatments for some children

with ADHD. As more genetic data becomes increasingly available, there will be ever-expanding opportunities to investigate how genetic variability in neurotransmitter transporters contributes to human disease and to design more rational treatment for these diseases.

See also: **Bioenergetics:** Membrane Transport, General Concepts; **Lipids Carbohydrates Membranes and Membrane Proteins:** Ion Channel Protein Superfamily; **Metabolism Vitamins and Hormones:** Glucose/Sugar Transport in Mammals.

Further Reading

- Bowton E, Saunders C, Erreger K, et al. (2010) Dysregulation of dopamine transporters via dopamine D2 autoreceptors triggers anomalous dopamine efflux associated with attention-deficit hyperactivity disorder. *Journal of Neuroscience* 30: 6048–6057.
- DeFelice LJ and Galli A (1998) Electrophysiological analysis of transporter function. *Advances in Pharmacology (New York)* 42: 186–190.
- Hahn MK and Blakely RD (2007) The functional impact of SLC6 transporter genetic variation. *Annual Review of Pharmacology and Toxicology* 47: 401–441.
- Jardetzky O (1966) Simple allosteric model for membrane pumps. *Nature* 211: 969–970.
- Mazei-Robison MS, Bowton E, Holy M, et al. (2008) Anomalous dopamine release associated with a human dopamine transporter coding variant. *Journal of Neuroscience* 28: 7040–7046.
- Sonders MS and Amara SG (1996) Channels in transporters. *Current Opinion in Neurobiology* 6: 294–302.
- Torres GE and Amara SG (2007) Glutamate and monoamine transporters: New visions of form and function. *Current Opinion in Neurobiology* 17: 304–312.

Neurotrophin Receptor Signaling

R S Kenchappa and B D Carter, Vanderbilt University Medical School, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Jennifer J. Gentry and Bruce D. Carter, volume 3, pp. 41–45, © 2004, Elsevier Inc.

Glossary

Apoptosis A form of cell death involving a series of programmed biochemical events, it is also called programmed cell death.

c-Jun-N-terminal kinase (JNK) Kinase enzyme, responsive to stress stimuli and phosphorylates several substrates.

Neurotrophin A member of a family of homologous growth factors that promote the survival and differentiation of neurons.

Nuclear factor kappa B (NF- κ B) A transcription factor.

Phosphatidyl-inositol 3 kinase (PI3K) Kinase that phosphorylates inositol lipids on the third position of the sugar.

PLC- γ An enzyme that cleaves phosphatidyl inositol (4,5) bisphosphate to release inositol-3-phosphate and diacyl glycerol (DAG).

Proneurotrophin Uncleaved form of neurotrophins.

p75NTR The 75-kDa neurotrophin receptor.

Receptor A cell surface protein specifically binds to particular substrate (e.g., ligand).

Trk The tyrosine kinase neurotrophin receptor.

The neurotrophins are a family of dimeric, secretory proteins, necessary for nervous system development and function. The members of the neurotrophin family include nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), neurotrophin 3 (NT3), and neurotrophin 4 (NT4), and each is essential for the survival of specific, but overlapping, populations of neurons during development. Given that nearly half of the neurons generated during development undergo programmed cell death, it is clear that neurotrophins have a major role in determining the ultimate population size for a given group of neurons. The neurotrophins not only promote survival, but also regulate cellular differentiation, axon guidance, synaptogenesis and plasticity, myelin formation, and neurite branching and elongation. Hence, this family of trophic factors has a wide spectrum of actions that are key for the proper sculpting of the mammalian nervous system. In addition to their role in development, this family of proteins also has important effects on maintenance of the nervous system, including processes such as learning and memory, neurogenesis, and response to injury. Like other secretory proteins, neurotrophins are generated from precursor proteins (proneurotrophins). However, these proneurotrophins are not simply transient intermediates in the biosynthesis of the mature protein, they can be secreted and are functionally active, promoting apoptotic cell death. This article focuses on the molecular signaling pathways activated by the neurotrophins and how these signals regulate neuronal survival and death.

The broad range of biological functions exerted by neurotrophins reflects the variety of signaling pathways that are influenced by activation of their receptors. The neurotrophins bind to two unrelated receptor types, a family of tyrosine kinase receptors, the Trks, and a member of tumor necrosis factor (TNF) family, the p75 neurotrophin receptor (p75NTR). Neurotrophin binding to the Trks promotes survival and differentiation while activation of p75NTR can also enhance cell viability or, paradoxically, induce apoptosis. Although activation of the Trks or p75NTR can initiate discrete signal transduction pathways, the two receptor types are frequently

coexpressed on a given cell and can function in a complex. In addition to the Trk/p75NTR complex, the p75NTR can interact with several other transmembrane proteins, forming various complexes. Thus, it is the manifold interactions between these receptors and the cross-talk between their signaling pathways that ultimately determines the fate of the cell. These receptor interactions, the signaling pathways activated, and their functional implications are discussed.

Neurotrophin Signaling through Trk Receptors

The most well-studied actions of neurotrophins are mediated through activation of a family of tyrosine kinase receptors, the Trks. There are three members of the Trk family, TrkA, TrkB, and TrkC. Each receptor binds to specific members of the neurotrophin family preferentially with NGF having high affinity for TrkA, BDNF, and NT4/5 for TrkB, and NT3 for TrkC. In addition, NT3 can also activate other Trk receptors with less specificity. Binding of a neurotrophin to a Trk receptor triggers receptor dimerization and activation of intrinsic tyrosine kinase activity. The intracellular domains (ICDs) of the Trk receptors contain 10 evolutionarily conserved tyrosine residues, three of which are in an autophosphorylation loop of the tyrosine kinase domain. Similar to other receptor tyrosine kinases, phosphorylation of these tyrosine residues results in increased Trk tyrosine kinase activity. Phospho-tyrosine residues in the ICD serve as docking sites for proteins containing Shc-homology 2 (SH2) or phospho-tyrosine-binding domains, including the enzyme phospholipase C- γ (PLC- γ) and the adaptor proteins Shc, fibroblast growth factor receptor substrate 2 (Frs-2), rAPS, and SH2-B. The major pathways activated by the Trk receptors are the Ras/MAP kinase pathway, the PI3 kinase/Akt pathway, and the PLC- γ pathway and their downstream effectors (MAP kinase, mitogen-activated protein kinase).

Phosphorylation of TrkA at Y785 and similar sites on TrkB and TrkC recruits the cytoplasmic enzyme PLC- γ , which

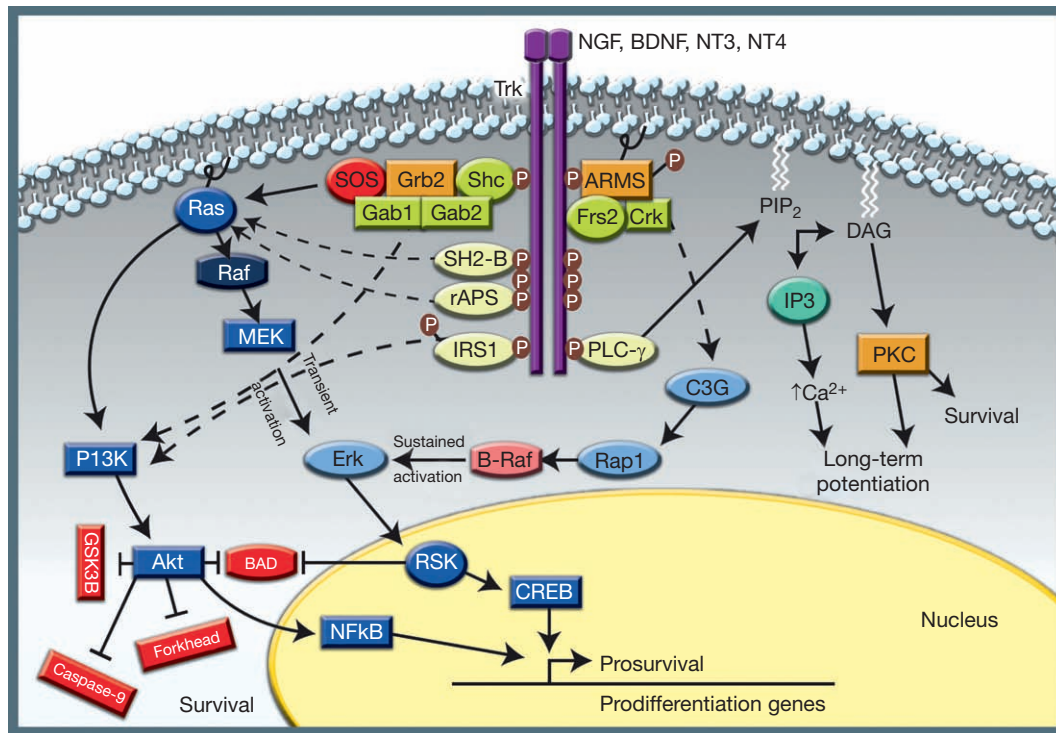


Figure 1 Neurotrophin signaling through Trk receptors. Neurotrophin binding to Trk receptors phosphorylates tyrosine residues in the intracellular domain and increases tyrosine kinase activity. Three major pathways get activated in response, the Ras/MAP kinase pathway, the PI3 kinase/Akt pathway, and the PLC- γ pathway, and these regulate cell survival, differentiation, and long-term potentiation.

associates with Trk receptors through its SH2 domain and is then phosphorylated by the receptor. Active PLC- γ cleaves phosphatidylinositol 4,5-bisphosphate to release the second messengers inositol 1,4,5-trisphosphate (IP3) and diacylglycerol (DAG) (Figure 1). DAG activates specific isoforms of protein kinase C (PKC), while IP3 triggers the release of Ca^{2+} from intracellular stores, which activates a variety of pathways, including calmodulin kinases and Ca^{2+} -regulated isoforms of PKC. The maintenance of calcium homeostasis has been well documented as key for regulating neuronal survival. Indeed, a role for PKC in the prosurvival effects of the neurotrophins has been suggested; however, this pathway also has been implicated in regulating spatial learning and memory. Analysis of transgenic, knock-in mice lacking the docking site for PLC- γ on the TrkB receptor has not revealed any defects in neuronal survival, but the animals do display deficiencies in hippocampal long-term potentiation, which is a model for spatial memory. The activation of PLC- γ also modulates the activity of TRP channels, a family of cation channels involved in thermal and mechanical sensation.

Another signaling cascade activated by the Trk receptor is the Ras–MAP kinase pathway. Stimulation of this pathway occurs through the binding of several different adaptors to the Trk receptors. The transient activation of Ras is mediated by the adaptor protein Shc, which is recruited to a phosphorylated tyrosine residue (Y490 on TrkA), leading to docking of the adaptor proteins Gab1, Gab2, and Grb2. Grb2 associates with the guanine-nucleotide-exchange factor Son of sevenless, which facilitates the activation of Ras, a guanosine triphosphate (GTP)-binding protein that activates several pathways.

Additional scaffold proteins, including SH2-B, adaptor protein with pleckstrin homology and Src homology-2 domains (APS), and insulin receptor substrate-1 (IRS1), also provide links to Ras. The activated, GTP-bound form of Ras recruits the serine/threonine kinase Raf to the plasma membrane, promoting its activation. Raf then phosphorylates MEK1 and/or MEK2 and these kinases phosphorylate the MAP kinases Erk1 and Erk2 (Figure 1). Ras also promotes the activation of the MAP kinase Erk5, through a cascade involving the upstream kinase MEK5.

The ability of neurotrophins to promote differentiation appears, in part, to depend on prolonged Erk1/2 activation, as opposed to the transient activation of the kinases induced by other growth factors such as EGF. The sustained activation requires a different adaptor protein, Frs-2. Frs-2, activated by binding to phosphorylated Y490 on Trk receptors, provides a binding site for the adaptor proteins Grb2 and Crk. Crk binds to the guanine-nucleotide-exchange factor C3G, which then stimulates the small GTP-binding protein Rap1. Alternatively, Crk can also associate with Trk receptors through binding to the transmembrane protein ankyrin repeat-rich membrane-spanning adaptor (ARMS). Upon neurotrophin binding to the Trk receptor, ARMS is phosphorylated, which creates a binding site for Crk, leading to Rap1 activation through C3G. The kinase B-Raf is then activated by Rap1, which results in sustained Erk1/2 activation (Figure 1).

The Erks, activated by neurotrophins/Trk receptors, translocate into the nucleus and regulate gene expression directly by phosphorylating transcription factors such as cAMP response element-binding (CREB), Egr-1, and Elk-1 or indirectly through the phosphorylation of a variety of substrates including the

ribosomal S6 kinase (RSK), which can then phosphorylate and inactivate the proapoptotic protein Bad. RSK also phosphorylates the transcription factor CREB, resulting in its activation and the subsequent upregulation of downstream genes to regulate cell survival and differentiation of neurons.

Another Trk-mediated signal that promotes neuronal survival and neuroprotection is the activation of the lipid kinase, phosphatidylinositol 3 kinase (PI3-kinase). Active Ras can lead to PI3-kinase stimulation; however, this kinase is also activated independent of Ras through Shc-Grb2 or through IRS1. Neurotrophin binding to Trks recruits Gab1/2 to phosphorylated Grb2, resulting in binding, phosphorylation, and activation of PI3-kinase. Trk receptor activation also leads to phosphorylation of IRS1, which permits activation of PI3-kinase. PI3-kinase produces P3-phosphorylated phosphatidylinositides and these serve to indirectly activate the serine/threonine kinase Akt. Akt phosphorylates and regulates several signaling proteins that are critical factors controlling cell survival. It inhibits the proapoptotic proteins BAD, forkhead transcription factors, GSK3 β and caspase 9, and induces the activation of the prosurvival transcription factor nuclear factor kappa B (NF- κ B). Several studies have demonstrated a requirement for the PI3-kinase-Akt pathway in neurotrophin-mediated survival; however, this pathway also promotes protein translation, neurite branching, and neuronal differentiation (Figure 1).

Adding to the complexity of Trk signaling is the existence of several splice variants of all of these receptors. Among these isoforms, both TrkB and TrkC exist as truncated receptors with short cytoplasmic motifs lacking the tyrosine kinase domain. Although there is evidence that loss of the kinase domain can function to abrogate Trk signaling, such as activation of PLC- γ , by competitively binding the available neurotrophin or forming inactive complexes with full-length receptors, there is also evidence that these truncated Trks may function to display bound neurotrophin to adjacent cells. In addition, recent findings suggest that the truncated T1 isoforms of TrkB and TrkC receptors can mediate signaling on their own, including the release of calcium stores and the regulation of the small GTPase Rac1, respectively. A second type of splice variant observed for all of the Trks is a form of the receptor lacking a small portion of the extracellular domain in the juxtamembrane region. Expression of these isoforms more stringently restricts the responsiveness of a given Trk for its cognate neurotrophin. For example, the TrkA isoform with the insert is activated by NT3 in addition to NGF. This may be one mechanism to insure the elimination of neurons making connections to an inappropriate target tissue.

One important aspect of neurotrophin signaling that needs to be considered is the spatial limitations of a neuron. Pioneering work by Rita Levi-Montalcini demonstrated that neuronal survival is critically dependent on their targets, which are often far removed from the cell soma. Typically, it is the axonal tip that encounters the neurotrophins, while the regulation of survival involves transcriptional activity; hence, the signaling must be retrogradely transported some distance back to the neuron soma. It has long been known that neurotrophins are retrogradely transported; however, it was only recently demonstrated that it is essential that the factor be transported with the active receptor, although this requirement has been contended by Campenot and colleagues. TrkA and NGF are not transported alone, rather, NGF-TrkA signaling leads to endocytosis of a

ligand-receptor complex and formation of signaling endosomes (signalosomes) containing numerous signaling components. These signalosomes are then transported retrogradely to the neuron cell body, where gene expression is regulated.

The mechanisms regulating the endocytosis and trafficking of the Trk receptors is a topic of intense investigation. The binding of neurotrophin to a Trk receptor stimulates internalization through a clathrin-mediated process; however, endocytosis has also been shown to occur through a novel protein, Pincher. The internalization and subsequent trafficking of Trk receptors is regulated by ubiquitination by E3 ubiquitin ligases. For example, the E3 ligase Nedd4-2 is phosphorylated upon TrkA activation, which leads to TrkA ubiquitination and down regulation. There remain many unknown mechanistic details of Trk internalization and the subsequent vesicular sorting.

Neurotrophin Signaling through p75 Neurotrophin Receptor

The p75NTR was the first neurotrophin receptor discovered, identified as a low-affinity receptor for NGF and subsequently shown to bind all the neurotrophins with equal affinity. This receptor plays an important role in regulating cell death during neuronal development and in response to nervous system injury. In addition, p75NTR signals modulate myelination, synaptic plasticity, neurite growth, cell-cycle regulation, and the invasiveness of tumor cells (Figure 2).

The p75NTR is the founding member of the tumor necrosis receptor family, which is comprised of single-pass transmembrane proteins containing cysteine-rich repeats in their extracellular domain and having a region of weak homology in their ICD referred to as a Death Domain, which mediates protein-protein interaction. In neurons, p75NTRs have been shown to exist as dimers, linked through a highly conserved cysteine²⁵⁷ in the transmembrane domain of the receptor that forms a disulfide bond. Binding of dimeric neurotrophins to preexisting p75NTR dimers leads to conformational rearrangement of the receptor, thereby initiating downstream signaling.

p75NTR contributes to a plethora of signaling complexes. It associates with several coreceptors, including the Trk receptors, Sortilin, Nogo-receptor, Lingo-1 and Ephrin A, and it mediates the response to multiple ligands, including NGF, BDNF, NT3, NT4, proneurotrophins, myelin-associated glycoprotein (MAG), Nogo, and oligodendrocyte myelin glycoprotein (OMgp) (Figure 2). Given this diversity of receptor complexes and ligands, it is not surprising that p75NTR can activate multiple signaling pathways with important physiological consequences. Nevertheless, the molecular mechanisms by which the receptor mediates the many downstream responses remain poorly understood. In an attempt to understand how the receptor transduces its signals, a number of proteins have been identified that can bind to its ICD, including neurotrophin receptor-interacting factor (NRIF), TNF α receptor-associated factor 6 (TRAF6), Rho-GDI, p75NTR-associated death executor, the neurotrophin receptor interacting melanoma antigen gene (MAGE) family members neurotrophin receptor-interacting MAGE homolog (NRAGE) and Necdin, Schwann cell factor 1 (SC1), receptor-interacting factor 2 (RIP2), brain-expressed-X-linked 1 (Bex1), and the actin bundling protein Fascin (Figure 2).

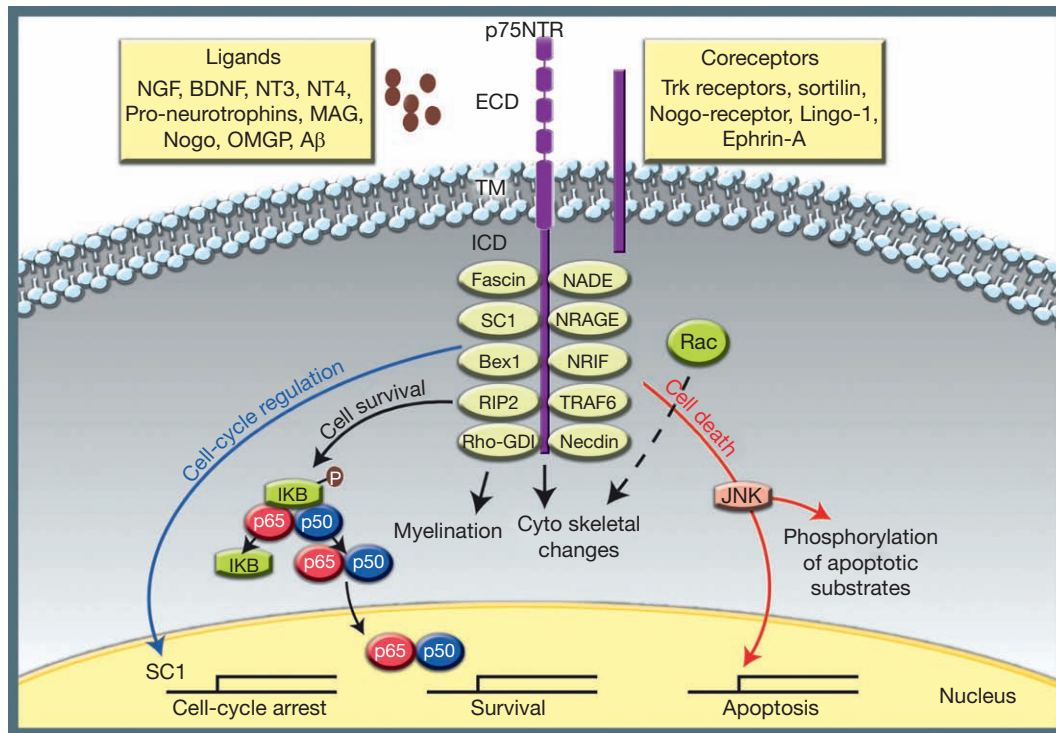


Figure 2 Neurotrophin signaling through the p75 neurotrophin receptor. Binding of either mature-neurotrophins, proneurotrophins, or nonneurotrophin ligands to the p75NTR extracellular domain leads to the recruitment of multiple adaptor proteins to the intracellular domain. These intracellular interactors initiate a wide range of signaling pathways, depending on cellular context and existence of coreceptors. Signaling pathways activated by the p75NTR regulate cell survival, cell death, myelination, cytoskeletal changes, cell cycle, long-term depression, and cell migration.

The first-recognized function of neurotrophin binding to the p75NTR was to facilitate the prosurvival signals generated by the TrkA receptor. When coexpressed, TrkA and p75NTR form a high-affinity-binding complex that increases the specificity of neurotrophin binding to TrkA and enhances tyrosine kinase signaling and the subsequent trophic effects. However, in specific contexts, p75NTR can also promote survival independent of the Trks. One of the best-characterized antiapoptotic signals generated by receptors in TNF receptor superfamily is the activation of the transcription factor, NF- κ B. This transcription factor can also be activated by p75NTR and was shown to mediate a survival signal in neurons and glia. This signal involves the recruitment of RIP2 and TRAF6 to the ICD of the receptor after neurotrophin binding leading to activation of NF- κ B (Figure 2).

The most surprising cellular response to p75NTR activation is the induction of programmed cell death. Although neurotrophins were discovered and best characterized for their ability to promote neuronal survival, it is now well established that neurotrophin binding to p75NTR in the absence of Trk signaling can induce an apoptotic response in specific cellular contexts. Furthermore, the proforms of the neurotrophins are selective p75NTR ligands and have been shown to induce apoptosis in many cell types. Like most-secreted peptides, the neurotrophins are initially translated containing an amino terminal domain that is proteolytically removed to form the mature, active protein. However, Hempstead and colleagues found that proNGF and proBDNF bind selectively to p75NTR, but not to TrkA or TrkB, and can potently induce apoptosis.

The proneurotrophins activate p75NTR through binding to a receptor complex that includes p75NTR and the VPS10P-domain receptor, Sortilin. The prodomain of proNGF or proBDNF binds to Sortilin, while the mature portion of the neurotrophin binds to the p75NTR. Although Sortilin is required for the proNGF or proBDNF binding, it is not clear how it may contribute to the apoptotic signal. The mechanisms regulating secretion of the proneurotrophins are not known; however, proNGF was detected in the cerebral spinal fluid of rats after lesion of the cortico-spinal tract and antibodies to the prodomain attenuated the resulting death of the cortical neurons. These findings suggest that proNGF may be released in response to certain neuronal injuries.

Numerous insults to the nervous system upregulate the expression of p75NTR and, in several instances, the receptor has been shown to mediate the accompanying apoptosis; for example, in hippocampal neurons after the induction of seizures, cortico-spinal neurons after axotomy and in oligodendrocytes after spinal cord injury. In addition, increased p75NTR expression has been observed in a wide variety of neuropathologies, such as focal ischemia, axotomy, stroke, Alzheimer's disease, and motor neuron disease, leading to speculation that p75NTR may be involved in these neuropathologies. P75NTR is also involved in regulating the naturally occurring cell death of specific neuronal populations during the development of the mammalian nervous system. Apoptosis mediated by p75NTR has been demonstrated to occur *in vivo* in the developing rodent spinal cord, retina, forebrain, and sympathetic ganglia.

Given the critical role p75NTR plays in regulating cellular viability during development and after injury, the molecular mechanisms that mediate the apoptotic signal are under intensive investigation. Like many members of the TNF receptor superfamily, p75NTR activates the stress kinase, c-Jun N-terminal kinase (JNK) and this signal is necessary for the receptor to induce apoptosis. Several p75NTR-interacting proteins have been linked to JNK activation, such as NRAGE, TRAF6, and NRIF. Genetic deletion of NRAGE, NRIF, or TRAF6 prevented the receptor from activating the kinase, suggesting that all three interactors contribute to this pathway, although the mechanistic details remain to be determined (Figure 3).

Cell death induced by JNK activation has been shown to occur through both transcription-dependent and independent mechanisms; for example, JNK activates members of the AP-1 family of transcription factors, leading to upregulation of proapoptotic genes such as Bim, Bak, and Fas ligand, and it also promotes apoptosis through phosphorylation of cell death factors such as p53, BAD, and Bim. Similarly, p75NTR-mediated cell death was demonstrated to involve JNK phosphorylation of Bim, Bad, and p53. However, protein synthesis inhibitors can block p75NTR-dependent apoptosis, suggesting that transcription may have a role (Figure 3).

A role for transcription in p75NTR-mediated apoptosis was also indicated by the fact that NRIF, a p75NTR-interacting protein that also can bind DNA, can be found in the nucleus following receptor activation by proapoptotic ligands. Nuclear localization of NRIF was found to be essential for p75NTR-mediated cell death, because a point mutant of NRIF, unable

to translocate to the nucleus, failed to reconstitute receptor-mediated apoptosis in *nrf1*^{-/-} neurons. The mechanism by which activation of the receptor facilitates NRIF nuclear translocation is reminiscent of signaling activated by the Notch receptor. Upon binding an apoptotic ligand, the receptor undergoes proteolytic cleavage first by TNF α -convertase enzyme (TACE) and then by the γ -secretase complex, thereby releasing the ICD. The liberated ICD allows for nuclear translocation of NRIF and subsequent apoptosis (Figure 3). The genes regulated by NRIF have yet to be identified. In addition, the mechanism by which p75NTR cleavage is regulated and how it relates to JNK activation remains unknown.

Proteolysis of the receptor by γ -secretase also plays a role in the regulation of neurite outgrowth by p75NTR. The myelin proteins Nogo, MAG, or OMgp bind to the Nogo receptor (NgR), a glycosylphosphatidylinositol-linked protein with no ICD, and this promotes the association of the NgR with p75NTR and another transmembrane protein termed Lingo-1. The formation of this complex recruits the guanine-nucleotide-exchange protein Kalirin9 and the Rho inhibitor Rho-GDI, resulting in activation of the small GTP-binding protein RhoA. This signaling pathway requires cleavage of the p75NTR, although how receptor processing facilitates the cascade is not clear. The activation of Rho causes reorganization of the actin-cytoskeleton, leading to growth cone collapse and the inhibition of axonal growth.

The ability of p75NTR to regulate the actin-cytoskeleton also plays a role in the migration and invasiveness of some tumor cells. Recent findings revealed that the most invasive cells in human gliomas express high levels of the p75NTR and

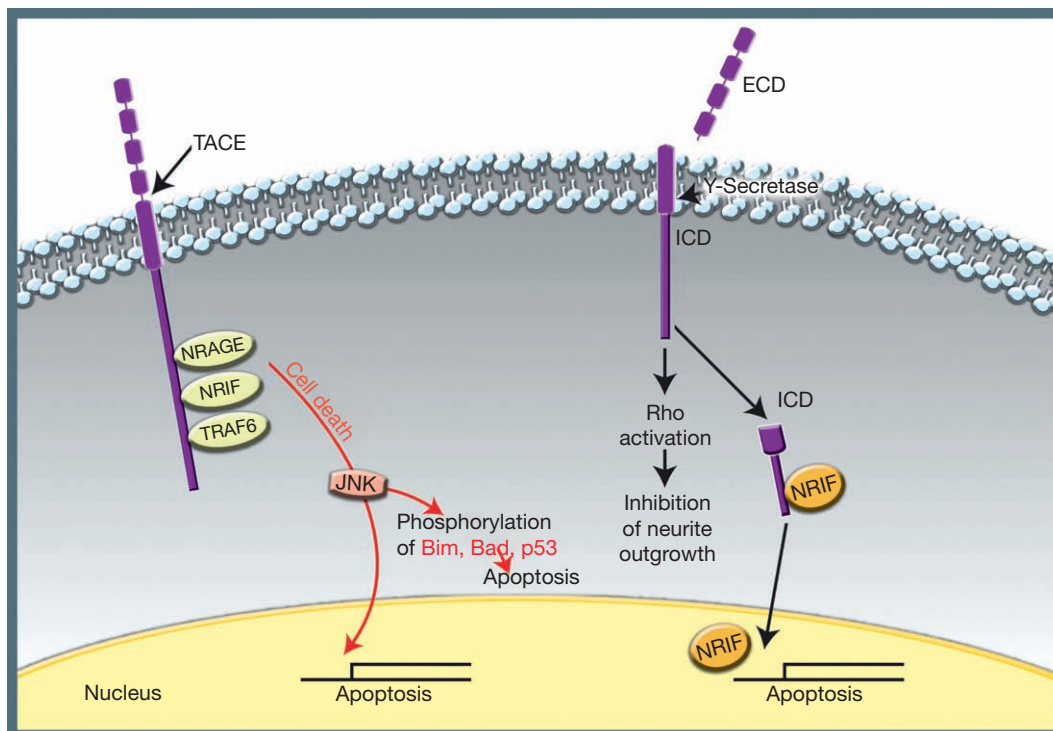


Figure 3 Proapoptotic signaling by the p75 neurotrophin receptor. Activation of the p75NTR by proapoptotic ligands activates the JNK pathway to induce apoptotic cell death. In addition, p75NTR undergoes regulated proteolytic cleavage by TACE and γ -secretase to release its intracellular domain, which is required for NRIF nuclear translocation and receptor-mediated cell death.

that expression of this receptor was sufficient to convert non-invasive gliomas to highly invasive tumors when xenografted into mice. The expression of the receptor in glioma cells was sufficient to induce RhoA activation and the formation of actin filaments. Similarly, the receptor is highly expressed in aggressive, invasive, malignant melanomas, and neurotrophin binding to p75NTR on these cells *in vitro* increased their invasiveness and promoted their survival. Neurotrophin-mediated migration of melanoma cells *in vitro* involved receptor binding to the actin bundling protein Fascin (Figure 2). Together, these findings suggest that p75NTR might be a target for developing therapeutic strategies for these deadly cancers.

In contrast to its action in melanoma and glioma cells, in prostate cells the p75NTR has been suggested to function as a tumor suppressor, promoting cell death and preventing cell proliferation. Interestingly, several of the receptor-interacting proteins have been shown to regulate the cell cycle, including NRIF, SC1, and Bex1. Indeed, SC1 can bind directly to the promoter for cyclin E and suppress its transcription. Given the widespread expression of p75NTR during embryonic development, particularly on many progenitor cells, it is provocative to consider the possibility that this receptor may control the population size of progenitors not only through apoptotic signaling, but also by modulating their rate of proliferation.

In addition to the aforementioned pathways, neurotrophin binding to the p75NTR has been shown to initiate a lipid-based signal. The receptor stimulates acid sphingomyelinase, which cleaves sphingomyelin to produce the second messenger ceramide. An increase in cellular ceramide has been linked to the promotion of cell death as well as survival by the p75NTR. In addition, this lipid can modulate a wide variety of signaling pathways, such as ERK, PI3 kinase, JNK, and NF- κ B.

Cross-Talk between p75NTR and Trk Receptor Signaling

In many neurons, Trk receptors are coexpressed with the p75NTR; thus, any neurotrophin binding to a Trk receptor will also activate the p75NTR. As these receptors can initiate divergent signaling pathways with opposing consequences, there must be mechanisms in place that ultimately allow the transduction of a consistent biological response. When both receptors are present, they interact to form a selective, high-affinity Trk-p75NTR complex. The interaction of p75NTR and Trk increases binding affinity for mature neurotrophins and enhances Trk tyrosine kinase activity, resulting in more robust

activation of downstream survival signals. This enhanced signaling is particularly important given the limiting amounts of neurotrophins available at target tissues. However, to promote survival, Trk must also prevent p75NTR from initiating its apoptotic cascade. Although the mechanisms by which this suppression occurs have yet to be fully elucidated, it is known that activation of the PI3-kinase–Akt pathway can inhibit JNK activation. For example, Akt can bind the scaffolding protein JNK-interacting protein 1 and displace members of the JNK cascade. In addition, both receptors can associate with TRAF6 (TrkA interacts via the adaptor p62), and this E3 ligase promotes the activation of NF- κ B, leading to upregulation of prosurvival genes. Similarly, ARMS bridges p75NTR and TrkA and leads to prolonged ERK activation. How ARMS affects p75 signaling, though, is not known. Clearly, there is much yet to learn about the complex relationship between the signaling pathways initiated by these two types of neurotrophin receptor.

See also: [Signaling: Mitogen-Activated Protein Kinase Family; Phospholipase C.](#)

Further Reading

- Chao MV (2003) Neurotrophins and their receptors: A convergence point for many signaling pathways. *Nature Reviews Neuroscience* 4: 299–309.
- Dechant G and Barde YA (2002) The neurotrophin receptor p75 (NTR): Novel functions and implications for diseases of the nervous system. *Nature Neuroscience* 5: 1131–1136.
- Ginty DD and Segal RA (2002) Retrograde neurotrophin signaling: Trk-ing along the axon. *Current Opinion in Neurobiology* 12: 268–274.
- Glebova NO and Ginty DD (2005) Growth and survival signals controlling sympathetic nervous system development. *Annual Review of Neuroscience* 28: 191–222.
- Hempstead BL (2002) The many faces of p75NTR. *Current Opinion in Neurobiology* 12: 260–267.
- Huang EJ and Reichardt LF (2001) Neurotrophins: Roles in neuronal development and function. *Annual Review of Neuroscience* 24: 677–736.
- Huang EJ and Reichardt LF (2001) Trk receptors: Roles in neuronal signal transduction. *Annual Review of Biochemistry* 72: 609–642.
- Lee FS, Rajagopal R, and Chao MV (2002) Distinctive features of Trk neurotrophin receptor transactivation by G protein coupled receptors. *Cytokine and Growth Factor Reviews* 13: 11–17.
- Nykjaer A, Willnow TE, and Petersen CM (2005) P75NTR – live or die. *Current Opinion in Neurobiology* 15: 49–57.
- Reichardt LF (2006) Neurotrophin-regulated signaling pathways. *Philosophical Transactions of the Royal Society B* 361: 1545–1564.
- Willnow TE, Petersen CM, and Nykjaer A (2008) VPS10P-domain receptors – regulators of neuronal viability and function. *Nature Reviews Neuroscience* 9: 899–909.

Nicotinamide Nucleotide Transhydrogenase

J Rydström, University of Gothenburg, Gothenburg, Sweden

© 2013 Elsevier Inc. All rights reserved.

Glossary

Membrane protein A membrane-associated protein, in this context an integral membrane protein spanning the membrane.

NAD(H) Oxidized (NAD^+) and reduced (NADH) nicotinamide adenine dinucleotide, a common cofactor in cell metabolism, especially catabolism.

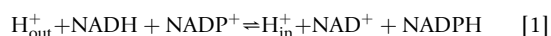
NADP(H) Oxidized (NADP^+) and reduced (NADPH) nicotinamide adenine dinucleotide phosphate, a common cofactor in cell metabolism, especially anabolism.

Proton pump A group of membrane proteins catalyzing a translocation of protons across the membrane, driven by either redox energy, an ion gradient, or adenosine triphosphate.

Transhydrogenases A group of enzymes composed of two subgroups: a soluble flavine-containing type and a membrane-bound and proton-translocating type. Both catalyze the transfer of a hydride ion between NAD(H) and NADP(H).

Introduction

Proton-pumping nicotinamide nucleotide transhydrogenase (TH; E.C. 1.6.1.2) is an integral membrane protein found in most species except some yeasts, plants, and certain bacteria. It is located in the inner membrane of mitochondria and in the plasma membrane of bacteria. The enzyme catalyzes the reversible reduction of nicotinamide adenine dinucleotide phosphate (NADP^+) by a hydride ion, donated by NADH (the reduced form of nicotinamide adenine dinucleotide), which is linked to proton translocation across the membrane according to the reaction



In the absence of an electrochemical proton gradient (Δp), the reaction from left to right (the forward reaction) is approximately fivefold slower than the reverse reaction. Under these nonphysiological conditions, the overall reaction proceeds to an equilibrium with an equilibrium constant of close to 1. In the presence of a Δp , that is, under more physiological conditions, the forward reaction is increased 5- to 10-fold and the apparent equilibrium constant is increased to ~ 500 , with one proton translocated per NADPH (the reduced form of nicotinamide adenine dinucleotide phosphate) formed. The net effect of the TH reaction is therefore probably to provide NADPH at the expense of NADH and Δp . Thus, this suggests important roles of TH in mitochondrial/cellular redox regulation including biosynthesis, detoxification (via the glutathione/thioredoxin systems), and apoptosis. TH is globally expressed in eukaryotes, that is, human, mouse, and *Caenorhabditis elegans*.

Proton-translocating THs from eukaryotes are homodimeric proteins with a relatively large hydrophilic domain composed of the NAD(H)-binding domain I (dI) and the NADP(H)-binding domain III (dIII); part of dI is homologous to alanine dehydrogenase (Figure 1). The hydrophobic part of the enzyme containing the proton channel, denoted domain II (dII), is composed of a variable number of transmembrane helices depending on species. The *Escherichia coli* TH is composed

of an α -subunit and a β -subunit, the former subunit containing dI and helices 1–4 of dII and the latter subunit containing dIII and helices 6–14 of dII. Intact *E. coli* TH has the composition $\alpha_2\beta_2$, that is, it is a tetramer in which $\alpha + \beta$ corresponds to the mitochondrial single polypeptide. dI and dIII can be expressed separately and purified in an active state and, when combined, show catalytic activity. In order to elucidate the coupling mechanism of TH, the enzyme mainly from bovine, human, and *E. coli* has been extensively studied in the purified forms, in cytosolic (plasma membrane) vesicles, as well as reconstituted in liposomal membranes. dI and dIII from bovine, human, and *Rhodospirillum rubrum*, as well as a complex between dI and dIII, have been structurally determined using X-ray crystallography and nuclear magnetic resonance (NMR).

Structure-Function of the Substrate-Binding Domains

Recent advances in structural resolution of the hydrophilic dI and dIII by X-ray crystallography and NMR spectroscopy have dramatically increased our knowledge about the structure-function relationships of these domains and their interactions. Thus, the crystal structure of the NAD^+ form of dI from *R. rubrum* solved by Jackson and co-workers, and in the NADH form from the same source later by Hatefi and co-workers, shows that it involves a dimer of a 40-kDa subunit, denoted as A and B. Each subunit comprises two subdomains, dI.1 and dI.2, both of which have a Rossmann nucleotide-binding fold composed of a sheet of parallel β -strands flanked by α -helices. The two domains are arranged on the sides of a deep cleft. However, in the crystal structure of the dimer only dI.2(A) contains bound NAD^+ positioned so that its nicotinamide moiety is exposed in the cleft. By contrast, dI.2(B) site contains disordered NAD^+ or no ligand. The arrangement of the monomers in the dimer is such that the nucleotide-binding sites in dI.2(A) and dI.2(B) are pointing away from each other. The potential nucleotide-binding sites in dI.1 (A) and dI.1(B) have no known ligands. It is obvious that in the dI dimer, dI.2(A)

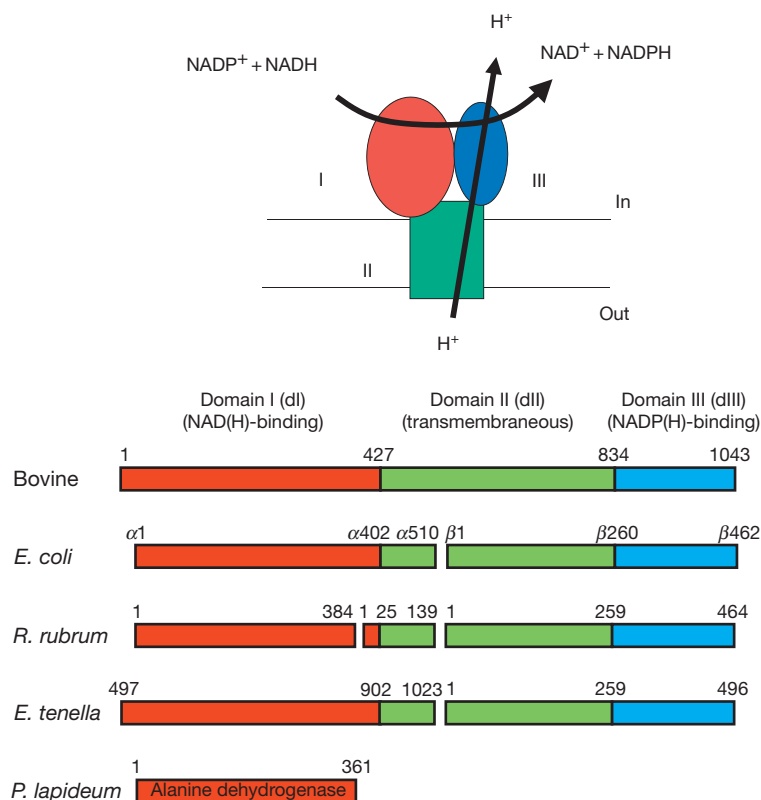


Figure 1 Structural organization of proton pumping THs.

and dI.2(B) must have different conformations, because only dI.2(A) contains well-ordered NAD^+ . In a dimeric structure of dI with bound NADH, other minor but significant structural changes are observed. The structure of the corresponding dI from *E. coli* TH, recently solved by Krengel and co-workers, was nearly identical to that of the *R. rubrum* enzyme.

Human and bovine dIII have been structurally resolved by Jackson and co-workers and Hatefi and co-workers, respectively, using X-ray crystallography. These are structurally close to identical. The structures involve an alternating α/β -structure with a central β -sheet core, and a Rossmann fold in which NADP^+ is tightly bound at the C-terminal end of the β -sheet. The molecular mass of dIII is 20 kDa, it is a monomer, and it is isolated with predominantly tightly bound NADP^+ ; the K_d of the NADP(H) bound is in the nM– μM range. A striking difference, however, as compared with other NAD(P) -binding enzymes is that the NADP^+ bound to dIII is turned approximately 180° , that is, the nicotinamide ring is close to α -helix 3 in *E. coli* dIII (corresponding to α -helix B in *R. rubrum* dIII) (Figure 2). A number of hydrogen bonds stabilize NADP(H) in its binding site, involving, for example, Asp392, Lys424, Arg425, and Ser426. Loops D and E may indeed contribute to the binding and have been proposed to be involved in the Δp -dependent release of NADPH . Interactions between dI and dIII and their regulation by NADP(H) / NAD(H) have been investigated thoroughly by Karlsson and co-workers using NMR spectroscopy.

A crystal structure of a heterotrimer dI_2 –dIII complex from *R. rubrum* showed that the dI and dIII structures are essentially identical to those determined separately. However, the dI_2 –dIII

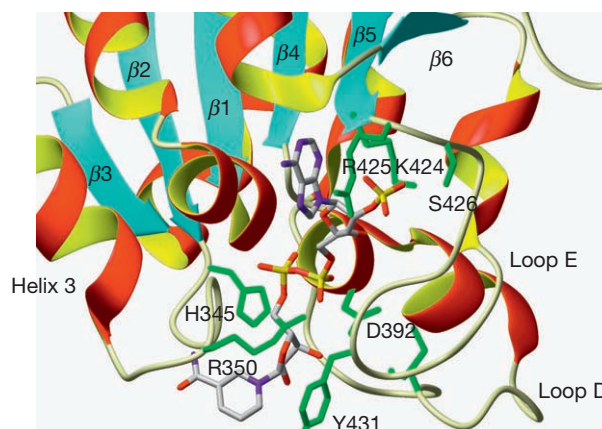


Figure 2 The NADP(H) -binding region of dIII. Amino acid residues and loops indicated are directly or indirectly essential in binding NADP(H) .

structure revealed that dIII interacts with the dI_2 (B), that is, the NAD(H) -binding site that contains disordered NAD^+ rather than well-ordered NAD^+ . This may be interpreted as indicating that when dIII (with bound NADP^+) is interacting with dI_2 , it binds the dI_2 (B) that has a disorganized NAD^+ , not the dI_2 (A) that has a well-ordered NAD^+ . However, a complicating fact is that, even though an NAD(H) molecule could be modeled in the dI_2 (B) site, the two nicotinamide rings are still separated by some 4–6 Å, making direct hydride transfer difficult. A possible mechanistic solution to this problem is that the nicotinamide ring of the modeled NAD(H) is

changed from an *anti* to a *syn* conformation, leading to a more open exposure of the nicotinamide ring in the cleft and a direct interaction with the nicotinamide ring of NADP⁺. Indeed, this rotation may be the very reason why NAD⁺ is more mobile in dI.2(B). An additional factor influencing the interaction between dI and dIII, and thus hydride transfer, is the amino acids that make up the interface between the domains. One of these, Gln132 in dI, appears to be essential for activity because it works as a tether during the catalytic cycle, holding both NAD(H) and NADP(H) together in the right conformation through H-bonds.

Mechanism of Hydride Transfer

Transfer of the hydride ion between the 4A-hydrogen (*pro-R*) of NADH and the 4B (*pro-S*)-hydrogen of NADP⁺, while the two substrates are bound to dI and dIII, respectively, occurs directly without any known intermediate and without exchange with bulk protons. Even though NAD⁺/NADH are rapidly exchanging with the NAD(H)-binding site in dI throughout the reaction cycle, NADH especially shows different affinities in the open (ordered and high affinity) and closed (disordered and low affinity) states. These affinity changes, and the associated *anti/syn* transitions of NAD(H), are apparently induced by subtle conformational changes in the dI dimer. However, several pieces of evidence together indicate that regulation of hydride transfer by, for example, an electrochemical proton gradient, Δp , occurs mainly at the level of dIII, rather than dI.

Essential information in this context was derived from the fact that separately expressed and purified dIII contains tightly bound NADP(H) which, due to a limiting release of the product NADP(H), catalyzes approximately two orders of magnitude slower reverse and forward reactions in the presence of purified dI than the wild-type enzyme. Isolated dIII also catalyzes a very fast so-called cyclic reaction, that is, reduction of 3-acetyl-pyridine-NAD⁺ (an NAD⁺ analog) by NADH, mediated by the bound NADP(H). This reaction involves reduction of bound NADP⁺ by NADH, release of the NAD⁺ formed, binding of 3-acetyl-pyridine-NAD⁺, reduction of 3-acetyl-pyridine-NAD⁺ by the bound NADPH, and release of 3-acetyl-pyridine-NADH. Indeed, it was reasonable to assume that the properties of the isolated dIII reflected an intermediate, the occluded state, in the reaction cycle of the intact TH. In reaction [1], this is interpreted to indicate that the activity of the forward reaction is limited by NADPH release, a limitation that is alleviated by Δp , allowing the enzyme to be catalytically fully active.

A general reaction scheme for forward reaction catalyzed by the *E. coli* TH may be outlined as follows (Figure 3). Starting with the resting state (state A), NADP⁺ binds to dIII in the β -subunit generating state B, in which the proton channel and the protonatable group X2 become accessible to the periplasmic space; in state C, the enzyme is protonated at X2, and the β -subunit subsequently changes conformation to state D. In state D, the proton moves from X2 to X1 concomitant with hydride transfer from NADH to NADP⁺, generating state E. States D and E correspond to the occluded state. Following a conformational change in dIII, an opening of dIII, and a less

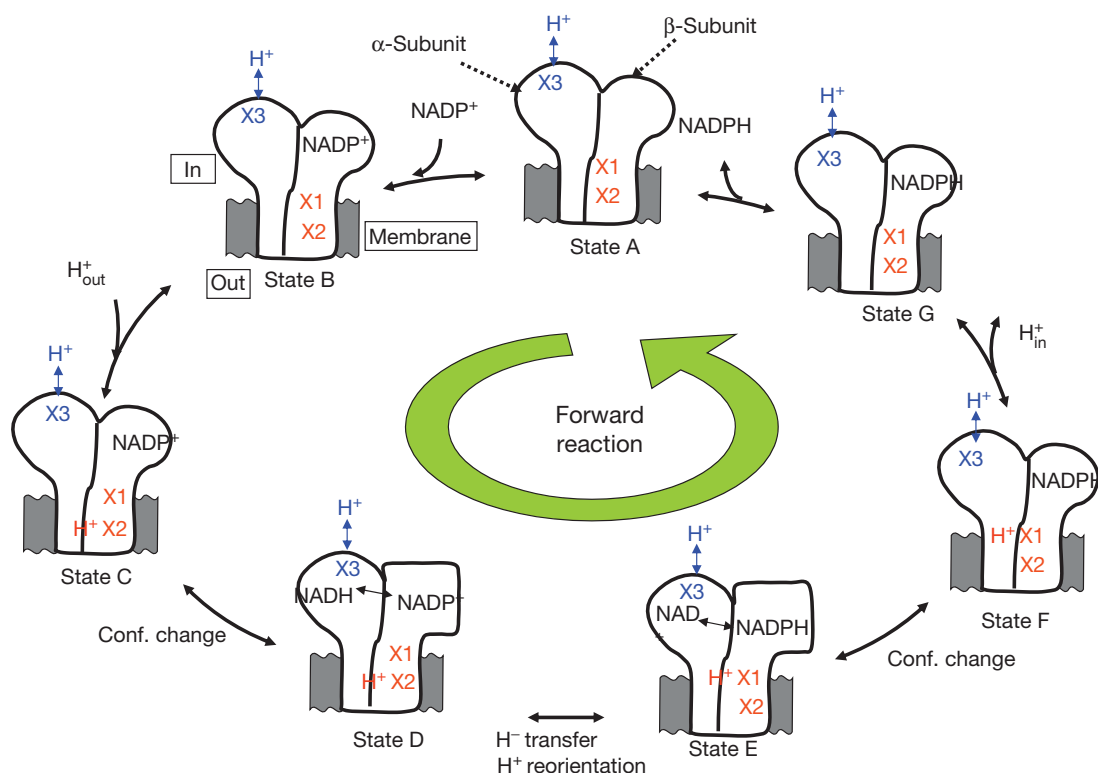


Figure 3 Reaction cycle of the forward TH reaction. X1 and X2 denote two protonatable residues in dII, possibly His91 and Asn222, respectively. X3 is a protonatable group in dI unrelated to the proton channel.

tight NAD(H) (state F), the proton dissociates on the cytosolic side generating state G. After dissociation of NADPH, the enzyme returns to the resting state A. During the entire cycle NAD(H) can freely bind to and dissociate from dI. X3 indicates a group on dI that may be regulated by the cytosolic pH.

As further discussed subsequently, at any given time, only half of the sites in the active dimeric enzyme (in the *E. coli* enzyme, $\alpha_2\beta_2$) seem to be available for, for example, NADP(H) binding and reaction with dicyclohexylcarbodiimide. In a recent modification of this principle, an alternating site mechanism has been proposed in which the two dIIIs alternate between an open and a closed (occluded) state.

Membrane Topology and Helix Packing of the Membrane Domain

dII of the mitochondrial TH was earlier predicted to be composed of 14 transmembrane α -helices, whereas dII of the *E. coli* TH has been shown to be composed of 13 transmembrane α -helices (Figure 4). However, in contrast to the mitochondrial enzyme, *E. coli* TH consists of two subunits, the α - and β -subunits, both of which contribute to dII. For sometime, the membrane topology of the *E. coli* enzyme was unclear, especially regarding the region in dII where the two subunits interact. Subsequently, it was shown that the α - and β -subunits can be fused at the gene level, generating a mitochondrial-like TH, provided that the peptide linker is long enough, that is, as long as the extra helix peptide in the mitochondrial TH; shorter linkers lead to truncated proteins with diminished activities. The C terminus of the α -subunit and the N terminus of the β -subunit are therefore located on different sides of the membrane (Figure 4).

Despite the obvious importance of dII in the overall TH reaction, the structure of this domain has subsequently received comparatively little attention. However, a systematic effort to elucidate the packing of dII of the *E. coli* enzyme using cysteine cross-linking of double cysteine mutants in

the cysteine-free background has been carried out by Rydström and co-workers, with one cysteine located in the α -subunit and the other at various positions in the β -subunit. The results show that, within the $\alpha + \beta$ unit, helices 2 and 4 are close to helix 6, and helix 3 and the C terminus of the α -subunit are close to helix 7. In the $\alpha_2\beta_2$, helices 2 and 4, as well as helix 6, are close to the same helices in the second subunit.

The Proton Channel

Being a proton pump, the membrane domain of THs must contain at least one proton channel, and considering the symmetry of the active $\alpha_2\beta_2$ structure, most likely two identical channels, one in each β -subunit. That the channel resides in the β -subunit is obvious due to the lack of suitably conserved protonatable residues in the α -subunit. However, residues in the α -subunit may contribute to the proton channel. dII of the *E. coli* TH contains few conserved protonatable amino acids, which may form part of a proton channel. Based on the properties of especially the positively charged H91K and N222R mutants (mimicking protonated His91 and Asn222, respectively), located in helix 9 and helix 13 (Figure 4), Bragg and co-workers made the important discovery that these mutants contain tightly bound NADP(H), reminiscent of isolated dIII. Indeed, this was the first piece of evidence indicating that dII not only communicates with dIII, but that the degree of protonation of specific residues, that is, His91 and Asn222, in dII also induces conformational changes that affect the dissociation constant for binding of NADP(H) to dIII. Ser139 is another potential component of the proton channel, with Glu85, Asp213, and Glu124 being possible peripheral residues of the channel, located close to the membrane surfaces.

Some mutants of the conserved His91, Asn222, and Ser139 residues still catalyze proton pumping although at reduced rates. This may be due to a compensatory effect of the remaining protonatable residues in the channel. In this context, it should be stressed that a proton channel *per se* does not require

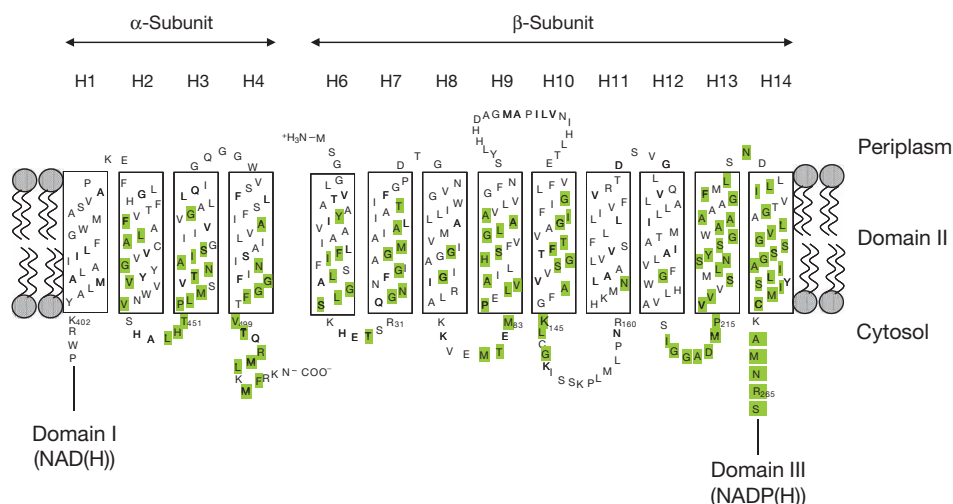


Figure 4 Membrane topology of *E. coli* TH. Residues in green background are at least 80% conserved among 61 THs. The topology was determined as described by Mueller J and Rydström J (1999) The membrane topology of proton-pumping *Escherichia coli* transhydrogenase determined by cysteine labeling. *Journal of Biological Chemistry* 274: 19072–19080.

protonatable residues. It would be sufficient with a chain of bound water molecules in a channel across the membrane to allow protons to jump from one water molecule to another. Protonation/deprotonation of protonatable amino acids in the channel induces conformational changes in the channel which link proton passage through the channel to catalytic events in often distantly located domains.

Coupling Mechanism

One of the main mechanistic challenges in past and present TH research is to understand how a Δp can be converted into a high $[NADPH][NAD^+]/[NADH][NADP^+]$ ratio and, conversely, how a high $[NADPH][NAD^+]/[NADH][NADP^+]$ (or a $[NADP^+][NADH]/[NADPH][NAD^+]$) ratio can drive proton pumping. Since the difference in standard redox potential between the NADPH/NADP⁺ and NADH/NAD⁺ pairs is negligible, the main contribution to the redox potential/free energy is derived from the above substrate/product ratios. Because of the water-soluble nature of the TH substrates, and the apparently large distance between the substrate-binding sites and the membrane, it has been generally accepted that some kind of conformational coupling is involved. Such a mechanism is often described as resembling that of adenosine triphosphatase (ATPase)/synthase in which conformationally dependent binding changes are essential. This was also experimentally demonstrated earlier by extensive Δp -dependent changes of K_m values for especially NAD(H). As described in the previous section, the single most important finding in this context was that mutating His91 and Asn222 to more positive amino acids leads to tightly bound NADP(H) in the intact mutant enzyme. An important contribution was also the demonstration of tightly bound NADP(H) in isolated dIII, which was assumed to represent an intermediate in the reaction cycle of TH, denoted the occluded state, which led to the proposal that both directions of the TH reaction, but especially the Δp -stimulated forward reaction, is limited by dissociation of NADP(H). The term 'binding change' mechanism is frequently used in order to stress the role of dissociation/binding of NADP(H).

Despite the known crystal structures of dI, dIII, and the dI–dIII complex, the interface between dII and dI + dIII is unknown, that is, little is known regarding the communication between dII and dIII. Most cytosolic loops of dII are poorly conserved, which makes it unlikely that dII–dIII signaling occurs via the surfaces of these domains. However, the peptide linking the C-terminal end of dII with dIII is quite conserved, and is highly mobile during catalysis as indicated by the NADP(H)-dependent increased trypsin sensitivity of Arg265 close to transmembrane helix 14 (Figure 4), indicating that this region is potentially important. Subsequently, Arg265 has been suggested to form a functional salt bridge with Asp213, constituting an essential component (hinge region) of a new coupling mechanism. However, communication between this proposed salt bridge and the NADP(H)-binding site remains unknown, although loops D and E close to the NADP(H)-binding site in dIII may be important. As already mentioned, proton-translocating THs, which are all dimers, may work according to an alternate-site-binding change mechanism. According to this mechanism, only half of the intact TH is catalytically active at any given time.

Physiological Role

In 1959, Klingenberg and Slenczka found that the mitochondrial redox level of NADP(H) was at least 90%, and that of NAD(H) ~50–60%. The forward TH reaction (reaction [1]), driven by Δp , is largely responsible for this high redox level of NADPH which, as expected, is sensitive to uncouplers. This also explains the fact that NADPH-dependent detoxification (through glutathione peroxidase, glutathione reductase, and reduced glutathione) of organic peroxides in intact mitochondria is more effective with NADH-linked substrates than with NADPH-linked substrates. Thus, endogenous NADP⁺ together with NADH (generated by the added substrate, e.g., β -hydroxybutyrate) provides a higher redox level of NADPH through TH than NADPH-linked substrates. NADPH also regulates the redox levels of thioredoxin and protein-bound thiols. Organic peroxides generated by, for example, oxidative stress, lead to a rapid oxidation of mitochondrial-reduced glutathione through the glutathione peroxidase system, which, in turn, leads to a massive release of Ca²⁺ ions, presumably through mitochondrial anion channels, for example, the inner mitochondrial anion channel (IMAC) and the permeability transition pore (MPTP). These channels may, in turn, contain redox-sensitive protein thiols. Therefore, the high mitochondrial redox level of NADP(H), maintained by TH, is of crucial importance for detoxification, Ca²⁺ homeostasis, cell thiol regulation, and biosynthesis. In liver, but probably not in heart, TH-generated NADPH in the mitochondrion contributes to cytosolic NADPH through transport in the form of (iso)citrate generated by the NADP-dependent isocitrate dehydrogenase (NADP-ICDH). That TH indeed regulates the redox level of the glutathione pool and, thus the defense against oxidative stress, was demonstrated using a knockout of TH in the nematode *C. elegans*.

A different hypothesis for the physiological role of TH involves a regulation of the citric acid cycle, especially the NAD and the NADP-dependent isocitrate dehydrogenases. The NAD enzyme (NAD-ICDH) is irreversible and strongly regulated by, for example, adenosine diphosphate, whereas NADP-ICDH is reversible and apparently unregulated. Oxidative decarboxylation of isocitrate to α -ketoglutarate, carbon dioxide, and NADH by the NAD-ICDH can be linked to the NADP-ICDH by a reductive carboxylation of α -ketoglutarate, carbon dioxide, and NADPH back to isocitrate. The NADH generated by the NAD-ICDH, together with NADP⁺, is thus used by TH and Δp to generate NADPH that drives the NADP-ICDH. This nonproductive cycle uses one proton/cycle to slow down or regulate the flux through this step in the citric acid cycle. However, there is no obligatory simultaneous occurrence of TH and NADP-isocitrate dehydrogenase in some species, which casts some doubt on this potentially important regulatory system.

Theoretically, a reversal of reaction [1] can also be used to generate a Δp under, for example, ischemic conditions. However, it is generally believed that unless a continuous source of NADP⁺-reducing substrates as well as NAD⁺-oxidizing substrates are available, the Δp generated would be transient and short lived.

More recently, the C57BL/6J mouse strain has been shown to be a natural/spontaneous knockout of the TH gene (nicotinamide nucleotide transhydrogenase (NNT)). These mice

show a number of interesting properties such as glucose intolerance, that is, a decreased insulin production in response to a glucose load, reminiscent of a diabetes-like condition, a decreased cancer frequency, decreased resistance to oxidative stress, and obesity. Knockout of the mitochondrial superoxide dismutase (SOD2) in the C57BL/6J background results in mice with severely decreased heart function and a life span of only 1–2 days.

See also: [Bioenergetics: Glutathione Peroxidases.](#)

Further Reading

- Althage M, Bizouarn T, Kindlund B, et al. (2004) Cross-linking of transmembrane helices in proton-translocating nicotinamide nucleotide transhydrogenase from *Escherichia coli*: Implications for the structure and function of the membrane domain. *Biochimica et Biophysica Acta* 1659: 73–82.
- Aon MA, Cortassa S, Maack C, and O'Rourke B (2007) Sequential opening of mitochondrial ion channels as a function of glutathione redox thiol status. *Journal of Biological Chemistry* 282: 21889–21900.
- Arklblad EL, Tuck S, Pestov NB, et al. (2005) A *Caenorhabditis elegans* mutant lacking functional nicotinamide nucleotide transhydrogenase displays increased sensitivity to oxidative stress. *Free Radical Biology and Medicine* 38: 1518–1525.
- Bergkvist A, Johansson C, Johansson T, Rydstrom J, and Karlsson BG (2000) Interactions of the NADP(H)-binding domain III of proton-translocating transhydrogenase from *Escherichia coli* with NADP(H) and the NAD(H)-binding domain I studied by NMR and site-directed mutagenesis. *Biochemistry* 39: 12595–12605.
- Bizouarn T, Althage M, Pedersen A, et al. (2002) The organization of the membrane domain and its interaction with the NADP(H)-binding site in proton-translocating transhydrogenase from *E. coli*. *Biochimica et Biophysica Acta* 1555: 122–127.
- Bizouarn T, Fjellström O, Meuller J, et al. (2000) Proton translocating nicotinamide nucleotide transhydrogenase from *E. coli*. Mechanism of action deduced from its structural and catalytic properties. *Biochimica et Biophysica Acta* 1457: 211–228.
- Bragg PD and Hou C (1999) *Escherichia coli*. *Biochimica et Biophysica Acta* 363: 182–190.
- Buckley PA, Jackson JB, Schneider T, et al. (2000) Protein–protein recognition, hydride transfer and proton pumping in the transhydrogenase complex. *Structure* 8: 809–815.
- Cotton NP, White SA, Peake SJ, McSweeney S, and Jackson JB (2001) The crystal structure of an asymmetric complex of the two nucleotide binding components of proton-translocating transhydrogenase. *Structure* 9: 165–176.
- Glavas NA and Bragg PD (1995) The mechanism of hydride transfer between NADH and 3-acetylpyridine adenine dinucleotide by the pyridine nucleotide transhydrogenase *Escherichia coli*. *Biochimica et Biophysica Acta* 1231: 297–303.
- Hatefi Y and Yamaguchi M (1996) Nicotinamide nucleotide transhydrogenase: A model for utilization of substrate binding energy for proton translocation. *FASEB Journal* 10: 444–452.
- Hoek JB and Rydstrom J (1988) Physiological roles of nicotinamide nucleotide transhydrogenase. *Biochemical Journal* 254: 1–10.
- Jackson JB, White SA, Quirk PG, and Venning JD (2002) The alternating site, binding change mechanism for proton translocation by transhydrogenase. *Biochemistry* 41: 4173–4185.
- Johansson T, Oswald C, Pedersen A, et al. (2005) X-ray structure of domain I of the proton-pumping membrane protein transhydrogenase from *Escherichia coli*. *Journal of Molecular Biology* 352: 299–312.
- Mather OC, Singh A, van Boxel GI, White SA, and Jackson JB (2004) Active-site conformational changes associated with hydride transfer in proton-translocating transhydrogenase. *Biochemistry* 43: 10952–10964.
- Meuller J and Rydstrom J (1999) The membrane topology of proton-pumping *Escherichia coli* transhydrogenase determined by cysteine labeling. *Journal of Biological Chemistry* 274: 19072–19080.
- Prasad GS, Wahlberg M, Sridhar V, et al. (2002) Crystal structures of transhydrogenase domain I with and without bound NADH. *Biochemistry* 41: 12745–12754.
- Rydstrom J (2006) Mitochondrial NADPH, transhydrogenase and disease. *Biochimica et Biophysica Acta* 1757: 721–726.
- Sundaresan V, Yamaguchi M, Chartron J, and Stout CD (2003) Conformational change in the NADP(H) binding domain of transhydrogenase defines four states. *Biochemistry* 42: 12143–12153.

Nitric Oxide Signaling

M A Marletta, University of California, Berkeley, CA, USA

Published by Elsevier Inc.

This article is reproduced from the previous edition, volume 3, pp. 62–65, © 2004, Elsevier Inc.

Glossary

Domain A compact region of protein structure that is a stable entity.

Heterodimer A protein composed of two nonidentical subunits.

Kinase An enzyme that transfers a phosphoryl group from a donor (typically adenosine triphosphate) to

an acceptor (often a protein in cell signaling pathways).

Neurotransmitter A chemical messenger released or synthesized at a synapse.

Phosphatase An enzyme that removes a phosphoryl group.

Radical A chemical species with an unpaired electron.

The basic components of nitric oxide (NO) signaling involve nitric oxide synthase (NOS) to synthesize NO and the soluble isoform of guanylate cyclase (sGC) to trap NO. Once activated by NO, sGC converts guanosine triphosphate (GTP) to guanosine 3'/5'-cyclic monophosphate (cGMP), leading to cGMP-dependent physiological responses such as vasodilation. Termination of the cGMP signal involves a cyclic nucleotide phosphodiesterase (PDE). There are multiple PDE isoforms, and selective inhibition of specific isoforms has proven to be clinically useful (e.g. sildenafil (viagra) in the treatment of erectile dysfunction). sGC must also be turned off; however, the *in vivo* mechanism for this remains unknown. Catalytic activity of constitutive NOS isoforms involved in signaling is controlled by Ca^{2+} and calmodulin (CaM) via receptor-mediated events that trigger Ca^{2+} release from intracellular stores or uptake from external sources.

Biological Processes Controlled by NO

First defined as endothelium-derived relaxing factor, the biological effects of NO have now been definitively characterized in a number of tissues. Smooth muscle is often the target of NO action, leading to vasodilation in blood vessels (Figure 1) and penile erection via NO synthesis in the corpus cavernosum, two well-established functions. Smooth muscle relaxation in the gut is another important function for NO, playing a key role in gastrointestinal tract motility. NO synthesis in the brain is also well established, though the exact role in learning and memory is still under debate. In the central nervous system (CNS), NO acts as a somewhat atypical neurotransmitter in that it is not synthesized and stored and then released with the appropriate signal, but is synthesized on demand. Constitutive NOSs are involved in all of these responses and are under the control of Ca^{2+} and CaM. Receptor-mediated events that trigger Ca^{2+} release or uptake turn on the specific NOS isoform in the tissue. NOS has also been found in skeletal muscle, in the myocardium, and in many other cells, suggesting that it will play an even more extensive role in physiological control than is currently established. The function of NO in the myocardium is particularly complex, although it appears that novel therapeutics will be an outcome of sorting out the complexity.

NO Synthase

The catalytic activity of the NOS isoforms (constitutive) involved in signaling is controlled by Ca^{2+} and CaM. After stimulation by the appropriate external signal, an increase in intracellular-free Ca^{2+} occurs, which then leads to a Ca^{2+} –CaM complex. Each of the four EF hands of CaM binds one Ca^{2+} , and then this Ca^{2+} –CaM complex binds to NOS, thereby activating NOS to synthesize NO. The NOS reaction is shown in Figure 2. The enzyme converts L-arginine to citrulline and NO. Co-substrates for the reaction include nicotinamide adenine dinucleotide phosphate-oxidase (NADPH) and O_2 . NOS is composed of two domains: a reductase domain with homology to cytochrome P450 reductase and a heme domain. The reductase domain contains two bound flavins, flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN), and with analogy to P450 reductase, the electron transfer route is from NADPH to FAD and then to FMN. The heme domain contains a cytochrome P450-type heme (ferric protoporphyrin IX coordinated to cysteine thiol), one equivalent of 6(R)-tetrahydro-L-biopterin (H_4B), and the L-arginine binding site. Crystal structures of the heme domain show it to be dimeric with a tetra-thiolate zinc center at the dimeric interface using two cysteines from each subunit. Catalytic chemistry at the heme domain is dependent on electron transfer from the flavins in the reductase domain, which is triggered by binding of the Ca^{2+} –CaM complex to NOS.

The catalytic chemistry takes place at the P450 heme site; in fact, structures of the oxygenase domain that have been solved in the presence of L-arginine show the substrate positioned over the heme. H_4B also participates in electron transfer at the heme site, using unprecedented one-electron chemistry. The exact details of the involvement of this required bound cofactor are still emerging. As shown in Figure 2, L-arginine is first converted to N^G-hydroxy-L-arginine (NHA) and then NHA is converted to citrulline and NO. NADPH and O_2 are required in both steps of the reactions, as is H_4B . The mechanism of the reaction is complicated and not fully understood, but clearly involves oxidative chemistry at the P450 heme with participation of H_4B .

The Ca^{2+} –CaM complex exerts control over the NOS reaction catalyzed by the constitutive cell-signaling isoforms,

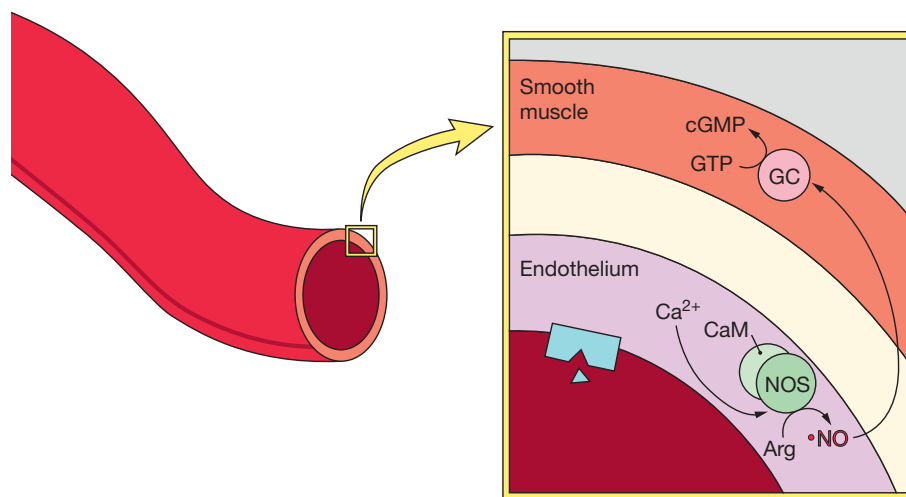


Figure 1 Nitric oxide function in blood vessel dilation. In response to a receptor-mediated event (shown schematically), endothelial NOS is activated by Ca^{2+} and CaM, leading to the synthesis of NO. NO then diffuses into the adjacent smooth muscle and activates the soluble isoform of guanylate cyclase and the synthesis of cGMP. cGMP, through a signaling cascade, then leads to smooth muscle relaxation.

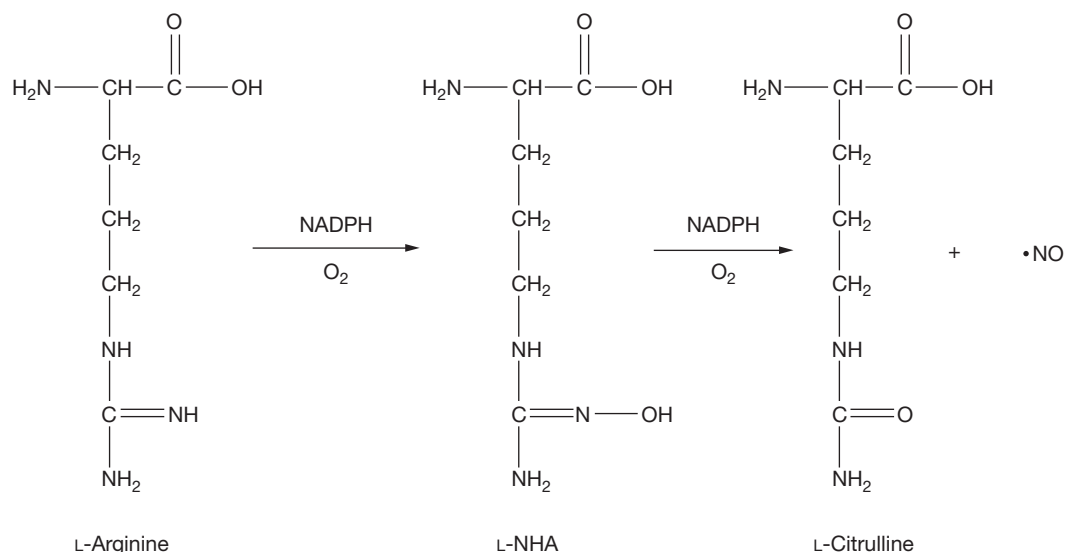


Figure 2 The reaction catalyzed by nitric oxide synthase. The substrate for the reaction is the amino acid L-arginine, which is converted to L-NHA and then to citrulline and $\bullet\text{NO}$. Although it has never been shown, it is assumed that the stereochemistry of citrulline at the α -carbon remains unchanged.

endothelial NOS and neuronal NOS, by controlling electron transfer from the reductase domain to heme domain. Without this required process, the reaction does not take place, therefore, Ca^{2+} -CaM exerts complete control over the reaction. NO is toxic, and so it is absolutely crucial in NO signaling to strictly control the amount of NO that is made. Tight regulation over Ca^{2+} concentrations in the cell provides this required level of control. In general, nM concentrations of NO are synthesized by the generator cell, and pM concentrations act in the target cell. inducible NOS (iNOS) was first described as a Ca^{2+} -CaM-independent isoform, but subsequent studies showed that Ca^{2+} -CaM copurified with this isoform. It is, therefore, constitutively active. Because the function of iNOS is in the host response to infection, a NOS isoform that is constitutively

active and produced at the site of infection conforms to this biological role.

A number of inhibitors of NOS have been developed. The initial molecules were simple derivatives of the substrate L-arginine; however, other nonarginine-related structures were later found that were derived largely from library screens in the pharmaceutical industry, and some demonstrate significant isoform selectivity.

Chemistry of NO

NO is unstable in aqueous, aerobic solution. The stable end products of NO decomposition are nitrite (NO_2^-) and nitrate

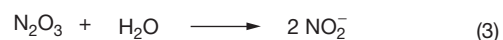


Figure 3 Relevant solution chemistry of nitric oxide. NO reacts with O_2 in aqueous solution to produce nitrogen dioxide ($\bullet\text{NO}_2$) (eqn [1]). $\bullet\text{NO}_2$ then reacts with another molecule of $\bullet\text{NO}$ to produce dinitrogen trioxide (N_2O_3) (eqn [2]). N_2O_3 can then hydrolyze to yield NO_2^- (eqn [3]) or react with protein or low-molecular-weight thiols to make the S-nitroso derivative (eqn [4]). The S-nitrosated protein or small molecule is capable of homolytically cleaving to generate the thiol radical and $\bullet\text{NO}$ (eqn [5]).

(NO_2^-). The reaction sequence that leads to these products is shown in **Figure 3**. This solution chemistry consumes NO as it traverses a path from the generator cell to the target cell, hence the amount that arrives at the target cell is always lower. An intermediate along the reaction pathway is dinitrogen trioxide (N_2O_3). N_2O_3 can react with water as shown to produce NO_2^- or react with other cellular nucleophiles. The reaction with cysteine residues in proteins or with small cellular sulfur nucleophiles such as glutathione leads to the formation of S-nitrosated species as shown in **Figure 3**. S-nitrosated proteins and small molecules are known to decompose to the sulfur radical and NO in a so-called homolytic cleavage (**Figure 3**); certain biological conditions can accelerate the reaction. The significance of this chemistry is discussed below. NO has 11 valence electrons and, therefore, has one unpaired electron. The radical nature of NO dictates other chemistry, such as the facile reaction of NO with superoxide to give peroxynitrite ($^-\text{OONO}$), which then converts very quickly to NO_3^- , and the reaction with oxyhemoglobin to give methemoglobin and NO_3^- , again with the intermediacy of $^-\text{OONO}$. Peroxynitrite is a potent oxidant and has been shown to be important in various types of tissue damage associated with the formation of high levels of NO. The decomposition of relatively high concentrations of NO in tissue that occurs under inflammatory conditions leads to nitration chemistry, most frequently seen on tyrosine residues.

Guanylate Cyclase

The soluble isoform of sGC is the receptor for NO in target cells. Activation of sGC leads to the formation of cGMP, and then through a cascade involving cGMP-dependent kinases, a lowering of intracellular Ca^{2+} occurs, leading to effects such as smooth muscle relaxation (e.g., vasodilation). sGC is a heterodimeric protein composed of two subunits. The best characterized is the physiologically occurring $\alpha 1\beta 1$ heterodimer, first isolated from lung tissue. The β -subunit contains a unique

ferrous heme cofactor that is the receptor for NO. This porphyrin, despite having the same ligation as the globins, does not bind oxygen; rendering it ideally suited to bind NO is the presence of a high-competing concentration of oxygen. Upon binding of NO, sGC is activated 300-fold over the basal activity, and a burst of cGMP is produced. cGMP then acts through cGMP-dependent kinases to bring about physiological responses such as vasodilation and other smooth muscle relaxation responses.

Tools to manipulate sGC are emerging and include 1H-[1,2,4]oxadiazole[4,3-a]quinoxalin-1-one (ODQ), which inhibits sGC by oxidation of ferrous porphyrin to the ferric oxidation, and a family of activators based on the structure of YC-1 (3-(5'-hydroxymethyl-2'-furyl)-1-benzyl indazole). This compound is a weak activator of sGC; however, YC-1 and CO activate sGC to the same level as NO, and YC-1 acts synergistically with NO. A class of activators based on the YC-1 structure that directly activate and are more potent than YC-1 are entering clinical trials.

Non-cGMP NO-Signaling Pathways

As mentioned previously, N_2O_3 is an intermediate that forms during the aerobic solution decomposition of NO. N_2O_3 is a potent nitrosating agent that reacts with the ω -amino group of lysine and the thiol of cysteine. Formation of S-nitrosated cysteines in proteins can cause a change in the biological activity of the protein. Although increasing studies are directed toward understanding the chemistry and biological activity of the S-NO bond and the formation of these bonds with small molecule thiols and protein thiols, critical aspects remain to be determined. For example, the formation of R-S-NO could occur via the solution reaction of N_2O_3 with the thiol or by an enzyme-catalyzed process. The former path would appear to be too inefficient, and an enzyme to catalyze the latter path has not been isolated. Other hypotheses advanced involve the reversible formation of R-S-NO, where again the R group could be either a low-molecular-weight species or a protein similar to phosphoryl transfer. The kinase and phosphatase equivalents are yet to be isolated. Confounding all these hypotheses is the chemical reaction of R-SH with N_2O_3 , which would be facilitated much beyond what would be expected to occur when nonbiological concentrations of NO are used.

There is another important outcome from the formation of S-nitrosated proteins and low molecular weight compounds like glutathione. The S-nitroso bond is known to homolytically fragment, forming NO and a thiyl radical. The NO so derived can then function to activate sGC as described above. This mode of generating NO is attractive in that it represents a semi-stable form of this signaling agent and could be a way to store, transport, and deliver NO. There is no question that the decomposition of NO in aerobic aqueous solution will lead to S-nitrosation; rather, the issue is whether the amounts formed *in vivo* are physiologically significant.

See also: **Bioenergetics: Calcium Signaling: NO Synthase; Heme Synthesis; Protein/Enzyme Structure Function and**

Degradation: Flavins; Heme Proteins; **Signaling:** Neurotransmitter Transporters.

Further Reading

Bredt DS and Snyder SH (1994) Nitric oxide: A physiologic messenger molecule. *Annual Review of Biochemistry* 63: 175–195.

Denninger JW and Marletta MA (1999) Guanylate cyclase and the NO/cGMP signaling pathway. *Biochimica et Biophysica Acta* 1411: 334–350.

Marletta MA (1994) Nitric oxide synthase: Aspects concerning structure and catalysis. *Cell* 78: 927–930.

Marletta (2001) Another activation switch for endothelial nitric oxide synthase: Why does it have to be so complicated? *Trends in Biochemical Sciences* 26: 519–521.

Moncada S, Palmer RM, and Higgs EA (1991) Nitric oxide: Physiology, pathophysiology, and pharmacology. *Pharmacological Reviews* 43: 109–142.

N-Linked Glycan-Processing Glucosidases and Mannosidases

T Suzuki, RIKEN Advanced Science Institute, Wako, Japan

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Linda O. Tremblay and Annette Herscovics, volume 3, pp. 46–49, © 2004, Elsevier Inc.

Glossary

Calnexin/calreticulin Endoplasmic reticulum (ER)-resident molecular chaperones/lectins that specifically bind to monoglucosylated *N*-glycans.

Cargo receptors Receptors for cargo proteins concentrated in budding vesicles. ERGIC-53, for example, is believed to be a cargo receptor for certain glycoproteins.

EDEM A protein that has a homology with ER α -mannosidase I, which is involved in glycoprotein-specific ER-associated degradation (ERAD). Recent studies have suggested that EDEM protein indeed retains α -mannosidase activity, which appears to act preferably on the C-arm of α 1,2-linked mannose.

ER-associated degradation (ERAD) If proteins newly synthesized in the ER cannot fold properly, such misfolded proteins are retrotranslocated from the ER into the cytosol, and finally they will be degraded by the proteasome in the cytosol. This process is well conserved from yeast to humans.

ER quality-control compartment (ERQC) An ER-derived compartment, which could segregate misfolded proteins in a microtubule-dependent fashion, thereby preventing them from interfering with ER functions. Some glycosidases (e.g., ER α -mannosidase I) or lectins (e.g., calnexin/calreticulin) are reported to be concentrated in the ERQC.

Malectin ER-resident lectins found in mammalian cells, it has a unique carbohydrate-binding specificity towards diglucosylated *N*-glycans.

OS-9/XTP3-B Putative degradation lectins involved in the recognition of misfolded glycoproteins in the ER. They are homologs of yeast protein, Yos9, and contain a mannose 6-phosphate receptor homology (MRH) domain.

UGGT A glucosyltransferase situated in the lumen of the ER. It only acts on glycoprotein substrates that are particularly unfolded, thereby serving as a 'folding sensor' in the ER quality-control system.

The early processing of *N*-glycans, occurring in the endoplasmic reticulum (ER) and early Golgi, is a relatively conserved event among eukaryotes. In mammalian cells, multiple α -glucosidases/ α -mannosidases are involved in this process. Recent evidence also suggests that these processes determine the fate of carrier proteins that go through the secretory pathway. In such a 'glycan-dependent' protein quality-control system, the subtle difference of structures can be dictated by carbohydrate-binding proteins (lectins) to regulate their folding, secretion, and/or degradation.

α -Glucosidases

Eukaryotic cells synthesize *N*-linked glycans by so-called '*en bloc*' transfer of, in most cases, the Glc₃Man₉GlcNAc₂ oligosaccharides (Figure 1) built on dolichol pyrophosphate onto the Asn-Xaa-Ser/Thr site (where Xaa can be any amino acid except Pro) of the nascent polypeptide in the ER lumen by oligosaccharyl transferases (OSTs).

After the transfer of glycans onto proteins, the first glycan-processing event that occurs is the trimming of the glucoses by ER α -glucosidases (Figure 2).

α -Glucosidase I is a type II membrane protein that removes the outermost α 1,2-linked glucose residue from a glycoprotein (residue a in Figure 1). A defect in the α -glucosidase I gene in humans, named congenital disorder of glycosylation (CDG) type IIb, has been reported in a neonate with various severe phenotypes, indicating that the removal of the terminal glucose is essential for proper human development.

α -Glucosidase II, on the other hand, is a soluble luminal enzyme acting on the inner α 1,3-linked glucoses (residues b

and c in Figure 1). This enzyme is composed of two subunits, α and β ; while the α subunit contains a catalytic activity, the β chain contains a C-terminal KDEL (Lys-Asp-Glu-Leu), which serves as an ER-retrieval signal, as well as a putative lectin domain (MRH domain; mannose-6-phosphate receptor homolog domain). Both subunits are essential for the full functionality of this enzyme.

Glucosidase inhibitors, such as castanospermine or 1-deoxynojirimycin, are often used to inhibit α -glucosidases I/II, while bromoconduritol is used as a selective inhibitor for α -glucosidase II.

Endo α -Mannosidase

Endo α -mannosidase is an enzyme present in the ER–Golgi intermediate compartment (ERGIC) and in the Golgi. The enzyme can convert Glc_{1–3}Man₉GlcNAc₂ to Man₈GlcNAc₂ (Figure 2), and therefore provides a backup processing pathway that can bypass the need for ER α -glucosidases I/II. This enzyme appeared relatively late in evolution and its occurrence is restricted to the chordate phylum, suggesting that this bypass route emerges to ensure the conversion of the high-mannose type to complex-type glycans in these animals.

Processing α -Mannosidases

After deglucosylation by ER α -glucosidases I/II, the next step in *N*-linked glycan processing involves the removal of four α 1,2-linked mannosidases to give rise to Man₅GlcNAc₂ (Figure 2). Processing α -mannosidases have been classified

EDEM, the third class of class I α -mannosidases

A protein called 'EDEM' (ER degradation enhancing α -mannosidase-like protein) is the third member of class I α -mannosidases. The Htm1/Mnl1 protein (an EDEM homolog) in yeast as well as EDEM proteins (EDEM1–3) in mammalian cells appear to be luminal, soluble proteins in the ER, except that the topology of EDEM1 protein is still controversial as it has been initially described as a type II transmembrane protein in COS cells. These results could be explained by variations in signal peptide cleavage in different cell lines. EDEM1 (but not the Htm1/Mnl1 protein in yeast) was originally discovered as an unfolded protein response (UPR)-inducible gene, implying its role in tolerance against ER stress. EDEM2/3 are subsequently identified by a sequence homology search with EDEM1, and their involvement in ER-associated degradation process has been assessed. Now, expression of all EDEM proteins is shown to be induced upon activation of the ER stress pathway.

With respect to the enzyme activity of EDEM proteins, several lines of evidence suggest the presence of α 1,2-mannosidase activity, which prefers the cleavage of branch C (residue i in Figure 1) in EDEM1/3 proteins as well as in Htm1/Mnl1, despite the lack of evidence for enzyme activity in purified proteins. The demannosylation of the C branch by EDEM proteins is suggested to play an important role in glycan-dependent ER-associated degradation machinery, as the exposed mannose residue appears to be recognized by degradation lectins, OS-9 and/or XTP-3B protein (see later).

Class II α -Mannosidases Involved in N-Linked Glycan Processing

Golgi α -mannosidases II

After the formation of the $\text{Man}_5\text{GlcNAc}_2$ structure by Golgi α -mannosidases I, a single GlcNAc residue is added by the action of uridine diphosphate-*N*-acetyl glucosamine (UDP-GlcNAc): α -3-D-mannoside β -1,2-*N*-acetylglucosaminyltransferase I (GnTI) (Figure 2). The $\text{GlcNAcMan}_5\text{GlcNAc}_2$ glycan is further demannosylated by class II Golgi α -mannosidases and Golgi α -mannosidase II/IIx (Man2A1 and Man2A2, respectively). Golgi α -mannosidase II is a type II membrane protein and does not require cations for enzyme activity. It was previously shown that Golgi α -mannosidase II physically associates with GnTI, suggesting that the protein–protein interaction plays a role in its Golgi localization. Mice lacking the *MAN2A1* gene develop a dyserythropoietic anemia concurrent with loss of erythrocyte complex-type *N*-glycans. This phenotype resembles human congenital dyserythropoietic anemia type II (HEMPAS, hereditary erythroblastic multinuclearity with a positive acidified-serum lysis test).

It was surprising that, in *MAN2A1*-null mice, some none-rythroid cells still produced complex-type *N*-glycans, suggesting the occurrence of an alternative processing pathway. The gene encoding α -mannosidase IIx (Man2A2) was found in the human genomic library by cross-hybridization using a complementary DNA (cDNA) for *MAN2A1* as a probe. This enzyme also is a type II membrane protein. Targeted disruption of *MAN2A2* in mice resulted in male infertility. This phenotype is believed to be due to the absence in the testis of a GlcNAc-terminated complex-type *N*-glycan required for adhesion between germ cells and Sertoli cells.

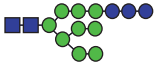
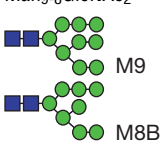
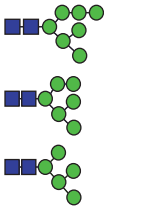
Recently, it has shown that both Golgi α -mannosidase II and IIx require prior action of GnTI. The double knockout (KO) mice of these enzymes were found to have much more severe phenotypes, that is, perinatal death, than either single KO mice. Structural analysis of glycans revealed that the double KO mice completely lack complex type *N*-glycans, implying that Golgi α -mannosidase IIx can serve as the sole alternative enzyme for Golgi α -mannosidase II.

Recognition Molecules (Lectins) for Distinct Glycan Structures

It has been shown that the processing of *N*-linked glycans can determine the fate of secretory glycoproteins in the eukaryotes. To achieve this purpose, there are many luminal lectins with different specificities in the ER/ER–Golgi intermediate compartments (Table 1).

Cells utilize *N*-glycans as a 'tag' to be recognized by various different lectins, which could result in several distinct consequences for the substrate glycoproteins, that is, retention in the ER, promotion of folding, exit from the ER for vesicular transport, or degradation. Therefore, the proper processing of

Table 1 Oligosaccharides and lectins/enzymes recognizing these structures

	Major lectins	Glycosidases/ glycosyltransferases
$\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ 		α -Glucosidase I
$\text{Glc}_2\text{Man}_9\text{GlcNAc}_2$ 	Malectin	α -Glucosidase II
$\text{GlcMan}_9\text{GlcNAc}_2$ 	Calnexin/ calreticulin	α -Glucosidase II
$\text{Man}_{9-8}\text{GlcNAc}_2$ 	Cargo receptors ^a (ERGIC-53; VIPL; VIP36)	UGGT (M9) ^b ER α -mannosidase I (M9) ^b EDEM (M8B) ^b Golgi α -mannosidases I (M8B) ^b
$\text{Man}_{7-5}\text{GlcNAc}_2$ 	OS-9/XTP3-B	Golgi α -mannosidases I ER α -mannosidase I in ER quality-control compartment? EDEM? GlcNAc transferase I (Man5)

^aCargo receptors generally have broader binding specificity and are reported to bind to glycans with more diverse structures.

^bPreferred (or predicted to be preferred) glycan structure.

N-linked glycans in the early secretory pathway by glycosidases is very critical for the quality control machinery of the secretory proteins.

Concluding Remarks

The N-linked glycan is one of the most conserved, commonly used modifications for secretory proteins. Therefore, it may not be surprising that the eukaryotic cells have evolved a highly sophisticated, N-glycan-dependent recognition system where the conditions of carrier proteins (e.g., folded vs. misfolded) are dictated by a subtle difference in glycan structure. In this regard, 'processing glycosidases' actually do more than processing, that is, creating signals for glycan-recognition systems. Complete understanding of this glycan-dependent quality control system still requires in-depth characterization of players, not only the glycosidases but also the lectins/glycosyltransferases involved in this process.

See also: **Cell Architecture and Function:** Endoplasmic Reticulum-Associated Protein Degradation; **Lipids Carbohydrates Membranes and Membrane Proteins:** Glycoprotein Folding and Processing Reactions; Glycoproteins, N-Linked.

Further Reading

- Aebi M and Helenius A (2004) Role of N-linked glycans in the endoplasmic reticulum. *Annual Review of Biochemistry* 73: 1019–1049.
- Aebi M, Bernasconi R, Clerc S, and Molinari M (2009) N-Glycan structures: Recognition and processing in the ER. *Trends in Biochemical Sciences* 35: 74–82.
- Akama TO, Nakagawa H, Wong NK, et al. (2006) Essential and mutually compensatory roles of α -mannosidase II and α -mannosidase IIx in N-glycan processing *in vivo* in mice. *Proceedings of the National Academy of Sciences of the United States of America* 103: 8983–8988.
- Clerc S, Hirsch C, Oggier DM, et al. (2009) Htm1 protein generates the N-glycan signal for glycoprotein degradation in the endoplasmic reticulum. *Journal of Cell Biology* 184: 159–172.
- Hebert DN, Bernasconi R, and Molinari M (2010) ERAD substrates: Which way out? *Seminars in Cell and Developmental Biology* 21: 526–532.
- Herscovics A (1999) Importance of glycosidase in mammalian glycoprotein biosynthesis. *Biochimica et Biophysica Acta – General Subjects* 1473: 96–107.
- Hosokawa N, Kamiya K, and Kato K (2010) The role of MRH domain-containing lectins in ERAD. *Glycobiology* 20: 651–660.
- Hosomi A, Tanabe K, Hirayama H, et al. (2010) Identification of Htm1(EDEM)-dependent, Mns1-independent ERAD pathway in *Saccharomyces cerevisiae*. Application of a novel assay for glycoprotein ERAD. *Journal of Biological Chemistry* 285: 24324–24334.
- Kanehara K, Kawaguchi S, and Ng DTW (2007) The EDEM and Yos9p families of lactic-like ERAD factors. *Seminars in Cell and Developmental Biology* 18: 743–750.
- Lederkremer GZ (2009) Glycoprotein folding, quality control and ER-associated degradation. *Current Opinion in Structural Biology* 19: 515–523.
- Molinari M (2007) N-Glycan structure dictates extension of protein folding or onset of disposal. *Nature Chemical Biology* 3: 313–320.
- Moremen KW (2002) Golgi α -mannosidase II deficiency in vertebrate systems: Implications for asparagines-linked oligosaccharide processing in mammals. *Biochimica et Biophysica Acta – General Subjects* 1573: 225–235.
- Olivari S and Molinari M (2007) Glycoprotein folding and the role of EDEM1, EDEM2 and EDEM3 in degradation of folding-defective glycoproteins. *FEBS Letters* 581: 3658–3664.
- Pearse BR and Hebert DN (2010) Lectin chaperones help direct the maturation of glycoproteins in the endoplasmic reticulum. *Biochimica et Biophysica Acta – Molecular Cell Research* 1803: 684–693.
- Quan EM, Kamiya Y, Kamiya D, et al. (2008) Defining the glycan destruction signal for endoplasmic reticulum-associated degradation. *Molecular Cell* 32: 870–877.
- Satoh T, Chen Y, Hu D, et al. (2010) Structural basis for oligosaccharide recognition of misfolded glycoproteins by OS-9 in ER-associated degradation. *Molecular Cell* 40: 905–916.
- Schachter H (2007) Processing of protein-bound N-glycans. In: Kamerling JP, Boons G-J, Lee YC, Suzuki A, Taniguchi N, and Voragen AGJ (eds.) *Comparative Glycoscience. From Chemistry to Systems Biology*, vol. 3, pp. 11–32. Oxford: Elsevier.
- Schallus T, Jaeckh C, Feher K, et al. (2008) Malectin: A novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein N-glycosylation. *Molecular Biology of the Cell* 19: 3404–3414.
- Suzuki T, Tanabe K, and Funakoshi Y (2007) Folding and quality control of glycoproteins. In: Kamerling JP, Boons G-J, Lee YC, Suzuki A, Taniguchi N, and Voragen AGJ (eds.) *Comparative Glycoscience. From Chemistry to Systems Biology*, vol. 3, pp. 129–149. Oxford: Elsevier.
- Yoshida Y and Tanaka K (2010) Lectin-like ERAD players in ER and cytosol. *Biochimica et Biophysica Acta – General Subjects* 1800: 172–180.

Nonhexameric SF1 DNA Helicases/Translocases

E Antony and T M Lohman, Washington University School of Medicine, St. Louis, MO, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Timothy M. Lohman, volume 1, pp. 618–623, © 2004, Elsevier Inc.

Glossary

Helicase An enzyme that utilizes nucleoside 5'-triphosphates to separate the two strands of the DNA double helix in order to form the single-stranded DNA (ssDNA) intermediates that are required for DNA metabolism (i.e., replication, recombination, and repair).

Processivity A measure of the average number of base pairs that can be unwound by a helicase before it dissociates from the DNA.

Translocase An enzyme that moves with either a 3'–5' or 5'–3' directionality along ssDNA using adenosine triphosphate (ATP) hydrolysis.

Functions

Helicases were originally identified as proteins capable of separating duplex DNA during replication, but have since been shown to function in a plethora of cellular responses, including displacement of proteins from DNA and RNA, chromatin remodeling, resolving Holliday junctions during recombination, and other nucleic acid conformational alterations. Unidirectional movement along single-stranded DNA (ssDNA) is central to the activity of SF1 helicases. For many SF1 enzymes, DNA unwinding and ssDNA translocation are separable activities due to their mechanism of action: a monomeric form of the enzyme is able to translocate along ssDNA, whereas an oligomeric form or multiple enzyme molecules are needed to observe DNA helicase activity.

The *Escherichia coli* Rep helicase is involved in replication, most likely playing a role in replication restart. Rep was first identified as being required for replication of bacteriophage ϕ X174. Rep in complex with the phage ϕ X174 gene A protein is highly processive, being able to unwind the entire genome (>6000 bp). However, in the absence of this accessory protein, Rep unwinds DNA with significantly lower processivity. *E. coli* UvrD, also known as helicase II, is involved in nucleotide excision repair and methyl-directed mismatch repair of DNA, as well as in plasmid replication. In the absence of accessory proteins, the unwinding processivity of UvrD is also relatively low. However, UvrD must be able to unwind at least 1000 bp processively during methyl-directed mismatch repair. The *Bacillus stearothermophilus* PcrA helicase is structurally similar to UvrD and Rep, and functions in DNA replication in complex with the replication initiator protein RepD which enhances its processivity and DNA unwinding activity. Therefore, the processivity of each of these helicases can be increased through interactions with accessory proteins.

Most DNA helicases show a preference for unwinding duplex DNA possessing an ssDNA flanking region or tail *in vitro*, and generally display a defined directionality of unwinding with respect to the backbone polarity of the ssDNA tail that flanks the duplex DNA. Helicases that initiate unwinding more efficiently on DNA substrates with a 3'-ssDNA tail are referred to as '3'–5' helicases', whereas those that prefer DNA substrates possessing a 5'-ssDNA tail are referred to as '5'–3' helicases'.

The *E. coli* Rep, UvrD, and *B. stearothermophilus* PcrA helicases are 3'–5' helicases, whereas the phage T4 Dda protein is a 5'–3' helicase. The *E. coli* RecBCD helicase is a unique bipolar helicase possessing two SF1 helicase subunits which move with opposite directionalities along ssDNA, with RecB translocating 3'–5' and RecD translocating 5'–3'. These two motors move in the same net direction since each moves on one of the two complementary strands of the DNA double helix. It has now been demonstrated that monomers of each of these enzymes contain all that is necessary for unidirectional translocation along ssDNA, although the mechanistic basis for what governs directionality is not known. At high-protein concentrations, *E. coli* UvrD can also initiate DNA unwinding at a nick *in vitro*, which is the biologically important site for initiation of unwinding in its roles in methyl-directed mismatch repair and excision repair.

Helicases are allosteric enzymes, many of which function as oligomeric assemblies. The class of hexameric DNA helicases exemplifies such oligomerization. The general importance of oligomerization for the function of SF1 helicases is less clear and is the subject of current study. Although there is evidence that a monomer of the phage T4 Dda helicase, an SF1 helicase, displays limited helicase activity *in vitro*, monomers of Rep UvrD and PcrA do not display helicase activity *in vitro*, but must oligomerize in order to unwind DNA *in vitro*. The role of this oligomerization may be to activate the enzyme and to provide the functional enzyme with multiple DNA-binding sites that are important for increasing the processivity of DNA unwinding.

Structural Features of SF1 DNA Helicases

Primary Structures

Helicases have been classified into six families and superfamilies (SF1–SFVI) based on conserved amino acid sequence patterns. The SF1 and SFII helicases function as monomers or dimers, whereas the SFIII–SFVI members are predominantly hexamers. The SF1 superfamily (containing Rep, UvrD, and PcrA) is defined by seven conserved regions of primary structure, referred to as 'helicase motifs'. The only regions of sequence similarity that are shared uniformly among all of the helicase families are

motifs I and II, which correspond to the Walker A and B boxes that form part of the nucleoside-5'-triphosphate-binding site. The presence of these two motifs is necessary but not sufficient for a protein to have helicase activity. The remaining conserved regions of the SF1 helicases are involved in binding of nucleoside-5'-triphosphate and/or DNA.

Crystal Structures of SF1 Helicases

Crystal structures are available for several SF1 DNA helicases, including *B. stearotherophilus* PcrA, *E. coli* Rep, and *E. coli* UvrD (Figure 1). The apo form of PcrA as well as in complex with adenosine diphosphate (ADP) was solved at 2.5 and 2.9 Å resolution, respectively. Structures of *E. coli* Rep bound to ssDNA and in complex with ssDNA and ADP were solved at 3.0 and 3.2 Å, respectively. Structures of UvrD in complex with a 3'-ss-dsDNA junction, and in the presence of ADP-MgF₃ or the nonhydrolyzable adenosine triphosphate (ATP) analog, AMPPNP, were solved at 3.0, 2.2, and 2.6 Å, respectively.

The Rep (673 amino acids), PcrA (675 amino acids), and UvrD (720 amino acids) proteins are structurally similar and are shown schematically in Figure 1. These SF1 proteins consist of two domains (1 and 2), separated by a cleft; domains 1 and 2 are further composed of subdomains 1A, 2A, 1B, and 2B (Figures 1 and 2). Subdomains 1A and 2A are composed of central β -sheets flanked by α -helical regions and are homologous to each other as well as to the nucleoside-5'-triphosphate-binding domain of the *E. coli* RecA protein. Subdomains 1A and 2A both contain large insertions which form the α -helical 1B and 2B domains.

Crystal structures of *E. coli* Rep, UvrD, and *B. stearotherophilus* PcrA in conjunction with mutagenesis studies on several helicases indicate that amino acid residues in the seven conserved SF1 helicase motifs, I–VI and IA are involved in either nucleotide and/or DNA binding. As shown in Figure 2, motifs I, IA, II, and III are contained within domain 1A, motifs V and VI are contained within domain 2A, whereas domain IV is found at the cleft formed between domains 1A and 2A. In general, ADP interacts with residues in motifs I and IV, whereas ATP interacts with residues within or near all of the motifs, with the apparent exception of motif IA.

In the crystal structures, ssDNA is bound with a defined orientation with respect to the polarity of the sugar-phosphate

backbone and the 3'–5' direction runs from domain 1 to domain 2 across the cleft between the two domains, with the 3' end of the ssDNA located in a groove formed between subdomains 1A and 1B. Several of the helicase motifs appear to be involved in the transmission of information between the ATP and ssDNA-binding sites within a monomer. Motif V residues are also likely to be involved in both ssDNA binding as well as ATP binding through interactions with the γ -phosphate. Motif VI interacts directly with the γ -phosphate of ATP, as well as with motif IV. The above summary considers only interactions that occur within a single monomer or subunit. However, since oligomerization of Rep, UvrD, and PcrA is functionally important for helicase activity *in vitro*, important allosteric interactions also might occur between two monomers.

Open and Closed Conformations

The asymmetric unit of the Rep-ssDNA crystal structure shows two molecules of Rep bound to a single molecule of dT(pT)₁₅. The conformations of the two Rep monomers are strikingly different with respect to the orientation of the 2B subdomain relative to the other three subdomains (Figures 1(a) and 1(b)). These two conformations, referred to as closed and open, respectively, differ by a 130° rotation of the 2B subdomain about a hinge region connecting it to the 2A subdomain. It has been speculated that the 2B subdomain might be part of the interface between subunits within the Rep dimer. The essential nature of the 2B subdomain in Rep was tested by making a *rep* gene construct in which the coding region for the entire 2B domain was deleted and replaced by three glycines. The resulting Rep Δ 2B protein retains helicase activity *in vitro* and can also support ϕ X174 phage replication *in vivo*, thus ruling out models invoking an essential role for the 2B domain in the helicase activity of Rep. In fact, a monomer of the Rep Δ 2B protein displays limited helicase activity *in vitro* in stark contrast to a wild-type Rep monomer which displays no helicase activity *in vitro*. This indicates that individual Rep monomers do possess all that is needed to unwind duplex DNA. However, the 2B subdomain appears to inhibit the helicase activity of the monomer and, thus, may have a regulatory function. This inhibition is relieved upon formation of a Rep or UvrD dimer. In the crystal structures of UvrD and PcrA

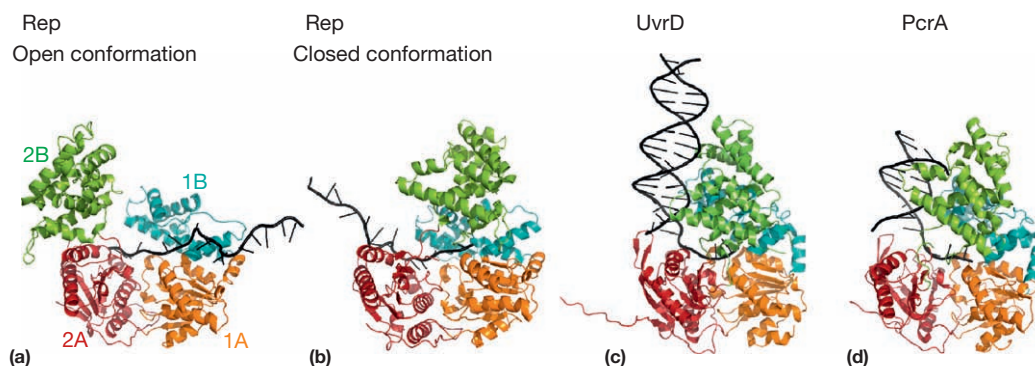
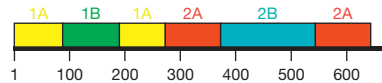


Figure 1 Structure of SF1 helicases bound to DNA. Subdomains 1A (orange), 1B (blue), 2A (red), and 2B (green) are indicated. In all the structures, the 3' end of the ssDNA resides near domain 1. In the two Rep structures, the open and closed conformations of the 2B domain and their position relative to the ssDNA are depicted.



Structural domains:



Helicase motifs:

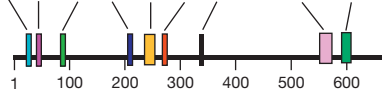


Figure 2 The domain structure of *E. coli* Rep monomer. The positions of the subdomains and the helicase motifs within the primary structure of Rep are also shown. The color scheme for the individual motifs corresponds to that shown in Figure 1. Modified from Korolev S, Hsieh J, Gauss GH, Lohman TM, and Waksman G (1997) Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell* 90: 635–647.

monomers bound to duplex DNA with a 3' ssDNA tail, residues in the 2B subdomain make contact with the duplex DNA and it has been inferred that these contacts are important for helicase activity; however, since PcrA, Rep, and UvrD monomers are unable to unwind DNA, the exact role of this interaction is not clear. Recently, UvrD and PcrA monomers have been shown to bind with specificity to ss–dsDNA junctions possessing a 5' ssDNA tail and to initiate translocation from the junction. Furthermore, mutations in the 2B subdomain affect this junction specificity. Hence, the 2B subdomain duplex DNA interaction may function to load the monomeric forms of these enzymes at ss–dsDNA junctions so that they can use their ssDNA translocation activity rather than their helicase activity.

ssDNA Translocation by Monomers of SF1 Helicases

Monomers of PcrA, Rep, and UvrD are able to translocate with biased directionality (3'–5') along ssDNA in an ATP-dependent reaction. During translocation, these monomeric enzymes hydrolyze one ATP per DNA base translocated. However, since monomers of Rep, UvrD, and PcrA are unable to unwind duplex DNA *in vitro*, it is clear that ssDNA translocation alone is insufficient for DNA helicase activity for these helicases, and that dimerization is required. By contrast, the

phage T4 Dda enzyme has the ability to unwind DNA as a monomer but its unwinding processivity is enhanced when multiple monomers operate on the DNA.

DNA Unwinding by SF1 Helicases

Protein Oligomerization Is Required for Helicase Activity *In Vitro*

Single-turnover DNA-unwinding studies indicate that Rep monomers are unable to unwind DNA *in vitro* and that Rep oligomerization is required for initiation of DNA helicase activity *in vitro*. Single-molecule fluorescence techniques have shown that a monomer of Rep uses ATP hydrolysis to translocate toward the ss–dsDNA junction, but then displays futile ATP-dependent conformational fluctuations, followed by dissociation of the monomer. DNA unwinding is initiated only if a functional oligomeric helicase is formed. Significantly, partial dissociation of the oligomeric helicase during unwinding leaves an inactive Rep monomer, resulting in a stalled complex. This stalled complex can be resolved in two ways. Dissociation of the remaining bound monomer leads to rewinding of the DNA duplex; however, re-initiation of unwinding can occur upon re-formation of the functional helicase by binding of additional Rep protein. These results show that a Rep monomer is unable to sustain DNA unwinding after the active Rep oligomeric complex disassembles. This also suggests that the low unwinding processivity observed for Rep DNA unwinding *in vitro* may be due to the relative instability of the functional dimer.

Initiation of DNA unwinding *in vitro* also requires a dimeric UvrD complex in which one subunit is bound to the ss–dsDNA junction, while the second subunit is bound to the 3' ssDNA tail. Since the assay used in these UvrD studies relies on complete unwinding of an 18-base-pair (bp) duplex, these results cannot exclude the possibility that monomers might unwind a short stretch of DNA duplex; however, processive unwinding of an 18-bp duplex requires UvrD oligomerization. Therefore, simple unidirectional translocation of a UvrD, Rep, or PcrA monomer along ssDNA is not sufficient for helicase activity.

Mechanisms for DNA Unwinding and Translocation by SF1 Helicases

Whereas it seems likely that SF1 helicases share some features of their unwinding mechanisms, it is still too early to conclude whether a single mechanism applies to all SF1 helicases. In fact, there are clear differences among some of the SF1 helicases studied to date. Although Rep, UvrD, and PcrA must oligomerize to function as helicases *in vitro*, a monomer of the phage T4 Dda helicase possesses limited unwinding activity *in vitro*. It is therefore possible that some SF1 helicases function as monomers to unwind short stretches of duplex nucleic acids, while oligomerization may be necessary to activate some SF1 helicases or to enhance their processivity. Of course, interaction of a helicase with an accessory protein may also modify the need for the helicase to oligomerize by providing a second DNA-binding site. It is also likely that oligomerization serves as the means by which the translocation and helicase activities of Rep, PcrA, and UvrD can be separated. That is, monomers of

these enzymes may actually function as translocases rather than helicases and oligomerization is needed to turn them into active helicases.

Most proposed mechanisms for DNA unwinding and helicase translocation assume the functional helicase to possess at least two DNA-binding sites. Differences in current models for DNA unwinding center on three aspects: (1) whether translocation of the helicase along ssDNA is sufficient for helicase activity or whether the helicase also interacts directly with the duplex DNA; (2) whether the two DNA-binding sites are contained within a single polypeptide; and (3) whether the same DNA-binding site remains as the lead site (inch-worm model) or whether multiple sites alternate as the lead subunit (rolling or hand-over-hand model).

Inch-Worm Mechanisms for ssDNA Translocation by SF1 Monomers

Based on the crystal structures of PcrA bound to DNA, a PcrA monomer has been proposed to unwind DNA by an inch-worm mechanism (Figure 3). Yet, unwinding by PcrA, Rep, and UvrD monomers *in vitro* has not been observed. However, aspects of this model may explain the mechanism of directional translocation along ssDNA. In the structure, the 2A domain contacts the ds-ssDNA junction and the 1A domain is bound to the ssDNA region. The inch-worm model proposes conformational changes in the 1A and 2A domains coupled to ATP binding and hydrolysis enables processive movement of the monomer on DNA. In the absence of ATP, the 2A domain binds weakly to dsDNA whereas the 1A domain has a high affinity for ssDNA. Binding of ATP in the cleft between the two domains results in closure of the space between the two domains and is coupled to a switch in the DNA-binding affinities, that is, 2A binds tightly to DNA whereas 1A binds with weaker affinity. ATP hydrolysis reverses the DNA-binding affinities and results in a net movement of one base per ATP hydrolyzed.

Interestingly, a recent structure of the *Deinococcus radiodurans* RecD2, an SF1 translocase with 5'-3' directionality indicates that it binds to ssDNA with the orientation as do Rep, PcrA,

and UvrD, all of which are 3'-5' translocases. Therefore, the basis for the different directionalities of translocation is not due to a different binding orientation. Rather, it has been suggested that the protein conformational changes that drive translocation are simply reversed.

Active versus Passive Mechanisms of DNA Unwinding

One operational distinction among DNA unwinding models is whether the helicase participates directly in destabilizing the duplex DNA (active mechanism), or whether the helicase interacts solely with the ssDNA and waits for transient fluctuations in the duplex to form a ssDNA region (end fraying) onto which the helicase can then translocate (passive mechanism). Active mechanisms would generally involve direct binding of the helicase to the duplex DNA (or the duplex-ssDNA junction), in addition to the ssDNA, although a torsional mechanism is possible in which binding of the helicase to both single strands might unwind the duplex without direct duplex interactions.

Tests of a passive mechanism of unwinding have been made for *E. coli* Rep and UvrD helicases using non-natural DNA substrates in which a region of the ssDNA adjacent to the duplex was reversed in its backbone polarity or replaced by polyethylene glycol. The segments of polyethylene glycol or reversed polarity ssDNA would prevent a helicase from translocating up to the ss-dsDNA junction, thus a helicase operating by a passive mechanism would not be expected to unwind such DNA substrates. However, Rep and UvrD can unwind such DNA molecules, ruling out a passive mechanism for these helicases.

Dimeric, Subunit Switching Models

Active, Rolling or Hand-Over-Hand Models for a Dimeric Helicase

An active, rolling mechanism (Figure 4(a)) has been proposed for how a Rep dimer might unwind DNA and translocate along

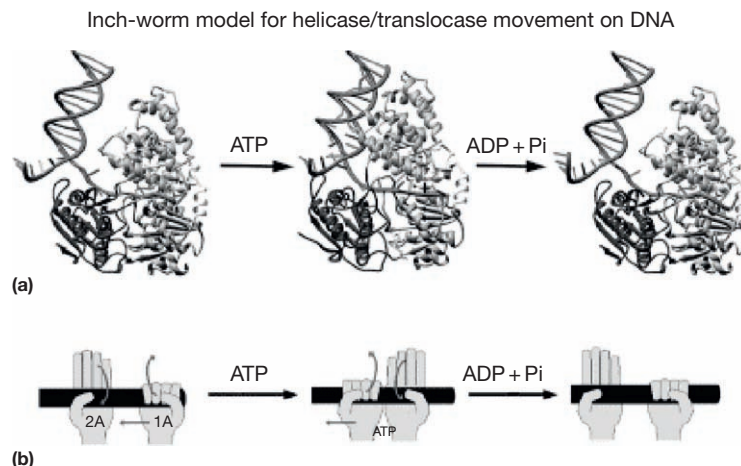


Figure 3 Inch-worm mechanism of helicase/translocase movement on DNA. The two hands represent the 2A and 1A domains respectively. Reprinted from Soultanas P and Wigley DB (2000) DNA helicases: 'Inching forward'. *Current Opinion in Structural Biology* 10: 124–128, with permission from Elsevier.

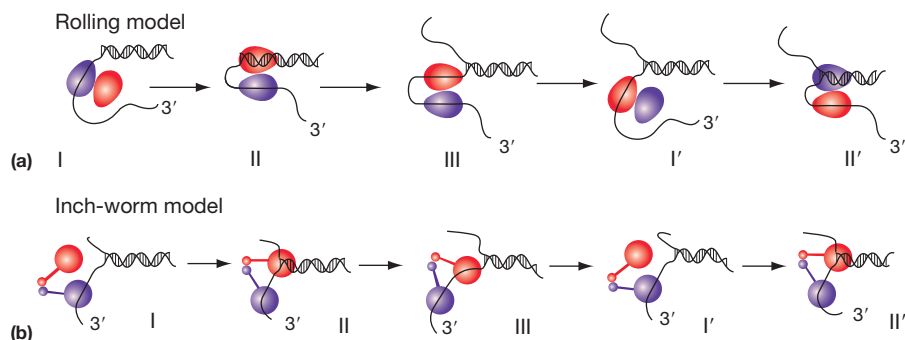


Figure 4 Models for DNA unwinding and translocation by a dimeric helicase. (a) Active, rolling or hand-over-hand model, and (b) dimeric inch-worm model.

the DNA filament. In this model, the individual subunits alternate as the lead subunit, which binds to duplex DNA, followed by unwinding. However, all of the evidence upon which this model was based is also consistent with a dimeric inch-worm model shown schematically in [Figure 4\(b\)](#).

Dimeric Inch-Worm Models

The only difference between the rolling and dimeric inch-worm models is that the same subunit remains as the lead subunit in the dimeric inch-worm model. In fact, recent evidence with both Rep and UvrD favours a dimeric inch-worm model rather than a rolling or hand-over-hand model. Based on the fact that monomers of Rep and UvrD are able to translocate along ssDNA with biased (3'–5') directionality, it seems most likely that in a dimeric inch-worm model, the trailing subunit of the dimer might maintain continuous contact with the 3' ssDNA, possibly providing the motor component of the helicase, whereas the leading subunit would interact with the ss–dsDNA junction. However, such details are still speculative and need to be tested.

Protein Displacement by SF1 Translocases

Apart from their ability to translocate and unwind DNA, SF1 helicases also function in DNA repair by removing recombination factors from DNA. UvrD removes RecA from ssDNA and this activity is also conserved in its yeast (Srs2) and human (RTEL1) orthologs. Translocation on ssDNA is most likely required for this activity as it pertains to RecA/Rad51 nucleoprotein filaments formed on ssDNA. Apart from the motor activity of the translocase, specific protein–protein interactions are required for this displacement activity. The domains/residues that mediate such interactions are yet to be identified in UvrD or Rep. As mentioned above, the 2B domain is inhibitory to the DNA unwinding activity of Rep and it remains to be tested if this domain has a role in displacing other proteins from DNA. Since a monomer efficiently translocates on ssDNA, it might be the active form of the translocase during protein displacement, but a protein bound on DNA could serve as a potent blockade and the requirement for multiple translocases cannot be ruled out.

Summary

It is now clear that monomers of SF1 helicases are able to translocate along ssDNA with biased directionality in an ATP-dependent reaction. These monomers must use some sort of inch-worm mechanism for this translocation. However, monomers of Rep, UvrD, and PcrA are not able to initiate DNA unwinding *in vitro*, despite the fact that these monomers can translocate along ssDNA with directional bias. Therefore, the ability to translocate with unidirectional bias along ssDNA is not generally sufficient for a protein to have helicase activity. The 2B subdomain of Rep is not required for its helicase activity *in vitro*; in fact, this domain inhibits the ability of a Rep monomer to unwind DNA *in vitro*. Rep, UvrD, and PcrA require oligomerization in order to initiate DNA unwinding *in vitro*, thus the inhibition of helicase activity by the 2B subdomain of Rep appears to be relieved through oligomerization.

See also: [Molecular Biology: DNA Helicases: Hexameric Enzyme Action; DNA Mismatch Repair in Bacteria; DNA Replication Fork, Bacterial; DNA Replication: Initiation in Bacteria.](#)

Further Reading

- Fairman-Williams ME, Guenther UP, and Jankowsky E (2010) SF1 and SF2 family matters. *Current Opinion in Structural Biology* 20: 313–324.
- Lohman TM and Bjornson KP (1996) Mechanisms of helicase-catalyzed DNA unwinding. *Annual Review of Biochemistry* 65: 169–214.
- Lohman TM, Tomko EJ, and Wu CJ (2008) Non-hexameric DNA helicases and translocases: Mechanisms and regulation. *Nature Reviews Molecular Cell Biology* 9: 1–10.
- Modrich P (1994) Mismatch repair, genetic stability, and cancer. *Science* 266: 1959–1960.
- Myong S and Ha T (2010) Stepwise translocation of nucleic acid motors. *Current Opinion in Structural Biology* 20: 121–127.
- Patel SS and Picha KM (2000) Structure and function of hexameric helicases. *Annual Review of Biochemistry* 69: 651–697.
- Pyle AM (2008) Translocation and unwinding mechanisms of RNA and DNA helicases. *Annual Review of Biochemistry* 37: 317–336.
- Singleton MR, Dillingham MS, and Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annual Review of Biochemistry* 76: 23–50.

Nonhomologous End Joining in Bacteria

E J Bartlett and A J Doherty, University of Sussex, Brighton, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Crystal structure High-resolution structural data derived from protein crystals.

Double-strand break Severe DNA damage where both strands of the helix break.

Micro-homology Short stretches of matching sequences in otherwise noncomplementary DNA strands used to promote break repair.

Nonhomologous end joining The repair of double-strand breaks without the use of a homologous DNA template.

Double-Strand Break Repair

DNA is exposed to a number of endogenous and exogenous agents, which can cause major structural damage. Double-strand breaks (DSBs) are the most lethal form of DNA damage as they prevent replication of the genetic 'blueprint' of life. Specific types of damaging agents, including ionizing radiation (IR) and reactive oxygen species (ROS), are capable of causing both strands of the DNA helix to be severed. All forms of cellular life maintain molecular pathways for repairing such dangerous lesions. The two major pathways are homologous recombination (HR) and nonhomologous end joining (NHEJ). HR requires an identical DNA donor copy of the broken molecule in order to accurately replicate the genetic information. NHEJ facilitates the direct joining together of broken DNA ends in the absence of a template. NHEJ can repair 'clean' breaks or remodel severely damaged termini to reconnect DNA ends. This process can be achieved by harnessing short regions of complementary bases, known as 'microhomologies', to form a connection between the ends, which is then cleaned up, if required, by polymerases and nucleases, prior to resealing by DNA ligases.

Different NHEJ Enzymes Are Required for Different Breaks

NHEJ is found in many organisms and typically the system is composed of a DNA end-binding protein (Ku), a ligase, a polymerase, and a nuclease. Ku recognizes DSBs and grasps the ends of the broken strands. Since DNA damage occurs in many different ways, the effects on the molecule can greatly differ. If the DNA duplex is neatly broken into two and the ends remain complementary, then the phosphate backbone can be rejoined directly by a specific ligase enzyme. Sometimes DSBs have single-stranded overhanging bases, which cannot be directly ligated (Figure 1). In this case, a nuclease activity removes or resects bases from the end until the two termini can be reconnected, often by microhomology. If there are any remaining gaps in the DNA, a specific NHEJ polymerase can fill them in, before the ligase finalizes the repair.

Cell-Cycle-Dependent DNA Break Repair

In eukaryotic cells, HR is the dominant DSB repair pathway in the S and G₂ phases of the cell cycle. HR is an error-free method

of replacing lost genetic material using the sister chromatid as a template. However, only a single copy of the genome is accessible during the nondividing stationary phase (G₁/G₀) of the cycle. If a break occurs in this phase, NHEJ is required to directly repair the DSB without relying on an intact template. For this reason, NHEJ can produce inaccurately repaired DNA, although the extent of the lost genetic information is thought to depend on the nature of the break. In eukaryotes, NHEJ often fails to replace lost nucleotides at the site of the DSB at either end. The ability of the NHEJ enzymes to remodel cleaved ends also inherently introduces variation to the site of repair, but this is preferable to cell death.

Bacterial NHEJ

For many years, the NHEJ repair pathway was thought to be exclusive to eukaryotes; however, in the last decade, it has been discovered in most species of bacteria and some archaea. Mammalian NHEJ requires five core components for NHEJ; DNA Ligase IV/XRCC4 complex, XLF, the heterodimer Ku70/80, DNA PK, and Artemis. By contrast, most bacterial NHEJ systems appear to require just two proteins: ligase D and Ku. In bacterial species such as *Mycobacteria tuberculosis* and *Pseudomonas aeruginosa*, ligase D (LigD) is a multifunctional enzyme, capable of various DNA remodeling and repair activities. These proteins are responsible for maintaining stability of the bacterial genome during stationary phase, which might encompass prolonged periods of nutrient starvation. Sporulating bacteria, such as *Bacillus subtilis*, also require NHEJ repair in the haploid spores. Both quiescent mycobacteria and *Bacillus* spores are not actively replicating, but are still accumulating significant prolonged DNA damage. HR is not a viable repair mechanism for this scenario; so, NHEJ is maintained as a practical solution to preserve an intact genome. Interestingly, not all bacterial species encode NHEJ proteins, and *Escherichia coli* is a notable example, but even these organisms appear to utilize a ligase-dependent process to repair DSBs in stationary phase.

Bacterial Ku

All Ku proteins form dimeric ring-like structures, capable of sliding over DNA ends. Ku does not bind to specific genetic sequences, nor make extensive contacts with the phosphate

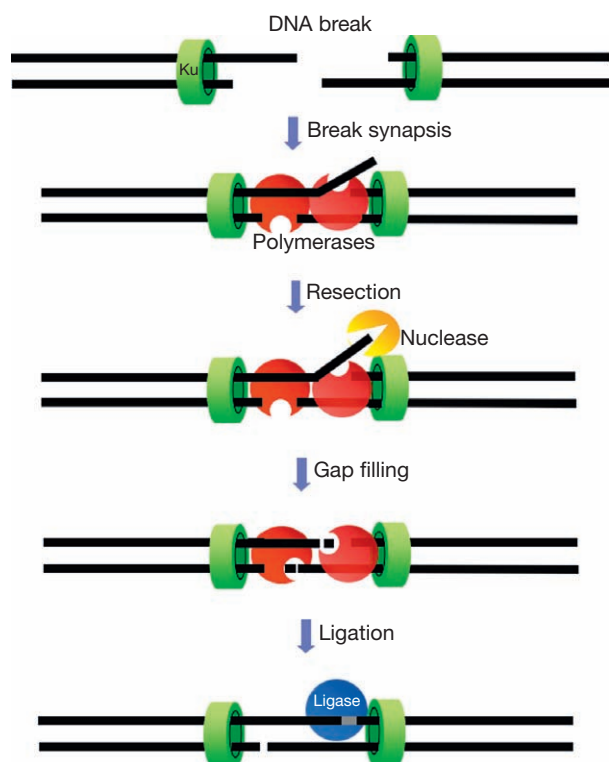


Figure 1 Mechanism of NHEJ in prokaryotes. Ku homodimers bind to the ends of the DNA break and recruit LigD via the polymerase. The polymerase domain of LigD specifically binds to a 5' phosphate and, together with Ku, promotes end synopsis. The nuclease and polymerase activities of LigD, and possibly other factors, process the break termini, if required, to restore complementary ends. Finally, sealing of the nicks by ligase domain repairs the break.

backbone. Instead, Ku fits around the major and minor groove contours of the DNA, analogous to a nut on a bolt. Ku is thought to initially recognize and bind broken DNA termini and then recruit LigD for end processing, gap filling, and ligase activities. Although no crystal structure of bacterial Ku exists, deletion mutants have demonstrated that Ku binds to the polymerase domain of LigD, and greatly stimulates end joining *in vitro*. Bacterial Ku is usually found as a single gene and exists as homodimers. It is much smaller than the eukaryotic Ku proteins and does not include the vWA and SAP domains that make Ku 70/80 distinct. These additional domains allow eukaryotic Ku to behave as a scaffold protein and interact with the variety of factors in the more complex NHEJ system. The eukaryotic Ku70/80 heterodimer was likely the result of a gene duplication of a Ku homodimer at some point in evolutionary history and examples of such operon-encoded duplications exist in bacteria.

Bacterial DNA LigD

Ku and LigD represent a functional two-component NHEJ system in bacteria. Plasmid repair has been achieved *in vitro* utilizing just Ku and LigD, and deletions of both *in vivo* significantly reduce the frequency of NHEJ repair events. Deletion of

LigD in bacteria is nonlethal, unlike the essential LigA which is used during DNA replication. LigD is not necessary for survival under normal growth conditions; however, LigD null mutants (Δ LigD) in mycobacterium are sensitized to IR and have a lower survival rate than wild-type cells. In some bacteria, LigD operates as a compact multifunctional repair enzyme, with three distinct domains each with related catalytic activities; ligase, polymerase, and nuclease. The order of these three domains is not strictly conserved throughout different bacterial species and many LigDs possess only ligase and polymerase domains. It should also be noted that many prokaryotes encode these activities on distinct genes that are commonly operon associated.

The ligase domain of LigD is the largest of the three subunits and is used exclusively for the final nick-sealing step of NHEJ. Mutation of the catalytic residues in the ligase domain prevents successful NHEJ. The polymerase domain exhibits a remarkable versatility of nucleotide incorporation. Nucleotides can be added to blunt ends without template (terminal transferase), in gaps in the sequences, and opposite damaged bases. This NHEJ polymerase is not a member of the family X polymerases, like the human NHEJ polymerases μ , λ , β , and TdT. Instead, the bacterial polymerase is a member of the archaeo-eukaryotic primase superfamily. The implication of this discovery is that the NHEJ polymerase may have evolved into the end-joining role from a primase.

Several crystal structures of different bacterial NHEJ polymerases have been solved, including one bound to DNA. These structures have demonstrated key amino acid residues that contact the DNA in different places and those which guide the incoming nucleotide into the active site ready for addition. The crystal structure of a complete LigD has not been solved, although the isolated ligase domain has been elucidated. The structure of phosphoesterase/nuclease (PE) domain has also yet to be solved, which may hold the key to its specific role in NHEJ. The PE domain is capable of removing both nucleotides and phosphate groups from the ends of DNA, although loss of these activities does not affect NHEJ in plasmid repair assays; therefore, its exact role remains to be established. It is thought that an unidentified nuclease may provide nucleolytic activity in bacterial NHEJ.

Mechanism of Prokaryotic NHEJ

In the event of a DSB, Ku rapidly binds the broken ends, protecting them from exonucleolytic degradation. LigD is then recruited to the break by Ku via its polymerase domain. Recent crystallographic and biochemical studies have determined that two polymerases can work in concert to bring the DNA ends into direct contact, using specific surface loops to act as DNA end 'matchmakers' (Figure 1). This polymerase-mediated synopsis process utilizes microhomologies between the two ends to encourage base pairing. Depending on the nature of the damage to the DNA termini, bases may be removed by nucleolysis or added by the polymerase to promote end pairing by microhomology. Once the DNA strands are back in contact, polymerase and nuclease activities remove unpaired nucleotides and fill in the remaining gaps. Finally,

the continuity of the DNA is resorted by the ligase, which seals the remaining nicks.

NHEJ in Viral Infection

NHEJ is now known to be conserved throughout all forms of life, following its discovery in viruses. In 2006, mycobacteriophages were shown to possess functional Ku homologs. These viral proteins share the DNA-binding abilities and ring-like structure of the prokaryotic and eukaryotic orthologs. Following infection of *Mycobacterium smegmatis*, viral Ku proteins were shown to interact with the host LigD. The Ku proteins bind to the ends of the linear phage DNA and stimulate LigD to ligate the ends to produce a circularized genome, thus allowing rolling-circle DNA replication. Bacteriophage Mu encodes another Ku-like protein known as 'Gam'. Like the eukaryotic and bacterial homologs, Gam operates as a homodimer and binds to DNA. Gam is a much more distant relative of the Ku family and does not have a role in NHEJ repair but appears to protect the linear genome from degradation following injection into the host. It is plausible that this protective role of Ku-like proteins may be significant in enabling lateral gene transfer within bacteria, including phage genetic material. Ku may have initially existed as a phage protein, which became incorporated into bacteria and eventually evolved to participate in DNA break repair processes.

Roles for NHEJ in Bacteria

Many species of bacteria employ NHEJ to repair DSBs produced by exposure to IR; however, excessive exposure to IR is an unlikely environmental concern for prokaryotes. Homologs of Ku and LigD are present in a large variety of bacterial species; therefore, it is likely that these genes have been retained for a specific purpose. Bacterial survival of the extreme DNA-damaging effects of gamma irradiation coincides with improved survival of dehydration. Desiccation and dry heat are a more significant source of DSBs for bacteria. Preservation of a repair

pathway that can repair DNA breaks, produced under harsh environmental conditions, without the need for a DNA template is therefore highly beneficial. NHEJ provides this increased desiccation resistance which is seen in *M. smegmatis* and *B. subtilis* among others.

Mycobacterial infection of human hosts potentially requires a robust NHEJ system for survival, since any persistence is met with an immune defense. Mycobacterial cells subside into latent infection after the initial acute phase in human tuberculosis (TB). Macrophages cause oxidative damage to pathogens, capable of promoting DSBs in invading DNA as part of the immune response. Should Ku and LigD prove integral to the ability of *M. tuberculosis* to repair DSBs while in quiescence, development of inhibitors might have significant effects on the virulence of the bacteria.

See also: **Molecular Biology:** DNA Ligases: Mechanism and Functions; DNA Mismatch Repair and Homologous Recombination.

Further Reading

- Bowater R and Doherty AJ (2006) Making ends meet: Repairing breaks in bacterial DNA by non homologous end-joining. *PLoS Genetics* 2: e8.
- Brissett NC, Pitcher RS, Juarez R, et al. (2007) Structure of a NHEJ polymerase-mediated DNA synaptic complex. *Science* 318: 456–459.
- Della M, Palmbo PL, Tseng HM, et al. (2004) Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science* 306: 683–685.
- Gong CL, Bongiorno P, Martins A, et al. (2005) Mechanism of nonhomologous end-joining in mycobacteria: A low-fidelity repair system driven by Ku, ligase D and ligase C. *Nature Structural and Molecular Biology* 12: 304–312.
- Mahaney BL, Meek K, and Lees-Miller SP (2009) Repair of ionizing radiation-induced DNA double-strand breaks by non homologous end-joining. *Biochemical Journal* 417: 639–650.
- Pitcher RS, Brissett NC, and Doherty AJ (2007) Nonhomologous end-joining in bacteria: A microbial perspective. *Annual Review of Microbiology* 61: 259–282.
- Pitcher RS, Wilson TE, and Doherty AJ (2005) New insights into NHEJ repair processes in prokaryotes. *Cell Cycle* 4: 675–678.
- Shuman S and Glickman MS (2007) Bacterial DNA repair by non-homologous end joining. *Nature Reviews* 5: 852–861.
- Weller GR, Kysella B, Roy R, et al. (2002) Identification of a DNA nonhomologous end-joining complex in bacteria. *Science* 297: 1686–1689.

Nonhomologous End Joining in Eukaryotes

E Weterings, Nevada Cancer Institute, Las Vegas, NV, USA

D J Chen, University of Texas Southwestern Medical Center, Dallas, TX, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Apoptosis Also referred to as 'programmed cell death', an event that leads to loss of cellular morphology, loss of attachment, nuclear fragmentation, chromatin condensation, and chromosome fragmentation, ultimately resulting in cell death.

Cell-cycle arrest Cessation of cell division in response to cellular damage. Regulated by checkpoint proteins in the G1/S, intra-S, and G2/M stages of the cell cycle.

Hypomorphic mutation Mutation leading to reduction (but not annihilation) of the gene product's activity.

Ligation Formation of a phosphodiester bond between the 3' OH group of a DNA strand and the 5' PO₄ group of another DNA strand, linking both strands, catalyzed by a ligase.

Lymphocytes Type of white blood cells, divided into T-cells, B-cells, and natural killer (NK) cells. B-cells respond to pathogens by the secretion of antibodies (humoral immune system); T-cells mediate the release of cytokines and cytotoxins, which induce apoptosis of infected target cells (cellular immune system).

Cellular Responses to the Onset of DNA Double-Strand Breakage

The DNA helix is a vulnerable molecule. Chemical modifications of the DNA backbone and nucleotides are continuously introduced by both intra- and extracellular factors, despite the biological necessity for genetic integrity and robustness. Such modifying factors include oxygen radical species, reactive metabolites, and ionizing radiation, all of which have the capacity to modify the sugar-phosphate backbone of the DNA molecule in such a way that it breaks apart. When both strands of the helix are severed in close proximity, a DNA double-strand break (DSB) occurs. The presence of a DSB, in turn, can cause the chromosome to physically fall apart.

Chromosome breakage caused by the presence of a DSB leads to an uneven distribution of chromosomes over the daughter cells during mitosis, resulting in deletion or translocation of potentially critical genes, such as oncogenes or tumor suppressor genes. In order to prevent this, eukaryotic cells have developed different defense mechanisms, each working on a different level of the cellular organization. These mechanisms include cell-cycle arrest, the induction of apoptosis, and the actual repair of the DSB.

The uneven distribution of chromosome fragments during mitosis is primarily prevented by arrest of the cell cycle. Eukaryotic cells are equipped with checkpoints in the major phases of mitosis (G1, S, G2, and M), which are triggered by the presence of genomic breaks. Activation of a cell-cycle checkpoint introduces a delay in the mitotic progression during which the cell can repair its genomic damage through one or several of the available DNA repair pathways. Alternatively, a cell can initiate a process called 'apoptosis' or controlled cell death, which is believed to annihilate cells that are damaged to such an extent that adequate repair is unlikely.

Although these cellular responses to DNA breakage are mediated by an extensive cascade of proteins, one enzyme appears to play a particularly important role during the regulation of several of these processes: the ataxia telangiectasia

mutated (ATM) kinase. This serine/threonine protein kinase is activated shortly after the introduction of a DSB and is thought to be responsible for the phosphorylation of the nucleosome subunit H2AX – an important regulatory step during the initiation of cell-cycle arrest. In addition, ATM activates the cell-cycle-regulating protein Chk2 and the apoptosis-mediating protein p53. Finally, growing evidence suggests that the ATM kinase activity is important during actual DSB repair as well. In the light of this wide variety of regulatory functions, ATM is often referred to as a sensory protein.

Those cells that do not undergo apoptosis and are properly halted in their cell-cycle progression have two distinctly different pathways at their disposal for the proper repair of DSBs. A set of proteins and enzymes constituting the homologous recombination (HR) pathway mediates removal of a DNA fragment containing the break followed by resynthesis of the removed fragment. This process requires a homologous template to guide the DNA synthesis step and is therefore predominantly active during cycle phases in which sister chromosomes are present (S and G2). Contrarily, the nonhomologous end-joining (NHEJ) pathway is characterized by direct repair of the break which does not require a homologous template.

NHEJ: The Process and the Enzymes

In order to directly couple two pieces of DNA together without the need for template guided synthesis, it is necessary that the ends of the DNA fragments are first brought into close proximity. This initial step is followed by preparation of the DNA ends in such a manner that they become compatible for final ligation. The NHEJ process can, therefore, be roughly divided into an early phase, which consists of recognition and tethering of the DNA termini, and a later phase in which further processing and ligation take place. However, it should be noted that the core enzymes of the early phase interact with those of the later phase and that several subprocesses of tethering and processing

may actually overlap, causing the borders between the early and late phase to be less than clearly drawn.

It is generally accepted that the first factor to associate with both termini of the DNA fragments is a heterodimer consisting of an 86-kDa and a 73-kDa subunit – the Ku70/80 protein. Crystallography studies have revealed that the assembled Ku70/80 dimer assumes a ring-shaped structure which snugly fits the exposed DNA terminus. This unique structure–function relationship is responsible for the exceptionally high affinity of the Ku70/80 molecule for open DNA ends. The main body of the ring is composed of the central regions of the Ku70 and the Ku80 molecules, whereas the protruding carboxy-termini may be involved in stabilization of the Ku–DNA interaction or in regulation of other NHEJ factors.

The complex formed by the association of Ku70/80 with both DNA termini functions as the building basis for the entire NHEJ assembly. One of the factors attracted by Ku70/80 and generally associated with the early NHEJ phase is a 460-kDa serine/threonine kinase called ‘DNA-dependent protein kinase catalytic subunit’ (DNA-PK_{CS}). The kinase activity of DNA-PK_{CS} depends on binding of this enzyme to both Ku70/80 and the DNA termini. As such, DNA-PK_{CS} is truly the catalytic subunit of a greater assembly, consisting of the whole terminal DNA–Ku–DNA–PK_{CS} complex.

The DNA-PK_{CS} protein is considered a good candidate for the role of tethering factor – the core protein that mediates juxtaposition of the two DNA ends prior to processing and ligation. Both atomic force and electron microscopy studies have provided insight into the possible organization of the DNA–Ku–DNA–PK_{CS} complex, although these have not yet led to a definitive and unequivocal understanding of the positioning of the large DNA-PK_{CS} molecule at the DNA termini. The most likely predictions, however, are dimeric structures in which the DNA fragments are juxtaposed by two DNA-PK_{CS} molecules, each associated with its own Ku70/80 dimer and DNA terminus (see Figure 1). The existence of such structures is also suggested by a series of biochemical experiments.

The function of the DNA-PK_{CS} kinase activity during NHEJ is a matter of contemporary dispute. Deletion studies have shown that impairment of kinase domain results in diminished DSB repair, suggesting an important function. The quintessential substrate of the DNA-PK_{CS} kinase, however, is yet to be determined. In fact, it is likely that several NHEJ core factors are phosphorylated, activated, or regulated in some fashion by DNA-PK_{CS}.

One of the best-researched substrates for DNA-PK_{CS} is the DNA-PK_{CS} molecule itself. This protein possesses at least 16 autophosphorylation sites, which are serine or threonine residues located in three main areas of the protein’s structure: clustered around either amino-acid residue 2056 or 2609, or located in the carboxy-terminal catalytic domain. Convincing evidence exists that autophosphorylation of DNA-PK_{CS} alters the structural conformation of the DNA–Ku–DNA–PK_{CS} complex.

Although, at present, there are no details available on the specific function of each individual autophosphorylation site, it has become apparent that overall phosphorylation of the DNA-PK_{CS} molecule results in increased accessibility of the DNA ends to modifying enzymes such as nucleases and ligases. This finding has given rise to a model in which the large

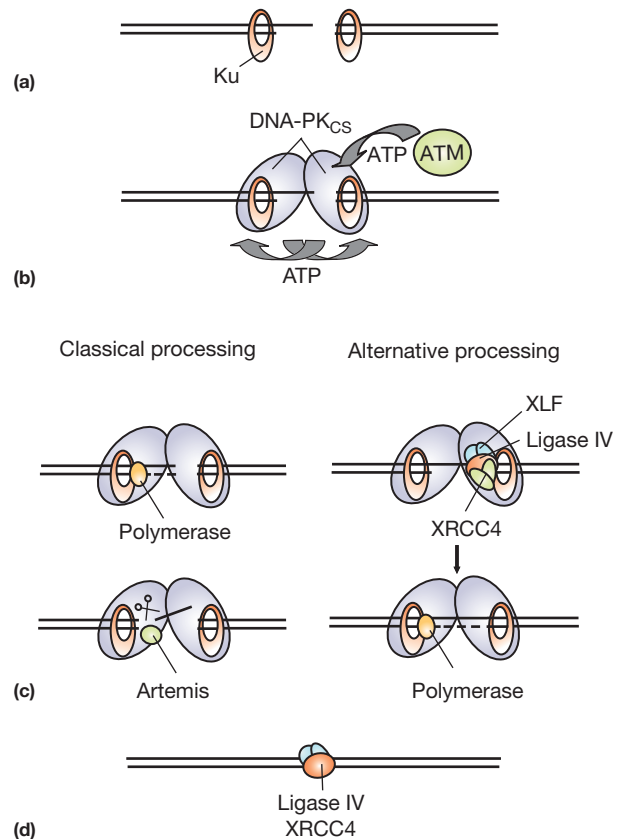


Figure 1 Working model for nonhomologous end-joining. (a) Recognition of the DSB by Ku70/80. (b) Tethering of the DNA ends, followed by phosphorylation and autophosphorylation of DNA-PK_{CS}. (c) Processing of the DNA ends via different routes. (d) Ligation of the DSB.

DNA-PK_{CS} molecule functions as a gatekeeper – tethering the DNA ends and protecting them against improper degradation by sterically blocking access, until the two DNA termini have approached each other closely enough for DNA-PK_{CS} autophosphorylation to occur in trans (see Figure 1). This latter event renders the DNA ends available for further processing and subsequent ligation.

Recently, it has been suggested that autophosphorylation is not the only one responsible for the initiation of conformational changes in the DNA-PK_{CS} molecule. Phosphorylation of the 2609 and 2647 DNA-PK_{CS} residues may be mediated by another factor, the ATM kinase. ATM is quickly activated after the introduction of a DSB and is known to facilitate signal transduction events that play a role in the cell-cycle arrest events that accompany the repair of DNA breaks. It is now becoming clear that in addition to its function in signal transduction, ATM may appear at the stage of the actual NHEJ process as well. It is conceivable that ATM-mediated phosphorylation of certain DNA-PK_{CS} residues may be required in conjunction with DNA-PK_{CS} autophosphorylation in order to establish the proper set of conformational changes necessary for further modifications of the DNA ends. Such a double requirement for NHEJ progress may ensure proper alignment of the NHEJ process and related pathways such as cell-cycle arrest.

After the DNA termini are brought in close proximity and the tethering complex has assumed a configuration that allows access of other NHEJ factors, the DNA ends must be processed in order to yield a substrate for the final ligation step. To this end, the NHEJ process utilizes a set of enzymes, none of which can be considered to be specialized NHEJ factors; instead, they appear to be derived from the pool of generally available genome maintenance enzymes as needed. For instance, the addition of 5' phosphate groups as necessary for ligation is mediated by the multifunctional mammalian polynucleotide kinase and removal of 3' phosphoglycolates by a host of NHEJ-unrelated enzymes, including the endonuclease APE1 and the tyrosyl-DNA phosphodiesterase TDP1.

Generally, events that create DSBs do not leave a readily ligatable blunt-ended break. Commonly encountered structures will possess either 3' or 5' single-strand overhangs of variable length. The most elementary processing of such overhang structures is accomplished by either filling or resection (see [Figure 1](#)). Several factors have been reported to mediate the filling of overhangs during NHEJ by using the single-strand sequence as a template to synthesize the complete double-strand structure. Polymerases μ and λ are both thought to be recruited by the NHEJ process, although it is unclear how this recruitment is organized or regulated. Polymerase μ has been shown to interact with the Ku heterodimer, making it likely that the DNA–Ku scaffold is able to attract this processing enzyme to the repair complex. This mode of recruitments has also been observed for the human terminal deoxynucleotidyl-transferase (TdT) enzyme, which catalyzes the addition of nucleotides to DNA ends.

Knowledge of factors that mediate the resection of overhangs during NHEJ is presently limited to the Artemis nuclease. This enzyme was originally identified as a critical player during the process of V(D)J recombination, catalyzing the nucleolytic opening of hairpin structures. Because Artemis deficiency leads not only to impairment of V(D)J recombination, but also to radiation sensitivity, it was soon recognized that this nuclease may also be involved in the processing of DNA termini during NHEJ. The credibility of this notion is strengthened by the fact that Artemis is a substrate for phosphorylation by the two NHEJ kinases ATM and DNA-PK_{CS}, although the biological significance of this interaction is still far from obvious.

Artemis is known to physically associate with the DNA-PK_{CS} molecule. This association changes the catalytic activity of Artemis – residing in the amino-terminal region of the protein – from exonucleolytic to endonucleolytic and is, therefore, considered to be an important prerequisite for the processing function of Artemis during both NHEJ and V(D)J recombination. The Artemis nuclease is most likely recruited to the repair complex through its interaction with the DNA-PK_{CS} molecule, which, in turn, is anchored at the DNA break by association with the DNA–Ku complex. Recent evidence suggests that Artemis activity may be modulated through the carboxy-terminus of the Ku80 molecule, either by direct interaction or by alteration of DNA-PK_{CS} kinase or autophosphorylation activity. This intricate team play demonstrates once more the importance of the DNA–Ku complex as supporting scaffold for the entire NHEJ process.

After proper modification of the DNA ends, ligation will be mediated by a complex of two molecules: DNA ligase IV and

the X-ray cross-complementing group 4 (XRCC4). The amino-terminal region of ligase IV harbors its adenosine triphosphate (ATP)-utilizing catalytic domain and shares a marked homology with mammalian ligases I and II. The carboxy-terminus of ligase IV is very different from those of other mammalian ligases and contains two BRCT (BRCA1 carboxy-terminus like) motifs which possibly facilitate interaction of ligase IV with other proteins. Quite unexpectedly, however, the BRCT motifs do not govern the interaction of ligase IV's most prominent partner XRCC4. This is thought to be mediated by a stretch of approximately 30 nucleotides in between the two BRCT domains. The XRCC4 protein has a mushroom-shaped, dimeric structure consisting of a globular amino-terminal head and a long carboxy-terminal stalk. The crystal structure of the ligase IV/XRCC4 complex reveals a conformation in which two XRCC4 molecules are positioned side by side, while one ligase IV molecule associates with the extending carboxy-terminal stalks of the XRCC4 proteins.

Ample evidence suggests a direct interaction between the Ku70/80 heterodimer and the ligase IV/XRCC4 complex, possibly mediated by the BRCT motifs in the ligase IV molecule. This interaction markedly stimulates the ligase activity of the complex approximately by one order of magnitude. Therefore, specific recruitment and activation of the ligase IV/XRCC4 complex at a DNA break is most likely regulated by the DNA–Ku scaffold. The attachment of the ligase IV/XRCC4 complex to the NHEJ machinery may be further strengthened by binding of the amino-terminal heads of the XRCC4 molecules to the DNA helix.

It is at present not clear during which stage of the NHEJ process the ligase IV/XRCC4 complex joins the repair efforts. Classic models tend to include association of ligase IV/XRCC4 with the repair machinery as the last step before ligation. In the past years, however, these models have been disregarded as incomplete and evidence is accumulating that the ligase IV/XRCC4 complex may play an integral part in earlier phases of the NHEJ pathway. These functions notably include alternative processing of single-strand overhangs, in which ligase IV/XRCC4 and Ku cooperate to accomplish tethering of noncomplementary overhangs. Such intermediate structures can be converted into double-strand structures during an integrated processing/joining operation (see [Figure 1](#)). Possibly, this mode of alternative processing may circumvent the necessity to resect single-strand overhangs in a subset of repair events and may contribute to eliminate the loss of genetic information during NHEJ.

The notion that alternative processing routes may take place has considerably gained credibility after the recent discovery of a seventh NHEJ key protein: XRCC4-like factor (XLF), also known as Cernunnos. This relatively small protein bears a structural resemblance to the XRCC4 molecule and interacts with both Ku70/80 and ligase IV/XRCC4, stimulating ligation of noncompatible DNA ends. Quite possibly, XLF supports the alternative processing routes mediated by the interplay between Ku and ligase IV/XRCC4, thereby indirectly decreasing the total contribution of processing events that result in nucleotide loss. Resolution of the XLF crystal structure revealed that this protein is theoretically capable of associating with the ligase IV/XRCC4 complex in a 2:2:1 XRCC4:XLF:ligase IV ratio.

Accessory Proteins to the NHEJ Process

In addition to its key enzymatic machinery, the NHEJ process appears to utilize several proteins that are accessory to this pathway. For instance, the breast cancer type 1 susceptibility protein (BRCA1) is known to somehow stimulate NHEJ. Although the exact nature of BRCA1 involvement in NHEJ is unclear, several studies suggest that this protein is able to suppress DNA end processing, thereby reducing the loss of genetic information and increasing the fidelity of repair.

A complex of three proteins (MRE11, Rad50, and NBS1), known under the abbreviation MRN, plays a central role in the process of HR, but may also assist the NHEJ pathway in a subset of repair events. The Rad50 component of the MRN complex provides two extending, flexible arms that may be involved in tethering of DNA fragments. Although the NHEJ core enzymes provide a mechanism that supports DNA tethering, several authors have suggested that MRN may actually compete for available DNA termini throughout the repair process. Such interchangeability of factors between NHEJ and HR shades the borderline between these two pathways, but simultaneously provides insight into possible regulatory mechanisms that mediate the choice between DSB repair paths.

Function of NHEJ in V(D)J Recombination and Telomere Maintenance

The importance of the NHEJ pathway is not limited to the repair of accidentally introduced DSBs. Functionality of the NHEJ core enzymes is indispensable for two other major cellular processes: V(D)J recombination and telomere maintenance.

The process of V(D)J recombination takes place in developing T- and B-lymphocytes, where it is responsible for the rearrangement of immunoglobulin (Ig) or T-cell receptor (Tcr) gene segments. This generation of novel combinations accounts for the immune systems' vast potential to recognize a wide range of immunogenic threats, and is, therefore, of crucial importance for the development of both cellular and humoral immunity. New, active Ig or Tcr genes are created by the combination of three different types of gene segments: variable (V), diversity (D), and joining (J) segments. This requires the introduction of an incision at the edge of the segment and the subsequent religation of this segment with its selected partner through a process that is essentially identical to the repair of a DSB (see Figure 2).

The borders of the gene segments are marked by recombination signal sequences which are recognized by the two nucleolytic enzymes that mediate the incision: the RAG1 and RAG2 proteins (recombination-activating gene, RAG). The incision made by the RAG proteins typically leaves hairpin structures at the termini of the segments, which are called 'coding ends'. Before recombination and ligation of the gene segments can take place, these coding ends need to be opened by the Artemis–DNA-PK_{CS} dyad. Hairpin opening during V(D)J recombination requires activation of the Artemis endonuclease activity by association of this enzyme with the DNA-PK_{CS} kinase. Subsequent ligation of the cut and processed gene segments is carried out by the NHEJ core enzymes, which tether and join the DNA fragments as if repairing a DSB.

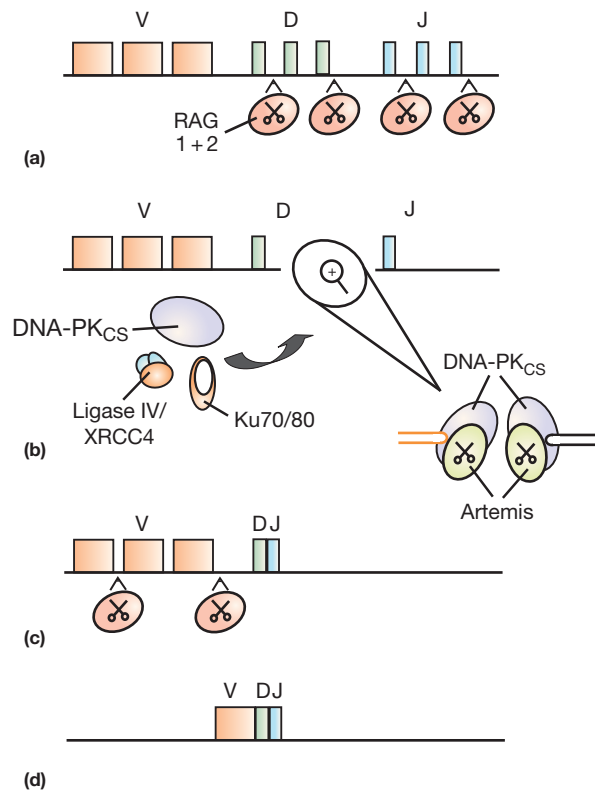


Figure 2 Working model for V(D)J recombination. (a) Creation of incisions at the recombination signal sequences by RAG1 and RAG2. (b) Processing of hairpin structures at the coding ends and ligation facilitated by the NHEJ core enzymes. (c, d) Completion of the V(D)J recombination process.

The involvement of the NHEJ core enzymes Ku70/80 and DNA-PK_{CS} in telomere maintenance is yet another testimonial for the general importance of this set of enzymes. Telomeres are defined as specialized structures at the ends of chromosomes which function to protect the chromosomes from degradation, recombination, or fusion. They consist of stretches of repeats of the TTAGGG sequence, which associate with telomere binding proteins and fold into a T-loop structure. Both Ku70/80 and DNA-PK_{CS} have been found to associate with the telomeres, most likely through direct binding of Ku to telomere repeat binding proteins. Disruption of this association by (null) mutation of the Ku or DNA-PK_{CS} proteins results in loss of telomere equilibrium. More specifically, the absence of Ku70/80 causes shortening of the telomeres, whereas the absence of DNA-PK_{CS} leads to elongation of the telomere sequence. Fusion of telomeres is observed after disruption of either Ku or DNA-PK_{CS}.

The mechanism by which these NHEJ factors mediate telomere protection is at present far from clear. The same can be said about the biological significance of telomere alteration, which is associated with such contrasting events as cellular senescence, aging processes, and tumorigenesis. The somewhat surprising involvement of NHEJ factors in the prevention of telomere joining events (theoretically the very opposite of the NHEJ core function: joining) will remain to be the target of intensive research for the immediate future.

NHEJ Enzymes in Human Diseases

Human diseases associated with NHEJ defects are invariably rare. Null mutations in most of the core NHEJ genes – with the notable exception of Artemis – are not observed in humans, indicating that the resulting phenotypical presentations may be too severe to be compatible with (human) life. Consequently, patients normally harbor only a hypomorphic mutation which allows for at least a residual activity of the affected factor. The few patients who were diagnosed with NHEJ defects generally suffer from a compromised immune system because corruption of the V(D)J process results in impairment of both cellular and humoral immunity; a condition that is often referred to as severe combined immune deficiency syndrome (SCID). When caused by mutations in NHEJ genes, the SCID condition is normally accompanied by increased sensitivity to ionizing radiation and is called radiosensitive SCID (RS-SCID).

At present, no patient with a Ku70 or Ku80 defect has been identified. The absence of a viable human phenotype may be indicative for the importance of the DNA–Ku scaffold during NHEJ and V(D)J recombination. Knock-out mouse models, however, are not embryonically lethal and display an RS-SCID phenotype, growth retardation, and a predisposition for the development of hematological malignancies. These presentations are in good agreement with the theoretical expectations for a human clinical phenotype should it ever be discovered.

Only one case of a hypomorphic DNA-PK_{CS}-defective RS-SCID patient has been documented so far. This patient presented with opportunistic oral and respiratory tract infections since the first months of life, which ultimately resulted in respiratory distress. The cause was determined to be a single missense mutation of the 3062 residue of DNA-PK_{CS}, thought to impair activation of Artemis and, consequently, the proper processing of coding end hairpins during V(D)J recombination. Patients that harbor mutations in the actual catalytic domain of DNA-PK_{CS} have as of yet not been reported.

Viable Artemis defects are relatively much more common and constitute the large majority of human RS-SCID cases. Both null- and hypomorphic mutations have been identified in human patients, all causing hypersensitivity to opportunistic infections, mainly of the respiratory tract. Null-mutations obliterate the maturation of B- and T-lymphocytes (T[−]B[−] SCID),

whereas hypomorphic mutations lead to profound but less severe lymphocytopenia. In addition, genomic instability and predisposition to the development of B-cell lymphomas are commonly associated with Artemis mutations. The existence of Artemis null-mutation patients suggests that the Artemis protein is not absolutely essential to life.

Hypomorphic mutation in the ligase IV gene results in LIG4 syndrome, which is characterized by immunodeficiency and often by microcephaly and growth retardation. The severity of these symptoms is reciprocally proportional to the residual ligase activity allowed by the mutation. A recent study has shown that reduction of ligase IV activity can result in prolonged processing of hairpin ends during V(D)J recombination, resulting in the loss of nucleotides around the joint which subsequently interferes with lymphocyte maturation. The phenotype of XLF patients resembles that of the LIG4 syndrome and is characterized by RS-SCID, dysmorphic features, growth retardation, and microcephaly.

Defects in the DSB signaling protein ATM give rise to the neurodegenerative disease ataxia telangiectasia (AT), which is characterized by the appearance of pronounced vasculature in the eye whites, loss of Purkinje cells in the cerebellum, and, consequently, marked impairment of motor skills. The aberrations that are at the basis of AT onset can be either hypomorphic or null mutations. The disease usually becomes apparent in the toddler years and is often, but not always, accompanied by increased sensitivity to ionizing radiation. Although ATM is not directly involved in NHEJ-mediated V(D)J recombination, AT patients often have a compromised immune system, the symptoms of which do not quite reach the severity of a SCID condition. Like most other NHEJ defects, AT is often accompanied by genomic instability and an increased incidence of hematologic malignancies. The neurological manifestations of the disease may be explained by non-NHEJ functions of the ATM protein, including its involvement in apoptosis and cell-cycle regulation (Table 1).

Inhibitors of the NHEJ Process and Radiation: Combination Therapies

Alleviation of the primary tumor burden is a common starting point for the treatment of most cancers, including metastasized

Table 1 The nonhomologous end-joining core factors

<i>NHEJ factor</i>	<i>Mol. weight^a</i>	<i>Discovery</i>	<i>Crystal structure</i>	<i>Primary interactions</i>	<i>Associated human disease</i>
Ku70	73	1981 ^b	Yes	Ku80	No patients known
Ku80	86	1981 ^b	Yes	Ku70	No patients known
DNA-K _{CS}	469	1990 ^c	No	Ku, Artemis	RS-SCID
Artemis		2001	No	Ku, DNA-PK _{CS}	RS-SCID, genomic instability, lymphomas
XRCC4	38	1990	Yes	Ligase IV, Ku	Single nucleotide polymorphism is associated with oral cancer
Ligase IV	96	1995	No	XRCC4	RS-SCID, LIG4 syndrome, growth retardation, genomic instability, leukemia
XLF	33	2006	Yes	Ku, XRCC4	RS-SCID, growth retardation, genomic instability
ATM	350	1995	No	P53, H2AX, Chk2, Nbs1	Ataxia telangiectasia

^aMolecular weights reflect literature consensus and are often based on measurements of apparent behavior.

^bOriginal discovery; Ku was not linked to DNA-PK_{CS} until 1990.

^cCoining of the name DNA-PK based on earlier observations (1985, 1988, 1990).

cancers which will require the additional inclusion of a chemotherapeutic regimen. Reduction of the primary tumor is commonly obtained either by surgical means or, more often, by radiation therapy. As modern radiology techniques have become precise enough to deliver a calibrated radiation dose within a specified spatial and temporal perimeter, researchers have been developing compounds that enhance the biological effects of the radiation exposure. Combining a systemic, subcytotoxic dose of a radio-sensitizing compound with strictly localized radiation delivery could theoretically result in a high-potential, tumor-specific therapy.

Because impairment of the NHEJ pathway increases sensitivity to ionizing radiation, current investigations are directed toward small-molecule inhibitors of NHEJ enzymes. At present, these inhibitors are mostly targeting the active sites of the ATM and DNA-PK_{CS} proteins. Because the kinase domains of ATM and DNA-PK_{CS} are structurally different from those of the main stream kinases, the cytotoxic characteristics of its inhibitors should be relatively mild in the absence of ionizing radiation.

Classical inhibitors of ATM and DNA-PK_{CS} include caffeine and wortmannin, a furanosteroid metabolite of a *Penicillium* fungus. However, the medicinal use of these natural compounds was found to be limited due to their low specificity and biological availability. A second generation of small molecular compounds was subsequently designed based on the structure of a compound called LY294002, which led to the development of the specific ATM inhibitor KU-55933 and the DNA-PK_{CS} inhibitor NU7441. At present, the feasibility of inclusion of these and other compounds in existing radiation protocols is being evaluated.

See also: **Molecular Biology:** DNA Damage: Alkylation; DNA Mismatch Repair and Homologous Recombination; DNA Mismatch Repair and the DNA Damage Response; DNA Mismatch Repair in Bacteria; DNA Oxidation; Nonhomologous Recombination: Bacterial Transposons.

Further Reading

- Broccoli D (2004) Function, replication and structure of the mammalian telomere. *Cytotechnology* 45: 3–12.
- de Villartay JP (2009) V(D)J recombination deficiencies. *Advances in Experimental Medicine and Biology* 650: 46–58.
- Friedberg EC, Walker GC, Siede W, et al. (2005) *DNA Repair and Mutagenesis*. Washington, DC: ASM Press.
- Lieber MR, Lu H, Gu J, and Schwarz K (2008) Flexibility in the order of action and in the enzymology of the nuclease, polymerases, and ligase of vertebrate non-homologous DNA end joining: Relevance to cancer, aging, and the immune system. *Cell Research* 18: 125–133.
- O'Connor MJ, Martin NM, and Smith GC (2007) Targeted cancer therapies based on the inhibition of DNA strand break repair. *Oncogene* 26: 7816–7824.
- Reaper PM, d'Adda di Fagagna F, and Jackson SP (2004) Activation of the DNA damage response by telomere attrition. *Cell Cycle* 3: 543–546.
- San Filippo J, Sung P, and Klein H (2008) Mechanism of eukaryotic homologous recombination. *Annual Review of Biochemistry* 77: 229–257.
- Shrivastav M, De Haro LP, and Nickoloff JA (2008) Regulation of DNA double-strand break repair pathway choice. *Cell Research* 18: 134–147.
- Soulas-Sprauel P, Rivera-Munoz P, Malivert L, et al. (2007) V(D)J and immunoglobulin class switch recombinations: A paradigm to study the regulation of DNA end-joining. *Oncogene* 26: 7780–7791.
- Weterings E and Chen DJ (2008) The endless tale of non-homologous end-joining. *Cell Research* 18: 114–124.

Nonhomologous Recombination: Bacterial Transposons

B T Hoang and M G Chandler, CNRS, Toulouse, France

© 2013 Elsevier Inc. All rights reserved.

Glossary

DDE transposases Transposases in which the catalytic site includes two aspartate (D) and one glutamate (E) amino acid residues.

S transposases Transposases that resemble serine (S) site-specific recombinases.

Transposases Enzymes that catalyze transposition.

Y transposases Transposases that resemble tyrosine (Y) site-specific recombinases.

Y1 transposases Transposases that resemble rolling circle replicases and in which the catalytic site includes a single tyrosine (Y) amino acid residue. Member of the HUH protein family.

Y2 transposases Transposases that resemble rolling circle replicases and in which the catalytic site includes two tyrosine (Y) amino acid residues. Member of the HUH protein family.

Transposable elements (TEs) are small functional segments of DNA capable of translocation into DNA sites within and between genomes and play an important role in horizontal gene transfer. A major division between TEs is whether they translocate via RNA or DNA intermediates. Prokaryotic TEs are probably exclusively limited to those that move using DNA intermediates. They are capable of sequestering and transmitting a variety of genes involved in accessory cell functions, such as resistance to antimicrobial agents, catabolism of unusual compounds, and pathogenicity, virulence, or symbiotic functions. They also promote deletions and inversions and their insertion can activate or extinguish gene expression. TEs can constitute an important fraction of bacterial genomes (both chromosomes and plasmids). For example, one class of TE, bacterial insertion sequences, represents ~7% of the 4 Mbp *Bordetella pertussis* chromosome and ~40% of the 220 kbp *Shigella flexnerii* virulence plasmid pWR100.

Structural Similarities and Diversity

All TEs encode a cognate transposase (Tpase) or recombinase, which recognizes their ends and catalyzes the individual DNA cleavages and strand transfer, events necessary for transposition. In bacteria, the term 'transposon' (Tn) is limited to elements that carry genes specifying additional functions to those involved in transposition. A large number of TEs, which do not include a distinct phenotypic trait, are known as insertion sequences (ISs).

Definition of Transposases

Different types of elements can be defined by the nature of their cognate recombinases. In transposition reactions, these enzymes are known as 'transposases' (Tpases). The majority falls into a class generally called 'DDE transposases' because of the presence of three key amino acids, two aspartate residues and a glutamate, as part of the active site. Although DDE enzymes represent the majority of known Tpases, at least four

additional types of Tpase have been identified. One is similar to rolling circle replicase used by certain bacteriophages and plasmids. These are known as 'Y2 transposases' since their catalytic mechanism involves two highly conserved tyrosine residues. A second, the 'Y1 transposases', is related to the Y2 group but uses a single catalytic tyrosine. The other two are members of the site-specific recombinase families. However, while these Tpases show site specificity for the ends of the donor Tn, they appear more flexible than classic site-specific recombinases in the target DNA sequences they recognize and use. They are known as Y-Tpases and S-Tpases to distinguish them from tyrosine and serine recombinases.

Types of Transposable Elements

Insertion sequences

The smallest autonomous TEs known are the bacterial ISs and are present in almost all eubacterial and archaeal genomes. They are genetically compact (750–2500 bp), generally flanked by two partially identical and inverted sequences (terminal inverted repeats: IRs) and generally encode only functions involved in their mobility. The Tpase gene is composed of one or two open reading frames (*orfs*) occupying almost the entire length of the IS (**Figure 1(a)**). The IR proximal to the Tpase promoter is defined as IRL. In many ISs with two *orfs*, Tpase is produced as a fusion protein using a programmed –1 translational frameshift between the *orfs* similar to that observed in generating the pol polyprotein in some retro-viruses.

Our perception of ISs is gradually changing. While they were thought to encode only the Tpase gene, an increasing number are now being identified which include passenger genes such as transcription regulators, methyltransferases, and antibiotic resistance genes. These elements, simpler than known Tns, are called transporter ISs (tISs). Moreover, the number of IS91 family members has expanded enormously, and the Tpase has now been identified associated with a large number of antibiotic resistance genes in an extensive group of elements called IS with a common region (ISCR). The common region is the IS91-like Tpase. Thus, the dividing line between Tns and ISs is becoming less clear.

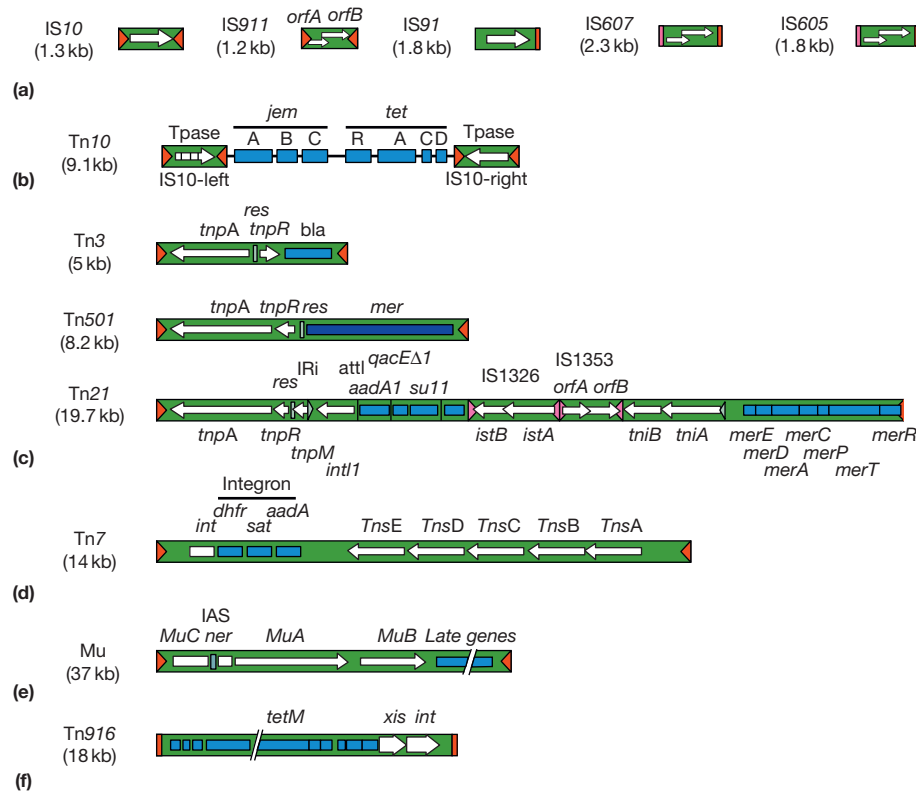


Figure 1 Diversity of bacterial transposable elements. The organization of different types of bacterial TE is shown. Terminal repeats (red triangles), transposon ends (red boxes in the case of IS91, IS605, IS607 and Tn916 to indicate that they are not inverted repeats), genes whose products are involved in transposition, site-specific integration or recombination (int, xis, R) functions (white boxed arrows), resistance genes, and accessory genes (conjugative transfer or phage functions) (blue boxes) are indicated. Gray boxes indicate recombination sites for resolution of cointegrates (*res*) or integron capture of gene cassettes (*att*). *bla*: resistance to β -lactams; *tet*: resistance to tetracycline; *mer*: mercury resistance.

Nearly 4000 bacterial ISs are known. These have been divided into 20–25 groups or families defined by their genetic organization, similarities in their IRs and Tnpase sequences, and their preference for given target DNA sequences. The majority of these families encode a DDE Tnpase. Another (IS91) encodes a Y2 enzyme, a third (IS607) encodes an S-Tnpase while a fourth (IS605/IS608) specifies a Y1 enzyme (Figure 1(a)). This classification will undoubtedly continue to evolve as more are characterized.

Miniature inverted repeat transposable elements

Several bacterial genomes harbor many copies of small DNA segments. Depending on the bacterial species, these are variously called *Rep*, *Box*, *Rup*, or *Correia*. Some, such as *Correia*, resemble ISs, which have been deleted for the Tnpase. While others, such as *Rep*, show structural similarities with members of quite a different family, IS200/IS605. Interestingly, the ends often resemble those of members of the IS630 family and like these elements are flanked by a directly repeated TA dinucleotide. These were first described in eukaryotic systems. Some of these derivatives also include passenger genes and are called MICs.

Compound transposons

Antibiotic resistance genes in bacteria are often flanked by a pair of similar IS elements. Flanking ISs can mobilize any

interstitial DNA segment. These structures are known as ‘compound transposons’. They are widespread and can include resistance to many different antibacterials, to genes involved in bacterial–host interactions as well as to other accessory genes, which are important to the host under some conditions of growth. Classical examples are tetracycline resistance Tn, Tn10 (flanked by two IS10 elements in inverse repeat) (Figure 1(b)), the neomycin resistance Tn, Tn5 (flanked by two IS50 elements in inverted repeat) and the chloramphenicol resistance Tn, Tn9 (flanked by two copies of IS1 in direct repeat). In some cases, the autonomy of the flanking ISs may be reduced in various ways to favor transposition of the entire structure.

Unit transposons

Another class of bacterial Tns are the unit Tns such as members of the Tn3 family (Figure 1(c)). Here, additional genes are contained within the transposable element itself. Large numbers of such Tns have been described carrying antibiotic resistance genes or genes involved in catabolism of unusual substances (including xenobiotics). Tn3-family transposons share very similar IRs (~40 bp) and carry highly related DDE Tnpase genes. They carry a specific site (*res*) and encode a site-specific recombinase (*tnpR*), which promotes recombination in the cointegrate between the two elements at their *res* sites. This leads to separation of donor and target molecules and

leaves a single copy of the Tn embedded in each. Many variations on this basic structure have been observed. They include derivatives, which encode only the T_pase with no additional genes and therefore might be included within the IS group.

Tn3-family elements can accommodate and accumulate additional genes. For example, Tn21 (Figure 1(c)) evolved by insertion of a now defective Tn into an ancestral mercury resistance Tn, which has additionally acquired two ISs. Tn21 also carries a genetic trap, which permits the acquisition of additional genes: the integron system. Integrons include an integrase gene, related to the λ -recombinase of bacteriophage λ . It also carries a flanking site equivalent to the λ attachment site, *att*, at which integration occurs. The acquired genes are generally promoterless and are located on small cassettes, which integrate site specifically into the *att* site where a resident promoter is appropriately located to drive expression.

Tn7

Transposon Tn7 (Figure 1(d)) is unusual in that it encodes five proteins necessary for its transposition and uses two transposition pathways: a site-specific insertion pathway and a more random pathway, which recognizes aspects of DNA structure in the target. One of the transposition enzymes (TnsB) is a DDE T_pase while a second (TnsA) shows similarity to restriction endonucleases. Two additional proteins (Tns D and E) are involved in target choice for each of the two pathways while the fifth (TnsC) provides a bridge between target protein and the T_pase. Tn7 also carries an integron system, which has permitted it to acquire resistance to trimethoprim among other antibiotic resistances. Other members of this group including different resistance genes have also been described.

Transposable bacteriophage

The first example of this type to be described was the temperate mutator phage Mu (Figure 1(e)), so-named because it inserts in a relatively random way and generates mutations during infection. Many different examples of mutator phages have been identified as either phage particles or during annotation of bacterial genome sequences. They are extremely large (>40 kbp) and relatively complex since they must encode for the enzymes and structural proteins involved in the bacteriophage lifestyle. They transpose using DDE enzymes.

Conjugative transposons

These were first discovered in Gram-positive bacteria as elements capable of transmitting tetracycline resistance from cell to cell and between bacterial species. Conjugative Tns (ICE for integrative and conjugative elements) use a site-specific recombination reaction for integration and excision. The enzymes are generally Y-T_pases, similar to the site-specific recombinases used by phage λ or the integrons. However, whereas λ and integrons show pronounced sequence specificity for integration, conjugative Tns are much less exacting in their choice of integration sites. They are large and, in addition to the genes involved in integration and excision and in antibiotic resistance, they carry genes required for conjugative transfer. Tn916 (Figure 1(f)) and Tn1545 are the best-characterized Tns of this type. Several examples of ICE encoding S-transposases have also been described.

Mechanism

Transposition requires precise cleavage at the Tn ends and their coordinated transfer into a target molecule. This is achieved through assembly of a nucleo-protein complex of defined architecture: the transpososome. This structure includes the T_pase, the Tn ends and target DNA. The signals for recognition and processing by T_pase are located in the Tn ends. Most bacterial ISs carry relatively simple IRs with a single T_pase-binding site. Other Tns show a complex constellation of sites often arranged differently at each end (e.g., phage Mu and Tn7).

DDE Enzymes

The majority of known T_pases is structurally related and carry a conserved triad of acidic amino acids, the DDE motif, which coordinate two divalent metal ions central to catalysis. Cleavage at the 3' ends of the element by an attacking nucleophile (generally H₂O) occurs to expose a free 3' OH group (Figure 2(a)). This nucleophile then attacks a 5' phosphate group in the target DNA in a single-step transesterification reaction (Figure 2(b)). The reaction, analyzed in detail for bacteriophage Mu, HIV, IS10, and IS50, occurs as an inline nucleophilic attack with chiral inversion of the phosphate group. The two metal ions are thought to act respectively as a Lewis acid and base in the reaction.

Concerted transfer of both Tn ends into the target site, while maintaining the correct strand polarity, results in joining each Tn strand to opposite target strands. Since the position of attack in target DNA is generally staggered, this generates short, complementary, single-strand regions, flanking the newly inserted element (Figure 2(c)). Repair by the appropriate host machinery generates a short duplication (direct repeat or DR) of defined length in the flanking target DNA (Figure 2(d)). The reactions do not require an external energy source nor do they involve a covalently linked enzyme-substrate intermediate. The donor Tn strand performs the cleavage-ligation steps in the target DNA and no cleaved target molecule is detected in the absence of strand transfer.

DDE enzymes cleave only one of the two DNA strands. An important feature of the transposition reaction is therefore the way in which the 5' end (second strand) is processed. Some Tns do not process the second strand and are, therefore, not separated from the donor molecule. Strand transfer generates a structure, which can be used as a replication fork. Replicative integration generates cointegrates or replicon fusions where the donor and target replicons are fused and separated at each junction by a directly repeated copy of the element (e.g., bacteriophage Mu, the Tn3 family; Figure 2, left column). Other elements shed their donor DNA prior to insertion. For some elements, the 3' OH generated by the first hydrolysis attacks the second strand to generate a hairpin form which then undergoes hydrolysis to regenerate an active 3' OH (e.g., IS10 and IS50; Figure 2, second column). Other TEs, at present generally of eukaryotic origin, generate an intermediate in which a hairpin structure is similarly formed on the flanking DNA. Yet others possess a second enzyme dedicated to cleavage of the second strand

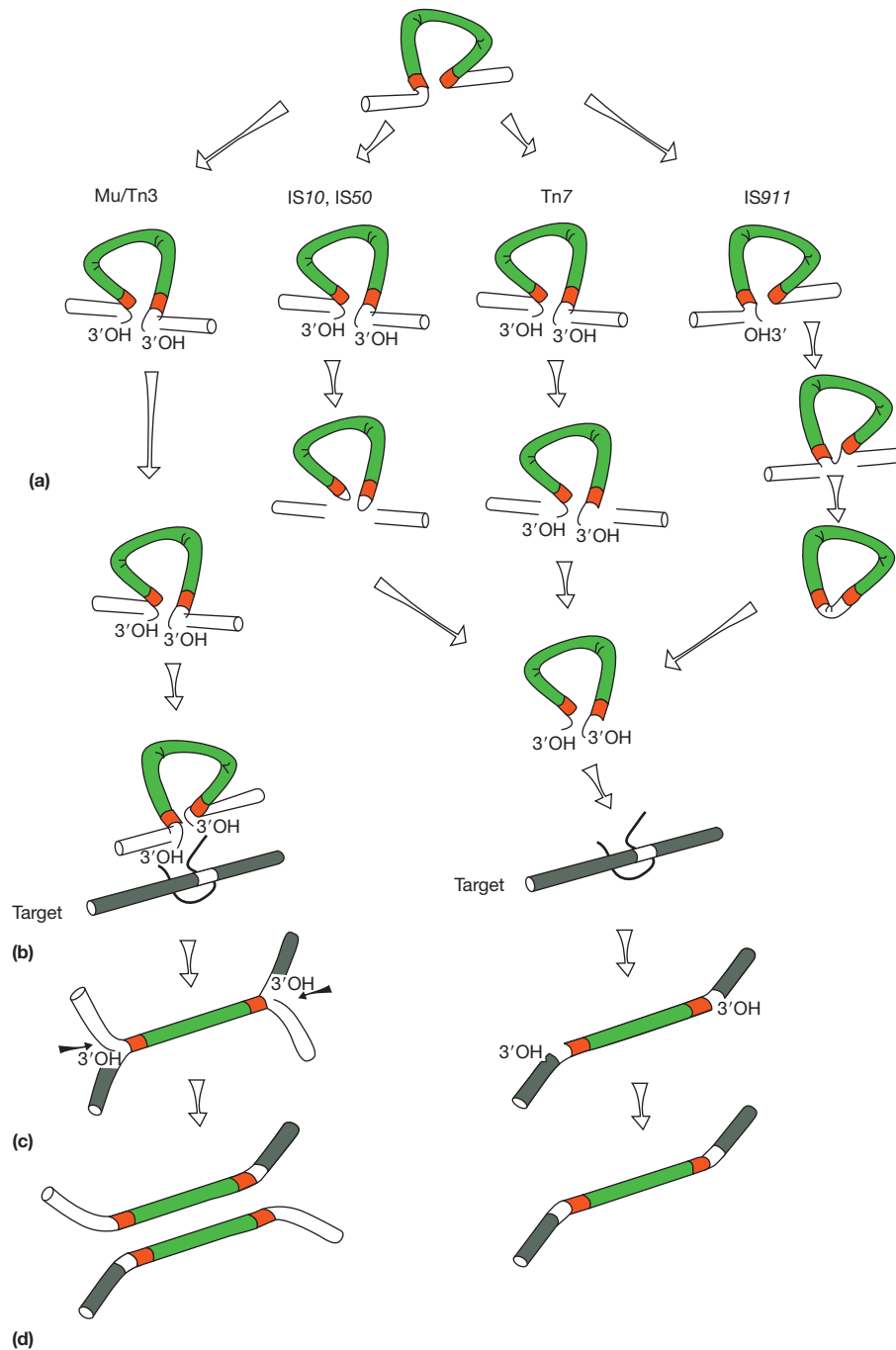


Figure 2 Transposition mechanism of DDE elements. Transposon, inverted repeats, and flanking sequences are shown as green, red, and white, respectively. In target DNA (gray), direct repetitions (DR) are shown in white. Liberated 3'OH groups involved in subsequent strand joining reactions are shown (a, b). Left column: Replicative transposition as the result of single strand cleavage and transfer to create a replication fork (c). 2nd column: In the *IS10* (*IS50*) reaction pathways the 3'OH liberated in the initial hydrolysis attacks the opposite strand to generate a hairpin intermediate, liberating the transposon from its flanking donor DNA (a). A second hydrolysis regenerates the 3' OH (b). 3rd column: Double strand cleavage using two enzymes (e.g., *Tn7*). 4th column: For members of the *IS3* family (e.g., *IS911*), the 3'OH from a single-end hydrolysis directs a strand transfer reaction to the same strand to the other end of the element (a). The branched single-strand circle is resolved into a transposon circle by the host. Single-strand hydrolysis at each 3' end (b) generates a linear transposon, which can integrate in a target.

(e.g., *Tn7*; [Figure 2](#), third column). For members of several bacterial *IS* families (e.g., *IS3*, *IS21*, and *IS30*), the 3'OH liberated at one end of the *IS* attacks the other end on the same strand, generating a form in which one *IS* strand is

circularized. The entire *Tn* is then resolved into a free transient circular form by replication or recombination of the second strand. The *Tn* circle then undergoes efficient integration after single-strand cleavage at each end ([Figure 2](#), fourth column).

Non-DDE Transposases

The mechanism of transposition mediated by the other classes of T_pase is less well detailed than that of the DDE T_pases.

Y2 transposases

The Y2 T_pases, encoded by the IS91 family, show significant identity to the phage and plasmid rolling circle replicases in the regions covering their active sites. The characteristics of transposition are strikingly different from those of DDE-mediated events. ISs of this type do not exhibit terminal IRs, do not generate flanking direct repeats and insert into a conserved tetranucleotide, which is required for further transposition events. The right (but not the left) end and the tetranucleotide are essential for transposition. Moreover, deletion of the left end can result in one-ended transposition. In this case, the inserted fragment of donor DNA is flanked at one end (constant end) by the right end and at the other end by a conserved tetranucleotide sequence present elsewhere in the donor plasmid (variable end). In addition, circular single-stranded derivatives of IS91 have been observed and may represent transposition intermediates. While these data clearly support a polarized rolling circle mechanism, little is known about the strand transfer, which would assure transfer from donor to target site.

Y1 transposases

More is known about Y1 T_pases. These are found in the atypical IS200/IS605 family and are among the smallest known (~150 amino acids). In contrast to classical ISs but like the IS91 family, these elements do not carry IRs at their ends but small (20–25 nucleotides) subterminal hairpin structures located at some distance from the cleavage sites which are recognized by the Y1 T_pase. Cleavage occurs using a catalytic tyrosine residue as the nucleophile to generate a 5′ phosphotyrosine bond between T_pase and its DNA substrate. Transposition requires single-strand substrates and uses single-strand DNA intermediates. It proceeds in two steps: excision of a single strand circle followed by its insertion 3′ to a specific tetra- or penta-nucleotide which is also needed for further transposition. The reaction is strand specific (the top strand is active while the complementary bottom strand is refractory to the T_pase) and both excision and insertion occur preferentially at replication forks or on single DNA substrates created by DNA repair processes.

Y- and S-transposases

Although site-specific recombination reactions directed by S- and Y- recombinases are known in exquisite detail, the activities of the corresponding T_pases remain largely uninvestigated. By analogy, and in contrast to the DDE enzymes, it is assumed that catalysis involves the formation of a covalent enzyme substrate intermediate (Figure 3). In both cases, the element is excised precisely from the donor molecule.

For the Y-transposases, this would involve nucleophilic attack of the target phosphodiester bond by the active site tyrosine to generate a 3′-phosphotyrosyl bond exposing a 5′OH (Figure 3(a), left column). In a second step, the 5′OH attacks the phosphotyrosine bond at the opposite end to promote strand exchange (Figure 3(b), left column). Note that the

reaction of only one strand is shown here. Sequential reactions of cleavage and transfer on the opposite strand would then generate a circular intermediate (Figure 3(c), left column). Insertion is thought to occur by reversal of these reactions (Figure 3(d), left column).

For the S-transposases, recombination occurs by a concerted four-strand cleavage (Figure 3(a), right column) in which a catalytic serine residue attacks the target phosphodiester bond to generate a 5′ phosphoserine bond and to expose a 3′OH (Figure 3(b), right column). Strand transfer to generate a circular intermediate would occur by nucleophilic attacks of the covalent DNA-enzyme bonds by the 3′ hydroxyl ends generated by cleavage (Figure 3(c), right column). A second round of similar recombination reactions (Figure 3(d), right column) would then lead to insertion of the circular transposition intermediates (Figure 3(e)).

Target Choice

The choice of an insertion site can have important implications both for the Tn and for the host. Insertion into a specific target DNA sequence can provide a secure refuge for the Tn. For example, by using a specific sequence in a highly conserved gene in such a way that insertion does not disrupt the gene, transposon Tn7 is assured an insertion site, which does not affect the fitness of the host and occurs in many bacterial species. On the other hand, in excluding insertions into other DNA sequences, this strategy precludes the capacity of the element to generate mutations or to activate or sequester genes and thus to generate genetic diversity. Different transposons exhibit a variety of different strategies for selecting a target. On a genome scale, most transposons appear to insert randomly. However, when analyzed at the nucleotide level, most show some degree of target preference. Insertion can be sequence specific, with target sequence lengths varying from a few to many nucleotides, or show regional preferences, for example, GC- or AT-rich DNA segments, probably reflecting more global parameters such as local DNA structure. Other factors which have also been implicated are: the degree of supercoiling, vegetative and conjugative (transfer) DNA replication (e.g., both Tn7 and members of the IS200/IS605 family are preferentially inserted into extant replication forks), transcription and protein-mediated targeting to, or exclusion from, transcriptional control regions. In addition, some transposons show a preference for insertion close to sequences resembling the transposon end. Others exhibit the converse behavior: exclusion of insertions close to a resident end. This phenomenon, transposition immunity, can act over many kilobases of DNA. For Tn7 and Mu, it has been shown that this is accomplished by a second Tn-encoded protein (in both cases a DNA dependant ATPase), which binds target DNA. Briefly, for phage Mu, the T_pase MuA is capable of stimulating dissociation of the DNA-dependent ATPase, MuB from DNA. Thus, MuA bound to a Tn end resident in the target molecule is thought to provoke disassembly of MuB from neighboring DNA with a probability, which decreases as the distance increases.

Clearly, many insertion preferences must reflect constraints in the architecture of the synaptic complexes formed between the Tn ends, the target DNA and the T_pase.

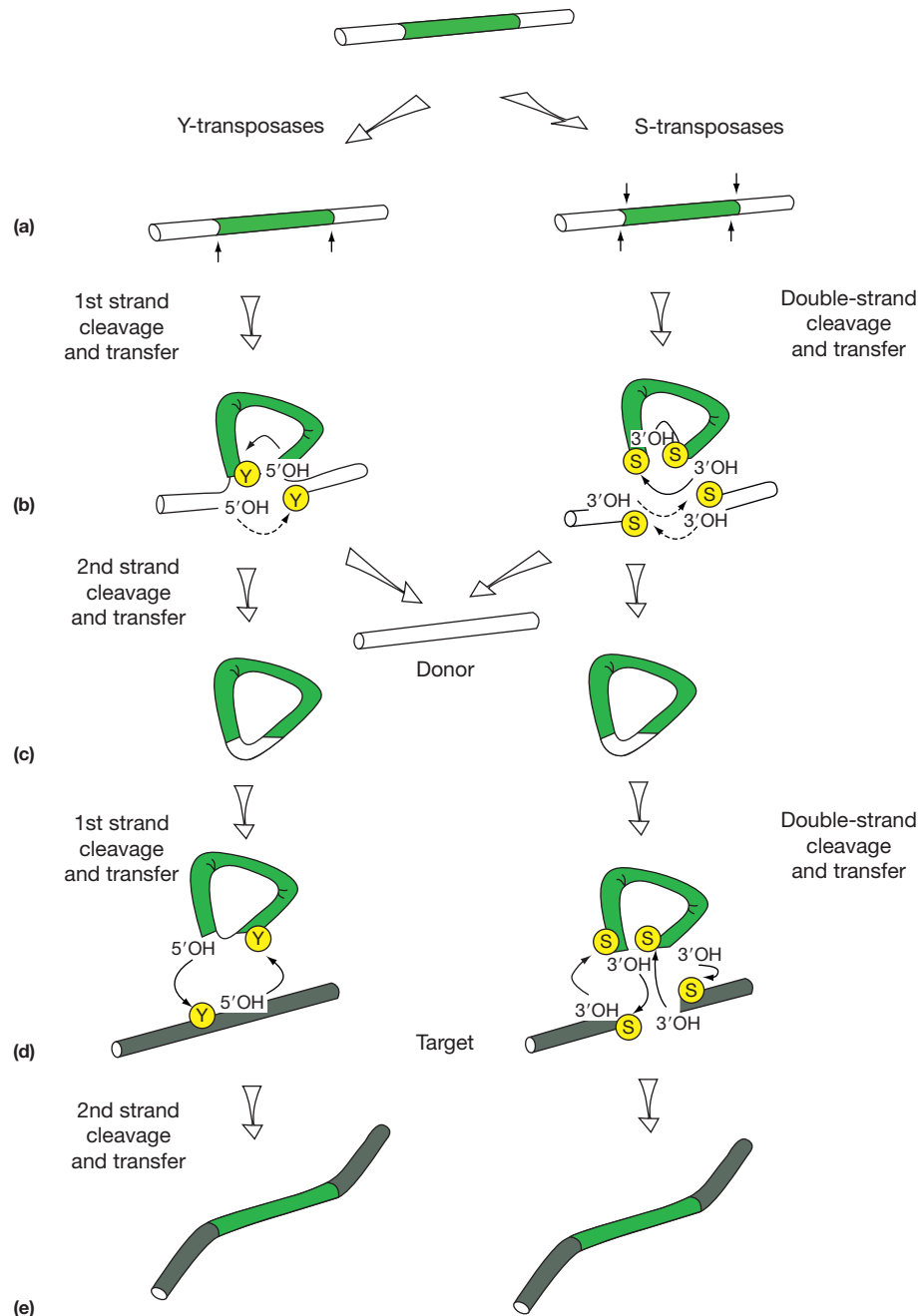


Figure 3 Reaction mediated by Y-transposase and S-transposases. Left column: Transposon and flanking sequences are shown as green and white, respectively. The active site residue Y or S is shown within a yellow circle. Y-transposase nicks one transposon strand (here on bottom strand) at both ends (a) and forms covalent 3' phosphotyrosine intermediates (b). The liberated 5' OH attacks the phosphotyrosine linkage at the other end (arrows) to circularize one transposon strand (b). The same reaction is performed on the top strand to regenerate the donor and a circular intermediate (c). Further cleavage and transfer (d) is presumed to result in an insertion into target DNA (e). In contrast to site-specific recombination, the Y-transposase exhibits low sequence specificity for the target site. Right column: S-transposase catalyzes concerted double-strand breaks at both ends (a), forming covalent 5' phosphoserine intermediates (b). After a switch of strands, the 3' OHs attack the phosphoserine bonds in the opposite end (b), resulting in a circular intermediate (c). The same process (d) permits insertion of the element in a target (e). For both pathways, strand transfers reconstituting donor molecule (b and c) are shown by dotted arrows.

Regulation

Transposable elements are very highly regulated and have evolved a variety of perspicacious regulatory mechanisms. Apart from the temperate transposable bacteriophages, which

possess a highly efficient transposition mechanism during their lytic development, transposition is not an efficient event. A given transposable element generally possesses an ensemble of regulatory mechanisms, which can differ in detail from even closely related elements.

Regulation can be at the gene expression level or at the level of T_pase activity. A combination of weak promoters and poor ribosome-binding sites can limit transcription and translation while secondary structures, which sequester translation initiation sites and involve RNA sequences upstream and downstream of the T_pase promoter, can be present to prevent accidental activation by impinging transcription (IS10, IS5). Some elements produce a small antiRNA also capable of sequestering translation initiation signals, which acts as a negative control component (IS10). Another mechanism to control transposition activity has been documented in elements whose transposition involves a circular or a dimeric intermediate (IS911, IS2, IS30, IS21). Here, the weak endogenous promoter can provide low levels of T_pase to enable low-frequency formation of the Tn circle. However, the IR–IR junction formed in both Tn dimers and circles creates a strong promoter by virtue of a –10 promoter element oriented inward at one IR and a –35 element oriented outward in the other. This boosts T_pase synthesis resulting in cleavage, disassembly of the junction promoter and insertion of the Tn, which then again assumes a low level of T_pase expression from the indigenous promoter. Transcription termination signals within the T_pase *orf* or posttranslational processing might also regulate T_pase expression, although these mechanisms have not been investigated in detail. Another relatively common regulatory mechanism involves programmed translational frameshifting (IS1 and IS3 families). Here, the T_pase is distributed over two consecutive *orfs*, and frameshifting serves to generate a fusion protein including the products of the upstream *orf*, which generally carries DNA-binding and multimerization functions, and the downstream *orf*, which can include the catalytic site. The product of the upstream frame in these cases generally serves a regulatory role. For certain elements, posttranslational proteolysis might also be used to generate small regulatory proteins. IS50 also generates a protein, Inh, using a second promoter located downstream from the T_pase promoter and representing an N-terminal truncated derivative of the T_pase. Inh acts as a T_pase inhibitor.

Regulation can also occur at the level of T_pase activity. The differential arrangement of T_pase-binding sites at each of phage Mu and Tn7 (and perhaps members of the IS21 family) permits a functional distinction and presumably favors the use of one left and one right end *in vivo*. Moreover, in many cases, T_pase bound at one end catalyzes strand cleavage at the partner end. This ‘trans’ cleavage assures that the correct complex has been assembled before transposition can take place. For Tn7, the cleavage and strand transfer reactions require the presence of the target DNA molecule.

An additional level of control of T_pase activity is reflected in the observation that these enzymes often prefer to act on the copy of the element from which they are synthesized. In other words, they exhibit a preference for activity in *cis*. Several nonexclusive explanations for this behavior are apparent and *cis* activity is likely to involve a combination of different effects. One invokes cotranslational binding in which a nascent T_pase protein is capable of correctly folding its DNA-binding domain and binding to Tn ends before synthesis of the entire protein is accomplished. This is a mechanism which appears to act in the case of IS911 and probably for many other ISs. A second model invokes T_pase stability and it has been demonstrated in at least

one case, IS903, that the Lon protease plays a crucial role in *cis* activity. There are also strong indications that protein chaperones such as GroE S/L and DnaK/J are involved in regulation at this level.

Other host factors can also intervene at different regulatory levels. DAM methylase is involved in regulating T_pase expression and the recombination activity of the ends both IS10 and IS50. DNA architectural proteins such as IHF, Fis, and Hu have also been implicated in assuring the correct assembly of the transpososome. Moreover, IHF has been implicated in driving the molecular transitions leading to strand cleavage in IS10. Other factors such as RecBC, DnaA, Acyl carrier protein, and ribosomal protein L29 have also been implicated in the transposition behavior of various elements, but the exact details remain to be elaborated.

Eukaryotic Transposons

Although this section is devoted to prokaryotic elements, it is important to note that eukaryotes generally carry large numbers of both intact and truncated transposable elements. These constitute over 40% of the human genome and somewhat less in Arabidopsis. There are many similarities between prokaryotic and eukaryotic elements. A major group is the Tc/mariner superfamily whose members are similar to bacterial ISs. They carry terminal inverted repeats and encode a DDE or, in this case, a DDD group T_pase and resemble the IS630 family of bacterial elements. Many eukaryotic genomes, especially those of plants, harbor extremely high numbers of truncated derivatives, similar to those found in bacteria, in which the T_pase gene appears to have been deleted leaving both IRs. These elements have also been called miniature inverted repeat transposable elements (MITEs) and can presumably be mobilized by the full length copies of the elements located elsewhere in the genome.

Another major group, which is limited to eukaryotes, are the retro-viruses and LTR-retro-Tns, which pass through an RNA intermediate during their (viral) life cycle. Integration into the host genome occurs through a double-stranded cDNA intermediate and uses DDE group integrase (IN).

The immunoglobulin V(D)J recombination system represents another important element in vertebrates. Here, rearrangements involve two host proteins, RAG1 and RAG2. The RAG1 recombinase is related to DDE T_pases. An interstitial DNA segment (signal sequence) is excised to permit the rejoining of two flanking sequences (coding sequence). In a similar way to IS10 and IS50, recombination involves the formation of a hairpin intermediate. In this case, however, the hairpin is generated on the flanking DNA rather than on the excised fragment.

A third large family of eukaryotic elements is the hAT superfamily. Although the nature of their T_pase is not clear and little work has yet been undertaken concerning their transposition (which includes *hobo* and *Hermes* from insects, *Ac* and *Tam3* from plants), it seems probable that they have adopted a V(D)J-type of recombination mechanism.

Eukaryotes also harbor other families of IS-like elements including the P element of *Drosophila melanogaster* whose mode of transposition has received much attention due to its use in

mutagenesis, the CACTA family, which includes the plant transposons Tam1 and Tam1, and the En/Spm family.

Other newly recognized classes of eukaryotic elements are the helitrons, which resemble rolling circle Tns of the IS91 family, and Cryptons, DIRS1, and kangaroo elements, which appear to encode Y-transposases.

In addition to eukaryotic elements, which transpose using a DNA intermediate including the retro-viruses, a large class of Tns, which transpose by a completely novel mechanism, has also been identified. These are the target-primed (TP) retro-Tns in which an RNA intermediate is copied into its DNA target via reverse transcription.

See also: **Molecular Biology: Nonhomologous Recombination: Retrotransposons.**

Further Reading

- Chandler M and Fayet O (1993) Translational frameshifting in the control of transposition in bacteria. *Molecular Microbiology* 7(supplement 4): 497–503.
- Craig NL (1997) Target site selection in transposition. *Annual Review of Biochemistry* 66: 437–474.

- Craig NL, Craigie R, Gellert M, and Lambowitz A (eds.) (2002) *Mobile DNA II*. Washington, DC: ASM Press.
- Curcio MJ and Derbyshire KM (2003) The outs and ins of transposition: From mu to kangaroo. *Nature Reviews: Molecular Cell Biology* 4: 865–877.
- Guynet C, Achard A, Ton Hoang B, et al. (2009) Resetting the site: Redirecting integration of an insertion sequence in a predictable way. *Molecular Cell* 34: 1–8.
- Hallet B and Sherratt DJ (1997) Transposition and site-specific recombination: Adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiology Reviews* 21(supplement 2): 157–178.
- Haren L, Ton-Hoang B, and Chandler M (1999) Integrating DNA: Transposases and retroviral integrases. *Annual Review of Microbiology* 53: 245–281.
- Hickman AB, Chandler M, and Dyda F (2010) Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Critical Reviews in Biochemistry and Molecular Biology* 45: 50–69.
- Mizuuchi K (1992) Polynucleotidyl transfer reactions in transpositional DNA recombination. *Journal of Biological Chemistry* 267(supplement 30): 21273–21276.
- Mizuuchi K (1997) Polynucleotidyl transfer reactions in site-specific DNA recombination. *Genes Cells* 2(supplement 1): 1–12.
- Rice P, Craigie R, and Davies DR (1996) Retroviral integrases and their cousins. *Current Opinion in Structural Biology* 6(supplement 1): 76–83.

Relevant Website

<http://www-IS.biotoul.fr> – IS Finder.

Nonhomologous Recombination: Retrotransposons

X Qi and S Sandmeyer, University of California, Irvine, Irvine, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Siew Loon Ooi and Jef Boeke, volume 3, pp. 80–86, © 2004, Elsevier Inc.

Glossary

Bacteriophage A virus or phage that can infect bacteria.

Epigenetic Inheritable properties that are caused by mechanisms other than changes in the underlying DNA sequence.

Exaptation An evolutionary phenomenon that a trait is selected for one function and adapts to perform another function.

Extrachromosomally primed (EP)

retrotransposon Retrotransposon that is reverse transcribed from genomic RNA template prior to association of the cDNA with the integration site.

Hybrid dysgenesis The high rate of mutation in germline cells of *Drosophila* strains resulting from a cross of males with autonomous P elements (P Strain/P cyotype) and females that lack P elements (M Strain/M cyotype). Hybrid dysgenesis causes sterility in the F1 progeny.

Integrase Enzyme that recombines phage or transposable element DNA into a host genome through a polynucleotide transesterification reaction.

Intrinsic immunity A set of cellular-based anti-viral defense mechanisms, notably genetically encoded proteins which specifically target eukaryotic retroviruses. Intrinsic immune proteins are usually expressed at a constant level, allowing a viral infection to be halted quickly.

Retrotransposon Transposon that replicates through the reverse transcription of a RNA copy.

Reverse transcriptase (RT) Enzyme that uses a RNA template to synthesize DNA

RNA interference (RNAi) Pathway through which the Dicer complex generates RNAs of 21–30 nt in length that target other RNAs for translational repression or degradation via the argonaute subunits of the RNA-induced silencing complex (RISC). RNAi can also lead to repression of transcription through modification of DNA or chromatin.

Target-primed (TP) retrotransposon Retrotransposon that reverse transcribes its genomic RNA at the site of integration primed from 3'-OH in host DNA at the integration site.

Target site duplication (TSD) Short DNA sequences present in one copy at the site of integration and directly repeated on both sides of the integrated copy of the transposon. This flanking direct repeat is generated during repair of the gaps created by staggered strand-transfer sites of the transposon DNA to host DNA.

Transposon Nonviral sequences of DNA mobilized through RNA- or DNA-based mechanisms.

Wild type Versions, for example, of transposons or organisms that are in the native state and implied to be functional.

Introduction

Transposons or mobile DNAs are sequences that can move within the genome of a single cell. Mobile genetic elements were discovered by Barbara McClintock in the 1940s, for which she was awarded the Nobel prize in 1983.

Transposons are classified as retrotransposons or DNA transposons based on the mechanisms by which they mobilize. Retrotransposons or class I transposons are transcribed into RNA, reverse-transcribed into a complementary DNA (cDNA), and integrated into the host genome. DNA or class II transposons mobilize via DNA replication-coupled integration into a new site (replicative transposition) or via excision of element DNA and integration at a new site (cut-and-paste). In either case the junctions with host DNA are offset by 4–9 nt on the two host DNA strands resulting in a characteristic direct repeat of the offset sequence flanking the inserted element. This repeat referred to as the target-site duplication (TSD) is a signature of transposon insertions. Insertion site patterns of elements range from highly specific to relatively nonspecific, depending upon the element. These patterns affect the impact of elements on host genome function.

RNA and DNA elements are found in all three domains of living organisms, *Archaea*, *Bacteria*, and *Eukarya* (Figure 1). In model organisms with relatively compact genomes, including bacteria (*Escherichia coli*), yeast (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*), fruit fly (*Drosophila melanogaster*), nematodes (*Caenorhabditis elegans*), and mustard weed (*Arabidopsis thaliana*), transposable elements comprise less than 10% of the genome, with variation in the representation of class I and class II elements. For example, *Saccharomyces*, *Drosophila*, and *Arabidopsis* contain mostly class I elements, but bacteria and worms contain mainly class II elements. It is not understood what accounts for this difference in representation of elements. However, one important distinction is that expansion of class I element copy number is independent of host cell replication. As a result, in some organisms, retrotransposons that have undergone expansion can become a significant factor contributing to larger genome size. For example, the human genome, which is 10 times larger than the fruit fly genome, has less than twice as many genes, and is almost half retrotransposons. In some plants there is an even greater proportion of retrotransposons. For example, the maize genome is 100 times larger than the fly genome and is 85% retrotransposons.

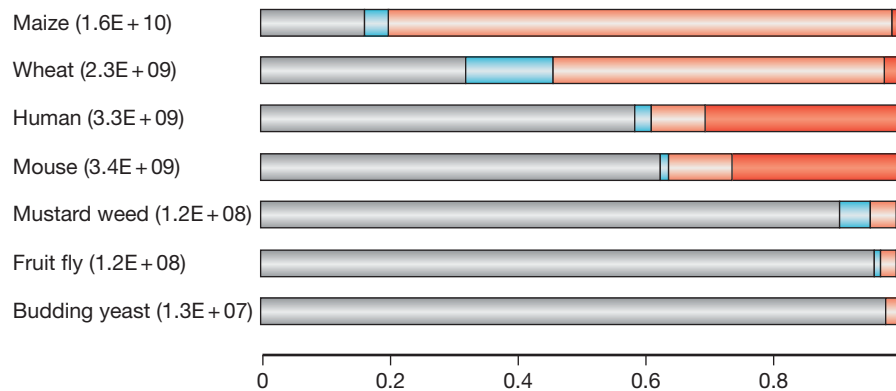


Figure 1 Percentages of model organism and human genomes composed of retrotransposons (class I) and DNA transposons (class II). Organisms are maize, *Zea mays* ssp. *mays* L; wheat, *Aegilops tauschii*; human, *Homo sapiens*; mouse, *Mus musculus*; mustard weed, *Arabidopsis thaliana*; fruit fly, *Drosophila melanogaster*; budding yeast, *Saccharomyces cerevisiae*. Genome sizes in bp are included in parentheses for each species. Percentages of transposons are shown as bars in various lengths, with the scale bar shown at the bottom. Color scheme: gray for non-transposon regions, light red for LTR retrotransposons, dark red for non-LTR retrotransposons and blue for DNA transposons.

Retrotransposon Diversity

Retrotransposons can be categorized based on differences among retrotransposition mechanisms which correlate with differences among reverse transcriptase (RT) and retrotransposon genome structures (Figure 2). RT itself is the most conserved of retrotransposon proteins. Mechanistically, there is a fundamental distinction between retrotransposons for which reverse transcription is primed from retrotransposon-associated RNA (extra-chromosomally-primed (EP)) and those for which reverse transcription is primed from host DNA (target-primed (TP)). Examples of the former class include direct, long terminal repeat (LTR) retrotransposons (retrovirus-like), and tyrosine recombinase (YR) elements. Examples of the latter class include the site-specific R2-type elements, long interspersed elements (LINEs), Penelope-like elements (PLEs), and group II introns of bacteria and organelles of lower eukaryotes and higher plants. These elements are discussed in more detail below.

Two additional bacterial classes of sequences, segments of which are replicated by reverse transcription, are worthy of mention and provide a hint of the richness of biology contributed by reverse transcription-related processes. These are bacterial retrons and diversity-generating retroelements (DGRs). Retrtons have similarity to the EP elements described above. The most extensively characterized retron is *E. coli* multicopy single-stranded DNA (msDNA). Retrtons encode RT but reverse transcribe only a segment of the retron transcript primed from a guanidylate 2'-hydroxyl in the template RNA itself and then degrade the template portion of the RNA leaving a product in which RNA and DNA are covalently linked. In the genomic form, retrtons typically occur as passengers in temperate bacteriophage, and the mechanism through which the extra-chromosomal copies integrate into phage or genomic DNA is not known, nor is any functional benefit of the retron known.

DGR elements have similarity to the TP elements described above. The DGR encodes two copies (template repeat and variable repeat) of a subdomain of a *Bordetella* phage tropism protein as well as RT. Reverse transcription of the template repeat generates a cDNA which is hypervariable at each

adenine position. This copy is recombined back into the genome at the site of the variable region copy in a homology-dependent process. DGR confers advantage to the phage in responding to changes in availability of its *Bordetella* host receptor protein.

TP and EP families of retrotransposons each include members that are incapable of autonomous retrotransposition (nonautonomous) in addition to the self-mobilizing elements (autonomous), but for which the RNA transcript can be mobilized *in trans* by proteins encoded in autonomous elements. Interestingly, in the case of TP elements – the proteins of which function efficiently *in trans* to the element genomic RNA – nonautonomous elements are either degenerate copies of active elements or highly-abbreviated copies containing minimal sequences required *in cis* in the RNA and cDNA for replication and integration, respectively. In the case of EP retrotransposons, the proteins of which function most efficiently *in cis* to the genomic RNA, certain elements such as LINEs support mobilization of small unrelated sequences transcribed by RNA Pol III.

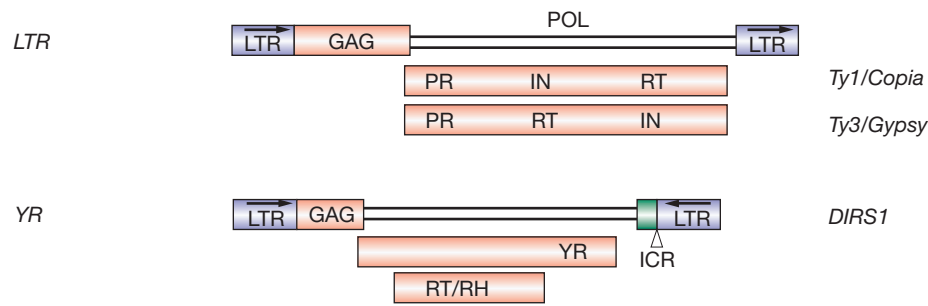
We have mentioned here the major classes of elements and some fundamentally different pathways by which they retrotranspose. Prokaryotic cells were used to exemplify nonconventional retrotransposon systems that allow directed feedback from RNA into the genomes of host organisms. One of the distinguishing properties of reverse transcription is its error-prone nature, and of course, as genomic parasites, retrotransposons are relatively free to diversify. The advent of high-throughput sequencing has driven a rapid increase in transposon discovery including novel varieties, suggesting that we have only begun to sample the diversity of these fascinating denizens of our genomes.

Extrachromosomally Primed Retrotransposons: Replication and Integration

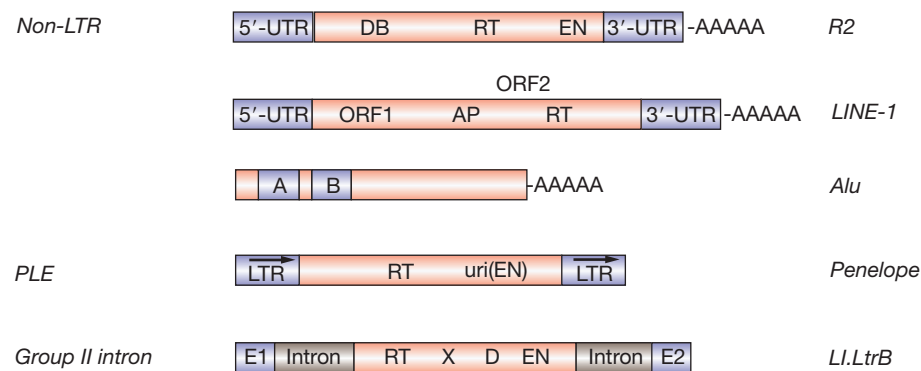
LTR Elements

EP retrotransposons are transcribed into RNA, which templates production of proteins required for retrotransposition and serves as the template for reverse transcription. LTR elements

Extrachromosomally-primed elements



Target-primed elements



Bacterial retroelements

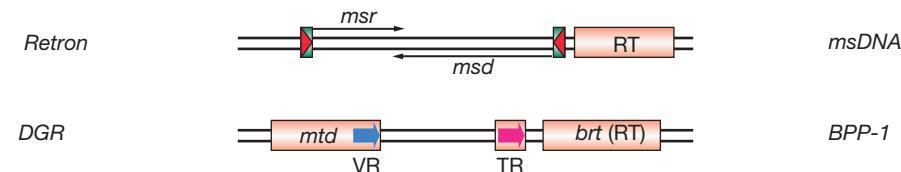


Figure 2 Genome structures of representative retrotransposons. Extrachromosomally-primed retrotransposons include LTR retrotransposons (e.g., *S. cerevisiae* Ty1 and Ty3) and tyrosine recombinase (YR) retrotransposons (e.g., *Dictyostelium discoideum* DIRS1). Target-primed retrotransposons include non-LTR retrotransposons – R2, LINES, and SINES (e.g., *Drosophila simulans* R2, *H. sapiens* LINE-1 (autonomous) and Alu (non-autonomous)) – Penelope-like retrotransposons (PLEs) (e.g., *Drosophila virilis* Penelope), and group II introns (e.g., *Lactococcus lactis* LI.LtrB). Bacterial retrotransposons include retrons (e.g., *E. coli* msDNA; retron RT-coding region is not to scale) and diversity-generating retroelements (DGRs) (e.g., *Bordetella* BPP-1).

are named after the LTRs flanking one (*S. pombe* Tf1) or two open reading frames (ORFs) (*S. cerevisiae*, Ty1, Ty3) in the DNA copy. These upstream and downstream repeats contain the promoter and terminator signals, respectively, for production of the genomic RNA. The central ORF(s) encode polyproteins that are processed into structural (capsid and RNA binding protein) and enzymatic (protease (PR), RT, and integrase (IN)) proteins necessary for retrotransposition (Figure 2). The budding yeast Ty retrotransposon (e.g., Ty1 and Ty3) replication cycle (Figure 3) illustrates replication of an LTR retrotransposon. Levels of the polyproteins are tightly regulated, resulting in a high ratio of structural to catalytic protein

that facilitates formation of virus-like particles (VLPs). In the case of single ORF elements, this ratio is maintained by regulated processing. In the case of overlapping ORFs, translation of the first ORF occurs normally, but similar to retroviruses, the second ORF is translated dependent upon programmed frameshifting that occurs at the downstream end of the first ORF. Because frameshifting is not efficient, a relatively low amount of structural protein fused to catalytic protein is produced. This not only means that protein is produced in appropriate proportions, but catalytic protein domains are included in the VLP by virtue of their fusion to structural protein. The RNA-binding domain of the structural protein binds to motifs in genomic

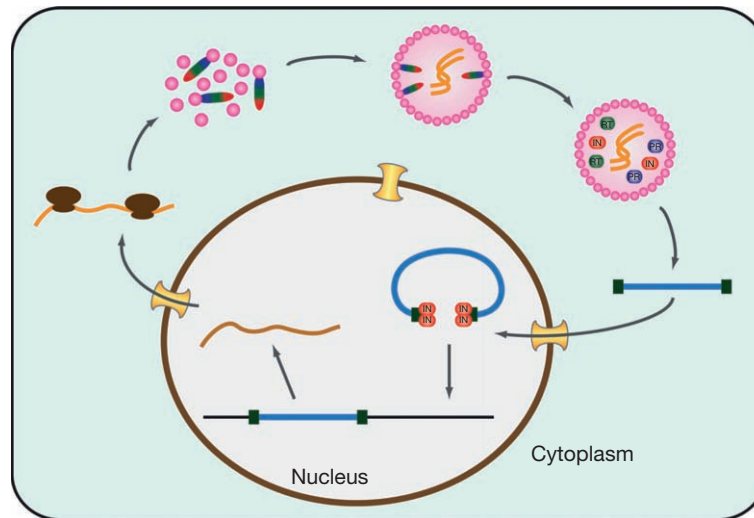


Figure 3 The replication cycle of a Ty LTR retrotransposon. Ty is transcribed in the nucleus to produce an almost full-length, polyadenylated RNA that is exported to the cytoplasm. This RNA serves translation template and genomic RNA functions. The upstream frame is designated *GAG* and downstream frame is *POL*, named after the analogous retrovirus ORFs. It is translated into Ty proteins: Gag and Gag-Pol that assemble around the Ty genomic RNA to form virus-like particles (VLPs). Upon assembly, PR cleaves the Gag3 into capsid and RNA-binding proteins and Gag-Pol polyprotein into PR, RT, and IN within the VLPs. Reverse transcription occurs associated with VLPs. The preintegration complex is imported into the nucleus, and IN mediates integration into new chromosomal locus, thereby creating a new copy of the retrotransposon.

RNA resulting in its assembly into VLPs. VLP formation activates the encoded PR which processes structural protein into capsid and RNA-binding protein and the downstream domain into PR, RT, and IN.

In association with VLP components, genomic RNA is reverse transcribed into cDNA through a process similar to that used by retroviruses. Transcription of the genomic RNA initiates in the upstream LTR 5' of the position of termination in the downstream LTR resulting in a RNA containing only one complete LTR copy. Therefore, reverse transcription must regenerate complete LTR sequences at each end. For most characterized LTR retrotransposons, the primer for reverse transcription is a host tRNA that anneals to a short complementary sequence just outside of the LTR sequence near the 5'-end of the genomic RNA. Through a process involving terminal transfers of cDNA intermediates, reverse transcription regenerates complete LTRs at each end of a product duplex cDNA. In a reaction catalyzed by the IN protein, hydroxyl groups at the two 3'-ends of this extrachromosomal duplex DNA attack phosphodiester bonds on the two target DNA strands. Energy is conserved in this polynucleotide trans-esterification reaction and adenosine triphosphate (ATP) is not required. Retrotransposon IN proteins have an active-site motif in common with other enzymes that catalyze polynucleotide trans-esterification reactions, such as DNA transposon recombinases. DNA copies of elements that are integrated in this way terminate in an inverted repeat in the sense strand. This is because each end must be similarly recognized by the IN protein. In addition to this structure, the specific sequence 5'TG...CA3' is also highly conserved. Because the target DNA is not only duplex, but helical, the joints between the element and the host on each strand are slightly offset, creating gaps in the product DNA on the nonjoint strands. These gaps are filled in and ligated by host enzymes, generating characteristic TSDs and completing the integration reaction.

Based on the similarities of retrovirus and LTR retrotransposon genome structure and function, ancient retroelements related to Ty3/gypsy may have captured host genes, adapted them for functions to enable cellular egress and entry, and emerged as retroviruses. Historically, LTR retrotransposons were divided groups designated according to characterized members in yeast and *Drosophila*, respectively: Ty1/copia, and Ty3/gypsy. Ty1 and Ty3 groups differ from each other on the order of the proteins encoded in *pol* gene. Based on this distinction and coding sequence similarity, retroviruses most closely resemble the Ty3/gypsy group. Specific elements in this group (e.g., the gypsy element of *Drosophila*) encode envelope proteins that have been shown to mediate extracellular infection, thus blurring the distinction between retrotransposons and retroviruses. Subsequently, an additional group was discovered in animals and also designated after founding members *pao/BEL*. The list of LTR retrotransposons is likely to continue to grow.

A high proportion of defective full-length LTR retrotransposons is observed in genomes and could theoretically be mobilized *in trans* by proteins encoded by wild-type elements. Many isolated LTRs also exist in genomes. These are generated by recombination between intact retrotransposon terminal LTRs (resulting in deletion of the intervening coding sequence). However, because the primer is typically not contained in the LTR, they do not generate transcripts with the appropriate sequences to allow reverse transcription and therefore are not mobilized by *trans*-acting autonomous functions. Families of nonautonomous, LTR elements with large internal deletions that appear to have been mobilized have been identified in plants. These include large retrotransposon derivatives (LARDs), which contain long noncoding sequences; terminal-repeat retrotransposons in miniature (TRIMs), which contain only very short sequences; and *Morgane*, which contains *pol*-like partial sequences flanked by LTRs.

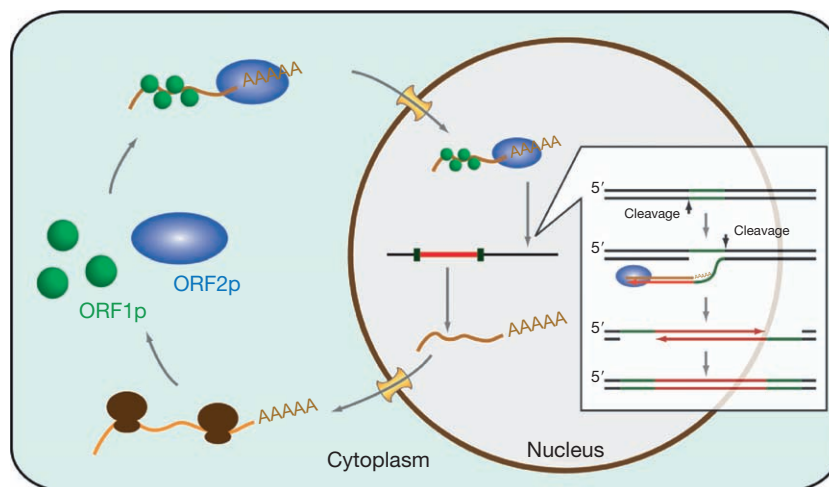


Figure 4 The replication cycle of human LINE-1 (L1) non-LTR retrotransposon. L1 RNA is transcribed and transported to the cytoplasm, where it is translated into ORF1p and ORF2p. The L1 proteins and RNA assemble into a ribonucleoprotein particle which is then transported into the nucleus. L1 copies its RNA into DNA at a new locus via a process called target-primed reverse transcription (TPRT).

Tyrosine Recombinase Elements

The LTRs of EP retrotransposons are typically direct repeats. However the LTRs of YR EP retrotransposons also occur as inverted repeats. YR elements encode structural protein and RT, but rather than a retrovirus-like IN, they encode an enzyme more similar to a bacteriophage lambda IN. This IN mediates integration of lambda bacteriophage at specific attachment sites in the bacterial genome via an intermediate in which DNA is covalently attached to an active site tyrosine. The DIRS1 retroelement from the slime mold *Dictyostelium discoideum* is representative of the YR class of retrotransposons. DIRS1 RNA is thought to be reverse transcribed into a circular cDNA by a mechanism involving internal complementary repeats (ICRs). The circular cDNA is proposed to be integrated into the host genome by YR-catalyzed recombination YR.

TP Retrotransposons: Replication and Integration

R2-Type Elements

The most extensively studied R2-type elements are found in the repeated ribosomal RNA genes of insects where the genomic RNA is actually processed from the rRNA. A generic version of TP retrotransposition is illustrated in Figure 4. The transcript has 5'- and 3'-untranslated regions flanking a single ORF that encodes a multifunctional protein containing an RT domain followed by an EN domain with similarity to type II restriction endonucleases (ENs). Genomic RNA is imported into the nucleus complexed with the multifunctional protein. The EN activity creates a 3'-end from which reverse transcription of the genomic RNA is primed. In a less well-understood reaction, the second strand of target DNA is also nicked and used to prime synthesis of the second strand of the integrated copy. Depending on the position of the second-strand nick with respect to the first-strand nick, either a TSD or a target-site deletion results. Studies of the R2 element led to formation of the TP reverse transcription (TPRT) model of replication for the broader class of TP elements.

Long- and Short-Interspersed Elements (LINEs and SINEs)

In contrast to the R2-type elements, LINEs are found in a broad distribution of species and have a preference for AT-rich sequences. Full-length LINE elements are 4–7 kb in length. Genomic RNA is produced by transcription from an internal RNA Pol II promoter. The predominant human LINE element, L1, is a well-studied example. L1 contains 5'- and 3'-untranslated regions flanking two ORFs. The upstream ORF encodes a protein (ORF1p) with nonspecific RNA-binding and structural domains and the downstream ORF encodes a protein (ORF2p) with apurinic endonuclease (AP) and RT domains. The mechanism of L1 integration is generally similar to the TPRT mechanism described above for R2. However, because insertions are templated by a polyadenylated RNA, they are distinguished from R2 and LTR retrotransposon insertions by the presence of a downstream poly (A) tract on the coding strand. It is notable that most L1 integrants are 5'-truncated and nonfunctional for further replication. This property has been attributed to non-processive L1 reverse transcription and to microhomology between L1 and the target site. There are more than 5×10^5 copies of L1 in the human genome, of which less than 100 are capable of autonomous retrotransposition. Because LINE proteins act predominantly on the RNA from which they are translated, they do not complement the mobility of defective LINES efficiently.

Nonautonomous short interspersed elements (SINEs) are noncoding sequences of 80–400 bp. They contain internal RNA Pol III promoter elements and are derived from ancestral Pol III-transcribed genes for structural RNAs. Interestingly, the SINE-source Pol III-transcribed genes differ among species. The major SINE in primates is the Alu element. These are derived from tandem repeats of the gene for the signal recognition particle 7SL RNA. In many species, SINEs are derived from tRNA genes. There are $\sim 1.4 \times 10^6$ copies of Alu in the human genome, which comprise greater than 10% of the genome. These nonautonomous elements are mobilized by proteins encoded by autonomous LINES. Exactly how they overcome the *cis*-bias of LINES is not known, but it may have to do with

the association of the Alu signal recognition particle RNA with translating ribosomes and nascent LINE proteins. Similar to LINES, these elements display a downstream poly(A) track on the coding strand and are flanked by TSDs of variable lengths. A second active class of primate SINE, called SVA, is about 2 kb in length and represents a composite of a hexameric repeat, a fragmentary Alu-like sequence repeat, a region containing a variable number of tandem repeats of 30 to 50 bp, and part of the envelope gene and LTR of a human endogenous retrovirus. Insertions of SVA sequence also have a downstream poly(A) tract on the coding strand. There are ~5000 SVA elements in the human genome. Similar to Alu elements, they are also thought to be mobilized by LINE activity. Together LINES and SINES constitute almost half of the human genome.

Penelope-Like Elements

PLEs are exemplified by Penelope of *D. virilis* and Athena of *Bdelloid rotifers*. Intact copies contain a single ORF encoding two domains, RT and EN, although the latter is not consistently present. The PLE RT is very distantly related to those of other elements and can be disrupted by an intron, which is a surprising feature for a retrotransposon. Although PLE can have LTRs, in either direct or inverted orientation, many copies are 5'-truncated and have variable-length TSDs. The 5'-truncation of characterized copies and variable presence of the EN domain together with the concentration of this element in telomeric regions in some organisms is consistent with a TP mechanism of replication.

Prokaryotic and Organellar Retrotransposons

TP retrotransposons include the group II introns found in bacterial, as well as organellar genomes of lower eukaryotes (mitochondria) and plants (mitochondria and chloroplasts). These elements mobilize via efficient site-specific retrohoming into double-stranded DNA or inefficient retrotransposition into single-stranded DNA sequences resembling retrohoming sites and associated with replication forks. Mobilization by the former mechanism requires the multifunctional intron-encoded protein (IEP) that has four domains RT, X, DNA-binding (D), and EN. RT and X provide maturase activity that stabilizes the intronic RNA structure required for splicing; X and EN domains are required for retrohoming. EN cleaves the nonspliced DNA strand to create the site for 3'-end priming of the reverse transcribed copy template by the integrated intronic RNA. The RT domain is required for reverse transcription of the intronic RNA. The intronic RNA is subsequently degraded by host RNase H activity. Retrotransposition mobilization into single-stranded DNA is EN independent as reverse transcription of the intron template is primed from lagging-strand replication intermediates.

Integration Site Preference

Genome-wide investigation of retrovirus insertion sites showed that specific viruses exhibit bias toward the promoter or transcribed regions of genes. In contrast to retroviruses,

many retrotransposons display a high degree of insertion site preference, possibly reflecting a requirement to be tolerated by the host for long periods. Integration patterns range from relatively little specificity to a high degree of sequence specificity. These differences are based on the proteins mediating integration, including site-specific restriction enzyme-like EN proteins and IN proteins that interact with host transcription factors or chromatin proteins or in the absence of IN or EN functions, targeting patterns based on availability of DNA ends. Although there may be other influences, it is safe to say that targeting reflects the influence of EN or IN proteins associated with the retrotransposon.

Extrachromosomally Primed Element Targeting

At the most extreme end of the spectrum, the *Drosophila* LTR retrotransposon ZAM IN recognizes a specific target sequence and, similar to a type II restriction enzyme, integrates adjacent to the sequence 5'CGCGCg3' (lowercase 'g' indicates a 50% occurrence of that base). Compared to retrovirus IN homologs, many LTR retrotransposon IN proteins, including those of fungal elements, have extended carboxyl-terminal domains, and several have been shown to mediate interaction and therefore targeting by host transcription factors or chromatin proteins. The *S. cerevisiae* retrotransposons Ty1–4 insert preferentially at Pol III-transcribed genes. Ty1, 2, and 4 are found within 750 bp upstream of these genes, while Ty3 IN interacts directly with Pol III transcription factor subunits and integrates into the transcription initiation site. Because Pol III promoters are internal, this mechanism of specificity tends to insure against detrimental effects of gene disruption. Similarly, the *S. pombe* Tf1 IN interacts with leucine zipper transcription factors causing Tf1 to target the promoter regions of Pol II-transcribed genes. Other elements target heterochromatic regions via interactions with chromatin proteins. Ty5 IN interacts with the Sir4-silencing protein found at silent mating-type loci and telomeres. The fungal MAGGY element IN contains a chromo-domain within the carboxyl-terminal domain of IN that interacts with methylated H3K9, biasing MAGGY integration toward heterochromatic regions. Overall, these patterns are consistent with minimizing disruption of host cell functions and conducive to retrotransposon proliferation.

TP Retrotransposon Targeting

Diverse mechanisms underlie the insertion patterns of TP retrotransposons. At one extreme is the R2-type non-LTR element which encodes an EN domain related to type II restriction ENs that creates the priming site by cleaving adjacent to its binding site within the rDNA repeats of host organisms. Lesser specificity is illustrated by the LINE type of non-LTR elements which derive EN function from an AP. These elements display bias for AT-rich sites favoring the consensus 5'TTTT/AA3' with cleavage at T/A. This pattern allows pairing of the template poly(A) tail with the priming site. Group II intron RNAs integrate by specific reverse splicing at homing sites. However, in this case, the reaction is mediated by an RNA structure stabilized by the IEP maturase rather than by an enzyme as in the case of the site-specific R2-type elements.

Other classes of TP elements occur in forms that can transpose in the absence of EN activity. The insertion patterns of these elements are influenced by the availability of DNA structures conducive to reverse transcription priming. This mechanism is dramatically illustrated by the autonomous TART element of *Drosophila*. Because DNA polymerase requires a free 3'-OH primer end, there must be a special mechanism for replication of the 5'-ends of linear chromosomal DNA. In most eukaryotic species, this function is performed by telomerase. Telomerase is an RT related to non-LTR retrotransposon RT. It extends the 3'-ends of chromosomes by copying a short guide RNA with which it is complexed. However, telomerase is lacking in *Drosophila*. Rather, retrotransposons such as TART exploit the 3'-OH of chromosomal DNA as the primer, and in so doing extend chromosomes and provide the telomeric replication buffer. TART elements occur associated with defective Het-A elements and are thought to complement Het-A mobility *in trans*. Some isolates of PLE lack EN and are similarly found in telomeric regions. TP elements lacking EN can also exploit available lagging strand 3'-ends at replication forks. The best-understood example of this is group II intron retrotransposition.

Overall, both EP and TP retroelements display some degree of target specificity. This is likely to reflect, particularly in lower eukaryotes with a greater density of coding sequence, the benefits to a genomic parasite of minimizing disruption of host functions.

Regulation of Retrotransposons

Because retrotransposon amplification is potentially harmful, host cells have developed extensive mechanisms for controlling retrotransposition. These operate at the levels of transcription, mRNA stability, translation, and protein stability and modification. Only EP LTR retrotransposons have been identified in the nuclear DNA of the single-cell model organism *S. cerevisiae*. The EP LTR retrotransposon Ty1 and Ty3 illustrate some of these fundamentally different types of regulation. For example, Ty3 is expressed only during the brief window of the yeast mating cycle. Transcription is low in haploid cells and repressed in diploid cells, but is induced more than 50-fold in mating haploid cells. Transposition occurs after diploid cells begin to replicate. Ty1 transcripts are the most abundant mRNAs in *S. cerevisiae*, but production of active VLPs is downregulated by anti-sense RNA and proteolytic degradation. DNA maintenance factors also control retrotransposition of Ty elements. This reflects cDNA degradation by some DNA damage surveillance factors as well as stimulation of Ty1 transposition by DNA repair checkpoints. The heterochromatin-targeting Ty5 illustrates yet another level of regulation. In this case, stress relaxes the stringency of its targeting, allowing it to proliferate to novel insertion sites. Thus, these elements in a primitive eukaryote transpose during brief developmental windows and during times of stress, potentially favoring evolutionary persistence of both host and element.

A mechanism of retrotransposon regulation which is extremely widespread, but for which functions do not occur in *S. cerevisiae*, is RNAi. Retrotransposons have a propensity to generate double-stranded RNA. Some of these can be attributed to sense and antisense promoters within the retrotransposon

itself. However, as discussed above, in many organisms preferential insertion of retrotransposons results in chromosomal clusters of elements. Because insertions can occur in both orientations, readthrough transcription even at a low level can generate anti-sense RNAs available to duplex with sense RNAs. This is particularly true of LTR retrotransposons which have promoters in the upstream and downstream LTR sequences. These duplex retrotransposon-related RNAs provide substrates for the Dicer complex, which cleaves duplex RNAs into RNAs 21–30 nt in length. Generation of these RNAs is potentially amplified by cellular RNA-dependent RNA polymerases. One strand of the processed duplex is taken into the RNA-induced silencing complex (RISC) and used as a guide RNA by the Argonaute (Ago) subunit to target complementary RNA sequences for post-transcriptional degradation or through less well-understood translational repression mechanisms. Furthermore, specialized Argonaute proteins, including the Piwi Argonaute proteins, are active in controlling retrotransposition in germ cells where it would have profound consequences in the form of heritable insertions.

Animal cell methylation of CpG in DNA and chromatin modification, particularly methylation of lysine 9 of histone 3 (H3K9), are important mechanisms for repression of transcription. Because these modifications of DNA and chromatin can feedback to target maintenance of modifications, they are heritable without changes in coding DNA. Such regulation is referred to as epigenetic regulation. Epigenetic regulation is important for cellular differentiation during development. A particular type of epigenetic regulation is imprinting, in which maternal and paternal alleles are differentially methylated. EP LTR retrotransposons and TP LINE retrotransposons are repressed by DNA and chromatin modification. Furthermore, they may be differentially repressed in maternal and paternal germ cells. Mutations in components of the RNAi pathway relieve at least some of this repression, indicating that RNA components of this pathway can direct modification of both DNA and chromatin to retrotransposon sequences.

Retrotransposons may also have driven development of cell-based or intrinsic immunity functions. One example is the apolipoprotein B RNA-editing catalytic polyprotein (APOBEC) family of proteins, members of which have been shown to post-transcriptionally suppress LINE and SINE retrotransposition. These proteins deaminate cytosine in single-stranded cDNA and at least some of their effect is attributed to this modification of cDNA intermediates. Additional mechanisms of intrinsic immunity have been identified that have anti-retrovirus activity, and some of these may act on retrotransposons as well.

Significance

Effects of Retrotransposons on Genome Structure and Function

In spite of the numerous ways in which retrotransposition can be controlled, retrotransposons have clearly expanded their numbers on an evolutionary timescale, particularly in some complex animal lineages. This might suggest that there has been some advantage to the host organism in retaining retrotransposons and even retrotransposition. What might that be? Some retrotransposons have assumed roles as chromosomal structures. One compelling example is the retrotransposons

clustered in centromeric regions that allow for heterochromatinization of large amounts of DNA so that it can segregate during cell division. A second example is the telomeres in which chromosome extension by retrotransposon TPRT actually substitutes for telomerase in maintenance of chromosome ends. Because retrotransposons move relatively rapidly compared to other genes, they have potential to contribute to speciation through several mechanisms. First, they act to increase the potential of genomes for expansion of gene families and associated acquisition of novel functions. One mechanism of this expansion is reverse transcription of host mRNAs and integration of novel intronless pseudogenes into the genome. Second, the availability of a large pool of repeated sequences allows ectopic recombination between nonhomologous regions of the genome that can cause deletions, inversions, and translocations. Finally, and more directly, differential accumulation of retrotransposons between individuals of a single species can result in activation of transposition when individuals with different transposon content are crossed which in turn can generate sterile progeny, a phenomenon known as hybrid dysgenesis.

Effects of Retrotransposons on Host Gene Expression

Stability of the host–parasite relationships between organisms and their retrotransposons depends upon the ability to regulate retrotransposon expression and activity. Many elements are responsive to signals allowing expression only in specific contexts – mating, stress, or specific developmental windows are a few examples. However, mechanisms for significantly downregulating either most elements or all elements most of the time are present in host cells and it is interesting to speculate that some of these mechanisms have arisen specifically in response to retrotransposon activity. Among the most interesting of these is the ability of cells to sense the presence of repeated sequences and suppress their expression through RNAi. Not surprisingly, expression of retrotransposons can also affect host genes. This can be observed on two levels. First, the physical presence of retrotransposons can have direct effects on expression of associated genes. In the case of LTR retrotransposons, the promoter or terminator sequences in the LTRs can sponsor fortuitous promotion or termination of neighboring gene transcription. Second, however, elements which themselves are targets of inactivation can confer that regulation on neighboring genes causing them to be variably subject to inactivation, a phenomenon referred to as position effect variegation. Additionally, inactivated LTRs can act as insulators, blocking flanking enhancer activity. In a less well-understood example, LINE elements are proposed to have a role in perpetuating spreading of RNA-mediated inactivation of the X chromosome in mammals to confer dosage compensation. Some retrotransposons are differentially expressed in sperm and oocytes and are speculated to be involved in imprinting of associated regions in which genes derived from parental and maternal sources are differentially expressed.

Insertion of elements directly into coding regions is clearly potentially disastrous for gene function. However, because LINEs are relatively poorly transcribed and SINE elements can contain inverted sequences, even insertion into intronic regions can have profound consequences for expression of the

interrupted host gene. Transcription elongation, polyadenylation, alternative splicing, and editing of resulting chimeric transcripts can be affected. However, even this seemingly drastic disruption could have been exploited in higher eukaryotes. Recent experiments have shown that LINE retrotransposition is activated during specific windows of neurogenesis and it is implicated in neuronal plasticity coincident with that period of development.

Repeated sequences including retrotransposons can cause genomic variation via ectopic recombination or retrotransposition. The current activity of both LINEs and SINEs in germline cells means not only that they affect gene expression over evolutionary time frames but that spontaneous insertions can cause heritable disease. There are more than 50 diseases now attributable to retrotransposed elements. The availability of high-throughput sequencing will make it possible to better understand the involvement of these elements in somatic diseases, including cancer, which can result from genetic instability.

Gene Functions Related to Retrotransposons and Retroviruses

An obvious mechanism for retrotransposon modification of host genome-coding potential is reverse transcription and integration of genes to create new copies or pseudogenes that lack introns. These superfluous copies provide genetic raw material for evolution of new gene functions.

The co-evolution of DNA and RNA transposons and retroviruses with their hosts has inevitably resulted in the exaptation of genes from retrotransposon parasites and retroviruses by their hosts. Several gene families have arisen from conservation of the Gag gene. Members of this class in which the nucleocapsid zinc finger is retained are speculated to be transcription factors. The ability of DNA and RNA transposons to recombine DNA has also been co-opted by the cell. Examples include the CENP-B centromeric protein, which interacts with histone deacetylase, and the RAG1 recombinase, which is required to generate immunoglobulin diversity during V(D)J recombination.

Laboratory and Clinical Applications of Retrotransposons

Retrotransposons are sophisticated tools for retrospective and prospective *in vivo* studies. Because of their ability to integrate into and modify genomes, retrotransposons and retroviruses have made unique contributions to phylogenetic analysis and genetic engineering. Retrotransposons move frequently relative to the time frame of speciation; furthermore, the error-prone nature of RT causes a relatively high frequency of changes and can add further resolution to probes of phylogenetic relationships. In more directed approaches, activation of transposition is used to map gene functions. Mutant libraries can be screened for desirable phenotypes and mutant genes readily identified through the retrotransposon tag. If appropriately high levels of retrotransposition can be achieved in organisms with simple genomes, pools of genomic DNA can be sequenced to identify regions that lack insertions and therefore must be essential under the experimental condition. Finally, in cases in which the gene of interest is known, retrotransposons and

retroviruses have proved to provide extremely powerful vectors for introduction of specific test genes into host organisms.

Perspectives

The catalog of retrotransposon sequences is expanding rapidly, accelerated by the advent of high-throughput sequencing of the genomes of diverse species. This has led to new appreciation of the variety of these fascinating elements and the potential impact they have on genomic functions. The very recent discovery of unanticipated mechanisms of gene regulation, such as RNAi, that relate to host control of retrotransposons suggests that many exciting discoveries remain to be made.

See also: [Molecular Biology: DNA Polymerases: Reverse Transcriptase Integrase, and Retrovirus Replication](#); [Nonhomologous Recombination: Bacterial Transposons](#).

Further Reading

- Beauregard A, Curcio MJ, and Belfort M (2008) The take and give between retrotransposable elements and their hosts. *Annual Review of Genetics* 42: 587–617.
- Cordaux R and Batzer MA (2009) The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics* 10: 691–703.
- Curcio MJ and Derbyshire KM (2003) The outs and ins of transposition: From Mu to Kangaroo. *Nature Reviews Molecular Cell Biology* 4: 865–877.
- Dewannieux M, Esnault C, and Heidmann T (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nature Genetics* 35: 41–48.
- Eickbush TH and Jamburuthugoda VK (2008) The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Research* 134: 221–234.
- Kazazian HH Jr. (2004) Mobile elements: Drivers of genome evolution. *Science* 303: 1626–1632.
- Lambowitz AM and Zimmerly S (2004) Mobile group II introns. *Annual Review of Genetics* 38: 1–35.
- Levin HL (2002) Newly identified retrotransposons of the Ty3/gypsy class in fungi, plants, and vertebrates. In: Craig NL, Craigie R, Gellert M, and Lambowitz AM (eds.) *Mobile DNA II*, pp. 684–705. Washington, D.C.: American Society for Microbiology.
- Malik HS and Eickbush TH (1999) Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. *Journal of Virology* 73: 5186–5190.
- Sandmeyer SB, Aye M, and Menees TM (2002) Ty3: A position-specific gypsylike element in *Saccharomyces cerevisiae*. In: Craig NL, Craigie R, Gellert M, and Lambowitz AM (eds.) *Mobile DNA II*, pp. 663–682. Washington, D.C.: American Society for Microbiology.
- Slotkin RK and Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* 8: 272–285.
- Symer DE and Boeke JD (2010) The everlasting war dance between retrotransposons and their metazoan hosts. In: Kurth R and Bannert N (eds.) *Retroviruses: Molecular Biology, Genomics and Pathogenesis*, pp. 1–33. Norfolk, UK: Caister Academic Press.
- Voytas DF and Boeke JD (2002) Ty1 and Ty5 of *Saccharomyces cerevisiae*. In: Craig NL, Craigie R, Gellert M, and Lambowitz AM (eds.) *Mobile DNA II*, pp. 631–662. Washington, D.C.: American Society for Microbiology.
- Winckler T, Szafranski K, and Glockner G (2005) Transfer RNA gene-targeted integration: An adaptation of retrotransposable elements to survive in the compact *Dictyostelium discoideum* genome. *Cytogenetic and Genome Research* 110: 288–298.

Nuclear Calcium in Synapse-to-Nucleus Communication: Common Regulator of Persistent Adaptations in the Nervous System

H Bading, University of Heidelberg, Heidelberg, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

Calcium Principal second messenger in the control of gene expression by electrical activity in neurons.

CRE-binding protein (CREB) CRE-interacting transcription factor that is a target of cAMP and calcium signaling pathway. Mediates gene induction upon synaptic activity and nuclear calcium transients.

CREB-binding protein (CBP) Transcriptional co-activator regulated by cAMP and calcium signaling pathway; interacts with CREB and many transcription factors thereby conferring calcium and cAMP inducibility.

Cyclic adenosine monophosphate (cAMP) response element (CRE) DNA regulatory element present in the promoter of many genes.

Mitogen-activated protein (MAP) kinases/extracellular signal-regulated kinases (ERK1/2) Important intracellular signaling pathway; activated by electrical activity-induced submembranous calcium signals.

Serum response element (SRE) DNA regulatory element regulated by signaling pathways that are activated by growth factors, serum, or calcium; nuclear target of MAP kinases (ERK1/2) signaling pathway.

Spatial calcium signaling Electrical activity can induce calcium signals in different intracellular compartments; submembranous, cytoplasmic, and nuclear calcium activate gene expression through different mechanisms and have distinct genomic targets.

Neurons initiate many changes in gene expression in response to a synaptic or electrical stimulus. Calcium is the principal second messenger bridging the gap between the synapse and the nucleus. To couple synaptic events to specific genomic responses, neurons exploit the spatial diversity of calcium signals associated with the stimulus. Dendritic calcium signals, activating the mitogen-activated protein (MAP) kinase (extracellular signal-regulated kinase, ERK1/2) signaling cascade, stimulate gene expression mediated by the serum response element (SRE). Calcium signals in the cell nucleus activate the transcription factor cyclic adenosine monophosphate (cAMP) response element (CRE)-binding protein (CREB) and the co-activator CREB-binding protein (CBP). Nuclear calcium may, in particular, initiate and govern gene expression-dependent forms of neuronal adaptations, including acquired neuroprotection and the formation of long-term memory.

Calcium Ions, Key Regulators of Electrical Activity-Dependent Gene Expression

The generation of calcium signals following electrical activation is a fundamental property of neurons that controls many processes in the developing and the adult nervous system. Activity-induced increases in the intracellular calcium concentration result from calcium entering neurons from the extracellular space through ligand and/or voltage-gated ion channels; these calcium transients can be amplified through calcium release from intracellular calcium stores. The *N*-methyl-D-aspartate (NMDA) receptor, a calcium permeable, glutamate-gated ion channel, plays a particularly important role in the mammalian central nervous system. Calcium entry through NMDA receptors, triggered by electrical activity, stimulates mechanisms that affect synaptic connectivity and neuronal network behavior leading to

information storage; NMDA receptor-mediated calcium entry can also promote neuronal survival or cause cell death. Important for the consolidation of these activity-induced responses is that calcium signals, generated initially at the plasma membrane, are transduced to the cell nucleus where they stimulate the expression of particular genes. Thus, neurons use calcium ions to couple electrical signals to specific genomic events. Among the target genes of calcium signaling are neurotrophic factors, cytoskeletal proteins, and transcription factors (see below).

Spatial Calcium Signaling in Synapse-to-Nucleus Communication

The use of calcium as a universal signal mediator raises the issue of specificity: what mechanism allows neurons to use a single second messenger to convert a diverse range of electrical stimuli into distinct gene expression responses? Neurons appear to achieve this by using a calcium code: a given electrical stimulus is transformed into a calcium signal with particular spatial and temporal properties. Thus, an activity-induced neuronal calcium signal is a signature of the stimulus; it functions as its intracellular representation and links the stimulus to a gene expression response. The amplitude and the duration of the calcium signal, two features that transcriptional regulators decode, primarily control the magnitude of the increase in transcription. The site of calcium entry also matters; indeed, calcium entry through synaptic NMDA receptors and through extrasynaptic NMDA receptors has opposing effects on the activation of a crucial transcription factor (see below). However, most importantly, the genomic response depends on which part of the neuron is invaded by the calcium transient (Figure 1). For example, a weak synaptic input may generate a calcium signal that is confined to the submembranous space in

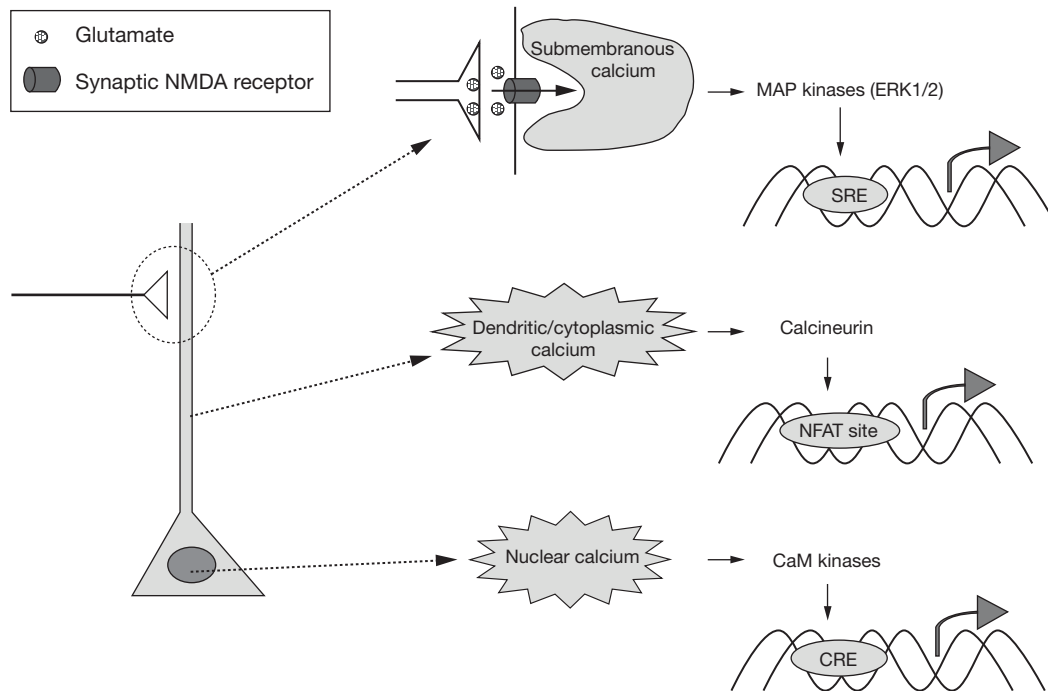


Figure 1 Schematic drawing of a neuron and the intracellular signaling pathways that mediate the induction of gene expression following neuronal activity. Three spatially distinct calcium signals (submembranous, cytoplasmic, and nuclear) are illustrated. Each type of calcium signal couples to a specific effector mechanism with distinct genomic targets. The decoding, by transcriptional regulators, of spatial information contained in activity-induced calcium transients allows neuronal impulse patterns to specify gene expression responses.

the immediate vicinity of the site of calcium entry; following a stronger stimulus, calcium increases may reach further into the dendritic tree. Under certain conditions, the calcium signal can invade the cell body and the nucleus. It is important for the specification of the gene expression response exactly which subcellular compartment (i.e., submembranous space, dendritic tree and somatic cytoplasm, and cell nucleus) undergoes a change in the calcium concentration. Thus, the ‘dialog between genes and synapses’ (to use a phrase from Eric Kandel) exploits the spatial and temporal diversity of calcium transients associated with electrical activation. Transcriptional regulation and the neuronal responses depend on how calcium enters the neurons, the amplitude of the signal, how long the signal lasts, and what subcellular compartment it invades.

Regulators of Calcium-Induced Gene Expression

Much knowledge of the mechanisms of electrical activity-dependent gene expression comes from studies of the *c-fos* gene. This gene is induced, without the need for new protein synthesis, within minutes after the stimulation. The *c-fos* gene encodes a DNA-binding protein; together with members of the Jun transcription factor family it forms the AP-1 complex that is involved in the regulation of many genes. *c-fos* was the first gene found to be induced by membrane depolarization and calcium influx into excitable cell lines and primary neurons.

SRE and CRE

Analysis of the *c-fos* promoter revealed that calcium influx-induced gene transcription is mediated by two distinct DNA

regulatory elements: the CRE and the SRE. The CRE and the SRE were previously known to mediate gene induction upon increasing intracellular levels of cAMP or following the addition of serum to serum-starved fibroblasts, respectively. Analysis of calcium regulation of SRE-dependent and CRE-dependent transcription uncovered the importance of the intracellular localization of the calcium signal: an increase in cytoplasmic calcium activates gene expression mediated by the SRE; by contrast, activation of the CRE and the CREB require increases in the nuclear calcium concentration (see below). Thus, a single second messenger can, depending on its intracellular localization, induce distinct responses.

NFAT and DREAM

Calcium-responsive regulators of transcription, other than the SRE and CREB, include the nuclear factor of activated T-cells (NFAT) and the downstream regulatory element (DRE)-binding protein, DREAM (DRE-antagonist modulator). NFAT is important in the immune system; it mediates in activated T-cells the induction of interleukin 2 expression, a critical step in the response to antigen stimulation. DREAM is a multifunctional calcium-binding protein that can act as a repressor of gene transcription. The importance of spatial calcium signaling in the regulation of the SRE, the CRE, NFAT, and DREAM is outlined in the following section.

Calcium Signals in Different Subcellular Compartments

Calcium brings about biological responses through activating effector molecules such as protein kinases or protein

phosphatases. Exactly which calcium-regulated molecule is stimulated depends on the subcellular localization of the calcium transient.

Submembranous Calcium Signals

A calcium microdomain restricted to the submembranous space near synaptic NMDA receptors serves as the on-switch of the MAP kinase (ERK1/2) cascade following synaptic activity. This signaling pathway is a major route for intracellular communication. The actions of MAP kinases (ERK1/2), both in the vicinity of calcium entry sites and in the nucleus, are involved in controlling many different brain functions ranging from learning to neuronal survival and light-induced shifts in the circadian rhythm.

One important target of the MAP kinase (ERK1/2) cascade is the nucleus. The MAP kinase (ERK1/2) signal propagates to the nucleus independently of global (i.e., cell-wide) increases in the calcium concentration and stimulates gene expression mediated by the SRE. It can also lead to the phosphorylation of CREB on its activator site, serine 133. Although this phosphorylation is necessary for inducing CREB-mediated transcription it is not sufficient; a second regulatory event is required that involves stimulation of the activity of the CREB coactivator CBP by nuclear calcium and nuclear calcium/calmodulin (CaM)-dependent protein kinases (see below). Nevertheless, the MAP kinase (ERK1/2) cascade acts as an auxiliary pathway that promotes at least one of several activating steps that control CREB-dependent gene expression. CREB target genes include *c-fos* and the brain-derived neurotrophic factor (BDNF) (see below).

Dendritic/Cytoplasmic Calcium Signals

The second calcium pool with a role in transcription regulation by neuronal activity is localized to the dendrites and the cytoplasm of the cell body. One of the targets of dendritic/cytoplasmic calcium is calcineurin (CaN). CaN is a cytoplasmic serine/threonine phosphatase that, upon activation by calcium/CaM, catalyzes the dephosphorylation of NFAT which unmasks an NFAT nuclear localization signal. The activated CaN/NFAT complex is subsequently imported into the nucleus to activate gene transcription. Possible target genes of the CaN/NFAT-signaling module are genes involved in regulating calcium homeostasis; this includes the IP₃ receptor and plasma membrane calcium adenosine triphosphatases (ATPases) (calcium pumps).

Once calcium concentrations have returned to basal levels, the NFAT/CaN complex dissociates, NFAT is exported from the nucleus and transcription is rapidly shut off. Thus, the magnitude of the transcriptional response is a function of the duration of the calcium signal; only prolonged increases in intracellular calcium efficiently activate NFAT-mediated gene expression. In addition, increases in nuclear calcium are required to maintain NFAT in the nucleus and to sustain transcription. Such calcium signals (i.e., elevated calcium plateaus in the cytoplasm and the nucleus) are typically seen in antigen-stimulated T-cells that require NFAT activation for the immune response (see above). In neurons, synaptic activity often generates only transient increases in intracellular calcium that may induce a less robust NFAT activation.

Nuclear Calcium Signals

The third player in electrical activity-dependent gene expression is nuclear calcium. Synaptically evoked calcium transients can be amplified by calcium release from intracellular stores and are propagated, possibly in the form of a regenerative wave, to the nucleus. Nuclear calcium is a central regulator of neuronal gene expression. It activates nuclear calcium/CaM-dependent protein kinases and is the principal regulator of gene expression by the CREB/CBP complex. Activation of gene expression by CREB/CBP requires at least two regulatory events: phosphorylation of CREB on its activator site serine 133, which allows CBP to be recruited to CREB. The second event is the stimulation of CBP activity. The first regulatory step can be catalyzed by several protein kinases, including kinases of the MAP kinase (ERK1/2) signaling pathway (triggered by submembranous calcium; see above) and nuclear calcium-induced CaM kinases. Induction of CBP activity, however, depends upon increases in the nuclear calcium concentration and the activation of nuclear CaM kinases.

The only other second messenger known to stimulate CBP activity and to increase CREB/CBP-mediated gene expression is cAMP. Global increases in the intracellular concentration of cAMP stimulate the cAMP-dependent protein kinase (PKA); upon binding of cAMP to the PKA regulatory subunit, the catalytic subunit of PKA translocates to the cell nucleus to activate the CREB/CBP complex. A large body of literature suggests that cAMP and PKA are important for certain long-term, gene expression-dependent forms of plasticity and learning in *Aplysia californica* and *Drosophila melanogaster*. Mammalian neurons are also capable of increasing intracellular levels of cAMP following synaptic activity through a process that involves calcium/CaM-dependent adenylyl cyclases. However, electrical activity does not appear to elevate global levels of cAMP sufficiently high to stimulate the CREB/CBP complex, although it may increase cAMP and PKA activity locally leading to the phosphorylation of neurotransmitter receptors or other proteins in the postsynaptic space. Thus, the cAMP-PKA system may be important for localized dendritic signaling in the mammalian nervous system, but the stimulation of CREB/CBP-dependent gene expression following synaptic activity appears to be controlled by nuclear calcium signals.

CBP is a coactivator of CREB but it also interacts with several other DNA-binding proteins thereby conferring calcium inducibility. One example is the transcription factor c-Jun that can function as a calcium-regulated activator. Thus, control of CBP activity by nuclear calcium and nuclear CaM kinases provides a general mechanism through which electrical activity regulates an entire class of transcription factors and, consequently, many genes. The generation of nuclear calcium transients activating CBP-containing complexes is likely to be important for activity- and gene transcription-dependent forms of neuronal adaptation including acquired neuroprotection, information storage, and memory formation. A transcriptome study revealed that in mouse hippocampal neurons, nuclear calcium is one of the most potent signals in neuronal gene expression. The induction or repression of nearly 200 activity-regulated genes is dependent upon nuclear calcium signaling. The nuclear calcium-regulated gene pool contains several genes that can provide neurons with a robust

neuroprotective shield and renders them more resistant to cellular stress and toxic insults. In line with the concept that most (perhaps all) activity-regulated, persistent adaptations require that stimulus-associated calcium signals invade the cell nucleus is the observation that nuclear calcium signaling is critical for the formation of long-term memory in *D. melanogaster* and in mice.

A second type of transcriptional regulation by nuclear calcium involves DREAM. Rather than being controlled by calcium/CaM-dependent enzymes, the ability of DREAM to modulate transcription is directly controlled by nuclear calcium. DREAM is a calcium-binding protein that, in the absence of calcium, interacts with the DRE and represses transcription. Increases in nuclear calcium lead to the formation of a calcium/DREAM complex; calcium-bound DREAM dissociates from the DNA allowing transcription to take place. Potential DREAM-binding sites are present in many different genes; the involvement of DREAM is best documented for the prodynorphin gene.

CREB Shut-Off Pathway

The genomic responses (and their effects on survival and plasticity programs) induced by calcium flux through synaptic NMDA receptors are antagonized by a signaling pathway stimulated by the addition of glutamate to extracellular medium. Bath glutamate application stimulates both synaptic and extrasynaptic NMDA receptors as well as other types of glutamate receptors. Pharmacological experiments revealed that calcium flux through extrasynaptic NMDA receptors is responsible for this nuclear signaling mechanism that negatively regulates gene expression; this pathway is also coupled to neuronal cell death. The target transcription factor inactivated following calcium entry through extrasynaptic NMDA receptors entry is CREB; the CREB shut-off is brought about by a rapid dephosphorylation of CREB on its activator site, serine 133 (Figure 2).

Extrasynaptic NMDA Receptors Antagonize Nuclear Calcium Signaling and BDNF Gene Expression

After calcium entry through extrasynaptic NMDA receptors, one of the consequences of the CREB shut-off is the suppression of BDNF gene expression. BDNF, a neurotrophic factor, activates intracellular signaling pathways through binding to TrkB, a receptor tyrosine kinase; BDNF also acts as a fast neurotransmitter and can rapidly depolarize neurons. BDNF has many functions in neuronal plasticity and cell survival. BDNF gene transcription is electrical activity dependent and controlled, antagonistically, by two calcium-signaling pathways: calcium flux through synaptic NMDA receptors (or through L-type voltage-gated calcium channels) stimulates BDNF gene transcription; by contrast, calcium flux through extrasynaptic NMDA receptors antagonizes increases in BDNF expression. The BDNF gene's promoter is complex and provides binding sites for many transcription factors; however, a key factor is CREB, which may explain the opposing action of synaptic and

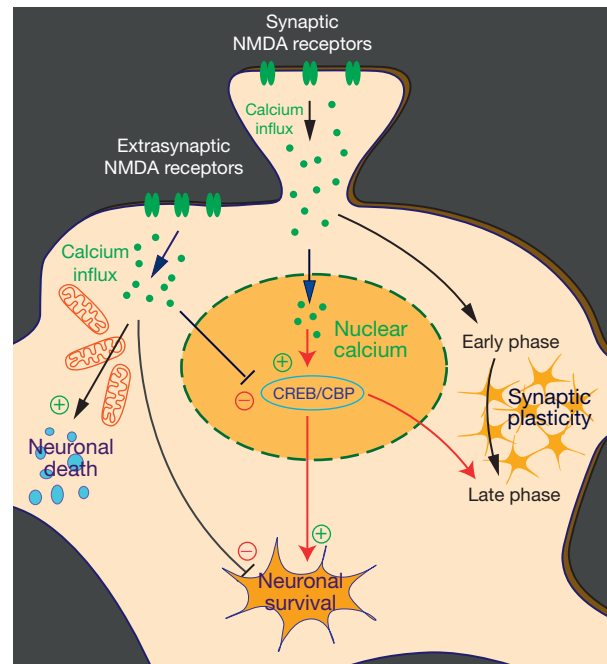


Figure 2 Schematic illustration of the role of synaptic and extrasynaptic NMDA receptors in neuronal gene expression, plasticity, survival, and cell death.

extrasynaptic NMDA receptors on influencing BDNF gene transcription. Activation of synaptic NMDA receptors (or L-type voltage-gated calcium channels) strongly promotes CREB activity, whereas CREB function is shut-off following activation of extrasynaptic NMDA receptors; those receptors trigger a CREB shut-off signal and antagonize increases in BDNF transcription induced by calcium entry through synaptic NMDA receptors. This underscores the importance of the intracellular localization of the calcium signal for the specification of the neuronal response.

See also: [Bioenergetics: Calcium-Modulated Proteins \(EF-Hand\); Calcium-Sensing Receptor \(CaSR\).](#)

Further Reading

- Bading H (2000) Transcription-dependent neuronal plasticity: The nuclear calcium hypothesis. *European Journal of Biochemistry* 267: 5280–5283.
- Hardingham GE and Bading H (2010) Synaptic versus extrasynaptic NMDA receptor signalling: Implications for neurodegenerative disorders. *Nature Reviews Neuroscience* 11: 682–696.
- Kandel ER (2001) The molecular biology of memory storage: A dialogue between genes and synapses. *Science* 294: 1030–1038.
- West AE, Griffith EC, and Greenberg ME (2002) Regulation of transcription factors by neuronal activity. *Nature Reviews Neuroscience* 3: 921–931.
- Zhang SJ, Zou M, Lu L, et al. (2009) Nuclear calcium signaling controls expression of a large gene pool: Identification of a gene program for acquired neuroprotection induced by synaptic activity. *PLoS Genetics* 5: e1000604.

Nuclear Compartmentalization

K P Smith and J B Lawrence, University of Massachusetts Medical School, Worcester, MA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Cajal bodies Spherical nuclear bodies (1–5/nucleus) enriched in messenger RNA (mRNA), ribosomal RNA (rRNA), and small nuclear RNA (snRNA) processing factors.

Nuclear compartment A nonmembrane-bound subnuclear structure in which a variety of factors accumulate to facilitate specific nuclear processes.

Nucleus A membrane-bound cellular structure containing DNA which controls the cell's growth, metabolism, and replication.

PML nuclear bodies Spherical nuclear structures (10–30/nucleus) rich in a variety of factors, many of which are involved in transcription regulation.

SC35 domains Twenty to forty irregular speckles in the nucleus in which numerous factors required for mRNA metabolism accumulate.

Survival of motor neuron (SMN) gems Nuclear bodies enriched in SMN protein and other factors involved in small nuclear ribonucleoproteins (snRNP) assembly. Gems often associate or coincide with Cajal bodies.

Introduction

The orchestrated expression of the genome's ~35 000 genes is largely regulated by epigenetic changes to chromatin and nuclear structure, involving the formation of more dense largely inactive heterochromatin versus more open and more expressed euchromatin. In somatic cells, the genome is packaged at a higher level into large, cytologically visible domains of euchromatin and repressive heterochromatin, which tend to occupy different regions of the nucleus. Euchromatin occupies the more internal nucleoplasm, and this large euchromatic compartment is further partitioned into several nonchromatin subcompartments enriched in different RNA metabolic and regulatory factors. By contrast, heterochromatin forms a compartment of compact DNA at the nuclear periphery, where many inactive sequences have been localized in mammalian cells.

Although it is beyond the scope of this article, it should be noted that our lab and several others have studied the organization of individual genes with distinct nuclear domains, using fluorescence *in situ* hybridization (FISH) methods to visualize single genes and their nuclear RNAs. In sum, evidence indicates that nuclear localization of specific genes with distinct nuclear compartments contributes to their regulation and/or efficient expression of RNAs, but this alone does not determine the expression status of a gene. For example, some active genes are at the nuclear periphery near heterochromatin, whereas some inactive genes will be in the more euchromatic regions of the nucleoplasm. Nonetheless, an important theme in nuclear structure is that the organization of specific genomic loci is clearly nonrandom with respect to many of the nuclear compartments described below. This article focuses on summarizing the main features of these various nuclear substructures themselves, beginning with the nuclear periphery and heterochromatic compartment and then the various domains and bodies throughout the large internal euchromatic compartment.

The Nuclear Periphery

The Nuclear Envelope and Lamina

The cell nucleus is bounded by a bilayered nuclear envelope (NE) that separates the nucleus from the cytoplasm. This

structure is comprised of the inner and outer membranes and the nuclear pores, which allow communication between the nucleus and the rest of the cell. Internal to the NE is a filamentous layer, the lamina. Several lamin proteins form the lamina which encircles the nuclear periphery at the interface between chromatin and the inner nuclear membrane (Figure 1), but lamins are also found more dispersed throughout the nucleoplasm. Lamins are required for transcription and DNA replication and are encoded by three separate genes, lamin B1, B2, and A/C, which produce several alternatively spliced isoforms. The lamina contains numerous factors that interact with the lamins, including emerin, lamin B receptor (LBR), and lamina associated peptides (LAPs) 1 and 2, all of which also associate with chromatin. Via these chromatin interactions, the NE and lamina serve as an anchoring structure for nuclear architecture, positioning chromosome territories in interphase.

Peripheral Heterochromatic Compartment

The prevalence of heterochromatin formation in genome regulation is manifest cytologically, generating cell-type specific architectural arrangements of a prominent nuclear heterochromatic compartment. Discrete regions of densely packaged heterochromatin are typically compressed at the nuclear and nucleolar periphery (Figure 2), and also are present in smaller nucleoplasmic blocks, in cell-type specific patterns. Constitutive heterochromatin is comprised of DNA that is rich in repetitive sequences, such as centromeric satellites, whereas facultative heterochromatin consists of sequences that are euchromatic in some cell types. Heterochromatin is typically late replicating (Figure 2) and most often has certain biochemical features, such as deacetylation on histone H4 and methylation on histone H3, as well as increased DNA methylation. Although not all inactive genes localize in a heterochromatic compartment, inactive genes and certain chromosomes tend to reside at the periphery in certain cell types. Inactivation of one X chromosome in mammalian females is a pre-eminent example of heterochromatin formation, controlled by the large noncoding X (inactive)-specific transcript (XIST) RNA, which coats and induces silencing of the X chromosome, positioning

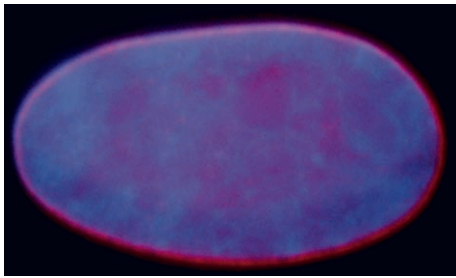


Figure 1 Lamin B1 staining (red) illuminates the nuclear lamina which encircles the nuclear DNA. The nucleus is stained blue using the DNA stain 4',6-diamidino-2-phenylindole (DAPI).

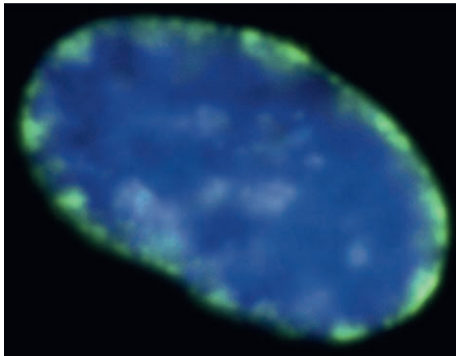


Figure 2 Incorporation of BrdU (green) into late replicating DNA delineates the peripheral heterochromatin that abuts the nuclear lamina.

it in the heterochromatic compartment at the nuclear or nucleolar periphery.

Chromosome Territories and Transcription Factories

With the development of FISH techniques, it became possible to visualize the distribution of DNA from an individual chromosome in interphase nuclei. This revealed that, despite the lower condensation of most interphase DNA, the interphase chromosome occupies a discretely bordered territory within nuclear structure (Figure 3). Thus, the DNA of one chromosome is not substantially intermingled with that of other chromosomes, although there may be some interconnections in the outer region of each territory. Although each chromosome occupies a distinct territory, there is no clearly visible space (devoid of chromatin or other ultra-structures) that defines the border of these territories; thus, they are only apparent by hybridization of a DNA library from a specific chromosome.

Several labs have investigated whether chromosome territories are organized in a specific three-dimensional (3D) position or in a particular order. Although in some specialized cells of certain species this appears to be the case, in mammalian cells there are preferred regions for specific chromosomes, but generally not a precise order of specific 3D coordinates. For example, larger chromosomes and less gene-dense chromosomes tend to be closer to the nuclear periphery, whereas smaller and more gene-rich chromosomes tend to be more central (including ribosomal (rDNA) bearing chromosomes that are with nucleoli). Another major question that has been investigated concerns the organization of genes within each chromosome territory. Several studies indicate that canonical

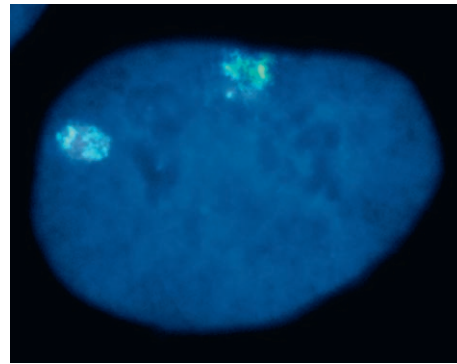


Figure 3 Fluorescence hybridization with a library of DNA from an individual chromosome defines an interphase chromosome territory for each of the two homologous chromosomes.

protein-coding genes have a propensity to reside in the outer $\sim 1/3$ of the chromosome territory. Whether the position of genes in the territory changes with their activity has been a subject of some debate, it appears that certain highly active gene clusters may loop out away from their territory upon activation, but this is not seen for most genes studied, as many show no clear difference in position within the territory relative to activity. Interestingly, on the inactive X chromosome, genes studied to date position at the very outer border of the chromosome territory, largely outside the dense heterochromatic core of the chromosome (the Barr Body). This suggests that this condensed, inactive chromosome makes more sharply defined an underlying organization that may be present on all chromosomes, with loose compartmentalization of coding genes peripheral to a core of noncoding more repeat-rich DNA.

Transcription factories are smaller structures within nuclei that are generally defined by the punctuate distribution of RNA polymerase II into approximately 1000–2000 very small discrete foci in most cells which is just within the limits of resolution of light microscopy. These transcription foci may be more prevalent near the peripheral zone of chromosome territories, near splicing factor-rich SC35 domains or other regions of high activity. While it is not entirely clear if these $0.1 \mu\text{m}$ foci represent a distinct structural compartment, a few studies have raised the possibility that coordinately regulated genes can cluster at these foci.

The Nucleolus

The nucleolus is the best-known subcompartment of euchromatin, as it was long ago easily visualized by phase microscopy and has long been known to be the site of highly active ribosomal RNA (rRNA) synthesis. In fact, while it was not widely appreciated initially, the nucleolus established the precedent for the compartmentalization of genes and RNA metabolic factors into an organized structure devoted to a certain function: a ribosomal subunit factory. The rDNA genes that form the nucleolus are on 10 different human chromosomes, and although the number and arrangement of nucleoli differ in cell-type specific patterns, in all cases genes on separate chromosomes must congregate within these structures for this common purpose. Ribosomal genes and RNA are transcribed

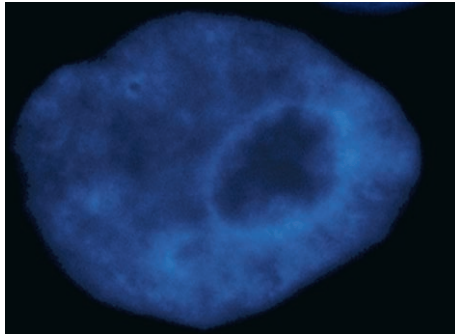


Figure 4 The nucleolus is typically visible as a dark region of low DNA density by DAPI DNA stain, due to the abundance of ribosomal RNA and proteins assembling throughout most of the nucleolus.

in small regions of the nucleolus known as the 'fibrillar center (FC)' and 'dense fibrillar component (DFC)'. The bulk of the nucleolus is the granular component (GC), where rRNA and proteins assemble to the ribosomal subunit (Figure 4). In addition to its well-established importance in the efficient production of rRNAs and ribosome assembly, evidence indicates that the nucleolus likely has expanded functions related to RNA metabolism or possibly even heterochromatin formation, and thus the nucleolus is increasingly considered a multi-functional nuclear suborganelle.

Splicing Factor Compartments, Speckles, or SC35 Domains

In the eukaryotic nucleus, the distribution of pre-messenger RNA (mRNA) splicing factors is not uniform, but is markedly concentrated at 15–30 sites, with lower levels of factors diffused throughout the nucleoplasm (Figure 5). Various referred to as 'speckles', 'SC35 domains', or 'splicing factor compartments (SFCs)', these irregular but discrete domains are most frequently visualized with an antibody directed against the spliceosome assembly factor SC35. Each domain, 0.5–3.0 μm in diameter, corresponds largely, if not entirely, to ultrastructures termed 'interchromatin granule clusters (IGCs)'. In addition to spliceosome assembly factor SC35, these domains contain numerous splicing factors and SR proteins, as well as poly(A) RNA, poly(A) RNA-binding protein, several factors implicated in RNA export and in some studies RNA polymerase II and lamins. Recently, it has been shown that nuclear paraspeckle assembly transcript 2 (NEAT2) RNA, one of the most abundant noncoding RNAs in the nucleus, also is markedly enriched in SC35 speckles. In addition, all of these components are more diffusely distributed throughout the nucleus. Although the positions of the domains themselves are relatively immobile, they are dynamic structures in that factors within them are in rapid flux dependent upon their phosphorylation state. It is likely that ~100 different proteins involved in pre-mRNA metabolism localize, at least transiently, to these domains.

Because the IGCs show less label with uridine incorporation, it had been thought that these regions would not contain mRNA or pre-mRNA but might be just storage sites unassociated with active genes. However, many (about half) of the active genes studied have been shown to position directly at

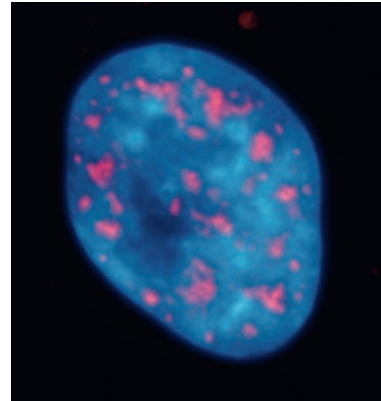


Figure 5 Splicing factor rich SC35 domains as visualized in a human fibroblast nucleus immunostained for the spliceosome assembly factor SC35 (red).

the periphery of SC-35 domains, and RNAs from highly expressed genes have been found within them. The organization of these genes with domains is locus-specific and cell-type-specific. Evidence now suggests that most, if not all, splicing factor domains associate with numerous genes and may contain mRNAs derived from them, but they may simultaneously serve as reservoirs or reassembly sites of RNA metabolic factors. The placement of specific genes near these concentrated domains of RNA metabolic factors suggests a specific nuclear organization of the gene relative to the domain which would facilitate the rapid reassembly and use of large RNA metabolic complexes for highly expressed and complex genes (Figure 6).

Paraspeckles

Paraspeckles were so named due to their close juxtaposition to the larger SC-35-rich speckles (Figure 7). Paraspeckles number between about 5 and 15 per cell. While they are present in mammalian somatic cells, similar to SC35 domains, they are not found in human embryonic stem cells. Paraspeckles, 0.5–1.0 μm in size, were initially identified due to their concentration of certain proteins implicated in RNA editing, including PSP1, PSF, and p54/NONO. While these proteins are also present at lower levels throughout the nucleoplasm, it was later shown that NEAT1 long-noncoding RNA is exclusively localized to paraspeckles, and it provides the structural underpinnings, setting a precedent for an architectural role for noncoding RNA in forming a cytological scale nuclear sub-compartment. While the function(s) of paraspeckles is being investigated, there is evidence that they are involved in adenosine-inosine editing of RNAs and in the nuclear retention of mRNAs for which the nuclear export is regulated in relation to cell signals.

Promyelocytic Leukemia Nuclear Bodies, ND10

Mammalian nuclei contain on the order of 10 spherical structures termed 'promyelocytic leukemia (PML) nuclear bodies' (PML NBs) (also termed PODs, ND10, or Kremer bodies) (Figure 8). These 0.2–1 μm bodies were first observed by electron microscopy (EM) and later found to contain the PML

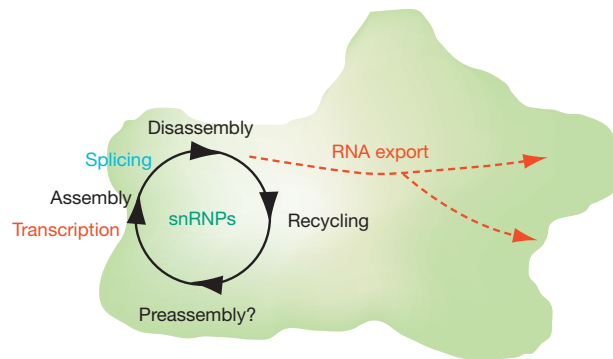


Figure 6 SC35 domain model. Domains may function by coupling the completion of mRNA maturation and release for mRNA export with the recycling/preassembly of factor complexes. Because the RNA metabolic machinery requires interaction of a large number of different factors, their concentration at a site would facilitate recycling and rapid reuse for expression by adjacent genes. Adapted from Johnson CV, Primorac D, McKinstry M, et al. (2000) Tracking COL1A1 RNA in osteogenesis imperfecta. Splice-defective transcripts initiate transport from the gene but are retained within the SC35 domain. *Journal of Cell Biology* 150: 417–431, and Hall LL, Smith KP, Byron M, and Lawrence JB (2006) Molecular anatomy of a speckle. *The Anatomical Record Part A* 288: 664–675.

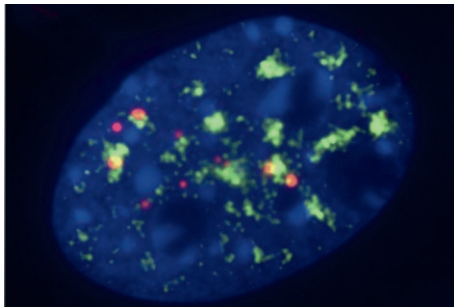


Figure 7 Paraspeckles are marked by their concentration of NEAT1 RNA (red) and are smaller structures that typically lie adjacent to SC35 domains, which are illuminated here by their enrichment for NEAT2 RNA (green).

protein. Interest in this nuclear body, and nuclear structure in general, significantly increased with the discovery that the gene-encoding PML is fused to the retinoic acid receptor gene in the t(15;17) translocation of acute promyelocytic leukemia (APL). The resulting PML-retinoic acid receptor fusion protein causes a breakdown of PML domains, a major hallmark of APL. PML bodies are also targeted by many DNA viruses in early infection, and are the site of viral transcription initiation. Subsequently, the viral early proteins localize to and eventually disrupt PML NBs. The number and size of PML NBs is also regulated by interferons and cellular stress, such as oxidative stress and DNA damage. In addition to the classical round PML NBs, recent studies have noted highly unusual PML structures in human embryonic stem cells.

The lynchpin of the PML NB is the PML protein, essential for PML body formation. The PML protein contains a RING finger, B-box coiled-coil (RBCC) motif that allows PML to interact with many proteins. The PML protein itself is involved in antiviral defense by impairing viral replication, inhibiting

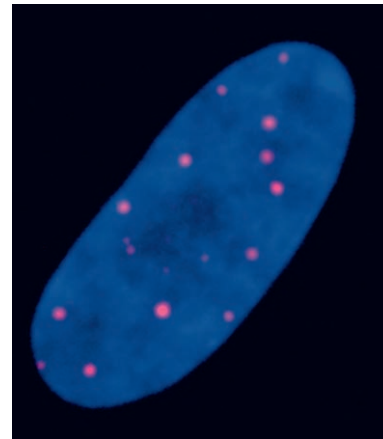


Figure 8 PML nuclear bodies. Human fibroblast nucleus stained with DAPI (blue) has numerous PML nuclear bodies visualized with an anti-PML antibody (red).

synthesis of viral mRNAs or by inducing apoptosis in infected cells. Although it does not appear that these antiviral activities occur in the PML NB, many viruses target disruption of the PML NB as a means of evading the cell's defense mechanisms.

The myriad PML NB proteins are involved in a number of different cellular processes, including tumor suppression, cell-cycle control, DNA repair, senescence, apoptosis, immune response, and transcription regulation. A few of the over 160 proteins which concentrate within PML NBs, such as Sp100 and death-associated dead protein (Daxx)1, are present under all conditions, but many others localize under certain conditions, in addition to lower levels in the nucleoplasm. Some proteins do not colocalize with all PML NBs or in all cell types, suggesting that different PML NBs may have distinct functions.

The sheer variety of proteins found in PML bodies has made attribution of a single function almost impossible. One theory is that PML NBs are nuclear depots from which proteins are recruited in response to stimuli such as viral attack or other cell stress. PML NBs could also be sites for degradation of some nuclear proteins, or act as a platform for posttranslational modifications such as phosphorylation, acetylation, and SUMOylation. In fact, over 40% of the factors found in PML NBs are SUMOylated, and small ubiquitin-like modifier (SUMO)-specific ligases and proteases have been detected in these bodies, suggesting that PML NBs may be SUMOylation hot spots. PML has been found to affect the activity of some stem cell related factors via posttranslational modification and, thus, may have a role in stem cell self-renewal.

Given their potential to sequester or modify transcription factors, it has been proposed that PML NBs affect chromatin architecture and transcription regulation.

Cajal (Coiled) Bodies, Histone Locus Bodies, and Gems

Cajal Bodies

First described in 1903 as nucleolar accessory bodies, the Cajal (coiled) bodies (CBs) are usually seen as roughly spherical 0.1–2.0 μm structures numbering one to five per nucleus (see [Figure 9\(a\)](#)), varying in number and size during the cell cycle and in different cell types. The Cajal bodies are not seen in all

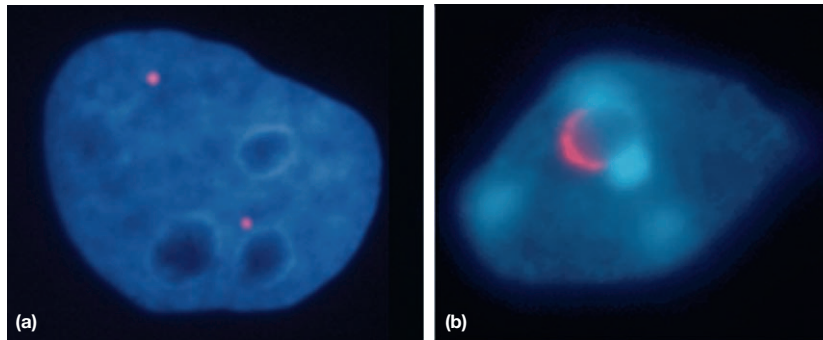


Figure 9 Cajal bodies, seen using an anti-coilin antibody (red) can appear as round nucleoplasmic bodies in highly proliferative cultured cells like the cervical carcinoma cell HeLa (a) or can be found to cap the nucleolus in a tissue section of mouse brain neurons (b). DNA in nuclei is stained with DAPI.

cells or tissues, but are especially prominent in highly proliferative cells such as cancer cells or metabolically active cells like neurons. The neuronal CBs, as originally identified, are not round nucleoplasmic bodies, but are found capping the nucleolus as seen in **Figure 9(b)**. This localization illustrates the close association between the nucleolus and the CBs. Whether the CBs capping the nucleolus are functionally distinct from the nucleoplasmic bodies most often seen in cultured cells is unknown.

The coilin protein was first identified using human auto-immune serum and has become the marker most often used to visualize CBs. Although relatively little is known about the function of coilin, studies have indicated that it is the link between CBs and the nucleolus. In addition to coilin, numerous small nuclear RNAs (snRNAs) and proteins of both nuclear and nucleolar origin are enriched in the CBs. These include factors involved in ribonucleoprotein maturation, RNA polymerase assembly, spliceosome assembly, and telomere elongation.

CBs have been shown to spatially associate with specific gene loci in the nucleus. They associate with several U snRNA gene loci (U1, U2, U4, U11, and U12) in a manner suggesting that CBs could have relationship to expression and nuclear metabolism of these snRNAs. The CBs also contain pre-U2RNA, implicating them in snRNA modification or export. In addition, Cajal bodies contain small Cajal body-specific RNAs (scaRNAs) which catalyze the methylation and pseudouridylation of snRNAs after the assembled small nuclear ribonucleoproteins (snRNPs) are transported into the nucleus. Cajal bodies also contain the core telomerase components and can transiently associate with telomeres in S phase of the cell cycle, suggesting a possible link to telomere elongation.

Survival of Motor Neuron Gems in Relation to CBs

Modification of snRNAs is likely not the sole function of CBs. The disease spinal muscular atrophy results from deletion of a gene encoding the survival of motor neuron (SMN) protein. This protein is consistently enriched in foci termed 'SMN gemini of Cajal bodies (gems)', since in the cell line in which they were discovered, gems were often very closely associated or even coincident with CBs (see **Figure 10(a)**). However, subsequently in many cultured cells studied, SMN and coilin

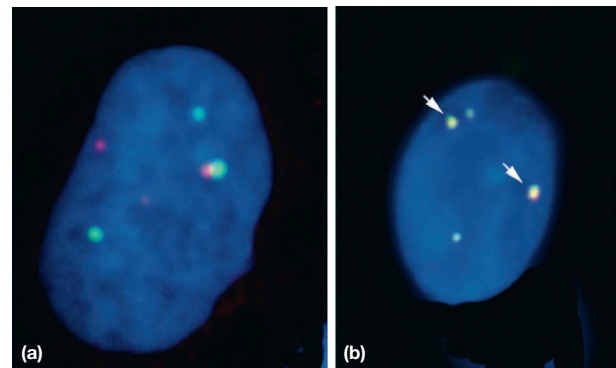


Figure 10 SMN gems are associated with CBs. By dual staining with anti-SMN antibody (red) and anti-coilin antibody (green), we see that SMN gems are separate from, but often associated with, CBs in some cells such as HeLa (a), but are most often found to be completely coincident with CBs (arrows) as seen in the breast cancer cell HCC1937 (b). Nuclei are stained with DAPI.

staining appear coincident in CBs (**Figure 10(b)**), so SMN is generally thought of as a component of CBs and gems are not distinct nuclear bodies. At a molecular level, SMN protein complexes with a number of other factors in a structure referred to as an 'assemblysome'. SMN accompanies newly assembled snRNPs back into the nucleus, but it is not clear if this process is related to their localization in CBs. However, this role for SMN is consistent with the idea that CBs may be involved in the recycling and/or biogenesis of splicing factors. An extension of this hypothesis is that CBs are sites of preassembly of pols I, II, and III transcription/RNA processing complexes or transcriptosomes.

Histone locus bodies

Once thought to be CBs, histone locus bodies (HLB) have been found to be related but distinct subnuclear structures. HLBs are spatially associated with the histone gene clusters on chromosomes 1 and 6. Although HLBs contain concentrations of many factors also found in CBs, such as coilin, they also have several unique factors, many of which are involved in histone mRNA processing. Among these are U7snRNA, the U7 snRNPs Lsm10 and Lsm11, stem-loop-binding protein (SLBP), FADD-like IL-1 β -converting enzyme (FLICE)-associated huge protein

(FLASH), nuclear protein, ataxia telangiectasia (NPAT), negative elongation factor, and simplekyn. Although the function of HLBs is unclear, their components and localization suggest that they may be sites of histone mRNA processing.

Conclusion

While this article has focused on some of the most-studied non-nucleolar compartments of mammalian nuclear structure, nuclear subdomains are still being discovered, and there is already evidence for the existence of multiple other compartments about which less is known. These include the perinucleolar compartment (PNC), Sam68 bodies, cleavage bodies, polycomb bodies, and Oct1/PTF/transcription (OPT) domains. Some of these structures (PNC) are hallmarks of cancer cells. While the nuclear compartments described here all have different components and are implicated in distinct functions, a common theme is that these compartments serve to concentrate and sort factors and RNAs involved in similar processes, presumably to enhance the efficiency and/or regulation of those processes within the nucleus. While in most cases, proteins that concentrate in a particular domain or body are also present at low levels in other regions of the nucleoplasm, the concentration of numerous factors involved in the same aspect of gene expression or particular type of RNA metabolism will facilitate the formation of large, complex macromolecular structures. Alternatively, in some cases, the concentration at a particular site may serve as a means to sequester factors away from interactions with other factors or specific areas of the genome. While these various compartments comprise relative stable nuclear structures, often the factors within these structures are dynamic, showing rapid movements within or between nuclear compartments. In addition, these various domains and bodies may change in number, shape, or specific components during the cell cycle and different cellular states, including tissue-type differentiation. Essentially all of these nuclear compartments disassemble and reassemble during mitosis. The consideration of how specific gene loci may associate with or be impacted by particular structures is important and complex, but it is clear that the organization of the genome is intimately integrated with at least some of these nuclear non-chromatin compartments. Finally, it is interesting to note that some bodies and domains appear to interact frequently with each other, such as the association of PML bodies, CBs, splicing factor (SC-35) speckles, and paraspeckles. The diversity of subcompartments within nuclear structure makes clear that the bland, homogeneous vision of the interphase nucleus that persisted for many decades is now firmly dispelled. The interphase nucleus reflects the functional organization of the genome associated with a multitude of subcompartments each of which is implicated in

efficient expression or regulation of one or more of the myriad functions of cell nuclei.

See also: [Cell Architecture and Function: Chromosome Organization and Structure, Overview](#); [Nuclear Lamina](#); [Molecular Biology: mRNA Polyadenylation in Eukaryotes](#); [Nuclear Organization, Chromatin Structure, and Gene Silencing](#); [Nucleolus, Overview](#).

Further Reading

- Bond CS and Fox AH (2009) Paraspeckles: Nuclear bodies built on long noncoding RNA. *Journal of Cell Biology* 186(5): 637–644.
- Borden KL and Culjkovic B (2009) Perspectives in PML: A unifying framework for PML functions. *Frontiers in Bioscience* 14: 497–509.
- Clemson CM, Hutchinson JN, Sara SA, et al. (2009) An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Molecular Cell* 33: 717–726.
- Cohen TV, Hernandez L, and Stewart CL (2008) Functions of the nuclear envelope and lamina in development and disease. *Biochemical Society Transactions* 36: 1329–1334.
- Cook PR (2010) A model for all genomes: The role of transcription factories. *Journal of Molecular Biology* 395(1): 1–10.
- Cremer T and Cremer M (2010) Chromosome territories. *Cold Spring Harbor Perspectives in Biology* 2: 1–22.
- Dechat T, Adam SA, Taiman P, Shimi T, and Goldman RD (2010) Nuclear lamins. *Cold Spring Harbor Perspectives in Biology* 2(11): a000547.
- Hager GL, McNally JG, and Misteli T (2009) Transcription dynamics. *Molecular Cell* 35(6): 741–753.
- Hall LL and Lawrence JB (2010) XIST RNA and architecture of the inactive X chromosome: Implications for the repeat genome. *Cold Spring Harbor Symposia on Quantitative Biology* 75: 345–356.
- Hall LL, Smith KP, Byron M, and Lawrence JB (2006) Molecular anatomy of a speckle. *The Anatomical Record Part A* 288: 664–675.
- Johnson CV, Primorac D, McKinstry M, et al. (2000) Tracking COL1A1 RNA in osteogenesis imperfecta. Splice-defective transcripts initiate transport from the gene but are retained within the SC35 domain. *Journal of Cell Biology* 150: 417–431.
- Kolb SJ, Battle DJ, and Dreyfuss G (2007) Molecular functions of the SMN complex. *Journal of Child Neurology* 22: 990–994.
- Lallemand-Breitenbach V and de Thé H (2010) PML nuclear bodies. *Cold Spring Harbor Perspectives in Biology* 2: 1–17.
- McKeown PC and Shaw PJ (2009) Chromatin: Linking structure and function in the nucleolus. *Chromosoma* 118: 11–23.
- Negorev D and Maul GG (2001) Cellular proteins localized at and interacting within ND10/PML nuclear bodies/PODs suggest functions of a nuclear depot. *Oncogene* 20: 7234–7242.
- Nizami Z, Deryusheva S, and Gall JG (2010) The Cajal body and histone locus body. *Cold Spring Harbor Perspectives in Biology* 2: 1–12.
- Pederson T (2010) The nucleolus. *Cold Spring Harbor Perspectives in Biology* 2: a003889.
- Spector DL and Lamond AI (2011) Nuclear speckles. *Cold Spring Harbor Perspectives in Biology* 3: a000646.
- Wilson KL and Berk JM (2010) The nuclear envelope at a glance. *Journal of Cell Science* 123: 1973–1978.
- Zhong S, Salomon P, and Pandolfi PP (2000) The transcriptional role of PML and the nuclear body. *Nature Cell Biology* 2: E85–E90.

Nuclear Factor Kappa B

T D Gilmore, Boston University, Boston, MA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Apoptosis A programmed cell death pathway that is used by an organism to eliminate unwanted cells or to maintain organ size.

Cytokines Factors that regulate the growth and activity of blood cells and lymphocytes.

Proteasome A cellular complex composed of several proteases that degrade ubiquitinated proteins.

Protein kinase An enzyme that can add a phosphate group onto specific residues in a target protein, which then affects the activity of the target protein.

Signal transduction A molecular process by which a signal (often an extracellular compound) effects a cellular response. This is usually a multicomponent step-wise

intracellular pathway that is initiated by binding of an extracellular ligand to a cell-surface receptor, which sets into motion a cascade of cytoplasmic protein–protein interactions and enzymatic activities that ultimately lead to the activation of a transcription factor which then regulates the expression of a biologically relevant set of target genes.

Transcription factor A protein or protein complex that binds to specific DNA sequences to regulate (increase or decrease) the rate of transcription of a gene.

Ubiquitin A small (~8.6 kDa) polypeptide chain that can be conjugated to substrate proteins, which often targets them for degradation by the proteasome or serves as a posttranslational modification to regulate protein activity.

Introduction

Nuclear factor kappa B (NF κ B) is a family of structurally related DNA-binding proteins that act as transcription factors to regulate the expression of a wide variety of genes involved in many key cellular and organismal processes. The discovery of mammalian NF κ B is generally attributed to Ranjan Sen and David Baltimore, who were studying the regulation of expression of a specific immunoglobulin gene; however, it is now clear that earlier researchers studying an avian retroviral oncoprotein (v-Rel) and a protein (Dorsal) required for the establishment of *Drosophila* embryonic polarity were also studying NF κ B activity. This article focuses primarily on the mammalian NF κ B system, although most of the discussion is also applicable to the NF κ B system in other vertebrates and invertebrates. NF κ B activity is controlled by a multi-component signal transduction pathway: that is, in most cells NF κ B is inactive in the cytoplasm until one of many inducers initiates the signal transduction pathway that ultimately converts an inactive form of NF κ B into an active nuclear DNA-binding protein. Misregulation of NF κ B activity has been implicated in a number of human diseases, including several chronic inflammatory diseases and cancer, and the NF κ B pathway is used by several viruses as part of their pathology or replication. As such, the NF κ B pathway is the subject of intense focus for pharmacological intervention.

NF κ B Proteins and Protein Structure

The NF κ B transcription factor family is comprised of several evolutionarily conserved proteins that are related through a protein domain, usually called the 'Rel homology domain (RHD)', which has sequences required for DNA binding, dimerization, and nuclear localization (Figure 1(a)). In simple marine organisms, such as sea anemones, corals, and sponges,

there is a single NF κ B protein, while in *Drosophila*, there are three NF κ B proteins (Dorsal, Dif, and Relish). In vertebrates, the NF κ B protein family is comprised of five proteins: p50/p105, p52/p100, c-Rel, RelA (also called p65), and RelB. There is also an avian retroviral oncoprotein, v-Rel, which is a member of this family.

All NF κ B proteins contain the ~300-amino-acid RHD toward their N termini; however, NF κ B proteins can be subdivided into two distinct groups based on sequences C-terminal to the RHD. One group (including p50/p105, p52/p100, and Relish) contains C-terminal sequences that fold back to inhibit the RHD, and thus these C-terminal sequences must be removed to release the active DNA-binding protein (i.e., p50 or p52 from the precursors p105 and p100, respectively). A second group (c-Rel, RelA, RelB, Dorsal, and Dif) contains C-terminal sequences that are not cleaved and are required for these proteins to activate transcription.

NF κ B proteins bind to 9–10 bp DNA sites, called ' κ B sites', located in the regulatory regions of many cellular and viral promoters. To bind DNA, NF κ B proteins must form dimers. These dimers can, with few exceptions, contain any combination of homodimer or heterodimer, although the most common NF κ B dimer in many cell types is a p50–RelA dimer (Figure 1(b)). The affinity of the various NF κ B subunits for one another and the affinity of the various dimers for specific target DNA sites can vary greatly. These two considerations provide an array of diversity in the NF κ B dimers that are found in different cells, on the DNA sites that can be recognized by different NF κ B dimers, and in the genes (and consequently cellular responses) that these dimers regulate.

Regulation of NF κ B Activity

In most cells, NF κ B is present in the cytoplasm in an inactive state, due to interaction with a family of inhibitor proteins,

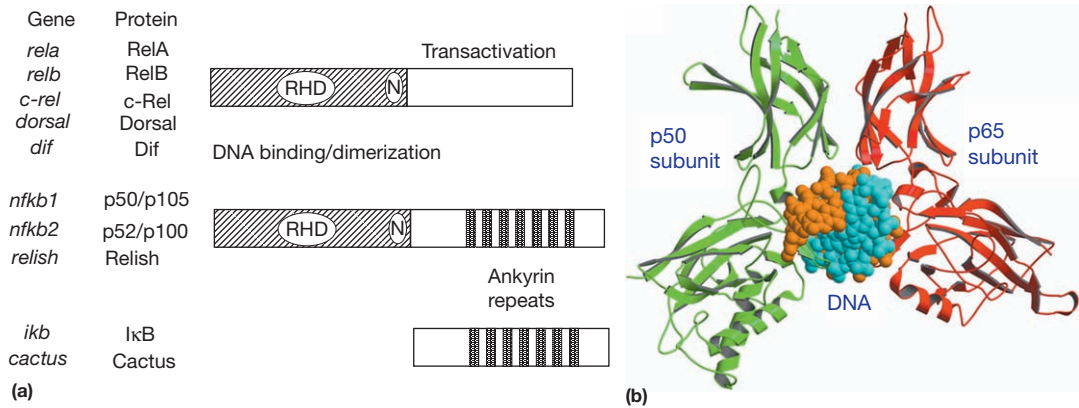


Figure 1 NFκB proteins. (a) Shown are the generalized structures, protein names, and genes encoding the two classes of NFκB proteins and the IκB proteins. RHD, Rel Homology domain; N, nuclear localization signal; vertical bars, ankyrin repeats. (b) X-ray crystal structure of the NFκB p50–RelA (p65) dimer on DNA. Figure provided courtesy of Gourisankar Ghosh (University of California, San Diego).

Table 1 Some inducers of NFκB activity

General category	Examples
Bacteria	<i>Salmonella</i> , <i>Staphylococcus</i> , Lipopolysaccharide
Viruses	HTLV-1, Epstein–Barr virus, hepatitis B virus
Cytokines	Interleukins, tumor necrosis factor
Oxidative stress	Hydrogen peroxide, reperfusion ischemia
Growth factors/ hormones	Platelet-derived growth factor, epidermal growth factor, insulin, transforming growth factor
Drugs	Various chemotherapeutic agents

termed ‘inhibitor of κB (IκB)’ proteins. IκB proteins include IκBα, IκBβ, IκBε, IκBζ, Bcl-3, and the C-terminal sequences of the NFκB precursor proteins p105 and p100. All IκB proteins contain multiple ~30 amino acid repeats called ankyrin repeats, which are protein–protein interaction domains. Through these ankyrin repeats, a single IκB protein binds directly to protein sequences in the RHD of an NFκB dimer. Binding of an IκB to NFκB inhibits the activity of the dimer in two ways: it causes NFκB to be in the cytoplasm and it blocks NFκB DNA binding. In some cases, especially with IκBζ and Bcl-3, the NFκB–IκB complex can still bind to DNA and the IκB protein serves as a transcriptional co-activator.

A variety of inducers (Table 1), including many cytokines, cell stresses, and viral or bacterial infections, can activate NFκB. Often the pathway is initiated by binding of an extracellular ligand to a cell-surface receptor; for example, tumor necrosis factor binding to its receptor or a bacterial or viral antigen to a Toll-like receptor. Engagement of the receptor by ligand causes receptor clustering, which promotes the recruitment of adaptor molecules to the intracellular domain of the receptor complex. In almost all cases, NFκB activity is then induced through activation of a cytoplasmic kinase complex called ‘IκB kinase (IKK)’. IKK is composed of three subunits: two related kinases (IKKα and IKKβ) and a scaffold protein called NEMO. Activation of the IKK complex generally requires phosphorylation of residues in the activation loop of the IKKβ kinase, and probably also posttranslational modifications of NEMO. Activated IKK then phosphorylates the IκB protein, which, in turn, acts as a signal for IκB to be degraded by the ubiquitin–proteasome

pathway. The liberated NFκB dimer can then enter the nucleus and bind to specific sites on DNA to activate target gene expression. Activation of NFκB is usually transient (often lasting no more than 30 min) primarily because one of the NFκB target genes is the gene encoding IκB. Therefore, newly synthesized IκB can cause NFκB to be resequenced in an inactive form in the cytoplasm, thus returning the pathway to its original inactive state.

Although the overall pathway of activation is similar (as in Figure 2), there are actually two distinct pathways for activation of NFκB, depending on the nature of the IκB protein and IKK subunit involved. In one pathway (the classical or canonical pathway), a separate IκB molecule, such as IκBα bound to p50–RelA, is phosphorylated by the IKKβ subunit, which leads to the complete proteolysis of IκBα. In a second NFκB activation pathway (called the alternative or noncanonical pathway), which involves a complex such as p100–RelB, a C-terminal site in p100 is phosphorylated by IKKα, which then leads to the partial proteolysis of the C-terminal ankyrin repeat sequences of p100, ultimately yielding an active p52–RelB dimer, which can also translocate to the nucleus to regulate gene expression.

The activity of NFκB is further modulated by mechanisms in addition to interaction with IκB proteins. For example, NFκB proteins can undergo posttranslational modifications (such as phosphorylation, acetylation, and methylation) that affect their activity, they can interact with a variety of other proteins (including many other transcription factors), and the IKK proteins can also have effects on chromatin-binding proteins to affect gene expression from κB site-containing promoters.

Genes and Biological Processes Regulated by NFκB

Normal Biological Roles for NFκB

There are over 300 genes whose transcription is known to be directly regulated by NFκB activity, and these genes are involved in a multitude of biological processes (Table 2). However, a variety of genetic and cellular studies have established that one of the key and evolutionarily conserved functions of NFκB is in regulating the innate immune response.

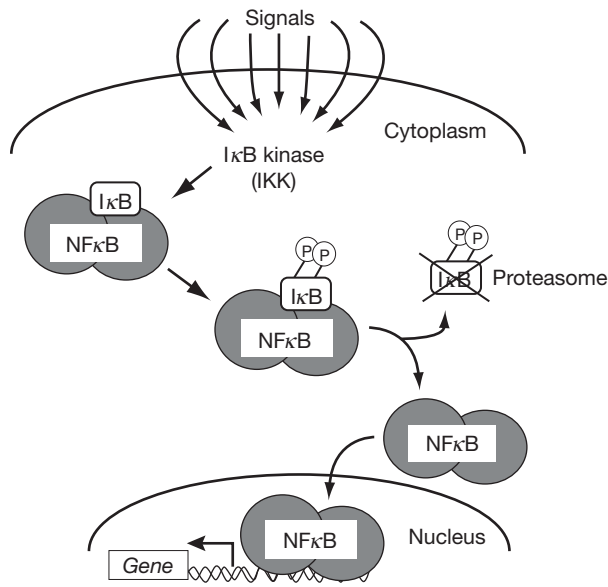


Figure 2 The NFκB signal transduction pathway. Under most circumstances, NFκB dimers (e.g., p50–RelA) are located in the cytoplasm of cells bound to inhibitory IκB proteins. Upon receiving an inducing signal, an IκB kinase (IKK) is activated. Active IKK phosphorylates IκB, and phosphorylated IκB then serves as a substrate for the ubiquitin–proteasome degradation pathway. Free NFκB can then enter the nucleus to increase or decrease specific target gene transcription. The system is eventually returned to its resting state through an auto-regulatory system, in that NFκB turns on expression of the gene encoding IκB, and newly synthesized IκB then re-sequesters NFκB in the cytoplasm.

Table 2 Some biological processes and relevant genes regulated by NFκB

Biological process	Relevant NFκB target genes
Angiogenesis	Vascular endothelial growth factor (VEGF)
Apoptosis	Bcl-X1, Bcl-2, IAP, TRAF, Fas
Cell adhesion	ICAM-1, VCAM-1
Cell growth	Cyclin D1, c-Myc
Stress	Cyclooxygenase, superoxide dismutase, nitric oxide synthetase
Innate immunity	Interleukins, interferon, tumor necrosis factor
Viral promoters	HIV-1, cytomegalovirus

In *Drosophila*, the Dorsal, Dif, and Relish proteins are induced to enter the nucleus by bacterial and fungal pathogens. These fly NFκB proteins then directly increase the expression of a family of genes called ‘cecropins’, which encode antifungal and antibacterial peptides. In mammals, many genes that encode antibacterial peptides, cytokines, and interferons are rapidly induced after infection with various viral and bacterial pathogens. Moreover, mice in which individual genes for c-Rel, RelB, p50, and p52 are disrupted (knockout mice) have defects in T- and B-cell-mediated immune responses. For example, c-rel knockout mice have B cells with greatly reduced survival and proliferation in response to antigen stimulation.

NFκB proteins are, however, also involved in the control of several other genes and processes that are not directly related to

Table 3 Some diseases associated with chronic NFκB activation

Arthritis
Asthma
Cancer
Diabetes
Inflammatory bowel disease
Ischemia/reperfusion damage
Sepsis

immunity. These include certain developmental processes, apoptosis, cell growth, and cell stress. In terms of development, NFκB activity (as directed by Dorsal) is required for the proper establishment of dorsal–ventral polarity in the early *Drosophila* embryo. In mammals, NFκB activity is involved in proper limb development, due to the regulation of a gene called *Twist*.

NFκB transcription factors also regulate cell survival in many contexts, either by promoting or, more often, by inhibiting apoptosis. For example, NEMO, IKKβ, and RelA are each required to protect embryonic liver cells from apoptosis induced by tumor necrosis factor, and knockout mice missing each of these factors die embryonically due to severe liver degeneration. Thus, the genes encoding several inhibitors of apoptosis are activated by NFκB.

NFκB and Disease

Aberrant NFκB activity has been implicated in several human diseases (Table 3), including many cancers and chronic inflammatory diseases, such as inflammatory bowel disease, asthma, and arthritis.

Many cancer cells have constitutively active, nuclear NFκB activity. In certain B-cell malignancies, notably diffuse large B-cell lymphoma, Hodgkin’s lymphoma, and multiple myeloma, NFκB activity is dysregulated due to either amplification of the c-rel gene, mutations that inactivate the gene encoding IκB, or mutations in genes encoding upstream activators of NFκB signaling. Furthermore, several oncogenic human viruses, such as Epstein–Barr virus (B-cell lymphoma and nasopharyngeal carcinoma), HTLV-1 (T-cell leukemia), and hepatitis B virus (liver cancer), encode proteins that activate NFκB. However, in many other cancers one finds chronic nuclear NFκB activity, which is usually made up of p50–RelA dimers, in the apparent absence of mutations in the NFκB signaling pathway or associated viral infection. That increased NFκB activity is important for the growth and survival of several cancers is verified by studies showing that inhibition of NFκB activity often causes the tumor cells either to die (undergo apoptosis) or to become more sensitive to apoptosis-inducing agents.

Because NFκB controls the expression of many cytokines, it is also consistently active in many situations involving acute or chronic inflammation. Moreover, in several animal model systems inhibition of NFκB activity can reduce inflammation. For example, mice with knockouts in the genes encoding p50 or c-Rel are resistant to experimentally induced arthritis. Finally, several human immunodeficiency diseases are caused by mutations that reduce activation of NFκB in response to pathogen infection; for example, mutations in the gene

Table 4 Some inhibitors of NFκB signaling

<i>Inhibitor</i>	<i>Molecular target</i>
Aspirin (high doses)	IKK
Curcumin	IKK
Glucocorticoids	RelA
Green tea compounds	IKK
Bortezimib	Proteasome
Sulindac	IKK

encoding NEMO are often the cause of the human immunodeficiency disease incontinentia pigmentosum.

Pharmacological Regulation of NFκB Activity

Because of the role of increased NFκB activity in inflammatory diseases and cancer, NFκB has received much attention as a molecular target for pharmacological intervention. As a multi-component pathway, NFκB activation can be inhibited at a variety of levels, including activation of IKK, IKK-mediated phosphorylation of IκB, ubiquitination and proteasome-mediated degradation of IκB, nuclear translocation of NFκB, and DNA binding and transactivation by NFκB. Indeed, NFκB inhibitors that act at each of these steps have been described (partial list in Table 4). Nevertheless, to date, most research effort has focused on identifying NFκB inhibitors that target the IKK complex, due to its central role in controlling the NFκB pathway and to successes in identifying specific kinase inhibitors in other systems.

Several natural products with anticancer or anti-inflammatory activities have been shown to inhibit activation of NFκB. These natural NFκB inhibitors include gold compounds (anti-arthritis), green tea (anticancer), curcumin (anticancer and anti-inflammation), and various plant extracts (anti-inflammation). In addition, some commonly used anti-inflammatory compounds, such as glucocorticoids, ibuprofen, and aspirin, have been reported to have anti-NFκB activity in

experimental model systems; however, in these cases, it is not clear that their biological effects in people are exerted through inhibition of NFκB.

Because NFκB is constitutively active in many cancers, clinical trials are underway with several compounds known to inhibit NFκB. These trials have reported promising results in the use of bortezimib, a small-molecule inhibitor of the proteasome, for the treatment of multiple myeloma, an NFκB-dependent B-cell cancer. Such compounds inhibit activation of NFκB by blocking proteasome-mediated degradation of IκB.

See also: [Bioenergetics: Cell Death by Apoptosis and Necrosis](#); [Protein/Enzyme Structure Function and Degradation: Proteasomes, Overview](#); [Ubiquitin System](#); [Ubiquitin-Like Proteins](#).

Further Reading

- Baud V and Karin M (2009) Is NF-κB a good target for cancer therapy? Hopes and pitfalls. *Nature Reviews Drug Discovery* 8: 33–40.
- Beyaert R (ed.) (2003) *Nuclear Factor-κB: Regulation and Role in Disease*. The Netherlands: Kluwer Academic.
- Ghosh G, Van Duyn G, Ghosh S, and Sigler PB (1995) Structure of NF-κB homodimer bound to a κB site. *Nature* 373: 303–310.
- Ghosh S (ed.) (2006) *Handbook of Transcription Factor NF-kappaB*. Boca Raton, FL: CRC Press.
- Gilmore TD (ed.) (2006) NF-κB: From basic research to human disease. *Special Issue: Oncogene* 25: 6680–6899.
- Hayden MS and Ghosh S (2008) Shared principles in NF-κB signaling. *Cell* 132: 344–362.
- Kawai T and Akira S (2007) Signaling to NF-κB by Toll-like receptors. *Trends in Molecular Medicine* 13: 460–469.
- Sen R and Baltimore D (1986) Inducibility of κ immunoglobulin enhancer-binding protein NF-κB by a post-translational mechanism. *Cell* 47: 921–928.
- Silverman N and Maniatis T (2001) NF-κB signaling pathways in mammalian and insect innate immunity. *Genes and Development* 15: 2321–2342.

Relevant Website

<http://www.nf-kb.org> – Gilmore TD (2009): Rel/NF-κB transcription factors.

Nuclear Genes Involved in Mitochondrial Function and Biogenesis

A Tzagoloff, Columbia University, New York, NY, USA

C L Dieckmann, University of Arizona, Tucson, AZ, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 100–104, © 2004, Elsevier Inc.

Glossary

ATP synthetase/hydrolase An enzyme of the inner membrane able to form adenosine triphosphate (ATP) from adenosine diphosphate (ADP) and inorganic phosphate by utilizing the energy of the proton gradient generated during the oxidation reactions catalyzed by the respiratory chain complexes. It can also form a proton gradient by hydrolyzing ATP to ADP and inorganic phosphate.

Chaperones Proteins that promote the folding or unfolding of proteins; in the present context, they also facilitate association of subunit polypeptides into hetero-oligomeric proteins.

Eukaryotes Organisms with cells having a nucleus enclosed by a membrane.

Organelle A structure or membrane-enclosed compartment of a cell.

pet mutants Viable mutants able to derive ATP from glycolysis but not from oxidative phosphorylation.

Proteome Denotes all of the proteins of an organism or compartment of the cell.

Respiration In the biochemical sense used here, it is the process by which the electrons and protons extracted from reduced fuels such as carbohydrates, fats, and proteins are transferred to oxygen.

rho zero mutants Mutants devoid of mitochondrial DNA.

TIM complex An organized assembly of proteins that mediates the translocation of proteins from the TOM complex to the inner membrane or matrix of mitochondria.

TOM complex An organized assembly of proteins that mediates the translocation of a protein from the cytoplasm to the intermembrane space of mitochondria.

Only a small fraction of the protein constituents of mitochondria is encoded in the small vestigial genome of the organelle. Most of the genes governing the biogenesis of mitochondria reside in the chromosomal DNA of the nucleus. Referred to as *PET* genes, they represent a sizable fraction of the total genetic information in the nucleus. *PET* genes code for the enzymes making up the different metabolic pathways housed in mitochondria, for the substrate and protein transport systems, for the transcriptional, translational machineries involved in expression and regulation of mitochondrial genes, and for proteins governing fusion and fission of mitochondria.

Historical Background

All life depends on adenosine triphosphate (ATP), the universal currency of biological energy. The chemical energy inherent in the structure of this molecule is used to support a large number of different of phenomena including chemical, osmotic, electrical, and mechanical work. Most of the ATP in eukaryotic organisms is supplied by the mitochondrion, a complex cellular organelle composed of two membranes (outer and inner) and two soluble compartments (intermembrane space and matrix). Biogenesis of mitochondria occurs by replenishment and addition of newly synthesized proteins and phospholipids to the pre-existing organelle. In dividing cells, there is a net increase of mitochondrial mass, which is distributed among the progeny. This process is aided by the ability of mitochondria to undergo fusion and fission. Mitochondria, therefore, are not synthesized *de novo* but rather serve as a template for their own faithful replication. The genetic determinants of mitochondria are

located in two physically separate compartments of the cell. Most of the information is provided by the *PET* genes of the nucleus, but a small number of genes present in mitochondrial DNA (mtDNA) are also indispensable for maintenance of fully functional mitochondria.

The discovery of an extra-nuclear genetic element (*rho*) in yeast by Boris Ephrussi and the later demonstration by Gottfried Schatz that this factor corresponds to a unique DNA present in mitochondria were the foundations for much of the subsequent work aimed at understanding the function of this genome. Despite the large variability in the organization, mode of expression, and gene compositions of mtDNA among different members of the plant and animal kingdoms, some general statements can be made, nonetheless. Close to 3000 different mitochondrial genomes have been sequenced as of this writing. In all instances, at least some components of the mitochondrial translational apparatus such as the genes for the endogenous ribosomal RNAs and transfer RNAs (tRNAs) are encoded in mtDNA. Another important attribute of this genetic system is that it codes for a small number of mitochondrial proteins that encode proteins crucial for the respiratory and ATP-generating capacities of the organelle. In most organisms, the protein-coding genes of mtDNA specify several subunits of cytochrome oxidase, cytochrome *b*, and one or more subunits of the proton-translocating ATP synthetase/hydrolase. The *rho* zero mutants lacking this set of genes produce mitochondria that are defective in respiration and oxidative phosphorylation because they lack the genes for catalytically important components of the respiratory chain, ATP synthetase, and the ribosomal RNA and tRNAs needed for translation of the endogenous messenger RNA *s* (mRNAs).

A much larger share of the information essential for maintaining the structural and functional integrity of mitochondria is contributed by *PET* genes, which are located in the chromosomal DNA of the nucleus.

Origin of Nuclear Gene Products with Functions in Mitochondria

The endosymbiont hypothesis, first proposed by Lynn Margulis in 1968, was a landmark advance in our thinking of how cells evolved in response to major changes in their environment. According to this hypothesis, many organelles of modern-day eukaryotic cells, including mitochondria and chloroplasts, are stripped-down products of what originally were free-living bacteria. In the case of mitochondria, a nucleated cell dependent on glycolysis as its sole source of ATP was invaded by a bacterium with an aerobic mode of energy metabolism. Such unions, when advantageous to the host, became fixed during evolution. There is compelling evidence to support the bacterial origin of subcellular structures that inhabit eukaryotic cells. Most notably is the presence in mitochondria and chloroplasts of autonomously replicating genomes, and of all the enzymatic machinery needed to express the encoded genes into their respective RNA and protein products. Significantly, the transcriptional and translational machineries of mitochondria and chloroplasts are much more related to their bacterial counterparts than to those operating in the cytoplasm of eukaryotic cells.

The extent to which the mitochondrial genome contributes toward the propagation of the organelle has been severely reduced due to a continual transfer of genetic information to the host's nuclear DNA. Today, the existence of mitochondria relies in large measure on the expression of gene products encoded in nuclear DNA. The loss of mitochondrial autonomy is clearly evident from the properties of cells that have been depleted of mtDNA. Such genome-less variants termed *rho* zero mutants have been created in yeast and mammalian (e.g., human) cells. Despite the complete absence of mitochondrial DNA, *rho* zero mutants contain subcellular structures with the characteristic morphology and most of the metabolic pathways of mitochondria. These observations point to the nonessentiality of mtDNA for the perpetuity of mitochondria, but do not minimize the importance of this genome for their functional competence.

Number of Mitochondrially Related Nuclear Genes

The number of distinct proteins in yeast mitochondria is still not known precisely but based on large-scale subcellular localization studies with the unicellular yeast *Saccharomyces cerevisiae*, 700–800 is a reasonable estimate. This corresponds to approximately 10–12% of the different proteins known to exist in this organism. Of course, this number will vary in more complex eukaryotes, where depending on their tissue of origin, mitochondria may have become highly specialized with a lesser functional complexity. For example, in mammals, mitochondria of organs such as heart and muscle have simpler metabolic profiles than those of liver and consequently have fewer proteins.

Yeast as a Model for Mitochondrial Studies

For practical reasons, *S. cerevisiae* has become an important model for the study of *PET* genes. A vast arsenal of genetic and biochemical tools is directly applicable to the analysis of mitochondria in this organism. Also available are the complete sequences of the 16 nuclear chromosomes and of mtDNA. Finally, because of its status as a facultative anaerobe, this yeast is able to tolerate most mutations that impair mitochondrial function as long as it is supplied with a fermentable source of carbon and energy.

Most of what we presently know about *PET* genes stems from three different sources. An important body of information has been derived from traditional studies of the intermediary metabolism carried out by mitochondria. In some instances, such studies have predated the discovery of mitochondria as the cellular repository of the pathways in question. Genetic analysis has played an equally important role. Respiratory defective mutants of yeast have been instrumental in defining many nuclear genes that impinge on all aspects of mitochondrial function, biogenesis, and human pathologies. More recently, the high-throughput proteomic approach has further augmented the list of mitochondrially related gene products, most with yet-to-be-understood functions.

Gene Functions

Based on the phenotype of the respective mutants, *PET* genes can be either essential or nonessential. Nonessential genes, which make up the larger class, are defined by *pet* mutants. Such strains are viable but because of their genetic lesions are unable to produce ATP by means of oxidative phosphorylation. The products of *PET* genes have functions that are related either directly or indirectly to mitochondrial metabolism, but not to the continuity of mitochondria as a structural compartment of the cell. The second, less abundant class consists of essential genes needed for the survival of the cell. Their products provide functions relevant to the structural integrity of mitochondria or to some essential aspect of metabolism. Mutations in the latter prevent growth on both fermentable and nonfermentable substrates.

Nonessential *PET* Genes

These genes encompass a wide range of mitochondrial functions. The most important, from a quantitative point of view, are involved in the expression of the products encoded in mtDNA (Figure 1). They consist of proteins that function in replication of mtDNA, and transcription and processing of mitochondrial RNA. In yeast, they also code for all except one of the mitochondrial ribosomal proteins, as well as most of the tRNA synthetases and translation initiation, elongation, and termination factors. A second important group of genes code for proteins that catalyze most of mitochondrial metabolism. This includes enzymes of the tricarboxylic acid (TCA) cycle, fatty acid metabolism, some aspects of amino acid biosynthesis, and a number of other pathways needed for the synthesis of important cofactors such as hemes A and B, coenzyme Q, and lipoic acid.

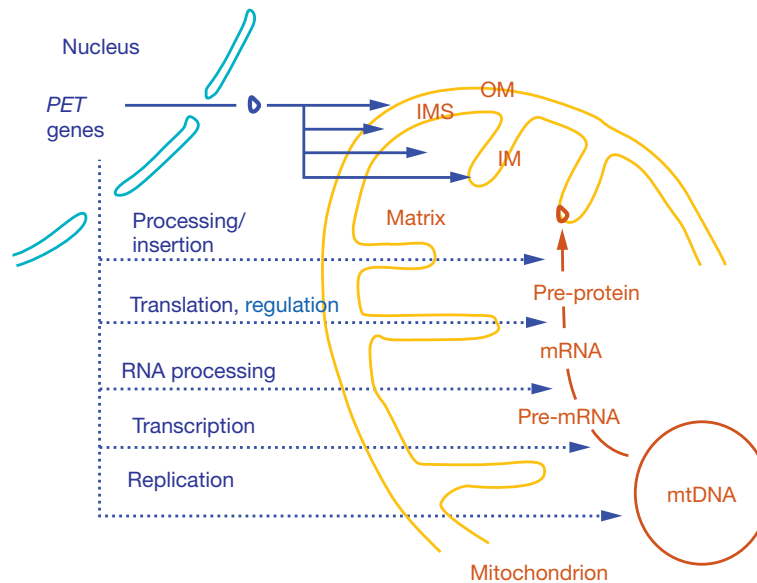


Figure 1 Dependence of mitochondrial gene expression on products of *PET* genes. The dashed blue arrows show some of the functions contributed by nuclear DNA for expression of the mitochondrially encoded subunits of the ATPase and respiratory enzymes. In yeast the pre-mRNAs can be multigenic transcripts that are matured by excision of intronic sequences and endonucleolytic cleavages. Some proteins are synthesized as precursors (pre-protein) with amino-terminal sequences that are removed by specific proteases. The solid blue arrows denote *PET* gene products targeted to the outer membrane (OM), intermembrane space (IMS), inner membrane (IM), and internal matrix compartment. The broken arrows indicate nuclear factors that act at different stages of mitochondrial gene expression and promote insertion of the resultant products into the inner membrane.

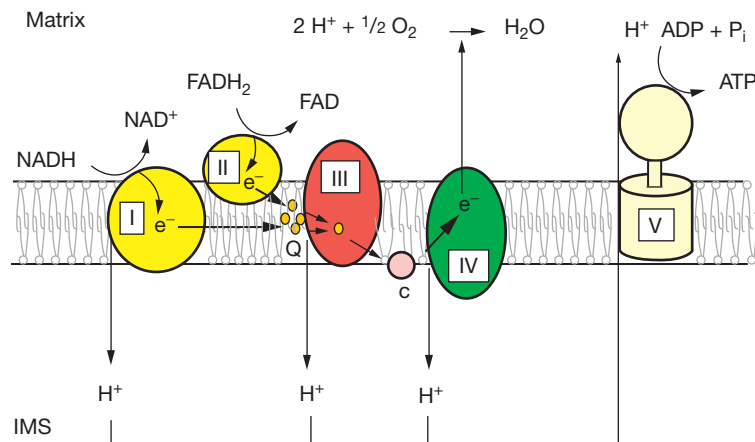


Figure 2 Schematic representation of the respiratory complexes and ATP synthetase in the mitochondrial inner membrane. The four complexes of the respiratory chain are NADH-coenzyme Q reductase or complex I (yellow), succinate-coenzyme Q reductase or complex II (yellow), coenzyme QH₂-cytochrome c reductase or complex III (red), and cytochrome c oxidase or complex IV (green). Coenzyme Q is denoted by the small orange circles and cytochrome c by the pink circle. The ATP synthetase/hydrolase or complex V is shown with the F₁ ATPase protruding into the matrix compartment. In addition to catalyzing the transfer of electron from NADH and FADH₂ to oxygen, three of the complexes promote the vectorial transfer of protons from the matrix to the intermembrane space (IMS). Protons are recycled by the ATP synthetase during synthesis of ATP from ADP and inorganic phosphate.

The ATP synthetase/hydrolase and the enzymes making up the respiratory chain (Figure 2) are all hetero-oligomeric proteins, with more than a dozen different subunit polypeptides in most cases. Assembly of the subunits into the final holoenzymes is a complex process that requires the assistance of nuclearly encoded factors. Some of the chaperones act by facilitating a topologically correct insertion of specific subunits into the inner membrane, by preventing nonproductive

interactions of hydrophobic proteins, or by still other unknown mechanisms. Because these chaperones target specific subunits, they represent a substantial fraction of the total number of mitochondrial proteins. For example, some dozen nuclear mutants have been reported to be cytochrome oxidase deficient due to mutations in proteins that affect late events in assembly of this single enzyme. Recent evidence from studies of yeast has uncovered another group of important proteins

that regulate translation of mitochondrial gene products so that they are balanced with their functional partners expressed by the nucleo-cytoplasmic genetic system. Finally, many *PET* genes, coding for novel proteins, still lack functions that can be assigned at this time.

Essential *PET* Genes

The highly regulated metabolism of eukaryotic cells depends on the sequestration of the enzymes making up the various pathways in their appropriate compartments. Proteins translated in the cytoplasm but destined to function in one of the internal compartments of mitochondria are imported and sorted by means of the TOM and TIM translocases of the outer and inner membranes, respectively. Mutations in components of either of these protein transport systems are lethal to the cell for several reasons. An immediate effect of the transport defect is a mis-localization of proteins in the cytoplasm. This abrogates the accretion of newly synthesized material by mitochondria, resulting in their eventual loss.

Mitochondria house proteins that function in pathways essential for the overall economy of the cell. For example, certain steps of heme B biosynthesis occur in mitochondria. Since heme B is a prosthetic group of enzymes (e.g., P450 hydroxylases) other than those involved in respiration, cells unable to produce heme B are not viable. This also applies to certain enzymes of amino acid biosynthesis located in mitochondria. In animals, carbamoyl synthetase and ornithine transcarbamylase, two important enzymes of arginine biosynthesis, catalyze their respective reactions in the mitochondrial matrix. Mitochondria also house the pathway for producing the iron-sulfur prosthetic groups of a large number of redox proteins that function not only in mitochondria but also in the cytosol.

Some nuclear genes code for proteins that are apportioned between the cytoplasm and mitochondria. This is true for some tRNA modification enzymes and for aminoacyl-tRNA synthetases. In yeast, six mitochondrial aminoacyl-tRNA synthetases are expressed from nuclear genes that also code for the same cytoplasmic enzymes. Mutations in these essential genes, therefore, result in cell lethality.

Regulation

Many *PET* genes of yeast are regulated at the level of transcription by oxygen and carbon source. Oxygen stimulates the synthesis of heme in mitochondria mainly at later steps of the pathway catalyzed by oxygenases. Heme activates the binding of the Hap1 transcriptional activator to the upstream activation sites of some genes for respiratory chain proteins such as cytochrome *c*, thereby activating their transcription. Glucose has the opposite effect. It suppresses transcription of genes involved in alternative fuel usage by means of a complex regulatory pathway called catabolite repression. In the absence of glucose, repression is relieved and transcription of glucose-responsive

genes is enhanced by the Hap2/3/4/5 transcriptional activator complex.

Mitochondrial gene expression is regulated primarily at the posttranscriptional level by products of *PET* genes. Mitochondrial RNA polymerase recognizes a simple promoter and there is little regulation of its activity except that it is more abundant in the absence of glucose and the presence of oxygen. Most mitochondrial genes are transcribed as long multi-genic RNAs. Points of regulation include processing of 5' and 3' ends of RNA to the mature length, excision of introns, and RNA stability. Initiation of translation is regulated by factors that attract ribosomes to the correct internal translational start sites of mitochondrial mRNAs. Posttranslation modification and assembly of the constituent subunits of the respiratory and ATP synthetase complexes also depend on the intervention of factors encoded by *PET* genes subject to these regulatory systems.

Human Disorders

A growing list of human pathologies has been related to mitochondrial defects and the responsible mutations mapped to RNA and protein-coding genes of mtDNA. Such mutations manifest a wide range of clinical phenotypes including encephalomyopathies, cardiomyopathy, deafness, diabetes, optic neuropathy, and dystonia. Similar presentations are elicited by mutations in nuclear DNA that affect mitochondrial metabolism either directly through lesions in the structural genes for enzymes of the TCA cycle, fatty acid oxidation, and mtDNA integrity or indirectly by interfering with metabolite and protein import, assembly of respiratory chain complexes, metal homeostasis, and protein turnover. In view of their central role in oxidative metabolism, a process dependent on a large number of nuclear *PET* genes, future studies of model organisms such as yeast can be expected to substantially increase the number of mitochondrially related diseases.

See also: [Bioenergetics: Giant Mitochondria \(Megamitochondria\); Mitochondrial Auto-Antibodies; Mitochondrial Genes and Their Expression: Yeast.](#)

Further Reading

- Alberts B, Bray D, Lewis J, et al. (1994) *Molecular Biology of the Cell*, 3rd edn. New York: Garland Publishing.
- Ernster L and Schatz G (1981) Mitochondria: A historical review. *Journal of Cell Biology* 91: 227s–255s.
- Margulis I (1981) *Symbiosis in Cell Evolution*. New York: W. H. Freeman.
- Neupert W and Brunner M (2002) The protein import motor of mitochondria. *Nature Reviews Molecular Cell Biology* 3: 555–565.
- Reinders J, Zahedi RP, Pfanner N, Meisinger C, and Sickmann A (2006) Toward the complete yeast mitochondrial proteome: Multidimensional separation techniques for mitochondrial proteomics. *Journal of Proteome Research* 5: 1543–1554.
- Scheffler IE (1999) *Mitochondria*. New York: Wiley-Liss.
- Tzagoloff A and Dieckmann CL (1990) *PET* genes of *Saccharomyces cerevisiae*. *Microbiological Reviews* 54: 211–225.

Nuclear Lamina

B M Paschal, University of Virginia, Charlottesville, VA, USA

J B Kelley, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Bryce M. Paschal, volume 3, pp. 92–95, © 2004, Elsevier Inc.

Glossary

CaaX A four-amino-acid peptide motif that directs the C-terminal isoprenylation of a protein. It is named for its consensus sequence: C, cysteine; a, an aliphatic amino acid; and X, any amino acid.

Chromatin Complex of DNA, histones, and nonhistone proteins contained within the nucleus of eukaryotic cells. Heterochromatin is transcriptionally inactive.

Electron microscopy A type of microscopy that resolves fine structural details of cells by using a focused beam of electrons to penetrate the specimen.

Integral membrane protein Membrane-bound protein containing one or more regions that span the hydrophobic core of the phospholipid bilayer.

Lamina Fibrous network of lamins and associated proteins on the inner surface of the nuclear envelope. Lamins are not related to extracellular laminin proteins.

Nuclear envelope Double membrane structure that surrounds the nucleus.

Nuclear pore complex Specialized channels that span the nuclear envelope, which are responsible for bidirectional transport of macromolecules between the nucleus and the cytoplasm.

Structure of the Nuclear Envelope

The first descriptions of the nuclear envelope were provided by nineteenth-century biologists, notably Fleming, who used light microscopy to examine the nuclei of plants and protozoa. A century later, the application of electron microscopy to the analysis of a variety of intracellular organelles revealed the unique structure and organization of the nuclear envelope. The nuclear envelope consists of two membranes, an inner membrane and an outer membrane, each of which is a bilayer of lipids and proteins (Figure 1). The intermembrane space, or lumen, is a site of calcium storage in the cell, and special channels in the membrane facilitate calcium release and re-uptake upon physiological stimulation. The outer nuclear membrane is continuous with the endoplasmic reticulum membrane, thus there is continuity between the lumen of the nuclear envelope and the lumen of the endoplasmic reticulum.

Structure of the Nuclear Lamina

In 1966, Fawcett published a seminal electron microscopy study showing the presence of a “fibrous lamina on the inner aspect of the nuclear envelope” in several cell types. The organization of the lamina as a meshwork of 10-nm-diameter filaments adjacent to the inner nuclear membrane was later revealed in samples prepared from frog oocyte nuclei (Figure 1).

Composition of the Lamina

The development of biochemical procedures for isolating nuclei from cells and for preparing the lamina–nuclear pore complex fraction from nuclei set the stage for identifying proteins that comprise the lamina.

Lamins

The major proteins of the nuclear lamina, termed lamins, are members of the intermediate filament protein superfamily. The molecular weights of lamins range from 60 to 70 kDa depending on the isoform, and all lamins have related primary and secondary structures. The amino- and carboxyl-terminal domains of lamins are globular, the latter encoding a nuclear localization signal. Like other members of the intermediate filament superfamily, lamin proteins contain a central rod domain composed of four α -helical segments that mediate coiled-coil interactions resulting in the formation of lamin dimers (Figure 2). Lamin dimers assemble into tetramers, which, in turn, assemble into 10-nm-diameter filaments that comprise the nuclear lamina. The approximately orthogonal arrangement of lamin filaments observed by electron microscopy implies the presence of proteins that mediate interfilament cross-links; however, the proteins and interactions that establish and maintain this higher-order structure have not yet been identified.

Lamin Genes

Lamins are classified as A type and B type. The B-type lamins are expressed in most cell types, whereas A-type lamins, which include lamin C, are generally expressed in differentiated tissues. In some tissues, lamin A/C is not expressed until after birth. There is a single lamin A gene (*LMNA*) in mammals that, through alternative splicing, produces transcripts for two lamin A isoforms and one lamin C isoform. There are two lamin B genes (*LMNB1* and *LMNB2*) that together encode three lamin B isoforms. The phylogenetic distribution of lamins extends to invertebrates, but lamins are absent from plants and fungi.

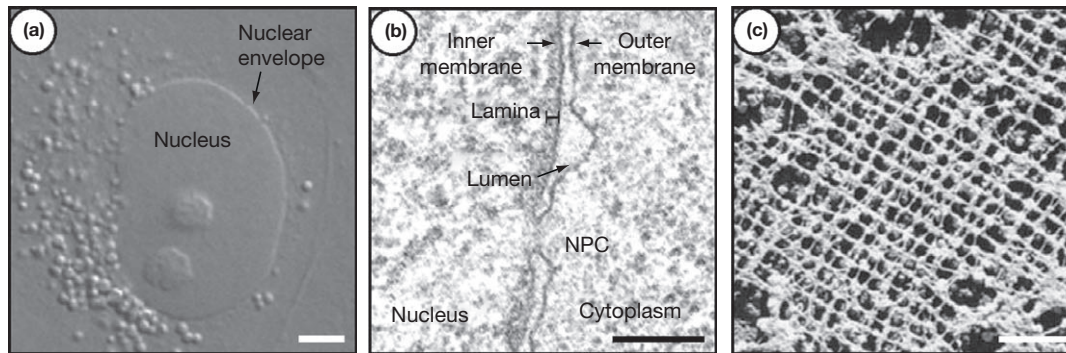


Figure 1 The nuclear membrane viewed by light and electron microscopy. (a) Differential interference contrast microscopy of a cell showing the nuclear envelope. Scale = 5 μm . (b) Thin section electron microscopy showing a small region of the nuclear envelope and associated structures. Scale = 0.2 μm . (c) Metal-shadowed nuclear lamina from a frog oocyte viewed by electron microscopy. Scale = 0.5 μm . (c) Modified from Aebersold U, Cohn J, Buhle L, and Gerace L (1986) The nuclear lamina is a meshwork of intermediate-type filaments. *Nature* 323: 560–564.

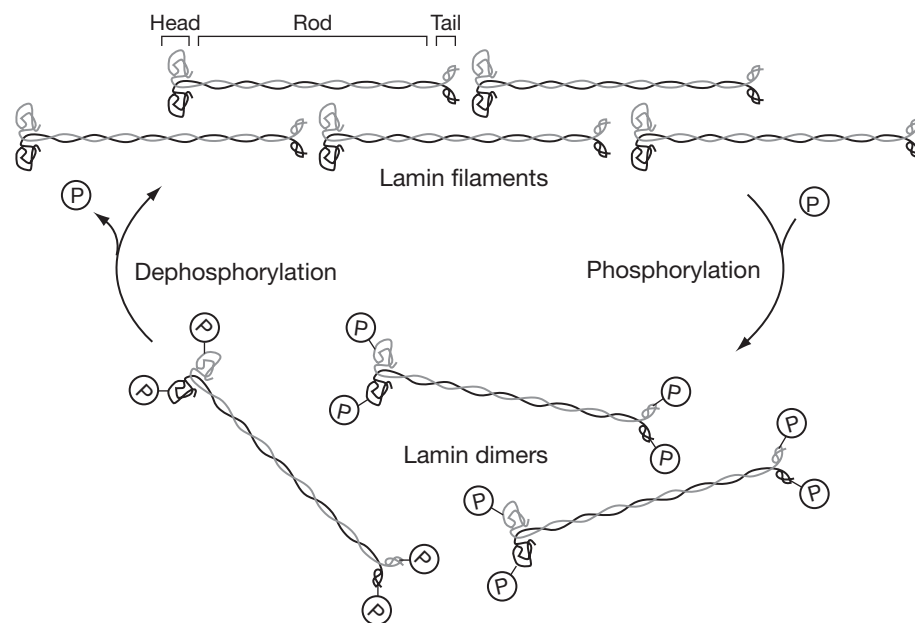


Figure 2 Structure of the nuclear lamina is regulated by phosphorylation. Lamin filaments undergo phosphorylation by kinases in mitosis, causing disassembly into lamin dimers. At the end of mitosis, the lamins are dephosphorylated by phosphatases, which return the lamins to an assembly competent state.

Lamin-Associated Proteins

A group of integral membrane proteins, which are unique to the inner nuclear membrane, bind directly to nuclear lamins and play critical roles in modulating lamina structure and function. This group includes LAP2, Emerin, Man1, and LBR. These proteins are anchored in the inner nuclear membrane by hydrophobic, membrane-spanning segments, while the rest of the protein projects into the nucleoplasm, where it contacts the lamina. Lamin-associated proteins also interact with DNA-binding proteins such as BAF and GCL that mediate chromatin-attachment to the nuclear membrane. In addition, lamin-associated proteins are part of a network of polypeptides that link the nucleoplasm to the actin filaments in the cytoplasm.

Functions of the Lamina

The nuclear lamina performs both structural and regulatory functions. These functions are not mutually exclusive, because alterations in structure can change the capacity of the nuclear lamina to regulate nuclear functions. Characterization of the regulatory functions of the nuclear lamina is a relatively recent area of investigation, and this knowledge base should expand dramatically in the next several years.

Nuclear Envelope Assembly after Mitosis

Disassembly of the nuclear envelope and lamina occurs at prometaphase, a period of the mitotic cell cycle that precedes

chromosome segregation. The nuclear envelope fragments into membrane vesicles and the lamin filaments are disassembled into lamin dimers. Near the end of mitosis in late anaphase, the nuclear envelope-derived vesicles and lamin proteins are reassembled around the chromosomes of the two new daughter cells. The reassembly pathway involves multiple steps, beginning with the binding of lamins and vesicles to the surface of chromosomes. The vesicles then fuse to generate membranes. The incorporation of lamins into the lamina continues after membrane formation by virtue of nuclear import of lamins through newly formed nuclear pore complexes.

Nuclear Size

Experimental manipulation of the nuclear reassembly pathway has shown that lamins and lamina-associated proteins are necessary for the several-fold increase in volume of the nucleus during the cell cycle. The lamina is, therefore, a critical regulator of nuclear size.

Nuclear Shape and Rigidity

The lamina forms a cage-like structure within the nucleus and is tethered to the inner nuclear membrane. As such, the lamina plays a major role in defining nuclear shape. A striking example of how lamins determine the shape of the nucleus is lamin B3, a lamin isoform that was shown by Furukawa and Hotta in 1993 to be responsible for the unusual hook shape of nuclei in spermatocytes. The lamina also provides rigidity that resists deformation and aids in restoration of shape following deformation. Nuclei from the nematode *Caenorhabditis elegans*, rendered lamin deficient, show little resistance to deformation, and a failure to restore nuclear shape.

Positioning of Nuclear Pore Complexes

Nuclear pore complexes span both the inner and outer nuclear membranes. Integration of nuclear pore complexes into the membrane is mediated by GP210, an integral membrane protein. The relative spacing of nuclear pore complexes in the two-dimensional plane of the nuclear envelope is maintained by the nuclear lamina. B-type lamins mediate this spacing through direct interactions with nuclear pore complexes. Loss of this functionality results in a failure to maintain the proper distribution of nuclear pore complexes in the nuclear envelope.

Regulation of Gene Expression

Lamins can have a positive or negative influence on gene expression, depending on the context. Lamins and lamina-associated proteins bind to chromatin, resulting in the attachment of a subset of chromatin to the inner nuclear membrane which can be observed by microscopy. Because there is a strong correlation between localization of chromatin at the nuclear periphery and gene silencing, the nuclear lamina has historically been viewed as a scaffold for anchoring transcriptionally inactive heterochromatin. Lamin B is highly concentrated at the

inner nuclear membrane, and, together with several DNA-binding proteins that repress transcription, mediates the negative influence on transcription at the nuclear periphery. By contrast, it has been shown that lamin A can turn on expression of muscle-specific genes, a clear indication of positive effects on transcription. Lamin A is found at both the inner nuclear membrane and throughout the nucleoplasm, and the latter pool is thought to promote transcription. Thus, whether lamins influence transcription positively or negatively is dependent on the particular lamin protein, and its subnuclear distribution.

Regulation of DNA Replication

Lamins play an important role in DNA replication by facilitating the recruitment of proteins needed for initiation of replication. The original link between lamins and DNA replication was made by Newport and colleagues in 1990, using a system that reconstitutes nuclear assembly *in vitro*. Depletion of lamin B from the system resulted in small nuclei that failed to initiate DNA synthesis. Subsequent experiments have implicated lamina-associated proteins in DNA replication as well.

Posttranslational Modification of the Lamina

The lamin proteins are subject to modifications that profoundly affect function.

Phosphorylation

Gerace and Blobel showed in 1980 that phosphorylation is used as a molecular switch to control the assembly properties of lamins. Disassembly of lamin filaments at prometaphase is driven by lamin phosphorylation (Figure 2). Lamin disassembly occurs specifically at prometaphase because the kinase that mediates lamin phosphorylation is activated during the mitotic phase of the cell cycle. The phosphorylated lamins are subsequently dephosphorylated by phosphatases, rendering them competent for reassembly into filaments in late anaphase and telophase.

Isoprenylation

A cysteine-containing motif in the carboxyl terminus of lamin B, called a CaaX box, is the site for attachment of an isoprenyl group. This lipid moiety inserts into the inner nuclear membrane and mediates the stable membrane attachment of lamin B. During mitosis, lamin B remains stably associated with nuclear-envelope-derived vesicles. The isoprenylated carboxyl terminus of lamin A is removed by proteolysis. Lamin C is not isoprenylated because the CaaX box is not encoded in the lamin C transcript.

Proteolysis

Lamins are regulated by proteolysis in two functionally distinct pathways. After translation, the carboxyl-terminal 18 amino acids of lamin A are removed by the lamin-A-specific protease



Figure 3 Images of children that carry an HGPS mutation in lamin A. The pictures were taken when the children were 4 and 20 years of age (left and right panels, respectively), and are courtesy of the Progeria Research Foundation.

Zmpste24. This cleavage reaction removes the isoprenylated carboxyl-terminus and the capacity of lamin A to associate directly with membranes. During apoptosis, a group of proteases called caspases cleave the lamins at multiple sites. Lamin cleavage by caspases is commonly used as a biochemical indicator of apoptosis.

Lamins in Human Disease

Mutations in genes encoding proteins of the nuclear lamina cause human diseases, collectively called laminopathies. More than 300 mutations have been identified in *LMNA*, the gene encoding the lamin A/C proteins. Remarkably, different *LMNA* mutations give rise to at least a dozen different diseases, many with distinct phenotypes. The clinical severity of a particular *LMNA* mutation frequently varies, even among members of a given family. This indicates that penetrance of the phenotype can be highly dependent on the genetic background.

Emery–Dreifuss Muscular Dystrophy

Particular mutations in the genes encoding either lamin A or emerin proteins can cause Emery–Dreifuss muscular dystrophy. Patients display tendon contractures, degeneration of skeletal muscle, and electrical conduction problems in cardiac muscle. *LMNA* mutations are also linked to familial partial lipodystrophy, which causes partial loss of fat tissue and late-onset insulin-resistant diabetes. A major unanswered question is how laminopathies are manifest in specific tissues because lamins are expressed in all cell types. Currently, there are two hypotheses addressing the tissue-specific basis of laminopathies, particularly those affecting muscle function. The structural hypothesis proposes that muscular dystrophies result from fragile nuclei that are susceptible to muscle contraction-induced physical damage. The gene-expression hypothesis is based on the concept that lamin deficiencies alter nuclear architecture in a manner that changes the pattern of gene expression. Tissues affected by laminopathies are derived

from the mesoderm, suggesting lamin A plays a particularly critical role in development of this cell lineage.

Hutchinson–Gilford Progeria Syndrome

In 2003, research groups in the United States and France discovered that mutations in *LMNA* are responsible for the premature aging disease progeria. A rare disease with an incidence of 1 in 4 million births, phenotypes associated with progeria are manifest after birth, and include reduced subcutaneous fat, alopecia, and complications of the cardiovascular system (Figure 3). Children with progeria die of heart attack and stroke at an average age of 13.

Progeria mutations in the *LMNA* gene generate a form of the lamin A protein (termed Progerin) that is defective for posttranslational modification. Progerin lacks the proteolytic cleavage site that normally releases lamin A from its membrane anchorage mediated by an isoprenyl group. Constitutive attachment of Progerin to the membrane of the inner nuclear envelope is the source of dominant effects on nuclear morphology and function. The mechanism by which Progerin effects are transduced throughout the cell are still under investigation.

See also: **Cell Architecture and Function:** Mitosis; Nuclear Pores and Nuclear Import/Export; **Molecular Biology:** Chromatin: Physical Organization.

Further Reading

- Aebi U, Cohn J, Buhle L, and Gerace L (1986) The nuclear lamina is a meshwork of intermediate-type filaments. *Nature* 323: 560–564.
- Broers JLV, Ramaekers FCS, Bonne G, Yaou RB, and Hutchison CJ (2006) Nuclear lamins: Laminopathies and their role in premature ageing. *Physiological Reviews* 86: 967–1008.
- Dechat T, Pfliegerhaer K, Sengupta K, et al. (2008) Nuclear lamins: Major factors in the structural organization and function of the nucleus and chromatin. *Genes and Development* 22: 832–853.
- Stewart CL, Roux KJ, and Burke B (2007) Blurring the boundary: The nuclear envelope extends its reach. *Science* 318: 1408–1412.

Nuclear Organization, Chromatin Structure, and Gene Silencing

M W Vitalini, G Dialynas, and L L Wallrath, University of Iowa, Iowa City, IA, USA
S R Mackey, St. Ambrose University, Davenport, IA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Chromosome territories Specific regions within the nucleus that are occupied by a given chromosome.

Epigenetics Heritable changes in gene expression that are not due to alterations in DNA sequence.

Euchromatin Gene-rich regions of the genome that decondense during interphase of the cell cycle to allow for transcription.

Heterochromatin Gene-poor regions of the genome that possess repetitive elements and remain condensed throughout the cell cycle.

Histones Small basic proteins that assemble into nucleosomes, the fundamental packaging unit of chromatin.

Lamina The network of proteins, including lamins, that lines the inner side of the nuclear envelope.

Nuclear envelope A layer of two membranes that encapsulates the nucleus and is punctuated with pores.

Nuclear Envelope

A distinguishing feature of the eukaryotic cell is the presence of a nucleus. The nucleus is partitioned from the cytoplasm by the nuclear envelope, which is composed of an inner and outer membrane (Figure 1). The nuclear envelope is studded with nuclear pores that allow for shuttling of macromolecules between the nucleus and cytoplasm. The inner side of the nuclear envelope is lined with a network of intermediate filament proteins called lamins, which interact with additional proteins that possess an LEM domain (a conserved stretch of amino acids first identified in the proteins LAP2, emerin, and MAN1) (Figure 1). Collectively, these proteins make up the nuclear lamina, which provides structural support for the nucleus, plays a role in DNA replication, and participates in signal transduction. LEM domain proteins also interact with the small DNA-binding protein barrier-to-autointegration factor (BAF) that connects the lamina with chromatin, providing anchor points for organization. The lamina contains SUN domain proteins (possessing a conserved amino acid sequence first identified in Sad1p and UNC-84) that span the inner nuclear membrane and connect with KASH domain proteins (possessing a conserved amino acid sequence first identified in the proteins Klarsicht, ANC-1, and Syne Homology) within the peri-nuclear space (Figure 1). KASH domain proteins interact with cytoskeletal components such as actin (Figure 1). Thus, a series of protein–protein interactions link the cytoplasm and nucleus, which allows for signals to be transmitted between the cellular compartments.

Chromatin Packaging

A function of the nucleus is to house the genomic DNA. The DNA is wrapped around histones and associated with nonhistone chromosomal proteins that collectively make up chromatin. The smallest unit of chromatin is the nucleosome, which is composed of 146 base pairs (bp) of DNA wrapped twice around an octamer of the core histone proteins H2A, H2B, H3, and H4. In most eukaryotes, a variable region of linker

DNA associates with linker histone H1 between histone octamers. Interactions among nucleosomes are thought to facilitate chromatin folding, which is responsible for condensing the DNA to ratios of 1:10 000. Chromatin condensation allows chromosomes that are centimeters in length to fit into a nucleus that is 6 μm (1/1500th of a centimeter) in diameter, a major packaging feat.

Depending on the cell type, different regions of the genome are loosely packaged into transcriptionally competent euchromatin, allowing for gene expression. Other regions are highly packaged into transcriptionally silent heterochromatin. In nearly all cells, repetitive DNA sequences located near the centromeres of chromosomes are packaged into constitutive heterochromatin that remains condensed throughout the cell cycle. In contrast, facultative heterochromatin contains genes that are silenced but can be reactivated by repackaging into euchromatin at specific times in development or upon environmental stimulation. The inactive X-chromosome in mammalian females, also known as the Barr body, is an example of facultative heterochromatin. In somatic cells, one of the two X-chromosomes of the female is inactivated upon differentiation. This inactivation equalizes the amount of gene product produced from the X-chromosome in males and females.

Atop the hierarchy of chromatin packaging are epigenetic modifications of the DNA and histones that influence chromatin compaction and nuclear processes, including gene expression, DNA replication, and DNA repair. Epigenetic literally means ‘above the genes’ and refers to the addition of specific chemical groups to DNA, such as cytosine methylation, and posttranslational modifications to histones, such as acetylation of lysine residues on histone tails. These modifications mark genomic regions and can persist through cell divisions, but do not alter the underlying DNA sequence. DNA methylation can attract specific proteins that regulate gene silencing. Likewise, posttranslational modifications on histones can serve as docking sites for specific effector proteins that promote open or closed chromatin structures. Combinations of histone modifications are proposed to code for specific biological functions such as gene activation, gene repression, and chromatin condensation. Importantly, these epigenetic marks are reversible,

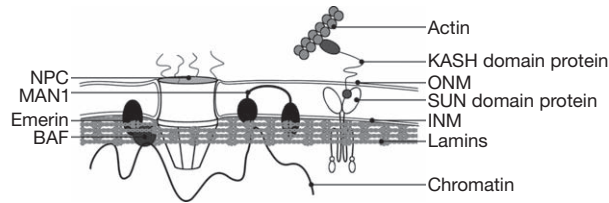


Figure 1 Diagram of the nuclear envelope. The nuclear envelope is composed of an outer nuclear membrane and an inner nuclear membrane that is lined with lamina. Nuclear pores punctuate the membrane and serve as conduits for the transfer of macromolecules between the nucleus and cytoplasm. LEM domain-containing proteins such as MAN1 and emerin associate with the inner nuclear membrane. The small DNA-binding protein BAF associates with LEM domains and connects the lamina to chromosomes. SUN domain proteins span the inner nuclear membrane and link to KASH domain proteins, which associate with cytoskeletal components such as actin. NPC, nuclear pore complex; ONM, outer nuclear membrane; INM, inner nuclear membrane.

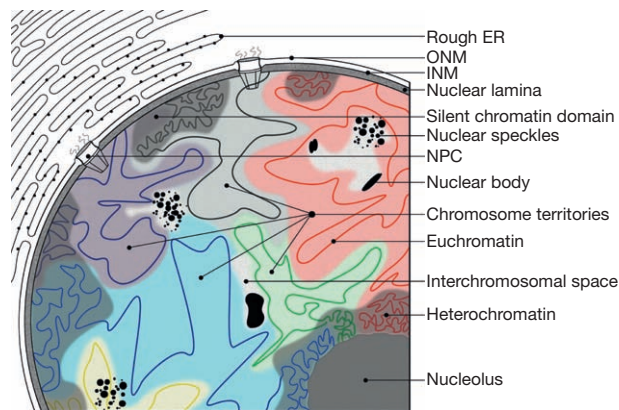


Figure 2 Diagram of nuclear organization. Lightly shaded areas represent chromosome territories in which chromosomes (solid color lines) are confined. Darkly shaded overlays represent domains of heterochromatin, which are frequently at the nuclear periphery. Many types of nuclear bodies occupy interchromosomal spaces and are thought to represent storage sites for proteins and sites of specific nuclear processes.

which distinguishes them from permanent changes in the DNA sequence. This epigenetic plasticity allows the genome to respond to developmental cues and changes in the environment by altering gene expression.

Nuclear Organization and Gene Expression

Chromosomes are not randomly arranged within the nucleus, but exist within defined three-dimensional regions called chromosome territories (Figure 2). In the late 1800s and early 1900s, Carl Rabl and Theodor Boveri were among the first scientists to note the territorial organization of chromosomes. In the 1980s, the advent of chromosome painting technology provided the ability to visualize individual chromosomes by hybridization with fluorescent DNA sequences that recognize specific chromosomes. The results of such analyses supported a

territorial organization of chromosomes. Interchromosomal space makes up the nuclear volume that is not occupied by chromosomes and can vary in amount depending on the cell type. The interchromosomal space contains various types of nuclear bodies, which act as localized storage houses for nuclear factors and nucleoprotein assembly factories that carry out processes such as RNA splicing. The spatial relationship among human chromosome territories was recently determined using a molecular technique involving ligation of DNA segments that are in close proximity to one another, coupled with new massive sequence technologies. The results of these experiments generated a genome-wide map of chromosomal interactions at the 1-Mb level. The data strongly supported the existence of chromosome territories and demonstrated that the entire genome is arranged as a 'fractal globule' rather than an 'equilibrium globule', which suggests that the chromosomes are arranged as interdigitated territories, not intertwined strands.

The radial position of a chromosome (position along the axis that runs from the center of the nucleus to the nuclear edge) within the nucleus of interphase animal cells correlates with gene activity. This correlation is best illustrated by comparative studies of human chromosomes 18 and 19. Chromosome 19 is gene rich and resides within the nuclear interior, whereas gene-poor chromosome 18 is associated with the nuclear periphery. In support of the nuclear periphery as a silent zone, the majority of heterochromatin within many types of cells resides in close proximity with the nuclear envelope. This finding suggested the possibility that nuclear envelope-associated proteins might play a role in chromatin packaging and chromosome organization. Consistent with this idea, mutations in mammalian genes encoding lamins result in reduced levels of constitutive and facultative heterochromatin as well as altered gene expression.

To directly study the role of the nuclear periphery in gene expression, several research groups have designed experiments using a variety of model systems to tether a chromosomal region to the nuclear envelope. Tethering is accomplished by utilizing two molecular components. The first component is a cell line or transgenic organism that carries reiterated binding sites for the *Escherichia coli* LacI repressor protein (*lac* operator repeats) positioned upstream of a reporter gene. The second component is a transgene that expresses a fusion protein consisting of a DNA-binding domain of the LacI repressor and a nuclear envelope protein. Upon expression, the fusion protein binds upstream of the reporter gene. This association results in tethering of the chromosomal region possessing the reporter gene to the nuclear periphery via the LacI nuclear envelope fusion protein (Figure 3).

Using the fruit fly *Drosophila* as a model, association of a reporter gene with a lamin protein resulted in localization of the reporter to the nuclear envelope and silencing (Figure 3). As a negative control, association of a LacI-GFP (GFP, green fluorescent protein) with the reporter did not result in peripheral positioning or cause gene silencing. Association with LacI-lamin caused limited silencing, as expression of neighboring genes was unaffected. Results similar to these have been observed in mammalian cells where tethering with nuclear envelope proteins such as emerin, Lap2 β , and lamin showed differential effects on expression of neighboring genes on the chromosome. The lack of silencing in some cases might be due

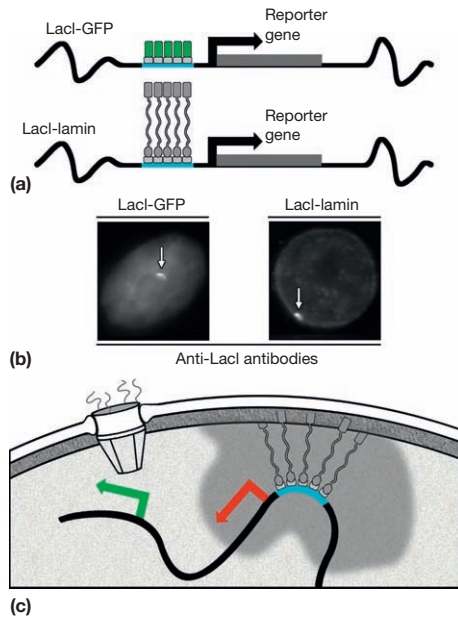


Figure 3 Molecular tethering experiments used to direct reporter genes to the nuclear periphery. (a) Diagram of reporter genes inserted within the genome. The reporter gene has *lac* operator repeats (blue line) positioned upstream of the promoter (bent arrow) of a reporter gene. The *lac* operator repeats can be bound by a LacI-GFP fusion protein (green, top) as a negative control or a LacI-lamin protein (gray, bottom) as the experimental situation. (b) Nuclear localization of the reporter gene. Salivary gland nuclei from *Drosophila* larvae expressing the fusion proteins were stained with anti-LacI antibodies and the location of the reporter gene is indicated by a focus of intense staining (arrow). Note that association with LacI-lamin resulted in the localization of the reporter gene to the nuclear periphery. (c) Diagram of the effect of tethering LacI-lamin. The reporter gene was transcriptionally silenced (red bent arrow). However, a closely linked gene (green bent arrow) retained the ability to be expressed, despite close proximity to the nuclear envelope.

to proximity to nuclear pores. In yeast, high levels of gene expression are observed when genes are located close to nuclear pores. Differences in gene expression upon localization to the nuclear periphery might also be due to the local epigenetic chromatin landscape. Treatment with the drug trichostatin A, which inhibits histone deacetylase proteins, results in the accumulation of acetyl groups on lysine residues of histone tails and increased expression of genes associated with the nuclear envelope. Taken together, these studies demonstrate that proximity to the nuclear periphery is not sufficient for gene silencing, but may promote a silent state if conditions are favorable.

The position of genes within the nucleus is dynamic. Activation by developmental signals and/or environmental stimulation sometimes correlates with the movement of genes to the exterior edge of chromosomal territories, the interchromosomal space, or migration to the nuclear interior. This movement depends on nuclear actin and/or myosin, suggesting that these nucleoskeletal components play a role in the chromatin repositioning that occurs during transcriptional activation or repression.

Nuclear Position and Disease

The organization of chromosomes within the nucleus impacts human disease. Exposure to radiation, chemicals, and other environmental factors can induce double-strand breaks in chromosomal DNA. The cell uses multiple pathways to repair this damage. Nonhomologous end joining is a pathway that joins the broken ends of chromosomes and requires little to no DNA sequence homology. The net outcome of the repair process may be a chromosomal translocation in which segments from two different chromosomes are covalently linked. Given that the repair process involves physical association of the two broken ends, the breakpoints of the translocations that occur are a reflection of the spatial arrangement of chromosomes in the three-dimensional nucleus.

Given the nonrandom distribution of chromosomes within the nucleus, it follows that genes reside in specific positions. A large-scale characterization of gene positions has been undertaken in normal and invasive breast cancer cells during interphase of the cell cycle, the stage in which transcription occurs. The comparative data showed that a small subset of the genes underwent reorganization upon cancer transformation. Interestingly, these genes were not physically linked on the same chromosome. Thus, gene positioning might represent a new marker that could be used to identify cancer cells among a population of normal cells or serve as a means of distinguishing between cancer subtypes.

Conclusions

The astute observations made by early researchers led to the discovery that the nucleus is not a homogeneous collection of confined molecules. Within its spherical structure resides a highly complex organization of chromosomes and protein complexes, each with their assigned three-dimensional position. This organization is imposed by connections with the nuclear envelope and a proposed nuclear substructure. Chromosomes exist in territories that are, to a certain degree, related to the overall gene activity of the chromosome. Genes within a given territory can shift in location upon activation or repression. As a general rule, the nuclear interior is a place of gene activity, whereas the nuclear periphery is generally, but not always, a location of gene silencing. Chromosome and gene positioning is dynamic and can change upon alterations in activity and with disease.

See also: [Molecular Biology: Regulation of Chromatin Dynamics; Signaling: Melanocortin System.](#)

Further Reading

- Allis DC, Jenuwein T, Reingerg D, and Caparros M-L (eds.) (2007) *Epigenetics*. New York, NY: Cold Spring Harbor Laboratory Press.
- Chow J and Heard E (2009) X inactivation and the complexities of silencing a sex chromosome. *Current Opinion in Cell Biology* 21: 359–366.
- Chuang C-H, Carpenter AE, Fuchsova B, et al. (2006) Long-range directional movement of an interphase chromosome site. *Current Biology* 16: 825–831.
- Cremer T and Cremer M (2010) Chromosome territories. *Cold Spring Harbor Perspectives in Biology* 2: a003889.

- Croft JA, Bridger JM, Boyle S, et al. (1999) Differences in the localization and morphology of chromosomes in the human nucleus. *Journal of Cell Biology* 145: 1119–1131.
- Dialynas G, Speese S, Budnik V, Geyer PK, and Wallrath LL (2010) The role of *Drosophila Lamin C* in muscle function and gene expression. *Development* 137: 3067–3077.
- Dundar M, Ospina JK, Sung M-H, et al. (2007) Actin-dependent intranuclear repositioning of an active gene locus *in vivo*. *Journal of Cell Biology* 179: 1095–1103.
- Finlan FE, Sproul D, Thomson I, et al. (2008) Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genetics* 4: e10000039.
- Kumaran RI and Spector DL (2008) A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *Journal of Cell Biology* 180: 51–65.
- Lieberman-Aiden E, van Berkum NL, Williams L, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326: 289–293.
- Mahy NL, Perry PE, and Bickmore WA (2002) Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *Journal of Cell Biology* 159: 753–763.
- Matarazzo MR, Boyle S, E'Esposito M, and Bickmore WA (2007) Chromosome territory reorganization in a human disease with altered DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America* 104: 16546–16551.
- Meaburn KJ, Gudla PR, Khan S, Lockett SJ, and Misteli T (2009) Disease-specific gene repositioning in breast cancer. *Journal of Cell Biology* 187: 801–812.
- Meaburn KF, Misteli T, and Soutoglou E (2007) Spatial genome organization in the formation of chromosomal translocations. *Seminars in Cancer Biology* 17: 80–90.
- Reddy KL, Zullo JM, Bertolino E, and Singh H (2008) Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* 452: 243–247.
- Wilson KL and Berk JM (2010) The nuclear envelope at a glance. *Journal of Cell Science* 123: 1973–1978.

Nuclear Pores and Nuclear Import/Export

A Lange and A H Corbett, Emory University School of Medicine, Atlanta, GA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Anita H. Corbett, volume 3, pp. 109–114, © 2004, Elsevier Inc.

Glossary

Cargo Macromolecular substrate to be transported between the nucleus and the cytoplasm.

Classical nuclear localization signal (cNLS) The most common form of nuclear localization signal comprised of one or two clusters of basic amino acids.

GTPase Protein that binds to and hydrolyzes the nucleotide guanosine triphosphate (GTP).

Heterogeneous nuclear ribonucleoprotein (hnRNP) Abundant poly(A) + RNA-binding protein.

Messenger RNA (mRNA) Class of RNA that carries the information from the DNA to the cytoplasm and serves as the informational blueprint for protein synthesis.

Nuclear export signal (NES) An amino acid sequence within a protein that targets that protein for export from the nucleus.

Nuclear localization signal (NLS) An amino acid sequence within a protein that targets that protein for import into the nucleus.

Nuclear pore complex (NPC) The large proteinaceous channel through which macromolecular cargoes are transported in and out of the nucleus.

Nucleoporins Proteins that make up the NPC.

Ran A small GTPase that regulates the directionality of nuclear transport.

RanGAP Cytoplasmic protein that enhances the GTPase activity of Ran.

RanGEF Nuclear protein that facilitates the nucleotide exchange on Ran.

Ribosomal RNA (rRNA) Class of RNA that serves as a structural and catalytic component of ribosomes.

Transfer RNA (tRNA) Class of RNA that is charged with amino acids to decode the mRNA for protein synthesis.

Transport receptor/karyopherin/importin/exportin The receptors that bind to and recognize protein and RNA cargoes that are transported into or out of the nucleus.

U small nuclear RNA (U snRNA) Small RNAs that participate in splicing.

Nuclear Pores

All transport between the nucleus and the cytoplasm occurs through large protein channels that are embedded in the nuclear membrane. These channels, which are called nuclear pore complexes (NPCs), are composed of proteins referred to as nucleoporins. Small molecules such as water and ions can diffuse through NPCs, but larger cargoes (>40 kDa) such as proteins and RNA require the participation of soluble transport receptors.

Structure

NPCs are large (~30 MDa in yeast and ~60 MDa in vertebrates) structures that span the double membrane surrounding the nucleus and provide a transport channel between the nucleus and the cytoplasm. Many different studies have led to a model for the structure of the NPC (Figure 1). The core of the NPC consists of a cylinder with eightfold rotational symmetry, which perforates the nuclear envelope and surrounds a central, hourglass-shaped channel. In addition to this central channel, there are filaments that extend into the cytoplasm and a basket structure that extends into the nucleoplasm. Cargoes that move through the nuclear pores probably interact first with the peripheral structures, move through the central channel, and are then released into their target compartment.

Composition

Sophisticated biochemical studies have revealed the identity of the nucleoporin proteins (termed Nups) that make up the

NPC. Surprisingly, these studies reveal that both the budding yeast and vertebrate nuclear pores are composed of only approximately 30 distinct proteins that are largely conserved from budding yeast to humans (Figure 2 and Table 1). The NPC is large because most of these pore components are present in 16–32 copies per NPC. Many of these nucleoporins contain a characteristic repeat sequence of phenylalanine glycine (FG) repeats. These FG-repeat nucleoporins line the central channel of the pore and provide the conduit for movement of macromolecules through the pore complex.

Transport Mechanisms: Protein Trafficking

Transport through nuclear pores occurs in both directions, from the nucleus to the cytoplasm and from the cytoplasm to the nucleus. The desired direction of transport generally depends on the cargo to be transported. The cell has developed mechanisms to mark cargoes for transport and regulate the direction of the transport processes.

Targeting Signals

As with most intracellular targeting mechanisms, trafficking between the nucleus and the cytoplasm depends on amino acid sequences within the protein cargoes to be transported. However, unlike other targeting mechanisms such as mitochondrial and endoplasmic reticulum targeting, nucleocytoplasmic trafficking signals are not cleaved and instead remain

an integral part of the protein. This mechanism can allow the cargo proteins to undergo multiple rounds of import and/or export. The nuclear-targeting signal within the cargo protein generally mediates a physical interaction with a soluble nuclear transport receptor, which targets the cargo to the nuclear pore and mediates interactions with nucleoporins to facilitate translocation through the NPC.

Canonical signals for both nuclear protein import and nuclear protein export have been identified. Nuclear import signals are generally termed 'nuclear localization signals'

(NLSs), while signals that target proteins for export from the nucleus are termed 'nuclear export signals' (NESs). Putative NLS and NES signals in candidate proteins are typically identified through computer searches of protein sequences, but it is essential that their function be verified through experimental methods.

Two classes of NLS have been described in great detail: the classical NLS (cNLS) and the proline-tyrosine (PY)-NLS. The canonical cNLS consists of either a single cluster of basic amino acid residues (monopartite) or two clusters of basic amino acid residues separated by a 10-amino acid-linker sequence (bipartite). There is evidence, however, that the linker sequence in functional bipartite cNLS sequences can vary and contain up to 20 amino acids. The monopartite cNLS sequence is typified by the NLS found in the simian virus 40 (SV40) large T antigen (PKKKRKV) and the bipartite cNLS is typified by the nucleoplasmin NLS (KRPAATKKAGQAKKKK). The PY-NLS consists of an upstream basic or hydrophobic region followed by an R/H/KX₂₋₅PY motif. The C-terminal proline and tyrosine residues lend their name to this targeting signal, but there is evidence that other residues including phenylalanine, histidine, and methionine can also function in the final position of a PY-NLS and that the entire signal can accommodate considerable sequence variability due to contributions from multiple-binding regions. Other NLS sequences undoubtedly exist and have even been proposed, but none have been characterized in detail or generalized into predictive consensus sequences. The classical NES is comprised of a series of hydrophobic amino acids, generally leucine, isoleucine, or valine. Numerous

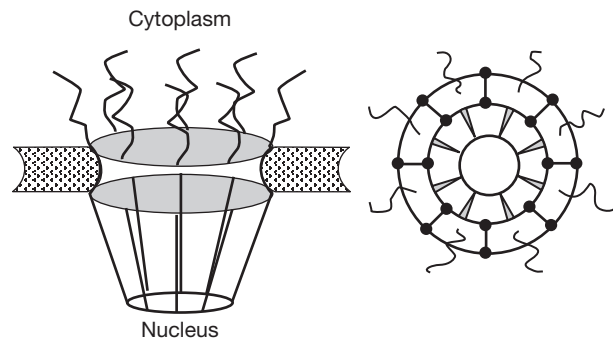


Figure 1 Schematic of the nuclear pore complex (NPC). Left: a schematic of the pore in the plane of the nuclear membrane showing the cytoplasmic filaments and the nuclear basket. Right: a schematic view of the pore looking down from the cytoplasmic face. The eightfold rotational symmetry of the pore is evident.

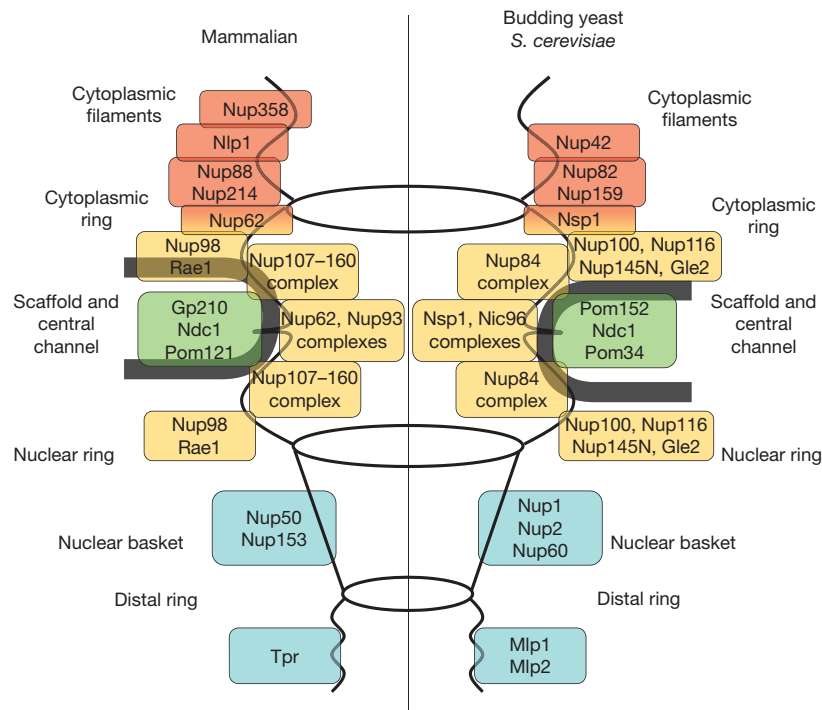


Figure 2 The nucleoporin proteins or Nups that make up the nuclear pore are highly conserved. A schematic of the nuclear pore with the cytoplasmic side at the top and the nuclear basket at the bottom is shown. The names of the Nups located in each region of the nuclear pore are indicated. The left of the schematic provides the names of budding yeast proteins and the corresponding vertebrate proteins are indicated on the right (see also Table 1).

Table 1 NPC subunit homologs based on similarities in sequence and function^{a,b,c}

Mammalian	Budding yeast (<i>S. cerevisiae</i>)	Location within NPC	Motifs/domains	Predicted function
Nup50/Npap60	Nup2	Nuclear	FG	Transport
<i>Nup62 complex</i> : Nup45, Nup54, Nup58, Nup62 ^d	<i>Nup1 complex</i> : Nsp1 ^d , Nup49, Nup57	Symmetric	FG, coiled-coil	Structure; transport
Nup88	Nup82	Cytoplasmic	β -Propeller, coiled-coil	Structure
<i>Nup93 complex</i> : Nup35/53, Nup93, Nup155, Nup188, Nup205	<i>Nic96 complex</i> : Nic96, Nup53, Nup59, Nup157, Nup170, Nup188, Nup192	Symmetric	β -Propeller, α -solenoid, FG	Structure; transport
Nup98	Nup100, Nup116, Nup145N	Symmetric	FG, Nup98 fold	Transport
<i>Nup107–160 complex</i> : ALADIN, Nup37, Nup43, Nup75/85, Nup96, Nup107, Nup133, Nup160, Sec13, Seh1	<i>Nup84 complex</i> : Nup84, Nup85, Nup120, Nup133, Nup145C, Sec13, Seh1	Symmetric	β -Propeller, α -solenoid	Scaffold
Nup153	Nup1, Nup2, Nup60	Nuclear	FG	Structure; transport
Nup214	Nup159	Cytoplasmic	β -Propeller, coiled-coil, FG	Structure
Nup358	–	Cytoplasmic	FG	Structure; transport
Gp210	Pom152	Transmembrane	TMH	Transport
Ndc1	Ndc1	Transmembrane	TMH	Structure
NLP1/CG1	Nup42	Cytoplasmic	FG	Structure; transport
Pom121	–	Transmembrane	TMH, FG	Structure
RAE1/Gle2	Gle2	Symmetric	β -Propeller	Transport
Tpr	Mlp1, Mlp2	Nuclear	Coiled-coil	Structure; transport
–	Pom34	Transmembrane	TMH	Structure

^aSome Nups share sequence similarity with multiple Nups.

^bA dash (–) indicates that no homolog has been identified.

^cFG, FG repeat domains; ND, not determined, TMH, transmembrane helix.

^dNsp1 (Nup62) is symmetrically located as part of the Nsp1 (Nup62) complex, and also localizes to the cytoplasmic face individually.

variations on this theme have been identified and, thus far, the best consensus sequence for an NES is LxxxLxxLxL; however, the leucines in this motif can be spaced differently and can be substituted with virtually any hydrophobic amino acid.

Transport Receptors

Targeting signals within cargo proteins are recognized by soluble receptors that direct those cargoes to the nuclear pore for transport. In general, these receptors form a family of structurally related molecules that are referred to as importins (for import receptors) and exportins (for export receptors) or as transport receptors or karyopherins. The family of related receptors is generally referred to as the importin β or karyopherin β family. One feature of these transport receptors is that they are modulated in such a way that they bind tightly to their cargo in the compartment where the cargo is picked up; however, following translocation through the NPC to the compartment where the cargo is delivered, the transport receptor undergoes a conformational change that leads to efficient release and delivery of the cargo. This switch in binding affinity is regulated by the small guanosine triphosphatase (GTPase), Ran.

Although most transport cargoes bind directly to their cognate transport receptor, there are examples where an adaptor protein mediates this interaction. The best example of this occurs in nuclear protein import of cargoes bearing a cNLS sequence. The cNLS within these cargoes is recognized by an adaptor protein called importin α or karyopherin α .

The importin/karyopherin α recognizes and binds to the cargo and enters the nucleus as a trimeric import complex consisting of cNLS cargo, importin/karyopherin α , and a conventional transport receptor termed importin/karyopherin β .

Once cargo is released from the transport receptors within the delivery compartment, all the receptors are recycled for subsequent rounds of transport. Import receptors are recycled back to the cytoplasm and export receptors are recycled back to the nucleus. This recycling of receptors ensures that a single receptor can mediate multiple rounds of transport.

The Ran GTPase Cycle

The small GTPase Ran serves as a molecular switch that modulates the directionality of nuclear transport. As with other GTPases, Ran can exist in two forms: bound to guanosine diphosphate (GDP) (RanGDP) or bound to GTP (RanGTP). It is the compartmentalization of these two distinct forms of Ran that regulates interaction between cargo and receptor and imparts directionality on nuclear transport. This compartmentalization is achieved through intracellular separation of the two important regulators of the Ran GTPase cycle. The Ran GTPase-activating protein (RanGAP), which enhances Ran-mediated GTP hydrolysis, is located in the cytoplasm and the Ran guanine nucleotide exchange factor (RanGEF), which facilitates exchange of GDP for GTP on Ran, is located in the nucleus. As RanGAP is in the cytoplasm, any RanGTP that enters the cytoplasm is rapidly converted to RanGDP.

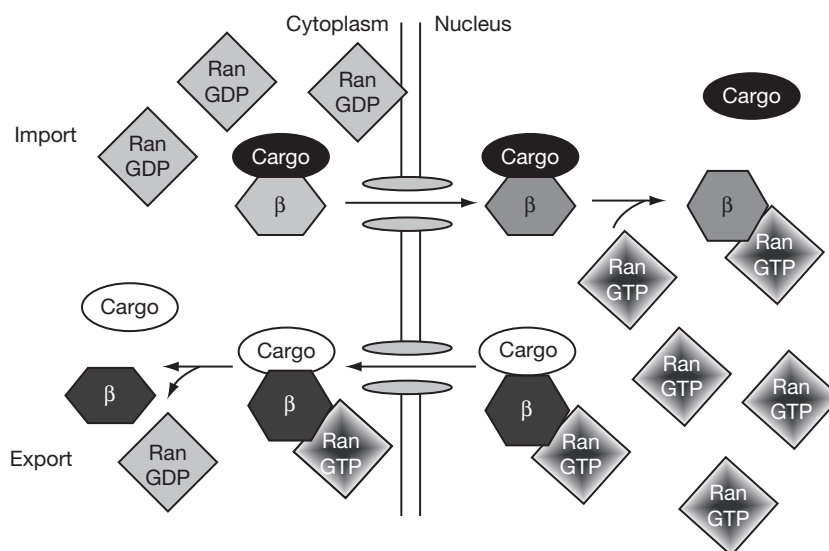


Figure 3 Ran regulates the assembly of import and export complexes. Top: for transport into the nucleus (import), cargoes are recognized in the cytoplasm by an importin/karyopherin β receptor to form an import complex. Once this complex is translocated into the nucleus, it encounters the nuclear RanGTP, which binds tightly to the importin/karyopherin β receptor and causes a conformational change that releases the import cargo. Bottom: for transport out of the nucleus (export), cargoes are recognized in the nucleus by an importin/karyopherin β receptor. However, in contrast to the import complex, the export complex can only form in the presence of RanGTP. Thus, export complexes are obligate trimeric complexes that consist of the export cargo, the importin/karyopherin β export receptor, and RanGTP. Once the export complex is translocated to the cytoplasm, it encounters the RanGAP. In the presence of RanGAP, the RanGTP within the complex is converted to RanGDP and the complex dissociates, resulting in release of the export cargo into the cytoplasm. Hence, by controlling complex assembly and disassembly, different forms of Ran confer directionality on the nuclear transport machinery.

Therefore, the level of RanGDP exceeds the level of RanGTP in the cytoplasm. By contrast, as RanGEF is in the nucleus, RanGDP is rapidly converted to RanGTP and the level of RanGTP is high compared to the level of RanGDP in the nucleus.

The different nucleotide-bound states of Ran regulate the flow of protein cargoes into and out of the nucleus by regulating assembly of import and export complexes (Figure 3). Cargoes to be transported into the nucleus bind to their cognate import receptors in the cytoplasm. These complexes are translocated into the nucleus, where levels of RanGTP are elevated. There, RanGTP binds to the import receptor causing a conformational change that releases the import cargo into the nucleus. By contrast, export cargoes bind to their export receptors only in the presence of RanGTP. Therefore, export complexes are obligate trimeric complexes consisting of the export receptor, the export cargo, and RanGTP. As with import, the export complex is translocated through NPCs to the cytoplasm. In the cytoplasm, the export complex encounters RanGAP and the bound RanGTP is converted to RanGDP. This conversion disassembles the export complex and leads to release of the export cargo into the cytoplasm.

Classical NLS-Mediated Protein Import

The best-understood nuclear transport process is the import of protein cargoes that contain a classical basic NLS (Figure 4). Thus, this process can be used most readily to illustrate the steps that occur when a transport cargo is moved into or out of the nucleus. Historically, nuclear protein import was divided into two steps, docking at the nuclear pore, an energy-independent step, and translocation into the nucleus, an energy-dependent step. Advances in our understanding of the process and the players

now lead us to define at least five distinct steps for import of a cargo that contains a classical NLS: (1) recognition and binding of the cNLS cargo to the importin/karyopherin α -importin/karyopherin β heterodimeric import receptor in the cytoplasm; (2) targeting to the nuclear pore through interactions between the importin/karyopherin β nuclear transport receptor and the nuclear pore; (3) translocation through the pore via transient interactions between importin/karyopherin β and the FG-repeat containing nucleoporins; (4) delivery into the nucleus where RanGTP binds to importin/karyopherin β to cause a conformational change that releases importin/karyopherin α and the cNLS cargo; and (5) recycling of importin/karyopherin α to the cytoplasm in a heterotrimeric complex with an export receptor and RanGTP and recycling of importin/karyopherin β , presumably in complex with RanGTP. Note that the only energy expenditure in this process occurs when the karyopherin proteins are recycled to the cytoplasm and the accompanying RanGTP is hydrolyzed.

Transport Mechanisms: RNA Trafficking

Multiple classes of RNAs, including messenger RNA (mRNA), transfer RNA (tRNA), micro-RNA (miRNA), ribosomal RNA (rRNA), and U small nuclear RNA (U snRNA), are transcribed within the nucleus and then transported to the cytoplasm. As compared with proteins, RNAs require extensive processing before they reach their mature form and are ready to exit the nucleus. This requirement means that RNA export is intimately linked to RNA processing within the nucleus.

The export of tRNA, U snRNA, miRNA, and rRNA follows pathways analogous to nuclear protein export. In each case, the

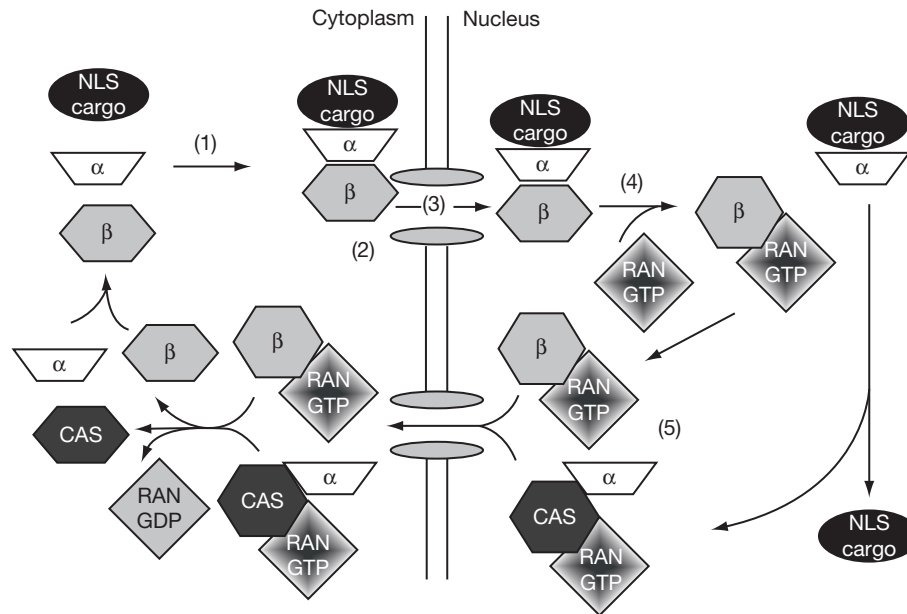


Figure 4 Model for classical NLS-mediated protein import into the nucleus. The best-understood transport mechanism is that of classical nuclear protein import, which is mediated by importin/karyopherin β and the cNLS-binding adapter, importin/karyopherin α . This process can be divided into at least five distinct steps as shown in the model: (1) recognition and binding of the cNLS cargo to the importin/karyopherin α -importin/karyopherin β heterodimeric import receptor in the cytoplasm; (2) targeting to the nuclear pore; (3) translocation through the pore; (4) delivery into the nucleus; and (5) recycling of importin/karyopherin α and importin/karyopherin β to the cytoplasm.

RNA is recognized either directly or indirectly by an importin/karyopherin β family nuclear transport receptor and exported to the cytoplasm in a Ran-dependent manner. mRNAs are exported to the cytoplasm as messenger ribonucleoprotein complexes through a mechanistically distinct pathway.

tRNA Export

tRNAs are synthesized in the nucleus by RNA polymerase III (Pol III) and undergo a number of maturation steps such as splicing, modification of the 5' and 3' termini, and aminoacylation prior to export to the cytoplasm where they are required for protein translation. Mature tRNAs are preferentially recognized in the nucleus by the importin/karyopherin β receptor, exportin-t/Los1/PAUSED, and are exported in complex with RanGTP. The NES signals in tRNAs require a specific 3' overhang and secondary and tertiary structures within the RNA to form, ensuring that only fully processed tRNAs are exported to the cytoplasm. Interestingly, Los1 and PAUSED are not essential in *Saccharomyces cerevisiae* or *Drosophila melanogaster*, respectively, suggesting that parallel pathways for tRNA export likely exist. Two additional importin/karyopherin β family members, Msn5, a receptor mainly involved in export of proteins that regulate transcription, and exportin-5, a receptor mainly involved in miRNA export, have been implicated in these alternative pathways.

miRNA Export

miRNAs are short, noncoding RNA sequences that are involved in posttranslational regulation of gene expression. They bind

complementary sequences in target mRNAs, resulting in translational silencing. The primary transcripts for miRNAs are transcribed by Pol II or Pol III and contain a stem-loop that is excised in the nucleus by an RNase called Drosha. This stem-loop, called pre-miRNA, is exported to the cytoplasm by the importin/karyopherin β , exportin-5, in complex with RanGTP. Following RanGTP hydrolysis, the freed pre-miRNA is cleaved by an RNase called Dicer to eventually form a single-stranded miRNA, which is incorporated into the RNA-induced silencing complex.

U snRNA Export

Many small nuclear RNAs, such as U snRNAs, are involved in mediating or regulating posttranslational RNA-processing events such as splicing. Most of these RNAs are transcribed in the nucleus by Pol II or Pol III as pre-snRNA, then capped on the 5' end and exported to the cytoplasm. Once in the cytoplasm, snRNAs associate with the protein components of small nuclear ribonucleoproteins and are reimported into the nucleus to perform their function. Export of U snRNA is mediated by the classical protein export receptor, Crm1/exportin-1, but requires an adaptor consisting of the cap-binding complex and the NES-containing protein, PHAX.

rRNA Export

Eukaryotic ribosomes, which translate mRNAs into proteins, are comprised of a large (60S) subunit and a small (40S) subunit. These subunits are generated in the nucleolus and each contains a number of rRNAs and up to ~50 proteins.

The 60S subunit is exported from the nucleus by Crm1/exportin-1 in a RanGTP-dependent manner. The leucine-rich NES recognized by Crm1/exportin-1 during the 60S subunit export is provided by the adaptor Nmd3, a *trans*-acting factor that associates with the 60S subunit during biogenesis. Due to its size, it is likely that other transport factors, such as Rrp12 or Mtr2 (a protein involved in mRNA export), are also involved in the export of the 60S subunit. Export of the 40S subunit is less well understood, but Crm1/exportin-1 and the Ran cycle are required. Several subunit components and assembly factors have been suggested to provide the NES for 40S subunit export, but no definitive mediator or adaptor has been identified.

mRNA Export

mRNAs are transcribed by Pol II in the nucleus, capped at the 5' end, spliced, and polyadenylated at the 3' end prior to export to the cytoplasm for translation. During these modification steps, various processing factors are recruited to the pre-mRNA and many of these proteins remain bound to the transcript, providing signposts that proper processing has occurred. Many of these heterogeneous nuclear ribonucleoproteins or RNA-binding proteins shuttle between the cytoplasm and the nucleus and escort the poly(A) + mRNA as it is exported through the NPC. Therefore, mRNAs are not exported to the cytoplasm as naked nucleic acids, but rather as RNA–protein complexes. It is generally agreed that mRNA export machinery recognizes signals within the proteins of these complexes rather than the mRNA itself.

Several export receptors have been implicated in the export of at least a subset of mRNA transcripts. The heterodimeric Mex67-Mtr2 (TAP-p15 in humans) complex can bind directly to mRNA, but relies on recruited mRNA-binding adaptor proteins such as Yra1/Aly/REF or members of the TREX complex to mediate interaction with mRNA. Alternatively, a small subset of cellular mRNA transcripts containing AUI-rich elements, which recruit NES-containing adapter proteins, are exported by the classical NES receptor, Crm1/exportin-1. Export of

intron-containing human immunodeficiency virus (HIV) transcripts from the nucleus is also dependent on Crm1/exportin-1 and is mediated by the HIV protein Rev.

See also: **Molecular Biology:** Messenger RNA Degradation in Bacteria; mRNA Polyadenylation in Eukaryotes; **Signaling:** Mitogen-Activated Protein Kinase Family; Neurotransmitter Transporters; Ran GTPase.

Further Reading

- Cook A, Bono F, Jinek M, and Conti E (2007) Structural biology of nucleocytoplasmic transport. *Annual Review of Biochemistry* 76: 647–671.
- Damelin M, Silver PA, and Corbett AH (2002) Nuclear protein transport. *Methods in Enzymology* 351: 587–607.
- Dasso M (2002) The Ran GTPase: Theme and variations. *Current Biology* 12: R502–R508.
- Goldfarb DS, Corbett AH, Mason DA, Harreman MT, and Adam SA (2004) Importin alpha: A multipurpose nuclear-transport receptor. *Trends in Cell Biology* 14: 505–514.
- Gorlich D and Kutay U (1999) Transport between the cell nucleus and the cytoplasm. *Annual Review of Cell and Developmental Biology* 15: 607–660.
- Kohler A and Hurt E (2007) Exporting RNA from the nucleus to the cytoplasm. *Nature Reviews Molecular Cell Biology* 8: 761–773.
- Lange A, Mills RE, Lange CJ, et al. (2007) Classical nuclear localization signals: Definition, function, and interaction with importin alpha. *Journal of Biological Chemistry* 282: 5101–5105.
- Lee BJ, Cansizoglu AE, Suel KE, et al. (2006) Rules for nuclear localization sequence recognition by karyopherin beta2. *Cell* 126: 543–558.
- Lei EP and Silver PA (2002) Protein and RNA export from the nucleus. *Developmental Cell* 2: 261–272.
- Lim RY, Aebi U, and Fahrenkrog B (2008) Towards reconciling structure and function in the nuclear pore complex. *Histochemistry and Cell Biology* 129: 105–116.
- Mosammaparast N and Pemberton LF (2004) Karyopherins: From nuclear-transport mediators to nuclear function regulators. *Trends in Cell Biology* 14: 547–556.
- Suntharalingam M and Wente SR (2003) Peering through the pore: Nuclear pore complex structure, assembly, and function. *Developmental Cell* 4: 775–789.
- Weis K (2002) Nucleocytoplasmic transport: Cargo trafficking across the border. *Current Opinion in Cell Biology* 14: 328–335.

Nucleolus, Overview

T Pederson, University of Massachusetts Medical School, Worcester, MA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Nucleolus organizer A genetic locus containing genes for ribosomal RNA.

Perinucleolar compartment A discoid, cap-like structure situated on the surface of nucleoli in most malignant cells.

Proteomics An emergent field of protein sequence determination in complex mixtures, based on a

high-resolution analytical method in which the protein components of a subcellular complex or organelle, first separated by electrophoresis or liquid chromatography, are proteolytically cleaved (usually by trypsin) into fragments that are then resolved by matrix-assisted, laser-desorption ionization time-of-flight mass spectrometry.

Background

The nucleolus was one of the earliest intracellular structures (and the first within the nucleus) to be observed by microscopy, around 1860. Throughout the early twentieth century, cytologists were struck by the consistent presence of nucleoli inside the nuclei of virtually all cells examined, although the number and shape of nucleoli were quite variable (Figure 1). In the 1930s, it was discovered that the formation of the nucleolus is determined by a distinct genetic locus. This genetic element, termed the nucleolus organizer, was found in the 1960s to consist of repeated genes for ribosomal RNA (rRNA). It is an interesting historical fact that the rRNA genes were not only the first ones from eukaryotes to be observed at high resolution in the electron microscope, but also the first genes to be physically isolated, well before the advent of recombinant DNA techniques.

Although a considerable amount of cytochemical and ultrastructural research had been done on nucleoli up through the 1950s, a major advance occurred when they were first isolated by cell fractionation in the mid-1960s. This opened the door to the investigation of their protein and RNA components, and accelerated the study of rRNA synthesis and processing.

Structure

Although the ultrastructural (electron microscopic) appearance of nucleoli is highly variable, a tripartite organization is generally observed. Fibrillar centers (FCs) are one or more rather small foci within a nucleolus that have a distinctive appearance. Each of these FCs is surrounded (in three dimensions) by a typically much larger zone, called the dense fibrillar component (DFC). The remainder of the nucleolus, or at least most of the remainder, displays a distinctly more granular appearance, and is, thus, called the granular component (GC).

Functional Organization

The tripartite ultrastructural landscape of the nucleolus has been functionally defined (Figure 2). The FCs are the zones where the active rRNA genes themselves are located. The nascent transcripts

of pre-rRNA extend out into the juxtaposed DFC, which also contains intermediate RNA species in the rRNA-processing pathway. Also present in the DFC are numerous (100–200 or more) small RNA species (and their associated proteins) that serve as guide RNAs for the site-specific modification (2'-O-ribose methylation and pseudouridine formation) of many internal nucleotides in pre-rRNA. The GC consists of more mature, assembled ribosomal particles. It is not yet entirely clear whether certain stages of ribosome maturation are sharply restricted to the DFC or the GC or instead have some degree of overlap between them.

Amplified Nucleoli and Missing Nucleoli

One of the most dramatic aspects of the biology of the nucleolus is its selective amplification in the oocytes of certain organisms. In these cells, the rRNA genes are amplified into a large number (typically thousands) of extrachromosomal copies, which form supernumerary nucleoli. Through their activity, the growing oocyte builds up a large reservoir of maternal ribosomes to sustain early embryonic development. Amplified nucleoli are found in the oocytes of several phyletic groups, but have been most extensively studied in urodele and anuran Amphibia (salamanders and frogs).

One of the ways the function of the nucleolus in ribosome synthesis was first revealed was by an elegant experiment that involved a frog mutant in which one chromosomal set of rRNA genes is missing. Half the oocytes of females that are heterozygous for this mutation have no rRNA genes whatsoever and, consequently, cannot form amplified nucleoli, nor can anucleolate embryos descended from them synthesize new ribosomes. Accordingly, these embryos die at a time during development when the maternal supply of ribosomes becomes limited. This experiment is one of the classics in the field of eukaryotic gene expression.

Clinical Connections

Two aspects of the nucleolus are significant clinically. For many years, pathologists have used nucleolar morphology

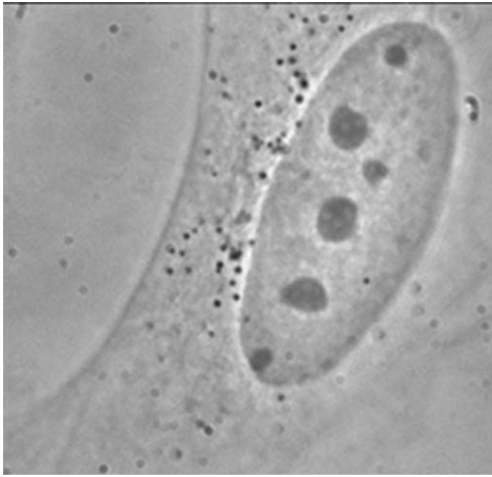


Figure 1 Phase-contrast micrograph of a portion of a living HeLa (human cervical carcinoma) cell. The dark, spherical objects within the nucleus are the nucleoli. This micrograph was taken by Laura Lewandowski in the author's laboratory. Magnification approximately $\times 3000$.

and number as an important guide in the diagnostic classification of certain malignancies. In addition, a specialized structure called the perinucleolar compartment (PNC) sits as a cap-like disk on the surface of the nucleoli in most malignant cells (both tumor cell lines and actual patient biopsies). The PNC seems to be involved in RNA metabolism, and contains a set of small RNAs that are transcribed by RNA polymerase III. However, its possible causative relationship to malignancy is not known.

The Plurifunctional Nucleolus

Since the 1990s, a number of clues have begun to suggest that the nucleolus may have functions beyond ribosome synthesis. One line of evidence consists of the observation in nucleoli of various molecules that have no known connection with ribosome biosynthesis. For example, several proteins related to mitogenic cell signaling and cell-cycle progression have been found in nucleoli, although these various observations have not yet been drawn together into a coherent framework. In budding yeast, the nucleolus of interphase cells contains a complex that is released and promotes a late stage of mitosis after the nucleolus breaks down, but whether this complex has some role within the interphase nucleolus as well is not known. The fact that so many of the various ribosome-unrelated molecules and complexes that are increasingly being observed in the nucleolus are related to cell growth is quite intriguing, and this is an important area for future work.

It has also become clear that the life cycle of certain viruses involves the nucleolus in some way. For example, in the replication of the human immunodeficiency virus, certain forms of viral messenger RNAs are associated with the nucleolus during a particular stage in the infectious cycle.

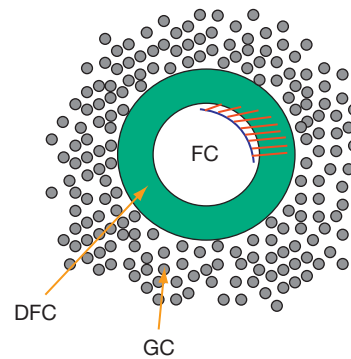


Figure 2 A schematic diagram illustrating the relationships of the three major components of the nucleolus. An FC contains the active rRNA genes (blue) with the nascent transcripts of pre-rRNA (red) extending into the surrounding DFC, embedded in the GC. See text for further details. Reproduced from Huang S (2002) Building an efficient factory: Where is pre-rRNA synthesized in the nucleolus? *Journal of Cell Biology* 157: 739–741, with permission from Rockefeller University Press.

However, the functional significance of this and other observations of viral connections with the nucleolus are not presently known.

A third nonribosomal function of the nucleolus is its likely participation in the biosynthesis of the signal-recognition particle (SRP). The SRP is an RNA–protein complex that ensures the topologically correct synthesis of membrane and secreted proteins. Both the RNA component of the SRP and some of the SRP's protein components are present in the nucleolus in yeast, amphibian oocytes (i.e., in the amplified nucleoli), and mammalian cells, suggesting that this translation-related RNA–protein complex originates in the nucleolus, as do the ribosomes themselves.

Current Trends and Conclusion

At the present time, studies on the nucleolus are being conducted by nucleolar specialists on the one hand, and by an increasing number of cell biologists on the other, who are encountering a nucleolar connection in their work. A nucleolar protein called nucleostemin has been implicated in the biology of embryonic stem cells, and a reorganization of the nucleolus in donor somatic cell nuclei appears to be a key factor in the cloning of mammalian animals. Meanwhile, on the analytical front, proteomics studies of purified human nucleoli have identified ~ 350 proteins that are highly enriched in the isolated fraction (Figure 3). Many of these proteins have no obvious relationship to ribosome biosynthesis. One interesting possibility is that perhaps all of these apparently nonribosome-related nucleolar components are indeed related to ribosome biosynthesis after all, not in the constitutive pathway of ribosome synthesis itself (the classical model), but as regulatory components of cellular growth control circuits that intersect in the nucleolus and, thus, with the biogenesis of the protein synthesis machinery (a neoclassical model).

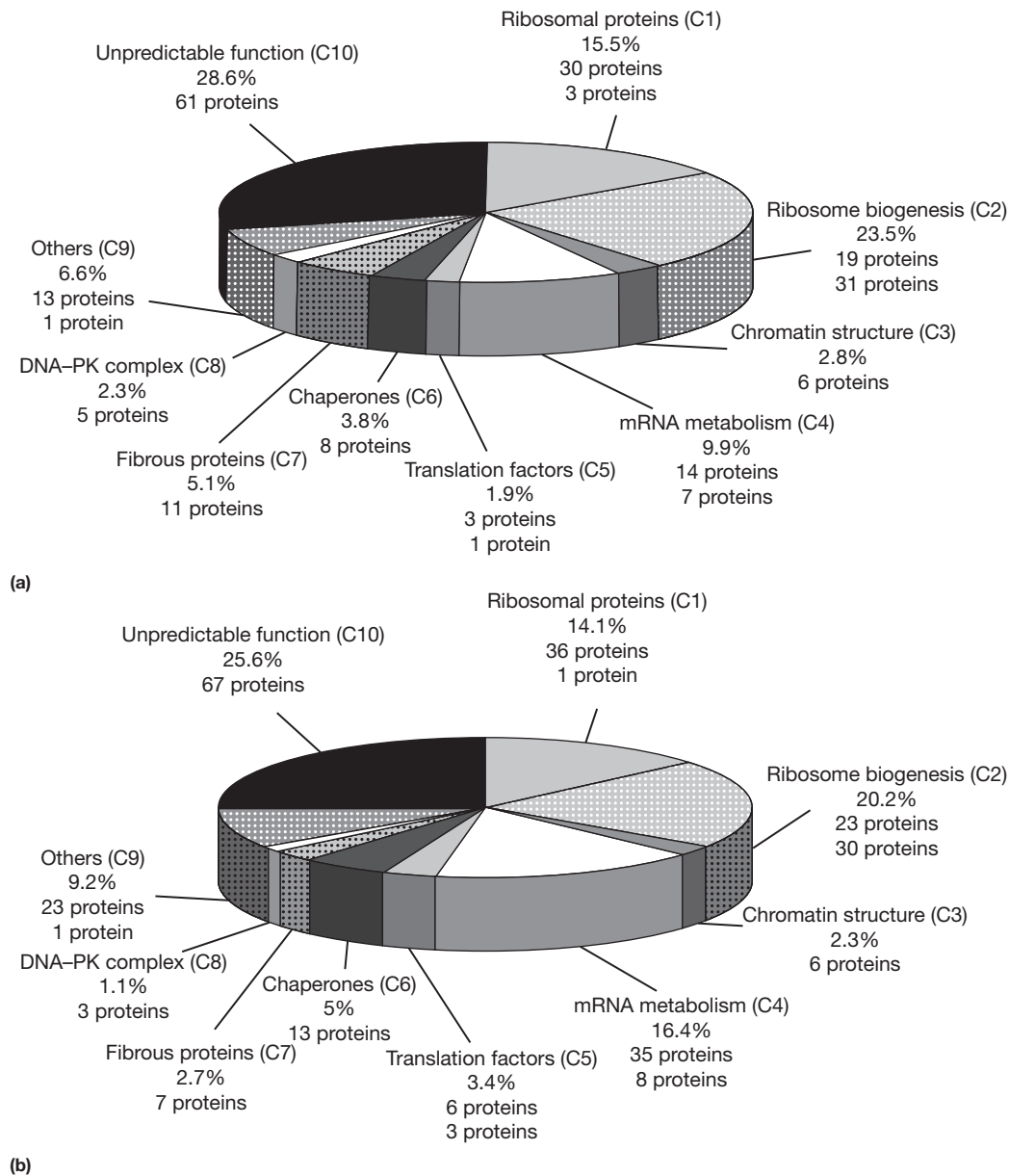


Figure 3 Relative occurrence of various classes of proteins in nucleoli purified from HeLa cells, as determined by proteomics. The two panels present the results from two independent studies: (a) The results from Scherl A, Coute Y, Deon C, et al. (2002) Functional proteomic analysis of human nucleolus. *Molecular Biology of the Cell* 13: 4100–4109 and (b) shows the results from Andersen JS, Lyon CE, Fox AH, et al. (2002) Directed proteomic analysis of the human nucleolus. *Current Biology* 12: 1–12. The figure is reproduced from Scherl A, Coute Y, Deon C, et al. (2002) Functional proteomic analysis of human nucleolus. *Molecular Biology of the Cell* 13: 4100–4109. For further details, see also Pederson T (2002). Proteomics of the nucleolus: More proteins, more functions? *Trends in Biochemical Sciences* 27: 111–112.

Further Reading

Andersen JS, Lyon CE, Fox AH, et al. (2002) Directed proteomic analysis of the human nucleolus. *Current Biology* 12: 1–12.

Brown DD and Gurdon JB (1964) Absence of rRNA synthesis in the anucleolate mutant of *X. laevis*. *Proceedings of the National Academy of Sciences of the United States of America* 51: 139–146.

Hernandez-Verdun D, Roussel P, and Gebrane-Younes J (2002) Emerging concepts of nucleolar assembly. *Journal of Cell Science* 115: 2265–2270.

Huang S (2002) Building an efficient factory: Where is pre-rRNA synthesized in the nucleolus? *Journal of Cell Biology* 157: 739–741.

Olson MO, Dundr M, and Szébeni A (2000) The nucleolus: An old factory with unexpected capabilities. *Trends in Biochemical Sciences* 10: 189–196.

- Pederson T (1998) The plurifunctional nucleolus. *Nucleic Acids Research* 17: 3871–3876.
- Pederson T (2002) Proteomics of the nucleolus: More proteins, more functions? *Trends in Biochemical Sciences* 27: 111–112.
- Pederson T (2009) The nucleolus. In: Spector DL and Misteli T (eds.) *The Nucleus*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Pederson T and Politz JC (2002) The nucleolus and the four ribonucleoproteins of translation. *Journal of Cell Biology* 148: 1091–1095.
- Pederson T and Tsai RY (2009) In search of nonribosomal nucleolar protein function and regulation. *Journal of Cell Biology* 184: 771–776.
- Scherl A, Coute Y, Deon C, et al. (2002) Functional proteomic analysis of human nucleolus. *Molecular Biology of the Cell* 13: 4100–4109.

Nucleotide Excision Repair, Bacterial: The UvrABCD System

B Van Houten, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; Hillman Cancer Center, Pittsburgh, PA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 134–142, © 2004, Elsevier Inc.

Glossary

DNA polymerase A protein or protein complex which synthesizes a daughter strand of DNA using a parental strand as a template.

Helicase A protein, which actively separates the two strands of DNA using the energy of ATP hydrolysis.

Nucleotide excision repair (NER) A highly conserved multistep process in which several protein machines identify and remove bulky damage from DNA using a dual incision mechanism.

SOS response A number of DNA metabolism genes controlled by the LexA repressor, which is cleaved during genotoxic stress.

uvr genes Genes encoding subunits of NER proteins, which confer sensitivity to killing by ultraviolet light.

Zinc finger A DNA-binding motif consisting of a molecule of zinc bound to four cysteines found in many DNA-repair proteins.

This article is dedicated to Lawrence (Larry) Grossman, PhD, 1924–2006 who was the University Distinguished Service Professor of Biochemistry in the Department of Biochemistry and Molecular Biology in the Johns Hopkins Bloomberg School of Public Health. He was an inspiration, friend, and a 'worthy opponent'. DNA repair is intrinsic to the evolution of life, and organisms have evolved a large number of complex systems to ensure the integrity of their hereditary material. DNA constantly undergoes chemical modification as a result of its intrinsic chemical instability, as well as endogenous and environmental stresses. The most versatile DNA repair mechanism is nucleotide excision repair (NER), first discovered in *Escherichia coli* in 1964 independently by Setlow and Carrier and Boyce and Howard-Flanders. This bacterial system, like its eukaryotic counterpart, has a very broad spectrum of damage recognition and operates according to the same basic principles in all organisms:

1. repair proteins form macromolecular assemblies in order to locate and activate damage utilizing a number of adenosine triphosphate (ATP)-hydrolysis-driven nucleoprotein rearrangements;
2. the DNA is incised on both sides of damage; and
3. the incised fragment is displaced with concomitant resynthesis followed by ligation.

NER is coupled to other cellular processes, notably, transcription. The sequencing of the complete genomes of over 100 different bacterial species indicates that the NER system is a highly conserved process.

UvrABC Endonuclease

The NER pathway consists of five basic steps: damage recognition, incision, excision, repair synthesis, and ligation. In most bacteria, the first steps of this process are carried out by an ensemble of three proteins encoded by the *uvrA*, *uvrB*, and *uvrC* genes in an ATP-dependent series of reactions. During the

initial characterization of these proteins, the name of this system was shortened to 'UvrABC endonuclease', which makes both the 5' and 3' incisions to a lesion in a 'dual incision' reaction (Figure 1). The UvrABC endonuclease possesses a broad spectrum of substrate specificity and is capable of acting on a wide variety of unrelated DNA damage (Table 1).

Genetics and Evolutionary Conservation of uvr Genes

It was photobiologists who discovered NER from analyses of ultraviolet (UV) radiation survival curves in some bacterial, most notably *E. coli*. Rather than observing a first-order decline in survival of cells as a function of UV dose, there was a shoulder in the survival curves at lower doses. Some *E. coli* mutants, which lacked this shoulder, were sensitive to low doses of UV and were isolated. All these mutants map at one of three loci on the *E. coli* chromosome designated *uvrA* (92 min), *uvrB* (17 min), and *uvrC* (41.5 min), and, are hence, unlinked genes. In species of *Bacillus*, the *uvrA* and *uvrB* genes are within 50 bp of each other on the chromosome.

The *uvr* genes were first cloned, sequenced, and amplified from *E. coli*, which led to a greater understanding of the roles of three proteins in the process of damage recognition and incision. Subsequently, large-scale sequencing efforts have revealed that *uvr* genes are highly conserved among the bacteria species.

uvrA

uvrA is one of a series of genes collectively referred to as SOS genes, which are induced to increased levels of transcription by agents that cause DNA damage. A specific binding site for the LexA repressor protein specific for the operator promoter region of the *uvrA* gene (the so-called LexA box or SOS box) has been identified. Sequence analyses reveal that the *uvrA* gene (2.82 kbp) translates into 940 amino acids ($M_r = 103\,874$ Da) and contains two ATP-binding motifs, two zinc-finger motifs, and a helix–turn–helix motif.

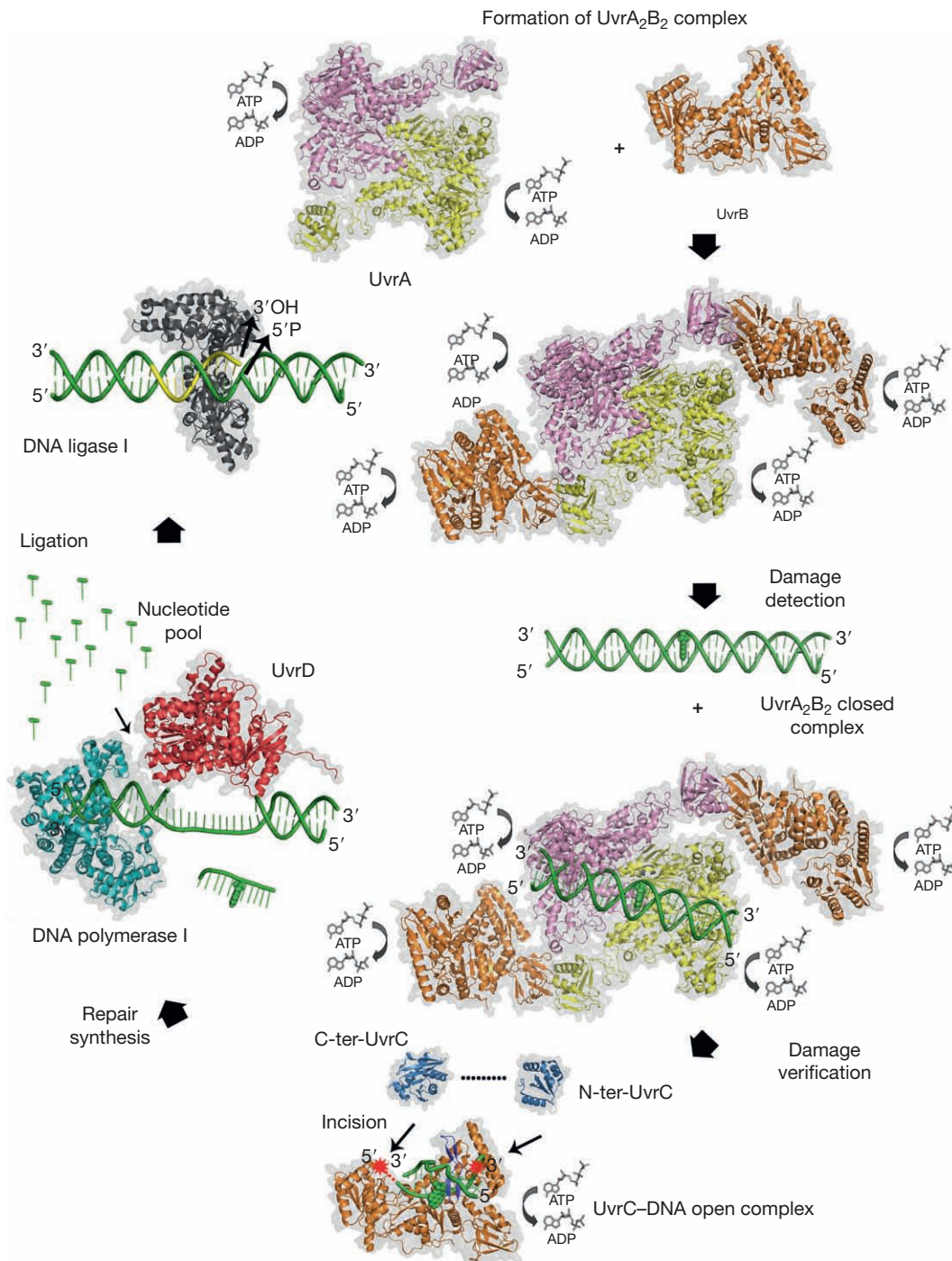


Figure 1 Nucleotide excision repair pathway in bacteria. Dimeric UvrA protein (light green and pink) (PDB ID:2R6F) hydrolyzes both ATP and GTP, and forms a complex with UvrB, orange (PDB ID: 2FDC). During damage detection, the UvrA₂B₂ complex (PDB ID for the contact interface:3FPN) first searches for damage-induced distortions in the DNA helix which are recognized by UvrA. DNA containing a benzo[a] pyrene diol epoxide-N₂-guanine adduct is shown in green. This activates the ATPase activity of UvrB and UvrA transfers the damaged DNA to UvrB. During damage verification, the β -hairpin of UvrB (shown in purple) inserts between the two strands of DNA, clamping the damaged strand forming a stable pre-incision complex. Binding and hydrolysis of ATP by UvrB are essential for recruitment of UvrC. The N-terminal endonuclease domain of UvrC (PDB ID:1YCZ) initiates the cut 4–5 nucleotides 3' to the damaged site followed by the 5' cut by C-terminal endonuclease domain of UvrC (PDB ID:2NRR) 8 nucleotides away from the lesion (both domains are shown in dark blue). Working together UvrD (red) (PDB ID:2IS1) unwinds the DNA and releases the oligonucleotide containing the lesion, while DNA polymerase I (turquoise) (PDB ID:2HHQ) synthesizes the missing strand (shown in yellow). Finally, DNA ligase I, dark gray (PDB ID:1DGS), seals the nick and completes the repair patch. All the protein structures are displayed using the α -carbon trace showing β -sheets and α -helices, with the van der Waal surface shown in light gray. The UvrA₂B₂ complex was created from three individual structures of UvrB (PDB ID:2FDC), UvrA (PDB ID:2R6F), and the contact interface of UvrA and UvrB (PDB ID:3FPN). The UvrB–DNA complex was modified from the UvrB–DNA co-crystal structure (PDB ID:2FDC).

Table 1 Range of substrates of the bacterial NER UvrABC system

Type	Lesion	UvrABC	ΔM_r	Properties
Single base modification	Thymine glycol	++	+34	
	Dihydrothymine	0	+2	
	Benzo[a]pyrene adduct	+++	+171	T_M lowered
	Anthramycin adduct	+++	+191	T_M elevated
	Cross-linked triple strand	+ [72]	~+3000	
	O^4 -alkyl thymine	+	+14	
	O^6 -methyl guanine	+	+14	
	N^6 -methyl adenine	0	+14	
	Psoralen adduct	+++	+185	Positive kink
	Base removed (AP site)	+	~-130	
Cross-links, intra-strand	<i>cis</i> -Pt adduct	+++	+227	Negative kink
	Pyrimidine dimer	++	0	
	6-4 photoproduct	+++	0	
	<i>cis</i> -Pt adduct	++	+227	
Cross-links, inter-strand	Nitrogen mustard adduct	+	+69	
	Psoralen bisadduct	+++	+185	Unwound
	dsDNA	0	0	
Natural bases	A-tracts	0	0	Bent
	Mismatches, loops	0 ^a	0	
	Caffeine complex	—	+194	Intercalator
Noncovalent modifications	Ditercalinium complex	++ ^b	+500	Bisintercalator

^aSome mismatches are recognized with very low efficiency (<1%).

^bFutile cycle of spurious repair and subsequent formation of the complex at another site.

uvrB

The *uvrB* gene is also a member of the SOS regulon and is inducible by DNA damage. The gene is transcribed from two overlapping promoters called P1 and P2. A LexA protein-binding site is present in the P2 promoter region. Transcription from P2 is inhibited by LexA repressor protein, while that from P1 is unaffected. Sequence analysis revealed that the *uvrB* gene is 2019 bp, which translates into 673 amino acids (76.6 kDa). The *uvrB* gene product has a consensus Walker-type A nucleotide-binding motif, helicase motifs II–VI as well as coiled-coil consensus sequences.

uvrC

In contrast to *uvrA* and *uvrB*, *uvrC* is not inducible by DNA damage and is not a member of the SOS regulon, although its cellular location is altered during SOS. The sequence analysis shows that the size of *uvrC* is 1830 bp encoding a polypeptide of 610 amino acids (66 kDa).

Cho

It was quite surprising that a new UvrC homolog, Cho, could be discovered in an organism for which the complete genome sequence was known since 1997. The gene encoding Cho was found to be UV-inducible by Hanawalt and Brown using a high-density DNA microarray. Careful inspection of this gene revealed homology to the T_{ev}-homing endonucleases and also the N terminus of UvrC. Goosen and co-workers cloned and overexpressed the Cho protein, and found that the Cho protein will incise a DNA adduct at the tenth phosphodiester bond 3' to the damaged site. This cutting at a distance was shown to

help incise at bulky lesions, which might otherwise block the normal UvrC 3' incision. They also found that UvrC can complete the incision process by efficient cleavage 5' to the 3' incised adduct. Cho is rather rare in the bacterial kingdom (Figure 2).

Properties of UvrA, UvrB, and UvrC Proteins

UvrA

The UvrA protein is a DNA-independent ATPase that binds DNA and is a member of the ABC transporter superfamily characterized by dual ATP-binding sites. Soon after UvrA was sequenced, it was noticed that the gene probably arose through multiple duplication events in which a four-cysteine zinc finger was placed between the Walker type A and B nucleotide-binding motifs. Site-directed mutagenesis studies show directly that UvrA has an ATPase activity located in a region centered at lysine residues of two Walker type A nucleotide-binding motifs. UvrA also hydrolyzes guanosine triphosphate (GTP). ATPase activity of UvrA is modulated by DNA. High concentrations of UvrA protein and the binding energy of ATP (or poorly hydrolyzable ATP- γ -S) favor dimerization of the protein. UvrA dimers are DNA-binding species. The binding of UvrA to undamaged DNA is 10^3 – 10^4 -fold weaker than to the damaged one. The specificity for damaged DNA is abolished by ATP- γ -S, while it enhances nonspecific binding. The dissociation rate for UvrA from damaged sites is fast in the presence of ATP. ATP hydrolysis increases the specificity of binding to the damaged DNA, but lowers the equilibrium-binding constant by stimulating dissociation. The crystal structure of a UvrA dimer by Verdine and co-workers indicated that each UvrA

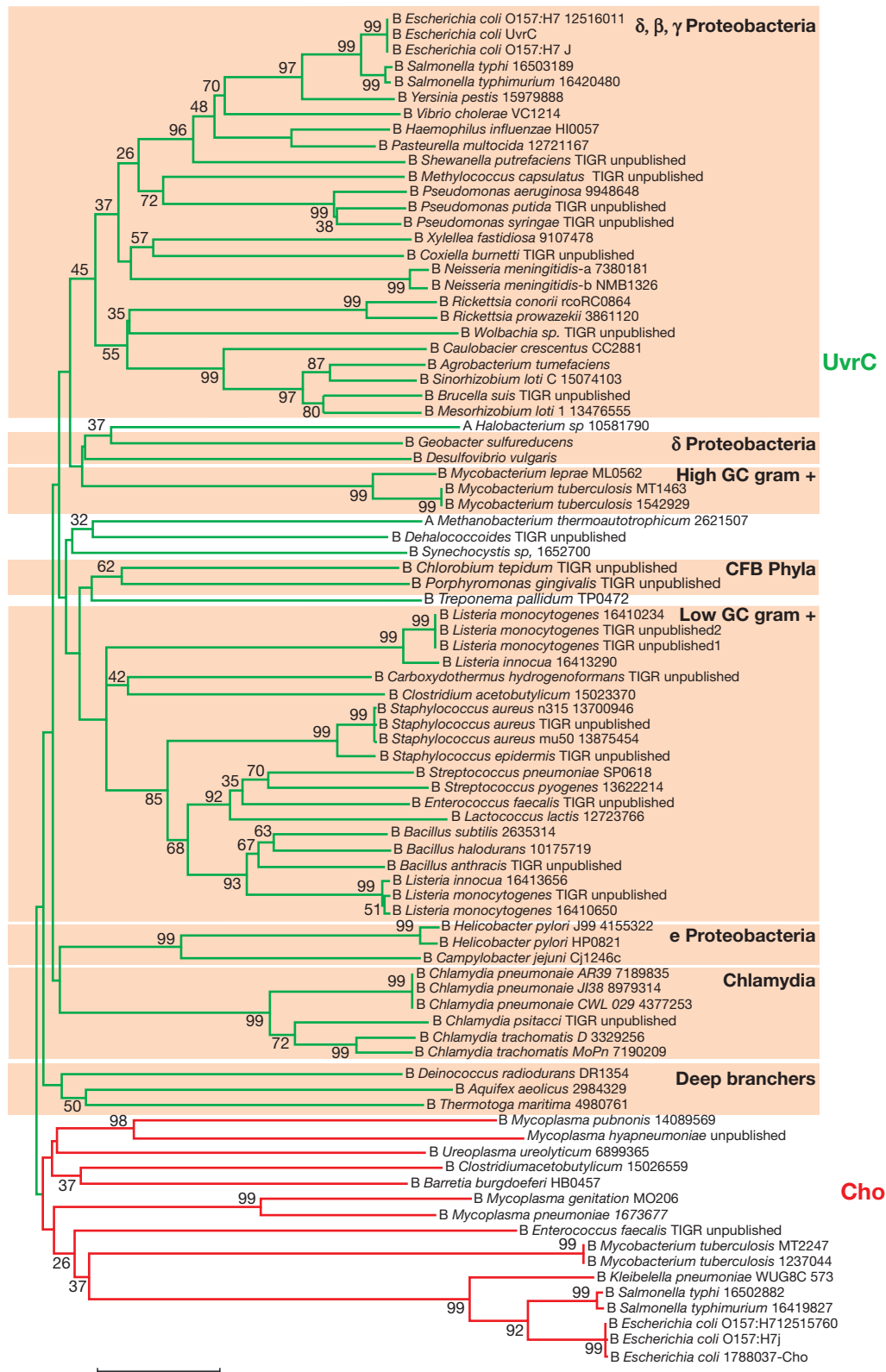


Figure 2 Phylogenetic tree of UvrC and Cho in various bacterial species. Phylogenetic tree of UvrC and Cho homologs. Homologs were identified in complete or nearly complete genomes by blastp or blastx searches. Protein sequences were aligned using clustalx. Bootstrap values, a measure of statistical support for particular groupings of genes to the right of the number, are shown on the tree when >40%. Reproduced from Van Houten B, Eisen JA, and Hanawalt PC (2002) A cut above: Discovery of an alternative excision repair pathway in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 99(5): 2581–2583.

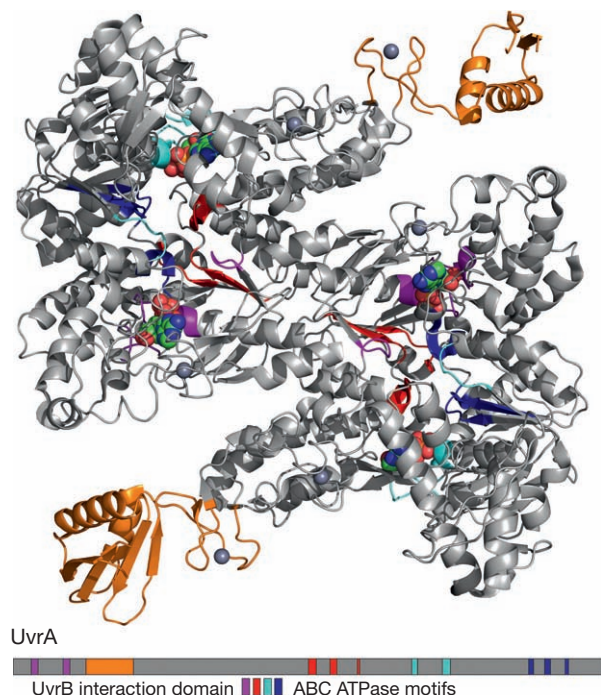


Figure 3 Crystal structure of UvrA. The UvrA dimer structure (PDB ID 2R6F): the UvrB-binding domain is shown in orange. For the first nucleotide-binding domain (NBDI), the Walker A and Q loop are shown in magenta; the Walker B, H loop, and signature sequence are shown in red. For the NBDII, the Walker A and Q loop are shown in cyan, the Walker B, H loop, and signature sequence are shown in blue. The Zn ions and ADP molecules are shown as CPK models.

monomer contains three zinc atoms per molecule, at least one of which is required for DNA binding (Figure 3). Van Houten and colleagues have shown that the C-terminal zinc-finger of UvrA is absolutely essential for damage recognition, but not DNA binding. The ATP-binding sites are created by intramolecular interaction of Walker A from one motif with the other Walker B motif.

UvrB

The purified protein has no detectable ATPase activity, although it binds ATP with a $K_d \sim 1$ mM. However, ATPase activity is observed when UvrB interacts with UvrA in the presence of DNA. This dramatic appearance is due to activation of the cryptic ATPase when UvrB forms an UvrAB–DNA complex. When UvrB is proteolyzed in a specific region of the C terminus, the resulting ~ 70 -kDa protein is referred to as UvrB*. The ‘cryptic UvrB ATPase’ functional in UvrB* is activated by single-stranded DNA or chaotropic salts. UvrB can bind to double-stranded DNA only in the presence of UvrA, and binds to single-stranded DNA with a K_d in the micromolar range. The amino acid sequence of the UvrB protein shows homology with two limited stretches of the sequence of the UvrC protein. Three groups have independently solved the crystal structure of UvrB. The structure has revealed that UvrB is folded into four separate domains (Figure 4). Domain 1a includes four helicase motifs: Ia, Ib, II, and III that are required for ATPase activity.

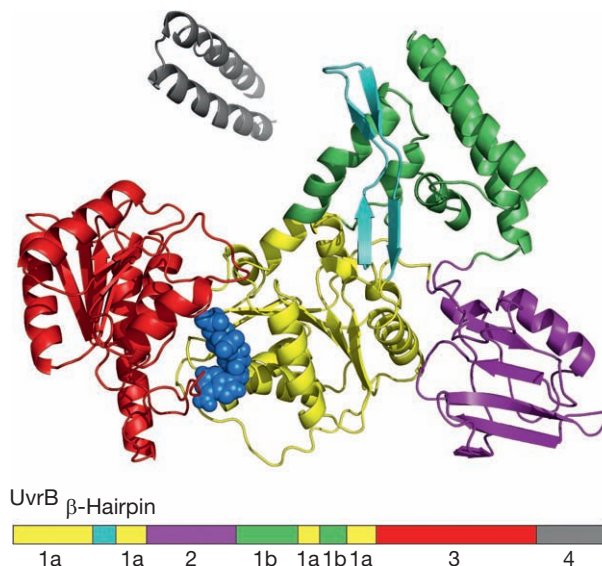


Figure 4 Structure of UvrB. The UvrB structure (PDB ID 2FDC): domain 1a is shown in yellow, 1b in green. Domain 2 is labeled as UvrA interacting domain and shown in magenta. Domain 3 is shown in red, and a separate domain 4 peptide structure is shown in gray. The β -hairpin motif in domain 2 is shown in cyan. ATP is shown in dark blue. UvrB belongs to the helicase superfamily II with seven helicase motifs in domain 1a and domain 3. Below the structure, the sequence of each protein is indicated and color coded as in its respective structure.

Domain 2 is important for UvrA interaction and domain 3 contains helicase motifs, IV, V, and VI. Sandwiched between domains 1a and 1b is an important β -hairpin motif which is believed to provide important DNA contacts. Further details of the UvrB structure and how it provides function are discussed below.

UvrC

UvrC is a single-stranded DNA-binding protein and binds to it with a relatively high affinity. UvrC is a catalytic subunit of the UvrABC endonuclease. Mutations in the C-terminal region of UvrC are extremely sensitive to UV and defective in 5'-incision activity. The 314 C-terminal amino acids of UvrC are sufficient to support dual incision of damaged DNA by the UvrABC endonuclease. UvrC protein associates with a UvrB–DNA complex at damaged site. Recently, the C-terminal part of UvrC was solved by multidimensional nuclear magnetic resonance and shown to consist of two helix–hairpin–helix (HhH) motifs as predicted previously. Kaptein and co-workers found that this C-terminal region of the protein bound avidly to double-strand/single-strand junctions like those found in bubble substrates containing more than six unpaired bases. Interestingly, this is exactly the number of base pairs believed to be unwound during the formation of the UvrB–DNA complex. Most recently, Kisker and colleagues have determined the crystal structure of the N-terminal domain responsible for the 3' incision and the C-terminal domain responsible for the 5' incision (Figure 5).

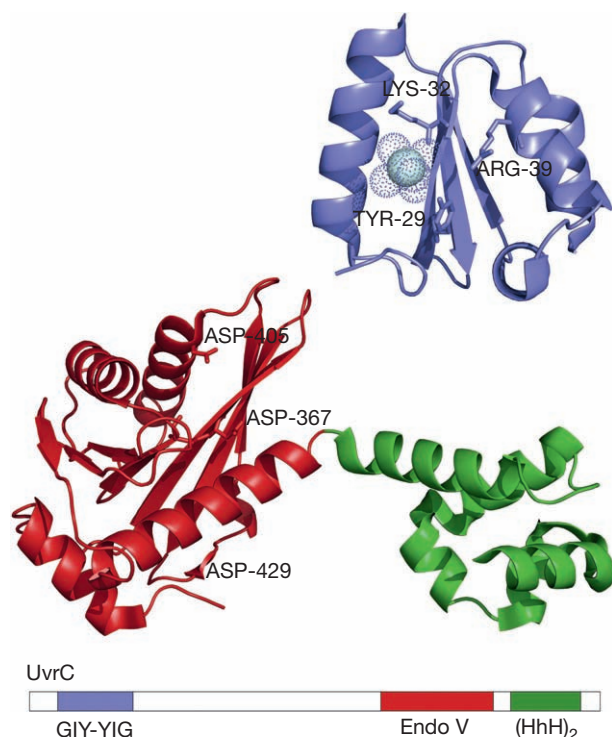


Figure 5 Structure of UvrC nuclease domains. The two distinct endonuclease centers of UvrC (PDB IDs:1YD5, 2NRR): the N-terminal GIY-YIG family nuclease domain is shown in blue and the residues which are essential for the cleavage are labeled. The metal and surrounded water molecules are shown as a CPK model. The C-terminal endonuclease domain is shown in red and the tandem HhH domain is shown in green. Residues that are necessary for the 5' incision are labeled.

Damage Recognition and Incision

The broad substrate specificity of the UvrABC endonuclease is achieved through a complex molecular mechanism (Figure 1) identifying some general changes in DNA conformation and dynamics rather than chemically modified bases *per se*.

Molecular Mechanism of Damage Recognition

Molecular recognition proceeds through a series of coordinated Uvr protein-protein and nucleoprotein intermediates characterized by different composition and architecture.

UvrA dimerization

UvrA forms a dimer in solution with apparent $K_d \sim 10^{-8}$ M. Nucleotide (ATP) binding to UvrA drives UvrA dimerization, whereas the hydrolysis of ATP or the presence of adenosine diphosphate (ADP) drives monomerization. The UvrA dimer is the active form in binding to undamaged or damaged DNA. The formation of a dimer results in a conformational change of the UvrA protein. UvrA is believed to recognize overall changes in the structure of the DNA helix, and has been found to bend the DNA by about 55° upon binding and unwind the DNA by as much as 3 bp upon DNA binding.

UvrA₂B₂ complex in solution

The purified UvrA protein associates in solution with UvrB protein (domain 2) in an ATP-dependent manner. The N-terminal region of about 150 amino acids of UvrA are involved in this interaction (Figure 3). UvrB has two UvrA-binding sites: one within amino acids 115–250 and another within C-terminal amino acids 573–673. However, UvrB* which lacks these C-terminal amino acids can also bind UvrA in solution. The apparent stoichiometry of the UvrA–UvrB complex in solution is UvrA₂UvrB₁.

UvrA₂–DNA binding

The UvrA dimer binds both damaged and undamaged DNA. ADP decreases the UvrA-binding affinity two- to threefold and ATP is not required for the specific binding to a damaged site. ATP- γ -S quantitatively inhibits the specific binding while enhancing nonspecific binding. The equilibrium constant for nonspecific binding of UvrA is $(0.7\text{--}2.9) \times 10^5 \text{ M}^{-1}$. The apparent binding affinity of UvrA₂ to damaged DNA is in the range $0.07\text{--}1.0 \times 10^9 \text{ M}^{-1}$, depending on the damage type. However, affinity of UvrA₂ to damage does not generally correlate with the efficiency of the entire incision reaction.

UvrA₂B₂–DNA complex

This complex provides for more productive damaged DNA-binding intermediates than UvrA dimer. Binding of UvrA₂B₂ to DNA results in locally unwound undamaged-DNA between 180° and 220° . Furthermore, it has been proposed that the DNA is wound around the UvrB molecule. The formation of the UvrA₂B₂–DNA complex results in a multifold stimulation of ATPase activity apparently through activation of a ‘cryptic’ ATPase of UvrB. Using atomic force microscopy, Goosen and co-workers have shown that UvrA and UvrB can form UvrA₂UvrB₂ complex on DNA. Van Houten and colleagues using single-molecule fluorescence has also shown that the UvrA dimer can serve as a scaffold for binding two UvrB molecules independently. Once the UvrAB complex actively engages the damage, UvrA (and possible one molecule of UvrB) dissociates leaving a long-lived stable UvrB–DNA complex.

Formation of the UvrB–DNA intermediate

Previous work by Grossman and co-workers have suggested that the UvrAB complex can track along the DNA helix using a limited helicase activity to create waves of negative and positive supercoiling behind and in front of the damaged site, respectively. This is consistent with the helicase fold of UvrB (shown in Figure 4). However, work from both the Rupp and Van Houten laboratories have argued that the UvrAB complex cannot dissociate damaged DNA strands even as small as 26 bases, and that the formation of the UvrB–DNA complex leads to a destabilization of the DNA helix. However, neither UvrB by itself nor in complex with UvrA can unwind long tracks of DNA, which is a typical property of proteins with true helicase activity. A recent study by Kad and Van Houten using single-molecule fluorescence of UvrA and UvrB labeled with quantum dots has shown that UvrA searches for DNA damage using a three-dimensional search, and can hop several microns from DNA duplex to DNA duplex. Surprisingly, the addition of UvrB

to UvrA changes the searching mode to one-dimensional sliding on DNA. Thus, UvrB is not only important for damage verification, but also changes the search mode for damage from three dimensions to one-dimensional sliding, rapidly increasing the rate of damage recognition.

Damage-Sensing Mechanism: Padlock Model of Damage Recognition by UvrB

The UvrABC endonuclease does not recognize a specific chemical group or structure in the damaged nucleotide nor does it, in all likelihood, solely recognize a specific backbone deformity in the duplex induced by the varied and unrelated genotoxic chemicals (Table 1). Initial damage recognition may occur through direct UvrA–DNA interactions. This triggers structural changes of the preceding UvrA₂B₁–DNA complex, so that UvrA₂ dissociates, while UvrB is loaded onto a damaged site. Even though the cocrystal structure of the UvrB–DNA complex has not yet been solved, this complex has been observed in footprinting experiments by electron microscopy and gel mobility shifts and the complex is isolated by gel filtration. The formation of this complex requires ATP hydrolysis by both UvrA and UvrB and occurs after the recognition of a damage site by the UvrAB complex. As noted above, Goosen and co-workers have shown that the formation of the UvrB pre-incision complex at damaged sites results in a conformational change of both UvrB and DNA. The DNA molecule in this complex is kinked and the DNA wrapped around the UvrB molecule. Helicase motifs and the β -hairpin motif of UvrB seem to play a role in UvrB–DNA binding at damaged sites (Figure 4). The rate of UvrB loading to DNA is low ($k_{on} = 6 \cdot 10^{-4} \text{ M s}^{-1}$), suggesting that this step may be limiting for damage-specific incision of DNA. Kisker and co-workers have proposed a padlock model of UvrB binding to DNA to help reconcile the earlier helicase-unwinding model of Grossman with the known helicase fold of UvrB. In this model, the UvrAB complex uses the UvrB ATPase to make limited motion along the DNA, causing transient opening and closing of the DNA, in which it is believed that the β -hairpin is inserted through the helix. If a DNA lesion is encountered by UvrB, there is a conformational change in the protein–DNA complex leading to the release of UvrA. In this precision complex, UvrB is believed to lock down the damaged between the β -hairpin and the wall of domain 1B (Figure 6). In this locked padlock model, the damaged strand wraps in front of the β -hairpin and the hydrophobic tyrosine side-changes at the base of the β -hairpin are believed to bind to or facilitate the damaged nucleotide to flip out of the DNA helix. In this locked-down conformation, ATP binding is believed to lead to a large motion of domain 3 which clamps down on the DNA, resulting in further deformation of the DNA helix leading to strained structure that is recognized by UvrC.

Dual Incision

The binding of UvrC to the pre-incision UvrB–DNA complex results in the incision of DNA on both sides of the lesion in the damaged strand. DNA sequence context and the type of

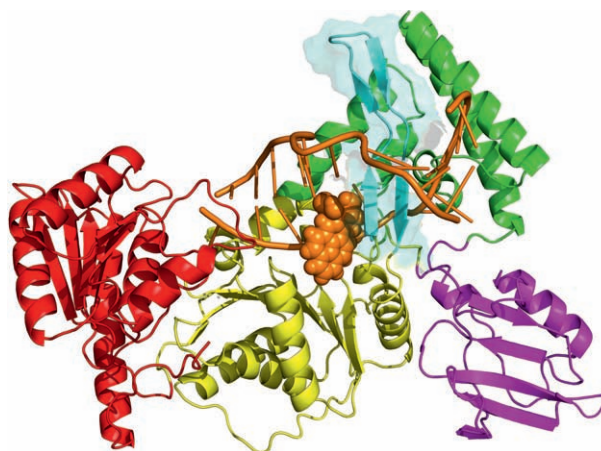


Figure 6 Hypothetical model of UvrB–DNA complex showing proposed damage-recognition pocket. Domains 1b and 3 are indicated as in Figure 4. The proposed conformation of DNA is shown as phosphate backbone in orange with the damaged strand extending under the β -hairpin which is depicted in cyan.

damage affect the precise location of the 5' and 3' incision sites. Usually, it occurs at the 8th phosphodiester bond 5' to the damage and 4th or 5th phosphodiester bond 3' to the damage. The C-terminal coiled-coil motif of UvrB (amino acids 636–668) and the N-terminal homologous regions of UvrC (amino acids 214–239) are required for the formation of UvrBC complex (Figure 5). Site-directed mutagenesis of the UvrC and UvrB proteins has led to the conclusion that the incision process is not a concerted one. The 3' incision is affected by the N terminus of the UvrC protein, and this incision precedes the 5' nicking, catalyzed by the C-terminus of UvrC protein. There is a conformational change of UvrBC–DNA complex following 3'-incision. The C-terminal region of UvrC (amino acids 555–610) is involved in DNA binding and 5'-incision. Dual incision *per se* does not require ATP hydrolysis; ATP- γ -S can substitute ATP during this step.

Postincision Steps

Following dual incision, the UvrBC complex remains bound to damaged DNA. As a result of the stability of the postincision complex, UvrABC endonuclease does not turnover in an *in vitro* system, which is biochemically reconstituted with only Uvr proteins. Restoration of the DNA primary structure and integrity as well as turnover of the Uvr proteins is accomplished in a series of postincision steps, which require products of two additional genes, *uvrD* and *polA*.

UvrD

Its gene maps at 84 min of *E. coli* chromosome. UvrD + is a part of the SOS regulon and is induced by DNA damage. The product of this gene is a 3'-5' DNA helicase II – a 75-kDa protein, which is a DNA-dependent ATPase. DNA helicase II is required for the removal of incised damage as a 12–13-mer oligonucleotide fragment. It also effects the release of UvrC protein, but not UvrB from the postincision complex.

PoiA

Its gene maps at 85 min of *E. coli* genetic map. It encodes a 109-kDa protein, DNA polymerase I. The UvrB protein is not displaced by DNA helicase II and remains bound to gapped DNA. In the presence of deoxyribonucleoside-5'-triphosphates, DNA polymerase I catalyzes repair synthesis, resulting in the filling in of the gap and displacement of the UvrB protein. The size of *in vitro* repair patch is usually ~12 nucleotides in length. Molecular interactions of Uvr proteins, DNA helicase II, and DNA polymerase I during postincision events are not presently known. *E. coli* strains, deficient in *uvrD*⁺ and *polA*⁺, are usually significantly less sensitive to the killing effects of UV radiation than Uvr-deficient strains, indicating their auxiliary role in repair. DNA ligase completes NER by catalyzing joining of the last 3'-nucleotide of the patch with the rest of polynucleotide chain.

Transcription-Coupled Repair

Transcription-repair coupling (TRC) allows cells to select some regions of the genome for NER in preference to others. In *E. coli*, as a consequence of a 436-fold induction of

β -galactosidase operon, 70% of the UV-induced pyrimidine dimers are removed from the transcribed strand of the induced operon within 5 min, whereas only 50% of dimers are removed from the nontranscribed strand after 20 min of repair. This selective removal of pyrimidine dimers from the transcribed strand of a gene is abolished in the absence of significant levels of transcription. The mechanism of TRC in *E. coli* is a complex one, and may include several subpathways. In one of them, TRC is achieved through the action of transcription-repair coupling factor (TRCF). TRCF is a product of mutation frequency decline (*mfd*⁺) gene, which maps at 25 min on the *E. coli* chromosome. The cloned *mfd*⁺ gene is translated into a 1148-amino-acid protein of ~130 kDa. The Mfd protein has consensus Walker type A nucleotide-binding motif, and, indeed, is a relatively weak ATPase. The Mfd protein can nonspecifically bind dsDNA (and less efficiently ssDNA) in an ATP binding-dependent manner, the ATP hydrolysis promoting its dissociation. The amino acid sequence of Mfd reveals motifs, which are characteristic of a number of DNA and RNA helicases. However, *in vitro* purified Mfd does not show either DNA or RNA helicase activity. The N-terminal 1–378 residues of Mfd, having a 140-amino-acid region of homology with UvrB, bind UvrA protein (Figure 7). The Mfd protein can also bind different

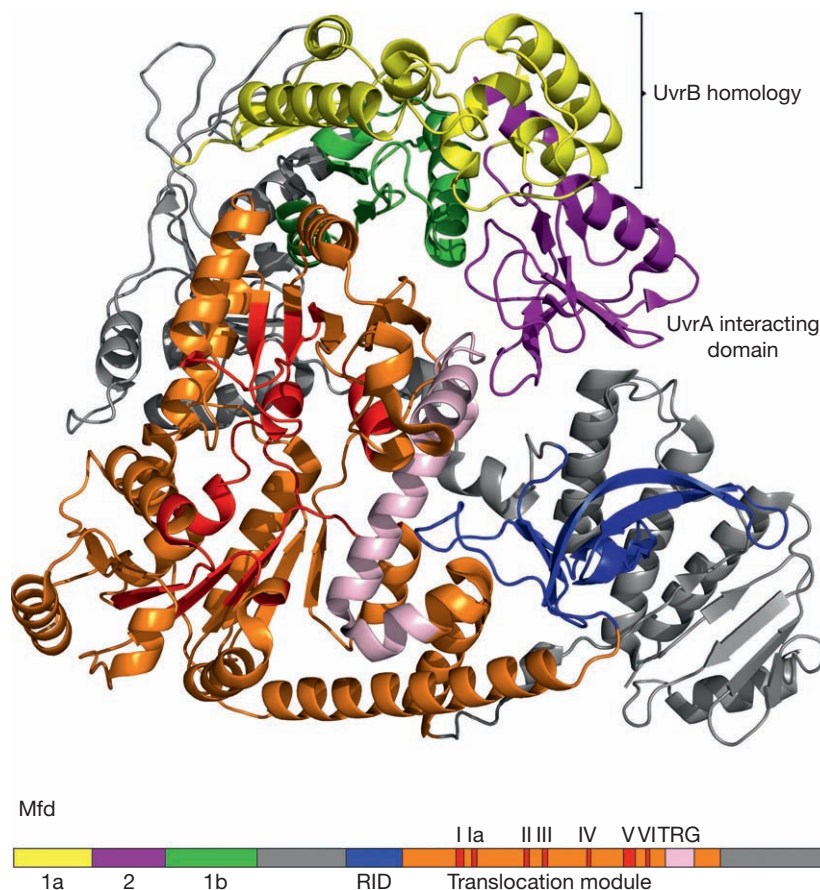


Figure 7 Crystal structure of Mfd. The Mfd structure (PDB ID 2EYQ): the domains that are homologous to UvrB homolog are shown in the same color scheme as for the UvrB molecule (1a in yellow, 2 in magenta, and 1b in green), and domain 2 is labeled as UvrA interacting domain. The translocation domain of Mfd, which is shown in orange, contains all seven SF2 helicase motifs which are shown in red. The RNAP interacting domain is shown in blue, and the TRG motif in light pink.

forms of *E. coli* RNA polymerase (RNAP). Amino acids 370–571 of Mfd are involved in this binding. *In vitro*, transcription inhibits NER of damage in a transcribed strand, whereas it has no effect on the coding strand. This inhibition is thought to result from RNAP stalled at the site of damage. Mfd is able to release stalled RNAP in an ATP-hydrolysis-dependent manner. Moreover, Mfd actually stimulates NER on the transcribed strand, so that it becomes faster than on the nontranscribed one. Based on all these observations, it is concluded that TRCF-Mfd carries out preferential repair of the transcribed strand by (1) releasing RNAP stalled at damaged site, and (2) recruiting the UvrA₂B₂ complex to damaged site through the high-affinity interaction with UvrA protein. The TRC in *E. coli* is influenced by the rate and conditions of growth as well as the level of gene transcription. In the LacZ gene, the Mfd-dependent mechanism of preferential repair dominates at basal levels of transcription, whereas Mfd-independent mechanisms prevail at induced transcription levels.

Conservation of NER from Bacteria to Humans

NER of bacterial, yeast, and mammalian cells have been reconstituted *in vitro* with purified proteins and defined damaged DNA substrates. Although it would appear that it takes nearly 25 proteins in a human cell to complete the process achieved by six bacterial proteins, there is a high degree of conservation of function from bacteria to eukaryotes. In fact, the recent structure of xeroderma pigmentosum group D indicates that it shares a common fold with UvrB and may have a similar mode of damage recognition. Thus, NER system of bacteria will certainly serve as a primary model for structure–function studies for years to come.

See also: **Protein/Enzyme Structure Function and Degradation:** Zinc Fingers; **Molecular Biology:** DNA Damage: Alkylation; DNA Polymerase I, Bacterial; Nonhexameric SF1 DNA Helicases/Translocases; DNA Helicases: Hexameric Enzyme Action.

Further Reading

- Goosen N and Moolenaar GF (2001) Role of ATP hydrolysis by UvrA and UvrB during nucleotide excision repair. *Research in Microbiology* 152(3–4): 401–409.
- Grossman L and Thiagalingam S (1993) Nucleotide excision repair, a tracking mechanism in search of damage. *Journal of Biological Chemistry* 268: 16871–16874.
- Kad NM, Wang H, Kennedy GG, Warshaw DM, and Van Houten B (2010) Collaborative dynamic DNA scanning by nucleotide excision repair proteins investigated by single-molecule fluorescence imaging of quantum dot labeled proteins. *Molecular Cell* 37: 702–713.
- Karakas E, Truglio JJ, Croteau D, et al. (2007) Structure of the C-terminal half of UvrC reveals an RNase H endonuclease domain with an Argonaute-like catalytic triad. *EMBO Journal* 26(2): 613–622.
- Mellon I and Hanawalt PC (1989) Induction of the *Escherichia coli* lactose operon selectively increases repair of its transcribed DNA strand. *Nature* 342: 95–98.
- Pakotiprapha D, Inuzuka Y, Bowman BR, et al. (2008) Crystal structure of *Bacillus stearothermophilus* UvrA provides insight into ATP-modulated dimerization, UvrB interaction, and DNA binding. *Molecular Cell* 29(1): 122–133.
- Sancar A (1996) DNA excision repair. *Annual Review of Biochemistry* 65: 43–81.
- Sancar A and Rupp WD (1983) A novel repair enzyme: UVRABC excision nuclease of *Escherichia coli* cuts a DNA strand on both sides of the damaged region. *Cell* 33: 249–260.
- Skorvaga M, Theis K, Kisker C, and Van Houten B (2002) β -Hairpin motif of UvrB is essential for DNA binding, damage processing and UvrC mediated-incisions. *Journal of Biological Chemistry* 277(2): 1553–1559.
- Theis K, Chen PJ, Skorvaga M, Van Houten B, and Kisker C (1999) Crystal structure of UvrB provides insight into the mechanism of nucleotide excision repair. *EMBO Journal* 18(24): 6899–6907.
- Truglio J, Croteau DL, Van Houten B, and Kisker C (2006) Prokaryotic nucleotide excision repair: The UvrABC system. *Chemical Reviews* 106: 233–252.
- Truglio JJ, Karakas E, Rhau B, et al. (2006) Structural basis of DNA damage recognition and processing by UvrB. *Nature Structural and Molecular Biology* 13: 360–364.
- Truglio JJ, Rhau B, Croteau DL, et al. (2005) Structural insights into the first incision reaction during nucleotide excision repair. *EMBO Journal* 24: 885–894.
- Van Houten B (1990) Nucleotide excision repair in *Escherichia coli*. *Microbiology Reviews* 54: 18–51.
- Van Houten B, Croteau DL, DellaVecchia MJ, Wang H, and Kisker C (2005) 'Close fitting sleeves': DNA damage recognition by the UvrABC nuclease system. *Mutation Research* 577: 92–117.
- Van Houten B, Eisen JA, and Hanawalt PC (2002) A cut above: Discovery of an alternative excision repair pathway in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 99(5): 2581–2583.
- Wolski SC, Kuper J, Hänzelmann P, et al. (2008) Crystal structure of the FeS cluster containing nucleotide excision repair helicase XPD. *PLoS Biology* 6: e149.

Nucleotide Excision Repair: Biology

E C Friedberg, University of Texas Southwestern Medical Center, Dallas, TX, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 143–147, © 2004, Elsevier Inc.

Glossary

Damaged DNA DNA in which the covalent structure, nucleotide sequence, or conformation has been altered.

DNA repair A collection of cellular processes whereby damaged DNA is restored to its normal chemistry, nucleotide sequence, conformation, and function.

Eukaryotes One of the kingdoms of life, more highly evolved than bacteria, the cells of which contain a defined nucleus.

Oligonucleotides Fragments of one or two DNA strands that can be of any size.

Transcriptionally active genes Genes that are in the active process of being transcribed into RNA.

The name nucleotide excision repair (NER) derives from the fact that during this process, damaged bases are excised from the genome as oligonucleotide fragments. This distinguishes NER from other types of excision repair such as base excision repair (BER), during which damaged bases are excised as free bases, and mismatch excision repair, during which mismatched nucleotides are removed from the genome as mononucleotides. The oligonucleotide fragments are generated by specific incision of the damaged DNA strand on either side of sites of base damage (bimodal incision) (Figure 1). The gaps (~30 nucleotides in length) generated by oligonucleotide excision are filled in by repair synthesis (Figure 1), using the intact opposite strand as an informational template. When the last nucleotide is inserted, the remaining nick is sealed by DNA ligase (Figure 1).

Modes of NER

NER transpires in cells in various subforms. Its operation on transcriptionally silent regions of the genome and on the nontranscribed strand of transcriptionally active genes is often referred to as global NER, to distinguish it from NER that specifically operates on the transcribed strand of transcriptionally active genes, so-called transcription-coupled NER (TC-NER). Additionally, both global NER and TC-NER are characterized by incision of the affected DNA strand on either side of sites of base damage (Figure 1); some lower eukaryotes support a form of NER that involves a single incision 5' to the site of base damage. This process is referred to as specialized NER.

The Biology of NER

Distribution in Nature

As we ascend the evolutionary ladder and encounter increasingly more complex organisms, our information about the cellular biology of NER becomes less precise, even though paradoxically our biochemical understanding has taken leaps and bounds in recent years. Several aspects of the cell biology of higher cells contribute to this paucity of information. Perhaps most importantly, both the structure and organization of the eukaryotic genome are much more complex than that of

prokaryotes. Eukaryotic genomes are considerably larger, and the requirements for packaging such large amounts of DNA into the limited confines of the cell nucleus have led to a complex structural organization. A consideration of the distribution of DNA damage in eukaryotes must take into account that nuclear DNA exists in intimate association with both histones and nonhistone chromosomal proteins, and that the structure of chromosomes reflects various levels of folding and coiling of the basic chromatin structural unit, the nucleosome.

Our present understanding of how chromosome structure and nucleosome conformation influence the enzymology of DNA incision and of postincision events during NER in eukaryotes is still scanty. It is certain that structural elements do, in fact, limit the access of repair enzymes to sites of base damage, and that specific perturbations of chromatin structure are necessary to facilitate NER.

Conventional NER involving the bimodal incision of damaged DNA is operationally and fundamentally similar in all organisms, from bacteria to humans. However, there are significant biochemical differences between NER in prokaryotes and eukaryotes. Additionally, the biochemistry of global NER and TC-NER differs in important details during the initial stage of the process. Specifically, during global NER, particular proteins are required for the recognition of base damage in DNA. In contrast, during TC-NER, the process of arrested RNA polymerase II transcription is believed to provide the mechanism of damage recognition.

The NER Machinery

NER operates through the assembly of a multiprotein complex. In bacteria, this complex comprises about six polypeptides. However, in both lower and higher eukaryotes, as many as 30 proteins are required for NER. These comprise proteins that participate in (1) the recognition of base damage, (2) bimodal incision of DNA, (3) oligonucleotide fragment incision, (4) repair synthesis of the gap generated in DNA, and (5) DNA ligation (Figure 1). In eukaryotes, the RNA polymerase II, basal transcription factor IIH (TFIIH), is an integral component of the NER machinery. As noted below, mutations in some of the genes encoding TFIIH subunits can result in disease states with transcription defects.

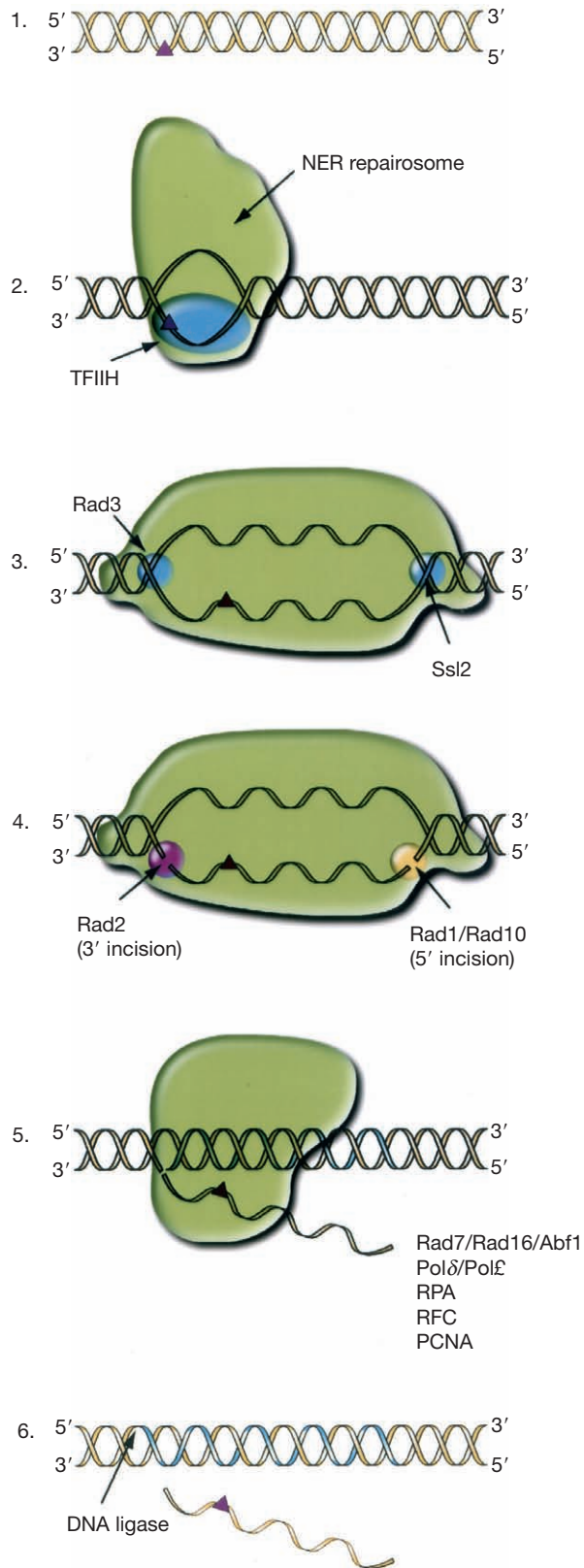


Figure 1 The general mechanism of nucleotide excision repair (NER) in eukaryotes. In the figure, yeast is used as an example. The red triangle represents some type of base damage that is recognized by the transcription-independent NER system. The binding of multiple proteins

Substrate Specificity

NER operates primarily on types of base damage that generate conformational distortion of the DNA helical structure. The so-called (6-4) photoproduct, a highly distortive type of base damage in DNA that results from exposure to ultraviolet (UV) light (Figure 2), as well as cyclobutane pyrimidine dimers (Figure 2), are premier examples of substrates for NER. The NER machinery does not recognize small loops or mismatches in DNA, and small chemical adducts such as those produced by methylation and the oxidation of bases are preferentially recognized by the BER process. The precise mechanism of substrate recognition during NER in eukaryotes remains to be determined.

Detection of NER *In Vivo* and *In Vitro*

Studies on NER in mammalian cells have historically utilized a variety of increasingly refined experimental strategies and techniques, many of which have been adapted from methods initially developed for studies with bacteria. Some of these procedures measure the damage-specific incision of DNA, either directly or indirectly. Others measure the physical excision of damaged nucleotides, repair synthesis of DNA, or the rejoining of strand breaks (DNA ligation). A particularly widely used technique for monitoring NER in living cells is referred to as the unscheduled DNA synthesis (UDS) assay, which identifies repair synthesis of DNA. In this procedure, cells in monolayer culture are exposed to a DNA-damaging agent such as UV radiation and the medium is supplemented with a radiolabeled precursor for DNA synthesis. Repair synthesis of DNA associated with NER is identified by a delicate autoradiographic stippling of nuclei in cells that are outside the S phase of the cell cycle (hence UDS) (Figure 3).

Kinetics of NER

Although the application of many of the experimental procedures mentioned above supports the existence of NER in higher organisms, the relative kinetics of specific events associated with NER, such as DNA incision, the excision of specific lesions from high-molecular-weight DNA, and repair synthesis in mammalian cells, are still controversial. In part, these controversies have arisen from the use of different experimental procedures, which, though designed to measure the same biological endpoints, sometimes have optimal reliabilities at different extents of DNA damage and under different

at or near the site of base damage in an ordered step-wise fashion generates a large multiprotein complex, the NE repairosome, which includes the RNA polymerase II basal transcription factor TFIIH. Two DNA helicase subunits in TFIIH (Rad3 and Ssl2 in yeast) facilitate the generation of a bubble in the DNA that is ~30 nucleotides in length and that flanks the site of base damage. The double-stranded/single-stranded DNA junctions are then recognized by two structure-specific endonucleases (Rad2 and Rad1/10). These cut the DNA duplex only on the damaged strand, generating an oligonucleotide fragment ~30 nucleotides in length. The Rad7/Rad16/Abf1 protein complex then facilitates displacement of the oligonucleotide as a free DNA fragment carrying the damaged base. Replicative DNA polymerases (δ and/or ε) together with accessory proteins (RPA, RFC, and PCNA) fill in the gap formed and DNA ligase seals the damaged DNA strand to restore complete covalent integrity.

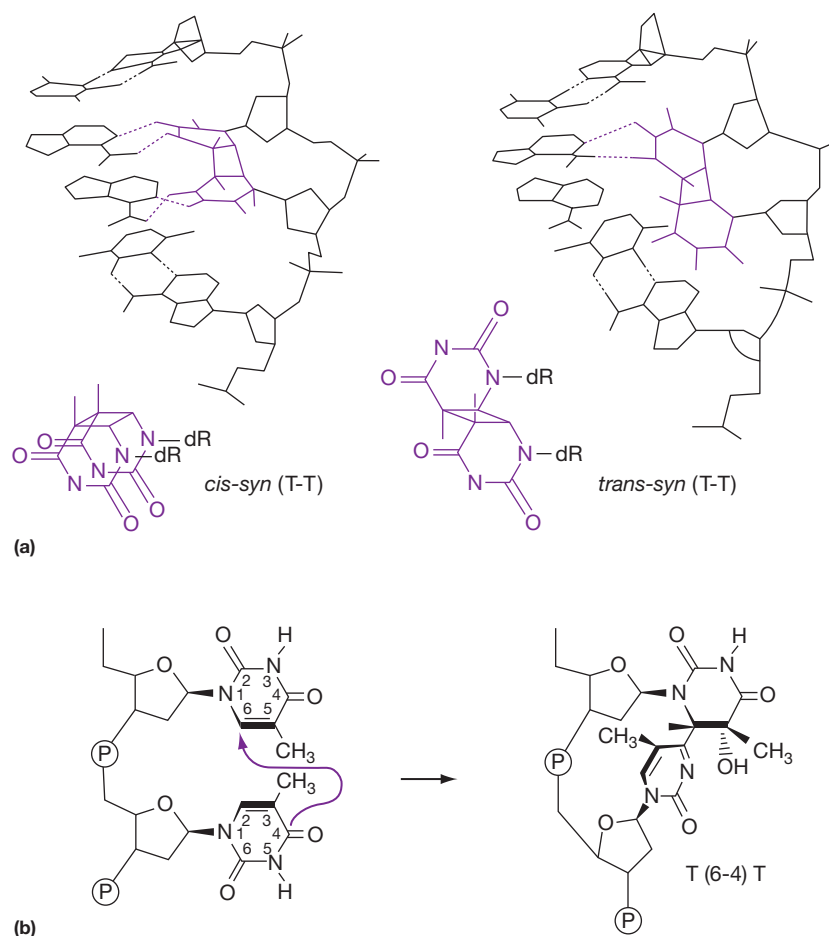


Figure 2 Chemical structures of cyclobutane pyrimidine dimers (CPD) (a) and (6-4) photoproducts (b). Both of these photoproducts are generated in large amounts when DNA is exposed to UV light, including sunlight.

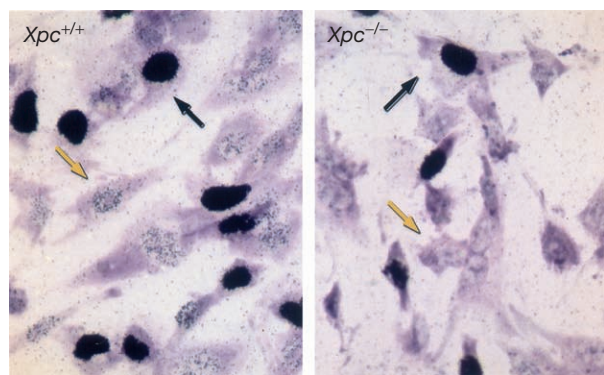


Figure 3 The identification of repair synthesis during NER in mammalian cells. The repair synthesis step shown in Figure 1 can be monitored in mammalian cells in monolayer culture. Mouse wild-type ($Xpc^{+/+}$) and NER-defective ($Xpc^{-/-}$) cells are exposed to UV radiation and allowed to incubate in the presence of [3H] thymidine. Cells in S phase of the cell cycle undergo intense autoradiographic labeling by incorporating the radiolabeled thymidine (black arrows). However, cells undergoing repair synthesis during NER incorporate much less label which can be identified as a delicate autoradiographic stippling (yellow arrow, $Xpc^{+/+}$). These grains are greatly reduced or absent in the NER-defective $Xpc^{-/-}$ cells.

experimental conditions, and hence may not be strictly comparable. Nonetheless, some of these apparent experimental inconsistencies may actually reflect real distinctions in NER associated with different cell types, different functional states of cells, particular regions of the genome, and different types of DNA damage. It is generally acknowledged that (6-4) photoproducts are removed from the genome of human, and from particular rodent cells, more rapidly than are pyrimidine dimers. It can take many hours before the latter lesions are fully cleared from the genome.

Defective NER and Human Disease

Several human hereditary diseases that result from defective NER have been identified. The so-called classical xeroderma pigmentosum (XP) is an autosomal-recessive disease that results from defects in both global NER and TC-NER. Affected individuals are extremely sensitive to sunlight and have a markedly increased risk of developing skin cancers. A clinically indistinguishable form of XP, called the XP variant form, derives from mutations that affect a different biological response to UV radiation-induced DNA damage. Individuals with mutations in genes required for both NER and RNA

polymerase II transcription can develop the clinical features of both XP and a different disease called Cockayne syndrome (CS). Pure CS (unaccompanied by XP) arises from defects in the CSA or CSB genes that encode proteins required for TC-NER. Unlike XP, CS does not lead to an increased risk of skin cancer in human subjects. Defects in subunits of TFIIH can result in a spectrum of diseases characterized by developmental and neurological defects with or without defective NER, the disease trichothiodystrophy being a notable example.

Mouse Models for Defective NER

Mouse models for each of the human hereditary diseases mentioned above have been generated by conventional targeted-gene replacement, and sometimes by more refined strategies. In general, these mice have proven to mimic the human disease well and have provided much valuable information about the relationships between defective NER and cancer predisposition and pathogenesis.

See also: **Bioenergetics:** Glutathione Peroxidases; **Cell Architecture and Function:** Unfolded Protein Responses; **Molecular Biology:** RecQ Helicase Systems; Ribosome Assembly; Riboswitches; **Signaling:** Gi Family of Heterotrimeric G Proteins; mTOR and its Downstream Targets.

Further Reading

- Batty DP and Wood RD (2000) Damage recognition in nucleotide excision repair of DNA. *Gene* 241: 193–204.
- de Boer J and Hoeijmakers JHJ (2000) Nucleotide excision repair and human syndromes. *Carcinogenesis* 21: 453–460.
- Friedberg EC (2001) How nucleotide excision repair protects against cancer. *Nature Reviews Cancer* 1: 22–33.
- Friedberg EC, Walker GC, Siede W, et al. (1995) *DNA Repair and Mutagenesis*, 2nd edn. Washington, DC: ASM Press.
- Norbury CJ and Hickson LD (2001) Cellular responses to DNA damage. *Annual Review of Pharmacology and Toxicology* 41: 367–401.

Nucleotide Excision Repair in Eukaryotes

N Goosen, Leiden University, Leiden, The Netherlands

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Laura A. Lindsey-Boltz and Aziz Sancar, volume 3, pp. 130–133, © 2004, Elsevier Inc.

Glossary

Chromatin Compacted structure of DNA and proteins. DNA wraps around histones to form nucleosomes. A further compaction of the nucleosomes leads to various levels of condensation of the chromosomal DNA.

DNA helicase Enzyme that uses energy of adenosine triphosphate hydrolysis to separate the two strands of a DNA duplex molecule.

DNA ligase Enzyme that repairs single-stranded breaks in the DNA backbone by linking the 3' OH of the ribose to the 5' phosphate of the adjacent nucleotide.

Nuclease Enzyme that cleaves the bond between the ribose and the phosphate in the backbone of a DNA strand.

Photoproduct Cross-link between two adjacent pyrimidine bases in the DNA caused by ultraviolet irradiation. The two main photoproducts are cyclobutane pyrimidine dimers (CPDs) and 6–4 photoproducts.

Proliferating cell nuclear antigen (PCNA)

Clamp Multimeric protein forming a ring-like structure that is loaded on double-stranded DNA. It serves to anchor DNA polymerase on its substrate during DNA replication.

WD40 repeat Structural motif of about 40 amino acids often terminating in a Trp-Asp (W-D) dipeptide. WD40 repeat proteins contain 4–16 repeats that form a β -propeller structure.

Introduction

Genomic DNA is constantly threatened by endogenous and environmental factors causing damage to the bases or backbone of the DNA. If left unrepaired, such damage may block important processes such as DNA replication or transcription, leading to cell death. Alternatively, erroneous bypass of a lesion by the DNA polymerase may lead to mutations. To counteract the deleterious effects of DNA damage, a variety of DNA repair mechanisms have evolved. Among these repair mechanisms, nucleotide excision repair (NER) is unique in its capacity to repair a large variety of structurally unrelated lesions. It can remove cross-links induced by ultraviolet (UV) light, such as cyclobutane pyrimidine dimers (CPDs) and 6–4 photoproducts (6–4PP), and it also eliminates many types of bulky carcinogen–DNA adducts. The biological importance of NER is evident from the clinical features of patients with hereditary defects in this repair pathway, which include extreme sun sensitivity, elevated occurrence of skin cancer, neurological degeneration, and developmental delays.

Although eukaryotes and bacteria use a similar cascade of events to repair damaged DNA via the NER pathway, the proteins involved are not evolutionary related. Apparently, the two similar DNA repair systems have evolved independently in both kingdoms of life. The eukaryotic NER factors (listed in Table 1) have been characterized extensively both in human and in the budding yeast *Saccharomyces cerevisiae*, and the complete repair reaction has been reconstituted *in vitro* with the purified proteins (Figure 1).

Damage Recognition

For initial damage recognition during NER, two subpathways are used: transcription-coupled repair (TCR, discussed next) and global genome repair (GGR).

The GGR pathway is initiated by the XPC complex, which scans the DNA for aberrant structures. In this complex,

the 125-kDa XPC protein is associated with HHR23B, a 58-kDa homolog of the yeast Rad23 and the 18-kDa centrosomal factor centrin-2. The XPC protein is the actual damage-recognition subunit of the complex, whereas the two other subunits have a regulatory function by protecting XPC from degradation. XPC preferentially binds to DNA regions with impaired Watson–Crick base pairing, which can be indicative for presence of a damaged nucleotide. The co-crystal structure of the yeast homolog Rad 4 with DNA containing a damaged site revealed the mode of DNA binding (Figure 2). The main portion of the protein binds the double-stranded part of the DNA at the 3' side of the lesion. In addition, two β -hairpin domains of the protein fold into a hand-like structure that holds a small region of single-stranded DNA opposite the lesion. The Rad4 protein does not make any contact with the lesion itself, as can be expected for a protein that recognizes helix distortions caused by many different types of damage.

The level of active XPC in the nucleus is tightly regulated and responds at different levels to challenging conditions. Upon genotoxic stress, expression of XPC is enhanced at the transcriptional level; the protein is protected from degradation by interaction with HHR23B and centrin-2; DNA-binding activity of XPC is enhanced by posttranslational modification; and XPC is prevented from shuttling to the cytoplasm and accumulating in the nucleus.

As the main determinant for XPC binding is the presence of unpaired bases in the DNA, lesions that induce only a minor distortion are poorly recognized. An example of such a lesion is the UV-induced CPD. Efficient recognition of a CPD by XPC therefore requires an additional factor, UV-DDB. The UV-DDB complex, which consists of two proteins, DDB1 (127 kDa) and DDB2 (48 kDa), specifically binds to the two major UV photolesions CPD and 6–4PP. In contrast to XPC, UV-DDB does make contacts with the damaged bases by extruding them from the helix into a binding pocket of the DDB2 protein (Figure 2). The interaction of UV-DDB with the photoproduct subsequently facilitates the binding of XPC.

Table 1 The human nucleotide excision repair (NER) factors and their yeast homologs

Human	Yeast	Function
Core NER-factors		
XPC-complex	Rad4-complex	Damage recognition
XPC	Rad4	
hHR23B	Rad23	
Centrin-2	Rad33	
TFIIH	TFIIH	Damage verification
XPD	Rad3	Open-complex formation
XPB	Rad25	
p62	Tfb1	
p52	Tfb2	
p44	Ssl1	
p34	Tfb4	
p8	Tfb5	
XPA	Rad14	Coordination, stabilization (damage verification?)
RPA	RPA	Stabilization open complex
RPA1	Rfa1	
RPA2	Rfa2	
RPA3	Rfa3	
XPG	Rad2	3' Incision
XPF-ERCC1	Rad1-Rad10	5' Incision
Global genome repair (GGR)		
UV-DDB	-	Damage recognition
DDB1		
DDB2 (XPE)		
-	Rad7-Rad16	
Transcription-coupled repair (TCR)		
CSA	-	
CSB	Rad26	

Yeast does not contain any homologs of the DDB proteins. Possibly, in this organism, other proteins can stimulate the binding of Rad4 to photoproducts, but such proteins have not yet been identified. However, yeast does contain another factor that is specifically required for the initial steps in GGR, a complex of the Rad7 and Rad16 proteins. The exact function of this complex is not yet known, but it has been proposed that it acts by remodeling the chromatin, thereby facilitating access of repair factors to damaged sites, which are obscured by nucleosomes.

Damage Verification and Formation of the Open Complex

Upon recognition of a mispaired DNA region by XPC, the next step in the repair reaction is to verify if indeed a lesion is present in this mispaired region. This is accomplished by the seven-protein TFIIH core complex. TFIIH was originally identified as a basal transcription factor required for RNA polymerase, to initiate transcription. The key components of TFIIH are the DNA helicases XPD and XPB. DNA helicases use the energy from adenosine triphosphate (ATP) hydrolysis for translocation along one of the strands of a DNA helix, thereby separating the two strands of the duplex. For transcription, this strand separation is required to generate a transcription bubble in which RNA polymerase can incorporate RNA nucleotides opposite one of the DNA strands. Moreover, in NER, the helicase activity of TFIIH also leads to a further opening of the DNA around the lesion creating a 10–20-bp bubble region. During translocation along the DNA strand, the helicases are arrested at a damaged site. In this way, TFIIH can not only confirm the presence of a lesion, but also discriminate which of

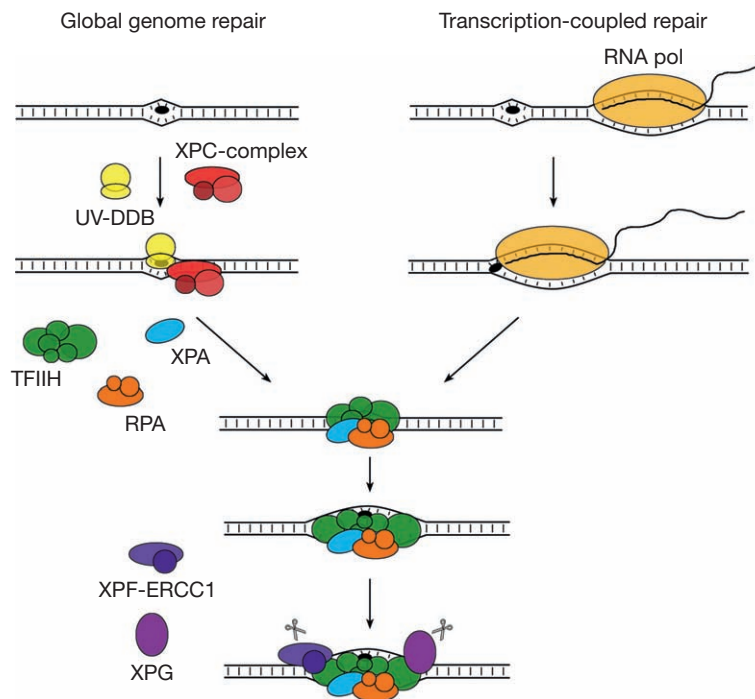


Figure 1 Two subpathways for eukaryotic nucleotide excision repair. In global genome repair (GGR), the damage is initially detected by the XPC complex. In transcription-coupled repair (TCR), RNA polymerase acts as a sensor for DNA damage. Both pathways are followed by a common cascade of events where the presence of the damage is verified by TFIIH, the DNA around the lesion is opened up by TFIIH, XPA, and RPA, and the DNA is incised by XPF-ERCC1 and XPG.

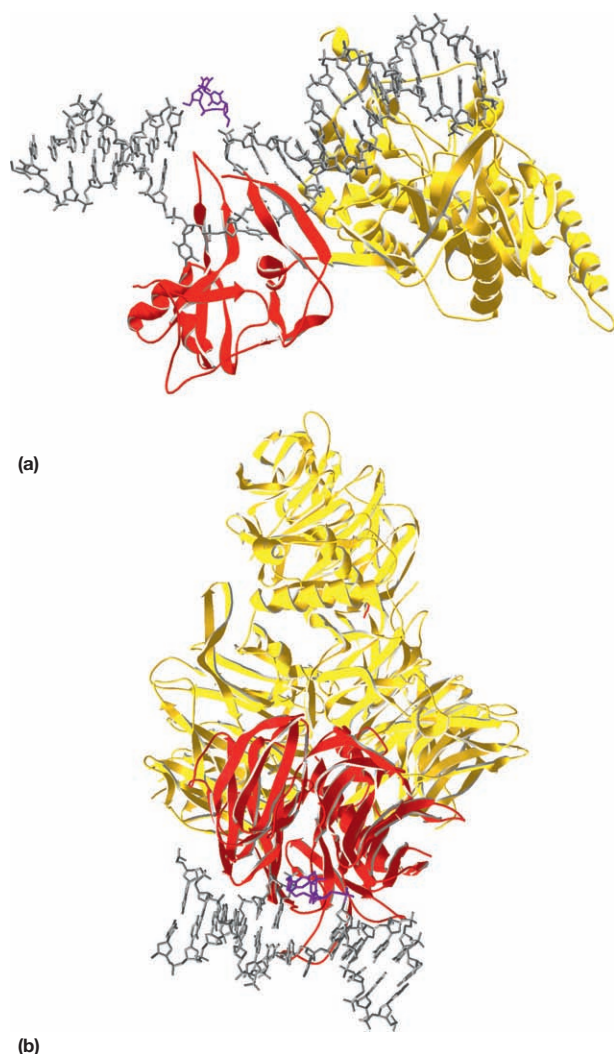


Figure 2 (a) Structure of the yeast Rad4 protein bound to DNA containing a photoproduct. Two β -hairpin domains bind to the DNA strand opposite the lesion. The photoproduct is flipped away from the protein. (b) Structure of the UV-DDB complex bound to DNA containing a photoproduct. The photoproduct is flipped into a pocket of the DDB2 subunit.

the two DNA strands is damaged. Mutational analysis has indicated that it is mainly the XPD helicase that serves as the damage-verification factor.

The formation of the open complex around the lesion is further stabilized by binding of RPA and XPA. RPA, consisting of three subunits, is a single-stranded DNA-binding factor that prevents reannealing of the separated DNA strands. The main function of the 31-kDa XPA protein is to coordinate the assembly of the different repair factors in the open complex. The XPA protein, however, was also shown to preferentially bind to sharply bent DNA structures. Therefore, this protein might serve as yet another damage-verification factor by probing the damage-induced bendability of the DNA.

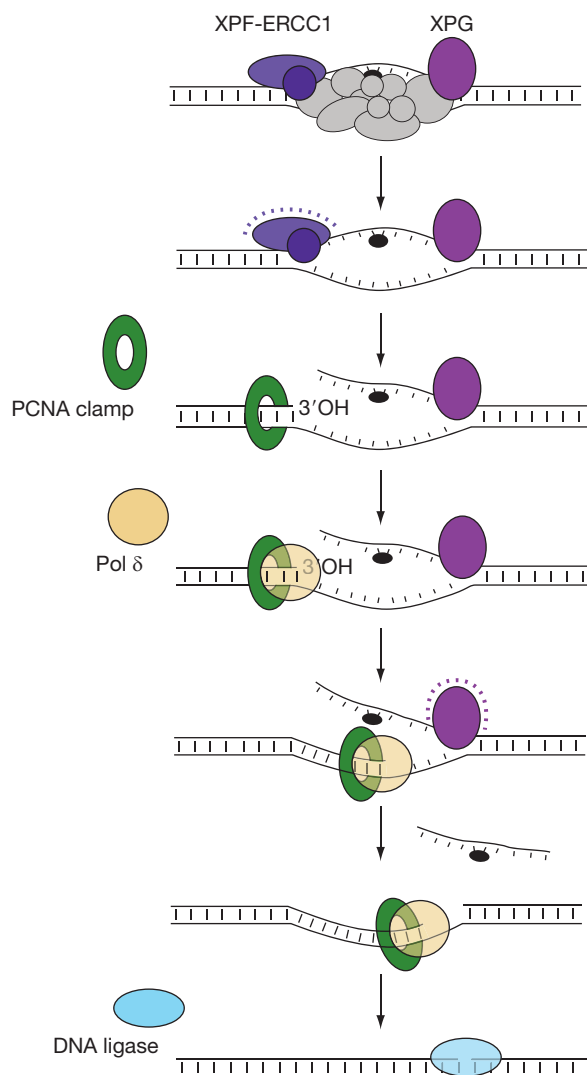


Figure 3 Incision and DNA synthesis occur in a coordinated manner. XPF-ERCC1 induces the first incision on the 5' side of the lesion. The PCNA clamp is loaded on the DNA, and DNA polymerase extends the free 3' end. When DNA polymerase has proceeded halfway through the repair patch, XPG is triggered to incise the DNA at the 3' side of the lesion. Upon completion of the DNA repair patch, DNA ligase seals the remaining nick. For clarity, the other repair factors are omitted from the drawings.

Incision of the Open Complex

The creation of the open complex by the concerted action of the repair proteins discussed above leads to the formation of distinct junctions between double- and single-stranded DNA on both sides of the lesion. These junctions are specifically recognized by two structure-specific nucleases, XPG and the ERCC1-XPF complex. Depending on the actual size of the bubble in the open complex, XPG incises the DNA 2–8 nt from the 3' side and ERCC1-XPF 15–24 nt from the 5' side of the lesion. Interaction with repair proteins present in the open complex ensures that each nuclease will be positioned at the proper junction. Simultaneous action of both nucleases,

followed by simple release of the damage-containing oligonucleotide, would result in a large (24–32 nt) single-stranded gap. Such large gaps create a high risk in the cell for double-stranded DNA breaks and should be avoided.

Therefore, the processes of incision and resynthesis of DNA are tightly coordinated (Figure 3). The 5' incision by the XPF subunit of the ERCC1/XPF complex occurs first, creating a free 3'-OH group. The clamp loader RFC subsequently loads the proliferating cell nuclear antigen (PCNA) clamp near the junction, as is the case in DNA replication, after which polymerase δ initiates DNA synthesis from the free 3' end. Only after the DNA polymerase has proceeded halfway through the repair patch, the XPG protein is triggered to induce the 3' incision. This subsequently allows completion of the repair synthesis by polymerase δ after which the final nick is sealed by DNA ligase.

Transcription-Coupled Repair

In TCR, the RNA polymerase that moves along the DNA during RNA synthesis is used as the initial damage sensor. The elongating RNA polymerase is blocked when a lesion is present in the transcribed strand (Figure 1). Specific TCR factors recognize the stalled RNA polymerase, which subsequently leads to the recruitment of the core NER factors. Because the scanning of DNA by an elongating polymerase is much more efficient than by the association/dissociation of damage-specific proteins, the repair of the transcribed strand in any given gene is always faster than the repair of the nontranscribed strand.

Two proteins have been identified in humans, which are specifically involved in the coupling of repair to a stalled RNA polymerase, CSA, and CSB, but their precise functions are as yet unknown. The CSA protein (44 kDa) consists largely of multiple WD40 repeats, a feature commonly found in proteins that coordinate multiprotein complex assemblies. No

functional homologs of CSA have been identified in yeast. The CSB protein (168 kDa) and its functional yeast homolog Rad26 are related to the SWI/SNF family of ATP-dependent chromatin remodelers. This remodeling activity might serve to push the RNA polymerase back on its template (without dissociating it from its template), thereby allowing interaction of the repair factors with the damage site. The XPC protein is dispensable for TCR, most likely because the transcription bubble generated by RNA polymerase can be directly bound by the damage-verification factor(s) TFIIH and XPA. Strikingly, the XPC homolog from yeast, Rad4, is required both for GGR and for TCR. Apparently, in yeast, the transcription bubble does not allow direct binding of TFIIH and/or XPA.

See also: **Bioenergetics:** Spastic Paraplegia; **Molecular Biology:** DNA Mismatch Repair in Bacteria; Nonhomologous Recombination: Bacterial Transposons; Nuclear Organization, Chromatin Structure, and Gene Silencing; Sliding Clamps in DNA Replication: *Escherichia coli* β -Clamp and PCNA Structure; **Signaling:** Gi Family of Heterotrimeric G Proteins; mTOR and its Downstream Targets.

Further Reading

- Fousteri M and Mullenders LHF (2008) Transcription-coupled nucleotide excision repair in mammalian cells: Molecular mechanisms and biological effects. *Cell Research* 18: 73–84.
- Friedberg EC, Walker GC, Siede W, et al. (2006) *DNA Repair and Mutagenesis*, 2nd edn. Washington, DC: ASM Press.
- Gillet LC and Scharer OD (2006) Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chemical Reviews* 106: 253–276.
- Goosen N (2010) Scanning the DNA for damage by the nucleotide excision repair machinery. *DNA Repair* 9: 593–596.
- Sugasawa K (2009) Regulation of damage recognition in mammalian global genomic nucleotide excision repair. *Mutation Research: Fundamental and Molecular Mechanisms of Mutagenesis* 685: 29–36.

Olfactory Receptors

S I Korsching, University of Cologne, Cologne, Germany

Published by Elsevier Inc.

This article is reproduced from the previous edition, volume 3, pp. 149–154, © 2004, Elsevier Inc.

Glossary

Chemosensation Any chemical sense, such as olfaction or taste. Very ancient, occurs in bacteria where it is used for chemotaxis.

Glomeruli Globular neuropil, which comprises synapses of olfactory receptor neurons.

Heptahelical A receptor with seven transmembrane helices, often, but not exclusively a G-protein-coupled receptor.

Odorant A chemically pure odor, that is, a single compound that activates an odorant receptor.

Olfactory code The neuronal representation of odors insofar as it is relevant to the encoding of these odors. Often confused with the neuronal representation, which is more commonly studied.

Pheromone An odorant that mediates intraspecies communication, eliciting neuroendocrine changes and stereotyped behavioral repertoires.

Olfactory receptors mediate the primary interaction of the olfactory brain with the external world. Any odor stimulus is initially represented as activation of one to many different olfactory receptors. Vice versa, anything that binds and activates an olfactory receptor is as per definition an odor, or odorant, as the single compounds are often called. Olfactory receptors constitute one of the largest families of G-protein-coupled receptors. Many large and small clusters of olfactory receptors are distributed throughout the genome. Olfactory receptors are expressed in a highly specific manner and monogenic expression is the general rule, that is, one neuron–one receptor. The molecular receptive range has been analyzed for several olfactory receptors. Generally, olfactory receptors have been found to exhibit a somewhat relaxed specificity toward ligands. However, some receptors, in particular, those for pheromones, show very high specificity. Olfactory receptors signal through specialized trimeric G proteins to open calcium-permeable nonspecific cation channels.

Olfactory Receptors Are G-Protein-Coupled Heptahelical Receptors

From a modest origin in simple eukaryotes such as yeast, G-protein-coupled heptahelical receptors (**Figure 1**) have evolved into the largest superfamily known. As the name indicates, signaling occurs through (trimeric) G proteins, with particular families making use of different G proteins. These receptors transmit biological effects of many hormones,

neurotransmitters, chemokines, and diverse sensory stimuli. Chemosensation constitutes an ancient function for this family. Already in yeast a dichotomy between sensing nutrients and sensing pheromones is found.

Whereas simple animals such as nematodes possess a single chemosensory sense, in higher animals several specialized systems have evolved, using olfactory and gustatory receptors, respectively. Olfactory receptors may recognize volatile or water-soluble odorants, depending on the species (e.g., mammals vs. fish), and the defining difference to taste receptors appears not to be the nature of the ligand, but the outreach of the system. Taste is generally considered to be a contact sense, whereas olfaction delivers long-range information.

Olfactory Receptors Segregate in Several Highly Divergent Groups

Thirteen years ago, the first higher eukaryote chemosensory receptors – mammalian odorant receptors (ORs) proper – have been identified. Linda Buck, then a postdoctoral fellow in Richard Axel's laboratory, used short motifs conserved in previously known heptahelical receptors to amplify a group of rat ORs from nasal complementary DNA via polymerase chain reaction. Like all heptahelicals, ORs exhibit the canonical seven-transmembrane domain structure (**Figure 1**) and the G-protein-interaction motif. Peculiar are the relatively short N and C termini. OR are a highly divergent group, and amino acid homologies below 20% have been reported within one

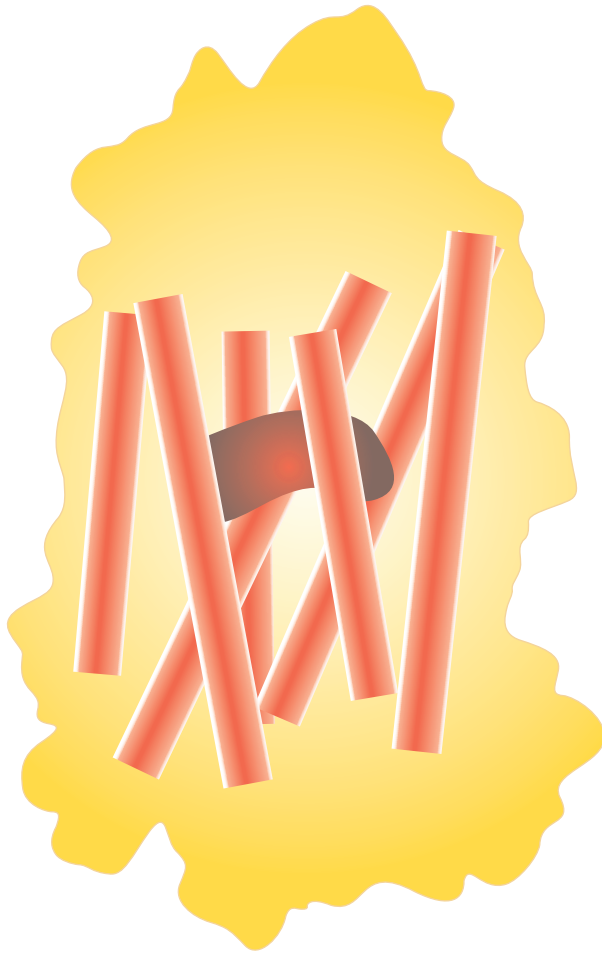


Figure 1 Schematic representation of an olfactory receptor. No crystallographical data are available for olfactory receptors. It is assumed that their structure is analogous to that of rhodopsin and the adrenergic receptor, which are more extensively studied. This is supported by some mutational analysis and molecular dynamics modeling. The scheme shown here is based on the structure of rhodopsin, whose crystal structure has been solved in 2000. Seven-transmembrane regions are depicted as cylinders. The binding site for ligands is centrally situated, between several transmembrane domains.

species. Between phyla, homologies have been so low as to be undetectable. Thus, in insect and nematode, chemosensory receptors have not been identified via homology searches, but by looking for orphan heptahelical receptors in either single-cell libraries or genomic databank searches.

Two classes of distantly related receptors have been found in the mid-1990s. They are designated V1R and V2R family, after their occurrence in the mammalian vomeronasal or accessory olfactory system. V1R receptors share the short N and C termini with the OR receptors, whereas the V2R exhibit a large extracellular N-terminal domain, and show some sequence homology to the Ca^{2+} -sensing receptor and the metabotropic glutamate receptor, a neurotransmitter receptor. In mice, the V2R family comprises ~100 genes and the V1R family contains ~150 genes, which segregate into 12 subfamilies designated V1Ra, V1Rb, etc. Humans and other hominoids have nearly no VR receptors as their accessory olfactory system is derelict and

their VR genes suffer ongoing pseudogenization. No intact human V2R is remaining and less than a handful of intact V1R genes have been identified in extensive genome searches.

It is commonly assumed but not rigorously shown that OR deal with common odors whereas VR are specialized for detecting pheromones – odors which are relevant in a social context as in kin recognition and mating.

It should be mentioned that there may be another, radically different group of olfactory receptors, a type of membrane-bound guanylate cyclases with a single transmembrane domain. Members of this family are expressed in specific subsets of olfactory receptor neurons. However, no ligands have been identified for them so far.

Gene Structure and Genomic Arrangement of Olfactory Receptors

Vertebrate OR and V1R feature an intronless coding region of ~1 kb that may be preceded by one to several noncoding introns. Alternative splicing is a frequent phenomenon for OR. Insect olfactory receptors and vertebrate V2R contain several introns within the coding region, as is common for other G-protein-coupled receptors.

The OR family size approaches 1300 identified genes in mouse, a species with a completely sequenced genome. About one-fifth are pseudogenes, a high percentage, which has people speculating about a putative role of the pseudogenes in the evolution of the gene family. The human OR gene family is of similar size; however, the percentage of pseudogenes is much higher, so that only ~350 OR appear functional in humans. Clusters of related genes are a frequent feature and indicate recent expansions of the respective gene lineages. Sometimes, other genes may be interspersed, a notable example being the intercalation of major histocompatibility complex (MHC) genes with OR reported for the mouse (but not for the human) genome. Overall, olfactory receptors are widely spread throughout the genome, asserting the ancient origin of this gene family.

Family sizes are smaller for lower vertebrates and insects – for *Drosophila* 60 receptor genes have been identified in the complete genome, mosquitoes possess ~80 olfactory receptor genes. For fish estimates range ~100 different OR. However, the occurrence of (smaller) clusters and the overall widespread distribution of receptor genes are equivalent to higher vertebrates. Surprisingly, the nematode *Caenorhabditis elegans* boasts ~1100 different chemosensory receptor genes (including 30% pseudogenes).

Monogenic and Monoallelic Expression of Olfactory Receptor Genes

Nematode chemosensory neurons, like taste receptor cells, express many different receptor genes per cell. Such bundling of receptors may be efficient if many different chemicals should lead to very few behavioral outcomes – mainly the choice between stay here and get out of here or swallow this and spit it out, respectively. As soon as a chemical sense is used for more sophisticated distinctions, it is imperative that more selective means of expression are used.

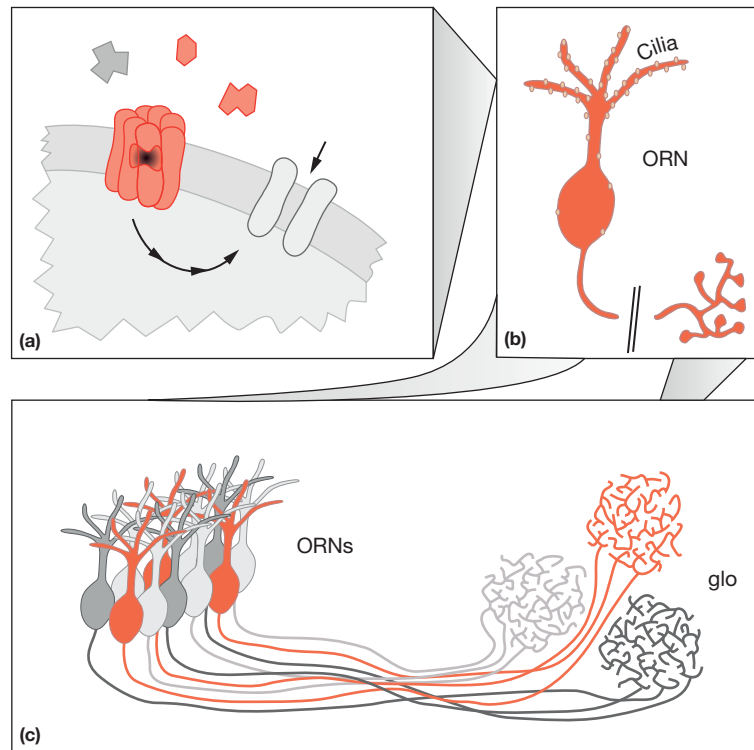


Figure 2 Olfactory receptors are part of the olfactory information processing pathway. (a) Olfactory receptors bind odorants and initiate a several-step signal-processing cascade. As endpoint, various channels are opened, leading to depolarization of the olfactory receptor neuron. (b) The olfactory receptors are mainly expressed on the cilia (or microvilli, not shown) of olfactory receptor neurons. These neurons generate action potentials in their axons and propagate them toward their terminals in the so-called glomeruli (glo). Glomeruli are neuropil structures and constitute the first relay station (synapse) in the olfactory-processing pathway. (c) Ciliated olfactory receptor neurons expressing the same odorant receptor are distributed within the sensory surface, but converge onto a common glomerulus. Microvillar receptor neurons such as those expressing the *V1R* and *V2R* genes converge onto several microglomeruli (not depicted here).

Accordingly, the expression of vertebrate and insect olfactory receptors is highly regulated. One particular olfactory receptor neuron expresses only a single type of OR gene or VR gene (Figure 2). This very restricted expression has been christened ‘monogenic expression’. In fact, for rodents it has been shown that expression is restricted even to one allele of a particular OR or VR gene, reminiscent of the monoallelic expression of immunoglobulin genes.

Some exceptions occur to the rule of one neuron–one gene. Notably a particular insect olfactory receptor, OR83b, has been found to be coexpressed in about two-thirds of all olfactory receptor neurons. This receptor is also highly unusual in that it is strongly conserved between many insect species. It is suspected that such atypical receptors may not have a role in odor detection, but serve another, so far unknown function.

Olfactory Receptor Gene Expression is Regulated by Proximal and Distal Control Regions

A stringent test for specificity of expression relies on the convergence of olfactory receptor neurons expressing the same receptor gene into a single glomerulus at a roughly fixed position (Figure 2). In several cases, such a restricted expression pattern could be reproduced faithfully by reporter gene

constructs containing variable segments of DNA flanking the coding region. It has been possible to narrow down control elements to regions of one to a few kilobases, but specific enhancer sites, transcription factors, or combinations thereof are just beginning to be identified. In some cases, all control elements appear to be situated close to the gene, whereas for others OR essential regulatory elements have been found in distances of ~100 kb from the coding region.

Selection of an OR gene for expression is a multilevel process. First, a choice is made for a particular subgroup of genes. Such subgroups feature spatially or temporally limited expression. In a second step, a particular OR is selected in a stochastic manner and then exerts negative feedback regulation onto all other ORs. Thus, monogenic expression of ORs is ensured.

Ligand Spectra Have Been Established for a Handful of Olfactory Receptors

Although hundreds of olfactory receptors have been cloned to date, ligands are only known for a small percentage of them. The difficulty of guessing potential ligands from among the enormous pool of potential odorants has been a major problem. Also, correct expression in heterologous systems has turned out to be unusually complicated, and has only been



Figure 3 Two types of olfactory processing: combinatorial vs. monospecific. To the left, the combinatorial representation of odorants is depicted. Different odors – here symbolized by various odor sources (skunk cabbage, skunk, rose, and lily) – activate overlapping, but distinctive combinations of olfactory receptors. This is also valid for the individual odor components, the odorants. Each odorant activates a particular subset of all olfactory receptors and thus each odorant generates a unique, if overlapping, fingerprint of olfactory receptor activation. To the right, monospecific representation of odorants is shown. A single compound activates a highly specific receptor very selectively, as is often observed for pheromones. Depicted here is a female silk moth, source of an attractive pheromone that is detected with extremely low threshold by the male of that species.

possible for some receptors. Most of the receptor molecules synthesized in heterologous systems are not transported to the plasma membrane, but remain in the endoplasmic reticulum, possibly because some unknown auxiliary molecules are missing. Indeed, for nematode chemosensory receptors, a chaperone has been shown to be essential for proper protein translocation, and V2R receptors are coexpressed with MHC class 1b molecules that appear to serve as escort molecules. For two OR genes, it has been possible to use a viral transfection method to express the recombinant OR in olfactory receptor neurons. In several cases, knockouts of particular olfactory receptors have served to identify receptor ligands, as first shown for the *C. elegans* ODR10 diacetyl receptor. The binding of ligands is not measured directly, but evaluated by a downstream parameter such as a change in membrane potential, or, more commonly, a rise in intracellular calcium levels.

Usually olfactory receptors respond best to a small group of highly related compounds and with lower sensitivity to further odorants. Often a particular functional group is preferred, but in some cases ligands with other functional groups may also activate the same receptor. Many strong ligands identified so far contain an aldehyde or keto group. The most extensive ligand analysis has been performed for the rat OR I7, which has *n*-octanal as the main ligand. Hydrocarbon chain length is another molecular feature that influences ligand efficacy. In the vast majority of cases, longer hydrocarbon chains result in more avid binding. From psychophysical experiments, it is known that also the shape and rigidity of the ligand molecule determine the olfactory sensation. Enantiomers of known ligands often, but not always, show differential responses. Taken together, a particular olfactory receptor requires a particular combination of molecular features in its cognate ligand/s.

Combinatorial versus Monospecific Representation of Odorants

Due to the monogenic expression of olfactory receptors, the ligand specificity of the olfactory receptor determines the odor

responses of the corresponding olfactory receptor neurons (Figure 2). Due to the convergence of olfactory neurons expressing the same receptor in single glomeruli (Figure 2), the ligand specificity of the olfactory receptor determines the odor responses in these glomeruli. This results in a rather direct map of the ligand specificities of the olfactory receptor repertoire on a two-dimensional surface in the corresponding brain region (the olfactory bulb in mammals, the antennal lobe in insects). Interestingly, mammalian OR influence the choice of a target glomerulus of olfactory receptor neurons as shown by gene-swap experiments. Insect olfactory receptors, on the other hand, appear not to be involved in selection of the proper target site.

Two complementary principles of odorant representation emerged from ligand binding studies *in situ* and in heterologous systems. First, many odorants activate a small subset of the large repertoire of ORs. Vice versa, any particular OR gets activated by a handful of different compounds, not just by a single compound. This results in a unique combination of activated receptors for each individual odor compound (Figure 3). The theoretical coding capacity of such a combinatorial representation is immense and much higher than needed – even for millions of different odors. In fact, the neuronal representation is sparse – the large majority of receptors do not respond to any given odorant – and thus suboptimal with respect to coding capacity. Nevertheless, humans (trained perfumers, that is) can distinguish and identify several thousand different odors, more than 10-fold the number of the functional human OR repertoire.

An alternative principle of neuronal representation seems to be realized for a smaller group of chemicals, including many pheromones. The high sensitivity and specificity of several such olfactory receptors (observed in mammals, fish, and insects) appears to indicate a monospecific interaction. A well-known example is the male/female attraction in silk moth, *Bombyx mori* (Figure 3). To integrate labeled line coding with hard-wired behavior appears straightforward. The coding capacity of such a system is low – it equals the number of different olfactory receptors devoted to this strategy – but sufficient for the limited

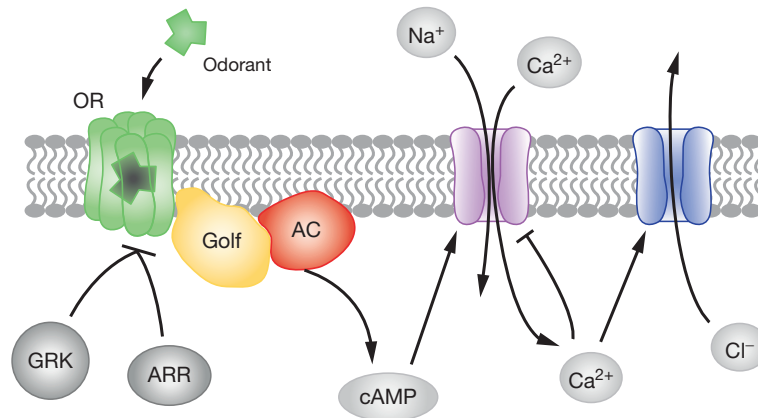


Figure 4 Signal transduction and termination of odor responses. A schematic representation of the main signaling pathway. AC, adenylate cyclase type III; ARR, β -arrestin-related molecules; G_{olf} , an α -subunit somewhat specific for the olfactory nervous system; GRK, G-protein-coupled receptor kinase; OR, odorant receptor. The cyclic nucleotide-gated channel is permeable for sodium and calcium, but gets blocked by calcium. GRKs phosphorylate the odorant receptor, which subsequently binds β -arrestin, leading to a complete stop of the signal transduction cascade.

number of potential pheromones. An example for a receptor/ligand pair is the mouse V1Rb2 receptor that specifically recognizes 2-heptanone.

Signal Transduction of Olfactory Receptors

All heptahelical olfactory receptors signal through trimeric G proteins. A controversy surrounding the signal transduction by OR has been more or less resolved. Vertebrate OR signal through G_{olf} , a specialized α -subunit, which in turn activates adenylate cyclase type III (Figure 4). The resulting cyclic adenosine monophosphate (cAMP) activates a cyclic nucleotide-gated channel that is a rather nonselective cation channel. Such channels are widespread in the nervous system, their particular gating properties are set by the subunit composition (A2A4B1b in the case of the mammalian olfactory channel). A considerable portion of inward current is carried by calcium (Figure 4). The resulting influx of calcium opens a calcium-gated chloride channel, resulting in chloride efflux, that is, in further depolarization and thus amplification of the signal (mammals, amphibians, and fish). Finally, the odor-induced signal is switched off by G-protein-coupled receptor kinase 3 (GRK3) and binding of arrestin-related molecules to the phosphorylated OR (Figure 4). Even with a constant odor stimulus the olfactory sensation is downregulated within seconds, in part, due to the peripheral mechanisms mentioned here.

The signal transduction cascade has not been worked out as well in invertebrates. Insects appear to also use cAMP as second messenger. Lobster may additionally employ IP₃ as a second signal transduction system that opens IP₃-gated channels situated in the plasma membrane. Such channels have been described in olfactory receptor neurons of many species, including fish and rodents, but appear to be a minor constituent compared to the number of cyclic nucleotide-gated channels. The effects described for IP₃ on olfactory signal

transduction in mammals and fish may reflect a modulatory role of this second messenger, conceivably in setting the basal activity level of olfactory receptor neurons. When all the evidence is weighed, a completely independent second signaling pathway for ORs appears not plausible.

Signal transduction for olfactory receptors of the VR type differs both from OR and between V1R and V2R. The emerging picture is that V1R signal via the G_{12} α -subunit and V2R via the G_o α -subunit. Finally, a transient receptor potential channel, TRP2, is activated by diacylglycerol and allows the influx of cations including calcium.

See also: [Signaling: Cyclic Nucleotide-Regulated Cation Channels; G-Protein-Coupled Receptor Kinases and Arrestins; Taste Receptors.](#)

Further Reading

- Breer H (2003) Olfactory receptors: Molecular basis for recognition and discrimination of odors. *Analytical and Bioanalytical Chemistry* 377: 427–433.
- Crasto C, Singer MS, and Shepherd GM (2001) The olfactory receptor family album. *Genome Biology* 2: 1027.
- Halpern M and Martinez-Marcos A (2003) Structure and function of the vomeronasal system: An update. *Progress in Neurobiology* 70: 245–318.
- Keller A and Vosshall LB (2003) Decoding olfaction in *Drosophila*. *Current Opinion in Neurobiology* 13: 103–110.
- Kim J and Carlson JR (2002) Gene discovery by E-genetics: *Drosophila* odor and taste receptors. *Journal of Cell Science* 115: 1107–1112.
- Korsching S (2002) Olfactory maps and odor images. *Current Opinion in Neurobiology* 12: 387–392.
- Nagao H, Yamaguchi M, Takahashi Y, and Mori K (2002) Grouping and representation of odorant receptors in domains of the olfactory bulb sensory map. *Microscopy Research and Technique* 58: 168–175.
- Ronnett GV and Moon C (2002) G proteins and olfactory signal transduction. *Annual Review of Physiology* 64: 189–222.
- Touhara K (2002) Odor discrimination by G protein-coupled olfactory receptors. *Microscopy Research and Technique* 58: 135–141.
- Young JM and Trask BJ (2002) The sense of smell: Genomics of vertebrate odorant receptors. *Human Molecular Genetics* 11: 1153–1160.

O-Linked GlcNAc Biosynthesis and Function

K Sakabe and G W Hart, Johns Hopkins University, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Background

The post-translational modification of nuclear and cytoplasmic proteins on Ser and Thr residues by the addition of a single sugar moiety β -N-acetylglucosamine (O-GlcNAc) through an O-glycosidic linkage was originally described in 1984. The dogma at that time limited covalent modification of proteins by carbohydrates to occur solely in the extracellular space and within the ER, Golgi, and other subcellular organelles. Unlike the other types of glycosylation, O-GlcNAc is usually not elongated to more complex structures, making it a unique sugar modification. Since its discovery, O-GlcNAc has been found in some bacteria, some protozoa, and all metazoans.

The addition of O-GlcNAc to proteins, termed O-GlcNAcylation, is analogous to phosphorylation in its use as a signaling intermediate and regulatory mechanism. Indeed, global changes of O-GlcNAc occur during different stages of cell cycle, as well as upon mitogen stimulation. Like phosphorylation, O-GlcNAc is abundant and dynamically cycles on and off polypeptides. Because phosphorylation and O-GlcNAcylation can compete for the same or for adjacent Ser and Thr residues on a protein, these two post-translational modifications display a complex relationship: (1) phosphorylation and O-GlcNAcylation can compete for the same residue, which can determine protein function, and (2) the presence of O-GlcNAc or O-phosphate on one site can influence the occupancy of an adjacent site, the adjacent site being near either spatially in three-dimensional space or on the peptide sequence level. In addition, the enzymes that control O-GlcNAc cycling are regulated by phosphorylation, and some of the enzymes controlling phosphate cycling are regulated by O-GlcNAcylation.

O-GlcNAcylated proteins represent many different functional classes, indicative of the variety of roles that O-GlcNAc plays within the cell. Virtually, every cellular process is represented as follows: structural proteins such as keratins and intermediate filaments, signaling molecules including kinases and adaptor proteins, the translational machinery, the transcriptional machinery including coactivators and RNA polymerase II, the proteasomal degradation machinery, metabolic enzymes, chaperones, and nuclear pore proteins.

While many different kinases and phosphatases are found within the cell, there is only one set of O-GlcNAc cycling enzymes. UDP-GlcNAc:polypeptide transferase or O-GlcNAc transferase (OGT) catalyzes the transfer of the N-acetylglucosamine from uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) to the hydroxyl group of specific Thr and Ser residues. β -D-N-acetylglucosaminidase or O-GlcNAcase, on the other hand, hydrolyzes the sugar moiety from the protein. Despite having only a single catalytic subunit, both of these enzymes are specifically targeted to their substrates by forming transient interactions with many other proteins to form a functional holoenzyme complex.

Biosynthesis of UDP-GlcNAc

The activated sugar donor, UDP-GlcNAc, is an example of a molecule exquisitely controlled by different pathways of cellular metabolism, including nucleotide, glucose, amino acid, and fatty acid metabolism. In addition, the formation of UDP-GlcNAc is sensitive to the overall energy status of the cell. The relative levels of each component determine the availability of UDP-GlcNAc for modifying different proteins, making this sugar donor an important sensor of overall nutrient status for the cell.

UDP-GlcNAc is synthesized via the hexosamine biosynthetic pathway. Approximately 2–5% of the glucose taken up by the cell is shunted through this pathway to generate amino sugars and activated amino sugars. The rate-limiting enzyme of the hexosamine biosynthetic pathway glutamine-fructose-6-phosphate-transaminase, or GFAT, controls how much glucose is diverted into sugar nucleotide synthesis (Figure 1).

O-GlcNAc Transferase

OGT is evolutionarily conserved from *Caenorhabditis elegans* to *Homo sapiens*, showing greater than 60% amino acid identity and 80% amino acid similarity between species. In fact, amino acid identity is around 99% when comparing sequences from higher-order species such as human, mouse, and rat. This degree of sequence conservation between species suggests an evolutionary constraint in maintaining primary sequence and tertiary structure. Further highlighting the importance of this enzyme, ablating this gene is lethal for mice, fruit flies, and plants, demonstrating the essential nature of this post-translational modification.

Although the predominant form of OGT is an enzyme of 110 kDa (ncOGT), two other isoforms of OGT have been identified, arising from different splice variants: a 103 kDa isoform targeted to the mitochondria (mOGT) and a 78 kDa short form of OGT (sOGT; Figure 2). While the C-terminal catalytic portion of the enzyme is identical between isoforms, they differ in the N-terminal tetratricopeptide repeat (TPR) domain. TPRs are structural motifs comprised of 34 amino acids that mediate protein–protein interactions or multimeric complex formation. TPRs are typically found in tandem repeats. OGT is no different in that the number of repeats ranges from 2, in the short isoform, to 13, for the full-length protein. mOGT also contains a mitochondrial localization sequence. In addition, OGT contains a binding domain for phosphatidylinositol 3,4,5-triphosphate (PIP3), a signal intermediate generated upon activation of growth factor receptors, allowing OGT to associate with the phospholipid bilayer of the plasma membrane in a signal-dependent manner.

No consensus sequence is known for OGT; however, a Pro residue within one to three amino acids of the modification

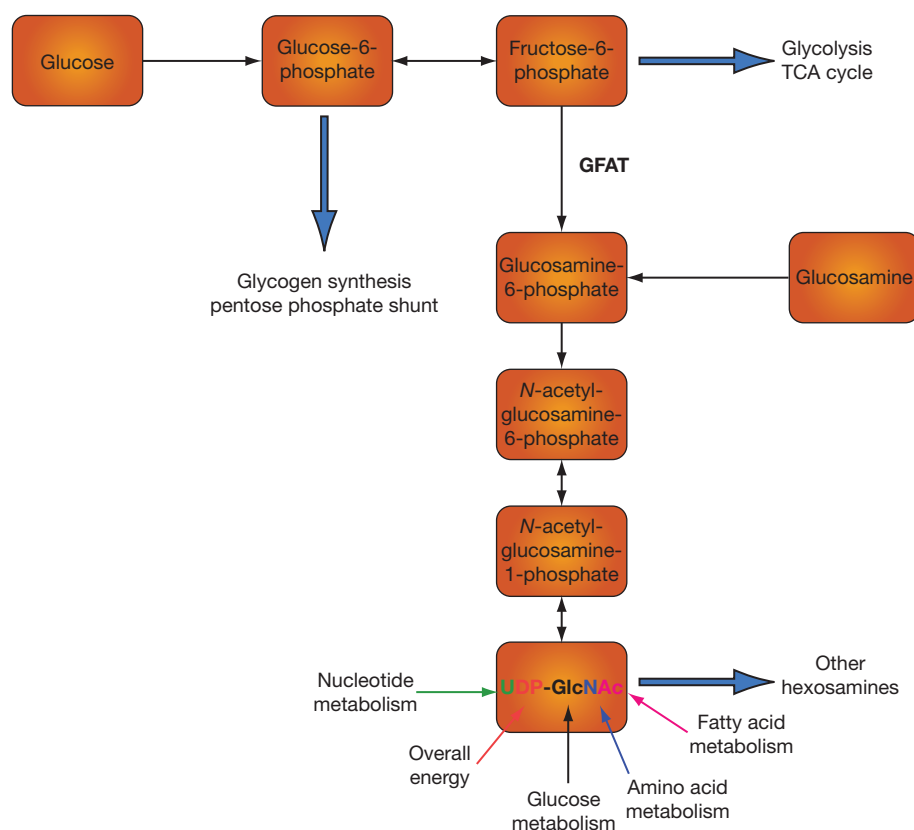


Figure 1 The hexosamine biosynthetic pathway. The diagram depicts how glucose is metabolized into UDP-GlcNAc.

site occurs frequently. Recent structural analyses have also provided insight into how the two domains of OGT function to O-GlcNAcylate substrates. A composite structure of the TPR domain of human OGT with a bacterial ortholog of OGT, *Xanthomanas campestris* OGT (XcOGT), reveals a corkscrew-like structure with the TPRs forming an extended superhelix and the catalytic portion adopting a typical nucleotide sugar-dependent glycosyltransferase fold. The inner portion of the TPR superhelix is lined with Asn residues, forming potential sites for interacting proteins. Interestingly, the TPR proximal to the catalytic domain faces the active site, thereby creating a surface by which substrates can interact. This conformation leaves the remainder of the TPR domain available for protein binding. Sequential deletion of the TPRs reduces the ability of OGT to O-GlcNAcylate protein substrates while retaining its ability to modify short peptide substrates. These observations suggest that OGT interacts with its substrate through its TPR domain and that the TPRs form a platform that can recruit target molecules to OGT. Moreover, the majority of different targeting proteins, or proteins that direct OGT to O-GlcNAcylate substrates specifically in a signal-dependent manner, interact with OGT through the TPR domain. This model explains how OGT is able to O-GlcNAcylate many different types of proteins and, in fact, this model is reiterated throughout nature as phosphatases use targeting subunits to direct them to specific substrates.

O-GlcNAcase

The enzyme that catalyzes the hydrolysis of the O-GlcNAc moiety, O-GlcNAcase, is a monomeric enzyme of 916 amino acids. The N-terminus contains the catalytic activity with a putative acyltransferase domain on its C-terminus. A shorter isoform resulting from an alternative splicing site yields a truncated enzyme lacking approximately 230 amino acids from its C-terminus and is identical to the full length enzyme except for the last 15 amino acids (Figure 3). Both forms are expressed ubiquitously.

Purification of O-GlcNAcase from cow brain shows that O-GlcNAcase is found within a high-molecular-weight complex suggesting an association with other proteins for regulation and specificity. In *in vitro* O-GlcNAcase assays, O-GlcNAcase can cleave GlcNAc from glycopeptides and from the chemical substrate paranitrophenyl-GlcNAc, but not efficiently from proteins, indicating that auxiliary factors are also required by O-GlcNAcase for substrate recognition.

In addition, O-GlcNAcase is a substrate for proteolytic cleavage by one of the executioner caspases, that is, Caspase-3. Upon initiation of apoptosis, O-GlcNAcase is cleaved between the catalytic domain and its C-terminus. These two halves remain associated and retain activity, so it is still unclear why O-GlcNAcase is cleaved during apoptosis.

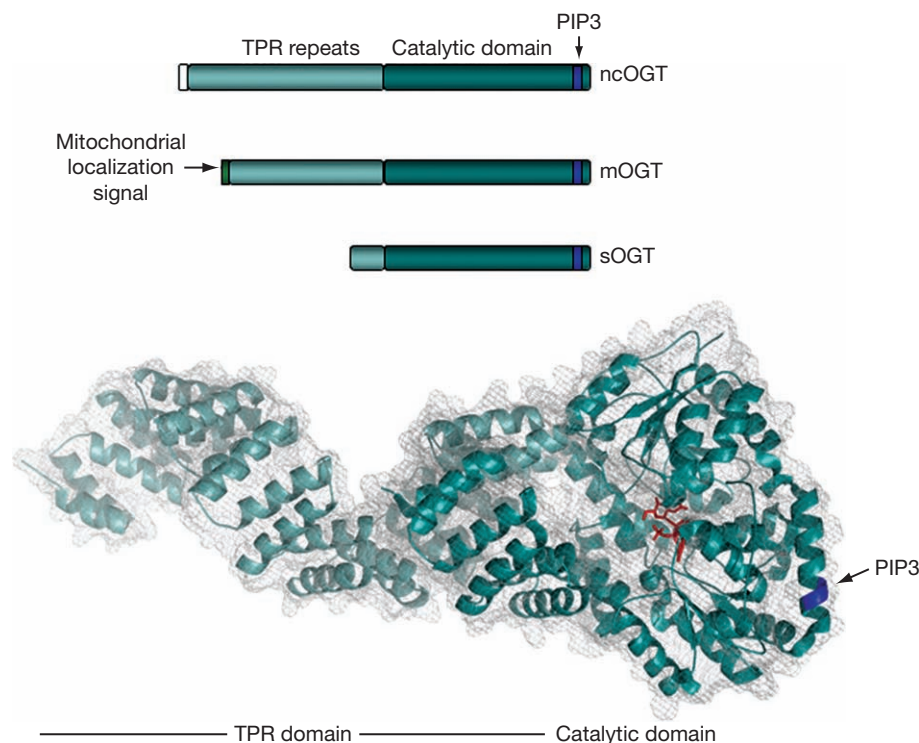


Figure 2 The structure of OGT. The upper portion depicts the domain architecture of the three OGT isoforms. The lower portion is a composite image for the TPR domain of human OGT overlaid (PDB:1W3B) with the bacterial ortholog XcOGT in complex with UDP-GlcNAc phosphonate analog in red (PDB:2JLB). The dark blue indicates the location of the PIP3-binding domain. The vast alignment algorithm was used to align the human OGT and XcOGT. MacPymol was used to create structural images.

Function and Physiological Importance

Modification with O-GlcNAc leads to a variety of outcomes for the protein, including regulation of enzyme activity, protein–protein interaction, and altered degradation. A variety of functions will be discussed below in different cellular contexts.

One of the largest classes of proteins modified with O-GlcNAc are components of the transcriptional machinery, such as transcription factors, chromatin remodeling enzymes, as well as RNA polymerase II itself. O-GlcNAcylation can affect the ability of certain transcription factors to interact with the transcriptional machinery, such as STAT5A, which requires O-GlcNAcylation for interaction, or cAMP response element-binding protein (CREB), where O-GlcNAcylation decreases affinity to basal transcription factors. In addition, it can affect the ability of a transcription factor to bind its cognate sequence, as in the case of PDX-1, where O-GlcNAcylation enhances its DNA-binding affinity. O-GlcNAc can also regulate chromatin architecture by modifying chromatin remodeling enzymes. For example, O-GlcNAcylation of MLL5 activates its histone lysine methyltransferase activity, which in turn activates transcription from certain promoters. Finally, O-GlcNAcylation can affect the subcellular location of some transcription factors. By determining nuclear or cytoplasmic localization, O-GlcNAc can enhance or repress transcription, respectively.

O-GlcNAc also plays a role in regulating protein degradation. For example, O-GlcNAcylation can protect proteins from degradation, such as for the transcription factors like the

estrogen receptor, the oncogene c-myc, and the tumor suppressor p53. Many proteins of the proteasome are O-GlcNAcylated, and increasing O-GlcNAc levels on the proteasome inhibits its adenosine triphosphatase (ATPase) activity, thereby reducing proteasome-mediated degradation.

The ability of the cell to tolerate various insults is mediated by the stress response pathway. Global O-GlcNAc levels increase during stress, and preventing this increase decreases cell viability. The protective effect of O-GlcNAc during stress is partly through the upregulation of heat shock proteins.

Misregulation of O-GlcNAc occurs in numerous disease pathologies such as type II diabetes, Alzheimer's disease, and cancer. Hyperglycemia and insulin resistance are hallmarks of type II diabetes. Since generation of UDP-GlcNAc is sensitive to the amount of glucose and glucosamine available, hyperglycemic conditions are believed to increase UDP-GlcNAc levels within the cell and thus increase the amount of O-GlcNAcylated proteins. Increasing O-GlcNAc levels genetically or pharmacologically causes insulin resistance; therefore, OGT appears to play a key role in the pathology of type II diabetes.

A major component of neurofibrillary tangles associated with Alzheimer's disease is the microtubule-associated protein tau. In the diseased state, tau becomes hyperphosphorylated. Under normal conditions, tau is modified with O-GlcNAc; it is hypothesized that by decreasing the amount of this modification on tau, perhaps due to defective glucose metabolism in the neuron, tau becomes hyperphosphorylated and forms the aggregates associated with Alzheimer's disease.

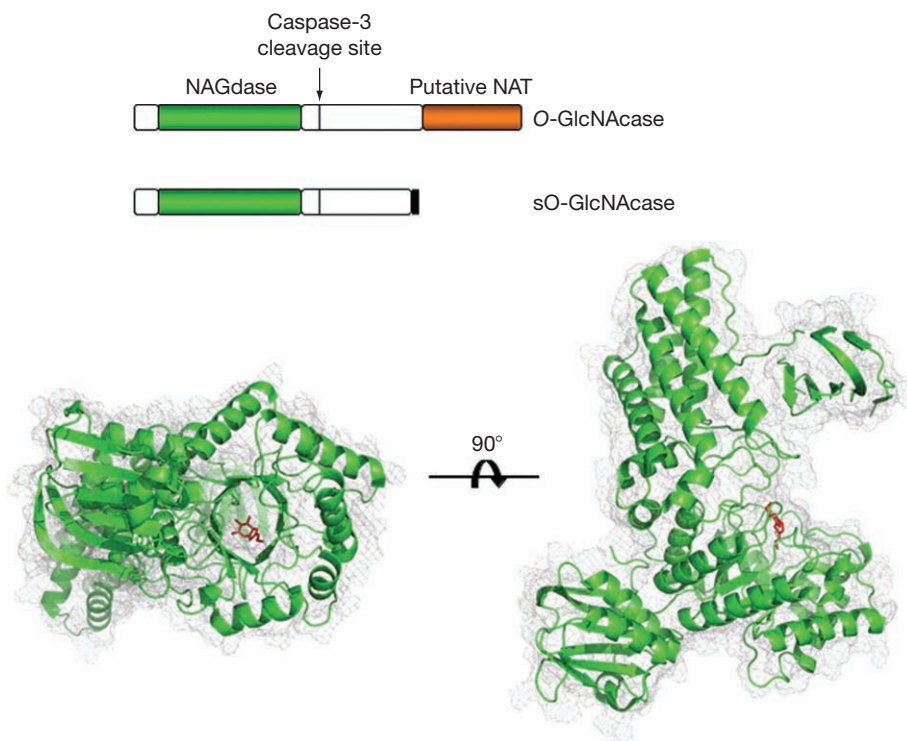


Figure 3 The structure of *O*-GlcNAcase. The upper portion shows the domain architecture of *O*-GlcNAcase. NAGdase domain contains the hexosaminidase activity. The orange domain is the putative acetyltransferase activity domain. The sO-GlcNAcase contains a unique 15-amino-acid sequence on its C-terminus, which is indicated in black. The lower portion of the figure shows the crystal structure of one of the bacterial orthologs for *O*-GlcNAcase in complex with the inhibitor Nag-thiazoline in red (PDB:2J4G). The ortholog contains only the hexosaminidase domain. The left-hand portion is looking down the catalytic domain while the right-hand portion is a 90° rotation on the *x*-axis. MacPymol was used to generate structural images.

Finally, a decrease in *O*-GlcNAc has been seen in cancerous tissue. In fact, increased *O*-GlcNAcase activity and decreased *O*-GlcNAc levels correlated with tumor grade. In addition, Thr 58 of *c*-Myc, an oncoprotein, is the major *O*-GlcNAc site on this oncogene protein and is also a mutational hot spot in Burkitt's lymphoma. In quiescent cells, *c*-Myc is mainly *O*-GlcNAcylated at Thr 58, but this site becomes rapidly phosphorylated upon growth stimulation.

In summary, *O*-GlcNAc is an abundant and dynamic post-translational modification in many organisms, including all metazoans and viruses that infect them. It has a complex interplay with phosphorylation and in turn plays a key role in regulating cell signaling, transcription, and cytoskeletal functions.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Glycoprotein-Mediated Cell Interactions, Nonmucin-Type O-Linked; Glycoproteins, N-Linked; Glycoproteins, Plant; Mucin Family of Glycoproteins; Proteoglycans; Siglecs; Sugar Nucleotide Transporters; **Metabolism Vitamins and Hormones:** Gluconeogenesis; Glucose/Sugar Transport in Mammals; Glycolysis Overview.

Further Reading

- Chatham JC, Not LG, Fulop N, and Marchase RB (2008) Hexosamine biosynthesis and protein O-glycosylation: The first line of defense against stress, ischemia, and trauma. *Shock* 29: 431–440.
- Copeland RJ, Bullen JW, and Hart GW (2008) Cross-talk between GlcNAcylation and phosphorylation: Roles in insulin resistance and glucose toxicity. *American Journal of Physiology – Endocrinology and Metabolism* 295: E17–E28.
- Hanover JA, Krause MW, and Love DC (2010) The hexosamine signaling pathway: *O*-GlcNAc cycling in feast or famine. *Biochimica et Biophysica Acta* 1800: 80–95.
- Hart GW, Housley MP, and Slawson C (2007) Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* 446: 1917–1022.
- Lazarus BD, Love DC, and Hanover JA (2009) *O*-GlcNAc cycling: Implications for neurodegenerative disorders. *International Journal of Biochemistry and Cell Biology* 41: 2134–2146.
- Martinez-Fleites C, He H, and Davies GJ (2010) Structural analyses of enzymes involved in the *O*-GlcNAc modification. *Biochimica et Biophysica Acta* 1800: 122–133.
- Torres C-R and Hart GW (1984) Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. *Journal of Biological Chemistry* 259: 3308–3317.
- Zachara NE and Hart GW (2006) Cell signaling, the essential role of *O*-GlcNAc! *Biochimica et Biophysica Acta* 1761: 599–617.

Opioid Receptors

P Y Law and H H Loh, University of Minnesota, Minneapolis, MN, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

G-protein-coupled receptors (GPCRs) Membrane receptors that transduce their signals via the heterotrimeric G proteins consisting of α - and $\beta\gamma$ -subunits. Activation of these receptors results in the dissociation of GDP from and association of GTP to the α -subunit. The association of GTP causes the dissociation of the heterotrimeric protein into the α - and $\beta\gamma$ -subunits leading to subsequent effectors' activation or inactivation. The enzyme GTPase within the α -subunit will hydrolyze the GTP to GDP resulting in the reassociation of the α - and $\beta\gamma$ -subunits, thus terminating the signaling of the receptor.

Intronic sequences Sequences within the gene that are removed by RNA splicing to generate the final mature RNA products of a gene that could include those that generate proteins, ribosomal RNA, or transfer RNA.

Mixed agonist–antagonists Opioid ligands that could have full agonist activity in one opioid receptor type but also act

as antagonist in another opioid receptor type. Therefore, their pharmacological profiles resemble the activation of one receptor type and inactivation of another receptor type.

Regulator of G-protein signaling (RGS) Multifunctional, GTPase activating protein or protein structural domains that promote GTP hydrolysis by the α -subunit of the heterotrimeric G proteins, thereby accelerating the termination of the GPCR signaling.

β -Arrestins Cellular proteins involved in the dampening of the cellular responses as a result of agonist-induced activation of the GPCR. Agonist-induced activation of GPCR leads to phosphorylation of the receptor and facilitating the β -arrestin binding. The bound β -arrestin could attenuate the G-protein interaction with the receptor, thus preventing the signal transduction. But the bound β -arrestin could also serve as scaffold for other cellular proteins, such as MAP kinase or clathrin, thus resulting in the transduction of alternative signals and the endocytosis of the receptor.

Opioid receptors are transmembrane proteins located at the synaptic membranes that serve as targets of an efficacious pain relief drug: morphine. The identification of these binding sites at neural membrane for plant alkaloids has led to the discovery of the first class of endogenous drug molecules, that is, the enkephalins and endorphins. These opioid receptors belong to the superfamily of G-protein-coupled receptor (GPCR), and upon activation, the activities of multiple cellular effectors are modulated, leading to inactivation of the ascending and activation of the descending pain pathways. Eventually, pain perception is altered. Unfortunately, with the good therapeutic actions of morphine, there are also bad side effects of the drug, for example, tolerance and addiction. Thus, the molecular mechanism in the cellular regulation of these opioid receptors, and the designs of specific agonists for these opioid receptors to be ideal analgesic agents are of great interest in pain medicine development.

Historical Perspectives

For centuries, extracts from the opium poppy, *Papaver somniferum*, have been prescribed for pain relief, to control cough, and to treat diarrhea. However, in addition to its medicinal uses, due to its psychological effects, man has abused opium since the time of the ancient Greeks. The isolation of the principal alkaloid by the German pharmacist, Friedrich W Serturner in 1805, which he named morphine after the Greek god of dreams, Morpheus, has enabled pharmacologists to determine the relationships between structures and their analgesic activities. From the relatively low concentration of morphine needed to relieve pain in 70% of the patients (10 mg

70 kg⁻¹ body weight) to the stereoselective requirement for analgesic action of the drug, it is apparent that morphine acts by high affinity binding to proteins, or specific receptor, located in the central nervous system (CNS). The presence of a specific morphine receptor was not demonstrated until 1973, when three research groups – Lars Terenius, Solomon Snyder, and Eric Simon, following the guideline established by Avram Goldstein – independently reported the presence of high affinity and stereoselective binding sites at the synaptic membrane preparations of rodent brain. The affinities of various opiate analgesics for these binding sites and the location of these sites correlated well respectively with the *in vivo* potencies and sites of action of the drug. Thus, the era of the opiate receptor is launched. Immediately, it is apparent that these mammalian CNS binding sites could not be evolved for plant alkaloids. This dichotomy was resolved by the isolation of the endogenous opioid peptide, enkephalin, by John Hughes and Han Kosterlitz in 1975 followed later by the identification of the proenkephalin A gene. The subsequent identification of two other gene products, β -endorphin deriving from the pro-opiomelanocortin gene and dynorphin deriving from the *prodynorphin* or *proenkephalin B* gene, suggests that the *in vivo* actions of morphine and its congeners must be a reflection of these endogenous opioid peptides actions. The presence of multiple endogenous opioid peptides genes also suggests the probable existence of multiple opioid receptors.

Discovery of Multiple Opioid Receptors

In 1976, Bill Martin and his colleagues hypothesized the presence of multiple opioid receptors. Concluding from the

physiological responses such as blood pressure and respiration, Martin and co-workers coined the multiple opioid receptors terminology based on the prototypic ligands within each group, with morphine being the mu (μ) agonist, ketocyclazocine being the kappa (κ) agonist, and SKF10 047 being the sigma (σ) agonist. The existence of multiple opioid receptors was further demonstrated by measuring the amount of an opiate antagonist to reverse the agonist effect in two *in vitro* bioassay systems, the guinea pig (GP) ileum and the mouse vas deferens. These bioassays were responsible for the identification of the fourth type of opioid receptor, the delta (δ) with enkephalin being the prototypic agonist. Additional studies such as the cross-tolerance measurements also supported the existence of multiple opioid receptors. In the cross-tolerance studies, animals or the *in vitro* GP ileum were treated with an opioid receptor-selective agonist. Chronic exposure to the agonist resulted in a loss in response, or tolerance development. However, activities of agonists that are selective for other opioid receptors remained unaltered, or no cross-tolerance was developed. These *in vivo* and *in vitro* experiments support the original hypothesis of multiple opioid receptors.

The implication of the existence of multiple opioid receptors by the animals' studies and bioassays was supported initially by the direct receptor-binding experiments reported by Law and Loh, KJ Chang, Hans Kosterlitz, and their colleagues. The development of receptor-selective antagonists, such as naltrindole (δ) and nor-binaltorphimine (κ), and the development of ligands that would alkylate receptor selectively, such as β -funaltrexamine (μ), have enabled investigators to characterize the properties of these multiple opioid receptors biochemically further by generating membrane preparations that contain single-receptor type. Only after the molecular cloning of the δ -opioid receptor, concurrently by the groups of Chris Evans and Brigitte Kieffer in 1992, followed by the cloning of μ - and κ -opioid receptors in 1993 by others, is the hypothesis of multiple opioid receptors firmly established. These multiple opioid receptors are products of three separate receptor genes and with high amino acid sequence homology (>60%). Although subtypes of these opioid receptors have been postulated, various gene products that exhibit the reported ligand selectivity of the various opioid receptor subtypes have not been identified. Also, from additional pharmacological and the molecular cloning studies, the sigma (σ) receptor does not belong to the opioid receptor family. Hence, the μ -, κ -, and δ -opioid receptors are the sites where endogenous opioid peptides and exogenous opiate alkaloids exert their actions.

Distribution and Pharmacology of Multiple Opioid Receptors

Prior to the cloning of the opioid receptors, the distribution of the receptors within various brain regions or tissues was determined with autoradiographic studies using radioactive ligands selective for a specific receptor. These autoradiographic results were later substantiated and extended with the *in situ* hybridization studies using the cloned receptor messenger RNA (mRNA) and with the histochemical or immunofluorescence studies using the receptor-specific antibodies. In general, these data indicate that the location of the multiple opioid receptors

coincides with the pharmacological actions of the drugs. These opioid receptors are mainly localized in the limbic system that are involved in the control of emotion and reward behaviors, the ascending and descending pain pathways that include the different laminae layers of cortex, thalamus, periaqueductal gray, midbrain median raphe and the dorsal horn of the spinal cord, and specific brain regions that are known to control locomotion, emesis, cough, and temperature.

The distributions of the μ - and δ -opioid receptor are very similar but distinct. Although these receptors do not colocalize in the pre- or post-synaptic membrane of the same neuron, they are observed to be located in the same components of the limbic systems such as the nucleus accumbens and the amygdala. Their locations at the cortex, the 4th ventricle and at the substantia gelatinosa of the spinal cord are the reasons why both μ - and δ -opioid receptor-selective drugs have strong spinal and supraspinal analgesic effects. Similarities in the distribution of these two opioid receptors suggest that the μ - and δ -opioid receptors have a similar pharmacological profile (Table 1). There are reports using δ -opioid receptor-selective antagonist or using mice that the δ -opioid receptor has been genetically deleted that implicate the role of δ -opioid receptor in modulating both the acute and chronic effects of the μ -opioid receptor selective agonist.

On the other hand, distribution of the κ -opioid receptor in the CNS is distinct from that of μ - and δ -opioid receptors. The most striking difference is the uniform pre-synaptic location of the κ -opioid receptor in the caudate putamen area. Such location could be the explanation for the observed κ -opioid-agonist-induced decrease in locomotion and the decrease of dopamine release in this brain area. The inhibition of dopamine release is the basis for the dysphoric effects of the κ -opioid agonists. The brainstem and spinal cord distribution of the κ -opioid receptor resemble that of μ -opioid receptor. Hence, κ -opioid agonists could produce spinal analgesia (Table 1).

The identification of these three types of opioid receptor could not account for the various receptor subtypes reported by the pharmacological studies. Such discrepancy could stem from the nonselective nature of the ligands for the receptors, or that there might be multiple receptor molecules generated from the three opioid receptor genes. The coding sequences of opioid receptors are not continuous within their respective genes. There are gaps, or intronic sequences, within the receptor genes. Thus, it is possible during the maturation of the

Table 1 Pharmacological effects associated with the multiple opioid receptor

	μ	δ	κ
Analgesia	Supraspinal Spinal	Supraspinal Spinal	Spinal
Pupil constriction	++	++	–
Respiratory depression	+++	++	+
Diuresis	Antidiuresis	–	++
GI	Constipate	Constipate	–
Smooth muscle	Spasm	Spasm	–
Behavior/effect	Euphoria Sedation ++		Dysphoria Sedation +
Physical dependent	++	++	+

messenger RNA (mRNA), different portions of the gene are spliced together, splice variants, resulting in different receptor protein structures and pharmacological properties. It is also possible that these three opioid receptors could form oligomers. The hetero-oligomers, that is, oligomers formed with different opioid receptors, will definitely exhibit different pharmacological properties from the homo-oligomers, that is, oligomers formed with same opioid receptor, and from the monomer. Opioid receptor existing in oligomeric forms have been implicated in numerous studies in cell model studies. Whether the three opioid receptors do exist in oligomeric forms at the synapse remains one of the questions to be addressed.

The multiple opioid receptors are not confined to the CNS. The presence of the receptors in the gastrointestinal tract accounts for the constipative effect of μ - and δ -opioid agonists. The presence of the κ -opioid receptor in the cardiac myocytes could be the reason for the bradycardia effect observed with the κ -opioid agonists, though the bradycardia effect of the opioid agonists is mainly associated with the agonist actions at the brainstem medulla. Opioid receptors have been reported also to be present in the immune cells. The true identities of these receptors are yet to be identified, whether they are the μ -, δ -, or κ -opioid receptor types or the nonclassical opioid receptors. Nevertheless, the pronounced effects that morphine has in the immune responses of animals, and the fact that intravenous drug users are one of the high-risk groups for HIV have led to many studies, and the effects of various opioid agonists and antagonists on the functions of immune cells cannot be ignored.

Opioid Ligands Selectivity for the Multiple Opioid Receptors

With the cloning of the opioid receptors and the subsequent generation of mice in which the transcription of selective opioid receptor gene could be disrupted, it is possible now to delineate the opioid receptor types that various opioid ligands could activate. By expressing the three cloned receptors individually into cells that do not express opioid receptor endogenously, it is possible to generate models that could determine the affinities of the opioid ligands to a specific receptor type accurately. Likewise, by genetically eliminating the *in vivo* expression of an individual opioid receptor type, the loss or gain in functions of a specific opioid agonist can be evaluated *in vivo*. These data in combination with the receptor-selective antagonists studies have determined the selectivity of various opioid ligands and their functions, as summarized in Table 2.

It is apparent that morphine and its congeners are agonists in the μ -opioid receptor. Although in the *in vitro* cell line, models that express the δ -opioid receptor morphine has partial agonist activities, morphine does not exhibit measurable analgesic activities in mice that do not express the μ -opioid receptor. Since these animals have intact δ - and κ -opioid receptor systems, it is unequivocal that morphine pharmacological effects are mediated via the activation of μ -opioid receptor.

Other opiate alkaloids have been shown to exhibit nonselectivity toward these three opioid receptors. The oripavine derivatives (etorphine and diprenorphine) and the benzomorphans (ketocyclazocine and bremazocine) have been shown

Table 2 Relative receptor selectivity of various opioid ligands and their functions in respective receptors

Opioid ligands	μ	δ	κ
Bemazocine	AG	AG	AG
Buprenorphine	PA		AN
Butorphanol	PA		AG
Diprenorphine	AN	PA	AN
Ethylketocyclazocine		AG	AG
Etorphine	AG	AG	AG
Fentanyl	AG		
β -Funaltrexamine	AN		AG
Levorphanol	AG	PA	
Methadone	AG		
Morphine	AG	PA	
Nalorphine	AN	AN	AG
Naloxone	AN	AN	AN
Naltrexone	AN	AN	AN
Naltrindole		AN	
Nor-Binaltorphimine			AN
NTB		AN	
Pentazocine	PA		AG
Spiradoline			AG
Sufentanyl	AG		
U50,488			AG
U69,593			AG

AG, agonist; PA, partial agonist; AN, antagonist.

by many to have similar affinities for all three opioid receptors. This is surprising since ketocyclazocine was used by Bill Martin to define one of the multiple opioid receptors, the κ -opioid receptor. Such dichotomy may be resolved by the fact that some opiate alkaloids have different activities in different opioid receptor types. For example, nalorphine exhibits antagonistic properties in the μ - and δ -opioid receptors but possesses agonistic activities in the κ -opioid receptor. Another example is pentazocine that possesses both κ -agonistic and μ -antagonistic properties. The mixed agonist–antagonist properties of various alkaloids are being exploited in the clinic so as to minimize the side effects of the drug, most noticeably, the addictive liability of the opiate analgesics.

The endogenous ligands, noticeably the products of the three peptides genes – enkephalin, β -endorphin, and dynorphin do not appear to have specificity toward these multiple opioid receptors. Initially, enkephalin and β -endorphin have been considered to be the prototypic ligands for the δ - and μ -opioid receptors, respectively. Dynorphin, with the positively charged amino acids, arginine and lysine, has been shown to preferentially interact with the κ -opioid receptor. Since then, it has been shown that Met⁵-enkephalin has equal affinity for the μ - and δ -opioid receptor, and that dynorphin has high affinity for the δ -opioid receptor. Only when the peptides are restricted in their rotation is receptor selectivity observed, for example, the D-penicillamine derivatives of enkephalin, DPDPE. This peptide has been shown *in vitro* to be selective for the δ -opioid receptor and with minimal activity in the μ -opioid receptor. Hence, the *in vivo* effect of DPDPE has been the hallmark for the δ -opioid receptor activation. However, recent *in vivo* studies using mice lacking the μ - or the δ -opioid receptor, clearly suggest that the DPDPE analgesic effect is mediated by activation of

μ -opioid receptor. This conflicting *in vitro* and *in vivo* receptor selectivity of the endogenous peptides and their derivatives requires further detailed investigation.

Molecular Mechanism of Multiple Opioid Receptor Function

A general feature of the opioid agonist is the inhibition of neurotransmitter release. Although there are instances that opioid agonist could increase the neurotransmitter release, it is due to the location of the receptor and the disinhibition of the neurotransmitter release. An excellent example is the dopamine release within the nigrostriatal loop. κ -Opioid agonist inhibits dopamine release by acting directly at the presynaptic dopaminergic terminals, whereas μ - and δ -opioid agonists increase the release by inhibiting the GABAergic interneurons that inhibit the dopamine release. Thus, in either case, the activation of opioid receptor results in the direct inhibition of synaptic transmission.

From the cloning studies, it is unequivocal that the opioid receptors belong to the superfamily of proteins, the GPCRs and the subfamily of rhodopsin-like receptors. One common feature of these receptors is that they contain hydrophobic regions that would span the membrane 7 times. Thus, these receptors are commonly known as 'seven-transmembrane receptors'. Another common feature is that their signals are mediated by promoting the dissociation of the nucleotide guanosine diphosphate (GDP) from the α -subunits of the heterotrimeric proteins. The dissociation of GDP allows the association of guanosine triphosphate (GTP) to the α -subunits resulting in the dissociation of the heterotrimeric proteins into α - and $\beta\gamma$ -subunits. These subunits will then control the activities of various intracellular or membrane-bound proteins and enzymes. For the multiple opioid receptors, the GTP/GDP-binding proteins involved in their actions are the pertussis-toxin-sensitive Gi/Go proteins.

In order for opioid receptors to regulate the neurotransmitter release, two of the proteins that the receptors regulate are the voltage-dependent Ca^{2+} channels and the inward rectifying K^{+} channels. By inhibiting the Ca^{2+} channels and thus reducing the influx of Ca^{2+} , opioid agonists could regulate the fusion of vesicles containing the neurotransmitters with the synaptic membrane. Also opioid agonists could activate the voltage-dependent inward rectifying K^{+} channels and thus hyperpolarizing the synaptic membrane. By controlling these two ions channels, activation of the opioid receptors would reduce the excitability of the membrane and reduce the transmitter release.

Other prominent proteins and enzymes that are regulated by the multiple opioid receptors are the enzymes responsible for cyclic adenosine monophosphate (cAMP) synthesis, adenylyl cyclase, and that which is involved in the long-term cellular adaptational events, the mitogen-activated protein kinase (MAP kinase). Opioid receptors inhibit the adenylyl cyclase activity, thereby reducing the intracellular cAMP content. This second messenger could control the activities of a protein kinase that, in turn, could regulate the activities of other proteins, such as the ion channels. The MAP kinase has been known to participate in the cellular proliferation and

apoptotic events. Thus, it is probable that MAP kinase activation by the multiple opioid receptors could signify their involvement in long-termed potential and learning, two of the probable mechanisms for the basis of chronic drug actions.

The complexity of opioid receptor signaling lies not only in the integration of all the signals from these effectors activated by the receptor, but also in the signal duration, intensity, and the pathway selected in signaling. There are accumulative evidences in support of the aggregation of cellular proteins or scaffolding of signaling molecules upon agonist activation of the receptor. Some of these molecules are protein kinases that could phosphorylate the receptor or G proteins leading to blunting of the signals. There are proteins such as regulator of G-protein signaling (RGS) that affect the GTP hydrolysis thereby altering the duration and magnitude of the G protein activity. Opioid receptor could interact also with the cellular protein β -arrestin, which in addition to terminating the original receptor signals, could initiate new signals due to its ability to serve as scaffold for other cellular proteins. Thus, depending on the conformation of the agonist-receptor complex, and the ability to recruit various cellular proteins to the receptor complex, different opioid agonists have shown to exhibit agonist-dependent pathway-selective signaling, or biased agonism. This is best demonstrated in the receptor-mediated activation of MAP kinase, in which morphine and fentanyl, two agonists used extensively in the clinics, required either G protein or β -arrestin to activate the kinase. The consequence of such pathway-selective activation of MAP kinase is manifested in the different genes being regulated by the two opiate agonists. Such difference has implications in the chronic usage of the drugs.

Summary and Future Direction

The similarities in the sequence and the molecular action of these multiple opioid receptors are striking. Hence, it is unclear why nature would evolve three different genes coding for three different proteins from the same ancestral gene so as to carry out the same functions. To complicate the picture, there are three endogenous peptides that exhibit selectivity but not specificity among these multiple opioid receptors. It is still unresolved why an endogenous opioid peptide such as Met⁵-enkephalin that exhibits *in vitro* agonistic properties in μ -opioid receptor would not produce analgesic effect *in vivo*. The physiological function of the endogenous ligands remains elusive also. There is no debate that endogenous peptides and opiate alkaloids have different recognition motifs within these opioid receptors, and that the cellular regulations of the peptide- and alkaloid-receptor complexes are different. Whether such differences could contribute to the differences in the pharmacological responses of these drugs or simply the spatial distribution of the receptors remains to be investigated.

In the past, the approach to eliminate the undesirable side effects of the opiate alkaloids was to design drugs that would have selectivity toward one specific opioid receptor. A classic example is the development of drugs that would have κ -opioid receptor selectivity, because the dysphoric effect of κ -opioid receptor activation would reduce the addictive liability of the drug. Unfortunately, drugs marketed thus far do not exhibit the ultimate property of opiate analgesic, that is, efficacious

pain relief with no side effects. The problem lies mainly with the presence of multiple opioid receptors and the selectivity of the drugs for these receptors. There are attempts to develop bivalent compounds, that is, drug molecules that have two pharmacophores recognizing two different receptors. By bridging an agonist and antagonist together with a spacer arm, the idea is to have the bivalent compound activate one receptor resulting in the pain relief, while the antagonist part of the molecule will inactivate a different receptor resulting in the blunting of the side effects. Such an approach is attractive and shows promise in blocking the tolerance development, but not the other side effects of opiate agonists. Hence, new approaches should be taken in the future for development of pain-management paradigms. With the advent of gene therapy, one could envision the use of receptor engineering in the treatment of pain. If an opioid-receptor mutant could be identified that exhibits distinct phenotype from that of the endogenous receptors, and that this receptor mutant could be delivered to the site of action, that is, substantia gelatinosa of the spinal cord, then activation of such mutants by a drug that normally would not activate the endogenous receptors, for example, an opiate antagonist, would result in pain relief. Such an approach would not only provide specificity, but also would eliminate the side effects of the opiate drugs; but this approach could only be

accomplished when the molecular basis for the action of the multiple opioid receptors is fully elucidated.

See also: **Molecular Biology:** DNA Damage: Alkylation; DNA Helicases: Hexameric Enzyme Action; DNA Polymerase I, Bacterial; Nonhexameric SF1 DNA Helicases/Translocases; **Protein/Enzyme Structure Function and Degradation:** Zinc Fingers.

Further Reading

- Galandrin S, Oligny-Longpré G, and Bouvier M (2007) The evasive nature of drug efficacy: Implication for drug discovery. *Trends in Pharmacological Sciences* 28: 423–430.
- Law PY, Wong YH, and Loh HH (2000) Molecular mechanisms and regulation of opioid receptor signaling. *Annual Review of Pharmacology and Toxicology* 40: 389–430.
- Minami M and Satoh M (1995) Molecular biology of the opioid receptors: Structures, functions, and distributions. *Neuroscience Research* 23: 121–145.
- Pasternak GW (1988) Multiple morphine and enkephalin receptors and the relief of pain. *Journal of the American Medical Association* 259: 1362–1367.
- Reisine T and Pasternak G (1996) Opioid analgesics and antagonists. In: Hardman JG, Limbird LE, Molinoff PB, Ruddon RW, and Gilman AG (eds.) *The Pharmacological Basis of Therapeutics*, 9th edn., pp. 521–555. New York, NY: McGraw Hill.
- Simon EJ and Giannini TL (1993) Opioid receptor multiplicity: Isolation, purification, and chemical characterization of binding sites. In: Herz A (ed.) *Opioid I, Handbook of Experimental Pharmacology*, vol. 104, pp. 3–36. Berlin: Springer.

ORAI Store-Operated Calcium Channel

Y Zhou and P G Hogan, La Jolla Institute for Allergy and Immunology, La Jolla, CA, USA

A Rao, La Jolla Institute for Allergy and Immunology, La Jolla, CA, USA; Children's Hospital Boston, Boston, MA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

ORAI The bona fide pore-forming subunit of the store-operated calcium channel, named after the keepers of heaven's gate in Greek mythology. The ORAI proteins have three isoforms, ORAI1, ORAI2, and ORAI3, each of which contains four transmembrane domains.

Store operated calcium entry (SOCE) A calcium ion entry mechanism in the plasma membrane activated by the

depletion of calcium ion from the internal calcium store in the endoplasmic reticulum.

Stromal interaction molecule (STIM) An ER-resident protein responsible for sensing ER calcium depletion via its luminal domain, and communicating with the store-operated ORAI calcium channels in the plasma membrane through its cytoplasmic domain.

Brief History and I_{CRAC}

Ca^{2+} is a universal and versatile messenger that regulates diverse biological processes, including triggering of life during fertilization and the perishing of cells through programmed cell death or apoptosis. To control the speed, amplitude, and spatio-temporal pattern of cytosolic Ca^{2+} signals, the cell draws on Ca^{2+} from two sources: the extracellular space where free Ca^{2+} concentration is as high as 1–2 mM and internal Ca^{2+} stores, mainly the endoplasmic reticulum (ER), with free luminal Ca^{2+} at a level of ~ 400 – 600 μ M. The cytoplasmic-free Ca^{2+} concentration ($[Ca^{2+}]_i$) in resting cells is tightly maintained at 50–100 nM through a long-term balance between Ca^{2+} influx and extrusion of Ca^{2+} via plasma membrane Ca^{2+} -ATPases (PMCA) and, in some cells, exchangers. In the short term, transient Ca^{2+} signals are shaped by pumping Ca^{2+} into the ER lumen through the sarco-endoplasmic Ca^{2+} -ATPases (SERCAs), by sequestering Ca^{2+} in mitochondria, and by Ca^{2+} binding to sites in proteins and other molecules. Changes in $[Ca^{2+}]_i$ can be initiated by the ligand engagement of plasma membrane receptors coupled to phospholipase C (PLC), an enzyme family that hydrolyzes the membrane phospholipid phosphatidylinositol-4,5-bisphosphate to produce the second messengers diacylglycerol and inositol-1,4,5-trisphosphate (IP_3). IP_3 subsequently binds to the IP_3 receptor in the ER membrane and triggers the release of Ca^{2+} from the ER lumen into the cell cytoplasm. In general, release of Ca^{2+} from internal stores is not sufficient by itself to promote long-term responses, such as the activation of nuclear factor of activated T cells (NFATs) and cytokine production by immune cells. With an ongoing stimulus, the initial rise of $[Ca^{2+}]_i$ triggered by store depletion is often followed by a second, sustained elevation of intracellular Ca^{2+} level (Figure 1). By the 1980s, the development of membrane-permeable Ca^{2+} indicator dyes and the patch-clamp technique provided the tools to examine these processes mechanistically.

In 1986, Putney proposed that the depletion of ER Ca^{2+} stores could evoke sustained Ca^{2+} influx across the plasma membrane of nonexcitable cells. Physiologically, this sustained Ca^{2+} influx is part of a chain of events following receptor activation, but it is a process separable experimentally from

receptor engagement, generation of second messengers, or the rise in $[Ca^{2+}]_i$. Observations on many types of cells were crystallized into the notion of capacitative Ca^{2+} entry, later termed store-operated Ca^{2+} entry (SOCE). Within a few years, electrophysiological studies established that T lymphocytes and mast cells indeed express Ca^{2+} channels that are opened in response to store depletion. The direct link between store depletion and Ca^{2+} current was demonstrated by use of pharmacological triggers such as the Ca^{2+} ionophore ionomycin, which empties the ER stores but bypasses receptor activation and the generation of IP_3 ; thapsigargin and other inhibitors of the SERCA pump, which have a similar effect by upsetting the balance between passive leakage of Ca^{2+} from ER stores and active uptake; and the chelator tetrakis-(2-pyridylmethyl)ethylene-diamine (TPEN), which lowers free Ca^{2+} concentration by binding Ca^{2+} in the ER lumen, without causing any net movement of Ca^{2+} out of the ER. The inward rectifying current recorded from T cells and mast cells, known as calcium release-activated calcium current (I_{CRAC}), is distinguished from other Ca^{2+} currents by the following unique fingerprint (also see the comparison in Table 1):

- I_{CRAC} is solely triggered by physiological stimuli leading to the partial depletion of ER Ca^{2+} stores or by experimental manipulations that mimic store depletion.
- The calcium release-activated calcium channel (CRAC) is highly selective for Ca^{2+} over monovalent cations such as Na^+ and K^+ , with a permeability ratio (P_{Na}/P_{Ca}) of $>1000:1$. This property sets the CRAC channel apart from many other Ca^{2+} -permeable channels, such as the transient receptor potential C (TRPC) channels that have a P_{Na}/P_{Ca} between 0.1:1 and 10:1.
- The single CRAC channel Ca^{2+} current amplitude is estimated to be as low as 4–5 fA, or roughly 10^4 Ca^{2+} ions flowing through the channel per second, at -110 mV transmembrane potential and 20 mM external Ca^{2+} , representing a stronger electrochemical force driving Ca^{2+} influx than the channel would experience physiologically. In comparison, the L-type Ca^{2+} channel, which resembles the CRAC channel in its Ca^{2+} selectivity, can carry $\sim 5 \times 10^5$ Ca^{2+} ions per second across the plasma membrane.

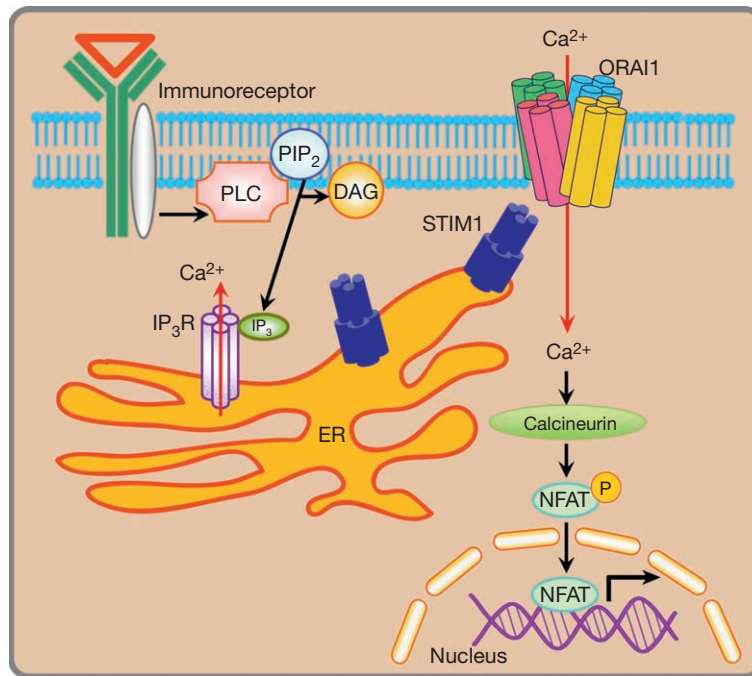


Figure 1 Scheme of SOCE mediated by the ORAI1–STIM1 pathway in lymphocytes. Upon engagement of immunoreceptors, the production of IP₃ leads to the activation of IP₃R and release of Ca²⁺ from internal stores. The decrease of ER Ca²⁺ concentration is subsequently sensed by the EF-hand-containing ER Ca²⁺ sensor STIM1, which in turn activates the ORAI1 CRAC channels through its interaction with the pore-forming subunit ORAI1, followed by a second phase of intracellular Ca²⁺ elevation and further activation of downstream effectors, including calcineurin, NFAT, and NFAT-dependent gene expression.

Table 1 Comparison of biophysical properties of the ORAI1 channel, L-type Ca_v channels, and TRPC channels

	<i>ORAI1</i> channel	<i>L-type Ca_v</i> channel	<i>TRPC</i> channel
Unitary conductance	<0.1 pS	3–20 pS	16–66 pS
Ion selectivity (P _{Ca} /P _{Na})	>1000	>1000	0.1–10
Modes of activation	Store operated	Voltage operated	Inositol lipid breakdown

- The CRAC channel exhibits distinct fast (100 ms timescale) and slow inactivation processes in response to the elevation of intracellular Ca²⁺.

Molecular Basis of SOCE: Roles of Stromal Interaction Molecules (STIM) and ORAI

The discovery of I_{CRAC} led to intense interest in the molecular identity of channels and regulators involved in SOCE. However, it was only with the advent of RNA interference (RNAi)-screening techniques that two protein families were clearly implicated in SOCE, the ER-resident stromal interaction molecules (STIM1 and STIM2) and the plasma-membrane ORAI proteins (ORAI1–ORAI3). Within these families, it is primarily STIM1 and ORAI1 that underpin I_{CRAC} in T lymphocytes and mast cells.

STIM Proteins

Early studies of STIM1 had provided evidence that STIM1 was a tumor suppressor or a transmembrane protein of bone marrow stromal cells involved in communication with B cells, and had placed a fraction of STIM1 in the plasma membrane. In 2005, limited RNAi screens performed with *Drosophila* S2 cells and HeLa cells identified *Drosophila* Stim and human STIM1 and STIM2 as proteins whose depletion diminished SOCE. STIM1 has an ER-luminal domain of ~22 kDa after cleavage of its signal peptide, a single transmembrane segment, and a cytoplasmic domain of ~51 kDa (Figure 2(a)). In contrast to expectations from the earlier work, STIM1 is predominantly localized in the ER, and it is generally accepted that ER-resident, but not plasma membrane STIM1, controls CRAC channel opening. ER-resident STIM1 carries out two basic functions: serving as a sensor of free Ca²⁺ in ER stores via its ER-luminal domain and communicating to ORAI Ca²⁺ channels in the plasma membrane through its cytoplasmic domain.

STIM1 ER-luminal domain: Sensing ER Ca²⁺ depletion and initiating STIM1 oligomerization

The ER-luminal portion of STIM1 contains an ER retention sequence, a canonical Ca²⁺-binding EF-hand, and a sterile alpha motif (SAM) domain. The Ca²⁺ sensing by STIM1 in cells is conveniently indicated by the redistribution of the diffusible STIM1 to form numerous individual spots, or puncta, adjacent to the plasma membrane following depletion of ER Ca²⁺ stores. The measured dissociation constant for Ca²⁺ binding to the STIM1 luminal domain is ~500–600 μM, a

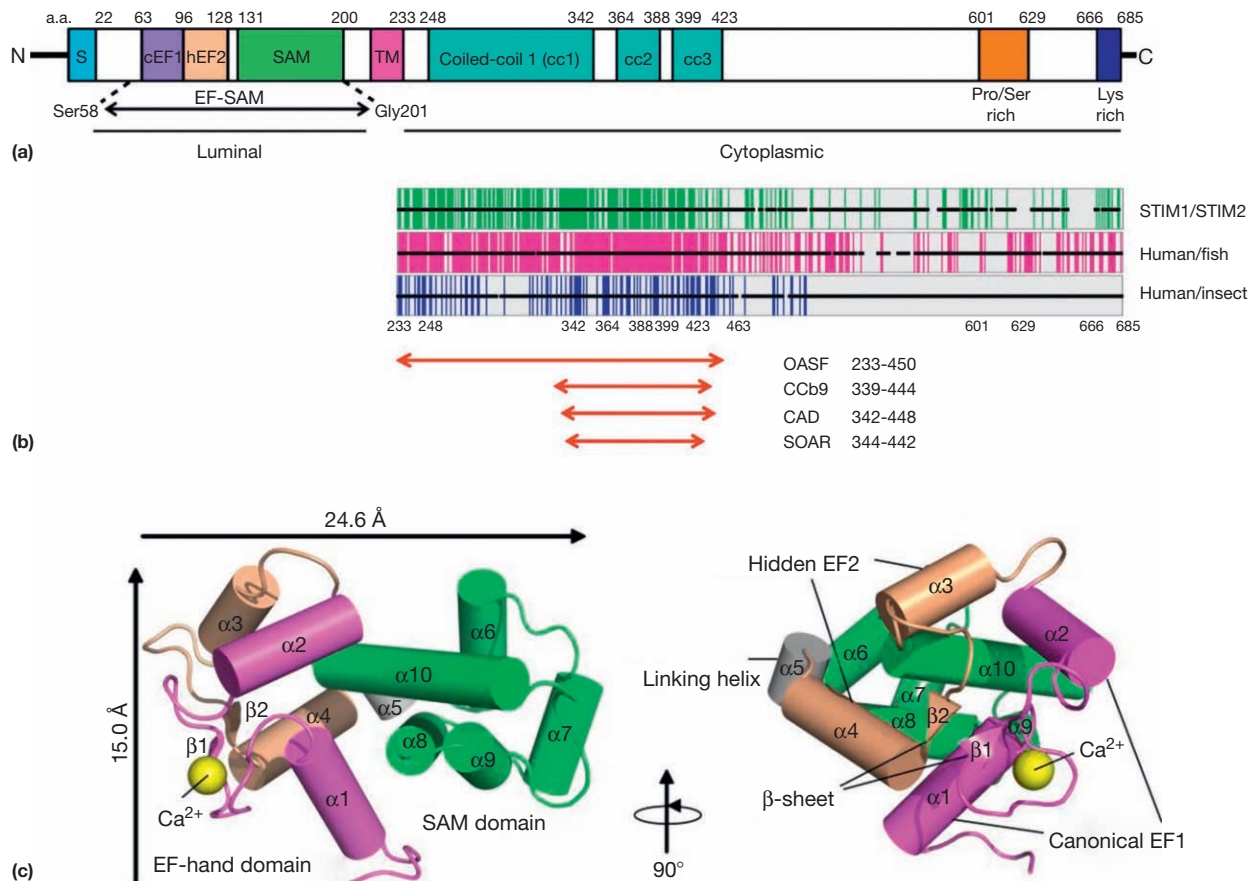


Figure 2 Structure and properties of STIM1. (a) Domain architecture of human STIM1, including the signal peptide (S), the canonical Ca²⁺-binding EF-hand 1 (cEF1), the hidden atypical EF-hand 2 (hEF2), the sterile alpha motif (SAM) domain, the transmembrane domain (TM), three putative coiled-coil regions (cc1, cc2, and cc3), the proline- and serine-rich region, and the polybasic lysine-rich region at the C terminus. The luminal EF-SAM fragment with determined NMR structure (panel c) is indicated. The numbers above the diagram indicate the approximate boundaries of the domains. (b) Sequence conservation in the cytoplasmic fragment of STIM1. The horizontal black bar represents the human STIM1 sequence, with gaps introduced to maintain alignment with human STIM2 or STIM proteins from other species as indicated. Vertical green lines indicate sequence identity between human STIM1 and STIM2 at corresponding positions; vertical magenta or blue lines indicate sequence identity among human STIM1 and at least four of five fish orthologs, or among human STIM1 and at least two of three insect STIM proteins, respectively. (c) Solution structure of the EF-SAM fragment determined by NMR spectroscopy. Cylinders represent alpha helices. (magenta, the canonical Ca²⁺-binding EF-hand 1 with bound Ca²⁺ (yellow sphere); beige, the hidden atypical EF-hand 2; green, the SAM domain). Two views related by a 90° rotation are shown. (a) Adapted from Hogan PG, Lewis RS, and Rao A (2010) Molecular basis of calcium signaling in lymphocytes: STIM and ORAI. *Annual Review of Immunology* 28: 491–533. (b) Adapted from Zhou Y, Meraner P, Kwon HT, et al. (2010) STIM1 gates the store-operated calcium channel ORAI1 *in vitro*. *Nature Structural & Molecular Biology* 17: 112–116. (c) Adapted from Stathopoulos PB, Zheng L, Li GY, Plevin MJ, and Ikura M (2008) Structural and mechanistic insights into STIM1-mediated initiation of store-operated calcium entry. *Cell* 135: 110–122.

value that is comparable to the free Ca²⁺ concentration in the ER lumen.

An NMR structure of a recombinant fragment of human STIM1 encompassing most of the STIM1 luminal domain (residues 58–201, termed 'EF-SAM', Figure 2(c)) reveals the presence of a second hidden non-Ca²⁺-binding EF-hand paired with the upstream canonical EF-hand. The paired EF-hands engage and stabilize the helix-rich SAM domain through extensive hydrophobic interactions. The EF-SAM domain readily undergoes a Ca²⁺-dependent transition from monomer to a mixture of dimers and larger aggregates *in vitro*: with Ca²⁺ bound, it exists as monomer in solution, whereas the dissociation of Ca²⁺ triggers unfolding and oligomerization of the EF-SAM domain. Mutations predicted to disable Ca²⁺ binding or to disrupt the interaction between the second

EF-hand and the SAM domain led to unfolding of the EF-SAM domain *in vitro* and caused constitutive formation of STIM1 puncta and Ca²⁺ influx in cells. Further, induced oligomerization of an engineered STIM1, in which the native luminal domain had been replaced by domains that will interact in the presence of a rapamycin analog, also triggered formation of puncta and activation of I_{CRAC}. The dimerization or oligomerization of STIM1 in cells during Ca²⁺ sensing is evidenced by increased fluorescence resonance energy transfer (FRET) between cyan fluorescent protein (CFP)-STIM1 donor and yellow fluorescent protein (YFP)-STIM1 acceptor upon store depletion, a step that precedes the appearance of STIM1 puncta near the plasma membrane. Thus, both *in vitro* and cell-based assays convincingly establish the physiological role of the STIM1 luminal domain in sensing changes in ER luminal

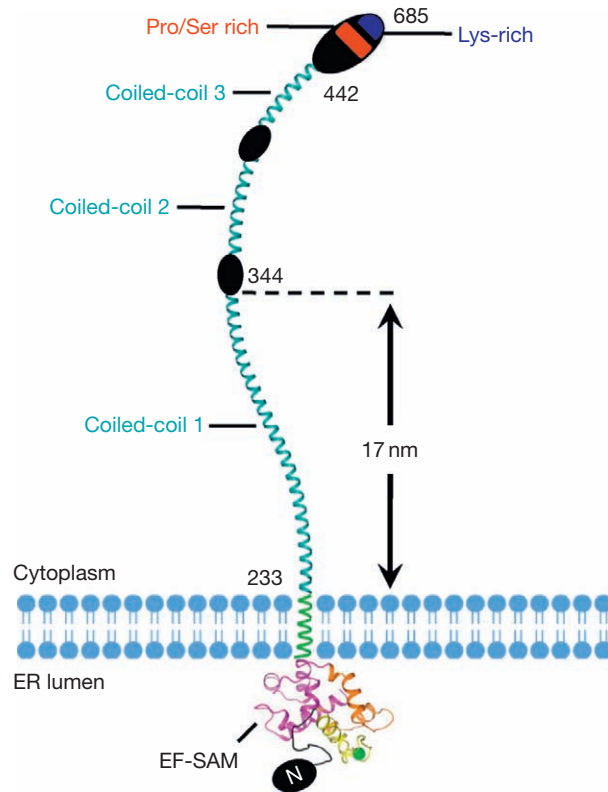


Figure 3 Cartoon representation of STIM1. The cytoplasmic fragment of STIM1 contains three putative coiled-coil regions (cyan), a proline- and serine-rich region (red), and a polybasic tail (blue). The coiled coil is able to span a distance of 8–17 nm that separates ER from plasma membrane. Adapted from Hogan PG, Lewis RS, and Rao A (2010) Molecular basis of calcium signaling in lymphocytes: STIM and ORAI. *Annual Review of Immunology* 28: 491–533.

Ca^{2+} concentration and in initiating the oligomerization of STIM1 and its movement to puncta.

STIM1 cytoplasmic domain: Coupling to and opening ORAI1 Ca^{2+} channel

The STIM1 cytoplasmic domain (residues 233–685) consists of three putative coiled-coil regions, a proline/serine-rich domain, and a poly-basic segment at the C terminus (Figure 3). This domain is responsible for direct coupling to ORAI1 and for triggering opening of the ORAI1 channel. The C-terminal poly-basic segment is believed to contribute to STIM1 targeting to ER-plasma membrane contacts by an interaction with negatively charged polyphosphoinositides such as PIP_2 and PIP_3 . N-terminal and C-terminal truncations of the STIM1 cytoplasmic domain have further mapped a minimal fragment of STIM1 that can directly interact with and activate overexpressed ORAI1. The minimal region, termed ‘STIM1-ORAI activating region (SOAR), CRAC activation domain (CAD), or ccb9, encompasses roughly residues 344–442, and overlaps with the second and third predicted coiled coils. Notably, the minimal region effective in activating CRAC current is positioned approximately 110 residues past the single transmembrane segment that tethers STIM1 to the ER. Thus, one likely function of the

first coiled coil is to span the $\sim 8\text{--}17$ nm separating ER from plasma membrane, but this does not rule out additional functions, such as participation in a conformational change that enables STIM1 to activate CRAC channels. The STIM1 cytoplasmic region and some of its fragments form dimers or tetramers in solution and bind directly to peptide segments in both the N- and C- terminal cytoplasmic regions of ORAI. Binding of recombinant STIM1 is sufficient to open the channel.

STIM2: Maintaining basal cytoplasmic and ER Ca^{2+} levels

The other isoform of human STIM proteins, STIM2, resembles STIM1 in its primary sequence (Figure 2(b)), domain architecture, and cellular localization. It also echoes the basic functional properties of STIM1. Recombinant STIM2 luminal domain binds Ca^{2+} with an affinity suitable for sensing ER Ca^{2+} levels and undergoes Ca^{2+} -dependent monomer-to-oligomer transition *in vitro*. In mammalian cells, STIM2 redistributes to puncta at ER-plasma membrane appositions, and activates ORAI1 channels when the proteins are coexpressed. However, although STIM2 can engage in the same signaling pathway as STIM1, in many cases it gives little evidence of contributing to SOCE. STIM2 at endogenous levels does not support SOCE in STIM1^{-/-} T cells and even overexpression of STIM2 only partially rescues SOCE. STIM2 redistributes in response to release of Ca^{2+} from ER stores in HEK293 cells responding to physiological stimulation, but it does not efficiently activate Ca influx. The established role for STIM2 is to maintain basal cytoplasmic and ER Ca^{2+} levels, an apparently humble occupation, yet absence of STIM2 has dramatic physiological and pathophysiological effects. STIM2^{-/-} murine T cells display severe deficits in the sustained translocation of NFAT to the nucleus and in cytokine production. In addition, STIM2 evidently has a key role in hypoxic neuronal cell death. STIM2^{-/-} neurons show increased survival under hypoxic conditions, and STIM2^{-/-} mice are protected from ischemia-induced neurological damage.

ORAI Proteins

ORAI1 was pinpointed as necessary for I_{CRAC} through the convergence of genome-wide *Drosophila* RNAi screens and a genetic approach that traced a familial severe combined immunodeficiency (SCID) syndrome to a point mutation in ORAI1. Mutations that altered ion selectivity indicated that ORAI was a CRAC channel subunit, and it is now considered to be the only essential channel subunit. The ORAI1 channel opens in response to the signal conveyed by STIM1, conducts Ca^{2+} selectively, and further communicates to downstream effectors.

Membrane topology and assembly of ORAI1

The human ORAI proteins have three isoforms, ORAI1, ORAI2, and ORAI3, among which ORAI1 is most thoroughly studied. ORAI1 monomer is a $\sim 33\text{-kDa}$ plasma membrane proteins with four transmembrane segments and intracellular N and C termini (Figure 4). The membrane topology of ORAI was established by (1) the intracellular location of N-terminal and C-terminal epitope tags and the extracellular location of an epitope tag introduced into the TM3–TM4 loop, (2) verification that an N-glycosylation site in the TM3–TM4 loop is

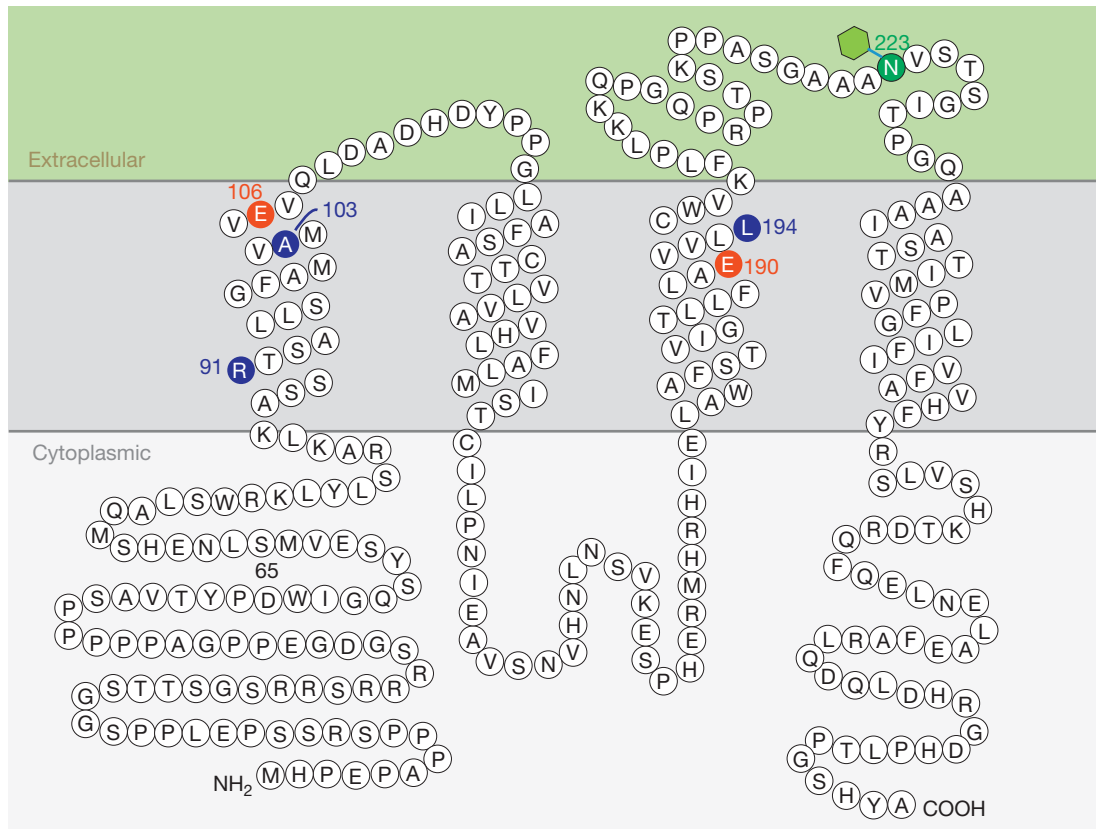


Figure 4 A snake-like diagram of human ORAI1. ORAI1 is a plasma membrane protein with four transmembrane (TM) segments. E106 (red) in TM1 is the main Ca^{2+} -binding site and directly coordinates Ca^{2+} ions traversing the channel. Another conserved glutamate residue E190 (orange) in TM3 has been implicated in Ca^{2+} selectivity, but does not directly coordinate Ca^{2+} ions. Disease-associated mutations R91W, which affect channel function, and A103E, and L194P, which affect protein stability, are shown in blue. The hexagonal structure attached to N223 (green) indicates the glycosylation site. Adapted from Hogan PG, Lewis RS, and Rao A (2010) Molecular basis of calcium signaling in lymphocytes: STIM and ORAI. *Annual Review of Immunology* 28: 491–533.

glycosylated and therefore extracellular, and (3) evidence that substitutions in the TM1–TM2 loop alter channel blockade by extracellular lanthanide ions.

ORAI1 is generally thought to assemble into a tetrameric CRAC channel. The best evidence comes from electrophysiological studies on channels made from concatenated ORAI1 monomers. A tandem tetramer of ORAI1 is functional as a selective Ca^{2+} channel, and fails to incorporate additional ORAI monomers when challenged by coexpression of a dominant negative ORAI1 subunit. The tetrameric structure of the ORAI1 channel is further corroborated by single-molecule photobleaching studies of green fluorescent protein-tagged *Drosophila* Orai and human ORAI1 channels sparsely expressed in the plasma membrane, in which four individual bleaching steps were often resolved, albeit only for activated Orai in the case of the *Drosophila* channel. In addition, chemical cross-linking studies indicate that both *Drosophila* Orai and human ORAI1 can form tetramers in the native membrane.

Structural determinants of Ca^{2+} permeation, selectivity, and gating

Like other ion channels, ORAI1 provides an energetically favorable ion conduction pathway that allows specific ions to traverse the lipid bilayer. The pore of the ORAI1 channel is

capable of selectively discriminating among different ions, particularly choosing Ca^{2+} over the more abundant Na^{+} of physiological solutions. And, as with many other channels, pore opening and closing (gating) of the ORAI1 channel is tightly and precisely regulated. Accelerated biochemical and electrophysiological studies have started to provide insights into the pore architecture and gating of this unique Ca^{2+} channel:

- The ORAI1 channel is assembled as a tetramer or a higher oligomer with the transmembrane helix 1 (TM1) centrally located. One face of each TM1 helix lines the pore.
- The short loop between TM1 and TM2 forms a flexible outer vestibule above the mouth of the pore. La^{3+} , but not Ca^{2+} , can bind tightly to this loop.
- The negatively charged glutamate residue E106 in TM1 plays a key role in Ca^{2+} permeation and contributes to the high Ca^{2+} selectivity of ORAI1 channels. Most likely, four glutamate residues at position 106 of a tetrameric ORAI channel directly coordinate Ca^{2+} ions traversing the channel, in much the way that the EEEE locus in the P-loop coordinates Ca^{2+} in the pore of the L-type voltage-gated Ca^{2+} channel. The pore (with an estimated diameter of 3.8 Å) sharply narrows at or below the level of E106, thereby shaping the ion selectivity.

- Although a conserved glutamate residue (E190) in TM3 has been implicated in Ca^{2+} selectivity, it is not directly involved in the coordination of Ca^{2+} . The E190Q substitution alters ion selectivity and widens the pore size from 3.8 to 7.0 Å, but probably does so through an allosteric change in the selectivity filter.
- STIM1 directly gates ORAI1 channels by physically interacting with both the C-terminal region of ORAI1 and an N-terminal segment immediately preceding TM1. Considerable circumstantial evidence suggests that this N-terminal segment may be involved in channel gating. A movement of TM1 might contribute to the channel gating, through either a displacement of TM1 or a reorientation of TM1 that brings E106 into a geometry allowing ions to travel through the channel. A deep and detailed understanding of gating mechanism, nevertheless, requires the determination of the atomic structure of ORAI1 and/or the ORAI1–STIM1 complex.

Orai2 and Orai3

The other two isoforms of ORAI proteins, ORAI2 and ORAI3, share remarkable sequence similarity to ORAI1 in the four transmembrane segments. Key residues equivalent to E106 and E190 in ORAI1 that contribute to the Ca^{2+} selectivity are conserved in these two proteins. In addition, the N- and C-terminal segments implicated in the STIM1–ORAI1 interaction are conserved. These channels are selective for Ca^{2+} over Na^+ . However, they differ in more subtle channel properties, such as the prominence of fast inactivation and the response to the pharmacological agent 2-aminoethoxydiphenyl borate (2-APB). For example, 2-APB at low concentrations has been found to activate store-operated ORAI1 and ORAI2 currents, but 2-APB at 75 μM is an inhibitor. By contrast, 2-APB at 75 μM directly activates the ORAI3 channel, independent of either Ca^{2+} store depletion or STIM1.

ORAI3 channels account for store-operated Ca^{2+} entry in human MCF7 adenocarcinoma cells. ORAI proteins are also capable of forming heteromeric channels. A pentameric assembly of ORAI1 and ORAI3 subunits is reported to form arachidonate-regulated Ca^{2+} -selective (ARC) channels that respond to arachidonic acid rather than to store depletion. It has also been proposed that ORAI proteins form heteromeric channels that are less Ca^{2+} selective than the homomeric channels. The full repertoire and physiological function of channels that incorporate ORAI2 and ORAI3 subunits remain to be explored.

STIM–ORAI Signaling

That ORAI and STIM are major players in the SOCE pathway has been established by (1) reconstitution of SOCE and CRAC current in the SCID T cells by expression of ORAI1; (2) production of monster Ca^{2+} -selective currents with the characteristics of I_{CRAC} when STIM1 and ORAI1 are overexpressed in mammalian cells; (3) evidence that various fragments derived from the STIM1 cytoplasmic domain are capable of decorating the plasma membrane and activating constitutive CRAC currents in ORAI1-expressing cells; and (4) successful recapitulation of Ca^{2+} flux through the pore of

ORAI1 channels induced by the STIM1 cytoplasmic fragments *in vitro*. Through extensive biochemical, structural, electrophysiological, and cell biological studies in the past 5 years, a wealth of information has been accumulated to support the following model of ORAI channel activation (Figure 5) and regulation.

Steps in SOCE

In the resting state, STIM1 is homogeneously distributed throughout the ER, whereas ORAI is distributed throughout the plasma membrane. Upon store depletion, the dissociation of Ca^{2+} from the STIM1 canonical EF-hand leads to partial unfolding and exposure of hydrophobic regions in the STIM1 luminal domain, which ultimately causes the oligomerization of STIM1. STIM1 oligomers then migrate toward ER-plasma membrane appositions to form STIM1 puncta within tens of seconds. This step is independent of ORAI expression and involves at least the interaction of the STIM1 polybasic C terminus with PIP_2 and PIP_3 in the plasma membrane. Subsequently, STIM1 oligomers recruit ORAI1 to puncta, most likely through a direct protein–protein interaction involving the ORAI C-terminus and a cytosolic region of STIM1. Finally, STIM1 oligomers open ORAI channels, possibly through a physical interaction with a membrane-proximal N-terminal segment (contained within residues 65–91) of ORAI1. The stoichiometry of the STIM–ORAI complex remains an open question that might be eventually resolved through structural determination of the conformations and oligomerization states of full-length STIM1 and ORAI1.

Modulation of STIM–ORAI Signaling

Influx of Ca^{2+} results in a high local Ca^{2+} concentration near the cytoplasmic opening of the ORAI channel. A feedback inhibition mechanism known as ' Ca^{2+} -dependent fast inactivation' limits Ca^{2+} overloading into the cells. This process is demonstrated experimentally as a decline in I_{CRAC} over 200–500 ms during large hyperpolarizing voltage steps that trigger substantial Ca^{2+} influx. The molecular determinants of this regulatory mechanism include (1) a stretch of negatively charged sequence C-terminal to the SOAR or CAD domain in STIM1, (2) a calmodulin-binding site in the membrane-proximal N-terminal region of ORAI1 that can be engaged by calmodulin in a Ca^{2+} -dependent manner, (3) the intracellular loop of ORAI1 that links TM2 and TM3 of ORAI1, and (4) the selectivity filter itself.

Ca^{2+} influx must be tightly regulated, and one regulator of SOCE has been identified through a proteomic approach. The protein termed 'CRACR2A' is a cytosolic Ca^{2+} -sensing protein that contains two EF-hand domains and extensive coiled-coil regions. CRACR2A has been shown to facilitate the clustering of ORAI1 and STIM1 in puncta through direct interactions with both ORAI1 and STIM1. The ternary complex falls apart when the cytosolic Ca^{2+} rises. Given the limited amounts of ORAI and STIM in a physiological context, it is speculated that CRACR2A plays a role in stabilizing the active CRAC channel complex upon store depletion.

The only reported physiological case in which SOCE is strongly suppressed or inactivated is during cell division,

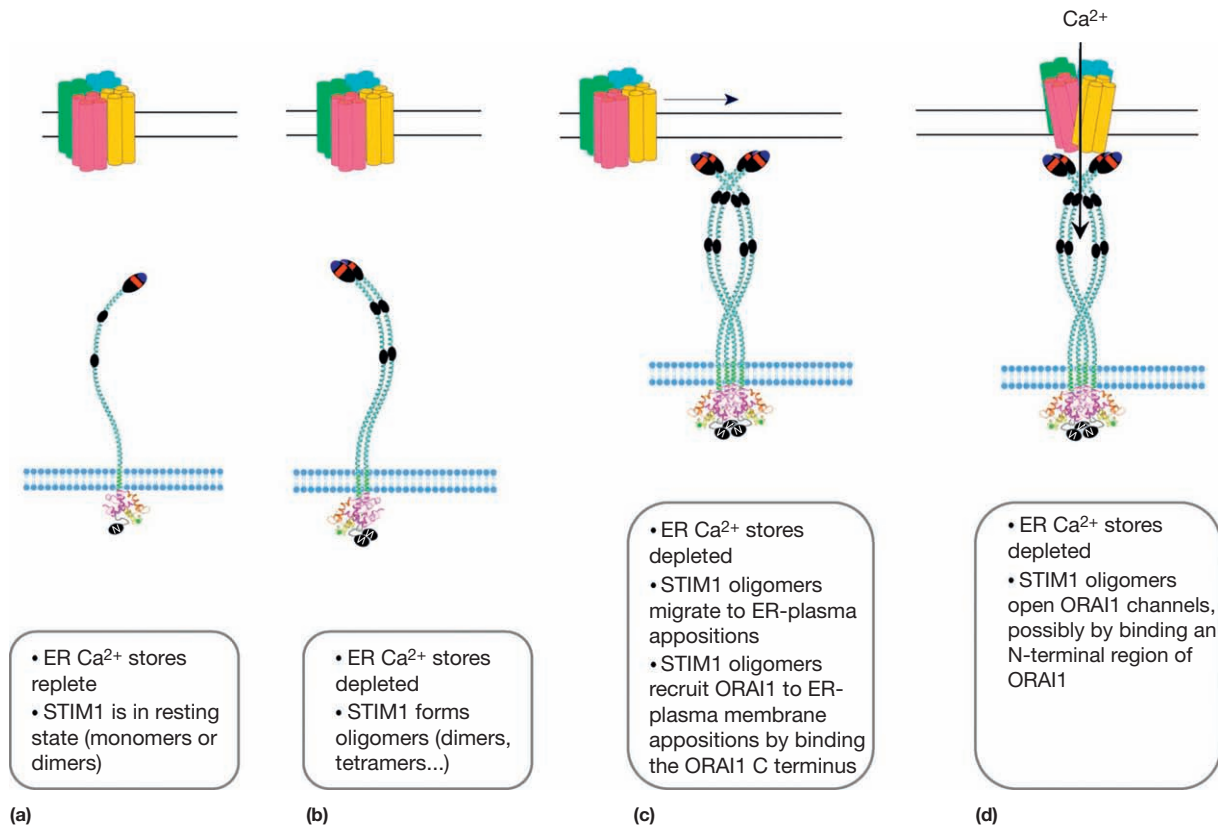


Figure 5 Major steps in SOCE mediated by ORAI1 and STIM1. (a) Schematic of STIM1 and ORAI1 in resting state, when ER Ca^{2+} stores are replete. Although their oligomerization states in resting cells are not yet fully defined, STIM1 is depicted as a monomer and ORAI1 as a tetramer for convenience. (b) STIM1 forms oligomers when ER Ca^{2+} stores are depleted. Again, the exact oligomerization state of STIM1 in activated cells remains unknown. (c) STIM1 oligomers undergo redistribution and migrate toward ER-plasma membrane appositions. STIM1 oligomers further recruit ORAI1 to ER-plasma membrane appositions by directly interacting with the C-terminal tail of ORAI1. (d) STIM1 oligomers open ORAI1 channels possibly through a second contact with an N-terminal region of ORAI1. Adapted from Hogan PG, Lewis RS, and Rao A (2010) Molecular basis of calcium signaling in lymphocytes: STIM and ORAI. *Annual Review of Immunology* 28: 491–533.

particularly during the M phase of the cell cycle. One recent study on mitotic HeLa and mitotic HEK293 cells indicates that phosphorylation of STIM1 may account for suppressed SOCE during mitosis. Two mitosis-specific phosphorylation sites in STIM1 (residues S486 and S668) have been mapped by taking a proteomic approach. Substitution of these sites with alanine to prevent phosphorylation has been shown to rescue mitotic SOCE. Nevertheless, results from another study on *Xenopus* oocytes during meiosis argue that phosphorylation cannot be the sole method of SOCE inactivation. In that study, it was shown that both the internalization of ORAI1 and the inhibition of STIM1–STIM1 interaction have a role in uncoupling store depletion from SOCE activation during meiosis.

STIM–ORAI Signaling and Human Disease

CRAC channel function is defective in a few human patients with inherited SCID. T cells from these SCID patients display a marked deficiency in cytokine production, loss of SOCE, loss of I_{CRAC} , and failure of efficient nuclear translocation of NFAT. The genetic basis has been traced back in four families to mutations in either ORAI1 or STIM1. The identified mutations

in ORAI1 have included a missense loss-of-function mutation (R91W), a frameshift that causes loss of ORAI1 protein expression, and two independent missense mutations, A103E and L194P, that also lead to loss of ORAI1 expression. A fourth inherited immunodeficiency was traced to insertion of an adenine in the third exon of STIM1 that results in a frameshift and premature termination, and absence of STIM1 protein.

Although loss of CRAC channel function is an uncommon cause of primary immunodeficiency, these cases have been illuminating in terms of CRAC channel signaling. Both ORAI1 and STIM1 are widely expressed in different tissues, yet the major clinical consequences of defective CRAC channels are surprisingly limited to the immune system, skeletal muscle, and certain ectodermally derived tissues. The main clinical manifestation of ORAI1 and STIM1 loss-of-function mutations is severe immunodeficiency, marked by frequent and severe infections early in life. The most prominent associated non-immune disorders are a congenital, nonprogressive myopathy, and ectodermal dysplasia. Thus, drug candidates that specifically target the STIM1–ORAI1 pathway have the potential to suppress immune function selectively and may lack the toxicity of current immunosuppressive agents, such as cyclosporine A and FK506.

See also: **Bioenergetics:** Calcium in the Regulation of Gene Expression; Calcium Oscillations; Calcium-Modulated Proteins (EF-Hand); IP₃ Receptors; Transient Receptor Potential Channels; Voltage-Gated Ca²⁺ Channels.

Further Reading

- Cahalan MD (2009) STIMulating store-operated Ca²⁺ entry. *Nature Cell Biology* 11: 669–677.
- Feske S, Gwack Y, Prakriya M, et al. (2006) A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature* 41: 179–185.
- Feske S, Picard C, and Fischer A (2010) Immunodeficiency due to mutations in ORAI1 and STIM1. *Clinical Immunology* 135: 169–182.
- Hogan PG, Lewis RS, and Rao A (2010) Molecular basis of calcium signaling in lymphocytes: STIM and ORAI. *Annual Review of Immunology* 28: 491–533.
- Hogan PG and Rao A (2007) Dissecting I_{CRAC}, a store-operated calcium current. *Trends in Biochemical Sciences* 32: 235–245.
- Hoth M and Penner R (1992) Depletion of intracellular calcium stores activates a calcium current in mast cells. *Nature* 355: 353–356.
- Liou J, Kim ML, Heo MD, et al. (2005) STIM1 is a Ca²⁺ sensor essential for Ca²⁺-store-operation-triggered Ca²⁺ influx. *Current Biology* 15: 1235–1241.
- Parekh AB (2010) Store-operated CRAC channels: Function in health and disease. *Nature Reviews Drug Discovery* 9: 399–410.
- Parekh AB and Putney JW, Jr. (2005) Store-operated calcium channels. *Physiological Reviews* 85: 757–810.
- Putney JW, Jr. (1986) A model for receptor-regulated calcium entry. *Cell Calcium* 7: 1–12.
- Rao A (ed.) (2009) Special issue: Calcium signaling in immune cell regulation. *Immunological Reviews* 231: 5–288.
- Roos J, DiGregorio PJ, Yeromin AV, et al. (2005) STIM1, an essential and conserved component of store-operated Ca²⁺ channel function. *Journal of Cell Biology* 169: 435–445.
- Vig M, Peinelt C, Beck A, et al. (2006) CRACM1 is a plasma membrane protein essential for store-operated Ca²⁺ entry. *Science* 312: 1220–1223.
- Zhang SL, Yeromin AV, Zhang XH, et al. (2006) Genome-wide RNAi screen of Ca²⁺ influx identifies genes that regulate Ca²⁺ release-activated Ca²⁺ channel activity. *Proceedings of the National Academy of Sciences of the United States of America* 103: 9357–9362.
- Zweifach A and Lewis RS (1993) Mitogen-regulated Ca²⁺ current of T lymphocytes is activated by depletion of intracellular Ca²⁺ stores. *Proceedings of the National Academy of Sciences of the United States of America* 90: 6295–6299.

Organization of the Bacterial Nucleoid

C J Dorman, Trinity College Dublin, Dublin, Republic of Ireland

© 2013 Elsevier Inc. All rights reserved.

Glossary

Chromosome domains Discrete loops of bacterial chromosomal DNA, approximately 10 kilobases in circumference, thought to be pinched closed by domain boundary elements such as nucleoid-associated proteins.

Nucleoid The region of the bacterial cell that is occupied by the genetic material and its associated proteins and RNA. Unlike the nucleus of a eukaryotic cell, the bacterial nucleoid is not bounded by a membrane.

Sigma factor The protein component of bacterial RNA polymerase that recognizes and binds to the transcriptional promoter in DNA, carrying out the isomerization step that converts the closed transcription complex to an open one

during transcription initiation. The sigma factor is ejected from RNA polymerase during the elongation phase of transcription.

Topoisomerase An enzyme that alters the linking number of DNA by a mechanism involving DNA strand cleavage, strand passage, and religation.

Transcription factories (or foci) Accumulations of the most active transcription promoters in the cell (usually the promoters of ribosomal RNA genes) together with RNA polymerase. When RNA polymerase is linked to a fluorescent marker, these accumulations are detectable by microscopy. Rapidly growing bacterial cells typically have two or three of these transcription factories.

Nucleoid-Associated Proteins

The histone proteins in eukaryotes play a critical role in organizing the DNA in the nucleus and in modulating gene expression. When proteins of similar molecular mass were discovered that performed analogous tasks in bacteria, these were initially named 'histone-like' proteins. This name is going out of use, chiefly because its implication of amino acid sequence similarity to eukaryotic histones is misleading. The bacterial proteins are, in general, quite distinct from the histones in their primary structure and are now usually referred to as 'nucleoid-associated proteins', or NAPs. However, both terms may be found in use in the literature, sometimes in the same article.

There are approximately 12 different NAPs in *Escherichia coli* and they form a disparate group of proteins. Some of the better characterized ones are listed in Table 1. NAPs are abundant and contribute to bacterial chromatin through their DNA-binding activities. Some bind to specific sequences in DNA, a trait that is shared with transcription factors. Others have no obvious DNA sequence preferences among the sites to which they bind but interact instead with particular structures, such as curved DNA. Upon binding, the NAPs influence the path of the DNA in different ways. Some, such as the leucine-responsive regulatory protein, Lrp, wrap the DNA, others, such as the integration host factor (IHF), bend the DNA quite severely, while yet others (e.g., H-NS, the histone-like nucleoid-structuring protein) can engage in the formation of DNA-protein-DNA bridges or can polymerize along the DNA (Figure 1). These DNA manipulations lend themselves to the compaction of the DNA in the nucleoid. They are thought to act in concert with molecular crowding and with negative supercoiling of the DNA to reduce the diameter of the nucleoid in ways that allow it to fit in the cell while allowing the DNA to be replicated and to participate in transcription and recombination.

NAPs and Gene Expression

The NAPs influence gene expression in a number of ways, chiefly at the level of transcription, and this, in turn, has consequences for the structure of the nucleoid. NAPs with well-defined binding sites in DNA frequently are discovered to behave as transcription factors, either enhancing or inhibiting the activities of promoters. The Lrp, IHF, and Crp proteins are examples of NAPs that perform this function. The Crp protein is a particularly interesting case because it was described originally as a transcription factor, but more recently has been shown by genome-wide surveys of Crp-DNA interactions to bind to many hundreds of sites across the nucleoid, a characteristic of NAPs.

The DNA-wrapping protein HU does not have a clear consensus binding site in DNA (unlike its close relative IHF), but it can influence transcription through its ability to facilitate the formation of small loops in DNA that promote regulatory interactions at target promoters. The factor for inversion stimulation (Fis) controls the expression of hundreds of genes and is especially important in optimizing the transcription of the genes that code for the translational machinery of the cell. It has the ability to organize local domains of supercoiled DNA and can facilitate transcription through a combination of local DNA topological adjustments and direct interactions with RNA polymerase.

Transcription has the potential to influence the structure of the nucleoid. When a fluorescent protein is attached to RNA polymerase, accumulations of highly transcribed genes can be seen in individual cells by microscopy. These bodies are called 'transcription foci' or 'transcription factories' and they are created by the driving together of the genes and operons that encode the components of the translation machinery of the cell. This happens in cells that are growing rapidly, if growth is halted the foci disperse. The forces that create and maintain the foci are still not fully understood but the phenomenon shows that

Table 1 Major nucleoid-associated proteins of *Escherichia coli*

Protein	Gene	Molecular mass (kDa)	Remarks
Dps	<i>dps</i>	19	Dps (DNA protection, starvation): abundant in stationary phase; also expressed in bacteria experiencing oxidative stress; general DNA-binding activity; does not influence transcription
Fis	<i>fis</i>	11.2	Fis (factor for inversion stimulation): homodimeric DNA-binding protein; binds to degenerate consensus sequence; bends DNA by between 40° and 90°; organizes local DNA topology; multifunctional; and is involved in transcription, DNA replication, recombination, and transposition. Strong growth-phase regulation varies protein number/cell over 200–100 000 range.
H-NS	<i>hns</i>	15.6	H-NS (histone-like nucleoid structuring): oligomeric DNA-binding protein; lacks a consensus sequence for DNA binding but has an affinity for curved DNA; compacts DNA and constrains supercoils; important transcription regulator (usually a repressor); estimated to be 20 000–60 000 H-NS monomers/cell.
HU	<i>hupA</i> <i>hupB</i>	9.5 9.5	HU (histone-like from strain 'U93'): abundant heterodimeric DNA-binding protein; 60 000 dimers/cell; lacks a consensus sequence for DNA binding; can constrain DNA supercoils.
IHF	<i>ihfA</i> <i>ihfB</i>	11.2 10.6	IHF (integration host factor): heterodimeric DNA binding and bending protein; homolog of HU but binds to well-conserved consensus sequence; bends DNA by up to 180°; between 17 000 and 34 000 monomers/cell.
Lrp	<i>lrp</i>	18	Lrp (leucine-responsive regulatory protein): degenerate consensus binding site sequence; 6000 molecules per cell; dimers, octamers, hexadecamers; activity can be modulated by binding leucine.

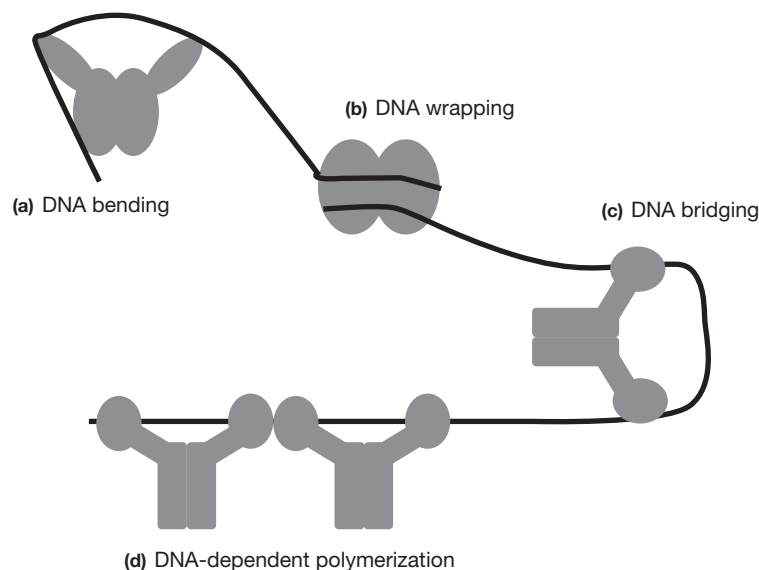


Figure 1 The impact of nucleoid-associated proteins on DNA structure. NAPs have a variety of effects on the DNA to which they bind and the principal ones are illustrated here. (a) DNA bending results from binding of many proteins, including NAPs such as IHF. This protein is a heterodimer (Table 1) that contacts the DNA at three places. Two beta ribbons extending from the dimer have proline amino acids at their apices that intercalate in the minor groove of the DNA, causing the duplex to bend by up to 180°. The resulting U-turn can facilitate long-range interactions in DNA. (b) DNA wrapping by proteins such as the Lrp protein compact the DNA and also displace other DNA-binding proteins with alternative modes of DNA binding, such as the bridging activity of H-NS. (c) DNA bridging is promoted by proteins such as H-NS that are dimeric and have a discrete DNA-binding domain that can be placed in contact with either different parts of the same DNA molecule (as shown here) or with two separate DNA molecules. The resulting bridge can act as an effective blockade to the movement of RNA polymerase and so impede the process of transcription. (d) Oligomerization of an NAP. The H-NS protein can also bind to a nucleation site from which it oligomerizes along DNA. The resulting protein–DNA complex can impede binding by other proteins, such as RNA polymerase, and so silence gene expression. Recent data suggest that H-NS can cycle between its DNA–protein–DNA bridging and oligomerization modes in response to the intracellular concentration of divalent cations such as magnesium. The physiological significance of this switching is currently unclear.

the nucleoid has a dynamic structure that is subject to remodeling in response to changes in the physiology of the bacterium.

DNA Supercoiling

The DNA in *E. coli* is underwound about its helical axis and this has important consequences for the structure of the nucleoid

and for the molecular transactions that take place within it. Its underwound state places the DNA under torsional stress and it responds by writhing about its own helical axis, creating the supercoiled form of the already coiled duplex. Supercoiling contributes to the compaction of the DNA in the nucleoid.

The tension in the underwound DNA represents a source of energy that can cause parts of the duplex to form single-stranded bubbles due to breakage of the hydrogen bonds

between pairs of bases. This facilitates transcription and other molecular processes that require the DNA to become transiently single stranded. It is important to note that the process of transcription imparts supercoiling to the DNA template, partly explaining why rapidly growing bacteria with highly active genes and operons have DNA that is more underwound than quiescent cells. DNA supercoiling has to be carefully managed so that it is maintained within limits that are optimal for the survival of the cell. Special enzymes called topoisomerases perform this management function (Table 2).

DNA gyrase is a type II topoisomerase that converts relaxed DNA to a negatively supercoiled form by making double-stranded breaks in the DNA duplex and passing an intact portion of the same DNA molecule through the gap, resealing it thereafter (Figure 2). Gyrase uses adenosine triphosphate (ATP) as a source of energy to complete each supercoiling reaction. This links gyrase activity to the ATP-generating processes of the cell, making gyrase activity rise and fall with the metabolic flux. Thus, rapidly growing cells will have an ATP content that is favorable for high levels of gyrase activity; cessation of growth produces cells in which DNA is relaxed.

DNA topoisomerase I opposes the supercoiling activity of DNA gyrase by relaxing negatively supercoiled DNA. It does this through breakage of one DNA strand, allowing the energy stored in the supercoiled DNA to unwind the gapped molecule by a swiveling action (Figure 2). If gyrase activity is absent, topoisomerase I will remove negative supercoils from the DNA in the nucleoid wherever it can, potentially leading to full relaxation of the chromosome.

NAPs and Topoisomerases

When a protein wraps DNA around its surface, it can preserve the supercoiled or writhed path of the DNA while, at the same time, relieving the tension due to the underwinding of the DNA molecule. It is estimated that about half of the DNA in *E. coli* is complexed with protein, leaving the other half to experience the torsional stress associated with underwinding of protein-free DNA. In this way, the NAPs can modulate the consequences of DNA supercoiling. They can also influence the interaction of topoisomerases with DNA. For example,

Table 2 DNA topoisomerases of *Escherichia coli*

Enzyme	Gene	Molecular mass (kDa)	Remarks
Topoisomerase I	<i>topA</i>	97	Type I enzyme; relaxes negatively supercoiled DNA; cuts one strand of DNA and binds to cut site via a 5'-phospho-tyrosine covalent bond; Mg^{2+} dependent.
DNA gyrase (topoisomerase II)	<i>gyrA</i> <i>gyrB</i>	105 (GyrA) 95 (GyrB)	Type II enzyme; negatively supercoils relaxed DNA (requires ATP); relaxes positively supercoiled DNA; makes a transient double-stranded cut in DNA and binds via a 5'-phosphotyrosine bond Mg^{2+} dependent.
Topoisomerase III	<i>topB</i>	73.2	Type I enzyme; can relax negatively supercoiled DNA; decatenase; Mg^{2+} dependent.
Topoisomerase IV	<i>parC</i> <i>parE</i>	75 (ParC) 70 (ParE)	Type II enzyme with strong sequence homology to gyrase; decatenase; relaxes negative supercoils but cannot create them; requires ATP and Mg^{2+} .

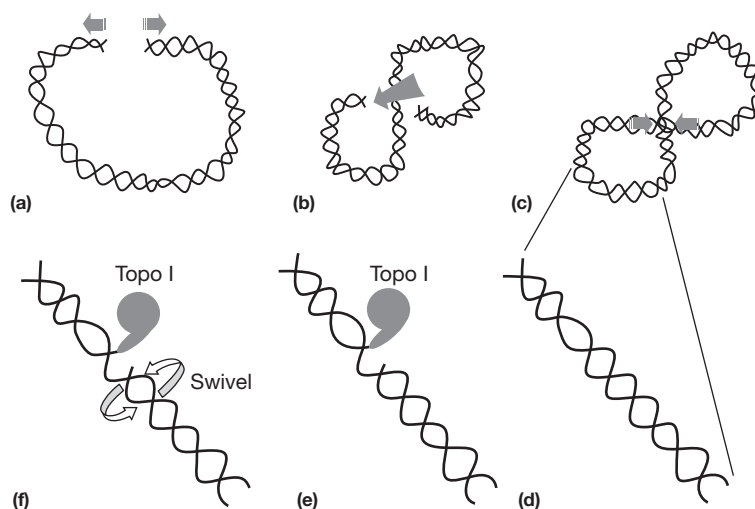


Figure 2 Bacterial topoisomerases and DNA supercoiling. DNA gyrase (not shown) creates a double-stranded break (a) in a circular DNA molecule, passes part of the intact molecule through the resulting gap (b), and then reseals the gap (c) introducing a superhelical turn into the molecule. Part of this supercoiled molecule (d) is bound by DNA topoisomerase I which cuts one DNA strand and becomes covalently attached to one site of the single-stranded gap (e) allowing the free end to rotate around the intact strand (f) releasing negative supercoiling by a swiveling action that relaxes the DNA.

the Fis protein can prevent gyrase and topoisomerase I from over-supercoiling or over-relaxing the DNA template, respectively. This protein appears to play a useful buffering function in the nucleoid by helping to optimize supercoiling levels adjacent to key genes, such as those that code for components of the translation apparatus. Fis also regulates the genes that encode gyrase (*gyrA* and *gyrB*) and topoisomerase I (*topA*): it is a repressor of the *gyrA* and *gyrB* genes and has a dual activator/repressor role at *topA*. Finally, the transcriptional promoter that is responsible for expression of the *fis* gene that encodes the Fis protein is activated by DNA supercoiling and inhibited by DNA relaxation.

The Domain Structure of the Nucleoid

Early experiments with DNA-nicking agents indicated that the chromosome of *E. coli* was organized in a series of independent domains within the nucleoid. A nick in one DNA strand at one site might be expected to result in the relaxation of the whole circular supercoiled chromosome through swiveling of the cut strand around the intact one. However, only limited and localized loss of supercoiling was observed, suggesting that isolators of some kind divide the chromosome into domains that are topologically independent. Genetic experiments and investigations using restriction enzymes to cut the genome in known locations have confirmed the existence of chromosomal domains and these are now estimated to have an average length of 10 kilobase pairs. However, it seems clear that the boundaries of these domains are not fixed and that the number of boundaries varies with the growth rate of the cell. Rapidly growing bacteria have about 400 domains but this number declines sharply as growth ceases. Much effort has been expended in attempting to identify the components of the cell that form the domain boundaries. At the time of writing, NAPs are among the leading contenders, with the Fis and H-NS proteins being especially prominent candidates. Their ability to form DNA–protein–DNA bridges adds to their attraction as potential isolators.

It has become apparent very recently that the nucleoid has a superstructure consisting of four macrodomains. These are located around the origin of chromosome replication (OriC), the termination region (Ter), and the flanking regions that connect these two. The MatP protein is associated exclusively with the Ter region, the SImA protein binds within the other three, while the SeqA protein is associated with GATC sequences in the OriC macrodomain. The integration of the macrodomain superstructure with the NAP-associated microdomain structure is currently a subject of intense research.

Bacterial Physiology, DNA Supercoiling, and NAPs

It is clear that the structure of the nucleoid is dynamic and that it varies as bacteria grow. This is not surprising when one considers that the features of the cell that impart structure to the nucleoid (transcription, molecular crowding, DNA supercoiling, and NAPs) are themselves dynamic. Experiments with *E. coli* and related bacteria where the organisms are grown in batch culture with aeration at their optimum temperature for

growth in a medium that supports a generation time of about 20 min reveal that the superhelicity of the DNA undergoes a series of transitions from relaxed in the inoculum to negatively supercoiled during the exponential phase of growth to relaxed again upon the onset of stationary phase. Under the same growth conditions, the population of NAPs changes. As exponential growth begins, the Fis protein increases to 100 000 copies, making it transiently the most abundant NAP in the cell (Table 1). This event coincides with a demand for ribosomes and other parts of the translational apparatus. The promoters of the relevant genes have a dependency on Fis for optimal activity. In addition, Fis can feed back negatively onto the genes coding for gyrase, exerting a braking action on the expression of topoisomerase that introduces negative supercoils into DNA just as their level peaks in the cell. The H-NS protein is expressed at a relatively constant rate throughout growth and inhibits transcription throughout the genome. However, the H-NS-to-DNA ratio is critical and even minor disturbances to this can alter the global gene expression pattern. This omnipresent protein must be opposed if the hundreds of genes that H-NS represses are to be expressed and this is achieved through a wide variety of molecular mechanisms that include a role for the Fis protein as an H-NS antagonist. As the bacteria enter the stationary phase of growth, the levels of Fis decline to negligible values and another NAP, Dps, becomes dominant. This protein, which appears to have no detectable role in controlling transcription, binds to the chromosome and protects it from chemical damage during the pseudohibernation period of late stationary phase. DNA in these cells is relaxed and transcription is generally repressed, except for a special set of stationary-phase-specific genes whose promoters are read by a dedicated stationary phase sigma factor called RpoS, an RNA polymerase component that can function well with a DNA template that lacks the highly energized state that is characteristic of the chromosome in exponentially growing bacteria. Thus, it may be seen that nucleoid structure and cellular physiology are intimately linked and that the features of the nucleoid that influence its structure also affect directly the gene expression program of the cell and vice versa.

See also: **Cell Architecture and Function:** Chromosome Organization and Structure, Overview; **Molecular Biology:** DNA Supercoiling; DNA Topoisomerases: Type I; DNA Topoisomerases: Type II; Nuclear Organization, Chromatin Structure, and Gene Silencing.

Further Reading

- Dame RT and Dorman CJ (2010) *Bacterial Chromatin*. Heidelberg: Springer.
- Dillon SC and Dorman CJ (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology* 8: 185–195.
- Higgins NP (2005) *The Bacterial Chromosome*. Washington, DC: American Society for Microbiology.
- Thanbichler M, Viollier PH, and Shapiro L (2005) The structure and function of the bacterial chromosome. *Current Opinion in Genetics and Development* 15: 153–162.
- Travers A and Muskhelishvili G (2005) DNA supercoiling – a global transcriptional regulator for enterobacterial growth? *Nature Reviews Microbiology* 3: 157–169.

Oxygenases

O Hayaishi, Osaka Bioscience Institute, Osaka, Japan

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 178–182, © 2004, Elsevier Inc.

Glossary

Dehydrogenase The enzyme that catalyzes electron transfer from a substrate to an acceptor.

E.C. numbers and trivial names Enzymes are classified and numbered by the International Enzyme Commission. In this list, the names of the enzymes are formal names, whereas trivial names are more common names or informal names.

For example, catechol 1,2-dioxygenase is a formal name while pyrocatechase is a trivial name.

Oxidase The dehydrogenase that utilizes molecular oxygen as an immediate electron acceptor.

Oxygenase The enzyme that incorporates molecular oxygen into substrates.

Oxygenases are a group of oxidative enzymes that catalyze the direct addition or fixation of molecular oxygen into various substrates. The terms ‘mono’ and ‘di’ oxygenases are generally assigned to the enzymes catalyzing the incorporation of either one or two atoms of oxygen per mole of substrate, respectively. Prior to the discovery of oxygenases in 1955, the essential characteristics of biological oxidation processes was believed to be the removal or transfer of electrons or hydrogen atoms from a substrate to an appropriate acceptor. These enzymes were termed ‘dehydrogenases’ and when oxygen molecules serve as the immediate electron acceptor, the enzymes have been called ‘oxidases’. Functionally, dehydrogenases and oxidases are mainly involved in energy metabolism whereas oxygenases play major roles in the anabolism and catabolism of biological materials as well as synthetic compounds.

Background

In 1932, Professor Heinrich Wieland, a German Nobel laureate, authored a book entitled *On the Mechanism of Oxidation*, in which he proposed the famous ‘dehydrogenation theory’. According to this theory, the principle of biological oxidation processes is the activation and transfer of hydrogen atoms, or their equivalent, from the substrate molecule to an appropriate acceptor such as coenzymes and various dyes. Oxygen molecules may, in some instances, serve as the immediate electron acceptor, and if so, then these enzymes are termed ‘oxidases’. Thus, according to Professor Wieland, molecular oxygen accepts hydrogen atoms and is reduced to water or hydrogen peroxide. However, it is never incorporated into the substrate. When the overall reaction is envisaged as the addition of oxygen, the oxygen atoms are always derived from water molecule rather than from atmospheric oxygen. For example, when aldehydes are converted to acids, aldehydes are hydrated first and then dehydrogenation occurs so that the sum is the addition of oxygen to the substrate X to form XO, but this oxygen atom is derived from water rather than from molecular oxygen. Thus, the direct addition of molecular oxygen to a substrate was considered completely irrelevant to biological oxidation.

In 1955, a set of experiments was performed in which pyrocatechase from a pseudomonad was incubated with its substrate catechol and a heavy oxygen isotope, ^{18}O (oxygen-18), the latter being in its molecular form in air in one flask or in the form of water in the other. The product, muconic acid, was isolated and analyzed for its ^{18}O content by mass spectrometry. Contrary to the then generally held belief, the results clearly demonstrated that the oxygen atoms incorporated into the product molecules were derived exclusively from molecular oxygen and not from the oxygen of the water molecule. Concurrently and independently, the phenolase complex of the mushroom was shown to incorporate one atom of molecular oxygen into a substrate, 3,4-dimethylphenol, to form 4,5-dimethylcatechol; whereas the other atom of oxygen was reduced to water. These findings together with subsequent work by others established that oxygen fixation did occur in biological systems, and revealed a new biological role of molecular oxygen; thus, the new concept of biological oxygenation was introduced. These enzymes that catalyze the direct incorporation of molecular oxygen were then named oxygenases. At first, these oxygenase reactions were generally thought to be rather unusual and were limited to only primitive living organisms such as soil bacteria or mushrooms. It took several years before the ubiquitous existence of oxygenases was confirmed in many laboratories; during that time, a large number of oxygenases were isolated from animals, plants, and microorganisms and were shown to play important roles in the metabolism of various nutrients, hormones, and neurotransmitters as well as synthetic compounds.

Nomenclature, Classification, and General Properties of Oxygenases

Dioxygenases

Dioxygenases are defined as enzymes catalyzing reactions in which both atoms of molecular oxygen are incorporated into substrates. In many instances where one substrate can act as the oxygen acceptor, the term ‘intramolecular dioxygenase’ may be used; pyrocatechase (catechol 1,2-dioxygenase) (EC 1.13.11.1) is a typical example. The dioxygenases acting upon two

acceptor substrates may be referred to as 'intermolecular dioxygenases'. One of the two substrates for the latter type has so far been invariably 2-oxoglutarate. Hypoxia-inducible factor 1 is a global regulator of cellular and systemic O_2 homeostasis in animals. Prolyl hydroxylase, also known as procollagen-proline: α -ketoglutarate 4-dioxygenase (EC 1.14.11.2), the enzyme that hydroxylates specific prolyl residues of the collagen chains during their translation and requires 2-oxoglutarate as a co-substrate in addition to ascorbate and Fe^{2+} , serves as an oxygen sensor. More recent evidence indicates that asparaginyl hydroxylase that hydroxylates the asparagine residues by a similar mechanism also serves as a direct oxygen sensor. A third class of dioxygenases include various enzymes that require nicotinamide adenine dinucleotide (NADH) or nicotinamide adenine dinucleotide phosphate (NADPH) as an electron donor. Although it is quite possible that reactions of this class may involve a simple dioxygenation reaction followed by a reductive step, these two processes may be coupled in such a way as to justify a separate category. The formation of catechol from anthranilate is an example of this type of reaction.

Prostaglandin endoperoxide synthase (EC 1.14.99.1) is a unique intramolecular dioxygenase catalyzing the sequential addition of two oxygen molecules to a substrate molecule arachidonic acid to produce prostaglandin G_2 followed by the peroxidase reaction to generate prostaglandin H_2 . This enzyme is also called cyclooxygenase, abbreviated as COX, and is of great clinical significance as discussed below.

Some dioxygenases, such as tryptophan 2,3-dioxygenase, contain heme as the sole prosthetic group; whereas others, such as pyrocatechase, contain nonheme iron, or like quercetinase, contain copper as the prosthetic group.

Monooxygenases

Monooxygenases are defined as a group of enzymes that catalyze the incorporation of one atom of molecular oxygen into a substrate, with the other being reduced to water. Therefore, these enzymes are sometimes referred to as mixed function oxidases or mixed function oxygenases. The electron donors that serve as coenzymes for these enzymes include reduced forms of pyridine nucleotides, flavin nucleotides, cytochromes, ascorbic acid, and pteridine derivatives. In some cases, the substrate itself may serve as an electron donor.

Internal Monooxygenases

The simplest type of monooxygenase catalyzes the incorporation of a single atom of molecular oxygen concomitant with the reduction of the other oxygen atom by electrons derived from the substrate. As the reducing agent is internally supplied, these enzymes may be referred to as internal monooxygenases. The first of these enzymes to be crystallized was the lactate oxidative decarboxylase from *Mycobacterium phlei*. This enzyme catalyzes the conversion of lactate to acetate with the incorporation of one atom of oxygen into acetate, the evolution of 1 mol of CO_2 , and the reduction of one atom of oxygen to water.

External Monooxygenases

Although the internal monooxygenases do not require external reducing agents, more common types of monooxygenases require various types of electron donors. The electron donor (DH_2) serves as a basis for the subclassification of external monooxygenases. Some examples are as follows:

1. *Flavoprotein monooxygenases with reduced pyridine nucleotides as DH_2 .* Salicylate 1-monooxygenase, a flavoprotein, is an example of this type of enzyme and catalyzes the formation of catechol from salicylate.
2. *Heme-containing monooxygenases.* Aryl 4-monooxygenase (liver microsomal cytochrome p-450) catalyzes the hydroxylation of a variety of substrates with reduced flavin as DH_2 .
3. *Heme-containing monooxygenases with a reduced iron-sulfur protein as DH_2 .* Camphor 5-monooxygenase is a typical example of this type of enzyme.
4. *Pteridine-linked monooxygenases.* Phenylalanine-4-monooxygenase catalyzes the formation of tyrosine from phenylalanine. Nitric oxide (NO) is a ubiquitous bioactive substance and a unique messenger molecule. It is synthesized from L-arginine by NO synthase (NOS), (EC 1.14.13.39), a unique monooxygenase that catalyzes two consecutive monooxygenase reactions and that requires NADPH, tetrahydrobiopterine, heme, and Ca^{2+} /calmodulin and contains flavin adenine dinucleotide and flavin mononucleotide. The product of the reaction is L-citrulline.
5. *External monooxygenases with ascorbate as DH_2 .* Dopamine β -monooxygenase is an example of this type of enzyme.
6. *External monooxygenases with another substrate as DH_2 .* Monophenol monooxygenase is an example of this type. In this case, dopa may be considered as the electron donor in the reaction.

Chemistry of Oxygen Fixation Reactions

Molecular Oxygen as a Substrate

Oxygenases utilize two different species of substrate, namely, molecular oxygen and the oxygen acceptor, which is called the substrate and may be either an organic or an inorganic compound. It has not yet been clarified whether gaseous oxygen or oxygen dissolved in water is utilized by these enzymes. It is generally believed that the latter type of oxygen is different from the former. In water, oxygen is considered to exist largely in a dimerized form, O_2-O_2 because of its biradical nature, and probably forms a charge transfer complex with a water molecule. Whatever form is taken by the oxygen dissolved in water, the type of oxygen that serves as a substrate for oxygenases is in rapid equilibrium with gaseous oxygen under normal experimental conditions and can be clearly distinguished through the use of $^{18}O_2$ from the oxygen in water molecules or other compounds in the reaction mixture.

Role of the Substrate That Acts as Oxygen Acceptor

As for the acceptor for molecular oxygen, a great variety of both organic and inorganic compounds can be oxygenated.

In general, oxygen-rich and/or hydrophilic compounds such as carbohydrates are not favorable substrates for oxygenases because these usually have many reactive groups containing oxygen, such as the hydroxyl, carbonyl, or formyl, and their biochemical function does not require further oxygenation. On the other hand, hydrophobic compounds such as lipids and aromatic compounds are often metabolized by oxygenases. These oxygen-deficient compounds generally require oxygenation in order to become biologically active or more soluble in water. By contrast, purines and pyrimidines with their hydrophilic ring systems are usually hydroxylated by the addition of water, followed by dehydrogenation.

Reaction Mechanisms and the Nature of Active Oxygen

Evidence generated from a number of laboratories has indicated that the enzyme binds oxygen only in the presence of substrate to form the ternary complex, ESO_2 , in which oxygen and substrate interact to form a product. Such a ternary complex was postulated in 1964 on the basis of binding experiments, but more direct experimental evidence was not available until 1967 when a ternary complex of tryptophan 2,3-dioxygenase-tryptophan- O_2 was demonstrated by spectrophotometric experiments. Since that time similar oxygenated intermediates have also been observed, such as those with protocatechuate 3,4-dioxygenase (an iron-sulfur protein dioxygenase), lysine monooxygenase (a flavoprotein), and cytochrome P-450. In each case, the enzyme must bind the organic substrate first, before it can be oxygenated, in contrast to oxygen-carrying pigments such as hemoglobin and hemoerythrin, which are freely and reversibly oxygenated in the absence of any effector or substrate. In fact, several lines of evidence indicate that the substrate tryptophan binds specifically to the heme coenzyme of tryptophan 2,3-dioxygenase; as a consequence, the state of the heme is altered in such a way that its reactivity toward ligands is increased by several orders of magnitude.

The nature of the so-called active form of oxygen in the above-mentioned ternary complexes has been one of the most difficult questions in this field. All oxygenase-catalyzed reactions are exothermic and are therefore irreversible. Nevertheless, molecular oxygen is a rather inert compound and at room temperature reacts slowly with the substrate compounds in the absence of enzymes. This low kinetic reactivity is usually explained on the basis that molecular oxygen is in a triplet ground state. The direct reaction of a triplet molecule with organic molecules in the singlet state is electronically spin forbidden, and therefore a substantial activation energy is required. For this reason, singlet oxygen has been suggested as a likely intermediate in many oxygenase-catalyzed reactions; however, definitive evidence is so far unavailable. It has recently been discovered that during enzymic hydroxylation of aromatic substrates, the substituent (deuterium, tritium, chlorine, bromine, etc.) displaced by the entering hydroxyl group migrates to an adjacent position in the aromatic ring. On the basis of extensive studies of this phenomenon, called the 'NIH shift', the active oxygen species involved in certain monooxygenases was postulated to be oxenoid.

On the other hand, evidence has appeared indicating that O_2^- , that is, the superoxide anion, may be the active form of

oxygen in the case of indoleamine 2,3-dioxygenase, and also in the hydroxylation reactions catalyzed by hepatic cytochrome P-450. It is, however, uncertain whether or not the superoxide anion is the general form of active oxygen in all oxygenase-catalyzed reactions or whether these enzymes represent a new class of enzyme that utilizes the superoxide anion rather than molecular oxygen as an oxygenating agent.

Biological Function of Oxygenases

Although oxidases and dehydrogenases are mainly involved in energy metabolism, namely, the generation of adenosine triphosphate, oxygenases play important roles in biosynthesis, transformation, and degradation of essential metabolites such as amino acids, lipids, sugars, porphyrins, vitamins, and hormones. They also play a crucial role in the metabolic disposal of foreign compounds such as drugs, insecticides, and carcinogens. Furthermore, they participate in the degradation of various natural and synthetic compounds by soil and airborne microorganisms in nature and are, therefore, of great significance in environmental sciences. Therefore, the significance of biological oxygen fixation in medicine, agriculture, and microbiology, and also in food technology, cosmobiology, public health problems, and biochemistry in general, has now been well established.

For example, certain oxygenases fulfill exceptionally important biological functions as exemplified by prolyl and asparaginyl hydroxylases as oxygen sensor as well as NO synthase as already mentioned. Several other examples are being briefly described.

Ribulose 1,5-bisphosphate carboxylase/oxygenase (EC 4.7.1.39), also known as 'rubisco', catalyzes the CO_2 fixation as well as dioxygenase reaction and is the most abundant naturally occurring catalyst in nature. The global fixation of O_2 by rubisco can be estimated to be $\sim 10^{11}$ metric ton (t) per year.

Tryptophan 2,3-dioxygenase (EC 1.13.11.11), also known as tryptophan pyrrolase, catalyzes the pyrrole ring cleavage by the insertion of two atoms of oxygen to produce L-formylkynurenine. This enzyme is present in the liver and is highly specific for L-tryptophan.

On the other hand, indoleamine 2,3-dioxygenase (EC 1.13.11.42), also known as IDO, shows broad substrate specificity including L- and D-tryptophan, L- and D-5-hydroxytryptophan, serotonin, and so forth and is widely distributed in almost all organs and tissues except in the liver. IDO is induced by interferon and has been implicated in the defense mechanism against infection by depleting tryptophan from the infected tissues and cells. More recently, allogenic fetal rejection was reportedly prevented by tryptophan catabolism by IDO.

The prostaglandin endoperoxide synthases I and II (COX-1 and COX-2) are the major targets of nonsteroidal anti-inflammatory drugs including aspirin, indomethacin, ibuprofen, and the new COX-2 inhibitors. These drugs reduce inflammation, pain, and fever and also fatal thrombotic events, colon cancer, and Alzheimer's disease.

Another common dioxygenase involved in lipid metabolism and widely distributed in both animals and plants is lipoxygenase (EC 1.13.11.12), which is also known as LOX,

lipoxidase, lipoperoxidase, or carotene oxidase. It is a non-heme iron-containing dioxygenase and catalyzes the formation of hydroperoxy derivative of unsaturated fatty acids. Of particular interest is the 5-LOX that catalyzes the dioxygenation of arachidonic acid to produce corresponding *cis,trans*-diene hydroperoxide followed by the dehydration to generate leukotrien A₄, another biologically important lipid mediator.

Unspecific monooxygenase is a heme-thiolate enzyme commonly known as cytochrome P-450, hepatic microsomal monooxygenase, xenobiotic monooxygenase, or aryl-4-monooxygenase (EC 1.14.14.1), and catalyzes hydroxylation of a wide variety of both aromatic and aliphatic compounds. Monooxygenases also catalyze a seemingly diverse group of reactions including epoxide formation, dealkylation, decarboxylation, deamination, and N- or S-oxide formation. Although the overall reactions catalyzed by various monooxygenases appear grossly unlike one another, the primary chemical event is identical, because these processes are all initiated by the incorporation of one atom of molecular oxygen into the substrate.

Oxygenases in Evolution

Primitive Earth is believed to have started with an oxygen-free anaerobic environment. Therefore, the minimum manifestation of life is capable of occurring in the absence of molecular oxygen, as exemplified by the strictly anaerobic microorganisms, which survive and grow solely in the absence of molecular oxygen. With the appearance of oxygen in the Earth's atmosphere, molecular oxygen became a more useful and efficient tool as the terminal electron acceptor. Oxygenases then appeared gradually with the subsequent development of extensive and complicated metabolic pathways for essential components of cell structures and of regulatory mechanisms of metabolism and growth. The appearance of aerobic form of life paved the way for oxygenase-catalyzed reactions by which complicated messengers and structural compounds

such as sterols, prostaglandins, and neurotransmitters are efficiently synthesized, while cytochrome P-450 plays an essential role in the metabolism of thousands of drugs, carcinogens, and other synthetic compounds and is probably the largest family of proteins that has been cloned and characterized to date. In an evolutionary sense, oxidases preceded oxygenases but the latter appears to be a more advanced oxygen-utilizing catalyst associated with the most advanced and highly specialized life processes.

See also: [Bioenergetics: Cytochrome P-450](#); [Signaling: Dopamine Receptors](#).

Further Reading

- Bruick RK and McKnight SL (2001) Oxygen sensing gets a second wind. *Science* 295: 807–808.
- Groves JT (2003) The bioinorganic chemistry of iron in oxygenases and supramolecular assemblies. *Proceedings of the National Academy of Sciences of the United States of America* 100: 3569–3574.
- Hayaishi O (ed.) (1962) *Oxygenases*. New York and London: Academic Press.
- Hayaishi O (ed.) (1974) *Molecular Mechanism of Oxygen Activation*. New York, NY: Academic Press.
- Hayaishi O, Katagiri M, and Rothberg S (1955) Mechanism of the pyrocatechase reaction. *Journal of the American Chemical Society* 77: 5450–5451.
- Ishimura Y, Nozaki M, Yamamoto S, et al. (eds.) (2002) *Oxygen and Life-Oxygenases, Oxidases and Lipid Mediators*. Amsterdam: Elsevier.
- Mason HS, Fowles WL, and Peterson E (1955) Oxygen transfer and electron transport of the phenolase complex. *Journal of the American Chemical Society* 77: 2914–2915.
- Nozaki M, Yamamoto S, Ishimura Y, et al. (eds.) (1982) *Oxygenases and Oxygen Metabolism*. New York, NY: Academic Press.
- Purich DL and Allison RD (eds.) (2002) *The Enzyme Reference: A Comprehensive Guidebook to Enzyme Nomenclature, Reactions, and Methods*. San Diego, CA: Academic Press.
- Solomon EI, Decker A, and Lehnert N (2003) Non-heme iron enzymes: Contrasts to heme catalysis. *Proceedings of the National Academy of Sciences of the United States of America* 100: 3589–3594.

P2Y Receptors

G R Dubyak, Case Western Reserve University, Cleveland, OH, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Agonist A molecule that binds to a receptor resulting in functional activation of the receptor.

Antagonist A molecule that binds to a receptor and keeps the receptor in a functionally inactive state.

Effector protein An intracellular protein, such as an enzyme or ion channel, whose activity is regulated by interaction with G proteins. Effector proteins regulate the production of second messenger signals.

G protein An intracellular protein that becomes active when it binds guanosine triphosphate (GTP) and becomes inactive when it hydrolyzes the bound GTP to guanosine diphosphate (GDP). G proteins act as signal transduction proteins that couple cell-surface receptors to the regulation of intracellular effector proteins.

G-protein-coupled receptor (GPCR) A cell-surface receptor protein for an extracellular signaling molecule; GPCR

reversibly interacts with G proteins to activate intracellular signal transduction pathways. There are thousands of different GPCRs that selectively bind a wide variety of extracellular signaling molecules, including many hormones and neurotransmitters.

P2Y receptors A family of related GPCR that become active when bound to extracellular nucleotides such as adenosine triphosphate (ATP), adenosine diphosphate (ADP), or uridine triphosphate (UTP).

Second messenger A small intracellular molecule that is transiently formed in response to the activation of an effector protein by a cell-surface receptor, such as a GPCR. Formation of the second messenger acts to relay the signal from the cell-surface receptor to the interior of the cell.

P2Y receptors are intrinsic plasma membrane proteins that mediate functional responses of intact cells to extracellular nucleotides, including adenosine triphosphate (ATP), adenosine diphosphate (ADP), uridine triphosphate (UTP), uridine diphosphate (UDP), and UDP-glucose. P2Y receptors belong to the superfamily of G-protein-coupled receptors (GPCRs), an exceptionally large group of structurally and functionally related gene products that play crucial roles in the transduction of signals between and among different types of cells. In response to the binding of extracellular nucleotides, P2Y receptors undergo conformational changes that facilitate their direct interaction with, and activation of, G proteins. In turn, the activated G proteins couple P2Y receptors to the regulation of enzymes and ion channels that facilitate generation of second messengers and intracellular signaling cascades used for the acute or chronic regulation of specific cellular functions. Appreciation of the physiological or pathological functions of P2Y receptors requires some understanding of the general mechanisms by which nucleotides are used as extracellular signaling molecules.

Extracellular Nucleotides as Signaling Molecules

In addition to their crucial roles in intracellular energy metabolism and nucleic acid synthesis, nucleotides, such as ATP, and

nucleosides, such as adenosine, play important roles as extracellular signaling molecules. Burnstock and his colleagues generated many of the initial hypotheses regarding the roles of purines in nonadrenergic, noncholinergic neurotransmission. They proposed that extracellular nucleotides and nucleosides are utilized for signal transduction at nerve endings in diverse tissues. Implicit in this concept of purinergic transmission or signaling was a requirement that ATP (or other nucleotides) be released and metabolized in a highly localized manner at sites of cell-to-cell communication. Given the emphasis on the role of purines in this neuronal signaling, initial characterization focused on describing the release of nucleotide/nucleosides at neuron-to-neuron synapses, neuron-to-tissue varicosities, or the immediate locale of neuroendocrine cells. Such studies verified that neurons and neuroendocrine cells can release ATP via standard mechanisms involving exocytosis of nucleotides co-packaged with biogenic amines or other neurotransmitters in specialized vesicles or granules.

However, subsequent research has resulted in the cloning and characterization of at least 15 distinct ATP/nucleotide receptor genes (including the P2Y receptors), four adenosine receptor genes, and a dozen different ecto-nucleotidase genes. Analysis of the expression of these genes revealed that most vertebrate cell types express one or more subtypes of nucleotide or nucleoside

receptor along with one or more of the ecto-enzymes used for metabolizing extracellular nucleotides. Because most cells lack direct proximity to nerve synapses or neuroendocrine cells, the identification of additional sources of extracellular nucleotides is a current area of investigation and speculation. Recent studies indicate that stressed or damaged cells in many tissues can locally release nucleotides via a variety of pathways ranging from irreversible cytolysis to nonlytic export through nucleotide-permeable channels. In turn, the released nucleotides activate autocrine or paracrine signaling cascades that permit those tissues to adapt to the particular stress or injury that precipitated release of the cellular nucleotides.

Regardless of the mechanism by which ATP or other nucleotides are released from cells, their half-life within extracellular compartments is generally brief due to rapid catabolism by ecto-nucleotidases. Some of these ecto-enzymes serially convert released nucleotide triphosphates (e.g., ATP) to their diphosphate (e.g., ADP) and monophosphate (AMP) forms. Other ecto-nucleotidases dephosphorylate the nucleotide monophosphates (e.g., AMP) to their corresponding nucleosides (e.g., adenosine). These nucleosides can be reaccumulated and reconverted to intracellular nucleotides via highly conserved nucleotide salvage pathways. In the case of adenosine, nucleosides can additionally function as agonists for G-protein-coupled nucleoside receptors that are distinct from the P2Y nucleotide receptors. Thus, the local release of ATP may initially activate ATP-selective P2Y receptor subtypes. Although this hydrolysis of ATP to ADP by ecto-ATPases would act to terminate signaling by these particular P2Y subtypes, this same extracellular metabolic reaction would result in activation of ADP-selective P2Y receptor subtypes. Further metabolism to adenosine might result to activation of G-protein-adenosine receptors in the same or adjacent cells. In these ways, ATP release in intact tissues can trigger the serial activation of multiple P2Y and adenosine receptor subtypes to generate complex patterns of positive or negative feedback among the various cell types that comprise the tissue. As a further complication, many cells export nucleotide diphosphokinases that catalyze transphosphorylation reactions between the various extracellular nucleotide tri- and diphosphates. Thus, released ATP may additionally drive the phosphorylation of ambient UDP to locally generate UTP that is a selective agonist for yet other P2Y receptor subtypes.

General Structure of the P2Y Receptors

As members of the GPCR superfamily, all P2Y receptors are intrinsic membrane proteins that consist of a single polypeptide chain with certain invariant structural domains. These include: (1) an extracellular amino terminal tail; (2) seven transmembrane-spanning segments (usually alpha helices); (3) three short extracellular loops and three variably sized intracellular loops that connect transmembrane segments; and (4) an intracellular carboxyl terminal tail. The exact protein domains of P2Y receptors that facilitate selective recognition and binding of the extracellular nucleotide agonists have not been identified as yet. Significantly, no single P2Y domain contains amino acid sequences similar to the well-defined nucleotide-binding sites of the many intracellular proteins that interact with ATP (e.g., ATPases, protein kinases, and nucleic acid polymerases).

Functional Classification of the P2Y Receptors

Genes encoding P2Y receptors have been identified in the genomes of all vertebrate species examined thus far. By contrast, no obvious P2Y receptor orthologs are apparent in the completely sequenced genomes of the insect, *Drosophila melanogaster*, or the nematode worm, *Caenorhabditis elegans*. At present, genes encoding eight distinct P2Y receptor subtypes have been characterized in a broad range of tissues and cell types from vertebrate organisms. These subtypes can be classified based either on their selective pharmacological recognition of different nucleotides or on their selective functional coupling to particular classes of heterotrimeric G proteins. [Table 1](#) summarizes both the pharmacological and functional properties of the eight P2Y receptor subtypes.

Based on their functional coupling to G proteins, effector proteins, and second messenger cascades, P2Y receptors can be broadly subdivided into three major groups. The P2Y1, P2Y2, P2Y4, and P2Y6 receptor subtypes predominantly couple to G_q -family G proteins with consequent activation of phosphatidylinositol-specific phospholipase C effector (PI-PLC) enzymes resulting in accumulation of inositol trisphosphate, diacylglycerol, Ca^{2+} as second messengers. The P2Y12, P2Y13,

Table 1 Classification of P2Y family nucleotide receptors

Receptor subtype	G protein–effector cascade	Selectivity for physiological nucleotide agonists	Tissue/cellular expression
P2Y1	$G_q \rightarrow PI-PLC$	$ADP > ATP \gg UDP, UTP$	Platelets; blood vessels, brain
P2Y2	$G_q \rightarrow PI-PLC$	$ATP = UTP \gg ADP, UDP$	Airways; exocrine glands; white blood cells; blood vessels, brain
P2Y4	$G_q \rightarrow PI-PLC$	$UTP \gg ATP, UDP, ADP$ (human) $UTP = ATP \gg UDP, ADP$ (rat, mouse)	Small intestine; inner ear
P2Y6	$G_q \rightarrow PI-PLC$	$UDP > UTP > ADP \gg ATP$	Gallbladder; airway
P2Y11	$G_q \rightarrow PI-PLC$ and $G_s \rightarrow AC$	$ADP > ATP \gg UDP, UTP$	Exocrine glands; kidney; white blood cells
P2Y12	$G_i \rightarrow AC/others$	$ADP > ATP \gg UDP, UTP$	Platelets; brain astrocytes
P2Y13	$G_i \rightarrow AC/others$	$ADP > ATP \gg UDP, UTP$	Brain astrocytes
P2Y14	$G_i \rightarrow AC/others$	$UDP-glucose \gg UTP, ATP, UDP, ADP$	White blood cell precursors; stem cells

and P2Y14 subtypes activate G_i -family G proteins. As a result, stimulation of these latter receptors most often leads to inhibition of adenylyl cyclase effector enzymes and reduced intracellular levels of cyclic AMP. However, depending on cellular background, these G_i -coupled P2Y receptor subtypes may also stimulate the activity of G-protein-regulated K^+ channels and/or inhibit activity of N-type voltage-gated Ca^{2+} channels. Finally, the P2Y11 receptor represents a subtype that can couple to both G_q - and G_s -dependent signaling pathways; the activation of G_s allows P2Y11 receptors to stimulate adenylyl cyclase effector enzymes and thereby induce accumulation of cyclic AMP. Thus, activation of the different P2Y receptor subtypes can variously trigger rapid changes in Ca^{2+} , cyclic AMP, lipid second messengers, and plasma membrane potential. In turn, these acute signals can be integrated within cell-specific transduction networks to regulate particular protein kinases, protein phosphatases, proteases, and transcription factors.

Pharmacological Classification of P2Y Receptors

P2Y receptors can also be categorized by their pharmacological selectivity for particular nucleotide agonists. The P2Y1, P2Y12, and P2Y13 exhibit very high selectivity for ADP over ATP and are virtually insensitive to nonadenine nucleotides. The P2Y2, P2Y4, and P2Y6 subtypes can be effectively activated by uridine nucleotides albeit with subtype-selective preferences. P2Y2 receptors show high affinity for UTP and ATP, but not the corresponding diphosphates. P2Y4 receptors are most potently stimulated by UTP over ATP (but this is species dependent), while P2Y6 receptors exhibit a preference for UDP. In general, P2Y11 receptors are activated by ATP or ATP, and less effectively by nonadenine nucleotides. The P2Y14 subtype, the most recently identified member of the P2Y family, exhibits a very high selectivity for sugar conjugates of UDP, such as UDP-glucose. Given the complex and overlapping affinities of the P2Y subtypes for naturally occurring nucleotides, considerable effort is being directed toward generation of synthetic nucleotide agonists and non-nucleotide antagonists that selectively target particular P2Y subtypes. Such subtype-selective reagents may have high potential utility as therapeutic agents in diseases that involve hyperfunctional or hypofunctional P2Y receptors.

Important Physiological or Pathological Functions of P2Y Receptor Subtypes

Understanding of the most important physiological roles for P2Y receptors has been limited by: (1) the relatively recent identification of most P2Y receptor subtypes (post 1995); (2) the current paucity of subtype-selective drugs; and (3) the rather broad expression of P2Y receptors in most tissue types. However, gene-targeting approaches have been recently used to generate knockout mice that lack the expression of P2Y1, P2Y2, P2Y4, P2Y6, or P2Y12 receptors. These mice (or tissues from the mice) exhibit significant deficits in several crucial biological responses.

Mice lacking either the P2Y1 or P2Y12 receptor are characterized by abnormalities in the blood clotting response (also known as hemostasis) that result in prolonged bleeding times. This reflects the synergistic roles of the P2Y1 and P2Y12

receptors in mediating the aggregation of platelets (the key blood cell type in the clotting reaction) in response to ATP/ADP released from damaged blood vessels. Because platelets also release large amounts of ATP/ADP during aggregation induced by other hemostatic factors (such as thrombin), P2Y1 and P2Y12 play additional roles in amplifying and extending overall blood clotting response. However, inappropriate platelet aggregation (thrombosis) can lead to interruption of normal blood flow (ischemia) to critical organs, such as the brain or heart, and thereby contribute to acutely life-threatening conditions, including stroke and heart attack. For this reason, drugs that specifically block or antagonize P2Y1 and/or P2Y12 receptors are currently being tested for the treatment of thrombotic diseases in humans.

Analyses of other knockout mice strains have revealed important roles for the P2Y2 receptor that are highly expressed in airway cells and the P2Y4 receptors expressed in the jejunal section of the small intestine. Multiple P2Y receptor subtypes are expressed in the epithelial cells from organs such as the airways, lungs, kidneys, and intestines. The major function of epithelial cells and tissues is to act as both a barrier and selectively permeable interface between very different biological compartments. For example, airway epithelial cells comprise an interface between the inspired air compartment (which contains microbes, small particles, and volatile compounds from the outside environment) and the blood compartment. In addition to physically blocking the entry of these airborne microbes, particles, compounds into the blood, airway epithelial cells secrete water, salt, mucus, and other protective proteins onto their extracellular surface that directly faces the inspired air compartment. While the secreted mucus solution acts to trap the microbes and particles, ciliated protrusions that extend from the epithelial cells beat in a synchronized manner to push the mucus solution – plus the entrapped microbes and particles – back to the nose and upper throat for eventual removal by sneezing or coughing. Activation, by ATP or UTP, of the P2Y2 receptors in different airway epithelial cells can coordinately regulate: (1) the secretion of the water, different salts, and mucus in sufficient amounts to generate mucus solutions with appropriate viscosity, fluidity, and composition; and (2) the frequency and direction of the ciliary beating required for the appropriate retrograde flow of the mucus back to the upper airways. In airways from the P2Y2 receptor knockout mice, these important epithelial functions are not appropriately regulated. Analysis of small intestine function in P2Y4 receptor–receptor mice has indicated a similar role for this P2Y subtype in the regulation of salt and fluid secretion by intestinal epithelial cells.

P2Y2 receptor deficiency also results in reduced clearance of apoptotic T-cell progenitors in the thymus because ATP released from the apoptotic thymocytes acts as a find-me signal for the P2Y2 receptor-dependent chemotaxis of phagocytic macrophages. The absence of P2Y6 receptors also perturbs normal macrophage function by reducing the production of inflammatory cytokines in response to endotoxin/lipopolysaccharide.

Future studies of additional organ- and tissue-specific functions in mice lacking these and other P2Y receptor subtypes will undoubtedly reveal other important *in vivo* roles for these receptors. Insights from these animal studies, coupled with the

development of subtype-selective antagonists, should facilitate delineation of how these receptors function in human physiology and pathology.

See also: [Signaling: Adenylyl Cyclases](#); [G Protein Signaling Regulators](#); [Neurotransmitter Transporters](#).

Further Reading

Abbraccio MP, Burnstock G, Boeynaems JM, et al. (2006) International union of pharmacology LVIII: Update on the P2Y G protein-coupled nucleotide receptors:

From molecular mechanisms and pathophysiology to therapy. *Pharmacological Reviews* 58: 281–341.
Bhatt DL and Topol EJ (2003) Scientific and therapeutic advances in antiplatelet therapy. *Nature Reviews – Drug Discovery* 2: 15–28.
Dubyak GR (2003) Knock-out mice reveal tissue-specific roles of P2Y receptor subtypes in different epithelia. *Molecular Pharmacology* 63: 773–776.
Elliott MR, Chekeni FB, Trampont PC, et al. (2009) Nucleotides released by apoptotic cells act as a find-me signal to promote phagocytic clearance. *Nature* 461: 282–286.
Lazarowski ER, Boucher RC, and Harden TK (2003) Mechanisms of release of nucleotides and integration of their actions as P2X- and P2Y-receptor activating molecules. *Molecular Pharmacology* 64: 785–795.
Zimmermann H (2000) Extracellular metabolism of ATP and other nucleotides. *Naunyn-Schmiedeberg's Archives of Pharmacology* 362: 299–309.

p53 Family

C E Barton and J A Pietenpol, Vanderbilt University, Nashville, TN, USA

Published by Elsevier Inc.

This article is a revision of the previous edition article by Jamie Hearnese and Jennifer Pietenpol, volume 3, pp. 192–195, © 2004, Elsevier Inc.

Glossary

Apoptosis Programmed cell death; an active form of cell suicide.

Cell cycle The ordered set of events by which a cell grows and divides.

Cell-cycle checkpoint Point in the cell-division cycle where progression through the cycle can be halted until

conditions are suitable for the cell to proceed to the next stage.

Oncogene A gene whose product has the ability to transform a normal cell into a cancer cell.

Tumor suppressor A gene that negatively regulates cell division and when lost allows the cell to progress through the cell cycle in an unrestricted manner.

p53 is a tumor-suppressor gene that is mutated in over half of all human cancers. The tumor-suppressive functions of p53 are dependent upon its activity as a transcription factor. This protein binds DNA in a sequence-specific manner to regulate the expression of select genes involved in a range of cellular processes including growth arrest, apoptosis, DNA repair, autophagy, senescence (irreversible growth arrest), and angiogenesis (new blood vessel formation). The p53 protein is activated in response to various cellular stresses and functions to suppress the transformation of a normal cell into a cancer cell. In addition, p63 and p73 are two transcription factors that are highly similar to p53 in both DNA sequence and protein structure. Collectively, p53, p63, and p73 are commonly referred to as the p53 family. Like p53, p63 and p73 act to regulate growth arrest and apoptosis. However, unlike p53, p63 and p73 also play important roles during early development.

Functional Domains of p53

The p53 protein was originally identified in 1979, and its name is derived from the apparent molecular weight of the protein (p53 = protein of 53 kDa). This protein consists of 393 amino acids and is divided into three functional domains: (1) an N-terminal transactivation domain; (2) a DNA-binding domain; and (3) an oligomerization domain (Figure 1(a)).

Transactivation Domain

The transactivation domain is located at the amino-terminus of the protein sequence (amino acids 1–73). Because the p53 gene contains multiple transcriptional start sites, the p53 protein can either contain the transactivation domain (denoted p53) or lack a full-length transactivation domain (Figure 1(a)). The full-length p53 protein is the primary isoform expressed in human cells, and this isoform is responsible for its tumor-suppressor function. Because they lack a full-length transactivation domain, $\Delta 40$ p53 and $\Delta 133$ p53 can inhibit the tumor-suppressor function of the full-length p53. This inhibition can occur by the $\Delta 40$ p53 and $\Delta 133$ p53 oligomerizing with full-length p53 and blocking its ability to activate transcription, or by independently binding p53 DNA response elements and

competing with full-length p53 for these gene promoters. Numerous proteins can bind the transactivation domain of p53 to regulate its function. This region is highly phosphorylated by kinases (Figure 1(b)), and this phosphorylation is important for controlling the stability and activity of p53 after various stimuli (see section Phosphorylation).

DNA-Binding Domain

The central region of the p53 protein, the DNA-binding domain (amino acids 100–293), is responsible for mediating the interaction between p53 and its respective DNA-binding sites. A majority of tumor-derived p53 mutations occur in the DNA-binding domain. The residues most commonly mutated in the DNA-binding domain either directly contact the DNA or are required for the structural integrity of the protein. Since p53 primarily functions as a transcription factor, these mutations result in a loss of function of the protein.

Oligomerization Domain

The carboxy-terminus of the p53 protein (amino acids 293–393) contains an oligomerization domain, which is required for the formation of transcriptionally active dimers and tetramers. The oligomerization domain also contains nuclear localization and export sequences that are required for transport into and out of the nucleus, respectively. The carboxy-terminus is also posttranslationally modified by multiple other proteins that regulate p53 stability and transcriptional activity (Figure 1(b)).

p53-Mediated Signaling

p53 functions as a sequence-specific transcription factor regulating select genes that are involved in numerous signaling pathways, including cell-cycle progression, apoptosis, autophagy, senescence (irreversible growth arrest), DNA repair, and angiogenesis (new blood vessel formation) (Figure 2). The p53 consensus binding site consists of two copies of the DNA sequence RRRCWWGYYY which are separated by a 0–13 base spacer region (R = A/G, W = A/T, Y = T/C). This binding site is

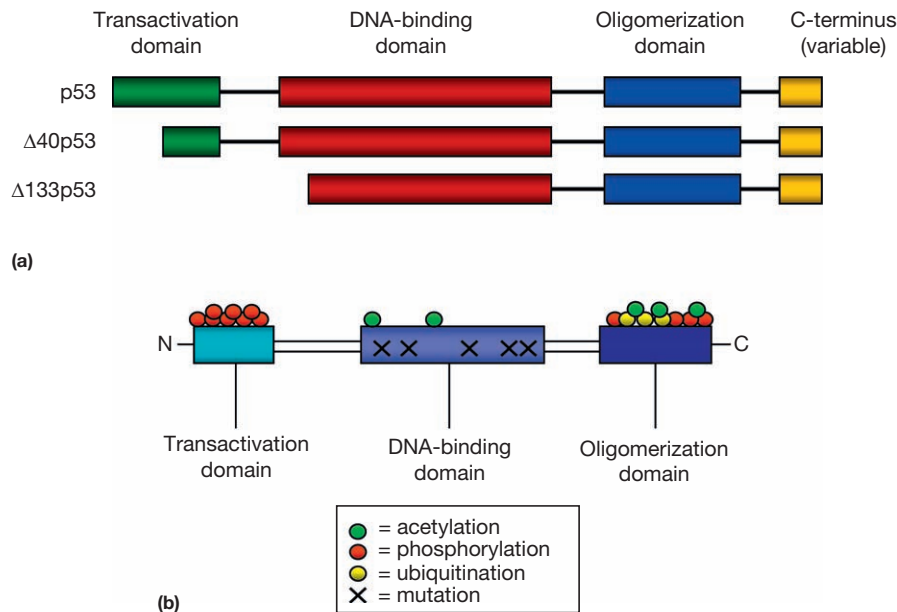


Figure 1 (a) Full-length p53 protein had several functional domains, including a transactivation domain, a DNA-binding domain, and an oligomerization domain. Differential RNA splicing can produce multiple protein isoforms that differ in the C-terminus of the protein. Alternative promoter usage produces p53 transcripts that differ only in the length of transactivation domain encoded. (b) p53 biochemical activities are affected by posttranslational modifications and mutations. The amino-terminus contains the transactivation domain that can be phosphorylated in response to certain types of cell damage. The central region of the protein contains the DNA-binding domain that is frequently mutated in human tumors. The oligomerization domain mediates the formation of p53 dimers and tetramers, is present at the carboxy-terminus of the protein, and can be modified posttranslationally by phosphorylation, acetylation, and ubiquitination.

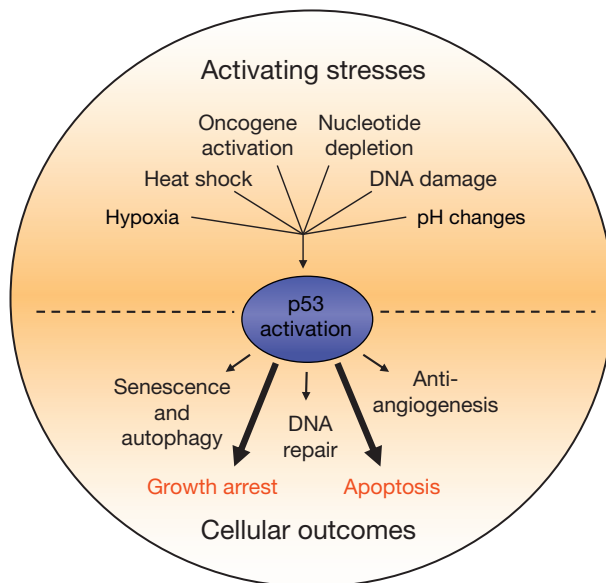


Figure 2 p53 signaling pathways. p53 is activated by many cellular insults (activating stresses) that affect numerous pathways yielding various cellular outcomes. The two primary, most well-characterized outcomes of p53 activation are growth arrest and/or apoptosis (thick arrows). However, p53-regulated genes also take part in numerous other signaling pathways (thin arrows).

present in a majority of p53 target gene promoters and/or regulatory regions. The degeneracy of the consensus DNA binding site at eight of the 10 bases is thought to contribute to the selectivity of p53 toward certain genes in response to

various stresses. p53 can serve as a transcriptional activator or repressor, depending on the specific gene being regulated.

Growth Arrest and Apoptosis

Throughout life, human cells are exposed to agents that can induce DNA damage. Eukaryotic cells have evolved a series of surveillance pathways termed 'cell-cycle checkpoints' to ensure that cells copy and divide their genomes with high fidelity during each replication cycle. These checkpoint signaling pathways execute several important tasks, such as rapid induction of cell-cycle delay, activation of DNA repair, maintenance of cell-cycle arrest until repair is complete, re-initiation of cell-cycle progression after completed repair or initiation of apoptosis if the damage is irreparable. These checkpoints provide a fail-safe mechanism by which cells are repaired or eliminated before the damaged DNA is transferred to the daughter cells. Loss of normal cell-cycle checkpoint signaling is a hallmark of tumor cells, and p53 is the most frequently altered cell-cycle checkpoint signaling molecule. One of the major functions of the p53 tumor-suppressor protein is to modulate cellular responses to stress and induce cell-cycle arrest or death, as appropriate (Figure 2). Once a cell has been stressed or damaged, p53 initiates cell-cycle arrest allowing time for the damage to be repaired or for the stress to dissipate. p53 halts cell growth by selectively activating certain target genes, including p21^{Waf1/Cip1}, a protein that inhibits other proteins required for cell-cycle progression. If cell damage is extensive and cannot be repaired or the stress is prolonged, p53 will then activate genes to mediate apoptosis. p53 is responsible for both the activation of pro-apoptotic genes as well as the repression of

anti-apoptotic genes, and the array of p53-regulated genes involved in apoptosis is quite extensive. Although a given cell in the human body would not benefit from undergoing apoptosis, this process is sometimes necessary for the benefit of the whole organism to prevent the development of cancer.

Other Pathways Regulated by p53

p53 function is not limited to growth arrest and apoptosis. p53 also regulates genes involved in DNA repair, senescence, cellular migration, autophagy, and angiogenesis (Figure 2). Although the relationships between p53 and these various pathways are not completely understood, p53 controls these cellular functions through transcriptionally regulating downstream target genes. Future discoveries of novel p53 target genes and protein interactions will further elucidate the signaling networks in which p53 is involved.

p53 Regulation

p53 must act immediately after cell stress occurs and, thus, it is critical that the regulatory mechanisms for this protein are rapid. Following stress, the levels of many proteins in the cell are dictated by transcriptional regulation of the genes encoding the proteins; however, this is not the primary mechanism for p53. Instead, p53 regulation occurs predominantly at the protein level, which allows for rapid induction of p53 activity.

Activating Stresses

In a normal unstressed cell, the p53 protein is present at low levels due to the short half-life of the protein. However, p53 protein levels increase if the cell experiences a stress. There are numerous types of cellular stress and damage that can activate the p53 protein including lack of oxygen (hypoxia), elevated/lowered temperatures (heat/cold shock), DNA damage, metabolic changes, nucleotide depletion, pH changes, and oncogene activation (Figure 2). Cellular stresses trigger signaling pathways that converge on p53, resulting in stabilization and activation of the protein.

Posttranslational Modifications

Posttranslational modification events are dynamic processes that change protein stability and activity by altering protein–DNA and protein–protein interactions. Depending on the stress, p53 is posttranslationally modified with various moieties resulting in diverse effects on the protein's stability, activity, and specificity of DNA binding (Figure 1(b)).

Ubiquitination

Ubiquitination is the addition of ubiquitin molecules to lysine residues of a protein. Following ubiquitination, most proteins are targeted to the 26S proteasome for degradation. This is the mechanism used to rapidly turn over the p53 protein. The ubiquitination system involves numerous proteins, but the specificity of the system depends on the specific E3 ubiquitin ligase enzyme employed, which attaches an ubiquitin molecule to the correct substrate. The mouse double minute 2 (Mdm2) protein is the predominant E3 ligase that can

ubiquitinate p53, and is a key regulatory protein involved in controlling p53 protein levels.

Phosphorylation

Phosphorylation of proteins at serine, threonine, and/or tyrosine residues is a common regulatory mechanism in the cell. p53 is phosphorylated at select serine and threonine amino acids in the amino- and carboxy-terminal regions of the protein in response to select cellular stresses. These stresses include exposure of cells to byproducts of cellular metabolism, ultraviolet (UV) radiation, and drugs used in cancer therapy. Numerous kinases have been identified that phosphorylate p53. The specific kinase activated is dependent upon the stress, and each kinase is responsible for phosphorylating a distinct residue(s) on the protein. For example, p53 is phosphorylated on serine-15 and serine-20 after DNA damage and this modification is important for p53 activity. Phosphorylation of the p53 protein can yield numerous outcomes including protein stabilization, increased transcriptional activity, and alterations in subcellular localization.

Acetylation

Acetylation is a modification that can occur on lysine residues of a protein. Acetylation of a transcription factor can affect the protein's transcriptional activity, either positively or negatively, depending on the specific factor. In response to DNA damage, known transcriptional coactivators acetylate p53 at certain carboxy-terminal lysines, and this modification is important for the transcriptional activation of p53.

Protein–Protein Interactions

There are numerous proteins that regulate p53 through a physical interaction. Many of these proteins are involved in post-translationally modifying the protein, but there are other mechanisms of p53 protein regulation that involve protein–protein interactions. These interactions can both positively and negatively regulate p53.

Mdm2 is a protein that can bind and ubiquitinate p53, as mentioned above. This p53–Mdm2 interaction results in two outcomes. The first is that Mdm2 inhibits the transcriptional activity of p53 by binding to the amino-terminal transactivation domain of the p53 protein, blocking other interactions that are necessary for p53 activity. The second outcome is the ubiquitination of p53 by MDM2, which results in export of p53 from the nucleus and its subsequent degradation. Mdm2 is responsible for maintaining p53 at low levels in a normal unstressed cell, as well as attenuating the p53 response after stress. Interestingly, Mdm2 is a p53 transcriptional target upregulated by p53 in response to cellular insults, forming an autoregulatory negative feedback loop, which is the primary mechanism of negative regulation for p53.

Of note, viral proteins can also interact with the p53 protein to affect its activity. For instance, the human papilloma virus (HPV) type 16 and 18 E6 proteins negatively regulate p53 in a manner similar to Mdm2. The E6 protein binds to a protein complex containing p53, which leads to p53 ubiquitination and subsequent degradation. As a result, there is a significant correlation between the presence of HPV and multiple human cancers, in particular cervical, head and neck, and anal cancers.

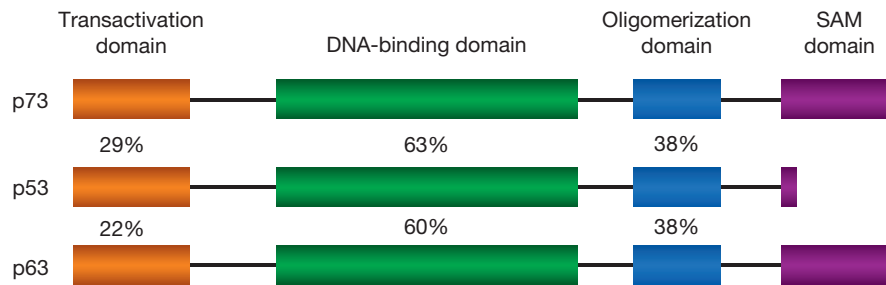


Figure 3 The p53 family consists of p53, as well as its ‘siblings’ p63 and p73. p63 and p73 are highly similar to p53 structurally in that they both consist of a transactivation domain, a DNA-binding domain, and an oligomerization domain. In addition, p63 and p73 contain a C-terminal sterile-alpha motif (SAM) domain that is used primarily for protein–protein interactions. The amino acid identity between p53 and its family members is significant, with the highest sequence homology present in the DNA-binding domain.

In addition to protein interactions that negatively regulate p53 activity, there are interactions that have a positive impact on the protein as well. One example of this is the human p14^{ARF} protein. This protein is a tumor suppressor activated in response to oncogenic stimuli. Once activated, p14^{ARF} binds to the p53–Mdm2 complex, which disables Mdm2. This protects p53 from Mdm2-mediated degradation, resulting in p53 activation and subsequent growth suppression or apoptosis of cells that contain activated oncogenes.

p53 Family Members

p53 is a member of a newly identified family of related proteins that includes p73 and p63. p73 was discovered in 1997, with p63 quickly following in 1998. The three family members exhibit a high degree of sequence similarity and conserved structural domains (Figure 3). However, unlike p53, it appears that the functions of p73 and p63 are not entirely restricted to tumor suppression.

p63 and p73 Protein Structure

Like p53, both the *p63* and *p73* genes have two transcriptional start sites, that allow for the production of multiple proteins from a single gene that either contain (TA) or lack (ΔN) the transactivation domain at the amino terminus. Differential RNA splicing creates additional isoforms that differ at the carboxy terminus. Due to the lack of a functional transactivation domain, ΔNp63 and ΔNp73 isoforms act primarily as transcriptional repressors and can block the activity of other family members by the formation of protein–protein interactions (Figure 4).

p63 and p73 During Development

p63 and p73 each play important, yet very different, roles during development. p73 is involved in neuronal signaling, immune cell apoptosis, and inflammation. p63 is required for proper development of complex epithelial tissues, such as the epidermis, mammary tissue, and prostate and salivary glands. In addition to development, p63 also controls processes such as cellular adhesion and migration.

p63 and p73 in Cancer

Due to the presence of multiple protein isoforms, the roles of each p53 family member in cancer is complex and, often,

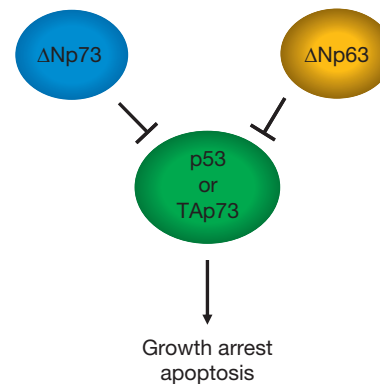


Figure 4 General model for the p53 family signaling in cancer. p53 and TAp73 function as tumor suppressors by activating downstream target genes that function in growth-inhibitory processes, such as cell cycle arrest and apoptosis. In some tumor cells, ΔNp63 or ΔNp73 isoforms can block the activity of p53 and TAp73 isoforms through competition for sequence-specific DNA-binding sites or by direct interaction with the protein, respectively. These antagonistic activities are hypothesized to prevent p53- and TAp73-mediated growth arrest and apoptosis, promoting ‘unchecked’ proliferation of cancer cells.

confusing. p53 acts as a tumor suppressor to block the malignant transformation of a cell, and the *p53* gene is mutated in over 50% of human tumors. TAp73 isoforms also maintain a role as tumor suppressors, due to their ability to activate a number of p53 target genes. p63, on the other hand, is often expressed as the ΔNp63 isoform and can block the ability of its other family members to act as tumor suppressors, thus giving ΔNp63 an oncogene-like activity (Figure 4).

Cancer Therapeutics Involving p53, p63, and p73

Since p53 is mutated in over 50% of all human tumors, many therapeutic strategies have been proposed to target p53 and its family signaling pathway, including: (1) transfer of normal p53 genes into mutant p53-containing cancer cells, (2) introduction of small molecules that convert mutant p53 proteins to proteins that are structurally normal in function, (3) identification of small molecules that mimic the function of p53 target genes, and (4) disruption of protein–protein interactions that negatively regulate p53 activity. In addition, small compounds that either (1) block the ability of ΔNp63 to inhibit

p53 and/or TAp73, or (2) induce the pro-apoptotic activity of TAp73 in tumors may also be successful in blocking cancer progression. One example of this is the use of rapamycin, a drug that blocks the activity of the mammalian target of rapamycin (mTOR) protein kinase. Because mTOR functions to negatively regulate the function of TAp73, compounds such as rapamycin can increase the stability and activity of p73 and activate apoptosis in human cancers. Given its important role in tumorigenesis, deciphering the function, regulation, and signaling pathways of the p53 family will continue to be a major focus of cancer-based research.

See also: **Bioenergetics:** Cell Death by Apoptosis and Necrosis; **Cell Architecture and Function:** Cell Cycle: DNA Damage Checkpoints; **Molecular Biology:** DNA Base Excision Repair Pathways; Transcription-Coupled DNA Repair Overview.

Further Reading

- Alberts B, Bray D, Lewis J, et al. (eds.) (1994) *Molecular Biology of the Cell*, 3rd edn. New York, NY: Garland Publishing.
- Appella E and Anderson CW (2001) Post-translational modifications and activation of p53 by genotoxic stresses. *European Journal of Biochemistry* 268: 2764–2772.
- El-Deiry WS (1998) Regulation of p53 downstream genes. *Seminars in Cancer Biology* 8: 345–357.
- Faridi R, Zahra A, Khan K, and Idrees M (2011) Oncogenic potential of human papillomavirus (HPV) and its relation with cervical cancer. *Virology Journal* 8: 269.
- Giaccia AJ and Kastan MB (1998) The complexity of p53 modulation; emerging patterns from divergent signals. *Genes and Development* 12: 2973–2983.
- Gillison ML, Chaturvedi AK, and Lowy DR (2008) HPV prophylactic vaccines and the potential prevention of noncervical cancers in both men and women. *Cancer* 113: 3036–3046.
- Lane PL and Lain S (2002) Therapeutic exploitation of the p53 pathway. *Trends in Molecular Medicine* 8: S38–S42.
- Leslie M (2011) Brothers in arms against cancer. *Science* 6024: 1551–1552.
- May P and May E (1999) Twenty years of p53 research: Structural and functional aspects of the p53 protein. *Oncogene* 18: 7621–7636.
- Pietsch EC, Sykes SM, McMahon SB, and Murphy ME (2008) The p53 family and programmed cell death. *Oncogene* 27: 6507–6521.
- Rosenbluth JM, Mays DJ, Jiang A, Shyr Y, and Pietenpol JA (2011) Differential regulation of the p73 cistrome by mammalian target of rapamycin reveals transcriptional programs of mesenchymal differentiation and tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America* 5: 2076–2081.
- Rosenbluth JM, Mays DJ, Pino MF, Tang LJ, and Pietenpol JA (2008) A gene-signature-based approach identifies mTOR as a regulator of p73. *Molecular and Cellular Biology* 19: 5951–5964.
- Stewart ZA and Pietenpol JA (2001) p53 signaling and cell-cycle checkpoints. *Chemical Research in Toxicology* 14: 243–263.
- Yang A and McKeon F (2000) p63 and p73: p53 mimics, menaces, and more. *Nature Reviews Molecular Cell Biology* 1: 199–207.

Parathyroid Hormone/Parathyroid Hormone-Related Protein Receptor

T J Gardella, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

G (guanine nucleotide-binding)-protein-coupled receptor (GPCR) An integral membrane protein characterized by seven membrane-spanning helical domains and the capacity to activate heterotrimeric G proteins in response to agonist binding.

Parathyroid hormone (PTH) An 84-amino-acid secreted polypeptide that functions as the major regulator of calcium-ion concentrations in blood and extracellular fluids; acts on bone and kidney cells.

PTH-related protein (PTHrP) A 141-amino-acid polypeptide that acts in a paracrine fashion to control development of the skeleton, heart, teeth, mammary glands, and other tissues. It is highly expressed in breast milk and is the usual mediator of hypercalcemia of malignancy.

PTH/PTHrP receptor (PPR, or PTH-1 receptor) A class 2 GPCR that mediates the actions of PTH and PTHrP. Functionally expressed in osteoblasts, renal proximal, and distal tubule cells and a variety of tissues during development.

Definition

The parathyroid hormone (PTH)/PTH-related protein (PTHrP) receptor (PPR) is a 593-amino-acid, class 2 G-protein-coupled receptor (GPCR) that mediates the actions of PTH, a peptide hormone that functions as a major regulator of blood-ionized calcium and phosphate levels, and PTHrP, a key developmental protein. The PPR is expressed in bone osteoblasts and in renal proximal and distal tubule cells, the key target cells of PTH, and in mesenchymal cells of developing tissues (e.g., skeleton, heart, teeth, and mammary glands), the key target cells of PTHrP. The PPR interacts with the bioactive (1–34) regions of its two peptide ligands, PTH and PTHrP, via a bipartite mechanism that involves initial docking interactions to the amino-terminal extracellular domain (ECD) of the receptor, and subsequent signaling interactions to the extracellular loops (ECLs) and seven transmembrane domain helical region of the receptor. The PPR couples strongly to the G_{α_s} /adenylyl cyclase (AC)/3',5'-cyclic-adenosine monophosphate (cAMP)/protein kinase A (PKA) signaling cascade, and can also activate the $G_{\alpha_q/11}$ /phospholipase C (PLC)/inositol triphosphate (IP_3)/intracellular calcium (iCa^{2+})/PKC pathway. The PPR has also been shown to activate the $G_{\alpha_{12/13}}$ /RhoA/PLD/phosphatidic acid/PKC pathway, as well as the map kinase pathway, the latter via G-protein-dependent and G-protein-independent mechanisms. Jansen's metaphyseal chondrodysplasia and neonatal-lethal Blomstrand's chondrodysplasia are rare disorders of skeletal development and calcium-ion homeostasis that are caused by activating and inactivating mutations in the PPR gene, respectively. PTH peptides have anabolic effects on bone, and the bioactive PTH(1–34) fragment, as well as the intact PTH(1–84) peptide, are now in use to treat osteoporosis.

Background and Physiology

Molecular Cloning of the PPR

The complementary (c)DNA encoding the PTH/PTHrP receptor was first isolated in 1991. Two cell lines were the sources for the messenger RNA (mRNA) – the opossum kidney line, OK,

and the rat osteosarcoma cell line, ROS17/2.8. The coding sequence of the PPR revealed the classical seven transmembrane domain architecture of the GPCR super family (Figure 1). It was clear that the PPR was most homologous to the recently cloned receptors for secretin and calcitonin, and that together these receptors formed a distinct GPCR subgroup, termed the class 2 (or family B) GPCRs. This GPCR subgroup is comprised of 15 distinct peptide hormone receptors, and is well-conserved evolutionarily, having representatives present in the protochordate tunicate, *Ciona intestinalis*, the arthropod, *Drosophila melanogaster*, and the nematode, *Caenorhabditis elegans*. The cloned PPR expressed in COS-7 cells binds PTH and PTHrP with similar affinities, and each ligand activates the G_{α_s} /AC/cAMP/PKA and G_{α_q} /PLC/ IP_3 / iCa^{2+} -signaling pathways with similar potencies. The same receptor is, thus, thought to mediate most, if not all, of the biological actions of the two distinct peptide ligands.

Parathyroid Hormone

Discovery of PTH

PTH was identified as a key calcium-regulating hormone in the 1920s, when it was demonstrated that extracts of the parathyroid glands could modulate blood calcium levels and correct the tetany induced by parathyroid gland removal. The active hormone was purified from glands in 1959 and the complete 84-amino-acid sequence of bovine PTH was determined in 1970. In 1971, the N-terminal PTH(1–34) fragment was synthesized and shown to be equipotent to native PTH in *in vitro* and *in vivo* assay systems. Physiological studies performed during this period established that PTH is the most important regulator of ionized calcium (Ca^{2+}) in the blood and extracellular fluids, and that it acts on bone, where it stimulates Ca^{2+} release from the mineralized matrix, and kidney, where it stimulates Ca^{2+} reabsorption from the glomerular filtrate. Related studies established that the secretion of PTH from the parathyroid glands is precisely regulated in negative-feedback fashion by the extracellular Ca^{2+} concentration; thus, increases in blood Ca^{2+} result in decreased PTH secretion and decreases in blood Ca^{2+} result in increased

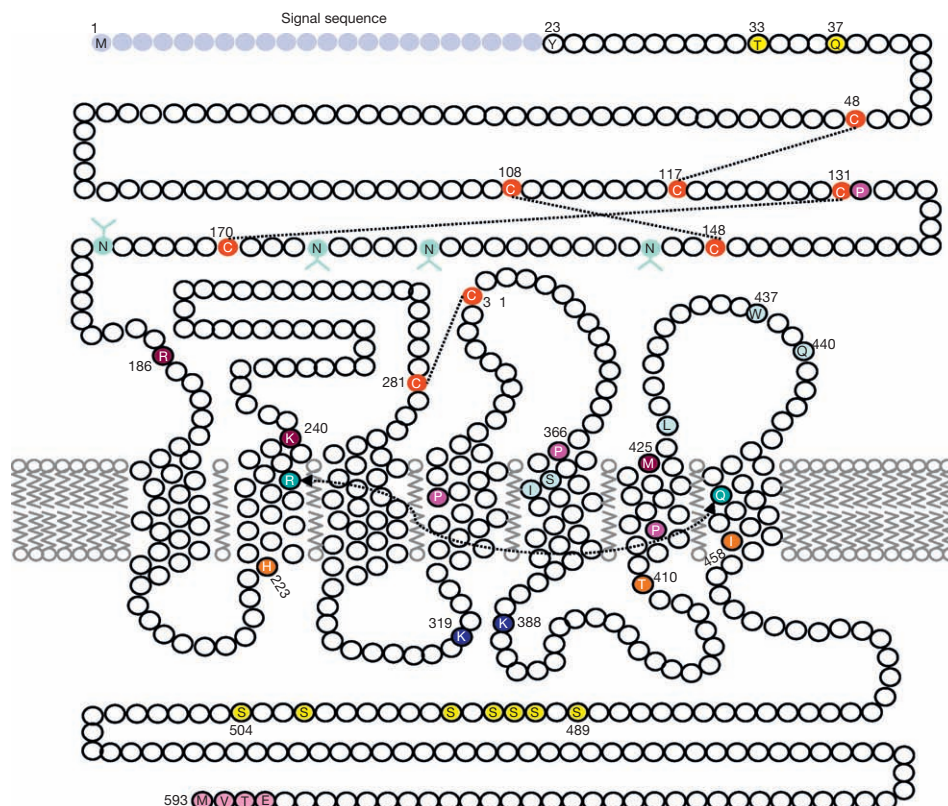


Figure 1 The human PTH/PTHrP receptor. Shown is a schematic of the PPR embedded in the lipid bilayer with selected features identified, including the conserved extracellular cysteines and their disulfide linkage arrangement; the glycosylated asparagines (N); Thr33 and Gln37, which contribute to PTH(15–34)-binding affinity; the probable cross-linking sites for ligand residues 13 (Arg186), 19 (Lys240), and 1 and 2 (Met425); conserved prolines (P), including Pro132, which is mutated to Leu in Blomstrand's chondrodysplasia; a putative interhelical interaction between Arg233 and Gln451; Ser370, Ile371, Leu427, Trp437, and Gln440, at which mutations alter responsiveness to position 2-modified PTH(1–34) antagonist ligands (also includes Met425); His223, Thr410, and Ile458, which are mutated in Jansen's chondrodysplasia to result in constitutive-signaling activity; Lys319 and Lys388, at which mutations impair G_q and G_s and G_q coupling, respectively; serines (S) that are phosphorylated upon agonist activation, and four C-terminal residues (Glu590–Met593) involved in NHERF binding.

secretion. This Ca^{2+} -mediated regulation of PTH secretion is determined by the actions of the calcium-sensing receptor, a class 3 GPCR expressed on the surface of parathyroid cells. The regulation is a highly sensitive process, such that blood Ca^{2+} levels are strictly maintained within the range of 1.1–1.3 mM. PTH is also a key regulator of blood phosphate, the principal Ca^{2+} counter-ion, and acts to suppress the reabsorption of phosphate in the kidney.

PTH effects in bone

In bone, the PPR is expressed by the bone-forming osteoblasts, as well as by osteocytes, the terminally differentiated stage of osteoblasts that are enclosed within the lacunae–canaliculi system of mineralized bone. Studies in osteoblasts have shown that a variety of responses to PTH, including alterations in the rates of cell proliferation, differentiation, and apoptosis, as well as changes in expression patterns for a number of genes encoding transcription factors, cytokines, and bone-matrix factors. These effects on osteoblasts can result in increased rates of bone formation. PTH, however, also induces osteoblasts to increase their expression of RANK-Ligand, a potent osteoclastogenic factor, and to decrease their expression of osteoprotegerin (OPG), a secreted RANK-L decoy receptor. The resulting increase in

the RANK-L:OPG ratio causes an increase in the differentiation of cells of the osteoclast lineage, and stimulates the bone-resorbing functions of mature osteoclasts. This indirect effect of PTH on osteoclasts, via the osteoblasts, likely contributes strongly to the calcium-releasing effects that the hormone has on bone.

By modulating the activities of both osteoblasts (directly) and osteoclast (indirectly), PTH plays a major role in the bone-remodeling process. The mechanisms that control the balance between the anabolic and catabolic effects of PTH on bone are complex and not fully understood; however, it is well documented that intermittent exposure to PTH (once-daily injection) results in net bone formation, whereas continuous exposure (infusion) results in net bone resorption. Recent data have implicated an important role for the canonical Wnt/ β -catenin-signaling pathway in mediating the bone-anabolic effects of PTH on osteoblasts. Recent data also suggest that PTH can stimulate bone formation through effects on osteocytes, in which PTH suppress the expression of the *Sost* gene, which encodes the secreted glycoprotein sclerostin, which can function as an inhibitor of the Wnt/ β -catenin-signaling pathway in osteoblasts. The sum cellular and molecular pathways that mediate the anabolic and catabolic effects of PTH

on bone are clearly complex and full understanding will require much further study. In any case, the anabolic effects that intermittent PTH can have on the skeleton are clear, and indeed were confirmed by a large clinical study completed in 2001, which showed that once-daily administration of recombinant human (rh) PTH(1–34) significantly reduces the risk of bone fracture in patients with osteoporosis. This peptide is, thus, now marketed in the U.S. by Eli Lilly Co. as *Forteo*TM and intact rhPTH(1–84) is marketed in Europe by Nycomed as *Preotact*TM, as treatments for this disease.

PTH effects in the kidney

The PPR is expressed by cells in the proximal and distal portions of the renal-convoluted tubule. In distal tubule cells, PTH stimulates the transcellular reabsorption of calcium from the urine, in part by affecting the luminal surface expression and activity of the calcium channel protein TRPV-5 (transient receptor potential vanilloid-5). In proximal tubule cells, PTH suppresses the surface expression of the type II sodium-dependent phosphate transporters, NPT2a and NPT2c, and thereby decreases the reuptake of phosphate from the tubular filtrate. Also in the proximal tubule, PTH stimulates the synthesis of 1,25-dihydroxy-vitaminD₃, via upregulation of the 25-hydroxyvitaminD₃-1- α -hydroxylase gene. The increased levels of activated vitaminD₃, in turn, promote calcium and phosphate absorption at the intestine, and also result in increased expression of the phosphate-regulating cytokine, fibroblast growth factor-23 (FGF23) in osteoblasts and osteocytes. PTH, together with vitaminD₃ and FGF23, form a central hormonal feedback axis that functions to maintain the extracellular concentrations of calcium and phosphate at the steady-state levels required for normal biological function.

PTHrP

Discovery of PTHrP

In the late 1980s, much effort was devoted to identifying the hypercalcemia-causing factor that is released by many tumors and, in 1987, the protein factor responsible, now termed PTHrP, was purified from tumor extracts and partially sequenced. This sequence data, in turn, led to the cloning of the corresponding cDNA, which revealed a protein of 141 amino acids showing clear N-terminal homology to PTH, as eight of the first 13 residues of the two proteins were identical (**Figure 2**). Although PTHrP(1–34) peptides were shown to mimic the calciotropic actions of PTH on bone and kidney cells, it seemed clear that PTHrP was not a calcium-regulating hormone, as its concentration in the normal circulation was too low (barely detectable), and its mRNA was found in a wide variety of tissues (e.g., brain, lung, heart, adrenals, spleen, liver, bladder, skin, stomach, breast, placenta, ovaries, and testes), in both adult and fetal life. These observations suggested that PTHrP acts in a paracrine manner, possibly in development.

Genetic manipulations of PTHrP in mice

The above hypotheses regarding PTHrP action were largely confirmed in 1994 when mice having homozygous deletion of the PTHrP gene were generated and found to die at birth with defects in endochondral bone formation, namely, premature mineralization and short limbs. Subsequent studies in mice determined that PTHrP also plays a role in the

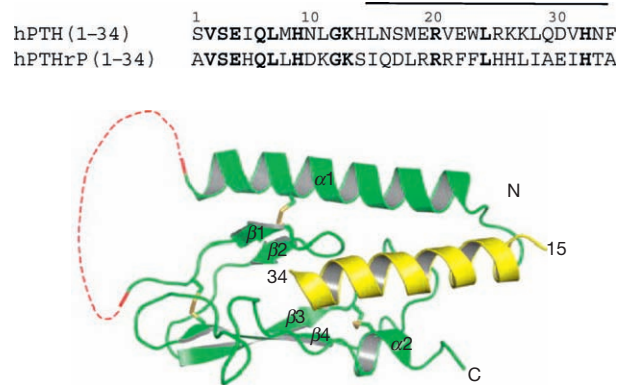


Figure 2 PTH and PTHrP. Shown is an alignment of the bioactive (1–34) regions of human PTH and human PTHrP, with conserved residues in bold-faced type; the bar indicates the C-terminal amphipathic α -helical region, residues 15–34. Also shown is the X-ray crystal structure, in ribbon representation, of the PPR-ECD (green, the exon E2 region not resolved is shown as a red dash) in complex with the PTH (15–34) domain (yellow), as reported by Pioszak and Xu in 2008.

development of the mammary glands, hair follicles, and the heart. In 1996, mutant mice lacking the PPR gene were generated, and this established that the PPR mediates the developmental actions of PTHrP, as the PPR-deficient mice exhibited nearly the same perinatal lethality and developmental defects as those lacking the PTHrP gene.

PTHrP actions in the growth plate

Extensive studies on the cartilaginous growth plates of the fetal tibia and femur in mice genetically altered for PTHrP expression demonstrate that the protein acts in a paracrine fashion to control cell differentiation and proliferation programs in the developing tissue. PTHrP is expressed by cells in the periarticular region of the growth plate and acts on prehypertrophic chondrocytes of the central zone to delay their terminal differentiation into hypertrophic chondrocytes. This action of PTHrP in the growth plate is further regulated by at least several other morphogenetic proteins, including Indian hedgehog (Ihh), which is produced by the prehypertrophic chondrocytes and stimulates chondrocyte proliferation. Ihh also induces, as part of a feedback regulatory network, the expression of PTHrP by the periarticular cells. Similar paracrine actions of PTHrP on PPR-expressing cells are also likely to operate in the mesenchyme of other developing tissues (e.g., heart, mammary, and hair follicles).

Structure–Activity Relationships in PTH and PTHrP

Functional Domains of PTH(1–34) and PTHrP(1–34)

Both PTH(1–34) and PTHrP(1–34) peptides bind to and activate the PPR with affinities and potencies in the nanomolar range, and the two ligands interact with largely overlapping sites in the receptor. For both ligands, the major determinants of PPR-signaling potency reside within the N-terminal (1–14) regions, while those for binding affinity reside in the C-terminal (15–34) portions. Based on this physical separation of functional domains, N-terminally truncated analogs, such as

[Leu¹¹,DTrp¹²]PTHrP(7–34), have been developed as PTH-receptor antagonists. The PTH(15–34) fragment represents the minimum-length-binding fragment ($K_{\text{dissociation}} \sim 1 \mu\text{M}$). Within this domain, Trp-23, Leu-24, Leu-28, and Val-31 function as key determinants of binding, and form the hydrophobic face of an amphipathic α -helix. Within the N-terminal signaling domain of PTH, the highly conserved Val-2, plays a key role in PPR activation, as its deletion or replacement by bulky amino acids (e.g., Arg or Trp) confers antagonist behavior to the ligand.

Short amino-terminal fragments, such as PTH(1–14), exhibit little or no binding or signaling activity; however, N-terminal fragment analogs with multiple, affinity-enhancing amino acid substitutions, such as the conformationally constraining amino acid, α -aminoisobutyric acid (Aib) at positions 1 and 3, were developed that exhibit considerably higher potency. For example, [Aib^{1,3}, Gln¹⁰, Har¹¹, Ala¹², Trp¹⁴]PTH(1–14)NH₂ is $\sim 100\,000$ -fold more potent than native PTH(1–14) for stimulating cAMP formation in cell-based assay systems ($\text{EC}_{50}\text{s} = 2 \text{ nM}$ vs. $200 \mu\text{M}$) and is, thus, as potent as PTH(1–34). A modified PTH(1–14) analog similar to the one above was used by researchers at Bristol Meyers Squibb Co. to screen a compound library for potential PPR mimetic ligands, and in 2007 they reported the discovery of a small molecule that competitively inhibits the binding of the modified PTH(1–14) peptide. The compound likely binds within the activation pocket of the receptor and, thus, suggests the potential for conversion, via medicinal chemistry, into a potent PPR mimetic agonist, of which none, so far, has been reported.

Structural Studies of PTH Ligands

Solution-phase nuclear magnetic resonance (NMR) studies performed on PTH(1–34) analogs have suggested a bihelical conformation in which a short N-terminal helix (approximately residues 3–10) is joined via a turn or flexible linker to a longer more stable C-terminal helix. Some NMR studies have indicated that the bihelical peptide can adopt a U-shaped structure, whereas an X-ray crystal structure of PTH(1–34) reported in 2001 revealed a single, slightly bent α -helix extending from Val-2 to His-32. The bioactive conformation of either PTH or PTHrP, as found in the receptor-bound state, is currently uncertain.

Structure–Activity Relationships in the PPR

Molecular Architecture of the PPR

The amino-terminal domain

The PPR contains an amino-terminal ECD of approximately 167 amino acids, excluding the 23-amino-acid signal sequence (Figure 1). This ECD is Asn-glycosylated at four sites and contains six cysteine residues that are highly conserved among the class 2 GPCRs and are predicted to form a disulfide-bond network that stabilizes a protein fold used in common by these receptors. Present in the PPR ECD, but not any other class 2 GPCR, is a 45-amino-acid segment, Ser61–Gly105, which is encoded by a separate exon, exon 2, which can be deleted without an apparent effect on function. The X-ray

crystallographic structure of the isolated, bacterially produced PPR ECD was reported in 2008. The structure, shown in Figure 2, was obtained as a complex with the PTH(15–34) fragment. In 2009, the same group reported the structure of the ECD in complex with the PTHrP(12–34) fragment; the structure was similar, but not identical to the PTH-bound structure. The overall protein fold of the ECD consists of two layers of beta-sheet and two bracing α -helical domains. The fold is stabilized by the predicted disulfide-bond network. The tertiary structures of the ECDs of several other class 2 GPCRs, including the corticotropin-releasing factor receptor and the pituitary-AC-activating peptide receptor, have also been determined by crystallography or NMR methods, and these structures indeed confirm the use of the same tertiary fold. In the structure of the PPR ECD, the exon 2 domain was not resolved due to disorder. In both the PTH(15–34)- and PTHrP(12–36)-bound complexes, the ligand binds as an α -helix in a central groove of the ECD (Figure 2). Relative to the PTH-bound helix, the PTHrP helix bends slightly at the midpoint, such that the C-terminal segments of the two ligands form distinct sets of contacts with the receptor.

The juxtamembrane domain

The juxtamembrane (J) region of the PPR is comprised of the heptahelical bundle and the interconnecting loops. The structure of this domain is likely to bear at least some similarity to the heptahelical core regions of the class 1 GPCRs, rhodopsin and the β 2-adrenergic receptor, the only GPCRs for which X-ray crystal structures are available. Computer models of the PPR have been generated based on the rhodopsin structure, but the nearly complete lack of sequence homology between the class 1 GPCRs and the PPR limits their reliability. One structural feature predicted for the PPR J domain, and likely present in most if not all GPCRs, is a disulfide bond between ECL-1 and ECL-2 (Figure 1). Within the transmembrane helices of the PPR, there are a number of residues that are well-conserved in the class 2 GPCR family and are, thus, likely to help define the topology and functionality of the J-domain region. These include the prolines in TMs 4, 5, and 6, expected to induce kinks in the helices, and polar residues, such as Arg233 in TM2 and Gln451 in TM7, that are likely to contribute to networks of inter- and intrahelical side-chain interactions. The three intracellular loops (ICLs) of the PPR, particularly ICL2 and ICL3, play important roles in coupling to G-protein-signaling pathways.

The Ligand Interaction Mechanism in the PPR

The two-site model

The ligand-binding mechanism for the PPR has been analyzed by complementary biochemical and pharmacological approaches, including photochemical cross-linking of ligand–receptor complexes. The overall results suggest a two-site model, by which the C-terminal-binding domain of PTH(1–34) first interacts with the PPR ECD to provide initial docking energy, and then the amino-terminal domain of the ligand interacts with the PPR J domain, which, in turn, induces the conformational changes involved in receptor activation and G-protein coupling (Figure 3). As for most GPCRs, the ternary complex comprised of ligand–receptor and G protein is thought to be a high affinity, but transient state. Recent studies, however, suggest that certain ligands for the PPR, such as PTH(1–34) but

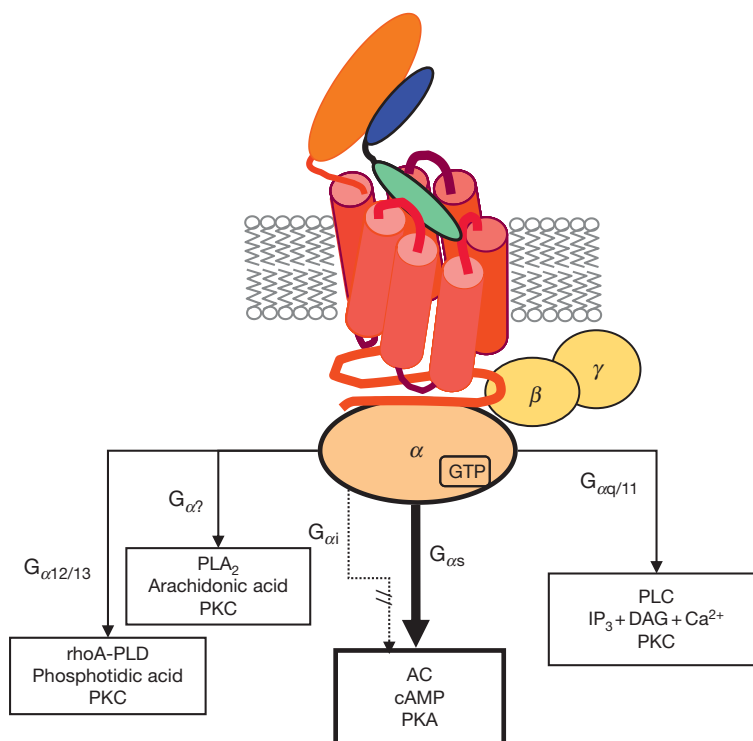


Figure 3 Ligand-binding and G-protein-coupling mechanisms for the PTH/PTHrP receptor. The ligand, PTH(1–34) or PTHrP(1–34), is shown to interact with the PPR via a bipartite mechanism in which the ligand's C-terminal portion (blue) interacts with the receptor's amino-terminal ECD (orange) to provide binding affinity, and the ligand's N-terminal portion (green) interacts with the receptor's ECLs and heptahelical portion (red) to induce receptor activation. The activated PPR can couple to multiple-signaling pathways via several different heterotrimeric G proteins. It couples most strongly, and in most cells, to $G_{\alpha s}$ to stimulate the AC/cAMP/PKA pathway. The PPR can also couple to the $G_{\alpha q}$ /PLC/ IP_3 / Ca^{2+} /PKC and $G_{\alpha 12/13}$ /RhoA/PLD/phosphatidic acid/PKC pathways, and some evidence suggests coupling to a pertussis toxin-sensitive G protein, presumably $G_{\alpha i}$, to inhibition of AC/cAMP, as well as to the activation of PLA_2 , although a cognate G protein for this latter pathway has not been identified.

not the C-terminally truncated, modified PTH(1–14) analogs mentioned above, nor PTHrP(1–36), can form a high-affinity state with the PPR, even when a G protein is not coupled to the receptor, and that this mode of stable binding can lead to prolonged signaling activities.

ECD domain interactions

The crystal structures of the PPR-ECD in complex with the PTH- or PTHrP-binding domains show that the ligand domain binds as an amphipathic α -helix and fits into a central groove of the protein fold. The hydrophobic face of the ligand helix, as formed largely by Trp-23, Leu-24, and Leu-28, makes extensive contacts with hydrophobic sites on the floor and walls of the groove. Relative to PTH, the PTHrP helix bends slightly at about position 27, such that the C-terminal portions of the two ligands diverge and make distinct sets of contacts. For both ligands, the highly conserved Arg20 side-chain fits into a pocket formed by a network of hydrophobic and polar groups, mostly near the receptor's N-terminus. Most, if not all, of the contact residues of the ligand have been shown to be important for binding by structure-activity relationship studies. The identified ECD contact site for ligand position 23, Trp in PTH and Phe in PTHrP, is also in agreement with photoaffinity cross-linking studies, which showed that a PTHrP(1–36) analogs modified at position 23 with photolabile benzoylphenylalanine (Bpa), cross-linked to the PPR segment

Thr33–Leu40, near the receptor's N-terminus (Figure 1). In the crystal structure, ligand residue 23 makes hydrophobic contacts to Leu41 of the receptor. The binding of the ligand's (15–34) domain to the ECD is likely to result in the N-terminal (1–14) domain of the ligand being positioned appropriately for binding to the receptor's membrane-embedded region.

J-domain interactions

Cross-linking studies have shown that positions 1 and 2 of bound PTH and PTHrP lie in proximity to Met425 at the extracellular end of TM6. Interestingly, the Bpa2-modified cross-linking ligand behaves as an antagonist/inverse agonist on the wild-type PPR, but the mutation of Met425→Leu causes it to behave as a partial agonist. These and related mutational data involving residues in TM5 suggest that the highly conserved Val-2 in the ligand plays a critical role in receptor activation, and does so by interacting with residues located at the extracellular ends of transmembrane domain helices five and six in the receptor.

Several other contact sites in the PPR J domain have been identified by cross-linking analyses, including Arg186 located at the border of the ECD and TM 1, which probably contacts ligand residue 13, as well as Lys240 at the extracellular of TM2, the probable contact site for ligand residue 19. The overall results suggest a possible model in which the N-terminal portion of the ligand, extending approximately from residue 1 to

19 and forming an α -helix, lies transversely along the extracellular surface of the receptor, and that key residues in the ligand helix, namely, Val-2, Ile-5, and Met-8, project into the membrane-embedded core of the receptor to induce receptor activation. Consistent with such a model, N-terminal PTH analogs that contain the enhancing modifications described above exhibit the same potency on a PPR mutant construct that lacks the ECD (PPR-delNt) as they do on the intact PPR. Although these studies suggest that the PPR ECD is not essential for receptor activation, the ECD is clearly required for achieving full potency and affinity with unmodified PTH(1–34), as this ligand is 100-fold weaker on PPR-delNt, as it is on the intact PPR.

PPR Signaling and Regulation

Conformational changes

As for all GPCRs, it is presumed that PPR activation involves movements in the TM helices and ICLs which result in increased accessibility of the receptor's cytoplasmic surface to G proteins; however, the exact conformational changes that occur are still largely undefined. A study using Zn(II)-metal-ion chelation strategies involving histidine-substituted PPR mutants suggests that TM3 and TM6 move away from each other during G_{α_s} activation. A similar TM3–TM6 movement has also been described for the β_2 -adrenergic receptor as well as for rhodopsin, suggesting similarity in the activation mechanisms used by the class 1 and class 2 GPCRs.

In 2003, a series of studies was initiated to investigate the receptor activation and binding mechanisms using optical kinetic approaches. These studies were based on Forster resonance energy transfer (FRET) methodology and utilized fluorescently modified receptors labeled with a variant of green fluorescent protein (GFP). To assess receptor activation, a cyan variant, CFP, was inserted in the receptor's third ICL, and a yellow variant, YFP, was added to the C-tail. In the basal state, these YFP and CFP groups lie in close enough proximity for a FRET response to occur, such that excitation illumination of the CFP ($\lambda \sim 430$ nm) results in increased fluorescent emission by the YFP ($\lambda \sim 530$ nm). Addition of PTH caused a decrease in the FRET signal, indicating a separation or reorientation of the YFP and CFP groups. The FRET responses were monitored in living cells and in real time, and revealed that PPR activation occurs on the time scale of ~ 1 s. Studies on the mechanisms of PTH(1–34) binding, performed using a fluorescently tagged ligand and a PPR-containing GFP in the ECD exon 2 region, revealed a two-step-binding process, for which the slower phase paralleled the time course of receptor activation. The slower phase was similarly paralleled by the FRET response observed between a YFP-tagged PTHR and CFP-tagged G protein. These studies in general suggest that a key, rate-limiting step in PPR-mediated signal transduction is the conformational change induced in the receptor as the N-terminal portion of the ligand engages the J-domain region of the receptor.

A few specific residues in the IC loops of the PPR involved in G-protein coupling have been identified by conventional mutational methods. These include Lys388 in ICL3, at which mutations impair both AC/cAMP (G_{α_s}) and PLC/IP₃ (G_{α_q}) signaling and Lys319 in ICL2, at which mutations selectively impair PLC/IP₃ (G_{α_q}) signaling. A binding site for the

G-protein β/γ subunit has also been mapped to the N-terminal proximal portion of the receptor's C-terminal tail.

The carboxyl-terminal tail

Molecular and cellular studies have shown that the PPR carboxyl-terminal tail plays a major role in signal desensitization/termination and receptor internalization. Upon agonist activation, several serines within the N-terminal-proximal portion of the tail (Figure 1) are rapidly phosphorylated, most likely by G-protein-specific receptor kinase-2. This phosphorylation promotes the binding of β -arrestin-1 and/or -2, and the internalization of the ligand–receptor complex into endosomal vesicles, most likely via a clathrin-coated pit-mediated process. At least some of the internalized PPR can recycle back to the cell membrane, presumably in a dephosphorylated, resensitized form. The PPR CT also mediates interaction with cytoskeletal scaffolding proteins, including the Na⁺/H⁺ exchanger regulatory factor (NHERF) family of proteins. The C-terminal four amino acids (Glu–Thr–Val–Met) of the receptor mediate this interaction by binding to one of the PDZ (postsynaptic density protein (PSD95), *Drosophila disc large tumor suppressor* (Dlg1), zonula occludens-1 protein (zo-1)) domains of the NHERF proteins. In some cells, the binding of NHERF proteins to the PPR has been shown to modulate the signaling pathways utilized by the receptor, resulting in increased PLC signaling and diminished AC/cAMP signaling. It also can reduce the lateral diffusion of the receptor on the cell surface, and the extent of receptor internalization following PTH activation, effects which involve linkage to the actin cytoskeletal network.

The PPR in Human Disease

Jansen's Metaphyseal Chondrodysplasia

This rare autosomal-dominant disease is characterized by hypercalcemia (reflecting the homeostatic functions of PTH in bone and kidney) and short-limbed dwarfism (reflecting the morphogenetic roles of PTHrP in the growth plates). In 1995, a mutation in the PPR gene, which changed His223 at the intracellular terminus of TM2, to Arg, was identified in an affected individual, and the mutation was shown to confer high basal cAMP-signaling activity to the receptor when analyzed in transfected COS-7 cells. Subsequently, three other activating mutations in the PPR were identified in patients with Jansen's disease, and each of these alters a residue at or near the intracellular termini of a TM helix: Thr410→Pro and Thr410→Arg in TM6 and Ile458→Arg in TM7. In transfected COS-7 cells, the N-terminally truncated antagonist, [Leu¹¹,DTrp¹²]PTHrP(5–36), behaves as an inverse agonist and, thus, suppresses constitutive cAMP signaling, on the mutant PPRs bearing the His223→Arg or Thr410→Pro mutation, whereas the [Bpa²]PTHrP(1–36) behaves as a selective inverse agonist on the His223→Arg mutant. These findings suggest a potential mode of treatment for patients with this disease.

Blomstrand's Chondrodysplasia

This rare embryonic lethal condition is characterized by a severely overcalcified fetal skeleton and short limbs. Three different PPR-inactivating mutations have been identified in this recessive disorder: a mRNA-splicing mutation that results

in an 11-amino acid deletion in TM5; a frameshift mutation in ECL2, and a missense mutation (Pro132→Leu) in the ECD. The phenotype seen in Blomstrand's chondrodysplasia mirrors that seen in mice having homozygous deletion of the PPR gene.

Conclusions

The PPR mediates the biological actions of two key proteins – PTH (calcium and phosphate homeostasis) and PTHrP (bone and tissue development). The so-called PTH-2 receptor reported by Usdin and co-workers in 1995 is 51% identity to the PPR at the amino acid level, and while it exhibits some overlap in ligand-binding pharmacology with the PPR, it likely interacts specifically with the hypothalamic peptide, TIP39. The PTH2R/TIP39 system is thought to play a neuro-endocrine role, possibly in nociception, as well as a role in spermatogenesis, rather than in PTH- or PTHrP-related biology. Genetic defects in the PTH/PTHrP/PPR system result in a number of diseases of calcium-ion homeostasis and tissue development. Further unraveling the molecular mechanisms by which the PPR interacts with PTH and PTHrP is an important goal of fundamental biological research, and could likely lead to new therapies for diseases of bone and mineral metabolism.

See also: **Signaling:** Angiotensin Receptors; Bradykinin Receptors; Calcitonin Gene-Related Peptide and Adrenomedullin Receptors; G Protein Signaling Regulators; G_q Family; Gs Family of Heterotrimeric G Proteins; Melanocortin System; Opioid Receptors; Vasopressin/Oxytocin Receptor Family.

Further Reading

- Bergwitz C and Juppner H (2010) Regulation of phosphate homeostasis by PTH, vitamin D, and FGF23. *Annual Review of Medicine* 61: 91–104.
- Gardella TJ (2009) Mimetic Ligands for the PTHR1: Approaches developments, and considerations. *JBMS BoneKEy* 6(2): 71–85.
- Gardella TJ, Juppner H, Brown EM, Kronenberg HM, and Potts JT Jr. (2010) Parathyroid hormone and parathyroid hormone-related peptide in the regulation of calcium homeostasis and bone regulation. In: DeGroot L (ed.) *Endocrinology*, 6th edn., vol. 1, part VI, ch. 56, pp. 1040–1073. Philadelphia, PA: WB Saunders.
- Jilka RL (2007) Molecular and cellular mechanisms of the anabolic effect of intermittent PTH. *Bone* 40: 1434–1446.
- Kramer I, Keller H, Leupin O, and Kneissel M (2010) Does osteocytic SOST suppression mediate PTH bone anabolism? *Trends in Endocrinology and Metabolism* 21(4): 237–244.
- Vilardaga JP, Bunemann M, Feinstein TN, et al. (2009) GPCR and G proteins: Drug efficacy and activation in live cells. *Molecular Endocrinology* 23: 590–599.

Pentose Phosphate (Hexose Mono Phosphate) Pathway

B R Landau, Case Western Reserve University, Cleveland, OH, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 211–215, © 2004, Elsevier Inc.

Glossary

NADPH Abbreviated form of nicotinamide adenine dinucleotide phosphate; the electron donor in reductive biosynthesis.

Pentose phosphate pathway Pathway by which NADPH is generated for reductive biosynthesis and ribose-5-P for nucleotide and nucleic acid formation.

Transaldolase Enzyme in the pentose phosphate pathway transferring three carbon units via the formation of a Schiff base.

Transketolase Enzyme in the pentose phosphate pathway having thiamine as cofactor and transferring two carbon units.

History

In 1931, Warburg demonstrated glucose-6-P's oxidation to 6-phosphogluconate with the formation of NADPH. In the following decade, Warburg, Lipmann, Dickens, and others found that the gluconate was decarboxylated to CO₂ and to a pentose phosphate that could then be reconverted to hexose-6-P. Details of the pathways were elucidated in the 1950s by Cohen, Racker, Horecker, and their associates. **Figure 1** shows the pentose phosphate pathway.

Reactions

Oxidative Segment

Glucose-6-P oxidation is catalyzed by glucose-6-P dehydrogenase, yielding 6-phosphogluconolactone and nicotinamide adenine dinucleotide phosphate (NADPH) (**Figure 2**). The spontaneous hydrolysis of the lactone forms 6-phosphogluconate, catalyzed by a specific lactonase. The gluconate is oxidatively decarboxylated to yield another NADPH and ribulose-5-P. The overall reaction of this segment is then $\text{glucose-6-P} \rightarrow \text{CO}_2 + 2\text{NADPH} + \text{ribulose-5-P}$.

Nonoxidative Segment

The ribulose-5-P is isomerized to ribose-5-P and also epimerized to xylulose-5-P (**Figure 3**). Transfer of the top two carbons as a unit from the xylulose-5-P to ribose-5-P, catalyzed by transketolase, yields seven-carbon-containing sedoheptulose-7-P and glyceraldehyde-3-P. Transfer of the top three carbon unit as a unit from the sedoheptulose-7-P to the glyceraldehyde-3-P, catalyzed by transaldolase, yields fructose-6-P and four-carbon-containing erythrose-4-P. Transfer of the top two carbons as a unit from another xylulose-5-P to the erythrose-4-P, again catalyzed by transketolase, yields fructose-6-P and glyceraldehyde-3-P. Thus, the overall reaction of this segment is $3 \text{ pentose-5-P} \leftrightarrow 2 \text{ fructose-6-P} + \text{glyceraldehyde-3-P}$.

Transketolase requires thiamine pyrophosphate as a cofactor. The two carbon unit is transferred via its addition to the thiamine ring of the vitamin. The mechanism is similar to that in the oxidative decarboxylation of pyruvate, catalyzed by the

thiamine pyrophosphate containing E1 subunit of the pyruvate dehydrogenase complex. In the transfer of the three carbon unit, catalyzed by transaldolase, a Schiff base is formed between the ketone group of the sedoheptulose-7-P and a lysine residue at the active site of the enzyme. The mechanism is similar to that in fructose-1,6-bisphosphatase aldolase catalysis.

Overall Balance and Carbon Fate

On balance, 3 glucose-6-P (18 carbons) are decarboxylated to yield three CO₂ and 3 pentose-5-P (15 carbons) (**Figure 4**). The rearrangements, catalyzed by transketolase and transaldolase produce 2 fructose-6-P (12 carbons) and glyceraldehyde-3-P (3 carbons). The CO₂ contains carbon 1 of the glucose-6-P (**Figure 5**). The origin of the top 3 carbons of the 2 fructose-6-P are carbons 2 and 3 of the glucose-6-P.

Mode of Operation

Shunt or Cycle

Since glucose-6-P is converted to fructose-6-P and glyceraldehyde-3-P, the pentose phosphate pathway has been viewed as an alternate route for the conversion of glucose-6-P to intermediates of the glycolytic pathway. Therefore, the pathway has also been called the hexose monophosphate shunt. Since the fructose-6-P can be isomerized to glucose-6-P that can re-enter the pentose phosphate pathway, the pathway has also been viewed as a cycle and called the pentose cycle.

Regulation

The dehydrogenation of glucose-6-P, the first step, is considered to be the rate-limiting step in the oxidative segment. Control is through the activity of glucose-6-P dehydrogenase regulated by the concentration of nicotinamide adenine dinucleotide phosphate (NADP). NADPH competes with the NADP for binding to the enzyme, thus tightly coupling NADPH production to utilization. When NADPH is utilized (e.g., in fatty acid synthesis), the NADPH is oxidized, increasing NADP and stimulating flux through the oxidative segment.

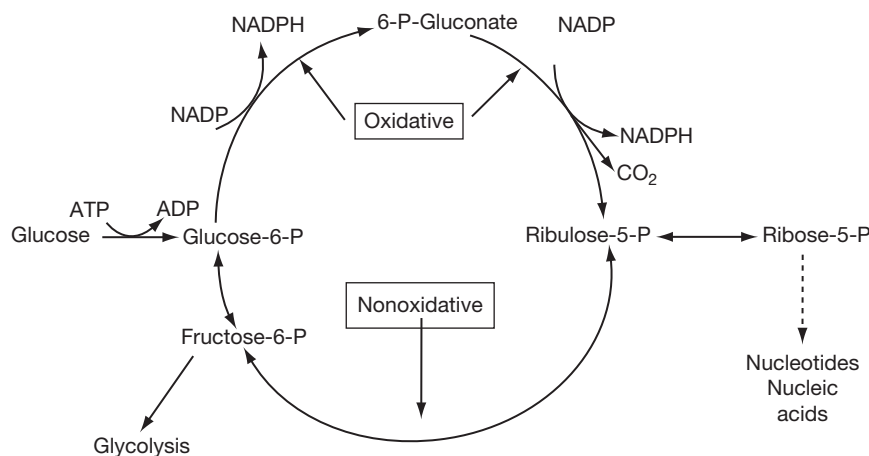


Figure 1 Depiction of the pentose phosphate pathway as a cycle composed of oxidative and nonoxidative segments.

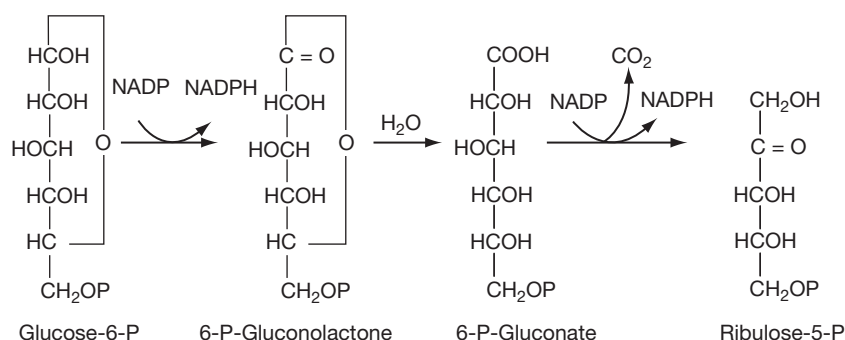


Figure 2 Reactions of the oxidative segment of the pentose phosphate pathway.

Flux through the nonoxidative segment is determined by the concentration of pentose-5-P. The activity of the pathway is coordinated with that of the pathways utilizing its products. Recent studies identify xylulose-5-P in the pathway as mediating the effect of carbohydrate feeding on the glycolytic pathway and regulating the enzymes required for fatty acid and triglyceride synthesis.

Other Functions

The pentose phosphate pathway is the route by which pentose from food is metabolized. Erythrose-4-P in some organisms is used as an intermediate in the synthesis of the aromatic ring of the amino acids. D-Xylulose formed in the glucuronic acid pathway is further metabolized via the pentose phosphate pathway.

Quantitation

Methods

Estimates of the activity of the pathway rest on measurements of the activities of its enzymes, particularly of glucose-6-P dehydrogenase, and of concentrations of intermediates in the pathway, as well as yields of labeled products after

administering labeled substrate and intermediates of the pathway. Specifically, carbon-labeled glucoses have been used to quantitate the amount of glucose utilized by the pathway relative to other pathways utilizing glucose. The quantitations depend on (1) the fact that carbon 1 of every molecule of glucose-6-P entering the pathway is oxidized to CO₂, while via glycolysis and the tricarboxylic acid cycle, carbons 1 and 6 are oxidized to CO₂ to essentially the same extent, or (2) the randomization of carbons 2 and 3 of glucose-6-P occurring in the nonoxidative segment (Figure 5).

The amounts of glucose oxidized and that utilized by the pathway are different. Thus, while the three molecules of glucose-6-P that enter the pathway (Figure 4) are oxidized, and 2 glucose-6-P equivalents are reformed, only a net of 1 glucose-6-P is utilized. It is converted to CO₂ and glyceraldehyde-3-P. If, as can occur in the liver, 2 glyceraldehyde-3-P that are formed are converted to 1 glucose-6-P, the balance becomes 6 glucose-6-P → 12 NADPH + 6 CO₂ + 5 glucose-6-P. Hence, 6 glucose-6-P are oxidized, but only 1 glucose-6-P is utilized (completely oxidized to CO₂).

Estimates

The pathway is present in all cells. The highest activities are found in tissues where lipogenesis is prominent, for example,

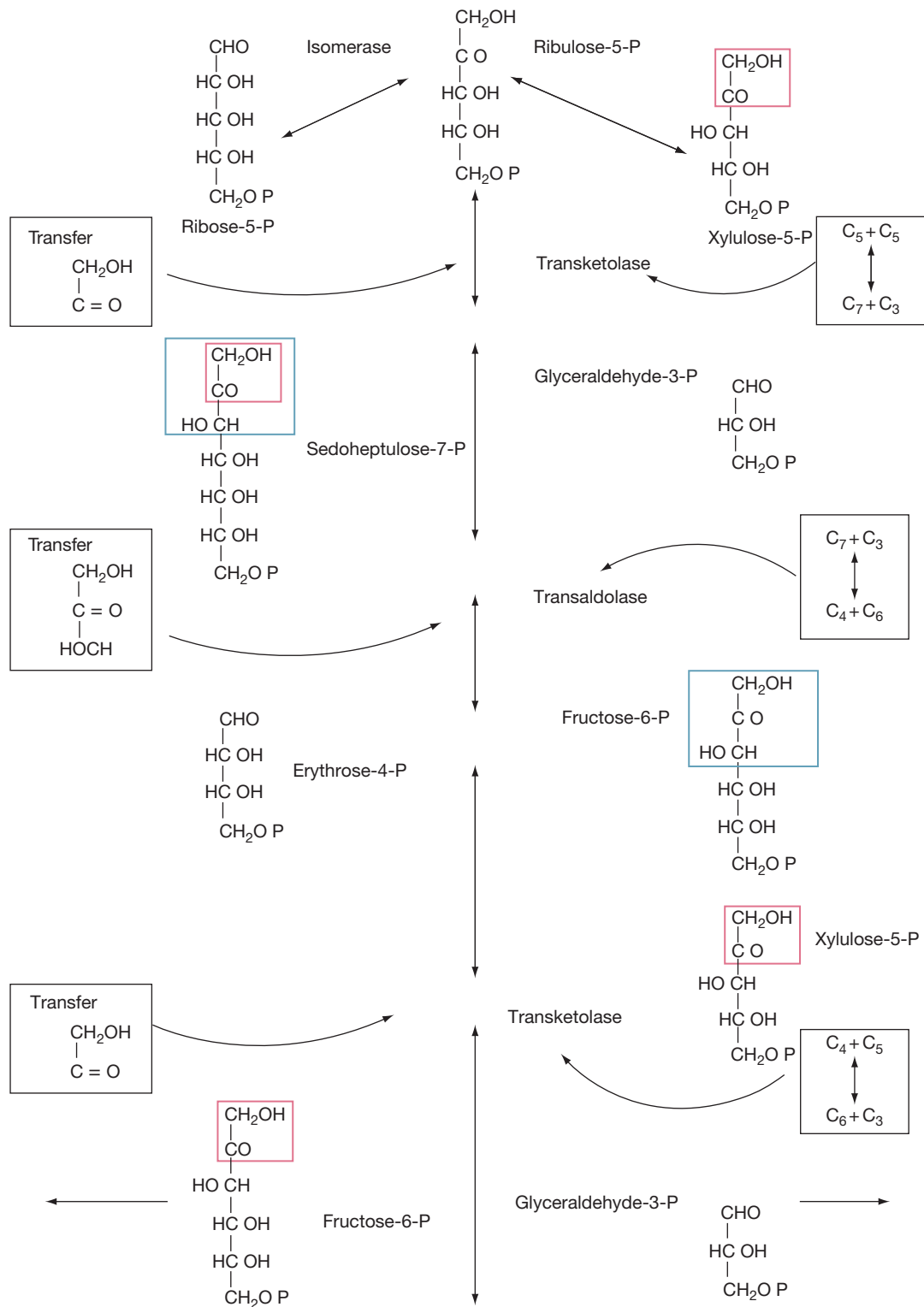


Figure 3 Reactions of the nonoxidative segment of the pentose phosphate pathway.

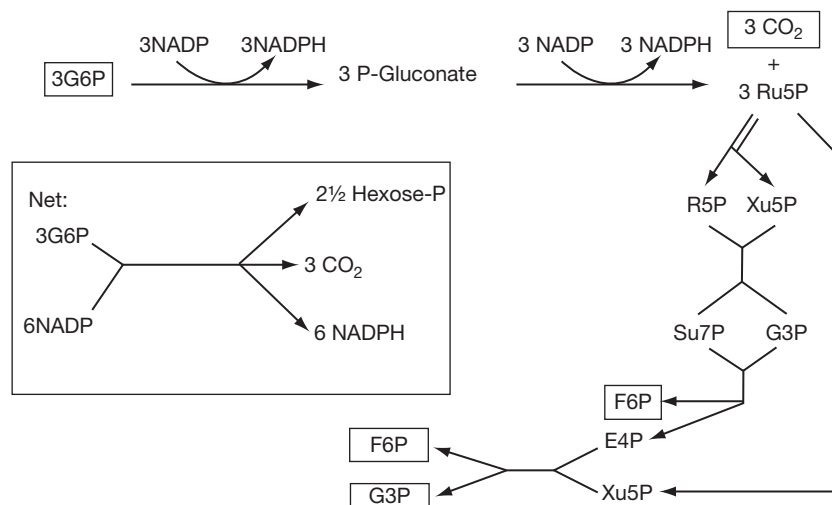


Figure 4 Chemical balance in the reactions of the pentose phosphate pathway.

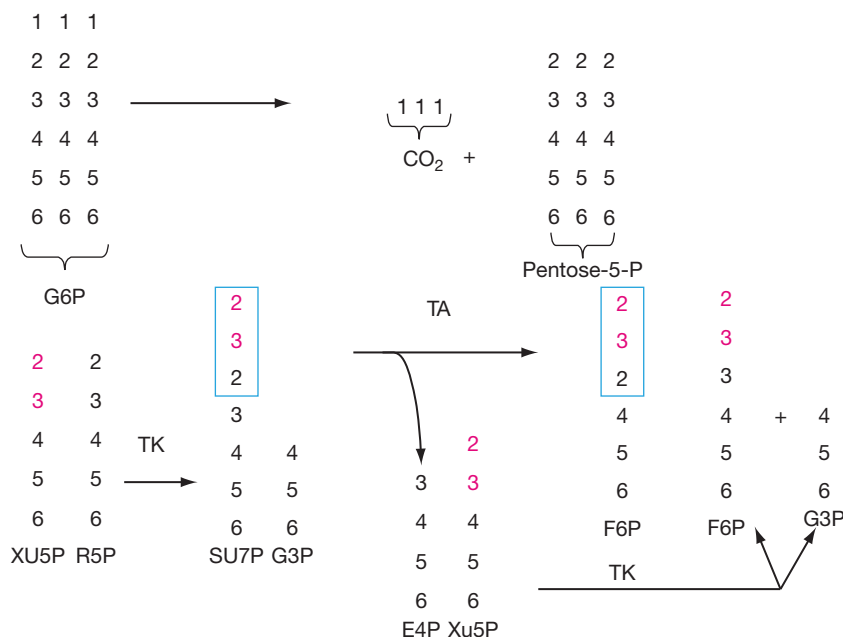


Figure 5 Fate of the carbons of the glucose-6-P (G6P) in its metabolism via the pentose phosphate pathway.

liver, mammary gland, particularly during lactation, and adipose tissue. The pathway is also active in the reproductive glands, adrenal gland, and red blood cells. In those tissues 10–20% of glucose utilization may be via the pathway. The pathway has relatively high activity in growing tissue, for example, in fetal and newborn tissues and tumors.

Clinical Importance

Maintaining the integrity of the red blood cells depends on having an adequate supply of reduced glutathione to protect the cells from toxic peroxides, as well as maintain hemoglobin

in a reduced state. Several hundred million people, mainly in Africa, Asia, and the Mediterranean, have a genetic deficiency of glucose-6-P dehydrogenase. Drugs that increase the formation of peroxide, infections, and other causes of oxidative stress can induce acute hemolysis and hence anemia in those individuals.

The Wernicke-Korsakoff syndrome is a neuropsychiatric disorder occurring in individuals with a dietary deficiency of thiamine. Findings are eyeparalysis, abnormal gait, confabulation, and loss of memory. A genetic defect in transketolase has been suggested as the reason that the syndrome is only observed in a small portion of those with thiamine deficiency. A genetic abnormality in transketolase has recently been found in a man with liver cirrhosis.

See also: [Lipids Carbohydrates Membranes and Membrane Proteins: Sugar Nucleotide Transporters](#).

Further Reading

- Cabezas H, Raposo RR, and Melendez-Havia E (1999) Activity and metabolic roles of the pentose phosphate cycle in several rat tissues. *Molecular and Cellular Biochemistry* 201: 57–63.
- Horecker BL (2002) The pentose phosphate pathway. *Journal of Biological Chemistry* 277: 47965–47971.
- Magnusson I, Chandramouli V, Schumann WC, et al. (1988) Pentose pathway in human liver. *Proceedings of the National Academy of Sciences of the United States of America* 85: 4682–4685.
- Massillon D, Chen W, Barzilai N, et al. (1998) Carbon flux via the pentose pathway regulates the hepatic expression of the glucose-6-phosphatase and phosphoenolpyruvate carboxykinase genes in conscious rats. *Journal of Biological Chemistry* 273: 228–234.
- Mehta A, Mason PJ, and Vulliamy TJ (2000) Glucose-6-phosphate dehydrogenase deficiency. *Bailliere's Clinical Haematology* 13: 21–38.
- Schenk G, Duggleby RG, and Nixon PF (1998) Properties and functions of the thiamin diphosphate dependent enzyme transketolase. *International Journal of Biochemistry and Cell Biology* 30: 1297–1318.
- Veech RL (2003) A humble hexose monophosphate pathway metabolite regulates short- and long-term control of lipogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 100: 5878–5880.
- Verhoeven NM, Huck JH, Roos B, et al. (2001) Transaldolase deficiency: Liver cirrhosis associated with a new inborn error in the pentose phosphate pathway. *American Journal of Human Genetics* 68: 1086–1092.
- Wood T (1985) *The Pentose Phosphate Pathway*. Orlando, FL: Academic Press.

Peptide Amidation

M J Niciu, R E Mains, and B A Eipper, University of Connecticut Health Center, Farmington, CT, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Large dense core vesicles/secretory

granules Membrane-enclosed 300–500 nm diameter peptide-containing storage organelles formed from the *trans*-Golgi network in neuroendocrine cells.

P-type ATPase A family of pumps whose members use the energy of adenosine triphosphate (ATP) to

transport cations uphill across membranes, undergoing a cycle of phosphorylation and dephosphorylation.

Type I integral membrane protein A protein that crosses a phospholipid bilayer once, with its N-terminus projecting into the extracellular environment/lumen of the secretory pathway and a cytosolic C terminus.

Generation of a C-terminal α -amide group on a peptide occurs by hydroxylation and cleavage of a C-terminal glycine residue and is a prevalent posttranslational modification essential for the production of many of the peptides that serve as hormones and neurotransmitters. Approximately half of all bioactive peptides are α -amidated, including gonadotropin-releasing hormone (GnRH), corticotropin-releasing hormone (CRH), thyrotropin-releasing hormone (TRH), oxytocin, vasopressin, calcitonin, gastrin, cholecystokinin (CCK), neuropeptide Y (NPY), substance P, and pituitary adenylate cyclase-activating polypeptide (PACAP). The α -amide group prevents ionization of the C terminus, which may increase receptor binding, hydrophobicity, and/or half-life. In vertebrates, C-terminal α -amidation is accomplished via the activity of peptidylglycine α -amidating monooxygenase (PAM), a bifunctional enzyme localized to large dense core vesicles in neuroendocrine tissues.

Biochemistry and Enzymology of PAM

PAM is composed of two protease-resistant catalytic cores targeted to the luminal compartment of secretory granules, peptidylglycine α -hydroxylating monooxygenase (PHM; EC 1.14.17.3) and peptidyl- α -hydroxyglycine α -amidating lyase (PAL; EC 4.3.2.5). PHM catalyzes the stereospecific α -hydroxylation of all C-terminally glycine-extended peptide substrates. PAL then catalyzes cleavage of the N–C bond of the α -hydroxyglycine, generating the α -amidated peptide product and glyoxylate (Figure 1).

PAM Structure

Bifunctional PAM is a type I integral membrane protein whose structure is highly conserved in vertebrate phylogeny. From N to C terminus, it contains a signal sequence, proregion, PHM catalytic core, noncatalytic spacer region (exon A), PAL catalytic core, transmembrane domain, and cytosolic domain. The hydrophobic signal sequence is cleaved cotranslationally in the endoplasmic reticulum; the prosequence is removed in a post-*trans*-Golgi network compartment. Tissue-specific endoproteolytic cleavage at pairs of basic amino acids can remove the proregion, separate the two catalytic domains, and release the catalytic domains from the transmembrane

domain. Poorly conserved linker regions follow both catalytic domains and are susceptible to cleavage by exogenous proteases. Additionally, alternative splicing can generate soluble, monofunctional PHM or soluble, bifunctional PAM. Soluble PAM proteins are secreted with their amidated product peptides, while integral membrane PAM proteins undergo endocytosis, guided by routing determinants and phosphorylation at multiple sites in the cytosolic domain (Figure 2).

PAM Substrates

Virtually any peptide with a C-terminal glycine residue can serve as a PAM substrate. PAM generates α -amides of all 20 amino acids. In addition, PAM catalyzes the amidation of nonpeptidergic substrates such as fatty acyl glycines, glutathione, the aspirin metabolite salicylurate, bile acid glycine conjugates, and leukotriene C₄. In addition to stereospecific α -hydroxylation of glycine-extended precursors, purified PHM can also catalyze N-dealkylation, O-dealkylation, and sulfoxidation of nonpeptidergic substrates.

Peptidylglycine α -Hydroxylating Monooxygenase

Essential cofactors

PHM requires the presence of molecular oxygen, a single electron donor (generally ascorbic acid or vitamin C), and copper.

Ascorbic acid (vitamin C)

Ascorbate is transported across the plasma membrane via a Na⁺-dependent transporter enriched in neuroendocrine tissue, SVCT2. How cytosolic ascorbate reaches the lumen of the secretory pathway is currently unclear. Nevertheless, concentrations of ascorbate are 5- to 10-fold higher in the lumen of the secretory pathway than in the cytosol. The millimolar concentrations of ascorbate in the luminal compartment ensure that luminal copper is reduced, as required for the enzymatic cleavage of molecular oxygen by PHM. In this oxidation–reduction reaction, PHM converts 2 mol of ascorbate into 2 mol of semidehydroascorbate, which disproportionate to form dehydroascorbate and ascorbate. Other single-electron reductants can substitute for ascorbate to provide reducing equivalents for Cu²⁺; consistent with this, no ascorbate-specific binding site has been identified on PHM.

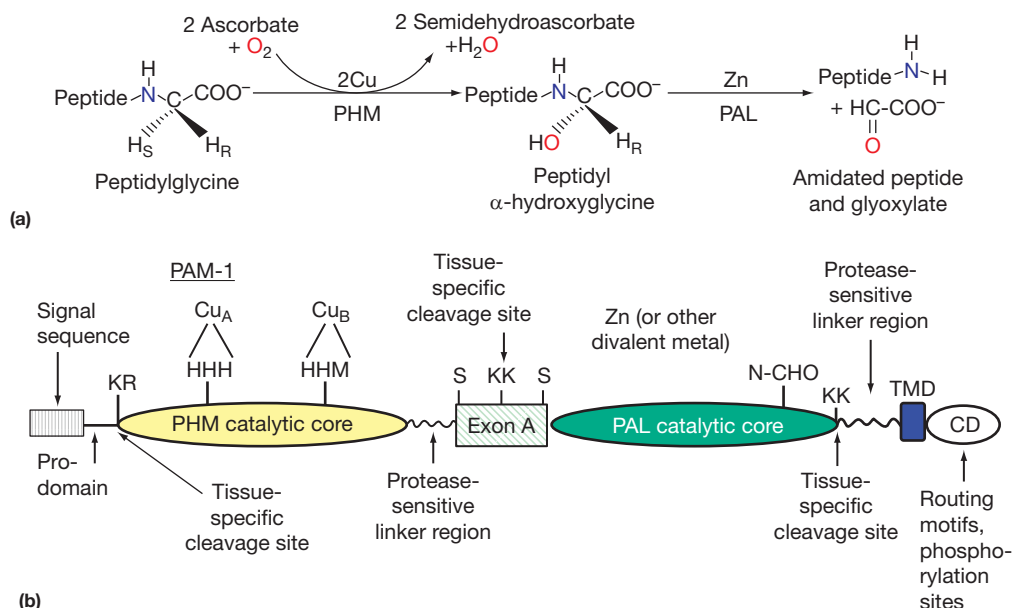


Figure 1 PAM reaction and structure. (a) Bifunctional PAM-catalyzed conversion of glycine-extended precursor into hydroxyglycine intermediate and amidated product plus glyoxylate. (b) Domains of integral membrane PAM-1 with salient features: KR and KK lysine–arginine and lysine–lysine dibasic cleavage sites; H, copper-binding histidine; M, copper-binding methionine; N-CHO, N-linked glycosylation site; TMD, transmembrane domain; CD, cytosolic domain, a multiply phosphorylated unstructured domain.

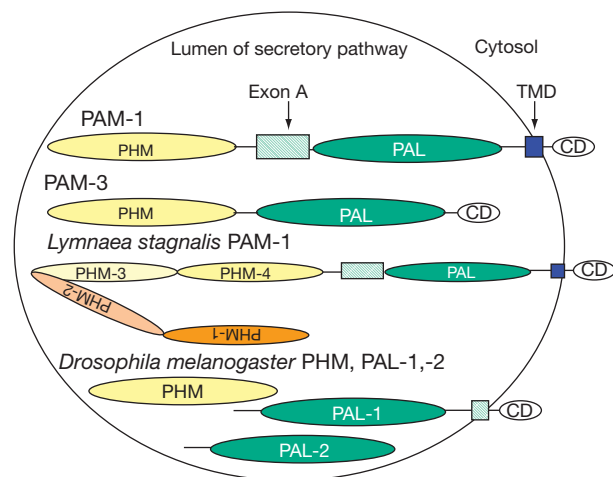


Figure 2 PAM in secretory granules. Soluble and membrane PAM isoforms from rat, pond snail, and fruit fly in mature secretory granule. Alternative mRNA splicing at hot spots generates soluble PAM-3, which lacks exon A and the TMD; it is secreted with the amidated product. *Lymnaea stagnalis* membrane PAM contains four functional PHM domains and an exon A-like linker sequence. *Drosophila* PHM and PAL-1/2 are encoded on separate genes. PAL-1 is a soluble protein and may be secreted with processed peptide. PAL-2 contains a transmembrane domain and may be recycled via endocytosis to post-Golgi complex compartments. CD, cytosolic domain, a multiply phosphorylated unstructured domain.

Copper

Copper is required in trace amounts by a limited group of cuproenzymes. However, in excess, copper can support the Fenton reaction, generating reactive oxygen species, which attack cellular constituents with diffusion-limited kinetics. To avoid this problem, copper transporters and copper

chaperones deliver this critical metal to these cuproenzymes while preventing toxic intracellular actions. The transport of Cu^{1+} across the plasma membrane requires Ctr1 in mammals. Atox1, a cytosolic copper chaperone, delivers Cu^{1+} to ATP7A and ATP7B, homologous P-type ATPases that pump Cu^{1+} into the lumen of the secretory pathway, where it can be loaded onto luminal cuproproteins or excreted from the cell.

Structure and reaction mechanism PHM binds two moles of copper, each of which plays a distinct role in the α -hydroxylation reaction. Composed almost entirely of β strands, the catalytic core of PHM consists of two structurally homologous domains (Figure 3). Mutagenesis and phylogenetic comparisons identify a limited number of conserved residues. The copper-binding sites are conserved: three histidine residues in the N-terminal half of PHM bind Cu_A , also known as Cu_H , and two histidines plus one methionine residue in the C-terminal half bind Cu_B , also known as Cu_M . Although reducing equivalents donated by both Cu_A and Cu_B are essential to the reaction, the two Cu ions are separated by a 10 Å solvent-filled cleft. The glycine-extended peptide substrate binds in a conserved, hydrophobic substrate-binding pocket located near Cu_B , placing the α -carbon of the glycine adjacent to molecular oxygen bound to Cu_B , with the C terminus of the glycine residue salt bridged to a conserved arginine residue. Several hydrophobic residues near the active site are conserved, along with proline and glycine residues that define the loop regions. The molecular oxygen consumed in the reaction binds only to the enzyme/peptidylglycine complex, and it is not yet clear how electron transfer is accomplished. The reaction is tightly coupled, with little generation of reactive oxygen species, and the two copper ions do not move much closer to each other during catalysis.

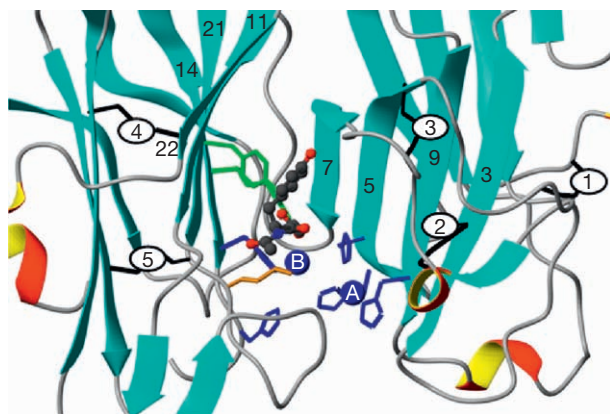


Figure 3 PHM active site. The structure determined by crystallization of oxidized PHMcc with its Ac-diiodo-Tyr-Gly substrate is shown to illustrate the structural role played by the disulfide bonds. Histidine ligands to Cu_A and Cu_B are blue; the methionine ligand to Cu_B is gold; substrate-binding residues are green; the substrate is shown in ball and stick format. The disulfide bonds in PHM are numbered (except for bond 1, they are totally conserved): 1, Cys⁴⁷–Cys¹⁸⁶; 2, Cys⁸¹–Cys¹²⁶, links $\beta 3$ to the $\beta 5$ – $\beta 6$ loop; 3, Cys¹¹⁴–Cys¹³¹, links $\beta 5$ to the $\beta 6$ – $\beta 7$ loop; 4, Cys²²⁷–Cys³³⁴, links $\beta 13$ to $\beta 22$; 5, Cys²⁹³–Cys³¹⁵, links $\beta 19$ to $\beta 21$. From Prigge ST, Mains RE, Eipper BA, and Amzel LM (2000) New insights into copper monooxygenases and peptide amidation: Structure, mechanism and function. *Cell and Molecular Life Sciences* 57: 1236–1259.

Pharmacological inhibition Disulfiram, used clinically in the treatment of chronic alcoholism, inhibits PHM activity *in vivo* via its ability to chelate copper. When assayed *in vitro*, after the addition of exogenous copper, the catalytic core of rat PHM displays paradoxically enhanced activity after *in vivo* disulfiram administration. A mechanism-based inhibitor of PHM, trans-styrylacetic acid (phenylbutenoic acid), has also been identified.

Peptidyl- α -Hydroxyglycine α -Amidating Lyase

Limited trypsin digestion and analysis of peptidyl- α -hydroxyglycine α -amidating lyase (PAL) truncation mutants revealed a stable catalytic core with no subdomains. The single N-linked glycosylation site near the COOH terminus of PAL is utilized, although glycosylation is without effect on the catalytic activity of PAL. The catalytic core of PAL is a six-bladed beta-propeller. An active site Tyr sits close to a Zn ion, which is coordinated to 3 His residues. Calcium plays a structural role, and the N- and C-termini of PAL are latched together, constraining the location of PHM. No reliable mechanism-based small-molecule inhibitors of PAL have been characterized. The reaction catalyzed by PAL is most similar to the reaction catalyzed by ureidoglycolate lyase.

Evolutionary Relationships

The process of peptide amidation is highly conserved. In invertebrates such as *Drosophila melanogaster* and *Caenorhabditis elegans*, PHM and PAL are encoded by separate genes; a single PHM gene may be accompanied by more than one

PAL gene. PHM and PAL homologs are not found in the genome of *Saccharomyces cerevisiae*. Intriguingly, *Lymnaea stagnalis*, a molluskan pond snail, expresses a huge membrane PAM with four independent, catalytically active PHM domains followed by a single PAL domain. *Schistosoma mansoni*, a nematode and potent human pathogen, expresses a heavily glycosylated, soluble PHM with atypical catalytic features as well as PAL. The copper-binding ligands, substrate-binding site, disulfide bonds, and selected glycine and proline residues of PHM are conserved across species.

The cofactor requirements of PHM demonstrate homology to another copper, ascorbate and dioxygen-requiring enzyme, dopamine- β -monooxygenase (DBM; EC 1.14.17.1). DBM converts dopamine to norepinephrine in selected neurons and adrenal chromaffin cells. The catalytic core of PHM is 32% identical to a 291-amino-acid stretch in DBM. The Cu_A and Cu_B -binding motifs and disulfide-bonding patterns of PHM are conserved in DBM. *Cnidarians*, such as the sea anemone and *Hydra*, use amidated peptides, not catecholamines such as norepinephrine for neural transmission, suggesting that PHM is the more ancestral enzyme. cDNA library screening of human senescent fibroblasts revealed the presence of a third sequence with significant homology to PHM, monooxygenase X (MOX; X13255). No substrates or hydroxylated products have yet been identified for this putative monooxygenase.

PAM Expression

PAM expression is most prevalent in heart atria, the pituitary gland, and peptide-producing neurons of the central and enteric nervous systems. Nevertheless, PAM expression is widespread, as transcripts can be detected in airway and olfactory epithelia, smooth muscle cells, endothelium, brain ependymal cells, astrocytes, and fibroblasts.

During development, PAM is first detected immunohistochemically at embryonic day 9 (E9) in the putative cardiogenic region and at E13 in peptide-producing neurons. PAM is also highly expressed in limb mesoderm and mesenchyme flanking the nasal, maxillary, palatal, and dental epithelia during embryonic tissue fusion and remodeling. Most *Drosophila* PHM knockout flies die as young larvae, during the molting process, with malformation of the mouth hooks. A role for PAM in these areas during development is not yet clear.

PAM Molecular Genetics

The human PAM gene is located on chromosome 5q14-21 and the mouse *Pam* gene is on chromosome 1. The rat PAM gene contains at least 28 exons encompassing >160 kb. At least two putative promoters upstream of the translational start site and proregion in exon 2 have been identified. PHM is encoded by exons 3–14 (excluding exon 13). The alternately spliced noncatalytic region between PHM and PAL is encoded by exon 16, which is bordered by at least 40 kb of intronic sequence. PAL is encoded by exons 17–24. Most of the intron/exon junctions in rat, which preferentially produce bifunctional PAM messenger RNA (mRNA) transcripts, and

D. melanogaster, which contains separate PHM and PAL genes, are conserved. Interestingly, rat-breeding stocks vary in their alternative splicing of PAM: Harlan Sprague-Dawley rats eliminate the exon A linker in their mRNA transcripts in pituitary, while Charles River Sprague-Dawley rats retain this domain.

Dysregulated Peptide Amidation

Due to the plethora of physiological functions regulated by amidated peptides, deficits in peptide amidation have very detrimental consequences for the organism.

Overexpression of Amidated Peptides

Many amidated peptides, for example, PACAP, NPY, and gastrin-releasing peptide (GRP), serve as autocrine or paracrine growth factors. Numerous tumors of neuroendocrine origin secrete amidated peptides that serve this function, and PAM expression is concomitantly upregulated. Amidated peptides secreted by neuroendocrine tumors disrupt normal physiological function. For example, small cell lung carcinomas often secrete GRP, while amidated isoforms of parathyroid hormone related protein (PTHrP) sustain bone growth and produce the hypercalcemia of malignancy. Gastrin-secreting tumors (gastrinomas) are frequently encountered in Zollinger–Ellison syndrome, which presents clinically as peptic ulceration; vasoactive intestinal polypeptide-secreting tumors (VIPomas) of pancreatic or duodenal origin, present with voluminous watery diarrhea in Verner–Morrison syndrome. Hence, pharmacological inhibition of PAM activity may be a rational therapeutic strategy in the treatment of such amidated peptide-secreting tumors.

Impaired Peptide Amidation

Based on assays of amidated product peptides, peptide amidation is impaired in the context of dietary deficiencies of ascorbate or copper, genetic deficiencies of copper-translocating proteins, or targeted gene disruption of the PAM gene.

Impaired copper delivery

A reduction in copper delivery to the secretory pathway results in a diminished ability to produce amidated peptides. This can be the result of nutritional deficiencies of copper or ascorbate or genetic malfunctioning of copper transporters. In mammals, ATP7A receives Cu^+ from the cytosolic chaperone Atox1 and transports it into the lumen of the secretory pathway, where the catalytic cores of cuproenzymes, such as PHM, DBM, and tyrosinase, reside. Mutations in ATP7A result in the X-linked copper-deficiency disorder, Menkes disease, which presents clinically with profound neurodegeneration. The mottled brindled (Mo^{Br}) mouse, which contains a six base pair in-frame deletion in ATP7A, serves as the mouse model of Menkes disease. Although PAM protein expression is unchanged, the concentrations of amidated neuropeptides, for

example, joining peptide (JP), CCK, and PACAP, are reduced in the pituitary gland and central nervous system of afflicted males relative to age-matched wild-type mice.

Targeted gene disruption

Insertion of a P-element containing transposon within the *D. melanogaster* PHM transcriptional unit eliminated PHM expression and resulted in early larval lethality with relatively few morphological defects. PAM null mice do not survive past embryonic day 14.5, with the most noticeable defect being intractable edema. No PHM or PAL enzymatic activity is detectable in homogenates of null embryos, and heterozygotes display half the wild-type level of PHM and PAL activity in age-matched wild-type mice. Thus, PAM is the only enzyme responsible for C-terminal peptide α -amidation *in vivo*. Mice heterozygous for PAM display increased anxiety-like behavior and are unable to maintain body temperature when kept in a cold environment.

See also: **Bioenergetics:** P-Type Pumps: Copper Pump; **Metabolism Vitamins and Hormones:** Vitamin C; **Signaling:** Calcitonin Gene-Related Peptide and Adrenomedullin Receptors; Vasopressin/Oxytocin Receptor Family.

Further Reading

- Bousquet-Moore D, Prohaska JR, Nillni EA, et al. (2010) Interactions of peptide amidation and copper: Novel biomarkers and mechanisms of neural dysfunction. *Neurobiology of Disease* 37: 130–140.
- Francisco WA, Blackburn NJ, and Klinman JP (2003) Oxygen and hydrogen isotope effects in an active site tyrosine to phenylalanine mutant of peptidylglycine α -hydroxylating monooxygenase: Mechanistic implications. *Biochemistry* 42: 1813–1819.
- Iwai N, Martinez A, Miller M-J, et al. (1999) Autocrine growth factor loops dependent on peptidyl α -amidating enzyme as targets for novel tumor cell growth inhibitors. *Lung Cancer* 23: 209–222.
- Katopodis AG and May SW (1990) Novel substrates and inhibitors of peptidylglycine α -amidating monooxygenase. *Biochemistry* 29: 4541–4548.
- Kolhekar AS, Bell J, Shiozaki EN, et al. (2002) Essential features of the catalytic core of peptidyl- α -hydroxyglycine α -amidating lyase. *Biochemistry* 41: 12384–12394.
- Liang WJ, Johnson D, and Jarvis SM (2001) Vitamin C transport systems of mammalian cells. *Molecular Membrane Biology* 18: 87–95.
- Miller LA, Baumgart LE, Chew GH, et al. (2003) Glutathione, S-substituted glutathiones, and leukotriene C_4 as substrates for peptidylglycine α -amidating monooxygenase. *Archives of Biochemistry and Biophysics* 412: 3–12.
- Prigge ST, Mains RE, Eipper BA, and Amzel LM (2000) New insights into copper monooxygenases and peptide amidation: Structure, mechanism and function. *Cell and Molecular Life Sciences* 57: 1236–1259.
- Puig S and Thiele DJ (2002) Molecular mechanisms of copper uptake and distribution. *Current Opinion in Chemical Biology* 6: 171–180.
- Rhames FC, Murthy NN, Karlin KD, and Blackburn NJ (2001) Isocyanide binding to the copper(I) centers of the catalytic core of peptidylglycine monooxygenase (PHMcc). *Journal of Biological Inorganic Chemistry* 6: 567–577.
- Stevenson TC, Ciccotosto GC, Ma X-M, et al. (2003) Menkes protein contributes to the function of peptidylglycine α -amidating monooxygenase. *Endocrinology* 144: 188–200.

Periplasmic Electron-Transport Systems in Bacteria

D J Richardson, University of East Anglia, Norwich, UK

G Sawers, John Innes Centre, Norwich, UK

R J M Van Spanning, Free University of Amsterdam, Amsterdam, The Netherlands

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 231–238, © 2004, Elsevier Inc.

Glossary

Electrogenic Property of a transmembrane electron-transport enzyme that generates a proton-motive force (PMF).

Heme A tetrapyrrole with a coordinated iron that can mediate electron transfer when bound to proteins. In *c*-type cytochromes the heme is covalently bound to the protein.

Iron–sulfur cluster A protein cofactor that comprises of inorganic iron and sulfur, which together form a cluster that is often involved in electron transfer.

Periplasm A subcellular compartment that lies between the inner (cytoplasmic) and outer membrane of Gram-negative bacteria.

Proton-motive force A transmembrane gradient of protons ($\Delta\rho$) that has an electrical component ($\Delta\psi$) and a chemical component ($\Delta\rho H$). Its magnitude is normally given in units of millivolts and can be calculated from $\Delta\rho = \Delta\psi - 60 \Delta\rho H$.

Respiratory Electron-Transport Systems

In mammalian mitochondrial respiration, electrons flow from low-potential cytoplasmic electron donors (e.g., NADH or succinate) to the high-potential electron acceptor oxygen. The site of oxygen reduction is the *aa*₃-type cytochrome *c* oxidase that is located within the cytoplasmic membrane (Figure 1). The key postulate of P. Mitchell's chemiosmotic theory is that many energy-consuming reactions catalyzed by integral membrane proteins, such as the adenosine triphosphate synthase, are driven by the energy contained within a transmembrane-proton electrochemical gradient (proton-motive force (PMF)), which is generated by the electron-transport system. Proton translocation is achieved via membrane-associated electron-transfer proteins. Mitchell described a redox loop mechanism whereby two electrons are transferred from the positive (P, or bacterial periplasmic) side of the membrane to the negative (N, or bacterial cytoplasmic) side of the membrane. There they combine with two protons to reduce a quinone to quinol. The quinol then diffuses back across the membrane where it is reoxidized at the P face releasing protons (Figure 1(b)). This movement of negative charge from the P to the N side of the membrane, and counter-movement of protons, produces a PMF. Structure–function studies of two key integral membrane enzymes of respiration (the structurally defined cytochrome *bc*₁ complex and the *aa*₃-type cytochrome *c* oxidase) illustrate yet another three different mechanisms that couple electron transfer to PMF generation. The cytochrome *bc*₁ complex utilizes a proton-motive Q-cycle mechanism that results in a q^+/e^- ratio of 1 (Figure 1(a)). The *aa*₃-type cytochrome *c* oxidase employs two different mechanisms to build up the PMF. Reduction of oxygen at the catalytic site involves proton transfer from the cytoplasm and electron transfer from a periplasmic donor. As a consequence, these charge movements from opposite sides of the membrane give rise to a net charge separation of 1 per electron transferred, $1q^+/e^-$. In addition, it utilizes a proton-pumping mechanism that results in one positive charge being moved across the membrane for each electron ($1q^+/e^-$). Therefore, the

stoichiometry of charge movements in the oxidases adds up to $2q^+/e^-$ transferred from donor to oxygen. It is notable that neither the proton pump mechanism nor the Q-cycle mechanism of energy conservation was originally envisaged by Mitchell.

Bacterial Periplasmic Respiratory Electron-Transport Systems That Are Dependent on the Cytochrome *bc*₁ Complex

The mammalian mitochondrial-type of electron-transport system is present in many aerobic and facultative anaerobic bacteria, for example, the well-studied soil-denitrifying bacterium *Paracoccus denitrificans*. However, many bacteria additionally have alternative electron-transport systems that utilize periplasmic electron donors and acceptors. In some cases, the periplasmic oxido-reductases will integrate easily into a mitochondrial-type of electron-transport chain. For example, in *P. denitrificans* the periplasmic oxido-reductases for the reduction of nitrite or nitrous oxide, which are important reactions of denitrification in the biogeochemical nitrogen cycle, can accept electrons via the cytochrome *bc*₁ complex (Figure 2). Due to the Q-cycle mechanism of the latter enzyme, the overall electron-transport pathway from quinol to nitrite and nitrous oxide is electrogenic, although the terminal nitrite and nitrous oxide reductases themselves are not. As a consequence, the electron transport from quinol to nitrite is less coupled ($q^+/e^- = 1$) (Figure 2) than electron transfer from quinol to oxygen ($q^+/e^- = 3$) (Figure 1(a)).

Periplasmic Electron-Transport Systems That Do Not Depend on the Cytochrome *bc*₁ Complex

Many bacteria (e.g., *Escherichia coli*) do not have genes encoding a cytochrome *bc*₁ complex. Using periplasmic electron donors and acceptors independently of the cytochrome *bc*₁

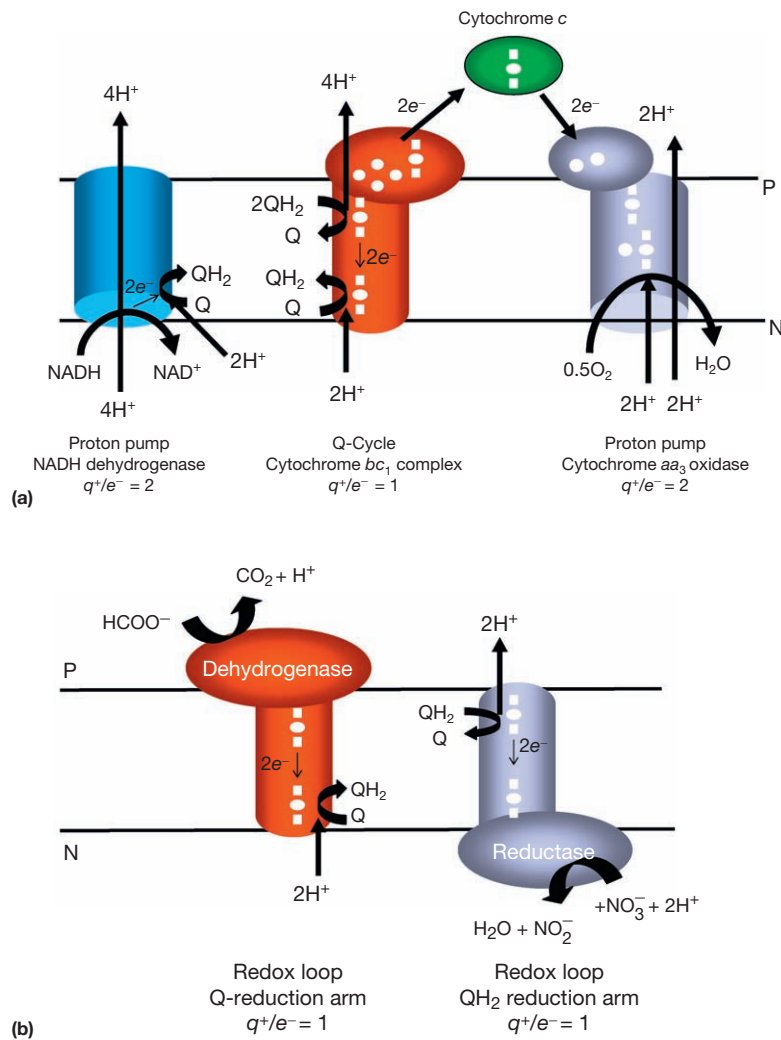


Figure 1 Mechanisms of energy conservation in respiration. (a) A basic mammalian mitochondrial electron-transport chain illustrating proton pumps and Q-cycle mechanisms of energy conservation. (b) The organization of redox loops using periplasmic electron donors found in some cytochrome bc_1 complex-independent bacterial respiratory systems.

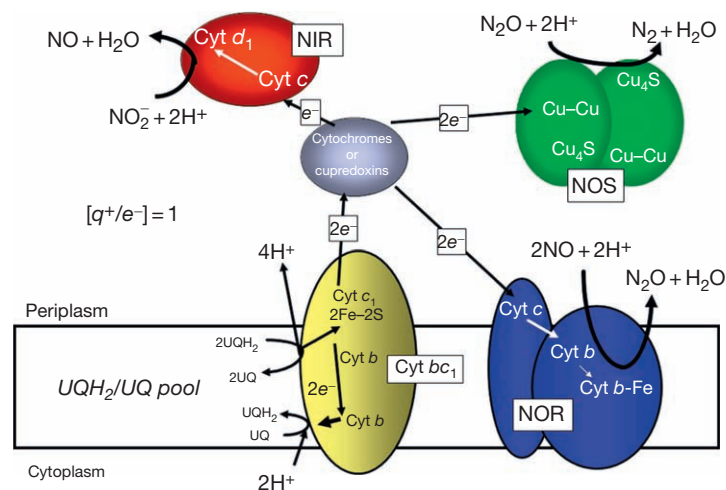


Figure 2 Cytochrome bc_1 complex-dependent periplasmic and membrane-bound oxido-reductases involved in denitrification in *Paracoccus* species. NIR, cytochrome cd_1 nitrite reductase dimer; NOR, nitric oxide reductase; NOS, nitrous oxide reductase.

complex can raise interesting questions about the bioenergetics of these systems, particularly with respect to the means by which the electron-transfer process is coupled to the generation of a transmembrane electrochemical gradient. The molecular nature of some of these coupling mechanisms has been illuminated by recent structure–function studies on key anaerobic enzymes.

Some of these transmembrane enzymes possess two heme *b* groups at opposite sides of the membrane, which allows transmembrane electron transfer from a periplasmic to a cytoplasmic side. During oxidation of their substrate, two protons are released at the periplasmic side of the membrane while two electrons move inward via the hemes *b* to the cytoplasmic side of the membrane. There they are taken up by the acceptor along with two protons from the cytoplasm. This classical Mitchellian redox loop mechanism results in a net charge separation of one per electron transferred through the enzyme from donor to acceptor molecule ($1q^+/e^-$). Both dehydrogenases, which oxidize their substrate at the periplasmic side and reduce quinone at the cytoplasmic side, and terminal oxido-reductases, which oxidize quinol at the periplasmic site and reduce the acceptor at the cytoplasmic side, may have a redox-loop-type architecture. Examples of such electrogenic dehydrogenase are the HydABC-type hydrogenase and FdhN of *E. coli*, homologs of which are present in many bacteria and archaea. An example of such an electrogenic terminal oxido-reductase is membrane-bound nitrate reductase (NarA). Periplasmic or membrane-bound enzymes that are not designed to

contribute to the generation of a PMF will dissipate the redox free energy as heat. These enzymes receive their electrons and protons required for reduction of their allocated electron acceptor at the same side of the membrane, either the periplasm (trimethylamine N-oxide reductase (TOR), dimethylsulfoxide reductase (DMS), periplasmic nitrate reductase (Nap), the cytochrome *c* nitrite reductase (Nrf)) or the cytoplasm (succinate dehydrogenase and fumarate reductase). Electron transfer in pathways that are made up of a redox-loop-type dehydrogenase or oxido-reductase along with a nonelectrogenic oxido-reductase or dehydrogenase, respectively, will yield a net proton movement of 1 from inside to outside during electron transfer from donor to terminal electron acceptor ($q^+/e^- = 1$) (Figure 3). The most efficient free-energy transducing pathway of electron transfer from donor to acceptor requires the participation of a reductase, which both have a redox-loop-type architecture (Figure 3). The FdhN to NarA respiratory chain of *E. coli* and the Hyd to NarA (electron transfer from formate or hydrogen to nitrate, $q/e^- = 2$) are the classical systems used to illustrate such an efficient proton-motive redox loop.

Formate Dehydrogenase N

Formate is an important and widespread electron donor available to bacteria in anaerobic environments. During anaerobic growth with nitrate, *E. coli* can express a membrane-bound

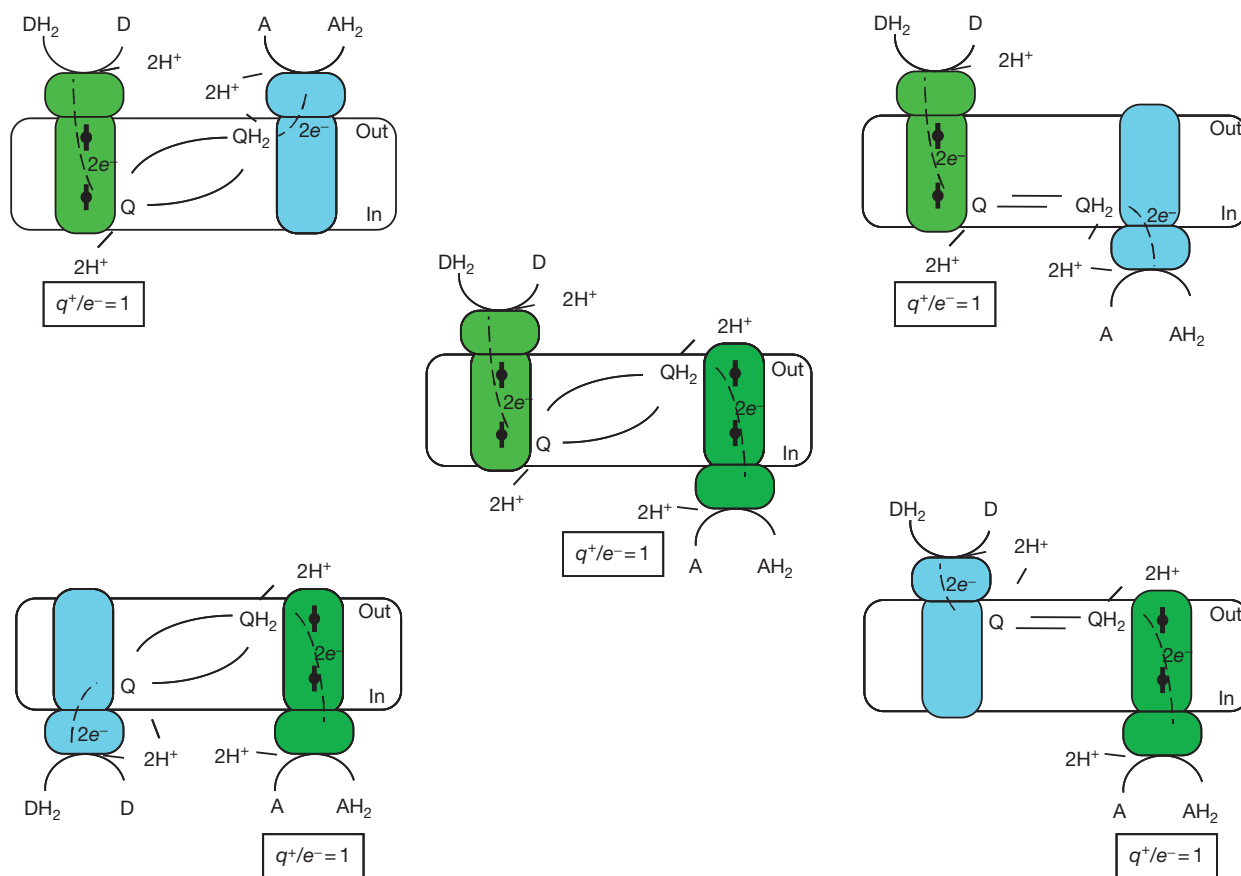


Figure 3 Examples of $q/e^- = 1$ and $q/e^- = 2$ redox loop mechanisms.

formate dehydrogenase (FdhN), which oxidizes formate in the periplasmic compartment and for which a high-resolution structure has been determined. The enzyme is a homo-trimer, with each monomeric unit being composed of an integral membrane di-heme cytochrome subunit (FdhI), a membrane-anchored periplasmic ferredoxin subunit (FdhH), binding four [4Fe-4S] clusters, and a periplasmic catalytic subunit (FdhG), which binds a [4Fe-4S] cluster and catalyzes formate oxidation at a Se-Mo-*bis*-guanine dinucleotide (Se-Mo-MGD) cofactor. Electrons extracted from formate at the periplasmic active site pass down a 90-Å redox ladder to a menaquinone reductase site at the cytoplasmic face of the membrane. The redox ladder consists of eight redox centers, Se-Mo-*bis*-MGD, five iron-sulfur clusters, and two hemes. Each of these is within 12 Å of its nearest neighbor, and this serves to ensure rapid electron transfer (Figure 4). In considering the whole electron-transfer ladder from the Se-Mo-*bis*-MGD to the menaquinone, a notable feature is that the intermediary electron carriers usually function as one-electron-transfer centers. However, formate oxidation and menaquinone reduction are two-electron reactions. Thus, the Se-Mo-*bis*-MGD plays a key role in gating electron transfer from formate into the redox ladder and the Q-reductase site is the key to gating the reduction of menaquinone.

Much of the FdhN redox ladder is periplasmic, being located in the extra-membranous periplasmic FdhGH subunits (Figure 4), but the two membrane-intrinsic hemes, bound within the four-helix bundle of FdhI, are critical to charge separation across the membrane. The midpoint redox potential at pH 7 of the formate/CO₂ couple is *c.* -420 mV and that of the MQH₂/MQ couple is *c.* -80 mV. This large, *c.* 340 mV, potential drop allows efficient electron transfer against a negative-inside membrane potential of *c.* 200 mV. The movement of one electron from the periplasmic (P) face to the

cytoplasmic (C) face of the membrane and uptake of one proton at the C face gives a coupling stoichiometry of $q^+/e^- = 1$ (Figure 4).

Hydrogenase

Hydrogen is another important periplasmic electron donor for bacteria in anoxic environments. The Hya and Hyb enzymes of *E. coli* catalyze anaerobic hydrogen oxidation. There are no structures for these enzymes, but they are thought to be organized in a manner similar to FdhN. They are composed of three different subunits: an integral membrane di-heme subunit in which the hemes are stacked across the membrane; a periplasmic iron-sulfur protein (anchored by a single trans-membrane helix); and a periplasmic catalytic subunit that binds a nickel-iron cofactor at the active side. The bioenergetics of periplasmic hydrogen oxidation are thus predicted to be very similar to that described for formate, with electrons extracted for hydrogen in the periplasm passing through a multi-cofactor redox ladder and combining with protons to reduce quinone at the cytoplasmic face, thus giving $q^+/e^- = 1$.

Nitrate Reductase A

In *E. coli*, nitrate reductase A can serve as a second proton-motive arm in electron transfer from the periplasmic active site of FdhN or Hya (Figure 4). NarA is a three-subunit enzyme composed of NarGHI. NarG contains the *bis*-MGD molybdopterin cofactor at its catalytic site and a [4Fe-4S] cluster. NarH contains four additional iron-sulfur centers (one [3Fe-4S] and three [4Fe-4S]). NarGH are located in the cytoplasm and associate with NarI, an integral membrane protein of five transmembrane helices with the N-terminus and two

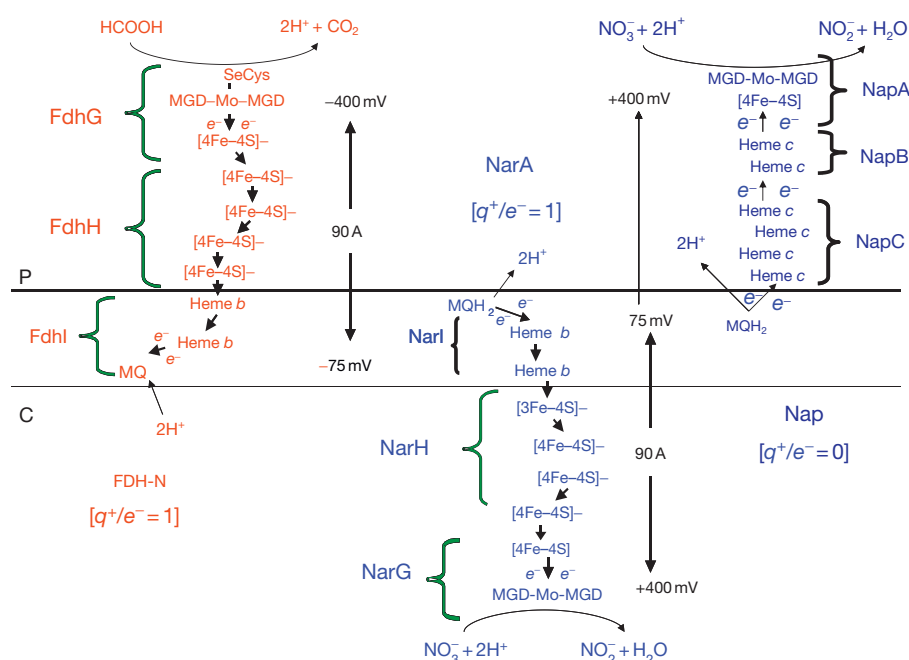


Figure 4 The organization of redox ladders involved in periplasmic formate-dependent nitrate reduction in *E. coli*. FdhN, formate dehydrogenase N; NarA, membrane-bound nitrate reductase A; Nap, periplasmic nitrate reductase.

inter-helix loops facing the periplasm. This subunit carries two hemes *b* stacked across the membrane. Oxidation of quinol occurs at the periplasmic side of NarI where the protons are released and the two electrons are moved from the outside low-potential heme *b* to the inside high-potential heme *b*. This charge separation makes the enzyme electrogenic, in that it contributes to the generation of a proton electrochemical gradient across the membrane (two charge separations during transfer of two electrons to nitrate; $q^+/e^- = 1$). Nar is organized in a manner very similar to FdhN, especially with respect to the equivalent FeS-containing NarH and FdhH subunits and Mo-*bis*-MGD NarG and FDHG catalytic subunits). However, the topology of the two systems is opposite (Figure 4), but the redox ladder arrangement of redox centers apparent in FdhN is also present in NarA. The midpoint redox potentials of individual cofactors in this ladder have been determined and suggest that both NarG and FdhN are part of a growing number of electron-transfer complexes in bacteria that contain a mixture of endergonic and exergonic electron-transfer steps. As in FdhN, the potential difference of *c.* 480 mV between the MQH₂/MQ and nitrate/nitrite redox couples is sufficient to drive electron transfer across the membrane against the negative-inside PMF (Figure 4).

Nonproton Motive Cytochrome *bc*₁ Complex-Independent Periplasmic Electron Transfer

In the FdhN–NarA electron-transport system, both the quinone-reducing and quinol-oxidizing arms of the electron-transport chain are proton motive and, when operating together, they yield $q/e^- = 2$. However, in principle, only one arm needs to be proton motive in order for a net PMF $q^+/e^- = 1$ to be generated. Examples of this can be seen in a periplasmic formate–nitrate electron-transport system in which nitrate is reduced by the Nap rather than NarA (Figure 3). Such a system operates under nitrate-limited growth conditions in *E. coli*.

The Nap

Nap is widespread in many species of denitrifying and non-denitrifying proteobacteria, including both *P. denitrificans* and *E. coli*. Nap is normally found as part of a three-subunit

electron-transfer system composed of NapABC. NapA is a periplasmic catalytic subunit with a [4Fe–4S] cluster and a *bis*-MGD cofactor. The crystal structure of *Desulfovibrio desulfuricans* NapA reveals that the overall organization is similar to that of the catalytic subunits of the Mo-*bis*-MGD formate dehydrogenases. NapB is a periplasmic di-heme cytochrome *c*, with both hemes having *bis*-histidinyl ligation. NapC has a single N-terminal transmembrane helix that anchors a globular periplasmic domain with four *c*-type hemes to the periplasmic surface of the cytoplasmic membrane (Figure 4). All four hemes have relatively low midpoint redox potentials (in the region of 0 to –250 mV) and have *bis*-histidinyl axial ligation.

NapC may be made up of two subdomains, each containing a di-heme pair, and it belongs to a large family of bacterial tetra-heme and penta-heme cytochromes. These cytochromes have been proposed to participate in electron transfer between the quinol/quinone pool and periplasmic redox enzymes such as the TOR, DMS, fumarate reductase, nitrite reductases, and hydroxylamine oxido-reductase. For an electron-transport pathway from quinol to nitrate through Nap, the enzyme receives electrons from membrane-embedded quinols, which are donated to the tetra-heme periplasmic domain of NapC. From there they flow via the hemes in NapB and the [4Fe–4S] cluster in NapA to the *bis*-MGD-containing catalytic site of NapA where the two-electron reduction of nitrate to nitrite is catalyzed. As discussed above for NarA and FdhN, this electron-transfer system also contains a mixture of endergonic and exergonic electron-transfer steps but it is expected that none of the eight redox centers involved will be more than 14 Å apart, thus ensuring rapid electron transfer. However, electron transfer from quinol to nitrate via Nap is nonelectrogenic ($q^+/e^- = 0$) since both the electrons and the protons required for the reduction of nitrate to nitrite are taken up at the same side of the periplasmic face of the membrane.

Multiheme *c*-Type Cytochromes and Periplasmic Electron Transfer

Multiheme *c*-Type Cytochromes

In *E. coli* three other illustrative examples of nonelectrogenic, quinol-oxidizing periplasmic electron-transport systems that

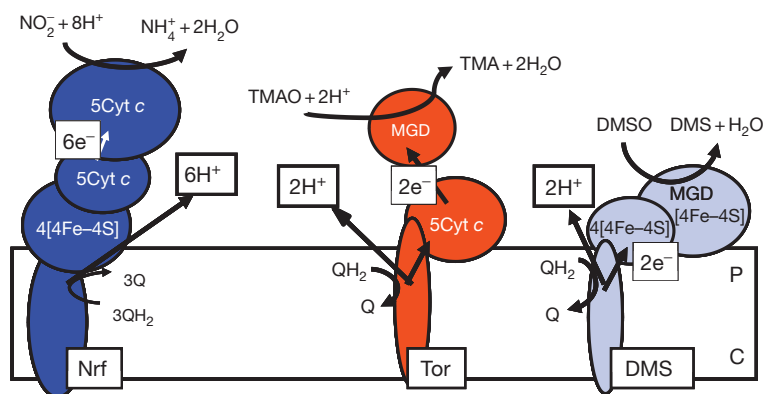


Figure 5 The organization of some nonelectrogenic periplasmic electron-transport systems in *E. coli*. Nrf, cytochrome *c* nitrite reductase; TOR, trimethylamine-N-oxide reductase; DMS, dimethylsulfoxide reductase.

include extended redox ladders are illustrated in Figure 5. The Nrf is a four-subunit electron-transport system that involves 14 redox centers (10 *c*-type hemes and four [4Fe–4S] clusters). The TOR is a two-subunit system comprising six redox centers (Mo–bis-MGD and five *c*-type hemes). Finally, the DMS is a three-subunit system containing an integral membrane subunit that is not involved in charge separation, and two periplasmic subunits that bind five redox centers (Mo–bis-MGD and four [4Fe–4S] clusters). Consideration of the Nap, Nrf, and TOR systems has revealed the participation of di-, tetra-, and penta-heme cytochromes in electron transfer. In mammalian mitochondria, the P-facing cytochrome *c* involved in electron transfer is a small mono-heme protein. However, multiheme cytochromes are a common feature of bacterial periplasmic electron transport. Other examples not mentioned so far include: (1) the octa-heme hydroxylamine oxido-reductase of nitrifying bacteria (e.g., *Nitrosomonas europaea*) and tetra-heme cytochrome *c*554, which, together with a tetra-heme NapC homolog CycB, mediate nonelectrogenic electron transfer from hydroxylamine to quinone; (2) the tetra-heme fumarate reductase of *Shewanella* species that, together with the NapC homolog CymA, reduces fumarate to succinate using electrons drawn from the quinol pool. Structural analysis of these cytochromes and other multiheme cytochromes such as the penta-heme NrfA reveals that the hemes are closely packed together to allow rapid electron transfer and are frequently arranged in pairs in which the porphyrin planes are nearly parallel or nearly perpendicular.

Iron (III) Respiration and Intermembrane Periplasmic Electron Transport

Bacterial iron respiration is the process in which the Fe(III) acts as the terminal electron acceptor in an energy-conserving bacterial respiratory electron-transport chain. It is an anaerobic process that provides a bacterium with a means of respiration when oxygen is absent and is distinct from the process of iron assimilation, in which iron is taken up into the cell by energy-consuming systems and incorporated into cell biomass. Although not a property of *E. coli*, the capacity for Fe(III) respiration is phylogenetically widespread and the ecological impact in microoxic and anoxic soils and sediments is considerable. A particular problem for Gram-negative bacteria is that at neutral pH the speciation of Fe(III) is complex as the cation is present either as insoluble polynuclear oxo/hydroxo-bridged complexes or as soluble organic chelates. Thus, Fe(III) is frequently presented to bacteria in a range of insoluble forms. The best-characterized group of Fe(III)-respiring bacteria are species of the genus *Shewanella*, which are widespread facultative anaerobes of the gamma proteobacterium group. *Shewanella* species can synthesize a number of periplasmic terminal respiratory reductases, including nitrate, nitrite, fumarate, and TORs. During growth under iron-respiring conditions, *de novo* synthesis of a number of multiheme *c*-type cytochromes that may play a role in the Fe(III) respiration process occurs and a number of genetic loci that encode multiheme cytochromes have been implicated as being important for Fe(III) respiration. Critically, these include an inner membrane tetra-heme cytochrome CymA (a NapC homolog), a small periplasmic 10-kDa tetra-heme cytochrome (STC); periplasmic deca-heme

cytochromes (e.g., MtrA); and outer-membrane cytochromes with deca-heme domains on the cell surface. The different cellular locations of these multiheme cytochromes is critical to the transfer of electrons from the inner membrane across the periplasm to the outside of the outer-membrane where direct reduction of the insoluble extracellular substrate can occur. The multiheme nature of the cytochromes may allow the proteins to form a multi-redox-centered electron wire that will be even longer than the redox ladders discussed for FdhN. Similar systems may also operate during the respiration of other insoluble respiratory substrates, such as manganese.

Synthesis of Periplasmic Electron-Transport Systems

The periplasmic location of the cofactor-containing periplasmic electron-transfer proteins raises the question as to how the organism assembles these proteins. This question is particularly relevant to a large protein complex with multiple redox centers, such as FdhN. In the case of periplasmic cytochromes *c*, the protein is translocated in an unfolded state via the Sec pathway and heme is covalently attached using periplasmic or membrane-associated assembly proteins. However, many other classes of periplasmic redox protein possess a signal sequence, processed from the mature enzyme, that includes a twin arginine motif. This motif directs the protein to the twin arginine translocase (Tat), which transports prefolded proteins with assembled redox cofactors. Some Tat substrates, including FdhGH, are very large and consequently the pore of the Tat apparatus must be ~100 Å in diameter to accommodate the largest substrates. The structure of the Tat complex is not yet known but there must be a gating mechanism, which maintains an ionic seal upon transport of these large substrates. The translocation process is also energy-driven, consuming PMF. This illustrates how the overall consideration of the biochemistry and energetics of a periplasmic electron-transfer system must take into account not only the operation of the system but also its biogenesis.

See also: **Bioenergetics:** Cytochrome *bc*₁ Complex (Respiratory Chain Complex III); Cytochrome *c*; F_X, F_A, and F_B Iron–Sulfur Clusters in Type I Photosynthetic Reaction Centers; Heme Synthesis; Iron–Sulfur Proteins; **Protein/Enzyme Structure Function and Degradation:** Heme Proteins.

Further Reading

- Bertero MG, Rothery RA, Palak M, et al. (2003) Insights into the respiratory electron transfer pathway from the structure of nitrate reductase A. *Nature Structural and Molecular Biology* 10: 681–687.
- Einsle O, Stach P, Messerschmidt A, et al. (2000) Cytochrome *c* nitrite reductase from *Wolinella succinogenes*. Structure at 1.6 Å resolution, inhibitor binding, and heme-packing motifs. *Journal of Biological Chemistry* 275: 39608–39616.
- Ferguson SJ (2002) Keilin's cytochromes: How bacteria use them, vary them and make them. *Biochemical Society Transactions* 29: 629–640.
- Jormakka M, Tornroth S, Byrne B, and Iwata S (2002) Molecular basis of proton motive force generation: Structure of formate dehydrogenase-N. *Science* 295: 1863–1868.
- Page CC, Moser CC, Chen X, and Dutton PL (1999) Natural engineering principles of electron tunnelling in biological oxidation–reduction. *Nature* 402: 47–52.

- Palmer T and Berks BC (2003) Moving folded proteins across the bacterial cell membrane. *Microbiology* 149: 547–556.
- Potter L, Angove H, Richardson D, and Cole J (2001) Nitrate reduction in the periplasm of Gram-negative bacteria. *Advances in Microbial Physiology* 45: 51–112.
- Richardson DJ (2000) Bacterial respiration: A flexible process for a changing environment. *Microbiology* 146: 551–571.
- Richardson DJ and Sawers G (2001) PMF through the redox loop. *Science* 295: 1842–1843.
- Sawers G (1994) The hydrogenases and formate dehydrogenase of *Escherichia coli*. *Antonie van Leeuwenhoek* 66: 57–88.
- Simon J (2002) Enzymology and bioenergetics of respiratory nitrite ammonification. *FEMS Microbiology Reviews* 26: 285–309.
- Van Spanning RJ, Delgado MJ, and Richardson DJ (2003) The nitrogen cycle: Denitrification and relationship to N_2 fixation. In: Newton W (ed.) *Nitrogen Fixation*, pp. 1888–2001. Dordrecht: Kluwer.

Peroxidase Biochemistry and Redox Signaling

A Bindoli and M P Rigobello, University of Padua, Padua, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Glutathione (GSH) A tripeptide

(γ -glutamylcysteinylglycine) found in high concentrations in cells. It is involved in redox processes, being reversibly transformed to the corresponding disulfide (GSSG).

Oxidases Enzymes able to oxidize various substrates using molecular oxygen.

Oxidative stress Condition dependent on an imbalance between reactive oxygen species (ROS) production and the capacity of cells to detoxify them readily using their enzymatic (superoxide dismutase, catalase, and various

peroxidases) and nonenzymatic defenses (vitamin E and other antioxidants).

Peroxidases Enzymes which reduce hydrogen peroxide to water, or other hydroperoxides to the corresponding alcohols. They can also use hydrogen peroxide to oxidize substrates of various kinds.

Reactive oxygen species Some are free radicals. They include superoxide anion ($O_2^{\bullet-}$), hydrogen peroxide (H_2O_2), singlet oxygen (O_2^*), and hydroxyl radical ($\bullet OH$).

Redox signaling Cell signaling mediated by oxidation/reduction processes.

Cellular Respiration and Production of Reactive Oxygen Species

Cellular respiration sustains aerobic life and involves the oxidation of nutrients, with the final production of carbon dioxide and water. During this process, oxidation energy is captured in the form of adenosine triphosphate (ATP) molecules. Most of the oxygen is reduced to water by cytochrome *c* oxidase in a four-electron process. However, a small percentage of oxygen (1–3%) can be converted to the superoxide anion ($O_2^{\bullet-}$) in a mono-electronic process (Figure 1), depending on the leak of single electrons, mostly released by the mitochondrial and microsomal electron transport chains. However, several other sources of superoxide anion are also active in the cell. Superoxide undergoes dismutation to hydrogen peroxide (H_2O_2), either spontaneously or in a superoxide dismutase-catalyzed reaction. In the presence of transition metals such as iron or copper, the formation of the strong oxidizing hydroxyl radical ($\bullet OH$) is favored (Figure 1). Superoxide anion, hydrogen peroxide, and hydroxyl radical are collectively called reactive oxygen species (ROS) and considered the main agents responsible for oxidative stress. Hydrogen peroxide is a relatively stable oxidant and, when its concentration in cells increases, it must be controlled by several enzymes, in particular, catalase and the peroxidases dependent on glutathione or thioredoxin (Trx). However, several peroxidases use H_2O_2 for defensive or biosynthetic purposes. Lastly, it is increasingly recognized that H_2O_2 should be included among the cell-signaling molecules such as nitric oxide, calcium, and cyclic adenosine monophosphate (cAMP). Hence, peroxidases, enzymes capable of breaking down H_2O_2 to H_2O , act as modulators of the level of hydrogen peroxide, controlling its function as a second messenger.

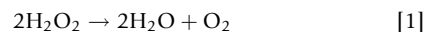
Classification, Distribution, Reaction Mechanism, and Physiological Role of Peroxidases

Hydrogen peroxide, generated either by dismutation of the superoxide anion or by the action of several oxidases, is

reduced to water by a large number of enzymes including catalase and peroxidases. The latter also act on several organic hydroperoxides which are reduced to their corresponding alcohols. Peroxidases are distinguished according to their active site cofactor, which may be a heme (heme peroxidases) or a redox-active cysteine or selenocysteine residue (nonheme peroxidases).

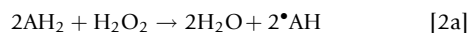
Heme Peroxidases

Heme peroxidases are characterized by the presence of a heme cofactor at their active site. Catalase is the best-known member of the hydrogen peroxide-decomposing heme enzymes, being found in all aerobic organisms and almost exclusively located in the peroxisomes. Animal catalase is a four-subunit enzyme, each of which has a Fe(III)-heme, which is oxidized in its catalytic cycle (Figure 2). Catalase decomposes hydrogen peroxide at the highest turnover rate known, acting on two molecules of hydrogen peroxide according to the reaction



Therefore, catalase is a special case of peroxidase, in which the reductant is a second molecule of hydrogen peroxide and the overall process is a dismutation reaction. Heme peroxidases constitute a very large group of peroxidases, which are divided into two superfamilies. The first includes plant, fungal, and bacterial peroxidases, further subdivided into three classes (I to III) according to distribution and amino acid sequence homologies. The second superfamily comprises mammalian peroxidases, including myeloperoxidase (MPO), lactoperoxidase (LPO), eosinophil peroxidase (EPO), and thyroid peroxidase (TPO). Although heme is noncovalently bound in plant and fungal peroxidases, in mammalian peroxidases heme is covalently linked to the protein.

The reaction mechanism of heme peroxidases proceeds by a two-electron oxidation reaction involving electron donor substrates (AH_2 or X^-) according to the general reactions



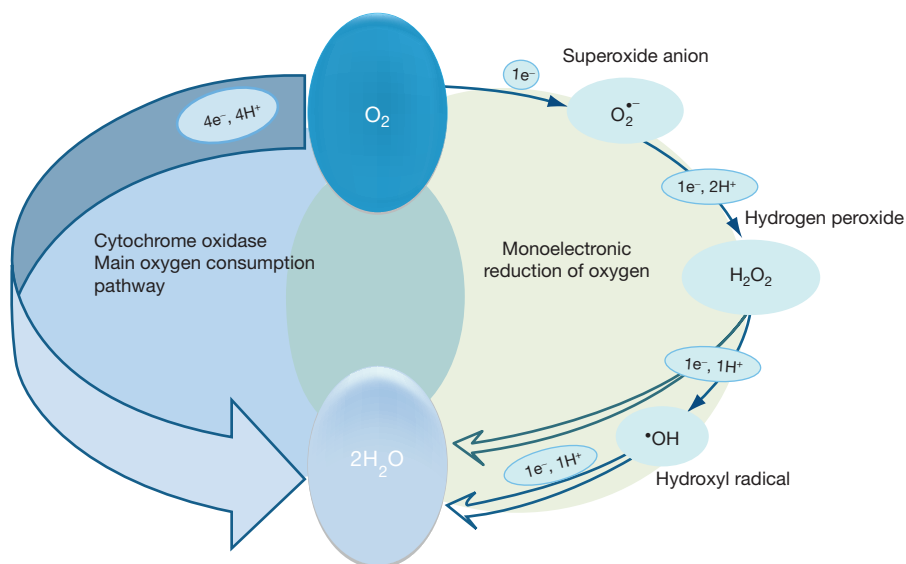


Figure 1 Reduction of oxygen and formation of oxygen-free radicals. Oxygen is largely reduced to water by the cytochrome oxidase complex, with the concomitant generation of energy in the form of ATP (left). However, a little oxygen can be reduced univalently, with the formation of reduced intermediates (right). The first reduction product is superoxide anion ($\text{O}_2^{\bullet-}$). Hydrogen peroxide is the two-electron reduced form of oxygen. In the presence of reduced iron ions, hydrogen peroxide undergoes the Fenton reaction ($\text{H}_2\text{O}_2 + \text{Fe(II)} \rightarrow \text{Fe(III)} + \text{OH}^- + \bullet\text{OH}$), leading to the formation of the highly oxidant hydroxyl radical ($\bullet\text{OH}$). Lastly, hydroxyl radical can be reduced to water. Other metals, particularly copper, can catalyze the formation of $\bullet\text{OH}$ from H_2O_2 .

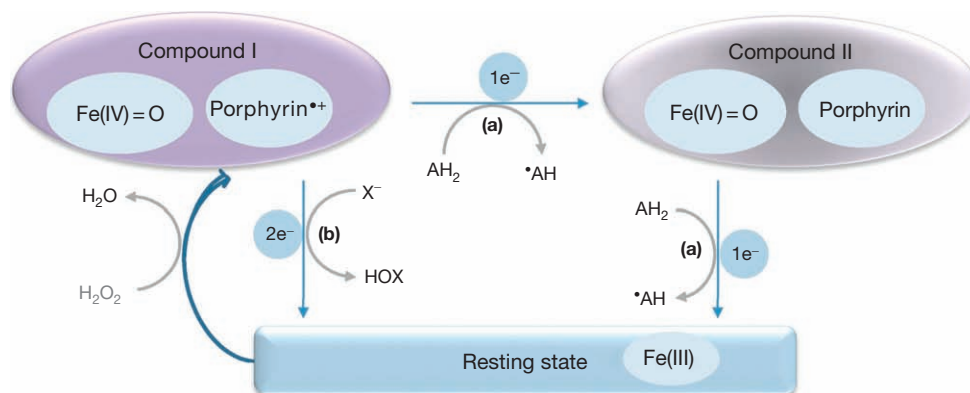
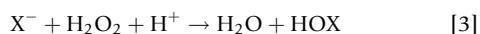


Figure 2 Catalytic mechanism of heme peroxidases. The first step is oxidation of Fe(III) in the active site of peroxidase at the resting state to compound I. Fe(IV)=O and $\text{Porphyrin}^{\bullet+}$ denote oxoferryl species and the porphyrin radical cation, respectively. In pathway (a), the electron donor substrate, generically indicated as AH_2 , transfers electrons in a two-step process, forming $\bullet\text{AH}$ radicals at each step. In the first step, compound II forms. The two radicals generated after oxidation of the substrate can dismutate or polymerize. Peroxidases use several substrates such as phenolics, thiols, reduced pyridine nucleotides, and several other chemically unrelated substances. Purified peroxidases are also able, in the presence of hydrogen peroxide, to oxidize many synthetic organic molecules. This pathway is observed in several plant peroxidases, e.g., horseradish peroxidase and ascorbate peroxidase. In pathway (b), electron-donor substrates (X^-) transfer two electrons to compound I, restoring the enzyme to its Fe(III) resting state. This occurs with halides ($\text{X}^- = \text{Cl}^-, \text{Br}^-, \text{I}^-$) and thiocyanate ($\text{X}^- = \text{SCN}^-$), which are transformed to hypohalous acids (HOCl , HOBr , and HOI) and hypothiocyanite (OSCN^-), respectively. Several animal peroxidases (myeloperoxidases, eosinophil peroxidases, lactoperoxidase, and thyroid peroxidase) follow this mechanism. In the case of catalase, hydrogen peroxide reduces compound I to the resting state Fe(III).



or



Substrates AH_2 or X^- are oxidized at the end of the reaction (A or HOX).

The reaction mechanism of catalase and peroxidases is a step process, as recognized by changes in electronic spectra. In the resting state, heme peroxidases contain a Fe(III) in their active site. Upon interaction with hydrogen peroxide, which removes two electrons, an oxoferryl porphyrin cation radical ($\text{Fe(IV)=O} \cdots \text{Porphyrin}^{\bullet+}$), called compound I, is formed. The oxidizing capacity of compound I resides on the oxoferryl

moiety and the one-electron oxidized heme radical cation ($\text{Por}^{\bullet+}$) (Figure 2). In some peroxidases, the electron can be transferred from amino acid residues such as tyrosine or tryptophan instead of porphyrin, resulting in amino acid radicals. Once formed, compound I can oxidize a reducing donor in a two-step or a single process (Figure 2). The first leads to compound II, in which the heme iron is still in oxoferryl form, but in which the porphyrin radical cation has been reduced by a substrate (AH_2). In turn, the substrate is oxidized to a free radical form ($\text{AH}^\bullet$). In the second step, compound II reacts with another molecule of the hydrogen donor (AH_2), forming the resting Fe(III) intermediate and then another substrate-free radical ($\text{AH}^\bullet$). The two substrate radicals may undergo different reactions, including a dismutation reaction [reaction 2b]. The type of reaction mechanism described occurs in several plant peroxidases such as horseradish peroxidase and ascorbate peroxidase. However, heme peroxidases can also catalyze direct two-electron oxidation of substrates, such as halides and thiocyanate, to the corresponding hypohalogenous acids and hypothiocyanite, respectively (Figure 2). This occurs in mammalian peroxidases such as MPO, LPO, and EPO. For instance, in MPO, the interaction of compound I with the largely available Cl^- forms the oxidizing species HOCl and restores the peroxidase to the resting state. In this case, only compound I is needed as an intermediate in the catalytic process (Figure 2), which occurs according to the following reaction:



MPO can also oxidize I^- which, however, is less concentrated than Cl^- in biological fluids. Although it can also oxidize Br^- , this reaction is probably not relevant *in vivo*. Hydrogen peroxide can further convert compound I to compound II, but the latter is inefficient in oxidizing chloride. In catalase, one molecule of hydrogen peroxide oxidizes the catalytic center to compound I, whereas the second molecule reduces the enzyme back to its resting state. Hydrogen peroxide therefore acts as both a reducing and an oxidizing agent.

Plant peroxidases oxidize phenolic compounds, lignin precursors, and other metabolites in order to accomplish a wide range of physiological processes, including defense against pathogens, lignin formation, cross-linking of cell-wall components, and metabolism of auxin.

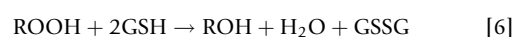
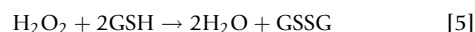
Specialized functions have been described for some animal tissue peroxidases. MPO, LPO, and EPO take part in unspecific immune defense; TPO acts in thyroid hormone biosynthesis. They are all able to catalyze the oxidation of halogens to form specific products of significant biological interest. MPO is present in neutrophils and monocytes and is stored in cytoplasmic granules. The latter fuse with the phagocytic vacuole and deliver the enzyme, which generates the highly reactive hypochlorous acid, easily contributing toward destroying pathogens. EPO is released by eosinophils, which represent substantial protection against parasites. EPO uses a catalytic mechanism similar to that of MPO, but prefers bromide (Br^-) instead of chloride. LPO is released from glands, and is found in milk, saliva, the mucus of the respiratory tract and tears, and provides a barrier against pathogenic microorganisms entering the body. LPO chiefly oxidizes thiocyanate (SCN^-), found in milk and saliva, to hypothiocyanite (OSCN^-), which is toxic to many bacteria.

TPO plays a crucial role in thyroid hormone biosynthesis. This peroxidase catalyzes the oxidation of the iodide ion (I^-) to a species (probably I^+) capable of iodinating tyrosines in a two-electron process. However, TPO can also catalyze the formation of iodotyrosyl radicals in one-electron oxidation of iodotyrosines.

Nonheme Peroxidases

The nonheme peroxidases include glutathione peroxidases and peroxiredoxins, the final links connecting the hexose monophosphate pathway to hydrogen peroxide removal.

Glutathione peroxidase reduces hydrogen peroxide, fatty acid hydroperoxides, and synthetic hydroperoxides, with the concomitant oxidation of glutathione (GSH) to the corresponding disulfide (GSSG) according to the reactions



Several isoforms of the glutathione peroxidase family have been described. Five contain selenium at their active site, that is, GPx1–GPx4 and GPx6. The latter is found as a selenoprotein in humans and a cysteine enzyme homolog in rodents. The reaction mechanism of selenium glutathione peroxidases involves oxidation of the selenol ($-\text{SeH}$), in the active site, to a selenenic acid ($-\text{SeOH}$) (Figure 3). This reaction is facilitated by the easy deprotonation of selenocysteine at physiological pH (the pK_a of selenocysteine is about 5.2). Selenol is therefore present in the ionized form ($-\text{Se}^-$), which readily reacts with hydrogen peroxide and other hydroperoxides. In the second step, one molecule of reduced glutathione reacts with selenenic acid, forming a mixed disulfide (Enzyme-Se-SG). This species is then reduced by a second molecule of GSH which completes the redox catalytic cycle, restoring the active site selenol (Figure 3).

GPx1–3 are tetramers, and each monomer contains a selenium atom at the active site. Instead, GPx4 is a monomer and therefore contains only one selenium atom. GPx1 is the classic glutathione peroxidase and is essentially dedicated to general protection against oxidative stress. In fact, in comparison with normal mice, mice lacking GPx1 are more sensitive to toxins which increase oxidative stress. In the cell, GPx1 is located in both cytosol and mitochondria. GPx2 (intestinal glutathione peroxidase) is primarily expressed in the gastrointestinal tract, but may also be found in the liver. It plays a protective role in the removal of peroxides in ingested food or generated in the intestine. GPx3 is a glycoprotein mostly expressed in the proximal tubule of the kidney. It is found, as an extracellular enzyme, in blood plasma (plasma glutathione peroxidase), the aqueous humor of the eye, amniotic fluid, and lung lining fluid. GPx4 has the distinctive capacity to reduce esterified fatty acid hydroperoxides such as oxidized phospholipids of biological membranes and lipoproteins and is therefore also called ‘phospholipid hydroperoxide glutathione peroxidase’ (PHGPx). GPx4 is highly expressed in the testis, where it is involved in sperm maturation, acting as a catalyst in the oxidation of protein thiol groups in the head of the sperm. GPx5 is a Se-independent glutathione peroxidase present in epididymis. In GPx5, selenocysteine is substituted by a cysteine. The expression of GPx6 is restricted to the olfactory epithelium.

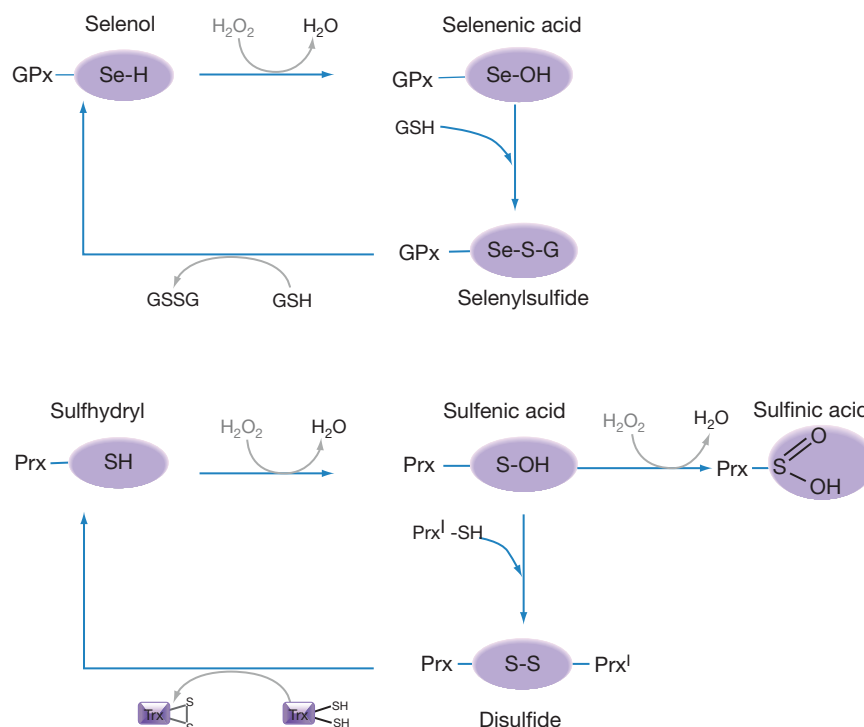
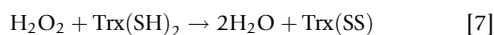


Figure 3 Comparison of catalytic mechanisms of glutathione peroxidases and peroxiredoxins GPx-SeH denotes glutathione peroxidase with its selenol group in the active site. Prx-SH represents peroxiredoxin with the peroxidatic cysteine (see text). Prx'-SH indicates another subunit of peroxiredoxin or another portion of the same subunit containing the resolving cysteine (see text). Hydrogen peroxide interacts with a selenol in the active site of glutathione peroxidase or with a reactive sulfhydryl in the active site of peroxiredoxin, forming a selenenic acid and a sulfenic acid, respectively. Selenenic and sulfenic acids react with another thiol group, forming a selenylsulfide and a disulfide, respectively, which may be further reduced by glutathione or thioredoxin, restoring the resting form of the two enzymes. Sulfenic acid in peroxiredoxin may be further overoxidized to sulfinic acid.

Peroxiredoxins are a family of peroxidases able to reduce hydrogen peroxide and other hydroperoxides in a reaction mostly coupled with the oxidation of Trx, according to the following reactions:



Trx(SH)₂ and Trx(SS) denote reduced and oxidized Trx, respectively.

In general, the donor-reducing substrate is Trx, but other substrates such as glutaredoxin and glutathione may also be involved. Mammalian cells express at least six isoforms of peroxiredoxins. Prx1, -2, and -6 are found in the cytosol, whereas Prx3 is typical of mitochondria. Prx5 is expressed in peroxisomes and endoplasmic reticulum, but also in mitochondria. According to the number and position of cysteine residues involved in the catalytic cycle, peroxiredoxins are organized into three subgroups, called typical 2-Cys peroxiredoxins (Prx1–4), atypical 2-Cys peroxiredoxin (Prx5), and 1-Cys peroxiredoxin (Prx6). All peroxiredoxins exhibit a similar basic cyclic catalytic mechanism, which essentially reproduces that shown for glutathione peroxidases (Figure 3). A major difference is that a cysteine, instead of selenocysteine, is the catalytic residue (Figure 3).

Prx1–5 need two cysteines for their catalytic cycle, a peroxidatic cysteine (C_P-SH) which reacts with hydrogen peroxide or

organic hydroperoxides and forms a sulfenic acid (C_P-SOH), and a resolving cysteine which forms a disulfide after reaction with the sulfenic moiety. In Prx5 (atypical Prx), the peroxidatic and resolving cysteines are located in the same molecule, whereas in Prx1–4 (typical Prx), each of the two catalytic cysteines is located in adjacent subunits. In the final step, Prxs are restored to their initial reduced state by hydrogen donors such as Trx. Peroxiredoxin 6 contains only the peroxidatic cysteine and, as resolving cysteine, uses one of other proteins or small molecules such as glutathione. The reactivity of the peroxidatic cysteine of peroxiredoxins is facilitated by the presence of positively charged groups which help to dissociate the thiol group, therefore forming a thiolate able to react with hydrogen peroxide. However, in addition to the low pK_a of the cysteine reacting with hydrogen peroxide, stabilization of the transition-state intermediate by a proper environment is also required. Although the reaction constants of peroxiredoxin with hydroperoxides are lower than those of glutathione peroxidases and catalase, their very high cellular abundance compensates for their lower catalytic efficiency.

At relatively high hydrogen peroxide concentrations, eukaryotic peroxiredoxins are distinctively subjected to overoxidation of their peroxidatic cysteine. In this process, the sulfenic moiety is oxidized to a sulfinic group and the enzyme is inactivated (Figure 3). It has been suggested that this inactivation fulfills a regulatory role in hydrogen peroxide signaling of eukaryotes. Interestingly, the presence of an ATP-dependent

enzyme, sulfiredoxin, able to reduce overoxidized peroxiredoxins back, stresses the importance of overoxidation of peroxiredoxins. Peroxiredoxins can undergo oligomerization, with the formation of ring-shaped complexes. However, the typical 2-Cys peroxiredoxins occur as homodimers and, in either their reduced or overoxidized forms, they can form decamers or dodecamers. The reduced oligomers are still redox active, whereas the overoxidized complex exhibits chaperone-like properties, helping to prevent protein aggregation.

In conclusion, heme and nonheme peroxidases make use of different active sites, such as the iron porphyrin prosthetic group, selenocysteine, or a deprotonated (reactive) cysteine. Other less common peroxidases containing vanadium or manganese at their active sites have also been described. This large variety of enzymes with active sites having different components and structures but with the unique function of removing H_2O_2 and other hydroperoxides emphasizes the physiological importance of this process and indicates that the cell cannot rely upon a single catalytic system when coping with peroxides.

Hydrogen Peroxide as a Signaling Molecule and the Controlling Role of Peroxidases

ROSs are not only responsible for oxidative stress, but, particularly at low levels, also considered to play an important role in normal cell metabolism, acting as modulators of redox-regulated processes. Several upstream effectors, such as growth factors, cytokines, insulin, and other hormones, are all able to activate signaling cascades by inducing the formation of superoxide and hydrogen peroxide. In turn, peroxidases endowed with the distinctive ability to scavenge hydrogen peroxide play a regulatory function in H_2O_2 signaling.

Glutathione peroxidases and peroxiredoxins are the terminal enzymes of the glutathione and Trx systems, respectively (Figure 4). The upstream reducing agent for both systems is reduced nicotinamide adenine dinucleotide phosphate (NADPH), which is kept reduced in the cytosol by the

reactions occurring in the hexose monophosphate pathway. In mitochondria, reducing equivalents are transferred from reduced nicotinamide adenine dinucleotide (NADH) to nicotinamide adenine dinucleotide phosphate ($NADP^+$) by the membrane-bound enzyme transhydrogenase. Glutamate and isocitrate dehydrogenases, capable of reducing $NADP^+$ directly, are further sources of this reduced coenzyme. In addition to NADPH, the glutathione system is also composed of glutathione reductase which, by keeping glutathione reduced, delivers electrons to glutathione peroxidase and to other proteins such as glutaredoxin and glutathione S-transferase (Figure 4). The Trx system is formed of NADPH, thioredoxin reductase (TrxR), and Trx (Figure 4). TrxRs are homodimeric proteins found in cytosol, mitochondria, and nucleus, characterized by a selenocysteine residue at their active site. TrxR reduces Trx in a reaction coupled with the oxidation of NADPH. Trx is a small redox protein with two cysteines in its active site, cycling between the reduced dithiol form ($Trx(SH)_2$) and the oxidized disulfide ($Trx(SS)$). It acts as an efficient reductant for several substrates and is a major donor of reducing equivalents for peroxiredoxins. Glutathione and Trx systems play an apparently similar role, but they act independently. An important difference is that glutathione, the main nonprotein thiol of the cell, is found at millimolar concentrations, whereas Trx occurs in the micromolar range.

One indication of the important role played by hydrogen peroxide in redox signaling is provided by the fact that overexpression of glutathione peroxidases or peroxiredoxins decreases the growth factor and cytokine-induced signaling, thus indicating that peroxidases can control H_2O_2 fluxes and hence the efficiency of the redox signal. This has been well described for platelet-derived growth factor receptor (PDGFR) and protein tyrosine phosphatase (PTP). Upon activation of the receptor, hydrogen peroxide is transiently produced and inactivates PTP molecules located close to the receptor. Consequently, phosphorylation is preserved and the growth signal is switched on. However, recruitment of peroxiredoxin by the receptor complex and the consequent removal of hydrogen

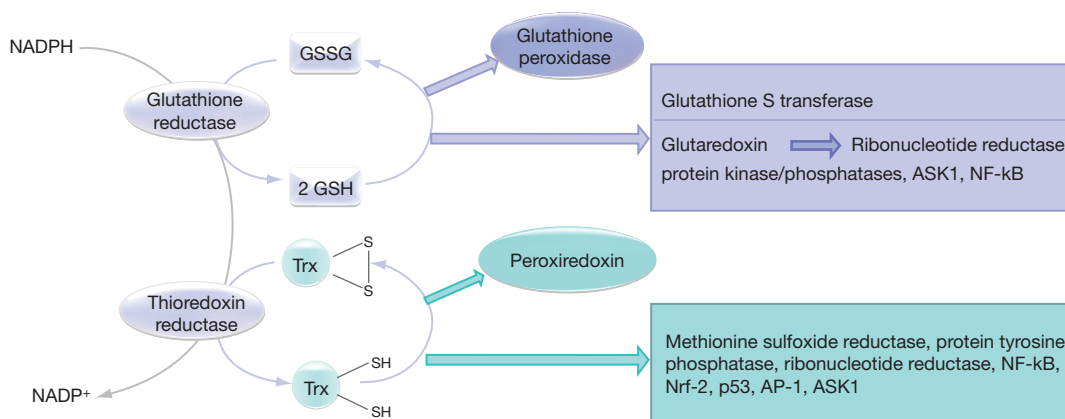


Figure 4 Glutathione and thioredoxin systems and their targets. The glutathione system is formed of glutathione, NADPH, and glutathione reductase; the thioredoxin system is composed of thioredoxin, NADPH, and thioredoxin reductase. Glutathione reductase and thioredoxin reductase transfer reducing equivalents from NADPH to glutathione and thioredoxin, respectively. Reduced glutathione acts as a substrate of glutathione peroxidase and of other enzymes such as glutathione S-transferase and glutaredoxin, which in turn, reduces ribonucleotide reductase, NF-κB, and other molecules. Similarly, thioredoxin, in addition to peroxiredoxin, can also transfer reducing equivalents to methionine sulfoxide reductase, ribonucleotide reductase, protein tyrosine phosphatase and other factors.

peroxide activate PTP, with a decrease in the growth signal. Similarly, peroxiredoxin prevents inactivation of the lipid phosphatase, phosphate and tensin homolog deleted from chromosome-10 (PTEN), induced by hydrogen peroxide.

Peroxiredoxins control not only the concentration of hydrogen peroxide but also the redox state of Trx, which is an important signaling event. For instance, after specific inhibition of TrxR and in the presence of hydrogen peroxide, Trx is almost completely oxidized by peroxiredoxin. In addition, oxidized Trx cannot be reduced back, as the inhibited TrxR is unable to transfer electrons from NADPH. Oxidized Trx can affect the functions of signaling molecules and alter membrane permeability. A well-known example is regulation of the apoptosis signal-regulated kinase-1 (ASK-1), belonging to the mitogen-activated protein kinase (MAP kinase) signaling pathway. ASK-1 is inhibited by reduced Trx, but this effect is lost upon oxidation of Trx. Inhibition of TrxR also leads to complete oxidation of Prx3, which may be responsible for increased permeabilization of mitochondrial membranes and the consequent release of proapoptotic factors. Lastly, overoxidation of peroxiredoxin, associated with its multimerization, address peroxiredoxins to a chaperone function which assists cells in recovery from the consequences of oxidative stress.

Peroxiredoxins can act as switches of oxidizing messages, as they are able to oxidize not only their thiol substrates but also other protein thiols. In this way, the highly reactive thiols of peroxiredoxins act as sensors of hydrogen peroxide and transfer this effect to other regulatory proteins which are otherwise unable to sense peroxides. In this case, the peroxidatic cysteine of peroxiredoxins, after oxidation to sulfenic acid, can interact with the cysteine of a transcription factor, leading to its activation. Trx regenerates both peroxiredoxin and the transcription factor to the native thiol forms. This thiol-oxidizing function of peroxidases has also been observed for Gpx4 in the formation of the capsula of spermatozoa.

In addition to removing hydrogen peroxide, peroxidases and catalase can also modulate cell signaling by association with a large number of regulatory proteins in a manner not necessarily dependent on their peroxidatic properties.

Concluding Remarks

Peroxidases are widely distributed peroxide-decomposing enzymes which carry out many physiological functions, not restricted to simple protective effects against potential peroxide toxicity. In addition to protective action, they also have

biosynthetic and immunological properties and modulate hydrogen peroxide signaling. They possess various catalytic sites such as heme, selenol, and thiol, with a unique reactivity toward peroxides. The extensive presence, wide distribution, and evolution of different mechanisms of catalysis of peroxidases reveal the importance of controlling the concentration and fluxes of hydrogen peroxide in living organisms.

See also: **Bioenergetics:** Mitochondria: Source and Target of Free Radicals; Oxygenases; Superoxide Dismutase; **Metabolism** **Vitamins and Hormones:** Pentose Phosphate (Hexose MonoPhosphate) Pathway; **Protein/Enzyme Structure Function and Degradation:** Disulfide Bond Formation; Selenoprotein Synthesis.

Further Reading

- Banerjee R (2008) *Redox Biochemistry*. Hoboken, NJ: Wiley.
- Bindoli A, Fukuto JM, and Forman HJ (2008) Thiol chemistry in peroxidase catalysis and redox signaling. *Antioxidants and Redox Signaling* 10: 1549–1564.
- Brigelius-Flohé R (1999) Tissue-specific functions of individual glutathione peroxidases. *Free Radical Biology and Medicine* 27: 951–965.
- Fourquet S, Huang M-E, D'Autreaux B, and Toledano MB (2008) The dual functions of thiol-based peroxidases in H₂O₂ scavenging and signaling. *Antioxidants and Redox Signaling* 10: 1565–1575.
- Furtmüller PG, Zederbauer M, Jantschko W, et al. (2006) Active site structure and catalytic mechanisms of human peroxidases. *Archives of Biochemistry and Biophysics* 445: 199–213.
- Halliwel B and Gutteridge JMC (2007) *Free Radicals in Biology and Medicine*, 4th edn. Oxford: Oxford University Press.
- Hofmann B, Hecht HJ, and Flohé L (2002) Peroxiredoxins. *Biological Chemistry* 383: 347–364.
- Kang SW, Rhee SG, Chang T-S, Jeong W, and Choi MH (2005) 2-Cys peroxiredoxin function in intracellular signal transduction: Therapeutic implications. *Trends in Molecular Medicine* 11: 571–578.
- Neumann CA, Cao J, and Manevich Y (2009) Peroxiredoxin I and its role in cell signaling. *Cell Cycle* 8: 4072–4078.
- Passardi F, Cosio C, Penel C, and Dunand C (2005) Peroxidases have more functions than a Swiss army knife. *Plant Cell Reports* 24: 255–265.
- Rhee SG, Chae HZ, and Kim K (2005) Peroxiredoxins: A historical overview and speculative preview of novel mechanisms and emerging concepts in cell signaling. *Free Radical Biology and Medicine* 38: 1543–1552.
- Stone JR and Yang S (2006) Hydrogen peroxide: A signaling messenger. *Antioxidants and Redox Signaling* 8: 243–270.
- Ursini F, Maiorino M, Brigelius-Flohé R, et al. (1995) Diversity of glutathione peroxidases. *Methods in Enzymology* 252: 38–53.
- Wood ZA, Schröder E, Robin Harris J, and Poole LB (2003) Structure, mechanism and regulation of peroxiredoxins. *Trends in Biochemical Sciences* 28: 32–40.

Peroxisomal Metabolism

S J Gould, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Published by Elsevier Inc.

The peroxisome is a single membrane organelle present in virtually all eukaryotic cells. While the enzymatic composition and metabolic function of peroxisomes can vary from cell to cell, tissue to tissue, and organism to organism, there are a few common themes to peroxisomal metabolism in mammalian cells. First, peroxisomes catalyze the oxidation of fatty acids. In humans, peroxisomes contain two fatty acid β -oxidation pathways as well as a fatty acid α -oxidation pathway. These pathways are essential for degrading very long chain fatty acids ($\geq C22$), 3-methyl branched-chain fatty acids, and for synthesizing bile acids. These peroxisomal pathways do not fully degrade their substrates, but instead reduce them to medium-chain length fatty acids that are exported for further β -oxidation in the mitochondrion. As for the oxidation of fuel fatty acids (e.g., palmitate, oleate), this is primarily a mitochondrial role in mammals and other animals, though peroxisomes are the sole site of all fatty acid β -oxidation in yeast and plants. Second, mammalian peroxisomal enzymes catalyze several early steps in the synthesis of ether-linked lipids. Third, inactivating mutations in the genes that code for key enzymes and transporters involved in peroxisomal β -oxidation, α -oxidation, or ether lipid synthesis cause severe neurological disorders, including X-linked adrenoleukodystrophy, D-bifunctional protein deficiency, rhizomelic chondrodysplasia punctata, and Refsum disease. Fourth, defects in peroxisome biogenesis impair all three of these core peroxisomal metabolic processes and can cause the most severe of all peroxisomal diseases, Zellweger syndrome.

Peroxisome Biogenesis

The metabolic functions of peroxisomes are dependent upon the proper biogenesis of the organelle and its constituent proteins. Peroxisomes lack nucleic acids and all peroxisomal proteins are encoded by nuclear genes. In general, peroxisomal enzymes and metabolite transporters are synthesized in the cytoplasm and imported posttranslationally into the peroxisome, though there are differences in the import of peroxisomal membrane proteins (PMPs) and peroxisomal matrix proteins, including most peroxisomal enzymes.

Biogenesis of Metabolite Transporters and Other PMPs

PMPs contain *cis*-acting peroxisomal targeting signals, mPTSs. These targeting signals are relatively large, multipartite signals that contain at least one transmembrane domain and one high-affinity binding site for PEX19, a chaperone/import receptor for newly synthesized PMPs (peroxisome biogenesis factors are known as 'peroxins', and referred to by the acronym PEX, with a dedicated number for each factor). PEX19 binds newly synthesized PMPs in the cytoplasm, maintains their

solubility in the cytoplasm, and delivers them to the outer surface of the peroxisome by binding to PEX3, an integral PMP that functions as a docking factor for PEX19–PMP complexes. The next step in PMP import is poorly understood, but involves the dissociation of PMPs from PEX19 and their insertion into the peroxisome membrane. While there is some evidence that certain peroxins that reside in the peroxisome membrane can traffic to peroxisomes via the endoplasmic reticulum (ER), peroxisomal metabolite transporters appear to follow the direct cytoplasm-to-peroxisome route of PMP import that is mediated by PEX19 and PEX3.

Given their central roles in PMP import, it is not surprising that loss of either PEX19 or PEX3 blocks PMP import and the eventual loss of peroxisomes from the cell. In mammalian cells, loss of PEX16 causes a similar phenotype, and while PEX16 does not appear to be required for import of most PMPs, it does play an essential role in the import of PEX3 into the peroxisome membrane.

Biogenesis of Peroxisomal Enzymes and Other Matrix Proteins

In contrast to the PMPs, peroxisomal matrix proteins (including most peroxisomal enzymes) have to be transported across the peroxisome membrane. These peroxisomal matrix proteins possess different types of PTSs, the PTS1 and PTS2. Unlike the poorly conserved, multipartite mPTSs that target PMPs into the peroxisome membrane, the PTS1 and PTS2 are short, discrete peptide motifs. The PTS1 is a tripeptide located at the extreme C-terminus and typically has the sequence serine–lysine–leucine_{COOH} or a conservative variant. Almost all peroxisomal matrix proteins possess a PTS1. The PTS2 is a nonapeptide-targeting signal located at or near the N-terminus, and mediates the import of only a very small number of peroxisomal enzymes. These targeting signals are recognized by the PTS1 receptor PEX5 and the PTS2 receptor PEX7.

PEX5 and PEX7 bind their cargoes of newly synthesized peroxisomal enzymes in the cytoplasm and then deliver them to the peroxisome via high-affinity binding to a specific docking factor, the peroxin PEX14. Peroxisomal matrix protein import also requires at least a dozen additional peroxins that catalyze the translocation of peroxisomal enzymes into the peroxisome lumen, as well as the recycling of PEX5 and PEX7 to the cytoplasm. These include a pair of AAA adenosine triphosphatases (ATPases) (PEX1 and PEX6), a PEX6-docking factor (PEX15/PEX26), ubiquitin-conjugating enzymes (PEX4/UBC5a,b,c), a PEX4-docking factor (PEX22), RING-containing PMPs that function as ubiquitin E3 ligases (PEX2, PEX10, and PEX12), PEX14-associated PMPs (PEX13 and PEX17), specialized chaperones for PEX7 (PEX18/20 or a specialized PEX5 isoform), as well as peroxins of less-defined function and conservation (PEX8, PEX9, PEX21, etc.).

Diseases of Peroxisome Biogenesis

Inactivating mutations in the genes encoding key factors involved in PMP import or peroxisomal matrix protein import cause Zellweger syndrome, a severe inherited disorder that is characterized by loss of all major peroxisomal metabolic functions (fatty acid β -oxidation, fatty acid α -oxidation, bile acid synthesis, and ether-phospholipid synthesis), an array of neurologic, hepatic, and renal abnormalities, and neonatal death. Hypomorphic mutations cause a spectrum of related diseases (e.g., neonatal adrenoleukodystrophy and infantile Refsum disease) that present with slightly less severe symptoms. The term 'Zellweger spectrum' is used to describe patients along this continuum of genetic, cell biological, and biochemical defects in peroxisome biogenesis. Current evidence indicates that these patients suffer from defects in 12 genes (*PEX1*, *PEX2*, *PEX3*, *PEX5*, *PEX6*, *PEX10*, *PEX12*, *PEX13*, *PEX14*, *PEX16*, *PEX19*, and *PEX26*) and that defects in *PEX1* are responsible for disease in nearly two-thirds of Zellweger spectrum patients, in part, because of two relatively common *PEX1* alleles in patients of European descent.

In contrast to the Zellweger spectrum disease, inactivating mutations in *PEX7* causes a distinct disorder known as 'rhizomelic chondrodysplasia punctata (RCDP)' type 1, which is characterized by rhizomelia, ichthyosis, and pronounced mental retardation. This disease is defined by an inability to import PTS2-targeted peroxisomal proteins, and this leads to selective defects in peroxisomal fatty acid α -oxidation and ether-phospholipid synthesis, which reflects the PTS2- and PEX7-dependent import of one key enzyme in each of these pathways: phytanoyl-CoA 2-hydroxylase and alkylldihydroxyacetone phosphate synthase, respectively. Interestingly, isolated defects in ether-phospholipid synthesis enzymes present a phenocopy of type 1 RCDP and are referred to as RCDP type 2 and 3 (see below).

Peroxisome Division

The most common route for the formation of new peroxisomes is the budding/scission of pre-existing peroxisomes to generate one or more peroxisomes. In humans, this mode of peroxisome biogenesis appears to involve peroxins of the *PEX11* family (humans express three *PEX11*-like genes) and the dynamin-like guanosine triphosphatase (GTPase) *DLP1*. Targeted disruptions of the *PEX11 α* and *PEX11 β* genes have been generated in mice, and loss of the *PEX11 β* gene displayed a number of Zellweger spectrum-like phenotypes, including impaired neuronal migration, hepatic abnormalities, and neonatal death. Neither *PEX11* gene appeared to be required for cell viability, whereas *DLP1* activity is essential for cell survival, which likely reflects its critical role in mitochondrial fission, and potentially in other aspects of cellular membrane dynamics. Interestingly, an isolated dominant mutation in *DLP1* caused a unique human disease with defects in both peroxisomal and mitochondrial function, and defects in both mitochondrial and peroxisomal dynamics, an array of clinical phenotypes, including some that resembled those of Zellweger spectrum patients.

Formation of Peroxisomes *De Novo*

Peroxisomes can also be formed *de novo*, at least under certain conditions. This phenomenon has only been definitively

observed under specialized laboratory conditions in which cells lacking *PEX19*, *PEX3*, or *PEX16* subsequently receive a functional copy of the defective gene, either by gene delivery or by cell fusion. The pathway of *de novo* peroxisome biogenesis appears to be quite slow, raising questions as to whether it makes a substantive contribution to peroxisome formation under physiologically relevant conditions. Nevertheless, *de novo* peroxisome formation is a ubiquitous property of eukaryotic cells and has been reported in organisms as diverse as yeast and human cells. Current models propose that *de novo* peroxisome formation involves the release of vesicles from the ER, which subsequently mature into peroxisomes that are capable of subsequent growth and division. As for whether defects in *de novo* peroxisome biogenesis affect peroxisome metabolism, this question is difficult to resolve since the key factors in *de novo* peroxisome formation are *PEX19*, *PEX3*, and *PEX16*, peroxins that also play direct and essential roles in PMP import.

Metabolic Functions of Peroxisomes

The purpose of the peroxisome is to catalyze specific metabolic processes in ways that complement the other metabolic activities of the cell and enhance the fitness of the organism. While the existence of unique peroxisomal metabolic activities has been recognized for more than four decades, the past decade has witnessed a systematic analysis of peroxisomal protein content, enzymatic activities, and genetically impaired cells and organisms that have revealed the full scope, significance, and limitations of peroxisomal metabolism. Nevertheless, certain details of peroxisomal metabolism still remain to be resolved, particularly in regard to metabolite transport across the peroxisome membrane.

Peroxisomal Fatty Acid β -Oxidation

Peroxisomal fatty acid β -oxidation occurs in organisms as diverse as yeast, plants, invertebrates, and vertebrates. In fact, the peroxisomal fatty acid β -oxidation pathway is the sole fatty acid β -oxidation pathway in yeast and plants, as these organisms lack a mitochondrial fatty acid β -oxidation pathway. Although cells can express multiple isoforms of each enzyme, the core of the peroxisomal fatty acid β -oxidation pathway consists of these four basic steps:

1. Oxidation of a fatty acyl-CoA to a 2-enoyl-CoA. In the peroxisomal pathways, this reaction is catalyzed by an acyl-CoA oxidase, an enzyme that couples the oxidation of acyl-CoAs to the reduction of molecular oxygen into hydrogen peroxide (H_2O_2). This step represents the major enzymatic difference between the peroxisomal and mitochondrial fatty acid β -oxidation pathways. This difference has significant bioenergetic consequences, as the free energy of acyl-CoA oxidation by the peroxisomal pathway is lost to the environment, whereas the free energy of acyl-CoA dehydrogenation by the mitochondrial pathway can be coupled to mitochondrial electron transport and ATP synthesis, yielding an additional 1–2 ATP per turn of the fatty acid β -oxidation spiral.

2. The second step in the pathway is the hydration of 2-enoyl-CoAs to 3-hydroxyacyl-CoAs, catalyzed by enoyl-CoA hydratase. Peroxisomal forms of enoyl-CoA hydratase are expressed as multifunctional peroxisomal proteins (e.g., D-bifunctional protein, or DBP).
3. 3-Hydroxyacyl-CoA dehydrogenase activity. This enzyme catalyzes the NAD^+ -dependent dehydrogenation of 3-hydroxyacyl-CoAs, yielding the corresponding 3-ketoacyl-CoA and $\text{NAD(H)} + \text{H}^+$.
4. Thiolase-catalyzed cleavage of the 3-ketoacyl-CoAs, releasing acetyl-CoA (or propionyl-CoA in the case of 2-methyl-branched chain fatty acids) and an acyl-CoA that is two carbon units shorter than the initial substrate (or three carbon units shorter in the case of 2-methyl branched chain fatty acids).

In addition to these core enzymatic steps of fatty acid β -oxidation, peroxisomes contain a number of ancillary proteins that play important roles in fatty acid oxidation. These include:

1. Metabolite transporters that import acyl-CoA molecules from the cytoplasm to the peroxisome lumen, such as the PXA1/2 proteins of yeast and the ABCD1–4 proteins of human cells. These are members of the ATP-binding cassette (ABC) superfamily of ATPases.
2. Acyl-CoA synthetases that can activate certain free fatty acids to their acyl-CoA derivatives.
3. An adenosine triphosphate/adenosine diphosphate (ATP/ADP) carrier that supports the intraperoxisomal activation of selected fatty acids to their fatty-acyl-CoA derivatives.
4. Enzymes that rearrange double bonds prevent the oxidation of unsaturated fatty acids. These include 3-enoyl-CoA isomerase, which converts 3-enoyl-CoAs into 2-enoyl-CoAs that substrates of the hydratase enzyme. This isomerase is essential for the oxidation of unsaturated fatty acids with double bonds at odd-numbered carbons. For unsaturated fatty acids with double bonds at even numbered fatty acids, two additional enzymes are required: a dienoyl-CoA isomerase and a dienoyl-CoA reductase, which act together to generate 3-enoyl-CoAs, which can be converted to 2-enoyl-CoAs substrates by the 3-enoyl-CoA isomerase.
5. Racemase activities that convert 2R-branched chain fatty acids, which are not substrates for peroxisomal acyl-CoA oxidases, into their 2S stereoisomers, which are the only form that can be degraded by the peroxisomal enzymes.
6. Acetyl-CoA export mechanisms (e.g., carnitine acetyltransferase) that transport acetyl groups from the peroxisome lumen to the cytoplasm, a necessary first step for their subsequent oxidation in the mitochondrion.
7. Metabolite shuttles (e.g., malate dehydrogenase-based shuttles in yeast, lactate dehydrogenase-based shuttles in mammalian cells) that (a) regenerate the free NAD^+ required for continued dehydrogenation of 3-hydroxyacyl-CoAs and (b) move the reduced form of electron carriers out of the peroxisome, a first step toward their subsequent use in electron transport and ATP synthesis.

The peroxisomal fatty acid β -oxidation pathways in animals have a somewhat restricted metabolic role. This is perhaps most clear in humans and other mammals, where the peroxisomal fatty acid β -oxidation pathways play little, if any, role in

the oxidation of the major fuel fatty acids (e.g., C16 and C18 fatty acids) but are essential for the oxidation of very long chain fatty acids (e.g., C24, and especially C26), bile acid precursors such as di- and trihydroxycholestanic acid (DHCA and THCA, respectively), pristanic and other 2-methyl branched-chain fatty acids, as well as long-chain dicarboxylic fatty acids. Moreover, the peroxisomal fatty acid β -oxidation pathways in mammals are thought to shorten fatty acids to medium chain length rather than reduce them entirely to their acetyl-CoA (and propionyl-CoAs in the case of 2-methyl branched-chain fatty acids). Consistent with this model, peroxisomes contain a carnitine-octanoyltransferase that facilitates the export of medium-chain length fatty acids from peroxisomes for subsequent oxidation in the mitochondrion.

In contrast to animals and certain other organisms, the peroxisomal fatty acid β -oxidation pathway is the sole mechanism for converting fatty acids into acetyl-CoA in yeast and plants. This means that the peroxisomal pathway plays a key bioenergetic role in these organisms, particularly under conditions when fatty acids are the sole energy source. This occurs normally during germination of oil seeds, which use the energy released by peroxisomal fatty acid oxidation and acetyl-CoA production, together with the subsequent oxidation of acetyl-CoA by the mitochondrial tricarboxylic acid cycle, to power their germination and early stages of development. It also occurs in yeast, perhaps most commonly in parasitic/pathogenic yeast, and also in yeast that grow near natural oil seeps, oil spills, and even in fuel tanks of diesel-powered vehicles.

Diseases of Peroxisomal Fatty Acid β -Oxidation

Despite the somewhat restricted role for the peroxisomal fatty acid β -oxidation pathways in humans and other animals, these pathways are still critical for human metabolism and development. This point is perhaps most obvious from the severe phenotypes that result from inactivating mutations in the genes encoding human peroxisomal β -oxidation enzymes. For example, inactivation of the human *ACOX1* gene, which encodes one of the human acyl-CoA oxidases, causes hypotonia, seizures, neurological abnormalities, and premature death, even though it only affects the oxidation of very long chain fatty acids. Inactivation of the gene encoding the D-bifunctional enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase (*HSD17B4*) gene causes an even more severe set of biochemical and clinical phenotypes, as it impairs the oxidation of almost all peroxisomal fatty acid β -oxidation substrates. Even defects in the peroxisomal 2-methyl-acylCoA racemase (encoded by the *AMACR* gene) result in clinical abnormalities, including a progressive loss of vision and other neurological abnormalities.

The most common human defect in peroxisomal fatty acid β -oxidation is caused by mutations in *ABCD1*, a gene located on the X-chromosome. This gene encodes an integral PMP that functions as a transporter for acyl-CoAs and is essential for the import of very long chain fatty acids into the peroxisome. Not surprisingly, inactivating mutations in *ABCD1* cause a pronounced accumulation of very long chain fatty acids throughout the body. Clinically, mutations in *ABCD1* can result in a spectrum of clinical phenotypes ranging from a

severe childhood disease, X-linked adrenoleukodystrophy, in which affected males develop normally until 5–8 years of age, and then undergo a rapid and progressive neurological deterioration that typically results in death within a few months, to a milder, later onset forms of the disease such as adrenomyeloneuropathy. Intriguingly, both the same mutations in *ABCD1* can give rise to dramatically different clinical phenotypes even in the same family, indicating that other loci and/or environmental factors can have a significant impact on the progression of this disease. As noted previously in the section on peroxisome biogenesis, defects in the import of PMPs or peroxisomal matrix enzymes can also cause significant defects in peroxisomal fatty acid β -oxidation.

Peroxisomal Contributions to Bile Acid Synthesis

As noted above, one class of substrates that is handled exclusively by the peroxisomal β -oxidation enzymes are the cholesterol-derived bile acid precursors DHCA and THCA. These large, lipophilic, 2-methyl branched chain fatty acids are first esterified with CoA, imported into the peroxisomal matrix, and then subjected to β -oxidation, release propionyl-CoA and choyl-CoA. Peroxisomes also contain a bile acyl-CoA: amino acid acyltransferase that transfers the bulky cholic acid moiety onto the amino acids taurine or glycine. This generates the bile acids that are subsequently exported from the peroxisome into the cytoplasm, and from the cytoplasm into the bile duct for subsequent delivery into the digestive tract.

Peroxisomal Fatty Acid α -Oxidation

Acting together, the core and ancillary enzymes of the peroxisomal and mitochondrial β -oxidation systems allow animals to oxidize a wide variety of dietary fatty acids. However, certain dietary fatty acids are resistant to β -oxidation, particularly 3-methyl substituted fatty acids such as phytanic acid (3,7,11,15-tetramethyl hexadecanoic acid). Nevertheless, animal cells rapidly oxidize these fatty acids due to the combined actions of (1) an exclusively peroxisomal fatty acid α -oxidation pathway, which shortens phytanic by one carbon and generates the 2-methyl branched-chain fatty acid known as pristanic acid and (2) the peroxisomal and mitochondrial fatty acid β -oxidation pathways, which degrade pristanic acid to acetyl-CoA and propionyl-CoA units.

Fatty acid α -oxidation pathway is a multistep process catalyzed by four peroxisomal enzymes:

1. Phytanic acid is activated to phytanoyl-CoA by a peroxisome-associated phytanoyl-CoA synthetase. There is some debate as to the location of this enzyme, and if it is located on the outer surface of the peroxisome, it seems likely that a metabolite transporter (possibly of the ABCD1–4 family) also plays a role in fatty acid α -oxidation.
2. The peroxisomal matrix enzyme phytanoyl-CoA 2-hydroxylase, a dioxygenase, converts phytanoyl-CoA hydroxylase to 2-hydroxyphytanoyl-CoA in a reaction that requires 2-oxoglutarate and oxygen.
3. 2-Hydroxyphytanoyl-CoA is degraded by 2-hydroxyphytanoyl-CoA lyase, releasing pristanal and formyl-CoA, which spontaneously decomposes to yield free CoA and formate, which subsequently evolves as CO₂.

4. Pristanal dehydrogenase catalyzes the NAD⁺-dependent oxidation of pristanal to pristanic acid, which is subsequently activated to pristanoyl-CoA, which is then subject to peroxisomal fatty acid β -oxidation.

As discussed above, the mammalian peroxisomal β -oxidation pathway only partially shortens its substrates, and this can be seen in the oxidation of pristanoyl-CoA (2,6,10,14-tetramethylpentadecanoyl-CoA) as well: three rounds of peroxisomal β -oxidation generate acetyl-CoA, 2 molecules of propionyl-CoA, and the branched-chain fatty acid 4,8-dimethylnonanoyl-CoA, which is then exported from the peroxisome for further oxidation in the mitochondrion. However, it is currently unclear whether this is mediated by the free fatty acid, a carnitine ester, or some combination of the two.

Refsum Disease

The importance of the fatty acid α -oxidation pathway in humans is evident from the severe disease that results from inactivating mutations in the phytanoyl-CoA 2-hydroxylase gene. Loss of this enzyme activity causes Refsum disease, which is characterized by early-onset retinitis pigmentosa, as well as a spectrum of more variable symptoms including deafness, neuropathy, ataxia, and heart disease. These manifestations are caused by the pronounced accumulation of phytanic acid, reaching levels as high as 30% of total body fatty acids. The phytanic acid that accumulates in this disease is of complex dietary origin. Humans acquire most of their phytanic acid through the consumption of meat, milk, and milk products (butter, cream, etc.) from cattle, goats, sheep, bison, yaks, camels, and other ruminants, sources that are rich in both phytanic acid and its precursor, the fatty alcohol phytol. As for the primary source of phytanic acid/phytol, it derives from the phytol side chain of chlorophyll. Phytol is cleaved from the chlorophyll molecule by rumen microorganisms (relatively little is released by nonruminant digestion of chlorophyll-containing foods). The free phytol generated in the rumen is typically oxidized to phytanic acid and incorporated into triglycerides and phospholipids. Consistent with this dietary origin for phytanic acid/phytol, dietary therapies that minimize the ingestion of ruminant-derived foods have therapeutic benefits for Refsum disease patients.

Ether–Phospholipid Biosynthesis

The classic structure of phospholipids is based on glycerol-3-phosphate esterified with fatty acids on its 1 and 2 carbons. However, ~15–20% of phospholipids in humans have a distinct structure in which a fatty alcohol is attached to the 1 carbon through an ether linkage. These ether-linked phospholipids are synthesized by the concerted action of enzymes that are located in the peroxisome, cytoplasm, and ER. The earliest enzymes in the pathway are located in/on the peroxisome, and these peroxisomal enzymes generate the alkyl-dihydroxyacetone phosphate (alkyl-DHAP) that serves as the essential precursor for ether–phospholipid synthesis. The key steps are:

1. A peroxisomal glycerol-3-phosphate dehydrogenase that oxidizes glycerol-3-phosphate to generate DHAP.

2. An intraperoxisomal DHAP acyltransferase that catalyzes the formation of 1-acyl-DHAP from DHAP and a fatty acyl-CoA. The acyl-DHAP released by this enzyme is one of the two substrates for the synthesis of alkyl-DHAP.
3. An NADPH-dependent fatty acyl-CoA reductase, located on the outer surface of peroxisomes, reduces acyl-CoAs to fatty alcohols, which are the other key substrate for making alkyl-DHAP.
4. Alkylidihydroxyacetone phosphate synthase, located in the peroxisome lumen, catalyzes the formation of 1-alkyl-DHAP using 1-acyl-DHAP and a fatty alcohol as its substrates and also releases a free fatty acid.
5. Alkyl-DHAP reductase, located on the cytoplasmic surface of both peroxisomes and the ER, catalyzes the NAD^{+} -dependent formation of 1-alkyl-glycerol-3-phosphate (1-alkyl-G3P).

In subsequent ER-localized reactions, 1-alkyl-G3P is converted to 1-alkyl-2-acyl-G3P, and after several additional steps, to 1-alkyl-2-acyl-glycerophosphatidylcholine and 1-alkyl-2-acyl-glycerophosphatidylethanolamine that make up the bulk of ether-linked phospholipids in mammalian cells.

RCDP is Caused by Defective Ether Phospholipid Synthesis

As with the other main peroxisomal metabolic functions, inactivating mutations in the genes encoding DHAP acyltransferase of alkyl-DHAP synthase cause severe human disease: RCDP type 2 and type 3. These diseases are clinically similar to RCDP type 1, which is caused by defects in the PTS2 import receptor, PEX7, and present with rhizomelia, ichthyosis, and pronounced mental retardation. The phenotypic similarities between type 1, 2, and 3 RCDP indicates that type 1 RCDP is caused primarily by the defect in ether-phospholipid synthesis. While the clinical similarities between a defect in PTS2 protein import and a defect in ether-phospholipid synthesis might not be immediately obvious, they make more sense when one considers that (1) alkyl-DHAP synthase is a PTS2-targeted protein and is imported into peroxisomes in a PEX7-dependent manner and (2) Refsum disease is a later onset, milder disease than RCDP and, thus, the consequences of defective fatty acid α -oxidation (phytanoyl-CoA 2-hydroxylase is also imported in a PTS2- and PEX7-dependent manner) in the RCDP type 1 patients are overshadowed by the more severe consequences of defective ether-phospholipid synthesis.

Additional Peroxisomal Enzymatic Functions

Peroxisomes also contain several enzymes that perform important cellular functions, yet lie beyond the scope of the core metabolic pathways outlined above. These include a large number of H_2O_2 -producing oxidases, including L- α -hydroxyacid oxidase, sarcosine/pipecolic acid oxidase,

D-amino acid oxidases, and (in nonprimates) urate oxidase. Complementing these peroxisomal oxidases are an array of antioxidative peroxisomal enzymes, including catalase, peroxiredoxin V, Cu/Zn-superoxide dismutase, epoxide hydrolase, and a glutathione S-transferase, as well as ancillary redox reactions that balance electron flow within the peroxisome and between the peroxisome and the cytoplasm (e.g., isocitrate dehydrogenase). Such systems presume the existence of specific metabolite carriers and/or a selective permeability of the peroxisome membrane to the requisite metabolites, both of which have empirical support. Although there has always been strong interest in the concept that the peroxisome's primary role is to serve as a protected site for the metabolism of reactive oxygen species, this hypothesis has thus far failed to receive strong genetic support, as disruption of catalase has little, if any, phenotypic consequence in mice. In fact, the only human disease that is caused by a defect in one of these additional peroxisomal enzymes is primary hyperoxaluria type 1. This disease, which is characterized by lethal accumulation of calcium oxalate kidney stones, is caused by defects in the peroxisomal alanine:glyoxylate aminotransferase, which uses alanine to donate an amino group to glyoxylate, releasing the glycine and pyruvate. In the absence of peroxisomal alanine:glyoxylate aminotransferase activity, glyoxylate is instead converted to oxalate, with severe pathological consequences.

See also: [Metabolism Vitamins and Hormones: Fatty Acid Metabolism and Cancer](#).

Further Reading

- Antonovkov VD, Grunau S, Ohlmeier S, and Hiltunen JK (2010) Peroxisomes are oxidative organelles. *Antioxidants and Redox Signaling* 13: 1–13.
- Deleille HK, Alves R, and Schrader M (2009) Biogenesis of peroxisomes and mitochondria: Linked by division. *Histochemistry and Cell Biology* 131: 441–446.
- Fagarasanu A, Fagarasanu M, and Rachubinski R (2007) Maintaining peroxisome populations: A story of division and inheritance. *Annual Review of Cell and Developmental Biology* 23: 321–344.
- Ferdinandusse S, Denis S, Faust PL, and Wanders RJA (2009) Bile acids: The role of peroxisomes. *Journal of Lipid Research* 50: 2139–2147.
- Girzalsky W, Platta HW, and Erdmann R (2009) Protein transport across the peroxisome membrane. *Biological Chemistry* 390: 745–751.
- Matsuzaki T and Fujiki Y (2008) The peroxisomal membrane protein import receptor PEX3 is directly transported to peroxisomes by a novel PEX19p- and PEX16p-dependent pathway. *Journal of Cell Biology* 183: 1275–1286.
- Sacksteder KA and Gould SJ (2000) The genetics of peroxisome biogenesis. *Annual Review of Genetics* 34: 623–652.
- van den Bosch H, Schutgens RBH, Wanders RJA, and Tager JM (1992) Biochemistry of peroxisomes. *Annual Review of Biochemistry* 61: 157–197.
- Wanders RJA and Waterham HR (2006) Biochemistry of mammalian peroxisomes revisited. *Annual Review of Biochemistry* 75: 295–332.
- Weller S, Gould SJ, and Valle D (2003) Peroxisome biogenesis disorders. *Annual Review of Genomics and Human Genetics* 4: 165–211.
- Wierzbicki AS (2007) Peroxisomal disorders affecting phytanic acid α -oxidation: A review. *Biochemical Society Transactions* 35: 881–886.

Peroxisome Proliferator-Activated Receptors

M C Sugden, P W Caton, and M J Holness, Queen Mary University of London, London, UK

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Mary C. Sugden, Edward A. Sugden, and Mark J. Holness, volume 3, pp. 239–245, © 2004, Elsevier Inc.

Glossary

Fibrates A group of compounds that act as agonists for peroxisome proliferator-activated receptor alpha (PPAR α).

Peroxisome proliferator-activated receptors

(PPARs) These are ligand-activated transcription factors,

which alter gene expression to confer distinct properties of lipid handling to the tissues in which they are found, with an overall systemic lipid-lowering effect.

Thiazolidinediones (TZDs) A group of compounds that act as agonists for PPAR γ .

Lipids (triacylglycerol (TAG), cholesterol, phospholipids, fatty acids (FAs), and ketone bodies) are essential for many aspects of cellular processes, while also being linked to many pathological processes. Due to the varying demands of individual tissues for lipids, a highly regulated lipoprotein highway transports lipids to and from the periphery. The peroxisome proliferator-activated receptors (PPARs), α , γ , and δ , act as primary 'metabolic' transcription factors, responding to FAs and other metabolic signals to alter gene expression and thereby to confer distinct properties of lipid handling to the tissues in which they are found, with an overall systemic lipid-lowering effect. Thus, the PPARs act as a team with a common goal, but possess different attributes, including regulation of fat breakdown by oxidation (PPAR α and PPAR δ), regulation of adipocyte differentiation and fat storage (PPAR γ), together with other functions that include regulation of cholesterol fluxes into and out of cells (PPAR γ and PPAR δ).

Lipids are essential for many aspects of cellular processes, including acting as substrates for energy production, steroid hormones, and cellular structures. However, the availability and utilization of the various lipids must be carefully regulated because they are also linked to many pathological processes, including the development of obesity. The most familiar biological lipids are triacylglycerol (TAG), which serve as principal energy stores, and cholesterol, both of which are present in the circulation at detectable levels. Because the hydrophobic lipids do not dissolve in water, moving them around the body requires transport mechanisms that are highly regulated. Due to the variable requirements for these compounds from a wide range of tissues, a lipoprotein highway exists that allows transport of the hydrophobic lipids with water-soluble protein escorts enabling the movement of TAG and cholesterol to and from the periphery. TAG-enriched lipoproteins, chylomicrons (containing TAGs made from the reesterification of newly absorbed FA and monoacylglycerol (MAG) in the gut) and very-low-density lipoproteins (VLDLs) (containing TAG fabricated in the liver), deliver TAG to peripheral tissues for storage and energy production. The utilization by individual tissues of TAG in VLDL and chylomicrons is carefully regulated through the expression of lipoprotein lipase (LPL), which is anchored to the outside of the cell in which it is expressed. LPL liberates FA and MAG in TAG for uptake by the tissue.

Consequently, individual tissues can regulate their uptake of TAG-derived FA by altering the level of expression of LPL. At the same time, low-density lipoproteins (LDLs) deliver cholesterol to tissues so that they can maintain membrane integrity and, in some instance, use it for the production of biologically important compounds, such as steroids. Cholesterol uptake by cells is also regulated, in this case by the expression of the LDL receptor. In addition to the lipid highway from the liver/gut to peripheral tissues, a counter-flow operates in the opposite direction, whereby high-density lipoproteins (HDLs) transport surplus cholesterol from cells, such as those of the vessel wall, to the liver, where it can be redeployed to form bile. Gridlock in these lipid transfer and utilization systems leading to the accumulation of VLDL and LDL in the circulation is integral to the development of metabolic diseases, including type 2 diabetes mellitus (T2DM) and cardiovascular disease.

PPARs: A Three-Pronged Attack

PPARs are transcription factors, activated by binding metabolic ligands, which exert their effects at the level of DNA to alter gene expression. There are three PPAR subclasses (α , γ , and δ); each of these plays a specific role in metabolic regulation. PPARs were originally identified by their ability to induce peroxisome proliferation, and this led to the current nomenclature. PPAR α was first identified as the intracellular target for a particular class of rodent hepatocarcinogens, which include clofibrate. Clofibrate, a member of the fibrate amphipathic carboxylic acid family, was found to be a potent inducer of proliferation of peroxisomes, cell organelles found in high quantities in the liver, in addition to increasing FA oxidation and lowering lipid levels. Several other 'peroxisome proliferating' agents were subsequently cloned. However, it has since emerged that the PPARs do not induce peroxisome proliferation in man, but have an array of additional functions, most notably as master regulators of genes involved in metabolic control, particularly those involved in whole-body lipid homeostasis. Inappropriately, the name PPAR has remained in common use! The PPAR subtypes are distributed nonuniformly in tissues. PPAR α is expressed in a wide range of tissues and is highly expressed in liver, kidney, and heart. PPAR γ has a

more limited tissue distribution and is highly expressed in adipocytes and macrophages. PPAR δ is widely expressed. After the PPARs are activated by binding specific ligands and/or by covalent modification (phosphorylation), they bind to another receptor, the 9-*cis* retinoic acid-activated receptor (RXR), and conjoined and active, recruit coactivators and together bind to peroxisome proliferator response elements to modify the expression of specific genes. This leads to an altered complement of enzymes in the cell, with a resultant alteration in cellular functional characteristics. Through this mechanism, each PPAR can, once activated, confer distinct properties of lipid handling to the tissues in which it is found, with an overall systemic lipid-lowering effect. The individual PPARs act to reinforce each other's activities and also have complementary individual activities, all of which are concerned with the removal of excess lipid from the blood. Acting as a team with a common goal but differing attributes, the functions of the PPARs include regulation of fat breakdown by oxidation (PPAR α and PPAR δ), regulation of adipocyte differentiation and fat storage (PPAR γ), together with others that include regulation of cholesterol fluxes in to and out of cells (PPAR γ and PPAR δ). The molecular messengers that bind to and activate the individual PPARs (the endogenous ligands) are not all identified. However, the close involvement of fat metabolism and lipid homeostasis with the PPAR family members arises, in part, because certain FAs and other lipid-derived molecules, including eicosanoids and arachidonic acid derivatives, can activate PPARs and are known to be primary natural ligands for these molecules. For example, PPAR α is activated by polyunsaturated FAs, and PPAR δ by long-chain FAs. Thereby, these lipid-signaling molecules can have power over their own fates by binding to PPARs and changing the expression of genes involved in modulating the patterns of tissue FA handling. The physiological importance of the PPARs, thus, resides in lipid management, and the clinical importance of drugs acting via the PPARs lies in their ability to mimic nature's own messengers when the task of lipid management becomes too onerous and harmful lipids accumulate. It is now established that the fibrates (e.g., clofibrate) and the antidiabetic insulin-sensitizing thiazolidinediones (TZDs) both exert their systemic lipid-lowering effects via the PPARs. In this article, we focus on how these receptors modulate lipid homeostasis in metabolically active tissues.

Peroxisome Proliferator-Activated Receptor Alpha

Too much FA, TAG, or cholesterol circulating in the blood can be hazardous to health. High-fat diets, particularly those rich in very-long-chain FAs and polyunsaturated FAs, induce FA oxidation via PPAR α . Synthetic ligands for PPAR α , in addition to the fibrate drugs, include phthalate ester plasticizers, herbicides, and leukotriene D4 receptor antagonists. The fibrate PPAR α agonists (clofibrate, gemfibrozil, fenofibrate, bezafibrate, and ciprofibrate) have been in clinical use for over 40 years for the treatment of dyslipidemia.

The liver occupies a central position in whole-body lipid handling. In the case of FAs, the liver is capable of either oxidizing, or storing FAs depending on the prevailing requirements.

FAs are stored after conversion to the TAG component of VLDL, and this function can be increased when FA delivery is higher than that required to fuel hepatic adenosine triphosphate (ATP) requirements via β -oxidation. Incoming FAs, derived from the diet or endogenous stores, are reesterified and stored in the fed state. During starvation, FAs, released from adipose tissue, are taken up by the liver and oxidized to create the water-soluble lipid-derived fuels, the ketone bodies (which are easily transported and oxidized by peripheral tissues), while TAG is incorporated into VLDL which is released into the circulation. These latter metabolic transformations of incoming FAs are enhanced by PPAR α , the primary PPAR subtype expressed in liver. PPAR α regulates specific genes that code for regulatory proteins involved in almost every stages of hepatic FA utilization, including FA uptake into the cell across the cell membrane (FA transport proteins), its retention within the cell, and its oxidation leading to ATP production and ketone body formation (Figure 1). It also plays a role in lipoprotein formation and cholesterol synthesis. PPAR α , therefore, fulfills a critical role in the regulation of hepatic FA handling and, indeed, is essential for the upregulation of the capacity of the liver for FA utilization in starvation or when lipid delivery is increased by a diet high in fat. This adaptive role is best illustrated by the inadequate response to high-fat feeding seen in PPAR α -deficient mice, the Jack Sprats of lipid biochemistry. These animals exhibit an impaired ability to upregulate hepatic FA oxidation in response to a high-fat diet, despite increases in FA supply, and the liver becomes engorged with lipid. Similarly, PPAR α -deficient mice are unable to adapt to starvation, where delivery of FA from adipose tissue is increased. Secondary to impaired FA oxidation, compromised hepatic energy generation results in a low blood glucose concentration (hypoglycemia) when PPAR α -deficient mice are fasted. Hypoglycemia arises by virtue of the necessity to continue to use glucose at high rates, rather than to be able to switch to FA or (in extrahepatic tissues) ketone bodies as alternative energy fuels, and because hepatic FA oxidation allows hepatic glucose synthesis (gluconeogenesis) through the generation of ATP and essential cofactors. The accumulation of large amounts of lipid in livers of PPAR α -deficient mice, either on fasting or when they are maintained on a high-fat diet, demonstrates the critical importance of PPAR α -linked functions in matching hepatic FA delivery to FA clearance when the systemic delivery of lipid is increased. Hence, the TAG-lowering action of fibrates and related PPAR α -targeted drugs in patients with hyperlipidemia is ascribed to increased FA clearance via oxidation, primarily but not exclusively in liver. Within the liver, PPAR α is also heavily involved in the regulation of the synthesis and catabolism of lipoproteins, including some steps of the HDL synthetic pathway. Increased diversion of FA into oxidation decreases the availability of fatty acyl-coenzyme A (CoA) substrates for TAG synthesis, and thereby decreases VLDL secretion by the liver. In addition, PPAR α inhibits expression of apolipoprotein (apo) C-III (a protein that may inhibit the TAG-hydrolyzing action of LPL) and so enhances the hydrolysis of TAG contained in lipoproteins, and increases the hepatic uptake of TAG-enriched lipoprotein remnants. This decreases plasma TAG levels further and increases the transfer of other surface components of TAG-rich particles to HDL.

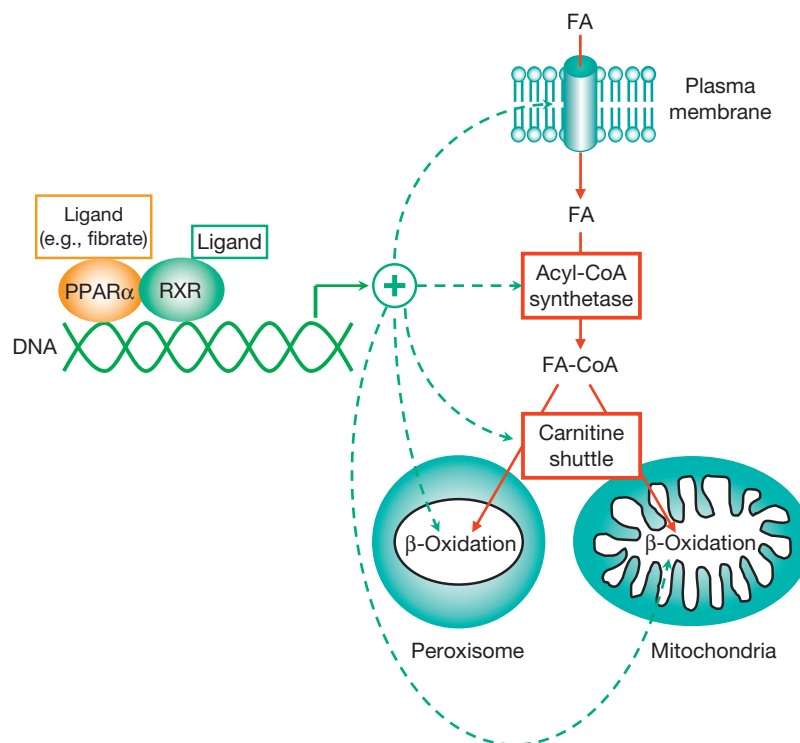


Figure 1 Mechanisms whereby activation of PPAR α regulates fatty acid uptake, sequestration in the cell and oxidation.

More recently, PPAR α studies have focused on elucidating the regulatory mechanisms governing PPAR α -mediated induction of gene expression. In particular, the essential role of PGC1 α has been described. PGC1 α interacts with PPAR α to induce expression of PPAR α target genes, for example, that encoding the rate-limiting enzyme for the transport of FA into the cellular organelles where they are further metabolized. This process is enhanced by several other regulators, such as sirtuin 1 (SIRT1) and lipin-1. SIRT1 is an enzyme involved not only in the regulation of cellular metabolism, but also in the responses to stress and the modulation of life span. SIRT1 activation in mice protects them against obesity induced by feeding a high-fat diet. Importantly, activation of an axis involving SIRT1 and an enzyme that senses metabolic stresses such as fuel starvation (adenosine monophosphate-activated protein kinase; AMPK) increases FA oxidation by covalent modification of their targets. SIRT1 is reported to deacetylate and activate PGC1 α , leading to induction of PPAR α -mediated expression of FA-oxidation genes, while liver-specific SIRT1 knockout mice develop a fatty liver when fed a high-fat diet. However, in a separate study, SIRT1-mediated induction of PGC1 α was shown to induce gluconeogenesis, while knockdown of SIRT1 and PGC1 α lowered blood glucose in normal and diabetic mice. These two studies highlight the problems of PPARs as a therapeutic tool for treatment of metabolic syndrome.

Another protein, lipin-1, is implicated in the regulation of cellular lipid homeostasis in several tissues, including liver. Lipin-1 regulates both intracellular lipid synthesis (through functioning as an enzyme, phosphatidate phosphatase-1 (PAP-1), producing an important glycerolipid building

block) and lipid clearance (by activating lipid regulation of gene expression). The role of lipin-1 in transcriptional coactivation was uncovered in a mouse model of PGC1 α deficiency. Within this latter context, lipin-1 expression is induced by PGC1 α and additionally lipin-1 also plays a key role in PPAR α regulation, functioning to induce PPAR α gene expression as well as interacting with and amplifying PGC1 α -PPAR α -mediated gene expression. The involvement of lipin-1 in this regulatory circuit leads to further confusion regarding the suitability of PPAR agonists as therapies for T2DM. Lipin-1 has been reported to increase TAG synthesis and is increased in diabetic mouse models, while knockdown of lipin-1 corrects insulin-resistant mice. Moreover, lipin-1 interacts with other nuclear factors, including hepatocyte nuclear factor 4 α (HNF4 α), a key transcription factor for induction of gluconeogenic genes.

Further studies will attempt to elucidate the mechanisms by which PPAR α -PGC1 α mediate gene expression, with particular focus on the roles of SIRT1 and lipin-1. This is of particular clinical relevance given that agonists of both PPAR α and SIRT1 are currently under investigation as therapies for T2DM.

As well as in liver, PPAR α is expressed abundantly in other tissues that are normally characterized by high rates of lipid oxidation. It is also present in the artery wall in monocytes/macrophages, vascular smooth muscle cells, and endothelial cells. We focus here on the role of PPAR α in kidney and heart, tissues which have a high-energy demand related to their physiological functions. In these tissues, PPAR α activation again enhances the expression of genes involved in FA uptake, activation, and oxidation. PPAR α -deficient mice subjected to ischemia/reperfusion injury respond with greater renal cortical

necrosis and impaired kidney function compared with wild-type controls, indicating the vital role of PPAR α in this tissue. PPAR α is found predominantly in the renal cortex. Like liver, the renal cortex is a site of gluconeogenesis. Altered renal gluconeogenesis is increasingly recognized as contributing to fasting hyperglycemia in diabetes mellitus, and assumes increasing importance for maintenance of glycemia after very prolonged food deprivation. Hence, as in liver, PPAR α -facilitated gluconeogenesis driven by enhanced FA oxidation is likely.

Although functionally quite distinct from liver and kidney, there is ample evidence for an important role for PPAR α in cardiac function in some circumstances. In adulthood, although the heart is a metabolic omnivore using most available fuels, lipids constitute the predominant energy substrate. However, during fetal and early life, in cardiac hypertrophy and in the failing heart, glucose utilization assumes greater importance, possibly as an adaptation to limited oxygen availability. Oxygen is essential for ATP production from FAs, but glucose oxidation needs less oxygen per mole than FA oxidation and glucose utilization can proceed, albeit less efficiently, in the absence of oxygen. The downside is that less ATP is generated per mole of glucose oxidized and when excess dietary lipid can be ingested, as in life after birth, the potential exists for myocardial lipid accumulation and lipotoxicity. Accordingly, the importance of PPAR α in the adult heart appears relatively limited when the dietary lipid supply or endogenous lipid production is restricted, but PPAR α assumes an important role in cardiac lipid management when lipid delivery to the heart is increased. PPAR α -deficient mice provided with free access to a high-carbohydrate low-fat diet do not show any obvious cardiac abnormalities, but they accumulate myocardial lipid on starvation or if a high-fat diet is provided. This indicates a vital function for PPAR α in allowing cardiac FA oxidation to occur at a rate sufficient to prevent the harmful accumulation of lipid within the heart itself.

Changes in the activity of the PPAR α signaling pathway to gene expression characterize many common myocardial diseases. Whether changes in cardiac PPAR α signaling are adaptive or causally related to myocardial pathology remains controversial, but there is no doubt that poor lipid management can cause a broken heart! Cardiac PPAR α expression and activity are suppressed in the pathologically hypertrophied heart. The resultant decrease in the capacity for FA oxidation is associated with increased rates of glucose utilization and reversion to the fetal pattern of substrate use. Although causality between the reciprocal changes in lipid and glucose utilization has not been established, this metabolic switch may serve to preserve ventricular function within the context of pressure overload. The expression and activity of both PPAR α and RXR are also suppressed in the hypoxic cardiac myocyte. To date, there are no reports of worsening cardiac function in humans with cardiac hypertrophy through the use of the PPAR α agonist fibrates. It is interesting that the heart is closely associated with adipose tissue in adult man, and epicardial and pericardial depots can represent a significant fraction of the total mass of the heart, suggesting a possible role as a local FA source. Both neutral lipid and FA adversely affect myofibrillar function. An intriguing possibility is that cardiac adipose tissue may act as a potential local sink for FA, mopping up FA when they are delivered in excess of the cardiac ATP requirement or when

they cannot be cleared through oxidation, for example, during cardiac hypertrophy.

Epidemiological studies reveal that people with diabetes are at an extraordinary high risk for the development of cardiovascular disease. The heart in uncontrolled diabetes is constrained from switching to glucose oxidation and relies almost exclusively on lipid as an energy source. Cardiac-specific overexpression of PPAR α , in the absence of alterations in lipid-fuel delivery, enhances the myocardial capacity for FA oxidation. This, however, is maladaptive, because many of the metabolic abnormalities of the diabetic heart, including decreased glucose utilization and ventricular dysfunction, are recapitulated. In this situation, the profile of myocardial TAG species is similar to the lipid species in the circulation, suggesting that the expanded lipid reservoir reflects increased uptake incompletely balanced by a corresponding increase in oxidation. When mice with cardiac-specific PPAR α overexpression are made diabetic, they develop a more severe cardiomyopathy than wild-type controls. By contrast, PPAR α null mice are protected from this cardiomyopathy.

Peroxisome Proliferator-Activated Receptor Gamma

PPAR γ is the predominant molecular target for the insulin-sensitizing TZDs and for specific prostanoids, such as 15-deoxy- $\Delta^{12,14}$ prostaglandin J2. It is abundantly expressed in fat cells (adipocytes), macrophages, the placenta, and the fetal heart. The role of PPAR γ as a critical regulator of adipocyte development and fat storage is perhaps most well established of the functions of the PPARs. Adipocytes expand as animals fatten, then shrink when the stored TAGs within them are mobilized because of food scarcity or, in the case of Mrs Sprat or the Atkins diet, lack of carbohydrate. A key feature of the fat cell is its ability to release the FA derived from TAG into the bloodstream: most other tissues are less altruistic and use FA derived from their endogenous TAG. Detectable changes in gene expression in response to PPAR γ activation in the fat cells of adipose tissue include the induction of genes that allow them to take up and accumulate lipid (Figure 2). Through this action, PPAR γ contributes to lowering of circulating FA and TAG levels and, by sequestering lipid in the adipocyte, reduces the burden of excessive lipid delivery to tissues other than adipose tissue, including heart, skeletal muscle, kidney, and liver, where its accumulation could have deleterious effects. PPAR γ can, therefore, enhance whole-body insulin action even though it is predominantly expressed in adipose tissue (which makes a relatively minor contribution to insulin-stimulated glucose disposal in lean individuals) by limiting the supply of lipid to other insulin-sensitive tissues and reducing the ability of FA to oppose insulin's actions in these tissues.

In addition to effects on genes involved in adipocyte lipid storage, TZDs increase adipocyte expression of the insulin-sensitive glucose transporter GLUT4, which allows increased use of circulating glucose for the synthesis of glycerol-phosphate which, in turn, allows FA esterification to form storage TAG (Figure 2). Moreover, the TZDs may actually induce the fat cell to express enzymes that promote glycerol released from TAG breakdown (lipolysis) to be reclaimed or synthesized from circulating lactate, a process termed glyceroneogenesis

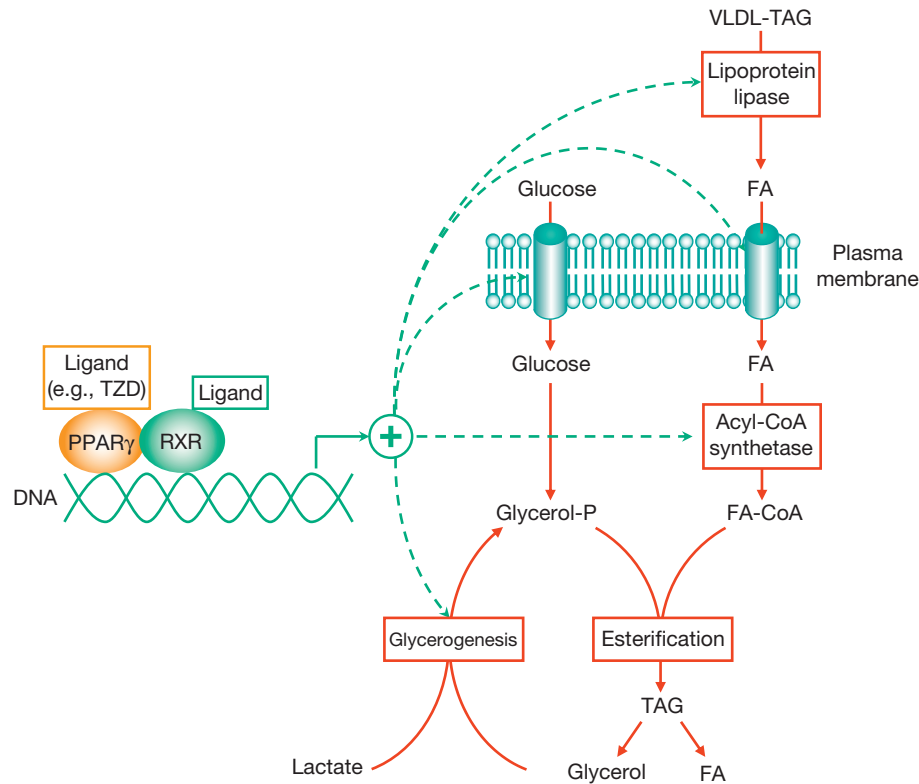


Figure 2 Mechanisms whereby activation of PPAR γ regulates fatty acid uptake, sequestration and storage, and glucose uptake and glycerogenesis in adipocytes.

(Figure 2). The presence of glycerol kinase allows glycerol released via the hydrolysis of stored TAG to be converted to glycerol phosphate which, in turn, allows reincorporation of released FA, or of FA derived from exogenous lipoproteins, into storage TAG, albeit at an energetic cost to the adipocyte. However, in obese people, as much as half of the body mass can be adipose tissue due to excessive enlargement of adipocytes and the presence of increased numbers of adipocytes. Obese people invariably become glucose intolerant and insulin resistant, as well as developing other obesity-related complications that increase morbidity and mortality. “We fat all creatures else to fat us, and we fat ourselves for maggots” (Hamlet). It follows that the long-term use of TZDs could actually exacerbate obesity and insulin resistance, despite its more acute beneficial insulin-sensitizing action due to adipose-tissue lipid sequestration.

An additional action of PPAR γ activation in adipose tissue is altered rates of secretion of biologically active adipokines (signaling molecules secreted by adipocytes), including leptin and tumor necrosis factor- α (TNF α), which can modulate insulin sensitivity in nonadipose tissue. Both leptin and TNF α have a role in obesity. Samples of adipocytes from obese people produce much more TNF α than those from lean people. Leptin is a molecule that informs the brain that adipose tissue is replete with lipid and signals a need to cease eating. PPAR γ suppresses the production of both leptin and TNF α by human adipose tissue, thereby fooling the adipocyte that enough (lipid) is not enough, and allowing the adipocyte to continue to accumulate even more lipid.

Effects of the TZDs to ameliorate hyperglycemia and hyperlipidemia in fatless mice demonstrate that, although

predominantly expressed in adipose tissue, PPAR γ activators may exert their effects via PPAR γ in tissues in addition to adipose tissue. PPAR γ activators have been reported to be cardioprotective against ischemic insult and to modify the cardiac hypertrophic growth response to pressure overload. However, most cardiac effects are likely to be indirect because of the low level of expression of PPAR γ in cardiac myocytes themselves in adulthood, although the adipose tissue surrounding the heart itself may be targeted directly and act as a lipid buffer. The beneficial versus adverse effects of PPAR γ activators, however, remain controversial.

Most kinds of mammalian cells synthesize adequate cholesterol for their needs, but the human diet usually contains more than enough to supply our cells. Only the liver can degrade and eliminate cholesterol, so any excess has to be transported there from other tissues. LDL particles are rich in cholesterol. Some LDL leak through the thin lining of major vessels, where they become trapped under the endothelial cells, a process facilitated by hypertension (high blood pressure). Trapped LDL molecules attract macrophages, immune cells that can take up and degrade LDL. Some trapped LDL react with oxygen in the blood. Unfortunately, the oxidized LDL molecules appear to be able to enter the macrophages in almost unlimited quantities, accumulating as pale, greasy particles that resemble the appearance of foam. Death of the lipid-bloated macrophages (foam cells) releases their lipid content forming fatty streaks in the arteries. The endothelial cells covering the fatty streaks can become damaged and are replaced with fibrous plaques, narrowing the arteries. In the final stages of cardiovascular disease, the stiff narrow arteries ulcerate, increasing the risk of forming a

blood clot, which will plug blood flow completely. The increased risk of cardiovascular disease in diabetic patients has prompted study of PPAR γ function in lipid-laden macrophages. It has been reported that activation of PPAR γ with the TZDs promotes cholesterol efflux from macrophages, improving the status of the atherosclerotic lesion, whereas macrophages lacking PPAR γ are defective in cholesterol efflux and display accelerated lesion progression. Activation of PPAR γ by components of oxidized LDL (9-hydroxyoctadecadienoic acid, 13-hydroxyoctadecadienoic acid), however, enhances macrophage lipid accumulation through induction of a scavenger receptor/transporter (FA translocase/CD36).

Peroxisome Proliferator-Activated Receptor Delta

Compared with PPAR α and PPAR γ , relatively little is known about PPAR δ (also called PPAR β or FA-activated receptor), except that it is relatively ubiquitously expressed. Nevertheless, the importance of PPAR δ is established by the observation that most PPAR δ -deficient mice die early in embryonic development due to a placental defect and those few that do survive have a reduced fat mass. PPAR δ may also be involved in the cellular differentiation and inflammatory response of the epidermis. Importantly, PPAR δ has been shown to mediate VLDL signaling to gene expression in the roving macrophages. Exposure of macrophages to VLDL results in enhanced expression of a lipid-droplet-coating protein that is implicated in lipid storage in these cells. Studies of PPAR δ -deficient macrophages reveal that the TAG components of VLDL, released by LPL, serve as direct ligands for PPAR δ to induce this protein, with accompanying lipid accumulation within the macrophage. These new data reveal a pathway whereby VLDL itself can directly regulate gene expression in atherosclerotic lesions.

Pharmacological activation of PPAR δ lowers fasting TAG and LDL levels while increasing levels of HDL (the cardioprotective lipoprotein). It is unclear which tissue is targeted although skeletal muscle, in which PPAR δ is highly expressed, is suspected. Slow-twitch oxidative skeletal muscles (e.g., the postural muscles and muscles used during endurance exercise) are a major site of FA catabolism, particularly during starvation and long-term exercise, when circulating lipid delivery increases. However, most evidence suggests that signaling via PPAR α impacts predominantly on the regulation of lipid gene expression and fat burning in fast-twitch skeletal muscles, recruited during explosive exercise such as sprinting, even though fast-twitch muscles do not normally oxidize FA as avidly as slow-oxidative muscles. Whereas PPAR α ablation in the mouse results in abnormally high accumulation of neutral lipid in heart and liver in response to interventions increasing FA delivery, such as starvation, starvation of PPAR α -deficient mice leads to only minor abnormalities of skeletal-muscle FA metabolism. The lack of accumulation of neutral lipid has been attributed to the findings that PPAR δ is the major type of PPAR found in mouse skeletal muscle. PPAR δ may, therefore, assume greater importance in the modulation of lipid metabolism than PPAR α in skeletal muscle, a conclusion supported by findings that PPAR δ messenger RNA expression levels in muscle are dramatically increased on fasting, under conditions where FA oxidation by skeletal muscle is increased.

Interestingly, PPAR δ is also relatively abundant in the heart and exposure of cardiac myocytes to either PPAR α - or PPAR δ -specific agonists leads to significant induction of PPAR target genes involved in FA uptake and oxidation, and both PPARs can be activated by FAs within this system.

Current and Potential Therapeutic Roles for PPAR Activators

For some time, PPAR activators have been used for the treatment of dyslipidemia, particularly when the patients are likely to develop heart disease. PPAR α also allows protection at the vascular atherosclerotic site; this is particularly efficacious in patients with dyslipidemia due to T2DM. As a result of the many benefits of PPAR γ activation with respect to maintenance of glucose and lipid homeostasis described above, TZDs are currently licensed for use in the management of T2DM. The first TZD to be approved, troglitazone, is no longer in use due to liver toxic effects. However, two other TZDs, pioglitazone and rosiglitazone, have been shown to lower blood glucose and are currently approved as antihyperglycemic agents. However, recent controversy has emerged regarding the safety of the TZD class of drugs, particularly rosiglitazone. Side effects of TZDs include worsening of heart failure, fluid retention, and weight gain, and so treatment using this drug is closely monitored. Indeed, TZDs are contraindicated in patients with certain types of heart failure. Moreover, a recent study comparing clinical effects of TZDs reported significantly increased relative risks of a heart attack or cardiovascular death with rosiglitazone. This revelation led to the precautionary removal of rosiglitazone from use in the Veterans Affairs Diabetes Trial (VADT), a large-scale clinical trial of antihyperglycemic therapies. Despite this, the final analysis of the VADT and two other similar trials, Action to Control Cardiovascular Risk in Diabetes (ACCORD) and Action in Diabetes and Vascular Disease (ADVANCE), seemed to vindicate the use of rosiglitazone demonstrating no adverse effects and, in some cases, lower rates of cardiovascular events. Nevertheless, the controversy surrounding the safety of TZDs as therapies for T2DM seems likely to continue, particularly given recent changes to licensing laws requiring that T2DM therapies demonstrate long-term cardiovascular benefit as well together with blood glucose-lowering effects.

In light of the difficulties surrounding PPAR γ agonists, glitazars, which activate both PPAR γ and PPAR α , are currently under investigation as T2DM therapies. These agonists aim to utilize the insulin-sensitizing properties of PPAR γ and the lipid-lowering properties of PPAR α . Two of these compounds – muraglitazar and tesaglitazar – reached phase III clinical trials before being rejected due to safety concerns. More recently, another dual PPAR agonist, aleglitazar has been reported as an effective therapy for T2DM and, importantly, has demonstrated an improved safety profile compared to TZDs.

Last Words

The question is raised as to why some nonadipose tissues express multiple PPARs, for example, PPAR γ and PPAR δ in macrophages and PPAR α and PPAR δ in muscle. Activation of all three PPARs appears to decrease systemic lipid availability

and to decrease lipid (TAG and/or cholesterol) storage in non-adipose tissue. One possibility is that each individual PPAR activates overlapping but distinct downstream metabolic pathways. Perhaps more likely, it may be that the physiological ligands (presumably upstream lipid-based messengers) for activation of each of the PPARs are unique and that such ligand distinctness allows specificity of cellular downstream metabolic responses. PPARs are, thus, jacks of all trades, and masters of (at least) one! Although their individual specific physiological roles still require much further investigation, and there is still much to be learned, it is clear that, as ligand-activated transcription factors, the PPARs represent attractive targets for the development of selective therapeutic agents for pharmacological modulation of lipid metabolism. Treating dyslipidemia and T2DM and their extensive ranges of associated complications places a massive strain on the health budget, both in Western societies and now even in the developing world where the incidence of these diseases is increasing.

See also: [Cell Architecture and Function: Peroxisomes](#); [Lipids](#); [Carbohydrates](#) [Membranes and Membrane Proteins](#): [Lipases](#).

Further Reading

- Finck BN and Kelly DP (2008) Peroxisome proliferator-activated receptor alpha (PPARalpha) signaling in the gene regulatory control of energy metabolism in the normal and diseased heart. *Journal of Molecular and Cellular Cardiology* 34: 1249–1257.
- Francis GA, Annicotte J-E, and Auwerx J (2003) PPAR alpha effects on the heart and other vascular tissues. *American Journal of Physiology – Heart and Circulatory Physiology* 285: H1–H9.
- Fruchart JC, Staels B, and Duriez P (2001) PPARs, metabolic disease and atherosclerosis. *Pharmacological Research* 44: 345–352.
- Kersten S, Desvergne B, and Wahli W (2000) Roles of PPARs in health and disease. *Nature* 405: 421–424.
- Lee CH, Olson P, and Evans RM (2003) Lipid metabolism, metabolic diseases, and peroxisome proliferator-activated receptors. *Endocrinology* 144: 2201–2207.
- Purushotham A, Schug TT, Xu Q, et al. (2009) Hepatocyte-specific deletion of SIRT1 alters fatty acid metabolism and results in hepatic steatosis and inflammation. *Cell Metabolism* 9: 327–338.
- Reue K and Brindley DN (2008) Thematic review series: Glycerolipids. Multiple roles for lipins/phosphatidate phosphatase enzymes in lipid metabolism. *Journal of Lipid Research* 49: 2493–2503.
- Rodgers JT, Lerin C, Gerhart-Hines Z, and Puigserver P (2008) Metabolic adaptations through the PGC-1 alpha and SIRT1 pathways. *FEBS Letters* 582: 46–53.

Peroxisomes

S Joshi and S Subramani, University of California San Diego, La Jolla, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Suresh Subramani, volume 3, pp. 246–250, © 2004, Elsevier Inc.

Glossary

Autophagy Degradation of cytosol and organelles by protein turnover involving the yeast vacuole or the lysosome in mammals.

Biogenesis Process of assembly.

Microbodies Another name for peroxisomes and similar organelles (glyoxysomes and glycosomes).

Organelle A subcellular, membrane-enclosed compartment performing specialized functions.

Peroxisomal matrix Lumen of the peroxisome.

Peroxisome A subcellular organelle involved in many lipid metabolic pathways.

Pexophagy Peroxisome turnover by selective autophagic degradation.

Ubiquitination A highly conserved posttranslational protein modification involving covalent attachment of a 76 amino acid ubiquitin (Ub) polypeptide to the substrate.

Vacuole or lysosome Organelle in which protein turnover and recycling occurs. The organelle is called the vacuole in yeast and the lysosome in mammalian cells.

Peroxisome-Like Organelles

Peroxisomes are related to specialized organelles called glycosomes in parasites such as Trypanosomes, and to plant glyoxysomes, but are unrelated to hydrogenosomes, mitochondria, and chloroplasts. Collectively, peroxisomes, glyoxysomes, and glycosomes are also referred to as microbodies. The Woronin body (named after Russian biologist Mikhail Woronin) is a peroxisome-derived microbody found near the hyphal-dividing septum in filamentous Ascomycete fungi such as *Neurospora crassa*. This organelle with a dense core is involved in plugging of the postwound septal pores restricting cytoplasmic loss at the injury sites.

Peroxisome Origin and Distribution

Peroxisomes exist in all eukaryotes, from single- and multicellular microorganisms, to plants and animals. Unlike mitochondria, nuclei, and chloroplasts, peroxisomes have no DNA. Consequently, all their proteins are encoded by nuclear genes. Christian de Duve proposed that peroxisomes may have been the first endosymbionts (allowing the cells to withstand rising free molecular oxygen in the Earth's atmosphere) which subsequently lost their DNA. However, the evidence for an endosymbiont origin is much weaker than it is for mitochondria or chloroplasts and it now appears that they may have been formed *de novo* from the endoplasmic reticulum (ER).

Like other organelles, peroxisomes divide and segregate to distribute themselves within cells and to provide daughter cells with peroxisomes. Proper peroxisome distribution requires organelle movement on cytoskeletal (actin in plants and fungi and microtubules in mammals) filaments. Unlike nuclei and the Golgi apparatus, but similar to mitochondria or chloroplasts, peroxisome division is uncoupled from the cell cycle.

Functions of Peroxisomes

The principal function of peroxisomes is to house many metabolic pathways that are involved in various aspects of lipid metabolism. These include enzymes involved in the degradative oxidation (e.g., β -oxidation of very long-chain fatty acids, 2-methyl-branched fatty acids, dicarboxylic acids, leukotrienes, bile acid intermediates and cholesterol side chains, and both α - and β -oxidation of 3-methyl-branched-chain fatty acids); the early steps in the synthesis of either glycerolipids or plasmalogens; the formation of bile acids, dolichol, and cholesterol; the catabolism of purines, polyamines, and amino acids; and the detoxification of reactive oxygen species such as hydrogen peroxide, superoxide anions, and epoxides. In methylotrophic yeasts, peroxisomes are also involved in the metabolism of methanol and methyl amines. Glycosomes contain the glycolytic enzymes, in addition to enzymes common to most peroxisomes, whereas plant glyoxysomes have some or all of the glyoxylate pathway enzymes. Peroxisomes in the leaves of plants also participate in photorespiration.

Despite the fragility of the organelle during biochemical purification, the peroxisome membrane is impermeable to small molecules such as NAD(H), NADP(H), acetyl-CoA, and even protons *in vivo*. Consequently, it is not surprising that the peroxisomal membrane has a number of transporters for fatty acids, fatty-acyl-CoA esters, metabolites, and adenosine triphosphate (ATP).

Human Peroxisome-Related Disorders

Currently, there are about 20 peroxisomal disorders known, many of which are fatal. These diseases affect either a single metabolic enzyme or the assembly of the organelle itself (Table 1). Almost all of the genes involved in the human

Table 1 Human peroxisomal disorders involving metabolism and biogenesis

<i>Disease</i>	<i>Peroxisomal enzyme affected</i>
<i>Peroxisomal metabolic disorders</i>	
Pseudoneonatal adrenoleukodystrophy	Acyl-CoA oxidase (Acox1)
Multifunctional protein 2 (MFP2) deficiency	MFP2 involved in β -oxidation of very long chain and 2-methyl-branched fatty acids
Peroxisomal thiolase deficiency	3-Ketoacyl-CoA thiolase
X-linked adrenoleukodystrophy	ALDP (transporter)
Rhizomelic chondrodysplasia punctata Type 2	Dihydroxyacetonephosphate acyltransferase
Rhizomelic chondrodysplasia punctata Type 3	Alkyl-dihydroxyacetonephosphate synthase
Refsum's disease (Classical)	Phytanoyl-CoA hydroxylase
Glutaric aciduria Type 3	Glutaryl-CoA oxidase
Hyperoxaluria Type I	Alanine:glyoxylate aminotransferase
Acatalsaeamia	Catalase
Mevalonic aciduria	Mevalonate kinase
Di/trihydroxycholestanoic acidaemia	Trihydroxycholestanoil-CoA oxidase (Acox2)
Adult-onset sensory motor neuropathy	2-Methylacyl-CoA racemase
<i>Disease</i>	<i>Peroxisomal genes affected</i>
<i>Peroxisome biogenesis disorders</i>	
Zellweger syndrome	<i>PEX1, PEX2, PEX3, PEX5, PEX6, PEX10, PEX12, PEX13, PEX16, PEX19</i>
Neonatal adrenoleukodystrophy	Several <i>PEX</i> genes (<i>PEX1, PEX5, PEX6, PEX10, PEX12, PEX13</i>)
Infantile Refsum's disease	Several <i>PEX</i> genes (<i>PEX1, PEX2, PEX5, PEX12</i>)
Rhizomelic chondrodysplasia punctata Type I	<i>PEX7</i>
Neonatal microcephaly	<i>DLP1</i>

peroxisome biogenesis disorders (PBDs) are now known – they fall within the *PEX* genes required for peroxisome biogenesis. As many *PEX* genes are similar in yeasts and humans, a better understanding of the functions of these genes in yeast can be applied to understand human peroxisome disorders. Moreover, mouse models for human PBDs also offer the promise of better insights in understanding the symptoms of these diseases, their diagnoses, and eventual therapeutic intervention.

Biogenesis of Peroxisomes

As peroxisomes lack DNA, all their proteins must be imported from genes encoded in the nucleus. Most of the proteins that reside in the peroxisome matrix and membrane are synthesized on free ribosomes in the cytosol and then imported posttranslationally to the organelle. Typically, peroxisome biogenesis involves the formation of the peroxisomal membrane followed by the import of matrix proteins leading to

further proliferation of these organelles. Around 32 *PEX* genes, encoding proteins named peroxins, along with dynamins (and related proteins) are necessary for the biogenesis of the organelle. Most of these genes are found in multiple organisms and many are conserved in humans (Table 2). Though the general principles of biogenesis appear to be common to organisms across the evolutionary spectrum, there are indeed organism-specific variations.

Peroxisomal Matrix Protein Import

Peroxisomal matrix protein import differs from other protein translocation systems in two ways. First, fully folded proteins translocate across the peroxisomal membrane, and second, it is assumed that transient assembly of peroxisomal importomer occurs in response to the signal. Proteins destined for the peroxisome matrix have one or more peroxisome-targeting signals (PTSs). Most matrix polypeptides have a conserved, C-terminal, tripeptide PTS1 (-SKL in the one letter amino acid code, or its conserved variants). Others have an N-terminal or internal sequence termed PTS2 {(R/K)(L/V/I)X₅(H/Q)(L/A)}_n. A few proteins, such as *Saccharomyces cerevisiae* acyl-CoA oxidase, either have no canonical PTS1 sequence or have one that is dispensable, suggesting that they may possess other, as yet undefined, features that allow them to be targeted to the peroxisome lumen.

Matrix proteins synthesized in the cytosol are bound by cytosolic receptors – Pex5p in the case of PTS1, and Pex7p in the case of PTS2 (see Figure 1). These receptor-cargo complexes then move to the peroxisome membrane where they dock with protein subcomplexes that are in or on the membrane. Two such complexes are a docking subcomplex, comprised minimally of peroxins Pex8p, Pex13p, Pex14p, and Pex17p, and a really interesting new gene (RING)-protein subcomplex, consisting of three RING proteins Pex2p, Pex10p, and Pex12p (and other yeast peroxins, such as Pex3p and Pex8p). The RING proteins have a characteristic zinc-binding domain and are members of a protein family whose first member was called RING. There is evidence that the protein subcomplexes are dynamic (e.g., the docking and RING-protein subcomplexes can come together as a larger complex during import). The PTS receptors, Pex5p and Pex7p, are believed to shuttle from the cytosol to the peroxisome membrane or lumen of the organelle, where they release their cargo, before they return back to the cytosol for another round of import. This is referred to as the extended shuttle model for matrix protein import. The RING-protein subcomplex is not involved directly in the translocation of proteins into peroxisomes, but provides E3 ligase activity for ubiquitination of Pex5p and Pex20p during receptor/coreceptor recycling to the cytosol. Many other peroxins and chaperones, such as Djplp, Hsp70, and Hsp40, are implicated in matrix protein import, but their precise roles are still under investigation.

The PTS2 pathway, which is dependent on the receptor, Pex7p, requires different additional proteins depending on the organism of its origin. In *S. cerevisiae*, the redundant proteins, Pex18p and Pex21p fulfill this function, whereas in *Yarrowia lipolytica*, *Neurospora crassa*, and *Pichia pastoris*, Pex20p, the PTS2 coreceptor, is needed, and in mammals an alternatively spliced

Table 2 Proteins involved in peroxisome biogenesis*Proteins involved in the formation of preperoxisomal vesicles and peroxisomes from the ER*

Pex3p	A ~40 kDa integral PMP conserved from yeast to humans that binds Pex19p and is defective in CG12 of the PBDs. Needed for assembly/stability of the RING-domain subcomplex comprised of Pex2p, Pex10p, Pex12p in yeast. Traffics through the ER.
Pex16p	A ~45-kDa integral PMP found in <i>Y. lipolytica</i> that binds Pex19p and is defective in CG9 of the PBDs. Traffics through the ER. Facilitates import of some peroxisomal proteins such as catalase and thiolase. Inhibits peroxisome division. No homolog found in <i>S. cerevisiae</i> . More common in higher eukaryotes.
Pex19p	A ~33-kDa farnesylated protein of yeasts and humans. Predominantly, cytoplasmic/partly peroxisomal. Chaperone and import factor for newly synthesized, class I integral PMPs. Recognizes some, but not all, mPTSs. It is defective in CG14 of the PBDs.

Proteins involved in matrix protein import

Pex1p	A ~100–150 kDa ATPase (AAA family) conserved from yeast to humans. Interacts with Pex6p and other peroxins. Involved in PTS receptor recycling from the peroxisome membrane to the cytosol during import. Defects in Pex1p are the most common cause of the PBDs (CG1).
Pex2p	A ~40-kDa integral PMP with a carboxy-terminal, cytoplasmically-exposed, zinc-binding RING domain. Has been identified in yeasts and humans, interacts with Pex3p, Pex10p, and Pex12p. Has E3 ligase function for Ubc4-dependent polyubiquitination of Pex5p during receptor recycling. Defective in CG10 of the PBDs.
Pex4p	A ~20–24 kDa peroxisome-associated ubiquitin-conjugating enzyme that interacts with Pex22p. Has been identified in several yeast species, but not in any metazoan. Involved in mono-ubiquitination of Pex5p.
Pex5p	A ~70-kDa, predominantly cytoplasmic/partly peroxisomal protein that is conserved from yeast to humans. It is the PTS1 receptor and has a PTS1-binding domain in its carboxy-terminal half. It interacts with several peroxins (Pex7p, Pex8p, Pex10p, Pex12p, Pex13p, and Pex14p) and is defective in CG2 of the PBDs. In yeast, it is monoubiquitinated by Pex4p during receptor recycling and polyubiquitinated by Ubc4/Ubc5 in a quality-control step.
Pex6p	A ~100-kDa (AAA family) ATPase conserved from yeast to humans. Interacts with Pex1p and is involved in PTS receptor recycling from the peroxisome membrane to the cytosol during import. Defective in CG4 of the PBDs.
Pex7p	A ~40-kDa, WD40-repeat-containing protein that is the PTS2 receptor. Defective in CG11 of the PBDs.
Pex8p	A variably sized (60–80 kDa), peripheral, but intraperoxisomal protein that interacts with Pex5p and the docking subcomplex; found only in fungi. It is an intraperoxisomal organizer of the peroxisomal import machinery in <i>S. cerevisiae</i> but not in <i>P. pastoris</i> . It plays a weak role in cargo release <i>in vitro</i> .
Pex9p	A ~44-kDa integral PMP found only in the yeast <i>Y. lipolytica</i> .
Pex10p	A ~35-kDa integral PMP with a carboxy-terminal, cytoplasmically exposed, zinc-binding RING domain. Conserved from yeast to humans, interacts with Pex2p, Pex3p, Pex5p, and Pex12p, and is defective in CG7 of the PBDs.
Pex12p	A ~40-kDa integral PMP with a carboxy-terminal, cytoplasmically exposed, zinc-binding RING domain. Conserved from yeast to humans, interacts with Pex3p, Pex10p, and Pex12p and is defective in CG10 of the PBDs. Has E3 ligase function for Pex4p-dependent monoubiquitination of Pex5p during receptor recycling.
Pex13p	A ~44-kDa integral PMP with a carboxy-terminal, cytoplasmically exposed SH3 domain. Conserved from yeast to humans, interacts with Pex5p and Pex14p, and is defective in CG13 of the PBDs.
Pex14p	A ~40-kDa PMP conserved from yeast to humans that interacts with Pex5p, Pex8p, Pex13p, and Pex17p. Proposed to be part of the minimal translocon for import of PTS1 proteins, along with Pex5p.
Pex15p	A ~44-kDa integral, tail-anchored PMP identified only in <i>S. cerevisiae</i> . Interacts with Pex6p.
Pex17p	A ~25-kDa integral PMP that interacts with Pex14p. Has been identified only in fungi.
Pex18p	A ~31-kDa soluble protein involved only in PTS2-protein import. It is highly similar to Pex21p, and acts as a PTS2 co-receptor. Identified only in <i>S. cerevisiae</i> .
Pex20p	A ~46-kDa soluble protein involved only in PTS2-protein import. Identified only in some fungi that use this protein as a PTS2 coreceptor instead of Pex18p/Pex21p in <i>S. cerevisiae</i> .
Pex21p	A ~31-kDa soluble protein involved only in PTS2-protein import, is highly similar to Pex18p, and might act as a PTS2 coreceptor. Identified only in <i>S. cerevisiae</i> .
Pex22p	A ~20-kDa integral PMP of yeasts that interacts with Pex4p and anchors it on the peroxisomal membrane.
Pex23p	A ~46-kDa integral PMP. Identified only in <i>Y. lipolytica</i> .
Pex24p	A ~61-kDa integral PMP found in yeasts; required for the proper localization of some, but not all, PMPs and matrix proteins.
Pex26p	A ~34-kDa integral PMP found in higher eukaryotes. Anchors Pex1p and Pex6p to peroxisomal membrane forming heteromeric AAA ATPase complexes required for import. Defective in PBD patients (CG8).

Proteins involved in controlling peroxisome size and number

Pex11p	A ~25-kDa integral PMP required for normal peroxisome abundance and division. Many species contain several <i>PEX11</i> genes.
Pex25p	A ~45-kDa <i>S. cerevisiae</i> PMP that regulates peroxisome size and number.
Pex27p	A ~44-kDa peripheral <i>S. cerevisiae</i> PMP that regulates peroxisome size and number. Interacts with Pex25p.
Pex28p	A ~66-kDa integral PMP of yeasts that regulates peroxisome size, number, and distribution. Acts at steps upstream of Pex30p, Pex31p, and Pex32p.
Pex29p	A ~63-kDa integral PMP of yeasts that regulates peroxisome size, number, and distribution. Acts at steps upstream of Pex30p, Pex31p, and Pex32p.
Pex30p/31p/32p	Dysferlin domain containing PMPs of yeasts that regulate peroxisome size, number, and distribution. Partially redundant with each other. Acts at steps downstream of Pex28p and Pex29p.

(Continued)

Table 2 (Continued)

Vps1p/Dnm1p	Dynamin-related proteins (DRPs) and GTPases required for fission and actin cytoskeleton organization. Vps1p/Dnm1p found in yeasts while DLP1 is a homolog found in mammals. A patient has been described with a heterozygous, dominant-negative mutation in DLP1.
Mdv1p/Caf4p	Conserved, WD repeats-containing adaptor proteins required for DRP recruitment in yeasts. Interacts with Dnm1p/Vps1p and Fis1p.
Fis1p	A ~18-kDa conserved, TPR repeat-containing protein required for localization of Mdv1p/Caf4p and DRPs.
Rho1p	A ~23-kDa conserved GTPase belonging to Ras-like protein subfamily, involved in actin assembly on peroxisomes. Interacts with Pex25p and Pex30p in yeast.
<i>Proteins involved in peroxisome inheritance</i>	
Inp1p	A ~47-kDa fungal PMP involved in retention of peroxisomes in mother and daughter cells.
Inp2p	A ~81-kDa fungal PMP involved in peroxisome movement through interaction with myosin class V motor, Myo2p.
Myo2p	A ~180-kDa protein that is an actin-based motor that propels peroxisome movement.

CG, complementation group; SH3, Src-homology; PMP, peroxisomal membrane protein.

Note: Molecular weights of the proteins are those of *S. cerevisiae* unless mentioned otherwise.

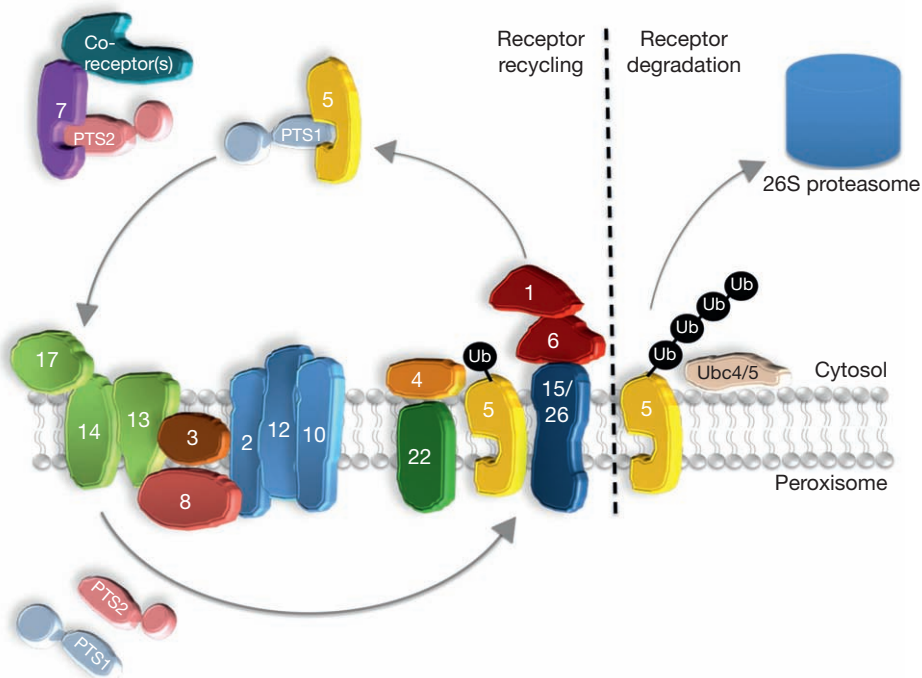


Figure 1 Overview of the peroxisomal matrix protein import cascade. Protein import occurs in four steps as outlined: (1) recognition of PTS-containing (cargo) proteins by the import receptors (Pex5p for PTS1 and Pex7p/coreceptor for PTS2) in the cytosol followed by delivery of the receptor–cargo complexes to the peroxisomal membrane; (2) translocation of the complex to the peroxisome lumen through a translocon formed by the docking (Pex13p, Pex14p, Pex17p, and Pex8p) subcomplex, which associates with the RING (Pex2p, Pex10p, and Pex12p) subcomplex peroxins; (3) delivery of the cargo by disassembly of the receptor–cargo complex to the peroxisomal lumen; and (4) receptor recycling back to the cytosol. The PTS1 receptor Pex5p, and the PTS2 coreceptor, Pex20p, (not shown) are monoubiquitinated by Pex4p or polyubiquitinated by Ubc4/Ubc5. AAA peroxins Pex1p and Pex6p (anchored to membrane by Pex15/Pex26p) dislocate monoubiquitinated Pex5p and Pex20p from the membrane and recycle them back to cytosol for further rounds of import. During this process, Pex5p and Pex20p are deubiquitinated. Polyubiquitinated Pex5p and Pex20p are degraded by the 26S proteasome. Numbers indicate the corresponding Pex proteins. Ub indicates ubiquitination.

form of Pex5 (Pex5L) is necessary for PTS2 import. However, the docking and RING-proteins are required in common for both PTS1 and PTS2 import pathways, leading to the current view that a common translocation machinery is involved for both these pathways.

Examples of organism-specific variations in the general scheme of biogenesis include the apparent lack of the entire PTS2 pathway (PTS2 proteins and Pex7p, the PTS2 receptor) in worms (*Caenorhabditis elegans*), and the dependence of the PTS2 import pathway on the PTS1 receptor, Pex5, in

mammals, but not in yeasts. However, even where such differences exist, the underlying molecular mechanism is similar. This is illustrated by the point that the proteins Pex18p and Pex21p from *S. cerevisiae*, Pex20p from *Y. lipolytica*, and Pex5L in mammals all have a conserved motif that allows them to interact with Pex7p and/or PTS2 cargo to facilitate the PTS2 import pathway.

The import receptors are released from the peroxisomal membrane back to the cytosol for another round of import at the end of receptor cycle. The PTS1 receptor (Pex5p) and the PTS2 coreceptor (Pex20p) undergo monoubiquitination by a

peroxisome-associated ubiquitin-conjugating enzyme, Pex4p, before they are recycled back to the cytosol from the peroxisome membrane. This receptor-recycling step requires Pex4p, its peroxisomal membrane anchor Pex22p, the RING sub-complex proteins as E3 ligases for PTS receptor/coreceptor ubiquitination and the AAA ATPases, Pex1p and Pex6p. In the absence of any component of the receptor recycling machinery (Pex1p, Pex4p, Pex6p, and Pex22p), Pex5p and Pex20p are polyubiquitinated and degraded by the proteasome, as part of a quality-control step.

Unlike the transport of unfolded proteins into other organelles such as the ER and mitochondria, folded and oligomeric proteins can be transported across the peroxisomal membrane. How such large multisubunit complexes are transported across the membrane is unknown, but the translocon in the peroxisomal membrane may be assembled transiently, using primarily the PTS receptors and Pex14p. Other proteins associated with Pex14p may improve the efficiency of protein translocation across the peroxisome membrane in unknown ways.

Import of Peroxisomal Membrane Proteins

In yeast, peroxisomal membrane protein (PMP) import requires the peroxisomal integral membrane protein, Pex3p, and Pex19p, a soluble factor that docks with it. Yeast mutants lacking either of the two peroxins fail to form peroxisomes. PMPs have one or more sequences (mPTSs) that direct them to the peroxisomal membrane with the correct topology. Although a dozen or so mPTSs have been defined in several yeast and mammalian PMPs, they have no simple consensus sequence. However, most of these PMPs bind Pex19p, a chaperone and import factor for PMPs. Recent findings suggest that the PMPs undergo anterograde transport via the ER to peroxisomes. Many peroxins (e.g., Pex3p, Pex16p, and Pex30p) are transported in this fashion. It has been postulated that these PMPs concentrate on ER and then migrate to peroxisome via ER-derived vesicles that then fuse with membranes of preexisting peroxisomes to allow growth.

Most mutants affecting the import of either peroxisomal matrix or membrane proteins have organelle remnants, in some but perhaps not all, organisms.

Regulation of Peroxisome Number, Volume, and Contents

In response to environmental signals, cells regulate peroxisome number, volume, and contents, allowing them to respond to various factors such as environmental or metabolic stresses, cell division, segregational imbalances, removal of damaged peroxisomes, and developmental reprogramming. Various pathways govern the control of peroxisome number and size. The first pathway is peroxisome division, in which a preexisting peroxisome divides by fission, regulating peroxisome number incrementally. The second pathway is 'peroxisome proliferation', in which the induction of peroxisomes leads to many new organelles. This pathway allows a rapid increase of organelle number as compared to division. The final pathway is 'peroxisome turnover', during which peroxisome numbers are

controlled by general (autophagic) or specific (pexophagic) processes.

Peroxisomes can be induced to proliferate in many organisms in response to metabolic needs. Examples include the induction of proliferation of hepatic peroxisomes by fibrates drugs, phthalate plasticizers, and xenobiotics, or the induction of peroxisomes in the methylotrophic yeast, *Pichia pastoris*, by methanol or oleate. The contents of the organelle are also responsive to the environment, as illustrated by the fact that peroxisomes of yeasts grown on oleate have induced levels of the fatty-acid β -oxidation enzymes, whereas methylotrophic yeasts grown on methanol have elevated levels of alcohol oxidase and dihydroxyacetone synthase. Peroxisome volume can also be regulated by proteins such as Pex11p, Pex25p, and Pex27p–32p and various dynamin-related proteins.

Excess peroxisomes can be turned over by autophagic mechanisms involving the lysosome in mammals, or its yeast equivalent, the vacuole.

Division and Proliferation of Peroxisomes

The division of peroxisomes is compatible with two models. The first model proposes that peroxisomes arise by budding and fission from preexisting peroxisomes, and may be the one that is used in normal, mitotically dividing cells. The other model is that peroxisomes arise either *de novo* or from some other reservoir of membranes such as the ER. Mature peroxisomes are then generated from this membrane reservoir via a variety of biogenesis intermediates. This model may be more applicable to the proliferating peroxisomes and to the regeneration of peroxisomes in *pex* mutants complemented by the affected gene. Peroxisome division is initiated by the signal from within. In *Yarrowia lipolytica*, the signal for the initiation of division is generated either by remodeling of peroxisomal membrane lipids or via metabolites generated by peroxisomal fatty-acid β -oxidation. Proteins involved in peroxisome division and inheritance to daughter cells (in yeast) are shown in Table 2.

Acknowledgment

This work was supported by an NIH grant DK41737.

See also: **Bioenergetics:** Mitochondrial Membranes, Structural Organization; Protein Import into Mitochondria; **Cell Architecture and Function:** Autophagy in Fungi and Mammals; **Metabolism** **Vitamins and Hormones:** Cholesterol Synthesis and Regulation; Fatty Acid Oxidation; Fatty Acid Structure and Synthesis; Peroxisomal Metabolism; Photosynthesis; **Protein/Enzyme Structure Function and Degradation:** AAA-ATPases; Proteasomes, Overview; Protein Degradation; Ubiquitin System; **Signaling:** Peroxisome Proliferator-Activated Receptors.

Further Reading

Girzalsky W, Saffian D, and Erdmann R (2010) Peroxisomal protein translocation. *Biochimica et Biophysica Acta* 1803(6): 724–731 (Epub 2010 January 13).

- Guo T, Gregg C, Boukh-Viner T, et al. (2007) A signal from inside the peroxisome initiates its division by promoting the remodeling of the peroxisomal membrane. *Journal of Cell Biology* 177(2): 289–303.
- Hettema EH and Motley AM (2009) How peroxisomes multiply. *Journal of Cell Science* 122(14): 2331–2336.
- Hettema EH and Tabak HF (2000) Transport of fatty acids and metabolites across the peroxisomal membrane. *Biochimica et Biophysica Acta* 1486: 18–27.
- Ma C and Subramani S (2009) Peroxisomal matrix and membrane protein biogenesis. *IUBMB Life* 61(7): 713–722.
- Manjithaya R, Nazarko TY, Farré JC, and Subramani S (2010) Molecular mechanism and physiological role of pexophagy. *FEBS Letters* 584(7): 1367–1373 (Epub 2010 Jan 17).
- Perry RJ, Mast FD, and Rachubinski RA (2009) Endoplasmic reticulum-associated secretory proteins, Sec20p, Sec39p and Dsl1p are involved in peroxisome biogenesis. *Eukaryotic Cell* 8(6): 830–843 (Epub 2009 Apr 3).
- Rabu C, Schmid V, Schwappach B, and High S (2010) Biogenesis of tail-anchored proteins: The beginning for the end? *Journal of Cell Science* 122(20): 3605–3612.
- Schlüter A, Real-Chicharro A, Gabaldón T, Sánchez-Jiménez F, and Pujol A (2010) Peroxisome DB 2.0: An integrative view of the global peroxisome metabolome. *Nucleic Acids Research* 38 (Database issue): D800–805 (Epub 2009 Nov 5).
- Steinberg SJ, Dodt G, Raymond GV, et al. (2006) Peroxisome biogenesis disorders. *Biochimica et Biophysica Acta* 1763(12): 1733–1748 (Epub 2006 Sep 14).
- Van den Bosch H, Schutgens RBH, Wanders RJA, and Tager JM (1992) Biochemistry of peroxisomes. *Annual Review of Biochemistry* 61: 157–197.

Relevant Website

<http://www.peroxisomedb.org> – PeroxisomeDB Home.

Phage Display for Protein Binding

H B Lowman, NGM Biopharmaceuticals, Inc., San Francisco, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Degenerate codon A three-nucleotide (triplet) sequence encoding more than one amino acid, realized by synthesis of a pool of triplets having more than one nucleotide at one or more positions, or by the combination of a pre-synthesized set of triplet codons that may be used as a block for further synthesis. Degenerate codons may be represented in shorthand according to standard abbreviations, for example, reactive nitrogen species for (adenosine (A) or guanosine (G)) followed by (A, G, cytidine (C), or thymidine (T)), followed by (G or C).

Elution Process of dissociation which causes the dissociation of bound phage from an immobilized target molecule, for example, using low pH or chaotropic agents.

Enrichment Process by which peptide or protein variants with specific binding properties are amplified from a diverse library. An enrichment ratio of the number of phage eluted

from an immobilized target divided by the number eluted from a control is often used to measure enrichment.

Enzyme-linked immunosorbent assay (ELISA) An assay for binding affinity, typically in which a target molecule is coated onto a plastic plate, a second molecule (the analyte) is added, and a third molecule is used to detect binding of the analyte.

Helper phage A bacteriophage which supplies the necessary gene products for packaging of a phagemid construct and is usually deficient in packaging itself into virions.

Panning Process of selecting phage for binding to a target protein, especially when the target is coated onto an immunosorbent plastic plate.

Phagemid A DNA vector constructed from the combination of replication and other genetic elements from a plasmid with replication, packaging, and other genetic elements from the genome of a bacteriophage.

Applications of Phage Display

Phage display has found many applications in the discovery, engineering, and analysis of peptides and proteins including (1) discovery of novel peptide mimotopes and antibodies for binding to proteins and other targets, (2) optimization, affinity maturation, or mutagenic scanning of antibodies and other proteins for reagent or therapeutic use, (3) identification of specific peptide substrates for proteases and other enzymes, and (4) preparation of specific nanoparticle materials. In short, it is an effective means of producing billions of mutated protein variants and selecting from the pool those variants with specific binding properties. Typically, filamentous phage such as Fd, M13, and related phages that infect *Escherichia coli* bacteria are used as vectors for phage display of peptide or protein diversity libraries; however, other species such as T7 phage may be used.

Phage display has contributed in the protein engineering stages of development of biological therapeutics such as pegvisomant (Somavert® hGH (human growth hormone) antagonist), adalimumab (Humira® anti-tumor-necrosis-factor (TNF) α monoclonal antibody), ranibizumab (Lucentis® anti-vascular-endothelial-growth-factor (VEGF) monoclonal antibody Fab), romiplostim (Nplate® thrombopoietin-receptor agonist peptibody), and motivizumab (anti-RSV (respiratory syncytial virus) monoclonal antibody). Many more biotherapeutics appear to be in the pipeline as a result of more extensive application of the technology to antibodies and other proteins. In addition, phage particles themselves, having been selected for specific binding properties, have found application as nanoparticles in novel materials-science applications, including the production of phage-based lithium batteries.

Phage Display Vectors

Phage display is a rapid means for discovery of novel peptides and proteins from large libraries of genetically engineered variants. The process begins with construction of a phage or phagemid vector that encodes either an initial protein scaffold or a random-sequence peptide as a fusion protein with a phage coat protein. Typical constructs involve a promoter region driving expression of the fusion protein in which a secretion signal sequence (SS) is followed by the display gene and all or part of a phage coat protein, for example, g3p or g8p (proteins encoded by gene-3 or gene-8 of bacteriophage M13). Alternatively, with appropriate modifications, the displayed protein may be fused at the C-terminus of the coat protein or it may be linked to a modified phage protein through formation of a specific disulfide bond. Filamentous phage(mids) are particularly useful vectors because their structure readily adapts to accommodate the size of DNA packaged into the virion.

Polyvalent and Monovalent Phage Display

When many copies of a fusion protein are displayed on each virion, each particle may simultaneously bind to several molecules of a surface-immobilized target. In this format, called polyvalent display, even intrinsically weakly binding variants appear to bind tightly because of the avidity effect. As this effect can mask tighter binding (higher intrinsic affinity) interactions, constructs that limit the display of the fusion protein to yield virions that usually display not more than one copy were developed. This format, called 'monovalent display', facilitates stringent selections for very high-affinity interactions. Statistically, monovalent display vectors allow for display of multiple

copies of the fusion protein by a small fraction of the virion population, but the difference between monovalent and polyvalent formats can be dramatic in terms of avidity effects. The two formats have been exploited to: (1) identify novel, low-affinity binding peptides in a polyvalent system, followed by (2) affinity maturation of the low-affinity peptides through construction of secondary libraries and selection in a monovalent system.

Phage Vectors

Phage vectors consist of an essentially complete phage genome, often M13 phage, into which is inserted DNA encoding the protein or peptide of interest (Figure 1). Typically, the remainder of the phage genome is left unchanged and provides the other gene products needed for the phage life cycle. For facile growth and propagation of phage in *E. coli*, the function of the fusion protein in the life cycle of the phage must be maintained, and this places constraints on the types of fusions that are usually produced using phage vectors. For example, large proteins may be poorly displayed and/or may interfere with phage packaging or infection.

Phagemid Vectors

Phagemid vectors consist of a typical plasmid construct into which a phage origin of replication (ORI), including the phage-packaging region of DNA from the phage genome, and a gene for the displayed fusion protein have been inserted (Figure 1). In this case, the phagemid DNA can be replicated independently of phage production for mutagenesis and amplification. However, for phage display of the protein, the remaining phage gene functions must be supplied by another construct, usually a helper phage, which is packaged into virions with lower efficiency than the phagemid itself. When

E. coli cells transfected with the phagemid, grown under selective pressure with an appropriate antibiotic, are infected with helper phage at an appropriate multiplicity of infection (MOI), both constructs undergo replication, but mature virions predominantly contain phagemid rather than helper-phage DNA. However, both types of particles can display the fusion protein; thus, antibiotic selection of cells containing the desired phagemid is needed. These examples indicate the most commonly used types of phage and phagemid vectors; however, other monovalent and polyvalent displaying phage and phagemid constructs utilizing several different phage coat proteins have also been described.

Phage Libraries

A phage library may begin with a wild-type protein expressed in a phage(mid) system or with a phage(mid) vector, into which DNA-encoding completely novel peptides have been inserted.

Construction of Libraries

The generation of sequence diversity in a phage-display library is usually accomplished by typical site-directed mutagenesis techniques, such as oligo-directed single-stranded template mutagenesis, polymerase chain reaction (PCR) mutagenesis, or cassette mutagenesis using restriction enzymes and DNA ligase to cut out a segment of DNA and replace it with a synthetic piece. Highly diverse DNA sequences can be generated using standard solid-phase deoxynucleotide synthesis techniques in which individual nucleotides are randomized by adding to the growing polynucleotide chain a mixture (which may be equimolar or of a biased, specified composition) of all four nucleotides, or a subset of adenosine (A),

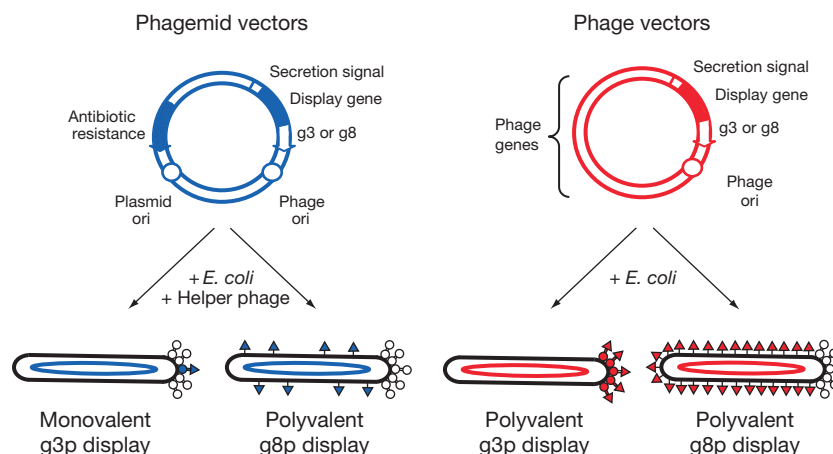


Figure 1 Construction of phagemid and phage vectors. Phagemid and phage DNA constructs are shown with appropriate origins of replication (ori), a gene for the protein of interest (shaded) fused the gene-3 or gene-8 encoded coat proteins of filamentous M13 phage and preceded by a secretion-signal sequence, as well as a gene for antibiotic resistance. The remaining genes of the phage vector or helper phage encode other proteins. The DNA is packaged as a single strand in the virion (shown in cartoon form) with many copies of the g8 protein along the length of the particle, a few copies of the g3 protein (two-lobe structure) at one end, and other structural proteins (not shown). The g3p or g8p fusion proteins are shown as filled triangles attached to the corresponding coat protein. In monovalent phage display, expression of the fusion is statistically limited so that very few particles contain more than one copy of the fusion protein, while in polyvalent display, a range of fusion-protein densities is possible.

guanosine (G), cytidine (C), or thymidine (T), rather than a single nucleotide (Table 1). The degeneracy of the genetic code makes possible the generation of chemically related subsets of codons depending on the amino acids encoded, or specific codons may be chosen using triplet codon synthesis to construct the oligonucleotides (Table 2). In some cases, the use of very small codon subsets encoding two to four amino acids have been useful for the generation of diversity libraries yielding novel binding properties or enhanced binding affinities. Once oligonucleotides with appropriate diversity have been designed and synthesized, mutagenesis is carried out with the phage or phagemid vector on an appropriate scale to yield a sufficient number of DNA molecules (moles of product) to represent the desired diversity as described below. The mutagenesis products are transformed into *E. coli* and propagated in culture (along with helper phage in the case of phagemid vectors) to yield a library of peptide- or protein-displaying phage particles.

Design of Diversity Libraries

The key consideration in the design of diversity libraries is the combinatorial complexity of the DNA and protein variants resulting from incorporation of random nucleotides at a given number of sites. For display of a complementary DNA (cDNA) library of genes derived from natural sources, the complexity of the library is determined by the complexity of the RNA source. For synthetic diversity libraries, the level of complexity can be specified according to the purpose of the experiment. For example, using neocarzinostatin naphthoate synthase (NNS) codon degeneracy (Table 2), all 20 commonly used L-amino acids are encoded by 32 codons. With one NNS codon randomized in this way, the library has a theoretical diversity of 32 at the DNA level and 20 at the protein level. However, this complexity grows quickly as a larger number of mutagenesis sites, n , are simultaneously randomized. For example, with 32 codons and $n = 6$ amino-acid positions, the theoretical DNA diversity is about 1.1×10^9 while the protein diversity is about 6.4×10^7 . Protein sequence diversity across additional sites can be increased by using 20 synthetic

triplet codons to encode the 20 amino acids; in this case, with $n = 7$ sites, both DNA and protein diversity would be about 1.3×10^9 . The experimentally obtained complexity or diversity of the library is limited to the number of DNA molecules that can be produced at the synthesis and mutagenesis stage, subsequently transformed into *E. coli*, and packaged into a solution of phage(mid) particles. Therefore, the number of randomized codons or the diversity of the degenerate codons may be restricted in order to obtain libraries that completely represent the designed diversity. At the virion level, phage particles are only soluble to a level of $\sim 10^{14}$ particles ml^{-1} (10^{17} l^{-1} ; or 170 nM), presenting an ultimate practical limitation for phage libraries to be produced and used at the laboratory scale. Additional limitations are imposed by the process of transformation of the phage or phagemid construct into *E. coli*, so that libraries are more commonly on the order of 10^9 to 10^{11} as measured by the total number of bacterial transformants obtained after mutagenesis and transformation using electroporation or other techniques.

Phage-Binding Selections

Phage libraries yield useful and interesting variants through a process of selection and amplification analogous to affinity-based protein purification (Figure 2).

Preparation of Packaged Virions

Once the products of random mutagenesis have been transformed into *E. coli*, phage amplification occurs through propagation of the infected culture, which in the case of phagemid libraries also includes a helper phage. Since filamentous phage particles are not lytic, large numbers of phage particles are produced during a few hours of culture growth. As the displayed protein is secreted, it is incorporated into the coat of the virion. Once a large number of phage have been produced, they are separated from the *E. coli* host by centrifugation and are often further purified using a series of precipitation and resuspension steps. The final phage stock is a solution of phage particles that are ready for appropriate binding selections.

Phage Binding and Elution

Early phage-binding selections used a simple process for separating variant protein phage that bind to a target molecule from those that do not bind. The target molecule was coated directly onto immunosorbent plastic plates or affixed to chemically activated resins. A variation on this technique involves binding of phage to a labeled target, such as a biotinylated protein, in solution, then capturing the bound phage through use of an affinity matrix that binds to the labeled target. Thereafter, the phage library is allowed to bind to the immobilized target and then removed. Nonbinding and weakly binding phage variants are washed away and the remaining tightly bound variants are eluted using affinity-selective reagents (e.g., a competing protein), mild to severe chaotropic agents, ranging from high/low pH buffers to urea, or site-specific proteases. Following elution, phage can be amplified by infecting fresh *E. coli* cells, and the

Table 1 Nucleotide abbreviations for degenerate codon design

Abbreviation	Nucleotide
A	Adenosine
G	Guanosine
C	Cytidine
T	Thymidine
R	A, G
Y	C, T
B	C, G, T
D	A, G, T
H	A, C, T
K	G, T
M	A, C
V	A, C, G
W	A, T
N	A, G, C, T
S	G, C

Table 2 Codon degeneracy and the generation of diversity in phage-display libraries

<i>DNA codon</i>	<i>Amino acid</i>	<i>NNS</i>	<i>NYC</i>	<i>NHS</i>	<i>ARS</i>	<i>TMC</i>	<i>Tri*</i>
AAG	Lys (K)	X		X	X		
AAC	Asn (N)	X		X	X		X
AGG	Arg (R)	X			X		
AGC	Ser (S)	X			X		X
ACG	Thr (T)	X		X			
ACC	Thr (T)	X	X	X			X
ATG	Met (M)	X		X			X
ATC	Ile (I)	X	X	X			X
GAG	Glu (E)	X		X			
GAC	Asp (D)	X		X			
GGG	Gly (G)	X					
GGC	Gly (G)	X					
GCG	Ala (A)	X		X			X
GCC	Ala (A)	X	X	X			
GTG	Val (V)	X		X			
GTC	Val (V)	X	X	X			
CAG	Gln (Q)	X		X			X
CAC	His (H)	X		X			
CGG	Arg (R)	X					
CGC	Arg (R)	X					X
CCG	Pro (P)	X		X			X
CCC	Pro (P)	X	X	X			
CTG	Leu (L)	X		X			X
CTC	Leu (L)	X	X	X			
TAG	Stop**	X		X			
TAC	Tyr (Y)	X		X		X	
TGG	Trp (W)	X					X
TGC	Cys (C)	X					X
TCG	Ser (S)	X		X			
TCC	Ser (S)	X	X	X		X	
TTG	Leu (L)	X		X			
TTC	Phe (F)	X	X	X			
AAA	Lys (K)						X
AAT	Asn (N)						
AGA	Arg (R)						
AGT	Ser (S)						
ACA	Thr (T)						
ACT	Thr (T)						
ATA	Ile (I)						
ATT	Ile (I)						
GAA	Glu (E)						X
GAT	Asp (D)						X
GGA	Gly (G)						
GGT	Gly (G)						X
GCA	Ala (A)						
GCT	Ala (A)						
GTA	Val (V)						
GTT	Val (V)						X
CAA	Gln (Q)						
CAT	His (H)						X
CGA	Arg (R)						
CGT	Arg (R)						
CCA	Pro (P)						
CCT	Pro (P)						
CTA	Leu (L)						
CTT	Leu (L)						
TAA	Stop						
TAT	Tyr (Y)						X
TGA	Stop						
TGT	Cys (C)						

(Continued)

Table 2 (Continued)

DNA codon	Amino acid	NNS	NYC	NHS	ARS	TMC	Tri*
TCA	Ser (S)						
TCT	Ser (S)						
TTA	Leu (L)						
TTT	Phe (F)						X

The 64 possible triplets are shown, along with the corresponding encoded amino acids in the first two columns. The remaining columns show examples of degenerate codons (see [Table 1](#)) encoding all amino acids with 32 triplets (NNS), and other selected sets that vary in amino acid composition, including an example of a synthetic triplet codon set (*), whereby a specific set of 20 codons are used to encode 20 amino acids. Often a *supE* strain of *E. coli* is used for phage propagation, in which case the amber stop codon (**) is translated as Gln (Q).

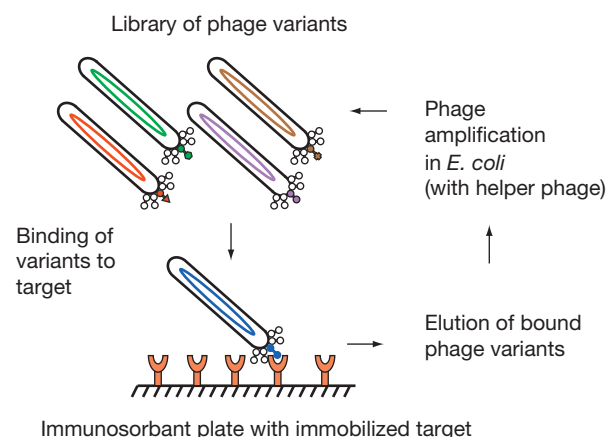


Figure 2 Affinity-selection process for binding to an immobilized target protein using a monovalent phagemid vector. One round of binding selection or panning is shown. A population of phage particles, each displaying a different peptide or protein variant, is allowed to bind to an immobilized target. After low-affinity variants are washed away, high-affinity variants are eluted and propagated to produce an enriched pool of high-affinity variants, which may be subjected to further rounds of selection.

resulting population again subjected to binding selection. This cycle is typically repeated for several rounds, followed by analysis of selectants through DNA sequencing, which may yield nonrandom occurrences of amino acids at particular positions (i.e., a consensus sequence), or through direct binding analysis as described later. As in enzyme-linked immunosorbent assay (ELISA) experiments, nonspecific binding leads to artifacts in phage display and can be reduced using proteins or detergents that block these typically low-affinity interactions. However, large diversity libraries sometimes contain peptides that can bind to the blocking agent itself. An enrichment ratio, representing the number of phage binding to the intended target versus the number binding to the matrix or blocking agent, is often used to monitor specificity during multiple rounds of binding selection.

Phage-Binding Analysis

The individual clones selected through phage display are often further analyzed in the form of the recombinant phage particles. Beyond its use in selecting relatively rare protein variants with specific binding properties from diversity libraries of

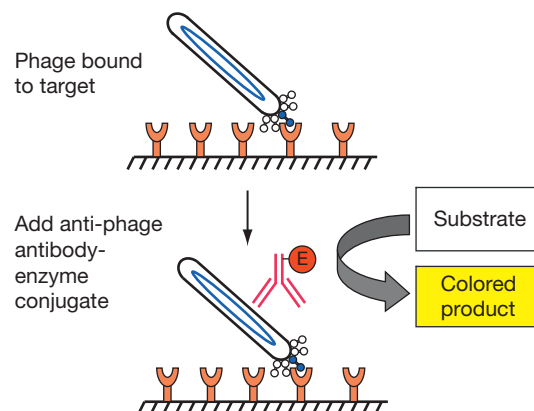


Figure 3 Binding assays using phage display. Binding of phage-displayed protein variants is analyzed as in a typical ELISA experiment, with binding of phage to an immobilized target followed by detection with an antiphage antibody conjugated to an appropriate enzyme such as horseradish peroxidase. Relative amounts of bound phage are measured by the amount of color produced when an appropriate substrate is added.

millions to tens of billions of combinatorial variants, phage display is also used to analyze protein–protein-binding interactions without the need to separately purify large numbers of protein variants. For example, to map the binding site on a phage displayed protein for a target protein, the target protein may be coated on an ELISA plate and mutants or libraries of mutants constructed for phage display. In the case of point-mutation analysis, a phage stock of each mutant is prepared and titrated as in a typical protein-binding ELISA ([Figure 3](#)). Alternatively, using libraries of protein variants, the outcome of sequencing many selected variants may be analyzed to determine statistically which amino acids are preferred or disallowed at multiple positions in the protein while maintaining the protein's function. The utility of phage display has been demonstrated using many systems in which the results of phage selections and binding analysis have been translated into functional proteins or peptides whose biochemical properties correspond closely to their phage-displayed forms.

See also: **Protein/Enzyme Structure Function and Degradation:** Biotinylation of Proteins; Protein Data Resources; Protein Folding and Assembly; Two-Hybrid Protein–Protein Interactions.

Further Reading

- Barbas CF III, Burton DR, Scott JK, and Silverman GJ (eds.) (2001) *Phage Display: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Bostrom J, Yu S-F, Kan D, et al. (2009) Variants of the antibody Herceptin that interact with HER2 and VEGF at the antigen binding site. *Science* 323: 1610–1614.
- Bradbury AR and Marks JD (2004) Antibodies from phage antibody libraries. *Journal of Immunological Methods* 290: 29–49.
- Clackson T and Lowman HB (eds.) (2004) *Phage Display: A Practical Approach*. Oxford, UK: Oxford University Press.
- Danner S and Belasco JG (2001) T7 phage display: A novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. *Proceedings of the National Academy of Sciences of the United States of America* 98: 12954–12959.
- Dower WJ and Mattheakis LC (2004) *In vitro* selection as a powerful tool for the applied evolution of proteins and peptides. *Current Opinion in Chemical Biology* 6: 390–398.
- Fellouse FA, Wiesmann C, and Sidhu SS (2004) Synthetic antibodies from a four-amino-acid code: A dominant role for tyrosine in antigen recognition. *Proceedings of the National Academy of Sciences of the United States of America* 101: 12467–12472.
- Ferrara N, Damico L, Shams N, Lowman H, and Kim R (2006) Development of ranibizumab, an anti-vascular endothelial growth factor antigen binding fragment, as therapy for neovascular age-related macular degeneration. *Retina* 26: 859–870.
- Frampton JE and Lyseng-Williamson KA (2009) Romiplostim. *Drugs* 69: 307–317.
- Kay BK, Winter J, and McCafferty J (eds.) (1996) *Phage Display of Peptides and Proteins*. San Diego, CA: Academic Press.
- Lowman HB and Wells JA (1993) Affinity maturation of human growth hormone by monovalent phage display. *Journal of Molecular Biology* 234: 564–578.
- Messing JM (1983) New M13 vectors for cloning. *Methods in Enzymology* 101: 20–78.
- Nam KT, Kim D-W, Yoo PJ, et al. (2006) Virus-enabled synthesis and assembly of nanowires for lithium ion battery electrodes. *Science* 312: 885–888.
- Osbourne J, Groves M, and Vaughn T (2005) From rodent reagents to human therapeutics using antibody guided selection. *Methods* 36: 61–68.
- Roelfsema F, Biermasz NR, Pereira AM, and Romijn JM (2006) Nanomedicines in the treatment of acromegaly: Focus on pegvisomant. *International Journal of Nanomedicine* 1: 385–398.
- Sidhu SS (ed.) (2005) *Phage Display in Biotechnology and Drug Discovery*. Boca Raton, FL: CRC Press.
- Sidhu SS, Lowman HB, Cunningham BC, and Wells JA (2000) Phage display for selection of novel binding peptides. *Methods in Enzymology* 328: 333–363.
- Smith GP (1985) Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science* 228: 1315–1317.
- Weiss GA and Penner RM (2008) The promise of phage display: Customized affinity and specificity. *Analytical Chemistry* 80: 3082–3089.
- Wu H, Pfarr DS, Johnson S, et al. (2007) Development of motavizumab, an ultrapotent antibody for the prevention of respiratory syncytial virus infection in the upper and lower respiratory tract. *Journal of Molecular Biology* 234: 564–578.
- Yáñez J, Argüello M, Osuna J, Soberón X, and Gaytán P (2004) Combinatorial codon-based amino acid substitutions. *Nucleic Acids Research* 32: e158.

Phagocytosis

G Wright, K Srinivasan, and C Janetopoulos, Vanderbilt University, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Axenic In biology, axenic describes a culture in which an organism is grown in the absence of any other contaminating organisms.

Phagolysosome A membrane-bound organelle that is formed when the phagosome fuses with a lysosome. After fusion the contents within the phagosome are digested with the enzymes contained within the lysosome.

Phosphoinositol A negatively charged phospholipid that is located within the cytoplasmic leaflet of the plasma membrane. The inositol ring can be phosphorylated on the 3-OH, 4'-OH, and 5'-OH groups.

Wortmannin A furanosteroid metabolite from the fungi *Talaromyces (Penicillium) wortmannii*, which is a covalent inhibitor of phosphoinositide 3 kinases and has been shown to also inhibit phosphoinositide 4 kinases.

Introduction

Eukaryotic cells are separated from the external environment by a semi-permeable phospholipid bilayer, called the cell, or plasma, membrane. Water and certain ions pass freely across the selectively permeable cell membrane. However, a process involving membrane invagination and subsequent engulfment and formation of internal endocytic membrane vesicles is required to internalize larger macromolecules. This process, termed 'endocytosis', involves restructuring and mobilizing large amounts of lipids and phospholipids, polymerizing actin, and redistributing membrane proteins to form and transport the endocytic membrane vesicles. There are two classes of endocytosis: phagocytosis and pinocytosis. Phagocytosis (equivalent to the cell eating) is the substrate-induced internalization of macromolecules from the external environment. Professional phagocytes, such as the social amoeba *Dictyostelium discoideum*, and mammalian macrophages, neutrophils, and dendritic cells ingest particles as small as a single molecule, and up to the size of filamentous bacteria and yeast cells. Pinocytosis (cell drinking) is the internalization of external liquid and solutes. There are two types of pinocytosis: the first type (macropinocytosis) engulfs macromolecules varying from 0.2 to 10 μm in diameter and the second (micropinocytosis) engulfs macromolecules and particles less than 0.2 μm in diameter. Micropinocytosis engulfs macromolecules through small vesicles that form from regulated curvature of the membrane and assembly of scaffolding proteins such as clathrin and caveolin. Macropinocytosis and micropinocytosis will not be covered in this article. The reader is referred to previously published reviews listed in the further reading section.

Endocytosis is an extraordinarily robust and dynamic cellular process. Amazingly, *D. discoideum* cells internalize via endocytosis an amount of membrane equal to their entire surface area every 45 min, macrophages every 30 min, and fibroblasts every 2 h. Endocytosis occurs randomly all over the cell surface during macropinocytosis and locally by substrate-induced phagocytosis. Therefore, phagocytosis normally requires a receptor to recognize the substrate to be engulfed, whereas macropinocytosis is the indiscriminate gulping of fluid into the cell and is receptor independent. A common structure involved in these processes

is a concave, cup-shaped membrane invagination that involves actin and other membrane-associated proteins on the cytosolic side of the cup, which later become the phagosome or macropinosome. This article focuses on the signaling and feedback mechanisms of the formation of the phagosome.

Phagocytosis is an essential process for the innate immune/inflammatory response as white blood cells dispose of foreign invaders. Macrophages also use this process in their garbage-collecting mission as they engulf senescent or apoptotic cells and cell debris. Phagocytosis is also essential for nonaxenic *D. discoideum* to feed and survive in nature. Much of our understanding of phagocytosis has been elucidated from these amoebae and from mammalian macrophages, neutrophils, and dendritic cells. As such, a significant portion of this article is devoted to studies performed on these cell types. Phagocytosis consists of multiple stages: (1) substrate binding to the cell surface through a transmembrane receptor or possibly attachment to cell-surface adhesion proteins, (2) activation of a signaling cascade that leads to the spatiotemporal modulation of F-actin polymerization, (3) actin-dependent pseudopod extension to facilitate the engulfment and internalization of the particle followed by vesicle trafficking from endolysosomal compartments (Figure 1), and (4) subsequent removal of the actin coat from the internalized phagosome and fusion and fission to form the mature phagolysosome. The phagolysosome is the result of the fusion of the phagosome with a lysosome in which the lysosomal enzymes are released and other lytic compounds are activated to digest the phagocytosed particle.

Receptor Recognition of Particles for Engulfment in Mammalian Systems

The first stage of phagocytosis involves receptor recognition of a particle. While the surface receptor that stimulates phagocytosis has not yet been identified in *D. discoideum*, in mammalian cells several surface receptors have been studied. In these cells, the most well-known phagocytic stimulating receptor is the Fc receptor (FcR). The FcR binds to the Fc domain of immunoglobulins such as immunoglobulin G (IgG). During FcR-mediated phagocytosis, the internalized

phagosomes form by a receptor-mediated, zipper-like, progression of the membrane and cytoskeleton over the particle. Initially, IgG coats the particle and binds the FcR on the phagocytic cell. The binding of the IgG with the FcR causes a conformational change in the receptor, which initiates a signaling cascade that activates actin polymerization and progression of the membrane over the particle. This advancement of actin and membrane over the particle results in more FcRs binding other IgG molecules. This elicits a further response, which continues the advancement of the actin and membrane over the particle. The locally modulated activation of the actin via IgG-bound FcR results in the phagosome taking on the shape of the particle they ingest.

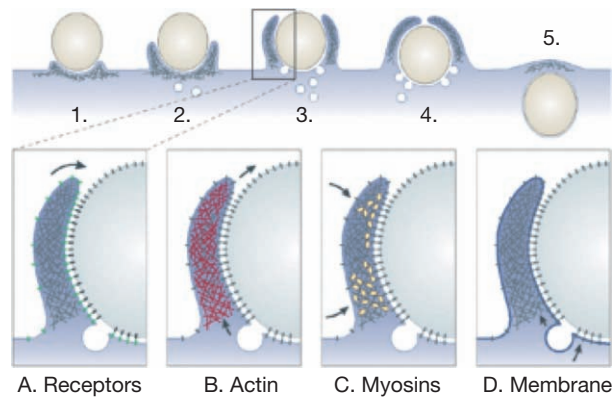


Figure 1 Phagocytic cup formation, extension and engulfment of a particle in a macrophage. The first steps (steps 1 and 2) of phagocytosis involve binding of the substrate or particle and the formation of the phagocytic cup. The extension of the phagocytic cup (step 3) involves the coordinated functions of receptors, actin, myosins, and membrane dynamics (A, B, C, and D). As the phagosome reaches closure (step 4) the particle is beginning to be internalized. Finally, the phagosome is formed and the particle is completely engulfed and internalized (step 5). At this point, the receptors and membrane are recycled and the actin and myosins dissociate from the phagosome. Reprinted by permission from Macmillan Publishers Ltd.: Swanson JA (2008) Shaping cups into phagosomes and macropinosomes. *Nature Reviews in Molecular Cell Biology* 9: 639–649, copyright 2008.

Involvement of Actin, Actin-Binding, and Actin-Regulating Proteins in Phagocytosis

Mammalian and *D. discoideum* cells enrich actin-polymerizing proteins (Wiskott–Aldrich syndrome protein (WASP), suppressor of cAMP receptor/WASP-family verprolin homologous protein (Scar/WAVE), and Arp2/3 complexes), actin cross-linking proteins (fimbrin, actin binding protein-120 (ABP120), α -actinin, etc.), actin monomer-binding proteins (profilin, cofilin, and Srv2/cyclase-associated protein (CAP)), membrane anchoring proteins (talin, comitin, ponticulin, and hisactophilin), F-actin fragmenting and capping proteins (severin and actin interacting protein-1 (Aip-1)), and contractile motor proteins (conventional and unconventional myosins) to the phagocytic cup at the onset, during, and until the particle has become totally surrounded by membrane. Interestingly, in mammalian cells, phagosomes enrich these proteins to the advancing phagocytic cup and dissipate at the base of the cup forming a belt-like structure during the progression of the phagosome (Figure 2).

In mammalian and *D. discoideum* cells, the membrane anchoring proteins such as talin are implicated in phagocytic substrate adhesion and in maintaining and modulating membrane/actin interactions throughout the cup formation and engulfment processes. Comitin also associates with endosomal vesicles and Golgi and is thought to be involved in membrane transport and turnover during phagocytosis in mammalian and *D. discoideum* cells. However, in *D. discoideum* there is no evidence that membranes from other organelles are diverted to the phagocytic cup or phagosome. It is apparent that the phagosome is predominately composed of plasma membrane lipids and proteins.

The actin cross-linking proteins Coronin and ABP-120 are involved in cross-linking multiple filaments of F-actin to form a mesh-like array. These proteins appear to play an important role in regulating the rate of phagocytosis. When either is knocked out in the *D. discoideum* cells, the rate of phagocytic cup formation and engulfment are drastically decreased.

The actin monomer-binding proteins comitin and profilin are important for actin turnover in the formation and maturation of the phagosome. Comitin positively affects phagocytosis, and profilin has an apparent negative role in phagocytosis.

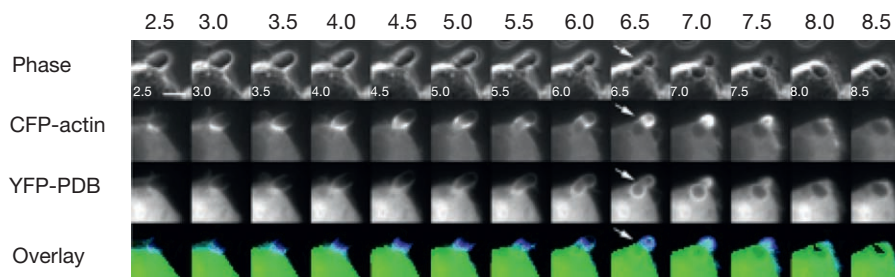


Figure 2 Actin and actin-associated proteins are enriched at the phagocytic cup and dissipate as the membrane forms around the particle in erythrocytes. Simultaneous imaging of CFP-actin and YFP-PBD (p21 binding domain of PAK1) within a macrophage during phagocytosis of an IgG-coated erythrocyte. At 2.5 min, CFP-actin is enriched at the phagocytic cup and as time progresses to 6.5 min the actin seems to form a belt-like structure around the erythrocyte (white arrow). At 8.5 min the phagosome is closed and the CFP-actin has dissipated from the phagosome. Reproduced from Hoppe AD and Swanson JA (2004) Cdc42, Rac1, and Rac2 display distinct patterns of activation during phagocytosis. *Molecular Biology of the Cell* 15: 3509–3519, with permission from American Society for Cell Biology.

The interplay between the modulatory effects of these proteins provides a sufficient pool of monomeric actin for polymerizing F-actin during phagocytosis in mammalian and *D. discoideum* cells.

In mammalian and *D. discoideum* cells the actin polymerizing proteins Scar/WAVE, WASp, and Arp2/3 are essential for forming and extending the phagocytic cup around the molecule. Scar/WAVE and WASp proteins are known to bind profilin, actin, phosphoinositides, and the Arp2/3 complex. These proteins are essential signaling components that link small guanosine triphosphate (GTP)-binding proteins to the polymerization of actin. For instance, Scar/WAVE and WASp have been shown to interact with the small GTP-binding proteins Rac1 and Cdc42, both of which remodel the actin cytoskeleton. The Arp2/3 complex, on the other hand, directly regulates the polymerization of F-actin.

Regulation of Phagocytosis by Heterotrimeric and Small GTP-Binding Proteins

In *D. discoideum*, phagocytosis is primarily modulated by large heterotrimeric GTP-binding proteins and small GTP-binding proteins. Heterotrimeric G-proteins are posttranslationally modified with lipid moieties that allow them to interact with the inner leaflet of the plasma membrane. The heterotrimer also exhibits GTPase activity and interacts with serpentine transmembrane receptors. These G-protein-coupled receptors (GPCRs) interact with the three subunits in the heterotrimeric G-protein complex: the $G\alpha$ subunit, $G\beta$ subunit, and the $G\gamma$ subunit. Stimulated receptors catalyze the exchange of guanosine diphosphate (GDP) for GTP on the $G\alpha$ subunit. During this exchange process the $G\beta\gamma$ dimer separates from the $G\alpha$ monomer, and the heterotrimer dissociates. The activated GTP-loaded $G\alpha$ monomer and $G\beta\gamma$ dimer signal to various downstream signaling and effector proteins. The GTPase activity of the $G\alpha$ subunit inactivates the protein by hydrolyzing the GTP to GDP. The inactivation can be further modulated by regulator of G-protein signaling (RGS) proteins. Upon hydrolysis, the GDP-loaded $G\alpha$ monomer re-associates with a $G\beta\gamma$ dimer and the GPCR. In mammalian cells, 16 $G\alpha$ proteins, five $G\beta$ proteins, and 14 $G\gamma$ proteins have been identified, allowing a large number of potential combinations of trimeric forms to assemble. In contrast, in *D. discoideum* there are 14 $G\alpha$ proteins, one $G\beta$ protein, and one $G\gamma$ protein. Evidence supporting the involvement of the heterotrimeric G-proteins in phagocytosis was first conducted in *D. discoideum*. Disruptions of the *D. discoideum* $G\beta$ and $G\alpha 4$ genes will result in attenuation, but not abolition, of phagocytic efficiency. Therefore, the $G\beta$ subunit and the $G\alpha 4$ subunit are necessary, but not sufficient, for phagocytosis in *D. discoideum*.

The small GTP-binding proteins primarily differ from large heterotrimeric G-proteins in that they are monomeric and are smaller in molecular weight than the $G\alpha$ subunits. These small GTP-binding proteins associate with the plasma membrane by controlled covalent modification with lipid moieties. The small GTP-binding proteins are activated by guanine nucleotide exchange factors (GEFs) and are inhibited by GAP proteins (GAP, GTPase activating protein). There are several families of small GTP-binding proteins, which include the Ras family proteins (signal transduction), Rho/Rac family proteins

(cytoskeleton), Rab/Arf family proteins (vesicular trafficking), and the Ran family proteins (nuclear import/export). The focus here is on the Ras and Rho/Rac families of small GTP-binding proteins that are involved in phagocytosis.

Ras family proteins are involved in global cellular signal transduction including changes in the actin cytoskeleton. In *D. discoideum* cells, chromosomal disruption of RasS leads to a decrease in phagocytosis and macropinocytosis suggesting a role for Ras regulation in phagocytosis. In mammalian cells, overexpression of wild-type Rap1 or of a constitutively active form of Rap1 results in an increased rate of phagocytosis. Moreover, expression of a dominant negative form of Rap1 decreased the rate of internalization of the particles. This suggests an importance of Rap1 in the rate of phagocytosis of particles.

Rho/Rac family proteins are primarily involved in the regulation of actin cytoskeletal dynamics. In particular, while Rac1, RhoA, and Cdc42 in mammalian macrophages are involved in phagocytosis, in *D. discoideum* cells it is the mammalian Rac1 homologs Rac1A–C and the mammalian RhoA homologs RacC, RacB, and RacG that are important for this process. In fact, Rac1A–C and RacB in *D. discoideum* are critical for efficient phagocytosis. The constitutively active mutant forms of these proteins greatly inhibit the rate of phagocytosis and the rate of efflux of ingested markers. This suggests that these proteins negatively regulate phagocytosis and downstream processing of the phagosome. However, overexpressing RacC and expressing a constitutively active mutant of RacG increases the rate of phagocytosis. These results suggest that these proteins positively regulate phagocytosis.

The Rab/Arf family of small GTP-binding proteins has been shown to regulate vesicle trafficking within cells. In mammalian cells, studies of macrophage phagosome maturation have identified 18 Rab proteins involved in phagosome maturation and the biogenesis of the phagolysosome. Specifically, activated Rab5a is rapidly recruited transiently to nascent phagosomes. Rab7 is also recruited to the phagosome, but recruitment of Rab5a is a prerequisite for this event. The phagosomes' acquisition of Rab7 allows the phagosomes to fuse with late endosomes and lysosomes. The phosphoinositide 3 kinase (PI3K) modulates the recruitment of both of these proteins. Active Rab11 is also localized to nascent phagosomes and is required for recycling and retrieval of phagosomal content from the plasma membrane. Rab21 and 22 transiently localize to phagosomes containing latex beads similar to Rab5. The presence of Rab14 and 22 on phagosomes seems to be important for the delay of phagolysosomal biogenesis.

In *D. discoideum* cells, the Rab7 and Rab11 GTP-binding proteins are essential for early and late phagosome maturation and with exocytosis of unutilized material. Although these proteins have less of a role in the recognition of the particle and in the formation of the phagocytic cup, they are still essential for efficient phagocytosis. The production of phosphoinositide (4,5) biphosphate (PI(4,5)P₂) is also important for Rab7 localization and phagocytosis. This suggests that Rab7 may be recruited to the maturing phagosome via PI(4,5)P₂.

Regulation of Phagocytosis by Phosphoinositides and Phosphoinositide-Modifying Enzymes

Phosphoinositides such as PI(4,5)P₂ and phosphoinositide (3,4,5) phosphate (PI(3,4,5)P₃) are potent secondary messengers

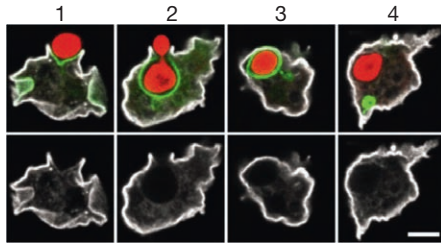


Figure 3 Localization of PH-CRAC-GFP (a $\text{PI}(3,4,5)\text{P}_3$ -binding probe) during early, middle stages of *D. discoideum* phagocytosis of an FITC-labeled yeast cell. Images of shallow (1) and deep (2) phagocytic cups and newly formed (3) and early (4) phagosomes. PH-CRAC-GFP is labeled in green, the yeast particle in red, and H36 antibody is labeled in white. H36 antibody binds to the H36 transmembrane protein on the external surface of the plasma membrane. The time course shown is 10 min and the scale bar is 4 μm . Reproduced from Mercanti V, Charette SJ, Bennett N, et al. (2006) Selective membrane exclusion in phagocytic and pinocytic cups. *Journal of Cell Science* 119: 4079–4087, with permission from Company of Biologists.

and are involved in the modulation of the actin cytoskeleton and membrane trafficking in most eukaryotic cells. $\text{PI}(3,4,5)\text{P}_3$ is important for pseudopod extension around the particle where it transiently accumulates – rapidly decreasing once the particle has been completely enclosed by the membrane. Moreover, $\text{PI}(3,4,5)\text{P}_3$ accumulations are followed by an increase in $\text{PI}(3,4)\text{P}_2$, which peaks shortly after phagosome closure and dramatically decreases during the internalization process. The phosphoinositide levels within the cell are modulated by the activities of phosphoinositide kinases, phosphoinositide phosphatases, and phospholipases.

The formation of the phagocytic cup on the inner leaflet of the plasma membrane is primarily dependent on Rho/Rac and Arp2/3 complex activation and independent of PI3Ks. PI3Ks phosphorylate $\text{PI}(4,5)\text{P}_2$ on the 3 position of the inositol ring to form $\text{PI}(3,4,5)\text{P}_3$. Although PI3K activity is not required for phagocytic cup formation, it is essential for pseudopod extension around the particle and for phagosome maturation. PI3K activity can be visualized in living cells by expressing a fluorescently tagged pleckstrin homology (PH) domain containing protein that binds to $\text{PI}(3,4,5)\text{P}_3$ (Figure 3).

During phagocytosis, phosphoinositide phosphate 5 kinase (PIP5K) is displaced from phagosomal membrane and its product $\text{PI}(4,5)\text{P}_2$ is hydrolyzed. $\text{PI}(4,5)\text{P}_2$ degradation is essential for preventing actin polymerization and its elimination is essential for phagosome closure and scission. $\text{PI}(3,4,5)\text{P}_3$ production by PI3K is essential for phospholipase C (PLC) activation and recruitment to the membrane. The PH-domain of PLC- γ , a PLC isoform in mammalian cells, binds to $\text{PI}(3,4,5)\text{P}_3$ on the membrane. The anchoring of PLC to the membrane via $\text{PI}(3,4,5)\text{P}_3$ increases the local concentration of PLC on the membrane and increases PLC activity for $\text{PI}(4,5)\text{P}_2$. PLC converts $\text{PI}(4,5)\text{P}_2$ into inositol (3,4,5) phosphate (IP_3) and diacylglycerol (DAG). PI3K inhibition by Wortmannin in macrophages leads to accumulation of $\text{PI}(4,5)\text{P}_2$ and PIP5K on the membrane in the forming phagosome and eventually inhibits phagocytosis. The accumulation of surface charge on the phagosomal membrane increases the binding of PIP5K and as the $\text{PI}(4,5)\text{P}_2$ is degraded by PLC the surface charge is decreased and PIP5K dissociates from the phagosomal membrane.

Phagocytosis in macrophages is negatively regulated by the phosphoinositide phosphatases PTEN (phosphatase tensin homolog) and SHIP-1 (Src homology 2 (SH2) domain-containing inositol-5-phosphatase 1 (SHIP-1)). PTEN removes the phosphate from the 3-position on the inositol ring and SHIP-1 removes the phosphate from the 5-position on the inositol ring. The inhibition and deletion of each of the phosphatases increase the rate of phagocytosis in macrophages. Alternatively, the deletion of PTEN in *D. discoideum* cells decreases phagocytosis and a SHIP-1-like protein has no apparent effect on phagocytosis.

PLC, which converts $\text{PI}(4,5)\text{P}_2$ into inositol (3,4,5) phosphate (IP_3) and DAG, is essential for phagocytosis. IP_3 increases the levels of intracellular calcium, which is important in regulating phagocytosis in macrophages. In mammalian cells, $\text{PI}(4,5)\text{P}_2$ increases in the extended regions of the membrane engulfing the particle and, as PLC is recruited to the phagosome, DAG is increased at the base of the phagocytic cup. Also, DAG activates protein kinase C (PKC) which regulates many enzymes involved in phagosome maturation.

Phagocytosis is a highly conserved mechanism considering the evolutionary distance of *D. discoideum* and humans. Many of the molecular components are shared and serve a wide range of biological processes. These attributes make a classic eukaryotic organism such as *D. discoideum* a valuable tool in the investigation of these mechanisms.

See also: **Bioenergetics:** Phosphatidylinositol-3-Phosphate; **Cell Architecture and Function:** Actin Assembly/Disassembly; Actin Organization; Actin-Related Proteins; Rho GTPases and Actin Cytoskeleton Dynamics; **Lipids Carbohydrates Membranes and Membrane Proteins:** Endocytosis; Membrane Fusion; **Signaling:** Immunoglobulin (Fc) Receptors; Inositol Phosphate Kinases and Phosphatases; Phosphatidylinositol Bisphosphate and Trisphosphate; Phosphoinositide 4- and 5-Kinases and Phosphatases; Rab Family; Ras Family; Small GTPases.

Further Reading

- Alberts B, Johnson A, Lewis J, et al. (2007) *Molecular Biology of the Cell*, 5th edn. New York, NY: Garland Press.
- Alderem A and Underhill DM (1999) Mechanisms of phagocytosis in macrophages. *Annual Review in Immunology* 17: 593–623.
- Bozzaro S, Bucci C, and Steinert M (2008) Phagocytosis and host–pathogen interactions in *Dictyostelium* with a look at macrophages. *International Review of Cell and Molecular Biology* 271: 253–289.
- Cardelli J (2001) Phagocytosis and macropinocytosis in *Dictyostelium*: Phosphoinositide-based processes, biochemically distinct. *Traffic* 2: 311–320.
- Fairn GD, Ogata K, Botelho RJ, et al. (2009) An electrostatic switch displaces phosphoinositide phosphatases from the membrane during phagocytosis. *Journal of Cell Biology* 187(5): 701–714.
- Greenberg S and Grinstein S (2002) Phagocytosis and innate immunity. *Current Opinion in Immunology* 14: 136–145.
- Hoppe AD and Swanson JA (2004) Cdc42, Rac1, and Rac2 display distinct patterns of activation during phagocytosis. *Molecular Biology of the Cell* 15: 3509–3519.
- Mercanti V, Charette SJ, Bennett N, et al. (2006) Selective membrane exclusion in phagocytic and pinocytic cups. *Journal of Cell Science* 119: 4079–4087.
- Pollard TD (2007) Regulation of actin filament assembly by Arp2/3 complex and formins. *Annual Review in Biophysics and Biomolecular Structure* 36: 451–477.
- Stephens L, Ellison C, and Hawkins P (2002) Roles of PI3Ks in leukocyte chemotaxis and phagocytosis. *Current Opinion in Cell Biology* 14: 202–213.
- Swanson JA (2008) Shaping cups into phagosomes and macropinosomes. *Nature Reviews in Molecular Cell Biology* 9: 639–649.

Pheromone Receptors (Yeast)

J B Konopka, Stony Brook University, Stony Brook, NY, USA

J W Thorner, University of California, Berkeley, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Agonist A ligand that activates a receptor to evoke a signal.

G-protein-coupled receptor (GPCR) Possesses a characteristic seven-transmembrane-segment primary structure with its N-terminal end facing the extracellular milieu and its C-terminal end facing the cytosol.

Ligand A substance, such as a mating pheromone, that binds to a receptor molecule.

Mating projection A term used to describe the distinctive type of polarized morphogenesis observed in yeast treated with mating pheromone. A haploid cell with this

shape is commonly referred to as a shmoo. This structure forms the conjugation bridge that connects mating cells.

Transmembrane domain A region of a protein that spans a membrane bilayer. These domains are typically present in an ordered structure, such as an α -helix or β -sheet, which allows the polar carbonyl groups of polypeptide chains to be shielded from the hydrophobic core of the membrane bilayer. In the case of the GPCRs whose crystal structures have been determined, each of the transmembrane domains forms an α -helical structure.

Pheromone Receptors Are Necessary for Conjugation

Pheromone signaling occurs only between cells of different mating type. In *Saccharomyces cerevisiae*, there are two mating types, which are defined by the combination of pheromone and receptor they produce (Figure 1). Haploid cells of the α -mating type secrete the α -factor pheromone and express in their plasma membrane, the receptor (Ste2) for α -factor; conversely, haploids of the α -mating type produce α -factor pheromone and express at their surface the receptor (Ste3) for α -factor. The α -factor is an unmodified 13-residue peptide (major isoform is WHWLQLKPGQPMY, in the one-letter amino acid code) that is processed from a larger precursor and secreted from cells via the standard secretory pathway. By contrast, α -factor is a 12-residue lipopeptide (major isoform is YIIKGLFWDPAC) that is modified by farnesylation of the side chain and methyl esterification of the carboxyl group of its C-terminal Cys residue, and then exported from cells through the action of a dedicated adenosine triphosphate-driven transporter (Ste6). Cell-type identity is established by genes at the mating type locus (MAT) on chromosome III, which encode transcriptional regulators that control expression of the genes for the corresponding pheromones and receptors. Gene products required for signaling downstream of the receptors are identical and are expressed in both MAT α and MAT α cells. The genomes of other ascomycetes (i.e., those that pack their spores into an ascus), including the fission yeast *Schizosaccharomyces pombe* and the human pathogen *Candida albicans*, encode homologs of both Ste2 and Ste3. However, in certain basidiomycetes, including the plant rust *Ustilago maydis* and the human pathogen *Cryptococcus neoformans*, the pheromone receptors in the different mating types are all related to Ste3. Furthermore, other groups of fungi, such as many basidiomycetes (including mushrooms), have more than two and often many combinatorial possibilities for distinct mating types and, hence, correspondingly more complex regulation.

Binding of pheromone to a receptor triggers a signal transduction pathway necessary for eliciting the physiological

responses required to differentiate the cells and prepare them for mating. These responses include arrest of cell division in G1 to synchronize the cell cycles of the mating partners, and transcriptional induction of genes that specify proteins required for agglutination, specialized morphogenesis, cell fusion, and nuclear fusion. These proteins and the capacities they confer permit the cells to conjoin, fuse, and form a diploid zygote.

Pheromone Receptors Stimulate a Signal Transduction Pathway

The pheromone receptors induce mating by stimulating a G-protein-initiated signaling pathway that has been most extensively studied in *S. cerevisiae* and will, thus, serve as a model for description of similar pheromone-stimulated signaling events that are likely to occur in other organisms. Upon binding of the cognate pheromone, an activated pheromone receptor stimulates exchange of guanosine triphosphate (GTP) for guanosine diphosphate (GDP) in the α -subunit (Gpa1) of the coupled heterotrimeric G protein. The ensuing conformational change in GTP-bound Gpa1 dissociates its associated G $\beta\gamma$ complex (Ste4–Ste18). Gpa1 remains membrane-associated because its mature N-terminal residue (Gly2) is N-myristoylated and Cys3 is S-palmitoylated. The free G $\beta\gamma$ complex, which is also firmly anchored in the plasma membrane by both C-terminal S-farnesylation and carboxymethylation of the C-terminal Cys and S-palmitoylation of the penultimate Cys in Ste18, now exposes surfaces that are able to bind and recruit to the plasma membrane, the proteins needed for subsequent events (Figure 2). Studies on pheromone signaling in yeast helped to first establish that the G $\beta\gamma$ subunits can transduce a signal and are not just regulators of the G α subunit.

One protein that binds to G $\beta\gamma$ is Ste5, which acts as a scaffold by associating with the three protein kinases – Ste11 (MEKK or MAPKKK), Ste7 (MEK or MAPKK), and Fus3 (ERK or MAPK) – of a mitogen-activated protein kinase (MAPK)

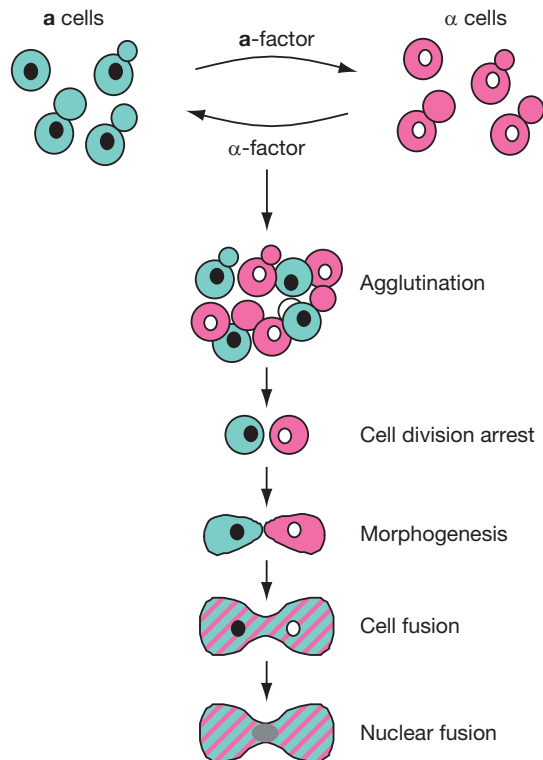


Figure 1 Landmark events during conjugation of *S. cerevisiae*. Intercellular communication between cells of mating type **a** and cells of mating type **α** induces the indicated stages of conjugation, as described in the text.

cascade. The existence of such a cascade was first demonstrated unequivocally in *S. cerevisiae* and was critical for defining highly related kinase cascades that play widespread roles in the regulation of human cells. Phosphorylation and activation of Ste11 require the action of yet another protein kinase, Ste20. Ste20 is recruited to the membrane in two ways. The activator of Ste20, the small GTPase, Cdc42 (a protein of about 21 kDa), is tethered to the membrane because it is geranylgeranylated and carboxymethylated on its C-terminal Cys and, in its GTP-bound state, binds to the N-terminal regulatory domain of Ste20. Thus, Ste20 was the first member of the family of protein kinases known as p21-activated protein kinases (PAKs) to be recognized. In addition, Ste20 is localized specifically near the rest of the pheromone response machinery at the plasma membrane because a motif near the C terminus of Ste20 binds to Gβγ. The third component recruited to the plasma membrane by the released Gβγ complex is another scaffold protein (Far1) that carries the guanine-nucleotide exchange factor (Cdc24) that stimulates conversion of Cdc42 from its GDP to its GTP-bound state. Once activated via this phosphorylation cascade, Fus3 is released from Ste5, translocates to the nucleus, and relieves negative regulation of a transcription factor (Ste12) that is poised on the promoters of pheromone-responsive genes, thereby inducing the expression of those genes. Although GTP-bound Gpa1 may also engage downstream effector(s), these seem to make only a minor contribution to pheromone response in *S. cerevisiae*, but activated Gα subunits may play more important roles in the mating process in other fungi.

Complete sequence analysis of the *S. cerevisiae* genome, accomplished in 1996, identified all of the genes that contain potential Ste12-binding sites in their promoter regions; subsequent chromatin immunoprecipitation demonstrated that many of these promoters are indeed occupied by Ste12 before induction, and with the advent of DNA microarrays, all genes actually induced by pheromone have been identified. Not surprisingly, this set of genes includes those needed for cell-cycle arrest, agglutination, morphological changes, cell fusion, and nuclear fusion. Interestingly, many signaling components, including the pheromones and receptors, as well as additional gene products that function in adaptation (downregulation) of the response are induced as part of the highly organized intercellular interplay that occurs between mating cells. Collectively, these gene products control the spatial and temporal events of the mating process, as well as recovery from those changes if conjugation does not occur successfully (Figure 2).

Pheromone Receptors Sense Spatial Gradients

In *S. cerevisiae*, the pheromone receptors have been shown to play a key role in sensing the position of the mating partner and providing proper spatial orientation for the subsequent polarized morphogenesis that enables cells of opposite mating type to grow toward each other, thereby enhancing the probability of their encounter and subsequent fusion. The ability of a yeast cell to track a gradient of pheromone emanating from a cell of the opposite mating type requires dynamic interaction between signaling and adaptation, particularly at the level of the pheromones and receptors. For example, exposure of MATα cells to a-factor increases transcription of the α-factor genes, whereas exposure of MATa cells to α-factor induces expression of an extracellular protease that degrades α-factor in the medium. Another aspect of this interplay involves removal of ligand-bound receptors by endocytosis and their replacement by newly synthesized receptors, which are deposited at the site on the surface immediately adjacent to the partner cell (because the oriented morphogenesis mentioned above directs secretory vesicles to their destination). This behavior sustains the highest degree of receptor activation in the region nearest to the site on the surface that is exposed to the highest concentration of pheromone emanating from the closest potential partner. Such localized receptor signaling is thought to establish and then reinforce the spatial cues that stimulate properly oriented actin polymerization and other aspects of the directed morphogenesis necessary for conjugation. Interestingly, the pheromone receptors in *S. cerevisiae* also appear to contribute directly to cell identity. There are residues in the N-terminal region of Ste2 that are not required for pheromone induction of signaling events *per se*, yet are needed for a MATa cell to be efficiently recognized by a MATα cell as an appropriate partner for mating and fusion.

Pheromone Receptors Are Regulated at Several Levels

Control of pheromone receptor expression and function is key for proper intercellular communication and also for allowing cells to return to normal growth in the event that mating does not occur. These controls have been best studied for the

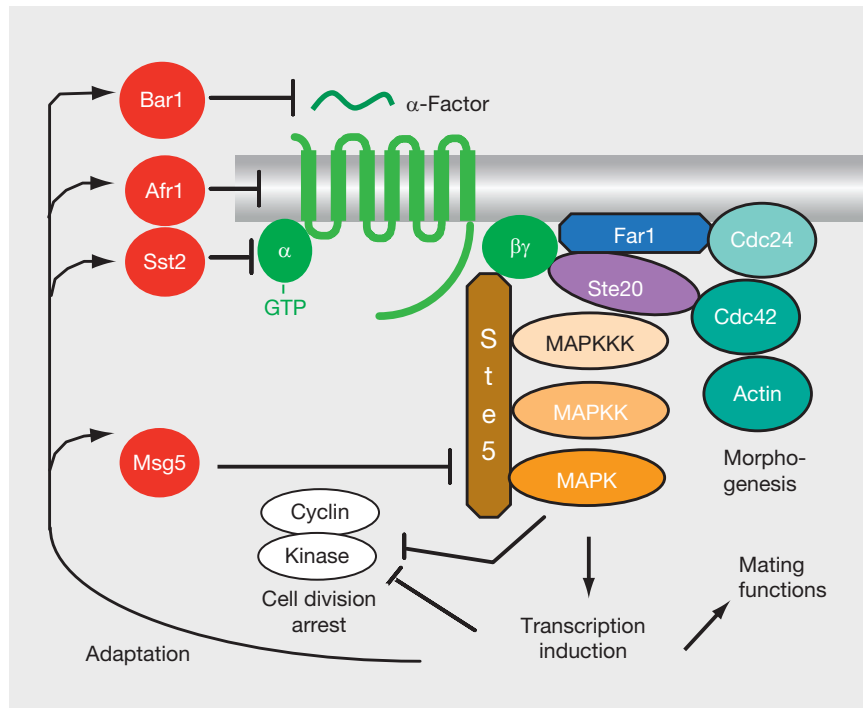


Figure 2 Schematic representation of the mating pheromone response pathway in *S. cerevisiae*. A diagram showing the main components involved in transmitting the pheromone signal in *MAT α* cells. A similar pathway operates in *MAT α* cells. Activated pheromone receptors stimulate the G α -subunit to release the G $\beta\gamma$ -subunits. The free G $\beta\gamma$ -subunits then recruit a scaffold protein, Ste5, to the membrane. Ste5 carries the components of a MAP kinase cascade and places them in proximity to another protein kinase, Ste20. The free G $\beta\gamma$ complex also recruits another scaffold protein, Far1, which carries the activator (Cdc24) of the small GTPase, Cdc42, which, in turn, is the activator of Ste20. Ste20 is further concentrated at these sites because its C terminus also binds to free G $\beta\gamma$ -subunits. Conjunction of these components achieved by the pheromone-dependent and receptor-initiated release of G $\beta\gamma$ results in activation of the protein kinase cascade, and ultimately stimulates the pheromone-responsive transcription factor, Ste12, in the nucleus. Activation of Cdc42 also promotes actin polymerization and other aspects of morphogenesis to promote highly polarized growth. Negative regulators of the pathway (shown in red on the left side) are also induced and constitute a negative feedback loop. Bar1 is a secreted protease that degrades α -factor in the medium. Afr1 negatively regulates receptor function. Sst2 is an RGS protein that promotes hydrolysis of the GTP bound to the G α -subunit (Gpa1). Msg5 is a phosphatase that dephosphorylates and inactivates the terminal MAPK (Fus3).

S. cerevisiae α -factor receptor (Ste2). The C-terminal cytosolic domain of Ste2 is the target for several adaptation mechanisms (and similar mechanisms are known to regulate the α -factor receptor Ste3). First, once ligand-bound, rapid phosphorylation of Ste2 within the C-terminal sequence appears to inhibit the ability of the receptor to continue to signal. The conformational change caused by pheromone binding makes the C-terminal segment of the receptor more accessible to certain protein kinases (casein kinase I family members) that are constitutive residents of the plasma membrane. Second, phosphorylation creates a determinant that leads to recognition by a ubiquitin ligase complex, which covalently attaches a single ubiquitin to each of several lysine residues (especially Lys337) in the C terminus (Figure 3). This monoubiquitination triggers recruitment of the receptor to clathrin-coated pits where it is endocytosed and removed from the cell surface. Third, genetic studies indicate that a pheromone-induced protein that localizes to the neck of the mating projection, Afr1, prevents Ste2 function, thereby helping to confine receptor signaling to the leading edge of the projection.

Other components of the pheromone-signaling pathway also undergo pheromone-induced downregulation. Indeed, synthesis of the most-significant negative regulator of pheromone

signaling, Sst2, is highly induced by the pheromone pathway. Sst2 was the first regulator of G-protein-signaling (RGS) protein to be recognized, making it an important model for understanding RGS function. In yeast, Sst2 promotes hydrolysis of the GTP bound to Gpa1, thereby deactivating G α and promoting its reassociation with G $\beta\gamma$. Interestingly, Sst2 functions in concert with Ste2. Sst2 binds via its Dishevelled/Egl-10/Pleckstrin (DEP) domains to the cytoplasmic C-terminal tail of Ste2, thereby tethering this regulator close to its substrate. Phosphorylation of Ste2 is then thought to release Sst2, preventing it from being endocytosed with Ste2 and degraded as described above. Additional regulators exert feedback control on receptor-mediated signaling both upstream and downstream of the receptor (Figure 2). For example, the genes encoding Bar1, a pepsin-like aspartyl protease that cleaves α -factor, and Msg5, a dual-specificity MAPK-specific phosphatase that dephosphorylates and thereby inactivates Fus3, are also induced in response to pheromone.

Pheromone Receptor Biogenesis

The most is known about synthesis of Ste2 because it has been the most extensively studied. Mature Ste2 is inserted into the

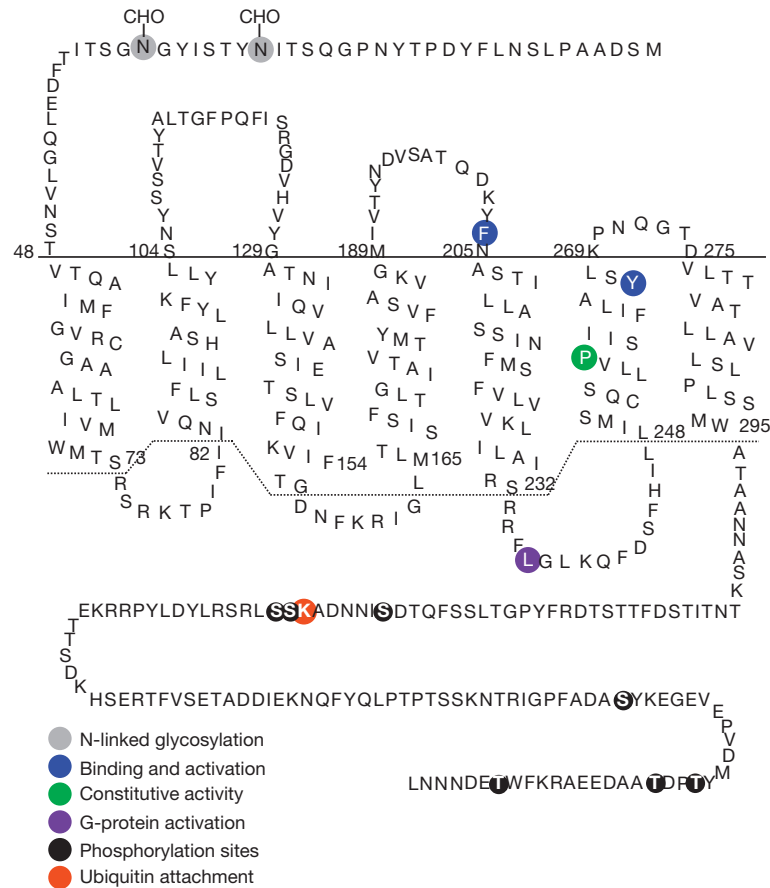


Figure 3 Primary structure of *S. cerevisiae* Ste2 (α-factor receptor). The amino acid sequence of Ste2, in the one-letter code, is arranged to represent the seven membrane-spanning elements in their presumed topology across the plasma membrane. Note that some membrane-spanning segments are predicted to be longer than others, suggesting that the longer transmembrane domains are tilted relative to the plane of the membrane, as has been observed in the structures of other GPCRs (i.e., rhodopsin). Residues discussed in the text that have key functional properties in receptor signaling and adaptation are shaded according to their role in ligand recognition and receptor activation (blue), conformational constraint (green), G-protein activation (purple), phosphorylation (black), and ubiquitination (red). Two sites in the extracellular amino terminus are modified by N-linked glycosylation (CHO) are shaded gray. The consensus site for attachment of Asn-linked oligosaccharide (N-x-T or N-x-S) at residue 25 is used close to 100% of the time, whereas that at residue 32 is used only about 50% of the time, and that at residue 46 is not used at all, presumably because it is too close to the membrane for recognition by the eight-subunit oligosaccharyltransferase enzyme complex that transfers the mannose-rich core oligosaccharide from its membrane-bound dolichol carrier to N-X-T or N-X-S motifs in secreted and membrane-localized proteins. Certain naturally occurring neutral polymorphisms exist in the primary structure of Ste2, specifically residue 269 is E in some *S. cerevisiae* strains, but K in others.

plasma membrane such that it spans the membrane 7 times (Figure 3). Biogenesis of Ste2 commences with its import into the endoplasmic reticulum (ER) via a hydrophobic, uncleaved, N-terminal signal peptide, with concomitant addition of two Asn-linked oligosaccharides, which are further modified and extended with mannose-rich outer chains during transport of Ste2 through the Golgi complex. Although N-glycosylation does not seem to be critical for receptor function, it may direct or stabilize proper folding of Ste2. While still in the secretory pathway, Ste2 self-associates to form dimers or perhaps larger oligomers. Mutational studies and cross-linking analysis indicate that TMD1 is especially important for this self-association. Oligomerization appears to be important for efficient trafficking of Ste2 to the plasma membrane and may be important for the signaling competent state of Ste2, but oligomerization does not appear to be modulated by ligand binding. Receptors that fold improperly in the ER are ultimately degraded by ER quality-control mechanisms; those misfolded receptors that make it

to the Golgi compartment get sorted to the vacuole (equivalent of the mammalian lysosome) for degradation. It is not clear where or when properly folded and oligomerized receptors first encounter and couple to the quiescent form of the $G\alpha\beta\gamma$ heterotrimer (Gpa1–Ste4–Ste18). However, some studies indicate that receptors are present in significant excess over their cognate G-protein; if so, there is a significant reservoir of receptors at the plasma membrane that are not associated with the heterotrimeric G protein in preactivation complexes primed for the arrival of their pheromone ligand.

Pheromone Receptors Bind Ligand via Contacts with the Extracellular Ends of the Transmembrane Segments

The ligand-binding site in the pheromone receptors (and in other seven-transmembrane-segment G-protein-coupled

receptors (GPCRs)) are conceivably rather complex, in that they have the potential to be composed of determinants that reside within the seven-transmembrane segments, and/or the three extracellular loops and the extracellular N-terminal extension (or combinations of these elements). The lipid modification of α -factor has prevented the application of standard methods for analyzing its binding to Ste3. Hence, the binding of α -factor to Ste2 has been examined in much greater detail. A combination of genetic and biochemical studies strongly implicates residues near the extracellular ends of the transmembrane segments (TMDs) as forming the α -factor-binding pocket. These observations suggest a model for receptor activation in which ligand binding to the ends of the TMDs could influence the relative packing or overall structural arrangement of all or part of the seven-helix bundle. The resulting conformational change could then be relayed to the intracellular side, and thereby communicated to the associated G protein to promote its release of GDP and binding of GTP. Two residues in Ste2 that are necessary for pheromone-induced receptor activation are Phe204, situated near the end of TMD5, which is required for high-affinity α -factor binding, and Tyr266, located near the end of TMD6, which, although not important for binding affinity *per se*, is nonetheless needed for subsequent signaling. The extracellular ends of TMDs 5 and 6 are, of course, directly connected to their intracellular ends, and the third intracellular loop of Ste2, which spans the intracellular ends of TMDs 5 and 6, has been implicated in G-protein activation. This situation provides a ready explanation for how conformational change upon pheromone binding could be directly coupled to the associated G protein. Current studies suggest that residues near the extracellular ends of several other TMDs in Ste2 are also involved in ligand binding, indicating that the binding pocket for α -factor is composed of a microdomain formed by the ends of multiple TMDs. Likewise, the crystal structures of antagonist-bound forms of two mammalian GPCRs (β 2-adrenergic receptor and A2A adenosine receptor) show these drugs nest within the palisade created by the exocellular ends of several TMDs.

Pheromone Receptor Activation May Involve Relief of Conformational Constraint

Current models for signaling by GPCRs suggest that, in the absence of ligand, the receptor is maintained in the off-state in a restricted conformation and that ligand binding induces conformational changes that result in greater flexibility. Insights about the conformational state of Ste2 that accompanies its activation have been provided by analysis of constitutively active mutants that signal in a ligand-independent manner. These mutations primarily alter residues within the TMDs in a manner expected to perturb α -helix packing in the membrane, suggesting that constitutive activity can result from changes in the packing of the bundle of transmembrane α -helices. One very interesting example is the high degree of constitutive receptor activity caused by substitutions that alter the conserved Pro258 residue in TMD6 of Ste2. Pro residues have the unique property of introducing a kink in a TMD because -X-Pro- bonds lack an amide hydrogen and, therefore, cannot form the main chain H-bond interaction that stabilizes

the rest of the backbone of an α -helix. Thus, substitution of Pro258 with other residues should affect the structure of TMD6. Moreover, TMD6 is the segment whose intracellular end is the most proximal to those residues in the third intracellular loop most strongly implicated in G-protein activation. Hence, the effect of substitution of Pro258 in TMD6 may reflect the conformational change induced within TMD6 that normally occurs only upon ligand binding. Interestingly, in *C. neoformans*, a pheromone receptor-like protein Cpr2 is necessary for proper morphogenesis during mating, but seems to function in a ligand-independent manner; interestingly, this protein contains Leu (instead of Pro) at the equivalent position in TMD6, perhaps explaining its constitutive-signaling activity. In this regard, it is also noteworthy that the majority of mammalian receptors of this class also contain Pro at a similar position in TMD6. Although mutation of Pro in TMD6 does not always lead to constitutive activity, it has frequently been implicated in receptor activation, suggesting mechanistic similarities in the activation process from yeast to humans. Comparison of the predicted structure of Ste2 with the well-studied photoreceptor rhodopsin has further identified other microdomains in common between the fungal pheromone receptor Ste2 and mammalian GPCRs. These comparisons also support the idea that a general underlying mechanism for activation of a wide range of receptors involves the movement of TMDs 1, 2, 3, and 4 away from the TMDs 5, 6, and 7.

Pheromone Receptors Promote G-Protein Activation via the Third Intracellular Loop

The residues on the cytoplasmic surface of pheromone receptors that contact the G protein and promote GDP-GTP exchange on the $G\alpha$ -subunit have not been pinpointed precisely. Analysis of chimeras between Ste2 and Gpa1 indicates that merely bringing the receptor and G protein into close proximity is not sufficient to stimulate GDP-GTP exchange because ligand stimulation is still required for activation of downstream signaling by such receptor- $G\alpha$ hybrids. As discussed above, current evidence suggests that pheromone binding promotes a conformational change at the extracellular ends of the transmembrane helical bundle in Ste2 that is propagated to the intracellular ends of the transmembrane segments, thereby probably exposing previously buried residues that promote G-protein activation. The identification of the critical residues has been challenging to identify by mutagenesis studies, suggesting that the interaction may involve multiple contacts between the receptor and G protein. However, several mutagenesis studies indicate that, in Ste2, the third intracellular loop certainly plays a key role. Consistent with this view, biochemical studies have demonstrated conformational changes in the third intracellular loop that are induced upon ligand binding. Remarkably, functional ligand-dependent Ste2 can be reconstituted by coexpressing halves of the receptor that are split in the middle of the third intracellular loop. Thus, it is possible that the residues closest to where TMDs 5 and 6 exit the plasma membrane are the most important for G-protein activation, assuming that the residues closest to the cut ends of such split receptors are the least likely to remain in their native condition. However, substitution of Leu236 (with His) very

near to the 'middle' (at least in terms of primary sequence) of the third loop, yields a Ste2 mutant that is strongly defective in promoting signaling yet displays an essentially wild-type level of surface expression, an essentially wild-type ability to bind α -factor, and an essentially wild-type level of association with the $G\alpha\beta\gamma$ heterotrimer, suggesting it is specifically defective in activation of the associated G protein. Interestingly, Ste2 (L236H) still undergoes ligand-stimulated endocytosis, indicating that the conformational changes and the determinants recognized by the endocytic machinery are distinct from those required for G-protein activation.

One major difficulty in understanding of the mechanism of G-protein activation is the fact that, although Ste2 and Ste3 both act through the same G protein (Gpa1–Ste4–Ste18), Ste2 and Ste3 do not share any recognizably significant tracts of sequence similarity, even at the intracellular ends of TMDs 5 and 6 or within the third intracellular loop. In this same regard, certain GPCRs from other organisms, including mammals, have been expressed in *S. cerevisiae* and, in response to the appropriate agonist, are able to activate, albeit to different extents, the endogenous yeast G protein. Again, this cross-species functionality occurs even though the heterologous receptors do not share any obvious sequence identity with the yeast pheromone receptors. Similarly, Ste2 expressed in mammalian cells is able to activate signaling in response to α -factor. Although the ability of diverse receptors from yeast to man to activate the same set of G proteins suggests a common mechanism underlying receptor and G-protein activation, the structural determinants of that common action are not readily apparent. Nevertheless,

continued functional studies of the yeast pheromone receptors are likely to continue to shed light on how this important class of receptors operates.

See also: [Protein/Enzyme Structure Function and Degradation: Ubiquitin-Like Proteins; Ubiquitin System.](#)

Further Reading

- Ballon DR, Flanary PL, Gladue DP, et al. (2006) DEP-domain-mediated regulation of GPCR signaling responses. *Cell* 126: 1079–1093.
- Bardwell L (2005) A walk-through of the yeast mating pheromone response pathway. *Peptides* 26: 339–350.
- Celic A, Connelly SM, Martin NP, and Dumont ME (2004) Intensive mutational analysis of G protein-coupled receptors in yeast. *Methods in Molecular Biology* 237: 105–120.
- Chen RE and Thorner J (2007) Function and regulation in MAPK signaling pathways: Lessons learned from the yeast *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta* 1773: 1311–1340.
- Dohlman HG and Slessareva JE (2006) Pheromone signaling pathways in yeast. *Science STKE* 364: cm6.
- Knaus M, Wiget P, Shimada Y, and Peter M (2005) Control of cell polarity in response to intra- and extracellular signals in budding yeast. *Novartis Foundation Symposium* 269: 47–54; (discussion 54–58 and 223–230).
- Madhani HD (2007) *From α to α : Yeast as a Model for Cellular Differentiation*, 115 pp. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Naider F and Becker JM (2004) The α -factor mating pheromone of *Saccharomyces cerevisiae*: A model for studying the interaction of peptide hormones and G protein-coupled receptors. *Peptides* 25: 1441–1463.

Phosphatidylinositol-3-Phosphate

J V Virbasius and M P Czech, University of Massachusetts Medical School, Worcester, MA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Autophagy Literally, 'self-eating'; a process of intracellular degradation of biomolecules and cellular components by which the constituents (e.g., amino acids) can be recycled by the cell.

Endocytosis The process by which regions of plasma membrane and any embedded proteins are brought inside the cell by formation of a membrane-bound intracellular vesicle.

Endosome The membrane-bound intracellular vesicle formed by the process of endocytosis.

Phagocytosis Uptake of particles such as cell fragments or bacteria by engulfing and then internalizing them in a membrane vesicle.

Phagosome The membrane-bound intracellular vesicle formed by the process of phagocytosis.

Phosphatidylinositol (PtdIns) A membrane lipid consisting of a pair of membrane embedded fatty acid chains linked to a surface-exposed inositol head group. The hydroxyl groups of the inositol may be phosphorylated by PI kinases at the 3', 4', and 5' positions.

Phosphatidylinositol is a polar phospholipid which comprises a small fraction of the total membrane lipids in cells. The hydroxyl moieties of the inositol head group may be reversibly modified by transfer of a phosphate group from adenosine triphosphate (ATP) to one or more of three positions. This reaction is catalyzed by the class of enzymes referred to as phosphatidylinositol kinases (PI kinases). Specific PI kinases can produce three monophosphorylated phosphatidylinositols, PtdIns(3)P, PtdIns(4)P, and PtdIns(5)P, in intact cells. These can be further phosphorylated by PI kinases to yield polyphosphoinositides such as PtdIns(4,5)P₂ and PtdIns(3,4,5)P₃, which function in regulating plasma membrane activities and cell signaling cascades, respectively. Although it is likely that PtdIns(4)P and PtdIns(5)P have important functions, molecular targets of these phosphoinositides have yet to be characterized. In contrast, through both genetic and biochemical studies our understanding of PtdIns(3)P functions and the PI 3-kinases which specifically produce PtdIns(3)P are well advanced. In recent years, a large number of proteins that interact with PtdIns(3)P have been discovered and their roles in various cellular processes, including membrane trafficking, autophagy, and signaling, have been elucidated.

Synthesis and Turnover of PtdIns(3)P

Although PtdIns(3)P synthesis was not thought to be acutely activated by extracellular stimuli as is the case for certain other phosphoinositides, PtdIns(3)P content of membranes appears to be dynamically regulated and spatially restricted in intracellular compartments. Current data indicate that PtdIns(3)P is found primarily, and perhaps exclusively, in endosomal membranes (Figure 1). These membranes are vesicular structures which delimit compartments derived from endocytic internalization of patches of membrane from the cell surface. Mechanisms for generation of PtdIns(3)P must therefore be activated concurrently with or shortly after endocytosis, while turnover of PtdIns(3)P by hydrolysis or additional phosphorylation must occur during the progression of endosomes toward

their cellular fates. These fates include recycling to the plasma membrane and fusion with lysosomes. Cellular proteins embedded in these membrane patches can likewise be targeted for recycling back to the plasma membrane where they may resume their functions, or for degradation in the lysosome. Many studies have shown that PtdIns(3)P and its molecular targets play critical regulatory roles in such membrane trafficking pathways of endosomes, and thus help to determine the cellular distribution and fate of membrane proteins such as receptors for growth factors.

Synthesis of PtdIns(3)P by PI 3-Kinases and Phosphoinositide Phosphatases

In the yeast genome, a single gene encoding a PtdIns 3-kinase has been identified. *VPS34* was originally identified in a screen for mutants that failed to properly direct a protein degradation enzyme to the vacuole, the organelle with functions similar to the lysosome of animal cells. The 100-kDa protein encoded by this gene, *Vps34p*, has a domain highly similar to mammalian PI 3-kinase and was subsequently shown to have PI 3-kinase activity. *Vps34p* exists in a complex with a 150-kDa serine/threonine kinase, *Vps15p*, which serves to target the *Vps34p* enzyme to membranes where its substrate phosphatidylinositol resides. Phosphorylation of *Vps34p* by the protein kinase activity of *Vps15p* is required for full activity of the *Vps34p* PI 3-kinase. A homologous PI 3-kinase complex exists in all eukaryotes studied. The mammalian *Vps34* PI 3-kinase (also referred to as a class III PI 3-kinase) appears to be the predominant enzyme which produces PtdIns(3)P in cells. Although class I and II PI 3-kinases are also capable of phosphorylating PI *in vitro*, their contribution to PtdIns(3)P production *in vivo* has not been well established.

PtdIns(3)P may also be produced by the action of phosphoinositide 4-phosphatases such as INPP4A and INPP4B on PtdIns(3,4)P₂, or of phosphoinositide 5-phosphatases including Sac3 on PtdIns(3,5)P₂. The Sac3 ortholog in yeast, Fig4p, appears to be regulated in concert with the yeast PtdIns(3)P

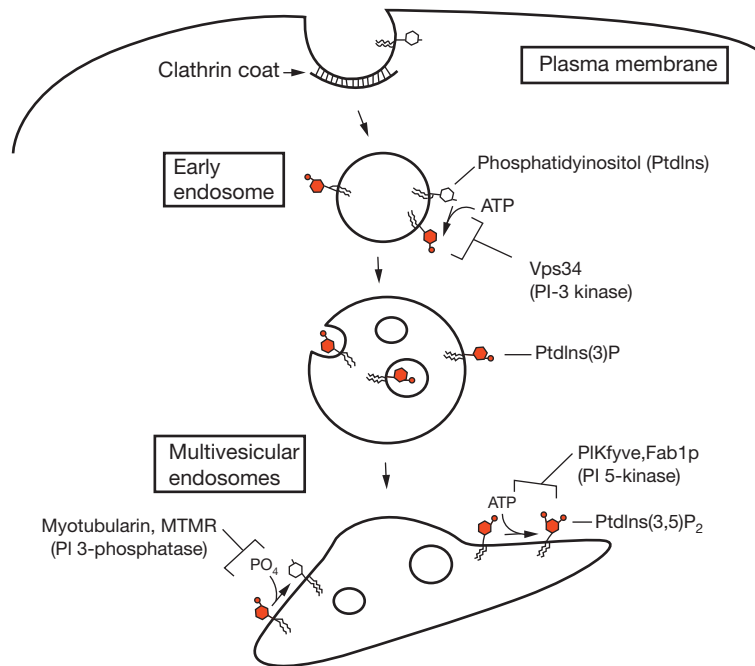


Figure 1 Production and turnover of PtdIns(3)P in endosomal membranes. In the process known as endocytosis patches of plasma membrane invaginate as a result of formation of clathrin coats. Subsequent scission from the plasma membrane and uncoating of these vesicles produce early endosomes where PtdIns(3)P is produced by the Vps34 PI-3 kinase. PtdIns(3)P also persists on the limiting membrane as well as intraluminal vesicles of the multivesicular endosome, a product of fusion of early endosomes. During subsequent maturation, the PtdIns(3)P may be hydrolyzed by the action of myotubularin family PI-phosphatases, or further phosphorylated by the PIKfyve/Fab1p PI 5-kinases to produce PtdIns(3,5)P₂. These reactions may also take place in the intraluminal vesicles.

5-kinase Fab1p (see below), suggesting that the synthesis of PtdIns(3)P and PtdIns(3,5)P₂ are in a tightly coupled cycle in the yeast endomembrane system.

Dephosphorylation of PtdIns(3)P

PtdIns(3)P may be converted back to phosphatidylinositol by the action of a 3-phosphatase, an enzyme which catalyzes the hydrolysis of the phosphate group on the 3' position of the inositol ring. A number of phosphoinositide phosphatases have been characterized in yeast and mammalian systems, but, to date, activity against PtdIns(3)P has been attributed only to members of the myotubularin family. This family consists of at least 16 genes in humans with one putative homolog in yeast. Of the human proteins, nine have functional phosphatase domains; the others have similar protein sequences but apparently no catalytic activity. These proteins also have potential phosphoinositide-binding PH domains while two have FYVE domains likely to specifically target PtdIns(3)P. Details regarding the activity of myotubularin proteins in cells have yet to be reported. However, mutations in a number of human myotubularin genes have been linked to neuromuscular diseases pointing to significant roles in the development or function of neural and muscle tissues.

Phosphorylation of PtdIns(3)P

In addition to hydrolysis of the 3-phosphate group by phosphatases, PtdIns(3)P levels may be modulated by additional

phosphorylation. It is known that significant levels of PtdIns(3,5)P₂ are present in yeast and mammalian cells. This phosphoinositide appears to be formed primarily by phosphorylation of PtdIns(3)P at the 5' position of the inositol ring. The Fab1 protein of yeast and a mammalian homolog called PIKfyve possess this enzymatic activity. Both of these proteins are also able to bind specifically to PtdIns(3)P via an FYVE domain. Conversion of PtdIns(3)P to PtdIns(3,5)P₂ not only would serve to terminate PtdIns(3)P-dependent cellular functions, but may also contribute to cellular responses to osmotic stress or other signals as well as to maturation of endosomes.

Molecular Targets and Cellular Functions of PtdIns(3)P

Protein Domains that Bind to PtdIns(3)P

Two structurally unrelated protein domains have been shown to bind to PtdIns(3)P. The first to be discovered was the FYVE domain (named for the first four proteins in which it was recognized, Fab1p, YOTB, Vac1p, and EEA1). The domain is characterized by a RING finger structure which binds two Zn molecules. It is found in five yeast proteins and 35 human proteins. Binding of FYVE domains to PtdIns(3)P-containing membranes relies on a combination of nonspecific membrane interactions and a groove which specifically accommodates the phosphate group of PtdIns(3)P while sterically excluding other phosphoinositide head groups. To date, all FYVE domains studied bind exclusively to PtdIns(3)P.

The phox homology (PX) domain was first recognized in two components of the phagocyte nicotinamide adenine

dinucleotide phosphate (NADPH) oxidase (phox) complex (p40phox and p47phox). It is structurally unrelated to the FYVE domain except for a somewhat similar binding pocket which accommodates the inositol 3-phosphate head group. However, comparison of the crystal structures of the p40phox PX domain bound to dibutanoyl PtdIns(3)P and the FYVE domain of early endosome antigen-1 (EEA1) bound to the head group inositol 1,3-bisphosphate indicates that both proteins utilize basic residues to bind both phosphate and hydroxyl groups in the lipid head group. All 15 PX domains in yeast bind PtdIns(3)P with varying affinities. Of the mammalian proteins characterized, many bind PtdIns(3)P strongly, while several have lower affinity for other, and in some cases several, phosphoinositides.

Recently, the GRAM-like, ubiquitin-binding on Eap45 (GLUE) of the Vps36 subunit of the ESCRT sorting complex (ESCRT, endosomal sorting complex required for transport) has been shown to be required for recognition of PI(3)P-containing membranes. This domain shows little similarity to either the PH or FYVE domain.

PtdIns(3)P-Binding Proteins in Endosomal Membrane Trafficking

In accordance with the predominant localization of PtdIns(3)P to membranes of the endosomal system, several PtdIns(3)P-binding proteins have been shown to function in various steps of endosome fusion, sorting, and maturation (Table 1). In mammalian cells, a major function of these events is to direct membrane components to the lysosome, whereas sorting of proteins to the vacuole is a major function in yeast.

Following internalization of plasma membrane patches to form small membrane vesicles denoted as early endosomes, a process of maturation is initiated which involves fusion of early endosomes into larger multivesicular bodies. These multivesicular bodies form specialized subdomains marked by the recruitment of different sets of proteins. Eventually these specialized domains and their cargo proteins form new compartments, which may then be returned to the plasma membrane or delivered to lysosomes or the trans-Golgi complex.

Table 1 Cellular functions of PtdIns(3)P and some associated proteins

Cellular function	Domain	PtdIns(3)P-binding proteins
Endosome fusion/maturation	FYVE	EEA1, rabenosyn-5, PIKfyve/Fab1p
Golgi-vacuole trafficking (yeast)	FYVE PX	Vps27p, Vac1p, Fab1p Grd19p, Mvp1p, Vps5p, Vam7p
Receptor sorting/targeting to lysosome	FYVE PX GLUE	Hrs Sorting nexins Vps36p
Signal transduction	FYVE PX	SARA CISK
Phagosome function/maturation	FYVE PX	EEA1 p40 ^{phox}

This process is controlled in part by various members of the small guanine triphosphatase (GTPase) Rab family. Rab5 in particular is required for fusion of early endosomes and it accomplishes this in part by stimulating the production of PtdIns(3)P via Vps34, with the resulting recruitment of PtdIns(3)P-binding proteins to the endosomes. Such proteins include EEA1 and rabenosyn-5, which both bind Rab5 itself as well as PtdIns(3)P via a FYVE domain. Once recruited to the endosome through these mechanisms, EEA1 and rabenosyn-5 appear to function in the tethering, docking, and ultimate fusion of early endosomes.

The ESCRT machinery conserved in yeast and consists of several multiprotein complexes and is critical in the trafficking of ubiquitinated cargo proteins and maturation of endosomes. The FYVE-domain protein Hrs binds PtdIns(3)P and is a subunit of the ESCRT-0 complex. Likewise, binding of PtdIns(3)P (and likely other phosphoinositides) via the GRAM-like ubiquitin binding in EAP45 (GLUE) domain is required for the function of the ESCRT-II subunit Vps36.

The sorting nexin (SNX) family containing up to 25 human proteins and the yeast homologs Mvp1p, Vps5p, Vps17p, and Grd19p all contain a PX domain. All of the yeast SNX proteins and some of the mammalian counterparts have been found to bind preferentially or specifically to PtdIns(3)P, and most characterized SNX proteins localize at least partially to endosomes. Consistent with the importance of PtdIns(3)P in vacuolar protein sorting in yeast, Vps5p and Vps17p have a role in the retrieval of the vacuolar protein sorting receptor Vps10p, while Mvp1p and Grd19 participate in Golgi to vacuole trafficking. In mammalian cells, roles of the various SNX proteins are being assigned, and the emerging picture suggests that different sorting nexins may be required for regulating the distinct fates of endosomes, that is, recycling to the cell surface, delivery to lysosomes, or delivery to the trans-Golgi.

PtdIns(3)P in Phagocytosis and Bacterial Invasion

Many of the PtdIns(3)P-binding proteins involved in the endocytic process are likely involved in the similar processes of phagocytosis and host cell invasion by bacteria and parasites. Phagosomes, like endosomes, form by invagination of patches of the plasma membrane followed by 'pinching off' and sealing to form a self-contained intracellular compartment. PtdIns(3)P appears transiently on phagosomes in a manner which parallels its appearance and disappearance in endosomes. Phagosomal PtdIns(3)P is apparently generated by the targeting and activation of the VPS34 PI 3-kinase. As in early endosomes, the PtdIns(3)P synthesized in phagosomes recruits FYVE- and PX-domain containing proteins including EEA1. Likewise, in the process of secretion-mediated invasion by *Salmonella*, the membrane vacuole which envelops the invading bacterial cells accumulates PtdIns(3)P. In certain cell types, recruitment of the NADPH oxidase to phagosomes may occur via PX domains in its p40 and p47 subunits. This mechanism may contribute to defense against pathogens since the NADPH oxidase produces reactive oxygen species toxic to invading bacterial cells. Disruption of this recruitment mechanism by pathogenic bacteria enhances their ability to escape host cell defenses.

PtdIns(3)P and Signaling from the Endosomal System

Recently, it has become clear that the endosomal system is an essential component of signal transduction pathways regulating proliferation, survival, differentiation, and stress responses. Current evidence suggests that growth factor receptors, many of which are internalized into endosomes following binding of their ligands, can continue to generate signals to downstream effectors as they transit through the endosomal system. Furthermore, internalization and delivery of such receptors to lysosomes for degradation is an important mechanism for downregulation of signaling in a variety of systems.

The FYVE-domain protein, SARA, localizes to early endosomes via its association with PtdIns(3)P. SARA also binds to Smad2, a downstream target of the transforming growth factor- β (TGF β) receptor. Association of the SARA–Smad complex appears to be essential for efficient phosphorylation of Smad by the internalized receptor. The related FYVE-domain protein endofin may play a similar role in Smad1 binding and bone morphogenetic protein (BMP) signaling. Thus, PtdIns(3)P in endosomes can serve to assemble signaling complexes in which the local concentration of components is increased by association with specific intracellular membranes. The FYVE domain-containing protein homolog of the yeast protein Vps27p (Hrs) may function in a similar manner in the TGF β signaling pathway. However, in its function in the ESCRT complex, Hrs may promote the internalization and ultimate degradation of growth factor receptors. Thus, there is precedent for the involvement of PtdIns(3)P-binding proteins in attenuation as well as enhancement of growth factor signals.

Endosomal localization via PtdIns(3)P binding may also serve to target or sequester components of signaling pathways several steps removed from the initial receptor signal. The PX domain-containing protein kinase CISK (also denoted SGK3) is localized to endosomes and may function in this way. Like its related family members Akt and SGK, CISK is activated by regulatory protein kinases in response to growth factor-mediated regulation of PtdIns 3-kinases that produce PtdIns (3,4,5)P₃. Based on its endosomal disposition, it seems likely that the substrates of the CISK protein kinase are similarly localized. Few specific substrates have been identified but an effect on hair follicle growth and development in knockout mice suggests an important role for CISK in epithelial and follicle development.

PtdIns(3)P may also serve as a component of the cellular stress response. In yeast and mammalian cells, PtdIns(3,5)P₂ is formed in response to osmotic stress by the FYVE-domain

proteins Fab1p and PIKFYVE, respectively. Effectors of PtdIns (3,5)P specific for this function have not been identified.

Vps34 and Mammalian Target of Rapamycin Signaling and Control of Autophagy

The mammalian target of rapamycin (mTOR) protein kinase is a key component in control of cell growth in response to nutrients, and affects processes such as protein synthesis that are highly dependent on cellular stores of energy and raw materials. Recent studies have implicated the class III PtdIns 3-kinase Vps34 as a positive regulator of the mTOR response to amino acids. Conversely, mTOR is a negative regulator of autophagy, frequently a response of cells to starvation conditions. Surprisingly, Vps34 has also been shown to play an essential role in autophagy, while the myotubularin family PtdIns 3-phosphatase jumpy appears to act as a negative regulator of this process. These results suggest that control of PtdIns(3)P levels, perhaps at specific cellular locations such as membrane systems involved in autophagy, may be key control points for autophagy as well as cellular responses to nutrient levels.

See also: [Signaling: Mitogen-Activated Protein Kinase Family; Phosphoinositide 4- and 5-Kinases and Phosphatases; Phosphoinositide-Dependent Protein Kinases; Phosphatidylinositol Bisphosphate and Trisphosphate.](#)

Further Reading

- Backer JM (2008) The regulation and function of class III PI3Ks: Novel roles for Vps34. *Biochemical Journal* 410: 1–17.
- Brummell JH and Grinstein S (2003) Role of lipid-mediated signal transduction in bacterial internalization. *Cellular Microbiology* 5: 287–297.
- Clague MJ, Urbe S, and de Lartigue J (2009) Phosphoinositides and the endocytic pathway. *Experimental Cell Research* 315: 1627–1631.
- Cullen PJ (2008) Endosomal sorting and signalling: An emerging role for sorting nexins. *Nature Reviews Molecular Cell Biology* 9: 574–582.
- Czech MP (2003) Dynamics of phosphoinositides in membrane retrieval and insertion. *Annual Review of Physiology* 65: 791–815.
- Lemmon MA (2003) Phosphoinositide recognition domains. *Traffic* 4: 201–213.
- Raiborg C and Stenmark H (2009) The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature* 458: 445–452.
- Wishart MJ and Dixon JE (2002) PTEN and myotubularin phosphatases: From 3-phosphoinositide dephosphorylation to disease. *Trends in Cell Biology* 12: 579–585.

Phosphatidylinositol Bisphosphate and Trisphosphate

A Toker, Harvard Medical School, Boston, MA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 266–271, © 2004, Elsevier Inc.

Glossary

PH domain A protein domain composed of approximately of 100 amino acids which can specifically bind to PtdIns-4,5-P₂ and PtdIns-3,4,5-P₃.

Phosphatidylinositol A membrane phospholipid, which is a precursor to all known phosphoinositides and is composed of an inositol head ring and a fatty-acid tail.

Phosphoinositide A phospholipid second messenger, which is found in intracellular membranes and which can be phosphorylated on the inositol head ring at the D-3, D-4, and D-5 positions resulting in seven known phosphorylated species.

Phosphoinositide kinase An enzyme which can use phosphoinositides as substrates, and can phosphorylate the inositol head ring at either the D-3, D-4, or the D-5 positions.

Phosphoinositide phosphatase An enzyme which can specifically remove the D-3, D-4, or D-5 phosphate from phosphorylated phosphoinositides.

Phospholipase An enzyme which can hydrolyze the phosphodiester bond in phospholipids and produce soluble inositol lipids and membrane-bound diacylglycerol; type C phospholipases specifically use phosphoinositides as substrates.

Phosphatidylinositol (PtdIns) is a minor membrane phospholipid, which is found in all eukaryotic cells. It is the precursor of a family of phosphorylated lipids, which are collectively known as phosphoinositides, and rather than serving a structural role in cellular membranes, they act as functional second messengers to transduce biochemical signals in the cell. This article discusses in detail the synthesis, hydrolysis, and cellular functions of phosphatidylinositol bisphosphate and trisphosphate.

Phosphoinositides are composed of a water-soluble inositol ring, which can be phosphorylated at each of three positions (D-3, D-4, and D-5), and a greasy fatty-acid tail which is attached to a glycerol and linked to the inositol ring by a phosphodiester bond (Figure 1). The concerted actions of phosphoinositide kinases, phosphoinositide phosphatases, and phospholipases on phosphatidylinositol lead to the production of seven additional phosphoinositides which are known to exist in cells (Figure 2). Of these, phosphatidylinositol-4,5-bisphosphate (PtdIns-4,5-P₂) and phosphatidylinositol-3,4,5-trisphosphate (PtdIns-3,4,5-P₃) are the best understood and characterized. Both the synthesis and the degradation of PtdIns-4,5-P₂ and PtdIns-3,4,5-P₃ are tightly regulated, and this ensures efficient and coordinated signaling through these second messengers. Biochemical and genetic studies have implicated both PtdIns-4,5-P₂ and PtdIns-3,4,5-P₃ in numerous cellular responses, such as cell growth, cell survival, cell motility and modulation of the actin cytoskeleton, gene transcription, and vesicle trafficking (Figure 3). The importance of these phosphoinositides in cell biology is exemplified by the fact that deregulation in the metabolism or mechanism of action of these lipids often results in human diseases such as cancer and type 2 diabetes.

Phosphatidylinositol-4,5-Bisphosphate

Synthesis

Phosphatidylinositol-4,5-bisphosphate (PtdIns-4,5-P₂) is often considered to be one of the most important functional lipids

in eukaryotic cells, because it not only has a functional second messenger role *per se*, but can also serve as an intermediate for the production of additional lipid mediators. It can serve as a substrate for phosphoinositide 3-kinase (PI 3-K) to produce phosphatidylinositol-3,4,5-trisphosphate (PtdIns-3,4,5-P₃), it can be hydrolyzed by phospholipases to produce diacylglycerol (DAG) and soluble inositol phosphates, and it can also be dephosphorylated by D-4 and D-5 lipid phosphatases (Figure 2). Although the bulk of PtdIns-4,5-P₂ is found at the plasma membrane, other distinct pools of PtdIns-4,5-P₂ have been reported. For example, there exists a nuclear phosphoinositide cycle where PtdIns-4,5-P₂ production has been detected. However, there are relatively modest, if any, changes in the total levels of PtdIns-4,5-P₂ upon cellular stimulation with agonists such as growth factors or hormones. This is difficult to reconcile with the second messenger hypothesis, which dictates that a lipid mediator must rapidly accumulate to amplify downstream signaling. Therefore, it is likely that the constitutive levels of PtdIns-4,5-P₂ change locally at discrete intracellular sites, such that the accumulation of this lipid is regulated both spatially and temporally. In addition, there are two other isomers of phosphatidylinositol bisphosphate: PtdIns-3,4-P₂ and the recently discovered PtdIns-3,5-P₂. The function of these two phosphoinositides is not entirely clear, but they are likely to mediate certain responses downstream of PI 3-K.

The synthesis of PtdIns-4,5-P₂ is controlled by the sequential action of phosphoinositide 4-kinases, which use phosphatidylinositol (PtdIns) as substrate and produce PtdIns-4-P, and phosphoinositide phosphate kinases (PIPKs), which can use either PtdIns-4-P or PtdIns-5-P as substrates. In both cases, the product is PtdIns-4,5-P₂ (Figure 2). Much is known about the enzymes which produce PtdIns-4,5-P₂. There are two families of PIPKs, type I and type II. Type I kinases preferentially phosphorylate PtdIns-4-P at the D-5 position, and this appears to be the major route of PtdIns-4,5-P₂ synthesis. However, it was recently discovered that type II PIPKs can phosphorylate

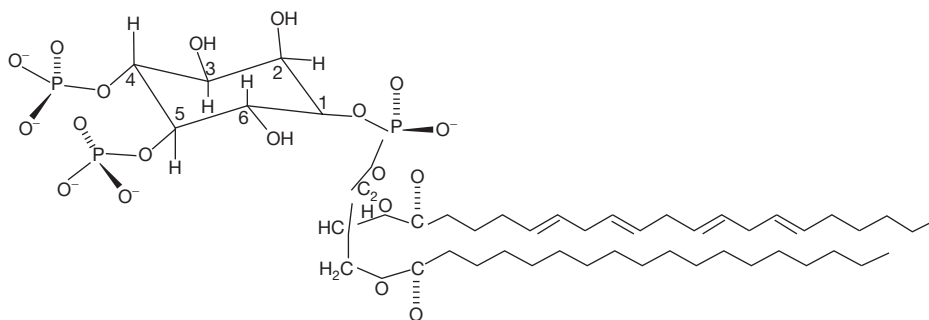


Figure 1 Structure of phosphatidylinositol-4,5-bisphosphate. The structure shows PtdIns-4,5- P_2 , which contains an inositol head group connected to a diacylglycerol via a phosphodiester linkage. The fatty acid moiety is typically stearyl-arachidonyl. The inositol head group can be phosphorylated at one of three positions, D-3, D-4, and D-5. Shown are phosphate groups at the D-4 and D-5 position in PtdIns-4,5- P_2 . Phosphoinositide 3-kinases phosphorylate the D-3 position of PtdIns-4,5- P_2 to produce PtdIns-3,4,5- P_3 .

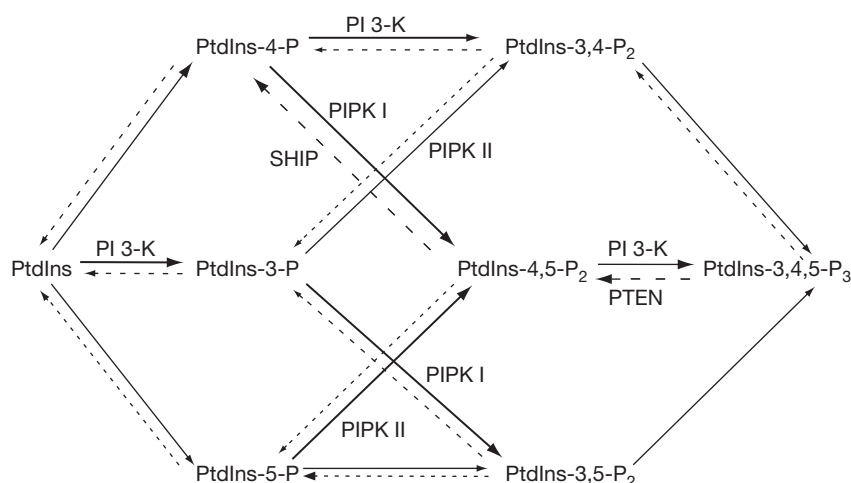


Figure 2 Phosphoinositide metabolism. The pathways responsible for the production of all known phosphoinositides are shown. The concerted actions of phosphoinositide kinases, such as type I and type II phosphoinositide 4-phosphate and 5-phosphate kinases (PIPKs) and phosphoinositide 3-kinases (PI 3-K), as well as phosphatases (PTEN and SHIP), leads to the production of seven distinct phosphoinositides, from the precursor phosphatidylinositol (PtdIns). Although all of the forward (solid arrows, mediated by kinases) and reverse (broken arrows, mediated by phosphatases) reactions are possible, the pathways which are likely to predominate in cells are shown in bold arrows.

PtdIns-5-P at the D-4 position, and so it is likely that in some instances, this represents an important pathway for the generation of PtdIns-4,5- P_2 . Several forms of both type I and type II PIPKs exist in mammalian cells, and homologs are also found in lower eukaryotes, such as yeasts, highlighting the importance of these enzymes in biology. The crystal structure of a type II enzyme has been solved, and has revealed an elongated, flat surface covered with highly basic, positively charged amino acids, which are ideally positioned for interaction with the negatively charged phosphoinositides such as PtdIns-5-P. Finally, a separate route of synthesis for PtdIns-4,5- P_2 is the dephosphorylation of PtdIns-3,4,5- P_3 by D-3 phosphoinositide phosphatases, though it is not clear what fraction of the bulk PtdIns-4,5- P_2 pool is accounted for by this mechanism. Regardless, PIPKs represent the major pathway leading to the production of PtdIns-4,5- P_2 , and as such their activities are tightly controlled.

Hydrolysis

PtdIns-4,5- P_2 is hydrolyzed by phosphoinositide-specific phospholipases. These enzymes, also known as type C phospholipases (PLC), hydrolyze PtdIns-4,5- P_2 into two important second messengers: the membrane-bound fatty acid moiety, also known as DAG, and the soluble portion, which is an inositol trisphosphate (Ins-1,4,5- P_3 , when PtdIns-4,5- P_2 is the PLC substrate). Each of these products has important roles in lipid signaling. DAG is the activator for members of the protein kinase C family of serine/threonine kinases, which are recruited to the membrane at the site of DAG production. Ins-1,4,5- P_3 plays an equally important role in lipid signaling as it is responsible for the release of Ca^{2+} from internal stores, such as the endoplasmic reticulum.

Although four distinct families of PLC exist in mammalian cells, they all share a similar catalytic mechanism employing

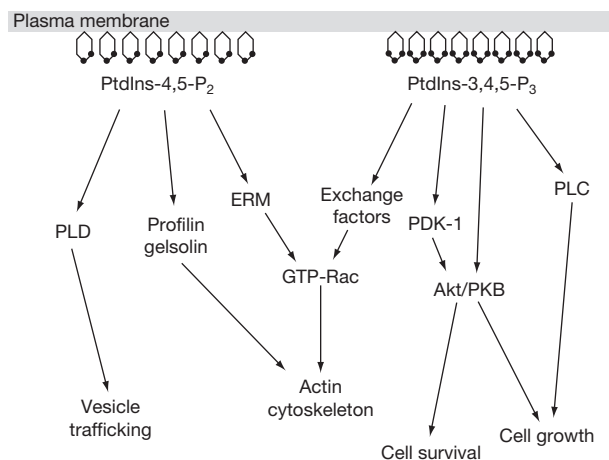


Figure 3 Targets of PtdIns-4,5- P_2 and PtdIns-3,4,5- P_3 . A large number of proteins whose activities are directly regulated by either PtdIns-4,5- P_2 or PtdIns-3,4,5- P_3 exist. A few of these are depicted and serve to illustrate that the binding of the phosphoinositide to the protein can alter its cellular location as well as directly influencing catalytic activity when enzymes are the targets. These interactions are responsible for the activation of secondary signaling pathways, which ultimately lead to numerous physiological responses, as depicted. In many cases, PtdIns-4,5- P_2 and PtdIns-3,4,5- P_3 have overlapping functions, due to the activation of distinct effector molecules, such as the remodeling of the actin cytoskeleton, which occurs as a consequence of both PtdIns-4,5- P_2 - and PtdIns-3,4,5- P_3 -dependent signaling.

calcium as a cofactor. All members of the PLC family use PtdIns-4-P and PtdIns-4,5- P_2 as substrates. PLC β enzymes are regulated by binding to both $\beta\gamma$ - as well as α -subunits of heterotrimeric GTP-binding proteins. PLC δ also appears to be regulated by G proteins, although the mechanism is unclear, and it is also reported that increases in Ca^{2+} alone are sufficient to activate PLC δ . Phospholipases belonging to the PLC γ family are activated by the binding of SH2 domains to phosphotyrosine-containing sequences in receptor tyrosine kinases, and as such participate in growth-factor-dependent responses such as cell growth. Tyrosine phosphorylation of PLC γ is also important in controlling catalytic activity. Very recently, a new PLC subtype has been described, PLC ϵ , and has been shown to be a novel effector of the GTP-bound form of the Ras GTPase. The diversity and complexity of the PLC family underscore the importance of PtdIns-4,5- P_2 hydrolysis and DAG/Ins-1,4,5- P_3 production in mediating cell proliferation, differentiation, and cell motility.

Cellular Function

In addition to serving as the substrate for additional lipid second messengers, PtdIns-4,5- P_2 is an important lipid mediator in its own right. Numerous studies have shown that PtdIns-4,5- P_2 can modulate the activity of proteins which regulate actin assembly, leading to rearrangements of the actin cytoskeleton (Figure 3). For example, PtdIns-4,5- P_2 has been shown to promote actin polymerization by directly interacting with actin-binding proteins such as profilin. Similarly, actin-capping proteins, such as CapZ, prevent spontaneous actin assembly by capping the free ends of actin filaments,

and the binding to PtdIns-4,5- P_2 disrupts this capping event leading to filament assembly. The list of actin-binding proteins whose activities are also regulated by PtdIns-4,5- P_2 includes gelsolin, cofilin, filamin, and vinculin. The small GTPase Rac also plays an important role in the remodeling of the actin cytoskeleton, and in some instances this requires PtdIns-4,5- P_2 . For example, the Ezrin, Radixin, and Moesin (ERM) family of proteins, a family of Rac effector molecules, are also PtdIns-4,5- P_2 -binding proteins which regulate dynamic actin assembly, again linking PtdIns-4,5- P_2 function to the actin cytoskeleton. Rac, in the active GTP bound form, also interacts with PIP kinases, which are responsible for PtdIns-4,5- P_2 synthesis. Recent studies in cells have provided further evidence that PtdIns-4,5- P_2 is a critical regulator of the actin cytoskeleton. For example, expression of type I PIPK enzymes in cells leads to a dramatic reorganization of the actin cytoskeleton, and PIPK I appears to be targeted to focal adhesions, sites of cellular contact with the extracellular matrix. This targeting provides clear evidence for the localized accumulation of PtdIns-4,5- P_2 in the cell. Studies such as these, carried out in mammalian cells, have also been corroborated in yeasts, whereby genetic ablation of PIP kinases (such as Mss4p) leads to abnormalities in the actin cytoskeleton, which can be rescued by the introduction of human PIPK isoforms.

PtdIns-4,5- P_2 is also an important cofactor for the activity of a number of signaling enzymes. The best example is for phospholipase D (PLD), an enzyme responsible for the production of phosphatidic acid (PA) in the cell. PLD requires PtdIns-4,5- P_2 for efficient catalytic function. As certain PIP kinases also depend on PA for enzymatic activity, and PtdIns-4,5- P_2 stimulates PLD leading to PA production, the net result is a positive amplification loop which leads to amplification of PA and PtdIns-4,5- P_2 synthesis at specific intracellular compartments where these enzymes are located. Both PIPKs and PLD have been implicated in trafficking and secretion, and as such, their products are directly involved in controlling vesicle formation, as well as endocytosis (Figure 3).

Phosphatidylinositol-3,4,5-Trisphosphate

Synthesis

Unlike its substrate PtdIns-4,5- P_2 , phosphatidylinositol-3,4,5-trisphosphate (PtdIns-3,4,5- P_3) is nominally absent in unstimulated cells and rapidly accumulates in response to virtually all known growth factors, hormones, cytokines, and other stimuli. In most cells, production of PtdIns-3,4,5- P_3 typically peaks within seconds to a few minutes after stimulation, and declines to near basal levels after 1–2 h. A second peak of PtdIns-3,4,5- P_3 has been observed in cells between 4 and 8 h after stimulation with mitogens such as platelet-derived growth factor (PDGF). This second peak of PtdIns-3,4,5- P_3 is responsible for efficient cell-cycle progression in fibroblasts, specifically entry into the S phase where DNA synthesis occurs. Although the bulk of PtdIns-3,4,5- P_3 is produced at the plasma membrane, it is also known that there are discrete intracellular locations where this phosphoinositide is also found, particularly the nucleus. PI 3-Ks are responsible for the production of PtdIns-3,4,5- P_3 in higher eukaryotes. These enzymes phosphorylate the D-3 position of the inositol head ring, and

depending on the substrate, four distinct products can result: PtdIns-3-P, PtdIns-3,5-P₂, PtdIns-3,4-P₂, and PtdIns-3,4,5-P₃ (Figure 1). Although different forms of PI 3-Ks exist, each with unique substrate selectivity, the class Ia enzymes are responsible for the production of PtdIns-3,4,5-P₃ in cells. These PI 3-Ks are composed of a p85 regulatory subunit and a p110 catalytic subunit. The regulatory subunit binds to specific phosphotyrosine-containing sequences in the cytosolic domains of activated receptor tyrosine kinases, via a tandem repeat of two Src-Homology 2 (SH2) domains. This interaction facilitates the activation of the catalytic subunit, whose activity is also increased by a direct interaction with the activated form of the small GTPase Ras. The newly activated PI 3-K converts PtdIns-4,5-P₂ to PtdIns-3,4,5-P₃. In addition, class Ib PI 3-Ks are also capable of producing PtdIns-3,4,5-P₃, and these kinases are typically activated downstream of heterotrimeric G protein-coupled receptors, such as the thrombin receptor in platelets. Class Ib enzymes are also composed of two subunits, a p120 catalytic subunit and a p101 regulatory subunit, which confers sensitivity to Gβγ. Class Ib enzymes also require activated Ras for full catalytic function. Thus, the primary route of synthesis for PtdIns-3,4,5-P₃ in cells is through the activation of class I PI 3-Ks, which use PtdIns-4,5-P₂ as substrate.

PtdIns-3,4,5-P₃ Phosphatases

As the production of PtdIns-3,4,5-P₃ is critical for a number of physiological responses in cells, mechanisms exist which mediate the termination of PI 3-K and PtdIns-3,4,5-P₃ signaling. PI 3-K lipid products, including PtdIns-3,4,5-P₃, are not degraded by phospholipases, but instead are downregulated by at least two different types of phosphoinositide phosphatases. SHIP1 and SHIP2 (SH2-containing inositol 5'-phosphatases) are lipid phosphatases which specifically remove the D-5 phosphate from both PtdIns-3,4,5-P₂ and PtdIns-3,4,5-P₃ (Figure 2). The action of SHIP1/2 on PtdIns-3,4,5-P₃ results in the production of PtdIns-3,4-P₂, and although this can impair signaling downstream of PI 3-K, PtdIns-3,4-P₂ is able to mediate certain responses which overlap with PtdIns-3,4,5-P₃. Despite this, loss of SHIP2 increases insulin sensitivity in responsive tissues such as muscle and fat. Phosphatase and tensin (PTEN) homolog deleted on chromosome 1 is a D-3 phosphatase which dephosphorylates PtdIns-3,4,5-P₃ with PtdIns-4,5-P₂ as the product (Figure 2). PTEN is very effective at terminating PI 3-K and PtdIns-3,4,5-P₃ signaling, and this is evidenced by the fact that mutations in the PTEN gene have been detected in a large fraction of advanced human tumors. For this reason, PTEN is referred to as a tumor-suppressor gene. Thus, efficient termination of PtdIns-3,4,5-P₃ signaling is of critical importance for intracellular signaling, and deregulation of this signal can have adverse consequences for cellular responses, often leading to diseases such as diabetes and cancer.

PtdIns-3,4,5-P₃-Binding Proteins

The mechanism by which PI 3-K activates downstream secondary signaling pathway is by the recruitment of cytosolic proteins to the plasma membrane at the site of PtdIns-3,4,5-P₃ production. This recruitment can have a number of consequences; in addition to causing a change in the cellular

location of the protein, it can induce a conformation change, and in the case of enzymes, it can increase (and sometimes decrease) enzymatic activity. Typically, a combination of these events results from the PtdIns-3,4,5-P₃-protein interaction. A number of protein domains have evolved, which specifically recognize phosphoinositides, particularly PtdIns-3,4,5-P₃. The first described example is the pleckstrin homology (PH) domain, a 100-amino-acid sequence found in several signaling proteins which participate in PI 3-K responses. PH domains are found in serine-threonine protein kinases (e.g., Akt/PKB and PDK-1), tyrosine kinases (e.g., Tec family kinases such as Btk), and guanine nucleotide exchange factors of small GTP-binding proteins (e.g., SOS, Grp-1, and pREX), and GTPases themselves (e.g., dynamin). In some of these cases, the PtdIns-3,4,5-P₃-protein interaction is well understood. For example, recruitment of Akt/PKB to the plasma membrane occurs in response to PtdIns-3,4,5-P₃ production. The PH domain of Akt/PKB directly binds to PtdIns-3,4,5-P₃ with high affinity and selectivity, and this leads to a conformational change in the protein kinase, which leads to unfolding of the PH domain, revealing a sequence in the catalytic domain known as the activation loop. The PDK-1 kinase, which also has a PH domain and is recruited to the membrane simultaneously with Akt/PKB, can now phosphorylate Akt/PKB at a critical threonine residue in this loop. The phosphorylated Akt/PKB is now fully catalytically competent and can itself phosphorylate downstream protein substrates. This elegant mechanism is recapitulated in a large number of other protein kinases, which are also regulated by PtdIns-3,4,5-P₃ or PDK-1, or both (e.g., protein kinase C and S6-kinases).

Recently, a distinct phosphoinositide-binding domain was described, and termed the phox (PX) domain. This 130-amino-acid domain is also found in many important signaling proteins, and was originally described in the p40^{phox} and p47^{phox} subunit of the NADPH-oxidase superoxide-generating complex in phagocytic cells. Although PX domains do not appear to bind to PtdIns-3,4,5-P₃, they bind with high affinity to PtdIns-3-P and PtdIns-3,4-P₂ and thus are important in mediating signaling downstream of these PI 3-K lipids. Another domain which has been shown to bind PtdIns-3,4,5-P₃ is the SH2 domain. Although less well understood, binding of PtdIns-3,4,5-P₃ to the SH2 domain of the phospholipase PLCγ-1, for example, leads to an increase in its enzymatic activity. Thus, several domains have evolved in higher eukaryotic cells which have the ability to bind with distinct affinities and selectivity to PtdIns-3,4,5-P₃, and other phosphoinositides. This interaction has a profound effect on downstream signaling and cellular responses.

Cellular Function

Numerous biochemical, pharmacological, and genetic studies have demonstrated that PI 3-K, and its product PtdIns-3,4,5-P₃, regulates a multitude of cellular responses. These include cell growth, protection from cell death (apoptosis), remodeling of the actin cytoskeleton, cell migration and invasion, metabolism and insulin responses, cell-cycle regulation, and gene transcription (Figure 3). It is beyond the scope of this article to discuss all of these in any detail, but some notable examples are given. Activation of the protein kinase Akt/PKB by the

PI 3-K pathway regulates cellular survival mechanisms, as well as insulin-dependent metabolism. Most proteins which mediate Akt/PKB-dependent cellular survival are inhibited by the phosphorylation event. For example, phosphorylation of the Forkhead family of transcription factors (e.g., FKHR-L1) by Akt/PKB inhibits their translocation to the nucleus due to the creation of a binding site for 14-3-3, a family of molecular chaperones. This retention of FKHR-L1 in the cytoplasm inhibits the transactivation of genes intimately associated with cell death, and the net effect is an increase in survival. Similarly, Akt/PKB phosphorylation of the pro-apoptotic protein Bad also facilitates 14-3-3 binding, and this prevents Bad from binding to the anti-apoptotic proteins Bcl-2 and Bcl-X_L, effectively preventing apoptosis. Akt/PKB can also regulate both cell-cycle progression and gluconeogenesis/glycolysis by phosphorylating glycogen synthase kinase-3 (GSK-3). This phosphorylation inhibits the elevated constitutive activity of GSK-3, which can no longer phosphorylate proteins such as Myc, the cell-cycle regulator cyclin D, and glycogen synthase. Thus, Akt/PKB signaling results in the activation of several pathways which are normally inhibited by GSK-3.

Not all PI 3-K responses are mediated by Akt/PKB. In fact, there are several other kinases which are activated by PtdIns-3,4,5-P₃ signaling, and which have important roles in cell physiology. For example, the S6-kinases, S6K-1 and S6K-2, regulate protein synthesis and insulin responses. The protein kinase C family of kinases can also be regulated by PtdIns-3,4,5-P₃-dependent mechanisms, and have been shown to mediate both cell growth and survival mechanisms, though these are not as well understood as the Akt/PKB pathway. The Tec family kinase Bruton's tyrosine kinase (Btk), which binds to PtdIns-3,4,5-P₃ via its PH domain, is important for B cell development, and importantly, genetic mutations within its PH domain have been shown to give rise to certain immunodeficiencies in humans.

Finally, regulation of the dynamics of the actin cytoskeleton is also achieved by PtdIns-3,4,5-P₃-dependent responses, and this leads to increased cell migration of many cell types (Figure 3). One important mechanism by which PtdIns-3,4,5-P₃ regulates such responses is by activating the small

GTP-binding protein Rac, known to contribute to actin remodeling and directional migration in response to chemotactic stimuli. Small GTPases, such as Rac, are activated by binding to GTP, and this is stimulated by the activation of Rac-specific guanine nucleotide exchange factors, such as SOS, Vav, and, in particular, pREX. These exchange factors have PH domains which in some cases have been shown to bind to PtdIns-3,4,5-P₃, leading to an increase in exchange activity toward Rac. In summary, the diverse cellular responses attributed to PI 3-K signaling are explained by the fact that there exist numerous signaling proteins which interact with PtdIns-3,4,5-P₃. This binding has a profound consequence on the function of the protein, leading to the activation of downstream pathways.

See also: **Bioenergetics:** Phosphatidylinositol-3-Phosphate; **Signaling:** Diacylglycerol Kinases and Phosphatidic Acid Phosphatases; Inositol Phosphate Kinases and Phosphatases; Phosphoinositide 4- and 5-Kinases and Phosphatases; Phosphoinositide-Dependent Protein Kinases; Phospholipase C; Phospholipase D.

Further Reading

- Anderson RA, Boronenkov IV, Doughman SD, Kunz J, and Loijens JC (1999) Phosphatidylinositol phosphate kinases, a multifaceted family of signaling enzymes. *Journal of Biological Chemistry* 274: 9907–9910.
- Cantley LC (2002) The phosphoinositide 3-kinase pathway. *Science* 296: 1655–1657.
- Cullen PJ, Cozier GE, Banting G, and Mellor H (2001) Modular phosphoinositide-binding domains – their role in signalling and membrane trafficking. *Current Biology* 11: R882–R893.
- Leslie NR, Biondi RM, and Alessi DR (2001) Phosphoinositide-regulated kinases and phosphoinositide phosphatases. *Chemical Reviews* 101: 2365–2380.
- Rhee SG (2001) Regulation of phosphoinositide-specific phospholipase C. *Annual Review of Biochemistry* 70: 281–312.
- Toker A (2002) Phosphoinositides and signal transduction. *Cellular and Molecular Life Sciences* 59: 761–779.
- Vivanco I and Sawyers CL (2002) The phosphatidylinositol 3-kinase AKT pathway in human cancer. *Nature Reviews Cancer* 2: 489–501.
- Wishart MJ and Dixon JE (2002) PTEN and myotubularin phosphatases: From 3-phosphoinositide dephosphorylation to disease. Phosphatase and tensin homolog deleted on chromosome ten. *Trends in Cell Biology* 12: 579–585.

Phosphofructokinase-2/Fructose Bisphosphatase-2

D M Raben, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Anomer Cyclic stereoisomers of saccharides with different configurations at the hemiacetal or hemiketal (anomeric) carbon. The configuration is designated as either α or β .

ATP (adenosine triphosphate) A molecule consisting of the nitrogenous base adenine linked to the sugar ribose and contains three phosphate groups attached to the ribose. ATP, present in all living cells, serves as the principal energy source of energy for cells.

cAMP (cyclic adenosine monophosphate) A cyclic form of AMP where a single phosphate is attached to both the third and fifth carbon of the ribose sugar linked to adenine. It is important in transmitting intracellular, as well as extracellular, signals induced by many agonists. cAMP binds to and activates the cAMP-dependent protein kinase (PKA).

Gluconeogenesis The synthesis of glucose from noncarbohydrate sources occurring primarily in the liver and to a small extent in the kidney. It is stimulated during starvation or intense exercise.

Glycolysis (glycolytic pathway) The anaerobic metabolism of glucose to pyruvate. This metabolism generates two

molecules of ATP and one molecule of nicotinamide adenine dinucleotide (NADH) per glucose molecule. The NADH may be used to generate energy in the form of ATP and the pyruvate may enter the tricarboxylic acid cycle to yield NADH for energy production under aerobic conditions.

GTP (guanosine triphosphate) A molecule consisting of the nitrogenous base guanine linked to the sugar ribose. The ribose contains three phosphate groups attached to the 5' carbon. GTP is another source of metabolic energy and is involved in protein synthesis and a number of signal transduction cascades.

K_m A kinetic parameter that indicates the affinity of an enzyme for its substrate and often called the Michaelis constant. It is defined as the concentration of substrate that results in half the maximal rate (V_{max}) of the enzyme-catalyzed reaction.

Protein kinase A (PKA) An enzyme that catalyzes the phosphorylation of proteins on specific serine or threonine residues in response to cAMP.

V_{max} The maximal initial velocity of an enzyme-catalyzed reaction that occurs at saturating substrate concentrations.

Discovery of PFK-2/FBPase-2

As with many discoveries, the identification of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PFK-2/FBPase-2) was preceded by the discovery of its product fructose-2,6-bisphosphate (F2,6BP). This sugar was first identified as the mediator responsible for stimulating hepatic gluconeogenesis. It had been known for some time that gluconeogenesis was induced in liver by elevated levels of glucagon. Emil van Shaften and Henri-Gery Hers demonstrated that rat liver lysates contained an acid-labile phosphoric ester that greatly stimulated 6-phosphofructo-1-kinase (PFK-1). The concentration of this compound in isolated hepatocytes was significantly increased by glucose and inhibited by glucagon. They then showed that this molecule was F2,6BP. Following this discovery, Hers and Uyeda identified PFK-2/FBPase-2 in rat liver as the bifunctional enzyme responsible for the generation and degradation of F2,6BP.

General Structure of PFK-2/FBPase-2

In mammalian tissues, PFK-2/FBPase-2 is a single protein with the PFK-2 activity located at the N-terminal end and the FBPase-2 located at the C-terminus. Structural analyses

demonstrate that the catalytic domains are flanked by regulatory regions at both the N and C termini. The enzyme functions as a dimer with protein-protein contacts being made at the regulatory regions at the N-terminus only. In contrast to this structure, the yeast PFK-2 and FBPase-2 enzymes are separate proteins but likely function as dimers.

Enzymology of PFK-2

While PFK-2 shares the same substrate as PFK-1, fructose-6-phosphate (F6P), the reaction catalyzed by PFK-2 differs from that catalyzed by PFK-1 in four important aspects. First, PFK-1 catalyzes the transfer of the γ -phosphate of adenosine triphosphate (ATP) to carbon-1 (C-1) of F6P generating F1,6BP while PFK-2 catalyzes the transfer of this ATP phosphate to C-2 yielding F2,6BP. Second, PFK-1 prefers the α -anomer of F6P, while PFK-2 prefers the β -anomer. Third, PFK-1 is strongly inhibited by ATP while PFK-2 is not sensitive to this nucleotide. Finally, PFK-2 is inhibited by both of its products, adenosine diphosphate (ADP) and F2,6BP, which are activators of PFK-1.

The complete catalytic mechanism of PFK-2-mediated phosphorylation of F6P has not been elucidated. Important features of the reaction have, however, been identified. In general, PFK-2 activity is much weaker than that of PFK-1.

This may be due to the fact that specific residues involved in accelerating catalysis in PFK-1 are absent in PFK-2. The pH optimum of the enzyme-catalyzed reaction is between 8 and 9, requires magnesium, and prefers ATP as the phosphate donor but uses guanosine triphosphate (GTP). Inhibition studies indicate that the PFK-2-mediated transfer of phosphate from ATP to F6P occurs via a sequential-ordered mechanism with ATP binding prior to F6P binding. Binding sites for ATP and F6P have been identified besides residues critical for catalysis.

Enzymology of FBPase-2

Much more is known about the reaction mechanism used for FBPase-2-mediated hydrolysis. The enzyme has a pH optimum between 5.5 and 6.5 and does not require any divalent cations. FBPase-2 specifically hydrolyzes the phosphate on C-2 of F2,6BP and will not hydrolyze C-6 of either F2,6BP, F1,6BP, or glucose-1,6-bisphosphate. It is, however, inhibited by both its substrate and product: F6P and F2,6BP. Interestingly, the FBPase-2 reaction mechanism involves an intermediate in which F6P is covalently bound to a histidine residue on the enzyme. This covalent intermediate has been useful in structural studies designed to elucidate the reaction mechanism which appears to resemble, but is not identical to, the mechanism used by serine proteases. Both involve a catalytic triad in which three residues work together to form an intermediate in which the substrate is covalently bound to the enzyme followed by hydrolysis of this intermediate to release the product from the enzyme. In the case of serine proteases a serine residue on the enzyme is covalently bound to the catalytic intermediate, whereas in FBPase-2 a phosphohistidine is formed and F6P is released. The dissociation of the F6P product appears to be the rate-limiting step.

Isoforms of PFK-2/FBPase-2

Since the discovery of PFK-2/FBPase-2 in mammalian tissues, a number of mammalian isoforms have been identified from four genes which include alternatively spliced products. In all isoforms, the catalytic domains are highly conserved while the N and C termini are most divergent. Most tissues contain varying amounts of different isoforms with one isoform predominating. For example, in liver approximately 10% of the total PFK-2/FBPase-2 is the skeletal muscle isoform. Even within the predominant liver isoform, there are two different species, a 54–55-kDa species and a 58-kDa species, with only the 58-kDa species being regulated by phosphorylation (see below). In addition to these mammalian isoforms, yeast contains PFK-2 and FBPase-2 activities but in this case the enzymes are monomeric although they likely work as dimers.

In addition to sequence differences, kinetic properties of the kinase and bisphosphatase activities also differ among the various isoforms. While the substrate affinities (measured as a K_m) and maximum velocities (V_{max}) vary between isoforms, there are some notable differences. Comparison of bovine brain, liver, and testes shows that these isoforms have similar kinetic properties for PFK-2 in terms of K_m for F6P, but the liver and muscle FBPase-2 enzymes have K_m values for F2,6BP that

are approximately 100–1000-fold lower than the other isoforms. Based on V_{max} , the skeletal muscle also appears to have the lowest kinase activity and high phosphatase activity, while the placental isoform has the highest kinase activity with very little phosphatase activity.

Regulation and Physiological Roles of Specific Isoforms in Specific Tissues

PFK-2/FBPase-2 is exquisitely poised to regulate the levels of F2,6BP. F6P exerts a substrate-level control over PFK-2 and is a product inhibitor of FBPase-2. The K_m of PFK-2 and K_i of FBPase-2 for F6P are within the physiological concentration ranges of this phosphorylated sugar in different tissues, suggesting that small changes in the levels of F6P result in significant reciprocal alterations in these two enzymatic activities. Such alterations have different effects in different tissues, depending on the isoforms and relative level of the endogenous PFK-2/FBPase-2 activities.

Perhaps the best-studied form of PFK-2/FBPase-2 regulation involves the phosphorylation of the enzyme. Both the liver and heart isozymes are subject to regulation by phosphorylation. The brain isoform is phosphorylated but this phosphorylation does not appear to affect enzymatic activities. Muscle and testis isoforms are not phosphorylated. A liver isoform (58 kDa) was the first one shown to be regulated by a protein kinase A (PKA)-mediated phosphorylation stimulated by the glucagon-induced second messenger cyclic adenosine monophosphate (cAMP). Further studies showed that PFK-2/FBPase-2 is phosphorylated on a specific serine residue, serine-32, near the N-terminus. This phosphorylation leads to inhibition of PFK-2 and activation of FBPase-2. For PFK-2 at physiological pH, this is primarily due to an increase in the K_m (i.e., lower affinity) for F6P, without affecting the K_m for ATP, and only slightly perturbing V_{max} . In contrast, this phosphorylation activates FBPase-2 by increasing the V_{max} of the enzyme without affecting the K_m for F2,6BP. It is believed that this effect is largely due to an increase in the rate of F6P dissociation from FBPase-2 which enhances the rate at which the phosphohistidine intermediate is hydrolyzed. Consistent with the critical role of this enzyme in regulating liver carbohydrate metabolism, phosphorylation of the enzyme can be reversed by specific, regulated protein phosphatases (protein phosphatase 2A and 2C) which serves to activate PFK-2 while inhibiting FBPase-2. On the other hand, phosphorylation of the heart isoform modulates only the PFK-2 activity without affecting FBPase-2. Importantly, in stark contrast to the liver isoform, phosphorylation of the heart isoform activates PFK-2. In bovine and rat hearts, phosphorylation lowers the K_m (i.e., increases the affinity) for F6P. The bovine heart isoform may be phosphorylated by PKA and the rat heart isoform can be phosphorylated and modulated by either PKA or the calcium/calmodulin-dependent protein kinase.

The fact that PFK-2 and FBPase-2 are located near the N and C termini respectively on the same polypeptide begs the question as to how both activities of the liver enzyme are affected by a phosphorylation near the N-terminus. While the exact mechanism underlying the reciprocal regulation of these two activities has not been resolved, a current view is that in the

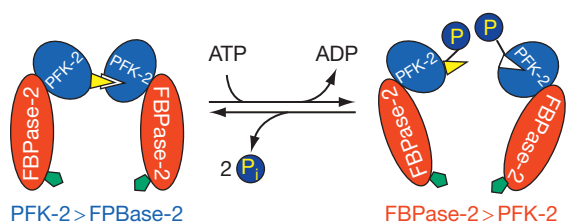


Figure 1 Reciprocal regulation of dimeric PFK-2/FBPase-2 in liver by phosphorylation. The blue indicates the N-terminal region containing PFK-2 activity and the red indicates the C-terminus containing the FBPase-2 activity. The triangle at the N-terminus and the pentagon at the C-terminus represent regulatory regions.

nonphosphorylated state, the N-termini of two polypeptides interact and activate the kinase while the C-terminal regions are separated (Figure 1). This is consistent with the crystal structure of the rat testis PFK-2/FBPase-2. Phosphorylation is believed to induce a conformational change in the enzyme resulting in the separation of the N-termini which inactivates the kinase while activating the phosphatase.

The above regulation provides for short-term, acute regulation for rapid, transient responses. The enzyme is also subject to long-term regulation mediated by gene expression. Insulin and dexamethazone both stimulate an increase in the level of PFK-2/FBPase-2 gene expression. Consistent with this, levels of liver PFK-2/FBPase-2 are depressed during starvation or diabetes and increase upon re-feeding or administration of insulin, respectively. In addition, levels of PFK-2/FBPase-2 decrease following partial hepatectomy and increase during liver regeneration.

The differences in isoforms and their regulation suggest different roles for each isoform. This is most dramatically seen when comparing the regulation of the liver isoform with that of the heart isoform. Liver not only consumes energy, but also stores energy mainly in the form of glycogen. It is also the only tissue, besides kidney to a small extent, which engages in gluconeogenesis to supply glucose to other glucose-dependent tissues in times of need. In that, as previously mentioned, the liver isoform is exquisitely poised to reciprocally regulate the level and activities of PFK-2/FBPase-2. This regulation is sensitive to glucose levels or hormones involved in regulating carbohydrate metabolism such as insulin (which stimulates glycolysis) and glucagons (which induces gluconeogenesis). On the other hand, heart muscle is a constant energy-consuming tissue and, therefore, engages in glycolysis but not gluconeogenesis. Indeed, the heart does not respond to glucagon but does respond to another hormone that signals through PKA, epinephrine. It is not surprising, therefore, that the regulated heart isoform of PFK-2/FBPase-2 responds to phosphorylation

by increasing PFK-2 activity as this hormone would signal a need for increased glycolytic flux. The increased PFK-2 activity would lead to an increase in F2,6BP levels, thereby increasing glycolysis via activation of PFK-1. The regulation of liver and heart isoforms exemplifies how differential regulation of specific isoforms can serve different roles, depending on the needs of various tissues in which they are expressed. This further illuminates the significant role PFK-2/FBPase-2 plays in appropriately regulating carbohydrate metabolism.

The role of PFK-2/FBPase-2 in pancreatic beta cells is much less clear. The isoform in these cells is similar to that found in the heart. The pancreatic beta cells release insulin in response to glucose. It has long been recognized that this release is at least in part regulated by the metabolism of glucose. The first regulated step in the metabolism of glucose is its phosphorylation catalyzed by an isoform of the enzyme hexokinase. Interestingly, the pancreatic isoform of PFK-2/FBPase-2 may exert an effect on insulin secretion by its effect on the hexokinase in these cells, also called an isoform of glucokinase. This effect may be independent of either PFK-2/FBPase-2 enzymatic activities establishing yet another mechanism through which this diverse and interesting enzyme affects glucose metabolism and homeostasis.

See also: **Metabolism Vitamins and Hormones:** Gluconeogenesis; Glycolysis Overview; **Protein/Enzyme Structure Function and Degradation:** Enzyme Reaction Mechanisms: Stereochemistry.

Further Reading

- Arden C, Hampson LJ, Huang GC, et al. (2008) A role for PFK-2/FBPase-2, as distinct from fructose 2,6-bisphosphate, in regulation of insulin secretion in pancreatic beta-cells. *Biochemical Journal* 411(1): 41–51.
- Baltrusch S, Langer S, Massa L, Tiedge M, and Lenzen S (2006) Improved metabolic stimulus for glucose-induced insulin secretion through GK and PFK-2/FBPase-2 coexpression in insulin-producing RINm5F cells. *Endocrinology* 147(12): 5768–5776.
- Kurland IJ and Pilkis SJ (1995) Covalent control of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase: Insights into autoregulation of a bifunctional enzyme. *Protein Science* 4(6): 1023–1037.
- Okar DA, Manzano A, Navarro-Sabate A, et al. (2001) PFK-2/FBPase-2: Maker and breaker of the essential biofactor fructose-2,6-bisphosphate. *Trends in Biochemical Sciences* 26(1): 30–35.
- Pilkis SJ, Claus TH, Kurland IJ, and Lange AJ (1995) 6-Phosphofructo-2-kinase/fructose-2,6-bisphosphatase: A metabolic signaling enzyme. *Annual Review of Biochemistry* 64: 799–835.
- Pilkis SJ, el-Maghrabi MR, and Claus TH (1988) Hormonal regulation of hepatic gluconeogenesis and glycolysis. *Annual Review of Biochemistry* 57: 755–783.

Phosphoinositide 4- and 5-Kinases and Phosphatases

S F Bairstow, Baxter Healthcare Corporation, Round Lake, IL, USA

M W Bunce, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

R A Anderson, University of Wisconsin, Madison, WI, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Inositol A six-carbon cyclic alcohol with hydroxyl residues present at each carbon.

Kinase An enzyme that catalyzes the transfer of a phosphate group from adenosine triphosphate to a second substrate.

Phosphatase An enzyme that hydrolyzes a phosphate group from a substrate.

Phosphoinositide A phospholipid molecule composed of inositol and diacylglycerol.

Phosphoinositides are one of the many phospholipid families found within membrane bilayers of eukaryotic cells. Site-specific phosphorylation of phosphoinositides by lipid kinases results in the generation of multiple phosphatidylinositol phosphate isomers. These mono- or poly-phosphorylated isomers either can be further modified by additional enzymes or can directly modulate numerous cellular functions. Because of the potential for phosphoinositides to alter a variety of cellular processes, their synthesis must be strictly regulated. A number of phosphoinositide kinases and phosphatases have been identified and characterized. By adding or removing phosphate groups at specific positions along the inositol ring, these kinases and phosphatases dictate the signaling capacity of phosphoinositides.

Phosphoinositides

Structure

The basic structure of phosphoinositides consists of a six-carbon inositol ring linked by a phosphodiester bond to diacylglycerol (DAG) (Figure 1). The two fatty acid tails of the DAG moiety anchor phosphoinositides into the various membrane structures throughout eukaryotic cells, such as the endoplasmic reticulum (ER), Golgi apparatus, and plasma membrane. *In vivo*, phosphoinositides can be phosphorylated at the 3-, 4-, and 5-hydroxyl positions along the inositol ring in all possible combinations.

Nomenclature

Phosphatidylinositol (PI) is the initial substrate for a number of phosphoinositide kinases that phosphorylate the hydroxyl groups along the inositol ring. Mono- and polyphosphorylated phosphoinositides are named with respect to the positions of the inositol ring that are phosphorylated. Phosphorylation of PI at the 4-hydroxyl position generates PI 4-phosphate (PI4P); a subsequent phosphorylation at the 5-hydroxyl position generates PI 4,5-bisphosphate (PI(4,5)P₂). PI can be singly or severally phosphorylated at the 3-, 4-, and 5-hydroxyl positions to yield PI mono-, bis-, or trisphosphate molecules.

PI 4-Kinases

Subfamilies

PI 4-kinases use PI as a substrate to generate PI4P. PI 4-kinases are classified as either type II or type III kinases based on their structural and kinetic properties. Two mammalian isoforms termed α and β have been identified for each PI 4-kinase subfamily. Type III PI 4-kinases are structurally and kinetically similar to PI 3-kinases. PI 3-kinases and type III PI 4-kinases have homologous catalytic domains, and type III PI 4-kinases are even inhibited by known PI 3-kinase inhibitors, such as wortmannin. Type II PI 4-kinases do not share these structural or catalytic similarities with PI 3-kinases.

Cellular Functions

The higher-order species such as PI(4,5)P₂ and PI(3,4,5)P₃ have received significant attention, whereas PI4P has typically been thought of simply as a metabolic intermediate in the synthesis of other phosphoinositides. However, a number of functional studies have demonstrated important roles for PI 4-kinases *in vivo*. Type II PI4-kinases have been linked to roles in both secretion and endocytosis. Type III PI4-kinases have been most thoroughly characterized in the budding yeast *Saccharomyces cerevisiae*. Stt4 and Pik1, the yeast homologs of mammalian type III α - and β -isoforms, respectively, are targeted to distinct subcellular locations and perform nonredundant functions. Pik1 is targeted to the nucleus and Golgi. The nuclear function of Pik1 is unknown, but several studies have indicated that Pik1 is required for normal secretion from the Golgi. PI 4-kinase β , the mammalian homolog to Pik1, has also been implicated in Golgi function. Stt4 localizes primarily to the plasma membrane and ER and has been linked to actin cytoskeleton organization and cell wall stability. The PI4P generated specifically by Stt4 can be further phosphorylated by Mss4, the yeast homolog of type I PI phosphate kinases, or dephosphorylated by the phosphoinositide phosphatase Sac1. The interplay between Stt4, Mss4, and Sac1 controls actin dynamics in *S. cerevisiae*. As mentioned above, Pik1 and Stt4 are functionally nonredundant. A loss of activity in either PI 4-kinase causes a unique phenotype that cannot be rescued by increasing the activity of the other kinase. These observations suggest that Pik1 and Stt4 each generate

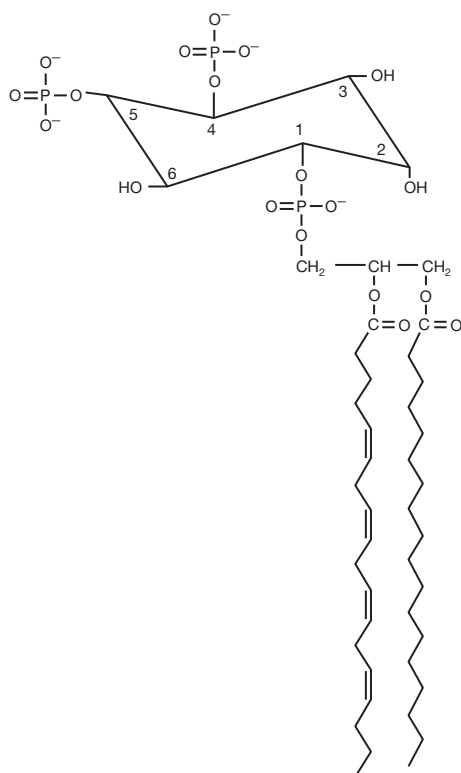


Figure 1 Phosphoinositide structure. Phosphoinositides are comprised of a six-carbon inositol ring linked by a phosphodiester bond to diacylglycerol. The inositol ring is numbered counterclockwise starting at the phosphodiester linkage. *In vivo*, phosphoinositides are phosphorylated at the 3, 4, and 5 positions of the inositol ring; phosphatidylinositol 4,5-bisphosphate (PI(4,5)P₂) is pictured.

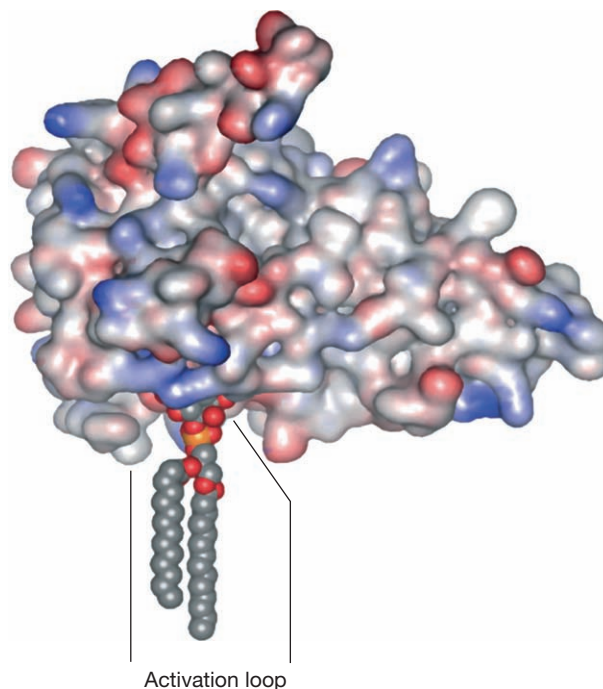
discrete pools of PI4P at different subcellular locations, and that these pools of PI4P are utilized by unique sets of effectors.

PI Phosphate Kinases

PI phosphate kinases (PIP kinases) generate PI bisphosphate (PIP₂) from PI monophosphate isomers. PIP kinases are classified into type I, II, or III subfamilies based on their structural and kinetic properties. Members of all three subfamilies have highly conserved kinase domains toward the C-terminus of the proteins, with more variable flanking sequences. Each PIP kinase subfamily has unique substrate specificity dictated by a region at the C-terminal end of the kinase domain named the activation loop. The activation loop is nearly identical for different isoforms within a subfamily but divergent between subfamilies (Figure 2).

Type I PIP Kinases

Type I PIP kinases are PI4P 5-kinases, that is, they use PI4P as a substrate to generate PI(4,5)P₂. Three mammalian type I PIP kinase isoforms termed α , β , and γ have been identified. These isoforms share a highly conserved kinase domain with variable amino and carboxy termini. The variable regions regulate enzymatic activity and direct the subcellular localization of each



Y	R	F	V	K	K	L	E	H	S	W	K	A	L	V	H	D	G	D	T	V	.	.	.	.	.	S	Hu type I α
Y	R	L	M	K	K	L	E	H	S	W	K	A	L	V	Y	D	G	D	T	V	.	.	.	.	.	S	Hu type I β
Y	R	F	I	K	K	L	E	H	T	W	K	A	L	V	H	D	G	D	T	V	.	.	.	.	.	S	Hu type I γ
Y	R	L	L	K	K	M	E	H	T	W	K	A	I	L	H	D	G	D	T	I	.	.	.	.	.	S	Skittles
Y	S	V	M	K	K	L	E	T	F	W	K	S	L	R	H	D	T	K	L	V	.	.	.	.	.	S	Sp Mss4p
Y	S	V	V	K	K	V	E	H	L	W	K	G	I	N	H	S	.	D	S	V	I	.	.	.	.	S	Sc Mss4p
Y	D	A	K	K	K	A	A	H	A	A	K	T	V	K	H	G	A	G	A	E	I	.	.	.	.	S	Hu type II α
Y	D	T	K	K	K	A	A	H	A	A	K	T	V	K	H	G	A	G	A	E	I	.	.	.	.	S	Hu type II β
Y	D	A	K	K	K	A	A	H	A	A	K	T	V	K	H	G	A	G	A	E	I	.	.	.	.	S	Hu type II γ
Y	G	V	K	K	Q	A	A	K	A	A	K	T	V	K	Y	G	S	N	V	D	G	I	.	.	.	S	Dm type II
F	T	W	D	K	K	L	E	S	W	V	K	E	K	G	L	V	G	G	A	S	V	I	K	Q	P	T	Sc Fab1p
Y	T	W	D	K	K	L	E	S	W	V	K	E	K	G	L	V	G	R	G	.	.	.	P	E	P	T	Sp Fab1p
Y	T	L	D	K	K	L	E	S	W	V	K	I	V	A	I	P	G	A	.	.	.	H	L	P	T	Ce Fab1p	

Figure 2 PIP kinase structure and the activation loop. The crystal structure for the human type II β PIP kinase isoform is shown above with PI5P modeled into the active site. Although the activation loop was not resolved in the crystal structure, its position relative to the active site is illustrated. A sequence alignment of type I, type II, and type III PIP kinase activation loops is shown below the crystal structure. The activation loop is highly conserved among members of each subfamily but divergent between subfamilies. Hu, human; Sp, *S. pombe*; Sc, *S. cerevisiae*; Dm, *D. melanogaster*; Ce, *C. elegans*.

isoform through unique protein–protein interactions. These three isoforms also have alternatively spliced forms, further enhancing the diversity of targeting for these kinases. Type I PIP kinases are best known for their ability to alter the organization of the actin cytoskeleton. PI(4,5)P₂ regulates the activity of many proteins that process actin at the plasma membrane. Mss4, the yeast homolog of mammalian type I PIP kinases, is required for proper actin cytoskeleton organization and cell wall integrity. Yeast lacking Mss4 activity can be rescued by expressing mammalian type I PIP kinase isoforms, illustrating that type I PIP kinases are functionally conserved in eukaryotes. Mammalian type I PIP kinases localize to sites of cortical actin dynamics, such as lamellipodia and focal adhesions, and

modulate the assembly of these structures in a PI(4,5)P₂-dependent manner. The type I γ isoform is targeted to focal adhesions by an association with the protein talin and is tyrosine phosphorylated in a focal-adhesion-kinase (FAK)-dependent manner. Kinase-dead PIP kinase I γ mutants block focal adhesion assembly, reinforcing the requirement for PI(4,5)P₂ in focal adhesion signaling. Additional roles for the I γ isoform have also been recently demonstrated in clathrin-mediated endocytosis and trafficking and assembly of E-cadherin (a transmembrane cell-cell adhesion protein) adherens junctions. More recently, a splice variant of the I α isoform has been shown to have an important role in the regulation of messenger RNA (mRNA) processing in the nucleus.

Type II PIP Kinases

Type II PIP kinases are PI5P 4-kinases, generating PI(4,5)P₂ from PI5P in a catalytically distinct reaction from type I PIP kinases. Like type I PIP kinases, the type II subfamily is comprised of α -, β -, and γ -isoforms, with the II α isoform being the most catalytically active. In contrast to the well-described physiological functions for type I PIP kinases, *in vivo* functions for type II PIP kinases remain largely undefined. Although type I and type II PIP kinases both produce PI(4,5)P₂, the two subfamilies are functionally nonredundant. No homologs for mammalian type II PIP kinases have been identified in *S. cerevisiae*, so type II PIP kinases may have specialized functions in higher eukaryotes. One such function could be to modulate receptor-mediated signaling, as the type II β isoform has been identified in a complex with the p55 TNF α transmembrane receptor. Alternatively, the II γ isoform is thought to be involved in regulation of vesicular trafficking from the Golgi apparatus. Interestingly, the II α and II β isoforms have been shown to heterodimerize, resulting in targeting of the II α isoform to both the cytoplasm and the nucleus. Recent work has identified a novel role for the II β isoform in the regulation of nuclear stress response signaling pathways. Downregulation of a pool of nuclear II β following cellular stress causes the accumulation of PI5P in the nucleus, which induces pro-apoptotic signaling by activating both the tumor suppressor p53 and p38 MAP kinase stress response pathways. Interestingly, PI5P, not PI(4,5)P₂, appears to be the potent signaling molecule in this system, and a key function for nuclear II β under nonstressed conditions appears to be the removal of nuclear PI5P via its conversion to PI(4,5)P₂.

Type III PIP Kinases

Type III PIP kinases are PI3P 5-kinases, generating PI(3,5)P₂ from PI3P. The *S. cerevisiae* protein Fab1 was the first type III PIP kinase to be identified. In yeast, Fab1 and PI(3,5)P₂ are involved in vacuole morphology, membrane trafficking, and stress response. Type III PIP kinases were not initially identified in mammalian cells, and *in vitro* studies suggested that type I isoforms can generate PI(3,5)P₂ and thus may perform similar functions to those of Fab1 in yeast. More recently, however, proteins homologous to ScFab1 have been identified in mammalian cells. PIKfyve (phosphoinositide kinase for five position containing a Fyve finger) is one example. These mammalian enzymes have similar *in vivo* substrate specificity to

the *S. cerevisiae* Fab1 enzyme and appear to regulate membrane trafficking in a manner comparable to ScFab1. Surprisingly, PIKfyve has been shown to phosphorylate PI to form PI5P *in vitro*, providing a putative mechanism for the direct synthesis of PI5P *in vivo* (as opposed to generation via dephosphorylation of higher order phosphoinositides).

PI Phosphatases

PI phosphatases are enzymes that can selectively dephosphorylate both lipid phosphoinositides and soluble inositol polyphosphates. As a family, they can be subdivided by their catalytic properties and substrate specificity.

Phosphoinositide 4-Phosphatases

The phosphoinositide 4-phosphatases (inositol polyphosphate 4-phosphatases) are a family of enzymes that dephosphorylate the 4-position of the inositol ring in a magnesium-independent manner. The members of this family can utilize PI(3,4)P₂, Ins(3,4)P₂, and Ins(1,3,4)P₃ as substrates *in vitro*, generating PI3P, Ins3P, and Ins(1,3)P₂, respectively (though Ins(3,4)P₂ and Ins(1,3,4)P₃ are dephosphorylated to a much lesser extent). Two genes have been identified in humans corresponding to type I and type II 4-phosphatases. These genes can be alternatively spliced, resulting in expression of several variant phosphatases (type I α and I β and type II α and II β isoforms, with the β isoforms expressed with hydrophobic C termini). Both types share a 37% total sequence identity and a conserved CKSAKDRT consensus sequence. This C-X₅-R consensus sequence is consistent with other magnesium-independent phosphatases.

Cellular Functions of 4-Phosphatases

The PI-3 kinase (PI3K) signaling pathway impacts cell growth, glucose metabolism, and apoptosis. Activation of the PI3K pathway results in the generation of PI(3,4,5)P₃ and PI(3,4)P₂, which stimulate downstream effectors such as protein kinase B/Akt and PI3K-dependent kinase 1. Phosphoinositide 4-phosphatases are thought to participate in regulation of the PI3K pathway by converting PI(3,4)P₂ to PI3P, thus attenuating the PI3K signal and subsequently deactivating PI3K effectors. An association identified between a PI 4-phosphatase and the PI3K p85 regulatory subunit supports this putative function for PI 4-phosphatases. Recently, the type I α isoform has been implicated in regulation of megakaryocyte and fibroblast proliferation, in addition to modulating PI3K signaling in endosomes and at the plasma membrane.

Phosphoinositide 5-Phosphatases

Phosphoinositide 5-phosphatases (inositol polyphosphate 5-phosphatases) dephosphorylate the 5-position of the inositol ring in a magnesium-dependent manner. This family collectively can utilize four substrates: PI(4,5)P₂, PI(3,4,5)P₃, Ins(1,4,5)P₃, and Ins(1,3,4,5)P₄. PI 5-phosphatases contain a bipartite consensus sequence, (F/I)WXGD_XN(F/Y)R, followed several amino acids downstream by (R/N)XP(S/A) (W/Y) (C/T)DR(I/V) (L/I). PI 5-phosphatases are classified into four groups based on their substrate specificity.

Group I 5-Phosphatases

Group I 5-phosphatases use $\text{Ins}(1,4,5)\text{P}_3$ and $\text{Ins}(1,3,4,5)\text{P}_4$ as substrates. Type I 5-phosphatase, one member of this group, is thought to play a role in the regulation of cellular calcium levels. $\text{Ins}(1,4,5)\text{P}_3$ (IP_3) is a pivotal regulator of cellular calcium levels and is generated by the cleavage of PIP_2 by phospholipase C. The resulting IP_3 binds to the IP_3 receptor, triggering a release of calcium ions from intercellular stores. Type I 5-phosphatase decreases this signal by dephosphorylating IP_3 into $\text{Ins}(1,4)\text{P}_2$.

Group II 5-Phosphatases

Members of group II 5-phosphatases can utilize all four of the substrates mentioned above. The synaptojanins, type II 5-phosphatase (INPP5B), and the oculocerebrorenal syndrome of Lowe (OCRL) gene product, OCRL-1, are all members of this group. In total, nine variants have been described in humans. OCRL is an X-linked disorder resulting in mental retardation, renal dysfunction, and congenital cataracts. Although the pathogenesis for these symptoms is not currently known, OCRL-1 dysfunction results in an increase in $\text{PI}(4,5)\text{P}_2$ levels, as observed in cells cultured from individuals with this disorder. Interestingly, OCRL-1 shares 51% identity with type II 5-phosphatase. However, this enzyme cannot compensate for the lack of OCRL-1 activity, possibly due to a difference in substrate specificity or tissue localization. OCRL-1 does contain a Rho-GAP (guanosine triphosphatase-activating protein) domain, which facilitates its targeting to the *trans*-Golgi network via interaction with the small G-protein Rac. The synaptojanins, on the other hand, have functions distinct from OCRL-1 and type II 5-phosphatase. These proteins contain N-terminal SAC1 domains, homologous to the yeast Sac1p phosphoinositide phosphatase. Synaptojanin 1 has been shown to function in recycling of vesicles at neuronal synapses and it has also been linked as a candidate gene to bipolar disorder. Synaptojanin 2, on the other hand, has been shown to be required for clathrin-coated pit formation and has been shown to be involved in cell migration via regulation of lamellipodia formation. This protein is alternatively spliced and some variants have been shown to have functions overlapping that of synaptojanin 1, while other variants have been localized to distinct cellular compartments, such as the outer mitochondrial membrane.

Group III 5-Phosphatases

Group III 5-phosphatases are active toward $\text{PI}(3,4,5)\text{P}_3$ and $\text{Ins}(1,3,4,5)\text{P}_4$, which have a 3-position phosphate group. Src homology domain containing inositol phosphatases (SHIPs), SHIP1 and SHIP2, are important members of this group. These proteins share 50% sequence identity and similar

substrate specificity, but differ in their tissue expression levels. SHIP1 is primarily expressed in hematopoietic tissues while SHIP2 is expressed in nonhematopoietic tissues. Recent work has demonstrated an importance of SHIPs in signal transduction. SHIP2 has been shown to associate with specific tyrosine kinase receptors, such as the EGF receptor. This places the phosphatase in an ideal location to attenuate the increased $\text{PI}(3,4,5)\text{P}_3$ and $\text{PI}(3,4)\text{P}_2$ levels generated by PI3K, which is also recruited to these receptors. SHIP2 mouse knockout experiments support SHIP2's role in PI3K signal termination and also suggest a link between SHIP2 activity and insulin resistance, a primary factor in type 2 diabetes.

Group IV 5-Phosphatases

Group IV phosphatases are the least characterized of the four groups. One member, phosphoinositide 5-phosphatase type IV (INPP5E) in humans or pharbin in rats, has been identified. This phosphatase utilizes $\text{PI}(3,4,5)\text{P}_3$ as its primary substrate, and has also been shown to dephosphorylate $\text{PI}(4,5)\text{P}_2$ and $\text{PI}(3,5)\text{P}_2$. Expression levels are highest in the brain for both humans and mice, with lower expression in other tissues. This phosphatase has been implicated in glucose metabolism and control of food intake.

See also: [Metabolism Vitamins and Hormones: Diabetes;](#)
[Signaling: Inositol Phosphate Kinases and Phosphatases;](#)
[Phosphatidylinositol Bisphosphate and Trisphosphate;](#)
[Phosphoinositide-Dependent Protein Kinases.](#)

Further Reading

- Balla A and Balla T (2006) Phosphatidylinositol 4-kinases: Old enzymes with emerging functions. *Trends in Cell Biology* 7: 351–361.
- Barlow CA, Laishram RS, and Anderson RA (2010) Nuclear phosphoinositides: A signaling enigma wrapped in a compartmental conundrum. *Trends in Cell Biology* 20: 25–35.
- Blero D, Payrastre B, Schurmans S, and Erneux C (2007) Phosphoinositide phosphatases in a network of signaling reactions. *European Journal of Physiology* 455: 31–44.
- Clarke JH, Wang M, and Irvine RF (2010) Localization, regulation, and function of type II phosphatidylinositol 5-phosphate 4-kinases. *Advances in Enzyme Regulation* 50: 12–18.
- Heck JN, Mellman DL, Ling K, Sun Y, and Wagoner MP (2007) A conspicuous connection: Structure defines function for the phosphatidylinositol-phosphate kinase family. *Critical Reviews in Biochemistry and Molecular Biology* 42: 15–39.
- Kwiatkowska K (2010) One lipid, multiple functions: How various pools of $\text{PI}(4,5)\text{P}_2$ are created in the plasma membrane. *Cellular and Molecular Life Sciences* 67: 3927–3946.
- Majerus PW and York JD (2009) Phosphoinositide phosphatases and disease. *Journal of Lipid Research* 50(supplement): S249–S254.
- Schiil NJ and Anderson RA (2009) Out, in and back again: $\text{PtdIns}(4,5)\text{P}_2$ regulates cadherin trafficking in epithelial morphogenesis. *Biochemical Journal* 418: 247–260.

Phosphoinositide-Dependent Protein Kinases

N R Leslie and C P Downes, University of Dundee, Dundee, UK

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 292–296, © 2004, Elsevier Inc.

Glossary

Kinase An enzyme that catalyzes the transfer of a phosphate group from ATP onto another molecule.

Phosphatase An enzyme that catalyzes the hydrolytic removal of phosphate from another molecule and its release as free phosphate.

Phosphoinositide A phospholipid in which the polar head group is inositol or a phosphorylated derivative of inositol.

The cells of multicellular organisms respond to external stimuli, such as growth factors, cytokines, neurotransmitters, and many hormones, through regulated changes in the levels of cellular signaling molecules, termed second messengers. The inositol phospholipids (collectively known as phosphoinositides) are constituents of cell membranes that play important roles as second messengers in signal transduction and membrane trafficking. Many of the downstream effects of phosphoinositide signaling are mediated through reversible protein phosphorylation and in this article the principal focus is upon protein kinases that function as downstream targets of the phosphoinositide 3-kinase (PI 3-kinase) lipid signaling pathway.

Phosphoinositide Signaling – An Overview

Phosphoinositides are a small family of differentially phosphorylated membrane-localized lipid second messengers. These lipids comprise the hydrophilic cyclic polyol, myo-inositol (Ins), linked via a diester phosphate in the 1-position to diacylglycerol, the hydrophobic membrane anchor. Several families of lipid kinases and phosphatases selectively insert and remove monoester phosphates in three of the remaining positions of the inositol ring, generating the eight currently known phosphoinositide species. This article focuses on some of the most intensively studied phosphoinositide signaling pathways initiated when growth factors, cytokines or insulin, and related hormones trigger the activation of one or more of the type I PI 3-kinases. These enzymes phosphorylate a small proportion of the cellular phosphatidylinositol 4,5 bisphosphate (PtdIns(4,5)P₂) to generate PtdIns(3,4,5)P₃ (often abbreviated to PIP₃). The latter is further metabolized by 5-phosphatases to PtdIns(3,4)P₂ or is returned to the PtdIns(4,5)P₂ pool by the tumor-suppressor lipid phosphatase PTEN. PtdIns(3,4,5)P₃ activates downstream signaling through proteins which possess a phosphoinositide binding domain capable of distinguishing this target lipid within a membrane that contains at least 100-fold higher levels of the parent lipid PtdIns(4,5)P₂. Several modular protein domain families have now been identified that include members able to bind to specific phosphoinositides, such as pleckstrin homology (PH), Phox homology (PX), FYVE, and ENTH domains.

Other domains, such as FERM, and PDZ domains may also function as phosphoinositide-binding motifs, although evidence for their selectivity between different phosphoinositides is currently lacking. In almost all cases presently identified, the PtdIns(3,4,5)P₃-specific-binding motifs mediating PI 3-kinase-dependent signaling are members of the PH domain family.

Protein Kinase B/Akt

Since the 1990s, it has become evident not only that PI 3-kinase signaling plays a pivotal role in regulating many diverse and significant cellular processes, including cellular proliferation, survival, size, and motility, but also that many of the downstream effectors of PtdIns(3,4,5)P₃ are protein kinases. The best evidence for regulation by PtdIns(3,4,5)P₃ exists for three classes of kinases that carry PH domains able, specifically, to bind this lipid: the Ser/Thr kinases, PKB (or Akt), PDK1, and the Tec family Tyr kinases. Space limitations prevent us from addressing enzymes such as GSK3 or PAK, which can clearly be regulated by the PI 3-kinase pathway, though not by directly binding to PtdIns(3,4,5)P₃, and these enzymes can also be regulated by PI 3-kinase-independent pathways. The PI 3-kinase signaling pathway has been dissected extensively using biochemical and genetic approaches. Significantly, the stimulated production of PtdIns(3,4,5)P₃ by PI 3-kinase enzymes is conserved in species from slime molds, through nematodes and flies to vertebrates. In species of all of these groups, genetic studies indicate that a principal (though not exclusive) cellular mediator of this signaling pathway is protein kinase B (PKB).

PKB, also known as Akt, has a C-terminal serine/threonine kinase domain and an N-terminal PH domain that binds with high specificity to PtdIns(3,4,5)P₃ and PtdIns(3,4)P₂. The enzyme has been shown to be activated by phosphorylation of two key residues, threonine 308, in the activation loop (or T-loop) of the kinase domain, and serine 473 in the C-terminal tail. These phosphorylation events in turn are believed to alter the conformation of PKB, greatly enhancing its activity, and although much of this work has focused on PKB, it now appears that a number of related kinases may be regulated by analogous mechanisms. Although PtdIns(3,4,5)P₃ and

PI(3,4)P₂ do not directly enhance the activity of PKB *in vitro*, they play a critical role in the phosphorylation and activation of PKB, by recruiting PKB to the plasma membrane where it is activated, and by opening the conformation of the protein to allow phosphorylation. Thus, it seems that in a normal cellular context, the activity state of PKB is largely governed by cellular levels of PtdIns(3,4,5)P₃ and PI(3,4)P₂.

PDK1 and the Regulation of the AGC Kinases

3-Phosphoinositide-dependent kinase 1 (PDK1) has been identified as the activating kinase that phosphorylates PKB in a PI 3-kinase dependent process at Thr 308 *in vitro* and *in vivo* (see Figure 1). This protein serine/threonine kinase also has a PH domain that specifically binds PtdIns(3,4,5)P₃ and PI(3,4)P₂ and is found both in the cytosol and on the plasma membrane. Significantly, after its initial identification as a PKB kinase, subsequent work has shown that it also phosphorylates and activates many other related kinases. A point of current controversy is whether the function of PDK1 itself is regulated through mechanisms such as phosphorylation or translocation to the plasma membrane. It has been proposed that PDK1 translocates to the plasma membrane in response to stimulated rises in cellular PtdIns(3,4,5)P₃ levels, as does PKB. However, this conclusion is not supported by other work, and as PDK1 has an affinity approximately 10-fold greater for these lipids than does PKB, an alternative idea is that a pool of cellular PDK1 may be constitutively associated with the plasma membrane and that its activity is not normally regulated by changes in the levels of these lipids. It has also been proposed that PDK1 is activated in stimulated cells by tyrosine

phosphorylation, possibly by a Src family kinase, although, once more, other work argues against this, and phosphorylation has not been observed in response to physiological stimuli, or in cells expressing physiological levels of the relevant kinases. Most of this work on the regulation of PDK1 function would suggest that if it is to be viewed as a phosphoinositide-regulated kinase, this regulation occurs through direct or indirect effects of PtdIns(3,4,5)P₃ on its substrates, rather than on PDK1 itself.

PKB exists as three isoforms in human cells (PKB α , β , and γ), but is itself a part of a diverse structurally related family of protein kinases, which are phosphorylated by PDK1 at a residue in the T-loop of each kinase. Although this PDK1-mediated phosphorylation plays a role in the activation of each of these kinases, in contrast to PKB, the phosphorylation of other family members is not directly responsive to cellular PtdIns(3,4,5)P₃ levels, and alternative signaling inputs seem to be required to activate these other kinases strongly. To allow efficient activation, most of these related kinases require the interaction of a hydrophobic motif, positioned C-terminal to their kinase domain, with a hydrophobic pocket on PDK1. In some cases, for example, S6K1 and SGK, this hydrophobic motif contains a serine or threonine that must be phosphorylated to allow interaction with and activation by PDK1. In these cases, it appears that this phosphorylation is in turn mediated by an unknown PI 3-kinase-dependent kinase. In contrast, in place of this serine or threonine residue, PKC ζ and PRK2 have an acidic residue, which appears to mediate phosphorylation-independent interaction with, and activation by, PDK1. PKC ζ , and probably the other atypical PKC kinases, PKC ι and PKC λ , are also unusual in that there is some evidence for the direct activation of the kinase through interaction with PtdIns(3,4,5)P₃.

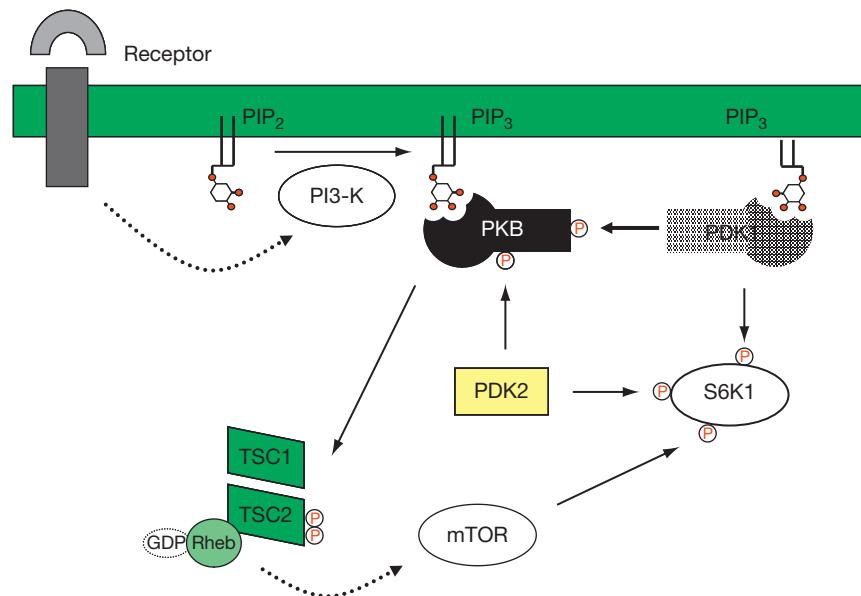


Figure 1 Mechanism of activation of PKB, mTOR, and S6K1 by PI 3-kinase. A model is shown for the activation of downstream kinases by a transmembrane receptor activating PI 3-kinase. Molecules are shown in the stimulated conformation. In unstimulated cells, PI 3-kinase is less active, PIP₃ levels are low, PKB is cytosolic, and along with TSC2 and S6K1, is unphosphorylated and inactive. Active TSC2 promotes the conversion of GTP to GDP by Rheb, although the mechanism by which this leads to the activation of mTOR signaling is unclear. Direct phosphorylation events are shown by bold arrows and indirect mechanisms of activation are represented by dotted arrows. Phosphates are represented by a circled letter P.

However, these kinases lack a PH domain or other identified PtdIns(3,4,5)P₃-binding motif, and the mechanism and physiological relevance of this apparent activation is unclear.

PDK2 – The Unidentified Hydrophobic Motif Kinase

The term PDK2 was used originally to describe the unidentified kinase responsible for the phosphorylation of PKB at the hydrophobic motif serine 473. Since then, although several candidates have been proposed, such as ILK, MAPKAP-K2, NEK6, and both PKB and PDK1 themselves, the identity of PDK2 is still not clear. What is clear, however, is that most of the AGC family kinases become phosphorylated at a serine or threonine residue in a hydrophobic motif C-terminal to the kinase domain, and that this phosphorylation plays a critical role in the activation of each enzyme. It is quite possible that different PDK2s exist for different AGC family kinases; it appears that a substantial component of the PI 3-kinase-dependent activation of kinases, such as SGK and S6K1, may be mediated by an as-yet unknown phosphoinositide-dependent hydrophobic motif kinase. As a consequence of the mechanisms of activation of this family of kinases, increases in cellular PtdIns(3,4,5)P₃ levels seem sufficient to activate PKB, since PtdIns(3,4,5)P₃ and PH domain-dependent colocalization of PKB and PDK1 appears to be the principal mechanism of regulation of this kinase. In contrast, most other PDK1-regulated kinases seem to require regulated docking with PDK1 through a protein–protein interaction to mediate activation in the cytosol, and it is this docking step that may be regulated by different upstream pathways, some of which may themselves involve phosphoinositide-dependent kinases.

Much of the interest in PKB stems from its role in mediating many of the effects of PI 3-kinase activation. Evidence now indicates that a substantial portion of the effects of elevated PtdIns(3,4,5)P₃ levels on cell-cycle progression, survival, growth, and insulin responses are mimicked by activation of PKB alone. In contrast, most PI 3-kinase-dependent cytoskeletal regulation appears to be PKB independent, but may rather result from the interaction of PtdIns(3,4,5)P₃ with the PH domains of guanine–nucleotide-exchange factors for Rho family GTPases. These findings go some way to explain why isoforms of PKB have been found to be overexpressed in a range of human tumor types, and why activation of PKB activity appears to be an even more common feature of tumors.

Regulators of Cellular Growth – mTOR and S6K1

Work addressing the stimulation of cell growth by PtdIns(3,4,5)P₃ signaling implicates two protein kinases as key players in this process, the mammalian target of rapamycin (mTOR), and ribosomal protein S6 kinase 1 (S6K1, also known as p70S6K). While S6K1 seems to regulate translation through several mechanisms, mTOR has been proposed to play a central role in cellular growth control, as its activity is responsive to nutrients and ATP and is able to regulate a range of downstream processes related to cell growth, including transcription, tRNA

and ribosome synthesis, autophagy, and nutrient transporter function. Although recent work has shed significant light on the regulation of both of these kinases, particularly the inputs from PtdIns(3,4,5)P₃ signaling, their regulation appears complex and many questions remain unanswered. mTOR activity now appears to be inhibited in nongrowing cells by the small GTPase Rheb, and a protein complex of two proteins, TSC1 and TSC2, the latter of which can function as a GTPase-activating protein for Rheb. Upon activation of PI 3-kinase and PKB, TSC2 is phosphorylated by PKB, Rheb becomes inactive (in the GDP bound state) causing activation of mTOR. S6K1 is then believed to be stimulated through phosphorylation by activated mTOR (in addition to other activating inputs such as PDK1). This model is represented in [Figure 1](#).

The study of S6K1, and many other kinases that are affected by PtdIns(3,4,5)P₃ signaling, with the exception of PKB, highlights a possible role for the level of basal PtdIns(3,4,5)P₃ present in unstimulated growing cells in culture. For example, S6K1 can be potently activated by G-protein-coupled receptors under conditions where PtdIns(3,4,5)P₃ levels do not increase. Thus, PtdIns(3,4,5)P₃ can play a permissive rather than stimulatory role in G-protein-coupled receptor-mediated S6K1 activation. Significantly, a similar role for basal PtdIns(3,4,5)P₃ has been proposed in the regulation of other molecules, such as Ras and PKCδ.

One important piece of evidence that cell-growth responses stimulated by mTOR and S6K1 are downstream effectors specifically of PtdIns(3,4,5)P₃ is the sensitivity of PTEN null ES cells, MEFs, and tumors to the immunosuppressant rapamycin. Deletion of the PtdIns(3,4,5)P₃ specific phosphatase, PTEN, elevates PtdIns(3,4,5)P₃ levels and activates downstream PI 3-kinase-dependent signaling without stimulating other pathways. Since these cells are unusually sensitive to growth inhibition by the mTOR (and thus S6K1) inhibitor rapamycin, this suggests that these pathways are significantly activated by increased PtdIns(3,4,5)P₃ levels, and that this is important for cell growth and tumor development.

Tec Family Kinases

The Tec family of tyrosine kinases has been implicated in signaling by lymphocyte antigen receptors. These kinases have an N-terminal PtdIns(3,4,5)P₃ binding PH domain and a C-terminal tyrosine kinase domain. They also contain a proline-rich region, and SH3 and SH2 domains in the central region of the proteins. It is believed that their activation requires recruitment to the plasma membrane by interaction of PtdIns(3,4,5)P₃ with the PH domain, and subsequent activation by members of the Src family of tyrosine kinases through phosphorylation of the kinase domain activation loop. These kinases are rather restricted in the range of tissues where they are expressed, most being hematopoietic specific, and appear to play a role most significantly in lymphocyte signaling. This picture has stemmed from the original identification of the first family member, Bruton's tyrosine kinase (Btk), as the gene mutated in a human immunodeficiency. The family is now known to include Btk, Tec, Itk, Bmx, and the slightly divergent Rlk. Interestingly, in Rlk, the PH domain has been

replaced by an N-terminal palmitoylation and membrane targeting signal, leading to PI 3-kinase-independent activation.

Perspectives

This article has focused largely upon well-studied kinases that are regulated through $\text{PtdIns}(3,4,5)\text{P}_3$ and $\text{PI}(3,4)\text{P}_2$ and the PI 3-kinase signaling pathway, and their roles in regulating cellular proliferation, survival, and growth. However, evidence for the regulation of kinases in these and other cellular signaling pathways by other phosphoinositides is emerging. For example, serum and glucocorticoid-inducible protein kinase 3 (SGK3 or CISK) is localized to endosomes through a PX-domain-mediated interaction, probably via $\text{PtdIns}(3)\text{P}$. Although tethering to endosomes appears to be required for CISK to function, there is also evidence that its catalytic activity is controlled through stimulated type I PI 3-kinases, and hence via a mechanism that also involves $\text{PtdIns}(3,4,5)\text{P}_3$. Since the human genome contains ~48 genes encoding proteins, many unstudied, with both a protein kinase domain, and a domain recognized as a potential phosphoinositide-binding domain (42 with PH domains and 6 with PX domains), it is quite possible that novel kinases, which are regulated by any of the

phosphoinositides, exist and play a role in the regulation of quite distinct cellular processes.

See also: **Bioenergetics:** Phosphatidylinositol-3-Phosphate; **Signaling:** Inositol Phosphate Kinases and Phosphatases; Phosphatidylinositol Bisphosphate and Trisphosphate; Phosphoinositide 4- and 5-Kinases and Phosphatases.

Further Reading

- Biondi RM and Nebreda AR (2003) Signalling specificity of Ser/Thr protein kinases through docking-site-mediated interactions. *Biochemical Journal* 372: 1–13.
- Kozma SC and Thomas G (2002) Regulation of cell size in growth, development and human disease: PI3K, PKB and S6K. *Bioessays* 24: 65–71.
- Lawlor MA and Alessi DR (2001) PKB/Akt: A key mediator of cell proliferation, survival and insulin responses? *Journal of Cell Science* 114: 2903–2910.
- Leslie NR and Downes CP (2002) PTEN: The down side of PI 3-kinase signalling. *Cell Signal* 14: 285–295.
- Marygold SJ and Leever SJ (2002) Growth signaling: TSC takes its place. *Current Biology* 12: R785–R787.
- Vanhaesebroeck B and Alessi DR (2000) The PI3K–PDK1 connection: More than just a road to PKB. *Biochemical Journal* 346: 561–576.
- Vanhaesebroeck B, Leever SJ, Ahmadi K, et al. (2001) Synthesis and function of 3-phosphorylated inositol lipids. *Annual Reviews in Biochemistry* 70: 535–602.

Phospholipase C

F Sekiya[†], National Institutes of Health, Bethesda, MD, USA

Published by Elsevier Inc.

This article is a revision of the previous edition article by Fujio Sekiya, Yeun Ju Kim and Sue Goo Rhee, volume 3, pp. 301–305, © 2004, Elsevier Inc.

Glossary

Diacylglycerol (DAG) A product of phosphatidylinositol 4,5-bisphosphate (PIP₂) cleaved by phospholipase C (PLC). It serves as a second messenger activating some members of protein kinase C (PKC)s, and is readily converted to phosphatidic acid in the cell by the action of diacylglycerol kinases.

Inositol 1,4,5-triphosphate (IP₃) The other product of PIP₂ cleavage by PLC. IP₃ is a short-life second messenger. It is readily dephosphorylated by specific phosphatases or phosphorylated by the IP₃ kinase.

Pleckstrin homology (PH) domain A domain structure that occurs in a wide range of proteins involved in intracellular signaling or as constituents of the cytoskeleton.

PH was originally found in pleckstrin (platelet C-kinase substrate protein). It binds to phosphoinositides such as PIP₂ and/or PIP₃. Because of its high affinity and specificity toward PIP₂, fusion proteins of PLC- δ 1 PH domain with fluorescent protein are now widely used to monitor the location and dynamics of PIP₂ in living cells.

Src homology 2/3 (SH2/3) domain Domain structures initially identified in the protooncoprotein *src*. They facilitate protein–protein interaction and occur in various signaling proteins. SH2 domain (~100 residues) recognizes phosphorylated Tyr residues in target peptides, whereas SH3 domain (~70 residues) binds to peptide sequences rich in proline.

Phospholipase C (PLC) is an enzyme that hydrolyzes a glycerophospholipid at the phosphodiester bond between the glycerol backbone and the phosphate group. All known eukaryotic PLCs utilize only phosphoinositides (phosphatidylinositol and its phosphorylated derivatives) as substrates; hence, they are called phosphoinositide-specific PLCs and are frequently just referred to as PLC. In signal transduction, activation of PLC is one of the earliest events following binding of various agonists to cell-surface receptors, and leads to diverse and profound consequences. Upon its activation, PLC hydrolyzes phosphatidylinositol 4,5-bisphosphate (PIP₂), a minor but essential constituent of plasma membranes, to modulate numerous PIP₂-dependent cellular processes. The two PIP₂ cleavage products, inositol 1,4,5-trisphosphate (IP₃) and 1,2-diacylglycerol (DAG), are critical second messengers to mediate calcium signaling in animal cells. PLC is not a single entity but a family of isozymes with various structures. Mechanisms for regulating PLC isozymes are also diverse. To date, 13 different mammalian PLCs have been identified, and they are grouped into six subfamilies based on their primary structures.

Physiological Roles of PLC

The earliest observation of PLC action was made by Hokin and Hokin in the 1950s. They discovered that stimulation of cells could induce rapid metabolism of inositol phospholipids. In 1975, Robert Michell published a hypothesis connecting phosphoinositide breakdown and cellular calcium mobilization. Functions of IP₃ and DAG as second messengers were established in the mid-1980s by Michael Berridge and Yasutomi Nishizuka, respectively. It was these landmark discoveries that brought special attention to PLC as the key enzyme to generate the two second messengers.

The soluble part of the PIP₂ cleavage product, IP₃, binds to its intracellular receptors on the endoplasmic reticulum (ER) (calcium storage site) and causes its calcium ions to be released into the cytoplasm. The other cleavage product, DAG, stays in the plasma membrane to recruit and activate protein kinase C (PKC) isozymes. Activation of certain PKC isozymes requires binding of both DAG and calcium ions. The activated PKC elicits a wide range of cellular responses through phosphorylation of many kinds of protein substrates (Figure 1).

Besides generating second messengers, modulating the population of PIP₂ in the plasma membrane is another function of PLC, which upon its activation quickly consumes large fractions of PIP₂ pools. PIP₂ constitutes only a minor phospholipid population and it is almost exclusively present in the inner leaflet of plasma membranes. Its turnover is rapid even in resting cells, yet its cellular levels are strictly kept constant. While IP₃ and DAG function as second messengers only in animal cells, PIP₂ is essential for all eukaryotic organisms to maintain many cellular processes through its interaction with a wide variety of proteins. First, it can modify activities of various proteins, such as the membrane-acting enzyme, phospholipase D, and the membrane-embedded transient receptor potential (TRP) channels, just to name a few. Second, it provides a platform to form functional signaling complexes. Many of the proteins constituting these complexes possess PIP₂-binding modules, such as pleckstrin homology (PH), FERM (F for 4.1 protein, E for ezrin, R for radixin, and M for moesin), and epsin N-terminal homology (ENTH) domains. Thus, reduction of cellular PIP₂ directly alters metabolic or signaling status. Third, the phospholipid is a focal point at which plasma membranes are tethered to cytoskeletal networks. Therefore, spatio-temporally controlled PIP₂ turnover is a necessary step for cells to regulate their morphology and motility.

[†]Deceased.

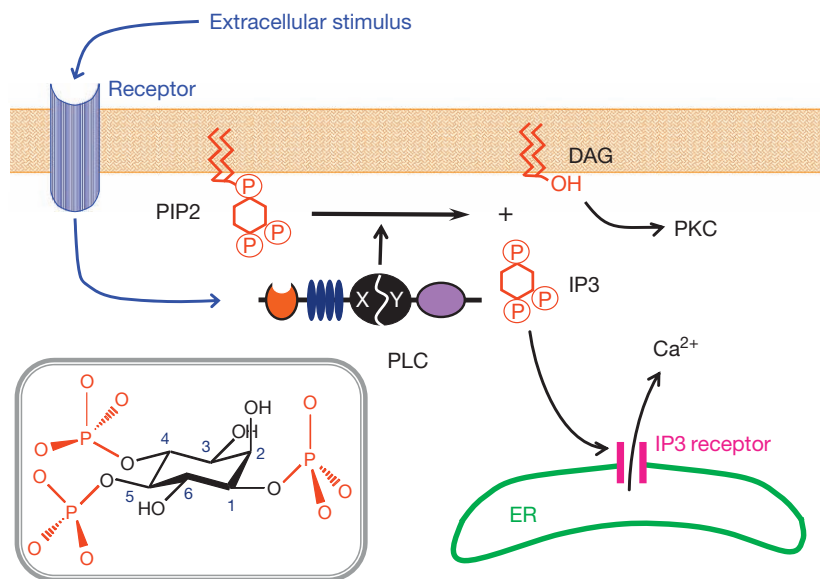


Figure 1 A scheme for the action of PLC in the cell. For abbreviations, see text. Inset shows chemical structure of IP3.

Given such diverse and profound consequences, PLC activity is under tight regulation and its activation is a central mechanism of cellular signaling in many instances. A large number of hormones, growth factors, and neurotransmitters rely on PLC to transmit their signals across plasma membranes.

Biochemistry of PLC

Structure of PLC

PLC has been found in yeasts, slime molds, plants, and animals, and is probably present in all taxa of eukaryotes. All eukaryotic PLCs appear to be evolved from a single archetype, and it is in the animal kingdom that PLC diversified. Thirteen different genes encoding PLC isozymes have been identified in mammals, and they are grouped into six subfamilies, named as β , γ , δ , ϵ , ζ , and η (Figure 2). Animals, from nematode *Caenorhabditis elegans* to fly to human, have similar sets of isozymes, but only the primitive ζ -like PLCs are found in other eukaryotes. The protein originally termed PLC- α was later found to be a proteolytic fragment of PLC- δ 1. Some isozymes have alternatively spliced variants at the messenger RNA (mRNA) level, and the total number of mammalian PLC polypeptides exceeds 20.

All PLC isozymes possess common structural features, indicating that they were derived from a common ancestor. All nonanimal PLCs, as well as PLC- ζ , consist of, from N- to C-terminal, four copies of EF-hands, regions called X and Y, followed by a C2 domain. The δ isozymes possess a PH domain in the N-terminus in addition to the structure described above. The other four types of PLC isozymes contain additional regulatory domain(s) in addition to those shown for the δ -subtype (Figure 2). PH domains are ~ 120 -residue units, which often function as membrane-binding sites through interaction with phosphoinositides. EF-hands are the calcium-binding motifs identified in calmodulin, but it is unclear whether PLC's EF-hands actually bind calcium ions. C2 domains (~ 120 residues) are the structures originally

found in PKC as calcium-dependent lipid-binding modules. X and Y regions are unique to eukaryotic PLCs and their sequences are highly conserved. Evidence shows that the two regions together form a discrete catalytic unit, though they are separated by a linker in the primary structure.

Mechanism of Catalysis

Calcium, at the physiological intracellular ranges, is required for PLC to catalyze PIP2 hydrolysis. However, with other inositides as substrates, much higher calcium ion concentrations are needed, a condition considered to occur only in test tubes. Based on studies of the crystal structures of a PLC- δ 1-IP3 complex and other biochemical data, a catalytic mechanism has been proposed. In this mechanism, the PLC-bound calcium ion is directly interacting with the IP3 to activate the 2-hydroxy group of inositol ring that in turn attacks the 1-phosphate, and leads to the cleavage of the head group to form a 1,2-cyclic phosphate intermediate and DAG. The cyclic derivatives are subsequently hydrolyzed to acyclic inositolphosphates. The positive residues located in the active site pocket are positioned to interact with the phosphate groups at the 4- and 5-positions of the inositol ring, explaining the preference of PLC for PIP2 over phosphatidylinositol or phosphatidylinositol 4-phosphate. The structure also indicates that the pocket cannot accommodate phosphoinositides with phosphate at the 3-position (e.g., phosphatidylinositol 3,4,5-trisphosphate), explaining why 3-phosphorylated inositides are not a substrate of PLC.

The X-Y linkers of PLC isozymes vary in their length and exhibit no noticeable sequence similarities to each other, but many, if not all, of them contain clusters of acidic amino acids. A recent crystallographic study of PLC- β 2 has shown that the acidic linker is unstructured and covers the catalytic cleft, presumably blocking access for the substrates. Removal of the acidic residues resulted in creation of unregulated, constitutively active enzymes. It was thus postulated that PLCs are autoinhibited by occlusion of the active site by the X-Y linker.

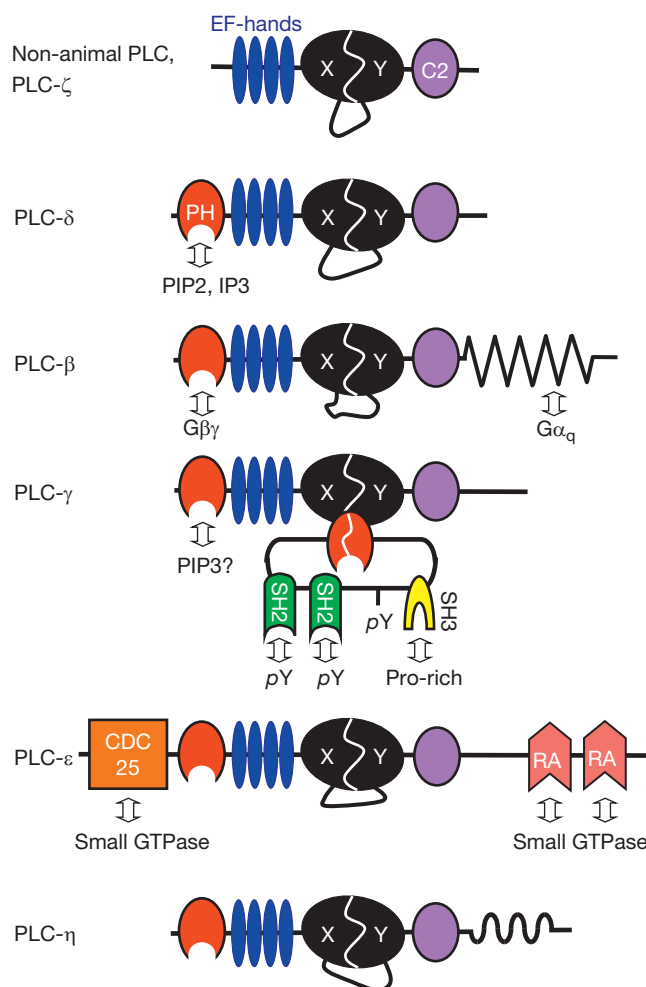


Figure 2 Structural organizations of PLC isozymes. Known ligands of domains of PLCs are shown. *pY*, phosphotyrosine; Pro-rich, proline-rich sequence.

Various kinds of acidic compounds are, when incorporated into substrate vesicles/micelles, known to stimulate PLC activities *in vitro*: Perhaps electrostatic repulsion by these negatively charged chemicals could push away the acidic linker to negate the autoinhibition.

PLC Isozymes and Their Regulation Mechanisms

PLC- β

Signals of many chemical and physical stimuli are received by cells via receptors that couple to heterotrimeric G proteins. A large portion of such stimulation induces PIP₂ hydrolysis by activating PLC- β isozymes. Beta isozymes ($\beta 1 \sim \beta 4$; ~ 150 kDa) are unique in having long C-terminal extensions that are absent in other classes of isozymes (Figure 2).

Upon activation of a receptor, the receptor-associated G protein heterotrimer, $\alpha\beta\gamma$, dissociates into α - and $\beta\gamma$ -subunits. Members of α -subunits of the Gq family (α_q , α_{11} , α_{14} , and α_{16}), but not other classes of α -subunits (Gs, Gi/o, and G12/13), directly activate PLC- β . In addition, $\beta\gamma$ -complexes can also stimulate PLC- β . The α_q - and $\beta\gamma$ -subunits appear to interact with different regions in the PLC- β molecule: α_q is recognized

by the C-terminal region unique to β -isozymes, whereas the $\beta\gamma$ -subunit binds to the N-terminal PH domain. In most cells, the Gi family of G protein is abundant compared to other classes. Therefore, it is reasonable to assume that only Gi/o-coupled receptors can provide significant amounts of $\beta\gamma$, and hence agonists that induce the activation of Gi/o are responsible for $\beta\gamma$ -mediated activation of PLC- β .

It has been shown that four isozymes of this subfamily are expressed in different tissues and they exhibit different sensitivities to the G protein subunits. PLC- $\beta 1$ and $\beta 3$ are widely expressed in various tissues and are more sensitive to α_q -mediated activation relative to other isozymes. PLC- $\beta 2$ expression is restricted to blood cells and is most sensitive to the activation by $\beta\gamma$ among the four PLC- β s. It has recently been shown that Rho-family small guanosine triphosphatases (GTPases) (Rac1, Rac2, and Cdc-42, but not RhoA) are able to activate PLC- $\beta 2$ via binding to its PH-domain. PLC- $\beta 3$, but not PLC- $\beta 1$, can be activated by the small GTPases albeit weakly. PLC- $\beta 4$ expression is only seen in some parts of the brain and in the retina.

Genes for all four β -isozymes have been disrupted in mice, and none of those single knockouts appeared to be fatal, probably due to their overlap functions. However, various

phenotypes have been reported with knockouts in mice. They include: PLC- β 1, epilepsy; PLC- β 2, dysfunctions in neutrophils and platelets; PLC- β 3, hyperresponse to opioid stimulation; and PLC- β 4, ataxia and impairment in visual perception.

PLC- γ

The two isozymes of this subfamily (γ 1 and γ 2; ~140 kDa) possess a long linker sequence between the X and Y regions. The linker consists of split halves of a PH domain flanking two Src-homology 2 (SH2) domains and an Src-homology 3 (SH3) domain. SH2 domain is phosphotyrosine-containing peptide sequences, while SH3 domain interacts with proline-rich sequences. Although separated by a large insertion, halves of the split PH domain assume a canonical PH fold. The linker region also contains a Tyr residue (Tyr783 of human PLC- γ 1), which is conserved in all PLC- γ proteins in various kinds of animals and of which phosphorylation is essential for the activation (Figure 2).

One major class of stimuli to activate PLC- γ is that of growth factors, such as platelet-derived growth factor, fibroblast growth factor, and epidermal growth factor. Receptors for these growth factors are transmembrane proteins with an extracellular part to bind their cognate ligands and an intracellular part containing intrinsic protein tyrosine kinase (PTK) activity and a tail with autophosphorylation sites. Ligand binding activates the PTK, which then recruits and phosphorylates PLC- γ . The other class of receptors that lead to activation of PLC- γ is immunoreceptors found in hematopoietic cells, such as the antigen-recognizing T-cell or B-cell receptors. These receptors possess no enzymatic activity. However, they, upon ligand binding, induce the formation of supramolecular complexes composed of the receptor themselves, adaptors, and cytoplasmic PTKs. PTKs belonging to the Src-family (Lck, Fyn, etc.), the Syk-family (Syk and ZAP-70), and the Tec-family (Btk, Itk, etc.) have been implicated in PLC- γ activation.

The long X-Y linker plays a central role in the regulation of PLC- γ . Upon activation, growth factor receptors possessing intrinsic PTK activity (or the supramolecular complexes containing PTKs) undergo autophosphorylation to provide docking sites for many proteins required for the downstream signaling. PLC- γ binds these phosphotyrosine-containing docking sites via its N-terminal SH2 domain, and its conserved Tyr residue is phosphorylated by the PTK. Phosphorylated PLC- γ then expresses its otherwise autoinhibited phospholipase activity via the intramolecular interaction between the phosphorylated Tyr residue and the C-terminal SH2 domain. Thus, the two SH2 domains collaborate via different functions during the activation. Whether the split PH domain has any function beyond its apparent structural importance to organize the X-Y linker is unclear, but recent study reported that Rac small GTPases interacts with and activate PLC- γ 2, but not γ 1, via this module. Physicochemical and biochemical properties of the SH3 domain have been extensively studied. A number of proteins of diverse classes/functions have also been identified as binding partners for the SH3 domain, but the physiologic significance for many of these interactions is yet to be established.

PLC- γ 1 is expressed ubiquitously and disruption of its gene in mice is lethal at an early embryonic stage. PLC- γ 2 is expressed in a set of hematopoietic lineages including B cells,

platelets, neutrophils, and mast cells. Knockout of γ 2 is not lethal but abolishes maturation of B cells, resulting in immunodeficiency, and causes dysfunctions of Fc-receptor-mediated signaling in platelets and mast cells.

PLC- δ

PLC- δ isozymes (δ 1, δ 3, and δ 4; ~90 kDa) have simpler structures compared to other isozymes (Figure 2). The protein found as PLC- δ 2 in the cow turned out to be a homolog of PLC- δ 4 identified in humans and the mouse. The delta isozymes lack obvious regulatory domains that facilitate coupling to receptors, suggesting that they are probably not activated directly by hormone receptors. PLC- δ has greater calcium sensitivity than β - and γ -isozymes. It is feasible that PLC- δ is secondarily activated when cytoplasmic calcium concentrations are elevated, perhaps following the activation of other classes of PLC. The PH domain of PLC- δ binds to PIP2 and it can also interact with IP3 at a higher affinity: this domain is thought to mediate tethering of PLC- δ to plasma membranes. When PIP2 hydrolysis occurs, one of its products, IP3, causes dissociation of the enzyme from membranes, thus terminating the reaction. This mechanism constitutes a negative feedback regulation system for this isozyme.

Gene targeting of PLC- δ isozymes has been conducted. No single PLC- δ isozyme knockout was fatal. The appearance of PLC- δ 1-deficient mice closely resembled that of nude mice. The isozyme appears to be necessary for commitment of skin stem cells. PLC- δ 3 null mice had no apparent phenotype, but combined deficiency in both PLC- δ 1 and - δ 3 was embryonic lethal. Male PLC- δ 4-deficient mice were infertile. It was postulated that this isozyme is required for the acrosome reaction of sperms.

PLC- ϵ

PLC- ϵ is the biggest PLC (>200 kDa) thus far recognized. It contains a CDC25-like guanine nucleotide exchange factor (GEF) domain in the N terminus, and two copies of Ras-binding motifs (RA domains) following the C2 domain (Figure 2). It has been shown that (1) PLC- ϵ is capable of hydrolyzing PIP2 both *in vitro* and *in vivo*, (2) the CDC25-like domain indeed has a GEF activity and is able to activate Ras and Rap GTPases, and (3) the RA domains bind to active GTP-forms of Ras or Rap. Because of the presence of domain structures that interact with small GTPases of the Ras superfamily, it seems likely that regulation of PLC- ϵ involves this class of proteins. Small GTPases seem likely to function as both regulators (upstream of PLC- ϵ) and effectors (downstream). Studies reported so far are still not sufficient to portray the exact mechanism by which this isozyme functions in the cell.

PLC- ϵ -null mice are viable but show defects in cardiac development and function.

PLC- ζ

A novel type of PLC named PLC- ζ (~70 kDa) was recently identified as a sperm-specific isozyme. PLC- ζ has a very primitive structure. It lacks the N-terminal PH domain (Figure 2)

and rather resembles PLCs found in plants or yeasts. A phylogeny analysis revealed that it is the least divergent from a hypothetical precursor among all mammalian PLCs. It has been shown that this sperm-specific isozyme is responsible for calcium oscillation in fertilizing eggs.

PLC- η

PLC- η s (η 1 and η 2; 115 kDa) are the newest member of the PLC family. PLC- η contains a C-terminal extension with a unique sequence in addition to the common structure. The regulatory mechanism for this new isozyme is mostly unknown, but an activation mechanism via G protein $\beta\gamma$ -subunit has been suggested. Both PLC- η isozymes are highly expressed in neuron-enriched regions of the brain.

See also: **Bioenergetics:** Inositol Trisphosphate and Calcium Signaling; IP_3 Receptors; Lipid Signaling and Ion Channels;

Signaling: Phosphatidylinositol Bisphosphate and Trisphosphate; Protein Kinase C Family.

Further Reading

- Carpenter G and Ji Q-s (1999) Phospholipase C- γ as a signal-transducing element. *Experimental Cell Research* 253: 15–24.
- Cockcroft S (2006) The latest phospholipase C, PLC η , is implicated in neuronal function. *Trends in Biochemical Sciences* 31: 4–7.
- Harden TK, Hicks SN, and Sondek J (2009) Phospholipase C isozymes as effectors of Ras superfamily GTPases. *Journal of Lipid Research* 50: S243–S248.
- Irvine RF (2003) 20 years of Ins(1,4,5)P₃, and 40 years before. *Nature Reviews Molecular Cell Biology* 4: 586–590.
- Rhee SG (2001) Regulation of phosphoinositide-specific phospholipase C. *Annual Review of Biochemistry* 70: 281–312.
- Suh PG, Park JI, Maonzoli L, et al. (2008) Multiple roles of phosphoinositide-specific phospholipase C isozymes. *BMB Report* 41: 281–312.
- Williams RL (1999) Mammalian phosphoinositide-specific phospholipase C. *Biochimica et Biophysica Acta* 1441: 255–267.

Phospholipase D

W Ali and M A Frohman, Stony Brook University, Stony Brook, NY, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Mary M. LaLonde and Michael A. Frohman, volume 3, pp. 306–313, © 2004, Elsevier Inc.

Glossary

Brefeldin-A ADP-ribosylated substrate (BARS) Plays a critical role in the fission step of COPI vesicle formation from Golgi.

Coat protein or COPI An ADP ribosylation factor (ARF)-dependent adaptor protein involved in membrane traffic.

NF- κ B (nuclear factor-kappaB) A heterodimeric protein composed of different combinations of members of the Rel family of transcription factors.

Piwi-interacting RNA (piRNA) The largest class of small RNA molecules that is expressed in animal cells.

RNA interference (RNAi) A system within living cells that takes part in controlling which genes are active and how active they are.

Small interfering RNA (siRNA) Sometimes known as short interfering RNA or silencing RNA, is a class of double-stranded RNA molecules, 20–25 nucleotides in length, that play a variety of roles in biology.

SNARE (an acronym derived from “SNAP (Soluble NSF Attachment Protein) REceptor”) proteins These mediate vesicle fusion, that is, the exocytosis of cellular transport vesicles with the cell membrane at the porosome or with a target compartment (such as a lysosome).

Son of Sevenless (Sos) A guanine nucleotide exchange factor (GEF) that activates monomeric GTPases such as Ras and Rac1.

Introduction

Phospholipase D (PLD) enzymes catalyze transphosphatidylation at the phosphodiester bonds found in a select but wide variety of substrates. Transphosphatidylation classically involves exchanging head groups of glycerophospholipids, but the range of substrates acted upon by the extended PLD superfamily is more extensive than just lipids. The most frequent reaction mediated by PLDs is cleavage of a substrate by hydrolysis, but in some cases, use of a nonwater lipid donor combines two lipids to result in synthesis of a new lipid.

The most widely studied PLD activity involves cleavage of the membrane lipid phosphatidylcholine (PC) to yield free choline and phosphatidic acid (PA; [Figure 1](#)). PA is an extensively studied bioactive lipid second messenger that effects changes in many cellular-signaling pathways and functions, depending on the cell type and subcellular location of PA generation. PA uses multiple mechanisms to achieve its effects, including acting as a lipid anchor to recruit proteins to sites of active signaling, functioning as an activator for selected enzymes, serving as substrate to generate other bioactive-signaling lipids, or even facilitating membrane curvature based on its cone-shaped physical structure.

The signaling pathways regulated by PLD and PA connect to many different cell biological processes, and from there to many different physiological processes, the latter of which are much less well defined at present. The cell biological processes include many different types of membrane vesicle trafficking and dynamic organization of the cytoskeleton, the control of cell growth and apoptosis, and interaction of the cell with its environment. These processes affect exocytosis, endocytosis, membrane biogenesis, cell migration, cell adherence, cell survival, and proliferation, which, in turn, affect physiological events such as viral budding from the Golgi complex, survival of

bacteria in the respiratory tract in insects, sporulation in yeast, stress responses in plants, phototransduction in flies, intersegmental vessel formation in zebrafish, and mitochondrial fusion, platelet activation, and neutrophil migration in mice.

PLD Structure

PLD superfamily members are defined by the presence of a consensus sequence, HXK(X)4D(X)6GSXN, known as the HKD domain, which comprises half of the catalytic site. The PLD catalytic core is composed of two HKD motifs that juxtapose to create the active site. Most superfamily members encode two HKD domains, and crystal structures of bacterial family members suggest that PLD proteins assemble as monomers. There are a few members of the superfamily that have only one domain, and accordingly need to dimerize to manifest catalytic activity. Both HKD domains must be intact for PLD to be catalytically active. The His residues mediate the transphosphatidylation reaction.

Prokaryotic PLDs encode just the PLD catalytic domain and function as soluble enzymes. MitoPLD, a mammalian PLD similar in size to prokaryotic PLDs, has, in addition, a leader sequence that targets it to mitochondria. Other superfamily members in yeast, plants, and animals, though, have added many types of domains and motifs to achieve targeted subcellular localization and regulated activity. PC-specific yeast, mammalian, and plant PLDs encode PX (phox) and PH (Pleckstrin homology) domains that function to recruit the PLDs to specific subcellular membranes through selectively binding to certain phospholipids or proteins. Similarly, these PLDs encode a short, basic amino acid-rich motif that binds phosphatidylinositol 4,5 bisphosphate (PI4,5P₂), which regulates both PLD localization and activation. By contrast, most plant

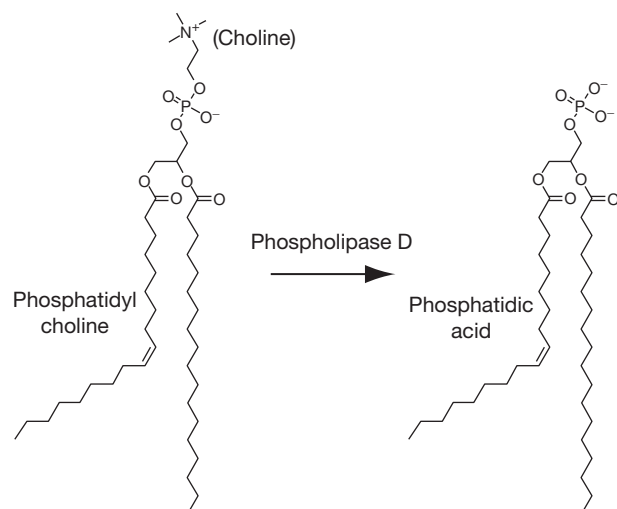


Figure 1 PLD hydrolysis of phosphatidylcholine to generate phosphatidic acid.

PLDs have a conserved C2 domain that enables regulation of the PLDs by calcium binding. Finally, mammalian PLD1 has binding sites for protein kinase C (PKC) and the small guanine triphosphatase (GTPase) families of adenosine diphosphate ribosylation factors (ARF) and Rho proteins, which regulate PLD1 activity through the interactions.

Posttranslational modifications are not essential for PLD activity. Yeast and mammalian PLDs undergo palmitoylation in their PH domain, which similarly is not necessary for enzymatic activity, but instead affects subcellular localization. Mammalian PLD also undergoes ubiquitination, which may control its degradation. The most interesting posttranslational modification is phosphorylation, because PKC is well established as an activator of PLD. However, PKC activates PLD1 via an adenosine triphosphate (ATP)-independent mechanism through interaction of the PKC regulatory domain with the amino terminus of PLD1. In fact, in the presence of ATP, PLD activity is diminished, suggesting that PKC-mediated phosphorylation plays a modest inhibitory role on PLD enzymatic activity. Nonetheless, *in vivo*, phosphorylation by other serine/threonine or tyrosine kinases may promote PLD activity.

Roles for PLDs in Cells and Organisms

Bacterial PLD

Yersinia pestis is a pathogenic bacterium transmitted by flea bite that causes the plague. A number of factors contribute to the virulence of *Y. pestis*, one of which is a PLD family member, Yersinia murine toxin (Ymt). Ymt has two HKD domains and has been shown to catalyze the cleavage of PC, phosphatidylserine, and phosphatidylethanolamine *in vitro*. Ymt as the name implies, is toxic to rats and mice, causing hypotension and vascular collapse when injected as a purified protein, and hypothetically as a result of its release from lysing bacteria during tissue invasion. However, bacteria lacking Ymt are not significantly less toxic to mice, and the Ymt protein is expressed at higher levels by the bacterium at 26 °C than at 37 °C, suggesting that its primary

role might take place in its flea host. In fact, Ymt, through a mechanism not currently well understood, appears necessary for survival of the bacterium in the flea digestive tract.

Many pathogenic bacteria also express Nuc, a highly divergent PLD family member with only one HKD domain that dimerizes to function as an endonuclease. Nuc is exported into the periplasmic space and may be released during bacterial invasion to promote movement of the bacteria through damaged tissues.

Yeast PLD (Spo14)

Yeast PLD selectively hydrolyzes PC to generate PA. Yeast PLD was identified in *Saccharomyces cerevisiae* as being required for sporulation, and hence denoted *spo14*. Extensive studies revealed a requisite role of PLD in the prospore membrane formation step that leads to creation of the spore's plasma membrane and cell wall. More specifically, Golgi-derived secretory membrane vesicles traffic properly to the meiotic plaque, an electron-dense scaffold on the outer surface of the spindle body, but the vesicles then fail to fuse to form the prospore membrane. Mechanistically underlying this failure, the PA produced by *spo14* has been shown to recruit spo20, a t-SNARE (soluble NSF attachment protein receptor (SNARE)) required for the vesicle fusion step, to the site of prospore membrane formation. Spo20 encodes a 40-residue basic amino-acid-rich domain that specifically binds to PA, and use of this domain has been made to create a fluorescent PA sensor that identifies PA in living yeast and mammalian cells. *Spo14* and PA also undertake additional roles during sporulation, but these are less well understood.

PLD activity is not essential during vegetative yeast growth, but becomes required when other components of the secretory machinery, such as Sec14, are absent. Although the exact role is complicated, PLD appears to be a component of a checkpoint mechanism that monitors levels of cellular PC.

Zebrafish

Zebrafish have two PLD genes that closely correspond to the two classic, PC-specific mammalian genes, PLD1 and 2. *pld1* transcripts are present throughout development, and knocking down the *pld1* gene using morpholinos results in an intersegmental vessel-developmental defect.

Drosophila

Drosophila have a single PLD gene that is expressed throughout development and in adulthood. Flies lacking PLD (PLDnull flies) are viable, but less so than PLD heterozygous mutants. A prolonged cellularization phase is observed in PLDnull flies, as well as morphological defects that have not been well specified. Overexpression of PLD during development also results in morphological defects and lethality.

PLD involvement in photoreceptors

Drosophila have compound eyes that are made up of hundreds of photoreceptor cells. Each photoreceptor cell is made up of a cell body and a rhabdomere. Photoreceptor transduction occurs in the rhabdomere which is a specialized compartment

that has polarized vesicle trafficking. Overexpression of *Drosophila* PLD produces PA that disrupts rhabdomere formation.

PA is important in the signal transduction pathway used for photoreceptors. Photoactivation of rhodopsin causes it to change its conformation to become metarhodopsin. Metarhodopsin then activates PLC β , which hydrolyzes PI(4,5)P₂ to generate diacylglycerol (DAG), which, in turn, opens transient receptor potential channels, causing changes in membrane potential and creation of a visual signal. The signal is then terminated by DAG kinase-mediated phosphorylation of the DAG to PA, which is then cycled back to reform PI(4,5)P₂. PLDnull flies have normal-appearing photoreceptors but are 50-fold less sensitive to light, indicating a role for PLD in the phototransduction-cycling process. PLD-generated PA may also play a role in protection of the rhabdomeres under low-light conditions by undergoing dephosphorylation to generate small amounts of DAG.

Plant PLD

PLD was first identified in plants. Up to 12 isoforms of PLD can be found in some plants, such as *Arabidopsis thaliana*. Plant PLDs are divided into subfamilies based on their Ca²⁺ requirement for phospholipid binding, as discussed earlier. Most plant PLDs hydrolyze phosphatidylethanol and phosphatidylglycerol in addition to PC to produce PA.

PLD production of PA in plants is involved in growth, development, and responses to stress. Many of the isoforms have overlapping functions; however, they each seem to have predominant responsibility for a specific stress response. PLD activity increases during periods of nutrient starvation (nitrogen, phosphorus, and potassium) to promote growth and phospholipid hydrolysis. Hyperosmotic stress also induces several PLD isoforms to promote lipid degradation on internal membranes while protecting lipid loss at the plasma membrane. PLD activity may also protect from leaf water loss. PLD can be activated in response to hormones such as ABA, which causes PLD to interact with G α s. Other hormones that stimulate PLD activity include auxin, a developmental hormone, and ethylene, a hormone that promotes senescence. Like mammalian PLD, plant PLD directs localized PA production and can interact with proteins directly. Plant PLD has been shown to interact with actin, microtubules, and cardosina.

Mammalian PLD

Overview

Approximately 30 years ago, it was noted that prolonged exposure of cells to agonists that stimulated PLC to cleave PI4,5P₂ into inositol triphosphate (IP₃) and DAG yielded only brief production of IP₃, but sustained production of DAG. This implied that there was a second pathway through which DAG could be produced, leading to the proposal and discovery of mammalian PLD as an additional source of DAG subsequent to dephosphorylation of newly formed PA. Formation of DAG through this pathway has since been shown to be important in cardiac muscle preischemic conditioning, and in activation of a DAG-responsive RAS exchange factor in T cells.

PLD activity was subsequently noted to increase in many types of cells in response to agonists that stimulated G-protein-

coupled receptors and receptor tyrosine kinases, with the highest levels in secretory cells. Primary alcohols (ethanol and 1-butanol), which competitively suppress PA production by diverting PLD activity to instead make phosphatidylalcohol, were shown to suppress agonist-stimulated secretion. These observations led to a general proposal that PLD is involved in regulated exocytosis. The idea was further supported by the finding the small G-protein ARF, which promotes budding of membrane vesicles from the *trans*-Golgi, is a major activator of PLD.

Cloning of the two classic isoforms of mammalian PLD, PLD1 and PLD2, revealed both overlapping and distinct properties for the enzymes. PLD1 and PLD2 are approximately 50% identical, have the same structure, and similarly hydrolyze PC to generate PA. Most cell types express both isozymes, but the relative expression of each varies between tissues. Both enzymes cycle through a series of subcellular sites. In most cells, PLD1 is found on a diverse collection of perinuclear membrane vesicles, but then translocates to the plasma membrane upon cell stimulation and returns to the perinuclear vesicles by passing through endosomes. PLD1 is observed in many cell types to colocalize with lysosomes, the Golgi and endoplasmic reticulum (ER). In some cells, the majority of PLD1 is found in the steady state on the plasma membrane, but this is more common for PLD2. PLD2 enters the cell via endocytosis upon receptor stimulation, and cycles through endosomes and the Golgi before returning to the plasma membrane. Overexpression or RNA interference (RNAi)-mediated knockdown of PLD1 affects regulated exocytosis in many cell types, whereas overexpression or RNAi-mediated knockdown of PLD2 is more frequently associated with alterations in the actin cytoskeleton and shape of cells. Nonetheless, in some studies, PLD1 has been associated with signaling-induced cell-shape changes and PLD2 has been linked to membrane vesicle trafficking at the Golgi.

Activation

PKC, ARF, and Rho family members synergistically activate PLD1 by binding to individual regions on the protein (Figure 2). PLD2, on the other hand, is constitutively active *in vitro*, and appears to be regulated *in vivo* through repression, which is countered during signaling events. It is not clear how the repression and activation works, but one mechanism

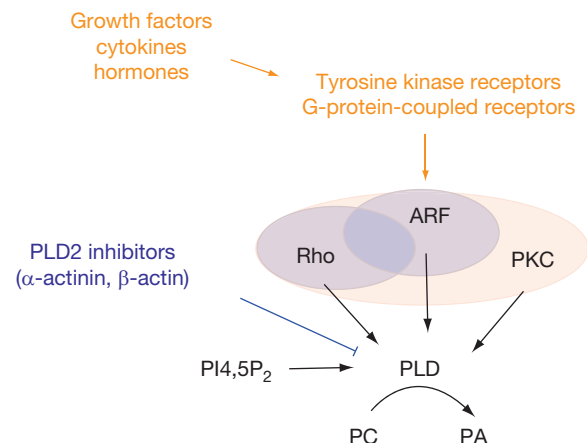


Figure 2 PLD is activated by Rho and ARF GTPases, PKC, and PI4,5P₂ to produce PA from PC.

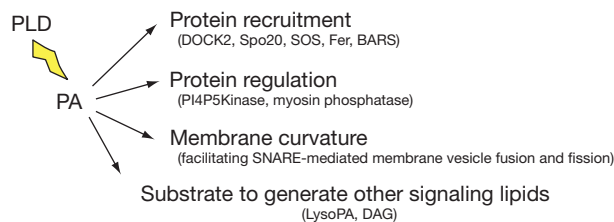


Figure 3 Cell biological roles for PA.

suggested is that F-actin inhibition of PLD2 may be terminated by binding of ARF to the protein. Alternately, both isoforms require PI4,5P₂ as a coactivator, and it has been suggested that limiting the access of freely available PI4,5P₂ may also serve to regulate the timing and spatial restriction of PLD and/or PLD1 activity.

Function of PA

In mammalian cells, PA has been shown to trigger dynamic relocation of proteins to specific subcellular membrane compartments (Figure 3). For example, PA produced in NIH3T3 cells in response to epidermal growth factor (EGF) stimulation and activation of PLD2 recruits Son of Sevenless (SOS), an exchange factor for the small G-protein oncogene Ras, to the plasma membrane, leading to GTP loading of Ras and cellular transformation. Similarly, in neutrophils, chemokine stimulation leads to PLD activation at the leading edge of the neutrophils, local production of PA, and focused recruitment of DOCK2, an exchange factor for Rac1, which then stimulates F-actin polymerization and polarized cell migration. PA also recruits the Fer and Raf1 kinases, Rac1, and a variety of other proteins to specific subcellular membrane compartments. Finally, PA produced by PLD2 in the Golgi has been reported to recruit brefeldin-A ADP-ribosylated substrate (BARS), a protein that triggers scissioning of Coat protein (COPI) retrograde-transport vesicles.

PA has also been shown to activate proteins, such as PI4P5Kinase, which generates PI4,5P₂, or to serve as substrate for PA phosphatases to generate DAG, or PLA2 enzymes to generate lysoPA, a potent mitogen. Finally, conversion of PC to PA, which has a smaller headgroup, has been proposed to promote negative membrane curvature. Negative curvature is important in membrane vesicle fusion and fission events, lowering the activation energy required to complete the process.

PA regulation of PLD

PA activates PI4P5Kinase which makes PI4,5P₂, and it helps localize ARF and Rho GTPases during actin reorganization. PI4,5P₂ and the GTPases are PLD stimulators, leading in theory to even more production of PA. Conversely, PA also stimulates ARF GAPs, which deactivate ARF GTPases. Multiple such subtle ways to regulate PLD activity are used to exert fine control over PA production.

Tools

In the presence of a primary alcohol such as 1-butanol, PLD will hydrolyze the alcohol preferentially. For this reason, butanol has historically been frequently employed as a means of inhibiting PLD. However, alcohol causes dramatic changes in membrane lipid composition and fluidity that can affect

phospholipid synthesis and receptor-stimulation events. More recently, lipase-inactive PLD alleles have gained popularity as dominant negative inhibitors, and RNAi-mediated knockout of PLD has become routine. Knockout mice and cells have recently been generated and studies on them are beginning to come out.

For short-term studies, small molecule inhibitors for PLD are now available. 5-Fluoro-2-Indolyl des-chlorohalopemide (FIP1), an analog of a pan-PLD inhibitor identified in a screen at Novartis, is a potent inhibitor of PLD1 and PLD2 (IC₅₀ at sub- or low-nM concentrations) that exhibits little cytotoxicity, no off-target effects thus far, and sufficient bioavailability *in vivo* to be usable for animal studies. Isoform-selective analogs have also been developed using combinatorial chemistry, and although somewhat less potent and well studied, offer the option to distinguish between the roles undertaken by the PLD1 and PLD2 isoforms in complex cell settings.

Sensors to detect localized production of PA in cells have also been developed, by fusing PA-binding domains from proteins such as the Raf1 kinase or yeast Spo20 to GFP. Such sensors have been used to image production of PA on macrophage phagosomes and on endocytosing vesicles upon which Ras–Raf1 kinase signaling is taking place.

Cell Biological Processes Regulated by PA

Actin cytoskeleton rearrangement

PLD is involved in dynamic localization of proteins necessary for actin reorganization. Reorganization of the actin cytoskeleton is necessary for many cellular responses such as cell spreading and migration, as well as membrane vesicle trafficking. PA helps localize ARF GTPases and Rho family GTPases. ARF GTPases are GTP-binding proteins involved in vesicle trafficking and actin cytoskeletal rearrangements. Of the six known ARF proteins, ARF1 and ARF6 are highly implicated in PLD-mediated actin rearrangement. ARF1 is associated with Golgi complex formation and is known to be required for coated vesicle formation, whereas ARF6 is associated with endosomal membrane trafficking and actin cytoskeletal rearrangements. Rho GTPases are another group of GTP-binding proteins that can regulate actin polymerization. PLD interactions with Rho family GTPases are independent of ARF interactions, although, in some cases, Rho interactions have been shown to enhance ARF-mediated effects. Rho GTPases are involved in formation of lamellipodia, sheet-like actin structures used to form a leading edge during cell migration, filopodia, which are finger-like projections of the cytoskeleton, and stress fibers used for cell adhesion and lagging edge cell migration. Rho-family GTPase members that are important in PLD-mediated actin rearrangement include Rac1 and Cdc42.

F-actin has been shown to stimulate PLD activity. Concentration of PA at filaments attract and orient ARF6 and Rac1 which direct actin filament formation while stimulating PLD to produce more PA. PLD is also involved in steps prior to GTPase activity. DOCK2, an exchange factor for Rac1, localizes to the leading edge of the plasma membrane in response to PA production, leading to Rac1 activation, F-actin polymerization, and lamellipodia formation. PLD2 activity also regulates myosin activity through recruitment and inhibition of myosin phosphatase, an inhibitor of myosin activity.

Finally, studies on PLD1 knockout mice, which are otherwise viable and overtly normal appearing, have indicated that integrins on platelets lacking PLD1 are activated poorly, leading to decreased clot formation and resulting in protection from pulmonary embolisms and strokes. It is not clear whether the requirement for PLD1 generation of PA is acutely required during thrombus formation, or whether the phenotype reflects long-term PLD1 deficiency.

Vesicle trafficking

Subcellular localization of PLD has indicated that both isozymes are frequently present where the formation of vesicles occurs, such as at the plasma membrane, the Golgi apparatus, and in the endosomal pathway. Generally, PLD is stimulated during vesicle trafficking. PLD-produced PA promotes vesicle formation due to its ability to encourage membrane curvature and to recruit proteins like dynamin and BARs that mediate vesicle scissioning. PA also facilitates SNARE-mediated vesicle fusion. PLD involvement in specific vesicle trafficking events is described below.

Endocytosis

PLD has been proposed to be involved in endosomal recycling; treatment of HeLa cells with 1-butanol results in accumulation of recycling endosomal membrane tubules. Inhibiting the PLD activators PKC and RalA or overexpression of dominant negative PLD1 or PLD2 or PLD RNAi inhibits the internalization of epidermal growth factor receptor (EGFR). PLD2 activity has also been linked to endocytosis of μ -opioid receptors and metabotropic glutamate receptors 1 and 5, whereas PLD1 activity has been associated with B-cell receptor trafficking. Endocytosis of receptors on the plasma membrane can occur through a clathrin-dependent or -independent manner. ARF6 has been shown to associate with the clathrin-independent endosomal pathway and endosomal recycling vesicles. Interestingly, it has recently been demonstrated that PLD1 and 2 bind dynamin in its GTP-bound form and stimulate its GTPase activity, which is required during vesicle scissioning in clathrin-dependent endocytosis.

Exocytosis

Inducers of exocytosis also stimulate PLD at the plasma membrane. Isolated chromaffin cells have been used extensively to study exocytosis and PLD stimulation. Chromaffin cells are neuroendocrine cells found in the adrenal gland. PLD1 appears to be most involved in exocytosis because dominant negatives and small interfering RNA (siRNA) for PLD1, but not PLD2, impair exocytosis. There is evidence that syntaxin, a SNARE membrane associated protein, may provide a polybasic-binding site for PA, allowing PA to concentrate at specific sites. Alternately, PA may help to recruit syntaxin or affect its function. The major interacting partners for PLD1 in exocytosis are ARF6, RalA, Rac1, and RSK2. ARF6 regulates exocytosis in chromaffin and PC12 cells, and PLD1 activation by ARF6 promotes this process. It has been proposed that ARF6 controls the spatiotemporal production of PA to restrict its formation to contact sites between granules and the plasma membrane. SNARE-mediated vesicle fusion is enhanced by the presence of PA on the t-SNARE-containing

membrane surface. PLD1 traffics to the plasma membrane on v-SNARE-containing membrane vesicles, but then binds to and is activated by PI4,5P₂ on the t-SNARE-containing membrane surface, where it then produces PA to recruit v-SNAREs, which have a region of basic amino acids known to be critical for the fusion process. RalA has been shown to interact directly with PLD1 where it can enhance ARF6 stimulation. There are also suggestions that Rac1 is involved in spatiotemporally regulating exocytosis by PLD1. Lastly, RSK2, which is part of the S6 kinase family, has been shown to phosphorylate PLD1, and this phosphorylation is essential to exocytosis in chromaffin cells.

Golgi

Primary alcohols inhibit protein export from the ER to the Golgi and the release of vesicles from the *trans*-Golgi network. The formation of coated vesicles is also inhibited by ethanol. PA dephosphorylation to form DAG, which can activate PKC, has also been shown to regulate vesicle trafficking at the Golgi apparatus.

Signal transduction

In addition to vesicle trafficking, PLD-produced PA serves as a signal transducer that can recruit proteins. The negative charge of PA can attract and orient other proteins for activity during signal transduction events.

Epidermal growth factor receptor

Growth factors have been shown to stimulate PLD activity. The EGF receptor dimerizes upon stimulation by its ligand EGF, resulting in phosphorylation that attracts signal-transducing proteins such as Src and Grb2. PLD is activated by this hub of proteins. EGFR signals are altered by endocytic internalization, which has been shown to be PLD mediated.

Mammalian target of rapamycin

Mitogen signaling activates PLD. Increased PLD activity is associated with cell survival. These effects may be mediated via PA's interactions with mTOR, the mammalian target of rapamycin. mTOR is a kinase that transduces cell survival signals by phosphorylating downstream targets. PA promotes mTOR phosphorylation of S6 kinase and Akt. PA binds to the internal FKBP12 domain of mTOR in a competitive manner with rapamycin. PA likely stabilizes mTOR complexes of mTOR and Raptor, which together form the mTORC1 complex, and mTOR and Rictor, which form mTORC2. It is likely that mTORC1 and mTORC2 have different binding affinities for PA.

PLD and cancer

There is emerging evidence that PLD is upregulated in cancer, which makes PLD a potential therapeutic target. PLD is activated in response to growth factors and participates in mitogen and cell survival signal transduction. PLD1 transcription may be increased in breast cancer in response to nuclear factor- κ B (NF κ B) binding to the PLD promoter region. PLD, often in conjunction with PI3K, is overactive in a number of cancers, and PLD2 driver mutations have been identified in

breast cancer. Ras, which is frequently mutated in cancer to be constitutively active, activates RalA which can then activate PLD. The major mitogen-signaling pathway that ties PLD to cancer involves ERK, AKT, mTOR, and PKC activation. Overactive PLD simultaneously impedes the effectiveness of rapamycin treatment of cancer through increasing mTOR activity. Inhibition of PLD has been shown to inhibit ERK, AKT, and mTOR phosphorylation in lung cancer. PLD activity is associated with vascularization and cell motility.

MitoPLD

A defining property of being included in the PLD family is the attribution of having an HKD motif. A BLAST search identified a protein with a single HKD motif that also contains a mitochondrial-localization signal at its N-terminus. This protein, denoted mitoPLD, is thought to play a role in mitochondrial fusion. MitoPLD is anchored in the outer membrane of the mitochondria and dimerizes to form an active catalytic complex that cleaves cardiolipin to form PA. Overexpression of mitoPLD results in mitochondrial aggregation, while knockdown or dominant negative expression of mitoPLD results in mitochondrial fragmentation. These and other observations indicate that PA produced by mitoPLD at the mitochondrial membrane is involved in mitochondrial fusion. Intriguingly, mice and flies lacking mitoPLD (the *Drosophila* homolog is known as Zucchini) exhibit blocks in germ-cell development during meiosis. Although the mechanism remains unclear, the PA on the mitochondrial surface appears to be required for biogenesis of a specialized form of RNAi (Piwi-interacting RNA (piRNA)) that is also generated at the mitochondrial surface or on juxtaposed structures.

Conclusions

The delineation of roles for PA is undergoing rapid evolution as small molecule inhibitors and mice deficient in the PLD isoforms are being characterized. Inhibition of PLD may prove useful in a variety of disease settings, as will be determined in coming years.

See also: **Molecular Biology:** Small RNAs in Bacteria; **Signaling:** Nuclear Factor Kappa B.

Further Reading

- Bi K, Roth MG, and Ktistakis NT (1997) Phosphatidic acid formation by phospholipase D is required for transport from the endoplasmic reticulum to the Golgi complex. *Current Biology* 7: 301–307.
- Choi SY, Huang P, Jenkins GM, et al. (2006) A common lipid links Mfn-mediated mitochondrial fusion and SNARE-regulated exocytosis. *Nature Cell Biology* 8: 1255–1262.
- Elvers A, Stegner D, Hagedorn I, et al. (2010) Impaired alpha(IIb)beta(3) integrin activation and shear-dependent thrombus formation in mice lacking phospholipase D1. *Science Signal* 3: ra1.
- Foster DA (2009) Phosphatidic acid signaling to mTOR: Signals for the survival of human cancer cells. *Biochimica et Biophysica Acta* 1791(9): 949–955.
- Huang H, Gao Q, Peng X, et al. (2011) piRNA-associated germline nuage formation and spermatogenesis require MitoPLD pro-fusogenic mitochondrial-surface lipid signaling. *Developmental Cell* 20: 376–387.
- LaLonde MM, Janssens H, Rosenbaum E, et al. (2005) Regulation of phototransduction responsiveness and retinal degeneration by a phospholipase D-generated signaling lipid. *Journal of Cell Biology* 169: 471–479.
- Lee CS, Kim KL, Jang JH, et al. (2009) The roles of phospholipase D in EGFR signaling. *Biochimica et Biophysica Acta* 1791: 862–868.
- Li M, Hong Y, and Wang X (2009) Phospholipase D- and phosphatidic acid-mediated signaling in plants. *Biochimica et Biophysica Acta* 1791: 927–935.
- McDermott M, Wakelam MJ, and Morris AJ (2004) Phospholipase D. *Biochemistry and Cell Biology* 82: 225–253.
- Nakanishi H, Morishita M, Schwartz CL, et al. (2006) Phospholipase D and the SNARE Sso1p are necessary for vesicle fusion during sporulation in yeast. *Journal of Cell Science* 119: 1406–1415.
- Nishikimi A, Fukuhara H, Su W, et al. (2009) Sequential regulation of DOCK2 dynamics by two phospholipids during neutrophil chemotaxis. *Science* 324: 384–387.
- Scott SA, Selvy PE, Buck JR, et al. (2009) Design of isoform-selective phospholipase D inhibitors that modulate cancer cell invasiveness. *Nature Chemical Biology* 5: 108–117.
- Su W, Yeku O, Olepu S, et al. (2009) FIPI, a phospholipase D pharmacological inhibitor that alters cell spreading and inhibits chemotaxis. *Molecular Pharmacology* 75: 437–446.
- Sung T-C, Roper RL, Zhang Y, et al. (1997) Mutagenesis of phospholipase D defines a superfamily including a trans-Golgi viral protein required for poxvirus pathogenicity. *EMBO Journal* 16: 4519–4530.
- Yang J-S, Gad H, Lee SY, et al. (2008) COPI vesicle fission: A role for phosphatidic acid and insight into Golgi maintenance. *Nature Cell Biology* 10: 1146–1153.
- Zhao C, Du G, Skowronek K, Frohman MA, and Bar-Sagi D (2007) Phospholipase D2-generated phosphatidic acid couples EGFR stimulation to Ras activation by Sos. *Nature Cell Biology* 9: 706–712.

Phospholipid Synthesis in Yeast

G-S Han and G M Carman, Rutgers University, New Brunswick, NJ, USA

Published by Elsevier Inc.

This article is reproduced from the previous edition, volume 3, pp. 321–325, © 2004, Elsevier Inc.

Glossary

Biochemical regulation Process of controlling the activity of enzymes.

Genetic regulation Process of controlling the expression of genes.

Phospholipid A lipid molecule that contains two fatty acyl chains that are esterified to the *sn*-1 and *sn*-2 positions of a glycerol phosphate backbone. Different

phospholipid molecules (e.g., PC, PE) are distinguished by the water-soluble head group (e.g., choline, ethanolamine) that is attached to the phosphate moiety of the molecule.

Saccharomyces cerevisiae A species of yeast that is commonly used in genetic and biochemical research. It serves as an excellent model eukaryotic organism because it is easily cultured and genetically manipulated.

Phospholipids are essential molecules that contribute to the structural definition of the cell. They also play a role as signaling molecules that participate in the regulation of cellular processes. The yeast *Saccharomyces cerevisiae* has been used as a model system to study phospholipid synthesis and its regulation in eukaryotes. Molecular, genetic, and biochemical studies have shown that phospholipid synthesis is regulated in a coordinated fashion. The mechanisms that govern this regulation mediate the messenger RNA (mRNA) and protein levels of the biosynthetic enzymes, as well as their activity and localization.

Phospholipid Composition

The major phospholipids found in the membranes of *S. cerevisiae* include phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylinositol (PI), and phosphatidylserine (PS) (Figure 1). Mitochondrial membranes also contain phosphatidylglycerol (PG) and cardiolipin (CL). The composition of phospholipids in the cell can vary dramatically when culture conditions are altered. However, the average charge of the membrane phospholipids remains relatively constant. Genetic and biochemical regulation of the enzymes in the phospholipid biosynthetic pathways can compensate for changes in the levels of phospholipids of one charge by orchestrating parallel changes in the levels of phospholipids of the opposite charge.

Phospholipid Synthesis

Phospholipid synthesis is a complex process containing many branch points. A key molecule in this process is phosphatidate (PA). It plays a central role in phospholipid synthesis as the precursor of all phospholipids synthesized via the cytidine diphosphate (CDP)–diacylglycerol (DAG), CDP–choline, and CDP–ethanolamine pathways (Figure 2). In the CDP–DAG pathway, PA is converted to CDP–DAG, which is then utilized for the synthesis of PS, PE, and PC. CDP–DAG is also utilized for the synthesis of PI and CL. In the CDP–choline

and CDP–ethanolamine pathways, DAG is generated from the dephosphorylation of PA, and is utilized for the synthesis of PC or PE when cells are supplemented with choline or ethanolamine, respectively. The DAG derived from PA is also used for the synthesis of triacylglycerols. The nucleotide cytidine triphosphate (CTP) is another important precursor in the synthesis of phospholipids. CTP, which is synthesized from UTP, is the direct precursor of the activated, energy-rich phospholipid pathway intermediates CDP–DAG, CDP–choline, and CDP–ethanolamine (Figure 2).

Synthesis of PA

PA is synthesized from glycerol 3-phosphate via two acylation reactions. The first acylation step, which results in the formation of lysoPA, is catalyzed by the *GAT1*- and *GAT2*-encoded glycerol 3-phosphate acyltransferases. Deletion of either gene does not affect cell viability, but deletion of both genes results in a lethal phenotype. Acylation of lysoPA, which is catalyzed by the *SLC1*-encoded lysoPA acyltransferase, generates PA. In addition to the *de novo* synthesis, PA can be produced from phospholipid turnover: (1) hydrolysis of PC by the *SPO14*/*PLD1*-encoded phospholipase D; (2) phosphorylation of DAG by CTP-dependent DAG kinase (whose structural gene is not yet known); and (3) dephosphorylation of diacylglycerol pyrophosphate (DGPP) by the *DPP1*- and *LPP1*-encoded lipid phosphate phosphatase enzymes.

CDP–DAG Pathway

PS, PE, and PC are synthesized via reactions in the CDP–DAG pathway. The committed step in this pathway is the synthesis of CDP–DAG from PA and CTP. This reaction is catalyzed by the essential enzyme CDP–DAG synthase, which is encoded by the *CDS1* gene. CDP–DAG is then converted to PS by the *PSS1*/*CHO1*-encoded PS synthase. This enzyme displaces the cytidine monophosphate (CMP) moiety of CDP–DAG with serine. PS is then decarboxylated to PE by the *PSD1*- and

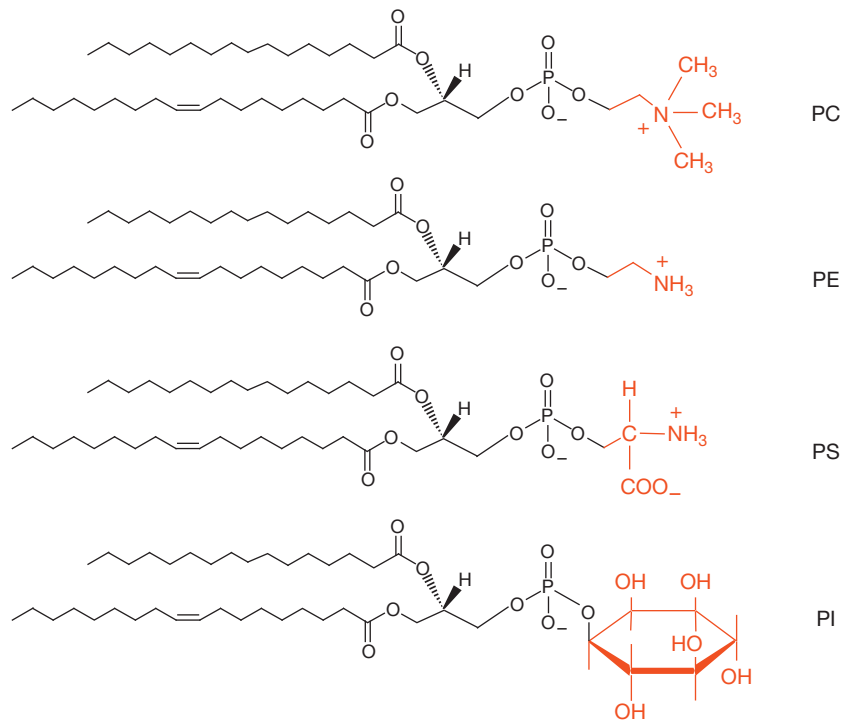


Figure 1 Major phospholipids in *S. cerevisiae*. The phospholipid headgroups (i.e., choline, ethanolamine, serine, and inositol) are shown in red.

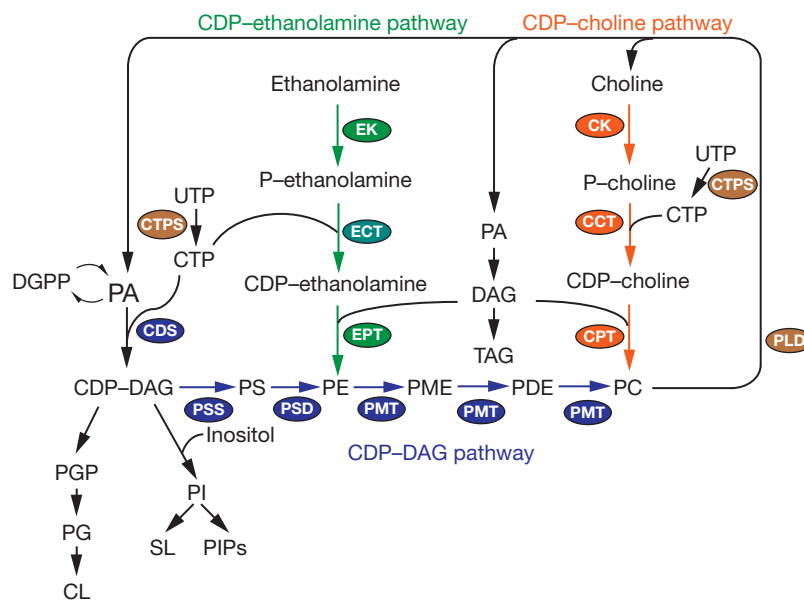


Figure 2 Pathways for the synthesis of the major phospholipids in *S. cerevisiae*. The pathways shown for synthesis of phospholipids include the relevant steps discussed in the text. The CDP-DAG, CDP-choline, and CDP-ethanolamine pathways are indicated. The enzymes CDP-DAG synthase (CDS), PS synthase (PSS), PS decarboxylase (PSD), and the PE methyltransferases (PMT) catalyze reactions that lead to the formation of PC by the CDP-DAG pathway (shown in blue). Choline kinase (CK), choline-P cytidyltransferase (CCT), and choline phosphotransferase (CPT) catalyze reactions that lead to the formation of PC by the CDP-choline pathway (shown in red). Ethanolamine kinase (ET), ethanolamine-P cytidyltransferase (ECT), and ethanolamine phosphotransferase (EPT) catalyze reactions that lead to the formation of PE by the CDP-ethanolamine pathway (shown in green). The reactions catalyzed by CTP synthetase (CTPS) and phospholipase D (PLD) are shown in brown. A more comprehensive description that includes the synthesis of PA and additional steps in these pathways may be found elsewhere. PA, phosphatidate; CDP-DAG, CDP-diacylglycerol; PS, phosphatidylserine; PE, phosphatidylethanolamine; PME, phosphatidylmonomethylethanolamine; PDE, phosphatidyltrimethylethanolamine; PC, phosphatidylcholine; DAG, diacylglycerol; TAG, triacylglycerol; PGP, phosphatidylglycerophosphate; PG, phosphatidylglycerol; CL, cardiolipin; SL, sphingolipids; PIPs, polyphosphoinositides, DGPP, diacylglycerol pyrophosphate.

PSD2-encoded PS decarboxylases. The *PSD1*-encoded PS decarboxylase is responsible for >90% of the cellular PS decarboxylase activity, while the *PSD2*-encoded enzyme is responsible for ~10% of the enzyme activity. Finally, PE is converted to PC via three sequential methylation reactions. The first methylation reaction is catalyzed by the *PEM1/CHO2*-encoded PE methyltransferase, whereas the last two methylation reactions are catalyzed by the *PEM2/OPI3*-encoded phospholipid methyltransferase.

CDP-Choline and CDP-Ethanolamine (Kennedy) Pathways

When cells are supplemented with choline and ethanolamine, PC and PE are also synthesized via the CDP-choline and CDP-ethanolamine pathways, respectively. In the CDP-choline pathway, choline is phosphorylated with adenosine triphosphate (ATP) to form phosphocholine by the *CKI1*-encoded choline kinase. Phosphocholine is then converted to CDP-choline by the reaction catalyzed by the *CCT1/PCT1*-encoded phosphocholine cytidylyltransferase. In the last step, PC is synthesized from CDP-choline and DAG by the *CPT1*-encoded choline phosphotransferase. The enzyme reactions in the CDP-ethanolamine pathway are similar to those in the CDP-choline pathway. The *EKI1*-encoded ethanolamine kinase catalyzes the phosphorylation of ethanolamine to phosphoethanolamine. Phosphoethanolamine is then converted to CDP-ethanolamine by the *ECT1*-encoded phosphoethanolamine cytidylyltransferase. In the last step, PE is synthesized from CDP-ethanolamine and DAG by the *EPT1*-encoded ethanolamine phosphotransferase. The Kennedy pathways become essential when the enzymes in the CDP-DAG pathway are defective or repressed. Mutants defective in the synthesis of PS, PE, or PC require choline for growth in order to synthesize PC via the CDP-choline pathway. Mutants defective in the synthesis of PS and PE can also synthesize PC if they are supplemented with ethanolamine. Ethanolamine is converted to PE via the CDP-ethanolamine pathway, and the PE is subsequently methylated to form PC via the CDP-DAG pathway. The CDP-choline pathway was once viewed as an auxiliary or salvage pathway used by cells when the CDP-DAG pathway was compromised. However, it is now known that the CDP-choline pathway contributes to PC synthesis even when wild-type cells are grown in the absence of exogenous choline. The PC synthesized via the CDP-DAG pathway is constantly hydrolyzed to form free choline and PA through the reaction catalyzed by the *SPO14/PLD1*-encoded phospholipase D. The free choline is incorporated back into PC via the CDP-choline pathway, and the PA is recycled into PC and other phospholipids via CDP-DAG and DAG.

Other Pathways from CDP-DAG

The essential *PIS1*-encoded PI synthase enzyme catalyzes the displacement of the CMP moiety of CDP-DAG with inositol, resulting in the formation of PI. In this reaction, the water-soluble inositol is provided either by receptor-mediated transport from the growth medium or by *de novo* synthesis from

glucose 6-phosphate. When cells are grown in the absence of inositol, glucose 6-phosphate is converted to inositol 1-phosphate, which is catalyzed by the *INO1*-encoded inositol 1-phosphate synthase. Inositol 1-phosphate is then dephosphorylated to form inositol by the *INM1*-encoded inositol monophosphatase. PI is itself a branch point in phospholipid biosynthesis: it is further converted to multiple phosphorylated forms (polyphosphoinositides) by PI kinases, or its phosphoinositol moiety is transferred to ceramide to form sphingolipids. In the CL pathway, CDP-DAG is converted to phosphatidylglycerolphosphate (PGP) by displacement of CMP from CDP-DAG with glycerol 3-phosphate. This reaction is catalyzed by the *PGS1/PEL1*-encoded PGP synthase. PGP is then converted to PG by PGP phosphatase whose structural gene has not yet been identified. In the last step, CL is synthesized from PG and CDP-DAG by *CLS1/CRD1*-encoded CL synthase.

Regulation of Phospholipid Synthesis

Phospholipid synthesis is regulated by a number of factors. These include water-soluble phospholipid precursors (e.g., inositol and choline), nucleotides (e.g., ATP and CTP), lipids (e.g., PA, CDP-DAG, and sphingolipids), growth phase, and phosphorylation. The regulation of phospholipid synthesis by these factors is complex and mediated not only by genetic mechanisms that control the synthesis of mRNA and protein, but also by biochemical mechanisms that modulate enzyme activity.

Genetic Regulation

The effect of inositol supplementation on phospholipid synthesis is a representative example of genetic regulation. When wild-type cells are grown in the presence of inositol, the levels of PS, PE, and PC decrease, while the level of PI increases. These changes in phospholipid composition result from the coordinated regulation of enzyme activities in phospholipid synthesis. Genetic regulation by inositol supplementation affects the enzyme activities by controlling protein content at the transcriptional level. Thus, inositol supplementation represses the transcription of *INO1* and other genes in the CDP-DAG and CDP-choline pathways, whereas inositol starvation stimulates the transcription of the coregulated genes. In many instances, the repressive effect of inositol is enhanced by the addition of choline or ethanolamine to the growth medium. All of the coordinately regulated genes contain a *cis*-acting element, inositol-sensitive upstream activation sequence (UAS_{INO}), in the promoter region. The UAS_{INO} element has the core consensus-binding site (CANNTG) for basic helix-loop-helix (bHLH) proteins. Ino2p and Ino4p are members of the bHLH family of DNA-binding proteins. They bind the UAS_{INO} element as a heterodimer and activate transcription. In addition, Ino2p (but not Ino4p) contains a transcriptional activation domain and is subject to autoregulatory control. Thus, levels of Ino2p appear to be limiting for the formation of the Ino2p/Ino4p heterodimer. By contrast, Opi1p represses the transcription of UAS_{INO}-containing genes. Opi1p contains a leucine

zipper and two glutamine-rich domains that are required for Opi1p repressor activity. Opi1p mediates its negative regulatory activity through the UAS_{INO} element, but not by direct interaction. Instead, *in vitro* data indicate that Opi1p interacts with DNA-bound Ino2p within the leucine zipper domain of Opi1p. In addition, the global repressor Sin3p interacts with the N-terminal region of Opi1p. Studies using mutant alleles of *INO2*, *INO4*, *OPI1*, and *SIN3* support a model whereby these interactions play a role in the expression of UAS_{INO}-containing genes *in vivo*.

In addition to inositol, the transcription of the UAS_{INO}-containing genes is also regulated by growth phase. Cells entering stationary phase reveal the repression of the *INO1* gene expression even in the absence of inositol from the growth media. The stationary phase repression of *INO1* transcription is also controlled by Opi1p. Deletion of the *OPI1* gene results in the derepression of *INO1* transcription in stationary phase.

Some enzymes in phospholipid synthesis are regulated in an opposite fashion: they are upregulated by inositol and derepressed in stationary phase. For example, inositol supplementation results in an increase in the level of DGPP phosphatase (encoded by *DPP1*) activity in both exponential and stationary phase cells. The activity is greater in stationary phase cells when compared with exponential phase cells, and the inositol- and growth phase-dependent regulation of DGPP phosphatase are additive. Analyses of DGPP phosphatase mRNA and protein levels indicate that transcriptional regulation is responsible for the control of the enzyme activity. This regulation is not mediated through a UAS_{INO} element, as the element does not exist in the *DPP1* promoter.

Biochemical Regulation

Alterations in phospholipid synthesis in response to environmental changes are also regulated by biochemical mechanisms. In addition to its role in regulating enzyme expression, inositol regulates the activity of phospholipid biosynthetic enzymes. Inositol inhibits PS synthase activity as a noncompetitive inhibitor, and regulates PI synthase activity through substrate availability. Since both enzymes share the common substrate CDP-DAG, inositol-mediated regulation of PS synthase activity contributes to the synthesis of PI at the expense of PS.

Phosphorylation is a mechanism by which an enzyme activity may be regulated, and indeed several phospholipid biosynthetic enzymes are phosphorylated. For example, the CTP synthetase, choline kinase, and PA phosphatase enzymes are phosphorylated and activated by protein kinase A, whereas the phosphorylation of PS synthase by protein kinase A results in the inhibition of activity. These phosphorylation reactions favor the synthesis of PC via the CDP-choline pathway.

Lipids regulate some phospholipid biosynthetic enzymes. For example, PS synthase activity is stimulated by PA and

inhibited by CL and DAG. The PA phosphatase enzyme is stimulated by CL, CDP-DAG, PI, and DGPP. PS synthase and PA phosphatase are also inhibited by the sphingoid bases phytosphingosine and sphinganine.

Nucleotides regulate the activity of some phospholipid biosynthetic enzymes. For example, choline kinase is subject to allosteric regulation by its substrate ATP and its product adenosine diphosphate (ADP). ATP promotes the oligomerization of choline kinase, whereas ADP inhibits its activity by affecting the catalytic properties of the enzyme and the apparent affinity of the enzyme for substrates ATP and choline. CTP inhibits PS synthase activity by chelating its cofactor manganese ions. CTP synthetase is regulated by CTP product inhibition. The CTP-mediated inhibition of CTP synthetase activity increases the positive cooperativity of the enzyme for UTP. CTP synthetase with a mutation at Glu¹⁶¹ is less sensitive to CTP product inhibition, and cells carrying the mutant CTP synthetase accumulate elevated levels of CTP. Because CTP inhibits PS synthase activity, the higher levels of CTP result in an increase in the utilization of the CDP-choline pathway at the expense of the CDP-DAG pathway.

In this article, we have only scratched the surface in discussing the complex mechanisms that govern the synthesis of membrane phospholipids. Moreover, once phospholipids are synthesized, they must be transported to the various organelles in the cell. These processes are complex, and in some cases, they regulate phospholipid synthesis. The availability of sequence information of the yeast genome continues to facilitate studies on the molecular, genetic, and biochemical mechanisms that regulate phospholipid synthesis.

See also: **Bioenergetics:** Phosphatidylinositol-3-Phosphate; **Lipids Carbohydrates Membranes and Membrane Proteins:** Lipid Bilayer Structure; **Signaling:** Inositol Phosphate Kinases and Phosphatases; Phosphatidylinositol Bisphosphate and Trisphosphate; Phosphoinositide 4- and 5-Kinases and Phosphatases; Phosphoinositide-Dependent Protein Kinases; Phospholipase D.

Further Reading

- Carman GM and Henry SA (1999) Phospholipid biosynthesis in the yeast *Saccharomyces cerevisiae* and interrelationship with other metabolic processes. *Progress in Lipid Research* 38: 361–399.
- Daum G, Lees ND, Bard M, and Dickson R (1998) Biochemistry, cell biology, and molecular biology of lipids of *Saccharomyces cerevisiae*. *Yeast* 14: 1471–1510.
- Greenberg ML and Lopes JM (1996) Genetic regulation of phospholipid biosynthesis in *Saccharomyces cerevisiae*. *Microbiological Reviews* 60: 1–20.
- Henry SA and Patton-Vogt JL (1998) Genetic regulation of phospholipid metabolism: Yeast as a model eukaryote. *Progress in Nucleic Acid Research* 61: 133–179.
- Wu W-I and Voelker DR (2002) Biochemistry and genetics of interorganelle aminoglycerophospholipid transport. *Seminars in Cell and Developmental Biology* 13: 185–195.

Photoinhibition and Photoprotection in Plants, Algae, and Cyanobacteria

G M Giacometti and T Morosinotto, Università di Padua, Padua, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Light harvesting complexes (LHCs) LHCs are proteins belonging to a multigenic family found in almost all photosynthetic eukaryotes. They bind several pigment molecules and are associated to both photosystems I and II. They are responsible for increasing light harvesting efficiency but members of the multigenic family are also involved in photoprotection.

Nonphotochemical quenching (NPQ) When exposed to intense illumination, photosynthetic organisms activate a mechanism for thermal dissipation of energy absorbed in excess. This response is called nonphotochemical quenching, although the term 'feedback de-excitation' has also been proposed. The nomenclature originates from the fact that NPQ is generally measured by monitoring *in vivo* fluorescence and its activation is detected indirectly by a quenching of emitted fluorescence.

Photoinhibition When exposed to intense illumination, photosynthetic organisms experience oxidative stress due to absorption of excess excitation energy. The oxidative species generated react with proteins and pigments of the photosynthetic apparatus. Photoinhibition is defined as the decrease in photochemical efficiency experienced in response to intense illumination due to radiation damages. Photosystem II is the major target of this phenomenon.

Photoprotection When exposed to intense illumination, photosynthetic organisms may experience oxidative stress due to excess excitation energy absorbed. Photosynthetic organisms are able to activate several distinct regulatory

mechanisms to avoid damages from excess illumination. This kind of response is generally defined as photoprotection.

Photosystem I Photosystem I is a multisubunit complex. It exploits light energy to catalyze the electron transport from plastocyanin to ferredoxin. By working in series with photosystem II, it is responsible of light-driven electron transport from water to the final acceptor NADP^+ .

Photosystem II Photosystem II is a multisubunit complex binding multiple pigment molecules, both chlorophylls and carotenoids. Photosystem II exploits light energy to catalyze the most fundamental reaction in the biosphere, the oxidation of water to molecular oxygen and reducing equivalents.

Protein D1 D1 protein is a central subunit of photosystem II. It binds most of the cofactors fundamental for electron transfer within photosystem II. Under illumination, it experiences a very high turnover rate and is central in the repairing cycle of photosystem II.

Xanthophylls and xanthophyll cycle Xanthophylls are oxygenated carotenoid which are commonly present in photosynthetic apparatus of most photosynthetic organisms. Carotenoids provide photosynthetic organisms with constitutive protection from oxygen reactive species and oxidative stress. However in many organisms, the carotenoid composition changes upon intense illumination. Most commonly, in strong light the xanthophyll violaxanthin is reversibly converted into zeaxanthin. Accumulated zeaxanthin is active in photoprotection.

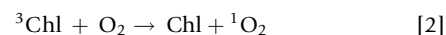
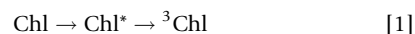
Light Energy Transduction

Solar radiation is the primary source of energy for the biosphere, and oxygenic photosynthesis is the process by means of which light is converted into the chemical energy that living organisms exploit for their needs. When a photon of visible light is absorbed by the pigments of a light-harvesting component of the photosynthetic apparatus, its energy is converted into electronic excitation which can be transmitted to nearby pigments of the antenna moiety, until it is trapped by the special chlorophyll *a* contained in the reaction centers. Here, electronic excitation is converted into charge separation, one electron being transferred from the excited *chl a* (*chl a*^{*}) to a nearby acceptor molecule. This process takes only a few picoseconds, as it has to compete with the thermal relaxation of molecular excitation. Light, like most forms of energy, can be dangerous when supplied in large amounts.

Oxygen, the By-Product of Photosynthesis, Can Be Harmful

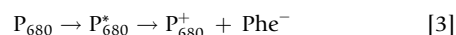
In the photosynthetic apparatus, light absorption in the antenna moiety and charge separation in the reaction centers take place in the presence of molecular oxygen, which is a byproduct of the electron transport activity of photosystem II (PSII), and this creates a potential risk of oxidative stress. Electron transfer reactions in the presence of dioxygen can give rise to the production of highly reactive activated oxygen forms such as singlet oxygen ($^1\text{O}_2$), superoxide ($\text{O}_2^{\bullet-}$), and hydroxyl radical ($\text{OH}^\bullet$).

In particular, the very reactive $^1\text{O}_2$ can easily be created by light in the presence of a photosensitizer such as chlorophyll, which is the main pigment of the photosynthetic apparatus:



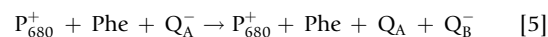
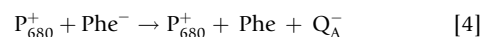
Chlorophyll is very efficient in absorbing visible light, and its excited state Chl^* has a lifetime of a few nanoseconds, long enough to allow excitation transfer within the light-harvesting complex and charge separation in the reaction centers. However, if the excitation energy is not used promptly, a spin transition may take place by intersystem crossing to the lower energy state ^3Chl (reaction [1]). The chlorophyll triplet state, with its much longer lifetime (in the microsecond time range), has plenty of time to react with the ground state of molecular oxygen and to produce its activation to singlet state (reaction [2]). $^1\text{O}_2$ is a highly oxidative agent which can react with the pigments, lipids, and proteins of the photosynthetic apparatus, causing oxidative damage and loss of their function.

Therefore, light is both indispensable to and dangerous to photosynthesis. Photosystem II is continually damaged during photosynthetic electron transport, and the extent of damage increases with light intensity. The main damaging agent is $^1\text{O}_2$, produced at the PSII reaction center, and the mechanism for its production is charge recombination. When the $\text{Chl } a$ at the reaction center (P_{680}) is excited, an electron is transferred to the first acceptor pheophytin (Phe), leading to a charge separated state:

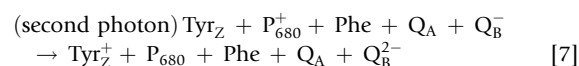
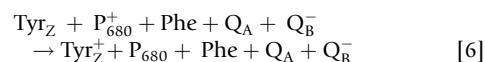


In the physiological electron transfer chain, the charge-separated state is stabilized by secondary electron transfer reaction to the first stable quinone acceptor (Q_A) in 200–400 ns [4]. Quinone Q_A is firmly bound to the D_2 protein of the reaction center and acts as a one-electron acceptor. In the subsequent 300–500 μs , the electron is taken up by the two-electron acceptor Q_B , loosely bound to the D_1 protein of the D_1/D_2 heterodimer of the PSII reaction center [5]. When Q_B becomes doubly reduced by a second electron transfer reaction, driven by a second absorbed photon, two protons are taken up from the stromal solution, and the resulting quinole (QH_2) leaves its

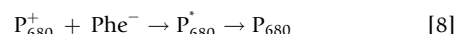
binding site on the D_1 protein for the lipidic phase of the thylakoid membrane (Figure 1):



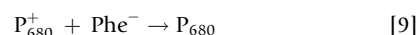
Meanwhile, the charge separation is further stabilized by an electron transfer from the donor side of PSII, that is, Tyr_Z (Y_Z) of the D_1 protein:



Reaction [3], radical pair formation, is an endo-ergonic reaction driven by photon energy, whereas all other electron transfer reactions are downstream, toward increasing redox potentials. Nevertheless, all of them are reversible, and each reaction may proceed backward with a certain degree of probability. In particular, in reaction [3], the charge-separated state may recombine through different pathways. One is an activated pathway:



where charge recombination leads to the excited state of P_{680} , which may eventually relax to ground state by fluorescence emission (retarded fluorescence) or thermal dissipation. Another possibility is direct recombination to ground state:



Both these reactions lead to dissipation of absorbed light energy, thereby decreasing the photosynthetic yield, but they neither activate molecular oxygen nor constitute any threat to the photosynthetic apparatus. However, reaction [9] may also take an alternative route: if the charge-separated state $\text{P}_{680}^+ + \text{Phe}^-$,

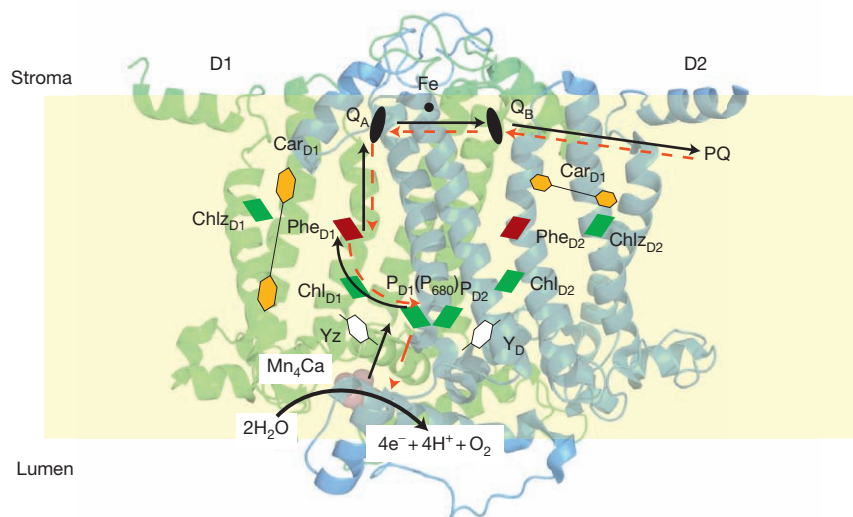
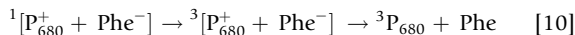
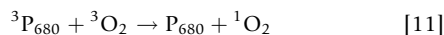


Figure 1 Arrangement of redox cofactors according to PSII structure. Electron transport routes are indicated.

which is normally in its singlet state $^1[P_{680}^+ + \text{Phe}^-]$, is given enough time, it may re-phase its spin state to the lower-lying triplet state $^3[P_{680}^+ + \text{Phe}^-]$ by intercrossing system. This charge-separated state may recombine to the triplet state of P_{680} :



As described above, the triplet state of a chlorophyll molecule can easily activate molecular oxygen to its singlet state:



Unlike the situation of the chlorophyll molecules, which act as absorbing pigments in the light-harvesting apparatus, the triplet state of which is efficiently scavenged by carotenoids, in the special case of the P_{680} chlorophyll of the PSII reaction center, there is no protection from carotenoids, as those located at the reaction center are too far away from P_{680} to be efficient scavengers of its triplet state (Figure 1).

Routes [9] and [10] of charge recombination have approximately the same probability of occurring, and charge recombination at the PSII reaction center therefore involves the production of singlet oxygen with a quite high degree of probability.

Photosynthetic organisms exploit solar radiation to support their metabolism. If illumination is low, as in deep water or under dense leaf canopies, light is a limiting factor for growth, and photosynthetic organisms have evolved a system of light-harvesting complexes (LHCs) which increases the capacity for efficient harvesting of light energy. These antenna systems are composed of pigment-protein complexes which absorb light energy and transfer the excitation to reaction centers, where photochemical reactions occur. Antenna systems vary in different groups of photosynthetic organisms, adapting to the environmental conditions encountered. In cyanobacteria,

they are soluble proteins called phycobilisomes, and bind phycoerythrin and phycobillin as pigments. In plants and eukaryotic algae, antenna proteins are all membrane-embedded proteins belonging to the multigenic Lhc family and bind Chl and carotenoids (Figure 2).

Solar radiation in a natural environment is not constant, and while in some conditions it may be limiting, in others, as at midday in full sunlight, it may saturate photosynthetic capacity. Light harvested in excess causes an increase of the excited states in the antenna pigments. This excess excitation is dangerous for the cells, because it may lead to the formation of reactive oxygen species and eventually photoinhibition. Plants, algae, and cyanobacteria have all evolved the capacity for efficiently repairing radiation damage and recovering from photoinhibition. However, they have also evolved mechanisms to dissipate excessive energy safely and to reduce damage to the photosynthetic apparatus by any reactive oxygen species. All these responses go under the name of photoprotection.

Photoinhibitory Damage to PSII

Most authors agree that singlet oxygen production at the PSII reaction center is the main mechanism of photoinhibition, which may be defined as the loss of photosynthetic activity undergone by oxygenic photosynthetic organisms (plants, algae, and cyanobacteria) when light energy is absorbed at a rate exceeding the overall rate of electron transfer through the photosynthetic electron transfer chain.

In these conditions, the quinone pool in the photosynthetic membrane and the quinone acceptors of PSII become fully reduced. At high irradiance, primary acceptor Q_A is stable in its reduced form Q_A^- , and a charge-separated state, formed by photon absorption, can only recombine, as the probability of double Q_A reduction is very low, even at very high irradiance. Thus, with increasing light intensity, singlet oxygen production at the PSII reaction center also increases, causing oxidative damage and loss of photosynthetic activity.

The D_1 Protein of the PSII Reaction Center is Selectively Degraded

It has long been known that the D_1 protein of the PSII reaction center has a very high turnover compared with other chloroplast proteins, and that its turnover rate increases with light intensity. This has been interpreted as evidence that the D_1 protein is the main target of light-induced photooxidation. According to the most popular model, the damaged D_1 protein is recognized by a set of specific proteases which cut it to pieces. A complex repair cycle, involving new synthesis of D_1 and its insertion into the PSII complex, brings the damaged PSII unit back to renewed full activity.

The D_1 Protein as a Sacrificial 'Fuse' System

A different interpretation has also been proposed: degradation of the D_1 protein is not due to the fact that it is preferentially damaged by the oxidative action of singlet oxygen. There is no

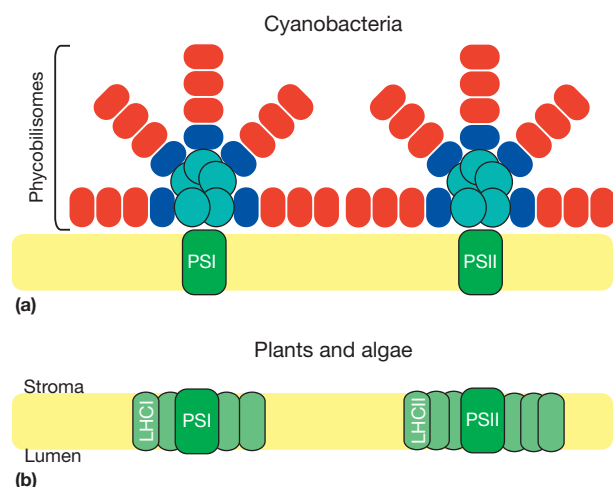


Figure 2 Sketch of antenna complexes in various photosynthetic organisms. (a) Cyanobacteria antenna is composed of soluble pigment-binding complexes, called phycobilisomes, which are associated with both PSI and II. (b) In plants and algae, the antenna system is composed of membrane proteins belonging to the LHC multigenic family. In plants and green algae, distinct subunits, LHCI and LHCII, are specifically associated with PSI and II. In other groups of algae, some antenna proteins may be associated with both photosystems.

explanation, based on experimental evidence, why singlet oxygen should preferentially damage the D₁ protein and not other nearby proteins, for example, D₂, which is symmetrically located with respect to P₆₈₀, should also be damaged to same extent. Nor is there any explanation on how proteases would be activated to recognize and degrade the damaged protein. Thus, protease degradation of D₁ may be a protective mechanism, activated by some suitable signal such as the reduction level of the chloroplast or increased concentration of singlet oxygen. Degradation of D₁, disrupting the integrity of the reaction center, would stop the production of singlet oxygen, thus protecting the remaining components of the PSII complex from irreversible destruction.

Charge Recombination to Ground State as a Protecting Mechanism

The yield of charge recombination, with the formation of ³P₆₈₀ and ¹O₂, is modulated by a change in the midpotential of redox couple Q_A/Q_A⁻. This is because the efficiency of the ¹O₂ producing pathway depends on competition between the formation of the triplet state of the radical pair ³[P₆₈₀⁺ + Phe⁻], forward electron transfer (reaction [4]), and direct recombination of the charge-separated states ¹[P₆₈₀⁺ + Phe⁻] and [P₆₈₀⁺ + Q_A⁻] to form the ground state of P₆₈₀. These processes, which do not lead to production of singlet oxygen, depend on the redox potential of Q_A and on the free energy gap between Phe⁻ and Q_A⁻.

In some conditions, which may be of physiological significance, such as high light flux, conformational adjustments of reaction center proteins – induced, for instance, by acidification of the thylakoid lumen – the midpoint potential of Q_A may increase by several dozen millivolts, hindering forward electron transfer from Q_A to Q_B and favoring direct charge recombination to the ground state of P₆₈₀. It has been experimentally shown, in *in vitro* experiments, that the extent of photodamage suffered by PSII depends on the midpoint potential of Q_A/Q_A⁻.

The Mechanism of Photoinhibition of PSII is Still Controversial

Although the molecular mechanisms for PSII photoinhibition have been studied intensively for decades, there is still significant controversy in the literature as regards the primary source of photodamage to PSII. The main controversy is between authors who support the model based essentially on singlet oxygen forming in the various events of charge recombination and those who assign photodamage to the visible light absorbed by the manganese cluster of the oxygen-evolving complex (OEC) on the donor side of PSII.

In the latter case, it is proposed that the primary event in photodamage to PSII by both UV-B and visible light is due to the photosensitizing action of the Mn cluster. This is based on the similarity of the action spectrum for impairment of the water splitting activity, as monitored by light-induced electron transport from H₂O to dichlorophenolindophenol (DCPIP) in the presence of an intact preparation of thylakoid membranes,

with the absorption spectrum of a dimeric Mn(III/IV) model compound. According to this model, the first harmful action of a high photon flux is impairment of OEC activity, which leaves primary donor P₆₈₀ permanently in its oxidized form P₆₈₀⁺. This type of phenomenon, called 'donor-site photodamage', had already been observed in the case of photoinhibition induced by UV-B light. In such conditions, the light absorbed by chlorophylls starts exerting its harmful action on the electron transfer chain, but degradation of the D1 protein is not due to the formation of singlet oxygen, which is instead involved in impairment of the repair cycle.

Photosystem I Light Damage

PSI is also a potential source of radical species under high light. However, PSI normally undergoes far less light-induced damage than PSII. First of all, charge recombination at the PSI reaction center, with dissipation of energy as heat, is an efficient process here and thus provides an efficient protection mechanism when light is in excess.

Nevertheless on the PSI acceptor side, reduced ferredoxin or Fe-S clusters can be directly oxidized by O₂, leading to the production of O₂^{•-}. This superoxide radical species is rapidly converted into H₂O₂ by superoxide dismutase (SOD). Efficient dismutation of superoxide is critical to protect PSI from photodamage, and *Arabidopsis* mutants in which SOD activity was impaired do show pronounced photosensitivity.

H₂O₂ produced by SOD may be converted into O₂ and H₂O by catalase, which is found in plant peroxisomes but not in chloroplasts. In these organelles, H₂O₂ produced by SOD is rapidly consumed by ascorbate peroxidase, which synthesizes a mono-dehydroascorbate radical (MDA). MDA is extremely reactive and can be reduced by MDA reductase by photosynthetic electrons originating from PSI. Altogether, these reactions are called the water–water cycle, because electrons generated from water oxidation at PS II are used to reduce O₂ back to water at PSI. The water–water cycle eliminates excess electrons and thus plays a photoprotective role, as has been shown in mutants lacking ascorbate peroxidase, which have increased photosensitivity. However, this cycle also contributes to the generation of a ΔpH and thus contributes to balancing the ratio between reducing power and ATP.

As a further protection mechanism, photodamage to PSI has been shown to affect first the antenna complexes, while photochemical activity is preserved. Thus, the PSI protection strategy is quite different from that of PSII: in the former, light damage first affects the light-harvesting antenna, which acts as a screen against harmful agents, leaving the reaction center intact and still capable of performing charge separation. In contrast, the PSII reaction center is damaged first, but is then efficiently and continually restored by the so-called 'repair cycle'.

Dissipation of Excess Singlet States: Nonphotochemical Quenching

The PSII reaction center is continually subjected to oxidative stress and, besides its own protection strategy, can continually repair itself. However, self-repairing is a complex, energetically expensive process. An additional strategy has therefore evolved

in oxygenic organisms, consisting of dissipating excess energy absorbed by the antenna (LHCII) before it is transferred to the reaction center. Reducing the concentration of excited states in the LHCII decreases the probability of generating reactive oxygen species at the reaction center.

An increase in light flux rapidly activates thermal dissipation of chlorophyll excitation, in a process called nonphotochemical quenching (NPQ). This capacity is particularly important in plants, to reduce light damage during the rapid changes in illumination which are common in natural environments. The mechanism is active at the LHC of PSII and involves a light-induced decrease in PSII fluorescence yield (quenching).

When light harvesting exceeds the capacity for light utilization in assimilatory reactions, an accumulation of protons in the thylakoid lumen, associated with electron transport in the photosynthetic chain, causes a rapid decrease in pH which triggers NPQ. This mechanism relies essentially on protein conformational changes, does not require changes in gene expression, and can be activated within seconds of the increase in illumination. The capacity for activating NPQ has been identified in several organisms such as plants, algae, and cyanobacteria, although with great differences in the molecular mechanisms and proteins involved. The differences are due to the fact that the antenna systems are different in these organisms and therefore so are their regulatory mechanisms. As already mentioned, the antenna systems of plants and eukaryotic algae are composed of membrane-embedded proteins belonging to the LHC multigenic family, whereas in cyanobacteria, soluble pigment-binding proteins, called phycobilisomes, are associated with the reaction centers (Figure 2).

Most of the available information comes from plants, in which NPQ activation is known to depend on the LHC-like protein PSBS. Several pieces of evidence support the hypothesis that PSBS does not quench the Chl singlet excited states by itself, but works as a sensor of luminal pH, which turns on NPQ in neighboring antenna proteins, where the quenching of $^1\text{Chl}^*$ actually occurs. This is possible because, under strong illumination, low luminal pH induces protonation of two glutamate residues of PSBS, thereby inducing a conformational change in the antenna proteins which in turn increase the rate of thermal dissipation of chlorophyll excited single states. Full activation of NPQ is also suggested to involve protein re-organization within the thylakoid membrane.

Whereas NPQ in plants is recognized to involve PSBS and antenna proteins, the pigment molecules involved and the precise mechanism of energy dissipation are still debated. Fluorescence quenching has been proposed to depend on direct energy transfer from chlorophylls to carotenoids. This transfer may be possible in antenna complexes after a conformational change occurring in the NPQ condition. An alternative hypothesis suggests that quenching occurs via a charge transfer within a chlorophyll–carotenoid dimer. In this case, during NPQ, a $[\text{Chl}^{\bullet-} \text{carotenoid}^{\bullet+}]$ radical pair state is first generated and then quickly recombines to the ground state, thereby dissipating excitation energy as heat.

Nonphotochemical Quenching in Algae

Although less information is available, most species of algae can also activate NPQ. Their antenna systems are also

composed of members of the LHC multigenic family, although different algal groups have different specific composition. Interestingly, PSBS protein has never been detected in algae so far, thus raising the question as to how NPQ is activated in them. In one green alga, *Chlamydomonas reinhardtii*, another protein, LHCSR, has been identified as responsible of NPQ. LHCSR also belongs to the LHC family, but does not show any similarity with PSBS. LHCSR is widespread among algae and, although molecular evidence is still missing, LHCSR is probably responsible for NPQ activation in many algal groups.

Instead, LHCSR is not found in vascular plants, and some mosses are the only known land organisms retaining this protein. Mosses also have PSBS like all other land plants, suggesting that, during land colonization, photosynthetic organisms lost the LHCSR-dependent NPQ typical of algae after acquiring that dependent on PSBS typical of plants.

It is still not clear whether LHCSR works as a modulator of other antenna proteins, like PSBS, or whether it is a direct quencher of excitation energy. The second hypothesis may be the correct one since, unlike PSBS, LHCSR directly coordinates chlorophylls and carotenoids and may, in principle, act as a quencher of excitation energy.

Nonphotochemical Quenching in Cyanobacteria

A response called NPQ has also been described in cyanobacteria, although its presence was long debated in the past. As already mentioned, light-harvesting apparatus in cyanobacteria is completely different than that in plants and algae, and NPQ does have very different features. In cyanobacteria, NPQ is specifically activated by blue light and not by all visible light, as in plants and algae. However, in both cyanobacteria and plants, NPQ consists of the quenching of antenna fluorescence emission, and antenna complexes (in this case, phycobilisomes) have been shown to be necessary for NPQ activation. NPQ in cyanobacteria also requires the presence of orange carotenoid protein (OCP) which induces quenching by interacting with phycobilisomes in a still unclarified mechanism. OCP is a soluble protein which coordinates a carotenoid molecule found only in cyanobacteria. It appears to act as both photoreceptor and mediator of a photoprotective energy dissipation mechanism, which decreases the amount of energy reaching both photosystems from the phycobilisomes.

Xanthophyll Cycle

Low luminal pH is a key signal of excessive illumination and is very important in regulating photosynthesis. In plants and algae, in addition to NPQ, on a minute timescale, it also activates the conversion of the carotenoid violaxanthin into zeaxanthin. This reaction is catalyzed by a luminal enzyme, violaxanthin de-epoxidase (VDE), a member of the lipocalin family of proteins. VDE is activated at low pH, which induces a conformational change and a change in its location, with its association with the thylakoid membranes where the substrate violaxanthin is found.

The presence of zeaxanthin modulates the photosynthetic apparatus and increases photoprotection at various levels. It strongly enhances NPQ, and zeaxanthin-binding antenna complexes are more efficient in dissipating excess energy as

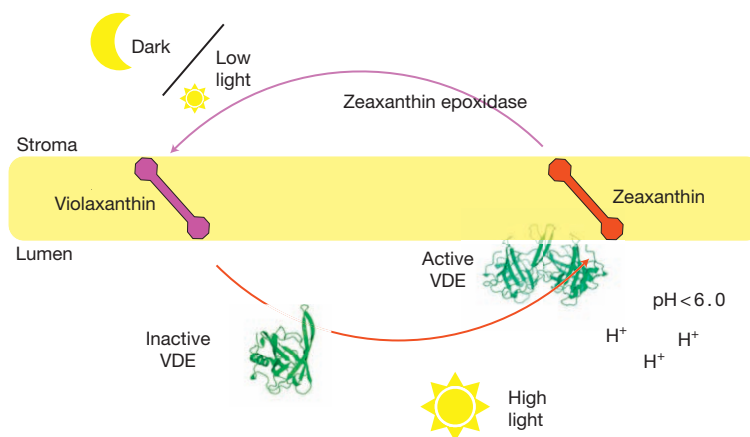


Figure 3 Sketch of xanthophyll cycle. Under strong illumination, xanthophyll violaxanthin is converted into zeaxanthin by luminal enzyme violaxanthin de-epoxidase (VDE). VDE is activated by decrease in lumenal pH resulting from strong illumination; it dimerizes and associates with membrane, where substrate is found. In low or no light, zeaxanthin is converted back to violaxanthin by stromal enzyme zeaxanthin epoxidase.

heat. However, zeaxanthin is also particularly efficient as a scavenger of reactive oxygen species, since it may dissipate any singlet oxygen which may form.

When light goes back to normal intensities, zeaxanthin is converted back to violaxanthin by zeaxanthin epoxidase, closing the so-called xanthophyll cycle (Figure 3).

See also: **Bioenergetics:** Chlorophylls and Carotenoids; Chloroplast Signaling and Redox Control; Chloroplasts; Dynamic Behavior of Photosystem II Light Harvesting; Light-Harvesting Complex I and II: Pigments and Proteins; Photosystem I; Photosystem II: Assembly and Turnover of the D1 Protein; Photosystem II: Redox and Protein Components; Photosystem II: Water Oxidation, Overview.

Further Reading

Eberhard S, Finazzi G, and Wollman FA (2008) The dynamics of photosynthesis. *Annual Reviews of Genetics* 42: 463–515.

Fufezan C, Gross CM, Sjödin M, et al. (2007) Influence of the redox potential of the primary quinone electron acceptor on photoinhibition in photosystem II. *Journal of Biological Chemistry* 282(17): 12492–12502.

Holt NE, Fleming GR, and Niyogi KK (2004) Toward an understanding of the mechanism of nonphotochemical quenching in green plants. *Biochemistry* 43(26): 8281–8289.

Krieger-Liszka A (2005) Singlet oxygen production in photosynthesis. *Journal of Experimental Botany* 56(411): 337–346.

Krieger-Liszka A, Fufezan C, and Trebst A (2008) Singlet oxygen production in photosystem II and related protection mechanism. *Photosynthesis Research* 98(1–3): 551–564.

Li Z, Wakao S, Fischer BB, and Niyogi KK (2009) Sensing and responding to excess light. *Annual Reviews of Plant Biology* 60: 239–260.

Ohnishi N, Allakhverdiev SI, Takahashi S, et al. (2005) Two-step mechanism of photodamage to photosystem II: Step 1 occurs at the oxygen-evolving complex and step 2 occurs at the photochemical reaction center. *Biochemistry* 44(23): 8494–8499.

Vass I and Cser K (2009) Janus-faced charge recombinations in photosystem II photoinhibition. *Trends in Plant Sciences* 14(4): 200–205.

Photoreceptors

K-W Yau, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Circadian photoentrainment The shifting of the (endogenous) circadian clock by light so that it is synchronized with the ambient light–dark cycle.

Depolarization Phenomenon in which the electrical potential inside the cell relative to outside becomes more positive compared to the resting state.

Hyperpolarization The electrical potential inside the cell relative to outside becomes more negative compared to the resting state.

Nonselective-cation channel An ion channel that allows only cations (largely Na^+ , K^+ , and Ca^{2+}) to go through but does not discriminate much between them.

Phototransduction The cellular mechanism that transduces light into an electrical signal in the photoreceptor.

Photoreceptors are cells that can detect light by virtue of a light-absorbing pigment they contain. In most cases, photoreceptors are neurons in the nervous system, and they signal light to the organism for behavioral changes. In some cases, however, photoreceptors are simply modified epithelial cells, such as melanophores in lower animals that can change color by virtue of aggregation or dispersion of intracellular pigmented granules in response to illumination. In humans, all photoreceptors known to date are present in the retina of the eyes. In lower vertebrates such as birds and fish, photoreceptors are also present in the pineal gland and elsewhere in the brain (deep-brain photoreceptors). In some lizards and frogs, photoreceptors are present in the parietal eye (also called the third eye or the frontal organ). In invertebrates, photoreceptors are present in diverse locations either singly or in small or large clusters to form well- or ill-defined eyes. In this article, the discussion is confined to vertebrate photoreceptors in the nervous system.

Retinal Photoreceptors

The well-known retinal photoreceptors in all vertebrates are the rods and cones. Rods are extremely sensitive to light, being able to detect and signal the absorption of a single photon; they are responsible for dim-light vision. Cones are much less sensitive to light, and they also adapt to steady light much more effectively than rods; they are responsible for bright-light vision. In most animals, including humans, rods are predominant in the retina. In occasional diurnal species (such as ground squirrel), however, cones are very abundant.

Rods

In mammals, all rods in a retina are identical. Morphologically, they can be distinguished into the outer segment, the inner segment, the cell body, and the synaptic terminal (Figure 1).

Outer Segment

The outer segment is a modified cilium that has become a highly expanded structure, with hundreds of internalized

membrane disks stacked on top of each other and oriented transverse to the longitudinal axis of the cell. The membrane disks are chock-full of the rod visual pigment, rhodopsin, a transmembrane protein. Rhodopsin consists of rod opsin, the apoprotein, covalently linked to the chromophore, 11-*cis*-retinaldehyde (a derivative of vitamin A). In the mammalian retina, there is a single class of rods, all with the same opsin. The wavelength of maximum absorption (λ_{max}) of rhodopsin is near 500 nm (blue-green monochromatic light).

Inner Segment

The inner segment contains much of the metabolic machinery of the cell, and gives rise to the outer segment. The part of the inner segment adjacent to the outer segment is packed with mitochondria, reflecting the heavy metabolic load on the cell.

Cell Body

The cell body contains primarily the nucleus of the cell.

Synaptic Terminal

The synaptic terminal contacts second-order retinal neurons, including the bipolar and horizontal cells. It releases glutamate as the neurotransmitter. The synapse has a ribbon structure that is typical of most sensory neurons.

Cones

In humans, there are three types of cones, each with a different cone opsin covalently linked to 11-*cis*-retinaldehyde. The three cone pigments have λ_{max} at ~430 nm (blue), 530 nm (green), and 560 nm (yellow), respectively. The functional interactions between the three cone types allow us to perceive colors. The morphology of cones is very similar to that of rods, except that the membrane disks in the outer segment are continuous with the plasma membrane.

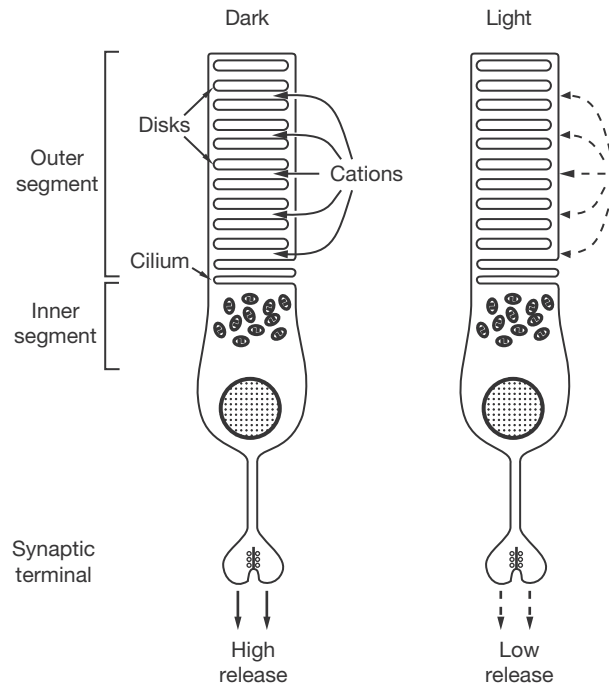


Figure 1 Diagram to show the overall morphology of a retinal rod photoreceptor and its response to light. See **Figure 2** for the phototransduction mechanism underlying the response. Reproduced from Yau K-W (1994) Phototransduction mechanism in retinal rods and cones. *The Friedenwald lecture. Investigative Ophthalmology and Visual Science* 35: 10–32.

Phototransduction

The process by which light triggers an electrical response (the signal that neurons use to communicate with each other) in rods and cones is called phototransduction. It is similar in both photoreceptor classes, and is now very well understood (**Figure 2**). There are nonselective cation channels on the plasma membrane of the outer segment that open (i.e., conduct ions) upon binding intracellular cyclic guanosine monophosphate (cGMP) (cGMP-gated channels). In darkness, the free cGMP concentration in the outer segment is relatively high, thus keeping some of these cGMP-gated channels open. These open channels depolarize the cell and sustain a continuous release of glutamate from the synaptic terminal. Upon absorbing a photon, the 11-*cis*-retinaldehyde in a visual pigment molecule isomerizes to all-*trans*-retinaldehyde, as a result of which the opsin moiety undergoes several extremely rapid conformational changes, one of which (meta-II state) is catalytically active. It activates a G protein called transducin, which, in turn, stimulates a cGMP-phosphodiesterase to hydrolyze cGMP. As a result, the free cGMP concentration decreases, and the cGMP-gated channels close, producing a membrane hyperpolarization that reduces or completely stops neurotransmitter release from the synapse. This signal is picked up by the second-order retinal neurons. The inward, depolarizing membrane current through the open cGMP-gated channels in darkness, in part, involves Ca^{2+} (the rest is mostly Na^{+}), which is steadily pumped out of the cell by an antiporter involving exchange of external Na^{+} for internal Ca^{2+} and K^{+} . The dark Ca^{2+} influx decreases or stops in the light because

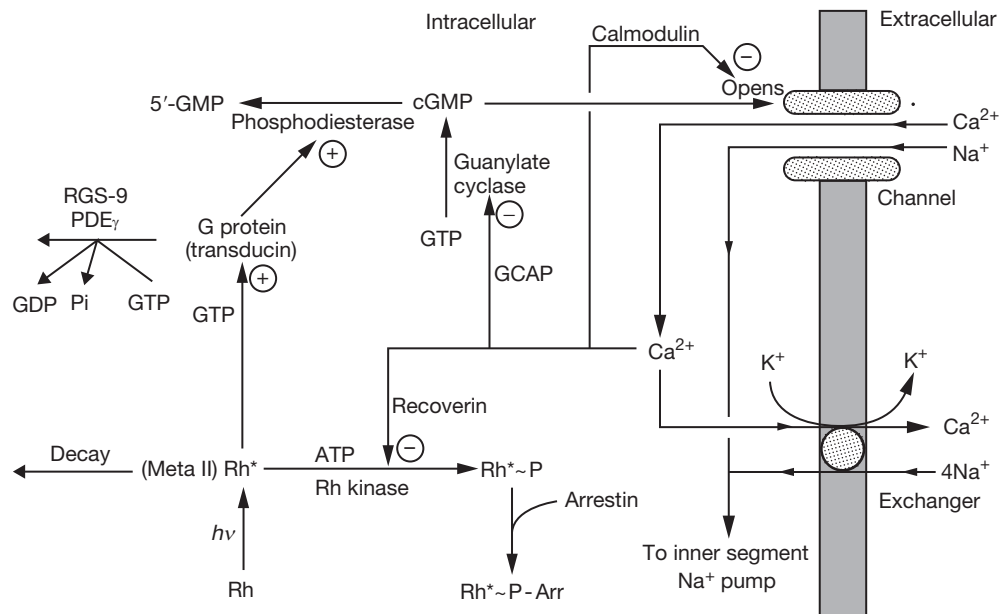


Figure 2 Phototransduction mechanism underlying the light response in rods. GCAP, guanylate-cyclase-activating protein; $h\nu$, photon; Rh, rhodopsin; Rh^* , photoactivated rhodopsin; $\text{Rh}^*\sim\text{P}$, phosphorylated form of rhodopsin; RGS-9, regulator of G-protein-signaling isoform 9; PDE_γ , inhibitory (γ) subunit of the phosphodiesterase; +, stimulation or positive modulation; –, inhibition or negative modulation. RGS-9, together with its associated proteins R9AP (RGS-9-anchoring protein) and $\text{G}_{\beta 5}$ (a previously orphaned member of the β -subunit subfamily of G proteins), form the so-called GTPase-Activating-Protein (GAP) complex. Modified from Koutalos Y and Yau K-W (1996) Regulation of light sensitivity in retinal rods. *Trends in Neurosciences* 19: 73–81.

the cGMP-gated channels close, but the Ca^{2+} efflux through the Na/Ca, K exchange continues, as a result of which the free Ca^{2+} in the outer segment decreases. This Ca^{2+} decrease triggers negative feedback on multiple targets in the phototransduction machinery, leading to an attenuation of the initial cGMP decrease induced by light, and hence reopening of some of the cGMP-gated channels. This negative feedback constitutes the key mechanism for active light adaptation by the photoreceptor, and is particularly important for cones because they function in bright light. The active, meta-II state of the pigment is deactivated by phosphorylation followed by binding of a protein called arrestin, which caps the pigment activity. After a delay, the meta-II decays into the inactive meta-III state and eventually splits into apo-opsin and all-*trans*-retinaldehyde (i.e., the pigment is bleached). The latter is reduced to all-*trans*-retinol (i.e., the corresponding alcohol), then leaves the photoreceptor to the neighboring retinal pigment epithelial cells, where all-*trans*-retinol is reisolomerized/reoxidized to 11-*cis*-retinaldehyde. The 11-*cis*-retinaldehyde returns to the rod and cone outer segments, where it recombines with opsin to form a functional pigment molecule again.

Retinal Nonrod/Noncone Photoreceptors

Very recently, a different class of retinal photoreceptors has been discovered. These are a small subset of retinal ganglion cells, third-order retinal neurons that send light signals to the brain via their axons (optic-nerve fibers). The activity of the great majority of retinal ganglion cells is synaptically driven by the second-order retinal neurons, which, in turn, receive inputs from rods and cones. About 1% of the ganglion cells, however, are also intrinsically photosensitive. They express the pigment, melanopsin, which makes these cells photosensitive (Figure 3). Light depolarizes these cells, leading to the firing of action potentials. The phototransduction mechanism underlying this depolarizing response is still unclear, although a phospholipase C (PLC) pathway is likely involved. Although most retinal ganglion cells convey light information for image-forming vision, such as used in the detection and recognition of objects, the intrinsically photosensitive ganglion cells convey light information mostly for nonimage-forming (accessory) visual functions, such as circadian photoentrainment and the pupillary light reflex. Accordingly, their axons project primarily to brain nuclei involved in these functions.

Most recently, it has been found that some horizontal cells in fish are also intrinsically photosensitive. This is presumably for fine modulation of retinal signal processing. Melanopsin (and possibly additional pigments) appears to be also the pigment underlying the intrinsic photosensitivity. The same may be true in birds.

Parietal-Eye Photoreceptor

The parietal-eye photoreceptor bears remarkable morphological similarity to rods and cones (especially cones). However, there are several interesting differences between them. First, even in isolation, the parietal-eye photoreceptor continues to respond to light after intense illumination, suggesting that

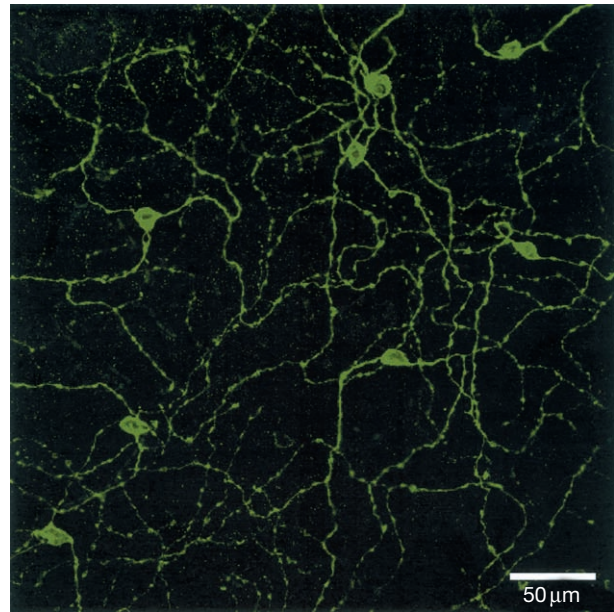


Figure 3 Immunofluorescent labeling of a flat-mounted rat retina with an antimelanopsin antibody. Green color shows the labeling. Stacked confocal images. Reproduced from Hattar S, Liao H-W, Takao M, Berson DM, and Yau K-W (2002) Melanopsin-containing retinal ganglion cells: Architecture, projections and intrinsic photosensitivity. *Science* 295: 1065–1070.

there is a chromophore-recycling system within the cell that allows autonomous pigment regeneration, a feature absent in rods and cones. Second, there appear to be two pigments in the same cell, optimally responding to blue and green light, respectively. The green pigment when active depolarizes the cell, whereas the blue pigment when active hyperpolarizes the cell. The phototransduction mechanism is very similar to that in rods and cones, except that, although the blue pigment activates a cGMP-phosphodiesterase, the green pigment inhibits it. Thus, there is a chromatic antagonism in a given cell. Depending on the relative intensities of green and blue light, the cGMP level rises or falls, causing the cGMP-gated channels to open or close, hence a depolarization or hyperpolarization. The exact function of the parietal eye is unclear, but it may have to do with informing the animal about the progression of the day.

Pineal Photoreceptors

In birds and fish, the pineal gland is intrinsically photosensitive. Morphologically, the pineal photoreceptors roughly resemble rods and cones. Light elicits a hyperpolarizing response from these cells, with an underlying transduction mechanism also similar to that in rods and cones. The pineal gland releases the hormone melatonin, with light curtailing this release. In mammals, the intrinsically light-insensitive pineal also releases melatonin; in this case, however, the curtailment of melatonin release by light is not direct, but indirect via the suprachiasmatic nucleus, a target of optic-nerve fibers.

Deep-Brain Photoreceptors

At least in birds, there is evidence that certain cells in the brain are intrinsically photosensitive. Also, immunocytochemistry and *in situ* hybridization studies suggest that opsin proteins are present in small populations/clusters of neurons in the diencephalon. No other details are known at present.

See also: [Signaling: Cyclic Nucleotide-Regulated Cation Channels](#).

Further Reading

Baylor DA (1987) Photoreceptor signals and vision. The Proctor lecture. *Investigative Ophthalmology and Visual Science* 28: 34–49.

- Cheng N, Tsunenari T, and Yau K-W (2009) Intrinsic light response of retinal horizontal cells of teleosts. *Nature* 460: 899–903.
- Hattar S, Liao H-W, Takao M, Berson DM, and Yau K-W (2002) Melanopsin-containing retinal ganglion cells: Architecture, projections and intrinsic photosensitivity. *Science* 295: 1065–1070.
- Koutalos Y and Yau K-W (1996) Regulation of light sensitivity in retinal rods. *Trends in Neurosciences* 19: 73–81.
- Vigh B, Manzano MJ, Zadori A, et al. (2002) Nonvisual photoreceptors of the deep brain, pineal organs and retina. *Histology and Histopathology* 17: 555–590.
- Xiong WH, Solessio EC, and Yau K-W (1998) An unusual cGMP pathway underlying depolarizing light response of the vertebrate parietal-eye photoreceptor. *Nature Neuroscience* 1: 359–365.
- Yau K-W (1994) Phototransduction mechanism in retinal rods and cones. The Friedenwald lecture. *Investigative Ophthalmology and Visual Science* 35: 10–32.
- Yau K-W and Hardie RC (2009) Phototransduction motifs and variations. *Cell* 139: 246–264.

Photosynthesis

R C Leegood, University of Sheffield, Sheffield, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

CO₂-concentrating mechanism A biochemical (in terrestrial plants) or biophysical pump (in many aquatic photosynthetic organisms) which concentrates CO₂ around RuBP carboxylase-oxygenase (Rubisco) to reduce photorespiration.

Photorespiration The wasteful process by which glycolate 2-phosphate, which results from the oxygenation reaction of Rubisco, is recycled to the Benson–Calvin cycle.

Reaction center A multisubunit protein complex situated in the photosynthetic membrane containing the reaction center chlorophylls as well as other components involved in electron transfer, which together convert the energy of sunlight into a usable chemical form.

Photosynthetic Structures

Leaves contain many different types of cells, including stomata, which regulate gas exchange through the epidermis, and the photosynthetic palisade and mesophyll cells, which are thin walled and have abundant air spaces for the diffusion of gases. Chloroplasts are the organelles which carry out photosynthesis within these cells. The chloroplast has a double-membrane envelope, which regulates transport of metabolites and proteins with the cytosol. It contains thylakoids, a system of membranes, which, in turn, contain the components of photosynthetic light harvesting (including chlorophyll), electron transport, and adenosine triphosphate (ATP) synthesis. The thylakoids form sacs with an internal lumen. Thylakoid sacs can occur singly or in stacks, called grana, which resemble a pile of coins. The soluble phase of the chloroplast is the stroma, which contains enzymes of a wide range of biosynthetic processes (Figure 1).

Harvesting Light

Light Harvesting

The first step in photosynthesis is light absorption by pigments. Besides various types of chlorophyll, these pigments include carotenoids, and open-chain tetrapyrrole bilin pigments found in, for example, cyanobacteria. Chlorophyll is a pigment based on a tetrapyrrole ring, rather like hemoglobin, except that it contains magnesium rather than iron. The ring is linked to a long side chain. Chlorophyll absorbs blue and red light, while it transmits in the green, and hence appears green. In plants, two types of chlorophyll, *a* and *b*, increase the range of wavelengths absorbed. The light energy absorbed by chlorophyll molecules either can be lost as heat or fluorescence, or can be transferred between adjacent chlorophyll molecules by resonance transfer. Chlorophyll and carotenoids associate with proteins in light-harvesting complexes. The light-harvesting complexes have been crystallized from both bacteria and plants. A light-harvesting complex acts like an antenna, similar to a satellite dish, feeding photons into the reaction centers,

which contain a dimeric form of chlorophyll, where charge separation occurs. The concept of an antenna arose from the discovery in 1932, by Emerson and Arnold, that only one CO₂ molecule was produced from about 2500 chlorophyll molecules after a short flash of light.

At a higher level of organization, the leaf and its cells are also adapted to harvest light efficiently. In shade, the photosynthetic apparatus is spread out in large thin leaves, to increase the area of light capture and to allow light to penetrate adequately, and there is more light-harvesting chlorophyll per antenna. Within the leaf, epidermal cells can act to focus light, elongated palisade cells act as light guides, whereas mesophyll cells reflect light, acting like a hall of mirrors and increasing the distance that photons travel, thereby increasing the chance that they will be intercepted by an antenna complex.

Energy Capture

Reaction centers are multisubunit protein complexes situated in the photosynthetic membrane. These complexes contain the reaction center chlorophylls as well as other components involved in electron transfer. The reaction center is the core of the photosynthetic process, converting the energy of sunlight into a usable chemical form. Reaction centers carry out light-driven electron-transfer reactions that result in charge separation across the photosynthetic membrane. Rapid electron transfer to secondary acceptors is necessary to prevent recombination of these separated charges. There are two classes of reaction centers with different terminal electron acceptors, those with Fe₄S₄ clusters (type 1) and those with pheophytin/quinones (type 2). In plants, green algae, and cyanobacteria, both reaction centers are present (in photosystems 1 and 2, respectively), whereas photosynthetic bacteria have only one type of reaction center (e.g., type 1 in green sulfur bacteria and type 2 in purple bacteria). In plants, photosystem 2 contains a chlorophyll dimer which, when in the excited state (P680*), is an extremely strong reducing agent. An acceptor molecule, a quinone, Q, becomes reduced, leaving the positively charged chlorophyll dimer (P680⁺). This is an extremely strong oxidant, so strong that it can extract electrons from water.

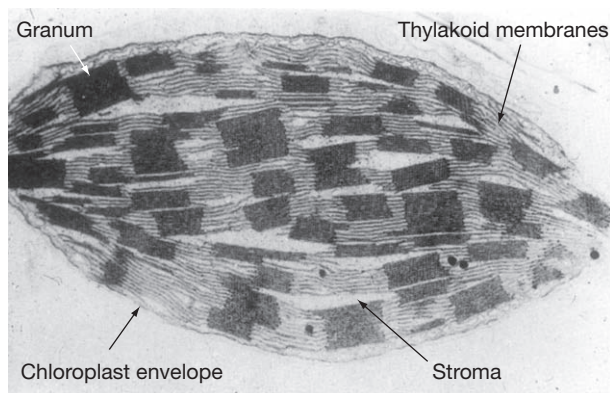


Figure 1 Electron micrograph of a chloroplast from maize.

Water Oxidation

Water oxidation is a unique feature of photosystem 2. Water is the electron donor for photosynthetic electron transport. Each photon absorbed by P680 enables it to extract one electron from a manganese-containing enzyme, in an oxygen-evolving complex. Once four electrons have been extracted, this complex can, in turn, oxidize water, releasing O_2 to the atmosphere, as well as releasing $4H^+$ to the thylakoid lumen. Photosynthetic bacteria utilize other sources, such as H_2S or organic compounds, as electron donors.

Energy Dissipation and Reactive Oxygen Species

Although light is required for photosynthesis, too much light can be harmful. The generation of highly reactive intermediates in oxygenic photosynthesis can lead to the generation of reactive oxygen species. To protect the photosynthetic apparatus from oxidative damage, photosynthetic systems possess antioxidant systems that scavenge reactive oxygen species, as well as mechanisms that regulate photosynthesis to minimize their production. These mechanisms are particularly important in high light, when the absorption of light energy exceeds a plant's capacity for CO_2 fixation (e.g., when CO_2 uptake is limited by stomatal closure brought about by water stress). For example, besides acting as accessory pigments in the antenna, carotenoids also have a photoprotective role. When there is an excess of light energy, carotenoids can quench the excited triplet state of chlorophyll before it reacts with oxygen, forming destructive singlet-state oxygen. Carotenoids can also regulate energy flow in the antenna by dissipating excitation of the chlorophylls as heat (nonphotochemical chlorophyll fluorescence quenching). Energy dissipation is associated with the accumulation of the carotenoid, zeaxanthin, which is interconverted with another carotenoid, violaxanthin, in the xanthophyll cycle. Physical changes, such as chloroplast movements within cells and heliotropic leaf movements, can also reduce or enhance light absorption, and photorespiration can dissipate excess photosynthetic energy.

Electron Transport and ATP Synthesis

There are two photosystems, 1 and 2, involved in photosynthesis in plants, algae, and cyanobacteria. The structures of

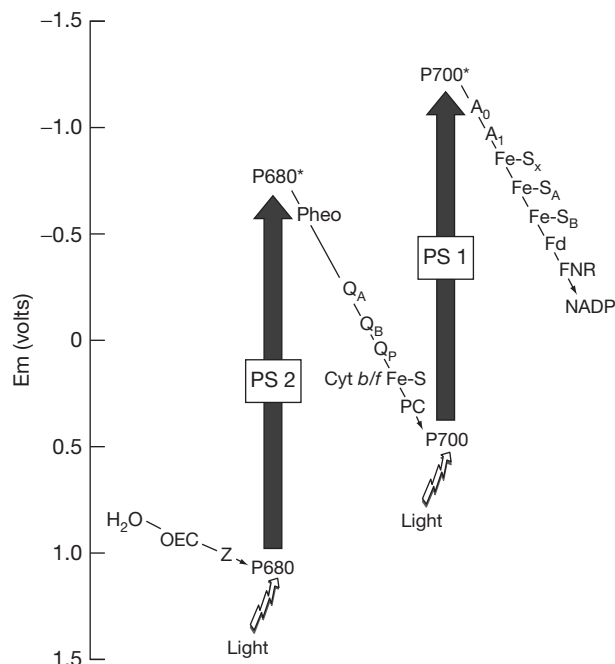


Figure 2 The Z-scheme of photosynthetic electron transport, showing the mid-point redox potentials of electron donors, acceptors, and carriers.

both have been determined. In 1960, Hill and Bendall published the Z scheme, which describes how the two photosystems and the electron carriers are organized with respect to their redox potentials (Figure 2). Electrons from photosystem 2 reduce a number of intermediate electron carriers, such as plastoquinone (PQ), the cytochrome *b/f* complex (including the Rieske iron-sulfur center), and plastocyanin (PC). The reaction center of photosystem 1 (P700) receives its electrons from photosystem II via these carriers and thus oxidizes Q. Photosystem 1 reduces an initial acceptor, A_0 , then a series of Fe-S acceptors and ferredoxin (Fd, another iron-sulfur protein) and, finally, nicotinamide adenine dinucleotide phosphate (NADP) is reduced to NADPH. Electrons may also cycle around photosystem 1, leading to ATP generation, as also occurs in some photosynthetic bacteria.

Electron transport components are organized within the thylakoid membrane so that, as electrons flow from water to NADP, protons are transferred from the stroma to the thylakoid lumen (Figure 3). This is possible because the key component linking the two photosystems is PQ, which is a hydrogen carrier, carrying both an electron (e^-) and a proton (H^+). Electrons from photosystem II reduce PQ, together with protons from the stroma, while oxidation of PQ releases electrons to the cytochrome *b/f* complex and protons to the thylakoid lumen. This proton shuttle is driven by photosynthetic electron transport. Protons are also released to the thylakoid lumen when water is oxidized. These processes decrease the pH in the stroma (from pH 7 in the dark to pH 8 in the light) and acidify the lumen (to about pH 5). A pH gradient is therefore created, which is used to drive ATP synthesis (chemiosmosis). For each molecule of ATP synthesized, $4H^+$ move from the lumen to the stroma. Movement of protons through

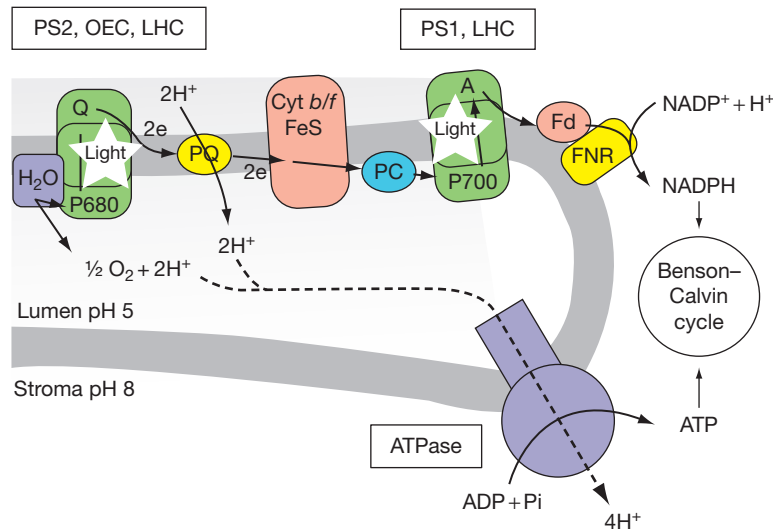


Figure 3 Organization of photosynthetic electron transport within the thylakoid membrane, showing how proton production and translocation generate a proton gradient between the thylakoid lumen and the stroma.

the ATPase causes its central crankshaft-like core to rotate. This rotational movement causes conformational changes in the active sites of other subunits, which result in the synthesis of ATP from adenosine diphosphate (ADP) and Pi.

Carbon Fixation and Photorespiration

The chloroplast is one of the plant cell's most important factories. ATP and NADPH are utilized to drive biosynthetic processes in the chloroplasts, including the reduction of sulfate and nitrite and the synthesis of lipids and amino acids, in addition to CO₂ fixation. All these require an array of transporters at the chloroplast envelope.

Benson–Calvin Cycle

Although often termed the 'dark reaction' of photosynthesis, CO₂ fixation can only proceed in the light, both because it requires photosynthetically generated ATP and NADPH and because many of the enzymes which catalyze the process are themselves activated by light. The Benson–Calvin cycle (or reductive pentose phosphate pathway) is the only mechanism in plants and algae, which can catalyze the net fixation of CO₂, although some bacteria have an alternative mechanism. The Benson–Calvin cycle (Figure 4) was discovered in the early 1950s by Melvin Calvin and co-workers, who supplied algae, such as *Chlorella*, with ¹⁴CO₂. The first product of photosynthesis, containing 100% of the radioactivity at zero time, was a three-carbon compound, glycerate 3-phosphate (hence the term 'C₃' plants). The primary acceptor was identified as ribulose 1,5-bisphosphate (RuBP). The Benson–Calvin cycle comprises three phases: (1) carboxylation by RuBP carboxylase-oxygenase (Rubisco) – RuBP is a five-carbon sugar phosphate which is carboxylated to form two molecules of glycerate-3-P; (2) reduction of glycerate-3-P to triose-P (which requires ATP and NADPH); and (3) regeneration of the acceptor, RuBP, from triose-P in a sugar–phosphate shuffle, which

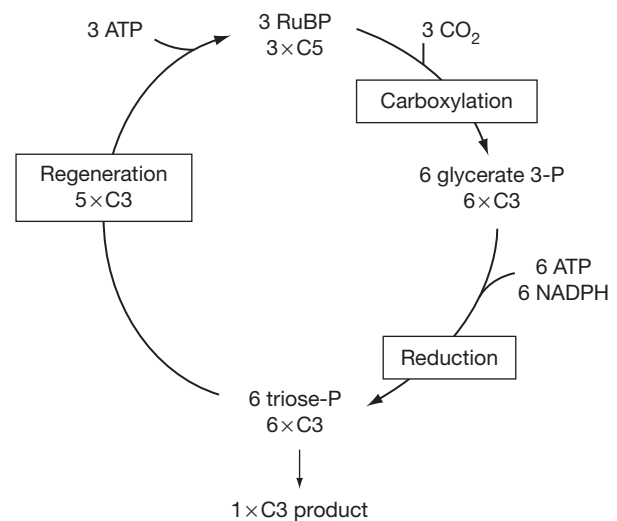
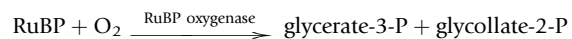


Figure 4 A simplified scheme for the Benson–Calvin cycle, showing the carboxylation of RuBP by Rubisco, reduction of glycerate 3-P to triose-P, and regeneration of the initial acceptor from triose-P.

requires ATP. Triose-P is a simple sugar phosphate produced by the Calvin cycle. Triose-P that is not needed for regeneration of RuBP can be either converted to starch in the chloroplast or exported to the cytosol, via the phosphate translocator, to make sucrose.

Photorespiration

Rubisco is an inefficient enzyme. It has a slow catalytic turnover rate, and about half the soluble protein in leaves is Rubisco, making it the most abundant protein in nature:



Rubisco also catalyzes a side reaction with oxygen, an inevitable consequence of its reaction mechanism. Rubisco evolved in the photosynthetic bacteria billions of years ago in an atmosphere, which was richer in CO_2 and depleted in O_2 in comparison with the present atmosphere. The ratio of oxygenation to carboxylation by Rubisco depends upon the relative concentrations of CO_2 and O_2 and oxygenation increases as the temperature increases. Photorespiration is the process by which two molecules of glycolate-2-P, which is not a metabolite of the Benson–Calvin cycle, are retrieved and converted into two molecules of an amino acid, glycine (C2), and then into one molecule of glycerate-3-P (C3). This involves shuttling of metabolites among the chloroplasts, cytosol, peroxisomes, and mitochondria. During this process, one-quarter of the carbon in glycolate-2-P is lost as CO_2 and ammonia are liberated. Photorespiration is therefore a wasteful process because it both reduces carbon gain and dissipates photosynthetic energy (because CO_2 is first fixed and released again). There is, therefore, a selection pressure on plants to reduce the rate of photorespiration by means of CO_2 -concentrating mechanisms, so as to improve their carbon economy. This is particularly strong when high rates of photorespiration are favored, as at high temperatures. However, a future doubling of ambient CO_2 will reduce photorespiration by about 50%, and it would be completely eliminated by a fivefold rise in atmospheric CO_2 .

CO_2 -Concentrating Mechanisms

C4 Photosynthesis

C4 plants include many tropical grasses and are among the world's most important crop species (maize and sugarcane). Although small in terms of total number of flowering plant species (3%), they constitute about 50% of the 10 000 grass species. Their productivity is high, and C4 grasses in savannah regions (15% of the Earth's vegetated surface) are responsible for about 20% of global photosynthesis. C4 plants have a distinctive leaf anatomy (Kranz anatomy), with chloroplast-rich bundle-sheath cells, which form a gas-tight cylinder surrounding the vascular bundle. A CO_2 pump (the C4 cycle) takes CO_2 from the mesophyll and transfers it into the bundle sheath, which contains Rubisco and the enzymes of the Benson–Calvin cycle (Figure 5). The process raises the concentration of CO_2 in the bundle sheath, and is sufficient to saturate Rubisco with CO_2 and to eliminate photorespiration. Like all pumps, the C4 cycle requires an input of energy in the form of ATP. Recently, two terrestrial plants have been shown to have single-celled C4 photosynthesis, a phenomenon only otherwise known in a few aquatic angiosperms and some diatoms.

The mechanism of photosynthesis in C4 plants was elucidated in the 1960s by Hatch and Slack in Australia. C4 plants are so called because the first product of CO_2 fixation is a C4 organic acid, oxaloacetate, formed by the carboxylation of phosphoenolpyruvate (PEP) by PEP carboxylase. The oxaloacetate is converted to other C4 acids (malate or aspartate) and transferred to the bundle sheath. Transport of metabolites between the mesophyll and bundle sheath occurs by diffusion via plasmodesmata. In the bundle sheath, the C4 acids are decarboxylated to generate CO_2 , and a C3 compound returns

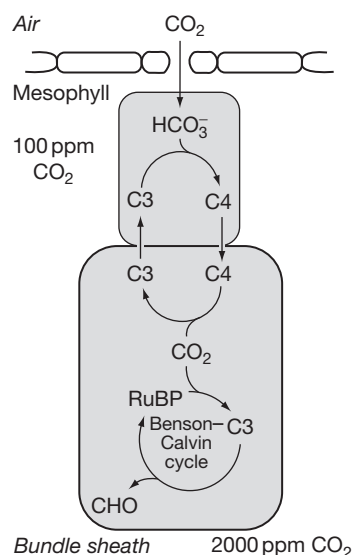


Figure 5 A simplified scheme for the mechanism of C4 photosynthesis, showing how the C4 cycle shuttles C4 acids into to bundle sheath, where they are decarboxylated to raise the CO_2 concentration in the vicinity of Rubisco and thereby suppress photorespiration.

to the mesophyll. The mechanism of decarboxylation differs, with NADP-malic enzyme in the chloroplast (maize), NAD-malic enzyme in the mitochondria (millet), or PEP carboxykinase in the cytosol (e.g., guinea grass).

C4 plants are commonly found in warm- to high-temperature environments, such as tropical grasslands, where photorespiratory rates would be high in C3 plants. C4 plants have double the water-use efficiency of C3 plants because photosynthesis can operate at low intercellular concentrations of CO_2 , and hence at lower stomatal conductances. Nitrogen-use efficiency is also improved because Rubisco is used more efficiently, due to the suppression of photorespiration.

Aquatic Photosynthesis

The aquatic environment poses serious problems for the acquisition of CO_2 . First, the diffusion of CO_2 in water is 10 000 times slower than in air. Second, inorganic carbon is present in water as CO_2 , HCO_3^- , and CO_3^{2-} , the relative amounts of which depend upon the pH of the water. Most inorganic carbon in alkaline waters is therefore present as carbonate and bicarbonate. Photosynthesis by aquatic plants, algae, and phytoplankton may raise the local O_2 concentration and enhance photorespiration. Aquatic organisms (cyanobacteria, micro- and macro-algae, and aquatic angiosperms) have developed means of taking up dissolved inorganic carbon (CO_2 and HCO_3^-) actively from their environment. They can concentrate it internally around Rubisco in specialized structures, such as carboxysomes and pyrenoids, and thus suppress photorespiration.

Crassulacean Acid Metabolism

Leaves are faced with a difficult task. Leaf cells have thin walls and the intercellular spaces are saturated with water vapor.

As leaves assimilate CO₂ (5000 l of air are needed to manufacture 1 g of sucrose), they need to restrict the amount of water lost from the leaf via the stomata. Therefore, there is a conflict between the requirements for CO₂ fixation and water conservation. Crassulacean acid metabolism (CAM) is a photosynthetic adaptation to periodic water supply, occurring in plants in arid regions (e.g., cacti) or in tropical epiphytes (e.g., orchids and bromeliads). CAM plants close their stomata during the day and take up CO₂ at night, when the air temperature is lower. Water loss can be lowered by an order of magnitude. CAM occurs in between 5% and 10% of plants and is always associated with succulence, at least at a cellular level.

Although the biochemistry is similar to that of C₄ plants, two carboxylations are now separated in time rather than in space. Malic acid is synthesized from carbohydrates, via PEP carboxylase, at night, and is stored in the vacuole. During the day, malate is decarboxylated, via PEP carboxykinase or NAD (P)-malic enzymes in the cytosol, and CO₂ is refixed via the Benson–Calvin cycle and carbohydrates are reformed. This process occurs behind closed stomata. The internal concentration of CO₂ is raised as high as 10 000 ppm, which also suppresses photorespiration. CAM plants show a high degree of metabolic flexibility. Seedlings and well-watered plants may show little or no CAM and perform C₃ photosynthesis, opening their stomata during the day. This allows increased carbon gain during periods of water availability or during seedling establishment. Water or salt stress can then induce CAM, switching on gene expression for synthesis of the component enzymes.

See also: **Bioenergetics:** Chlorophylls and Carotenoids; Chloroplasts; Green Bacteria: The Light-Harvesting Chlorosome; **Metabolism Vitamins and Hormones:** Photosynthetic Carbon Dioxide Fixation; **Protein/Enzyme Structure Function and Degradation:** AAA-ATPases.

Further Reading

- Blankenship RE (2002) *Molecular Mechanisms of Photosynthesis*. Oxford: Blackwell Science.
- Edwards GE, Franceschi VR, and Voznesenskaya EV (2004) Single-cell C₄ photosynthesis versus the dual-cell (Kranz) paradigm. *Annual Review of Plant Biology* 55: 173–196.
- Giordano M, Beardall J, and Raven JA (2005) CO₂ concentrating mechanisms in algae: Mechanisms, environmental modulation, and evolution. *Annual Review of Plant Biology* 56: 99–131.
- Li Z, Wakao S, Fischer BB, and Niyogi KK (2009) Sensing and responding to excess light. *Annual Review of Plant Biology* 60: 239–260.
- Nelson N and Yocum CF (2006) Structure and function of photosystems I and II. *Annual Review of Plant Biology* 57: 521–565.
- Rutherford AW and Faller P (2001) The heart of photosynthesis in glorious 3D. *Trends in Biochemical Sciences* 26: 341–344.
- Sage RF and Monson RK (1999) *C₄ Plant Biology*. San Diego, CA: Academic Press.
- Spreitzer RJ and Salvucci ME (2002) Rubisco: Structure, regulatory interactions, and possibilities for a better enzyme. *Annual Review of Plant Biology* 53: 469–475.
- Sproviero EM, Gascón JA, McEvoy JP, Brudvig GW, and Batista VS (2007) Quantum mechanics/molecular mechanics structural models of the oxygen-evolving complex of photosystem II. *Current Opinion in Structural Biology* 17: 173–180.
- von Ballmoos C, Wiedenmann A, and Dimroth P (2009) Essentials for ATP synthesis by F₁F₀ ATP synthases. *Annual Review of Biochemistry* 78: 649–672.
- Walker DA (1992) *Energy, Plants and Man*. Brighton: Packard Publishing. (Also available on CD or as a free download from Oxygraphics, Brighton, UK.)
- Weber APM, Schwacke R, and Flügge U-I (2005) Solute transporters of the plastid envelope membrane. *Annual Review of Plant Biology* 56: 133–164.

Photosynthetic Carbon Dioxide Fixation

M J Paul, Rothamsted Research, Harpenden, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Calvin cycle, reductive pentose phosphate pathway, or photosynthetic carbon reduction cycle Series of reactions that reduce CO₂ into triose phosphate in the chloroplast from which sugars, amino acids, and other plant products are synthesized.

Chloroplast Organelle in which light reactions and Calvin cycle take place.

Feedback regulation Regulates carbon flow through the Calvin cycle in response to the demand for photosynthate in end-product synthesis. It can operate both at the level of recycling inorganic phosphate back to the Calvin cycle and

at the level of sugar-regulated gene expression of the photosynthetic apparatus.

Feedforward regulation Enables the Calvin cycle to respond sensitively to light by affecting the synthesis of Calvin cycle enzymes and their activity.

Light reactions They drive electron flow which leads to the reduction of nicotinamide adenine dinucleotide phosphate (NADP) to reduced form of nicotinamide adenine dinucleotide phosphate (NADPH) and synthesis of adenosine triphosphate (ATP) necessary for the operation of the Calvin cycle.

Light Provides Energy (Adenosine Triphosphate) and Reducing Power (Reduced Nicotinamide Adenine Dinucleotide Phosphate) for CO₂ Fixation

Photons in the range of 400–700 nm – except green light, which is reflected – drive photosynthesis through provision of adenosine triphosphate (ATP) and reduced form of nicotinamide adenine dinucleotide phosphate (NADPH). Light also regulates the Calvin cycle by the thioredox system (see later). Photons interact with electrons in the chlorophylls of light-harvesting antenna in thylakoid membranes of chloroplasts. There are two light-absorbing photosystems that cooperate in the reduction of nicotinamide adenine dinucleotide phosphate (NADP) by electrons derived from water. Energy is passed from these to reaction centers of photosystem II, where electrons are excited and passed to an acceptor and down an electron transport chain to photosystem I. Electrons and protons pass to NADP, yielding NADPH. The open reaction center of photosystem II is filled with electrons from water, from which O₂ is evolved. A pH gradient across the thylakoid membrane from H⁺ accumulation in the thylakoid lumen provides the impetus for ATP synthesis by ATP synthase in photophosphorylation (Figure 1).

The Path of Carbon in Photosynthesis

The sequence of reactions that fix CO₂ into triose phosphate using ATP and NADPH from these light reactions was determined using C14 radioisotopes by Melvin Calvin and colleagues, for which the Nobel prize for chemistry was awarded. This sequence of reactions is now known as the Calvin cycle, reductive pentose phosphate pathway, or photosynthetic carbon reduction cycle. It consists of three distinct phases. First, the fixation of CO₂ into 3-phosphoglyceric acid (3-PGA) (carboxylation). Second, the reduction of 3-PGA into

the triose phosphate, glyceraldehyde-3-phosphate (GAP). The triose phosphate is exported to the cytosol for sucrose synthesis or remains in the chloroplast for starch synthesis or the other biosynthetic reactions that occur there. Five-sixths of the carbon fixed, however, is retained within the cycle itself to allow regeneration of the CO₂ acceptor, ribulose 1,5-bisphosphate (RuBP), in the third, regenerative phase of the cycle (Figure 2).

Carboxylation Phase

Carboxylation is catalyzed by ribulose bisphosphate carboxylase/oxygenase (Rubisco), which can constitute up to 50% of the soluble protein in a leaf and is probably the Earth's most abundant protein. The combination of CO₂ with RuBP, a five-carbon compound, yields two molecules of the three-carbon compound 3-PGA. Many enzymes will bind to molecules in addition to the ones they have primary affinity for, if they are present at high enough concentrations. Present-day atmospheric O₂ concentration is 1000 times higher than that of CO₂, and oxygenation of RuBP also occurs. This produces one molecule of 3-PGA and one molecule of phosphoglycolate (glycolate 2-P). The formation of phosphoglycolate represents carbon lost to the Calvin cycle and a series of reactions serve to return this carbon by converting phosphoglycolate to 3-PGA in a process termed as 'photorespiration' because CO₂ is evolved (Figure 1). These reactions involve the peroxisome where glycolate is converted into glycine, and mitochondria where glycine is converted into serine where CO₂ and ammonia are evolved. Serine then returns to the peroxisome for conversion to 3-PGA and exports back to the chloroplast. Ammonia is reassimilated through a reaction with glutamate and ATP to form glutamine catalyzed by glutamine synthetase in chloroplasts. Glutamine synthetase then catalyzes the formation of glutamate from glutamine and α -ketoglutarate. Oxygenation

Chloroplast

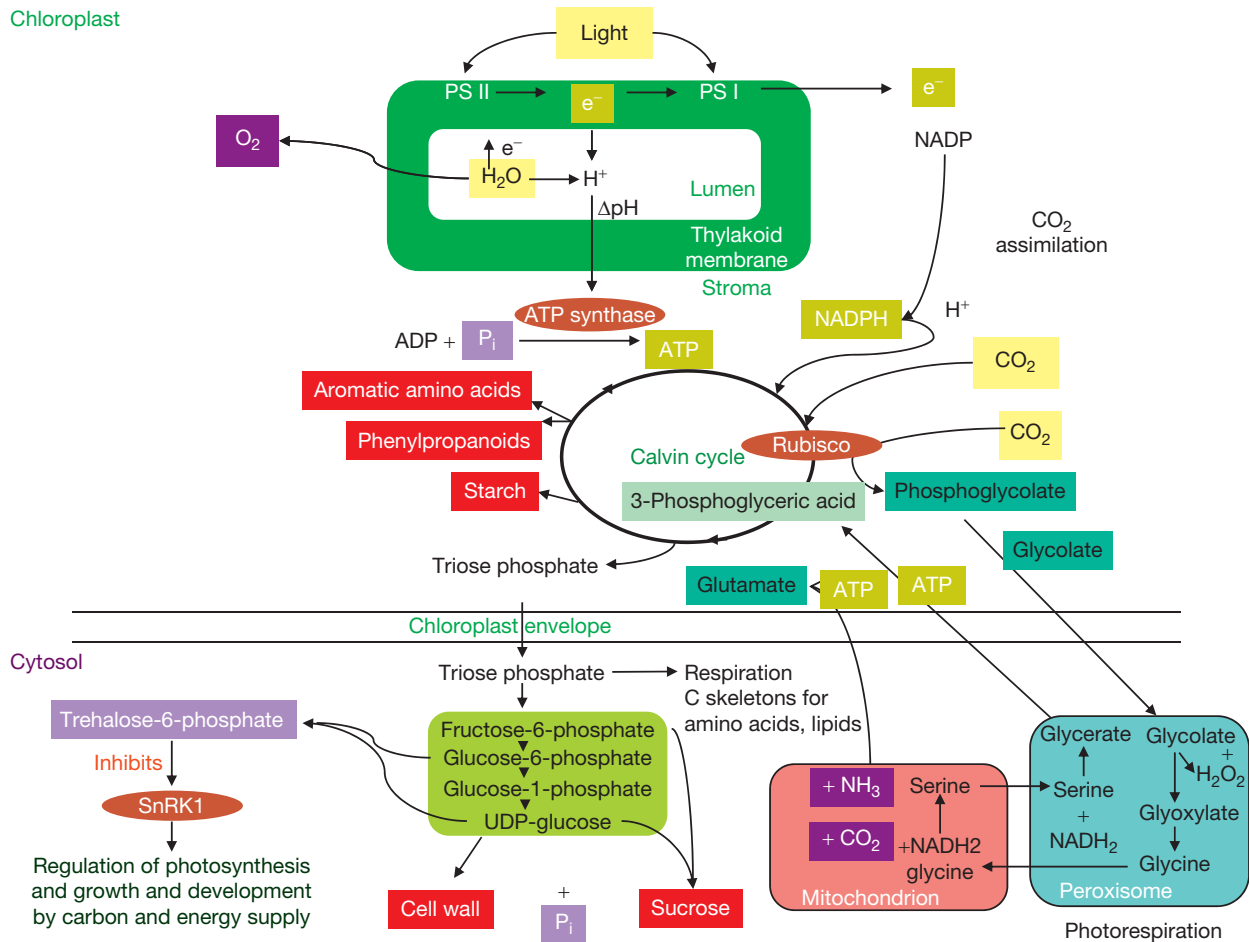


Figure 1 Schematic representation of photosynthetic metabolism summarizing the principal components of light reactions, CO₂ fixation into the Calvin cycle, photorespiration, and the synthesis of end products (sucrose and cell wall). Light causes feedforward regulation of the Calvin cycle. Feedback regulation by carbon fixed in photosynthesis occurs through regulation of the protein kinase, SnRK1.

of Rubisco and the retrieval of carbon from phosphoglycolate are wasteful of carbon and ATP and are also costly in terms of investment in the catalytic machinery required. Early evolution of Rubisco occurred when there was no oxygen in the atmosphere. When oxygen appeared more than a billion years later, the complexity of the Rubisco protein may have made it too difficult to eliminate oxygenase activity linked to phosphoglycolate production.

Supplementary Carboxylation Pathways – C₄ and Crassulacean Acid Metabolism

Some plants have evolved a supplementary carboxylation reaction catalyzed by phosphoenolpyruvate carboxylase where minimal oxygenation occurs. Such plants are known as C₄ plants because the first products of CO₂ fixation are carboxylic acids (oxaloacetic acid and malic acid) containing four carbon atoms. A high concentration of CO₂ generated from the decarboxylation of these acids at the site of the Calvin cycle minimizes the oxygenation reaction of Rubisco, and hence suppresses photorespiration. Other specialized plants known as Crassulacean acid metabolism (CAM) plants fix CO₂ into C₄

acids at night, which are then decarboxylated during the day. These metabolic adaptations of C₄ and CAM ensure a high daytime intercellular CO₂ concentration, which in addition to the benefit of maximizing carboxylation maintains a small stomatal aperture, conserving water, an overriding selection force in some environments. In CAM plants, stomata can be shut for much of the day and is observed in more extreme xerophytes such as cacti and succulents. The C₄ pathway is a major evolutionary success accounting for about a quarter of terrestrial CO₂ fixation. C₄ plants dominate tropical grasslands and savannahs, and include the important crops – sugarcane, maize, and sorghum.

Reduction Phase

The 3-PGA produced in the carboxylation phase (and returned from photorespiration) is reduced to triose phosphate, GAP, in two steps. The first, a phosphorylation step, requires ATP. The 1,3-bisphosphoglycerate formed is then reduced with NADPH to GAP. The high ATP and NADPH concentrations associated with photosynthesis shift these reactions in favor of triose phosphate.

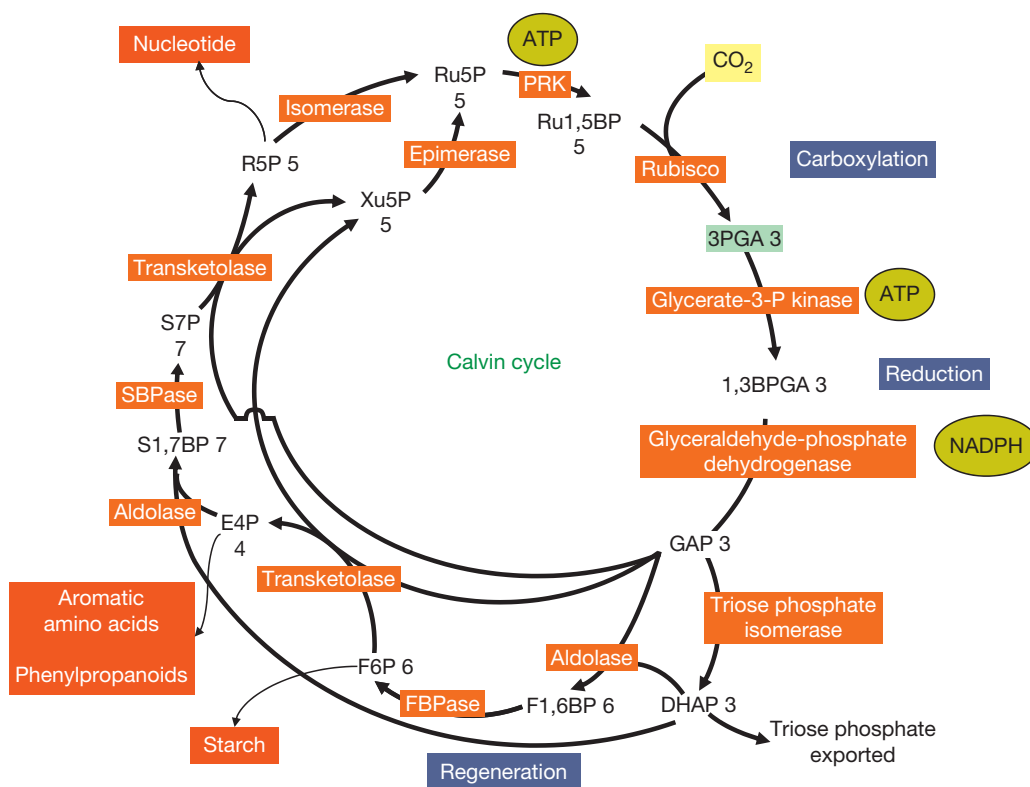


Figure 2 Calvin cycle of reactions that incorporate CO_2 into triose phosphate and which regenerate ribulose-1,5-bisphosphate (Ru1,5BP) for continued CO_2 fixation. Numbers after compounds indicate number of carbon atoms. 3-PGA, 3-phosphoglyceric acid; 1,3 BPGA, 1,3 bisphosphoglyceric acid; GAP, glyceraldehyde-3-phosphate; DHAP, dihydroacetone phosphate; F1,6BP, fructose-1,6-bisphosphate; F6P, fructose-6-phosphate; E4P, erythrose-4-phosphate; S1,7BP, sedoheptulose-1,7-bisphosphate; S7P, sedoheptulose-7-phosphate; R5P, ribose-5-phosphate; Ru5P, ribulose-5-phosphate; Xu5P, xylulose-5-phosphate; Rubisco, ribulose-1,5-bisphosphatase; FBPase, fructose-1,6-bisphosphatase; SBPase, sedoheptulose-1,7-bisphosphatase; PRK, phosphoribulokinase.

Regeneration Phase

Carbon leaves the Calvin cycle for the synthesis of other products, such as triose phosphate for the synthesis of sucrose in the cytosol. Five-sixths of the carbon fixed, however, is retained within the cycle itself to resynthesize the acceptor molecule for CO_2 fixation, RuBP to keep the cycle going. In this series of reactions, five trioses are converted into three pentoses. The first part involves the conversion of GAP to fructose-6-phosphate (F6P). Then the carbon atoms in F6P, GAP, and dihydroxyacetone phosphate are rearranged into the pentose, ribulose-5-phosphate (Ru5P). In the final step, Ru5P is converted into RuBP in an ATP-requiring step. Similar to the rest of the cycle, this phase is highly regulated and can strongly affect the ultimate rate of CO_2 fixation. An important branch point in this segment of the cycle, catalyzed by transketolase links primary and secondary metabolism. The substrates and products of transketolase can supply carbon (erythrose-4-phosphate) for the synthesis of aromatic amino acids and phenylpropanoids (compounds important for flavor, texture, and color), nucleotide synthesis (R5P), and starch (F6P), in addition to the regeneration of RuBP.

Regulation of the Calvin Cycle

Metabolic regulation enables the balance between substrate and product of enzyme-catalyzed reactions to be maintained

so that ordered metabolic flow can occur in response to developmental requirements and environment. The Calvin cycle is regulated by light, which ensures its temporal separation from other reactions, such as the oxidative pentose phosphate pathway, also present in the chloroplast stroma, which would compete for the same substrates. Light regulates the cycle through provision of ATP and NADPH, which in turn regulate those steps dependent on them. Light also causes an increase in the pH of the chloroplast stroma to 8, the pH optimum for Calvin cycle enzymes. Pumping of protons into the thylakoids is also coupled with flow of magnesium necessary for enzyme activity into the stroma. A further way that light regulates the cycle is through photosynthetic electron transport. Reducing equivalents from photosystem I via ferredoxin–thioredoxin reductase reduce the thiol groups of GAP dehydrogenase (GAPDH), fructose-1,6-bisphosphatase (FBPase), sedoheptulose-1,7-bisphosphatase (SBPase), and phosphoribulokinase (PRK), increasing their activity. The net effect of the ferredoxin–thioredoxin system is the upregulation of the Calvin cycle in the light and downregulation in the dark, whereas the oxidative pentose phosphate pathway and glycolysis are regulated in the opposite manner in response to light and dark. Rubisco is activated by CO_2 and magnesium in a process called carbamylation, a process promoted by an activating protein, Rubisco activase. Rubisco activase also facilitates the release from the active site of Rubisco of sugar phosphates, such as RuBP and 2-carboxyarabinitol 1-phosphate,

a transition state analog of the carboxylation reaction of Rubisco. Rubisco activase is controlled in turn by ATP/adenosine diphosphate (ADP) ratio and is necessary for high rates of photosynthesis. Enzymes of the Calvin cycle are associated as enzyme complexes. For example, GAPDH and PRK are linked by the protein CP12. The precise role of the association of this complex is not yet known, but through aggregation into a supramolecular complex is likely to play a specific role in light/dark regulation.

Are Some Calvin Cycle Enzymes More Important in Regulating Carbon Flux than Others?

The equilibrium position of reactions catalyzed by Rubisco, NADP-GAP dehydrogenase, SBPase, FBPase, and PRK, because of their regulation, lies very much in favor of the end product. These enzymes are termed as irreversible enzymes because of the large amount of energy that would be required to reverse the reaction. Such enzymes have traditionally been seen as key enzymes because once carbon has passed through the reaction catalyzed by them it cannot return. Since the late 1980s, genetic modification has enabled individual enzymes to be targeted through the use of antisense RNA to decrease the synthesis of target enzyme. Lines of transgenic plants exhibiting progressive decreases in activity of an enzyme have enabled metabolic control analysis to be undertaken and flux control coefficients for individual enzymes assigned. For the Calvin cycle a change in enzyme activity can then be related to flux through the pathway by measuring CO₂ uptake by leaves using infrared gas analysis. The flux control coefficient is given as $C = \delta J/J$, where J is the original flux and δJ the changed flux. This coefficient can also be written as $C = \delta E/E$, where E is the original enzyme activity and δE the changed enzyme activity.

Therefore, if a 50% decrease in enzyme activity achieves a 25% decrease in CO₂ uptake, then the flux control coefficient is 0.5 (25/50); essentially, half the control of the pathway resides with that enzyme. If a 50% decrease achieves no change in flux, then the flux control coefficient is zero and the enzyme is in excess under those conditions. Over the course of a day, and during the life cycle of a plant, an enzyme may exhibit a range of flux control coefficients. Variation during the course of a day will be caused largely by changes in illumination, whereas over the life cycle by nitrogen availability and changes in light environment. Both light and nitrogen affect the amount of enzyme synthesized, and light also affects enzyme activity to different extents for different enzymes, so relative contributions to flux control change. Other aspects of environment, for example, temperature and phosphorus nutrition, are also influential.

Calvin Cycle Enzymes That Most Limit Flux

The first enzyme for which detailed flux control analysis was conducted using transgenics was Rubisco. Quite surprisingly at the time, in view of the perceived importance of this enzyme, the first data showed that removing half of it with

antisense resulted in little change in CO₂ fixation. However, when plants were grown at high light conditions, the flux control coefficient rose and, combined with low nitrogen, could reach 0.8. Rubisco is the enzyme that imposes the greatest potential limitation of Calvin cycle carbon fixation. SBPase is another enzyme with a potentially high flux control coefficient at high light. A major surprise was that non-regulated equilibrium enzymes, aldolase and transketolase, could have flux control coefficients around 0.3 under high light conditions. Other enzymes in the cycle when decreased up to 50% have showed no large impact on flux. The most extreme example of this is PRK, where a decrease up to 85% has no impact on CO₂ fixation. However, the flux control coefficient for this enzyme could increase to 0.25 if plants grown in shady conditions were suddenly exposed to high light.

The important message to come from this work is that both highly regulated and unregulated equilibrium enzymes could control flux, and the control exerted by individual enzymes varies according to the conditions. The work supported the von Caemmerer and Farquhar 1981 model, which proposed that photosynthesis was colimited between carboxylation (Rubisco activity) and the capacity to regenerate RuBP. Control was seen as poised between carboxylation and regeneration, but under saturating light, the balance would shift toward Rubisco and under low light toward regeneration. The research using antisense confirmed this assumption and gave useful indications as to which enzymes in the regeneration phase would be most likely to increase rates of CO₂ fixation if their activities could be increased. Thus, there would most likely be benefit if one or more of activities of SBPase, aldolase, or transketolase could be increased. Increasing amounts of Rubisco may be problematic because of the large nitrogen investment required.

Genetic Modification of Photosynthesis to Improve Growth and Yield

Direct Targeting of Calvin Cycle

In support of the findings of flux control analysis, expression of a bifunctional SBPase/FBPase from a cyanobacterium resulted in increased photosynthesis and biomass production. This was the first report of an increase in activity of one or more Calvin cycle enzymes benefiting photosynthesis and biomass production. There is also evidence that overexpressing the plant SBPase improves photosynthetic rates. Improved photosynthesis has also been achieved through expressing a cyanobacterial gene in plants involved in HCO₃⁻ accumulation, which releases CO₂ within the leaf for CO₂ fixation and hence improves carboxylation. Attempts are being made to improve crop productivity with superior forms of Rubisco from algae and through introduction of the C4 pathway. These former targets constitute direct targeting of the Calvin cycle. Another way to potentially increase CO₂ fixation is through targeting mechanisms that use the reduced carbon produced in photosynthesis downstream of the Calvin cycle. It has been known that the use of carbohydrate can regulate photosynthesis through the so-called feedback mechanisms. If such

mechanisms could be targeted this may represent a means to increase photosynthesis.

Feedback Regulation of the Calvin Cycle

In addition to the sophisticated control mechanisms that relate CO₂ fixation to light availability (feedforward regulation), feedback mechanisms relate Calvin cycle activity to the demand for the end products of photosynthesis. Inorganic phosphate (P_i) is a component of phosphorylated intermediates in pathways and is liberated when end product, for example, sucrose, is synthesized (sucrose contains no P_i). This recycling of P_i is necessary to ensure that metabolic flow can continue. Photophosphorylation and ATP synthesis are particularly sensitive to P_i supply. Slow recycling of P_i caused by slow end-product synthesis at low temperature, for example, can restrict the recycling of P_i and so limit photosynthesis.

Demand for end product is also determined by the genetic potential of the plant for growth and by other environmental factors, particularly nitrogen, which facilitates high growth rates. The assimilation of nitrogen into amino acids requires carbon skeletons derived from triose phosphates supplied by the Calvin cycle. An impairment of nitrogen supply slows growth, which can lead to a buildup of sugar in leaves, because photosynthates are no longer being used in growth and nitrogen assimilation. It was shown in 1990 that seven maize photosynthetic genes were repressed by glucose or sucrose in a maize protoplast system. Further evidence of control of photosynthetic gene expression by sugars has been obtained where sugars were made to accumulate in leaves due to genetic modification, environmental treatments (high CO₂ and low N₂, in particular), and through direct feeding of sugars to leaves. There have been advances recently in understanding mechanisms through which sugars regulate photosynthetic gene expression. One such mechanism involves a protein kinase called sucrose nonfermenting-1-related protein kinase 1 (SnRK1) of the sucrose nonfermenting-1 (SNF1)-related group of conserved protein kinases, found in all eukaryotes that includes adenosine monophosphate (AMP)-activated protein kinase (AMPK) of mammals and SNF1 of yeast. These kinases are activated by low-energy supply through AMP/ATP ratio and inactivated by reduced carbon – glucose (yeast), glycogen (mammals), or trehalose-6-phosphate (plants). Under conditions of low energy and low carbon availability, these protein kinases regulate processes at the level of gene expression and through regulation of enzyme activity by phosphorylation. They prevent starvation by inhibiting processes that consume ATP, such as biosynthetic processes (anabolism), and through activating processes that mobilize carbon and energy such as carbohydrate, protein, and lipid breakdown (catabolism). In plants, photosynthesis is also upregulated and growth is inhibited by active SnRK1 (Figure 1). This is an important means through which plant growth and development are regulated by carbon and energy supplied by photosynthesis, and means of regulating photosynthesis itself in the context of whole plant carbon and energy balance. These protein kinases are indispensable for the growth and survival of plants and their importance is increasingly being appreciated in the healthy functioning of all organisms.

Photosynthetic CO₂ Fixation of Whole Plants

It is clear that photosynthesis is regulated locally not only by factors such as light, but also by more distant effects at the whole plant level. This is important, as we have seen for feedback regulation, where slow use of carbon in growing regions can lead to feedback regulation by accumulation of sugars in leaves. In terms of the relationship between photosynthesis and productivity, regulation of photosynthesis must be considered at the whole plant level. Stomatal density of leaves can also be controlled by long-distance signals. Mature leaves detect CO₂ concentration, which controls the stomatal development of young leaves. Leaf area is another important consideration. A central strategy of photosynthetic regulation of whole plants is to balance photosynthetic rate of leaves with the overall leaf area of the plant. Surveys of variations in photosynthesis per unit leaf area of many crops demonstrate a negative relationship with leaf area. Many environmental factors, particularly light, and the genetic makeup of the plant determine the trade-off between photosynthetic rate of individual leaves and leaf area. For example, plants grown in low light invest in larger leaf area to maximize light interception rather than by high investment of photosynthetic machinery per unit leaf area. Growth rate and productivity are more closely related to leaf area than the photosynthetic rate of individual leaves. This is reflected in crop breeding for yield, which has increased leaf area often at the expense of photosynthetic rate. Large leaf area combined with use of fertilizers and control of pests and diseases has maximized photosynthesis per unit of land for high crop yields. Improvements in yields in this way may be reaching a limit, so advances in improvement of photosynthetic rate combined with large leaf area may be the way forward to increasing photosynthesis and for future productivity gains. The complexity of photosynthetic regulation at the pathway and whole plant level show that there are many things to consider in trying to genetically engineer photosynthesis and productivity. However, improved understanding of the regulation of photosynthesis provides room for optimism.

See also: **Bioenergetics:** ATP in Plant Mitochondria: Substrates, Inhibitors, and Uncouplers; Chlorophylls and Carotenoids; Chloroplasts; Ferredoxin; **Metabolism Vitamins and Hormones:** Pentose Phosphate (Hexose MonoPhosphate) Pathway; Photosynthesis.

Further Reading

- Calvin M (1961) *The Path of Carbon in Photosynthesis, Nobel Lectures, Chemistry 1942–1962*. Amsterdam: Elsevier.
- Lake JA, Quick WP, Beerling DJ, and Woodward FI (2001) Signals from mature to new leaves. *Nature* 411: 154.
- Leegood RC, Sharkey TD, and von Caemmerer S (eds.) (2000) *Advances in Photosynthesis: Physiology and Metabolism*, vol. 9. Dordrecht: Kluwer.
- Lieman-Hurwitz J, Rachmilewicz S, Mittler R, Marcus R, and Kaplan A (2003) Enhanced photosynthesis and growth of transgenic plants that express *icbB*, a gene involved in HCO₃⁻ accumulation in cyanobacteria. *Plant Biotechnology* 1: 43–50.
- Miyawaga Y, Tamoi M, and Shigeoka S (2001) Over expression of a cyanobacterial fructose-1,6-bisphosphatase/sedoheptulose-1,7 bisphosphatase in tobacco enhances photosynthesis and growth. *Nature Biotechnology* 19: 965–969.

- Osborne CP and Beerling DJ (2006) Nature's green revolution: The remarkable evolutionary rise of C4 plants. *Philosophical Transactions of the Royal Society B* 361: 173–194.
- Tcherkez GG, Farquhar GD, and Andrews TJ (2006) Despite slow catalysis and confused substrate specificity, all ribulose biphosphate carboxylases may be nearly perfectly optimised. *Proceedings of the National Academy of Sciences of the United States of America* 103: 7246–7251.
- Von Caemmerer S and Farquhar GD (1981) Some relationships between the biochemistry of photosynthesis and the gas exchange of leaves. *Planta* 153: 376–387.
- Zhang Y, Primavesi LF, Jhurrea D, et al. (2009) Inhibition of SNF1-related protein kinase1 activity and regulation of metabolic pathways by trehalose 6-phosphate. *Plant Physiology* 149: 1860–1871.
- Zhu XG, de Sturler E, and Long SP (2007) Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: A numerical simulation using an evolutionary algorithm. *Plant Physiology* 145: 513–526.

Photosystem I

I Grotjohann and P Fromme, Arizona State University, Tempe, AZ, USA

Published by Elsevier Inc.

Glossary

Chloroplast Organelle of the plant cell, where photosynthesis takes place.

Electron donor/acceptor Molecules that donate or accept electrons in a redox reaction.

Electron transfer chain Sequential arrangement of pigments within the PS I, which transport the electron across the membrane.

Excitation energy transfer Transfer of the energy absorbed by the antenna pigments to the site of charge separation.

Photosystem I and II Large membrane protein complexes that perform the first reaction of energy conversion, the light-induced charge separation.

Thylakoid Membrane system, located inside the chloroplasts. The site where the proteins of the electron transfer chain are located and the primary processes of photosynthesis take place.

Definition

Photosystem I is a large membrane protein complex involved in the process of oxygenic photosynthesis, catalyzing the light-induced transmembrane electron transfer from plastocyanin or cytochrome *c*₆ to ferredoxin or flavodoxin.

Introduction

Oxygenic photosynthesis, catalyzed by plants, algae, and cyanobacteria, is the major process on Earth that converts light energy into chemical energy. It provides the major energy source for all higher life on the Earth and changed the atmosphere of planet Earth 2.5 billion years ago by using water as an electron source, leading to the evolution of oxygen. The light reactions of photosynthesis are catalyzed by two large membrane protein complexes, photosystem I and II (PSI and PSII), which act in series and are coupled by the cytochrome *b*₆*f* complex. PSI catalyzes the light-induced transmembrane charge separation across the photosynthetic membrane from plastocyanin or cytochrome *c*₆, located at the inside of the photosynthetic membrane (lumen of the thylakoids), to the soluble electron carrier protein ferredoxin, located at the outside (stroma or cytosolic side) of the thylakoids. The electrochemical gradient formed across the photosynthetic membrane by the light reactions in PSI and PSII drives synthesis of adenosine triphosphate (ATP) by the ATP synthase. The electrons are finally used for reduction of hydrogen, stored in the form of nicotinamide adenine dinucleotide phosphate and ATP, which are used for CO₂ fixation in the Calvin cycle.

Structure and Function of PSI: Overview

PSI is a large protein complex. In cyanobacteria, it is a trimer with a molecular weight of 1 million daltons. One monomer of PSI consists of 12 proteins and 127 cofactors. The structure of cyanobacterial PSI was determined at 2.5 Å resolution. It is shown in [Figure 1\(a\)](#).

The Cofactors

The cofactors are essential for the function of light absorption, energy transfer, and electron transfer. One functional unit of PSI contains 96 chlorophylls, the pigments that give plants their green color. The chlorophylls play a vital role in light absorption, energy transfer, and catalyze the primary charge separation and electron transfer reactions. The second prominent cofactors are 22 carotenoids, orange molecules, which serve as additional antenna pigments and protect PSI from photodamage under high light conditions. In addition to the chlorophylls and carotenoids, PSI contains three [4Fe4S] clusters and two phylloquinones, which are involved in electron transfer as well as four lipids and one Ca²⁺ ion, which play major structural roles.

The Proteins

The most important proteins in PSI are the two large protein subunits PsaA and PsaB. They form the functional and structural core of PSI and harbor all membrane intrinsic cofactors of the electron transport chain as well as 79 of the 90 antenna chlorophylls. The core of PSI consists of 10 transmembrane helices of PsaA and PsaB that surround the electron transfer chain like a fence. This core is represented by the five C-terminal transmembrane helices of PsaA/B. This part of PSI shows some structural but no sequence homologies to the core subunits of PSII and the purple bacterial reaction center (RC), which gives a hint that the photoreaction centers may all have evolved from a common ancestor. This heart of PSI is flanked on both sides by six transmembrane helices that correspond to the N-terminal antenna domain of PsaA/B (see [Figure 1\(c\)](#)). This antenna domain of PSI shows homologies to the major antenna proteins of PSII, fueling the suggestion that the large subunits of PSI, PsaA and PsaB, may have evolved by gene fusion of an ancient RC protein with an antenna protein. Seven small membrane intrinsic subunits surround the PsaA/B core of PSI. These small subunits, which contain only one to three transmembrane helices each, stabilize the antenna

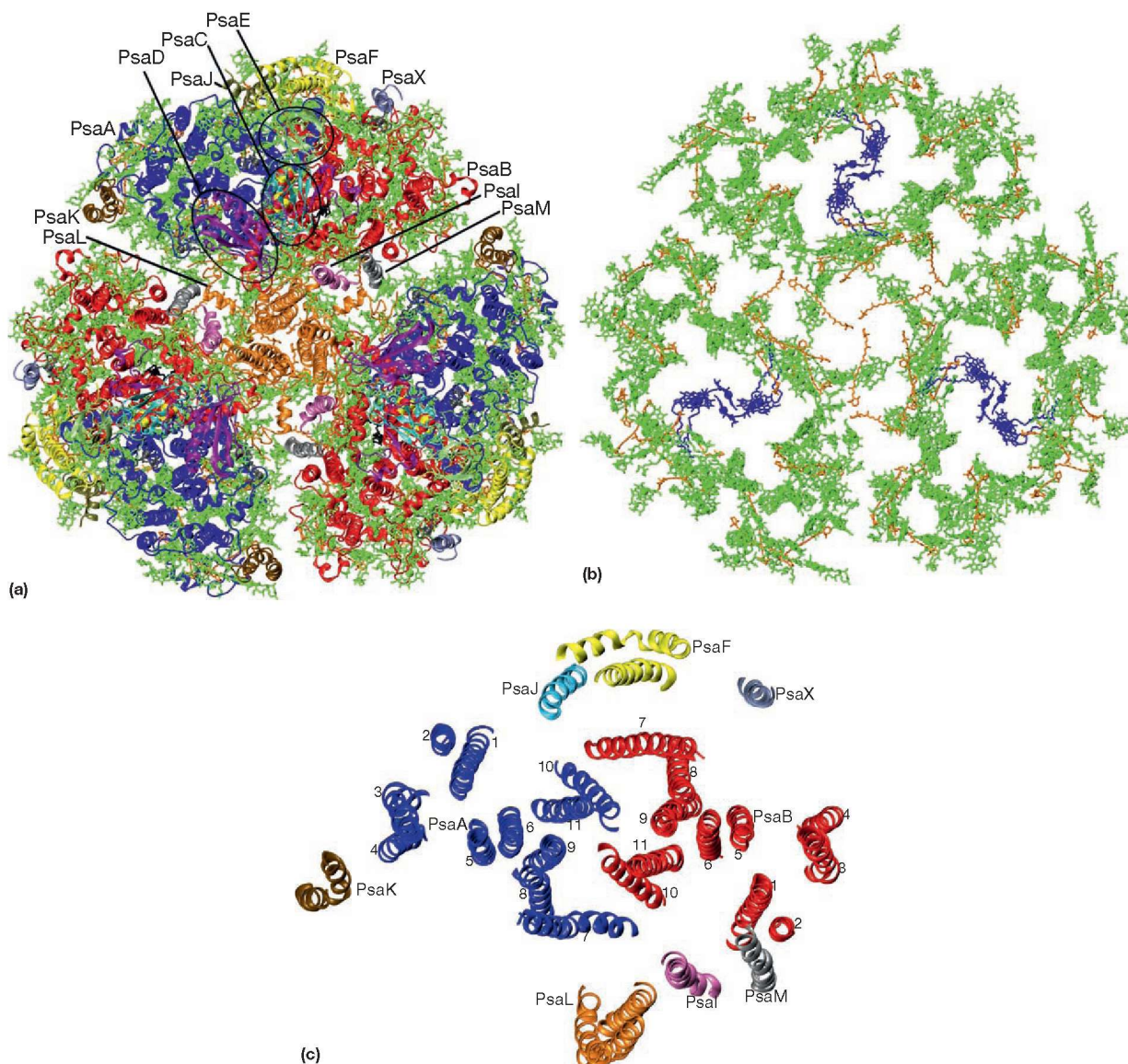


Figure 1 Photosystem I (PSI) from *Thermosynechococcus elongatus*. (a) View on a PSI trimer as seen from the cytosolic side. The protein subunits of one PSI monomer are marked. Nine of 12 subunits of a monomer are membrane intrinsic, with the largest subunits PsaA and PsaB among them. The three membrane-extrinsic subunits PsaC, PsaD, and PsaE form a stromal hump, and black circles mark their main bodies in the image. Note that PsaD forms a clamp across PsaC, which has not been marked. (b) The core-antenna network of a PSI trimer. Chlorophylls are shown in green, except those in the electron transfer chain, which are depicted in blue. Carotene molecules are orange. (c) The transmembrane helices of the nine membrane-intrinsic subunits of one PSI monomer as seen from the cytosol. The transmembrane helices of PsaA and PsaB have been numbered from N-terminus to C-terminus for orientation.

system and mediate trimerization of PSI. While the PsaA/B core of PSI is highly conserved between plants, algae, and cyanobacteria, the small membrane extrinsic subunits show variations between different photosynthetic kingdoms. In cyanobacteria, trimerization is mediated by subunits PsaL, I and M, while one of the additional plant specific subunit, PsaH, binds to PsaL and maintains PSI in a monomeric state in higher plants and green algae, where the PSI monomer is flanked on one site by four peripheral light-harvesting complexes I (LHCI), Lhca1-4. The structure of PSI from pea has been solved as a supercomplex with four LHCI proteins bound

to a monomeric PSI. The structure of this supercomplex is shown in Figure 2.

The soluble electron carrier proteins plastocyanin/cytochrome c_6 dock to PSI at its luminal site. Cytochrome c_6 is very likely the evolutionary older electron donor to PSI and can be found in most cyanobacteria and algae, some of which also contain plastocyanin, while in plants plastocyanin serves as the sole electron donor to PSI. The docking site for cytochrome c_6 /plastocyanin is located at a relatively shallow, 20-Å-deep luminal indentation of PSI, which guides the electron donor proteins to the docking site in close proximity to

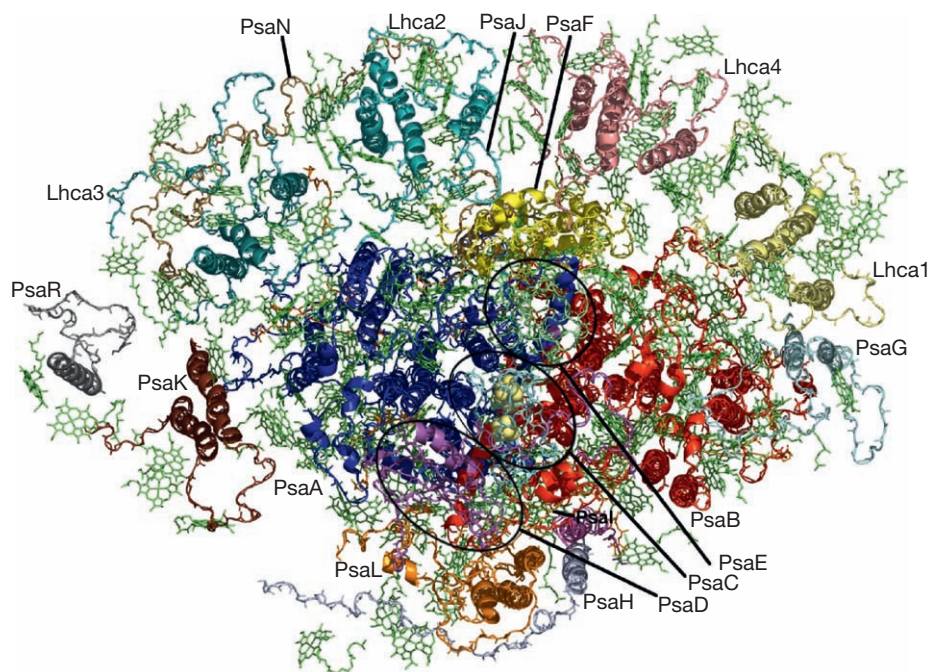


Figure 2 PSI from *Pisum sativum* as seen from the stroma. The core subunits (PsaA, PsaB, PsaC, PsaD, PsaE, PsaF, PsaI, PsaJ, PsaK, and PsaL) closely resemble their cyanobacterial counterparts, shown in [Figure 1\(a\)](#). The plant-specific subunits, PsaG and PsaH, play roles in binding LHCI and LHCII, respectively. The subunit PsaR has been putatively assigned and awaits further characterization.

P700, which in turn catalyzes the primary charge separation. This docking site only involves PsaA and PsaB and is mainly hydrophobic in cyanobacteria. However, an extension of a luminal helix of the subunit PsaF provides an additional interaction site in plants and algae that allows for a more stable complex formation by electrostatic interactions.

While the luminal side of PSI is relatively flat, it extends the membrane by 90 Å at the stromal side, a structural feature which is often referred to as the stromal hump. Three membrane extrinsic subunits, PsaC, PsaD, and PsaE, form the stromal hump which provides the docking site for the stromal electron carrier ferredoxin or flavodoxin. The latter replaces ferredoxin under iron deficiency. PsaC, which is located in the center of the hump, is the most important of the three extrinsic subunits, as it carries the terminal FeS clusters, named F_A and F_B , which transfer the electron from PSI to ferredoxin/flavodoxin. PsaE and PsaD stabilize the stromal hump and mediate docking of ferredoxin/flavodoxin. The stromal hump with the putative docking site is shown in [Figure 3](#).

The Electron Transfer Chain

The electron transport chain of PSI is well conserved between plants and cyanobacteria and consists of six chlorophylls, two phylloquinone molecules, and three [4Fe4S] clusters. The cofactor arrangement is shown in [Figure 4](#). The organic cofactors are arranged in the form of two branches, which show a twofold axis running through the first FeS cluster F_X . They are named A- and B-branch, which refers to the preferential coordination of the cofactors by the two large subunits, PsaA and PsaB. The electron transfer starts with the primary charge

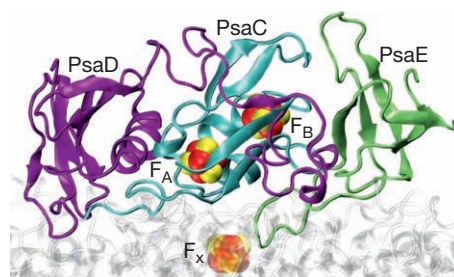


Figure 3 The cytosolic subunits PsaC, PsaD, and PsaE from *T. elongatus*. PsaC encloses two iron-sulfur clusters, F_A and F_B . PsaD forms a clamp across PsaC. The iron-sulfur cluster F_X , which is situated between the large membrane-intrinsic subunits PsaA and PsaB, is shown for orientation.

separation. Excitation energy transfer from the antenna chlorophylls leads to excitation of P700 to the excited state P700*, which catalyzes the primary charge separation. P700 is very likely represented by the first pair of chlorophylls, which consist of one chlorophyll *a* molecule in the B-branch and the C13 epimer of chlorophyll *a* in the A-branch. However, there is also spectroscopic evidence that P700* may be delocalized over more than just two chlorophylls and that charge separation may also be able to start from one of the accessory chlorophylls. The electron is transferred along either the A- or B-branch over a chain of electron carriers that include the second and third symmetry-related pairs of chlorophylls, the phylloquinones, and the FeS cluster F_X , until it reaches the terminal FeS clusters F_A and F_B , which are coordinated by PsaC. From F_B , the electron is finally transferred to ferredoxin. The nature and distribution of electrons along the two branches is an

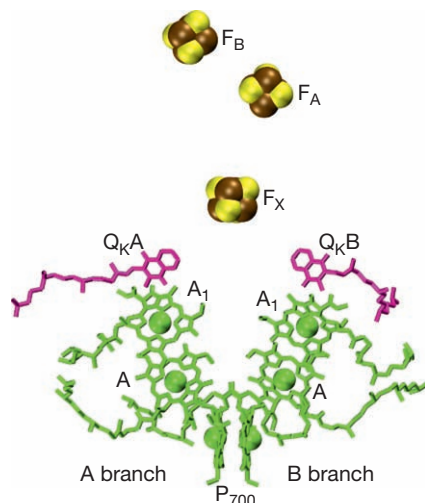


Figure 4 The electron transfer chain (ETC) of PSI. After charge separation at P_{700} , an electron moves along one of the branches of chlorophylls and phyloquinones and further along the iron–sulfur clusters, until the iron–sulfur cluster F_B reduces the soluble electron acceptor Ferredoxin.

open and vividly discussed topic of current research. Results of site-directed mutagenesis in combination with optical and electron paramagnetic resonance (EPR) spectroscopy evidence show that both electron transfer branches are active and that the ratio of A- versus B-branch electron transfer varies between organisms. There is also evidence that the electron transfer step from the phyloquinones to F_X is rate limiting for electron transfer in PSI, with the B-branch electron transfer step from the phyloquinone to F_X being 10 times faster than the electron transfer along the A-branch. The reason for this difference in electron transfer rates is under investigation. One explanation is the presence of two different lipids in close proximity to the electron transport chain. In the faster B-branch, this lipid is a neutral galactolipid, while at the A-branch a negatively charged phospholipid is located in proximity to the phyloquinone, which could increase the activation energy barrier for the electron transfer between the phyloquinone and F_X at the A branch. The first FeS cluster F_X is a rare example of a $4Fe_4S$ cluster coordinated by two different proteins. Two cysteine residues of PsaA and PsaB coordinate the four Fe atoms in the $[4Fe_4S]$ cluster F_X . F_X is located at the twofold axis that relates PsaA and PsaB. In addition to its role in electron transfer, F_X also plays a very important role in the assembly of PSI. From F_X , the electron is transferred subsequently to the FeS clusters F_A and F_B , located in the extrinsic subunit PsaC. PsaC shows strong structural homology to soluble ferredoxins that contain two $[4Fe_4S]$ clusters and it has therefore been proposed that it may share the same evolutionary roots with these soluble electron carriers, where a previously soluble protein became part of the larger multi-protein complex of PSI. Further evidence for this hypothesis is provided by the comparison of PSI with primitive anaerobic homodimeric photoreaction centers found in heliobacteria and green sulfur bacteria. The photoreaction centers in these organisms consist of a homodimeric RC core, where an extrinsic protein containing two $4Fe_4S$ clusters is only weakly attached.

The Antenna System

The core-antenna system of PSI in cyanobacteria consists of 90 chlorophyll *a* molecules and 22 carotenoids per monomer. The antenna systems of the three monomers in the PSI trimer (see [Figure 1\(b\)](#)) are tightly interconnected so that excitation energy transfer also involves transfer of excitation energy from one monomer to the next. The chlorophylls form a clustered network, where nearly all chlorophylls are located at distances <15 Å to the next neighbor, which ensures high efficiency of excitation energy transfer. The carotenoids are distributed all over the antenna system. They serve two functions. They serve as additional antenna pigments and are very important for photo-protection, as they quench dangerous triplet states of chlorophylls and dissipate the excess energy as heat, thereby preventing the formation of reactive oxygen species.

The quantum efficiency of the excitation energy transfer from the antenna to P_{700} is $>99.9\%$ at room temperature. This core-antenna system has been highly conserved over 1.5 billion years of evolution as more than 90% of the core-antenna chlorophylls have the same orientation and location in plant and cyanobacteria.

Most chlorophylls of the core-antenna system are coordinated by the large subunits PsaA and PsaB. It has been proposed that these subunits have evolved by a gene fusion between an antenna protein with the protein that coordinated the electron transfer chain, a situation that can still be found in PSII, where both functions are found on separate protein chains. The fusion of the two proteins allowed for a development of a very efficient antenna system in PSI, which contains a joint RC and antenna system. The whole core-antenna system is, structurally and functionally, tightly coupled to the RC core. The chlorophylls of the core antenna can be divided nearly half and half: (1) into antenna chlorophylls that are coordinated by the N-terminal six-transmembrane helices of PsaA/PsaB and the peripheral subunits and (2) the chlorophylls that are located close to the RC core which are coordinated by the C-terminal five-transmembrane helices of PsaA and PsaB. These include the two so-called ‘connecting’ antenna chlorophylls, which are coordinated by helix 11 and are located inside the fence of the 10 C-terminal transmembrane helices that surround the RC core of PSI.

The major function of the small subunits is to stabilize the core-antenna system and provide a connection to interact with peripheral antenna systems. The subunits that coordinate core-antenna chlorophylls are PsaI, PsaJ, PsaK, PsaL, and PsaX. It should also be noted that one of the phospholipids in PSI coordinates an antenna chlorophyll.

In plants, PSI is a monomer and forms a complex with four tightly bound peripheral antenna proteins: the LHCs Lhca1, Lhca2, Lhca3, and Lhca4. The structure of the PSI–LHCI complex of pea has been determined at 3.4 Å and reveals the architecture of the complex (see [Figure 2](#)). The Lhca proteins flank the PSI core at the site of the complex where subunits PsaF and PsaJ are located. The PSI–LHCI complex contains a total of 173 chlorophylls: 109 chlorophylls are assigned to the core of PSI, 55 chlorophylls are coordinated by the LHCI proteins, while nine chlorophylls are located at the interface between the LHCI antenna proteins and the core-antenna system of PSI. Plants undergo state transitions, during which the

light-harvesting complex II (LHCII) proteins can move from PSII to PSI. There is evidence from electron microscopic studies that the LHCII trimer may dock to PSI at the opposite site to the LHCI belt to the plant-specific subunit PsaH, which is located in close vicinity to PsaL.

Most cyanobacteria contain genes for two different peripheral antenna systems. Under nutrient replete conditions, phycobilisomes, which are large membrane extrinsic antenna complexes, serve as peripheral antenna systems for both PSI and PSII. They can move very fast between the photosystems in response to changing light conditions. In addition, many cyanobacteria also contain a membrane intrinsic antenna system consisting of rings of the IsiA protein that surround the PSI trimer. These IsiA rings dramatically increase the size of the antenna system in PSI under conditions of iron deficiency, which is the dominant environmental limitation in most aquatic environments. The system contains up to two rings of IsiA proteins surrounding the PSI core, with the inner ring consisting of 18 and the outer ring of 25 IsiA subunits. These 43 extra antenna proteins lead to a size of the PSI–IsiA double ring complex of 3.2 million daltons, which includes 855 chlorophyll molecules. The chlorophyll network in this giant antenna system is functionally even more tightly coupled to the core of PSI than the chlorophyll network in the plant PSI–LHCI complex.

Summary

PSI is a large membrane protein complex, present in plants, algae, and cyanobacteria. In cyanobacteria, it forms a trimer, where each monomer consists of 12 proteins and 127 cofactors. PSI catalyzes the light-induced electron transfer across the photosynthetic membrane from plastocyanin or cytochrome c_6 to ferredoxin or flavodoxin by a chain of electron carriers, consisting of six chlorophylls, two phylloquinones, and three 4Fe4S clusters. PSI contains a tightly coupled core-antenna system of 90 chlorophylls and 22 carotenoids, which is highly

conserved between plants and cyanobacteria and transfers the excitation energy to the core of PSI. In higher plants, a monomer of PSI forms a complex with four light harvesting proteins, while in cyanobacteria either membrane extrinsic phycobilisomes or (under iron deficiency) rings of the IsiA protein further increase the absorption cross section of PSI.

See also: [Bioenergetics: Dynamic Behavior of Photosystem II Light Harvesting](#); [Mitochondrial Calcium Transport: Historical Aspects](#).

References

- Amunts A, Toporik H, Borovikova A, and Nelson N (2010) Structure determination and improved model of plant photosystem I. *Journal of Biological Chemistry* 285: 3478–3486.
- Bibby TS, Nield J, and Barber J (2001) Iron deficiency induces the formation of an antenna ring around trimeric photosystem I in cyanobacteria. *Nature* 412: 743–745.
- Boekema EJ, Hifney A, Yakushevska AE, et al. (2001) A giant chlorophyll–protein complex induced by iron deficiency in cyanobacteria. *Nature* 412: 745–748.
- Chauhan D, Folea IM, Jolley CC, et al. (2011) A novel photosynthetic strategy for adaptation to low-iron aquatic environments. *Biochemistry* 50: 686–692.
- Dashdorj N, Xu W, Cohen RO, Golbeck JH, and Savikhin S (2005) Asymmetric electron transfer in cyanobacterial photosystem I: Charge separation and secondary electron transfer dynamics of mutations near the primary electron acceptor A_0 . *Biophysical Journal* 88: 1238–1249.
- Jensen PE, Bassi R, Boekema EJ, et al. (2007) Structure, function and regulation of plant photosystem I. *Biochimica et Biophysica Acta* 1767: 335–352.
- Jordan P, Fromme P, Witt HT, et al. (2001) Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* 411: 909–917.
- Li Y, van der Est A, Lucas MG, et al. (2006) Directing electron transfer within photosystem I by breaking H-bonds in the cofactor branches. *Proceedings of the National Academy of Sciences of the United States of America* 103: 2144–2149.
- Muller MG, Slavov C, Luthra R, Redding KE, and Holzwarth AR (2010) Independent initiation of primary electron transfer in the two branches of the photosystem I reaction center. *Proceedings of the National Academy of Sciences of the United States of America* 107: 4123–4128.
- Romberger SP and Golbeck JH (2010) The bound iron–sulfur clusters of type-I homodimeric reaction centers. *Photosynthesis Research* 104: 333–346.

Photosystem II: Assembly and Turnover of the D1 Protein

S Sirpiö, E Kanervo, and E-M Aro, University of Turku, Turku, Finland

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Eira Kanervo and Eva-Mari Aro, volume 3, pp. 363–366, © 2004, Elsevier Inc.

Glossary

D1 protein turnover Replacement of a nonfunctional D1 protein in the photosystem II (PSII) complex.

Photoinactivation Light-induced inactivation of the PSII electron transfer and water splitting.

PSII (photoinhibition) repair cycle Cycling of the PSII complexes between the grana- and stroma-exposed thylakoid membranes during the turnover of the D1 protein.

Biogenesis and Assembly of a New PSII Complex

Cooperation between Chloroplast and Nucleus

The PSII complex of eukaryotic organisms contains both nuclear and chloroplast-encoded gene products, as is the case with other photosynthetic protein complexes in the thylakoid membrane. This indicates that a tight cooperation exists between the organelle and nucleus in building up the PSII complex. From at least 25 genes encoding the PSII proteins, 15 genes reside in the chloroplast genome, rest of them locating in the nucleus.

Targeting the chloroplast-encoded PSII proteins

The majority of the PSII core proteins are chloroplast encoded. These comprise the D1, D2, CP43, CP47, cyt *b559 α* , and cyt *b559 β* proteins encoded by the *psbA*, *psbD*, *psbC*, *psbB*, *psbE*, and *psbF* genes, respectively. In addition, the PSII core proteins include the *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbTc*, and *psbZ* gene products of low molecular weight (LMW). Translation of the chloroplast-encoded PSII core proteins takes place on the surface of the thylakoid membrane. Initiation of translation of many chloroplast-encoded proteins is regulated by nuclear-encoded messenger RNA (mRNA)-binding proteins. To what extent these proteins are involved in targeting the ribosome–mRNA complexes to the thylakoid membrane is, however, unclear. Insertion of the chloroplast-encoded PSII proteins to the thylakoid membrane occurs both co- and post-translationally, and several soluble and membrane proteins are involved in assisting these processes. Cotranslational membrane insertion is typical for the PSII core proteins with several transmembrane helices, such as the D1, D2, CP43, and CP47 proteins.

Targeting the nuclear-encoded PSII proteins

The nuclear-encoded PSII core monomer proteins are three LMW subunits, namely the *psbW*, *psbX*, and *psbTn* gene products. In addition, proteins of the light-harvesting chlorophyll *a/b*-binding protein complex II (LHCII) and the oxygen-evolving complex (OEC), composed of the gene products of *psbO*, *psbP*, *psbQ*, and *psbR*, are nuclear encoded. All the nuclear-encoded proteins are synthesized as precursors on cytosolic ribosomes. These precursors are subsequently targeted to the chloroplast via an N-terminal transit peptide, which is proteolytically cleaved after import into the chloroplast. The nuclear-encoded

precursor proteins targeted to the thylakoid lumen contain a bipartite N-terminal transit peptide, the last part of which is removed after the arrival of the transported protein into the lumen. The luminal proteins, including those for the OEC, use both the Sec machinery and the aminoglycoside-modifying enzyme (Δ pH) pathway in their insertion and translocation to the luminal side of the thylakoid membrane.

Assembly of a New PSII Complex

General features

During the past decade, significant progress has been made in elucidating the PSII assembly and turnover processes. Besides purely biochemical approaches, the assembly of PSII has been studied using plant and cyanobacterial mutants. Particularly useful have been the mutants deficient in a specific PSII subunit or carrying a mutation in a nuclear gene that affects the expression or assembly of a structural PSII subunit. In addition, studies on plastid development from etioplast to chloroplast have contributed to the understanding of PSII assembly. Furthermore, research on the light-induced turnover of the PSII complex has produced information that can be applied also for the research on the assembly process of a new PSII center.

Sequence of the assembly of PSII core, OEC, and LHCII proteins

Light is required for maturation of etioplasts to chloroplasts and for synthesis and assembly of most of the PSII subunits. However, some of the PSII subunits do not need light for synthesis, since they exist already in etioplasts. This is the case particularly with cyt *b559* and the OEC proteins. In the assembly of the PSII core proteins, the two-subunit cyt *b559* functions as the first assembly partner for the D2 protein. All these together form an initial complex that functions as a receptor for the cotranslational assembly of the precursor-D1 protein. The assembly process then continues by association of the *psbI* gene product and the CP47 protein to the PSII subcomplex. Subsequently, the *psbH* gene product, the CP43 protein, and several LMW subunits attach to the growing PSII subcomplex. Assembly of the additional LMW subunits, namely the *psbJ* and *psbL* gene products, is essential for subsequent proper association of the OEC proteins on the luminal side of the PSII core monomer. Finally, the rest of the LMW subunits, such as the

psbK, *psbW*, and *psbZ* gene products, are required for the last events of the assembly process. These events include dimerization of the PSII core monomer and the attachment of LHCII to the PSII complex, resulting in the formation of a PSII supercomplex. An array of such PSII supercomplexes exists in the grana membranes of chloroplasts.

Regulation of the PSII Assembly

Light is required for an efficient translation elongation and accumulation of the D1 protein as well as for synthesis of the CP43 and CP47 core proteins. In addition to light, regulation of synthesis and assembly of the PSII complex involves the availability of a variety of factors, including the ligation of pigments and cofactors (chlorophyll *a*, pheophytin, β -carotene, Fe, Mn, and plastoquinone) and the availability of the assembly partners.

The chloroplast-encoded core proteins, D1, D2, CP43, and CP47, are cotranslationally inserted into the thylakoid membrane. Pigments are thought to bind to these proteins during their cotranslational membrane insertion and concomitant association with assembly partner(s). Light is needed for chlorophyll biosynthesis at the level of conversion of protochlorophyllide to chlorophyll *a* but we lack the information on details of pigment binding, even though the importance of this event as a regulative step in the assembly of PSII has been known for a long time. It is clear, however, that the chlorophyll-binding proteins of PSII are stabilized by ligation of pigments. Most chlorophyll-binding proteins do not accumulate at all in a nonpigmented form.

Successful synthesis of the PSII core proteins requires the availability of assembly partners. Synthesis of the D2 protein is dependent on the presence of *cyt b559*. Furthermore, *cyt b559*/D2 subcomplex is indispensable for the synthesis of D1, and the *cyt b559*/D2/D1 subcomplex, in turn, functions as an assembly partner for CP47. From the PSII core proteins, CP43 seems to be synthesized quite independently.

Moreover, dozens of nuclear-encoded auxiliary protein are required for correct and efficient assembly of PSII. During the D1 protein translation and insertion into the membrane, the stromal proteins LPA1 and HCF136 as well as the luminal immunophilin CYP38 guide the proper folding of the D1 protein. On the other hand, LPA2 is responsible for the assembly of CP43 into PSII. The thylakoid lumen proteins are of utmost importance during the assembly of the PSII core complex and OEC proteins. These proteins include several luminal peptidyl-prolyl isomerases, OEC protein isoforms, and PSB27. Integration of the LHCII protein to the membrane is dependent on the presence of the thylakoid membrane protein ALB3.

Turnover of the D1 Protein of PSII

PSII Photoinactivation and the D1 Protein Damage

The PSII complex performs a remarkable task in splitting water molecules to oxygen and protons. Such oxidizing electron transfer reactions of PSII readily lead to the formation of highly reactive radicals that are potentially harmful to the protein moiety. Thus, PSII, when performing its normal function, is hazardous for itself.

The rate of PSII photoinactivation increases almost linearly with increasing light intensity. Under low light, the rate of photoinactivation is also low. The rate of inactivation correlates with the amount of photons absorbed and mediated to the reaction center. Upon inactivation, the PSII complex becomes unable to transfer electrons and split water molecules. A constant repair of inactivated complexes is required to guarantee the maintenance of a sufficient level of active PSII complexes for photosynthesis. The efficiency of repair is affected negatively by environmental stress factors, such as high light and the extremes of temperature. If the repair process cannot keep up with the rate of photodamage, nonfunctional PSII complexes start to accumulate. This situation leads to a measurable decrease in the rate of total photosynthesis.

The primary target of the photo-induced PSII damage is the reaction center protein D1. This fact reflects the grand design of PSII: only one protein of the multisubunit complex is primarily destroyed, whereas the rest of the proteins are rescued and used for reparation of the PSII complex. The primary photodamage is thought to include a conformational change and an oxidative damage to the D1 protein. These events, in turn, trigger the reaction center protein D1 to a substrate for proteolytic degradation. However, the actual nature of the primary photodamage is not yet properly understood.

PSII Photoinhibition Repair Cycle

Start of the repair process

Active PSII centers exist as dimers in the thylakoid membranes of grana stacks (Figure 1). After a light-induced damage to PSII, the LHCII antenna dissociates and monomerization of PSII occurs. PSII monomers then migrate from the grana to the stroma-exposed thylakoid membranes where a contact with the components acting in degradation and synthesis of the D1 protein is feasible. OEC dissociates from PSII and a partial disassembly of the PSII core proteins takes place. The stages from photodamage to degradation of the D1 protein are regulated by phosphorylation–dephosphorylation events of the PSII core proteins.

Degradation of the D1 protein

Degradation of the light-damaged D1 protein is a proteolytic process that occurs in a well-coordinated manner on the stroma-exposed thylakoid membranes, leaving the closest partner, the D2 protein, most often intact. The D1 protein is an integral membrane protein with five transmembrane helices. Between helices, there are loops both on the stromal and luminal sides of the thylakoid membrane. The loops are the preferential sites of the primary cleavage of an integral membrane protein, in general. In the case of the D1 protein, the stromal DE-loop shows the most important contribution to proteolysis, while the role of luminal AB- and CD-loops seems to be less important. The primary proteolysis of D1 produces fragments that are further degraded by the secondary proteolysis. Apparently, several enzymes of the FtsH and Deg protease families in concert degrade the D1 protein, however, the FtsH proteases being in dominant role. FtsH proteases are metalloproteases and form a heterohexameric complex on the stromal side of the thylakoid membrane. Deg proteases, most of which

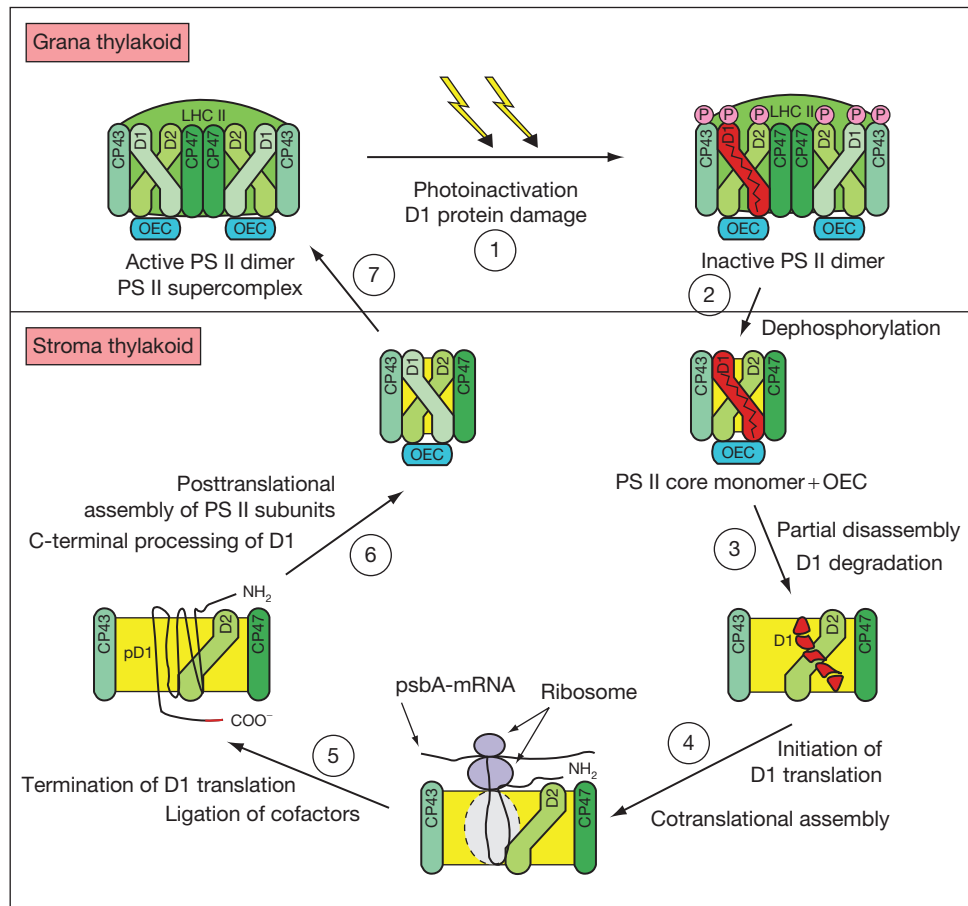


Figure 1 Model for the PSII photoinhibition repair cycle. (1) Photoinactivation of PSII and damage to the D1 protein. (2) Monomerization of the inactive PSII complex and its translocation from the grana thylakoids to the stroma-exposed thylakoids, followed by PSII core protein dephosphorylation. (3) Partial disassembly of the PSII core proteins and proteolytic degradation of the D1 protein. (4) Initiation of D1 translation and cotranslational assembly of the nascent D1 chain into the D1-depleted PSII core monomer. (5) Ligation of cofactors during translation elongation and termination of D1 translation. (6) C-terminal processing of the pD1 protein; posttranslational assembly of the PSII core subunits and OEC. (7) Translocation of the repaired PSII monomer from the stroma-exposed thylakoids to the grana thylakoids; dimerization of PSII and the attachment of the LHCII antenna. D1, D2, CP43, and CP47, chlorophyll *a/b*-binding proteins of the PSII core; LHCII, light-harvesting chlorophyll *a/b*-binding protein complex II; OEC, oxygen-evolving complex; pD1, precursor-D1 protein; P, phosphorylated protein.

are located in the thylakoid lumen, are endopeptidases and form a homotrimeric complex.

Synthesis of the D1 protein

Synthesis of the D1 protein during the PSII repair cycle shares features with synthesis of the D1 protein in the assembly of a new PSII complex (see above). The D1 protein is encoded by the *psbA* gene residing in the chloroplast genome. The *psbA* mRNA-ribosome complex is targeted to the stroma-exposed thylakoid membrane, where elongation of the nascent D1 chain takes place. The nascent D1 chain is cotranslationally inserted into the thylakoid membrane, where the D2 protein and cyt *b559* act as the first association partners. Elongation and concomitant membrane insertion of the D1 protein are controlled by a variety of factors. Such factors include association of the nascent chain with a translocon and availability of the assembly partners and pigment molecules. Further regulation on D1 synthesis is provided by transthylakoid proton gradient, reactive oxygen

species, and other redox-active chloroplast components. When all the five transmembrane helices have been synthesized, the precursor D1 protein releases from the ribosome. The fifth transmembrane helix traverses the membrane, finally having its carboxyl terminus in the thylakoid lumen.

Posttranslational events of the D1 synthesis and reactivation of the PSII complex

Posttranslational events of the D1 protein synthesis include several steps. First, during maturation of the D1 protein, the luminal processing protease CtpA removes the extension of 9–16 amino acids from the carboxyl terminus of the precursor D1 protein. This cleavage is necessary for ligation of the manganese cluster and reassociation of the OEC to the PSII complex. Finally, the properly assembled PSII monomer migrates back to the grana membranes, where PSII core dimerization and full reactivation of the PSII complex take place. The final reactivation stage also includes an association of the LHCII

antenna proteins to the PSII dimer. These final stages in the grana thus accomplish the PSII photoinhibition repair cycle providing an active PSII dimer for photosynthesis.

See also: **Bioenergetics:** Chloroplasts; Dynamic Behavior of Photosystem II Light Harvesting; Photosystem II: Water Oxidation, Overview; Photosystem II: Redox and Protein Components.

Further Reading

- Andersson B and Aro E-M (2001) Photodamage and D1 protein turnover in photosystem II. In: Aro E-M and Andersson B (eds.) *Regulation of Photosynthesis*, vol. 11, pp. 377–393. Dordrecht: Kluwer.
- Baena-González E and Aro E-M (2002) Biogenesis, assembly and turnover of photosystem II units. *Philosophical Transactions of the Royal Society London B* 357: 1451–1460.
- Herrmann RG and Westhoff P (2001) Thylakoid biogenesis and dynamics: The result of a complex phylogenetic puzzle. In: Aro E-M and Andersson B (eds.) *Regulation of Photosynthesis*, vol. 11, pp. 1–28. Dordrecht: Kluwer.
- Kato Y and Sakamoto W (2009) Protein quality control in chloroplasts: A current model of D1 protein degradation in the photosystem II repair cycle. *Journal of Biochemistry* 146: 463–469.
- Marín-Navarro J, Manuell AL, Wu JP, and Mayfield S (2007) Chloroplast translation regulation. *Photosynthesis Research* 94: 359–374.
- Mulo P, Sirpiö S, Suorsa M, and Aro E-M (2008) Auxiliary proteins involved in the assembly and sustenance of photosystem II. *Photosynthesis Research* 98: 489–501.
- Paulsen H (2001) Pigment assembly – transport and ligation. In: Aro E-M and Andersson B (eds.) *Regulation of Photosynthesis*, vol. 11, pp. 219–233. Dordrecht: Kluwer.
- Tikkanen M, Nurmi M, Kangasjärvi S, and Aro E-M (2008) Core protein phosphorylation facilitates the repair of photodamaged photosystem II at high light. *Biochimica et Biophysica Acta* 1777: 1432–1437.
- Van Wijk KJ (2001) Proteins involved in biogenesis of the thylakoid membrane. In: Aro E-M and Andersson B (eds.) *Regulation of Photosynthesis*, vol. 11, pp. 153–175. Dordrecht: Kluwer.
- Zerges W (2002) Does complexity constrain organelle evolution? *Trends in Plant Science* 7: 175–182.

Photosystem II: Redox and Protein Components

J Barber, Imperial College, London, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Light-harvesting system Proteins that bind chlorophyll, carotenoids, and phycobilins which absorb light and transfer excitation energy to the reaction center.

Oxidation The process of removing an electron to a chemical species.

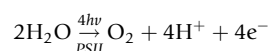
Photosystem Pigment–protein complex containing a light-harvesting system and a reaction center.

Radical Reactive chemical species with an unpaired electron.

Reaction center Protein subunits of the photosystem that bind redox active cofactors which use excitation energy to bring about charge transfer across the membrane.

Reduction The process of donating an electron to a chemical species.

Photosystem II (PSII) is a multisubunit protein complex embedded in the thylakoid membranes of oxygenic photosynthetic organisms: plants, algae, diatoms, and cyanobacteria. By using chlorophyll and carotenoid molecules and several redox active cofactors, it captures and uses solar energy to split water into molecular oxygen and reducing equivalents:



To produce a dioxygen molecule, four photons ($h\nu$) are needed to provide the energy to oxidize two water molecules. The dioxygen released maintains our oxygenic atmosphere and provides an ozone layer which protects us from the damaging effects of ultraviolet (UV) radiation. The reducing equivalents, which exit PSII as plastoquinol, are ultimately used, with the aid of additional light energy absorbed by photosystem I (PSI), to convert atmospheric carbon dioxide to the organic molecules that make up virtually all the biomass on our planet.

Morphology

Thylakoid Membrane

In chloroplasts of plants and green algae, the thylakoids are highly folded membranes having stacked (grana) and unstacked (stromal) lamellae regions. PSII is located mainly in the stacked regions. In cyanobacteria, diatoms, and other types of algae, the thylakoids do not have a clear differentiation into stacked and unstacked regions and PSII is more evenly distributed along the plane of the membrane. In all cases the thylakoid membrane encloses an inner aqueous compartment known as the lumen while the outer compartment is either the stroma of the chloroplast or cytoplasm of cyanobacteria referred to as the stromal side below.

Reaction Center

Reaction center (RC) is that part of a photosystem, like PSII, where light energy is converted to electrochemical potential energy. It is composed of protein-bound redox active cofactors that facilitate charge separation across the membrane. PSII exists *in vivo* either as a monomeric or a dimeric RC core

complex where the structure of the latter has been determined by X-ray crystallography (Figure 1).

Light-Harvesting Complexes

These pigment–protein complexes capture light energy and transfer it to the RC. They are often composed of several hundred pigment molecules, the exact number being dependent on growth conditions and type of organism. Various pigments are used so as to fully utilize the solar spectrum and for PSII these are different types of chlorophylls, phycobilins, and carotenoids. Some light-harvesting complexes (LHCs) are contained within the thylakoid membrane (intrinsic) and some attach to the surface (extrinsic) of the membrane.

Redox Active Cofactors

P680

P680 is composed of chlorophyll *a* molecule which, after excitation by the absorption of light to form P680*, gives up an electron to an acceptor, converting it to P680^{•+}. This radical cation has a redox potential estimated to be ~1.2 V or more, which is required to oxidize water. Its oxidation causes a bleach in its absorption spectrum centered at 680 nm. P680 is located on the luminal side of the PSII RC. Recent work suggests that the excited state P680* is delocalized over four symmetrically related chlorophyll molecules (depicted as P_{D1}, P_{D2}, Chl_{D1}, and Chl_{D2} in Figure 2) and that upon oxidation the hole becomes localized on a single chlorophyll (P_{D1}).

Pheophytin *a*

Pheophytin *a* (Pheo_{D1}) is the primary electron acceptor of PSII located toward the stromal side of the RC. It receives an electron from P680* in a few picoseconds to form a radical pair P680^{•+} Pheo^{•-}. The reduction of Pheo_{D1} can be detected as an optical absorption decrease at 422 nm and its reduction potential is ~-600 mV. The symmetrically related Pheo_{D2} (Figure 2) is not normally involved in electron transfer.

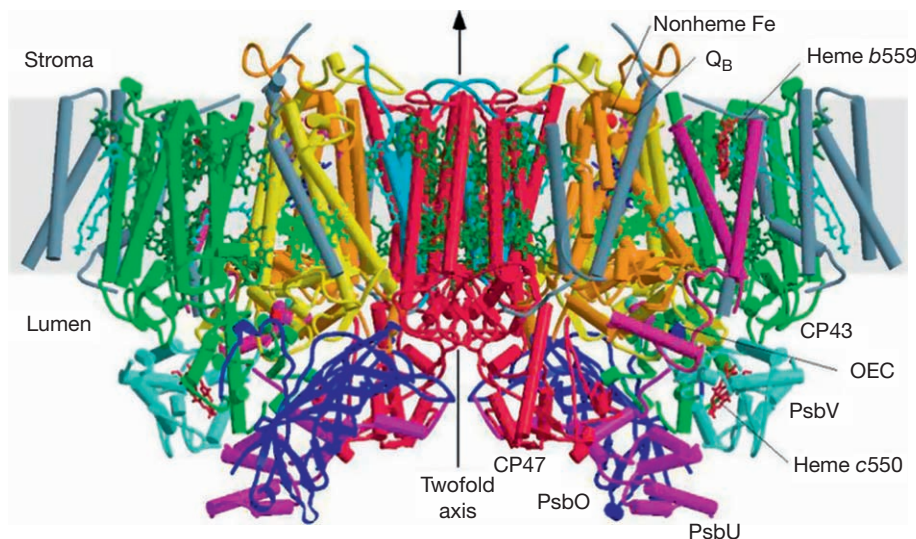


Figure 1 Side view of the PSII dimer perpendicular to the membrane normal derived from X-ray crystallography. Helices are represented as cylinders with D1 in yellow, D2 in orange, CP47 in red, CP43 in green, Cyt b559 in wine red, PsbL, PsbM, and PsbT in medium blue, and PsbH, PsbI, PsbJ, PsbK, PsbX, PsbZ, and the putative PsbN in gray. The extrinsic proteins are PsbO in blue, PsbU in magenta, and PsbV in cyan. Chlorophylls of the D1/D2 reaction center are light green, pheophytins are blue, chlorophylls of the antenna complexes are dark green, β -carotenes are in orange, hemes are in red, nonheme Fe is red, and Q_A and Q_B are magenta. The oxygen-evolving center (OEC) is shown as the red (oxygen atoms), magenta (Mn ions), and cyan (Ca^{2+}) balls. From Ferreira K, Iverson T, Maghlaoui K, Barber J, and Iwata S (2004) Architecture of the photosynthetic oxygen-evolving centre. *Science* 303: 1831–1838, with permission from AAAS.

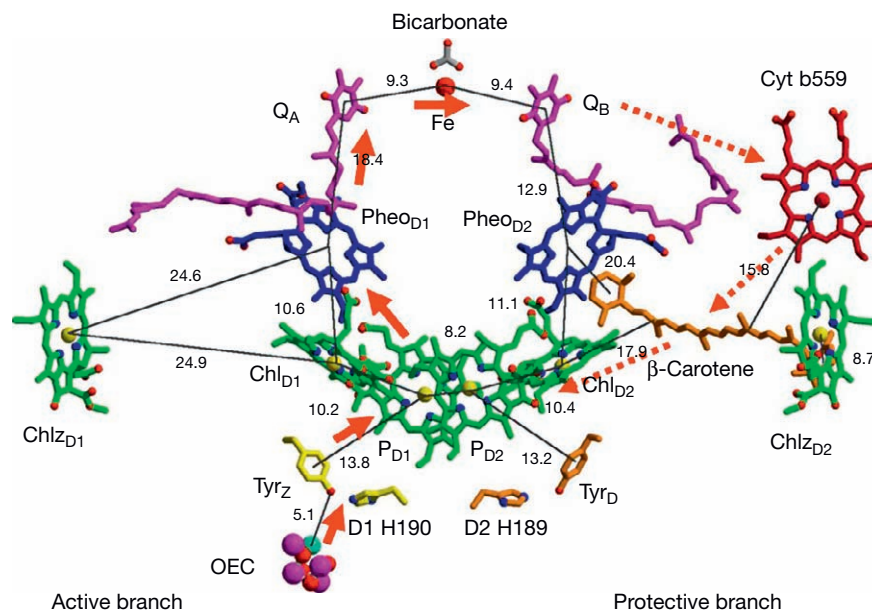


Figure 2 Cofactors involved in electron transfer visualized perpendicular to the internal pseudo-twofold axis. Coloring scheme is the same as Figure 1. The phytol tails of the chlorophylls and pheophytins have been removed for clarity. The side chains of Y_Z (D1-Tyr161) and D1-His190 are shown in yellow and Y_D (D2-Tyr160) and D2-His189 in orange. The four chlorophylls comprising P680 are in direct van der Waals contact, and other electron transfer distances are given in angstroms. The electron transfer pathway is indicated by arrows and distances are given in angstroms based on the 3.5-Å crystal structure. Excitation of the reaction center, via light absorption by chlorophylls (Chl) in the antenna, drives electron transfer from the cluster of four chlorophylls bound to the D1- and D2-proteins, to the pheophytin (Pheo) acceptor leading to the radical pair $P680^{*+} Pheo^{*-}$. The radical cation of P680 is localized on P_{D1} although the primary electron charge separation probably occurs from Chl_{D1} followed by hole migration, while the radical anion is located on $Pheo_{D1}$. The electron on $Pheo_{D1}^{*-}$ is rapidly donated to a firmly bound plastoquinone Q_A (shown in purple) and then transferred to a second plastoquinone Q_B (also shown in purple). This electron transfer is aided by the presence of a nonheme iron located midway between them. When the Q_B plastoquinone is fully reduced and protonated to plastoquinol (PQH_2), it diffuses from the Q_B -binding site into the lipid matrix of the membrane. On the electron donor side, $P680^{*+}$ is reduced by a redox active tyrosine (Tyr_Z) which then extracts electrons from the $Mn_4Ca^{2+}O_4$ -cluster that constitutes the oxygen-evolving center (OEC). These electron transfer processes occur mainly on the D1 side of the reaction center and the symmetrically related cofactors located on the D2 side (Chl_{D2} and $Pheo_{D2}$) are nonfunctional. Other cofactors shown, including the heme of cytochrome b559 (cyt b559 shown in red) and the β -carotene molecule (shown in brown), play a role in protecting PSII against photoinduced damage. From Ferreira K, Iverson T, Maghlaoui K, Barber J, and Iwata S (2004) Architecture of the photosynthetic oxygen-evolving centre. *Science* 303: 1831–1838, with permission from AAAS.

Q_A

Q_A is a bound plastoquinone (PQ) molecule which undergoes a single reduction by PheoD1 $^{*+}$ in ~ 200 ps, forming the charge transport state P680 $^{*+}$ Pheo Q_A^+ where the redox potential of the semi-plastoquinone (Q_A^-) is ~ -150 mV.

Q_B

Q_B is also a PQ which can accept two electrons successively from Q_A^- and together with two protons (H^+) from the stroma or cytoplasm to form plastoquinol (PQH $_2$). In this fully reduced form, PQH $_2$ is released from the Q_B -binding site of the PSII RC and diffuses into the lipid matrix of the thylakoid membrane until it is oxidized by the cytochrome b_6f complex. The reducing equivalents are further energized by a second light reaction which occurs in PSI so as to reach a redox potential sufficient to reduce carbon dioxide.

Tyrosines Y_Z and Y_D

Y_Z is a redox active tyrosine 161 of the D1 protein which reduces P680 $^{*+}$. In so doing, it loses a H^+ from its phenolic group to the nearby H-bonded histidine 190 of the D1 protein (see Figure 2) and the resulting species is a neutral tyrosine radical $Y_Z^\bullet$. Symmetrically related to Y_Z is Y_D tyrosine 160 of the D2 protein (see Figure 2). Although this tyrosine is also H-bonded to D2His189 and can be oxidized by P680, it is not directly involved in water oxidation.

Chl $_{D1}$ and Chl $_{D2}$

Chl $_{D1}$ and Chl $_{D2}$ are two peripheral chlorophyll a molecules which under some circumstances are oxidized by P680 $^{*+}$ and are located close to the A/B transmembrane helices of the D1 and D2 proteins, respectively. The relatively long distance from the P680 chlorophylls suggests that the electron transfer, which is in the millisecond timescale, is facilitated by β -carotene molecules, one of which is shown in Figure 2. The function of this redox activity seems to be to protect PSII against oxidative damage.

Hemes

There are two hemes in the PSII of cyanobacteria, the high potential cytochrome b559 and the low potential cytochrome c550. The former can be oxidized by P680 with the electron transfer being facilitated by a β -carotene molecule positioned midway between it and P680 (see Figure 2). This millisecond electron transfer, like that associated with Chl $_{D1}$ and Chl $_{D2}$, seems to play a protective function for PSII. The heme of cytochrome b559 is ligated by two histidines equally provided by the PsbE and PsbF proteins. In contrast, cytochrome c550 is not normally redox active, is located in the extrinsic PsbV protein, and is found only in PSII of cyanobacteria and red algae. Cytochrome b559 is conserved in PSII of all types of oxygenic photosynthetic organisms.

Mn $_4$ Ca $^{2+}$ -cluster

A cluster of four manganese atoms together with a Ca $^{2+}$ atom makes the catalytic center (Mn $_4$ Ca $^{2+}$ -cluster), where the oxidation of two substrate water molecules occurs. The removal of

4e and 4 H^+ requires four successive oxidations of P680, and there is evidence that $Y_Z^\bullet$ may act as an H^+ as well as an electron acceptor and therefore is directly involved in the water oxidation reaction.

The highest resolution X-ray crystallography (1.9 Å) has revealed that Mn $_4$ Ca-cluster is organized as a 'distorted chair', with an asymmetric cubane-like structure composed of three Mn atoms, one Ca $^{2+}$ ion, and four oxygen atoms in the corners of this structure. The fourth Mn (Mn4, or the dangler Mn) is located outside of the cubane and is linked to Mn1 and Mn3 of the cubane via two di- μ -oxo bridges involving the fifth oxo and a bridging oxo of the cubane. This general geometry of cubane has been suggested previously based on lower 3.5-Å resolution diffraction. Four water molecules are associated with the cluster, two ligated to Mn4 and the other two to the calcium ion. The remaining ligands are provided by the D1 and CP43 proteins.

Bicarbonate

A bicarbonate provides bidentate ligation to the nonheme iron with further histidine ligands provided equally by the D1 and D2 proteins. Although tentatively modeled into an earlier crystal structure of PSII as a ligand to the Mn $_4$ Ca $^{2+}$ -cluster, there is no further evidence to support this, even though there is some support for this anion playing a role in the water oxidation process.

Chloride

Two chloride ions are located in the vicinity of the Mn $_4$ Ca-cluster but not in ligating distance. One site is close to D2Lys317 and backbone nitrogens of D1Glu333 and the other close to CP43Glu354 and the backbone nitrogens of D1Asn338. They seem to play a role in maintaining the coordination environment of the metal cluster while at the same time facilitate the transport of water and protons to and from the catalytic site.

Proteins

The RC core complex of PSII is composed of a large number of intrinsic and extrinsic proteins encoded by *psb* genes (see Table 1). The RC core is serviced by LHC proteins, the number and type of which vary between different types of organisms. Plants and green algae have their LHC proteins encoded by *cab* genes while red algae and cyanobacteria contain phycobiliproteins encoded by *apc* and *cpc* genes. Here we confine ourselves to PSII RC core proteins encoded by the *psb* genes (Figure 3) but note that there are differences between plants and cyanobacteria.

PsbA-D1 Protein

After C-terminal posttranslational modification, this protein has a molecular mass of ~ 38 kDa and contains ~ 350 amino acids depending on species. It has five transmembrane α -helices (A–E), DE stromal surface α -helix, and CD lumenal surface α -helix. It binds the majority of the cofactors involved in PSII-mediated electron transport: Y_Z is Tyr161, P680 $^{*+}$ is chlorophyll a ligated by His198, called P $_{D1}$, Pheo $_{D1}$ is H-bonded by Tyr126 and Glu130, Q_B via interactions with

Table 1 Psb protein subunits of PSII RC core complex

Gene	Subunit	Mass (kDa)	No. of transmembrane α -helices
<i>psbA</i> (c)	D1	38.021 (S)	5
<i>psbB</i> (c)	CP47	56.278 (S)	6
<i>psbC</i> (c)	CP43	50.066 (S)	6
<i>psbD</i> (c)	D2	39.418 (S)	5
<i>psbE</i> (c)	α -cyt b559	9.255 (S)	1
<i>psbF</i> (c)	β -cyt b559	4.409 (S)	1
<i>psbH</i> (c)	H protein	7.697 (S)	1
<i>psbI</i> (c)	I protein	4.195 (S)	1
<i>psbJ</i> (c)	J protein	4.116 (P)	1
<i>psbK</i> (c)	K protein	4.283 (S)	1
<i>psbL</i> (c)	L protein	4.366 (S)	1
<i>psbM</i> (c)	M protein	3.755 (P)	1
<i>psbN</i> (c)	N protein	4.722 (T)	1
<i>psbO</i> (c)	33-kDa O protein	26.539 (S)	0
<i>psbP</i> (n)	23-kDa P protein	20.210 (S)	0
<i>psbQ</i> (n)	16-kDa Q protein	16.523 (S)	0
<i>psbR</i> (n)	R protein	10.236 (S)	0
<i>psbS</i> (n)	S protein	29.197 (S)	4
<i>psbT</i> (c)	T _C protein	3.849 (S)	1
<i>psbT</i> (n)	T _n protein	5.000 (A)	0
<i>psbU</i>	U protein	10.491 (Sy)	
<i>psbV</i>	V protein	15.121 (Sy)	0
<i>psbW</i> (n)	W protein	5.928 (S)	1
<i>psbX</i> (n)	X protein	4.225 (S)	1
<i>psbY</i> (c)	Y protein	4.202 (Sy)	1
<i>psbZ</i> (c)	Z protein	6.541 (S)	2
<i>ycf 12</i>	Ycf protein	4.143 (Sy)	1

These proteins are products of the *psbA* to *psbZ* plus *ycf 12* genes that occur in oxygenic organisms.

In eukaryotic organisms, the *psb* genes are located in either the chloroplast (c) or nuclear (n) genes.

The molecular masses of the mature PsbA–PsbZ/Ycf12 proteins are calculated from the protein sequences reported in the SwissProt database using the MacBioSpec program (Sciex Corp., Thornhill, ON, Canada) for spinach (S), pea (P), tobacco (T), *Arabidopsis* (A), and *Synechocystis* PCC6803 (Sy).

His 215 and Ser 264. The accessory Chl_{D1} is coordinated by a water ligand H-bonded to Thr179. Mn₄Ca²⁺-cluster is ligated by Asp170, Glu189, His332, Glu 333, Asp342, and Ala344. With the exception of Glu189 and His332, all other D1 protein ligands are bidentate carboxylate. The D1 protein also binds a nonheme iron, which is normally redox inactive and positioned between Q_A and Q_B, and is ligated by His215 and His272. The D1 protein also contributes to the binding of two chloride molecules near to the Mn₄Ca²⁺-cluster. Another important property of the D1 protein is that it turns over more rapidly than any other protein in the thylakoid membrane. This remarkable feature is linked to the fact that PSII is susceptible to photoinduced damage leading to photoinhibition and a lowering of photosynthetic efficiency. The degradation, synthesis, and reinsertion of this protein into the RC core complex represent a very important aspect of the dynamics of PSII.

PsbB-CP47

This highly conserved PSII RC protein consists of ~500 amino acids and has a molecular mass of ~56 kDa depending on

species. It is often known as CP47 and has six transmembrane α -helices (I–VI) arranged in a ring consisting of three pairs with N- and C-termini exposed at the stromal surface (Figure 3). The luminal loop joining transmembrane helices V and VI is large, consisting of 200 amino acids having two long and four short helices as well as three β -sheets. It binds 16 chlorophyll *a* molecules and five β -carotenes. Most of the Chls are ligated to His residues located in the transmembrane helices and form an LHC system for the RC. Where they are not, three water molecules act as ligands. The large luminal loop functions indirectly in the water oxidation reaction by interacting with the D2 protein which is adjacent to it. Depletion of the *psbB* gene and a wide range of site-directed mutational studies have emphasized the absolute requirement of the PsbB protein in PSII assembly and function.

PsbC-CP43

After posttranslational processing, the PsbC-CP43 protein, depending on species, contains ~470 amino acids and has a molecular mass of ~50 kDa. It is homologous with PsbB (CP47) in that it has six transmembrane α -helices arranged in the same way but located on the D1 side of the RC (see Figure 3). It also contains a considerable number of histidine residues which are involved in ligating some of the 13 Chl_a bound to this protein. Where they are not, CP43Asn39 and two water molecules act as ligands. Like CP47, CP43 binds three β -carotene molecules and has a large luminal loop joining helices V and VI, which is slightly smaller, having ~150 amino acids. It contains two long and three short helices one of which is a 3₁₀ helix consisting of the highly conserved GGETMRFW motif where E is Glu354 which provides a bidentate ligand to the Mn₄Ca-cluster and R is Arg357 which is within the second coordination sphere of the metal cluster. It also differs from CP47 in that in plants (but not in algae and cyanobacteria) its N-terminal threonine can be reversibly phosphorylated. Moreover, CP43 seems to be more weakly associated with PSII compared with that of CP47, a feature which may be important when the D1 protein is degraded and replaced. Despite these differences, CP43, like CP47, acts as an LHC for PSII and its presence is also necessary for water splitting activity since it provides amino acids which are intimate with the catalytic center.

PsbD-D2 Protein

The PsbD-D2 protein is homologous to the D1 protein. Although it is slightly larger than the D1 protein, having in the region of 353 amino acids and a molecular mass of ~39.5 kDa, depending on species it has five transmembrane helical segments organized in the same way as the D1 protein (see Figure 3). Compared with the D1 protein it is involved to a lesser extent in binding active cofactors. There are two chlorophylls P_{D2} and Chl_{D2} which together with P_{D1} and P_{D1} form the cluster constituting P680* and are ligated to His197 and a water molecule, respectively. The binding of Q_A involves at least His214, Phe261, and possibly Trp253 and Leu267. The D2 protein also provides ligands for the nonheme iron being His214 and His268. Normally the D2 protein does not turn over rapidly like the D1 protein but, under extreme conditions (e.g., irradiation with damaging UV light), it will be replaced by a repair process.

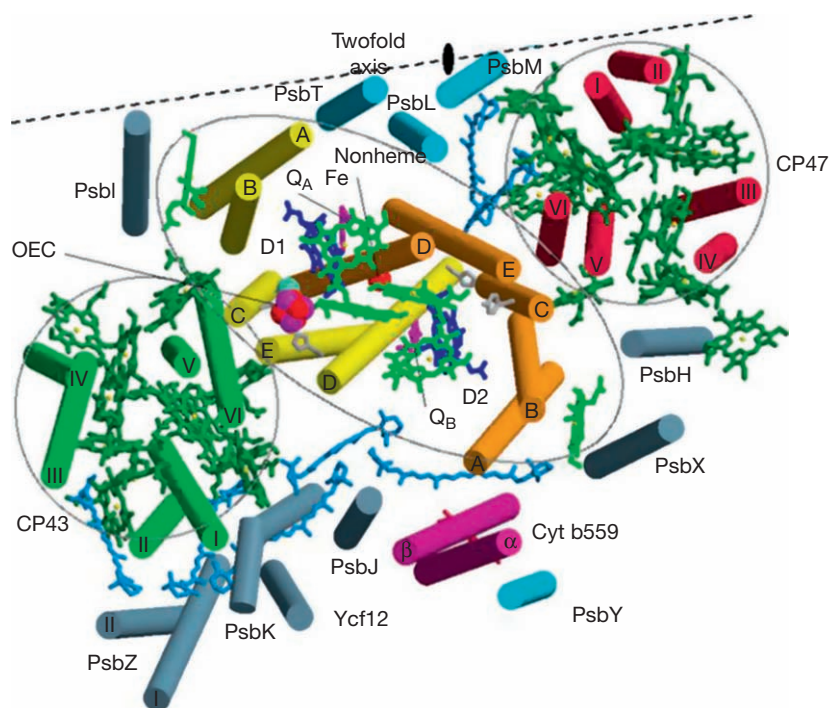


Figure 3 View of the PSII monomer along the membrane normal from the luminal side showing the positioning and transmembrane helices of the intrinsic protein subunits. They are related to same proteins in the other monomer by the twofold axis shown as a dotted line and running along the monomer/monomer interface. The pseudo-twofold axis perpendicular to the membrane plane passing through the nonheme Fe relates the transmembrane helices of the D1/D2 heterodimer, the low molecular subunits, PsbI and PsbX and CP43 and CP47 as emphasized by the black lines encircling these subunits. Coloring is the same as in Figure 1. From Ferreira K, Iverson T, Maghlaoui K, Barber J, and Iwata S (2004) Architecture of the photosynthetic oxygen-evolving centre. *Science* 303: 1831–1838, with permission from AAAS.

PsbE and PsbF

These are the α - and β -subunits respectively of cytochrome b559. After processing, the PsbE and PsbF contain 82 and 38 amino acids in most higher plants, and have molecular masses of ~ 9.3 – 4.4 kDa. Both subunits have a single transmembrane helix with their N-termini exposed to the stromal surface.

The α -subunit is characterized by having a long C-terminus of ~ 44 amino acid residues extending from the luminal surface of the membrane, while the β -subunit has essentially no luminal domain. Of considerable importance is that each subunit contains a conserved histidine residue which is located within the membrane-spanning region toward the stromal surface. These two residues form axial ligands to the heme of the cytochrome. Although it had been argued by some that there are two cyt b559 per RC, all X-ray crystal structures find only one. This cytochrome is located on the D2 side of the RC (see Figure 3). The deletion of the *psbE* gene results in a lack of assembly of PSII. On the other hand, replacement of the His residue in the α -subunit can lead to an assembled and functionally active PSII complex even when no heme is ligated to the PsbE and PsbF proteins. Normally the heme has a high redox potential of ~ 0.4 V but, under some circumstances, this potential is lowered to 0.15 V or less. The chemical basis of this redox shift and its significance, if any, are unclear. Under some circumstances, usually when the water splitting reaction is inhibited such as at low temperature, the heme of cyt b559 is oxidized by $P680^{+\bullet}$. Recent structural models show the heme to be at least $40 \text{ \AA}$ from P680, which is too far to account for

the millisecond time constant for its oxidation. Accordingly, it has been suggested that a β -carotene molecule associated with the D2 protein acts as a redox intermediate between the cyt b559 and $P680^{+\bullet}$. The reduction of the oxidized heme may involve Q_A^- , Q_B^- , or $\text{Pheo}^{\bullet-}$ but again there is a distance problem, with the required electron donation occurring over distances of $28 \text{ \AA}$ or more, according to the recent X-ray structures. It seems likely that cyt b559 plays a protective role in minimizing photodamage of the RC.

PsbG

A chloroplast gene was initially called *psbG* but it was shown later that the product of this gene was not a PSII protein but rather a component of a chloroplast-located nicotinamide adenine dinucleotide phosphate (NADPH)/quinone oxidoreductase. This gene is now known as *ndhK*.

PsbH (H-Protein)

In higher plants, the PsbH protein contains 72 amino acids and has a calculated mass of ~ 7.7 kDa. In cyanobacteria the protein is slightly smaller (~ 6.5 kDa) with a truncated N-terminus and as such does not have the N-terminal threonine residue which is conserved in all higher plant and green algal sequences which is the site for reversible phosphorylation. The functional significance of the phosphorylation process is unclear. The higher plant the PsbH protein is

characterized by having a single transmembrane helix with a long N-terminal region consisting of ~41 amino acids at the stromal surface, which reduces to ~30 residues in the case of cyanobacteria. It is located adjacent to the PsbX protein on the D2/CP47 side of the RC (see Figure 3). The *psbH* gene can be deleted from the cyanobacterium *Synechocystis* without impairing the assembly of PSII and photoautotrophic growth. The knockout mutant was, however, more sensitive to photoinhibition than the wild type. This sensitivity seemed to be due to the partial inhibition of the D1 protein repair process rather than to an increase in photochemical damage. In contrast, when the *psbH* gene was deleted in *Chlamydomonas reinhardtii*, PSII did not assemble.

PsbI (I-Protein)

The PsbI protein has a molecular mass of ~4.2 kDa. In most eukaryotes, this protein contains 35 amino acids and is predicted to have a single transmembrane helix with a short N-terminal region at its stromal end. It is located close to the D1/D2 heterodimer, being adjacent to the β -helix of the D1 protein (see Figure 3). Its position is related to the PsbX protein by the same pseudo-twofold axis which relates the transmembrane helices of the D1 and D2 proteins. Interestingly, the mature protein retains the initiating N-formyl group at its N-terminal methionine residue. The *psbI* gene has been deleted in the green alga *Chlamydomonas* and the cyanobacterium *Synechocystis*, without inhibiting PSII assembly and photoautotrophic growth. Thus, the function of the PsbI protein remains unknown.

PsbJ (J-Protein)

In most organisms, the *psbJ* gene is located in a gene cluster also containing the *psbE*, *psbF*, and *psbL* genes. The PsbJ protein is highly conserved, consists of 39 amino acids, and has a calculated molecular mass of 4.1 kDa. It has one transmembrane helix and located adjacent to the PsbE/F proteins of Cyt b559 (see Figure 3) and forms a channel for Q/PQH₂ diffusion in and out of PSII. The *psbJ* gene has been deleted in cyanobacteria to generate a knockout mutant which assembles PSII at a lower level than the wild type and has an impaired rate of Q_A to Q_B electron transfer. Nevertheless, they can grow photoautotrophically but at a rate slower than WT.

PsbK (K-Protein)

This ~43-kDa protein is highly conserved and has 37 amino acid residues. It has a single transmembrane helix with a kink due to the presence of a proline. Of particular note is that in higher plants, 24 amino acids are posttranslationally cleaved from the initial gene product. In cyanobacteria, eight amino acids are removed after translation. These pre-sequences probably bring about the insertion of the protein into the thylakoid membrane such that its N-terminus is on the luminal side. PsbK is located on the outer surface of the PSII RC core, close to helices I and II of CP43 and clustered with the PsbZ and Ycf12 proteins (see Figure 3). Deletion of the *psbK* gene in cyanobacteria has very little effect on photoautotrophic growth and PSII activity. In the case of green algae, however, the deletion of the

gene destabilized the PSII complex and the transformant was unable to grow photoautotrophically.

PsbL (L-Protein)

This highly conserved PSII protein contains 37 amino acids and has a calculated molecular mass of ~4.4 kDa. In most organisms, its gene is located in the cluster together with *psbE*, *psbF*, and *psbJ*. It consists of a single transmembrane helix and is located, together with PsbM and PsbTc, at the monomer-monomer interface of the dimeric core complex (see Figure 3). Inactivation of the *psbL* gene in cyanobacteria results in a loss of PSII-mediated oxygen evolution and the transformant is unable to grow photoautotrophically. It has also been reported that PsbL is required for normal functioning of the Q_A site based on studies with isolated PSII complexes. This conclusion was reinforced by the finding that PsbL is required for the oxidation of Y_Z by P680^{•+}. As a consequence, the primary quinone acceptor is destabilized by the increased probability of rapid recombination between Q_A⁻ and P680^{•+}.

PsbM (M-Protein)

The PsbM protein has one transmembrane helix with a very short N-terminal extension on the stromal side. It contains 33 amino acids and has a molecular mass of ~3.7 kDa. It is located with PsbL and PsbTc at the monomer-monomer interface of the dimeric core complex (see Figure 3). However, deletion of the *psbM* gene in cyanobacteria is not sufficient to prevent dimer formation which only occurs when *psbTc* is also deleted.

PsbN (N-Protein)

This protein is predicted to be in both cyanobacterial and plant PSII. However, it has not been detected directly in plant PSII and it is not present in the crystal structure of PSII isolated from cyanobacteria. PsbN has a predicted single transmembrane helix. Deletion of both *psbN* and *psbH* genes from cyanobacterium *Synechocystis* caused no effect other than those observed in the absence of *psbH* alone. Thus, the function and location of *psbN* are unknown.

PsbO (O-Protein)

PsbO is often called the 'manganese-stabilizing' protein. Between higher plants and cyanobacteria this protein is highly conserved containing, after processing, 241–247 residues. Although the mature PsbO protein is often referred to as the 33-kDa protein, its calculated molecular mass is ~26.5 kDa. It is an extrinsic protein having as its main body a β -barrel composed of eight antiparallel β -strands. A very large loop joining β -strands 1 and 2 provides a head domain which binds to the luminal surfaces of the D1, D2, CP43, and CP47 proteins (see Figure 1). The β -barrel is not hollow but full of bulky side chains. It plays an important role in stabilizing the Mn₄Ca-cluster and maintaining an optimal environment for water oxidation to occur. Various studies indicate that it does so by controlling Ca²⁺ and Cl⁻ levels at the catalytic site, but it does not bind Mn directly. The deletion of the *psbO* gene in

cyanobacteria does not prevent water oxidation unless the *psbV* gene is also deleted while the equivalent single *psbO* deletion in green alga inhibits the assembly of a functional PSII core complex.

PsbP (P-Protein)

PsbP is a 23-kDa extrinsic protein. After processing, this protein consists of ~186 amino acids with a calculated molecular mass of ~20 kDa. It consists mainly of β -sheets. Although it is a well-established component in PSII of higher plants and green algae, its presence in cyanobacteria is not stoichiometric with PSII and not present in the crystal structures. Like PsbO, its function seems to be to optimize the Ca^{2+} and Cl^- levels needed for the water oxidizing reaction and is located in the vicinity of this 33-kDa protein to which it binds.

PsbQ (Q-Protein)

The PsbQ mature protein contains ~49 amino acids and, like PsbP, is located close to the 33-kDa protein and the Mn_4 -cluster. Its core consists of a four helical bundle. Like PsbO and PsbP, it seems to be involved in optimizing the ionic environment necessary for oxygen evolution. PsbQ, however, is not found in the crystal structure of cyanobacterial PSII, although a *psbQ*-like gene is present and the protein has been detected in PSII preparation at stoichiometric levels. Its binding to PSII requires the presence of PsbO and PsbP.

PsbR (R-Protein)

The role of PsbR is unknown. It has a molecular mass of 10.2 kDa and consists of ~99 amino acids. It seems to be an extrinsic protein of plant PSII, and is bound relatively tightly to the luminal surface in the vicinity of the water splitting site. Whether it has a transmembrane helix is a matter for debate. It does not exist in cyanobacteria.

PsbS (S-Protein)

The PsbS protein consists of ~205 amino acids and has an apparent molecular mass of 22 kDa, although its calculated mass is ~29 kDa. It is predicted to have four transmembrane helices. Helices I and III, and II and IV, are homologous indicating that the protein is derived from internal gene duplication. Sequence homology studies suggest that PsbS is related to Chla/Chlb-binding antenna Lhcb1–6 proteins (*cab* gene products) and is likely to be a chlorophyll-binding protein. A functional role for PsbS, therefore, could be to act as a pigment chaperonin which aids the incorporation of chlorophyll molecules into the various pigment-binding proteins of PSII. This protein has also been implicated in the protection of PSII against photoinduced damage (photoinhibition). It does not exist in cyanobacteria.

PsbT_C (T_C-Protein)

The chloroplast-encoded PsbT_C protein is also found in cyanobacteria. It contains 30–34 amino acids, has a molecular weight of ~3.9 kDa, and retains the initiating *N*-formyl group

on its N-terminal methionine residue. It has a single transmembrane helix which is located in the N-terminus of the protein. The *psbT_C* gene was formerly known as *ycf8* and is located close to the *psbB* gene in the chloroplast genome. The first identification of this subunit was made by comparing wild type and a mutant of *Chlamydomonas* that lacked the *psbT_C* gene. This work also showed that the PsbT_C is a PSII protein and is required for optimal activity under high-light conditions so as to prevent photoinhibition. It is located with PsbM and PsbL in the dimerization domain (see Figure 3), and the deletion of its gene together with that of PsbM results in monomerization of the dimeric PSII core.

PsbT_N (PsbT_N-Protein)

The PsbT_N is a nuclear-encoded hydrophilic 5-kDa protein which copurifies with PsbO. Although its function is unknown, it seems to be an extrinsic protein located on the luminal surface of PSII. The mature PsbT_N protein consists of ~28 amino acids in higher plants and the presence of two cysteine residues suggests that it contains a disulfide bridge. It is not present in cyanobacteria.

PsbU (U-Protein)

This is a cyanobacterial all α -protein extrinsically located on the luminal surface of PSII close to the 33-kDa protein (see Figure 1). It has an apparent molecular mass of ~11 kDa and binds no cofactors. It is also found in red algae but not in plants and green algae.

PsbV (V-Protein)

PsbV is a heme-containing protein known as cytochrome c550 and is found only in cyanobacteria and red algae. It has a molecular mass of 15 kDa and is an extrinsic protein found on the luminal surface of PSII close to PsbO and PsbU (see Figure 3). It plays no direct role in water oxidation, although it does help to provide a $\text{Ca}^{2+}/\text{Cl}^-$ environment at the catalytic site. It does not, however, bind directly to PsbO but indirectly via PsbU, which forms a bridge.

PsbW (W-Protein)

The PsbW protein is found in PSII cores isolated from higher plants and green algae. A cyanobacterial PsbW-like gene has also been documented which has a low homology with the eukaryotic equivalent, and therefore there is some uncertainty whether or not cyanobacteria contain the PsbW. In any event, it is not present in the cyanobacterial PSII crystal structures. In higher plants and green algae, it is nuclear-encoded and is processed to a mature protein having an apparent molecular mass of 6.1 kDa and consists of 54–56 amino acids. It is predicted to have one membrane-spanning region with its N-terminus located on the luminal side of the membrane. This orientation is consistent with the pre-sequence of the PsbW protein having characteristics of a target sequence typical for luminal proteins. The function of PsbW is unknown.

PsbX (X-Protein)

The PsbX protein is nuclear-encoded in higher plants and green algae, and consists of ~42 amino acids. In higher plants, its synthesis is tightly regulated by light. The mature PsbX protein has a molecular mass of ~4.1 kDa and a single transmembrane helix having its N-terminus in the lumen. It has been detected in the oxygen-evolving PSII core complex of higher plants and cyanobacteria. In the latter, it is located close to the D1/D2 heterodimer and close to PsbH, being adjacent to the B-helix of the D2 protein. Its position is related to the PsbI protein by the same pseudotwofold axis which relates the transmembrane helices of the D1 and D2 (see [Figure 3](#)). The deletion of the *psbX* gene in cyanobacteria did not inhibit PSII assembly and photoautotrophic growth. However, these studies suggest that PsbX may play a role in the electron transfer functions of Q_A and Q_B .

PsbY (Y-Protein)

This protein is located in PSII of plants, algae, and cyanobacteria. It has a molecular mass of ~4.2 kDa, consists of 36–39 amino acid residues, and has one transmembrane segment. The deletion of the *psbY* gene in cyanobacteria does not impair the water oxidation process, probably because of its peripheral location which is close to PsbE (see [Figure 3](#)). The function of this subunit is unclear and was only detected in one of the crystal structures of cyanobacterial PSII.

PsbZ

This 6.8-kDa protein contains 62 amino acids and forms two transmembrane helices. It is located at the peripheral edge of the PSII complex adjacent to PsbC (CP43) where it clusters with the Psb K, PsbJ, and Ycf12 proteins (see [Figure 3](#)). This cluster probably functions to stabilize the binding of pigments, particularly β -carotene.

Ycf12 (Also Called Psb30)

Subunit Ycf12 is situated next to the PsbK, PsbJ, and PsbZ proteins (see [Figure 3](#)), having a single transmembrane helix with its N-terminus in the lumen. It directly interacts with PsbK and PsbZ and coordinates a Ca^{2+} ion. A cyanobacterial mutant with the *ycf12* gene deleted possesses oxygen-evolving activity comparable with wild type at normal light intensities.

Conclusion

Here, 27 different Psb protein subunits have been described that are found in the PSII RC core (20 intrinsic and seven extrinsic). In a few cases, the subunits are specific either to plants and green algae or to red algae and cyanobacteria. In addition, there may be some other candidates which have yet to be characterized in detail including some like Ycf48, Psb27, and Psb29 which seem to be involved in PSII assembly and maintenance. Although the X-ray structures of the cyanobacterial PSII RC core have unambiguously assigned all proteins in the electron density maps, none of these proposed assembly proteins are present. In addition to protein assignments, the most recent X-ray structures of the cyanobacterial PSII RC core have also identified, per monomer, the positions of 35 chlorophylls, 11 all-*trans* β -carotenes, at least 20 lipids (six monogalactosyldiacylglycerols (MGDGs), five digalactosyldiacylglycerols (DGDGs), four sulfoquinovosyldiacylglycerols (SQDGs) and five phosphatidylglycerols (PGs)), 15 detergent molecules, and more than 1300 water molecules of which almost all are located in the two layers located on the stromal and luminal sides.

See also: [Bioenergetics: Chlorophylls and Carotenoids](#); [Chloroplasts](#); [Dynamic Behavior of Photosystem II Light Harvesting](#); [Light-Harvesting Complex I and II: Pigments and Proteins](#); [Photosystem II: Assembly and Turnover of the D1 Protein](#); [Photosystem II: Water Oxidation, Overview](#).

Further Reading

- Barber J (2003) Photosystem: II. The engine of life. *Quarterly Reviews of Biophysics* 36: 71–89.
- Barber J, Nield J, Morris EP, Zheleva D, and Hankamer B (1997) The structure, function and dynamics of photosystem two. *Physiology of Plant* 100: 817–827.
- Blankenship RE (2002) *Molecular Mechanisms of Photosynthesis*. Chichester: Blackwell.
- Ferreira K, Iverson T, Maghlaoui K, Barber J, and Iwata S (2004) Architecture of the photosynthetic oxygen-evolving centre. *Science* 303: 1831–1838.
- Guskov A, Kern J, Gabdulkhakov A, et al. (2009) Cyanobacterial photosystem II at 2.9 Å resolution and the role of quinones, lipids, channels and chloride. *Nature Structural and Molecular Biology* 16: 334–342.
- Hankamer B, Morris EP, Nield J, Carne A, and Barber J (2001) Subunit positioning and transmembrane helix organisation in the core dimer of photosystem II. *FEBS Letters* 504: 142–151.
- Loll B, Kern J, Saenger W, Zouni A, and Biesiadka J (2005) Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II. *Nature* 438: 1040–1044.
- Ort DR and Yocum CF (1996) *Oxygenic Photosynthesis. The Light Reactions, Advances in Photosynthesis*, vol. 4. Dordrecht: Kluwer.
- Umena YK, Kawakami K, Shen J-R, and Kamiya N (2011) Crystal structure of a oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* 473: 55–61.

Photosystem II: Water Oxidation, Overview

F Rappaport, Centre National de la Recherche Scientifique, Paris, France

A Boussac, Centre National de la Recherche Scientifique, Gif-sur-Yvette, France

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Fabrice Rappaport and Pierre Joliot, volume 3, pp. 375–380, © 2004, Elsevier Inc.

Glossary

Chlorophyll excited state State that is formed after absorption by a chlorophyll molecule of a light quantum. At the molecular orbital level, it corresponds to the hopping of an electron from the highest occupied molecular orbital to the lowest unoccupied molecular orbital.

Oxidation De-electronation reaction or abstraction of an electron from a molecule.

Reduction Electronation reaction. An electron transfer reaction occurs between an electron donor

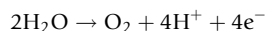
(which is oxidized) and an electron acceptor (which is reduced). In the case of photosynthetic water oxidation, two molecules of water (H_2O) are oxidized into molecular oxygen and serve as an electron source.

S_i states Successive states, as defined by B Kok, formed during the accumulation in the catalytic center of the water-splitting enzyme of the oxidizing power required to split water into dioxygen. The transition from state S_i to state S_{i+1} reflects the abstraction of one electron from the catalytic center.

Photosystem II and the Water-Splitting Enzyme

The Photosystem II Complex

Water oxidation into dioxygen (O_2) follows the equation



where H^+ and e^- stand for protons and electrons, respectively. Water is obviously a very stable compound and its oxidation requires the availability of strongly oxidizing species (at pH 7.0 the midpoint potential of the $\text{H}_2\text{O}/\text{O}_2$ couple is 810 mV). This oxidizing power is formed at the expense of light absorption and the conversion of light energy into electrochemical energy. The action spectrum of oxygen emission showed that it is sensitized by pigments associated with photosystem II (PS II). Photosystem II is a large pigment protein complex located in the thylakoid membrane of the chloroplast, in green algae or higher plants or in the cytoplasmic membrane in cyanobacteria. As illustrated in [Figure 1](#), the core of PS II is composed of two homologous polypeptides (D_1 and D_2 , D for diffuse electrophoretic profile), each of which consists of five transmembrane helices. It bears various redox cofactors, including a chlorophyll dimer ($\text{P}_\text{A}\text{P}_\text{B}$), two chlorophyll monomers (B_A and B_B), two pheophytins (Ph_A and Ph_B), and two quinones Q_A and Q_B . Among the various photosynthetic reaction centers, PS II is unique because it bears a tetra-manganese cluster which, coupled to a redox active tyrosine Y_Z , is responsible for water splitting into oxygen. The core of the reaction center is surrounded by pigment (chlorophyll and carotenoid)-binding proteins, which increase by a factor of ≈ 200 its absorption cross section. Thus, under normal solar radiance, one photon is absorbed by PSII every ≈ 5 ms. The absorption of a photon by one of the chlorophyll borne by the pigment-binding proteins results in the formation of an excited state of a chlorophyll, that is, an electron jumps from the highest occupied molecular orbital to the lowest unoccupied orbital of higher energy. Provided other pigments are close enough, this excitation energy may hop from one pigment to another and then be funneled to the core of the reaction center and

more precisely to the primary donor B_A (denoted B_A^* in the excited state). The singularity of B_A resides in the fact that in its close vicinity one finds an electron acceptor (Ph_A). The excitation energy, by weakening the binding energy of the electron by the energy of the transition between the two molecular orbitals, allows the electron transfer reaction between B_A^* and this nearby acceptor to occur, yielding the charge-separated state ($\text{B}_\text{A}^+\text{Ph}_\text{A}^-$). This state is then further stabilized by successive electron transfer steps between the different redox compounds. At room temperature, the first oxidized cofactor is essentially P_A . It is reduced in a few tens of nanoseconds by the redox active Y_Z (Y161 of the D_1 polypeptide), which is, in turn, reduced at the expense of the water-splitting enzyme. As shown in [Figure 1](#), despite the arrangement of the various redox cofactors in two symmetrical branches with respect to a pseudo C_2 symmetry axis perpendicular to the membrane plane, electron transfer only proceeds down one of these two branches indicating that function does not strictly follow structure.

The Water-Splitting Enzyme

As just described, conversion of light energy into electrochemical energy is a one-electron process. However, water oxidation into dioxygen is a four-electron process. Taking advantage of the possibility to trigger the light-induced electron-transfer reaction by short light flashes, P. Joliot showed that the oxygen emission yield is maximum after the third, seventh, eleventh... flash of a series ([Figure 2](#)). This oscillating pattern with a periodicity of four was modeled by B. Kok who postulated, on this basis, the existence of a charge-storing device allowing, in a single catalytic unit associated with PSII, the accumulation of the four oxidizing equivalents needed to split water into oxygen. In this model, illustrated in [Figure 2](#), the catalytic center has five different oxidation states (S_0 , S_1 , S_2 , S_3 , and S_4). Four successive light-induced oxidation triggers the four successive transitions between the S_i and S_{i+1} states. Upon S_4 formation, oxygen is produced and released while S_0 is

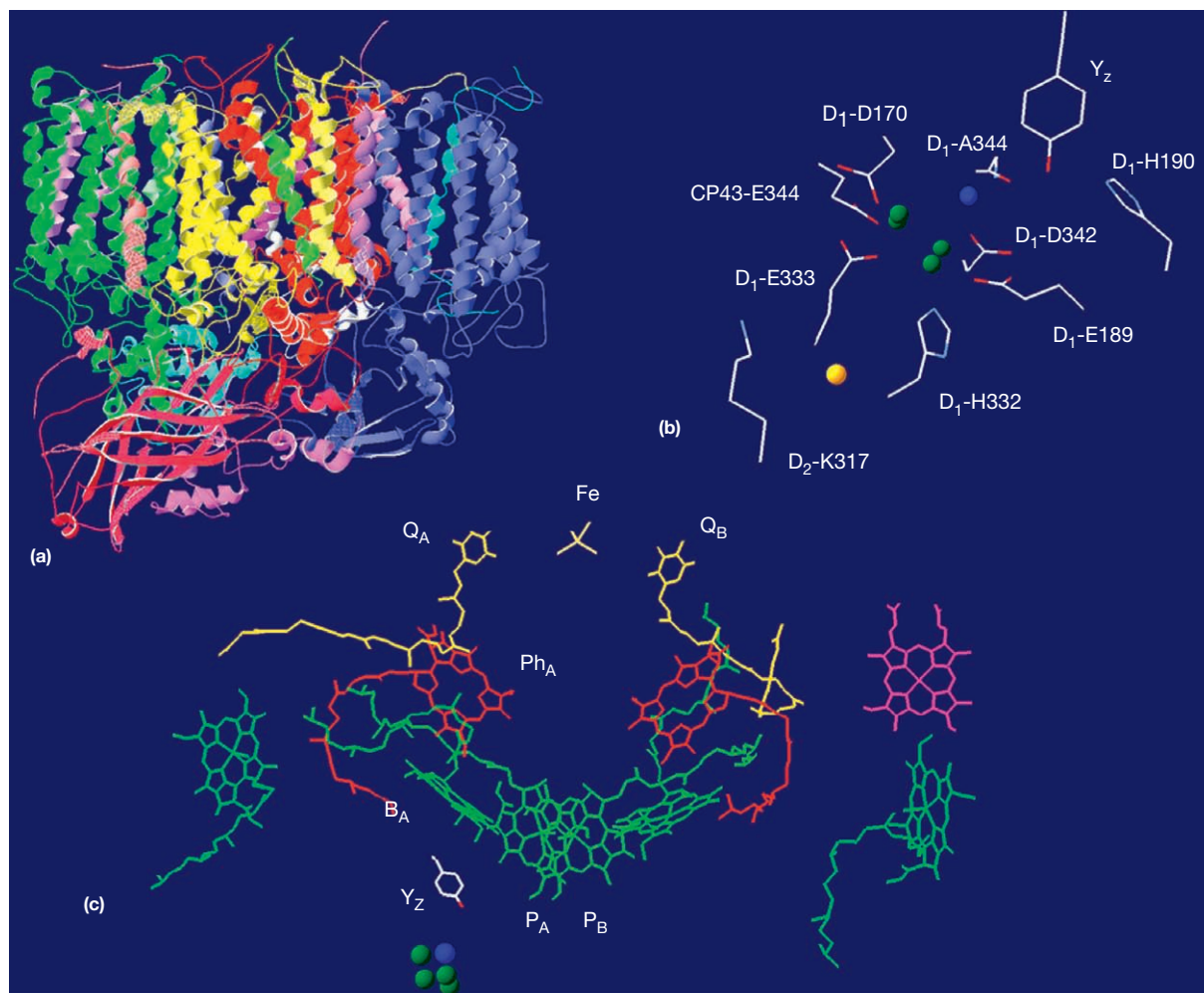


Figure 1 A structural model of photosystem II. (a) The 17 membranous proteins and the three extrinsic proteins. (b) Some amino acids of the first coordination sphere of the Mn₄Ca-Cl cluster (Mn in dark green, Ca in dark blue, and Cl in yellow). (c) Structural arrangement of the redox-active components of photosystem II. Based on X-ray diffraction studies by Ferreira et al. (2004) Architecture of the photosynthetic oxygen-evolving center. *Science* 303: 1831–1838 and Loll B, Kern J, Saenger W, Zouni A, and Biesiadka J (2005) Towards complete cofactor arrangement in the 3.0 angstrom resolution structure of photosystem II. *Nature* 438: 1040–1044, PDB entries 1S5L, 2AXT and 3BZ1.

regenerated. The beauty of the Kok model resides in its ability to fully describe the function of the enzyme without any assumption neither on the chemical nature of the transient states it involves nor on the precise mechanism underlain by the successive transition between the S states. This accounts for its success as a working model ever since it was enounced. Simultaneous to the discovery of this charge-storing device, G. Cheniae and co-workers evidenced the requirement for manganese atoms in water splitting. The determination of a stoichiometry of four Mn atoms per PSII complexes provided the chemical basis for the catalytic center where the accumulation of the oxidizing equivalent produced by light-induced charge separation occurs. Together with this manganese cluster, the catalytic center bears two ions, Ca²⁺ and Cl[−]. In the absence of either one of these two ions, the lower S states (S₀ and S₁) may undergo oxidation but the advancement of the Kok cycle cannot be completed and water splitting is inhibited. The water-splitting enzyme further comprises extrinsic polypeptides (Figure 1(a)), which stabilizes the assembled tetranuclear cluster.

The Assembly of the Manganese Cluster

The ligands involved in the Mn₄Ca cluster binding to the protein backbone are provided by the polypeptides D₁ and CP43 (Figure 1(b)). The assembly of the Mn cluster requires Mn^{II}, in the Mn(OH)⁺ or Mn(OH)₂ form, Ca²⁺, Cl[−], and light. The absorption of light by PSII produces the oxidant Y_z^{ox}, which oxidizes the Mn^{II} ions. The first oxidation of a Mn^{II} has a high quantum yield, unlike the following steps which are needed for the oxidation-induced association of the stable tetra nuclear cluster.

The Thermodynamic and Kinetic Features of the Water-Splitting Enzyme

Before attempting to address the mechanism of water oxidation, one should question the thermodynamic and kinetic issues of this reaction. Obviously, water is a very stable molecule under physiological conditions, while its oxidation into

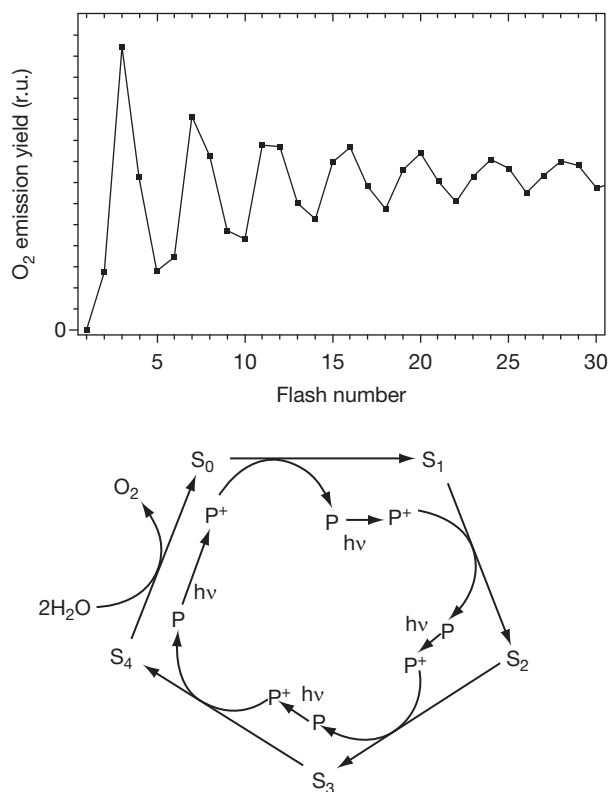


Figure 2 (Top) The O_2 emission yield as a function of the flash number in a series. Note that this yield depends on the flash number with a periodicity of four. (Bottom) The charge-storing device allowing accumulation of the oxidizing power needed to split water as modeled by Kok.

molecular oxygen is likely to involve the transient formation of highly reactive species such as oxo-radicals, hydrogen peroxides, etc. The enzyme must meet the following requirements: accumulate enough oxidizing power to not only perform a very demanding reaction with respect to common biological redox potential but also prevent side reactions, which are thermodynamically made possible by the accumulation, in a single catalytic site, of such an oxidative power.

The Redox Potential of the Different S States

Until now, the redox potentials of the different S states have not been determined by direct redox titration. They could only be inferred from the redox potentials of other cofactors found in the PS II reaction centers combined with the free-energy change associated with the electron transfer reaction between the S states and these cofactors. There are thus large error bars on the value, which can be found in the literature. The redox potential of the P^+/P couple, which drives the catalytic cycle, has been estimated to ≈ 1.25 V, making the P^+ species the strongest oxidant found in biology. With an estimated midpoint potential of ≈ 0.75 V, the S_1/S_0 couple lies ≈ 0.5 V below the P^+/P couple, whereas the S_2/S_1 , S_3/S_2 , and S_4/S_3 midpoint potentials are estimated to ≈ 1 V. The overall free energy available to oxidize water is thus ≈ 3.75 V. This is sufficient to

drive the reaction at neutral pH (the minimal driving force required to reversibly oxidize water at pH 7 is 0.81 V per electron) but becomes low when the reaction has to occur at a pH close to 5, a value often quoted when the photosynthetic chain is at work.

The Rates of Oxidation of the Various S States

The rates of the various electron transfer reactions between the oxidized tyrosine Y_Z^{ox} and the S-states S_i are in the tens or hundreds of microseconds for the three lower S states. The S_4 to S_0 transition is significantly slower, in the millisecond time range, and concomitant with O_2 production and release. Based on the distance between the electron acceptor Y_Z^{ox} and the Mn cluster (7 Å) combined with the free-energy changes associated with these reactions, these rates appear much too slow for the reaction to be kinetically limited by the electron transfer proper. This has been taken as an indication of either redox-induced structural changes or/and a coupling with other processes rather than electron transfer such as proton transfer. Interestingly, a lag phase of a few tens of microseconds preceding the decay of S_4 into S_0 , which is accompanied by O_2 release, might evidence such events.

The Structure of the Catalytic Center and the Redox Events during the Reaction Cycle

Despite the tremendous amount of effort put into its elucidation, the structure and the valence of the Mn_4Ca cluster are still matters of strong debate. Indeed, the group of Yachandra has shown that the geometry and ligand environment of the Mn_4Ca cluster in the crystal structure suffers from the progressive reduction by the X-ray beam of the native high valence manganese cluster back to the Mn^{II} state and from a concomitant change in the delicate structural arrangement. Therefore, the combination of various spectroscopic methods such as electron paramagnetic resonance (EPR) and X-ray absorption spectroscopy (XAS) will still be required to provide a set of possible structures in the relaxed S_1 state and in the other higher oxidation states.

The Structure of the Manganese Cluster

The light-induced formation of the S_2 state results in a remarkable EPR signal, which was first reported by Dismukes and Siderer. This signal shows very rich hyperfine structures (it is commonly referred to as the 'multiline' signal), which have been taken as the spectroscopic signature of a mixed valence dinuclear $Mn^{III}-Mn^{IV}$ cluster. The multiline signal has provided experimental basis for modeling the structure and the valence state of the Mn_4Ca cluster in the S_2 state. EXAFS, pioneered by M. Klein, has provided valuable data on the distances separating the heavy atoms such as Mn or Ca, which are found in the catalytic center. These data put severe constraints on the possible structures. From both the EPR and EXAFS data, a dimer of di- μ -oxo $Mn^{III}-Mn^{IV}$ dimers has long been favored. However, several other structures with oxo-bridged Mn atoms are compatible with the data. Although not settled yet, the debate seems to be converging to structures

with three strongly magnetically coupled Mn and one weakly coupled Mn resulting in a mixed-valence cluster commonly called a 3 + 1 structure.

The Redox Events during the Catalytic Centers

As mentioned earlier, EPR and X-ray spectroscopies have also generously fed the debate on the valence state of the tetranuclear cluster during the catalytic cycle. Mn edge absorption spectroscopy has provided strong evidences for oxidation of a Mn atom during the S_0 to S_1 and S_1 to S_2 transitions. EPR and (less clearly) ultraviolet (UV) spectroscopies supported this view. It is now generally agreed that the valences of the Mn ions are $Mn^{III}_3Mn^{IV}$ in S_0 , $Mn^{III}_2Mn^{IV}_2$ in S_1 , and $Mn^{III}Mn^{IV}_3$ in S_2 . Based on a kinetic analysis, the S_4 is equivalent to the $[S_3Y_Z^{ox}]$ state. This seems to ground the accumulation of four oxidizing equivalents within the catalytic center (which would then include Y_Z , the Tyr161 of D1 in Figure 1, as a partner) before water is oxidized. Yet, the assignment of the substrate of the oxidation associated with the S_2 to S_3 transition is still an open question. This ambiguity mainly comes from the fact that depending on the group involved in these studies, EXAFS shows either a shift in the absorption edge consistent with a Mn oxidation, or little, if any, shift in the absorption edges, ruling out this latter assignment and favoring the formation of a radical species. Unfortunately, other spectroscopic methods such as EPR or UV absorption did not provide either unequivocal result. Recently, the EPR signature of the S_3 state has been characterized and corresponds to a spin = 3 state. As a consequence, the oxidation associated with the S_2 to S_3 transition likely occurs at the level of the cluster, that is, it involves either a Mn ion or a ligand belonging to the first coordination sphere (a bridging O atom or an amino acid ligand).

The Mechanism of Water Oxidation

In the above-mentioned model, oxidation of a bridging O atom is equivalent to the formation of a radical species. The possibility arises that this O atom originates from a water molecule which could therefore be oxidized during the S_2 to S_3 transition. If, on the contrary, the three successive steps leading to the formation of S_3 only involve Mn oxidation, an obvious conclusion is that water oxidation only takes place after the formation of S_4 (i.e., $S_3Y_Z^{ox}$). As discussed, this alternative could not be settled by spectroscopic approach. However, recent experiments on substrate binding tilt the scale toward the latter and further constrain the inventiveness when attempting to draw possible mechanism for water splitting.

Substrate/Bulk Water Molecule Exchange

In a series of very elegant studies, Messinger, Hillier, and Wydrzynski have used isotopically labeled water ($H_2^{18}O$) and measured the release of either O_2^{34} or O_2^{36} after a rapid mixing with PSII in the presence of unlabeled water. Most interestingly, they observed that substrate water may be exchanged with bulk water up to the S_3 state. This has been, until now, the strongest result in favor of a O–O bond-formation step occurring after S_3 is formed and thus during the S_4 to S_0 transition.

Proton Release during the Catalytic Cycle

The oxidation of two water molecules releases not only oxygen as a by-product but also four protons. From the pioneering work of Fowler, it is known that proton release accompanies the S_i to S_{i+1} transitions, which precedes the final S_4 to S_0 step during which O_2 is released. However, it is commonly agreed that these proton release events do not result from water chemistry. Indeed, proton transfers are often associated with redox changes either when the species which undergo the redox change is itself protonated or when the change in electrostatic potential resulting from the redox change shifts the affinity of a nearby acid group for proton. The two nonexclusive processes are likely to be at work, simultaneously or not, during the S_0 to S_1 , S_1 to S_2 , and S_2 to S_3 steps. The different groups which are deprotonated during these steps would be reprotonated by the protons produced during water oxidation. Although they most probably do not originate from water splitting, the proton release events occurring during the first three steps do play an important role in the overall water chemistry. They obviously compensate the charge accumulation which would otherwise result from three successive electron abstractions from a single catalytic unit. They control the thermodynamics of the enzyme by modulating the redox potential of the successive intermediate. In addition, the phenolic oxygen of the tyrosine side chain is protonated when Y_Z is in the reduced state, whereas Y_Z^{ox} is deprotonated. Its reduction must thus be accompanied by its reprotonation. This led G. Babcock to propose that Y_Z could act as H atom abstractor undergoing a simultaneous reprotonation and reduction at the expense of water molecule. Although it seems to be agreed that this process does not occur on each step, as previously proposed, this model is largely resorted to for the higher S-states transitions.

The Mechanisms for O_2 Formation

As discussed above, many issues regarding the catalytic cycle are still obscure. This may explain why there are almost as many mechanistic models for water splitting as groups studying this fascinating enzyme. It is beyond the scope of this article to present them all. As described, Mn atoms are oxidized during some, if not all, transitions between S states. Obviously, many models have incorporated these results and are based upon Mn-centered redox chemistry. Currently, three possible mechanisms are considered for the formation of the O = O double bond: (1) the coupling between two O bridging two Mn ions, (2) a mechanism which involves a terminal oxyl radical bound to a Mn ion (or bridging two Mn ions), and (3) a nucleophilic attack from a water molecule bound to the Ca^{2+} ion on the second water molecule bound to a Mn ion. In all cases, the two cofactors Ca^{2+} and Cl^- play a crucial role. Last, the tyrosine Y_Z could play a crucial catalytic role. As discussed above, the S_4 state is equivalent to the $[S_3Y_Z^{ox}]$ state so that Y_Z may be considered as a full member of the catalytic center. The close proximity between Y_Z and the Mn cluster supports this view.

Perspective

Obviously, important questions such as when and how does water oxidation chemistry come into play are still unsettled.

However, as illustrated by the recent crystallographic structures, the expected improvement in our knowledge of PSII at the atomic level should clarify if not solve some of the remaining ambiguities. Yet, it will not put an end to the sparking development of models and concepts, since identifying the different intermediate states during the overall reaction is a constantly renewed challenge, which is likely to be taken up by the field. Understanding the water chemistry became recently a renewed challenge for the biochemists and chemists owing to the increasing demand of new CO₂-free fuels.

See also: **Bioenergetics:** Bioenergetics: General Definition of Principles; Dynamic Behavior of Photosystem II Light Harvesting; Photosystem II: Redox and Protein Components.

Further Reading

- Cramer WA and Soriano GM (2000) Thermodynamics of energy transduction in biological membranes. *Biophysics Textbook*. <http://www.biophysics.org/Portals/1/PDFs/Education/cramer.pdf>.
- Debus RJ (2007) Protein ligation of the photosynthetic oxygen-evolving center. *Coordination Chemistry Reviews* 252: 244–258.
- Diner BA and Babcock GT (1996) Structure, dynamics, and energy conversion efficiency in photosystem II. In: Ort DR and Yocum CF (eds.) *Oxygenic Photosynthesis: The Light Reactions*, pp. 213–247. Dordrecht: Kluwer.
- Ferreira KN, Iverson TM, Maghlaoui K, Barber J, and Iwata S (2004) Architecture of the photosynthetic oxygen-evolving center. *Science* 303: 1831–1838.
- Guskov A, Kern J, Gabdulkhakov A, Broser M, Zouni A, et al. (2009) Cyanobacterial photosystem II at 2.9 Å resolution and the role of quinones, lipids, channels and chloride. *Nature Structural and Molecular Biology* 16: 334–342.
- Herrero C, Lassalle-Kaiser B, Leibl W, Rutherford AW, and Aukauloo A (2008) Artificial systems related to light driven electron transfer processes in PSII. *Coordination Chemistry Review* 252: 456–468.
- Hoganson CW and Babcock GT (1997) A metalloradical mechanism for the generation of oxygen from water in photosynthesis. *Science* 277: 1953–1956.
- Lavergne J and Joliot P (2000) Thermodynamics of the excited states of photosynthesis. *Biophysics Textbook* <http://www.biophysics.org/Portals/1/PDFs/Education/cramer.pdf> (accessed March 2011).
- Lavergne J and Junge W (1993) Proton release during the redox cycle of the water oxidase. *Photosynthesis Research* 38: 279–296.
- Lubitz W, Reijerse EJ, and Messinger J (2008) Solar water-splitting into H₂ and O₂: Design principles of photosystem II and hydrogenases. *Energy and Environmental Science* 1: 15–31.
- Nicholls DG and Ferguson SJ (2002) *Bioenergetics* 3. London: Academic Press.
- Rutherford AW and Boussac A (2004) Water photolysis in biology. *Perspectives. Science* 303: 1782–1783.
- Siegbahn PEM and Blomberg MRA (2010) Quantum chemical studies of proton-coupled electron transfer in metalloenzymes. *Chemical Reviews* 110: 7040–7061.
- Umena Y, Kawakami K, Shen JR, and Kamiya N (2011) Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* 473: 55–65.
- Yano J, Kern J, Sauer K, et al. (2006) Where water is oxidized to dioxygen: Structure of the photosynthetic Mn₄Ca cluster. *Science* 314: 821–825.

Plasma-Membrane Calcium Pump: Structure and Function

M Brini, University of Padua, Padua, Italy

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Emanuel E. Strehler, volume 3, pp. 385–389, © 2004, Elsevier Inc.

Glossary

Adenosine triphosphate (ATP) The biochemical molecule utilized by the cells as energy source.

Alternative splicing A posttranscriptional process responsible for the maturation of a primary gene transcript and the generation of different messenger RNAs (mRNAs). Translation of these differently spliced mRNAs leads to protein products (isoforms) that differ only in the region affected by the alternative splice.

Calcineurin A serine/threonine phosphatase controlled by cellular calcium.

Calmodulin A calcium-binding protein expressed in all eukaryotic cells.

Calpains Proteases responsible for Ca^{2+} -regulated degradation of specific proteins.

Caspases Protein family responsible for degradation of specific proteins that contain recognizing sequences.

Their activation leads to signals of programmed cell death.

Cochlear hair cells Sensory cells in the organ of Corti of the inner ear, in synaptic contact with sensory as well as efferent fibers of the auditory nerve.

Isoform A protein structurally and functionally similar to a given protein.

Mechanoelectrical transduction (MET) channels Channels located in hair-cell stereocilia and permeant to Ca^{2+} .

PDZ domain A common structural domain of 80–90 amino acids found in the signaling proteins of bacteria, yeast, plants, viruses, and animals.

Plasma membrane One biological membrane separating the interior of a cell from the outside environment.

Stereocilia Apical structures of the hair cells of the inner ear that transduce mechanical energy into electro-chemical messages.

Structural and Regulatory Characteristics

The PMCA pump belongs to the family of P-type ATPases, which utilizes adenosine triphosphate (ATP) energy in the form of a phosphorylated enzyme intermediate (hence P-type) formed between the γ -phosphate of hydrolyzed ATP and an invariant D-residue in a highly conserved sequence of the pump molecules: SDKTGT[L/I/V/M][T/I/S]. The pump operates with a 1:1 Ca^{2+} /ATP stoichiometry as a Ca^{2+} : H^{+} exchanger: the matter of Ca^{2+} / H^{+} stoichiometry is still controversial. The pump has high Ca^{2+} affinity and a multitude of agents that modulate it. The K_d of the pump for Ca^{2+} , which is 10–20 μM in the resting state, decreases to less than 1 μM following calmodulin interaction. Acidic phospholipids also increase the Ca^{2+} affinity of the pump. Even if the molecular mechanism of the activation by phospholipids is obscure, it could be physiologically meaningful: it has been calculated that in the membrane environment the pump is probably permanently activated by acidic phospholipids to about 50% of its maximal activity.

Structurally, the pump is predicted to consist of 10 transmembrane domains, two large intracellular loops, and of N- and C-terminal cytoplasmic tails (Figure 1). The 90-residue N-terminal tail contains a consensus binding site for the 14-3-3 protein, which plays an inhibitory role on pump activity. The first cytosolic loop between transmembrane domains 2 and 3 is considered the transducer domain. It contains one of the two sites that mediate the activation by acidic phospholipids and one of the two sites for the autoinhibitory interaction with the calmodulin-binding domain. This loop also contains site A, which is one of the two main sites of the

alternative splicing that generates pump variants. Interestingly, the A-splice-site-insert also plays a key role in the apical targeting of PMCA.

The cytosolic loop that connects transmembrane domains 4 and 5 contains the catalytic center that includes the ATP-binding site and the aspartate residue that forms the acyl phosphate intermediate. It also contains the second binding site for the C-terminal calmodulin-binding domain. The C-terminal tail contains the main regulatory sites for the activity of the pump: the calmodulin-binding domain (which also binds acidic phospholipids), consensus sites for protein kinases A (PKA) and C (PKC), and high-affinity allosteric Ca^{2+} -binding sites. The calmodulin-binding domain of the PMCA pump acts as an autoinhibitory domain, binding to receptor sites in the first and second cytoplasmatic loops of the pump. Calmodulin would remove the domain from the binding sites, freeing the pump from autoinhibition.

The C-terminal region of the pump is also the target of isoform-specific phosphorylations by PKA and PKC. Phosphorylation could influence the regulation of the pumps either directly (e.g., by de-inhibiting it) or indirectly (e.g., via interference with calmodulin regulation). Recent studies have shown that PKC affects the various PMCA isoforms and splicing variants in different ways according to variations in the C-terminal sequence: the phosphorylation may activate the pump, inhibit it, or have no effects. As for PKA, its consensus site has only been described in isoform 1 of the pump. The PMCA pump is also a substrate of intracellular proteases, as its C-terminal tail contains target sequences for the Ca^{2+} -dependent protease calpain and for caspase 1 and 3. Calpain cleaves the pump in the calmodulin-binding sequence, leading

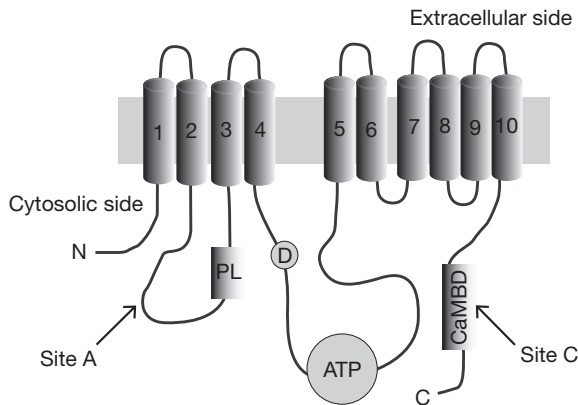


Figure 1 Topography model of the plasma membrane Ca^{2+} -ATPase. The 10 putative transmembrane domains are numbered and indicated by shadow boxes, PL indicates the phospholipid-binding domain downstream site A of alternative splicing, D indicates the catalytic aspartate. ATP and CaMBD indicate the ATP-binding site and the Calmodulin-binding domain, which contains the site C of alternative splicing.

to permanent activation of the pump. Both activation and inactivation have been described for caspases 1 and 3 action on PMCA2 and 4.

The C-terminal domain of the pump also contains an important site of alternative splicing, site C. C-site splicing occurs in all pump isoforms (albeit with varying degrees of complexity) and causes the inclusion of one (or two) additional exons, or of a portion of exons. The insertion has structural and functional consequences: it leads to the truncation of the pump due to a frame shift in the coding region, and alters, for instance, the pH sensitivity of the interaction of calmodulin with the calmodulin-binding domain. In general, the truncated isoforms have much decreased affinity for calmodulin. Surprisingly, however, the full length and the truncated variants expressed in living cells appear to have the same ability to restore the cytosolic Ca^{2+} transient generated by cell stimulation: the finding suggests that *in vivo* calmodulin may not be the only factor regulating PMCA activity.

The C-terminal domain also mediates the interaction with PDZ (postsynaptic density protein, PSD-95; Drosophila disk large tumor suppressor, Dlg; Zonula occludens-1, ZO-1) domains of different proteins, and mediates PMCA dimerization (which also activates the pump). The interaction with PDZ domains is apparently limited to the full-length splice isoforms, as the C-terminally truncated variants lack the consensus site for its binding. PMCA has been reported to interact with PDZ-domain-containing proteins involved in the recruitment and maintenance of membrane proteins in specific membrane domains, such as members of the membrane associated guanylate kinase (MAGUK) family, that is, synapse-associated protein (SAP), calcium-calmodulin-dependent serine protein kinase (CASK), the Na/H exchanger regulatory factor 2, PISP, and the Ania3/Homer protein. These proteins could contribute to targeting and retention of PMCA in specialized domains. The increased local pump concentration could regulate Ca^{2+} signaling. The PDZ-binding domain also directs the interaction of one of the PMCA-isoforms (4b) with nitric oxide synthase 1 (NOS-1) in a ternary complex with alpha-syntrophin.

The double interaction would thus downregulate the production of NO by decreasing Ca^{2+} in the immediate vicinity of the synthase: that is, the pump could function as a modulator of signal transduction, in addition to regulating cell Ca^{2+} .

Isoforms and Tissue Distribution

The four basic PMCA isoforms (the PMCA is the product of four separate genes) have tissue-specific expression. While PMCA1 and PMCA4 are expressed in most tissues, PMCA2 and 3 are found in a restricted number of tissues, among them in brain, striated muscle, and the mammary gland. In particular, PMCA2 is abundant in specialized cells such as cerebellar Purkinje neurons and cochlear hair cells, and it is also present in the uterus and in lactating mammary glands. It is also expressed in significant amounts in the liver and kidney. PMCA3 distribution is more restricted, the choroid plexus expressing significant amounts of it.

The transcript of each of the four genes encoding PMCA pumps is subject to alternative splicing: the sites where it occurs are named A, B, and C. Of the large number of splice variants theoretically possible, about 30 have been detected at the RNA or protein levels (Figure 2). Alternative splicing at site A and C occurs for all four isoforms (however, PMCA1 is never spliced at site A). Alternative splicing at site B has been described only for human isoform 1 and 4 and it could be artifactual (see below). Splice site A is located upstream of one of the regulatory binding sites for acidic phospholipids (see above) and involves up to three exons that can be optionally inserted or excluded. In PMCA3 and 4, a single exon may be included or excluded in the mature messenger RNA (mRNA), thus generating the x and z splice variants, respectively. The situation in PMCA2 is more complex: in the human pump, three exons of 33, 60, and 42 nt, alternatively used, could in principle originate eight different variants. However, only four of them have so far been detected as mRNAs in a variety of tissues. Variant w includes all three exons, variant z excludes all three exons, variant x includes the 42 nt exon, and variant y includes the 33 and 60 nt exons. This alternative splicing was shown to be essential in the apical targeting of PMCA and could thus also be involved in the regulation of the pump distribution in specialized portions of the plasma membrane (directly or through the interaction with other molecular determinants). Despite the splice site A being contiguous to one of the two phospholipid-binding domains, the insertion of additional exons does not affect the phospholipid regulation of PMCA isoforms. As mentioned, the splicing at site B may be an artifact, as its occurrence *in vivo* has been a point of controversy. It has been proposed to remove a C-terminal 108 nt segment, leading to the loss of the 10th transmembrane domain and causing a reorganization of the pump topology. Splice site C is located in the C-terminal of the PMCA protein, within the calmodulin-binding domain: the inclusion of a large exon (154–175 nt) alters the C-terminal half of the domain in all isoforms and changes the reading frame of the remaining C-terminal tail. This splicing is thus responsible for the generation of variant a, truncated after the first 12 residues of the calmodulin-binding domain. The variant has reduced responsiveness to calmodulin with respect to the full-length

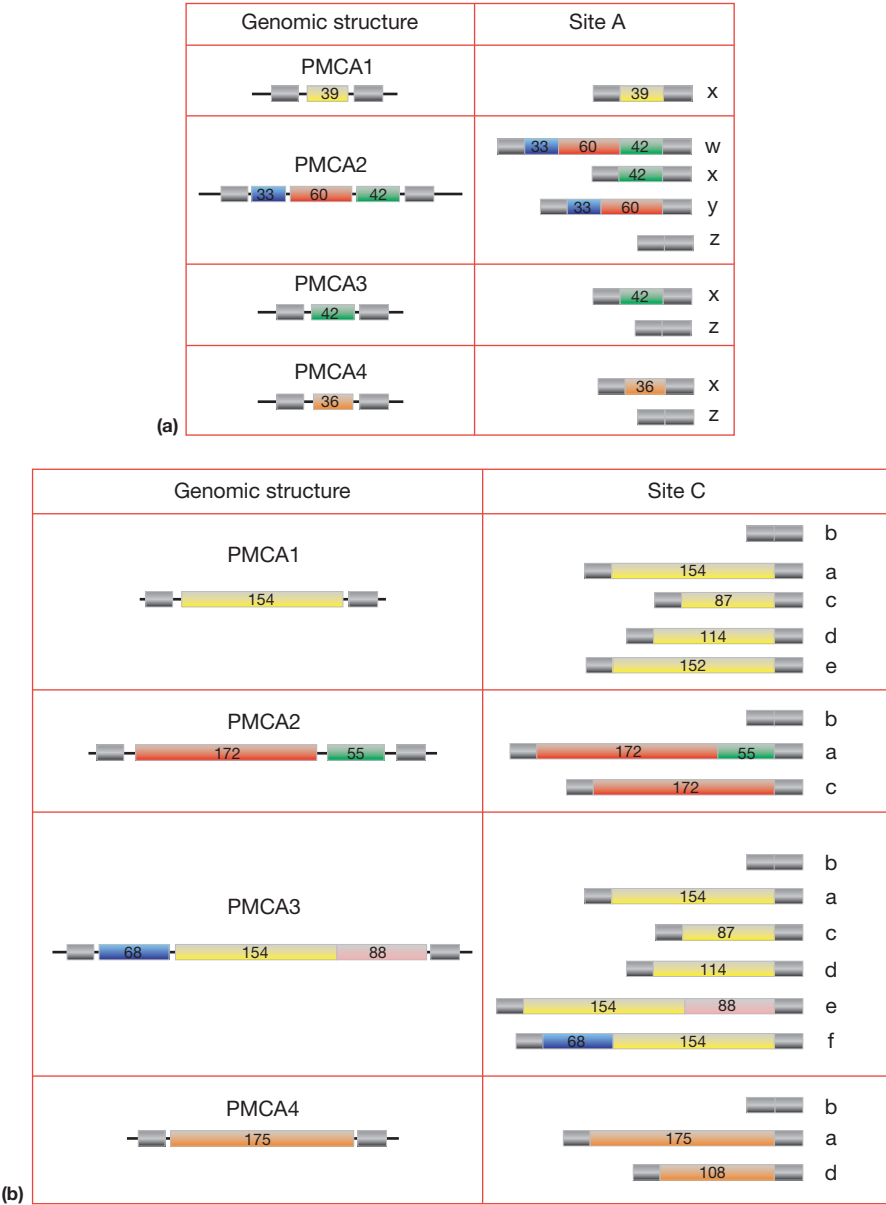


Figure 2 Linear representation of the alternative splicing options at site A (a) and site C (b) of the PMCA transcripts. Exons are indicated by shadow boxes and introns by the black lines. The numbers in the boxes represent the nucleotide number of each exon.

variant (termed ‘variant *b*’). Lower Ca-calmodulin affinity and a significantly elevated basal pumping activity are general characteristics of the *a* splice variants in comparison with the *b* variants.

Interestingly, the expression of the PMCA pumps is transcriptionally regulated by Ca^{2+} itself. The phenomenon has been studied on cultured neurons and is correlated to their maturation. In cerebellar granule cells, as Ca^{2+} concentration in the cytoplasm increases, the pattern of expression of the PMCA pumps undergoes a dramatic rearrangement: isoforms 2 and 3 become strongly upregulated, whereas isoform 1 undergoes a splicing switch which generates high levels of the truncated *a* variant. By contrast, PMCA4 disappears in a process that is calcineurin dependent. The reprogramming of the transcription of PMCA is essential to the survival of the

developing neurons: the regulation of PMCA expression may thus be critical to the survival of cells exposed to pathological increases in intracellular Ca^{2+} . Parallel amplification of Ca^{2+} influx and efflux pathways may enable differentiated neurons to precisely localize Ca^{2+} signals in time and space. Similarly, in rat hippocampal neurons, transcripts for all isoforms of the PMCA pump have been found to be strongly upregulated during the second week in culture, the overall increase in Ca^{2+} extrusion systems being accompanied by changes in the expression and cellular localization of different isoforms. The accumulation of PMCA in dendrites and dendritic spines coincided with the functional maturation of these neurons, underlining the importance of the proper spatial organization of the Ca^{2+} extrusion systems in synaptic function and development.

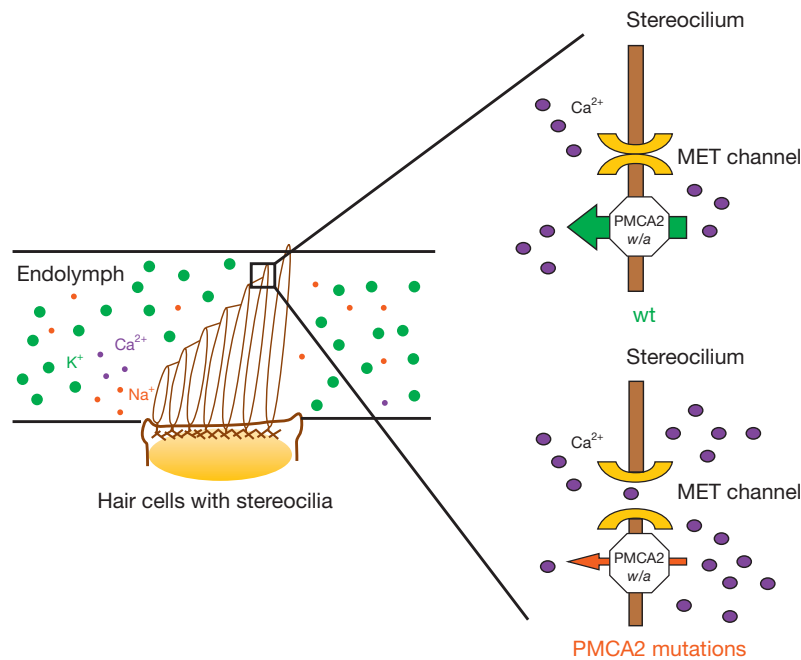


Figure 3 Schematic model of the hair bundle. The plasma membrane Ca^{2+} pump and the mechanotransduction channel (MET) are indicated. PMCA2 mutations reduce the ability of the PMCA to extrude Ca^{2+} , decreasing the endolymphatic Ca^{2+} concentration, thus increasing the open probability of MET channels.

Role in Physiology and Pathology

PMCA pumps are expressed ubiquitously; thus, they are unlikely candidates as molecules that could mediate cell-specific processes. However, the existence of so many PMCA variants (basic isoforms and alternative spliced forms), the finding on knockout mice for the specific PMCA isoforms, and studies on the mechanisms modulating their cell/tissue-specific distribution indicate that PMCA pumps may play important roles as signaling molecules, in addition to having a constitutive role as Ca^{2+} -housekeeping enzymes.

The ablation of isoforms 2 and 4 in mice has indeed revealed that they are important in specialized biological processes. PMCA2 plays an essential role in the hearing process, since it dissipates with peculiar kinetics the Ca^{2+} transients generated by the opening of the mechanoelectrical transduction (MET) channels in the stereocilia of cochlear hair cells (Figure 3). Mutations in the PMCA2 pumps in hereditary deafness impair the longer-term export of Ca^{2+} from hair cells, but not their ability to respond to a sudden increase of Ca^{2+} . This suggests that the mutated pumps reduce the extracellular Ca^{2+} gradient in the endolymph surrounding the hair cells, compromising the adaptation process responsible for the control of the opening of MET channels. Interestingly, the PMCA2 isoform has another peculiar function in the mammary gland: its activity is responsible for the high content of Ca^{2+} in the milk. PMCA2 is strongly upregulated in lactating mammary gland, and PMCA2 null mice have 40% less Ca^{2+} in the milk.

PMCA4 is crucial to male fertility: in mice, its ablation reduces sperm motility, probably resulting from Ca^{2+} overload and mitochondrial damage. Interestingly, PMCA4 is also essential for the modulation of the activity of the calcium/calmodulin-dependent neuronal nitric oxide synthase (nNOS)

which controls nitric oxide production, which is important in the regulation of excitation–contraction coupling of the heart. It has been shown that PMCA4b also regulates cardiac contractility *in vivo* through its interaction with nNOS.

PMCA defects have been described in a large number of disease conditions, among them those linked to the oxidative stress in several tissues, particularly brain, in brain ischemia, diabetes, atherosclerosis, aging, neurodegenerative diseases, and various disturbances of Ca^{2+} metabolism. PMCA defects have also been reported in various cancer types, in line with accumulating evidence on the involvement of Ca^{2+} homeostasis dysfunction in the process of malignant transformation. The genetic diseases involving the PMCA pump are best characterized: a number of dysfunctions linked to the ablation, or to partial disruptions, including point mutations, of its genes, have been described in mice. So far, the only identified spontaneous human disease related to a PMCA pump (isoform 2) defect is the form of hereditary deafness described above.

Genetic Manipulations of PMCA and Sodium Calcium Exchanger (NCX)

Knockout mice have been developed and the phenotypes studied for PMCA1, PMCA2, and PMCA4, but not yet for PMCA3. As this isoform is widely expressed in developing embryos, it may be essential for normal gestation.

The ablation of the PMCA1 gene has resulted in early embryonic lethality in homozygotes, underlying the essential role of PMCA1 in the early stages of development and in organogenesis, in line with its suggested housekeeping function. Heterozygotes, by contrast, appeared healthy; however, loss of one copy of the PMCA1 gene in mice with a PMCA4 null

background led to increased propensity to apoptosis in the smooth muscles of blood vessels, possibly due to Ca^{2+} overload. The ablation of the PMCA2 gene revealed its involvement in the hearing process: the phenotype of the PMCA2 knockout mice was characterized by balance impairment and hearing loss. The study of the vestibular inner ear revealed the absence of the otoconia and the progressive degeneration of the hair cells beginning 10 days after birth. The PMCA2 pump is probably also important to general neuronal Ca^{2+} homeostasis since PMCA2 null mice showed a significant reduction in the number of spinal cord motor neurons and abnormalities in Purkinje neurons. However, the balance defect was apparently not due, as could in principle have been expected, or was only partially due, to a cerebellar dysfunction. It was apparently related to the absence of the Ca-carbonate crystals of the otoconia embedded in the otolithic membrane overlying the sensory epithelium that sense gravity and linear acceleration in the inner ear. In keeping with the high level of expression of PMCA2 in the mammary gland, the ablation of its gene, in addition to generating the hearing loss phenotype, also strongly reduced the concentration of Ca^{2+} in the milk.

The ablation of the PMCA4 gene failed to cause a very evident general pathological phenotype, but local defects were present. This is interesting, as PMCA4 is also widely expressed, and has also been proposed to play a housekeeping role. However, it has now become evident that PMCA4 plays more specialized roles, and is not only essential for the general function of controlling Ca^{2+} homeostasis in all cells. One prominent defect caused by PMCA4 dysfunction was male infertility, reflecting the dominance of the PMCA4 pump in the testis, where it represents more than 90% of the total PMCA protein. The ablation of the PMCA4 gene produced other localized dysfunctions, for instance, in the modulation of the Ca^{2+} signals in B-lymphocytes and in the contractility of vascular or bladder smooth muscles.

Important physiological functions of the PMCA4 isoform were revealed by studies on mice overexpressing it rather than by ablating it. Early studies in which the PMCA4 pump had been expressed specifically in the myocardium of the rat, and in vascular smooth muscle cells in mice, had indicated a role in the modulation of myocardial growth and hypertrophy. They have also indicated a role in the regulation of the peripheral vascular tone, as the mice displayed increased peripheral blood pressure.

Concluding Remarks

To respond dynamically to the changing needs of Ca^{2+} signaling, cells must be able to precisely control the type, amount, localization, and activation of Ca^{2+} transporters. This control is performed in two ways: through the variable expression of Ca^{2+} transporters with different biochemical characteristics and through the fine modulation of their function by expressing differently active isoforms in response to the local or temporal cell demands. The cell also modulates the activity of its transporters by choosing specific molecular partners: in the

case of the PMCA pump, the interaction with different proteins contributes to regulate the activity of the pump such as in the case of calmodulin or 14-3-3 protein, but also to generate a special micro-environment where Ca^{2+} concentration can be finely modulated to optimize the activity of other enzymes, such as in the case of the cardiac NOS. Thus, the PMCA pump not only contributes to maintain the general Ca^{2+} homeostasis but also operates as an essential actor in specific signal transduction pathways.

Acknowledgments

The original work described in this review by the author has been supported over the years by grants from the Italian Ministry of University and Research (PRIN 2003, 2005, and 2008), the Telethon Foundation (Project GGP04169), and the Italian National Research Council (CNR).

See also: [Bioenergetics: Calcium Signaling: Calmodulin-Dependent Phosphatase](#); [Cell Architecture and Function: Caspases and Cell Death; Tubulin and Its Isoforms](#); [Protein/Enzyme Structure Function and Degradation: AAA-ATPases](#); [Signaling: Calcium/Calmodulin-Dependent Protein Kinases](#).

Further Reading

- Brini M (2009) Plasma membrane Ca^{2+} -ATPase: From a housekeeping function to a versatile signaling role. *Pflügers Archiv – European Journal of Physiology* 457: 657–664.
- Brini M and Carafoli E (2009) Calcium pumps in health and disease. *Physiological Reviews* 89: 1341–1378.
- Carafoli E, Genazzani A, and Guerini D (1999) Calcium controls the transcription of its own transporters and channels in developing neurons. *Biochemical and Biophysical Research Communications* 266(3): 624–632.
- Cartwright EJ, Oeandry D, and Neynes L (2009) Physiological implications of the interaction between the plasma membrane calcium pump and nNOS. *Pflügers Archiv* 457: 665–671.
- Ficarella R, Di Leva F, Bortolozzi M, et al. (2007) A functional study of plasma-membrane calcium-pump isoform 2 mutants causing digenic deafness. *Proceedings of the National Academy of Sciences of the United States of America* 104: 1516–1521.
- Monteith GR, McAndrew D, Faddy HM, and Roberts-Thomson SJ (2007) Calcium and cancer: Targeting Ca^{2+} transport. *Nature Reviews Cancer* 7: 519–530.
- Monteith GR and Roufogalis BD (1995) The plasma membrane calcium pump – a physiological perspective on its regulation. *Cell Calcium* 18(6): 459–470.
- Prasad V, Okunade GW, Miller ML, and Shull GE (2004) Phenotypes of SERCA and PMCA knockout mice. *Biochemical and Biophysical Research Communications* 322: 1192–1203.
- Schultz JM, Yang Y, Caride AJ, et al. (2005) Modification of human hearing loss by plasma-membrane calcium pump PMCA2. *New England Journal of Medicine* 352: 1557–1564.
- Street VA, McKee-Johnson JW, Fonseca RC, Tempel BL, and Noben-Trauth K (1998) Mutations in a plasma membrane Ca^{2+} -ATPase gene cause deafness in deafwaddler mice. *Nature Genetics* 19: 390–394.
- Strehler EE and Zacharias DA (2001) Role of alternative splicing in generating isoform diversity among plasma membrane calcium pumps. *Physiological Reviews* 81: 21–50.
- Toyoshima C, Nakasako M, Nomura H, and Ogawa H (2000) Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* 405: 647–655.

Plastocyanin

E L Gross, Ohio State University, Columbus, OH, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 390–393, © 2004, Elsevier Inc.

Glossary

Cytochrome *b₆f* complex Large multisubunit complex located in the thylakoid membrane that mediates electron transport between plastoquinone and plastocyanin (PC). Cyt *f*, a member of this complex, donates electrons directly to PC.

P700 Chlorophyll dimer in the reaction center of PSI which is oxidized by light and reduced by PC.

Photosystem I Large multisubunit complex located in the thylakoid membrane which accepts electrons

from PC and donates them to ferredoxin. It contains the primary electron donor P700 and acceptors A₀, A₁, and Fx.

PsaA, PsaB, and PsaF Protein subunits of photosystem I (PSI).

Thylakoid Portion of the inner membrane system of the chloroplast where the light reactions take place and where the cytochrome *b₆f* complex and PSI are located.

The photosynthetic electron transport chain of higher plants, algae, and cyanobacteria consists of two light reactions (photosystem II (PSII) and photosystem I (PSI)), which act in series to produce molecular oxygen, reduce nicotinamide adenine dinucleotide phosphate (NADP⁺) to nicotinamide adenine dinucleotide phosphate-oxidase (NADPH), and synthesize adenosine triphosphate (ATP). The cytochrome *b₆f* complex is a large multisubunit complex involved in the transfer of electrons from PSII to PSI. Plastocyanin is an ~10-kDa Cu protein which acts as a mobile electron carrier, located in the lumen of the chloroplast thylakoid, which shuttles electron from cytochrome *f* (Cyt *f*) in the cytochrome *b₆f* complex to P700 in PSI. It can be replaced by an Fe-containing cytochrome (cytochrome *c₆*) in some green algae and cyanobacteria under conditions of copper deficiency.

Plastocyanin Structure

The Copper Center

Plastocyanin (PC) is a β -sheet protein (Figure 1) with a Cu atom that acts as the redox center. The Cu atom is coordinated to four amino acid residues (H87, H37, S84, and M92) in a distorted tetrahedral geometry. The distorted tetrahedral geometry of the Cu center together with the protein matrix promotes efficient electron transfer between Cyt *f* and P700. Electrons enter and leave PC via the H87 ligand to the copper, which is located on the surface of the PC molecule (Figures 1 and 2).

The Surface of the PC Molecule

Higher plant and green algal PCs

Figure 2(a) depicts the surface of spinach PC which is a typical representative of higher plant and green algal PCs. There are two important features. The first is the patch of hydrophobic residues (green in Figures 2(a) and 2(b)) that surround the H87 ligand to the copper. These residues – including G10, L12, G34, F35, P36, L62, P86, G89, and A90 – are either conserved or replaced with similar residues in all PCs. They provide a

hydrophobic binding surface for both Cyt *f* and PSI (see below). The second feature consists of two patches of negatively charged residues surrounding Y83 (yellow in Figures 2(a) and 2(b)). Negatively charged residues are shown in red and positively charged ones are shown in blue in Figure 2. The lower patch consists of anionic residues #42–44 and either #45 or #79. The upper patch consists of anionic residues #59–61 in most higher plant PCs. In a few higher plant species and in all green algae, one of the anionic residues in the upper patch is deleted and is replaced by a glutamate at position #85. Also, all higher plant and green algal PCs contain an additional anionic residue at either position #51 or #53. Thus, there are a total of eight anionic residues that contribute to large negatively charged electrostatic field surrounding Y83.

Cyanobacterial PCs

Cyanobacterial PCs are similar to those from higher plants and green algae in that a patch of hydrophobic residues surrounds the H87 ligand to the Cu (Figure 2(b)) (residues from cyanobacterial PCs are numbered so as to be consistent with those from higher plant PCs). However, the distribution of charged residues on the surface of cyanobacterial PCs is quite different from that observed for higher plants and green algae. The only conserved anionic residue is D42. Instead, there are a series of positively charged residues on PC including R88, K44, and K51 that contribute to a positively charged electrostatic field surrounding Y83.

The Interaction of PC with Cyt *f*

The Structure of Cyt *f*

Cyt *f* is also a β -sheet protein with two domains. The heme is located on the large domain. The sixth ligand to the heme is the amino group of Y1 which may provide a pathway of electron transfer from the heme to H87 on PC. Y1 and the heme are surrounded by a patch of highly conserved hydrophobic residues. Cyt *f*s from higher plants and green algae have a highly conserved basic patch consisting of five residues (K58, K65,

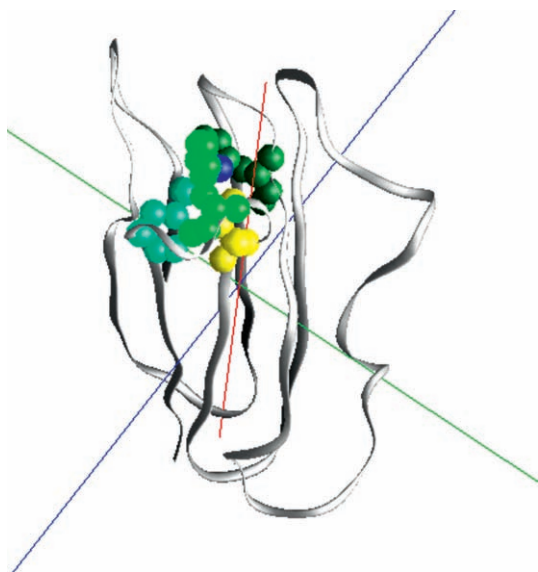


Figure 1 The structure of spinach plastocyanin: the copper center. The copper atom in PC (dark blue) is coordinated to two histidines (H87: light green; H37: dark green), C84 (yellow) and M92 (cyan). PC structures are taken from the Protein Data Bank from Berman HM, Westboro J, Feng Z, et al. (2000) The Protein Data Bank. *Nucleic Acids Research* 28: 235–242, and displayed using the Program GRASP from Nicholls A and Honig B (1991) A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. *Journal of Computational Chemistry* 12: 435–445.

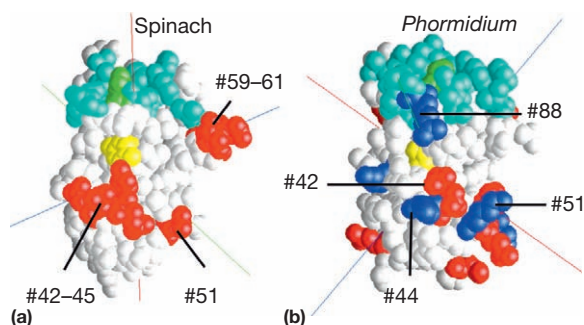


Figure 2 Surface residues on PC. H87, green; Y83, yellow; hydrophobic, cyan; cationic, blue; anionic, red: (a) spinach PC and (b) cyanobacterial PC from *P. lamosum*. Other conditions are the same as for Figure 1.

K66, K187, and R209 in turnip Cyt *f*). The negatively charged residues on higher plant and algal PCs are attracted to the positive patch on Cyt *f*. By contrast, Cyt *f* from the cyanobacterium *Phormidium lamosum* has a large number of negatively charged residues on its surface that attract the positively charged residues on cyanobacterial PCs.

Complex Formation between PC and Cyt *f*

Figure 3 depicts a complex between spinach PC and turnip Cyt *f* determined by nuclear magnetic resonance (NMR). Two types of interactions are observed. First, there is an electrostatic attraction between cationic residues on Cyt *f* (blue) and the anionic residues surrounding Y83 on PC (red). These

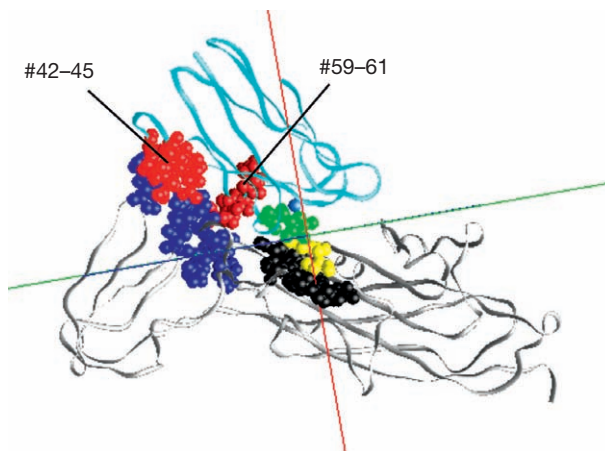


Figure 3 Complex formed between turnip Cyt *f* and spinach PC. Backbone of Cyt *f*, gray; backbone of PC, cyan. Other colors are the same as for Figure 2. Model 1 taken from Ubbink M, Ejdeback M, Karlsson BG, and Bendall DS (1998) The structure of the complex of plastocyanin and cytochrome *f*, determined by paramagnetic NMR and restrained rigid-body molecular dynamics. *Structure* 6: 323–335.

electrostatic interactions serve to bring the two molecules together and to orientate them properly. Evidence for electrostatic interactions comes from cross-linking, chemical modification, and mutagenesis experiments. Second, hydrophobic residues surrounding the heme and Y1 on Cyt *f* interact with those surrounding the H87 ligand and the Cu on PC. These hydrophobic interactions anchor the two proteins in the proper orientation for efficient electron transfer from the heme and Y1 on Cyt *f* to H87 and the Cu on PC. Complexes formed between *Phormidium* Cyt *f* and *Phormidium* PC show a greater influence of hydrophobic as opposed to electrostatic interactions.

The Interaction of PC with PSI

The Structure of PSI

Recently, the structure of PSI from *Synechococcus elongatus* has been determined at a resolution of 2.5 Å by X-ray crystallography. A simplified model of PSI is presented in Figure 4. Two regions of PSI have been implicated in PC binding. The first region includes the two largest subunits, PsaA and PsaB. These two subunits form a heterodimer containing P700 as well as the early electron acceptors A_0 , A_1 , and F_x . The second region includes the small subunit PsaF.

The Involvement of the PsaA and PsaB Subunits in Hydrophobic Interactions with PC

The PsaA–PsaB heterodimer has a 10-Å depression which has been postulated to form a binding pocket for PC and cyt c_6 (Figure 4). Two surface-exposed hydrophobic helices line the binding pocket. One is part of PsaA and extends from residue #646 to residue #664; the second is located on PsaB and extends from residue #Y621 to residues #N639. Also, W655 on PsaA and W631 on PsaB are located in these helices and point toward the binding pocket. It has been postulated that these residues may be involved in the transfer of electrons

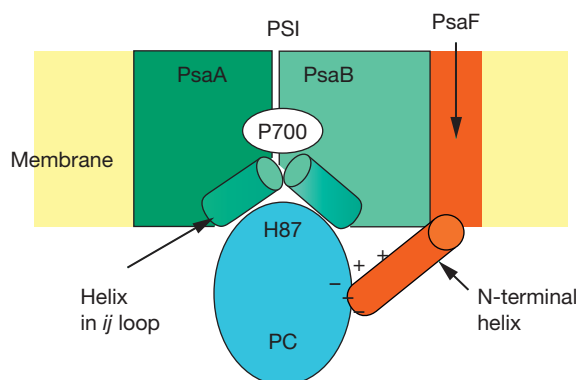


Figure 4 Model of the interaction of PC with PSI in higher plants and algae. The interactions of PC (blue) with (1) the PsaA and PsaB subunits of PSI (green) and (2) the N-terminal positively charged helix of PsaF (red) are shown. The hydrophobic helices in the *ij* loops of PsaA and PsaB that have been postulated to be part of the binding pocket are also shown.

from PC to P700. The observation that two mutants of residues located in the surface-exposed PsaB helix (W627F and E613N) caused significant inhibition of electron transport from PC to P700 support the hypothesis that these helices participate in the binding of PC to PSI.

Involvement of the PsaF Subunit in Electrostatic Interaction between PC and PSI in Higher Plants and Green Algae

The PsaF subunit in green plants and algae has been implicated in PC binding to PSI (Figure 4). The evidence for this is as follows. First, removal of PsaF by either detergent treatment or mutagenesis inhibits the fast phase of electron transfer from PC to P700. This rapid phase has a half-time of ~11 ps and represents electron transfer from PC to P700 in preformed PC–PSI complexes. Second, mutagenesis of residues #42–45 in the lower patch of PC inhibited the fast phase of electron transfer. By contrast, mutagenesis of the upper patch (residues #59–61) had very little effect. The N-terminal portion of PsaF consists of an amphipathic helix in which six lysine residues are lined up on one side of the helix. Third, mutagenesis of the lysine residues in this region of PsaF inhibited cross-linking and electron transfer from PC to P700; the greatest effect was observed for K23Q. Fourth, PC can be chemically cross-linked to PsaF. It was found that residues from the lower acidic patch of PC (#42–45) were cross-linked to residues #10–24 on PsaF, and residues from the upper acidic patch (#59–61) were cross-linked to residues #24–51 on PsaF.

The situation is different in cyanobacteria. Cyanobacteria show no rapid phase of electron transfer from PC to P700. In fact, the rate of electron transfer in cyanobacteria is ~2 orders of magnitude lower than that observed in green

plants or algae. Also, removal of PsaF has no effect. An examination of PsaF in cyanobacteria shows that it lacks the N-terminal lysine-rich helix found in higher plants and green algae. However, when a chimeric PsaF, made up of the N-terminal sequence from *Chlamydomonas* PsaF and the C-terminal sequence from the cyanobacterium *Synechococcus*, was placed in *Synechococcus* cells, rapid electron transfer was observed between *Chlamydomonas* PC (and cyt *c₆*) and P700 which had the same characteristics as observed in *Chlamydomonas* cells. However, there is some evidence for electrostatic effects in cyanobacteria as removal of the negative charge of D42 on a cyanobacterial PC increased the rate of electron transfer to P700, whereas removal of the positive charge of R88 decreases it.

In conclusion, binding of PC to PSI in green plants and algae is similar to its binding to Cyt *f* in that electrostatic forces steer PC (and cyt *c₆*) to the hydrophobic binding sites on PSI (Figure 4) where electron transfer to P700 occurs in a rapid and efficient manner.

See also: **Bioenergetics:** Cytochrome *c*; *F_x*, *F_A*, and *F_B* Iron–Sulfur Clusters in Type I Photosynthetic Reaction Centers; Photosystem I; Structure–Function of the Cytochrome *b₆f* Complex of Oxygenic Photosynthesis.

Further Reading

- Berman HM, Westbrook J, Feng Z, et al. (2000) The Protein Data Bank. *Nucleic Acids Research* 28: 235–242.
- Fromme P, Jordan P, and Krauss N (2001) Structure of photosystem I. *Biochimica et Biophysica Acta* 1507: 5–31.
- Gross EL (1996) Plastocyanin: Structure, location, diffusion, and electron transfer mechanisms. In: Ort D and Yocum C (eds.) *Oxygenic Photosynthesis: The Light Reactions*, pp. 413–429. Dordrecht: Kluwer Academic Publishers.
- Hippler M, Drepper F, Roach J-D, and Muhlenhoff U (1999) Insertion of the N-terminal part of PsaF from *Chlamydomonas reinhardtii* into photosystem I from *Synechococcus elongatus* enables efficient binding of algal plastocyanin and cytochrome *c₆*. *Journal of Biological Chemistry* 274: 4180–4188.
- Hippler M, Reichert J, Sutter M, et al. (1996) The plastocyanin binding domain of photosystem I. *EMBO Journal* 23: 6376–6384.
- Hope AB (2000) Electron transfers amongst cytochrome *f*, plastocyanin and photosystem I. Kinetics and mechanisms. *Biochimica et Biophysica Acta* 1456: 5–26.
- Kannt A, Young S, and Bendall DS (1996) The role of acidic residues of plastocyanin in its interaction with cytochrome *f*. *Biochimica et Biophysica Acta* 1277: 115–126.
- Martinez SE, Huang D, Szczepaniak A, Cramer WA, and Smith JL (1994) Crystal structure of the chloroplast cytochrome *f* reveals a novel cytochrome fold and unexpected heme ligation. *Structure* 2: 95–105.
- Nicholls A and Honig B (1991) A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. *Journal of Computational Chemistry* 12: 435–445.
- Sigfridsson K (1998) Plastocyanin, an electron-transfer protein. *Photosynthesis Research* 57: 1–28.
- Ubbink M, Ejdebäck M, Karlsson BG, and Bendall DS (1998) The structure of the complex of plastocyanin and cytochrome *f*, determined by paramagnetic NMR and restrained rigid-body molecular dynamics. *Structure* 6: 323–335.

Platelet-Activating Factor Receptor

K M Howard, University of Nevada Las Vegas School of Dental Medicine, Las Vegas, NV, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Katherine M. Howard and Merle S. Olson, volume 3, pp. 394–398, © 2004, Elsevier Inc.

Glossary

Autocoid An autopharmacologic substance that has potent local activity.

G-protein-coupled receptor A class of receptors for a diverse set of ligands including hormones, lipid inflammatory mediators, and chemokines that use associated trimeric G proteins for intracellular signal transduction.

Lipopolysaccharide A component of the cell wall of Gram-negative bacteria containing both lipid and

carbohydrate moieties that instigates many innate immune responses such as secretion of cytokines, activation of macrophages, and expression of leukocyte adhesion molecules. Synonymous with endotoxin.

Platelet-activating factor (PAF) A lipid autocoid.

Zymosan An insoluble protein–carbohydrate complex from the cell wall of yeast.

The platelet-activating factor (PAF) receptor is a member of the G-protein-coupled seven-transmembrane receptor superfamily. PAF receptor stimulation and the subsequent intracellular signaling events are involved in both physiological and pathophysiological processes but mainly participate in transcellular communication during inflammatory events. PAF receptor activation is instigated by the binding of PAF (1-*O*-alkyl-2-acetyl-*sn*-glycero-3-phosphocholine), a potent lipid mediator, which produces biological responses at very low molar concentrations (10^{-10} M).

Platelet-Activating Factor

Although PAF was first identified in the 1970s as a soluble mediator responsible for platelet aggregation when IgE-sensitized basophils were challenged with antigen, researchers have since established that PAF has diverse and potent biological activities. In addition to platelet aggregation, PAF can activate monocytes/macrophages, induce hypotension and bronchoconstriction, stimulate glycogenolysis, and participate in ovulation and ovoidimplantation. In addition, PAF has a profound role in acute inflammation, allergic disorders, and endotoxic and anaphylactic shock. Unlike other cytokines and inflammatory mediators, PAF is not synthesized and stored in a preformed state; rather, it is rapidly synthesized in response to specific stimuli. Cells that participate in inflammatory reactions including monocytes/macrophages, polymorphonuclear neutrophils, eosinophils, basophils, and platelets all rapidly produce PAF upon stimulation. Most cells that produce PAF also express PAF receptors. Therefore, PAF can operate in an autocrine, paracrine, or juxtacrine fashion. For instance, stimulated monocyte/macrophages release a majority of the PAF synthesized into the fluid phase and this PAF can interact with PAF receptors on its own cell surface (autocrine) or can bind to PAF receptors on other inflammatory cells in the immediate vicinity (paracrine). In the case of endothelial cells and leukocytes, synthesized PAF remains partially cell associated. For endothelial cells, this PAF is

presented on the cell surface for the juxtacrine activation of infiltrating leukocytes. The biological responses resulting from PAF receptor activation occur as a consequence of direct interactions of the mediator with receptors on target cells or through indirect secondary signaling effects.

In 1979, Hanahan and colleagues defined the chemical structure of this novel ether phospholipid, and the chemical structure of the lipid is presented in [Figure 1](#). The PAF molecule contains several specific structural features which are necessary for its potent biological activity. PAF contains an *O*-alkyl ether residue at the *sn*-1 position of the glycerol backbone. The ether-linked side chain can vary in chain length and degree of saturation but the 16:0 moiety is predominant. The 18:0-, 17:0-, and 18:1-containing molecules also occur naturally, but these molecules exhibit several orders of magnitude-less biological activity. There is an acetyl moiety at the *sn*-2 position of PAF and removal of this acetate group by the enzyme PAF acetylhydrolase generates lysoPAF which is biologically inactive. In addition to the strict structural requirements at the *sn*-1 and *sn*-2 positions, PAF requires a polar head group containing choline at the *sn*-3 position for specific PAF binding and functional activation of the PAF receptor. Both PAF synthetic and degradative enzymes recognize these unique structural features of PAF.

PAF is synthesized from 1-*O*-alkyl containing glycerolphosphocholines enriched with arachidonic acid at the *sn*-2 position. There are two distinct synthetic routes for PAF production, a remodeling and a *de novo* pathway. The *de novo* pathway produces low levels of PAF and has been attributed to the formation of constitutive levels of PAF for normal physiological actions. The remodeling pathway is the predominant route of PAF production by cells and tissues in response to systemic or local injury and trauma. Such stimulants for PAF generation can be microbial agents such as lipopolysaccharide (LPS) and zymosan or biological agents such as cytokines, chemotactic peptides, thrombin, and calcium ionophores.

Stimulating the remodeling pathway generates PAF in a two-step process. PAF synthesis begins with the activation of

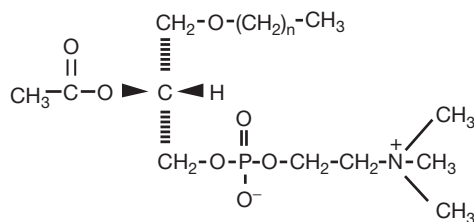


Figure 1 Chemical structure of PAF.

cytoplasmic phospholipase A_2 (cPLA $_2$) which preferentially cleaves the arachidonate at the *sn*-2 position. Therefore, the reaction releases arachidonic acid during the formation of lyso-PAF. The arachidonic acid is converted to eicosanoids via the lipoxygenase or cyclooxygenase pathways. Production of PAF occurs concomitant with the production of the eicosanoids which also contains far-reaching biological activities. The essential role of cPLA $_2$ in the synthesis of PAF has been confirmed from cPLA $_2$ knockout animals. In the second step of PAF generation, lysoPAF is converted to PAF by the actions of acetyl-CoA-lysoPAF acetyltransferase (LysoPAFAT). After nearly two decades of effort, the LysoPAFAT was recently cloned and characterized. This bifunctional enzyme also contains lysophosphatidylcholine acyltransferase activity and, therefore, participates in not only PAF synthesis but also membrane phospholipid synthesis. Inflammatory stimuli switches the activity of the enzyme from predominately membrane biogenesis to PAF synthesis.

Although the synthesis of PAF is tightly controlled, PAF-like species capable of binding and activating the PAF receptor can arise from the oxidation of unsaturated phosphatidylcholine. These ligands of the PAF receptor arising from oxidative damage of membrane phospholipids contribute to dysfunctional PAF receptor signaling and to the pathophysiological events associated with numerous disease entities.

PAF and PAF-like phospholipids are rendered biologically inactive via the actions of specific Ca^{+2} -independent phospholipases A_2 called 'PAF acetylhydrolases'. Three unique PAF acetylhydrolases are known to exist: two intracellular and one secreted form. These acetylhydrolases play a crucial role in terminating inflammatory responses elicited by both PAF and PAF-like phospholipids. Elevated plasma levels of the secreted PAF acetylhydrolase (also called 'lipoprotein-associated phospholipase A_2 ; Lp-PLA $_2$ ') were observed in patients with coronary artery disease but the exact role of plasma PAF-acetylhydrolase in cardiovascular diseases and atherosclerosis remains controversial.

Many of the biological and pathophysiological actions of PAF were initially investigated using specific PAF receptor antagonists. These PAF receptor antagonists include compounds with structural resemblance to PAF such as CV-3988 (phospholipid analogs) and nonstructural antagonists such as WEB 2086 and BN 50739. Classical ligand-binding experiments demonstrated specific PAF binding in numerous tissues and cells, including platelets, neutrophils, macrophages, mononuclear leukocytes, eosinophils, Kupffer cells, epithelial cells, and endothelial cells.

The unique structural features of PAF, the existence of PAF receptor antagonists, and direct binding data utilizing radiolabeled PAF, as well as biochemical evidence for the diverse

activation of second-messenger signal cascades, all provided evidence for the existence of specific PAF receptors. Confirmation that PAF signaling occurs through receptor-mediated mechanisms was affirmed in 1991 with the molecular identification of the first receptor cloned for a lipid mediator by Shimizu and colleagues.

PAF Receptor Structure

Numerous attempts by investigators to isolate the PAF receptor proved unsuccessful. The eventual cloning of the PAF receptor was accomplished in an elegant strategy involving expression cloning of a guinea pig lung complementary DNA (cDNA) library in *Xenopus laevis* oocytes. The initial cDNA library, constructed from size-fractionated poly (A) RNA, was converted *in vitro* into cRNA using the phage DNA as a template. Pools of cRNA were introduced into oocytes and a PAF-induced electrophysiological detection of G-protein-coupled phosphatidylinositol (PI) turnover was assayed. A single PAF-receptor cDNA sequence isolated by the cloning strategy predicted a protein of 342 amino acids and a calculated molecular mass of ~ 39 kDa. However, immunological detection of the PAF receptor on Western blots revealed an apparent MW of 69 kDa suggesting the presence of posttranslational modifications of the receptor. In fact, two potential N-linked glycosylation sites were identified in the PAF receptor cDNA sequence. Hydrophathy plot analyses of the deduced PAF receptor amino acid sequence predicted the presence of seven transmembrane-spanning domains within the PAF receptor. A schematic representation of the PAF receptor is presented in **Figure 2**. Through site-directed mutagenesis studies, several key features of the PAF receptor have been determined. Similar to other G-protein-coupled receptors, the ligand-binding site is defined in three dimensions by amino acids from several of the transmembrane-spanning regions. Mutations in membrane-spanning segments II (N58A and D63A), III (N100A, T101A, and S104A), and VII (D289A) increased affinity of PAF binding while mutations in V (H188A), VI (H248A, H249A, and Q289A), and VII (Q276A and T278A) lowered the affinity of PAF for the receptor. The third intracellular loop of the receptor is coupled to PI turnover. The C-terminal cytoplasmic tail of the receptor containing a cluster of serine and threonine residues are candidate sites for phosphorylation by G-protein receptor kinases, and are critical for agonist-induced desensitization.

Subsequent to the isolation of the guinea pig cDNA, cDNA for numerous other species was isolated by cross hybridization. The PAF receptor maintains $\sim 74\%$ sequence identity between the various species identified to date. Because the PAF receptor messenger RNA (mRNA) is present in only subsets of cells, Northern blot hybridizations have limitations in detecting PAF receptor expression in some tissues. The relative tissue abundance among different species varies but, in general, lung, neutrophils, monocytes, macrophages, placenta, small intestine, heart, liver, kidney, brain, and spleen have detectable PAF receptor expression.

The isolation of human PAF receptor cDNAs identified two unique 5'-noncoding regions designated transcript 1 and transcript 2. The tissue- and cell-type distribution of these two

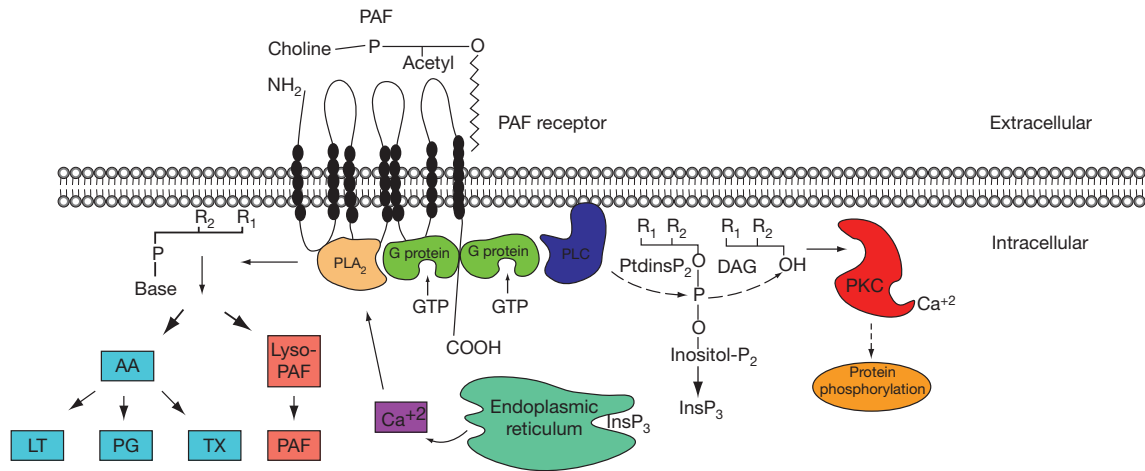


Figure 4 PAF receptor intracellular signaling pathways. See text for abbreviations.

is controlling the availability of PAF. This mechanism includes regulation of both PAF synthesis and release, and the rapid inactivation of PAF by PAF acetylhydrolases. Other mechanisms of regulation of this signaling system are intrinsic to the PAF receptor and include homologous desensitization to repeated or sustained administration of PAF. Desensitization of the PAF receptor leads to decreased PAF-induced cellular responses in guanosine triphosphatase (GTPase) activity, IP_3 production, and calcium mobilization. The carboxy terminus of the PAF receptor is the target for phosphorylation required to desensitize the receptor. PAF receptor desensitization may occur by uncoupling of the receptor to G proteins and by ligand-induced internalization of the receptor.

As opposed to the fairly rapid process of controlling PAF availability and PAF receptor responsiveness, long-term regulation of the PAF receptor includes alterations in PAF receptor expression. Transient elevation of cyclic adenosine monophosphate (cAMP) reduces expression of the PAF receptor. This downregulation is accompanied by a decreased responsiveness to a prolonged exposure of PAF. Conversely, priming of macrophages with low levels of LPS can induce a threefold increase in PAF receptor transcript levels, thereby amplifying subsequent response to stimulation.

PAF Receptor Signal Transduction

The wide range of biological responses instigated by PAF receptor activation can be attributed to complicated intracellular signaling mechanisms. Calcium, cAMP, IP_3 , and diacylglycerol (DAG) are all proven second messengers for PAF signaling. **Figure 4** presents an overview of PAF receptor-mediated signaling pathways. Heterotrimeric G proteins are intricately involved in the signal transduction events generated by PAF receptor occupancy. G proteins bind guanine nucleotides, are activated by guanosine triphosphate (GTP), and possess intrinsic GTPase activity. Binding of PAF activates the associated G protein by exchanging GTP for guanosine diphosphate (GDP). The PAF receptor is linked to more than one G protein because some processes are sensitive to pertussis toxin treatment, while

others are resistant. In fact, this G-protein utilization also can differ in different cell types. The activated G proteins, in turn, activate polyphosphoinositide turnover mediated by phospholipase C (PLC). The subsequent generation of inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol mobilizes intracellular Ca^{+2} and activates protein kinase C (PKC), respectively. PAF receptor-associated G-protein activation also leads to the activation of cytoplasmic PLA_2 with the subsequent cleavage of arachidonic acid (AA) and the formation of leukotrienes (LT), prostaglandins (PG), and thromboxanes (TX). More recently, investigators demonstrated broader activation of kinases and phospholipases following PAF receptor activation. For instance, PAF receptor activation has been shown to activate mitogen-activated protein kinase (MAPK), PI 3-kinases (PI3K), protein tyrosine kinases, and G-protein receptor kinase.

Experiments in both PAF receptor knockout animals and transgenic animals overexpressing PAF receptors support the pathophysiological role attributed to PAF receptor signaling. Deletion of PAF receptors in mice drastically attenuates systemic anaphylactic responses after antigen challenge, whereas mice overexpressing PAF receptors exhibit hypersensitivity to bacterial endotoxin. It is not surprising then that PAF receptor activation is being investigated as a contributing factor in numerous disease entities including asthma, atherosclerosis, inflammatory bowel disease, ischemia-reperfusion injury, endotoxemia, and allergic disorders.

See also: **Signaling:** Eicosanoid Receptors; G_{12}/G_{13} Family; G_i Family of Heterotrimeric G Proteins; G-Protein-Coupled Receptor Kinases and Arrestins; G_q Family; G_s Family of Heterotrimeric G Proteins.

Further Reading

- Chao W and Olson MS (1993) Platelet-activating factor: Receptors and signal transduction. *Biochemical Journal* 292: 617–629.
 Honda Z, Ishii S, and Shimizu T (2002) Platelet-activating factor receptor. *Journal of Biochemistry* 131: 773–779.

- Honda Z, Nakamura M, and Miki I (1991) Cloning by functional expression of platelet-activating factor receptor from guinea-pig lung. *Nature* 349: 342–346.
- Ishii S, Kuwaki T, Nagase T, et al. (1998) Impaired anaphylactic responses with intact sensitivity to endotoxin in mice lacking a platelet-activating factor receptor. *Journal of Experimental Medicine* 187: 1779–1788.
- Ishii S, Nagase T, and Shimizu T (2002) Platelet-activating factor receptor. *Prostaglandins and Other Lipid Mediators* 68–69: 599–609.
- Ishii S and Shimizu T (2000) Platelet-activating factor (PAF) receptor and genetically engineered PAF receptor mutant mice. *Progress in Lipid Research* 39: 41–82.
- Izumi T and Shimizu T (1995) Platelet-activating factor receptor: Gene expression and signal transduction. *Biochimica et Biophysica Acta* 1259: 317–333.
- Nagase T, Ishii S, Katayama H, et al. (1997) Airway responsiveness in transgenic mice overexpressing platelet-activating factor receptor – roles of thromboxanes and leukotrienes. *American Journal of Respiratory and Critical Care Medicine* 156: 1621–1627.
- Prescott SM, Zimmerman GA, Stafforini DM, and McIntyre TM (2000) Platelet-activating factor and related lipid mediators. *Annual Review of Biochemistry* 69: 419–445.
- Shindou H, Hishikawa D, Nakanishi H, et al. (2007) A single enzyme catalyzes both platelet-activating factor production and membrane biogenesis of inflammatory cells. *Journal of Biological Chemistry* 282: 6532–6539.
- Stafforini DM, McIntyre TM, Zimmerman GA, and Prescott SM (2003) Platelet-activating factor, a pleiotrophic mediator of physiological and pathological processes. *Critical Reviews in Clinical Laboratory Sciences* 40: 643–672.

Platelet-Derived Growth Factor Receptor Family

M V Kovalenko and A Kazlauskas, Harvard Medical School, Boston, MA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 399–406, © 2004, Elsevier Inc.

Glossary

Adaptor A protein linking together proteins that are otherwise unable to interact.

Autocrine signaling A type of signaling in which the cell responds to a stimulus (e.g., growth factor) that is produced by the same cell.

Cell cycle A tightly regulated sequence of events exhibited by dividing cells. It consists of the following phases: Gap 0 (G0, temporary or permanent quiescence); Gap 1 (G1, growth and preparation of the chromosomes for replication); S (synthesis of DNA); Gap 2 (G2, growth and preparation for mitosis); and M (mitosis).

Chemotaxis The ability of cells to migrate toward a gradient of chemoattractant.

Growth factor A substance that needs to be present in the environment of responsive cells in order for them to divide.

Ligand A molecule (or ion) that can interact with a receptor.

Mitogen A substance (e.g., growth factor) that can initiate cell division.

Paracrine signaling A type of signaling in which the cell responds to a stimulus (e.g., growth factor) that is produced by a nearby cell.

Pericytes The pluripotent cells surrounding blood vessels.

Receptor A molecule (mostly of protein nature) able to detect the presence of a specific stimulus (ligand) and give rise to signaling events leading to physiological response(s).

Transphosphorylation A phosphorylation event in which two protein kinases phosphorylate each other (e.g., within the dimer).

Platelet-derived growth factor (PDGF) is a potent mitogen, chemoattractant, and survival factor for cells of mesenchymal origin (such as fibroblasts, smooth muscle cells (SMCs), or glial cells). PDGF exists as a number of isoforms that initiate signaling via two closely related receptor tyrosine kinases (RTKs) named α - and β -receptors. PDGFs and their receptors (PDGFRs) are important for embryonic and postnatal development. Knockout mice lacking either PDGFs or receptors develop severe defects in various tissues and organs, and die before or shortly after birth. In adult organisms, PDGFs participate in wound healing, regulation of blood vessel tone, and maintenance of the interstitial fluid pressure. PDGF signaling is also involved in the pathogenesis of various proliferative diseases (certain tumors, atherosclerosis and restenosis, and fibrotic conditions). PDGFs and their receptors have been best characterized in mammals. However, there is evidence that PDGFR-like proteins are expressed in lower vertebrates, and PDGF-like growth factors have been found in molluscs, annelids, sea urchins (*Lytechinus pictus*), and *Drosophila*. Taken together, these findings indicate that PDGFs and their receptors play an important role in a variety of biological processes throughout eukaryotic species.

PDGF Isoforms

At present, four genes encoding different PDGF chains are known: A, B, C, and D. Biologically active PDGFs exist as disulfide-bonded homodimers designated AA, BB, CC, and DD. A and B chains form a heterodimeric PDGF AB (Figure 1). The history of the discovery of PDGFs dates back as early as 1974, when mitogenic activity of whole blood serum was linked to the presence of platelets. PDGF AB was the first to

be purified and biochemically characterized a few years later, followed by PDGF BB and AA. Cloning of PDGF A and B cDNAs and determination of the structure of corresponding genes were completed in the 1980s.

New members of the PDGF family, PDGF C and PDGF D, were found only recently by searching the database of human-expressed sequences. PDGFs C and D have a two-domain structure with N-terminal CUB domain and C-terminal PDGF/VEGF (core) domain, separated by a hinge region (Figure 1). No heterodimers involving C or D chains have been detected. The unique feature of these two new PDGFs is the requirement for proteolytic cleavage of the CUB domain upon secretion in order to achieve biological activity. Thus, latency may be the reason why these growth factors were not originally detected by functional assays.

PDGF Receptors

PDGFs exert their biological functions by binding to two isoforms of PDGF receptors, α and β , with different degrees of affinity. Both receptors are transmembrane tyrosine kinases, composed of extracellular, transmembrane, and intracellular parts. The extracellular part consists of five immunoglobulin-like domains that are involved in ligand binding (domains I–III) and receptor dimerization (domain IV). The intracellular kinase domain of the PDGF receptors is split into two by ~100-amino acid insert (Figure 1).

PDGF A chain binds specifically to α -receptor, and PDGF B can bind to both α and β . PDGF C was originally described as a ligand for PDGF α -, but not β -receptor; however, it was later shown to bind PDGF β -receptor in cells expressing both α - and β -isoforms. PDGF D has been reported to be a β -receptor

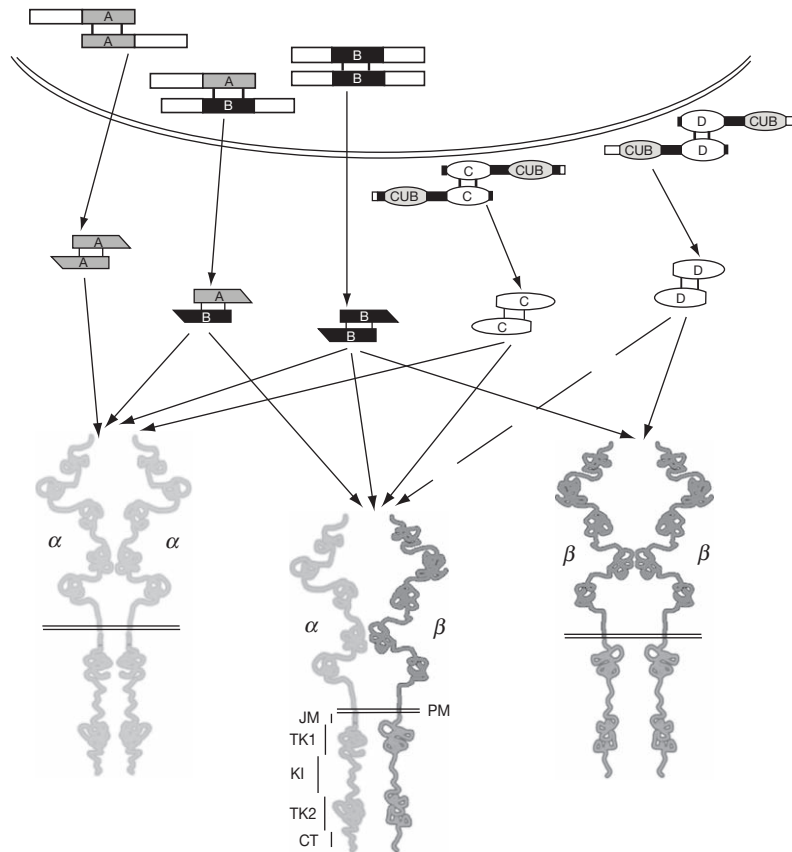


Figure 1 Platelet-derived growth factors and their receptors. PDGFs A and B are synthesized as precursors that are proteolytically processed before secretion. PDGFs C and D are secreted in latent form and activated by proteolytic cleavage of the CUB domain. All PDGF isoforms dimerize prior to the proteolytic cleavage. Specificity of binding of mature active PDGFs to their receptors is shown; dashed arrow indicates the lack of consensus regarding the ability of PDGF DD to cause the formation of a heterodimeric receptor. Domain structure of the PDGF receptor α -subunit (identical to that of β -subunit) is presented. The extracellular part of the receptor consists of five Ig-like domains. PM, plasma membrane; JM, juxtamembrane domain; TK1 and TK2, proximal and distal parts of tyrosine kinase domain; KI, kinase insert; and CT, carboxyl-terminal tail.

ligand, but its ability to bind α -receptor in α/β -expressing cells remains controversial. Interactions of PDGFs with α - and β -receptors are summarized in **Figure 1**. A bivalent dimer, PDGF molecule binds to two receptor subunits causing them to dimerize. Upon dimerization, PDGF receptors become rapidly phosphorylated on multiple tyrosine residues. One of them (regulatory tyrosine) is located in the second part of the kinase domain and is important for receptor kinase activity (Tyr857 in β -receptor and, by homology, Tyr849 in α -receptor). Most of the other phosphorylation sites lie in noncatalytic parts of the receptor (**Figure 2**). The exact sequence of events during receptor activation and formation of signaling complex is unknown. By analogy to other RTKs, it is likely that ligand binding and subsequent dimerization of the receptor induces a conformational change that facilitates transphosphorylation of regulatory tyrosine residues within the dimer. Importantly, the regulatory tyrosine lies within activation loop, the region that is conserved among RTKs. As shown for the insulin receptor, the activation loop blocks the catalytic site in nonphosphorylated receptor, whereas upon ligand-induced phosphorylation it moves away providing access to the substrate. It is possible that the PDGF receptor activation occurs via a similar mechanism, although there is no direct evidence for this in the absence of the crystal structure of PDGF receptors.

Recently, ligand-bound dimerized PDGF β -receptor was found to be less susceptible to dephosphorylation by protein tyrosine phosphatases (PTPs) than the monomeric receptor. It is suggested that this phosphatase protection may contribute to maintenance of the activated (phosphorylated) state of the receptor.

Proteins Associated with the PDGF Receptors and PDGF-Driven Signaling Pathways

A total of 13 tyrosine residues in β -receptor and 11 in α -receptor get phosphorylated upon PDGF stimulation. Phosphorylation of most of the tyrosines within PDGF receptors results in the creation of docking sites for a variety of proteins, many of which in turn get phosphorylated upon association with the receptors (**Figure 2**). These proteins include enzymes (phosphatidylinositol 3'-kinase (PI3-kinase), phospholipase C- γ 1 (PLC γ 1), SH2 domain-containing protein tyrosine phosphatase-2 (SHP-2), GTPase-activating protein of Ras (RasGAP), Src family kinases (SFKs), transcription factors (members of STAT family), or adaptor proteins, such as Grb2, Grb7, Shc, Shb, Nck, or Crk, linking the receptor to signaling proteins further downstream. Some proteins that are recruited to the PDGF receptors are

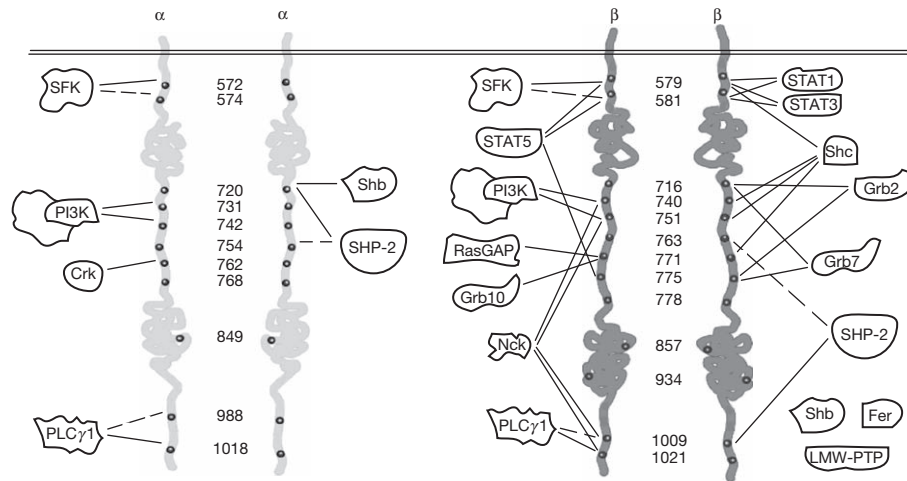


Figure 2 Intracellular domains of $\alpha\alpha$ and $\beta\beta$ homodimers of the PDGF receptor. Phosphorylated tyrosine residues are shown (dark circles). Numbers indicate their positions in the human PDGF receptor sequence. Association of signaling proteins with phosphorylation sites of the activated PDGF receptors is shown by solid lines (major binding sites) or dashed lines (additional sites). Proteins binding with high affinity (e.g., PI3-kinase, PLC γ 1, and RasGAP) bind to the receptor via one or two phosphotyrosine residues on the receptor, whereas low-affinity binding (e.g., Nck, Shc, or Shb binding to β -receptor) involves multiple sites. Shb was found to bind to most of the phosphorylated tyrosines on the PDGF β -receptor. Fer and LMW-PTP bind to as yet unidentified sites on the PDGF β -receptor.

both adaptors and enzymes, for example, tyrosine phosphatase SHP-2 is able to recruit Grb2 via its phosphorylated C terminus. Association is mediated in most cases by SH2 (Src homology 2) domains of these proteins and is remarkably specific, each phosphorylation site having its own binding partner(s). This specificity is based on the ability of SH2 domains to differentially recognize amino-acid sequence following the phosphorylated tyrosine. Many receptor-binding proteins also contain SH3 (Src homology 3), PTB (phosphotyrosine binding), or PH (pleckstrin homology) domains that may facilitate recruitment of other yet unidentified signaling molecules to the complex.

Cellular Responses Mediated by the PDGF Receptors

Assembly of the PDGF receptor complex initiates a number of signaling pathways leading to cellular responses (Figure 3): chemotaxis, proliferation, differentiation (in certain cell types), and protection from apoptosis. One pathway may lead to several different responses, whereas some of the pathways are redundant and converge on the same cellular effect.

Proliferation

Proliferation is the major and most well-studied cellular effect induced by PDGFs. Virtually all signaling proteins binding to the receptors have been implicated in proliferation, but not all of them have to be present in the signaling complex at the same time to induce the response.

Ras-dependent pathway

The mitogenic pathway that involves Ras, a small GTPase, can be activated through multiple ways in response to PDGF. The complex of Grb2 (adaptor protein) and Sos1 (nucleotide exchange factor for Ras) can bind to the β -receptor either directly or via phosphorylated Shc or SHP-2. This association

brings the complex to the plasma membrane where it activates Ras. Activated Ras GTPase interacts directly with Raf-1 serine/threonine kinase, in turn activating it. Raf-1 is the first kinase in the MAP kinase cascade. Elevation of its activity is followed by sequential activation of MEK and MAP kinases, Erk1 and 2, which are translocated to the nucleus and phosphorylate a number of transcription factors.

As for the PDGF α -receptor, Grb2 is able to bind to it via SHP-2, but there is no evidence of its involvement in Ras activation. Therefore, pathways other than the Ras-mediated one are likely to contribute to α -receptor-mediated proliferation.

The Ras pathway is believed to be counteracted by RasGAP that binds to the PDGF β -receptor (but not to α -receptor) and is able to enhance GTPase activity of Ras, thus converting it from active (GTP-bound) to inactive (GDP-bound) state. The extent to which this negative signaling affects proliferation may depend on many factors (cell type and environmental conditions).

Src

Src (p60^{c-src}) has been shown to be important for PDGF-driven mitogenesis. However, its association with the PDGF receptors and early PDGF-induced elevation of activity are dispensable for proliferative response and may have a different function. Yet, Src is required at multiple time points during PDGF-dependent cell-cycle progression. Thus, elevation of Src activity early in G1 leads to increase of *c-myc* expression, which is required for transition through G1. Importance of Src has also been confirmed for late G1 and for the exit from M-phase of the cell cycle.

PLC γ 1

PLC γ 1 binds to the PDGF receptors upon ligand stimulation and becomes phosphorylated and activated. Increasing the catalytic activity of PLC γ 1 leads to production of inositol 1,4,5-trisphosphate (IP3) and diacylglycerol (DAG). IP3 is involved in regulation of cytoplasmic Ca²⁺ levels, which,

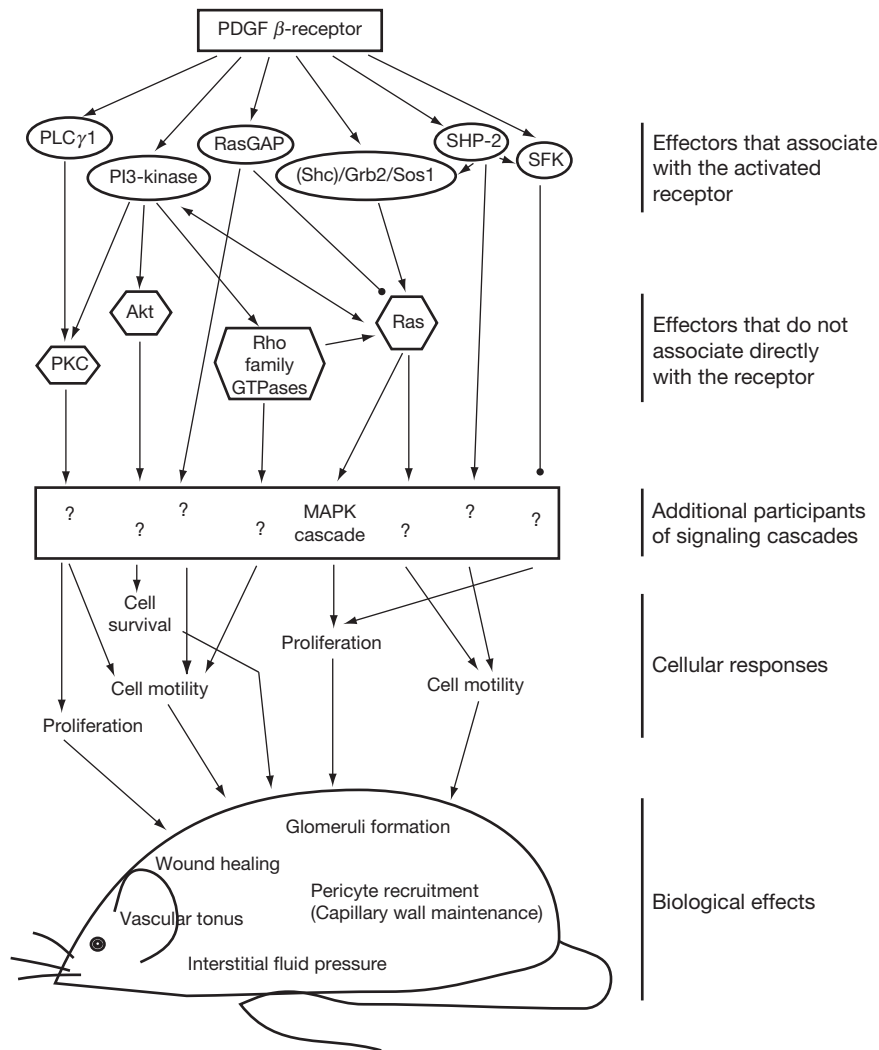


Figure 3 Major events initiated by the PDGF β -receptor and culminating in biological responses. Upon ligand binding, activated PDGF β -receptor recruits enzymes and adaptor proteins that start a number of distinct signaling pathways, some of which are shown on the figure. Note that not all of them are initiated by the same receptor dimer or in the same cell. Arrows represent either physical or functional interactions along the pathways. The least defined part of the signaling cascade is the one that bridges the receptor-proximal events with proteins that carry out cellular responses. PDGF is necessary but not sufficient to achieve biological effects. In a living organism, other growth factors and hormones, as well as cell–cell and cell–matrix interactions within tissues and organs make indispensable contributions.

together with direct action of DAG, can contribute to activation of certain members of the PKC family. In particular, DAG has been shown to activate protein kinase C- ζ (PKC- ζ), which then activates Raf-1 and, consequently, the rest of the MAP kinase cascade, leading to proliferation. PLC γ 1 is involved in PDGF-stimulated proliferation mediated by β -receptor, whereas in the case of α -receptor, it is important mostly for the motility response.

PI3-kinase

PDGF-dependent PI3-kinase activation is critical for mitogenesis in most cell types. Prolonged exposure to PDGF, needed to induce cell-cycle progression in quiescent cells, causes two distinct peaks of PI3-kinase activity, one within minutes and another following a delay of several hours. Only the second peak was found to be important for the cell-cycle progression,

whereas the early increase of activity is required only for immediate responses to PDGF, such as chemotaxis.

How this biphasic increase of PI3-kinase activity is linked to the cell-cycle machinery still remains an open question. Continuous treatment with PDGF can be substituted by two separate pulses of the growth factor, the first getting cells out of G0 to early G1 and the second pushing them through late G1 and into S phase. Early elevation of *c-myc* and sustained activation of the MAP kinase pathway can replace the first pulse of PDGF, but is not sufficient to drive the cell through late G1. To complete cell-cycle progression, a properly timed second peak of PI3-kinase activity (second pulse of PDGF or just the addition of PI3-kinase lipid products) is required. It is hypothesized that early *c-myc* and Erk activation triggers expression of new protein(s) whose interaction with the late PI3-kinase products is critical for S phase entry.

Chemotaxis

Most receptor-expressing cells exhibit chemotaxis upon exposure to PDGF. Migration of pericytes in response to SMC-released PDGF plays a critical role in capillary formation, which is particularly important during embryonic development. In the process of wound healing, PDGF causes chemotaxis of neutrophils and macrophages, as well as fibroblasts and SMCs.

α - and β -PDGF receptors differ in their ability to mediate chemotaxis. PDGF β -receptor homodimer as well as α - β heterodimer potently stimulate it. However, the ability of α - α homodimer to drive chemotaxis remains controversial and may depend on the cell type. This difference may be explained by variations in composition of the signaling complex assembled by different receptor dimers, which may be cell-type specific.

PI3-kinase and Ras pathways

Chemotaxis involves dynamic remodeling of cytoskeleton, including rearrangement of actin filaments and microtubules, changes of focal adhesions, and formation of lamellipodia or filopodia. These processes are controlled, in part, by the Rho family of small GTPases (RhoA, Rac1, and Cdc42), which are activated by the PDGF receptors via different pathways. The PI3-kinase pathway is considered to be the major one. Activation of the Rho family of small GTPases is dependent on PI3-kinase lipid products. The Ras pathway that can be initiated by the PDGF receptor independently from PI3-kinase pathway, leads to chemotactic response via activation of Rac1, another small GTPase Ral, and in some cell types, MEK1 and stress-activated kinase p38. Interestingly, PI3-kinase is able to interact directly with Ras and activate it, whereas both Ras and Rho family GTPases can activate PI3-kinase, forming a positive feedback loop. Thus, reciprocal regulation of Ras and PI3-kinase provides an example of interaction between different PDGF-dependent pathways.

RasGAP, a negative regulator of Ras, has been shown to inhibit chemotaxis. However, in some cell types, RasGAP has also been shown to have a positive impact on motility that is most likely independent from Ras.

PLC γ 1

PLC γ 1 promotes chemotaxis in many cell types, but the knowledge about what happens downstream of PLC γ 1 is only starting to emerge. Upon PDGF stimulation, PLC γ 1 activates sphingosine kinase through IP₃ formation and subsequent release of Ca²⁺ from intracellular stores. Sphingosine kinase product, sphingosine-1-phosphate (SPP), is a ligand for EDG-1, a G-protein-coupled receptor (GPCR). Stimulation of EDG-1 by SPP leads to activation of Rho family GTPases, which, in turn, results in remodeling of actin cytoskeleton and cell-matrix adhesion sites. This is a new example of receptor crosstalk, where an RTK requires activation of a GPCR to achieve a cellular response.

PKC isoforms that are activated via both PI3-kinase and PLC γ 1 pathways are also involved in chemotactic signaling, as chemotaxis is reduced when PKCs are inhibited or downregulated.

SHP-2 and low-molecular-weight phosphatase

Two other positive mediators of PDGF-induced chemotaxis are tyrosine phosphatases SHP-2 and low-molecular-weight

phosphatase (LMW-PTP), both directly associated with the phosphorylated PDGF receptor and activated by PDGF-dependent phosphorylation. Involvement of SHP-2 in chemotactic signaling may occur via the Ras pathway, since SHP-2 can serve as an adaptor protein for Ras binding. LMW-PTP is likely to act through dephosphorylating and inactivating one of its cellular targets, p190RhoGAP, a negative regulator of Rho GTPase, thus leading to increased activation of Rho and subsequent cytoskeletal rearrangements.

SFKs

The role of SFKs in PDGF-induced chemotaxis is not clear. A pathway leading to chemotaxis is initiated at SFK-binding sites in the juxtamembrane domain of the PDGF α -receptor, since mutating these tyrosines to phenylalanine residues results in chemotaxis inhibition. However, it is not yet clear whether it is SFKs or other signal transduction molecules binding to the same sites that are responsible for cell migration, since in triple SFK knockout cells, PDGF-AA-dependent cell migration is intact.

It is necessary to note that $\alpha\alpha$ - and $\beta\beta$ -homodimers of the PDGF receptor cause different, although partially overlapping, cellular responses, due to the differences in their ability to bind signaling proteins (Figure 2). For instance, α -receptor homodimer has been found to send a less potent mitogenic signal than the β -receptor homodimer. Additionally, in some cell types (SMC, human foreskin fibroblasts), it fails to mediate a chemotactic response to PDGF AA and inhibits chemotaxis induced by other agents. PDGF β -receptor homodimer formation generally results in higher mitogenicity and chemotaxis. The heterodimeric α - β receptor is believed to have unique signaling properties, due to the altered phosphorylation pattern of both α - and β -subunits, which results in recruitment of a different subset of signaling proteins. In particular, the heterodimer mediates a stronger mitogenic response than either α - α or β - β complexes do. Distinct signaling properties of different PDGF receptor dimers are based not only on different intrinsic properties of receptor isoforms, but also on the type of cells where these receptors are expressed and on varying experimental conditions.

Negative Regulation of the PDGFR Signaling

Down-regulation mechanisms described for the PDGFR include internalization, ubiquitin-mediated proteolysis (degradation), and dephosphorylation by PTPs. Their relative contribution to the termination of signaling appears to be different.

Internalization and Degradation

Internalization of the PDGF receptors, mediated by clathrin-coated pits, occurs shortly after ligand stimulation and is dependent on the receptor kinase activity. Intracellular trafficking of the PDGF β -receptor involves PI3-kinase, but is not entirely elucidated yet. Internalized receptors can be recycled and go back to the cell surface or can be directed to lysosomal degradation, which actually results in receptor downregulation. However, before they are degraded, internalized receptors remain associated with downstream proteins and continue to signal. For this reason, internalization and subsequent

trafficking can be considered a late signaling stage rather than merely the method of receptor inactivation.

Additionally, PDGFRs can be targeted for cytoplasmic degradation in proteasomes via polyubiquitination. Ubiquitination of the receptors occurs upon ligand binding and is strictly dependent on receptor autophosphorylation. Generally, multi-ubiquitin chain is added to proteins by E3–E2 ubiquitin ligase system and after that ubiquitinated proteins are recognized by the 26S proteasome and degraded. c-Cbl, a E3 ubiquitin ligase, has been shown to be phosphorylated upon PDGF stimulation and is therefore likely to contribute to the PDGFR ubiquitination and subsequent proteolysis. Proteolytic degradation can be considered a true mechanism of receptor downregulation that terminates signaling through decreasing the number of the PDGF receptor molecules.

Dephosphorylation by PTPs

PTP-dependent dephosphorylation of the PDGF receptors represents another level of regulation of PDGF-induced signaling. Thus, dephosphorylation of the regulatory site or nonspecific removal of phosphate from all tyrosine residues would shut down all signaling, whereas site-specific dephosphorylation would selectively block corresponding signaling pathways. LMW-PTP, for instance, has been shown to act both ways. This PTP selectively interferes with two pathways initiated by the PDGF β -receptor: Src-dependent induction of *c-myc* and STAT1/3-mediated *c-fos* expression. It can also affect kinase activity of the PDGF β -receptor by dephosphorylating Tyr857, providing a general negative regulation of all downstream signals.

Unlike LMW-PTP that acts on membrane-bound PDGF β -receptor, PTP1B was found to dephosphorylate the receptor upon its internalization on the surface of the endoplasmic reticulum, shutting down signaling at a later time point, yet before the receptor gets degraded or recycled to the plasma membrane.

The ability of SHP-2 to act as both positive and negative regulator of PDGFR signaling puts it apart from other receptor-directed PTPs. It is able to dephosphorylate selectively, tyrosines 771 and 751 on the β -receptor, potentially turning off RasGAP and PI3-kinase pathways. On the other hand, SHP-2 can serve as an adaptor for Grb2 binding, therefore initiating Ras-mediated signaling. Additionally, SHP-2 has been implicated as a positive regulator in the stimulatory effect of integrins on PDGF receptor signaling.

An increase of intracellular reactive oxygen species (ROS) is an early PDGF-dependent event mediated by PI3-kinase and Rac1, and it was shown to be important for mitogenic signaling. Since ROS reversibly inhibits PTPs, it is likely that this event helps to prevent the negative effect of PTPs during early stages of PDGF signaling.

Developmental Role of the PDGF Receptors

At very early stages of embryonic development (before gastrulation), both PDGF α -receptors and PDGF A chain are expressed throughout the embryo, suggesting that PDGF AA

may be acting in an autocrine manner to stimulate cell proliferation. At later stages, PDGF β -receptor and PDGF B chain appear, and both receptors and growth factor chains start to show a distinct expression pattern. The expression of PDGF C largely overlaps with that of PDGF A, and information about PDGF D expression in embryo is still incomplete. PDGF receptors are found mainly in cells of mesodermal origin, with some exceptions (e.g., central nervous system), whereas PDGFs are present in adjacent ecto- and endodermal derivatives. This distribution indicates that, during morphogenesis, PDGFs act in a paracrine manner to induce migration of receptor-expressing mesenchymal cells.

As shown by genetic analysis in mice, interruption of PDGF signaling leads to morphogenic defects and embryonic or perinatal lethality. PDGFR α -null mice die before birth, many (but not all) at a very early stage. Those available for analysis have myotomal abnormalities, rib fusions, defects in neural crest derivatives, and incomplete cephalic closure. PDGF A knockouts demonstrate an emphysema-like lung condition characterized by the absence of mature alveoli. This defect originates due to a lack of migration of PDGF α -receptor expressing alveolar SMC progenitors in the walls of alveolar sacs. Both PDGF A and PDGFR α -null mice also display hypoplasia of skin mesenchyme and abnormal intestinal villi formation. The most pronounced phenotypic features of the PDGF B and PDGF β -receptor knockout mice are defects in kidney glomeruli and capillary microaneurisms. Both result from failure of capillary branching and dilation of capillary walls in the absence of PDGF-dependent recruitment of pericytes and kidney mesangial cells. Other abnormalities include dilation of heart and large arteries, thrombocytopenia, and anemia.

See also: [Cell Architecture and Function: Rho GTPases and Actin Cytoskeleton Dynamics](#); [Signaling: Chemotactic Peptide/ Complement Receptors; Phospholipase C; Ras Family; Src Family of Protein Tyrosine Kinases](#).

Further Reading

- Betsholtz C, Karlsson L, and Lindahl P (2001) Developmental roles of platelet-derived growth factor. *BioEssays* 23: 494–507.
- Heldin CH, Östman A, and Rönnstrand L (1998) Signal transduction via platelet-derived growth factor receptors. *Biochimica et Biophysica Acta* 1378: F79–F113.
- Östman A and Heldin CH (2001) Involvement of platelet-derived growth factor in disease: Development of specific antagonists. *Advances in Cancer Research* 80: 1–38.
- Pietras K, Östman A, Sjöquist M, et al. (2001) Inhibition of platelet-derived growth factor receptors reduces interstitial hypertension and increases transcapillary transport in tumors. *Cancer Research* 61: 2929–2934.
- Rosenkranz S and Kazlauskas S (1999) Evidence for distinct signaling properties and biological responses induced by the PDGF receptor α and β subtypes. *Growth Factors* 16: 201–216.
- Smits A and Funa K (1998) Platelet-derived growth factor in primary brain tumours of neuroglial origin. *Histology and Histopathology* 13: 511–520.
- Takehara K (2000) Growth regulation of skin fibroblasts. *Journal of Dermatological Science* 24(supplement 1): S70–S77.

Porphyrin Metabolism

M Shepherd, University of Kent, Canterbury, UK

A E Medlock and H A Dailey, University of Georgia, Athens, GA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Mark Shepherd and Harry A. Dailey, volume 3, pp. 415–419, © 2004, Elsevier Inc.

Glossary

Chelation The removal of free metal ions from solution by an organic chemical.

Herbicide Chemical agent used to curtail plant growth.

Plastid Generic name for chloroplasts and chloroplast precursors.

Singlet oxygen A reactive oxygen radical with a single unpaired electron.

Introduction

Porphyrins are cyclic tetrapyrroles that perform critical roles in a variety of biological systems. Examples of important tetrapyrroles include heme, siroheme, and heme d₁ (all with chelated iron), chlorophyll (with chelated magnesium), coenzyme B₁₂ (which possesses only three meso carbons between pyrrole rings, is a corrin rather than a porphyrin, and contains chelated cobalt), and factor F₄₃₀ (with chelated nickel). Heme, as a prosthetic group in various hemoproteins, is necessary for oxygen transport and storage as part of hemoglobin and myoglobin. It is essential for electron transport as part of various cytochromes and is also required by mixed function oxidases as a part of cytochrome P₄₅₀. In addition, heme is needed for the decomposition and production of hydrogen peroxide as a cofactor of catalase and peroxidase, respectively. Recently, heme has been shown to play a regulatory role in several central cellular processes, including transcription, translation, and intermediary metabolism. Heme d₁ is a cofactor in dissimilatory nitrite reductase, while siroheme serves as the cofactor in both assimilatory nitrite and sulfite reductases. Neither heme d₁ nor siroheme is synthesized by mammals.

Several types of chlorophylls exist which differ with respect to the substituents on the tetrapyrrole ring. Chlorophylls are employed by photosynthetic organisms to absorb energy for photosynthesis. This process harnesses the vast majority of energy consumed by living organisms and only chemolithotrophic bacteria are independent of this energy source. Chlorophylls are required in the first stage of photosynthesis, whereby visible radiation excites an electron in the porphyrin that, upon returning to a lower energy state, provides the energy needed for the synthesis of adenosine triphosphate (ATP) and nicotinamide adenine dinucleotide phosphate (NADPH).

Coenzyme B₁₂ consists of cobalamin (vitamin B₁₂), a corrin ring, which is linked to dimethylbenzimidazole and 5' deoxyadenosine via its central cobalt atom. The bond between the cobalt and the latter substituent is a weak cobalt–carbon bond, the only example known in biology, which catalyzes intramolecular rearrangements by cleaving homolytically to generate a free radical. Examples include methylations in methionine synthesis and the reduction of ribonucleotides to deoxyribonucleotides. Finally, there is a nickel(II) tetrapyrrole, factor F₄₃₀, which is found in methanogenic bacteria and is involved in methane metabolism.

Biosynthesis of Porphyrins

Figure 1 illustrates the first three steps of tetrapyrrole biosynthesis which are common to all the tetrapyrroles mentioned above. All of these molecules have a core structure of four pyrrole rings, designated A–D, as shown on the structure of heme in Figure 1. The carbon atoms are numbered 1–20 starting in ring A and ascending in a clockwise direction around the perimeter of the porphyrin ring.

Formation of δ -Aminolevulinic Acid

The first common intermediate in tetrapyrrole biosynthesis is δ -aminolevulinic acid (ALA). Two pathways for its synthesis are found in nature: the C₄ pathway in animals, fungi, and some bacteria (α -group of proteobacteria) including *Rhodospirillum rubrum*, and the C₅ pathway in plants, algae, archaea, and most other bacteria including cyanobacteria. In the C₄, or Shemin pathway, ALA is formed by the condensation of glycine and succinyl coenzyme A (CoA) in a reaction catalyzed by the enzyme ALA synthase (ALAS). In the C₅ pathway an NADPH-dependent glutamyl-transfer RNA (tRNA) reductase produces glutamate 1-semialdehyde (GSA) from glutamyl-tRNA, which is subsequently converted to ALA by the replacement of the aldehyde moiety with an amide group. The C₅ pathway is thought to be ancestral to the Shemin pathway as bacteria with the former lack the α -ketoglutarate–dehydrogenase complex to make succinyl CoA. This group includes cyanobacteria such as *Synechocystis*, which may explain why plants today retain the C₅ pathway.

ALA to Uroporphyrinogen III

In all organisms studied to date, the conversion of ALA to uroporphyrinogen III proceeds via exactly the same steps. This process is summarized in Figure 1. The first step involves the condensation of two ALA molecules to form porphobilinogen with the release of two molecules of H₂O. This reaction is catalyzed by porphobilinogen synthase, also known as ALA dehydratase. Hydroxymethylbilane synthase, also known as porphobilinogen deaminase, catalyzes the condensation of four molecules of porphobilinogen in the order A, B, C, and D (the letters representing the rings of the ultimate tetrapyrrole product).

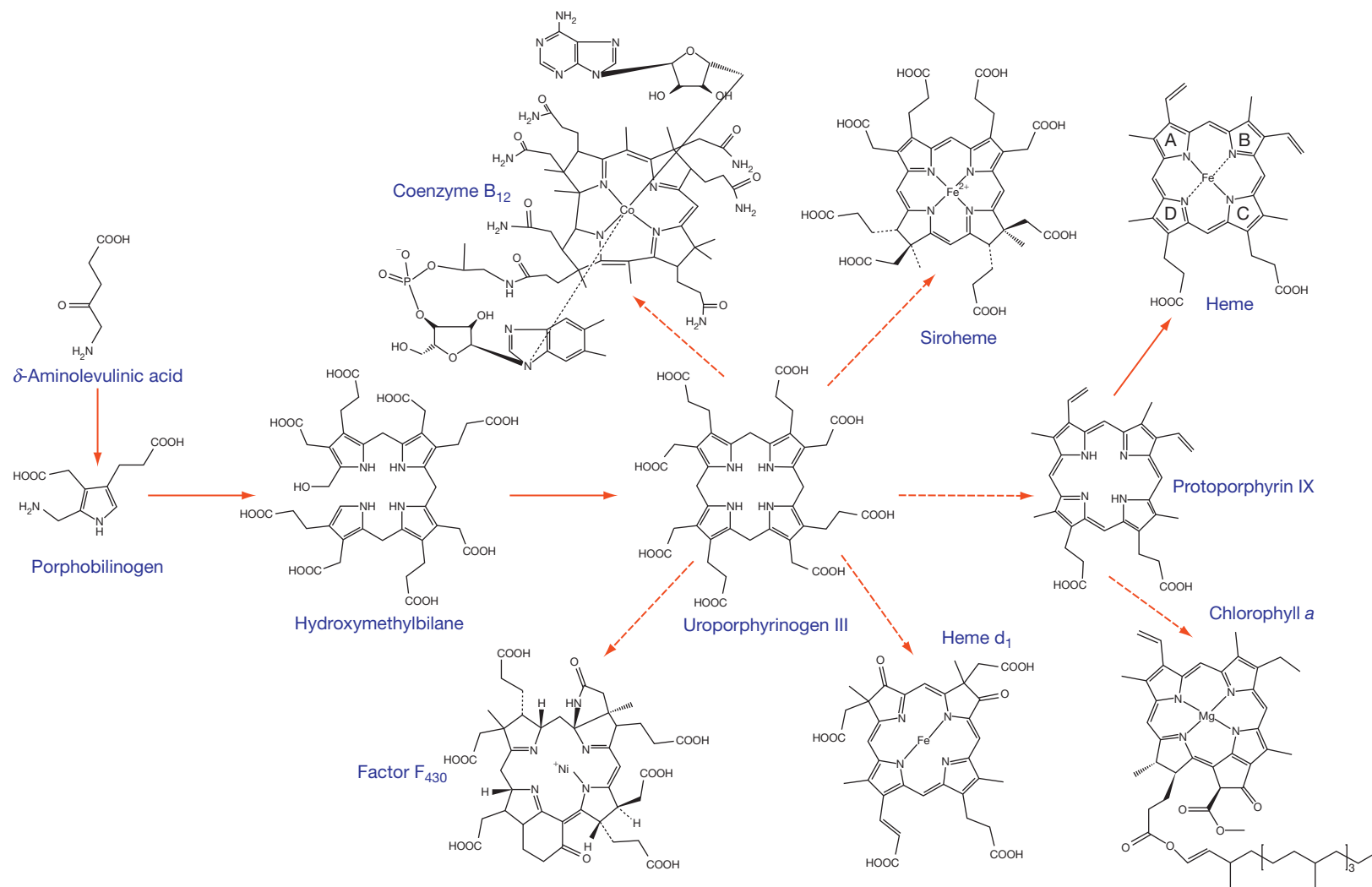


Figure 1 The porphyrins. The first three steps common to all porphyrin biosynthesis and the ultimate products of this branched metabolic pathway.

This condensation gives rise to the first linear tetrapyrrole, hydroxymethylbilane, also known as preuroporphyrinogen.

The enzyme hydroxymethylbilane synthase synthesizes its product by building on a dipyrromethane cofactor (a modified dipyrrole) which is posttranslationally attached to a cysteine residue of the enzyme. The first porphobilinogen molecule condenses with the cofactor and further porphobilinogen molecules are subsequently added in a head-to-tail fashion to form hydroxymethylbilane. The enzymatic formation of the first macrocyclic tetrapyrrole, uroporphyrinogen III, is catalyzed by the enzyme uroporphyrinogen III synthase. During this reaction, the D pyrrole ring of hydroxymethylbilane undergoes an inversion via a spiro intermediate. As free hydroxymethylbilane spontaneously reacts to form the biologically inactive isomer, uroporphyrinogen I, the enzyme must be present during or immediately after the release of hydroxymethylbilane from hydroxymethylbilane synthase to ensure that the correct isomer, uroporphyrinogen III, is formed.

Heme and Chlorophyll Biosynthesis

The biosynthesis of heme and chlorophyll proceeds along a common pathway until a branch point is reached at protoporphyrin IX (Figure 2). The formation of protoporphyrin IX from uroporphyrinogen III is achieved via a series of decarboxylation and oxidation steps. Uroporphyrinogen decarboxylase catalyzes the removal of a carboxyl group from each pyrrole ring to form coproporphyrinogen III. This is converted to protoporphyrinogen IX via the action of two enzymes, coproporphyrinogen oxidase and protoporphyrinogen oxidase. During aerobic growth both enzymes use oxygen as the terminal

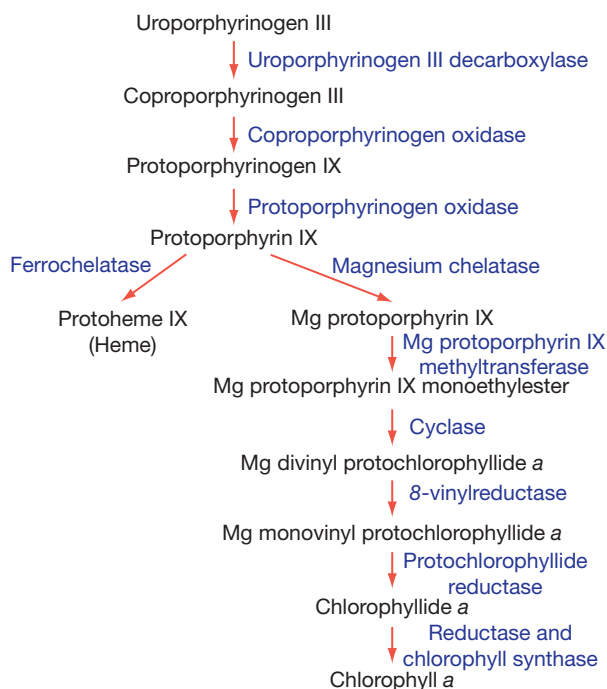


Figure 2 Heme and chlorophyll biosynthesis. The branching of the heme and chlorophyll biosynthetic pathways. The enzymes that catalyze the individual steps are shown in blue.

electron acceptor. However, anaerobic organisms have been shown to possess oxygen-independent isozymes and some facultative anaerobes encode both types of coproporphyrinogen oxidase. The pathway, as described above, is conserved for almost all organisms studied, except *Desulfobrio vulgaris* and *Methanosarcina barkeri*, in which protoheme is formed from uroporphyrinogen III indirectly via precorrin-2.

The next step involves the insertion of a divalent cation into the porphyrin ring. Ferrochelatase catalyzes the insertion of an iron(II) ion while magnesium chelatase, as its name suggests, catalyzes the insertion of a magnesium(II) ion. Ferrochelatase distorts the porphyrin ring slightly to allow for proton abstraction and iron(II) insertion. The insertion of magnesium requires the presence of MgATP and the participation of all three subunits of the heterotrimeric magnesium chelatase.

Subsequent steps of the chlorophyll biosynthetic pathway involve methylation and cyclization of the propionate group on ring C. After the reduction of the vinyl group on ring C, a light-dependent reaction takes place. This is catalyzed by protochlorophyllide reductase (POR) and is one of the only two reactions known to be light dependent, the other being catalyzed by DNA photolyase. This reaction can also proceed via a light-independent reaction, catalyzed by dark POR (DPOR). The formation of chlorophylls requires the addition of a hydrophobic tail to the propionate group on ring D via an esterification reaction. In the cases of chlorophyll and bacteriochlorophyll, these long-chain alcohols are phytol and geranylgeraniol, respectively.

Coenzyme B₁₂, Siroheme, Heme d₁, and Factor F₄₃₀ Biosynthesis

Figure 3 summarizes the synthesis of coenzyme B₁₂, siroheme, heme d₁, and factor F₄₃₀ from uroporphyrinogen III. The first step in synthesis of any of these porphyrins involves the methylation of uroporphyrinogen on rings A and B to form precorrin-2. Reduction to form sirohydrochlorin and subsequent ferrochelation yields siroheme, the prosthetic group for both nitrite and sulfite reductases. Factor F₄₃₀ synthesis requires the cyclization of substituents on rings B and D and the chelation of nickel(II), which is bound to the pyrrole nitrogen on ring C. This final step is thought to be catalyzed by an ATP-dependent nickel chelatase. The steps involved in the synthesis of heme d₁ are not yet known beyond the formation of precorrin-2.

The formation of corrin coenzyme B₁₂ involves a branched pathway consisting of several reactions. Figure 3 shows that the synthesis of cob(II)yrinate a,c-diamide can occur by either

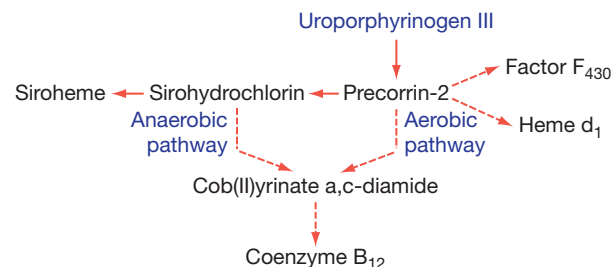


Figure 3 Coenzyme B₁₂, siroheme, heme d₁, and factor F₄₃₀ biosynthesis.

an anaerobic or an aerobic pathway. Both pathways involve the methylation, reduction, and isomerization of precorrin-2, along with the insertion of a central cobalt atom. The anaerobic chelatase has been compared to ferrochelatase as it is encoded by a single gene and does not require MgATP. The aerobic cobalt chelatase is similar to the magnesium chelatase involved in chlorophyll biosynthesis, as it is heterotrimeric and requires MgATP for catalysis. Subsequent formation of coenzyme B₁₂ proceeds via alkylations, phosphorylation, an internal rearrangement, and the coordination of an adenosyl group to the central cobalt atom.

Regulation

As δ -ALA is the universal precursor for porphyrin synthesis, the supply of this metabolic intermediate ultimately governs whether or not porphyrin synthesis may take place. However, there are several examples where other points in the pathway are subject to regulation. Heme plays a key role as a cofactor for many redox reactions in several different cell types and requires iron as a cofactor. From this, it is clear that the regulation of heme biosynthesis must be dealt with on a global scale among different cell types and subcellular compartments. The model of negative feedback regulation mediated by a single regulator (heme) at a single site (ALAS) remained dogma for many years. Only after compelling evidence was generated to support regulation at other sites was it accepted that heme biosynthesis is subject to more complex control.

In plants multiple tetrapyrroles are produced including heme, chlorophyll, and siroheme. Regulation of these cofactors along with the proteins that bind them is critical due to their reactive nature. Main control points in plants include the formation of δ -ALA, the metal chelation reactions, and steps in siroheme synthesis. The biosynthesis and regulation of heme d₁, cofactor B₁₂, and factor F₄₃₀ are more poorly understood. As mammals do not synthesize these cofactors, little work has been performed on the regulation of these pathways.

Regulation of Heme Biosynthesis

Role of ALAS

In animal cells two ALAS genes exist that are classified by their cellular localization. ALAS-E (erythroid; or ALAS-2) is induced during erythroid differentiation and ALAS-1 encodes ALAS in other tissues. ALAS-1 is also referred to as ALAS-H (housekeeping) or ALAS-N (nonerythroid). The primary sequence of both ALAS proteins is highly conserved. However, significant differences exist between the regulatory regions of the genes and the untranslated regions of their respective messenger RNAs (mRNAs), which are manifested in the tissue-specific expression of the two enzymes.

Regulation at sites other than ALAS

Although it is accepted that ALAS is the major site of regulation, several examples of pathway control exist at the level of gene expression. In mammals, all the genes that encode heme biosynthetic enzymes possess both housekeeping- and erythroid-specific promoter elements, although only ALAS is represented by two distinct genes.

Several examples of regulation of enzymes found later in the heme biosynthetic pathway have been identified. A good example of regulation in response to an external stimulus is exhibited by the coproporphyrinogen oxidase of the nitrogen-fixing bacterium *Azorhizobium caulinodans*. The enzyme responsible for nitrogen fixation, nitrogenase, requires the absence of oxygen to convert dinitrogen into ammonia. This organism is thought to possess both aerobic and anaerobic coproporphyrinogen oxidases and switches to the anaerobic isozyme in response to falling oxygen levels. Regulation of the antepenultimate enzyme in heme biosynthesis by the electron transport chain is seen in *Escherichia coli*. In this organism a respiratory electron transport chain component has been shown to act as the electron acceptor for protoporphyrinogen oxidase. This reaction is catalyzed by an enzyme distinct from the oxygen-dependent protoporphyrinogen oxidase.

Deficiencies in Porphyrin Synthesis

Porphyrias

Porphyrias are inherited diseases caused by point mutations or deletions in the genes encoding heme biosynthetic enzymes. Individuals with these disorders are usually able to synthesize normal amounts of heme, but accumulate pathway intermediates which cause the clinical symptoms of porphyria. These diseases can be divided into three groups: (1) cutaneous porphyrias, characterized by photosensitization; (2) acute porphyrias, where individuals experience abdominal pain, psychiatric disorders, constipation, vomiting, and paralysis; and (3) mixed porphyrias, which manifest themselves as a combination of both. Figure 4 illustrates the porphyrias that result from deficiencies in specific steps of the pathway.

Herbicides

The porphyrin biosynthetic pathway has been targeted for inhibition by certain herbicides to curtail the production of chlorophyll. Diphenyl ether-type herbicides, acifluorfen in particular, have been shown to act as potent inhibitors of protoporphyrinogen oxidase, the last enzyme before the branchpoint of heme and chlorophyll biosynthesis. These molecules compete for the tetrapyrrole-binding site on the enzyme, which impedes the binding of protoporphyrinogen. Hence, protoporphyrin production is inhibited, leading to reduced levels of heme and chlorophyll. As a result, plants treated with such herbicides appear chlorotic (yellow colored) due to their ability to produce carotenoids but not chlorophylls. The action of these herbicides is characterized by the accumulation of protoporphyrinogen IX. This diffuses out of the plastids/mitochondria and is oxidized to protoporphyrin IX via nonspecific plant peroxidases. This mislocation of protoporphyrin IX leads to phototoxicity through the evolution of singlet oxygen species.

Porphyryn Catabolism

Much of the work on porphyrin catabolism has focused on heme and chlorophyll degradation in both prokaryotes

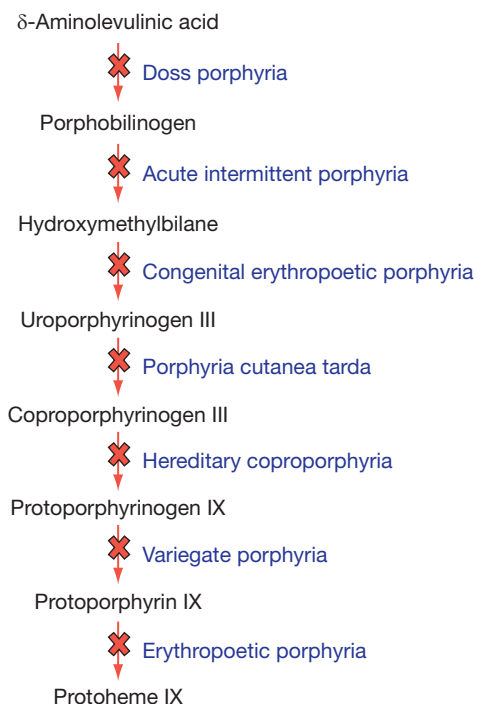


Figure 4 The porphyrias. The porphyric disorders which result from deficiencies in specific steps of the heme biosynthetic pathway are shown in blue.

and eukaryotes. While the role of chlorophyll catabolism appears mainly for the recovery of Mg and macrocycle degradation, heme catabolism serves multiple purposes. Many pathogenic prokaryotic organisms sequester exogenous heme and catabolize this as a source of iron to fulfill their metabolic needs. In plants and cyanobacteria, the heme macrocycle is cut open releasing the iron and the linear tetrapyrrole is used to make the light-harvesting pigments phycobilins and phytychromobilins.

In the metazoans heme catabolism serves not only to degrade heme, but also to form compounds that serve other purposes. The main source of heme for catabolism in higher animals, including man, is senescent erythrocytes that liberate large amounts of hemoglobin during hemolysis. Catabolism of heme is needed for iron recycling, as well as for the degradation of a potentially harmful molecule. In nonplacental animals heme is catabolized to the water-soluble compound biliverdin, but in placental animals, biliverdin is converted to the poorly soluble bilirubin. Bilirubin in mammals is glycosylated, thereby making it soluble. This conjugated bilirubin is then excreted via the kidney. Inability to excrete bilirubin may result in the disease states called hyperbilirubinemias. In premature neonates which lack the glycosylating enzyme this may result in kernicterus. Three inherited disorders of bilirubin excretion are Gilberts syndrome, and Crigler–Najjar syndrome type I and II.

The catabolic reactions are catalyzed by the action of two enzymes: heme oxygenase and biliverdin IX α reductase. The products of these reactions are the physiologically important compounds bilirubin, which at an appropriate concentration and body location can serve as an effective antioxidant

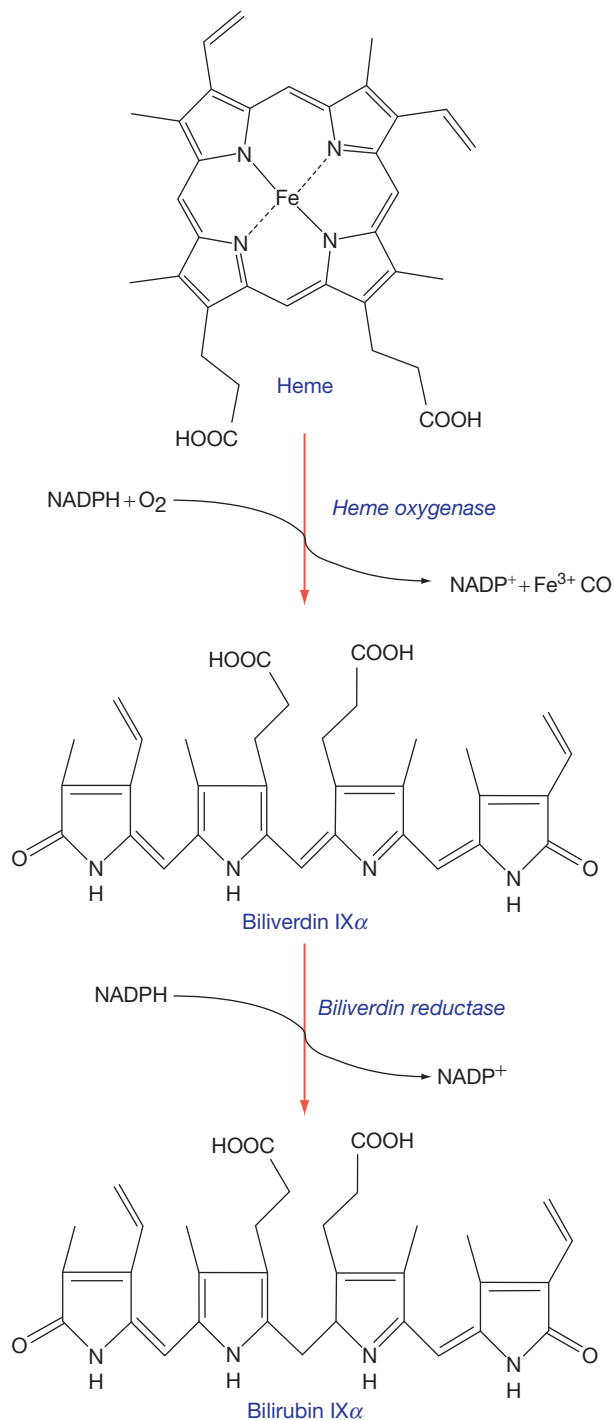


Figure 5 Heme catabolism. The conversion of heme to bilirubin. The enzymes that catalyze each step are shown in italics in blue.

and carbon monoxide, a neuronal messenger. Heme catabolism is the only reaction in mammals that produces carbon monoxide. **Figure 5** shows the degradation of heme to bilirubin IX α . The catabolites of heme degradation biliverdin IX α and bilirubin IX α are colored compounds, green and yellow, respectively, accounting for the color progression observed in a healing hematoma.

Interestingly, chlorophyll catabolism, for the most part, produces colorless and nonfluorescent products. Similar to what is seen in heme degradation by heme oxygenase, the key step in chlorophyll degradation appears to be an oxygenolytic cleavage of the porphyrin macrocycle between rings A and B (see [Figure 1](#)). The enzyme responsible for this reaction has yet to be identified as having the role(s) of any of the catabolites. Little is known in terms of the degradation of other porphyrins.

See also: **Bioenergetics:** Chlorophylls and Carotenoids; Cytochrome P-450; Heme Synthesis; **Metabolism Vitamins and Hormones:** Folate and Vitamin B₁₂ Function; **Protein/Enzyme Structure Function and Degradation:** B₁₂-Containing Enzymes; Heme Proteins.

Further Reading

- Beale SI (1999) Enzymes of chlorophyll biosynthesis. *Photosynthesis Research* 60: 43–73.
- Cox TM, Jack N, Lofthouse S, et al. (2005) King George III and porphyria: An elemental hypothesis and investigation. *Lancet* 366: 332–335.
- Dailey HA (1997) Enzymes of heme biosynthesis. *Journal of Biological Inorganic Chemistry* 2: 411–417.
- Smith K (1975) *Porphyrins and Metalloporphyrins*. Amsterdam: Elsevier.
- Tripathy BC, Sherameti I, and Oelmüller R (2010) Siroheme: An essential component for life on earth. *Plant Signaling and Behavior* 5: 14–20.
- Warren MJ and Smith AG (eds.) (2009) *Tetrapyrroles: Birth, Life and Death*. Austin: Landes Bioscience; Springer.

Potassium Channels in the Inner Membrane of Mitochondria in Various Organisms: From Unicellular Eukaryotes to Higher Plants and Mammals

I Szabó and M Zoratti, University of Padua, Padua, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Conductance (G) The ratio between the electric current and the applied voltage. The conductance is expressed in siemens ($S = \text{ampere/volt (A/V)}$).

Electrochemical proton gradient An electrochemical gradient has two components: (1) the electrical component is caused by a charge difference across the lipid membrane ($\Delta\Psi$) and (2) a chemical component is caused by a differential concentration of ions across the membrane (ΔpH). The combination of these two factors determines the thermodynamically favorable direction for an ion movement across a membrane. It is commonly expressed in

millivolts (mV) and in mitochondria its value corresponds to $-180/-200$ mV (negative on matrix side).

Electrophysiology Comprises various methodologies to study the electrical properties of biological membranes in organelles, cells, and tissues. It involves measurements of voltage change or electric current on a wide variety of scales from single ion channel proteins to whole organs.

Mitoplasts Swollen mitochondria with only remnants of the outer membrane are present and the mitoplasts thus become suitable for electrophysiological patch clamp experiments.

Mitochondria are bioenergetic adenosine triphosphate (ATP)-producing organelles of endosymbiotic origin, formed by an outer membrane, permeable to metabolites and ions, an inner membrane (IMM), where oxidative phosphorylation takes place, an intermembrane (periplasmic) space, and the matrix, enclosed by the IMM, where important metabolic processes such as the citric acid cycle and fatty acid beta-oxidation occur. The mitochondrial inner membrane has long been considered to be poorly permeable to cations and anions, since the strict control of inner mitochondrial membrane permeability is crucial for efficient ATP synthesis. Over the past 30 years, however, it has become clear that various ion channels are present in the IMM along with well-characterized carriers (antiporters and uniporters). Ion channels can transport up to 10^7 ions per second, and the electrochemical driving forces for ion movement across the IMM are of considerable magnitude, given the very high negative membrane potential across this membrane (in energized mitochondria $\Delta\Psi_m$ is around -180 mV) created by the respiration-driven efflux of protons from the matrix during oxidative phosphorylation. Channel proteins therefore are expected to be present in extremely low abundance and/or to be open for very short times to maintain the low permeability to ions required to exploit the proton motive force for ATP generation. Nonetheless, these channels are crucial to the mechanism of energy supply and are a decisive factor in determining whether a cell lives or dies. In fact, increased mitochondrial membrane permeability due to activation of mitochondrial megachannel/permeability transition pore (PTP) is a hallmark of cell death. Mitochondrial ion channels for Ca^{2+} , K^+ , or anions have been functionally and pharmacologically characterized at the single-channel level, not only in isolated mitochondria but also in intact mitochondria, in intact cells, and in some cases at whole-organ level.

Potassium fluxes (the K^+ cycle) can be monitored by classical bioenergetic techniques. These include mitochondrial swelling assays (since net uptake of K^+ salts will be

accompanied by osmotically obligated water, resulting in matrix swelling), measurements of changes in mitochondrial redox potential, respiration, and transmembrane potential, and direct determination of K^+ uptake into mitochondria using potassium-selective electrodes. Furthermore, pathways allowing potassium flux can be revealed directly by patch-clamp measurements on swollen mitochondria (mitoplasts) conserving only remnants of the outer membrane or by electrophysiological planar lipid bilayer recordings of channels in proteoliposomes containing mitochondrial membrane proteins.

The Mitochondrial Potassium Cycle

By using these methods, it was established that electrogenic proton ejection by the respiratory chain generates a high negative electrical membrane potential, which drives electrophoretic K^+ influx through potassium-selective ion channels. The ones that have been identified so far are listed below and are shown in [Figure 1](#). K^+ is the major monovalent cation in the cytoplasm and is consequently an important ion mediating changes in volume when the K^+ conductance of the inner membrane is altered. As mentioned above, net uptake of K^+ salts will be accompanied by water, resulting in matrix swelling. Excess matrix K^+ is then ejected by the electroneutral K^+/H^+ antiporter which is regulated to sense volume changes and uses the energy stored in the proton gradient to expel K^+ . A continuous electrophoretic influx of cations followed by their release via an electroneutral mechanism would result in a futile cycle dissipating the membrane potential. Respiring mitochondria in potassium chloride media undergo massive swelling upon addition of the potassium ionophore valinomycin (which allows the rapid electrophoretic uptake of K^+). Simultaneous addition of excess nigericin (which promotes the electroneutral exchange of H^+ and K^+) does prevent swelling but causes permanent uncoupling (where electron

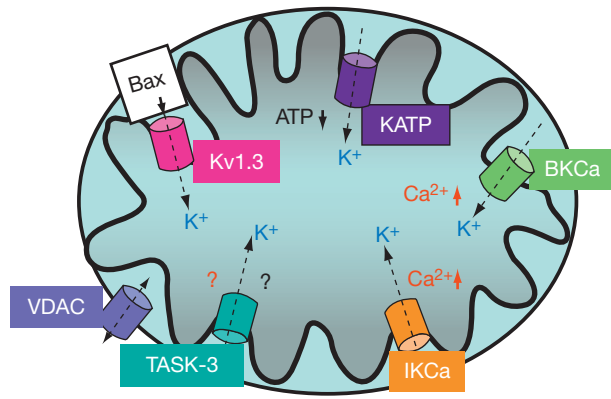


Figure 1 Potassium channels in the inner mitochondrial membrane. Two calcium-activated channels (BKCa and IKCa), two-pore potassium channel TASK-3, an ATP-inhibited channel (KATP), and Kv1.3, which can be blocked by a lysine residue (K128) of outer-membrane-inserted Bax molecule, are shown. In the outer membrane, the voltage-gated anion channel, VDAC, is indicated.

transport through components of the respiratory chain is not coupled to ATP production). In intact mitochondria, energy dissipation is limited by the fact that channels and electrophoretic uptake pathways (uniports) are tightly regulated. The K^+ cycle thus can potentially modulate the tightness of coupling between respiration and ATP synthesis in mitochondria, thereby maintaining a balance between energy supply and demand in the cell. It also ensures mitochondrial volume regulation, preventing excessive matrix swelling, thus maintaining the structural integrity of the organelle, or contraction, which would inhibit respiration. Finally, the K^+ cycle seems to regulate mitochondrial reactive oxygen species (ROS) production. This process is not yet completely understood. Some experimental results suggest that depolarization of the IMM and increase of the respiratory rate induced by opening of a K^+ channel may increase the leak of electrons from the respiratory chain to form superoxide anion, that is, ROS. On the other hand, activation of potassium channels has also been reported to reduce ROS production. Both effects might take place, depending on the tissue under consideration.

The Mitochondrial ATP-Dependent Potassium Channel

An ATP-sensitive potassium channel has been described in mitochondria of various tissues in mammals as well as in unicellular organisms and in plant mitochondria and has been proposed to correspond to the well-studied potassium uniport. The channel is selective for K^+ over Na^+ and in the presence of Mg^{2+} it is inhibited by ATP added to the cytoplasmic side of the IMM. The ATP-inhibited channel can be reactivated by the plasma membrane KATP channel openers, cromakalim and diazoxide, the latter being much more potent with the mitochondrial than the plasma membrane channel. Other inducers of mitoKATP activity include ROS, guanosine triphosphate (GTP), guanosine diphosphate (GDP), and uridine diphosphate (UDP), while in

addition to ATP and adenosine diphosphate (ADP), NO, quinine, 5-hydroxydecanoic acid, glybenclamide, palmitoyl-CoA, and oleyl-CoA inhibit mitoKATP.

The detailed pharmacological characterization, initially performed mainly on rat liver mitochondria, later allowed determination of the presence of mitoKATP in various organisms. In the nonphotosynthesising free-living amoeboid protist (*Amoebozoia*), *Acanthamoeba castellanii*, the electrophysiological (single-channel properties of the channel reconstituted into planar lipid membrane) and pharmacological (effect of specific modulators on bioenergetics of isolated mitochondria) profiles of the mitoKATP channel display similarities to mammalian plasma membrane ATP-sensitive K^+ channels (KATP channels). Swelling experiments have been used to assess the effects of some of the above-mentioned substances, known to modulate the mammalian mitoKATP channel, on mitochondria of protozoan parasites trypanosomatids, *Trypanosoma cruzi* (which causes Chagas' disease) and *Crithidia fasciculata*. MitoKATP activity was identified also in mitochondria of the nematode worm *Caenorhabditis elegans*. In higher organisms, mitoKATP channel activity, revealed by electrophysiology or classical bioenergetic techniques, is present in mitochondria of liver, heart, brain, kidney, skeletal muscle, and human T lymphocytes. Furthermore, in plant mitochondria, which possess a very active K^+ cycle, a K^+ uniport pathway with characteristics of mitoKATP has been observed to sustain an ion flow sufficient to uncouple respiration. Very recently, the first patch-clamp experiments on plant mitochondria identified mitoKATP also at the level of single-channel activity.

Concerning the molecular identity of mitoKATP, a definitive identification is not yet available. As for its plasma-membrane-located counterpart (KATP channel), inward rectifying potassium channel subunits (Kir6.1 or Kir6.2), together with the regulatory subunit sulfonylurea receptor (SUR), have been proposed to give rise to mitoKATP activity. A comparative genomics/proteomics approach may help molecular identification in view of the fact that mitoKATP exist throughout eukaryotes.

MitoKATP is involved in the regulation of mitochondrial ionic homeostasis (see above). Importantly, a cytoprotective effect of mitoKATP activation has been observed, possibly due to the modulation of the rate of ROS generation in mitochondria. In accordance, mitoKATP has been proposed to play a role in ischemic preconditioning (IP) consisting of brief sublethal ischemic periods that are capable of protecting against subsequent massive ischemia. Similarly, in plant mitochondria, mitoKATP channel is proposed to be involved in the prevention of ROS formation.

Calcium-Dependent Potassium Channels

Two types of calcium-activated potassium channels have been described in various systems: mitoBKCa (big conductance potassium channel) and mitoIKCa (intermediate conductance channel). MitoBKCa (KCa1.1), displaying a conductance of 100–300 pS, has been observed by direct patch clamping of mitoplasts of mammalian cells as well as in planar lipid bilayer experiments. The channel has been identified in glioma cell

lines as well as in ventricular cells, rat skeletal muscle, and brain. Evidence for its presence in mitochondria has also been provided by Western blot, electron microscopy, and immunofluorescence microscopy. The channel is activated by μM concentration of calcium, and is blocked by specific inhibitors charybdotoxin, iberiotoxin, and paxilline. Concerning the molecular nature of mitoBKCa, it seems to coincide with the PM-located BKCa and/or its alternatively spliced forms.

MitoBKCa has been proposed to be activated under pathophysiological conditions that increase mitochondrial Ca^{2+} uptake, thus preventing excessive mitochondrial Ca^{2+} accumulation (by partially depolarizing the IMM, the driving force for calcium influx is reduced) and may also play a physiological role to fine-tune mitochondrial volume and/or Ca^{2+} accumulation under conditions of increased cardiac workload. Opening of mitoBKCa, similarly to mitoKATP, has been reported to protect against damage to the heart and other organs caused by ischemia and reperfusion. Interestingly, mitoBKCa activity has been revealed also in potato, tomato, and maize mitochondria by swelling measurements, and a large-conductance Ca^{2+} -activated potassium channel has been described in potato tuber mitochondrial fractions incorporated into an artificial planar lipid membrane. The mitoBKCa channel in plant mitochondria may function as a possible signaling link between matrix calcium levels and the mitochondrial membrane potential.

The mitoIKCa (KCa3.1) has been recently observed by patch-clamping mitoplasts isolated from human colon carcinoma cells. This channel has been detected in other cell types, namely in HeLa of human cervix adenocarcinoma origin and in mouse embryonic fibroblasts. The channel is selectively inhibited by clotrimazole and TRAM-34. The biophysical and pharmacological properties of the mitoIKCa and of the PM IKCa channels of the same cells are indistinguishable. MitoIKCa may have a protective role similar to that proposed for mitoBKCa and mitoKATP channels; however, so far its physiological role has not been studied in detail.

MitoKv1.3

Kv1.3 is a highly selective channel expressed in the plasma membrane of various cells. Its presence in the mitochondria has been reported in T lymphocytes, macrophages, astrocytes, and hippocampal neurons by several groups, using patch clamp and immunochemistry. The molecular identification of the channel activity was obtained thanks to specific inhibitors margatoxin and ShK, and to a genetic model consisting in Kv1.3-less CTLL-2 cytotoxic lymphocytes and CTLL-2 cells stably expressing Kv1.3. Importantly, a hyperpolarization of the mitochondrial membrane potential could be recorded upon inhibition of mitoKv1.3 with specific drugs, indicating that in energized mitochondria the channel is normally active.

The above model pointed out a role of mitoKv1.3 in apoptosis. MitoKv1.3 was identified as a novel target of the proapoptotic Bcl-2 family protein, Bax, and a physical interaction between the two proteins was shown to occur in apoptotic cells. Incubating Kv1.3-positive isolated mitochondria with Bax or Kv1.3-inhibiting Margatoxin and ShK triggered apoptotic events

including membrane potential changes, ROS production, and cytochrome c release, whereas Kv1.3-deficient mitochondria were resistant. After insertion of activated Bax in the outer mitochondrial membrane, the highly conserved Bax lysine 128 is located in the intermembrane space and mimics a crucial lysine in Kv1.3-blocking toxins. Mutation of Bax at K128 (BaxK128E) abrogated its effects on Kv1.3 and mitochondria also in intact cells, indicating a toxin-like action of Bax on Kv1.3 to trigger at least some of the mitochondrial changes typical for apoptosis. The crucial role of Kv1.3 in cell death was demonstrated by the fact that its knockdown in human peripheral blood lymphocytes impaired apoptosis induction by various stimuli.

Two-Pore Potassium Channel TASK-3

Recently TASK-3, a two-pore potassium channel known to reside in the plasma membrane, was identified in mitochondria of melanoma and keratinocyte cells by immunochemical and molecular biology methods. The functional properties as well as the physiological function of the TASK-3 channel in the inner mitochondrial membrane remain to be investigated.

Open Questions

In summary, the mitochondrial potassium channels listed above have a tissue-dependent expression, and the ones for which molecular candidates exist appear to represent subpopulations of channels present also in the plasma membrane. This raises questions concerning the mechanism of differential targeting of the channels and also casts some doubts on the effective presence of these channels in mitochondria, even though an analogous multiple localization has been established for various other proteins (e.g., kinases). The several mitochondrial proteomic studies published so far have failed to recognize mitochondrial channels except mitochondrial porin. This presumably reflects technical difficulties ascribable to the hydrophobic nature and especially to the low abundance of channel proteins. Genetic approaches may be useful to obtain conclusive evidence on the molecular identity and localization of some of the IMM channels. Using a biochemical approach, the observation that the abundance of a protein in various membranous fractions increases along with that of mitochondrial markers, while markers of contaminating membranes (PM and endoplasmatic reticulum) decrease, may be considered good evidence for a mitochondrial location. Last but not least, observation of a channel by patch clamp in swollen mitoplasts, that is, in the inner membrane, directly proves its presence in that membrane. A comparative proteomic/genetic approach combined with a systematic RNA interference screening may give the link between the genes coding for potassium channels and the mitochondrial potassium channels.

See also: **Bioenergetics:** Mitochondrial Channels; Mitochondrial Permeability Transition Pore.

Further Reading

- Bernardi P (1999) Mitochondrial transport of cations: Channels, exchangers, and permeability transition. *Physiological Reviews* 79: 1127–1155.
- Jarmuszkiewicz W, Matkovic K, and Koszela-Piotrowska I (2010) Potassium channels in the mitochondria of unicellular eukaryotes and plants. *FEBS Letters* 584: 2057–2062.
- O'Rourke B (2007) Mitochondrial ion channels. *Annual Review of Physiology* 69: 19–49.
- Szabò I, Zoratti M, and Gulbins E (2010) Contribution of voltage-gated potassium channels to the regulation of apoptosis. *FEBS Letters* 584: 2049–2056.
- Szewczyk A, Kajma A, Malinska D, et al. (2010) Pharmacology of mitochondrial potassium channels: Dark side of the field. *FEBS Letters* 584: 2063–2069.
- Zoratti M, De Marchi U, Gulbins E, and Szabò I (2009) Novel channels of the inner mitochondrial membrane. *Biochimica et Biophysica Acta Bioenergetics* 1787: 351–363.

Pre-tRNA and Pre-rRNA Processing in Bacteria

Z Li, Florida Atlantic University, Boca Raton, FL, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Sidney R. Kushner, volume 3, pp. 420–424, © 2004, Elsevier Inc.

Glossary

Double-stranded RNA RNA structure that is formed either inter- or intra-molecularly between two complementary sequences. The base pairs that are normally found in RNA include A-U, G-C, and G-U. A stable RNA double strand may exist even when the complementarity is short or not continuous.

Polycistronic RNA transcript RNA that contains more than one functional RNA species. Polycistronic RNAs are common in bacteria, including primary transcripts of operons encoding rRNA–tRNA, tRNA, tRNA–mRNA, and mRNA. Polycistronic RNAs consist of a leader sequence preceding the first gene, a trailer sequence following the last gene, and spacer sequences in between the genes.

Ribonuclease (RNase) Enzyme that degrades a RNA molecule by cutting phosphodiester bonds. An endoribonuclease cleaves RNA at internal locations. An exoribonuclease removes one nucleotide at a time in either 3'→5' or 5'→3' direction.

RNA polarity RNA backbone is made of ribose sugars linked together by phosphodiester bonds at the 5' and 3' carbon atoms of ribose rings, conferring RNA directionality, or polarity. A cellular RNA usually contains one or three free phosphate groups at the 5'-end, and a free hydroxyl group at the 3'-end.

Sedimentation coefficient (S) The S value, determined by the mobility of a particle in a centrifugal field, is a measure of the size of the macromolecule or a complex of multiple macromolecules. The higher the S value the larger the molecule.

Transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) constitute the large majority of cellular RNA in bacteria. In exponentially growing *Escherichia coli* cells, rRNAs and tRNAs account for nearly 80% and 15% in total RNA, respectively. The rRNAs and numerous ribosomal proteins form large complexes known as ribosomes, which are the machinery for protein synthesis. tRNAs work in protein synthesis as carriers of specific amino acids that are added to defined peptide sequences by the ribosome. Both tRNAs and rRNAs are very stable, being functional in generations of cells once they are made. Growing bacterial cells constantly synthesize large amount of tRNAs and rRNAs to keep up with the need for protein synthesis.

A universal feature for all tRNAs and rRNAs is that they are transcribed as precursors that require extensive processing to form the mature, functional forms. In bacteria, the primary transcripts contain extra sequences at both 5'- and 3'-termini of the mature RNA sequences. The extra sequences are removed precisely by specific ribonucleases. This article describes nucleolytic processing pathways that are best understood in model bacteria such as *E. coli* and *Bacillus subtilis*. In addition, numerous nucleotides are modified post-transcriptionally in both tRNAs and rRNAs, which is not discussed in this article.

Processing of Pre-tRNAs

Mature bacterial tRNAs are relatively short molecules ranging from 73 to 95 nucleotides in length. The tRNAs fold into cloverleaf structures at the two-dimensional level (Figure 1), and adopt inverted L structures at the three-dimensional level. Such structures are likely formed prior or during processing. The numbers of tRNA species range from 36 to over 100 in bacteria. In most cases, multiple tRNAs can be charged by one

of the 21 amino acids. The 3'-end of mature tRNAs contain the invariable trinucleotide cytidylate–cytidylate–adenylate (CCA). Specific amino acids can be covalently attached to the 3'-terminal A residue. In bacteria, the CCA is generally encoded by the tRNA gene. However, many bacteria contain CCA-less tRNA genes, and CCA is added post-transcriptionally to these tRNAs.

Our knowledge of tRNA maturation is most advanced in *E. coli*, an organism in which most, if not all, ribonuclease activities have been identified and studied in detail. Bacterial ribonucleases can be divided into endoribonucleases and exoribonucleases. Endoribonucleases responsible for intracellular RNA processing and degradation generally cleave RNA at specific sites, generating 5'-phosphate and 3'-hydroxyl end groups in the resulting fragments. All of the eight known exoribonucleases in *E. coli* digest RNA from the 3' to 5' direction, generating nucleoside-5'-monophosphates. However, exoribonucleases digesting RNA from 5' to 3' exist in bacteria, such as RNase J1 in *B. subtilis*.

Structure of Primary tRNA Transcripts

The large numbers of tRNA genes in each bacterium are organized into various transcriptional units that are located throughout the genome. For instance, tRNA genes in *E. coli* may exist as monocistronic genes, as operons containing a group of tRNA genes transcribed together, and as part of the operons that also contain genes encoding rRNAs or mRNAs. In *Mycoplasma genitalium*, a model bacterium for minimum genome, 27 of the 36 tRNA genes are arranged into five polycistronic tRNA operons, and the rest are probably monocistronic.

As illustrated in Figure 1, the primary tRNA transcripts may contain extra sequences at both 5'- and 3'-termini of each

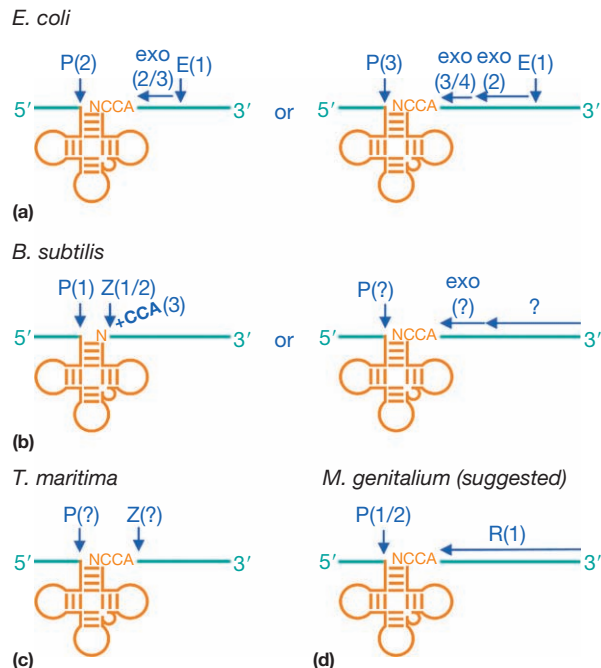


Figure 1 tRNA processing pathways in various bacteria. The primary transcripts contain extra sequences (marked in dark green lines) at both ends of the mature tRNAs (colored in orange). These extra sequences are removed by various endo- or exo-ribonucleases that are indicated in blue color. The suggested order of RNase actions is indicated by numbers in the parenthesis. Processing of most pre-tRNAs in *E. coli* (a) is initiated by RNase E cleavage in the 3'-trailer sequence. If the 3'-extra sequence is long (right part), it can be shortened exonucleolytically by RNase II or PNPase. Final maturation of the 3'-end requires exonucleolytic trimming, mainly by RNase T and PH and less efficiently by RNase D, II, and BN. If the 5'-extra sequence blocks 3'-maturation such as by forming base pairs, RNase P may cleave first, followed by exonucleolytic trimming of the unblocked 3'-end. Otherwise, both ends can be matured simultaneously. In *B. subtilis* (b), CCA-less pre-tRNAs are cleaved by RNase P to generate the mature 5'-end and by RNase Z downstream of the discriminator (left part). CCA is then added by tRNA nucleotidyl transferase. CCA-containing pre-tRNAs (right part) undergo exonucleolytic maturation at the 3'-end by RNase PH, and possibly also RNase R and YhaM. A long 3'-trailer may be shortened prior to maturation by the exoribonucleases. In *T. maritima* (c), RNase P and Z cleavages generate mature termini from CCA-containing pre-tRNAs. (d) Exonucleolytic trimming by RNase R can generate mature 3'-end for *M. genitalium* tRNA. Maturation of the 3'-end does not require prior removal of the 5'-extra sequence by RNase P. Modified from Li Z, Gong X, Joshi VH, and Li M (2005) Co-evolution of tRNA 3' trailer sequences with 3' processing enzymes in bacteria. *RNA* 11: 567–577, with permission from Cold Spring Harbor Laboratory Press.

tRNA. A monocistronic tRNA transcript contains a leader sequence from transcription start site to the residue upstream of the 5'-end of mature tRNA, and a trailer sequence from the residue immediately downstream of the 3'-terminus of tRNA to the end of the transcript. The transcription terminator sequences are usually copied into the 3'-trailers of tRNA transcripts. Polycistronic transcripts contain an additional spacer sequence between each of the mature RNA. The leader, spacer, and trailer sequences are removed by ribonucleases in sequential orders during tRNA processing.

Processing at the 5'-End of Bacterial tRNA Precursors

The 5'-leaders of tRNAs are removed by an endoribonuclease RNase P, an ubiquitous essential enzyme. In most cases, RNase P is a complex formed by an RNA and a protein. The RNA subunit catalyzes the cleavage of the 5'-leader sequence precisely upstream of the mature 5'-end. RNase P is found in all organisms except in *Aquifex aeolicus* where RNase P has not been identified but similar activity exists.

RNase P recognizes tRNA precursor by interacting with the pre-tRNA at multiple sequence and structural domains. Some key interactions include the formation of base pairing of RNase P RNA with the discriminator and CCA sequence of tRNA. These interactions place the 5'-end of tRNA at the catalytic center of the RNase P RNA. The endonucleolytic cleavage by RNase P releases the leader and generates 5'-phosphate and 3'-hydroxyl end groups in the products. In *E. coli*, RNase P cleaves 5'-leader less efficiently if a long 3'-trailer is also present in the pre-tRNA.

Processing of tRNA 3'-End in *E. coli*

Processing at the 3'-end is more complicated because bacteria have developed different pathways to handle this problem. In *E. coli*, multiple enzymes are involved in removing the 3'-trailers (Figure 1(a)).

The 86 tRNAs identified in *E. coli* contain encoded 3'-terminal CCA in all of them. In most cases, the endoribonuclease RNase E carries out the first step of tRNA maturation by cleaving in the 3'-trailers or spacer sequences in the primary tRNA transcripts, separating individual pre-tRNAs. RNase E recognizes and cleaves at A-U-rich positions that are usually a few nucleotides downstream of the 3'-end of tRNAs (Figure 1(a), left). The intermediates with short 3'-trailers are then cleaved efficiently by RNase P to generate the mature 5'-end. The few 3'-extra residues are removed by exoribonucleases, mainly RNase T and PH. Other exoribonucleases, including the tRNA-specific RNases D and BN (also known as RNase Z and can act as an endoribonuclease), and the nonspecific RNase II, remove the 3'-extra residues less efficiently. These enzymes may contribute to normal tRNA 3'-maturation and become essential in this process in the absence of RNase T and PH. In a few cases, RNase E cleaves further downstream (Figure 1(a), right). The resulting longer 3'-trailer sequences are shortened by the nonspecific exoribonucleases, RNase II and polynucleotide phosphorylase (PNPase), prior to the actions of RNase P, T, and PH.

An interesting point of the exonucleolytic 3'-maturation is that all the processing exoribonucleases are sensitive to the double-stranded structure of the amino acid stem. The structure may serve to stop the exonucleolytic trimming reaction at the end of CCA. In most cases, the stem in pre-tRNAs ends at the base pair between the first residue at the mature 5'-end and the residue upstream of the discriminator. For such precursors, the exonucleolytic trimming reactions may occur simultaneously with RNase P action. Some tRNA precursors contain extended double strand in which the discriminator and CCA sequence may form base pairs with part of the 5'-leader. In such cases, the 5'-leader must be removed first to

unblock the 3'-region before final exonucleolytic maturation of the 3'-end.

Maturation of tRNA 3'-End in Other Bacteria

The maturation pathway found in *E. coli* may also be employed by a number of other bacteria, mainly in the proteobacteria β - and γ -subdivisions, and occasionally in the α -subdivision. These bacteria contain RNase E as well as three tRNA-specific exoribonucleases RNase T, PH, and D, with a few exceptions in which at least one of the three enzymes exists.

The CCA-containing and CCA-less tRNAs in *B. subtilis* appear to be matured by different mechanisms at the 3'-end (Figure 1(b)). The 3'-end of the CCA-less pre-tRNAs is matured in two steps. First, the endoribonuclease RNase Z cleaves the pre-tRNA after the discriminator residue, similar to the way its eukaryotic counterparts work. The CCA sequence is then added by tRNA nucleotidyl transferase. It is interesting that RNase Z is sensitive to the first C residue in CCA sequence, making it unlikely to cleave CCA-containing pre-tRNAs. In addition, RNase Z cleavage at the 3'-end may occur before the removal of 5'-leader, although it can be less efficient in the presence of a long 5'-leader. The 3'-trailer sequences in CCA-containing pre-tRNAs are probably shortened first by activities not yet identified. The shorter 3'-tails are then removed exonucleolytically. RNase PH seems to be the most important RNase in this final step. Other exoribonucleases, such as RNase R, PNPase, and YhaM, may also contribute to 3'-exonucleolytic maturation of CCA-containing tRNAs.

The bacterium *Thermotoga maritima* presents a simpler tRNA processing pathway. The RNase Z in *T. maritima* cleaves at a position immediately after the encoded CCA, generating mature 3'-end directly (Figure 1(c)).

RNase Z is present throughout bacterial, archaeal, and eukaryotic kingdoms. The eukaryotic counterparts of this RNase seem to cleave pre-tRNAs immediately after the discriminator residue. It is not clear why RNase Z enzymes in different bacteria work differently in tRNA 3'-processing. As mentioned above, RNase Z in *B. subtilis* efficiently cleaves CCA-less pre-tRNAs but is inhibited by CCA sequence. RNase Z in *T. maritima* appears to have evolved to cleave directly at the mature 3'-end of CCA-containing pre-tRNAs. RNase Z in *E. coli* was initially identified as an exoribonuclease RNase BN. It is now known to work as both endo- and exoribonuclease. *E. coli* RNase Z is also sensitive to CCA sequence, and it works poorly for 3'-maturation of *E. coli* tRNAs. Consistent with its role in tRNA maturation, RNase Z is essential for *B. subtilis* viability but is dispensable for *E. coli* growth.

Mycoplasma species must have developed a unique pathway for tRNA 3'-processing as none of the 3'-processing enzymes described above exists in this group of bacteria. The only exoribonuclease that seems to be present in mycoplasmas is an RNase belonging to the RNase II/R family. The RNase R ortholog in *M. genitalium* has some catalytic features similar to those of *E. coli* RNase R, such as the ability to efficiently degrade structured RNA. Remarkably, purified RNase R from *M. genitalium* can generate mature 3'-end from a pre-tRNA *in vitro*, making it a candidate for the maturation of tRNA 3'-end. In contrast, RNase R from *E. coli* degrades such

pre-tRNA completely. It is likely that mycoplasmas have adopted a novel mechanism to remove the 3'-trailers of pre-tRNAs by a single exoribonuclease in one step (Figure 1(d)).

Processing of Pre-rRNAs

Bacteria contain 5S, 16S, and 23S rRNAs, where S is the sedimentation coefficient used as a measure of RNA size. The 5S, 16S, and 23S rRNAs in *E. coli* are 120, 1542, and 2904/2905 nucleotide in length, respectively. The length of the rRNAs may vary slightly in other bacteria. The rRNAs fold into complicated secondary and tertiary structures, most likely during or prior to their maturation. One of each rRNA species and numerous ribosomal proteins are assembled into a ribosome, with 5S and 23S in the large (50S) ribosome subunit and 16S in the small (30S) ribosome subunit. The rRNAs not only provide the structural core of the ribosome, but may also be catalytically active in protein synthesis reactions. For instance, the RNA component of the 50S subunit may catalyze peptide bond formation. Pre-rRNA processing shares some of the RNases that are involved in tRNA maturation.

Structure of Primary rRNA Transcripts

The three rRNAs are usually encoded in the same operon and are co-transcribed together. In most bacterial genomes, the rRNA operons are duplicated to ensure the production of sufficient rRNA transcripts. In *E. coli*, there are seven rRNA operons (*rrn*), each containing rRNA genes in the order of 16S, 23S, and 5S. There are one or two tRNA genes between 16S and 23S rRNAs. At the distal region, there are two more tRNA genes downstream of 5S rRNA in the *rrnC* and *rrnH* operons, and one tRNA gene and one additional 5S rRNA gene in the *rrnD* operon. In *B. subtilis*, 16S, 23S, and 5S rRNAs are co-transcribed from 10 *rrn* operons. Majority of the primary transcripts also contain tRNAs that are located upstream of 16S rRNA, in between 16S and 23S, or downstream of 5S. In *M. genitalium*, the 16S, 23S, and 5S rRNAs are encoded by a single operon with short spacer regions in between the rRNAs.

As shown in Figure 2, the primary rRNA transcripts in *E. coli* and *B. subtilis* may contain extra sequences flanking each of the functional RNA species. These include the sequence upstream of the 16S rRNA, the spacer region between each of the rRNA or tRNA species, and the sequence downstream of the most distal RNA ended by a terminator. The rRNAs may be folded right after they emerge during transcription, and are associated with ribosomal proteins. It can be expected that many of the nucleolytic processing and ribosome assembly events may interfere with each other and can be co-transcriptional. Importantly, the precursor sequences form extended double-stranded regions at the termini of each rRNA species. As described below, these double-stranded regions are important structural determinants for various processing events.

Primary Processing Reactions

In *E. coli* and *B. subtilis*, the 30S primary rRNA transcripts are first processed by RNase III, an endoribonuclease that

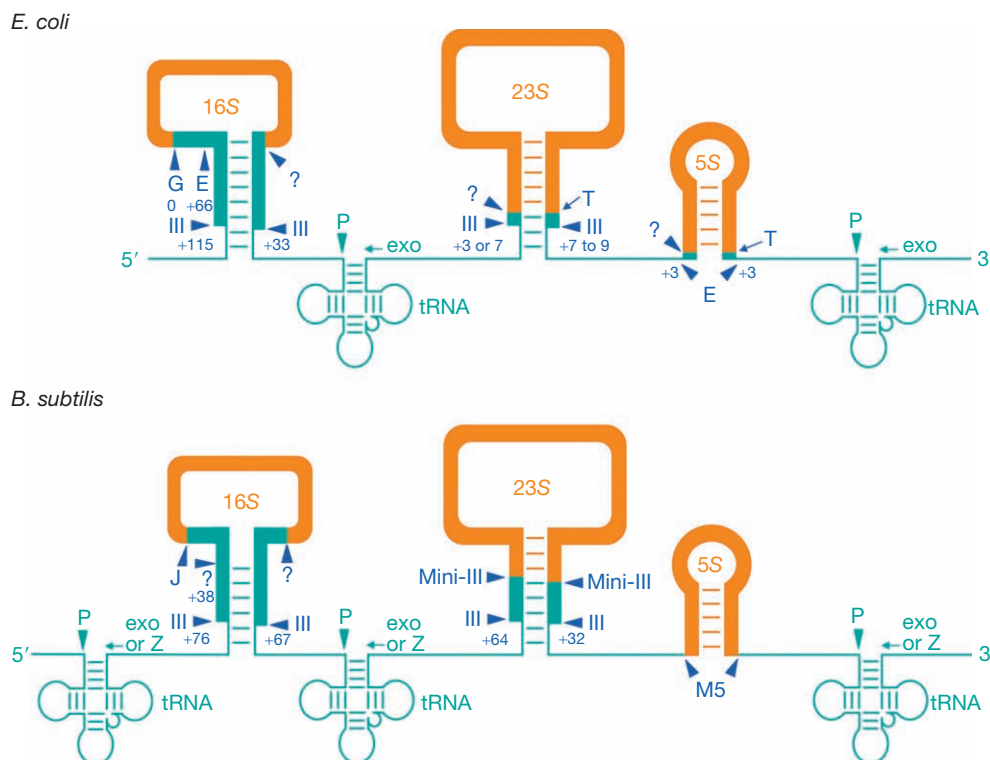


Figure 2 rRNA maturation pathways in *E. coli* (upper part) and *B. subtilis* (lower part). The rRNA sequences (colored in orange) are separated by extra sequences (marked in green lines) in the 30S primary transcripts. The RNases involved in rRNA maturation at indicated positions are colored in blue. Note the similar organization of rRNA transcripts and primary processing step in both bacteria. *E. coli* and *B. subtilis* differ in secondary processing by employing different sets of RNases. Modified from Li Z and Deutscher MP (2004) Exoribonucleases and endoribonucleases. In: Curtiss R, III (ed.) *EcoSal—Escherichia coli and Salmonella: Cellular and Molecular Biology*, ch. 4.6.3. Washington, DC: ASM Press.

recognizes and cleaves double-stranded RNAs (Figure 2). RNase III cleaves both RNA strands in the stem flanking 16S and 23S rRNAs, separating the three rRNA species into fragments which undergo further processing to form mature RNAs. This process is so efficient in exponentially growing cells that essentially no 30S RNA is detectable at steady state. Other bacteria may also employ RNase III cleavage as the primary step of processing because RNase III is widely distributed in bacteria and rRNAs are organized in similar transcripts. The subsequent processing events are relatively well studied in *E. coli* and *B. subtilis*.

Generation of Mature rRNAs in *E. coli*

16S rRNA

RNase III cleavages in the stem flanking 16S rRNA generate a precursor of 17S in size, containing 115 extra residues at the 5'-end and 33 at the 3'-end (Figure 2, upper part). Maturation of the 5'-terminus is dependent on the sequential actions of two homologous endoribonucleases, RNase E and RNase G. First, RNase E cleavage at 66 residues upstream of 16S removes 49 nucleotides, generating a 16.3S product. Subsequently, RNase G removes the remaining 66 nucleotides rapidly, generating the mature 5'-end. When RNase E is inactive, RNase G is able to remove the entire 115 residues at a much reduced rate. In the absence of RNase G, the 16.3S product is slowly

converted to a product containing four to five extra residues at the 5'-end by RNase E. No other activity can process the 115 leader sequence when both RNase E and RNase G are missing. Maturation of 5'-end likely occurs in partially or completely assembled ribosomes as purified ribosomes containing the 5' extra residues can be cleaved by RNase E and G at the correct sites.

Maturation of the 3'-end of 16S rRNA is thought to be endonucleolytic, even though the activity responsible for this event has not yet been identified. In supporting this idea, inactivation of the exoribonucleases responsible for 3'-maturation of other RNAs has no effect on the 33 extra residues. In addition, the 33 residues appear to be removed in one step, forming no intermediate between the 3'-mature terminus and the +33 precursor. It is also known that 3'-processing is slightly inhibited by the blockage of 5'-maturation.

23S rRNA

The RNase III cleavage produces processing intermediates of 23S rRNA containing seven to nine extra residues at the 3'-end and either three or seven extra residues at the 5'-end (Figure 2, upper part). Maturation of the 3'-end requires the exoribonuclease RNase T. In the absence of RNase T, little mature 3'-end is made, and products ranging from one to six extra residues accumulate. These products must be converted from RNase III cleavage products by other exoribonucleases. In cells deficient

in RNase PH, D, II, and BN in combination, but retaining RNase T, the majority of 23S rRNA is matured at the 3'-end, indicating that RNase T plays a major role in generating the mature 3'-end. In such cells, slightly more pre-23S rRNA containing seven to nine extra residues at the 3'-end was detected, suggesting that the missing exoribonucleases may be responsible for shortening these products before RNase T action. Interestingly, RNase T removes the last extra residues much more efficiently from pre-23S rRNA in the ribosome than from isolated pre-23S rRNA, suggesting that 3'-maturation occurs during or after ribosome assembly.

Removal of the three or seven extra residues at the 5'-end occurs independently of 3'-maturation. Limited evidence suggests that 5'-maturation is probably carried out by an endoribonuclease other than RNases III, E, or G. It is also known that generation of mature 5'-terminus requires protein synthesis conditions.

5S rRNA

RNase III cleavage around 23S rRNA releases a downstream product containing 5S rRNA (Figure 2, upper part). For most rRNA transcripts, this processing intermediate is 9S in size. The 9S RNA is rapidly cleaved by RNase E at sites flanking 5S sequence, generating a pre-5S RNA molecule that contains three extra residues at each end. The 5'-extra residues are thought to be removed by an exonucleolytic activity. However, this 5'- to 3'-exoribonuclease has never been identified, and it is also likely that an endoribonuclease makes sequential cleavages to generate the mature 5'-terminus. Removal of the 5'-extra residues is independent of 3'-maturation of 5S RNA.

Maturation of the 3'-end is dependent on the exoribonuclease RNase T. RNase T can remove all three extra residues. Cells having RNase T contain essentially only mature 5S rRNA, regardless of whether other exoribonucleases are present or not. In the absence of RNase T, other exoribonucleases can only remove the residue at the +3 position, leaving a product with two extra residues at the 3'-terminus. Similar to 23S rRNA, 5S rRNA is matured at the 3'-end by RNase T much more efficiently in the ribosome than in isolated form.

RNase T is one of the key exoribonucleases for 3'-maturation of tRNA and other small, stable RNAs. This enzyme is absolutely required for 3'-processing of 5S and 23S rRNAs, and is the only RNase responsible for tRNA end turnover. A common feature present in all of these RNAs is the stable stem structures formed between the 5'- and 3'-terminal sequences (Figure 3). As mentioned above, the 3'- to 5'-exoribonucleases are sensitive to double strands in the RNA and they tend to stop trimming when the 3'-end becomes close to the terminal stem. Most of the RNAs undergoing exonucleolytic 3'-maturation contain four single-stranded residues at the mature 3'-end located downstream of the stem. Thus, the terminal stem structure may serve as a 'molecular ruler' for exonucleolytic digestion to stop at the correct downstream position. RNase T seems to be able to go closer to the double-stranded region than other exoribonucleases (Figure 3). This ability explains why it is the only enzyme that can generate mature 3'-end for 5S and 23S rRNAs that contain a one- and two-residue overhang, respectively. For similar reasons, RNase T is the only enzyme that can trim the uncharged CCA sequence of tRNA species for certain amino acids. Such tRNAs containing shortened CCA sequence are nonfunctional and can be repaired by tRNA nucleotidyl transferase.

Generation of Mature rRNAs in *B. subtilis*

16S rRNA

The pre-16S rRNA generated by RNase III cleavage contains 67 extra residues at the 3'-end and 76 or 140 extra residues at the 5'-end from rRNA transcripts encoded by different *rnm* operons. Removal of the extra residues follows similar steps as described in *E. coli*; however, different ribonucleases may be used in *B. subtilis*. In fact, RNase E and G, the enzymes responsible for 5'-maturation of 16S RNA in *E. coli*, are not present in *B. subtilis*. Instead, *B. subtilis* employs an as-yet unidentified enzyme(s) to cleave the 76 extra residues at +38 or 140 extra residues at +102. These shorter products are then removed by an essential enzyme RNase J1 to generate the mature 5'-terminus (Figure 2, lower part). Interestingly, RNase J1 is a large protein exhibiting both endoribonuclease and 5'- to

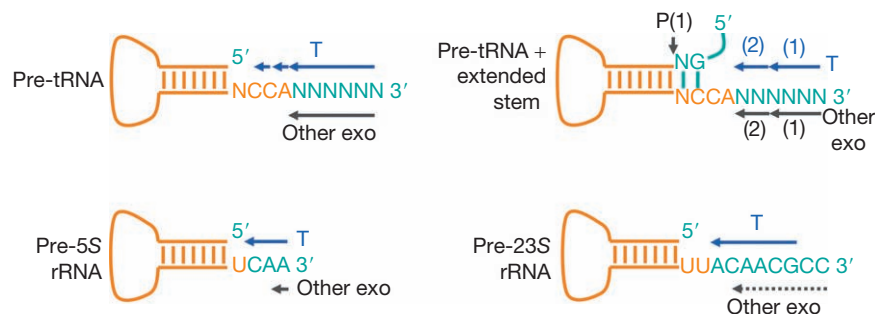


Figure 3 3'-Exonucleolytic trimming is hindered by RNA double-stranded structures in *E. coli* RNAs. The 3'-extra residues (marked in dark-green) in pre-tRNA, pre-23S, and pre-5S rRNA sequences (colored in orange) are removed by exoribonucleases during the final maturation step. The exoribonucleases are generally sensitive to RNA structures, and their trimming stops when the 3'-end is close to the stem formed between the termini of the RNAs. In all cases, RNase T (colored in blue) can trim residues closer to the stem than other exoribonucleases. This property of RNase T is consistent with its role as a major 3'-maturation enzyme for tRNAs, and as the sole RNase for the 3'-maturation of 23S and 5S rRNAs as well as for tRNA end turnover. Modified from Li Z and Deutscher MP (2004) Exoribonucleases and endoribonucleases. In: Curtiss R, III (ed.) *EcoSal—Escherichia coli and Salmonella: Cellular and Molecular Biology*, ch. 4.6.3. Washington, DC: ASM Press.

3'-exoribonuclease activities. Maturation of the 5'-end of 16S rRNA is probably carried out by the exonucleolytic activity of this enzyme. The RNase(s) responsible for 3'-maturation is yet unidentified.

23S rRNA

RNase III cleavage produces a pre-23S rRNA containing 64 extra residues at 5'-end and 32 extra residues at 3'-end. These extra residues present in a long double-stranded region are removed by another double-strand-specific endoribonuclease Mini-III, which cleaves both sides of the stem to generate mature termini in one step (Figure 2, lower part). However, Mini-III is not essential for 23S rRNA maturation. It has been recently reported that in *B. subtilis* cells lacking Mini-III, RNase J1 and the 3'- to 5'-exoribonucleases RNase PH and YhaM generate mature 5'- and 3'-termini, respectively. RNase J1 removes the 5'-extra residues in a way consistent with its exonucleolytic activity.

5S rRNA

Pre-5S rRNA is released from the primary rRNA transcripts by RNase III cleavage. This processing intermediate extends from RNase III cleavage site at the 3'-side of 23S rRNA to the 3'-terminus of the entire transcript. 5S rRNA is matured endonucleolytically by RNase M5 (Figure 2, lower part). RNase M5 is essential for maturation of 5S rRNA but is dispensable for cell viability.

Generation of Mature rRNAs in Other Bacteria

Unlike the ubiquitous primary processing by RNase III, the secondary processing pathways vary among bacteria. For instance, RNase E-dependent maturation of 16S and 5S rRNAs may only occur in relatively small number of bacteria as this enzyme seems to be limited in proteobacteria. RNase G, and in some cases enzymes such as both RNase E and G, is present in a broader range of bacteria including proteobacteria, actinobacteria, chlamydiae, clostridia, quifecales, thermotogales, and the *Chlorobium*–*Flavobacterium* group. However, it remains to be determined if the 5'-end of 16S rRNA is matured by RNase G or the E/G-like enzyme in these bacteria, especially when RNase E is not present.

RNase J1 is widely distributed throughout the bacterial and archaeal kingdoms, and is a good candidate for 5'-maturation of 16S rRNA in such organisms. *Sinorhizobium meliloti* harbors genes encoding both RNase E and J1. Interestingly, it was recently shown that RNase J, but not RNase E, is responsible for the 5'-end maturation of 16S rRNA in *S. meliloti*.

Mini-III generates mature ends for 23S rRNA, and RNase M5 does the same to 5S rRNA. Mini-III is confined in Firmicutes and the Cyanobacteria. In bacteria lacking Mini-III, 23S rRNA can be matured by exonucleolytic activities, such as RNase J1 for the 5'-end and RNase T, PH, and YhaM for the 3'-end. RNase M5 is confined in low G + C Gram-positive bacteria.

It was recently suggested that the exoribonuclease RNase R may be responsible for the maturation of 3'-termini of 16S and 5S rRNAs in *Pseudomonas syringae*. In the absence of RNase R, 16S and 5S rRNA products that contain extra residues at the 3'-ends accumulate at low temperature (4 °C).

The Consequences of Defective Processing

tRNAs

Defects in pre-tRNA processing would generate tRNA molecules that are processed incompletely or incorrectly. Such tRNAs are most likely nonfunctional because of the strict requirements for correct tRNA structure in protein synthesis. For instance, aminoacylation absolutely depends on the CCA sequence at the 3'-end of mature tRNA. Cells die if the major activities of tRNA processing fail, such as when either RNase P or RNase E is inactivated. However, a number of mechanisms may prevent more trivial defects in tRNA processing from becoming lethal. In most cases, there is more than one specific tRNA for each amino acid. If one specific tRNA is not processed correctly, others may still be functional. The 3'-maturation mechanism that involves multiple redundant exoribonucleases can avoid incomplete processing due to the failure of one RNase. Consistent with this notion, none of the tRNA maturation exoribonucleases RNase T, PH, D, II, or BN is essential in *E. coli*. Deficiency in one or a combination of up to four of these RNases slows down cell growth to levels depending on the importance of the missing enzymes in 3'-maturation, whereas inactivation of all five leads to cell death. If the CCA sequence is accidentally shortened or removed, tRNA nucleotidyl transferase can repair it to restore tRNA function.

rRNAs

Defective processing of pre-rRNA appears to be better tolerated than that of pre-tRNA because incompletely processed rRNAs are sometimes found in ribosomes that are active in protein synthesis. When the primary processing fails, cells may use alternative processing activities. In *E. coli* lacking RNase III, the primary transcripts can be cleaved by RNase E, G, and P. However, these cleavages are not as efficient as those carried out by RNase III, leading to the accumulation of primary rRNA transcripts in RNase III mutants. The cleavage products generated by the alternative enzymes can also be difficult to process in secondary processing reactions. In fact, 23S rRNA is not completely matured in RNase III mutants. Another example is that the 3'-termini of 23S and 5S rRNAs are not matured in RNase T-deficient cells. The immature rRNAs must be functional because mutants lacking RNase III or RNase T are viable.

Unlike in *E. coli*, RNase III is essential in *B. subtilis*. However, RNase III mutation leads to the accumulation of lower levels of primary rRNA transcripts in *B. subtilis* than in *E. coli*, probably due to the availability of more effective alternative processing activities in *B. subtilis*. In the absence of RNase III, RNase J1, Mini-III, and M5 can still generate mature rRNAs.

Quality Control

Accumulation of incompletely or incorrectly processed RNAs may be deleterious to cells. Such RNAs may be removed by RNA quality-control mechanisms. For instance, a mutant tRNA^{Trp} in *E. coli* is processed inefficiently and the majority is degraded at the precursor level. PNPase, a processive phosphorolytic exoribonuclease, and poly(A) polymerase play important roles in the degradation of this mutant tRNA. In the absence of both enzymes, a product containing the tRNA

and the entire 31-residue 3'-trailer accumulates to high level. Poly(A) polymerase probably facilitates degradation by adding a poly(A) tail to the stem present at the end of the 3'-trailer. The poly(A) tail may help to recruit exoribonucleases that are sensitive to stem structures.

See also: **Molecular Biology:** Messenger RNA Degradation in Bacteria; Messenger RNA Processing in Eukaryotes; Ribosome Assembly; Ribosome Structure.

Further Reading

- Condon C (2003) RNA processing and degradation in *Bacillus subtilis*. *Microbiology and Molecular Biology Reviews* 67: 157–174.
- Condon C and Putzer H (2002) The phylogenetic distribution of bacterial ribonucleases. *Nucleic Acids Research* 30: 5339–5346.
- Deutscher MP (2009) Maturation and degradation of ribosomal RNA in bacteria. *Progress in Molecular Biology and Translational Science* 85: 369–391.
- Li Z and Deutscher MP (2004) Exoribonucleases and endoribonucleases. In: Curtiss R, III (ed.) *EcoSal—Escherichia coli and Salmonella: Cellular and Molecular Biology*, ch. 4.6.3., pp. 1–107. Washington, DC: ASM Press.
- Li Z, Gong X, Joshi VH, and Li M (2005) Co-evolution of tRNA 3' trailer sequences with 3' processing enzymes in bacteria. *RNA* 11: 567–577.
- Li Z, Pandit S, and Deutscher MP (1999) RNase G (CafA protein) and RNase E are both required for the 5' maturation of 16S ribosomal RNA. *EMBO Journal* 18: 2878–2885.
- Mathy N, Bénard L, Pellegrini O, et al. (2007) 5'-to-3' exoribonuclease activity in bacteria: Role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell* 129: 681–692.
- Pellegrini O, Nezzar J, Marchfelder A, Putzer A, and Condon C (2003) Endonucleolytic processing of CCA-less tRNA precursors by RNase Z in *Bacillus subtilis*. *EMBO Journal* 22: 4534–4543.
- Redko Y, Bechhofer DH, and Condon C (2008) Mini-III, an unusual member of the RNase III family of enzymes, catalyses 23S ribosomal RNA maturation in *B. subtilis*. *Molecular Microbiology* 68: 1096–1106.
- Redko Y and Condon C (2010) Maturation of 23S rRNA in *Bacillus subtilis* in the absence of Mini-III. *Journal of Bacteriology* 192: 356–359.

Primary Biliary Cirrhosis

M M Kaplan, Tufts University School of Medicine, Boston, MA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Cholestasis A liver disorder in which bile flow is impeded. Clinical signs of cholestasis are a disproportionate increase in the serum alkaline phosphatase or gamma glutamyl transpeptidase compared to the serum alanine aminotransferase (ALT) or aspartate aminotransferase (AST). The symptom most characteristic of cholestasis is pruritus.

Hepatic encephalopathy A disturbance in consciousness that occurs in patients with advanced liver disease. Manifestations can range from subtle changes in personality to frank coma.

Micellar Pertaining to a colloid particle formed by the aggregation of small molecules. Micelles in bile contain bile acids, phospholipids, and cholesterol.

Mitochondria Organelles important in energy metabolism that are found in nucleated cells throughout the body.

Ursodeoxycholic acid A naturally occurring bile acid, the major constituent of bear bile, comprises approximately 1–2% of human bile and is used to treat primary biliary cirrhosis.

Primary biliary cirrhosis (PBC) is a chronic cholestatic liver disease that is characterized by the continuing destruction of interlobular bile ducts, progressive fibrosis, and, if not treated, the eventual development of cirrhosis and liver failure. In contrast to patients seen more than two decades ago, most of whom had symptoms of fatigue, pruritus (itching), or jaundice, the majority of PBC patients who are diagnosed now are asymptomatic. The prognosis of PBC has improved greatly during the past decade. This is due to the two factors: its detection at earlier stages and the efficacy of treatment with ursodeoxycholic acid (ursodiol, UDCA), particularly in patients who have early-stage disease. Patients with early-stage PBC treated with ursodiol have a normal life expectancy for up to 20 years after diagnosis. Survival of patients with histologically more advanced PBC, stages III and IV, is not improved with ursodiol. There is general agreement that all patients with PBC should be treated with ursodiol, but there is no consensus about the treatment of patients who do not respond fully to ursodiol. The author has found that the addition of colchicine and methotrexate to ursodiol in those patients who failed to respond fully to ursodiol improved symptoms, biochemical tests of liver function, and liver histology in 80% of nonresponders.

Epidemiology

PBC was once considered a rare disease. However, increased awareness of PBC has led to earlier and more frequent diagnoses. In some series, it accounted for almost 2% of deaths due to cirrhosis. In a study from England, there was an annual incidence of 5.8 cases per million population and a point prevalence of 54 cases per million. The prevalence in the US is at least that high. Estimates place the number between 50 000 and 100 000 patients. The etiology is unknown but is most likely related to altered immunoregulation. The high but not total concordance of PBC in identical twins suggests that some triggering event is required to initiate PBC in a genetically

susceptible individual. The prevalence of PBC in families with one affected member is estimated to be 1000 times greater than that in the general population. A close association of PBC with genetic variants at the human leukocyte antigen (HLA) class II interleukin (IL)12A and IL12RB2 loci has recently been reported. A number of environmental causes have been implicated in the pathogenesis of PBC, including bacteria, viruses, smoking, and exposure to halogenated hydrocarbons such as those found in some cosmetics. Molecular mimicry has been the most widely proposed mechanism for the initiation of PBC. The most robust data supporting a role for molecular mimicry implicate a unique ubiquitous, xenobiotic-metabolizing gram negative bacterium, *Novosphingobium aromaticivorans*, that is found in soil throughout the world. This bacterium has four lipoyl domains with homology strikingly similar to the human lipoylated autoantigens that have been implicated in the pathogenesis of PBC. These lipoyl domains can be found by polymerase chain reaction (PCR) in approximately 20% of all individuals, in both normals and those with PBC, and are capable of metabolizing environmental estrogens to active estradiol. Almost 100% of patients with antimitochondrial antibody (AMA)-positive PBC have very high antibody titers against the lipoyl domains of *N. aromaticivorans*. Some of the most compelling evidences for an environmental factor as a triggering factor in the etiology of PBC have been derived from epidemiological studies that demonstrated significant geographic clustering of PBC cases in urban areas in northeast England and near waste disposal sites in New York City.

Etiology

Autoantibodies, Antimitochondrial Antibody

PBC is considered to be a model autoimmune disease. Of the patients with PBC, 90–95% are women, a sex distribution similar to that of other autoimmune diseases such as systemic lupus erythematosus. There is an increased incidence of

autoantibodies, including antimitochondrial antibodies, antinuclear antibodies, and antimicrobial antibodies. AMA is detected in 95% of patients with PBC and 98% of patients with a positive AMA have PBC. The major autoantibody found in PBC patients has been called anti-M2. It is principally directed against the dihydrolipoamide acyltransferase component (E2) of the branched keto-acid dehydrogenase and pyruvate dehydrogenase complexes on the inner mitochondrial membrane. Other antimitochondrial antibodies that have been described in PBC, anti-M4, anti-M8, and anti-M9 are most likely artifacts and were not found in a study that employed highly purified human mitochondrial proteins as antigens. The relationship between antimitochondrial antibodies and immunologic bile duct injury is not clear. The mitochondrial antigens are not tissue specific, and their intracellular location should reduce their immunogenicity. Repeated studies have shown that there is no correlation between the presence or titer of AMA and the severity or course of PBC.

Mitochondrial Antigens and the Cellular Immune System

Although there is little evidence that AMA plays a direct role in the pathogenesis of PBC, an interaction between the mitochondrial antigens and the cellular immune system appears to be important. The mitochondrial antigens, normally found in the inner mitochondrial membrane of all cells, are aberrantly expressed in the luminal surface of biliary epithelial cells only in patients with PBC and at an early stage of disease. CD4+ and CD8+ cells are present in high concentration in the portal triads of patients with PBC and often surround and infiltrate necrotic bile ducts. The E2 mitochondrial antigens stimulate interleukin-2 production by peripheral blood mononuclear cells and by T cells cloned from liver biopsies of PBC patients. E2 antigens are recognized by both helper and cytotoxic T cells in PBC. Cytotoxic T cells appear to mediate bile duct epithelial cell necrosis. Autoreactive T lymphocytes isolated from PBC patients are specific for PDC-E2. CD4+ cells that recognize this epitope are 100–150 times more concentrated in liver and regional lymph nodes compared to peripheral blood in patients with PBC while autoreactive CD8+ are 10–15 times more concentrated in liver compared to blood. These cloned CD8+ T cells produce interferon (IFN)- γ in response to PDC-E2. The immunodominant epitope recognized by major histocompatibility complex (MHC) class II-restricted CD4+ T cells in patients with PBC includes amino acids 163–176 of PDC-E2. The immunodominant epitope recognized by MHC class I-restricted CD8+ T cells is similar, amino acids 159–167 of PDC-E2. This region of PDC-E2 also overlaps with the region to which AMAs bind. These data suggest that bile duct epithelial cells expressing this epitope are targeted by CD4+ and CD8+ T lymphocytes in patients with PBC.

Why bile duct epithelial cells express this epitope is still unknown. One possibility is that immunoglobulin IgA that recognizes PDC-E2 is complexed to it as IgA is transcytosed from the basal to apical side of the bile duct epithelial cell en route to its secretion in bile. Serum IgA has been found at mitochondrial surfaces in bile duct epithelial cells and

immune complexes containing PDC-E2, and IgA has been found in bile duct lumens in patients with PBC.

Association with Other Autoimmune Diseases

There is an increased association of PBC with other autoimmune diseases, such as thyroiditis, rheumatoid arthritis, the CREST syndrome (CREST stands for calcinosis, Raynaud's syndrome, esophageal dysmotility, sclerodactyly, telangiectasia), Raynaud's disease, autoimmune chronic hepatitis, and Sjögren's syndrome. Although circulating immune complexes have been demonstrated in some PBC patients, it is unlikely that they play a role in pathogenesis. Only a small percentage of PBC patients have immune complexes. Some PBC patients have an IgM which can fix complement in the absence of demonstrable antigen binding and will give false positive results with many of the assays used to detect immune complexes.

Pathophysiology

The signs and symptoms of PBC are due to the long-standing cholestasis. There is a gradual destruction of small intrahepatic bile ducts that results in a decreased number of functioning bile ducts within the liver. This, in turn, causes retention within the liver of substances that are normally secreted or excreted into bile, such as bile acids, bilirubin, and copper. The increased concentration of some of these substances, for example, bile acids, causes further damage to liver cells. Other substances regurgitate from the liver into blood and soft tissues and cause symptoms such as pruritus. The identity of the substance or substances that cause pruritus is unknown. It is not any of the naturally occurring primary and secondary bile acids. Most data suggest that it is some substance that is secreted into bile and that binds tightly to cholestyramine, a nonabsorbed, quaternary ammonium resin. Possible pruritogens include the naturally occurring opioids and lysophosphatidic acid. Serum levels of endogenous opioids (endorphins and enkephalins) are increased in PBC, and opioid antagonists will relieve itching in some patients. A recent study demonstrated that the activity of autotaxin, the enzyme that converts lysophosphatidylcholine to lysophosphatidic acid, was significantly and uniquely elevated in the sera of patients with cholestasis and pruritus and that lysophosphatidic acid induced scratching when injected into mice and stimulated isolated neurons *in vitro*. The striking hyperlipidemia and xanthoma formations in some PBC patients are also a consequence of long-standing cholestasis. The impaired secretion of bile in patients with clinically advanced PBC causes a diminished concentration of bile acids within the intestinal lumen. The bile acid concentration may fall below the critical micellar concentration and thus be inadequate for complete digestion and absorption of neutral triglycerides in the diet. This causes the striking fat malabsorption seen in some jaundiced PBC patients as well as malabsorption of the fat-soluble vitamins A, D, E, and K. Some PBC patients with Sicca syndrome also have pancreatic insufficiency. The pathogenesis of the osteopenic bone disease that occurs in at least 25% of PBC patients is still unclear. Some data suggest that high concentrations of

serum bilirubin may impede osteoblast function. The clinically important bone disorder is osteoporosis, not osteomalacia, as was once believed.

Pathology

PBC is characterized by portal inflammation, the destruction of interlobular bile ducts, progressive fibrosis, and the eventual development of cirrhosis. PBC is divided into four histologic stages, I–IV. The disease is believed to progress from stage I to stage IV. However, the liver is not affected uniformly in PBC. One biopsy specimen may demonstrate lesions ranging from normal to stage IV. By convention, the histologic stage is the highest stage seen in a biopsy. There are two widely used staging systems, those of Scheuer and Ludwig, both well-known liver pathologists. In the Scheuer classification, stage I is defined by the presence of florid, asymmetric destructive bile duct lesions within portal triads (Figures 1(a) and 1(b)). These lesions are irregularly scattered throughout the portal triads and are often seen only on large surgical biopsies of the liver. The damaged bile ducts are often surrounded by dense collections of lymphocytes, histiocytes, plasma cells, eosinophils, and occasionally true giant cells, a pathologic lesion that is called a granuloma (Figure 1(a)). In the Ludwig classification, stage I is defined by the localization of inflammation to the portal triads. In stage II in the Scheuer classification, there are

reduced numbers of normal bile ducts within some portal triads and increased numbers of poorly formed bile ducts with irregularly shaped lumens in others, a lesion called atypical bile duct hyperplasia (Figure 2). Mononuclear cell inflammation occurs in portal triads and at the beginning of portal fibrosis. In stage II in the Ludwig classification, the inflammation and fibrosis extend beyond portal triads and into the surrounding parenchyma. A diminished number of bile ducts in an otherwise unremarkable appearing needle biopsy of the liver should alert one to the possibility of PBC. In stage III in both classifications, fibrous septa now extend beyond triads and form portal-to-portal bridges (Figure 3). The portal triads are otherwise similar to those in stage II. Stage IV represents the end stage of the lesion, frank cirrhosis, and regenerative nodules. It may be difficult to distinguish this late lesion from other types of cirrhosis. A paucity of normal bile ducts in areas of scarring suggests PBC (Figure 4).

Clinical Findings

The earliest and most common complaint is fatigue. Fatigue, however, is nonspecific and is found in many diseases. The earliest specific complaint in PBC is pruritus, usually worse at bedtime. In some patients, the pruritus begins during the third

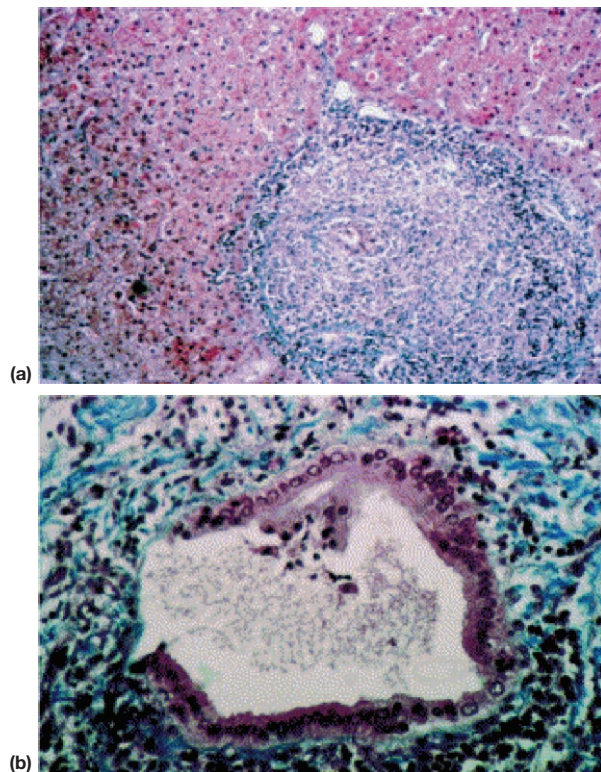


Figure 1 (a) Stage I florid bile duct lesion in primary biliary cirrhosis. A damaged bile duct is at the center of a granuloma (Masson trichrome, $\times 124$). (b) Stage I florid bile duct lesion in primary biliary cirrhosis. Epithelial cells have been dislodged from one quadrant of a bile duct (Masson trichrome, $\times 310$).

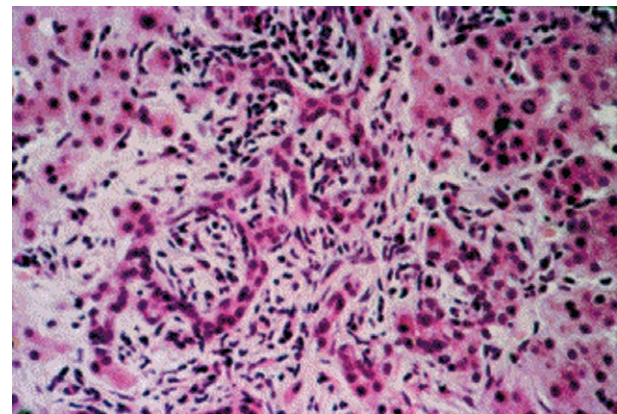


Figure 2 Stage II primary biliary cirrhosis. Atypical bile duct hyperplasia. Bile ducts are tortuous. None are cut in cross section and there are no visible bile duct lumens (hematoxylin and eosin, $\times 310$).

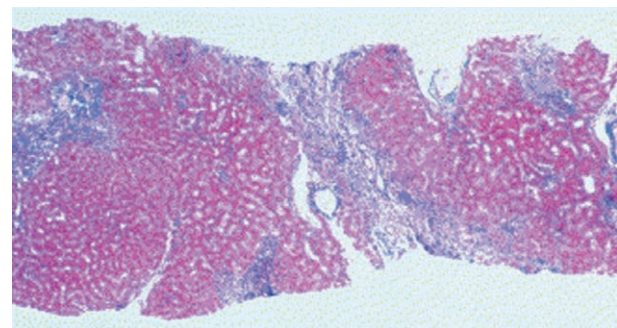


Figure 3 Stage III primary biliary cirrhosis. Adjacent portal triads are linked by septae that contain mononuclear inflammatory cells and scar tissue (Masson trichrome, $\times 310$).

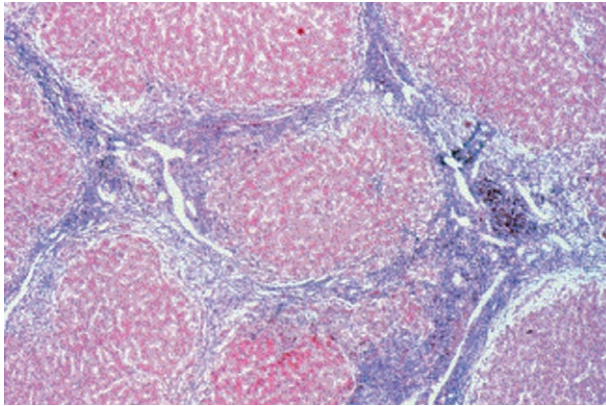


Figure 4 Stage IV primary biliary cirrhosis. Cirrhotic liver removed at the time of liver transplantation performed because of end-stage primary biliary cirrhosis. Nodule formation is evident and there are no visible bile ducts (Masson trichrome, $\times 54$).

trimester of pregnancy and persists after delivery. Physical examination may reveal hepatomegaly and increased skin pigmentation. The pigment is melanin, not bilirubin, at this early stage. Excoriations, caused by intense itching and an uncontrollable urge to scratch, may be widespread. Jaundice usually presents later in the course of the disease. Kayser–Fleischer rings have been observed in a few PBC patients with longstanding jaundice. Unexplained weight loss may also occur. Rarely, patients with PBC may go undetected until late in the course of the disease when they present with severe jaundice, ascites, or bleeding from esophageal varices. During the course of the disease, some patients with jaundice may develop malabsorption and complain of nocturnal diarrhea, frothy, bulky stools, or weight loss in the face of voracious appetite and increased caloric intake. Fewer than 5% of patients will develop xanthomas (a cutaneous tumor composed of lipid-laden foam cells) but xanthelasma (a planar xanthoma of the eyelids) are more common. Bone pain may occur in PBC patients with osteoporosis, as well as spontaneous collapses of vertebral bodies and hairline fractures of ribs. Fractures of long bones are less common.

Laboratory Tests

Liver function tests reveal a cholestatic pattern. The serum alkaline phosphatase and gamma glutamyl transpeptidase (GGT) are often strikingly elevated. The serum aminotransferases are less elevated and may be normal early in the course of PBC. Serum albumin and prothrombin times are characteristically normal early in the course of the disease while IgM levels are often elevated. The elevated alkaline phosphatase is usually of liver origin. 5'-Nucleotidase and GGT levels parallel the alkaline phosphatase. The serum bilirubin is normal early in the course of the disease but becomes elevated in most patients as the disease progresses. Both direct and indirect fractions are increased. Despite strikingly high serum cholesterol values, occasionally 800–1600 mg dl⁻¹, atherosclerosis is uncommon, most likely because serum high-density lipoprotein (HDL) cholesterol levels are also elevated. AMA is positive in 95% of patients. Serum ceruloplasmin is elevated, in contrast

to Wilson's disease, another disorder with increased retention of copper within the liver. There is an increased incidence of hypothyroidism in PBC that is not always reflected by routine tests of thyroid function. Elevated thyroid-stimulating hormone (TSH) levels will usually identify such individuals.

Diagnosis

The combination of a disproportionately elevated alkaline phosphatase and a positive AMA is sufficient evidence to make a diagnosis of PBC as long as the liver biopsy is consistent with this diagnosis. It is prudent to demonstrate that the bile ducts are patent. Ultrasonography is usually adequate for this purpose. Endoscopic cholangiography (ERCP) is recommended if there is a suspicion of bile duct obstruction on imaging tests or if the AMA is negative in a patient with inflammatory bowel disease. The ERCP is done in this setting to detect primary sclerosing cholangitis. A percutaneous needle biopsy of the liver will confirm the diagnosis, allow histologic staging, and provide a baseline for follow-up biopsies that are done to evaluate the efficacy of medical treatment.

Treatment

Treatment of Symptoms

Pruritus

The nonabsorbed resin, cholestyramine, 4–8 g twice daily with meals, will relieve pruritus in most patients. Commonly prescribed anti-itch medicines such as antihistamines are only helpful early in the course of PBC when itching is not severe. Rifampin, 150 mg twice daily, will control itching in most patients who fail to respond to cholestyramine. Opioid antagonists, such as naloxone and naltrexone, may be effective in the small number of PBC patients who fail to respond to ammonium resins and rifampin.

Fatigue

We noted that modafinil (Provigil), 100 mg in the morning, relieved fatigue in five consecutive PBC patients who had incapacitating fatigue. This observation has now been confirmed in two larger observational studies. In our later study, 43 of 527 PBC patients (7%) had incapacitating fatigue and modafinil relieved fatigue in 70% of the 42 patients who took it. The effectiveness of modafinil appears to be durable in most patients who benefited from it.

Malabsorption

Some patients with PBC and the Sicca syndrome may have concomitant pancreatic insufficiency. This can be treated with pancreatic enzyme preparations given orally. Malabsorption of fat-soluble vitamins is irregular and unpredictable, but deficiencies of vitamins A, D, E, and K may occur in patients who are chronically jaundiced. Patients with low levels of a specific vitamin should be given supplements of it orally.

Anemia and osteoporosis

PBC patients may develop iron-deficiency anemia. This often reflects unrecognized gastrointestinal (GI) blood loss. Upper

endoscopy is indicated to detect esophageal varices or congestive gastropathy. If present, treatment with nonspecific beta blockers such as propranolol or nadolol is effective. Until recently, liver transplantation was the only proven effective treatment for osteoporosis. Now, bisphosphonates such as alendronate and risedronate together with supplemental calcium and vitamin D appear to be equally effective and are recommended in PBC patients who have osteopenia confirmed by bone density studies.

Treatment of the Underlying Disease Process

PBC, if untreated, is a progressive disease that eventually ends in liver failure and the need for liver transplantation. Median survival for symptomatic patients without treatment is approximately 7–10 years. Patients who are asymptomatic at the time of diagnosis have a longer life expectancy than symptomatic patients.

Medical treatment

UDCA, 12–15 mg kg⁻¹ body weight daily, is the only drug approved by the Food and Drug Administration (FDA) for the treatment of the PBC. It decreases serum levels of alkaline phosphatase, alanine aminotransferase (ALT), aspartate aminotransferase (AST), low-density lipoprotein cholesterol and IgM, reduces the risk of developing esophageal varices, and slows histologic worsening. It is safe and well tolerated. Colchicine, 0.6 mg twice daily, and methotrexate, 0.25 mg kg⁻¹ body weight per week in three divided doses given at 12-h intervals, have been effective in the author's experience when given to patients who do not respond adequately to UDCA. However, the efficacy of combination therapy (UDCA, colchicine, and methotrexate given together) has not been proved in randomized placebo-controlled trials.

Initiating treatment early in the course of PBC is important because medical therapy is less likely to help once cirrhosis and portal hypertension develop. Liver transplantation is the only effective treatment in patients with clinically advanced disease. The author's current approach is based on the observation that the response to medical treatment in PBC varies among patients. Treatment is initiated with UDCA. Colchicine is added if patients have not responded adequately to UDCA after 1 year. Methotrexate is added if patients have not responded fully to the combination of UDCA and colchicine after another year. An adequate response to treatment consists of resolution of pruritus, a decrease in serum alkaline phosphatase to values that are <50% above normal, and improvement in liver histology. Methotrexate is discontinued after 1 year if patients fail to respond to it. Patients are then maintained on UDCA and colchicine.

Liver transplantation

Liver transplantation has improved the survival of PBC patients with cirrhosis and liver failure. Patients who are candidates for liver transplantation are those with cirrhosis and (1) persistent

jaundice that is unresponsive to medical treatment, (2) ascites, (3) bleeding from esophageal varices, (4) hepatic encephalopathy, and/or (5) intractable pruritus. One-year survival after liver transplantation is greater than 90% in most transplant centers. Recurrence of PBC after liver transplantation occurs in about 10% of patients after 5 years and in 20% of patients after 10 years. Recurrence is more common if tacrolimus is used as an immunosuppressive agent rather than cyclosporine. We have observed that the treatment with UDCA and colchicine has decreased serum levels of bilirubin, alkaline phosphatase, ALT, and AST and has improved liver histology in patients with recurrent PBC.

See also: **Bioenergetics:** Mitochondrial Auto-Antibodies;
Metabolism Vitamins and Hormones: Bile Salts.

Further Reading

- Bonis PAL and Kaplan M (1999) Methotrexate in primary biliary cirrhosis unresponsive to ursodeoxycholic acid: An observational study in 10 patients. *Gastroenterology* 117: 395–399.
- Corpechot C, Carrat F, Bahr A, et al. (2005) The effect of ursodeoxycholic acid therapy on the natural course of primary biliary cirrhosis. *Gastroenterology* 128: 297–303.
- Gan SJ, de Jongh M, and Kaplan MM (2009) Modafinil in the treatment of debilitating fatigue in primary biliary cirrhosis: A clinical experience. *Digestive Diseases and Sciences* 54: 2242–2246.
- Gershwin ME, Ansari AA, Mackay IR, et al. (2000) Primary biliary cirrhosis: An orchestrated immune response against epithelial cells. *Immunological Reviews* 174: 210–225.
- Kaplan MM, Bonder A, Ruthazer R, and Bonis PAL (2010) Methotrexate in patients with primary biliary cirrhosis who respond incompletely to treatment with ursodeoxycholic acid. *Digestive Diseases and Sciences* 55: 3207–3217.
- Kaplan MM, DeLellis R, and Wolfe H (1997) Sustained biochemical and histological remission of primary biliary cirrhosis in response to medical treatment. *Annals of Internal Medicine* 126: 682–688.
- Kaplan MM and Gershwin ME (2005) Primary biliary cirrhosis. *New England Journal of Medicine* 353: 1261–1273.
- Kita H, Matsumura S, He XS, et al. (2002) Quantitative and functional analysis of PDC-E2 specific autoreactive cytotoxic T lymphocytes in primary biliary cirrhosis. *Journal of Clinical Investigation* 109: 1231–1240.
- Kremer AE, Marten JJ, Kulik W, et al. (2009) Increased serum autotaxin activity in cholestatic pruritus. *Hepatology* 50: 376A.
- Lindor K (2007) Ursodeoxycholic acid for the treatment of primary biliary cirrhosis. *New England Journal of Medicine* 357: 1524–1529.
- Lindor KD, Gershwin ME, Poupon R, et al. (2009) AASLD guidelines: Primary biliary cirrhosis. *Hepatology* 50: 291–308.
- Pares A, Caballeris L, and Rodes J (2006) Excellent long-term survival in patients with primary biliary cirrhosis and biochemical response to ursodeoxycholic acid. *Gastroenterology* 130: 715–720.
- Poupon RE, Lindor KD, Cauch-Dudek K, et al. (1997) Combined analysis of randomized controlled trials of ursodeoxycholic acid in primary biliary cirrhosis. *Gastroenterology* 113: 884–890.
- Prince M, Chetwynd A, Newman W, Metcalf JV, and James OFW (2002) Patient survival and symptom progression in a large geographically based cohort of patients with PBC: Follow up for up to 28 years. *Gastroenterology* 123: 1044–1051.
- Selmi C, Balkwill DL, Invernizzi P, et al. (2003) Patients with primary biliary cirrhosis react against a ubiquitous xenobiotic-metabolizing bacterium. *Hepatology* 38: 1250–1257.

Prions of Yeast and Fungi: Proteins Acting as Genes

R B Wickner, National Institutes of Health, Bethesda, MD, USA

Published by Elsevier Inc.

Glossary

Amyloid A filamentous form that many proteins can assume, characterized by high β -sheet content, protease-resistance, and yellow-green birefringence on staining with Congo red.

Chaperone A protein that aids the folding (or refolding or renaturation) of another protein.

Prion A protein that transmits a trait or disease without the need for an accompanying nucleic acid.

Prion domain Part of a protein that determines the prion properties of the full protein.

Yeast and Fungal Prions: Genetic Criteria

The word 'prion' means an infectious protein, by whatever mechanism. This idea arose from studies of the transmissible spongiform encephalopathies (TSEs) of mammals, believed to be due to an infectious form of a protein called 'PrP'. Two nonchromosomal genes of *Saccharomyces cerevisiae*, called [URE3] and [PSI⁺], were identified as prions, because their unusual genetic properties were inconsistent with their being nucleic acid replicons, but were just what one would expect for a prion (Figure 1). First, each could be efficiently eliminated (cured) from the cells, but, from the cured cells, could again arise, at low-frequency, cells carrying the nonchromosomal gene. Second, overproduction of the normal form of the protein was shown to increase the frequency with which the prion form arose *de novo*. In addition, it was the protein whose overproduction induced the appearance of the nonchromosomal gene, neither the RNA nor the gene in high copy number. Finally, the chromosomal gene encoding the protein was, of course, necessary for the propagation of the prion, but the phenotype of mutants in the chromosomal gene was the same as that in the presence of the prion.

Known Prions of Yeasts and Fungi

[URE3] has each of these properties as a prion of the Ure2 protein, a regulator of nitrogen catabolism in yeast. [URE3] cells are inappropriately derepressed for the enzymes and transporters needed for the assimilation of poor nitrogen sources. Similarly, [PSI] has the properties expected of a prion of the Sup35 protein, a subunit of the translation termination factor. [PSI⁺] cells show increased rates of read-through of termination codons (nonsense suppression). A prion of *Podospora anserina*, called [Het-s], is unique in performing a function for this fungus, being required for an antiviral defense mechanism called 'heterokaryon incompatibility'. A third *S. cerevisiae* prion, called [PIN⁺], was discovered by its ability to promote the *de novo* development of the [PSI⁺] prion and is an altered form of Rnq1p. While [URE3] and [PSI⁺] make their presence known by the absence of the normal form of Ure2p and Sup35p, respectively, both [Het-s] and [PIN⁺] produce a phenotype by the presence of the abnormal forms of HET-s and Rnq1p, respectively, similar to the case with the mammalian

TSEs. Swi1p and Cyc8p, two transcription regulatory proteins, were suspected to be capable of prion formation, because, not only do they have Q/N-rich domains similar to the prion domains of Ure2p and Sup35p, but the overproduction of either protein mimics the Pin ([PSI]-inducibility) effect of the [PIN] prion, resulting in the rare seeding of Sup35p to form the [PSI] prion. Mot3p was found to form a prion by studying proteins with Q/N rich domains. [ISP⁺] is an intranuclear prion of the transcriptional regulator Sfp1. It was then found that the full-length Mca1 protein, a metacaspase homolog, could be a prion. Careful studies showed that each could indeed form a prion as judged by the three genetic criteria discussed above.

Prions do not have to be based on self-propagating amyloids. The [BETA] prion is simply the active form of yeast vacuolar protease B (Prb1p). This protease is synthesized as an inactive precursor protein, and under certain circumstances, its active form is required for the processing of the inactive precursor to the active form by proteolytic cleavage. While having no connection with amyloid, this prion has all of the same genetic properties and is indeed an infectious protein. The crippled growth non-Mendelian genetic element of *P. anserina* may likewise be a prion based on self-action of a set of MAP kinases.

Prion Domains and Functional Domains

In Ure2p and Sup35p prion proteins, an N-terminal segment is responsible for the prion properties of the protein, while the normal functions of the protein are carried out by both the N-terminal and C-terminal domains (Figure 2). The prion domain can propagate the prion completely independent of the remainder of the protein, and can even confer prion properties if fused to unrelated reporter protein. The prion domain of Ure2p is necessary for the stability of the protein against degradation, while the C-terminal domain is the nitrogen regulation domain. The Sup35p prion domain functions normally in the degradation of the 3' polyA structure of all messenger RNAs (mRNAs), and is thus necessary for normal mRNA turnover, while the C-terminal domain functions with Sup45p in translation termination. The HET-s protein, Rnq1p, Swi1p, and Cyc8p also have special domains responsible for their prion properties.

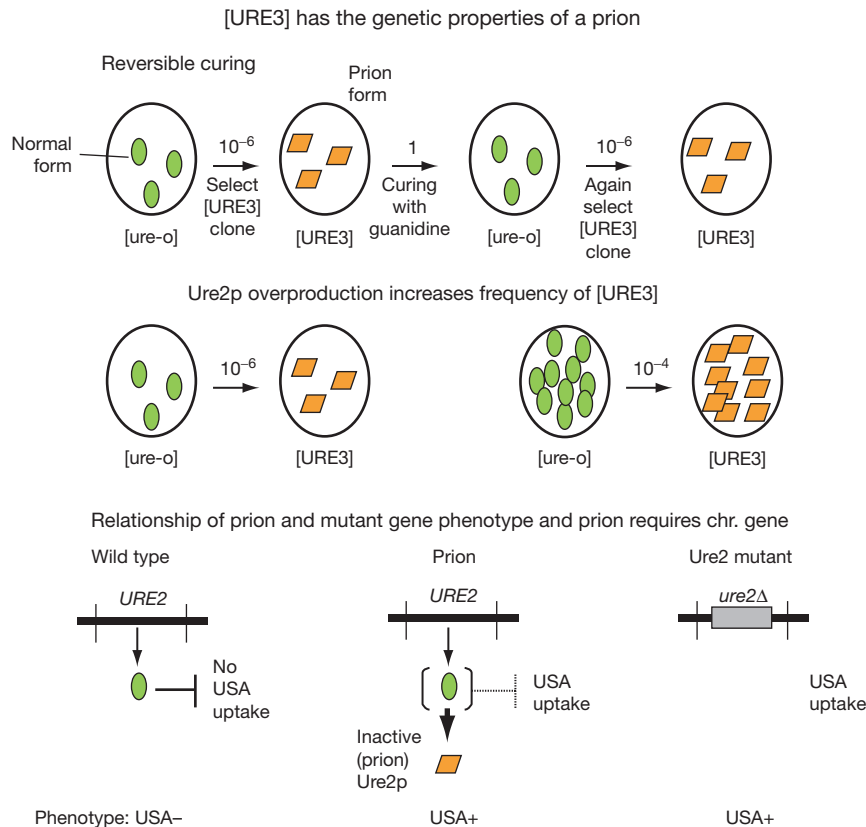


Figure 1 Genetic properties identify prions. Cells cured of a prion can give rise to cells again carrying the prion (top). Overproduction of the prion protein increases the frequency with which the prion arises (middle). A mutant in the prion protein gene can have a similar phenotype to the presence of the prion because both lack the normal form of the protein (bottom). Here, the phenotype of prion or mutant is the ability to take up ureidosuccinic acid (USA), because Ure2p normally inhibits transcription of *DAL5*, which encodes a transporter capable of taking up USA. Phenotypes are shown at the bottom as USA+, or USA-.

Except for HET-s and PrP, the prion domains are all rich in asparagine or glutamine residues, and these runs of N or Q are important for the prion properties of the proteins. However, neither the HET-s protein nor mammalian PrP have regions rich in N or Q (Figure 2). The N-, Q-rich prion domains of Ure2p, Sup35p, and Rnq1 resemble the polyQ of Huntingtin whose amyloid formation is the central event in Huntington's disease.

Yeast, Fungal Prions [PSI⁺], [URE3], [PIN⁺], and [Het-s] Are Self-Propagating Amyloidoses

Amyloid is a special form of protein structure that is characteristic of a number of human diseases. Amyloid is a filamentous structure high in β -sheet content, with β -strands oriented perpendicular to the long axis of the filaments. Amyloid filaments display a characteristic yellow-green birefringence on staining with the aromatic dye Congo red and are partially resistant to protease digestion, probably reflecting their very stable β -sheet structure. Demonstration of aggregation is always simpler than that of amyloid structure, but it is evident that not all aggregates are amyloid. Amyloid formation by the A β peptide is the central event in Alzheimer's disease,

while amyloid of amylin is a prominent feature of late-onset diabetes. Parkinson's disease features intracellular amyloid formation by α -synuclein, and amyloid of serum amyloid A protein is a complication of many long-term infectious and inflammatory processes.

Sup35p, Ure2p, and HET-s each form amyloid *in vitro*. Moreover, the characteristics of the amyloid formed and the requirements for its formation closely reflect the corresponding properties and conditions of prion formation *in vivo*. For example, extracts of [PSI⁺] cells can prime the formation of amyloid by Sup35p much better than can extracts of normal cells. The prion domain of Ure2p induces amyloid formation by the full-length native Ure2p *in vitro* just as the prion domain produced *in vivo* induces *de novo* generation of the [URE3] prion. Indeed, filaments of Ure2p and Sup35p have been visualized specifically in [URE3] and [PSI⁺] cells, respectively. Most important, the amyloid filaments formed *in vitro* from recombinant HET-s, Sup35, Ure2, or Rnq1 proteins are highly infectious for yeast, transmitting the corresponding prion to the cells.

The mechanism by which amyloid formation inactivates Ure2p is apparently not by a conformational change of the nitrogen regulation domain of the molecule. For example, the glutaredoxin activity of the nitrogen regulation domain is unaffected by amyloid formation, suggesting that being

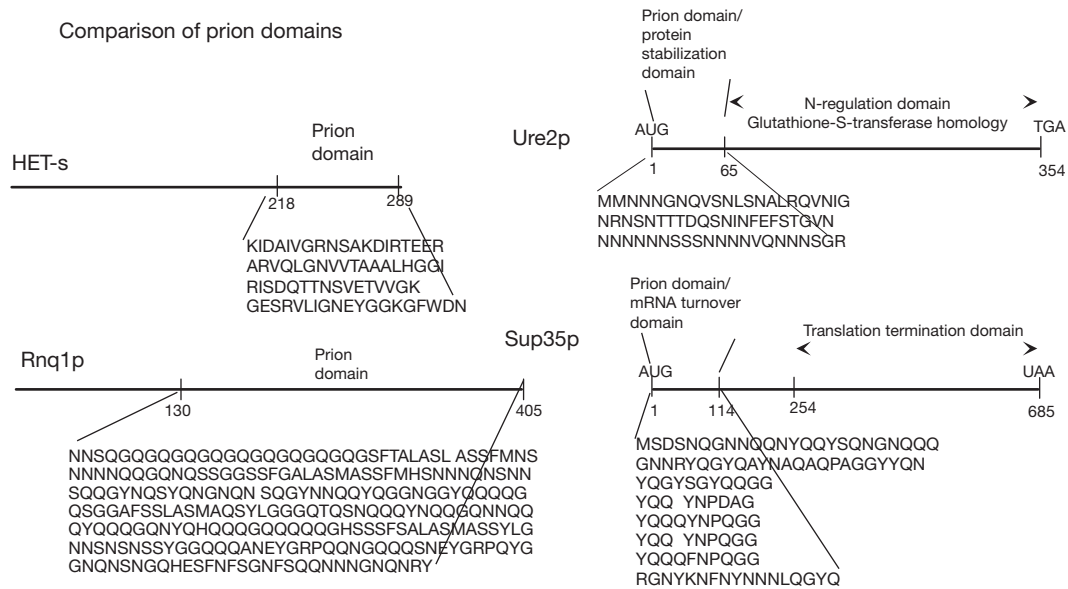


Figure 2 Prion domains of Ure2p, Sup35p, Rnq1p, and HET-s. While the prion domains of Ure2p, Sup35p, and Rnq1p are rich in N and Q, that of HET-s is rich in neither. Normal functions of the prion domains of Sup35p and Ure2p have been identified.

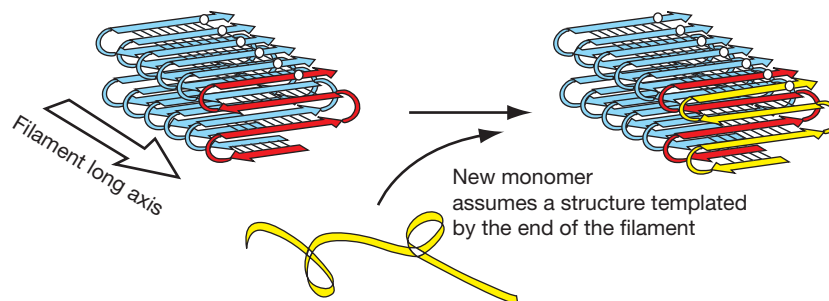


Figure 3 The in-register parallel β -sheet architecture of the prion domains of Sup35p, Ure2p, and Rnq1p. Each residue forms a line along the long axis of the filaments (white circles may be, e.g., Val35 of each molecule). Such rows of Asn or Gln residues form a line of peptide-like hydrogen bonds (the ‘beta zipper’). A row of Ser or Thr residues can form inter-residue bonds as well. A row of Leu, Val, Ile, Phe, Tyr, or Trp residues can have hydrophobic interactions along the fiber. These interactions force the new monomer (yellow) to assume the same structure as the molecule on the end of the fiber (red). This templating results in inheritance of structure, just as DNA templating results in inheritance of sequence.

in the amyloid filament sterically blocks access to binding Gln3p (the target of Ure2p action) or imposes a diffusion limitation.

The structures of the amyloid filaments have been analyzed using solid-state nuclear magnetic resonance (NMR). Amyloid of the HET-s protein is a β -helix, with each molecule making two turns of the helix. The Sup35, Ure2, and Rnq1 prion domains each form amyloids with an in-register parallel β -sheet architecture (Figure 3). This in-register parallel β -sheet architecture can explain how a single protein sequence can transmit its conformation to another molecule, and how any of several such conformations (prion variants) can be stably propagated by the same protein sequence. The sheet is folded along the long axis of the filament, and it is proposed that different locations of the folds, or different extents of the beta sheet structure along the sequence, can produce different variants. The in-register parallel β -sheet architecture allows proteins to act as a template to transmit their

structure, in analogy to the ability of DNA strands to template their sequence.

[PSI] and [URE3] Are Diseases, but [HET-s] Carries Out a Normal Function

Amyloidoses of humans are generally associated with disease states, and the TSEs, in particular, are uniformly fatal. However, the Pmel17 melanosome protein forms amyloid filaments that are involved in melanin biogenesis, and amyloids on the surface of bacterial and fungal cells function in adhesion and cell protection. The [URE3] prion is associated with slow growth under most culture conditions. [PSI⁺] and [PIN⁺] do not confer any consistent phenotype on otherwise normal cells, although over half of [PSI⁺] variants make cells very sick or dead. Searches by three groups for [PSI⁺] in wild strains

have failed to find it, supporting the notion that it is a disease, like the mammalian counterparts. The Ure2 and Sup35 prion domains have normal functions unrelated to prion formation, indicating that they are not conserved for the purpose of forming a prion. Indeed, prion-forming ability is not conserved, even within *S. cerevisiae*, where a quarter of wild isolates in one study have a large deletion in their Sup35 prion domain such that they cannot form a prion. Furthermore, one species of *Saccharomyces* has a Ure2p that also cannot form prions. The prion domains of Sup35 and Ure2 of *Saccharomyces* species have diverged much more rapidly than have the other parts of these molecules. Because prion propagation requires self-templating of peptides with identical sequence, this divergence has the effect of partially blocking prion transmission, and, as has been suggested for the M/V polymorphism at residue 129 of human PrP, this divergence may have been selected because it provides protection from the harmful effects of catching a prion.

A colony of a filamentous fungus is not a pile of cells (as in yeast), but rather a syncytium – cells connected by cytoplasmic bridges. When two compatible colonies of *Podospora* or other filamentous fungi grow together, they fuse cellular processes (hyphae) thereby joining the two colonies into one, called a ‘heterokaryon’. This allows the colonies to share nutrients, but has the danger that a virus infecting one will spread by cytoplasmic mixing throughout the other. To minimize this risk, filamentous fungi have a system, called ‘heterokaryon incompatibility’, which tests for identity between the partners at about a dozen polymorphic chromosomal loci. Nonidentity of alleles at a single locus is sufficient to abort the fusion process and a barrier to further fusion is formed. Most of these *het* loci are genetically unremarkable. The *het-s* locus of *Podospora* can have either of two alleles, *het-s* or *het-S*. A colony with the *het-s* allele is only properly incompatible with a *het-S* colony if the HET-s protein is in an amyloid form. This amyloid form is infectious and is the [Het-s] prion. The HET-S protein, which differs by only 14 amino acid residues from the HET-s protein, does not form amyloid, and residue 33 is critical in this difference.

Thus, the [Het-s] prion is the (so far) unique case of a prion that is necessary for a normal function, heterokaryon incompatibility. Consistent with this view is the fact that most wild strains of *P. anserina* with the *het-s* allele carry the [Het-s] prion. However, [Het-s] also mediates a meiotic drive system, in which the *het-s* allele is preferentially inherited in *het-s* X *het-S* crosses as a result of lethality of *het-S* [Het-s] spores. Thus, the [Het-s] prion may ultimately prove to be a disease like the yeast prions.

Prion Generation

[PIN⁺] was discovered as a nonchromosomal gene whose presence was necessary for the efficient induction of the *de novo* appearance of [PSI⁺] by overexpression of Sup35p. [PIN⁺] was found to be a prion of the Rnq1 protein, which had already been shown capable of a self-propagating aggregation *in vivo*. In fact, [URE3] is also capable of [PIN⁺]-like activity. This showed that one prion could promote the generation of another. This may be because all of these prions

([PSI⁺], [PIN⁺], and [URE3]) are based on amyloid of asparagine-glutamine rich segments. Although cross-priming of polymerization is not as efficient as self-priming (the usual propagation reaction), it is far more efficient than the truly spontaneous development of a prion. A similar, though weaker, effect of [PIN⁺] on [URE3] generation has also been observed.

Prion generation is also diminished by the Hsp70-group chaperones Ssb1p and Ssb2p. Both are ribosome associated, and are believed to help nascent peptides fold properly. It will be of great interest to understand the mechanisms by which Ssb’s block prion formation.

Chaperones and Prion Propagation

Chaperones play a central role in the propagation of yeast prions. Hsp104 is a disaggregating chaperone critical in the cellular reaction to heat shock, and either increasing or decreasing its level adversely affects propagation of [PSI⁺]. Amyloid aggregates are clumpy and may all remain in one of the daughter cells if they form a single lump in the mother cell. Modest levels of Hsp104 carry out the partial fragmentation of these filaments insuring that both daughter cells receive one or more seeds. Each of the other yeast amyloid-based prions likewise requires Hsp104 for propagation, but only [PSI⁺] is cured by Hsp104 overproduction. The mechanism of curing by Hsp104 overproduction is not known, but may be an effect on segregation of amyloid to daughter cells. Inhibition of Hsp104 by mM concentrations of guanidine cures all the amyloid-based yeast prions.

Different Hsp70s and Hsp40s have positive and negative functions with different prions. For example, overexpression of the Hsp40 protein, Ydj1, can cure [URE3] by its interaction with Hsp70s, and the Hsp40 Sis1p is needed by all prions tested. In spite of their close homology, one cytosolic Hsp70 promotes one prion while inhibiting another. Clearly, the detailed interactions of each prion with chaperones will be different, and this remains an important area to investigate. There is circumstantial evidence for a role of chaperones in the mammalian TSEs, but direct evidence has so far been limited to the yeast systems.

The importance of the discovery of the role of chaperones in yeast prions goes far beyond these systems and even the mammalian prion diseases. Amyloidoses are common (almost universal) diseases of old age, and one suspects that one of the primary roles of the heat-shock proteins is to defend against amyloid formation. It will be critical to learn how these systems can be marshaled to contend with these otherwise largely intractable human amyloidoses. A screen for chemical agents that cure yeast prions has turned up two candidates that also cure tissue culture cells of scrapie. These two compounds act on a ribosomal chaperone activity. It may be possible to extend this approach to finding general anti-amyloid drugs.

Conclusions

The ability of proteins to bear hereditary information recalls the earlier discovery by Sonneborn of cortical inheritance in *Paramecium*, in which the surgically altered arrangement of

cilia on the surface of the cell was shown to be faithfully inherited by the progeny of the original cell. Self-propagating protein structures in the former case are closely analogous to self-propagating organellar structures in the latter. Efforts to find other self-propagating organellar structures have not yet met with any clear successes, and the mechanisms operating in the cortical inheritance phenomenon have not yet been elucidated.

There are doubtless many prions because evolution is not a finished job. As there are useful (non-prion) amyloids, there are likely to be useful prions (perhaps [Het-s]), but which, if any, is not yet certain. Retroelements are selfish parasites, but they are DNA, and cells may occasionally be selected in which a particular retroelement has been harnessed to do the cell's work. This does not diminish their primary role as parasites.

Most of the prions discussed in this article appear to be based on the self-propagating amyloid of the corresponding protein. Of course, most amyloidoses are not prions, although nearly all have been shown to be self-seeding *in vitro*. Prion propagation requires not only expansion of amyloid within one individual cell or compartment, but also a mode of transmission. It is clear in several systems that fully formed amyloid is the infectious protein form, refuting suggestions that only some oligomer intermediate is the infectious material, although it is evident that the shorter the fiber, the higher the ratio of growing ends to mass of material.

The direct parallels between these yeast, fungal, and mammalian prions and amyloidoses have opened new areas for investigation of these very difficult human diseases.

See also: [Protein/Enzyme Structure Function and Degradation: Amyloid](#); [Prions Overview](#).

Further Reading

- Chernoff YO, Lindquist SL, Ono B-I, Inge-Vechtomov SG, and Liebman SW (1995) Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi+]. *Science* 268: 880–884.
- Du Z, Park K-W, Yu H, Fan Q, and Li L (2008) Newly identified prion linked to the chromatin-remodeling factor Swi1 in *Saccharomyces cerevisiae*. *Nature Genetics* 40: 460–465.
- Kushnirov VV and Ter-Avanesyan MD (1998) Structure and replication of yeast prions. *Cell* 94: 13–16.
- Liebman SW and Derkatch IL (1999) The yeast [PSI+] prion: Making sense out of nonsense. *Journal of Biological Chemistry* 274: 1181–1184.
- Masison DC and Wickner RB (1995) Prion-inducing domain of yeast Ure2p and protease resistance of Ure2p in prion-containing cells. *Science* 270: 93–95.
- McGlinchey R, Kryndushkin D, and Wickner RB (2011) Suicidal [PSI+] is a lethal yeast prion. *Proceedings of the National Academy of Sciences of the United States of America* 108: 5337–5341.
- Patel BK, Gavin-Smyth J, and Liebman SW (2009) The yeast global transcriptional co-repressor protein Cyc8 can propagate as a prion. *Nature Cell Biology* 11: 344–349.
- Roberts BT and Wickner RB (2003) Heritable activity: A prion that propagates by covalent autoactivation. *Genes and Development* 17: 2083–2087.
- Saupe SJ (2000) Molecular genetics of heterokaryon incompatibility in filamentous ascomycetes. *Microbiology and Molecular Biology Reviews* 64: 489–502.
- Sharma D and Masison DC (2009) Hsp70 structure, function, regulation and influence on yeast prions. *Protein and Peptide Letters* 16: 571–581.
- Shewmaker F, Wickner RB, and Tycko R (2006) Amyloid of the prion domain of Sup35p has an in-register parallel β -sheet structure. *Proceedings of the National Academy of Sciences of the United States of America* 103: 19754–19759.
- Tycko R (2006) Molecular structure of amyloid fibrils: Insights from solid-state NMR. *Quarterly Reviews of Biophysics* 1: 1–55.
- Wickner RB (1994) Evidence for a prion analog in *Saccharomyces cerevisiae*: The [URE3] non-Mendelian genetic element as an altered URE2 protein. *Science* 264: 566–569.
- Wickner RB, Liebman SW, and Saupe SJ (2004) Prions of yeast and filamentous fungi: [URE3], [PSI+], [PIN+], and [Het-s]. In: Prusiner SB (ed.) *Prion Biology and Diseases*, pp. 305–372. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Wickner RB, Shewmaker F, Kryndushkin D, and Edsles HK (2008) Protein inheritance (prions) based on parallel in-register β -sheet amyloid structures. *BioEssays* 30: 955–964.

Prions Overview

O Bannach and D Riesner, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Cedric Govaerts and Fred E. Cohen, volume 3, pp. 437–440, © 2004, Elsevier Inc.

The Prion as an Unprecedented Agent

The discovery of prions was inseminated by the research on the agent of the scrapie disease in sheep. Although transmissibility of the scrapie agent was suggested already 250 years ago, it was shown by inoculation experiments in sheep not until 1936 by J. Cuiil  and P. J. Chelle. Searching for a virus, however, did not lead to the detection of typical viral features; instead, highly enigmatic properties of the agent were demonstrated, such as the absence of particles in the electron microscope, no immune response during the infection, and high resistance of the agent against chemical and physical treatment. As early as 1966, Tikvah Alper and her colleagues concluded from inactivation studies using ionizing and ultraviolet (UV)-irradiation that the target size of the scrapie infectious agent of 50–150 kDa is too small for a virus, and the inactivation spectrum is more characteristic of a protein. In the following, Stanley B. Prusiner performed systematic analyses using not only chemical and physical but also enzymatic procedures. He summarized the results in two groups: (1) procedures which modify or destroy nucleic acids do not abolish scrapie infectivity and (2) procedures which modify or destroy proteins abolish infectivity. Many researchers discussed the scrapie agent as an unconventional virus but Prusiner came to the conclusion that the agent could not be a virus but a novel and unprecedented proteinaceous type of an agent, which he termed ‘prion’ (proteinaceous infectious particle).

As a major component of the prions a hydrophobic and, in mild detergents, insoluble protein of 33–35 kDa could be highly enriched. In a collaborative research project of the laboratories of Prusiner, L Hood, and Ch. Weissmann using the Syrian hamster as the experimental animal, the protein was cloned and shown to be a host-encoded protein, later called ‘prion protein (PrP)’. In a similar approach, Bruce W Chesebro and colleagues identified the corresponding protein in mice.

The prion protein from infectious material is partially resistant to degradation by proteinase K; it is N-terminally truncated at amino acid 89/90 leading to PrP 27–30, which retains full infectivity. The prion protein was identified as encoded in the host genome, and the protein was found indeed also in the noninfected organism. It is, however, very sensitive to proteinase K digestion, so that a clear biochemical distinction between two isoforms of the prion protein could be drawn: the cellular prion protein PrP^C in the noninfected organism, and the scrapie isoform PrP^{Sc} as the major component of the purified agent. It should be emphasized that proteinase K resistance of PrP was used as the first biochemical marker for infectivity, and often PrP^{Sc} and PrPres (for resistance) were used synonymously. Later, however, it was shown that infectivity and proteinase K resistance do not correlate in all cases; therefore, PrP^{Sc} is used for the property of infectivity and PrPres for proteinase K resistance, respectively.

Properties of the Two Isoforms of PrP

In addition to PK-resistance, more parameters were identified characteristically different for PrP^C and PrP^{Sc}. They are listed in [Table 1](#). Of particular relevance in the context of the scrapie history is the highly aggregated state of PrP^{Sc} because it is the physico-chemical basis of the resistance of prions to proteinase K degradation and deactivation procedures.

PrP was available in large amounts only when expressed recombinantly in *Escherichia coli*; when conditions were established (low pH) to solubilize PrP in high concentrations, the molecular structure of PrP in the noninfectious form could be solved by Kurt W  thrich and his co-workers applying high resolution nuclear magnetic resonance (NMR) spectroscopy. The structure is shown schematically in [Figure 1\(a\)](#). The C-terminal half of the molecule (residues 125–231) is folded in a well-defined globular structure comprising three α -helices and a short antiparallel β -strand, whereas the N-terminal half is more or less flexible. The globular part of the structure was confirmed by X-ray crystallography. Under more physiological conditions like neutral pH, low concentration of lipids and detergents, or membrane attachment, PrP tends to form dimers with additional secondary structure of the amino acids 90–125.

The insolubility of PrP^{Sc} prevented so far its structural determination with molecular resolution. A combination of molecular dynamic calculations, fiber diffraction studies, and biochemical data from chemical cross-linking, ligand accessibility, etc., led to a structural model (cf. [Figures 1\(b\)–1\(d\)](#)) which, however, should be regarded as preliminary. The model clearly reflects the fibrillar structure as known from electron microscopy of prion rods, and the intermolecular interactions are established by stacks of parallel β -helices.

The Prion Model

The concept of prions for scrapie infectivity was under debate for many years, and the unusual resistance against deactivation was interpreted by many researchers in terms of unconventional viruses or so-called virinos possessing only a very small not yet detected nucleic acid. Quantitative analysis by Prusiner, Riesner, and co-workers of nucleic acids possibly essential for scrapie infectivity resulted in the determination of a limit, that is, that less than one nucleic acid molecule with more than 20 nucleotides was present in one infectious unit; consequently viral or viroid-like nucleic acids could be excluded from being crucial for scrapie infectivity. Also, the direct evidence for the proteinaceous nature of the agent was broadened continuously during the last 20 years and particularly the generation of synthetic prions (cf. below) can be regarded as an experimental proof so that we consider at present only the prion concept as

valid for scrapie infectivity. A simple form of prion amplification is depicted in **Figure 2(a)**. PrP^C is expressed from a nuclear single copy gene and presented as a glycolipid-anchored protein on the outer cell membrane. If it comes into contact – in a direct or indirect manner – with PrP^{Sc}, it is enforced to switch over into the PrP^{Sc}-conformation. Breaking of the oligomeric or polymeric PrP^{Sc} generates more and new attachment and conversion sites for PrP^C leading to exponential growth of PrP^{Sc} subunits in prion particles. In summary, multiplication of prion infectivity is modeled by continuous synthesis of PrP^C and induced conformational conversion into PrP^{Sc}.

Table 1 Isoforms of the prion protein

Cellular isoform (PrP ^C)	Scrapie isoform (PrP ^{Sc})
Not infectious	Infectious
α -Helical	β -Sheet-rich
Monomer	Aggregated
Soluble (in mild detergents)	Insoluble
Proteinase K sensitive	Proteinase K resistant

Conformational Transitions and Synthetic Prions

The conformational transitions of PrP were studied in order to simulate the transition of the cellular isoform PrP^C into the pathogenic isoform PrP^{Sc}. Quite different conversion systems have been applied, among them lowering the pH, adding fairly high concentrations of denaturants like urea and GndHCl, or low concentrations of the detergent sodium dodecyl sulfate (SDS). In **Figure 3** the SDS conversion system is shown. By lowering the SDS concentration from 0.2%, which is still submicellar, to 0.01% or after washing out SDS, different conformational states were established; from monomeric partially denatured α -helical, via dimeric renatured α -helical and oligomeric β -sheet-rich to an highly aggregated β -sheet-rich but polymorphic state. By adding 0.15 M NaCl, an intermediate, partially denatured state can be established from either the dimeric α -helical or the oligomeric β -sheet-rich state. This intermediate state is essential because after several weeks of gently shaking, fibrils are formed spontaneously. If small amounts of prions are added as seeds, fibrils are formed within hours depending upon the concentration of prion seeds and PrP substrate. The kinetics can be described quantitatively

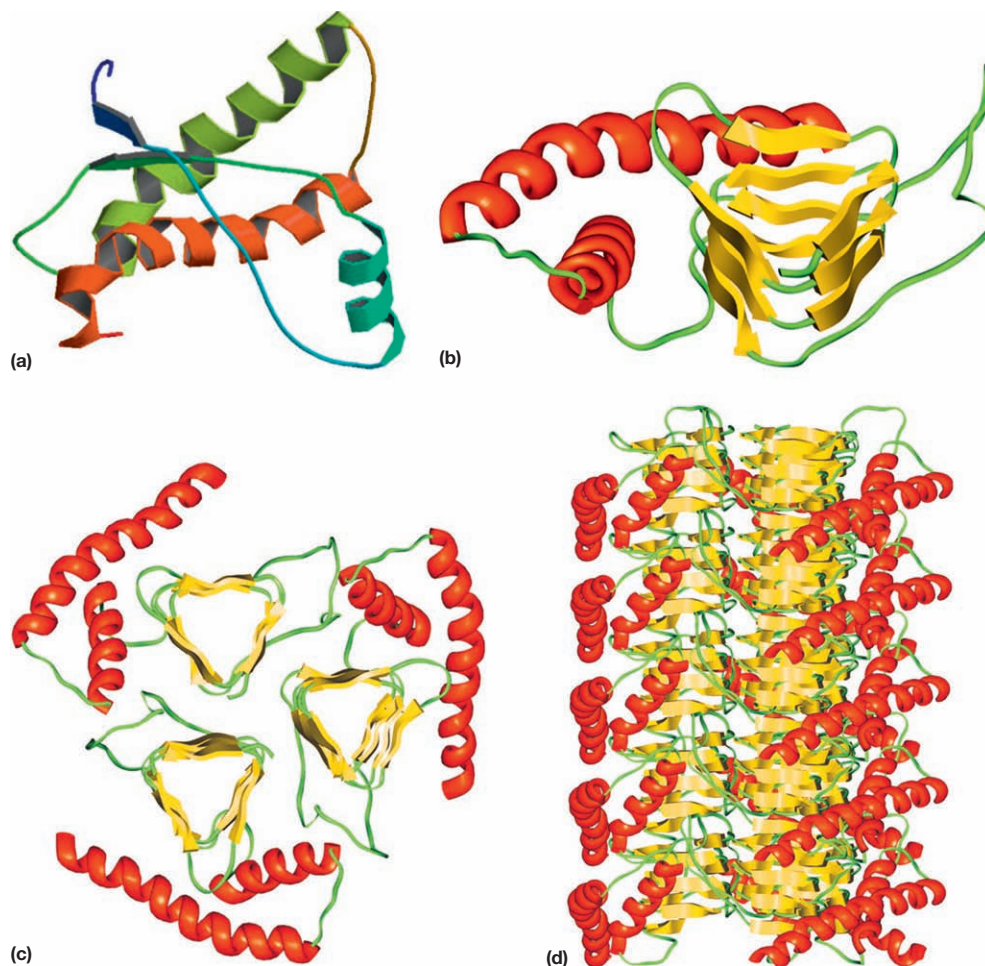


Figure 1 Structures of the prion protein. (a) Solution structure of the prion protein (90–231) from Syrian Goldhamster, (b) model for monomeric PrP^{Sc}, (c) trimeric PrP^{Sc}, and (d) PrP^{Sc} fibril.

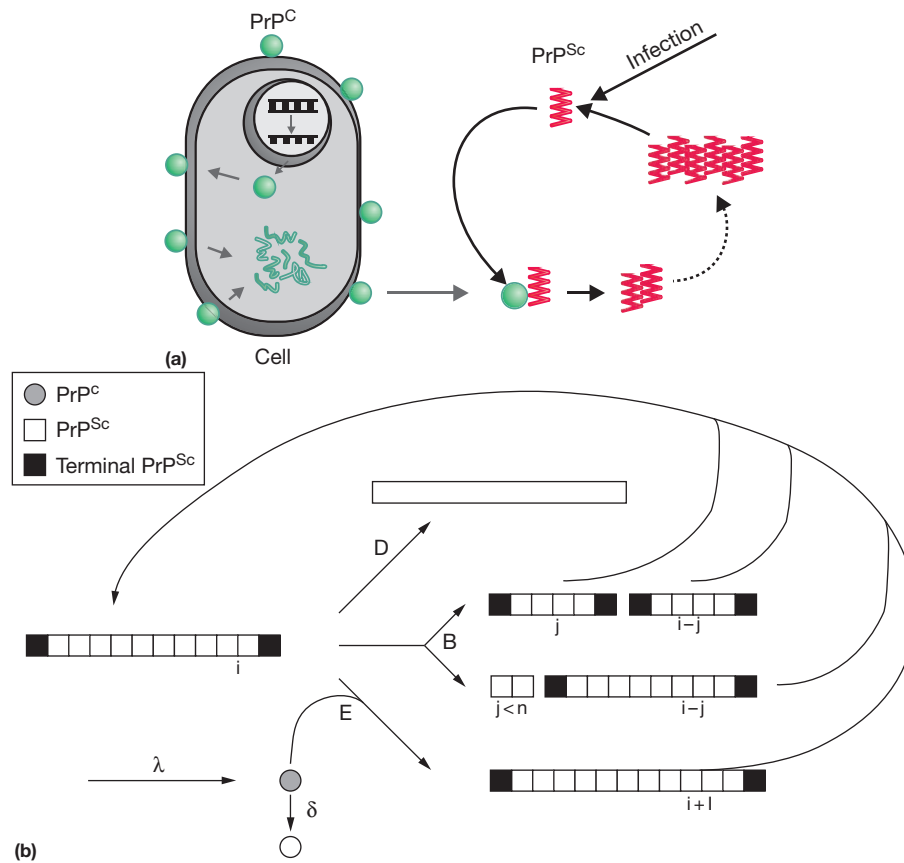


Figure 2 Models for prion propagation and amplification. (a) PrP^C is expressed from a single copy nuclear gene and attached via a glycolipid-anchor to the outer cell membrane. Upon contact with PrP^{Sc}, PrP^C adopts conformation of PrP^{Sc}. (b) Mechanistic model of seeded fibrilization. D, degradation; B, breakage; E, elongation; λ , synthesis of monomeric PrP; δ , degradation of monomeric; n , minimal size of an oligomer active as seed.

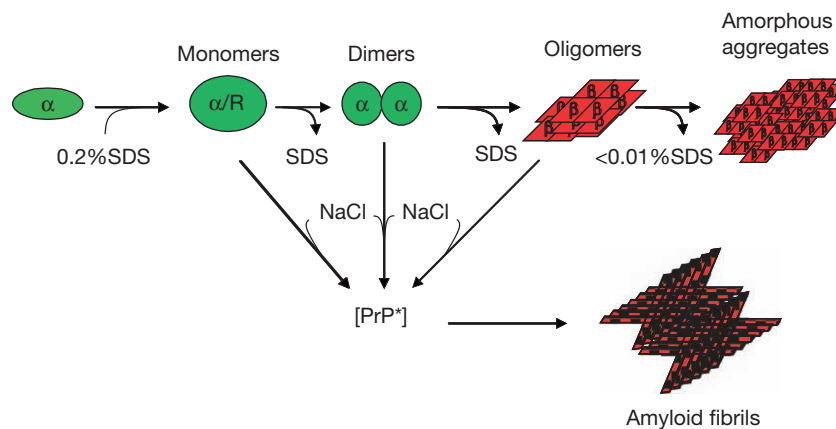


Figure 3 *In vitro* conversion of PrP. Dilution of SDS causes a structural shift from monomeric PrP into amorphous PrP aggregates. Addition of sodium chloride favors transition into a high-energy intermediate promoting fibril formation.

in the mechanistic model of Figure 2(b). It should be noted that the method of protein misfolding cyclic amplification (PMCA) from Claudio Soto and co-workers follows a similar mechanistic model with the main difference, that growing of fibrils and breaking of fibrils are established experimentally in different steps.

Generation of infectivity from PrP which was never in contact with an infected animal or – even better – synthesized recombinantly in *E. coli* would be the ultimate proof for the prion model. In Prusiner's group, fibrils with recombinant PrP were generated spontaneously *in vitro* and used as an inoculum for transgenic mice. The mice became sick after a long

incubation time showing that the infectivity of the fibrils was very low, but from these mice the disease could be transmitted to wild-type mice with shorter incubation times. The conclusion of the experiments, that is, to generate infectivity with fibrillar PrP *de novo*, has been confirmed in the meantime by other groups using different procedures. In summary, it is possible today to produce synthetic prions.

Prion Strains

Host specificity of prions is determined primarily by the sequence of *Prnp*. Interpreting the prion model in a simplistic manner, one might expect that after inoculating prions from a foreign host and passing the species barrier the properties of prions in the new host are characterized only by the *Prnp* sequence of the new host. That is evidently not the case; by contrast, different phenotypes, so-called prion strains, are observed in one and the same host species, that is, genotype of *Prnp*. Prion strains are characterized by individual incubation times, spatial distribution of brain lesions, and behavioral peculiarities. Biochemical and biophysical analyses revealed also significant differences in properties like cleavage sites by PK-digestion, insolubility, resistance against deactivation procedures, and others. All of these features are transmitted over many passages. The molecular basis of the strain phenomenon is not clear. A second information molecule like a nucleic acid has been postulated many times but was never found experimentally. Rather, one has to assume that several pathogenic conformations like PrP^{Sc}₁ or PrP^{Sc}₂, and even more exist and induce again PrP^{Sc}₁ or PrP^{Sc}₂, respectively, in the next passage. It is not known on what level of structure hierarchy the strain-specific structural features are encoded. Without assuming additional cellular factors, one might suggest that subtle differences in the secondary structure are stabilized by the tertiary and, in particular, by the quaternary structure which then would act again as a structure-specific seed.

Physiology and Pathophysiology of PrP and Prions

The physiological role of PrP^C may be manifold and depends upon the cell type where it is expressed. Activities reported in the literature are: copper binding due to the His-containing N-terminal repeats with the consequences of copper endocytosis and homeostasis, superoxide dismutase (SOD) and antioxidant activity, furthermore signaling in different pathways, adhesion in conjunction with neural cell adhesion molecule (NCAM), and anti-apoptotic activity.

The pathophysiological role of PrP^{Sc} or prion infection is generally described as neurodegeneration. Vacuolar degeneration of the gray matter with PrP^{Sc} deposition is observed which led to the name spongiform encephalopathy. The location of pathological changes can be the cerebral or the cerebellar cortex depending upon the host species and the prion strain.

Molecular models of the pathology have to take into account both PrP^C and PrP^{Sc}. The loss of function of PrP^C due to conversion into PrP^{Sc} is not very probable as primary pathological event because transgenic PrP^{0/0} mice, that is, with knocked out PrP gene, do not show symptoms. Overexpression

of PrP^C, however, can lead to disease although not a typical prion disease. Therefore, one has to assume a toxic effect of PrP^{Sc}, but at present it is not clear whether intra- or extracellular PrP^{Sc} deposition itself or its labile and soluble intermediates or even its interaction with PrP^C is the primary pathological event. The molecular and cellular pathways leading to neurodegeneration are present-day themes of active research and deserve particular attention because of the general relevance for protein misfolding diseases. Among the afflicted cellular activities are apoptosis, endosomal–lysosomal activity, complement activation, and synaptic and dendritic function, without claiming completeness.

Scrapie Prions: The Transmissible Archetype of Protein Misfolding Diseases

A somewhat simplified but logic line in research can be drawn from historic scrapie to present-day Alzheimer's disease research. The first step was the experimental transmission of scrapie in sheep as mentioned above. A clear step forward was the observation of William J Hadlow that the neuropathology, that is, the spongiform degeneration of the brain and plaque deposition, of scrapie in sheep and the Kuru disease among the Fore people in Papua New Guinea looked very similar. Carleton Gajdusek and colleagues had described the Kuru disease as transmissible and could transmit experimentally the disease to monkeys. Briefly later, it became evident that also the Creutzfeldt–Jakob disease (CJD) belongs to that group of disease. When the first cases of bovine spongiform encephalopathy (BSE) in Great Britain were identified as a new disease, it was soon recognized as a prion disease. Very enigmatic, however, remained the etiology of CJD, which can be transmissible, for example, as iatrogenic accidents, or sporadic or even familial. Only the development of the prion hypothesis by Prusiner delivered a unifying concept: (1) the gene of the prion protein *Prnp* could be identified in every host accessible to the disease; (2) differences in the sequence of *Prnp* are the molecular basis for the observed species barriers although not understood quantitatively up to now; and (3) all familial cases of CJD are 100% correlated with a mutation in *Prnp*. The prion amplification model of Figure 2(a) is in accordance with the assumption that a mutation in *Prnp* could facilitate the conversion which then would lead to amplification as in the transmissible case. The sporadic cases would represent the very rare event of a spontaneous yet thermodynamically possible switch from PrP^C to PrP^{Sc}, which then would be amplified. As the prion model is today generally accepted as a molecular model, the diseases mentioned so far are called 'prion diseases'. A list of the most common natural prion diseases is shown in Table 2. Experimental prion diseases were transmitted to mice, hamsters, and bank voles.

The protein folding conversion of PrP and its connection with a disease suggested similarity to other neurodegenerative diseases like Alzheimer's disease, Parkinson, Huntington, and others. In those cases, specific proteins – others than PrP – are pathologically refolded and lead to disease-associated aggregates. These similarities led to the general concept of protein misfolding diseases; some are listed in Table 3. One should note, however, that only the prion diseases are transmissible,

Table 2 Natural prion diseases

Disease	Host	Etiology
Scrapie	Sheep	Natural transmission
Creutzfeldt–Jakob disease	Humans	90% sporadic Transmissible Familial
BSE (bovine spongiform encephalopathy)	Cattle	Transmission from meat and bone meal
CWD (chronic wasting disease)	Deer Elk	Natural transmission

Table 3 Protein misfolding diseases (selected)

Disease	Protein	Manifestation	Organ
Prion	Prion protein	cf. Table 2	Brain
Alzheimer	Abeta Tau	Sporadic Genetic	Brain
Parkinson	α -Synuclein	Sporadic Genetic	Brain
Huntington	Huntingtin	Genetic	Brain
Type II diabetes	Amylin (IAPP)	Sporadic	Pancreas
Injection-localized amyloidosis	Insulin	Sporadic (iatrogenic)	Skin Muscles

whereas the sporadic and genetic etiology is common to most protein misfolding diseases. That was also the reason why only in prion research an experimental animal model was available which facilitated the research significantly. When, however, the other protein misfolding disease could be established in transgenic mice, which overexpressed the corresponding protein, a new quality of research in the general protein misfolding disease was introduced.

Applied Aspects

Epidemic and Diagnosis

About 200 000 cattle were afflicted during the BSE epidemic in the United Kingdom and to a much lesser extent in other countries. In 1996, it became evident that BSE can be transmitted to humans. The disease was manifested in humans as variant CJD, a particular strain of CJD, and about 200 victims had to be regretted. The peak of the BSE epidemic was 1992, that of variant CJD in 2000, and both have dropped since then to nearly not significant numbers. The source of the infection with BSE was contaminated meat and bone meal, that of variant CJD consumption of cattle products, most probable of those containing high proportion of brain material.

In order to fight against the cattle epidemic and, in particular, against the danger of human disease, there was an urgent need for diagnostic procedures. Initially, safety measures were taken by testing all cattle in the slaughterhouses older than 24 months. Later, the age limit was raised stepwise to 30, 36,

and 42 months; still, cattle with clinical signs are always tested before entering the food chain.

All commercial tests were based on the property of partial PK-resistance of PrP^{Sc} in contrast to complete PK-sensitivity of PrP^C. The tests were post-mortem analyses of brain material; different kinds of enzyme-linked immunosorbent assay (ELISA) tests and Western-blot tests were offered. In recent years, much effort is taken to develop *ante mortem* diagnostics with blood or blood fractions as test material. Most promising is the method of PMCA (cf. above) and single particle counting methods, but both methods are not in routine use yet.

Post-mortem diagnosis in humans is similar to that in animals, but only of neuropathological and not of therapeutic interest. As a reliable *ante mortem* test from body fluids is not available, a combination of surrogate markers, imaging procedures like the MRI of the brain and behavioral tests are used with, however, very limited success in the early and even less in the preclinical stages of the disease.

Therapy

In spite of the large effort of several research groups, no therapeutic drug is presently available or applied in a larger clinical trial. A few medical treatments in the orphan drug mode are carried out presently with quinacrine and tetracycline; both are drugs that are licensed for other disease treatments. The molecular approaches to develop new drugs are directed to block or revert the conversion of PrP^C to PrP^{Sc}, to inhibit the synthesis of PrP^C, or to prevent the initial contact of PrP^C and PrP^{Sc}.

See also: [Molecular Biology: Prions of Yeast and Fungi: Proteins Acting as Genes](#); [Protein/Enzyme Structure Function and Degradation: Amyloid](#).

Further Reading

- Aguzzi A, Sigurdson C, and Heikenwaelder M (2008) Molecular mechanisms of prion pathogenesis. *Annual Review of Pathology* 3: 11–40.
- Castilla J, Saa P, Morales R, et al. (2006) Protein misfolding cyclic amplification for diagnosis and prion propagation studies. *Methods in Enzymology* 412: 3–21.
- Caughey B and Baron GS (2006) Prions and their partners in crime. *Nature* 443: 803–810.
- Govaerts C, Wille H, Prusiner SB, and Cohen FE (2004) Evidence for assembly of prions with left-handed beta-helices into trimers. *Proceedings of the National Academy of Sciences of the United States of America* 101: 8342–8347.
- Hörnlimann B, Riesner D, and Kretschmar HA (eds.) (2006) *Prions in Humans and Animals*. Berlin: De Gruyter.
- Kovacs GG and Budka H (2008) Prion diseases: From protein to cell pathology. *American Journal of Pathology* 172: 555–565.
- Prusiner SB (1982) Novel proteinaceous infectious particles cause scrapie. *Science* 216: 136–144.
- Prusiner SB (1998) Prions. *Proceedings of the National Academy of Sciences of the United States of America* 95: 13363–13383.
- Riek R, Hornemann S, Wider G, et al. (1996) NMR structure of the mouse prion protein domain PrP(121–321). *Nature* 382: 180–182.
- Safar JG, Kellings K, Serban A, et al. (2005) Search for a prion-specific nucleic acid. *Journal of Virology* 79: 10796–10806.

Processivity Clamps in DNA Replication

N Y Yao and M O'Donnell, Rockefeller University, Howard Hughes Medical Institute, New York, NY, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Megan J. Davey and Mike O'Donnell, volume 3, pp. 441–446, © 2004, Elsevier Inc.

Glossary

Processivity The ability of an enzyme to catalyze more than one turnover before releasing the product or the substrate.

Replicase Protein machinery that copies a strand of DNA. Chromosomal replicases typically contain a core polymerase, a processivity clamp, and a clamp loader.

Replisome Protein machinery that copies both strands of DNA. Replisomes include the replicase (polymerase, clamp, and clamp loader) as well as a helicase to unwind DNA. A priming activity is required for one strand.

Clamp and Clamp Loader Structures

Comparative structures of sliding clamps from bacteria, eukaryotes, and archaeal cells are shown in [Figure 1](#). Bacterial clamps, called β subunit, are homodimers, while eukaryotic and archaeal proliferating cell nuclear antigen (PCNA) clamps are homotrimeric. Regardless of their oligomeric state, all clamps have a pseudo-sixfold symmetric appearance based in the repetition of a domain of similar folding pattern. Thus, the protomers of dimeric clamps are composed of three repeated domains and the protomers of trimeric clamps are comprised of two repeated domains. This architectural similarity among all clamps suggests that the sliding clamp evolved early, and was present in the last common ancestor before cells diverged into the three different domains of life. The same can be said for the clamp loader.

The structure of clamp loaders from *Escherichia coli* and yeast has been solved to high resolution ([Figures 2\(a\)](#) and [2\(b\)](#)). In each case, the five subunits essential for clamp loader function are arranged in a circular spiral shape. The clamp loader of *E. coli*, referred to as γ complex, has two additional subunits (χ and ψ) that are not required for clamp loading, although they facilitate the reaction. The five clamp-loading subunits consist of three γ subunits and one each of the δ and δ' subunits, all of which are members of the AAA+ family of ATPases (ATPases associated with diverse cellular activities). Only the γ subunits contain ATP-binding sites. The γ , δ , and δ' subunits are composed of three domains; the two N-terminal domains contain the AAA+ region of homology, while the C-terminal domains are tightly packed together and form the major pentameric contacts.

The recent crystal structure of $\gamma_3 \delta \delta'$ bound to a primed DNA reveals details of how it interacts with a primed site in a structure specific, yet sequence-independent, fashion ([Figure 2\(a\)](#)). The N-terminal domains (domain 1) are arranged in a spiral that matches the pitch of duplex DNA and they form attachments to the phosphodiester backbone of the template strand of the duplex. The circular arrangement of the subunits thus forms a central chamber into which duplex DNA fits. The C-terminal domains (domain 3) form an uninterrupted collar which prevents DNA from continuing in a straight path. However, there is a distinct gap

in the side of the clamp loader located between the AAA+ domains of the δ and δ' subunits, and the single-strand (ss) DNA bends to exit through this gap. The sharp bend required to exit the central chamber of the clamp loader cannot be accommodated by a rigid duplex, and it is this requirement for a sharp bend that confers structure-specific binding to a primed site by clamp-loading machines. Specifically, the ssDNA portion of the primed site has the requisite flexibility needed to make the sharp bend for exit of the DNA from the central chamber.

The structure of the eukaryotic clamp loader, referred to as replication factor C (RFC), has been solved in complex with the PCNA clamp ([Figure 2\(b\)](#)). Unlike the five clamp-loading subunits of the bacterial clamp loader, each of the five RFC subunits is encoded by a distinct gene. However, each RFC subunit is an AAA+ family member and shows homology to the bacterial clamp loader subunits. Accordingly, RFC is structurally and functionally similar to *E. coli* γ complex. The PCNA clamp binds to the N-terminal domains of RFC subunits and a model of DNA bound to the RFC–PCNA complex indicates that DNA located in the central chamber of RFC is positioned to pass through the center of the PCNA ring. This satisfying structural arrangement suggests that, upon opening the clamp, the clamp loader will bind a primed site in a way that positions DNA through the clamp.

Biochemical studies have shown that ATP binding is sufficient for clamp opening and DNA binding in both bacterial and eukaryotic systems. ATP hydrolysis is required for clamp closing and this ejects the clamp loader from the DNA, leaving the clamp topologically linked to DNA as illustrated in [Figure 2\(c\)](#). The RFC–PCNA structure in [Figure 2\(b\)](#) contains ATP γ S, yet PCNA remains closed in the RFC–PCNA structure. Presumably, this closed state of PCNA is due to ATP site mutations in RFC that prevent hydrolysis during crystal growth. The PCNA ring interacts with three of the five RFC subunits because the N-terminal surface of the clamp loader is not flat, but instead assumes a spiral shape. Molecular simulation studies indicate that the PCNA clamp opens into a right-handed lock-washer configuration that may fit on the N-terminal spiral surface of RFC when the clamp is opened. A similar structure is implied by electron microscopy particle reconstruction studies of an archaeal RFC–PCNA–DNA complex.

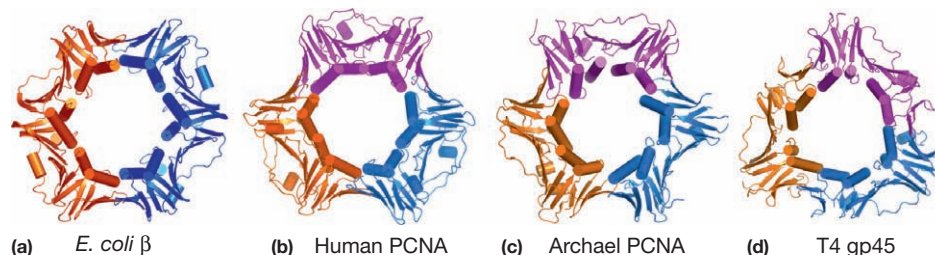


Figure 1 Sliding clamps from various walks of life. (a) *E. coli* (bacterial) β (PDB ID, 2POL), (b) human (eukaryotic) PCNA (PDB ID, 1AXC), (c) *Sulfolobus solfataricus* (archaeal) PCNA (PDB ID, 2IX2), and (d) bacteriophage T4 gp45 protein (PDB ID; 1CZD).

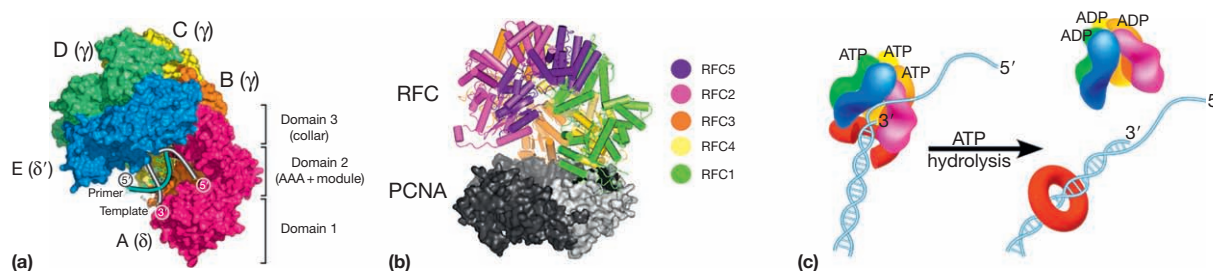


Figure 2 Clamp loaders are heteropentamers. (a) *E. coli* γ complex bound to a primed template (PDB ID; 3GLF). (b) Yeast RFC bound to PCNA (PDB ID; 1SXJ). (c) Summary of ATP use in the clamp-loading reaction.

Polymerase Switching on Clamps during Replication

All cells contain multiple DNA polymerases and repair factors that function with the sliding clamp. The fact that sliding clamps are homooligomers has given rise to the tool-belt hypothesis in which multiple polymerases attach to one clamp simultaneously. In fact, the binding of more than one protein to the same clamp has been demonstrated experimentally. However, the clamp can only accommodate one DNA template and therefore the different polymerases attached to the clamp must time-share for the single DNA template. The process by which different enzymes share, or switch, with one another is suggested by the crystal structure of *E. coli* β bound to a primed site (Figure 3). The structure shows that DNA passes through the ring at a sharp angle, $\sim 22^\circ$ relative to the C2 axis perpendicular to the plane of the ring. Both strands of the duplex are bound by residues on loops that project from the face of the clamp (see Figure 3(a)). The loops of only one of the two identical subunits of the β clamp can bind to DNA at a given time. Thus, the DNA likely isomerizes between the two protomers, and this DNA flipping from one protomer to another may provide the motions needed to switch the primed site among different polymerases attached to the same clamp as illustrated in Figure 4(a).

Experiments to address the use of different polymerases with the clamp have demonstrated that they can exchange with one another through an intermediate complex in which two DNA polymerases are bound to the same clamp. An early demonstration of polymerase switching utilized the phage T4 system in which mutant and wild-type T4 gp43 DNA polymerases traded places on the same clamp during ongoing replication. Similar studies in *E. coli* showed that two different polymerases, the Pol III replicase and Pol IV translesion bypass enzyme, simultaneously bind to β and trade places on the

DNA. The latter study demonstrated that when Pol III stalls, Pol IV takes over the primed site. Pol III recruits the primed site back from Pol IV under moving conditions.

Polymerase Exchange on Sliding Clamps during DNA Damage

One unexpected feature of the β -DNA structure is an interaction of the ssDNA at the primed site as illustrated in Figure 3(b). The ssDNA binds into a hydrophobic pocket on the surface of the clamp. The ssDNA is guided into the binding pocket by a channel lined with positively charged residues and by two conserved tyrosine residues that stack with nucleotide bases. This hydrophobic binding pocket is the same site to which DNA polymerases attach to the clamp. Therefore, it seems unlikely to be occupied by DNA while the clamp is functioning with the DNA polymerase. However, if a DNA polymerase dissociates from the clamp, this binding pocket may serve the function of a placeholder, helping to hold the clamp at the primed site where it can later be joined with a DNA polymerase (see Figure 4(b)). This placeholder action has been demonstrated in single-molecule studies and may be useful during polymerase exchange reactions in which the clamp is unoccupied for a period of time during the switching process.

Recent experiments address the response of replication forks to high levels of translesion (TLS) polymerases which are induced inside cells upon DNA damage. DNA damage in *E. coli* leads to an SOS response in which many different proteins are induced to high levels. These proteins generally help the cell to cope with DNA damage. Among the earliest proteins to be induced are Pol II and Pol IV, TLS polymerases that function with the β clamp to extend DNA over various DNA blocks. An *in vivo* study showed that induction of Pol IV

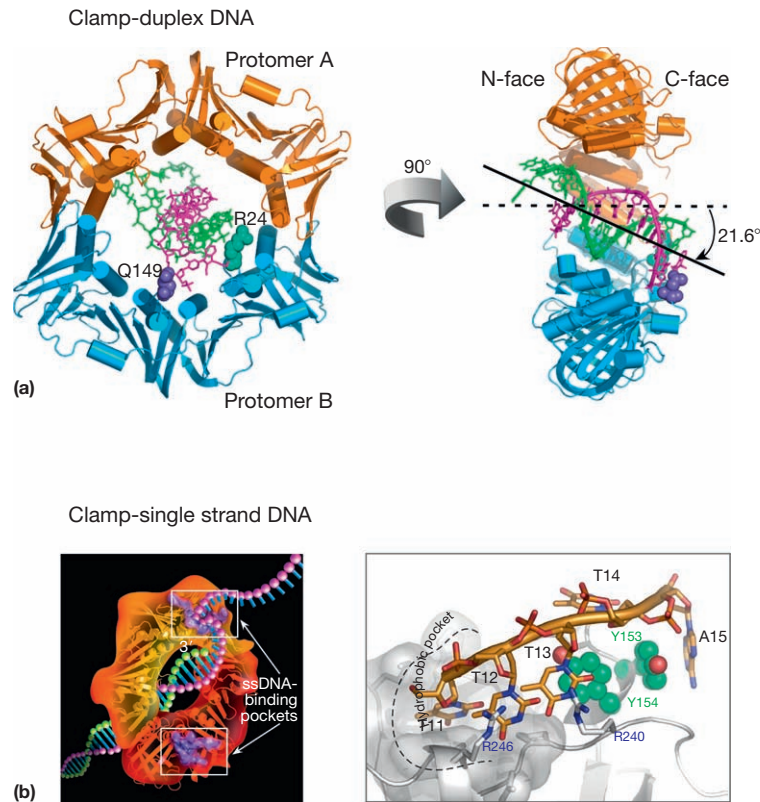


Figure 3 Interactions of the β clamp with DNA. (a) Interaction of *E. coli* β with double-strand DNA of a primed template reveals DNA is highly tilted as it passes through the ring. Residues R24 and Q149 are located on surface loops and these residues of one subunit bind both strands of DNA (R24 and Q149 are shown as a space filling representation). (b) Interaction of β with ssDNA at the primed site. The illustration depicts binding of the ssDNA to a surface pocket in β . The inset shows the detailed structure of ssDNA binding to the pocket. Tyrosines 153 and 154 stack with nucleotide bases. Adapted from figures 3(a) and 4(c) of Georgescu RE, Kim SS, Yurieva O, et al. (2008) Structure of a sliding clamp on DNA. *Cell* 132: 43–54.

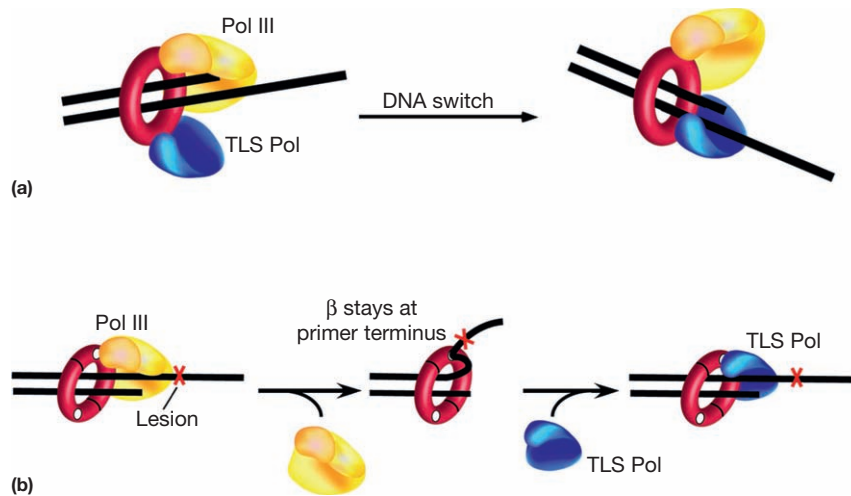


Figure 4 Predicted functions of DNA interaction with the clamp. (a) The tilted DNA binds residues of one β protomer (see [Figure 3\(a\)](#)), yet these same residues are present on the other protomer of β . Thus, the DNA likely isomerizes from one protomer to the other, resulting in large movements of the primed template junction. These movements may facilitate switching of the primed site between different DNA polymerases bound to the same ring. (b) The ssDNA binding site may serve a placeholder role, holding β to a primed site if no DNA polymerase is bound to it. Adapted from Figure 7 of Yao NY and O'Donnell M (2008) Replisome dynamics and use of DNA trombone loops to bypass replication blocks. *Molecular Biosystems* 4: 1075–1084.

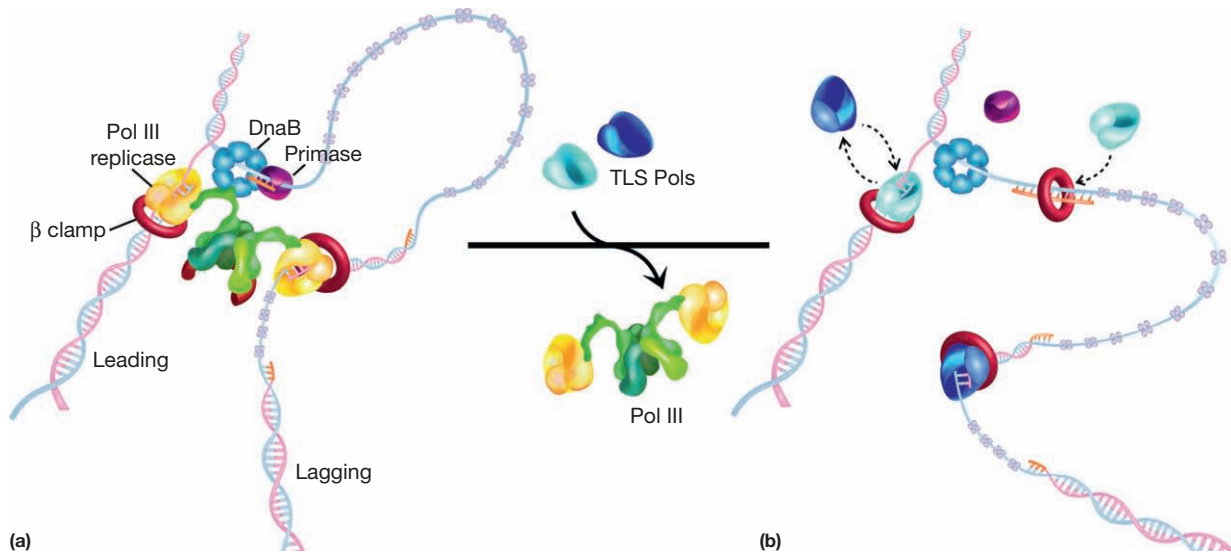


Figure 5 The replisome can morph into different architectures. (a) A replisome consisting of Pol III holoenzyme attached to DnaB, β clamps, and primase. Leading and lagging strand syntheses are coupled by the dimeric structure of Pol III holoenzyme. (b) Displacement of Pol III holoenzyme by TLS polymerases such as Pol II and Pol IV results in an uncoupled replisome. The TLS polymerases act in a distributive fashion by trading places on β clamps.

halts replication forks in a β -dependent fashion. *In vitro* studies reveal that both Pol II and Pol IV rapidly dislodge Pol III from β in the context of the moving replication fork, even in the absence of a DNA blocking lesion.

The fact that Pol III can be dislodged from the replication fork is somewhat surprising given the numerous interactions of the Pol III replicase with other components of the replication fork (Figure 5(a)). At a replication fork, Pol III is embedded in a large multi-protein complex referred to as Pol III holoenzyme, which contains two molecules of Pol III attached to one clamp loader. The clamp loader within Pol III holoenzyme contains τ subunits in place of the γ subunits. This is possible because both τ and γ are encoded by the same gene (*dnaX*). The τ subunit is the full-length product of the *dnaX* gene, while γ is truncated by a translational frameshift. The extra C-terminal 24 kDa in τ consists of two domains that appear to connect to the clamp-loading domains via a flexible linker. The C-terminal domains on the τ subunits are known to bind Pol III and also the hexameric DnaB helicase that encircles the lagging strand. The τ subunits also bind directly to DNA, further cementing Pol III holoenzyme into the replication fork. In addition, Pol III holoenzyme contains a subunit called χ that binds to SSB (single-strand binding protein), providing yet another point of contact between Pol III holoenzyme and the replication fork. These several contacts may be expected to hold Pol III to the replication fork.

Despite these several connections, Pol II and Pol IV are capable of rapidly taking over β clamps from Pol III on both the leading and lagging strands of the moving replication fork, as illustrated in Figure 5(b). Due to the slow speed of Pol II and Pol IV, the replication fork slows dramatically, from about 1000 basepairs (bp)/s to 10 bp/s (Pol II) or 1 bp/s (Pol IV) depending on which polymerase occupies the fork. In the cell, these slow-moving replication forks may give more time for DNA damage to be repaired by normal repair systems, thereby

diminishing the probability that a replisome will encounter a DNA lesion. Study of these slow replication forks reveal that the translesion DNA polymerases, Pol II and Pol IV, rapidly exchange on the clamp, and that the clamp is unoccupied some of the time. Thus, translesion polymerase action within the translesion replisome is highly distributive. Under these conditions, the β clamp-ssDNA interaction may come in very useful in the role of a placeholder for the β clamp.

Questions for Future Studies

It has been nearly 20 years since the circular nature of the *E. coli* β clamp, the first protein known to bind DNA topologically, was discovered. On the one hand, we know a great deal about the structure and function of DNA clamp processivity factors. On the other hand, their use by many different polymerases and repair factors has expanded the list of important questions about how these clamps function. In both bacteria and eukaryotic systems, the replicating polymerase can bind tight to its cognate sliding clamp, yet hop from one clamp to the next. This polymerase hopping mechanism was originally thought to function only on the lagging strand enabling extension of multiple Okazaki fragments by one DNA polymerase. However, recent studies suggest that polymerases hopping among clamps may be used as a general strategy to bypass other types of blocks, such as lesions or RNA polymerase. Further studies should elucidate if clamps circumvent barriers to replication by mediating polymerase hopping around the barrier. The detailed mechanism by which different polymerases recruit the DNA template while bound to a clamp with another polymerase is also still largely a mystery. Clamps can also be modified. The eukaryotic PCNA clamp is modified by ubiquitin in response to DNA damage; the function of this modification is not clear. Eukaryotes also possess alternative clamp

loaders, one of which loads an alternative clamp (911 clamp) onto a 5' single-strand/double-strand junction for activation of a kinase involved in the DNA damage checkpoint response. Clearly, many exciting questions about sliding clamp function remain for future studies.

Acknowledgments

The authors are grateful to grants from the NIH, and a grant from HHMI in support of these studies.

See also: **Molecular Biology:** DNA Helicases: Hexameric Enzyme Action; DNA Polymerase β , Eukaryotic; Sliding Clamps in DNA Replication: *Escherichia coli* β -Clamp and PCNA Structure.

Further Reading

- Benkovic SJ, Valentine AM, and Salinas F (2001) Replisome-mediated DNA replication. *Annual Review of Biochemistry* 70: 181–208.
- Bowman GD, O'Donnell M, and Kuriyan J (2004) Structural analysis of a eukaryotic sliding DNA clamp–clamp loader complex. *Nature* 429: 724–730.
- Courcelle J, Khodursky A, Peter B, Brown PO, and Hanawalt PC (2001) Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. *Genetics* 158: 41–64.
- Dalrymple BP, Kongsuwan K, Wijffels G, Dixon NE, and Jennings PA (2001) A universal protein–protein interaction motif in the eubacterial DNA replication and repair systems. *Proceedings of the National Academy of Sciences of the United States of America* 98: 11627–11632.
- Ellison V and Stillman B (2003) Biochemical characterization of DNA damage checkpoint complexes: Clamp loader and clamp complexes with specificity for 5' recessed DNA. *PLoS Biology* 1: E33.
- Fujii S and Fuchs RP (2004) Defining the position of the switches between replicative and bypass DNA polymerases. *EMBO Journal* 23: 4342–4352.
- Georgescu RE, Kim SS, Yurieva O, et al. (2008) Structure of a sliding clamp on DNA. *Cell* 132: 43–54.
- Johnson A and O'Donnell M (2005) Cellular DNA replicases: Components and dynamics at the replication fork. *Annual Review of Biochemistry* 74: 283–315.
- Kong XP, Onrust R, O'Donnell M, and Kuriyan J (1992) Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: A sliding DNA clamp. *Cell* 69: 425–437.
- McHenry CS (2003) Chromosomal replicases as asymmetric dimers: Studies of subunit arrangement and functional consequences. *Molecular Microbiology* 49: 1157–1165.
- Napolitano R, Janel-Bintz R, Wagner J, and Fuchs RP (2000) All three SOS-inducible DNA polymerases (Pol II, Pol IV and Pol V) are involved in induced mutagenesis. *EMBO Journal* 19: 6259–6265.
- Simonetta KR, Kazmirski SL, Goedken ER, et al. (2009) The mechanism of ATP-dependent primer-template recognition by a clamp loader complex. *Cell* 137: 659–671.
- Stukenberg PT, Turner J, and O'Donnell M (1994) An explanation for lagging strand replication: Polymerase hopping among DNA sliding clamps. *Cell* 78: 877–887.
- Sutton MD and Walker GC (2001) Managing DNA polymerases: Coordinating DNA replication, DNA repair, and DNA recombination. *Proceedings of the National Academy of Sciences of the United States of America* 98: 8342–8349.
- Yang J, Zhuang Z, Roccasecca RM, Trakselis MA, and Benkovic SJ (2004) The dynamic processivity of the T4 DNA polymerase during replication. *Proceedings of the National Academy of Sciences of the United States of America* 101: 8289–8294.
- Yao NY and O'Donnell M (2008) Replisome dynamics and use of DNA trombone loops to bypass replication blocks. *Molecular Biosystems* 4: 1075–1084.
- Yao NY and O'Donnell M (2009) Replisome structure and conformational dynamics underlie fork progression past obstacles. *Current Opinion in Cell Biology* 21: 336–343.

Prostaglandins and Leukotrienes

W L Smith, University of Michigan Medical School, Ann Arbor, MI, USA

R C Murphy, University of Colorado Medical School, Denver, CO, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 452–456, © 2004, Elsevier Inc.

Glossary

5-Lipoxygenase The oxygenase that converts arachidonic acid and related 20 carbon essential fatty acids to leukotrienes.

Aspirin A nonsteroidal anti-inflammatory drug that inhibits cyclooxygenase by acetylating an active site serine residue and blocking productive binding of arachidonic acid.

Cyclooxygenase Common name of the oxygenase enzyme that converts arachidonic acid and related 20 carbon essential fatty acids to prostanoids.

FLAP 5-Lipoxygenase activating protein, an accessory protein that is required for 5-lipoxygenase functioning in intact cells.

Leukotriene Those oxygenated derivatives of 20 carbon essential fatty acids formed through the action of 5-lipoxygenase and containing a conjugated triene system of double bonds.

Montelukast A drug used in the treatment of asthma that inhibits the binding of peptidoleukotrienes to the CysLT₁ receptor.

Prostanoid Those oxygenated derivatives of 20 carbon essential fatty acids formed through the action of prostaglandin endoperoxide H synthase.

Prostanoid Structures

Figure 1 shows the biosynthetic relationships among the prostanoids formed from AA. The letters following the abbreviation PGs indicate the nature and location of the oxygen-containing substituents present in the cyclopentane ring; the subscript indicates the number of carbon–carbon double bonds in the side chains. Greek subscripts denote the orientation of ring hydroxyl groups (e.g., PGF_{2α}). Prostanoids known as isoprostanes are formed from AA by nonenzymatic peroxidation and consequently contain many stereoisomers. Interestingly, the amounts of isoprostane metabolites in urine are greater than prostanoids formed enzymatically.

Prostanoid Biosynthesis

Eicosanoids are not stored in cells but are formed on demand in response to extracellular hormonal stimuli (e.g., bradykinin, angiotensin, and thrombin) that increase cell Ca²⁺ concentrations. Prostanoid formation requires three enzymatic steps: (1) mobilization of AA (or 2-arachidonylglycerol (2-AG)) from membrane phospholipids; (2) conversion of arachidonate (or 2-AG) to PGH₂ (or 2-PGH₂-glycerol); and (3) isomerization of PGH₂ (or 2-PGH₂-glycerol) to one of the major prostanoids by a terminal synthase. Mobilization of AA involves cytosolic phospholipase A₂ (cPLA₂) which translocates to the endoplasmic reticulum (ER) and nuclear membrane (NM) when intracellular Ca²⁺ concentrations rise and cleaves AA from a phospholipid on the cytosolic surface of the membrane. Newly released AA moves, methyl end first, to the luminal half of the bilayer where it enters the active site of a prostaglandin endoperoxide H synthase (PGHS) and is oxygenated to PGG₂ and its 15-hydroperoxyl group is then reduced to form PGH₂; PGHSs are also called cyclooxygenases (COXs) and there are two isoforms (PGHS-1 and -2; COX-1 and -2). Once formed, PGH₂ is isomerized by a synthase located in the ER

(e.g., TXA synthase and PGE synthase). Recently, 2-AG has been found to be converted to 2-PGH₂-glycerol by PGHS-2. This latter intermediate can be converted to the 2-prostanlyl-glycerol derivatives but not 2-thromboxane-glycerol. 2-AG itself is probably formed from phosphatidylcholine by phospholipase C and monoacylglycerol lipase.

PGHS Catalysis and Inhibition

Figure 2 is a model of the COX and peroxidase active sites of ovine PGHS-1. To initiate oxygenation of AA at the COX site, the heme group at the peroxidase site must first be oxidized by a hydroperoxide. This causes formation of an oxidized heme intermediate that is reduced by an electron from Tyr385, thereby generating a Tyr385 radical. This radical abstracts a hydrogen from AA so that it can react with oxygen (**Figure 1**). The guanidino group of Arg120 binds to the carboxylate group of AA.

PGHS-1 is a constitutive enzyme purified in 1975 and its cDNA cloning has been reported in 1988. PGHS-2 is an inducible enzyme discovered in 1991 as an immediate early gene product in phorbol ester-activated murine 3T3 cells and in *v-src*-transformed chicken fibroblasts. PGHS-2 contains a unique 18 amino acid insert near its carboxyl terminus that regulates the degradation of the protein. The reason for the existence of the two PGHS isoforms is not known. Each protein functions as a homodimer with a monomer mass of 72 kDa. PGHSs are integral membrane proteins which instead of having transmembrane helices have a cluster of four amphipathic helices that interact with only one face of the lipid bilayer.

PGHSs are inhibited by nonsteroidal anti-inflammatory drugs (e.g., aspirin, ibuprofen, and naproxen), which compete with AA for binding to the COX site of PGHSs. Prostaglandin synthesis mediated by PGHS-2 can be inhibited by anti-inflammatory steroids, which block the synthesis of PGHS-2 protein. Drugs called COX-2 inhibitors (e.g., celecoxib) target PGHS-2 specifically and are used as anti-inflammatory and analgesic agents.

Aspirin, acetylsalicylic acid, binds the COX site of PGHSs and acetylates Ser530 causing irreversible inhibition of PGHSs (Figure 2). Low doses of aspirin are used prophylactically to reduce coronary thrombosis; at low doses, aspirin inhibits PGHS-1 in platelets blocking thromboxane formation. PGHS inhibitors appear to reduce mortality from colon cancer.

PGHS-1 and PGHS-2 Gene Expression

There are separate genes for PGHS-1 and PGHS-2 located on different chromosomes. They have similar intron/exon structures

but the PGHS-2 gene is 8 kb and the PGHS-1 gene is 22 kb. Little is known about the regulation of PGHS-1 gene expression. PGHS-2 gene transcription can be induced by cytokines and growth factors that function through multiple response elements in the PGHS-2 gene promoter. There is much interest in this because of the potential role of altered PGHS-2 expression in carcinogenesis and inflammatory diseases.

PGH₂ Metabolism

Different cells form different prostanoids. PGE₂, PGD₂, PGF₂, PGI₂, and TXA₂ syntheses from PGH₂ are catalyzed by PGE, PGD, PGF, PGI, and TXA synthases, respectively; all of these enzymes catalyze isomerization reactions. PGF₂ formation involves a two-electron reduction of PGH₂. PGI and TXA synthases are cytochrome P-450s. PGE synthase is a member of the MAPEG family of transmembrane ER proteins and requires reduced glutathione for activity. There are two PGD synthases: a hematopoietic type is glutathione dependent and has been isolated from spleen and a lipocalin type is glutathione independent and occurs in brain.

Prostanoid Catabolism

The initial step in the inactivation of PGE₂ is oxidation of the 15-hydroxyl group to a 15-keto group catalyzed by 15-hydroxyprostaglandin dehydrogenase. Further catabolism involves reduction of the double bond between C-13 and C-14, β -oxidation, and ω -oxidation.

Prostanoid Actions

There are pharmacologically distinct receptors for each of the known prostanoids. In the case of PGE₂, four different prostaglandin E (EP) receptors have been identified and designated as EP1, EP2, EP3, and EP4 receptors. EP1 is coupled through G_q to the activation of phospholipase C, EP2 and EP4 are coupled

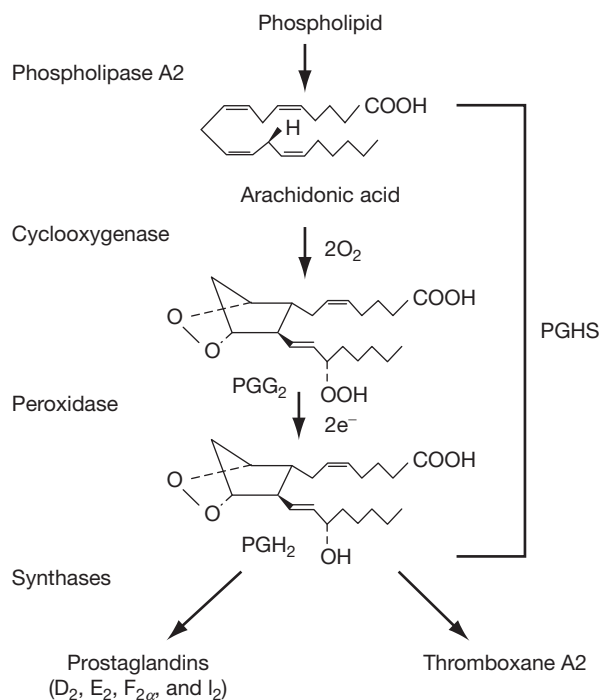


Figure 1 Biosynthesis of common prostanoids from AA.

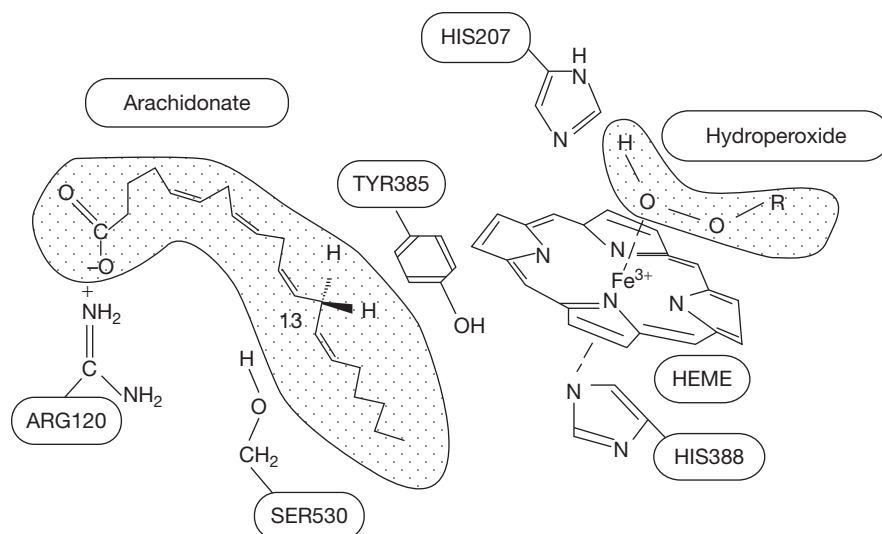


Figure 2 Model of the cyclooxygenase and peroxidase active sites of the ovine PGHS-1.

via G_s to the stimulation of adenylate cyclase, and EP3 receptors are coupled via G_i to inhibition of adenylate cyclase.

EP3 receptors are involved in the development of fever, EP2 and EP4 function in bone resorption, and EP1 receptors are involved in chemically induced colon cancer. Prostanoids can activate some isoforms of peroxisomal proliferator-activated receptors (PPARs). PGI_2 can be involved in PPAR δ -mediated responses such as decidualization and apoptosis.

Leukotrienes and Lipoxygenase Products

Leukotrienes were discovered in 1979 during the search for the chemical structure of the "slow reacting substance of anaphylaxis." Leukotrienes are produced by the action of 5-lipoxygenase (5-LO) (Figure 3) which catalyzes both an oxygen insertion to form 5-HpETE and a dehydration reaction to form leukotriene A_4 (LTA_4). LTA_4 is converted to the biologically active leukotrienes LTB_4 and LTC_4 by LTA_4 hydrolase and LTC_4 synthase, respectively. LTB_4 is a potent chemotactic and chemokinetic agent for human polymorphonuclear leukocytes, whereas LTC_4 constricts specific smooth muscle (e.g., bronchial smooth muscle) and mediates leakage of vascular fluid in the process of edema. The formation of LTC_4 versus LTB_4 is controlled by the expression of LTA_4 hydrolase or LTC_4 synthase by specific cell types. For example, the human neutrophil expresses LTA_4 hydrolase and produces LTB_4 ,

whereas mast cells and eosinophils express LTC_4 synthase and produce LTC_4 .

5-Lipoxygenase

Human 5-LO (E.C.1.13.11.34) is a nonheme iron-containing protein of 673 amino acids. It catalyzes the abstraction of a hydrogen atom from C-7 of AA and insertion of O_2 at C-5 to generate 5(S)-HpETE (Figure 3). The LTA_4 synthase activity of 5-LO then catalyzes removal of 10 pro-R hydrogen atoms from 5(S)-HpETE through a second redox cycle and an internal rearrangement of double bonds to form LTA_4 . LTA_4 has a half-life of less than 10 s at pH 7.4. Purified 5-LO requires Ca^{2+} , ATP, fatty acid hydroperoxide, and phosphatidylcholine in addition to the AA and O_2 substrates. Ca^{2+} facilitates the association of 5-LO with internal membranes. The need for an increase in intracellular Ca^{2+} concentrations differentiates 5-LO from other lipoxygenases. 5-LO oxygenates AA at the nuclear member (NM). When Ca^{2+} concentrations increase in neutrophils and mast cells, 5-LO and cytosolic PLA $_2$ become associated with the nuclear membrane where another protein required for LTA_4 synthesis – 5-lipoxygenase activating protein (FLAP) – is found. In other cells 5-LO is constitutively associated with the NM, whereas in alveolar macrophages 5-LO is present in the nucleus. The drug zileuton inhibits 5-LO.

The 5-LO gene is large (c. 80 kb) and has 14 exons. It is on human chromosome 10. The human 5-LO gene promoter

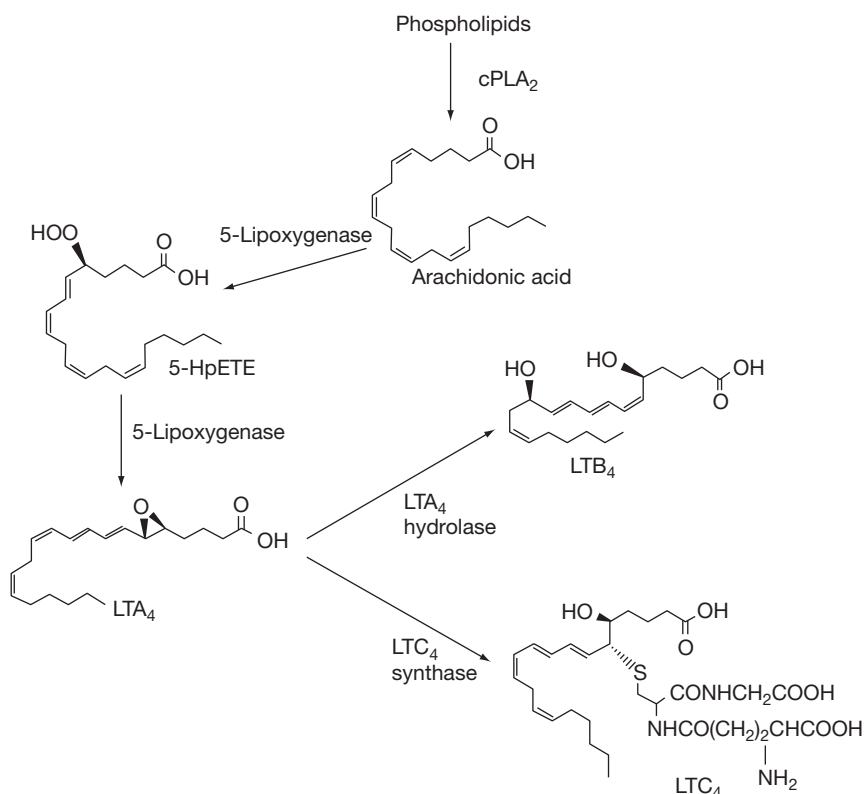


Figure 3 Biochemical pathway of the conversion of AA into the biologically active leukotrienes. AA released from phospholipid by cPLA $_2$ is metabolized by 5-LO to 5-hydroperoxyeicosatetraenoic acid (5-HpETE) and leukotriene A_4 (LTA_4) which is then converted to LTB_4 by LTA_4 hydrolase or conjugated to glutathione to form LTC_4 by LTC_4 synthase.

contains Spl and EGR-1-binding sites at -88 to -212 bp upstream of the translation start site. Genetic polymorphisms in humans are found in Spl-binding sites.

5-Lipoxygenase Activating Protein

FLAP was discovered as a novel protein essential for leukotriene synthesis which was the target of an investigational drug from Merck (MK886). FLAP is an integral membrane protein of the NM containing 161 amino acids. The function of FLAP is unclear but it does bind AA analogs suggesting that it may transfer AA to 5-LO. LTC₄ synthase has 31% amino acid identity to FLAP with a highly conserved region putatively involved in AA binding.

LTA₄ Hydrolase and LTC₄ Synthase

LTA₄ hydrolase catalyzes the stereochemical addition of water to C-12 of LTA₄ to form the neutrophil chemotactic factor LTB₄. LTA₄ hydrolase contains 610 amino acids and one essential zinc atom. The enzyme is structurally related to zinc metalloproteases and exhibits protease activity; bestatin and captopril inhibit the enzyme. LTA₄ and LTC₄ synthases are found in many cells including those which do not contain 5-LO. It is believed that such cells can cooperate in LTB₄ and LTC₄ transcellular synthesis by transferring LTA₄ from one cell to another. LTA₄ hydrolase is localized in the cytosol. To metabolize the unstable LTA₄, LTA₄ hydrolase must either come in contact with the NM during LTA₄ formation or a carrier protein must transfer LTA₄ to LTA₄ hydrolase. The crystal structure of LTA₄ hydrolase has been determined.

LTC₄ synthase (E.C.2.5.1.37) conjugates reduced glutathione (γ -glutamyl-cysteinyl glycine) to LTA₄. LTC₄ synthase is located on the NM. The enzyme has significant sequence homology to microsomal glutathione-S-transferases and FLAP. LTC₄ synthase is found predominantly in mast cells, macrophages, eosinophils, and monocytes. MK-886, which interacts with FLAP, also inhibits LTC₄ synthase. The gene for LTC₄ synthase is located on human chromosome 5.

Leukotriene Metabolism

LTB₄ can be oxygenated to 20-hydroxy-LTB₄ by specific cytochrome P-450s of the CYP4F family. 20-Hydroxy-LTB₄ can be further metabolized to 20-carboxy-LTB₄ by CYP4F3 or to 20-oxo-LTB₄ by alcohol dehydrogenase then to 20-carboxy-LTB₄ by fatty aldehyde dehydrogenase. An alternative pathway involves an initial oxidation of the 12-hydroxy group to a 12-oxo moiety followed by reduction of the conjugated dienone and the 10,11 double bond. A secondary catabolic pathway, β -oxidation,

can occur from the C-1 carboxyl moiety of LTB₄ resulting in the loss of the C-5 hydroxyl group and from the 20-carboxy terminus of 20-carboxy-LTB₄. Individuals with genetic deficiencies in peroxisomal metabolism (Zellweger disease) that reduce β -oxidation or in aldehyde dehydrogenase (Sjogren-Larsson syndrome) excrete intact LTB₄ and 20-hydroxy-LTB₄. LTC₄ catabolism involves peptide cleavage reactions involving glutamyl transpeptidase and various dipeptidases to yield LTD₄ and LTE₄, both of which have biological activity. The sulfur atom of sulfidopeptide leukotrienes can be oxidized by reactive oxygen species. More specific metabolic processing of the sulfidopeptide leukotrienes includes ω -oxidation by cytochrome P-450 followed by β -oxidation from the ω terminus.

Leukotriene Biology and Leukotriene Receptors

LTB₄ functions in inflammatory processes through its chemotactic and chemokinetic affect human polymorphonuclear leukocytes. LTB₄ induces the adherence of neutrophils to vascular endothelial cells and enhances the migration of neutrophils (diapedesis) into extravascular tissues. The biological activity of LTB₄ is mediated through two specific G-protein-coupled receptors termed BLT₁ and BLT₂. LTC₄ and its peptide cleavage products LTD₄ and LTE₄ mediate bronchial smooth muscle contraction in asthma and cause edema. There are two G-protein-linked receptors for cysteinyl leukotrienes called CysLT₁ and CysLT₂. CysLT₁ is found in bronchial and intestinal smooth muscle; LTD₄ and LTC₄ activate CysLT₁. Montelukast, pranlukast, and zafirlukast inhibit the CysLT₁ receptor in humans. CysLT₂ receptors are abundant in the heart and in vascular endothelial cells.

See also: **Bioenergetics:** Cytochrome P-450; **Signaling:** Eicosanoid Receptors; Phospholipase C.

Further Reading

- Evans J, Ferguson A, Mosley R, and Hutchinson J (2008) What's all the FLAP about?: 5-Lipoxygenase-activating protein inhibitors for inflammatory diseases. *Trends in Pharmacological Sciences* 29: 72–78.
- Funk C (2005) Leukotriene modifiers as potential therapeutics for cardiovascular disease. *Nature Reviews Drug Discovery* 4: 664–672.
- Murphy R and Gijon M (2007) Biosynthesis and metabolism of leukotrienes. *Biochemical Journal* 405: 379–395.
- Rouzer CA and Marnett LJ (2009) Cyclooxygenases: Structural and functional insights. *Journal of Lipid Research* 50: S29–S34.
- Smith WL (2008) Nutritionally essential fatty acids and biologically indispensable cyclooxygenases. *Trends in Biochemical Sciences* 33: 27–37.
- Smith WL and Murphy RC (2008) The eicosanoids: Cyclooxygenase, lipoxygenase, and epoxygenase pathways. In: Vance DE and Vance JE (eds.) *Biochemistry of Lipids, Lipoproteins and Membranes*, 5th edn., pp. 331–362. Amsterdam: Elsevier.

Proteases in Blood Clotting

J D Kulman and E W Davie, University of Washington, Seattle, WA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 457–463, © 2004, Elsevier Inc.

Glossary

Gla The amino acid gamma-carboxyglutamic acid.

Hemophilia Any of several hereditary blood coagulation disorders in which the blood fails to clot normally.

Hemostasis The stoppage of blood flow characterized by vasoconstriction and the formation of a platelet plug and polymerized fibrin.

Platelet A small, non-nucleated, cell body found in the blood that functions to promote blood clotting.

Protease Any of various enzymes that catalyze the hydrolytic breakdown of proteins into peptides or amino acids.

Proteolysis The hydrolytic breakdown of proteins into smaller proteins, peptides, or individual amino acids.

Transglutaminase Any of various enzymes that catalyze the covalent linkage of protein-bound glutamine and lysine amino acid side chains.

Vitamin K A fat-soluble vitamin that promotes blood coagulation and prevents hemorrhage.

Warfarin A synthetic organic compound, derived from the natural product coumarin, that antagonizes the effect of vitamin K.

Zymogen The inactive precursor of an enzyme, converted into an active enzyme by proteolysis.

The biochemical pathways leading to thrombin formation are now well understood, as are the mechanisms by which thrombin promotes and modulates hemostasis. In response to vascular damage, thrombin is generated from an inactive precursor by a series of biochemical reactions collectively known as the coagulation cascade (Figure 1). Thrombin subsequently catalyzes a series of reactions that lead to platelet aggregation, fibrin polymerization, and, ultimately, feedback downregulation of the coagulation cascade itself (Figure 2). The enzymes involved in these pathways are structurally and functionally related to trypsin, the prototypical serine protease of the digestive system. However, in contrast to trypsin, the coagulation proteases are characterized by a complex modular architecture that facilitates interactions with one another, with other protein substrates, as well as with protein cofactors. These interactions serve to restrict the proteolytic reactions of the blood coagulation cascade to coagulation-promoting surfaces at sites of vessel injury.

The Coagulation Cascade

The Extrinsic Pathway

Almost since the 1960s, the coagulation cascade has been envisaged as a series of stepwise reactions in which inactive forms of serine proteases are converted to active forms by limited proteolysis in a process known as ‘zymogen activation’. The coagulation cascade is initiated when tissue factor, an integral membrane protein located in the tissue adventitia, is exposed to circulating blood following a vascular injury (Figure 1). The binding of tissue factor to factor VII or to the active serine protease factor VIIa (trace amounts of which circulate in the blood) initiates the extrinsic pathway of coagulation. The tissue factor–factor VIIa complex converts factors IX

and X to their active forms (IXa and Xa). Feedback amplification is achieved when inactive factor VII bound to tissue factor is converted to its active form (VIIa) by factors VIIa, IXa, and Xa, resulting in an enhanced generation of factor Xa. The factor Xa thus produced generates thrombin from its inactive form (prothrombin) in the presence of calcium, the nonenzymatic cofactor factor Va, and anionic phospholipid (primarily phosphatidylserine (PS)) exposed on the surface of activated platelets or other cells. Thrombin generated in the extrinsic pathway initiates an additional feedback amplification loop by generating factors Va and VIIIa by limited proteolysis. These surface-bound protein cofactors serve as scaffolds for the assembly of complexes that facilitate the production of factors Xa and thrombin (the Xase and prothrombinase complexes, respectively).

The Intrinsic Pathway

Thrombin enhances flux through the coagulation cascade by activating the intrinsic pathway of coagulation. In this arm of the cascade, thrombin generates factor XIa from its zymogen form (XI) by limited proteolysis. Factor XIa then generates factor IXa by a similar mechanism. The precise physiological role of the intrinsic pathway has been the subject of considerable debate, particularly because defects or deficiencies of factor XI are mild, and are associated with a form of hemophilia that is quite distinct from classical hemophilia. Bleeding due to factor XI deficiency is typically associated only with major surgery or trauma, whereas deficiencies of factor VIII (hemophilia A) or factor IX (hemophilia B) give rise to spontaneous soft tissue and joint bleeding. Thus, whereas the extrinsic pathway is required in most cases of vessel injury, the intrinsic pathway appears to be necessary for sustained thrombin generation in response to severe hemostatic challenges.

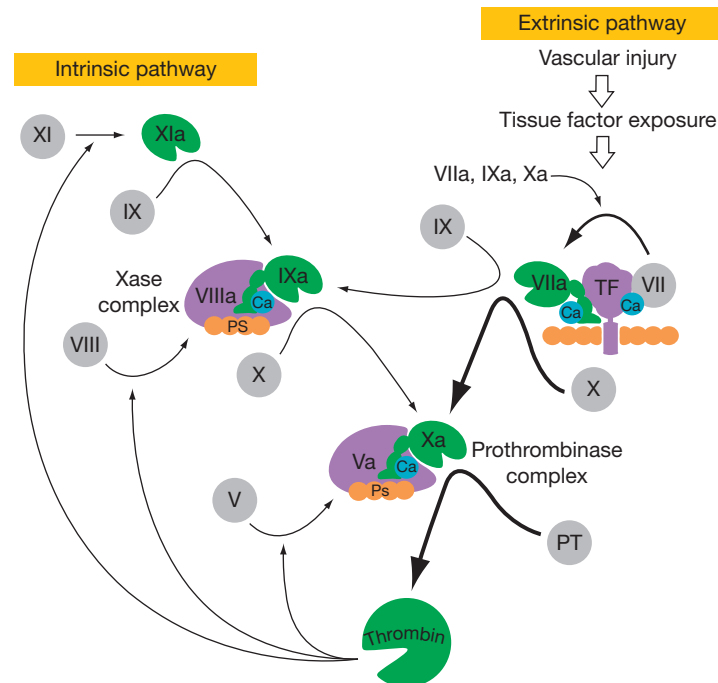


Figure 1 The blood coagulation cascade. Inactive cofactors and inactive (zymogen) forms of serine proteases are represented by gray circles. Active proteases are shown in green. Active protein cofactors are shown in purple. Phospholipid membranes are shown in orange. Proteolytic reactions are indicated by solid arrows. Ca, divalent calcium ions; PS, membrane surface-exposed phosphatidylserine; TF, tissue factor; PT, prothrombin.

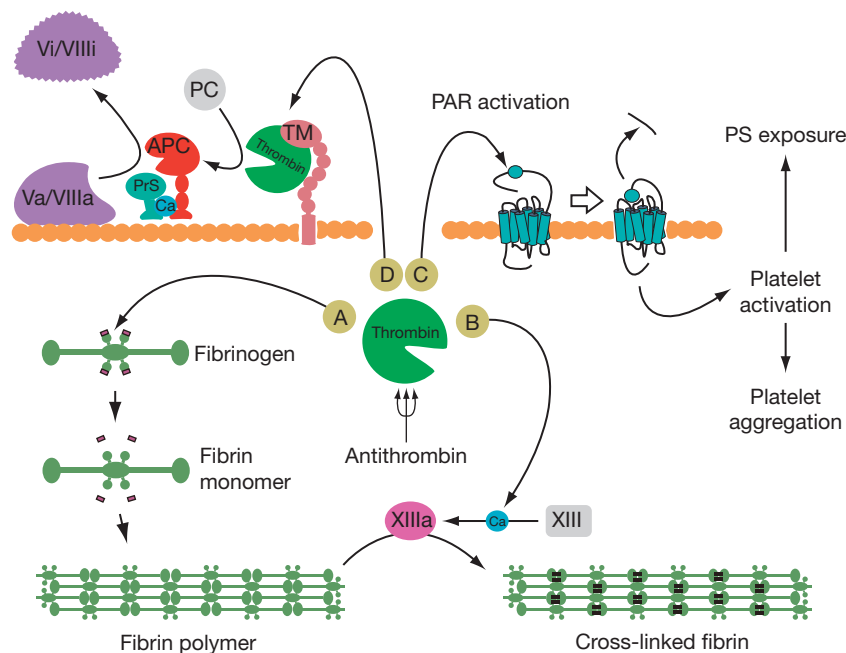


Figure 2 Effects of thrombin. Arrow A: Thrombin cleaves short fibrinopeptides from the central region of fibrinogen. The newly exposed termini spontaneously associate with the terminal domains of adjacent molecules allowing formation of polymeric fibrin. Arrow B: Proteolytic cleavage of factor XIII converts it into an active transglutaminase (XIIIa) that chemically cross-links adjacent fibrin molecules, providing structural integrity to the clot. Arrow C: Thrombin cleaves the extracellular region of protease-activated receptors (PARs) on the platelet surface. This modification exposes a tethered ligand that auto-activates the PAR, leading to platelet activation. Activated platelets aggregate to form a hemostatic plug and expose phosphatidylserine (PS), providing a procoagulant surface for the coagulation cascade. Arrow D: Feedback downregulation of the coagulation cascade. Thrombin binds to thrombomodulin on the surface of endothelial cells. This association alters the substrate specificity of thrombin such that it cleaves protein C (PC) to form activated protein C (APC). In the presence of calcium and the cofactor protein S (PrS), APC promotes anticoagulation by proteolytically converting factors Va and VIIIa to their inactive forms (Vi and VIIIi).

The Role of Thrombin

Thrombin is a multifunctional enzyme that employs limited proteolysis to effect a variety of biochemical reactions at the site of vascular injury (Figure 2). Thrombin has many physiological roles, such as promoting the formation of a cross-linked fibrin clot, the activation and aggregation of platelets, and, ultimately, the attenuation of the clotting cascade.

Fibrin Polymerization

Fibrinogen is one of the more abundant proteins in the blood, circulating at a concentration of $\sim 3 \text{ mg ml}^{-1}$. The fibrinogen molecule consists of six individual polypeptide chains arranged roughly in a dumbbell shape, with the amino termini localized to a central core and the carboxy termini located at each of the two ends. The carboxy terminal end comprises binding pockets that are unoccupied in the circulating fibrinogen molecule. Thrombin converts fibrinogen to fibrin by selectively cleaving short peptides (fibrinopeptides) from the central core region, thereby exposing new amino termini that interact with the carboxy terminal binding pockets on adjacent fibrin molecules (Figure 2, arrow A). In this way, fibrin monomers associate to form thin fibrils that thicken to form an insoluble fibrin clot.

Fibrin Cross-Linking by Factor XIIIa

When fibrin is formed, it accelerates the conversion of factor XIII to XIIIa in a reaction catalyzed by thrombin in the presence of calcium ions (Figure 2, arrow B). Factor XIIIa is unique among the coagulation enzymes in that it is not a serine protease; rather, it is a transglutaminase that covalently links the side chains of lysine and glutamine residues present in the carboxy terminal regions of adjacent fibrin molecules. This cross-linking imparts structural integrity and rigidity to the clot.

Platelet Activation

Thrombin activates platelets by limited proteolysis of proteins called protease-activated receptors (PARs). PARs constitute a small subset of a much larger family of G-protein-coupled receptors (GPCRs). GPCRs are seven-pass transmembrane proteins that transduce extracellular signals via intracellular activation of heterotrimeric G-protein signaling pathways. The PARs are unique among the GPCRs in that they utilize a tethered ligand mechanism in which limited proteolysis within the extracellular region of the receptor unmasks a protein sequence that auto-activates the receptor itself (Figure 2, arrow C). In platelets, the proteolysis of PARs 1 and 4 by thrombin results in the expression of the hallmarks of platelet activation, namely, morphological change, degranulation, the expression of surface proteins required for aggregation, and the translocation of PS from the inner to the outer leaflet of the platelet membrane.

The exposure of PS has dramatic consequences for amplification of the coagulation cascade. Under normal cellular homeostasis, PS is tightly restricted to the inner leaflet of the plasma membrane and, as such, is not exposed to the general circulation. However, the exposure of PS (whether through

translocation on the platelet surface or by mechanical disruption of cells lining the vasculature) provides a surface for the recruitment of the components of the coagulation cascade, including the serine proteases and factors V and VIII. Viewed in this light, platelet activation and subsequent PS exposure provide yet another positive feedback loop for the generation of thrombin.

Attenuation of the Coagulation Cascade

The unregulated activation of the coagulation cascade would have serious consequences for an organism, including the occlusion of blood vessels and the formation of clots at locations other than the site of injury. Numerous serine protease inhibitors are present in the blood, and serve to restrict coagulation to sites of vessel injury. The most abundant among these is antithrombin, which is present at a concentration ~ 3 times that of prothrombin. Antithrombin rapidly and irreversibly inactivates thrombin, particularly in the presence of heparan sulfate, a proteoglycan present in the blood vessel wall. Antithrombin has also been shown to inhibit factors IXa, Xa, and XIa.

In an alternative mechanism for attenuating the coagulation cascade, thrombin initiates a pathway that leads to the destruction of factors Va and VIIIa. In this pathway, thrombin binds to thrombomodulin, an integral membrane protein located on the vascular endothelium (Figure 2, arrow D). As the name implies, thrombomodulin modulates the specificity of thrombin such that it generates activated protein C (APC), by limited proteolysis of protein C. APC then inactivates both factors Va and VIIIa by minor proteolysis in a reaction that requires calcium ions, anionic phospholipid, and the protein cofactor, protein S.

Structure and Function of the Blood-Clotting Proteases

Similarity to Trypsin

The proteases involved in blood coagulation are functionally and structurally related to trypsin, the prototypical serine protease of the digestive system. Like trypsin, the proteases involved in blood coagulation (thrombin, APC, and factors VIIa, IXa, Xa, and XIa) constitute a class of serine proteases that cleave their target proteins after basic (arginine and lysine) amino acid residues. However, trypsin cleaves nonspecifically after basic residues, whereas proteases of the coagulation cascade recognize and cleave only after basic residues surrounded by specific sequences within their targets. As is the case with the clotting proteases, trypsin exists in a zymogen form (trypsinogen) that is activated by limited proteolysis. Thus, while trypsin does not possess the high degree of substrate specificity of the coagulation proteases, it shares the same general mode of zymogen activation and mechanism of catalysis.

The three-dimensional representation of trypsinogen solved by X-ray crystallography reveals a two-lobed structure resembling a clamshell with a cleft running between the two individual halves (Figure 3). The primary catalytic machinery of trypsin, the so-called 'catalytic triad', consisting of histidine, aspartic acid, and serine residues, resides within this cleft. (It is the catalytic serine residue from which the serine proteases

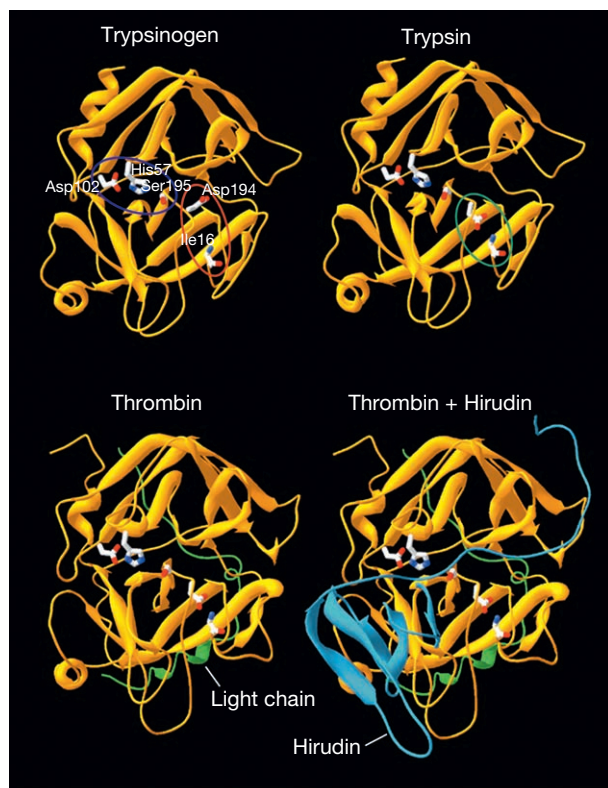


Figure 3 Related structures of trypsinogen, trypsin, and thrombin. The structure of trypsinogen (upper left panel) reveals clamshell topology with a cleft harboring the active site. Residues His57, Asp102, and Ser195 constitute the catalytic triad (blue oval). In the inactive (zymogen) form of the enzyme, the side chain of Asp 194 does not contact the amide nitrogen (blue atom) of Ile16 (bounded by red oval). The preceding 15 residues have been omitted from the structure to provide ease of interpretation. The structure of trypsin (upper right panel) illustrates the structural change upon zymogen activation. Here, the newly formed amino terminus at Ile16 forms a salt bridge with the side chain of Asp194. The structure of thrombin (lower left panel) illustrates its overall topological similarity with that of trypsin. The light chain of thrombin is shown in green. The structure of thrombin bound to the leech anticoagulant protein hirudin (lower right panel) demonstrates the molecular basis for hirudin's anticoagulant effect. Hirudin (depicted as a blue ribbon) blocks the active site of thrombin and extends an arm around thrombin to block its fibrinogen-binding exosite.

derive their name.) Trypsinogen is catalytically inactive until it is cleaved just prior to an isoleucine residue located near the amino terminus of the molecule. This proteolytic event triggers a subtle rearrangement of the molecule, driven by the generation of an electrostatic interaction between the positively charged amino group at the new amino terminus and a negatively charged aspartic acid residue adjacent to the catalytic serine residue. The formation of this salt bridge alters the structure of trypsin and renders it catalytically active.

The structure of thrombin is remarkably similar to that of trypsin (Figure 3), as are those of the other coagulation proteases. Thrombin shares with trypsin its overall clamshell structure, its catalytic triad, and its mode of zymogen activation upon limited proteolysis. However, in addition to the trypsin-like catalytic domain (heavy chain), thrombin possesses an

additional polypeptide (the light chain) – a property shared by the other clotting proteases. In addition, thrombin possesses a secondary binding site on its surface that confers specificity for both fibrinogen and PAR1. The position of this so-called ‘fibrinogen-binding exosite’ is illustrated in the structure of thrombin complexed with the leech salivary protein hirudin, which exerts its anticoagulant effect by simultaneously blocking access to the active site and occupying the fibrinogen-binding exosite with a long arm (Figure 3).

Gla Domains

The presence of both a catalytic heavy chain and a light chain is a common feature of the activated coagulation proteases. However, in contrast to thrombin, which has a relatively short light chain, the light chains of factors VIIa, IXa, Xa, and APC are longer and are characterized by the presence of multiple modular protein domains. (The crystal structure of factor IXa illustrates this arrangement; Figure 4(a).) When viewed in the same orientation as previously depicted structures of trypsinogen, trypsin, and thrombin, the serine protease domain with its characteristic catalytic triad is clearly visible. However, this orientation obscures the light chain, which consists of an amino terminal Gla domain followed by two tandem modules that bear structural and sequence similarity to epidermal growth factor (EGF1 and EGF2). The modular structure of the factor IXa light chain is visible when the protein is rotated (Figure 4(b)).

The Gla domain derives its name from a modification of glutamate (Glu) residues that occur during the biosynthesis of factors VII, IX, and X, prothrombin, and protein C. The amino terminal 45 residues of these proteins contain between 9 and 13 Glu residues that are converted to γ -carboxyglutamate (Gla) residues in a reaction that requires vitamin K as a cofactor. Gla residues are capable of binding divalent ions, particularly calcium. The three-dimensional structure of factor IXa, determined in the absence of calcium, reveals an amino terminal Gla domain that is relatively disordered, with Gla residues projecting into solution. However, at physiological calcium levels found in the blood, the Gla residues coordinate with calcium and drive the Gla domain into its native conformation (Figure 4(c)). In this form, the Gla domain is organized around a nearly planar array of calcium ions and projects a hydrophobic ‘thumb’ into solution. It is believed that the hydrophobic thumb associates with the hydrophobic regions of membranes, while the positively charged calcium ions associate electrostatically with the negatively charged head groups of PS exposed on coagulation-promoting surfaces such as the surface of activated platelets.

Gla domains are essential for localizing factors VIIa, IXa, and Xa, and APC, as well as their zymogen forms, to PS on the surface of activated platelets located at sites of vascular injury. (Thrombin is unique in that its Gla domain is removed from the light chain during the proteolytic activation of prothrombin.) The essential role of Gla domains and the Gla residues that they bear in directing the components of the coagulation cascade to PS-containing surfaces is underscored by the anticoagulant effect of warfarin, a potent vitamin K antagonist. Warfarin inhibits the conversion of Glu residues to Gla residues, rendering the clotting factors incapable of binding to the

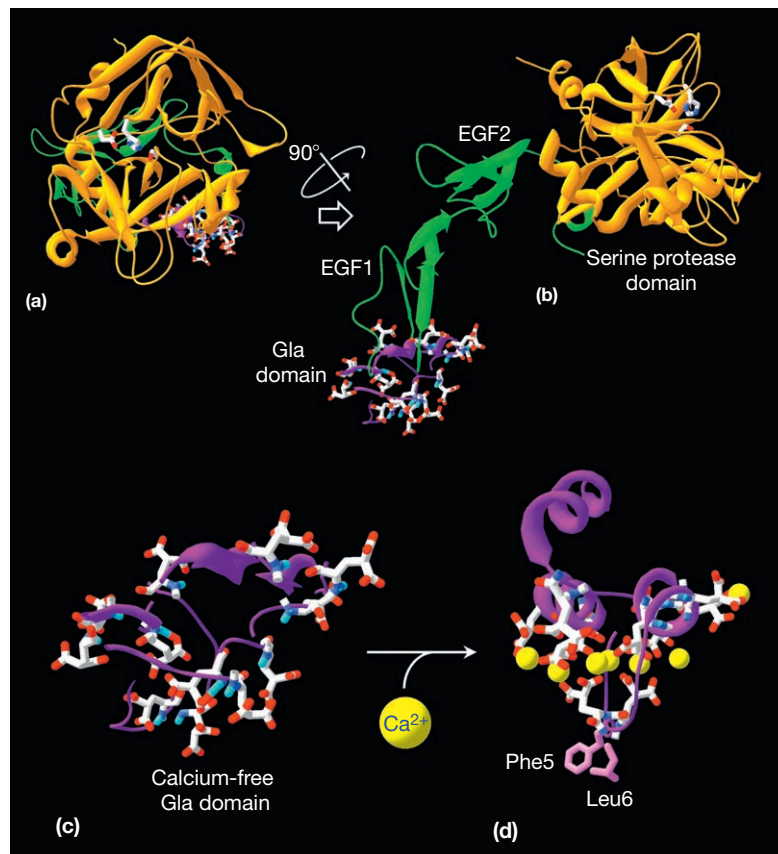


Figure 4 The role of Gla domains in the blood-clotting proteases. (a) Structure of factor IXa showing the serine protease domain with amino acid residues of the catalytic triad highlighted. In this view, the light chain, consisting of a Gla domain and two tandem EGF domains, is obscured. The factor IXa structure depicted was determined in the absence of calcium ions. (b) Rotation of the structure reveals the amino terminal Gla domain as well as the EGF domains. (c) Closeup of the Gla domain of calcium-free factor IXa. In the absence of calcium, the Gla domain is relatively disordered with the Gla residues projecting outward into solution. (d) The calcium-bound form of a Gla domain. The coordination of calcium ions by Gla residues within the core of the domain drives the exposure of a hydrophobic 'thumb' consisting of a phenylalanine and a leucine residue. This region of the domain contacts the hydrophobic region of membranes. The nearly planar arrangement of positively charged calcium ions may provide specificity for negatively charged phospholipids such as PS.

platelet surface. Originally employed at high doses as a rodent poison, warfarin is currently one of the most widely prescribed drugs for depressing the coagulation cascade in patients who are at risk of heart attack or stroke.

See also: [Signaling: Platelet-Activating Factor Receptor.](#)

Further Reading

Brandstetter H, Bauer M, Huber R, Lollar P, and Bode W (1995) X-ray structure of clotting factor IXa: Active site and module structure related to Xase activity and hemophilia B. *Proceedings of the National Academy of Sciences of the United States of America* 92: 9796–9800.

Davie E, Fujikawa K, and Kisiel W (1991) The coagulation cascade: Initiation, maintenance, and regulation. *Biochemistry* 30: 10364–10370.

Esmon C (2000) Regulation of blood coagulation. *Biochimica et Biophysica Acta* 1477: 349–360.

Fehlhammer H, Bode W, and Huber R (1977) Crystal structure of bovine trypsinogen at 1.8 angstrom resolution: II. Crystallographic refinement, refined crystal structure and comparison with bovine trypsin. *Journal of Molecular Biology* 111: 415–438.

High K and Roberts H (eds.) (1995) *Molecular Basis of Thrombosis and Hemostasis*. New York, NY: Marcel Dekker.

Macfarlane S, Seatter M, Kanke T, Hunter G, and Plevein R (2001) Proteinase-activated receptors. *Pharmacological Reviews* 53: 245–282.

Scarborough R, Naughton M, Teng W, et al. (1992) Tethered ligand agonist peptides: Structural requirements for thrombin receptor activation reveal mechanism of proteolytic unmasking of agonist function. *Journal of Biological Chemistry* 267: 13146–13149.

Soriano-Garcia M, Padmanabhan K, de Vos A, and Tulinsky A (1992) The Ca²⁺ ion and membrane binding structure of the Gla domain of Ca-prothrombin fragment 1. *Biochemistry* 31: 2554–2566.

Vitali J, Martin P, Malkowski M, et al. (1992) The structure of a complex of bovine alpha-thrombin and recombinant hirudin at 2.8 angstrom resolution. *Journal of Biological Chemistry* 267: 17670–17678.

Proteasomes, Overview

M Rechsteiner, University of Utah, Salt Lake City, UT, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 464–468, © 2004, Elsevier Inc.

Glossary

Hybrid proteasome A 20S proteasome with a regulatory complex (RC) at one end and a proteasome activator at the other, essentially a 26S proteasome bound to a proteasome activator.

Immuno-proteasome A 20S proteasome in which each of the three active β -subunits are replaced by interferon- γ -inducible catalytic subunits.

Proteasome activators Single polypeptide chains or small protein complexes that bind 20S proteasomes and stimulate peptide hydrolysis.

Regulatory complex (RC) A particle containing six adenosine triphosphatases (ATPases) and 12 other subunits that binds the 20S proteasome to form the 26S proteasome, the eukaryotic cell's major ATP-dependent protease.

Ubiquitin A small highly conserved protein that can be covalently attached to other proteins and to itself. Chains of ubiquitin target proteins for destruction by the 26S proteasome.

Proteasomes are cylindrical, multisubunit proteases found in eukaryotes, eubacteria, and archaeobacteria. Eukaryotic proteasomes come in two sizes, the 20S proteasome and the considerably larger adenosine triphosphate (ATP)-dependent 26S proteasome formed when the 20S proteasome binds a regulatory complex (RC). The 26S proteasome is the central protease in the ubiquitin (Ub)-mediated proteolytic pathway and is essential for a vast array of cellular processes, including cell-cycle traverse, transcription control, regulation of enzyme levels, and antigen presentation. Both 20S and 26S proteasomes can associate with protein complexes that activate peptide hydrolysis and may serve to localize the enzymes within cells.

The 20S Proteasome

Structure

The 20S proteasome is a cylindrical particle consisting of 28 subunits arranged as four rings of heptamers stacked upon one another (see [Figure 1](#)). The simpler archaeobacterial enzyme is composed of just one type of α -subunit and one type of β -subunit which form the end rings and the two central rings, respectively. Proteolysis is the province of the β -subunits, and their active sites face a central chamber within the cylinder. The α -subunit rings seal off the proteolytic chamber from the external solvent making the 20S proteasome a perfect protease to have among the critical proteins that comprise nucleus and cytoplasm. Unless a native protein is forced into the proteasome interior, it will be impervious to the enzyme. Moreover, proteasome β -subunits are not catalytically active unless they are present in the holoenzyme.

Eukaryotic proteasomes maintain the overall structure of the archaeobacterial enzyme, but they exhibit a more complicated subunit composition. There are seven different α -subunits and at least seven distinct β -subunits. Although current evidence indicates that only three of its seven β -subunits are catalytically active, the eukaryotic proteasome cleaves a wider range

of peptide bonds. The archaeobacterial enzyme, with its 14 identical β -subunits, preferentially hydrolyzes peptide bonds following hydrophobic amino acids and is therefore said to have chymotrypsin-like activity. By contrast, the eukaryotic proteasome contains two copies each of trypsin-like, chymotrypsin-like, and post-glutamyl-hydrolyzing subunits. For this reason, it is capable of cleaving almost any peptide bond having difficulty only with Pro-X, Gly-X, and, to a lesser extent, with Gln-X bonds. There is a further complication in higher eukaryotes. In immune tissues, or after exposure to interferon γ , the three catalytically active β -subunits found in the proteasomes of most organs are replaced by β -subunits with differing substrate specificities. Thus, in higher vertebrates there are constitutive and immuno-proteasomes. The latter are thought to play an important role in class I antigen presentation.

Enzyme Mechanism and Proteasome Inhibitors

Whereas most proteases use serine, cysteine, or metals to cleave peptide bonds, the proteasome employs N-terminal threonines in an unusual catalytic mechanism. The N-terminal threonines, generated by self-removal of short peptide extensions from the active β -subunits during assembly of the 20S particle, act as nucleophiles during peptide hydrolysis. There are highly specific inhibitors of the proteasome. The fungal metabolite lactacystin and the bacterial product epoxomicin covalently modify the active site threonines and inhibit the enzyme. Both compounds are commercially available; other inhibitors include vinylsulfones and various peptide aldehydes that are generally less specific.

The 26S Proteasome

Whereas bacteria may possess as many as five ATP-dependent proteases, the 26S proteasome is the only ATP-dependent protease discovered so far in the nuclear and cytosolic

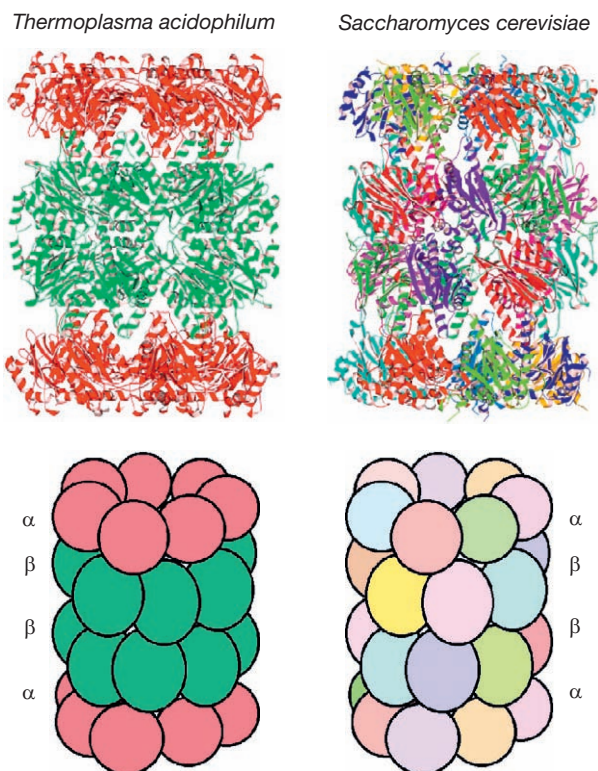


Figure 1 20S proteasomes. At the top are ribbon diagrams of an archaeobacterial proteasome (left) and a eukaryotic proteasome (right) as revealed by X-ray crystallography. Schematic representation of their subunit arrangements is shown below each crystal structure. Note that the archaeobacterial proteasome is assembled from 14 copies of a single α -subunit and 14 copies of the same β -subunit. By contrast, the yeast 20S proteasome consists of two copies each of seven different α - and seven different β -subunits. From Lowe J, Stock D, Jap B, et al. (1995) Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* 268: 533–539; and Groll M, Ditzel L, Lowe J, et al. (1997) Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386: 463–471.

compartments of eukaryotic cells. As the 20S proteasome's internal cavities are inaccessible to intact proteins, openings must be generated in the enzyme's outer surface for proteins to be degraded. A number of protein complexes have been found to bind the proteasome and stimulate peptide hydrolysis. The most important of the proteasome-associated components is the 19S RC, for it is a major part of the 26S ATP-dependent enzyme that degrades ubiquitylated proteins in eukaryotic cells (see **Figure 2**). As its name suggests, this 19S particle is roughly the same size as the 20S proteasome, except that it is a more complex protein assemblage containing 18 different subunits. Among the RC subunits are six ATPases, an isopeptidase, a subunit that recognizes polyUb chains, and a number of subunits whose functions are yet to be determined. RC subunits are arranged in two large subcomplexes, called the base and the lid. The ATPases and two other subunits comprise the base that sits directly on the end rings of the 20S proteasome; the lid is separated from the base by what appears to be a cavity. Overall, the RC looks much like a Chinese dragonhead in negatively stained electromagnetic (EM) images.

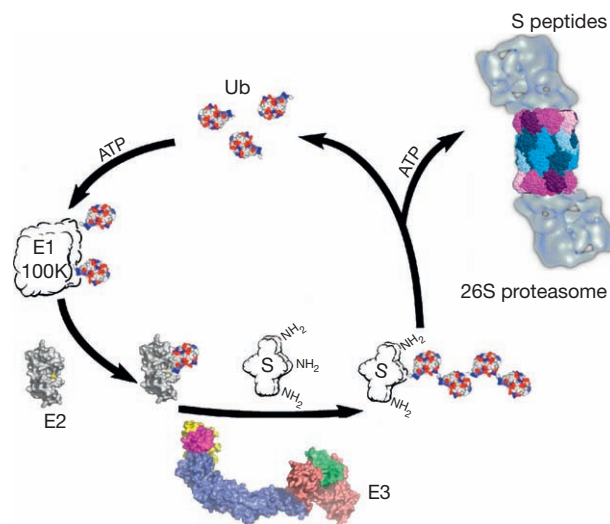


Figure 2 Schematic representation of ubiquitin (Ub) activation and adenosine triphosphate (ATP)-dependent proteolysis of conjugated substrates. Ub molecules are activated by an E1 enzyme in an ATP-dependent reaction, transferred to a cysteine residue (yellow) on an E2 or Ub carrier protein, and subsequently attached to a substrate protein (S) by an E3 or Ub ligase. Note that chains of Ub are generated on the substrate and these are recognized by the 26S proteasome depicted in the upper right at 1/20th scale.

Presumed Mechanism

The six ATPases in the 19S RC are members of the large family of AAA nucleotidases, whose members share the common property of altering the conformation or association state of proteins. In the case of the 19S RC ATPases, they are thought to act as unfoldases and translocases. Presumably, ubiquitylated proteins are captured by the RC after which the ATPases unfold the bound substrate protein and thread it through the α -ring of the 20S proteasome for its subsequent degradation in the central proteolytic chamber (see **Figure 3**). Most of the resulting peptide fragments are 5–10 residues in length, but fragments as long as 35 amino acids can be present. Furthermore, in some cases, the 26S proteasome only partially degrades the substrate protein, releasing instead functionally processed domains.

The mechanism by which the 26S proteasome recognizes ubiquitylated proteins has not been firmly established. Ub chains containing four or more monomers are preferred substrates, and one RC subunit, called S5a, has been shown to bind polyUb. However, deletion of the gene encoding this subunit in budding yeast has only a minor impact on intracellular proteolysis. Hence, there must be other subunits or additional mechanisms, by which the 26S enzyme recognizes polyUb substrates. In this regard, there are recent reports that the 26S proteasome interacts directly with Ub ligases or ligase-associated proteins that are ubiquitylated but not degraded. It should be noted that some nonubiquitylated proteins are also degraded by the 26S proteasome.

Tissue and Subcellular Distribution of Proteasomes

Proteasomes are found in all organs of higher eukaryotes, but the degree to which the composition of proteasomes and their

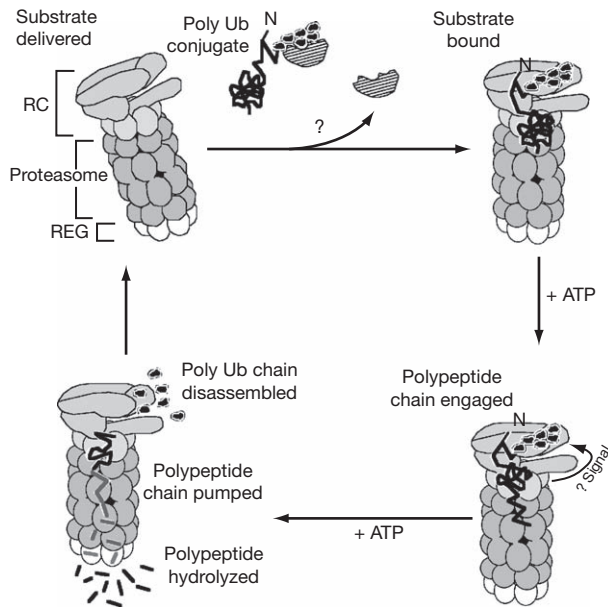


Figure 3 Hypothetical reaction cycle for the 26S proteasome. A polyubiquitylated substrate is delivered to the 26S proteasome, possibly by chaperones (step 1). Substrate is bound by polyUb recognition components of the regulatory complex (RC) until the end of the polypeptide chain is engaged by the adenosine triphosphatases (ATPases) (step 2). As the polypeptide chain is unfolded and pumped down the central pore of the proteasome, a signal is conveyed to the metallo-isopeptidase to remove the polyUb chain (step 3). The unfolded polypeptide is eventually degraded within the inner chamber of the proteasome (step 4) and peptide fragments exit the enzyme.

activators varies among tissues is largely unexplored territory. At the subcellular level, 26S proteasomes are present in cytosol and nucleus where they appear to be freely diffusible. They are not found in the nucleolus or within membrane-bound organelles other than the nucleus. When large amounts of misfolded proteins are synthesized by a cell, they often accumulate around the centrosome in what are called 'aggresomes'. Under these conditions, 26S proteasomes, chaperones, and proteasome activators also redistribute to the aggresomes, presumably to refold and/or degrade the misfolded polypeptides.

Physiological Importance

Deletion of yeast genes encoding 20S proteasome and 19S RC subunits is usually lethal, indicating that the 26S proteasome is required for eukaryotic cell viability. Known substrates of the 26S proteasome include transcription factors, cell-cycle regulators, protein kinases, etc. – essentially, most of the cell's important regulatory proteins. Surprisingly, even proteins secreted into the endoplasmic reticulum are returned to the cytosol for degradation by the 26S proteasome. Given the scope of its substrates, it is hardly surprising that in higher eukaryotes, the Ub–proteasome system affects a vast array of physiological processes ranging from cell-cycle traverse to circadian rhythms to learning. The Ub–proteasome system impacts a number of human neurological diseases that include Parkinson's, Huntington's, and Alzheimer's. Mutations in components of the Ub–proteasome pathway also cause various other diseases.

Proteasome Activators

In addition to the RC, there are two protein complexes, 11S regulator (REG) $\alpha\beta$ and REG γ , and a single polypeptide chain, PA200, which bind and stimulate peptide but not protein degradation by the proteasome (see Figure 4). A fourth protein, called ecm29 in yeast, associates with the proteasome. Like the RC, proteasome activators bind the ends of the 20S proteasome, and, importantly, they can form mixed or hybrid 26S proteasomes in which one end of the 20S proteasome is associated with a 19S RC and the other is bound to a proteasome activator. This latter property raises the possibility that proteasome activators serve to localize the 26S proteasome within eukaryotic cells. At present, we do not know whether activation, or targeting, or something else is the primary function of proteasome activators.

REGs or PA28s

There are three distinct subunits of the REG called α , β , and γ . REG α and REG β form hetero-heptamers found principally in the cytoplasm, whereas REG γ forms a homo-heptamer located in the nucleus. REG α and β are abundantly expressed in immune tissues, whereas REG γ is highest in brain. In addition to their different locations, the REGs differ in their activation properties. REG $\alpha\beta$ activates all three proteasome active sites; REG γ only activates the trypsin-like subunit. There is reasonably solid evidence that REG $\alpha\beta$ plays a role in class I antigen presentation, but we have no idea what REG γ does, especially since REG γ knockout mice have almost no phenotype.

The crystal structure of REG α reveals seven subunits that form a donut-shaped structure with a central aqueous channel, and the structure of an REG–proteasome complex provides important insight into the mechanism by which REG α activates the proteasome. The carboxyl tail on each REG subunit fits into a corresponding cavity on the α -ring of the proteasome, and loops on the REG subunits cause N-terminal strands on several proteasome α -subunits to reorient upward into the aqueous channel of the REG heptamer. These movements open a continuous channel from the exterior solvent to the proteasome central chamber.

PA200

A new proteasome activator, called PA200, was recently purified from bovine testis. Human PA200 is a nuclear protein of 1843 amino acids that activates all three catalytic subunits with some preference for the peptidyl–glutamyl peptide hydrolyzing (PGPH) active site. Homologs of PA200 are present in budding yeast, worms, and plants. A single chain of PA200 can bind each end of the proteasome, and, when bound, PA200 molecules look like volcanoes in negatively stained EM images. Mutation of yeast PA200 results in sensitivity to the DNA-damaging agent bleomycin, and evidence from both yeast and mammals indicates that PA200 is involved in DNA repair.

ECM29

Another proteasome-associated protein, called ecm29p, has been identified in several recent proteomic screens. It has

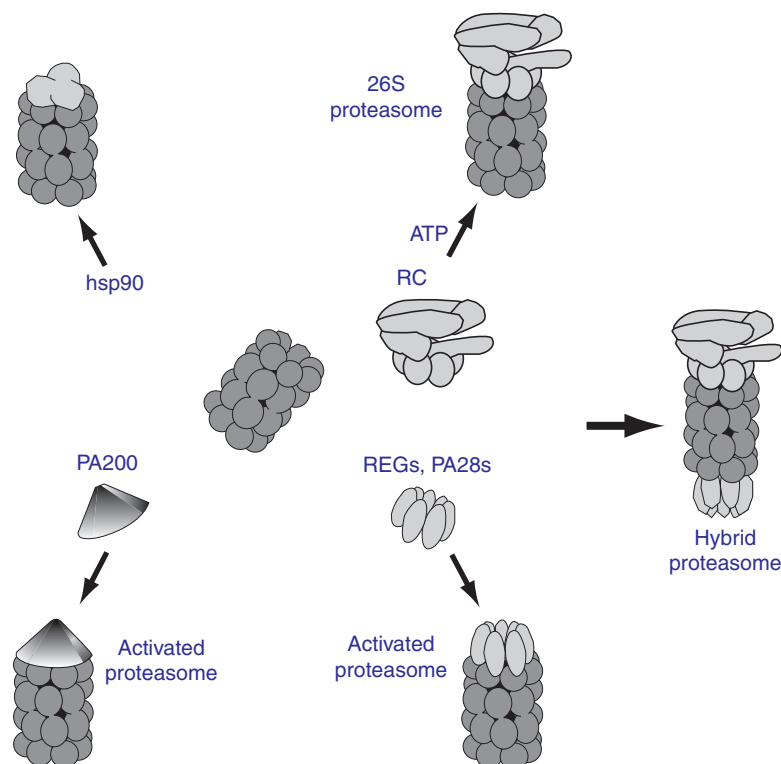


Figure 4 Schematic representation of the 20S proteasome assembling with various activator proteins (regulatory complex (RC), REGs, and PA200) and with the chaperone hsp90, a protein that inhibits the enzyme.

been reported that ecm29p serves to stabilize the yeast 26S proteasome, but the human homolog of ecm29p is found predominately associated with secretory and endocytic membranes, as per an unpublished observation by Gorbea and his associates, a location suggesting a role in secretion rather than 26S proteasome stability. Ecm29p clearly associates with the 26S proteasomes; whether it activates proteasomal peptide hydrolysis is unknown.

Protein Inhibitors of the Proteasome

Two proteins have been found to suppress proteolysis by the proteasome: one of these is PI31 and the other is the abundant cytosolic chaperone, Hsp90. Both may affect how the proteasome functions in class I antigen presentation. PI31 is a 30-kDa proline-rich protein that inhibits peptide hydrolysis by the 20S proteasome and can block activation by both RC and REG $\alpha\beta$. Although surveys of various cell lines show PI31 to be considerably less abundant than RC or REG $\alpha\beta$, when overexpressed, PI31 is reported to inhibit class I antigen presentation by interfering with the assembly of immuno-proteasomes. A number of studies have shown that Hsp90 can bind the 20S proteasome and inhibit its chymotrypsin-like and PGPH activities. Interestingly, inhibition by Hsp90 is observed with constitutive but not with immuno-proteasomes, a finding consistent with proposals that Hsp90 shuttles immuno-proteasome-generated peptides to the endoplasmic reticulum for class I presentation.

Summary

The 20S proteasome was discovered in 1980 and its 26S version 6 years later. Since the 1980s, a great deal of knowledge has been gained about these two enzymes and their central importance in eukaryotic cell physiology. Still, there is much more to discover, for example, the crystal structure of the 19S RC and the mechanism by which the 26S proteasome degrades its substrates; how the 26S proteasome is itself regulated; and the extent to which proteasomal components vary among tissues in higher eukaryotes.

See also: **Cell Architecture and Function:** 26S Proteasome: Structure and Function; **Protein/Enzyme Structure Function and Degradation:** Chaperonins; Ubiquitin System; Ubiquitin-Like Proteins.

Further Reading

- Bochtler M, Ditzel L, Groll M, Hartmann C, and Huber R (1999) The proteasome. *Annual Review of Biophysics and Biomolecular Structure* 28: 295–317.
- Groll M, Ditzel L, Lowe J, et al. (1997) Structure of 20S proteasome from yeast at 2.4Å resolution. *Nature* 386: 463–471.
- Hampton RY (2002) ER-associated degradation in protein quality control and cellular regulation. *Current Opinion in Cell Biology* 14: 476–482.
- Hartmann-Petersen R, Seeger M, and Gordon C (2003) Transferring substrates to the 26S proteasome. *Trends in Biochemical Sciences* 28: 26–31.

- Hershko A and Ciechanover A (1998) The ubiquitin system. *Annual Reviews in Biochemistry* 67: 425–479.
- Hill CP, Masters EI, and Whitby FG (2002) The 11S regulators of 20S proteasome activity. *Current Topics in Microbiology and Immunology* 268: 73–89.
- Lowe J, Stock D, Jap B, et al. (1995) Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* 268: 533–539.
- Taylor JP, Hardy J, and Fischbeck KH (2002) Toxic proteins in neurodegenerative disease. *Science* 296: 1991–1995.
- Voges D, Zwickl P, and Baumeister W (1999) The 26S proteasome: A molecular machine designed for controlled proteolysis. *Annual Reviews in Biochemistry* 68: 1015–1068.
- Yewdell JW and Bennink JR (2001) Cut and trim: Generating MHC class I peptide ligands. *Current Opinion in Immunology* 13: 13–18.

26S Proteasome: Structure and Function

F Förster and E Sakata, Max Planck Institute of Biochemistry, Martinsried, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

AAA ATPases ATPases associated with a variety of cellular activities are part of the AAA+ superfamily of ring-shaped P-loop NTPases. AAA proteins are characterized by a highly conserved module of about 240 residues, which contains several conserved motifs, such as the Walker A and B motifs. AAA proteins consists of an N-terminal non-ATPase module, considered to be the primary substrate recognition site, followed by either one or two AAA domains (D1 and D2).

DUBs Deubiquitylating enzymes perform proteolytic cleavage of ubiquitin-protein bonds. DUBs are divided into five superfamilies: (1) Jab1/Mov34/MPN (JAMM) motif containing metalloproteases (e.g., Rpn11); (2) ubiquitin-specific processing proteases (USP/UBP, e.g., Ubp6); (3) ubiquitin C-terminal hydrolyases (UCH, e.g., UCH37); (4) ovarian tumor (OTU); and (5) Machado–Joseph disease (MJD).

Ntn hydrolases N-terminal nucleophile hydrolases are enzymes with a conserved fold and a single-residue

N-terminal catalytic residue, threonine, serine, or cysteine, which mediates the nucleophilic attack on the substrate.

19S regulatory particle 19S regulatory particles bind to both ends of the 20S core particle. They regulate substrate access to the 20S core particle, for example, based on polyubiquitylation of substrates.

20S core particle The 20S core particle is situated in the center of the 26S proteasome and harbors the proteolytically active Ntn hydrolases in its core. The cylinder-shaped complex is assembled of four stacked rings, the seven-membered alpha and beta rings, both present in two copies.

Ubiquitin The highly conserved and ubiquitous, eukaryotic protein of 76 amino acids can be linked to the ϵ -amino group of a lysine residue in a protein via its C-terminal carboxyl group (ubiquitylation). Ubiquitin monomers can form polyubiquitin chains by binding to one of the seven lysines in ubiquitin, most frequently K-48.

Introduction

The 26S proteasome is an abundant complex in eukaryotes, primarily located in the nucleus, the cytoplasm, and to a lesser extent at the endoplasmic reticulum (ER) membrane. It is composed of at least 33 different subunits and is composed of two modules: a highly conserved 20S core particle (CP) containing a central catalytic cavity and one or two 19S regulatory particles (RPs). The 26S proteasome is the most downstream element of the ubiquitin proteasome pathway (UPP). In this pathway, a cascade of three enzymes, E1 ubiquitin-activating enzyme, E2 ubiquitin-conjugating enzyme, and E3 ubiquitin ligase performs attachment of a polyubiquitin tag to the substrate, which then serves as the degradation signal for the 26S proteasome. The UPP is the major route used by eukaryotic cells for the disposal of misfolded or damaged proteins and for controlling the lifespan of proteins. As a consequence, it regulates a plethora of fundamental cellular processes, such as protein quality control, DNA repair, or signal transduction. Due to its involvement in many intracellular pathways, for example, tumor suppression by p53, the 26S proteasome has emerged as a significant drug target.

To achieve substrate-specific degradation, the 26S proteasome is highly regulated. The intricate means of regulation are reflected in a complex molecular architecture where one or two copies of asymmetric 19S RPs bind to the end(s) of the barrel-shaped CP. In this article, we summarize current knowledge on the structure of the 26S proteasome and its regulation mechanisms to perform its cellular functions.

The 20S Proteasome

20S proteasomes are present in all domains of life, that is, archaea, bacteria, and eukaryotes. Whereas the 20S proteasome is conserved and essential in all archaea and eukaryotes sequenced to date, it is only found in few bacteria, mostly actinomycetes where it is not essential. In most bacteria, ClpP, which is evolutionary more distantly related to the proteasome, has an analogous function to the 20S proteasome.

Molecular Architecture

The 20S proteasomes share a highly conserved quaternary structure for species: two types of subunits, α and β , each constitute heptameric rings and collectively form a barrel-shaped complex, arranging in an $\alpha/\beta/\beta/\alpha$ -ring fashion (Figure 1). The holocomplex measures 15 nm in length and 11 nm in diameter. A channel traverses the particle from end to end and widens into three internal cavities, each ~ 5 nm in diameter. The complexity of each of the rings ranges from homo-heptameric (e.g., in the archaeon *Thermoplasma acidophilum*) to hetero-heptameric in all eukaryotes. When composed of homo-heptameric rings, the proteasome displays a sevenfold rotational symmetry (C_7) along the cylinder axis and C_2 symmetry orthogonal to the symmetry axis, whereas the eukaryotic proteasome preserves the C_2 symmetry. According to their location within the α and β rings, subunits are sequentially numbered $\alpha 1/\beta 1$ through $\alpha 7/\beta 7$. The structure-based nomenclature differs from the classification of subunits prior to

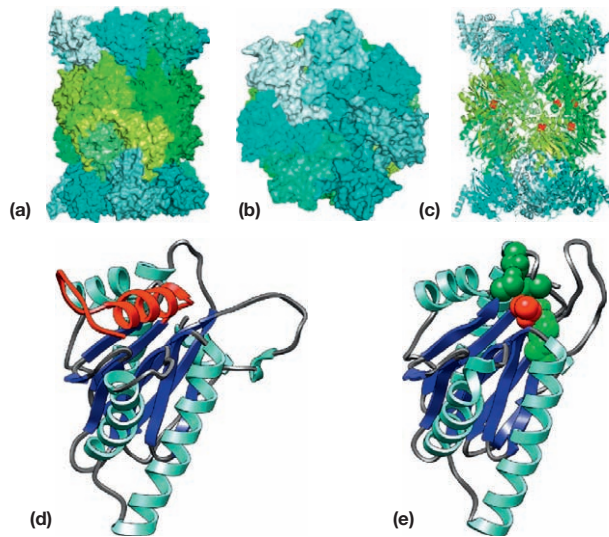


Figure 1 Structure of the 20S proteasome. (a) Surface view of crystal structure of the *Saccharomyces* CP (blue: α -subunits, green: β -subunits). (b) The same structure as in (a) seen from the top. (c) The same structure as in (a) displayed as a cartoon view. The 2 $\times$ 3 active sites are marked in red. (d) Fold of the *Thermoplasma acidophilum* 20S proteasome α -subunit (cyan: α -helix, blue: β -sheets). The N-terminal helix H0 that blocks the active site is shown in red. (e) Fold of the *Thermoplasma* 20S proteasome β -subunit. The N-terminal, catalytically active Thr residue is shown in red. Moreover, Glu17 and Lys33 (light green) are involved in the catalytic mechanism and Ser129, Asp166, and Ser169 (dark green) maintain conformational stability of Thr1.

structure determination, for example, the $\alpha 1$ subunit is referred to as ‘alpha type-6 proteasome subunit’ in databases, such as Uniprot.

Both types of subunits, α and β , are members of the N-terminal nucleophile (Ntn) hydrolase superfamily. The proteasome subunits are well conserved: the different subunits have at least 50% similarity, while there is only $\sim 30\%$ similarity between the seven subunits within the different subunit types in same species, and even less among the α - and β -subunits. Ntn hydrolases consist of two central antiparallel β -sheets, which are flanked on either side by α -helices and open at one end to form the active-site cleft (Figures 1(d) and 1(e)). The α -subunits possess an N-terminal extension of the Ntn core structure, filling the active-site cleft, which renders α -subunits proteolytically inactive (Figure 1(d)).

In the assembled proteasome, the inner cavity or nanocompartment formed by the two β -rings harbors the active sites (Figure 1(c)). Thus, access to the active sites is prohibited by the inactive α -rings forming a pore entrance to the catalytic chamber. This regulatory principle is also referred to as ‘self-compartmentalization’. ClpP has a simpler molecular architecture than the 20S proteasome: it consists only of the two β -rings and the shielding α -subunits are absent.

Active Site

A characteristic feature of Ntn hydrolases is a single-residue N-terminal active site (Thr/Ser/Cys). In eukaryotes, only three of the seven different subunits contain the catalytic threonine

residue and display proteolytic activity ($\beta 1$, $\beta 2$, and $\beta 5$), whereas subunits $\beta 3$, $\beta 4$, $\beta 6$, and $\beta 7$ lack one or more of the critical residues and are inactive (Figure 1(c)). The threonine residue acts as both proton acceptor and donor during hydrolysis: it donates a proton to its own α -amino group and attacks the carbonyl carbon of the substrate. After cleavage the α -amino group donates the proton to the nitrogen of the scissile peptide. Proton transfer is possibly mediated by a water cluster centered in the vicinity of the N-terminus. There is no consensus of residues in the vicinity of the active site among the Ntn hydrolases. Nevertheless, several residues, all highly conserved and in close proximity to Thr1, are critical for activity of the 20S proteasome, as consistently shown by site-directed mutagenesis with archaeal, bacterial, yeast, and mammalian proteasomes (Figure 1(e)). Interestingly, the different active sites of the CP subunits display different catalytic activity: $\beta 5$ has a chymotrypsin-like active site (preferential cutting after hydrophobic residues), $\beta 2$ displays trypsin-like degradation (basic residues), and $\beta 1$ is caspase like (acidic residues).

As is the case with all known Ntn hydrolases and many other proteases, proteasomal β -subunits are synthesized as an N-terminally extended, catalytically inactive precursor to prevent premature proteolytic activity. For the formation of the active sites, the so-called propeptides must be removed posttranslationally.

Immunoproteasome and Thymoproteasome

One notable feature of the CP is the tissue-specific replacement of proteolytic subunits, resulting in the so-called immunoproteasome and thymoproteasome, which generate different sets of peptides. Interferon gamma (IF γ) induces expression of the orthologs $\beta 1i$, $\beta 2i$, and $\beta 5i$, which are then part of the immunoproteasome. It contributes to the efficient production of peptide epitopes, which are then displayed at the cell surface by the major histability complex class I to cytotoxic T lymphocytes. In Thymus, $\beta 5i$ can be exchanged by a further ortholog, $\beta 5t$, which then gives rise to the thymoproteasome.

20S Proteasome Assembly

The assembly pathway of the 20S proteasome seems to vary for different species. Prokaryotic 20S proteasomal subunits can assemble autonomously into mature form. In *Rhodococcus*, α - and β -subunits form heterodimers prior to assembling to a $\alpha_{14}\beta_{14}$ half-proteasome. By contrast, *Thermoplasma* α -subunits form heptameric rings, which may then serve as a template for the assembly of the β -ring, also resulting in a half-proteasome. The propeptide serves as an intramolecular chaperone during the assembly process: the crystal structure of the *Rhodococcus* half-proteasome indicates that the pro-peptides link the β -subunits to two neighboring α -subunits, thus increasing the affinity of the β -subunits to the α -ring by increasing the area of intersubunit contacts.

On the other hand, the eukaryotic CP requires a set of chaperones for its assembly because of the higher complexity of its subunit composition: seven different α subunits and seven different β subunits are placed in a defined arrangement. Four distinct assembly chaperones (yeast: Pba1/human homolog: PAC1, Pba2/PAC2, Pba3/PAC3, and Pba4/PAC4) and Ump1

are required for the formation of CP. The Pbas form heteromeric dimers, Pba1–Pba2 and Pba3–Pba4, which bind specific α -subunits and promote assembly of the α -rings. For the assembly of β -subunits, the α -ring serves as a scaffold, on which β subunits assemble sequentially to form a half-proteasome aided by Ump1: Ump1 likely binds to the center of the assembled α -ring, where it coordinates the processing of β -subunits, assembly of the β -ring, and dimerization of half-proteasomes. Moreover, dimerization is intramolecularly chaperoned by the C-terminal tail of the $\beta 7$ subunit. After completed CP assembly, the mature CP degrades the propeptide, Ump1, and PAC1–PAC2.

CP Gate

The N-terminal tails of the α -subunits block access to the inner proteolytic compartment of the CP, even to small peptides and unfolded proteins. Therefore, the CP behaves, with respect to folded proteins, as a latent protease, which is dependent on regulatory complexes or activators that open the gate to the CP for its catalytic activity. Most common are ATP-dependent activators: the eukaryotic 19S RP contains six AAA–ATPase subunits that assemble to a hetero-hexameric ring. In archaea, the proteasome activating nucleotidase (PAN), a homo-hexameric AAA–ATPase complex, attaches to the α -ring and induces gate

opening. Likewise the homo-hexameric AAA–ATPase ARC activates the 20S proteasome in bacteria (AAA ATPase forming ring-shaped complex, ARC). These ATPase proteins possess a hydrophobic HbYX motif at their C-terminus, which binds to the hydrophobic pockets of the α -ring and induces gate opening.

In eukaryotes, two further RPs exist in addition to the 19S RP that open the gate to the CP in an ATP-independent manner: the heptameric complex PA28/11S regulates access to the immunoproteasome, and the resulting holocomplex generates the peptides presented by the major histocompatibility class (MHC) I complex. In addition, Blm10/PA200 binds to the CP in a similar way as PA28/11S, but its function is not resolved to date.

The 26S Proteasome

The 19S Regulatory Particle and Its Subcomplexes

The 19S RP is a multi-protein complex of approximately 900 kDa. It consists of 19 different canonical subunits: six regulatory particle AAA ATPases (Rpt1–6), and 13 regulatory particle non-ATPases (Rpn1–Rpn15); the proteins Rpn4 and Rpn14 were erroneously notated as integral subunits, but later turned out to be a proteasomal transcription factor and an assembly chaperone, respectively (Figure 2). Moreover, numerous

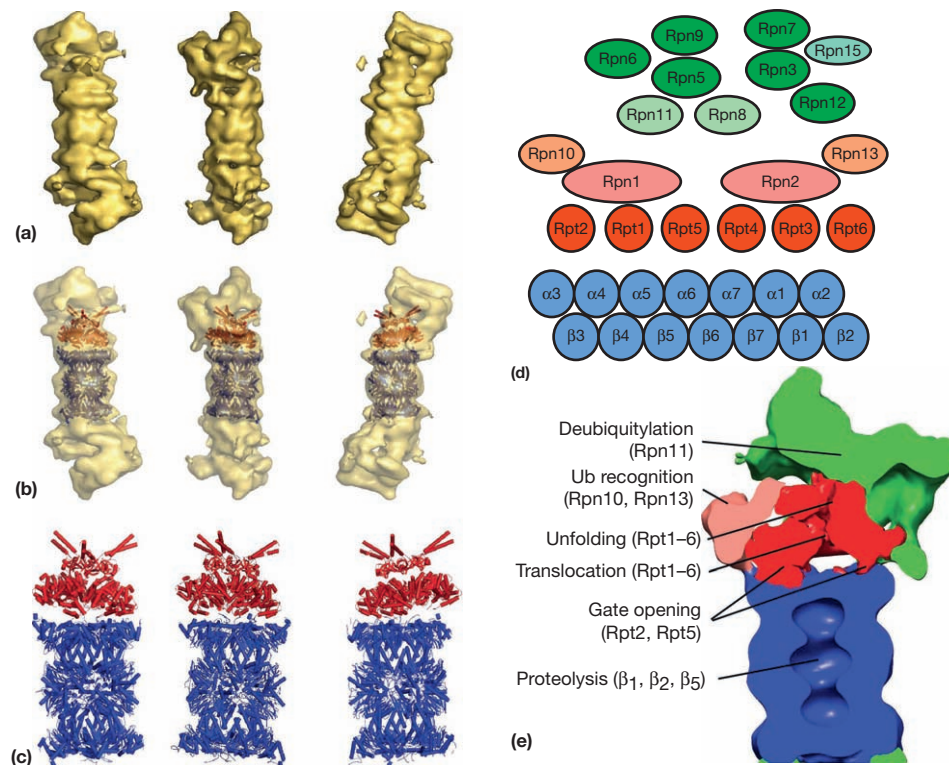


Figure 2 Structure of the 26S proteasome. (a) Cryo-EM map of the *Drosophila melanogaster* 26S proteasome resolved to 20 Å resolution. Three different views, each rotated by 120° around the vertical axis, are shown. (b) Atomic models of the CP (blue) and the AAA–ATPases (red) fitted into the map. (c) 2x magnified view of the 20S proteasome in complex with the AAA–ATPases. The pseudo sevenfold rotational symmetry axis of the CP and the pseudo threefold axis of the AAA–ATPase hexamer deviate substantially: the center of the AAA–ATPase hexamer is shifted by approximately 30 Å with respect to the CP axis and the AAA–ATPase axis is tilted by approximately 10°. (d) Canonical 26S proteasome subunits: subunits of CP (blue), base (dark red: ATPases, light red: PC-repeat containing subunits), and lid (dark green: PCI-module containing subunits, light green: MPN domain containing subunits, cyan: Rpn15) are ordered as suggested by protein–protein interactions. (e) Approximate locations of the different steps involved in substrate degradation.

proteasome interacting proteins (PIPs) interact transiently with the 26S proteasome. Among the widely studied PIPs are the deubiquitylating enzymes (DUBs), Ubp6 and UCH37 (in most eukaryotes, but no homolog in yeast), and the ubiquitin receptors, Rad23, Dsk2, and Ddi2. The 19S RP easily dissociates into a base complex comprising the AAA-ATPases Rpt1–6, Rpn1, Rpn2 and Rpn13, and often Rpn10, and a lid subcomplex containing the remaining non-ATPase subunits.

Base

The AAA-ATPases are in the heart of the RP. The AAA-ATPases assemble to a hetero-hexameric ring in the topology Rpt1/Rpt2/Rpt6/Rpt3/Rpt4/Rpt5. Structural studies of the archaeal homolog of Rpt subunits, PAN, provided implications for Rpt structure. The ATPase domains of the Rpt subunits form a ring that directly binds the α -subunits of the CP (Figures 2(b) and 2(c)). The N-terminal coiled coils and the succeeding oligosaccharide binding (OB) domains assemble to a smaller ring (N-ring), which is placed concentrically on top of the ATPase-ring, that is, distal to the CP. The ATPase-ring of Rpt subunits is stabilized by binding of the two largest RP subunits, Rpn1 and Rpn2 (~1000 residues), suggesting both function as scaffolding proteins. Both proteins, Rpn1 and Rpn2, consist to a large extent of small bihelical repeats, proteasome cyclosome (PC) repeats, which possibly wrap around the AAA-ATPase hexamer. Rpn13 and Rpn10 are known as Ub receptors, which bind to Rpn2 and Rpn1, respectively. In addition to the canonical subunits, two non-integral DUBs, UCH37 and Usp14/Ubp6, are often found substoichiometrically in 26S proteasome preparations. Ubp6 binds to Rpn1 and UCH37 binds to the C-terminal domain of Rpn13. *Saccharomyces cerevisiae* Rpn13 lacks the UCH37 binding domain and does not possess a UCH37 homolog.

Lid

The lid complex comprises proteins of two families: Rpn3, Rpn5, Rpn6, Rpn7, Rpn9, and Rpn12 contain a characteristic module, which can be found in subunits of the proteasome, COP9 complex, and eIF3 (proteasome COP9 eIF3 (PCI) module). Rpn8 and Rpn11 share an Mpr1 and Pad1 in the N-terminus (MPN) domain, which is also found in COP9 and eIF3. Among these proteins, only the function of Rpn11 is well understood: Rpn11 is a DUB and responsible for ubiquitin removal from substrates to enable ubiquitin recycling. Since the deubiquitylating activity of Rpn11 is ATP-dependent, Rpn11 is only active in the assembled RP. The MPN domain of Rpn8, which is unable to bind Zn ions, is catalytically inactive, whereas Rpn11 shows DUB activity. Interestingly, Rpn11 has a further function: the C-terminus is essential for the maintenance of tubular mitochondrial morphology. On the other hand, the function of the PCI module containing proteins is unknown. Nevertheless, the nucleic acid interacting winged helix motif, which is part of the PCI module, may hint on a role in DNA repair, possibly in conjunction with the ~70 residue peptide Rpn15.

Assembly of 19S RP

The base and lid are assembled independently and these two complexes sequentially assemble to the 19S RP. Similar to the eukaryotic CP, the assembly of the base is mediated by four

chaperones: in yeast, Rpn14, Nas2, Nas6, and Hsm3 mediate base assembly (human homologs: Rpn14, p27, Gankyrin, and S5b). All these chaperones bind to specific ATPase subunits, resulting in the formation of three intermediate precursors, which form the base complex together with Rpn1, Rpn2, and Rpn13. These chaperones are dissociated from the proteasome during the 19S RP maturation.

Compared to the base, the mechanism of the lid assembly is less characterized. Yeast genetic studies indicate that Rpn5, Rpn6, Rpn8, Rpn9, and Rpn11 forms a core module, followed by the integration of another module consisting of Rpn3, Rpn7, and Rpn15. Association of Rpn12 completes the lid complex. The complexity of the subunit composition suggests involvement of not yet identified chaperones.

Regulatory Mechanism of the 19S RP

The mechanism of the 19S RP involves (1) recognition of polyubiquitylated substrates, (2) trimming ubiquitin chains, (3) opening the gate to the CP, (4) unfolding substrates, and (5) actively translocating them to the CP in an ATP-dependent manner. These different steps are cooperatively coupled and even the precise sequence of events is not entirely resolved.

Substrate recognition

Recognition of ubiquitylated substrates requires Ub receptors. Several ubiquitin receptors of the proteasome are identified: two of these receptors, Rpn10 and Rpn13, are canonical subunits of proteasome, whereas other known Ub receptors are categorized as PIPs: Rad23 (human homolog: hHR23), Dsk2 (Kar1), and Ddi1.

Rpn10 binds ubiquitin via its C-terminal ubiquitin interaction motif (UIM), which adopts an α -helical conformation. Rpn13 binds ubiquitin through a pleckstrin homology (PH) domain: in contrast to UIMs, ubiquitin binds to loop regions rather than secondary structural elements of the PH domain. The weakly associated shuttling Ub receptors Dsk2, Rad23, and Ddi1 contain an ubiquitin-like domain (Ubl) and an ubiquitin-associated domain (UBA) to recruit ubiquitylated substrates to the proteasome. These Ub receptors bind to ubiquitylated substrates via their UBA domain and bind to the proteasomal subunits, Rpn1, Rpn10, and Rpn13, through their Ubl domain. The interplay between these Ub receptors is not completely understood.

Deubiquitylation

Ubiquitin chains conjugated to substrates can be trimmed by DUBs before the substrate is translocated into the CP. Deubiquitylation ensures Ub recycling, but may also downregulate degradation by shortening polyubiquitin chains below the length initiating degradation. The lid subunit Rpn11 exhibits its deubiquitylating activity when coupled with ATPase subunits, suggesting that Rpn11 cleaves ubiquitin chains from the substrate immediately prior to translocation (Figure 2(e)). Two other associated DUBs, the Ub C-terminal hydrolases (UCHs), Ubp6 (human homolog: Usp14) and Uch37, act further upstream in the UPP. The UCHs may downregulate degradation by trimming of Ub chains, but they are also required for degradation of some substrates.

Gate opening

In archaea, the C-terminal residues of PAN were shown to induce opening of the CP gate. It was shown that gate opening requires three C-terminal residues: a hydrophobic residue, a tyrosine, and a residue of any type (HbYX). However, in eukaryotes not all Rpt subunits containing this C-terminal HbYX motif are able to open the CP gate: Rpt2 and Rpt5 are the only eukaryotic AAA subunits, of which the C-termini were shown to open the CP gate. Crystallographic and cryo-EM studies indicate that the HbYX motif binds to a hydrophobic pocket of the CP α -ring.

Translocation and unfolding

Translocation through the pore of the AAA-ATPase ring is coupled to the unfolding process. The AAA-ATPase N-ring was shown to exhibit chaperone function, i.e., it unfolds substrates (reverse chaperone activity). The C-terminal domains of the AAA-ATPases form the canonical AAA-ATPase ring with a very narrow channel, where a conserved aromatic-hydrophobic (Ar- ϕ) loop is located. The ATP-driven movement of the Ar- ϕ loop could pull down substrate proteins by trapping hydrophobic residues. Thus, the power stroke of the proteasomal ATPase is generated by ATP hydrolysis in a single subunit, reminiscent of the bacterial ATP-dependent protease system HslU/HslV.

26S Proteasome Regulation in the Cell

UPP in the Cell

The UPP is responsible for degradation of proteins that reside in the cytosol and the nucleus, but also membrane and secretory proteins, which are degraded by a specialized branch of the UPP, that is, ER associated degradation (ERAD), as well as some mitochondrial proteins. The UPP functions not only as a disposal machinery for misfolded proteins and for proteins damaged by stress or aging, but also as a regulator to control the lifespan of many short-lived proteins. Importantly, the 26S proteasome does not degrade all substrates completely, but it can also perform limited proteolytic processing to activate certain transcription factors, such as human NF κ B and NF κ B2. In addition to its function in ubiquitin-dependent proteolysis, the 26S proteasome was also found to degrade several non-ubiquitylated substrates, for example, ornithine decarboxylase (ODC), c-Jun, I κ B α , and p21^{Cip1}.

UPP is pivotal for cellular homeostasis because it regulates the abundance of most cellular proteins (proteomics studies indicate >75% of the entire proteome) directly or indirectly. Many PIPs beyond those discussed above are involved in proteasomal regulation in a cell, which are still far from being comprehensively elucidated; proteomics studies suggest that more than 400 proteins directly interact with the proteasome.

Proteasome and Chaperones

Molecular chaperones and the 26S proteasome seem to interact physically and cooperate in cellular protein quality control. For example, CHIP, a co-chaperone of Hsp70 and Hsp90, has E3 ubiquitin-ligase activity and mediates degradation of certain substrate proteins by the 26S proteasome. Bag-1, another Hsp70-interacting protein, was also shown to stimulate

substrate ubiquitylation and to bind directly to the 26S proteasome via its UBL domain. Thus, proteins that resist Hsp70 mediated folding may be directly tagged and degraded by recruited 26S proteasomes. The chaperone Hsp90 was shown to be required for assembling 26S proteasomes and maintaining them assembled, which may in turn regulate proteasomal activity.

Proteasome and E3 Ligases

Very little is known about the molecular mechanisms of substrate shuttling from the E3 ubiquitin ligase, where the final step of ubiquitylation takes place, to the 26S proteasome. Free diffusion of ubiquitylated substrates to the 26S proteasomes would be consistent with the rather high-binding constants observed *in vitro*. However, misfolded substrates may need to be disposed more rapidly to prevent aggregations. Indeed, E3 ligase complexes, such as Ufd4, APC, Hul5, and Ltn1 were shown to interact physically with the proteasome, indicating spatial coordination of polyubiquitylation and degradation for the respective substrates.

Degradation from Organelles

The ER does not possess its own protease(s): instead it relies on the cytosolic 26S proteasome through ERAD, which is a sub-branch of the UPP. Erroneous secretory proteins and membrane proteins are recognized and transported to the cytosol where they are ubiquitylated by membrane-associated E3 ligases. Key for subsequent degradation of almost all known substrates is the AAA-ATPase Cdc48 (human homolog: p97). In complex with its cofactors Ufd2 and Npl4, Cdc48 presumably actively escorts substrates to the 26S proteasome in an ATP-dependent manner. In ERAD, the entire mechanism of substrate recognition, translocation to the cytosol, ubiquitination and translocation to the 26S proteasome seems to be spatially and temporally highly coordinated. Since Cdc48 is also involved in the degradation of many cytosolic substrates, cytosolic degradation of many substrates might be highly coordinated as well. Moreover, proteasomal degradation of substrates from mitochondria may function analogous to ERAD, as indicated by the involvement of membrane-associated AAA-ATPases.

Summary

As the endpoint of the UPP, the 26S proteasome is involved in a plethora of intercellular pathways. In concert, the ubiquitylation machinery, Ub receptors, deubiquitylating enzymes, and translocases ensure substrate-specific regulated degradation. Structural and functional studies of the 26S proteasome revealed some details of how this giant machinery functions. However, these studies raise many further questions about 26S proteasome. How is the ubiquitylated substrate transferred into the ATPase ring? It is also puzzling how deubiquitylating enzymes remove the polyubiquitin chain with proper timing prior to protein translocation. Structure determination of the 26S proteasome will provide clues on the sequence of intermolecular events during protein degradation. Structural and biochemical studies of the 26S proteasome together with its transient interaction

partners will undoubtedly be the next challenge after obtaining a draft of the static architecture of the 26S proteasome to resolve the complex regulation of proteasomal degradation.

See also: [Cell Architecture and Function: Endoplasmic Reticulum-Associated Protein Degradation](#); [Protein/Enzyme Structure Function and Degradation: AAA-ATPases](#); [Proteasomes, Overview](#); [Protein Degradation](#); [Ubiquitin-Like Proteins](#); [Ubiquitin System](#).

Further Reading

- Besche HC, Peth A, and Goldberg AL (2009) Getting to first base in proteasome assembly. *Cell* 138: 25–28.
- Buchberger A, Bukau B, and Sommer T (2010) Protein quality control in the cytosol and the endoplasmic reticulum: Brothers in arms. *Molecular Cell* 40: 238–252.
- Finley D (2009) Recognition and processing of ubiquitin–protein conjugates by the proteasome. *Annual Reviews in Biochemistry* 78: 477–513.
- Förster F, Lasker K, Nickell S, Sali A, and Baumeister W (2010) Towards an integrated structural model of the 26S proteasome. *Molecular Cellular Proteomics* 9: 1666–1677.
- Glickman MH and Ciechanover A (2002) The ubiquitin–proteasome proteolytic pathway: Destruction for the sake of construction. *Physiological Reviews* 82: 373–428.
- Hershko A and Ciechanover A (1998) The ubiquitin system. *Annual Reviews in Biochemistry* 67: 425–479.
- Hochstrasser M (2009) Origin and function of ubiquitin-like proteins. *Nature* 458: 422–429.
- Pick E, Hofmann K, and Glickman MH (2009) PCI complexes: Beyond the proteasome, CSN, and eIF3 Troika. *Molecular Cell* 35: 260–264.
- Rock KL, York IA, Saric T, and Goldberg AL (2002) Protein degradation and the generation of MHC class I-presented peptides. *Advances in Immunology* 80: 1–70.
- Saeki Y, Toh-E A, Kudo T, Kawamura H, and Tanaka K (2009) Multiple proteasome-interacting proteins assist the assembly of the yeast 19S regulatory particle. *Cell* 137: 900–913.
- Striebel F, Kress W, and Weber-Ban E (2009) Controlled destruction: AAA+ ATPases in protein degradation from bacteria to eukaryotes. *Current Opinion in Structural Biology* 19: 209–217.
- Tai HC and Schuman EM (2008) Ubiquitin, the proteasome and protein degradation in neuronal function and dysfunction. *Nature Reviews Neurosciences* 9: 826–838.
- Tanaka K (2009) The proteasome: Overview of structure and functions. *Proceedings of the Japanese Academy, Ser. B. Physical and Biological Sciences* 85: 12–36.
- Voges D, Zwickl P, and Baumeister W (1999) The 26S proteasome: A molecular machine designed for controlled proteolysis. *Annual Reviews in Biochemistry* 68: 1015–1068.
- Wójcik C and DeMartino GN (2003) Intracellular localization of proteasomes. *International Journal of Biochemistry and Cell Biology* 35: 579–589.

Proteinase-Activated Receptors

R Ramachandran, K K Hansen, and M D Hollenberg, University of Calgary, Calgary, AB, Canada

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Kristina K. Hansen and Morley D. Hollenberg, volume 3, pp. 543–548, © 2004, Elsevier Inc.

Glossary

Agonist Compound that stimulates a cellular receptor to give a biological response.

Antagonist Compound that binds to a cellular receptor which prevents a biological response from occurring.

Antibody Protein produced by specialized B cells after stimulation by an antigen (foreign substance), which acts specifically against an antigen in an immune response.

Glycosylation sequence A three amino acid sequence where the first amino acid is asparagine, the second

is any amino acid except proline, and the third is serine or threonine.

Proteinase Enzyme which selectively catalyzes the hydrolysis of peptide bonds.

Proteinase-activated receptor (PAR) A G-protein-coupled receptor that is activated upon cleavage of the N-terminal domain by a proteinase.

Proteinase-activated receptor activating peptide (PAR-AP) Synthetic peptide that causes activation of a PAR.

Proteinase-activated receptors (PARs) are members of a novel family of seven-transmembrane G-protein-coupled receptors (GPCRs) that were first discovered in the early 1990s. To date, four members of this family (named PAR₁, PAR₂, PAR₃, and PAR₄ in order of their discovery) have been identified and their expression has been described in various organs and across several species. PARs are unique among GPCRs in having a mechanism of activation that involves enzymatic cleavage of the receptor N-terminal domain to reveal a self-activating ligand. Thus, PAR cleavage reveals a new receptor N-terminus, often described as the 'tethered ligand', which binds to another site of the receptor to initiate signaling (Figure 1(a)). Additionally, synthetic peptides (PAR-activating peptides; PAR-APs), with sequences derived from the tethered ligand, can act as soluble ligands for the PARs (Figure 1(b)). Tethered ligand or PAR-AP interaction with the receptor leads to G-protein coupling and initiation of a complex signaling cascade, resulting in various physiological or patho-physiological responses.

Structure

Proteinase-activating receptors are comprised of about 400 amino acids (Table 1) with a sequence homology of approximately 30% between the PARs. A key structural element for the PARs, the tethered ligand, is revealed upon enzymatic cleavage at an arginine or lysine residue in the receptor N-terminus (Table 1). Much evidence has supported a key role for extracellular loop 2 of PAR₁ and PAR₂ in the activation of PARs by PAR-APs. In PAR₁, a sequence of about 10 amino acids in the N-terminal extracellular domain (residues 83–93 in human PAR₁) is also involved in signaling either by the tethered ligand or by the PAR-AP. For activation by proteinases, however, the importance of extracellular loop 2 is less clearly understood. All four PARs possess a number of potential N-linked glycosylation sequences on their extracellular domains. The human genes encoding PAR₁, PAR₂, and PAR₃ are all located on chromosome 5q13, while PAR₄ is located on chromosome 19p12.

All of the PAR genes have similar overall structures, including a single intron upstream of the tethered-ligand coding sequence.

PAR Activating and Inactivating Proteinases

Thrombin, a key coagulation cascade serine proteinase, was the first proteinase shown to activate PARs (Figure 1(a)). In fact, PAR₁ was originally described as the thrombin receptor. It is now known that PAR₁, PAR₃, and PAR₄ are all targeted by thrombin. Thrombin, however, does not activate PAR₂. The most common activator of PAR₂ is trypsin, an enzyme that can also activate PAR₄.

Proteinases can also interact with PAR receptors to cleave the receptor downstream of the activation cleavage site (Figure 1(d)) resulting in a disarming of the receptor to activation by activating proteinases. For example, the neutrophil enzymes elastase, proteinase 3, and cathepsin G cleave PAR₁ on platelets and endothelial cells downstream of the thrombin cleavage site, making the cells refractory to further activation by thrombin.

Other serine proteinases that have been shown to target PARs to activate or inactivate the receptors include plasmin, granzyme A, factors VIIa/X/Xa/TF, mast cell tryptase, tissue kallikreins (hK14, hK5, and hK6), MMP-1, and dust mite proteinases (Der p3 and Der p9).

Agonist Peptides

Peptides (PAR-APs) with sequences derived from those of the tethered ligands can be used in place of proteinases to selectively activate PARs *in vitro* or *in vivo* (Figure 1(b)). Unlike thrombin and trypsin, which can act on multiple PARs, PAR-APs selective for PAR₁, PAR₂, and PAR₄ have been developed. The sequences of the most selective PAR-APs for these three receptors, as designated by the one-letter codes for each amino acid, are TFLLR-NH₂ (PAR₁), SLIGRL-NH₂ (PAR₂), and

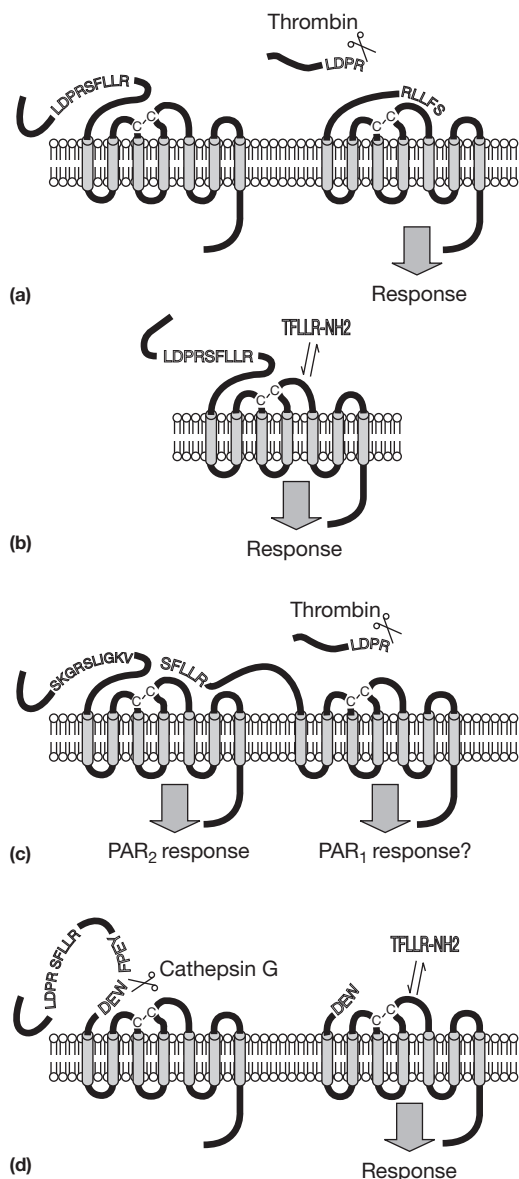


Figure 1 Mechanisms of activation and inactivation of PARs. (a) The functional PARs cloned to date contain within their receptor N terminus a serine protease cleavage/activation site that, once cleaved by a serine protease, results in the unmasking of a cryptic tethered-ligand sequence. The revealed tethered-ligand sequence is believed to bind to and subsequently activate the receptor. The example provided (a) is for human PAR₁ activation by thrombin. (b) PARs may be activated independently of proteolytic unmasking of the tethered-ligand sequence by small synthetic peptides (PAR-APs) corresponding to the first five or six amino acids of the tethered-ligand sequence. Activation by the receptor-selective PAR₁AP, TFLLR-NH₂, is shown. (c) Intermolecular activation of PAR₁ or PAR₂ may occur following PAR₁ activation, whereby the unmasked tethered-ligand of a PAR₁ receptor activates an uncleaved neighboring PAR₂ or PAR₁ receptor. (d) Inactivation of PARs may occur by proteinases cleaving the receptor C-terminal to the tethered-ligand sequence, thus amputating the tethered ligand from the receptor body. These amputated receptors can no longer be activated by proteinases but are still responsive to exogenously applied PAR-APs. Note: a putative cysteine palmitoylation site in the C-terminal tail (present in all PARs except human PAR₄) may result in a fourth intracellular loop (not shown) that could play a role in receptor signaling. Reproduced, with permission, from Hollenberg MD and

AYPGKF-NH₂ (PAR₄), respectively (Table 1). The peptide SFLLRN-NH₂, which is identical to the proteinase revealed tethered-ligand sequence of human PAR₁, activates both PAR₁ and PAR₂. Although SLIGRL is the tethered-ligand sequence of PAR₂ in both the rat and mouse but not the human, the peptide SLIGRL-NH₂ is a more potent agonist for human PAR₂ than the human tethered-ligand-derived peptide SLIGKV-NH₂. Similarly, AYPGKF-NH₂ is 10 times more potent as a human PAR₄ agonist than GYPGQV-NH₂ or GYPGKF-NH₂, the sequences represented by the tethered ligands of the human and mouse receptors, respectively. Specific agonist peptides for PAR₃ have not been described, although a peptide with the sequence corresponding to the proteolytically revealed sequence of PAR₃ can activate PAR₁ and PAR₂.

Turning the Signal Off: Desensitization, Internalization, and Resensitization

As the tethered ligand, once exposed, cannot diffuse away from the receptor, as can agonists for traditional GPCRs, the PARs require special mechanisms to silence the signal. Two mechanisms are involved: (1) a rapid desensitization of the signal, in keeping with other GPCRs that are rapidly desensitized; and (2) receptor internalization, via mechanisms involving phosphorylation of the receptor C-tail, recruitment of β -arrestins, and targeting to clathrin-coated pits for degradation. Once activated by the tethered ligand, PARs signal to cells primarily via their interactions with a variety of G proteins (G_q, G_i, and G_{12/13}) that act as transducers of the signal. The arrestins play a key role in regulating the G-protein-signaling process. In addition, via G-protein-independent mechanisms that also involve arrestin interactions, PARs (especially PAR₂) can be internalized on a signaling scaffold. By this arrestin-dependent process, PAR₂ can signal via G-protein-independent, beta-arrestin-dependent pathways. Thus, the PARs can signal via versatile mechanisms, and the arrestins play important roles in their desensitization, internalization, and G-protein-independent signaling processes.

Once cleaved and internalized, the activated/desensitized PARs, even if they return to the cell surface, can no longer be proteinase activated. Thus, it is essential that the cleaved/activated receptors be replenished at the cell surface with intact receptors to resensitize the cells for proteinase activation. For the PARs, there are intracellular reservoirs of preformed receptors, that when called upon by cell signaling, can translocate to the cell surface, thereby replenishing a cell's store of surface-accessible PARs. The mechanisms that regulate this interesting cycle in the life of the PARs are an area of considerable interest.

Receptor Dimerization and Intermolecular Activation of PARs

In addition to the ability of proteinases and agonist peptides to activate an individual PAR directly, one PAR can be activated via a neighboring PAR in the cell membrane. For example,

Table 1 Summary of human protease-activated receptors

	PAR ₁	PAR ₂	PAR ₃	PAR ₄
Amino acid composition	425	397	374	385
Cleavage site for receptors	LDPR ⁴¹ ↓SFLLR	SKGR ³⁶ ↓SLIGKV	LPIK ³⁸ ↓TFRGAP	PAPR ⁴⁷ ↓GYPGQV
Proteinase activators	Thrombin	Trypsin, tryptase	Thrombin	Thrombin
Selective peptide agonists	TFLLR-NH ₂	SLIGRL-NH ₂	None known	AYPGKF-NH ₂

Reproduced, with permission, from Hollenberg MD and Compton SJ (2002) International Union of Pharmacology. XXVIII. Proteinase-activated receptors. *Pharmacological Reviews* 54: 203–217.

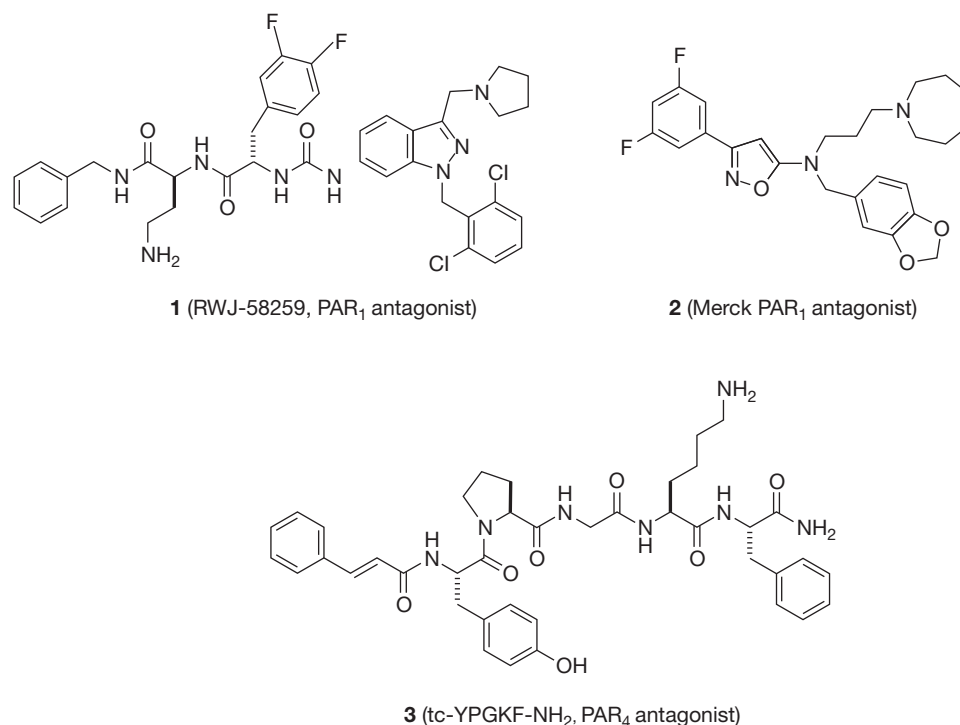


Figure 2 PAR antagonists. Reproduced, with permission, from Hollenberg MD and Compton SJ (2002) International Union of Pharmacology. XXVIII. Proteinase-activated receptors. *Pharmacological Reviews* 54: 203–217.

it has been demonstrated that the proteolytically revealed tethered ligand of PAR₁ can, in turn, activate PAR₂ (Figure 1(c)). Thus, theoretically, PAR₂ can generate a response to thrombin without being directly activated by the enzyme. Further, by acting as a docking site for thrombin, PAR₁ and PAR₃ can potentially facilitate thrombin's ability to activate PAR₄. In this manner, the PARs are acting as coreceptors, the one for the other. To date, although able to act as a coreceptor for PAR₄, PAR₃ has not been found to generate an intracellular signal on its own. Recent evidence has also pointed to the ability of PARs to dimerize. Use of resonance energy-transfer techniques in live cells have revealed that PAR₁ can dimerize with both PAR₃ and PAR₄, while indirect evidence exists for PAR₁ dimerization with PAR₂, as detailed above.

Antagonists

For understanding the physiological effects of the PARs fully, and for targeting the pathology of thrombin-mediated cardiovascular disorders, selective antagonists that inhibit activation

of PARs (by both proteinases such as thrombin and by activating peptides) are of great value. Thus, considerable work has been done to develop PAR-selective antagonists. Activation of PAR₁, by either thrombin or a PAR₁-AP, is inhibited by compounds 1 and 2 (Figure 2). More advanced PAR₁-antagonist compounds are currently in clinical development for the treatment of atherosclerosis, ischemia, myocardial infarction, and cerebrovascular accident (ClinicalTrials.gov Identifier: NCT00526474). PAR₂ inhibitors have been very challenging to develop. A PAR₁-related peptide, FSLLRY, was found to block trypsin, but not PAR₂-activating peptide triggering of PAR₂, but the potency of the compound was too low to be considered as a useful antagonist for studies done *in vivo*.

The peptide, LIGK, was found to inhibit both trypsin and peptide-dependent activation of PAR₂ and was, therefore, used as a basis to synthesize a peptidomimetic PAR₂ antagonist, ENMD-1068 (N-1-3-methylbutyryl-N-4-6-aminohexanoyl-piperazine) (Figure 1). This compound was used *in vivo* to diminish joint swelling in an animal model of arthritis, which is known to involve PAR₂ activation. However, the potency of ENMD-1068 is also too low to be considered

for further use *in vivo*. More recently, two other peptidomimetic PAR₂ antagonists, K12940{(N-[1-(2,6-dichlorophenyl)methyl]-3-(1-pyrrolidinylmethyl)-1Hindol5yl)aminocarbonyl}-glycyl-L-L-phenylalanine-N-benzylamide and K-14585 {(N-[1-(2,6-dichlorophenyl)methyl]-3-(1-pyrrolidinylmethyl)-1Hindol5-yl)aminocarbonyl}-glycyl-L-L-phenylalanine-N-benzylamide, have been shown to inhibit PAR₂-dependent responses *in vitro* and *in vivo*. But, again, the potency of these compounds is insufficient for their further use as a therapeutic agent *in vivo*. Thus, although the development of PAR₁ antagonists for potential clinical use has been successful, much more work must be done to develop clinically useful PAR₂ antagonists.

The development of PAR₄ antagonists is just beginning. Its activation in rodent platelets can be inhibited by compound 3 (tc-YPGKF-NH₂; Figure 2), but this peptide PAR₄ antagonist, that blocks rodent PAR₄ may not block PAR₄ from other species. A small-molecule PAR₄ antagonist, YD3 [1-benzyl-3-(ethoxycarbonylphenyl)-indazole], has been found to block PAR₄ activation of human platelets, but has been found to have off-target activity (e.g., blocking the rat platelet adenosine diphosphate receptor) in rodent platelets and may, therefore, not be of general utility to evaluate PAR₄ function in systems other than the human platelet. A novel approach to developing PAR₄ antagonists has targeted the intracellular receptor domains that interact with G proteins to trigger signaling. Cell-penetrating peptides (N-palmitoylated, to promote internalization), termed 'pepdudins', have been designed to disrupt PAR-G-protein interactions. Effective PAR₄ inhibition has been demonstrated with the use of the pepducin P4pal-i1, an N-palmitoyl cell-penetrating peptide modeled on the third intracellular loop of PAR₄. Pepducins-targeting PAR₁ have also been described, but no work has yet been done to see if targeting PAR₃ might block PAR activity.

PAR-targeted antibodies may also prove of use as systemic receptor antagonists. As an alternative to small-molecule PAR antagonists, antibodies that restrict the access of proteinases to the activation/cleavage site have proved of value. Thus, a PAR₂-targeted antibody that blocks trypsin activation of this receptor has proved of value to attenuate joint swelling in an *in vivo* model of arthritis.

Distribution

PARs have been detected in a variety of tissues by northern blot analysis. The brain, lung, heart, stomach, colon, kidney, small intestine, liver, pancreas, testes, prostate, uterus, placenta, trachea, thymus, lymph node, skeletal muscle, and adrenal gland have all been shown to have at least one of the PARs. Table 2 summarizes the tissues and cells in which each PAR is found. It is important to note, however, that cellular expression is not necessarily conserved across species. For example, human platelets express PAR₁ and PAR₄, while mouse platelets express PAR₃ and PAR₄.

Detection Methods

Expression of PARs in tissue and cells is characterized with standard methods of detection including reverse-transcriptase/polymerase-chain reaction (RT-PCR), northern blot analysis,

and immunohistochemical analysis. A number of specific antibodies targeting the PARs have been developed and are available commercially. For PAR₁, antibodies targeting different epitopes within the receptor include ATAP-2, WEDE-15, and SPAN-12, as well as a number of other commercially available antibodies. Similarly, PAR₂ antibodies targeting different sites within the receptor have been described and include B5, SLAW-A, and Sam-11, as well as others. A number of polyclonal antibodies targeting PAR₃ and PAR₄ are also commercially available.

Successful ligand-binding assays have been described for PAR₁ using the PAR-AP-derived radioligand [3H]haTRAP (alanine-parafluorophenylalanine-arginine-cyclohexylalanine-homoarginine-[3H] phenylalanineamide). For PAR₂, two different tritiated ligand probes, 2-furoyl-LIGRL[N(3H)propionyl]-O-NH₂ and 2-furoyl-[H³]LIGRL-NH₂, have been developed. The ornithine-containing PAR₂ ligand-binding probe has also been derivatized with a fluorescent tag (2-furoyl-LIGRL[N-(Alexa Fluor 594)-O]-NH₂) for use in binding studies and for visualizing the receptor on cells and tissue.

Bioassays

Measurement of Calcium Signaling

All of PARs 1, 2, and 4 can couple to Gq and mobilize calcium from intracellular stores through activation of phospholipase C, which then converts phosphatidyl inositol 4,5-bisphosphate to inositol 1,4,5-trisphosphate (IP₃) and diacylglycerol. The IP₃ causes calcium to be released from the endoplasmic reticulum, thus increasing the intracellular calcium levels. Measurement of this increase in intracellular calcium concentration provides a good measure of receptor expression and function. Calcium mobilization can be measured in cells preincubated with a fluorescent cell-permeable calcium-sensitive dye (Fluo-4, Fluo-3, Fura-2, etc.). Activation of PARs results in an elevation in cytoplasmic calcium levels, which can be measured using a fluorescent spectrophotometer, plate reader, or microscope. To confirm whether or not an unknown agonist or proteinase activates a certain PAR, cross-desensitization experiments can be performed. For example, PAR-expressing cells can be treated first with a PAR-specific agonist (maximum dose) to desensitize the specific PAR receptor population in those cells. This pretreatment is then followed by addition of the unknown agonist that of itself can cause an increase in intracellular calcium. If a calcium response is no longer observed for the unknown agonist, but still present for a PAR-unrelated compound (e.g., lysophosphatidic acid), it can be concluded that the unknown agonist indeed does act via that particular PAR receptor. By contrast, if the PAR-desensitization protocol does not prevent the unknown agonist from triggering a calcium response, it can be concluded that it stimulates the cell through another PAR-unrelated mechanism.

Tissue Bioassays (Contraction or Relaxation)

To better understand how the activation of PARs leads to a biological response, blood vessel and intestinal smooth muscle strips may be treated with PAR-APs to monitor the contractile or relaxant activities. The relaxation of rat aorta via the endothelium-dependent nitric oxide-mediated pathway

Table 2 Localization and potential roles of the human proteinase-activated receptors

	PAR ₁	PAR ₂	PAR ₃	PAR ₄
Tissue distribution ^a (northern blot analysis)	Brain, lung, heart, stomach, colon, kidney, testis	Prostate, small intestine, colon, liver, kidney, pancreas, trachea	Heart, kidney, pancreas, thymus, small intestine, stomach, lymph node, trachea	Lung, pancreas, thyroid, testis, small intestine, placenta, skeletal muscle, lymph node, adrenal gland, prostate, uterus, colon
Cellular expression	Platelets, endothelium, vascular smooth muscle, leukocytes, GI tract epithelium, fibroblasts, neurons, mast cells	Endothelium, leukocytes, GI tract epithelium and lung, airway and vascular smooth muscle, neurons, mast cells, keratinocytes, lung fibroblasts, renal tubular cells	Airway smooth muscle, platelets	Platelets, megakaryocytes
Known physiological role	Platelet activation			Platelet activation
Potential physiological roles	Pro-inflammatory, embryonic development, regulation of vascular tone	Pro-inflammatory, mediator of nociception, airway protection, regulation of vascular tone	Cofactor for PAR ₄	Platelet activation

GI, gastrointestinal.

^aInformation for PAR₂, PAR₃, and PAR₄ is provided for human tissues; for PAR₁, distribution is recorded in rat tissues, for which more extensive data are available than in human tissues. Reproduced, with permission, from Hollenberg MD and Compton SJ (2002) International Union of Pharmacology. XXVIII. Proteinase-activated receptors. *Pharmacological Reviews* 54: 203–217.

can be measured. Similarly, contraction of the rat gastric longitudinal muscle or the endothelium-free aorta tissue can be assessed. For the bioassay experiment, a tissue is suspended between force-transducer calipers under a resting tension in an organ bath. Changes in tension are detected by measuring the displacement transducer signal coupled to an amplifier and a data acquisition system. Known compounds that cause a relaxant or contractile response, other than the PAR-activating agonists, are first added to the organ bath to verify the responsiveness of the tissue. For example, acetylcholine is known to relax precontracted aortic rings via an endothelium-dependent process, and potassium chloride is known to cause a large contractile response in smooth muscle from the stomach or blood vessels by acting directly on the tissue to cause depolarization, and thus contraction. The response to the PAR-APs added to the organ bath is then monitored and compared to the responses generated by KCl (contraction) or acetylcholine (relaxation). As in the calcium-signaling assay, cross-desensitization experiments can be performed to assess the potential contribution of each PAR to the biological response.

Platelet Aggregation

Platelets are known to express PARs 1, 3, and 4, but not PAR₂. Activation of PARs on platelets causes aggregation. Using a light-transmission aggregometer, platelet aggregation can be monitored by measuring the turbidity of a suspension of platelets. Simply by adding agonists to a suspension of platelets, aggregation can be quantified either as a percentage increase in light transmission relative to that of the initial platelet suspension, or as a percentage of the maximal aggregation, versus time. This assay is commonly used to determine whether or not a compound is an agonist or an antagonist of PAR₁ and PAR₄. For example, if an antagonist is added before a specific PAR-AP, the aggregation of the platelets will be diminished.

In addition to monitoring the aggregation response, calcium signaling can also be measured in platelets in response to PAR activation as described in the section Measurement of Calcium Signaling.

Physiological Roles of PARs

PAR₁ plays an aggregatory and secretory role in human platelets and a role as a regulator of endothelial cell function in blood vessels. PAR₁ has also been shown to have an important role in the process of narrowing the arteries after vascular injury. PAR₂ has been suggested to be involved in the cardiovascular, pulmonary, and gastrointestinal systems, as well as having an important role in the settings of inflammation and pain sensation. As the only role of PAR₃ shown to date is that it functions as a cofactor for thrombin-mediated activation of PAR₄, its physiological role is the same as for PAR₄. PAR₄, like PAR₁, has been shown to be responsible for the activation of platelets in humans and other species, and is also thought to be involved in interactions between the endothelium and leukocytes. Further, PAR₄ is implicated as a receptor involved in nociception and analgesia. There are many other physiological roles of PARs that have yet to be discovered.

See also: [Signaling: G Protein Signaling Regulators; G-Protein-Coupled Receptor Kinases and Arrestins; Phospholipase C.](#)

Further Reading

Adams MN, Ramachandran R, Yau MK, et al. (2011) Structure, function and pathophysiology of protease activated receptors. *Pharmacology and Therapeutics* 130: 248–282.

- Coughlin SR (2000) Thrombin signalling and protease-activated receptors. *Nature* 407: 258–264.
- Defea K (2008) Beta-arrestins and heterotrimeric G-proteins: Collaborators and competitors in signal transduction. *British Journal of Pharmacology* 153: S298–S309.
- Hollenberg MD and Compton SJ (2002) International Union of Pharmacology. XXVIII. Proteinase-activated receptors. *Pharmacological Reviews* 54: 203–217.
- Ishihara H, Connolly AJ, Zeng D, et al. (1997) Protease-activated receptor 3 is a second thrombin receptor in humans. *Nature* 386: 502–506.
- Kahn ML, Zheng YW, Huang W, et al. (1998) A dual thrombin receptor system for platelet activation. *Nature* 394: 690–694.
- Leger AJ, Covic L, and Kuliopulos A (2006) Protease-activated receptors in cardiovascular diseases. *Circulation* 114: 1070–1077.
- MacFarlane SR, Seatter MJ, Kanke T, Hunter GD, and Plevin R (2001) Proteinase-activated receptors. *Pharmacological Reviews* 53: 245–282.
- Nystedt S, Emilsson K, Wahlestedt C, and Sundelin J (1994) Molecular cloning of a potential proteinase activated receptor. *Proceedings of the National Academy of Sciences of the United States of America* 91: 9208–9212.
- O'Brien PJ, Molino M, Kahn M, and Brass LF (2001) Protease activated receptors: Theme and variations. *Oncogene* 20: 1570–1581.
- Ossovskaya VS and Bunnett NW (2004) Protease-activated receptors: Contribution to physiology and disease. *Physiological Reviews* 84: 579–621.
- Ramachandran R and Hollenberg MD (2008) Proteinases and signalling: Pathophysiological and therapeutic implications via PARs and more. *British Journal of Pharmacology* 153: S263–S282.
- Ramachandran R, Noorbakhsh F, Defea K, and Hollenberg MD (2012) Targeting proteinase-activated receptors: Therapeutic potential and challenges. *Nature Reviews Drug Discovery* 11: 69–86.
- Rasmussen UB, Vouret-Craviari V, Jallat S, et al. (1991) cDNA cloning and expression of a hamster alpha-thrombin receptor coupled to Ca^{2+} mobilization. *FEBS Letters* 288: 123–128.
- Steinhoff M, Buddenkotte J, Shpacovitch V, et al. (2005) Proteinase-activated receptors: Transducers of proteinase-mediated signaling in inflammation and immune response. *Endocrine Reviews* 26: 1–43.
- Vergnolle N (2004) Modulation of visceral pain and inflammation by protease-activated receptors. *British Journal of Pharmacology* 141: 1264–1274.
- Vu TK, Hung DT, Wheaton VI, and Coughlin SR (1991) Molecular cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell* 64: 1057–1068.
- Xu WF, Andersen H, Whitmore TE, et al. (1998) Cloning and characterization of human protease-activated receptor 4. *Proceedings of National Academy of Sciences of the United States of America* 95: 6642–6646.

Protein Carboxyl Esterification

E Perez, Signum Biosciences, Monmouth Junction, NJ, USA

J B Stock, Princeton University, Princeton, NJ, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Shilpa G. Lalchandani and Jeffrey B. Stock, volume 3, pp. 474–477, © 2004, Elsevier Inc.

Glossary

Methylesterase Enzyme responsible for demethylating the target protein by hydrolysis of the methyl ester.

Methyltransferase Enzyme that transfers a methyl group from the universal methyl donor *S*-adenosylmethionine, SAM, to a carboxyl group in a target protein.

Introduction

The methylation and demethylation reactions are catalyzed by two types of enzymes: methyltransferase (MTase) and methylesterase (MEase). The MTase transfers methyl groups from the universal methyl donor *S*-adenosylmethionine (SAM), to a carboxyl group in a target protein. This converts a negatively charged carboxylate anion on the protein surface into a neutral methyl ester with concomitant formation of *S*-adenosylhomocysteine (SAH). The methyl group is subsequently removed by an MEase that catalyzes the hydrolysis of the methyl ester to restore the carboxylate anion on the target protein, with the concomitant production of methanol:

MTase : SAM + protein → SAH + methylated protein

MEase : Methylated protein + H₂ → methanol + protein

The central role of carboxyl methylation in sensory receptor function was first evidenced by the discovery that membrane receptor proteins are methylated and demethylated during bacterial chemotaxis. The involvement of methylation in sensory regulation was extended to vertebrate systems by the finding that RAS-related and heterotrimeric G proteins were subject to carboxyl methylation and demethylation. Finally, an exhaustive search for additional systems of carboxyl methylation revealed a second major regulatory locus in eukaryotic cells that involves methylation and demethylation of phosphoprotein phosphatase 2A (PP2A), the major phosphoprotein phosphatase in the brain.

To understand how carboxyl methylation works in order to regulate cellular responses to extracellular signals, it is informative to consider the role of another type of protein modification chemistry, the phosphorylation, and dephosphorylation of protein side chains. As with carboxyl methylation, protein phosphorylation is controlled by two types of converter enzymes. A kinase catalyzes the transfer of a phosphoryl group from the universal phosphodonor, adenosine triphosphate (ATP), to an acceptor residue at the surface of a target protein with concomitant formation of adenosine diphosphate (ADP); and a phosphoprotein phosphatase catalyzes the hydrolysis of the protein phosphoester with concomitant production of inorganic phosphate. Phosphorylation generally acts as a switch to control the activity of a target protein. The first example of this regulatory mechanism was obtained through studies of the regulation of glucose metabolism in muscle and liver. There, it

is shown that the release of glucose from glycogen in response to epinephrine involves the activation of a specific kinase that turns on glycogen phosphorylase by transferring a phosphoryl group from ATP to a serine hydroxyl side chain in the enzyme. This modification is later reversed by a phosphoprotein phosphatase to turn off glycogen phosphorylase under resting conditions, when glucose needs to be stored. Subsequent research has shown that regulation by phosphorylation and dephosphorylation is a ubiquitous mechanism. A typical vertebrate genome encodes upward of a thousand different protein kinases whose activities are controlled by different regulatory signals, and it has been estimated that a quarter of the proteins in a typical vertebrate cell are subject to regulation by phosphorylation and dephosphorylation.

One might suppose that carboxyl methylation and demethylation would provide the same type of switching mechanism as phosphorylation and dephosphorylation; however, this is not the case. Whereas phosphorylation is quite common, carboxyl methylation is rare and present at only approximately 2% of the level of phosphorylation. There appear to be only three regulatory protein MTase/MEase enzyme systems: one in prokaryotes that controls chemotaxis receptors and two in eukaryotes that regulate G protein and PP2A. Moreover, in these instances, carboxyl methylation does not function as an on/off switch, but serves instead to modulate the output of a regulatory network of kinase/phosphatase switches to optimize the response to a given stimulatory input.

Carboxyl Methylation in Bacterial Chemotaxis

Virtually all motile prokaryotes use the same basic mechanism to control their motor activities in response to attractant and repellent stimuli. Generally, each cell has a signal transduction apparatus composed of a cluster of thousands of α -helical hair-like fibers that traverse the cytoplasmic membrane (**Figure 1**). The portion of each fiber located at the outside surface of the membrane acts as a sensory receptor, interacting with specific attractant and repellent stimuli in the extracellular medium. The portion of the fibrous array inside the cell controls the activity of an associated protein kinase that catalyzes the transfer of a phosphoryl group from ATP to a small protein in the cytoplasm that subsequently interacts with the motor apparatus to produce a chemotaxis response. Although the

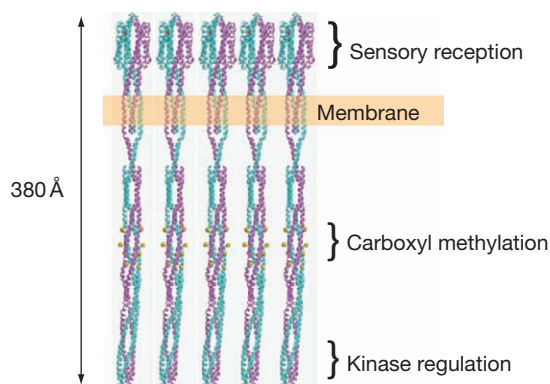


Figure 1 Model of the chemotaxis receptor fibers. Transmembrane methylatable receptors are employed as sensors in bacterial chemotaxis.

cytoplasmic domains of receptor fibers and their associated kinase/phosphatase signaling enzymes are highly conserved, the extracellular sensing domains are variable, depending on the particular stimuli with which they interact. The number of different sensory receptor fibers encoded in a given genome ranges from just a few, for example, five in the case of *Escherichia coli*, to several dozen, for example, 46 in *Vibrio cholerae*.

Stimulus–response coupling through the chemotaxis phosphorelay signaling apparatus is modulated by carboxyl methylation. Sequences within the cytoplasmic portions of each fiber between the membrane spanning domain and the kinase regulatory domain contain a series of at least four glutamate residues that are subject to methylation and demethylation by specific MTase and MEase enzymes. The degree of methylation of the glutamyl carboxylate side chains controls the link between stimulus binding to the extracellular sensory portions of the fibrous array and the intracellular kinase. In most bacteria, lower levels of methylation act to increase the cell's sensitivity to attractant stimulation. The level of methylation is controlled by several parameters including the availability of SAM and feedback from the kinase, as well as by conformational changes in the fibers induced by stimulus interactions with the extracellular sensory receptor domains. Since changes in methylation occur at a much slower rate than changes in kinase activity, methylation provides an adaptive mechanism, whereby the sensory apparatus adjusts the phosphorylation response to fit a given background stimulus intensity. The result is analogous to light–dark adaptation in vertebrate visual systems, where sensitivity is shifted in response to background light intensity.

Carboxyl Methylation of Eukaryotic Signal Transduction Proteins

Responses of eukaryotic cells to extracellular signals are generally mediated by sensory receptor proteins that traverse the cytoplasmic membrane. Portions of these proteins at the cell surface interact with hormones and other stimuli. These interactions cause structural changes in portions of the proteins that extend into the cytoplasm, and these changes act in turn to cause associated G proteins to exchange bound guanosine diphosphate (GDP) for guanosine triphosphate (GTP). The G

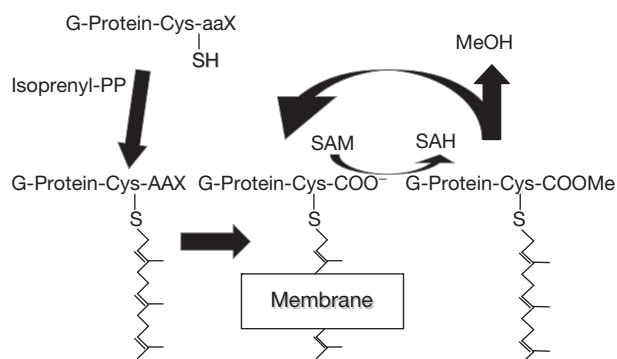


Figure 2 Sequence of processing events involved in isoprenylcysteine carboxyl methylation. The methylated molecules are soluble GTPases in the cytosol, which are targeted to the membranes by a series of posttranslational C-terminal modifications that include prenylation, C-terminal proteolytic cleavage, and carboxyl methylation.

proteins in their GTP-bound state are activated to produce second messengers such as cyclic adenosine monophosphate (cAMP) that activate protein kinases, which regulate key enzymatic activities by phosphorylating serine and threonine side chains. The phosphoryl groups are subsequently removed through the action of phosphoprotein phosphatases, such as PP2A. These types of signal transduction pathways are modulated by the carboxyl methylation and demethylation of alpha carboxyl groups at the C termini of the G proteins and PP2A. Thus, although the chemistry is somewhat different, carboxyl methylation modulates receptor-mediated responses in eukaryotes just as in bacterial cells.

Carboxyl Methylation of Ras-Related and Heterotrimeric G Proteins

A majority of carboxyl-methylated proteins in eukaryotic cells are Ras-related and are heterotrimeric G proteins. These signal transduction proteins are targeted to membranes, where they relay information from sensory receptors. Membrane targeting depends on a series of posttranslational C-terminal modifications that include prenylation, proteolytic cleavage, and carboxyl methylation (Figure 2). The initial event, prenylation, depends on a characteristic C-terminal tetrapeptide, termed a 'CaaX motif', consisting of a cysteine, C, followed by two aliphatic amino acids, aa, and a final amino acid that is variable, X. The prenyl group is either a C-15, farnesyl or a C-20, geranylgeranyl, polyisoprenoid, depending primarily on the nature of the C-terminal X residue. In either case, prenylation causes the G protein to become associated with endoplasmic reticulum (ER) membranes, where a membrane-associated protease catalyzes the cleavage of the aaX sequence, leaving a prenylcysteine alpha carboxyl group at the C terminus. This is the site of methylation by a specific MTase and of demethylation by one or more MEase activities.

G-protein methylation is catalyzed by an integral membrane protein that is localized in the ER. This MTase, termed 'isoprenylcysteine methyltransferase (ICMT)', has a broad specificity for prenylcysteine alpha carboxyl groups. First identified in *Saccharomyces cerevisiae* (STE14p), ICMTs have

since been shown to be conserved from yeast to humans. ICMTs are not protein specific – they even catalyze the methylation of small-molecule amino-acid analogs, such as *N*-acetyl-*S*-farnesylcysteine or *N*-acetyl-*S*-geranylgeranyl cysteine. Functional characterizations of ICMT interactions with methyl-accepting substrates have been performed primarily with these analogs, which have recently been shown to possess anti-inflammatory properties by altering G-protein signaling. The rate of methylation is controlled at several levels: G-protein activation leads to increases in methylation, and elevated levels of SAH lead to product inhibition. Methylation is not required for receptor-mediated G-protein signaling *per se*, but rather acts to modulate stimulus–response coupling. The physiological importance of isoprenylcysteine carboxyl methylation has been demonstrated with ICMT^{−/−} knockout mice, which die at embryonic day 11.5, designating *Icmt* as an essential gene for mouse embryonic development. Furthermore, use of *Icmt*-deficient and conditional *Icmt* mutant fibroblasts inhibited both Ras- and Raf-induced oncogenic transformation and led to mislocalization of Ras proteins.

Carboxyl Methylation of Phosphoprotein Phosphatase 2A

PP2A is generally isolated as a heterotrimeric complex consisting of three subunits: A, B, and C. The 65-kDa, A subunit serves as a scaffold for the association of the 36-kDa catalytic C subunit and regulatory B subunits. Although AC dimers are abundant in tissue, ABC trimers of the PP2A enzyme are prevalent *in vivo*. Although only two isoforms of the A and C subunits have been reported so far in humans, there are several different unrelated families of B subunits. These are, in large part, responsible for the specificity of PP2A toward the thousands of different potential phosphoprotein substrates in

vertebrate tissues. It has been estimated that PP2A is responsible for roughly half of the total phosphoprotein phosphatase activity. PP2A is one of the most abundant enzymes in brain, where, among other functions, it plays a critical role in the regulation of axonal microtubule assembly.

The PP2A catalytic subunit has a highly conserved C-terminal extension, the last six residues, TPDYFL, being identical in all PP2As. The alpha carboxyl group of the C-terminal leucine residue is the site of methylation and demethylation (Figure 3). Methylation is essential for the assembly of PP2A holoenzymes. The discovery of the PP2A methylation system led to the purification and characterization of specific enzymes responsible for PP2A methylation and demethylation.

The PP2A MTase was initially purified from bovine brain as a soluble 38-kDa monomer. The enzyme appears to be completely specific for the methylation of PP2A AC dimers. PP2A methyl esters are hydrolyzed by the action of a specific MEase with similar specificity. The PP2A MEase has been purified from bovine brain as a soluble 46-kDa monomer. The MTase and MEase are highly conserved in species ranging from yeast to human, and the structures regulating this reversible methylation reaction have been solved (Figure 3).

Methylation appears to function primarily to regulate the assembly of PP2A heterotrimeric holoenzymes (Figure 3). Differential methylation of PP2A has been observed during cell cycle progression, suggesting that methylation might be altering the subcellular localization of PP2A. Selective microtubule-associated PP2A is differentially regulated during the cell cycle, implying that PP2A is regulated by methylation. PP2A methylation seems to play an important role in growth control. Mutant yeasts that completely lack PP2A MTase activity are resistant to rapamycin, and are sensitive to antibiotics, such as nacotazol, that inhibit microtubule function. It has been proposed that, in humans, product inhibition of PP2A by SAH in brain may help explain

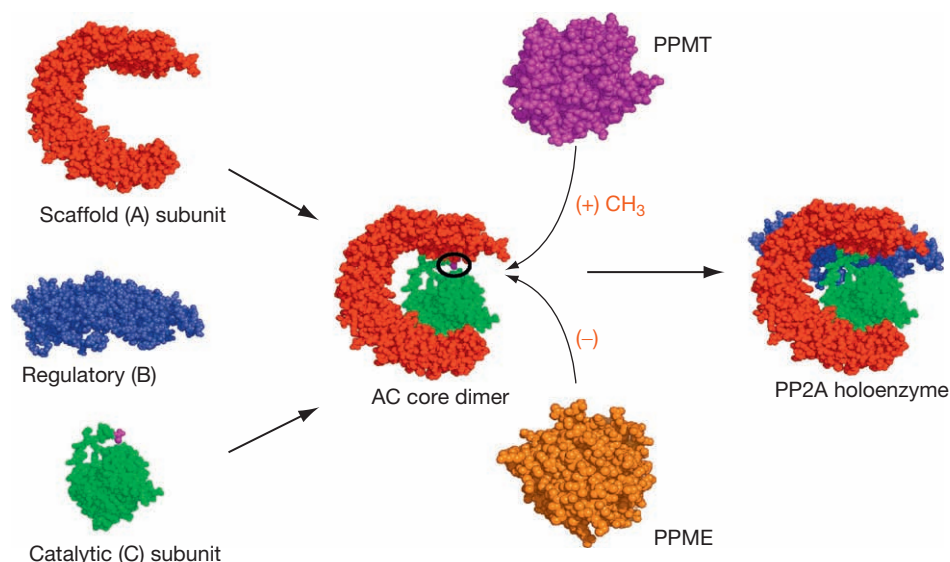


Figure 3 Carboxyl methylation regulates PP2A by controlling the association of regulatory B subunits. The PPMT enzyme binds to unmethylated AC and catalyzes the transfer of a methyl group from SAM to the C subunit. The methylated AC dimer is then able to bind regulatory B subunits. Methylated AC dimers are subject to demethylation by a specific methylesterase, PME. The Pymol software was used to generate the figure with the following PDB ID (PP2A: 3DW8, PPMT: 1RJD, PPME: 3C5V).

why elevated homocysteine levels are a risk factor for Alzheimer's disease since homocysteine can condense with adenosine to generate SAH. Furthermore, low levels of AC-dimer methylation in Alzheimer's disease correlate well with deficiencies in PP2A activity, leading to hyperphosphorylated tau, and the formation of a neurofibrillary tangle.

See also: **Cell Architecture and Function:** Eukaryotic Chemotaxis; **Signaling:** G Protein Signaling Regulators; G₁₂/G₁₃ Family; G_q Family; Gs Family of Heterotrimeric G Proteins; Ras Family.

Further Reading

- Anderson JL, Henriksen BS, Gibbs RA, and Hrycyna CA (2005) The isoprenoid substrate specificity of isoprenylcysteine carboxylmethyltransferase: Development of novel inhibitors. *Journal of Biological Chemistry* 280(33): 29454–29461.
- Bergo MO, Leung GK, Ambroziak P, et al. (2001) Isoprenylcysteine carboxyl methyltransferase deficiency in mice. *Journal of Biological Chemistry* 276(8): 5841–5845.
- Djordjevic S, Stock AM, Chen Y, and Stock JB (1999) Protein methyltransferases in signal transduction. In: Cheng X and Blumenthal RM (eds.)
- S-Adenosylmethionine-Dependent Methyltransferases: Structure and Functions*, pp. 149–183. Singapore: World Scientific.
- Gordon JS, Wolanin PM, Gonzalez AV, et al. (2008) Topical *N*-acetyl-*S*-farnesyl-L-cysteine inhibits mouse skin inflammation, and unlike dexamethasone, its effects are restricted to the application site. *Journal of Investigative Dermatology* 128(3): 643–654.
- Leulliot N, Quevillon-Cheruel S, Sorel I, et al. (2004) Structure of protein phosphatase methyltransferase 1 (PPM1), a leucine carboxyl methyltransferase involved in the regulation of protein phosphatase 2A activity. *Journal of Biological Chemistry* 279(9): 8351–8358.
- Li Z and Stock JB (2009) Protein carboxyl methylation and the biochemistry of memory. *Biological Chemistry* 390(11): 1087–1096.
- Perez E, West AH, Stock AM, et al. (2004) Discrimination between different methylation states of chemotaxis receptor Tar by receptor methyltransferase CheR. *Biochemistry* 43(4): 953–961.
- Sontag E, Hladik C, Montgomery L, et al. (2004) Downregulation of protein phosphatase 2A carboxyl methylation and methyltransferase may contribute to Alzheimer disease pathogenesis. *Journal of Neuropathology and Experimental Neurology* 63(10): 1080–1091.
- Winter-Vann AM and Casey PJ (2005) Post-prenylation-processing enzymes as new targets in oncogenesis. *Nature Reviews Cancer* 5(5): 405–412.
- Wright LP and Philips MR (2006) Thematic review series: Lipid posttranslational modifications. CAAX modification and membrane targeting of Ras. *Journal of Lipid Research* 47(5): 883–891.
- Xing Y, Chen Y, Stock JB, Jeffrey PD, and Shi Y (2008) Structural mechanism of demethylation and inactivation of protein phosphatase 2A. *Cell* 133(1): 154–163.

Protein Data Resources

N Furnham and J M Thornton, European Bioinformatics Institute, Cambridge, UK
J S Garavelli, University of Delaware, Newark, DE, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Philip E. Bourne, volume 3, pp. 478–483, © 2004, Elsevier Inc.

Glossary

Curation The process of evaluating, validating, and annotating experimental or derived data.

Hidden Markov model A statistical model in which a system is assumed to be a Markov process (with fixed conditional transition probabilities between system states) but with unobserved state.

Ontology In information science, a formal representation of the knowledge within a domain by

an agreed hierarchical set of concepts forming a controlled vocabulary, with the definitions relating those concepts forming nonrecursive, nested subsets.

Proteomics The study or analysis of the set of proteins expressed by an organism, their relationship to the genes coding for them, and to the physiological and pathological processes affecting their structure and function.

Primary Sequence Data

Humorists have observed that as we mature professionally, we learn more and more within an increasingly specialized area of knowledge until ultimately we know almost everything there is to know about nothing at all. The knowledge being acquired about proteins is suffering a similar fate. The rising flood of protein sequence translations coming from genome sequencing results necessarily means that a smaller and smaller fraction of those sequences can actually be studied experimentally as proteins. The computer is now one of the most powerful tools we have for studying proteins. From the computer-automated sequencing of polynucleotides, through the assembly of polynucleotide reads into genomes, through the interpretation of the genes into transcripts and translations, to the comparison of protein sequences and their contexts in the effort to discern their meaning, every step is done with computers. The biological databases play a central role in this computer-driven flow of biology. The biological databases, made possible by computer technology, provide access to the primary sources of biological information, the primary sequences. Although the first bioinformatics resource was published as a book, *The Atlas of Protein Sequence and Structure* in 1974, and most of the various bioinformatics databases produced CDROMs at one time, the access and distribution provided by bioinformatics resources today is entirely through the Internet, either by web browsers, by file transfer (ftp) of selected database segments or entries, or by server-produced query responses in computer-readable format.

Genomic Sequence and Context Resources

Genome sequences are today the starting point for comprehending biological knowledge. Now that a large number of whole genomes are available for comparison studies, it is not just the polynucleotide sequences, but also their contexts and interrelationships that database providers must attempt to

capture and present. This is currently the largest field in bioinformatics. The international collaboration taking the central role in collecting and presenting polynucleotide sequences from single genes to whole genomes consists of the National Center for Biotechnology Information (NCBI) producing GenBank, the European Bioinformatics Institute (EBI) of the European Molecular Biology Laboratory producing the Nucleotide Archive (ENA, formerly EMBL-Bank), and the Center for Information Biology and DNA Data Bank of Japan producing the DDBJ. All of the polynucleotide sequences deposited with each of these collaborators are shared and redistributed by all of them, but each in their own legacy format or set of formats.

The contextual relationship of genes is important in understanding how they are expressed and function, but most especially in how they evolve. In predicting the function of a particular protein, it is important to know whether the genes being compared are considered to be orthologous (derived from genes with ostensibly the same function in different species) or paralogous (derived from genes that may have acquired different functions after gene duplication). The availability of whole genomes provides a more direct means of assessing whether homologous sequences are orthologous with the same function, or paralogous with related but different functions. Assembled genomic information is available from several resources. The EBI produces Ensembl, which is focused on vertebrate genomes, and Ensembl Genomes, focused on nonvertebrate species and presently organized in five divisions, metazoa, protists, bacteria, plants, and fungi. The NCBI produces the Genome Database, and the Database of Genomic Structural Variation. Another useful resource in this area is eggNOG, a database of clustered orthologous groups with functional annotation.

Protein Primary Sequence Resources

UniProt, the universal protein resource, is a comprehensive resource for protein sequence and annotation data. It is produced by the UniProt Consortium consisting of the European

Bioinformatics Institute, the Swiss Institute of Bioinformatics (SIB), and the Protein Information Resource (PIR). The PIR produced the original *Atlas of Protein Sequence and Structure* and later the first bioinformatic database, the Protein Sequence Database. In 1986, SIB began producing Swiss-Prot, later being joined by the EBI in co-producing Swiss-Prot and TrEMBL. In 2002, the three institutes formed the UniProt Consortium to pool their resources and expertise. The primary database produced by the Consortium is the UniProt Knowledgebase (UniProtKB) produced in two divisions, the curated Swiss-Prot and the automatically annotated TrEMBL. Other databases are UniProt Reference Clusters (UniRef), the UniProt Archive, and the UniProt Metagenomic and Environmental Sequences database, a repository of unannotated metagenomic and environmental protein sequence data. The resource provides a unified web portal maintained jointly by each Consortium member. The most current listing of database formats and web access tools is published in the annual Database issue of *Nucleic Acids Research*.

The NCBI produces RefSeq, a database of nonredundant translated protein sequences. Basic local alignment search tool (BLAST) protein sequence searches are performed on the RefSeq nonredundant database of all coding sequence translations from GenBank, along with UniProt and the Protein Data Bank (PDB). The Protein Clusters database is a representative set of nearly identical RefSeq sequences that are encoded by complete prokaryotic, mitochondrial, and chloroplast genomes organized taxonomically. The Protein Clusters database is used as a basis for genome comparison at the NCBI and can be used in simplified BLAST searches. The most current descriptions of the protein sequence database resources produced by the NCBI are also published in the annual Database issue of *Nucleic Acids Research*.

Protein Modification and Secondary Structure Resources

Although most protein sequence information in the primary structure resources is derived from the translation of polynucleotide sequences, those translations tell us nothing about how those protein structures are altered by posttranslational processing, and almost all proteins are processed in some way, usually beginning with the removal of the initiator methionine. Typical protein processing includes the cleavage of signal and activation peptides, the covalent modification of amino acid side chains, of N- and C-terminals, and of the polypeptide backbone. The tight and sometimes highly specific binding of metals and metal complexes also changes the covalent structure.

There are a number of web resources for predicting certain particular types of modifications, such as phosphorylations or glycosylations, or for signal or targeting sequences. Especially notable are resources such as those developed by the Research Institute of Molecular Pathology Bioinformatics Group in Vienna, Austria, and by the Center for Biological Sequence Analysis at the Technical University of Denmark. The RESID Database of Protein Modifications is a resource providing a comprehensive collection of the chemical structure of natural protein modifications that have been reported, and includes the controlled vocabulary used to describe those modifications in the feature annotations of the UniProt Knowledgebase.

Proteomics and Protein Interaction Resources

After genome sequencing data, the largest growth in bioinformatics has been in two areas that are usually grouped together as representative of proteomics, protein interaction data and proteome mass-spectrometry data. The International Molecular Exchange (IMEx) consortium of protein interaction databases includes IntAct, DIP, MINT, MPact, MatrixDB, MPIDB, and BioGRID. Protein interaction data in the entire core IMEx databases are accessible using the Proteomics Standard Initiative (PSI) Common Query Interface. This interaction data include the interactions of proteins with other proteins and other molecules, important experimental approaches in studying protein function. Similarly, the ProteomeXchange consortium consists of the proteomics repositories, the EBI PRIDE (Proteomics Identifications Database), the NCBI Peptidome, the Institute for Systems Biology PeptideAtlas, the Proteome Commons Tranche repository and the Global Proteome Machine Organization gpmDB. The ProteomeXchange consortium aims to exchange mass spectrometry-based proteomics data and to ensure data availability in all member databases. Data derived from large-scale proteomics experiments are stored either in databases with curation guaranteeing annotation to some minimum acceptable level, or in data repositories with no curation and depositors usually required to meet standards for minimum acceptable annotation, such as the minimal information about a proteomics experiment (MIAPE).

In order to perform the mass-spectrometric identification of peptides for proteome analysis, it is necessary to construct libraries of expected peptide sequences. If the peptide library contains large numbers of sequences from sources that are not contributing to the experimental sources, then the rate of false-positive identifications becomes unacceptably high. The UniProt Consortium had produced the International Protein Index (IPI) as a source of splice form-derived peptide sequences for human and other model organisms. The UniProt Consortium is committed to providing an ongoing resource for proteome-derived peptide identifications, but plans have not been finalized for how this resource will be structured.

Three-Dimensional Protein Data Resources

The Protein Data Bank

From the early days of macromolecular atomic structure determination, researchers have publicly shared results of their data as atomic models. In the mid-1970s, this was formally organized into the PDB, a computer-based repository for macromolecular structures. Today, the PDB is split over three sites RCSB PDB (USA), PDBj (Japan), and PDBe (Europe) and coordinated under the World Wide PDB (wwPDB). It holds descriptions of the atomic models derived from data collected from X-ray crystallographic, nuclear magnetic resonance (NMR), electron microscopy, fiber diffraction, neutron diffraction, and solution scattering experiments. Data are held in two formats: as a prescribed text formatted file and as a controlled condensed format called the 'crystallographic information file (CIF)'. It also holds some other data derived from the experiment such as the structure factors from X-ray experiments. This

allows the electron density used to generate the deposited atomic model to be calculated. In addition, details of the experiment itself are cataloged. Each entry is given a unique four-character alphanumeric code. Researchers are encouraged to deposit their atomic models and experimental data by requirements from scientific journals for a deposition code prior to consideration for publication.

All three sites provide access to the deposited data in the original agreed deposition format. In addition, each provides alternative views and modes of access, including dynamic creation of personal queries and result views. Likewise, resources such as PDBSum bring together additional data resources, checking, visualization, and computation of derived features in the context of the protein structure and present this in a variety of data views.

Protein 3D Models Predicted from Sequence

In comparison to the number of protein sequences that are cataloged, there is a paucity of experimentally determined structural information. This is not only due to the sheer volume of protein sequences, combined with the time and resources required to gather structural data, but also because so many proteins are not amenable to current structure determination experiments. To overcome this, a wide range of computational methods and pipelines have been developed to predict the 3D structure from sequence. Resources exist which hold predicted models automatically generated by such prediction methods, for example, ModBase, as well as online prediction services and metasearchers that take a poll from running a number of predictions from individual resources. Developments in this field can be followed via the bi-annual Critical Assessment of Structure Prediction (CASP), a community-wide experiment that aims to assess the ability to predict protein structure.

Protein Structure Classification and Integration with Sequence Data Resources

There are almost as many approaches to protein classification as there are researchers in the field. The first project driving bioinformatics was the collection of protein sequences so they could be aligned and compared. Working with comparison tools like Fasta and BLAST, and alignment tools like Clustal and Needleman-Wunsch, it became apparent that most proteins could be viewed as consisting of modules, commonly referred to as 'domains', that persisted across broad taxonomic ranges, occasionally being tandemly repeated in the same protein, duplicated in paralogous proteins, shuffled with other domains, and reorganized in linear assemblies to form different proteins. With experience, most researchers were confident that their sequence comparison methods could recognize protein domains that were probably related by evolution down to about 30% similarity. As more protein conformations were determined during the 1980s and 1990s, it was observed that most of these sequence domains possessed conserved folds and conformations, and that they represented structural domains. These structural domains were conserved and could be recognized as conformationally similar even in sequences with as little as 20% similarity. Today, a number of protein

classification schemes using powerful detection methods are cooperating to provide insights into how evolution has shaped the structures of proteins.

Class, architecture, topology, and homologous superfamily (CATH) is a hierarchal classification scheme of protein structures. Proteins are segmented into structural domains that are assigned to homologous superfamilies (groups of domains probably related by evolution). The classification is based on four criteria, class (determined according to the secondary structure composition and packing within the structure), architecture (the overall shape of the domain structure), fold or topology (sharing the same overall shape and connectivity of secondary structures in the domain core), and homologous superfamily. Another approach is used by structural classification of proteins (SCOP), in which the hierarchal classification also reflects both structural and evolutionary relatedness. The principal levels are family (with a clear evolutionarily relationship), superfamily (low sequence identities, but whose structural and functional features suggest a common evolutionary origin), and fold (proteins sharing the same major secondary structures in the same arrangement and with the same topological connections). Pfam uses an approach designed to scale with growth in the number of new protein sequences. Seed alignments are created, many automatically, from relatively stable and consistent genetic translations. The seed alignments are used to build profile hidden Markov models (HMMs) that are used to search sequence databases for potential homologs. Those homologs that score above curated thresholds are included in the alignment for iterative refinement.

InterPro is a resource integrating the results of classification methods of 11 InterPro consortium member databases (PROSITE, Pfam, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, PANTHER, Gene3D, PRINTS, HAMAP, and ProDom). The member databases use different methods based on string regular expressions, fingerprints, profiles, or HMMs to develop diagnostic signatures for protein domains and families. Together, the different signature recognition methods, coupled with statistical and biological significance tests, allow any failings in the individual methods to be recovered. A particular method for detecting sequences belonging to a certain class may miss some true examples, false negatives. Using a combination of methods increases overall sensitivity by decreasing the false-negative rate. However, this is done at the cost of increasing the false-positive rate. Taking into account the intersection of the results of the different methodologies increases the reliability of the classification and its sensitivity. The basis set of sequences for the InterPro classification methods are updated with every UniProtKB release and then used for the next round of classification of UniProtKB entries.

Protein Function Resources

Annotating and Ascribing Protein Function

Most, if not all, resources ascribe some form of annotation for the protein data they present. Annotations fall into two categories: those determined from experimental observation and those predicted using computational methods. Annotations of experimental observations are obtained either at the time of

deposition from the depositor or, more often, from extensive searching of associated literature by highly trained curators from the data resource. Emerging text-mining techniques now permit the automatic annotation using the scientific literature. Where either time to search the literature or experimental data is lacking, many data resources use computational techniques to infer functions for nonannotated entries from the manually curated entries. Most methods for automated annotation use evolutionary relationships either at the sequence level or at the structural level. Accuracy of the annotations decline as the evolutionary distances between proteins increases. Moreover, incorrect or inappropriate annotations can be easily passed on by these inference methods resulting in a phenomenon termed 'genome rot'. It is vital therefore that the provenance of the annotation be recorded and considered.

Functional Annotation Resources

The first organized attempt to classify proteins by function came when the Enzyme Commission (EC) was set up by the International Union of Biochemistry in 1955 to produce a classification and standard nomenclature for enzyme activities. The four-level numeric identifiers and standard names produced by the EC used by all bioinformatics resources were published first as books, and are now available in tables on the web with some text searching capability. In recent years, the rising flood of biochemical information has carried with it a small proportion of enzyme activity reports that has nonetheless overwhelmed the capabilities of the EC with its very limited staff to cope with the demands. Enzyme activity reports in the modern era often fail to provide many of the clear experimental results and delimitations that were expected in enzyme research before 1990, making the work of the EC more difficult. One approach to providing a bioinformatics resource for enzyme information has been the building of relational databases with web access like ENZYME Database and IntEnz with more powerful search capabilities and links to publication, chemical, and biological resources.

The appearance of the first complete genome sequences created a need for a systematic organization of biological cellular processes so that the functions of genes could be more precisely described within the limits of experimental results with varying degrees of ambiguity. Fortunately, mathematical logic, information science, and linguistic research had already produced the means of expressing ambiguity with precision – ontologies, formal representations of knowledge by a hierarchical set of concepts within a domain and the ordering relationships between those concepts. Work began on construction of the Gene Ontology (GO) in 1998 by a consortium of researchers studying the genome of three model organisms. It has three hierarchical controlled vocabulary components organized along orthogonal lines of conceptual categorization: cellular component, molecular function, and biological process. Along with the usual verbal description of what has been determined about a protein's function and role in an organism, the UniProt Knowledgebase also provides the more precise GO terms that can be associated with each protein. The GO

Consortium website provides public access to all aspects of its work from downloading of the database to participation in discussions about the definitions.

Protein Expression, Interaction, and Localization

Another important catalog of protein information relates to their expression and interaction in a cell or tissue. There are a number of experimental protocols that can characterize protein expression and interaction, including microarray analysis, yeast two hybrid, co-immunoprecipitation, and others. The results of these experiments are collated and presented in a number of resources including IntAct, BioGRID, MINT, and others. In addition to understanding interactions, other resources catalog protein-signaling pathways as well as the biochemical pathways in which protein enzymes participate. Such resources include REACTOME, KEGG, and BRENDA. To further understanding of a protein's context within a cell, other resources detail the cellular location of proteins. One of the most innovative of these has been the Human Protein Atlas, which provides a systematic exploration of the human proteome by generating fluorescently tagged antibodies against all human proteins and using them to immunohistochemically label proteins in tissues and cells, which can be observed using confocal microscopy. This is implemented in a high throughput manner using cell microarrays, allowing analysis of different cell lines and states such as those from cancer patients. The expression profiles and confocal images are presented through a web portal.

Availability and Maintenance of Resources

Changing Modes of Access to Resources

From their beginnings in the 1970s, the biological databases have depended on the availability of computer database technology. Scientific users seldom have to become involved in database programming, and probably prefer that such aspects remain invisible to them. How scientists have accessed those database resources has, however, changed most dramatically with computer technology – from books, magnetic media, and CDROMs to internet web browsing, servers, and file transfers. The biological sequences remain the primary objects of study, but as the volume of the sequence data has increased beyond comprehension, scientists have increasingly had to rely on biological database resources to provide the annotations to help them find and understand the significance of the sequences they want to study.

The ever-increasing number of protein resources (a summary of some of them described in this article is given in [Table 1](#)), and biological data resources in general, has raised issues regarding the longevity and on-going maintenance of these resources. Once a resource has been developed, it can be difficult to find funding to maintain it. The movement of the resources' developers as they follow their careers can also exacerbate the problem of maintaining a resource's presence. In addition, the expected large increase in data from the second and third generation of sequencing technologies as well as the big biology projects currently being undertaken will generate questions about how the data is to be maintained. To begin to address this, a European initiative has been started to construct

Table 1 Major protein resources

- Protein Sequence Databases
 - UniProtKB – comprehensive collection of determined and translated protein sequences with annotations that are manually curated, or automatically derived from manually curated homologous sequences (<http://www.uniprot.org/>)
 - RefSeq – integrated, non-redundant, annotated set of translated protein sequences, including genomic DNA, and transcripts (<http://www.ncbi.nlm.nih.gov/refseq/>)
 - GenBank – annotated collection of all publicly available DNA sequences (<http://www.ncbi.nlm.nih.gov/genbank/>)
 - ENA – annotated collection of all publicly available DNA sequences (<http://www.ebi.ac.uk/embl/>)
 - DDBJ – annotated collection of all publicly available DNA sequences (<http://www.ddbj.nig.ac.jp/>)
 - Entrez – multidatabase search engine (<http://www.ncbi.nlm.nih.gov/sites/gquery>)
 - Ensembl and Ensembl Genomes – comparative genome databases for vertebrates and other eukaryotic species (<http://www.ensembl.org/> and <http://www.ensemblgenomes.org/>)
 - eggNOG – database of orthologous groups of genes (<http://egglog.embl.de/>)
- Proteomic Data Resources
 - PRIDE – proteomics data resource providing a repository for protein and peptide identifications together with the evidence supporting these identifications (<http://www.ebi.ac.uk/pride/>)
 - The IMP-IMBA Bioinformatics group – portal to site predictions for GPI-anchors, N-myristoylations, prenylations, and protein kinase A phosphorylations (<http://mendel.imp.ac.at/>)
 - CBS Prediction Servers – site predictions for a number of different types of glycosylation, phosphorylation, and signal sequence processing (<http://www.cbs.dtu.dk/services/>)
 - RESID – comprehensive collection of annotations and structures for protein modifications including amino-terminal, carboxyl-terminal, and peptide chain cross-link posttranslational modifications (<http://www.ebi.ac.uk/RESID/> and <http://www.ncifcrf.gov/RESID/>)
- Protein Structure Databases
 - wwPDB – Database and tools for access PDB data (<http://www.wwpdb.org/>)
 - PDBSum – Summaries of all protein structures (<http://www.ebi.ac.uk/pdbsum/>)
 - ModBase – Database of Comparative Protein Structure Models (<http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi>)
 - LiveBench – assessment of current protein structure prediction servers (<http://meta.bioinfo.pl/livebench.pl>)
- Protein Classification
 - CATH – Hierarchical classification of protein structural relationships (<http://www.cathdb.info/>)
 - SCOP – provides a detailed and comprehensive description of the structural and evolutionary relationships between all proteins whose structure is known Structural Classifications of Proteins (<http://scop.mrc-lmb.cam.ac.uk/scop/>)
 - PFAM – Database of protein domain families represented by multiple sequence alignments and hidden Markov models (<http://pfam.sanger.ac.uk/>)
 - GENE3D – structural annotation for proteins in the main sequence databases based on CATH classification (<http://gene3d.biochem.ucl.ac.uk/Gene3D/>)
 - InterPro – integrated database of predictive protein ‘signatures’ used for the classification and automatic annotation of proteins and genomes (<http://www.ebi.ac.uk/interpro/>)
- Protein Function, Interaction and Location
 - GO – The Gene Ontology used to annotate gene and gene products (<http://www.geneontology.org/GO.database.shtml>)
 - KEGG Database – integrated database resource consisting of 16 main databases, broadly categorized into systems information, genomic information, and chemical information (<http://www.genome.jp/kegg/>)
 - ENZYME Database – a repository of information relative to the Enzymes Commission characterisation of an enzymes function using Nomenclature Committee of the International Union of Biochemistry and Molecular Biology definitions (<http://expasy.org/enzyme/>)
 - IntEnz – resource focused on enzymatic reactions and chemical nomenclature in collaboration with the ENZYME database (<http://www.ebi.ac.uk/intenz/>)
 - IntAct – proteomics data resource for curated experimental evidence of protein interactions with proteins and other molecules, and repository for submitting interaction data (<http://www.ebi.ac.uk/intact>)
 - BIOGRID – data repository compiled through curation of protein and genetic interactions from major model organism species (<http://thebiogrid.org/>)
 - MINT – database of experimentally verified protein–protein interactions (<http://mint.bio.uniroma2.it/mint/Welcome.do>)
 - REACTOME – curated pathways database encompassing many areas of human biology (<http://www.reactome.org/>)
 - BRENDA – an enzyme specific database cataloging many aspects of protein enzymes (<http://www.brenda-enzymes.info/>)
 - Human Protein Atlas – database detailing expression and localization of proteins in a large variety of normal human tissues, cancer cells, and cell lines (<http://www.proteinatlas.org/>)

The table presents some of the major protein resources as mentioned within the text and their current online location. This is by no means an exhaustive list of resources available; one web source, the Online Bioinformatics Resources Collection (OBRC) of the Health Sciences Library System, University of Pittsburgh (<http://www.hsls.pitt.edu/guides/genetics/obrc>) maintains a database containing annotations and links for more than 2600 bioinformatics databases and software tools.

and operate a sustainable infrastructure for biological information in Europe to support life science research. Its aim is to secure funding from governmental, industrial, charitable, and intergovernmental agencies to provide an infrastructure to integrate, maintain, and manage current and future European biological data resources.

Maintaining Access to Primary Publications in an Electronic Era

Another major change in how science has been done in the past 5 years is how that science is published. First, all major scientific journals are now published in on-line form. One

study of academic library journal usage found that within 3 years of the introduction of an on-line format for a journal, the on-line use could be expected to outstrip print use by a factor of ten. Academic libraries no longer have to find ever more shelf space for their print journals. In addition, with the emergence of the free-access model as a viable economic alternative for publication, the costs of journal subscriptions are not rising as fast, while the number of journals has continued to increase.

Some major scientific journal publishers have taken on the expense of converting print editions of their journals from before publishable electronic versions became available. This is an important development, and not only for the academic purposes of teaching the history of the biomolecular sciences from its inception, but also for the database annotators who now have greater in-depth access to the sources they need to provide the contexts of discovery along with the data. Especially, with the emergence of ontologies for controlled vocabulary development, having good on-line access to both current and more historic information is important for the work of database annotators.

The Role of Biocuration, Establishing, and Maintaining Protein Data Resources

Although the computer is a powerful tool for exploring protein structure, it is only a tool. The enterprise of constructing scientific databases critically depends on having curators and annotators who are both scientifically trained and skilled at conveying understanding clearly and concisely. A significant development for the work of database annotators has been the founding of the International Society for Biocuration. One of the objectives of the biocurators is to assist databases by

promoting standards for database annotation, by establishing best practices guidelines, and by adopting or improving standard operating procedures. They also want to publicize the work of biocurators for the scientific community and public funding agencies, and secure sustainable support for bioinformatic resources that are essential to research. A professional society will promote the relationship of biocurators with the communities of researchers, publishers, and database users to ensure that the databases biocurators produce meet the specific needs of each community. It will also foster the training of students in the work of biocuration.

See also: [Molecular Biology: Genome-Wide Analysis of Gene Expression.](#)

Further Reading

- Baxeianis AD (2009) The importance of biological databases in biological discovery. *Current Protocols Bioinformatics*, Chapter 1, Unit 11.
- Dutta S, Burkhardt K, Young J, et al. (2009) Data deposition and annotation at the worldwide protein data bank. *Molecular Biotechnology* 42: 1–13.
- Fitch WM (1970) Distinguishing homologous from analogous proteins. *Systematic Zoology* 19: 99–113.
- Konagurthu AS and Lesk AM (2010) Cataloging topologies of protein folding patterns. *Journal of Molecular Recognition* 23: 253–257.
- Muller J, Szklarczyk D, Julien P, et al. (2010) eggNOG v2.0: Extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Research* 38: D190–D195.
- Obst O (2003) Patterns and costs of printed and online journal usage. *Health Information and Libraries Journal* 20: 22–32.
- Sayers EW, Barrett T, Benson DA, et al. (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 38: D5–D16.
- UniProt Consortium (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Research* 38(database issue): D142–D148.

Protein Degradation

A L Goldberg, Harvard Medical School, Boston, MA, USA

Published by Elsevier Inc.

Glossary

AAA ATPase (adenosine triphosphatases (ATPases) associated with diverse cellular activities) A distinct family of ATPases characterized by the presence of highly conserved domains in their primary sequence, hexameric structure, and protein unfolding activity.

Adenosine triphosphatase (ATPase) A type of enzyme that hydrolyzes ATP and uses the energy released to drive

biological processes such as pumping ions, unfolding proteins, and moving molecules.

Folding A process in which newly synthesized polypeptides attain their native three-dimensional structures.

Molecular chaperones Specialized proteins, many are ATPases, that prevent protein aggregation and may function to facilitate proteins attaining their native conformations; some can promote protein unfolding and solubilization of aggregated proteins.

Nearly all proteins within cells and most in the extracellular space are continually turning over, being synthesized from amino acids and, after widely divergent life spans, are broken down back to amino acids. The term 'protein degradation' refers to several intracellular processes by which proteins are broken down to their basic constituents. Protein degradation occurs in all living species, from bacteria to humans, and is mediated by the sequential action of several endo- and exopeptidases. Because cells use significant amount of energy for new protein synthesis, the continual destruction of proteins might appear highly wasteful. However, this process serves several important functions that are essential for life. (1) The degradation of intracellular proteins is a highly selective and tightly regulated process that is required for removal of many regulatory proteins (transcription factors and signal transducers) and many key rate-limiting enzymes, whose rapid degradation is essential for precise regulation of metabolic pathways and for maintaining cellular homeostasis. Unlike most regulatory mechanisms (e.g., phosphorylation and ubiquitylation), peptide bond cleavage is an irreversible process, and protein degradation therefore acts as a unidirectional biological switch. (2) The only way that cells can reduce the steady-state level of a particular protein is by proteolytic degradation, and removal of certain proteins permits cells to adapt their composition to changes in the cellular environment and new physiologic conditions. This kind of adaptation occurs in mammals; for example, in the liver during fasting when enzymes for glycogen production are degraded and enzymes required for gluconeogenesis are induced. (3) Under fasting conditions, protein breakdown, especially in skeletal muscle, also serves as a source of amino acids that provides the organism with essential precursors for gluconeogenesis, protein biosynthesis, or energy production. (4) The continuous breakdown of cell proteins functions also as an important quality-control mechanism that selectively removes abnormal misfolded proteins that may arise by mutations, postsynthetic damage (e.g., by oxygen radicals), biosynthetic errors (e.g., misincorporation of amino acids and premature termination of translation), or failures of normal folding or assembly of multimers. The frequency of such events in cells is uncertain in part because the resulting misfolded proteins are rapidly degraded. (5) In higher

vertebrates, protein degradation is also critical for the function of the immune system. The mechanisms defending against both intracellular pathogens (e.g., viruses) and ones in the extracellular space (e.g., bacteria) depend on the continual production of peptide antigens by intracellular proteolytic pathways.

In eukaryotic cells, there are two principal pathways for complete degradation of proteins: the endocytic-lysosomal and the ubiquitin-proteasome pathways. The former occurs primarily within the membrane-enclosed autophagic vacuoles, and the latter operates in the soluble cytosol and nucleus (Figure 1). These pathways involve very different enzymes and serve different functions in the cell. Additional proteolytic systems are also present in some organelles (e.g., mitochondria or chloroplasts). Bacteria and archaea do not have lysosomes or ubiquitin but, nevertheless, possess soluble and membrane-bound proteases, many of which have homologs that catalyze proteolysis in mitochondria and chloroplasts. Most extracellular proteins (such as plasma proteins) also continuously turn over, mainly by uptake and degradation inside cells. One critical feature of all these intracellular degradative systems is that they require metabolic energy, unlike the typical proteases functioning in the extracellular space. As explained below, energy-dependent targeting processes (e.g., protein ubiquitylation) help insure exquisite selectivity, and the adenosine triphosphate (ATP)-dependent proteolytic enzymes allow efficient protein degradation.

Endocytic-Lysosomal Pathway

This pathway is responsible for degradation of extracellular proteins, most membrane proteins, some organelles, and some intracellular proteins. The principal proteolytic component of this pathway where most protein degradation takes place is the lysosome (or the vacuole in yeast).

Lysosomes

Lysosomes are single membrane-bound spherical organelles in the cytoplasm. They were discovered by Christian de Duve and

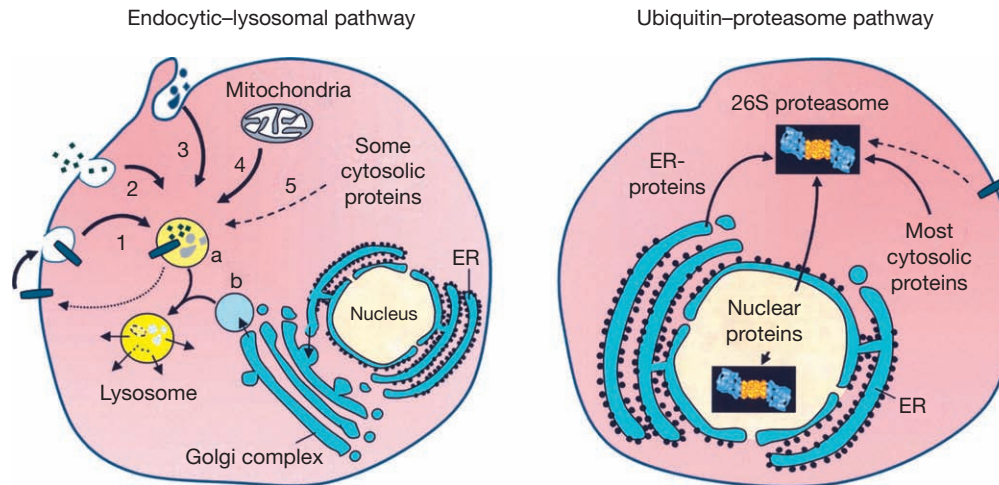


Figure 1 Pathways of protein breakdown in mammalian cells. Most intracellular proteins in the cytosol, nucleus, and ER are degraded by the ubiquitin–proteasome pathway (right panel). Extracellular and autophagocytized proteins and most membrane proteins are broken down within the endocytic–lysosomal compartment (left panel). Substrates are delivered to this pathway by receptor-mediated endocytosis (1), pinocytosis (2), phagocytosis, (3), autophagy (4), and a selective process for some cytosolic proteins containing specific amino acid motifs (5). Some internalized membrane proteins are recycled back to the plasma membrane (dotted arrows). After the internalized material is enclosed in vesicles (a), the vesicle fuses with primary lysosomes (b) derived from the Golgi complex and ER to form secondary lysosomes where final degradation takes place. Degradation products are transported across the lysosomal membrane into the cytosol. Some plasma membrane proteins are broken down by the ubiquitin–proteasome pathway, and some cytosolic proteins are degraded in lysosomes (dashed arrows). Although worn-out mitochondria are cleared by autophagy, mitochondrial proteins are selectively broken down by their own ATP-dependent proteases.

co-workers in the mid-1950s and are found in all mammalian cells except red blood cells. Lysosomes are not found in higher plants, but the plant cell vacuole performs some of their degradative functions. Lysosomes play an important role in innate immunity, adaptation to stress or starvation, recycling of cytoplasmic organelles, development, cell differentiation, and some forms of programmed cell death.

Uptake of substrates

Lysosomes receive substrates for degradation by fusing with other intracellular vesicles that are formed by endocytosis (endosomes, pinosomes, and phagosomes) or autophagy. Endocytosis refers to the process of uptake of extracellular material or membrane proteins by invagination of the cell membrane, followed by formation of vesicles inside the cell in the endosomal (ESCRT, endosomal sorting complex required for transport) pathway. Before these vesicles reach the lysosomes (pH 4.5–5.0), ingested material moves from the less acidic early endosomes (pH 6.0–6.5) to the more acidic late endosomes (pH 5.0–6.0). There are three major forms of endocytosis: (1) receptor-mediated endocytosis (uptake of membrane proteins, such as cell surface receptors and their ligands), (2) pinocytosis (ingestion of extracellular liquid with its solute molecules), and (3) phagocytosis (engulfing of large particles such as bacteria or apoptotic bodies by specialized cells such as neutrophils, macrophages, or amoeba) (Figure 1). Pinocytosis and phagocytosis are not selective bulk processes, but receptor-mediated endocytosis is a highly specific, often regulated cellular process that is used, for example, by most cells to take up extracellular products (e.g., lipoproteins and transferrin) or to reduce the levels of surface receptors (downregulation) after the receptors are occupied by ligands, or to internalize antigens using antibodies on the cell surface as receptors

(e.g., in B lymphocytes). Autophagy is highly regulated, but generally a nonselective process that digests most intracellular material (Figure 1). It serves an adaptive role in response to nutritional deprivation and stress and is responsible for selective digestion of aggregated cytosolic proteins and redundant or damaged intracellular organelles, such as mitochondria and peroxisomes. In addition, some cytosolic proteins that contain specific amino acid motifs (e.g., KFERQ) are transported into lysosomes selectively through a specific energy-, chaperone-, and receptor-mediated process (termed ‘chaperone-mediated autophagy’) that is activated by nutrient deprivation (Figure 1).

Lysosomal enzymes

Lysosomes contain at least 50 different hydrolytic enzymes, which catalyze the breakdown of proteins, nucleic acids, carbohydrates, and lipids into their components. The isolation of lysosomal hydrolases in this vesicular compartment is essential for protecting against nonspecific degradation of key cell constituents. However, during necrotic cell death, lysosomes release their content, which helps destroy the cell. Release of lysosomal enzymes outside of cells causes severe inflammation and contributes to the symptoms of some diseases (e.g., gout). Most lysosomal enzymes are optimally active at an acidic pH of 4–6, and the pH within activated lysosomes is maintained in this range by a membrane adenosine triphosphatase (ATPase) that functions as a hydrogen ion pump. Most lysosomal proteases are called ‘cathepsins’, and most belong to the family of cysteine proteases. The most abundant proteases are cathepsins D and L (endopeptidases) and cathepsins B and H (exopeptidases). The acidic milieu in lysosomes not only keeps lysosomal enzymes in their most active state, but it also helps denature protein substrates, which enhances their

susceptibility to proteolytic digestion. Once the protein substrates are degraded, the products of their digestion are transported across the membrane into the cytosol, where they are further metabolized.

Additional roles of the endocytic pathway

The endocytic-lysosomal (ESCRT) pathway plays an important role in maintaining the structural integrity of the plasma membrane by continuously recycling internalized membranes and some membrane receptors back to the cell surface. In addition, in specialized antigen-presenting cells (macrophages, dendritic cells, and B cells), some 15–20 residue peptides derived from the breakdown of extracellular proteins escape complete destruction and become bound in late endosomes to specialized membrane receptors, called ‘major histocompatibility (MHC)’ class II molecules. These complexes are transported to the cell surface for presentation to a subpopulation of T lymphocytes responsible for generation of humoral immune responses against non-native proteins (e.g., ones from pathogens).

Ubiquitin-Proteasome Pathway

Most cytosolic and nuclear proteins are degraded by the ubiquitin-proteasome pathway (Figure 2). This system catalyzes the breakdown of most long-lived proteins, which comprise the bulk of proteins in cells, and of short-lived proteins that regulate a wide variety of essential cellular processes, ranging from cell-cycle progression to signal transduction and gene transcription. In addition, this pathway catalyzes the selective elimination of misfolded, or mutated proteins, which are continuously produced due to imperfections in protein biosynthesis and folding or to postsynthetic damage. A critical clue to the discovery of the nonlysosomal ubiquitin-proteasome pathway by Hershko, Ciechanover, Goldberg, and collaborators was the finding that the breakdown of intracellular proteins was ATP dependent, which was surprising because the peptide bond cleavage by proteases is thermodynamically favored and does

not require energy. Biochemical dissection of this pathway in the early 1980s revealed that it consists of two ATP-dependent processes: the conjugation of the small protein ubiquitin to proteins, which marks them for degradation, and the hydrolysis of ubiquitylated proteins by the very large proteolytic complex, the 26S proteasome (Figure 2) (S refers to Svedberg units – the velocity at which the protein sediments in an ultracentrifuge).

Ubiquitin Conjugation

Formation of polyubiquitin chains

In order to be degraded by 26S proteasomes, most intracellular proteins must be first conjugated to a chain of ubiquitin molecules. This process ensures that only selected substrates are degraded in a regulated manner and prevents the uncontrolled degradation of other cellular proteins. Ubiquitin is a 76-residue globular protein, which is present only in eukaryotic cells, and not in bacteria and archaea, where highly selective protein degradation does occur, but through ATP-dependent proteolytic enzymes that function independently of ubiquitin. Conjugation of ubiquitin to other proteins and the formation of polyubiquitin chains is mediated by the sequential action of three types of enzymes, called E1 (ubiquitin-activating enzyme), E2s (ubiquitin-conjugating enzymes), and E3s (ubiquitin-protein ligases). In this process, the E1 enzyme activates ubiquitin in an ATP-dependent reaction by forming a highly reactive thiolester linkage between the C-terminus of ubiquitin and the thiol group in the active site of the E1. The activated ubiquitin is then passed from the E1 to a sulfhydryl group on one of the cell’s approximately 30 E2s. Next, the activated ubiquitin molecules are linked to a lysine side chain (or in some cases, to the N-terminal amino group) on the substrate protein, directly from the E2 as occurs with the Ring family of ubiquitin ligases. In the case of the homologous to E6-AP carboxy terminus (HECT) family of E3s, there is a transfer of activated ubiquitin from the E2 to the E3 before transfer to the substrate. A single E2 may function with multiple E3s to

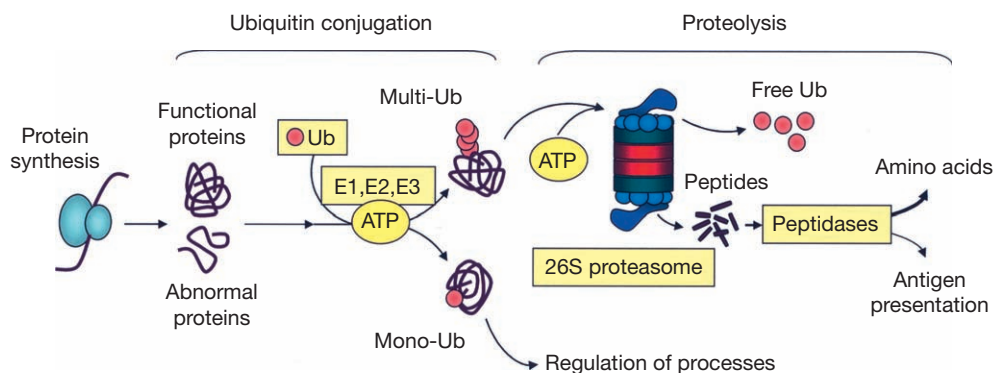


Figure 2 Ubiquitin-proteasome pathway for degradation of intracellular proteins. In eukaryotic cells, the site for degradation of most intracellular proteins is a large proteolytic particle, termed the ‘26S proteasome’. Protein substrates destined for degradation by the 26S proteasome are first polyubiquitylated in ATP-dependent manner in a series of reactions involving at least three groups of enzymes (see text for details). Polyubiquitylated substrates are rapidly hydrolyzed by the 26S proteasome and monomeric ubiquitin is recycled by the action of deubiquitinating enzymes. The energy of ATP is required for ubiquitin conjugation, binding, and unfolding of a substrate and its translocation into the inner cavities of the proteasome. Most peptides produced by proteasomes are further degraded to amino acids by endo- and exopeptidases in the cytosol and nucleus, but a small fraction of peptides escapes complete hydrolysis and are utilized for MHC class I antigen presentation. Some proteins are conjugated to only a single ubiquitin molecule. This modification does not mark proteins for degradation by proteasomes but is utilized for regulation of cellular processes, such as endocytosis and transcription.

provide specificity in a combinatorial fashion. Cells contain more than 600 distinct E3, which are the main determinants of substrate specificity and are capable of recognizing a few or multiple protein substrates that share specific degradation signals (often termed 'degrons'). E3s are soluble or membrane-bound proteins; some are monomeric enzymes containing distinct domains for binding a substrate and a specific E2 enzyme, while others are very complex oligomeric structures in which different subunits subserve different functions required for highly specific substrate recognition and efficient ubiquitylation. A polyubiquitin chain is synthesized by linking additional ubiquitin molecules to the lysine residues on the previously conjugated ubiquitin. Ubiquitin contains seven lysine residues, all of which can be involved in chain formation. Proteasomal degradation generally involves polyubiquitin chains where the ubiquitin molecules are mainly linked to lysine 48. However, ubiquitin chains in cells that are composed of K63 chains serve other regulatory roles and in the endocytic pathway target membrane proteins to lysosomes. For example, ubiquitylation of cell surface receptors (e.g., growth factor receptors) with K63 chains results in their internalization and lysosomal degradation. Once a protein is conjugated to K48 or other type of chain containing four or more ubiquitins, it is rapidly degraded by 26S proteasomes.

Mechanisms of substrate recognition

Ubiquitylation of proteins is triggered by various features of the substrates that act as signals (so-called degrons) that are recognized by the E3 and thus cause degradation of the protein. These structural features can be constitutive elements in the substrate itself (e.g., an atypical specific N-terminal amino acid residue or a short internal sequence of amino acids) or can be acquired covalent modifications. Phosphorylation of substrates is one of the frequently employed acquired signals that regulate protein ubiquitylation. In many instances, this modification leads to direct recognition of a substrate by the specific E3. However, in some cases, phosphorylation of a substrate can inhibit its ubiquitylation. Phosphorylation-dependent ubiquitylation is responsible for the degradation of I κ B during the inflammatory response and of the various cyclins that control progression through the cell cycle. Other types of posttranslational protein modifications can either prevent or stimulate ubiquitylation, including acetylation, glycosylation, or attachment of the small ubiquitin-like protein, SUMO-1. An interesting posttranslational modification that induces ubiquitylation of a substrate is the oxygen-dependent hydroxylation of a specific proline residue in the transcription factor hypoxia-inducible factor-1 (HIF1)- α , whose stabilization in hypoxic conditions plays an important role in activating the transcriptional adaptation to hypoxia. Under normoxic conditions, two prolines in HIF1- α are readily hydroxylated, leading to its recognition and ubiquitylation by a specific E3. Another mechanism for substrate recognition involves the association of the substrate with an adapter protein that is required for its recognition and ubiquitylation by the specific E3 enzyme. This strategy is particularly exploited by oncogenic strains of the human papilloma virus, which uses the cellular E3, E6AP, and the viral adapter, E6, to cause destruction of the critical host factor p53. The selective degradation of unfolded proteins in mammals involves their initial binding to specific molecular

chaperones (e.g., Hsp70), which recognizes exposed hydrophobic domains on proteins and then the chaperone-dependent E3, termed 'CHIP', ubiquitylates them. For most substrates and most E3s, the degradation signals have not yet been defined nor are the functions of most ubiquitin ligases.

Deubiquitination

Ubiquitin conjugation to cellular proteins is a reversible process and the removal of a ubiquitin moiety is catalyzed by deubiquitinating enzymes, which are cysteine proteases or metalloproteases. Three such enzymes are associated with the 26S proteasome and catalyze the disassembly of the ubiquitin chain during degradation of ubiquitin-conjugated proteins. They must play important roles in recycling monomeric ubiquitin for use in new conjugation reactions, and in preventing the accumulation of free polyubiquitin chains that could act as competitive inhibitors of the binding of ubiquitylated substrates to the proteasome. There are almost 100 deubiquitinating enzymes (often termed 'DUBs') in the human genome. These enzymes also probably serve as a proofreading system, which insures that no erroneous ubiquitylation of substrates takes place. However, there are a growing number of these DUBs that function to deubiquitinate specific substrates and to oppose the actions of specific E3s. However, mutations in certain deubiquitinating enzymes can cause neuronal degeneration, and others have been shown to contribute to the occurrence of certain cancers.

Protein Degradation

The 26S proteasome

Proteasomes are abundant evolutionarily highly conserved, large multimeric complexes found in the cytosol and nucleus of all eukaryotic cells. The 26S proteasome contains about 60 subunits, is about 2.4 MDa in size and is responsible for the degradation of ubiquitylated proteins. It consists of the 20S core proteasome and one or two 19S regulatory particles (Figure 3) and comprises 1–2% of the proteins in typical mammalian cells.

The 20S core particle

In eukaryotes, the 20S proteasome is a hollow cylinder and consists of four stacked rings, each containing seven homologous subunits. There are two identical outer α -rings and two identical inner β -rings (Figure 3). Three of the subunits in each inner β -ring contain the proteolytic active sites. The outer α -rings surround the narrow openings into the 20S particle at either end, through which substrates can enter through a gated channel into the inner proteolytic chamber. These openings are tightly regulated, and protein access is enabled through the association of α -rings with different regulatory particles. This architecture with active centers isolated within the particle clearly evolved to prevent the uncontrolled destruction of cell proteins. The six proteasomal active sites in the β -rings are confined to a central chamber within the cylinder where proteins are degraded. Each of the three active subunits has a different catalytic activity: one cleaves preferentially after basic amino acids in the protein substrate (the trypsin-like site), one after large hydrophobic residues (the chymotrypsin-like site), and one after acidic residues (the caspase-like site).

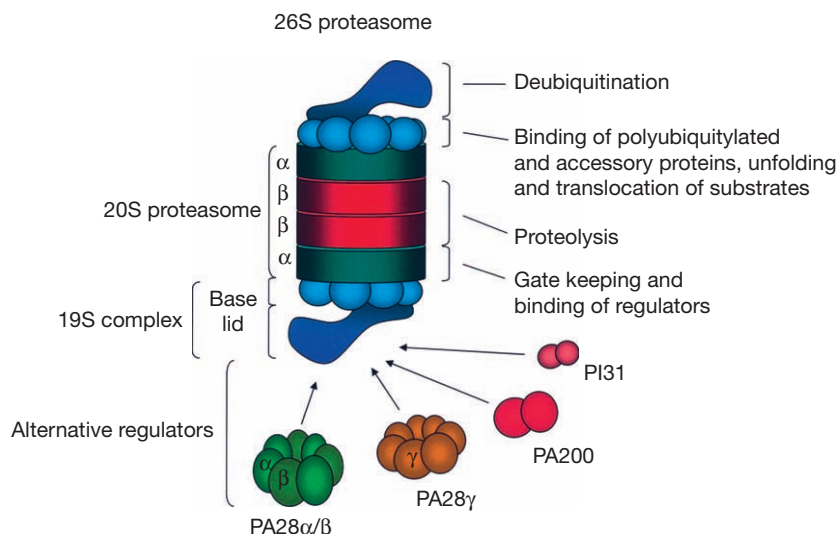


Figure 3 The 26S proteasome and its regulatory components. The 26S proteasome consists of the 20S core particle and one or two regulatory 19S particles. Different structures comprising the 26S proteasome carry out distinct functions (as detailed on the picture). In addition to 19S regulators, eukaryotic 20S proteasomes can also associate with other ring-forming complexes, such as PA28 α/β or PA28 γ , or simpler proteins (PA200 or PI31). PA28 α/β is a heptameric ring composed of α and β subunits. It promotes opening the channel enclosed by α -rings of 20S proteasomes. PA28 γ is related to α and β subunits of PA28 α/β and forms a homoheptameric complex that binds to 20S proteasomes and activates mainly the cleavages after charged residues in peptide substrates. It is predominantly found in the nucleus and, in some cells, is downregulated by interferon γ . Its biological function is still not well understood. PA200 is a recently discovered 20S proteasome-interacting protein of 200 kDa that is localized in the nucleus. The mechanism of its association with proteasomes is not well understood, but like PA28, it activates only peptide hydrolysis by proteasomes. It appears to play a role in DNA repair by recruiting proteasomes to damaged DNA. PI31 is a homodimeric protein that was discovered by its ability to inhibit peptide hydrolysis by 20S proteasomes *in vitro*. Recent *in vivo* data suggest that this protein does not affect proteasome-mediated proteolysis but rather interferes with MHC class I antigen presentation by slowing maturation of immunoproteasomes.

The proteasomes belong to a novel mechanistic class of proteases. The active site nucleophile of the proteasome is the hydroxyl group of the threonine at the amino terminus of the β -subunit. The existence of this new type of proteolytic mechanism, not utilized by other known proteases, has enabled the synthesis of highly specific proteasome inhibitors. They are now very valuable, widely used as research tools (e.g., MG132), and are also widely used (Bortezomib/Velcade) in the clinic to treat multiple myeloma and other hematologic cancers.

19S regulatory particle

The 19S particle consists of at least 18 different subunits and associates with α -rings on one or both ends of the 20S proteasome. Two functionally different substructures of this complex have been distinguished: the base and the lid (Figure 3). The base, which touches and regulates the α -rings of the 20S core particle, contains eight polypeptides including six ATPases that form a ring (Rpt1-6). Once the ubiquitin chain binds to the ubiquitin-receptor subunits (Rpn10 and Rpn13), an unfolded end of the polypeptide interacts tightly with the ATPases, and this step seems to commit the protein for degradation. These ATPases use the energy of the ATP to unfold protein substrates, open the gate in α -rings, and promote rapid translocation of the substrates into the 20S core particle for degradation. The six ATPases function in a cyclic manner and cleave two ATPs in each step causing in each step major conformational changes that drive proteasome function. For example, when a subunit binds ATP, its C-terminal residues dock into intersubunit

pockets in the 20S to cause gate-opening (like a key-in-a-lock). The base of the 19S particle also contains deubiquitinating enzymes and binding sites for ubiquitin-like domains of various shuttling factors that seem to function in bringing ubiquitylated substrates to the proteasome. One of these deubiquitinating (Usp14) has regulatory functions and when it is engaged in disassembling the ubiquitin chain, it triggers gate opening in the 20S and ATP hydrolysis to promote protein degradation. The lid contains eight non-ATPase subunits with still unknown functions.

Immunoproteasome and hybrid proteasome

In higher vertebrates, a small fraction of the peptides generated by proteasomes is not completely degraded by cytosolic peptidases to free amino acids, but instead functions in immune surveillance. They are transported into the endoplasmic reticulum (ER), where they associate with specific membrane-bound MHC class I molecules and are subsequently transported to the cell surface. The immune system is continually surveying the surface of all cells for non-native peptides presented on MHC Class I molecules that may originate from viral or mutated proteins. If cells carrying such non-native peptides are encountered, they are rapidly destroyed by cytotoxic T-lymphocytes. Proteasomes found in most of the tissues (the so-called 'constitutive proteasomes') contribute to the generation of basal levels of antigenic peptides in the absence of infection. However, during inflammation, two additional types of proteasomes appear to play a major role in enhancing MHC class I antigen presentation and are termed 'immunoproteasomes'

and 'PA28-hybrid proteasomes'. In lymphoid tissues (e.g., spleen and thymus), immunoproteasomes are present exclusively, but they are induced in other tissues during infection in response to the cytokine interferon γ . Immunoproteasomes are formed by replacement of all three catalytic β -type subunits in constitutive proteasomes with alternative proteolytically active β -subunits. The alternative immunosubunits possess qualitatively different cleavage specificities than constitutive β -subunits and appear to play a major role in increasing the supply of peptides appropriate for MHC class I antigen presentation.

Hybrid PA28-proteasomes are formed when 20S immunoproteasomes are capped on one side with the 19S particle and on the other side with an alternative protein complex, called 'PA28 α/β ' (Figure 3). Most probably, in these hybrid particles, ubiquitylated substrates initially bind to the 19S regulator, and peptide products are released at the opposite side through the PA28 α/β complex which promotes the opening of the peptide exit channel enclosed by the α -rings and also favors the generation of peptide fragments particularly suitable for MHC class I antigen presentation during inflammation.

Mechanism of protein degradation

Although the exact sequence of events in which proteins are degraded by 26S proteasomes is not fully understood, a rough outline of steps can be provided. Upon binding of the polyubiquitylated substrate to its receptors and a loose terminus of the protein to the ATPases, the substrate is unfolded in an energy-dependent manner by the mechanical actions of the 19S ATPases' conformational changes, and then the unfolded polypeptides are translocated into the 20S's inner cavity. A key event is the ATP-dependent opening of the gate in α -rings allowing translocation of the unfolded substrate into the proteasome. Successful translocation of the substrate follows disassembly of polyubiquitin chain by the three 19S deubiquitinating enzymes. Translocation can occur from its amino- or carboxyl-terminus, depending on the structure of these ends, or even by an internal polypeptide loop which can enter the axial channel to permit initial endoproteolytic cleavage. In some cases, proteasomes degrade substrates only partially, yielding biologically active protein fragments, as occurs in the generation of the p50 subunit of the transcription factor NF κ B from a larger precursor and in the processing and activation of some membrane-bound transcription factors.

Once in the central chamber of the 20S particle, the substrate is cleaved by six catalytic subunits into small peptides ranging from 2 to 25 residues in length. Unlike traditional proteases, the proteasome does not simply cleave a protein and release the two fragments; on the contrary, a protein substrate in the proteasome is generally cut multiple times until small peptides are generated. This behavior ensures that partially digested proteins do not accumulate within cells. The product release may be facilitated in some cells by PA28 α/β , as well as by hydrophobic products themselves, which can trigger gate opening in α -rings. In eukaryotic cells, the peptide products generated by proteasomes are rapidly hydrolyzed into amino acids by the sequential actions of endopeptidases and aminopeptidases. The whole process of degradation of one molecule of the protein substrate requires several hundred molecules of ATP and is completed in a matter of seconds or

minutes, depending on the size of a substrate. Generally, substrates must be first ubiquitylated in order to be degraded by proteasomes. However, 26S proteasomes can degrade certain substrates *in vivo* (e.g., ornithine decarboxylase associated with cofactor antizyme) and many polypeptides *in vitro* (e.g., denatured casein and aged calmodulin) in ubiquitin-independent manner. It still remains uncertain whether *in vivo* there are more proteins that are degraded in ubiquitin-independent manner.

ER-Associated Degradation

Newly synthesized proteins that are destined to be membrane components or to be selected are translocated into the lumen of the ER or are inserted into the ER membrane. The ER contains an abundant set of chaperones and enzymes that assist newly synthesized proteins in attaining their correct conformations and activity, but sometimes, this maturation process fails and misfolded proteins are generated. Such abnormal proteins are effectively removed so as not to aggregate and form potentially toxic inclusions. This disposal process is often called 'ER-associated degradation (ERAD)'. In it, misfolded proteins are retro-translocated from the ER back into the cytosol, and after ubiquitylation on the membrane, are degraded by 26S proteasomes. This export is necessary because the ER does not contain proteolytic enzymes necessary for protein degradation. Certain ER-associated enzymes (e.g., 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase) are continually being degraded by the ERAD pathway. A key component in extracting the ubiquitinated mutant proteins from the ER is the p97 ATPase (also called 'valosin-containing protein (VCP)', which somehow helps in the delivery of the protein to the 26S proteasomes. ERAD is also used by some viruses to downregulate MHC class I molecules in the ER (e.g., cytomegalovirus) as a means of viral evasion of the immune system. The ERAD pathway is particularly important in several human diseases, such as cystic fibrosis or certain forms of α 1-antitrypsin deficiency in which the mutant proteins are rapidly degraded and therefore unable to function (Table 1).

Proteolysis in Mitochondria

As in the cytosol and nucleus, proteins in mitochondria also have distinct half-lives and mitochondria can selectively degrade abnormally folded proteins. Mitochondria as a whole can be selectively degraded in lysosomes by autophagy (a process called 'mitophagy'). Their degradation seems to be targeted by a ubiquitin-ligase, parkin, which is mutated in a form of Parkinson's disease. However, this process cannot account for the different turnover rates of regulatory or misfolded mitochondrial proteins. Since the matrix and the inner membrane of mitochondria are not accessible to cytosolic proteases, these sites harbor their own proteolytic system, which consists of several evolutionarily conserved ATP-dependent proteases. This system has dual functions: it mediates the complete degradation of nonassembled or misfolded proteins, and it enables selective proteolysis of specific normal regulatory proteins that are important for the maintenance of mitochondrial function. The matrix of mammalian mitochondria contains two kinds of

Table 1 Disturbances of ubiquitin-dependent proteolysis in human disease

<i>Protein substrate</i>	<i>Disease</i>	<i>Mechanism</i>
<i>Increased proteolysis</i>		
p53	Cervical carcinoma	E6/E6-AP-mediated degradation of p53
p27 ^{Kip1}	Colorectal/breast Ca.	Enhanced degradation of a cell-cycle inhibitor
I κ B	Inflammation	Activation of NF κ B transcription factor
CFTR	Cystic fibrosis	Point mutation in CFTR, misfolding, and degradation through ERAD
MHC class I heavy chain	Cytomegalovirus infection	Downregulation of MHC class I by viral proteins
Most muscle proteins	Muscle atrophy, cachexia	Upregulation of an SCF E3, Atrogin-1, and the RING finger E3, MurF-1
<i>Decreased proteolysis</i>		
HIF1- α	von Hippel–Lindau disease	Mutation in SCF E3, VHL
ENaC	Liddle syndrome	Mutation in HECT E3, NEDD4
Receptor tyrosine kinases	Cancer	Mutation in RING finger E3, c-Cbl
Mitochondria	Parkinson's disease	Mutation in RING finger E3, Parkin
α -Synuclein	Parkinson's disease	Overexpression of HECT E3, Nedd4
Unknown	Angelman syndrome	Mutation in HECT E3, E6-AP
α 1-antitrypsin	Liver disease in some α 1-antitrypsin deficient patients	Decreased degradation of mutated α 1-antitrypsin by ERAD pathway
Epstein–Barr virus nuclear antigen 1 (EBNA1)	Epstein–Barr virus-associated malignancies	Immune evasion due to lack of EBNA1 Degradation

soluble proteases that are homologous to eubacterial Lon and ClpXP proteases (see below) and are responsible for degradation of matrix proteins. The mitochondrial inner membrane of all eukaryotic cells has two proteases that belong to the AAA superfamily of ATPases, termed 'i-AAA' and 'm-AAA'. These enzymes are complex structures with a native molecular mass of about 1 MDa composed of smaller subunits of 70–80 kDa. They expose their catalytic sites toward either the matrix or intermembrane space and have chaperone-like activity that is required for protein unfolding and subsequent degradation at proteolytic sites. Substrates of these proteases are proteins located in the inner membrane, and their degradation is essential for the maintenance of oxidative phosphorylation. Impaired function of these proteases causes neurodegeneration in humans and severe growth defects in yeast.

Proteolysis in Bacteria and Archaea

Roles of Proteolysis in Bacteria

Although bacteria and archaea do not have lysosomes, autophagic vacuoles, ubiquitin, or 26S proteasomes, they carry out very selective and highly regulated degradation of intracellular proteins. These prokaryotes contain several distinct ATP-dependent proteases that mediate the degradation of intracellular proteins. These enzymes serve functions similar to those of the protein degradation apparatus in eukaryotic cells. Bacteria selectively destroy abnormal proteins, including those arising from mutations, gene fusions, misfolding, chemical damage, or genetic engineering. In fact, successful expression of many recombinant proteins in *Escherichia coli* has been hampered by the action of these enzymes, and several bacterial strains lacking some of these proteases have been developed to increase the stability of cloned foreign proteins. Exposure to harsh environmental conditions, such as high temperature (heat shock), induces the genes which encode several molecular chaperones and the ATP-dependent proteases (Lon, ClpP, and HslUV) to

refold or degrade unfolded proteins that may be generated and protect against the buildup of these potentially toxic polypeptides. During exponential growth, most bacterial proteins are quite stable entities. However, even during normal growth, certain polypeptides with major regulatory functions are degraded very rapidly. For example, the specific component of bacterial RNA polymerase, so called sigma factor σ^{32} , has a half-life of a few minutes in *E. coli* under normal conditions, but it becomes transiently stabilized during heat shock, when it mediates the transcription of heat shock genes. A general enhancement of protein breakdown occurs in nongrowing cells deprived of nutrients in order to provide the starving organism with a source of amino acids for new protein synthesis or energy production. In many Gram-positive bacteria, prolonged starvation triggers sporulation, which is also accompanied by the marked degradation of most normal cellular proteins to provide amino acids for the expression of spore-specific genes.

Bacterial ATP-Dependent Proteases

Many eubacterial and archaeal ATP-dependent proteases are ancestors of more complex eukaryotic enzymes. Archaea and actinomycetes, unlike most eubacteria, for example, contain proteasomes that have similar architecture as eukaryotic 20S proteasomes, but are composed of 14 copies of a single α and a single β subunit. These particles function in protein breakdown in association with the ATPase complex, PAN, a hexameric-ring structure that is homologous to the six ATPases found in the base of the 19S component of eukaryotic proteasome. PAN has the capacity to unfold globular proteins and promote their entry into the 20S particle. The major ATP-dependent proteases in eubacteria are the cytoplasmic, the serine proteases La/Lon, 'caseinolytic proteases' ClpAP and ClpXP, the cytoplasmic threonine protease HslUV, and the inner membrane zinc-metalloprotease FtsH. These proteases

all have homologs in eukaryotic mitochondria of some species (see above). Bacterial ATP-dependent proteases are multimeric enzymes (400–700 kDa) that have distinct substrate specificities and subunit composition. They are composed of proteolytic domains and regulatory ATPase domains (e.g., Lon and FtsH) or separate regulatory ATPase complexes (e.g., ClpA, ClpX, and HslV), which actually determine substrate specificities. The simpler proteases, La/Lon and FtsH, contain these two domains in a single polypeptide chain. Lon consists of six identical 87-kDa protein subunits and was the first ATP-dependent protease described (by Chung and Goldberg in 1981). Other proteases are structurally closer to the eukaryotic 20S proteasomes (such as HslUV or ClpP), but their proteolytic subcomplexes are smaller. For example, the HslUV contains a central proteolytic chamber (HslV) composed of two hexameric rings with the active sites facing inward. Two ring-forming HslU complexes have ATPase activity and associate with proteolytic HslV cylinders on both ends to form a four-ring 20S proteasome-like particle capable of binding, unfolding, and degrading bacterial proteins. Besides proteases, the degradation of many proteins *in vivo* also requires chaperones of the Hsp70 and chaperonin (GroEL) families as cofactors that serve in the recognition of unfolded substrates. Like the proteasome, these ATP-dependent proteases generate short peptides of 10–40 amino acids, and they are degraded to amino acids by energy-independent cytosolic peptidases.

Selection of Substrates

Because bacteria lack ubiquitin and any homologous mechanism for covalent modification of substrates to be degraded, the ATPase-ring components (e.g., HslU, ClpA, or ClpX) of bacterial proteases determine their substrate specificity. With some exceptions, the exact recognition motifs on bacterial proteins are not well defined. However, special sequences on the N- and C-termini of cell proteins have been identified that lead to their recognition by ClpXP and may become particularly important in conditions when overall protein degradation is accelerated. All bacteria selectively degrade prematurely terminated proteins. A specific transfer RNA (tRNA) incorporates the tagging peptide SsrA (11-residue stretch of amino acids) into nascent polypeptide chain if ribosomes are stalled. The presence of an SsrA tag at the C-termini of a protein leads to its rapid hydrolysis by ClpXP and FtsH proteases. In addition to this recognition motif, in the case of multiprotein complexes, some proteins become substrates if present in cells without their natural binding partners, because the motifs that are recognized may be more accessible. Finally, chaperones may also provide some assistance in increasing susceptibility of substrates to bacterial proteases.

Targeting Proteolysis for Drug Development

Although a number of lysosomal diseases are known, the major ones are not due to defects in lysosomal proteases but rather

to failure to degrade polysaccharides or lipids due to lack of a specific lysosomal hydrolase. Commonly, these diseases are called 'lysosomal storage diseases' because undegraded material progressively accumulates in lysosomes. A number of major malignant, neurological, metabolic, and viral diseases are caused by the gain or loss of function of the ubiquitin–proteasome proteolytic system (Table 1). Since the disturbances in the ubiquitin–proteasome system underlie the pathogenesis of many human diseases, efforts to develop drugs that target this system are being pursued. The proteasome inhibitors are the first success in these efforts. In fact, the inhibitor, Bortezomib/Velcade, has proved to be very effective in the treatment of multiple myeloma, a cancer of the plasma cells. This agent has prolonged the lives of thousands of patients and its wider applications and other proteasome inhibitors are under intense investigation. However, more selective and more effective therapies are desirable, and efforts to develop small molecule inhibitors that target events upstream of proteasomes (e.g., ubiquitin ligation or deubiquitination and substrate recognition) may eventually allow the stabilization of desirable proteins or the clearance of specific pathogenic proteins.

See also: Protein/Enzyme Structure Function and Degradation: Aminopeptidases; Aspartic Proteases; Proteases in Blood Clotting; Signaling: B-Cell Antigen Receptor; T-Cell Antigen Receptor.

Further Reading

- Ciechanover A (2005) Proteolysis: From the lysosome to ubiquitin and the proteasome. *Nature Reviews Molecular Cell Biology* 6: 79–87.
- Cuervo AM (2008) Autophagy and aging: Keeping that old broom working. *Trends in Genetics* 24: 604–612.
- Finley D (2009) Recognition and processing of ubiquitin–protein conjugates by the proteasome. *Annual Review of Biochemistry* 78: 477–513.
- Glickman MH and Ciechanover A (2002) The ubiquitin–proteasome proteolytic pathway: Destruction for the sake of construction. *Physiological Reviews* 82: 373–428.
- Goldberg AL (2010) Bortezomib's scientific origins and its tortuous path to the clinic. In: Anderson KC, Richardson PG, and Ghobrial I (eds.) *Bortezomib in the Treatment of Multiple Myeloma*, pp. 1–27. Basel: Springer.
- Gottesman S (1996) Proteases and their targets in *Escherichia coli*. *Annual Reviews in Genetics* 30: 465–506.
- Hoeller D and Dikic I (2009) Targeting the ubiquitin system in cancer therapy. *Nature* 458: 438–444.
- Käser M and Langer T (2000) Protein degradation in mitochondria. *Seminars in Cell and Developmental Biology* 11: 181–190.
- Kisselev AF and Goldberg AL (2001) Proteasome inhibitors: From research tools to drug candidates. *Chemistry and Biology* 8: 739–758.
- Peth A, Uchiki T, and Goldberg AL (2010) ATP-dependent steps in the binding of Ub conjugates commit them to degradation by the 26S proteasome. *Molecular Cell* 40: 671–681.
- Rabl J, Smith DM, Yu Y, et al. (2008) Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. *Molecular Cell* 30: 360–368.
- Rock KL, York IA, Saric T, and Goldberg AL (2002) Protein degradation and the generation of MHC class I-presented peptides. *Advances in Immunology* 80: 1–70.

Protein Folding and Assembly

D P Goldenberg, University of Utah, Salt Lake City, UT, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Amyloid A polymeric and fibrous form of a protein, usually formed by polypeptides that have undergone unfolding or misfolding.

Domain A compact structure composed of 50–150 amino acid residues that can fold independently of other parts of a protein.

Folding intermediate A partially folded conformational state that may form transiently during the conversion of an unfolded protein to the native state.

Molten globule A partially folded state that is characterized by the presence of regular secondary structure (usually

α -helices), but lacking well-defined interactions between side chains distant in the sequence.

Native state The well-defined folded structure of a protein usually associated with biological activity.

Unfolded state An ensemble of rapidly interconverting conformations with little or stable structure, usually generated by incubating a native protein with a denaturant or at elevated temperature. *In vivo*, polypeptides that have not yet folded into their native conformations may resemble artificially unfolded proteins.

Introduction

The folding and assembly of newly synthesized proteins represents a key step in the expression of genetic information, in which one-dimensional information is converted into three-dimensional structure. The central importance of this process has long been recognized, and a wide variety of approaches – experimental, theoretical, and computational – have been applied to studying folding both *in vitro* and *in vivo*. The goals of this research have been to understand the physical principles that determine the three-dimensional structures of proteins, as well as the biological aspects of the folding process, including the coupling of folding with other cellular processes and the consequences of misfolding.

The past decade has seen particularly rapid progress in the application of new experimental techniques, including single-molecule methods, to the study of protein folding, as well as advances in computational techniques. Although much remains to be learned, and some aspects are still controversial, there is a growing consensus regarding the general features of the folding process, and the goal of this article is to provide an overview of our current understanding of both the physical-chemical features of folding reactions and some of the biological aspects.

Another important advance in the past several years has been the recognition that many proteins, or parts of proteins, do not actually fold into stable structures. In some cases, these intrinsically disordered proteins do fold when they assemble with other molecules, but, in other cases, the protein may function without taking on a unique structure. While this class of proteins lies outside of the scope of this article, they are the subject of much current research and new biological implications are likely to emerge in the coming years.

Thermodynamics of Protein Folding

The thermodynamic stability of a protein is defined by the equilibrium between the native folded form and the highly

disordered unfolded state that is induced by elevated temperatures or chemical denaturants, such as urea or guanidinium chloride (GuHCl). Relatively small monomeric proteins, composed of 50–150 amino acid residues, typically unfold and refold efficiently and cooperatively, so that the process can be described as a simple equilibrium:



where N and U represent the native and unfolded states (Figure 1). While the native state usually has a well-defined three-dimensional structure, the unfolded state is best thought of as a broad distribution of rapidly interconverting conformations.

Under physiological conditions, the free-energy change for unfolding (ΔG_u) for a small monomeric protein typically lies in the range of 5–15 kcal mol⁻¹. This relatively small preference for the native protein reflects a delicate balance between three major factors: (1) a large loss of entropy as the chain is converted from a highly disordered state to an ordered state, (2) the transfer of nonpolar groups from an aqueous environment to a nonpolar environment in the protein interior, and (3) the formation of stabilizing interactions within the protein, including hydrogen bonds, salt bridges, and van der Waals interactions.

During the 1980s and 1990s, a great deal of effort was devoted to measuring the stabilities of genetically modified proteins in which single amino acid residues were replaced so as to delete one or a few stabilizing interactions. Taken together, these studies revealed that proteins are surprisingly tolerant of amino acid replacements, so that most mutations destabilize the native state by only a few kilocalories per mole or less, and substitutions that completely prevent folding are quite rare. Single substitutions that remove hydrogen bonds or nonpolar interactions cause, on average, roughly similar degrees of destabilization (usually of the order of 1–3 kcal mol⁻¹), implying that both types of interactions play important roles in stabilizing native proteins.

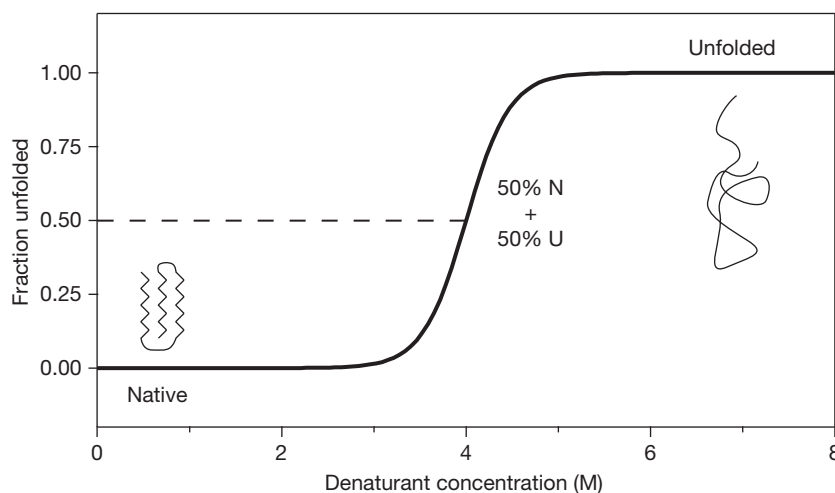


Figure 1 A cooperative protein unfolding–refolding transition. In a typical experiment, the native protein is incubated in solutions with increasing concentrations of denaturant such as urea or guanidinium chloride, and the fraction of unfolded molecules is determined by monitoring a spectroscopic signal, such as UV absorbance, fluorescence, or circular dichroism. For small monomeric proteins, the transition is usually highly cooperative, so that only the fully folded and unfolded forms are significantly populated and the fraction unfolded represents the relative concentrations of the two forms.

Folding Kinetics and Mechanisms

Levinthal's Paradox

Much of the work on folding mechanisms, both experimental and theoretical, has been motivated by the realization that the number of possible conformations accessible to a polypeptide chain is so immense that random sampling to find the structure with minimum free energy would take a ridiculously long time. For instance, if a chain is composed of 100 residues, and each residue can take on 10 possible conformations, then the total number of conformations is 10^{100} . If the various conformations are randomly sampled at a rate of 10^{13} per second (approximately the rate of rotation for a covalent bond), then the total time required to sample all of the conformations would be approximately 10^{80} years!

The calculation outlined above is frequently identified as the 'Levinthal paradox', in recognition of an insightful (but difficult to find!) 1969 paper by Cyrus Levinthal. Although this argument undoubtedly represents a great oversimplification, since many of the conformations assumed to be accessible to the chain will, in fact, be excluded by steric clashes among atoms in different residues, the calculation suggests that there must be some mechanism by which the search for the native conformation is limited.

Folding and Unfolding Rates

The rates at which different proteins fold and unfold vary tremendously. Even considering only relatively small proteins composed of a single domain, experimentally determined folding rate constants cover at least the range from 0.1 to 10^5 s^{-1} . While it might be expected that smaller and thermodynamically more stable proteins would fold the fastest, folding rates for unrelated proteins are poorly correlated with size or stability. Instead, a major determinant of folding rate appears to be the topological complexity of the native structure. While true topological knots are rarely found in folded proteins, different

structures vary with respect to the relative prevalence of interactions formed between closely and distantly spaced residues, and faster folding is often associated with topologies in which short-range interactions are more prominent. The physical basis for this correlation is not yet fully understood, but it strongly suggests that the rate-determining step in folding is associated with formation of the correct topology, as opposed, for instance, to formation of a local structure that then nucleates rapid folding of the rest of the chain.

Unfolding rates also vary greatly among different proteins and under different solution conditions. For some proteins, unfolding may be a surprisingly common event, even under physiological conditions. For instance, if the stability of a protein, ΔG_u , is 5 kcal mol^{-1} at 25°C , and the folding rate is 10^3 s^{-1} , the unfolding rate is predicted to be about 0.5 s^{-1} , so that 99% of the molecules in a population will undergo an unfolding event over a period of 10 s. Since unfolded proteins are typically much more sensitive to proteolysis than are folded proteins, spontaneous unfolding may be an important factor in determining the rate of protein turnover *in vivo*. Transient unfolding may also contribute to formation of aberrant structures, such as the amyloid fibers described later.

Folding Intermediates

Following in the tradition of biochemists who have successfully dissected other processes by identifying and characterizing discrete intermediates, many investigators have studied kinetic intermediates that accumulate transiently before protein-folding reactions reach equilibrium. While there is considerable diversity in the intermediates observed in the folding of different proteins, there appears to be common features among many of them.

Generally, the most easily detected intermediates form very rapidly (on the millisecond timescale or faster) after the unfolded protein is transferred to conditions that favor the native state. Often, the intermediate appears to contain

ordered secondary structure, especially α -helices, but there is little evidence of well-defined interactions between the side chains from different segments of the chain. The average dimensions of the chain in these intermediates typically lie between those of the native protein and the fully unfolded state in denaturants. Kinetic intermediates of this type also resemble partially unfolded states that can sometimes be generated under special solution conditions, including low pH for some proteins, or when a ligand (such as a metal ion or heme group) is removed from a native protein. These states, typically characterized by the presence of regular secondary structure but the absence of stable tertiary interactions, are described as molten globules.

As the most frequently observed partially folded states often interconvert rapidly with the fully unfolded protein, establishing whether or not these species are actually productive intermediates in folding, as opposed to products of side reactions, is quite difficult. As a consequence, there is presently a great deal of uncertainty and disagreement about the significance of these states for understanding folding mechanisms. Some authors believe that the formation of intermediates is important for limiting the search for the native conformation, while others argue that they are simply compact forms of the denatured state and do not play any special role in directing the chain toward the native state.

Transition States

It might seem surprising that the concept of a transition state, derived from studies of simple chemical reactions, could be usefully applied to a process as complex as protein folding. However, the folding kinetics of many proteins can be well described by a simple model in which the rate is determined by an equilibrium between the unfolded state and a higher energy state that is rapidly converted to the native conformation (Figure 2). The transition state can be envisioned as an

ensemble of partially ordered conformations that collectively have an equal probability of being converted to either the native or fully unfolded states.

By their very nature, transition states cannot be characterized directly by spectroscopic or other physical methods. Instead, their properties must be inferred from the results of experiments in which the kinetics of folding and unfolding are compared after perturbing either the structure of the protein or its environment. For instance, replacing an amino acid residue that contributes to the stability of the transition state is expected to increase the free-energy barrier for folding and, thereby, decrease the folding rate. On the other hand, a substitution that weakens an interaction that is present in the native protein but not the transition state is expected to increase the unfolding rate, without affecting the folding kinetics.

Detailed mutational analyses of transition states have now been carried out for several proteins. In general, it has been found that substitutions that destabilize the native protein most often increase the unfolding rate more than they decrease the folding rate, suggesting that most interactions that are present in the native protein are significantly weakened in the transition state. In addition, the relatively few residues that appear to play specific roles in stabilizing the transition state are often separated in the amino acid sequence but lie relatively close together in the native protein. These patterns suggest that the transition state ensemble may have a topology similar to that of the native protein, but not the network of cooperative interactions found in the fully folded structure (Figure 2). Poised with the correct topology, the molecules in the transition state can not only rapidly form these interactions, but also unfold rapidly.

Collectively, studies of folding intermediates and transition states over the past decade generally support a view of folding in which there is considerable breadth in the distribution of conformations representing different stages of the process, a view that has also emerged from computational simulations of folding. The conformational heterogeneity of these states makes detailed structural analysis very difficult, and there is much about folding mechanisms that still remains hidden from experimental view. More detailed descriptions of folding can, in principle, be derived from computational simulations, and the rapidly increasing power of computers and the development of new algorithms is now making it possible to carry out detailed molecular dynamics simulations of folding reactions. The past several years have seen much more extensive collaborations between experimentalists and computational scientists working on protein folding, and this development is already leading to important new insights into both the details of specific folding reactions and the general principles of the process.

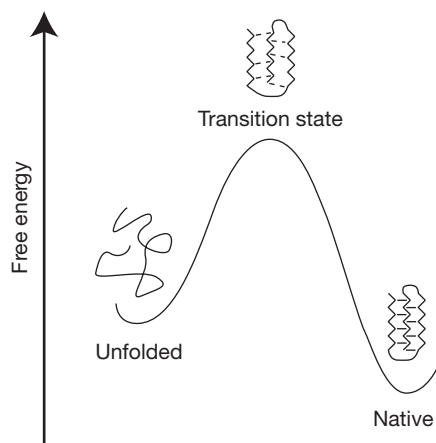


Figure 2 Hypothetical energy profile for a protein-folding reaction. Currently available data, especially from mutational analyses of folding kinetics, suggest that the transition state for folding and unfolding is an ensemble of conformations that possess much of the topology of the native protein, but in which many of the native interactions are missing or weakened. The transition state ensemble is poised to fold completely or unfold rapidly.

More Complex Folding Reactions

Although the great majority of protein-folding experiments and simulations has been carried out with small monomeric proteins, most proteins found in nature are larger and have more complex structures. Most polypeptides form multiple domains when they fold, and many assemble into structures composed of multiple subunits. In addition, some proteins are stabilized by

disulfide bonds, which are formed during the folding process, and others contain tightly bound ligands such as metal ions, heme groups, or cofactors required for enzyme activity.

Proteins with Multiple Domains

Early in the study of protein structure, it was recognized that most proteins composed of more than 100–200 amino acid residues are typically organized into domains in which segments of 50–150 residues are folded into compact structures that are relatively independent of one another. Although a precise definition of a domain is somewhat elusive, the organization of a protein into substructures is often quite apparent from inspection of the three-dimensional structure or by computer analysis (Figure 3). Experimentally, domains can often be defined by limited treatment with a protease, since the polypeptide segments separating domains are usually much more sensitive to digestion than the regions within domains. The functions of larger proteins are often divided among their substituent domains, with different domains typically interacting with different molecules. It is clear from comparisons of protein structures and sequences that the same domains have frequently been exchanged among different proteins by genetic recombination, and the domain architecture of a protein can often be deduced by comparison of its sequence with those of known domains.

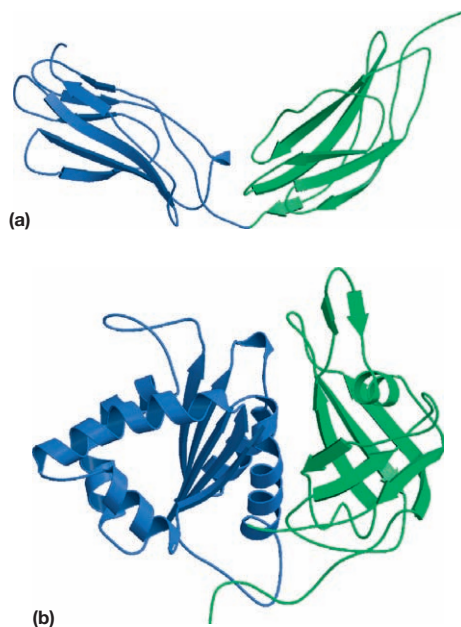


Figure 3 Domain structures of two proteins. In each drawing, the domains are distinguished by color. (a) Two domains of a cellular adhesion molecule, neuroglian from *Drosophila melanogaster*. The intact protein contains nine other domains, three of which are similar to the two shown here, which are members of the fibronectin III domain family. (b) Ferredoxin reductase from spinach chloroplasts. In this structure, the two domains do not resemble one another and interact much more extensively than the two fibronectin domains shown in (a). (a) Drawn from the atomic coordinates of entry 1CFB in the Protein Data Bank. (b) Drawn from the atomic coordinates of entry 1FNB in the Protein Data Bank. The drawings in Figures 3 and 4 were made with the computer programs MOLSCRIPT and Raster3d.

When isolated from the rest of the protein, individual domains typically retain their native conformations and undergo cooperative unfolding transitions similar to those observed with smaller proteins. Furthermore, when the intact protein unfolds, the resulting spectroscopic signals often resemble a simple superposition of those for the isolated domains. In these cases, the domains appear to be truly independent, and studying their folding individually likely provides a good description of the folding mechanism for the intact protein. In other cases, however, the domains may interact more intimately, so that the entire structure unfolds cooperatively. The folding mechanism may also involve interactions among the domains before their individual structures are fully defined.

A relatively recent advance has been the use of single-molecule techniques, especially atomic force microscopy and laser traps, to monitor the mechanically induced unfolding (stretching) of multi-domain proteins. These studies have provided direct evidence for the modular structures of these proteins, as well as new insights into the elastic properties of proteins.

Disulfide-Coupled Folding

Many proteins, especially among those that function extracellularly, are stabilized by disulfide bonds, covalent bonds between the sulfur atoms of cysteine (Cys) residues that are formed by oxidation of the Cys thiol groups. *In vivo*, disulfide bonds are formed in concert with folding after the polypeptide is synthesized and translocated to either the lumen of the endoplasmic reticulum or the intermembrane space of mitochondria in eukaryotic cells, or the periplasmic space of bacteria. In all three of these compartments, disulfides in the newly synthesized proteins are formed by exchange with disulfides present in special proteins that act as catalysts of the process.

In a landmark series of experiments in the early 1960s, C B Anfinsen and his colleagues demonstrated that the four disulfide bonds of bovine ribonuclease A could be induced to reform spontaneously *in vitro* after the protein had been fully reduced and unfolded in urea. The regeneration of active enzyme, even after the covalent constraints on the polypeptide chain had been broken, provided a particularly clear demonstration that the three-dimensional structure was encoded by the amino acid sequence.

When refolding of a reduced protein is initiated, disulfides typically form quite randomly, leading to a distribution of molecules with different disulfide bonds that can rapidly interconvert. The formation of an initial disulfide, however, can often favor conformations that then bring other Cys residues together to form a second disulfide. These conformations can also be stabilized by noncovalent interactions, leading to a high degree of cooperativity among the various interactions.

As more of the three-dimensional structure forms and is stabilized by additional disulfides, the ability of the molecule to undergo additional disulfide formation and rearrangement steps can become severely constrained, leading to the accumulation of species that contain most or all of the native structure but lack one or more disulfides. *In vivo*, disulfide isomerase enzymes can catalyze the slow rearrangements necessary to convert the trapped molecules to the native form, or may promote alternative pathways that avoid the kinetic traps.

Ligand Binding

Yet another way in which folding can be influenced is through the binding of ligands such as metal ions or prosthetic groups. In most cases, ligands have little tendency to interact with unfolded proteins, so that the molecule must be at least partially folded before the interaction takes place. A thermodynamic consequence of this is that the stability of the folded protein is enhanced by the presence of ligand and depends on both the affinity of the interaction and the concentration of free ligand in solution. If the binding site is formed only when the protein is fully folded, binding may not greatly alter the folding process either *in vitro* or *in vivo*. On the other hand, interactions of ligands with partially folded species could favor some pathways over others and might prevent aberrant interactions with other molecules.

Proteins that require the binding of reactive metal ions, such as those of iron, copper, or nickel, present a particularly interesting problem during folding *in vivo*. As these ions are highly toxic, their free concentrations are maintained at very low levels by transport and storage proteins that bind tightly to them. The ions must then be transferred to other proteins for which they are required for activity, such as catalysis. Although the details are not yet well understood, it appears that quite elaborate systems involving multiple transfer proteins have evolved to facilitate the safe transfer of ions to their proper sites.

Some prosthetic groups, such as the heme group of cytochrome *c*, are covalently bonded to the polypeptide chain. For these proteins, the covalent modifications, and the enzymes that catalyze them, are likely to be intimately involved in the folding process.

Molecular Chaperones

While many proteins refold spontaneously and efficiently *in vitro*, it is now known that folding *in vivo* is often facilitated by molecular chaperones, proteins that interact transiently with the polypeptide chain and either promote folding or decrease the likelihood of nonproductive interactions. The best-characterized molecular chaperones include members of the GroEL/GroES and HSP70 families. Members of both groups utilize the energy of adenosine triphosphate (ATP) hydrolysis to couple conformational changes in the chaperone to a cycle of substrate protein binding and release, during which the substrate folds into its native conformation. Other examples of molecular chaperones include the catalysts of disulfide formation and exchange and the metal chaperones mentioned above. Although the chaperones can increase the efficiency of folding, it does not appear that they influence the final structure. Their discovery, therefore, has not required a revision of the fundamental idea that the folded structure of a protein is determined by the amino acid sequence. However, we now have a much greater appreciation of protein folding as a biological process that is intimately coupled to numerous other intracellular events.

Assembly of Oligomeric Proteins

Only a minority, perhaps 20%, of proteins contain a single polypeptide chain, the remainder being assemblies of multiple

subunits. Dimeric and tetrameric complexes are the most common, but assemblies of 20–50 subunits are also known. The assembly of subunits into larger complexes can offer a variety of functional advantages, including mechanisms for regulating the activities of enzymes, the creation of novel binding sites at subunit interfaces, and the coordination of multiple related functions in a single complex.

Multimeric proteins can be assembled from a single type of subunit, resulting in a homooligomer, or they can be heterooligomers composed of two or more different kinds of chains. Quite often, these structures are highly symmetrical. Dimers of identical polypeptide chains are especially common, and dimers are often the basic units of higher-order structures, such as tetramers, hexamers, etc.

The types of interactions that hold the subunits of oligomers together are essentially the same as those found within monomeric proteins or the individual subunits themselves, including hydrogen bonds, ionic, and van der Waals interactions. However, surveys of the distribution of different interaction types in various protein environments have revealed some differences between those found between and within subunits, with ionic interactions being more common in the interfaces between subunits.

Many oligomeric proteins have been successfully refolded and assembled *in vitro* after being unfolded in denaturants, though the reactions are often slower and less efficient than the folding of small monomeric proteins. In some cases, the rate of refolding is dependent on protein concentration, indicating that the rate is determined by intermolecular association reactions. In other cases, the reaction kinetics are limited by intramolecular rearrangements that occur either before or after subunit association.

These differences in kinetic behavior suggest that oligomeric structures may form by a variety of mechanisms. In order to assure specificity, the initial interactions between two subunits presumably require that one or both of them first takes on at least a partially folded conformation. In some cases, the interaction may only occur after the subunits are fully folded, so that folding and assembly are distinct processes. This mechanism requires that the individual subunits have compact self-contained structures similar to those of monomeric proteins, as epitomized by the four subunits of hemoglobin (Figure 4(a)). For some oligomers, however, the subunits form extensive interactions with one another along large interfaces that make it unlikely that the individual subunits could fold completely before assembling. In extreme cases, the subunits are actually wrapped around one another, as observed in the structure of the tailspike protein of bacteriophage P22 (Figure 4(b)). These more complex arrangements of subunits likely lend the structures a great deal of kinetic stability, but possibly at the cost of more complex folding mechanisms involving unstable intermediates that may be prone to aberrant misfolding or aggregation reactions.

Protein subunits can also form complexes with either ribose nucleic acid (RNA) or deoxyribose nucleic acid (DNA), as in ribosomes or the nucleosomes of eukaryotic chromatin. These interactions are often mediated by electrostatic interactions between Arg or Lys residues of the proteins and the negatively charged phosphate groups of the nucleic acids, but nonpolar interactions can also play a role. Just as protein subunits can

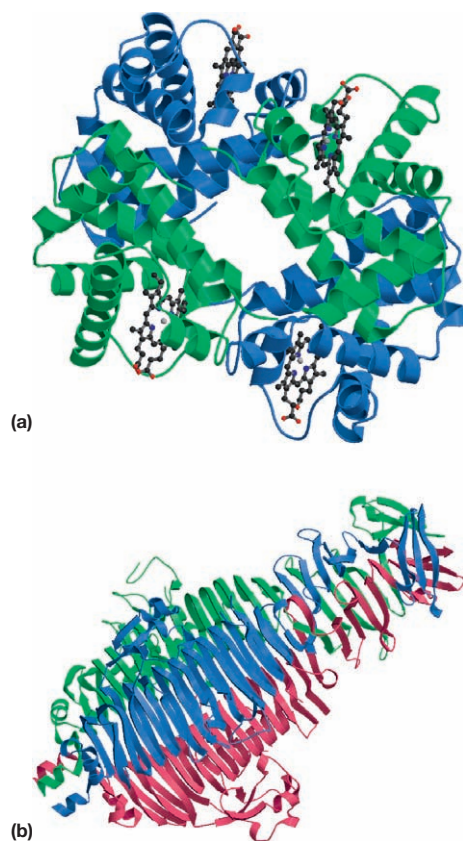


Figure 4 Subunit structures of two proteins. (a) Human hemoglobin, a tetramer formed of two types of subunits, colored green and blue in the drawing. Each of the subunits is a compact and self-contained structure. The heme groups associated with each of the four chains are also shown in a ball-and-stick representation. Drawn from entry 2HHB in the Protein Data Bank. (b) A fragment of the trimeric tailspike protein of bacteriophage P22. The intact protein includes an additional domain at the amino-terminus of the chain (at the lower left corner of the drawing), by which the tailspike binds the rest of the virus particle. In this structure, the individual subunits interact extensively with one another and are intertwined, requiring an assembly mechanism in which the chains do not fold completely until after they have associated. Drawn from entry 1TSP of the Protein Data Bank.

associate with one another either before or after being fully folded, the formation of protein–nucleic acid complexes may involve folding processes that take place after association as well as before.

Protein Misfolding and Disease

Over the past decade, defects in protein folding or assembly have become recognized as important factors leading to a variety of diseases. These include cystic fibrosis, some forms of emphysema, and a large number of conditions associated with the formation of long fibrillar protein structures, the amyloid diseases. Mutations that destabilize the native protein may lead to premature degradation or the formation of aberrant assemblies, which, in turn, may lead to a deficiency for

an important activity or the formation of structures that are actually toxic.

At present, the amyloid diseases appear to be the most important category of folding pathologies. Diseases associated with amyloid formation include neurodegenerative conditions such as Alzheimer's and Parkinson's diseases and amyotrophic lateral sclerosis, as well as several systemic amyloid diseases. In an intriguing minority of diseases, particles of amyloid fibers from one individual can promote fibril formation in another, providing a mechanism of infectivity that does not involve either bacteria or viruses. These infectious agents have been named prions and the diseases they are associated with include scrapie (in sheep) and bovine spongiform encephalopathy (BSE or mad cow disease), as well as the human disease Kuru and a variant of Creutzfeldt–Jacob disease that is believed to be derived from BSE. However, the transmission of an altered physiological state by a protein amyloid is not at all restricted to mammals, as a very similar process has been found in the unicellular yeast, *Saccharomyces cerevisiae*, and this discovery has provided an important experimental system for analyzing the prion phenomenon.

A surprisingly large number of proteins are now known to form amyloid-like fibers under at least some *in vitro* conditions, and it may be that the ability to form these structures is a nearly universal property of polypeptide chains. Remarkably, there is a great deal of similarity among the structures of fibers formed from proteins with very different sequences and native structures. The fibers are generally about 50–100 Å in diameter and a fraction at least of the polypeptide chain has a β -strand conformation, with the strands oriented so that their backbones are approximately perpendicular to the fiber axis.

Fibril formation *in vitro* is typically favored by conditions that destabilize the normal folded conformation but do not fully unfold and solubilize the polypeptide. *In vivo*, conversion to the amyloid state may be promoted by inherited mutations, environmental stress, or, as noted above, seeding by an existing fibril fragment. At present, it is difficult to establish a clear mechanistic link between fibril formation and the physiological changes associated with amyloid diseases. In some cases, the amounts of protein deposited in fibrils are so large, as much as several kilograms in some patients, that these structures are clearly able to interfere with normal organ function. Smaller fibrils may similarly be toxic at the cellular level. In other cases, it may be that diversion of essential proteins into fibrils leads to functional deficiencies.

Although most known examples of amyloid formation are associated with detrimental effects, there is a growing list of cases in which the amyloid form of a protein has a distinct function. These include regulation of gene expression by the *Saccharomyces* proteins mentioned above, formation of melanin granules in the human epidermis, and storage of proteins in secretory granules.

As amyloid fibril formation usually competes with normal folding, investigators are actively pursuing the possibility that these diseases might be prevented or treated by agents that favor productive folding or disfavor unfolding, for instance, by binding to the native structure. These developments provide a vivid link between the basic science of protein folding and an important medical problem.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Glycoprotein Folding and Processing Reactions; **Protein/Enzyme Structure Function and Degradation:** Chaperonins; Disulfide Bond Formation.

Further Reading

- Bartlett AI and Radford SE (2009) An expanding arsenal of experimental methods yields and explosion of insights into protein folding mechanisms. *Nature Structural and Molecular Biology* 16: 582–588.
- Brockwell DJ and Radford SE (2007) Intermediates: Ubiquitous species on folding energy landscapes. *Current Opinion in Structural Biology* 17: 30–37.
- Chiti F and Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. *Annual Reviews in Biochemistry* 75: 333–366.
- Cobb NJ and Surewicz WK (2009) Prion diseases and their biochemical mechanisms. *Biochemistry* 48: 2574–2585.
- Daggett V and Fersht A (2003) The present view of the mechanism of protein folding. *Nature Reviews Molecular Cell Biology* 4: 497–502.
- Dill KA, Ozkan SB, Shell MS, and Weikl TR (2008) The protein folding problem. *Annual Reviews in Biophysics* 37: 289–316.
- Gsponder J and Babu MM (2009) The rules of disorder or why disorder rules. *Progress in Biophysics and Molecular Biology* 99: 94–103.
- Hartl FU and Hayer-Hartl M (2009) Converging concepts of protein folding *in vitro* and *in vivo*. *Nature and Structural Molecular Biology* 16: 574–581.
- Levinthal C (1969) How to fold graciously. In: DeBrunner J and Munck E (eds.) *Mossbauer Spectroscopy in Biological Systems*, pp. 22–24. Urbana, IL: University of Illinois Press.
- Mamthambika BS and Bardwell JC (2008) Disulfide-linked protein folding pathways. *Annual Reviews in Cell Developmental Biology* 24: 211–235.
- Riemer J, Bulleid N, and Herrmann JN (2009) Disulfide formation in the ER and mitochondria: Two solutions to a common process. *Science* 324: 1284–1287.
- Wilson CJ, Apiyo D, and Wittung-Stafshede P (2004) Role of cofactors in metalloprotein folding. *Quarterly Reviews in Biophysics* 37: 285–314.

Protein Import into Mitochondria

J M Herrmann, University of Kaiserslautern, Kaiserslautern, Germany

W Neupert, Max-Planck-Institute of Biochemistry, Munich-Martinsried, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

Oxa1 translocase Protein complex of the inner membrane of mitochondria that facilitates membrane insertion of both nuclear and mitochondrially encoded proteins.

Preprotein Precursor form of a mitochondrial protein; preproteins often contain presequences that have to be removed by mitochondrial processing peptidases.

Presequence Typically N-terminal extension on a protein that serves as mitochondrial targeting signal.

Translocase of the outer membrane (TOM) Multisubunit translocation complex of the outer membrane of mitochondria consisting of receptor subunits and a protein-translocating channel.

Translocases of the inner membrane (TIMs) Oligomeric translocation complexes of the inner membrane of mitochondria.

Mitochondria are essential organelles of eukaryotic cells that accommodate nearly 10–20% of the cellular proteome. They are made up of two membrane systems that divide the organelle into two aqueous subcompartments: the intermembrane space, between the outer and the inner membrane, and the matrix, which is enclosed by the inner membrane. The vast majority of mitochondrial proteins are encoded by the nuclear genome and synthesized on cytosolic ribosomes as precursors (also called preproteins). These proteins contain targeting information for specific sorting of each polypeptide to its respective mitochondrial subcompartment. Protein import into mitochondria is achieved by the concerted action of translocation complexes located in both membranes of the organelle.

Mitochondrial Targeting Signals

Targeting signals direct mitochondrial precursor proteins to their particular destination in the organelle. In addition, these sequences can prevent folding of precursor proteins in the cytosol and thereby maintain them in a transport-competent conformation. Proteins of the mitochondrial matrix and many inner membrane proteins are made with N-terminal extensions which are proteolytically removed by proteases following translocation into the mitochondrial matrix. These matrix targeting signals (MTSs) are referred to as presequences. Most mitochondrial proteins carry such N-terminal presequences. In addition, less defined internal signals are used for mitochondrial targeting, typically by proteins of the outer membrane, the intermembrane space, and by some inner membrane proteins.

Presequences

Presequences generally comprise 15–80-amino-acid residues that have the potential to form amphipathic α -helices with one hydrophobic and one positively charged face. These signals are both necessary and sufficient for mitochondrial targeting. The addition of presequences to unrelated proteins typically leads to targeting of the fusion proteins to mitochondria. The MTSs direct the extended polypeptide in an

N-to-C-direction across both mitochondrial membranes. Examples were reported in which MTSs are located not at the N-terminus but at the C-terminus or in the interior of a preprotein. In the latter case, the internal signal forms an amphipathic hairpin loop which precedes import of the N- and C-terminal parts of the preprotein. In many cases, N-terminal presequences can be identified by computer algorithms which predict the intracellular location of proteins.

Internal Targeting Signals

Many mitochondrial proteins contain less defined targeting signals which are a part of the mature polypeptide. These are referred to as internal signals. For example, proteins of the solute carrier family of the inner membrane are made up of modules of pairs of transmembrane domains which are imported as loop-like structures. The regions flanking these transmembrane domains are critical for the uptake of these proteins by the mitochondria. The nature of these signals and how they are deciphered are, however, not well understood.

Proteins of the outer membrane are often anchored in the lipid bilayer by terminal transmembrane domains. A moderate degree of hydrophobicity and flanking positive charges specify these proteins for mitochondria and differentiate them from other tail-anchored proteins of the cell. Sorting of β -barrel proteins into the outer membrane by the TOB/SAM (topogenesis of mitochondrial outer-membrane β -barrel proteins/sorting and assembly machinery) complex is directed by a specific signal sequence close to their C-terminus.

Little is known about the targeting information that directs proteins into the intermembrane space of mitochondria. Many of these proteins are small and bind cofactors such as metal ions or heme. Several proteins of the intermembrane space contain structural disulfide bonds. The oxidation of cysteinyl thiols is mechanistically coupled to the import of these proteins. The amino acid sequences in proximity to these cysteine residues follow specific patterns and represent sorting signals to the intermembrane space. These signatures are referred to as mitochondrial intermembrane space sorting sequences (MISSs) or intermembrane space-targeting signals (ITSSs).

Transport to and across the Outer Membrane

A number of cytosolic factors were identified that bind to mitochondrial preproteins following their synthesis. Among these factors are members of the general chaperone system, such as Hsp70 and Hsp90, and components that appear to play a more specific role in the targeting of mitochondrial proteins. These factors keep preproteins in an import-competent conformation and accompany them to the surface of mitochondria.

The Translocase of the Outer Membrane

The outer membrane harbors surface receptors that expose binding sites for MTSS to the cytosol (see Table 1). These receptors are part of a multimeric protein translocase, called the translocase of the outer membrane (TOM) complex. In fungi, the TOM complex comprises the receptor subunits Tom70, Tom22, and Tom20. These surface receptors direct preproteins to the so-called 'general insertion pore' of the outer membrane, the main constituent of which is Tom40. Tom40 is thought to be present in a β -barrel structure in the outer membrane, similar to bacterial porin proteins, and by oligomerization forms protein-conducting channels. Three small additional subunits, Tom5, Tom6, and Tom7, appear to play a role as structural components and might be involved in regulating the interaction of the receptors with Tom40. In mammals, the TOM complex consists essentially of a similar set of homologous components. High-resolution electron microscopy of isolated TOM complexes revealed two membrane-traversing cavities of 2.2 nm, which most likely represent the preprotein-conducting channels in the outer membrane. Following translocation of preproteins through the general insertion pore, preproteins associate with the *trans* site of the TOM complex. This *trans*-binding site is formed by the intermembrane space domains of Tom40 and perhaps Tom22. An increasing binding affinity from *cis* to *trans* sites on the TOM complex was suggested to drive translocation of preproteins across the outer membrane into the intermembrane space.

Insertion into the Outer Membrane

In addition to its translocation function, the TOM complex facilitates the insertion of proteins into the lipid bilayer of the outer membrane (Figure 1, pathway 1). Although the molecular mechanism of this process is poorly understood, a critical role of the TOM complex was shown for membrane integration and assembly of TOM subunits, as well as for the insertion of other outer membrane proteins. The most abundant protein of the outer membrane, porin, and several tail-anchored proteins, such as the apoptosis regulator B-cell lymphoma 2 (bcl-2), require cytosolic receptor domains of TOM subunits for mitochondrial targeting. However, whereas integration of porin requires the general insertion pore in the TOM complex, bcl-2 was reported not to use the TOM channel.

β -Barrel proteins such as porin and the pore-forming constituent of the TOM complex, Tom40, are inserted into the outer membrane from the intermembrane space (Figure 1, pathway 2). Membrane insertion of these proteins is mediated by a specialized complex, named TOB or SAM complex. This insertion machinery is conserved from bacteria to mitochondria

Table 1 Components of the mitochondrial import machinery

Components		Function
<i>Fungi</i>	<i>Mammals</i>	
<i>Outer membrane, TOM complex</i>		
Tom70, Tom71	Tom70	Surface receptor
Tom40	Tom40	Pore formation
	Tom34	Surface receptor, Tom70 homolog
Tom22	Tom22	Surface receptor
Tom20	Tom20	Surface receptor
Tom7	Tom7	Structural or regulatory component
Tom6	Tom6	Structural component
Tom5	Tom5	Receptor function
<i>Outer membrane, TOB/SAM complex</i>		
Tob55/Sam50	Tob55/Sam50	Pore formation
Tob38/Sam35	Metaxin1/2	Structural or regulatory component
Mas37/Sam37	Metaxin1/2	Structural or regulatory component
Mdm10		Function unclear
Mim1		Function unclear
<i>Intermembrane space, transport from TOM to TIM22</i>		
Tim13	Tim13	Receptor and/or chaperone function
Tim12		Receptor and/or chaperone function
Tim10	Tim10 a, b	Receptor and/or chaperone function
Tim9	Tim9	Receptor and/or chaperone function
Tim8	DDP1, DDP2	Receptor and/or chaperone function
<i>Intermembrane space, protein disulfide array system</i>		
Mia40	Mia40	Mia40
		Receptor and oxidoreductase
Erv1	ALR (Erv1)	Sulfhydryl oxidase
Hot13	Hot13	Metal-binding factor
<i>Inner membrane</i>		
<i>TIM23 complex</i>		
Tim50	Tim50	Potentially receptor subunit
Tim44	Tim44	mtHsp70 binding
Tim23	Tim23	Receptor and channel activity
Tim21		Regulation of presequence binding
Tim17	Tim17	Function unclear
Tim16/Pam16	Tim16/Pam16	Regulation of import motor
Tim14/Pam18	Tim14/Pam18	Regulation of import motor, J-protein
Pam17		Function unclear
<i>TIM22 complex</i>		
Tim54		Function unclear
Tim22	Tim22	Channel activity
Tim18		Function unclear
<i>Further factors</i>		
Oxa1	Oxa1	Protein insertion from the matrix
Mba1		Protein insertion from the matrix
Imp1, Imp2	Imp1, Imp2	Sorting signal peptidases
<i>Matrix</i>		
Ssc1, Ecm10	mtHsp70	Chaperone, ATP-dependent precursor binding
Mge1	mtGrpE	Nucleotide exchange factor for mtHsp70
MPP	MPP	Processing of presequences
MIP	MIP	Processing of presequences
Mdj1	hTid (mtDnaJ)	Folding of imported proteins
Hsp60	Cpn60	Folding of imported proteins
Hsp10	Cpn10	Folding of imported proteins

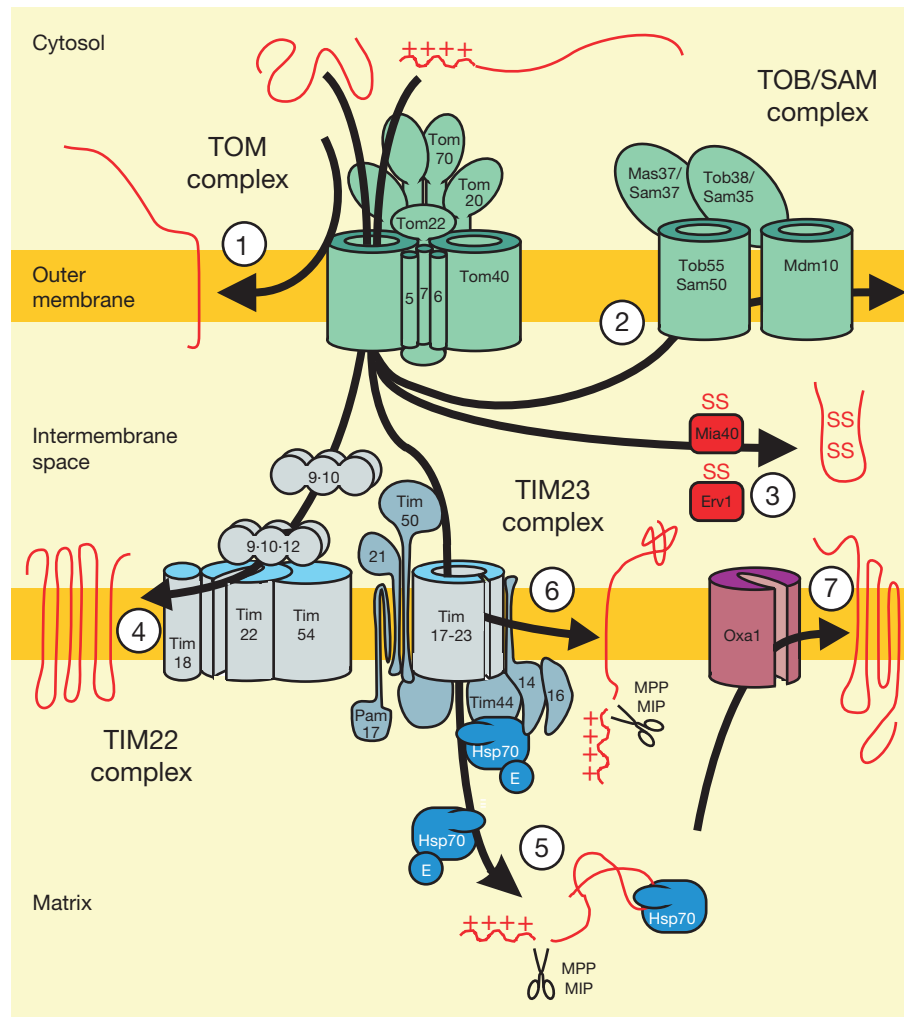


Figure 1 Protein import pathways in mitochondria. (1) Proteins of outer membrane that contain α -helical transmembrane domains are synthesized on cytosolic ribosomes and bind to the receptor subunits of the TOM complex before they are integrated into the outer membrane or further transported into the other mitochondrial subcompartments. (2) β -Barrel proteins of the outer membrane are initially transported through the TOM complex before they insert into the outer membrane in a process facilitated by the TOB/SAM complex. (3) Many proteins of the intermembrane space contain disulfide bonds. These are formed by interaction of these proteins with the oxidoreductase Mia40. Mia40 is maintained in a functional, oxidized state by the sulfhydryl oxidase Erv1. (4) Members of the solute carrier family and some Tim subunits traverse the outer membrane, forming loop-like structures. They are escorted to the TIM22 translocase in the inner membrane by complexes of small Tim protein complexes in the intermembrane space. (5) Matrix proteins contain presequences that mediate translocation through the TOM channel and interact with the TIM23 complex in the inner membrane. The presequences are transported across the inner membrane in a membrane potential-dependent manner, and removed by processing peptidases. Further translocation across the inner membrane requires the action of the ATP-dependent import motor. Matrix proteins then fold with the assistance of molecular chaperones. (6) Single-spanning inner membrane proteins can be arrested at the TIM23 translocase and laterally integrated into the lipid bilayer of the inner membrane. (7) Conservatively sorted proteins are first transported into the matrix and form sorting intermediates bound by mtHsp70. Subsequently, these proteins are inserted into the inner membrane in a membrane potential-dependent reaction. The inner membrane protein Oxa1 is required for membrane integration of several of these proteins.

and chloroplasts. Its central component, Tob55/Sam50, is also of β -barrel structure and homologous to the Omp85 protein that is essential for outer membrane biogenesis in prokaryotes.

Translocation into the IMS

The intermembrane space is a compartment of very small volume since the outer and inner membranes are in very close proximity. Only a relatively small number of components

of the intermembrane space are known, most of which are required for mitochondrial biogenesis or have a function in energy metabolism. A prominent example is cytochrome *c*, which plays a role in apoptosis apart from its function in shuttling electrons between complexes III and IV. Apocytochrome *c* is imported by the TOM complex and associates in the intermembrane space with the protein cytochrome *c* heme lyase (CCHL). CCHL represents a docking site for apocytochrome *c* and catalyzes its conversion to holocytochrome *c* by covalent addition of a heme group. Holocytochrome *c* is not able to

traverse the TOM channel and, thus, remains in the intermembrane space. Like cytochrome *c*, many intermembrane space proteins bind cofactors and may be imported by a similar mechanism.

Many proteins of the intermembrane space contain cysteine residues which form structural disulfide bonds. The import of these proteins relies on a specialized thiol oxidation machinery that consists of the oxidoreductase Mia40 and the sulfhydryl oxidase Erv1 (Figure 1, pathway 2). Mia40 serves as a receptor that traps incoming substrate proteins and facilitates their oxidation. Erv1 maintains Mia40 in the oxidized state. Both proteins are highly conserved among eukaryotes.

Another group of intermembrane space proteins is synthesized with N-terminal presequences followed by sorting signals. These signals are arrested at the level of the inner membrane. Proteases such as the Imp1/Imp2 protease at the outer face of the inner membrane then cleave after the sorting signals and the mature parts of the precursors are released into the intermembrane space. These sorting signals can be transplanted to unrelated passenger proteins to direct these into the intermembrane space. Examples of this type of proteins are cytochrome *b*₂ or cytochrome *b*₅ reductase.

A further group of soluble proteins of the intermembrane space is made without a cleavable and so-far-identified internal targeting signal. Examples are the cytochrome *c* and *c*₁ lyases, adenylate kinase and creatine kinase.

Translocation into the Matrix

The mitochondrial matrix contains a large number of proteins, many of which fulfill important functions in metabolic processes, energy production, or the replication and expression of the mitochondrial genome. These proteins in most contain N-terminal presequences and are imported on a common import route (Figure 1, pathway 5).

The Translocase of the Inner Membrane of Mitochondria 23 Complex

The translocase of the inner membrane of mitochondria 23 (TIM23) complex consists of Tim50, Tim23, Tim21, and Tim17, proteins which are integrated into the inner membrane exposing functional domains to the intermembrane space side and Tim44, Tim16/Pam16, Tim14/Pam18, and Pam17 which are associated with the former constituents at the matrix face. Tim23 and Tim17 share sequence similarity and are likely to form the translocation pore in the inner membrane. In yeast, the N-terminus of Tim23 protrudes into the outer membrane and thereby may mediate contact of the TOM and the TIM23 translocases. Tim50 exposes a large domain to the intermembrane space and plays a role in guiding preproteins from the TOM to the TIM23 complex. The transfer of preproteins from the TOM to the TIM23 complex was suggested to be further regulated by Tim21. Binding of the presequence to the TIM23 complex was proposed to lead to a membrane potential-dependent gating of the channel, which allows the presequence to traverse the inner membrane. This translocation reaction is driven by the membrane potential, which probably supports the movement of the positively

charged presequences to the negatively charged matrix face of the inner membrane. In the matrix, the presequences are bound by the chaperone mtHsp70, which prevents backsliding of the translocation intermediate. MtHsp70 is recruited to the TIM23 translocase by the Tim44 subunit of the complex which warrants a high activity of mtHsp70 at the exit of the TIM23 channel and, thus, the efficient trapping of the incoming polypeptide. The adenosine triphosphate (ATP)-dependent interaction of mtHsp70 with the preprotein is regulated by two DnaJ-like cofactors Tim16/Pam16 and Tim14/Pam18, and by the nucleotide exchange factor, mtGrpE/Mge1. This process is responsible for the ATP dependence of the mitochondrial import process. Binding of the preprotein to mtHsp70 is followed by the release of mtHsp70 from Tim44. Spontaneous unfolding of segments of the precursor on the outer face of the outer membrane, reversible sliding of the polypeptide chain in the TOM-TIM23 translocation channel, and repeated cycles of trapping by and release from mtHsp70 lead to vectorial movement of the preprotein into the matrix. This mechanism is referred to as Brownian ratchet model of mitochondrial protein import, as the essential role of mtHsp70 is to prevent backsliding of the preprotein. It was suggested that in addition to its trapping function, mtHsp70 undergoes an intramolecular conformational change and thereby mechanically pulls the bound preprotein into the matrix. The validity and significance of this motor model are still under debate.

Protein Folding and Assembly

In the matrix, the presequences are typically removed by processing peptidases. The mitochondrial processing peptidase (MPP) consists of two conserved subunits which convert most matrix proteins into their mature form. In addition, some preproteins are further processed by the mitochondrial intermediate peptidase, which removes another eight N-terminal amino acid residues. The mature proteins are then allowed to fold into their native structure. This process is facilitated by mtHsp70 which cooperates in this process with the DnaJ homolog Mdj1/hTid. Folding of some matrix proteins depends on the chaperonin complex formed by Hsp60 and Hsp10, which is homologous to the GroEL/GroES complex of the bacterial cytosol.

Protein Insertion into the Inner Membrane

The inner membrane contains a large number and diversity of proteins and belongs to the protein-richest membranes of eukaryotic cells. Inner membrane proteins reach their destination on several different insertion routes. At least three different pathways are known that lead precursor proteins into the inner membrane.

Insertion via the TIM23 Translocase

A number of monotopic inner membrane proteins are integrated into the inner membrane following translocation arrest at the level of the TIM23 complex (Figure 1, pathway 6). This sorting route was named the 'stop-transfer' pathway. Most of these proteins are of N_{in} – C_{out} topology, but some examples

were reported which are oriented in an $N_{out} - C_{in}$ topology by internal loop-forming presequences.

Insertion from the Matrix

Alternatively, proteins can be inserted into the inner membrane following complete translocation into the matrix (**Figure 1**, pathway 7). This transport pathway was denoted conservative sorting, since the process employs a bacterial-like insertion reaction from the matrix. The examples for this sorting path are inner membrane proteins of prokaryotic origin. Not only the direction of the membrane integration reaction resembles that in prokaryotes, but also the topogenic signals and the components that catalyze the insertion process. Conservatively sorted proteins adhere to the positive-inside rule according to which negative charges flanking the transmembrane segments are preferentially translocated across a membrane. A central player in the insertion process of these proteins is the inner membrane protein Oxa1 which is conserved from bacteria to mitochondria. Besides its role in the membrane integration of conservatively sorted proteins, the Oxa1 translocase mediates the insertion of mitochondrial translation products into the inner membrane.

Insertion via the TIM22 Translocase

A third group of inner membrane proteins is inserted by an alternative TIM complex, the TIM22 translocase (**Figure 1**, pathway 4). This pathway is used by members of the solute carrier family, a prominent member of which is the adenosine diphosphate (ADP)/ATP carrier, and by some TIM subunits. These proteins do not contain presequences but internal signals. Upon translocation through the TOM complex, these proteins associate with the hexameric Tim9·Tim10 and/or

Tim8·Tim13 complexes of the intermembrane space which belong to the family of 'small Tim proteins'. These complexes escort the hydrophobic precursor proteins from the TOM complex through the intermembrane space to the TIM22 complex. The TIM22 complex comprises the membrane-embedded subunits Tim54, Tim22, and Tim18 to which a Tim9·Tim10·Tim12 hexamer is associated. Tim22 is related to Tim17 and Tim23, and forms the translocation pore of the complex. Mutations in the human Tim8 homolog, DDP1, lead to the deafness dystonia (Mohr-Tranebjaerg) syndrome, a progressive neurodegenerative human disorder.

See also: **Bioenergetics:** Mitochondrial DNA; Mitochondrial Dynamics; Mitochondrial Genes and Their Expression: Yeast; Mitochondrial Membranes, Structural Organization; Mitochondrial Metabolite Carrier Protein Family; Mitochondrial Outer Membrane and the VDAC Channel; Nuclear Genes Involved in Mitochondrial Function and Biogenesis; **Protein/Enzyme Structure Function and Degradation:** Protein Folding and Assembly.

Further Reading

- Chacinska A, Koehler CM, Milenkovic D, Lithgow T, and Pfanner N (2009) Importing mitochondrial proteins: Machineries and mechanisms. *Cell* 138: 628–644.
- Koehler CM (2004) New developments in mitochondrial assembly. *Annual Review of Cell and Developmental Biology* 20: 309–335.
- Kuhn A, Stuart R, Henry R, and Dalbey RE (2003) The Alb3/Oxa1/YidC protein family: Membrane-localized chaperones facilitating membrane protein insertion? *Trends in Cell Biology* 13: 510–516.
- Neupert W and Herrmann JM (2007) Translocation of proteins into mitochondria. *Annual Review of Biochemistry* 76: 723–749.
- Rehling P, Brandner K, and Pfanner N (2004) Mitochondrial import and the twin-pore translocase. *Nature Reviews Molecular Cell Biology* 5: 519–530.
- Riemer J, Bulleid N, and Herrmann JM (2009) Disulfide formation in the ER and mitochondria: Two solutions to a common process. *Science* 324: 1284–1287.

Protein Kinase C Family

A C Newton, University of California, San Diego, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Diacylglycerol The membrane-retained lipid backbone released from phospholipids following activation of appropriate phospholipases, enzymes that hydrolyze phospholipids. Diacylglycerol is considered a second messenger because it transfers information from stimuli such as hormones to protein kinase C, which transduces the signal by phosphorylating protein substrates.

Kinase The class of enzymes that covalently transfer phosphate from adenosine triphosphate (ATP) to hydroxyl groups of proteins.

Phosphorylation The covalent attachment of phosphate from the cellular energy currency, ATP to hydroxyl residues of proteins, a modification that changes the properties of the protein.

Protein kinase C (PKC) is a family of enzymes that has a central role in transducing information from external stimuli to cellular responses. Members of this family of serine/threonine kinases respond to signals that cause lipid hydrolysis. PKC isozymes phosphorylate an abundance of substrates, leading to both short-term cellular responses such as regulation of membrane transport and long-term responses such as memory and learning.

Historical Perspective

PKC was discovered in the late 1970s by Yasutomi Nishizuka and colleagues at Kobe University, Japan. Their initial discovery was of a constitutively active enzyme that required only Mg^{2+} for activity (and hence was named protein kinase M, PKM). Further studies revealed that PKM was a proteolytic product of a full-length enzyme; this enzyme was named PKC because its enzymatic activity could be released by a Ca^{2+} -dependent protease. The subsequent discovery that PKC is activated by the phospholipid hydrolysis product, diacylglycerol, was a major finding in biology: it provided the molecular mechanism for how lipid hydrolysis, discovered 25 years earlier to be triggered by stimuli such as acetylcholine, couples to cellular signaling pathways.

But the discovery that catapulted research on PKC to the forefront of cellular signaling was the finding that it is the receptor for the potent tumor-promoting phorbol esters. Phorbol esters are present in the milky sap exuded from plants of the Euphorbiaceae family; the oil from the seeds of one member of this family, in particular *croton tiglium*, has particularly strong irritant properties and, as such, has been used over the millennia for purposes as varied as poison for hunting arrows to medicinal purposes. In the 1960s, the active ingredient in the oil was found to be a family of diesters of the tetracyclic diterpene phorbol. Phorbol esters were shown to be extremely potent tumor promoters, and classic studies revealed that painting phorbol esters on the skin of mice allowed otherwise subthreshold amounts of carcinogens to promote tumors. The finding that PKC is the direct molecular target of phorbol esters placed this enzyme at the center of signaling pathways that control normal cell function and carcinogenesis.

PKC Family Members

There are 10 mammalian isozymes of PKC that share in common a carboxyl-terminal kinase domain linked to an amino-terminal regulatory moiety (Figure 1). The regulatory moiety, in turn, contains a number of functional modules and it is the composition of these functional modules that further defines the three subfamilies of the PKC isozymes. These modules are an autoinhibitory-pseudosubstrate sequence that maintains the enzyme in an inactive conformation, and one or two membrane-targeting modules that direct PKC to the membrane following generation of the appropriate second messengers. Specifically, the C1 domain binds diacylglycerol and phorbol esters and the C2 domain binds Ca^{2+} ; each event promotes the binding of the respective domain to membranes.

Conventional PKC isozymes (α , βI , βII , and γ), have a C1 and a C2 domain and respond to both diacylglycerol and Ca^{2+} . Novel PKC isozymes (δ , ϵ , η , and θ) have a C1 domain that binds diacylglycerol, but an impaired C2 domain that does not bind Ca^{2+} . These isozymes respond to increases in cellular diacylglycerol but not Ca^{2+} . Note that a change from a Tyr to Trp in the C1b domain of novel isozymes, compared to conventional isozymes, confers an order of magnitude higher affinity for diacylglycerol, allowing these isozymes to respond to signals that elevate only intracellular diacylglycerol. Atypical PKC isozymes (ζ and ι/λ) have an impaired C1 domain and no C2 domain and bind neither diacylglycerol nor Ca^{2+} . Rather, they have a PB1 domain that mediates protein:protein interactions. Thus, stimuli that elevate intracellular diacylglycerol activate conventional and novel PKC family members, with conventional isozymes being additionally regulated by Ca^{2+} . Atypical isozymes are regulated by other mechanisms such as phosphorylation and protein interactions.

PKC Phosphorylation

Before PKC is competent to signal, it must first be processed by a series of ordered phosphorylations. These phosphorylations depend on two kinases: the phosphoinositide-dependent

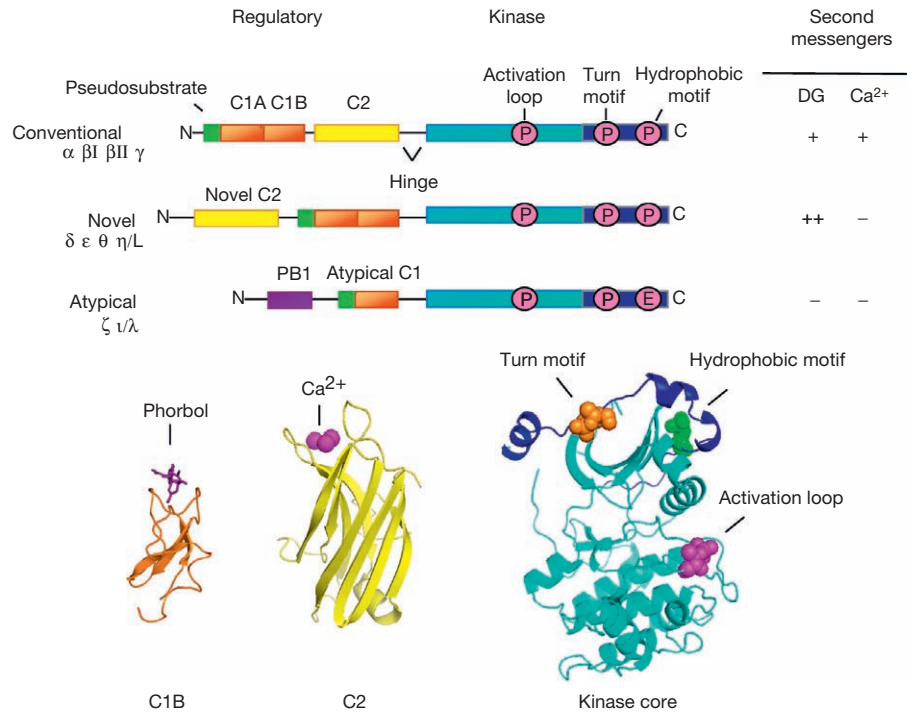


Figure 1 Primary structure and domain composition of protein kinase C family members. The amino-terminal regulatory moiety contains the autoinhibitory-pseudosubstrate sequence (green), the C1 domain, which binds diacylglycerol/phorbol esters (orange; present as a tandem repeat in conventional and novel protein kinase C isozymes), and the C2 domain, which binds Ca²⁺ (yellow). The C2 domain in novel protein kinase Cs and the C1 domain in atypical protein kinase Cs are non-ligand-binding variants. Atypical isozymes contain a PB1 domain that mediates protein:protein interactions (purple). The carboxyl-terminal catalytic moiety contains the kinase core which has three phosphorylation sites, one on the activation loop segment (pink circle) and two on the carboxyl-terminal tail (blue): the turn motif (orange circle) and the hydrophobic motif (green circle; the negatively charged amino acid glutamate (E) occupies the phospho-acceptor position of the hydrophobic motif in atypical protein kinase Cs). The second messenger sensitivity is shown on the right of each isozyme subgroup; note that novel isozymes have a 10-fold increased affinity for diacylglycerol compared to conventional protein kinase C isozymes, allowing them to respond to signals that elevate only diacylglycerol and not both diacylglycerol and Ca²⁺. The 3D structures of the domains are shown below the primary structure.

kinase, PDK-1, and the mammalian target of rapamycin (mTORC2) complex, which comprises the kinase mTOR, rictor, and a number of other interacting partners. These kinases have pivotal positions in cell signaling: PDK-1 catalyzes the activating phosphorylation to many other PKs, including the prosurvival kinase, Akt/PKB and mTORC2 plays an essential role in allowing Akt/PKB to become fully phosphorylated. Both kinases are necessary (but neither is sufficient) for the processing of PKC by phosphorylation, and lack of either one results in unphosphorylated PKC that is unstable and becomes degraded. Although the precise mechanism of phosphorylation is still being resolved, it appears that the first phosphorylation is catalyzed by PDK-1 on a conserved segment near the entrance to the active site referred to as the activation loop (Figure 1), an event that structures the active site for substrate binding and catalysis. The phosphorylation of the activation loop by PDK-1 triggers two tightly coupled phosphorylations at two conserved positions in the carboxyl terminus, the turn motif and hydrophobic motif (Figure 1). These latter phosphorylation events depend on mTORC2, but whether by direct phosphorylation or because mTORC2 structures or positions PKC for C-terminal phosphorylations (which are mediated by intramolecular autophosphorylation *in vitro*) remains to be resolved. These phosphorylations lock PKC in its mature and

catalytically competent conformation. It is this species of PKC that is activated by lipid hydrolysis and transduces signals.

PKC Translocation

Mature (i.e., phosphorylated) PKC is typically localized to the cytosol where it bounces on and off the membrane by diffusion-controlled mechanisms (although it is often scaffolded at specific locations, thus, increasing the local concentration for more effective signaling; see below). It is maintained in an inactive conformation because the pseudosubstrate sequence occupies the substrate-binding cavity. For conventional PKC isozymes, generation of Ca²⁺ and diacylglycerol target PKC to the membrane by binding the C2 and C1 domains and, thus, tethering the enzyme to membranes. The membrane-bound species adopts an active conformation by removal of the pseudosubstrate from the substrate-binding cavity, allowing substrate binding, phosphorylation, and downstream signaling. Finding a membrane-embedded ligand (diacylglycerol) by diffusion from the cytosol is a low-probability event, and, in the case of conventional PKCs, nature has chosen a clever mechanism to increase the efficiency of this. Binding of Ca²⁺ to the C2 domain essentially pretargets conventional PKC to the

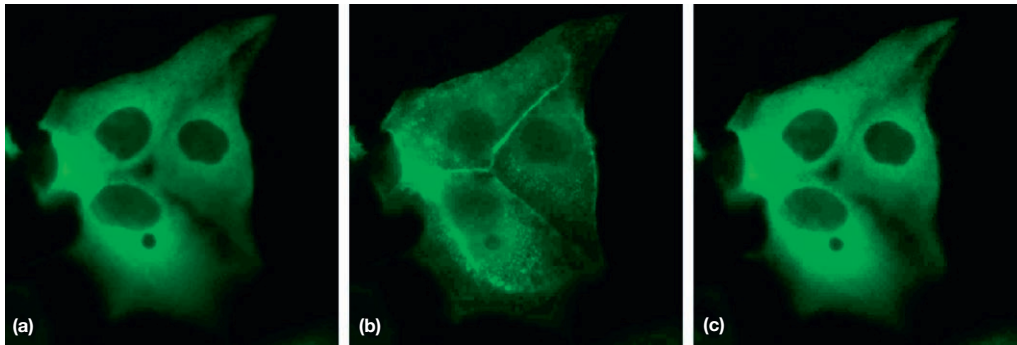


Figure 2 Protein kinase C was visualized in cells by expression of a construct of protein kinase C fused to a naturally-fluorescent protein from the jellyfish *Aequorea victoria*, the green fluorescent protein (GFP). (a) Shows that protein kinase C is localized to the cytosol in unstimulated MDCK cells (diffuse fluorescence throughout the cell); (b) shows that protein kinase C translocates to the membrane (strong fluorescence intensity at cell periphery) following 1 min of treatment with the agonist UTP, which induces phospholipid hydrolysis and generation of the two second messengers for protein kinase C: Ca^{2+} and diacylglycerol. (c) Shows that protein kinase C has redistributed back to the cytosol 5 min after UTP treatment; second messenger levels have returned to resting levels. Images courtesy of Jon Violin.

membrane (primarily plasma membrane, where the C2 domain binds phosphatidylinositol-bis-phosphate (PIP_2)), where it can now initiate a much more effective search for its membrane-embedded ligand, diacylglycerol. As a consequence, conventional PKCs translocate to membranes approximately one order of magnitude faster than novel PKC isozymes, which do not have the advantage of pretargeting by a Ca^{2+} -responsive C2 domain. Nonetheless, novel isozymes are able to respond to diacylglycerol because their C1b domain is tuned to be an order-of-magnitude more responsive to diacylglycerol; these isozymes preferentially bind Golgi, which is unusually rich in diacylglycerol.

The advent of fluorescent methodologies has allowed imaging of PKC translocation and activity in real time in living cells (Figure 2). Notably, targeted reporters have allowed the visualization of kinase translocation, activity, and second messenger levels at precise intracellular locations, from specific membranes to protein scaffolds. In general, PKC translocation and activity mirror the generation of its second messengers. For example, histamine stimulation of HeLa cells results in oscillations in PKC substrate phosphorylation that are phase-locked with Ca^{2+} oscillations. Conventional PKC isozymes translocate rapidly to plasma membrane with kinetics mirroring Ca^{2+} elevation; diacylglycerol retains these isozymes at the plasma membrane and their activity here decays in concert with diacylglycerol. Novel isozymes translocate more slowly to Golgi membranes, where their activity is sustained because of sustained diacylglycerol at the Golgi.

PKC Scaffolds

Correct subcellular location is essential for normal signaling by PKC. An abundance of scaffold proteins that tether PKC near its substrates, activators, and regulatory proteins, such as phosphatases, has been described, most notably receptors for activated C kinases. The importance of correct subcellular location is perhaps best illustrated in the *Drosophila* visual cascade, where mutants lacking the PKC-binding scaffold, InaD, are defective in visual transduction because components of the signaling cascade are mislocalized.

PKC Downregulation

Prolonged activation of PKC by treatment of cells with phorbol esters results in degradation of PKC, a phenomenon referred to as downregulation. In fact, prolonged treatment of cells with phorbol esters is a commonly used approach to deplete cells of all except atypical PKCs (these are resistant to phorbol ester-dependent downregulation because they do not bind phorbol esters). The first step in downregulation is the dephosphorylation of the hydrophobic motif site by the recently discovered phosphatase PH-domain leucine-rich repeat protein phosphatase, an event that shunts PKC to a detergent-insoluble fraction where it is further dephosphorylated, ubiquitinated, and degraded via a proteosomal pathway. The molecular chaperone HSP70 protects PKC from downregulation by allowing rephosphorylation of the enzyme and sustaining its signaling lifetime.

PKC Signaling

PKC phosphorylates an abundance of substrates, including membrane proteins, cytoskeletal proteins, cytosolic proteins, and nuclear proteins. It plays key roles in diverse biological processes, many involving cell-survival pathways. Indeed, its role in cancer is becoming increasingly apparent. Not only are levels of specific isozymes frequently altered in cancer, but also driver mutations in several PKC isozymes have recently been identified in human tumors, underscoring the key role these enzymes play in maintaining cellular homeostasis.

Genetic deletion of specific isozymes results in subtle phenotypic differences, suggesting functional redundancy of the isozymes. It is interesting to note, however, that animals deficient in PKC isozymes are frequently deficient in adaptive responses. For example, mice lacking PKC ϵ have reduced anxiety and have reduced tolerance to alcohol. Mice lacking PKC γ have reduced pain perception and mice lacking PKC β II have reduced learning abilities and an impaired immune response. This theme carries over to the molecular level where many of the substrates of PKC are receptors which become desensitized following phosphorylation by PKC.

Isozyme-specific functions have been most clearly delineated for novel and atypical PKC isozymes. For example, PKC δ activation has been shown to play a role in apoptosis. PKC θ plays a key role in immune responses, and mice deficient in this isozyme have impaired T-cell signaling and interleukin 2 production. Defined functions have also been established for PKC ζ : this isozyme is required for maintenance of cell polarity and, in addition, regulates cell growth, DNA synthesis, and activation of the transcription factor, nuclear factor-kappa B (NFkB).

The amplitude of the PKC signal depends on (1) the cellular levels of PKC, which are controlled by phosphorylation and dephosphorylation, (2) the acute activity of PKC, which is controlled by second messengers, and (3) the subcellular localization of PKC, which is controlled not only by membrane interactions, but also by protein scaffolds. Each mechanism is precisely controlled. Dysregulation at any step results in pathophysiological states, notably cancer, where it is becoming

increasingly apparent that this family of isozymes plays key roles in controlling cell growth and survival.

See also: **Signaling:** Calcium/Calmodulin-Dependent Protein Kinases; Glycine Receptors; Natriuretic Peptides, Their Receptors and Therapeutic Applications; Neurotransmitter Transporters; Phospholipase C; Phospholipase D.

Further Reading

- Cameron AJ, De Rycker M, Calleja V, et al. (2007) Protein kinases, from B to C. *Biochemical Society Transactions* 35: 1013–1017.
- Gallegos LL and Newton AC (2008) Spatiotemporal dynamics of lipid signaling: Protein kinase C as a paradigm. *IUBMB Life* 60: 782–789.
- Griner EM and Kazanietz MG (2007) Protein kinase C and other diacylglycerol effectors in cancer. *Nature Reviews Cancer* 7: 281–294.
- Newton AC (2010) Protein kinase C: Poised to signal. *American Journal of Physiology* 298: E395–402.

Protein N-Myristoylation

E D Sonnenburg and J I Gordon, Washington University School of Medicine, St. Louis, MO, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 527–531, © 2004, Elsevier Inc.

Glossary

Heterogeneous acylation Nmt-catalyzed attachment of the following acyl chains having somewhat lower hydrophobicity than C14:0: tetradecenoate (C14:1^{Δ5}), tetradecadienoate (C14:2^{Δ5,8}), or laurate (C12:0). At present, this modification is only known to occur in retinal photoreceptor cells.

N-myristoylproteins This diverse group of covalently modified proteins use their acyl chain to promote a variety of readily reversible interactions with various cellular membranes or with other proteins. The stability of their interactions can be regulated by other modifications, including conformational changes that affect the presentation of the myristoyl moiety, posttranslational

acylation at other sites with other lipids, or charge–charge interactions mediated by protein side chain atoms.

N-myristoyltransferase (Nmt) An enzyme that catalyzes the linkage of myristate, via an amide bond, to the N-terminal glycine amine of cellular proteins. Catalysis occurs through the nucleophilic addition–elimination reaction. The acyl chain length specificity of the enzyme has been highly conserved during evolution, while its peptide substrate specificity has diverged among species. Nmt is a target for the development of drugs against fungal and other human pathogens.

Protein N-myristoylation The covalent attachment of myristate, a 14-carbon saturated fatty acid (C14:0), to the N-terminal glycine residue of eukaryotic proteins.

Protein N-myristoylation refers to the covalent attachment of myristic acid, a 14-carbon saturated fatty acid (C14:0), to the N-terminal glycine of proteins. Linkage occurs via an amide bond and takes place as proteins are being synthesized. N-myristoylproteins have varied intracellular destinations, and are involved in myriad cellular functions ranging from signal transduction to protein and vesicular trafficking. N-myristoylproteins are encountered in members of all kingdoms of the eukaryotic domain (Protist, Fungi, Plant, and Animal), but are not produced by members of Bacteria or Archaea. MyristoylCoA: protein N-myristoyltransferase (Nmt), E.C. 2.3.1.97, a member of the GCN5 acetyltransferase (GNAT) superfamily, is responsible for catalyzing the transfer of myristate from myristoylCoA to proteins. While the acylCoA substrate specificity of Nmt has been highly conserved during evolution, its peptide substrate specificities have diverged among eukaryotes.

Protein N-myristoylation has relevance to a number of diseases. Genetic and pharmacologic studies have shown that Nmt activity is essential for the survival of a number of fungi that cause systemic infections in immunocompromised humans. These organisms include *Candida albicans* and *Cryptococcus neoformans*. Protein N-myristoylation is also critical to the life cycle of viruses, such as human immunodeficiency virus-1. Nmt levels are elevated in human gastrointestinal tract malignancies. Other pathogens, including *Plasmodium falciparum*, the parasite that causes malaria, have NmTs. Therefore, understanding the contributions of myristate to protein function, elucidating the structural basis for the divergent protein substrate specificities of orthologous NmTs, characterizing the myristoyl transfer reaction, and identifying biologically active, species-selective Nmt inhibitors may provide new therapeutic strategies and agents.

N-Myristoylproteins

Function of N-Myristoylation

Myristate exposed on the surface of a protein increases its lipophilicity, allowing transient, low-affinity interactions with cellular membranes or other proteins. This feature makes N-myristoylation an attractive choice for molecules involved in a variety of signaling cascades, such as protein kinases, kinase substrates, and protein phosphatases.

Myristate is not sufficiently hydrophobic to allow stable anchorage to cellular membranes. Therefore, most N-myristoylproteins utilize an additional mode of attachment. One supplementary membrane tether is provided by the reversible posttranslational covalent attachment of palmitate, a 16-carbon saturated fatty acid (C16:0), to cysteine (S-palmitoylation). Alpha subunits of heterotrimeric G proteins provide examples of this type of dual acylation. Other N-myristoylproteins, such as the protein tyrosine kinase Src, use positively charged exposed residues to interact electrostatically with negatively charged groups in membrane lipids. Some N-myristoylproteins regulate their association with membranes through reversible exposure of their myristoyl moiety. Recoverin, a retinal photoreceptor protein involved in light adaptation, illustrates this ‘myristoyl-conformational switch’: its myristoyl group is unfurled when the protein binds calcium; in the absence of calcium, the acyl chain is encased in a hydrophobic pocket.

Heterogeneous acylation is another device used by N-myristoylproteins to achieve regulated membrane association. Heterogeneous acylation refers to Nmt-catalyzed linkage of tetradecanoate (C14:0), tetradecenoate (C14:1^{Δ5}), tetradecadienoate (C14:2^{Δ5,8}), or laurate (C12:0) to a substrate protein. At present, this phenomenon has only been described in

retinal photoreceptor cells. Augmented production of these acylCoAs, and/or increased access of Nmt to these species, is among the mechanisms that have been invoked to account for the observed cell-lineage-specificity of heterogeneous acylation.

Examples of heterogeneously acylated proteins include recoverin, the α -subunit of the G protein transducin (Gtz), guanylyl cyclase activating protein, and the catalytic subunit of cyclic adenosine monophosphate (cAMP)-dependent protein kinase. Lauric acid and C14 fatty acids, containing a *cis* double bond between C5–C6 (C14:1^{Δ5}) or C5–C6 and C8–C9 (C14:2^{Δ5,8}), have reduced hydrophobicity compared to C14:0. A less hydrophobic acyl chain may be required for these proteins to properly operate in photoreceptor cells, either because of their extraordinarily high concentrations of membranes, and/or because the proteins require short-lived anchorage to membranes to rapidly transduce responses in visual signaling pathways.

Substrate Selection

Most proteins targeted for cotranslational N-myristoylation have an N-terminal consensus sequence of Met–Gly–X–X–X–Ser/Thr. The initiator methionine must first be removed from the nascent protein by another enzyme, methionylaminopeptidase, to expose the acceptor glycine. However, this consensus sequence is too general to allow definitive identification of Nmt substrates. Furthermore, in at least one case involving the pro-apoptotic protein BID, an internal myristoylation site is exposed after posttranslational proteolytic cleavage. **Figure 1** provides a more detailed recognition consensus sequence that has been developed from an analysis of known N-myristoylproteins, as well as currently available X-ray crystal structures of Nmt.

There are a number of ways to establish that a protein is N-myristoylated, including metabolic labeling with tritiated or iodinated myristate, mass spectrometric analysis of the purified protein, or *in vitro* Nmt assays that contain the purified acyltransferase, myristoylCoA, and a peptide encompassing the N-terminal 8–15 residues of the candidate protein substrate. A current listing of experimentally confirmed as well as candidate N-myristoylproteins can be found at the website address listed in the Relevant Website section.

Escherichia coli containing a dual plasmid expression system has also been used widely to determine whether a protein is an Nmt substrate, and to study the functional significance of its myristoyl group. The system takes advantage of the fact that this bacterium is able to generate myristoylCoA but does not

have any endogenous Nmt activity. The eukaryotic protein modification is recapitulated in this prokaryote by using one plasmid with an inducible promoter to direct expression of an Nmt, and another plasmid, containing a different inducible promoter, to express a candidate Nmt substrate. Production of Nmt is initiated first. Synthesis of the protein substrate is then turned on to allow for its cotranslational acylation. Incorporation of myristate can be verified by metabolic labeling or by mass spectrometry. The properties of the N-myristoylated protein can be compared and contrasted to that of the nonmyristoylated isoform produced in *E. coli* lacking the Nmt expression vector.

N-Myristoyltransferase

Thus far, a total of 68 known and postulated Nmts have been identified from 63 species. Four mammalian genomes (*Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, and *Bos Taurus*), plus a plant genome (the mustard weed, *Arabidopsis thaliana*), contain two Nmt genes (designated type I and type II). Type I Nmts show a high degree of conservation among one another, as do the type II acyltransferases. Type I and II enzymes diverge most prominently at their N termini. These N-terminal domains are not required for catalytic activity *in vitro* but may help direct the acyltransferase to different intracellular sites. For example, the N terminus of type I human Nmt appears to be involved in its targeting to the ribosome.

Reaction Mechanism

The N-myristoyltransferase reaction mechanism has been characterized using the *Saccharomyces cerevisiae* enzyme as a model. *S. cerevisiae* Nmt contains 455 amino acids, is monomeric, and has no known cofactor requirements. Steady-state kinetics, isothermal titration calorimetrics, and X-ray crystallographic studies indicate that the enzyme has an ordered Bi–Bi reaction mechanism. The apo-Nmt first binds myristoylCoA to form a Nmt:myristoylCoA binary complex. MyristoylCoA binding allows subsequent binding of a nascent protein substrate to generate a Nmt:myristoylCoA:peptide ternary complex. Following the chemical transformation step (conversion of the enzyme:substrate complex to the enzyme:product complex), CoA and then myristoylpeptide are released. Pre-steady-state kinetic studies indicate that the rate-determining step occurs after the chemical transformation, most likely involving release of myristoylpeptide.

Catalysis occurs through a classic nucleophilic addition–elimination reaction where the nucleophilic N-terminal Gly



Figure 1 Consensus sequence for recognition of protein substrates by N-myristoyltransferase (Nmt) (*side chain volume compensation allowed; fungal Nmts allow hydrophobic residues; § some fungal Nmts prefer small residues; and || often S or T). Figure based on Maurer-Stroh S, Eisenhaber B, and Eisenhaber F (2002) N-terminal N-myristoylation of proteins: Refinement of the sequence motif and its taxon-specific differences. *Journal of Molecular Biology* 317: 523–540, 541–557.

amine of the peptide substrate attacks the polarized thioester carbonyl of myristoylCoA. Several elements in the enzyme's active site facilitate this reaction. An oxyanion hole, formed by the backbone amides of two conserved residues, Phe170 and Leu171, polarizes the reactive carbonyl of myristoylCoA. The C-terminal carboxylate (Leu455 in the yeast enzyme) functions as a catalytic base to deprotonate the Gly1 ammonium to a nucleophilic amine. The amine rotates 180° along Ψ to reduce its distance from the thioester carbonyl of myristoylCoA. H-bonding interactions between the Gly1 amine, Asn169, and Thr205 help direct the reaction trajectory, and facilitate nucleophilic attack of the polarized carbonyl. The oxyanion hole, together with the H-bonding network formed by Asn169 and Thr205, stabilizes the developing tetrahedral intermediate. With subsequent collapse of the intermediate, CoA is extruded. The Gly1 amine is deprotonated, while the thiolate-leaving group of CoA is reprotonated.

Three-Dimensional Structure of Nmt: Insights about Substrate Recognition/Acquisition and Release

Presently, three X-ray crystal structures of *S. cerevisiae* Nmt have been reported: a binary complex containing bound myristoylCoA (Protein Data Bank (PDF) accession 1IIC), plus two ternary complexes – one with a peptide substrate and S-(2-oxo)pentadecylCoA (a nonhydrolyzable myristoylCoA analog containing a methylene interposed between the sulfur of CoA and the carbonyl of myristate; PDF 1IID) and the other with S-(2-oxo)pentadecylCoA and a dipeptide inhibitor (PDB, 2NMT). In addition, three structures of *C. albicans* Nmt are available: the apo-enzyme (PDB, 1NMT), the enzyme with bound myristoylCoA and peptidic inhibitor (PDB, 1IYK), and the enzyme with a bound benzofuran inhibitor (PDB, 1IYL).

The Nmt fold consists of a large saddle-shaped β -sheet with several α -helices located on either side (Figure 2). There is pseudo-twofold symmetry. The N-terminal half forms the myristoylCoA-binding site, while the C-terminal half contributes to peptide recognition. The N terminus is disordered.

Binding of myristoylCoA induces two important conformational changes. A 3_{10} helix (A' in Figure 2) is formed from part of the disordered N terminus, to complete the myristoylCoA binding site. There is also a change in the conformation of a loop connecting helix A and strand b (Ab loop), thereby opening a lid overlying the peptide binding site. The pantetheine group of CoA forms a key component of the peptide recognition site. Together, these observations indicate why myristoylCoA binding is a prerequisite for subsequent acquisition of a nascent protein substrate.

MyristoylCoA is bound to Nmt in a conformation that resembles a question mark (Figure 2). Nmt uses residues that produce bends at C1 and C6 of the bound myristoyl chain, as well as the floor of its acylCoA-binding pocket, to measure and properly position myristoylCoA. Available X-ray structures indicate that the two additional methylenes present in palmitoylCoA cannot be accommodated by the enzyme without adversely affecting positioning of its thioester carbonyl within the oxyanion hole.

The X-ray structures suggest that the extent of ordering of the Ab loop may correlate with the overall catalytic efficiency of different Nmt peptide substrates, and that the rate-limiting step

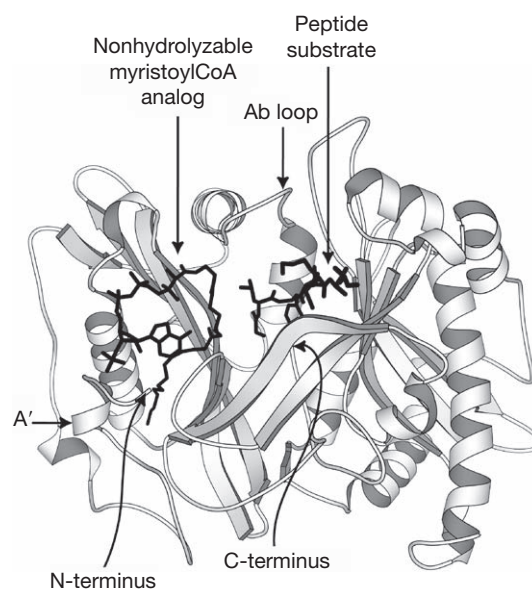


Figure 2 Ribbon diagram of *S. cerevisiae* Nmt. The enzyme is shown with a S-(2-oxo)pentadecylCoA, a nonhydrolyzable myristoylCoA analog and competitive inhibitor, and an octapeptide substrate (GlyLeuTyrAlaSerLysLeuAla) derived from the N terminus of ADP ribosylation factor 2 (Arf2). Binding of the myristoylCoA induces ordering of a 3_{10} helix (A'), a change in the conformation of the Ab loop, and completion of the peptide binding site. The C-terminal carboxylate (Leu455) functions as a catalytic base to deprotonate the Gly1 ammonium to a nucleophilic amine.

in the Nmt reaction is an isomerization involving reversal of the conformational changes that take place when myristoylCoA is bound. In other words, it indicates disordering of the 3_{10} A' helix and opening the lid formed by the Ab loop.

Therapeutics

Alignments of known NmTs disclose that their N-terminal halves are more conserved than their C-terminal halves. This finding is consistent with their shared specificity for myristoylCoA, and with the divergent nature of their protein substrate specificities. Exploiting the differences in peptide recognition among NmTs may be key to developing species-specific inhibitors. This divergence, and the fact that genetic experiments have established that pathogenic fungi such as *C. albicans* and *C. neoformans* require Nmt for their survival, makes N-myristoyltransferase an attractive target for developing new classes of fungicidal agents. The need for such agents is great, given the limited number of existing drugs that effectively kill these organisms, and the increasing resistance being encountered to these available compounds.

Several classes of Nmt-targeted antifungal agents have been identified to date. Peptidomimetic inhibitors have been produced starting from an octapeptide substrate representing the N terminus of *S. cerevisiae* adenosine diphosphate (ADP) ribosylation factor 2 (Arf2) (involved in vesicular trafficking). Benzofuran inhibitors have emerged from high-throughput screens of chemical libraries. These benzofurans have several advantages over the Arf-derived compounds – smaller size,

increased bioavailability, and resistance to degradation. Crystallographic studies of *C. albicans* Nmt revealed that one of the lead benzofurans was positioned in a hydrophobic pocket located within the enzyme's peptide substrate-binding site. Using this structural information, a more biologically active derivative, substituted with difluorophenoxy- and pyridyl-terminated side chains, was created that inhibits fungal growth in a rat model of systemic Candidiasis. Further screens for compounds that display specificity for fungal compared to human NmTs are likely to yield additional classes of fungicidal compounds in the future.

See also: **Protein/Enzyme Structure Function and Degradation:** Lipid Modification of Proteins: Targeting to Membranes; **Signaling:** Src Family of Protein Tyrosine Kinases.

Further Reading

- Ames JB, Ishima R, Tanaka T, et al. (1997) Molecular mechanics of calcium-myristoyl switches. *Nature* 389: 198–202.
- Devadas B, Freeman SK, McWherter CA, et al. (1998) Novel biologically active nonpeptidic inhibitors of myristoylCoA: Protein N-myristoyltransferase. *Journal of Medicinal Chemistry* 41: 996–1000.
- Ebiike H, Masubuchi M, Liu P, et al. (2002) Design and synthesis of novel benzofurans as a new class of antifungal agents targeting fungal N-myristoyltransferase. *Bioorganic and Medicinal Chemistry Letters* 12: 607–610.
- Farazi TA, Manchester JK, Waksman G, and Gordon JI (2001) Pre-steady state kinetic studies of *Saccharomyces cerevisiae* myristoylCoA: Protein N-myristoyltransferase reveal that a step after chemical transformation is rate limiting. *Biochemistry* 40: 9177–9186.
- Georgopapadakou NH (2002) Antifungals targeted to protein modification: Focus on protein N-myristoyltransferase. *Expert Opinion on Investigational Drugs* 11: 1117–1125.
- Johnson RS, Ohguro H, Palczewski K, et al. (1994) Heterogeneous N-acylation is a tissue- and species-specific post-translational modification. *Journal of Biological Chemistry* 269: 21067–21071.
- Lodge JK, Jackson-Machelski E, Toffaletti DL, Perfect JR, and Gordon JI (1994) Targeted gene replacement demonstrates that myristoylCoA protein N-myristoyltransferase is essential for the viability of *Cryptococcus neoformans*. *Proceedings of the National Academy of Sciences of the United States of America* 91: 12008–12012.
- Maurer-Stroh S, Eisenhaber B, and Eisenhaber F (2002) N-terminal N-myristoylation of proteins: prediction of substrate proteins from amino acid sequence. *Journal of Molecular Biology* 317: 541–547.
- Maurer-Stroh S, Eisenhaber B, and Eisenhaber F (2002) N-terminal N-myristoylation of proteins: Refinement of the sequence motif and its taxon-specific differences. *Journal of Molecular Biology* 317: 523–540.
- Weinberg RA, McWherter CA, Freeman SK, et al. (1995) Genetic studies reveal that myristoylCoA protein N-myristoyltransferase is an essential enzyme in *Candida albicans*. *Molecular Microbiology* 16: 241–250.
- Weston SA, Camble R, Colls J, et al. (1998) Crystal structure of the antifungal target N-myristoyltransferase. *Nature Structural and Molecular Biology* 5: 213–221.

Relevant Websites

- <http://www.jbc.org> – *Journal of Biological Chemistry*, The Biology and Enzymology of Protein N-Myristoylation, Data Supplement for Journal of Biological Chemistry: Volume 276, Issue 43, Page 39501.
- <http://www.rcsb.org> – Protein Data Bank: An information portal to biological macromolecular structures.
- <http://mendel.imp.univie.ac.at> – Research Institute of Molecular Pathology, supplementary material to N-terminal N-myristoylation of proteins.

Protein Palmitoylation

R J Deschenes, College of Medicine, Tampa, FL, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 3, pp. 532–535, © 2004, Elsevier Inc.

Glossary

DHHC–CRD motif A protein sequence motif that is associated with the class of palmitoyltransferase that carries out the S-palmitoylation of proteins.

Palmitoylation The process by which proteins become covalently modified by long-chain fatty acids related to palmitic acid.

Palmitoylthioesterase An enzyme that cleaves palmitate from S-palmitoylated proteins.

Protein acyltransferase (PAT) An enzyme that carries out the palmitoylation of proteins.

Palmitoylation is the covalent modification of proteins with 16 carbon acyl chains related to palmitic acid. Protein palmitoylation was first described in the 1970s and is now recognized as a major form of posttranslational modification in all eukaryotes, with a few examples in prokaryotes as well. In many cases, palmitoylation is a reversible modification, suggesting that the cell uses it to regulate protein function. This article focuses on the enzymology and functional significance of protein palmitoylation.

Covalent Modification of Proteins with Palmitoyl Groups

Palmitoylation refers to a family of protein modifications involving palmitic acid (Figure 1). The most common forms of palmitoylation are S-palmitoylation and N-palmitoylation. S-palmitoylation occurs on cysteine residues and involves a thioester linkage. In the cases where it has been investigated, S-palmitoylation is a reversible modification. Examples include receptors, signaling proteins, and scaffold proteins. N-palmitoylation reactions occur on the amino terminus or on the epsilon amino group of lysine. Reports of esterification of palmitate with serine and threonine residues (O-palmitoylation) have appeared for myelin proteolipids; however, other studies conclude that myelin proteolipids are S-palmitoylated. Protein palmitoylation is generally detected by labeling cells or cell extracts with [³H]-palmitate. Identification of the covalently linked lipid is done by chemical or enzymatic cleavage followed by high-performance liquid chromatography (HPLC) analysis. In most cases, palmitate is found, but in some cases, unsaturated forms of palmitate are also detected. An estimate of the total extent of protein palmitoylation in the cell is difficult to obtain because of the rapid reversibility of palmitoylation and the chemical lability of thioesters. Proteomic methods that account for posttranslational modifications have begun, but application to lipid modifications needs further development. However, based on the growing list of palmitoylated proteins reported, it is clear that palmitoylation is a significant posttranslational modification.

Palmitoyltransferases

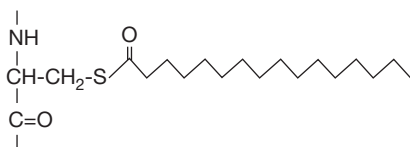
There has been a debate concerning whether palmitoylation is spontaneous or enzyme mediated. The thiol of cysteine is capable of serving as a nucleophile in a spontaneous S-acylation reaction using palmitoyl-CoA as the acyl donor. However, it is difficult to know whether spontaneous palmitoylation occurs *in vivo*. It is more likely that protein acyltransferase(s) (PATs) are responsible for palmitoylation under physiological conditions. In fact, PAT activities that cofractionate with the plasma membrane and endomembranes have been reported, suggesting that there is either ubiquitous expression of a single PAT or multiple PATs with distinct subcellular distributions. The recent discovery of protein PATs in model organisms has begun to shed light on this question.

Plasma membrane localization of Ras requires palmitoylation of a conserved cysteine located close to the C terminus. Palmitoylation is associated with the translocation of Ras from the endoplasmic reticulum (ER) to the plasma membrane where it functions in signaling. A genetic screen was carried out in yeast to identify mutants affecting Ras palmitoylation. Mutations in two genes, *ERF2* and *ERF4*, were identified. Lobo and co-workers found Erf2p and Erf4p form a complex that together carries out the enzymatic palmitoylation of Ras. The new enzyme is called Ras PAT. Erf2p has a conserved domain referred to as a DHHC–CRD motif (Pfam designation zfDHHC). Another yeast DHHC–CRD protein, Akr1p, has been shown to be the PAT for yeast casein kinase, Yck2p. Thus, Erf2p/Erf4p and Akr1p are founding members of a new class of S-PATs. The DHHC–CRD motif is found in six other yeast proteins, 23 human proteins, and all other organisms examined.

Not all palmitoylated proteins are intracellular. The *Drosophila* secretes signaling molecules, Hedgehog and Wnt, which are also palmitoylated. Again, genetic screens led to a candidate PAT. The gene goes by the names *skinny hedgehog* (*ski*), *sightless* (*sit*), or *rasp* depending on the group that isolated the mutation. A clue to the function came from a short region of sequence similarity with O-acyltransferases. Members of this family transfer fatty acids onto hydroxyl groups of nonproteinaceous targets. Mutation of residues in the conserved domain of *ski/sit/rasp* results in loss of Hedgehog palmitoylation and signaling.

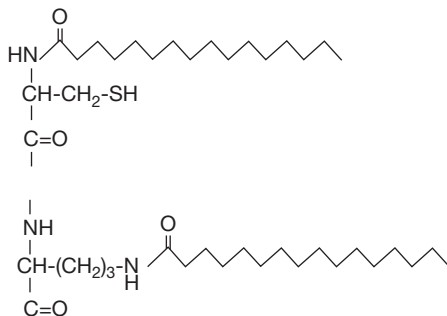
S-palmitoylation

Examples: G proteins and G-protein-coupled receptors, Ras, Src, PSD-95



N-palmitoylation

Examples: Hedgehog, G_s , adenylate cyclase (*Bordetella pertussis*)



O-palmitoylation

Example: myelin proteolipids

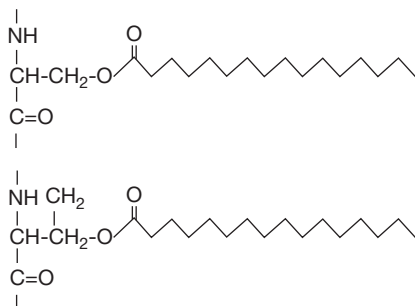


Figure 1 The three major classes of covalent attachment of palmitate in proteins. Representative examples of each type are listed.

Biochemical studies are still needed to confirm that *ski/sit/rasp* is a PAT and to elucidate the mechanism.

The Wnt proteins are cysteine-rich secreted glycoproteins that have recently been shown to be palmitoylated on conserved cysteine. The modification could be removed by an acylprotein thioesterase (APT1) consistent with it being a thioester. Sequence alignment of Wnt family members reveals that this is the first cysteine residue in the sequence that is conserved in all family members. Thus, palmitoylation is likely to be found on all Wnt signaling proteins. Enzymatic depalmitoylation of purified Wnt3a protein results in the loss of its ability to stabilize β -catenin in mouse L cells, a measure of the Wnt signaling pathway. *porcupine* (*por*) is a segment polarity gene in *Drosophila* that is required for processing and secretion of Wingless (Wg). In *por* mutants, Wg is confined to cells where it is synthesized. Like *skinny hedgehog* (*ski*)/*sightless* (*sit*)/*rasp*, porcupine shares sequence homology with O-acyltransferases making it an excellent candidate for the Wnt PAT.

Palmitoylthioesterases

Given the reversible nature of protein palmitoylation, thioesterases are predicted to play an important regulatory role. However, the first protein thioesterases identified, PPT1 and PPT2, are apparently involved in protein turnover. PPT1 found in lysosomes and mutations, which reduce or eliminate its expression, causes a severe neurodegenerative disorder, infantile neuronal ceroid lipofuscinosis (INCL). INCL is formally classified as a lysosomal-storage disorder. The neurological phenotype arising from a deficiency of PPT thioesterases is likely to be caused by the accumulation of palmitoyl peptides or palmitoylcysteine.

A cytosolic palmitoyl protein thioesterase, APT1, was discovered in rat liver extracts. APT1 was originally described as a lysophospholipase, but Duncan and Gilman showed that APT1 cleaves thioesters in acyl-CoAs and acylproteins, as well as oxyesters in lysolipids. In fact, APT1 exhibits a strong substrate preference for acylproteins over lipid substrates consistent with a role for APT1 as a regulator of protein thioacylation and not as a regulator of lipid metabolism. A number of substrates have been proposed for APT1, including G_{α} -subunits, Ras, and eNOS. APT1 orthologs have been identified in a large number of species, ranging from yeast to humans.

Palmitoylation and Protein Function

The functional significance of reversible palmitoylation depends on the protein under consideration. For peripheral membrane/cytosolic proteins, lipid addition can be the primary determinant of membrane recruitment. It can also specify a particular membrane or microdomain of a membrane. A theme that has emerged from many studies is that palmitoylation operates along with other lipid modifications or membrane association sequences. Ras proteins are a case in point. Farnesylation of the C-terminal *CaaX* box targets H-Ras and N-Ras to the cytosolic surface of the ER membrane where it is further processed by a *CaaX* box protease and methyltransferase. Plasma membrane localization of H-Ras requires palmitoylation by the Erf2p/Erf4p Ras PAT. Once at the plasma membrane, H-Ras reversibly associates with microdomains believed to be involved in signaling. There is intense interest in defining the nature of this microdomain and the role of palmitoylation in Ras signaling.

Reversible palmitoylation also influences the subcellular trafficking of integral membrane proteins. For example, receptor endocytosis and recycling is often controlled by palmitoylation status. In virus-infected cells, palmitoylation directs Env proteins to lipid rafts where viral assembly occurs. In addition to surface expression and stability, palmitoylation has been proposed to play a role in signaling of some receptors. Palmitoylation can effect signaling by altering effector binding. For example, mutations in the palmitoylation site on the cytoplasmic tail of the β -adrenergic receptor leads to uncoupling of the receptor, its cognate G protein.

Palmitoylation also plays a role in vesicle trafficking and membrane fusion. Palmitoyl-CoA is required for efficient vesicle fusion in cell free assays. It is generally believed that palmitoylation of a specific protein or proteins accounts for

this requirement. A possible candidate has emerged from the study of vacuole inheritance in yeast. Vac8p is a myristoylated and palmitoylated peripheral membrane protein that is required for vacuole inheritance. Palmitoylation, but not myristoylation, is required for Vac8p function. Palmitoylation of Vac8p may be involved in the interaction of Vac8p with vacuole-specific N-ethylmaleimide sensitive factor attachment protein receptor (SNARE) complexes.

Not all palmitoylation is involved with membrane-mediated processes. Mitochondrial methylmalonyl semialdehyde dehydrogenase is covalently modified by palmitate on an active site cysteine residue. Insertion of [125 I]-labeled analog of myristoyl-CoA at this site results in enzyme inhibition. This observation, together with other results that palmitoyl-CoA inhibits the activity of several mitochondrial enzymes, suggests a regulatory role of S-acylation in metabolism.

Palmitoylation Inhibitors

The role of palmitoylation in cellular processes has been investigated using inhibitors such as 2-bromo or 2-fluoropalmitate, cerulenin, and tunicamycin. Each inhibits protein palmitoylation, but their mechanisms of action may differ. Cerulenin inhibits fatty acid synthase, but analogs with selectivity toward PATs are being developed and tested. Tunicamycin is a fatty acyl-CoA analog and has been found to reduce protein palmitoylation without significantly affecting palmitoyl-CoA levels. In some cases, 2-fluoropalmitate has been shown to decrease uptake of [3 H]palmitate and thereby lower the levels of labeled palmitoyl-CoA levels and apparent protein palmitoylation. The recent identification of palmitoyltransferases should allow a more detailed characterization of these inhibitors and the development of new inhibitors with specificity toward the different classes of palmitoylated proteins.

Concluding Remarks

Protein palmitoylation has emerged as an important posttranslational modification used by cells to organize the subcellular distribution of proteins on membranes. The study of protein palmitoylation has entered a new phase with the identification of PATs in yeast and *Drosophila*. As the tools develop, the role

of protein palmitoylation, in the cellular physiology, should come into better focus. PATs may be excellent targets for the design of drugs that alter the subcellular distribution of receptors and signaling molecules.

See also: **Metabolism Vitamins and Hormones:** Fatty Acid Oxidation; Fatty Acid Structure and Synthesis; **Protein/Enzyme Structure Function and Degradation:** Protein N-Myristoylation; **Signaling:** Lysophospholipid Receptors; Ras Family.

Further Reading

- Bijlmakers MJ and Marsh M (2003) The on-off story of protein palmitoylation. *Trends in Cell Biology* 13: 32–42.
- Chamoun Z, Mann RK, Nellen D, et al. (2001) Skinny hedgehog, an acyltransferase required for palmitoylation and activity of the hedgehog signal. *Science* 293: 2080–2084.
- Duncan JA and Gilman AG (1998) A cytoplasmic acyl-protein thioesterase that removes palmitate from G protein alpha subunits and p21 (RAS). *Journal of Biological Chemistry* 273: 15830–15837.
- El-Husseini Ael D and Bredt DS (2002) Protein palmitoylation: A regulator of neuronal development and function. *Nature Reviews in Neuroscience* 3: 791–802.
- Hancock JF (2003) Ras proteins: Different signals from different locations. *Nature Reviews in Molecular Cell Biology* 4: 373–384.
- Lee JD and Treisman JE (2001) Sightless has homology to transmembrane acyltransferases and is required to generate active Hedgehog protein. *Current Biology* 11: 1147–1152.
- Linder ME (2001) Reversible modification of proteins with thioester-linked fatty acids. In: Tamanoi F and Sigman DS (eds.) *The Enzymes: Protein Lipidation*, vol XXI, pp. 215–240. San Diego, CA: Academic Press.
- Linder ME and Deschenes RJ (2003) New insights into the mechanisms of protein palmitoylation. *Biochemistry* 42: 4311–4320.
- Lobo S, Greentree WK, Linder ME, and Deschenes RJ (2002) Identification of Ras palmitoyltransferase in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* 277: 41268–41273.
- Micchelli CA, The I, Selva E, Mogila V, and Perrimon N (2002) Rasp, a putative transmembrane acyltransferase, is required for Hedgehog signaling. *Development* 129: 843–851.
- Nusse R (2003) Wnts and Hedgehogs: Lipid-modified proteins and similarities in signaling mechanisms at the cell surface. *Development* 130: 5297–5305.
- Resh MD (1999) Fatty acylation of proteins: New insights into membrane targeting of myristoylated and palmitoylated proteins. *Biochimica et Biophysica Acta* 1451: 1–16.
- Roth AF, Feng Y, Chen L, and Davis NG (2002) The yeast DHHC cysteine-rich domain protein Akrlp is a palmitoyl transferase. *Journal of Cell Biology* 159: 23–28.
- Thompson GA Jr. and Okuyama H (2000) Lipid-linked proteins of plants. *Progress in Lipid Research* 39: 19–39.

Protein Tyrosine Phosphatases

D J Pagliarini, F L Robinson, C A Worby, and J E Dixon, University of California, San Diego, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by David J. Pagliarini, Fred L. Robinson, and Jack E. Dixon, volume 3, pp. 536–542, © 2004, Elsevier Inc.

Glossary

Classic ‘protein tyrosine phosphatase’ (PTP) These enzymes are specific for phosphotyrosine substrates.

Dual-specific PTPs (DS-PTPs) In addition to having the ability to dephosphorylate phosphotyrosine substrates *in vitro*, DS-PTPs have been shown to utilize phosphoserine/threonine, RNA, inositol lipids, or phospho-glucans as their physiological substrates.

Kinase Enzymes that work opposite of phosphatases by catalyzing the addition of phosphate group to a target substrate.

Phosphatases Catalyze the removal of a phosphate group from a target substrate.

Receptor-like PTP (RPTP) PTPs possessing a transmembrane domain and one or two classic, phosphotyrosine-specific PTP domain.

Introduction

The process of reversible phosphorylation is perhaps the cell's most prevalent means of regulation at the molecular level. It has been estimated that up to 30% of all cellular proteins are phosphorylated, and phosphorylation has been shown to play a crucial regulatory role in such diverse cellular events as metabolism, growth and differentiation, vesicular transport, and gene transcription. Phosphorylation and dephosphorylation are carried out by kinases and phosphatases, respectively. There are currently predicted to be 518 kinases and ~125 phosphatases encoded in the human genome, further underscoring the overall importance of phosphorylation in molecular regulation. Phosphatases are generally divided into two main families based on their catalytic mechanism and substrate specificity: the protein phosphatases (PPs), which exclusively desphosphorylate serine and threonine residues, and the protein tyrosine phosphatases (PTPs), which can dephosphorylate tyrosine residues, and are the focus of this article. PTPs can be further classified into subfamilies based on (1) subcellular location (receptor vs. intracellular), (2) substrate preference, and (3) three-dimensional topology. In this article, we describe the different subfamilies of PTPs and their conserved catalytic mechanism. In addition, we briefly discuss human diseases that result from disrupted PTP signaling, and discuss the pursuit of PTPs as drug targets.

Catalytic Mechanism of PTPs

PTPs, regardless of subfamily, are predicted to use similar mechanisms for catalyzing the removal of a phosphate group. All PTPs possess a catalytic domain that contains the active-site signature motif cysteine- X_5 -arginine (CX_5R ; X = any amino acid), which constitutes the phosphate-binding pocket, or ‘P-loop’. The dephosphorylation reaction is initiated when the thiol nucleophile of the active-site cysteine attacks the phosphate group on the substrate molecule, forming a cysteinyl-phosphate intermediate (Figure 1). A conserved aspartic acid, usually residing outside of the P-loop, then acts as a general acid to aid in the ejection of the substrate molecule. In the second step of the

reaction, the phosphoenzyme intermediate is hydrolyzed by a water molecule that has been activated by the aspartate, thereby releasing free phosphate and regenerating the active enzyme.

Classification of PTPs

Intracellular PTPs

PTPs

PTPs can be classified as receptor-like (RPTPs) or intracellular based on their location within a cell, whereas all RPTPs to date exhibit specificity toward phosphotyrosine substrates (referred to as ‘classic’ PTPs); intracellular PTPs can be classic or dual-specific (DS-PTPs), having the ability to dephosphorylate serine/threonine, RNA, or inositol lipids in addition to pTyr (Figure 2). The major distinguishing factor between classic and DS-PTPs appears to be the depth of the substrate-binding pocket, with DS-PTPs having a shallower landscape less conducive to the binding of the long side chain of tyrosine. A series of structural studies over the past decade have revealed that all PTPs possess one of three distinct topologies (Figure 3). This first and largest topological family, referred to simply as PTPs, includes all of the RPTPs, classic PTPs such as PTP1B and the *Yersinia* PTP, and DS-PTPs such as vaccinia VH1-related (VHR) and phosphatase and tensin homolog (PTEN), among others (Figure 2).

CDC25 family

The CDC25 family is a set of DS-PTPs involved in the regulation of the cell cycle. Each enzyme of this family is highly specific for a particular serine or threonine residue on a cyclin-dependant kinase/cyclin complex. Although presence of the signature CX_5R motif suggests a common catalytic mechanism, the PTP domains of these enzymes share little sequence homology with other PTPs (Figure 3).

Low-molecular-weight PTPs

The third family of topologically distinct CX_5R phosphatases is the low-molecular-weight PTPs (LMW-PTPs). Conservation of the catalytic cysteine, arginine, and aspartic acid residues strongly suggests that these enzymes also utilize a similar

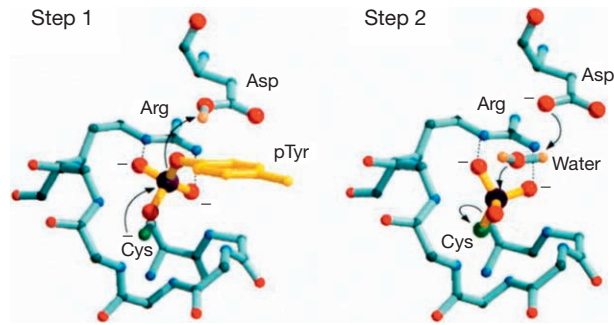


Figure 1 Catalytic mechanism of protein tyrosine phosphatases. Step 1: a thiolate anion of the active-site cysteine performs a nucleophilic attack on the phosphate of phosphoryl-tyrosyl substrates, resulting in the formation of a phospho-enzyme intermediate. An aspartic acid, acting as a general acid, donates a proton to the phosphate group, ejecting the tyrosine side chain. Hydrogen bonds between phosphate oxygen atoms and nitrogen atoms of the arginine guanido group promote phosphate binding and stabilize the transition state. Step 2: acting as a general base, aspartate activates a water molecule for a nucleophilic attack on the phosphorous atom. The phosphorous-sulfur bond is hydrolyzed, thus, regenerating the native enzyme. Reproduced from Denu JM, Stuckey JA, Saper MA, and Dixon JE (1996) Form and function in protein dephosphorylation. *Cell* 87: 361–364.

catalytic mechanism. As with the CDC25 family, however, LMW-PTPs show no homology with other PTPs, and a lack of any other domains outside of their small (~150 amino acids) PTP domain provides few clues to their *in vivo* function. All LMW-PTPs examined to date have been specific for tyrosine-phosphorylated substrates.

Receptor-Like PTPs

RPTPs possess an extracellular receptor-like domain, a single transmembrane domain, and one or two intracellular classic tyrosine-specific PTP domains. Where two PTP domains exist, the membrane proximal domain accounts for the majority of catalytic activity. Based on their active-site homology and common features of their extracellular domains, RPTPs can be further subdivided into eight subtypes: R1/R6, R2A, R2B, R3, R4, R5, R7, and R8 (Figure 2). The CD45 family (R1/R6) is expressed on nucleated hematopoietic cells and is thought to be involved in the positive regulation of antigen receptor signaling via dephosphorylation of the Src family kinases. RPTPs ρ , μ , κ , and λ (R2A) are marked by an extracellular meprin-A5 antigen-PTP μ (MAM) and intracellular cadherin-like domains. They also possess extracellular fibronectin type III-like repeats and an immunoglobulin-like domain on their extracellular region that are involved in cell adhesion. The leukocyte common antigen-related (LAR)-like PTPs (R2B) are primarily expressed in neuronal cells where they are thought to play an important role in neuronal development. They are marked by tandem repeats of immunoglobulin-like and fibronectin type III like domains resembling neuronal adhesion molecules. Members of the R3 subfamily, for example, the mammalian OST-PTP (OST, osteotesticular), possess multiple fibronectin type III repeats resembling cell-adhesion molecules such as collagen type VII, fibronectin, and tenascin. PTP α and PTP ϵ (members of R4) are marked by their relatively small,

highly glycosylated extracellular domains. PTP γ and PTP ζ represent the R5 subfamily, which have a carbonic anhydrase-like domain. RPTPs possessing a meprin domain in addition to immunoglobulin-like and fibronectin type III-like repeats comprise R6. Members of the R7 subfamily possess no additional extra- or intracellular domains, and the more evolutionarily distant R8 family members are predicted to be catalytically inactive.

PTPs and Human Disease

Involvement of PTPs in Mammalian Biology and Disease

At least 18 human protein tyrosine kinases (PTKs) have been shown to behave as onco-proteins when inappropriately activated by mutation. Given this, it was expected that PTPs, which can terminate tyrosine kinase signaling, might be found as genes frequently mutated in cancers. This has generally not proved to be the case. However, links between specific PTPs and human genetic diseases have been, and are continuing to be, established. In addition, numerous studies involving mice with targeted disruptions of individual PTP genes have provided *in vivo* evidence for the involvement of individual PTPs in specific biological processes. Abnormal PTP function has been implicated in a wide array of pathological conditions including diabetes, obesity, cancer, immune disease, neurodegeneration, and vascular disease. In this section, links between PTPs and human disease are briefly discussed. Studies of three PTPs that play essential roles in mammalian physiology are offered as examples. In addition, a summary of the results of PTP gene disruption experiments in mice is provided (Table 1). Finally, the development of PTP inhibitors for use as therapeutics and research tools is discussed.

PTP-1B

A number of studies over the last 5 years have indicated that PTP-1B plays a specific role in downregulating both insulin and leptin receptor signaling in mice. Studies from the laboratories of both Kennedy and Kahn have demonstrated that mice lacking PTP-1B are resistant to diabetes and diet-induced obesity, while displaying no other apparent physiological defects. Enhanced insulin sensitivity in PTP-1B-deficient mice likely results from hyperphosphorylation of the insulin receptor, as these mice display enhanced autophosphorylation of this protein in liver and muscle tissues. PTP-1B has been shown to associate with the insulin receptor PTK *in vivo* and both the insulin receptor and insulin receptor substrate 1 (IRS1) have been implicated as substrates for PTP-1B. Taken together, these data indicate a critical role for PTP-1B in turning off insulin signaling in sensitive mammalian tissues. The apparent *in vivo* specificity of PTP-1B function came as a surprise given the many other cellular processes in which this enzyme was thought to be involved. A number of pharmaceutical companies are attempting to generate specific inhibitors of PTP-1B for use as drugs to treat diabetes and obesity.

CD45

CD45 is an RPTP highly expressed in all nucleated hematopoietic cells. The extracellular portion CD45 contains fibronectin-like repeat sequences (Figure 2), is heavily glycosylated and

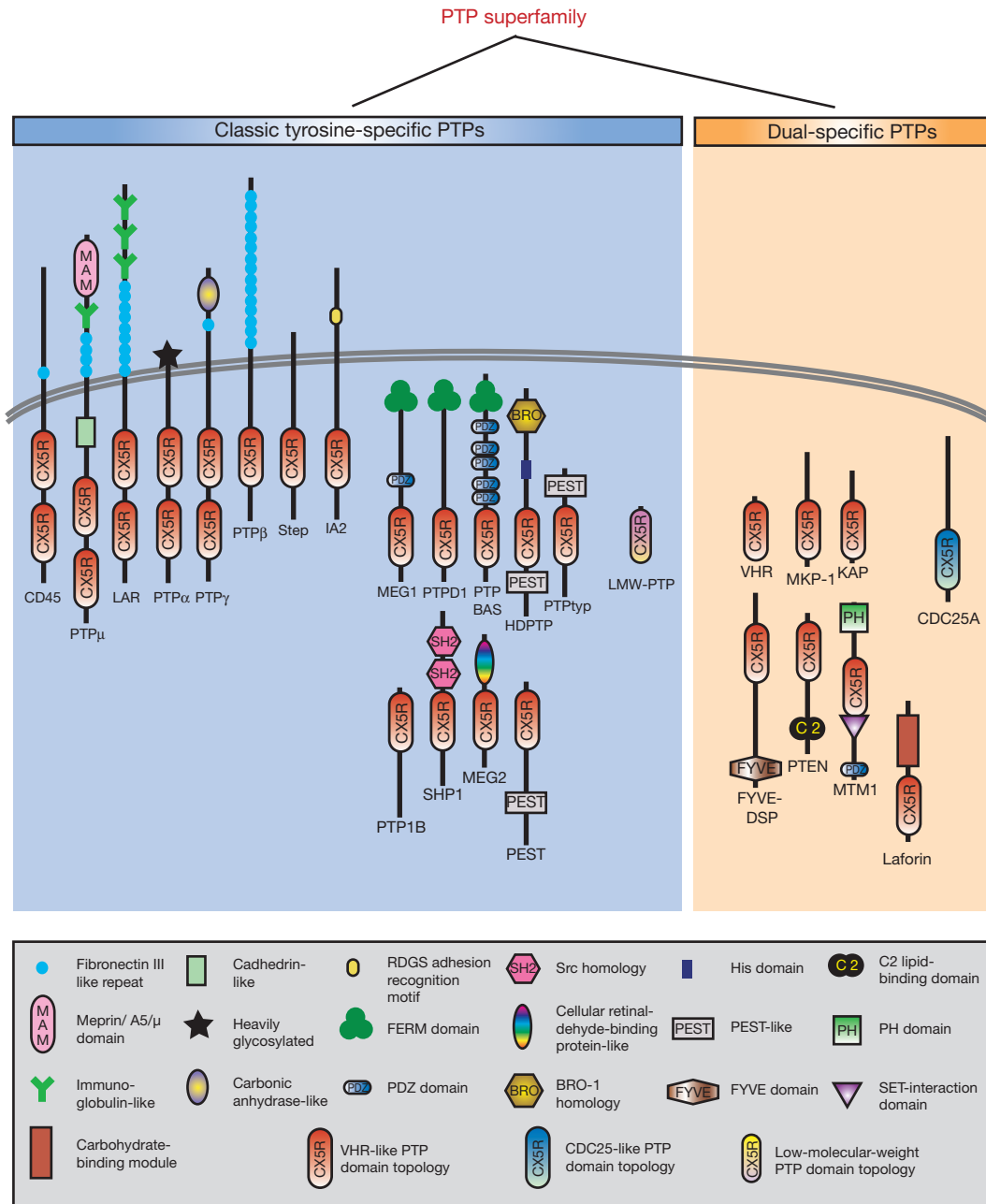


Figure 2 Schematic illustration of representative members of the PTP superfamily.

is involved in the homodimerization of the molecule. Ligands for the extracellular domain of CD45 have not been identified, but this segment of the protein is important for the regulation of phosphatase activity. The cytoplasmic portion of CD45 contains two PTP domains; however, only the membrane proximal PTP domain is catalytically active and required for CD45 function. Homodimerization of CD45 is believed to maintain the phosphatase domain in a low-activity state. CD45 mutations in either humans or mice result in a severe-combined immune deficient (SCID) phenotype, in which both B and T cells are affected (Table 1). Consistent with the SCID phenotype, studies in CD45-deficient mice have indicated a

role for this PTP as a critical positive regulator of immunoreceptor signaling in both B and T cells. In T cells, binding of an antigen to the extracellular portion of the T-cell receptor (TCR) results in the assembly of a cytoplasmic complex of TCR-associated-signaling proteins. The PTKs Lck and Fyn phosphorylate a number of components of the assembled TCR-signaling complex, leading to the recruitment and activation of the ZAP-70 PTK. Activated ZAP-70 participates in the transmission of stimulatory signals to downstream pathway components, ultimately leading to T-cell activation. The kinase activities of Lck and Fyn are negatively regulated by tyrosine phosphorylation on specific residues. By dephosphorylating these key inhibitory

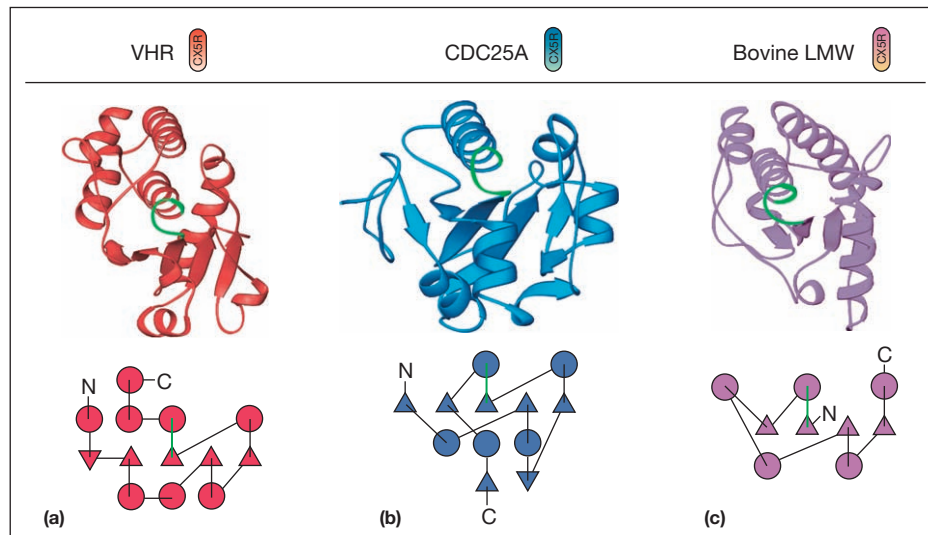


Figure 3 Crystal structures and topologies of representative members of each of the three topologically distinct PTP subfamilies. The amino (N) and carboxyl (C) termini of the proteins are indicated. Circles represent α -helices; triangles represent β -strands. The location of the P-loop possessing the signature CX₅R motif is highlighted in green.

Table 1 Summary of PTP gene disruption experiments in mice

PTP	Mouse gene disruption phenotype and references
CD45	Severe-combined immunodeficiency (Byth et al., 1996; Kishihara et al., 1993; Mee et al., 1999)
CD148	Embryonic lethal (day 11.5); extensive vascular defects (Takahashi et al., 2003)
Cdc25b	Female sterility due to permanent meiotic arrest of oocytes (Lincoln et al., 2002)
Cdc25c	No phenotypic defects identified (Chen et al., 2001)
HePTP	Enhanced ERK activation in lymphocytes, but no physiological defects identified (Gronda et al., 2001)
Laforin	Neurodegeneration and Lafora body accumulation; ataxia, epilepsy, and impaired behavioral responses (Ganesh et al., 2002)
LAR	Decreased basal forebrain cholinergic neuron size and hippocampal cholinergic innervation; delayed regenerative neurite outgrowth following nerve injury; impaired mammary gland development; abnormal glucose homeostasis (Ren et al., 1998; Schaapveld et al., 1997; Van der Zee et al., 2003; Van Lieshout et al., 2001; Xie et al., 2001; Yeo et al., 1997)
PTP-1B	Enhanced sensitivity to insulin, leptin, and growth hormone; obesity resistance (Cheng et al., 2002; Elchebly et al., 1999a; Gu et al., 2003; Klamman et al., 2000; Zabolotny et al., 2002)
PTP α	No gross phenotypic defects; reduced activation of Src and Fyn kinases; altered integrin signaling and cell migration (Ponniah et al., 1999; Su et al., 1999; Zeng et al., 2003; Zheng and Shalloway, 2001)
PTP β	No phenotypic defects identified (Harroch et al., 2000)
PTP δ	Impaired learning; altered hippocampal long-term potentiation (Uetani et al., 2000)
PTP ϵ	No gross phenotypic defects; abnormal macrophage responsiveness; hypomyelination of sciatic nerve neurons; abnormal voltage-gated potassium channel function (Peretz et al., 2000; Sully et al., 2001)
PTP σ	Neonatal or early lethality; developmental defects in neuronal and neuroendocrine cell lineages (Batt et al., 2002; Batt et al., 2003; Elchebly et al., 1999b; McLean et al., 2002; Meathrel et al., 2002; Wallace et al., 1999)
SHP-1	Early lethality (<12 weeks); severe hematopoietic dysregulation and autoimmunity (Kamata et al., 2003; Kozlowski et al., 1993; Shultz et al., 1993; Tsui et al., 1993)
SHP-2	Embryonic lethality (day 10.5) with gastrulation defects; enhanced MAP kinase activation (Arrandale et al., 1996; Saxton et al., 1997)
TC-PTP	Early lethality (<5 weeks); defective bone marrow function; impaired B and T cell function (Dupuis et al., 2003; Galic et al., 2003; Ibarra-Sanchez et al., 2001; You-Ten et al., 1997)

tyrosines, CD45 acts to prime Lck and Fyn for full activation, which is achieved through further phosphorylation. In this manner, CD45 is thought to function by setting an important threshold for T-cell activation.

Laforin

Mutations in the human gene for Laforin, a DS-PTP, have been shown to cause Lafora disease (LD), a progressive form of monoclous epilepsy. The hallmark of LD is the accumulation

of polyglucosan inclusions called Lafora bodies in the cytoplasm of cells of the brain, liver, kidney, skin, skeletal, and cardiac muscle. This human disease phenotype has been replicated in Laforin-deficient mice, through the work of Yamakawa and colleagues. In addition to its DS-PTP domain, Laforin is the only phosphatase in the animal kingdom that possesses a carbohydrate-binding module (CBM), which is required for its association with glycogen particles, and is responsible for targeting the phosphatase to the same subcellular location as

glycogen metabolism. LD was initially proposed to result from abnormal glycogen metabolism, but subsequent studies reported that all enzymes involved in glycogen metabolism from LD patients displayed normal activities. Therefore, Laforin plays a unique and previously unrecognized role in glycogen metabolism that was recently uncovered when Laforin was shown to possess the ability to dephosphorylate phosphoglucans, such as amylopectin and glycogen.

Plants also contain a DS-PTP with a CBM called starch excess 4 (SEX4). Mutations in SEX4 result in an increase in insoluble glucans, a cellular phenotype similar to that seen in LD patients. SEX4 also has the same biochemical properties as Laforin in that it binds glucans and possesses phosphatase activity that releases phosphate from glucans. Although our understanding of the role of glycogen phosphorylation is in its infancy, plant researchers have figured out a mechanistic cause and effect for starch phosphorylation. It appears that phosphorylation of starch solubilizes the outer surface allowing access to the degradation machinery. Then the phosphate is removed by SEX4 so that another round of degradation can occur. Although it follows that mutations in SEX4 result in increased insoluble glucans, it was surprising that this phenotype could be rescued by Laforin expression in plants, making SEX4 and Laforin functionally equivalent. Due to this rare intersection between neuroscience and plant starch metabolism, phosphoglucans can now be added to the physiological substrate list for the DS-PTPs.

PTPs as Drug Targets

Given the potential involvement of PTPs in many aspects of human biology and disease, there is considerable interest in the development of small molecule inhibitors of PTPs for use as therapeutics. In addition to their potential medical value, highly selective and potent PTP inhibitors would be powerful research tools for determining the functions of individual PTPs in mammalian cells. However, the development of specific PTP inhibitors has proved challenging. Like protein kinases, PTPs comprise a large family of enzymes that utilize a common chemical mechanism for catalysis. Accordingly, structure-based drug design will likely be critical to the development of highly specific PTP inhibitors.

Structural studies have suggested a strategy by which inhibitor specificity for individual PTPs might be achieved. Investigations with PTPs have indicated that residues adjacent to the pTyr-binding site play essential roles in achieving substrate specificity and catalytic efficiency. Contemporary work has taken advantage of unique properties of the PTP-1B molecule to develop inhibitors with high selectivity. PTP-1B possesses a second phosphate-binding site on the protein surface, adjacent to its active site. Zhang and colleagues have designed small nonpeptide molecules containing two nonhydrolyzable phosphate moieties. Such molecules inhibit PTP-1B at nanomolar concentrations and are highly selective for this enzyme. High affinity and specificity are achieved by the combined action of the two phosphate moieties, with one binding in the active site and the other at the adjacent site. Such 'bidentate' inhibitor compounds may serve as a model for the development of additional, highly specific

PTP inhibitors, as they target both the active site, which is highly conserved among PTPs, and adjacent portions of the structure, which are less highly conserved.

See also: Signaling: B-Cell Antigen Receptor; Serine/Threonine Phosphatases.

Further Reading

- Andersen JN, Mortensen OH, Peters GH, et al. (2001) Structural and evolutionary relationships among protein tyrosine phosphatase domains. *Molecular and Cellular Biology* 21: 7117–7136.
- Arrandale JM, Gore-Willse A, Rocks S, et al. (1996) Insulin signaling in mice expressing reduced levels of Syp. *Journal of Biological Chemistry* 271: 21353–21358.
- Batt J, Asa S, Fladd C, and Rotin D (2002) Pituitary, pancreatic and gut neuroendocrine defects in protein tyrosine phosphatase-sigma-deficient mice. *Molecular Endocrinology* 16: 155–169.
- Batt J, Cutz E, Fladd C, and Rotin D (2003) Apparent normal lung architecture in protein tyrosine phosphatase-sigma-deficient mice. *American Journal of Physiology – Lung Cellular and Molecular Physiology* 284: L214–L223.
- Blume-Jensen P and Hunter T (2001) Oncogenic kinase signalling. *Nature* 411: 355–365.
- Byth KF, Conroy LA, Howlett S, et al. (1996) CD45-null transgenic mice reveal a positive regulatory role for CD45 in early thymocyte development, in the selection of CD4⁺CD8⁺ thymocytes, and B cell maturation. *Journal of Experimental Medicine* 183: 1707–1718.
- Chen MS, Hurov J, White LS, et al. (2001) Absence of apparent phenotype in mice lacking Cdc25C protein phosphatase. *Molecular and Cellular Biology* 21: 3853–3861.
- Cheng A, Uetani N, Simoncic PD, et al. (2002) Attenuation of leptin action and regulation of obesity by protein tyrosine phosphatase 1B. *Developmental Cell* 2: 497–503.
- Denu JM and Dixon JE (1998) Protein tyrosine phosphatases: Mechanisms of catalysis and regulation. *Current Opinion in Chemical Biology* 2: 633–641.
- Dupuis M, De Jesus Ibarra-Sanchez M, Tremblay ML, and Duplay P (2003) Gr-1+ myeloid cells lacking T cell protein tyrosine phosphatase inhibit lymphocyte proliferation by an IFN-gamma- and nitric oxide-dependent mechanism. *Journal of Immunology* 171: 726–732.
- Elchebly M, Payette P, Michaliszyn E, et al. (1999a) Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science* 283: 1544–1548.
- Elchebly M, Wagner J, Kennedy TE, et al. (1999b) Neuroendocrine dysplasia in mice lacking protein tyrosine phosphatase sigma. *Nature Genetics* 21: 330–333.
- Galic S, Klingler-Hoffmann M, Fodero-Tavoletti MT, et al. (2003) Regulation of insulin receptor signaling by the protein tyrosine phosphatase TCTP. *Molecular and Cellular Biology* 23: 2096–2108.
- Ganesh S, Delgado-Escueta AV, Sakamoto T, et al. (2002) Targeted disruption of the Epm2a gene causes formation of Lafora inclusion bodies, neurodegeneration, ataxia, myoclonus epilepsy and impaired behavioral response in mice. *Human Molecular Genetics* 11: 1251–1262.
- Gentry MS, Dixon JE, and Worby CA (2009) Lafora disease: Insights into neurodegeneration from plant metabolism. *Trends in Biochemical Sciences* 34: 628–639.
- Gronda M, Arab S, lafrate B, Suzuki H, and Zanke BW (2001) Hematopoietic protein tyrosine phosphatase suppresses extracellular stimulus-regulated kinase activation. *Molecular and Cellular Biology* 21: 6851–6858.
- Gu F, Dube N, Kim JW, et al. (2003) Protein tyrosine phosphatase 1B attenuates growth hormone-mediated JAK2-STAT signaling. *Molecular and Cellular Biology* 23: 3753–3762.
- Harroch S, Palmeri M, Rosenbluth J, et al. (2000) No obvious abnormality in mice deficient in receptor protein tyrosine phosphatase beta. *Molecular and Cellular Biology* 20: 7706–7715.
- Ibarra-Sanchez MJ, Wagner J, Ong MT, Lampron C, and Tremblay ML (2001) Murine embryonic fibroblasts lacking TC-PTP display delayed G1 phase through defective NF-kappaB activation. *Oncogene* 20: 4728–4739.
- Jackson MD and Denu JM (2001) Molecular reactions of protein phosphatases – Insights from structure and chemistry. *Chemical Reviews* 101: 2313–2340.

- Kamata T, Yamashita M, Kimura M, et al. (2003) src homology 2 domain-containing tyrosine phosphatase SHP-1 controls the development of allergic airway inflammation. *Journal of Clinical Investigation* 111: 109–119.
- Kishihara K, Penninger J, Wallace VA, et al. (1993) Normal B lymphocyte development but impaired T cell maturation in CD45-exon6 protein tyrosine phosphatase-deficient mice. *Cell* 74: 143–156.
- Klaman LD, Boss O, Peroni OD, et al. (2000) Increased energy expenditure, decreased adiposity, and tissue-specific insulin sensitivity in protein-tyrosine phosphatase 1B-deficient mice. *Molecular and Cellular Biology* 20: 5479–5489.
- Kozlowski M, Mlinaric-Rascan I, Feng GS, et al. (1993) Expression and catalytic activity of the tyrosine phosphatase PTP1C is severely impaired in motheaten and viable motheaten mice. *Journal of Experimental Medicine* 178: 2157–2163.
- Li L and Dixon JE (2000) Form, function, and regulation of protein tyrosine phosphatases and their involvement in human diseases. *Seminars in Immunology* 12: 75–84.
- Lincoln AJ, Wickramasinghe D, Stein P, et al. (2002) Cdc25b phosphatase is required for resumption of meiosis during oocyte maturation. *Nature Genetics* 30: 446–449.
- McLean J, Batt J, Doering LC, Rotin D, and Bain JR (2002) Enhanced rate of nerve regeneration and directional errors after sciatic nerve injury in receptor protein tyrosine phosphatase sigma knock-out mice. *Journal of Neuroscience* 22: 5481–5491.
- Meathrel K, Adamek T, Batt J, Rotin D, and Doering LC (2002) Protein tyrosine phosphatase sigma-deficient mice show aberrant cytoarchitecture and structural abnormalities in the central nervous system. *Journal of Neuroscience Research* 70: 24–35.
- Mee PJ, Turner M, Basson MA, et al. (1999) Greatly reduced efficiency of both positive and negative selection of thymocytes in CD45 tyrosine phosphatase-deficient mice. *European Journal of Immunology* 29: 2923–2933.
- Peretz A, Gil-Henn H, Sobko A, et al. (2000) Hypomyelination and increased activity of voltage-gated K(+) channels in mice lacking protein tyrosine phosphatase epsilon. *EMBO Journal* 19: 4036–4045.
- Ponniah S, Wang DZ, Lim KL, and Pallen CJ (1999) Targeted disruption of the tyrosine phosphatase PTPalpha leads to constitutive downregulation of the kinases Src and Fyn. *Current Biology* 9: 535–538.
- Puius YA, Zhao Y, Sullivan M, et al. (1997) Identification of a second aryl phosphate-binding site in protein-tyrosine phosphatase 1B: A paradigm for inhibitor design. *Proceedings of the National Academy of Sciences of the United States of America* 94: 13420–13425.
- Ren JM, Li PM, Zhang WR, et al. (1998) Transgenic mice deficient in the LAR protein-tyrosine phosphatase exhibit profound defects in glucose homeostasis. *Diabetes* 47: 493–497.
- Saxton TM, Henkemeyer M, Gasca S, et al. (1997) Abnormal mesoderm patterning in mouse embryos mutant for the SH2 tyrosine phosphatase Shp-2. *EMBO Journal* 16: 2352–2364.
- Schaapveld RQ, Schepens JT, Robinson GW, et al. (1997) Impaired mammary gland development and function in mice lacking LAR receptor-like tyrosine phosphatase activity. *Developmental Biology* 188: 134–146.
- Shen K, Keng YF, Wu L, et al. (2001) Acquisition of a specific and potent PTP1B inhibitor from a novel combinatorial library and screening procedure. *Journal of Biological Chemistry* 276: 47311–47319.
- Shultz LD, Schweitzer PA, Rajan TV, et al. (1993) Mutations at the murine motheaten locus are within the hematopoietic cell protein-tyrosine phosphatase (Hcp) gene. *Cell* 73: 1445–1454.
- Su J, Muranjan M, and Sap J (1999) Receptor protein tyrosine phosphatase alpha activates Src-family kinases and controls integrin-mediated responses in fibroblasts. *Current Biology* 9: 505–511.
- Sully V, Pownall S, Vincan E, et al. (2001) Functional abnormalities in protein tyrosine phosphatase epsilon-deficient macrophages. *Biochemical and Biophysical Research Communications* 286: 184–188.
- Takahashi T, Takahashi K, St John PL, et al. (2003) A mutant receptor tyrosine phosphatase, CD148, causes defects in vascular development. *Molecular and Cellular Biology* 23: 1817–1831.
- Tsui HW, Siminovitch KA, de Souza L, and Tsui FW (1993) Motheaten and viable motheaten mice have mutations in the hematopoietic cell phosphatase gene. *Nature Genetics* 4: 124–129.
- Uetani N, Kato K, Ogura H, et al. (2000) Impaired learning with enhanced hippocampal long-term potentiation in PTPdelta-deficient mice. *EMBO Journal* 19: 2775–2785.
- Van der Zee CE, Man TY, Van Lieshout EM, et al. (2003) Delayed peripheral nerve regeneration and central nervous system collateral sprouting in leukocyte common antigen-related protein tyrosine phosphatase-deficient mice. *European Journal of Neuroscience* 17: 991–1005.
- Van Lieshout EM, Van der Heijden I, Hendriks WJ, and Van der Zee CE (2001) A decrease in size and number of basal forebrain cholinergic neurons is paralleled by diminished hippocampal cholinergic innervation in mice lacking leukocyte common antigen-related protein tyrosine phosphatase activity. *Neuroscience* 102: 833–841.
- Wallace MJ, Batt J, Fladd CA, et al. (1999) Neuronal defects and posterior pituitary hypoplasia in mice lacking the receptor tyrosine phosphatase PTPsigma. *Nature Genetics* 21: 334–338.
- Xie Y, Yeo TT, Zhang C, et al. (2001) The leukocyte common antigen-related protein tyrosine phosphatase receptor regulates regenerative neurite outgrowth *in vivo*. *Journal of Neuroscience* 21: 5130–5138.
- Yeo TT, Yang T, Massa SM, et al. (1997) Deficient LAR expression decreases basal forebrain cholinergic neuronal size and hippocampal cholinergic innervation. *Journal of Neuroscience Research* 47: 348–360.
- You-Ten KE, Muise ES, Itie A, et al. (1997) Impaired bone marrow microenvironment and immune function in T cell protein tyrosine phosphatase-deficient mice. *Journal of Experimental Medicine* 186: 683–693.
- Zabolotny JM, Bence-Hanulec KK, Stricker-Krongrad A, et al. (2002) PTP1B regulates leptin signal transduction *in vivo*. *Developmental Cell* 2: 489–495.
- Zeng L, Si X, Yu WP, et al. (2003) PTP alpha regulates integrin-stimulated FAK autophosphorylation and cytoskeletal rearrangement in cell spreading and migration. *Journal Cell Biology* 160: 137–146.
- Zhang ZY (2002) Protein tyrosine phosphatases: Structure and function, substrate specificity, and inhibitor development. *Annual Review of Pharmacology and Toxicology* 42: 209–234.
- Zheng XM and Shalloway D (2001) Two mechanisms activate PTPalpha during mitosis. *EMBO Journal* 20: 6037–6049.

Proteoglycans

S K Olson and J D Esko, University of California, San Diego, La Jolla, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Chondroitin sulfate A type of glycosaminoglycan defined by the disaccharide unit (GalNAc β 1,4GlcA β 1,3)_n, modified with ester-linked sulfate at certain positions and typically found covalently linked to a proteoglycan core protein.

Dermatan sulfate A modified form of chondroitin sulfate in which a portion of the β -D-glucuronic acid residues are epimerized to α -L-iduronic acid.

Glycan Generic term for a sugar or assembly of sugars, existing either in free form or attached to another molecule.

Glycosaminoglycan Polysaccharide side chains of proteoglycans or free complex polysaccharides composed of linear disaccharide repeating units, each composed of a hexosamine and a hexose or a hexuronic acid.

Glycosyltransferase An enzyme that forms a glycosidic linkage between a sugar and an acceptor.

Heparan sulfate A glycosaminoglycan defined by the disaccharide unit (GlcNAc α 1,4GlcA β 1,4/IdoA α 1,4)_n, containing N- and O-sulfate esters at various positions, and typically found covalently linked to a proteoglycan core protein.

Heparin A type of heparan sulfate made by mast cells that has the highest amount of iduronic acid and of N- and O-sulfate residues.

Hyaluronan A glycosaminoglycan defined by the disaccharide unit (GlcNAc β 1,4GlcA β 1,3)_n that is neither sulfated nor covalently linked to protein, referred to in older literature as hyaluronic acid.

Keratan sulfate A polylactosamine (Gal β 1,4GlcNAc β 1,3)_n with sulfate esters at C-6 of GlcNAc and galactose residues found as a side chain of a keratan sulfate proteoglycan.

Lectin A protein (excluding an antibody) able to bind glycans without causing modification of the glycan structure.

Proteoglycan A protein with one or more covalently attached glycosaminoglycan chains.

Sugar nucleotide Activated forms of monosaccharides, such as UDP-Xyl, UDP-Gal, UDP-GlcA, and UDP-GlcNAc, and UDP-GalNAc, typically used as donor substrates by glycosyltransferases.

Proteoglycan Structure

Glycosaminoglycans

Proteoglycans (PGs) consist of two main structural components – the glycosaminoglycan (GAG) chains and the protein cores.

Three classes of GAG chains are covalently attached to PG core proteins: chondroitin sulfate (CS)/dermatan sulfate (DS), heparan sulfate (HS)/heparin, and keratan sulfate (KS) (Figure 1). Their structure, assembly, and pattern of sulfation distinguish each type of GAG. In addition, vertebrates contain a fourth class of GAGs called hyaluronan (hyaluronic acid, HA), but unlike the other GAGs, HA is not covalently bound to a protein core.

Building blocks

Sugars used in GAG formation include the 5-carbon sugar xylose (Xyl) and the 6-carbon sugars galactose (Gal), glucuronic acid (GlcA), N-acetylglucosamine (GlcNAc), and N-acetylgalactosamine (GalNAc). In order to build the chains, individual sugars are added to a nascent chain in a stepwise manner. The enzymes (glycosyltransferases) that catalyze these reactions use activated nucleotide sugar donors, such as UDP-GlcNAc and UDP-GlcA for HA; UDP-Xyl, UDP-Gal, UDP-GlcA, and UDP-GlcNAc or UDP-GalNAc for HS/heparin and CS/DS; and UDP-Gal and UDP-GlcNAc for the repeating units of KS. PAPS (3'-phosphoadenyl 5'-phosphosulfate) is the universal sulfate donor. With the exception of UDP-Xyl, which is derived from UDP-GlcA in the endoplasmic reticulum (ER), all of the sugar nucleotides and PAPS assemble in the cytoplasm, and transporters shuttle them into the Golgi apparatus or endoplasmic reticulum. Addition of sugar and sulfate residues to the

growing chain releases UDP and 3'-phosphoadenyl 5'-phosphate (PAP), which appear to be converted to uridine monophosphate (UMP) and adenosine monophosphate (AMP) for transport out of the Golgi/ER and subsequent recycling in the cytoplasm.

Hyaluronan

Hyaluronan is structurally the simplest GAG. It consists of repeating disaccharide units of [GlcNAc β 1-4GlcA β 1-3]_n, and does not undergo any further modification. HA is unique among the GAGs in that it does not assemble on a protein core and it is made at the plasma membrane instead of the Golgi apparatus. This orientation facilitates its extrusion from the cell into the surrounding extracellular matrix (ECM). The ECM is the network of PGs, glycoproteins, and fibrous proteins that give strength, support, and structure to the cells embedded within it. HA chains are especially long, typically reaching 10⁴ disaccharides. Due to their length they can intertwine to form a mesh-like material with interesting chemical properties, such as water retention and high viscosity. The turgor properties of the vitreous of the eye, the umbilical cord, and rooster combs reflect the high concentration of HA in these tissues. Several proteins, known as hyalectins, contain a fold (link module) that allow specific interactions with HA. HA helps to organize the extracellular matrix, and can bind to signaling and adhesion receptors on the cell surface, affecting growth and migration.

Chondroitin sulfate/dermatan sulfate

Unlike HA, CS/DS chains (as well as HS/heparin) assemble while attached to a core protein. The chains initiate at serine (Ser) residues of protein cores through the addition of Xyl from UDP-Xyl, followed by addition of two Gal residues, and finally

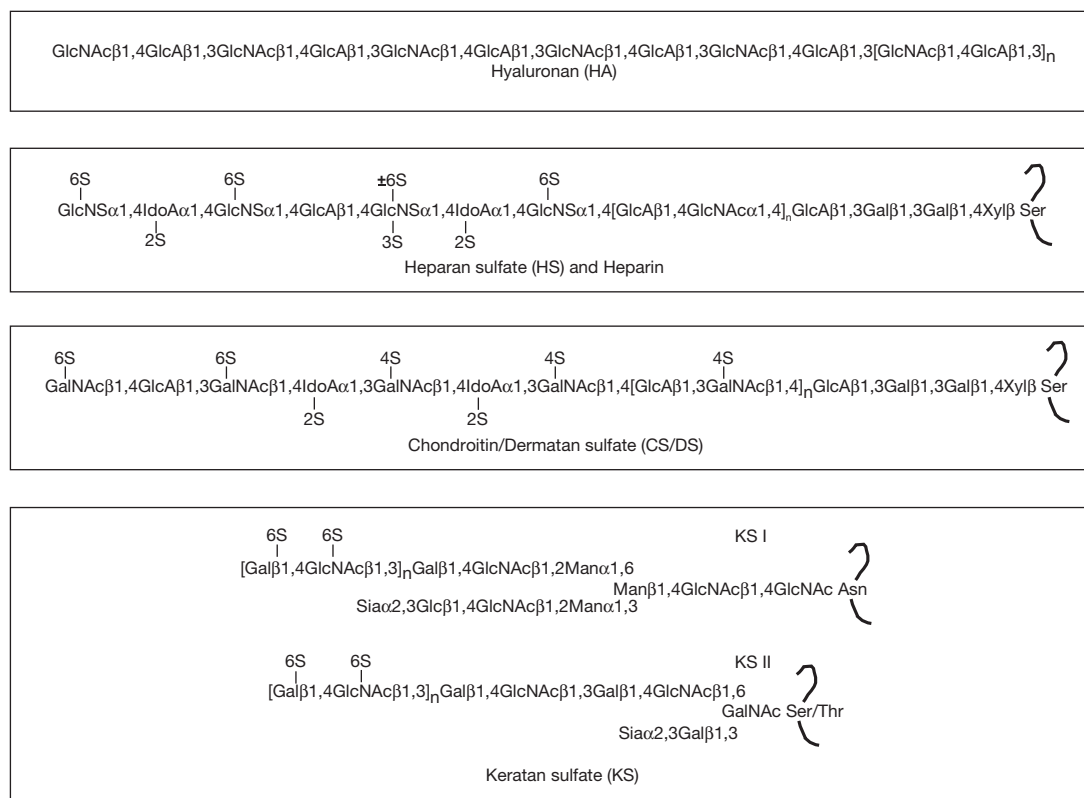


Figure 1 Classes of glycosaminoglycans. Hyaluronan (HA) is made up of repeating units of GlcNAc and GlcA. Heparan sulfate (HS) and chondroitin sulfate (CS) arise from a common linkage tetrasaccharide precursor linked to a protein core at serine (Ser) residues. HS consists of repeating GlcNAc and GlcA, while CS is composed of GalNAc and GlcA. Keratan sulfate (KS) is composed of GlcNAc and Gal residues attached either to an Asn-linked glycan on a glycoprotein (KSI) or to Ser or threonine (Thr) residues in a mucin-type glycoprotein.

GlcA from their corresponding sugar nucleotides. This linkage tetrasaccharide (GlcA β 1–3Gal β 1–3Gal β 1–4Xyl–) can undergo additional modifications, including sulfation of the Gal residues and phosphorylation of Xyl, which may play a regulatory role in the assembly process. Separate enzymes catalyze each step of the assembly of the linkage tetrasaccharide.

Attachment of GalNAc β 1–4 commits the nascent chain to become chondroitin. The stepwise addition of GalNAc β 1–4 and GlcA β 1–3, catalyzed by the chondroitin synthase enzyme-family, elongates the chain. An average chondroitin chain consists of 40 disaccharides, but the size varies in different tissues. The chondroitin backbone can be modified by sulfation of GalNAc units at C-4 and/or C-6, which tends to occur at most positions along the entire chain. In DS, a subset of D-GlcA units undergoes epimerization to L-IdoA. Epimerization of GlcA to IdoA occurs by stereochemical inversion at C-5 (the proton is axial in D-GlcA but planar in L-IdoA). The extent of epimerization is variable, ranging from 5% to 95% dependent on the tissue origin of the chains. IdoA, and to a lesser extent GlcA, can be sulfated at C-2. In invertebrates, such as *Caenorhabditis elegans* and *Drosophila melanogaster*, most of the disaccharides are nonsulfated.

CS/DS PGs are generally secreted into the extracellular matrix where they perform a variety of structural and developmental roles. The model organism *C. elegans* requires a non-sulfated form of chondroitin for the processes of eggshell assembly and epithelial cell invagination in the vulva. In

vertebrates, CS plays numerous structural roles in cartilage, tendon, and other organized extracellular matrices. In the nervous system, it appears to affect axonal outgrowth and guidance. Very few CS-binding proteins have been described as compared to HS/heparin-binding proteins. Some of the properties of chondroitin are related to its biophysical properties (hydrostatic and electrostatic effects). DS binds with certain growth factors and protease inhibitors involved in hemostasis.

HS/Heparin

HS and heparin synthesis initiates in a manner similar to CS/DS by the assembly of the linkage tetrasaccharide. The nascent chain becomes heparan upon addition of a GlcNAc α 1–4 residue. Polymerization of the chain then ensues by the stepwise addition of GlcA β 1–4 and GlcNAc α 1–4. As the chain elongates, it undergoes a number of modifications. Subsets of GlcNAc residues can be N-deacetylated, N-sulfated, and O-sulfated at the C-3 and C-6 positions. As in the case of CS/DS, epimerization of GlcA to IdoA can occur, but only at residues neighboring GlcNS. Additionally, the uronic acids can be sulfated at C-2. HS chains are generally modified in sections, with highly sulfated blocks interspersed between nonmodified regions. Variable deacetylation, epimerization, and sulfation patterns give rise to specific binding sites for a large number of ligands, including growth factors (fibroblast growth factors (FGFs), Wnts (members of the Wnt family of proteins related to wingless in flies), bone morphogenetic proteins (BMPs)),

anticoagulant proteins (antithrombin III, heparin cofactor II, which also binds DS, and thrombin), and matrix components (e.g., laminin, fibronectin, and some collagens). HS plays a vital role in numerous biological processes involved in cell growth, migration, and differentiation. A human disease called hereditary multiple exostoses (MHE), characterized by bony outgrowths (osteochondromas) on the long bones, is caused by mutation of the enzymes responsible for polymerizing the HS chain (EXT1 and EXT2).

Heparin is a form of HS that has potent anticoagulant activity due to its ability to bind and activate the protease inhibitor, antithrombin, by way of a unique pentasaccharide sequence. In general, heparin is shorter than HS (~14 kDa vs. ~75 kDa) and is much more extensively modified compared to HS, containing as much as 85% GlcNS units and 75% IdoA, whereas HS contains 40–50% GlcNS and 25–40% IdoA. However, like heparin sections of HS chains can be extensively modified by sulfation and epimerization. HS is made by virtually all cells, whereas heparin is made selectively by connective tissue-type mast cells and stored in secretory granules. Therapeutic heparin derives from porcine and equine tissue digests and is fractionated by precipitation methods to yield highly sulfated material with high anticoagulant activity. Therapeutic heparin is one of the largest commercial medicinal natural products in use worldwide. It also has many other useful properties related to blocking inflammation and infection, which are currently in development. Commercial heparin-Sepharose is often used in research laboratories to aid in the purification of proteins.

Keratan Sulfate

KS is classified as a glycoprotein as well as a PG. The chain consists of repeating Gal β 1–4GlcNAc β 1–3 units (*N*-acetylglucosamine), containing sulfate groups in the C-6 positions of Gal and GlcNAc residues. Two forms of KS PGs vary by their linkage to core proteins. In KSI the linkage of the poly-*N*-acetylglucosamine is identical to conventional Asn-linked glycoproteins, whereas in KSII the linkage is identical to that found in mucins (Figure 1). Like HS, the disaccharide units of KS can be unsulfated, monosulfated, or disulfated. KS is expressed in cartilage and a number of epithelial and neural tissues, playing a role in events as diverse as collagen fiber assembly, wound healing, and embryogenesis. They are also present in neurosecretory vesicles, suggesting unique roles in vesicle formation and neuronal activity.

Protein Cores

With the exception of HA, all GAGs are synthesized while attached to a protein core. Some PG cores serve functional roles independent of the associated GAG chains by binding to ECM molecules and the cytoskeleton. The number of GAG chains carried by individual PGs depends on the number of attachment sites, ranging from a single chain (decorin) up to 100 chains (aggrecan) (Table 1). Most PGs contain multiple glycosylation sites, any or all of which may be used. A loose consensus sequence has been identified for HS and CS glycosylation sites on the protein core. Xyl is added to Ser residues

that are followed by glycine (Gly), typically flanked by two or more acidic amino acids (Asp or Glu). The factors that regulate the efficiency of initiation are unknown. Evidence also exists that the presence of hydrophobic residues near the attachment sites, the density of attachment sites, as well as inhibitory sequences in the protein cores determine the selection of sites for the assembly of HS/heparin or CS/DS.

PG Families

More than 25 PGs have thus far been identified in mammalian systems (Table 1), and many fall into gene families based on sequence homologies. Some PGs carry a single type of GAG, whereas others can be hybrid molecules carrying HS and CS, or CS and KS. Many PGs also carry other types of O-linked and N-linked glycans. Some PGs are large multidomain molecules reaching masses of 400 kDa, with subdomains of varying functions (e.g., cell attachment, association with other ECM proteins, binding of growth factors, and intrinsic growth modulatory activities), while others have small masses around 10 kDa. While structurally and functionally diverse, they can be classified into the following families based on localization and homology. Note that invertebrates also express numerous PGs, including heparan sulfate proteoglycans (HSPGs) related to those found in vertebrates and chondroitin PGs that are unique.

Secreted and Matrix PGs

The majority of PGs secreted into the ECM carry CS or DS chains. This group includes the small interstitial PGs, the aggrecan family, and secretory granule PGs, such as serglycin. Perlecan, agrin, and collagen XVIII are the only known HSPGs specifically secreted into the ECM. However, some membrane HSPGs (e.g., syndecans and glypicans) can be shed from the cell surface and the ectodomains can become part of the ECM or accumulate in extracellular fluids.

Small Interstitial PGs

This class of PGs includes proteins with leucine-rich repeats flanked by cysteines in their central domain, the so-called small leucine-rich proteoglycans (SLRPs). The family has at least nine members, and all carry either CS/DS or KS GAG chains. Members include decorin and biglycan, which interact with collagen in the matrix. Much work to date has concentrated on understanding the role of these PGs in stabilizing and organizing collagen fibers. In decorin, this activity appears to be dictated by the PG core because removal of the GAG chains fails to affect fibrillogenesis. Mice lacking the decorin PG live to adulthood but show a fragile skin defect, and exhibit great variation in collagen fibril diameter and spacing. Human patients with a mutation in the decorin gene develop congenital stromal corneal dystrophy, characterized by opacity of the cornea due to irregular spacing of collagen fibrils. KSPGs also maintain the register of collagen fibers in the cornea, which is required for transparency and curvature. Humans with a mutation in the keratocan gene develop Cornea plana 2, a recessive disease in which patients exhibit severe far-sightedness due to defects in corneal curvature. Decorin is also able to bind

Table 1 Proteoglycan families

<i>Class</i>	<i>Family</i>	<i>Proteoglycan</i>	<i>Core mass (kDa)</i>	<i>Chain #</i>	<i>Chain type</i>	<i>Tissue</i>	<i>Human disease</i>
Matrix/secreted	Small interstitial	Decorin	36	1	CS	Connective tissue cells	Congenital stromal corneal dystrophy
		Biglycan	38	1–2	CS	Connective tissue cells	
		Endocan	20	1	CS	Endothelial cells	
		Epiphycan	35	2–3	CS	Cartilage	
		Fibromodulin	42	2–3	KS	Broad	
		Lumican	38	3–4	KS	Broad	
		Keratocan	38	3–5	KS	Broad	
		Mimecan/osteoglycin	25–35	2–3	KS	Broad, sulfated in cornea	
	Aggrecan	Osteoadherin	49	2–3	KS	Bone	Cornea plana
		Aggrecan	208–220	~100	CS/KS	Cartilage	
		Versican	265	12–15	CS	Fibroblasts	
		Neurocan	145	1–2	CS	Brain	
	Basement membrane	Brevican	96	0–4	CS	Brain	Schwartz–Jampel syndrome (skeletal dysplasia), Silverman–Handmaker dyssegmental dysplasia
		Perlecan	400	1–3	HS	Basement membranes	
		Agrin	212	2–3	HS	Basement membranes	
		Collagen XVIII	150	1–3	HS	Epithelial cells, basement membranes	
	Intracellular	Serglycin	10–19	10–15	Heparin/CS	Mast cells, macrophages	Multiple epiphyseal dysplasia
		SV2	80	1–3	KS	Synaptic vesicles	
	Miscellaneous	Collagen α 2IX	68	1	CS	Cartilage, vitreous humor	
		Phosphacan (splice variant of RPTPb)	176		CS/KS	Brain	
	Syndecans	Testicans 1–3	50	1–3	HS/CS	Testes, brain endothelium	
		Syndecans 1–4	31–45	1–3	HS/CS	Epithelial cells, fibroblasts	
	Glypicans	Glypicans 1–6	~60	1–3	HS	Epithelial cells, fibroblasts	
Membrane bound	Miscellaneous	Betaglycan	110	1–2	HS/CS	Fibroblasts	Thrombophilia, atypical hemolytic uremic syndrome 6
		Thrombomodulin	58	1	CS	Endothelial cells	
		CD44	37	1–4	CS	Lymphocytes	
		NG2	251	2–3	CS	Neural cells	
		Invariant chain	31	1	CS	Antigen-processing cells	
		Neuropilin-1	130	1	HS/CS	Endothelial cells	
		CD47	35		HS/CS	T-cells	

transforming growth factor beta (TGF- β), serving as a sink to keep the growth factor sequestered.

Aggrecan family

The Aggrecan family of PGs includes aggrecan, versican, brevican, and neurocan. All four members share these common features: the N-terminal domain is able to bind HA, the central region contains GAG attachment sites, and the C-terminal domain contains a C-type lectin domain (lectins are nonimmunoglobulin, nonenzymatic proteins that bind sugar chains). Aggrecan is the most well-known member of this family, since it is the major PG present in cartilage. It contains by far the largest number of GAG chains, as many as 100 CS and KS chains per protein. It complexes with HA in the ECM and draws water into cartilage, thereby providing the ability to resist compressive forces. Mutations in the human aggrecan gene have been linked to two different forms of spondyloepimetaphyseal dysplasia, a skeletal dysplasia characterized by short stature (dwarfism) and other problems associated with the importance of cartilage in bone development. Versican is the next largest member of the aggrecan family with a mass of 265 kDa, and it also undergoes alternative splicing events that generate smaller proteins. Roles in neural crest cell and axonal migration have been suggested. Human patients with mutations in the versican gene exhibit Wagner syndrome, a should be vitreoretinopathy that causes lesions of the vitreous of the eye, and occasionally retinal detachment and cataracts. Neurocan and brevican are both localized to the brain. Neurocan is expressed in the late embryonic central nervous system (CNS) and acts to inhibit neurite outgrowth. Brevican is expressed in the terminally differentiated CNS, particularly in perineuronal nets.

Basement membrane PGs

The basement membrane is an organized layer of ECM that separates epithelial and endothelial cells from the fibrous matrices of connective tissues. It consists of laminin, collagen, nidogen, and usually HSPGs. Basement membranes contain three types of large PGs – perlecan, agrin, and collagen XVIII – each containing one to three HS chains (though perlecan has been shown to carry CS on occasion, e.g., in rib cartilage). Perlecan is the largest known HSPG with a mass of 400 kDa and is a modular protein with multiple domains that serve numerous functions. It is present in the pre-implantation blastocyst and endothelial cells, and appears to play a role in embryogenesis and tissue morphogenesis. In humans, mutations in the perlecan gene cause Schwartz–Jampel syndrome (SJS) and Silverman–Handmaker dyssegmental dysplasia (DDSH), which are skeletal dysplasias of varying severity and can cause either short stature (SJS) or neonatal lethal short-limbed dwarfism (DDSH). These disorders highlight the importance of perlecan in cartilage and bone development. Agrin is a well-characterized PG that acts in the nervous system by aggregating acetylcholine receptors at the neuromuscular junction, and is additionally expressed in the renal tubules where it plays an important role in determining the filtration properties of the kidney. Mutations in the human agrin gene cause congenital myasthenic syndrome, which is characterized by muscle weakness due to defects at the neuromuscular junction. Collagen XVIII is prevalent in the liver in the space of Disse that separates the endothelial cells of the sinusoids from

the hepatocytes, and throughout the vasculature. In human patients, mutations in the collagen XVIII gene have been associated with the vitreoretinal degeneration, retinal detachment, and near-sightedness associated with type I Knobloch syndrome. Homologs of the basement membrane PGs exist in flies and worms.

Secretory granule PGs

Serglycin is the main PG localized to intracellular cytoplasmic secretory granules. It is synthesized by endothelial, endocrine, and hematopoietic cells, and it can carry CS or heparin chains. Serglycin is stored in secretory granules until the host cell becomes activated, resulting in secretion of granular contents into the surrounding environment. Serglycin may play a role in packaging granular contents, maintaining granule proteases in an inactive state, and regulating activities after secretion, such as coagulation, host defense, and wound repair. Heparin, the highly sulfated form of HS, is thought to be made exclusively on serglycin present in mucosal mast cells, but macrophages also produce a highly sulfated form of HS.

Membrane-Bound PGs

Many PGs reside in cell membranes as type I transmembrane proteins (e.g., syndecan family) or glycosylphosphatidylinositol (GPI)-linked proteins (e.g., glypican family).

Syndecans

In syndecans, a short hydrophobic domain spans the membrane, linking the larger extracellular domain containing the GAG attachment sites to a smaller intracellular cytoplasmic domain. There are four members in mammalian systems, whereas lower organisms such as flies and worms contain only one homolog. They can carry HS and/or CS chains, which bind a wide range of extracellular ligands such as growth factors and matrix molecules. These interactions likely account for the requirement of syndecans at numerous stages of embryonic development and in the adult animal. Syndecans can additionally coordinate signals from the extracellular environment to the intracellular cytoskeleton via their cytoplasmic tails. For example, binding of a ligand to a GAG chain can induce oligomerization of syndecans at the cell surface. This leads to recruitment of factors at the cytoplasmic tail of syndecan, such as kinases (e.g., c-Src), PDZ-domain proteins, or cytoskeletal proteins. The recruitment of cytosolic proteins in turn triggers a signal that leads to actin filament assembly. Proteolytic cleavage of the syndecans occurs, resulting in release of the ectodomains bearing the GAG chains. These ectodomains can have potent biological activity. Human genetic studies have identified single nucleotide polymorphisms (SNPs) in the syndecan-3 (SDC3) gene. In one study, a particular SNP was closely correlated with obesity in Korean subjects, suggesting that the role of SDC3 polymorphisms in progression to obesity may have ethnic differences. In mice, alterations of syndecan-1 in hepatocytes result in hypertriglyceridemia.

Glypicans

The glypican family of cell-surface PGs is involved in numerous aspects of animal development. The C-terminus is embedded

in the membrane by a hydrophobic GPI-linked anchor, and thus the glypicans do not have a cytoplasmic tail like the syndecans. The N-terminal portion of the protein is globular and has been shown to carry only HS chains *in vivo*. The glypicans bind a wide array of factors that are essential for the development and morphogenesis. Their roles in growth factor signaling potentially include both positive and negative regulation. HS chains are able to positively promote signaling by bringing a ligand and its receptor into close proximity. On the other hand, HS chains could negatively regulate signaling by binding a particular ligand and preventing its diffusion or interaction with a receptor.

Six glypican family members exist in mammals, two in flies, and one in worms. Glypican-3 (GPC3) is the best-studied member of the family. Human patients lacking functional GPC3 exhibit Simpson–Golabi–Behmel syndrome, which is characterized as an overgrowth disorder. The overgrowth phenotype suggests that GPC3 normally functions to inhibit cell proliferation, likely mediated at the level of growth factor signaling at the cell surface. Mutation of GPC6 causes omodysplasia, a condition where the arms and/or legs are stunted in growth and appear club-like. This condition is thought to be caused by misregulated growth factor signaling at the bone growth plate, preventing proper chondrocyte differentiation. Mutation of GPC1 decreases the size of the brain through regulation of fibroblast growth factor signaling in early neurogenesis.

Miscellaneous membrane PGs

A number of membrane PGs do not show homology to syndecans or glypicans, but are expressed on the surface of many different cell types. The CSPG NG2 is a surface marker expressed on stem cell populations that are able to divide throughout an organism's life, including cartilage chondroblasts, myoblasts, endothelial cells of the brain, and glial progenitors. It has been shown to play a role in angiogenesis and inhibits neurite outgrowth. CD44 is a transmembrane cell-surface receptor that plays a role in processes as diverse as immune function, axon guidance, and organ development. CD44 is considered a PG since it can carry GAG chains in certain cell types, such as lymphocytes. It is also an HA-binding protein. The GAG chains allow CD44 to interact with ECM proteins and growth factors. Neuropilin-1 is a subunit of receptor complexes that bind to semaphorins and vascular endothelial growth factors (VEGFs).

Processing

PGs are born in the ER, undergo glycosylation and sulfation in the Golgi, and appear at the cell surface or in the extracellular matrix. However, important steps of the PG life cycle also occur at the cell surface and other intracellular compartments.

Cell-Surface Desulfation

Although all of the sulfation reactions occur in the Golgi apparatus as the chains assemble prior to secretion, a family of cell-surface endosulfatases (Sulfs 1 and 2) has been identified that removes sulfate groups from the C-6 position of glucosamine residues in HS. Sulfs can promote Wnt growth

factor signaling through HS in a finely tuned manner by desulfating HS at the appropriate time and place. Sulfs also can decrease other growth factor signaling events.

Shedding and Heparanase Cleavage

PGs are turned over on a continual basis (Figure 2). Both the syndecans and glypicans undergo proteolytic cleavage and shedding at the cell surface, mediated by one or more of the matrix metalloproteases. This process appears to be regulated and can be induced by foreign organisms that exploit the ectodomains for colonization. Secreted enzymes called heparanases can cleave HS chains into fragments at defined sites, liberating small HS fragments. Cells might use heparanase during tissue remodeling and migration, for example, during metastasis.

Endocytosis

Cells internalize HA, CS/DS, and KS PGs by endocytosis, most likely by macropinocytosis. Incoming vesicles containing the PGs fuse with the lysosomes where sulfatases and exoglycosidases degrade the chains completely, while proteases degrade the core proteins. The liberated sugars (GlcNAc, GalNAc, Gal, and Man) and amino acids can then be recycled for the formation of new GAG chains and PGs, whereas the uronic acids and Xyl are thought to undergo oxidative catabolism since no

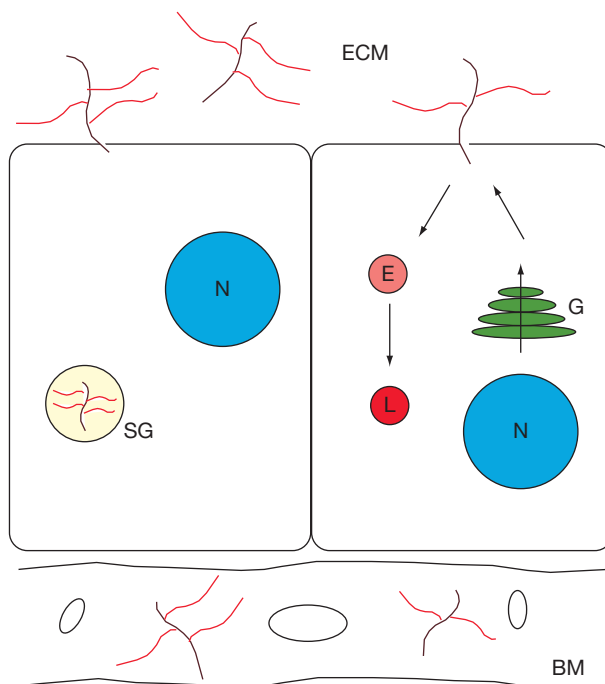


Figure 2 Life cycle of a proteoglycan. Proteoglycans are predominantly synthesized in the Golgi apparatus (G). A mature proteoglycan can be embedded in the plasma membrane, stored in secretory granules (SG), or secreted into the extracellular matrix (ECM) or basement membrane (BM). Degradation of proteoglycans begins with endocytic internalization, followed by cleavage of the GAG chains and protein core by glycosidases and proteases, respectively, in endosomes (E) and lysosomes (L). Nucleus is denoted as N.

salvage pathways for these sugars have been identified. Lysosomes also contain enzymes for degrading HA and CS/DS. Many ligands that bind to cell-surface PGs can enter the cells during the internalization process. In fact, this mechanism is a major route for clearance of triglyceride-rich lipoproteins in the circulation (remnant lipoprotein receptors). Endocytosis of ligands bound to PG receptors may be a general function of continuous turnover of the PGs. Disruption of this pathway leads to lysosomal-storage disorders known as the mucopolysaccharidoses (MPS), which have devastating health consequences for afflicted individuals.

See also: **Cell Architecture and Function:** Endoplasmic Reticulum-Associated Protein Degradation; **Lipids Carbohydrates Membranes and Membrane Proteins:** Glycosylation, Congenital Disorders of.

Further Reading

- Bernfield M, Gotte M, Park PW, et al. (1999) Functions of cell surface heparan sulfate proteoglycans. *Annual Review of Biochemistry* 68, 729–777.
- Bishop JR, Schuksz M, and Esko JD (2007) Heparan sulphate proteoglycans fine-tune mammalian physiology. *Nature* 446 (7139): 1030–1037.
- Bülow HE and Høbert O (2006) The molecular diversity of glycosaminoglycans shapes animal development. *Annual Review of Cell and Developmental Biology* 22, 375–407.
- Esko JD and Selleck SB (2002) Order out of chaos: Assembly of ligand binding sites in heparan sulfate. *Annual Review of Biochemistry* 71, 435–471.
- Funderburgh JL (2002) Keratan sulfate biosynthesis. *IUBMB Life* 54, 187–194.
- Gallagher JT (2001) Heparan sulfate: Growth control with a restricted sequence menu. *Journal of Clinical Investigation* 108, 357–361.
- Gorsi B and Stringer SE (2007) Tinkering with heparan sulfate sulfation to steer development. *Trends in Cell Biology* 17, 173–177.
- Habuchi H, Habuchi O, and Kimata K (2004) Sulfation pattern in glycosaminoglycan: Does it have a code? *Glycoconjugate Journal* 21, 47–52.
- Iozzo R (1998) Matrix proteoglycans: From molecular design to cellular function. *Annual Review of Biochemistry* 67, 609–652.
- Iozzo RV, Zoeller JJ, and Nyström A (2009) Basement membrane proteoglycans: Modulators *Par Excellence* of cancer growth and angiogenesis. *Molecules and Cells* 27 (5): 503–513.
- Kolset SO, Prydz K, and Pejler G (2004) Intracellular proteoglycans. *Biochemical Journal* 379, 217–227.
- Lamanna WC, Kalus I, Padva M, Baldwin RJ, Merry CL, and Dierks T (2007) The heparanome – the enigma of encoding and decoding heparan sulfate sulfation. *Journal of Biotechnology* 129, 290–307.
- Lindahl U (1999) What else can “heparin” do? *Haemostasis* 29, 38–47.
- Song H and Filmus J (2002) The role of glypicans in mammalian development. *Biochimica et Biophysica Acta* 1573, 241–246.
- Trowbridge JM and Gallo RL (2002) Dermatan sulfate: New functions from an old glycosaminoglycan. *Glycobiology* 12, 117R–125R.
- Varki A, Cummings R, Esko J, et al. (2009) (eds.) *Essentials of Glycobiology*, 2nd edn Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Vlodavsky I, Goldshmidt O, Zcharia E, et al. (2002) Mammalian heparanase: Involvement in cancer metastasis, angiogenesis and normal development. *Seminars in Cancer Biology* 12, 121–129.
- Volpi, N (Ed.), 2006. *Chondroitin Sulfate: Structure, Role and Pharmacological Activity*. Academic Press, Advances in Pharmacology, vol. 53, pp. 1–568. London.
- Vreys V and David G (2007) Mammalian heparanase: What is the message? *Journal of Cellular and Molecular Medicine* 11, 427–452.
- Yamaguchi Y (2000) Lecticans: Organizers of the brain extracellular matrix. *Cellular and Molecular Life Sciences* 57, 276–289.
- Yoneda A and Couchman J (2003) Regulation of cytoskeletal organization by syndecan transmembrane proteoglycans. *Matrix Biology* 22, 25–33.

Proton Pumping in the Respiratory Chain

V Rauhamäki and M Wikström, University of Helsinki, Helsinki, Finland

Published by Elsevier Inc.

Glossary

Antiporter A membrane-bound proteinaceous entity that catalyzes transmembrane transport of a molecule (or ion) in one direction with obligatory linkage to transport of another molecule (ion) in the opposite direction. Examples are the adenine nucleotide ($\text{ADP}^{3-}/\text{ATP}^{4-}$) and inorganic phosphate ($\text{H}_2\text{PO}_4^-/\text{OH}^-$) carriers of mitochondria and Na^+/H^+ exchangers of bacteria. Note that the former is electrogenic, whereas the two latter are electroneutral.

Electron carrier A protein, a prosthetic group in a protein, or a cofactor, that may exist in an oxidized and a reduced form, which differ by the absence or presence of an electron.

F_1F_0 ATP synthase The membrane-bound enzyme complex that catalyzes ATP synthesis driven by protonmotive force in mitochondrial and bacterial oxidative phosphorylation, and in photosynthesis.

Hydrogen carrier A protein, a prosthetic group in a protein or a cofactor, that may exist in an oxidized and a reduced form, which differ by the absence or presence of hydrogen (usually two hydrogen atoms, such as ubiquinone/ubiquinol).

Mitochondrial membrane Usually refers to the highly invaginated inner (cristae) membrane of mitochondria that harbors the protein complexes of oxidative phosphorylation, and which is relatively impermeable to ions.

Proton pump A membrane-bound enzyme complex that catalyzes translocation of protons across a membrane against a proton electrochemical gradient, driven by an exergonic reaction.

Protonmotive force Equals the electrochemical proton gradient across a membrane, being the sum of a concentration term (the pH-gradient) and an electric term (the membrane potential).

Q-cycle The ubiquinone cycle is the name of the mechanism by which the cytochrome bc_1 complex and the related photosynthetic cytochrome b_6f complex catalyze redox-linked generation of protonmotive force.

Redox couple The collective term for the oxidized and reduced forms of an electron or hydrogen carrier.

Redox loop A mechanistic principle proposed by Peter Mitchell, according to which redox reactions in respiration and photosynthesis are orientated in membranes in such a way as to transfer neutral hydrogen in one direction and negatively charged electrons in the opposite direction, leading overall to net translocation of protons.

Redox potential E_h (in volts, usually relative to NHE, the normal hydrogen electrode) reports the tendency of a redox couple to accept an electron. At thermodynamic equilibrium, all redox couples experience the same redox potential.

Aerobic organisms depend on the availability of molecular oxygen (O_2), which serves as a sink for electrons released upon oxidation of hydrocarbon foodstuffs in catabolism, a series of events referred to as cellular respiration. Transfer of these electrons to O_2 , reducing it to water in the final stage, is an exergonic process (i.e., it proceeds with decrease of free energy) that is catalyzed in the inner mitochondrial membrane, or in the cell membrane of aerobic prokaryotes, by transmembrane protein complexes that are collectively called the respiratory chain. The three enzymes forming the respiratory chain are able to convert energy freed from the electron-transfer reactions, including the chemistry of O_2 reduction to water, into an electrochemical proton gradient across the membrane, which drives the synthesis of adenosine triphosphate (ATP) (called oxidative phosphorylation) and membrane transport reactions. Here we briefly define the basic mechanistic principles by which the respiratory chain creates a proton gradient, which is the basis of primary energy conservation in aerobic organisms.

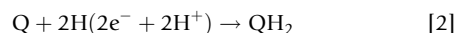
The Proton-Translocating Respiratory Chain, an Overall Perspective

The respiratory chain is a functional concept (Figure 1). Reducing equivalents are fed into the chain in the form of a reduced water-soluble cofactor, nicotinamide adenine dinucleotide

(NADH). NADH is produced from its oxidized counterpart, NAD^+ , especially in Krebs' cycle (also known as the citric acid or the tricarboxylic acid cycle) and during fatty acid oxidation. Dehydrogenases catalyze a reaction



where AH_2 is the hydrogenated metabolite substrate and A its oxidized form. Of the two hydrogen atoms removed from the substrate, a hydride ion is utilized for reduction of NAD^+ , while a proton is liberated into the medium due to the acid-base properties of NADH. At neutral pH, the midpoint oxidoreduction or redox potential of the NAD^+/NADH couple is -320 mV (this is the $E_{m,7}$ value), relative to the normal hydrogen electrode (NHE). NADH transfers its reducing equivalents to the first component of the respiratory chain, for example, complex I, which functions as an NADH-ubiquinone oxidoreductase. Ubiquinone, the acceptor of the reducing equivalents, is a highly apolar hydrogen carrier diffusing within membranes. The redox reaction of the ubiquinone/ubiquinol couple



has an $E_{m,7}$ of $c. 60$ mV. Ubiquinone is present in the inner mitochondrial membrane at a concentration far higher than the respiratory chain complexes. Under standard conditions ($\text{pH}=7$ and equal concentrations of oxidized and reduced

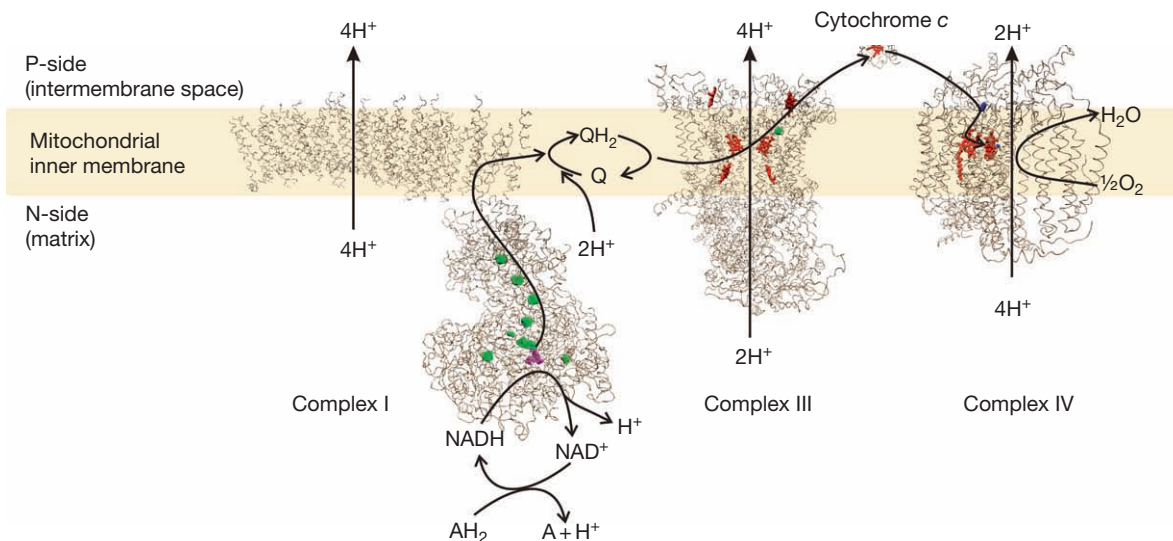


Figure 1 A schematic representation of the three proton-translocating enzyme complexes forming the respiratory chain. The flow of electrons through the complexes as well as the translocation of protons across the membrane is represented by black arrows. The structures used are as follows: complex I from *Thermus thermophilus* (PDB ID 3M9S); complex III (dimeric) from bovine heart (PDB ID 1BGY); cytochrome *c* from equine heart (PDB ID 1HRC); and complex IV from bovine heart (PDB ID 1OCC). AH_2 and *A* represent the reduced and oxidized forms of an NAD^+ -linked metabolic substrate (malate and oxaloacetate, for example).

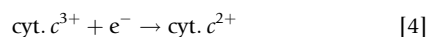
forms), there is hence a redox potential difference ($\Delta E_{m,7}$) of ~ 380 mV between the donor and acceptor ends of complex I. Or put differently, there is a driving force of this magnitude for the electrons to be transported from NADH to ubiquinone, the conversion to a standard Gibbs free energy change ($\Delta G'_0$) being of the form

$$\Delta G'_0 = -nF\Delta E_{m,7} \quad [3]$$

where n is the number of electrons to be transferred and F the Faraday constant, the approximate value of which is $23 \text{ kcal mol}^{-1} \text{ V}^{-1}$.

Ubiquinone may also be reduced by other routes than via complex I, for example, by succinate via the succinate dehydrogenase complex, also called complex II. However, complex II is not a part of the respiratory chain because the reaction catalyzed does not conserve redox energy as proton-motive force. Succinate dehydrogenase is the only enzyme of the Krebs' cycle that is bound to the inner mitochondrial membrane, the reason for which is the location of the hydrogen acceptor, ubiquinone.

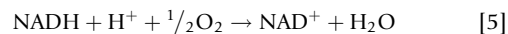
Ubiquinone is a point in the respiratory chain into which hydrogen is transferred from different branches of catabolism, as is the case for NAD^+/NADH . In animal mitochondria reduced ubiquinone is oxidized further by a single pathway, via the respiratory chain complexes III and IV (see below), whereas alternative routes of ubiquinol oxidation are found in plant mitochondria and in many aerobic bacteria. Complex III is a ubiquinol-ferricytochrome *c* oxidoreductase, also known as the cytochrome bc_1 complex. Here, the water-soluble hemoprotein ferricytochrome *c* is the acceptor, this time a pure electron acceptor



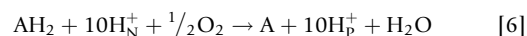
so that oxidation of the hydrogen in ubiquinol must result in net release of protons. The $E_{m,7}$ of cytochrome *c* is $c. 220$ mV when associated with the mitochondrial membrane. The last enzyme

in the respiratory chain, complex IV or cytochrome *c* oxidase, catalyzes oxidation of cytochrome *c* by molecular oxygen. The $E_{m,7}$ of the $\text{O}_2/\text{H}_2\text{O}$ redox couple is 810 mV under standard conditions (1 atm-O_2 , or $\sim 1.2 \text{ mmol l}^{-1}$ in pure water at 25°C).

The chemical reaction catalyzed by the mitochondrial respiratory chain



thus covers a standard redox potential span of $c. 1.15$ V from NADH to O_2 . The actual redox potential span (ΔE_h) is offset from $\Delta E_{m,7}$ by the concentration terms of the oxidized and reduced forms of reductant and oxidant and constitutes the driving force for proton translocation across the membrane, and thereby generation of an electrochemical proton gradient (i.e., protonmotive force). The stoichiometry of proton translocation in the respiratory chain is 10 H^+ per two redox equivalents, that is,



where AH_2 and *A* are again the reduced and oxidized forms of an NAD^+ -linked substrate, respectively. The maximum protonmotive force created by the mitochondrial respiratory chain is $c. 230$ mV, under which conditions the rate of respiration comes to a minimum and ATP synthesis stops. With a stoichiometric proton pumping efficiency of $5\text{H}^+/\text{e}^-$ and $5 \times 230 \text{ mV} = 1.15$ V, it follows that this is a state close to thermodynamic equilibrium between the driving and the driven reactions.

Redox Loop versus Proton Pump

In the respiratory chain, electrons, ultimately used to reduce molecular oxygen, travel through a series of redox centers with increasing electron affinities, while energy liberated during the

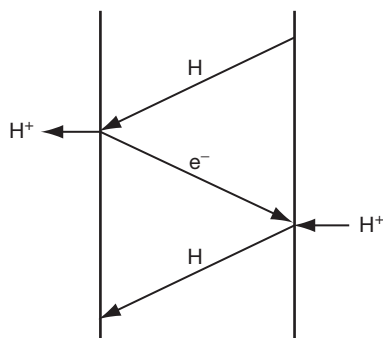


Figure 2 Schematic representation of the redox-loop mechanism. Sequential movement of neutral hydrogen atoms and charged electrons in opposite directions across the membrane results, in effect, in translocation of protons.

process is utilized for translocation of protons across the membrane from the mitochondrial matrix (negatively charged N-side) to the intermembrane space (positively charged P-side) (Figure 1). Peter Mitchell's chemiosmotic theory was based on the postulate of orientated redox chemistry in the respiratory chain in such a way that the neutral element of hydrogen is translocated from the N- to the P-side of the membrane, whereas the negatively charged electrons are translocated in the opposite direction, from P to N (Figure 2). In such a system, there is an alternate involvement of hydrogen and electron carriers as redox-active elements, of which ubiquinone and cytochrome *c* are examples, respectively. It is important to note here that in such a redox-loop arrangement of the respiratory chain, there is actually no proton translocation in the microscopic/mechanistic sense, even though the overall effect will be a transfer of hydrogen ions across the membrane. Instead, the body of the proton is transferred toward the P-side as the electrically neutral hydrogen atom (i.e., in conjunction with an electron), and the translocation of electrical charge is due to transfer of the electron in the opposite direction.

By contrast, in a redox-driven proton pump the chemical reaction is coupled to and drives true proton translocation across the membrane. Hence, the principle and the mechanistic requirements are very different from those of a redox loop. As it has turned out, only the protonmotive principle of the cytochrome *bc₁* complex is of the redox-loop type, employing a fairly complicated cyclic mechanism (see below). On the other hand, both complexes I and IV are true redox-linked proton pumps even though some redox-loop features are nevertheless present in the latter case.

Complex I or NADH: Ubiquinone Oxidoreductase

It has been established that two protons are translocated across the membrane per electron transferred from NADH to ubiquinone, when catalyzed by complex I:



We specify that reduction of NAD^+ by hydrogenated substrates in the mitochondrial matrix is followed by reduction of complex I by NADH, where subsequent to reduction of flavin

mononucleotide (FMN; the first acceptor of reducing equivalents in complex I; purple structure in Figure 1), the reducing equivalents are accepted by iron-sulfur (Fe/S) centers (green structures; Figure 1). The Fe/S centers act as pure electron carriers, so that ultimately the two protons released on the N-side of the membrane during oxidation of AH_2 via NAD^+ are consumed when ubiquinone is reduced (Figure 1).

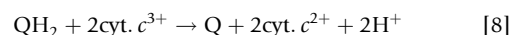
These protonations and deprotonations should not be confused with the translocation of two protons per electron across the membrane (or 4H^+ translocated/NADH oxidized), which is the proton-pumping activity of complex I and its contribution to generation of protonmotive force (Figure 1).

As has been revealed by electron microscopy and recent X-ray structures, complex I is L-shaped consisting of a membrane-intrinsic arm and an extrinsic domain attached to it on the N-side of the membrane (Figure 1). Moreover, it has been established that all known redox centers are located in the extrinsic domain, for example, the FMN cofactor and an array of 8 (in some species 9) Fe/S clusters. A binding site for ubiquinone has been located near the interface between the extrinsic and intrinsic domains.

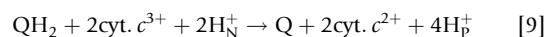
The membrane domain consists of a number of conserved subunits, three of which are homologous to Na^+/H^+ antiporters and hence likely to be involved in the proton-pumping mechanism. Very recent X-ray structures of the membrane domain show a remarkably long α -helix positioned parallel to the membrane, which has been proposed to act as a piston-like element in the redox-linked proton pump mechanism. Such a mechanic element is of interest also, because it is reminiscent of the mechanical coupling between proton translocation and ATP synthesis in the F_1F_0 ATP synthase, the enzyme responsible for phosphorylation of adenosine diphosphate (ADP) to form ATP. However, to date there is no direct evidence for such a function in complex I, and the mechanism of generation of protonmotive force is, in this case, far less elucidated than the corresponding mechanisms for complexes III and IV.

Complex III or Ubiquinol:Ferricytochrome *c* Oxidoreductase

Complex III is unique in the respiratory chain in that it switches from a hydrogen carrier as the donor (ubiquinol) to an electron carrier as the acceptor (ferricytochrome *c*), which means that the overall process must be associated with net release of $2\text{H}^+/2\text{e}^-$:



However, the actual reaction catalyzed is far more complicated than implied by eqn [8], which applies only in solution. Especially, it should be noted that the substrate protons of eqn [8] have no protonmotive function, because they are released from mitochondria into the strongly buffered intermembrane space (from bacteria into the outside world), and do not significantly contribute to either the ΔpH or membrane potential terms of the protonmotive force. The true reaction of complex III in the membrane can be described as



where two of the four protons released on the P-side are the substrate protons of eqn [8]. The question then arises as to how two protons are translocated across the membrane linked to the two-electron oxidation of ubiquinol by ferricytochrome *c* (eqn [9]). The solution to this problem is basically understood today and is a key part of the history of the field of molecular bioenergetics.

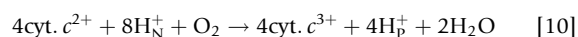
Complex III is characterized by having two ubiquinone binding sites, Q_o near the P-side of the membrane and Q_i near the N-side (Figure 3). Two *b*-type heme groups, b_L and b_H (for low and high redox potential, respectively), connect the sites from an electron-transfer point of view. A high-potential iron-sulfur center, the Rieske Fe/S site, has alternate positions, either near site Q_o or close to the acceptor cytochrome c_1 .

The two-electron donor ubiquinol (QH_2) is oxidized at the Q_o site by a unique mechanism in which the path of the two electrons is bifurcated; one electron is donated to the Fe/S center and the other to heme b_L in the vicinity. Simultaneously, the two protons of QH_2 are released to the P-side of the membrane. The low-potential heme b_L donates the electron to the high-potential heme b_H on the N-side and equilibrates with bound ubiquinone in the Q_i site next to it. The full catalytic cycle requires oxidation of a second molecule of QH_2 in the same manner as the first, whereby arrival of the second electron to the Q_i site leads to uptake of two protons from the N-side yielding QH_2 as the product. This so-called Q-cycle mechanism, which is summarized by eqn [9] (see Figure 3), has been well established for quite some time, and is fully consistent with the X-ray diffraction structures of complex III that were solved much later.

It is important to note that complex III is not a proton pump in the mechanistic sense, but a Mitchellian redox loop (see Figure 2), where QH_2 carries hydrogen from the N- to the P-side of the membrane, electrons are carried in the opposite direction via the *b*-type hemes, protons are released on the P-side when QH_2 is oxidized by pure electron acceptors, and taken up on the N-side when Q is reduced by the electron-donating heme b_H . From a bioenergetic viewpoint, it is also important that the effective stoichiometry of proton translocation is $1H^+/e^-$ (as it must be in a redox loop), even though $2H^+/e^-$ are released on the P-side of the membrane (see above).

Complex IV or Cytochrome *c* Oxidase

Contrary to complex III, the reaction catalyzed by complex IV uses an electron carrier as a donor of reducing equivalents (cytochrome *c*), but a hydrogen-accepting entity (O_2) as the acceptor. This means that the overall chemical reaction will consume $2H^+/2e^-$. In addition, the energy released in this reaction is used to translocate one proton across the membrane for every electron (the proton-pumping function):



The active site of complex IV consists of a binuclear heme-copper center (heme a_3/Cu_B) which binds O_2 in its ferrous/cuprous state in much the same general fashion as dioxygen is bound to heme iron in oxyhemoglobin and oxymyoglobin. Electrons are donated to this center, one at a time, from

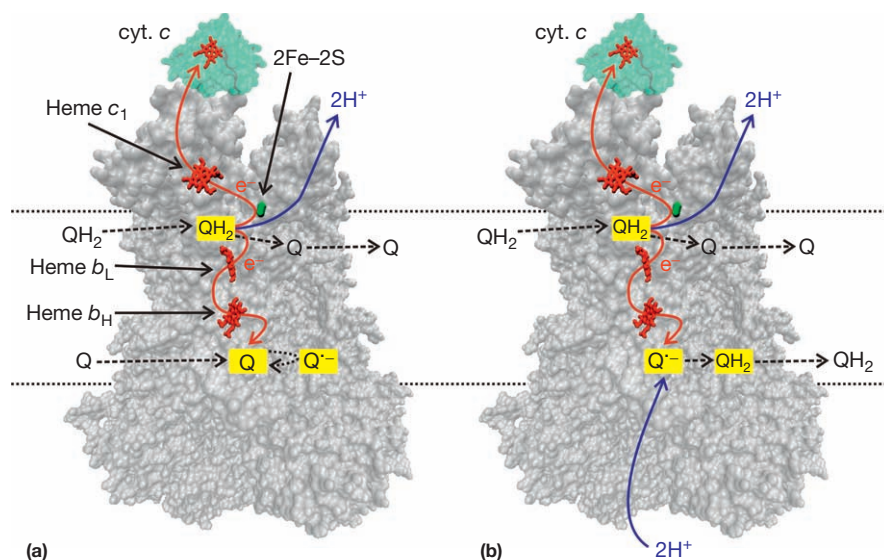


Figure 3 The Q-cycle mechanism. The structure of complex III is depicted in gray while the soluble cytochrome *c* is shown in green. Prosthetic groups of complex III are labeled and marked with black arrows. The binding sites for the quinone molecules are marked by yellow boxes; the Q_o -site resides close to the P-side of the membrane (above) while the Q_i -site is located near the N-side (below). The dashed arrows represent the movements of the quinone molecules. The flow of electrons is shown by red arrows, and the protons involved in the mechanism are represented by blue arrows. (a) A quinol molecule enters the Q_o -site while a quinone molecule occupies the Q_i -site. Oxidation of the quinol in the Q_o -site results in the release of two electrons, one of which is used to reduce quinone in the Q_i -site to semiquinone while the other travels via the Rieske Fe-S center and heme c_1 to the soluble cytochrome *c*. The two protons liberated in the process are released on the P-side of the membrane. (b) In the next step, the entrance of the second quinol molecule to the Q_o -site leads to the same reaction sequence which results in reduction of the semiquinone to quinol in the Q_i -site with uptake of two protons from the N-side of the membrane.

another heme (heme *a*) that lies in the vicinity, and which receives electrons from a bimetallic copper center, Cu_A , that lies at the P-side interface. Cu_A in turn, receives electrons from the donor, soluble cytochrome *c* (Figure 1).

It may be noted as a curiosity that complexes III and IV function together as one complete Mitchellian redox loop: oxidation of ubiquinol by ferricytochrome *c* catalyzed by complex III releases $1\text{H}^+/\text{e}^-$ on the P-side of the membrane (see above). The resulting electron is then transferred across part of the membrane to the heme a_3/Cu_B site of complex IV, where it reduces $1/4 \text{O}_2$ to $1/2 \text{H}_2\text{O}$ with the aid of one proton taken from the N-side. However, the Q-cycle mechanism in complex III ensures protonmotive activity of this complex by itself (see above), and complex IV adds a proton pumping activity that doubles its protonmotive efficiency so that it is $2\text{H}^+/\text{e}^-$, that is, twice that of complex III. Note that $2\text{H}^+/\text{e}^-$ is the effective protonmotive efficiency of complex IV even though only $1\text{H}^+/\text{e}^-$ is released on the P-side.

The mechanism by which protons are pumped by complex IV is electrostatic in nature. An electron in the heme *a* domain increases the proton affinity (pK_a) of a proton loading site (or pump site), which has not been definitely identified but has been suggested to be the A-ring propionate of heme a_3 . The increased pK_a leads to uptake of a proton from the N-side into that site, uptake from the P-side being too slow to compete. Loading the pump site with a proton increases the electron affinity ($E_{m,7}$) of the metals in the binuclear site, which results in transfer of the electron to that site, which further increases the pK_a of the pump site. However, electron transfer to the binuclear site causes an even larger increase of the pK_a of the binuclear center, which leads to uptake of a second proton to that site from the N-side of the membrane. The charge of the electron at the binuclear site is thus neutralized by the second proton, which is one of the four reaction steps in which O_2 is

reduced to two molecules of water at this location. Electrical neutralization of the binuclear site results in lowering of the initial proton affinity of the pump site, with the consequence that the proton in the latter is released by electrostatic repulsion to the P-side of the membrane. If it were released to the N-side, the pump mechanism would be compromised, and this is prevented by a gating function in which the position of the electron (at heme *a* or at the binuclear site) controls the kinetic barrier of proton transfer between the N-side of the membrane and the pump site.

See also: **Bioenergetics:** Bioenergetics: General Definition of Principles; Chemiosmotic Theory; F-ATPases; Membrane-Associated Energy Transduction in Bacteria and Archaea; Mitochondrial Membranes, Structural Organization; Respiratory Chain and Adenosine Triphosphate Synthase; **Metabolism Vitamins and Hormones:** Tricarboxylic Acid Cycle.

Further Reading

- Cramer WA and Soriano GM (2003) Thermodynamics of energy transduction in biological membranes. In: *Biophysics Textbook Online*. <http://www.biophysics.org/Portals/1/PDFs/Education/cramer.pdf>. (accessed September 2012).
- Efremov RG, Baradaran R, and Sazanov LA (2010) The architecture of respiratory complex I. *Nature* 465: 441–445.
- Mitchell P (1966) Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. *Biological Reviews of the Cambridge Philosophical Society* 41: 445–502.
- Nicholls DG and Ferguson SJ (2002) *Bioenergetics 3*. London: Academic Press.
- Trumpower BL (1990) The protonmotive Q cycle. *Journal of Biological Chemistry* 265: 11409–11412.
- Wikström M and Verkhovsky MI (2007) Mechanism and energetics of proton translocation by the respiratory heme-copper oxidases. *Biochimica et Biophysica Acta* 1767: 1200–1214.

Pteridines

S C Daubner, Texas A&M University, College Station, TX, USA

P F Fitzpatrick, University of Texas Health Science Center, San Antonio, TX, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Aromatic amino acid Aromatic compounds are cyclic compounds containing double bonds so that their three-dimensional structure is planar. Amino acids are small molecules, which consist of a central carbon atom, tetrahedrally bonded to a carboxylic acid group, an amino group, a hydrogen, and a fourth group, often called an R group; in the aromatic amino acids, it is this R group which is aromatic.

Heme A large planar heterocyclic compound consisting of four bridged pyrrole rings which binds

Fe in biological systems for the transport or reactivity of oxygen.

Heterocycle A ring-shaped compound made up of more than one kind of atom.

Pyrimidine and pyrazine Organic heterocyclic compounds consisting of a six-membered ring with two nitrogen atoms separated by one carbon in pyrimidine and two carbons in pyrazine.

Radical An atom or group of atoms possessing an odd (unpaired) electron.

The pteridines are a group of heterocyclic compounds containing a wide variety of substitutions on the basic compound pterin. Pterin itself is composed of a pyrazine ring and a pyrimidine ring; the pyrimidine ring has a carbonyl oxygen and an amino group. The biosynthesis of pterins begins with the molecule guanosine triphosphate (GTP); the enzyme which controls the conversion of GTP to pterin, GTP cyclohydrolase, is found in both prokaryotes and eukaryotes. Pterins were first discovered in the pigments of butterfly wings and they perform many roles in coloration in the biological world. Pterins also function as cofactors in enzymatic catalysis. Folates, 'conjugated' pterins which contain *p*-amino benzoic acid and glutamates in addition to the pterin ring system, are critical compounds in methyl transfer biochemistry. Tetrahydrobiopterin, the major unconjugated pterin in vertebrates, is involved in the hydroxylation of aromatic compounds and synthesis of nitric oxide. Molybdopterin – which contains pterin, a third heterocyclic ring containing an ether oxygen and two thiol substitutes, and molybdenum – is involved in biological hydroxylations, reduction of nitrate, and respiratory oxidation.

The name pteridine refers to the heterocyclic aromatic compound shown in [Figure 1](#), composed of a pyrimidine ring fused to a pyrazine ring. However, the word pteridines has come to stand for an entire group of naturally occurring compounds. Their structures are based upon pteridine, but they have substituents on the pyrimidine ring: an amino group or a carbonyl at position 2 and a carbonyl group at position 4 (position numbers as shown in [Figure 1](#)). Pterin contains the amino substituent at position 2; all pterins are derivatives of this molecule. Most of the naturally occurring pteridines are pterins. The first pterins to be studied, leucopterin and xanthopterin, were pigments isolated from butterfly wings, by Sir Frederick Gowland Hopkins at Cambridge University in the late nineteenth century. Their names are derived from the Greek words for wing, *pteron*, white, *leukos*, and yellow, *xanthos*.

Pterins are referred to as being conjugated or unconjugated. The conjugated pterins are folates, derivatives of folic acid. They contain a *p*-aminobenzoate substituent at the 6-position and one or more glutamate residues. The folates are discussed

elsewhere in this encyclopedia. As a general but not fast rule, folates are involved in one-carbon transfers in metabolism and pterins are involved in oxygenation reactions or in pigmentation.

Distribution of Pterins in Nature

Pterins are found in all living organisms, from bacteria, blue-green algae, and trypanosomes, to mammals; they have also been detected in the chloroplasts of spinach leaves and the cotyledons of pea sprouts. A wide variety of pterins exist in bacteria, such as the various pterins with three-carbon side chains of *Escherichia coli*, the glucuronidated pterins of *Bacillus subtilis*, the three ribityl pteridines of the marine bioluminescent bacterium *Photobacterium phosphoreum*, the molybdopterins of photosynthetic bacteria, and the tetrahydrodromethanopterins of methanogenic bacteria.

As pigments, pterins are not only found in butterfly wings, but also in moth heads, silkworm skin, grasshopper bodies, and honeybee larvae. A list of examples demonstrates the very wide use nature has made of pterins in coloration; diversity is found in species distribution, chemical structures, and hue. Unusual dimeric pterins are found in the eyes of the fruit fly, *Drosophila*. A blue pigment in the roundworm *Ascaris lumbricoides* contains biopterin. Pterins are common in the skins of amphibians and fish, and photolabile pterins are found in some amphibians' eyes. The red color of some salmon, the Siamese fighting fish, and the red lizard *Anolis*, is due to pterins, as is the green color of some snakes.

Clearly, pterins serve widely in coloration, but they are also critical compounds in metabolic pathways. In the vertebrates, the major pterin is tetrahydrobiopterin, which serves as a redox cofactor in enzymatic reactions utilizing molecular oxygen.

Biosynthesis of Neopterins and Biopterin

The synthesis of all pterins begins with the ring opening and reclosing of the nucleotide GTP by the enzyme GTP

cyclohydrolase (GTPCH). GTPCH has been isolated and sequenced from many sources. It is a decamer, each active site containing a zinc atom. As the rate-limiting enzyme of pterin biosynthesis, it is carefully regulated, both by transcriptional and posttranslational mechanisms. The catalytic domains of the *E. coli* and human enzyme are 37% identical in amino acid sequence, testifying to the ancient status of pterins in the living world.

The reaction catalyzed by GTPCH is shown in Figure 2. The purine ring system of guanine supplies all but two of the carbons in the ring system of pterin. The ribose moiety of GTP supplies the two new carbons for the pyrazine ring and the carbons of the erythrose side chain; one carbon from the guanine ring is released as formate. The pterin product is dihydroneopterin triphosphate. In eukaryotes, dihydroneopterin triphosphate is converted to dihydrobiopterin

(Figure 3). By contrast, dihydrobiopterin is not formed in most bacteria.

The structures of biopterin and its reduced forms are shown in Figure 3. Reduction of the 7,8 double bond yields 7,8-dihydrobiopterin (a tautomer of 7,8-dihydropterin is quinonoid dihydropterin). Further reduction at the 5,6 double bond gives tetrahydrobiopterin.

Pterins in Metabolism

The Aromatic Amino Acid Hydroxylases

One of the most-studied processes involving biopterin is the hydroxylation of the aromatic amino acids. This is catalyzed by a small family of enzymes containing phenylalanine (PheH), tyrosine (TyrH), and tryptophan (TrpH) hydroxylase. They contain an iron atom and utilize tetrahydrobiopterin and molecular oxygen to hydroxylate a fairly unreactive entity, an aromatic ring. The aromatic amino acid hydroxylases (AAHs) are only active in the Fe(II) state; one role of the tetrahydrobiopterin may be to keep the iron reduced. The major role of tetrahydrobiopterin is to activate the oxygen for eventual transfer to the aromatic amino acid substrate, probably through formation of a C4a-peroxy pterin intermediate (Figure 4).

Each of the AAHs catalyzes the rate-limiting step in its metabolic pathway: PheH in phenylalanine catabolism, TyrH in synthesis of the catecholamine neurotransmitters, and TrpH in serotonin synthesis. The three enzymes are homologous to a great extent in their catalytic domains (between any two there is 60% identity and 75% similarity). These enzymes as a rule are not found in bacteria, consistent with the general absence of biopterin from prokaryotes. However, PheH has been isolated from several prokaryotes, and from one, *Chromobacterium violaceum*, the sequence is 30% identical to the catalytic domain of human PheH. Two amino acid residues which bind

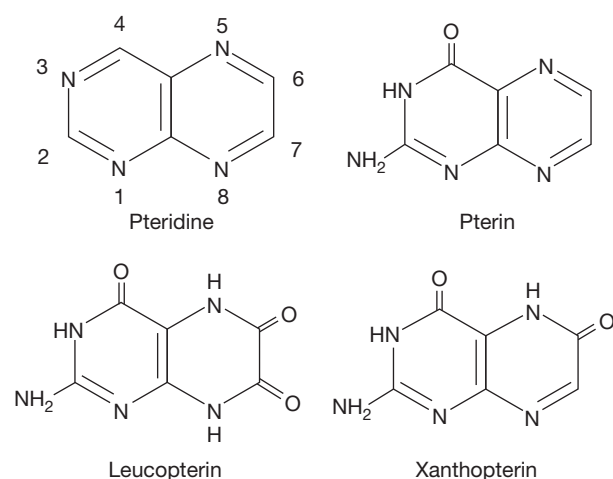


Figure 1 Structures of pteridine and three pteridines.

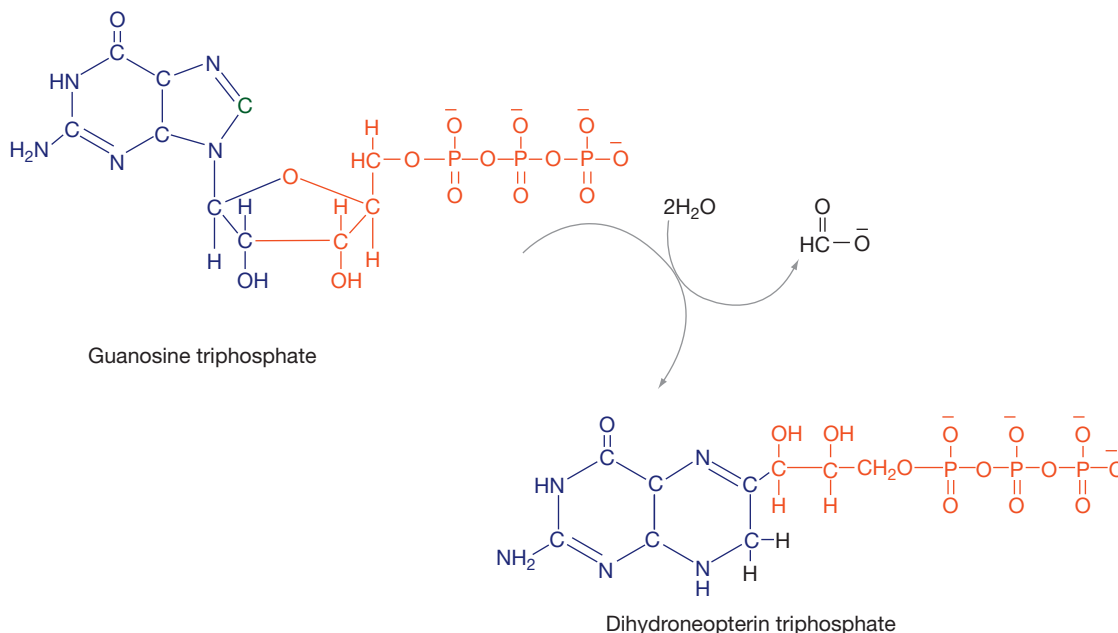


Figure 2 The first step in the synthesis of pterins from guanosine triphosphate. The atoms of GTP are color coded to show their positions in the resulting neopterin molecule. Note that one carbon of GTP, colored green, is lost as formate.

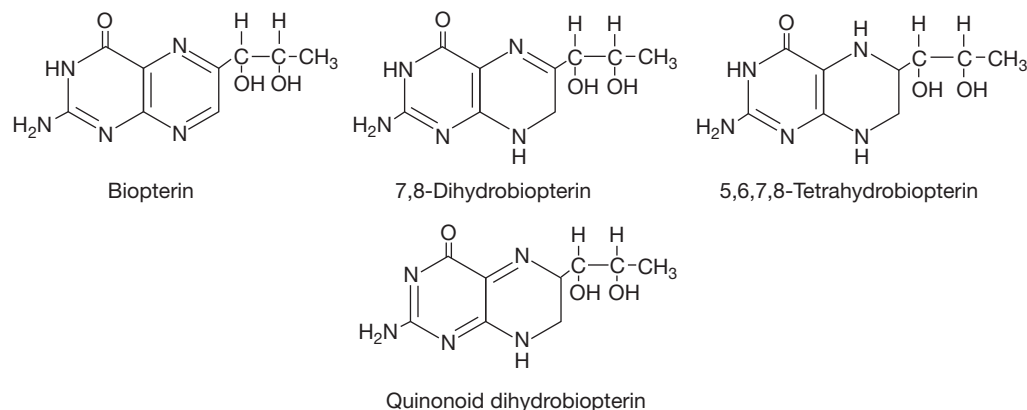


Figure 3 Structures of bioppterin and its reduced forms.

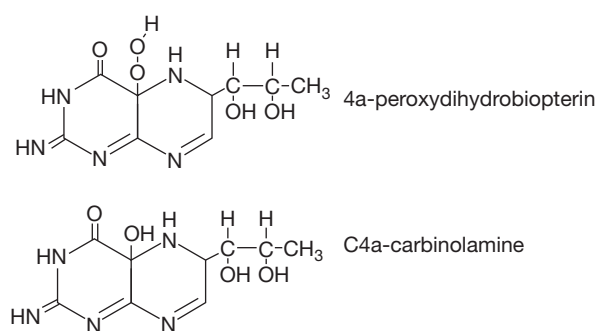


Figure 4 Peroxypterin and carbinolamine intermediates in the hydroxylation of aromatic amino acids

tetrahydrobioppterin (in TyrH, F300 and E332) are conserved in all the AAHs studied, suggesting a central role for bioppterin in the evolution of neurotransmitters and higher nervous systems.

The pterin product of the AAHs is C4a-carbinolamine dihydrobioppterin (Figure 4). It is released from the enzyme and is returned to the state needed for amino acid hydroxylation by the action of two more enzymes, C4a-carbinolamine dehydratase and dihydropterin reductase.

Nitric Oxide Synthase

Nitric oxide synthase (NOS) is responsible for the synthesis of nitric oxide, NO. NOS utilizes tetrahydrobioppterin, NADPH, and molecular oxygen to convert L-arginine to L-citrulline and NO (Figure 5). Nitric oxide is a gas and a highly toxic compound. Due to its toxicity, short half-life (1–5 s), and diffusibility, NO serves the mammalian immune system as a killer of foreign intruders and cancer cells. It is also a cellular signal through its stimulation of cGMP synthesis.

NOSs are dimers of identical subunits. NOS contains flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD), heme, and tetrahydrobioppterin. The tetrahydrobioppterin-binding site is positioned close to the heme to facilitate electronic interaction between them. The tetrahydrobioppterin-binding site is formed by residues in both subunits at the

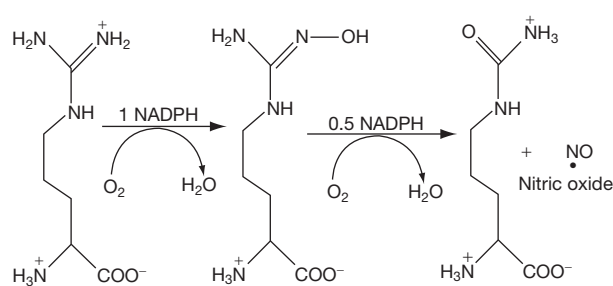


Figure 5 Reaction catalyzed by NOS.

interface where the subunits associate; the pterin ring is sandwiched between a tryptophan residue from one subunit and a phenylalanine from the other. In contrast to the two-electron loss in AAHs, the tetrahydrobioppterin of NOS loses only one electron generating a bioppterin radical. Bioppterin remains bound to NOS before and after catalysis, another difference between NOS and the AAHs.

Molybdopterin-Containing Enzymes

A very large group of enzymes contain molybdenum (or in a few cases tungsten) in conjunction with a unique pterin, molybdopterin. These enzymes function in general to transfer oxygen to or from a physiological molecule. They can be assigned to two categories based on the reaction catalyzed. The first group, the true hydroxylases, performs hydroxylation of aldehydes or aromatic rings; some of the enzymes in this category are milk xanthine oxidase, liver aldehyde oxidase, and bacterial enzymes such as formate dehydrogenase and nicotine dehydrogenase. The second category can be subdivided into (1) sulfite oxidases and enzymes, which reduce nitrate to enable a cell to utilize ammonia, and (2) bacterial enzymes, which function as terminal respiratory oxidases.

The structures of molybdopterin and the molybdenum cofactor are shown in Figure 6. The R group in prokaryotic systems can be guanine, cytosine, adenine, or hypoxanthine; in eukaryotes it is a hydrogen. The mode of binding the molybdenum cofactor to the protein differs from protein to protein in accord with the classification described above. The enzymes of

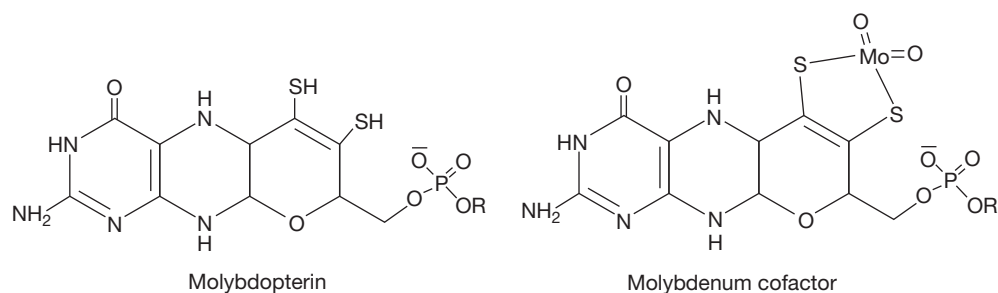


Figure 6 The structures of molybdopterin and the molybdenum cofactor.

the xanthine oxidase group contain one single molybdenum center per enzyme subunit, the sulfite oxidase family members are likely to have one molybdenum cofactor bound to a cysteine of the protein, and the final group can contain two molybdopterin coordinated to one molybdenum, which can be coordinated to a selenium, sulfur, or cysteine. The biosynthetic pathways of these pterins have not yet been resolved.

The molybdenum enzymes are a vast enough group that a discussion of their properties is inappropriate here. In general, they contain other redox centers in addition to their pterin cofactor, such as heme, iron–sulfur clusters, and flavin. The role of the pterin is postulated to include mediation of electron transfer to the other redox active centers and modulation of the reduction potential of the molybdenum. The pterin is not considered to be oxidized and reduced during catalysis, a distinct difference between these enzymes and the NOSs and AHHs.

Tetrahydrobiopterin and Human Disease

Tetrahydrobiopterin deficiency is associated with a rare variant of hyperphenylalaninemia, which cannot be treated by the low phenylalanine diet successful for typical phenylketonuria (PKU) patients. It is characterized by a deficit of catecholamine neurotransmitters and serotonin as well (the products of the AAHs), affecting the nervous system as well as the liver. It can

be caused by mutations in GTPCH or in the enzymes which recycle dihydrobiopterin. Decreased levels of tetrahydrobiopterin in cerebrospinal fluid have also been noted in diseases such as Parkinson's, autism, depression, and Alzheimer's.

See also: **Bioenergetics:** Calcium Signaling; NO Synthase; **Metabolism Vitamins and Hormones:** Amino Acid Metabolism; **Protein/Enzyme Structure Function and Degradation:** Flavins; Heme Proteins.

Further Reading

- Brown GM (1985) Biosynthesis of pterins. In: Blakley RL and Benkovic SJ (eds.) *Folates and Pterins, Chemistry and Biochemistry of Pterins*, vol. 2, pp. 115–154. New York, NY: Wiley.
- Fitzpatrick PF (2000) The aromatic amino acid hydroxylases. *Advances in Enzymology and Related Areas of Molecular Biology* 74: 235–294.
- Hille R (1996) The mononuclear molybdenum enzymes. *Chemical Reviews* 96: 2757–2816.
- Nixon JC (1985) Naturally occurring pterins. In: Blakley RL and Benkovic SJ (eds.) *Folates and Pterins, Chemistry and Biochemistry of Pterins*, vol. 2, pp. 1–42. New York, NY: Wiley.
- Rebel J, Auerbach G, Bader G, et al. (2003) Biosynthesis of pteridines. Reaction mechanism of GTP cyclohydrolase I. *Journal of Molecular Biology* 326: 503–516.
- Thony B, Auerbach G, and Blau N (2000) Tetrahydrobiopterin biosynthesis, regeneration and functions. *Biochemical Journal* 347: 1–16.
- Wei C-C, Crane BR, and Stuehr DJ (2003) Tetrahydrobiopterin radical enzymology. *Chemical Reviews* 103: 2365–2383.

P-Type Pumps: Copper Pump

Y Hatori and S Lutsenko, Johns Hopkins University, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Ilia Voskoboinik and James Camakaris, volume 3, pp. 561–564, © 2004, Elsevier Inc.

Glossary

Copper chaperone Soluble cytosolic protein that binds copper and transfer copper to other protein targets.

Cuproenzyme Enzyme that uses copper as a cofactor to perform its catalytic reactions.

Phosphointermediate Intermediate state of enzyme with a covalently bound phosphate.

P-type adenosine triphosphatases (ATPases) A class of membrane proteins that use the energy of adenosine

triphosphate (ATP) hydrolysis to transport ions across biologic membranes; unlike other ATP-driven transporters, the P-type ATPases form transient phosphointermediate.

Trafficking The process of protein changing its intracellular localization between intracellular compartments.

trans-Golgi network Terminal portion of the Golgi apparatus.

γ-Phosphate The terminal phosphate group in the chain of three phosphate moieties in ATP.

Biological Roles of Cu Pumps

Cu Pumps Are Universally Found in Biological Systems

The diverse family of P-type adenosine triphosphatases (ATPases) is composed of molecular pumps that translocate specific ions and, in some cases, phospholipids from the cytosol across cellular membranes utilizing the energy of adenosine triphosphate (ATP) hydrolysis. The Cu transporting P-type ATPases, or Cu pumps, form one of the largest phylogenetic clusters in the P-type ATPase family. Cu pumps are present in three kingdoms of life (prokaryotes, archaea, and eukaryotes), illustrating the essentiality of active Cu transport for life. In biological systems, the redox chemistry of Cu ions ($\text{Cu}^+ \leftrightarrow \text{Cu}^{2+}$) has found its productive use in electron transfer enzymes. However, the same chemistry could trigger nonspecific oxidation and radical-mediated damage. Cells and tissues avoid Cu toxicity through coordinated and well-controlled Cu transport processes: sufficient uptake, targeted delivery to specific compartments, and excretion of Cu excess. In the past decade, this homeostatic system has been a subject of intensifying scientific investigations due to accumulating evidence for the role of copper homeostasis in heart development, central nervous system (CNS) function, and tumor growth. These studies have firmly established that Cu pumps play a central role in the maintenance and regulation of Cu homeostasis.

Cu Pumps Are Essential for Human Health

Human cells express two homologous Cu pumps, ATP7A and ATP7B. ATP7A facilitates transports of Cu ions from enterocytes into a blood stream for further distribution to tissues; this represents an essential step in Cu absorption. Excess copper is removed by the liver where another Cu pump, ATP7B, facilitates the export of Cu into the bile. In both cases, Cu exits cells as a result of exocytosis of vesicles, which are filled with Cu by Cu pumps. Another important role of ATP7A and ATP7B is to deliver Cu ions to specific enzymes produced in the secretory pathway. ATP7A delivers Cu to copper-requiring enzymes such as peptidyl- α -monooxygenase, tyrosinase, lysyl oxidase,

and extracellular superoxide dismutase, whereas ATP7B delivers Cu to ceruloplasmin. The multiple roles of Cu pumps in Cu absorption, delivery to the secretory pathway, and excretion require trafficking between intracellular compartments (Figure 1). It is now well established that human Cu pumps respond to the increased Cu levels by relocating from the terminal portion of the Golgi apparatus – the *trans*-Golgi network (TGN) toward the vesicles in the vicinity of plasma membrane. Although ATP7A and ATP7B are highly homologous (having 54% primary sequence identity), their intracellular itineraries differ. For example, ATP7A in enterocytes relocates toward the basolateral membrane (the blood side), while hepatic ATP7B traffics toward the apical membrane to facilitate Cu efflux into the bile. In cells expressing both Cu pumps, they show distinct trafficking behavior (different response to copper and/or different polarity of movement), indicating that Cu pumps have specific structural elements regulating their trafficking behaviors.

Information on specific physiologic roles of ATP7A and ATP7B in tissues and cells expressing both Cu pumps is limited. Recent studies in the placenta and mammary gland indicate that in these tissues the expression and intracellular localization of Cu pumps are under hormonal control. Exposure to hormones such as estrogen and insulin stimulates trafficking of Cu pumps; the trafficking step is necessary to facilitate Cu delivery to the fetus or to return Cu to maternal circulation. By contrast, in kidneys and lungs, only one Cu pump traffics (ATP7A, and ATP7B, respectively) even though both transporters are expressed. These observations suggest that in some tissues Cu pumps may participate in copper storage by sequestering copper within the endocytic pathways rather than inducing copper export.

The human Cu pumps also play a role in cellular response to pathological conditions such as hypoxia and inflammation. Under these conditions, the cellular pathways associated with Cu pumps appear selectively activated resulting in a preferential flow of copper toward the secretory pathway. The critical role of Cu pumps in human health is best illustrated by the phenotypes of Menkes disease (MND) and Wilson's disease

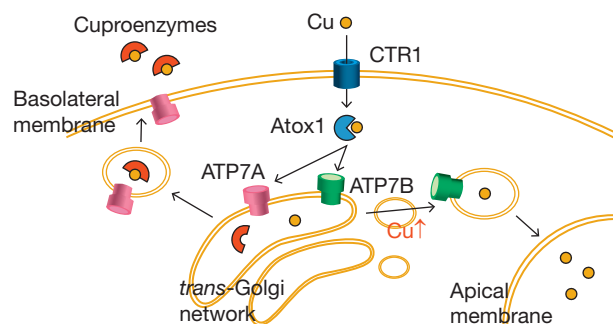


Figure 1 Cu pumps play a central role in copper homeostasis in human cells. Cu ions are transported into a cell through CTR1 and then delivered by a cytosolic Cu chaperone Atox1 to Cu pumps (ATP7A and ATP7B). Cu pumps located within the *trans*-Golgi network transfer copper from the cytosol to cuproenzymes located in the lumen of the TGN. When cellular Cu is elevated, ATP7B traffics from the TGN to a vesicular compartment in proximity to the apical membrane. By contrast, ATP7A traffics to vesicles at basolateral membrane. Cu is exported from cells as a result of vesicle fusion with the plasma membrane.

(WND) which are associated with inborn mutations in *ATP7A* and *ATP7B* genes, respectively. Defects in the Cu pumps lead to copper misbalance in multiple organs, causing systemic effects with the major impact on the CNS (MND) and/or liver (WND). The MND patients display severe neurological symptoms, mental retardation, vascular defects, and other developmental problems. These pathologies reflect the low activity of Cu-dependent enzymes caused by the impaired absorption of Cu through intestine and choroid plexus. Symptoms can be lessened by Cu injections into the blood, particularly if MND is identified at the presymptomatic stage (in newborns) and if the mutant *ATP7A* retains residual copper-transport activity. Mutations that completely inactivate *ATP7A* invariably lead to patient's death.

In contrast to MND, the WND symptoms are caused by abnormal accumulation of Cu, predominantly in the liver. In the CNS, the major damage is observed in the basal ganglia. The majority of WND patients present with either hepatic or neuro-psychiatric symptoms, suggesting important functions for *ATP7B* in corresponding organs. The liver is particularly affected reflecting the dominant role of *ATP7B* as a Cu exporter in this tissue. In the liver, *ATP7B* also delivers Cu ions to ceruloplasmin, an enzyme involved in iron homeostasis. Defect in this process results in the excretion of apo-ceruloplasmin, which is inactive and easily degraded in the serum (The low ceruloplasmin level serves a diagnostic marker for WND). The major WND symptoms can be greatly diminished by reducing Cu level in the body using Cu chelators and/or Zn supplementation as Zn interferes with Cu absorption.

Recent clinical studies have uncovered a new role for Cu pumps in regulating tumor growth and the resistance of tumors to chemotherapeutic drugs. Higher levels of *ATP7A* and particularly *ATP7B* expression have been detected in solid tumors that show increased resistance to cisplatin and other platinum-containing drugs. Although the relationship between Cu pumps and drug resistance seems firmly established, the precise mechanism through which Cu pumps modulate sensitivity to chemotherapeutics remains

uncertain. It was shown that *ATP7B* directly interacted with cisplatin *in vitro* and in cells, however, binding alone does not fully account for the protective effects of Cu pumps against the drug. Similarly, cisplatin can modulate the activity of *ATP7B*. However, the transport of cisplatin by Cu pumps is inefficient, and in the presence of copper the competition between these two potential substrates is not observed. Recent studies in cells and tissues deficient in Cu pumps activity indicate that the changes in copper metabolism associated with Cu-pump inactivation result in distinct transcriptional and metabolic changes (increased proliferation, altered lipid metabolism, and increased expression of other transporters). Therefore, metabolic re-adjustments due to altered levels of Cu pumps may contribute to drug resistance through indirect mechanisms.

Cu Pumps in Prokaryotes, Plants, and Fungi

Plants also utilize Cu pumps in the processes of copper absorption, delivery, and excretion. *Arabidopsis thaliana* has four genes (*AtHMA5*, *AtHMA6*, *AtHMA7*, and *AtHMA8*) encoding Cu pumps. *AtHMA6* and *AtHMA8* are closely related and elegantly collaborate to transport Cu into chloroplasts. *AtHMA6* transports Cu from cytosol across the plastid envelope into stroma where Cu/Zn superoxide dismutase (*AtCSD2*) binds transported Cu; *AtHMA8* transports Cu from stroma into the thylakoid lumen to supply Cu to plastocyanin, an electron-transferring enzyme in the photosystem. Two Cu pumps (*CtaA* and *PacS*) of *Synechocystis*, a cyanobacterium that shares evolutionary ancestry with chloroplasts, play analogous roles. This finding suggests that the double-tiered Cu transport system was established early in the evolution of the photosynthetic system. *AtHMA7* is thought to deliver Cu ions to the secretory pathway for incorporation into the ethylene receptor (*ETR1*). A single Cu atom in an *ETR1* homodimer is required for the high-affinity binding of a hormone ethylene and initiation of signaling cascade. Mutation of *AtHMA7* results in the loss of *ETR1*-dependent ethylene signaling, likely due to the lack of bound copper. Finally, *AtHMA5* is a candidate for Cu excretion, as suggested by a hypersensitivity of plants lacking *AtHMA5* to environmental Cu.

Yeast *Saccharomyces cerevisiae* has a single Cu pump (*Ccc2p*). Similar to human Cu pumps, *Ccc2p* is located in the TGN. It transports Cu to the lumen of the secretory pathway where Cu is incorporated into *Fet3p*, a Cu-containing ferroxidase required for iron absorption. Deletion of *Ccc2* leads to inactivation of *Fet3p* and poor growth under Fe-limiting conditions. The $\Delta Ccc2$ phenotype can be rescued by gene complementation using the Cu-pump complementary DNA (cDNA) from various species. This assay has been a valuable system for evaluation of Cu pumps transport activity. However, *Ccc2p* does not traffic from TGN or contribute to Cu export, indicating the important difference of Cu homeostasis between yeast and mammals. Cu homeostasis varies even among yeast strains. *Candida albicans*, a strain with high resistance to Cu, has a second Cu pump, *CRP1p*. *CRP1p* is expressed at the plasma membrane, excretes Cu ions from the cell, and contributes to high resistance against environmental Cu.

In prokaryotes, Cu pumps together with pumps for other transition metals (Zn, Pb, Cd, Co, and Ni) are the predominant

P-type pumps. *Enterococcus hirae* was the first model organism to study Cu homeostasis. A Cu-responsive operon in this bacterium harbors two genes for Cu pumps (CopA and CopB). CopA is thought to act as a Cu importer because cells lacking CopA cannot grow in the Cu-depleted media, while CopB exports Cu from the cell and is involved in Cu resistance. CopA homologs are present in plasma membrane of many other bacteria and archaea; most of them have Cu-export activities (unlike *E. hirae*) typically utilized to confer resistance to Cu. The homologs from various thermophiles (*Archaeoglobus fulgidus*, *Sulfolobus solfataricus*, *Aquifex aeolicus*, and *Thermotoga maritima*) have been successfully used in recent structural studies (see below).

Molecular Architecture and Functions of Cu Pumps

Characteristics of the P-type ATPases Shared by the Cu Pumps

Cu pumps share the general mechanism and architecture with other P-type ATPases. A typical Cu pump is a membrane protein with the N- and C-termini both oriented toward the cytosol. The α -helical transmembrane (TM) part of Cu pumps is involved in copper translocation, while ATP hydrolysis is mediated by its cytosolic domains (Figure 2). The majority of Cu pumps transport Cu ions in the outward direction (from the cytosol into a lumen or extracellular milieu) and follow the reaction cycle common to all P-type pumps, which is known as 'Post-Albers cycle'. According to this cycle, the pump assumes two major ground conformations, E1 and E2. The intramembrane Cu-binding sites have cytosolic access and high affinity in E1 while they have luminal access and low affinity in E2. Cu ions enter these binding sites from the cytosol in E1, become occluded after ATP hydrolysis and transient auto-phosphorylation (E1P intermediate), and jettisoned into the lumen during a subsequent protein isomerization to E2P. The phosphorylated intermediate is then cleaved through auto-dephosphorylation, generating E2 for the next reaction cycle. The ability of Cu pumps to hydrolyze ATP and form a

phosphorylated intermediate in a copper-dependent manner has been directly demonstrated. Extensive studies of the sarcoplasmic reticulum Ca^{2+} pump (SERCA) revealed a strikingly sophisticated domain movement underlying ATP-driven transport of Ca^{2+} . Recent analysis of Cu pumps has demonstrated similar domain movements that suggest the analogous pumping mechanisms for Cu and other P-type pumps.

Overall structure of Cu pumps is similar to other P-type ATPases, which is apparent from the low-resolution maps generated using the 2-D crystals of bacterial/archaeal Cu pumps (*Aquifex aeolicus* CtrA3 and *A. fulgidus* CopA). Cytosolic loops form four separate domains including the N-terminal metal-binding domain (N-MBD) and three domains common for all P-type pumps: A (actuator) domain, N (nucleotide-binding) domain, and P (phosphorylation) domain. The P and N domains are located between TM6 and 7, forming the catalytic core. The structure of the P domain (determined for CopA) is almost identical in all P-type pumps. It has a typical Rossmann fold with a parallel β -sheet sandwiched by short α -helices. During catalysis, the γ -phosphate of ATP is transferred to the carboxyl group of the invariant Asp located at the end of the central β -strand of the P-domain, producing transient phosphointermediates (i.e., E1P and E2P in the reaction scheme). The functionally important sequence motifs, DKTGTLT (phosphorylation site) and GDGXND (Mg^{2+} coordinating site), are invariant in the P-type pumps. The geometry of these key residues is preserved within haloacid dehalogenase (HAD) superfamily to which P-type ATPases belong, reflecting the importance of this arrangement for nucleophilic reactions catalyzed by the HAD superfamily members. The N domain is inserted in the P domain and consists of an antiparallel β -sheet sandwiched by two pairs of α -helices. This global fold is shared by all P-type pumps despite limited sequence homology. The N domain forms a docking site for the adenosine moiety of ATP. The A domain is located between TM4 and 5 and is composed of antiparallel β -strands with a β -hairpin structure. The A domain has a Gly-Glu motif (an invariant part of a more variable TGES motif), which is involved in

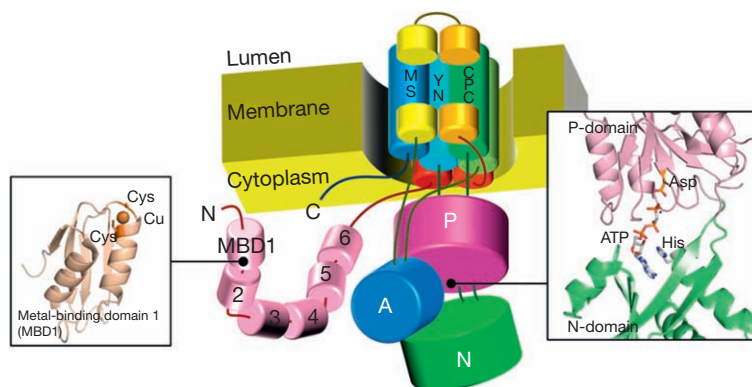


Figure 2 The major functional domains of human Cu pumps. The transmembrane portion of a Cu pump is composed of eight transmembrane segments (colored in spectrum, red to blue, according to their order from the N-terminus) that form two intramembrane copper-binding site(s) contributed by the sequence motifs, CPC, YN, and MxxS. The cytosolic catalytic core is composed of the P-domain (magenta) and the N-domain (green) and, together with the A-domain (turquoise), is responsible for enzymatic cycle (ATP binding, hydrolysis, phosphorylation, and dephosphorylation). The N-terminal domain has six metal (Cu)-binding subdomains (MBD1–6, light magenta). (Left inset) The structure of MBD1 with a Cu atom bound in the GMxCxxC motif. (Right inset) The local environment around bound ATP highlighting residues essential for binding. For clarity, some structural elements are not shown. The blue dot near the ATP indicates Mg^{2+} ion.

dephosphorylation ($E2P \rightarrow E2$), and may actuate rearrangement of connected TM helices, as in other pumps.

Distinct Functional Properties of Cu Pumps

Cu pumps have evolved to accommodate unique properties of Cu (particularly Cu^+) that are markedly different from those of other ions transported by the P-type pumps. First, Cu^+ prefers sulfur as a coordinating atom, whereas hard metal ions are predominantly associated with oxygen atoms. Generally, Cu^+ -sulfur interaction is stronger than the hard metal-oxygen interaction. This explains extremely high affinity observed for various Cu^+ -binding proteins including Cu pumps. Measurements for archaeal Cu pump CopA shows an apparent affinity of 1.1 fM^{-1} for one of the two intramembrane Cu-binding sites. The mechanism through which Cu pumps disassemble such an extremely stable binding site to dissociate bound Cu is presently unknown. A drastic conformational change, protonation by low luminal pH, and/or assistance of Cu-acceptor molecules may conceivably facilitate copper release.

Second, unlike H^+ , Na^+ , K^+ , Ca^{2+} , or Mg^{2+} , free Cu ions are mostly unavailable in the cytosol. In yeast, the number of free Cu ions was calculated to be below one per cell. In eukaryotes, cytosolic Cu exists largely in a complex with proteins including Cu chaperones, which carry Cu ions to specific protein acceptors. The human Cu pumps are thought to receive their Cu from the copper chaperone Atox1. This step is important but not obligatory for their activity, because the phenotype of mice lacking Atox1, although reminiscent of ATP7A loss, appears milder than that of mice lacking the Atox1 downstream targets ATP7A and ATP7B. Also, ATP7A senses copper and traffics in the *Atox1*^{-/-} cells, although with slower kinetics compared to control. Current data indicate that Atox1 transfers copper to the N-MBD of human Cu pumps, but little information is available on how copper reaches the TM Cu-binding sites.

Third, the turnover of Cu pumps is strikingly slower compared to other ion-pumping P-type ATPases ($\sim 30 \text{ s}^{-1}$ for Ca^{2+} pump and $0.1\text{--}3 \text{ s}^{-1}$ for Cu^+ pump). This may not be surprising since the Cu pumps do not generate significant ion gradients and their major role is to deliver copper for further utilization by the enzymes, which are synthesized at limited quantities and with defined kinetics. It is thought that the turnover of Cu pumps is regulated by their unique N-terminal domain. This domain is present in all Cu pumps, and it influences their reaction cycle via copper binding and copper-dependent interactions with other domains. Mutation and deletion of the N-terminal domain have different consequences on the reaction cycle in Cu pumps from different species, although the decrease of Cu transport activity is common.

Finally, unlike other P-type pumps which work predominantly in a single cellular location (plasma membrane, endoplasmic reticulum, and secretory pathway), the mammalian Cu pumps traffic within the cell and function in different intracellular compartments. Regulated trafficking allows Cu pumps to switch from their function in the biosynthesis of cuproenzymes to copper export and protection of cells from copper overload. Studies of the trafficking mechanism have identified several factors involved in regulation of this process

and suggested an important link between the Cu transport and a kinase-mediated signaling. The unique molecular functions of Cu pumps and their regulation can be better understood if one considers the unique structural elements of Cu pumps and compares those to other P-type pumps.

Unique Structural Features of Cu Pumps

TM domain. The TM domain is the place where ions are selectively bound, occluded, and then released in a vectorial manner. Therefore, this domain is the fundamental component of the ion-pumping machinery. The TM domain consists of 10 helices in most P-type pumps; however, the transition metal pumps including Cu pumps have eight TM helices. The membrane topologies and the estimated arrangements of the TM helices in the two-dimensional projection maps suggest that TM3–TM8 of the Cu pump are counterparts of TM1–TM6 of other P-type pumps. The TM1, 2 hairpin is specific to transition metal pumps. In Cu pumps, it frequently contains Met and His residues suggestive of copper-binding role, though their actual function is yet to be explored. High-resolution model of Cu pumps is not yet available; however, a detailed sequence analysis and mutagenesis using CopA from *A. fulgidus* determined that the Cu pumps have two Cu-binding sites in the TM region. Site I is constituted by Cys/TM6, Cys/TM6, and Tyr/TM7; while site II constituted by Asn/TM7, Met/TM8, and Ser/TM8. Consistent with this predicted coordination, the EXAFS spectra of bound Cu ions indicated trigonal geometries and the participation of nitrogen/oxygen atoms as well as sulfur. The residues predicted to coordinate copper are conserved within Cu pumps including human homologs (ATP7A and ATP7B). Mutation of Ser/TM8 of ATP7B has been identified with the WND phenotype.

Cytosolic catalytic domains. The N domain of a Cu pump (and other transition-metal pumps) is distinguished by the unique amino acid sequence and the distinct mode of nucleotide recognition compared to other P-type ATPases. The solution structure of the N domain of ATP7B pointed to residues forming the ATP-binding pocket (H1069, G1099, G1101, I1102, G1149, and N1150) as these residues showed chemical shifts upon binding of nucleotide. The ATP-coordinating residues are highly conserved in all Cu pumps but differ from the ATP-coordinating residues of other P-type pumps. Such difference in coordination environment may explain a tighter binding of nucleotide by the isolated N domain of ATP7B compared to the N domains of other P-type pumps and may contribute to a slow turnover rate of Cu pumps during their reaction cycle. Recently, the coordination environment and the position of ATP in the pocket became crystal clear with two atomic models of nucleotide-bound forms (archaea CopA and ATP7A). These studies highlighted an important role of the conserved Glu (in the ExSEHPL motif), which was found hydrogen bonded with the adenine ring.

The structures of the A domains of archaea (*A. fulgidus* CopA) and human ATP7A revealed their high similarity to those of other P-type pumps. However, the A domains of other P-type pumps in addition to a central β -sheet/ β -hairpin contain a small α -helical element. In SERCA, this α -helical part transmits the movement of the A domain to the TM helix regulating the cytosolic gating of the ion path. The Cu

pumps apparently lack this important helical element. This difference raises a question as ‘how do the Cu pumps regulate the gating of the ion path? Is there a gating?’ One suggested possibility is that the N-MBD associates with the A domain and substitutes the missing α -helical part with one of its helices.

N-terminal metal-binding domain. The N-MBD is a characteristic component of Cu pumps, which is involved in regulations of catalytic activity and protein trafficking. This domain is composed of one to six (mammalian Cu pumps) homologous subdomains. Each of the subdomains is, typically, 70–80 amino-acid residues long, has a ferredoxin-like $\beta\alpha\beta\beta\alpha\beta$ -fold, and houses a single Cu-binding site GMxCxC, in which two invariant Cys residues coordinate Cu(I). Cu binding induces rearrangement of the subdomains, which can be readily visible in different proteolytic patterns (limited proteolysis is often used to detect a conformational change) or by circular dichroism spectroscopy. In human Cu pumps, the tight packing of six N-terminal MBD is responsible for the maintenance of the reduced state of the Cu-binding sites and for conformational transitions of the entire domain in response to copper binding. In eukaryotic pumps, the NMBD is thought to serve as a docking site for copper chaperone Atox1. Copper transfer from Atox1 to N-MBDs initiates the transfer of copper to the TM domain. In archaeal CopA, the direct delivery of copper to the TM domain by the copper chaperone copZ was also shown.

Regulation of Mammalian Cu Pumps

Regulations of Cu Pumps via Trafficking

Copper chaperone Atox1 is likely to sense cytosolic copper and regulate copper occupancy of Cu pumps. Atox1 and N-terminal metal-binding subdomains have the same fold and the Cu-binding motif (GMxCxC). *In vitro*, Atox1 and the NMBD interact with each other and exchange bound Cu ions. The Cu-Atox1 complex stimulates the Cu-pump activity, while apo Atox1 removes copper from NMBD and downregulates Cu pump. The Atox1-mediated regulation of copper binding allows NMBD to assume different conformations which may be central for further regulation.

In response to the elevation of cellular Cu, the mammalian Cu pumps traffic from the TGN to vesicles and during vesicle fusion reach the plasma membrane (Figure 1). This process is reversible and Cu pumps return back to the TGN when copper levels decrease. So far, three major targeting motifs have been identified in human Cu pumps. The LL or LLL motifs in the protein C-termini play a role in endocytic retrieval of Cu pumps from the plasma membrane (ATP7A), as well as the exit from the TGN. The C-terminus of ATP7A, but not ATP7B, has a putative PDZ-binding motif, DTAL, which is required for the basolateral targeting. By contrast, ATP7B contains the FAFDNV-GYE sequence, which is an essential apical-targeting determinant and is located in the N-terminal segment before the first metal-binding site.

The mechanism through which changing Cu levels regulate the trafficking of Cu pumps is not fully understood, although conformational changes in Cu pumps in response to copper binding were reported. Mutagenic studies demonstrated that the loss of transport function results in the loss of trafficking,

indicating that the binding of ligands and/or saturation of lumen with copper play a significant role in the Cu pumps exit from the TGN. The trafficking ability of inactive mutants can be restored by trapping the pump in a specific conformation; however, in such mutants the sensitivity to Cu appears altered or lost resulting in constitutive trafficking under basal conditions. The current data are best accommodated by the model, where elevated Cu triggers conformational change in the NMBD, which stimulates the Cu-pump activity and induces further structural changes in the pump-making sites available to trafficking machinery. Lumenal signaling by copper (under conditions of saturation) may also contribute to the trafficking from the TGN. Also, other signals/pathways contribute to trafficking response. Insulin and estrogen promote ATP7A trafficking, but suppress ATP7B trafficking in placental cells; while activation of the N-methyl-D-aspartate (NMDA) receptor promotes trafficking of both ATP7A and ATP7B in hippocampal neurons, presumably by facilitating Ca^{2+} -dependent vesicle fusion. Current evidence suggests that kinase-mediated phosphorylation could be one of the unifying pathways that regulates trafficking of mammalian Cu pumps.

Kinase-Mediated Phosphorylation of Mammalian Cu Pumps

Kinase-mediated phosphorylation was reported for both ATP7A and ATP7B, and the current data point to the important role of kinase-mediated phosphorylation in regulating the activity and intracellular localization of these Cu pumps. Mammalian Cu pumps are phosphorylated at Ser residues by Mg^{2+} /ATP-dependent kinases in basal conditions and in response to the increase in the intracellular Cu level. Mass-spectrometric analysis of ATP7A and ATP7B identified multiple phosphorylation sites, located predominantly in the NMBD and C-terminus of Cu pumps. Specific function of each phosphorylation site remains to be explored; however, in ATP7A mutation of C-terminal phosphorylation sites results in mislocalization of Cu pumps. It is interesting that the kinase-mediated phosphorylation appears to be coupled to catalytic phosphorylation, a finding that may link together several events that lead to Cu pump exit from the TGN: copper binding, stimulation of transport, and phosphorylation.

See also: [Bioenergetics: Membrane Transport, General Concepts; Superoxide Dismutase.](#)

Further Reading

- Argüello JM, Eren E, and González-Guerrero M (2007) The structure and function of heavy metal transport P1B-ATPases. *Biometals* 20: 233–248.
- Banci L, Bertini I, Cantini F, et al. (2006) The Atx1–Ccc2 complex is a metal-mediated protein–protein interaction. *Nature Chemical Biology* 2: 367–368.
- Barry AN, Shinde U, and Lutsenko S (2010) Structural organization of human Cu-transporting ATPases: Learning from building blocks. *Journal of Biological Inorganic Chemistry* 15: 47–59.
- Boal AK and Rosenzweig AC (2009) Structural biology of copper trafficking. *Chemical Reviews* 109: 4760–4779.
- Cox DW and Moore SD (2002) Copper transporting P-type ATPases and human disease. *Journal of Bioenergetics and Biomembranes* 34: 333–338.

- Hatori Y, Lewis D, Toyoshima C, and Inesi G (2009) Reaction cycle of *Thermotoga maritima* copper ATPase and conformational characterization of catalytically deficient mutants. *Biochemistry* 48: 4871–4880.
- La Fontaine S, Ackland ML, and Mercer JF (2010) Mammalian copper-transporting P-type ATPases, ATP7A and ATP7B: Emerging roles. *International Journal of Biochemistry and Cell Biology* 42: 206–209.
- Lutsenko S, Barnes NL, Bartee MY, and Dmitriev OY (2007) Function and regulation of human copper-transporting ATPases. *Physiological Reviews* 87: 1011–1046.
- Toyoshima C (2009) How Ca^{2+} -ATPase pumps ions across the sarcoplasmic reticulum membrane. *Biochimica et Biophysica Acta* 1793: 941–946.
- Veldhuis NA, Gaeth AP, Pearson RB, Gabriel K, and Camakaris J (2009) The multi-layered regulation of copper translocating P-type ATPases. *Biometals* 22: 177–190.

P-Type Pumps: H⁺/K⁺ Pump

J M Shin and G Sachs, University of California, Los Angeles, CA, USA; VA Greater Los Angeles Healthcare System, Los Angeles, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Gastric H⁺,K⁺-ATPase An enzyme, which is one of the P₂-type ion-motive ATPases, that allows an electroneutral exchange of cytoplasmic protons for extracytoplasmic potassium.

α-Subunit of gastric H⁺,K⁺-ATPase Functional subunit consisting of 1034 amino acids with an M_r

of 114 kDa, which has 10 transmembrane segments.

β-Subunit of gastric H⁺,K⁺-ATPase A glycoprotein consisting of ~290 amino acids having one transmembrane segment and six or seven N-linked glycosylation sites in the extracytoplasmic region.

The gastric H⁺/K⁺ pump is a member of the P₂-type ion-motive adenosine triphosphatase (ATPase) family and is responsible for the elaboration of HCl by the parietal cell of the gastric mucosa. The gastric H⁺,K⁺-ATPase catalyzes H⁺ for K⁺ exchange by an electroneutral exchange of cytoplasmic protons for extracytoplasmic potassium, while the protein undergoes phosphorylation and dephosphorylation related to ion binding and release. The gastric H⁺,K⁺-ATPase is composed of two subunits, the catalytic or α-subunit having 10 transmembrane segments and the β-subunit having 1 transmembrane segment. The gastric H⁺,K⁺-ATPase α-subunit is a 100-kDa protein consisting of ~1034 amino acids and β-subunit is a glycoprotein with ~290 amino acids.

Structure of the Gastric H⁺,K⁺-ATPase

The α-Subunit of Gastric H⁺,K⁺-ATPase

The primary sequences of the α-subunits deduced from complementary DNA (cDNA) have been identified from pig, rat, and rabbit. The hog gastric H⁺,K⁺-ATPase α-subunit sequence deduced from its cDNA consists of 1034 amino acids with an M_r of 114 285 Da. The rat gastric H, K ATPase consists of 1033 amino acids with an M_r of 114 012 Da, and the rabbit gastric H, K ATPase consists of 1035 amino acids, showing an M_r of 114 201 Da. The degree of conservation among the α-subunits is extremely high (over 97% identity) among species. The gene sequence for human and the 5' part of the rat H⁺,K⁺-ATPase α-subunits have been determined showing that the human gastric H⁺,K⁺-ATPase gene has 22 exons and encodes a protein of 1035 residues including the initiator methionine residue (M_r = 114 047 Da). These H⁺, K⁺-ATPase α-subunits show high homology (~60% identity) with the Na⁺, K⁺-ATPase catalytic α-subunit. The putative distal colon H⁺,K⁺-ATPase α-subunit has also been sequenced and shares 75% homology with both the H⁺,K⁺- and Na⁺, K⁺-ATPases.

The gastric H⁺,K⁺-ATPase α-subunit has conserved consensus sequences along with the other P type ATPases, the sarcoplasmic reticulum Ca²⁺-ATPase and the Na⁺, K⁺-ATPase, for the ATP-binding site, the phosphorylation site, the pyridoxal 5'-phosphate-binding site, and the fluorescein isothiocyanate-binding

site. These sites are thought to be within the ATP-binding domain in the large cytoplasmic loop between transmembrane segments 4 and 5. Structurally, this enzyme can be divided into three domains, namely, cytoplasmic, membrane, and extracytoplasmic domains. The cytoplasmic domain contains four subdomains, a stalk domain, N-domain, P-domain, and A-domain. The N-domain or nucleotide-binding domain is a large cytoplasmic domain between M4 and M5 where ATP binds. The P-domain or phosphorylation domain is a cytoplasmic domain near M4 where there is phosphorylation and dephosphorylation. The A-domain or activation domain is a cytoplasmic domain between M2 and M3, which moves depending on conformational changes. The stalk domain contains the extension of the transmembrane segments into the cytoplasmic region and forms a link for the passage of ions into the membranes and for transmission of conformational changes from the cytoplasm. The gastric H⁺, K⁺-ATPase structure is shown in [Figure 1](#), based on a 6.5-Å resolution of two-dimensional (2D) crystals reported by Abe et al. in 2009.

There are 10 transmembrane segments in the H⁺,K⁺-ATPase α-subunit. The first four transmembrane pairs connected by their luminal loop were detected in a tryptic digest of the hog gastric H⁺,K⁺-ATPase. A peptide fragment beginning at gln¹⁰⁴ represents the H1/loop/H2 sector. The H3/loop/H4 sector was found at a single peptide peak beginning at thr²⁹¹, and the H5/loop/H6 sector at a peptide beginning at leu⁷⁷⁶. The H7/loop/H8 region was found in a single peptide fragment of 11 kDa, beginning at leu⁸⁵³.

Antibody 95 inhibits ATP hydrolysis in the intact vesicles, and appears to be K⁺ competitive. The sequence recognized by this antibody was between amino acid positions 529 and 561. As it inhibits the enzyme in intact cytoplasmic side out vesicles, this epitope must be cytoplasmic and it is close to the region known to bind the cytoplasmic reagent, fluorescein isothiocyanate, a lysine reagent, in the loop between H4 and H5. Antibody 1218 was shown to have its major epitope between amino acid positions 665 and 689 in the same N domain. A second epitope for mAb 1218 was also identified to be between amino acid positions 853 and 907. The monoclonal antibody 146 epitope was defined to be between positions 873 and 877 of the hog α-subunit. This is on or close to the extracytoplasmic face of H7.

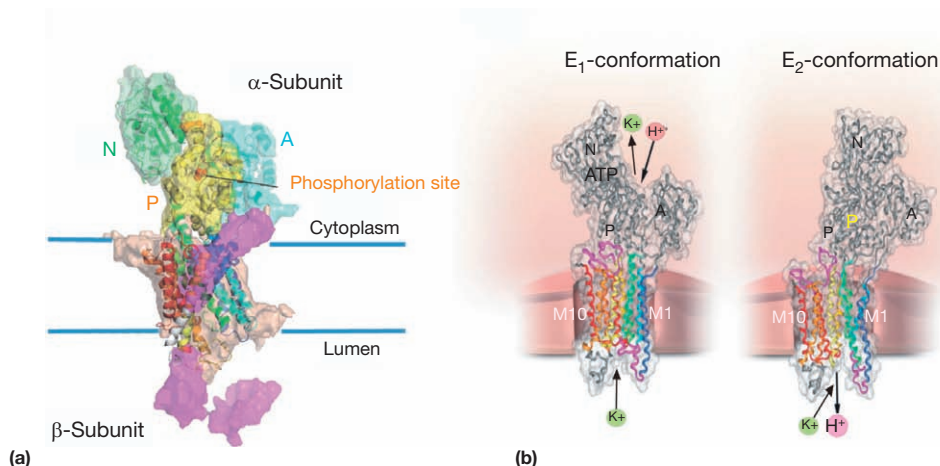


Figure 1 Proposed structure of the gastric H^+,K^+ -ATPase. (a) The α -subunit has three lobes, N (ATP binding), P (phosphorylation), and A (activation) regions in the cytoplasmic domain, and 10 transmembrane segments in the membrane domain. The gastric β -subunit has one transmembrane segment and a heavily glycosylated extracellular region. (b) Conformational changes of gastric H^+,K^+ -ATPase α -subunit. With binding of H^+ and MgATP to the E_1 form, there is transient association of the N and P lobes with phosphorylation of asp386 and a conformational change to E_2P so that the A domain associates with the P domain. This change is transmitted to the membrane domain in the region of TM6 (dark yellow) and H^+ binding to the cation-binding domain in the middle of the membrane domain is followed by release of H^+ and binding of K^+ from the luminal surface. There is then dephosphorylation and reformation of the E_1 conformation with release of K^+ to the cytoplasm.

The β -Subunit of Gastric H^+,K^+ -ATPase

The primary sequences of the β -subunits have been reported for rabbit, hog, rat, mouse, and human. The β -subunit with ~ 290 amino acids is a glycoprotein having one transmembrane segment located at the region between amino acid sequence positions 38 and 63 near the N terminus and six or seven N-linked glycosylation sites in the extracytoplasmic region. Using lectin-affinity chromatography, the H^+,K^+ -ATPase α -subunit was co-purified with the β -subunit, showing that the α -subunit interacts with the β -subunit. By cross-linking with low concentrations of glutaraldehyde, the β -subunit was shown to be closely associated with the α -subunit. In the extracytoplasmic domain, there are six cysteines that are linked through disulfide bonds. Reduction of the disulfides of the β -subunit inhibits the activity of the H^+,K^+ -ATPase. In the case of the Na^+,K^+ -ATPase, the effect of reducing agents on the ability of the enzyme to hydrolyze ATP and bind ouabain was quantitatively correlated with the reduction of disulfide bonds in the β -subunit. Seven putative N-glycosylated sites (AsnXaaSer and AsnXaaThr) are shown in rabbit H^+,K^+ -ATPase β -subunit, which are conserved in rats and humans. Six putative N-glycosylation sites are found in the hog gastric β -subunit. The N-linked oligosaccharides of the β -subunit of rabbit gastric H^+,K^+ -ATPase were identified. All N-linked AsnX(Ser/Thr) sequences at positions 99, 103, 130, 146, 193, and 222 were fully glycosylated. Asn 99 was modified exclusively with oligomannosidic-type structures, $Man_6GlcNAc_2$ – $Man_8GlcNAc_2$, and Asn193 has $Man_5GlcNAc_2$ – $Man_8GlcNAc_2$ and lactosamine-type structures. Asn 103, 146, 161, and 222 contain lactosamine-type structures. All the branches of the lactosamine-type structure were terminated with $Gal\alpha$ – $Gal\beta$ – $GlcNAc$ extensions. The role of the carbohydrate chains was studied using HEK-293 cells. The enzyme activity

was not affected by removal of any one of seven carbohydrates of the β -subunit; however, removal of all carbohydrate chains resulted in the complete loss of activity. Except for the second glycosylation site, each site was essential for trafficking from Golgi to plasma membrane, but not for endoplasmic reticulum to Golgi movement.

The function of the β -subunit is not clearly known, but this subunit appears to be required for proper assembly and targeting of the catalytic subunit. Perhaps the β -subunit has a function in maintaining the structure of the α -subunit to enable effective binding of K^+ ion to the outside face of the α -subunit. The H^+,K^+ -ATPase β -subunit has the sequence Phe–Arg–His–Tyr in its cytoplasmic domain. This tyrosine has been shown to be important for the retrieval of the H^+,K^+ -ATPase from the apical membrane of the parietal cells and to ensure its return to the tubulovesicular compartment in order to terminate the process of acid secretion. The participation of a tyrosine-based signal in the retrieval of the H^+,K^+ -ATPase suggests that this process involves interactions with adaptins and is mediated by clathrin-coated pit formation. In knockout mice, deficient in the H^+,K^+ -ATPase β -subunit, cells that express the H^+,K^+ -ATPase α -subunit had abnormal canaliculi and were devoid of typical tubulovesicular membranes.

Regions of Association in the Oligomeric Structure

There has been much suggestive evidence that the α – β heterodimeric H^+,K^+ -ATPase exists as an oligomer. Such evidence includes target size by irradiation and unit cell size of the enzyme in 2D crystals. Using Blue native gel separation and cross-linking with Cu^{2+} -phenanthroline, it was shown that the enzyme did indeed exist as an $(\alpha\beta)_2$ heterodimeric dimer. Membrane-bound H^+,K^+ -ATPase reacted with

Cu-phenanthroline to provide an α - α dimer. The site of Cu²⁺-oxidative cross-linking was at either cys⁵⁶⁵ or cys⁶¹⁶. Hence, this region of the α -subunit is in close contact with its neighboring α -subunit. No evidence was obtained for β - β dimerization. ATP prevents this Cu-phenanthroline-induced α - α dimerization. The M7/loop/M8 sector of the α -subunit is tightly associated with the β -subunit. An antibody mAb 146-14 recognizes the region of the α -subunit at the extracytoplasmic face of the M7 segment as well as the β -subunit, a finding consistent with the association found by column chromatography. Using yeast two-hybrid analysis, a fragment Leu855 to Arg922 of the α -subunit was identified to bind the β -subunit. This is at the entrance of the TM8 into the membrane domain.

Kinetics and Conformational Changes of the Gastric H⁺,K⁺-ATPase

The gastric H⁺,K⁺-ATPase exchanges intracellular hydrogen ions for extracellular potassium ions by consuming ATP. The H⁺ for K⁺ stoichiometry of the H⁺,K⁺-ATPase is not clear. There are reports of two different stoichiometries, one or two H⁺ per ATP hydrolyzed. In ion-tight vesicles studied at pH 6.1, the stoichiometry is 2H⁺:1ATP but at full pH gradient where the pH reaches 1.0, this has to fall to 1:1. Probably, protonation of a carboxylic acid in the second ion-binding site at the lower luminal pH is then maintained.

There are several steps in the reaction scheme of the H⁺,K⁺-ATPase. The enzyme is phosphorylated by ATP at asp³⁸⁶. The rate of formation of the phosphoenzyme (EP) and the K⁺-dependent rate of EP breakdown are sufficiently fast to allow the EP to be an intermediate in the overall ATPase reaction. The initial step is the reversible binding of ATP to the enzyme in the absence of added K⁺ ion, followed by an Mg²⁺ (and proton)-dependent transfer of the terminal (gamma) phosphate of ATP to asp³⁸⁶ of the catalytic subunit (E₁-P•H⁺). The Mg²⁺ remains occluded until dephosphorylation. The addition of K⁺ to the enzyme-bound acyl phosphate results in a two-step dephosphorylation. The faster initial step is dependent on the concentration of K⁺, whereas the slower step is not affected by K⁺ concentration. The second phase of EP breakdown is accelerated in the presence of K⁺ but, at K⁺ concentration exceeding 500 μ M, the ratio becomes independent of K⁺ concentration. This shows that two forms of EP exist. The first form E₂P is K⁺ sensitive and converts spontaneously in the rate-limiting step to E₁P, the K⁺ insensitive form. ATP binding to the H⁺,K⁺-ATPase occurs in both the E₁ and the E₂ state, but with a lower affinity in E₂ state (2000 times lower compared to E₁). The effects of H⁺ and K⁺ on formation and breakdown of EP were determined using transient kinetics. Increasing hydrogen ion concentrations on the ATP-binding face of the vesicles accelerate phosphorylation, whereas increasing potassium ion concentrations inhibit phosphorylation. Increasing hydrogen ion concentration reduces this K⁺ inhibition of the phosphorylation rate. Decreasing hydrogen ion concentration accelerates dephosphorylation in the absence of K⁺, and K⁺ on the luminal surface accelerates dephosphorylation. Increasing K⁺ concentrations at constant ATP decreases the rate of phosphorylation and increasing ATP

concentrations at constant K⁺ concentration accelerates ATPase activity and increases the steady-state EP level. Therefore, inhibition by cations is due to cation stabilization of a dephospho form at a cytosolically accessible cation-binding site. The reaction mechanism is similar to that of the Na⁺,K⁺-ATPase.

As Na⁺ is transported as a surrogate for H⁺ at high pH, it is likely that the hydronium ion, rather than the proton *per se* is the species transported. The ions transported inward from the outside face of the pump are Tl⁺, K⁺, Rb⁺, or NH₄⁺. Presumably, the change in conformation changes a relatively small ion-binding domain in the outward direction into a larger ion-binding domain in the inward direction.

The E₁ conformation of the H⁺,K⁺-ATPase binds the hydronium ion from the cytoplasmic side at high affinity. With phosphorylation, the conformation changes from E₁P•H₃O⁺ to E₂P•H₃O⁺ form, which has high affinity for K⁺ and low affinity for H₃O⁺ allowing release of H₃O⁺ and binding of K⁺ on the extracytoplasmic surface of the enzyme. The rapid breakdown of the E₂P form requires K⁺ or its congeners on the outside face of the enzyme. With dephosphorylation, the E₁K⁺ conformation is produced with a low affinity for K⁺, releasing K⁺ to the cytoplasmic side, allowing rebinding of H₃O⁺.

Fluorescein isothiocyanate binds to the H⁺,K⁺-ATPase, inhibiting ATPase activity but not pNPPase activity. Fluorescence of the fluorescein isothiocyanate (FITC)-labeled enzyme, representing the E₁ conformation, was quenched by K⁺, Rb⁺, and Tl⁺. The quenching of the fluorescence by KCl reflects the formation of E₂K⁺. FITC binds at lys⁵¹⁶ in the hog enzyme sequence. This FITC-binding site apparently becomes less hydrophobic when KCl binds to form the E₂•K conformation. In the E₁ form, the FITC region is relatively closer to the membrane and the extracytoplasmic loop relatively hydrophilic. With the formation of the E₂•K form, the FITC region is more distant from the membrane. These postulated conformational changes are, therefore, reciprocal in the two major conformers of the enzyme.

When the gastric H⁺,K⁺-ATPase was cleaved by Fe²⁺-catalyzed oxidation under various ligand, cleavage patterns were different between the different conformations. There are two Fe²⁺ cleavage sites. In Fe²⁺ site1, the parallel appearance of the fragments at ²³⁰ESE, near ⁶²⁴MVTGD, and at ⁷²⁸VNDS upon transition from E₁ to E₂(Rb) conformations were observed. Meanwhile, in Fe²⁺ site2, the fragment near ²⁹⁹HFIH was cleaved independent of conformational changes. These cleavage patterns were the same as those of the Na⁺,K⁺-ATPase. The cleavage data showed that the structural organization and changes in the cytoplasmic domains, association with E₁/E₂ transitions, are essentially the same for the H⁺,K⁺-ATPase, the Na⁺,K⁺-ATPase, and sr Ca-ATPase. N-domain where ATP binds inclines nearly 90° with respect to the membrane and the A-domain rotates by about 110° horizontally during the E₁ to E₂ conformation.

Functional Residues of the H⁺,K⁺-ATPase

When the gastric H⁺,K⁺-ATPase was digested by trypsin in the presence of high concentration of KCl, the tryptic membrane digest showed the presence of Rb⁺ occlusion, that is, stable binding of Rb⁺ in the membrane domain of the catalytic

subunit as found for the Na⁺, K⁺-ATPase. Some regions near the membrane such as the region between gly⁹³ and glu¹⁰⁴ near the M1 segment, the region between asn⁷⁵³ and leu⁷⁷⁶ near the cytoplasmic side of the M5 segment, and the region after the M8 segment, especially the region between ile⁹⁴⁵ and ile⁹⁶³ containing five arginines and one lysine, were K⁺ protected against digestion. Furthermore, when K⁺ was removed from this membrane digest, the M5/M6 hairpin was released from the membrane, showing that M5M6 membrane hairpin is stabilized by K⁺ ions. Proton pump inhibitors such as omeprazole, pantoprazole, lansoprazole, and rabeprazole all bind to Cys813 of M5M6, giving inhibition of activity.

Using site-directed mutagenesis of the gastric H⁺,K⁺-ATPase transfected in HEK293 cells, the M5M6 luminal loop was studied in terms of K⁺ access to the ion-binding domain. Mutations of M5, M5–M6, and M6 regions, such as P798C, Y802L, P810A, C813A or S, F818C, and T823V, and mutations of M7–M8 and M8, such as E914Q, F917Y, G918E, T929L, and F932L, reduced the affinity for SCH28080, a reversible proton pump inhibitor, up to 10-fold without affecting the nature of the kinetics. The L809F substitution in the loop between M5 and M6 resulted in about 100-fold decrease in inhibitor affinity. The C813T mutant showed ninefold loss of SCH28080 affinity. All these data suggest that the binding domain for SCH28080 is at the surface between L809 and C813, where the M5–M6 loop and the luminal end of M6 are located. Mutations of C813 and I816 in M6 and M334 in M4 also showed that the inhibitor binds close to the luminal surface of the enzyme. Mutations of the negatively charged amino acid residue in the α -subunit showed that the carboxyl group of the membrane-spanning domain is important for cation binding. Mutation of E820Q showed decreased K⁺ sensitivity and the dephosphorylation was not stimulated by either K⁺ or ADP, indicating that E820 might be involved in K⁺ binding and transition to the E₂ form of the H⁺,K⁺-ATPase. Mutation of E795 showed a decrease of the phosphorylation rate and the apparent ATP affinity, indicating that E795 is involved in both K⁺ and H⁺ binding. Mutation of E795 and E820 in M5 and M6 resulted in a K⁺-independent, SCH28080-sensitive ATPase activity, due to a high spontaneous dephosphorylation rate. Thus, the sixth transmembrane (M6) segment of the catalytic subunit plays an important role in the ion recognition and transport in the type II P-type ATPase families. When all amino acid residues in the M6 segment of gastric H⁺,K⁺-ATPase α -subunit were singly mutated with alanine, four mutants, L819A, D826A, I827A, and L833A, completely lost K⁺-ATPase activity. Mutant L819A was phosphorylated but barely dephosphorylated in the presence of K⁺, whereas mutants D826A, I827A, and L833A were not phosphorylated from ATP. Amino acids involved in the phosphorylation are located exclusively in the cytoplasmic half of the M6 segment and those involved in the K⁺-dependent dephosphorylation are in the luminal half. Several mutants such as I821A, L823A, T825A, and P829A partly retained the K⁺-ATPase activity accompanying the decrease in the rate of phosphorylation. Substituting three residues in the Na⁺, K⁺-ATPase sequence with their H⁺,K⁺-ATPase counterparts (L319F, N326Y, and T340S) and replacing the TM3–TM4 ectodomain sequence with that of the H⁺,K⁺-ATPase result in a pump that gives 50% of ATPase activity in the absence of Na⁺ at pH 6.

The cation selectivity of the Na⁺, K⁺- and H⁺,K⁺-ATPase would be generated through a cooperative effort between residues of the transmembrane segments and the flanking loops that connect these transmembrane domains. A model that results from these studies and the three-dimensional (3D) structure of the sarcoplasmic reticulum (SR) Ca-ATPase suggests that ion transport occurs between TM4, 5, 6, and perhaps TM8 and that the vestibule around cys813 is the binding domain of both compounds such as omeprazole and SCH28080.

Acid Secretion and the Gastric H⁺,K⁺-ATPase

The H⁺,K⁺-ATPase is present only in the gastric parietal cell and the intercalated cell of the distal renal tubule. In the resting parietal cell, it is present in smooth-surfaced cytoplasmic membrane tubules. Upon stimulation of acid secretion, the pump is translocated to the microvilli of the secretory canaliculus of the parietal cell. This morphological change results in a several-fold expansion of the canaliculus. There are actin filaments within the microvilli and the subapical cytoplasm. In the cytoskeleton system, there is abundance of microtubules among the tubulovesicles. Some microtubules appeared to be associating with tubulovesicles. A large number of electron-dense coated pits and vesicles were observed around the apical membrane vacuoles in cimetidine-treated resting parietal cells, consistent with an active membrane uptake in the resting state. Cultured parietal cells also undergo morphological transformation under histamine stimulation, resulting in great expansion of apical membrane vacuoles. Immunogold labeling of H⁺,K⁺-ATPase was present not only on the microvilli of expanded apical plasma membrane vacuoles but also in the electron-dense coated pits.

There is activation of a K⁺ and Cl[−] conductance in the pump membrane, which allows K⁺ to access the extracytoplasmic face of the pump. This allows H⁺ for K⁺ exchange to be catalyzed by the ATPase.

The covalent inhibitors of the H⁺,K⁺-ATPase that have been developed for the treatment of ulcer disease and esophagitis depend on the presence of acid secreted by the pump. These are acid-activated prodrugs that accumulate in the acid space of the parietal cell. Hence, their initial site of binding is only in the secretory canaliculus of the functioning parietal cell. These data also show that the pump present in the cytoplasmic tubules does not generate HCl.

The upstream DNA sequence of the α -subunit contains both Ca and cyclic adenosine monophosphate responsive elements in the case of the rat H⁺,K⁺-ATPase. There are gastric nuclear proteins present that bind selectively to a nucleotide sequence, GATACC, in this region of the gene. These proteins have not been detected in other tissues. Stimulation of acid secretion by histamine increases the level of messenger RNA (mRNA) for the α -subunit of the pump. Elevation of serum gastrin, which secondarily stimulates histamine release from the enterochromaffin-like cell in the vicinity of the parietal cell, also stimulates the mRNA levels in the parietal cell. H₂ receptor antagonists block the effect of serum gastrin elevation on mRNA levels. It seems, therefore, that activity of the H₂ receptor on the parietal cell determines, in part, the gene expression of the ATPase. It might be expected, therefore, that

chronic stimulation of this receptor would upregulate pump levels, whereas inhibition of the receptor would downregulate levels of the ATPase.

However, chronic administration of these H₂ receptor antagonists, such as famotidine, results in an increase in pump protein, whereas chronic administration of omeprazole (which must stimulate histamine release) reduces the level of pump protein in the rabbit. Regulation of pump protein turnover downstream of gene expression must account for these observations.

See also: **Bioenergetics:** Membrane Transport, General Concepts; **Cell Architecture and Function:** Actin Assembly/Disassembly; Actin-Capping and -Severing Proteins; Actin-Related Proteins; **Metabolism Vitamins and Hormones:** Amino Acid Metabolism.

Further Reading

- Abe K, Tani K, Nishizawa T, and Fujiyoshi Y (2009) Inter-subunit interaction of gastric H⁺,K⁺-ATPase prevents reverse reaction of the transport cycle. *EMBO Journal* 28: 1637–1643.
- Munson K, Lambrecht N, Shin JM, and Sachs G (2000) Analysis of the membrane domain of the gastric H⁺,K⁺-ATPase. *Journal of Experimental Biology* 203: 161–170.
- Munson K, Law RJ, and Sachs G (2007) Analysis of the gastric H⁺,K⁺-ATPase for ion pathways and inhibitor binding sites. *Biochemistry* 46: 5398–5417.
- Munson K, Vagin O, Sachs G, and Karlisch S (2003) Molecular modeling of SCH28080 binding to the gastric H,K-ATPase and MgATP interactions with SERCA- and Na,K-ATPases. *Annals of the New York Academy of Sciences* 986: 106–110.
- Vagin O, Munson K, Denevich S, and Sachs G (2003) Inhibition kinetics of the gastric H,K-ATPase by K-competitive inhibitor SCH28080 as a tool for investigating the luminal ion pathway. *Annals of the New York Academy of Sciences* 986: 111–115.

P-Type Pumps: Na⁺,K⁺-ATPase

H Haviv and S J D Karlish, Weizmann Institute of Science, Rehovoth, Israel

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Peter L. Jorgensen, volume 3, pp. 571–576, © 2004, Elsevier Inc.

Glossary

Electrical driving force A force derived from a potential difference that causes ion movement due to charge repulsion and attraction.

Free energy The amount of energy that can be converted to work in a physical system.

Hydrolysis A chemical reaction in which a compound is split in two by addition of a water molecule that is also split into hydrogen and hydroxyl.

Hypokalemia Lower than normal (<3.5 mEq l⁻¹) concentration of K⁺ in the blood.

Immunoglobulin-like fold A protein fold that is similar to mammalian antibody fold consisting of a sandwich of beta sheets arranged in a Greek key motif.

Phosphorylation/dephosphorylation Addition/removal of a phosphate group to/from an amino acid side chain in a protein.

Type II membrane proteins Proteins that span the membrane only once, and their carboxy terminus resides outside the cell.

Introduction

The Na⁺,K⁺-ATPase, or sodium/potassium pump, is a membrane protein that is ubiquitous in animal cell membranes. Na⁺,K⁺-ATPase was described first by J C Skou in 1957, a discovery that led to the award of the 1997 Nobel prize in chemistry. Na⁺,K⁺-ATPase couples the hydrolysis of one adenosine triphosphate (ATP) molecule to the transport of three Na⁺ ions out and two K⁺ ions into the cell, thus creating the characteristic concentration gradients of these ions across the cell membrane (typically: intracellular Na⁺, 10–20 mM; K⁺, 100 mM and extracellular Na⁺, 140 mM; K⁺, 4.5–5 mM). At rest, the pump converts 20–30% of the cellular ATP production to active Na⁺ and K⁺-transport in kidney, intestine, muscle, heart, central nervous system, and other tissues. The cell utilizes the Na⁺- and K⁺ gradients to drive numerous processes including volume regulation, electrolyte balance, and secondary transport of sugars, amino acids, protons, and many other metabolites. These functions maintain essential physiological processes such as neurotransmission, cardiac contraction, renal function, and intestinal absorption, etc.

The Na⁺,K⁺-ATPase is a member of the P-type family of cation pumps. The catalytic mechanism of P-type pumps involves phosphorylated active site aspartate intermediates. All P-type pumps are inhibited by vanadate. Other members of this family include sarcoplasmic reticulum Ca²⁺-ATPase (SERCA), plasma membrane Ca²⁺-ATPase, gastric H⁺,K⁺-ATPase, plant and yeast H⁺-pumps (type II, 10 trans-membrane segments), a variety of heavy metal pumps (type I, six trans-membrane + two extra trans-membrane segments), and several homologous gene products of unknown function. The human genome contains 38 P-type sequences, but only 17 of known function.

Subunits and Isoforms

The minimal functional unit of the Na⁺,K⁺-ATPase consists of two subunits, α and β , that are separate polypeptide chains

encoded by separate genes. The α subunit is the catalytic moiety of the pump, numbering about 1000 residues, arranged in 10 trans-membrane segments with three intracellular domains, and contains the binding sites for ATP, Mg²⁺, Na⁺, and K⁺, and the cardiac glycoside inhibitors. The β subunit, consisting of about ~300 residues, is a type II membrane protein with a single trans-membrane segment, and associates with very high affinity to α , to form an $\alpha\beta$ complex. The β ectodomain contains three conserved disulfide bridges and three glycosylation sites. The presence of a β subunit is unique for the K⁺-counter-transporting P-type ATPases, such as Na⁺,K⁺-ATPase and gastric and colonic H⁺,K⁺-ATPase. The β subunit is required for stabilizing the pump, targeting it to the cell membrane, and modulating K⁺ affinity.

In mammals, there exist four known isoforms of α and three of β , combinations of which are expressed and distributed in a tissue- and development-specific manner throughout the body. The isoforms are highly homologous proteins with 70–90% sequence identity. α_1 is found in nearly all tissues and is considered the housekeeping isoform. In the adult human heart, α_1 , α_2 , and α_3 are expressed but the most prominent one is α_1 . Skeletal muscles express α_1 and α_2 predominantly. In neurons, α_1 and α_3 are abundant, and α_1 and α_2 in glial cells. α_4 is expressed uniquely in sperm and is important for motility (Table 1). Each isoform has functional properties that optimally maintain the Na⁺ and K⁺ gradients in the different tissues. Table 1 summarizes the chromosomal location and tissue-specific expression of the isoforms.

Another type of membrane proteins, FXYD proteins, regulate Na⁺,K⁺-ATPase activity in a tissue- and isoform-specific manner. FXYD proteins are a family of seven accessory subunits (in mammals), named after a conserved FXYD motif, which have a single trans-membrane segment, and directly associate with the $\alpha\beta$ complex. FXYD proteins modulate functional properties of the pump, such as the affinity for Na⁺ or K⁺ ions, but are not essential for Na⁺,K⁺-ATPase activity (see Table 2 and Figure 1).

Table 1 Human genes of Na⁺,K⁺-ATPase proteins: Isoforms of α and β subunits are expressed from individual genes

<i>Isoform</i>	<i>Gene</i>	<i>Human chromosome</i>	<i>Amino acid residues</i>	<i>Cell or tissue-specific expression</i>
$\alpha 1$	ATP1A1	1p13.1	1023	Constitutive, ubiquitous, dominant in epithelia of kidney, glands, and intestine
$\alpha 2$	ATP1A2	1q23.2	1020	Brain, muscle, and heart
$\alpha 3$	ATP1A3	19q13.2	1013	CNS and brain
$\alpha 4$	ATP1A4	1q21–q23	1029	Testis and spermatozoa
$\beta 1$	ATP1B1	1q24.2	303	Ubiquitous, like the $\alpha 1$ -subunit
$\beta 2$	ATP1B2	17p13.1	290	Brain, adhesion molecule (AMOG), and liver
$\beta 3$	ATP1B3	3q23	279	Mostly in neural tissue, heart, kidney, and lung

Data from the genome database of National Center for Biotechnology Information.

Table 2 Human genes of Na⁺,K⁺-ATPase proteins: Small ion transport regulators of the FXYD family

<i>Isoform</i>	<i>Gene</i>	<i>Human chromosome</i>	<i>Amino acid residues</i>	<i>Cell or tissue-specific coexpression with α,β-unit of Na⁺,K⁺-ATPase</i>
FXYD1	PLM	19q13.1	92	Phospholemman co-expression, with $\alpha 1\beta 1$ or $\alpha 2$, reduces Na ⁺ affinity of Na ⁺ ,K ⁺ -ATPase in muscle
FXYD2	ATP1G	11q23	66	Two splice variants, $\gamma(a)$ and $\gamma(b)$. Co-expression with α,β reduces the Na ⁺ affinity of the Na ⁺ ,K ⁺ -ATPase
FXYD3	MAT-8	19q13.11	87	In kidney and colon coexpression with $\alpha 1$, CHIF increases the Na ⁺ affinity of the Na ⁺ ,K ⁺ -ATPase
FXYD4	CHIF	10p11	89	
FXYD5	RIC	19q12–q13.1	178	Phosphohippolin in brain and kidney In brain, FXYD7 reduces K ⁺ affinity of Na ⁺ ,K ⁺ -ATPase.
FXYD6	PHL	11q23.3	95	
FXYD7		19q13.13	80	

Reaction Mechanism

The reaction mechanism of Na⁺,K⁺-ATPase activity, commonly referred to as the ‘Post-Albers cycle’ depicted in [Scheme 1](#), is the generally accepted scheme to describe the sequence of coupled cation transport and biochemical reactions. The mechanism involves specific Na⁺-catalyzed phosphorylation and K⁺-catalyzed dephosphorylation reactions of the active site aspartic acid residue, and switching between two conformational states: E₁ and E₂. The E₁ conformation binds 3Na⁺ with high affinity and 2K⁺ ions with low affinity at the intracellular surface, while the E₂ conformation binds 2K⁺ ions with high affinity and 3Na⁺ ions with low affinity at the extracellular surface, respectively. During the transport cycle the ions become transiently occluded – inaccessible to the medium on either side of the membrane – resulting in the occluded states, E₁P(3Na) and/or E₂(2K). Efflux of 3Na⁺ ions from the cell is associated with phosphorylation by ATP followed by the conformational change E₁P–E₂P, while the influx of 2K⁺ ions is associated with rapid dephosphorylation of E₂P and the conformational change E₂(2K) → E₁, which is accelerated by binding of ATP with a low affinity.

In normal cells containing mitochondria, the free energy of ATP hydrolysis is close to 60 kJ mol^{−1} and the electrical driving force of the Na⁺,K⁺-ATPase is close to 600 mV (G/F ; where F is Faraday’s constant (96 485 C mol^{−1})). Most of the electrochemical work (~390 mV) is used for the outward translocation of three Na⁺ ions per cycle against the inside negative membrane potential and the 10-fold-steep chemical gradient for Na⁺ ions. Less energy (~40 mV) is used for the inward transport of two K⁺ ions per cycle since potassium in most cells is close to electrochemical equilibrium.

The stoichiometry of 3Na⁺/2K⁺ transported per ATP hydrolyzed is maintained over a wide range of electrolyte compositions. Due to the unequal quantities of charge transported in opposite directions, the pump produces an electric current, in which a net positive charge is extruded from the cell. This property of the ion pump is called ‘electrogenicity’. Thus, the pump is also sensitive to the membrane potential – the binding and release of one of the three Na⁺ ions is the major voltage-dependent step. Because the Na⁺ and K⁺ transport is sequential, it is possible to study separately the Na⁺- and K⁺-transport pathways and the associated charge movement.

Structure of the Na⁺,K⁺-ATPase

A proper understanding of the question “How does a cation-pump pump cations?” requires knowledge of the molecular structure. Since the essence of active Na⁺ and K⁺ pumping is the coupling of the phosphorylation and dephosphorylation steps to the E₁ and E₂ conformational changes, the crucial issues concern the nature of the ATP and cation-binding sites, and structural events accompanying the conformational changes.

The first structure of the heterodimeric P-type ATPase, Na⁺,K⁺-ATPase, isolated from pig kidney, at 3.5 Å, was published in 2007, 50 years after its discovery. This structure contains α , β , and FXYD2 (γ) subunits but only the α and trans-membrane segments of β and (γ) subunits were resolved. A very recently published crystal structure of the Na⁺,K⁺-ATPase from shark rectal gland at 2.4 Å resolution, also in (2Rb)₂–MgF₄^{2−} conformation (Rb⁺ is an analog of K⁺, MgF₄^{2−} is a phosphate analog), shows the subunits of α , β , and the FXYD protein

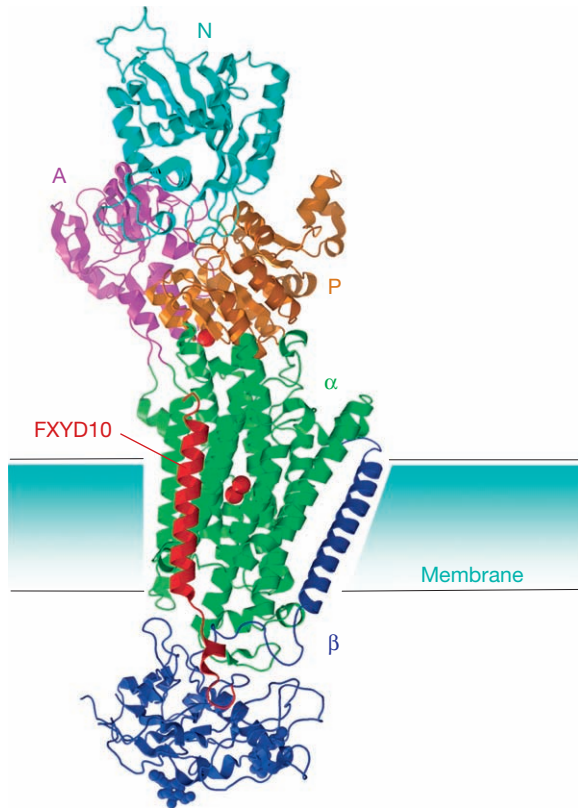
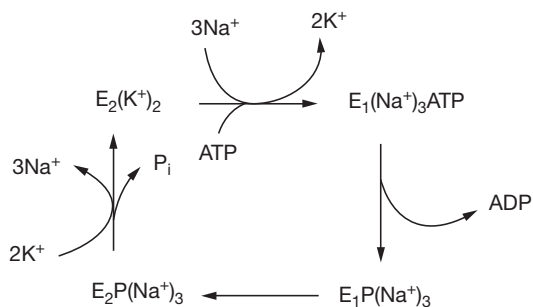


Figure 1 The structure of the shark rectal gland Na^+, K^+ -ATPase, rendered from the protein data bank (PDB) file 2zxe. The trans-membrane helices of the α subunit are colored green and the A, N, and P domains are colored magenta, cyan, and orange, respectively. The β subunit is in blue and FXD10 is red. The red spheres represent bound K^+ ions.



Scheme 1 The Post-Albers scheme.

(referred to as FXD10), their folds, secondary structures and side-chain rotamers, including the resolved extracellular domain of the β and FXD subunits (see Figure 1).

The α Subunit – The Cation-Pumping Mechanism

The α subunit of Na^+, K^+ -ATPase consists of head, stalk, and membrane sectors. The α subunit has 10 trans-membrane segments and the two Rb^+ ions are occluded approximately in the center of trans-membrane segments M4, M5, and M6. Helices

M4 and M6 are unwound in the middle to enable ion binding. The cytoplasmic part of the pump is made up of three distinct α domains: N for nucleotide binding, P for phosphorylating, and A for actuator. These domains are intimately involved in the pumping mechanism. The A-domain, connected to the M1–M3 helices with rather long linkers, works as the actuator of the trans-membrane gating mechanism that regulates cation binding and release. The P-domain contains the phosphorylated aspartate residue, Mg^{2+} -coordinating residues, and other residues that characterize the Na^+, K^+ -ATPase as members of the haloacid dehalogenase superfamily. The A-domain contains one of the signature sequences, the Thr–Gly–Glu–Ser (TGES) motif, which plays an important role in the processing of the aspartylphosphate intermediate. The P-domain has a wedge shape that is a key in the conversion of A-domain rotational movements to the vertical movements (i.e., perpendicular to the membrane) of the trans-membrane helices. The N-domain, formed by a long insertion between two parts of the P-domain, connected by flexible linkers, contains the residues for adenosine binding and those critical for bridging ATP and the P-domain.

Although we are concerned primarily with Na^+, K^+ -ATPase, it was the advent of the molecular structures of the rabbit sarcoplasmic endoplasmic reticulum calcium ATPase (SERCA1) in 2000 that provided the breakthrough in our understanding of the working of all cation pumps. The structure of Ca^{2+} -ATPase has now been determined in all major conformations of the catalytic cycle. Since the basic principles of the cation pumping mechanism of Na^+, K^+ -ATPase in Scheme 1 are shared by all P-type ATPases, a number of general insights have emerged.

In SERCA, the sequence of structural changes in the pump that constitute the different pumping stages are depicted in cartoon form (see Figure 2):

1. In the E_1 conformation, two Ca^{2+} are bound in the high-affinity sites formed by trans-membrane helices M4, M5, M6, and M8. The cytoplasmic gate is open and bound Ca^{2+} exchange with those in the cytoplasm. The M1 helix is deeply embedded in the lipid bilayer.
2. In E_1 , ATP binds and cross-links the P- and N-domains, so that the gamma-phosphate of ATP and a Mg^{2+} bind to the P-domain to bend it. The N-domain is fixed in a highly inclined position and makes contact with the A-domain in a strained position. The M1 helix is pulled up and bent so that the top of the trans-membrane part closes the cytoplasmic gate of the Ca^{2+} -binding sites.
3. Phosphoryl transfer of the active site aspartate allows dissociation of ADP from $\text{E}_1\text{-P}$, which triggers the opening of the N- and P-domain interface; the A-domain rotates so that the TGES motif wedges into the gap and interacts with the phosphorylation site in the P domain. The separation of N and P domains and creation of interactions between the cytoplasmic A and P and A and N domains is a major characteristic of the E_1P to $\text{E}_2\text{-P}$ conformational change.
4. These movements of the cytoplasmic domains are coupled to marked rearrangement of the trans-membrane helices M1–M6; large downward movements of M4, sharp bending of M5, and rotation of M6 destroy the Ca^{2+} -binding sites.

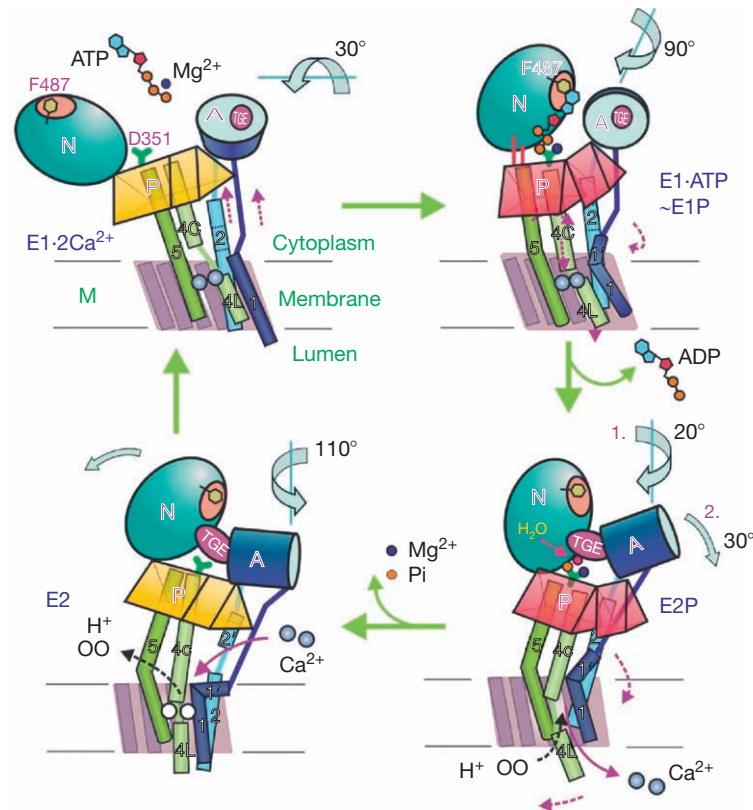


Figure 2 A cartoon illustration of the architecture of SERCA and the structural changes occurring during conformational transitions. See detailed description in text. Adapted from Toyoshima C (2009) How Ca^{2+} -ATPase pumps ions across the sarcoplasmic reticulum membrane. *Biochimica et Biophysica Acta* 1793: 941–946.

The lower sections of M1 and M2 push against M4, opening the luminal gate and releasing the bound Ca^{2+} into the lumen. 2H^+ ions then bind to the $\text{E}_2\text{-P}$ from the extracellular surface.

5. The glutamate residue in the TGES loop of the A-domain fixes a particular water molecule and catalyzes its attack on the aspartylphosphate. The release of the phosphate and Mg^{2+} unbends the P-domain. This in turn releases M1 and M2 so that M4L closes the luminal gate on the bound 2H^+ ions.
6. In the last step of the cycle, the reverse conformational change from E_2 to E_1 , driven by ATP binding to E_2 , brings the cation-binding sites back into contact with the intracellular medium, the N and P domains come into proximity and the A domain is displaced from the P domain.

In summary, the P- and N-domains change interfaces and thereby control the position of the A-domain, the actuator of trans-membrane gates. ATP, phosphate, Mg^{2+} , and Ca^{2+} are the modifiers of the interfaces. Energy barriers between the principal intermediate states seem to be comparable to the thermal energy; the enzyme has evolved to utilize thermal energy as the driving force for the pumping activity.

While the general mechanism of all P-type ATPase must be similar to that in Figure 2, a major difference between Na^+, K^+ -ATPase and Ca^{2+} -ATPase is the selectivity for the transported cations ($3\text{Na}^+/2\text{K}^+$ vs. $2\text{Ca}^{2+}/2\text{H}^+$), including the third

unique Na^+ site in Na^+, K^+ -ATPase. The cation-binding residues in the occlusion sites are similar for the Na^+, K^+ -ATPase and Ca^{2+} -ATPase, except two conserved residues. Apparently, it is differences in the positions of the side chains and water molecules that mainly define the cation selectivity in occlusion. There is also considerable overlap between the residues that coordinate the two K^+ ions in the E_2 structure of Na^+, K^+ -ATPase, and those expected to coordinate the first two Na^+ ions in the E_1 form (corresponding to the residues that bind Ca^{2+} in the E_1 structure of SERCA1), supporting the notion that both cations use the same occlusion cavity. The third sodium transport site, which is a unique feature of the Na^+, K^+ -ATPase is proposed to be located between M6, M8, and M9. The carboxy terminus of the alpha-subunit (KETYY) is contained within a pocket between trans-membrane helices, and may be a novel regulatory element controlling sodium affinity. Another interesting feature is a regulatory K^+ -binding site, located in the P domain, which may play a role in activation of dephosphorylation and stabilization of the dephosphorylated E_2 conformation.

The β Subunit, FXYD Proteins, and Other Insights

The recent structure of the Na^+, K^+ -ATPase provides an answer to one of the biggest conundrums in the field: the structure and interaction site of the β subunit. This cannot be inferred

from the structure of Ca²⁺-ATPase, which lacks a β subunit. The β subunit is visible with its single trans-membrane segment slanted at 32° to the membrane normal, forming numerous contacts with α M10 and M7. The β ectodomain covers the extracellular loops of the α subunit and can be roughly divided into two subdomains, C- and N-terminal, that form many salt bridges with residues in extracellular loops connecting α trans-membrane segments. The C-terminal ectodomain has an immunoglobulin-like fold. The β subunit also has a short intracellular domain, which was not resolved.

A unique feature of the α subunit of the Na⁺,K⁺-ATPase is an unwound part and kink in M7 and at the cytoplasmic end of M10, which interact with the β subunit. It has been proposed that the unwound part of α trans-membrane M7 and the interaction with β is of central importance to stabilize K⁺ binding within the protein, thus providing a possible explanation of the unique requirement of a β subunit in K⁺-transporting ATPases.

In addition to its stabilization role, the ectodomain of β 1 has also been strongly implicated in epithelial cell adhesion, and β 2 in neural adhesion in the rat brain. The C-terminal lobe of the ectodomain, with its immunoglobulin-like fold, may be crucial for the cell adhesion function, perhaps forming hetero- or homodimers with other cell adhesion molecules that, typically, also contain immunoglobulin-like domains involved in protein-protein interactions.

The single trans-membrane helix of the FXDY protein is close to and interacts with α M9. This has an interesting functional implication, since it is near the proposed location of the third Na⁺-binding site between M6, M8, and M9. Thus, interactions of the trans-membrane segments of FXDY proteins may modulate the binding affinity of Na⁺ ions via the unique third Na⁺ site. The α , β , and FXDY subunits also make contact together close to the outer membrane leaflet, where residues in both α and β subunits and the FXDY residues form a hydrophobic pocket, providing a rationale for this conserved motif.

A specifically bound cholesterol molecule was observed in the shark Na⁺, K⁺-ATPase structure. This molecule is fully resolved and resides in a crevice formed by the trans-membrane helices of β , α M7, and M5. It also plays a role in shielding the unwound part trans-membrane of M7 from bulk lipid and is likely to be involved in stabilizing α/β interactions. Cation pumps are embedded in the phospholipid bilayer of the membrane. Thus, the physical properties of the lipid bilayer can play an important role in their function and stability, for example, by variation in bilayer fluidity and so-called hydrophobic matching with the trans-membrane segments of the protein. In addition, there is also substantial evidence for direct interactions with phospholipids, such as phosphatidylserine and also cholesterol. The new structure provides direct evidence for such a direct interaction with cholesterol.

Physiological Functions of Na⁺,K⁺-ATPase

The α 1 β 1 Na⁺,K⁺-ATPase is constitutively expressed in most cells and it maintains the Na⁺/K⁺ gradients driving the active

transcellular transport in kidney, intestine, lung, and various glands. There is a close relationship between the capacity for reabsorption of Na⁺ in these tissues and the abundance of α 1 β 1 Na⁺, K⁺-ATPase in the basolateral membranes of the epithelial cells. The Na⁺,K⁺-ATPase in human kidneys hydrolyzes over 2 kg of ATP per day and over 600 g of sodium is reabsorbed per day. Accordingly, ouabain-binding data reveal 40–50 million Na⁺,K⁺-ATPase pumps per cell in thick ascending limbs of Henle (TALH) or distal convoluted tubules (DCTs) in kidney, as compared to a few hundred in red blood cells or a few thousand pumps in other nonpolarized cells. In all tubular cells, the Na⁺,K⁺-ATPase maintains electrochemical gradients for sodium ($\Delta\mu_{\text{Na}^+}$), the driving force for secondary active transport of other solutes and ions, such as H⁺, Ca²⁺, and Cl[−]. In the TALH, the Na⁺,K⁺-ATPase functions in concert with the Na⁺,K⁺,2Cl[−] cotransporter and channels for K⁺ or Cl[−] to bring about transcellular NaCl transport. NaCl is reabsorbed at very high rates, while luminal and cytoplasmic Na⁺ concentrations remain relatively high; 20–40 mM in the lumen of TALH. In TALH the γ -subunit (FXDY2) is expressed with the α 1 β 1 Na⁺,K⁺-ATPase. The α 1 β 1 Na⁺,K⁺-ATPase complex has a relatively low affinity for [Na⁺]_{cyt} and it is expressed at very high densities in the basolateral membranes of the TALH.

In the inner medullary collecting duct (IMCD), Na⁺,K⁺-ATPase functions are integrated with those of the sodium channel, epithelial Na channel (ENaC), and K⁺-channels to drive reabsorption of Na⁺ under the regulation of aldosterone. In the IMCD, the expression of FXDY4 or corticosteroid-induced factor (CHIF) is stimulated by aldosterone and the CHIF- α 1 β 1 Na⁺,K⁺-ATPase pump complex has a relatively high affinity for [Na⁺]_{cyt}, thus enabling the pump to reduce the concentration of Na⁺ to 2–5 mM in the tubular lumen and in the cytoplasm of the principal cells of the IMCD. Aldosterone also regulates Na⁺ reabsorption driven by α 1 β 1 Na⁺,K⁺-ATPase in epithelia of lung, glands, and colon, thus providing a major force in the regulation of Na⁺ metabolism and extracellular fluid volume.

Extensive analyses have been performed to clarify the tissue-specific gene expression and the molecular mechanisms of regulation of the α and β -subunit genes (Table 1). Experiments on animal models allow reverse genetic strategy to analyze the function of transport proteins. Targeted disruption of the α 1- and α 2-subunit isoforms in mice confirms that the Na⁺, K⁺-ATPase is essential for life of all mammalian cells. Egg cells lacking both copies of the ubiquitous α 1-isoform do not pass the blastocyst stage. Homozygote knockouts of the α 2-isoform results in death few minutes after birth of the mice apparently due to malformation of neural connections in the medulla oblongata. Knockout of one of the genes shows that α 1- and α 2-isoforms have different physiological functions. An increased force of contraction of the heart is seen in animals lacking one copy of the α 2-isoform, whereas the force of contraction is reduced after elimination of one copy of the α 1-isoform. There is a correlation between the expression of the α 2-Na⁺, K⁺-ATPase, and the Na⁺/Ca²⁺ exchanger in plasma-membrane domains overlaying the endoplasmic reticulum (ER). Elevation of local [Na⁺]_{cyt} may therefore increase the [Ca²⁺]_{cyt} concentration in the space of the cytosol near the ER and thus amplify the force of contraction. This may explain the specific role

of the α_2 -isoform as a regulator of calcium and the force of contraction in the heart.

In skeletal muscle, long-term regulation involves thyroid hormone-mediated changes of the expression level of Na⁺,K⁺-ATPase. Short-term regulation of the Na⁺,K⁺-ATPase is mediated either by changes in intracellular Na⁺ concentration or by hormone-mediated protein kinase reactions leading to changes of the Na⁺,K⁺-ATPase transport properties or in its cell-surface expression. Insulin causes translocation of α_2 Na⁺,K⁺-ATPase to the cell surface, and the α_2 Na⁺,K⁺-ATPase in skeletal muscle is involved in specific regulation in response to hypokalemia. Exercise stimulates exocytosis of α_1 - and α_2 Na⁺,K⁺-ATPase to the plasma membranes of muscle cells and this response is important for recovery of K⁺ after repetitive muscle contractions. Phospholemman (FXYP1) interacts with α_1 and α_2 Na⁺,K⁺-ATPase and modulates their properties to cause a decrease in the affinity to Na⁺. The presence of Na⁺,K⁺-ATPase with low affinity can also be important for the recovery process after muscle contractions. Skeletal muscles are the main intracellular store of K⁺ and the Na⁺,K⁺-ATPase in the sarcolemma (along with H⁺,K⁺-ATPase) is important for maintaining plasma K⁺ concentration in kidney and colon.

The isoforms α_1 – α_3 are expressed in the central nervous system. During embryogenesis, the α_2 Na⁺,K⁺-ATPase is expressed in most regions of the brain and is important in the modulation of neuronal activity in the neonate. In the adult, α_2 Na⁺,K⁺-ATPase is expressed primarily in astrocytes, where glutamate is transported in a sodium- and potassium-dependent fashion. FXYP7 has been identified as an isoform-specific regulator of the Na⁺,K⁺-ATPase in the brain. The properties of α_3 Na⁺,K⁺-ATPase pumps may allow reuptake of K⁺ at low extracellular activities in the brain.

Cardiac Glycosides

Cardiac glycosides are a family of steroids that bind and inhibit the Na⁺,K⁺-ATPase with high selectivity and affinity. Traditionally, these substances were extracted from plants such as *Digitalis purpurea* or foxglove (digitalis), *Strophanthus gratus* (ouabain) or amphibian skin (Bufo marinus-bufalin, marinobufagenin). Ouabain (Somali-waabaayo) was used as an arrow poison in hunting and war. Cardiac glycosides are produced endogenously in mammals as well as plants. Several different cardiac glycosides have been identified in mammalian tissues including the cardenolides ouabain and digoxin, and several bufodienolides including marinobufagenin (see Figure 3). Endogenous cardiac glycosides are intimately involved in control of blood pressure and renal natriuresis. Endogenous

ouabain is synthesized from steroid precursors in the zona glomerulosa and zona fasciculata cells of the adrenal cortex, similar to other steroid hormones.

Cardiac glycosides, particularly digoxin, have been used in the clinic for over 200 years. At therapeutic levels, they exert a positive inotropic effect (an increase in contractile force) on the heart muscle, thus improving circulation in cases of insufficient cardiac output. The positive inotropic mechanism of cardiac glycosides is thought to be mediated by a local increase of Na⁺ concentration in cardiac myocytes due to inhibition of the Na⁺,K⁺-ATPase. This local increase of [Na⁺]_{cyt} reduces extrusion of Ca²⁺ by the Na⁺/Ca²⁺ exchanger, leading to an increase in cytoplasmic Ca²⁺ concentration. Consequently, SERCA-mediated Ca²⁺ load in the ER is enhanced, and more Ca²⁺ is released from the ER into the cytoplasm following excitation. Through the actin/myosin system, the higher cytoplasmic Ca²⁺ causes an increase in force of contraction of cardiac muscle. Digoxin also increases vagal activity through its action on the central nervous system, thus decreasing the conduction of electrical impulses through the atrioventricular (AV) node. This negative chronotropic effect (slowing of heart rate) is important for its clinical use in different arrhythmias.

Experiments with genetically engineered mice support a central role of the α_2 isoform in mediating the inotropic effect of cardiac glycosides and preferential modulation of Ca²⁺ transients and SR Ca²⁺ release in cardiac myocytes, compared to α_1 . Digoxin, the principal cardiac glycoside used in the clinic today, has a narrow therapeutic window due to toxicity caused by Ca²⁺-overload. This overload, produced by excessive inhibition of the Na⁺,K⁺-ATPase in the myocytes, leads to arrhythmia and is potentially lethal. Thus, an α_2 -specific cardiac glycoside could exert the positive inotropic effect without the untoward side-effect of Ca²⁺ overload due to general cardiac Na⁺,K⁺-ATPase inhibition. Recent work shows that digoxin is itself moderately α_2 -selective.

A recently published high-resolution crystal structure of ouabain bound to the shark rectal gland Na⁺,K⁺-ATPase in the E₂(2 K) conformation provides insight into the inhibition mechanism of cardiac glycosides (Figure 4). In this structure, ouabain is seen to bind parallel to the membrane normal, wedged deep in a cavity formed by M1–M5, with its lactone ring embedded deep into the trans-membrane segments, in close proximity to the two occluded K⁺ ions. The structure shows how ouabain can block the ion pathway thereby preventing any transport. However, it is known that K⁺ and ouabain are antagonistic, and thus this model represents a low-affinity structure. A model of the high-affinity conformation, E₂P, suggests that high affinity for ouabain is created by the movement of the M1–M2 trans-membrane segments that forms

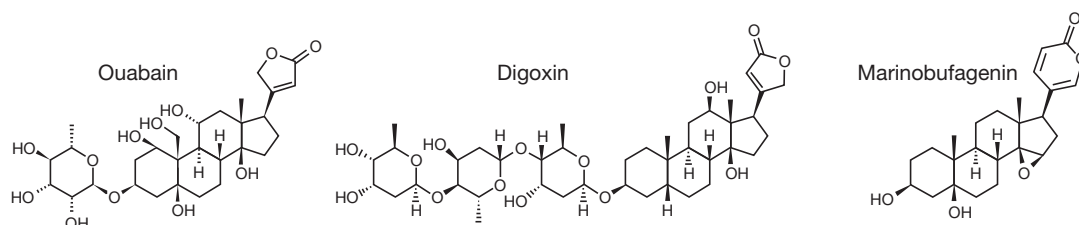


Figure 3 The chemical structures of ouabain, digoxin, and marinobufagenin.

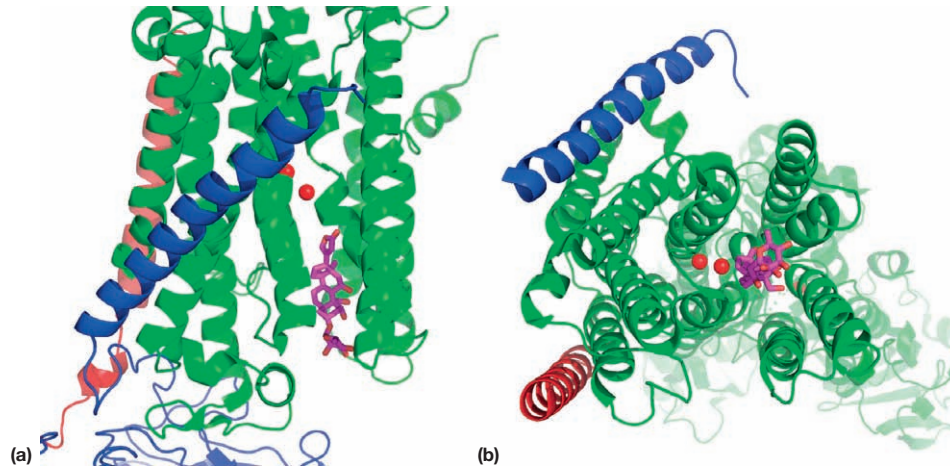


Figure 4 The structure of the shark rectal gland Na^+ , K^+ -ATPase in the $\text{E}_2(2\text{K})$ conformation with bound ouabain, rendered from the PDB file 3a3y. The coloring scheme is similar to Figure 1. Ouabain is rendered in magenta sticks. (a) side view of the trans-membrane stalk and (b) view into the cavity formed by M1–M5 from the extracellular direction.

a complementary surface with the bound steroid and prevents its dissociation. This complementary interface cannot be optimally formed in the presence of occluded K^+ that prevents the closure of the cavity formed by M1–M5 on the ouabain molecule, thus explaining the classical K-ouabain antagonism.

In addition to their classical roles as inhibitors of active Na^+ and K^+ transport, there is now compelling evidence that endogenous cardiac glycosides are hormones involved in signal transduction processes, and affect cell growth and proliferation, the Na^+ , K^+ -ATPase acting as the hormone receptor. One signal pathway shown to be involved in kidney and cardiac cell proliferation is ouabain-induced activation of phosphatidylinositol-3-kinase (PI3K) that interacts at the N-terminal end of α subunit, and then Ca^{2+} -dependent phosphorylation of Akt (protein kinase B). Another signaling pathway that ouabain was found to activate is the Src-EGFR-Ras-Raf-Erk cascade that controls gene expression, differentiation, and apoptosis. These effects of ouabain may involve protein–protein interactions, and the signaling complexes are located in microdomains or caveolae, similar to other cellular-signaling complexes. In cancer cells, these signaling pathways may also be involved in effects of cardiac glycosides to stimulate apoptosis, suggesting the possibility of use of cardiac glycosides in anti-cancer therapy – new uses for an old drug. In short, endogenous cardiac glycosides serve as regulators of blood pressure, as well as hormones that influence gene expression, exerting their effects through various cellular pathways, whether in direct response to the changes in Na^+ concentration or through protein–protein interaction networks.

See also: **Bioenergetics:** Membrane Transporters: Na^+ / Ca^{2+} Exchangers; Voltage-Dependent K^+ Channels.

Further Reading

- Blaustein MP, Hamlyn JM, and Pallone TL (2007) Sodium pumps: Ouabain, ion transport, and signaling in hypertension. *American Journal of Physiology. Renal Physiology* 293: F438; author reply F439.
- Garty H and Karlisch SJ (2006) Role of FXYD proteins in ion transport. *Annual Review of Physiology* 68: 431–459.
- Geering K (2008) Functional roles of Na, K-ATPase subunits. *Current Opinion in Nephrology and Hypertension* 17: 526–532.
- Jorgensen PL, Hakansson KO, and Karlisch SJ (2003) Structure and mechanism of Na, K-ATPase: Functional sites and their interactions. *Annual Review of Physiology* 65: 817–849.
- Kaplan JH (2002) Biochemistry of Na, K-ATPase. *Annual Review of Biochemistry* 71: 511–535.
- Morth JP, Pedersen BP, Toustrup-Jensen, et al. (2007) Crystal structure of the sodium–potassium pump. *Nature* 450: 1043–1049.
- Ogawa H, Shinoda T, Cornelius F, and Toyoshima C (2009) Crystal structure of the sodium–potassium pump (Na^+ , K^+ -ATPase) with bound potassium and ouabain. *Proceedings of the National Academy of Sciences of the United States of America* 106: 13742–13747.
- Schoner W and Scheiner-Bobis G (2007) Endogenous and exogenous cardiac glycosides: Their roles in hypertension, salt metabolism, and cell growth. *American Journal of Physiology Cell Physiology* 293: C509–C536.
- Shinoda T, Ogawa H, Cornelius F, and Toyoshima C (2009) Crystal structure of the sodium–potassium pump at 2.4 Å resolution. *Nature* 459: 446–450.
- Toyoshima C (2009) How Ca^{2+} -ATPase pumps ions across the sarcoplasmic reticulum membrane. *Biochimica et Biophysica Acta* 1793: 941–946.
- Xie Z and Askari A (2002) Na^+ / K^+ -ATPase as a signal transducer. *European Journal of Biochemistry* 269: 2434–2439.

Relevant Websites

- <http://www.mark-hilge.com> – A website with a good animation of the pumping cycle.
- <http://www.brenda-enzymes.info> – The sodium pump at the BRENDA enzyme data bank website.
- <http://enzyme.expasy.org> – The sodium pump at the SIB Bioinformatics Resource Portal.
- <http://www.pdb.org> – The sodium pump molecule of the month at the Protein Data Bank website.

P-Type Pumps: Plasma-Membrane H⁺ Pumps

A B Mason and C W Slayman, Yale University, New Haven, CT, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Fusicoccin A compound toxic to plants that is produced by the fungus *Fusicoccum amygdali*.

Glycolysis regulation 1 (Gcr1) Yeast transcription factor that acts together with Rap1 to increase expression of some glycolytic enzymes.

Proteasome A large cytosolic protease complex that is responsible for intracellular protein degradation.

Translation upstream factor 1/Repressor activator protein 1/General regulatory factor 1 (Tuf1/Rap1/Grf1) Yeast transcription factor involved in a variety of cellular processes including telomere regulation, chromatin silencing, and glycolysis.

Plasma-membrane H⁺ pumps (also known as H⁺-ATPases) are widely distributed throughout the fungal and plant kingdoms. They use the energy from adenosine triphosphate (ATP) hydrolysis to generate a sizable electrochemical gradient across the surface membrane, which in turn drives nutrient uptake into the cell via a series of H⁺-coupled co-transporters (**Figure 1 (a)**). The proton-pumping activity of the adenosine triphosphatase (ATPase) also contributes to the regulation of intracellular pH. The genes encoding representative members of the family have been cloned from budding (*Saccharomyces cerevisiae*) and fission (*Schizosaccharomyces pombe*) yeast, *Neurospora crassa*, several pathogenic fungi, and numerous plant species including *Arabidopsis thaliana*, *Nicotiana plumbaginifolia* (tobacco), and *Lycopersicon esculentum* (tomato).

Introduction

The fungal and plant plasma-membrane H⁺-ATPases belong to a ubiquitous family of cation pumps known as P-type ATPases, so named because they hydrolyze ATP via a covalent aspartyl-phosphate reaction intermediate. When the P-ATPases are classified based on sequence alignments, membrane topology, and the nature of the transported cation(s), H⁺-ATPases fall into a category designated as P₂ or type III.

Since 1986, when the *PMA1* genes of *S. cerevisiae* and *N. crassa* were cloned and sequenced, these two H⁺-ATPases have served as prototypes for the group as a whole. Yeast, in particular, has proven to be an extremely useful model system. A feature of considerable value in the molecular characterization of the yeast *PMA1* ATPase is its abundance in the plasma membrane, where it constitutes up to 25% of total protein. By virtue of the ease with which site-directed yeast mutants can be made and analyzed, an extensive collection of *pma1* mutants has been generated covering the entire length of the protein, and this resource continues to provide molecular details as to the structure and function of the H⁺-ATPase. Finally, the use of yeast as a vehicle for heterologous expression of other ATPases, such as the individual isoforms of plant H⁺-ATPase, has enabled researchers to study specific features of these enzymes that could not be analyzed in their natural setting. Recently, such studies culminated in the publication of the first high-resolution crystal structure of a P-type proton pump (see section **Structure**).

Just as the P-type ATPases of animal cells are of paramount importance to cellular homeostasis, fungal and plant H⁺-ATPases are essential enzymes whose function is fundamental to the viability of the host organism. As a cell surface component that is common to all fungi examined to date, the H⁺-ATPase has potential clinical significance as a target for antifungal drug development. Specific, high-affinity inhibitors would most likely be fungicidal and may not even be required to enter the cell, thus avoiding multidrug resistance (MDR) efflux pumps and reducing the likelihood of toxic effects on host cells. Plant H⁺-ATPases perform a multiplicity of tissue-specific roles that include nutrient uptake in roots, opening of leaf stomata, seed coat development, and pollen formation. A better understanding of the structure, function, and regulation of these enzymes will have far-reaching effects of agricultural and economic significance.

Structure

At the structural level, the fungal and plant plasma-membrane H⁺-ATPases are 100-kDa polypeptides that are deeply embedded in the lipid bilayer, requiring detergents for solubilization. The N- and C-termini of the H⁺-ATPase are both located in the cytoplasm. As in other P₂-type ATPases, each H⁺-ATPase has 10 hydrophobic transmembrane α -helices, four (M1–M4) at the N-terminal end and six (M5–M10) at the C-terminal end. This membrane (M) domain anchors the large (almost 400 amino acids) cytoplasmic loop in the center of the polypeptide containing the critical catalytic sites that bind ATP and form the aspartyl-phosphate intermediate.

The first three-dimensional (3D) glimpse of H⁺-ATPase structure came from cryoelectron microscopic maps of the *N. crassa* fungal enzyme at 8-Å resolution. Over the past few years, this low-resolution view has been extensively updated by comparison with X-ray crystallographic structures of the related mammalian sarcoplasmic reticulum Ca²⁺-ATPase (SERCA1a), for which there are now more than 20 different conformational snapshots. Although the level of amino acid identity between the H⁺-ATPases and rabbit SERCA1a is only around 20%, Kuhlbrandt and colleagues demonstrated that the amino acid sequence of the *Neurospora* H⁺-ATPase could be overlaid upon the Ca²⁺-ATPase template to give a 3D

structure with the same three well-defined cytoplasmic regions representing nucleotide-binding (N), phosphorylation (P), and actuator (A) domains. Much more recently, Pedersen and colleagues compared a 3.6 Å structure of the *A. thaliana* H⁺-ATPase (AHA2) enzyme with equivalent forms of the SERCA1a enzyme, revealing a strikingly similar folding pattern that likely reflects a structure conserved even between different subfamilies of P-type ATPases.

Reaction Mechanism

Like the intensively studied mammalian Na⁺-, K⁺-, and Ca²⁺-ATPases, the fungal and plant plasma-membrane H⁺-ATPases are considered to alternate between two major conformational states (E₁ and E₂) during their reaction cycle (Figure 1(b)). The scale and complexity of domain movements have now been comprehensively mapped in a series of crystal structures of the SERCA1a ATPase. This information, supplemented with details from the AHA2 structure that suggest an elegant and highly plausible model for proton transport across the membrane, is presented in a much simplified form in Figure 2.

Starting with the transient E₂ open conformation, protonation of the membrane-embedded acidic residue that acts as the critical acceptor/donor residue in the proposed H⁺ pathway (Asp⁶⁸⁴ in AHA2) is followed by nucleotide binding in the presence of Mg²⁺ as metal cofactor (E₁). This produces conformational changes that lead to phosphorylation of Asp³²⁹ in

AHA2 (or Asp³⁷⁸ in yeast PMA1) and closure of the structure, bringing the N and A domains into close contact (E₁P). Subsequent rotation of the A domain occurs during the E₁P to E₂P transition, opening a pathway for exit of the H⁺ into extracellular space. This allows for dephosphorylation of the P domain and transition back to the vanadate-sensitive E₂ form to complete the cycle. The physical scale of domain movements made during the reaction cycle is impressive. Measurements based on SERCA 1a structures indicate that the top of the N domain moves more than 50 Å in approaching the A domain, which then rotates 90° to interact with the P domain via a highly conserved TGES motif. Correct positioning and orientation of the A domain on the interactive surface of the wedge-shaped P domain depends on movements of the transmembrane helices perpendicular to the plane of the membrane that are transduced by flexing of the P domain itself.

Detailed information from the plant AHA2 structure suggests a transmembrane proton transport mechanism involving a water-filled cavity that is formed by partial unwinding of M4, and lined with polar or charged side chains contributed by conserved residues in M5 and M6. The proton enters the pathway between M1, M2, and M4, subject to a gating mechanism that relies on hydrogen bonding between Asn¹⁰⁶ and Asp⁶⁸⁴. Asp⁶⁸⁴, a residue located in the center of the membrane in all H⁺-ATPases and known to be essential for H⁺ translocation, is the likely proton-binding site. Upon de-protonation, Asp⁶⁸⁴ is

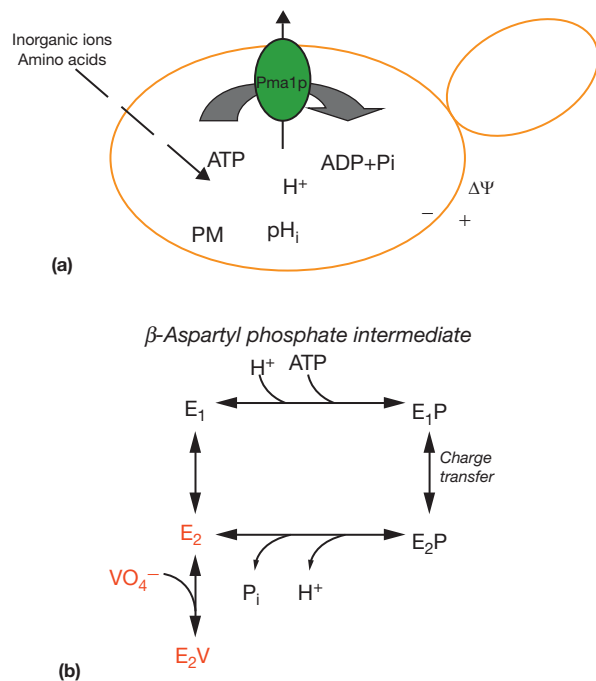


Figure 1 (a) The plasma-membrane H⁺-ATPase plays an essential role in yeast cell physiology. The pumping of protons (H⁺) across the membrane at the expense of ATP hydrolysis establishes the membrane potential (ΔΨ) required for nutrient import and helps to regulate intracellular pH. PM, plasma membrane. (b) The H⁺-ATPase reaction cycle. See text for explanation. VO₄⁻, vanadate, a potent P-type ATPase inhibitor.

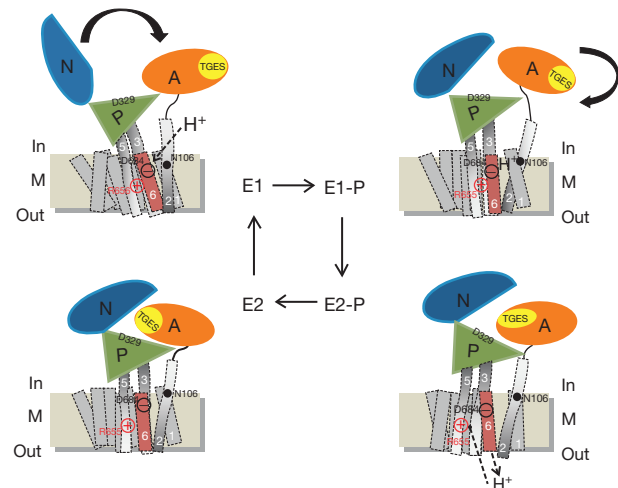


Figure 2 (A) Major domain movements that occur during conversion of the ATPase from the E₁ (cation-bound) to the E₂ (cation-free) conformation. All 10 transmembrane helices are shown but only five are numbered (white). M6, the helix shown in red, includes the embedded proton-binding Asp⁶⁸⁴ residue (—) at the center of a putative H⁺ pathway (see text for further details). Specific residues, including the phosphorylated Asp³²⁹, correspond to those of the plant H⁺-ATPase AHA2. During the E₂-P to E₂ transition, the conserved TGES amino acid motif is involved in dephosphorylation of the aspartyl phosphate (D329) and protection of the unoccupied nucleotide-binding site. N, nucleotide-binding domain; P, phosphorylation domain; A, actuator domain; M, membrane sector formed by 10 transmembrane helices. This cartoon depicts the reaction cycle of a plant H⁺-ATPase (AHA2) but incorporates information derived from the domain movements of the P-type Ca²⁺-ATPase of muscle sarcoplasmic reticulum, as determined by Toyoshima and colleagues.

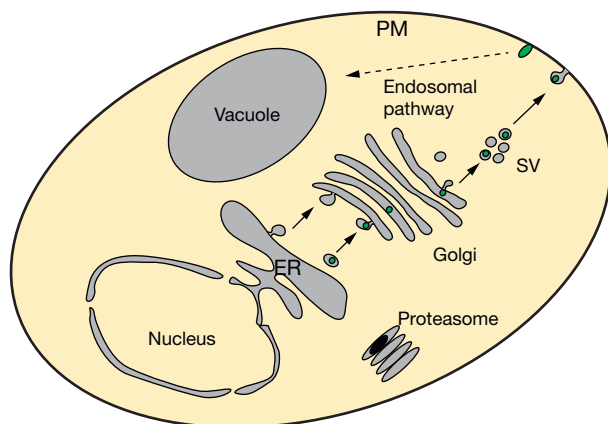


Figure 3 The yeast secretory pathway. ER, endoplasmic reticulum; SV, secretory vesicles; PM, plasma membrane.

protected from re-protonation by the close proximity of another conserved charged residue (Arg⁶⁵⁵), effectively preventing any retrograde proton transport. The positive charge also serves to stabilize the negatively charged residue, keeping the H⁺ pathway closed. This feature probably accounts for the extremely high membrane potentials generated by plant and fungal H⁺-ATPases compared to their mammalian cation-transporting counterparts. Protons, most likely in the form of hydronium ion (H₃O⁺) complexes, are conducted rapidly across the membrane along proton wires formed by a H-bonded network of water molecules.

Biogenesis

Biogenesis of the H⁺-ATPase begins with its synthesis in the rough endoplasmic reticulum (ER), from which it travels to the plasma membrane via the secretory pathway (Figure 3). Because of its abundance in the membrane, together with the ease of making site-directed mutants, the yeast H⁺-ATPase has served as a valuable model for studying this process.

While still in the ER, the ATPase adopts a fully folded structure that has the protease-resistant characteristics of the mature enzyme at the plasma membrane. The folded molecule is then incorporated into coat protein complex II (COPII) vesicles and shuttled to the *cis*-Golgi. Of particular relevance, a sub-population of COPII vesicles has been identified that preferentially recruits and packages the H⁺-ATPase by virtue of a specific coat protein, Lst1p. At least in principle, this mechanism provides a way to regulate biogenesis, underscoring the significance of the ATPase to the functioning of the cell as a whole.

There is much interest in the structure of the ATPase as it exits the ER. Oligomeric forms of the *Neurospora* H⁺-ATPase have been known for many years and were originally thought to be by-products of detergent treatment. More recent studies, however, have demonstrated the existence of oligomers of the yeast H⁺-ATPase early in biogenesis by three different methods: blue native gel electrophoresis, sucrose gradient centrifugation, and cross-linking of epitope-tagged polypeptides. Oligomer formation also provides a logical explanation for

the effects of dominant-negative *pma1* mutants, which cause co-expressed wild-type H⁺-ATPase to be retained in the ER. The functional significance of the oligomers is not yet clear, given an earlier demonstration that the *Neurospora* H⁺-ATPase is fully active when solubilized and reconstituted as a monomer. There has been speculation that the oligomers may play a special role in the biogenesis and/or storage of the ATPase but neither case has been proven definitively.

Successful transport of the yeast H⁺-ATPase to the plasma membrane depends on its incorporation into detergent-resistant lipid rafts, sterol- and sphingolipid-rich structures into which certain plasma-membrane proteins are segregated en route to the cell surface. Wild-type ATPase becomes associated with rafts prior to exiting the ER and failure to do so compromises its trafficking and stability. The key lipid factor in this process appears to be neither sterol nor sphingolipid *per se*, but rather the saturated C26 very long-chain fatty acid in ceramide, a major component of yeast sphingolipids. Substitution of C26 with shorter fatty acids destabilizes the lipid-ATPase complex, leading to mistargeting and degradation of the ATPase in the vacuole. These findings reinforce the importance of lipid-lipid and lipid-protein interactions for the functional organization of the plasma membrane and the entire lipid raft concept.

Many mutations in the *PMA1* gene that lead to misfolding of the H⁺-ATPase result in arrest of the polypeptide in the ER and massive accumulation of internal membranes. This represents the first level of quality control in biogenesis and, in most instances, is followed by re-direction of the offending ATPase via the ER-associated degradation (ERAD) pathway for digestion in the proteasome. Quality control screening is maintained, however, as the H⁺-ATPase progresses through the secretory pathway. There are several examples of mutant ATPases that escape ER quality control only to be recognized as aberrant and differentially processed at later stages of the secretory pathway by vacuolar proteases. These include mutants with as few as two amino acid substitutions and some with short C-terminal truncations. Although these forms are apparently delivered normally to the plasma membrane, they are short-lived compared to the wild-type enzyme, becoming ubiquitinated and then processed by the endocytic pathway for delivery to the vacuole. Information contained within a stretch of 20 amino acids toward the C-terminus of the yeast ATPase seems to be critical for ER export and stability of the enzyme once it reaches the plasma membrane. It is also apparent that the ATPase is not uniformly distributed throughout the plasma membrane but is restricted to certain compartments, segregated from some other protein components of the membrane such as the proton symporters Can1p and Fur1p.

Regulation

In addition to the major H⁺-ATPase gene (*PMA1*), *S. cerevisiae* and *S. pombe* each have a second gene (*PMA2*) that encodes a nearly identical protein. The expression level of *PMA2* is so poor that it is unable to support growth; however, the physiological function of the second ATPase remains unclear.

By contrast, plants possess multiple H⁺-ATPase genes. In *A. thaliana*, whole-genome sequencing has identified a family

of 11 such genes, most or all of which are expressed in a tissue-specific manner. Where tested, the corresponding H⁺-ATPase isoforms differ detectably in their catalytic behavior, suggesting that enzyme characteristics may be matched to the physiological requirements of individual tissues. While there is undoubtedly some overlap between tissue distribution and ATPase isoforms, there is strong evidence that particular *A. thaliana* H⁺-ATPase isoforms have tissue-specific roles in stomatal opening and closing (AHA1), pollen development (AHA3), and seed coat development (AHA10).

A diversity of environmental conditions and stress factors influence the biosynthesis and the activity of fungal and plant plasma-membrane H⁺-ATPases. In yeast, two mechanisms have been identified that both result in upregulation of the H⁺-ATPase by glucose. At the transcriptional level, addition of glucose to the growth medium leads to a rapid increase in messenger RNA (mRNA) for the H⁺-ATPase, and analysis of the promoter region of the *PMA1* gene has revealed active binding sites for several recognized transcription factors including translation upstream factor 1/repressor activator protein 1/general regulatory factor 1 (Tuf1/Rap1/Grf1) and glycolysis regulation 1 (Gcr1). It is almost certainly significant that the same factors promote transcription of genes encoding glycolytic enzymes, pointing to a coordinated set of mechanisms that enable rapid, efficient growth on glucose.

Because the yeast H⁺-ATPase has an unusually long half-life at the plasma membrane (~11 h), posttranslational mechanisms are required to modulate its activity rapidly. Indeed, within minutes of glucose addition, the yeast ATPase displays an increased V_{max}, a lowered K_m, and a more alkaline pH optimum. It has been confirmed that this ensemble of changes results from posttranslational phosphorylation near the C-terminus of the 100 kDa polypeptide. Initially, detailed mutational analyses linked the glucose-induced change in V_{max} with amino acids Arg⁹⁰⁹ and Thr⁹¹² and the change in K_m with Ser⁸⁹⁹ and Glu⁹⁰¹. Both pairs of residues constitute potential phosphorylation sites and both *in vitro* and *in vivo* evidence suggests that yeast protein kinase 2 (Ptk2)-mediated phosphorylation of Ser⁸⁹⁹ is responsible for the glucose-regulated K_m change. With respect to changes in V_{max}, the application of cutting-edge techniques in mass spectrometry has demonstrated that the sequential phosphorylation of Ser⁹¹¹ and Thr⁹¹² is directly linked with glucose-dependent activation of the ATPase. Phosphorylation of these two tandem residues is required for full activation of the ATPase in response to glucose; therefore, *in vivo*, the ATPase exists in three forms relative to C-terminal phosphorylation – P0, P1, and P2 (glucose activated). Truncation of the C-terminus leads to constitutive activation, suggesting that the domain normally behaves in an auto-inhibitory fashion. Studies using site-directed mutagenesis of the yeast H⁺-ATPase, coupled with suppressor analysis, suggest that the C-terminus interacts with the stalk region near the cytoplasmic surface of the membrane and also with the ATP-binding (N) domain.

In plants, a variety of physiological conditions, such as light, hormones, and salt concentration, can affect the expression of specific ATPase isoforms but precise transcriptional mechanisms have not yet been elucidated. As in yeast, however, the primary modulator of plant H⁺-ATPase activity is the C-terminus (or regulatory R domain) of the enzyme acting in

an auto-inhibitory fashion. In the plant H⁺-ATPase, the penultimate C-terminal residue (Thr⁹⁴⁷ in AHA2) is a phosphorylation target that, when modified, is bound by regulatory 14-3-3 proteins, a highly conserved family of acidic proteins expressed ubiquitously in eukaryotic cells. The 14-3-3 proteins bind as dimers, simultaneously activating the H⁺-ATPase and driving it into a dodecameric quaternary structure consisting of six ATPase units and six 14-3-3 molecules. The mode of action of the fungal toxin fusaric acid is to bind irreversibly to the H⁺-ATPase/14-3-3 complex, constitutively activating the enzyme such that leaf stomata are unable to close, causing the plants to wilt. The R-domain of the plant H⁺-ATPase, which is about 60 amino acids longer than that of the yeast enzyme, includes two short inhibitory sequences, each ~20 residues in length. Removal or mutagenesis of these sequences activates the ATPase, presumably by preventing the sequences from interacting with other sites on the enzyme and/or by improving accessibility of the C terminus for phosphorylation and subsequent 14-3-3 binding. As observed in yeast, mutations at sites distal to the C-terminus of the plant H⁺-ATPase can activate the enzyme. Since these mutations map to the actuator and phosphorylation domains it is likely that these domains house the inhibitory binding sites targeted by the C terminus.

Regulation of H⁺-ATPase activity is achieved through the coordinated action of multiple pathways. Our appreciation of the significance and complexity of this network will increase as our ability to explore cell physiology on a genome-, transcriptome- and proteome-wide basis improves.

Summary

Research into the structure, function, biogenesis, and regulation of fungal and plant plasma-membrane H⁺ pumps continues to provide a wealth of scientific information about the physiology of their host cells. Moreover, these pumps are invaluable as protein models not solely for P₂-type ATPases but for complex integral membrane proteins in general.

See also: **Bioenergetics:** ER/SR Calcium Pump: Function; P-Type Pumps: H⁺/K⁺ Pump; P-Type Pumps: Na⁺/K⁺-ATPase; Structure of P-Type Adenosine Triphosphatases; **Lipids Carbohydrates Membranes and Membrane Proteins:** Lipid Bilayer Structure; Lipid Rafts; Secretory Pathway.

Further Reading

- Buch-Pedersen MJ, Pedersen BP, Veierskov B, Nissen P, and Palmgren MG (2009) Protons and how they are transported by proton pumps. *Pflügers Archives* 457: 573–579.
- Duby G and Boutry M (2009) The plant plasma membrane proton pump ATPase: A highly regulated P-type ATPase with multiple physiological roles. *Pflügers Archives* 457: 645–655.
- Ferreira T, Mason AB, and Slayman CW (2001) The yeast Pma1 proton pump: A model for understanding the biogenesis of plasma membrane proteins. *Journal of Biological Chemistry* 276: 29613–29616.
- Kühlbrandt W, Zeelen J, and Dietrich J (2002) Structure, mechanism, and regulation of the *Neurospora* plasma membrane H⁺-ATPase. *Science* 297: 1692–1696.
- Lecchi S, Allen KE, Pardo JP, Mason AB, and Slayman CW (2005) Conformational changes of yeast plasma membrane H⁺-ATPase during activation by glucose:

- Role of threonine-912 in the carboxy-terminal tail. *Biochemistry* 44: 16624–16632.
- Lecchi S, Nelson CJ, Allen KE, et al. (2007) Tandem phosphorylation of Ser-911 and Thr-912 at the C-terminus of yeast plasma-membrane H⁺-ATPase leads to glucose-dependent activation. *Journal of Biological Chemistry* 282: 35471–35481.
- Lingwood D and Simons K (2010) Lipid rafts as a membrane-organizing principle. *Science* 327: 46–50.
- Morsomme P, Slayman CW, and Goffeau A (2000) Mutagenic study of the structure, function, and biogenesis of the yeast plasma membrane H⁺-ATPase. *Biochimica et Biophysica Acta* 1469: 133–157.
- Pedersen BP, Buch-Pedersen MJ, Morth JP, Palmgren MG, and Nissen P (2007) Crystal structure of the plasma membrane proton pump. *Nature* 450: 1111–1115.
- Schekman R (2002) SEC mutants and the secretory apparatus. *Nature Medicine* 8: 1055–1058.
- Toyoshima C (2009) How Ca²⁺-ATPase pumps ions across the sarcoplasmic reticulum membrane. *Biochimica et Biophysica Acta* 1793: 941–946.
- Toyoshima C, Nakasako M, Nomura H, and Ogawa H (2000) Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* 405: 647–655.
- Yatime L, Buch-Pedersen MJ, Musgaard M, et al. (2009) P-type ATPases as drug targets: Tools for medicine and science. *Biochimica et Biophysica Acta* 1787: 207–220.

Relevant Websites

- <http://www.pumpkin.au.dk> – Centre for Membrane Pumps in Cells and Disease, Aarhus University, Denmark.
- <http://www.iam.u-tokyo.ac.jp> – Toyoshima Lab, University of Tokyo, Japan.

Purple Bacteria: Electron Acceptors and Donors

A Adessi and R De Philippis, University of Florence, Florence, Italy

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Evaldas Katilius and Neal W. Woodbury, volume 3, pp. 582–585, © 2004, Elsevier Inc.

Glossary

Bacteriochlorophyll *a* Photosynthetic pigment composed of a tetrapyrrolic ring coordinating a Mg^{2+} ion, present in purple bacteria to operate anoxygenic photosynthesis.

Bacteriopheophytin *a* A molecule of bacteriochlorophyll *a* missing the central Mg^{2+} ion.

Intracytoplasmic membranes (ICMs) System of internal membranes originated from and continuous with the plasmatic membrane, organized in fingerlike intrusions, lamellae, tubules, or vesicles depending on bacterial species.

Membrane P and N faces The positive (P) and the negative (N) faces of the cell membrane, namely the sides that face the phase characterized by a positive or a negative potential.

Midpoint potential The midpoint potential, or $E_{m,7}$, is the potential at which the concentrations of the oxidized and reduced forms of the redox pair are equal, at pH 7.

Ubiquinone One of a class of aromatic diketones in which the carbon atoms of the carbonyl groups are part of the ring structure. This oil-soluble substance is present in plants, animals, and microorganisms as it is a component of the electron transport chain.

Introduction

Purple bacteria form a heterogeneous group of microorganisms capable of growing under anoxic conditions by anoxygenic photosynthesis. They can be divided into purple nonsulfur bacteria, which are able to grow both phototrophically and in darkness, and purple sulfur bacteria, all of them capable to grow in the light but only few of them in the dark. The recent publication of the first complete genome sequence of a purple nonsulfur bacterium, *Rhodospseudomonas palustris*, pointed out the metabolic versatility of this group of bacteria. Indeed, it was found that the genome contains all the genes needed for switching from chemotrophy to phototrophy, from organotrophy to litotrophy, and from heterotrophy to autotrophy. Under anoxic light conditions, *R. palustris* can use either the reducing power deriving from organic compounds, such as organic acids or aromatic compounds, or the electrons deriving from inorganic compounds, such as H_2 or $S_2O_3^{2-}$. When growing chemotrophically in the presence of oxygen, respiration occurs and the reducing power can derive, as well as under anoxic light conditions, from either organic or inorganic compounds. When neither light nor oxygen is present, anaerobic respiration can take place, using NO_3^- or N_2O as final electron acceptors. Moreover, purple sulfur bacteria are able to use sulfide as an electron donor, oxidizing it to elemental sulfur or even to sulfate, while only some nonsulfur bacteria can use S^{2-} , but with different enzymatic pathways.

Due to the remarkable complexity of the metabolism of purple bacteria, this article mainly focuses on the main electron acceptors and donors involved in the photosynthesis and in the respiration.

The Electron Transport Chain

The main stages of both photosynthesis and respiration processes occur in membranes. Therefore, membranes are the site

of adenosine triphosphate (ATP) generation for purple bacteria and in both processes ATP is formed as a result of the proton motive force generated along with electron translocation through the electron transport chain (ETC).

These two processes, however, take place in different sites of the purple bacterial membrane: the cytoplasmic membrane (CM) harbors the components of the respiratory apparatus, while the photosynthetic apparatus is located in the so-called intracytoplasmic membranes, which originate from the invagination of the CM and can take different forms throughout the purple bacteria group. Although oxygenic respiration and anoxygenic photosynthesis would appear to be two distinct and very different processes, they have many common issues.

Anoxygenic Photosynthesis

The specificity of purple bacteria is their ability to form their energy carrier (ATP) in the absence of oxygen by using sunlight as a source of energy. In this process, bacteriochlorophylls are both the primary electron donors and the final electron acceptors, as anoxygenic photosynthesis is a cycle (Figure 1).

A photon is absorbed by the light-harvesting complexes that funnel the excitation toward bacteriochlorophylls in the reaction center (RC) and charge separation occurs. This energy is used for the release of an electron which reduces the quinone Q. Once the quinone is doubly reduced (i.e., after a second photon is captured), it picks up protons from the cytoplasmic space and translocates through the membrane to reach the cytochrome bc_1 complex; here, electrons are addressed to the cytochrome c_2 (Cyt c_2), while protons are released in the periplasmic space. Cyt c_2 is then able to reduce the oxidized primary electron donors in the RC, thus closing the cycle. The protons accumulated in the periplasm form an electrochemical gradient which is used by the ATP synthase to generate ATP.

The electron cycle is opened only by the 'Δp-driven reversed electron transport' (indicated by the thin blue arrows in Figure 1), that is by the action of both the reduced

nicotinamide adenine dinucleotide (NADH) dehydrogenase working in the reversed way to reduce NAD^+ to NADH, and the succinate dehydrogenase also working backwards reducing fumarate to succinate. As these two enzymes are able to catalyze both the forward and reverse reactions, the force that drives the direction is the presence or absence of the products; in particular, the reversed reactions are ways to get rid of the possible excess of the reduced quinol. The reversed NADH dehydrogenase reaction is also the way to refurbish the cell with NADH-reducing equivalents.

Respiration

In the presence of O_2 , anoxygenic photosynthesis in purple bacteria is inhibited and the ATP is synthesized through cellular respiration (Figure 2).

The reducing power carried by NADH is transferred, by the action of NADH dehydrogenase, to a quinone. The quinone can also be reduced by the succinate dehydrogenase, but in this process there is no proton translocation. The reduced quinone migrates through the membrane and can be oxidized both by

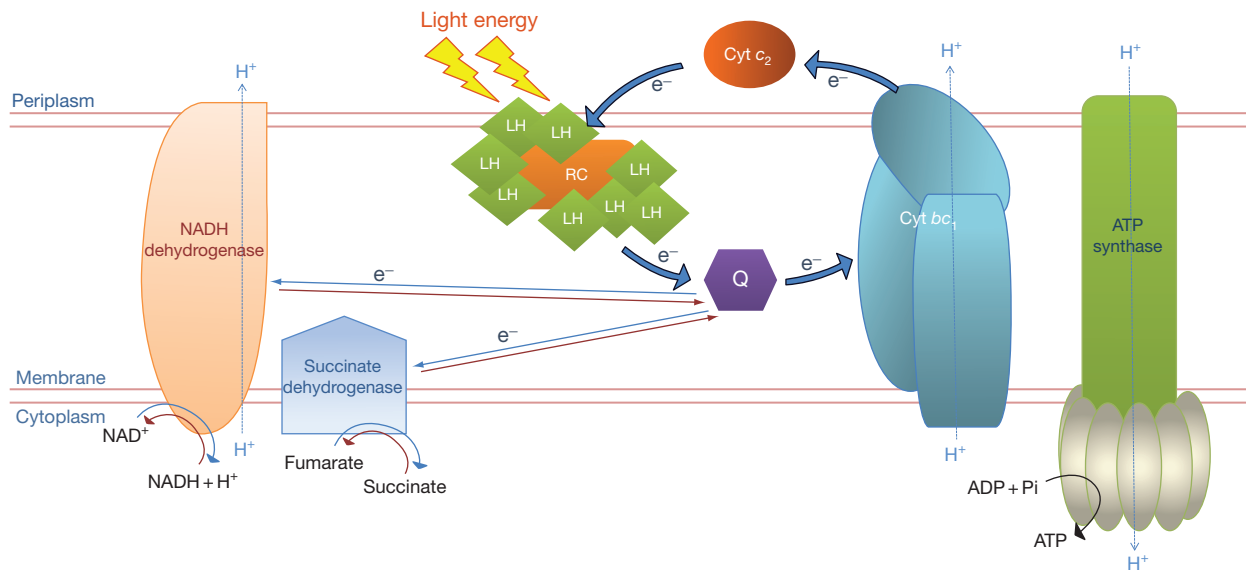


Figure 1 Schematic representation of the photosynthetic ETC. RC, reaction center, LH, light-harvesting complexes; Q, quinones; Cyt bc_1 , cytochrome bc_1 complex; Cyt c_2 , cytochrome c_2 . The photosynthetic electron cycle is indicated by the blue bold arrows. The Δp -driven reversed electron transport is indicated by the thin blue arrows; the red arrows indicate the oxidation of NADH and succinate and the reduction of quinones. The dotted arrows indicate the steps for proton translocation.

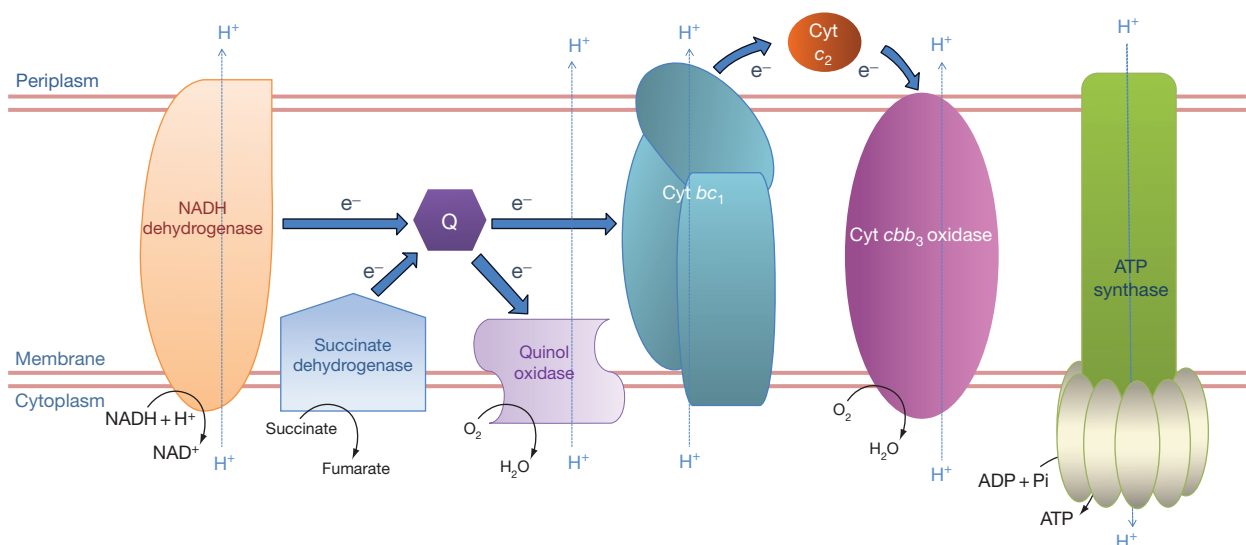


Figure 2 Schematic representation of the respiratory ETC. Q, quinones; Cyt bc_1 , cytochrome bc_1 complex; Cyt c_2 , cytochrome c_2 ; Cyt cbb_3 oxidase, cytochrome cbb_3 oxidase. The blue bold arrows indicate the electron flow as occurs in respiration. The dotted arrows indicate the steps for proton translocation.

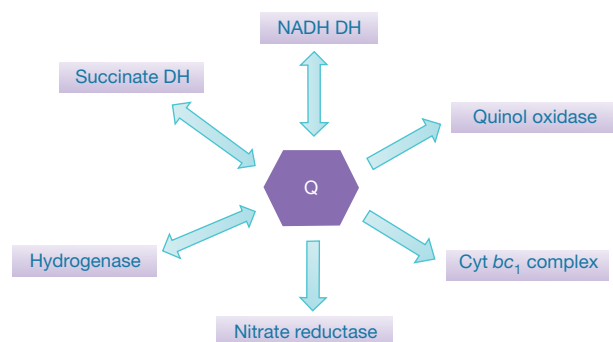


Figure 3 The quinone pool functions. Q, quinones; Cyt bc_1 complex, cytochrome bc_1 complex; NADH DH, NADH dehydrogenase; succinate DH, succinate dehydrogenase.

the quinol oxidase, with oxygen as the electron acceptor, or by the bc_1 complex, that functions (exactly as in photosynthesis) as a proton-translocating step and transfers the electrons to the Cyt c_2 . This soluble protein is, in respiration, oxidized by the Cyt cbb_3 oxidase which drives the reducing power to a molecule of O_2 . As in photosynthesis, the protons accumulated in the periplasm form an electrochemical gradient which is used by the ATP synthase to generate ATP.

The Role of the Quinone Pool

As discussed above, purple bacteria have a very versatile metabolism. The regulation of the different metabolic behaviors is headed by the RegA/RegB system, which regulates photosynthesis, respiration, and nitrogen and carbon fixation. This regulon gathers an integrated signal inclusive of environmental and cellular stimulations.

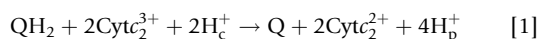
An important integrative signal coming from the cell is its redox state, which undergoes different values depending on the different metabolic conditions. This signal is mediated by the quinone pool, which holds a crucial role for the energetic processes in the cell, as it is schematically shown in Figure 3.

Electron Acceptors and Donors Common in Photosynthesis and Respiration

All the above-described processes take place by means of prosthetic groups and protein cofactors.

The Ubiquinol: Cytochrome c Oxidoreductase (or Cytochrome bc_1 Complex)

The ubiquinol:cytochrome c oxidoreductase is part, as previously shown, of both photosynthetic and respiratory ETC as its role is to oxidize the lipid-soluble quinol and transfer its electrons to c -type cytochromes. The overall reaction is the following (eqn [1]), where QH_2 indicates the reduced quinol, while H_c^+ and H_p^+ indicate protons in the cytoplasmic and in the periplasmic space, respectively:



As cytochromes are able to accept only one electron at a time, quinones have to release their reducing equivalents one by one, and this shift from two electrons to a one-electron process is managed by Cyt bc_1 thanks to its architecture and cofactors.

This protein complex is a dimer and each monomer is composed of three subunits and contains four redox-active metal centers forming two distinct ETC: a low potential chain that transfers electron from the Q_P (P = membrane P face) site to the Q_N (N = membrane N face) site through two b -type hemes, namely Cyt b_L (L = low potential) and Cyt b_H (H = high potential), and a high-potential chain, formed by a $[2Fe2S]$ cluster and a c -type heme covalently bound to the protein (Cyt c_1), that makes electrons flow from the Q_P site to the oxidized Cyt c_2 .

The quinol (QH_2) binds to the Q_P site where a bifurcation occurs. The electron that takes the low-potential chain easily passes from the Cyt b_L to the Cyt b_H , as the first has a midpoint potential of about -90 mV and the second of about $+50$ mV. This high-energetic step is necessary to compensate the otherwise unfavorable passage of a negative charge through the membrane from the P to the N face. Close to the Cyt b_H site is located the Q_N site, where a quinone is reduced to semiquinone by the reduced Cyt b_H . This process, which is unfavorable due to this potential difference (from $+50$ to -180 mV), occurs due to the protein-quinone interactions, which make the quinone potential more positive. The semiquinone is held in the Q_N site until a second electron takes the low-potential chain and reduces it to QH_2 .

The electron that takes the high-potential chain reduces the $[2Fe2S]$ cluster, which, in this state, shows a higher affinity to the Cyt c_1 that is almost $25 \text{ \AA}$ away. This affinity makes the protein literally turn close to Cyt c_1 , so that the electron can reach the c_1 -heme. The oxidized $[2Fe2S]$ cluster is then more affine to the Q_P site and turns close to it. Midpoint potentials have been measured for the $[2Fe2S]$ cluster, but the values depend dramatically on the occupant of the Q_P site; with the native Q bond, the midpoint potential is about $+320$ mV; Cyt c_1 has a midpoint potential of about $+230$ mV.

The reduced Cyt c_1 is able to transfer its electron to the Cyt c_2 that would close the photosynthetic cycle reducing the RC's P dimer.

Cytochromes

Cytochrome c_2 is a single-heme soluble periplasm-located protein and acts as an electron deliverer. It accepts electrons from the Cyt c_1 of the bc_1 complex and is present both in the photosynthetic process and in respiration, respectively, it reduces the P^+ dimer to P , which is thus able to start a new photosynthetic turn, or reduces the cytochrome cbb_3 oxidase and in some organisms (*Rhodobacter sphaeroides*) also the cytochrome aa_3 oxidase.

The underlying process for electron transfer at this step is Cyt c_2 docking: many models have been proposed and many kinetic studies have been carried out to determine the Cyt c_2 -RC interaction; it can be said that the docking is possible mainly due to electrostatic interactions, but short-range interactions and H-bonds play a role as well.

After the protein docking, the most likely electron-transfer pathway proceeds from the heme edge via Tyr L162 to the

bacteriochlorophyll dimer, in a microsecond-scale process. The heme midpoint potential varies between +280 and +360 mV, depending on purple bacteria strains.

The RC of *Blastochloris viridis* contains a nondissociating tetraheme cytochrome molecule that facilitates the electrons flowing from Cyt c_2 . The electron transfer between the cytochrome and primary donor is about 3 times faster in *Blc. viridis* than in *R. sphaeroides*. This may be due to a different path length (the distance between conjugated atoms on the heme and primary donor is 12.3 Å in *B. viridis* compared to 14.2 Å in *R. sphaeroides*).

Cytochromes other than Cyt c_2 are synthesized by purple bacteria in different metabolic phases: isocytochrome c_2 and cytochrome c_7 (the latter membrane-bound) are present preferably in the presence of O_2 , while Cyt c_2 transcription is increased in anoxic conditions, thus revealing that Cyt c_2 is mainly involved in photosynthesis, while iCyt c_2 and Cyt c_7 are involved in respiratory pathways.

NADH Dehydrogenase

NADH dehydrogenase is an L-shaped transmembrane complex, the largest of the bacterial electron-transfer complexes as is composed of 14 subunits; electron flow is mediated by a flavin mononucleotide (FMN) cofactor and 9 Fe/S centers, two [2Fe2S] and seven [4Fe4S] as described for *Thermus thermophilus*; one of the tetranuclear clusters is prokaryote specific.

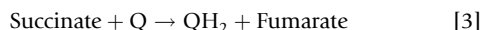
This complex transfers electrons from NADH (−320 mV) to ubiquinone (+90 mV). The primary electron acceptor is FMN (−380 mV), then the electron passes through the linear pathway created by seven [FeS] clusters. The potentials of six of them vary between −300 and −230 mV; the other has a pH-dependent midpoint potential ranging between −20 and −160 mV and is probably the last to be reduced. This enzyme mediates the overall reaction [2] by which four cytoplasmic protons are released in the periplasmic space:



This enzyme can also work in the reverse direction, in particular in phototrophic growth conditions in presence of a high QH_2/Q ratio, to synthesize NADH-reducing equivalents.

Succinate Dehydrogenase

The succinate dehydrogenase is a membrane-bound flavoprotein that is part of both the ETC and the tricarboxylic cycle; it transfers electrons from succinate to quinones, as indicated in reaction [3], and no proton translocation occurs.



This enzyme contains different cofactors that create a track for electrons; namely, a flavin adenine dinucleotide molecule, a [2Fe2S] cluster, a [4Fe4S] cluster, a [3Fe3S] cluster, and a heme b . Depending on the cell energetic needs, the enzyme can also catalyze the reverse reaction, as can happen during photosynthesis (see Figure 1).

Electron Acceptors and Donors Specific of Photosynthesis

The Electron Transfers in the RC

The RC of purple bacteria is an integral membrane protein complex composed of L, M, and H subunits. The H subunit contributes to the complex stability and is involved in proton binding and transfer, while L and M compose the core of the RC and embed nine cofactors, symmetrically disposed in an A and a B branch as shown in Figure 4: two bacteriochlorophyll a molecules that form a dimer (P) connecting the two branches, two monomeric bacteriochlorophyll a molecules (B_A and B_B), two bacteriopheophytin a molecules (H_A and H_B), two ubiquinone molecules (Q_A and Q_B), and an iron ion. Also, some authors suggest the presence of a carotenoid molecule located next to the B_B monomer. Bacteriochlorophylls and bacteriopheophytins can be substituted by some other molecules of the same class.

The midpoint potential of the cofactors is a critical point due to their own nature of electron transfer. The one directly measured is the midpoint potential of the P dimer, that is approximately +500 mV for the P^+/P couple. This positive potential makes it an optimal electron acceptor for Cyt c_2 .

When a photon is conveyed to the P dimer, an electron is raised to a higher-energy level in a femtosecond timescale; as Mg^{2+} is not able to release an electron, the electron comes from the tetrapyrrolic rings of the P dimer. The P^+/P^* couple has a very negative midpoint potential and is very unstable.

The excited dimer P^* transfers the electron, selectively through the A branch, to B_A (in about 3.5 ps, in *R. sphaeroides*). Then, the electron is transferred to H_A (in 0.9 ps, in *R. sphaeroides*); the overall transfer from P^* to H_A takes from 3 to 5 ps, depending on bacterial species. The electron transfer from H_A^- to Q_A takes 200 ps, while the reduction of Q_B to QH_2 takes place much more slowly, in about 100 μs. The oxidized P^+ is reduced back to P by an electron donated by Cyt c_2 : this transfer takes from 0.5 to 50 μs.

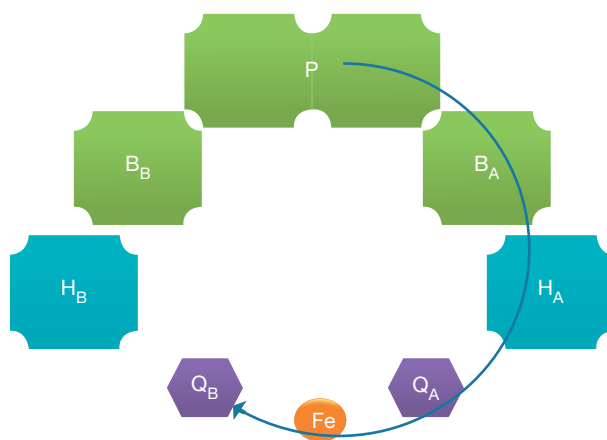


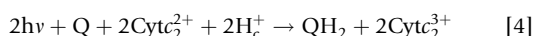
Figure 4 Schematic representation of the RC organization. P, dimer of bacteriochlorophyll a ; B_A and B_B , monomeric bacteriochlorophyll a molecules; H_A and H_B , bacteriopheophytin a molecules; Q_A and Q_B , ubiquinone molecules; Fe, iron ion. The arrow shows the electron path.

Purple bacteria export reducing equivalents in pair from the RC, and the molecule migrating from the RC to the other membrane proteins is the reduced quinol (QH₂).

Two electrons are needed to reduce the quinone to quinol, so, as the RC transfers electrons one by one, two quinones (Q_A and Q_B) are needed and act in a cycle in order to accumulate two reducing equivalents; after two RC turnovers a QH₂ is released from the Q_B site.

Q_A and Q_B have different features: Q_A is very tightly bound in its site and can be under the quinone and the free-radical anionic semiquinone form (Q_A^{•−}), while Q_B is reversibly bound and can be under the quinone, semiquinone, and quinol form within the cycle. What makes the main difference between Q_A and Q_B (that in some species are chemically identical – for instance, in *R. sphaeroides*) is the interaction with the protein in their binding sites.

The Q_A^{•−}/Q_A couple has a midpoint potential of about −180 mV. Then, the Q_A^{•−} migrates to the Q_B site and here the second electron reduces it to Q^{2−}. The protons are captured from the cytoplasm during the acceptor quinone cycle, and will be released in the periplasm by the cytochrome *bc*₁ complex to form the proton gradient needed for ATP synthesis. The overall reaction can be resumed as follows:



QH₂ is enough lipophilic to freely move throughout the membrane and transports its reducing power to different membrane-bound enzymes; in photosynthesis, it can reach not only the cytochrome *bc*₁ complex but also NADH dehydrogenase and succinate dehydrogenase, thus opening the cyclic photosynthetic process with the ‘reversed electron flow’. The reverse action of the NADH dehydrogenase is necessary to refurnish the cell of NADH.

Electron Acceptors and Donors Specific of Respiration Oxidases

Evidences have shown that purple bacteria can contain different types of terminal oxidases: some species (e.g., *R. sphaeroides*) contain the cytochrome *aa*₃ oxidase, the cytochrome *cbb*₃ oxidase, and a *bb*₃-type quinol oxidase; some other species (e.g., *R. capsulatus*) contain only the cytochrome *cbb*₃ oxidase and a *bb*₃-type quinol oxidase. In microorganisms that possess both cytochrome oxidases, the *aa*₃-type is predominant when the cells are grown aerobically, while the *cbb*₃-type is expressed mainly under microaerobic or photosynthetic conditions; otherwise, the *cbb*₃-type is the only cytochrome oxidase present.

All these proteins are part of the heme-copper oxidase (HCO) superfamily, but have significant differences in structure, in mechanisms, and in the quantity and quality of cofactors. For example, *aa*₃-type oxidases contain a binuclear copper cofactor named copper A (Cu_A), which is absent both in *cbb*₃-type and in quinol oxidases.

As Cyt *aa*₃ oxidase is not always present in purple bacteria, it is not discussed in this section.

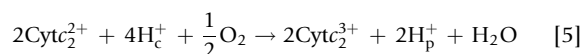
Cytochrome *cbb*₃ oxidase

Cyt *cbb*₃ oxidase, a member of the HCO superfamily, is constituted by four subunits and contains many cofactors: two

Table 1 Hemes of the Cyt *cbb*₃ oxidase with their midpoint potentials (data refer to *Pseudomonas stutzeri* Cyt *cbb*₃ oxidase)

Redox center	Midpoint potentials (mV)
heme <i>b</i>	+310/+265
heme <i>b</i> ₃	+225
heme <i>c</i>	+245/+215
heme <i>c</i>	+245/+205
heme <i>c</i>	+185/+105

b-type hemes, three *c*-type hemes, and a copper B (Cu_B). Cyt *cbb*₃ oxidase passes electrons from the Cyt *c*₂ to O₂-pumping protons in the periplasm following the reaction:



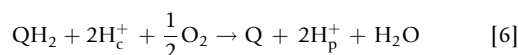
Cyt *cbb*₃ oxidase is composed by a catalytic subunit containing the active site formed of a high-spin iron atom coordinated by a *b*₃ heme and a Cu_B. Electrons are transferred to the active site by a second, low-spin, *b*-type heme.

Electrons get to the catalytic subunit through a path created by rising potential *c*-type hemes: soluble Cyt *c*₂ binds close to the first *c*-type heme and the electron flows through and reaches the active site. Potentials have been measured for the five hemes of a Cyt *cbb*₃ oxidase of *Pseudomonas stutzeri*; in Table 1 those values are reported.

As already mentioned, this enzyme can be present at low levels during photosynthesis that in purple bacteria takes place under anaerobic conditions. This evidence suggested another enzymatic function, that is the reduction of substrates other than oxygen.

Quinol oxidase

Quinol oxidase operates in order to equilibrate the QH₂/Q ratio, oxidizing quinols to quinones as indicated in the following reaction:



The catalytic subunit is organized as in Cyt *cbb*₃ oxidase and the reaction mechanism is the same, but the difference stands in the absence of the *c*-type hemes.

This enzyme pumps protons in the periplasmic space but less efficiently than the *bc*₁–*cbb*₃ system, the reason for what this oxidase is mainly expressed under microaerophilic condition to avoid an overreduction of the quinone pool.

It was shown, in *R. capsulatus*, that Cyt *cbb*₃ oxidase is a low-O₂-affinity oxidase operating under conditions of high-O₂ concentration, while quinol oxidase is a high-O₂-affinity oxidase, operating under conditions of low-O₂ concentration. In this way, it is possible that all the available oxygen is used as an electron acceptor before switching to lower-yielding acceptors such as nitrate, dimethylsulfoxide (DMSO) or arsenate in the anaerobic respiration process.

Many purple photosynthetic bacteria are able to use electron acceptors other than oxygen: some conventional substrates such as sulfur and nitrogen compounds, and some exotic substrates such as DMSO, trimethylamine-*N*-oxide, and even arsenate and halogenated aromatics. All the enzymes

Table 2 Enzymes involved in denitrification with their respective reactions and the midpoint potentials of the redox couples

Enzymes involved	Reactions	Midpoint potentials (mV)
Nar ^a or Nap ^b	$\text{NO}_3^- \rightarrow \text{NO}_2^-$	+420
Cu-Nir ^c or <i>cd</i> ₁ -Nir ^d	$\text{NO}_2^- \rightarrow \text{NO}$	+370
cNor ^e	$\text{NO} \rightarrow \text{N}_2\text{O}$	+1100
Nos ^f	$\text{N}_2\text{O} \rightarrow \text{N}_2$	+1350

^aNar, membrane-bound nitrate reductase.^bNap, periplasmic nitrate reductase.^cCu-Nir, copper-containing nitrite reductase.^d*cd*₁-Nir, heme containing Nir.^ecNor, cytochrome *c* oxidase NO reductase;^fNos, nitrous oxidoreductase.

catalyzing these reactions contain complex cofactors, such as molybdoptertine, [FeS] clusters, tetraheme cytochromes.

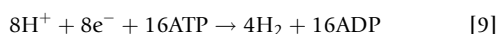
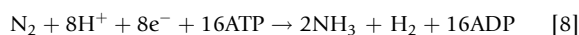
Other Electron Acceptors and Donors

Hydrogen

Hydrogen can be an electron donor for purple bacteria, oxidized by a membrane-bound hydrogenase, which functions as a ubiquinone oxydoreductase. The below reaction can take place in both directions, depending on the presence or absence of the substrates.



However, H₂ can also be produced in purple bacteria by nitrogenase which not only functions primarily as nitrogen-fixing enzyme (eqn [8]), but also functions (in the absence of molecular N₂) using protons as electron acceptors (eqn [9]) to dissipate the excess of reducing power in the cell:



For these reasons, the hydrogenase and nitrogenase are often synthesized together and are regulated by complex and integrated signals.

Nitrogen Compounds

Purple bacteria can use oxidized nitrogen compounds as electron acceptors, carrying out denitrification. Many are complete denitrifiers, while some are so-called partial denitrifiers, because of the four enzymes required for the nitrate reduction to N₂ not all are ubiquitous. Generally speaking, purple bacteria contain most commonly the periplasmic nitrate reductase (Nap), the copper-containing nitrite reductase (Cu-Nir), the cytochrome *c* oxidase NO reductase (cNor), and the nitrous oxidoreductase (Nos); a few contain the membrane-bound nitrate reductase (Nar), and the heme-containing Nir (*cd*₁-Nir), but none contains the quinol-oxidizing Nor (qNor). Table 2 summarizes the steps of denitrification as it occurs in purple bacteria.

Taking into account the already-mentioned capability of reducing molecular nitrogen to ammonia, N₂ can be

Table 3 Differences in substrates and end products between purple sulfur and nonsulfur bacteria

	Most frequent substrates	Most frequent end products
Purple nonsulfur bacteria		
Alpha-proteobacteria	Sulfide, thiosulfate	Sulfate, elemental sulfur
Beta-proteobacteria	No existing data	No existing data
Aerobic anoxygenic phototrophic bacteria ^a	Sulfide, thiosulfate, sulfite, elemental sulfur	Sulfate
Purple sulfur bacteria		
<i>Ectothiorhodospiraceae</i>	Sulfide, elemental sulfur, thiosulfate, sulfite (rare)	Sulfate
<i>Chromatiaceae</i>	Sulfide, elemental sulfur, thiosulfate ^b , sulfite ^b	Sulfate

^aSulfur oxidation does not occur in phototrophic conditions; only in the presence of organic substrates.^bNot used by all *Chromatiaceae*.

considered an electron acceptor, even if the process has an assimilatory aim.

Sulfur Compounds

Many anoxygenic phototrophic purple bacteria use inorganic sulfur compounds as electron donors for reductive carbon dioxide fixation during photolithoautotrophic growth.

The behavior of sulfur and nonsulfur bacteria is slightly different, as indicated in Table 3.

See also: **Bioenergetics:** Algal Hydrogen Production; Cytochrome Oxidases, Bacterial; Hydrogenases, Structure and Function; Purple Bacteria: Photosynthetic Reaction Centers; Quinones.

Further Reading

- Basak N and Das D (2007) The prospect of purple non-sulfur (PNS) photosynthetic bacteria for hydrogen production: The present state of the art. *World Journal of Microbiology and Biotechnology* 23: 31–42.
- Berry EA, Lee DW, Huang LS, and Daldal F (2008) Structure of the cytochrome *bc*₁ complex. In: Hunter CD, Daldal F, Thurnauer MC, and Beatty JT (eds.) *Advances in Photosynthesis and Respiration: The Purple Phototrophic Bacteria*, vol. 28, pp. 425–450. Dordrecht: Springer.
- Imhoff JF (2006) The phototrophic alpha-proteobacteria. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, and Stackebrandt E (eds.) *The Prokaryotes: Proteobacteria: Alpha and Beta Subclasses*, vol. 5, pp. 41–64. New York, NY: Springer.
- Klamt S, Grammel H, Straube R, Ghosh R, and Gilles ED (2008) Modeling the electron transport chain of purple non-sulfur bacteria. *Molecular Systems Biology* 4: 156.
- Larimer FW, Chain P, Hauser L, et al. (2004) Complete genome sequence of the metabolic versatile photosynthetic bacterium *Rhodospseudomonas palustris*. *Nature Biotechnology* 22: 55–61.
- Sander J and Dahl C (2008) Metabolism of inorganic sulfur compounds in purple bacteria. In: Hunter CD, Daldal F, Thurnauer MC, and Beatty JT (eds.) *Advances in Photosynthesis and Respiration: The Purple Phototrophic Bacteria*, vol. 28, pp. 595–622. Dordrecht: Springer.

- Shapleigh JP (2008) Dissimilatory and assimilatory nitrate reduction in the purple photosynthetic bacteria. In: Hunter CD, Daldal F, Thurnauer MC, and Beatty JT (eds.) *Advances in Photosynthesis and Respiration: The Purple Phototrophic Bacteria*, vol. 28, pp. 623–642. Dordrecht: Springer.
- Williams JC and Allen JP (2008) Directed modification of reaction centers from purple bacteria. In: Hunter CD, Daldal F, Thurnauer MC, and Beatty JT (eds.) *Advances in Photosynthesis and Respiration: The Purple Phototrophic Bacteria*, vol. 28, pp. 337–353. Dordrecht: Springer.
- Zannoni D (1995) Aerobic and anaerobic electron transport chains in anoxygenic phototrophic bacteria. In: Blankenship RE, Madigan MT, and Bauer CE (eds.) *Advances in Photosynthesis and Respiration: Anoxygenic Photosynthetic Bacteria*, vol. 2, pp. 949–971. Dordrecht: Springer.
- Zannoni D, Schoepp-Cothenet B, and Hosler J (2008) Respiration and respiratory complexes. In: Hunter CD, Daldal F, Thurnauer MC, and Beatty JT (eds.) *Advances in Photosynthesis and Respiration: The Purple Phototrophic Bacteria*, vol. 28, pp. 537–561. Dordrecht: Springer.

Purple Bacteria: Photosynthetic Reaction Centers

C R D Lancaster, Saarland University, Homburg (Saar), Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

Cytochrome c_2 A heme-containing hydrophilic protein that functions as an electron carrier in both photosynthesis and respiration.

Photosynthetic reaction center A membrane-intrinsic protein complex that carries out the primary photochemical event, the charge-separation process.

Purple bacteria Prokaryotes that catalyze photosynthetic electron transfer. Divided into

Rhodospirillaceae (nonsulfur) and *Chromatiaceae* (sulfur).

Ubiquinone (formerly called coenzyme Q) A quinone derivative with a variable length side chain of isoprene units (mostly 8–10). It occurs in the lipid core of the eukaryotic inner mitochondrial membranes and of bacterial cytoplasmic membranes and functions as a mobile, lipophilic carrier of electrons (and protons).

Photosynthetic Reaction Centers

Virtually all life on Earth ultimately depends on the ability of photosynthetic organisms to convert solar energy into biochemically amenable energy. Photosynthesis is performed by higher plants, algae, and photosynthetic bacteria. Photosynthetic organisms absorb photons for the generation and storage of energy and for the production of biomass. A large proportion of photosynthetically active organisms consists of anoxygenic photosynthetic bacteria. In contrast to the higher plants, algae, and cyanobacteria of oxygenic photosynthesis, which contain the two membrane-bound photosystems (PS) I and II, each of the anoxygenic photosynthetic bacteria has only one type of reaction center (RC). Although the iron–sulfur type RCs of heliobacteria and anaerobic green sulfur bacteria resemble that of PSI, the pheophytin–quinone type RCs of purple bacteria are more similar to the RC of PSII.

Purple bacteria find their ecological niche in deeper layers of stagnant bodies of water. In all purple bacteria, the photosynthetic pigments and the photosynthetic apparatus are located within a more or less extended system of invaginated intracytoplasmic membranes. Unlike PSII, however, the purple bacterial RC is incapable of extracting electrons from water; instead, it must oxidize inorganic or organic molecules available in the environment. According to their electron-donor requirements, sulfur and nonsulfur purple bacteria have traditionally been distinguished. In contrast to sulfur purple bacteria (*Chromatiaceae*, *Ectothiorhodospira*), nonsulfur purple bacteria (*Rhodospirillaceae*) do not require inorganic sulfur compounds, such as hydrogen sulfide, but instead use organic electron donors such as malate or succinate as electron donors. Most of what is known today about purple bacterial RCs results from studies on RCs from nonsulfur purple bacteria. These are currently the best-characterized membrane protein complexes.

Subunit Composition and Molecular Characterization

Most purple bacterial RCs contain four protein subunits (Figure 1), referred to as H, M, L, and C (a tetraheme

cytochrome c). Some, however, such as the RCs of *Rhodobacter* (*Rb.*) *sphaeroides*, *Rb. capsulatus*, and *Rhodospirillum* (*Rs.*) *rubrum*, contain only the H, M, and L subunits. The related RC of the green aerobic thermophilic bacterium *Chloroflexus* (*Cf.*) *aurantiacus* lacks the H subunit.

Generally, RCs from purple bacteria have been isolated and characterized from *Rhodopseudomonas* (*Rp.*) *viridis* (suggested by Hiraishi in 1997 to be referred to as *Blastochloris viridis*), *Rb. sphaeroides*, *Rb. capsulatus*, and a number of other purple bacteria. Variant RCs have been created by site-directed mutagenesis as well as classical herbicide-resistance selection procedures and have been isolated and characterized from *Rb. capsulatus*, *Rb. sphaeroides*, and *Rp. viridis*.

RC preparations have a nonheme iron and four magnesium-containing bacteriochlorophyll cofactors per RC, as measured by atomic absorption spectroscopy. In *Rb. sphaeroides* and *Rp. viridis*, these are bacteriochlorophyll *a* and bacteriochlorophyll *b*, respectively (Figures 1(a) and 2(a)). Those preparations with a tightly bound C subunit have four iron-containing heme groups which are covalently bound to the protein via thioether bonds formed with the Cys residues of the heme attachment site sequences Cys–X–Y–Cys–His and the His is the fifth ligand to the heme iron. In the *Rp. viridis* RC, the sixth ligands to heme1, heme2, and heme3 are Met residues (His–Met-ligated hemes), whereas the sixth ligand to heme4 is a second His (*bis*-His ligation). Apart from these four hemes, all other cofactors are noncovalently bound by the L and M subunits. The six-coordinated nonheme ferrous iron is coordinated by the Nε atoms of four His residues and the two Oε atoms of a Glu residue. The fifth ligands to the five-coordinated bacteriochlorophyll–magnesium ions are provided by the Nε atoms of His side chains. In addition to these metal-containing cofactors, there are two bacteriopheophytin groups, a carotenoid, and two quinones bound by the RC. In *Rb. sphaeroides*, these are bacteriopheophytin *a*, spheroidene, and ubiquinone-10, respectively, whereas *Rp. viridis* contains bacteriopheophytin *b*, 1,2-dihydroneurosporene, menaquinone-9, and ubiquinone-9 (Figures 1(a) and 2(a)–2(c)).

Apart from the availability of high-resolution crystal structures discussed below, one major reason why, despite its

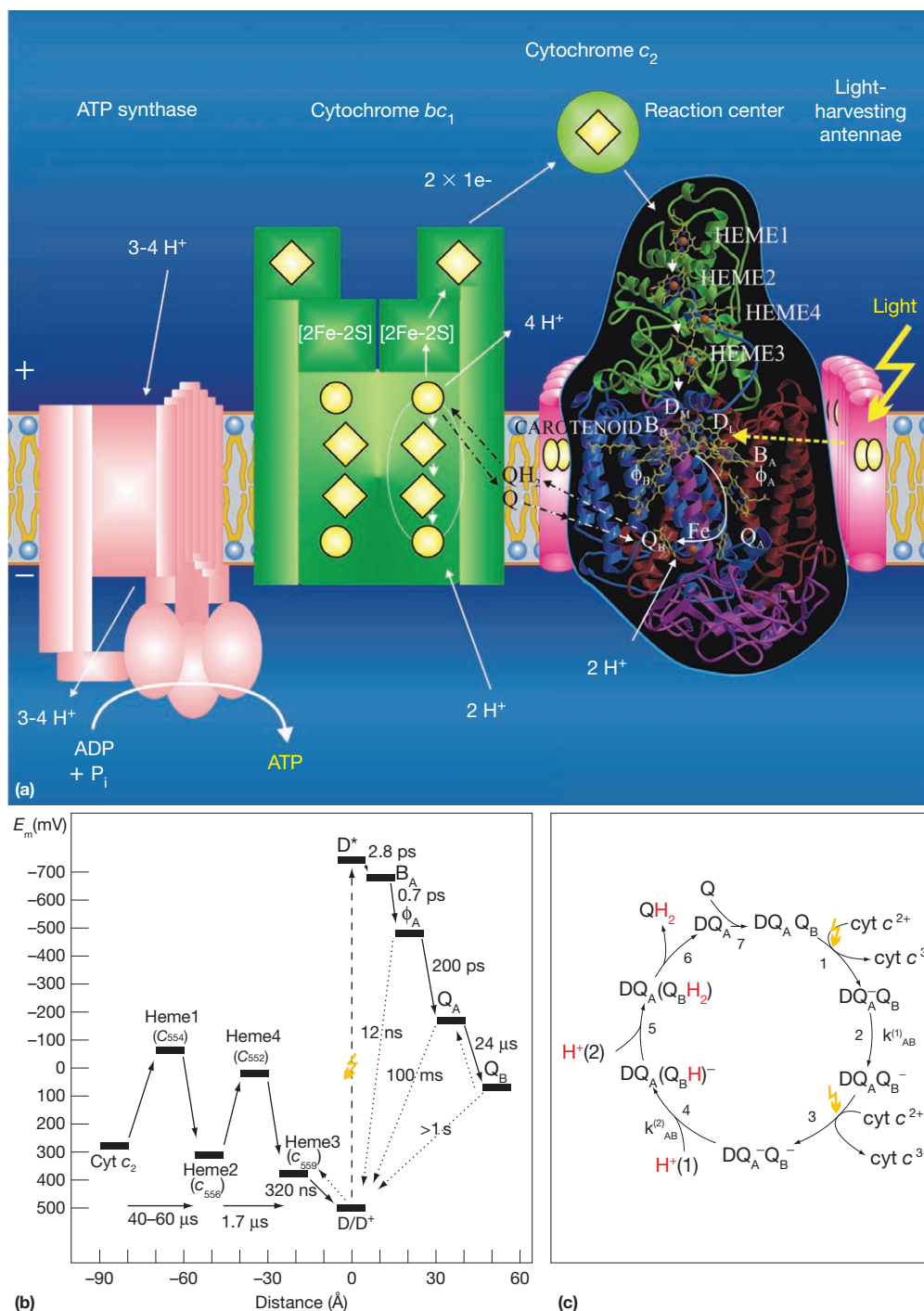


Figure 1 Structure and function of the photosynthetic RC. (a) Light-induced cyclic electron flow and the generation and utilization of a transmembrane electrochemical potential in the purple bacterium *Rp. viridis*. The structure of the *Rp. viridis* RC is represented schematically showing the heterotetramer of C, L, M, and H subunits as α traces in green, brown, blue, and purple, respectively, plus the 14 cofactors, which have been projected on to the molecule for better visibility. Also for the sake of clarity, the quinone tails are truncated after the first isoprenoid unit and the phytyl side chains of the bacteriochlorophyll and bacteriopheophytin molecules have been omitted, as have those atoms of the carotenoid molecule which were not observed in the electron density and assigned zero occupancy in the PDB entry 2PRC. Carbon, nitrogen, and oxygen atoms are drawn in yellow, blue, and red, respectively. (b) Equilibrium oxidation–reduction potentials of the *Rp. viridis* RC cofactors as compiled by Lancaster and Michel (see section Further Reading) as a function of intercofactor distance. The soluble electron donor protein cytochrome c_2 has been included as suggested in Figure 4(a). Reaction half-times indicated are taken from the references cited by Lancaster and Michel. The photochemical excitation is indicated by a dashed arrow and unphysiological charge recombination reactions are shown as dotted arrows. (c) Quinone reduction cycle. Reduced quinones are typeset in bold. Steps 2, 4, 5, and 6 are reversible (see text for details). (b) Reproduced from Lancaster CRD and Michel H (2001) Photosynthetic reaction centers. In: Messerschmidt A, Huber R, Poulos T, and Wiegardt K (eds.) *Handbook of Metalloproteins*, vol. 1, pp. 119–135. Chichester: Wiley, with permission from John Wiley & Sons.

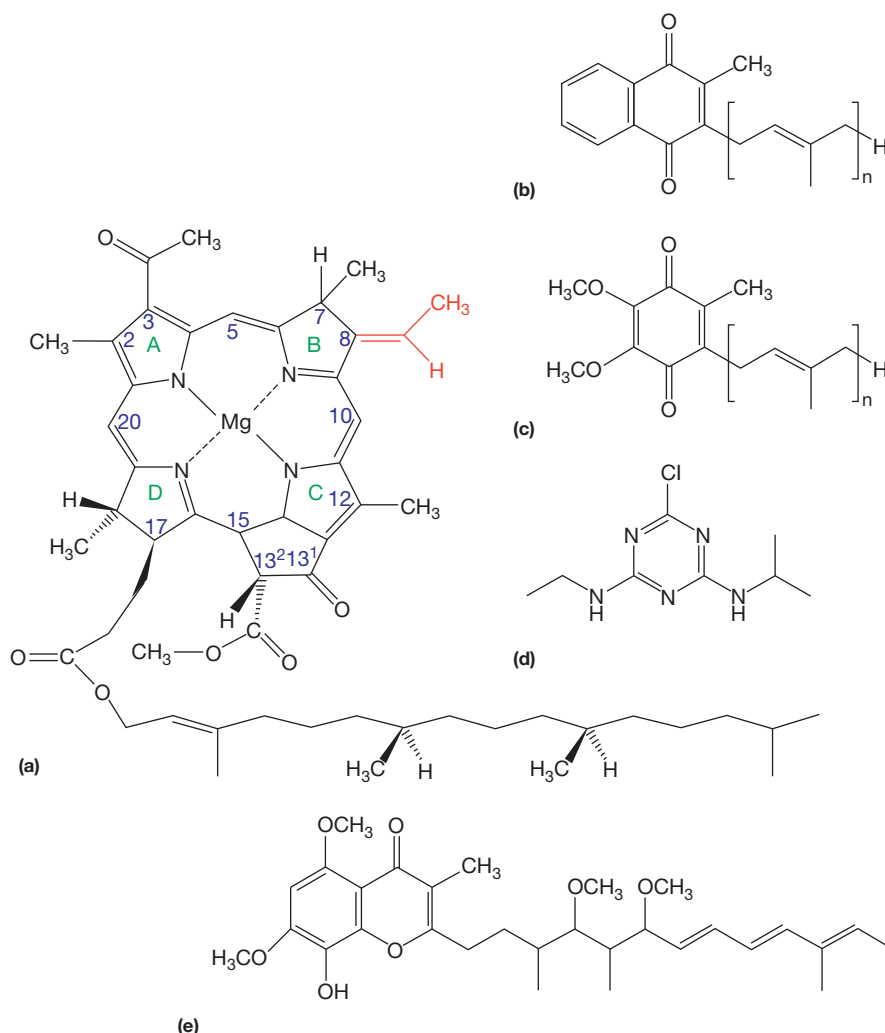


Figure 2 Chemical structures of RC cofactors (a–c) and inhibitors at the Q_B site (d,e). (a) Bacteriochlorophyll *b*, as bound in the *Rp. viridis* RC. The bacteriochlorophyll *a* bound in the *Rb. sphaeroides* RC differs by the presence of a C8 ethyl group instead of the C8 ethylidene group indicated in red. Bacteriopheophytins (*a* or *b*) are the metal-free variants of the bacteriochlorophylls (*a* or *b*) with two protons bonded to the nitrogens of the unsaturated pyrrole rings A and C. (b) Menaquinone-*n*. The native Q_A in the *Rp. viridis* RC is menaquinone-9. (c) Ubiquinone-*n*. The native Q_A in the *Rp. viridis* RC is ubiquinone-9. In the *Rb. sphaeroides* RC, both Q_A and Q_B are ubiquinone-10. (d) Atrazine (2-chloro-4-ethylamino-6-isopropylamino-*s*-triazine). (e) Stigmatellin A.

complexity, the purple bacterial RC has become the ‘hydrogen atom of protein electron transfer’ (W W Parson), is the richness of its characterization by optical absorption, electron paramagnetic resonance, electron-nuclear double resonance, Fourier-transform infrared, resonance Raman, fluorescence, Stark effect, and other types of spectroscopy (see section [Further Reading](#) for detailed overviews).

X-Ray Crystal Structures of RCs from Purple Bacteria

The photosynthetic RC from *Rp. viridis* was the first membrane protein complex for which an atomic structural model could be determined by X-ray crystallography (Protein Data Bank (PDB) entry 1PRC). The structure of the four-subunit *Rp. viridis* RC is shown schematically in [Figure 1\(a\)](#). The structure of the

RC from *Rb. sphaeroides* (not shown) would appear almost identical except for the C subunit at the top, which would be missing. The more recently determined structure of the RC from the sulfur purple bacterium *Thermochromatium tepidum* (PDB entry 1EYS) is very similar in overall architecture to that of *Rp. viridis*.

The *Rp. viridis* RC has an overall length of 130 Å in the direction perpendicular to the membrane. Parallel to the membrane, the maximum width is about 70 Å. The central core of the RC is formed by the L subunit and the M subunit, which possess five membrane-spanning segments each. Both subunits are closely associated and noncovalently bind 10 cofactors as detailed above and shown in [Figure 1\(a\)](#). Large parts of the L and M subunits and their associated cofactors are related by a twofold axis of symmetry perpendicular to the plane of the membrane. The H subunit is anchored to the membrane by

a single membrane-spanning helix and is attached to the LM core on the cytoplasmic side. On the periplasmic side, the C subunit with its four covalently bound heme groups is attached. The N-terminal diacylglycerol moiety which, in addition to the interaction with the LM core, anchors the C subunit in the membrane is not visible in the electron-density map.

The pigments form two symmetry-related branches, also shown in **Figure 1(a)**, each consisting of two bacteriochlorophylls, one bacteriopheophytin, and one quinone, which both cross the membrane starting from the special pair D of two closely associated bacteriochlorophylls (D_M and D_L) near the periplasmic side, followed by the accessory bacteriochlorophyll, B, one bacteriopheophytin, ϕ , and a quinone, Q. As shown in **Figure 1(a)**, only the branch more closely associated with L subunit is used in the light-driven electron transfer. It is called the A-branch, the inactive one is the B-branch. The active branch ends with the primary quinone Q_A , the inactive one with the secondary quinone Q_B . Halfway between both quinones, the nonheme iron is located. The carotenoid is in van der Waals contact with B_B and disrupts the twofold symmetry. In both species, the crystallographic temperature factors, which are a measure for the rigidity of the structure, are considerably higher along the B-branch than along the A-branch.

Functional Aspects

Oxidation–Reduction Potentials

The oxidation–reduction midpoint potentials, E_M , of the four heme groups follow the order low, high, high, low in the sequence or low, high, low, high if the hemes are ordered with decreasing distance (see **Figure 1(b)**) from the primary electron donor D. Theoretical analysis by the application of continuum electrostatics to the crystal structure of the C subunit has provided quantitative estimations of the factors contributing to the equilibrium E_M values of the four hemes. Specific residues and the propionic side chains on the hemes are calculated to strongly modulate the E_M values. The correct division into low- and high-potential hemes can be obtained by taking only the protein into account. Consideration of heme–heme interactions is required to reproduce the experimental data quantitatively.

The redox potentials of the other *Rp. viridis* RC cofactors are also included in **Figure 1(b)**. The potential of the excited state D^* is derived from the free energy difference between the lowest vibrational levels of D and D^* . Estimated from the wavelength of the absorption maximum, the dimer of bacteriochlorophyll *b* in *Rp. viridis* provides an energy of 1240 meV between D and D^* , that of the bacteriochlorophyll *a* dimer in *Rb. sphaeroides* one of 1380 meV. The redox potential of *Rb. sphaeroides* D has been increased from 505 to 765 mV by the introduction of three additional hydrogen bonds to the special pair bacteriochlorophylls by site-directed mutagenesis, thus destabilizing the oxidized state of the donor, D^+ .

The redox potential of Q_A is higher in the *Rb. sphaeroides* RC than in the *Rp. viridis* RC because of the different chemical nature of the quinones (ubiquinone vs. menaquinone). However, it is still 67 mV lower than that of *Rb. sphaeroides* Q_B , even though both cofactors are chemically identical, thus requiring a role of the protein in tuning the *in situ* redox potentials of the quinones.

Kinetics

Figures 1(a) and **1(b)** show schemes of the electron-transfer steps that occur in the purple bacterial RC. Light absorption leads to an excited primary donor D^* , from which an electron is transferred via the monomeric bacteriochlorophyll B_A (reaction half-time: 2.8 ps) and the bacteriopheophytin ϕ_A (700 fs) to Q_A in 200 ps, leading to the formation of $D^+Q_A^-Q_B$. Rereduction of D^+ by cytochrome c_{559} (heme 3) occurs in 320 ns. These processes are much faster than the subsequent proton uptake and interquinone electron-transfer reactions. Therefore, the first step of quinone reduction in the RC can be viewed as a photochemical cytochrome oxidation, giving rise to the radical state $DQ_A^-Q_B$ (**Figure 1(c)**). Rereduction of cytochrome c_{559}^+ by cytochrome c_{556} (heme 2) occurs in 1.7 μ s. The second step of quinone reduction involves the transfer of the first electron to Q_B in 17–25 μ s, resulting in the state $DQ_AQ_B^-$. The one-electron reduction of Q_B is not associated with direct protonation and the semiquinone species is anionic. However, the $Q_A^-Q_B \rightleftharpoons Q_AQ_B^-$ equilibrium constant is pH dependent, as electron transfer is accompanied by substoichiometric proton uptake due to the protonation of amino acid residues. Cytochrome c_{556}^+ is rereduced by cytochrome c_2 in 40–60 μ s. In species that lack the C subunit, for example, *Rb. sphaeroides*, the photooxidized special pair is directly rereduced by cytochrome c_2 in a biphasic reaction. The fast phase of ~ 1 μ s is attributed to intermolecular electron transfer, the slow phase of 100 μ s is limited by docking and reorientation of the cytochrome c_2 -RC complex. After a second photochemical cytochrome oxidation, the diradical state $DQ_A^-Q_B^-$ is formed at rates similar to those for the first electron transfer. Coupled transfer of the first proton and the second electron to Q_B^- leads to the monoprotonated, doubly reduced state $DQ_A(Q_BH)^-$. Transfer of the second proton for the formation of Q_BH_2 is kinetically indistinguishable from the first proton transfer in the wild-type RC and can only be resolved in the case of mutants with significantly retarded second proton-transfer rates.

Functional Derivatives

Substrate Analog and Inhibitor Complexes

In the original *Rp. viridis* RC structure (PDB entry 1PRC), the Q_B site was poorly defined because it was only partially occupied with the native ubiquinone-9 in the standard RC crystals. However, ubiquinone-2-reconstitution experiments have yielded crystals with full quinone occupancy of the Q_B site. Subsequent X-ray diffraction analysis and refinement have led to a well-defined Q_B -site model (2PRC), with the quinone bound in the proximal position, that is, close to the nonheme iron (hydrogen-bonded to its ligand His L190, see **Figure 3(a)**). In the RC structure with a Q_B -depleted Q_B site (3PRC, **Figure 3(b)**), refined at 2.4 Å, apparently five, possibly six, water molecules are bound instead of the ubiquinone head group, and a detergent molecule binds in the region of the isoprenoid tail. Using the structures 2PRC and 3PRC as references, the original data set 1PRC was reexamined. A more quantitative analysis of the original data resulted in 20% of the Q_B sites being occupied with quinone in the proximal site,



Figure 3 Derivatives at the Q_B site of the *Rp. viridis* RC. (a) Comparison of distal (1PRC_{new}, cyan) and proximal (2PRC, yellow) ubiquinone-binding sites. (b) Comparison of Q_B-depleted (3PRC, pink) and ubiquinone-2-occupied (2PRC, yellow) Q_B sites. (c) Comparison of stigmatellin binding (4PRC, gray) and ubiquinone-2 binding (2PRC, yellow). (d) Atrazine binding (5PRC, pink) compared to distal (1PRC_{new}, cyan) and proximal (2PRC, yellow) ubiquinone-binding sites. (e) Mechanistic implications of the structures 2PRC, 3PRC, 4PRC, and the revised model 1PRC_{new} for the events at the Q_B site within the reduction cycle of quinone to quinol. The numbering of the steps is analogous to that in [Figure 1\(c\)](#). Hydrogen atoms are drawn as small light gray spheres. Dashed green arrows symbolize quinone movements.

30% having quinone bound in a more distal site, not hydrogen-bonded to His L190 and further away from the nonheme iron (see [Figure 3\(a\)](#)), and half of the Q_B sites being empty or having the quinone unaccounted for. A further structure, the RC complex with the inhibitor stigmatellin (see [Figure 2\(e\)](#)), refined at 2.4 Å, indicates that additional hydrogen bonds stabilize the binding of stigmatellin over that of ubiquinone-2 (4PRC, see [Figure 3\(c\)](#)). The binding pattern

observed for the stigmatellin complex can be viewed as a model for the stabilization of a monoprotonated reduced intermediate (Q_BH or Q_BH^+). This indicates that the Q_B site is not optimized for Q_B binding, but for Q_B reduction to the quinol. In combination with the results of electrostatic calculations, these crystal structures can provide models for intermediates in the reaction cycle of ubiquinone reduction to ubiquinol (**Figure 3(a)**), as discussed below.

The Q_B site is also a well-established site of herbicide action. Over half of the commercially available herbicides function by inhibition of higher plants at the Q_B site of the D1 polypeptide of the PSII RC. A commercially very important class of herbicides are the triazines, which were introduced by J R Geigy S A in the 1950s. A prominent example is atrazine ([Figure 2\(d\)](#)), first reported in 1957. According to statistics from 1995, atrazine was used on approximately 67% of all US corn acreage, 65% of sorghum acreage, and 90% of sugarcane acreage. Another well-known triazine is terbutryn (2-*t*-butylamino-4-ethylamino-6-methylthio-*s*-triazine). X-Ray crystal structures of complexes of the RC with atrazine (PDB entry 5PRC) and terbutryn (PDB entry 1DXR) have been determined at 2.35 and 2.00 Å resolution, respectively. In both cases, three hydrogen bonds bind the distal side of the inhibitors to the protein, and four additional hydrogen bonds, mediated by two tightly bound water molecules are apparent on the proximal side, as shown for atrazine in [Figure 3\(d\)](#). In contrast to the proximal binding of stigmatellin ([Figure 3\(c\)](#)), the triazine inhibitor partially overlaps with both the distal and the proximal ubiquinone-binding sites ([Figure 3\(d\)](#)).

Mutants

Work on site-directed mutagenesis of photosynthetic RCs started with the RC from *Rb. capsulatus*. This species is genetically very well characterized and able to grow non-photosynthetically under aerobic conditions, as well as under anaerobic conditions using, for example, dimethylsulfoxide as an electron acceptor. Most importantly, under these latter conditions, the photosynthetic apparatus is fully induced. Unfortunately, the RC from *Rb. capsulatus* could not be crystallized, thus thwarting proper inspection for structural changes.

The closely related *Rb. sphaeroides* can be grown under similar nonphotosynthetic conditions, so that site-directed mutagenesis is also straightforward. As mentioned above, this RC is amenable to inspection by X-ray crystallography for structural changes. Many amino acids that were considered to be of importance for pigment binding, electron transfer, or proton transfer were changed in *Rb. sphaeroides* RCs (see section [Further Reading](#) for detailed overviews). In addition to mutagenesis, quinone or bacteriochlorin cofactors may be removed or replaced chemically with a wide range of similar compounds.

Site-directed mutagenesis of the RC from *Rp. viridis* is possible, but more difficult. *Rp. viridis* can grow only under photosynthetic and, very slowly, under microaerophilic conditions. However, under microaerophilic conditions, the photosynthetic apparatus is hardly induced and photosynthetic growth conditions exert a selection pressure for revertants and suppressor mutants if the RCs are functionally impaired. On the other hand, very interesting herbicide-resistant mutants were obtained by classical selection procedures, with mutations some of which would not have been made by site-directed mutagenesis.

Some of these herbicide-resistant mutants of the *Rp. viridis* RC have also been analyzed by X-ray crystallography. For instance, the mutation Tyr L222 → Phe unexpectedly leads to resistance against the herbicide terbutryn. In the wild type, Tyr

L222 forms a hydrogen bond with the peptide carbonyl oxygen of Asp M43. As this hydrogen bond is now missing, a stretch of the M subunit (M25–50) moves into a new position. The side chain of Phe L222 rotates by 90° into the herbicide-binding site, thereby preventing the binding of terbutryn by steric hindrance, but at the same time introducing sensitivity to two other classes of PSII inhibitors normally inactive in bacterial RCs, the ureas and phenolics.

Using site-directed mutagenesis, the highly conserved Tyr L162, positioned halfway between D and the proximal heme3 (cytochrome *c*₅₅₉) in the *Rp. viridis* RC (see [Figure 4\(a\)](#), below), was exchanged against a number of amino acids. All mutants grew photosynthetically. The redox potentials of D and *c*₅₅₉ were changed by the mutations. The structures of two mutants (Tyr L162 → Phe and Tyr L162 → Thr) were determined and found not to differ significantly from the wild-type structure. Analysis of the kinetics of electron transfer led to the conclusion that the tyrosine residue at position L162 is not required for fast electron transfer from *c*₅₅₉ to D⁺.

A more recent mutation of Arg C264 → Lys decreases the midpoint potential of heme3 (cytochrome *c*₅₅₉) from +380 to +270 mV, that is, below that of heme2 (+320 mV, see [Figure 1\(b\)](#)). In the structure of the mutant RC at 2.46-Å resolution, no remarkable differences were found apart from the mutated residue itself. The half-time of electron transfer between heme2 and heme3 was the same as in the wild type, indicating that the observed reaction rate is limited by the very uphill electron transfer from heme2 to heme4 (see [Figure 1\(b\)](#)).

Inter- and Intramolecular Electron and Proton Transfer and Catalytic Mechanism

Intramolecular Electron Transfer

The kinetics of light-induced electron transfer via Q_A to Q_B and of the rereduction of the special pair have been detailed above. The reasons for the unidirectionality of electron transfer along the active A branch and not along the inactive B branch, despite the twofold pseudosymmetry of the LM core of the RC, have been the subject of numerous theoretical and experimental investigations. Slight differences in geometry, differences in rigidity, and differences in the amino acid composition of the L and M subunits have been suggested to contribute to unidirectionality. The latter aspect was specified by the theoretical identification of a large electrostatic field favoring charge separation along the A branch. A major contributor to the potential gradient is Arg L103, whose positive charge is stabilized by different sections of the polypeptide backbone. This dipolar stabilization leads to a much longer-range effect of the positive charge than if it were stabilized by a counterion.

An experimental observation consistent with this electrostatic analysis is the finding that a site-directed double mutant in *Rb. capsulatus* RCs (Leu L212 → His; Gly L201 → Asp) appears to show significant electron transfer to ϕ_B . The first mutation leads to the incorporation of the β bacteriochlorophyll in the ϕ_A position and the second mutation introduces a negative charge close to B_A, thus making electron transfer down the A branch less favorable.

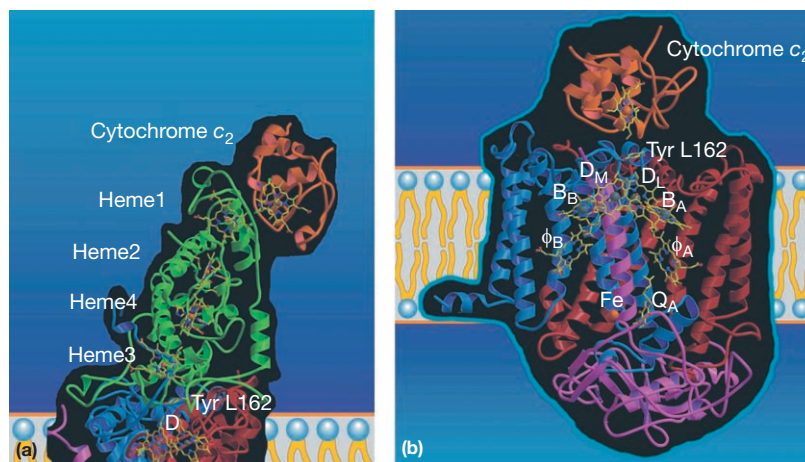


Figure 4 Cytochrome c_2 oxidation by the photosynthetic RCs of *Rp. viridis* and *Rb. sphaeroides*. (a) Reduction of the photooxidized tetraheme C subunit of the *Rp. viridis* RC (color-coding as in Figure 1(a)) by *Rp. viridis* cytochrome c_2 (orange). Theoretical docking as suggested from mutagenesis experiments. (b) Reduction of the special pair D in the *Rb. sphaeroides* RC (color coding of the L, M, and H subunits analogous to Figure 1(a)) by *Rb. sphaeroides* cytochrome c_2 (orange) as determined by X-ray crystal structure analysis (PDB entry 1L9B).

Quinone Reduction and Protonation

Both Q_A and Q_B sites are deeply buried within the photosynthetic RC complex, approximately 15 Å from the cytoplasmic surface. Proton transfer to the reduced quinone within the Q_B site could occur by protons moving along a chain of proton donors and acceptors by a proton wire, or hydrogen-bonded chain mechanism. Possible proton donors and acceptors are protonatable amino acid residues and water molecules. A number of the protonatable residues between the Q_B site and the cytoplasmic surface have been shown to be functionally relevant to the proton-transfer process by analysis of site-directed mutations and second site revertants. The observed effects can be due to the modification of the kinetics and thermodynamics of electron or proton transfer. Electrostatic calculations based on the X-ray structure coordinates of the RCs from *Rb. sphaeroides* and *Rp. viridis* led to the identification of residues that can contribute to the changes in equilibrium distributions of protons in the different redox states of the protein, thus helping to determine the role of the functionally important residues.

In combination with the results of electrostatic calculations, the crystal structures 3PRC, 1PRC_{new}, 2PRC, and 4PRC discussed above (cf. Figures 3(a)–3(d)) can provide models for intermediates in the reaction cycle of ubiquinone reduction to ubiquinol (see Figure 3(e)). The binding of the incoming Q_B to the distal site displaces some of the water molecules present in the empty pocket. The quinone ring is flipped around the isoprenoid tail and further water molecules are displaced for the Q_B to occupy the proximal position (step 1a in Figure 3(e)). This is the position in which neutral Q_B accepts an electron from Q_A^- . The hydrogen bonds donated to the quinone will automatically lead to a tighter binding of the negatively charged semiquinone Q_B^- compared to the neutral Q_B (step 2). Additionally, the side chain of Ser L223 can reorient by rotation of its χ_2 ($C\alpha-C\beta-O\gamma-H\gamma$) torsional angle, thus establishing an additional hydrogen bond to Q_B^- . Coupled to the transfer of the second electron, the first proton

is transferred, possibly via a transiently protonated Ser L223 – OH_2^+ , thus forming the monoprotinated, doubly reduced intermediate Q_BH^- (steps 3 and 4). After transfer of the second proton (step 5), movement of the quinol from the proximal to the distal position (step 6a) may be facilitated by increased stacking interactions of the aromatic ring systems with the Phe L216 ring and the diffusion of water molecules back into the pocket. The structures of these intermediates provide explanations for their relative binding affinities, as required for proper enzymatic function of the Q_B site. A rearrangement of hydrogen bonds, most prominently the reorientation of the Ser L223 side chain for Q_B reduction, as suggested by the scenario in Figure 3(e), has also been calculated by Alexov and Gunner to be necessary to make Q_B reduction more favorable than Q_A reduction. These local rearrangements may constitute the conformational changes deduced to be required for function by a variety of experiments.

Cytochrome c_2 Oxidation

All four hemes of the *Rp. viridis* RC tetraheme C subunit are located close enough to the surface of the protein to accept electrons from soluble cytochrome c_2 . Site-directed mutagenesis in another nonsulfur purple bacterium, *Rubrivivax gelatinosus*, has led to the identification of a patch of acidic residues immediately surrounding the distal low-potential heme 1 of the tetraheme C subunit that apparently forms an electrostatically favorable binding site for soluble cytochromes. Thus, all four hemes in the C subunit appear to be directly involved in the electron transfer toward the photooxidized special pair. Based on these findings, a model was proposed for the transient cytochrome c_2 -RC complex for *Rp. viridis* (see Figure 4(a)).

In the case of the *Rb. sphaeroides* RC, which lacks the C subunit (cf. section Subunit Composition and Molecular Characterization), the photooxidized special pair D^+ is directly rereduced by cytochrome c_2 . The structure of the cocomplex of

the *Rb. sphaeroides* RC and cytochrome c_2 has been determined by X-ray crystallography at 2.4-Å resolution (cf. **Figure 4(b)**). In these crystals, D^+ is reduced by cytochrome c_2 at the same rate as measured in solution indicating that the structure of the complex in the region of electron transfer is the same in the crystal and in solution. The binding interface can be divided into two domains. The first domain contributes to the strength and specificity of cytochrome c_2 binding and is a short-range interaction domain that includes Tyr L162 (cf. **Figure 4(b)**), and groups exhibiting nonpolar interactions, hydrogen bonding, and a cation- π interaction. The second is a long-range, electrostatic interaction domain that contains complementary charges on the RC and cytochrome c_2 . In addition to contributing to the binding, this domain may help steer the unbound proteins into the right orientation.

Relevance to PSII

Based on the determined structure of the purple bacterial RC, very specific sequence homologies, and azidoatrazine labeling, the RC core of higher plant PSII was proposed to be similar to the LM core of the bacterial RC, with the D1 and D2 proteins corresponding to the L and M subunits, respectively. This proposal could be verified experimentally. Recently, suitably designed, modified bacterial RCs have been shown to mimic tyrosine oxidation in PS II.

See also: **Bioenergetics:** Chlorophylls and Carotenoids; Green Bacteria: Chlorophyll Biosynthesis, Light-Harvesting, Reaction Centers, and Electron Transport; Heme Synthesis; Structure-Function of the Cytochrome b_6f Complex of Oxygenic Photosynthesis; **Metabolism Vitamins and Hormones:** Photosynthesis; Porphyrin Metabolism; Vitamin K: Biochemistry, Metabolism, and Nutritional Aspects; **Protein/Enzyme Structure Function and Degradation:** Enzyme Inhibitors; Heme Proteins; Substrate Binding, Catalysis, and Product Release.

Further Reading

- Blankenship RE, Madigan MT, and Bauer CE (eds.) (1995) *Anoxygenic Photosynthetic Bacteria*. Dordrecht: Kluwer Academic.
- Deisenhofer J and Michel H (1989) The photosynthetic reaction center from the purple bacterium *Rhodospseudomonas viridis* (Nobel Lecture). *EMBO Journal* 8: 2149–2170.
- Hoff AJ and Deisenhofer J (1997) Photophysics of photosynthesis – structure and spectroscopy of reaction centers of purple bacteria. *Physics Reports – Review Section of Physics Letters* 287: 1–247.
- Lancaster CRD (1999) Quinone-binding sites in membrane proteins: What can we learn from the *Rhodospseudomonas viridis* reaction centre? *Biochemical Society Transactions* 27: 591–596.
- Lancaster CRD (2008) Structures of reaction centers in anoxygenic bacteria. In: Renger G (ed.) *Primary Processes of Photosynthesis. Principles and Apparatus. Part 2: Reaction Centers/Photosystems, Electron Transport Chains, Photophosphorylation and Evolution*, pp. 5–56. Cambridge: Royal Society of Chemistry.
- Lancaster CRD and Michel H (2001) Photosynthetic reaction centers. In: Messerschmidt A, Huber R, Poulos T, and Wieghardt K (eds.) *Handbook of Metalloproteins*, vol. 1, pp. 119–135. Chichester: Wiley.
- Lancaster CRD, Hunte C, Kelley J III, Trumpower BL, and Ditchfield R (2007) A comparison of stigmatellin conformations, free and bound to the photosynthetic reaction center and the cytochrome bc_1 complex. *Journal of Molecular Biology* 368: 197–208.
- Michel-Beyerle ME (ed.) (1996) *The Reaction Center of Photosynthetic Bacteria. Structure and Dynamics*. Berlin: Springer.
- Moser CC, Page CC, Cogdell RJ, et al. (2003) Length, time, and energy scales of photosystems. *Advances in Protein Chemistry* 63: 71–109.
- Nabedryk E and Breton J (2008) Coupling of electron transfer to proton uptake at the Q_B site of the bacterial reaction center: A perspective from FTIR difference spectroscopy. *Biochimica Biophysica Acta* 1777: 1229–1248.
- Okamura MY, Paddock ML, Graige MS, and Feher G (2000) Proton and electron transfer in bacterial reaction centers. *Biochimica Biophysica Acta* 1458: 148–163.
- Parson WW (2003) Electron donors and acceptors in the initial steps of photosynthesis in purple bacteria: a personal account. *Photosynthesis Research* 76: 81–92.
- Parson WW (2008) Functional patterns of reaction centers in anoxygenic photosynthetic bacteria. In: Renger G (ed.) *Primary Processes of Photosynthesis. Principles and Apparatus. Part 2: Reaction Centers/Photosystems, Electron Transport Chains, Photophosphorylation and Evolution*, pp. 57–109. Cambridge: Royal Society of Chemistry.
- Wraight CA (2004) Proton and electron transfer in the acceptor quinone complex of photosynthetic reaction centers from *Rhodobacter sphaeroides*. *Frontiers in Bioscience* 9: 309–337. <http://www.bioscience.org> (accessed March 2011).

Pyridoxal Phosphate

M D Toney, University of California, Davis, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by David E. Metzler, volume 3, pp. 595–599, © 2004, Elsevier Inc.

Glossary

Aldol-type reaction Biochemical reactions related in their chemical mechanisms to the classical aldol condensation and cleavage reactions of organic chemistry.

Coenzymes Small organic molecules that function together with enzymes to catalyze specific types of chemical reactions.

Lysine unit A lysine residue, a molecule of the amino acid lysine incorporated into a protein molecule.

Substrate The substance which undergoes the reaction catalyzed by an enzyme.

Tetrahydrofolic acid A coenzyme formed by reduction of the vitamin folic acid.

Pyridoxal phosphate (PLP) is one of several small organic molecules known as coenzymes, compounds that are essential to all forms of life. Coenzymes provide chemical functionality to enzymes that the 20 proteinogenic amino acids cannot. PLP is employed by a large number of enzymes. The majority of these are involved in primary metabolism, in pathways for the synthesis or degradation of amino acids or other nitrogen-containing compounds. In free living microorganisms such as the common *Escherichia coli* laboratory bacterium, approximately 1.5% of the total genes encode PLP-dependent enzymes. The Enzyme Commission has cataloged more than 150 distinct catalytic activities that are PLP dependent. Thus, PLP-dependent enzymes are central to the nitrogen metabolism of all living systems.

Vitamin B₆ constitutes a family of chemical structures, as shown in Figure 1. PLP, the most common enzymatically active form of vitamin B₆, is the phosphoric acid ester of pyridoxal. It is synthesized by plants and by many microorganisms, but humans must obtain it (or the other forms) from dietary sources. The major form of vitamin B₆ in plants (and in vitamin pills) is the alcohol pyridoxol (pyridoxine). In tissues of our bodies (as well as of other organisms), this is converted to pyridoxol 5'-phosphate, which is oxidized to the aldehyde PLP (Figure 1). Some enzymes can convert PLP reversibly to pyridoxamine phosphate (PMP) by a process known as 'transamination' (Figure 2). All organisms contain PLP and PMP, most of which is bound to enzymes.

Transamination

Transamination is the most common reaction type catalyzed by PLP-dependent enzymes. It is a biologically important process by which living cells reversibly transfer the amino group from an amine (e.g., γ -aminobutyrate) or α -amino acid (e.g., aspartate) to an α -keto carboxylic acid (e.g., α -ketoglutarate). (Figure 2). In the human body, aminotransferases catalyze many different steps in pathways for the biosynthesis and breakdown of amino acids. For example, glutamic acid can donate nitrogen to α -keto acids to form amino acids such as alanine and aspartic acid.

The process occurs via two half reactions. In the first, the amino group is transferred from the amine or amino acid

to PLP to form PMP. In the second, the amino group from PMP is transferred to the α -keto acid (Figure 2). During the reactions, PLP and PMP are held tightly in the active sites of the enzymes. Transamination proceeds through covalent substrate-coenzyme intermediates known as 'aldimines' (Figure 3) and 'ketimines'. Formation of the covalent external aldimine between the substrate and PLP allows the electrophilic pyridine ring to stabilize, by resonance, carbanionic intermediates formed on the substrates.

The Variety of PLP-Dependent Reactions

Transamination is only one of many reactions of amino acids and amines that are catalyzed by PLP-dependent enzymes. Some of these are listed in Table 1. Groups *a*, *b*, and *c* in this table are defined by the cleavage of one of three possible bonds to C α of the external aldimine intermediate. The enzymes in group *a*, which includes transaminases, initiate reaction by removing a proton from C α of the substrate. In doing so, these enzymes form a carbanionic, 'quinonoid' intermediate (Figure 4). For several enzymes, this intermediate has been identified spectroscopically. The quinonoid intermediate reacts in different ways, depending on the reaction being catalyzed. Aminotransferases catalyze addition of a proton to the exocyclic carbon of PLP at the 4 position of the pyridine ring, leading to a ketimine intermediate. Overall, this is a 1,3-prototropic shift of a hydrogen. Subsequent hydrolysis of the ketimine intermediate yields the PMP form of the enzyme and the α -keto acid product. The transamination cycle is completed by a second α -keto acid reacting with the PMP enzyme through the reverse of the above sequence of steps.

Racemases, which remove a proton from C α of the external aldimine, add a proton back to the same carbon atom but from the opposite side of the planar quinonoid intermediate (Figure 4). The effect is to catalyze the interconversion of two stereoisomers known as the L- and D-forms of the amino acid. Bacteria form D-alanine from L-alanine using PLP-dependent alanine racemase. They utilize the D-alanine, as well as D-glutamic acid, in their cell walls. The D-amino acids provide protection from extracellular proteases that act on L-amino acids. Small amounts of D-serine are found in the human brain. It is formed

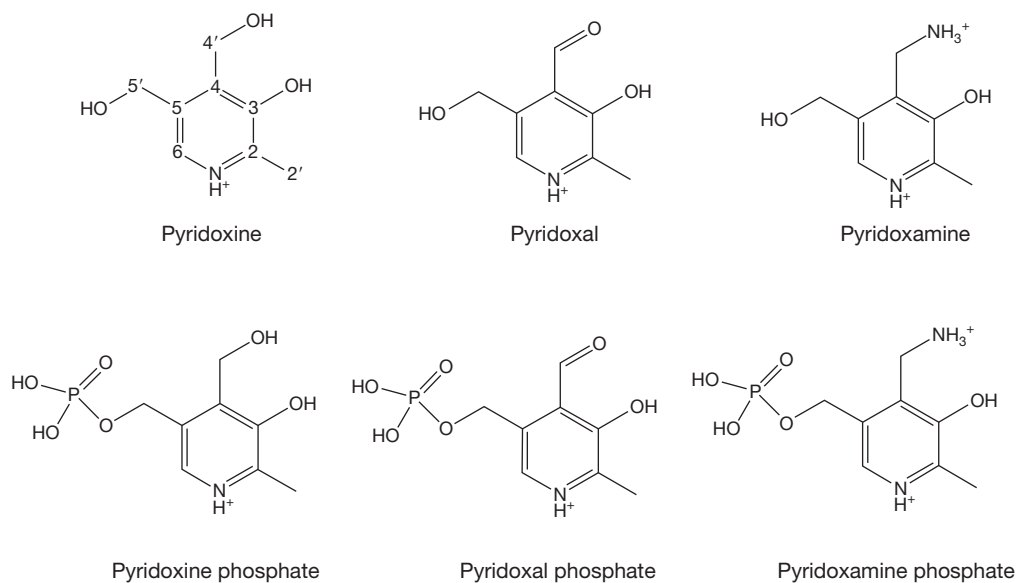


Figure 1 The family of vitamin B₆ structures. The numbering of the carbons is indicated on pyridoxine.

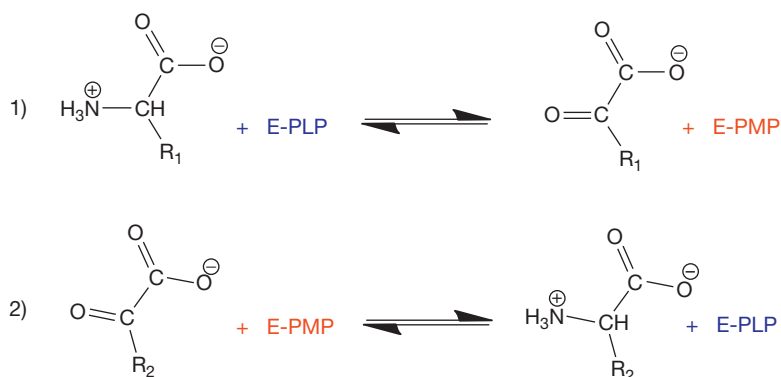


Figure 2 The transamination reaction by which amino groups are transferred from one carbon skeleton (in the form of an α-ketoacid) to another.

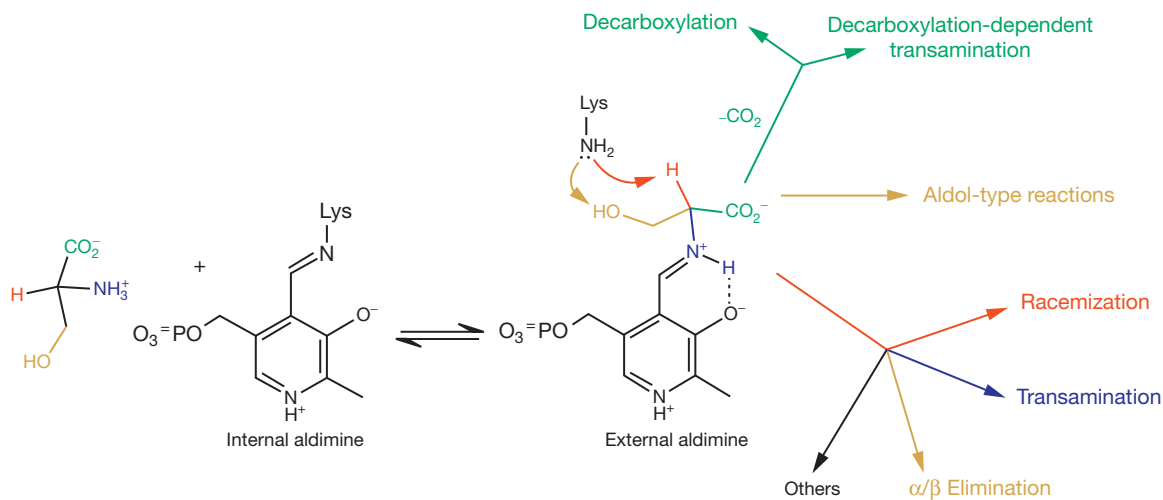


Figure 3 The variety of reactions catalyzed by PLP-dependent enzymes. The bonds, atoms, and reaction arrows are color coded to indicate those involved in each reaction type.

from L-serine by a PLP-dependent racemase that is structurally distinct from alanine racemase.

A third type of reaction of quinonoid intermediates (Figure 3) is nucleophilic addition to a carbon, which occurs

Table 1 A selection of enzymes that require pyridoxal phosphate as a coenzyme

- | |
|--|
| (a) Removal of α -hydrogen as H^+ |
| (1) Racemization |
| Alanine racemase |
| Serine racemase |
| (2) Cyclization |
| Aminocyclopropane carboxylate synthase |
| (3) Amino group transfer |
| Aspartate aminotransferase |
| Alanine aminotransferase |
| D-Amino acid aminotransferase |
| Branched chain aminotransferase |
| γ -aminobutyrate aminotransferase |
| (4) β -elimination or replacement |
| D- and L-serine dehydratases (deaminases) |
| Tryptophan indole-lyase (tryptophanase) |
| O-Acetylserine sulfhydrylase (cysteine synthase) |
| Tryptophan synthase |
| (b) Removal of α -carboxylate as CO_2 |
| Diaminopimelate decarboxylase |
| Glycine decarboxylase (requires lipoyl group) |
| Glutamate decarboxylase |
| Histidine decarboxylase |
| Dopa decarboxylase |
| Tyrosine decarboxylase |
| (c) Removal or replacement of side chain |
| Serine hydroxymethyltransferase |
| δ -aminolevulinic acid synthase |
| Serine palmitoyltransferase |
| (d) Gamma elimination and replacement |
| Cystathionine γ -synthase |
| Cystathionine γ -lyase |
| Threonine synthase |
| (e) Other reactions |
| Aspartate β -decarboxylase |
| Aminocyclopropanecarboxylate deaminase |

in the biosynthesis of the cyclic amino acid aminocyclopropane carboxylic acid, a precursor to the plant hormone ethylene, from S-adenosylmethionine.

A large group of enzymes catalyzes a fourth type of reaction, which involves elimination of a leaving group such as $-OH$ or $-SH$. This may be followed by nonenzymatic decomposition of the product with liberation of NH_3 and other products, or replacement of the leaving group by a new group. These reactions play many roles in both the biosynthesis and degradation of amino acids. A familiar example of such a decomposition reaction initiates the release of compounds giving the characteristic odor to crushed garlic.

A second major group of PLP-dependent enzymes, the decarboxylases, removes CO_2 from the substrate. These enzymes also function in both catabolic and biosynthetic pathways. In each case, the irreversible loss of CO_2 drives the pathway in the needed direction. The most common decarboxylases participate in the formation of neurotransmitters such as γ -aminobutyrate, dopamine, adrenaline, and serotonin.

A third major group of PLP-dependent enzymes catalyzes cleavage or formation of carbon-carbon bonds in amino acids. These reactions resemble aldol-type reactions of organic chemistry. An example is formation of serine from glycine and formaldehyde. The latter, being a toxic compound, is carried as a derivative of another coenzyme, tetrahydrofolic acid. The enzyme serine hydroxymethyltransferase (Table 1), which catalyzes this reaction, allows bacteria and plants to form serine from glycine and also allows mammals to form glycine from serine.

A fourth, large group of enzymes (d in Table 1) have more complex mechanisms. Examples include the decarboxylation of the β carbon of aspartate, cyclopropyl ring opening of aminocyclopropanecarboxylate, radical-dependent reactions, etc. One of these is a bacterial enzyme that converts the selenium-containing selenocysteine into L-alanine and elemental selenium (Se^0). Related enzymes in all organisms transfer sulfur, as S^0 , into important metal-sulfur clusters such as Fe_4S_4 or the $MoFe_7S_9$ cluster, which participates in conversion of N_2 to HN_3 by nitrogen-fixing bacteria.

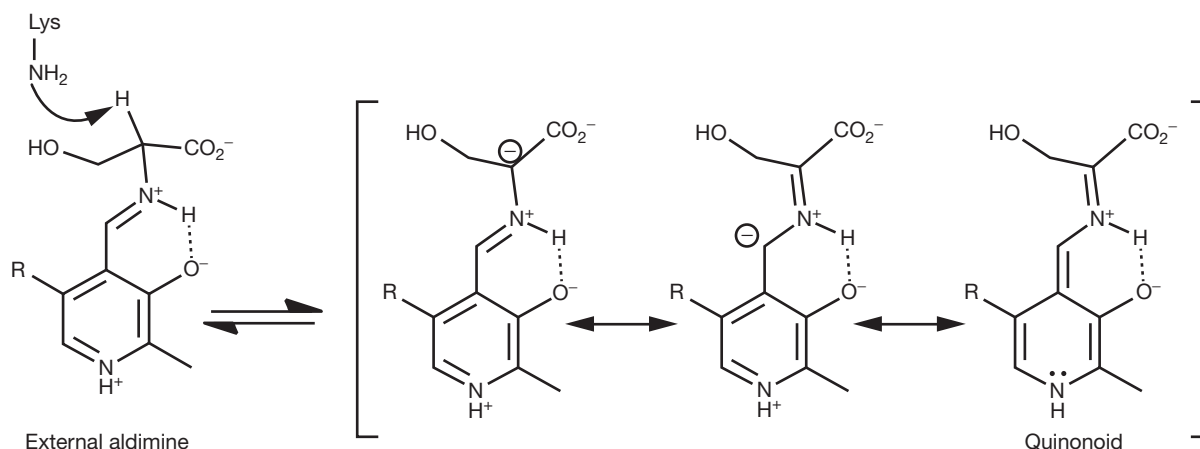


Figure 4 The formation of the carbanionic quinonoid intermediate is thought to be central to the reaction mechanisms of most PLP enzymes. It has a variety of fates, according to reaction type. In transamination, for example, it is protonated on C4' of PLP to give a ketimine intermediate that is hydrolyzed to the PMP enzyme form and a keto-containing product.

Chemical Mechanisms

The basic mechanistic steps are shown in [Figures 3](#) and [4](#); however, several steps are not shown for simplicity. Many individual chemical steps must occur in sequence for each reaction catalyzed, and PLP-dependent enzymes do a remarkable job of ensuring high fidelity given the complexity of the reactions they catalyze. In a few cases, such as with aspartate aminotransferase, the mechanistic details have been well characterized by use of spectroscopic techniques and by the determination of high-resolution structures by X-ray crystallography.

In the resting enzyme, PLP makes a Schiff base linkage to the ε -NH₂ group of lysine in the protein structure, which is called the 'internal aldimine'. Substrate binding often induces a conformational change in the protein, commonly leading to an enzyme-substrate complex in which the substrate is shielded from solvent. The first chemical step in all PLP-dependent reactions is the formation of a Schiff base with the substrate (external aldimine) via attack of the substrate amino group on the internal aldimine. It is from the external aldimine that the different reaction types diverge (i.e., the external aldimine is the only common intermediate for all PLP-dependent enzymes).

The lysine ε -NH₂ group that is freed from the internal aldimine can serve as a catalytic base that removes protons from the substrate in subsequent steps. In aminotransferases, for example, lysine 258 removes the proton at C α of the substrate and places it on C4' of PLP to form a ketimine. The structures, chemical groups involved, and individual steps vary from one enzyme to another and one reaction type to another. Most PLP-dependent reactions do not form PMP as an intermediate. However, virtually all PLP-dependent enzymes make use of an active site lysine as a general acid/base catalyst for multiple proton transfers.

The fundamental importance of the ring nitrogen at the bottom of the structure in [Figure 3](#) lies in its action as an electron-withdrawing group. In many PLP-dependent enzymes, a negatively charged carboxylate side chain forms a hydrogen bond to the N⁺-H at the bottom of the PLP ring. This carboxylate helps, through electrostatic interactions, to keep the ring protonated. On the other hand, the 3'-phenolate group opposes electron withdrawal from the substrate, perhaps assisting in other steps in the reaction such as interconversion

of internal and external aldimines. The 3'-phenolate group also helps to hold a proton on the Schiff base nitrogen as shown in [Figure 3](#).

In 1966, Dunathan proposed that the primary event in determining reaction specificity, cleaving the appropriate bond to C α of the external aldimine intermediate, is controlled by stereoelectronic effects. In other words, when a bond to C α is aligned parallel to the *p* orbitals of the conjugated aldimine/pyridine ring structure, that bond is more easily broken. This is so because the negative charge that develops during the bond-breaking process will be stabilized by resonance. Each enzyme must hold its substrate in such a way that the appropriate bond to the C α atom is parallel to the *p* orbitals of the conjugated π system. Thus, binding interactions between the enzyme and substrate are especially critical in maintaining reaction specificity in PLP-dependent enzymes.

See also: [Metabolism Vitamins and Hormones: Amino Acid Metabolism](#); [Protein/Enzyme Structure Function and Degradation: B₁₂-Containing Enzymes](#).

Further Reading

- Anthony C, John RA, and Wilmot C (eds.) (2003) Third international symposium on vitamin B6, PQQ, carbonyl catalysis, and quinoproteins. *Biochimica et Biophysica Acta* 1647: 1–408.
- Bender DA (1999) Non-nutritional uses of vitamin B6. *British Journal of Nutrition* 81: 7–20.
- Christen P and Metzler DE (1985) *Transaminases*. New York, NY: Wiley.
- Cooper AJL and Meister A (1989) An appreciation of Alexander E. Braunstein. The discovery and scope of enzymatic transamination. *Biochimie* 71: 387–404.
- Dolphin D, Poulson R, and Arramovic O (1986) *Vitamin B₆ Pyridoxal Phosphate Chemical, Biochemical, and Molecular Aspects*, vol. 2. New York, NY: Wiley.
- Dunathan HC (1966) Conformation and reaction specificity in pyridoxal phosphate enzymes. *Proceedings of the National Academy of Sciences of the United States of America* 55: 712–716.
- Jansonius JN (1998) Structure, evolution and action of vitamin B6-dependent enzymes. *Current Opinion in Structural Biology* 8: 759–769.
- Mehta PK and Christen P (1998) The molecular evolution of Pyridoxal-5'-phosphate-dependent enzymes. *Advances in Enzymology and Related Areas of Molecular Biology* 74: 129–184.
- Metzler DE (2001/2003) *Biochemistry. The Chemical Reactions of Living Cells*. San Diego, CA: Academic Press.
- Snell EE (1958) Chemical structure in relation to biological activities of vitamin B6. *Vitamins and Hormones* 16: 77–125.

Pyrimidine Biosynthesis and Degradation (Catabolism)

M Löffler and E Zameitat, Philipps-University, Marburg, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

Antipyrinimidine A structural analog that is taken up by cells and interferes with normal metabolism. In a broad sense, drugs which reduce the production of pyrimidine nucleotides.

Flavoenzyme Enzyme containing a flavin group as redox cofactor.

Metabolon Physically associated enzymes of a metabolic pathway to facilitate substrate channeling.

Nucleoside Consists of [1 base+1 (deoxy)ribose] such as uridine.

Nucleotide Consists of [1 base+1 (deoxy)ribose+1–3 phosphate groups] such as deoxythymidine triphosphate (dTTP) or NTP (nucleoside triphosphate; N, base not specified).

Salvage Recycling of preformed nucleosides by cells.

De Novo Synthesis

Formation of Uridine Monophosphate

Chemically, pyrimidines are heterocyclic six-membered ring structures containing two nitrogen atoms. The pyrimidine bases of nucleic acids possess an amino or hydroxyl group at position 4, and always an oxygen function at position 2. Thymine, in addition, has a methyl group at position 5. This gives rise to tautomeric structures for the bases uracil, cytosine, and thymine. The first pyrimidine nucleotide in cells, uridine monophosphate (UMP), is synthesized from CO₂, the glutamine amide group, aspartate, and phosphoribosyl-1-pyrophosphate (PRPP) (Figures 1 and 2). In contrast to the biosynthesis of purine nucleotides, the pyrimidine ring is assembled first and then linked to ribose phosphate to form a pyrimidine nucleotide. In higher eukaryotes, the cytosolic enzyme catalyzing formation of carbamoyl phosphate, carbamoylphosphate synthetase II (CPSase II) (Figures 2 and 1, no. 1), is distinctly different from carbamoylphosphate synthetase I, which is found in mitochondria as the first reaction of arginine biosynthesis and urea cycle in mammalian liver, and thus CPSase II is the first step committed to pyrimidine metabolism in higher animals. The cleavage of two adenosine triphosphate (ATP) is required to form the high-energy-phosphate bond of carbamoyl phosphate. The formation of *N*-carbamoylaspartate (aspartate transcarbamoylase (ATCase), no. 2) is the committed step for pyrimidine synthesis in prokaryotes. Next, the pyrimidine ring is cyclized on dehydration by dihydroorotase (DHOase, no. 3). In higher animals, the first three enzymes together with glutaminase (which provides the amino group from the side-chain amide group of glutamine) have been fused together (exon shuffling) during evolution to form a single polypeptide, CAD (CPSaseII + ATCase + DHOase). This protein is located mainly in the cytosol and to some extent in the nucleus. Plants, such as prokaryotes, possess these enzymes individually encoded and located in chloroplasts. Formation of orotate from dihydroorotate is catalyzed by the mitochondrial flavoenzyme dihydroorotate dehydrogenase (DHODH, no. 4) transferring the reducing equivalents to the proximal acceptor ubiquinone (Q, Figure 2) and final acceptor molecular oxygen in the

respiratory chain. Thus, the energy of dihydroorotate oxidation can be stored in ATP. Orotate phosphoribosyltransferase (OPRTase, no. 5) and orotidine decarboxylase (ODCase, no. 6) are fused on one bifunctional polypeptide, UMP synthase, operating in the cytosol of higher eukaryotes except in plants. Thus, only three genes (chromosomal location in Man: 2p21(CAD), 16q22 (DHODH), and 3q13 (UMP synthase)) code for the enzymes catalyzing the reaction of *de novo* UMP synthesis.

Regulation

Pyrimidine *de novo* synthesis is a demonstration that compartmentation within different parts of the cell – cytosol, mitochondria, and chloroplasts – enhances the mode of regulation and flexibility of a metabolic pathway. Because of gene duplication, eukaryotes now have the potential to compartmentalize the synthesis of carbamoylphosphate for either pyrimidines (cytosol), or arginine and urea (mitochondria). In humans, urea synthesis takes place only in the liver. The location of DHODH (Figure 2) in the inner mitochondrial membrane and its connection to the functional respiratory chain ensures the most efficient oxidation of dihydroorotate in aerobes. In turn, pyrimidine biosynthesis becomes a pacemaker for cell growth and proliferation under limited oxygen tension. Patients with acquired or inherited defects in mitochondria energetics, however, may suffer from a concomitant pyrimidine nucleotide starvation or imbalance. The two multienzyme polypeptides (CAD and UMP synthase) achieve a micro-compartmentation of pyrimidine *de novo* synthesis. Encoding five enzyme activities on two polypeptides simplifies the coordination and regulation of gene expression and activity. In addition, the direct transfer of a product from one active site to the next enzyme (channeling) considerably increases the efficiency of pyrimidine biosynthesis in higher eukaryotes. Evidence for the location of CAD and UMP synthase close to the mitochondria surface has been obtained in an electron microscopic study. A metabolon-like complex facilitating the movement of substrate and product between cytosol and mitochondrion would allow a rapid increase in *de novo* synthesis when needed. The molecular regulation of pyrimidine *de novo* synthesis occurs at step 1

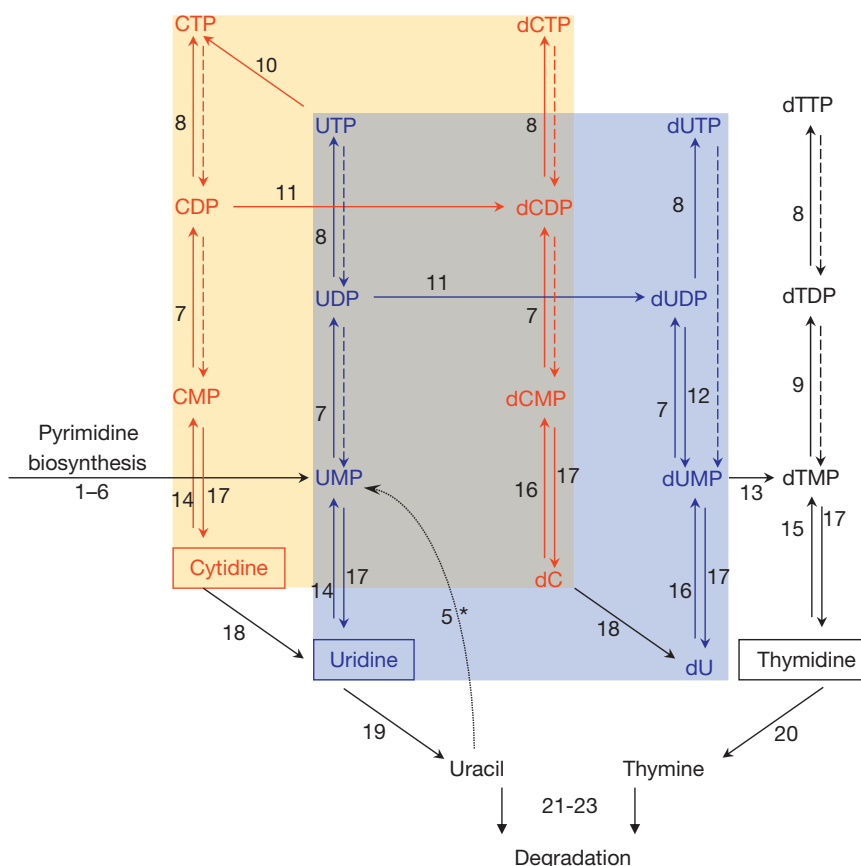


Figure 1 Biosynthesis, interconversion and degradation of pyrimidine nucleotides. The enzymes (given by numbers in the scheme) perform the *de novo* synthesis (1–6, go to [Figure 2](#)), interconversion of UMP to other ribonucleotides and deoxyribonucleotides (7–13), the salvage of preformed metabolites from dietary products or catabolism of nucleic acids (14–18), the first steps of catabolism (19–20) and the degradation (21–23, explained in [Figure 4](#)): 1, carbamoyl phosphate synthetase; 2, aspartate transcarbamoylase; 3, dihydroorotate dehydrogenase; 4, dihydroorotate dehydrogenase; 5, orotate phosphoribosyltransferase; 6, orotidine 5′-monophosphate decarboxylase; 7, uridine(pyrimidine) monophosphate kinases; 8, pyrimidine diphosphate kinases; 9, thymidine monophosphate kinase; 10, cytidine triphosphate synthetase; 11, ribonucleoside diphosphate reductase; 12, dUDP phosphohydrolase; 13, thymidylate synthase; 14, uridine/cytidine kinase; 15, thymidine kinase; 16, deoxycytidine/deoxyuridine kinase; 17, pyrimidine 5′-nucleotidases; 18, (deoxy)cytidine deaminase; 19, uridine phosphorylase; 20, thymidine phosphorylase; 21, dihydropyrimidine dehydrogenase; 22, dihydropyrimidinase; 23, ureidopropionase. → (pyro)phosphohydrolases, not specified here; 5*, uptake of 5-fluorouracil by orotate phosphoribosyl transferase in malignant cells.

of CAD. The CPSase II specific activity is low relative to that of the subsequent enzymes. It is allosterically activated by PRPP and feedback inhibited by uridine triphosphate (UTP). The CAD activity is tuned by phosphorylation at two sites: mitogen-activated protein (MAP) kinase action leads to more easily activated enzyme by reducing the efficacy of UTP and increasing that of PRPP, and phosphorylation at a second site by the cyclic adenosine monophosphate (cAMP)-dependent protein kinase A leads to a more easily inhibited enzyme. Hence, by losing feedback inhibition, carbamoyl phosphate synthesis is activated rapidly and sensitively in response to external growth-promoting signals. At the end of S-phase of the cell cycle, the inhibition by UTP is more pronounced, and the activation by PRPP is reduced. A modulation of ATCase activity is assured by a cooperative binding of carbamoyl phosphate and aspartate. An overflow into pyrimidine biosynthesis can occur from overabundant carbamoyl phosphate production in the case of urea cycle defects in liver: the carbamoyl phosphate leaves the mitochondria

and is used by the ATCase of CAD. Under these conditions, UMP synthase is the limiting enzyme, presumably because of an insufficient supply of PRPP for orotidine monophosphate (OMP) synthesis, causing orotate to accumulate and to be excreted (secondary orotic aciduria).

Deficiency and Interference

The only hereditary but rare defect of pyrimidine *de novo* synthesis diagnosed for certain in living humans is that of UMP synthase (orotate phosphoribosyltransferase (OPRT) domain or orotidine decarboxylase (ODC) domain) with different clinical manifestations, such as anemia or megaloblastic changes in bone marrow, immunodeficiency and gross orotate excretion in type 1 deficiency or orotidine plus orotic acid excretion, neurological deficits, and intellectual impairment in type 2 deficiency. Only recently, by exome sequencing, the presence of DHODH gene mutations was identified as a putative cause for a rare malformation disorder

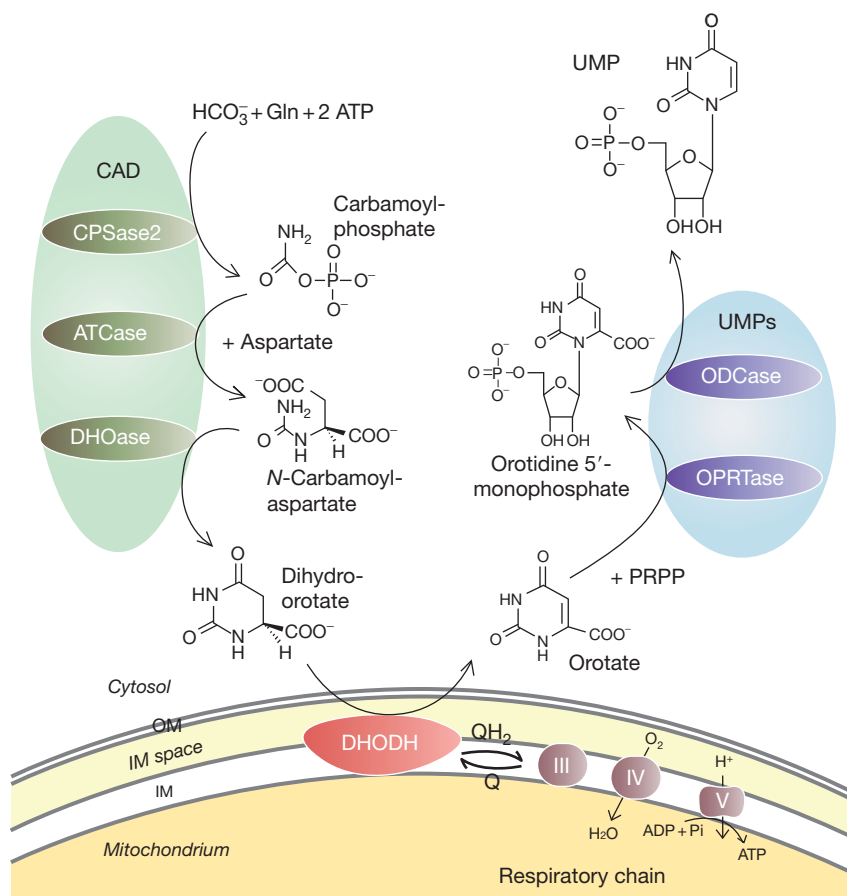


Figure 2 Compartmentation of pyrimidine *de novo* synthesis. In higher eukaryotes the multifunctional enzyme CAD (CPSase 2 + ATCase + DHOase, no. 1 + 2 + 3) and UMP synthase (OPRTase+ODCase, no. 5 + 6) are cytosolic, the monofunctional DHODH (no. 4) is in mitochondria and connected via ubiquinone (Q/QH₂) to the functional respiratory chain at the stage of complex III (III). Electrons are transferred to oxygen via cytochrome oxidase (IV). ATP synthetase (V). Note that in plants, no. 1, 2, 3 and no. 5 + 6 are in chloroplasts. In Gram-positive bacteria and anaerobic yeasts, all enzymes are cytosolic. In Gram-negative bacteria, the quinone-dependent DHODH is attached to the plasma membrane; in many parasites, it is located in mitochondria.

named 'Miller syndrome'. Such new strategies of genetic analyses will guide concerted diagnostics – which have been greatly improved for the detection of intermediates of pyrimidine metabolism in body fluids through tandem mass spectrometry – rather than wide-screening protocols for inborn errors of metabolism. A hereditary defect of the CAD enzyme is not known to date.

There are no specific inhibitors of CPSase of chemotherapeutic interest. *N*-Phosphonoacetyl-L-aspartate (PALA), a model inhibitor of ATCase in prokaryotes, was clinically disappointing because gene amplification leading to overproduction of CAD caused resistance of human cancer cells. Since parasites such as *Plasmodium* or *Toxoplasma* synthesize pyrimidine nucleotides via the *de novo* pathway only – unlike purine nucleotides which are synthesized via a salvage pathway – quite a lot of efforts have been made to solve the three-dimensional structure of these enzymes with the rationale to develop potent inhibitors of utmost specificity against these and other parasites. The drug design for DHOase inhibitors focuses on its thiol Zn²⁺ center.

Antimetabolites of UMP synthase with anticancer activity but limited by toxicity at high dosage to patients (e.g., 6-azauracil and pyrazofurin) have been shown to suppress reproduction of *Toxoplasma gondii* and *Plasmodium falciparum*. Some widely used drugs, for example, the calcium channel blockers nifedipine and nimodipine, were shown to suppress murine OMP decarboxylase and uridine kinase, comparable to barbituric acid as orotidine analog if it is converted to barbiturate ribonucleotide. Natural and chemical ubiquinone analogs, potent inhibitors against DHODH, failed to get clinical approval as anticancer drugs. Severe impairment of other ubiquinone-binding enzymes was assumed. The naphthoquinone atovaquone, a competitor for ubiquinone-dependent reactions in parasites, has been applied for treatment of opportunistic infections and malaria. The inhibition of DHODH by the isoxazol leflunomide, which was licensed in 2000, is central to the efficacy of this drug in treating rheumatoid arthritis and other autoimmune diseases. The present concept to find new drugs for the modulation of pyrimidine nucleotide pools in immune cells is concentrated on DHODH.

Interconversions in Pyrimidine Biosynthesis

Formation of Cytidine and Thymidine Nucleotides

Pyrimidine nucleoside di- and triphosphates are produced by ATP-dependent kinases (Figure 1, no. 7–9). Cytidine triphosphate (CTP) synthase (no. 10) catalyzes the formation of CTP from UTP using the amide group of glutamine in an ATP-dependent reaction (Figure 3(a)). Only at the time

of DNA replication, the deoxynucleotide pools increase noticeably. Nucleoside 5'-diphosphate reductase (ribonucleotide reductase, no. 11) catalyzes the formation of deoxyribose (exchange of H for OH at C2') of pyrimidine as well as purine ribonucleotides using dicysteinyl proteins of the thioredoxin/glutaredoxin superfamily of small redox proteins with final nicotinate adenine dinucleotide phosphate, reduced form (NADPH) as hydrogen donor (Figure 3(b)).

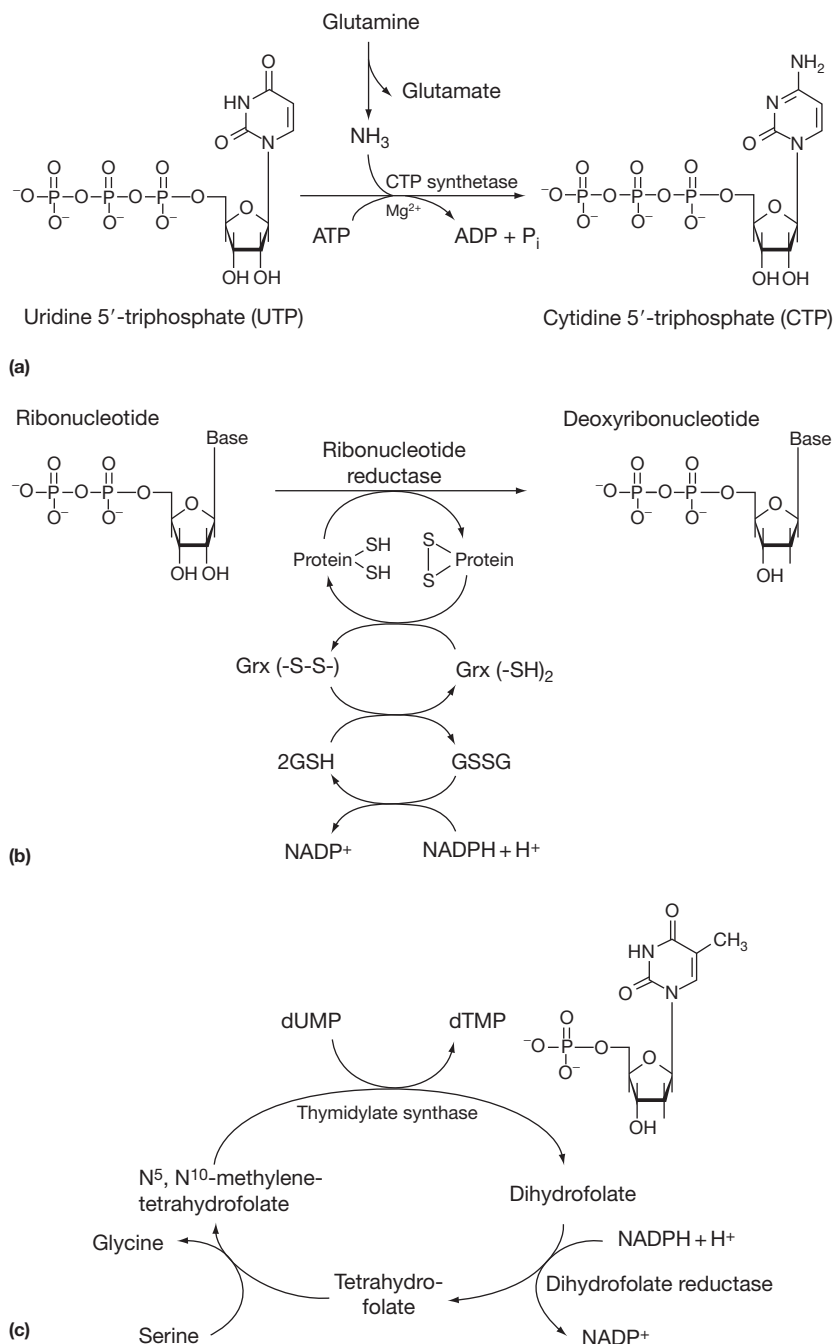


Figure 3 Formation of cytidine and thymidine nucleotides. (a) The synthesis of CTP from UTP by CTP synthetase. (b) Conversion of UMP to dUMP by ribonucleotide reductase, schematic presentation of the enzyme with affiliated glutaredoxin (Grx) and glutathione (GSH/GSSG) reductase catalyzed reactions. (c) The thymidylate synthase catalyzed reaction and regeneration of methylenetetrahydrofolic acid (THF) by means of dihydrofolate reductase.

The product deoxyuridine diphosphate (dUDP) is dephosphorylated to deoxyuridine monophosphate (dUMP) (no. 12). In case of deoxyuridine triphosphate (dUTP) formation, this can effectively be reconverted to dUMP by the deoxyuridine triphosphate nucleotidohydrolase (pyrophosphohydrolase, [Figure 1](#)) in order to prevent its incorporation into DNA. dUMP is used by thymidylate synthase (no. 13) to form deoxythymidine monophosphate (dTMP) by means of (N^5 , N^{10} -) methylenetetrahydrofolic acid (THF) as a one-carbon donor and reducing agent ([Figure 3\(c\)](#)).

Regulation

The enzymes of this part of pyrimidine synthesis are cytosolic. Their activity is strictly controlled by availability of substrates or by their products and other ribonucleotides and deoxyribonucleotides, respectively. This is a well-understood mechanism to ensure well-balanced pools of pyrimidine and purine (deoxy)ribonucleotides in cells. For example, CTP is a negative effector and the purine nucleotide guanosine triphosphate (GTP) is an activator of CTP synthetase, which catalyzes the rate-limiting step in the synthesis of cytidine nucleotides. This enzyme and CAD were described to be associated with the cytoskeleton, hence are able to move around according to change in cell shape. Ribonucleotide reductase (no. 11) of all eukaryotes and some bacteria consists of two nonidentical protein subunits, one containing the substrate-binding site and different effector-binding sites, the other a stable tyrosyl radical generated by the reaction of molecular oxygen at a nonheme binuclear iron center. This reaction gives the second control point for nucleic acid synthesis by O_2 in aerobes. Anaerobes have simpler monomeric or dimeric structures with different metal or coenzyme B12 centers. The key event in all ribonucleotide reductase proteins is the initiation of a transient thiyl radical at the substrate site by means of the radical at the second subunit. In eukaryotes, this enzyme is subject to a sophisticated control mechanism, in which an excess of one deoxyribonucleotide inhibits reduction of all other ribonucleotides. Effective inhibition by deoxyadenosine triphosphate (dATP) or deoxythymidine triphosphate (dTTP) explains the intrinsic toxicity of elevated levels of deoxyadenosine and thymidine to many organisms and also the immunodeficiency in patients with adenosine deaminase and purine nucleoside phosphorylase deficiency. ATP is the main activator of ribonucleotide reductase, and deoxycytidine triphosphate (dCTP) and dTTP are major positive or negative effectors for the interconversion and salvage of deoxyribonucleosides. The key regulatory mechanism for thymidylate synthase is the delivery of THF by dihydrofolate reductase ([Figure 3\(c\)](#)). In plants and parasitic protozoa, this enzyme is fused as a bifunctional polypeptide with thymidylate synthase.

Deficiency and Interference

Effective inhibitors of ribonucleotide reductase (e.g., the radical scavenger hydroxyurea, HU) are potent inhibitors of DNA synthesis and hence of replication. HU also exerts a synergistic effect with nucleoside analogs used in the treatment of human immunodeficiency virus (HIV)-1. This is attributed to reduction in intracellular deoxyribonucleotides coupled with interference

in antigen-dependent T-lymphocyte activation. However, the clinical application is disputed. It is noteworthy that healthy human lymphocytes do not pick up exogenous orotic acid, nor do they convert UTP to CTP. However, both these pathways are upregulated in malignancy. Most of the CTP synthetase inhibitors that have been developed and tested with experimental tumors are synthetic pyrimidine analogs (e.g., deazauridine and cyclopentenyl cytosine). Since some protozoan parasites have limited capacity for CTP formation, the application of CTP synthetase inhibitors was suggested for the treatment of sleeping sickness. Thymidylate synthase and dihydrofolate reductase are well-established target enzymes in cancer therapy using 5-fluorouracil and folate analogs (e.g., methotrexate) preventing THF regeneration. In cancer cells, the drug 5-fluorouracil can be activated to 5-F-UMP by OPRFase with PRPP (no. 5) or by uridine phosphorylase with ribose-1-phosphate (no. 9, in a reverse reaction).

Salvage

Phosphorylation and Dephosphorylation

The fact that hereditary orotic aciduria does respond to life-long supply with uridine but not to uracil indicated that pyrimidine nucleoside salvage can compensate for the lack of *de novo* synthesis in humans, if the dosage as food additive is in grams per day and increasing from childhood to adulthood. This is in direct contrast to purines where the gut mucosa contains a complement of enzymes which degrades all dietary purines to the end product, uric acid. Prominent salvage capacity is expressed in lymphoid cells, polymorphonuclear cells, and the central nervous system. Uridine, (deoxy)cytidine, and thymidine originating from daily nutrients or intracellular breakdown from nucleic acids are rescued from circulation by cytosolic (deoxy)ribonucleoside kinases that show a high level of specificity with respect to the base and sugar moieties, with ATP as the major phosphate donor (no. 14–16). In contrast to purines, the salvage of free pyrimidine bases at the expense of PRPP does not occur in mammalian tissue. Orotate as a natural component of dairy food is taken up and converted to UMP only by liver and erythrocytes. It can also stem from food additives such as zinc orotate and magnesium orotate, which are used for metal ion substitution therapy. Salvage of deoxycytidine is important in DNA repair. Deoxycytidine kinase is also responsible for activating some nucleoside analogs as well as natural deoxyribonucleosides – such as deoxyadenosine in adenosine deaminase deficiency. Pyrimidine and purine deoxyribonucleoside kinases seem to have evolved from a common progenitor multisubstrate kinase. Multisubstrate deoxyribonucleoside kinases were found in several insect species and shown to phosphorylate pyrimidines as well as purine (deoxy)nucleosides with high efficiency. The dephosphorylation of pyrimidine (deoxy)ribonucleoside monophosphates is catalyzed by 5'-nucleotidases (no. 17) which are expressed as specific isoforms in different tissues. This reaction is a prerequisite for transport and delivery of pyrimidines by the circulation. In cells, uridine and deoxyuridine can also be formed from cytidine and deoxycytidine by cytidine deaminase (no. 18).

Regulation

The balance between excretion and uptake is set by intracellular substrate cycles involving 5'-nucleotidases, kinases, and nucleoside transport through the plasma membrane of cells. Uridine is taken up effectively by a carrier-mediated concentrated transport process or by cotransport, by all human cells except erythrocytes.

Deficiency and Interference

Different mutants of the pyrimidine-5'-nucleotidase (no. 17) are known to cause enzyme deficiency. This is inherited as an autosomal recessive trait. The symptoms include nonspherocytic hemolytic anemia, splenomegaly, and hepatomegaly; gross elevations in erythrocyte UTP/CTP are characteristic findings. A different (acquired) deficiency of 5'-nucleotidase associated with hemolytic anemia has been described in patients with lead poisoning or thalassemia. Uridine 5'-nucleotidase superactivity was described to include developmental delay, recurrent infections, and hyperactivity of children. Daily treatment with uridine resulted in a remarkable improvement of the clinical condition, as was the case with hereditary orotic aciduria patients. Mutations in the gene of thymidine phosphorylase (no. 20) caused mitochondrial neurogastrointestinal encephalomyopathy in humans. Patients with mutations in the mitochondria-specific thymidine kinase-2 develop myopathy and depletion of muscular mitochondrial DNA. Pyrimidine (deoxy)ribonucleoside kinases are critical constituents for

interference by pharmacological agents; they are of key importance for the channeling of nucleoside analogs (such as arabinosyl cytosine, azidothymidine, and 2'/3'-dideoxycytidine) in cell metabolism of humans and pathogens. On the other hand, strategies for inhibition of pyrimidine biosynthesis by drugs can be anticipated in cells with high activities in the salvage reactions. This aspect underlies current research on structure and function of pyrimidine (deoxy)ribonucleoside transporters in the plasma and mitochondria membranes and their connection with salvage and biosynthesis of pyrimidines.

Catabolism of Pyrimidines

Degradation of Uracil and Thymine

The cleavage of uridine and thymidine by phosphorylases (Figure 1, no. 19–20; Figure 4) to the appropriate base and ribose-1-phosphate is the key to the catabolic pathway. Ribose-1-phosphate is transduced via ribose-5-phosphate to the pentose phosphate pathway of cells, whereas deoxyribose-1-phosphate is cleaved by deoxyribose-phosphate aldolase to acetaldehyde and glyceraldehyd-3-phosphate. Although the biochemical strategies used in the *de novo* and salvage pathways of pyrimidines seem to be similar in all organisms, this is not the case for the degradation of the pyrimidine bases. In most eukaryotes and some bacteria, the first step of uracil and thymine degradation is catalyzed by the flavoenzyme dihydropyrimidine dehydrogenase (DPD) (Figure 4, no. 21), which reduces the bases to dihydrouracil and dihydrothymine,

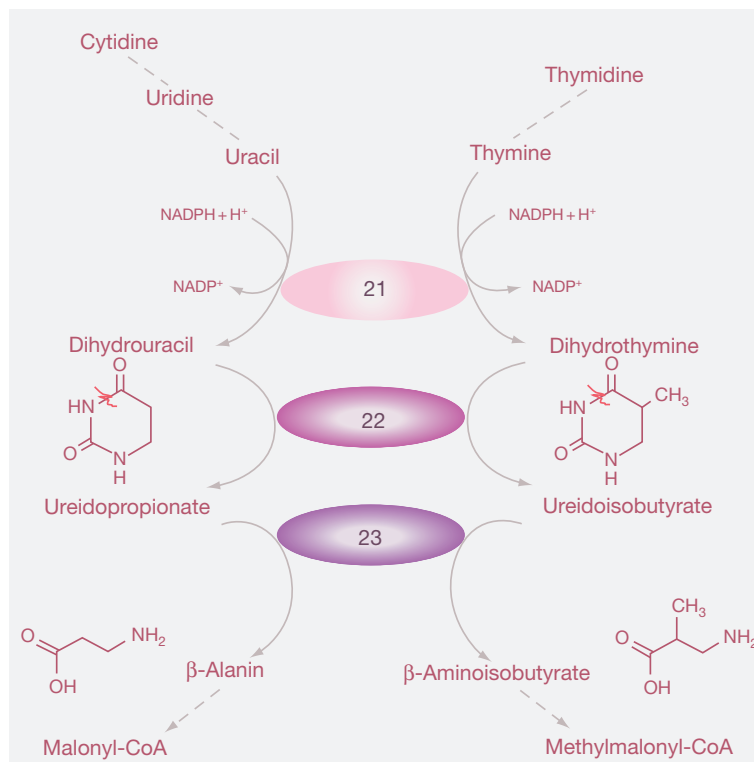


Figure 4 Degradation of pyrimidine bases. Three enzymes (given in numbers) perform the degradation of the pyrimidine ring. Dihydropyrimidine dehydrogenase (21), dihydropyrimidinase (22), and ureidopropionase (23). The formation of malonyl-CoA from β-alanine and methylmalonyl-CoA from β-aminoisobutyrate as well as their final degradation is by usual metabolism and not specified here.

respectively. Hydrolytical cleavage of the pyrimidine ring is performed by dihydropyrimidine amidohydrolase (dihydropyrimidinase, no. 22). Ureidopropionase (no. 23) eliminates CO₂ and NH₃ and produces β-alanin and β-aminoisobutyric acid. These metabolites are excreted in urine of humans or converted to malonyl-CoA and methylmalonyl-CoA, respectively, by different enzymes, as is different the further catabolism through the usual ways for odd-chain fatty acids or branched amino acids. Thus, pyrimidine degradation does not produce problems for the human body – rather it contributes to ATP production – in contrast to urate, which is produced from the purines and can precipitate, causing gout in humans. In some unicellular organisms, other catabolic pathways of uracil have been described, for example, an oxidative one with barbituric acid as intermediate, and malonic acid and urea as the end products.

Regulation

In mammals, dihydropyrimidine dehydrogenase activity (Figure 4, no. 21) catalyzes the rate-limiting step in the degradation pathway. It has been found in the cytosol of cells in many tissues of the body, whereas high activity of dihydropyrimidinase was associated with liver and kidney (Figure 4, no. 22) and that of ureidopropionase with liver only (Figure 4, no. 23). This tissue can guarantee the support with NADPH, required for the reduction of the pyrimidine ring, as well as the speedy and complete catabolism to CO₂ and ammonia.

Deficiency and Interference

The examination of urine and plasma samples for elevated levels of uracil and thymine or other intermediates, set first diagnoses of a complete or partial deficiency of pyrimidine degradation enzymes until molecular analyses allowed identification of quite different mutations. Convulsive disorders, motor retardation, epilepsy, or mental retardation was observed in patients with DPD or dihydropyriminase deficiency. The onset of clinical symptoms usually occurred during childhood. The fundamental importance of DPD for the degradation of pharmacological pyrimidine analogs became evident when patients with partial DPD deficiency suffered from severe neurotoxicity following 5-fluorouracil chemotherapy. β-Alanin is thought to function as neurotransmitter in view of its chemical similarity to the well-known γ-amino-butyric acid (GABA). The altered concentration of β-alanine and the exposure of the nervous system to high concentrations of uracil and thymine have been supposed as contributing factors to the

cerebral dysfunction. Therefore, inhibitors developed for the study of catabolic pathway enzymes are presently not of clinical relevance.

See also: **Bioenergetics:** Calcium, Biological Fitness of; Inositol Trisphosphate and Calcium Signaling; Membrane Transport, General Concepts; Mitochondrial Dynamics; Oxygenases; **Lipids** **Carbohydrates** **Membranes and Membrane Proteins:** Cell–Matrix Interactions; Glycosylphosphatidylinositol Anchors; Neoglycoproteins; **Signaling:** Dopamine Receptors.

Further Reading

- Arner ES and Eriksson S (1995) Mammalian deoxyribonucleoside kinases. *Pharmacology and Therapeutics* 67: 155–186.
- Carrey EA (1995) Key enzymes in the biosynthesis of purines and pyrimidines: Their regulation by allosteric effectors and by phosphorylation. *Biochemical Society Transactions* 23: 899–902.
- Connolly GP and Duley JA (1999) Uridine and its nucleotides: Biological actions, therapeutic potentials. *Trends in Biochemical Sciences* 20: 218–225.
- Evans DR and Guy HI (2004) Mammalian pyrimidine biosynthesis: Fresh insights into an ancient pathway. *Journal of Biological Chemistry* 279: 33035–33038.
- Follmann H (2004) Deoxyribonucleotides: The unusual chemistry and biochemistry of DNA precursors. *Chemical Society Reviews* 33: 225–233.
- Fox B and Bzik DJ (2002) *De novo* pyrimidine biosynthesis is required for virulence of *Toxoplasma gondii*. *Nature* 415: 926–929.
- Hatse S, DeClercq E, and Balzarini J (1999) Role of antimetabolites of purine and pyrimidine nucleotide metabolism in tumor cell differentiation. *Biochemical Pharmacology* 58: 539–555.
- Jones ME (1980) Pyrimidine nucleotide biosynthesis in animals: genes, enzymes, and regulation of UMP biosynthesis. *Annual Review of Biochemistry* 49: 253–279.
- Lee M, Maher MJ, and Christopherson RI (2009) Dihydroorotase inhibitors. In: Supuran CT and Winum JY (eds.) *Drug Design of Zn-Enzyme Inhibitors*, pp. 949–977. New York, NY: Wiley.
- Löffler M, Fairbanks L, Zameitat E, Marinaki T, and Simmonds A (2005) Pyrimidine pathways in health and disease. *Trends in Molecular Medicine* 11: 430–437.
- Piskur J, Sandrini MP, Knecht W, and Munch-Petersen B (2004) Animal deoxyribonucleoside kinases: Forward and retrograde evolution of their substrate specificity. *FEBS Letters* 560: 3–6.
- Piskur J, Schnackerz KD, Andersen G, and Björnberg O (2007) Comparative genomics reveals novel biochemical pathways. *Trends in Genetics* 23: 369–372.
- Schaefer CM, Schäfer MKH, and Löffler M (2010) Region-specific distribution of dihydroorotate dehydrogenase in the rat central nervous system points to pyrimidine *de novo* synthesis in neurons. *Nucleosides Nucleotides Nucleic Acids* 29: 476–481.
- Stubbe JA (1998) Ribonucleotide reductases in the twenty-first century. *Proceedings of the National Academy of Sciences of the United States of America* 95: 2723–2724.
- Van Kuilenburg ABP, Van Lenthe H, Löffler M, and Van Gennip AH (2004) Analysis of pyrimidine synthesis *de novo* intermediates in urine and dried filter stripes with HPLC tandem mass spectrometry. *Clinical Chemistry* 50: 2117–2124.
- Voet D and Voet JG (2004) *Biochemistry*, 3rd edn. New York, NY: Wiley.
- Webster D, Becroft DMO, Van Gennip AH, and Van Kuilenburg ABP (2001) Hereditary orotic aciduria and other disorders of pyrimidine metabolism. In: Scriver CR, et al. (ed.) *The Metabolic and Molecular Bases of Inherited Disease*, vol II, 8th edn., pp. 2663–2702. New York, NY: McGraw-Hill.

Pyruvate Kinase

K Uyeda, University of Texas Southwestern Medical Center & Veteran Affairs Medical Center, Dallas, TX, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Carbohydrate Response Element A DNA sequence within the promoter of a gene to which a specific transcription factor binds in response to excess carbohydrate.

Gene Promoter A region of a gene that responds to glucose, hormones, and other stimuli and binds a number of transcription factors that affect transcription.

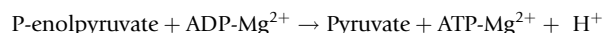
Liver Pyruvate Kinase An isoenzyme of pyruvate kinase that occurs mainly in liver.

Pyruvate Kinase M2 (fetal isoform) A fetal isoenzyme of pyruvate kinase M₂ may aid in promotion of cancer cell growth.

Transcription Synthesis of messenger RNA (mRNA) transcribed from a specific sequence of DNA of a gene.

PK Reaction

Pyruvate kinase (PK) (EC 2.7.1.40) catalyzes the final step of glycolysis:



The enzyme requires two equivalents of divalent cation and K⁺ for a full activity. One divalent cation complexes with adenosine triphosphate (ATP) or adenosine diphosphate (ADP) and the other cation binds to the protein and functional groups of pyruvate. The γ -phosphate of ATP bridges the two divalent cations.

Structure

All PKs from a variety of sources are tetramers (Mr = 240 kDa). Mammals express four isozymes designated as M1 and M2 for muscle type, L for liver, and R for erythrocytes. M1 muscle isozyme is the only PK exhibiting hyperbolic kinetics for PEP. However, in the presence of the allosteric ligand such as phenylalanine, it shows cooperative saturation for PEP. All other isozymes exhibit cooperativity with respect to PEP and are activated by fructose 1,6-diphosphate (Fru1,6-P₂) by lowering K_{PEP}.

PK has been isolated and characterized from a number of prokaryotes and from eukaryote tissues. Each consists of 500 amino acid residues. Crystal structures of the enzymes from cat and rabbit muscle, yeast, *Escherichia coli*, and *Leishmania mexicana* have been reported (1 and 2 (review)). They all exhibit a very similar three-dimensional (3-D) structure (Figure 1 (a)). Each subunit contains three domains: the A domain with the classic (α/β)₈ structure; the B domain with an irregular fold; and the C domain with an α - and β -organization (Figure 1 (b)). The eukaryotic enzymes contain additional short N-terminal peptides. The active site is situated on the C-terminal side of the A domain, facing the cleft between the A and B domains. The Fru 1,6-P₂ complex of yeast PK has revealed that the Fru 1,6-P₂-binding site is located in the C (regulatory) domain, which is 40 Å away from the catalytic center.

Regulation of Liver PK *in vitro*

Effect of pH

In alkaline pH, liver PK (LPK) exhibits a sigmoidal saturation curve with PEP, but at lower pH the enzyme shows Michaelis–

Menten kinetics. The transition between these two enzyme conformational states occurs at a narrow pH range of 6.8–7.3.

Allosteric Effectors

Fru 1,6-P₂ reduces the cooperative effects of PEP and converts the enzyme to a form exhibiting Michaelis–Menten kinetics with a higher affinity for the substrate. ATP and alanine are inhibitors of LPK, and these ligands raise the apparent K_{PEP} to higher values. The inhibitory effect of ATP and alanine is counteracted by PEP and Fru 1,6-P₂. Pig LPK is inhibited by phenylalanine more strongly than alanine.

Phosphorylation/Dephosphorylation

LPK was first shown to be regulated by phosphorylation catalyzed by cyclic adenosine monophosphate (cAMP)-dependent protein kinase (PKA), resulting in loss of the enzyme activity. The phosphate incorporation is 1 mol phosphate per mole enzyme subunit on the N-terminal Ser within the consensus PKA target sequence (ArgArgAlaSerPVal). When the phosphorylated LPK was dephosphorylated by histone phosphatase, the enzyme recovered full activity.

The major effect of phosphorylation is to raise the apparent K_{PEP} approximately threefold (from 0.3 to 0.8 mM). It also increases the Hill coefficient from 1.1 to 1.5. V_{max} remains the same for both enzyme forms. This inhibitory effect of phosphorylation is counteracted by Fru 1,6-P₂ and H⁺. Phosphorylation does not inhibit the rat LPK in the presence of 5 μM Fru 1,6-P₂ at pH 7.3, and the enzyme follows Michaelis–Menten kinetics with K_{PEP} of 0.04 mM. Phosphorylated LPK is more sensitive to inhibition by ATP and alanine than is the dephospho form.

Crystallographic Studies

Crystallographic and mutagenesis studies provided the mechanistic features for the allosteric transition in PK at the molecular level. The enzyme displays significant conformational changes involving all three domains of each subunit in the tetrameric enzyme in moving from the T- to R-state (Figure 2). This model is based on the X-ray structure determination of the allosteric *E. coli* PK in the inactive T-state and the M1 isozyme, which is an active R-conformation. In general, the allosteric

transition involves two types of movements: (1) the rotation of the B and C domains (17° and 15°) within each subunit and (2) a 16° rotation of each subunit within the tetramer. Thus, the allosteric activation of PK appears to involve the rotation of both domain and subunit. This model is further supported by the observation that the 3-D structure of each

domain is not altered by their movement, suggesting that they rotate as rigid bodies connected by flexible hinges. The transition from R- to T-state also causes significant changes in the active site including distortion of the PEP-binding site, suggesting the poor substrate affinity of the T-state. It also shrinks the cleft between the A and C domains preventing effector binding.

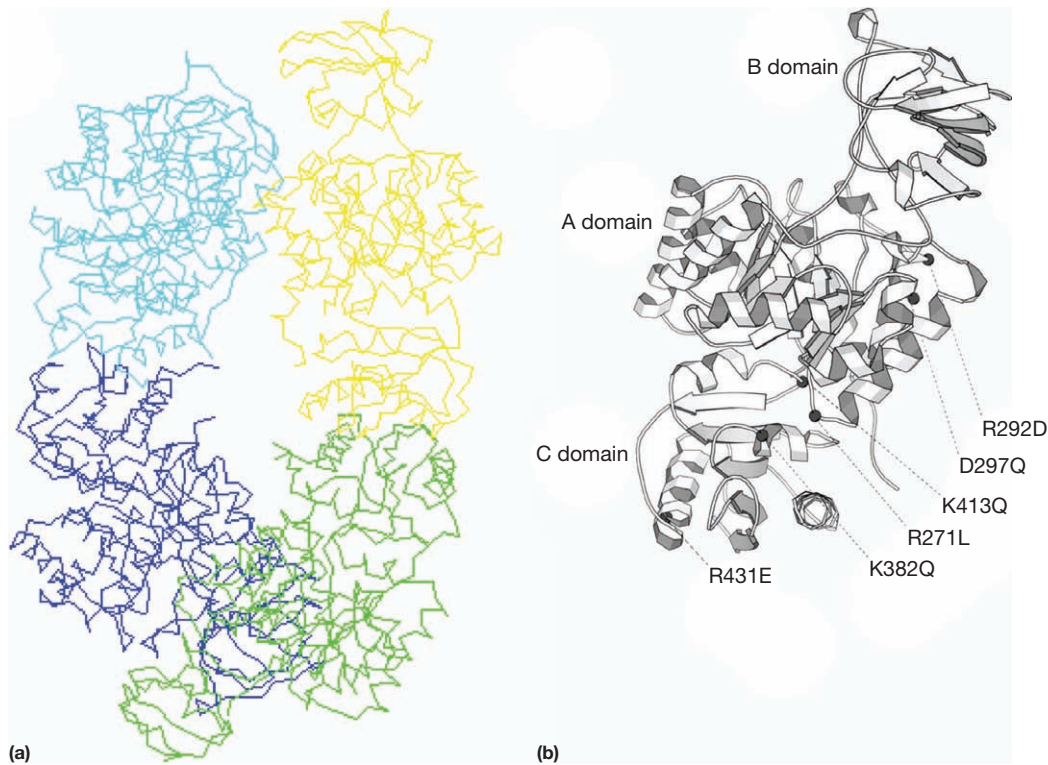


Figure 1 (a) Tetrameric *E. coli* PK in inactive T state. (b) *E. coli* PK subunit. The orientation is same as in (a). Reproduced from Valentini G, Chiarelli L, Fortin R, et al. (2000). The allosteric regulation of pyruvate kinase. *Journal of Biological Chemistry* 275: 18145–18152, with permission.

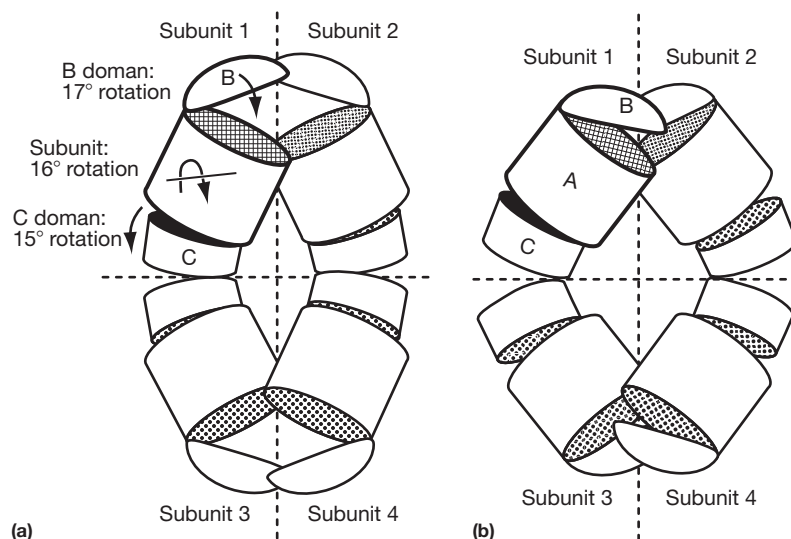


Figure 2 Schematic representation of the domain and subunit rotations in the T- (left) and R- (right) state transition of PK. Reproduced from Mattevi A, Bolognesi M, and Valentini G (1996). The allosteric regulation of pyruvate kinase. *FEBS Letters* 389: 15–19, with permission.

These kinetic changes induced by various ligands and by phosphorylation are the acute regulation of PK activity by hormonal and nutritional states. Experiments performed on intact animals confirm the regulation of PK *in vivo*. Intravenous administration of glucagon causes rapid inactivation of PK in liver. The effect of glucagon also raises cAMP concentration. Insulin causes rapid activation of PK.

Regulation of LPK Concentration by Gene Expression

Effect of Insulin, Glucagon, and Carbohydrate in Liver

In long-term regulation, LPK gene expression is controlled also by hormones and nutrients. Feeding mammals a diet rich in carbohydrate and low in fat leads to an immediate increase in the insulin levels and a fall in glucagon. Insulin and glucagon can activate and repress, respectively, the LPK transcription. These hormonal effects are mediated by direct response to the hormonal signals acting via membrane receptors.

In addition to increased insulin secretion, the excess glucose leads to activation of LPK gene transcription. This activation of gene expression is mediated by glucose, independent of insulin, and requires metabolism of glucose. Thus, glucose is not only a source of energy but also is able to stimulate glycolytic, and lipogenic enzyme genes independent of insulin.

Carbohydrate Response Element of LPK Promoter

The LPK gene promoter contains the glucose response element (−168 to −145 nucleotides upstream of open reading frame) and several other nuclear factor (NF)-binding sites including NF and HNF4. The transcription factor binding to the glucose response element was identified recently and termed carbohydrate response element-binding protein (ChREBP). In addition, it contains multiple phosphorylation sites, subject to regulation by protein kinases/phosphatase.

These kinetic changes induced by various ligands and by phosphorylation are the acute regulation of PK activity by hormonal and nutritional states. Experiments performed on intact animals confirm the regulation of PK *in vivo*. Intravenous administration of glucagon causes rapid inactivation of PK in liver. The effect of glucagon also raises cAMP concentration. Insulin causes rapid activation of PK.

PK M2 (Fetal Isoform) Mediates the Warburg Effect

For over a century now, it has been appreciated that O₂-deprived cells exhibit increased conversion of glucose to

lactate. Otto Warburg observed 70 years ago that cancer cells metabolize glucose more rapidly than normal cells under aerobic conditions. This has become known as the Warburg effect or an enhanced rate of aerobic glycolysis. The mechanisms for these effects are not understood. A recent report has demonstrated that a fetal isoform of PK, PKM2, binds phosphotyrosine (p-Tyr) peptides and may aid in promoting cancer cell growth. No other PK isoforms possess this property, and cancer cell lines exclusively express the PKM2 isoform. Investigators have found that the p-Tyr binding promotes the release of a tightly bound allosteric activator, fructose 1,6-bisphosphate, resulting in the inhibition of enzyme activity for PKM2. It has also been demonstrated that this inhibition of PKM2, by the p-Tyr binding, diverts some of the glucose metabolites from energy production via glycolysis/oxidative phosphorylation to anabolic processes, especially when cells are stimulated by growth factors. These specific metabolites have not been identified, but the results suggest that expression of p-Tyr PKM2 may be important for rapid growth of cancer cells.

See also: [Metabolism Vitamins and Hormones: Carbohydrate-Responsive Element Binding Protein; Gluconeogenesis; Glycolysis Overview; Insulin- and Glucagon-Secreting Cells of the Pancreas.](#)

Further Reading

- Christofk HR, Vander Hellen MG, Wu N, Asara JM, and Cantley LC (2008) Pyruvate kinase M2 is a phosphotyrosine-binding protein. *Nature* 452: 181–188.
- Girard J, Ferre P, and Foufelle F (1997) Mechanism by which carbohydrates regulate expression of genes for glycolytic and lipogenic enzymes. *Annual Review of Nutrition* 17: 325–352.
- Mattevi A, Bolognesi M, and Valentini G (1996) The allosteric regulation of pyruvate kinase. *FEBS Letters* 389: 15–19.
- Uyeda K and Repa J (2006) Carbohydrate response element binding protein, ChREBP, a transcription factor coupling hepatic glucose utilization and lipid synthesis. *Cell Metabolism* 4: 107–110.
- Uyeda K, Yamashita H, and Kawaguchi T (2002) Carbohydrate responsive element-binding protein (ChREBP): A key regulator of glucose metabolism and fat storage. *Biochemical Pharmacology* 63: 2075–2080.
- Valentini G, Chiarelli L, Fortin R, et al. (2000) The allosteric regulation of pyruvate kinase. *Journal of Biological Chemistry* 275: 18145–18152.
- Wool JO, Friesen RHE, White MA, et al. (2001) Structural and functional linkages between subunit interfaces in mammalian pyruvate kinase. *Journal of Molecular Biology* 312: 525–540.

Quinones

G Lenaz and M L Genova, Alma Mater Studiorum – Università di Bologna, Bologna, Italy

© 2013 Elsevier Inc. All rights reserved.

Glossary

Antioxidants Molecules capable of preventing the propagation of free radicals and, hence, protecting cellular and extracellular biomolecules from free-radical-induced damage, such as lipid peroxidation and genetic mutations in DNA. There are water-soluble antioxidants, such as ascorbic acid (vitamin C), and lipid-soluble antioxidants, such as α -tocopherol (vitamin E). Ubiquinol belongs to this latter category.

Chloroplasts Organelles present in photosynthetic higher plants and in eukaryotic algae that carry out the process of oxygen-evolving photosynthesis by extracting electrons from water, in a complex electron transfer chain containing plastoquinone among other components, and reducing CO_2 , thus, yielding carbohydrates by a complex synthetic process (Calvin cycle).

Coenzyme A nonprotein compound required for the catalytic action of an enzyme that can be strongly bound to the enzyme itself as a prosthetic group or participate as a substrate to the enzyme-catalyzed reaction. Several coenzymes cannot be synthesized *de novo* and derive from nutritionally supplied vitamins.

Free radical An atom or molecule having one or more unpaired electrons. For this reason, free radicals are usually very reactive, in the attempt to donate or receive an electron, and may damage biological molecules by establishing a chain reaction of propagation followed by permanent chemical

modification of the attacked molecules. Among free radicals are some of the so-called reactive oxygen species, derived by partial reduction of molecular oxygen during mitochondrial respiration and other biological processes.

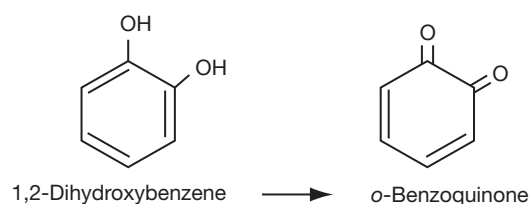
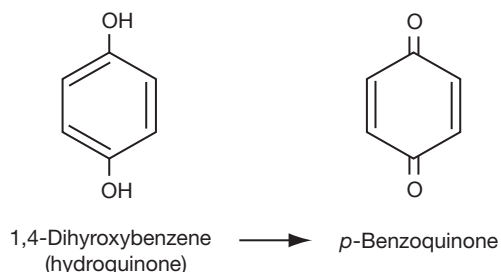
Mitochondria Organelles present in most eukaryotic cells carrying out the process of oxidative phosphorylation and several other metabolic pathways. Mitochondria consist of two membranes, the inner membrane being the site of the respiratory chain, delimiting two aqueous compartments (called the intermembrane space and the matrix).

Supercomplex A stable protein assembly consisting of different respiratory complexes immobilized by strong specific interactions that lead to a nonrandom supramolecular organization of the proteins in the membrane. The function of the supercomplexes appears not to be restricted to kinetic advantages in electron transfer along the respiratory chain, but they have a role in the stability and assembly of the individual complexes and in preventing excess oxygen radical formation. There is increasing evidence that disruption of the supercomplex organization leads to functional derangements responsible for pathological changes.

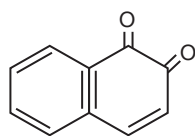
Vitamin A biological molecule, acting in small amounts (therefore, not a structural component), which cannot be synthesized by the body and, therefore, requires dietary supplementation. Many vitamins are part of coenzymes. Vitamins are classified as water soluble and lipid soluble.

Chemistry of Quinones

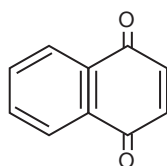
Quinones are diketones derived by oxidation of dihydroxyar- enes; the simplest quinones are those derived by benzenediols (1,2- and 1,4-dihydroxybenzene):



Also condensed arenes give quinones, for example, the naphthoquinones:

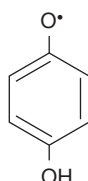


1,2-Naphthoquinone



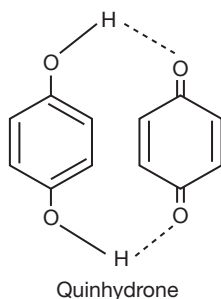
1,4-Naphthoquinone

The reversible oxidation of dihydroxyarenes to the respective quinones is a two-electron process with the intermediate formation of the radical form of semiquinones, for example,



p-Benzosemiquinone

The redox systems quinone/diphenol (or quinol) are widely present in nature. Also known is the charge-transfer complex quinhydrone, where quinone is the electron acceptor while diphenol is the electron donor:



Quinhydrone

Biological Benzoquinones: Coenzyme Q and Plastoquinone

Chemistry and Distribution

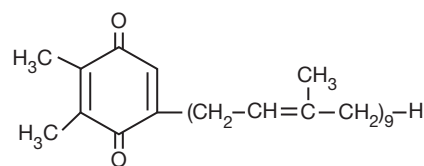
Coenzyme Q was discovered in 1957 by FL Crane in Wisconsin as a yellow oily substance, extracted from bovine heart mitochondria, and having redox properties. It is also called ubiquinone, initially from the European nomenclature. The bovine compound is 2,3-dimethoxy-5-methyl-6-decaprenyl-1,4-benzoquinone (coenzyme Q₁₀), but other homologs exist having polyprenyl chains of different lengths in the 6-position (Table 1). Also, some analogs with different substitutions in the benzene ring have been found in nature or synthesized.

Coenzyme Q homologs were initially found in the inner membrane of mitochondria of eukaryotic cells and in the plasma membranes of aerobic and photosynthetic bacteria and they were discovered to be essential components of respiratory and photosynthetic chains. A different analog, plastoquinone-9 (2,3-dimethyl-5-nonaprenyl-1,4-benzoquinone), was found in chloroplasts of higher plants and algae and in cyanobacteria and proved as an essential component of the photosynthetic apparatus (see structure below).

Table 1 Naturally occurring homologs and analogs of coenzyme Q

Homolog or analog	Molecular weight	Occurrence
Coenzyme Q ₆	590	<i>Saccharomyces cerevisiae</i> and other yeasts
Coenzyme Q ₇	659	Minor component of some yeasts
Coenzyme Q ₈	728	Some yeasts, <i>Escherichia coli</i> , and many other bacteria
Coenzyme Q ₉	794	Rat, mouse, shark, some insects, protozoa, molds, yeasts, basidiomycetes, <i>Pseudomonas fluorescens</i> , and other bacteria
Coenzyme Q ₁₀	862	Most mammals including man, birds, amphibia, most invertebrates, most photosynthetic bacteria, and some other bacteria
Rhodokinone-10	847	<i>Rhodospirillum rubrum</i>
Ubichromenol-10	862	Accompanies Q ₁₀ in minor amounts as a cyclization product
Plastoquinone-9	747	Chloroplasts of higher plants and algae

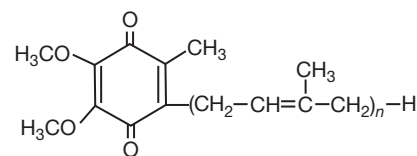
See text for chemical structures. Rhodokinone-10 is 2-methoxy-3-amino-5-methyl-6-decaprenyl-1,4-benzoquinone.



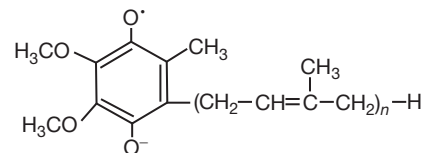
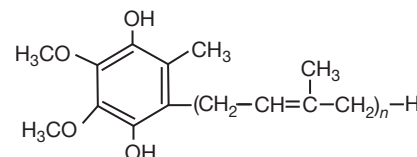
Plastoquinone-9

Soon after its discovery, coenzyme Q was also found to be present in other cellular membranes of eukaryotic cells, including Golgi apparatus, lysosomes, and plasma membrane.

The major redox forms of coenzyme Q, with their prevalent protonation states, are the following:

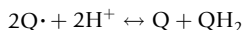


Ubiquinone (Q)

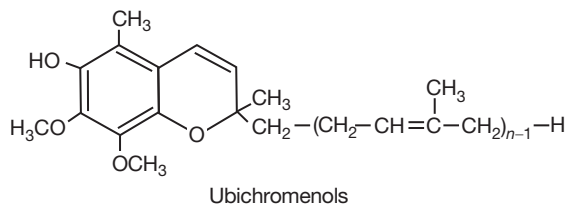

 Ubisemiquinone anion (Q^{•-})

 Ubiquinol (QH₂)

The standard redox potential of the ubiquinone/ubiquinol couple at pH 7 is ~+100 mV. The semiquinone radical is

intermediate in most redox reactions of coenzyme Q; however, it is very unstable and is rapidly dismutated to the quinone and quinol forms:



Under some conditions, ubiquinones undergo cyclization between the ring and the first isoprenoid unit of the side chain forming the respective ubiquinomenols:



Coenzyme Q homologs 1–12 are soluble in most organic solvents but not in water due to their isoprenoid sidechains. Only homologs 6–12 are obtained in crystalline form at room temperature. In water, coenzyme Q homologs form micelles or aggregates. They are soluble in lipids and disperse in monomeric form in phospholipid bilayers as well as in biological membranes.

Coenzyme Q absorbs light in the UV with a maximum at 275 nm in ethanol and at 270 nm in hydrocarbons when in oxidized form (extinction coefficient $15 \text{ mM}^{-1} \text{ cm}^{-1}$), and with a maximum at 290 nm in all organic solvents when in reduced form (extinction coefficient $4 \text{ mM}^{-1} \text{ cm}^{-1}$). The ubisemiquinone anion has a maximum at 318 nm (extinction coefficient $10.7 \text{ mM}^{-1} \text{ cm}^{-1}$) and minor peaks also in the visible spectrum. The semiquinones are best detected by electron spin resonance techniques.

The lateral mobility of coenzyme Q in phospholipid bilayers is very high: using the method of collisional fluorescence quenching with membrane-bound fluorescent molecules, a lateral diffusion coefficient was calculated $\sim 10^{-7} \text{ cm}^2 \text{ s}^{-1}$.

Such mobility seems essential for the function of the quinone in electron transfer chains in energy-transducing membranes.

Functions of Coenzyme Q

Electron transfer chains of mitochondria and bacteria

The best-known function of coenzyme Q is as a redox component of the mitochondrial respiratory chain, localized in the inner membrane. The structural organization of the mitochondrial oxidative phosphorylation (OXPHOS) system has received large attention in the past and most investigations led to the conclusion, accepted till recently, that the respiratory enzymatic complexes are randomly dispersed in the lipid bilayer of the inner membrane and functionally connected by fast diffusion of smaller redox components, coenzyme Q and cytochrome *c*. As shown in the standard scheme of the respiratory chain depicted in Figure 1.

Coenzyme Q was first described as a mobile diffusible component localized between complex I (reduced nicotinamide adenine dinucleotide (NADH) coenzyme Q reductase) or complex II (succinate coenzyme Q reductase) and complex III (ubiquinol cytochrome *c* reductase); in these systems, coenzyme Q is reduced by complex I or II and reoxidized by complex III. It was then found that coenzyme Q is also reduced by other mitochondrial enzymes, such as glycerol-1-phosphate dehydrogenase, electron transfer flavoprotein dehydrogenase (an enzyme involved in fatty acyl CoA oxidation), and dihydro-orotate dehydrogenase.

Due to its extremely hydrophobic nature, coenzyme Q is localized in the hydrocarbon interior of the lipid bilayer of the inner mitochondrial membrane, where its diffusion rate may be very high. In this way it behaves as a common pool of molecules, so that any molecule would be randomly reduced by any complex I or II and reoxidized by any complex III. Major observations in support of this model were the random distribution of the intramembrane particles observed

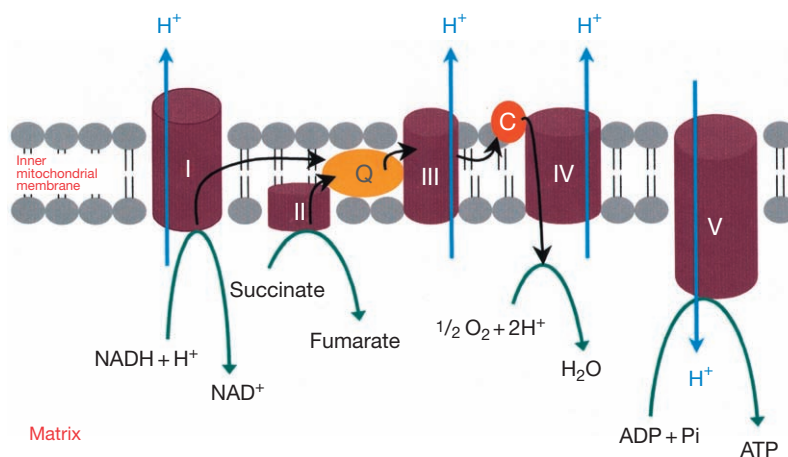


Figure 1 Schematic view of the mitochondrial oxidative phosphorylation system, showing the role of coenzyme Q as a shuttle of electrons between complex I or complex II and complex III. Other dehydrogenases using coenzyme Q as acceptor are not shown for simplicity. Roman numbers refer to the respective enzyme complexes. Black arrows indicate the flux of electrons, green arrows indicate chemical reactions, and blue arrows indicate proton translocation. During redox reactions catalyzed by complex I, III, and IV, protons are extruded forming a transmembrane gradient that is used by complex V (adenosine triphosphate (ATP) synthase) for synthesizing ATP.

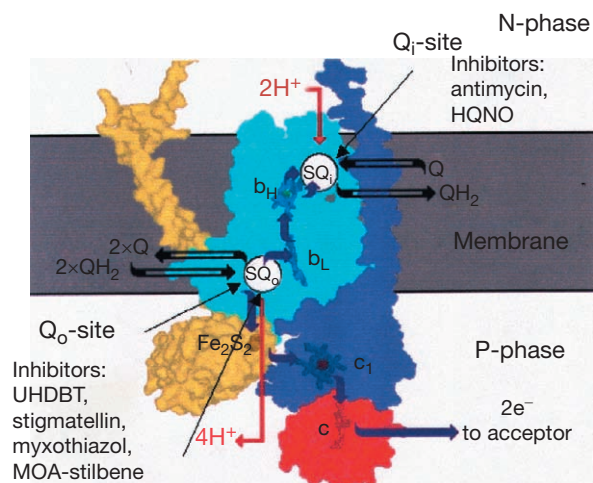


Figure 2 The Q-cycle as a mechanism of proton translocation in complex III of the mitochondrial respiratory chain. The scheme shows the Q-cycle in the context of the enzyme structure where the catalytic subunits are shown by their surfaces and made transparent so as to show the redox centers (cyan: cytochrome *b*; yellow: iron-sulfur protein; blue: cytochrome *c*₁). Quinol (QH₂) is oxidized in a bifurcated reaction at the Q₀-site of the complex located at the outer positive side of the membrane (P-phase): one electron is transferred to the high-potential chain, consisting of the Fe₂S₂ center and cytochrome *c*₁, and then to cytochrome *c* (red) that carries the electron to complex IV. As the semiquinone formed (SQ₀) is unstable, another electron is transferred through the low-potential chain of cytochrome hemes (*b*_L and *b*_H) to the Q₁-site where a quinone molecule is fully reduced to quinol. To provide the two electrons required at the Q₁-site, located at the inner negative side of the membrane (N-phase), the Q₀-site oxidizes two equivalents of quinol (2 × QH₂) in successive turnovers. Information about inhibitor specificity of the above-mentioned reactions is shown in the figure. The integration of the redox reactions with the release or uptake of protons in the aqueous phases (red arrows) allows the complex to pump protons across the membrane. Figure courtesy: Dr. A. R. Crofts, University of Illinois at Urbana-Champaign, IL, USA.

by freeze-fracture electron microscopy and the kinetic demonstration that coenzyme Q (as well as cytochrome *c*) behaves as a homogeneous pool. However, more recent investigations by native gel electrophoresis have shown the existence of supra-molecular associations of the respiratory complexes (super-complexes), confirmed by electron microscopy analysis and single-particle image processing. Flux control analysis has demonstrated that complexes I and III in mammalian mitochondria and complexes I, III, and IV in plant mitochondria kinetically behave as single units, suggesting the existence of substrate channeling within the supercomplexes.

Besides coenzyme Q dissolved in the lipid bilayer or in lipid domains within supercomplexes, coenzyme Q molecules bound to the individual complexes have also been described. These bound quinones may be important for the mechanism of outward proton translocation, required in the chemiosmotic mechanism of OXPHOS. The best-known mechanism is the Q-cycle, first proposed by the Nobel laureate Peter Mitchell (Figure 2).

Tight coenzyme Q binding to proteins (Q-binding proteins) in the respiratory complexes is required to stabilize the

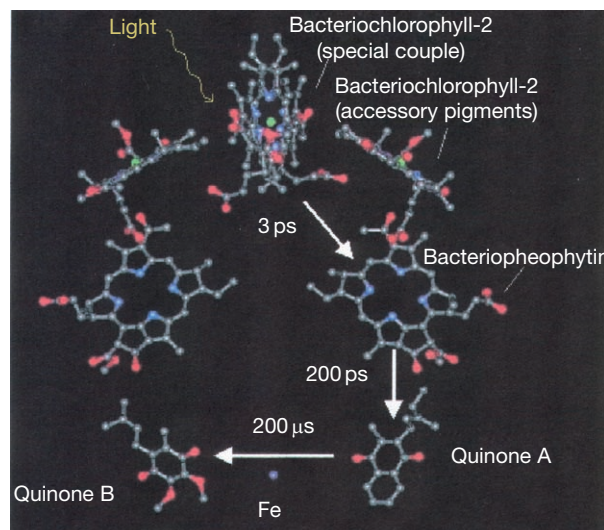


Figure 3 Bound quinones in the photosynthetic reaction center of *Rhodospseudomonas viridis*. Following excitation, the special pair of bacteriochlorophyll molecules donates an electron to a nearby bacteriopheophytin pigment (time constant: 3 ps). Subsequent transfer to the primary quinone (QA, a bound menaquinone) occurs with a half-time of 200 ps. Further electron transfer occurs from QA to the quinone acceptor QB, a bound ubiquinone occupying an active site near the cytoplasmic surface of the protein. QB accepts two electrons and two protons, then it undergoes exchange with the mobile pool of quinones in the lipid bilayer. With permission of Dr. R. Baxter, University of Chicago, IL, USA.

semiquinone form that is an essential intermediate in the mechanism of proton translocation.

Moreover, a well-known example of bound quinones is represented by the reaction center of photosynthetic bacteria that has been crystallized and resolved by X-ray diffraction together with its bound prosthetic groups (Figure 3). In bacteria, besides ubiquinone, menaquinone (a naphthoquinone) also may participate in electron transfer.

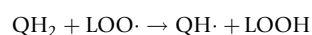
In chloroplasts of photosynthetic algae and higher plants, the function of ubiquinone is exploited by plastoquinone, present in bound form in photosystem II and free as a pool in the lipid bilayer.

Other electron transfer chains

Coenzyme Q of eukaryotic cells is present in most intracellular membranes and in the plasma membrane as well. Its function in these systems is better known in the plasma membrane oxidoreductase where the quinone appears to participate in transmembrane electron transfer between intracellular NADH or NADPH and extracellular acceptors including the ascorbic acid free radical, derived from the partial oxidation of ascorbic acid. This system may be involved in the antioxidant defenses of the cell because it regenerates ascorbic acid, an antioxidant.

Coenzyme Q as an antioxidant

Coenzyme Q and other quinones are powerful antioxidants in their reduced form. The ubiquinol reacts with peroxide radicals, such as the lipid peroxide radical LOO·:



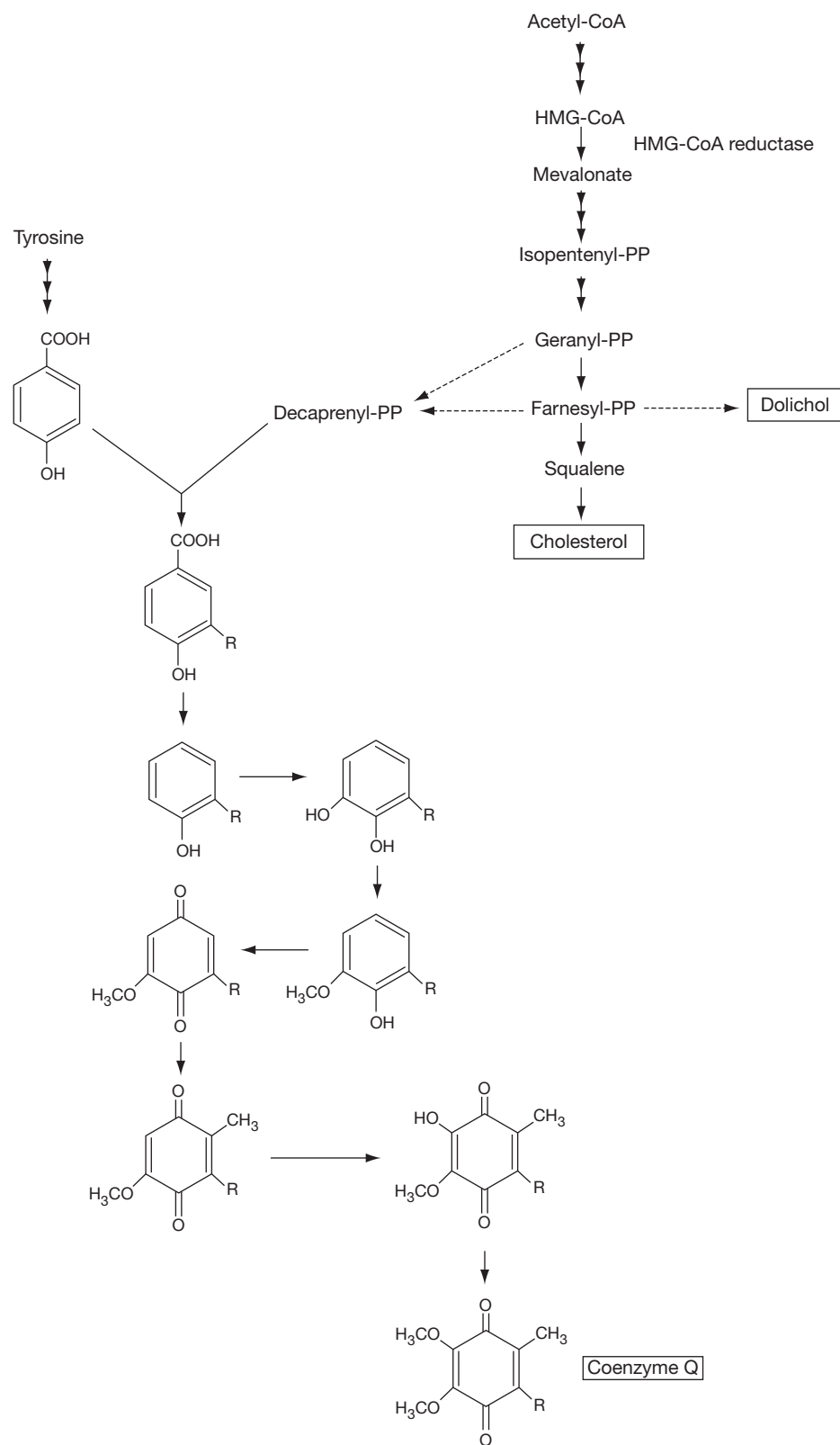
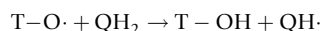
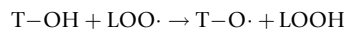


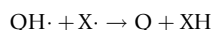
Figure 4 Simplified scheme of the biosynthesis of coenzyme Q. R, isoprenoid side chain; HMG-CoA, 3-hydroxy-3-methyl-glutaryl coenzyme A.

where LOOH is a lipid hydroperoxide.

In addition, it may be used to regenerate the reduced antioxidant form of vitamin E (α -tocopherol, T-OH) from the tocopheroxyl radical TO $\cdot$:



The semiquinone form of coenzyme Q, originated from its antioxidant action, is readily converted to the oxidized form by another radical:



The cell contains several enzymatic systems to reduce oxidized coenzyme Q (Q) back to its antioxidant form (QH $_2$). Among these are the cytosolic enzyme DT-diaphorase (NAD(P)H dehydrogenase (quinone) E.C.1.6.99.2; DT = formerly DPNH/TPNH, now NADH/NADPH), also found bound to the plasma membrane, a cytosolic NADPH quinone reductase, and microsomal NADH cytochrome b_5 reductase.

Biosynthesis of coenzyme Q

Coenzyme Q is synthesized by animals starting from the amino acid tyrosine which is converted to *p*-hydroxybenzoate; after insertion of the isoprenoid side chain in the 6-position, the ring is completed by a complex series of reactions involving hydroxylations and O-methylations. The isoprenoid side chain is synthesized starting from acetyl CoA through the mevalonate pathway and isopentenyl pyrophosphate, which is common to the synthesis of cholesterol, dolichol, and other polyprenyl compounds (Figure 4). It was found that cholesterol-lowering agents acting on the enzyme 3-hydroxy-3-methyl-glutaryl CoA reductase, which converts 3-hydroxy-3-methyl-glutaryl CoA to mevalonate, besides lowering plasma cholesterol, also lower coenzyme Q levels, suggesting caution in the use of these agents.

A complex of at least 10 proteins encoded by nuclear (COQ) genes synthesizes CoQ but its regulation is not clear. Since 1989, a growing number of patients with multisystemic mitochondrial disorders and neuromuscular disorders showing deficiencies of CoQ have been identified. CoQ deficiency caused by mutation(s) in any of the COQ genes is designated primary deficiency. Other patients have displayed other genetic defects

independent on the CoQ biosynthesis pathway, and are considered to have secondary deficiencies.

The biosynthesis of coenzyme Q also requires several vitamin-derived coenzymes and appears to be easily limiting on the normal levels of the quinone in the body. For this reason, dietary intake of coenzyme Q, as a vitamin-like supplement, has been recommended by many authors.

Biological *o*-Benzoquinones Derived from Catecholamines

Some *o*-quinones may be physiologically formed from catecholamine derivatives. Adrenaline may undergo oxidation and cyclization to adrenochrome in a multistep process, in which the main oxidant under physiological conditions is the superoxide anion (Figure 5); on the other hand, adrenochrome can be reduced to the corresponding semiquinone by NADPH-cytochrome P450 reductase in liver microsomes and by mitochondrial complex I in bovine heart; a redox cycle is then established in which the semiquinone reacts with O $_2$, producing superoxide and regenerating adrenochrome. Similar reactions occur with dopamine and dopaminochrome. Substantial evidence exists that aminochromes and their chemical derivatives initiate deleterious effects at the cellular level that affect the supply of energy for the maintenance of vital cellular processes, particularly in the heart.

Biological Naphthoquinones: Menaquinone and Vitamin K

Menaquinone is the quinone used for electron transfer by many bacteria, either alone or together with ubiquinone (Figure 3).

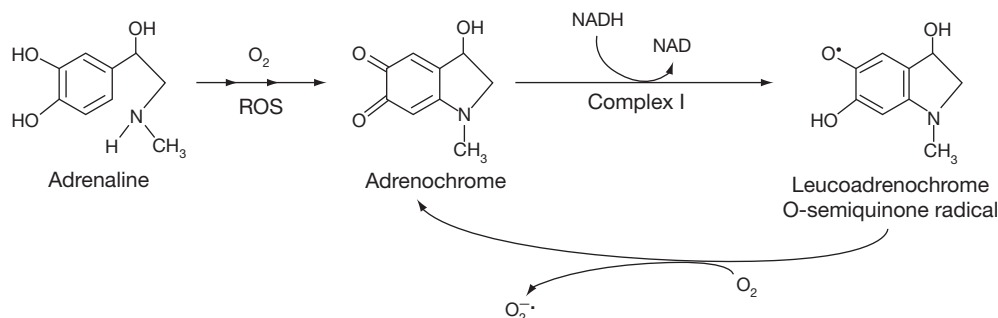
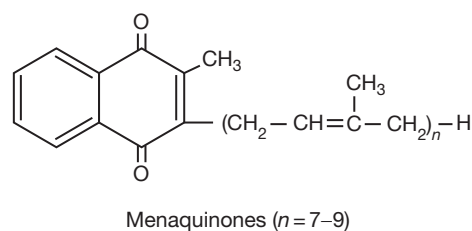


Figure 5 Redox cycling of catecholamines: the example of adrenaline is shown in the figure. Adrenaline is first oxidized to adrenochrome that is reduced by complex I by one electron transfer to the semiquinone that, on the other hand, auto-oxidizes generating superoxide and perpetuating the cycle.

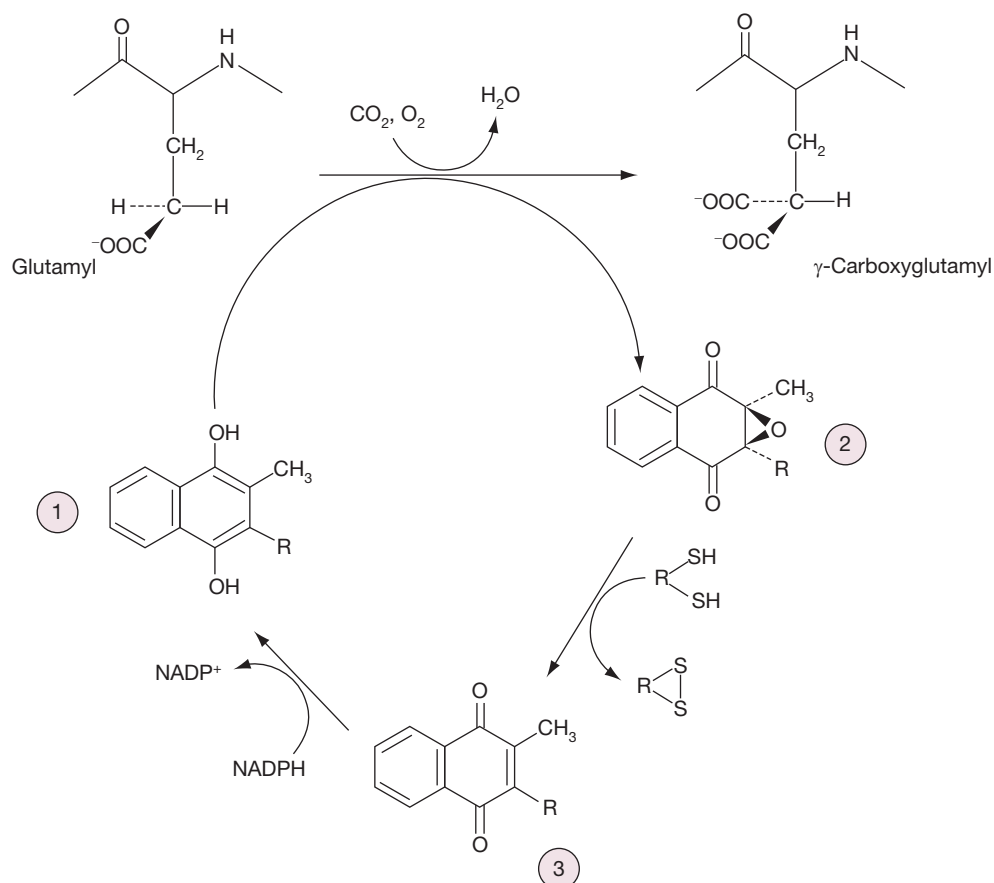
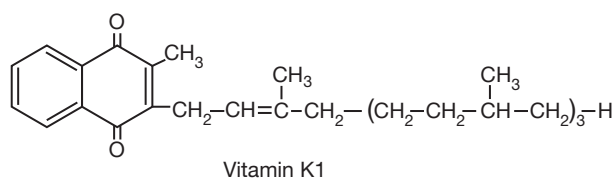


Figure 6 Mechanism of action of vitamin K in protein carboxylation. The γ -carboxylation reaction of protein-bound glutamate requires molecular oxygen and the reduced form of vitamin K (1); the hydroquinone is converted to the vitamin K epoxide (2) during the reaction. The vitamin K epoxide must be reconverted to the hydroquinone; this happens in two steps: in the first step, the epoxide is converted to the quinone form (3) using the coenzyme thioredoxin $[R(SH)_2]$; the quinone is then reduced to the hydroquinone in an NADPH-dependent reaction.

Vitamin K exists in two natural forms, phyloquinone (K1) and menaquinone (K2), with 4 and 6 isoprenoid units, respectively. In phyloquinone, three isoprenoid units are hydrogenated. Menadione (K3), without an isoprenoid chain, is a synthetic derivative.



Vitamin K is involved in the process of blood coagulation, being required for the synthesis of prothrombin and factors VII, IX, and X. Vitamin K is involved in posttranslational modifications of these proteins consisting in the γ -carboxylation of glutamate residues. Vitamin K is the coenzyme of the carboxylase enzyme and acts in its reduced hydroquinone form. The mechanism of action is shown in [Figure 6](#).

See also: **Bioenergetics:** Cytochrome bc_1 Complex (Respiratory Chain Complex III); Mitochondrial Permeability Transition Pore; Purple Bacteria: Electron Acceptors and Donors; Purple Bacteria: Photosynthetic Reaction Centers; Reactive Oxygen and Nitrogen Species: Interactions with Mitochondria and Pathophysiology; Respiratory Chain and Adenosine Triphosphate Synthase; Respiratory Chain Complex I; Respiratory Chain Complex II and Succinate: Quinone Oxidoreductases; Structure of Respiratory Complex I; Structure–Function of the Cytochrome b_6f Complex of Oxygenic Photosynthesis; **Metabolism Vitamins and Hormones:** Vitamin K: Biochemistry, Metabolism, and Nutritional Aspects.

Further Reading

- Behonick GS, Novak MJ, Neally EW, and Baskin SI (2001) Toxicology update: The cardiotoxicity of the oxidative stress metabolites of catecholamines (aminochromes). *Journal of Applied Toxicology* 21: s15–s22.
- Ebadi M, Marwath J, and Chopra R (eds.) (2001) *Mitochondrial Ubiquinone (Coenzyme Q₁₀): Biochemical, Functional, Medical, and Therapeutic*

- Aspects in Human Health and Diseases*, 2 vols. Scottsdale, AZ: Prominent Press.
- Haas RH (ed.) (2007) *Special Issue: The Role of Coenzyme Q in Cellular Metabolism: Current Biological and Clinical Aspects*, *Mitochondrion* 7(Supplement 1): S1–S180.
- Kagan VE and Quinn PJ (eds.) (2001) *Coenzyme Q: Molecular Mechanisms in Health and Disease*. Boca Raton, FL: CRC Press.
- Lenaz G (ed.) (1985) *Coenzyme Q: Biochemistry, Bioenergetics and Clinical Applications of Ubiquinone*. London: Wiley.
- Lenaz G and Genova ML (2009) Mobility and function of coenzyme Q (ubiquinone) in the mitochondrial respiratory chain. Review. *Biochimica et Biophysica Acta* 1787: 563–573.
- Lenaz G and Genova ML (2010) Structure and organization of mitochondrial respiratory complexes: A new understanding of an old subject. Review. *Antioxidants and Redox Signaling* 12: 961–1008.
- Littarru G (ed.) (2008) *Special Issue dedicated to the Fifth Conference of the International Coenzyme Q₁₀ Association – 50th Anniversary of CoQ₁₀ Discovery*, *BioFactors* 32(1–4): 1–275.
- Salviati AR, Hirano JS, and Navas P (2009) Coenzyme Q₁₀ deficiencies in neuromuscular diseases. Review. *Advances in Experimental Medicine and Biology* 652: 117–128.
- Many authors (2007) Special issue about biological and clinical aspects of the role of coenzyme Q in cellular metabolism. *Mitochondrion* 7s.
- Many authors (2008) Special Issue dedicated to the Fifth Conference of the International Coenzyme Q₁₀ Association – 50th anniversary of CoQ₁₀ discovery. *BioFactors* 32.

Relevant Website

<http://www.icqa.org> – International Coenzyme Q10 Association.

ENCYCLOPEDIA OF **BIOLOGICAL CHEMISTRY**

SECOND EDITION

Volume 4

R-Z

R

Rab Family

M W McCaffrey and A J Lindsay, University College Cork, Cork, Republic of Ireland

Published by Elsevier Inc.

Glossary

Effectors Proteins which are necessary to effect small guanine triphosphatase (GTPase) function. Many (putative) effectors of the ras superfamily have been described in the past decade.

Guanine nucleotide diphosphate (GDP) As for GTP but lacking the terminal (high energy phosphate bond/group).

Guanine nucleotide triphosphate (GTP) A compound – termed a purine – that comprises of two (carbon and nitrogen containing) cyclic rings, a five carbon sugar, and three phosphate groups.

GTPases Proteins capable of binding and hydrolyzing GTP by cleavage of its terminal phosphate. This reaction results in the protein being switched to the inactive state, that is, associated with GDP.

Ras Monomeric GTPases, first identified as oncogenes/oncoproteins, which function as central players in mitogenic signaling pathways.

Rho The first ras homolog to be identified, rho proteins control actin cytoskeleton organization and structure.

Discovery

The first Rab guanine triphosphatases (GTPases) were discovered in the late 1980s, through the pioneering work of the Tavittian group, in Paris, which set out to identify homologs of the ras oncogenes. They screened a brain complementary DNA (cDNA) library with degenerate oligonucleotides designed against the most conserved regions of G proteins (the G boxes), as well as known G proteins, new ras-related genes were identified in this screen. These were sufficiently different from ras and rho to be placed in a new family of the ras superfamily and were named the Rabs–Ras homolog from brain.

Sequencing revealed that these new Rab genes bore greatest similarity to two previously described yeast secretory mutant genes, *sec4* and *ypt1*. The fact that these two yeast orthologs had previously been implicated in membrane trafficking provided the first clues that the members of the Rab GTPase family function in membrane trafficking in higher eukaryotes. In the years that followed, several research groups (notably the Zerial group) extended the family to 11 then to between 30 and 40 distinct members. With the aid of information from the human genome project, up to 70 distinct mammalian Rab GTPases have now been identified.

Primary Structure and Evolution

The Rab family are the most divergent and extensive branch of the ras superfamily. They contain approximately 200 amino acids with a molecular mass between 20 and 25 kDa, and are

hydrophilic in nature (see [Figure 1](#) for a schematic representation of Rab functional domains). Like all G proteins, they contain well-conserved guanine nucleotide-binding domains (G boxes) but otherwise are quite divergent – particularly in their carboxy-terminal region. This region is known to be critical for specifying membrane localization.

Their extreme carboxy-termini also possess a characteristic conserved CC or CXC motif (C=cysteine and X=any amino acid). This motif is posttranslationally modified, in a process known as ‘isoprenylation’, by the addition of geranyl groups. The addition of a hydrophobic moiety on the cysteine results in a protein that is chemically competent for insertion into the lipid bilayer (see later for further details on posttranslational modifications).

By analogy with the Ras proteins, a further domain has been identified in the Rabs – the effector domain. This domain of approximately 10 amino acids was found to be critical for transmission of the Ras signal to its downstream machinery. Mutation of this effector domain abolishes growth signaling, independent of Ras GTP binding. Thus, this region was named the effector domain of ras. The equivalent region in the Rabs, also termed the ‘effector’ domain, is frequently mutated in studies that investigate Rab function.

By analogy with the oncogenic Ras mutants, a number of Rab mutations have been generated. The most frequently utilized activating mutant for the Rabs is at amino acid position approximately 70 and is generated by the conversion of a glutamine (Q) to a leucine (L) at this position. Such Rab mutants are deficient in GTPase activity and thus remain in the activated GTP-associated state (see [Figure 2](#) for GTP structure). A serine

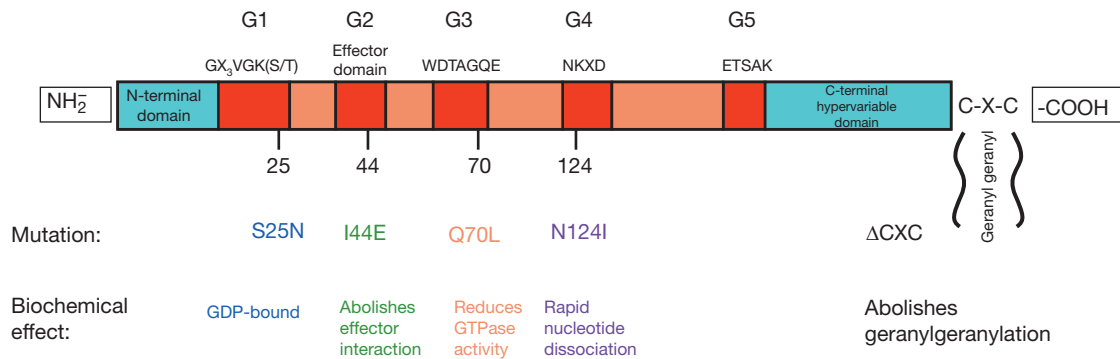


Figure 1 Schematic diagram of the general domain structure of Rab GTPases. Indicated are the commonly used Rab11 mutations and their biochemical consequences.

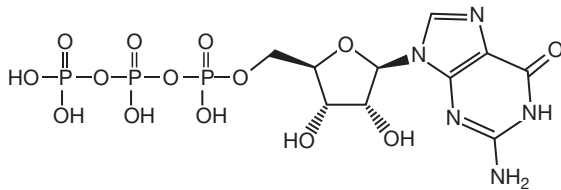


Figure 2 The structure of GTP.

(S) to asparagine (N) mutation, near amino acid 20, generates a dominant negative Rab mutant that is constitutively GDP bound and therefore inactive. Expression of these mutants in mammalian cells has revealed important Rab functional data.

In a study of the evolution and complexity of the core protein machinery involved in intracellular trafficking (coat complexes, Rab GTPases, SNAREs, and Sec1s) by the Scheller group, 60 distinct Rab genes were identified in humans. It was noted that there was a significant expansion of the Rab family from yeast to worms, and a further expansion from worms to mammals. A phylogenetic tree of the human Rab family is shown in [Figure 3](#). Of the four core membrane trafficking families compared, the Rabs were the only family that underwent such a significant expansion. It is likely that the higher degree of complexity in multicellular organisms is reflected by the greater requirement for Rabs, not only for more intricate regulation of intracellular trafficking but also for tissue specialization of membrane trafficking functions.

The Rab family can be divided into ~10 subfamilies. Representative members of these subfamilies are Rab1, Rab3, Rab4, Rab5, Rab6, Rab8, Rab11, Rab22, Rab27, and Rab40.

Posttranslational Modifications

Rab proteins are hydrophilic; thus, in order for membrane binding they must possess one, or two, C20 geranyl-geranyl lipid groups at their carboxy-terminus. The process by which they gain these lipid moieties is known as 'isoprenylation'. Newly synthesized, unprenylated, Rab GTPases are bound by Rab escort protein (REP) and delivered to geranylgeranyl transferase (GGTase). GGTase catalyzes the attachment of the lipid groups to two carboxy-terminal cysteine residues present on all

Rab GTPases. REP appears to be a specialized Rab GDI that can deliver newly synthesized, GDP-bound, Rab proteins to GGTase for isoprenylation, and then transport the modified protein to its target membrane. It is thought that Rab-GDP binds its host membrane via a protein receptor, which senses the Rab carboxy-terminal domain; however, very little evidence currently exists in the literature regarding the identity of such Rab receptors. Thus, while the amino acid sequence in the C-terminal region confers membrane specificity to a given Rab, the isoprenoid group confers chemical compatibility with the membrane, anchoring the otherwise hydrophilic protein in the hydrophobic lipid bilayer.

Cellular Localization

Rab proteins have now become standard markers for organelles of the secretory and endocytic pathways. Not all Rab GTPases have yet been sufficiently characterized such that the entire Rab localization pattern is known. However, for those that have been characterized, it is clear that they localize to distinct membrane organelles. For example, Rab1 localizes to the endoplasmic reticulum (ER)/cis-Golgi and Rab3 localizes to synaptic vesicles/chromaffin granules of the neuroendocrine system, while Rab5, Rab4, and Rab11 localize to distinct early endosomal membranes, Rab6 is located on the Golgi apparatus, Rab7 is on late endosomes, Rab22 localizes to endosomes and the Golgi, and Rab27 to melanosomes and lytic granules. [Figure 4](#) and [Table 1](#) indicate the known localization and trafficking steps controlled by Rabs characterized to date.

Functional Cycle

The functional cycle of Rab5 has been extensively characterized and is used here as a paradigm of all Rabs (see [Figure 5](#) for generalized Rab GTPase cycle). GDP-bound Rab5 is delivered to its target membrane by Rab guanine nucleotide dissociation inhibitor (Rab GDI). Rab guanine nucleotide exchange factor (Rab GEF) then catalyzes the exchange of the GDP for GTP, allowing Rab5 to adopt its active conformation. The GTP-bound Rab is now resistant to extraction by Rab GDI. Once membrane associated and in its active conformation, the Rab

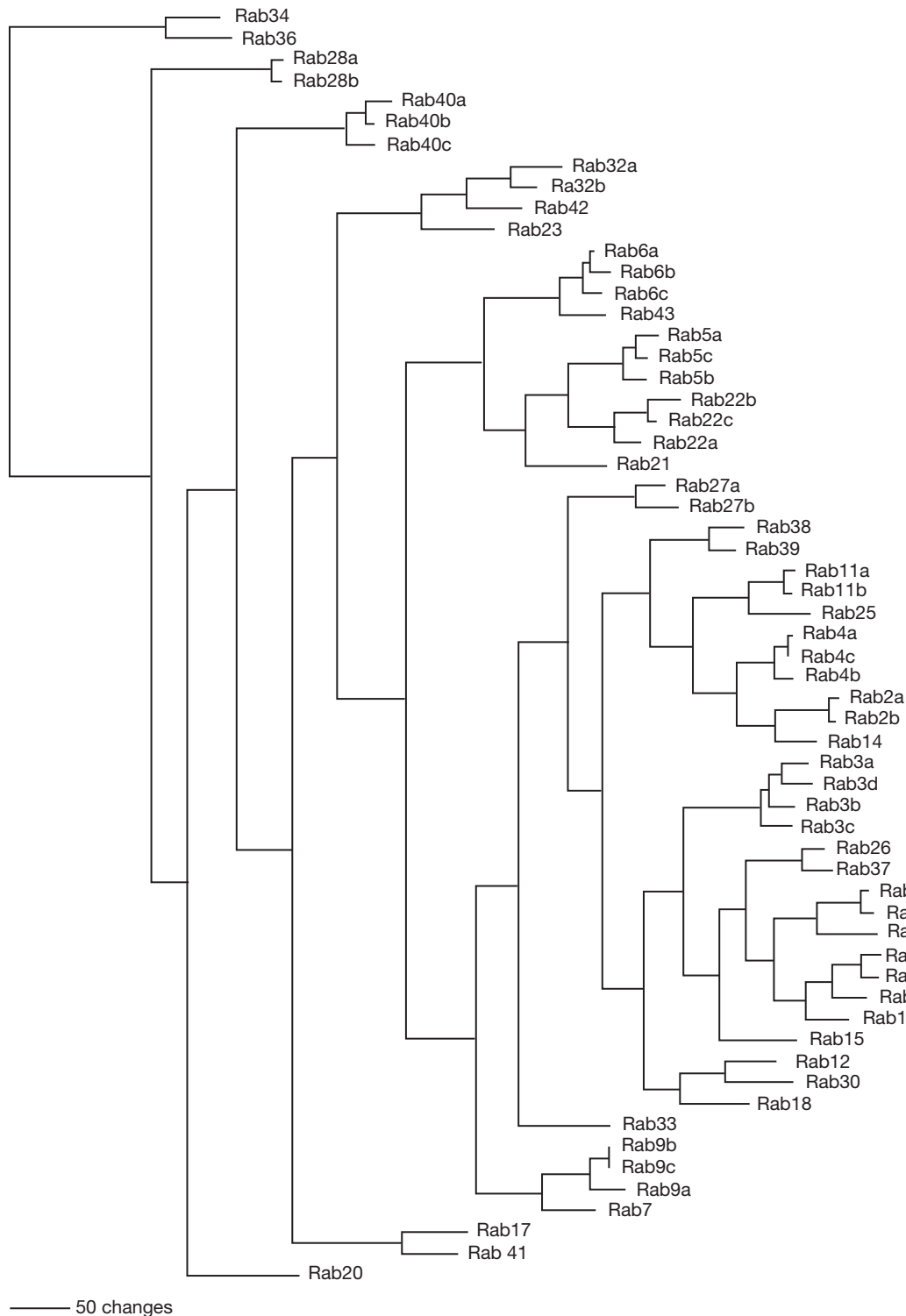


Figure 3 Phylogenetic tree of the human Rab GTPases. From Bock JB, Matern HT, Peden AA, and Scheller RH (2001) A genomic perspective on membrane compartment organization. *Nature* 409: 839–841.

protein is now competent to exert a controlling function on membrane trafficking. Rab5 mediates its function by recruiting a repertoire of specific effector proteins necessary for fusion of early endosomes. Upon completion of Rab function, the GTP

is hydrolyzed by the Rab GTPase enzymatic activity, releasing inorganic phosphate and resulting in an inactive Rab bound to GDP. This GTPase activity is stimulated by Rab GTPase activating protein (Rab GAP).

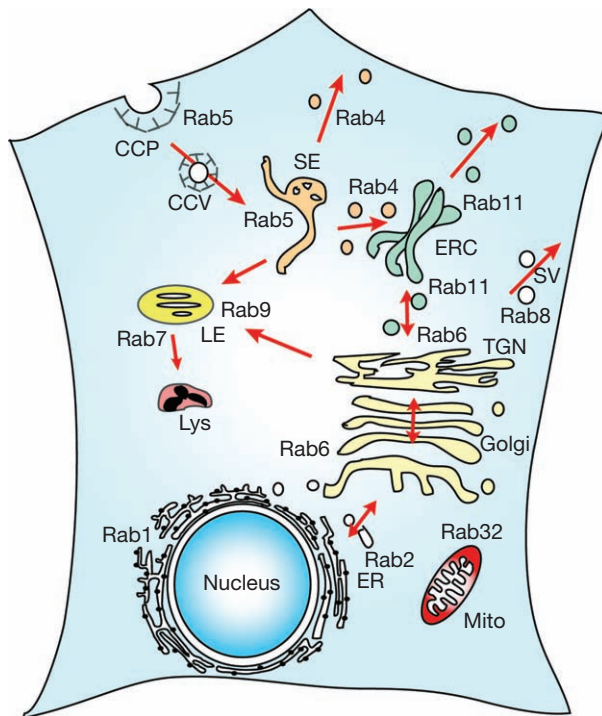


Figure 4 Localization of selected Rab GTPases and the membrane trafficking pathways they regulate. The endocytic pathway involves internalization of cargo in clathrin-coated vesicles (CCV) and delivery to the early/sorting endosome (SE). Cargo is then either sorted for transport along the degradative pathway to the late endosomes (LE) and lysosomes (Lys), or directly recycled back to the plasma membrane from the SE, or indirectly, via the endocytic recycling compartment (ERC). The biosynthetic pathway transports proteins from the endoplasmic reticulum (ER) through the Golgi complex from where they are delivered to the cell surface in secretory vesicles (SV). The endocytic and secretory pathways communicate through the *trans*-Golgi network (TGN).

Rabs have been variously implicated in vesicle budding from the donor membrane, transport along the cytoskeleton, docking to, and fusion with, the acceptor membrane. Though there is evidence for various Rabs of their involvement in these distinct events in membrane trafficking, it is not clear whether some Rabs function in vesicle budding exclusively, while others function in vesicle transport or docking, or whether a given Rab, in association with one specific set of effector proteins, functions in budding and in combination with other sets of effector proteins functions in transport, docking, or fusion.

Effectors

Several putative Rab effectors have been identified – primarily by two-hybrid library screening with GTPase deficient (dominant-positive) Rab mutants. While a survey of all of the Rab effectors reported in the literature is beyond the scope of this article, it is clear that there are likely to be many more Rab effectors than Rab GTPases. The GTP-bound form of Rab5 has been shown to interact specifically with at least 20 different proteins. Several Rab effectors have coiled-coil and/or phospholipid-binding domains (such as FYVE or C2 domains), while some are enzymes, or cytoskeletal (actin or tubulin) motor proteins. Identification of Rab effectors, and regulatory proteins, is a very important approach toward understanding the function of Rab GTPases and is likely to be an area of continued research for several years to come.

Rabs and Diseases

It is well appreciated that a substantial number of disease-causing genes are involved in membrane trafficking. A growing proportion of these genetic diseases are being attributed to mutations of Rab GTPases, or their interacting proteins.

Table 1 Localization, tissue distribution, and function of the human Rab GTPases

Rab GTPase	Localization	Tissue expression	Function
Rab34	Golgi	Ubiquitous	Secretory pathway
Rab36	Golgi	Unknown	Lysosome positioning
Rab28	Endosomes	Ubiquitous	Degradative pathway
Rab40	Golgi?	Unknown	Unknown
Rab32	Melanosomes	Spleen, mast cells, macrophages, bone marrow dendritic cells, melanocytes	Melanosome transport in melanocytes. Autophagic vacuole formation.
Rab42	Unknown	Unknown	Unknown
Rab23	Cilium	Widely expressed	Negative regulator of Sonic hedgehog pathway in mouse. Ciliogenesis
Rab6	Golgi	Ubiquitous	Retrograde Golgi transport
Rab43	Golgi and ER	Widely expressed	ER to Golgi transport
Rab5	Plasma membrane, CCVs, and early endosomes (EE)	Ubiquitous	Transport from the plasma membrane to EE
Rab22	EE, LE, Golgi	Ubiquitous	Communication between biosynthetic and endocytic pathways
Rab21	EE	Ubiquitous	Endocytosis, cell division
Rab27	Melanosomes and granules	Melanocytes, platelets, and lymphocytes	Transport of melanosomes and lytic granules
Rab38	Melanosomes	Melanocytes, lung, mast cells, RPE	Melanosome sorting, lung surfactant secretion

(Continued)

Table 1 (Continued)

<i>Rab GTPase</i>	<i>Localization</i>	<i>Tissue expression</i>	<i>Function</i>
Rab39	Golgi-associated organelles	Ubiquitous	Endocytosis
Rab11/Rab25	ERC and TGN	Ubiquitous	Endocytic recycling and transport to TGN
Rab4	EE	Ubiquitous	Endocytic recycling and degradation
Rab2	Golgi	Ubiquitous	Golgi to ER transport
Rab14	TGN and endosomes	Ubiquitous	Biosynthetic pathway
Rab3	Synaptic vesicles (SV)	Neurons	Regulation of neurotransmitter release
Rab26	Secretory granules	Parotid gland	Regulated secretory pathway
Rab37	Secretory granules	Mast cells	Mast cell degranulation
Rab1	ER/ <i>cis</i> -Golgi	Ubiquitous	ER to Golgi transport
Rab35	Plasma membrane, endosomes	Ubiquitous	Actin bundling, cytokinesis
Rab8	Endosomes, primary cilium	Ubiquitous	Regulates polarized membrane transport
Rab10	Endosomes, tubules	Ubiquitous	Phagosome maturation, GLUT4 transport
Rab13	TGN, recycling compartment	Epithelial cells	Transport of junctional proteins
Rab15	EE and RC	Ubiquitous	Endocytic recycling
Rab12	Small vesicles	Widely expressed	Degradative pathway
Rab30	Golgi	Widely expressed	Golgi structure
Rab18	ER and Golgi	Epithelial cells	ER to Golgi transport
Rab33	Golgi	Brain, lymphocytes, melanocytes	Autophagosome formation (Rab33B)
Rab9	LE	Ubiquitous	LE to TGN transport
Rab7	Late endosomes (LE)	Ubiquitous	Transport from LE to lysosomes
Rab17	Epithelial cells	Apical RC	Transport through apical RC
Rab41	Unknown	Widely expressed	Unknown
Rab20	Apical dense tubules in kidney tubule epithelial cells	Epithelial cells	Connexin 43 transport

RPE, retinal pigment epithelium.

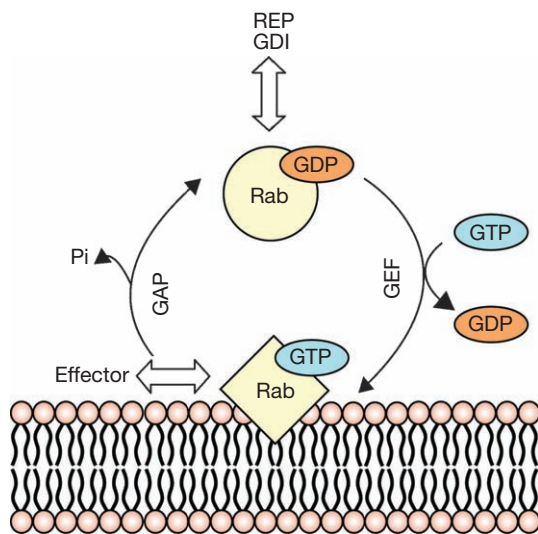


Figure 5 The Rab GTPase cycle. Rab GTPase cycle between GTP- and GDP-bound forms, allowing the Rab to adopt active and inactive conformations. The cycle is mediated by various accessory factors. A guanine-nucleotide exchange factor (GEF) exchanges the GDP for GTP. Conversion from the GTP- to the GDP-bound form is mediated by a GTPase-activating protein (GAP) which stimulates the Rab to hydrolyze bound GTP. The active GTP-bound form mediates its function through specific effector molecules. Newly synthesized cytosolic, GDP-bound Rab, interacts with Rab escort protein (REP) for membrane delivery. Subsequently, after each functional round, the GDP-bound Rab is extracted by the GDP dissociation inhibitor (GDI) from its host membrane and delivered to a new membrane location for another functional round.

The first example of Rab involvement in disease was in Griscelli syndrome – an autosomal recessive disorder that results in pigmentation defects. There are two variants of this disease, one that results in immunological defects, and the other is associated with neurological dysfunction. The former is caused by missense mutations in the Rab27a gene, and the latter is the result of mutations in its putative effector, myosin Va. Both genes are located side by side on chromosome 15. Rab27a functions to regulate the transport of melanosomes to the periphery of melanocytes, it also regulates the secretion of lytic granules in cytotoxic T lymphocytes. Hence, the lack of functional Rab27a results in the loss of pigmentation and in the uncontrolled activation of T lymphocytes. Mutation of myosin Va causes pigmentation defects, but since it is not involved in lytic granule movement does not cause immunological defects.

Choroideremia is an X-linked disease that results in the degeneration of the pigment epithelium of the eye. It is caused by a defect in the REP-1 gene. REP-1 is one of the two isoforms of GGTase that prenylate Rab GTPases. It is essential for the prenylation of Rab27a in the retinal pigment epithelium. REP-2 appears to be sufficient for the efficient geranyl-geranylation of all other Rab GTPases. Thus, the lack of functional REP-1 results in nonfunctional Rab27a and the onset of the disease. The pathology of this disease may be caused by the degeneration of the retinal epithelium due to deficient melanosome transport, and hence lack of protection during light exposure.

Tuberin is a tumor-suppressor gene that interacts with Rabaptin-5 and negatively regulates endocytosis by acting as a Rab5-GAP. Its inactivation causes tuberous sclerosis, a disease that results in malformations and tumors of the central

nervous system. Furthermore, mutations in the GDI1 gene, which encodes GDI- α , have been found in a subgroup of patients with X-linked nonspecific mental retardation. GDI- α is particularly abundant in brain tissue, thus defective membrane recycling of one or more Rab GTPases in brain synapses is likely to be the cause of this condition.

In recent years, the involvement of the Rab11/25 subfamily and their effector proteins in cancer and Rab7 in Charcot-Marie-Tooth neuropathies have been described. Specifically, Rab11a has been shown to be involved in low oxygen (hypoxia) stimulated invasion of breast cancer cells. Independently, Rab11c(Rab25) has been shown to determine the aggressiveness of breast and ovarian cancers and its increased expression correlates with poor prognosis. The major underlying basis for the role of Rab25 in cancer aggressiveness appears to be its ability to drive the recycling of the $\alpha 5\beta 1$ integrin to the plasma membrane resulting in increased random three-dimensional (3D) cell motility. Supportive of an underlying role of Rab11 GTPases in epithelial cancer aggressiveness is the fact that two Rab11 effectors – the Rab coupling protein (RCP) and Rab11-FIP3 are separately implicated in human cancers and additionally that RCP drives the recycling of the $\alpha 5\beta 1$ integrin and the epidermal growth factor receptor (EGF-R) to the plasma membrane leading to increased 3D cell motility and growth factor signaling.

Finally, Rab7 has been implicated in a group of autosomal dominant axonal disorders – namely, Charcot-Marie-Tooth

type 2B disease (CMT-2B). This neuropathy is characterized by severe sensory loss often leading to amputations. Rab7, a GTPase which controls late endosomal trafficking, has been implicated in CMT-2B through the identification of a number of missense mutations which render it in the dominant-positive (GTP-bound) form. The molecular basis for the role of Rab7 in CMT-2B is likely to be increased degradation rates of cargo proteins, such as the neurotrophin receptor as a consequence of increased Rab7 activity.

See also: **Cell Architecture and Function:** Rho GTPases and Actin Cytoskeleton Dynamics; **Signaling:** Ras Family; Small GTPases.

Further Reading

- Aridor M and Hannan LA (2002) Traffic jams II: An update of diseases of intracellular transport. *Traffic* 3: 781–790.
- Bock JB, Matern HT, Peden AA, and Scheller RH (2001) A genomic perspective on membrane compartment organization. *Nature* 409: 839–841.
- Stenmark H and Gillooly DJ (2001) Intracellular trafficking and turnover of phosphatidylinositol 3-phosphate. *Seminars in Cell and Development Biology* 12: 193–199.
- Woodman P (1998) Vesicle transport: More work for the Rabs? *Current Biology* 8: R199–R201.
- Zerial M and McBride H (2001) Rab proteins as membrane organisers. *Nature Reviews Molecular Cell Biology* 2: 107–119.

Ran GTPase

M S Moore, Baylor College of Medicine, Houston, TX, USA

Published by Elsevier Inc.

This article is reproduced from the previous edition, volume 3, pp. 635–639, © 2004, Elsevier Inc.

Glossary

GTPase A protein that can bind and hydrolyze GTP.

Guanosine 5'-diphosphate (GDP) A nucleotide analogous to adenosine diphosphate (ADP) that is composed of guanine linked to ribose and two phosphates.

Guanosine 5'-triphosphate (GTP) An energy-rich nucleotide analogous to adenosine triphosphate (ATP) that is composed of guanine linked to ribose and three phosphates.

Interphase A period or stage between two successive mitotic divisions of a cell nucleus.

Mitosis The sequential differentiation and segregation of replicated chromosomes in a cell nucleus that precedes complete cell division.

Nuclear pore complex (NPC) Large organelles of the nuclear envelope through which passage of proteins and RNAs between the cytoplasm and nuclear interior occurs.

Ras related nuclear protein (Ran) is a 25-kDa member of the Ras superfamily of small guanosine triphosphatases (GTPases). To date, a Ran homolog has been found in every nucleated cell of every eukaryotic organism on Earth, and Ran is essential for the viability of all those cells. During interphase, Ran plays an essential role in driving and regulating the active transport of cargo between the nucleus and cytoplasm. Ran also has critical functions during mitosis when Ran is essential for mitotic spindle assembly and placement as well as re-assembly of the nuclear envelope.

Small Guanosine Triphosphate-Binding Proteins as Molecular Switches

Ran is called a GTPase because it binds and hydrolyzes guanosine triphosphate (GTP). All small GTPases of the Ras superfamily cycle between GTP- and guanosine diphosphate (GDP)-bound states, and undergo a marked conformational change upon conversion from one state to the other. Thus, GTPases in the GTP-bound form can bind proteins that the GDP-bound form cannot, and vice versa. This key property confers on small GTPases the ability to function as molecular switches that cycle between active and inactive states, the GTP-bound form generally being the active form.

Ran represents its own family of small GTP-binding proteins in the Ras superfamily, of which it is the only member. The other families include the Ras, Rab, Arf, Rac, and Rho families that regulate signal transduction, vesicular transport, cytoskeletal organization, and cell proliferation. Ran's role is to act as a switch to regulate nuclear transport during interphase and spindle, and nuclear envelope assembly during mitosis.

Nucleocytoplasmic Transport

A basic necessity of the eukaryotic cell is the ability to transport macromolecules bidirectionally between the nucleus and

cytoplasm. The traffic between these two compartments is both heavy and complex with multiple classes of cargo being actively transported in both directions (**Figure 1**). Nuclear proteins such as transcription factors and histones are, like all proteins, made in the cytoplasm and must cross the nuclear envelope into the nuclear interior to carry out their function. Conversely, ribosomal subunits, transfer RNAs (tRNAs), and messenger RNAs (mRNAs) produced inside the nucleus must cross the nuclear envelope in the opposite direction into the cytoplasm in order to be used for protein translation.

Nucleocytoplasmic traffic is restricted to nuclear pore complexes (NPCs), which are large proteinaceous structures that bridge the double membrane of the nuclear envelope to form aqueous channels linking the cytoplasm and nucleoplasm. The vertebrate NPC is composed of ~30 different proteins, and a HeLa cell contains ~3000 NPCs. At present, all NPCs appear to be functionally identical with the same NPC capable of carrying out both import and export virtually simultaneously. While molecules up to ~50 kDa can diffuse through the NPC, the majority of nuclear proteins (both larger and smaller than 50 kDa) as well as the other classes of transport cargo appear to undergo signal-mediated active transport rather than diffusion.

There are two types of nuclear transport signals, nuclear localization sequences (NLSs) that mediate nuclear import and nuclear export sequences (NESs) that mediate nuclear export. There are multiple types of NLSs and NESs found on different types of cargo, and, in general, each type of signal is recognized by a different nuclear receptor/carrier. Most nuclear carriers are members of the karyopherin- β superfamily and there are over 20 members of this family in human cells. During nuclear import, a soluble carrier binds the NLS-containing cargo in the cytoplasm, then moves the cargo through the NPC and releases it in the nuclear interior. Nuclear export works essentially the same way in reverse, with an NES-containing cargo being recognized by its specific carrier inside the nucleus, movement of the export complex through the NPC, and disassembly of the complex and release of the cargo upon reaching the cytoplasm.

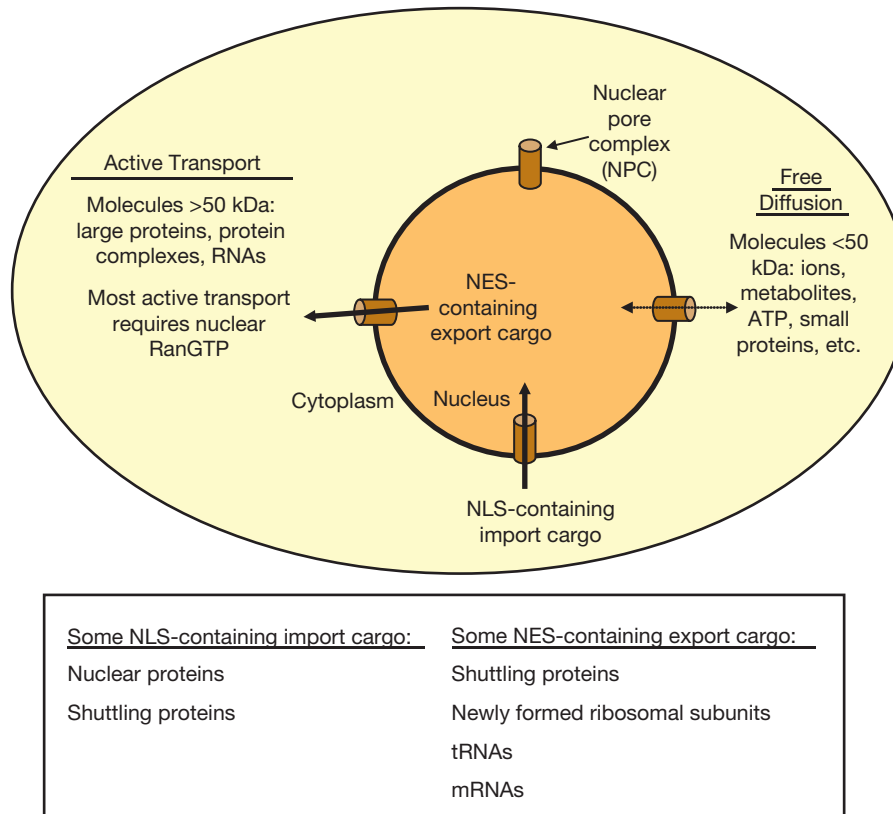


Figure 1 Overview of nucleocytoplasmic transport. Most nuclear transport requires nuclear RanGTP, however mRNA export does not. Carriers of the NTF2 family mediate mRNA export, and this transport pathway is Ran-independent. Active transport of the rest of the cargo listed does require nuclear RanGTP.

The Ran GTPase

The Cellular Gradient of RanGTP

A key player in the majority of the known nuclear transport pathways is the small GTPase Ran and, not incidentally, the main region of homology between all the karyopherin- β nuclear carriers is a RanGTP-binding domain. Inside the cell, Ran is primarily nuclear in spite of the fact it does not appear to contain an NLS in its sequence. This is because Ran is actively imported into the nucleus by a small protein called 'p10/NTF2', which is a member of a second class of nuclear carriers, the NTF2 family.

Importantly, whether Ran is in the cytoplasm or nucleus determines whether it is RanGTP or RanGDP (**Figure 2**). Inside the nucleus, Ran is converted to the GTP-bound form by the RanGEF (guanine nucleotide exchange factor), which stimulates RanGDP to release GDP and bind GTP. By contrast, the RanGAP (GTPase-stimulating protein) is located in the cytoplasm. The RanGAP rapidly converts any RanGTP entering the cytoplasm to RanGDP by stimulating Ran to hydrolyze its bound GTP.

The resulting concentration gradient of low RanGTP in the cytoplasm to high RanGTP in the nucleus is vitally important for controlling the directionality of nuclear transport by regulating the assembly and disassembly of karyopherin- β -containing transport complexes. Also, because the RanGEF is bound to chromatin, the concentration of RanGTP is highest near the chromatin where it is produced (**Figure 2(c)**). This

location becomes important during mitosis, when chromatin-localized RanGTP stimulates mitotic spindle and nuclear envelope assembly in the correct location around chromatin.

Cellular Functions of Ran

Nuclear Transport

The increasing RanGTP concentration from the cytoplasm to the nuclear interior is the primary factor that triggers the disassembly of transport complexes at their correct final destination. Members of the karyopherin- β family preferentially bind RanGTP with a much lower affinity for RanGDP and, upon binding RanGTP, the karyopherin- β carrier undergoes a conformational change that alters its cargo-binding domain. NLS receptors of the karyopherin- β family can only bind their cargo when they do not have bound RanGTP, which is their probable state in the cytoplasm. After movement across the NPC into the nuclear interior where the import complex encounters RanGTP, the carrier binds RanGTP and, as a result, releases its cargo into the nuclear interior (**Figure 3(a)**).

Export carriers of the karyopherin- β family behave just the opposite. These export carriers, in order to simultaneously bind their export cargo, must first bind RanGTP. After movement through the NPC, the cargo:export carrier:RanGTP complex encounters the RanGAP in the cytoplasm. Ran is stimulated to hydrolyze its bound GTP and, after hydrolysis, becomes RanGDP. The export carriers have a low affinity for

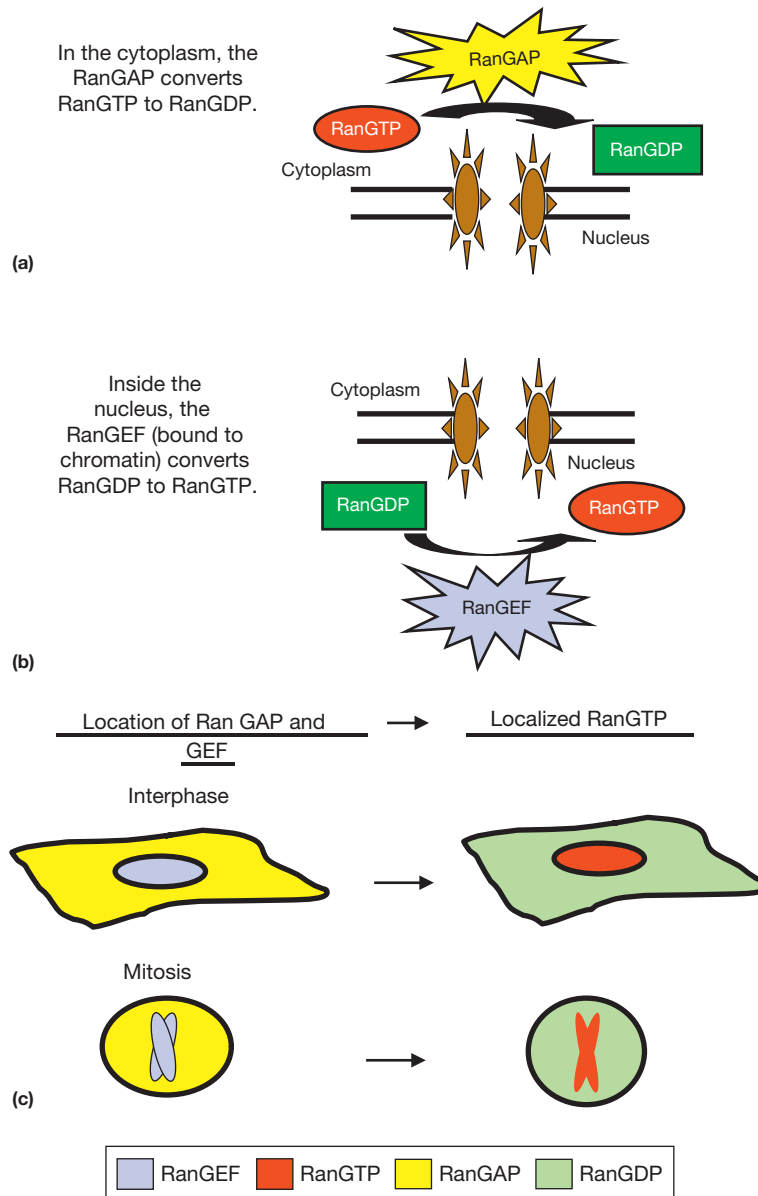


Figure 2 Establishment of the RanGTP gradient between the cytoplasm and nucleoplasm. (a) The RanGAP, which stimulates Ran to hydrolyze its bound GTP, is restricted to the cytoplasm. (b) The RanGEF, which stimulates Ran to release bound GDP and take up GTP, is found inside the nucleus bound to chromatin. (c) Restricting the RanGAP and the RanGEF to different compartments results in high levels of RanGTP inside the nucleus and low levels of RanGTP in the cytoplasm during interphase (top). The resulting gradient of RanGTP concentration between the cytoplasm and nucleoplasm is the key regulatory mechanism controlling the assembly and disassembly of nuclear import and export complexes in the correct cellular location (see Figure 3). Also, the RanGEF remains bound to chromatin throughout the cell cycle, and production of RanGTP at the chromatin surface in mitotic cells is required for proper placement of the mitotic spindle and re-assembly of the nuclear envelope around chromatin at the end of mitosis (bottom).

RanGDP and, as a result, both RanGDP and an export cargo dissociate from the export carrier in the cytoplasm.

Regulation of Spindle and Nuclear Envelope Assembly at Mitosis

The nuclear envelope vesiculates and breaks apart during mitosis in vertebrate cells, and then reforms after the daughter cells are divided. Accordingly, no nuclear transport occurs during this time because there is no nuclear envelope and the cytoplasmic

and nuclear contents are mixed. Ran, however, has additional functions critical to this period. Nuclear proteins containing an NLS are once again exposed to the NLS receptors and can be rebound by them. However, as during nuclear import, RanGTP can break this complex apart, releasing the proteins.

Many of the proteins responsible for assembly of the mitotic spindle at mitosis are nuclear during interphase and contain an NLS. To be active for spindle assembly, RanGTP must first release these spindle assembly factors from NLS receptors. Because the concentration of RanGTP is highest, close to the chromatin due to the RanGEF located there; active spindle

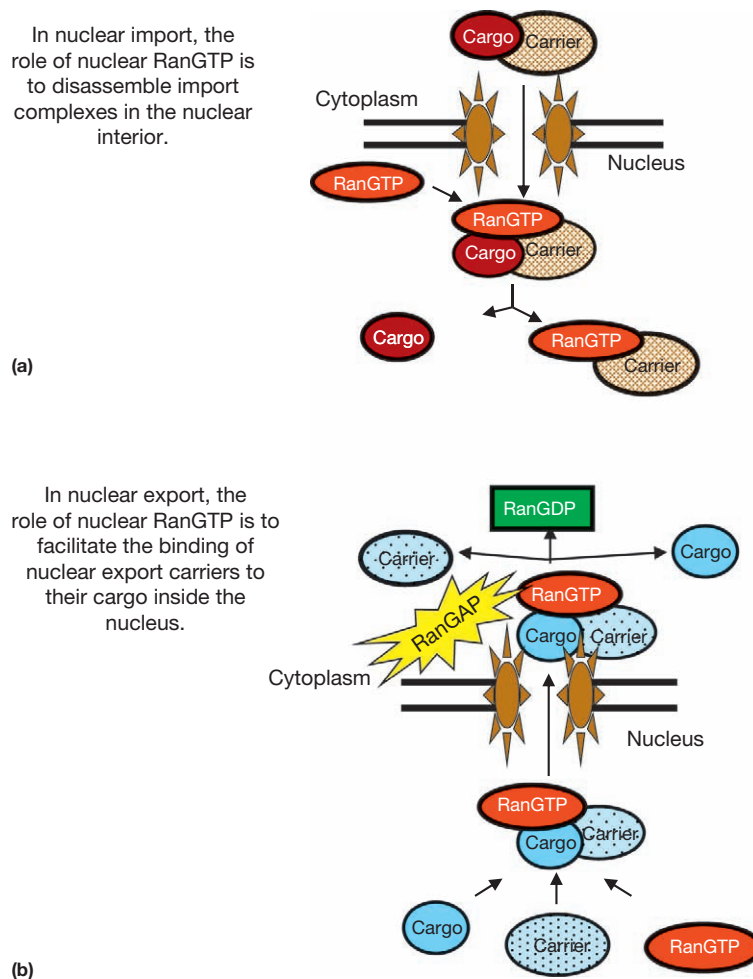


Figure 3 The RanGTP gradient across the NPC controls the assembly and disassembly of nuclear transport complexes. (a) Nuclear import carriers cannot bind import cargo and RanGTP simultaneously. After import of the import carrier: cargo complex through the nuclear pore complex (NPC), the carrier binds nuclear RanGTP. As a result of binding RanGTP, the import carrier releases the import cargo in the nuclear interior. (b) In contrast to nuclear import carriers, nuclear export carriers require bound RanGTP in order to bind their export cargo with high affinity. After export through the NPC, the export complex (RanGTP: export carrier: cargo) encounters the RanGAP in the cytoplasm that stimulates Ran to hydrolyze its bound GTP. Conversion of RanGTP to RanGDP results in the release of RanGDP and export cargo from the export carrier into the cytoplasm.

assembly factors (as opposed to inactive factors complexed with NLS receptors) are also most concentrated there. RanGTP is thus essential for spindle formation; both to release spindle assembly factors from NLS-binding carriers and for correct orientation of the spindle around chromatin.

As stated above, when a human cell enters mitosis at the beginning of cell division, the nuclear envelope fragments into vesicles. At the end of mitosis, after the two daughter cells have separated, these membrane vesicles bind to the decondensing chromatin and then fuse together laterally to re-form an intact nuclear envelope. RanGTP is essential for this fusion step, as is one particular nuclear carrier of the karyopherin- β family (importin- β), however, the exact mechanism is not yet known.

replication of chromatin during S phase and in regulation of kinetochore function throughout the cell cycle. It should be noted that in all of the multiple pathways in which Ran has been implicated, the general nature of the biochemical reaction carried out by Ran remains constant. Switching between RanGDP and RanGTP controls Ran's association with different proteins, and Ran's association with assorted proteins has different downstream effects depending on the protein involved, the cellular location where the interaction occurs, and the stage of the cell cycle when the interaction occurs. In summary, the Ran GTPase plays a powerful and irreplaceable role in the eukaryotic cell, both for constitutive viability and for cell division.

Other Roles of Ran

In addition to the cellular roles for Ran discussed above, Ran has also very recently been implicated in prevention of re-

See also: **Cell Architecture and Function:** Mitosis; Nuclear Pores and Nuclear Import/Export; **Signaling:** Arf Family; Ras Family.

Further Reading

- Arnaoutov A and Dasso M (2003) The Ran GTPase regulates kinetochore function. *Developmental Cell* 5: 99–111.
- Dasso M (2002) The Ran GTPase: Theme and variations. *Current Biology* 12: R502–R508.
- Kunzler M and Hurt E (2001) Targeting of Ran: Variation on a common theme? *Journal of Cell Science* 114: 3233–3241.
- Steggerda SM and Paschal BM (2002) Regulation of nuclear import and export by the GTPase Ran. *International Review of Cytology* 217: 41–91.
- Talcott B and Moore MS (1999) Getting across the nuclear pore complex. *Trends in Cell Biology* 9: 312–318.
- Weis K (2003) Regulating access to the genome: Nucleocytoplasmic transport throughout the cell cycle. *Cell* 112: 441–451.
- Yamaguchi R and Newport J (2003) A role for Ran-GTP and Crm1 in blocking re-replication. *Cell* 113: 115–125.

Ras Family

L A Quilliam, Indiana University School of Medicine, Indianapolis, IN, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Effector Proteins (typically enzymes) that guanosine triphosphate (GTP)-bound Ras proteins bind to elicit a biological response.

GTPase Describes the intrinsic GTP hydrolyzing enzymatic activity of Ras. Binds guanosine diphosphate (GDP) or GTP with high affinity and catalyzes the hydrolysis of GTP to GDP and phosphate.

GTPase-activating proteins (GAPs) Negative regulators of Ras that bind the GTPase and accelerate its ability to hydrolyze bound GTP.

Guanine-nucleotide-exchange factors Upstream positive regulators of Ras proteins that promote increased nucleotide release so that Ras can bind to GTP.

Isoprenylation Process of irreversible posttranslational covalent modification of the C terminus of proteins by 15 carbon farnesyl or 20 carbon geranylgeranyl isoprene-derived lipids.

Ras A high-affinity guanine-nucleotide-binding protein with intrinsic GTPase activity that acts as a molecular switch or biotimer by cycling between inactive GDP- and active GTP-bound states.

Physical Properties

Ras family proteins are often referred to as small or low-molecular-weight guanosine triphosphatases (GTPases). This is due to their 20–25-kDa size being smaller than the 38–52-kDa α -subunits of the heterotrimeric G proteins, another class of cell-signaling molecules that bind and hydrolyze GTP. Ras proteins consist of a core GTP-binding/GTPase domain that encompasses ~160 amino acids, followed by a hypervariable domain that is thought to act as a flexible linker, and a C-terminal CAAX (cysteine-aliphatic-aliphatic-other amino acid) motif that is the signal for posttranslational lipid modification, see [Figure 1](#). Addition of a 15-carbon farnesyl or 20-carbon geranylgeranyl isoprene group to the Cys, cleavage of the three C-terminal (AAX) residues, and methylation of the C-terminal COO-group help promote the membrane targeting of most Ras subfamily GTPases. Additional palmitoylation of nearby cysteine(s) or presence of many basic (arginine and lysine) residues in the hypervariable region helps maintain tight membrane association. Membrane association is essential for Ras protein function and mutation of the CAAX cysteine, preventing the series of posttranslational modifications, disables Ras function. Most Ras superfamily proteins are modified by a geranylgeranyl transferase that adds the 20-carbon lipid. However, H-, K-, and N-Ras are modified by a farnesyl transferase that recognizes a different CAAX sequence that adds a 15-carbon farnesyl lipid instead. Most commonly, a serine or methionine in the X position signals for farnesylation of the CAAX, whereas leucine or phenylalanine signals for geranylgeranylation. Because of this specificity, drugs that mimic the C terminus of Ras and/or the farnesyl lipid can block posttranslational modification and clinical trials have been performed to evaluate their use in cancer chemotherapy.

The GDP/GTP Cycle and Ras Mutations

Ras binds GTP and guanosine diphosphate (GDP), but not guanosine monophosphate, with very high affinity. The intrinsic rates of nucleotide exchange and GTP hydrolysis

by Ras *in vitro* are very slow (on the order of minutes to hours) and not practical for a signaling molecule that needs to rapidly respond to incoming stimuli. However, this is by design. Most Ras proteins are tightly regulated by one or more guanine-nucleotide-exchange factors (GEFs) that help turn it on and GTPase-activating proteins (GAPs) that subsequently help to inactivate it, see [Figure 2](#).

Ras GEFs contain a CDC25 homology domain that shares similarity with the yeast RAS1/2 GEF, CDC25 (not to be confused with the mammalian Cdc25 protein phosphatase). There are over 30 CDC25 homologs in the human genome. Some are highly specific for one Ras protein, while others catalyze nucleotide exchange and multiple family members. GEFs promote the release of GDP, then stabilize the nucleotide-free protein until GTP can subsequently bind and displace them. This occurs very rapidly *in vivo* due to the high affinity of Ras for GTP. Also because GTP is in a 10–20-fold excess over GDP, the GEF is predominantly displaced by GTP. Because of the ability of nucleotide-free Ras to bind tightly to GEFs, dominant inhibitory mutants have been designed that have low affinity for nucleotide. These bind tightly to GEFs and prevent them from activating the endogenous wild-type Ras, so blocking signaling pathways. These are useful tools for implicating Ras proteins or Ras pathways in the control of cellular events. The most commonly used is a serine to asparagine mutation of H-Ras residue 17 (H-Ras17N).

Although Ras is a GTPase, it has a poorly designed catalytic site. However, GAPs help stabilize the catalytic site and contribute a critical arginine residue that greatly accelerated nucleophilic attack of the GTP. One reason the α -subunits of heterotrimeric G proteins are bigger than Ras proteins is that they have an extra domain that contributes an 'arginine finger' resulting in their typically having a higher intrinsic GTPase activity than Ras. Mutations that block intrinsic and GAP-stimulated GTP hydrolysis result in Ras proteins being locked in their active GTP-bound state. Such mutations at residues 12, 13, and 61 are found in H-, K-, or N-Ras in various human tumors. Ras subfamily GAPs fall into two structurally distinct families; those that regulate Rap/Rheb and those that

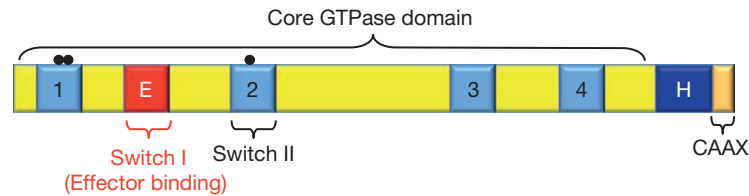


Figure 1 Linear diagram of Ras protein highlighting key sequences. Light-blue areas are highly conserved across the entire Ras superfamily and form the guanine-nucleotide-binding pocket. Closed circles indicate sites of activating mutations found in human cancers. Switch I and II domains change conformation upon GDP vs. GTP binding. E, effector binding loop that associated with downstream effectors; H, hypervariable domain, a 10–20-amino-acid linker that also contributes to membrane and target protein binding; CAAX, the terminal four residues that signal for posttranslational lipid modification of the cysteine residue. The identity of residue X dictates whether the G protein receives a 15- or 20-carbon lipid tail.

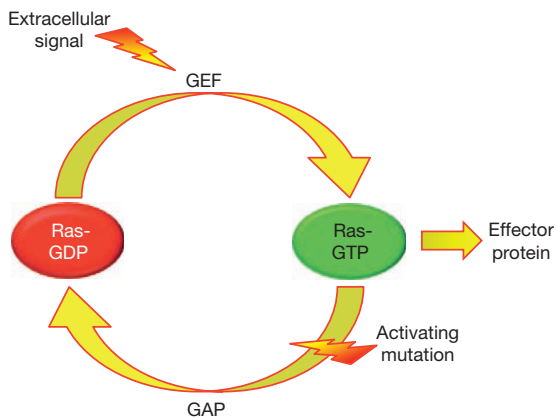


Figure 2 GDP/GTP-cycle activity of Ras family GTPases. Ligand binding induces receptor-mediated activation or membrane-recruitment of guanine-nucleotide-exchange factors (GEFs) which promote the release of GDP from Ras and acquisition of GTP from the cytosol. GTP binding promotes a conformational change permitting Ras to associate with downstream effector proteins to elicit a biological response. GTPase-activating proteins (GAPs) also recognize Ras-GTP and their binding promotes rapid acceleration of Ras' intrinsic GTPase activity. This converts Ras back to its inactive GDP-bound state from which the cycle can begin again. Mutations that block the intrinsic and GAP-stimulated GTPase activity lock Ras in its active, GTP-bound state.

regulate Ras and other GTPases. One of the Ras GAPs, neurofibromin, is the product of the NF1 tumor suppressor and, upon its loss, Ras GTP levels become elevated contributing to increased signaling that results in cancer and other proliferative diseases. Information on the better-characterized Ras subfamily members is provided below.

H-, K-, and N-Ras

H-, K-, and N-Ras represent the original and best-studied/-understood Ras proteins. Mutations of each of these proteins are found in human tumors with those in K-Ras being the most prevalent. K-Ras can have alternate splicing of its fourth exon with the '2' or 'exon 4B' version being the most predominant.

It was long known that Ras is involved in growth control downstream of growth factor receptors, and in 1993, an entire

pathway from cell-surface receptor to the nucleus was delineated. The binding of growth factors to their protein tyrosine kinase receptors leads to phosphorylation of the receptors on tyrosine residues. This enables the binding of various adapter proteins that contain modular Src homology 2 or SH2 domains that recognize specific primary sequences that contain phospho-tyrosine. One such molecule is Grb2 that also has two SH3 domains that bind to proline-rich sequences like those found in the Ras GEF, Sos. Binding of a preformed Grb2–Sos complex to receptor tyrosine kinase results in recruitment of Sos to the plasma membrane where it promotes GTP loading of Ras. In its active GTP-bound state, Ras now adopts a conformation that enables it to bind to the serine/threonine protein kinase, Raf. This, in turn, recruits Raf to the membrane, promotes a conformational change, and enables Raf to be phosphorylated by nearby kinase(s). Raf then phosphorylates MEK, which phosphorylates extracellular signal-regulated kinases (ERKs) 1 and 2. Phospho-ERKs translocate into the nucleus where they phosphorylate transcription factors such as Elk1, leading to increased gene expression. It is now known that a number of additional signals activate Ras. For example, Ras GEFs (such as GRF1 in brain and GRP1 in T cells) and GAPs (Capri, Rasal) can be activated by Ca^{2+} elevation and/or diacylglycerol. Additionally, G-protein-coupled receptors can trigger tyrosine phosphorylation to regulate the Grb2/Sos complex.

Growth factor receptors reside on the plasma membrane and Ras also localizes there. However, Ras is also found on the Golgi membrane and many of the events mediated by Ras occur at this intracellular site. One mechanism whereby activation of cell-surface receptors might signal to Ras at a remote intracellular membrane is via Ca^{2+} elevation, resulting in the activation of a Ca^{2+} -regulated exchange factor, as has been reported in T cells.

Binding of GTP triggers a conformational change in Ras in two regions called switch I and II. Switch I, which encompasses residues ~32–40 is the main determinant for effector binding. Conformation of this and switch II dictate which targets various Ras subfamily members associate with. Although a number of putative Ras effectors have been described, three have been particularly well characterized. These are the protein kinase Raf, phosphatidylinositol 3 kinase (PI3K), and RalGDS, an exchange factor for the Ras-related GTPase, Ral (see below). There are several isoforms of each of these effectors that have different tissue distribution and affinities for various Ras family members. However, each

of these effectors has a related ~60-residue Ras association (RA) or Ras-binding domain (RBD). As RBD/RA domains bind specifically to Ras-GTP, the isolated domains (typically fused to glutathione S-transferase and immobilized on glutathione-conjugated beads) can be used in the lab to precipitate Ras-GTP from cell lysates to gauge intracellular Ras activity. The Raf RBD is typically used to detect H-, K-, or N-Ras activity. Although most Ras activity is associated with increasing cell growth and transformation, several recently characterized Ras targets that include RASSF1 appear to have tumor-suppressing properties.

Ral

As indicated above, a family of four Ral GEFs (RalGDS, Rgl, Rgl2/Rlf, Rgl3/RPM) possess RBDs and are downstream effectors of Ras. The primary function of Ras here is in recruiting Ral-GEFs to the membrane where they can activate Ral. In addition to being regulated by Ras/RalGDS, Ral is activated by Ca^{2+} , possibly due to its ability to bind calmodulin and by a second class of GEFs (the RalGPS family) that lack an RA domain/Ras regulation.

Due to its regulation by Ras, initial studies on Ral focused on its ability to regulate tumorigenesis, and RalGDS has been established as an important mediator of Ras-induced transformation of human epithelial cells. Ral or its GEFs have been reported to regulate various transcription factors including c-Fos, c-Jun, STAT3, and AFX (a member of the forkhead family). Interestingly, Ral-GEFs are often more potent at inducing biological effects than activated (GTPase-defective) Ral mutants. This may be because (1) GTP/GDP cycling is required for Ral function or that (2) RalA and RalB act in concert to promote their biological effects. Depletion of RalA severely impairs the anchorage-independent proliferation of cancer cell lines, whereas RalB seems to be essential for the survival of a variety of tumor-derived cell lines as well as their ability to metastasize. Thus, RalA and RalB appear to cooperate to drive oncogenic transformation.

Ral proteins localize to intracellular vesicles in both endocytic and exocytic compartments and regulate the exocyst, a protein complex that directs vectorial targeting of vesicles from the Golgi to the basolateral membrane of epithelial cells. Ral is essential for this process in higher eukaryotes but, puzzlingly, Ral is not present in yeast where the exocyst was first identified. The ability of RalA to promote the movement of lipid rafts (cholesterol- and sphingolipid-rich membranes replete with key signaling molecules) to the plasma membrane is likely key to its ability to promote anchorage-independent growth. RalB also drives exocyst assembly during polarized cell migration. This along with its ability to increase the transcription of a survival factor, FLIP, contribute to its key role during tumor metastasis. The Ral proteins also regulate receptor-mediated endocytosis but this is achieved via a distinct effector, RalBP1/RLIP76. RalBP1 not only binds to proteins associated with endocytosis but also contains a GAP-like domain for the Rho family GTPase, CDC42. It is via this activity that it may also regulate cell morphology. Given these myriad functions, it is not surprising that Ral proteins can contribute to cell proliferation and cancer.

R-Ras 1, 2, and 3

R-Ras, TC21 (R-Ras2), and M-Ras (R-Ras3) are closely related to Ras (**Figure 3**) and have overlapping signaling and biological functions. However, despite the transforming activity of laboratory-engineered mutants and the identification of a few activating mutations in cultured cell lines, these GTPases appear to seldom be mutated in human cancers. R-Ras activates PI3K but does not regulate the Raf/ERK or RalGDS pathways. Distinct from H-Ras, R-Ras has been implicated in 'inside-out' signaling, increasing the ability of integrins to bind extracellular matrix. This enhancement of receptor-ligand binding results in increased cell attachment and spreading. One study has implicated Rap1 as a mediator of R-Ras action, whereas others show a requirement for the Rho family GTP-binding protein Rnd for R-Ras inactivation, two more examples of Ras protein cascades. Complexes of semaphorins and plexins that play important roles in development and axon guidance regulate R-Ras and orchestrate growth cone collapse.

Like R-Ras, TC21 and M-Ras also effectively activate PI3K. TC21 was recently implicated in the survival and proliferation of T and B cells via direct interaction with antigen receptors. Although M-Ras is most abundant in brain, its exact function is unclear. Despite preferentially activating PI3K, M-Ras can also impact Raf activation by regulation of its phosphorylation state.

Rap1 and 2

There are five members of the Rap cluster: Rap1A and 1B and Rap2A, 2B, and 2C of which Rap1A/B function is best understood. Rap1A/Krev-1 was isolated in 1988 due to sequence homology with Ras (Ras-proximate) and also due to its ability to revert the transforming activity of K-Ras. The revertant activity was believed due to Rap1A binding to Ras effectors and acting as a competitive antagonist of Ras. This indeed may occur in T cells where Rap needs to be turned off by a costimulatory receptor such as CD28 in order for T-cell receptor-induced activation of Ras to have its desired effect. However, the main role of Rap1 appears to be in the inside-out regulation of integrins. A Rap1 effector RAPL has recently been described in leukocytes, which is responsible for promoting integrin clustering but how Rap regulated adhesion of cells in solid tissues is still not fully understood. Although early studies described Rap1 and 2 localization to endocytic and Golgi membranes, Rap1 also localizes to epithelial and endothelial cell adherens junctions where it plays a role in junction maintenance. It is likely that Rap1 and 2 plays multiple cell type-specific roles.

Rap1 and 2 are regulated by the same GEFs and GAPs. Strikingly, at least 10 of the CDC25 family as well as the unrelated Smg GDS and DOCK4 proteins act as Rap GEFs suggesting that Raps play fundamental roles in cell biology that are triggered by a broad spectrum of extracellular stimuli. Signals from Ca^{2+} , diacylglycerol, cyclic adenosine monophosphate (cAMP), as well as receptor and nonreceptor protein tyrosine kinases can all activate Rap. Indeed the protein kinase A-independent effects of cAMP are due to activation of two Rap GEFs, Epac1 and 2, that contain cyclic nucleotide-binding

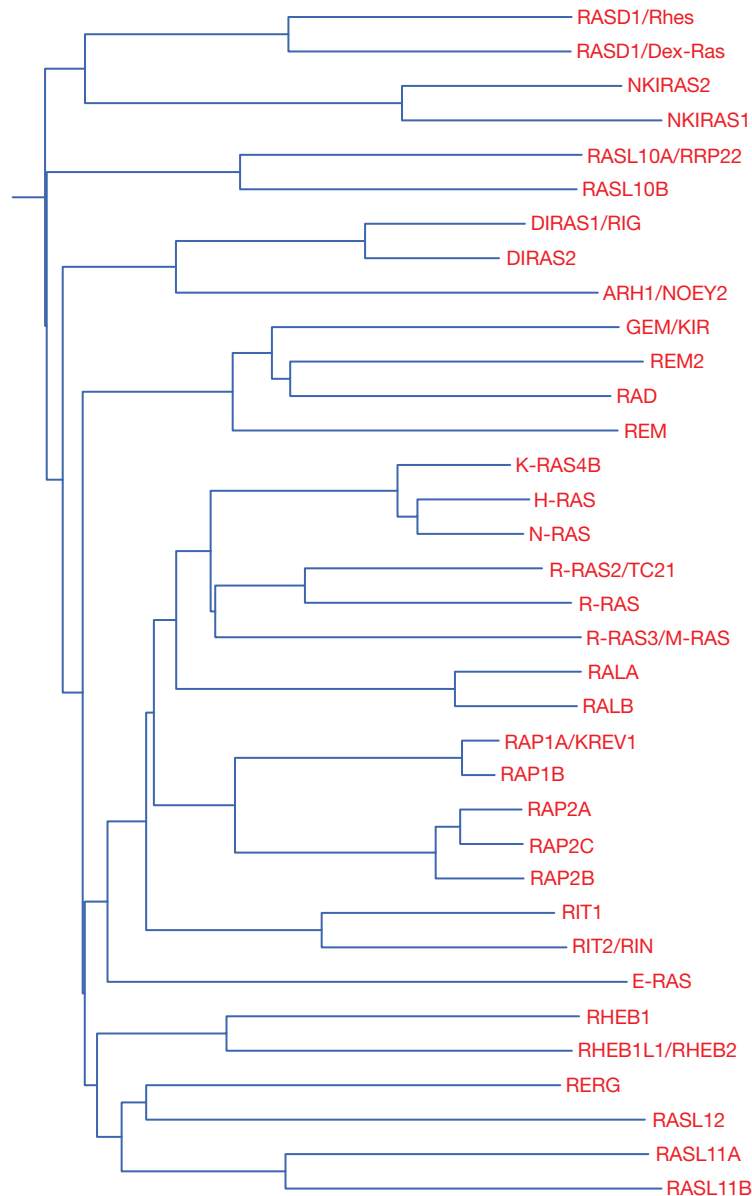


Figure 3 Dendrogram showing relatedness of mammalian Ras subfamily proteins. Commonly used names in literature are included as well as official gene names.

domains. Several Rap GEFs also contain RA domains, and are recruited to membranes as downstream effector of other GTPases. This suggests that Rap activation may be a requisite for other Ras family members to execute their biological functions. Rap1a and 1b null mice as well as work using small interfering (si) RNA have recently implicated Rap1 in endothelial cell adherens junction formation, angiogenesis, and a variety of leukocyte functions.

Rheb

Despite its name, Rheb (Ras homolog expressed in brain) is found in many tissues. However, it was discovered in 1995 as a

protein whose expression was induced by synaptic stimulation. Although no Rheb GEFs have been found to date, Rheb's GTPase activity is regulated by TSC2/tuberin, a tumor-suppressor protein whose C terminus resembles that of Rap GAPs. Following loss of TSC2 or mutation of its GAP domain, the loss of GAP activity results in the accumulation of Rheb GTP. This activates the mammalian target of rapamycin, mTOR, a kinase that stimulates protein synthesis in response to growth factors and nutrient abundance. Interestingly, Rheb is posttranslationally farnesylated similarly to H-, K-, and N-Ras. However, the use of farnesyl transferase inhibitors to block mTOR activation as a result of Rheb activation downstream of various tumor-suppressor proteins (TSC1, TSC2, PTEN, NF1, or LKB1) has not proved successful.

Other Ras Family Members

A number of additional GTPases exist (Figure 3) that are less well characterized but help exemplify the diversity of GTPase functions. These include the ARHI/DiRas and Rerg/RasL11 families that have each been shown to play inhibitory roles, the latter being suppressed in breast and prostate cancers, respectively. Critical conserved residues within their GTP-binding pockets diverge from those of most other Ras family members suggesting that these G proteins are constitutively GTP bound. The expression of these proteins can be regulated by various agonists suggesting that they are regulated at the level of gene expression rather than by GEFs and GAPs. Further, it is often their loss rather than mutation that contributes to cellular proliferation/transformation. Transcriptional regulation is not limited to these proteins; for example, RasD1/Dex-Ras expression is induced by corticosteroids, RasD2/Rhes by thyroid hormone and Rheb by estrogen and, as noted above, synaptic stimulation. Another branch of the Ras subfamily regulates transcription: NKIRas1/2 bind to the nuclear factor κ B/inhibitor of NF κ B (NF κ B/I κ B) complex and help mask a nuclear localization sequence so that this transcription factor complex cannot translocate to the nucleus and induce gene expression.

The function of the RGK branch of the Ras family (Rem1 and 2, Rad, and Gem/Kir) has until recently eluded investigators. However, they have now been shown to negatively regulate the Rho–Rho kinase pathway, helping to explain their ability to modulate the actin cytoskeleton. Additionally, they interact with and regulate voltage-dependent calcium channels.

Finally, several Ras family proteins lack the characteristic C-terminal CAAX box required for posttranslational lipid modification. These include the closely related Rit and Rin proteins as well as Rerg and NKIRas. Although the former two proteins have been reported to regulate PC12 (pheochromocytoma) cell differentiation, an event regulated by several Ras proteins via Raf and/or PI3K, Rit, and Rin appear to act via some alternative pathway(s). Interestingly, although Rit lacks any

known recognition signal for C-terminal lipid modification, proliferation and survival of Rit-transformed cells in low-serum growth medium is dependent on a farnesylated protein. As most prenylated proteins are G proteins, this suggests, once again, the possibility of cross talk between members of the Ras subfamily of GTPases exists.

See also: [Signaling: Arf Family](#); [Rab Family](#); [Ran GTPase](#); [Small GTPases](#).

Further Reading

- Berg JM, Tymoczko JL, and Stryer L (2002) *Biochemistry*, 5th edn. New York, NY: WH Freeman.
- Blum R, Cox AD, and Kloog Y (2008) Inhibitors of chronically active Ras: Potential for treatment of human malignancies. *Recent Patents on Anticancer Drug Discovery* 3: 31–47.
- Bodemann BO and White MA (2008) Ral GTPases and cancer: Linchpin support of the tumorigenic platform. *Nature Reviews Cancer* 8: 133–140.
- Boettner B and Van Aelst L (2009) Control of cell adhesion dynamics by Rap1 signaling. *Current Opinion in Cell Biology* 21: 684–693.
- Colicelli J (2004) Human RAS superfamily proteins and related GTPases. *Science STKE* 250: re13.
- Correll RN, Pang C, Niedowicz DM, Finlin BS, and Andres DA (2008) The RGK family of GTP-binding proteins: Regulators of voltage-dependent calcium channels and cytoskeleton remodeling. *Cellular Signalling* 20: 292–300.
- Hancock JF (2003) Ras proteins: Different signals from different locations. *Nature Reviews Molecular Cell Biology* 4: 373–384.
- Henis YI, Hancock JF, and Prior IA (2009) Ras acylation, compartmentalization and signaling nanoclusters. *Molecular Membrane Biology* 26: 80–92.
- Quilliam LA, Rebhun JF, and Castro AF (2002) A growing number of guanine nucleotide exchange factors is responsible for activation of Ras family GTPases. *Progress in Nucleic Acid Research and Molecular Biology* 71: 391–444.
- Rosner M, Hanneder M, Siegel N, Valli A, Fuchs C, and Hengstschnager M (2008) The mTOR pathway and its role in human genetic diseases. *Mutation Research* 659: 284–292.
- Takai Y, Sasaki T, and Matozaki T (2001) Small GTP-binding proteins. *Physiological Reviews* 81: 153–208.
- Weinberg RA (2007) *The Biology of Cancer*. New York, NY: Garland Sciences.

Reactive Oxygen and Nitrogen Species: Interactions with Mitochondria and Pathophysiology

B R Zelickson, S W Ballinger, L J Dell'Italia, J Zhang, and V M Darley-USmar, University of Alabama at Birmingham, Birmingham, AL, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Electron transport chain The site of oxidative phosphorylation in the mitochondrial inner membrane which couples the transfer of electrons from electron donors (such as NADH) to O₂ with the pumping of protons (H⁺) from the matrix to the intermembrane space, thus generating an electrochemical proton gradient which is used by adenosine triphosphate (ATP) synthase to produce ATP.

Flavin The prosthetic group in flavoproteins which is capable of undergoing oxidation–reduction reactions by accepting/donating either one or two electrons at once.

Heme The prosthetic group in hemoproteins which contains an iron atom at the center of a porphyrin ring and serves various roles, such as a source/sink of electrons during electron transport and redox reactions and the transportation of diatomic gases.

Iron–sulfur center Moiety in iron–sulfur proteins, often consisting of four iron atoms and four sulfur atoms arranged in a cubane-like structure, which participates in redox reactions, particularly in mitochondrial electron transport.

Krebs cycle Series of enzyme-catalyzed reactions occurring in the mitochondrial matrix which uses acetyl-coenzyme (CoA) and oxaloacetate to reduce three NAD⁺ to NADH and one Q to QH₂, while also producing one GTP and two CO₂.

Oxidative stress State in which there is an increase in oxidizing species, such as reactive oxygen species (ROS), and/or a decrease in antioxidant defenses, and often leads to protein, lipid, and DNA damage.

Redox center Moiety which is either oxidized or reduced during redox reactions (e.g., flavins, iron–sulfur centers, and hemes).

Reductive stress State of redox imbalance in which there is an excess in reducing species, leading to a reductive cytosolic environment (e.g., high NADH:NAD⁺ ratio).

Thiols Organosulfur compound containing a carbon-bound sulfhydryl group (e.g., cysteine and glutathione) which participates in redox biology and signaling through the formation of various oxidation states.

Introduction

The importance of reactive oxygen and nitrogen species (ROS/RNS) in physiology, as well as in pathology, and their interactions with mitochondria are well established. Originally, ROS/RNS were considered to be simply destructive and cytotoxic, but more recently it has become clear that they have several beneficial effects through their ability to activate specific cell-signaling pathways. ROS have been shown, at low levels, to induce cell proliferation and control the synthesis of endogenous antioxidant enzymes. Mitochondria are both a source and target for ROS and play a key role in redox cell signaling. Under pathological conditions, increased levels of mitochondrial ROS formation can lead to damage to DNA, lipids, and proteins. This can impair bioenergetic function, and, in some cases, initiate programmed cell death or apoptosis (Figure 1).

Nitric oxide (•NO) is the best-understood redox signaling molecule. It modulates cell function through its interaction with heme proteins such as soluble guanylate cyclase, cytochrome *c* oxidase, and hemoglobin. In addition, •NO can lead to protein modification through its ability to interact with ROS leading to S-nitros(yl)ation or oxidation. On the other hand, the reactions of ROS and RNS have been implicated in many chronic pathologies, particularly those involving inflammation. For example, high levels of reactive species have been shown to further the progression of various cardiovascular and neurodegenerative diseases.

Nitric oxide is produced by a family of enzymes called the nitric oxide synthases (NOSs) and mediates its cell signaling effects through protein modification and binding to heme iron. Despite its predominantly extra-mitochondrial origin, •NO is also known to modulate mitochondrial function at cytochrome *c* oxidase and to stimulate mitochondrial biogenesis by activation of soluble guanylate cyclase. The aim of this chapter is to discuss what reactive oxygen and nitrogen species are, how they are formed, and how they interact with the mitochondrion and give key examples of their biological effects in cardiovascular and neurodegenerative pathologies.

What Are Reactive Oxygen and Nitrogen Species?

The terms ROS/RNS are freely used in the biological literature; however, it is important to note that they do not have a specific chemical definition and may include free radicals and nonradical species. The general property they all share is that ROS/RNS react with biological molecules leading to changes in function that may contribute to cell signaling in a controlled fashion or to the pathology of a disease. It seems likely that the modification of protein cysteine residues is the basis of redox cell signaling and can be highly specific. In contrast, nonspecific oxidation reactions such as carbonylation are generally thought to contribute to protein damage, aggregation, and activation of the autophagic process.

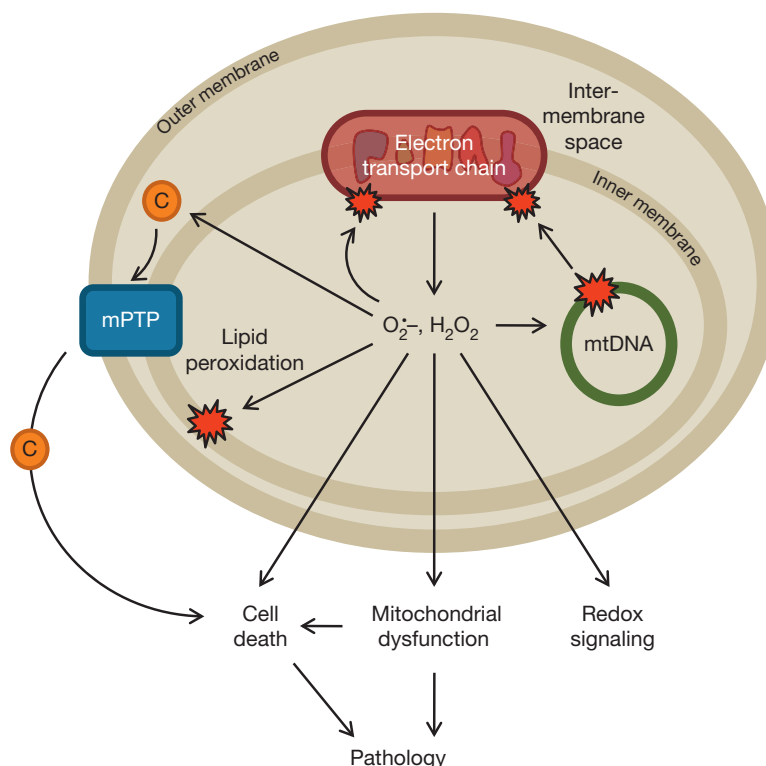


Figure 1 Effects of mitochondrial ROS. Mitochondrial ROS production can alter the function through a variety of pathways. Lower levels of ROS can act through several redox signaling mechanisms, while higher levels can directly damage mitochondrial proteins, leading to the further production of ROS. ROS can also damage mtDNA and induce lipid peroxidation in the membrane, and can also cause mitochondrial permeability transition pore (mPTP) opening, which leads to cytochrome *c* release and apoptosis induction.

Reactive Oxygen Species

The ROS are the partially reduced intermediates of oxygen and can be radical or nonradical molecules. Free radicals are denoted in a chemical formula using a superscript dot which is usually written next to the symbol of the atom which carries the unpaired electron. ROS include the superoxide radical anion (O₂^{•-}) and its protonated form, the hydroperoxyl radical (HO₂[•]), hydrogen peroxide (H₂O₂), and the hydroxyl radical (•OH). As discussed below, O₂^{•-} is formed by the mitochondrial respiratory chain at a number of sites within electron transport and the Krebs cycle. As the mitochondrion contains high levels of the enzyme superoxide dismutase (SOD), most of the O₂^{•-} that is formed is converted into H₂O₂.

Reactive Nitrogen Species

As with reactive oxygen species, this is a term commonly used by biologists to describe •NO-derived molecules which show biological activity. Nitric oxide is an uncharged free radical which can diffuse freely across membranes. It reacts rapidly with other free radicals and has a high affinity for ferrous heme groups. The RNS peroxynitrite (ONOOH) is formed from the reaction of superoxide with •NO and has been studied in depth. The reactivity of peroxynitrite is enhanced in biological systems through its reaction with carbon dioxide to form nitrosoperoxycarbonate, which can then undergo homolytic fission to generate the nitrogen dioxide (NO₂[•]) and carbonate

(CO₃^{•-}) radicals, both of which are reactive and can cause damage to biomolecules. For example, NO₂[•] can react with tyrosine residues or tyrosyl radicals on proteins, resulting in 3-nitrotyrosine formation. This adduct can be detected in a broad range of pathologies, and is often used as a marker of nitrative stress. Nitrated proteins have been detected in mitochondria isolated from a broad range of pathological conditions. Mitochondrial DNA damage is associated with several diseases associated with oxidative and nitrative stress, leading to increased mitochondrial ROS production and dysfunction.

Generation of ROS from Mitochondria

Mitochondria can generate ROS from a number of different redox centers in several metabolic pathways. The basic principle for ROS generation from these sites is shown in [Figure 2\(a\)](#). Electrons are transferred to a redox center which then acts as an intermediate carrier to another redox active member of the metabolic pathway. The more reduced the redox center is, the greater the potential production of ROS. The detailed mechanisms for ROS generation from each redox center are described below.

Complex I

NADH-ubiquinone oxidoreductase (complex I) consists of 45 subunits with a combined mass of close to 1 MDa. Complex I accepts two electrons from NADH, which is produced by the

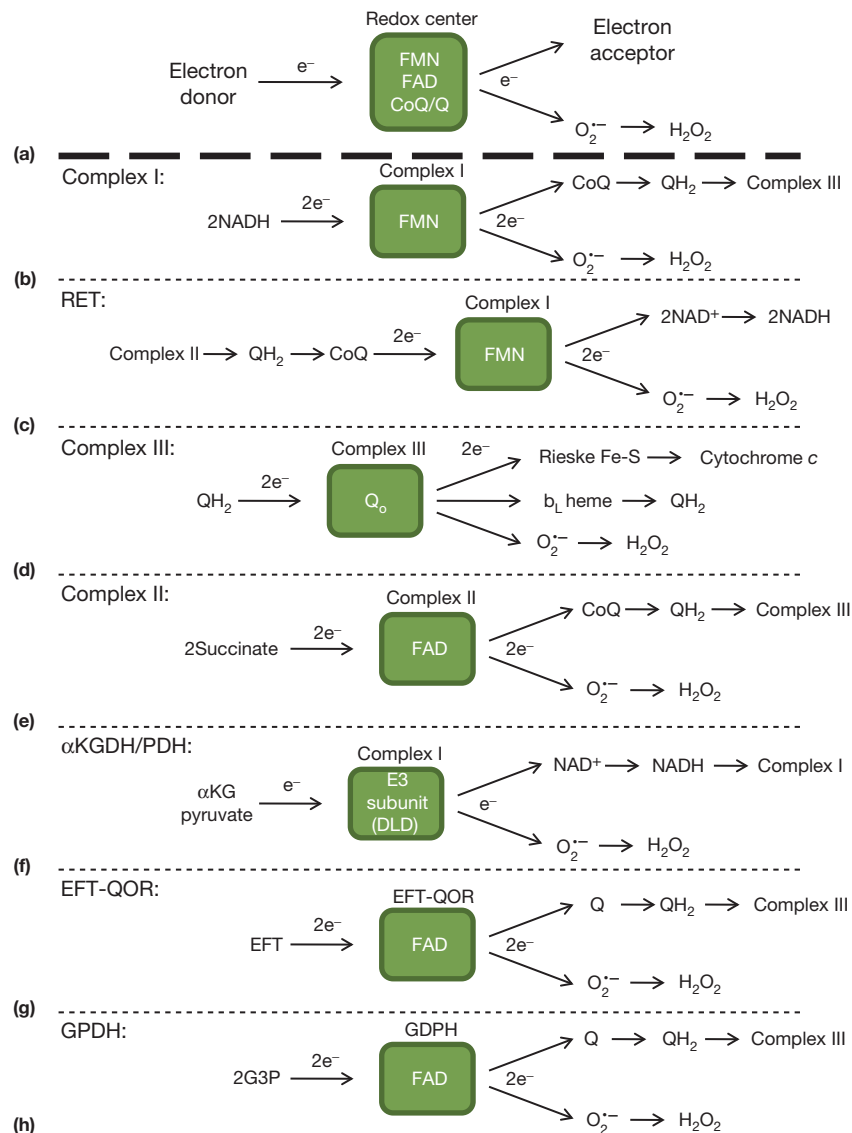


Figure 2 Mechanisms of ROS formation in the mitochondria. The sites of ROS production within the mitochondria all follow a similar mechanism. An electron donor donates an electron to a redox center, which then reduces the electron acceptor. However, electrons can also be transferred from the redox center to reduce oxygen into superoxide (a). Complex I oxidizes NADH through the FMN site, and can transfer the electrons to either the CoQ site to make ubiquinol (QH_2) or, if the CoQ site is inhibited, to O_2 to make superoxide (b). In reverse electron transport (RET), QH_2 produced by complex II can be transferred back into complex I through the CoQ site to the FMN, which can then either reduce NAD^+ back to NADH or reduce O_2 to form superoxide (c). Complex III uses QH_2 produced predominantly by complex I and II as the electron donor, transferring the electrons to the Rieske Fe-S and then on to cytochrome c1 and finally cytochrome c. Complex III also transfers electrons to the b_L heme, which participates in the Q cycle. However, when the Q cycle is inhibited, by antimycin A, for example, the electrons can leak out of the Q_o site to reduce O_2 to make superoxide (d). Complex II has also been shown to produce superoxide when electron transfer to the CoQ site is inhibited, or when the enzyme is solubilized, causing electrons to leak from the reduced FAD site to make superoxide (e). PDH and α KGDH use their substrates to transfer electrons through their E3 subunit to reduce NAD^+ ; however, electrons have also been shown to reduce O_2 to make ROS (f). EFT-QOR oxidizes EFT and transfers electrons through its FAD domain to reduce Q to QH_2 , but it has also been shown that it can also reduce O_2 to superoxide as well (g). GPDH reduces G3P via its FAD site to make QH_2 , although it has also been shown to leak electrons to make ROS as well (h).

Krebs cycle, via a flavin mononucleotide (FMN) and transfers them through a chain of seven iron-sulfur centers to the CoQ reduction site. There the electrons are transferred to the lipid-soluble electron carrier ubiquinone (Q), reducing it to ubiquinol (QH_2). One mechanism by which complex I produces ROS is by the reaction of O_2 with the fully reduced FMN

forming $O_2^{\bullet-}$ on the matrix side of the complex. The rate of $O_2^{\bullet-}$ production at the FMN can be increased by inhibiting the CoQ site using the small molecule inhibitor rotenone. Similarly, this reaction is dependent on the NADH/ NAD^+ ratio, which determines the proportion of FMN that is fully reduced. Under physiological conditions, $O_2^{\bullet-}$ generation

from complex I is sensitive to the redox state of the respiratory chain (Figure 2(b)).

Pathologically, there are several conditions which result in a reductive stress, causing the NADH/NAD⁺ ratio to increase, and this in turn results in increased O₂^{•−} production. Some of the best-understood examples of this condition are the reductive stresses associated with alcohol metabolism and ischemia.

An additional mechanism through which complex I can produce ROS is called reverse electron transport (RET) (Figure 2(c)). This occurs when the mitochondria are utilizing succinate as a substrate through complex II. Under conditions of low electron flux (when adenosine diphosphate (ADP) levels are low), the larger pool of reduced QH₂ is sufficient to reduce complex I at the CoQ site. This reverses the normal flow of electrons back through the complex to reduce the FMN site, thereby reducing NAD⁺ to form NADH, while also reducing O₂ to form O₂^{•−}. RET results in the generation of O₂^{•−} in the matrix and can be inhibited with rotenone treatment. Interestingly, RET generates the highest rate of O₂^{•−} formation detected thus far in the mitochondrion. Whether RET plays a physiological role in cell signaling through ROS is not clear.

Since complex I appears to play a central role in mitochondrial ROS formation, it is not surprising that it has been implicated in a number of pathological processes. For example, the complex I (and RET) inhibitor rotenone has been shown to induce Parkinson's disease-related pathologies in rodents. It may also be an important environmental toxicant which can contribute to Parkinson's disease pathogenesis in humans. Furthermore, many studies have consistently found decreases in complex I activity in Parkinson's disease brains. Other compounds have been shown to interact with complex I, modifying both enzyme activity and ROS formation. An important example is 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), originally discovered as a by-product of illegal heroin production, which was shown to cause Parkinsonism in drug users. MPTP toxicity is mediated by its metabolite MPP⁺, which is selectively taken up by dopaminergic neurons and inhibits complex I activity, leading to their eventual death.

Complex III

Complex III is approximately 240 kDa and is made up of 11 subunits, three heme groups, and the Rieske iron-sulfur (Fe-S) center. It functions by oxidizing QH₂ in the inner membrane and transferring the electrons to reduce cytochrome *c* in the intermembrane space. The rates of O₂^{•−} production from complex III are relatively low during RET compared to that of complex I. However, the further inhibition of complex III with the Q_o site inhibitor myxothiazol attenuates most of this O₂^{•−} production, which suggests that complex III does produce some O₂^{•−} during RET, albeit a small amount. Alternatively, when the Q_i site is inhibited using antimycin A, complex III produces greater amounts of O₂^{•−}, which are released to both sides of the inner membrane. This formation of O₂^{•−} is due to the reaction of O₂ with the QH bound to the Q_o site. This O₂^{•−} formation due to antimycin A can be abrogated by treatment with myxothiazol, which inhibits the Q_o site, suggesting that QH bound to the Q_o site is the main site of O₂^{•−} production (Figure 2(d)).

Complex II

Succinate dehydrogenase, also known as complex II, is embedded in the matrix side of the inner mitochondrial membrane and consists of four subunits. It contains a flavin adenine dinucleotide (FAD) site, three Fe-S clusters, a heme group, and a Q binding site. Complex II oxidizes succinate to fumarate as part of the Krebs cycle, which is coupled to the reduction of Q to QH₂. Complex II readily produces O₂^{•−} when solubilized, probably by the reaction of O₂ with the reduced FAD site, due to the lack of Q as an electron acceptor (Figure 2(e)). This can occur *in vivo*, and only when complex II has been damaged by oxidative stress or aging. There are also several genetic mutations of the various subunits of complex II which can cause increased O₂^{•−} production. Several of these mutations have been shown to lead to cancer, particularly cancers of neuroendocrine lineage. However, complex II produces very little O₂^{•−} during normal physiological conditions.

Other Sources of ROS Formation within the Mitochondrion

Other mitochondrial metabolic pathways, including the Krebs cycle, contain proteins with redox centers which are also capable of generating ROS. For example, the enzyme glycerol-3-phosphate dehydrogenase (GPDH) has been shown to produce O₂^{•−}. This enzyme participates in the glycerol-3-phosphate (G3P) shuttle, which provides a mechanism for the transport of electrons from cytosolic NADH to the electron transport chain, by oxidizing NADH on the cytosolic face of the outer mitochondrial membrane and reducing Q to QH₂ in the inner mitochondrial membrane. GPDH has been reported to have some G3P-dependent production of O₂^{•−}, most likely at the FAD site within the enzyme (Figure 2(h)). Approximately 70% of the O₂^{•−} produced is directed to the intermembrane space, while 30% is directed toward the matrix.

Similarly, α -ketoglutarate dehydrogenase (α KGDH), which is responsible for the conversion of α -ketoglutarate to succinyl CoA in a reaction which is coupled to the reduction of NAD⁺ to NADH, can generate O₂^{•−} when the NADH/NAD⁺ ratio is elevated. This occurs in the E3 subunit of the enzyme complex, which is a flavoprotein known as dihydrolipoamide dehydrogenase (DLD). Pyruvate dehydrogenase (PDH) has been shown to produce ROS via a similar mechanism (Figure 2(f)).

In addition, many of the key enzymes controlling fatty acid oxidation are located in the mitochondrial matrix and inner membrane. In this metabolic pathway, reduction of the electron transfer flavoprotein (ETF) results in the conversion of Q to QH₂ via the enzyme ETF:ubiquinone oxidoreductase (ETF-QOR). ETF-QOR is located on the mitochondrial inner membrane and contains an Fe-S cluster and an FAD site, and has been shown to produce O₂^{•−} directed toward the matrix when in the presence of palmitoyl carnitine as a substrate. ETF-QOR has also been shown to increase RET and so can enhance O₂^{•−} production from complex I (Figure 2(g)).

Can the Mitochondria Sense the Site of ROS Formation?

If mitochondrial ROS play a role in cell signaling, it implies that the site of ROS formation within the respiratory chain will lead to differential modification of mitochondrial proteins.

This is a relatively new concept, but in support of this idea it has been shown that ROS produced by complex III upon treatment with antimycin A induces the modification of protein thiols on complex III core protein I, which does not occur with exogenous hydrogen peroxide treatment or with the induction of RET. These modifications are reversible by the peroxiredoxin/thioredoxin system, as well as by glutathione, which further support their roles as mediators of redox signaling. Reversible thiol modification due to ROS production by specific respiratory chain complexes may be responsible for regulation of carbohydrate and fatty acid metabolism. Mitochondrial protein thiols may be more susceptible to modification than cytosolic protein thiols because of both the proximity to the source of ROS and the fact that the mitochondrial matrix has a high pH (~8), which will result in more thiols being in their reactive form, the thiolate anion.

Mitochondria have also been shown to sense ROS levels through cardiolipin oxidation. Cardiolipin is an abundant lipid which is found in the mitochondrial inner membrane. Cardiolipin participates in redox signaling due to its propensity for peroxidation in the presence of high levels of ROS. The peroxidation of cardiolipin then enables it to induce apoptosis by aiding in mitochondrial pore opening and cytochrome *c* release from the intermembrane space.

Can Mitochondrial ROS be Controlled?

The mitochondrion has several ways through which it can control the levels of ROS it produces. Superoxide can be converted to hydrogen peroxide by the enzyme SOD. The SOD located in the mitochondrial matrix contains a Mn prosthetic group and is known as SOD2 or MnSOD. The SOD located in the cytosol contains Cu/Zn prosthetic groups and is known as Cu/Zn SOD or SOD1. SOD1 may also be present in the mitochondrial inter-membrane space. SOD2 provides a major mechanism by which the mitochondrion can control the levels of ROS, through its ability to catalyze the dismutation of $O_2^{\bullet -}$ to hydrogen peroxide, which can be further metabolized to H_2O by peroxiredoxin, thioredoxin, or glutathione peroxidase. In addition, these systems can also help to remove other peroxides, such as those that have formed on lipids and proteins.

Mitochondria also control reactive oxygen species production through the ability of uncoupling proteins (UCPs) to increase proton leak and decrease the mitochondrial membrane potential and the redox state of sites of $O_2^{\bullet -}$ formation. UCP3 has also been suggested to lower mitochondrial levels of lipid peroxidation by causing fatty acid efflux out of the matrix when fatty acid levels are in excess.

Nitric Oxide and Mitochondria

Nitric oxide partitions into the mitochondrial inner membrane and inhibits cytochrome *c* oxidase by competing with O_2 for binding at the binuclear Cu_B /cytochrome a_3 site. $\bullet NO$ binding to complex IV is O_2 dependent, which may be beneficial because it can then allow O_2 to diffuse further into tissues, thus extending the O_2 gradient. However, excessive inhibition of mitochondrial respiration by $\bullet NO$ may lead to inhibition of

ATP formation and can be pathological. In addition to modulating mitochondrial function, low levels of $\bullet NO$ are now known to be able to induce mitochondrial biogenesis. This occurs through $\bullet NO$ activating soluble guanylate cyclase (sGC), which then produces cyclic GMP (cGMP). The cGMP then leads to the activation of PGC-1 α , which is the principal transcription regulator of mitochondrial biogenesis.

However, when $\bullet NO$ is produced in excess, it impacts on pathologies with an underlying bioenergetic defect, including ischemia reperfusion. One prominent example of $\bullet NO$ production in the central nervous system is associated with ischemic stroke. Initial brain damage is caused by lack of blood flow, oxygen, and nutrients. Further exacerbating the initial insult, injured ischemic brain cells release excitatory amino acids, such as glutamate, into the synapses between neurons. Excitatory amino acids bind to and induce over-excitation of neighboring neurons, leading to calcium influx. Both the primary ischemia itself and the secondary over-excitation of neighboring neurons followed by calcium influx induce NOS, resulting in overproduction of $\bullet NO$. The excessive $\bullet NO$ in neurons, in turn, causes DNA, protein, and organelle damage, and promotes additional cell death with a major contribution from mitochondrial calcium overload and bioenergetic failure.

Summary

In this short overview, we have outlined how ROS/RNS are formed as well as how they function in (patho)physiology. Although much is already known about the production and function of these reactive species, many aspects of how ROS and RNS are formed by and interact with the mitochondrion are still unclear, particularly *in vivo*. A greater understanding of how these species are formed, and how they affect mitochondrial and cellular function, is important for the development of mitochondrial therapeutics.

Acknowledgments

The authors are funded by T32 HL007918 (BRZ), R01 HL77419 (SWB), R01 HL94518 (SWB), R01 NS064090 (JZ), R01 ES010167 (VDU), R01 DK075865 (VDU), R01 AA013395 (VDU), and T32 HL007918 (VDU).

See also: **Protein/Enzyme Structure Function and Degradation:** Flavins; Heme Proteins.

Further Reading

- Ballinger SW (2005) Mitochondrial dysfunction in cardiovascular disease. *Free Radical Biology and Medicine* 38: 1278–1295.
- Ballinger SW, Patterson C, Knight-Lozano CA, et al. (2002) Mitochondrial integrity and function in atherogenesis. *Circulation* 106: 544–549.
- Brand MD, Affourtit C, Esteves TC, et al. (2004) Mitochondrial superoxide: Production, biological effects, and activation of uncoupling proteins. *Free Radical Biology and Medicine* 37: 755–767.
- Brookes PS, Yoon Y, Robotham JL, Anders MW, and Sheu SS (2004) Calcium, ATP, and ROS: A mitochondrial love-hate triangle. *American Journal of Physiology – Cell Physiology* 287: C817–C833.

- Cadenas E (2004) Mitochondrial free radical production and cell signaling. *Molecular Aspects of Medicine* 25: 17–26.
- Cooper CE and Giulivi C (2007) Nitric oxide regulation of mitochondrial oxygen consumption II: Molecular mechanism and tissue physiology. *American Journal of Physiology – Cell Physiology* 292: C1993–C2003.
- Dickinson DA, Levonen AL, Moellering DR, et al. (2004) Human glutamate cysteine ligase gene regulation through the electrophile response element. *Free Radical Biology and Medicine* 37: 1152–1159.
- Erusalimsky JD and Moncada S (2007) Nitric oxide and mitochondrial signaling: From physiology to pathophysiology. *Arteriosclerosis, Thrombosis, and Vascular Biology* 27: 2524–2531.
- Gutierrez J, Ballinger SW, Darley-Usmar VM, and Landar A (2006) Free radicals, mitochondria, and oxidized lipids: The emerging role in signal transduction in vascular cells. *Circulation Research* 99: 924–932.
- Hill BG, Dranka BP, Bailey S, Lancaster J, and Darley-Usmar V (2010) What part of NO don't you understand? Some answers to the cardinal questions in nitric oxide biology. *Journal of Biological Chemistry* 285: 19699–19704.
- Kagan VE, Bayir HA, Belikova NA, et al. (2009) Cytochrome *c*/cardiolipin relations in mitochondria: A kiss of death. *Free Radical Biology and Medicine* 46: 1439–1453.
- Murphy MP (2008) Targeting lipophilic cations to mitochondria. *Biochimica et Biophysica Acta* 1777: 1028–1031.
- Murphy MP (2009) How mitochondria produce reactive oxygen species. *Biochemical Journal* 417: 1–13.
- Pacher P, Beckman JS, and Liaudet L (2007) Nitric oxide and peroxynitrite in health and disease. *Physiological Reviews* 87: 315–424.
- Radi R, Cassina A, Hodara R, Quijano C, and Castro L (2002) Peroxynitrite reactions and formation in mitochondria. *Free Radical Biology and Medicine* 33: 1451–1464.
- Shiva S, Oh JY, Landar AL, et al. (2005) Nitroxa: The pathological consequence of dysfunction in the nitric oxide–cytochrome *c* oxidase signaling pathway. *Free Radical Biology and Medicine* 38: 297–306.

Recombination-Dependent DNA Replication

K N Kreuzer, Duke University Medical Center, Durham, NC, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Direct restart pathways Pathways that restart replication forks that are intact but have become stalled or blocked.

Gene conversion Nonreciprocal recombination event in which one allele is converted into the form of the second allele.

Homologous recombination Recombination between two nearly identical sequences that are not allelic (i.e., not at the same corresponding position on the chromosome).

Inducible stable DNA replication (iSDR) A pathway of recombination-dependent DNA replication in *Escherichia coli* that occurs without the normal initiator protein DnaA, but which does require recombination proteins (including RecA and RecBCD) as well as SOS induction.

Recombination-dependent replication (RDR) Extensive replication of a chromosome that is dependent on homologous genetic recombination and recombination proteins.

Recombinational restart The restart of broken replication forks by an RDR reaction.

Synthesis-dependent strand annealing (SDSA)

A double-strand break repair pathway in which one invading 3' end is extended by DNA polymerase using a homologous template; as the 3' end is extended, the newly synthesized strand at the back of the D-loop is extruded, and eventually pairs with the other broken 3' end to reattach the two broken ends to each other.

Early Models of Homologous Recombination Invoke Recombination-Dependent DNA Replication

One of the early molecular models for homologous recombination, called break-and-copy, invoked recombination-dependent DNA replication (RDR). One simple version of this model invoked a break in one molecule, with a resulting 3' broken end being used as primer for DNA replication. Another DNA molecule, of different parentage, is used as the template, resulting in a genetic crossover in the resulting DNA molecule.

As the field progressed, evidence accumulated in several systems for break-and-rejoin models without extensive DNA replication (but sometimes short gap filling reactions). Thus, break-and-copy models fell into disrepute for some time. We now understand that there are not only pathways of homologous recombination that do not involve extensive DNA replication, but also some very important recombination pathways that involve extensive DNA replication, as in the original break-and-copy models.

RDR Pathways in Bacteriophage

Most of the early data arguing for the existence of RDR came from studies of bacterial viruses. For example, recombination-deficient mutants of well-known bacteriophage such as T4, T7, and lambda were found to be partially defective in DNA replication. The interpretation of that replication defect was not straightforward, because recombination reactions can play indirect roles in viral DNA replication. For example, if recombination links multiple bacteriophage genomes together into a linear concatamer, a single initiation event from a conventional replication origin can result in much more DNA

replication than if the genomes are not linked. Furthermore, recombination proteins might play a role in converting a theta-form replication intermediate into a rolling circle, resulting in numerous replication products rather than just two.

Studies with bacteriophage T4, however, demonstrated that the involvement of homologous recombination proteins in DNA replication is very direct. At early times of infection, T4 DNA replication initiates at replication origins by a mechanism that involves a persistent RNA–DNA hybrid (R-loop), but this origin-dependent replication is turned off after several minutes. For the remainder of the infection cycle, all DNA replication initiates by an RDR mechanism in which a fully functional replication complex is assembled on a displacement loop (D-loop) (Figure 1). The D-loop is formed when the 3' single-stranded end of one phage DNA molecule invades a homologous region of another phage DNA molecule, or the other end of the same molecule (since T4 DNA is terminally redundant).

D-loop formation depends on the phage-encoded UvsX and UvsY proteins, acting in concert with the T4 single-stranded DNA binding protein gene product 32 (gp32). UvsX is a strand invasion protein, homologous to bacterial RecA and eukaryotic Rad51 protein, while UvsY plays the key role of loading UvsX onto gp32-coated single-stranded DNA (i.e., the 3' end).

Once a D-loop is formed, the phage protein gp59 is the critical component that directs the replication machinery to the D-loop. The gp59 protein preferentially binds branched DNA including D-loops and thereby stimulates the loading of the replicative helicase (gp41) onto D-loop substrates during *in vitro* replication reactions. Once the replicative helicase is loaded, more extensive template unwinding, loading of additional replication proteins, and extensive DNA replication becomes possible. Proteins that allow the loading of replicative helicases onto recombination intermediates are central to

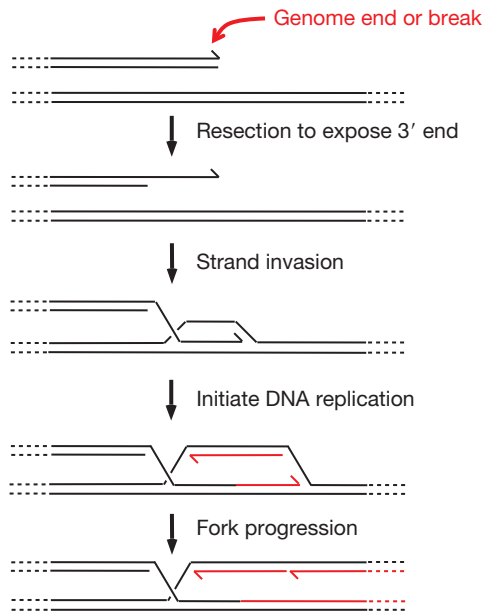


Figure 1 Model for bacteriophage T4 RDR. A single-stranded 3' end is generated from a genomic end or break in the first step, either by exonuclease resection or by prior DNA replication up to the end (not shown in the figure). The strand invasion step creates a D-loop, upon which replication is initiated after assembly of the replication complex. The dotted lines indicate extended regions of the phage chromosome, and all newly replicated DNA is in red.

the process of RDR, and the corresponding bacterial proteins will be discussed below.

There is now quite convincing experimental evidence for the T4 RDR pathway. A variety of physical and genetic experiments support the proposal that genome ends trigger DNA replication after a strand-invasion reaction. Furthermore, artificially introduced double-strand breaks (DSBs) dramatically stimulate the replication of homologous DNA *in vivo*. Finally, a robust *in vitro* system of RDR has been established using the T4 system and recapitulates the major features of *in vivo* RDR.

RDR in *Escherichia coli* and Other Bacteria

An apparent cellular RDR pathway, called inducible stable DNA replication (iSDR), was first uncovered by the late Tokio Kogoma and colleagues. They found that *Escherichia coli* cells deficient in normal origin-directed replication could nonetheless replicate if their SOS system for DNA damage response was activated. This DNA replication required the strand invasion protein RecA, and a variety of results led to the proposal that iSDR also required chromosomal DNA breaks induced by some component of the SOS system. The simplest model is that the induced DNA breaks lead to invasive 3' ends, and that resulting D-loops become the sites of assembly of new replication complexes (much like in Figure 1). Experiments with artificially induced DSBs have directly shown that recombinational repair of breaks can indeed trigger DNA replication in *E. coli*, providing compelling support for bacterial RDR.

A key protein for *E. coli* RDR is PriA, which, like T4 gp59, allows the loading of the replicative helicase (DnaB in this case) onto D-loop structures. *In vitro* experiments have directly demonstrated PriA-dependent RDR on D-loops generated by the bacterial RecA and RecBCD proteins, using the normal replicative DNA polymerase III holoenzyme complex. Several other *E. coli* proteins, including DnaC, PriB, and PriC, also participate in PriA-dependent RDR reactions. The importance of PriA in the survival of *E. coli* will be discussed later, when we consider the role of RDR in completing genome replication. It is interesting to note at this point, however, that mutational inactivation of PriA dramatically reduces phage P1-mediated generalized transduction and also conjugal recombination. Therefore, these seemingly well-studied genetic pathways, which had been thought for decades to occur by a break-and-rejoin-like mechanism, may actually occur by a mechanism similar to RDR.

Homologs of the *E. coli* PriA protein are found in most bacterial species, indicating widespread occurrence of RDR in bacteria. The best-studied PriA homolog is from *Bacillus subtilis*, and like the *E. coli* protein, *B. subtilis* PriA plays a major role in loading the replicative helicase. Interestingly, the extremely radioresistant bacterial species *Deinococcus radiodurans* has no recognizable PriA homolog (but could have a functionally equivalent protein). Nonetheless, *D. radiodurans* can reconstitute its genome after very extensive levels of DNA breakage by a repetitive process that uses recombination to prime DNA synthesis, sequentially building up longer and longer segments of the genome from the broken pieces.

RDR in Eukaryotic Cells

RDR has also been convincingly demonstrated in eukaryotes, with the most direct studies in yeast model systems. Several different genetic approaches provided strong evidence that RDR is a legitimate pathway of DNA break repair in the yeast *Saccharomyces cerevisiae*. One approach used diploid yeast cells with a site that can be cleaved to create a DSB on one of the two copies of a particular chromosome, and the arms of this chromosome were marked with a series of heteroalleles (i.e., the two chromosomes had different genetic markers). Most DSBs were repaired by a localized event that did not alter the genetic markers on the remainder of the chromosome arm. At a low frequency, however, the resulting diploid cell experienced an extensive gene conversion event in which all alleles in the arm distal to the DSB had been converted to the form that was originally located on the unbroken (homologous) chromosome. Apparently, the distal chromosomal fragment from the broken chromosome had been lost and replaced with a newly replicated copy of the distal arm from the unbroken chromosome. When the experimental setup was altered so that the distal fragment of the broken chromosome had no homology to the intact chromosome, wild-type cells efficiently utilized this RDR pathway to regenerate an intact chromosome. The only general mechanism to explain this extensive gene conversion of a large portion of a chromosome is that the centromere-proximal chromosomal end undergoes a strand-invasion reaction with the homologous chromosome, and the resulting D-loop triggers a new replication fork just as in the RDR pathways.

Another informative genetic approach that reveals RDR in yeast involves the transformation of linear fragments of DNA with selectable markers and ends that are homologous to yeast chromosomes. In this case, a new chromosome can be established from the transforming fragment when one of the DNA ends invades the homologous chromosome and triggers an extensive RDR reaction that copies the homologous chromosome right to the telomeric end. In some experiments, both ends of the transforming fragment trigger RDR reactions, and in other experiments, one of the two ends is engineered to generate a new telomere upon transformation into the yeast cells (and so only one end needs to be repaired by RDR).

In addition to these compelling genetic approaches, direct physical assays have also been used to demonstrate RDR reactions in yeast and study their detailed characteristics. Furthermore, protein requirements for yeast RDR have been uncovered using one or more of the above assays. As might be expected, all three major replicative DNA polymerases are involved, as are the replicative helicase (MCM complex) and recombination proteins including RAD51, RAD52, RAD54, RAD55, and RAD57.

In vivo studies of RDR in cells from metazoans are more difficult than in the simpler model systems described above, but evidence is mounting that RDR is also important in mammals and other multicellular eukaryotes. In *Drosophila*, an RDR-related pathway called synthesis-dependent strand annealing (SDSA) is able to replicate extensive chromosomal regions (albeit with an unusual replication mechanism). Also, recent studies with chicken cells showed a provocative requirement for DNA polymerase η during immunoglobulin gene rearrangement by homologous recombination. In parallel, another group demonstrated that polymerase η can catalyze DNA replication from D-loops formed by human strand invasion proteins, arguing that η could be a key factor in triggering RDR from D-loops. As will be described in more detail below, genetic studies of mammalian cells deficient in certain recombination proteins lends support to the proposition that RDR is an important pathway for maintaining genome stability in higher eukaryotes, and RDR-related pathways likely contribute to chromosomal rearrangements (e.g., in the generation of cancer). Furthermore, biochemical studies with *Xenopus* oocyte extracts provide strong evidence that RDR provides a pathway for restarting stalled or aborted replication forks in that system (see below).

A special role for RDR has also emerged in the field of telomere biology, with important studies in both yeast and human systems. Human somatic cells generally lack telomerase and thereby acquire shorter and shorter telomeres with each cell doubling, until a crisis is reached. In order to propagate further, most human tumor cells reactivate telomerase to allow complete chromosome replication. However, a subset does not reactivate telomerase, and yet show strong telomere DNA replication – these are called ALT, for alternative lengthening of telomeres (ALT cells can also be obtained during immortalization of cell lines *in vitro*). Telomere replication in these ALT cells is dependent on recombination proteins and must occur by some form of RDR, although the detailed mechanism is not known. Likewise, prolonged growth of telomerase-negative yeast cells leads to progressive telomere shortening and a crisis period with short telomeres.

Many such cells rebound from the crisis by activating one or another telomere replication pathway that depends on recombination proteins. These and other findings have led to the suggestion that RDR may have been the primordial eukaryotic telomere replication pathway, with telomerase evolving later.

RDR as a Backup Mechanism to Complete DNA Replication

Nearly 30 years ago, Ann-Marie Skalka proposed that recombination could provide a mechanism for creating new replication forks after a previous fork encountered a template nick. Her proposal dealt with the complexities of phage lambda DNA replication, and was not ascribed any broader significance for some time. However, many years later, scientists realized that a similar mechanism provides a means of restarting a replication fork that has been destroyed by a template nick. As diagrammed in Figure 2, when a replication fork encounters a

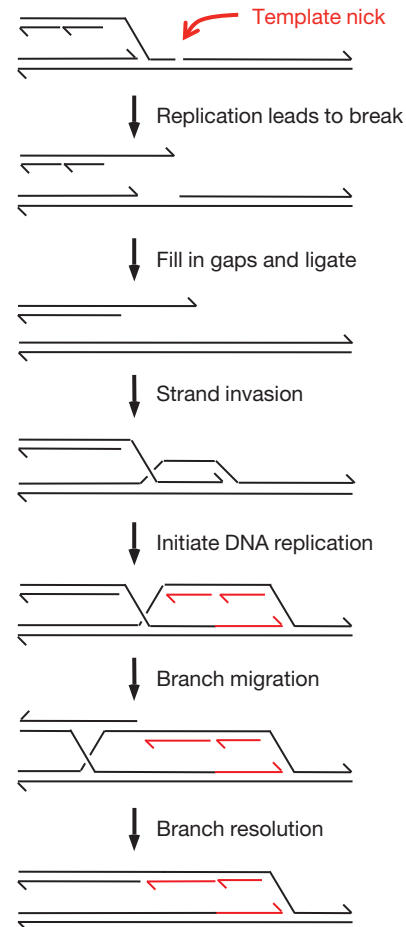


Figure 2 Replication fork breakage at a nick and restart by recombination. A template nick leads to a broken fork, which is then restarted by RDR. Note that one DNA strand has been flipped to the top in the molecule immediately after the arrow indicating branch migration (because DNA is a double helix but the drawing depicts only parallel lines, this flipping is just a convenience to make the drawing simpler).

template nick, one of the two arms can be broken off and replication thereby terminated. The broken end can be reconfigured into a replication fork by first forming a D-loop using a strand-invasion reaction, and then assembling a replication complex onto the D-loop. Subsequent branch migration and junction resolution can thereby regenerate a simple replication fork and allow the completion of replication and normal segregation of the products at the next cell division (Figure 2).

We now believe that this pathway explains why many recombination-deficient mutants of bacteria grow poorly – they have great difficulty completing DNA replication due to problems at the fork. Notably, PriA-deficient mutants of *E. coli*, whose primary defect is in replication fork restart, are nearly inviable and quickly accumulate suppressor mutations that improve their viability.

In vivo experiments have demonstrated that replication forks are indeed broken when they encounter a template nick. It is not clear how often replication forks encounter template nicks, but broken replication forks also arise by another pathway that could be even more common. In bacterial and phage systems, it has been demonstrated that certain stalled or blocked replication forks are prone to breakage. This breakage can be induced by the branched-DNA-specific nucleases that normally resolve Holliday junctions (i.e., *E. coli* RuvC and T4 endonuclease VII). Results in eukaryotic systems also indicate that some blocked forks are broken in a similar manner.

It may seem strange that blocked replication forks would be actively cleaved by recombination nucleases, and this is indeed a dangerous reaction. However, this cleavage is apparently under strict control, and may only be used as a last resort when a stalled or blocked replication fork cannot be restarted by other mechanisms (see section [Replication Fork Failure and Direct Restart Pathways](#)). By coupling fork cleavage with an RDR reaction, the replication fork can potentially be restarted and the cell saved from the dire consequences that result from trying to segregate a partially replicated DNA molecule in the next cell division.

Xenopus oocyte extracts provide a powerful and convenient system to study many aspects of cellular biology, including DNA replication. Strikingly, when such extracts are depleted of the Mre11/Rad50/Nbs1 (MRN) complex, which is involved in homologous recombination and damage signaling, broken replication forks are generated and complete replication is prevented. Thus, in the normal extracts, the MRN complex is presumably contributing to restart of stalled and/or broken replication forks by an RDR pathway.

With this perspective that recombination provides a backup mechanism to restart broken replication forks, it is worth considering an interesting question from evolutionary biology – how and why did homologous recombination evolve? It is now clear that homologous recombination can play an important role at the level of single cell survival, namely the completion of DNA replication. Furthermore, DNA replication was likely to be even more prone to failure early in the evolution of cells, before very sophisticated replication machineries and repair pathways had been fully developed. Therefore, many scientists in the field now believe that

homologous recombination first evolved as a means to complete cellular DNA replication, and that all the other wonderful benefits of recombination came later.

Replication Fork Failure and Direct Restart Pathways

The above description of replication fork restart is incomplete, focusing only on the restart of forks that end up being broken. As might be expected, other pathways are able to directly restart stalled or blocked replication forks without the dangerous step of breaking the fork. Indeed, bacterial cells have multiple pathways for directly restarting forks. Although these pathways need not involve any recombination of the DNA molecules, some of the proteins that are involved in direct restart pathways are nonetheless recombination proteins that play roles in recombinational restart or other homologous recombination pathways.

One of the important unanswered questions is the fate of the various components of the replication machinery when forks are blocked or stall. It is clear that replication fork restart pathways require proteins that can load the replicative helicase, and so by inference, the replicative helicase seems to be lost when forks stop prematurely. We do not know which other components of the replication machinery are lost upon fork failure. Because the replicative helicase is thought to encircle the lagging-strand template, it is presumably lost whenever that strand is broken. In other cases (e.g., breakage of the leading strand and fork arrest without breakage), it remains to be determined how exactly the helicase is lost, and for that matter, whether the helicase is lost in only a subset of replication fork failures.

There are many possible reasons why a replication fork might stop prematurely. For example, a particular nucleotide might be temporarily limiting, a noncovalently bound protein might be encountered on the template duplex, or one of the key proteins of the replication machinery might become damaged or displaced from the replication fork. By sidestepping the question of how and why the replicative helicase and other proteins might be lost from such a fork, it is easy to see how direct restart pathways can solve the immediate problem and allow a completion of DNA replication.

However, the true complexities and benefits of replication fork restart pathways only become obvious when we consider other problems that a replication fork can encounter, particularly the many different forms of template DNA damage. The most obvious problem is when the replication machinery encounters a damaged base on the leading-strand template. In this case, one can easily see the advantage of disassembling the normal replication machinery – namely, that an appropriate DNA repair complex can be recruited to fix the template damage or, if that fails, a translesion DNA polymerase can be recruited to incorporate something opposite the damaged base (Figure 3; left). After the repair or translesion synthesis event, the normal replication machinery would need to be assembled to resume replication, hence the need for a direct restart pathway (or an RDR event if the fork has become broken). This description of the behavior of the replication fork at sites of template DNA damage is very superficial and incomplete,

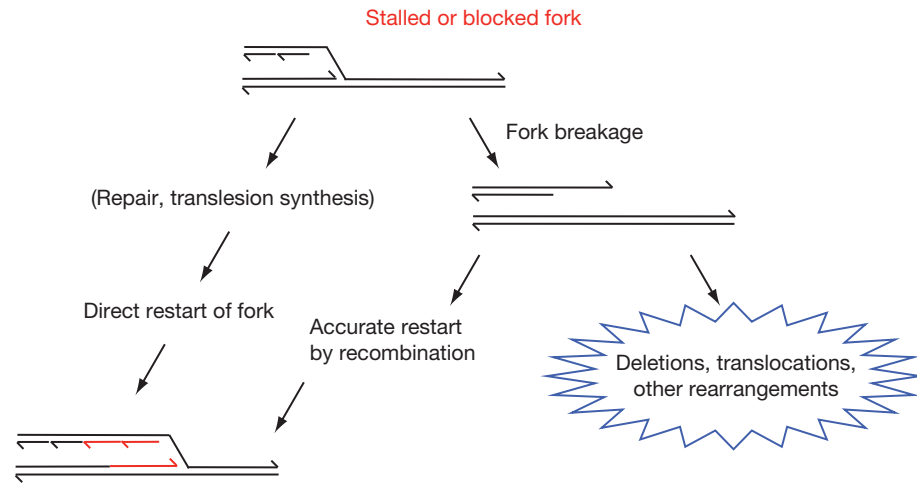


Figure 3 Pathways for restarting replication forks. Replication forks may stall due to the absence of nucleotides or problems with the replication machinery, or may be blocked due to template damage or bound proteins ahead of the fork. If the fork was blocked by template damage, repair of the damage or translesion synthesis may precede direct fork restart. Some of these pathways involve a process called replication fork regression, where the two newly synthesized strands unwind from their parental templates and rewind with each other, creating a Holliday junction at the fork (this process is not shown in the figure). If the fork was simply stalled, direct fork restart can occur without repair. In either case, some or all of the replication proteins need to be re-assembled on the fork to re-initiate replication. In the pathway on the right, the stalled or blocked fork is broken, for example, by a recombination nuclease. Accurate restart of the fork by RDR (as in [Figure 2](#)) can lead to a normal cell without any genetic damage, but failure of this accurate pathway can lead to chromosomal rearrangements.

in part because there are many forms of damage and multiple pathways of repair. The more serious limitation at this time, however, is that the field is just beginning to investigate what promises to be a very rich area, this interface where DNA replication, repair and recombination all intersect at troubled replication forks.

Disturbances in Fork Processing Lead to Genome Instability and Cancer Predisposition

DNA breaks are inherently dangerous. A broken fragment of a chromosome that lacks the centromere may be lost during mitosis to generate a terminal chromosomal deletion (assuming that the centromere-containing fragment acquires a new telomere). A broken end might engage in a nonhomologous end-joining reaction, or if the end is located within a repeat region, might undergo homologous recombination with another copy of the repeat. In both cases, chromosomal deletions, inversions, and translocations can result.

With an understanding that DNA replication can lead to DNA breaks, it is therefore not surprising to find a strong association between disturbances in replication fork processing and genome instability. For example, studies in *S. cerevisiae* have shown that mutational inactivation of genes involved in the recombinational restart of replication forks (RDR) leads to genome instability. Apparently, when the normal pathway of restarting broken forks is disturbed, the broken DNA can result in chromosomal rearrangements ([Figure 3](#), right).

Cells from individuals who suffer from certain human cancer predisposition syndromes (Werner's, Bloom's, and Rothmund-Thomson syndrome) show increased genome instability. Genetic analysis has traced these defects to genes

that encode proteins that are suspected to be involved in replication fork integrity, recombinational restart of broken forks, and/or direct restart pathways. All three of these proteins belong to the RecQ family of DNA helicases (named after the founding member, bacterial RecQ). Several other human cancer predisposition syndromes have been shown to have defects in other recombination or DNA damage response genes. The precise role of these various proteins are under active investigation, and it is likely that at least in some cases, genome instability results from a defect in recombinational restart of broken forks or from a higher incidence of replication fork breakage. Increased fork breakage could potentially arise from a defect in a protein that is involved in direct restart pathways, or from any mutation that leads to more frequent fork arrest.

Current models for the generation of chromosomal rearrangements, including translocations, inversions, and gene amplifications, often invoke reactions including RDR and/or replication template strand switching. Particularly in the case of large gene amplifications, extensive DNA replication by an RDR-like mechanism seems very likely. Thus, RDR appears to play an important role in the genome instability that is critical in the generation and progression of cancer, and as described above, RDR is also important in telomere cancer biology.

In summary, the processes of DNA replication, recombination, and repair participate in an intricate choreography that is necessary for the accurate and complete replication of the genome, which in turn is important for genome stability and for the prevention of cancer in humans. RDR is a central aspect of this choreography, directly connecting homologous recombination to DNA replication and allowing the completion of replication when the replication fork becomes broken.

See also: **Cell Architecture and Function:** Cell Cycle: DNA Damage Checkpoints; Chromosome Organization and Structure, Overview; Mitosis; **Molecular Biology:** DNA Replication Fork, Bacterial; DNA Replication Fork, Eukaryotic; DNA Replication: Initiation in Bacteria; DNA Replication, Mitochondrial; Homologous Recombination in Meiosis; Recombination: DNA Helicases and Nucleases; Recombination: DNA-Strand Transferases; RecQ Helicase Systems; Telomeres: Maintenance and Replication.

Further Reading

- Cox MM, Goodman MF, Kreuzer KN, et al. (2000) The importance of repairing stalled replication forks. *Nature* 404: 37–41.
- Gabbai CB and Marians KJ (2010) Recruitment to stalled replication forks of the PriA DNA helicase and replisome-loading activities is essential for survival. *DNA Repair* 9: 202–209.
- Haber JE (1999) DNA recombination: The replication connection. *Trends in Biochemical Sciences* 24: 271–275.
- Hickson I (2003) RecQ helicases: Caretakers of the genome. *Nature Reviews Cancer* 3: 169–178.
- Kogoma T (1997) Stable DNA replication: Interplay between DNA replication, homologous recombination, and transcription. *Microbiology and Molecular Biology Reviews* 61: 212–238.
- Kowalczykowski SC and von Hippel PH (2000) Special issue: The DNA replication–recombination interface. *Trends in Biochemical Sciences* 25: 155–206.
- Kreuzer KN (2005) Interplay between DNA replication and recombination in prokaryotes. *Annual Review of Microbiology* 59: 43–67.
- Lydeard JR, Lipkin-Moore Z, Sheu Y-J, et al. (2010) Break-induced replication requires all essential DNA replication factors except those specific for pre-RC assembly. *Genes and Development* 24: 1133–1144.
- McEachern MJ and Haber JE (2006) Break-induced replication and recombinational telomere elongation in yeast. *Annual Review of Biochemistry* 75: 111–125.
- McGlynn P and Lloyd RG (2002) Recombinational repair and restart of damaged replication forks. *Nature Reviews* 3: 859–870.
- Michel B, Boubakri H, Baharoglu Z, LeMasson M, and Lestini R (2007) Recombination proteins and rescue of arrested replication forks. *DNA Repair* 6: 967–980.
- Radding C (2001) Links between recombination and replication: Vital roles of recombination. *Proceedings of the National Academy of Sciences of the United States of America* 98: 8172.
- San Filippo J, Sung P, and Klein H (2008) Mechanism of eukaryotic homologous recombination. *Annual Review of Biochemistry* 77: 229–257.
- Sonoda E, Sasaki MS, Buerstedde J-M, et al. (1998) Rad51-deficient vertebrate cells accumulate chromosomal breaks prior to cell death. *EMBO Journal* 17: 598–608.
- Xu L and Marians KJ (2003) PriA mediates DNA replication pathway choice at recombination intermediates. *Molecular Cell* 11: 817–826.

Recombination: DNA Helicases and Nucleases

J Sowa and G Ira, Baylor College of Medicine, Houston, TX, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Crossing-over A product of homologous recombination in which the arms of recombining molecules were reciprocally exchanged. In meiosis crossing-over constitutes 20–50% of the recombination products and occurs mostly between homologous chromosomes while in growing cells crossing-over represents only <1–10% of the products and recombination involves mainly sister chromatids.

Heteroduplex DNA Double-stranded DNA created during recombination from two complementary strands derived

from two DNA molecules. Strand-exchange proteins mediate heteroduplex DNA formation. The length of heteroduplex DNA ranges from few hundred bps to several kbs.

Holliday junction (HJ) Junction of four strands formed between two homologous DNA molecules in recombination. Double and single HJs were demonstrated as intermediates of meiotic recombination. HJs are cleaved by nucleases called resolvases.

Introduction

There are many sources of the DNA damage which is repaired by recombination. Single-strand gaps and breaks and double-strand breaks (DSBs) occur spontaneously during replication or are induced by radiation or radiomimetic chemicals. Some types of cells induce programmed DNA breaks to initiate meiotic recombination, V(D)J recombination, or class switching in the immune system. Additionally, some lower organisms such as baker's yeast induce single DSBs to initiate mating-type switching. In higher eukarya DNA recombination is essential for life, indicating that in each cell cycle there is spontaneous damage that needs to be repaired by DNA recombination. DNA recombination generates new combinations of alleles and is necessary for chromosome segregation during meiosis. Both the mechanisms and proteins involved in recombination are well conserved among all organisms.

Models of DNA Recombination

The DNA recombination events which are best understood at the molecular level are those initiated by DSBs. In mitotic cells DNA breaks can be induced by site-specific nucleases, such as HO endonuclease in yeast or I-SceI in mammals, and in meiotic cells by type II topoisomerase-related protein Spo11. Several models summarized in Figure 1 have been proposed to explain DSB-induced homologous recombination at the molecular level. All recombination events are initiated by formation of 3' single-stranded DNA (ssDNA) at DSB ends. This process is called 5'-strand resection. In the classic DSB repair model both ends invade the homologous donor sequence and prime new DNA synthesis to form a double Holliday junction intermediate (dHJ) (Figure 1(a), left). This dHJ is then resolved to form either a crossover or a noncrossover product. dHJs have been demonstrated to be the major intermediate of recombination in meiotic cells where crossover products are very frequent, while in mitotic cells both crossovers and dHJs are rare. A dHJ can be processed into

noncrossover products by type I topoisomerase (Figure 1(a), middle); however, crossover frequency in the absence of type I topoisomerases increases only slightly, suggesting that this is not a major noncrossover pathway. An alternative model of recombination called synthesis-dependent strand annealing (SDSA) has been proposed to account for low crossover in mitotic cells. In this model only one end invades its template to prime new DNA synthesis. The newly synthesized strand is then displaced from the template molecule and anneals to the second end of a DSB to form noncrossover product exclusively (Figure 1(a), right). In this case, both newly synthesized strands are inherited by the molecule that had been broken, as has been demonstrated for noncrossover recombination in mitotic cells in yeast. All these models of recombination explain repair of two-ended DSBs. However, it is likely that most of the breaks created spontaneously in cells during replication are single ended as shown in Figure 1(b). A single-ended DSB is formed when a replication fork runs into a DNA nick or a gap. These are repaired by break-induced replication (BIR). In BIR strand invasion is followed by DNA synthesis (both lagging and leading strands) to the next replication fork or to the end of the chromosome. This process requires most of the proteins involved in DNA synthesis during regular replication (Figure 1(b)). Finally, when a DSB is flanked by direct repeats as shown in Figure 1(c), upon resection the two repeats can anneal to each other and the sequence between the repeats and one repeat is lost. This type of repair is called single-strand annealing (SSA) and does not require a strand invasion step. Some rare recombination events or recombination in some organisms have distinct features not described by these models. For example, in fission yeast, single HJs (sHJs) are major intermediates of meiotic crossover recombination. Other examples include recombination pathways called microhomology-mediated end joining or microhomology-mediated BIR that are initiated by annealing of very short 2–15 nt sequences and lead to genome rearrangements frequently observed in cancer cells. These are rare but important pathways that are uncharacterized genetically and will not be discussed here.

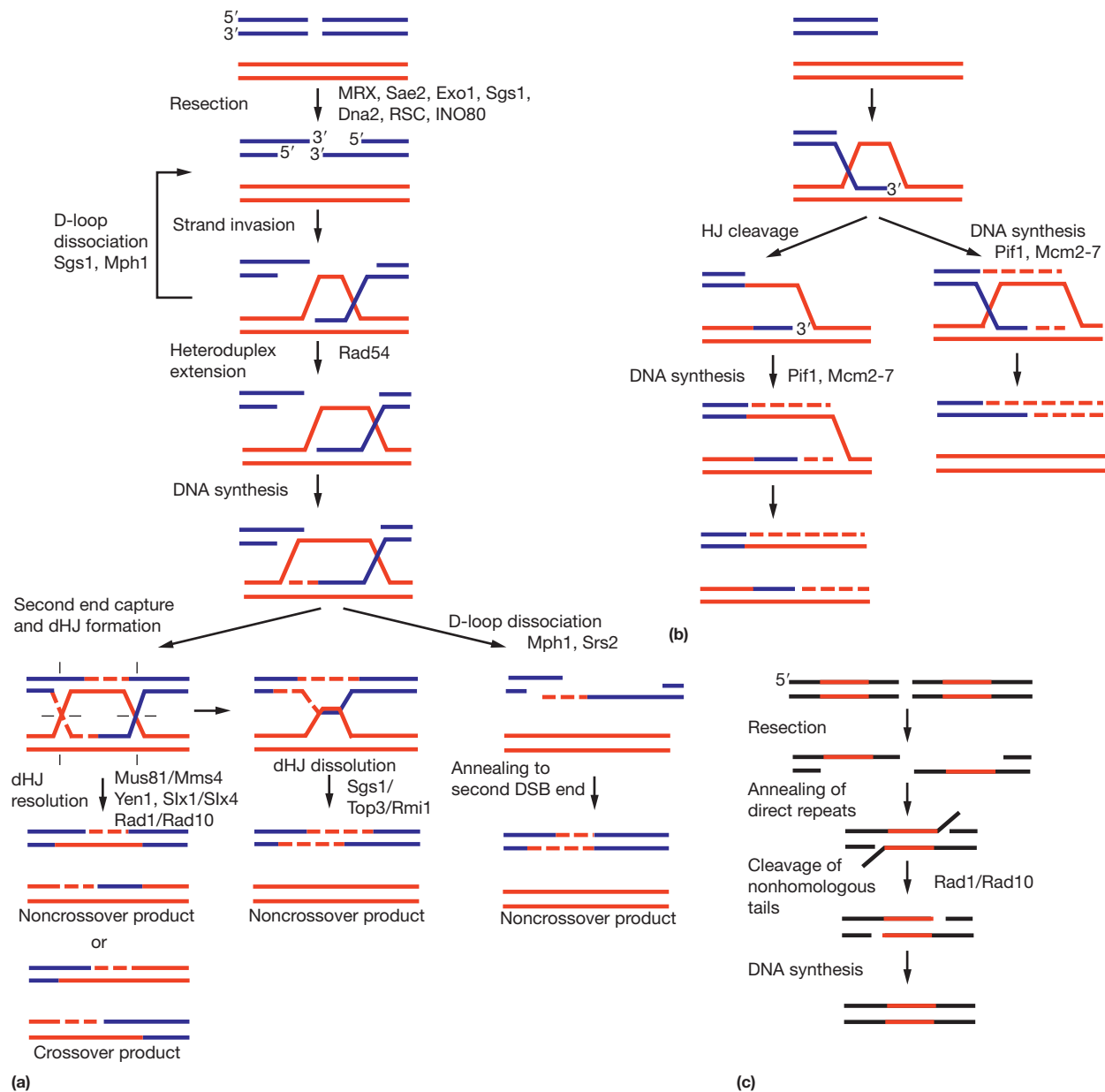


Figure 1 Models of DNA recombination. (a) Models of two-ended DSB repair. In all models 5'-strands are resected and 3' ends invade a homologous molecule. In the dHJ model (left), both ends invade the homologous molecule and prime new DNA synthesis leading to the dHJ formation and resolution. Both crossovers and noncrossovers can be formed as outcomes of the resolvase-mediated cleavage indicated with black lines. In the dHJ dissolution model (middle), two HJs migrate toward each other forming a single catenane that is processed by type I topoisomerase to form only noncrossover product. In SDSA model (right), only one end primes new DNA synthesis and the newly synthesized DNA is dissociated from template and anneals to the second DSB end forming exclusively noncrossover product. DNA translocases, helicases, and nucleases working in each step of repair are indicated. Only yeast enzymes are shown; human homologs and functional homologs are shown in [Table 1](#). (b) Break-induced replication model for repair of one-ended DSBs. DNA helicases important specifically for this process are shown. (c) Single-strand annealing model for repair of a DNA break flanked by direct repeats. The Rad1 nuclease cleaves nonhomologous tails after annealing of direct repeats.

DNA Helicases

As shown in [Figure 1](#), recombination involves multiple strand-unwinding and DNA cleavage steps. DNA helicases and nucleases mediate these processes and are listed in [Table 1](#). DNA helicases are the enzymes that couple adenosine triphosphate (ATP) hydrolysis to unwinding of double-stranded DNA

(dsDNA). Helicases belong to a larger group of proteins called translocases that translocate on DNA. Many translocases cannot unwind duplex DNA but still play an important role in DNA repair. Translocases are classified into many groups based on their signature motifs. All of them contain motifs responsible for the binding and hydrolysis of ATP, and an arginine finger that presumably senses ATP binding and hydrolysis and

Table 1 Enzymes with translocase or nuclease activity in homologous recombination in yeast and human

Step in DNA recombination	Yeast	Human	Activity
5'-Strand resection	Mre11–Rad50–Xrs2	MRE11–RAD50–NBS1	Mre11 is 3'-exonuclease and ssDNA endonuclease
	Sae2	CtIP	Sae2 ssDNA nuclease
	Exo1	EXO1	5'-Exonuclease
	Sgs1	BLM	3'-5' Helicase
	Dna2	DNA2	5'-Exonuclease, weak 3'-exonuclease
	INO80 complex	INO80 complex	Snf2 family translocase
	RSC complex	RSC complex	SWI/SNF family translocase
Homologous pairing and strand exchange	Rad54	RAD54	Snf2 family dsDNA translocase
Removal of nonhomologous tails	Rad1–Rad10–Saw1	ERCC1–XPF	Rad1 is a 3'-endonuclease
Heteroduplex extension	Rad54	RAD54	Translocase, branch migration
	Mer3	HFM1	3'-5' helicase
D-loop dissociation	Mph1	FANCM	Yeast-3'-5' helicase, human translocase
			3'-5' Helicase
	Sgs1	BLM	3'-5' Helicase
	Srs2 (?)	-	3'-5' Helicase
	-	RTEL-1	5'-3' Helicase
Resolution of Holliday junctions	Mus81–Mms4	MUS81–EME1	Cleaves nicked HJ's
	Yen1	Gen1	Cleaves HJ's
	Slx1–Slx4	SLX1–SLX4	Human complex cleaves HJs
	Rad1–Rad10	ERCC1–XPF	3'-Endonuclease

(?) indicates that there are conflicting observations.

signals for conformational changes essential for enzyme translocation. DNA helicases have either 3'–5' or 5'–3' polarity which means they translocate on the DNA strand they bind to in one or the other direction. Finally, DNA helicases may translocate on ssDNA or dsDNA. Additional domains within helicases confer binding specificity to different DNA structures or additional catalytic activities.

DNA Nucleases

DNA nucleases catalyze the cleavage of phosphodiester bonds between deoxyribose and the phosphate residue within the backbone of DNA strands. Nucleases that cleave near the ends of DNA molecules are called exonucleases while enzymes that cleave within a DNA strand and do not require a free DNA end for cleavage are called endonucleases. Some enzymes have both activities. As with helicases, exonucleases have either 3'–5' or 5'–3' polarity depending on the direction of DNA hydrolysis. Multiple nucleases play key roles in recombination, generating substrates and cleaving intermediates.

DSB End Resection: Helicases and Nucleases Generate ssDNA Substrate for DNA Recombination

The ends of DSBs need to be converted into ssDNA for recombination proteins to be loaded and for the DNA damage checkpoint to be activated. An average of 1–4 kb of ssDNA is generated on each side of a DSB during repair. It is then covered with strand-exchange protein Rad51, which is capable of genome-wide searching for homologous sequence. Checkpoint response proteins Mec1/Ddc2 in yeast and ATR/ATRIP in human bind ssDNA covered with ssDNA-binding proteins

RPA, and activate a kinase cascade that leads to cell-cycle arrest. In yeast Mre11–Rad50–Xrs2 (MRX complex, Mre11–Rad50–Nbs1 or MRN complex in human) together with Sae2 (CtIP in human) initiate the process of resection by generating short lengths of ssDNA and recruiting other nucleases that can generate longer ssDNA. Mre11 is an ssDNA endonuclease and 3'-exonuclease with polarity opposite to the polarity of resection. One possible mode by which Mre11 could generate short 3'-ssDNA is by cleavage of the 5'-strand of duplex DNA at some distance from the DSB ends, followed by its 3'-exonuclease-dependent degradation. Mre11 nuclease activity is essential for resection when the ends of DSBs are modified (e.g., by covalently bound proteins such as Spo11 or by adducts) in all eukarya that have been tested. In human and fission yeast, nuclease activity of Mre11 is required for resection even at clean (unmodified) DSB ends. By contrast, in budding yeast a nuclease mutant of *MRE11* has only a mild defect, while complete *MRE11* deletion has dramatic defect in resection. Mre11 nuclease is stimulated by Sae2, which also shows ssDNA nuclease activity *in vitro*. MRX, besides generating ssDNA itself, can also recruit nucleases Exo1 and Dna2 (in budding yeast) and Exo1 (in human) to DSB ends. This activity is Mre11 nuclease independent in yeast. Finally, MRX complex antagonizes excessive DSB end binding by Ku proteins that block resection. Exo1 is a 5'-exonuclease and has been shown to play a key role in generating longer ssDNA at DSBs during meiotic recombination and in processing stalled replication forks or ultraviolet (UV) light damage. In budding yeast mitotic cells, in addition to Exo1 there is also Dna2, a 5'-exonuclease involved in Okazaki fragment processing, which can resect DSB ends. While Exo1 can process 5'-strands by itself, Dna2 needs Sgs1 DNA helicase and RPA for resection. RPA stimulates 5'-strand cleavage by Dna2 and protects 3'-ends from Dna2 nuclease. Sgs1 forms a complex with topoisomerase Top3 and Rmi1 and

entire complex (called in yeast STR complex) but not Top3 catalytic activity is needed for normal resection. It is not yet clear why two resection pathways exist and whether Exo1 and Dna2 act entirely independently. Similarly to what is found in yeast, in humans Exo1 and the Sgs1 ortholog BLM are needed for resection while in frogs another Sgs1 ortholog called WRN and DNA2 nuclease are used. In human there are five Sgs1 orthologs, three of which (BLM, WRN, and RECQ5) play important roles in DNA recombination. Besides helicases and nucleases directly processing 5'-strands, at least two additional translocases that remodel chromatin stimulate resection. Yeast RSC complex recruits MRX complex to DSBs while INO80 likely displaces nucleosomes to stimulate resection (Table 1).

In prokaryotes several DNA helicase/nuclease complexes that resect DSBs have been described. In *Escherichia coli* a RecBCD heterotrimeric complex resects DSBs, where both RecB and RecD have helicase activity while RecC is also a nuclease. Both unwound strands are initially degraded until the complex reaches an 8-bp long sequence (5'-GCTGGTGG-3') called Chi, which is highly enriched in the *E. coli* genome. From the Chi sequence only the 5'-strand is degraded, exposing a long 3'-tail used to form RecA nucleofilament. RecB also directly helps to load the RecA protein at 3'-ssDNA. Similarly to eukaryotes, in *E. coli* RecQ helicase, an ortholog of Sgs1, together with RecJ, a 5'-3' ssDNA exonuclease, forms an ssDNA substrate for recombination. This presumably occurs at DNA damage where RecBCD has no access, such as DNA strand gaps. A functional homolog of RecBCD in *Bacillus subtilis* is AddAB, and in *Mycobacterium smegmatis* is AdnAB. AddAB is a complex of two nucleases and one helicase, while mycobacterial AdnA and AdnB both harbor helicase and nuclease domains. It remains to be determined if and how AddAB or AdnAB is controlled by sequences equivalent to Chi and whether like RecBCD they help to load RecA.

DNA Translocases in Strand Invasion, Branch Migration, and DNA Synthesis

When 3'-ssDNA is formed, Rad51 protein covers it, forming a nucleofilament, and mediates the strand-exchange process. Rad54, a member of the Snf2 family, can translocate on DNA (but has no helicase activity) and stimulate the strand-exchange process. It has been suggested that translocase activity could stimulate heteroduplex DNA formation by opening duplex donor DNA to facilitate interaction with the Rad51 nucleofilament. *In vitro* evidence suggests that Rad54 translocase plays several roles in recombination after heteroduplex DNA is formed. First, Rad54 stimulates Rad51 pairing activity and therefore extends heteroduplex DNA. Second, in yeast Rad54 and its homolog Rdh54 displace Rad51 from heteroduplex DNA and presumably facilitate DNA synthesis primed by invading 3'-ends. Finally, Rad54 promotes 3- and 4-strand branch migration. This activity can extend or shorten heteroduplex DNA at the D-loop stage or after full HJs are formed (Figure 1(a)). DNA synthesis by Pol δ in yeast in the context of D-loop is stimulated by a 5'-3' DNA helicase called Pif1, while extensive DNA synthesis during BIR requires the replicative helicase Mcm2-7.

Removal of Terminal Nonhomology by Yeast Rad1/ (Rad10) Nuclease

When the ends of invading strand are nonhomologous, DNA polymerase cannot use the ends for extension until the terminal nonhomology is removed. In yeast very short terminal nonhomologous sequences (≤ 10 nt) can be removed by the proofreading 3'-5' exonuclease activity of polymerase δ . Longer nonhomologous tails require a structure-specific 3'-endonuclease called Rad1 and its associated protein Rad10 (XPF-ERCC1 in human), which also plays a key role in nucleotide excision repair (NER). Nonhomologous tails are formed during the SSA recombination pathway upon resection and annealing of two repeats (Figure 1(c)). Besides Rad1 nuclease, two mismatch repair proteins, Msh2 and Msh3, stimulate removal of nonhomologous tails presumably by stabilizing invading 3'-strands carrying terminal nonhomology. Saw1 is another accessory protein which has been shown to recruit Rad1/Rad10 specifically for nonhomologous tail removal, but plays no role in NER. In addition, Slx4, but not its endonuclease partner Slx1 (discussed below), is essential for Rad1-Rad10 cleavage. Slx4 needs to be phosphorylated by DNA damage checkpoint proteins (Mec1/Tel1) to participate in nonhomologous tail removal. However, the specific function of Slx4 in Rad1 cleavage is not known. The same Rad1/Rad10 nuclease is used for removal of long nonhomologous sequences that form large loops in heteroduplex DNA.

Multiple DNA Helicases and Nucleases Involved in Resolution of Recombination Intermediates

Once 3'-ends invade the homologous molecule, they prime new DNA synthesis. Based on genetic and biochemical evidence, it is assumed that the majority of noncrossover products in mitotic and even meiotic cells result from unwinding of newly synthesized 3'-strands (D-loop dissociation) as described in SDSA model of recombination (Figure 1(a), right). Several DNA helicases have been shown to catalyze D-loop dissociation. When both 3'-strands invade donor molecule and initiate synthesis, dHJs are formed that can be resolved into crossover products. In this case, nucleases called resolvases, which cleave 4-stranded junctions, are required to complete repair. Resolvases have been well characterized in prokaryotes and were recently identified in eukaryotes as well.

Enzymes Disrupting D-Loops

Several DNA helicases have been proposed to promote the noncrossover pathway via D-loop unwinding. In budding yeast, mutants deficient in Srs2 or Mph1 helicases show increased crossovers; additionally, double mutants have a more than additive effect, suggesting that Srs2 and Mph1 play at least partially overlapping roles in preventing crossover formation. Fission yeast homologs of these helicases called Srs2 and Fml1 also limit crossover recombination in mitotic cells. Consistent with this, purified Mph1 and Fml1 helicases unwind Rad51-made D-loops *in vitro*. In addition, human FANCM, BLM, WRN, and RTEL-1 have been shown to

dissociate D-loops *in vitro*. As mentioned above, yeast Rad54 can also promote 3-stranded branch migration such as those at D-loops, raising the possibility that it unwinds D-loops as well. BLM-, RTEL-1-, and FANCM-deficient cells show increased crossover frequency, while the role of Rad54 in crossover regulation has not been established in cells because it is an essential and multifunctional protein in recombination. Disruption of D-loops before 3'-invading strands prime new DNA synthesis would simply inhibit recombination, and, as discussed below, some of these helicases prevent recombination under certain circumstances. How these helicases are regulated either to stimulate SDSA or to inhibit recombination, or in other words to unwind D-loops before or after the 3'-invading strand initiates new DNA synthesis, is not known but may be related to the stability of invading ends and/or length of heteroduplex DNA.

Resolution of HJs

When both 3'-ends invade and prime new DNA synthesis, dHJs are formed that can be resolved to either crossover or noncrossover product by resolvases. HJ resolvases are nucleases that nick two strands symmetrically at or near to the junction point. The nicked molecules are then repaired by DNA ligation, resulting in either crossover or noncrossover products (Figure 1(a), left). In prokaryotes, RuvC resolves HJ by symmetrical cleavage. In eukaryotes that do not have orthologs of RuvC (except mitochondrial Cce1), several enzymes were identified which resolve HJs. The first complex, called MUS81-EME1 in human and Mus81-Mms4 in budding yeast, is related to Rad1/XPF. It preferentially cleaves nicked HJs while intact HJs are poor substrates and their cleavage is not symmetrical. The second nuclease capable of cleaving intact HJs, and does it symmetrically, is Yen1 in budding yeast and its ortholog GEN1 in human. These enzymes belong to the Rad2/XPG nuclease family that includes XPG nuclease, Exo1, and Okazaki fragment-processing enzyme Fen1. The third eukaryotic enzyme able to cleave HJs, SLX1, was identified in humans. Human SLX1 forms a heterodimer with SLX4 and cleaves HJs symmetrically, while their fungal orthologs Slx1-Slx4 cleave HJs very poorly and asymmetrically. Finally, in *Drosophila melanogaster* MEI9, an ortholog of XPF that also forms a complex with an SLX4 ortholog called MUS312 in flies plays a crucial role in meiotic crossing over presumably by resolving HJs. Similarly, in *Caenorhabditis elegans* SLX4 and XPF orthologs, but not MUS81 or GEN1 orthologs, play a crucial role in meiotic recombination and the crossover pathway, while the function of SLX1 orthologs has not yet been established. As described above, SLX4 serves as a platform interacting with multiple endonucleases, possibly coordinating cleavage of different structures arising during repair.

While biochemical properties of the eukaryotic resolvases have been described carefully, their contribution to resolution of HJs *in vivo* remains unclear, as most model organisms tested so far seem to have distinct major proteins resolving HJs. *D. melanogaster* and *C. elegans* use mostly XPF nuclease in crossover recombination. In fission yeast, which lack a Yen1/GEN1 homolog, elimination of mus81 dramatically decreases crossover outcomes while noncrossovers remain unaffected. By contrast, in budding yeast elimination of Mus81 or Yen1

has little to no effect on crossovers in both meiotic and mitotic recombination, while mus81 yen1 double mutants show a significant decrease in crossovers and severe sensitivity to DNA damage. In higher eukaryotes MUS81/EME1 seems to play an even less significant role, because mutants lacking these enzymes show no defect in the meiotic cycle or sensitivity to ionizing radiation. However, again double mutants lacking both MUS81 and GEN1 are very sensitive to DNA damage. Taken together, it seems that Mus81 and Yen1 in yeast and MUS81 and GEN1 in human play overlapping roles in HJ resolution. It remains to be determined what the function of SLX1 and XPF in generating crossovers is, particularly in the absence of Mus81 and Yen1/Gen1 in yeast and human. More research is required to understand specific functions of each resolvase in mitotic and meiotic recombination.

Dissolution of HJs by Helicase/Topoisomerase Complex

In BLM mutants in human and *sgs1Δ* in yeast, the number of crossovers, particularly spontaneous sister chromatid exchanges, is increased. Similarly to resection mediated by these helicases, crossover suppression requires type IA topoisomerase, Top3, and Rmi1 components of the STR complex. What distinguishes the STR complex function in resection from its function in processing of dHJs is the catalytic activity of Top3. Strand passage activity is not required for resection; however, it is essential for crossover suppression. A mechanism proposed for this anticrossover activity suggests that BLM/Sgs1 helicases branch-migrate two Holiday junctions toward each other, thus creating a single catenane that is further resolved by type I topoisomerase into two separate DNA molecules to create a noncrossover product (Figure 1(a), middle). This activity is called dHJ dissolution and has been demonstrated in *in vitro* studies. As discussed above, BLM also dissociates D-loops, suggesting that it may suppress crossovers both by dissolution of dHJs and by unwinding D-loops.

DNA Helicases Inhibiting DNA Recombination

DNA recombination is an essential pathway of DSB repair. However, its use must be tightly regulated to limit recombination between very short repeats or excessive involvement during replication or other DNA transactions with ssDNA intermediates. Several DNA helicases prevent undesired DNA recombination and they act at different stages of recombination. Interestingly, these are the same helicases that, under different circumstances, promote recombination. In yeast at least three DNA helicases, Srs2, Sgs1, and Mph1, have both pro- and anti-recombinogenic functions. In mutants deficient in Srs2 and Sgs1, and particularly both Sgs1 and Mph1, spontaneous recombination is dramatically increased. Sgs1 might be particularly adept at inhibiting recombination between short homologous or partially homologous (homeologous) sequences, presumably by unwinding unstable heteroduplex DNA. Also, Sgs1 may suppress spontaneous DNA damage from initiating recombination by stabilizing stressed replication forks. On the molecular level, DNA helicases suppress homologous recombination in at least two different ways. First, they strip strand-exchange protein from ssDNA, and second they

unwind the initial intermediates of DNA recombination, the D-loops. DNA helicases able to strip Rad51 from ssDNA have been identified in bacteria, yeast, and higher eukaryotes. Yeast Srs2 and its *E. coli* homolog UvrD efficiently disrupt the pre-synaptic Rad51 (*EcRecA*) nucleofilaments, and thus inhibit the strand-exchange processes. Translocase (but not helicase) activity of Srs2, as well as its Rad51-interacting domain, appear to be essential for this process. One of the major functions of Srs2 as an anti-recombinase is to inhibit recombination in the context of replication forks to favor Rad6/Rad18-dependent postreplication repair. When recombination is the only means of repair, such as at DSB ends, Srs2 activity is antagonized by proteins called Rad52 and Rad55/Rad57 that mediate Rad51 loading. How cells regulate the antagonizing activities of Srs2 and Rad51 assembly mediators at different DNA substrates is not known. In mammals, RECQ5 helicase has been shown to possess similar activity to Srs2. Mouse mutants deficient in RECQ5 consistently show increased recombination rates, longer lasting Rad51 nuclear damage foci, genomic instability, and increased cancer predisposition. Additionally, BLM helicase can strip Rad51 from ssDNA *in vitro*; however, unlike RECQ5, BLM also inhibits recombination by unwinding Rad51-generated D-loops. Given multiple activities of mammalian BLM and yeast Sgs1, it is likely that a combination of several activities is important to prevent unscheduled recombination and genome rearrangements. Besides BLM, metazoan RTEL-1 and yeast Mph1 also unwind Rad51-made D-loops, an activity likely contributing to their anti-recombinase activities. Finally, Rad51 and its homolog expressed in meiotic cycle, Dmc1, can bind double-stranded DNA. However, these complexes do not produce recombination but may stimulate genomic instability. Several DNA translocases including Rad54, Rdh54, and Ulp1 disassemble such dsDNA-bound Rad51.

Nucleases in NHEJ

In NHEJ, two ends of a DSB are joined to restore the original sequence, and in most cases nuclease activity is not required for such repair. However, when the DNA ends are modified by adducts produced by ionizing radiation, or when hairpins or overhangs are created at DSB ends such as during V(D)J recombination, nuclease becomes essential for processing the DSB ends to allow subsequent ligation. Artemis has been identified in vertebrates as a major nuclease processing alternative DSB ends. Consequently, Artemis-deficient cells are radiosensitive and not capable of V(D)J recombination. Artemis belongs to the SNM1 family of nucleases that includes Pso2 in yeast and three nucleases in vertebrates, SNM1A (ortholog of yeast Pso2), SNM1B (known also as Apollo), and SNM1C (Artemis).

Artemis alone has a 5′–3′ exonuclease activity and an ssDNA endonuclease activity; when phosphorylated by and bound to DNA PKcs, it also possesses a dsDNA endonuclease activity and targets the junction of ssDNA and dsDNA. DNA PKcs is an essential protein kinase for NHEJ in higher eukarya, and is activated by binding to DNA. Artemis/DNA PKcs can act on a variety of substrates, including hairpins, 3′ and 5′ overhangs, and loops, and it generates either blunt ends or short 3′-overhanging ends. Processing of such structures by Artemis is crucial for repair by NHEJ, and some reports suggest that it may play a role in a fraction of homologous recombination.

Conclusions

DNA helicases and nucleases are key enzymes in DNA recombination. The molecular function of many of them has been identified, and their biochemical activities described. Individual helicases and nucleases play multiple roles during recombination, and the same step of recombination can be catalyzed by different helicases or nucleases. Also, helicases often have both pro- and anti-recombinogenic activities. Therefore, the next challenge is to identify mechanisms for the control of DNA helicases and nucleases by posttranslational modifications or by protein–protein interactions.

See also: **Molecular Biology:** DNA Mismatch Repair and Homologous Recombination; Homologous Recombination in Meiosis; Nonhexameric SF1 DNA Helicases/Translocases; Recombination: DNA-Strand Transferases; RecQ Helicase Systems.

Further Reading

- Chu WK and Hickson ID (2009) RecQ helicases: Multifunctional genome caretakers. *Nature Reviews Cancer* 9: 644–654.
- Dillingham MS and Kowalczykowski SC (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiology and Molecular Biology Reviews* 72: 642–671.
- Haber JE (2008) Evolution of models of homologous recombination. In: Lankenau D-H (ed.) *Genome Dynamics and Stability*, vol. 3, pp. 1–64. Berlin: Springer.
- Lyndaker AM and Alani E (2009) A tale of tails: Insights into the coordination of 3′ end processing during homologous recombination. *Bioessays* 31: 315–321.
- Marti TM and Fleck O (2004) DNA repair nucleases. *Cellular and Molecular Life Sciences* 61: 336–354.
- Mimitou EP and Symington LS (2009) Nucleases and helicases take center stage in homologous recombination. *Trends in Biochemical Sciences* 34: 264–272.
- Singleton MR, Dillingham MS, and Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annual Review of Biochemistry* 76: 23–50.

Recombination: DNA-Strand Transferases

W D Wright and W-D Heyer, University of California, Davis, Davis, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Floyd R. Bryant, volume 3, pp. 649–653, © 2004, Elsevier Inc.

Glossary

DNA joint molecule Two formerly distinct DNA molecules that are joined by base-pairing interactions and/or the DNA-strand transferase filament. Unlike heteroduplex DNA, this term includes synapsed DNA structures whose strands have not yet intertwined (paranemes) or where the heteroduplex DNA has been repaired by mismatch repair.

DNA translocase A motor protein whose adenosine triphosphatase cycle is coupled to movement on DNA.

Heteroduplex DNA The duplex DNA product of DNA-strand exchange, where the two component single strands were originally part of two separate DNA molecules.

Paralog A type of homolog where genes are related by duplication within an organism's genome that led to their specialization of functions.

Homologous recombination (HR) is highly reliant on the DNA-strand transferases that catalyze the key reactions of homology search and DNA-strand exchange. The prototype for this class of enzymes is RecA protein from *Escherichia coli*. Other members include UvsX from T4 bacteriophage, RadA from Archaea, Rad51 from eukaryotes, and the meiosis-specific Dmc1 protein, found in most, but not all, eukaryotes. Defects in DNA-strand transferases lead to a severe reduction in HR associated with DNA damage sensitivity as well as impaired replication and meiotic chromosome metabolism.

Homology Search and DNA-Strand Invasion by DNA-Strand Transferases

DNA-strand transferases catalyze homologous pairing between single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA, or duplex DNA), followed by the exchange of one strand of the dsDNA with the incoming ssDNA. The product of the DNA-strand exchange process is heteroduplex DNA, in which two single strands of DNA that were formerly from distinct DNA molecules form duplex DNA, displacing the original complementary duplex strand in the region where strand exchange has occurred. This DNA-strand invasion reaction is fundamentally different from DNA-strand annealing, where two ssDNA regions come together to form duplex DNA, because it requires DNA-strand displacement and the original base pairs to be broken. The DNA substrates required for recombination are typically the 3'-overhanging ssDNA from a processed DNA double-strand break (DSB) or ssDNA gaps formed during replication fork stalling (Figure 1) or other processes. The fundamental goal of a recombination reaction is a template switch to provide a 3'-OH primer a suitable template for DNA synthesis. This synthesis allows the cell to reference and recover sequence from an intact copy of a genetic locus when the other has been damaged.

DNA-Strand Transferase Filament Formation

The catalytic form of DNA-strand transferases is a specialized, right-handed helical filament that forms on ssDNA with

approximately six monomers per turn. The first challenge of HR is to assemble these nucleoprotein filaments so the homology search can begin. *In vivo*, the ssDNA substrate for filament formation is coated with ssDNA-binding protein (bacterial SSB, phage T4 Gp32, or eukaryotic RPA) as soon as it is generated. Although the binding of successive DNA-strand transferase monomers to filament ends is cooperative, competition for ssDNA binding with the ssDNA-binding proteins limits the initiation of filament formation, a process termed nucleation. In the cell, nucleation of the first monomers and the replacement of ssDNA-binding protein by growing DNA-strand transferase filaments are assisted by additional proteins, termed recombination mediators (see below). The eukaryotic DNA-strand transferases Rad51 and Dmc1 have added reliance on accessory proteins, as they lack the strong binding preference for ssDNA over dsDNA exhibited by RecA and UvsX. Consequently, *in vivo*, Rad51 or Dmc1 require assistance to be targeted to ssDNA, and filaments that have formed on dsDNA must be dissociated. The evolutionary constraints that led to these differences between DNA-strand transferases from eubacteria and eukaryotes are not understood yet, but likely are adaptations to the more complex genomes and regulatory requirements of eukaryotic organisms. Generation of ssDNA, nucleation, and extension of DNA strand transferase filaments are required to form the presynaptic nucleoprotein structure that is active for recombination, the generation of which is the primary challenge of the early steps of HR. This process is generally termed presynapsis, and the presynaptic filament is necessary for all subsequent steps of HR.

Homology Search and Paranemic Joint Formation

The presynaptic filament on ssDNA searches for homology in the genome. The homology search process is poorly understood, but a key feature is the stretching of the bound DNA in the confines of the filament by 150% relative to B-form DNA. Presumably, homology search involves many rounds of association/dissociation between the ssDNA filament and duplex DNA to sample homology by Watson–Crick base pairing. Finally, a region of

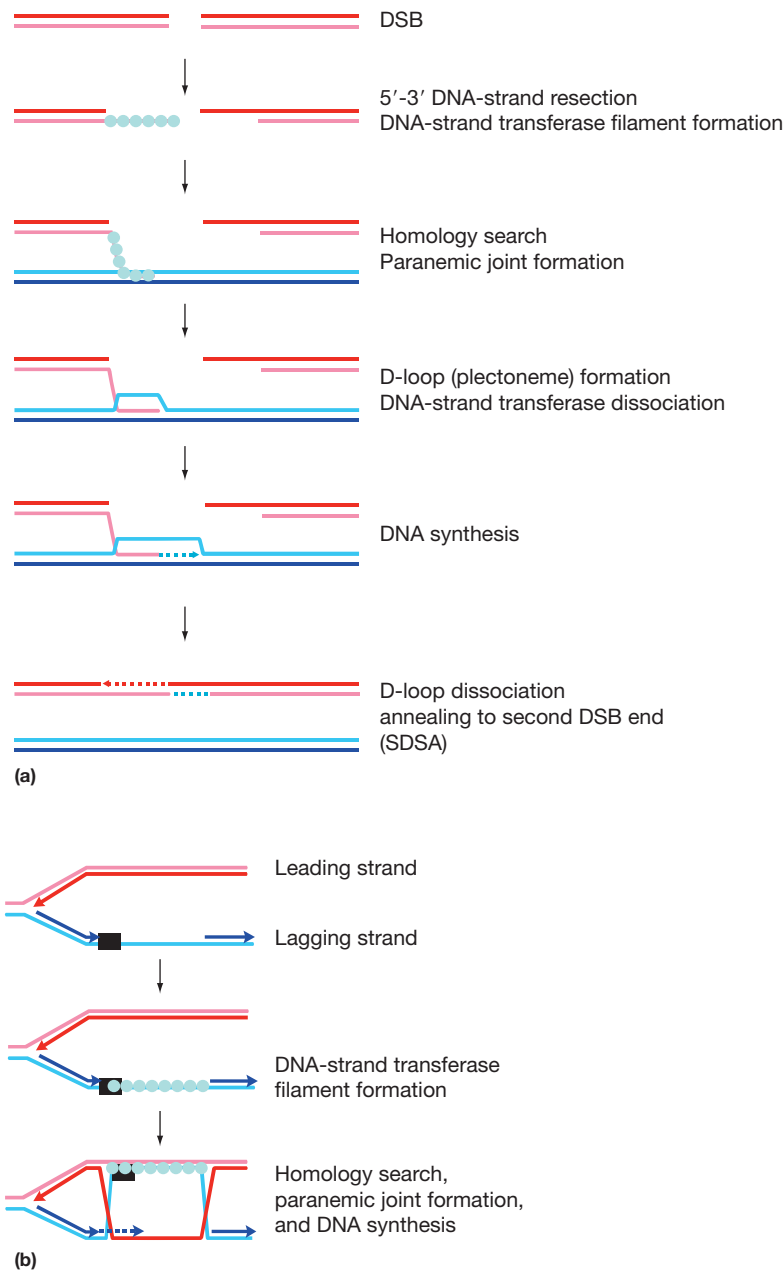


Figure 1 Involvement of DNA-strand transferases in DSB and gap repair by homologous recombination. (a) Double-strand break repair by synthesis-dependent strand annealing (SDSA). After DSB formation, nucleolytic DNA end-processing creates 3'-ended ssDNA that becomes the substrate for DNA strand transferase filament formation (spheres; not to scale or form). Homology search leads to the formation of a paranemic joint, which transitions to a plectonemic joint (D-loop) after strand intertwining. Upon dissociation of the DNA strand transferase from the heteroduplex product, the invading strand 3' end is extended by DNA polymerase. SDSA is one pathway of DSB repair, and alternative pathways differ only in the steps after DNA synthesis (see suggested readings). In SDSA, the extended invading ssDNA retracts from the D-loop and anneals to the complementary strand of the other DSB end. Synthesis to fill in the remaining gaps and sealing of the nicks by ligation results in a fully restored chromosome. (b) Gap repair by homologous recombination. Recombination from a gap on the lagging strand can displace the newly synthesized leading strand for use as a template for DNA synthesis past a blocking DNA lesion (black box). Notice that the initial joint molecule that forms is an obligatory paraneme because there is no free DNA end to allow strand intertwining. Reproduced with permission from Yu X, Jacobs SA, Westt SC, and Ogaw T (2001) Colloquium Paper: Domain structure and dynamics in the helical filaments formed by RecA and Rad51 on DNA. *Proceedings of the National Academy of Sciences of the United States of America* 98(15): 8419–8424. Copyright (2001) National Academy of Sciences, U.S.A.

DNA within the strand transferase-ssDNA filament is homologically aligned with duplex DNA, marking the beginning of the synapsis stage of HR. The initial synaptic complex is termed a paranemic (from Greek, literally 'beside thread') joint DNA

molecule, and it is stable only in the continued presence of the DNA-strand transferase protein. At this stage, there is no net intertwining of the incoming ssDNA with its complement that has been found in the dsDNA.

Plectonemic Joint Formation

After homology search and homologous alignment in the paranemic joint, the invading strand must intertwine with its complement in the duplex DNA, thereby displacing the original complementary strand in the duplex from base-pairing interactions. The resulting intermediate is known as a displacement loop (D-loop). The 3'-end must intertwine to generate a suitable primer/template junction that can be extended by a DNA polymerase. Once intertwined, the existence of the homologously paired joint DNA molecule is no longer dependent on the continued presence of the DNA-strand transferase filament, as the newly formed heteroduplex DNA stabilizes the joint DNA molecule. Such intermediates where strands from two homologous DNA molecules have intertwined are termed plectonemic (literally, 'knotted threads') joint DNA molecules. They have been experimentally shown to form after paranemic joints, their direct precursors.

The transition from a paranemic to plectonemic joint requires the presence of a free DNA end in either the incoming ssDNA or duplex DNA donor to allow the free rotation required for DNA-strand intertwining. Therefore, the initial joint molecule that forms from an ssDNA gap, where there is no free DNA end, is an obligatory paraneme. Conversely, the ssDNA substrate for DSB repair has a free 3'-end, such that paranemes are likely kinetic intermediates on the way to becoming plectonemic joint molecules.

Filament Dissociation

The continued presence of the DNA-strand transferase on the region of intertwined homology within the plectonemic joint (heteroduplex DNA) can actually inhibit further steps in HR, because it obstructs the 3'-ssDNA end from access by a DNA polymerase to initiate DNA repair synthesis. Therefore, the DNA-strand transferase must either self-dissociate or be removed by an external factor. Both types of turnover have been observed.

DNA-strand transferases are DNA-dependent ATPases, and the ATPase cycle modulates the affinity of the DNA-strand transferase for DNA. The adenosine triphosphate (ATP)-bound form has the highest affinity for DNA, and for RecA the adenosine diphosphate (ADP)-bound form has the lowest, even lower than the nucleotide-free form. ATP hydrolysis is thought to occur throughout the filament, but because of cooperative interactions between monomers within, dissociation occurs primarily from ADP-bound monomers at the ends. For DNA-strand transferases with higher ATPase activities such as RecA ($\sim 30 \text{ min}^{-1}$) and UvsX ($\sim 200 \text{ min}^{-1}$), this appears to be sufficient for autonomous turnover from heteroduplex DNA. However, eukaryotic Rad51 proteins have a relatively low ATPase activity (< 1 ATP hydrolyzed per minute), and their ADP-bound form has considerably higher affinity to DNA than RecA leading to very inefficient turnover from dsDNA. These biochemical differences between the RecA and Rad51 proteins explain the necessity for external turnover factors that dissociate Rad51 from dsDNA.

Restoration of Information and Chromosome Resolution

The fundamental purpose of homology search and DNA-strand invasion by the DNA-strand transferase filament is to position the 3'-end of a DSB onto a suitable template for repair DNA synthesis that ultimately leads to the restoration of chromosomal integrity (Figure 1(a)). Similarly in gap repair, DNA-strand exchange leads to a template switch that provides an intact template for a 3'-end that could not otherwise be extended due to a damaged template (Figure 1(b)).

Structure and Properties of DNA-Strand Transferase Filaments

General Properties of DNA-Strand Transferases

DNA-strand transferases share homology in their central core, called the RecA domain, that shows a considerable degree of

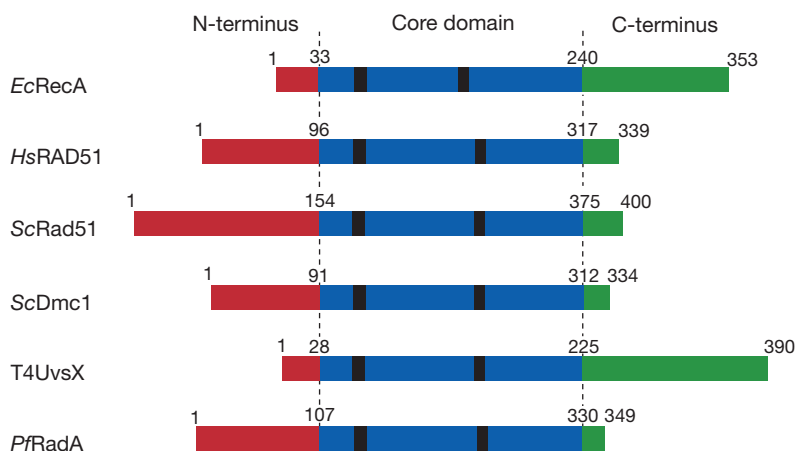


Figure 2 Schematic alignment of DNA-strand transferases from Eubacteria, Eukaryotes, Archaea, and T4 virus. Black boxes indicate the location of the conserved Walker A and B ATP-binding motifs within the core domain. Numbers denote amino acid residues. Sequence representations are approximately to scale. *Ec*, *Escherichia coli*; *Hs*, *Homo sapiens*; *Sc*, *Saccharomyces cerevisiae*; T4, T4 bacteriophage; *Pf*, *Pyrococcus furiosus*. Reproduced with permission from Yu X, Jacobs SA, Westt SC, and Ogaw T (2001) Colloquium Paper: Domain structure and dynamics in the helical filaments formed by RecA and Rad51 on DNA. *Proceedings of the National Academy of Sciences of the United States of America* 98(15): 8419–8424. Copyright (2001) National Academy of Sciences, U.S.A.

sequence conservation between species and contains the ATPase active site as well as residues involved in DNA binding (Figure 2). Amino and carboxy-terminal domains are also present which are more variable between species. RecA has an N-terminal domain (aa 1–33), followed by the core domain (aa 34–240), and a C-terminal domain which contains a DNA-binding site (aa 241–352). In comparison, Rad51 proteins have an extended N-terminal region containing a DNA-binding site and lack the C-terminal domain found in RecA.

ATP Hydrolysis within the Filament

ATP hydrolysis is coupled to changes in the conformation of the DNA-strand transferase filament. X-ray crystallography and electron microscopy studies have observed RecA and Rad51

filament structures with distinctly different pitches, correlating to the identity of the bound nucleotide cofactor (Figure 3). An ATP-bound-like state exists in an extended filament conformation with a pitch of about 94 Å/turn, stretching the DNA within to about 150% relative to B-form DNA. The ADP-bound state, however, is in comparison more compressed, with a pitch of 65–85 Å/turn and does not appreciably extend DNA. The extended filament, trapped by the use of slowly hydrolyzable ATP analogs, has been correlated with activity using *in vitro* assays such as the D-loop assay which measures plectonemic joint formation between an ssDNA and a homology-containing supercoiled plasmid (Figure 4(b)). Consequently, and because conditions favoring formation of the compressed filaments do not lead to DNA-strand exchange, the extended filament has been described as the ‘active’ form of DNA-strand transferase

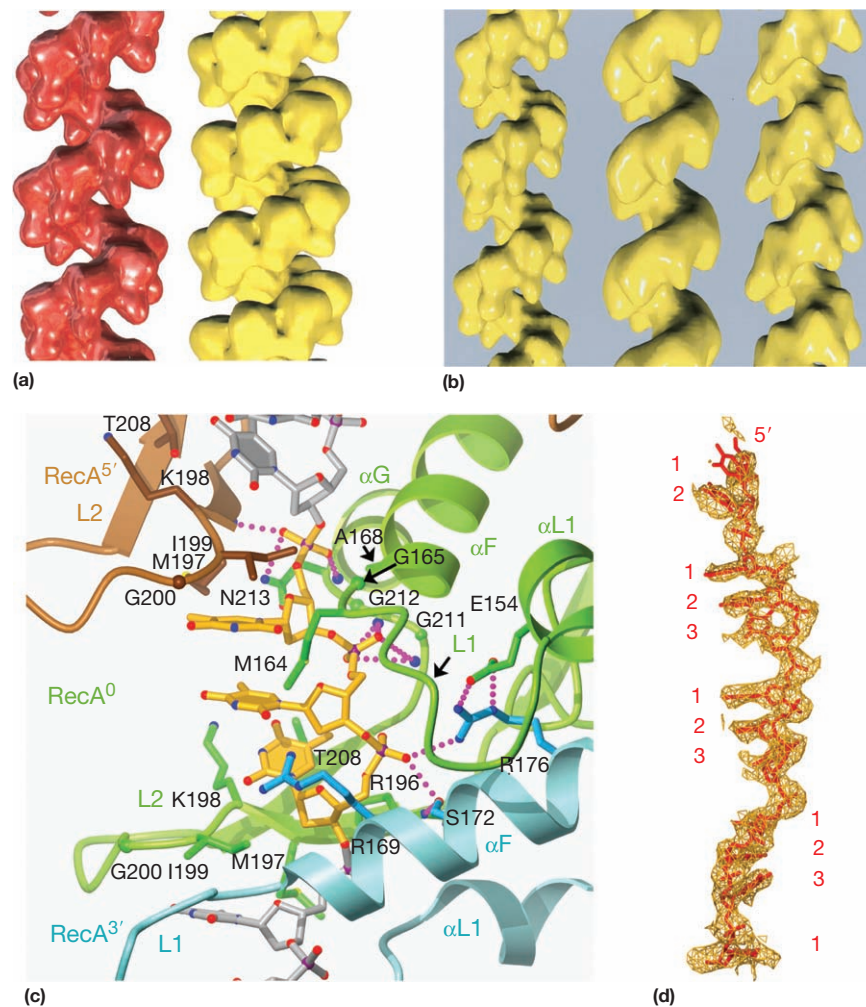


Figure 3 Structure of DNA-strand transferase filaments. (a) EM reconstructions of human RAD51 filaments in the extended (left, 99 Å pitch) and compressed (right, 76 Å pitch) conformations. (b) EM reconstructions of extended filament conformations of three DNA-strand transferases, left to right: human RAD51, *S. cerevisiae* Rad51, and RecA. (c) RecA crystal structure snapshot showing a nucleotide triplet bound by the primary monomer (RecA⁰). The triplet also makes contacts with RecA monomers 5' and 3' of the primary monomer. Notice the insertion of loop 2 (L2) residues between triplet base stacks. (d) RecA extended filament ssDNA electron density, with bases in triplets numbered 1–3. Notice the stretched and negatively twisted intertriplet base step between nucleotide triplets. (a, b) Modified from Yu X, Jacobs SA, West SC, et al. (2001) Domain structure and dynamics in the helical filaments formed by RecA and Rad51 on DNA. *Proceedings of the National Academy of Sciences of the United States of America* 98: 8419–8424. (c, d) Modified from Chen Z, Yang H, and Pavletich NP (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* 453: 489–484.

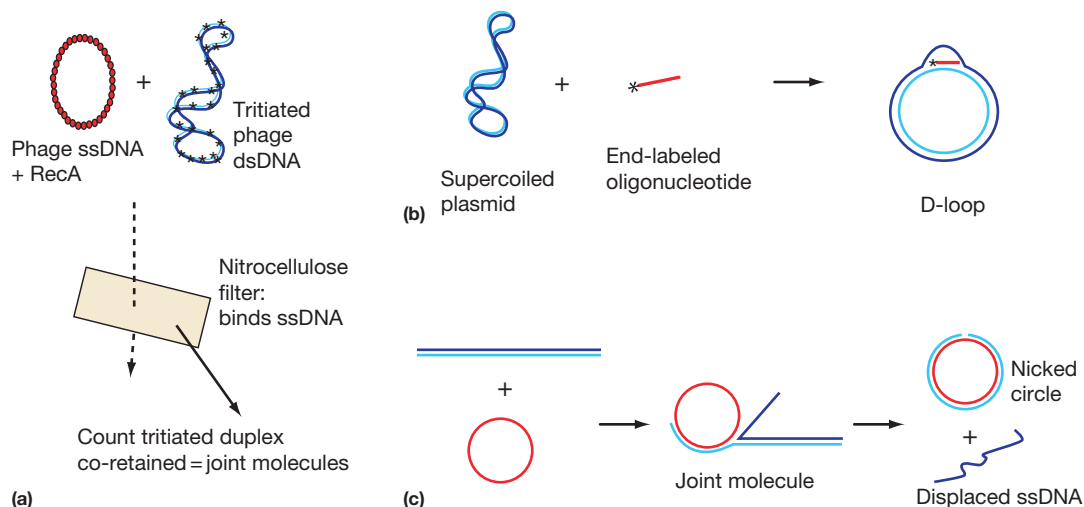


Figure 4 *In vitro* assays for DNA strand transferase activity. (a) Nitrocellulose filter assay for paranemic joints; (b) D-loop assay; and (c) DNA-strand exchange assay. See text for details.

filament, and the compressed form 'inactive'. However, single DNA molecule studies with Rad51 protein have shown that ATP hydrolysis, while leading to partial contraction of the filament, does not necessarily lead to DNA-strand transferase dissociation. The same filament is able to rebind ATP and again adopt a fully extended conformation. Therefore, it appears that ATP hydrolysis, ADP dissociation, and the rebinding of ATP lead to dynamic changes in the pitch of the DNA-strand transferase filament. It is not known if this plays a part in the DNA-strand transferase reaction, but DNA extension appears crucial for the homology search step.

Structure of the RecA-ssDNA Extended Filament

The X-ray crystal structure of RecA bound to ssDNA in the extended filament conformation (pitch of 94 Å/turn) provides key insights into the cooperative interactions within the filament between ATP, DNA, and adjacent monomers of RecA (Figure 3). Hydrogen bond networks link bound ssDNA to the ATPase active site, such that without DNA binding there would be no ordering of critical residues for ATPase activity. The active site is located at the RecA–RecA interface, with the ADP- AlF_4 (ATP transition state analog) completely buried between monomers, in contrast to the compressed filament structure where the nucleotide is partly solvent exposed. In the extended filament, the core domains undergo a rotation that creates a new interaction surface between monomers. Now, critical lysine residues from the 3'-adjacent RecA monomer, K248 and K250, are positioned to make charge-stabilized H-bonds to the ATP γ -phosphate and E96, another important residue thought to activate a water molecule for nucleophilic attack on the γ -phosphate. ATP hydrolysis destabilizes the inter-monomer interface, because it eliminates these favorable interactions and introduces a net charge in a buried environment.

Perhaps the most surprising information from the RecA-ssDNA crystal structure is the way in which the DNA is extended, which has profound implications for the homology search process. The DNA inside the filament is stretched to

5.1 Å per base (compared to 3.4 Å per base pair in free dsDNA), and untwisted to 18.5 bases per turn (compared to 10.5 base pairs per turn in free dsDNA). However, this stretching of DNA is not accomplished uniformly along the length of the DNA. Instead, each RecA monomer binds three nucleotides of DNA and holds them with close to B-DNA rise values (3.8–4.2 Å per nucleotide; B-DNA is 3.4 Å). Most of the 1.5-fold extension relative to B-DNA is accomplished by the stretched (7.1–7.8 Å rise) and negatively twisted (-42°) base steps between these nucleotide triplets, from the 3'-most base of one triplet to the first base of the next triplet. This stretching is accomplished by the insertion of hydrophobic residues in the L2 loop region into the base stacks between nucleotide triplets.

Implications of Filament Structure for the Homology Search

The structure of the RecA-ssDNA filament suggests that for greater than three nucleotides in dsDNA to be consecutively base-paired with nucleotides in the extended filament, the dsDNA too must extend (untwist) in the same pattern of triplet stacked bases intervened by unstacked and stretched intertriplet gaps. To accomplish this, RecA has a secondary DNA-binding site that is thought to interact with one strand of duplex DNA to destabilize the helix and allow bases to be sampled for complementarity. Because net unstacking of base pairs is necessary to pair consecutive triplets, this may favor correct homologous alignment by requiring that the next triplet of dsDNA paired be fully homologous in order to energetically favor base unstacking. Furthermore, triplet isolation limits cooperative base stacking interactions with nonhomologous bases as the heteroduplex region extends, requiring the selectivity be based more on Watson–Crick hydrogen bond interactions.

Although there are now crystal structures of RecA available bound to both ssDNA and dsDNA, one intermediate remains elusive – the three-stranded, synapsed paranemic joint. It is not known how the DNA strands are arranged in the joint molecule to allow homologous pairing without net intertwining of DNA strands. It is thought that the bases in the target duplex are

flipped out for sampling, that is, they temporarily and reversibly break their Watson–Crick hydrogen bonds with the other duplex strand, assume a new orientation, and sample base pairing with nucleotides inside of the DNA-strand transferase filament. The extension of bound DNA by the presynaptic filament reduces stacking interactions to facilitate base flipping, and the structure suggests that the basic sampling unit for base flipping and homology search is the nucleotide triplet.

Recombination Mediators and Other Factors

Nucleation is the rate-limiting step for Rad51 filament formation and, consistent with other filament-forming proteins such as actin and tubulin, nucleation is a key regulatory target involving recombination mediators and other protein factors. Some of these proteins may also be involved in maintaining filament stability to increase the steady-state level of active DNA-strand transferase filaments.

Mediators for Presynaptic Filament Formation

DNA-strand transferases are impeded from nucleating filaments when ssDNA-binding proteins saturate ssDNA. Recombination mediators were defined biochemically as proteins that allow filament formation of DNA-strand transferases on ssDNA coated by the cognate ssDNA-binding protein. The prototypical recombination mediator is UvsY protein from bacteriophage T4, which overcomes the block to ssDNA binding imposed by Gp32 SSB protein to allow UvsX filament formation. RecFOR provides this function in *E. coli* for RecA protein. In eukaryotes, the situation is more complex. *S. cerevisiae* Rad52, the ortholog of UvsY and RecO proteins, mediates Rad51 filament formation *in vitro*, but human protein does not. Instead, in eukaryotes containing the BRCA2 protein, the mediator function is occupied by this protein, which was identified as a central tumor suppressor protein for breast and ovarian cancers.

Rad54, a Specialized DNA Translocase Working with Rad51

Rad54 is a member of the Snf-2 family, a class of proteins that remodel specific protein–dsDNA complexes, such as nucleosomes or the complex between TATA-binding protein and the TATA box. Rad54 and its paralogs Rdh54 in *Saccharomyces cerevisiae* or RAD54B in human exhibit robust dsDNA-specific ATPase activity, and translocate on DNA (in the case of the yeast enzyme at 300 base pairs per second). Rad54 and its paralogs interact with Rad51 or Dmc1 and stimulate recombination in a species-specific manner. *In vitro*, Rad54 exerts multiple functions, including stabilization of the Rad51 filament on ssDNA (presynapsis), enhancing D-loop formation or DNA-strand exchange (synapsis), as well as branch migration and dissociation of the DNA-strand transferase filament on dsDNA (postsynapsis). Dissociation of Rad51 or Dmc1 is required to allow access by DNA polymerase to the invading 3'-end. In addition, Rad54 and Rdh54 are able to move nucleosomes on dsDNA. While Rad54 is critical for recombination *in vivo*, it is unclear which specific biochemical activity(ies) is most required for its *in vivo* function(s).

Other Eukaryotic DNA-Strand Transferase-Enhancing Proteins

In eukaryotes, new classes of proteins have been discovered that functionally and structurally interact with Rad51 or Dmc1. These include the Rad51 paralogs, Rad55–Rad57 as well as Shu1 and Psy3, in *S. cerevisiae*. In humans, RAD51 paralogs include five members – RAD51B, RAD51C, RAD51D, XRCC2, and XRCC3 – forming two main complexes (B–C–D–X2 and C–X3). In addition, the human breast and ovarian tumor suppressor BRCA2 bind multiple molecules of RAD51 and other proteins, including PALB2 and DSS1, that are expected to play a role in recombination. While it is known that these factors interact with Rad51, directly or indirectly, their exact function remains to be determined. The multitude of factors that interact with Rad51 and that are genetically implicated in recombination suggests complex regulation of the assembly, stability, and function of the Rad51–ssDNA filament.

Anti-Recombinases Target the Rad51-ssDNA Filaments for Disassembly

While recombination is an active DNA repair mechanism, inappropriate recombination can interfere with other cellular processes such as DNA replication or lead to toxic intermediates. In eukaryotes, anti-recombinases have been identified genetically that enforce negative control on HR. The Srs2 DNA helicase was identified genetically in *S. cerevisiae* as an anti-recombinase. Translocation of Srs2 dissociates Rad51 from ssDNA leading to disassembly of presynaptic filaments. Srs2 is the archetype for such an anti-recombinase activity, and other eukaryotic proteins, including FANCI, FBH1, RECQ5, and BLM, were shown to dissociate Rad51 from ssDNA. In summary, the DNA-strand transferase presynaptic filament appears to be controlled in a meta-stable balance between processes that control its assembly and disassembly.

In Vitro Assays for DNA-Strand Transferase Activity

Functional *in vitro* assays for DNA-strand transferases have generated a detailed mechanistic understanding of their function and have provided test systems for the identification and characterization of recombination mediators and other accessory proteins (Figure 4).

Filter-Binding Assay for Detection of Paranemic Joint Molecules

The formation of the initial product of synapsis, the paranemic joint, is difficult to study because of its metastable nature, existing only in the presence of the DNA-strand transferase filament. The D-loop and DNA-strand exchange assays (discussed below), while being invaluable for the study of DNA-strand transferase reactions, are limited in that they can only detect stable plectonemic joint molecules where the DNA strands have already intertwined. A filter-binding assay, developed for RecA protein, can detect paranemic joints because it does not require deproteinization of reactions (Figure 4(a)). RecA filaments are first formed on

circular ssDNA and then are paired with homologous supercoiled plasmid DNA that has been labeled with tritium. Lack of a free DNA end in the DNA substrates ensures that all joint molecules formed are paranemic, since they cannot intertwine. To detect the presence of joint molecules, the reactions are diluted in a high-salt buffer and passed through nitrocellulose filters. The filter selectively binds ssDNA and any tritiated plasmid that is complexed to it by RecA in paranemic joint DNA molecules. The level of joint molecules is determined by scintillation counting of the washed filters. A major limitation to this assay is that protein binding to dsDNA causes the plasmid to bind the filter regardless of it being complexed in a joint molecule with ssDNA or not. Therefore, the assay is not suitable for DNA-strand transferases that lack an ssDNA-binding preference, such as the Rad51 or Dmc1 proteins.

The D-Loop Assay

The D-loop assay (Figure 4(b)) is a versatile and sensitive way to detect DNA-strand transferase activity *in vitro*. The ssDNA substrate is commonly a synthetic oligonucleotide that has been 5'-end-labeled with ^{32}P , while the dsDNA homology donor is a small (~ 3 kb), negatively supercoiled plasmid. DNA-strand transferase filaments are first formed on the oligonucleotide. The supercoiled DNA is then added and D-loops form when the oligonucleotide is incorporated into the plasmid as a plectoneme. After deproteinization, the reactions are analyzed by agarose gel electrophoresis, where the fast-migrating free ssDNA is separated from the slow-migrating D-loops, allowing precise quantitation of the proportion of oligonucleotide found in the product D-loop. The deproteinization sidesteps the requirement of the DNA-strand transferase to release its product. Since the plasmid is covalently closed, the amount of DNA-strand exchange that can occur (heteroduplex DNA tract length) is limited by topological constraints. The dsDNA must be supercoiled for productive D-loop formation, as it aids DNA-strand transferase filament-mediated dsDNA unwinding and stabilizes the D-loop during gel electrophoresis.

The DNA-Strand Exchange Assay

Similar to the D-loop assay, the three-strand DNA exchange assay (Figure 4(c)) detects plectonemic joint molecules formed by DNA-strand transferases. However, the ssDNA substrate is relatively long (5.4 kb for the frequently used phiX174 virion form DNA) and circular, while the dsDNA partner is linear and of the same sequence (phiX174 Replication Form I). Consequently, the region of DNA-strand exchange can now be very long, up to the entire length of the DNAs.

First, DNA-strand transferase filaments are formed on the ssDNA with the help of the cognate ssDNA-binding protein to remove secondary ssDNA structure. Then, the dsDNA is added and the homology is aligned, likely via the formation of a paranemic joint. Prerequisite to the appearance of products on agarose gels, a plectonemic joint molecule forms, starting with DNA intertwining one strand of a dsDNA end with the incoming ssDNA to begin strand exchange. Lastly, polar branch migration extends the heteroduplex region around the circle, until DNA-strand exchange is complete, with nicked circular dsDNA

and displaced linear ssDNA as final products. Upon deproteinization of the reactions followed by agarose gel electrophoresis, two product species can be distinguished – joint molecules and nicked circles. The linear ssDNA product usually cannot be resolved from the circular ssDNA substrate. The joint molecules include all plectonemes that have begun but not completed strand exchange, whereas the nicked circles represent the product with the maximum extent of heteroduplex DNA.

Single DNA Molecule Studies

The development of techniques to track single DNA molecules and to visualize recombination proteins acting on them has advanced our understanding of DNA-strand transferase filament nucleation, growth, and disassembly. Study at the single-molecule level eliminates the averaging and consequent loss of information that results from experiments that employ a large ensemble of molecules. Many different techniques have the sensitivity to gain information on single molecules of DNA. One classical method employed to study DNA-strand transferases involves the use of linear DNA (usually linear 48-kb lambda dsDNA) end-bound to a 1- μm -diameter bead. The bead is held in place inside of a multi-channelled flow cell by an optical trap. By moving the stage, the bead and bound DNA can be transferred into different channels within the flow cell, thereby changing buffer conditions or adding additional proteins. Nucleoprotein filament formation, stability, and disassembly can be monitored by chemically labeling the DNA-strand transferase with a fluorophore. Alternatively, the DNA end opposite the bead is labeled with a fluorophore and DNA-strand transferase binding is monitored by the movement of the fluorophore that results from the extension of the DNA that accompanies DNA-strand transferase binding. Single-molecule approaches to study DNA-strand transferases are still in the early phase of their development, but have already provided unprecedented insights into their function.

DNA-Strand Transferases in Signaling and Translesion Synthesis

Although DNA-strand transferases perform a primary function in the mechanism of HR, they are also involved in other cellular processes, including DNA damage signaling and translesion DNA synthesis. In *E. coli*, the SOS damage response is dependent on the RecA-ssDNA filament to act as a coprotease to cleave and deactivate the LexA repressor. This, in turn, leads to the expression of many derepressed genes (SOS operon) required for the damage response. Independent of the SOS signaling function, RecA protein is also required for translesion DNA synthesis by DNA polymerase V. In *S. cerevisiae*, a single unrepairable DSB leads to a checkpoint-mediated G(2)/M arrest, but a phenomenon termed adaptation allows the cell cycle to resume even in the presence of the DSB. Adaptation requires Rad51 protein through an unknown mechanism involving its interaction with Rad52 protein. Thus, it appears that eubacteria and eukaryotes have evolved feedback mechanisms to monitor the status of DNA-strand transferase filaments as a way to assess genotoxic stress conditions.

See also: **Cell Architecture and Function:** Actin Assembly/Disassembly; Cell Cycle: DNA Damage Checkpoints; Meiosis; **Molecular Biology:** DNA Mismatch Repair and Homologous Recombination; DNA Replication Fork, Bacterial; DNA Replication Fork, Eukaryotic; Homologous Recombination in Meiosis; LexA Regulatory System; Nonhomologous End Joining in Eukaryotes; Recombination-Dependent DNA Replication.

Further Reading

- Beernink HT and Morrical SW (1999) RMPs: Recombination/replication mediator proteins. *Trends in Biochemical Sciences* 24: 385–389.
- Bianco PR, Tracy RB, and Kowalczykowski SC (1998) DNA strand exchange proteins: A biochemical and physical comparison. *Frontiers in Biosciences* 3: D570–D603.
- Chen Z, Yang H, and Pavletich NP (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* 453: 489–484.
- Cox MM (2007) Regulation of bacterial RecA protein function. *Critical Reviews Biochemistry and Molecular Biology* 42: 41–63.
- Egelman EH (2003) A tale of two polymers: New insights into helical filaments. *Nature Reviews Molecular Cell Biology* 4: 621–630.
- Ehmsen KT and Heyer WD (2008) Biochemistry of meiotic recombination: Formation, processing, and resolution of recombination intermediates. *Genome Dynamics and Stability* 3: 91–164.
- Forget AL and Kowalczykowski SC (2010) Single-molecule imaging brings Rad51 nucleoprotein filaments into focus. *Trends in Cell Biology* 20: 269–276.
- Gupta RC, Folta-Stogniew E, O'Malley S, Takahashi M, and Radding CM (1999) Rapid exchange of A:T base pairs is essential for recognition of DNA homology by human Rad51 recombination protein. *Molecular Cell* 4: 705–714.
- Heyer WD, Ehmsen KT, and Liu J (2010) Regulation of homologous recombination in eukaryotes. *Annual Review of Genetics* 44: 113–139.
- West SC (2003) Molecular views of recombination proteins and their control. *Nature Reviews Molecular Cell Biology* 4: 1–11.
- Yu X, Jacobs SA, West SC, et al. (2001) Domain structure and dynamics in the helical filaments formed by RecA and Rad51 on DNA. *Proceedings of the National Academy of Sciences of the United States of America* 98: 8419–8424.

RecQ Helicase Systems

G Mosedale and I D Hickson, University of Oxford, Oxford, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Crossover An exchange of genetic information either side of a double-strand break (DSB) during homologous recombination (HR).

Double holliday junction (DHJ) dissolution A reaction catalyzed by BLM/TopoIII α /RMI1 that results in a DHJ being processed into exclusively non-crossover products.

DSB DNA double-strand break.

Holliday junction (HJ) A four-stranded DNA HR intermediate.

Homologous recombination (HR) A process of DSB repair that utilizes homologous DNA sequence as a template for high-fidelity repair.

Late replication intermediate (LRI) It occurs when two replication forks converge.

Nonhomologous end joining (NHEJ) A DSB repair pathway that rejoins the broken DNA ends in a process that

does not require sequence similarity and is therefore error prone.

Sister chromatid exchange (SCE) A form of HR that can be visualized in metaphase cells and results from reciprocal exchange of genetic information between sister chromatids.

Synthesis-dependent strand annealing (SDSA) It can occur following the early invasion step of HR when a D-loop is extended and is then able to be annealed to the other end of the DSB due to sequence homology.

Topoisomerase An enzyme that catalyzes a DNA-strand passage event, either by creating a break in one strand of the DNA helix (type I) or by creating a DSB (type II), allowing another DNA strand to pass through.

Ultrafine bridge (UFB) A DNA bridge that can be visualized in cells during anaphase by staining with antibodies against BLM/TopoIII α /RMI1, but cannot be stained with conventional DNA dyes such as DAPI.

Introduction

Helicases are enzymes that utilize the energy derived from adenosine triphosphate (ATP) hydrolysis to separate the complementary strands of a DNA or RNA duplex. They are found in all organisms and their strand-separating action is essential for a wide range of biological processes, including DNA replication, transcription, and repair, as well as various aspects of RNA metabolism.

DNA helicases have been classified into superfamilies based on their sequence similarity. The two largest superfamilies, SF1 and SF2, share a characteristic helicase domain that is composed of seven motifs: I, Ia, II, III, IV, V, and VI. Motifs I and II make up the highly conserved Walker A and B boxes that are conserved in many ATPases. SF1 and SF2 include both DNA and RNA helicases.

Helicases can also be classified according to the directionality of their translocation along DNA/RNA. These enzymes often require a region of single-stranded DNA (ssDNA) to bind to, before initiating unwinding of an adjacent region of the duplex. The helicase enzyme then translocates unidirectionally along the ssDNA, coupling displacement of the complementary strand of the duplex with ATP hydrolysis. The directionality is defined by whether the enzyme translocates 3'–5' or 5'–3' along this ssDNA.

The distance that a single helicase molecule is able to travel along the DNA during the unwinding reaction is termed the 'processivity' of the enzyme, and it differs considerably among different helicases. The processivity and DNA-substrate specificity of a helicase can be enhanced by a number of factors. Additional domains outside of the helicase domain may promote binding to different DNA substrates and hence unwinding. Protein-binding partners and subsequent protein-complex

formation may stimulate the unwinding activity of the helicase. Moreover, posttranslational modifications may affect the processivity of unwinding.

Structural and biochemical studies have provided insight into the mechanism of DNA unwinding by helicases. The favored model is the 'inchworm' mechanism in which the helicase, either monomeric or oligomeric, acts in a manner analogous to that of a snow plough, translocating along the DNA, possibly one base pair at a time, and separating the DNA strands using a wedge-like structure to peel apart the DNA strands. The details of this mechanism vary between helicases but require the protein to be able to bind simultaneously to both ssDNA and duplex DNA at some point during the reaction. The alternative 'rolling-circle' mechanism assumes that the helicase acts as a multimer, with a conformational change occurring following or during ATP hydrolysis that stimulates the ability of the helicase to separate the DNA duplex.

The RecQ helicases are a family of 3'–5' helicases belonging to SF2. They are highly conserved through evolution and have been shown to play an important role in maintaining the stability of the genome. This family of helicases is discussed in more detail in this article.

RecQ Helicases

RecQ helicases are very well conserved throughout evolution, from prokaryotes to eukaryotes, with the number of RecQ helicases expressed in a given organism often correlating with the size of the genome; cells with larger genomes, such as in mammals, often express several, while unicellular organisms, such as yeast or bacteria, express one or sometimes none (Figure 1). This family is named after the prototypical member

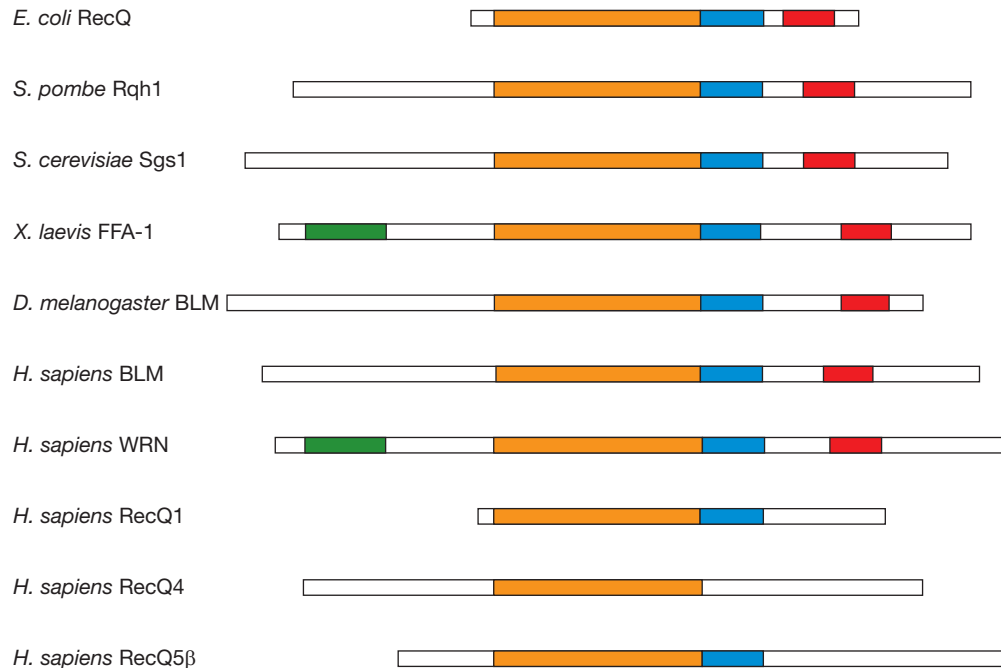


Figure 1 Comparison of the domain architecture of RecQ helicases. Colored areas represent conserved domains including the helicase domain (orange), RQC domain (blue), HRDC domain (red), and exonuclease domain (green).

in *Escherichia coli*, RecQ. The central helicase domain, containing the Walker A and B box motifs, is the defining feature of all RecQ helicases and is essential for the binding and hydrolysis of ATP. Surrounding domains differ between the various RecQ family members, but there is commonly an RecQ C-terminal domain (RQC domain) and a helicase and RNase D C-terminal domain (HRDC domain) present. The RQC domain is specific to RecQ helicases and is located immediately C-terminal to the helicase domain. It is composed of two subdomains, a so-called winged-helix motif, and a zinc-binding region, the combination of which is responsible for DNA binding and hence determining the substrate specificity of the different enzymes. The HRDC domain is also usually C-terminal to the helicase domain, but is not found in all RecQ helicases. The HRDC domain is also considered to play a role in mediating protein–DNA interactions. Other motifs are also present in some RecQ helicases; for example, the human WRN protein has an exonuclease domain. Posttranslational modifications, such as phosphorylation and sumoylation, play a central role in the regulation of activity of RecQ helicases.

Humans have five RecQ helicases, designated BLM, WRN, RecQ1, RecQ4, and RecQ5, all of which share the highly conserved central helicase region but differ in their ancillary domains. The main interest in RecQ helicases stems from the discovery that mutation in three of these human RecQ helicases leads to defined genetic disorders that result in human disease such as premature aging and cancer; mutation of BLM causes Bloom’s syndrome (BS), mutation of WRN leads to Werner’s syndrome (WS), and mutation of RecQ4 causes Rothmund Thomson syndrome (RTS). One common feature of each of these human RecQ disorders is their association with an increase in cancer predisposition. RecQ proteins therefore act as genome caretakers, protecting the genome from deleterious

changes that might cause tumor formation; lack of a RecQ helicase causes genome instability and an increased likelihood of oncogenic activation.

Yeast model systems have been used extensively to study the function of RecQ helicases. Sgs1 in *Saccharomyces cerevisiae* and Rqh1 in *Schizosaccharomyces pombe* have the characteristic domains seen in most RecQ helicases, including the central helicase domain, the RQC domain, and the HRDC domain. Yeast-deletion mutants of RecQ helicases provide a relatively quick way to investigate the genetic interactions between RecQ helicases and other pathways of DNA metabolism. The high level of conservation between yeast and human RecQ helicases suggests that the functions of these proteins are evolutionarily conserved, likely making findings in yeast applicable to human proteins as well.

RecQ Helicases and DNA Repair

The faithful replication of genetic information during the cell cycle is essential for cell viability. Failure to maintain the fidelity of DNA replication can result in chromosomal aberrations, such as translocations and deletions, which can promote tumorigenesis. Cells are constantly exposed to agents that damage DNA both from endogenous sources, such as free radicals generated within the cell, and from exogenous sources, such as ultraviolet light or chemicals. To combat this damage, cells have evolved numerous DNA repair pathways, which ensure that the DNA is repaired before the damage is replicated and transferred to daughter cells. RecQ helicases are a family of proteins that play a role in this maintenance of genome integrity by stabilizing damaged replication forks to prevent their irreversible demise.

A type of DNA damage that is particularly detrimental to the cell is the DNA double-strand break (DSB). DSBs can arise in the cell during DNA replication, when a replication fork encounters damage and collapses, or can occur directly following exposure to ionizing radiation. DSBs can be repaired by two major pathways in human cells: homologous recombination (HR) and nonhomologous end joining (NHEJ). HR is usually an error-free process of repair because it utilizes a homologous DNA sequence such as a sister chromatid. RecQ helicases have been shown to play a major role in suppressing recombination to ensure that illegitimate or aberrant recombination does not occur.

Bloom's Syndrome

BS is caused by loss of function of the protein BLM. BS is an extremely rare, autosomal recessive disorder, affecting less than 1 in 10 million people. The clinical phenotype of BS patients includes dwarfism, type II diabetes, immunodeficiency, reduced fertility, and a predisposition to the development of all types of cancers. It is this susceptibility to the range of cancers seen in the human population that makes the study of BS and the BLM protein of such interest to cancer researchers. Cells from BS patients have a number of hallmark characteristics including sensitivity to drugs that inhibit DNA replication and an increase in the level of sister chromatid exchanges (SCEs), a marker used in the diagnosis of the disease. SCEs are thought to arise via a form of HR and are usually aphenotypic as the two sister chromatids should have an identical DNA sequence. However, if SCE occurs unequally, or between nonidentical sequences, loss of heterozygosity can arise, leading to an increase in the likelihood of tumorigenesis if this occurs in tumor-suppressor genes. The presence of BLM is therefore required to limit this hyper-recombination phenotype. Inherited nonsense or frameshift mutations in the *BLM* gene usually lead to the production of a nonfunctional truncated BLM protein, which may be degraded or which is unable to undergo its normal cellular functions, thereby contributing to genome instability and subsequent cancer formation.

The BLM Protein

The BLM protein is an ATP-dependent 3'–5' DNA helicase that requires a short region of ssDNA to bind to before translocating in a 3' to 5' direction, and displacing the complementary strand. It has been shown *in vitro* that its preferred substrates are forked structures, two types of HR intermediates (DNA displacement loops (D-loops) and four-way Holliday junction (HJ) structures), and G-quadruplex DNA, a noncanonical DNA structure that can form in guanine-rich DNA (Figure 2). Besides being able to unwind these various DNA structures, BLM is also able to perform the opposite function, that of annealing complementary DNA strands together. This strand-annealing activity is independent of ATP hydrolysis; indeed, it cannot occur in the presence of ATP, and therefore the relevance of this annealing activity *in vivo* is as yet not clear. BLM is also able to branch-migrate the HJ recombination intermediate.

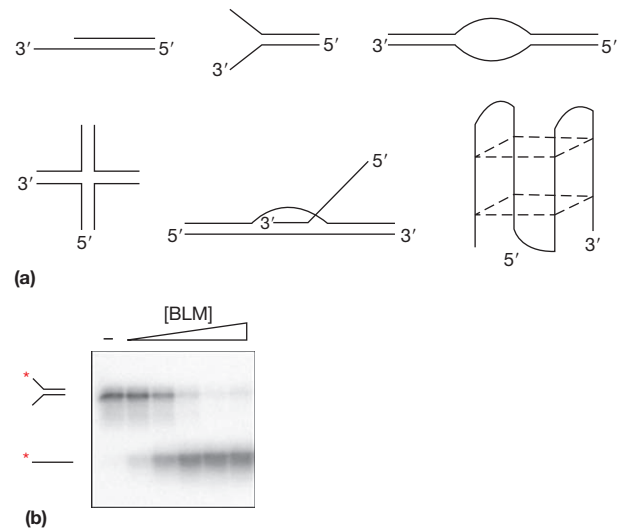


Figure 2 (a) Examples of DNA substrates for RecQ helicases. From top, left to right, a 3'-single-stranded DNA overhang; a forked duplex; a bubble structure; a Holliday junction; a displacement loop; and G-quadruplex DNA. (b) An example of the helicase activity of recombinant BLM protein. BLM is able to unwind a forked duplex into single-stranded DNA in a concentration-dependent manner. The asterisk represents the radiolabeled 5' phosphate.

BLM Binding Partners

BLM interacts with a number of proteins *in vivo*. It carries out many of its biological functions in conjunction with topoisomerase III α (TopoIII α), a type IA topoisomerase that has the ability to catalyze DNA-strand passage by creating a break within a region of ssDNA and rejoining this break once another DNA strand has been passed through it. BLM also forms a complex with a protein called RMI1 to form the BLM/TopoIII α /RMI1, or BTR, complex. Rmi1 has no detectable enzymatic activity of its own, but the presence of RMI1 stimulates the enzymatic activity of BLM and TopoIII α via a direct interaction between RMI1 and TopoIII α . A fourth component of the complex, RMI2, is considered to enhance the stability of the BTR complex, but has no effect on the enzymatic activity of BLM. BLM has also been shown to interact with the ubiquitous ssDNA-binding protein, replication protein A (RPA). RPA is a heterotrimeric protein essential for DNA replication and therefore cell survival. RPA can stimulate the DNA-unwinding activity of BLM.

In cell biological assays, BLM has been shown to localize to distinct nuclear foci following replication stress. These foci are thought to represent sites of DNA damage as many proteins involved in the DNA damage response colocalize there, including the key HR protein, Rad51, and the BRCA1 and Fanconi anemia proteins. The presence of BLM in these foci suggests a role for BLM in the replication-stress response, possibly by maintaining the integrity of stalled replication forks.

S. cerevisiae Sgs1

Many of the BLM–protein interactions seen in human cells are conserved in yeast. The budding yeast RecQ helicase, Sgs1, is most similar to the human RecQ helicase BLM, with many of the cellular features of BS cells being recapitulated in *sgs1 Δ*

cells. Like the hyper-recombination phenotype observed in BS cells, *sgs1Δ* cells have an increase in the level of recombination. Similarly, *sgs1Δ* cells are sensitive to drugs that inhibit DNA replication. Sgs1 also binds Top3 and Rmi1, the yeast orthologs of human TopoIII α and RMI1.

Werner's Syndrome

WS is an autosomal recessive disease caused by mutation of the *WRN* gene. WS is associated with premature aging and an increased likelihood of cancer. Cells from WS patients divide a limited number of times and enter senescence much earlier than normal cells, making cellular studies challenging. This premature senescence could provide the basis for the aging phenotype seen in WS patients. WS cells display an increased number of aberrant chromosomes, such as translocations and deletions, and are sensitive to replication inhibitors and DNA cross-linking agents.

The WRN protein not only contains the characteristic features seen in RecQ helicases, but also has a 3'-5' exonuclease domain that is similar to the exonuclease domain of *E. coli* DNA polymerase I. Most mutations in *WRN* result in the formation of a truncated protein that is unable to translocate to the nucleus. The WRN protein catalyzes similar DNA-unwinding reactions to BLM. It can unwind forked structures, D-loops, HJs, and G-quadruplex DNA (Figure 2), and this helicase activity can be stimulated by RPA. Additionally, WRN, like BLM, can branch-migrate HJ substrates.

Rothmund Thomson Syndrome

RTS is an autosomal recessive disease caused by mutation of the *RecQ4* gene. The main symptoms of this disease are skeletal and skin abnormalities, developmental retardation, and a predisposition to the development of osteosarcoma. As with other RecQ-deficient syndromes, RTS cells display an increased frequency of chromosomal aberrations, including translocations and deletions, suggesting a role for RecQ4 in maintaining genome stability.

The RecQ4 gene product is unique among RecQ helicases in lacking both the RQC domain and an HRDC domain. RecQ4 has ATPase activity, but initial studies showed no detectable helicase activity on a wide range of substrates. However, it is able to anneal complementary ssDNA and has been proposed to play a role during replication via its interaction with core replication factors. More recent studies indicate that RecQ4 might even possess two separate helicase domains and activities, and bind to proteins required for replication initiation.

Mutations in *RecQ4* also cause two other human genetic disorders: RAPADILINO and Baller-Gerold syndrome. Neither of these diseases is associated with an increased risk of cancer development and they are phenotypically distinct from each other and RTS, although they do have in common some aspects of defective development, such as skeletal malformation.

Anti-Recombinogenic Actions of RecQ Helicases

As mentioned earlier, the classical feature of cells from BS patients is a high level of spontaneous SCE. This suggests that

the presence of BLM is required to suppress this form of recombination. Similarly, *sgs1Δ* cells show an increase in recombination events, a feature that can be reversed following deletion of the key HR gene *RAD52*. The suppression of HR by RecQ helicases has been demonstrated in a number of ways. In biochemical assays, BLM and WRN have been shown to disrupt two structures that form early in HR, the Rad51 filament and the D-loop. In HR, following the formation of a DSB, an exonuclease resects the break to reveal a 3' ssDNA overhang (Figure 3). The Rad51 recombinase binds to this 3' ssDNA overhang and catalyzes invasion of this strand into homologous dsDNA, resulting in the formation of a D-loop. BLM has been shown to disrupt this Rad51 filament formation, thereby limiting HR. Strand invasion generates a D-loop. One possible outcome for a D-loop is that the invading ssDNA can be extended far enough to become a template for the so-called synthesis-dependent strand annealing (SDSA) pathway. This process does not involve an HJ intermediate and requires unwinding of the extended D-loop. BLM can catalyze D-loop unwinding, thereby facilitating SDSA and preventing crossover formation. Crossover products only form following HR during DSB repair and are defined by the exchange of the genetic material on either side of the break site, which can be deleterious to the stability of the genome. An alternative outcome for a D-loop involves the formation of HJ (Figure 3). If the second strand of the DSB is captured, a double Holliday junction (DHJ) will be formed. HJs can be resolved in prokaryotes and eukaryotes by HJ resolvases. These HJ resolvases make symmetrical nicks on two of the four strands of the HJ, which is followed by a ligation step. Resolvases create both crossover and non-crossover products with equal frequency, depending on the orientation of the nicks. In biochemical experiments using an oligonucleotide-based DHJ substrate, BLM and TopoIII α were also shown to be able to act on this DHJ structure, and to dissolve the structure into exclusively non-crossover products. This so-called DHJ-dissolution activity is specific to BLM and TopoIII α , and requires the helicase domain and HJ branch-migrating ability of BLM and the strand-passage activity of TopoIII α . RMI1 stimulates DHJ dissolution by interacting with TopoIII α and enhancing its strand-passage function. This DHJ dissolution activity of BLM could be one way in which BLM prevents crossing over and hence SCE.

WRN interacts with a number of proteins involved in DNA recombination and repair, including Ku70/Ku80, a heterodimeric protein complex that is essential for NHEJ, an error-prone pathway of DSB repair. The binding of the Ku complex to WRN stimulates the exonuclease activity of WRN. Lack of WRN in cells results in larger deletion tracts during NHEJ, suggesting a role for WRN in limiting NHEJ by either processing or stabilizing the DNA ends.

Pro-Recombinogenic Roles of RecQ Helicases

RecQ helicases can clearly play a role in many of the steps of HR, with most of these roles leading to suppression of recombination. However, recent data have shown that BLM and Sgs1 are able to affect one of the earliest steps of HR, the formation of the 3' ssDNA overhang. Initial processing of the DSB ends

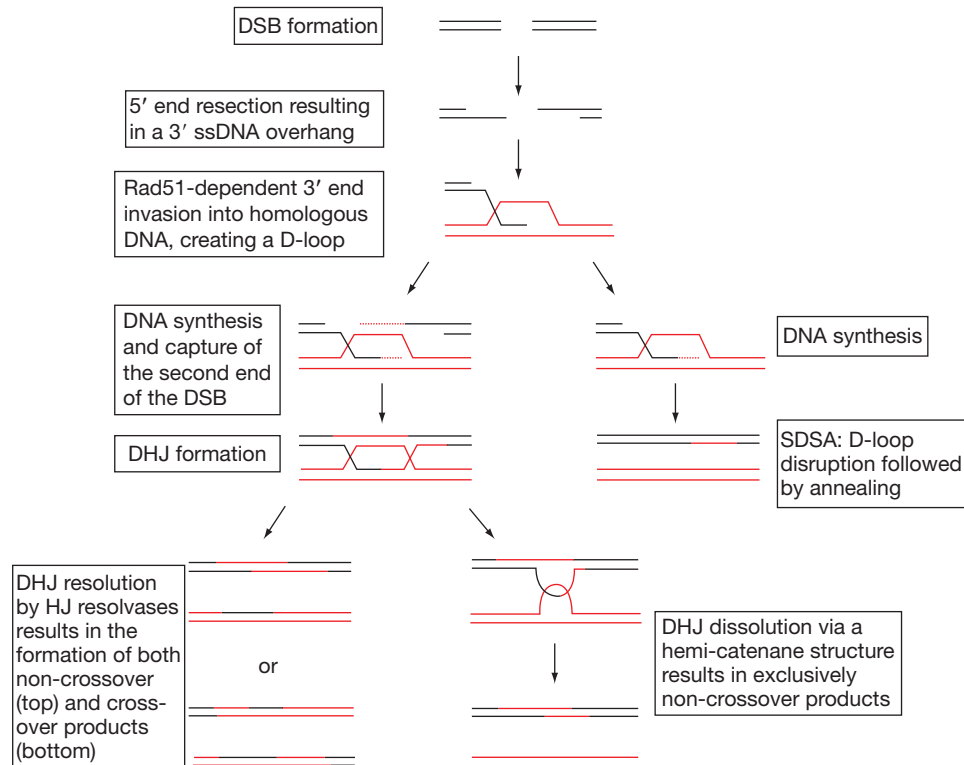


Figure 3 Schematic of the homologous recombination pathway of DSB repair. The RecQ helicases can act a number of steps within this process to limit crossover formation. BLM can disrupt Rad51 filament formation on the 3'-ssDNA overhang thereby preventing HR. BLM and WRN can unwind D-loop structures, a process that is required for synthesis-dependent strand annealing (SDSA) and produces only non-crossover products. Alternatively, if the second double-strand break (DSB) end has been captured and a double Holliday junction (DHJ) is formed, BLM can ensure that exclusively non-crossover products are generated by catalyzing DHJ dissolution in conjunction with TopoIII α and RMI1.

requires the Mre11–Rad50–Nbs1 complex (Mre11–Rad50–Xrs2 complex in yeast) and CtIP (Sae2 in yeast). Biochemical evidence using purified BLM protein indicates that BLM promotes 5' end resection over long distances by stimulating Exonuclease I (Exo I), thereby activating HR. However, in yeast, Sgs1 has been shown to be able to catalyze this resection step independently of Exo I, and through stimulating the activity of another exonuclease, Dna2, to promote 5' end resection. This role for BLM and Sgs1 in promoting the early resecting step of HR seems to be at odds with the recombination suppression action of RecQ helicases, but this discrepancy could be resolved if the resection catalyzed by BLM channels repair through SDSA, a process that gives exclusively non-crossover products.

RecQ Helicases in Meiosis

Studies in yeast have revealed an important role for RecQ helicases in a scheduled DSB repair process that occurs during meiosis. Genetic diversity is essential for evolution and is generated in meiosis during the exchange of genetic information between homologous chromosomes, termed 'crossing over'. Unlike in mitotically dividing cells, crossing over is promoted in meiosis. This genetic exchange is initiated by a programmed DSB that is processed into a DHJ structure. The Mus81–Mms4/Eme1 enzyme complex is an HJ resolvase responsible for

resolving at least some of these DHJs into crossover products during meiosis. In *S. cerevisiae*, *sgs1 Δ mus81 Δ* double mutant cells are synthetically lethal (i.e., they are unable to survive), suggesting a role for Sgs1 in meiotic recombination. These *sgs1 Δ mus81 Δ* cells are unable to separate their chromosomes during meiosis, presumably because some recombination structures remain unresolved in the absence of both Sgs1 and Mus81.

Roles of RecQ Helicases during Replication

RecQ helicases function during replication to maintain genome stability. One way in which they may do this is by unwinding roadblocks in the DNA that may otherwise cause replication fork (RF) stalling or collapse. BLM and WRN have been shown *in vitro* to be able to unwind non-B form DNA, such as G-quadruplex DNA, which forms in guanine-rich regions of the genome, thereby allowing replication to proceed unhindered. WRN is also able to unwind telomeric DNA that might form G-quadruplexes. Telomeres protect the end of the chromosome from being recognized as a DSB by the cellular DNA repair machinery. They are made up of the repetitive sequence (TTAGGG) which is 2–10 kb in length, with a 3' ssDNA tail at the very end. This tail is able to loop back and invade the telomeric sequence, creating a so-called t-loop, thereby protecting the DNA end. The 3' ssDNA tail is a substrate for telomerase, a specialized reverse transcriptase that is

able to maintain the length of the telomere during replication. WRN has been shown to be necessary to unwind these t-loops to allow replication of the telomeres to occur. In the absence of telomerase, cells maintain the length of their telomeres by the alternative lengthening of telomeres (ALT) pathway. In ALT, the 3' ssDNA end of a shortened telomere invades double-stranded telomeric DNA and uses this as a template for HR-mediated elongation. The source of the dsDNA is controversial but may be the same or another chromosomal end, or an extrachromosomal telomeric circle. *S. cerevisiae* cells lacking telomerase components have been shown to senesce earlier than wild-type cells. A very small proportion of these cells are able to escape senescence by utilizing an Sgs1-mediated HR pathway of ALT.

An additional role for BLM during replication was recently shown to be the restart of stalled RFs. RFs stall when they encounter a block in the DNA such as a DNA secondary structure, protein–DNA complexes, or specific DNA damage such as a DNA cross-link or nick. Stalled RFs have the potential to collapse, resulting in a single-ended DSB that is dependent on HR for repair. BLM has been shown to be able to stabilize stalled RFs and ensure replication progression, an activity that is dependent on the helicase activity of BLM and phosphorylation of a specific threonine residue on BLM (Thr 99), which is the target for the stress-activated kinase ATR.

Biochemical evidence has also indicated that BLM is able to catalyze RF regression (Figure 4). Fork regression is one way in

which stalled RFs can be stabilized, preventing their collapse. It occurs when the nascent DNA strands at an RF (which are complementary) are unwound from the parental strands and anneal to each other, creating a four-way junction analogous to an HJ. Once the regressed fork has been formed, the RF is stabilized to allow time for repair processes to remove or bypass the blocking lesion. BLM can restart replication either by direct reversal of the regressed fork or by forming a DHJ. Again, the helicase domain of BLM is required for this fork-regression activity. Theoretically, fork regression will only occur if the RF-stalling lesion is present on the leading strand. If the lesion is present on the lagging strand, a gap will be left in the nascent DNA that can be either filled postreplicatively or filled by recombination-mediated gap repair via DHJ formation (Figure 4). DHJ dissolution by BLM/TopoIII α will ensure that crossovers, and potential genome instability, do not occur.

S. cerevisiae has been used extensively to study RF stalling and recovery within a cell population. Treatment of *sgs1 Δ* , *top3 Δ* , or *rmi1 Δ* cells with the drugs hydroxyurea (HU, depletes nucleotide pools leading to RF stalling) or methylmethanesulphonate (MMS, an alkylating agent that damages DNA and causes DNA polymerase stalling) causes increased cell death compared to wild-type cells. A technique called two-dimensional gel electrophoresis can be used to analyze a specific region of DNA in yeast cells and has revealed that *sgs1 Δ* cells contain unresolved, X-shaped recombination intermediates following MMS exposure. A yeast strain carrying mutations

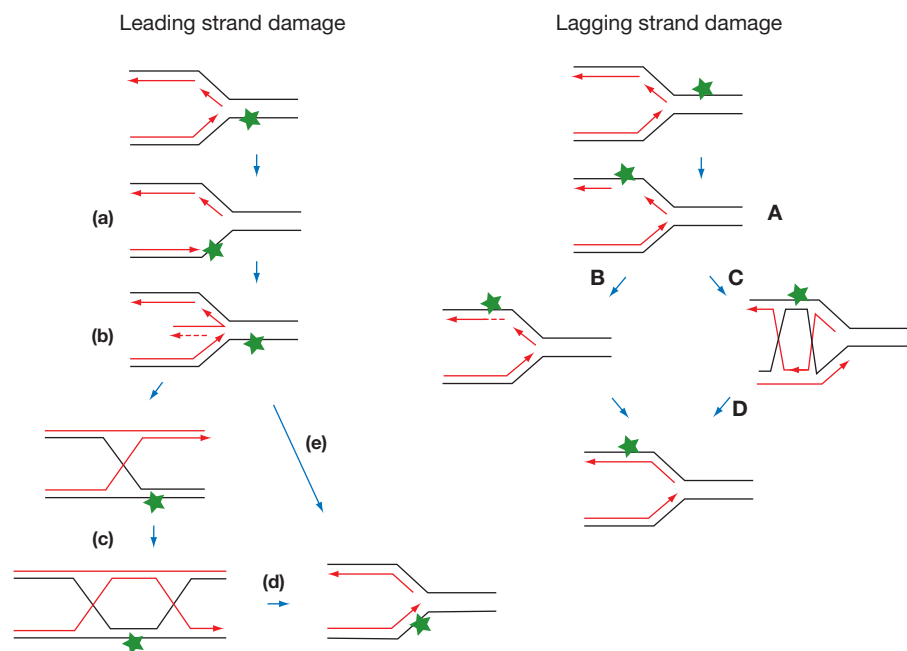


Figure 4 Potential roles for RecQ helicases in replication fork (RF) stability. (Left) If a lesion (green star) is encountered on the leading strand of an RF (a), leading and lagging strand polymerases may become uncoupled, allowing lagging strand DNA synthesis to continue. (b) RF regression, possibly catalyzed by BLM/WRN/Sgs1, allows template switching to occur with the nascent lagging strand acting as a template for leading strand synthesis. The RF can be restored either by DHJ formation (c) and DHJ dissolution, potentially catalyzed by BLM/TopoIII α /RMI1 (d), or by direct reversal of the regressed fork (e). The RF is maintained and the damage can be removed by specialized repair pathways. (Right) If an RF encounters damage on the lagging strand (A), replication is able to continue past the lesion due to the discontinuous nature of lagging strand synthesis. The resulting gap can be filled in by specialized polymerases that are able to synthesize across damaged DNA through a process termed 'translesion synthesis' (B), or a DHJ can be formed allowing the nascent lagging strand to use the nascent leading strand as a template (C). The RF can be restored following DHJ dissolution by BLM/TopoIII α /RMI1 (D).

in *SGS1* and the gene encoding the essential HR protein, Rad51, no longer contain these unresolved HR intermediates, indicating that Sgs1 is involved in the removal of these structures in a recombination-dependent manner.

Another role for RecQ helicases during replication has been suggested from studies using *E. coli* RecQ. A key question in cell biology is on how replication is completed when two RFs converge during S phase. These late replication intermediates contain intervening DNA that is yet to be replicated. By reconstituting *E. coli* DNA replication using purified proteins, it was shown that RecQ, together with Top3 and the ssDNA-binding protein SSB, is able to resolve this late replication structure via a decatenation reaction. The resulting ssDNA regions on both chromatids can then be filled in by the replicative polymerase. A similar reaction using BLM, TopoIII α , and RPA, the orthologous proteins in humans, may provide a similar result.

RecQ Helicases in Mitotic Cell Division

Studies using human cells have revealed the presence of DNA bridges that connect the separating sister chromatids during the anaphase of mitosis. These bridges are believed to represent unreplicated regions of chromosomes. Cells from BS patients have an increased frequency of these anaphase bridges, indicating that BLM has a role in ensuring the timely completion of DNA replication before the cells enter mitosis. Recently, a new class of anaphase bridges has been discovered, termed ultrafine bridges (UFBs). These UFBs can be stained along their length with antibodies against BLM, TopoIII α , and RMI1, and their number increases in cells lacking BLM. These UFBs could represent late replication intermediates that have not been resolved before the cells enter mitosis.

Conclusion

Cellular and biochemical approaches, in yeast and human cells, have provided a wealth of information on RecQ helicases in recent years. By identifying interacting proteins and revealing the activities of these enzymes, we have begun to understand how they function in the maintenance of genome stability. Their roles in suppression of illegitimate recombination and in maintaining the stability of stalled RFs are clearly crucial to ensure genome integrity and prevent tumorigenesis.

See also: **Molecular Biology:** DNA Helicases: Hexameric Enzyme Action; DNA Replication Fork, Eukaryotic; DNA Topoisomerases: Type I; Nonhexameric SF1 DNA Helicases/Translocases; Recombination: DNA Helicases and Nucleases; Recombination-Dependent DNA Replication; Telomeres: Maintenance and Replication.

Further Reading

- Chu WK and Hickson ID (2009) RecQ helicases: Multifunctional genome caretakers. *Nature Reviews Cancer* 9(9): 644–654.
- Friedberg E, Walker G, and Siede W (2006) Recombination and DNA repair. In: Craig NL (ed.) *DNA Repair and Mutagenesis*, 2nd edn. Washington, DC: American Society for Microbiology Press.
- Helleday T, Lo J, van Gent DC, and Engelward BP (2007) DNA double-strand break repair: From mechanistic understanding to cancer treatment. *DNA Repair* 6(7): 923–935.
- Hoeijmakers JH (2009) DNA damage, aging, and cancer. *New England Journal of Medicine* 361(15): 1475–1485.
- Karow JK, Chakraverty RK, and Hickson ID (1997) The Bloom's syndrome gene product is a 3'–5' DNA helicase. *Journal of Biological Chemistry* 272(49): 30611–30614.
- Wu L and Hickson ID (2003) The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* 426(6968): 870–874.
- Wu L and Hickson ID (2006) DNA helicases required for homologous recombination and repair of damaged replication forks. *Annual Review of Genetics* 40: 279–306.

Regulated Intramembrane Proteolysis

J Ye, University of Texas Southwestern Medical Center, Dallas, TX, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Jin Ye, Michael S. Brown, and Joseph L. Goldstein volume 3, pp. 665–670, © 2004, Elsevier Inc.

Glossary

Regulated intramembrane proteolysis (RIP) A process of signal transduction in which a membrane-bound protease cleaves its substrate within the lipid bilayer, thus allowing the cleaved protein fragment to function at a new location.

Rhomboid A family of intramembrane serine-proteases involved in the regulated intramembrane proteolysis (RIP) of type 1 transmembrane proteins.

Signal peptide peptidase (SPP) A family of intramembrane aspartyl-proteases involved in the RIP of type 2 transmembrane proteins.

Site-2 protease (S2P) An intramembrane metalloprotease involved in the RIP of type 2 transmembrane proteins.

γ -Secretase A protein complex containing an intramembrane aspartyl-protease catalytic subunit involved in the RIP of type 1 transmembrane proteins.

Regulated intramembrane proteolysis (RIP) is a newly recognized mechanism for signal transduction that involves the generation of regulatory molecules from membrane proteins. Such cleavage liberates cytoplasmic or luminal/extracellular fragments from transmembrane precursor proteins, allowing the cleaved fragments to function at a new location. RIP influences processes as diverse as cellular differentiation, lipid metabolism, immune defense, and the response to unfolded proteins, as originally addressed in a review article by Brown et al. In addition to its occurrence in animal cells, RIP has been observed in lower organism including bacteria, and, remarkably, the bacterial proteases are related evolutionarily to the ones used in animal cells. Proteins that are known to undergo RIP span the membrane bilayer at least once. The intramembrane cleavage of RIP is mediated by four different families of membrane-bound proteases: site-2 protease (S2P), γ -secretase, signal peptide peptidase (SPP), and rhomboid. Except for the RIP mediated by rhomboid, the intramembrane cleavage does not take place until the bulk of the protein on the extracytoplasmic (luminal or extracellular) face has been removed by a primary cleavage. This primary cleavage shortens the extracytoplasmic segment to less than 30 amino acids, which is a prerequisite for the secondary intramembrane cleavage. Numerous proteins are currently known to undergo RIP. In this article, we only discuss the RIP whose regulations and physiological/pathological importance are well characterized (Figure 1).

RIP Mediated by S2P Family

S2P is a polytopic membrane protein with the characteristics of a membrane-embedded zinc metalloprotease. Genes similar to the one encoding mammalian S2P have been identified in DNA sequences from multiple species, including archaea, bacteria, plants, and animals. The encoded proteins share a similar hydrophobicity profile that predicts a highly hydrophobic structure with multiple membrane-spanning regions. The NH₂-terminal portion of all of these proteins contains a sequence conforming to the HExxH consensus that is found in a large subfamily of zinc metalloproteases (where x is

typically a noncharged amino acid). In these proteins, the two histidines of the HExxH motif coordinate with a zinc atom, and the glutamate activates a water molecule, allowing it to make a nucleophilic attack on the peptide bond. Unlike its location in hydrophilic domains of classic zinc metalloproteases, the HExxH sequence of S2P is embedded in a highly hydrophobic segment. In addition to the HExxH motif, classic zinc metalloproteases contain a remote residue (tyrosine or aspartate) that provides an additional coordination bond for the zinc. In the S2P-like proteins, this function is believed to be filled by the aspartate of the amino acid sequence LDG, which also resides in a hydrophobic segment.

The crystal structure of the *Methanocaldococcus jannaschii* S2P, an archaeal S2P ortholog, provides a strong piece of evidence that S2P is indeed an intramembrane protease. The structure shows that the HExxH and LDG motifs are indeed embedded in transmembrane helices. The zinc ion, coordinated by the histidines in the HExxH motif and the aspartate in the LDG sequence, is sunk ~14 Å into the membrane from the cytosolic face. However, the active site surrounding the zinc ion is hydrated by a narrow water channel lined with polar residues including backbone carbonyls and charged side chains. These observations suggest that water molecules are capable to be delivered to the intramembrane active sites to catalyze the proteolytic reaction.

Proteins Cleaved by S2P in Animal Cells

The transmembrane helices cleaved by S2P adopt a type 2 orientation, that is, the NH₂-terminus of the transmembrane helices faces cytosol.

In animal cells, S2P is known to cleave three families of membrane-bound transcription factors: sterol regulatory element-binding proteins (SREBPs), activation transcription factor 6 (ATF6), and cAMP responsive element-binding protein 3 (CREB3). All of these proteins contain two conserved asparagine or proline residues in the middle of the transmembrane domain, located COOH-terminal to the cleavage site. These residues are required for S2P-catalyzed cleavage. They may serve as an NH₂-terminal cap that allows a portion of the

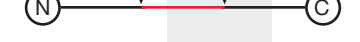
Membrane protein	Organism	Primary cleavage enzyme	Enzyme mediating rip	Regulator of cleavage	Sites of cleavage	Function or consequence of cleaved protein
<i>Rip by S2P family of proteases</i>					Extracytosolic Membrane Cytosolic	
SREBP	Mammals/ <i>drosophila</i>	S1P	S2P	Sterols/ phospholipids		Activates genes for lipid metabolism
ATF6	Mammals	S1P	S2P	ER stress		Activates genes for ER chaperones
CREB3L1	Mammals	S1P	S2P	BMP2		Activates transcription of collagen in osteoblast
CREB3L3	Mammals	S1P	S2P	IL-6		Activates genes for acute phase response
RseA	<i>E. coli</i>	DegS	YaeL	Unfolded proteins		Releases σ^E , which activates stress-induced genes
cAD1	<i>E. faecalis</i>	Signal peptidase	Eep	?		Stimulates mating
pro- σ^k	<i>B. subtilis</i>	None	SpoIVFB	SpoIVB		Activates genes for sporulation
<i>Rip by γ-Secretase (presenillin)</i>						
APP	Mammals	β -Secretase	γ -Secretase	?		Generates A β peptide
Notch	Mammals/ <i>drosophila</i>	TACE	γ -Secretase	Delta		Activates genes for differentiation
ErbB-4	Mammals	TACE	γ -Secretase	Heregulin		Phosphorylates nuclear proteins that inhibit cellular growth
<i>Rip by SPP family of proteases</i>						
MHC, class I	Mammals	Signal peptide	SPP	?		Generates peptide for antigen presentation by HLA-E
TNF α	Mammals	TACE	SPPL2a/b	LPS		Activates transcription of IL-12 in dendritic cells
<i>Rip by rhomboid family</i>						
Spitz	<i>Drosophila</i>	None	Rhomboid	Star		Generates ligand for EGF receptor
TatA	<i>P. stuartii</i>	None	AarA	?		Active TatA channel required for quorum sensing
					Mitochondria intermembrane space Inner membrane Mitochondria matrix	
Mgm1	<i>S. cerevisiae</i>	None	Rbd1	Mitochondria membrane potential		Maintain mitochondria morphology
HtrA2	Mammals	None	PARL	?		Inhibit apoptosis by preventing accumulation of BAX

Figure 1 Membrane proteins that undergo regulated intramembrane proteolysis (RIP) categorized according to the RIP protease. The cleaved protein fragments are highlighted in red. For simplicity, proteins are not drawn to scale, and we do not show the alternate pathway for APP processing, which involves cleavage by TACE. For Notch, we show only the transmembrane subunit.

transmembrane α -helix to unwind partially to expose the peptide bond for cleavage by S2P.

Sterol regulatory element-binding proteins

SREBPs are membrane-bound transcription factors that regulate the synthesis of cholesterol, a major component of membranes in mammalian cells. As shown in **Figure 1**, SREBPs are inserted in membranes of the endoplasmic reticulum (ER) in a helical hairpin fashion. The cytosolic NH₂-terminal domain is

a transcription factor of the basic-helix-loop-helix-leucine zipper family. The COOH-terminal domain, also cytosolic, forms a complex with a polytopic membrane protein called Scap. When cells are overloaded with sterols, the Scap/SREBP complex is trapped in the ER owing to its interaction with Insig proteins, which are also polytopic ER membrane proteins. The ER retention of SREBPs prevents their cleavage as proteases that cleave SREBPs reside in the Golgi complex. Upon deprivation of sterols, Scap is dissociated from Insig proteins, a reaction

allowing Scap to transport SREBPs from the ER to Golgi, where cleavage occurs. Inasmuch as SREBPs activate genes encoding enzymes of cholesterol synthesis, this sterol-mediated block in transport allows cholesterol to inhibit its own synthesis in a classic feedback fashion.

The first cleavage of SREBPs is catalyzed by site-1 protease (S1P), a membrane-bound serine protease that is oriented with its active site in the lumen of the Golgi complex. Cleavage by S1P separates the two transmembrane helices of SREBPs, but the NH₂-terminal fragment remains membrane bound until it is cleaved by S2P at a site that is three residues within the transmembrane segment (Figure 1). Even though the activity of S2P is not regulated directly by sterols, S2P does not act until the two transmembrane segments have been separated by S1P, which effectively brings this reaction under cholesterol control.

In *Drosophila*, orthologs of SREBP, Scap, S1P, and S2P are all expressed even though *Drosophila* cells do not produce sterols. *Drosophila* SREBP is processed by S1P and S2P in a reaction that requires Scap and presumably requires ER to Golgi transport. The major SREBP targets in *Drosophila* S2 cells are enzymes required for synthesis of saturated fatty acids, which are incorporated into the major phospholipid, phosphatidyl ethanolamine (PE). PE, rather than sterols, is the feedback regulator of SREBP cleavage in these cells.

Activation transcription factor 6

ATF6 is a type 2 membrane protein with a single transmembrane domain (Figure 1). Its NH₂-terminal cytosolic domain is a transcription factor of the bZIP family. Under resting condition, ATF6 is not cleaved as it is retained in the ER through its association with an ER chaperon, BiP. Upon accumulation of unfolded proteins in the ER, ATF6 is dissociated from BiP. Consequently, ATF6 is translocated from the ER to Golgi, where it is cleaved sequentially by S1P and S2P, the same proteases that process SREBP. The intramembrane proteolysis carried out by S2P liberates the NH₂-terminal domain of ATF6 from the membrane, allowing it to activate transcription of *BiP* and other genes whose products assist the folding of ER proteins. This regulatory pathway, called the unfolded protein response, helps cells to survive conditions of ER stress.

ATF6 β is a close homolog of ATF6 in mammalian cells. It is also cleaved by S1P and S2P in response to ER stress, but the cleaved nuclear form does not activate transcription of genes that facilitate folding of proteins in the ER. Instead, the cleaved nuclear form of ATF6 β inhibits transcription of genes activated by the nuclear form of ATF6 in a dominant negative manner. Thus, ATF6 β may serve as a transcriptional repressor functioning in part to regulate the strength and duration of ATF6 activity during the ER stress response.

cAMP responsive element-binding protein 3

There are five proteins (CREB3, CREB3-like 1 (CREB3L1), CREB3L2, CREB3L3, and CREB3L4) that belong to the CREB3 transcription factor family in mammalian cells. Similar to ATF6, all of them contain a single transmembrane helix with an NH₂-terminal cytosolic domain resembling a transcription factor of the bZIP family. Among these proteins, only the functions of CREB3L1 and CREB3L3 are characterized.

CREB3L1, also named as OASIS, plays an important role in bone development. In response to bone morphogenetic

protein-2 (BMP-2), a cytokine required for bone formation and osteoblast differentiation, CREB3L1 in osteoblast is cleaved sequentially by S1P and S2P. The cleaved NH₂-terminal fragment then enters the nucleus to drive transcription of *coll1a1*, the major collagen found in bone matrix (Figure 1). Cleavage of CREB3L1 can also be stimulated by drugs triggering ER stress in cultured cells. However, whether RIP of CREB3L1 is also activated by ER stress *in vivo* remained to be determined.

CREB3L3, also named as CREB-H, is exclusively expressed in livers. It plays an important role in innate immunity by producing acute phase proteins. Microbial infection stimulates the production of pro-inflammatory cytokines such as interleukin-6 (IL-6) by leukocytes. IL-6 induces ER stress in hepatocytes, which in turn triggers the RIP of CREB3L3 by S1P and S2P. The cleaved NH₂-terminal nuclear form then activates transcription of C-reactive protein and other acute phase proteins in livers (Figure 1). These acute phase proteins are then secreted out of livers to provide early defense against microbial infection.

Proteins Cleaved by S2P in Bacteria

RseA

The RIP of RseA in bacteria is closely related to the RIP of mammalian ATF6. In *Escherichia coli*, σ^E is the transcription factor that regulates extracytoplasmic stress response. Under nonstress conditions, σ^E is negatively regulated by RseA, an inner membrane protein with one transmembrane domain (Figure 1). The NH₂-terminal cytosolic domain of RseA binds to σ^E and is sufficient to inhibit σ^E activity. When unfolded proteins accumulate in the periplasmic space (similar to ER lumen in eukaryotic cells), RseA is first cleaved by DegS, a membrane-bound serine protease with its active site projecting into the periplasmic space. Cleavage by DegS removes the majority of the COOH-terminal periplasmic domain of RseA. YaeL, an ortholog of S2P, then cleaves the remaining part of RseA, after which RseA is rapidly degraded. This frees σ^E to associate with RNA polymerase and direct transcription of its target genes.

The similarity between mammalian and bacteria unfolded-protein response suggests that S2P-mediated RIP has been selected to cope with extracytoplasmic stress during evolution, although individual organisms use the mechanism in different ways: in mammalian cells RIP directly releases the active transcription factor from its substrate, whereas in bacteria RIP leads to degradation of a membrane-bound inhibitor of the responsible transcription factor.

cAD1 precursor

Enterococcus faecalis, a Gram-positive bacterium, secretes an eight-amino-acid peptide pheromone called cAD1, which induces a mating response in other enterococci that harbor a plasmid called pAD1 (Figure 1). Genomic sequencing revealed that the octapeptide pheromone is derived from a 143-amino-acid precursor that contains a signal peptide at the NH₂-terminus and is inserted into the plasma membrane with a type 2 orientation. The pheromone corresponds to the COOH-terminal eight amino acids of the signal peptide (Figure 1). The octapeptide is generated by cleavage of the

precursor at two sites, at the extracellular side of the membrane in a reaction carried out by signal peptidase and at a site in the middle of the transmembrane signal peptide by Eep, an S2P ortholog. This system differs from the others in that the active fragment is released by S2P into the extracellular space rather than the cytosol.

Pro- σ^K

Another prokaryotic relative of S2P is a *Bacillus subtilis* protein called spoIVFB. This protein contains an HExxH and an LDG motif in hydrophobic segments, but the overall hydrophobicity profile of SpoIVFB differs from those of the other family members. SpoIVFB is the protease that removes a membrane-embedded NH₂-terminal hydrophobic peptide from a transcription factor, pro- σ^K , thereby releasing the factor into the cytosol and allowing it to activate gene transcription. This process is necessary for completion of spore formation in response to nutrient deprivation. The membrane orientation of pro- σ^K is opposite to the type 2 orientation of other known S2P substrates, and it is the only S2P substrate that does not require a primary cleavage (Figure 1). These differences in catalytic requirements may account for the differences in structure between SpoIVFB and the other members of the S2P family.

RIP Mediated by γ -Secretase

γ -Secretase is a membrane-bound complex consisting of several proteins, including presenilin, nicastrin, anterior pharynx-1, and presenilin enhancer-2 protein. In mammalian cells, there are two presenilin isoforms encoded by separate genes: presenilin-1 and presenilin-2. These two proteins are highly homologous and they function redundantly in γ -secretase-mediated RIP. Presenilin contains two conserved aspartate residues that are located in two of its nine transmembrane helices at positions that are predicted to place them at the same depth in the membrane. This property is reminiscent of soluble aspartyl proteases, which contain two closely apposed aspartates that are required for activity. Experiments with active-site-modifying agents specific for aspartyl proteases provide evidence that presenilin contains the active proteolytic site. Thus, presenilin is considered to be the catalytic subunit in γ -secretase.

γ -Secretase cleaves transmembrane helices that adopt the type 1 topology, that is, the COOH-terminus of the transmembrane segments exposes to cytosol. Numerous proteins are cleaved by γ -secretase. Three of them will be discussed here in detail to illustrate the RIP mediated by γ -secretase.

Amyloid Precursor Protein

Amyloid precursor protein (APP) is a type 1 membrane protein with an NH₂-terminal extracellular domain and a COOH-terminal cytosolic tail (Figure 1). Processing of APP generates a toxic amyloid β peptide that is responsible for Alzheimer's disease. APP is first cleaved at a site that is 28 amino acids from the transmembrane helix by β -secretase, a membrane-bound aspartyl protease with an extracytoplasmic active site. This cleavage shears off the bulk of the extracellular domain of

APP, which then allows the protein to be cleaved within the membrane by γ -secretase. Intramembrane cleavage can occur at either of the two sites separated by two amino acids, leading to two different amyloid β peptides, designated A β _{1–40} and A β _{1–42}, which accumulate extracellularly. After the intramembrane γ -secretase cleavage, the cytosolic tail of APP enters the nucleus where it may affect transcription of genes.

In addition to cleavage by β -secretase, APP can be cleaved by tumour necrosis factor α (TNF α) converting enzyme (TACE), a membrane-bound metalloprotease, whose active site faces the extracellular surface. TACE cuts the extracytosolic domain of APP even closer to the membrane, leaving only 12 amino acids on the external surface. Like the β -secretase cleavage, TACE cleavage is followed by γ -secretase cleavage. In this case, the liberated fragment is too short to form an amyloid deposit, and thus cleavage by TACE does not lead to Alzheimer's disease.

Notch

The cell-surface receptor Notch is synthesized as a type 1 membrane protein (Figure 1) that is processed constitutively by a furin-like enzyme in the secretory pathway. The enzyme cleaves the precursor to generate two subunits: an extracellular subunit and a transmembrane subunit, which remain associated as a noncovalent heterodimer. The heterodimer travels to the cell surface and remains intact until it binds its ligand Delta, a membrane protein that resides on the surface of an adjacent cell. Binding leads to cleavage of the transmembrane subunit of Notch by TACE, the same enzyme that cleaves APP. TACE cleavage releases most of the extracellular portion of the transmembrane subunit of Notch along with the attached extracellular subunit. The shortened transmembrane subunit is then cleaved by γ -secretase within the membrane-spanning helix, liberating a cytosolic fragment. The cytosolic fragment translocates to the nucleus where it activates several genes whose products influence the fate of cells during development.

ErbB-4

ErbB-4 is a transmembrane receptor tyrosine kinase that regulates cell proliferation and differentiation. It adopts a type 1 orientation with an NH₂-terminal extracellular ligand-binding domain and a COOH-terminal cytosolic tyrosine kinase domain (Figure 1). Upon binding to its ligand, heregulin, the ErbB-4 ectodomain is cleaved by TACE, followed by intramembrane proteolysis carried out by γ -secretase. The intramembrane cleavage results in the release of the cytosolic fragment, which travels to the nucleus where it phosphorylates substrates that regulate cell growth.

RIP Mediated by SPP

Five human SPPs have been identified: SPP, SPPL2a, SPPL2b, SPPL2c, and SPPL3. SPPs are remotely homologous to presenilin. They also contain two conserved aspartate residues that form the catalytic active site within the predicted transmembrane domains. However, topology studies revealed that the orientation of these transmembrane regions is inverted related

to that of presenilin. As a result, SPPs cleave type 2 membrane proteins instead of type 1 membrane proteins, which are cleaved by presenilin. While SPP and SPPL3 locate in the ER, SPPL2a and SPPL2b reside in endosomes and lysosomes. The difference in their localization is consistent with the findings that these proteases do not share the same substrates.

The best-studied example of RIP mediated by SPP is the proteolytic processing of signal peptides from MHC class I molecules such as HLA-A, -B, -C, and -G. These molecules are expressed with a typical signal sequence for targeting to the secretory pathway. During their translocation through the ER membrane, the signal sequences are cleaved off from the pre-protein by signal peptidase. The cleaved signal peptides, which remain membrane-bound with a type 2 orientation, are then cleaved by SPP in the middle of the membrane to liberate the NH₂-terminal half of the signal peptides into the cytosol (Figure 1). These cytosolic fragments are then transported into the ER lumen where they bind to HLA-E, a nonclassical MHC class I molecule. The HLA-E/peptide complexes travel to the cell surface where they bind to cluster of differentiation 94 (CD94) receptors on natural killer (NK) cells and inhibit NK cell-mediated lysis. This pathway protects cells expressing normal MHC class I molecules from killing by NK cells.

SPP is also involved in the maturation of core protein encoded by hepatitis C virus (HCV). The core protein corresponds to the first 169 amino acids of the viral polyprotein. It is localized at cytosolic side of the ER membrane and is linked to the remainder of the viral protein via a signal peptide-like transmembrane sequence at its COOH terminus. After cleavage by signal peptidase in the ER lumen, SPP cleaves the signal peptide-like transmembrane domain. Consequently, core protein is released from the ER and binds to lipid droplet where it plays an important role in assembly of viral particles.

SPPL2a and SPPL2b play an important role for activated dendritic cells to produce interleukin-12 (IL-12). Upon activation by Lipopolysaccharide, dendritic cells produce TNF α that is anchored to plasma membrane with a type 2 orientation (Figure 1). Following cleavage by TACE to shed the majority of the extracellular domain, SPPL2a/b cleaves inside the transmembrane domain to release the intracellular domain of TNF α from membranes, allowing it to enter the nucleus to activate transcription of IL-12. IL-12 secreted from activated dendritic cells promotes differentiation of T-helper 1 (Th1) cells and is essential for the induction of Th1-mediated adaptive immunity.

RIP Mediated by Rhomboid

Apart from S2P, rhomboid is the only other intramembrane protease whose crystal structure has been solved. Rhomboid is an atypical serine protease that contains a serine and histidine catalytic dyad. It does not have the third aspartic residue in the active site to form the catalytic triad, which is the signature for most serine proteases. The structure shows that the active site of rhomboid is embedded in the transmembrane domains. The catalytic serine locates at the bottom of a funnel-shaped cavity that is lined with polar residues and opens to outside of a cell. This structure ensures that water is accessible for the intramembrane cleavage reaction.

Rhomboid belongs to a large gene family that is conserved throughout archaea, bacteria, yeast, plants, and animals, including humans. The function of rhomboid homologs in bacteria has only been determined for AarA in Gram-negative bacterium *Providencia Stuartii*. AarA is a protein required for bacteria intercellular signaling known as 'quorum sensing'. The substrate for AarA is TatA, which contains a single transmembrane domain at its NH₂-terminus (Figure 1). Cleavage within the transmembrane domain by AarA is required for the activation of TatA, allowing the protein to form a channel that exports proteins across the bacterial membrane into the periplasm for signaling.

In eukaryotic cells, rhomboids are divided into two major subfamilies: secretase rhomboids, which reside in the secretory pathway, and PARL-type rhomboids, which locate in mitochondrial. To date, secretase rhomboid has only been characterized in *Drosophila* where it initiates epidermal growth factor receptor (EGFR) signaling. Spitz is the main ligand for *Drosophila* EGFR. It is synthesized as a transmembrane precursor with an NH₂-terminal extracellular EGF-like domain and a short COOH-terminus cytosolic tail (Figure 1). Spitz is confined to the ER and is inert until it binds to another protein named star. Star escorts spitz from ER to Golgi where spitz is cleaved by rhomboid. The intramembrane cleavage liberates the extracellular EGF-like domain from the membrane, allowing it to be secreted out of the cell.

The functions of mitochondrial rhomboids have been characterized in both yeast and mammalian cells. In yeast, the major substrate for the mitochondrial rhomboid, Rbd1, is Mgm1, a dynamin-like GTPase. Mgm1 is required to maintain the proper mitochondria morphology, and this function requires both the full length protein inserted into inner mitochondrial membrane and a processed form located in the intermembrane space. The full-length Mgm1 is inserted into the inner mitochondrial membrane through a single transmembrane domain, with its NH₂ and COOH-terminus residing in intermembrane space and mitochondrial matrix, respectively (Figure 1). Interestingly, the Rbd1 cleavage site is not in this transmembrane domain. Instead, it is located in a secondary hydrophobic site that is initially located in the intermembrane space. With the energy provided by ATP hydrolysis, mitochondrial import motor pulls the original transmembrane domain out of the inner membrane into the matrix, allowing the secondary hydrophobic domain to be inserted into the membrane. This cleavage mechanism is thought to provide an opportunity to regulate the balance between the processed and full length Mgm1 in response to mitochondrial membrane potential, which provides the energy source for the shifting in the transmembrane segment.

Mgm1 is the yeast homolog of the mammalian protein Optic Atrophy 1 (OPA1). However, OPA1 is not a good substrate for mammalian mitochondrial rhomboid presenilins-associated rhomboid-like protein (PARL). Rather than maintaining the mitochondrial morphology, PARL controls apoptosis, and this function is achieved by its cleavage of HtrA2. HtrA2 is synthesized as an inactive precursor inserted into mitochondrial inner membrane in a way similar to the yeast Mgm1. Cleavage by PARL releases the NH₂-terminal fragment of HtrA2 into intermembrane space (Figure 1). This processed form of HtrA2 inhibits apoptosis by preventing the accumulation of activated

Bax, which initiates apoptosis. The site where HtrA2 is cleaved by PARL is outside the transmembrane domain, suggesting that the cleavage mechanism employed by PARL may be similar to that used by Rbd1 to cleave Mgm1.

See also: [Metabolism Vitamins and Hormones: Cholesterol Synthesis and Regulation](#); [Protein/Enzyme Structure Function and Degradation: Amyloid](#); [Secretases](#).

Further Reading

- Alba BM, Leeds JA, Onufryk C, Lu CZ, and Gross CA (2002) DegS and YaeL participate sequentially in the cleavage of RseA to activate the σ^E -dependent extracytoplasmic stress response. *Genes Development* 16: 2156–2168.
- Brown MS, Ye J, Rawson RB, and Goldstein JL (2000) Regulated intramembrane proteolysis: A control mechanism conserved from bacteria to humans. *Cell* 100: 391–398.
- Freeman M (2008) Rhomboid proteases and their biological functions. *Annual Review of Genetics* 42: 191–210.
- Golde TE, Wolfe MS, and Greenbaum DC (2009) Signal peptide peptidases: A family of intramembrane-cleaving proteases that cleave type 2 transmembrane proteins. *Seminars in Cell and Development Biology* 20: 225–230.
- Haze K, Yoshida H, Yanagi H, Yura T, and Mori K (1999) Mammalian transcription factor ATF6 is synthesized as a transmembrane protein and activated by proteolysis in response to endoplasmic reticulum stress. *Molecular Biology of the Cell* 10: 3787–3799.
- Lee JR, Urban S, Garvey CF, and Freeman M (2001) Regulated intracellular ligand transport and proteolysis control EGF signal activation in *Drosophila*. *Cell* 107: 161–171.
- Lemberg MK and Martoglio B (2002) Requirements for signal peptide peptidase-catalyzed intramembrane proteolysis. *Molecular Cell* 10: 735–744.
- McCarthy JV, Twomey C, and Wujek P (2009) Presenilin-dependent regulated intramembrane proteolysis and γ -secretase activity. *Cellular and Molecular Life Sciences* 66: 1534–1555.
- Ni C-Y, Murphy MP, Golde TE, and Carpenter G (2001) γ -Secretase cleavage and nuclear localization of ErbB-4 receptor tyrosine kinase. *Science* 294: 2179–2181.
- Rawson RB, Zelenski NG, Nijhawan D, et al. (1997) Complementation cloning of *S2P*, a gene encoding a putative metalloprotease required for intramembrane cleavage of SREBPs. *Molecular Cell* 1: 47–57.
- Rudner DZ, Fawcett P, and Losick R (1999) A family of membrane-embedded metalloproteases involved in regulated proteolysis of membrane-associated transcription factors. *Proceedings of the National Academy of Sciences of the United States of America* 96: 14765–14770.
- Selkoe DJ and Podlisny MB (2002) Deciphering the genetic basis of Alzheimer's disease. *Annual Review of Genomics and Human Genetics* 3: 67–99.
- Sinisa U and Shi Y (2008) Core principles of intramembrane proteolysis: Comparison of rhomboid and site-2 family proteases. *Current Opinion in Structural Biology* 18: 432–441.
- Ye J, Rawson RB, Komuro R, et al. (2000) ER stress induces cleavage of membrane-bound ATF6 by the same proteases that process SREBPs. *Molecular Cell* 6: 1355–1364.

Regulation of Body Weight by Malonyl-CoA in the CNS

S Rodriguez and M J Wolfgang, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Energy homeostasis The homeostatic regulation of energy balance.

Intermediary metabolite An intermediary step in a metabolic process.

Knockout mouse Term used for the homologous recombination and inactivation of an endogenous gene.

Neuropeptides Small protein signaling molecules secreted by neurons to facilitate endocrine neural communication.

Nutrient sensing The direct sensing and signaling by nutrients in contrast to endocrine signaling.

Introduction

Malonyl-CoA is a key intermediary metabolite in fatty acid synthesis. In *de novo* fatty acid synthesis, malonyl-coenzyme A (CoA) is the substrate that provides the primary carbon source for the formation of palmitate (C16) catalyzed by fatty acid synthase (FASN). Malonyl-CoA acts also as an allosteric inhibitor of carnitine palmitoyltransferase-1 (CPT1) and therefore fatty acid oxidation. The role of malonyl-CoA and fatty acids has been very well defined in lipogenic tissues. The role of *de novo* lipogenesis in the brain has been largely overlooked as the brain primarily utilizes glucose as its energy source and does not store energy in the form of lipid. Furthermore, neurons are thought to have little capacity for fatty acid oxidation. Our understanding of the neural fatty acid metabolism has been challenged by emerging data that suggest a new role for these molecules in the brain. Areas within the central nervous system (CNS), in particular, the hypothalamus, can sense peripheral circulating glucose and lipids and respond to fluctuations in these nutrients by modifying food intake or peripheral energy expenditure via altering the malonyl-CoA concentration.

Regulation of Cellular Fatty Acid Metabolism

In lipogenic tissues such as the liver, *de novo* lipogenesis is regulated by the activity of acetyl-CoA carboxylase (ACC). ACC, the rate-limiting and highly regulated step, carboxylates acetyl-CoA to form malonyl-CoA in an adenosine triphosphate (ATP)-consuming step. ACC is positively regulated by the tricarboxylic acid (TCA) intermediate citrate, which is also the carbon source for cytoplasmic acetyl-CoA after cleavage by ATP:citrate lyase. ACC is inhibited by phosphorylation by adenosine monophosphate (AMP)-activated protein kinase (AMPK). The activity of AMPK is enhanced by the binding of AMP and is highly responsive to the ratio of AMP:ATP. Therefore, when cellular energy is in deficit, fatty acid synthesis is turned off.

The product of ACC, malonyl-CoA, is used by FASN as the chain elongation unit for the reiterative reductive biosynthesis of long-chain fatty acids. Malonyl-CoA serves another important regulatory role in fatty acid metabolism by

allosterically inhibiting CPT1. CPT1 is the rate-limiting step in the oxidation of long-chain fatty acyl-CoAs as it regulates the entry of fatty acyl-CoAs into the mitochondrial matrix. The dual roles of malonyl-CoA as the carbon source for *de novo* fatty acid synthesis and inhibitor of oxidation ensure the proper regulation of fatty acid metabolism. In other words, malonyl-CoA provides the metabolic logic for fatty acid metabolism.

Humans with mutations in the liver isoform of CPT1, CPT1a, result in an inability of the liver to efficiently buffer blood glucose via gluconeogenesis. The oxidation of long-chain fatty acids in the liver upon fasting also enables the production of ketone bodies to provide energetic substrates for neural metabolism during prolonged fasting. Fatty acid oxidation in muscle is used as a primary energy source in the absence of insulin-stimulated glucose uptake. Muscle does not, however, have the ability to produce fatty acids *de novo* due to the lack of FASN. Muscle expresses a unique CPT1, CPT1b, and a unique ACC, ACCbeta. ACCbeta performs the same enzymatic reaction and is regulated in a similar fashion as the liver isoform, ACCalpha. The main difference is that ACCalpha is a soluble cytoplasmic enzyme and ACCbeta is tethered to the outer mitochondrial membrane and therefore juxtaposed to CPT1b. Therefore, the function of ACC and malonyl-CoA in muscle is solely to regulate the oxidation of exogenous fatty acids. To deplete and regulate malonyl-CoA, as no FASN exists, muscle expresses malonyl-CoA decarboxylase to decarboxylate malonyl-CoA to acetyl-CoA and CO₂.

The Hypothalamus Is Important for the Control of Food Intake and Energy Homeostasis

The control of energy homeostasis is critical for the maintenance and survival of any organism. Disruptions in energy homeostasis can often lead to chronic disease states. When energy intake exceeds expenditure, fatty acids in the form of triacylglycerol are stored in adipose tissue. Intricate control systems have evolved to maintain energy homeostasis by inhibiting food intake and increasing energy expenditure during periods of energy excess and increasing food intake and decreasing energy expenditure during lean times. More recently, the

mammalian CNS has garnered much attention as a site that regulates the energy-sensing needs of an organism by altering hunger/satiety and communicating this need directly via neural circuits to peripheral energy centers to mediate fuel utilization.

The hypothalamus is the specialized region within the brain which contains distinct nuclei to control energy homeostasis and food intake. Other areas of the brain also contribute to the control of energy homeostasis. Chemical, genetic, or physical lesioning of the hypothalamus results in severe deficiencies in body weight regulation. The arcuate nucleus, ventral medial nucleus, and paraventricular nucleus among others have been identified as the key centers that respond to changes in energy status by modifying the expression of specific neuropeptides that alter food intake and energy expenditure. Also, the alteration of synaptic connectivity in the hypothalamus has been shown to be an important component in body weight regulation. The hypothalamus can respond to circulating endocrine hormones including leptin, ghrelin, and insulin, secreted by peripheral tissues. The hypothalamus contains at least two distinct neuronal populations that integrate these circulating cues. One set of neurons express the orexigenic (feeding promoting) neuropeptides such as agouti-related protein (AgRP) and neuropeptide Y (NPY). The second neuronal population within the hypothalamus expresses the anorexigenic (feeding inhibiting) neuropeptides, such as alpha melanocyte-stimulating hormone (alpha MSH), and the cocaine and amphetamine-related transcript (CART). The paradigm follows: when energy intake exceeds expenditure, the expression of NPY and AgRP neuropeptides decrease and alpha MSH and CART increase; when energy expenditure exceeds intake, the expression of CART and alpha MSH neuropeptides increases and NPY and AgRP decreases. The first-order neurons in the hypothalamus are under the primary control of circulating hormones, and integrate signals to higher brain centers where this information is integrated and behavioral responses are formulated.

The Role of Central FASN in Regulating Body Weight

Fatty acids can be imported into the brain through passive diffusion across the blood-brain barrier or generated through *de novo* synthesis. The interest in fatty acids playing a role in food intake and energy expenditure arose from the discovery that inhibitors of FASN, C75 and cerulenin, cause a dramatic weight loss in animal studies. These inhibitors rapidly induce a reduction in food intake and body weight loss in obese mice. These inhibitors were efficacious for models of diet-induced obesity (DIO) and genetic forms of obesity such as the leptin (*ob/ob*) or leptin receptor (*db/db*) mutant mice. Elucidation of the mechanism of FAS inhibitors is of great interest in the development of pharmacological targets to counter obesity.

The effect of the inhibitors is rapid and suppresses food intake in less than 20 min. Long-term (30-day study) administration of a low-dose treatment of *ob/ob* mice resulted in an almost complete reversal of obesity. FASN inhibition studies have shown that central or peripheral administration rapidly affects the levels of hypothalamic neuropeptides that regulate feeding behavior. Concomitant with a decrease in food intake

is a decrease in the feeding promoting neuropeptides, NPY/AgRP, while levels of the food-inhibiting POMC/CART neuropeptides increase (POMC, proopiomelanocortin). These inhibitors were found to rapidly inhibit NPY production in the arcuate nucleus of mice even after 23 h of fasting, usually a time when NPY production is highly upregulated to stimulate food intake.

FASN has been shown to colocalize with NPY neurons in the hypothalamus. The loss of the FASN gene via homologous recombination is lethal with even heterozygotes showing a loss of viability. A conditional loss of function allele was made and FASN was specifically deleted from the hypothalamus. These mice phenocopied the effects of the inhibitor. The mice were hypophagic, lean, and had altered neuropeptide expression. Furthermore, these mice also exhibit an increase in energy expenditure mediated by an increase in activity. FASN inhibitors can further affect weight loss by increasing energy expenditure and whole-body oxygen consumption. Pair feeding experiments in which control mice are restricted to the intake of FASN inhibitor-treated mice demonstrate that treated mice lose more weight; they increase their energy expenditure in peripheral tissues, as measured by whole-body calorimetry studies. Central administration of FASN inhibitors to lean or obese mice rapidly activates fatty acid oxidation in skeletal muscle, therefore mediating a major impact in whole-body energy expenditure.

Hypothalamic Malonyl-CoA

Inhibiting FASN either pharmacologically or genetically reduces body weight. The inhibition of FASN causes an increase in its substrate malonyl-CoA. Since malonyl-CoA serves a signaling role in muscle, it was hypothesized that malonyl-CoA could be serving a signaling role in the hypothalamus to regulate body weight. Indeed, the simultaneous inhibition of ACC and FASN reverses the weight-reducing effects of FASN inhibitors. Furthermore, exogenous addition of malonyl-CoA decarboxylase to the hypothalamus results in an increase in body weight. Finally, malonyl-CoA is dynamically regulated by meal state. Fasting suppresses hypothalamic malonyl-CoA and refeeding causes a spike in hypothalamic malonyl-CoA concentration. These experiments strongly suggest that malonyl-CoA is a major nexus for the regulation of body weight (Figure 1).

Regulation of Hypothalamic Malonyl-CoA

The concentration of malonyl-CoA fluctuates between periods of nutritional surplus and shortage. The initial signal for this

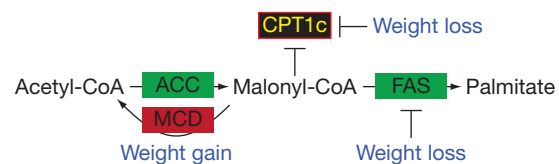


Figure 1 Phenotypic outcomes of modulating *de novo* fatty acid synthesis in the CNS. Inhibition of FASN results in weight loss, and promoting malonyl-CoA decarboxylation results in weight gain. Loss of CPT1c results in mice with a reduction in food intake and weight loss.

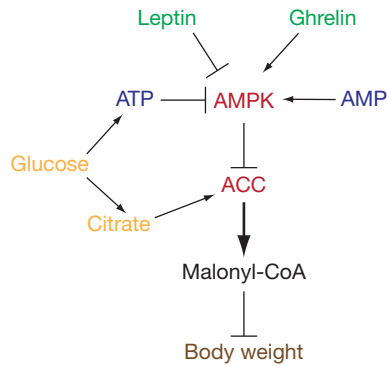


Figure 2 The regulation of hypothalamic malonyl-CoA. ACC is regulated by citrate as an allosteric activator and inhibited by AMPK phosphorylation. Glucose flux through the mitochondria produces both citrate and ATP. A high AMP:ATP ratio activates AMPK activity while a low AMP:ATP ratio is inhibitory. AMPK activity can also be modulated by anorexogenic and orexogenic endocrine molecules such as leptin and ghrelin.

nutrient-sensing pathway appears to be through the modulation of AMP-dependent protein kinase (AMPK) activity. AMPK differentially binds two molecules of AMP or ATP. ATP-bound AMPK is inactive while the AMP-bound form is activated by upstream kinases. Coupled with the action of adenylate kinase, the AMP:ATP ratio serves as a biochemical signal for AMPK. Changes in the hypothalamic AMP:ATP ratio and AMPK activity via differential nutrient paradigms impact feeding behavior and body weight, although it remains unclear what the cell-type-specific effect is. There is also cross-talk between endocrine signaling pathways in the hypothalamus such as leptin, ghrelin, and adiponectin and the nutrient-sensing pathway as all of these molecules effect AMPK activity. AMPK was first characterized as the kinase responsible for phosphorylating and thus inhibiting ACC and HMG-CoA reductase (HMG-CoA, 3-hydroxy-3-methylglutaryl-coenzyme A). Therefore, AMPK inhibits the rate-limiting steps in the synthesis of fatty acids and cholesterol. During times of low cellular energy, synthesis is inhibited. ACC catalyzes the biotin-dependent carboxylation of acetyl-CoA to malonyl-CoA in the highly regulated and committed step in fatty acid synthesis. Therefore, the activity of AMPK directly impacts the concentration of malonyl-CoA. Hypothalamic malonyl-CoA concentration serves as a satiety signal and is directly proportional to the activity of AMPK and ACC (Figure 2).

Carnitine Palmitoyltransferase-1c

Malonyl-CoA is essential for the regulation of mitochondrial fatty acid beta oxidation due to its role as an allosteric inhibitor of CPT1 enzymes. Neurons do not robustly oxidize fatty acids and most neurons have a low level of CPT1a and CPT1b. Neurons do have, however, a neuron-specific CPT1 gene, CPT1c. CPT1c is expressed throughout the brain and is expressed in

hypothalamic nuclei known to be important for body weight regulation. CPT1c binds malonyl-CoA at a K_d equivalent to CPT1a, but does not catalyze acyl transfer to carnitine. In fact, CPT1c is not localized to mitochondria, but is integral to the endoplasmic reticulum. Therefore, it has been hypothesized that CPT1c performs a novel enzymatic reaction that is essential to controlling body weight. When CPT1c is knocked out in mice, they have a low body weight and reduced food intake. This is consistent with CPT1c being the target of malonyl-CoA in the hypothalamus that controls body weight. Furthermore, it was shown that CPT1c knockout (KO) mice are susceptible to weight gain on a high fat diet and exhibited reduced whole-body energy expenditure. Taken together, malonyl-CoA regulation of CPT1c represents a new signaling paradigm in body weight regulation.

Conclusion

The control of body weight by intermediary metabolites of fatty acid biosynthesis is a new concept in nutrient signaling. Malonyl-CoA serves as a central signal in the CNS to convey organismal energy status and modify body weight accordingly. The *in vivo* physiologic role of this pathway has been established with nutrients and endocrine signals playing regulatory roles. Hypothalamic malonyl-CoA concentration can be affected by glucose concentration through modulating AMPK activity or by endocrine signals such as leptin and ghrelin. The mechanism by which malonyl-CoA effects change in body weight are in part through modulating the novel neuron-specific CPT1c. Finally, this newly identified pathway provides new therapeutic targets for the treatment of obesity.

See also: **Metabolism Vitamins and Hormones: Branched-Chain Amino Acids; Cholesterol Synthesis and Regulation; Fatty Acid Oxidation.**

Further Reading

- Chakravarthy MV, Zhu Y, Lopez M, et al. (2007) Brain fatty acid synthase activates PPARalpha to maintain energy homeostasis. *Journal of Clinical Investigation* 117: 2539–2552.
- Gao Q and Horvath TL (2007) Neurobiology of feeding and energy expenditure. *Annual Review of Neuroscience* 30: 367–398.
- He W, Lam TK, Obici S, and Rossetti L (2006) Molecular disruption of hypothalamic nutrient sensing induces obesity. *Nature Neuroscience* 9: 227–233.
- Lam TK, Schwartz GJ, and Rossetti L (2005) Hypothalamic sensing of fatty acids. *Nature Neuroscience* 8: 579–584.
- Loftus TM, Jaworsky DE, Frehywot GL, et al. (2000) Reduced food intake and body weight in mice treated with fatty acid synthase inhibitors. *Science* 288: 2379–2381.
- Wolfgang MJ, Kurama T, Dai Y, et al. (2006) The brain-specific carnitine palmitoyltransferase-1c regulates energy homeostasis. *Proceedings of the National Academy of Sciences of the United States of America* 103: 7282–7287.
- Wolfgang MJ and Lane MD (2006) The role of hypothalamic malonyl-CoA in energy homeostasis. *Journal of Biological Chemistry* 281: 37265–37269.

Regulation of Chromatin Dynamics

E-J E Hong and C L Peterson, University of Massachusetts Medical School, Worcester, MA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Bromodomain A distinct protein domain that is found in chromatin-related proteins in a variety of organisms from mammals to yeast. Bromodomain binds specifically to acetylated lysine residues on histone proteins.

Chromatin Name given to the DNA–protein complex present in the nucleus of eukaryotic cells.

Chromodomain The chromatin-organization-modifier domain, a conserved region of ~60 amino acids, originally identified in *Drosophila* modifiers of variegation. Recent studies indicate that this domain binds specifically to methylated lysine residues on histone proteins.

Nucleosome The repeating unit of chromatin composed of DNA wrapped around a histone octamer consisted of two copies each of H2A, H2B, H3, and H4.

Chromatin Basics

Eukaryotic genomes are organized by a set of positively charged proteins called histones. Two copies of four histones, H2A, H2B, H3, and H4, form an octamer by hydrophobic interaction through their histone-fold domains. Each octamer wraps around 147 bp of DNA (~1.8 turns) to form the basic unit of chromatin called the nucleosome. Each nucleosome has eight amino-terminal tails (one from each histone) and two carboxyl-terminal tails (one from each H2A) that protrude out of the core region. These tails are targets of posttranslational modifications that regulate nucleosome dynamics. In higher eukaryotes, linker histones (H1 and H5) bind to the entry–exit point of the nucleosomes, which can contribute to the stability of nucleosomes by preventing unwrapping of deoxyribonucleic acid (DNA). Nucleosomes complexed with linker histones are called chromatosomes, and they are prominent in highly condensed regions of a chromosome.

Different organisms and cell types show distinct spacing between nucleosomes. Average nucleosome spacing ranges from 10 to 50 bp. *In vitro*, an array of nucleosomes forms an extended, 10-nm fiber in low salt (e.g., tris with ethylenediaminetetraacetic acid (EDTA) (TE) buffer) that is cytologically visible as a beads on a string. In physiological ionic conditions (100 mM Na⁺, 1–2 mM Mg²⁺), an array of nucleosomes folds into a fiber of 30 nm in diameter. Linker histones stabilize the 30-nm fiber, and emerging works suggest that different lengths of linker DNA may alter the structure of the 30-nm fiber. Self-association of the 30-nm fibers has been proposed to lead to the formation of chromonema fibers with diameters ranging from 60 to 80 and 100 to 130 nm. Further condensation leads to thick 200–400 nm filaments that are characteristic of interphase chromatin. In addition, some regions of chromatin form a special structure called heterochromatin, whose underlying DNA is less accessible than those in euchromatin.

Nucleosomes provide a safe storage of long, linear DNA (2 m in human cells) and contribute to genomic integrity. However, the positional stability of nucleosomes limits the access of underlying DNA, and thus has an inhibitory effect on DNA metabolism that includes gene expression, repair of

DNA lesions, and DNA replication. To overcome this inhibition, cells utilize three strategies to change the packaging states of nucleosomes: (1) covalent modifications of histones, (2) incorporation of histone variants, and (3) ATP-dependent nucleosome remodeling (Figure 1). The interplay of these three mechanisms determines nucleosome dynamics and subsequent folding of chromatin, such that DNA is accessed/exposed in a regulated manner. The resulting eviction, movement, and/or restructuring of nucleosomes is called ‘chromatin remodeling’.

Covalent Modifications on Nucleosomes

Although covalent modifications are not limited to histone tails, the tails are rich with residues that are susceptible to various modifications. Lysines and arginines are subject to acetylation, methylation, citrullination, sumoylation, and ubiquitylation, whereas serines, threonines, and tyrosines are targets of phosphorylation.

Histone modifications may affect accessibility of DNA in three ways. First, they can change the physical properties of nucleosomes to alter the histone–DNA contacts. This could allow removal of nucleosomes by other proteins/remodelers more readily, and hence aid in creating a nucleosome-free region (NFR) over important *cis*-acting elements. Notably, histone tails do not make strong contributions to DNA–histone interactions within individual nucleosomes; so, this mechanism is more relevant to modifications within the core domains of the histones. Second, histone modifications can regulate chromatin folding and/or oligomerization of chromatin fibers by altering nucleosome–nucleosome interactions. Indeed, the histone tail domains are known to play essential roles in chromatin folding. Third, histone modifications can recruit/stabilize the binding of effector proteins that contain specialized domains that recognize defined chemical moieties. These effector domains are present in components of the transcription apparatus as well as structural/regulatory proteins that modulate chromatin structure. Notably, the effector domains are present in chromatin remodelers; thus, histone modifications can have a direct effect on recruitment of these proteins.

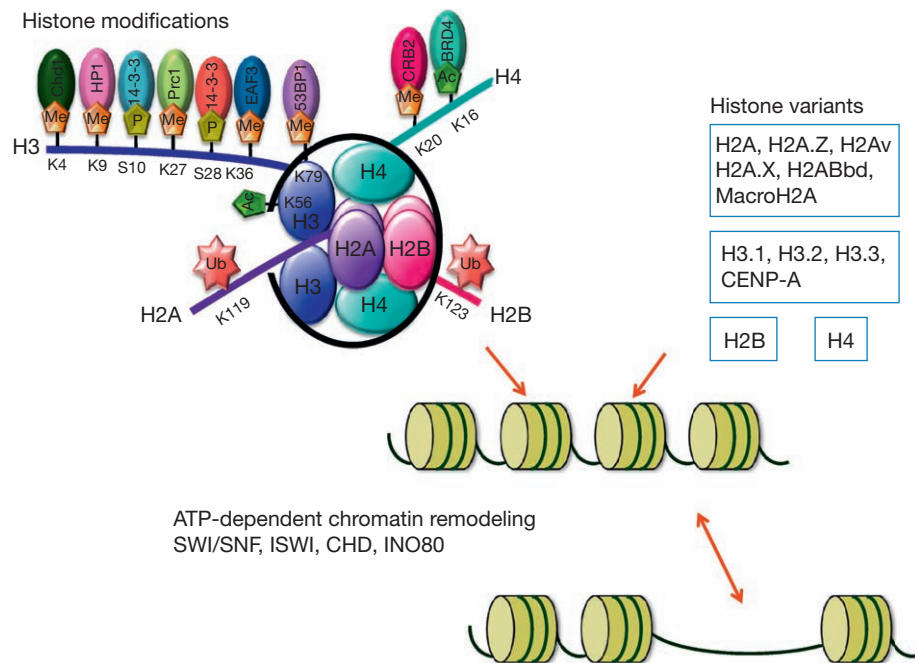


Figure 1 A diagram showing the dynamic interplay of histone modifications, histone variants, and ATP-dependent chromatin remodeling enzymes.

Acetylation

Histone acetylation involves the transfer of an acetyl moiety from acetyl-CoA to the ϵ -amino group of a lysine residue, resulting in neutralization of the positive charge. Yeast Gcn5 was the first nuclear acetyltransferase to be identified, and along with other histone acetyltransferases (HATs), p300 and PCAF, it functions as a transcriptional coactivator. Histone deacetylases (HDACs) remove the acetyl group from histones, and typically act as transcriptional repressors. Notably, establishment of heterochromatin in budding yeast requires removal of acetylation at H4K16 by a nicotinamide adenine dinucleotide (NAD)-dependent deacetylase silent information regulator 2 (Sir2). Thus, histone acetylation *in vivo* correlates with access to underlying DNA.

It has been proposed that the acetylation of certain internal lysines (e.g., H3K56) may affect nucleosome stability by reducing the affinity of histones for DNA. Although acetylation of histone tails does not affect nucleosome structure in a significant way, it may affect the mobility of nucleosomes by altering the interaction between the tail and the linker DNA. Acetylation also prevents folding of chromatin – a 10-nm nucleosomal array containing H4 that is homogeneously acetylated on lysine 16 shows a marked decrease in its ability to fold into a 30-nm fiber. Furthermore, H4K16Ac inhibits inter-fiber interaction and oligomerization of the nucleosomal arrays, which is thought to be a major way in which chromatin is organized in yeast. The H4 tail (16–20) interacts with the acidic patch on H2A of the neighboring nucleosome, and it is likely that the acetylation and subsequent charge neutralization of H4K16 disrupts its contact with H2A. In addition, the remodeling activity of the Imitation Switch (ISWI) protein (see next), which contributes to gene repression by facilitating regular spacing of nucleosomes, is attenuated by H4K16 acetylation.

Bromodomains are structural motifs that recognize acetylated lysines. Although the bromodomain has been shown to bind unmodified histone tails *in vitro*, this domain of 110–120 amino acids shows a strong preference for acetylated lysines. Structural studies of bromodomains by nuclear magnetic resonance and X-ray crystallography showed that this preference for acetylated lysines can be attributed to extra hydrogen bonds formed between the carbonyl oxygen of the acetyl group and the amino acid side chains that are present within the pocket of the four-helix bundle of bromodomains. The amino acid sequences following the acetylated lysine contribute to the specificity for different histone acetylations (e.g., H3K14Ac vs. H4K16Ac) by providing additional sites of interaction. The dissociation constant of a bromodomain for a single acetylated lysine shows relatively weak affinity (e.g., hTAF_{II}250 double bromodomain $K_d = 39 \mu\text{M}$ for H4K16Ac peptides); however, the presence of multiple acetylations can enhance affinity significantly (e.g., hTAF_{II}250 double bromodomain $K_d = 5.6 \mu\text{M}$ for H4K8Ac/H4K16Ac), demonstrating cooperation among multiple histone modifications. Bromodomains are important for gene expression, as they are present in many transcriptional activators and ATP-dependent chromatin remodelers that promote remodeled chromatin structures.

Methylation

Methylation involves transfer of a methyl group from S-adenosyl-L-methionine to lysine or arginine. Methylation/demethylation has been associated with a wide range of biological phenomena that include transcriptional regulation, heterochromatin formation, X-chromosome inactivation, as well as DNA methylation. Thus, methylation plays an important role in epigenetic regulation of the genome.

Most methylations are carried out by proteins that contain a highly conserved Su(var)3–9, Enhancer-of-zeste, Trithorax (SET) domain. However, the discovery that the Dot1/KMT4 methyltransferase, which targets H3K79, does not have a SET domain showed that methylation could be carried out via multiple mechanisms. Unlike acetylation, methylations occur in different states. Lysine methylation occurs in a monomethyl, dimethyl, or trimethyl state, whereas arginine methylations occur in monomethyl, asymmetric dimethyl, or symmetric dimethyl states. Different sites as well as different states lead to different readouts. Methylation at H3K4 correlates with active transcription, whereas H3K9 methylation is generally associated with gene repression. However, while H3K9me2/3 correlates with gene silencing in heterochromatin, H3K9me1 is seen over actively transcribed genes.

Methylation is decoded by different effector proteins that recognize the methylated moiety. One well-known example is the chromodomain. Chromodomains are typically 50–60 amino acids long and form an N-terminal, three-stranded, anti-parallel β -sheet packed against the C-terminal α -helix. A cleft between β 1 and β 3 creates a hydrophobic pocket composed of three aromatic residues (exact residues differ from mouse HP1 β to *Drosophila* Polycomb), to which the methylated lysine binds. Dissociation constants for chromodomains show a dynamic range – HP1 β bound to H3K9me2 peptide shows $K_d = 0.7 \mu\text{M}$, whereas Swi6 (HP1 of *Schizosaccharomyces pombe*) and Ctr4 (H3K9 methyltransferase in *S. pombe*) show $K_d = 10$ and $3 \mu\text{M}$, respectively. As is the case for acetylation, the dissociation constants displayed by chromodomains are in the μM range, and are not sufficient to recruit these proteins to modified nucleosomes. Thus, their recruitment may be mediated by other DNA-binding proteins, or by the combined affinity of multiple chromodomains for different histone modifications. Chromodomains do not recognize methylated lysine as a free amino acid, and the context within which the methylated lysine resides (i.e., the surrounding amino acids) is important for specificity. Similarly, the context within which the chromodomain resides within the histone tail is also important for recognition of different modifications. For example, the chromodomain of Polycomb binds to H3K27me2/3, whereas the chromodomain of heterochromatin protein 1 (HP1) binds to H3K9me2/3. Similarly, the chromodomain of human CHD1 binds to H3K4me2/3.

It was long believed that lysine or arginine methylation was an irreversible histone modification and that the only way to reverse this mark was by histone removal. However, the recent discovery of histone lysine demethylases suggests otherwise. Two classes of lysine demethylases exist to date: (1) lysine (K)-specific demethylase LSD1/KDM1 family and the (2) jumonji C domain-containing histone demethylase (JHDM) family. LSD1 oxidatively demethylates H3K4me and H3K9me using flavin adenine dinucleotide as a cofactor. The JHDM family represents a large class of demethylases (over 30 members in human) that contain a highly conserved jumonji C domain, which uses Fe(II) as a cofactor. *In vitro*, the rate of LSD1/KDM1-mediated catalysis is >200-fold faster than that of the JHDM family of demethylases, prompting a possibility that JHDM associates with unknown protein cofactors that may stimulate its activity *in vivo*.

Arginines are methylated by a large family of protein arginine methyltransferases, and are demethylated by either a

JHDM-like demethylase (JMJD6) or arginine deiminases (see later for citrullination). Arginine methylations have been implicated in nuclear-hormone-mediated transcriptional activation.

Phosphorylation

Phosphorylation appends a negatively charged phosphate moiety to the hydroxyl group of serines, threonines, or tyrosines. All four histones, as well as the linker histone H1, are targets of phosphorylation. Notably, histone H3 Ser10 is phosphorylated during mitosis, and this mark is associated with transcriptional activation, chromatin condensation, and mitotic progression. H3S10 phosphorylation provides a classic example of cross-talk between histone modifications. H3S10 phosphorylation occurs more readily on histones that are pre-acetylated at H3K14, and the coexistence of both marks is important for binding of the 14-3-3 effector protein to the phosphorylated moiety. The 14-3-3 protein then recruits the HAT MOF that acetylates H4K16. This event then results in recruitment of a bromodomain-containing BRD4 protein that brings the transcription elongation factor b, which aids in elongation of the RNA polymerase II. Thus, cooperation of multiple histone modifications works in concert to promote gene expression, providing multiple opportunities for regulation.

Similar cross-talk occurs for another binary methylation/phosphorylation switch. HP1 protein contains a chromodomain that binds to H3K9me2/3, and it plays an important role in transcriptional repression. When the nucleosome containing H3K9me2/3 and HP1 is phosphorylated at H3S10, it results in removal of HP1 proteins without altering the H3K9 methylation. This allows transcription of the underlying DNA without erasing the methyl mark, and may play an important role in transient transcription of the normally silent chromatin. Likewise, H3Y42 phosphorylation also excludes HP1 proteins from heterochromatin; thus, the mechanism by which multiple histone modifications modulate the effector proteins may be a common theme.

Ubiquitylation

Ubiquitylation is an unusual form of histone modification. Instead of a small chemical moiety, a 76-amino-acid-long peptide is attached to a lysine residue. Ubiquitylation occurs via sequential enzyme reactions: (1) ubiquitin-activating enzyme (E1) activates ubiquitin in an ATP-dependent manner, (2) activated ubiquitin is attached to a ubiquitin-conjugating enzyme (E2), and (3) ubiquitin is appended to the substrate using an E3 ubiquitin ligase as a scaffold that brings the E2 and the substrate in close proximity.

Histone H2B is ubiquitylated at K123 (K120 in higher eukaryotes) by a macromolecular complex containing Rad6 (E2) and Bre1 (E3) enzymes. H2B–K123 ubiquitylation is linked to transcriptional elongation, and it is reversed by ubiquitin proteases Ubp8 and Ubp10 in yeast. H2B–K123 ubiquitylation has served as another prime example of histone cross-talk, as this modification is required for H3K4 and H3K79 methylation.

Significantly more H2A (10–15%) is ubiquitylated than H2B (1–5%) in higher eukaryotes. H2A ubiquitylation is regulated by the polycomb group complexes, PRC1 and PRC2. The PRC2 complex interacts with the *cis*-acting element called polycomb response element, and methylates H3K27. H3K27me_{2/3} allows binding of the PRC1 complex, which contains an E3 ubiquitin ligase that mediates ubiquitylation at H2A–K119. H2A–K119ub results in enhanced loading of H1 to chromatin *in vitro* and *in vivo*, and subsequent formation of the chromatosome.

H3 is polyubiquitylated in rat-elongating spermatids, and this modification is thought to mediate degradation during the replacement of histones by transition proteins during spermatogenesis. In human cells, H3 and H4 are ubiquitylated by the CUL4–DDB–ROC1 complex, and this is thought to facilitate the cellular response to ultraviolet DNA damage by affecting nucleosome stability.

Sumoylation

Histone H4 is also sumoylated in human cells, and this mark is linked to gene silencing. Small ubiquitin-related modifier is 18% identical to ubiquitin, and slightly larger (11 kDa as opposed to 9 kDa). Sumoylated H4 is thought to contribute to gene silencing by recruiting HDACs and HP1. In budding yeast, all four histones are sumoylated. In particular, H2B–K6/K7, H2B–K16/17, H2A–K126, and five lysines on the H4 tail were identified as targets of sumoylation. H2B–K6/7/16/17A mutation leads to an increase in transcription at *TRP3*, *SUC2*, and *GAL1* by two- to threefold, suggesting that sumoylation is required for transcriptional repression at these loci. Sumoylation seems to occur throughout the genome, but is present at slightly higher level in sub-telomeric regions. It is hypothesized that sumoylation counteracts ubiquitylation and acetylation to inhibit transcription.

Citrullination

Peptidylarginine deiminases (PADs) convert arginine into a non-native amino acid citrulline. H2A, H3, and H4 are targets of citrullination. PAD4, which was first identified in human leukemia cells, can remove a methylated arginine by converting it into citrulline. This conversion has been shown to regulate genes that are targets of estrogen and p53, demonstrating a promoter-specific histone citrullination in gene regulation. Hypercitrullination of histones is also associated with chromatin decondensation in granulocytes/neutrophils, leading to speculation that it contributes to chromatin unfolding in a global manner.

Proteolysis

An unusual form of histone regulation is the regulation of histone-tail proteolysis. The H2A C-terminus is cleaved between Val114 and Leu115 during granulocyte differentiation. These cleaved H2A–H2B dimers display a lower affinity for H3–H4 tetramers, and this results in destabilization of the nucleosome. A similar phenomenon is observed in *S. cerevisiae*,

in which the N-terminus of H3 is cleaved prior to nucleosomal eviction from chromatin, suggesting that proteolysis promotes open chromatin structure by inducing histone turnover. During mouse embryonic stem-cell differentiation, 21–27 amino acids of the N-terminus of histone H3.2 is cleaved by a cysteine protease cathepsin L. The functional significance remains to be seen; however, such a mechanism could act as a rapid switch in erasing epigenetic marks.

Histone Variants

Histone variants are nonallelic isoforms of the canonical histones. Histone variants are expressed throughout the cell cycle, and their deposition into nucleosomes is not dependent on DNA replication or limited to S phase when canonical histones are incorporated into nucleosomes. As histone variants are also covalently modified, deposition of variants provides cells yet another layer of regulation in nucleosome dynamics.

Of the four core histones, the H2A histone has the most variants. As described above, H2A has an acidic patch that makes contact with the H4 tail of the neighboring nucleosome. This contact is important for nucleosome–nucleosome interactions that are required for chromatin folding. In H2A.Z, which shares ~60% identity to H2A, the acidic nature of this patch is increased. Consistent with this, H2A.Z stimulates the folding of nucleosomal arrays *in vitro*, while simultaneously inhibiting inter-fiber interaction. A crystal structure of an H2A.Z-containing nucleosome indicates a slight destabilization in the H2A.Z–H3 interaction, and the presence of a metal ion-binding pocket, which could affect linker histone binding. Insights from the crystal structure have also argued against the possibility of H2A.Z and H2A being in the same nucleosome. However, both homotypic (two H2A.Z–H2B dimers) and heterotypic (H2A–H2B and H2A.Z–H2B dimers) nucleosomes have been observed *in vivo*. The effect of H2A.Z on the stability of the nucleosome has been controversial, but one determining factor may be H3 – H2A.Z nucleosomes containing H3.3 are particularly unstable and salt-labile compared to those with H3.1 or H3.2. Thus, multiple variants may collectively determine nucleosome structure and stability. H2A.Z is incorporated into chromatin by the Swr1 complex (see next), and recent genetic evidence suggests that it may be removed by the Ino80 complex (see next). In yeast and humans, H2A.Z is present at the nucleosomes flanking the transcription start site, and it has been hypothesized that H2A.Z facilitates transcription. H2A.Z is monoubiquitylated at the inactive X chromosome of mammalian female cells, is sumoylated in *S. cerevisiae*, and is subject to multiple amino-terminal acetylations in many species.

H2ABbd is unique to mammals, and causes a significant change in nucleosome structure. Instead of 147 bp of DNA wrapping around a canonical octamer, nucleosomes containing H2ABbd wrap only 118–130 bp, resulting in a far less stable structure. H2ABbd lacks the C-terminal tail that contains the ubiquitylatable lysine residue and the acidic patch that interacts with histone H4. Consistent with this, nucleosomal arrays containing H2ABbd have reduced propensity to fold into 30-nm fibers *in vitro*. *In vivo*, H2ABbd is hypothesized to be associated with active chromatin.

MacroH2A is unique to vertebrates. While it has 64% identity to canonical H2A, it has a large nonhistone C-terminus, and is overall much larger in size than H2A. MacroH2A resides in heterochromatin, most notably in the inactive X chromosome of female mammals. MacroH2A confers resistance to chaperone-mediated histone exchange, ATP-dependent remodeling by SWItching defective/Sucrose NonFermenting (SWI/SNF), and RNAP II-dependent transcription; hence, macroH2A correlates with the closed chromatin state.

The H2A.X variant is present throughout genomes, at a density of ~ 1 H2A.X nucleosome per 10 H2A nucleosomes. H2A.X is quite similar to H2A, but contains a novel C-terminal motif (SQEL) that provides a site for phosphorylation. In budding yeast, the major form of H2A is similar in sequence to mammalian H2A.X. In response to formation of a DNA double-strand break (DSB), H2A.X is phosphorylated at a C-terminal serine residue (Ser139) by ATM, ATR, and DNA-PK kinases. H2A.X phosphorylation occurs over megabases of chromatin that flank a single DSB in mammalian cells (~ 50 kb in yeast), and is thought to provide binding sites for components of the DNA repair and cell-cycle checkpoint machineries.

Histone H3 has many variants: H3.1, H3.2, H3.3, and CENP-A. The H3.1 variant differs from H3.2 by one amino acid. H3.1 is considered to be the major H3, as it has the highest expression level. H3.2 is encoded by one gene (as opposed to 10 genes for H3.1), and H3.2 is enriched with repressive marks, such as H3K27me_{2/3} and H3K9me₂, suggesting that H3.2 may be enriched in heterochromatin. H3.3 differs from H3.1 by five amino acids, and is linked to active transcription, especially when in complex with H2A.Z. While incorporation of H3.1 and H3.2 to chromatin is coupled to replication, H3.3 is incorporated into chromatin by both replication-coupled and replication-independent mechanisms. CENP-A is a specialized H3 that is loaded at centromeres and provides a platform for kinetochore assembly. Human CENP-A has a less-flexible N-terminus compared to H3, and incorporation of CENP-A results in unwrapping of 7 bp of DNA near the entry/exit point. How these changes aid in assembly of special centromeric chromatin remains to be seen.

ATP-Dependent Nucleosome Remodeling

Remodeling enzymes use the energy from ATP hydrolysis to disrupt DNA–histone interactions. All remodeling enzymes contain at least one ATPase subunit and other subunits that determine the biological context, for example, subunits that determine the recruitment to specific loci, or those that regulate its ATPase activity. There are four families of remodeling enzymes, and they are classified by distinct features/domains within their ATPase subunit. They are the SWItching defective/Sucrose NonFermenting (SWI/SNF), chromodomain helicase DNA binding (CHD)/Mi-2, ISWI, and inositol requiring 80 (INO80) families.

SWI/SNF Family

SWI/SNF is one of the first chromatin remodeling complexes to be characterized. SWI/SNF is a multi-subunit complex

(11 subunits, ~ 1 MDa in yeast), whose members were initially identified in a genetic screen for positive regulators of transcription. It was later discovered that mutations in H3 and H4 (called the Switch INdependent (SIN) mutations) suppressed the transcriptional defects in SWI/SNF mutants, leading to the hypothesis that SWI/SNF modulates chromatin structure. Subsequent biochemical assays led to the discovery that SWI/SNF can disrupt nucleosomes. SWI/SNF can (1) reposition/slide nucleosomes translationally; (2) evict nucleosomes, albeit with low efficiency; and (3) evict H2A–H2B dimers on certain DNA templates (mouse mammary tumor virus).

In yeast, SWI/SNF and Remodeling Structure of Chromatin (RSC) represent this family of remodeling enzymes, and they are characterized by the following features on their ATPase subunits: DExx-HELICc ATPase domain with a short insertion, a bromodomain, and a Helicase-SANT-Associated (HSA) domain that interacts with actin-related proteins (ARPs). Genome-wide gene expression studies demonstrated that both SWI/SNF and RSC promote gene expression, and that they regulate largely nonoverlapping sets of genes. RSC is much more abundant than SWI/SNF, suggesting a broader role in chromatin metabolism. Consistent with this notion, RSC is essential for viability, while SWI/SNF is not. RSC is also required for chromosome segregation – RSC localizes to centromeres, and is required for kinetochore function. Moreover, RSC associates with chromosome arms in a cell cycle-dependent manner, and physically interacts with cohesin to promote sister chromatid cohesion.

Evolutionarily conserved set of complexes has been identified in higher eukaryotes. In flies, two complexes, BRM-associated proteins (BAPs) and polybromo-associated BAPs (PBAPs), share one ATPase subunit, Brahma. In mammals, Brg1/hBRM-associated factors (BAFs) and polybromo-associated BAFs (PBAFs) have distinct ATPase subunits (as in yeast) called hBRM and BRG1. Similar to their counterparts in yeast, SWI/SNF family members in higher eukaryotes promote transcription, although human SWI/SNF complexes have also been shown to facilitate transcriptional repression.

In *Drosophila*, subunits of SWI/SNF (Brahma, Moira, and Osa) are members of the Trithorax Group (TrxG) that antagonizes the transcriptional repressive effect of the Polycomb group proteins, in order to induce expression of homeotic (*HOX*) genes that specify cell fates during development. In addition, BRM is required for transcriptional activation of many genes on polytene chromosomes, and both PBAP and BAP localize to hyperacetylated regions on polytene chromosomes, emphasizing their general role in transcriptional activation.

In mammalian cells, BAF is required for the function of many transcriptional activators such as glucocorticoid receptor, estrogen receptor, retinoid receptor, C/EBP β , c-Myc, heat shock factor, and MyoD. Mice homozygous for mutation in either BRG1 or mSNF5/INI1, genes encoding subunits of SWI/SNF, die in early development prior to implantation. Mice heterozygous for the same mutations confer susceptibility to many tumor types, including glandular epithelial tumors and malignant rhabdoid tumors. Mutations in human SNF5/INI1 are causative for an aggressive, pediatric rhabdoid tumor as well as other types of neoplasms, highlighting the role of SWI/SNF as a tumor suppressor. In addition, the SWI/SNF subunit BAF250a is essential for pluripotency and self-renewal

in mouse embryonic stem cells (ESCs), while human SWI/SNF is involved in Myo-D-mediated muscle differentiation. Thus, SWI/SNF-mediated chromatin remodeling is important for gene expression, development, cell-cycle control, stem-cell maintenance, and differentiation.

In addition to its roles in transcription, the yeast SWI/SNF complex also facilitates the repair of DNA double-strand breaks within a condensed, heterochromatin context. *In vitro* data indicate that SWI/SNF action can evict the Sir3 heterochromatin protein from nucleosomal arrays, promoting formation of an early intermediate in homologous recombination. On the contrary, chromatosomes pose a different challenge. H1 linker histone inhibits remodeling by SWI/SNF and many other remodeling enzymes, and this inhibition can be relieved by H1 phosphorylation. More recently, work from the Becker group has shown that some ATP-dependent remodeling enzymes are proficient at remodeling chromatin that contains linker histones.

SWI/SNF has a single bromodomain within the Swi2/Snf2 ATPase subunit, which is important for localization of SWI/SNF to several genes. Consistent with this, histone acetylation by the Spt-Ada-Gcn5-acetyltransferase and NuA4 acetyltransferase was found to stabilize interaction between SWI/SNF and nucleosomes in a manner dependent on the Swi2/Snf2 bromodomain. RSC has multiple bromodomains, located on several subunits, and their collective affinity for multiple-acetylated histone moieties may aid the targeting/recruitment of RSC to genomic loci. Interestingly, electromagnetic (EM) imaging suggests that RSC displays different conformations in response to acetylated peptides (H3K9Ac and H3K14Ac), and thus acetylated nucleosomes may alter its structure/activity via its interaction with multiple bromodomains.

CHD/Mi-2 Family

The CHD family of ATPases contains the DExx-HELICc ATPase domain with a short insertion, and a tandem chromodomain that recognizes methylated lysines. There are two classes: monomeric CHD remodelers (in yeast and higher eukaryotes) and Mi-2/NuRD remodelers in a multi-protein complex (only in higher eukaryotes). Recruitment of human CHD1 may be aided by interaction between its tandem chromodomains and the active histone mark, H3K4me. Alternatively, the chromodomains in Mi-2 are hypothesized to bind DNA rather than methylated lysine, as a mutation in its chromodomain eliminates binding and remodeling of unmodified nucleosomes.

CHD remodeling enzymes have multiple functions. CHD1 acts as a monomer to slide and eject nucleosomes to promote transcription. Specifically, *drosophila* CHD1 (dCHD1) colocalizes with elongating RNAPII, as does yeast CHD1, which also physically interacts with elongation factors. In contrast, the multi-subunit complex Mi-2/NuRD, which contains an HDAC and methyl CpG-binding domain proteins, represses transcription. Moreover, dCHD1 can function in concert with the histone chaperone Nap1 to assemble chromatin *in vitro*. *In vivo*, dCHD1 was shown to deposit the histone H3.3 variant in the male pronucleus, further confirming its role in chromatin assembly. In yeast, however, CHD1 inhibits DNA replication, and thus the role of CHD1 may be tailored to different biological

contexts. Mice lacking CHD4/Mi-2 β , although viable, display hematopoietic and immune defects. CHD7 mutation in mice causes CHARGE syndrome, the genetic disorder associated with growth retardation, heart defects and deafness. Mice with a CHD8 mutation are not viable. Furthermore, CHD5 has been implicated as a tumor suppressor that activates expression of p16 and p19 associated with neuroblastoma in human.

ISWI Family

The ISWI family of remodeling enzymes (NURF, ACF, and CHRAC) contain the DExx-HELICc ATPase domain with a short insertion, and a SANT-SLIDE domain that serves as a nucleosome-recognition motif that binds to unmodified histone tails and DNA. A role for ISWI in general is to provide regular spacing of nucleosomes after transcription and DNA replication, the result of which is associated with a closed chromatin state. Recent studies suggest that ACF may act as a dimer to provide even spacing of nucleosomes. EM imaging suggests that two SNF2 molecules may associate with a single nucleosome, and it is hypothesized that the back-and-forth movement of the nucleosome from either direction may allow proper sensing of the length of linker DNA. In yeast, it was shown that ISWI (Isw2) increases nucleosome occupancy over intergenic regions to prevent inappropriate transcription initiation from cryptic sites. Moreover, ISWI can contribute to the eviction of TATA-binding protein (TBP) to prevent transcription initiation. However, another member of ISWI in higher eukaryotes, NURF, can disrupt nucleosomes to promote RNAPII activation. Notably, dNURF is a member of the Trithorax group that activates the expression of *HOX* genes during development in flies.

INO80 Family

The INO80 (inositol requiring 80) family of remodeling enzymes is different from other enzymes in many respects. First, they contain multiple ATPases. In addition to one large ATPase subunit (Ino80), there are two other AAA⁺ ATPases called Rvb1 and Rvb2, which form hetero-dodecamers. Second, the ATPase domain has a long insertion between the DExx and HELICc domains, which serves as a platform for binding Rvb1–Rvb2. Third, one member of this family, SWR1, can restructure nucleosomes by mediating the exchange of a nucleosomal H2A–H2B dimer for an H2A.Z–H2B dimer. In yeast, the Ino80 and Swr1 complexes are the two members of this class of remodeling enzymes.

Swr1 remodeling enzymes deposit H2A.Z–H2B dimers in a replication-independent manner at nucleosomes flanking the transcription start site (Figure 2). When present with H3.3 (the only H3 isoform in yeast), the H2A.Z-containing nucleosomes are thought to be less stable and more prone to disruption (see above), which may aid in transcriptional activation. At the sites of a DSB, Swr1 (together with NuA4 acetyltransferase) exchanges phosphorylated H2A.X (γ -H2A.X) with H2A.Z, which is believed to facilitate movement of nucleosomes during recombinational repair. To date, only the Swr1 complex has displayed dimer exchange activity, and it lacks other activities that could mobilize nucleosomes (e.g., sliding

or eviction). Thus, the Swr1 complex seems to be specialized for restructuring nucleosomes.

The Ino80 complex was first identified in a genetic screen for mutants defective in the transcriptional response to inositol depletion. The Ino80 complex is now known to control transcription of a subset of yeast genes, and to regulate efficient DNA double strand break repair, stabilize stalled replication forks, and promote sister chromatid cohesion. Thus, Ino80 plays a broad role in chromatin metabolism. The Ino80 complex contains 15 subunits, and the ATP-dependent nucleosome-sliding activity of Ino80 is hypothesized to be important for transcriptional activation. Like the Swr1 complex, the Ino80 complex is recruited to DSB chromatin in a manner dependent on phosphorylated H2A.X. Genetic evidence has led to a model in which the Ino80 complex might reverse the action of the Swr1 complex by exchanging H2A.Z–H2B dimers with H2A–H2B, hence restoring the chromatin state that was present prior to DSB formation. This restoration process appears to be important for eliminating the cell-cycle checkpoint response when cells encounter an unreparable DSB.

The INO80 family of remodeling enzymes is conserved in higher eukaryotes. In *Drosophila*, Pho-dINO80 and Tip60 complexes share similarities with yeast Ino80 and Swr1 complexes, respectively. In mammals, there are INO80, SRCAP, and TRRAP/Tip60 complexes. Notably, the mammalian Tip60–p400, which encodes an acetyltransferase, plays an important role in ESCs. A set of genes in the ESC contains bivalent nucleosomes that are decorated with both activating and repressive histone modifications (H3K4me3 and H3K27me3), and it is thought that this binary state allows genes to be poised for activation. Consistent with this notion, many of these genes are preferentially activated during ESC differentiation. It was recently demonstrated that Tip60–p400 localizes to, and acetylates the bivalent nucleosomes, which may be important for stem-cell maintenance. Similarly, H2A.Z, which is incorporated into chromatin by the SWR1 family member, was shown to localize to promoters of developmentally important genes in ESC. In lineage-committed cells, H2A.Z redistributes to a different set of genes, suggesting that H2A.Z is important for regulating gene expression during differentiation.

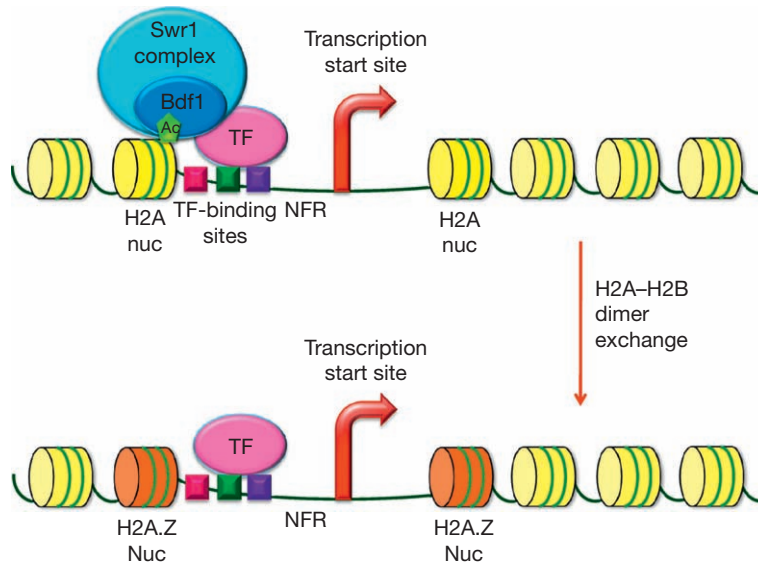


Figure 2 Swr1-dependent deposition of H2A.Z–H2B dimers into nucleosomes flanking the transcriptional start site. Histone acetylation by NuA4 and interaction with transcriptional activators guide Swr1 to the target loci.

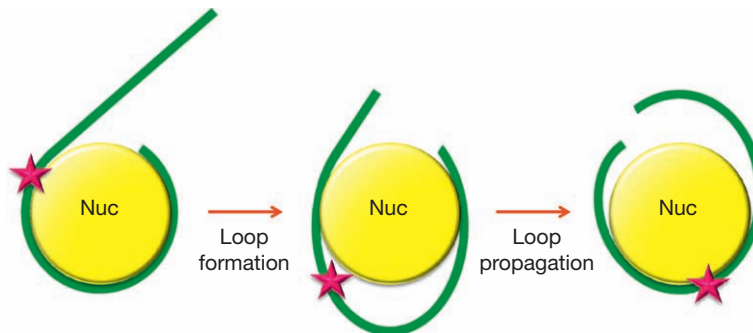


Figure 3 A diagram depicting a current model to explain the mechanism of how ATP-dependent chromatin remodeling enzymes slide a nucleosome translationally. See text for details.

Mechanism of Remodeling

Sliding and eviction of nucleosomes require disruption of 14 histone-DNA contacts. Many previous models have speculated that this disruption is mediated either by a conformational change of the nucleosome, or conversely by a conformational change in the remodeling enzyme. However, the classification of all the ATPases as members of the SF2 family of DEAD/H-box helicases/translocases (they lack a motif that allows separation of a DNA duplex, thus they do not have helicase activity) and the biophysical tests of this by single molecule experiments, led to the current model that remodeling enzymes are DNA translocases that pump DNA around histone octamers.

In a simplified model, the DNA-binding domain of the remodeling enzyme binds linker DNA near the entry-exit site, and the translocase domain binds a fixed place on the octamer, namely two DNA turns from the dyad, which is the central position of the nucleosome. The translocase domain then pumps DNA from this internal location, and generates a DNA loop (of varying sizes for different remodelers) near the dyad. The loop then propagates along the octamers, with the help of DNA-tracking domain of the remodeler, resulting in translational movement of the nucleosome (Figure 3). The directionality of the movement may be determined by the non-ATPase subunits. For example, *drosophila* ISWI enzymes (NURF, CHRAC, and ACF) all share one ATPase subunit ISWI. By itself, ISWI slides the nucleosome from the center to the end of a DNA fragment, while CHRAC (which contains ISWI with three other proteins) repositions the nucleosome from the end to the center of a DNA template. Recently, ACF was hypothesized to work as a dimer to coordinate bidirectional movement, raising the possibility that communication between the remodelers could determine directionality. However, cryo-EM studies indicate one nucleosome per macromolecular remodeling enzyme (e.g., SWI/SNF and RSC), suggesting other mechanisms for directionality.

Summary

Cells are challenged with the conundrum of having to package and safely store long strands of DNA, and at the same time, to

allow regulated usage of the packaged DNA for gene expression, DNA repair, and DNA replication. The dynamic interplay of three mechanisms (1) histone modifications, (2) incorporation of histone variants, and (3) ATP-dependent chromatin remodeling, achieves this feat, and allows cells to coordinate the accessibility of DNA with cellular metabolism and growth. Studies in the past decades have elucidated detailed mechanisms as to how chromatin dynamics is regulated. Notably, the advances of single-molecule studies have led to a new model for how ATP-dependent chromatin remodelers change nucleosomal position. Structural studies of remodeling enzymes have begun to emerge, although insights into how they bind nucleosomes have been difficult to achieve. In addition, most of the *in vitro* work has been done with mononucleosomes or unfolded arrays of nucleosomes, the substrates that do not necessarily reflect chromatin structure *in vivo*. Thus, future work to understand how the histone modifications and histone variants alter the higher-order structure of chromatin, and how ATP-dependent chromatin remodeling enzymes access DNA from folded and more condensed form of chromatin, is anticipated.

See also: **Molecular Biology:** Chromatin: Methyl-CpG-Binding Proteins; Chromatin: Nucleosome Positioning – the GAL Promoter; Chromatin: Physical Organization; Nuclear Organization, Chromatin Structure, and Gene Silencing.

Further Reading

- Campo EI and Reinberg D (2009) Histones: Annotating chromatin. *Annual Review of Genetics* 43: 559–599.
- Clapier CR and Cairns BR (2009) The biology of chromatin remodeling complexes. *Annual Review of Biochemistry* 78: 273–304.
- Lall S (2007) Primers on chromatin. *Nature Structural and Molecular Biology* 14: 1110–1115.
- Smith BC and Denu JM (2009) Chemical mechanisms of histone lysine and arginine modifications. *Biochimica et Biophysica Acta* 1789: 45–57.
- Taverna SD, Li H, Ruthenburg AJ, Allis CD, and Patel DJ (2007) How chromatin-binding modules interpret histone modifications: Lessons from professional pocket pickers. *Nature Structural and Molecular Biology* 14: 1025–1040.

Regulation of Gene Transcription by Hypoxia-Inducible Factor 1

G L Semenza, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Autophagy From the Greek, *auto* oneself, *phagy* to eat; an intracellular process involving the degradation of organelles and/or proteins through the formation of vesicles that subsequently fuse with lysosomes.

Coactivator A protein that interacts with DNA-binding factors to stimulate gene transcription. Coactivators interact with and stabilize the transcriptional initiation complex, which contains RNA polymerase II and associated factors. In addition, coactivators have histone acetyltransferase activity, which is required for the polymerase to access DNA within chromatin and transcribe it into RNA.

Hypoxia response elements (HREs) These are defined functionally as *cis*-acting DNA sequences that are sufficient,

when inserted into heterologous reporter gene (encoding, e.g., firefly luciferase), to mediate hypoxia-inducible transcription of the reporter gene in a hypoxia-inducible factor 1 (HIF-1)-dependent manner.

Micro-RNA (miR) Double-stranded RNA molecules of 21–23 bp that bind to nucleotide sequences with imperfect complementarity located in the 3'-untranslated regions of mRNAs. The binding of an miR to messenger RNA (mRNA) can either block translation of the mRNA into protein or induce degradation of the mRNA.

Mitophagy Some authors use this term to refer to the selective degradation of mitochondria by the process of autophagy.

Hypoxia-Inducible Factor 1 Mediates Transcriptional Responses to Hypoxia in All Metazoan Species

Cells require a constant supply of O₂ for use as the final electron acceptor in the mitochondrial electron transport chain that drives oxidative phosphorylation of adenosine diphosphate to form adenosine triphosphate (ATP). Transfer of electrons from cytochrome *c* oxidase (complex IV) to O₂ results in the formation of H₂O, whereas transfer of electrons from complex I or complex III to O₂ results in the formation of superoxide, which is converted into hydrogen peroxide. Reduced O₂ availability (hypoxia) can lead to cell death due to energy depletion, whereas increased O₂ availability (hyperoxia) can lead to cell death due to excessive production of reactive oxygen species (ROS). Thus, the challenge of oxygen homeostasis is to maintain cellular ATP levels without incurring excess ROS production.

Hypoxia-inducible factor 1 (HIF-1) regulates oxygen homeostasis in all metazoan species that have been analyzed to date, from *Caenorhabditis elegans* to *Homo sapiens*, by controlling many critical adaptive cellular and systemic responses to hypoxia, such as energy metabolism, erythropoiesis, and vascularization. In mammals, HIF-1 is required for proper embryonic development. It has also been shown to play key roles in the pathogenesis of common human diseases, including cancer, chronic lung disease, diabetes, heart failure, infectious and inflammatory diseases, ischemic heart disease, obstructive sleep apnea, ocular neovascularization, peripheral arterial disease, renal failure, rheumatoid arthritis, and stroke.

This article focuses on three aspects of HIF-1 biology and chemistry. First, the molecular mechanisms by which HIF-1 regulates gene transcription are discussed. Second, the wide variety of signal transduction pathways that regulate HIF-1 α protein levels and activity is reviewed. Finally, target genes activated by HIF-1 that maintain oxygen homeostasis by modulating the balance between oxidative and glycolytic metabolism in response to changes in cellular oxygenation are presented.

HIF-1 Functions Principally as a DNA-Binding Protein That Activates Gene Transcription

HIF-1 mediates these physiological effects by functioning as a DNA-binding protein that activates gene transcription. Purification of HIF-1 by DNA affinity chromatography revealed that it was a heterodimer, composed of HIF-1 α and HIF-1 β subunits. Both subunits contain helix-loop-helix (HLH) domain and PER-ARNT-SIM (PAS) homology domain (Figure 1) that mediate dimerization, which is required to juxtapose the basic domains that mediate DNA binding. HIF-1 binds to target genes at hypoxia response elements (HREs), which contain the core sequence 5'-(A/G)CGTG-3'. The first HRE was discovered in the 3'-flanking region of the *EPO* gene, which encodes erythropoietin, the hormone that controls red blood cell production. Subsequently, HREs were identified in 5'-flanking regions and intervening sequences of other HIF-1 target genes.

The binding of HIF-1 results in the recruitment of coactivator proteins, such as CREB binding (CBP) and p300, which is mediated by two transactivation domains (TADs) located in the carboxy-terminal half of HIF-1 α (TAD-N and TAD-C; see Figure 1). Approximately 250 direct HIF-1 target genes have been identified using the reporter gene assay described above and, more recently, by combining genome-wide chromatin immunoprecipitation (ChIP) and messenger RNA (mRNA) expression profiling (ChIP-chip) assays.

HIF-1 Triggers a Transcriptional Cascade

Among the direct target genes that are transcriptionally regulated by HIF-1, the most abundant class encodes other transcription factors. When these transcription factors bind to their target genes, a secondary wave of transcriptional responses to hypoxia occurs. When any given cell type is exposed to hypoxia, the transcription of several hundred genes increases and the

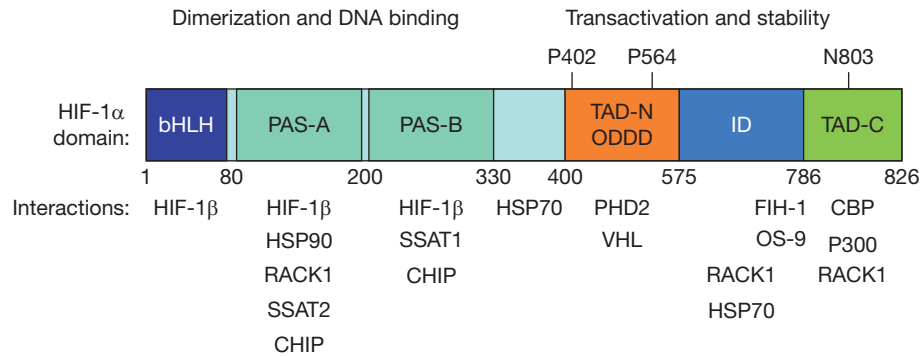


Figure 1 Functional domains of HIF-1 α and protein–protein interactions. bHLH, basic helix–loop–helix domain; PAS, PER-ARNT-SIM homology domain; TAD-N, amino-terminal transactivation domain; ODDD, oxygen-dependent degradation domain; TAD-C, carboxyl-terminal transactivation domain; P, prolyl residue; N, asparaginyl residue.

transcription of a similar number of genes decreases. Although both transcriptional activation and repression in response to hypoxia are HIF-1-dependent, only the genes subject to transcriptional activation are direct HIF-1 target genes. The transcriptional repression that is mediated by HIF-1 is indirect. One obvious mechanism involves direct HIF-1 target genes that encode transcriptional repressors. For example, HIF-1 binds directly to the *DEC1* gene. All of the *DEC1* target genes are secondary targets of HIF-1. Included among the targets for repression by *DEC1* is the *PPARG* gene, which encodes the transcription factor peroxisome proliferator-activated receptor gamma 2 (PPAR γ 2) – whose target genes represent tertiary targets of regulation by HIF-1. When indirect (i.e., secondary or higher order) target genes are taken into account, the number of genes regulated by HIF-1 increases from hundreds to thousands.

HIF-1 Regulates Genes Encoding Histone Demethylases

In addition to DNA-binding proteins, a second key group of transcriptional regulatory proteins includes those that modify chromatin. Among these, the histone acetylases/deacetylases and histone methylases/demethylases appear to play particularly important roles. *JMJD1A*, *JMJD2B*, *JMJD2C*, and *JARID1B* are direct HIF-1 target genes that encode histone demethylases, enzymes that remove methyl groups from lysine residues. As in the case of DNA-binding factors that are regulated by HIF-1, the altered expression of histone demethylases is likely to result in altered rates of transcription of dozens or even hundreds of genes in hypoxic cells. Depending on the specific histone and lysine residue involved, such demethylation can either increase or decrease transcription of the gene to which the histones are bound. Specific genes whose transcription is modulated under hypoxic conditions as a result of histone demethylation have not yet been identified. Nevertheless, these findings suggest that HIF-1 plays an important role in altering the epigenetic state of chromatin in response to changes in cellular oxygenation.

HIF-1 Regulates the Transcription of Micro-RNAs

One of the most remarkable recent advances in the field of molecular biology has been the identification of hundreds

of short micro-RNAs (miRNAs) that bind to the 3'-untranslated regions of mRNA molecules and thereby either block their translation into protein or induce their degradation. The expression of a growing number of miRNAs is regulated by hypoxia in an HIF-1-dependent manner via direct binding of HIF-1. By doing so, HIF-1 exerts a tremendous effect not only on mRNA levels but also on protein levels, that is, HIF-1 is an indirect regulator of translation. miRNA-induced mRNA degradation may represent another major mechanism by which HIF-1 can indirectly repress gene expression.

Transcriptional Regulation by HIF-1 α Monomers

All of the mechanisms described above by which gene and protein expression are directly or indirectly regulated involve binding of the HIF-1 α :HIF-1 β heterodimer to DNA. However, HIF-1 α has also been shown to regulate gene expression independently of HIF-1 β . It does so not by binding to DNA but by binding to other transcription factors. Transcription of the *MSH2* and *MSH6* genes, which encode subunits of the mismatch DNA repair protein MUTS α , is activated by promoter binding of SP1, which then recruits MYC. Under hypoxic conditions, HIF-1 α displaces MYC from binding to SP1, thereby blocking transcriptional activation. By binding to SP1 and blocking transcriptional activation, HIF-1 α functions as a co-repressor in this context, which provides another expression of mRNAs that may be reduced in response to hypoxia. Remarkably, HIF-1 α also functions as a co-activator of NOTCH-dependent transcription by direct binding to the NOTCH intracellular domain.

Regulation of HIF-1 α Levels and Transcriptional Activity

Given the complex and wide-ranging effects of HIF-1 α :HIF-1 β heterodimers and HIF-1 α monomers on gene expression, it is perhaps not surprising that HIF-1 α levels and activity are subject to complex and wide-ranging regulatory mechanisms that modulate HIF-1 α mRNA levels, protein synthesis, protein stability, and TAD function.

HIF-1 α mRNA Levels

HIF-1 α mRNA levels increase dramatically in tissues subjected to hypoxia or ischemia *in vivo*. However, because HIF-1 α mRNA levels are only transiently regulated by hypoxia in cultured cells, it has been difficult to study the molecular mechanisms involved. For example, it is not known whether such regulation occurs at the level of HIF-1 α mRNA synthesis or stability.

HIF-1 α Protein Levels

Regulation of HIF-1 α stability by acute hypoxia

HIF-1 α protein levels increase dramatically in tissues subjected to hypoxia or ischemia *in vivo* and in cultured cells subjected to hypoxia. Analysis of cultured cells revealed that proteasomal degradation of HIF-1 α is inhibited under hypoxic conditions. Delineation of the molecular basis for this response revealed a direct mechanism by oxygen modulates HIF-1 α levels. In the presence of adequate O₂ concentrations, HIF-1 α is hydroxylated on two proline residues (Pro402 and Pro564 in human HIF-1 α) by a family of dioxygenases (designated PHD1, PHD2, and PHD3) that utilize HIF-1 α , O₂, and α -ketoglutarate as substrates and generate prolyl-hydroxylated HIF-1 α , succinate, and CO₂ as products. In tissue culture cells, PHD2 is specifically required for maintenance of low HIF-1 α levels under nonhypoxic conditions. The protein OS-9 binds to both HIF-1 α and PHD2, thereby stabilizing complex formation and increasing the efficiency of the hydroxylation reaction. SIAH2 is a ubiquitin ligase that targets PHD1 and PHD3 for degradation. SIAH2 expression is induced by hypoxia, which may further contribute to loss of prolyl hydroxylase activity.

Prolyl hydroxylation of HIF-1 α is required for binding of the von Hippel–Lindau (VHL) protein, which recruits Elongin C and its associated ubiquitin-ligase subunits (Elongin B, RBX1, CUL2, and an E2 ubiquitin-conjugating enzyme). The protein SSAT2 binds to HIF-1 α , VHL, and Elongin C, thereby stabilizing complex formation and increasing the efficiency of the ubiquitination reaction. Polyubiquitination of HIF-1 α targets the protein for degradation by the proteasome. VHL-dependent ubiquitination is counteracted by VHL-deubiquitinating enzyme 2, which is itself a target for VHL-dependent ubiquitination and degradation.

Under hypoxic conditions, the catalytic activity of the prolyl hydroxylases is inhibited, the fraction of HIF-1 α that escapes ubiquitination increases, leading to decreased HIF-1 α degradation and rapid protein accumulation. The precise mechanism by which hypoxia inhibits prolyl hydroxylase activity is somewhat controversial. Under conditions of anoxia or severe hypoxia, the reduced substrate (i.e., O₂) availability is a sufficient explanation for loss of catalytic activity. Under more modest hypoxia that is typically observed under physiological conditions, it is not clear whether O₂ is limiting. Instead, hypoxia-induced ROS generation may lead to inhibition of hydroxylase activity as a result of oxidation of the Fe(II) ion located at the catalytic center of the enzyme. These mechanisms are not mutually exclusive.

Regulation of HIF-2 α stability by acute hypoxia

Based on its sequence similarity to HIF-1 α , a related protein that is designated HIF-2 α was identified by database searches.

HIF-2 α also dimerizes with HIF-1 β and binds to HIF-1 target genes. However, in contrast to HIF-1 α , which is ubiquitously expressed, HIF-2 α is expressed in a limited number of cell types and is only found in vertebrate species, where it appears to play important roles in specialized systems for O₂ delivery such as erythropoiesis. HIF-2 α also contains two proline residues that are hydroxylated by PHDs in well-oxygenated cells, leading to its ubiquitination by the VHL-Elongin C ligase complex, and proteasomal degradation.

Differential regulation of HIF-1 α and HIF-2 α stability by prolonged hypoxia

Kinetic analyses have revealed that HIF-1 α is strongly induced by acute hypoxia, with peak protein levels observed after 4–8 h and declining levels after 16–24 h. In contrast, HIF-2 α levels are induced modestly at 4–8 h and increase progressively at time points >24 h. A potential molecular mechanism underlying the differential effect of prolonged hypoxia is the recent identification of the HSP70–CHIP complex as an E3 ubiquitin ligase that binds to HIF-1 α and catalyzes its ubiquitination, leading to subsequent proteasomal degradation. Although HSP70 also binds to HIF-2 α , it does not efficiently recruit CHIP and, as a result, the HSP70–CHIP complex does not affect the stability of HIF-2 α protein.

Role of HSP90 in stability of HIF-1 α and HIF-2 α

Inhibitors of heat shock protein HSP90 such as 17-allylamino-geldanamycin (17-AAG), which are currently in clinical trials as anti-cancer agents, have been shown to induce the degradation of HIF-1 α and HIF-2 α in an O₂-, PHD-, and VHL-independent manner. 17-AAG blocks the binding of HSP90 to PAS-A subdomain of HIF-1 α , which is instead occupied by RACK1, which recruits Elongin C and its associated ubiquitin ligase complex, leading to HIF-1 α ubiquitination and degradation. Whereas VHL only binds to HIF-1 α and recruits the Elongin C complex in an O₂-dependent manner, RACK1 binding and recruitment of Elongin C complex is 17-AAG dependent but O₂ independent. The protein SSAT1 binds to both HIF-1 α and RACK1, stabilizing complex formation and increasing the efficiency of the ubiquitination reaction. It is remarkable that SSAT1 and SSAT2, which are homologs that share 46% amino acid identity, both bind to and promote the ubiquitination of HIF-1 α , but they do so by completely different mechanisms.

Regulation of HIF-1 α stability by small ubiquitin-related modifier

Conjugation of the 100-amino-acid small ubiquitin-related modifier (SUMO) protein to HIF-1 α promotes the binding of VHL even in the absence of prolyl hydroxylation. Mice that lack SENP1, which is a protease that removes conjugated SUMO from HIF-1 α , die of anemia due to EPO deficiency that results from excessive SUMO-dependent ubiquitination and degradation of HIF-1 α .

Regulation of HIF-1 α protein synthesis

In some normal cell types and many cancer cells, the rate of translation of HIF-1 α mRNA into protein is regulated by the activity of the protein kinase mTOR complex 1 (mTORC1), which activates ribosomal protein S6 kinase and inactivates eukaryotic initiation factor 4E binding protein 1, thereby

promoting the translation of a subset of cellular mRNAs. The effect of mTOR on HIF-1 α synthesis is particularly prominent in cancer cells with somatic mutations, leading to gain of function for activators of mTORC1 (e.g., phosphatidylinositol 3-kinase and AKT) or loss of function for inhibitors of mTORC1 (e.g., LKB1 and the TSC1/TSC2 complex) and such effects are observed under nonhypoxic conditions. In many cancer cells, the synthesis of HIF-1 α protein is dependent upon mTORC1 activation, such that treatment with rapamycin dramatically reduces HIF-1 α protein levels even under hypoxic conditions, providing a second example of an anti-cancer agent that inhibits HIF-1 α expression.

Regulation of HIF-1 α TAD function by hypoxia

Binding of the co-activators CBP and p300 to the TAD of HIF-1 α under nonhypoxic conditions is blocked by the hydroxylation of an asparagine residue (Asn-803 in human HIF-1 α) by another dioxygenase, designated FIH-1 (factor inhibiting HIF-1), which also uses O₂ and α -ketoglutarate as reaction substrates. Thus, both the half-life and transcriptional activity of HIF-1 α are regulated by O₂-dependent hydroxylation reactions. FIH-1 also hydroxylates HIF-2 α but appears to play a less important role in regulation of its target genes. Recently, the protein MINT3 was shown to interact with FIH-1 and block its hydroxylase activity.

Regulation of HIF-2 α TAD function

In tissue culture cells, the binding of HIF-2 α to target genes (as identified by ChIP assays) does not lead to transcriptional activation and evidence suggests the presence of a corepressor that blocks HIF-2 α TAD function. However, the identity of the putative corepressor and its mechanism of action remain unknown.

Regulation of HIF-1 α expression and activity by intermittent hypoxia

In patients with obstructive sleep apnea, pharyngeal soft tissue occludes the airway during sleep, resulting in episodes of apnea (cessation of breathing) that last 15–30 s, leading to hypoxemia (reduced arterial PO₂), which awakens the patient, who clears his/her airway and resumes breathing, then returns to sleep to repeat the cycle of hypoxia and reoxygenation many times over the course of each night. A major consequence of obstructive sleep apnea is systemic hypertension (high blood pressure), which increases the risk of heart attack, heart failure, and stroke. Animal and cell-based models of chronic intermittent hypoxia (CIH) reveal that the induction of HIF-1 α plays a key role in disease pathogenesis. CIH results in the generation by nicotinamide adenine dinucleotide phosphate (NADPH) oxidase of ROS, which activate Ca²⁺ signaling pathways involving: phospholipase C- γ and protein kinase C, which inhibit prolyl hydroxylase activity and induce mTOR activity, thereby markedly increasing HIF-1 α protein levels; and Ca²⁺-calmodulin kinase, which phosphorylates p300, thereby increasing its binding to the HIF-1 α TAD. Remarkably, whereas CIH increases HIF-1 α protein levels by the mechanisms described above, it decreases HIF-2 α levels by inducing calpain-1-dependent degradation of HIF-2 α . Mice that are heterozygous for an HIF-1 α -null allele do not develop CIH-induced hypertension, nor do mice treated with a calpain inhibitor, which also blocks

CIH-induced HIF-2 α degradation. Thus, dysregulation of both HIF-1 α and HIF-2 α is involved in the pathogenesis of CIH-induced hypertension.

Regulation of Energy Metabolism by HIF-1

Given the fact that HIF-1 is present in the simple roundworm *C. elegans*, which consists of only ~1000 cells and contains no specialized systems for O₂ delivery, it appears that the ancestral function of HIF-1 was to optimize O₂ utilization such that ATP generation was not accompanied by excess ROS production. Whereas it was well known that hyperoxia is associated with increased ROS production, recent studies have demonstrated that acute hypoxia leads to increased ROS levels, which are required to inhibit the prolyl hydroxylases that negatively regulate HIF-1. This is a classic stimulus-response pathway because under conditions of prolonged hypoxia, ROS levels actually decrease below those observed under nonhypoxic conditions and this decrease is HIF-1 dependent. In the absence of HIF-1 α , ROS levels remain increased and cause cell death within 72–96 h.

HIF-1 has been shown to mediate several different adaptive responses that lead to reprogramming of cellular energy metabolism. Perhaps the most subtle of these adaptations is a subunit switch in cytochrome *c* oxidase from the COX4-1 to the COX4-2 isoform. HIF-1 directly activates expression of the genes encoding COX4-2 and LON, a mitochondrial protease that is required for the degradation of the COX4-1 subunit. This subunit switch optimizes complex IV activity under hypoxic conditions; that is, it increases the efficiency by which electrons are transferred to O₂ by cytochrome *c* oxidase to form water rather than the premature escape of electrons to form superoxide. Remarkably, yeast also execute a homologous subunit switch but do so by an entirely different molecular mechanism, as they lack HIF-1, suggesting that this is an example of convergent evolution: an important problem was solved independently in two different lineages.

The COX subunit switch is likely to maintain redox homeostasis under conditions of mild hypoxia, whereas more severe hypoxia may require a reduction in the flow of electrons through the respiratory chain. To accomplish this, HIF-1 activates the expression of the *PDK1* gene, which encodes pyruvate dehydrogenase (PDH) kinase 1. PDK1 phosphorylates and inactivates the catalytic subunit of PDH, the enzyme that converts pyruvate to acetyl coenzyme A (CoA) for entry into the tricarboxylic acid (TCA) cycle, which generates reducing equivalents (in the form of nicotinamide adenine dinucleotide (NADH)) that are supplied to the respiratory chain. Forced expression of PDK1 rescues HIF-1 α -null cells by reducing ROS levels under hypoxic conditions.

Coincident with the reduction in oxidative metabolism, HIF-1 increases the expression of glucose transporters and glycolytic enzymes to increase the uptake of glucose and its conversion to lactate. This increased glycolytic flux is necessary because of the greatly reduced efficiency of ATP generation from glycolytic as compared to oxidative metabolism of glucose.

Although PDK1 prevents the formation of acetyl-CoA from pyruvate, mitochondrial fatty acid oxidation also generates

acetyl-CoA. Thus, an even more draconian measure to reduce ROS production is to reduce the mitochondrial mass of the cell. HIF-1 activates expression of the *BNIP3* gene, which encodes a mitochondrial protein that targets mitochondrial-selective autophagy. Forced expression of BNIP3 rescues HIF-1 α -null cells by increasing autophagy and reducing ROS levels under hypoxic conditions.

Most recently, HIF-1 was shown to induce the expression of the miRNA miR-210 that bound to the mRNA encoding iron-sulfur cluster assembly proteins ISCU1/2, which are required for assembly of the TCA cycle enzyme aconitase and complexes I, II, and III of the electron transport chain. This provides yet another mechanism by which HIF-1 can down-regulate mitochondrial metabolism under hypoxic conditions, underscoring the critical nature of this ancestral response to hypoxia.

See also: **Bioenergetics:** Bioenergetics: General Definition of Principles; Mitochondria: Source and Target of Free Radicals; Mitochondrial Dynamics; Nuclear Genes Involved in Mitochondrial Function and Biogenesis; Reactive Oxygen and Nitrogen Species: Interactions with Mitochondria and Pathophysiology; Respiratory Chain Complex IV; **Cell Architecture and Function:** Autophagy in Fungi and Mammals; **Metabolism Vitamins and Hormones:** Metabolic Control during Ischemia of the Heart; Roles of Micro-RNAs in Metabolism; **Molecular Biology:** Genome-Wide Analysis of Gene Expression; MicroRNAs in Eukaryotes; Regulation of Chromatin Dynamics; **Protein/Enzyme Structure Function and Degradation:** B₁₂-Containing Enzymes; Chaperonins; Protein Degradation; Ubiquitin System; Ubiquitin-Like Proteins; **Signaling:** von Hippel-Lindau (VHL) Protein.

Acknowledgments

Work in the author's laboratory is supported by grants from the National Cancer Institute; National Heart, Blood, and Lung Institute; National Institute of General Medical Sciences; and the Johns Hopkins Institute for Cell Engineering.

Further Reading

- Chan SY, Zhang YY, Hemann C, et al. (2009) MicroRNA-210 controls mitochondrial metabolism during hypoxia by repressing the iron-sulfur cluster assembly proteins ISCU1/2. *Cell Metabolism* 10: 273–284.
- Guzy RD and Schumacker PT (2006) Oxygen sensing by mitochondria at complex III: The paradox of increased reactive oxygen species during hypoxia. *Experimental Physiology* 91: 807–819.
- Hu CJ, Sataur A, Wang L, et al. (2007) The N-terminal transactivation domain confers target gene specificity of hypoxia-inducible factors HIF-1 α and HIF-2 α . *Molecular and Cellular Biology* 18: 4528–4542.
- Kaelin WG, Jr. and Ratcliffe PJ (2008) Oxygen sensing by metazoans: The central role of the HIF hydroxylase pathway. *Molecular Cell* 30: 393–402.
- Kulshreshtha R, Davuluri RV, Calin GA, and Ivan M (2008) A microRNA component of the hypoxic response. *Cell Death and Differentiation* 15: 667–671.
- Luo W, Zhong J, Chang R, et al. (2009) HSP70 and CHIP selectively mediate ubiquitination and degradation of hypoxia-inducible factor (HIF)-1 α but not HIF-2 α . *Journal of Biological Chemistry* 285: 3651–3663.
- Mole DR, Blancher C, Copley RR, et al. (2009) Genome-wide association of hypoxia-inducible factor (HIF)-1 α and HIF-2 α DNA binding with expression profiling of hypoxia-inducible transcripts. *Journal of Biological Chemistry* 284: 16767–16775.
- Nanduri J, Wang N, Yuan G, et al. (2009) Intermittent hypoxia degrades HIF-2 α via calpains resulting in oxidative stress: Implications for recurrent apnea-induced morbidities. *Proceedings of the National Academy of Sciences of the United States of America* 106: 1199–1204.
- Peers C, Haddad GG, and Chandel NS (eds.) (2009) *Annals of the New York Academy of Sciences, Vol. 1177: Hypoxia and Consequences from Molecule to Malady*. Boston, MA: Blackwell Publishing.
- Sakamoto T and Seiki M (2009) Mint3 enhances the activity of hypoxia-inducible factor-1 (HIF-1) in macrophages by suppressing the activity of factor inhibiting HIF-1. *Journal of Biological Chemistry* 284: 30350–30359.
- Semenza GL (ed.) (2007) *Special Issue on Hypoxia and Human Disease. Journal of Molecular Medicine* 85: 1293–1346 (<http://dx.doi.org/10.1007/s00109-007-0285-z>; <http://dx.doi.org/10.1007/s00109-007-0277-z>; <http://dx.doi.org/10.1007/s00109-007-0281-3>; <http://dx.doi.org/10.1007/s00109-007-0279-x>; <http://dx.doi.org/10.1007/s00109-007-0280-4>; <http://dx.doi.org/10.1007/s00109-007-0278-y>; <http://dx.doi.org/10.1007/s00109-007-0283-1>; <http://dx.doi.org/10.1007/s00109-007-0282-2>).
- Sen CK and Semenza GL (eds.) (2004) *Methods in Enzymology, Vol. 381: Oxygen Sensing*. San Diego, CA: Academic Press.
- Sies H and Brune B (eds.) (2008) *Methods in Enzymology, Vol. 435: Oxygen Biology and Hypoxia*. San Diego, CA: Academic Press.
- Xia X, Lemieux ME, Li W, et al. (2009) Integrative analysis of HIF binding and transactivation reveals its role in maintaining histone methylation homeostasis. *Proceedings of the National Academy of Sciences of the United States of America* 106: 4260–4265.
- Zhang H, Bosch-Marce M, Shimoda LA, et al. (2008) Mitochondrial autophagy is an HIF-1-dependent adaptive metabolic response to hypoxia. *Journal of Biological Chemistry* 283: 10892–10903.

Renewable Hydrogen from Biomass

S Thammannagowda, L Magnusson, J H Jo, P-C Maness, and M Seibert, National Renewable Energy Laboratory, Golden, CO, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Dark fermentation An anaerobic metabolic process in which sugar substrate, oxidized by bacteria via the glycolytic pathway, forms pyruvate, reductant, and adenosine triphosphate (ATP), from which mixed products, including ethanol, acetate, lactate, hydrogen, etc., can be produced in exothermic reactions.

Microbial electrolysis cell (MEC) An MEC consists of immobilized bacteria on an anode, which oxidize waste organic materials, with energy capture at an anaerobic cathode; a small applied over-potential is required to facilitate H₂ production.

Photofermentation A light-driven metabolic process in which organic acids are converted to hydrogen by photosynthetic bacteria.

Introduction

Hydrogen can be produced from biomass using thermochemical (pyrolysis and gasification), photobiological (photo-fermentation), electrochemical (microbial electrolysis), and dark fermentative approaches. Thermochemical routes represent nearer term and the others, midterm technologies, requiring additional research and development (R&D) and cost-reduction efforts. All of these are termed 'second-generation technologies' and are much more efficient in terms of land utilization than corn ethanol production. This article highlights fermentation approaches with a brief discussion of the enzymes, organisms and metabolic pathways, and processes involved in H₂ production from biomass.

Enzymes

Two types of redox enzymes catalyze biological H₂ production, hydrogenases (H₂ases) and nitrogenases (N₂ases). Hydrogenases are metallo-proteins that catalyze the oxidation or reduction of H₂ according to the following:



Based on the structure of the active site metal center, H₂ases can be classified as two major types, [FeFe]- and [NiFe]-H₂ases (Figure 1). Although unrelated phylogenetically, both contain CN[−] and CO ligands attached to the iron atoms of the binuclear metal center to efficiently catalyze the inter-conversion of H⁺ and H₂.

Nitrogenases catalyze the fixation of atmospheric N₂; come in forms that contain FeMo, FeV, or Fe-only co-factors; and can produce H₂ as a by-product of N₂ fixation, but at a high energy cost.

[FeFe]-Hydrogenases

[FeFe]-hydrogenases are primarily found in anaerobic bacteria and eukarya. The modular structure and domain organization of these H₂ases are diverse from a simple monomer (e.g., H₂ase I from *Clostridium pasteurianum*) to more complex trimers

or tetramers (e.g., nicotinamide adenine dinucleotide (phosphate) (NAD(P)H)-dependent [FeFe]-H₂ases from *Thermotoga maritima* and *Thermoanaerobacter tengcongensis*). [FeFe]-H₂ases are usually involved in H₂ production, but the periplasmic [FeFe]-H₂ase of *Desulfovibrio vulgaris* Hildenborough can function as an uptake hydrogenase. The active site (H-cluster) is composed of a [4Fe-4S] cluster coordinated to a unique binuclear Fe-Fe center via a conserved cysteine residue (Figure 1(a)). Maturation (functional assembly) of [FeFe]-H₂ases require a minimum of three auxiliary proteins, HydE, HydF, and HydG. The HydE and HydG maturases contain the CxxxCxxC motif, the three cysteines of which coordinate a [Fe₄S₄] cluster, and belong to the class of radical S-adenosylmethionine (SAM) enzymes. The HydF is also a metallo-protein with guanosine triphosphatase (GTPase) activity at its N-terminal region and binds a FeS cluster at its C-terminus. HypF alone is considered to be sufficient to deliver the CN[−] and CO ligands for activating the [FeFe]-H₂ase apo-enzyme.

NiFe-Hydrogenases

[NiFe]-hydrogenases are found in many bacteria, cyanobacteria, and archaea, but not in eukaryotes. The core enzyme is a hetero-dimer with the catalytic large subunit (c. 60 kDa) hosting the bimetallic NiFe-active site and the small subunit (c. 30 kDa) hosting the Fe-S clusters. The NiFe-active site is deeply buried within the large subunit and is coordinated to the protein via four conserved cysteines (Figure 1(b)). The small subunit contains up to three accessory Fe-S clusters, which mediate electron transfer from/to the catalytic site to/from the physiological electron mediator. The most common type is the membrane-bound, H₂-uptake, [NiFe]-H₂ase. Its orientation results in a release of protons in the periplasmic compartment that could potentially facilitate energy coupling to H₂ oxidation. The maturation of the catalytic center within the large subunit is based on the model developed in *Escherichia coli*. It is under the control of six genes in the *hyp* operon as well as an endopeptidase. The biosynthetic pathways for the CN[−] and CO ligands are different. The cyanide ligand derived from carbamoylphosphate (CP) is inserted via HypE and HypF, and the CO ligand may be synthesized from acetate or one of its precursors. HypC and HypD are involved in Fe

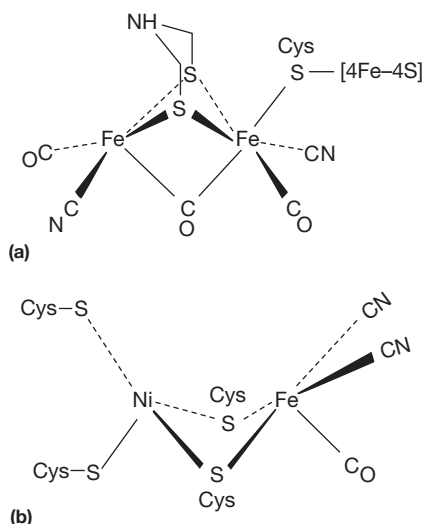
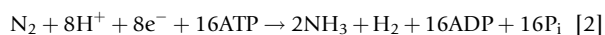


Figure 1 Active site of (a) [FeFe]-hydrogenases and (b) [NiFe]-hydrogenases.

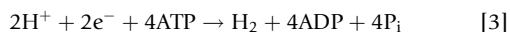
coordination, and HypA, HypB, and SlyD are responsible for nickel delivery in a GTP-dependent reaction. The last step is a proteolytic cleavage of the carboxyl terminus following Ni insertion.

Nitrogenases

In addition to H_2 ase, H_2 can also be produced by N_2 ase, a metallo-enzyme with the most common type harboring a molybdenum and iron metal cluster. Nitrogenase is de-repressed when the growth medium is limited in ammonium. It normally functions in converting atmospheric N_2 into ammonia, a nutrient supporting cell growth according to the following:



One notable feature common to all N_2 ases is the obligatory production of H_2 as a by-product. In the absence of N_2 , H_2 is produced as the sole by-product but consumes 4 mol of adenosine triphosphate (ATP) in the process according to



Consequently, N_2 ase-mediated H_2 production should and does proceed in an inert atmosphere (argon gas) devoid of N_2 to improve overall conversion efficiency. Nitrogenase contains two subcomponents, the MoFe-protein and the Fe-protein. MoFe-protein harbors the MoFe-metallo-cluster catalytic center where N_2 and proton reduction take place. The Fe-protein contains a single $[4Fe4S]$ cluster as well as an Mg-ATP-binding site for electron transfer and ATP hydrolysis. The requirement for large amounts of ATP (contrary to H_2 ases) to drive the uphill reaction, thus, decreases the overall efficiency of H_2 production. Another drawback of N_2 ases is its low turnover rate compared to H_2 ases. Yet, the large negative free-energy change ($\Delta G'^0 = -463 \text{ kJ mol}^{-1}$) accounts for its spontaneous and irreversible nature in H_2 production when compared to H_2 ase. Nitrogenase is present in almost all purple nonsulfur photosynthetic bacteria (PSB) (see section Photofermentation). These organisms, therefore, are ideal candidates

to convert waste organic acids (fermentation waste) to H_2 in the light (photofermentation).

Substrates

Biomass is a broad term including organic materials produced by living organisms or found in waste streams. Lignocellulosic biomass is comprised of cellulose (pure glucose polymer) as the main constituent along with lignin and hemicelluloses (mostly pentose sugars). Cellulose is the most abundant biopolymer on Earth, yet at present only about 4% of US energy comes from biomass. Most common forms of biomass are the wood and its waste, agricultural residues, and municipal solid waste (MSW). A few biomass sources are cultivated as energy crops such as switch grass, miscanthus, and poplar due to their fast growth rates, low water usage, and minimal nutrient requirements. In recent years, carbohydrate- and organic acid-enriched wastewater from various industries (e.g., cattle farm, food processing, dairy processing, molasses based, and starch hydrolysate) has also attracted attention as potential substrate for biofuel production in combination with waste reduction and water treatment. Utilization of all types of waste biomass will inevitably grow in the coming decades for the production of H_2 , methane, ethanol, and biodiesel.

Due to recalcitrance, lignocellulosic biomass has to undergo pretreatment prior to fermentation. Pretreatment can be carried out by several methods such as steam explosion, ammonia fiber explosion (AFEX), and chemical (acid or base) processing. Pretreatment breaks open the cell walls, producing two main fractions: a soluble fraction containing hemicelluloses and acid-soluble lignin, and a solid fraction containing lignin and cellulose (i.e., lignocellulose). Microbes are readily available to ferment lignocellulose or hemicellulose directly without subjecting either substrate to further enzymatic hydrolysis. The concept of 'consolidated bioprocessing', combining substrate hydrolysis with fermentation in one step, is an attractive and viable option to improve the overall techno-economic feasibility of fermentative H_2 production.

Organisms and Pathways

Dark Fermentation

Many anaerobic microbes including *Clostridia* and *Enterobacter* can carry out dark fermentation reactions, which convert various substrates such as sugars, amino acids, and fatty acids to H_2 , CO_2 , ethanol, and other reduced end products. *Enterobacter aerogenes* is capable of producing H_2 but uses costly glucose and sucrose as the substrates. Of particular interest are the clostridial microbes, many of which are cellulolytic and capable of metabolizing crystalline cellulose. This is accomplished by the assembly of an extracellular cellulosome system, consisting of about 30 catalytic subunits along with noncatalytic subunits for cellulose binding. Cellulosome sizes range from 650 kDa up to 100 MDa, with the latter aggregated as polycellulosomes. Among the clostridial microbes, the thermophilic *Clostridium thermocellum* has the highest reported rate of cellulose degradation. Yet, in general, the clostridial microbes are not genetically tractable, making it difficult for

metabolic pathway engineering. Apart from the *Clostridia*, several other mesophilic anaerobic bacteria such as *Acetivibrio cellulolyticus*, *Bacteroides cellulosolvens*, and many microbes isolated from rumen fluids have the cellulosomes. The list is expected to grow rapidly as more cellulosome-producing microbes are discovered with immense potential for biotechnological applications

Most clostridial microbes use the Embden–Meyerhof (glycolytic) pathway for sugar metabolism (Figure 2). The oxidation of pyruvate to acetyl-CoA via the pyruvate:ferredoxin oxidoreductase (step 3 in Figure 2) results in the concomitant production of H_2 using ferredoxin as the electron donor coupling to a FeFe- H_2 ase (step 4). The NADH generated during glycolysis can also couple directly to a different H_2 ase (step 6) or indirectly via ferredoxin (step 5) to catalyze H_2 production. Hydrogenase is presumed to help dissipate excess energy to maintain redox homeostasis in the organism, which may account for the presence of H_2 ase of different types (FeFe- and NiFe-) in a single organism as has been reported in various *Clostridia*, *Thermoanaerobacter*, and *Thermotoga*. Moreover, fermentative microbes must also rely on other pathways to regenerate $(NAD(P)^+)$, such as those NAD(P)H-consuming reactions that yield lactic and butyric acids, and ethanol as the more reduced end products. This mode of mixed-acid fermentation improves substrate utilization and cell growth, yet it compromises the molar yield of H_2 (mol H_2 /mol hexose). This explains the observed H_2 molar yield of between one and two by most laboratories, even though theoretically a maximal yield of four can be realized when acetate is the sole fermentation end product. Numerous strategies have been proposed to overcome this barrier, and some will be discussed later.

Photofermentation

Noxygenic, purple, nonsulfur PSB have attracted industrial interest for H_2 production from biomass because their

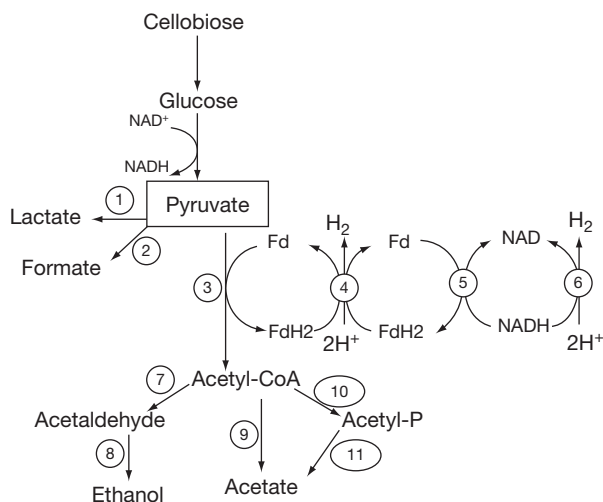


Figure 2 Embden–Meyerhof pathway in *C. thermocellum*. 1. Lactate dehydrogenase, 2. pyruvate:formate lyase, 3. pyruvate:ferredoxin oxidoreductase, 4. Fe–Fe-hydrogenase, 5. NADH:ferredoxin oxidoreductase, 6. NADH NiFe hydrogenase, 7. acetaldehyde dehydrogenase, 8. alcohol dehydrogenase, 9. acetate thiokinase, 10. phosphotransacetylase, 11. acetate kinase.

photosynthetic pathway does not produce O_2 , an inhibitor of both H_2 ase and N_2 ase function. The best-studied PSB belongs to the family of *Rhodospirillaceae* with photosynthetic antenna pigments absorbing light from 400 to 600 nm and from 800 to about 1000 nm. Harnessing only a type II photosystem, *Rhodospirillaceae* must rely on the presence of organic acids (vs. H_2O for oxygenic photosynthetic organisms) as the source of electrons for CO_2 fixation and for proton reduction. *Rhodospirillaceae* are known to utilize a diverse array of organic acids to support photosynthetic growth, with acetic, lactic, formic acids, propionic, and butyric acids being the most common. These compounds are typical waste by-products of dark fermentation. Most, if not all, PSB contain the nitrogenase enzyme, which under certain conditions can be directed solely toward H_2 production, with electrons derived from the oxidation of these organic acids, and ATP generated during the photosynthetic electron transport process. PSB, therefore, can work synergistically when integrated with a dark fermentation system (the latter supplies the required carbon substrates). This integration has dual benefits of addressing waste remediation from dark fermentation, while improving the overall H_2 molar yield.

Processes, Challenges, and Approaches

To improve the techno-economic feasibility of biological H_2 production on a commercial scale, many challenges need to be overcome. The first challenge is the cost of feedstock. Identifying a feedstock that can be easily cultivated, collected, and fermented has been the goal of many studies. Glucose and xylose are ideal substrates, yet they are prohibitively expensive on a commercial scale, because of the energy-extensive pretreatment required to release these simple sugars from complex lignocellulosic biomass. Consolidated bioprocessing using cellulolytic microbes is one of the most promising solutions, because it combines enzyme production, hydrolysis, and fermentation of complex lignocellulosic substrates in a single step.

The second challenge of fermentation is the low H_2 molar yield. Optimization of culture conditions, species selection, biomass retention, and metabolic engineering are all strategies used to increase the H_2 molar yield. Choosing an appropriate microorganism or consortium is key to efficient substrate degradation. Some extremophiles, for example, are capable of fermenting lignocellulosic materials at very high temperatures and are among the most efficient H_2 -producing organisms. An additional benefit to working at high temperature is to prevent contamination by other more ubiquitous, mesophilic methane-producing bacteria. The purpose of biomass retention is to increase cell density within the bioreactor, thereby increasing the rate of substrate conversion and H_2 production. Typical bioreactor configurations used to maintain high cell density include fed batch, immobilized cell membrane recycle, and various filtration device approaches. Metabolic engineering aims at blocking one or more of the competing pathways, thereby redirecting the reaction from formic, lactic, and butyric acid toward acetate production (this could potentially increase the molar yield of H_2 from 2 to 4 mol H_2 mol $^{-1}$ hexose).

The third challenge of fermentative H_2 production is the accumulation of waste organic acids (acetic, formic, lactic,

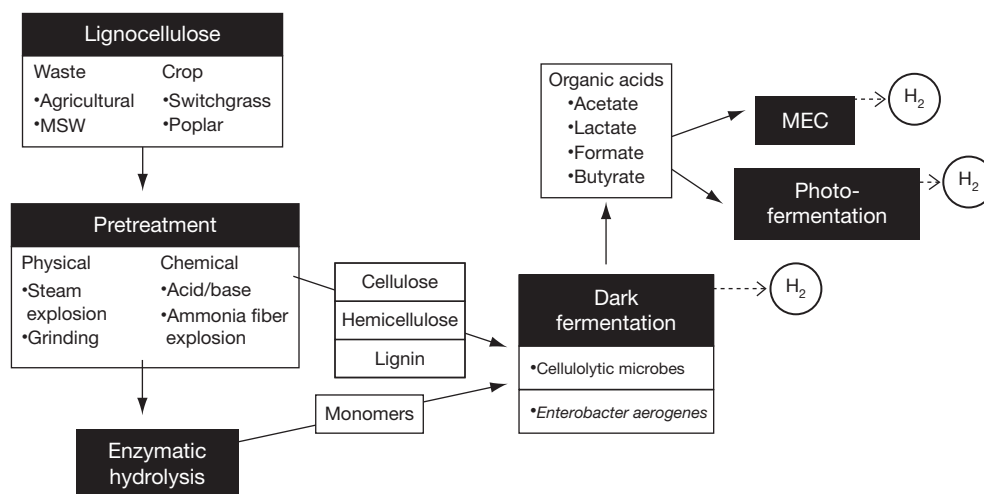


Figure 3 Process flow diagram of an integrated H₂ production process starting from lignocellulosic biomass.

butyric, etc.). The presence of waste acids not only lowers H₂ molar yields but also requires further treatment. These organic acids can be used as substrates for two other H₂ production processes. **Figure 3** illustrates an integrated, H₂ production process where lignocellulose is first converted to H₂ via dark fermentation, and the resultant organic acids are then converted to H₂ either in a microbial electrolysis cell (MEC) or by photofermentation. MECs recruit immobilized bacteria on an anode to oxidize the waste acids, with energy capture at an anaerobic cathode; a small applied over-potential is required to facilitate H₂ production. In one fermentation–MEC integrated system, the MEC was capable of removing 70–85% of the organic acids from cellobiose fermentation with a system H₂ molar yield near 9.95, the highest reported thus far. The second option is to integrate dark fermentation with photofermentation, in which the purple, nonsulfur PSB can photo-assimilate the waste acids and produce additional H₂. In this case, the highest reported combined molar yield is 7.1 mol H₂ mol^{−1} per glucose. In general, effluents derived from carbohydrate-fed (vs. protein-enriched) fermentation are more suitable for photofermentation due to lower carry-over ammonia concentration, an inhibitor of the nitrogenase function. However, the cost and durability of photobioreactors must be considered to ensure economical feasibility.

Epilog

Future studies will improve both productivity and cost issues by focusing on integrated processes that combine the use of cellulosic biomass rather than starch or simple sugars, dark fermentation using thermophilic bacteria (they exhibit faster rates of H₂ production), and either a photofermentation or an MEC to produce more H₂ (up to 4 mol H₂ mol^{−1} acetate

equivalent) from the organic acid products of the dark fermentation. At the moment, the highest reported molar yield of any integrated system is around 10 mol H₂ mol^{−1} hexose starting from cellobiose fermentation and converting 70–85% of the resulting organic acids to H₂ using an MEC.

Acknowledgments

The authors would like to thank the US Department of Energy's Office of Science (MS and PCM) and Fuel Cell Technologies Programs (MS and PCM) and the Korea Research Foundation Grant funded by the Korean Government (KRF-2008-357-D00155) (J.H.J) for support.

See also: **Metabolism Vitamins and Hormones: Glycolysis Overview; Photosynthesis.**

Further Reading

- Fontecilla-Camps JC, Volbeda A, Cavazza C, and Nicolet Y (2007) Structure/function relationships of [NiFe]- and [FeFe]-hydrogenases. *Chemical Reviews* 107: 4273–4303.
- Ghirardi ML, Maness P-C, and Seibert M (2008) Photobiological methods of renewable hydrogen production. In: Rajeshwar K, McConnell R, and Licht S (eds.) *Solar Generation of Hydrogen*, ch. 8, pp. 229–271. New York, NY: Springer.
- Hallenbeck PC and Ghosh D (2009) Advances in fermentative biohydrogen production: The way forward? *Trends in Biotechnology* 27: 287–297.
- Lalauette E, Thammannagowda S, Mohagheghi A, Maness P-C, and Logan BE (2009) Hydrogen production from cellulose in a two-stage process combining fermentation and electrohydrogenesis. *International Journal of Hydrogen Energy* 34: 6201–6210.
- Laurinavichene TV, Belokopytov BF, Laurinavichius KS, et al. (2010) Towards the integration of dark- and photo-fermentative waste treatment. 3. Potato as substrate for sequential dark fermentation and light-driven H₂ production. *International Journal of Hydrogen Energy* 35: 8536–8543.

Repair of G–T Mismatches by *Escherichia coli* Vsr and Eukaryotic DNA Glycosylases

M Lieb, University of Southern California, Los Angeles, CA, USA

A S Bhagwat, Wayne State University, Detroit, MI, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Margaret Lieb, volume 1, pp. 691–693, © 2004, Elsevier Inc.

Glossary

AP endonuclease An enzyme that makes incisions in DNA on the 5' side of either apurinic or apyrimidinic sites.

Deamination Hydrolytic replacement of an amino group (–NH₂) from a chemical compound with a hydroxyl group.

Dominant negative mutation A mutation that results in a protein that is inactive and, in addition, competes with an active form of the same protein.

Glycosylase (glycosidase) A hydrolytic enzyme that cleaves the bond between a sugar molecule and a nucleic acid base such as thymine.

Monomeric enzyme An enzyme composed of a single polypeptide.

Type II restriction endonuclease An enzyme that cuts both DNA strands at defined positions close to or within a specific base sequence.

5-Methylcytosine (5meC) is found at specific sites in the DNA of *Escherichia coli* and related bacteria, and also in eukaryotes, where it functions in the regulation of gene expression. Hydrolytic deamination of 5meC produces thymine (T) (Figure 1), which is mispaired with guanine (G) in the sister DNA strand. In *E. coli*, the very-short patch (VSP) repair system replaces the mismatched T with cytosine (C), thus preventing the occurrence of a C:G to T:A mutation during the next replication of the DNA (Figure 2). In the absence of VSP repair, spontaneous mutation at the sites of 5meC increases manyfold. VSP repair requires the specific endonuclease Vsr and also a DNA polymerase, pol I, and DNA ligase. MutS and MutL, proteins that are essential for the correction of mispairs arising during DNA replication, are required for efficient VSP repair. In mammals, 5meC occurs in CpG sequences. At least two different enzymes appear to preferentially excise T in a T:G mispair that arises in this context.

VSP Repair in *E. coli*

Discovery of VSP Repair in *E. coli*

Before the advent of DNA sequencing, the relative locations of different mutation sites on chromosomes were determined by performing crosses of different mutants and determining the frequency of wild-type recombinants. The frequency of recombination that requires DNA breakage generally is directly proportional to the distance between mutation sites on the parental genomes. Certain mutations (or markers) appear to recombine with other mutations in their vicinity more frequently than expected. This 'marker effect', which has also been referred to as 'high negative interference', was studied using bacteriophage lambda mutants. The markers in lambda (and in *E. coli*) that display unusually high frequencies of recombination were found to be mutations of 5meC to T. These mutations can be corrected by VSP repair in DNA heteroduplexes during recombination.

Structure and Regulation of *vsr*

A small gene, *vsr*, coding for only 156 amino acids is present in *E. coli* and related enteric bacteria, and is required for VSP repair. It is located adjacent to *dcm*, whose product adds a methyl group to the second cytosine in a 5' CC (A or T) GG sequence (Figure 2). The first seven codons of *vsr* overlap the 3' end of *dcm* and both genes are transcribed from a single promoter. Although their gene transcripts are present on the same messenger RNA, Dcm and Vsr proteins are expressed independently. In actively growing cultures, little Vsr is present in the bacteria, and mutation hot spots are found at the sites of 5meC. In nondividing bacteria, the concentration of Vsr increases significantly and, although deamination of 5meC continues at a rate that is time and temperature dependent, mutations do not accumulate at 5meC. As the bacteria are expected to divide infrequently under natural conditions, VSP repair is an efficient mechanism to prevent mutation at 5meC.

Biochemistry of Vsr

Vsr endonuclease is a monomeric protein whose overall structure and active site have similarities to type II restriction endonucleases. Vsr binds to DNA containing a T:G mispair resulting from the deamination of 5meC (Figure 2). It also binds, with lower affinity, to related sequences lacking the 5'C or the 3'G of the cognate sequence. Binding occurs in the presence of either magnesium or calcium ions, but cleavage occurs only in the presence of magnesium ion. Crystallographic studies have shown that Vsr spreads apart the DNA strands at the mispair site by inserting amino acid side chains, causing the DNA to bend. *In vitro* absence of 5meC from the strand opposite the mispaired T reduces both the rate of DNA binding and the efficiency of enzyme action. This effect has not been observed *in vivo*, where additional factors influence VSP repair.

In vitro studies show that Vsr nicks DNA 5' to the mismatched T. When large amounts of Vsr are present, the

endonuclease activity does not require an additional protein. However, in replicating bacteria, VSP repair activity is enhanced significantly by MutS and MutL. Since the absence of either protein, or of both, has the same effect, it is likely that these products act cooperatively, as they are known to cooperate in mismatch correction of replication errors (MMRs). MutS binds to T:G mispairs in any sequence context and recruits MutL resulting in a MutL–MutS complex that interacts with an endonuclease called MutH. Dominant negative mutations in MutS that block MMR also prevent VSP repair. Studies with purified proteins show that DNA cleavage by Vsr is stimulated by the presence of both MutS and MutL. This suggests that the MutL–MutS complex interacts similarly with both the Vsr and MutH endonucleases.

In addition to Vsr, VSP repair requires DNA polymerase I (Pol I), which has an intrinsic 5′–3′ exonuclease activity that digests away a small number of nucleotides starting at the nick immediately upstream of T. It simultaneously resynthesizes DNA in the 5′–3′ direction. As Pol I is a high-fidelity polymerase, it inserts a C in place of the T that it has removed from the mismatch. The actual number of bases removed is known

to be less than 10, and is perhaps as few as 1 or 2. This is why this repair is referred to as VSP. The repair is completed by DNA ligase, an enzyme that seals single-strand nicks in double-stranded DNA.

Mutagenesis by Vsr

Overproduction of Vsr increases the frequency of mutation at many sites in *E. coli*, mimicking the effect of mutations in *mutS* or *mutL*. Mutagenesis by Vsr can be attributed to its competition for a component of the MMR system. Vsr has been shown to interact with an amino terminal domain of MutL that also interacts with MutH. Thus, it is likely that the concentration of Vsr is maintained at a low level in dividing bacteria to prevent it from competing with MutH for binding to MutL. Vsr would also have a mutagenic effect if it corrected to C:G, a T:G mispair in which the G was a replication error in a newly synthesized strand. It should be noted that MutS has a higher affinity for T:G than for any other mispair. Therefore, it is not surprising that excess MutS reduces VSP repair by competing with Vsr for binding at T:G mispairs that result from deamination of 5meC.

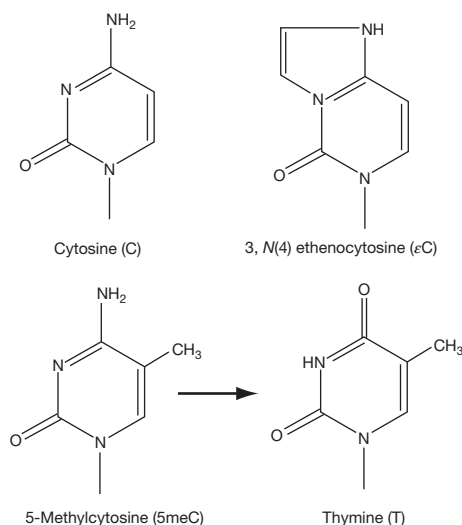


Figure 1 Chemical structures of cytosine and cytosine derivatives. 3,*N*(4) ethenocytosine is a mutagen produced in DNA by endogenous products of lipid peroxidation and by industrial pollutants such as vinyl chloride.

Methyl-CpG-Binding Domain 4 Protein in Eukaryotes

Structure and Biochemistry of Methyl-CpG-Binding Domain 4

The gene for methyl-CpG-binding domain 4 (MBD4) was identified in a search of the human sequence database for a methyl-CpG-binding domain. Mbd4 genes in both the human and mouse code for a ~580 amino acid protein with a methyl-CpG-binding domain near the amino-terminal end and a carboxy-terminal glycosylase domain. The glycosylase region of the protein is related to bacterial glycosylases such as MutY. The enzyme has no lyase activity, so that the removal of the mismatched base leaves behind an abasic site. The sugar-phosphate backbone of DNA can subsequently be cleaved by the human AP endonuclease HAP1, but other endonucleases may also perform this function. There is evidence suggesting that processing of the abasic site also involves interaction with MLH1, a homolog of bacterial MutL.

Specificity

Investigations with human MBD4 have shown that only T:G and U:G mispairs are good substrates for the enzyme.

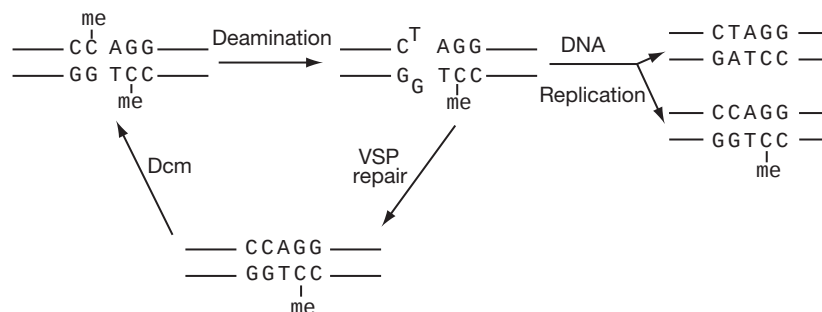


Figure 2 The life cycle of a dcm site in *E. coli*.

The human MBD4 enzyme binds preferentially to T:G mismatches in the meCpG/TpG context. However, removal of the mismatched base was also observed when U:G and T:G mismatches were in a nonmethylated substrate. T:G mismatches in other contexts are processed at a greatly reduced rate. The fact that U is removed from a mispair with G more rapidly than T suggests that removal of T resulting from deamination of meC may not be the primary function of MBD4.

Cancer Prevention

In the cells of the intestinal mucosa of mice containing inactive *Mbd4* genes there was a two- to threefold increase in C to T mutations in CpG sequences. In combination with mutations in the adenomatosis polyposis (*Apc*) genes, MBD4 inactivation increased the number of tumors and tumor progression. A 3.3-fold increase of C to T mutations at 5meC was also observed in the liver and spleen of mice lacking MBD4. Therefore, it is clear that MBD4 is a significant factor in the prevention of mutations leading to cancer.

Thymine-DNA Glycosylase

Discovery and Specificity of Thymine-DNA Glycosylase

Thymine-DNA glycosylase (TDG) was first isolated during the search for a Vsr-like activity in extracts of human cells. The enzyme has a higher affinity for U:G mispairs than for T:G, and also has significant affinity for 3,*N*⁴-ethenocytosine (Figure 1) paired with G. TDG has a strong preference for mispairs arising in the CpG context. However, there is as yet no direct evidence that it has an important role in preventing mutation resulting from deamination of 5meC.

Biochemistry of TDG

TDG has homologs in other eukaryotes, and also in *E. coli*. It has no apparent amino acid similarity to MBD4 or Vsr.

The enzyme is characterized by low abundance and slow action. TDG can remove from DNA the bases T or U that arise by deamination of 5meC or C, and are, thus, mispaired with G. It also excises 3,*N*⁴-ethenocytosine from DNA. Following removal of T or U from the DNA backbone, TDG remains bound to the newly created abasic site. Studies of human TDG *in vitro* have indicated that an apurinic exonuclease (HAP1) helps TDG turnover and cleaves the abasic sugar phosphate. Repair of the resulting nick in the DNA strand requires both a DNA polymerase and DNA ligase.

See also: [Molecular Biology: DNA Mismatch Repair in Bacteria](#); [DNA Restriction and Modification: Type II Enzymes](#).

Further Reading

- Bellacosa A (2001) Role of MED1(MBD4) gene in DNA repair and human cancer. *Journal of Cellular Physiology* 178: 137–144.
- Bhagwat AS and Lieb M (2002) Cooperation and competition in mismatch repair: Very short patch repair and methyl-directed mismatch repair in *Escherichia coli*. *Molecular Microbiology* 44: 1421–1428.
- Hardeland U, Bentele M, Jiricny J, and Schär P (2003) The versatile thymine DNA-glycosylase: A comparative characterization of the human, *Drosophila* and fission yeast orthologs. *Nucleic Acids Research* 31: 2261–2271.
- Heinze RJ, Giron-Monzon L, Solovyova A, et al. (2009) Physical and functional interactions between *Escherichia coli* MutL and the Vsr repair endonuclease. *Nucleic Acids Research* 37: 4453–4463.
- Lieb M and Bhagwat AS (1996) Very short patch repair: Reducing the cost of cytosine methylation. *Molecular Microbiology* 20: 467–473.
- Millar CB, Guy J, Sansom OJ, et al. (2002) Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* 297: 403–405.
- Wu P, Qiu C, Sohail A, et al. (2003) Mismatch repair in methylated DNA. *Journal of Biological Chemistry* 278: 5285–5291.

Respiration in Phototrophic Microorganisms

R Borghese and D Zannoni, University of Bologna, Bologna, Italy

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Hartmut Michel, volume 3, pp. 688–694, © 2004, Elsevier Inc.

Glossary

Cytochromes Redox proteins with an iron-containing porphyrin ring (heme).

Electrochemical proton gradient It defines the membrane-energized state in terms of electrical units.

Genetic regulatory element A molecule that regulates gene expression by interacting with DNA.

Phototroph An organism that converts radiant energy into chemical energy.

Proton motive force An energized state of the membrane resulting from the separation of charges across the membrane.

Respiration The process in which a compound is biologically oxidized by an electron acceptor (O_2 or an O_2 substitute) linked to generation of a proton motive force.

Anoxygenic Phototrophs

On a phylogenetic basis (16S rRNA analyses) phototrophic bacteria and their relatives are grouped in a class, the Proteobacteria, which is formed by several subclasses, named α , β , γ , δ , and ϵ . Facultative photosynthetic bacteria belong to α and β subclasses, for example, genera *Rhodobacter*, *Rhodospirillum*, *Rhodocyclus*, *Rubrivivax*, and *Erythrobacter*.

Metabolic Aspects of Facultative Anoxygenic Bacteria

Facultative phototrophs are probably the most metabolically flexible organisms of the microbial world. Species such as *Rhodobacter* (*Rba.*) *capsulatus* and *Rba. sphaeroides* can grow not only by aerobic respiration and photosynthesis using either organic or inorganic substrates but also by anaerobic respiration with trimethylamine-N-oxide (TMAO) or dimethyl sulfoxide (DMSO) as electron acceptors. Further, some strains of the species *Rhodopseudomonas* (*Rps.*) *palustris*, *Roseobacter* (*Rsb.*) *denitrificans*, and *Rba. sphaeroides* can respire nitrate (NO_3^-) reducing it into dinitrogen (N_2) via nitrite (NO_2^-), and, in some cases, also nitric oxide (NO) and nitrous oxide (N_2O). These energy-generating processes are catalyzed by oxido-reduction protein complexes forming composite electron transport chains. Apparently, not all the above summarized metabolic options are activated or can be available simultaneously in a single species; however, cells of *Rba. capsulatus* or *Rba. sphaeroides* grown photosynthetically in the presence of low oxygen tension (<1%) contain both photosynthetic and respiratory apparatuses.

Electron Transport Chains

Respiration in facultative phototrophs involves numerous redox carriers, namely: (1) transmembrane protein complexes such as nicotinamide adenine dinucleotide (reduced form) (NADH)- and succinate-quinone oxidoreductases (NQR and SQR, respectively), cytochrome (cyt) bc_1 or hydroquinone-cytochrome c oxidoreductase (QCR), quinol oxidase(s)

(QOX), and cyt cbb_3 and/or aa_3 oxidases (COX); (2) electron and/or proton carriers such as ubiquinones (UQ), cyt c_2 , HiPIP, and cyt c_y ; and (3) enzymes of the periplasmic space such as NO_3^- , NO_2^- , N_2O , and DMSO reductases. With O_2 as final electron acceptor, the NQR, QCR, and COX enzymes constitute three main coupling sites where the potential energy between the initial donor and the final acceptor molecules is released in small steps that are controlled by the differences between the redox mid-point potentials (E_m) of the redox couples involved, and efficiently coupled to the generation of an electrochemical proton gradient ($\Delta\mu_{H^+}$). Photosynthesis converts the radiant energy into chemical energy at the level of the photochemical reaction center (RC). This transmembrane protein complex generates a charge separation that is followed by a cyclic electron transfer involving quinone molecules (UQ-10), cyt bc_1 complex, and soluble cyt c_2 in addition to the membrane-bound cyt c_y , in the case of *Rba. capsulatus*. Under dark aerobic conditions, facultative phototrophs use a respiratory chain, closely related to that present in mitochondria. The last step of O_2 reduction into H_2O is catalyzed by two oxidases (not always present in a single species): a cyt c oxidase (of cbb_3 and/or aa_3 type), inhibited by μM cyanide, and a quinol oxidase (of bo type), inhibited by mM cyanide, branching the electron flow at the level of the UQ-10 pool. **Figure 1** shows a block scheme of the different electron-transport pathways operating in *Rba. sphaeroides*.

Interaction between the Diverse Electron Transport Chains

The interaction between the different bioenergetic chains of photosynthetic bacteria occurs by two nonexclusive mechanisms, namely: (1) an indirect mechanism, exerted by the proton motive force (PMF) as a back pressure on the rate of electron transport catalyzed by redox complexes and (2) a direct mechanism of interaction between electron carriers, for example, UQ-10 or cyt c_2 , that are engaged in multiple bioenergetic processes. In this way, the activity of a given bioenergetic chain would affect the redox state of the components in common and, consequently, the functioning of the other chains.

the phosphorylation of RegA which results in the activation of photosynthesis.

In the other well-studied species, *Rba. sphaeroides*, the sensing of the oxygen tension relies, instead, on the terminal oxidase *cbb₃*, which is able to detect changes in the flux of electrons. High electron flow, in the presence of oxygen, represses the PrrA/PrrB system which results in the induction of respiration and the repression of photosynthesis. On the contrary, low electron flow, due to low oxygen, stimulates PrrA/PrrB to induce the photosynthetic-anaerobic metabolism. Depending on its phosphorylated or dephosphorylated status, the RegA/RegB (PrrA/PrrB) couple can regulate the expression of a considerable number of genes involved in important metabolic processes in addition to respiration and photosynthesis: N₂ fixation, CO₂ fixation, and H₂-ase activity. The electron transport chain (ETC) genes that have been shown to be regulated by RegA/RegB (in *Rba. capsulatus*) and PrrA/PrrB (in *Rba. sphaeroides*) are the ones coding for COX and QOX, which are specific for the respiratory chain, and those coding for the *bc₁* complex, cyt *c₂* and cyt *c_v*, which participate in both respiration and photosynthesis (Figure 2). Other regulatory elements that participate in the regulation of respiratory ETC components are AerR, that operates in the presence, as well as in the absence, of O₂; HvrA and FnrL are anaerobic regulators only (Figure 2). Although the interaction of all the afore-mentioned regulatory elements is quite complex, it allows the fine tuning and controlled interplay of respiratory and photosynthetic activities.

Oxygenic Micrororganisms

Cyanobacteria are capable of oxygenic photosynthesis, differing in that from anoxygenic phototrophs. Cyanobacteria form one of the major phyla of Bacteria and they were most likely the first oxygen-evolving organisms on Earth changing the atmosphere from anoxic to oxic. Cyanobacteria are grouped into several morphological groups; however, most of the available biochemical and genetic data concern mainly unicellular genera such as *Synechococcus* and *Synechocystis*.

Respiration is by definition a membrane-bound process; in this respect, cyanobacteria contain three different types of membranes: (1) the outer membrane, typical of Gram negatives, which has no specific role in respiration; (2) the cytoplasmic membrane (CM) also called 'periplasmic membrane (PM)' and (3) the intracytoplasmic membranes (ICMs) also called 'thylakoid membranes (TMs)'. Both CM and TM contain respiratory redox complexes; electron microscopy also indicates that CM and TM might be connected, at least in *Synechococcus* sp. strain PCC 6301, although a correct picture of the membrane arrangement *in vivo* is lacking at present.

Electron Transport Pathways in Cyanobacteria

Respiratory and photosynthetic electron transports are intimately connected in two distinct bioenergetically active membranes, TM and CM. All photosynthetic electron transport is located in the TM, where photosynthesis and respiration share components (Figure 3). In addition, ample experimental

evidences indicate the presence of respiratory chain(s) in the CM. Cyanobacterial respiratory terminal oxidases (RTOs) have no direct function in photosynthesis and therefore can be considered the key enzymes of respiration. All cyanobacteria (investigated so far) contain several respiratory branches ending in different RTOs but their actual location in the membrane cell (TM, CM, or both) is far from being assessed. The best-characterized species is *Synechocystis* sp. strain PCC 6803, for which the complete genomic sequence is available. Three sets of genes for RTOs were found, the well-characterized *aa₃*-type cyt *c* oxidase (COX, encoded by *coxBAC*), a related set of genes also belonging to the heme-copper oxidase superfamily, termed 'alternate RTO (ARTO; encoded by *ctaCII-ctaDIIIEII*), and two genes (*cydAB*) encoding a putative cyt *bd*-type quinol oxidase (Cyd).

Two types of nicotinamide adenine dinucleotide phosphate (reduced form) (NAD(P)H) dehydrogenases have been found. One is a NADPH-type I dehydrogenase (NDH-1), that is encoded by *ndh* genes, which consists of about 12 subunits and contributes to a proton gradient ($\Delta\mu_{H^+}$) across the membrane. The second type of dehydrogenase is a NADH-type II dehydrogenase (NDH-2) consisting of a single subunit and probably not contributing to energy transduction. In *Synechocystis* sp. strain PCC 6803, three genes coding for NDH-2 (*ndbA,B,C*) are found.

The CM forms the inner boundary of the periplasmic space and is known to contain proteins typically associated with respiratory electron transport, such as NAD(P)H dehydrogenase of type II, succinate dehydrogenase (SDH), and terminal oxidases, presumably ARTO and/or QOX (quinol oxidases: Cyd or Cta II). Three intermingling pathways are dominant in *Synechocystis* sp. strain PCC 6803 thylakoids (TM), namely: (1) a linear electron transport from H₂O to NAD(P)H, catalyzed by photosystems I and II (PSI/PSII), (2) a respiratory transport from NAD(P)H and succinate to both COX of *aa₃*-type (Cta I) and QOX of *bd*-type (Cyd), and (3) a cyclic electron flow around PSI, that is, electrons at the acceptor side of PSI returning to the PQ pool by means of a soluble ferredoxin (Fd) (Figure 3). However, as generally seen in facultative anoxygenic phototrophs, (Figure 1) electrons can move from one pathway to another at the level of the PQ pool, *b₆/f* complex (homologous of the *bc₁* complex), and/or soluble carriers such as plastocyanin (PC) or cyt *C₆* (previously known as cyt *c-553*) (Figure 3). For example, in the absence of PSI, reducing equivalents generated by PSII are fed into COX, while in darkness, respiratory electrons are used to reduce the acceptor side of PSII if terminal oxidases are blocked. Results with mutants of *Synechocystis* sp. strain PCC 6803 impaired in several combinations of respiratory and photosynthetic redox complexes indicate that SDH is the main electron transfer pathway into the PQ pool and that type I and II NAD(P)H dehydrogenases might simply operate as regulators of NADP⁺ and NAD⁺ reduction levels. This suggests that respiration in cyanobacteria plays an important role in the control of the intracellular redox balance.

In general, the genes for components of the respiratory chain(s) are present in only one copy per chromosome even in those species having two respiratory chains (Figure 3). How one gene directs its gene product into two different membranes (CM and TM) is an intriguing yet unanswered question. Further, the amounts of several components of the respiratory

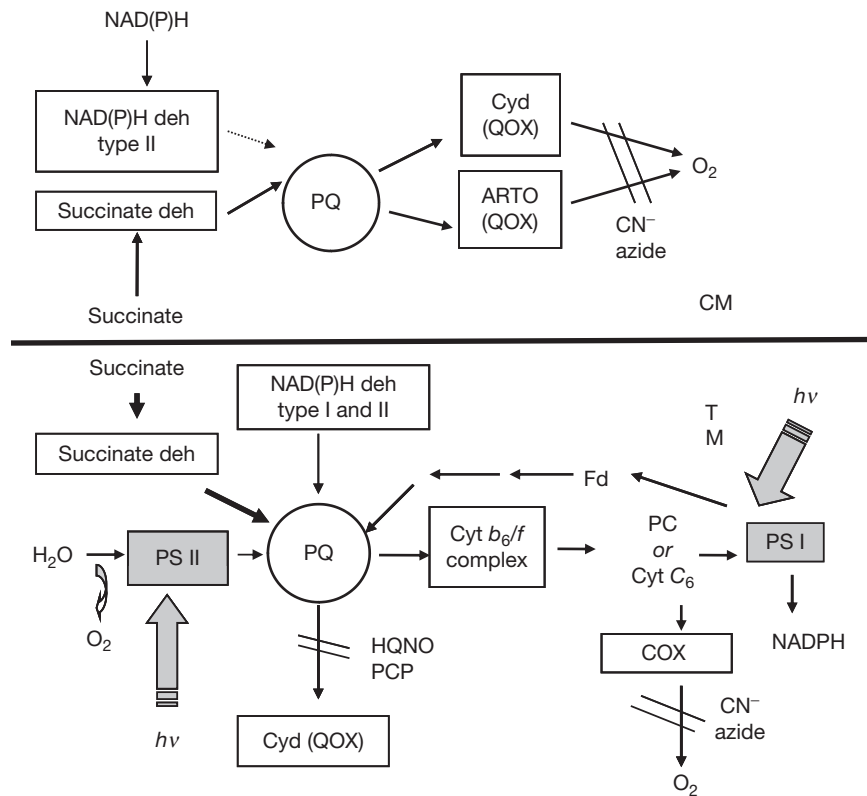


Figure 3 Working model of photosynthetic and respiratory electron transport chains in CM and TM of *Synechocystis* sp. strain PCC 6803. Membrane-bound redox complexes are indicated as rectangular boxes. A dotted arrow from NAD(P)H to PQ (CM redox chain) symbolizes the lack of evidence for electron flow at this level; the thickness of arrows in TM redox chains is indicative of their relative activities. Abbreviations: HQNO, 2-heptyl-4-hydroxy-quinoline-N-oxide; PCP, pentachlorophenol; CN^- , cyanide anion; PC, plastocyanin; PQ, plastoquinone pool; $h\nu$, radiant energy; see also text for further details.

chain(s) are regulated by external factors such as the concentration of Cu^{2+} for synthesis of cyt C_6 and PC or the ionic strength for cyt aa_3 -type oxidase synthesis. Unfortunately, the mechanisms of gene regulation are largely unknown at present and they will be important topics for future studies in respiration of oxygenic phototrophs.

See also: **Bioenergetics:** Cytochrome bc_1 Complex (Respiratory Chain Complex III); Respiratory Chain Complex II and Succinate: Quinone Oxidoreductases; Respiratory Chain Complex IV; **Metabolism Vitamins and Hormones:** Photosynthesis.

Further Reading

- Bernroither M, Zamocky M, Painer M, et al. (2008) Heme-copper oxidases and their electron donors in cyanobacterial respiratory electron transport. *Chemistry and Biodiversity* 5: 1927–1961.
- Schultze M, Forberich B, Rexroth S, et al. (2009) Localization of cytochrome b_6f complexes implies an incomplete respiratory chain in cytoplasmic membranes of the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochimica et Biophysica Acta* 1787: 1479–1485.
- Vermeglio A, Borghese R, and Zannoni D (2004) Interaction between photosynthesis and respiration in facultative phototrophs. In: Zannoni D (ed.) *Respiration in Archaea and Bacteria*, vol. 16, pp. 279–295. Dordrecht: Kluwer.
- Zannoni D, Schoepp-Cothenet B, and Hosler J (2009) Respiration and respiratory complexes. In: Hunter CN, Daldal F, Thurnauer MC, and Beatty JT (eds.) *The Purple Photosynthetic Bacteria*, pp. 537–561. Dordrecht: Springer Science.

Respiratory Chain and ATP Synthase

D G Whitehouse and A L Moore, University of Sussex, Brighton, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Binuclear center A binuclear center comprises two redox-active components in close proximity, for example, copper and cytochrome *a* or two-iron atoms linked by an oxygen atom that function as a single entity to mediate electron transfer.

Proton-motive force The energy contained in an electrochemical gradient of protons across the inner mitochondrial membrane/thylakoid membrane and composed of two components, the membrane potential

and pH gradient. It is the driving force for the transport of substrates across the membrane and for the synthesis of adenosine triphosphate.

Proton pump A transmembrane protein which mediates the active transport of protons across the inner mitochondrial membrane/thylakoid membrane.

Respiratory chain An assemblage of electron carriers normally arranged as a multicomponent complex and located in the inner mitochondrial membrane.

Components of the Respiratory Chain

The respiratory chain and adenosine triphosphate (ATP) synthase are, in eukaryotic cells, located in the inner membrane of mitochondria. Mitochondria are oval-shaped organelles, typically 1–2 μm long and 0.5 μm in diameter, whose principal cellular functions are to generate energy, in the form of ATP and carbon skeletons for biosynthetic purposes. The shape and appearance of mitochondria vary considerably and these organelles appear to be most numerous in mammalian tissues with a high-energy demand (heart, liver, muscle, and brain and in rapidly dividing plants cells). The respiratory chain functions to oxidize NADH ($+H^+$) and $FADH_2$ and reduce molecular oxygen to water. These functions are electron transfer (electron motive) processes and result in a substantial release of energy, which generates a proton-motive force that is used to drive the ATP synthase in the forward direction thereby synthesizing ATP for use in cellular reactions.

The conversion of the electron-motive force (NADH and $FADH$ oxidation) into a proton-motive force is carried out by three electron-driven proton-pumping respiratory chain complexes. These large transmembrane complexes contain numerous reduction–oxidation (redox) components that include flavins, iron–sulfur proteins, cytochromes, ubiquinone, and metal ions.

Flavins

Flavin mononucleotide (FMN) and flavin dinucleotide (FAD) are tightly bound (to their enzymes) cofactors that can accept (or donate) two electrons and two protons (to become fully reduced) or a single electron and proton to form the semiquinone intermediate.

Iron–Sulfur Proteins

Iron–sulfur (or nonheme iron) proteins contain iron atoms covalently bound to the apo-protein by cysteine sulfurs and to other iron atoms via acid labile sulfur bridges. In general, there are three types of iron–sulfur proteins, namely, a single iron atom coordinated to the sulphydryl groups of four

cysteine residues of the protein, a protein containing two irons and two inorganic sulfides ($2Fe-2S$), and four irons and four inorganic sulfides ($4Fe-4S$). Similar to the single iron cluster, both are coordinated by four cysteine residues. All three types of iron–sulfur proteins can act as single electron carriers.

Cytochromes

Cytochromes are proteins containing a prosthetic group called a 'heme'. The basis of the heme is a porphyrin, which consists of four pyrrole rings linked in a cyclic manner by methene bridges with a central coordinated iron atom that acts as a single electron carrier. There are at least four classes of cytochromes (cytochrome A–D) differentiated from each other by the ability to absorb light at different wavelengths (due to differences in the side chains on the porphyrin).

Ubiquinone

Ubiquinone or coenzyme Q is a hydrophobic quinone derivative that in mammalian mitochondria has a side chain of 10 five-carbon isoprene units (Q_{10}). Ubiquinone can be fully reduced to the quinol form by the acceptance of $2H^+$ and $2e^-$ (or the semiquinone derivative can be formed through the acceptance of a single proton and electron).

Respiratory Chain Complexes

The following respiratory chain complexes (I, III, and IV) are associated with proton translocation across the inner mitochondrial membrane.

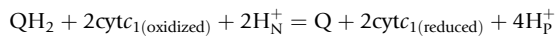
Complex I (NADH–Ubiquinone Oxidoreductase)

NADH–ubiquinone oxidoreductase is the largest (~ 1 MDa) and most complicated of the respiratory chain complexes, having ~ 43 subunits. Although it is well established that complex I contains FMN and 8–9 iron–sulfur proteins as redox centers, the exact route of electron transfer is uncertain. What is certain, however, is that the enzyme transfers two

protons and two electrons from NADH ($+H^+$) to ubiquinone, and electron transfer is accompanied by the translocation of protons across the membrane with a stoichiometry of $4H^+/2e^-$. Enzyme activity is potentially inhibited by rotenone and piericidin A. Although a crystal structure for complex I is yet to be solved, single particle electron microscopic studies suggest that the complex has an L-shape with the majority of the redox centers being located within the peripheral hydrophilic part of the enzyme. The exact mechanism whereby protons are translocated across the membrane is uncertain.

Complex III (Ubiquinone–Cytochrome *c* Oxidoreductase or the *bc*₁ Complex)

Complex III is a large (~250 kDa) multicomponent complex that catalyzes the oxidation of ubiquinol and the reduction of cytochrome *c*₁ and results in the vectorial transport of protons across the inner mitochondrial membrane. Redox centers include a 2Fe–2S protein (the Rieske center), two *b* cytochromes (cytochrome *b*_L and *b*_H), and cytochrome *c*₁. X-ray crystallography studies reveal that the enzyme is dimeric, with each of the two monomers being composed of at least 11 subunits (eight of which appear to have no catalytic role). Crystallography has also revealed that the monomers do not function independently but cooperate during electron transfer. Structural studies have also shown that the Rieske center shuttles between the quinone-binding site (of one of the monomers) in its oxidized state and cytochrome *c*₁ (of the second monomer) in its reduced state. Although the pathway of electron flow through complex III (known as the Q-cycle) is complicated, the net equation is



where N and P refer to the negative and positive sides, respectively, of the membrane (Figure 1).

The Q-cycle accommodates the fact that there is a mixture of two and one electron carriers and hence there must be a switch between ubiquinol (a two-electron carrier), the

cytochromes, and Rieske center (all of which are single-electron carriers). It also explains the well-documented observation that the H^+/e^- stoichiometry for the *bc*₁ complex is 4. Complex III is inhibited by antimycin A, myxothiazol, and stigmatellin.

Complex IV (Cytochrome *c*–Oxygen Oxidoreductase or Cytochrome *c* Oxidase)

Complex IV is a third large transmembrane enzyme (204 kDa) that catalyzes the oxidation of reduced cytochrome *c* and the reduction of molecular oxygen to water. It is also associated with the translocation of protons across the membrane and has at least 13 subunits of which only subunits I and II appear to play a role in catalysis. The complex contains three Cu atoms and two heme *a* (cytochrome *a* and cytochrome *a*₃). High-resolution structural studies reveal that subunit II contains two of the Cu atoms complexed with sulfur atoms to form a binuclear center (known as 'Cu_A center'). Subunit I contains the two cytochromes and the third Cu atom (Cu_B), which is closely associated with cytochrome *a*₃ forming a second binuclear center. Electron transfer through complex IV is from cytochrome *c* to the Cu_A center (which acts as a single-electron receptor), then to heme *a* (which is slightly closer to the Cu_A center than heme *a*₃) onto the heme *a*₃–Cu_B center and finally onto oxygen bound to heme *a*₃. For every four electrons passing through the complex, one molecule of oxygen is reduced to two molecules of water, four protons are consumed from the matrix, and four protons are translocated across the membrane. It should be noted that the four-electron reduction of molecular oxygen occurs as four single electron events and hence it is important that the potentially dangerous partially reduced intermediates such as superoxide anion are not released but remain tightly bound until completely reduced to water. Cytochrome *c* oxidase is potentially inhibited by cyanide, azide, and carbon monoxide, which bind at the oxygen-binding site. Nitric oxide is also a reversible inhibitor of this complex.

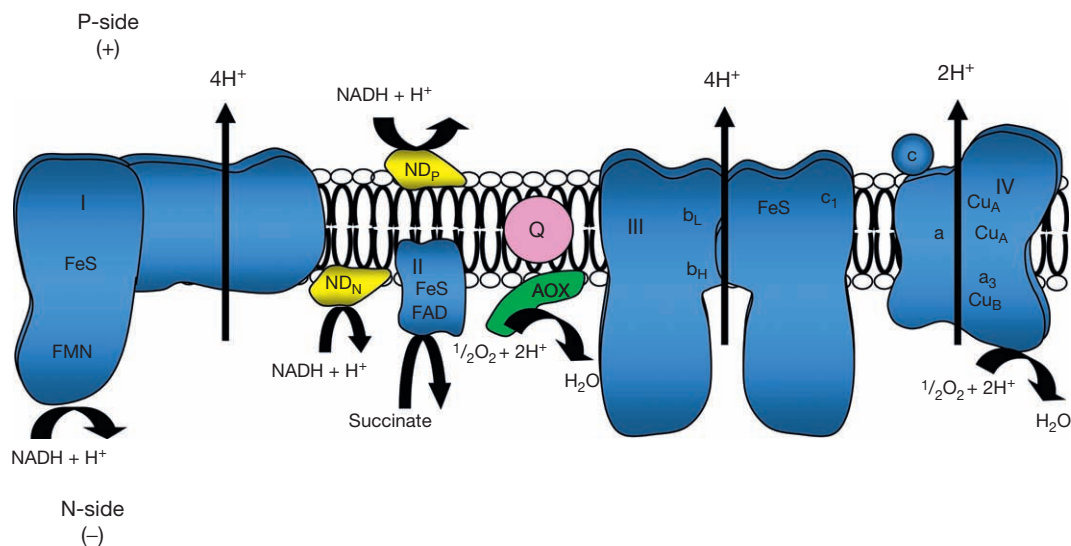


Figure 1 A schematic representation of the respiratory chain. The P- and N-sides of the membrane refer to the intermembrane space (positive) and matrix (negative), respectively. ND_P and ND_N refer to NADH dehydrogenases (such as the ETF, glycerol-3-phosphate, and externally located NADH dehydrogenases) located on the P- and N-sides, respectively. AOX is the alternative oxidase.

Other Ubiquinone-Reducing Enzymes

Animal, plant, and a number of protist mitochondria also contain a number of membrane-bound enzymes that can reduce ubiquinone. These enzymes do not transfer charge across the membrane, that is, protons are not vectorially transported across the inner mitochondrial membrane and are detailed in the following section.

Succinate dehydrogenase

Succinate dehydrogenase is the only membrane-bound enzyme of the citric acid cycle that interacts directly with the respiratory chain. Succinate dehydrogenase is much smaller than complex I (~140 kDa) being composed of four subunits. The two larger of these subunits are peripheral proteins that contain covalently bound FAD (on subunit I) and three iron-sulfur centers (on subunit II). These two large subunits are anchored to the inner surface of the inner membrane by two integral membrane polypeptides (subunits III and IV) in which a single heme *b* is located. Although the exact route of electron transfer from succinate to ubiquinone is uncertain, it is known that the reduction of ubiquinone by succinate is not associated with any charge movement.

Electron-transferring flavoprotein-ubiquinone reductase

The electron-transferring flavoprotein (ETF)-ubiquinone oxidoreductase is a globular protein located on the inner surface of the inner membrane and contains FAD, an Fe-S center, and a ubiquinone-binding site. The enzyme can accept reducing equivalents from a variety of dehydrogenases including those involved in fatty acid oxidation or amino acid and choline catabolism.

Glycerol-3-phosphate dehydrogenase

Glycerol-3-phosphate dehydrogenase is located on the outer surface of the inner membrane and catalyzes the oxidation of glycerol-3-phosphate to dihydroxyacetone phosphate and the reduction of ubiquinone. Although little is known about the structure or the route of electron transfer to ubiquinone, redox centers include FAD and probably an iron-sulfur center.

External NADH dehydrogenase

There are a number of rotenone-insensitive NADH dehydrogenases located on the outer and inner surface of the inner mitochondrial membrane in plant and yeast mitochondria. Again, there is little structural information apart from the presence of FAD and possibly an iron-sulfur center. Electron transfer is probably similar to that described for the ETF-ubiquinone oxidoreductase since the enzyme catalyzes the oxidation of NADH and the reduction of ubiquinone. In plants, the external NADH dehydrogenase (on the outer surface of the inner membrane) is regulated by cytosolic Ca^{2+} . These enzymes are not associated with proton pumping across the mitochondrial membrane.

Other Ubiquinol-Oxidizing Enzymes (the Alternative Oxidase)

Although mammalian mitochondria do not possess any other ubiquinol-oxidizing enzymes apart from cytochrome oxidase (complex IV), most plants, some yeasts, fungi, and trypanosomes possess a cyanide- and antimycin-insensitive alternative oxidase that catalyzes the oxidation of ubiquinol and reduces

molecular oxygen to water. Unlike cytochrome *c* oxidase, electron transfer does not result in proton translocation and the redox energy is released as heat. Spectroscopic and recent preliminary crystallographic studies indicate that the enzyme contains two nonheme iron atoms linked by an oxygen atom as its only redox carrier (di-iron protein) and is globular with two of its amphiphilic helices lying parallel to the membrane plane. Enzyme activity is insensitive to all complex IV inhibitors but can be potentially inhibited by salicylhydroxamic acid, alkyl gallates, and ascofuranone.

Organization of the Respiratory Chain

The respiratory chain complexes in mammalian and plant mitochondria are generally considered to be free to diffuse laterally and independently of each other within the plane of the membrane. They are connected electronically to each other by ubiquinone and cytochrome *c* with electron transfer occurring by diffusion-based collisions between these components. There is also evidence to suggest that the number of complexes within the membrane are unequal with the following ratio: complexes I:II:III:IV: cytochrome *c*: ubiquinone = 1:2:3:7:14:63. Both cytochrome *c* and ubiquinone are considered to function as common pools and hence act as mobile carriers mediating electron transfer between freely diffusible large respiratory complexes. In contrast to this fluid-state model, however, there is evidence for a solid-state model in which stable interactions are believed to occur between some of the respiratory chain complexes resulting in the formation of supercomplexes.

In both cases, however, for every $2e^-$ passing from NADH ($+H^+$) to oxygen to produce one molecule of water a total of 10 protons are released at the P-side of the membrane. Since four protons are required to synthesize and export one ATP, the amount of ATP formed by the oxidation of NADH is 2.5 (1.5 for succinate).

The ATP Synthase

The ATP synthase (or F_1F_0 ATPase and also referred to as 'complex V') uses the free energy of an electrochemical gradient of protons (or sodium ions) generated by the respiratory chain to synthesize ATP. The ATP synthases comprise a very large group of highly conserved enzymes that are found in the bacterial cytoplasmic membranes, the thylakoid membranes of chloroplasts, and the inner membranes of mitochondria. Most members of the group use H^+ as the coupling ion (the *Propionigenium modestum* enzyme is an example of the few ATP synthases that can use Na^+ as the physiological coupling ion).

Structure of the ATP Synthase

The ATP synthase is a miniature rotating motor and the simplest or prototype enzyme is that isolated from *Escherichia coli*. The enzyme has a bipartite structure composed of a membrane-bound (hydrophobic) F_0 component crossing the inner mitochondrial membrane and a water-soluble (hydrophilic) F_1 component that protrudes into the mitochondrial matrix.

The F_0 component

The F_0 entity is composed of three subunits (a , b , and c) with a stoichiometry of: $a:2b:9-12c$ and it functions as a proton channel. The proton channel is a ring of 9–12 c subunits (the number of subunits may depend on the carbon source used for growth) positioned in the membrane and tightly bound to one a -subunit plus two b -subunits of F_0 and also bound to the γ , δ , and ϵ subunits of the F_1 component of the synthase. The F_1 component is also attached to F_0 by the peripheral stalk or stator, an assembly of single copies of subunits b and d , oligomycin sensitivity conferral protein (OSCP), and F_6 .

The F_1 component

The F_1 component has five types of subunits (α , β , γ , δ , and ϵ) with a stoichiometry of $3\alpha:3\beta:\gamma:\delta:\epsilon$. The $\alpha_3\beta_3$ subunits are arranged as a hexamer of alternating α - and β -subunits with the γ -subunit projecting into the central cavity of this hexamer. The β -subunits contain the catalytic site for ATP synthesis/hydrolysis and bind the adenylates and inorganic phosphate.

ATP synthases from other sources all have very similar structures to the *E. coli* synthase except the mitochondrial enzyme. It contains an inhibitor protein (IF₁) and the OSCP factor. The mitochondrial δ -subunit is related to the ϵ -subunit of the *E. coli* synthase (Figure 2).

Mechanism of ATP Synthesis

ATP synthesis is driven by the rotation of the single γ -subunit, which extends into the center of the $\alpha_3\beta_3$ hexamer that contains the catalytic sites. As the single γ -subunit rotates so it alternately changes the conformation (and hence binding characteristics) of the catalytic sites. Changes in nucleotide affinity occur in the three β -subunits and arise from their

interactions with the γ -subunit. This results in the sequential binding of adenosine diphosphate (ADP) and P_i , followed by ATP synthesis and ATP release. This concept recognizes that the β -subunit can exist in three conformations: open, tight, or loose forms. These forms correspond to the binding, synthesis, and release of ATP and are driven by the rotation of the γ -subunit. The hydrolysis of ATP will drive the rotation of the γ -subunit in the opposite direction to that of synthesis.

Subunits c_{9-12} , γ , and ϵ constitute the rotor (rotating unit) and subunits α_3 , β_3 , δ , a , and b_2 constitute the stator or peripheral stalk (stationary unit). The proton-motive force drives rotation of the oligomeric ring of the c -subunits and of the γ -subunit in F_0 (since the γ -subunit is attached to the c -oligomer). Thus, protons moving from the P-side to the N-side of the inner membrane (i.e., down the electrochemical gradient) are converted into rotational torque of γ , which alters nucleotide binding at the β -catalytic site and results in the synthesis/hydrolysis of ATP. It is not the synthesis of ATP but its release from the synthase that is driven by the proton-motive force. Subunits γ and ϵ rotate relative to the $\alpha_3\beta_3$ hexamer to alter the nucleotide affinities (and hence binding) in the β -catalytic sites. Probably subunit γ of F_1 and two copies of subunit b in F_0 stabilize the synthase complex.

A simplified mechanism suggests that it is the flow of protons through c via a channel (or two half channels) that protonates an aspartate residue and the proton is released through the channel (or second half channel). Conformational changes during this cycle of protonation–deprotonation provide the basis to explain rotation. This protonated c subunit then rotates through 360° during ATP synthesis/hydrolysis and both the subunits γ and ϵ rotate by 120° for each ATP synthesized (or hydrolyzed) with the rotation being coupled to the protonation–deprotonation of three–four c subunits.

See also: **Bioenergetics:** ATP in Plant Mitochondria: Substrates, Inhibitors, and Uncouplers; Cytochrome bc_1 Complex (Respiratory Chain Complex III); Cytochrome c ; Iron–Sulfur Proteins; **Protein/Enzyme Structure Function and Degradation:** Flavins.

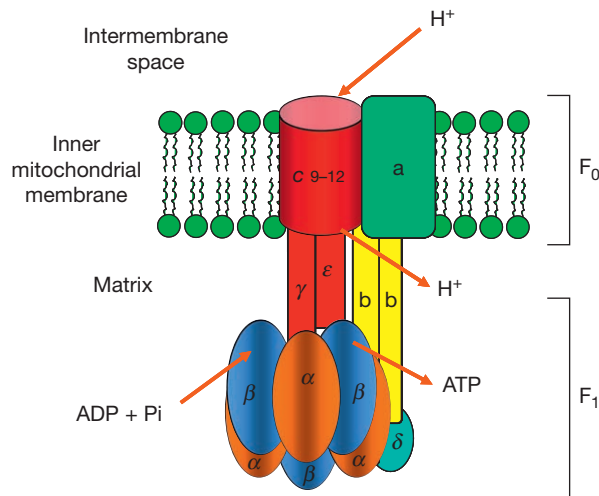


Figure 2 Diagram of the ATP synthase. Note that during ATP synthesis, subunits c , γ , and ϵ rotate as a complete unit with respect to the other (fixed) subunits.

Further Reading

- Abrahams JP, Leslie AG, Lutter R, and Walker JE (1994) The structure of F_1 -ATPase from bovine heart mitochondria determined at 2.8 Å resolution. *Nature (London)* 370: 621–628.
- Boyer PD (1997) The ATP synthase: A splendid molecular machine. *Annual Review of Biochemistry* 66: 717–749.
- Brandt U (2006) Energy converting NADH: Quinone oxidoreductase (complex I). *Annual Review of Biochemistry* 75: 69–92.
- Mitchell P (1979) Keilin's respiratory chain concept and its chemiosmotic consequences. *Science* 206: 1148–1159.
- Nelson DL and Cox MM (2000) *Lehninger Principles of Biochemistry*, 5th edn. New York: W.H. Freeman.
- Nicholls DG and Ferguson SJ (2008) *Bioenergetics 3*. London: Academic Press.
- Reesa DM, Leslie AGW, and Walker JE (2009) The structure of the membrane extrinsic region of bovine ATP synthase. *Proceedings of the National Academy of Sciences of the United States of America* 106: 21597–21601.
- von Ballmoos C, Wiedenmann A, and Dimroth P (2009) Essentials for ATP synthesis by F_1F_0 ATP synthases. *Annual Review of Biochemistry* 78: 649–672.

Respiratory Chain Complex I

U Brandt and V Zickermann, Goethe University, Frankfurt am Main, Germany

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by U. Brandt, volume 3, pp. 676–680, © 2004, Elsevier Inc.

Glossary

Accessory subunits Subunits of mostly eukaryotic respiratory chain complexes that are not considered to be essential for the bioenergetic core function, because they are not present in fully functional prokaryotic complexes. The function of most accessory subunits remains unclear, but involvement in assembly, stability, and regulation has been proposed.

Iron–sulfur clusters Nonheme iron prosthetic groups found in a wide range of redox enzymes. The reduced forms of binuclear and tetranuclear iron–sulfur clusters are paramagnetic and can be studied by electron-paramagnetic-resonance (EPR) spectroscopy.

Mitochondrial DNA Small genome that codes for 13 subunits of the respiratory chain complexes and adenosine triphosphate (ATP)-synthase. A typical mammalian cell

contains several hundreds of mitochondria and several thousands of mitochondrial genomes that are inherited exclusively through the oocyte cytoplasm.

Oxidative phosphorylation (OXPHOS) The fuel cell of aerobic metabolism. Process by which redox energy derived from reduced nicotinamide adenine dinucleotide (NADH) or reduced flavin adenine dinucleotide (FADH₂) is converted into the phosphorylation potential of ATP. Electron transport and ATP-synthesis are energetically coupled through an electrochemical proton potential across the inner mitochondrial membrane.

Reactive oxygen species (ROS) Superoxide (O₂^{•−}), hydrogen peroxide (H₂O₂), and hydroxyl radical (OH[•]) that may cause severe damage to lipids, proteins, and DNA. In mitochondria, ROS are generated by side reactions of the respiratory chain complexes.

Introduction to Complex I

Complex I (EC 1.6.5.3) is a multi-subunit respiratory chain reduced nicotinamide adenine dinucleotide (NADH):ubiquinone oxidoreductase that is found in the inner mitochondrial membrane of almost all eukaryotes and in the plasma membrane of many bacteria. It contains one molecule of flavin mononucleotide (FMN) and eight to nine iron–sulfur clusters as redox prosthetic groups.

Mammalian complex I couples the transfer of two electrons to the pumping of four protons across the mitochondrial inner membrane and thereby provides up to 40% of the driving force for adenosine triphosphate (ATP) generation by oxidative phosphorylation (OXPHOS). Complex I is the target of numerous pathological human mutations in both the mitochondrial and the nuclear genomes. Single electron transfer is part of its catalytic cycle and complex I is considered a major source for reactive oxygen species (ROS) that have been implicated in the etiology of neurodegenerative disorders and in aging.

Complex I Structure

Overall Shape

Electron microscopic analysis using single particles and two-dimensional (2D) crystals has demonstrated that the enzymes from mammalian and fungal mitochondria as well as eubacteria are L shaped with a hydrophobic membrane integral arm and a hydrophilic peripheral arm that protrudes into the mitochondrial matrix or the bacterial cytoplasm, respectively (Figure 1).

The X-ray structure of a peripheral arm fragment of complex I from the eubacterium *Thermus thermophilus* was solved at up

to 3.1-Å resolution. Lower-resolution crystallographic data permitted to construct an α -helical model for the major part of the membrane arm of complex I from *Escherichia coli* and to define the overall architecture of the complete enzyme complex from *T. thermophilus*. X-ray crystallographic analysis of complex I from the aerobic yeast *Yarrowia lipolytica* revealed the arrangement of functional modules in the much larger and more complex eukaryotic enzyme. Complex I from different species was shown to assemble into supercomplexes with respiratory complexes III and IV.

Central Subunits and Redox Clusters

The minimal form of complex I that is found in bacteria such as *T. thermophilus* typically consists of 14 subunits. These central subunits are conserved throughout in prokaryotes and eukaryotes and can be assigned to three main modules according to their bioenergetic core functions: (1) NADH oxidation (N-module), (2) ubiquinone reduction (Q-module), and (3) proton translocation (P-module) (Table 1).

The N-module and the Q-module contain the seven hydrophilic central subunits and all redox prosthetic groups namely one molecule of FMN and eight to nine iron–sulfur clusters. At present, their nomenclature is a matter of debate and there is only partial agreement on the assignment of electron-paramagnetic-resonance (EPR) signatures to individual iron–sulfur clusters (Table 1). The N-module resides in the distal and the Q-module in the proximal part of the peripheral arm (Figure 1). The primary electron acceptor FMN in the N-module is connected via a chain of iron–sulfur clusters with the ubiquinone reduction site near cluster N2 in the Q-module. For this cluster, a remarkable distance of about

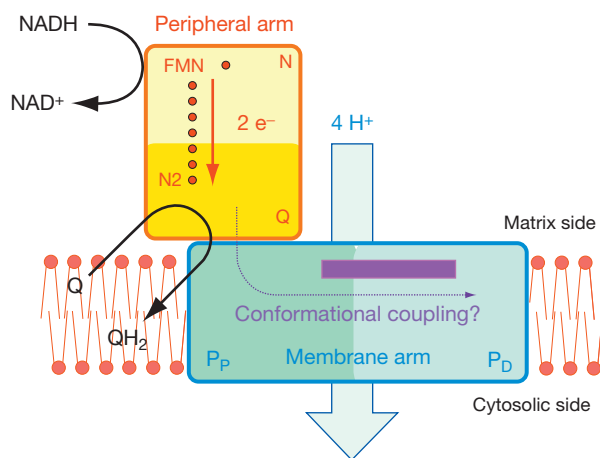


Figure 1 Schematic representation of mitochondrial complex I. The peripheral arm contains the seven nuclear-encoded central subunits and harbors all known redox prosthetic groups. The membrane arm contains the seven mitochondrially encoded subunits which are all highly hydrophobic. Eukaryotic complex I contains a number of additional accessory subunits distributed over both arms. A chain of iron–sulfur clusters (red spheres) connects the NADH oxidation site in the N-module with the ubiquinone reduction site near cluster N2 in the Q-module. One cluster may serve a special function in protection against formation of reactive oxygen species at the flavin site. The energy of the redox reaction drives vectorial proton translocation across the inner mitochondrial membrane. The proton translocation sites must be located in the membrane arm that is subdivided into a proximal P_P and a distal P_D domain. The architecture of complex I excludes direct coupling between electron transfer and proton pumping and rather suggests the transmission of energy via conformational changes. The helical element (magenta) connecting the P_P and P_D domains may play a crucial role in a conformational coupling mechanism.

30 Å to the membrane plane has been reported for eukaryotic complex I, very similar to the arrangement described for the prokaryotic enzyme. The flavin site of complex I has been identified as a major source of ROS. Cluster N1a in the N-module is not lined up in the long electron transfer chain with the other iron–sulfur clusters and might rather serve a critical function in protection against excessive ROS formation. The P-module contains the seven hydrophobic subunits ND1–ND6 and ND4L that are encoded by the mitochondrial genome in most eukaryotes. Based on recent structural evidence, the P-module can be subdivided into a proximal and a distal domain, P_P and P_D respectively. These two domains are connected by an extended helical element laterally associated with the membrane arm (Figure 1, magenta). This unprecedented structural feature has been suggested to serve as a transmission element for conformational energy (see section Proton Pumping).

Accessory Subunits

Eukaryotic complex I also contains a large number of additional so-called accessory subunits, resulting in a total number of more than 35 subunits in fungi and at least 45 subunits in mammals. While some of these subunits may simply serve as a scaffold to increase the stability of the enzyme, there are

indications that others may confer additional features specific to eukaryotic complex I. Some examples include the following: the acyl carrier protein (ACP) carries a phosphopantetheine prosthetic group and may function in the biosynthesis of lipoic acid or in mitochondrial lipid metabolism; several subunits of purified bovine complex I were found to be phosphorylated (42-kDa, ESSS, MWFE, B14.5a, B14.5b, and B16.6); phosphorylation at a cyclic adenosine monophosphate (cAMP)-dependent protein kinase consensus site of the AQDQ subunit stimulates complex I activity, presumably by promoting the import of this nuclear-encoded protein into mitochondria and assembly into the mature enzyme complex; the B16.6-kDa subunit is identical with the programmed cell-death-regulatory GRIM19 protein and may link complex I function to apoptosis; and the MWFE subunit of the membrane arm seems to be involved in complex I assembly. Overall, very little is known about the function of the accessory subunits.

A nomenclature established for the bovine enzyme by John Walker and co-workers has been generally accepted for these complex I subunits and their orthologs found in other organisms: Apart from some of the nuclear-coded subunits that were originally named by their molecular masses, the first four amino acids of the mature protein are used (Table 1). If the mature protein is N-terminally blocked, a number indicating the molecular mass preceded by the letter B identifies the subunit.

Complex I Assembly

The multistage assembly pathway of complex I is still only partially understood. In mutants of *Neurospora crassa*, deletion of two genes coding for CIA30 and CIA84 (for Complex I Associated) proteins led to complex I assembly defects. For human complex I, five assembly factors, NDUFAF1–NDUFAF4 and Ecsit have been identified. NDUFAF1 is the CIA30 homolog. In addition, Ind1 was shown to be essential for efficient assembly of complex I in the yeast *Y. lipolytica* and in humans. For identifying these factors, the complex one phylogenetic profile (COPP) approach, that is, the search for genes only present in species containing complex I, has proven very useful. The involvement in complex I assembly of several other proteins including the apoptosis-inducing factor (AIF) is unclear.

Complex I Evolution

Complex I was assembled from at least three preexisting structural and functional modules during evolution. The homology to [NiFe] hydrogenases from various archeal and eubacterial sources has been especially useful for the understanding of complex I.

Evolution of the Q-Module

Water-soluble hydrogenases from *Desulfovibrio* species consist of a large subunit which carries the [NiFe] site and is homologous to the 49-kDa subunit of complex I, and a small subunit which carries three iron–sulfur clusters and is homologous to the PSST subunit of complex I. In membrane-bound [NiFe] hydrogenases, exemplified by the enzyme encoded by the *ech* operon in *Methanosarcina barkeri*, the PSST homolog has

Table 1 The 14 central subunits of complex I; eukaryotic complex I contains a large number (31 in bovine heart) of additional accessory subunits

Subunit symbol					Redox prosthetic groups ^c	Functional or structural module ^d	Comments
<i>E. coli</i>	<i>T. thermophilus</i>	<i>Y. lipolytica</i> ^{a,b}	Bovine ^a	Human ^a			
NuoG ^e	Nqo3 ^e	NUAM	75 kDa	NDUFS1	N1b, 2 × [Fe ₄ S ₄]	N	NADH oxidation site, ROS formation protection against ROS?
NuoF	Nqo1	NUBM	51 kDa	NDUFV1	FMN, N3	N	
NuoE	Nqo2	NUHM	24 kDa	NDUFV2	N1a	N	
NuoD ^f	Nqo4	NUCM	49 kDa	NDUFS2		Q	
							Homologous to large subunit of soluble [NiFe] hydrogenase, ubiquinone binding site
NuoC ^f	Nqo5	NUGM	30 kDa	NDUFS3		Q	Homologous to part of small subunit of soluble [NiFe] hydrogenase, ubiquinone binding site
NuoI	Nqo9	NUIM	TYKY	NDUFS8	2 × [Fe ₄ S ₄]	Q	
NuoB	Nqo6	NUKM	PSST	NDUFS7	N2	Q	
NuoH	Nqo8	NU1M	ND1	ND1		P _P	Interface to peripheral arm, participates in inhibitor binding
NuoN	Nqo14	NU2M	ND2	ND2		P _P	Antipporter module
NuoA	Nqo7	NU3M	ND3	ND3		P _P	
NuoK	Nqo11	NULM	ND4L	ND4L		P _P	Antipporter module
NuoJ	Nqo10	NU6M	ND6	ND6		P _P	
NuoM	Nqo13	NU4M	ND4	ND4		P _D	Antipporter module
NuoL	Nqo12	NU5M	ND5	ND5		P _D	Antipporter module

^aSubunits ND1–ND6 and ND4L and NU1M–NU6M and NULM are encoded by mitochondrial genome.

^bTo avoid confusion and to simplify general discussions we use the nomenclature for bovine complex I also for central subunits of the enzyme from *Y. lipolytica*.

^cThe nomenclature and assignment of EPR signatures to individual [Fe₄S₄] clusters in the TYKY and 75-kDa subunits is discussed controversially.

^dCompare Figure 1.

^eIn many bacteria this subunit contains one additional [Fe₄S₄] cluster.

^fNuoC and NuoD are fused to a single polypeptide in *E. coli*.

suffered a C-terminal deletion abolishing two iron–sulfur clusters. Instead, an additional subunit carrying two iron–sulfur clusters in a ferredoxin-like fold has been acquired in this type of hydrogenase that is homologous to the TYKY subunit of complex I. Recent structural evidence confirmed the close relationship between water-soluble [NiFe] hydrogenases and the 49-kDa and PSST subunits of complex I. It follows that a significant part of the ubiquinone reduction site of complex I has been derived from the [NiFe] site in the large subunit, while cluster N2 corresponds to the proximal iron–sulfur cluster in the small subunit of water-soluble hydrogenases. Critical residues for binding of ubiquinone and inhibitors were identified by site-directed mutagenesis in the yeast *Y. lipolytica*.

Evolution of the N-Module

The electron input device of complex I is related to the nicotinamide adenine dinucleotide (NAD)⁺ reducing [NiFe] hydrogenase from *Ralstonia eutropha*, encoded by the *hox* operon. The 24- and 51-kDa subunits are homologous to HoxF and the first 200 residues of the 75-kDa subunit are homologous to HoxU. The C-terminal half of the 75-kDa subunit shows sequence similarity to a formate dehydrogenase from *Methanobacterium formicicum*.

Evolution of the P-Module

Four subunits of the membrane arm of complex I are related to subunits of a multi-subunit bacterial monovalent cation/proton antiporter encoded by the *mrp* operon. Subunits ND5, ND4L,

and ND2 correspond to subunits MrpA, MrpC, and MrpD, respectively, while subunit ND4 may have originated from a duplication of the ND2-encoding gene. In contrast to ND4L, the three largest hydrophobic subunits of complex I, ND2, ND4, and ND5, display a certain degree of structural similarity.

Origin and Variants of Complex I

According to an evolutionary scheme suggested by Hägerhall and colleagues, complex I and membrane-bound hydrogenases originate from a common ancestor that comprised subunits derived from [Ni–Fe] hydrogenases (see section [Evolution of the Q-Module](#)), the antiporter homologs (see section [Evolution of the N-Module](#)), and the membrane-bound subunit ND1. Loss of the [Ni–Fe]-binding site and acquisition of the hydrophobic subunits ND3 and ND6 generated a 11-subunit-membrane-bound protein complex that can be regarded as the last common ancestor of classical complex I and variants with different electron input modules.

Two archaeal proton-pumping enzymes, F₄₂₀H₂:methanophenazine oxidoreductase from *Methanosarcina mazei* and F₄₂₀H₂:menaquinone oxidoreductase from *Archaeoglobus fulgidus* accept electrons from coenzyme F₄₂₀, which is specific for methanogenic archaea. The metabolic function of complex I from cyanobacteria and the chloroplasts of higher plants is not known. It may catalyze electron transfer from either NADH, reduced nicotinamide adenine dinucleotide phosphate, or ferredoxin to plastoquinone.

Complex I from the bacteria *Helicobacter pylori* and *Campylobacter jejuni* lacks the 51- and 24-kDa subunits and contains a

modified version of the 75-kDa subunit. Therefore, it is likely that the electron donor to these enzymes is also not NADH.

Complex I Function

The overall reaction of respiratory chain complex I is the transfer of electrons from NADH to ubiquinone or menaquinone in some bacteria and the concomitant buildup of chemiosmotic membrane potential that in mammalian mitochondria contributes about 40% to the total proton-motive force generated during oxidative phosphorylation.

Electron Transfer and Redox Centers

The physiological reaction of complex I involves electron transfer from NADH via FMN onto a series of iron–sulfur clusters with E_{m7} values between -370 and -250 mV. The tetranuclear iron–sulfur cluster N2 ($E_{m7} -150$ mV) is generally regarded as the last redox center in this electron transfer chain and most likely is the immediate electron donor for ubiquinone. Electron transfer from NADH to artificial acceptors like ferricyanide or hexaammineruthenium(III)-chloride does not involve the quinone-binding site and can also be observed with a three-subunit subcomplex (flavoprotein, FP) of bovine heart complex I which consists of the peripheral arm subunits 51 kDa, 24 kDa, and 10 kDa.

Proton Pumping

Experimental data with intact mitochondria and submitochondrial particles indicate that complex I pumps protons with a stoichiometry of $4H^+/2e^-$. The mechanism how proton translocation is coupled to redox chemistry is still unknown. A large number of concepts and more or less detailed hypothetical mechanisms have been proposed over the last decades, but even the number of proton translocation sites remains an unresolved issue. Several lines of evidence indicate that the potential drop that releases the energy for proton pumping occurs during reduction of ubiquinone. A membrane-potential-dependent and inhibitor-sensitive semiquinone species in the vicinity of cluster N2 has been observed by Ohnishi and co-workers by EPR spectroscopy. The structural models recently obtained by X-ray crystallography demonstrate a spatial separation of the redox chemistry including the reduction of ubiquinone in the peripheral arm from the pumping modules in the membrane arm subunits. As there is evidence that even the distal membrane arm domain contributes to proton translocation, this architecture strongly suggests that complex I operates through a conformational coupling mechanism. The extended helical element connecting the two membrane arm domains (see section [Central Subunits and Redox Clusters](#)) has been suggested to serve a critical function for transmission of conformational energy.

Na⁺ Pumping by Bacterial Complexes I

It has been reported that in some bacteria complex I pumps sodium ions in addition or instead of protons across the plasma membrane, generating a sodium-motive force rather

than a proton-motive force. The relevance of these experimental data is discussed controversially. Most recently, a sodium/proton antiport activity has been suggested for some forms of bacterial complex I.

Inhibitors of Complex I

A plethora of structurally diverse hydrophobic compounds have been reported to inhibit complex I and are considered to interfere with ubiquinone binding. These inhibitors include naturally occurring antibiotics as well as synthetic insecticides and acaricides, also, a number of drugs such as amytal and meperidine, neurotoxins related to 1-methyl-4-phenyltetrahydropyridine (MPTP), and even detergents such as Triton X-100 and Thesit inhibit complex I. Kinetic studies suggested that these inhibitors can be grouped into three different classes that are represented by the well-known and widely used inhibitors DQA (2-decyl-4-quinazolinyl amine) and piericidin A (class I/A-type), rotenone (class II/B-type), and capsaicin (C-type). Except for capsaicin, all these inhibitors exert their inhibitory potential at I_{50} values in the nanomolar range. Competition experiments demonstrated that these inhibitors share one common binding domain with partially overlapping sites. In agreement with this observation, cross-resistance between DQA/piericidin A and rotenone is observed in inhibitor-resistant mutants.

Active–Deactive Transition

Mitochondrial complex I from a number of species is reversibly deactivated in the absence of substrates, and several catalytic cycles are needed to restore full activity. The selective exposition of a specific cysteine residue of subunit ND3 in the deactive form indicated that the active/deactive transition is linked to structural changes. The transformation into a dormant state under hypoxic conditions and its slow reversion to the fully active state during reoxygenation might be a protective mechanism against a burst of excessive ROS formation. Conversely, S-nitrosation of the deactive form in the presence of NO and oxygen radicals might be a pathophysiological event impeding the return of complex I into the active state.

Complex I in Disease

Genetic defects in complex I typically result in multisystem disorders with a wide spectrum of severities ranging from adult-onset exercise intolerance to fatal infantile lactic acidosis. Disease is often progressive. Exposure to complex I inhibitors has been described as one pathogenic mechanism in sporadic Parkinson's disease.

Mitochondrial Mutations

Deletions within the mitochondrial genome or point mutations in mitochondrial transfer RNA (tRNA) genes affect the function of several respiratory chain complexes simultaneously. Skeletal muscle biopsies from such patients typically show ragged red fibers caused by mitochondrial hypertrophy. The typical clinical phenotype of mitochondrial point mutations in

complex I genes is Leber's hereditary optic neuropathy (LHON), characterized by optic nerve atrophy or LHON plus dystonia (LDYT), characterized by additional lesions in the caudate and putamen. Ragged red fibers are not commonly observed in patients with isolated complex I defects.

Nuclear Mutations

Due to the large number of potential target genes, mutations affecting nuclear-coded subunits of complex I or assembly factors have been described only recently. The typical clinical picture is Leigh syndrome, a severe, early-onset neuromuscular disorder, associated with lactic acidosis, ataxia, hypotonia, spasticity, developmental delay, optic atrophy, and, in the terminal stage, degeneration of the basal ganglia. Sometimes, this is accompanied by hypertrophic cardiomyopathy. Such mutations were found in the 75-kDa, 51-kDa, 49-kDa, PSST, TYKY, and AQDQ subunits and in the assembly factor NDUF2.

Mechanisms of Complex I Pathology

Complex I dysfunction has several biochemical consequences which may be relevant for producing disease symptoms. (1) Feedback inhibition of the NADH-producing citric-acid cycle and pyruvate dehydrogenase complex results in elevated blood lactate levels. (2) Lowered OXPHOS efficiency will reduce ATP production. OXPHOS defects often show a remarkable degree of tissue and organ specificity. This has been explained by assuming different energetic thresholds, in descending order for optic nerve, neuronal cells, heart muscle, skeletal muscle, and liver. (3) Complex I is a major source for ROS and malfunctions of the enzyme may lead to increased oxidative stress as evident from the induction of manganese superoxide dismutase messenger RNA (mRNA) in complex I patients. ROS damage of complex I and the ND genes in the mitochondrial genome may result in a vicious cycle of mitochondrial damage. (4) Reduced mitochondrial energy production and elevated oxidative stress may lead to the opening of the mitochondria permeability transition pore, release of pro-apoptotic factors, and apoptotic cell death.

Poisoning of Complex I

The neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), a synthetic meperidine analog, is able to rapidly and persistently induce symptoms of acute Parkinsonism in humans and other primates. MPTP is metabolized to 1-methyl-4-phenylpyridinium (MPP⁺) by the action of glial cell monoamine oxidase B. As MPP⁺ is a substrate for the dopaminergic reuptake pathway, it accumulates in the nigrostriatal dopaminergic neurons, where it causes a specific reversible inhibition of complex I. Strikingly, highly selective neurodegeneration

of the nigrostriatal dopaminergic system was also observed following chronic exposure of rats to the common complex I inhibitor rotenone. These results support the view that complex I deficiency may also be involved in the etiology of idiopathic Parkinsonism.

See also: **Bioenergetics: Iron-Sulfur Proteins; Mitochondrial DNA; Protein/Enzyme Structure Function and Degradation: Flavins.**

Further Reading

- Arteni AA, Zhang P, Battchikova N, et al. (2006) Structural characterization of NDH-1 complexes from *Thermosynechococcus elongatus* by single particle electron microscopy. *Biochimica et Biophysica Acta* 1757: 1469–1475.
- Brandt U (2006) Energy-converting NADH: Quinone oxidoreductase (complex I). *Annual Review of Biochemistry* 75: 69–92.
- Carroll J, Fearnley IM, Skehel M, et al. (2006) Bovine complex I is a complex of 45 different subunits. *Journal of Biological Chemistry* 281: 32724–32727.
- Efremov RG, Baradaran R, and Sazanov LA (2010) The architecture of respiratory complex I. *Nature* 465: 441–445.
- Friedrich T and Scheide D (2000) The respiratory complex I of bacteria, archaea and eukarya and its module common with membrane-bound multisubunit hydrogenases. *FEBS Letters* 479: 1–5.
- Galkin A, Abramov AY, Frakich N, Duchon MR, and Moncada S (2009) Lack of oxygen deactivates mitochondrial complex I. *Journal of Biological Chemistry* 284: 36055–36061.
- Hunte C, Zickermann V, and Brandt U (2010) Functional modules and structural basis of conformational coupling in mitochondrial complex I. *Science* 329: 448–451.
- Janssen RJ, Nijtmans LG, van den Heuvel LP, and Smeitink JAM (2006) Mitochondrial complex I: Structure, function and pathology. *Journal of Inherited Metabolic Disease* 29: 499–515.
- Mathiessen C and Hägerhall C (2003) The "antiporter module" of respiratory chain complex I includes the MrpC/NuoK subunit – a revision of the modular evolution scheme. *FEBS Letters* 549: 7–13.
- McKenzie M and Ryan MT (2010) Assembly factors of human mitochondrial complex I and their defects in disease. *IUBMB-Life* 62: 497–502.
- Ohnishi T and Nakamaru-Ogiso E (2008) Were there any "misassignments" among iron-sulfur clusters N4, N5 and N6b in NADH-quinone oxidoreductase (complex I)? *Biochimica et Biophysica Acta* 1777: 703–710.
- Roessler MM, King MS, Robinson AJ, et al. (2010) Direct assignment of EPR spectra to structurally defined iron-sulfur clusters in complex I by double electron-electron resonance. *Proceedings of the National Academy of Sciences of the United States of America* 107: 1930–1935.
- Schägger H (2002) Respiratory chain supercomplexes of mitochondria and bacteria. *Biochimica et Biophysica Acta* 1555: 155–159.
- Tocilescu MA, Zickermann V, Zwicker K, and Brandt U (2010) Quinone binding and reduction by respiratory complex I. *Biochimica et Biophysica Acta* doi:10.1016/j.bbabo.2010.05.009.
- Zickermann V, Kerscher S, Zwicker K, et al. (2009) Architecture of complex I and its implications for electron transfer and proton pumping. *Biochimica et Biophysica Acta* 1787: 574–583.

Relevant Websites

- <http://www.mitomap.org> – MITOMAP: A Human Mitochondrial Genome Database.
- <http://www.scripps.edu/mem/ci> – The Scripps Research Institute; Complex I.

Respiratory Chain Complex II and Succinate:Quinone Oxidoreductases

C R D Lancaster, Saarland University, Homburg (Saar), Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

Flavin adenine dinucleotide (FAD) A riboflavin-containing hydrogen acceptor molecule in the Krebs cycle and a coenzyme of some oxidation–reduction enzymes.

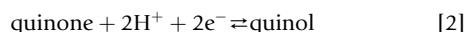
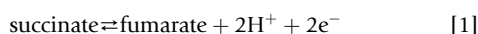
Fumarate A dicarboxylate anion ($^-\text{OOC}-\text{CH}=\text{CH}-\text{COO}^-$) that occurs as an intermediate compound in metabolic processes, most importantly in the citric acid cycle (Krebs cycle).

Heme group A complex organic red iron-containing pigment used by proteins to bind oxygen and/or electrons.

His ligand A histidine residue that participates in forming a complex around a central atom or ion, for example, the central iron of a heme group.

Succinate A dicarboxylate anion ($^-\text{OOC}-\text{CH}_2-\text{CH}_2-\text{COO}^-$) that occurs as an intermediate compound in metabolic processes, most importantly in the citric acid cycle (Krebs cycle).

Succinate:quinone oxidoreductases (SQORs; EC 1.3.5.1) are enzymes that couple the two-electron oxidation of succinate to fumarate (eqn [1]) to the two-electron reduction of quinone to quinol (eqn [2]):



They can also catalyze the opposite reaction, the coupling of quinol oxidation to the reduction of fumarate. Depending on the direction of the reaction catalyzed *in vivo*, the members of the superfamily of SQORs can be classified as either succinate:quinone reductases (SQRs) or quinol:fumarate reductases (QFRs). SQR and QFR complexes are anchored in the cytoplasmic membranes of archaeobacteria and eubacteria, and in the inner mitochondrial membrane of eukaryotes with the hydrophilic domain canonically extending into the cytoplasm and the mitochondrial matrix, respectively. SQR (respiratory complex II) is involved in aerobic metabolism as part of the citric acid cycle (Krebs cycle) and of the aerobic respiratory chain (Figure 1(a)). QFR participates in anaerobic respiration with fumarate as the terminal electron acceptor (Figure 1(b)), and is part of the electron transport chain catalyzing the oxidation of various donor substrates (e.g., H_2 or formate) by fumarate. These reactions are coupled via an electrochemical proton potential (Δp) to adenosine diphosphate phosphorylation with inorganic phosphate by adenosine triphosphate synthase.

SQOR Classification

Functional Classification

SQORs can be divided into three functional classes based on the quinone substrate and the *in vivo* function of the enzyme. Subclass 1 contains those enzymes that oxidize succinate (with an oxidation–reduction potential at pH 7, $E_{\text{M7}} = +25$ mV) and reduce a high-potential quinone, for example, ubiquinone ($E_{\text{M7}} = +90$ mV) or caldariella quinone ($E_{\text{M6.5}} = +103$ mV). The SQRs from mammalian mitochondria (respiratory complex II) and many prokaryotes belong to this group. Subclass 2 comprises all those enzymes that catalyze the oxidation of a

low-potential quinol, for example, menaquinol ($E_{\text{M7}} = -74$ mV) or rholoquinol ($E_{\text{M7}} = -63$ mV), and the reduction of fumarate. All QFRs studied so far belong to this subclass. Subclass 3 includes those enzymes that catalyze the oxidation of succinate and the reduction of a low-potential quinone, for example, menaquinone or thermoplasmaquinone. These subclass 3 enzymes are found in Gram-positive bacteria (e.g., *Bacillus subtilis*, *Paenibacillus macerans*, and *Bacillus licheniformis*) and archaeobacteria (e.g., *Thermoplasma acidophilum*).

Structural Classification

SQORs generally contain four protein subunits, referred to as A, B, C, and D. Subunits A and B are hydrophilic, whereas the subunits C and D are integral membrane proteins. Among species, subunits A and B have high sequence homology, while that for the hydrophobic subunits is much lower. Most of the SQR enzymes of Gram-positive bacteria and the QFR enzymes from epsilon-proteobacteria contain only one larger hydrophobic polypeptide (C), which is thought to have evolved from a fusion of the genes for the two smaller subunits C and D. Although subunit A harbors the site of fumarate reduction and succinate oxidation, the hydrophobic subunit(s) contain the site of quinol oxidation and quinone reduction.

Based on their hydrophobic domain and heme *b* content, SQORs can be classified into five types (Figure 2(a)). Type A enzymes contain two hydrophobic subunits and two heme groups, for example, SQR from the archaea *Archaeoglobus fulgidus*, *Natronomonas pharaonis*, and *Thermoplasma acidophilum*. Type B enzymes contain one hydrophobic subunit and two heme groups, as is the case for SQR from the Gram-positive bacteria *Bacillus subtilis* and *Paenibacillus macerans*, and QFR from the epsilon-proteobacteria *Campylobacter jejuni*, *Helicobacter pylori*, and *Wolinella succinogenes*. Examples for type C enzymes, which possess two hydrophobic subunits and one heme group, are SQR from mammalian mitochondria and from the proteobacteria *Paracoccus denitrificans* and *Escherichia coli*, and QFR from the nematode *Ascaris suum*. The QFR of *E. coli* is an example for a type D enzyme, which contains two hydrophobic subunits and no heme group. Finally, type E

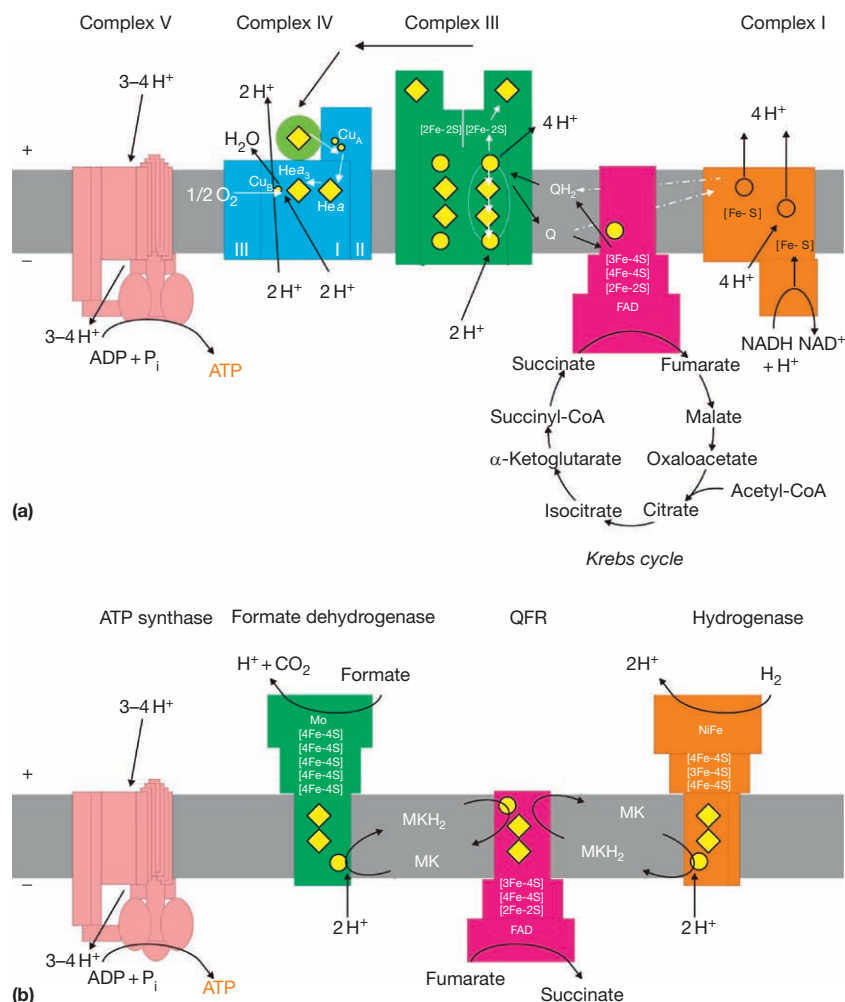


Figure 1 Electron flow and the generation and utilization of a transmembrane electrochemical potential in aerobic respiration (a) and anaerobic respiration (b). Modified from Lancaster CRD (2003) The role of electrostatics in proton-conducting membrane protein complexes. *FEBS Letters* 545: 52–60, with permission.

enzymes, such as SQRs from the archaea *Acidianus ambivalens* and *Sulfolobus acidocaldarius* and the cyanobacterium *Synechocystis*, and also the methylmenaquinol:fumarate reductases from the epsilon-proteobacteria *C. jejuni* and *W. succinogenes*, also contain no heme, but have two hydrophobic subunits very different from the other four types and more similar to those of heterodisulfide reductase from methanogenic archaea.

SQOR Structure and Function

Overall Description of the Structure

The first available crystal structures of SQORs (Figures 2(b)–2(d)) were those of two prokaryotic QFRs, both since 1999, and one prokaryotic SQR, since 2003. More recently, structures of mitochondrial complex II have become available, those of the porcine and avian enzymes in 2005 and 2006, respectively. The *E. coli* QFR (Figure 2(d)), determined at 3.3-Å resolution, belongs to the type D enzymes and the QFR of *W. succinogenes* (Figure 2(b)), refined initially at 2.2 Å and later at 1.78-Å resolution, is of type B. The structure of the SQR from *E. coli*

(Figure 2(c)) was reported at 2.6-Å resolution, those of the porcine and avian enzymes at resolutions of 2.4 and 1.8 Å, respectively. The latter three structures are very similar. Interestingly, *E. coli* QFR appears to be a monomeric complex of one copy each of the A, B, C, and D subunits, whereas *W. succinogenes* QFR is apparently a homodimer of two sets of A, B, and C subunits, and *E. coli* SQR is a homotrimer of three sets of A, B, C, and D subunits. At least for the B-type enzymes from *W. succinogenes*, *C. jejuni*, and *H. pylori*, it has been derived from analytical gel filtration and analytical ultracentrifugation experiments, that the homodimer is apparently also present in the detergent-solubilized state of the enzyme, implying that it is unlikely to be an artifact of crystallization. However, functionally all three enzymes appear to act as monomers, which is why only the monomeric complexes are shown in Figures 2(b)–2(d).

The Relative Orientations of Soluble and Membrane-Embedded SQOR Subunits

The structures of all three enzymes shown in Figures 2(b)–2(d) can be superimposed on the basis of the Cα positions of the

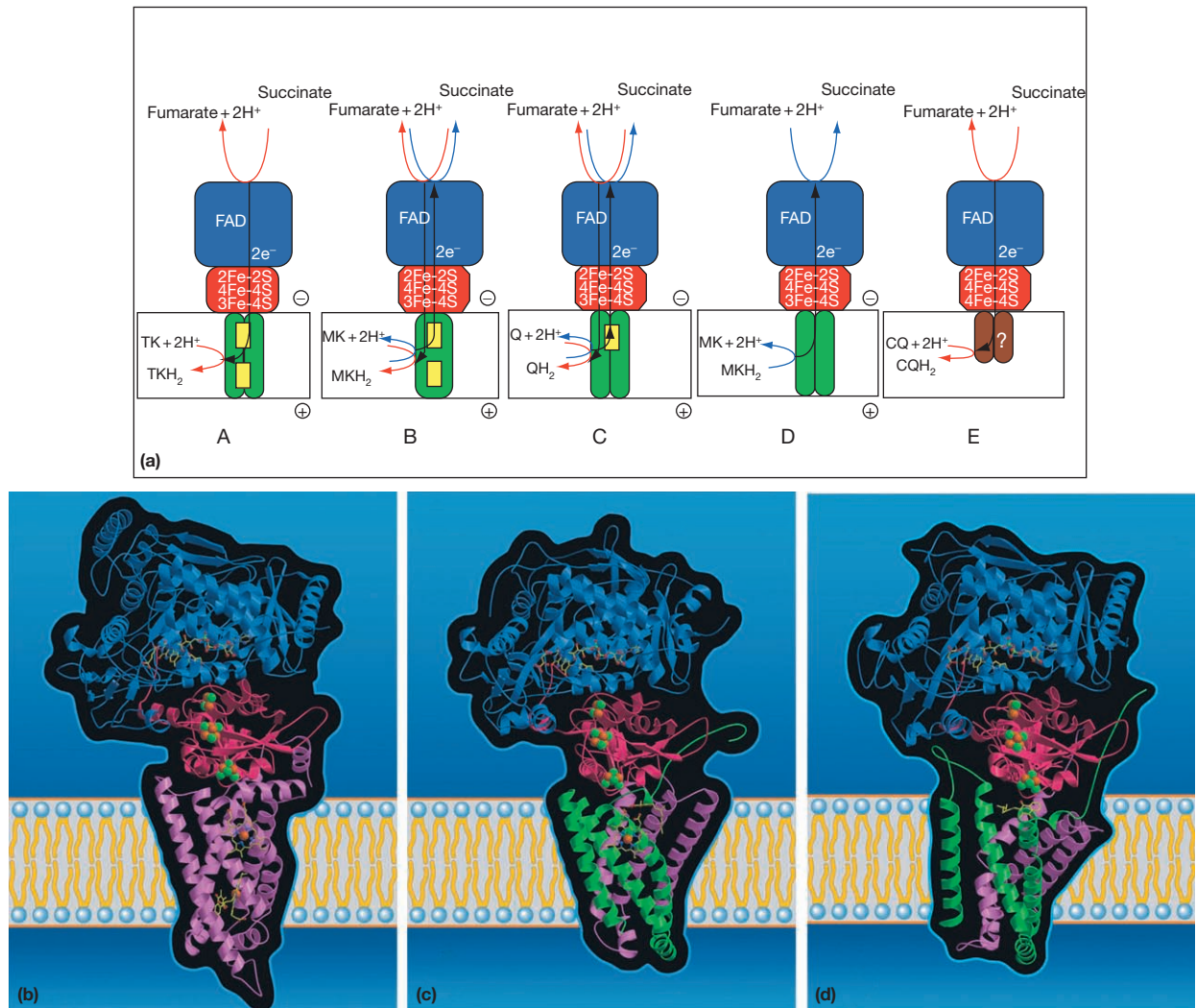


Figure 2 (a) Classification (A–E) of succinate:quinone oxidoreductases based on their hydrophobic domain and heme content. The hydrophilic subunits A and B are drawn schematically in blue and red, respectively, the hydrophobic subunits C and D in green. Heme groups are symbolized by small rectangles. The directions of the reactions catalyzed by SQR and QFR are indicated by red and blue arrows, respectively. Yellow rectangles symbolize the respective cytoplasmic or inner mitochondrial membrane bilayer. The positive (+) and negative (–) sides of the membrane are indicated. In bacteria, the negative side is the cytoplasm (inside) and the positive side is the periplasm (outside). For mitochondrial systems, these are the mitochondrial matrix and the intermembrane space, respectively. The type of quinone transformed *in vivo* is not necessarily unique for each type of enzyme. The examples given are thermoplasma-quinone (TK), menaquinone (MK), ubiquinone (Q), and caldariella quinone (CQ). (b–d) Three-dimensional structures of *W. succinogenes* QFR, a B-type SQOR ((b) determined by Lancaster et al.), *E. coli* SQR, a C-type SQOR ((c) determined by Yankovskaya et al.), and *E. coli* QFR, a D-type SQOR ((d) determined by Iverson et al.). The Cα traces of the A subunits are shown in blue, those of the B subunits in red, those of the C subunits in pink, and those of the D subunits in green. The atomic structures of the prosthetic groups are superimposed for better visibility. From top to bottom, these are the covalently bound FAD, the [2Fe–2S], the [4Fe–4S], and the [3Fe–4S] iron–sulfur clusters, (and, where present) the proximal and the distal heme b groups. Atomic color-coding is as follows: C, N, O, P, S, and Fe are displayed in yellow, blue, red, light green, green, and orange, respectively. Figures with atomic models were prepared with a version of Molscript (Kraulis PJ (1991) Molscript: A program to produce both detailed and schematic plots of protein structures. *Journal of Applied Crystallography* 24: 946–950.), modified for color ramping (Esnouf RM (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *Journal of Molecular Graphics and Modelling* 15: 132–134.), and rendered with the program Raster3D (Merritt EA and Bacon DJ (1997) Raster3D photorealistic molecular graphics. *Methods in Enzymology* 277: 505–524.) (a) Modified from Lancaster CRD (2001) Succinate:quinone oxidoreductases – what can we learn from *Wolinella succinogenes* quinol:fumarate reductase? *FEBS Letters* 504: 133–141, with permission.

conserved hydrophilic subunits A and B. In this superimposition, large parts of the membrane-embedded subunits can be aligned only in the case of *W. succinogenes* QFR and *E. coli* SQR, but not for *E. coli* QFR. However, in an alternate orientation, the transmembrane subunits of *E. coli* QFR and of *W. succinogenes*

QFR can be overlaid in spite of the respective absence and presence of the two heme groups. Consequently, the relative orientations of the hydrophilic subunits and the transmembrane subunits are similar in *W. succinogenes* QFR, *E. coli* SQR, and mitochondrial complex II and different in *E. coli* QFR.

Subunit A, and the Site of Succinate Oxidation/Fumarate Reduction

The flavoprotein or A subunit (64–73 kDa) of all described membrane-bound SQOR complexes contains a flavin adenine dinucleotide (FAD) prosthetic group covalently bound to a conserved His residue as an α -[N ϵ -histidyl]-FAD. The two most important domains of subunit A are the bipartite FAD-binding domain and the 'capping' domain, which is inserted between the two parts of the FAD-binding domain. The capping domain contributes to burying the otherwise solvent-exposed FAD isoalloxazine ring from the protein surface.

The binding site of succinate/fumarate is located between the FAD-binding domain and the capping domain next to the plane of the FAD isoalloxazine ring. The structures and results from site-directed mutagenesis suggest that the *trans* dehydrogenation of succinate to fumarate occurs by the combination of a hydride ion transfer from one succinate methylene group to the N5 position of the FAD and of a proton transfer from the other succinate methylene group to the side chain of a conserved Arg residue of the capping domain (Figure 3). All residues implicated in substrate binding and catalysis are conserved throughout the superfamily of SQORs, so that this reversible mechanism is considered generally relevant for all SQORs. Release of the product could be facilitated by movement of the capping domain away from the dicarboxylate site.

Subunit B, the Iron–Sulfur Protein

Generally, SQORs contain three iron–sulfur clusters, which are exclusively bound by the B subunit of 26–30 kDa. Enzyme types A–D contain one [2Fe–2S]^{2+,1+}, one [4Fe–4S]^{2+,1+}, and one [3Fe–4S]^{1+,0} cluster, whereas an additional [4Fe–4S]^{2+,1+} cluster apparently replaces the [3Fe–4S]^{1+,0} in the type E

enzyme. This subunit consists of two domains, an N-terminal 'plant ferredoxin' domain, in contact with subunit A and binding the [2Fe–2S] iron–sulfur cluster, and a C-terminal 'bacterial ferredoxin' domain, in contact with the hydrophobic subunit(s) and binding the [4Fe–4S] and the [3Fe–4S] iron–sulfur clusters. In general, these iron–sulfur clusters are coordinated by conserved Cys residues, although one of the ligands to the [2Fe–2S] cluster in *E. coli* SQR is an Asp. At the position corresponding to the fourth Cys of the [4Fe–4S] cluster, the [3Fe–4S] cluster contains a Leu (*W. succinogenes* QFR), Ile (*E. coli* SQR), Val (*E. coli* QFR), or Ser (*B. subtilis* SQR), whereas the introduction of a Cys into *E. coli* QFR could replace the native [3Fe–4S] by a [4Fe–4S] cluster, this was not the case for *B. subtilis* SQR.

The Integral Membrane Subunit(s) C (and D) and the Sites of Quinol Oxidation/Quinone Reduction

Type A, C, and D SQOR enzymes contain two hydrophobic polypeptides with three membrane-spanning helices each (numbered I, II, and III and IV, V, and VI, respectively). According to an evolutionary model proposed by Hägerhäll and Hederstedt in 1996, the large single hydrophobic polypeptides of type B SQOR enzymes with five membrane-spanning helices are thought to have evolved from the fusion of the genes for the two small hydrophobic polypeptides with concomitant loss of transmembrane helix III. This view is supported by the structural superpositions discussed above (section IIB). To a varying degree, all transmembrane segments are tilted with respect to the membrane normal.

The planes of both heme molecules bound by *W. succinogenes* QFR are approximately perpendicular to the membrane surface and their interplanar angle is 95°. The His axial ligands to the 'proximal' heme *b_p*, located toward the cytoplasmic

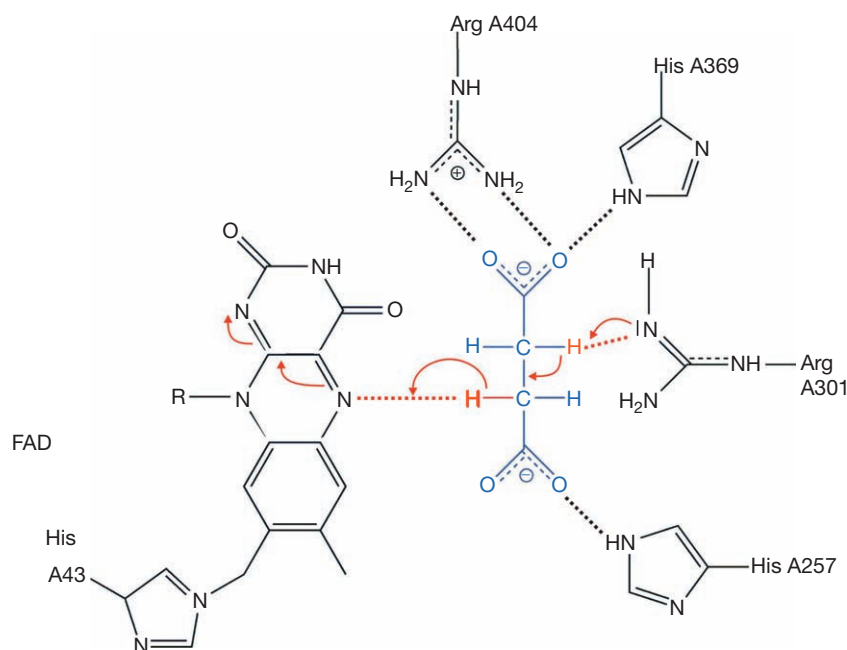


Figure 3 Possible mechanism of succinate oxidation/fumarate reduction. Modified from Lancaster CRD (2001) Succinate:quinone oxidoreductases – what can we learn from *Wolinella succinogenes* quinol:fumarate reductase? *FEBS Letters* 504: 133–141, with permission.

surface of the membrane and, thus, toward the [3Fe–4S] iron–sulfur cluster, are located on transmembrane helices II and V. Also, residues of the two other transmembrane helices I and IV interact with the propionate groups of heme b_p via hydrogen bonds and salt bridges, which underscores the structural importance of the bound heme. The axial ligands to the ‘distal’ heme b_D are located on helices I and IV. The binding of the two heme b molecules by an integral membrane protein four-helix bundle described here is very different from that described for other diheme-binding membrane protein complexes, such as the cytochrome bc_1 complex, where only two transmembrane segments provide two axial heme b ligands each, and also the membrane-bound hydrogenases and formate dehydrogenases, where one transmembrane helix provides two axial His ligands and two others provide one His ligand each. One consequence of this difference is that the distance between the two heme iron centers is distinctly shorter in *W. succinogenes* QFR (15.6 Å) than it is in the mitochondrial cytochrome bc_1 complex (21 Å) and in *E. coli* formate dehydrogenase-N (20.5 Å). The mode of heme binding for the single (proximal) heme of *E. coli* SQR is analogous to that described for heme b_p of *W. succinogenes* QFR.

The sites of quinol oxidation/quinone reduction have been localized in all three enzymes of known three-dimensional structure by a combination of X-ray crystallography, site-directed mutagenesis, functional characterization of resulting variant enzymes, and/or inhibitor-binding studies, respectively. In *E. coli* QFR, the menaquinol oxidation site is located proximally, close to the [3Fe–4S] cluster. In *E. coli* SQR, this position is occupied by the heme b group, but the ubiquinone reduction site is also located proximally, close to the [3Fe–4S] cluster, at an alternate position. In *W. succinogenes* QFR, the menaquinol oxidation site is located distally, close to heme b_D .

Electron and Proton Transfer

Electron Transfer

For the function of QFR, electrons have to be transferred from the quinol-oxidizing site in the membrane to the fumarate-reducing site, protruding into the cytoplasm. Conversely, for the function of SQR, electrons have to be transferred from the succinate oxidation site in subunit A to the quinone reduction site in the membrane. The linear arrangement of the prosthetic groups in the complexes shown in **Figures 2(b)–2(d)**, therefore, provides one straightforward pathway by which electrons could be transferred efficiently between the two sites of catalysis. It has been shown for other electron transfer proteins that physiological electron transfer between prosthetic groups occurs if the edge-to-edge distances relevant for electron transfer are shorter than 14 Å, but not if they are longer than 14 Å. In all three cases, this indicates that physiological electron transfer can occur between the prosthetic groups of one heterotrimeric or heterotetrameric complex, but not between the two (*W. succinogenes* QFR) or three (*E. coli* SQR) complexes in the respective homodimer or homotrimer.

Prior to the determination of the three-dimensional structures, because of its very low midpoint potential ($E_m < -250$ mV), the [4Fe–4S] iron–sulfur cluster had been suggested not to participate in electron transfer. However, the

determined low potential may be an artifact due to anti-cooperative electrostatic interactions between the redox centers. The position of the [4Fe–4S] cluster as revealed in the structures is highly suggestive of its direct role in electron transfer between the [3Fe–4S] cluster and the [2Fe–2S] cluster. Despite this major thermodynamically unfavorable step, the calculated rate of electron transfer is on a microsecond scale, demonstrating that this barrier can easily be overcome by thermal activation as long as the electron transfer chain components are sufficiently close to promote intrinsically rapid electron tunneling.

The positioning of the heme in *E. coli* SQR is apparently puzzling because it does not seem to be required for electron transfer between the catalytic sites. A possible explanation for this also explains why SQR is favored over QFR in *E. coli* when oxygen is present. Under aerobic conditions, reduced *E. coli* QFR produces large amounts of reactive oxygen species (ROS), including superoxide radical (O_2^-) and hydrogen peroxide (H_2O_2), which has been suggested to be caused by electron accumulation around the FAD. In sharp contrast, *E. coli* SQR reacts poorly with molecular oxygen, producing modest amounts of O_2^- and no H_2O_2 . On the basis of their crystal structure, Yankovskaya et al., argue that the heme group in SQR provides an electron sink when the quinone site is not occupied, thus preventing buildup of electrons around the FAD and subsequent ROS generation. However, this electron-sink mechanism is predicted to be less effective for mitochondrial SQRs because heme b has a lower-redox potential than in *E. coli* SQR.

Electron-Coupled Proton Transfer

In addition to the transfer of electrons, two protons are bound at the site of reduction and two protons are liberated at the site of oxidation (see **eqns [1] and [2]**). An overview of the current status of discussion of electron and proton transfer in SQORs is shown in **Figures 4(a) and 4(b)**. In mitochondrial complex II and other C-type enzymes, such as SQR from *P. denitrificans* and *E. coli*, electron transfer from succinate to ubiquinone does not lead to the generation of a transmembrane electrochemical potential Δp , because the protons released by succinate oxidation are on the same side of the membrane as those consumed by quinone reduction (**Figure 4(a)**). Similarly, in *E. coli* QFR, the protons released by proximal quinol oxidation are balanced by the protons consumed by fumarate reduction (**Figure 4(c)**). Succinate oxidation by menaquinone, an endergonic reaction under standard conditions, is catalyzed by a B-type SQOR in Gram-positive bacteria, for example, *B. subtilis*. There is experimental evidence indicating that the menaquinone reduction site in *B. subtilis* SQR is close to the heme b_D in an analogous position to the menaquinol oxidation site of *W. succinogenes* QFR. This arrangement of the catalytic sites of succinate oxidation and menaquinone reduction would allow succinate oxidation by menaquinone in *B. subtilis* to be driven by the electrochemical proton potential (**Figure 4(b)**) and there is indeed experimental evidence that this is the case. This is the analogous reaction to that suggested by the arrangement of the catalytic sites for *W. succinogenes* QFR, but in the opposite direction. Recent experimental results indeed indicate that *B. subtilis* SQR generates a proton potential when

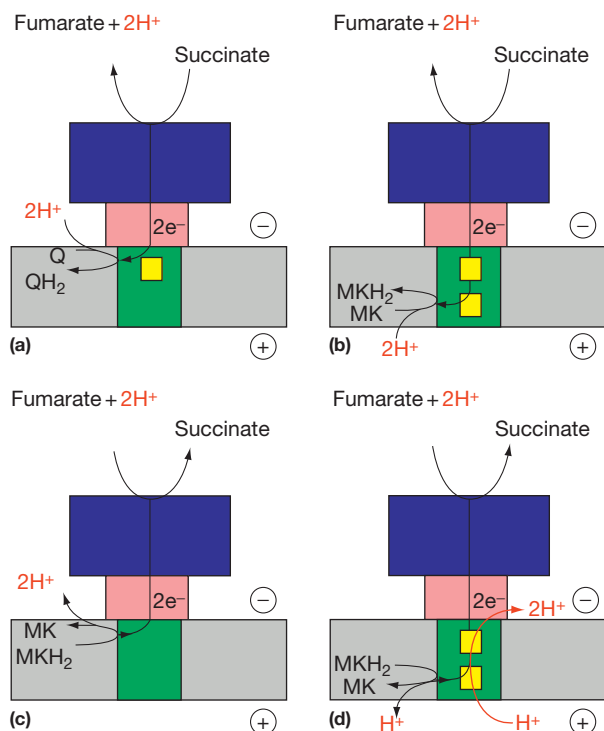


Figure 4 The coupling of electron and proton flow in succinate:quinone oxidoreductases in aerobic (a) and (b) and anaerobic respiration (c) and (d), respectively. Positive and negative sides of the membrane are described for Figure 2. (a, c) Electroneutral reactions as catalyzed by C-type SQR enzymes (a) and D-type *E. coli* QFR (c). (b) Utilization of a transmembrane electrochemical potential Δp as possibly catalyzed by A-type and B-type enzymes. (d) Electroneutral fumarate reduction by B-type QFR enzymes with the compensatory 'E-pathway'. Modified from Lancaster CRD (2001) Succinate:quinone oxidoreductases – what can we learn from *Wolinella succinogenes* quinol:fumarate reductase? *FEBS Letters* 504: 133–141, with permission.

functioning as a QFR. However, the experimental results on intact bacteria, with inverted vesicles or liposomes containing *W. succinogenes* QFR indicate that the oxidation of menaquinol by fumarate as catalyzed by *W. succinogenes* QFR is an electro-neutral process. To reconcile these experimental findings with the arrangement of the catalytic sites in the structure, it has been proposed that transmembrane electron transfer in dihemmic QFRs is tightly coupled to the compensatory, parallel transfer of protons across the membrane, thus balancing the protons released to the periplasm upon menaquinol

oxidation and the protons bound from the cytoplasm upon fumarate reduction (Figure 4(d)). A key residue in this proposed proton transfer pathway is a Glu residue located in the middle of the membrane that is conserved only in dihemmic menaquinol:fumarate reductases, but not in dihemmic succinate:menaquinone reductases. Experimental results proving this 'E-pathway hypothesis' have recently been obtained.

See also: Bioenergetics: Ferredoxin; Ferredoxin-NADP⁺ Reductase; F_X, F_A, and F_B Iron-Sulfur Clusters in Type I Photosynthetic Reaction Centers; Iron-Sulfur Proteins; Protein/Enzyme Structure Function and Degradation: Flavins.

Further Reading

- Ackrell BAC, Johnson MK, Gunsalus RP, and Cecchini G (1992) Structure and function of succinate dehydrogenase and fumarate reductase. In: Mueller F (ed.) *Chemistry and Biochemistry of Flavoenzymes*, vol. 3, pp. 229–297. Boca Raton, FL: CRC Press.
- Cecchini G (2003) Function and structure of complex II of the respiratory chain. *Annual Review of Biochemistry* 72: 77–109.
- Hägerhäll C (1997) Succinate:quinone oxidoreductases: Variations on a conserved theme. *Biochimica Biophysica Acta – Bioenergetics* 1320: 107–141.
- Hägerhäll C and Hederstedt L (1996) A structural model for the membrane-integral domain of succinate:quinone oxidoreductases. *FEBS Letters* 389: 25–31.
- Hederstedt L (2003) Complex II is complex too. *Science* 299: 671–672.
- Huang L-S, Shen JT, Wang AC, and Berry EA (2006) Crystallographic studies of the binding of ligands to the dicarboxylate site of complex II, and the identity of the ligand in the "oxaloacetate-inhibited" state. *Biochimica Biophysica Acta* 1757: 1073–1083.
- Iverson TM, Luna-Chavez C, Cecchini G, and Rees DC (1999) Structure of the *Escherichia coli* fumarate reductase respiratory complex. *Science* 284: 1961–1966.
- Juhnke HD, Hiltcher H, Nasiri HR, Schwalbe H, and Lancaster CRD (2009) Production, characterization and determination of the real catalytic properties of the putative 'succinate dehydrogenase' from *Wolinella succinogenes*. *Molecular Microbiology* 71: 1088–1101.
- Lancaster CRD (2002) Special issue on fumarate reductases and succinate dehydrogenases. *Biochimica et Biophysica Acta* 1553: 1–176.
- Lancaster CRD (2003) The structure of *Wolinella succinogenes* quinol: fumarate reductase and its relevance to the superfamily of succinate: quinone oxidoreductases. *Advances in Protein Chemistry* 63: 131–149.
- Lancaster CRD (2003b) *Wolinella succinogenes* quinol: fumarate reductase and its comparison to *Escherichia coli* succinate: quinone reductase. *FEBS Letters* 555: 21–28.
- Lancaster CRD, Kroger A, Auer M, and Michel H (1999) Structure of fumarate reductase from *Wolinella succinogenes* at 2.2 angstrom resolution. *Nature* 402: 377–385.
- Madej MG, Muller FG, Ploch J, and Lancaster CRD (2009) Limited reversibility of transmembrane proton transfer assisting transmembrane electron transfer in a dihaem-containing succinate:quinone oxidoreductase. *Biochimica et Biophysica Acta* 1787: 593–600.
- Sun F, Huo X, Zhai YJ, et al. (2005) Crystal structure of mitochondrial respiratory membrane protein complex II. *Cell* 121: 1043–1057.
- Yankovskaya V, Horsefield R, Tornroth S, et al. (2003) Architecture of succinate dehydrogenase and reactive oxygen species generation. *Science* 299: 700–704.

Respiratory Chain Complex IV

H Michel, Max Planck Institute of Biophysics, Frankfurt, Germany

© 2013 Elsevier Inc. All rights reserved.

Glossary

α -Proteobacteria A group of eubacteria containing many well-characterized bacterial species. The evolutionary ancestor of mitochondria is most likely a close relative of the *Rickettsiaceae* which belong to this group of bacteria.

Cyanobacteria Prokaryotes that perform oxygenic photosynthesis. They developed the water-splitting system

converting the atmosphere to an oxidizing one. They are considered to be the evolutionary ancestor of chloroplasts.

Cytochrome *c* A heme-containing hydrophilic protein that functions as an electron carrier in respiration.

Ubiquinol A hydrophobic coenzyme which carries hydrogen atoms in membranes. It can release protons and electrons separately. It forms ubiquinone upon oxidation.

Background

Aerobic, oxygen-consuming respiration is the major means of producing energy in animals, fungi, many yeasts, and bacteria. Oxygen itself is a side product of the oxygenic photosynthesis occurring in plants, algae, and cyanobacteria. These organisms produce reduced coenzymes (nicotinamide adenine dinucleotide phosphate (NADPH), quinols) and adenosine-5'-triphosphate (ATP) in the course of the photosynthetic light reaction. NADPH is used to fix carbon dioxide and to synthesize carbohydrates and other organic compounds during the photosynthetic dark reactions. This organic matter is consumed by heterotrophic organisms. NADH and quinols are formed during the degradation of foodstuff and oxidized in the respiratory chain. In the typical mitochondrial respiratory chain (Figure 1), NADH is oxidized by complex I (NADH dehydrogenase). The electrons are transferred onto ubiquinone; the ubiquinol formed is oxidized by complex III (cytochrome *bc*₁ complex, also called ubiquinol–cytochrome *c* oxidoreductase) which reduces cytochrome *c*. This electron carrier is oxidized again by complex IV (cytochrome *c* oxidase). The latter uses the electrons to reduce oxygen and to form water, thus closing the cycle. Complex II (succinate dehydrogenase) constitutes a parallel entry into the respiratory chain and is also part of the citric acid cycle.

The respiratory chains from aerobic bacteria contain many more membrane–protein complexes and are often branched. They may contain up to four different terminal oxidases reacting with molecular oxygen. The electron-donating substrates for the individual terminal oxidases may be cytochromes of the *c* type, various quinols, or even high-potential iron sulfur proteins. Most of them belong to the family of heme/copper-containing terminal oxidases. The exception is cytochrome *bd*, which is a ubiquinol oxidase and not related to the heme/copper-containing terminal oxidases. Cytochromes *bd* do not pump protons; however, electrons and protons, as in the heme/copper-containing oxidases, enter the active site of this enzyme from opposite sides of the membrane leading to a net translocation of four charges per oxygen molecule consumed.

Astonishingly, the superfamily of the heme/copper-containing terminal oxidases appears to be rather old. Family members are found in bacteria, as well as in archaea. They

splitting into these lineages during evolution predates the conversion of the atmosphere from a reducing one to an oxidizing one, which dates back to ~2 billion years, when cyanobacteria developed the water-splitting, oxygen-releasing photosystem II. It turned out that the nitric oxide reductase (nitrogen monoxide reductase, NO reductase) is closely related to the heme/copper-containing terminal oxidases. An iron atom in the NO reductases replaces a copper atom which is a part of the active site in the heme/copper-containing terminal oxidases. NO reductases appear to be unable to pump protons. The original task of both the heme/copper-containing terminal oxidases (which are *OO* reductases) and the NO reductases might have been detoxification. Molecular oxygen, being toxic for the organisms living under the reducing atmosphere, might have been created by photolysis of water by ultraviolet light and had to be destroyed.

Prosthetic Groups

The heme/copper-containing terminal oxidases minimally contain two heme groups and one copper atom as prosthetic groups. One of the heme groups is a high-spin heme. A copper atom, called Cu_B, is found in ~5Å distance. Cu_B and the high-spin heme form the so-called 'binuclear center'. During turnover, molecular oxygen is bound to the central iron atom of the high-spin heme between the heme and Cu_B and reduced there during the catalytic cycle of the enzyme. The second heme is a low-spin heme and acts as an intermediate electron carrier.

The optical absorbance spectra have been the signature for the cytochrome *c* oxidases. In the mitochondrial cytochrome *c* oxidases both hemes are of the A type (see Figure 2), which give rise to characteristic absorbance maxima at around 600 nm. The high-spin heme of the active site is called 'heme *a*₃', the low-spin one heme *a*. The term cytochrome *aa*₃ was used for the terminal oxidase of mitochondria. In many bacteria, the low-spin heme is a heme B, giving rise to a terminal oxidase of the *ba*₃ type. One ubiquinol oxidase of *Escherichia coli* contains heme O as the high-spin heme; therefore, it is also called 'cytochrome *bo*' or '*bo*₃'.

Cytochrome *c* oxidases contain another copper center, which is called Cu_A. It is actually formed by two copper atoms, which share one electron upon reduction. During

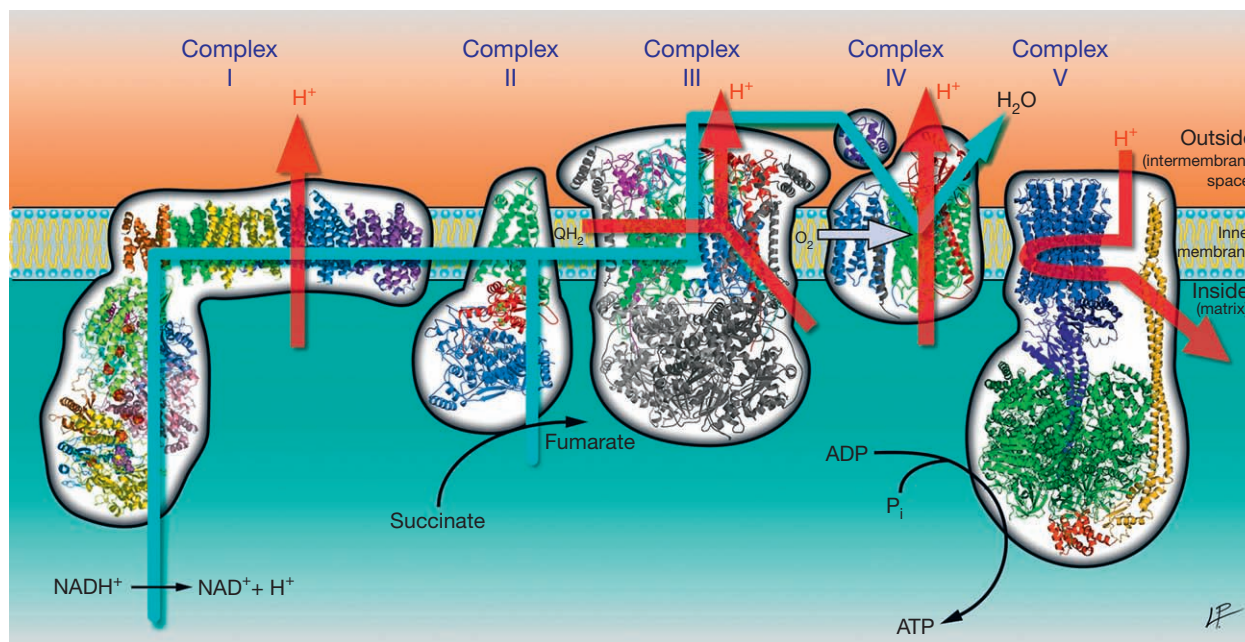


Figure 1 The mitochondrial respiratory chain. The five complexes of the respiratory chain are presented with their known structures in a functional context. The atomic structure of complex I or NADH dehydrogenase (left) is not known, but its shape has been determined by electron microscopy. It reduces ubiquinone to ubiquinol and pumps protons. Complex II (succinate dehydrogenase) is represented by a chain model. It also produces ubiquinol (QH₂). Ubiquinol is oxidized by complex III (cytochrome *bc*₁ complex). The cytochrome *bc*₁ complex releases protons at the outside and uses a quinone cycle to translocate additional protons across the membrane. The cytochrome *bc*₁ complex from yeast is shown. It reduces cytochrome *c*, which is oxidized by cytochrome *c* oxidase. The bacterial cytochrome *c* oxidase from *Paracoccus denitrificans* is shown. Cytochrome *c* oxidase pumps protons. The pumped protons flow back via the membrane part of complex V (ATP synthase). The backflow leads to a rotation of a subunit in the extra-membranous part of the ATP synthase, which is coupled to the synthesis of ATP. Proton flows are indicated by the red arrows. Figure drawn by P. Lastrico and E. Olkhova.

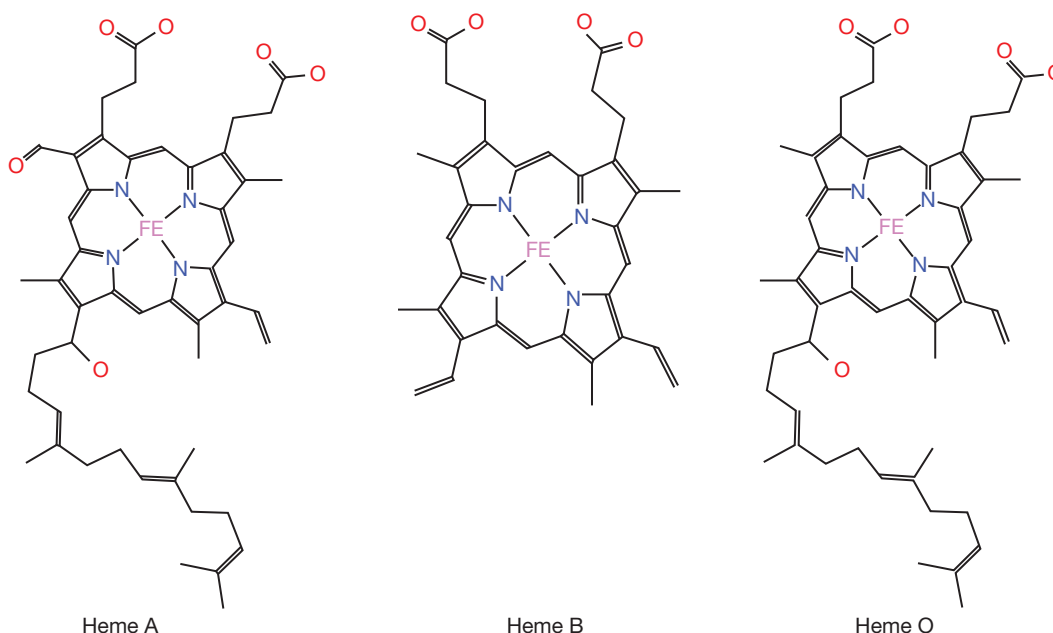


Figure 2 Chemical structures of the noncovalently bound hemes found in terminal oxidases. The structures of heme A, heme B, and heme O (from left to right) are shown.

turnover of the enzyme, Cu_A receives the electron from cytochrome *c* and transfers it to the low-spin heme. Ubiquinol oxidases do not possess this Cu_A center. Several terminal oxidases also contain a covalently bound cytochrome *c*. Depending on the nature of the other hemes these enzymes are called *caa*₃, or *cbo*, or *cbb*₃ oxidases. The terminal oxidases of the *cbb*₃ type operate at rather low-oxygen concentrations and, for example, are required for the nitrogen fixation in *Rhizobia* in order to keep the oxygen levels very low.

Cytochrome *c* oxidases also contain a magnesium or manganese atom close to the periplasmic (intermembrane) face of the enzyme. It is missing in the ubiquinol oxidases and does not appear to be involved in electron transfer.

Protein Composition

The mitochondrial cytochrome *c* oxidases possess a core consisting of three protein subunits. These subunits are encoded by the mitochondrial genome. Subunit I is the largest subunit and possesses between 500 and 600 amino acid residues. It binds heme *a* as well as the heme *a*₃-Cu_B binuclear center and is the major catalytic subunit. Subunit II provides the binding site for the Cu_A center and is anchored to the membrane by two N-terminal membrane-spanning helices. Subunit III possesses seven membrane-spanning helices but no prosthetic groups. Its function is unknown; it is not required for the catalytic activity. If its gene is deleted in bacteria, an active two-subunit enzyme is still found, albeit at a reduced level. If subunit III is removed, the enzyme becomes rather unstable under turnover conditions.

The mitochondrial cytochrome *c* oxidases contain up to eight additional smaller subunits, which are encoded by the nuclear genome. For some of the subunits, organ-specific subtypes exist; for example, there are liver and heart types of the so-called subunit VIa in the mammalian cytochrome *c* oxidase. It has been suggested that the presence of the liver or heart type determines the efficiency of proton pumping, and that only two protons per oxygen molecule are pumped by cytochrome *c* oxidase in liver, but four in heart. Subunit VIa also seems to be responsible for the dimer formation of cytochrome *c* oxidase from bovine heart mitochondria.

The bacterial cytochrome *c* oxidases of known structure contain one additional small protein subunit, called 'subunit IV'. It has no counterpart in the mitochondrial enzyme and its function is not known.

X-Ray Crystal Structures of Cytochrome *c* Oxidases and Other Terminal Oxidases

Protein Structure

Until now atomic structures, determined by X-ray crystallography, have been presented for the cytochrome *c* oxidases from the soil bacterium *Paracoccus denitrificans*, from the purple photosynthetic bacterium *Rhodobacter sphaeroides*, and from bovine heart mitochondria. These are rather typical closely related enzymes. Both *Paracoccus* and *Rhodobacter* are grouped within the α -proteobacteria, and the symbiotic ancestor of the mitochondria belonged to this group. In addition, the

structure of the aberrant *ba*₃-type cytochrome *c* oxidase (e.g., it does not possess a homolog of subunit III) from the bacterium *Thermus thermophilus*, and that of the ubiquinol oxidase (cytochrome *bo*₃) from *E. coli* have been determined.

In the cytochrome *c* oxidases the subunits I, II, and III form the core, which, viewed parallel to the membrane, has a trapezoidal shape, when seen from above it looks oval. The length of the trapezoid is about 90 Å, its height is 55 Å. It comprises 21 membrane-spanning helical segments. The globular polar domain of subunit II is attached to the narrow, periplasmic (intermembrane, respectively) side of the trapezoid (see Figures 1 and 3). Subunit I contains 12 membrane-spanning helices which are arranged in a threefold symmetric manner. In this kind of arrangement, three pores are formed. One contains the heme *a* with two histidine residues as axial iron ligands, another heme *a*₃ with one axial histidine ligand and Cu_B, and the third one is blocked by aromatic residues. Subunit II contains two N-terminal membrane-anchoring transmembrane helices. The globular domain contains a 10-stranded β -barrel similar to the copper proteins plastocyanin and azurin. The copper-binding site is in the same position, but to subunit II two copper atoms instead of one are bound. Subunit III possesses seven transmembrane helices. They form two bundles, one of five, the other of two transmembrane helices. Between them a V-shaped cleft exists, which is partially filled by lipid molecules. In the cleft at the center of the membrane, an entrance to a hydrophobic channel can be seen which leads to the active site. It has been proposed that this channel is a diffusion channel for oxygen because it leads to the active site, and because the solubility of oxygen in the hydrophobic center of the membrane is considerably higher than in the aqueous phase. Subunit III of the ubiquinol oxidase (cytochrome *bo*₃) from *E. coli* consists of five transmembrane helices only. The loss is compensated by the addition of two transmembrane helices to the 12 of subunit I.

Arrangement of the Prosthetic Groups

Cu_A receives an electron from cytochrome *c* at the periplasmic (intermembrane) surface of the membrane. Tryptophan 121 (*Paracoccus* numbering) appears to play an important role in the electron transfer. From Cu_A the electron is transferred to heme *a*, the distance of the lower Cu atom to the iron atom of heme *a* is ~20 Å. Heme *a*₃ lies in the membrane in the same height, so that electron transfer from heme *a* to heme *a*₃ occurs parallel to the membrane. The center-to-center distance of both hemes is 13–14 Å. The propionate side chains of both heme groups are oriented toward the external (periplasmic or intermembrane) side. Their charges are partially compensated by the formation of ion pairs with two arginine residues.

Cu_B is ~5 Å away from the heme *a*₃ iron, so that a very rapid electron transfer between heme *a*₃ and Cu_B has to be expected. It is a general problem in cytochrome *c* oxidase research that Cu_B is silent and there is no spectroscopic technique available which allows determining the redox state of Cu_B. Cu_B possesses three histidine ligands. One of them is covalently linked to a nearby tyrosine. The existence of a covalent histidine-tyrosine cross-link was unknown prior to the X-ray crystallographic structure determinations.

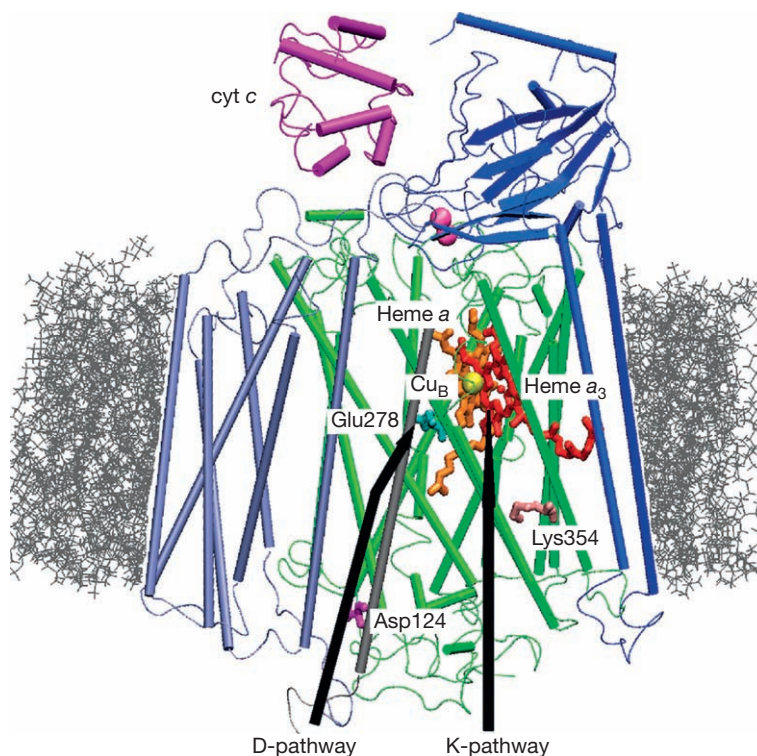


Figure 3 Cytochrome *c* oxidase in a bilayer membrane. The four subunits containing cytochrome *c* oxidase from *Paracoccus denitrificans* are shown together with its substrate cytochrome *c* (purple, top). Helices are represented as columns, β -strands as ribbons with arrows. Green: subunit I; dark blue: subunit II; iceblue: subunit III; gray: subunit IV; heme *a* is shown in orange, heme *a*₃ in red, Cu_B in yellow, and Cu_A in mauve. The lipid bilayer is represented in gray as atomic model. Figure prepared by E. Olkhova using the program VMD.

Proton Transfer Pathways

The structural knowledge combined with the results of previous site-directed mutagenesis experiments lead to the identification of two proton transfer pathways. One, called the K-pathway because of the existence of an essential lysine residue as a component of this pathway, leads straight from the cytoplasmic (matrix) surface into the binuclear site. The cross-linked tyrosine is a part of it. Mutation of the essential lysine residue to other residues prevents the initial reductions of the binuclear site during the catalytic cycle.

A second proton transfer pathway has been called the D-pathway because an aspartic residue at the protein surface is a part of it. A change of this aspartate residue to an asparagine still allows a (reduced) turnover of the enzyme, but proton pumping is abolished. The D-pathway is wider than the K-pathway, and many water molecules are involved. It leads straight up first, then changes direction at a water-filled cavity, which leads to a glutamate residue (E278 in *Paracoccus* numbering). The further pathway is unclear.

There must be a possibility for proton transfer to the active site, because the reaction cycle can be started from an artificially fully reduced enzyme in K-pathway mutants, proceeds quite normally, and water is formed in the active site. Proton transfer using a temporarily established chain of water molecules is a possibility. Alternatively, pumped protons can be transferred to the heme propionates and stored there in some parts of the catalytic cycle.

The existence of a third pathway, called H-pathway, has been suggested for the bovine heart mitochondrial cytochrome *c* oxidase. However, there is no evidence for it on the basis of mutagenesis experiments in bacterial enzymes and the proposal has met with a lot of skepticism.

The Catalytic Cycle and Proton Pumping

The catalytic cycle starts with a fully oxidized enzyme (O-state). All four prosthetic groups (Cu_A, heme *a*, heme *a*₃, and Cu_B) are oxidized. In cytochrome *c* oxidases an electron is first transferred from cytochrome *c* to Cu_A, and then to heme *a* with a time constant of 20–50 μ s. The published values for the electron transfer to heme *a*₃ from heme *a* vary greatly and most likely this electron transfer requires proton uptake via the K-pathway first. Most likely in this E-state, the single electron equilibrates over all prosthetic groups. A second electron has to be provided by cytochrome *c*. When the binuclear site is doubly reduced, an oxygen molecule can be bound to the iron atom of heme *a*₃; thus the oxygenated form, called compound A, corresponding, for example, to oxygenated myoglobin, is formed. This reacts to an intermediate state called P, which is characterized by an absorption maximum at 607 nm. P stands for peroxy because the original (incorrect) belief was that this form contains a peroxide dianion. After the input of a third electron into the enzyme a state absorbing visible light at

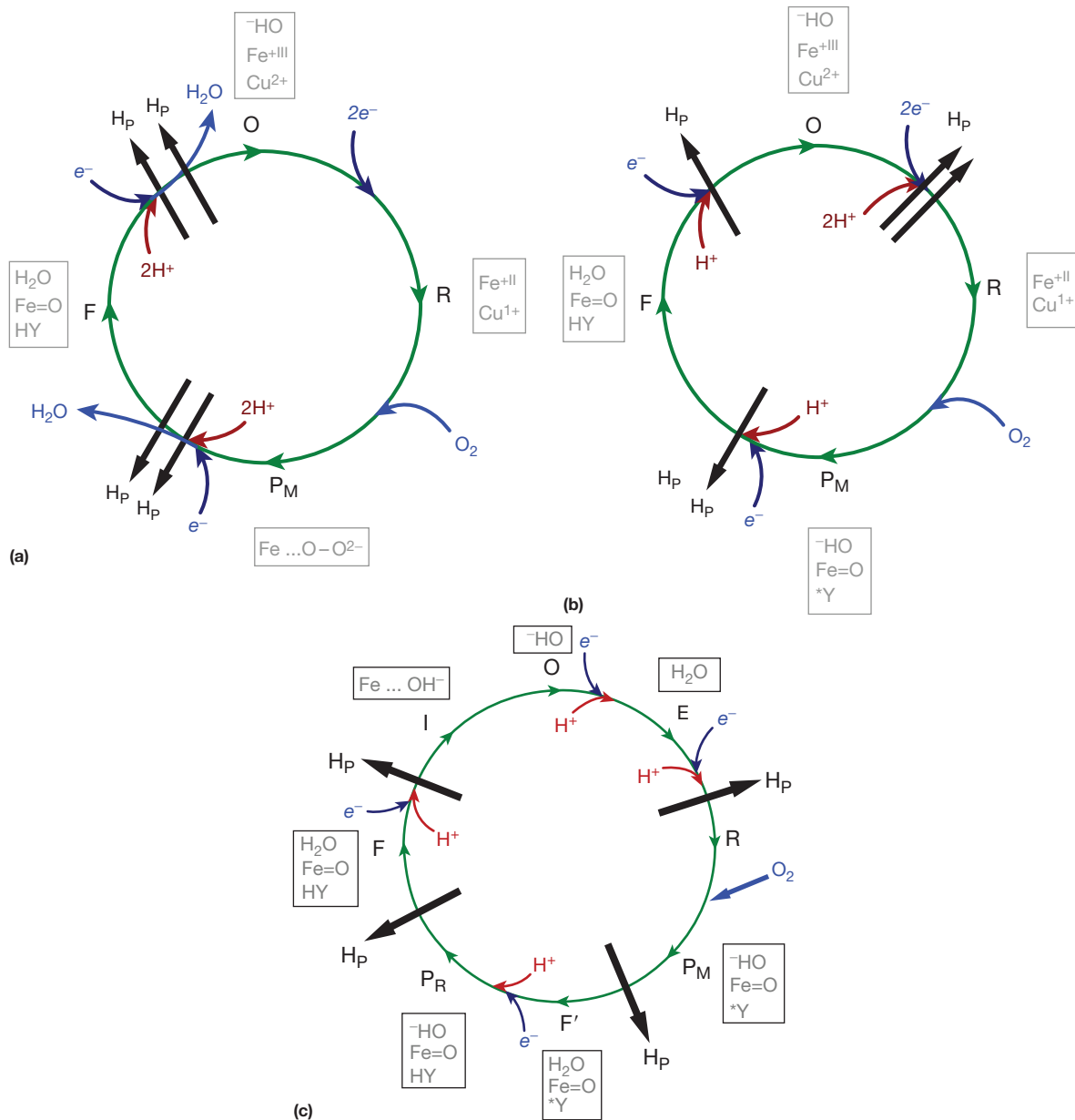


Figure 4 Catalytic cycles of cytochrome *c* oxidase. Classical, now obsolete, cycle with only four states. The P-state erroneously has a peroxide anion bound to the heme *a*₃-iron; two protons are pumped during the P- to F-, and during the F- to O-transitions under consumption of two protons each. It is clear now, that only one proton is consumed during each transition. (a) Adaptation of the classical cycle to recent findings. Two protons are taken up already during the initial two-electron reduction of the enzyme, and two protons are pumped in this phase (under turnover conditions). The P_M-state contains an oxoferryl-iron, a tyrosine radical, and a hydroxide in the binuclear center, only one proton is pumped during the P- to F-, and during the F- to O-transitions. (b) A recent cycle with two P- and two F-states. The protonations of the hydroxides in the binuclear sites of the P-states leads to proton pumping, each electron uptake is accompanied by a charge-compensating proton uptake. Tyrosine radicals are present in the two-electron reduced states, and tyrosines in the three-electron reduced states. One proton is pumped during the second reduction step, which could now be experimentally demonstrated. One electron is pumped during the F- to O-transition as a result of the protonation of the oxo-group at the heme *a*₃ iron (c).

580 nm is formed. This was believed to be the oxoferryl state and called F. After the fourth reduction step, the iron-bound oxygen atom was thought to be converted to a second water molecule under consumption of two protons and the O-state is formed again. Such a simple catalytic cycle is presented in Figure 4(a).

In recent years, much of the accepted knowledge had to be revised. First, it is clear now that there are two P- and F-states each, one each in the two-electron reduced enzyme, and one each in the three-electron reduced enzyme. It cannot be excluded that all of them are part of the catalytic cycle (Figure 4). Second, it has been convincingly demonstrated by

resonance Raman spectroscopy and chemical experiments that in the P-state the O=O double bond is already broken and that an oxoferryl state already exists. This finding causes the problem that four electrons are required to split the O=O double bond, but only three (two from the Fe^{II} to Fe^{IV} transition upon formation of the oxoferryl state, and one from Cu_B) are available from the prosthetic group. Although definite proof could not be presented yet, it is the general belief that the fourth electron (and a proton) is provided by the cross-linked tyrosine, which becomes a neutral tyrosine radical. The existence of the cross-link, which most likely is a side product formed by a tyrosine radical in one of the first turnovers, has to be considered as a hint for the occurrence of a tyrosine radical. Also, most likely, a hydroxide as a Cu_B-ligand is generated, and may cause the difference of the optical absorbance spectra between the P- and the F-states. The F-state in the three-electron reduced enzyme is most likely to contain water instead of the hydroxide, and the tyrosine radical has become a neutral tyrosine.

The question, which of the electron transfer steps are coupled to proton pumping, is currently a matter of intense debate. The already generally accepted proposal that only the P- to F-, and F- to O-transitions are linked to proton pumping with two protons pumped each, could not be maintained. There is good evidence that one proton is pumped upon the input of the second electron into the oxidized enzyme in the absence of oxygen, and it has also been proposed that under turnover conditions already the first reduction step is connected to proton pumping. For a final answer, further experiments are required. It is a general problem in cytochrome *c* oxidase research that many kinetic experiments are started from a fully reduced enzyme (all four prosthetic groups are artificially reduced). However, this state is not part of the catalytic cycle and the value of the conclusions drawn is limited. As long as it is unknown, when proton pumping occurs in the catalytic cycle, it is very difficult to find out the mechanism of proton pumping. A rather general mechanism of proton pumping, which has found many supporters, is based on the charge-compensation principle. This principle states that each electron transfer into the membrane to heme a is

charge-compensated by uptake of one proton from the opposite side of the membrane and not consumed in water formation. This proton is supposed to be electrostatically repelled and thus pumped by those protons, which are later taken up and consumed in water formation. The mechanism requires storage of the charge-compensating protons in a way that they cannot be used for water formation.

It is evident that still many experiments are required before the mechanism of proton pumping by cytochrome *c* oxidases can be considered to be known.

See also: [Bioenergetics: Cytochrome Oxidases, Bacterial; Proton Pumping in the Respiratory Chain; Respiratory Chain and Adenosine Triphosphate Synthase.](#)

Further Reading

- Ferguson-Miller S and Babcock GT (1996) Heme/copper terminal oxidases. *Chemical Reviews* 96: 2889–2907.
- Jü nemann S (1997) Cytochrome *bd* terminal oxidase. *Biochimica et Biophysica Acta* 1321: 107–127.
- Michel H (1999) Cytochrome *c* oxidase: Catalytic cycle and mechanism of proton pumping – a discussion. *Biochemistry* 38: 15129–15140.
- Michel H, Behr J, Harrenga A, and Kannt A (1998) Cytochrome *c* oxidase: Structure and spectroscopy. *Annual Review of Biophysics and Biomolecular Structure* 27: 329–356.
- Ostermeier C, Harrenga A, Ermler U, and Michel H (1997) Structure at 2.7 angstrom resolution of the *Paracoccus denitrificans* two-subunit cytochrome *c* oxidase complexed with an antibody F-V fragment. *Proceedings of the National Academy of Sciences of the United States of America* 94: 10547–10553.
- Pereira MM, Santana M, and Teixeira M (2001) A novel scenario for the evolution of haem-copper oxygen reductases. *Biochimica et Biophysica Acta* 1505: 185–208.
- Ruitenbergh M, Kannt A, Bamberg E, Fendler K, and Michel H (2002) Reduction of cytochrome *c* oxidase by a second electron leads to proton translocation. *Nature* 417: 99–102.
- Saraste M (1999) Oxidative phosphorylation at the *fin de siècle*. *Science* 283: 1488–1493.
- Saraste M and Castresana J (1994) Cytochrome-oxidase evolved by tinkering with denitrification enzymes. *FEBS Letters* 341: 1–4.
- Yoshikawa S, Shinzawa-Itoh K, Nakashima R, et al. (1998) Redox-coupled structural changes in bovine heart cytochrome *c* oxidase. *Science* 280: 1723–1729.

Retinoblastoma Protein (pRB)

N Singh, M Frolov, and N Dyson, Massachusetts General Hospital Cancer Center, Boston, MA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Nicholas Dyson and Maxim Frolov, volume 3, pp. 700–703, © 2004, Elsevier Inc.

Glossary

Apoptosis Programmed cell death.

Cell cycle The complete sequence of events needed for the production of a new daughter cell. A mitotic cell cycle has two major landmarks, S-phase and mitosis, which are separated by two gap phases, G1 (the period before DNA synthesis) and G2 (the period between the S- and M-phases).

Cell proliferation The overall increase in cell number that occurs when the rate of cell division is greater than the rate of cell death.

Compensation A change in function that allows one type of gene product to perform a function that is normally carried out by another type of gene product.

Redundancy The situation that occurs when two or more genes have a similar function and either gene is sufficient for a biological process.

The retinoblastoma protein (pRB) is the product of the retinoblastoma tumor susceptibility gene (*RB1*). The gene takes its name from the discovery that mutation of both copies of *RB1* is the critical rate-limiting event in the development of retinoblastoma, a rare form of cancer that affects very young children. Mutation of *RB1* is also found in a variety of cancers and pRB is thought to be functionally inactivated in most tumor cells. pRB is a nuclear protein, present in most cell types, which functions as a general negative regulator of cell proliferation. It is a member of a family of proteins (the pRB family) that have overlapping functions and are regulated by cyclin-dependent kinases. pRB is generally thought to function by controlling gene expression, and its best-known target is the transcription factor E2 promoter-binding factor (E2F).

The pRB Family of Proteins

pRB-related proteins are conserved in plants and animals. Mammalian cells contain three pRB-family members (pRB, p107, and p130) that share a similar domain structure and a low, but significant, degree of primary sequence homology. All three proteins contain a conserved pocket domain that provides binding sites for E2F proteins, and complexes that repress transcription. In addition, pRB, p107, and p130 contain many sites for phosphorylation by cyclin-dependent kinases (cdks). Although pRB, p107, and p130 are currently studied for their roles in the control of cell proliferation, these proteins were first discovered as cellular targets of the viral oncoprotein adenovirus E1A. The binding sites for pRB, p107, and p130 map to the regions of E1A that are needed for its oncogenic properties. Similar sequences are also present in the E7 proteins of human papilloma viruses and large T antigens of several polyoma viruses. In each case, genetic studies show that the ability of the viral proteins to bind to pRB/p107/p130 allows them to interfere with control of cell proliferation. The viruses are thought to use these interactions to create a cellular environment that is favorable for viral replication.

The RB Tumor Suppressor

RB1 was the first tumor suppressor gene to be identified and it is often described as a prototype for this class of cancer-related genes. Unlike oncogenes, whose activation promotes tumorigenesis, it is the inactivation of tumor suppressor genes that is linked to the development of cancer.

Individuals inheriting a mutant allele of *RB1* are strongly predisposed to develop retinoblastoma within the first 2 years of life. Patients, successfully treated for eye cancer, often develop osteosarcoma in later life. Inactivation of *RB1* appears to be essential for the development of both familial and sporadic retinoblastoma, and also occurs in many osteosarcomas. *RB1* mutation is common in small-cell lung cancers and has also been found in other tumors, including bladder, prostate, and breast carcinomas. pRB is broadly expressed and thought to control the proliferation of most cell types. Although *RB1* is mutated in only a subset of tumors, pRB is thought to be functionally compromised in most tumor cells by alternative mutations in the Rb pathway (see below).

Precisely why people who inherit a mutant allele of *RB1* are predisposed to certain types of cancer is still largely unclear. Part of the explanation may stem from the fact that pRB has a distinct pattern of expression: it is broadly expressed but its levels are strongly upregulated in specific cell types, particularly during cell-cycle exit. As described below, a second reason is that pRB shares overlapping functions with the related proteins p107 and p130. p130 is most highly expressed in differentiated cells, whereas p107 is highly expressed in populations of proliferating cells. The ability of p107 and p130 to compensate for the loss of pRB varies in different cellular contexts and also varies between species. This was illustrated by the unexpected finding that mice carrying a mutant allele of the *Rb* gene do not develop retinoblastoma but develop pituitary and thyroid tumors instead. The combined mutation of *Rb* and either of the related *p107* or *p130* genes allows tumors to develop in the murine retina. This suggests that p107 and p130 are able to compensate for the absence of pRB in the developing mouse retina in a way that is different from humans.

A third reason why pRB inactivation has tissue-specific consequences is that, in addition to its apparently universal role in E2F regulation (see below), pRB also interacts with several cell-type-specific factors. For example, pRB interacts with the osteoblast transcription factor CBFA1/RUNX2 to promote the expression of bone-specific genes, and this role may explain why pRB is important for suppression of osteosarcoma. Studies in mice suggest that the presence of pRB facilitates cell-cycle exit, promotes the expression of markers of differentiation, and helps to maintain the quiescent state. In some progenitors, the presence of pRB also helps cells choose between different cell fates.

The Cellular Function of pRB

Since it is the inactivation of pRB that is linked to the development of cancer, the function of pRB is most commonly viewed through the prism of the cellular changes that occur when pRB is removed. pRB-deficient cells are slightly smaller than wild-type cells and have defective cell-cycle control. In tissue culture assays, *Rb*^{-/-} cells are generally less responsive to signals that cause wild-type cells to stop proliferating, and they are more easily driven from quiescence into the cell cycle. *Rb*^{-/-} cells fail to arrest in G1, or in S-phase, in response to DNA damage. pRB-deficient cells also show a reduced sensitivity to p16^{INK4}-induced cell-cycle arrest and ras-induced senescence. Studies in primary cultures of human and mouse cells have revealed that *Rb*^{-/-} cells are genetically unstable. *Rb*^{-/-} cells have chromosome defects; they display high levels of chromosome instability and are prone to polyploidy or aneuploidy. Cells lacking pRB have an altered sensitivity to apoptosis, a change that involves both cell autonomous and nonautonomous changes and results from both the misexpression of apoptotic regulators and defects in cell-cycle arrest.

Analysis of the murine pRB-family has revealed an extensive degree of redundancy and compensation between family members. Each of the pRB family members has been inactivated in the mouse by homologous recombination. *Rb*-deficient mice die during embryogenesis, displaying elevated levels of apoptosis, inappropriate cell proliferation, and developmental defects in several tissues. Animals lacking two pRB family members are more severely affected than the single knockouts, and the inactivation of all three family members renders cells unresponsive to most types of G1 control. Triple knockout cells differentiate poorly, are readily immortalized, and easily transformed. Remarkably and unexpectedly, the early lethality of *Rb*^{-/-} animals has been shown to be largely an indirect consequence of a placental defect. When pRB is expressed specifically in the placenta, the rescue of placental function allows *Rb*-mutant embryos to survive close to birth. Although these mutant animals have defects in cell-cycle control, and in skeletal muscle development, the development of these animals is surprisingly normal. Experiments in tissue culture using cells derived from the knockout animals have shown that the sudden removal of pRB causes more severe cell-cycle control defects than those seen in cells derived from *Rb*^{-/-} animals. This difference is due, at least in part, to an upregulation of *p107* expression that occurs following the sustained absence of pRB and appears to allow p107 to compensate for the loss of pRB.

pRB and Cell-Cycle Regulation

pRB is a critical component of the cell-cycle control machinery, being a regulator of the cell cycle and, in turn, being regulated during cell-cycle progression. pRB is a long-lived protein (half-life > 16 h) and its activity is controlled by phosphorylation. When normal cells proliferate, pRB is inactivated by cdks, the kinases that drive progression through the phases of the cell division cycle. When cells cease to proliferate, the loss of cdk activity allows pRB to accumulate in its active form. Inactivation of pRB is not achieved by the phosphorylation of a single site but by the accumulation of phosphorylated residues at multiple sites that are scattered throughout the protein. A cluster of phosphorylation sites is found in the C-terminal fragments of pRB, p107, and p130 and the accumulation of negative charge in this domain is believed to drive a large conformation shift that regulates interactions with an assortment of cellular proteins. In a similar manner, phosphorylation of the pocket and amino-terminal domains is thought to promote the functional inactivation of pRB.

When quiescent cells are driven into the cell cycle, pRB is sequentially phosphorylated by multiple kinases. Cyclin D/cdk4 is believed to be the first kinase to act on pRB with cyclin E/cdk2, and perhaps cyclin A-dependent kinases completing the process. Phosphorylation of pRB causes it to migrate slowly on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels. This mobility shift is often used as an indicator of pRB activity, but it is an unreliable measure; pRB inactivation involves the cumulative effects of multiple phosphorylation events, and some modifications have no effects on the migration of the protein, whereas changes at others sites have a major effect.

Microinjection experiments on cultures of synchronized cells first showed that the ability of pRB to prevent cells from entering S-phase is restricted to a window within G1 phase of the cell cycle. Once cells progressed beyond a point in G1, ~2–4 h prior to S-phase, the introduction of pRB had no effect. The loss of a pRB-induced arrest correlates with the accumulation of cdk activity and corresponds approximately to the restriction point, a point in G1 where cells become irreversibly committed to progress through a complete cell cycle. pRB that has been phosphorylated by cdks *in vitro* is unable to arrest cells in G1. pRB is abruptly dephosphorylated at metaphase of mitosis and re-associates with chromatin.

Phosphorylation is not the only type of modification of pRB to be described. Under specific conditions, pRB can also be acetylated, methylated, ubiquitinated, or sumoylated, and these posttranslational modifications have been proposed to modulate specific aspects of pRB function. Although the details have yet to be fully elucidated, these additional modifications are thought to influence both the level of pRB phosphorylation and the types of cellular proteins that physically interact with pRB.

The Rb Pathway

pRB is a component of a critical regulatory pathway that is functionally inactivated in most tumor cells. This pathway comprises pRB, cyclin D/cdk4 (the kinase that phosphorylates

pRB in G1), and p16^{INK4} (a cdk inhibitor that antagonizes cyclin D/cdk4 activity and is induced in response to certain cellular stresses). Initial experiments showed that the overexpression of cyclinD/cdk4 phosphorylates pRB and prevents it from arresting the cell cycle. In addition, cells lacking pRB are largely insensitive to p16^{INK4}-induced cell-cycle arrest. Subsequent studies have revealed that changes in the levels of either the INK4 family (p16^{INK4A}, p15^{INK4B}, p18^{INK4C}, and p19^{INK4D}) or Cip/Kip families of cdk inhibitor proteins (p27^{kip1}, p21^{cip1}, and p57^{kip1}) can also have profound effects on the activity of pRB.

The p16^{INK4}/cyclin D/cdk4/pRB pathway is disrupted in a variety of ways in tumor cells. Retinoblastoma cells, for example, contain mutations in *RB1*. Many cervical cancers, on the other hand, express HPV E7 proteins that bind and inactivate all three of the pRB family members. Other cells contain lesions that elevate the activity of the cyclin D/cdk4 kinase (overexpression of cyclin D1, loss of the p16^{INK4} cdk inhibitor, or mutations of cdk4 that are resistant to p16^{INK4}). Although most tumor cells contain one of these changes, these lesions primarily occur in a mutually exclusive manner. This may mean that there is no further selection for changes in this group of genes once the p16^{INK4}/cyclin D/cdk4/pRB pathway is disrupted. However, it is also true that several types of tumors are associated with specific types of mutations (e.g., pRB mutations are the only types of lesions in this pathway found in retinoblastoma, while most breast cancers contain elevated levels of cyclin D1); hence, the relative importance of individual components of the pathway varies greatly between cell types.

Molecular Functions of pRB

More than 160 different cellular proteins have been reported to associate with pRB and many of these have been proposed to contribute to its tumor-suppressing properties. Most of these pRB-associated proteins are transcription factors and, although the list is diverse, in general pRB has been proposed to inhibit the activity of factors that are needed for cell proliferation, but to augment the activity of factors that promote cell differentiation.

The best-characterized property of pRB is its ability to repress E2F-dependent transcription. The E2F transcription factor allows the periodic expression of a large number of proliferation-related genes to be tightly coupled with cell-cycle position. E2F controls the expression of an extensive network of genes that includes essential components of the DNA synthesis machinery and proteins needed for cell-cycle progression, mitosis, checkpoint responses, and regulators of apoptosis. pRB blocks cell proliferation, at least in part, by blocking the expression of E2F-regulated genes.

E2F complexes are strong repressors of transcription in quiescent cells and during G1 phase of the cell cycle. As cells progress toward S-phase, these repressor complexes are disrupted and replaced by E2F activator complexes in a cdk-regulated process. All three pRB-family members associate with E2F proteins. In wild-type cells, p107 and p130 associate primarily with E2F4 and E2F5 and are components of E2F repressor complexes. In these complexes, the pRB-family members act as adapter proteins: binding to E2F with one surface

and to chromatin remodeling proteins through a second binding site (the LXCXE-binding cleft). In doing so, they allow a variety of enzymatic activities to be recruited to E2F-regulated promoters. In a similar manner, pRB also interacts with E2F4 or E2F5 and has the potential to recruit repressor complexes to DNA. In addition, pRB has the important ability to regulate E2F1, E2F2, and E2F3, the activator forms of E2F. pRB binds to an element that is buried within the transcriptional activation domain of the E2Fs, preventing them from activating transcription. This inactivation of the activator E2Fs is one of the first steps in the silencing of proliferation genes during cell-cycle exit. Biochemical studies have also shown that pRB that is recruited to DNA by its association with E2F is also able to prevent adjacent DNA-bound transcription factors from interacting with the basal transcription machinery.

Chromatin immuno-precipitation (ChIP) experiments have provided a great deal of information about the binding of E2F and pRB-family members to E2F-regulated promoters during cell proliferation and cell-cycle exit. Surprisingly, p107 and p130 are the pRB-family members that are most easily detected at most E2F targets in most cell types. While some studies have found pRB at E2F-regulated promoters in cultures of proliferating cells, others show that pRB is preferentially recruited to E2F targets in response to specific signals, such as p16-induced cell-cycle arrest, DNA damage, or following ras-induced senescence. Under these circumstances, pRB is thought to recruit histone deacetylase and/or histone methylase activities to DNA. These complexes repress transcription through modification of histone tails, changing the local chromatin structure to a form that is nonconductive to gene expression.

ChIP studies suggest that there is not a single group of promoters that is targeted by pRB, p107, and p130, nor is there a single mechanism of repression. Many different promoters can be repressed by pRB, p107, or p130, and the types of repressor complexes that are operative at these promoters can vary depending in part on whether the cells are proliferating, quiescent, undergoing senescence, differentiating, or responding to DNA damage. Currently, it seems unlikely that there is a single form of pRB-mediated transcriptional regulation. Instead, pRB-family members seem to be versatile proteins that are capable of recruiting several types of repressor activities to chromatin. A key goal for future studies will be to identify the underlying rules that can explain which types of repressor complexes are recruited to individual promoters and when the presence of the different complexes becomes functionally significant.

Although E2F is the best-known target, pRB's physical interactions with other cell-cycle regulators also help to promote normal cell-cycle control. pRB has been detected at origins of replication and reported to relocalize near to replication origins following DNA damage. In addition to repressing E2F-dependent transcription, pRB also promotes cell-cycle exit by increasing the levels of p27^{kip1}. pRB stabilizes p27^{kip1} by its binding to S-phase kinase-associated protein 2 (Skp2/p45) and preventing the SCF^{Skp2} ubiquitin ligase from targeting p27. This regulation is enhanced by pRB's ability to associate with the ubiquitin ligase anaphase-promoting complex and its activator cadherin-1/E-cadherin (APC-Cdh1) and to promote APC-Cdh1-mediated turnover of Skp2.

Genetic interactions between *Rb* and *E2F* genes have been demonstrated in multiple experimental systems and in species

as diverse as mice and fruit flies. Nevertheless, it is important to remember that pRB has many binding partners and precisely how much of pRB's functions in development and tumor suppression are mediated through its interaction with E2F is uncertain. The developmental defects of *Rb*^{-/-} mice can be partly suppressed by mutations in E2F genes (*E2f1* and *E2f3*), but rescue can also be seen with mutant alleles of the differentiation regulator, *Id2*, and the apoptotic regulators *Apaf1* and procaspase3 (*casp3*). In each case, individual mutant alleles give only a partial rescue of the *Rb*^{-/-} phenotype, suggesting that the *Rb*^{-/-} mutant phenotype is too complex to be corrected by manipulating a single pathway.

One of the ultimate goals of pRB research is to find ways to selectively kill *Rb*^{-/-} tumor cells. Excitingly, several mutations have been shown to reduce tumor phenotypes of *Rb*^{+/-} mice. These include mutant alleles of *E2f1*, *E2f3*, *E2f4*, *K-ras*, and *Skp2*. One of the most striking of these interactions has been observed with *Skp2*. *Skp2* is the product of an E2F-regulated gene and is aberrantly overexpressed in many human cancers. Studies in mice show that *Skp2* is not generally required for tumorigenesis or for animal development; nevertheless, the inactivation of *Skp2* suppresses tumorigenesis in the pituitary by inducing apoptosis in *Rb*^{-/-} cells. Additional studies in other model organisms show that there may be several additional pathways that can be targeted in a way that is synthetically lethal with the inactivation of pRB. However, studies of pRB function suffer from the general caveat that the precise role of pRB is heavily dependent on the context in which it is studied, and many of the genetic interactions reported for *Rb* are cell-type specific.

See also: **Bioenergetics:** Cell Death by Apoptosis and Necrosis;
Cell Architecture and Function: Cell-Cycle Controls in G₁ and G₀;

Molecular Biology: Chromatin: Physical Organization; Regulation of Chromatin Dynamics.

Further Reading

- Dyson N (1998) The regulation of E2F by pRB-family proteins. *Genes and Development* 12: 2245–2262.
- Goodrich DW, Wang NP, Qian YW, Lee EY, and Lee WH (1991) The retinoblastoma gene product regulates progression through the G₁ phase of the cell cycle. *Cell* 67: 293–302.
- Goodrich DW, Wang NP, Qian YW, Lee EY, and Lee WH (2006) The retinoblastoma tumor-suppressor gene, the exception that proves the rule. *Oncogene* 25(38): 5233–5243.
- Harbour JW and Dean DC (2000) The Rb/E2F pathway: Expanding roles and emerging paradigms. *Genes and Development* 14: 2393–2409.
- Robanus-Maandag E, Dekker M, van der Valk M, et al. (1998) p107 is a suppressor of retinoblastoma development in pRb-deficient mice. *Genes and Development* 12: 1599–1609.
- Sage J, Miller AL, Perez-Mancera PA, Wysocki JM, and Jacks T (2003) Acute mutation of retinoblastoma gene function is sufficient for cell cycle re-entry. *Nature* 424: 223–228.
- Sage J, Mulligan GJ, Attardi LD, et al. (2000) Targeted disruption of the three Rb-related genes leads to loss of G₁ control and immortalization. *Genes and Development* 14: 3037–3050.
- Sherr CJ (1996) Cancer cell cycles. *Science* 274: 1672–1677.
- Takahashi C and Ewen ME (2006) Genetic interaction between Rb and N-ras: Differentiation control and metastasis. *Cancer Research* 66(19): 9345–9348.
- Thomas SA, Carty DM, Piscopo JS, et al. (2001) The retinoblastoma protein acts as a transcriptional coactivator required for osteogenic differentiation. *Molecular Cell* 8: 303–316.
- van den Heuvel S and Dyson NJ (2008) Conserved functions of the pRB and E2F families. *Nature Reviews Molecular Cell Biology* 9: 713–724.
- Wang H, Bauzon F, Ji P, et al. (2010) Skp2 is required for survival of aberrantly proliferating Rb1-deficient cells and for tumorigenesis in Rb1^{+/-} mice. *Nature Genetics* 42(1): 83–88.
- Weinberg RA (1995) The retinoblastoma protein and cell cycle control. *Cell* 81: 323–330.
- Wu L, de Bruin A, Saavedra HI, et al. (2003) Extra-embryonic function of Rb is essential for embryonic development and viability. *Nature* 421: 942–947.

Retinoic Acid Receptors

M Petkovich, Queen's University, Kingston, ON, Canada

© 2013 Elsevier Inc. All rights reserved.

Glossary

Retinoic acid The derivative of the fat-soluble vitamin A that is active in the regulation of gene expression by activating nuclear retinoic acid receptors.

Transcription The process by which RNA polymerase II is recruited to the promoter of a gene and

enzymatically copies the base sequence of a gene into an RNA transcript.

Transcription factor A protein or collection of proteins that are involved in altering the efficiency of RNA polymerase II activity in transcribing a gene.

Retinoic acid receptors (RARs) are nuclear transcription factors that, once activated by retinoic acid (RA), regulate the expression of RA target genes, leading to changes in cell differentiation, cell proliferation, and apoptosis. The discovery of the first RAR in 1987 defined for the first time the mechanism of action of vitamin A outside of the visual cycle. Three RAR subtypes have been identified in the vertebrate genome and have been named RAR α , RAR β , and RAR γ .

RA and RA Signaling

RA is a signaling molecule derived from vitamin A (retinol) that is essential for normal embryonic development and health in the adult. RA is produced in a two-step enzymatic process involving the alcohol dehydrogenase-mediated conversion of retinol to retinaldehyde and, in a subsequent retinaldehyde dehydrogenase-catalyzed step, to RA. RA signaling is mediated by RARs.

RARs belong to the superfamily of nuclear hormone receptors and bear conserved structural similarities to other members of this family, including estrogen and androgen receptors, glucocorticoid receptor, thyroid hormone receptor, and vitamin D receptor. In general, members of this family can be distinguished by the linear arrangement of six more or less conserved functional domains designated A through F. Key functional domains of RARs include the RA-binding domain, the DNA-binding domain, and two separate transcription activation domains (see Figure 1). Although the DNA- and RA-binding domains of the RARs are similar between subtypes, the transcription activation domains vary considerably, possibly reflecting target gene-specific interactions. Several regions within RAR domains mediate specific interactions with retinoid X receptor subtypes (RXR α , RXR β , or RXR γ), forming an obligatory heterodimeric partnership required for activation of RA target genes. RAR/RXR heterodimers bind to specific DNA elements called retinoic acid response elements (RAREs) usually found in the promoters of such genes. The unique expression patterns of each RAR subtype in the embryo and the adult also support the specific functional role that each of the RARs fulfills. Murine knockout studies have revealed that although there is functional redundancy between RARs, each receptor is essential for normal physiology.

RAR Structure and Function

RARs share structural similarities with other members of the nuclear receptor family. The key functional domains of RARs include the RA-binding domain, the DNA-binding domain, transcriptional activation domains, and amino acid sequence motifs that are necessary for interactions between RARs and other transcription factors, including RXRs. RARs form heterodimeric complexes with RXRs in order to regulate gene expression.

DNA-Binding Domain

The hallmark of all nuclear receptors is the DNA-binding domain (domain C), which comprises a highly conserved, 66–68-amino-acid stretch encoding two zinc-binding fingers labeled C1 and C2. The DNA-binding domain of the RARs is 66 amino acids in length, and its structure has been determined by X-ray crystallographic studies. The two zinc fingers fold into a globular domain such that residues of the first zinc-binding finger (C1) make specific contacts with DNA through interactions between three conserved amino acid residues (P-box) with specific base residues in the major groove of the DNA double helix. The second zinc finger (C2) is more basic in nature, making contacts with the negatively charged phosphate backbone of DNA. This helps to stabilize the interaction between RARs and the DNA motifs to which they bind, RAREs, in complex with RXRs. This RAR/RXR association with the RARE is also dependent on the conformation of the specific RARE. Residues in C2 are critical for heterodimer formation.

RA-Binding Domain

The RA-binding domain (domain E) fulfills several important functions, including, of course, RA binding, along with transcription activation, and nuclear localization. The RA-binding domain binds both *all-trans* and *9-cis* stereoisomers of RA with high affinity. By contrast, the ligand-binding domain of RXRs binds only the *9-cis* RA isomer. These ligand-binding domains possess a series of 12 α -helical structures (H1–H12) that fold together to form the ligand-binding pocket. H12 contains the transcription activation domain AF-2. From X-ray crystallographic studies, it has been proposed that conformational changes which



Figure 1 Domain structure of RARs. All RARs and RXRs have domain structures similar to other members of the nuclear receptor family. Domains A–F are shown. DNA-binding domain (C) and ligand-binding domain (E) are highly conserved. Transcription activation functions are found in E (AF-2) and A (AF-1) domains.

occur as a result of the binding of RA alter the position of H12, and thus the ability of the AF-2 to interact with transcription factors that repress or activate gene transcription. In the mouse-trap model of ligand activation, H12 is extended into the solvent in the absence of RA but snaps shut once ligand enters the RA-binding pocket. These dynamic conformational reconfigurations result in differential interactions with cofactors that functionally link the receptor complex with the transcription initiation complex.

Other Functional Regions of RARs

The N-terminus of the RARs, comprising the variable domains A and B, contains another region important for transcriptional activation (AF-1). Interestingly, differential splicing/alternate promoter usage can give rise to multiple A domains for each subtype. This multiplicity of RAR isoforms and their potential to form partnerships with RXR subtypes create, at least conceptually, a remarkable number of combinatorial possibilities that may each have subtly different yet important contributions to RA target gene regulation. The B region also contains proline-rich motifs and phosphorylation sites for interactions with other proteins. Domain D, or the hinge region, has been described as a tether that connects DNA and ligand-binding functions. This domain also appears to possess a nuclear localization motif required to sequester receptors in the nucleus. Unlike steroid hormone receptors, which reside in the cell cytoplasm bound to heat-shock proteins that are released to the nucleus upon ligand binding, RARs and RXRs are complexed with DNA in their unliganded state. The role of the F-domain located at the extreme C-terminus of RAR is not known, but it contains phosphorylation sites that might have functional significance.

RA Response Elements

RAREs are DNA motifs to which RAR/RXR heterodimers can bind with high affinity. In most cases, these motifs comprise a direct repeat of the half-site consensus, PuGGTCA, separated by five base pairs; however, other configurations are also possible. These elements can be found in variations of number and orientation from within a few base pairs of transcription initiation sites to several thousand base pairs (usually upstream). These elements concentrate RAR/RXR heterodimers, which are normally in very low abundance, near transcription complexes.

RA Transcriptional Activation

RARs act as biosensors to detect RA in the local environment. The binding affinity of these receptors for RA is on the order of

10–30 nM. Once RAR-expressing cells are exposed to RA, changes in the pattern of gene transcription will occur. The role of transcription activators is to increase the efficiency of RNA polymerase II binding to the promoter and engage the successful transcription of a gene – and in the case of RARs, to do so only in the presence of RA. In unliganded form, the RAR/RXR heterodimer, when bound to an RARE on a target gene, can act to repress transcription (Figure 2). This complex recruits a specific type of transcription factor called a corepressor. Nuclear receptor corepressors found to interact with unliganded RAR include N-CoR/SMRT. The presence of RA causes allosteric modulation of the ligand-binding domain, shifting the cofactor binding preference from corepressor to coactivator. Thus, release of N-CoR/SMRT by RA results in the complex having increased affinity for coactivators such as the p140 and p160 subfamilies (e.g., steroid receptor coactivator (SRC)). Many coregulators that interact with RAR/RXRs and that also have other chromatin-remodeling functions, such as histone-deacetylase activity, kinase activity, or methylase activity, have been isolated. In addition to chromatin-remodeling functions, some of these cofactors correspond to factors critical for establishing a stable transcription initiation complex. The net effect of these interactions is increased or decreased transcription of target genes, depending on the promoter sequence, the local concentrations of factors binding to the promoter, and whether other signaling pathways are regulating the activities of any of these factors. In this way, the RA signal is processed in a manner that integrates the activities of many pathways of a given cell type.

RAR Function

Each of the murine RAR and RXR subtypes has been knocked out by homologous recombination in order to determine their functions during embryonic development and in the adult. Although there are multiple receptor isoforms, there are distinct roles for each defined by spatiotemporal distribution and promoter-specific activities. Throughout the murine life cycle, RAR α is ubiquitously expressed and may play a role in some general housekeeping functions. By contrast, RAR β transcript distribution indicates that this receptor may be involved in ontogenesis of the central nervous system during development and differentiation of epithelia in the adult. Similarly, the expression patterns of RAR γ are spatiotemporally regulated, suggesting that this receptor is involved in embryonic tail bud development, craniofacial morphogenesis, chondrogenesis, and maintenance of squamous epithelia. Among RXRs, RXR α is ubiquitously expressed, whereas RXR β and RXR γ are more restricted during embryogenesis.

Gene knockout studies confirmed distinct roles for each RAR and RXR subtype and also revealed a high level of functional redundancy. Whereas the RAR α - and RAR β -null mutants display some of the postnatal defects observed in vitamin A-deficient (VAD) mice (poor viability, growth deficiency, and male sterility), most of the RAR- or RXR-null single mutants exhibit only subtle or, in some cases, undetectable developmental abnormalities. By contrast, RAR double mutants exhibit reduced viability and recapitulate most symptoms of VAD.

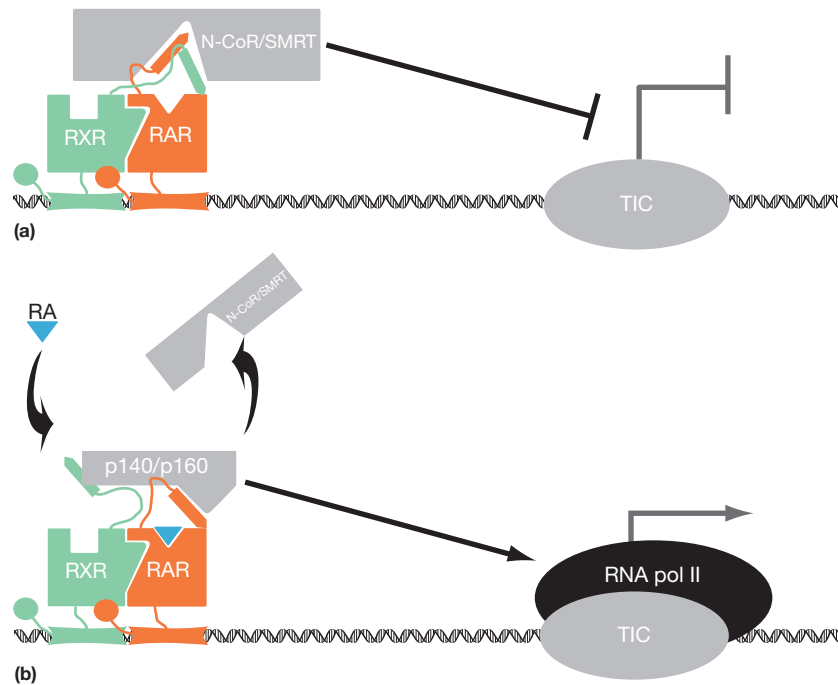


Figure 2 The RAR/RXR complex regulates transcription. (a) Transcriptional repression occurs when unliganded RAR/RXR heterodimer binds to an RARE in the promoter region upstream of the transcription start site of an RA-responsive gene. Unliganded receptors bind to corepressor factors such as N-CoR/SMRT that act to limit the function of the transcription initiation complex (TIC) through reorganization of chromatin structure. (b) Binding of RA ligand to the RXR/RAR complex induces a conformational change in the ligand-binding domain of RAR, causing the release of N-CoR/SMRT and allowing the recruitment of transcriptional coactivators such as the p160 complex. This coactivator complex promotes the binding of transcription accessory factors essential for the efficient engagement of RNA pol II activity.

$\text{RXR}\beta^{-/-}\text{RXR}\gamma^{-/-}\text{RXR}\alpha^{+/-}$ triple-mutant mice exhibit marked growth deficiency but, surprisingly, display no obvious congenital or postnatal abnormalities. However, mutation of both $\text{RXR}\alpha$ alleles is lethal. Therefore, it is likely that $\text{RXR}\alpha$ carries out most of the vital developmental and postnatal functions of the RXRs, or is at least sufficient to do so.

See also: **Metabolism Vitamins and Hormones:** Vitamin A (Retinoids); **Molecular Biology:** Gene Expression in Eukaryotes: RNA Polymerase II Structure; RNA Polymerase II and Its General Transcription Factors; RNA Polymerase II Elongation Control in Eukaryotes; **Protein/Enzyme Structure Function and Degradation:** Zinc Fingers.

Further Reading

- Chambon P (1996) A decade of molecular biology of retinoic acid receptors. *FASEB Journal* 10: 940–954.
- Clagett-Dame M and DeLuca HF (2002) The role of vitamin A in mammalian reproduction and embryonic development. *Annual Review of Nutrition* 22: 347–381.
- Dolle P (2009) Developmental expression of retinoic acid receptors (RARs). *Nuclear Receptor Signaling* 7: e006.
- Mark M, Ghyselinck NB, and Chambon P (2009) Function of retinoic acid receptors during embryonic development. *Nuclear Receptor Signaling* 7: e002.
- Rochette-Egly C and Germain P (2009) Dynamic and combinatorial control of gene expression by nuclear retinoic acid receptors (RARs). *Nuclear Receptor Signaling* 7: e005.
- Rosenfeld MG and Glass CK (2001) Coregulator codes of transcriptional regulation by nuclear receptors. *Journal of Biological Chemistry* 276: 36865–36868.
- Ross SA, McCaffery PJ, Drager UC, and De Luca LM (2000) Retinoids in embryonal development. *Physiological Reviews* 80: 1021–1054.

Rho GTPases and Actin Cytoskeleton Dynamics

P Villalonga, Universitat de les Illes Balears, Palma, Spain

A J Ridley, King's College, London, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Cell transformation A complex alteration of cellular behavior that is characteristic of cancer cells. It involves a reduction in the dependence on soluble mitogens and anchorage to the extracellular matrix in order for the cells to proliferate, together with loss of the ability to respond to growth-inhibitory signals both from the extracellular environment or arising from intracellular checkpoint controls.

Cytoskeleton Filaments consisting of protein polymers that are responsible for giving the cell its mechanical strength, controlling its shape, and driving cell movement. Three types of filaments are found in animal cells: actin filaments, microtubules, and intermediate filaments.

Leading edge The front of an actively migrating cell, which generally displays a fan-shaped meshwork of actin filaments that push forward the plasma membrane as they polymerize and grow.

Small guanosine triphosphatases (GTPases) Monomeric guanosine triphosphate (GTP)-binding proteins, as opposed to the transmembrane receptor-coupled family of heterotrimeric G proteins, which bind guanine nucleotides and have different conformations when bound to GTP or guanosine diphosphate (GDP). This enables them to act as molecular switches and transiently activate several effector proteins.

The dynamic regulation of the actin cytoskeleton plays a crucial role in the control of cell shape and movement, and, therefore, is essential for the development and function of multicellular organisms. Accordingly, actin cytoskeleton regulators modulate a broad array of cellular processes such as cell migration, tissue morphogenesis, inflammation, vesicular trafficking, secretion, and cancer cell invasion and metastasis. Members of the Rho family of small guanosine triphosphatases (GTPases) are major regulators of the actin cytoskeleton in all eukaryotic cells. Similar to other small GTPases, these proteins are molecular switches that are exquisitely regulated by several positive and negative regulators that can switch them either on or off. Once activated, they bind to and activate a number of downstream effectors, most of them directly implicated in the regulation of actin-related biochemical events (e.g., actin polymerization). This enables Rho GTPases to act as key players in processes that require the dynamic regulation of the actin cytoskeleton.

Rho GTPase Family: Overview and Regulation

Rho GTPases are found in all eukaryotic organisms and constitute a separate highly conserved family within the larger Ras superfamily of small GTPases. In mammals, there are 20 Rho GTPases divided into eight subfamilies (Figure 1). The prototypical Rho family members are RhoA, Rac1, and Cdc42, which have been the most extensively characterized. Some Rho family members show a tissue-specific pattern of expression, although most of them are expressed in the majority of mammalian cells. They usually localize to cellular membranes as a consequence of lipid modifications at their carboxy termini, such as prenylation. Similar to other GTPases, Rho proteins bind to guanine nucleotides and most cycle between an

active guanosine triphosphate (GTP)-bound state and an inactive guanosine diphosphate (GDP)-bound state. Exceptions to this are the members of the Rnd subfamily (Rnd1, Rnd2, and Rnd3/RhoE), RhoBTB subfamily (RhoBTB1 and RhoBTB2), and RhoH (also known as TTF), which are constitutively GTP-bound. Three major types of regulators modulate the activation levels of Rho GTPases: guanine-nucleotide-exchange factors (GEFs), GTPase-activating proteins (GAPs), and guanine-nucleotide-dissociation inhibitors (GDIs), as illustrated in Figure 2. In addition, some Rho GTPases are regulated by expression levels, phosphorylation, which usually affects their subcellular localization, or by ubiquitination and subsequent degradation.

Guanine-Nucleotide-Exchange Factors

GEFs promote the release and exchange of bound nucleotides on the GTPase, effectively leading to an increase in the number of GTP-bound molecules due to the higher intracellular concentration of GTP. GEFs are, therefore, positive regulators of Rho GTPases that promote their activation. In mammalian cells, 80 GEFs have been described, which are divided into two groups. The first group, the Dbl-related GEFs, share a highly conserved exchange factor domain termed Dbl-homology (DH), named after the first GEF identified, Dbl. They have an adjacent pleckstrin homology (PH) domain, which helps to position the DH domain on membranes in the vicinity of the GTPases. In some GEFs, the PH domain also contributes to Rho GTPase binding. Most Dbl-related GEFs have other protein-protein interaction domains such as SH2 and SH3 domains, allowing them to be part of multi-protein complexes. The second group of GEFs are the DOCK

family proteins, which contain a Dock homology region 2 (DHR2) exchange factor domain. They also have a DHR1 domain that can mediate membrane interaction, similar to the PH domains of Dbf-related GEFs. GEFs can be regulated by binding to other proteins or by phosphorylation. Rho GEFs exhibit different degrees of specificity *in vitro* toward Rho

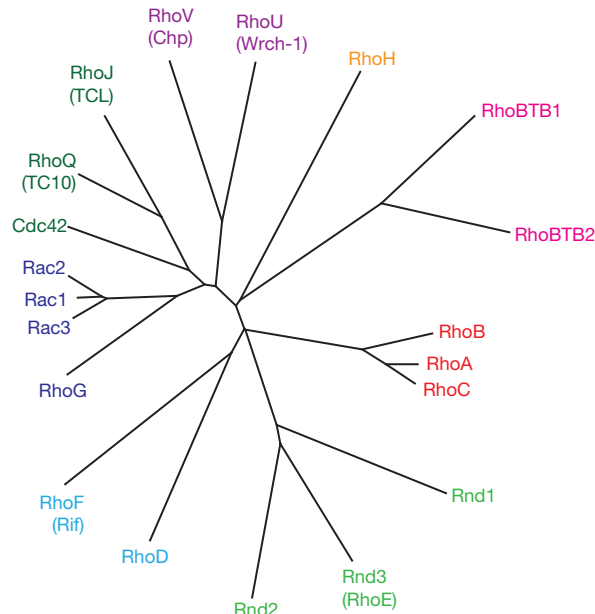


Figure 1 Phylogenetic tree of the Rho GTPase family. Rho GTPase family members can be grouped into eight subfamilies (shown in different colors) according to their sequence similarity.

GTPases, but it still has not been clearly established which GEFs regulate each of the Rho GTPase members *in vivo*.

GTPase-Activating Proteins

GAPs act by binding to the GTPase and enhancing hydrolysis of GTP, an otherwise slow and inefficient process, and by promoting the conversion of GTP to GDP they are negative regulators of Rho GTPases. Rho GAPs share in common a 140-amino-acid domain, termed the RhoGAP domain, which confers GAP activity. There are over 70 RhoGAPs in mammals. They are usually large proteins containing other domains involved in signaling and lipid and protein binding. These domains are thought to influence GAP function and localization in cells. Several RhoGAPs are known to be regulated by phosphorylation or lipid binding.

Guanine-Nucleotide-Dissociation Inhibitors

GDIs bind to some Rho GTPases and mask the prenyl groups at their C-termini, thus preventing their interaction with membranes, and with GEFs, GAPs, and downstream targets. For these reasons, they are considered to be negative regulators of Rho GTPases that sequester them in an inactive form in the cytoplasm. Signals that activate Rho GTPases will disrupt the GTPase/GDI complex and localize the GTPase to membranes. GDIs can be phosphorylated by a number of protein kinases, leading to their dissociation from GTPases and, consequently, to GTPase activation. There are three RhoGDIs in mammalian cells (α , β , and γ) that show differences in their pattern of expression, activity, and GTPase-binding specificity.

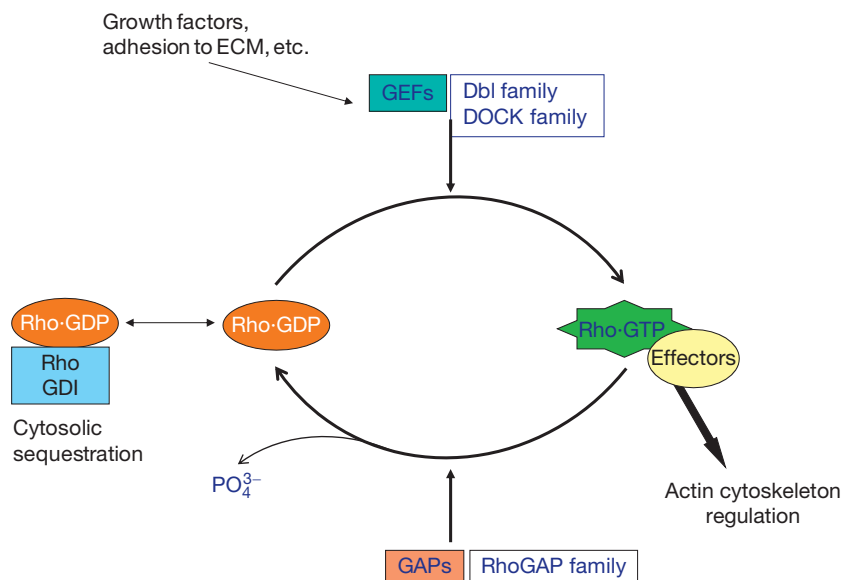


Figure 2 Rho GTPase regulation. Most Rho GTPases associate with membranes, and cycle between a GDP-bound inactive state and a GTP-bound active state. Some Rho GTPases can bind to RhoGDIs, which sequester them in the cytoplasm. Extracellular signals induce dissociation from RhoGDIs and the exchange of GDP for GTP, which is regulated by GEFs, thereby activating Rho GTPases and localizing them on membranes. In their GTP-bound form, Rho GTPases interact with their downstream effectors and regulate a variety of cellular responses, including changes to the actin cytoskeleton. Inactivation is achieved through GTP hydrolysis, which is promoted by GAPs binding to Rho GTPases.

Rho GTPase Family Activity: Downstream Targets

Rho GTPases exert their function by binding to and modulating the activity of an array of downstream effectors. Here, the manner in which Rho GTPases activate some of their most relevant targets in terms of actin cytoskeleton control will be described (Table 1).

Rho Effectors

The archetypical responses elicited by Rho (RhoA, RhoB, and RhoC) activation in cultured cells are the assembly of bundles of filamentous actin and myosin, termed stress fibers, and the formation of multimolecular complexes at the end of those fibers at sites of integrin-mediated cell adhesion, termed focal adhesions. Both effects can be largely attributed to one Rho effector, the serine/threonine kinase, Rho-kinase (ROCK). Rho-GTP binding to ROCK unfolds and relieves its kinase domain from an otherwise closed, autoinhibited, conformation. The two isoforms of ROCK phosphorylate a number of proteins involved in regulating the actin cytoskeleton, including the myosin-binding subunit of myosin light-chain phosphatase, thereby inhibiting its phosphatase activity and enhancing myosin light-chain phosphorylation and actomyosin-based contractility. Other Rho effectors contribute to Rho-mediated actin cytoskeleton regulation, such as mDia proteins (mammalian Diaphanous-related formins). mDia proteins bind as dimers to barbed (or plus) ends of

actin filaments and act as actin nucleators promoting the growth of parallel, unbranched, filaments.

Cdc42 Effectors

Activation of Cdc42 in most cells promotes the formation of filopodia, spike-like membrane protrusions induced by actin polymerization in parallel bundles. As with Rac-induced lamellipodia, these structures are associated with the leading edge of migrating cells. Cdc42 induces the formation of filopodia through the activation of several targets, including formins mDia formins, which are actin nucleators, and IRSp53, an adaptor protein with an inverse-Bin-amphipysin-Rvs (BAR) domain that bends membranes. Cdc42 also acts through its target N-Wiskott-Aldrich syndrome protein (N-WASP) and the hematopoietic-specific WASP to regulate vesicle trafficking and the formation of podosomes and invadopodia, actin-based structures at the plasma membrane that mediate localized degradation of extracellular matrix proteins. WASP proteins induce *de novo* actin polymerization through their binding to the Arp2/3 (actin-related protein) complex, which acts as an actin nucleator. Binding of active Cdc42 to WASP changes its conformation, relieving it from an autoinhibited state and exposing a previously masked domain, the VCA domain (verprolin-homology, cofilin-homology, and acidic region). Once exposed, this domain binds to G-actin through the V-domain and to the Arp2/3 complex through the CA-domain, strongly activating Arp2/3 complex-induced actin nucleation and polymerization. WASP also binds to the actin-binding protein profilin, which further enhances WASP-mediated actin filament formation through its adenosine triphosphate/adenosine diphosphate (ATP/ADP) exchange activity toward actin. Activation of WASP family proteins by Rho GTPases, such as WASP activation by Cdc42, is a typical example of downstream effector activation by binding to active GTPases (Figure 3).

Rac Effectors

In most mammalian cells, activation of Rac induces lamellipodia and membrane ruffles, both being broad sheet-like plasma membrane protrusions that extend over the substratum or upward, respectively. These protrusions are driven by a branching network of actively polymerizing actin filaments, and are usually found at the leading edge of migrating cells. The Rac effector pathway that leads most directly to actin polymerization involves WAVE proteins (wasp/verprolin homologous protein). WAVEs are members of the WASP family, and are part of a multiprotein complex mediating actin polymerization through activation of the Arp2/3 complex. Rac does not bind directly WAVEs but to CYFP4/Sra-1, which is part of the WAVE complex. Although the mechanism whereby Rac1 regulates the WAVE complex is different to Cdc42 stimulation of WASP, the consequence in both cases is to stimulate actin polymerization (see Figure 3). Other Rac effectors also contribute to actin polymerization and to extension of lamellipodia and membrane ruffles. For example, p21-activated kinases (PAKs) are a family of kinases that can be activated by Rac and Cdc42 and contribute to actin reorganization, for instance, through phosphorylation of LIM kinase, which phosphorylates and inhibits the actin-filament severing and depolymerizing protein cofilin, thereby increasing the level of polymerized actin.

Table 1 Rho GTPase, downstream effectors and their functions

<i>Rho GTPase</i>	<i>Downstream effectors</i>	<i>Function in cell biology</i>
RhoA, B, C	ROCK1,2	Actomyosin contractility: stress fiber and focal adhesion formation, tail retraction in migrating cells
	mDia	Actin nucleator promoting actin filament formation
Rac1, 2, 3	CYFP1/Sra-1	Part of WAVE complex, which stimulates actin polymerization through the Arp2/3 complex, inducing lamellipodium formation
	PAK1, 2, 3	Actin filament stability through LIMK-induced cofilin inhibition; focal adhesion turnover
Cdc42	WASP/N-WASP	Actin polymerization through Arp2/3 complex. Vesicle trafficking; invadopodia and podosomes
	IRSp53	Adaptor protein involved in filopodium formation
	mDia	Actin nucleator involved in filopodium formation
	Par6	Part of PAR complex involved in cell polarity
Rnd1, Rnd3	p190RhoGAP	Antagonize RhoA effects: disassembly of actin stress fibers and focal adhesions.
RhoG	ELMO/DOCK180	Lamellipodium extension, cell migration
RhoH	Unknown	Inhibits Rac1 signaling and migration in hematopoietic cells
RhoQ	Unknown	Involved in vesicle trafficking

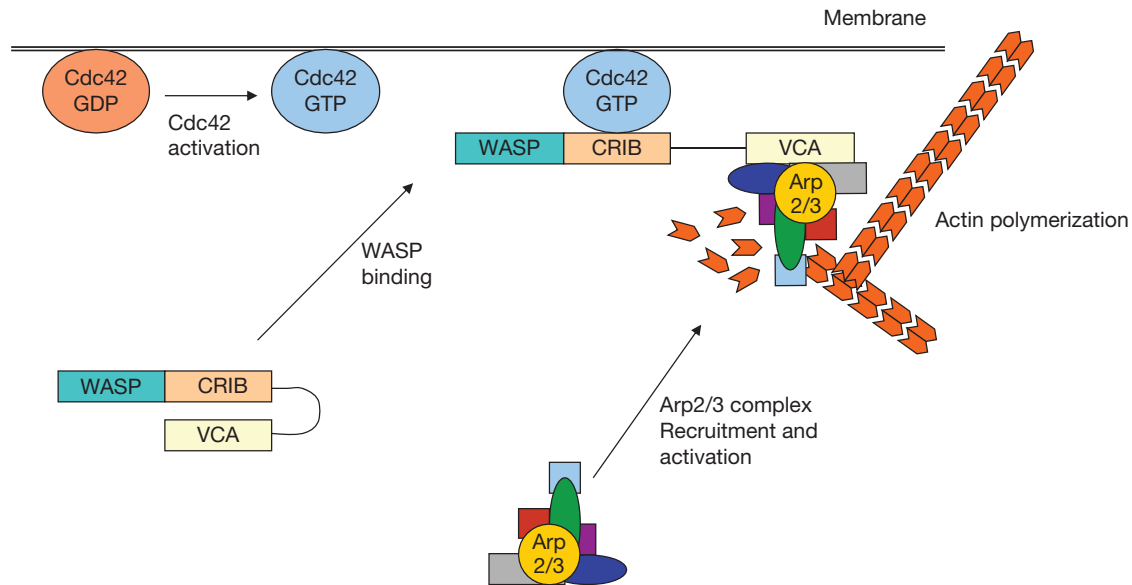


Figure 3 Model of effector activation by Rho GTPases: Cdc42-induced WASP activation promotes actin polymerization. When Cdc42 is inactive, WASP is in an auto-inhibited conformation. Upon Cdc42 activation, inactive WASP binds to Cdc42–GTP through its CRIB domain, relieving its inhibitory conformation and exposing its VCA domain. The VCA domain is now available to recruit and activate the Arp2/3 complex, stimulating actin polymerization.

Other Rho Family Members

Although RhoA, Rac1, and Cdc42 are the most extensively studied members of the Rho family, other Rho GTPases also regulate the actin cytoskeleton (Table 1). First of all, closely related isoforms of these proteins such as RhoB or Rac2 appear to have distinct roles in regulating the actin cytoskeleton and cell adhesion. The Rnd proteins are unusual examples of GTP-binding proteins, as they do not cycle between an active GTP-bound state and an inactive GDP-bound state. Rnd1 and Rnd3 have been shown to antagonize RhoA in its effects on the actin cytoskeleton, promoting stress fiber and focal adhesion disassembly in several different types of mammalian cells. RhoH seems to antagonize some of the Rac1 GTPase functions related to lamellipodium extension and hence cell migration. Other members have more similar effects to those of the prototypical Rho GTPases: RhoG signals in parallel to Rac to regulate lamellipodium extension and cell migration, whereas RhoJ/TC10 is more related to Cdc42 and regulates vesicle trafficking.

Rho GTPase Family: Functions in Cell Biology

The ability to regulate the actin cytoskeleton dynamically enables Rho GTPases to be involved in many cellular processes, the most obvious of them being cell migration. However, there are many other processes that require Rho GTPases, such as secretion, endocytosis, and cell proliferation.

Cell Migration

Cell movement requires an orchestrated spatio-temporal regulation of the actin cytoskeleton, which involves coordinated activation of several members of the Rho GTPase family. Rac isoforms are important to induce actin polymerization in the

leading edge of the cell, in the form of lamellipodia and membrane ruffles, pushing forward the plasma membrane and generating the driving force for cell motility. Rac activity is also crucial to form adhesions as cells move forward and interact with new extracellular matrix. Rho contributes to cell motility by providing actomyosin-based contractility necessary for cell body contraction and tail retraction, and also has a role at the leading edge in lamellipodium extension. Cdc42 induces actin polymerization in invadopodia, which are involved in degrading the extracellular matrix to allow cells to migrate through a three-dimensional environment. In addition, Rho GTPases affect microtubule dynamics, which is also important for cell migration. For example, Cdc42 is important for directional migration and chemotaxis in some cell types, predominantly by acting through the Par complex and polarizing the microtubule cytoskeleton. Although the principal roles of Rho GTPases in cell migration have been already established, there are important differences in the role that these proteins have in different cell types, such as epithelial cells versus macrophages. Furthermore, it is not yet fully understood how these proteins and their effectors are spatially and temporally regulated, and how such an exquisite level of coordination in their functions is achieved in cell migration.

Secretion and Endocytosis

Vesicle trafficking requires both the actin cytoskeleton and the microtubule network, and several Rho GTPases have been shown to affect different aspects of vesicle trafficking. For example, Cdc42 is involved in endocytosis and Golgi-to-endoplasmic reticulum (ER) transport, and this involves its downstream target N-WASP. In addition, three less-studied Rho family members, RhoB, RhoD, and RhoQ, have been found to localize to endosomes and modulate endosome dynamics.

Proliferation and Transformation

Both Rho and Rac isoforms have been shown to affect cell proliferation. Although Rho effectors are important for the actin ring contraction that drives cytokinesis, it is in the G₁ phase of the cell cycle where Rho GTPase function seems to be contribute to cell-cycle progression. Activation of both Rho and Rac has been linked to the activation of cyclin-dependent kinases by regulating the expression of D-type cyclins and cyclin-dependent kinase inhibitors, such as p21^{cip1} and p27^{kip1}. In contrast to conventional Rho GTPases, Rnd3 negatively regulates cell proliferation by inhibiting cyclin D1 expression. However, the downstream effectors that lie between Rho GTPases and cell-cycle regulators remain unknown. Interestingly, Rho GTPases also contribute to cell transformation. The expression and activity of several Rho GTPases are increased in cancer, correlating with tumor progression. Active mutants of Rho, Rac, and Cdc42 all have positive effects on cell transformation and anchorage-independent growth, although they are not known to be mutated in human tumors. However, their biological effects on cell proliferation and actin cytoskeleton place them as potentially important mediators of neoplastic transformation. For instance, their crucial role in cell motility has been shown

to influence the loss of epithelial polarity and invasiveness of cancer cells.

See also: [Cell Architecture and Function: Actin Assembly/Disassembly](#); [Actin-Capping and -Severing Proteins](#); [Actin-Related Proteins](#); [Lipids Carbohydrates Membranes and Membrane Proteins](#); [Cell-Matrix Interactions](#); [Endocytosis](#); [Signaling: Small GTPases](#).

Further Reading

- Heasman S and Ridley AJ (2008) Mammalian Rho GTPases: New insights into their functions from *in vivo* studies. *Nature Reviews Molecular Cell Biology* 9: 690–701.
- Jaffe AB and Hall A (2005) Rho GTPases: Biochemistry and biology. *Annual Review of Cell and Developmental Biology* 21: 247–269.
- Rossman KL, Der CJ, and Sondek J (2005) GEF means go: Turning on RHO GTPases with guanine nucleotide-exchange factors. *Nature Reviews Molecular Cell Biology* 6: 167–180.
- Tcherkezian J and Lamarche-Vane N (2007) Current knowledge of the large RhoGAP family of proteins. *Biology of the Cell* 99: 67–86.
- Vega FM and Ridley AJ (2008) Rho GTPases in cancer cell biology. *FEBS Letters* 582: 2093–2101.
- Visvikis O, Maddugoda MP, and Lemichez E (2010) Direct modifications of Rho proteins: Deconstructing GTPase regulation. *Biology of the Cell* 102: 377–489.

Ribosome Assembly

A Sahasranaman and J L Woolford Jr., Carnegie Mellon University, Pittsburgh, PA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Assembly factors Proteins required for assembly of ribosomes, which are not themselves components of mature ribosomes.

Assembly subcomplexes Stable, isolatable complexes of a few assembly factors and ribosomal proteins.

Ribosome assembly Folding and processing of ribosomal RNA, and association of ribosomal proteins with ribosomal RNA.

Ribosomal proteins Proteins that are present in mature, functioning ribosomes.

RNA helicases RNA-dependent ATPases that catalyze pairing or unpairing of RNA or alteration of RNA–protein interactions.

RNA processing Removal of internal or 5′ or 3′ flanking sequences by exo- and endonucleolytic processing.

RNA–protein interactions Noncovalent interactions between RNA and proteins.

Structural proofreading The process of surveilling misassembly of ribosomes.

Introduction

Ribosomes are ribonucleoprotein (RNP) nanomachines that synthesize proteins. The small subunit (SSU) brings messenger RNA (mRNA) and transfer RNA (tRNA) together in the decoding center, and the large subunit (LSU) catalyzes peptide bond formation in the peptidyl transferase center. Exquisite high-resolution structures of these ancient multimolecular machines have facilitated informed predictions about ribosome function. While their function is phylogenetically conserved, there are differences in the size and complexity of ribosomes in prokaryotes and eukaryotes. In prokaryotes and archaea, the 30S ribosomal subunit contains the 16S ribosomal RNA (rRNA) and 20 ribosomal proteins (r-proteins), while the 50S ribosomal subunit is composed of 23S and 5S rRNAs, and 34 r-proteins. Eukaryote ribosomes are more complex. The 40S subunit of eukaryotes is composed of 18S rRNA and 33 r-proteins, and the 60S subunit consists of 25S, 5.8S, and 5S rRNAs, and 47 different r-proteins. The presence of insertion elements in eukaryotic rRNAs (referred to as expansion segments), and 20–30 additional r-proteins, might account for observed mechanistic differences between prokaryotes and eukaryotes during the initiation and elongation phases of translation, thereby providing more flexibility for regulation of translation in eukaryotes.

In this article, we focus on what has been learned about how multiple, dynamic protein–protein, RNA–RNA, and RNA–protein interactions are established and rearranged during the assembly of ribosomes both *in vitro* and *in vivo*. Ribosome assembly requires coordination of several processes: synthesis, folding, modification, and nucleolytic processing of pre-rRNAs, and ordered binding of r-proteins to the pre-rRNA. In prokaryotes, ribosome assembly and function occur in the same compartment, whereas these processes are largely but not completely segregated in eukaryotes. Ribosome biogenesis begins with transcription of rDNA genes in the nucleolus, and subsequent maturation occurs upon release of assembling ribosomes into the nucleoplasm to enable export, and even in the cytoplasm, prior to establishing functional subunits.

Model Systems

Much has been learned about ribosome biogenesis from studies of bacterial ribosome assembly *in vitro* and from analysis of yeast ribosome synthesis *in vivo*. More recently, investigations of metazoans are revealing how dysregulation of ribosome biogenesis results in various diseases. Analysis of the nucleotide sequence of genomes from prokaryotic and eukaryotic model organisms demonstrates that many r-proteins are conserved. Whereas the primary sequence of rRNAs is not completely conserved between prokaryotes and eukaryotes, the secondary structure of rRNAs is conserved. Thus, the principles and mechanisms of ribosome assembly learned from bacteria and yeast will be applicable to other organisms.

Assembly Principles Learned from *In Vitro* Studies

Functional bacterial 30S and 50S ribosomal subunits can be reconstituted *in vitro* from rRNA plus each of the ribosomal proteins. Detailed studies of the biochemical and biophysical principles underlying ribosome assembly *in vitro* have provided useful paradigms to guide investigations of ribosome biogenesis *in vivo*.

One of the first principles revealed by these experiments is that assembly of r-proteins into ribosomal subunits occurs in a hierarchical manner. Reconstitution studies of bacterial 30S ribosomal subunits defined an ‘assembly map’ for the order of association of each r-protein with rRNA. Based on their order in the binding hierarchy, r-proteins are designated as primary (binding directly to rRNA), secondary (binding is dependent on primary binding proteins), or tertiary proteins (binding is dependent on secondary binding proteins). While such a dependency map has proved to be a very useful starting point, it does not provide any information on the kinetics of protein binding. Rather, the pathway of assembly may largely reflect the order in which each r-protein is most stably associated with rRNA. Additionally, the assembly map does not take into account either the folding pathway of 16S rRNA

independent of r-proteins or the role in ribosome assembly of nucleolytic processing of rRNA precursors that occurs *in vivo*. More recently, these aspects of 30S subunit assembly have been explored in greater detail using techniques such as chemical and hydroxyl radical probing of RNA conformation to assay changes in rRNP conformation with time, and pulse-chase/mass spectrometry (PC/MS) to determine kinetics of r-protein binding to rRNA.

A possible explanation for the observation of primary binding r-proteins in the assembly map is that their binding sites are created at least in part by the conformation adopted by 16S rRNA independent of r-proteins. In fact, consistent with this idea and the view that ribosomes have evolved from an RNA catalyst, it has been shown that the 5' domain of 16S rRNA can fold and form tertiary contacts in the absence of all r-proteins. This tertiary structure is presumably not completely stable in the absence of protein contacts and, thus, is most likely stabilized by the binding of r-proteins to sites that are created by transient conformations adopted by rRNA as it undergoes folding. Therefore, the second principle suggests that rRNA folding provides the basis of r-protein binding during ribosome assembly.

A third principle that has emerged from *in vitro* studies of bacterial ribosomes is that strengthening of r-protein binding with rRNA occurs as ribosome assembly progresses. Many r-proteins contact multiple nucleotides on rRNA. Time-resolved hydroxyl radical footprinting has revealed that some of these nucleotides are protected before others, suggesting that as ribosome biogenesis advances, r-proteins make more contacts with rRNA, thereby becoming more stably integrated with ribosomes.

The initial establishment of a few contacts between r-proteins and rRNA potentially stabilizes certain RNA conformations, and induces changes in rRNA conformation to provide binding sites for additional (secondary or tertiary) r-proteins. Therefore, r-protein binding consolidates RNA folding gains and imparts directionality to the assembly process. This suggests a model where assembly proceeds through an alternating series of RNA conformational changes and protein binding, which sequentially stabilize the final RNA structure. Kinetic data have revealed that assembly of mature 30S subunits occurs via multiple, parallel RNA folding pathways, all of which converge on the final product. This has led to the concept of an assembly landscape as opposed to a unique pathway over which assembly of ribosomal subunits occurs.

Finally, *in vitro* studies have revealed that ribosome assembly occurs in the 5' to 3' direction. PC/MS and chemical probing of secondary structure of rRNA have shown that r-protein binding to the 5' domain of rRNA can be observed before protein contacts are established with the 3' domain of rRNA. By contrast, hydroxyl radical footprinting assays showed an initial burst of nucleotide protection all along 16S rRNA indicating that RNA folding nucleates from many different sites spread across the RNA, and thus pointing to a lack of 5' to 3' directionality in assembly. These conflicting observations can be reconciled by taking into account the nature of the different experimental setups. PC/MS measures the kinetics of protein binding to rRNA, and as such only stable r-protein–rRNA interactions can be detected because r-proteins that bind weakly during the pulse can be washed off during the chase. Footprinting assays, on the other hand, can capture nucleotide protection by both weakly-interacting and stably-associated

r-proteins. Taken together, these data indicate that while ribosome assembly can nucleate across the entire rRNA, the final tight association of r-proteins with rRNA occurs in a 5' to 3' direction.

It is important to consider how these *in vitro* experiments might or might not reflect assembly *in vivo*. For example, when ribosome assembly occurs *in vivo*, it is coupled with transcription of rRNA, which also presumably contributes to the observed 5' to 3' directionality of r-protein association. rRNA may have evolved to fold 5' to 3' as it is synthesized, and to incorporate r-proteins thusly, thereby coupling assembly with transcription of rRNA. Furthermore, the requirement of a heating step during reconstitution of ribosomal subunits *in vitro* and the identification of a number of bacterial mutants that are defective in production of ribosomes indicate the need for *trans*-acting factors during ribosome biogenesis *in vivo*.

Eukaryotic Ribosome Biogenesis *In Vivo*

Synthesis and Processing of rRNA

In yeast, the 9.1-kb rDNA fragment encoding all four rRNAs is present as 100–200 tandem repeats on chromosome XII. The high copy number of these genes enables cells to produce large quantities of ribosomes when necessary, for example, ~40 ribosomes per second in rapidly dividing yeast cells. Transcription of the 35S precursor containing three of the four rRNAs is carried out by RNA polymerase I in the nucleolus, while pre-5S rRNA is independently transcribed by RNA polymerase III. The 35S pre-rRNA contains sequences encoding the 25S, 18S, and 5.8S rRNAs flanked by 5' and 3' external transcribed spacers (ETSs) and separated by internal transcribed spacers 1 and 2 (ITS1 and ITS2). The predicted secondary structures of the mature rRNAs are highly conserved. Although these features are somewhat less conserved for spacer sequences, it appears that their secondary structures are important for their removal. The gene encoding 5S rRNA is separated from the gene encoding 35S pre-rRNA by intergenic sequences (IGSs), which are transcribed by RNA polymerase II to synthesize, at low levels, noncoding RNAs. Transcription by RNA polymerase I leads to silencing of these noncoding RNAs, perhaps through a mechanism involving modification of histone acetylation. Concurrent with or immediately following transcription, the 35S pre-rRNA is covalently modified by the box C+D and box H+ACA snoRNPs. They are composed of snoRNAs, which guide the snoRNPs to sites of modification on rRNA by base pairing with the rRNA. Methylases and pseudouridylylases present in the box C+D and box H+ACA snoRNPs, respectively, then carry out modification of the appropriate nucleotides. These sites of modification are located within conserved, functionally important domains of mature rRNA.

The presence of three of the four mature rRNAs within one transcript, one of which ends up in the 40S subunit and the other two in the 60S subunit, presumably ensures that stoichiometric amounts of rRNAs are delivered to 40S and 60S subunits. Furthermore, proper timing of removal of the spacer sequences could potentially allow pre-rRNA processing to be subject to cues from the cellular environment to regulate ribosome biogenesis. A more interesting potential role for these spacers is to ensure proper folding of rRNA by enabling,

for example, base pairing between different mature rRNA sequences required for formation of functional ribosomes.

Three successive cleavages by unidentified nucleases at sites A_0 , A_1 , and A_2 generate the 20S pre-rRNA and 27SA₂ pre-rRNAs, thus separating the mature rRNAs of the 40S and 60S ribosomal subunits (Figure 1). The 20S pre-rRNA is exported to the cytoplasm where its 3' end is cleaved to form the 18S rRNA. This is the last known step in the formation of 40S ribosomal subunits. Processing of 27SA₂ pre-rRNA to form mature rRNAs of the 60S ribosomal subunit is more complex. Alternative pathways generate two forms of the 27SB pre-rRNA – the major, short form 27SB_S pre-rRNA (~85%) and the 27SB_L pre-rRNA (~15%), which is ~7 nt longer at its 5' end. In the major pathway, the 27SA₂ pre-rRNA is cleaved at the A_3 site by RNase MRP, followed by exonuclease digestion from the A_3 site to the B_{1S} site by the exonucleases Rat1, Xrn1, and Rrp17. In the alternative pathway, 27SA₂ pre-rRNA is endonucleolytically processed at the B_{1L} site to create the 27SB_L pre-rRNA. The processing of 27SB_S and 27SB_L pre-rRNAs to 25S rRNA and 5.8S_S or 5.8S_L rRNAs, respectively, is thought to be identical. Cleavage at site C_2 by an unknown endonuclease separates the precursors of the 5.8S and 25S rRNAs, the 7S pre-rRNA and the 25.5S pre-rRNA respectively. Subsequent 3' processing of 7S pre-rRNA by the exosome complex and 5' processing of 25.5S pre-rRNA by the exonucleases Rat1 and Xrn1 eventually lead to the formation of mature 5.8S and 25S rRNAs. While the identity of some exo- and endo-nucleases involved in pre-rRNA processing is known, many of these nucleases remain to be discovered.

Pre-Ribosomal Intermediates and the Roles of Assembly Factors

Initial studies revealed a crude outline of the assembly pathway (Figure 2). 35S pre-rRNA is present in 90S pre-ribosomes (pre-rRNP) that give rise to 43S and 66S pre-rRNP assembly intermediates. These precursor particles are then converted to the mature 40S and 60S ribosomal subunits, respectively. The 43S assembly intermediate contains the 20S pre-rRNA. There are at least six 66S assembly intermediates containing the 27SA₂, 27SA₃, 27SB_L or 27SB_S, 25.5S + 7S, and 25S + 5.8S rRNAs, respectively.

Each of these pre-rRNPs contains a subset of the r-proteins found in mature subunits, plus a number of nonribosomal proteins/assembly factors not present in mature functioning ribosomes. More than 200 such *trans*-acting factors required for assembly of ribosomes have been identified in yeast, using genetic screens or selections to identify mutants defective in ribosome biogenesis, or using proteomic approaches to purify and characterize assembly intermediates. Recently, metazoan counterparts of most of these proteins have been identified by mass spectrometry of proteins from purified nucleoli. These assembly factors include potential RNA-binding proteins, scaffolding proteins, RNA helicases/RNPases (DEXD/H box proteins), AAA+ ATPases, GTPases, kinases, and interestingly, ribosome-like proteins.

While initial characterization assigned most factors to one or more steps in pre-rRNA processing/subunit assembly, more

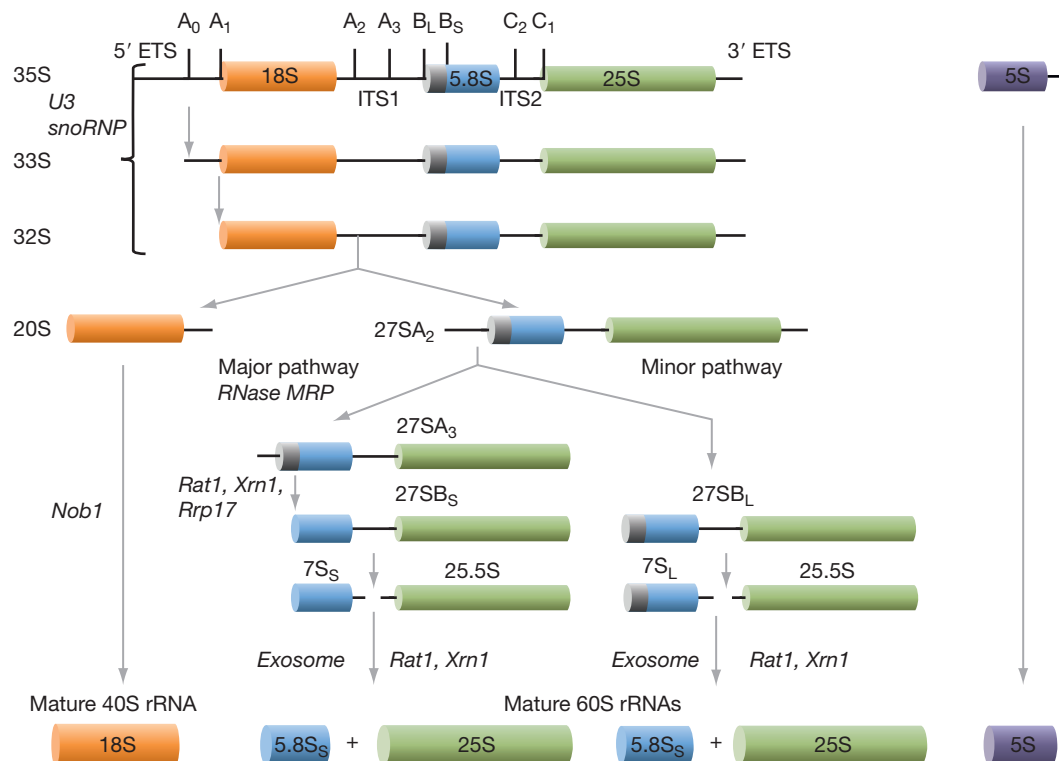


Figure 1 Pre-rRNA processing pathway in *Saccharomyces cerevisiae*. The initial 35S pre-rRNA transcript, synthesized by RNA polymerase I, contains sequences for mature 18S, 5.8S, and 25S rRNAs (cylinders) along with internal and external transcribed spacer sequences (thin horizontal lines). Sites on pre-rRNA at which processing occurs are indicated on the 35S pre-rRNA transcript. Processing intermediates that form during ribosome assembly are indicated. Exo- and endo-nucleases, where known, are mentioned alongside the step in which they function.

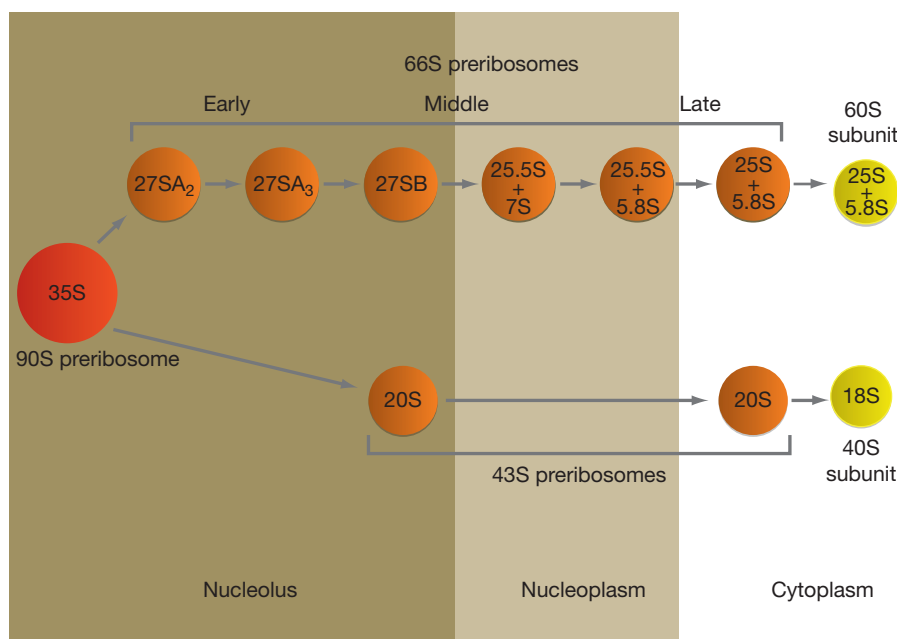


Figure 2 Pathway of assembly of 60S and 40S ribosomal subunits in *Saccharomyces cerevisiae*. The red circle represents 90S pre-ribosomes; subsequent 66S and 43S pre-ribosomes are shown as orange circles. The pre-rRNA processing intermediate present in each pre-ribosomal particle is indicated within the circles. Yellow circles represent mature 60S and 40S ribosomal subunits. Subcellular compartments within which different pre-ribosomal intermediates are formed are indicated.

specific roles of assembly factors have been experimentally verified in only a few cases. An interesting example is the kinase Hrr25, which is required for structural reorganization during the final stages of maturation of 40S ribosomal subunits. Phosphorylation and subsequent dephosphorylation of rpS3 by Hrr25 enable tighter binding of this r-protein to 40S precursors and reorganizes the beak structure, formed in part by 18S rRNA. Another example is the GTPase Bms1, which in its GTP-bound form has been shown to interact with Rcl1, a putative endonuclease, and U3 snoRNA, during an early step of 40S subunit assembly. GTP hydrolysis causes Bms1 to be released from pre-ribosomes leaving behind Rcl1 on pre-rRNPs. Thus, Bms1 is thought to function as a GTP-regulated switch that delivers Rcl1 to pre-ribosomes. By contrast, GTPases required for assembly of 60S ribosomal subunits are thought to function later in the assembly pathway and lead to dissociation and recycling of assembly factors. The same is true for AAA+ ATPases, the most well-studied of which is Rea1. Rea1 is thought to use the energy of adenosine triphosphate (ATP) hydrolysis within its ring domain to create a tensile force, which results in the release of several different assembly factors from two distinct regions of pre-ribosomes. Rea1-mediated dissociation of assembly factors occurs just prior to nuclear export of pre-ribosomes.

Compartmentalization of assembly: structural proofreading

Late steps in ribosome biogenesis in yeast occur in the cytoplasm, where a few remaining assembly factors are released to generate translation-competent ribosomal subunits. It is thought that structural proofreading monitors correct assembly of ribosomes, and ensures that incorrectly assembled pre-ribosomes are not exported to the cytoplasm. Such a quality-control

mechanism would couple export of ribosomal precursors with late maturation events. In principle, export factors could only bind to properly assembled ribosomal subunits. Furthermore, the presence of some assembly factors might prevent incompletely assembled ribosomes from functioning prematurely in translation. The export factor Nmd3 has been shown to interact genetically with the mRNA exporters Mex67/Mtr2, suggesting that Mex67/Mtr2 might also enable export of pre-60S ribosomal subunits through interactions with the nuclear pore complex (NPC). In addition to Mex67/Mtr2, the shuttling assembly factor Arx1 is also thought to facilitate export of pre-60S subunits, via interactions with the NPC. Both Nmd3 and Arx1 have been shown to accompany 60S ribosomal subunits to the cytoplasm (Figure 3).

The assembly factors Tif6, Mrt4, Alb1, and Rlp24/Nog1 also accompany pre-60S ribosomal subunits to the cytoplasm, where they are released in a step-wise manner by the activity of ATPases and GTPases, thereby generating functional 60S ribosomal subunits. The first step is initiated by the AAA+ ATPase Drg1, which releases the assembly factors Nog1 and Rlp24. The release of the ribosome-like protein Rlp24 enables association of the r-protein, rpL24. The Drg1-mediated release of assembly factors is followed by two events – the release of Arx1/Alb1 by Rei1/Jjj1/Ssa and the assembly of the P0 stalk. The dual-specificity phosphatase Yvh1 is required for the removal of another ribosome-like protein Mrt4, which in turn enables stable association of the ribosomal protein P0. Assembly of the P0 stalk is thought to recruit the GTPase Efl1 for release of Tif6. The release of Nmd3 by the GTPase Lsg1 and the r-protein rpL10 is thought to be the last event in the cytoplasmic maturation of 60S ribosomal subunits. With the release of Nmd3, the mature 60S ribosomal subunit is formed, ready to engage in translation.

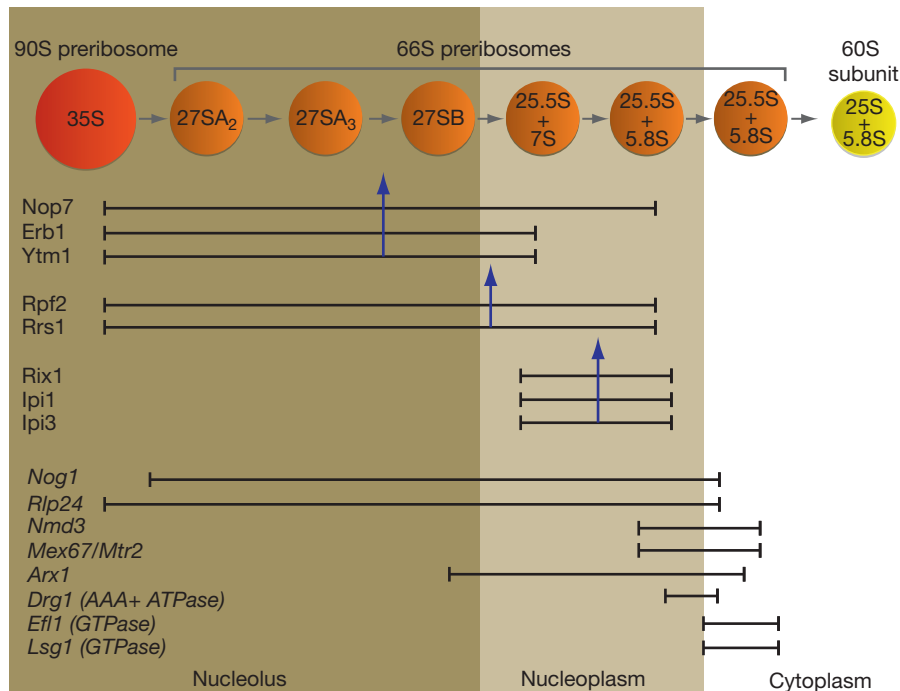


Figure 3 Association of assembly factors with pre-ribosomal intermediates. Pathway of assembly of 60S ribosomal subunits is shown. Each horizontal line spans all pre-ribosomal intermediates with which a specific assembly factor associates. Assembly factors (except those in italics) are organized into different physical subcomplexes. The blue vertical arrow indicates the pre-rRNA processing step for which each physical subcomplex is required. Assembly factors in italics are those required for structural proofreading. They do not form a physical neighborhood. Refer to text for details of step-wise dissociation of these late factors.

Role of Ribosomal Proteins in Ribosome Biogenesis

In yeast, most r-proteins are essential for growth, indicating either a more expansive role for r-proteins in eukaryotic ribosomes or else that some might have extraribosomal roles. Because of a preoccupation with assembly factors, with a few notable exceptions, the role of r-proteins in pre-rRNA processing and ribosomal subunit maturation went largely unstudied for a long time. Pioneering work by Milkereit and colleagues corrected this imbalance by analyzing the role of SSU r-proteins in the formation of 40S subunits. They found that most essential SSU r-proteins were required for formation of 18S rRNA and many were required for efficient export of pre-40S particles to the cytoplasm. Stable association of r-proteins with 43S pre-ribosomes is important for significant restructuring events such as formation of the head-like and body-like regions of 40S ribosomal subunits. Furthermore, consistent with *in vitro* studies of bacterial ribosome assembly, it has been shown that r-protein binding with pre-ribosomes is strengthened as pre-rRNA processing and assembly proceed. A study with the LSU r-proteins in yeast has shown that the essential r-proteins of the 60S subunit can also be grouped based on the pre-rRNA processing step(s) for which they are required.

Studying Physical Neighborhoods to Understand Ribosome Assembly

One approach to simplify the study of ribosome biogenesis, while taking into account the cooperativity that exists during the assembly process, is to understand how small physical

subcomplexes or neighborhoods are constructed within nascent ribosomes. Physical clusters likely contain molecules, many or all of which are required for the same step in ribosome assembly (Figure 3). A number of such physical subcomplexes have been identified, each of which is required for a specific step of maturation of 66S pre-rRNPs. By studying when members of a physical neighborhood associate with (and dissociate from) pre-ribosomes, how they interact with each other, and their interdependencies for association with pre-ribosomes, one can model how assembly factors and r-proteins within a physical neighborhood cooperate with each other to enable a specific step in maturation of ribosomal subunits to occur efficiently.

One of the emerging principles underlying the formation of pre-ribosomal particles is the stepwise or ordered binding of such discrete subcomplexes to pre-rRNA. This principle is exemplified by studies on the formation of 90S pre-rRNPs, where assembly of the t-UTP subcomplex leads to the ordered binding of five additional subcomplexes through two mutually independent assembly routes. This has given rise to a view of the 90S pre-ribosome as a modular structure that is constructed in a hierarchical fashion. Similar principles can be expected to underlie the formation of 66S pre-rRNPs. Constructing such large RNPs from multiple smaller subcomplexes, therefore, provides a means to simplify the assembly process.

Quality Control of Ribosomal Subunits

Given that ribosome assembly requires the coordinated functions of many proteins and RNAs, and constant structural

changes, there is a high probability that a certain fraction of ribosomes will be misassembled. Such aberrant particles might be detrimental to cells by interfering with protein synthesis directly or indirectly. Therefore, it is important that such misassembled ribosomal subunits are recognized, thereby recruiting quality-control machinery to degrade nonfunctional ribosomes. Indeed, such quality-control pathways to turn over dysfunctional mature ribosomes exist. rRNA mutations in the decoding or peptidyl transferase center (PTC) of ribosomes are recognized and subjected to nonfunctional rRNA decay (NRD). However, it appears that there are mechanistically distinct NRD pathways involved in turnover of 40S and 60S ribosomal subunits. The NRD of nonfunctional 40S subunits is similar to no-go decay (NGD) of mRNA. By contrast to 40S subunits, it has been shown that nonfunctional 60S ribosomes are subject to active ubiquitin-dependent degradation. This underscores the fact that turnover of 60S and 40S subunits occur independently of each other.

In addition to surveillance of the functionality of mature ribosomal subunits that occurs in the cytoplasm, it appears that there might be nuclear quality control processes in place to monitor aberrant pre-ribosomal assembly intermediates. Interestingly, when ribosome biogenesis is blocked, rRNA precursors fail to accumulate to high levels. It has been shown that the Trf4/5, Air1, Mtr4 polyadenylation (TRAMP) complex is involved in the polyadenylation of rRNA precursors that could lead to exosome-mediated degradation of aberrant intermediates.

Ribosome Biogenesis and Disease

Mutations in both r-proteins and assembly factors have been shown to result in a variety of genetic disorders. This came as a surprise because it has long been thought that such mutations would be lethal because of the essential nature of the proteins and the process in which they function. Furthermore, all these diseases show some degree of tissue-specificity, which is intriguing given that ribosomes are present in all cell types in the body. Mutations in assembly factor genes can act as the causative agent, or, in some cases they can exacerbate the effect of a mutation in a non-pre-ribosome associated gene product. These mutations result in a variety of diseases ranging from alopecia and male infertility to Dyskeratosis congenita and Prader-Willi syndrome. Dyskeratosis congenita is caused by a mutation in the *DKC1* gene, which encodes dyskerin, a pseudouridine synthase required for rRNA modification and is characterized by premature ageing and increased susceptibility to cancer. Mutations in the r-protein S19 have been shown to result in another syndrome called Diamond-Blackfan anemia, which typically also results in increased cancer susceptibility. Understanding the molecular mechanisms underlying diseases caused by defects in ribosome biogenesis is an important challenge for the future.

The link between ribosome biogenesis and cancer is complicated, and therefore not very well understood. rRNA synthesis is tightly coupled with cell-cycle progression, to ensure proper cell growth and proliferation. The tumor suppressor p53 is known to repress RNA polymerase I, and in cancer cells carrying mutations in p53, RNA polymerase I activity is derepressed, leading to increased rRNA synthesis. Similarly, r-proteins are also upregulated in cancer cells, thus amplifying the number of functional ribosomes in the cell. This would have the effect of increasing the translation rate in cells and therefore promoting transformation. Because r-protein expression is upregulated in cancer cells, it is still unclear how loss of an r-protein could contribute to neoplasm. It is conceivable that haploinsufficiency of r-proteins would affect ribosome biogenesis, resulting in a decrease in global translation of mRNAs, including those encoding tumor suppressor genes. Thus loss of r-proteins could result in cancer. Recently, it has been shown that haploinsufficiency of *RPS6* and *RPS7* (leading to reduced levels of these proteins) affects 40S subunit biogenesis and leads to p53 activation and apoptosis through a mechanism based on translation of RPL11. Under conditions of reduced *RPS6* and *RPS7*, even though there is a reduction in global translation levels, there is increased translation of mRNAs with a polypyrimidine tract at the transcription start (5'-TOP mRNAs). Since all mRNAs encoding r-proteins contain a 5'-TOP, the resulting overexpression of RPL11 would inhibit MDM2/HDM2, thereby activating p53 and promoting cell death and apoptosis. It is still unclear how haploinsufficiency of r-proteins can lead to apoptosis on the one hand and promote tumorigenesis on the other. The mechanisms underlying these complex signaling pathways, their relationship with each other, and their response to different cellular stimuli need to be investigated in greater detail.

See also: **Molecular Biology:** Nucleolus, Overview; Ribosome Structure; RNA Polymerase I and RNA Polymerase III in Eukaryotes; **Protein/Enzyme Structure Function and Degradation:** AAA-ATPases; Affinity Tags for Protein Purification; Mass Spectrometry of Native Complexes.

Further Reading

- Freed EF, Bleichert F, Dutca LM, and Baserga SJ (2009) When ribosomes go bad: Diseases of ribosome biogenesis. *Molecular BioSystems* 6: 481–493.
- Henras AK, Soudet J, Gerus M, et al. (2008) The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cellular and Molecular Life Sciences* 65: 2334–2359.
- Kressler D, Hurt E, and Baßler J (2009) Driving ribosome assembly. *Biochimica et Biophysica Acta* <http://dx.doi.org/10.1016/j.bbamcr.2009.10.009>.
- Panse VG and Johnson AW (2010) Maturation of eukaryotic ribosomes: Acquisition of functionality. *Trends in Biochemical Sciences* 35(5): 260–266.
- Strunk BS and Karbstein K (2009) Powering through ribosome assembly. *RNA* 15: 2083–2104.
- Sykes MT and Williamson JR (2009) A complex assembly landscape for the 30S ribosomal subunit. *Annual Review of Biophysics* 38: 197–215.

Ribosome Regulation by EF-G and EF-Tu

S T Gregory, Brown University, Providence, RI, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Poul Nissen and Jens Nyborg, volume 2, pp. 1–5, © 2004, Elsevier Inc.

Glossary

Elongation factor G (EF-G) The protein factor responsible for catalyzing the translocation step of protein synthesis.

Elongation factor TU (EF-Tu) The protein factor responsible for delivery of aminoacyl-tRNA to the ribosome.

Hybrid tRNA-binding states The binding state of peptidyl-tRNA, in which the anticodon end is in the A site of the 30S subunit and the acceptor end is in the P site of the 50S subunit, and deacyl-tRNA is in the 30S P site and 50S E site. This designation is relative to the classical tRNA-binding states, as defined originally, with the A site as the binding site for aminoacyl-tRNA and the P site as the binding site for peptidyl-tRNA.

Ribosome The universal particle responsible for protein synthesis in all organisms. It is composed largely of ribosomal RNA (rRNA), the folding of which to adopt an active conformation is aided by ribosomal proteins.

Substrate proofreading The rejection of noncognate or near-cognate substrates prior to polymerization, in contrast to product proofreading used by DNA polymerases to remove erroneously incorporated monomers from nascent polymer chains.

Transfer RNA (tRNA) RNA molecules of around 76 nucleotides that are the substrates of the ribosome. They contain triplet base sequences called anticodons that base pair with the corresponding codon of messenger RNA (mRNA) in the 30S subunit during translation. Amino acids are attached to the 3' end of tRNAs.

Translocation The movement of tRNAs and mRNA by one codon through the ribosome during protein synthesis.

tRNA selection The process by which the ribosome, together with EF-Tu, discriminates between aminoacyl-tRNAs based on the codon present in the ribosomal A site. This is also often referred to as decoding.

Introduction

Protein synthesis is a highly complex, multistep process that universally occurs on the 2.5-million-Da ribonucleoprotein particle called the ribosome. Ribosomes consist of two subunits of unequal size and are composed primarily of RNA. The most thoroughly studied ribosomes are those of bacteria, whose small or 30S subunit consists of a single 16S ribosomal RNA (rRNA) molecule and around 20 ribosomal proteins, and whose large or 50S subunit consists of single 23S and 5S rRNA molecules and around 30 ribosomal proteins. Ribosomes from eukaryotes are composed of slightly larger rRNA molecules with additional ribosomal proteins. These large RNA molecules fold into complex three-dimensional structures, aided by the ribosomal proteins, to produce functionally active ribosomal subunits. The two subunits associate with one another to produce a 70S ribosome, with the space between them forming three transfer RNA (tRNA) substrate-binding sites. The manner in which tRNAs bind to and move through the ribosome forms the essence of the translation process.

Protein synthesis on the ribosome occurs in the three main phases of initiation, elongation, and termination. Elongation is an iterative process and can itself be divided into the three steps of tRNA selection, peptide bond formation, and translocation. During each elongation cycle, a tRNA substrate moves sequentially through three tRNA-binding sites, designated A (for acceptor or aminoacyl), P (for peptidyl), and E (for exit). To enter the elongation cycle, an aminoacyl-tRNA must be selected by the ribosome in accordance with the particular messenger RNA (mRNA) codon triplet present in the A site. This is the rate-limiting step of the elongation cycle, as the

ribosome must discriminate the correct, or cognate, aminoacyl-tRNA from the intracellular pool of much more numerous near-cognate and noncognate aminoacyl-tRNAs. No option exists other than to randomly sample a large number of aminoacyl-tRNAs for each codon. Successful selection of an aminoacyl-tRNA is followed by peptide bond formation, in which the growing nascent polypeptide chain is transferred from the peptidyl-tRNA in the P site to the aminoacyl-tRNA now stably bound in the A site. This leads to hybrid tRNA-binding states, with peptidyl-tRNA in the 30S A site and the 50S P site (the so-called A/P state) and deacyl-tRNA in the 30S P site and 50S E site (the P/E state). Hybrid states formation is accompanied by a relative rotation of the two ribosomal subunits. To advance to the next codon, the ribosome then undergoes a translocation step, during which the tRNA substrates move their anticodon ends in concert with mRNA on the 30S subunit to adopt the classical tRNA-binding states (P/P and E/E). This movement is accompanied by a reversal of subunit rotation back to the ground state, producing a vacant A site that is now ready for the next cycle of tRNA selection programmed by the next mRNA codon.

It is an axiom of the RNA world hypothesis that the prebiotic ribosome was composed entirely of RNA and carried out each of the steps of protein synthesis in a protein-independent manner. Even after 4.5 billion years of evolution, the ribosome retains the ability to carry out each of the elongation steps without added protein factors. The enzymatic activity of the ribosome, peptide bond formation, is catalyzed by a peptidyl-transferase active site that is still composed of rRNA. The chemical step of peptide bond formation is extremely fast, and evolutionary processes have not produced protein factors

to accelerate it. However, all organisms produce protein factors that greatly enhance the rates of the previous and subsequent steps, tRNA selection and translocation.

While free aminoacyl-tRNAs will bind to the ribosome in a codon-dependent manner *in vitro*, they are delivered to the ribosome *in vivo* as part of a ternary complex with a G-protein called elongation factor Tu (EF-Tu) and guanosine triphosphate (GTP). This incorporation of a GTPase protein factor into the kinetic pathway allows an additional discrimination or substrate-proofreading step, thereby greatly increasing the accuracy of tRNA selection. Translocation can also occur in the absence of exogenous factors, albeit very slowly. *In vivo*, cells greatly accelerate this reaction using second G-protein called EF-G. Thus, EF-Tu exists primarily to greatly enhance decoding accuracy by providing a proofreading mechanism, and EF-G accelerates translocation that would otherwise be unacceptably slow for the demands of the modern cell. One striking observation is the similarity in the overall shape of the EF-Tu–aminoacyl-tRNA–GTP ternary complex and EF-G, and the way in which they interact with the ribosome (Figure 1).

EF-Tu and Initial Selection

The very first event in tRNA selection is a codon-independent step termed initial binding. Cognate ternary complexes of EF-Tu–aminoacyl-tRNA–GTP undergo initial binding in competition with the more numerous noncognate and near-cognate complexes. This binding state is highly labile and has not been observed directly by cryo-electron microscopy (cryo-EM) or X-ray crystallography and has been detected solely by kinetic measurements. As this step occurs prior to codon–anticodon interaction, it does not contribute directly to tRNA discrimination.

After initial binding is initial recognition, wherein the aminoacyl-tRNA samples the codon in the A site. Noncognate aminoacyl-tRNAs are rejected at this step, while cognate aminoacyl-tRNAs (and less efficiently, near-cognate aminoacyl-tRNAs) adopt a stable binding state referred to as the A/T state, in which aminoacyl-tRNA occupies the 30S A site while remaining bound to EF-Tu. The ribosome contributes directly

to codon recognition by employing an induced fit mechanism, where the stereochemistry of codon–anticodon base pairing is precisely monitored by specific interactions with 16S rRNA residues. Of particular importance are those involving residues G530, which flips from the *syn* to the *anti* configuration, and A1492/A1493, which flip out from their stacked position in helix 44. C1054 also interacts with the codon–anticodon helix. These conformational changes place these residues in position to fit into the minor groove of the three-base-pair helix formed by the mRNA codon and the aminoacyl-tRNA anticodon (Figure 2(b)). The binding energy of these interactions induces a global conformational change in the 30S subunit from an open to a closed form that accelerates the forward steps of decoding. Because the non-A-form geometry of their codon–anticodon complexes is less effective at stimulating domain closure than cognate ones, near-cognate aminoacyl-tRNAs are usually rejected during initial recognition. Mutations perturbing the equilibrium of this open–closed transition, such as those producing changes in 16S rRNA, ribosomal proteins S4, S5, or S12, can either increase or decrease decoding accuracy.

Binding of aminoacyl-tRNA simultaneously to EF-Tu and the codon on the 30S subunit in the A/T state requires a 30° bend in the tRNA body in comparison to its configuration when bound to the ribosome in the A/A state (Figure 2(b)) or in complex with EF-Tu free in solution. The binding energy of codon recognition is used both to drive 30S subunit domain closure and to induce this bend. A G24A base substitution in the elbow region of tRNA^{Trp}, which is expected to facilitate this bending, accelerates both the initial recognition and proofreading rate constants. This results in misreading of UGA codons as UGG (Trp) codons.

EF-Tu and GTP Hydrolysis

After codon recognition and domain closure, the stable A/T-binding state of aminoacyl-tRNA activates the GTPase activity of EF-Tu. Near-cognate complexes do not effectively induce GTPase activation and have a greater likelihood of being released prior to GTP hydrolysis. Once GTP hydrolysis occurs, EF-Tu adopts a conformation that is incompatible with binding

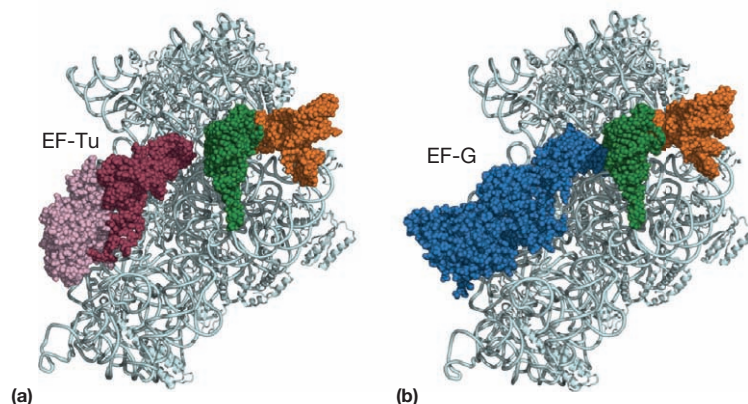


Figure 1 Elongation factors Tu and G interact with the ribosome in similar fashions. (a) The EF-Tu (light pink) delivers aminoacyl-tRNA (dark pink) to the A site of the ribosome as a ternary complex with GTP. (b) EF-G (blue) binds to the ribosome to promote translocation. In both panels, only the 30S ribosomal subunit (cyan) is shown, with the 50S subunit omitted for clarity. P-site (green) and E-site (orange) tRNAs are shown.

either the ribosome or aminoacyl-tRNA and is released as EF-Tu-GDP. Once free of EF-Tu, aminoacyl-tRNA cannot return to the A/T state, and in this way GTP hydrolysis kinetically drives decoding forward. Perhaps surprisingly, the ribosome does not take advantage of the 350-fold difference in thermodynamic stability of cognate versus near-cognate codon-anticodon complexes in tRNA discrimination, because GTP hydrolysis occurs long before binding equilibrium is reached.

How codon recognition stimulates GTPase activation is not entirely clear, given the distance (some 80 Å) between the decoding site on the 30S subunit and the GTPase active site of EF-Tu (Figure 3). A large number of ribosomal components interact with EF-Tu or aminoacyl-tRNA in the A/T state and it is difficult to determine which ones are responsible for activating GTPase differentially in response to cognate and near-cognate ternary complexes. One ribosomal component frequently implicated is 50S ribosomal protein L7/L12, which is present as a tetramer or hexamer, depending on the bacterial species

(L7 is an acetylated form of L12). The L7/L12 tetramer/hexamer forms a stalk protruding from the ribosome and is conformationally dynamic, and, therefore, has been poorly disordered in ribosome X-ray crystal structures. This stalk is thought to recruit ternary complex to the ribosome, and depletion of the stalk eliminates GTPase activation. Ribosomal protein L11, near the base of the stalk and close to the GTPase active site of EF-Tu, also forms part of the binding site for the antibiotic thiostrepton. On the 30S subunit, aminoacyl-tRNA in the A/T state makes direct contact with ribosomal protein S12. A large amount of genetic data implicates this protein in the decoding process, with some suggesting a role for this contact in GTPase activation. As no single component can monitor codon recognition and interact with the GTPase active site simultaneously, activation must result from the cooperative action of multiple ribosomal elements.

However GTPase activation occurs, GTP hydrolysis by EF-Tu is controlled by a hydrophobic gate, in which Val20

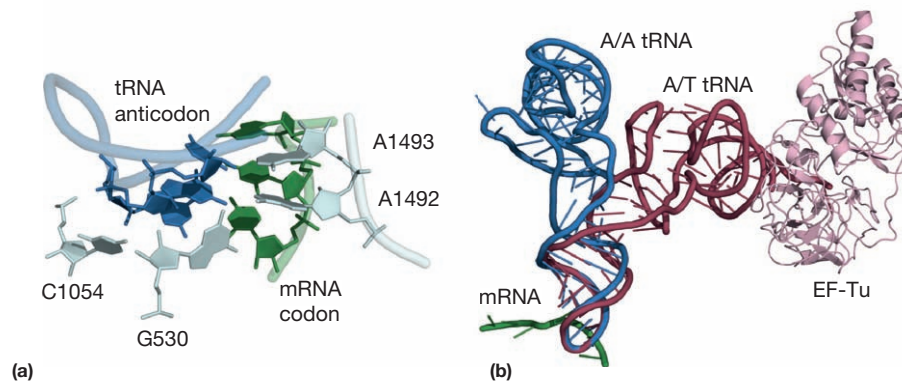


Figure 2 Codon recognition occurs in the ribosomal A site during tRNA selection. (a) 16S rRNA bases G530, C1054, A1492, and A1493 (cyan) change conformation, as part of an induced fit mechanism, to monitor Watson-Crick base pairing between the anticodon bases (blue) and the mRNA codon bases (green). Cognate codon recognition leads to a global conformational change in the 30S subunit. (b) Bending of aminoacyl-tRNA in the A/T state. This bending allows simultaneously codon recognition and interaction with EF-Tu. Base pairing between tRNA anticodon and mRNA codon occurs in the same manner in both the A/T (prior to proofreading) and A/A states (after proofreading and accommodation).

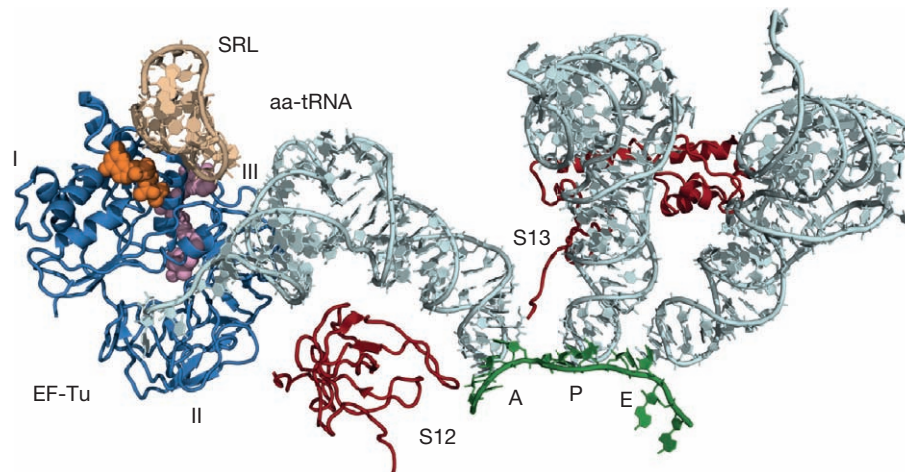


Figure 3 Details of ribosome interaction with EF-Tu-aminoacyl-tRNA-GTP in the A/T state. The acceptor helix of aminoacyl-tRNA interacts with ribosomal protein S12 (red). EF-Tu (blue) interacts with the sarcin-ricin loop (SRL) of 23S rRNA (tan). Kirromycin is shown as pink spheres, with GDP as orange spheres, while mRNA and tRNAs are colored green and cyan, respectively. Most of the ribosome is omitted for clarity.

and Ile60 move to allow putative catalytic residues Gly83 and His84 access to GTP. G2661 in the sarcin–ricin loop of 23S rRNA may hold His84 in an inactive conformation until cognate-codon recognition is signaled. Upon GTP hydrolysis, EF-Tu–GDP is released, leaving the aminoacyl-tRNA free to move into the A/A state in a process referred to as accommodation. Free EF-Tu–GDP then forms a complex with the elongation factor EF-Ts, GDP is exchanged for GTP, and EF-Tu–GTP reenters the pool of ternary complexes with aminoacyl-tRNA.

Substrate Proofreading

The ribosome achieves accuracy of tRNA selection by employing two sequential steps, the initial selection step described above and a substrate-proofreading step, kinetically separated by the irreversible hydrolysis of GTP by EF-Tu. The occasional near-cognate aminoacyl-tRNA that slips through initial selection to form a stable complex with the ribosome is usually rejected at this point. Although the mechanistic details of proofreading remain to be established, substrate proofreading by the ribosome is distinct from the product-proofreading mechanism utilized by DNA polymerases, in that it results in the rejection of incorrect substrates prior to, rather than after, polymerization. This two-step filtering mechanism greatly enhances the accuracy of tRNA selection, with the overall selectivity being the product of the selectivity of the individual steps.

An aminoacyl-tRNA that successfully passes through the proofreading step adopts the A/A classical tRNA-binding state in a process termed accommodation. In the A/A state, the acceptor end of the aminoacyl-tRNA is positioned in the peptidyltransferase active site on the 50S ribosomal subunit. Once peptide bond formation occurs, the ribosome has committed itself to incorporation of this amino acid into the growing polypeptide chain.

EF-G and Translocation

After tRNA selection and peptide bond formation, the ribosome undergoes a large conformational rearrangement, a relative rotation of the 30S and 50S subunits of about 6–10°. This rotation places the tRNAs in hybrid-binding states, in which the CCA ends of the A-site and P-site tRNAs have moved into the P and E sites, respectively, on the 50S subunit, while the anticodon ends remain in the A and P sites, respectively, on the 30S subunit. These hybrid states are adopted sequentially, with the P/E (30S P site/50S E site) hybrid state forming before the A/P (30S A site/50S P site) hybrid state. The next step is the concerted movement of the mRNA by one codon and the anticodon ends of the tRNAs into the 30S subunit P and E sites, and the reversal of the ratcheted state of the ribosome. This step is called translocation, and occurs spontaneously on the ribosome. The spontaneous reaction is very slow, and is accelerated about 50 000-fold by the activity of EF-G. Although EF-G hydrolyzes GTP, this catalytic activity is not absolutely essential for translocation, as EF-G–GDP or EF-G with a nonhydrolyzable GTP analog will also promote the reaction, albeit at a 50–100-fold lower rate. Instead, GTP hydrolysis actually precedes translocation and induces a conformation of EF-G that stabilizes an ‘unlocked’ configuration of the ribosome.

The unlocked configuration of the ribosome is associated with multiple structural changes, including the aforementioned ratcheting motion of the two subunits, a movement of the L1 stalk on the E-site side of the ribosome, and formation of the P/E hybrid tRNA-binding state. These three events are only loosely coupled to one another, and the role of EF-G may be to couple these events in order to make the ribosome competent for translocation. After translocation and release of EF-G–GDP, the ribosome reverts to the locked configuration, but with tRNAs now in the classical states (P/P and E/E), and is ready to select the next aminoacyl-tRNA.

Reverse translocation, with movement of tRNAs back into the hybrid states, occurs spontaneously in the absence of EF-G, indicating that movement of tRNAs in either direction through the ribosome in the unlocked configuration occurs by diffusion. EF-G favors movement in the forward direction, by preventing movement in the backward direction. This is consistent with the EF-G–ribosome co-crystal structure, which shows loop I of domain IV of EF-G stabilizing peptidyl-tRNA in the posttranslocation state via interactions with the minor groove of the codon–anticodon mini-helix in the P site (Figure 4). Thus, EF-G acts as a Brownian ratchet to prevent backward movement of the tRNA–mRNA complex.

EF-G–Ribosome Interactions

X-ray crystal structures of isolated EF-G or the EF-Tu–aminoacyl-tRNA–GTP ternary complex reveal an overall resemblance in shape, not only suggesting a common evolutionary origin of the two factors but also reflecting a similarity in the way they interact with the ribosome. This latter point is most obvious in cryo-EM reconstructions and X-ray crystal structures of these factors bound to the ribosome (refer once again to Figure 1). In comparison to EF-Tu, which adopts radically different conformations in the EF-Tu–aminoacyl-tRNA–GTP ternary complex and the EF-Tu–GDP binary complex, EF-G undergoes only limited changes in conformation. Domain III is particularly dynamic in its position relative to the rest of the factor, and was disordered in early EF-G crystal structures. Many mutations conferring resistance to fusidic acid, an antibiotic that freezes EF-G on the ribosome after GTP hydrolysis, are located at the interface between domain III and the rest of EF-G, suggesting that movement of domain III is important for EF-G function. EF-G–GTP binds to the ratcheted configuration of the ribosome in the pretranslocation state, with individual contacts occurring in an as yet unknown sequence. Like EF-Tu, EF-G undergoes a GTPase activation step.

After GTP hydrolysis and translocation, the ribosome returns to the nonratcheted, locked configuration with tRNAs in the classical binding states (P/P and E/E). A high-resolution X-ray crystal structure of this posttranslocation complex reveals many of the details of EF-G–ribosome interaction (Figure 4). Domain III simultaneously contacts the sarcin–ricin loop on the 50S subunit and ribosomal protein S12 on the 30S subunit. These contacts are probably essential, as EF-G lacking domain III has greatly decreased GTPase activity and is defective in translocation. Further, by interacting simultaneously with both subunits, domain III may assist EF-G in distinguishing between the locked and unlocked configurations of the ribosome.

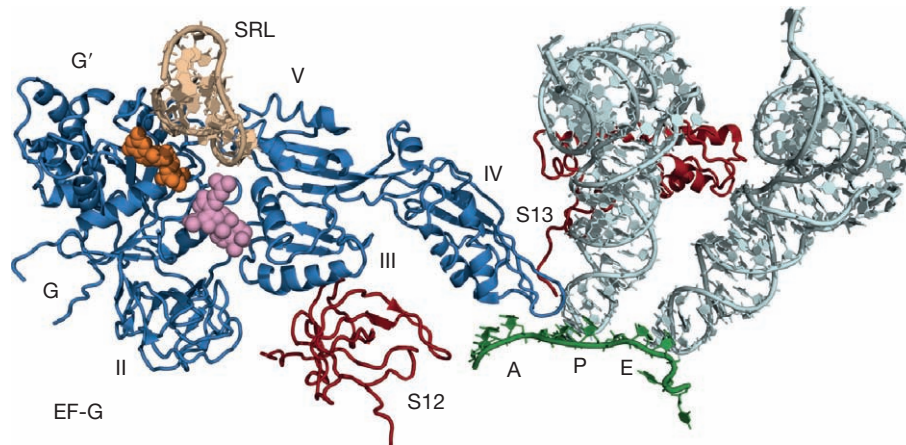


Figure 4 Details of EF-G–ribosome interactions after translocation. EF-G (blue) domain IV interacts with peptidyl-tRNA in the P site (cyan), while domain III interacts simultaneously with ribosomal protein S12 (red) and the sarcin-ricin loop (SRL) of 23S rRNA (tan). The C-terminal tail of ribosomal protein S13 (red) also interacts with the peptidyl-tRNA anticodon helix in the posttranslocation state (compare its position in [Figure 3](#)). Fusidic acid is shown as pink spheres, GDP as orange spheres, while mRNA and tRNAs are colored green and cyan, respectively. Most of the ribosome is omitted for clarity.

EF-G prevents reverse translocation primarily through its interactions with the 30S subunit. Domains III and IV, whose positions are spatially analogous to that of the aminoacyl-tRNA moiety of the EF-Tu–aminoacyl-tRNA–GTP ternary complex, sterically block backward movement of the tRNA–mRNA complex. Domain III interacts extensively with ribosomal protein S12, while domain IV interacts directly with both ribosomal protein S13 and peptidyl-tRNA in the P site. These interactions are accompanied by an altered conformation of the extended tail of S13, which itself now also interacts with the anticodon helix of peptidyl-tRNA ([Figure 4](#)). Mutations affecting ribosomal protein S12 or chemical modification of either S12 or S13, all stimulate spontaneous translocation, suggesting that EF-G stimulates translocation, in part, through its interaction with these proteins.

EF-G and Ribosome Recycling

In addition to its role in the elongation cycle, EF-G also participates in dissociation of the 70S ribosome into 30S and 50S subunits following translation termination. This is done in concert with ribosome recycling factor (RRF). RRF binds to the ribosome with deacylated tRNA in the P/E state. Following the action of EF-G, which requires hydrolysis of GTP and release of inorganic phosphate, the subunits dissociate and IF3 binds to the 30S subunit. The presence of IF3 prevents reassociation of the 30S and 50S subunits and promotes the release of deacylated tRNA and mRNA prior to the formation of a translation initiation complex. In mitochondria, two distinct EF-G proteins carry out translocation and recycling. Although complexes of 70S ribosomes and RRF have been solved by X-ray crystallography, posttermination complexes with both RRF and EF-G have not.

Conclusion

During the elongation phase of protein synthesis, tRNA selection and translocation are regulated by the action of two

GTPases that bind to the same region of the ribosome. In the case of tRNA selection, EF-Tu delivers aminoacyl-tRNA in a ternary complex with GTP to the A site of the ribosome. This factor kinetically divides tRNA selection into the two steps of initial recognition and proofreading by the irreversible hydrolysis of GTP. The use of two sequential steps results in an overall accuracy of decoding that is the product of the accuracy of the two individual steps. In the case of translocation, EF-G complexed with GTP stabilizes or induces an unlocked configuration of the ribosome. This unlocking entails a relative rotation of the 30S and 50S subunits, coincident with the formation of hybrid tRNA-binding states resulting from peptide bond formation. In the unlocked configuration, tRNAs and mRNA are capable of spontaneous movement through the ribosome. Although the spontaneous reaction is inherently reversible, the presence of EF-G in the ribosomal A-site prevents the backward movement of the tRNA–mRNA complex. Thus, EF-G acts as a Brownian ratchet to shift the equilibrium of the translocation reaction in the forward direction. It is, in part, through the use of these two protein factors that the ribosome achieves its remarkable accuracy and speed.

See also: [Molecular Biology: Ribosome Assembly; Ribosome Structure; Translation Elongation in Bacteria; Translation Initiation in Bacteria: Factors and Mechanisms; Translation Initiation in Eukaryotes: Factors and Mechanisms; tRNA Synthetases.](#)

Further Reading

- Agirrezabala X and Frank J (2009) Elongation in translation as a dynamic interaction among the ribosome, tRNA, and elongation factors EF-G and EF-Tu. *Quarterly Reviews of Biophysics* 42: 159–200.
- Frank J and Agrawal RK (2000) A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* 406: 318–322.
- Gao YG, Selmer M, Dunham CM, et al. (2009) The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science* 326: 694–699.
- Gromadski KB and Rodnina MV (2004) Streptomycin interferes with conformational coupling between codon recognition and GTPase activation on the ribosome. *Nature Structural and Molecular Biology* 11: 316–322.

- Horan LH and Noller HF (2007) Intersubunit movement is required for ribosomal translocation. *Proceedings of the National Academy of Sciences of the United States of America* 104: 4881–4885.
- Munro JB, Altman RB, Tung CS, et al. (2010) Spontaneous formation of the unlocked state of the ribosome is a multistep process. *Proceedings of the National Academy of Sciences of the United States of America* 107: 709–714.
- Ogle JM and Ramakrishnan V (2005) Structural insights into translational fidelity. *Annual Review of Biochemistry* 74: 129–177.
- Schmeing TM, Voorhees RM, Kelley AC, et al. (2009) The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. *Science* 326: 688–694.
- Wilden B, Savelsbergh A, Rodnina MV, and Wintermeyer W (2006) Role and timing of GTP binding and hydrolysis during EF-G-dependent tRNA translocation on the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* 103: 13670–13675.
- Zhang W, Dunkle JA, and Cate JH (2009) Structures of the ribosome in intermediate states of ratcheting. *Science* 325: 1014–1017.

Ribosome Structure

T M Schmeing, McGill University, Montreal, QC, Canada

© 2013 Elsevier Inc. All rights reserved.

Glossary

A site The ribosomal binding site for the incoming aminoacyl transfer RNA (tRNA), whose covalently bound amino acid will be the next addition to the growing peptide chain.

Decoding The multistep process in which the ribosome determines whether an incoming tRNA carrying the next amino acid to be added is correct or incorrect. If incorrect, the tRNA is rejected.

E site The ribosomal binding site for the exiting tRNA, which has already donated its amino acid to the peptide chain.

Messenger RNA (mRNA) A transient RNA copy of the DNA instructions coding for the protein to be synthesized, hence the genetic template for protein synthesis on the ribosome.

P site The ribosomal binding site for the peptidyl tRNA, which is covalently attached to the growing peptide.

Peptidyl transfer The chemical reaction between the free amino group of an amino acid and the esterified

carboxy terminus of a peptide, which results in the formation of a new peptide bond. The RNA-based ribosomal catalyst of this reaction is referred to as peptidyl transferase.

Ribosome The macromolecular complex of several large RNAs and about 50 small proteins that performs translation.

Ribozyme An enzyme which is made of RNA rather than protein.

Transfer RNA (tRNA) The substrate for the ribosome. One end of the L-shaped tRNA base pairs with mRNA, which allows ribosomal decoding, while the other end bears a covalently attached amino acid, which is added to the growing protein in the ribosome.

Translation The synthesis of protein, as determined by a messenger RNA template; it is catalyzed by ribosomes and ribosome-associated proteins, such as the initiation, elongation, and termination factors.

Translocation The movement of tRNAs and associated mRNA through the ribosome, from A site to P site, and from P site to E site.

Ribosome Function

Overview

In all organisms, template-directed protein synthesis is performed by the ribosome. The ribosome translates genetic information contained within its template, messenger RNA (mRNA), by joining amino acids carried by its substrates, transfer RNAs (tRNAs). The mRNA template is a linear sequence of nonoverlapping, three-nucleotide codons, each of which codes for a single amino acid to be added to the protein chain. There are 20 different kinds of amino acids, which become linked to tRNAs specific for each particular amino acid. The tRNAs are fairly large, L-shaped macromolecules with functionally important ends (Figure 1). One end, the aminoacyl end or 3' CCA terminus, is the site of covalent attachment of the amino acid. The other end features a three-nucleotide anticodon that forms base-pairing interactions with a complementary mRNA codon while in the ribosome.

The ribosome contains three distinct binding sites for substrate tRNAs, as well as a long path for mRNA (Figure 1(c)). The three tRNA-binding sites are named after the state of the tRNA that will be found at that site. The A site binds an incoming aminoacyl tRNA, the P site contains the peptidyl tRNA, and the E site contains an empty or exiting tRNA, which is deacylated. The ribosome also has binding sites for the many protein translation factors required for the rapid *in vivo* rates of protein synthesis.

Activities of the Ribosome

After the ribosome binds mRNA and initiator tRNA during translation initiation, new proteins are actually synthesized

during the elongation cycle of translation. The cycle begins with a peptidyl tRNA in the P site, and an empty A site. The elongation factor (EF)-Tu delivers a candidate aminoacyl tRNA to the A site, and the ribosome determines whether the candidate tRNA is a correct (or cognate) tRNA or one of the more numerous incorrect (noncognate) tRNAs. This gate-keeping function, known as 'decoding', is crucial for maintaining the accuracy of translation and thus of the genetic information contained in DNA.

Once an incoming aminoacyl tRNA has been accepted in the A site, its amino acid must be added to the peptide chain attached to the P-site tRNA. A new peptide bond is formed when the α -amino group of the aminoacyl-tRNA in the A site attacks the C-terminal carbonyl of the growing peptide attached to P-site tRNA. Peptide bond formation results in transfer of the peptide chain to the A-site tRNA and deacylation of P-site tRNA. Subsequent coordinated movement of the tRNA and mRNA relative to the ribosome (from right to left in Figure 1(c)), catalyzed by EF-G, restores a peptidyl tRNA in the P site and places a deacylated tRNA in the E site. During translocation, the ribosome must prevent slippage of the tRNA substrates relative to the mRNA template, which would alter the reading frame on mRNA and result in the synthesis of an incorrect and usually prematurely terminated protein. Translation terminates when a stop codon in the mRNA is reached. Stop codons are not recognized by tRNAs, but by protein release factors (RFs) which catalyze the hydrolysis of the peptidyl tRNA and release of the nascent protein.

Although EF-Tu and EF-G are required for the rapid *in vivo* rates of elongation, it is important to note that decoding, peptide bond formation, and translocation are all inherent

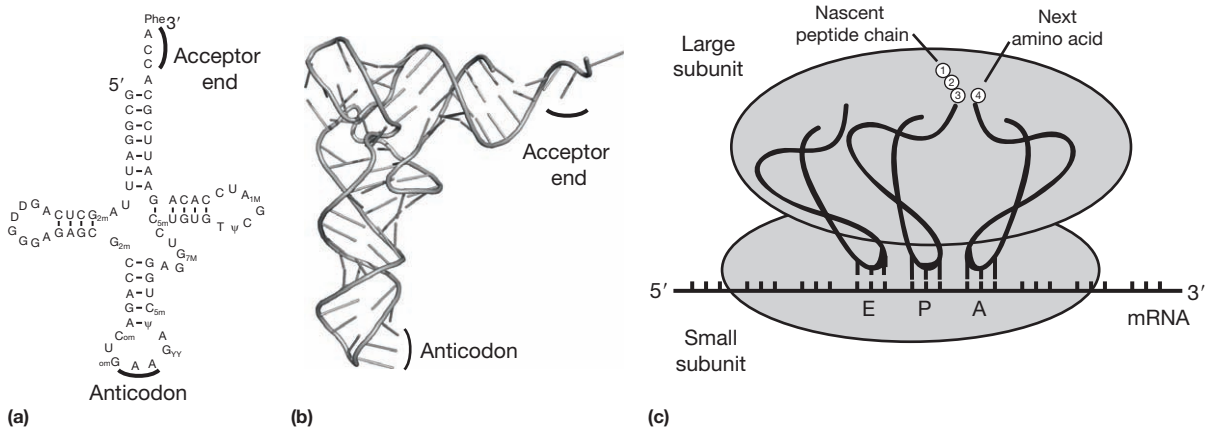


Figure 1 (a) Schematic view of a tRNA, the substrate of the ribosome. (b) Three-dimensional cartoon view of the structure of a tRNA. (c) Schematic view of a translating ribosome. Three amino acids (circles numbered 1–3) have already been joined to form the nascent peptide chain, whose C-terminal end is covalently attached to the tRNA in the P site. The next amino acid to be added (circle 4) is attached to a tRNA, which has just bound the ribosomal A site. A deacylated tRNA (left rectangle) is in the E site and will soon exit the ribosome. Vertical triplet bars represent codons in the mRNA, which base-pair to complementary anticodons in the tRNAs. The tRNA–mRNA complex moves through the ribosome from right to left, with each tRNA donating its covalently attached amino acid to the growing peptide chain as it passes through the peptidyl transferase site.

functions of the ribosome; they can be performed by the ribosome without the aid of EFs.

Ribosome Structure

Subunit Organization and Nomenclature

Ribosomes are large and complex enzymes: the simplest ribosomes from bacteria have a mass of some 2.5 million Da. All ribosomes consist of two loosely associated subunits of unequal size. In the bacterial ribosome, the large subunit is denoted the 50S, the small subunit 30S, and the entire ribosome 70S, according to their rates of sedimentation during ultracentrifugation. The bacterial small subunit consists of the 16S ribosomal (rRNA), which is approximately 1500 nucleotides, and around 20 small proteins (named S1, S2, etc.). The large subunit contains the 23S rRNA (~2900 residues), the 5S rRNA (~120 residues), and around 30 proteins (L1, L2, etc.). Eukaryotic ribosomes are even more massive, with the 60S large subunit (with 5S, 28S, and 5.8S rRNAs plus proteins) and the 40S small subunit (with 18S rRNAs plus proteins) making up the 80S ribosome.

In addition, eukaryotic cells contain mitochondrial ribosomes, and plant cells may also have chloroplast ribosomes.

In all ribosomes, the large subunit interacts with the CCA end of the tRNAs and contains the active site (the peptidyl transferase center (PTC)) which catalyzes peptide bond formation. The small subunit binds the anticodon end of the tRNAs, both monitoring (during decoding) and maintaining (during translocation) the tRNA–mRNA base-pairing interactions.

The Road to Ribosome Structure

The great size and complex composition of ribosomes were enormous barriers to efforts to determine the structural basis of ribosome function. The ribosome size is inconvenient for structural biology: it was, until recently, too large for X-ray

crystallography, but also rather too small for cellular electron microscopy (EM) methods. Many genetic, biochemical, and biophysical methods were developed over the years to probe ribosome structure, and divide and conquer strategies were able to determine high-resolution structures of small fragments of the ribosome, all of which yielded large amounts of useful data. However, the ultimate goal of revealing the chemical basis of protein synthesis would require visualizing intact ribosomes and ribosomal subunits at the high resolution possible.

Single particle EM led the way in the structural characterization of ribosomes. First, very low resolution images (~40 Å) from negatively stained EM collected in the 1970s showed the shapes and large structural features of the subunits. Then a big step forward occurred in 1995 when Joachim Frank and co-workers, using images of the ribosome suspended in vitreous ice (cryo-EM), visualized the ribosome at 25 Å. Shortly afterward, both the Frank and van Heel groups were able to identify the tRNAs in complex with the ribosome at similar resolution. Resolutions of EM reconstructions gradually improved, but it was X-ray crystallography that provided the decisive breakthroughs to atomic detail.

Crystals of ribosomes had been around for decades, the first being of a bacterial 50S reported by Wittmann, Yonath, and colleagues in 1980. In 1987, the Garber group in Puschino, Russia was the first to report 30S and 70S crystals. However, the technology and know-how to determine the structures of these large asymmetric molecules was lacking. By the late 1990s, this had all changed. The Steitz and Moore groups successfully calculated maps of the large subunit from *Haloarcula marismortui* at 9 Å resolution in 1998. By 1999, this was improved to 5 Å, and Ramakrishnan simultaneously reported a 5-Å structure of the small subunit from *Thermus thermophilus*. By the next year, atomic resolution was achieved for both crystal forms; these structures were completely built. The Yonath group also released a small subunit structure at the same time and at comparable resolution, and later determined the structure of a bacterial large ribosomal subunit. Suddenly, almost

all of the details of ribosomal architecture were clear, enabling many fundamental new insights into the structural basis for translation mechanisms.

The subunit structures were the culmination of a vast amount of work, but rather than capping off the field, they sparked a new dawn in ribosome study by structural, biochemical, single molecule, and biophysical techniques. In crystallography studies, these crystal forms of the ribosomal subunits were used to produce a myriad of co-crystal structures to study decoding, peptide bond formation, and antibiotic action. The subunit structures were also used to derive the phases required to solve structures of the 70S ribosome. The *T. thermophilus* 70S structure was determined to moderate resolution by the Noller and Yusopov group and then to high resolution by the Ramakrishnan group. The later 70S crystal form was subsequently used by Noller, Steitz, and Yusupov, while the Cate group concentrated on *Escherichia coli* ribosomes. Furthermore, the flexible areas of the ribosome which were not easily seen in some of these crystal structures, such as the L1 and L7/L12 stalks of the 50S subunit, have been determined in isolation to high resolution. More recently, structures of the ribosome have been solved in complex with translation factors, providing a major breakthrough in understanding how these work. Finally, the Ban and Yusupov groups achieved the next big step by determining crystal structures of the eukaryotic small subunit and 80S ribosome.

Although cryo-EM lost the race to atomic structures, it continues to be a relevant and useful tool for ribosome study, benefiting greatly from the X-ray structures to allow molecular interpretation of EM reconstructions. The achievable resolution has improved to near X-ray standards, and it is very powerful to visualize the many transient conformations required for ribosome function. In all, we now have a very detailed picture of the structure of the ribosome and how the ribosome changes to perform its many tasks.

General Features of Ribosomal Architecture

The overall architecture and composition of the ribosome and its functional centers reveal it to be essentially an RNA-based machine. The shape of both subunits is determined by the RNA

component: only one of the landmark features visible at low resolution (the L7/L12 stalk) is composed only of protein (Figure 2). Moreover, the overall distribution of the proteins is not uniform: the proteins are concentrated on the back and sides of the subunits, away from the most functionally important regions at the subunit interface. The protein component is also concentrated on the surface, rather than buried in the middle of the subunits; the ribosome thus literally has an RNA core. Many of the proteins contain both a globular component on the periphery of the subunits and one or more narrow extended peptide tails that penetrate deeply into crevices of the RNA core. Proteins are found to interact with RNA multistem junctions and tight bends in the RNA, which suggests particular functional roles in subunit assembly, in agreement with biochemical data.

It makes good sense that the ribosome is an RNA-based enzyme, whereas most other enzymes are made entirely of protein. After all, if proteins were made by a hypothetical ribosome composed entirely of protein, it would be hard to imagine how this protein-only ribosome could have evolved. An RNA-based ribosome solves this chicken-and-egg problem and fits well with speculations about an RNA world preceding the existence of large proteins and DNA.

There is one functionally important difference between the large-scale architectures of the 30S and 50S subunits. While the secondary structure domains of the 23S rRNA interact closely to form a single, nearly hemispherical mass, the 16S rRNA secondary structure domains are for the most part independent globular entities – true three-dimensional domains – that have relatively few interactions with each other. One domain of the 16S rRNA constitutes the head, another the platform, and another the bulk of the body. This architecture allows required conformational changes in the small subunit, such as rotation of the head which accompanies translocation. By contrast, the biggest movements of the large subunit are in peripheral helices (L1 stalk, Helix 69) or protein features (L7/L12 stalk).

In addition to information specifically about ribosome function, these structures provide an enormous boost to the understanding of RNA-protein, RNA-ion, and RNA tertiary interactions, since they vastly expand the database of known

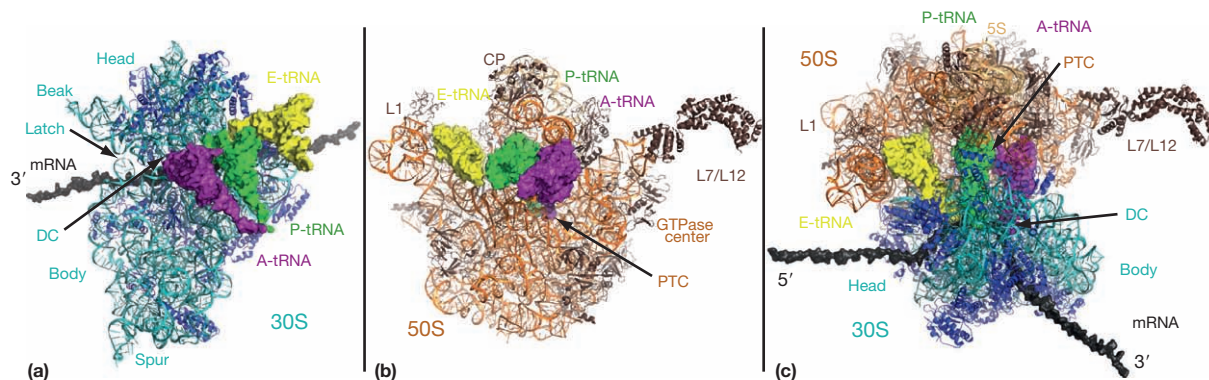


Figure 2 The three-dimensional structure of the ribosome. The 30S subunit, tRNAs and mRNA are shown in panel (a), the 50S subunit and the same tRNAs are shown in panel (b), and the 70S ribosome with tRNAs and mRNA is shown in panel (c). Landmarks are indicated in each panel. DC denotes decoding center, CP denotes central protuberance, and PTC denotes peptidyl transferase center. Reproduced from Schmeing TM and Ramakrishnan V (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461(7268): 1234–1242.

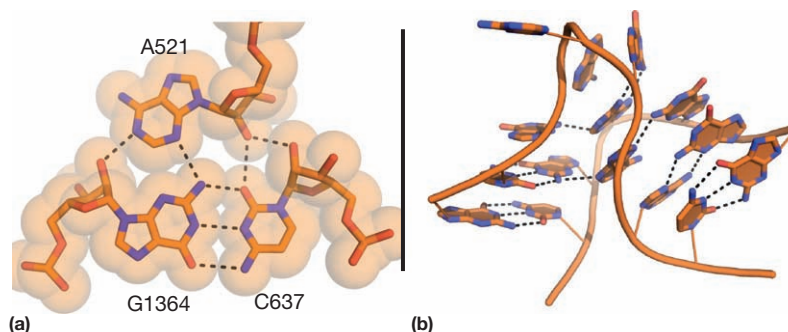


Figure 3 The kink turn (a) and the A-minor motif (b) are two RNA packing features seen repeatedly in the ribosome.

RNA structure. Two principles of RNA higher-order structure which were revealed are the A-minor motif and the kink turn (Figure 3). The A-minor motif is a common strategy for packing RNA helices, whereby patches of adenine residues pack their bases and 2' OH moieties against the minor groove of another RNA helix. Four different classes of this A-minor motif were defined, some of which had been previously observed in other structures. The kink turn is a common RNA structure of about 15 nucleotides which contains a kinked phosphodiester backbone and a sharp turn of an RNA helix. The kink turn is predictable from sequence information and is stabilized by interactions with proteins.

The Small Subunit and Decoding

A principal role played by the small subunit is the binding of mRNA and appropriate tRNAs (Figure 2(a)). The mRNA wraps around what would be the neck of the small subunit, between the head and the body. The mRNA pathway becomes a tunnel when noncovalent interactions between the head and shoulder domains form a so-called latch. The mRNA enters this tunnel behind the shoulder of the small subunit, and the nearby proteins S3 and S4 which form the entrance to the tunnel have been implicated in helping resolve secondary structure in the mRNA, thus imparting helicase function to the ribosome as part of translocation. The path then leads to the A site, responsible for decoding (see below), then to P and E sites. There is a turn in the mRNA pathway between the A and P sites which delineates the border of these sites, and the mRNA structure here is sharply kinked, the kink stabilized by a magnesium ion. The kink is thought to allow A and P site bound tRNAs to interact with adjacent nucleotides in the mRNA, prevent slippage of mRNA, and define the appropriate reading frame. Overlapping with and following on from E site, the mRNA passes the very 3' end of the 16S rRNA with which during the initiation phase of translation the Shine Delgarno sequence in the mRNA forms a helix. The mRNA then emerges from the 30S adjacent to the platform. Special functional sequences in the mRNA can interact with the ribosome near the platform to influence translation.

During elongation, incoming aminoacyl tRNAs bind the 30S A site, which is also called the 'decoding center'. The rate of protein synthesis is about 20 amino acids added per second, and there are substantially more incorrect (noncognate) aminoacyl-tRNAs present than correct (cognate) ones, so the ribosome must very quickly and robustly detect whether the

tRNA is appropriate to use. The 30S subunit plays a crucial role in tRNA selection by directly monitoring base-pairing interactions between the tRNA anticodon and the mRNA codon. In the presence of a cognate A-site tRNA, conformational changes occur in 16S rRNA residues A1492, A1493, and G530 (*E. coli* numbering), conserved elements of the head (helix 34) and body (the 530 loop), to sense the correctness of the codon-anticodon interaction. A1492, A1493, and G530 swing out and dock into the minor groove of the A-site codon-anticodon minihelix to test whether the geometry of the minihelix is correct (Figure 4). Monitoring geometry of the minihelix is an elegant way to determine whether the tRNA anticodon matches the codon, as it relies on the correctness of the base pair only, not of the sequence. The pairing is most stringently monitored at the first two positions of the codon, whereas the third wobble position has more flexibility. If the codon-anticodon matching is deemed correct, a more global conformational change occurs, moving the shoulder region of the 30S toward the decoding center. This larger-scale conformational change from an open to a closed form is involved in the signal for guanosine triphosphate (GTP) hydrolysis by EF-Tu, the EF that delivers the incoming candidate tRNAs to the A site (see below).

Many antibiotics (streptomycin, spectinomycin, tetracycline, hygromycin B, pactamycin, and edeine) bind to the 30S. Some, such as tetracycline, act by preventing binding of tRNA substrate. Interestingly, binding of error-inducing antibiotics of the aminoglycoside family causes some of the conformational changes seen in decoding, such as the swinging out of A1492-3. These aminoglycosides were thus proposed to induce miscoding by stabilizing a conformation very similar to the closed form adopted during normal translation.

The Large Subunit and Peptide Bond Formation

The large ribosomal subunit catalyses the key chemical event in protein synthesis, peptide bond formation. The catalytic active site is in the bottom of a deep cleft, open on one side to allow binding of tRNA substrates. The PTC leads to a long tunnel, through which the nascent peptide chain travels to leave the ribosome on the back side of the large subunit.

Although the ribosome is vast, all of the key elements required to catalyze the peptidyl transferase reaction are within ~ 15 Å of the reactive atoms. As with all enzymes, substrate positioning is vital for reaction. Precise substrate positioning is achieved when the proper binding of the 3' terminal

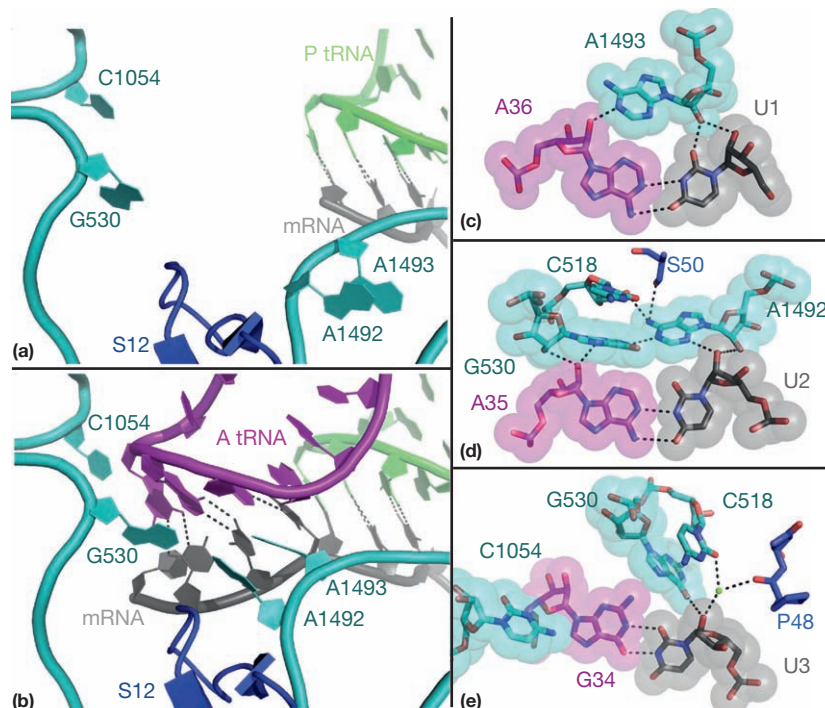


Figure 4 The decoding center of the 30S subunit. When the decoding center (a) is bound by a cognate tRNA (b) 16S residues A1492, A1493, and G530 rearrange to monitor the mRNA–tRNA interaction. (c–e) The specific interaction of the decoding center with the codon and anticodon at the three base pairs. Reproduced from Schmeing TM and Ramakrishnan V (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461(7268): 1234–1242.

CCA sequence of aminoacyl-tRNA to the A site of the PTC induces a reaction-competent conformation of the PTC and the peptidyl-tRNA, bound to the P site. The reactive α -amino group is oriented for nucleophilic attack by hydrogen bonds to the N3 of A2451 of 23S rRNA and the 2' OH of A76 of the peptidyl-tRNA. The ribosome does not appear to employ a metal ion or provide a near-neutral group for general acid–base catalysis. Rather, it now seems likely that the 2' hydroxyl of the peptidyl-tRNA's final nucleotide, A76, abstracts a proton from the alpha amino group of the aminoacyl-tRNA in a proton-shuttling mechanism (Figures 5(b) and 5(d)). In addition, the highly polarized transition state appears to be stabilized by a polar water molecule which is precisely positioned by the ribosome to act as a oxyanion hole. No ribosomal protein acts directly in this reaction, making the ribosome a ribozyme, but ribosomal proteins (L10e in archaea and L16 and L27 in bacteria) are nearby and help binding of tRNAs. Although this catalytic mechanism is supported by many studies, a caveat remains, as one set of experiments using cell extracts appears to show that a 2'-deoxy-A76 tRNA is competent for use in protein synthesis. The reasons for this discrepancy need to be clarified.

Although the ribosome is asymmetric, there exists a pseudo twofold symmetry at the heart of the PTC, relating the 50S A and P sites. This is likely a relic from when the proto-ribosome was formed by a duplication event of ~ 100 nucleotides of RNA, which first allowed bringing together of two similar substrates. The modern 23S rRNA results from subsequent asymmetric expansion of the nucleotide sequence through large insertion events.

The large ribosomal subunit is a common target for antibiotics. Although these antibiotics work in many different ways, two common themes seem to be prevention of tRNA binding (e.g., puromycin and sparsomycin) and interference with the nascent peptide chain in the exit tunnel (the macrocyclic family). Understanding the molecular interactions of these compounds with the ribosome is of great use in rational design of improved antibacterial drugs.

The bulk of the 50S is monolithic and moves very little during protein synthesis. Two large protuberances, the L1 arm and the L7/L12, do move. The L7/L12 arm is made of the L10 RNA-binding site, L10, and four to six copies of protein L12 (L7 is a modified form of L12). This huge feature waves around, probably to help recruit incoming tRNAs and translation factors to the A site. The L1 protein and the helix of 23S rRNA bind to perform the complementary function, changing conformation to facilitate release of the deacylated tRNA from the E site.

The 70S Ribosome and Translocation

Although codon recognition and peptide bond formation are fundamentally functions of the small and large subunits, respectively, the subunits work together in the context of the full 70S ribosome throughout elongation to perform protein synthesis. The subunits are held together by subunit bridges made by every possible combination of interaction: RNA–RNA, RNA–protein, protein–protein, water-mediated, and ion-mediated interactions. The tRNAs must travel through the ribosome during protein synthesis, and although this is

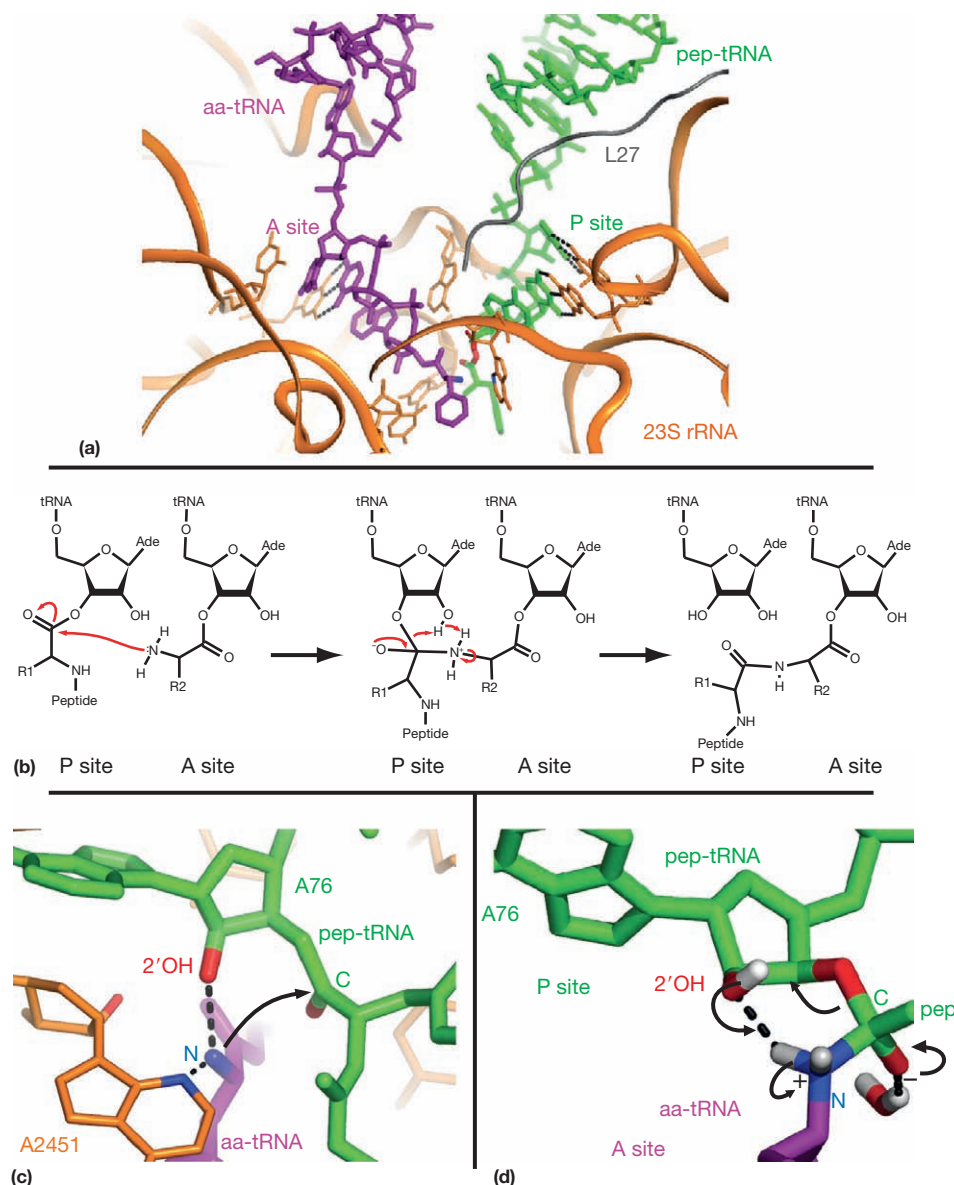


Figure 5 The peptidyl transferase center (a) and the reaction it catalyzes (b). The mechanism likely involves the 2' hydroxyl group of the peptidyl-tRNA both orienting the nucleophilic alpha amino group (c) and accepting a proton during the breakdown of the reaction intermediate (d). Reproduced from Schmeing TM, Huang KS, Kitchen DE, Strobel SA, and Steitz TA (2005) Structural insights into the roles of the 2' hydroxyl of the peptidyl-tRNA and water in the peptidyl transferase reaction. *Molecular Cell* 20(3): 437–448.

catalyzed *in vivo* by EF-G, translocation is an intrinsic function of the ribosome. After peptide bond formation, the first step of translocation is a movement of the 3' ends of the tRNA from A to P and P to E sites of the 50S subunit, independent of the stationary anticodon ends. This step can initially occur with little change in the ribosome conformation. In the next step, the relative orientation of the subunits changes markedly, as a ~6–8 degree rotation or ratcheting motion occurs and the head domain of the 30S subunit moves substantially. To complete translocation, the anticodon end of the tRNAs and the mRNA shift in the 30S subunit, while the ribosome returns to the un-ratcheted state. These ratcheting and unratcheting movements necessitate shifting, rearrangement, or breaking and reforming of many of the intersubunit bridges, and the

strengths of the interactions at the start, intermediate, and end states of ratcheting must be exquisitely balanced to allow ratcheting when required, while also holding subunits together. Furthermore, the three tRNA substrates are intimately cradled by subunit bridges in the intersubunit space in such a way that relative movement of the subunits can effect translocation of the tRNAs. Fully defining the steps of translocation is a very active area of research, with cryo-EM and single particle work leading the way.

The Ribosome and Translation Factors

High-resolution crystal structures of the ribosome in complex with the translation factors required to catalyze the *in vivo* rates

of protein synthesis were the next frontier after the subunit and 70S structures. Some of these proteins, despite being quite large, fit smoothly into the crystal forms of the unliganded ribosome. This allowed determination of co-crystal structure with initiation factor 1, EF-P, RF I and II, and ribosome recycling factor. Of these, the structures provided greatest insight into the function of EF-P and RFI and II.

EF-P is a translation factor that, although not required in a minimal cell-free translation system, promotes peptide bond formation and is conserved in all kingdoms of life. The crystal structure of the 70S with EF-P and initiator tRNA in the P site shows that EF-P binds to an area of the ribosome between the classical P and E sites and shifts the conformation of the L1 arm (**Figure 6(a)**). EF-P is nestled up against the P-site tRNA and makes many interactions with it, including close to the aminoacyl end. However, the conformation of the P-site tRNA is the same in absence and presence of EF-P, and thus the role of EF-P is simply to stabilize the tRNA in a conformation competent to participate in peptide bond formation. It is not clear whether EF-P is used to aid the formation of just the first peptide bond, or of each peptide bond made.

At termination, class I RFs (RFI and RFII) bind to ribosomes which contain stop codons in the 30S A site and catalyze peptide release at the PTC (**Figure 6(a)**). A series of crystal structures shows how the RFs use specific interactions between the protein and the mRNA codon to distinguish stop codons from sense codons. In the 50S, each RF inserts a

conserved glycine–glycine–glutamine sequence into the PTC to hydrolyze the nascent peptide, perhaps through coordination of the nucleophilic water and/or product stabilization. Computational simulations using these crystal structures provide complementary insight into the release mechanism.

Many of the essential EFs, such as IF2, EF-Tu, EF-G, and RF3, are GTP-hydrolyzing enzymes, which have a conserved guanosine triphosphatase (GTPase) domain and all bind to the same location on the ribosome, called the GTPase factor binding site, which is adjacent to the A site. In all the 70S crystals reported prior to 2009, the ribosomal protein L9 from an adjacent ribosome in the crystal protruded from that ribosome and blocked the GTPase factor binding site (not shown in these figures). The Ramakrishnan group genetically removed L9 from the *T. thermophilus* ribosome and were then able to find several new crystal forms that allowed co-complexes of the ribosome with GTPase factors EF-G and EF-Tu.

EF-G is a large, 78-kDa protein that catalyzes translocation. EF-G is seen bound at the factor-binding site, interacting with both the 30S and 50S including the L7/L12 stalk. Its G-domain is bound near the Sarcin Ricin Loop of 23S rRNA, and its domain IV inserts into the decoding center (**Figure 6(c)**). There, domain IV makes critical interaction with both the mRNA codon and the tRNA in the P site which serve to ensure that the mRNA and tRNA move together in translocation.

EF-Tu is the GTPase factor that delivers incoming aminoacyl-tRNA to the A site. When EF-Tu and aminoacyl tRNA are

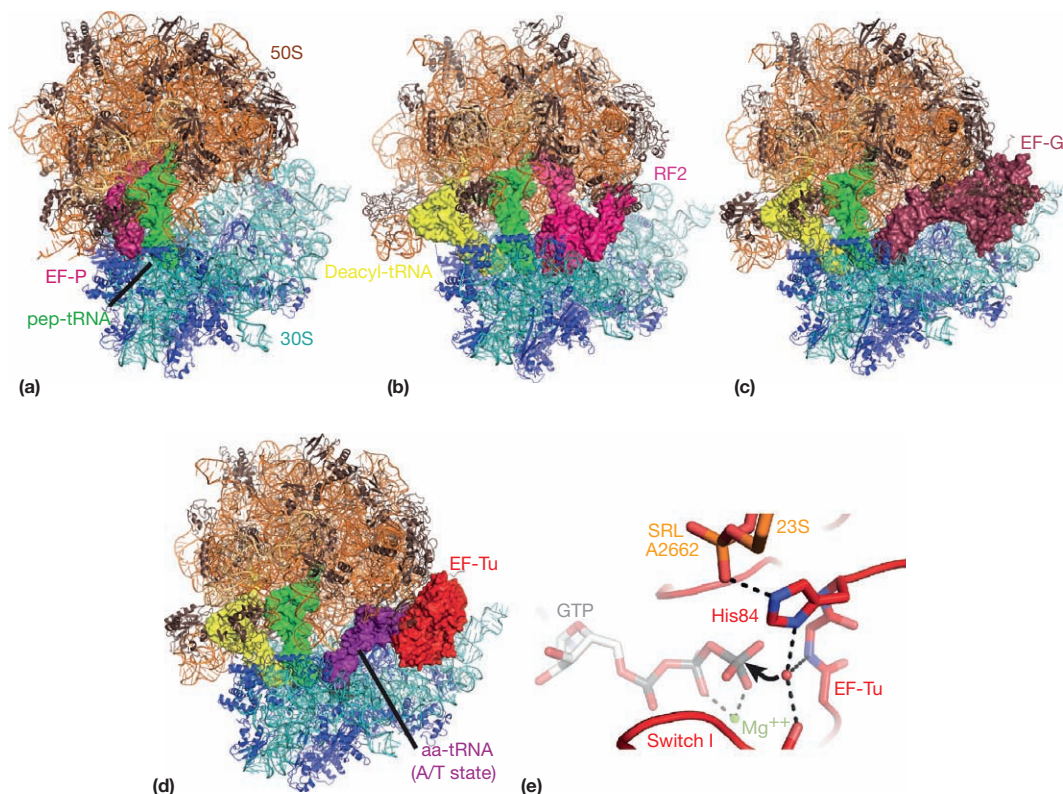


Figure 6 The 70S ribosome in complex with translation factors EF-P (a), RFII (b), EF-G (c), and EF-Tu (d). (e) GTP hydrolysis by EF-Tu is facilitated by the phosphate of 23S rRNA residue A2662, part of the sarcin ricin loop, helping His84 of EF-Tu to orient the reactive water molecule. (d, e) Reproduced from Schmeing TM, Voorhees RM, Kelley AC, et al. (2009) The crystal structure of the ribosome bound to EF-Tu and tRNA. *Science* 326(5953): 688–694, with permission from AAAS.

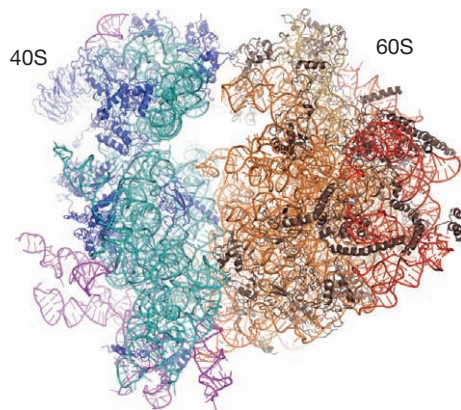


Figure 7 The 80S ribosome from Baker's yeast. The extensions of rRNA specific to eukaryotes are shown in purple for the small subunit and red for the large subunit. Reproduced from Ben-Shem A, Jenner L, Yusupova G, and Yusupov M (2010) Crystal structure of the eukaryotic ribosome. *Science* 330(6008): 1203–1209, with permission from AAAS.

bound to the ribosome, the body of the aminoacyl tRNA is distorted, so that it can interact with both the decoding center and the factor-binding site, where EF-Tu sits (**Figure 6(d)**). This is necessary so that the anticodon–codon matching can be tested before GTP hydrolysis leads to release of EF-Tu and acceptance of tRNA. Cognate tRNA binding produces the changes at the decoding center and the movement of the shoulder domain (see above, The Small Subunit and Decoding), which, along with conformational changes in the tRNA and EF-Tu transfer the signal that a proper codon–anticodon match has been made to the GTPase center of EF-Tu. The GTPase center is interacting with the 23S rRNA, and a phosphate group of the 23S rRNA positions histidine 84 of EF-Tu to coordinate a water molecule for nucleophilic attack of GTP (**Figure 6(e)**). After GTP hydrolysis and Pi release, EF-Tu leaves the ribosome, and aminoacyl-tRNA can be rejected by a proofreading mechanism or accepted into the A site of the PTC.

Structures of Eukaryotic Ribosomes

The most recent breakthroughs in ribosome structure are crystal structures of the eukaryotic small subunit in complex with initiation factors from the Ban group and a structure of the yeast 80S ribosome from the Yusupov group. Eukaryotic ribosomes are significantly larger than prokaryotic ribosomes, possessing more proteins and expansion segments that enlarge the rRNA. The 80S structure, solved at 4.2 Å resolution, shows that the extra mass of the 80S is located at the periphery of both subunits, while the core which contains the active sites required for the universal functions of the ribosomes is more

similar to prokaryotes (**Figure 7**). In addition, this 80S was determined in a ratcheted state (see above), which allows more insight into both conserved and eukaryotic specific subunit bridges. Furthermore, the central protuberance shows signs of mobility, which is less evident in prokaryotic structure. These structures lead the way for further eukaryotic structures to reveal mechanisms behind eukaryotic-specific events such as initiation and to exploit differences from the prokaryotic ribosome for rational antibiotic design.

Perspective

In the past 10 years, there has been a veritable boom in high-resolution structural information on the ribosome. The structures highlighted here, combined with vast amounts of critical biochemical, biophysical single particle, and mutagenesis work as well as important cryo-EM results, have fantastically improved our understanding of translation. As always, there are more questions to be answered, including those related to initiation and translocation complexes and the differences between prokaryotic and eukaryotic protein synthesis. Even so, it is amazing how much our knowledge of the ribosome continues to increase.

See also: **Molecular Biology:** Nucleolus, Overview; Ribosome Assembly; Translation Elongation in Bacteria; Translation Initiation in Bacteria: Factors and Mechanisms.

Further Reading

- Aitken CE, Petrov A, and Puglisi JD (2010) Single ribosome dynamics and the mechanism of translation. *Annual Review of Biophysics* 39: 491–513.
- Blanchard SC, Cooperman BS, and Wilson DN (2010) Probing translation with small-molecule inhibitors. *Chemistry and Biology* 17(6): 633–645.
- Dunkle JA and Cate JH (2010) Ribosome structure and dynamics during translocation and termination. *Annual Review of Biophysics* 39: 227–244.
- Frank J and Gonzalez RL (2010) Structure and dynamics of a processive Brownian motor: The translating ribosome. *Annual Review of Biochemistry* 79: 381–412.
- Korostelev A and Noller HF (2007) The ribosome in focus: New structures bring new insights. *Trends in Biochemical Sciences* 32(9): 434–441.
- Moore PB (2009) The ribosome returned. *Journal of Biology* 8(1): 8.
- Ramakrishnan V (2010) Unraveling the structure of the ribosome (Nobel Lecture). *Angewandte Chemie International Edition in English* 49(26): 4355–4380.
- Rodnina MV, Beringer M, and Wintermeyer W (2007) How ribosomes make peptide bonds. *Trends in Biochemical Sciences* 32(1): 20–26.
- Schmeing TM and Ramakrishnan V (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461(7268): 1234–1242.
- Steitz TA (2010) From the structure and function of the ribosome to new antibiotics (Nobel Lecture). *Angewandte Chemie International Edition in English* 49(26): 4381–4398.
- Yonath A (2010) Polar bears, antibiotics, and the evolving ribosome (Nobel Lecture). *Angewandte Chemie International Edition in English* 49(26): 4341–4354.

Riboswitches

A R Ferré-D'Amaré, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA

Published by Elsevier Inc.

Glossary

Aptamer Nucleic acid (DNA or RNA) molecule evolved artificially to bind to another molecule (nucleic acid, protein, or small molecule) with high affinity and specificity.

Kinetic control Said of a chemical or biochemical phenomenon whose outcome results from the relative rates of competing processes that underlie it (cf. thermodynamic control).

Pseudoknot A nucleic acid structure that results from nucleotides in the loop of a stem-loop structure forming

Watson–Crick base pairs with nucleotides outside of the stem-loop.

Riboswitch A domain of a messenger RNA (mRNA) that binds to a cellular small molecule with high specificity and regulates expression of the mRNA it is part of.

Ribozyme A catalytic RNA molecule.

Thermodynamic control Said of a chemical or biochemical phenomenon whose outcome results from the relative stabilities of the end results of competing processes that underlie it (cf. kinetic control).

Discovery

Genetic Regulation by Metabolite-Binding RNAs

In 1961, François Jacob and Jacques Monod suggested that the expression of genes that encode proteins could be regulated by a macromolecule that senses the intracellular concentration of a small-molecule metabolite. They proposed the name 'repressor' for this regulatory molecule, which would exert its function by binding to a specific nucleic acid sequence called the 'operator'. Although, at that time, Jacob and Monod speculated that the repressor could be either a protein or RNA, it was eventually shown that the particular repressor these researchers were investigating (the *lac* repressor) is a protein that binds to its cognate DNA sequence in the absence of allolactose. Nonetheless, this early postulation of an RNA genetic regulator that directly binds to small molecules presaged the discovery of riboswitches.

Antecedents

Several developments set the stage for the discovery of riboswitches. First, the invention of the systematic evolution of ligands by exponential enrichment (SELEX) methodology for the *in vitro* evolution of RNA led to the discovery of artificial RNAs (named 'aptamers') that are capable of binding other molecules with high affinity and specificity. Aptamers have been selected that recognize a variety of ligands, ranging from simple metabolites such as amino acids and nucleobases to complex secondary metabolites and even proteins. These aptamer selection experiments demonstrated that RNA has the capability of folding into structures that can recognize ligands with exquisite selectivity. Moreover, the results of these experiments also imply that such RNA sequences are relatively common in sequence space, because typical *in vitro* selection experiments can only sample a minuscule portion of the total number of possible RNA sequences of a given length (for instance, $\sim 10^{14}$ out of a total of $\sim 10^{60}$ possible sequences for a 100 nucleotide RNA). Second, a variety of mechanisms

for the control of gene expression that take place in *cis* at the messenger RNA (mRNA) level were discovered. Examples of these include transcriptional attenuation, where the efficiency of mRNA production is modulated by an mRNA-binding protein that is sensitive to the intracellular concentration of a metabolite, and translational attenuation, in which the efficiency of ribosome binding to an mRNA is controlled by a small-molecule-binding protein that binds to a regulatory domain of the mRNA. Third, an example of genetic regulation by attenuation that does not rely on a protein sensor was discovered. It was found that the 5'-untranslated region (UTR) of mRNAs encoding aminoacyl-tRNA synthetases (ARSs) in a number of Gram-positive bacteria can directly bind to the cognate transfer RNA (tRNA) of the ARSs they encode. These UTRs, characterized by a sequence and structural motif called the 'T box,' recognize both, the identity of the tRNA (through base pairing with the anti-codon sequence of the tRNA) and whether or not the tRNA is charged (aminoacylated). The T-box element functions as a genetic switch, promoting production of the ARS encoded in the mRNA of which they are part, when the concentration of uncharged cognate tRNA rises, and suppressing production of the ARS when aminoacylated tRNA becomes abundant.

Defining Characteristics of Riboswitches

Riboswitches were discovered in the first years of the twenty-first century, and embody aspects of RNA function that first became apparent from the three lines of research summarized above. First, riboswitches are RNA domains that bind to small molecule metabolites with high affinity and specificity. Although no natural riboswitch has been discovered to date that employs the same three-dimensional structure to bind to a small molecule as that used by an *in vitro*-evolved RNA, the affinity and selectivity of riboswitches for their cognate small molecules rival, and typically surpass, those of man-made aptamers that bind to the same ligands. Second, riboswitches function in *cis* at the mRNA level. These are domains of mRNAs

that modulate the expression of the mRNAs of which they are part. In this, they differ from the repressor of Jacob and Monod, since this latter type of gene regulation occurs in *trans* (i.e., regulation is effected by a molecule that is not covalently part of the mRNA), but riboswitch function is analogous to those of mRNAs regulated by attenuation. Third, like the T box, riboswitches are mRNA domains that can sense the intracellular concentration of a molecule directly, without employing a protein. Unlike the T box, which recognizes a macromolecule (tRNA), riboswitches respond to the concentration of small-molecule metabolites. This is an arbitrary distinction.

Heuristics and Their Consequences

All known riboswitches have sequence (and structural) elements that are highly conserved between multiple species. Some riboswitches are encoded in more than one genomic locus in an organism (i.e., they control multiple distinct mRNAs encoded by different genes or operons), and in such instances, a high level of conservation is also present within a species. Indeed, the majority of known riboswitches has been discovered by computational alignment of the sequences of noncoding mRNA elements (such as UTRs or introns). In some instances, previous genetic and biochemical experiments had already indicated that the aligned UTRs were responsible for regulation of the expression of the mRNA in response to the concentration of a particular metabolite. In other cases, the ability of the conserved mRNA element to regulate gene expression is unknown and has to be determined after discovery of the element. In either case, the identity of the physiologic ligand needs to be established experimentally and often involves educated guesses based on the nature of the proteins encoded by the mRNAs of which the putative riboswitch is part. These heuristics may have biased the types of riboswitches discovered to date, in that most known riboswitches are bacterial, and regulate the metabolism or transport of primary metabolites. Discovery of highly conserved noncoding mRNA elements is easier in bacteria, in which UTRs are rarely longer than a few hundred nucleotides. By contrast, UTRs and introns can be much longer in eukaryotic mRNAs. The number of sequenced bacterial genomes is also much larger than those of eukaryotes. In addition, the mRNAs of proteins involved in central metabolic pathways are much better annotated than those of proteins involved in other aspects of cellular function. Thus, the nature of the potential ligands for a newly discovered conserved UTR is more easily established for mRNAs encoding proteins that participate in primary metabolism. Finally, the reliance on sequence conservation as a heuristic necessarily excludes the discovery of riboswitches that are present in a single or only a few species.

Riboswitch Structures and Classes

Domain Structure

Riboswitches are not freestanding molecules; rather, they are functional domains embedded within mRNAs. Therefore, the first step in classifying them is to define their domain

borders. Biochemical analyses indicate that the borders of functional riboswitches extend beyond the segments with high phylogenetic conservation. Specifically, while the segments with high sequence conservation are often sufficient for recognition of the cognate small molecule, genetic regulation requires segments that are typically 3' of the conserved core. By analogy to *in vitro* selected RNAs, the conserved core is referred to as the 'aptamer domain'. Many riboswitch aptamer domains have been found to be autonomously folding and functional units, that is, these are RNA domains in the structural sense. The less conserved segments that are needed for regulation of gene expression are thought to transduce the small molecule-binding event to the gene expression machinery. These segments are collectively known as the 'expression platform'.

Riboswitch Ligands

Riboswitches can be classified based on their cognate ligands. The ligands of currently known riboswitches fall into five classes. First, riboswitches that respond to a variety of coenzymes and enzyme cofactors, including thiamine pyrophosphate (TPP), flavin mononucleotide (FMN), adenosylcobalamin, S-adenosylmethionine (SAM), and the molybdenum and tungsten cofactors Moco and Tuco, are known. Second, several riboswitches recognize purines and purine nucleosides, including adenine, guanine, preQ₁ (7-deaza-7-aminomethylguanine), and 2'-deoxyguanosine. A riboswitch that binds the bacterial second messenger cyclic diguanylate has also been discovered. Third, riboswitches that recognize amino acids include the glycine and lysine riboswitches. Fourth, one riboswitch that recognizes a carbohydrate, glucosamine-6-phosphate (GlcN6P), has been described. Fifth, riboswitches sensitive to the intracellular concentration of divalent cations have been described.

Aptamer Domain Structures

Classification of riboswitches based on their cognate ligands is unsatisfactory because several ligands are recognized by more than one type of riboswitch. For instance, five different types of riboswitches have been described that recognize SAM. Crystal structures have been determined of the aptamer domains of SAM riboswitches belonging to three of the five different types, and these demonstrated that not only are the folds of the RNAs different, but the interactions that the three types of RNAs employ to recognize SAM are distinctly different. Is a structural classification of riboswitches more meaningful? As of this writing, crystal structures of the ligand-bound conformations of a dozen riboswitch aptamer domains have been determined. Except for those riboswitches that are members of the same family (i.e., TPP riboswitches from *Escherichia coli* and *Arabidopsis thaliana*, or the adenine and guanine riboswitches, which are very closely related in sequence and nearly identical in structure), the structures of different riboswitches do not suggest evolutionary relationships between them, and therefore do not form a basis for a natural classification.

Crystal structures do allow classification of riboswitches based on RNA architectural principles. For instance, riboswitch

aptamer domains can be broadly divided into those that are assembled around a multi-helical junction and those that employ pseudoknotting. An example of the former is the FMN riboswitch, which is comprised of six stem loops radiating from a central junction. Five of the stem loops fold back on each other making a series of complex loop–helix and loop–loop interactions that give rise to the folded structure, which is pseudo-two-fold symmetric. A single FMN molecule is bound asymmetrically as a magnesium ion chelate at the core of the RNA (Figure 1(a)). Pseudoknots are nucleic acid structures that form when nucleotides in the loop of a stem–loop base pair with nucleotides outside of the stem–loop. This has the result of producing two helices that are connected by three nonhelical segments (at least formally, since some of these connectors can be as short as a single phosphodiester bond, i.e., be of length zero). The SAM-II (the second of the SAM riboswitch types to be described) is an example of a riboswitch that is a compact pseudoknot, in which the two constituent helices stack coaxially. The SAM molecule is recognized at a discontinuity internal to one of the two helices (Figure 1(b)). The *glmS* riboswitch recognizes GlcN6P and is an example of an RNA comprised of several pseudoknots. This RNA consists of a core domain that is sufficient for GlcN6P binding and a peripheral domain that stabilizes the core (Figure 1(c)). The core domain is made up of three helices that are connected by four crossovers, forming a double pseudoknot. The peripheral domain is connected to the core by an additional pseudoknot (resembling the SAM-II riboswitch). One of the loops of this peripheral domain is a long stem loop, so that the peripheral pseudoknot in effect functions as a helical junction. These three RNAs are but examples of the

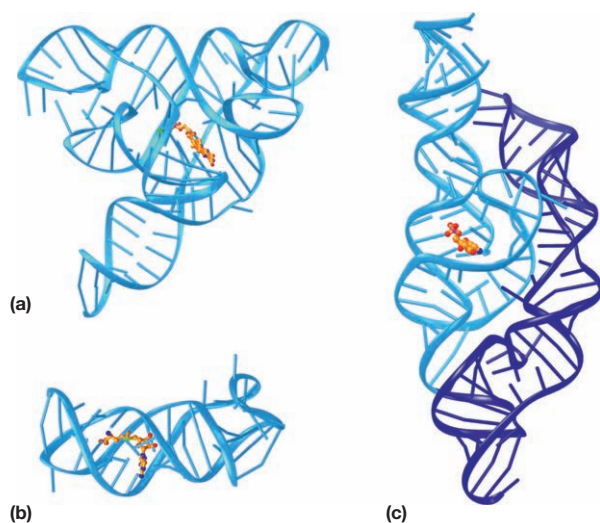


Figure 1 Examples of the three-dimensional structures of riboswitch aptamer domains. (a) The FMN riboswitch aptamer domain. RNA backbone is represented as a ribbon, and nucleobases as short cylinders. The bound FMN is as a ball-and-stick figure in contrasting colors, with the Mg^{2+} ion that chelates the phosphate of this ligand depicted as a green sphere. (b) The SAM-II riboswitch aptamer domain, represented as in the previous panel. (c) The *glmS* riboswitch represented as the previous panel with the core domain in cyan and the peripheral domain in dark blue.

diversity of structural solutions that nature has evolved to function as riboswitches.

Riboswitch Function

Small Molecule Recognition

The details of the molecular interactions employed by riboswitches to recognize their cognate small molecules are highly idiosyncratic. Nonetheless, the intrinsic chemical properties of RNA are reflected in two recurrent recognition strategies. First, planar heterocycles tend to be recognized in the same manner in which purine and pyrimidine nucleobases of an RNA pack in the interior of a molecule, that is, through a combination of stacking interactions and hydrogen bonding. This is most obvious in the case of riboswitches that bind to purines and purine nucleosides and nucleotides, and ligands such as SAM that contain a nucleotide moiety, but it extends to other ligands with different heterocycles. For instance, the thiamine moiety of TPP consists of a thiazole and an aminopyrimidine. The latter is recognized by the TPP riboswitch by making an interaction analogous to a sugar-edge base pair with a guanosine residue of the riboswitch, and by stacking it between two RNA nucleobases. The FMN riboswitch also recognizes its ligand in a base pair-like manner. In this case, one of the edges of the isoalloxazine ring of FMN makes an interaction analogous to a Watson–Crick base pair with an adenosine residue of the ribozyme. Second, negatively charged ligands are often recognized as cation chelates. A magnesium ion coordinates the phosphate of FMN bound to the riboswitch, and two magnesium ions chelate the pyrophosphate moiety of TPP bound to its cognate riboswitch. In both cases, the metal ions are partially solvated, and the inner-sphere-coordinated water molecules mediate many of the RNA–metal ion interactions. That is, the riboswitch recognizes neither the naked ligand nor the metal ion chelate, but the hydrated metal ion chelate. Another example of cation-mediated recognition is the lysine riboswitch, which employs a potassium ion to mediate interactions between the RNA and the carboxylate functional group of its ligand.

The importance of some of the riboswitch–small molecule interactions revealed by X-ray crystallography has been addressed experimentally. The adenine and guanine riboswitches are closely related structurally, and both recognize their ligands by completely enveloping them. Almost all the hydrogen-bonding groups of the bound purines are recognized by riboswitch functional groups. Mutagenesis experiments indicate that the specificity of the adenine and guanine riboswitches is controlled by a single pyrimidine residue at a position in the RNAs that makes a Watson–Crick-type interaction with the bound ligand. Thus, if the riboswitch contains a cytidine residue at this position, it will bind to guanine, and if the riboswitch contains a uridine at the same position, it will be specific for adenine. In other riboswitches, the specificity is not such a simple function of RNA sequence.

The importance of metal ion ligands for ligand recognition varies from riboswitch to riboswitch. The lysine riboswitch has a strong preference for potassium, since replacing this cation with sodium or magnesium ions reduces the affinity of the RNA for its ligand by 50- to 100-fold. *In vitro*, the TPP riboswitch will bind to its ligand when Mg^{2+} is replaced with

Mn^{2+} , or even Ca^{2+} or Ba^{2+} . Structure analysis of this RNA bound to thiamine monophosphate (TMP) demonstrated a striking degree of plasticity: when bound to TMP, which has one phosphate group less than the cognate ligand TPP, the riboswitch still employs two metal ions that make essentially the same contacts with the RNA as the two ions that chelate TPP. However, the RNA has to deform to bind to the smaller TMP using this strategy. Thermodynamic analysis shows that this accommodation results in considerably weaker binding of TMP ($K_d \sim 260$ nM) than TPP ($K_d \sim 8.4$ nM, both measured at 10 mM Mg^{2+} concentration).

The degree of burial of riboswitch ligands in the interior of the aptamer domain varies from case to case. The purine-responsive riboswitches almost completely envelop their ligands, whereas the c-di-GMP-responsive riboswitch buries the nucleobases of its ligand, but leaves parts of the phosphate-ribose backbone of c-di-GMP exposed to solvent. The TPP riboswitch buries the pyrophosphate and aminopyrimidine moieties at either end of its ligand, but makes very few interactions with the thiazole ring in the center of TPP. The affinity of riboswitches for their ligands is highly variable and presumably reflects the intracellular concentration range over which the riboswitch has to carry out its genetic control function. Regardless of the precise affinity, however, it appears that binding of riboswitches to their ligands is in general accompanied by folding of the RNA, at least at the local level. Thus, studies of riboswitch–ligand interactions by isothermal titration calorimetry show that binding is enthalpically driven. Since RNA folding is facilitated by divalent cations, analysis of binding experiments at higher concentrations of Mg^{2+} shows a reduction in the unfavorable entropy of binding. That is, preorganization of riboswitch aptamer domains facilitates ligand recognition. A particularly interesting elaboration of the mechanism of ligand binding is present in the glycine riboswitch. This regulatory RNA is comprised of two aptamer domains that can recognize glycine independently. However, in the full riboswitch, the two domains interact, resulting in positive cooperativity in glycine binding, with a Hill coefficient of ~ 1.6 .

Gene Regulation: an Example

The mechanisms by which ligand binding to a riboswitch results in alterations in gene expression, that is, the mode of action of the riboswitch expression platforms, are highly variable, but with the exception of the *glmS* riboswitch discussed separately below, most appear to function by adopting two mutually exclusive secondary structures as a result of ligand binding. The alternative structures adopted by the expression platform of the TPP riboswitch upstream of the *ThiM* gene in *E. coli*, which functions by translational attenuation, illustrate this (Figure 2). The aptamer domain lies upstream of the open reading frame (ORF), and it is transcribed first. As RNA polymerase transcribes the end of the aptamer domain, a decision is made based on the presence of sufficiently high concentrations of TPP. In the presence of TPP, the aptamer domain structure is stabilized, and this results in formation of a helix (or paired region) P1, between the initial and final segments of the aptamer domain. Stabilization of this helix P1 allows the subsequent transcript to fold into a long helix comprised of three coaxially stacked helices (P6, P7, and P8). Helix P8 is

formed by nucleotides that comprise the Shine–Delgarno element (required for ribosome binding and initiation of translation) and a complementary anti-Shine–Delgarno sequence. When the expression platform (which, in this case, includes sequences starting at the 3' side of P1 to the end of P6) is in this conformation, the mRNA is not translated. In the absence of sufficient TPP, the aptamer domain is not stabilized, and the nucleotides that form the 3' side of P1 in the TPP-bound form instead form a new helix (PX in Figure 2) by pairing with the anti-Shine–Delgarno nucleotides. This results in exposure of the Shine–Delgarno sequence, and translation of the mRNA.

Transcriptional and Translational Control

It is apparent from the example in Figure 2 that similar alternative folding strategies could be devised where the riboswitch allows translation to take place in the absence of cognate ligand, and indeed such examples of translation-on riboswitches exist in nature. Many bacterial riboswitches function as transcriptional switches. In those cases, the mutually exclusive secondary structures include a transcriptional terminator helix and a helix that sequesters one of the strands of the terminator (an anti-terminator helix). As in the case of riboswitches that function by translation attenuation, transcriptional riboswitches are also known that activate or repress transcription depending on ligand occupancy. Indeed, the TPP riboswitch has been found to function by all four mechanisms in bacteria, depending on its genetic context. The different examples of TPP riboswitch all share a highly conserved aptamer domain, but differ in their expression platforms.

Riboswitches and Alternative Splicing

Thus far, the TPP riboswitch is the only riboswitch discovered in eukaryotes. It has been found in introns of several genes that appear to encode proteins linked to TPP metabolism or transport in filamentous fungi, algae, and plants. In all cases examined, TPP riboswitches in eukaryotes appear to function by promoting alternative splicing as a function of TPP concentration. The specific consequences of alternative splicing vary from example to example. In some fungal genes, the TPP riboswitch is present in the first intron, and depending on whether this intron is spliced or not, the mRNA will encode small upstream open reading frames (uORFs) whose translation downregulates translation of the principal ORF encoded by the mRNA. In the case of some mRNAs from *A. thaliana*, the TPP riboswitch is present in the last intron of the pre-mRNA. When the intron is not spliced, a strong polyadenylation signal will be present close to the end of the ORF. When the intron is spliced, a weaker polyadenylation signal much further downstream of the ORF is employed. This weaker signal results in a less stable mRNA, possibly because of a process involving nonsense-mediated decay (NMD).

Kinetic versus Thermodynamic Control

Since riboswitches make gene control decisions at the mRNA level, it is possible, in principle, that they can function either by thermodynamic or by kinetic control. A riboswitch would be

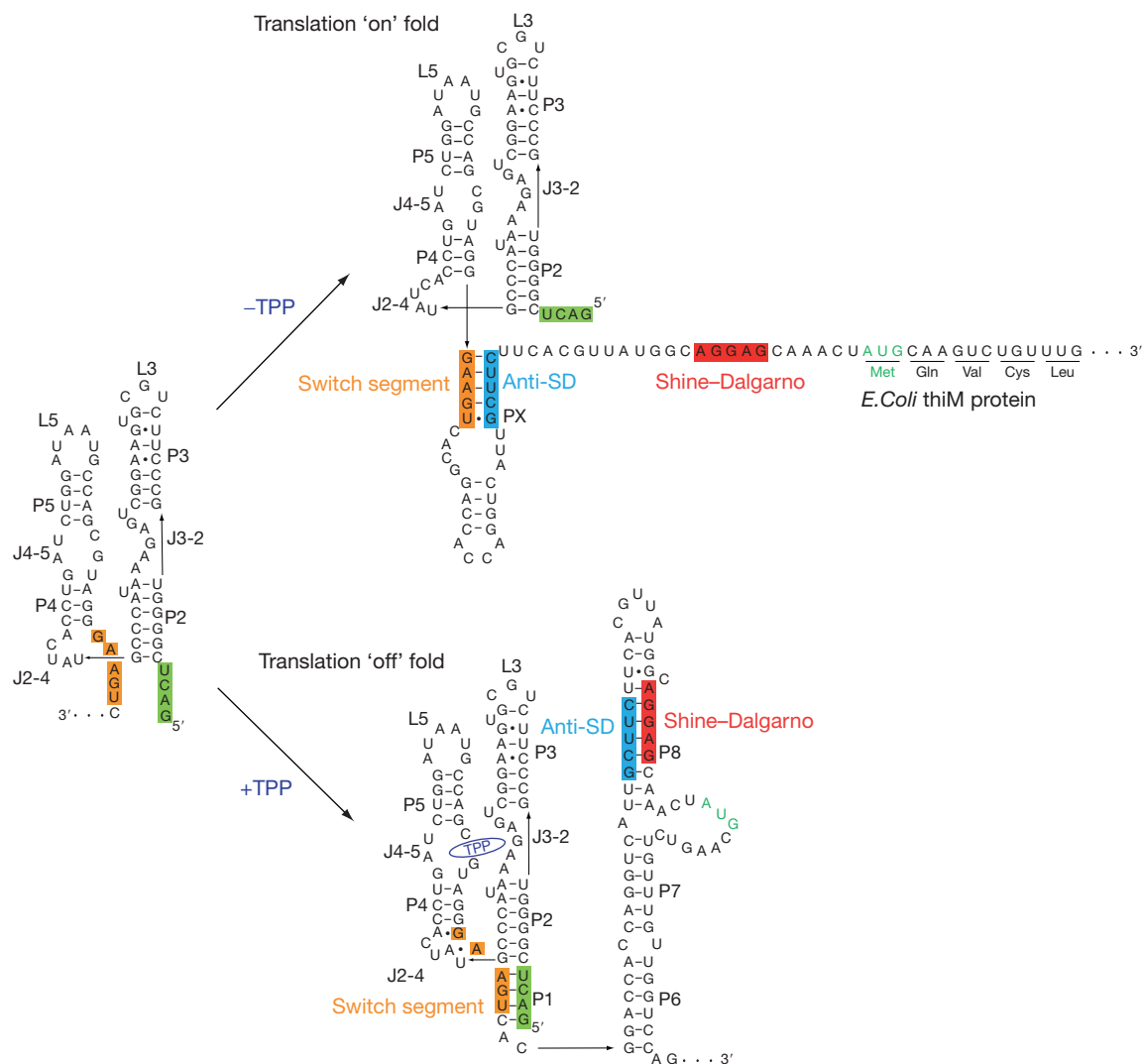


Figure 2 Schematic of the mechanism of translational control by two states of the TPP riboswitch from the *ThiM* gene of *E. coli*. Left, transcription of the mRNA generates the aptamer domain. In the presence of sufficient TPP (bottom) the aptamer domain is stabilized, and a helix that sequesters the Shine-Dalgarno sequence is formed. This results in downregulation of translation. In the absence of sufficient TPP to bind to the aptamer domain (top), the aptamer domain is not stable and the switch segment (orange) base pairs with the anti-Shine-Dalgarno sequence. This exposes the Shine-Dalgarno sequence allowing translation of the downstream ORF to take place.

under thermodynamic control if ligand binding and release are fast compared to the rate of mRNA synthesis, and therefore the ligand-bound and free forms of the riboswitch equilibrate quickly. Conversely, if the rate of mRNA synthesis is comparable to or faster than the rate of ligand binding, the switch decision occurs co-transcriptionally. Moreover, concentrations of ligand much higher than the equilibrium dissociation constant may be needed to throw the switch. Studies of the kinetics and thermodynamics of ligand binding by the FMN riboswitch demonstrated that the rate of ligand binding by the RNA is slow compared to the rate of transcription, such that unless the RNA polymerase is slowed down by encountering strategically positioned pause sites, the riboswitch cannot fold and bind FMN in time for its expression platform to adopt a secondary structure that is consistent with the liganded state of the aptamer domain.

Conformational Switching

Is ligand binding accompanied by a global folding transition of riboswitch aptamer domains, and not just a change in structure of the expression platform? That is, do ligand-free riboswitch aptamer domains adopt defined conformations, or are they partially or completely unfolded? Most structural studies of riboswitch aptamer domains have concentrated on their ligand-bound states. Recently, a number of studies of the ligand-free states of riboswitch aptamer domain have been reported. For instance, the TPP riboswitch has been examined by small-angle X-ray scattering (SAXS). In addition to providing global measurements of molecular size, SAXS can be used to generate low-resolution molecular envelopes. The molecular envelope of the ligand-bound state of the TPP riboswitch closely approximates the Y-shaped crystal structure of the

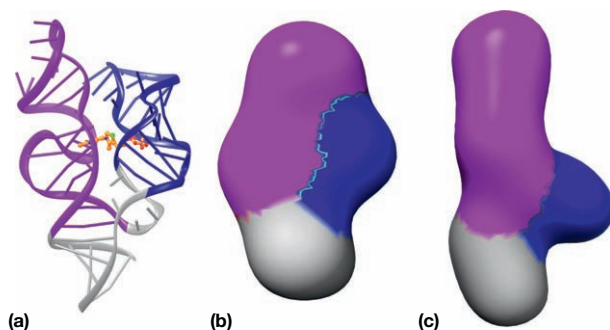


Figure 3 Conformational switch of the TPP riboswitch aptamer domain in solution. (a) Ribbon and stick representation of the crystal structure of the aptamer domain from the *ThiM* TPP riboswitch. (b) Low-resolution molecular envelope generated from SAXS data collected on the TPP-bound aptamer domain in solution. Portions of the reconstruction are colored to match the secondary structure elements in panel (a). (c) SAXS molecular envelope of the ligand-free state of the TPP riboswitch. The helix colored blue in panel (a) appears to have swung away from the red helix.

same state (Figures 3(a) and 3(b)). By contrast, the ligand-free state (which appears to be dominated by a single conformation, rather than a complex of states) is much more elongated than the crystal structure (Figure 3(c)). Thus, the TPP riboswitch aptamer domain exhibits distinct bound and free conformations. Other aptamer domains, such as those of the FMN and lysine riboswitches, appear to be pre-formed and do not exhibit global conformational changes concomitant with ligand binding.

The Special Case of the *GlmS* Riboswitch

The *glmS* riboswitch from Gram-positive bacteria functions by a mechanism that is thus far unique. This riboswitch resides in the 5'-UTR of the mRNA encoding the enzyme GlcN6P synthetase (*GlmS*). Binding of GlcN6P (its cognate ligand) to the riboswitch uncovers a latent self-cleavage activity of the RNA. In Gram-positive bacteria, intact mRNAs are capped by a 5' triphosphate. Activation of the *glmS* ribozyme by high levels of GlcN6P releases a short leader sequence that bears the triphosphate and exposes a new 5'-OH group in the ribozyme domain. This hydroxyl group is recognized by RNase J1, a ribonuclease that degrades the mRNA. Because the *GlmS* protein is unstable, degradation of the its mRNA results in downregulation of the enzymatic activity, thus completing a negative-feedback loop. The physiologic importance of the *glmS* ribozyme has been demonstrated by employing ribozyme-riboswitch mutants that lack GlcN6P-induced self-cleavage activity. *Bacillus subtilis* strains harboring *glmS* genes downstream of such catalytically defective riboswitches fail to sporulate under conditions that lead to sporulation of wild-type bacteria. Presumably, the differentiation program leading to sporulation requires precise regulation of GlcN6P levels.

Riboswitches as Drug Targets

Riboswitches are potential targets for the development of new antibacterial compounds for three principal reasons. First, the

currently known riboswitches are primarily bacterial. Thus, if riboswitches can be targeted specifically, the potential for undesired off-target interactions with human molecules is lessened. Second, riboswitches have evolved to bind small molecules with high affinity and specificity. Thus, they demonstrate that RNA (or at least some RNAs) can achieve highly specific binding. Third, riboswitches regulate essential metabolic pathways. One riboswitch is known that binds to a bacteria-specific second messenger. Thus, alteration of riboswitch function may significantly reduce the fitness of a bacterial pathogen. Indeed, several compounds whose anti-bacterial activity had been identified before the discovery of riboswitches have been shown subsequently to function by binding to riboswitches, and mutations in those riboswitches have been shown to diminish the effectiveness of these compounds.

See also: **Metabolism Vitamins and Hormones:** Roles of Micro-RNAs in Metabolism; **Molecular Biology:** Alternative Splicing; Gene Expression in Bacterial Systems: The *trp* Operon and Attenuation; Messenger RNA Degradation in Bacteria; Ribozymes and Evolution; Small RNAs in Bacteria; Transcription Termination; tRNA Synthetases.

Further Reading

- Baird NJ and Ferré-D'Amaré AR (2010) Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis. *RNA* 16: 598–609.
- Batey RT, Gilbert SD, and Montange RK (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* 432: 411–415.
- Blount KF and Breaker RR (2006) Riboswitches as antibacterial drug targets. *Nature Biotechnology* 24: 1558–1564.
- Collins JA, Irnov I, Baker S, and Winkler WC (2007) Mechanism of mRNA destabilization by the *glmS* ribozyme. *Genes and Development* 21: 3356–3368.
- Edwards TE, Klein DJ, and Ferré-D'Amaré AR (2007) Riboswitches: Small-molecule recognition by gene regulatory RNAs. *Current Opinion in Structural Biology* 17: 273–279.
- Grundy FJ and Henkin TM (1993) tRNA as a positive regulator of transcription antitermination in *B. subtilis*. *Cell* 74: 475–482.
- Jacob F and Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3: 318–356.
- Klein DJ and Ferré-D'Amaré AR (2006) Structural basis of *glmS* ribozyme activation by glucosamine-6-phosphate. *Science* 313: 1752–1756.
- Mandal M, Lee M, Barrick JE, et al. (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* 306: 275–279.
- Mironov AS, Gusarov I, Rafikov R, et al. (2002) Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria. *Cell* 111: 747–756.
- Nudler E and Mironov AS (2004) The riboswitch control of bacterial metabolism. *Trends in Biochemical Sciences* 29: 11–17.
- Serganov A (ed.) (2009) *Riboswitches Methods and Protocols*. New York, NY: Humana.
- Sudarsan N, Lee ER, Weinberg Z, et al. (2008) Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* 321: 411–413.
- Wickiser JK, Winkler WC, Breaker RR, and Crothers DM (2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Molecular Cell* 18: 49–60.
- Winkler W, Nahvi A, and Breaker RR (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419: 952–956.

Ribozymes and Evolution

N Lehman, Portland State University, Portland, OR, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Abiotic Formed in the absence of life.

Nucleotide A chemical subunit of nucleic acids (A, adenosine; C, cytidine; G, guanosine; U, uridine).

Phylogeny A tree-like representation of the ancestral-descendant relationships among species.

Polypeptide A polymer of amino acids that forms a protein.

Ribozyme An enzyme made of RNA.

RNA Ribonucleic acid, a polymer of nucleotide subunits.

Ribozymes, also known as ribonucleic acid (RNA) enzymes, have been known since the early 1980s when it was first discovered that RNA could catalyze chemical reactions. Previously, it had been thought that all enzymes were proteins, but it is now clear that RNA-directed catalysis plays many important cellular functions, and that numerous ribozymes can be found in organisms spanning the tree of life. Notably, for example, we now realize that the catalytic core of the ribosome itself, the cellular machinery for protein synthesis, consists of RNA such that peptide-bond formation is catalyzed by ribosomal RNA. These findings strongly suggest that RNA, or some RNA-like molecule, was an integral piece of the earliest biochemical processes to evolve on the Earth. In 1986, Gilbert coined the phrase 'RNA World' to invoke a time during the coalescence of life when RNA was responsible for most, if not all, of the genetic transformations of the planet's first organisms. Ribozymes have, thus, played an important role in biological evolution and have recently been used to perform evolution experiments in the laboratory.

Ribonucleic Acid

RNA is a polymer of ribonucleotides. RNA differs from DNA in three basic respects. First, RNA employs uracil as a nitrogenous base, in place of the thymine used in DNA. Second, RNA nucleotides possess a hydroxyl group at the 2' position, while DNA is deoxygenated at that position to a proton. Third, RNA is more often found single stranded than DNA, which is typically completely base paired into a double helix. All three of these features help engender RNA with catalytic capabilities, although they also tend to make RNA chains less chemically stable than DNA. From an informational point of view, both DNA and RNA have the capability to store and transmit biological information through complementary base pairing: C with G and T(U) with A.

Catalytic RNA

RNA as an Enzyme

Enzymes are molecules that can catalyze a chemical reaction by lowering the activation energy needed to transit from reactants

to products. Proteins, built up from amino acids of 20 or more varieties, are well known to perform catalysis in biological systems. RNA is built up from only four nucleotide varieties but still can perform catalysis. Similar to protein-directed catalysis, the basis for this is the acquisition of a three-dimensional structure that contains a catalytic pocket. Many ribozymes can speed up the rate of reaction 10^3 – 10^{11} -fold over the corresponding uncatalyzed rate and, in certain cases, can behave as true enzymes by 'turning over', catalyzing more than one reaction event.

Chemistry

In the catalytic pocket, RNA usually relies on metal-ion coordination and one or more chemical strategies to accomplish catalysis. In some cases, the 2'-hydroxy group itself can serve as a nucleophile to attack a chemical bond in the substrate molecule. In other cases, the chemical moieties on the nitrogenous bases (adenine, guanine, uracil, or cytosine) can act as acids or bases to stimulate a reaction. For example, it is now thought that peptide-bond formation in the heart of the ribosome is promoted by some type of interaction between amino acids and a phylogenetically invariant adenosine residue in the ribosomal RNA. In most cases, RNA must bind divalent metals such as magnesium to help properly position the substrate molecules in order to stabilize the transition state.

Classes of Natural Ribozymes

Since 1982, nine distinct types of naturally occurring ribozymes have been discovered (Figure 1). With the notable exception of the ribosome, these ribozymes are generally limited in their capacity to making and/or breaking phosphoester bonds that hold nucleic acids together. In addition to the list below, RNA plays an important role in the catalytic events promoted by ribonucleoproteins such as telomerases, spliceosomes, and editosomes. Other natural ribozyme will undoubtedly be discovered in the near future.

- 1 and 2. *Group I and Group II Introns*. These are large (>200 nucleotides) self-splicing introns: ribosomal RNA (rRNA) or transfer RNA (tRNA) introns that excise themselves from nascent RNA transcripts via ribozyme activity.

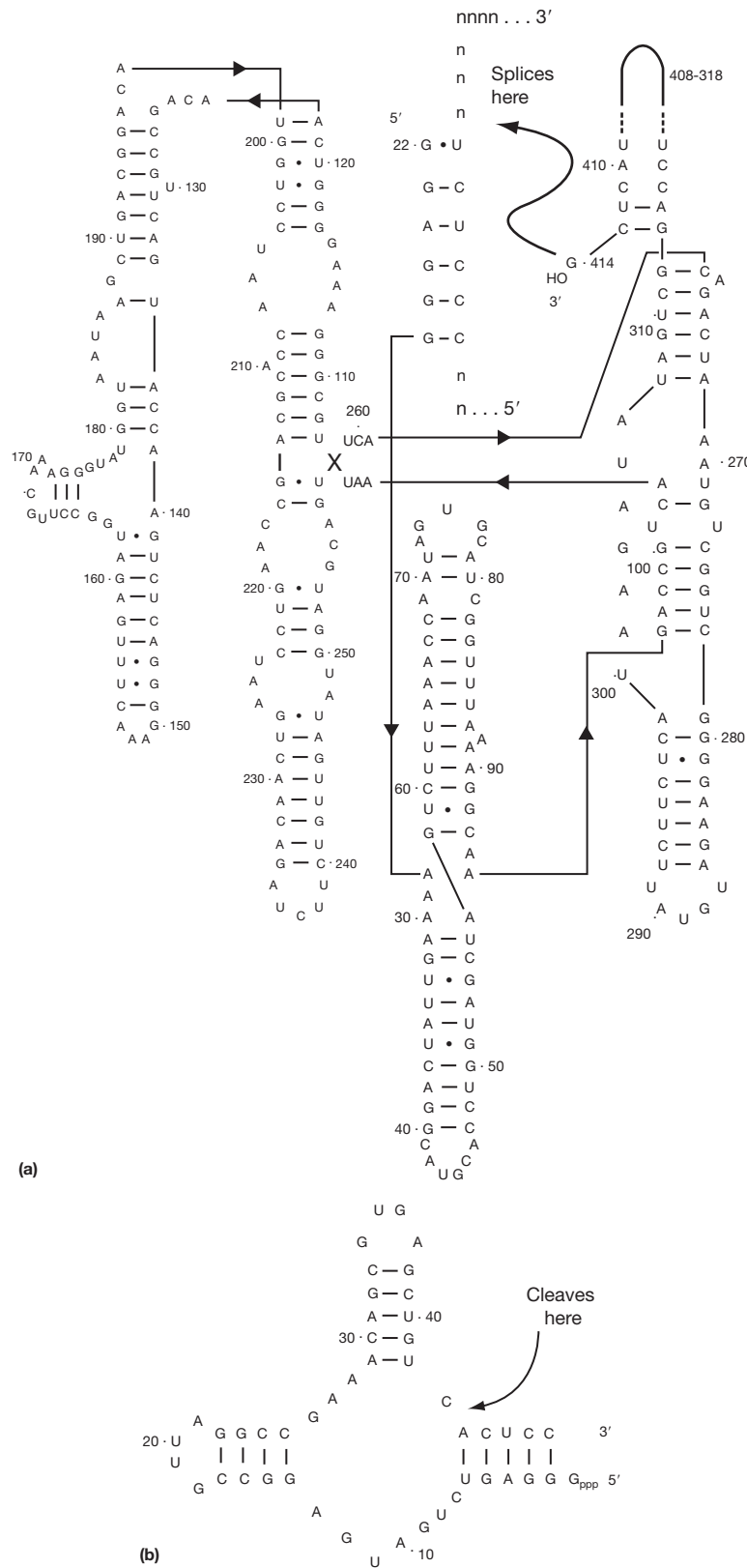


Figure 1 Some known ribozymes. (a) The group I intron from the ciliate *Tetrahymena thermophila*. Shown is the shortened version of the natural rRNA intron; this 414-nucleotide ribozyme binds a complementary RNA substrate (lowercase letters) via complementary base pairing and catalyzes phosphoester-bond cleavage and ligation reactions at the site indicated by the arrow. (b) One possible hammerhead ribozyme. This ribozyme self-cleaves at the site indicated by the arrow.

3. *RNase P*. RNase P is a large ribozyme that allows maturation of nascent tRNA transcripts and a few other small RNAs by removing the 5'-leader portion of those transcripts.
4. *Hammerhead ribozymes*. These small ribozymes, of which there are many throughout nature, possess three base-paired stems and catalyze self-cleavage of a RNA phosphodiester bond.
5. *Hairpin ribozymes*. Like the hammerheads, hairpin ribozymes, originally found in a tobacco RNA virus, self-cleave an RNA bond, but unlike the hammerheads, consist of four stems in two general domains.
6. *Hepatitis delta virus ribozyme*. The hepatitis delta virus (HDV) ribozyme is another viral-derived small ribozyme that catalyzes RNA bond cleavage, but it is slightly more complex, having a 'pseudoknot'-folded structure.
7. *VS ribozyme*. The VS motif is another small self-cleaving ribozyme, found first in a satellite plasmid of *Neurospora* mitochondria. This ribozyme, as is the case for all RNA enzymes, can also catalyze the reverse reaction, which here would be ligation of two RNA strands.
8. *Riboswitch ribozyme*. The *glmS* gene from Gram-positive bacteria produces upon transcription an RNA that can fold into two conformations depending on the presence or absence of a small metabolite molecule, glucosamine-6-phosphate. In one of these two conformations, the RNA cleaves itself, thereby halting translation of the RNA into a protein.
9. *The ribosome*. It is now understood that the RNA portion of ribosomes contains the catalytic function of the peptidyl-transferase center, which joins two amino acids together during the translation of messenger RNA into proteins.

Unnatural Ribozymes

Using the techniques of systematic evolution of ligands by exponential enrichment (SELEX), *in vitro* selection, and *in vitro* evolution (see below), many novel catalytic RNA sequences have been discovered (Figure 2), many of which promote chemistries not yet seen in natural ribozymes. The existence of these ribozymes demonstrates not only that RNA has a wide repertoire of catalytic capabilities, but that evolutionary forces are quite adept at molding RNA-based catalysis.

1. *Ligase ribozymes*. Isolated from random pools of RNA sequences, the ligase ribozymes were the first type of unnatural ribozymes to be fashioned. These RNAs catalyze phosphoester-bond formation between themselves and an exogenous oligoribonucleotide. Several classes of ligase ribozymes exist, with differing RNA folds. Ligases catalyze bond formation by a chemistry that is distinct from that of known natural ribozymes, but which is remarkably similar to the chemistry used by the protein enzyme RNA polymerase. In fact, some ligases can be engineered to behave as rudimentary RNA polymerases.
2. *Diels-Alderase ribozymes*. To move from phosphoester chemistry to carbon-carbon bond chemistry, it was necessary to derivatize one of the four bases, uridine, with a pyridyl moiety. This allowed the selection *in vitro* of a Cu^{2+} -dependent ribozyme that could catalyze a Diels-Alder cycloaddition reaction.
3. *Three- and two-base ribozymes*. Recently, ribozymes have been selected that use only three (i.e., A, G, and U)

nucleotides to perform catalysis. Also, if only U and 2,6-diaminopurine are available, a 2-base ligase ribozyme can be constructed through *in vitro* evolution.

4. *Other unnatural ribozymes*. Many other unnatural ribozymes exist and more are being constructed. These ribozymes can catalyze amide bond formation, N-alkylation, porphyrin metallation, thioester bond formation, and a host of other chemical reactions. Catalytic centers of these ribozymes can be as small as five nucleotides.

Phylogenetic Distribution of Ribozymes

Judging from the phylogenetic distribution of natural ribozymes (Figure 3) and from the central nature of the reactions that they catalyze, it is clear that catalytic RNA was likely present in the last common universal ancestor (LUCA) of contemporary life. Group I ribozymes, for example, are present in eukaryotes (e.g., the ciliated protozoan *Tetrahymena*) and in eubacteria, including some bacteria thought to occupy basal lineages in the tree of life (e.g., the hyperthermophilic bacterium *Thermotoga*). Ribozymes are wide spread in RNA viruses, group II introns have been isolated from deep-sea hydrothermal vent bacteria, and small ribozymes have been isolated from genomes in organelles such as mitochondria and chloroplasts. Also, the RNase P ribozyme is present in a variety of archaeal species.

RNA and the Origins of Life

In addition to the inference that ribozymes were a feature in the LUCA, there is much evidence and corresponding speculation that RNA, in particular catalytic RNA, was critical in the earliest life forms on the Earth. RNA, or an RNA-like molecule, could even have been the basis for the origins of life itself. Initially, this postulate was based on the observation that RNA was the only known polymer to possess simultaneously a genotype and a phenotype. Subsequently, more direct evidence of the antiquity of RNA has come to light. These data include the realization that the ribosome is a ribozyme, that RNA can catalyze a diverse array of reactions, a ligase-derived RNA can catalyze rudimentary RNA polymerization, RNA can catalyze recombination, and, importantly, that populations of RNA can evolve in a test tube and respond to environmental pressures in manners quite analogous to natural populations of organisms. Moreover, much is now known about plausible abiotic synthetic routes to RNA precursors such as purines, pyrimidines, carbohydrates, and phosphorylated compounds. For example, cationic clays such as montmorillonite can promote the polymerization of RNA-like monomers into RNA chains of 50 or more nucleotides in length.

In Vitro Evolution of RNA

Systematic Evolution of Ligands by Exponential Enrichment

SELEX is a powerful method to isolate RNA sequences that bind to particular substrates without necessarily reacting with them. Random pools of RNA, typically 40–100 nucleotides in

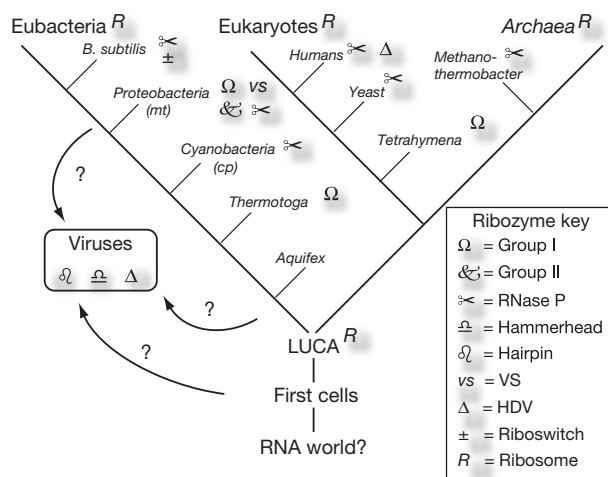


Figure 3 Distribution of ribozymes in the tree of life. The examples shown here are not exhaustive; branch lengths are not indicative of extinction events or amount of evolutionary change. Mitochondria (mt) and chloroplasts (cp) are included with their bacterial precursors. HDV-like ribozymes have been detected in the human genome.

the initial pool can either be a set of random RNA sequences, or a set of mutants based on a particular known wild-type sequence.

Evolution *In Vitro*

Evolution *in vitro* follows the strategy of selection *in vitro* except that in each round, mutations are deliberately introduced into the RNA population to provide additional variation on which natural selection can operate (Figure 4). The mutation can come from error-prone protein polymerase enzymes, from mutagenic polymerase chain reaction (PCR) or from sexual PCR. Evolution *in vitro* can in principle be carried out for an indefinite number of rounds, but in principle convergence on a particular catalytic solution, if one exists, typically occurs after 10–20 rounds and will occur in general accordance to the principles of population genetics.

Continuous Evolution *In Vitro*

The most recent modification of these RNA evolution techniques is to mimic a prebiological milieu in which all the components of evolution – the RNA, the nucleotides, the salts, the oligonucleotides, and the protein enzymes, if any – are simultaneously present in a single test tube. In this fashion, several rounds of evolution occur autonomously before equilibrium is reached. Serial dilution and fresh additions of

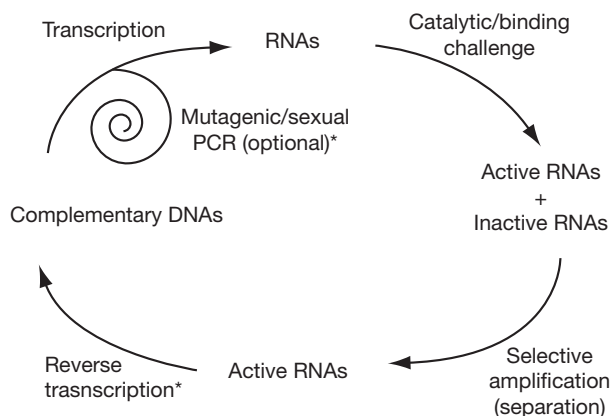


Figure 4 *In vitro* selection/evolution. A pool of RNA molecules, partially or completely randomized in nucleotide sequence, are challenged to perform a novel binding and/or catalytic event. As a result of successfully performing that event, RNA molecules become physically tagged in such a manner that they can be separated from unsuccessful RNA molecules. The successful RNAs then get converted to a corresponding DNA sequence through reverse transcription and back into RNA by forward transcription. A complete cycle is called a round, or a generation, and may or may not include mutagenic steps (indicated by asterisks) to introduce an evolutionary element into the process.

reagents can result in the extremely rapid evolution of RNA populations.

See also: [Molecular Biology: Catalysis by RNA](#), [Structural Themes: Group I Introns](#); [Pre-tRNA and Pre-rRNA Processing in Bacteria](#); [Riboswitches](#).

Further Reading

- Bartel DP and Unrau PJ (1999) Constructing an RNA world. *Trends in Cell Biology* 9: M9–M13.
- Bittker JA, Phillips KJ, and Liu DR (2002) Recent advances in the *in vitro* evolution of nucleic acids. *Current Opinion in Chemical Biology* 6: 367–374.
- Breaker RR and Joyce GF (1994) Inventing and improving ribozyme function: Rational design versus iterative selection methods. *Trends in Biotechnology* 12: 268–275.
- Gilbert W and de Souza SJ (1999) Introns and the RNA world. In: Gesteland RF, Cech TR, and Atkins JF (eds.) *The RNA World*, 2nd edn., pp. 221–231. Cold Spring Harbor, NY: Cold Spring Harbor Press.
- Joyce GF (2004) Directed evolution of nucleic acid enzymes. *Annual Review of Biochemistry* 73: 791–836.
- Lilley DMJ and Eckstein F (eds.) (2007) *Ribozymes and RNA catalysis: Introduction and primer*. In: *Ribozymes and RNA Catalysis*, pp. 1–10. London: Royal Society of Chemistry.
- Wilson DS and Szostak JW (1999) *In vitro* selection of functional nucleic acids. *Annual Review of Biochemistry* 68: 611–647.

Ribozyme Structural Elements: Group II Introns and the Spliceosome

O Fedorova, Yale University, New Haven, CT, USA

C Waldsich, MFPL, University of Vienna, Vienna, Austria

© 2013 Elsevier Inc. All rights reserved.

Glossary

Helicases Enzymes that move directionally along DNA or RNA duplexes, thereby separating (unwinding) the two nucleic acid strands in an ATP-dependent manner.

Intron A noncoding RNA region that interrupts a protein-coding mRNA sequence or intervenes non-protein-coding RNA sequences. Accordingly, an intron has to be removed from the precursor RNA by a process named splicing.

RNA folding The process that describes how an RNA undergoes the structural transition from the

unfolded disordered state to the native, functional conformation.

Self-splicing introns (group I and group II introns)

Catalyze their own excision from the precursor RNA through two sequential transesterification reactions.

Spliceosome A large complex that consists of small nuclear ribonucleoproteins (snRNPs) and facilitates splicing of most eukaryotic introns, so-called nuclear introns, which are not capable of self-splicing.

Splicing The process by which introns are excised out of the primary RNA transcript and exons (i.e., coding regions) are joined together to generate the mature RNA.

Introduction

Group II introns are large, multidomain RNAs that are capable of self-splicing. This abundant class of catalytic RNA is found in bacteria, lower eukaryotes and in organelles of higher eukaryotes. Although these introns do not share a lot of sequence conservation, their secondary structure organization is highly conserved, consisting of six major domains (DI–DVI), which radiate from a central wheel (**Figure 1**). Notably, group II introns can be subdivided into three main phylogenetic families, the IIA, IIB, and IIC introns, and two minor groups, the IID and IIE introns. Like most RNAs, group II introns have to adopt a specific 3D architecture to be functional. Numerous tertiary contacts are required to assemble the individual intron domains into the catalytically active conformation. *In vitro* high-salt and elevated temperature are typically necessary to promote self-splicing of group II introns. By contrast, in the cell these introns depend on protein cofactors, such as intron-encoded maturases or host-encoded splicing factors, to promote their own excision from the precursor RNA. A hallmark of this intron class is their catalytic versatility, with forward splicing, reverse splicing, and trans-splicing being the most notable examples. In the first step of splicing, the 2′OH group of a bulged adenosine in DVI, the so-called branch point, acts as a nucleophile to attack the 5′-splice site, resulting in the release of the upstream exon and in a covalent bond (2′–5′ linkage) between the 5′ end of the intron and the branch-point adenosine (**Figure 2(a)**). In the second step, the 3′OH group of the liberated 5′ exon attacks the 3′-splice site, resulting in exon ligation and the intron is released as lariat. Alternatively, to this branching pathway, a water molecule can act as a nucleophile in the first transesterification, which eventually results in ligated exons, but the intron is excised in a linear form (**Figure 2(b)**). Notably, in the case of the yeast mitochondrial intron ai5γ, the branching pathway is the predominant splicing mechanism in the cell, but the hydrolytic pathway does exist *in vivo* as well. However, there are group II introns

that, unlike ai5γ, do not possess a branch point, implying that these variants splice predominantly through hydrolysis. Importantly, the excised intron remains active and is able to catalyze reverse splicing into RNA and DNA substrates (**Figure 2(c)**), which represents the first step toward intron homing. This process, however, strictly depends on intron-specific maturases, which comprise RNA-binding, DNA endonuclease and reverse transcriptase motifs, and form an RNP complex with their intron. This astounding ability makes group II introns mobile genetic elements, which contribute to the genomic organization of their hosts. Another essential feature of these introns is their striking modularity. This characteristic had been used *in vitro* to create diverse group II intron-derived ribozymes of different size, catalytic proficiency, and domain content. Notably, individual domains can be efficiently assembled *in trans*. This striking phenomenon is also found in plant organelles, where two- or tri-partite introns naturally exist. These split introns form the functional structure and splice *in trans*. The modular nature of group II introns and their dominant splicing mechanism gave rise to the theory of group II introns sharing a common ancestor with the spliceosome – a large RNP machinery involved in nuclear intron splicing.

This article focuses on the structural elements and the folding mechanism of group II introns. For an in-depth discussion of group II intron evolution, their role as mobile elements and their protein collaborators, the reader is directed to recent articles on these topics by Belfort, Lambowitz, Michel, Pyle, Zimmerly, and their co-workers.

The Structural Organization of Group II Introns

Domain I

DI is the largest domain, which provides the scaffold for docking of other intron domains and the exon-binding sites (EBS1 and EBS2). Together with the catalytic center DV, which is firmly anchored in the DI scaffold through two long-range

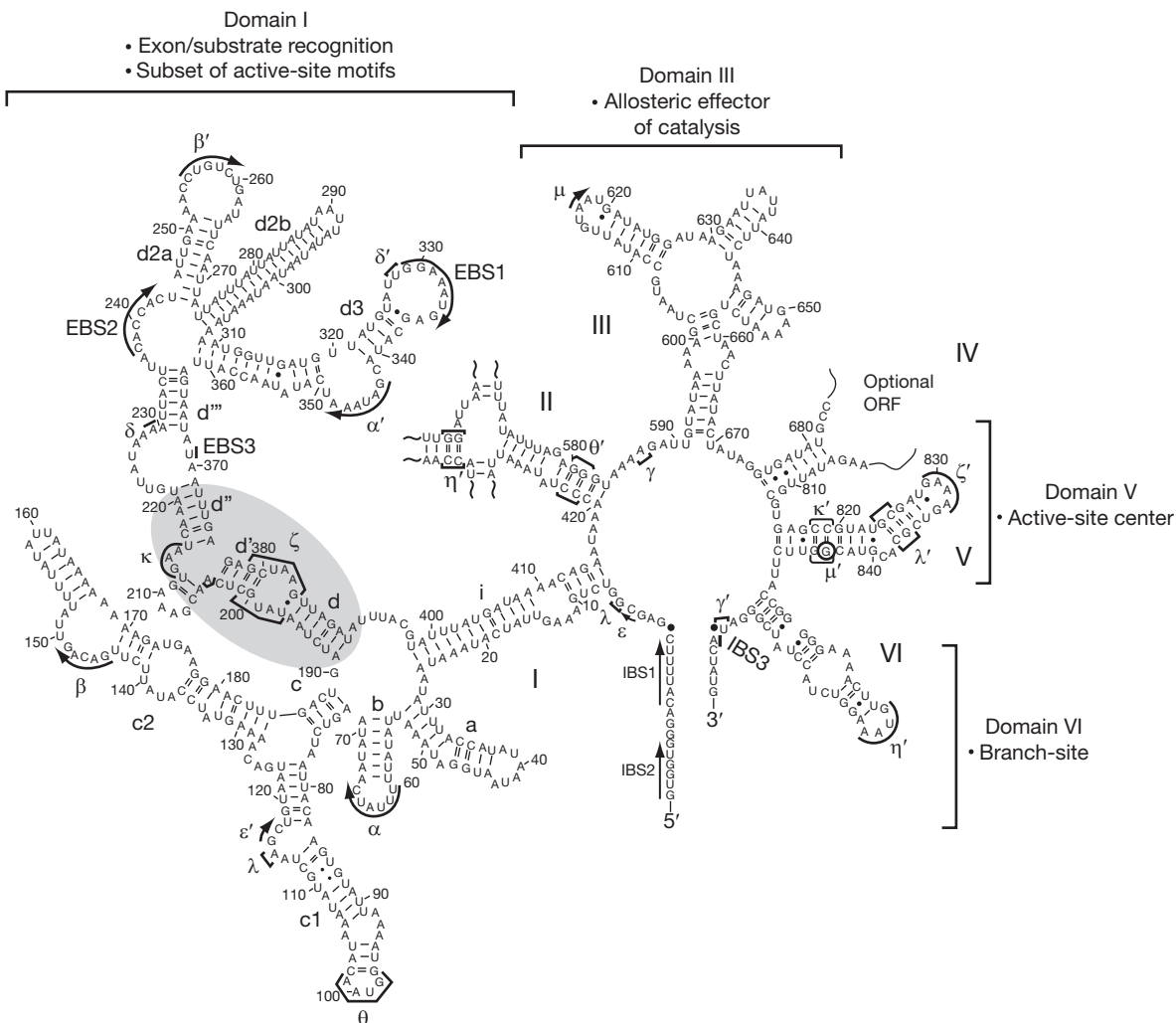


Figure 1 The conserved secondary structure organization of group II introns. The *Sc. coxI/ai5 γ* group IIB intron is shown as an example, whereby only the basal stems of DII and DIV are shown. Long-range tertiary interaction partners are indicated with Greek symbols. The κ - ζ element, which plays an important role in ai5 γ folding, is highlighted in gray. Some group II introns contain ORFs coding for a maturase protein within DIV.

interactions (κ - κ' and ζ - ζ'), DI forms the minimal group II ribozyme. DI does not only provide the platform for active-site assembly, but also harbors essential active-site constituents, like ε - ε' and λ - λ' . The DI architecture itself is characterized by the intradomain tertiary contacts α - α' and β - β' , of which only the former is shared by all group II introns, while the latter is not strictly conserved. The five-way junction from which the main substructures of DI protrude is another key feature of the DI structural organization, as it influences the spatial arrangement of critical elements such as the EBS sites, the DV docking region, and residues engaging in tertiary interactions (α , α' , β , β' , θ , ε' , and λ). Another intriguing DI feature is that, part of its secondary structure has a high contact order (e.g., stem d and stem i). In the case of the ai5 γ intron, formation of these helical structures depends on a protein cofactor in yeast mitochondria (Liebeg, Mayer, and Waldsich), while magnesium ions are essential for their formation of stem d *in vitro*. Given its central role in the intron architecture, it was not surprising to discover that DI is an independent folding unit, which adopts the same core structure by itself as within the intron

context (except for the active-site constituents), thereby providing a preorganized scaffold for subsequent docking of other intron domains.

Domain II

DII is not strictly required for group II intron catalysis. In group II introns from different organisms, this domain exhibits substantial variability in sequence and size. However, it is somewhat conserved with respect to tertiary interactions it forms with other domains. DII is involved in two long-range tetraloop-receptor type tertiary contacts with DI (θ - θ') and DVI (η - η'). The former is believed to stabilize the tertiary fold of the ribozyme, while the latter may play a role as a conformational switch between first and second steps of splicing. Biochemical data strongly indicate that the formation of η - η' contact is detrimental for the first step of splicing, while beneficial for the second step. Disruption of the η - η' contact severely impairs the second step of splicing, while the first step is either unaffected (in IIB introns) or enhanced (in IIA

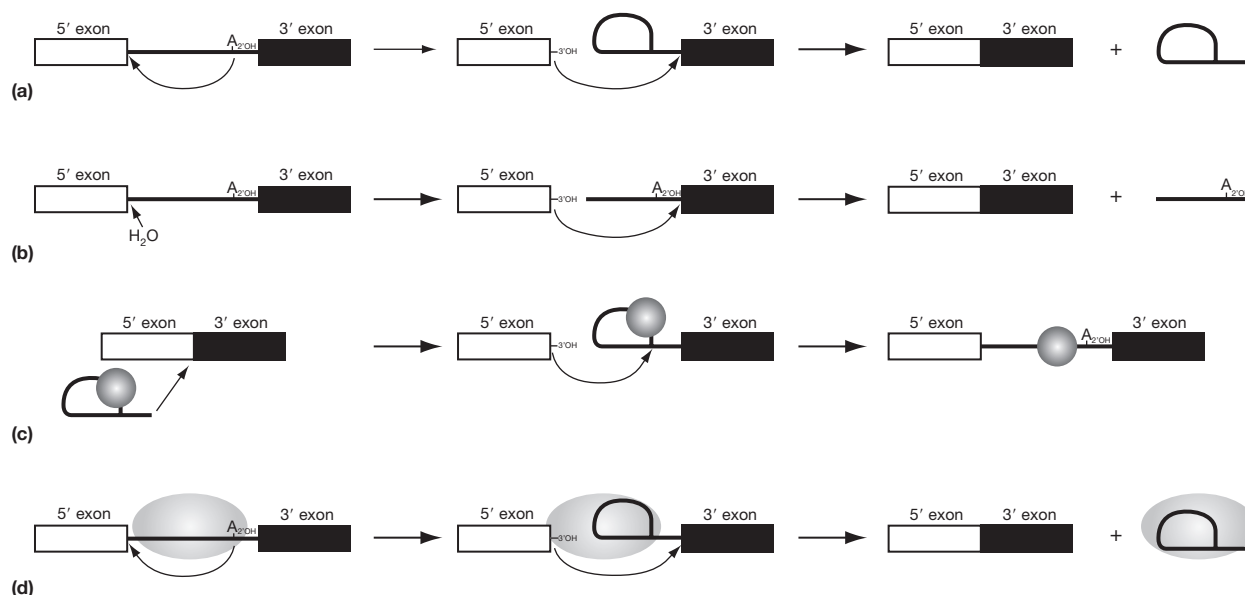


Figure 2 The splicing pathway of group II and nuclear introns. (a) The branching pathway of group II introns, in which the 2'OH of the branch-point adenosine makes the nucleophilic attack at the 5'-splice site and the intron is released in a lariat form. (b) The hydrolytic pathway of group II introns, in which water acts as a nucleophile in the first step of splicing and the intron is excised in a linear form. (c) Group II introns are capable of reverse splicing into RNA and DNA substrates. This is the first step toward intron homing, which depends on intron-encoded maturases (indicated as dark gray sphere). (d) Spliceosomal intron splicing, which is mechanistically similar to the branching pathway of group II introns (a). The light gray ellipse symbolizes the spliceosome, which catalyzes splicing of nuclear introns in eukaryotes. The upstream and downstream exons are shown as white and black boxes, respectively, while the intron is displayed as a black line.

and some IIB introns). By contrast, strengthening this contact nearly abolishes the first transesterification. These data suggest that for efficient group II intron splicing, $\eta-\eta'$ interaction must form after the first step. Another important function of DII is that its basal stem helps correctly position the joiner between DII and DIII (J2/3), which is an essential active-site element, in the catalytic core of the ribozyme.

Domain III

DIII is frequently referred to as a 'catalytic effector'. Although its presence is not absolutely necessary for the group II intron catalysis, it substantially accelerates all reactions catalyzed by the group II ribozyme. One of the most conserved elements of DIII (and the intron itself) is an A-rich internal bulge, which is essential for catalytic activity. It forms a series of noncanonical base pairs which, in group IIA and IIC introns, results in a loop E-like motif. Multiple biochemical data suggest that in IIB introns, the conserved A-rich bulge is located in close proximity to the 5'-splice site and the dinucleotide bulge of DV, which are important active-site elements. In group IIB introns, DIII structure includes a GUAAY pentaloop capping DIIIa stem. NAIS studies revealed an interaction ($\mu-\mu'$) between the two adenosines from this loop contacting the backbone of G844 (ai5 γ numbering) in DV, which is adjacent to the $\kappa-\kappa'$ contact between DV and DI. Recent crystallographic data suggest that in IIC introns DIII may adopt a slightly different orientation than that in IIB introns. However, both conformations are consistent with DIII acting as an allosteric catalytic effector, by stabilizing the active site via a complex network of interactions with other domains.

Domain IV

In some group II introns, DIV encodes an open reading frame for a maturase protein and contains a primary binding site for the maturase. This protein is intron specific, and it plays an important part in splicing, reverse-splicing, and retro-homing of the intron. In ORF-less introns, DIV has no specific function and is dispensable for catalysis.

Domain V

DV is the catalytic heart of a group II intron. It is the smallest and the most conserved domain of the ribozyme. DV is a 34-nucleotide hairpin, which contains an asymmetric dinucleotide bulge and is generally capped with a GNRA tetraloop. When DV is deleted, the catalytic activity of group II ribozymes is completely abolished. Mutational analysis, single atom substitutions and NAIM revealed multiple residues in DV that are critical for catalysis. Among these residues is the dinucleotide bulge and a so-called 'catalytic triad' AGC, which is located at positions 2–4 of the basal stem of DV. Although base identity of A and C residues of the catalytic triad is less critical than their participation in Watson–Crick base-pairing with the opposite strand, G is completely invariant. Single atom substitutions at this nucleotide suggest that its minor groove functionalities are important for ground-state interactions, while its major groove functional groups are critical for the chemical step of catalysis. However, the role of the catalytic triad in the group II ribozyme catalysis became fully understood only after the crystal structure of *O. iheyensis* (Oi.) intron was solved. The catalytic triad forms a triple helix, with J2/3 and one of the bulge nucleotides

(the 'catalytic triplex'), which helps create a negatively charged metal-binding pocket that harbors two magnesium ions with a potential role in catalysis.

A combination of phylogenetic analysis and biochemical studies revealed a series of long-range tertiary contacts involving DV. The GNRA tetraloop capping DV docks into a conserved 11-nucleotide receptor in DI (ζ - ζ'). In both IIA and IIB introns, DV is capped by a GNRA tetraloop. Interestingly, in IIC introns, a GANC tetraloop, which forms a novel type of interaction with its cognate receptor in DI, is found. The architecture of a GANC tetraloop is slightly different from the canonical GNRA tetraloop. The sugar pucker of its leading G residue is in a C2'-endo conformation instead of the more common C3'-endo. This results in the first base of the tetraloop being shifted to the minor groove of the helix, as opposed to the GNRA tetraloop structure, in which the first base stacks on the closing base pair. Consequently, the second through fourth bases of the GANC tetraloop stack on the closing base pair, instead of being pushed to the minor groove like in a GNRA tetraloop. These differences make it impossible for a GANC tetraloop to be inserted into receptor, in the same manner, as observed for a GNRA tetraloop-receptor interaction. The receptor structure is also different from that of a canonical 11-nucleotide motif. It is reduced to a three-nucleotide motif that does hydrogen-bond with the tetraloop but does not form the canonical AA-platform motif. In addition, three nucleotides, which in other introns belong to the ζ receptor, form a ribose zipper ω - ω' interaction with tandem sheared base pairs in a distal part of DI (close to EBS1 and α'). These interactions appear to be unique to the group IIC introns.

In addition to the ζ - ζ' contact, DV forms two more tertiary contacts with DI: κ - κ' and λ - λ' . The former, possessing some GNRA-like characteristics, plays a primarily structural role by helping position DV in the active site of the ribozyme. The latter, forming minor groove base triples, brings together catalytically important elements in DI and DV and helps anchor the 5'-splice site in the ribozyme active-site.

Domain VI

The main function of DVI is presenting the branch-point adenosine, which serves as a nucleophile in the branching pathway of the first step of splicing (Figure 2(a)), to the active site of the group II ribozyme. Branching occurs with exceptionally high fidelity; mutations designed to disrupt correct recognition of the branch site by the active site of the ribozyme generally result in impaired rather than cryptic branching. Surprisingly, aside from the branch-point adenosine, DVI is not very conserved. It has been proposed that correct selection of the branch site is determined by three factors: local geometry of the branch site, which is generally located below a conserved G-U pair, length of the DVI basal stem, and length of the linker between DV and DVI (J5/6). Recent UV-cross-linking studies have identified a purine-rich asymmetric internal loop between stems d'' and d''' (coordination loop) as a receptor for the branch site. The same study has also demonstrated that the branch site is located close to the 5'-splice site and J2/3 connecting DII and DIII, which is consistent with its role as a nucleophile in the first transesterification. A substructure similar to the coordination loop in IIB introns is also observed in

group IIC introns, although no information is available about its possible interaction with DVI. Intriguingly, there is no obvious alternative for the coordination loop in IIA introns, suggesting that they may use a different mechanism of the branch-point recognition.

Three-Dimensional Architecture and Structural Features of Group II Introns

Over 20 years have passed since the discovery of group II introns to the first crystal structure of this large ribozyme, which was solved by Pyle and co-workers in 2008. In the meantime, researchers relied on three-dimensional models obtained by using distance constraints from phylogenetic covariation and biochemical experiments. Although these models were successful in accurately positioning many of the key active-site elements, they could not depict the ribozyme active site at atomic resolution and predict all the tertiary interactions comprising the catalytic core. This void was finally filled by the crystal structure of the *Oi*. group IIC intron. It provides an invaluable insight into the overall three-dimensional organization of this intron class as well as the role of critical active-site elements.

As predicted previously from biochemical studies, the active-site elements are deeply internalized within the intron structure and protected by an outer shell formed by coaxially stacked helical regions of DI and DIII and basal stem of DII (Figure 3). DII and DIV are oriented away from the active site, which is consistent with their dispensable role in splicing. One of the key elements of the intron structural organization is the five-way junction in DI, which positions DI helices to form long-range tertiary interactions and create a scaffold for docking of other domains. This is consistent with previous findings, which identified many functionalities in this substructure as critical for intron folding. One of the most interesting features of the five-way junction is a T-loop motif, which is formed by the loop capping IA stem and intercalating adenosine residue A245 in JD/I (Figure 3). The T-loop forms a network of stacking interactions, which define the architecture of the junction. However, the most striking of all newly discovered motifs is the z-anchor, which is formed by bulged nucleotides 106–111 of subdomain IC and includes ϵ - ϵ' and λ - λ' contacts (Figure 3). It has been named 'z-anchor' because of two 160° kinks in the backbone between residues 109 and 111, which cause bases to alternate from side to side. This extended motif brings together DV, internal loop of stem I in DI and beginning of the intron, thus creating the active site. In addition to new motifs, the crystal structure confirmed the existence of predicted long-range tertiary contacts such as θ - θ' and the kissing-loop interaction α - α' .

Co-crystallization of the *Oi*. intron with exonic substrate helped elucidate exon recognition and coordination of metal ions in the ribozyme active site. The structure reveals the presence of two metal ions located at a 3.7-Å distance from each other in the catalytic core of the ribozyme (Figure 3), which is consistent with a two-metal ion mechanism of the group II intron catalysis. As these ions coordinate directly to the scissile phosphate of the ligated exons mimic, this suggests that these metals are involved in the catalytic mechanism.

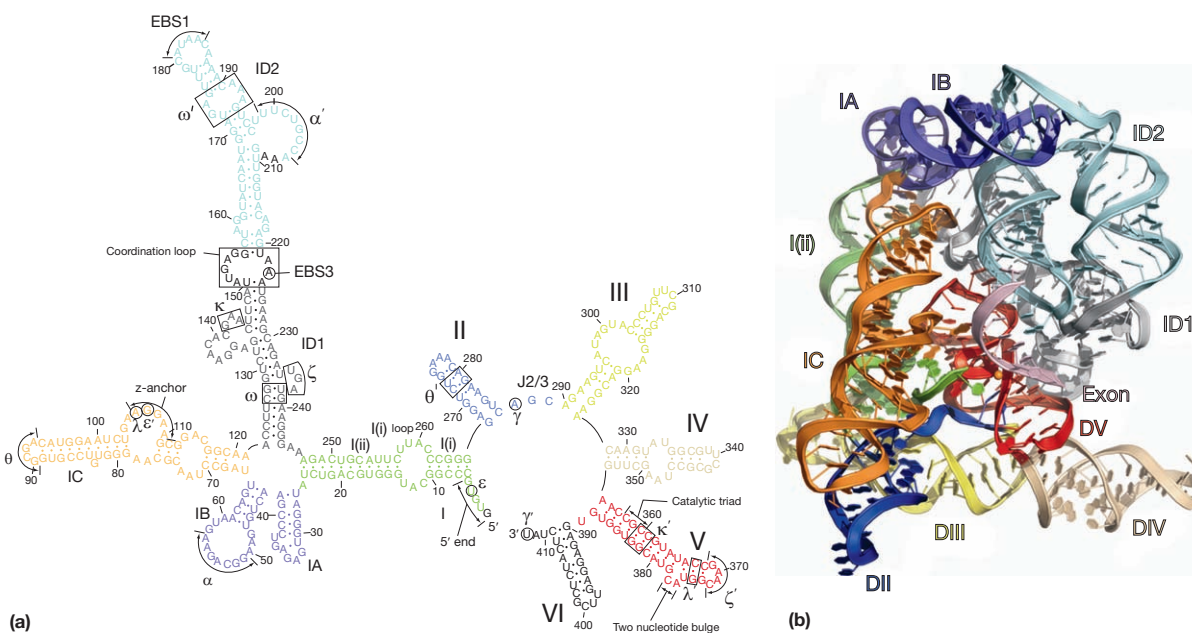


Figure 3 The *O. iheyensis* group IIC intron. (a) The secondary structure map and (b) the crystal structure of the *Oi.* group IIC intron. Roman numerals and Greek letters indicate domains and tertiary interaction partners, respectively. Color code is the same for both panels. Disordered nucleotides are shown in black. The atomic coordinates were published in Toor N, Keating KS, Taylor SD, and Pyle AM (2008) Crystal structure of a self-spliced group II intron. *Science* 320: 77–82; and Toor N, Rajashankar K, Keating KS, and Pyle AM (2008) Structural basis for exon recognition by a group II intron. *Nature Structural and Molecular Biology* 15: 1221–1222.

Group II Intron Folding and the Role of a Protein Cofactor

For a long time, it was assumed that characterizing the higher-order structure of group II introns might be impossible due to their size and their apparent instability. However, during the past 10 years, significant progress has been made in describing the architecture as well as the folding pathway of group II introns. Most of the work focused on the yeast mitochondrial group IIB intron *coxI/ai5γ*.

The Autonomous Folding Pathway of *ai5γ* Ribozyme *In Vitro*

Although protein cofactors are essential for intracellular structure formation of group II introns, these RNAs can fold autonomously *in vitro*. Thus far, only the maturase-less *ai5γ* intron has been subjected to detailed kinetic and equilibrium folding analyses at both non- and near-physiological conditions. A ribozyme derived from *ai5γ* folds directly to the native state through an ordered, hierarchical pathway devoid of kinetic traps. This ribozyme collapses slowly to a transient intermediate in DI, while subsequent docking of other intron domains is rapid (Figure 4(a)). A small substructure in DI, the κ - ζ element, which is also the DV docking site, was found to induce the DI collapse. The compact conformation is then locked in by intradomain interactions (α - α' , β - β'), allowing subsequent docking of DIII, DV, and DVI, thereby completing intron assembly. Folding of the κ - ζ element itself is stabilized by Mg^{2+} ions through a tertiary capture mechanism. The folding pathway was recently refined by a smFRET study, in which a novel intermediate (F) during transition from the I to N state was identified. Under near-physiological conditions,

the D1356 ribozyme collapses even more slowly to a compact, near-native state, but it is incapable of forming and/or maintaining the native, splicing-competent conformation (Figure 4(b)). This is consistent with a lack of splicing under such low-salt conditions.

Protein-Facilitated Group II Folding and Splicing

Group II intron splicing factors can be divided into two categories: intron-encoded maturases and host-encoded proteins. In the past 30 years, an incredible diversity of proteins that are recruited to assist intron splicing was identified. Most of these factors promote splicing indirectly. One prominent example is Mrs2p, which is a Mg^{2+} transporter and modulates the Mg^{2+} ion concentration in yeast mitochondria. However, only a handful of proteins with a direct role in splicing have been described. Our understanding of how such proteins assist intron folding and in turn splicing is almost exclusively limited to the studies on the nuclear-encoded protein Mss116p, which is essential for splicing of all yeast mitochondrial introns.

Mss116p belongs to the DEAD-box helicase family, which are known to be nonprocessive helicases and were recently suggested to be viewed as ATP-dependent RNA-binding proteins. Indeed, Mss116p displays unwinding and annealing activity *in vitro* and lowers the Mg^{2+} requirements for group II intron splicing *in vitro*. Notably, the ATPase activity of Mss116p is required to promote intron splicing, but it is still a matter of debate whether the helicase activity is critical for this process as well. Two different models have been proposed to explain Mss116p's role in *ai5γ* splicing. One model suggests that the protein acts in an RNA chaperone-like fashion by unwinding misfolded off-pathway intermediates. Alternatively, Mss116p

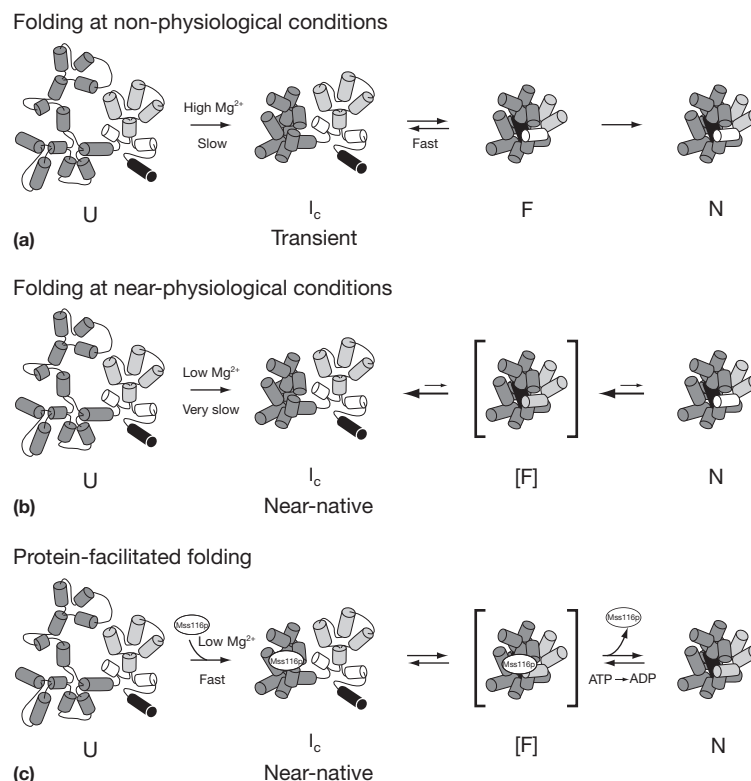


Figure 4 Folding of the ai5 γ intron *in vitro*. The ai5 γ folding pathway is depicted (a) at nonphysiological conditions, (b) at near-physiological conditions, or (c) in the presence of the DEAD-box helicase Mss116p. The cylinders represent individual intron elements. For simplicity, we only sketched DI in dark gray, DIII in light gray, and DV in black, while DII and DIV are in white and DVI and the exons were omitted. Mss116p is represented as an ellipse. The unfolded, compact intermediate, folded, and native states are abbreviated as U, I_c, F, and N, respectively. Notably, as the F state has not been studied at near-physiological conditions or in the presence of Mss116p, this state is hypothetical and therefore put in brackets.

might act more like a general RNA-binding protein and may in some cases stabilize folding intermediates. While the first reports on Mss116p's mechanism of action used kinetic assays to measure native state formation as a function of catalysis *in vitro*, no structural evidence was available until this year. First, it was shown that Mss116p accelerates the collapse of ai5 γ to the compact near-native state (U $\leftrightarrow$ I), in an ATP-dependent manner, by stabilizing early folding intermediates (Figure 4(c)). However, Mss116p does not stabilize the ai5 γ native state, but such stabilization is obtained from exon binding. ATP hydrolysis is required for recycling of the protein and Mss116p's release of the compact intron might be required to allow completion of intron assembly. Strikingly, the Mss116p-promoted ai5 γ folding pathway is highly reminiscent of that in the presence of high salt. The finding that a protein known for unwinding of RNA is also able to stabilize RNA structure broadens the spectrum of how DEAD-box proteins are able to shape RNA folding landscapes, and to influence RNP architecture and rearrangements – specifically in light of the potential function of helicases in nuclear intron splicing including the spliceosome assembly pathway.

Are Group II Introns and the Spliceosome Related?

The spliceosome is a large RNP machinery composed of snRNAs and a myriad of proteins and is essential for splicing of nuclear introns, which are not catalytically active (except for

some autocatalytic group I introns present in the nucleus of a few lower eukaryotes). Years ago, it was hypothesized that group II introns and the spliceosome share a common ancestor. This was mainly based on the mechanistic similarities of spliceosomal splicing and the branching pathway of group II introns, in which the introns are released in the lariat form (Figure 2). In group II introns, key elements of the active site are the catalytic triad AGC and the dinucleotide bulge of DV, as well as the junction J2/3 (Figure 5(a)), which from a triple helical scaffold. Counterparts of these essential elements are found in the spliceosome: The AGC triad is located in the U6 internal stem-loop (ISL), which is formed in the U2/U6 snRNA complex and resembles the group II intron catalytic center DV. The conserved ACAGAGA box, which is neighboring the ISL, is considered to be the J2/3 equivalent and plays crucial role at different stages in splicing. A tertiary interaction between the last G of the ACAGAGA sequence and the bulged A between helices 1a and 1b (containing the AGC triad) has also been proposed. Another important piece of evidence for a possible evolutionary relationship between group II introns and the spliceosome is the fact that the group II intron DV can functionally replace the U6atac ISL in an *in vivo* splicing assay. Despite all these commonalities, the evolutionary link between the spliceosome and autocatalytic group II introns remains an open question. In particular, it is still a matter of debate whether the spliceosome is indeed a ribozyme. Aside from snRNAs being the catalyst, it is also possible that the protein Prp8, which is located at the spliceosome active-site and has an

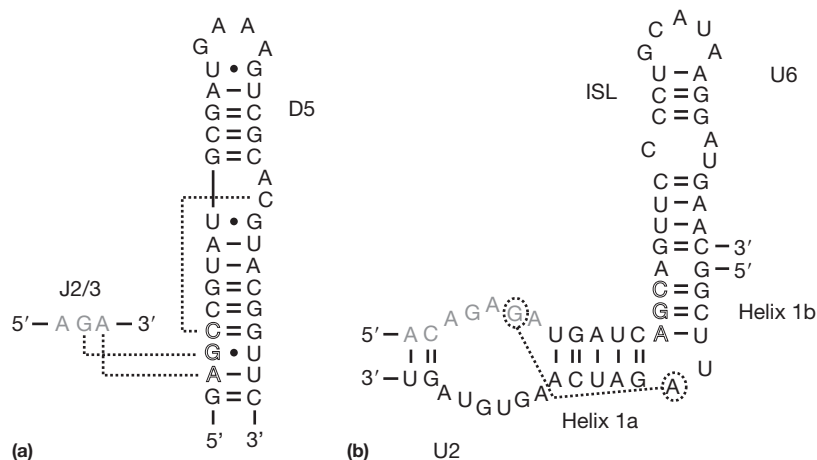


Figure 5 Key elements of the group II intron active site and their counterparts in the spliceosome. (a) DV, containing the AGC triad (white filled letters), and J2/3 (gray) of the *ai5γ* intron are shown. These highly conserved elements form a triple helix in the *Oi*. group IIC intron crystal structure (equivalent interactions that could potentially form in the *ai5γ* intron are indicated by dashed lines). (b) The *Saccharomyces cerevisiae* U6 internal stem-loop (ISL), which harbors the AGC triad (white filled letters), and the ACAGAGA box (gray), which is located upstream of the ISL, are shown. An interaction (dashed line) between the last G residue of this conserved box and the A of the dinucleotide bulge separating the U2/U6 intermolecular helices 1a and 1b had been identified.

RNase H-like fold, is the enzymatically active subunit of the spliceosome. Alternatively, it has also been suggested that the spliceosome might have two active sites, one composed of RNA and the other one of protein residues, each one responsible for a distinct step in splicing.

Conclusion

The atomic-resolution structure of a group II intron represents a major breakthrough that is likely to have the field undergo a significant boost in the coming years. In particular, it will be of great interest to obtain more detailed information on the conformational rearrangement during splicing. Also, the crystallized intron belongs to the IIC family, which is considered to be the ancestral type and is much smaller than IIA and IIB introns, possessing both unique and conserved structural features. Therefore, more structural highlights remain to be discovered. Since proteins assist group II splicing *in vivo*, the next challenge will be to obtain insights into the conformation of an intron complexed with its protein partner, such as a maturase or host-encoded splicing factor, at atomic level. As of now, there are no high-resolution structures of large spliceosomal components. Undoubtedly, the field will benefit from ongoing and future efforts to explore RNA–RNA and RNA–protein interactions, as well as associated rearrangements and dynamics within the spliceosome. Eventually, the open questions of whether the spliceosome is an RNA or protein enzyme and whether the splicing systems are indeed evolutionary-linked will be answered.

Acknowledgments

We are indebted to Kevin Keating and Anna Pyle for preparing Figure 3. O.F. thanks the Howard Hughes Medical Institute for funding. C.W. acknowledges the financial support from the Austrian Science Foundation FWF (grants P21017 and Y401).

See also: **Molecular Biology: Catalysis by RNA, Structural Themes: Group I Introns; Ribozymes and Evolution.**

Further Reading

- Beauregard A, Curcio MJ, and Belfort M (2008) The take and give between retrotransposable elements and their hosts. *Annual Review of Genetics* 42: 587–617.
- Bonen L and Vogel J (2001) The ins and outs of group II introns. *Trends in Genetics* 17: 322–331.
- Butcher SE (2009) The spliceosome as ribozyme hypothesis takes a second step. *Proceedings of the National Academy of Sciences of the United States of America* 106: 12211–12212.
- Fedorova O and Zingler N (2007) Group II introns: Structure, folding and splicing mechanism. *Biological Chemistry* 388: 665–678.
- Jurica MS (2008) Detailed close-ups and the big picture of spliceosomes. *Current Opinion in Structural Biology* 18: 315–320.
- Lambowitz AM and Zimmerly S (2004) Mobile group II introns. *Annual Review of Genetics* 38: 1–35.
- Lehmann K and Schmidt U (2003) Group II introns: Structure and catalytic versatility of large natural ribozymes. *Critical Reviews in Biochemistry and Molecular Biology* 38: 249–303.
- Michel F, Costa M, and Westhof E (2009) The ribozyme core of group II introns: A structure in want of partners. *Trends in Biochemical Sciences* 34: 189–199.
- Michel F and Ferat JL (1995) Structure and activities of group II introns. *Annual Review of Biochemistry* 64: 435–461.
- Pyle AM, Fedorova O, and Waldisch C (2007) Folding of group II introns: A model system for large, multidomain RNAs? *Trends in Biochemical Sciences* 32: 138–145.
- Pyle AM and Lambowitz AM (2006) Group II introns: Ribozymes that splice RNA and invade DNA. In: Gesteland RF, Cech TR, and Atkins JF (eds.) *The RNA World*, pp. 469–534. New York, NY: Cold Spring Harbor Laboratory.
- Solem A, Zingler N, Pyle AM, and Li-Pook-Than J (2009) Group II introns and their protein collaborators. In: Walter NG, Woodson SA, and Batey RT (eds.) *Non-Protein Coding RNAs*. Berlin: Springer.
- Toor N, Keating KS, and Pyle AM (2009) Structural insights into RNA splicing. *Current Opinion in Structural Biology* 19: 260–266.
- Valadkhan S (2007) The spliceosome: Caught in a web of shifting interactions. *Current Opinion in Structural Biology* 17: 310–315.
- Will CL and Luhrmann R (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Current Opinion in Cell Biology* 13: 290–301.

RNA Editing

C E Samuel, University of California, Santa Barbara, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Adenosine-to-inosine (A-to-I) editing The C6 deamination of adenosine (A) to yield inosine (I), which is recognized as G rather than A, and hence represents a kind of base substitution RNA editing.

Cytidine-to-uridine (C-to-U) editing The C4 deamination of cytidine (C) to yield uridine (U), which is a kind of base substitution RNA editing.

Deamination editing Base modification involving the replacement of the amino group of adenosine (A) with oxygen from water to generate inosine (I), or in the case of cytidine (C) to generate uridine (U).

Edited RNA RNA transcript in which the RNA nucleotide sequence differs from that of the encoding genome template sequence at one or more sites, either by base substitution or by nucleotide insertion/deletion.

RNA editing The process of posttranscriptional modification of an RNA nucleotide sequence by base substitution or by nucleotide insertion/deletion.

Unedited RNA RNA transcript in which the nucleotide sequence corresponds to that encoded by the genome template.

RNA editing is the posttranscriptional modification of an RNA nucleotide sequence at one or more positions. There are two general types of RNA editing, viz., substitution editing and insertion/deletion editing. RNA editing of either type leads to the formation of transcripts whose sequence differs from that of the genome template. Such RNA sequence differences between mature transcript and encoding genome represent a form of genetic recoding. The sequence changes generated by RNA editing are different from those arising from 5'-capping, 3'-polyadenylation, and splicing, additional RNA processing events that may also occur as part of mRNA biogenesis. RNA editing is widely observed in eukaryotic organisms and their viruses. Editing, like splicing, represents a form of processing that has the capacity to amplify genetic diversity and alter gene product function by modifying the information transfer process at the posttranscriptional level of gene expression.

Discovery of RNA Editing

RNA editing was discovered in a unicellular protozoan. The frameshifted cytochrome oxidase *coxII* transcript of trypanosome mitochondria was found to contain four inserted uridines that were not encoded by the genomic DNA. Then editing by nucleotide substitution was discovered in two mammalian messenger RNAs. The apolipoprotein B (apoB) mRNA in mouse intestine had a translation stop codon at a position in the mature mRNA, whereas the genomic DNA specified the amino acid glutamine; and the glutamate-gated ion channel (GluR-B) mRNA in mouse brain had the amino acid arginine in place of the genome-encoded glutamine at the same position. Several biochemical mechanisms have been shown to account for the various RNA editing events subsequently identified in a range of mRNA transcripts. These include A-to-I substitution editing; C-to-U substitution editing; U insertion/deletion editing; C insertion editing; and G insertion editing. In addition, 2'-O-methyl ribose nucleotide modification and uridine to pseudouridine conversion represent editing events

mediated by complexes of small nucleolar RNA (snoRNA) cofactor guides and associated proteins that modify ribosomal RNAs and other RNAs. This article focuses on A-to-I and C-to-U RNA editing that occur by selective RNA-specific deamination reactions that result in nucleotide substitutions. Genetic recoding by such substitution editing can produce gene products with altered sequence and hence altered function.

Adenosine-to-Inosine Substitution RNA Editing

RNA Substrates That Undergo A-to-I Editing

RNA editing by adenosine (A)-to-inosine (I) substitution is catalyzed by C6-adenosine deaminases that act on RNA substrates with double-stranded character (Figure 1(a)). A-to-I editing occurs in several eukaryotes including *Caenorhabditis* (nematode worm), *Drosophila* (fruit fly), and *Xenopus* (frog) in addition to mammals and their viruses. Among mammalian RNAs, transcripts in the brain that encode neurotransmitter receptors for L-glutamate and serotonin represent two of the best-understood cases of A-to-I editing. This is illustrated by the α -amino-3-hydroxy-5-methyl-isoxazole-4-propionate (AMPA) glutamate receptor subunit B (GluR-B). Exon 11 in unedited GluR-B transcripts possesses a genome-encoded glutamine (Q) codon (CAG). This exon codes for a hydrophobic domain (TM2) that loops into the cell's cytoplasmic membrane. The edited GluR-B mRNA possesses an arginine (R) codon (CIG) at the same codon position, because I is recognized as G by decoding ribosomes during the translational process. Editing at this site, known as the Q/R site, occurs almost completely in the brain of adult rodents. Q/R editing of GluR-B affects channel subunit assembly in a manner that results in reduced permeability to Ca^{2+} ions. A second A-to-I editing site in GluR RNA, termed the R/G site, involves the conversion of a genome-encoded arginine (R) codon (AGA) to a glycine (G) codon (IGA) in exon 13 of edited GluR-B, -C, and -D subunit transcripts. Editing at the R/G site controls the kinetic

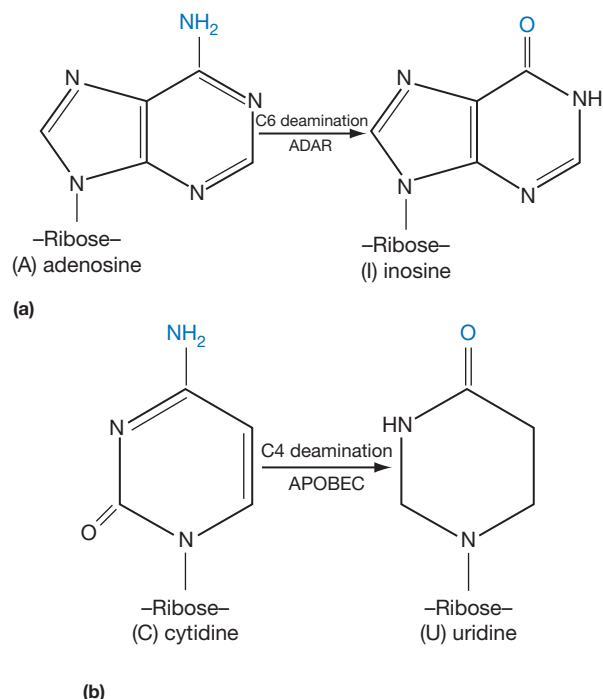


Figure 1 Substitution RNA editing by hydrolytic deamination. (a) A-to-I RNA editing involves the C6 deamination of adenosine to generate inosine catalyzed by ADAR enzymes. (b) C-to-U RNA editing involves the C4 deamination of cytidine to generate uridine catalyzed by APOBEC enzymes.

properties of AMPA receptor channels and leads to faster recovery rates from receptor desensitization.

Editing of GluR-B occurs at the pre-mRNA level within the nucleus, prior to splicing. RNA elements dictate the specificity of editing at the Q/R and R/G sites. The elements consist of unique *cis*-acting complementary inverted repeat sequences that are present in the intron adjacent to the exon possessing the edited adenosine. For the Q/R site in exon 11, the inverted repeat in intron 11 is predicted to form an imperfect duplex structure spanning the edited A at the Q/R site. A similar duplex RNA structure also exists within exon 13 containing the R/G site and the proximal portion of the adjacent intron 13. These exon-intron double-stranded RNA structures are absolutely essential for efficient editing at the Q/R and R/G sites of GluR-B.

The serotonin (5-hydroxytryptamine)-2C receptor (5-HT_{2c}R) provides another example of a neurotransmitter receptor whereby A-to-I editing of the pre-mRNA prior to splicing leads to amino acid substitutions in the encoded protein that alter activity. In the case of rat 5-HT_{2c}R pre-mRNA, four selective A-to-I editing sites denoted A, B, C, and D have been identified in exon 3 that specify three amino acid substitutions. The 5-HT_{2c}R genomic sequence specifies the transcript sequence ...AUA CGU AAU CCU AUU... that encodes ...*ile*-*arg*-*asn*-*pro*-*ile*... in the unedited RNA. Editing of the four sites leads to the sequence ...IUI CGU AIU CCU IUU... in fully edited RNA that encodes ...*val*-*arg*-*ser*-*pro*-*val*... in the receptor protein product. The three amino acid substitutions shown in italics created by A-to-I editing of 5-HT_{2c}R transcripts occur in the putative intracellular loop II of the receptor and they cause

a reduction in G-protein-mediated signaling. Here again, the selective editing of 5-HT_{2c}R pre-mRNAs is dictated by imperfect inverted repeat sequence that forms an RNA duplex between the 3'-end of exon 3 spanning the four editing sites and the 5'-end of the adjacent intron-3.

One of the challenges of A-to-I editing is to identify additional substrates. The GluR-B and 5-HT_{2c}R substrates in rodents and humans were identified serendipitously, and indeed so far such A-to-I editing events that give rise to recoding of translated sequences appear to be rare. Bioinformatic approaches for analysis of the human transcriptome, using cDNA and expressed sequence tag databases and the human genome sequence, identified several thousand candidate A-to-I editing sites. Most of these sites occur in noncoding regions of RNA, often in short repetitive Alu elements of intronic and 3'-untranslated regions of RNAs. The functional significance of A-to-I editing of Alu elements is not yet fully understood.

Editing sites have also been identified in another class of noncoding RNA, the microRNAs (miRs) and their precursors. In RNA interference (RNAi), miRs function as regulatory RNAs that can downregulate with high specificity the production of a targeted protein gene product through either mRNA degradation or translational suppression. A-to-I RNA editing can affect RNAi at two levels. A-to-I editing can impair the processing of miR precursor RNAs by the Dicer and Drosha nucleases of the RNAi machinery, thereby altering the amount of miR RNA produced. A-to-I editing can also alter the sequence of the mature miR product by substitution of I (G) for A, thereby leading to a different mRNA targeting in the RISC complex.

Adenosine Deaminases That Act on RNA

A-to-I editing events of substrates such as GluR-B and 5-HT_{2c}R are catalyzed by a multigene family of enzymes known as ADARs adenosine deaminases that act on RNA. Three mammalian genes are known to encode ADARs, two of which are catalytically active deaminases. ADAR1 and ADAR2 efficiently deaminate both synthetic and naturally occurring double-stranded RNA substrates. The ADAR1 and ADAR2 enzymes possess different activities for the different sites in the GluR-B and 5-HT_{2c}R transcript RNAs. ADAR2 deaminase efficiently edits the Q/R site of GluR-B, while the ADAR1 enzyme possesses little editing activity at Q/R. Both ADAR1 and ADAR2, however, are capable of editing the R/G site of GluR-B. For 5-HT_{2c}R pre-mRNA transcripts in rat brain, ADAR1 very efficiently edits the A site but is inactive for the D site, whereas ADAR2 efficiently edits the D site. ADAR1 and ADAR2 possess multiple copies of a canonical double-stranded RNA-binding motif (R, dsRBM) that are located in the central region of the ADAR proteins (Figure 2). The interferon inducible form of ADAR1 possesses, in addition to three dsRNA-binding motifs, two copies of a Z-DNA-binding motif (Z α and Z β) in the N-terminal region of the protein. The dsRNA- and Z-DNA-binding motifs of ADAR1 are distinct from the deaminase catalytic domain located in C-terminal region of the ADAR proteins. The dsRNA-binding motifs are presumed to play a role in the binding of substrate RNAs, but it is also possible that they play a role in regulation of enzymatic activity. For

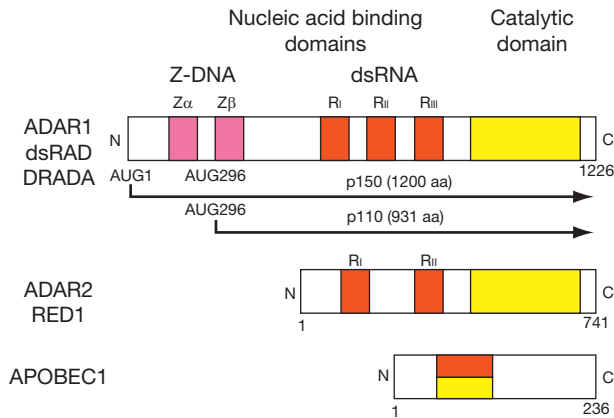


Figure 2 Schematic structure of human RNA-specific deaminases. ADAR1 and ADAR2 possess double-stranded RNA-binding (R) motifs (red) and Z-DNA-binding motifs (pink). The RNA-binding motif of APOBEC1 overlaps with the catalytic domain. The catalytic domains of ADAR and APOBEC are shown in yellow. Two size forms of ADAR1 are expressed, the larger p150 form that is interferon inducible and the smaller constitutively expressed p110 form. While ADAR2 and APOBEC1 are not known to be regulated by cytokines or growth factors, both APOBEC3G and 3F are interferon inducible.

example, adenovirus VA₁ RNA, a small and highly structured viral gene product transcribed by cellular RNA polymerase III, efficiently antagonizes ADAR deaminase activity.

The expression of ADAR enzymes in cells is a complex process. For example, in the case of the *ADAR1* gene in humans and mice, alternative promoters including one inducible by interferons together with alternative exon 1 splicing give rise to two differently sized enzyme isoforms, called p150 and p110 (Figure 2). The interferon inducible form of ADAR1, the p150 or long form, is found in both the cytoplasm and nucleus of cells. The constitutive p110 or short form, which lacks the N-terminal 295 amino acids that includes the Z α -binding motif, is found predominantly if not exclusively in the nucleus. Additional ADAR protein multiplicity is achieved by alternative splicing involving downstream exons, both in the case of ADAR1 and ADAR2. The full physiological significance of the splice variants of ADAR is not known; one possibility may relate to the substrate selectivity of the enzymes.

Cytidine-to-Uridine Substitution RNA Editing

RNA Substrates That Undergo C-to-U Editing

Cytidine (C)-to-uridine (U) substitution RNA editing is catalyzed by C4-cytidine deaminases that act on single-stranded RNA substrates (Figure 1(b)). C-to-U editing has been characterized in both plants and mammals. One of the best-understood examples is the mammalian mRNA transcript that encodes apolipoprotein B (apoB) that functions in lipid metabolism. There are two forms of apoB. The larger form (apoB100) is synthesized in the liver and assembled into lipoprotein particles that are secreted into the bloodstream and are important for transport of synthesized triglycerides and cholesterol. The carboxyl terminal region of apoB100 interacts with the low-density lipoprotein

receptor to remove LDL from circulation. Editing of apoB mRNA by deamination of a single cytidine changes a genome-encoded glutamine codon (CAA) to a translation termination codon (UAA). This C-to-U editing occurs in the small intestine but not the liver. ApoB RNA editing causes the synthesis of a truncated intestinal lipoprotein known as apoB48 that lacks the C-terminal region of apoB100, but otherwise is identical to the N-terminal part (48%) of apoB100. While apoB100 is associated with LDL protein particles, apoB48 lacks receptor-binding activity and is important for transport of dietary lipids.

Editosome Complex

C-to-U editing of apoB mRNA occurs in the nucleus and is selective for spliced transcripts rather than pre-mRNAs. Deamination is mediated by a multiprotein complex called an editosome. The editosome includes a protein that possesses catalytic activity (Apolipoprotein B Editing Catalytic subunit 1, APOBEC-1) together with one or more associated proteins that include an apobec-1 competence factor (ACF), which possesses RNA recognition activity. Unlike the ADAR enzymes, APOBEC does not possess a separate canonical RNA-binding motif, but rather RNA-binding activity appears to overlap with the catalytic domain residues (Figure 2). A tripartite RNA motif that includes efficiency elements, spacer, and a mooring sequence surrounds nucleotide C6666 of apoB mRNA and appears sufficient to confer the required specificity of the C-to-U editing reaction of ApoB RNA. ACF protein binds apoB RNA selectively at the region of the editing site and interacts with APOBEC-1 deaminase to mediate C-to-U editing activity. Multiple splice variants of the ACF protein are found in different human tissues, but their functional significance remains unknown. Similarly, several homologs of the APOBEC-1 have been identified, including ARCD1/APOBEC-2, AID, and phorbolin/ARCD2-7/APOBEC-3A to G.

The existence of multiple APOBEC-1 homologs and ACF variants suggests additional substrates and physiologic roles for C-to-U editing beyond that apoB mRNA. Two examples include C-to-U RNA editing by APOBEC-1 associated with a malignant phenotype, and dC-to-dU editing of viral DNA by APOBEC-3 cytidine deaminases that act as DNA mutators. In the example of the cancer, the cellular mRNA for the neurofibromatosis type 1 (NF1) protein encoded by NF1 tumor suppressor gene undergoes selective C-to-U RNA editing. The encoded NF1 protein, neurofibromin, functions as a Ras-GTPase-activating protein (GAP). Site-specific deamination of a cytidine (nucleotide C3916) in NF1 mRNA converts a genome-coded arginine codon (CGA) to a translation termination codon (UGA). The shortened NF1 open reading frame that results from this C-to-U editing has been observed in a subset of peripheral nerve sheath tumors from NF1 patients and is predicted to give rise to a truncated neurofibromin protein product deficient in GAP activity.

Substitution Editing of Viral RNAs

RNA editing may play a significant role in viral pathogenesis and affect the host response to viral infection. A number of viral RNAs, including those of avian leukosis virus (ALV), hepatitis delta virus (HDV), hepatitis C virus (HCV), human

herpes virus 8 (Kaposi sarcoma virus), lymphocytic choriomeningitis virus (LCMV), measles virus, and polyoma virus all show codon changes under certain infection conditions that are consistent with A-to-I editing. Among the best-characterized viral RNA substrates that undergo site-selective editing is that of HDV. Both ADAR1 and ADAR2 convert an amber translation stop codon (UAG) to a tryptophan codon (UIG) in HDV antigenomic RNA. This editing allows for the synthesis of two delta virus antigen forms, a short form used during viral replication and a long form necessary for assembly of viral particles. A-to-I editing of viral RNA genomes can also occur at multiple sites during lytic and persistent infections, giving rise to hypermutations of viral RNAs. One example of such hypermutations characteristic of A-to-I editing is that of measles virus (MV), where viral RNAs isolated from brains of persistently infected subacute sclerosing panencephalitis (SSPE) patients possess multiple sequence differences from the viral RNA genome that correspond to A-to-G and U-to-C substitution mutations. Because MV, a negative-strand RNA virus, replicates in the cytoplasm, it is presumed that ADAR1 is the responsible enzyme. A-to-I hypermutations are also seen in viral RNAs of polyoma virus, HCV and LCMV. The DNA virus vaccinia that replicates in the cytoplasm of cells encodes a protein product (E3L) that efficiently antagonizes ADAR1 deaminase activity, and mutations in E3L are known to affect viral pathogenesis in the mouse model.

Finally, C-to-U RNA editing like A-to-I editing may play a role in the interactions between viral pathogens and their hosts. Among the homologs of APOBEC-1, the APOBEC-3 family members, particularly 3G and 3F, function as innate cellular restriction factors with broad anti-retroviral activity. These deaminases catalyze cytidine deamination of retrovirus reverse transcripts that correspond to the negative-sense single-strand cDNA in infected cells. The resultant dC-to-dU deaminations lead to dG-to-dA hypermutations in viral positive-sense cDNA and the subsequent inhibition of virus growth. Although both 3G and 3F APOBECs restrict retrovirus infection, they display somewhat different sequence context specificity for deamination of deoxycytidine. Primate lentiviruses including human immunodeficiency virus (HIV-1), a retrovirus that is the cause of acquired immune deficiency syndrome (AIDS), encode a regulatory viral protein called vif that blocks the antiviral activity of the cellular APOBEC-3G and -3F proteins. Like ADAR1 p150, APOBEC-3G and -3F are both inducible by interferon. Alteration of the amount of APOBEC deaminase, accessory protein, or viral antagonist could modulate RNA editing or DNA mutator activity within cells.

Editing by Nucleotide Insertion/Deletion

A second general type of RNA editing, in addition to base substitution editing, is nucleotide insertion/deletion editing. Among the best-characterized examples of insertion/deletion editing is that of U-insertion and U-deletion. Uridine insertion/deletion editing occurs by an enzymatic cleavage-ligation mechanism. The specificity of U-insertion/deletion editing is determined by RNA-RNA base pairing between a complementary guide RNA (gRNA) and the target RNA (mRNA) substrate. The base-paired

RNA complex serves to position the enzymatic machinery at the site of U insertion or deletion, thereby guiding the modification process. Much has been learned from the trypanosome mitochondrial system about the features and requirements of the complementary gRNAs that act in trans to target enzyme complexes to the mRNA site of uridine insertion/deletion editing. The enzymatic cleavage-ligation mechanism of U insertion/deletion editing involves hybridization of gRNA to mRNA to form a gRNA:mRNA anchor duplex, after which the mRNA is specifically cleaved by a gRNA-dependent ribonuclease at the site of the gRNA:mRNA mismatch. For U-insertion editing, a 3'-terminal uridylyl transferase (TUTase) then adds U's to the 3'-end of the 5'-fragment guided by base pairing with the gRNA, followed by ligation of the two fragments by an RNA ligase. For U-deletion editing, a U-specific 3'-5'-exonuclease removes the bulged U residues, followed by ligation of the two RNA fragments. Finally, in addition to U-insertion editing as exemplified by the cytochrome oxidase mRNA in trypanosome mitochondria, C-insertion editing and G-insertion editing are also known to occur as illustrated by Physarum mitochondria and by negative-strand RNA viruses, respectively. In the case of paramyxoviruses including measles virus and Sendai virus, G-nucleotide insertion occurs during the RNA transcription process to generate edited P/V/C gene transcripts that include nontemplated Gs.

Roles of RNA Editing

Altering the information transfer process at the posttranscriptional level of gene expression by RNA editing represents an important strategy for amplifying genetic diversity and

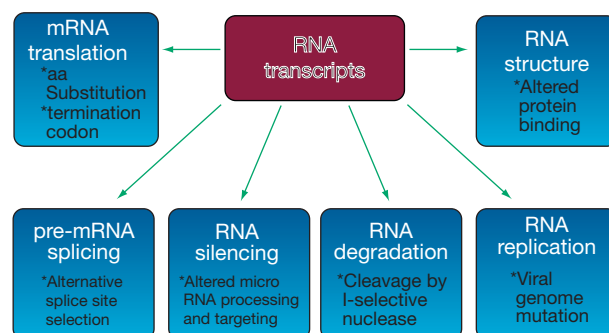


Figure 3 Established and potential roles of RNA editing by deamination. Edited RNA transcripts (A-to-I; C-to-U) possess different sequences than their unedited transcript counterparts. Hence, edited RNAs may display different functional activities than the corresponding unedited transcripts. Editing may alter processes including mRNA translation, by changing mRNA codons and hence proteins encoded by them. Editing may alter pre-mRNA splicing patterns by changing splice site-recognition sequences. Editing may alter micro-RNA production or activity, by changing RNA structure or gene targeting sequence recognition. Editing may affect RNA degradation by modifying RNA sequences involved in nuclease recognition. Editing may affect viral RNA genome genetic stability by changing template and hence product sequences during RNA replication. Finally, editing potentially may affect RNA structure-dependent activities that entail binding of RNA by proteins. Adapted from Samuel CE (2003) RNA editing minireview prologue. *Journal of Biological Chemistry* 278: 1389–1390.

modifying the functions of products encoded by an organism's genome. Some of the mechanistic roles that substitution editing through A-to-I and C-to-U deamination play, or are anticipated to play, in biological processes are summarized in [Figure 3](#). These biochemical mechanisms include effects of editing on mRNA translation, pre-mRNA splicing, RNA silencing, RNA degradation, RNA replication, and RNA structure that result from nucleotide substitution via deamination or nucleotide insertions. Site-specific editing may change the coding potential of mRNA transcripts, leading to proteins with altered function due to amino acid substitutions following A-to-I editing. This is exemplified by the GluR-B and serotonin-2c receptor proteins. Introduction or removal of translation termination codons may also occur, as exemplified by apoB and NF1 mRNAs where C-to-U editing generates UAA and UGA translation termination codons respectively, and HDV RNA where A-to-I editing converts an amber UAG stop to a tryptophan codon. ADAR2 edits its own transcript, to create an alternative splice acceptor site. A-to-I editing can affect RNA interference responses, either by altering the production process and hence amount of mature microRNA or by generating a microRNA with an altered targeting sequence. A novel ribonuclease selective for inosine-containing RNAs has also been described, consistent with the notion that A-to-I edited transcripts may be degraded preferentially relative to the unedited transcripts. For viral RNAs, deaminations in either the plus- or minus-strand RNA could lead to mutations in the encapsidated viral genome sequence. Finally, sequence changes resulting from editing may subsequently affect RNA structure and function, including protein–RNA interactions such as those necessary for protein kinase PKR ([Figure 3](#)).

See also: [Cell Architecture and Function: Nuclear Pores and Nuclear Import/Export](#); [Molecular Biology: Pre-tRNA and Pre-rRNA Processing in Bacteria](#); [tRNA Synthetases](#).

Further Reading

- Blanc V and Davidson NO (2003) C-to-U RNA editing: Mechanisms leading to genetic diversity. *Journal of Biological Chemistry* 278: 1395–1398.
- Decatur WA and Fournier MJ (2003) RNA-guided nucleotide modification of ribosomal and other RNAs. *Journal of Biological Chemistry* 278: 695–698.
- Gott JM and Emeson RB (2000) Functions and mechanisms of RNA editing. *Annual Review of Genetics* 34: 499–531.
- Jepson JE and Reenan RA (2008) RNA editing in regulating gene expression in the brain. *Biochimica et Biophysica Acta* 1779: 459–470.
- Malim MH (2009) APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philosophical Transactions of the Royal Society B* 364: 675–687.
- Nishikura K (2010) Functions and regulation of RNA editing by ADAR deaminases. *Annual Review of Biochemistry* 79: 321–349.
- Samuel CE (2001) Antiviral actions of interferons. *Clinical Microbiology Reviews* 14: 778–809.
- Samuel CE (2003) RNA editing minireview prologue. *Journal of Biological Chemistry* 278: 1389–1390.
- Seeburg PH and Hartner J (2003) Regulation of ion channel/neurotransmitter receptor function by RNA editing. *Current Opinion in Neurobiology* 13: 279–283.
- Simpson L, Sbicego S, and Aphasizhev R (2003) Uridine insertion/deletion RNA editing in trypanosome mitochondria: A complex business. *RNA* 9: 265–276.
- Toth AM, Zhag P, Das S, George CX, and Samuel CE (2006) Interferon action and the double-stranded RNA-dependent enzymes ADAR1 adenosine deaminase and PKR protein kinase. *Progress in Nucleic Acid Research and Molecular Biology* 81: 369–434.

RNA Polymerase I and RNA Polymerase III in Eukaryotes

R J White, Beatson Institute for Cancer Research, Glasgow, UK

Published by Elsevier Inc.

This article is reproduced from the previous edition, volume 3, pp. 759–762, © 2004, Elsevier Inc.

Glossary

Chromatin The DNA wrapped up with proteins, as found in chromosomes.

Promoter A stretch of DNA located near the transcription start site that influences the efficiency of transcription.

RNA polymerase The enzyme responsible for synthesizing RNA.

Transcription The process using DNA as template to synthesize a complementary RNA.

Transcription factor Protein that influences the ability of RNA polymerase to carry out transcription.

RNA polymerases are the enzymes responsible for synthesizing RNA, in a process called 'transcription'. They use DNA as a template, and the RNA they produce carries a faithful copy of the genetic information contained in the template. Whereas bacteria use a single RNA polymerase to transcribe all their genes, the nuclei of eukaryotes contain three. Each is responsible for transcribing a unique set of essential genes. RNA polymerase I synthesizes most of the ribosomal RNA (rRNA), RNA polymerase II synthesizes the protein-encoding messenger RNA (mRNA) and most small nuclear RNA (snRNA), while RNA polymerase III synthesizes a variety of small untranslated RNAs, including transfer RNA (tRNA) and 5S rRNA. An additional RNA polymerase is present in mitochondria, which carry a small DNA molecule of their own. This division of labor could have become necessary as genomic complexity increased during evolution. In actively growing cells, RNA polymerase I can be responsible for ~70% of all nuclear transcription, whereas RNA polymerases II and III contribute ~20% and 10%, respectively.

Genes Transcribed by RNA Polymerases I and III

Despite its very high activity, RNA polymerase I only transcribes a single type of gene, which encodes the large rRNA precursor molecule. This precursor is then processed into the mature 5.8S, 18S, and 28S rRNA components of ribosomes. A single molecule of each of these rRNAs, along with a 5S rRNA molecule and ~85 proteins, is incorporated into each ribosome. The pre-rRNA is encoded by multiple gene copies, the number of which varies between different organisms; mammals have ~150–200 copies arranged in tandem head-to-tail arrays.

The genes transcribed by RNA polymerase III are invariably very short, rarely exceeding 300 bp in length. Many of their products have essential functions, such as the 5S rRNA and tRNAs (required for protein synthesis), 7SL RNA (involved in intracellular trafficking of proteins as part of the signal recognition particle), and the U6, multidrug resistance-associated protein (MRP), and H1 RNAs (involved in the processing of mRNA, rRNA, and tRNA, respectively). Mammalian genomes also contain large numbers of short interspersed repeats

(SINEs) that are transcribed by RNA polymerase III, such as the Alu family in man. In addition, certain viruses carry genes that are transcribed by RNA polymerase III, such as the VA genes of adenovirus and the EBER genes of Epstein–Barr virus. The products of these viral genes can subvert the translational machinery of an infected cell, diverting it toward the more effective production of viral proteins.

Nucleoli

Within the nucleus, the most obvious landmarks under a light microscope are nucleoli, because of their high density. These form around the actively transcribed rRNA gene repeats and contain closely packed molecules of transcribing RNA polymerase I, each with a nascent pre-rRNA attached. Furthermore, ribosomal proteins assemble on the nascent transcript while it is being made. Although nucleoli have the appearance of organelles, they are in fact dynamic structures without membranes that disappear when transcription stops (e.g., during the general suppression of gene expression that accompanies mitosis). Transferring rRNA genes to a novel chromosomal location can result in the appearance of an ectopic nucleolus. When cells become very old, their nucleoli fragment; it has been suggested that nucleolar degeneration may be an important aspect of cellular aging.

The RNA Polymerase Molecules

RNA polymerases I, II, and III are large and complex proteins, consisting of 14, 12, and 17 subunits, respectively. Five of these subunits are shared, and a further two are found in RNA polymerases I and III, but not in RNA polymerase II. The two largest subunits of each enzyme are unique, but share substantial homology with each other and also with bacterial RNA polymerases; they fold together to provide the catalytic site of the enzyme. Besides these, RNA polymerases I, II, and III each have an additional set of unique subunits. Although most of the subunits are required for cell viability, their precise roles remain poorly characterized.

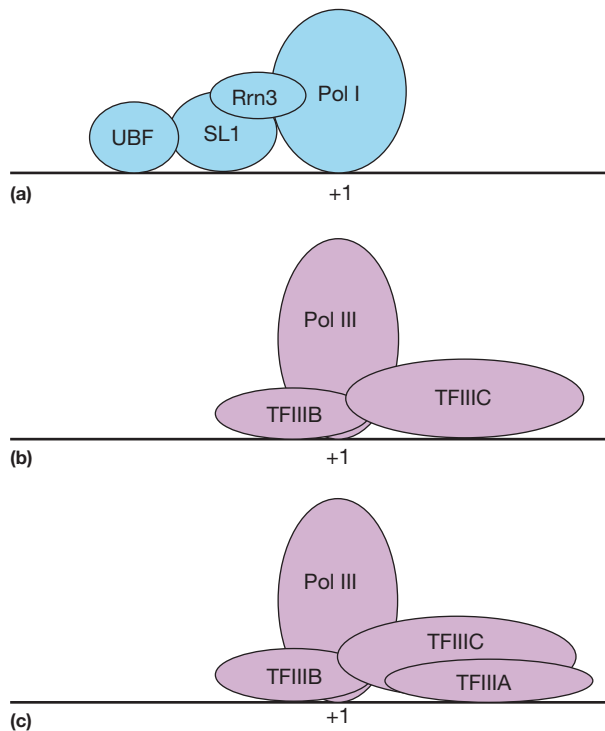


Figure 1 Simplified model of the transcription complexes used by RNA polymerases I and III. (a) UBF and SL1 assemble at the promoters of large rRNA genes and then recruit a preassembled complex containing Rrn3 and RNA polymerase I (Pol I), as well as additional factors not shown. (b) TFIIC binds directly to the promoters of most genes transcribed by RNA polymerase III, for example, tRNA genes. TFIIB is recruited by protein/protein interactions with TFIIC and then serves to bring RNA polymerase III (Pol III) to the transcriptional start site. (c) Similar events occur at 5S rRNA genes, but these are initiated by TFIIA, which binds promoter DNA and provides a scaffold for TFIIC assembly. +1 denotes where transcription initiates.

The RNA Polymerase I and III Transcription Apparatus

Despite their complexity, the nuclear RNA polymerases require considerable assistance to locate their genetic templates. Sets of transcription factor proteins assemble at promoter DNA sequences near the initiation sites of genes and then serve as molecular beacons which recruit the appropriate RNA polymerase.

For the large rRNA genes, promoter recognition is achieved by the homodimeric factor upstream binding factor (UBF) that binds across the promoter region upstream of the transcription start site. UBF bends the DNA and wraps it around itself to create a three-dimensional nucleoprotein structure that is recognized by SL1, the factor responsible for polymerase recruitment. RNA polymerase I is preassembled with a polypeptide called Rrn3 that serves as an essential bridge between the polymerase and SL1 (**Figure 1(a)**). Several additional transcription factors are also associated with RNA polymerase I and are recruited with it in a concerted manner. Although these principles apply to mammals and frogs, the machinery required for rRNA synthesis is rather different in yeast. Indeed, the RNA polymerase I transcription apparatus is much less well

conserved through evolution than those of RNA polymerases II or III; whereas human proteins can transcribe a yeast tRNA gene, they will not even recognize the large rRNA genes from mouse.

In the case of RNA polymerase III, two key transcription factors, TFIIB and TFIIC, play central roles in both yeast and man (**Figure 1(b)**). The former is responsible for recruiting RNA polymerase III and positioning it over the transcription initiation site; it also helps separate the two strands of the DNA helix, thereby facilitating the start of RNA synthesis. However, TFIIB is unable to recognize the DNA sequences of most RNA polymerase III promoters, but must be brought into position through protein-protein interactions with DNA-bound TFIIC. An unusual feature of this system is that the promoter sequences recognized by TFIIC are located downstream of the transcription start site, within the transcribed region. Although in most cases TFIIC binds directly to DNA, this is not possible for the 5S rRNA genes which have a distinct promoter arrangement. In these cases, an additional polypeptide, TFIIA, binds along the DNA and provides a protein scaffold to which TFIIC binds (**Figure 1(c)**). TFIIA contains nine tandem copies of the zinc finger DNA-binding domain. Indeed, it was the founder member of the zinc finger family, the largest group of transcription factors in organisms ranging from yeast to man (~900 members in humans).

Alu Genes and Retrotransposition

The human genome carries over a million copies of the Alu gene, which provides ~10% of our total DNA. Despite their abundance, only small numbers of Alu genes are actually transcribed, the remainder being silenced through epigenetic mechanisms, including methylation and incorporation into repressive chromatin. SINES propagate by retrotransposition, in which their RNA transcripts are copied into DNA and then inserted into novel genomic sites. These insertions occur preferentially into areas of active transcription, perhaps because of their more open and accessible chromatin structure. This increases the danger that integrating SINES will disrupt genes at the recipient loci, thereby having a mutagenic effect. Indeed, cases of haemophilia, neurofibromatosis, and breast cancer are believed to have arisen in this way. It is estimated that Alu insertions account for ~0.1% of all human genetic disorders. By contrast, retrotransposition is much more active in mice, where it may contribute ~10% of mutations.

RNA Polymerases I and III and Tumor Suppression

The RNA polymerase I and III transcription systems are both targeted directly by two key tumor-suppressor products, p53 and retinoblastoma (RB). Both these proteins bind to TFIIB and block its interactions with TFIIC and RNA polymerase III, thereby repressing transcription. RB also binds to UBF to inhibit synthesis of large rRNA, whereas p53 inactivates SL1. The ability of RB and p53 to suppress the output of RNA polymerases I and III may contribute to their capacity to restrain cell growth. RB has been described as a gatekeeper that only allows cells to grow and proliferate when conditions

are propitious. Inherited mutations in RB can cause the pediatric eye cancer RB. Indeed, it is widely believed that RB function must be compromised in some way before any tumor can develop. The most commonly mutated gene in human cancer is p53, which is famous as the guardian of the genome. If a cell's DNA gets damaged, for example, through exposure of skin to excessive doses of UV light, p53 gets activated and triggers a protective response; this can involve halting growth and cell-cycle progression until the damage has been mended or, under extreme circumstances, apoptotic death. These responses ensure that DNA-carrying potentially harmful mutations are not passed on to further generations of cells.

Links between Human Disease and RNA Polymerases I and III

RNA polymerases I and III together may contribute ~80% of nuclear transcription and provide ~95% of the total RNA in cells. This very high level of production is necessary to make the several million new ribosomes that are needed per generation to maintain protein synthetic capability. Because they are such major determinants of biosynthetic capacity, the activities of RNA polymerases I and III are closely linked to the rate of cellular growth. In hyperproliferative diseases such as cancer, there is generally an elevated rate of transcription by these polymerases. Indeed, pathologists use enlarged nucleoli (a sign of accelerated rRNA synthesis) as a diagnostic indicator of tumor formation. To a large degree, the elevated activity of RNA polymerases I and III in cancer cells is likely to reflect the inactivation of p53 and/or RB; since these tumor suppressors play a major role in restraining transcription by RNA polymerases I and III in healthy cells, their inactivation during cancer development will remove a major restraining influence and allow deregulated gene expression. These effects must be extremely widespread, since p53 is mutated in ~50% of human tumors and RB function may be compromised in all forms of malignancy. In addition to their release from these transcriptional repressors, the output of RNA polymerases I and III is also increased in some cancers through the action of certain oncogenic proteins. An example is c-Myc, which is deregulated in many tumor types and is believed to contribute to one in seven US cancer deaths. TFIIB is bound and activated by c-Myc, leading to elevated rates of RNA polymerase III transcription. Certain types of cancer also produce abnormally high levels of the transcription factors used by RNA polymerases I and III. For example, UBF is frequently overexpressed

in hepatocellular carcinomas, whereas TFIIC is unusually abundant in ovarian tumors.

Epstein–Barr virus is associated with gastric and nasopharyngeal carcinomas, Burkitt's lymphoma, Hodgkin's disease, and acquired immuno-deficiency syndrome (AIDS)-associated lymphoma. Its most highly expressed gene products are the EBER RNAs, which are synthesized by RNA polymerase III. The presence of these transcripts is used diagnostically as a test for Epstein–Barr virus. Indeed, very few other viral genes are expressed in Burkitt's lymphomas. Recent studies have shown that the EBER1 transcript can be sufficient for oncogenic transformation, a remarkable fact since it is only 167 nucleotides long. This provides the first reported example of an untranslated RNA that can transform cells.

An inherited-developmental disorder has been shown to result from mutations in the MRP gene that is transcribed by RNA polymerase III to produce a short RNA that is required for mitochondrial replication and for the processing of large rRNA. Families afflicted with this condition suffer from cartilage–hair hypoplasia, a complex syndrome involving short stature (adult heights of 104–149 cm), abnormal cartilage, and hypoplastic hair, sometimes with a compromised immune system. This disorder is especially common amongst the Old Order Amish of Pennsylvania.

See also: **Molecular Biology: Nucleolus, Overview; RNA Polymerase II and Its General Transcription Factors.**

Further Reading

- Batzar MA and Deininger PL (2002) Alu repeats and human genomic diversity. *Nature Reviews Genetics* 3: 370–380.
- Carmo-Fonseca M, Mendes-Soares L, and Campos I (2000) To be or not to be in the nucleolus. *Nature Cell Biology* 2: E107–E112.
- Felton-Edkins ZA, Kenneth NS, and Brown TRP (2003) Direct regulation of RNA polymerase III transcription by RB, p53, and c-Myc. *Cell Cycle* 2: 181–184.
- Geiduschek EP and Kassavetis GA (2001) The RNA polymerases II transcription apparatus. *Journal of Molecular Biology* 310: 1–26.
- Moss T and Stefanovsky VY (2002) At the center of eukaryotic life. *Cell* 109: 545–548.
- Paule MR (1998) *Transcription of Ribosomal Genes by Eukaryotic RNA Polymerase I*. Austin: Landes Biosciences.
- Paule MR and White RJ (2000) Transcription by RNA polymerases I and III. *Nucleic Acids Research* 28: 1283–1298.
- Schramm L and Hernandez N (2002) Recruitment of RNA polymerase III to its target promoters. *Genes and Development* 16: 2593–2620.
- White RJ (2001) *Gene Transcription: Mechanisms and Control*. Oxford: Blackwell Science.
- White RJ (2002) *RNA Polymerase III Transcription*. Austin: Landes Biosciences.

RNA Polymerase II and Its General Transcription Factors

J L Corden, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Published by Elsevier Inc.

This article is a revision of the previous edition article by Joan W. Conaway and Ronald C. Conaway, volume 3, pp. 763–765, © 2004, Elsevier Inc.

Glossary

Basal transcription factors Minimal set of auxiliary transcription factors required to support promoter-specific transcription initiation by polymerase II (pol II).

DNA helicase Enzyme that uses the energy of adenosine triphosphate (ATP) hydrolysis to unwind the two strands of DNA.

DNA template DNA that is transcribed by RNA polymerases.

Promoter DNA sequences that direct RNA polymerases to transcribe genes.

RNA polymerase Enzyme that catalyzes synthesis of RNA.

Synthesis of eukaryotic messenger RNA (mRNA) is catalyzed by RNA polymerase II (pol II). A set of five general transcription factors (GTFs) designated transcription factor IIB (TFIIB), TFIID, TFIIIE, TFIIF, and TFIIFH interact to promote binding of RNA pol II to gene regulatory sequences termed promoters and assist in the initiation of RNA synthesis.

RNA Pol II Structure and Function

RNA pol II is one of three multi-subunit DNA-dependent RNA polymerases present in all eukaryotic cells. Pol II transcribes mainly protein-coding genes, while pol I and pol III transcribe ribosomal RNA genes and small noncoding RNA genes, respectively. Transcripts formed by pol II in the nucleus are subsequently capped, spliced, polyadenylated, and transported to the nucleus where they are translated by ribosomes to form proteins.

Pol II is a multi-subunit enzyme of about 500 kDa that is highly conserved from simple to complex organisms. The catalytic mechanism of pol II has been determined from examination of high-resolution X-ray crystal structures of *Saccharomyces cerevisiae* pol II.

Rpb1 and Rpb2 are the largest subunits and they form the core of the enzyme with smaller subunits arranged around the surface. The main feature of the enzyme is a deep channel where RNA is synthesized. Approximately 12 bp of DNA are unwound during transcription and the single-stranded template is buried in this central channel where it forms a hybrid with approximately nine nucleotides at the 3'-end of the growing RNA chain. The nontemplate DNA strand is displaced to the surface of the enzyme. This arrangement creates a single-stranded bubble of DNA that slides through and around the central channel of the polymerase as DNA base pairs are progressively unpaired in front of the enzyme and then reform after transcription.

At the base of the channel is the active site of the polymerase that catalyzes the formation of phosphodiester bonds between the growing 3'-end of the transcript and an incoming ribonucleoside triphosphate (NTP). Selection of the correct NTP is determined by base pairing with the single-stranded DNA template in the active site. The catalytic site contains two magnesium ions, one chelated by aspartate residues in the

enzyme and the other by phosphates on the incoming NTP. The magnesium ions facilitate a nucleophilic attack of the 3'-hydroxyl of the transcript on the α -phosphorus of the substrate NTP. Formation of the phosphodiester bond is coupled with release of a pyrophosphate and translocation of the polymerase along the substrate placing the new 3'-hydroxyl in the active site. As the enzyme advances, nascent RNA starting about 10 bases downstream of the growing end is removed from the RNA:DNA hybrid and exits the polymerase through a small channel. The elongating pol II complex is extremely stable because after transcription starts the polymerase undergoes a conformational change clamping the template in the central channel.

Pol II Core Promoters

DNA sequences that signal the start of pol II transcription vary widely from gene to gene with the most common sequences constituting a core promoter that surrounds the transcription start site. The core promoter serves to orient pol II on the DNA and directs accurate initiation. Several different core promoter elements have been identified and some combination of these elements is present on many, but not all, pol II promoters. The TATA box contains the sequence TATA(A/T)A(A/T) and in metazoans is located 25–30 nucleotides upstream of the transcription start site. A second common core promoter element is the initiator (Inr) that contains a pyrimidine-rich consensus sequences YYAN(T/A)YY, where Y is a pyrimidine, N can be any of the four bases A, C, G, or T, and the first A is the transcription start site. The downstream promoter element (DPE) with a consensus (A/G)G(A/T)CGTG is located about 30 nucleotides downstream of the transcription start site. Some promoters contain none of these elements but may contain less common core promoter sequences such as the motif ten element, the downstream core element, or the TFIIB-responsive element.

Core promoters are insufficient for pol II initiation *in vivo* and most genes contain additional regulatory elements for one or more DNA-binding regulatory proteins. These gene-specific activators or repressors control frequency of initiation at the core promoter. Regulatory sequences within a few hundred nucleotides of the core promoter are called promoter proximal

elements, while more distal regulatory sequences, termed enhancers or silencers, can be located thousands of base pairs from the core promoter.

Pol II GTFs

Pol II alone is unable to initiate transcription and requires a set of auxiliary proteins termed GTFs including TFIIB, TFIID, TFIIF, TFIIE, and TFIIH. The GTFs work in a multistep process that begins with the assembly of a pol II pre-initiation complex at the core promoter. In the first step, TFIID binds to the core promoter setting up a DNA–protein complex that directs transcription in the proper orientation starting at the correct site. TFIID is a multi-subunit protein containing the TATA-box binding protein (TBP) and as many as 13 additional proteins designated TBP-associated factors (TAFs). TBP binds directly to the TATA box when present but is able to recognize a wide range of related sequences. TAFs also interact specifically with the core promoter DNA with TAF2 binding the Inr and TAF6 and TAF9 recognizing the DPE. In the next step of pre-initiation complex assembly, TFIIB recruits pol II to the TFIID–core promoter complex. TFIIB is a single polypeptide with several domains that interact independently with pol II and TFIID. The TFIID–TFIIB–polII complex is stabilized by the binding of TFIIF, which is required for the subsequent binding of TFIIE and then TFIIH to form the final pre-initiation complex. TFIIH is a nine-subunit factor containing both helicase and protein kinase activities.

Mechanism of Pol II Initiation

The fully assembled pre-initiation complex contains about 40 polypeptides bound on the core promoter DNA. The multiple DNA-binding proteins in this complex result in the DNA bending at the site of TBP binding, thereby positioning TFIIH near the start of transcription. A helicase activity present in TFIIH uses adenosine triphosphate (ATP) to unwind the core promoter DNA, resulting in a short segment of single-stranded template DNA entering the pol II active site channel. TFIIB interacts with this DNA, stabilizing the transcription bubble and assisting the positioning of the transcription start site in the pol II active site. Pol II then synthesizes the first few phosphodiester bonds resulting in a short transcript that remains bound to the DNA template. After the synthesis of about 10 bases, the 5′-end of the RNA chain is removed from the DNA:RNA hybrid and enters the RNA exit channel. At this point, the function of the GTFs is complete and pol II leaves the pre-initiation complex and begins the process of elongation. The nascent 5′ end of the mRNA is capped as it emerges from the RNA exit channel.

Termination of Pol II Transcription

Pol II transcripts generally span a single gene. The elongating pol II eventually reaches the end of the protein-coding region where it transcribes a sequence that signals the cleavage and subsequent polyadenylation of the nascent pre-mRNA. Cleavage of the nascent transcript separates the transcript from the elongation complex, leaving pol II with a short RNA containing an unprotected 5′-end. Exonucleases degrade this un-capped RNA more rapidly than pol II can elongate, thus leading to an unstable elongation complex that dissociates from the template DNA.

Conservation of Pol II and the GTFs

Pol II is one of three multi-subunit RNA polymerases common to all eukaryotes. These three polymerases are highly related to each other and to archaeal and bacterial RNA polymerase, especially in the regions of the largest subunits that constitute the active site. In addition, several of the smaller eukaryotic RNA polymerase subunits are shared by more than one type of polymerase, whereas bacterial RNA polymerase uses sigma factors to assist in the initiation reaction eukaryotes and archaea employ GTFs. Archaea use only three GTFs: TBP, TFIIB, and TFIIE. The genomes of eukaryotes encode the full complement of GTFs and, in addition, have additional genes encoding variants of the GTFs. For example, the human genome encodes several variants of TBP that can recognize different types of core promoters.

See also: [Molecular Biology: Gene Expression in Eukaryotes: RNA Polymerase II Structure](#); [RNA Polymerase I and RNA Polymerase III in Eukaryotes](#); [RNA Polymerase Reaction in Bacteria](#); [RNA Polymerase Structure, Bacterial](#); [Sigma Factors](#); [Transcription Termination](#).

Further Reading

- Cramer P, Bushnell DA, and Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292: 1863–1876.
- Gnatt AL, Cramer P, Fu J, Bushnell DA, and Kornberg RD (2001) Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292: 1876–1882.
- Hahn S (2009) Structural biology: New beginnings for transcription. *Nature* 462: 292–293.
- Juven-Gershon T and Kadonaga JT (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental Biology* 339: 225–229.
- Nudler E (2009) RNA polymerase active center: The molecular engine of transcription. *Annual Review of Biochemistry* 78: 335–361.
- Sikorski TW and Buratowski S (2009) The basal initiation machinery: Beyond the general transcription factors. *Current Opinion in Cell Biology* 21: 344–351.
- Thomas MC and Chiang CM (2006) The general transcription machinery and general cofactors. *Critical Reviews in Biochemistry and Molecular Biology* 41: 105–178.

RNA Polymerase II Elongation Control in Eukaryotes

D H Price, University of Iowa, Iowa City, IA, USA

Published by Elsevier Inc.

This article is reproduced from the previous edition, volume 3, pp. 766–769, © 2004, Elsevier Inc.

Glossary

DRB The P-TEFb inhibitor 5,6-dichlororibofuranosyl benzamidazole.

DSIF DRB sensitivity inducing factor, an N-TEF.

NELF Negative elongation factor, an N-TEF.

N-TEFs Negative transcription elongation factors.

P-TEFb Positive transcription elongation factor b composed of Cdk9 and a cyclin partner.

During the 1990s, it became clear that the elongation properties of RNA polymerase II are regulated and the factors responsible play a critical role in controlling gene expression in eukaryotic cells. One set of factors helps RNA polymerase II to negotiate pause and arrest sequences in the template, and to maintain an elongation rate of 1000–1500 nucleotides per minute. The initiation factor TFIIF (transcription initiation factor IIF) reduces RNA polymerase II pause times and stimulates the rate of elongation. The elongation factor S-II or transcription initiation factor IIS (TFIIS) rescues polymerase molecules that are blocked from further elongation at arrest sites. Another set of factors is responsible for an elongation control process that regulates the fraction of initiated polymerases that enter productive elongation to generate messenger RNAs (mRNAs). Negative transcription elongation factors (N-TEFs), such as 5,6-dichlororibofuranosyl benzamidazole (DRB) sensitivity inducing factor (DSIF) and negative elongation factor (NELF), slow the rate of elongation and restrict the polymerase to promoter proximal sequences. The positive transcription elongation factor, P-TEFb, overcomes the effect of the negative factors and allows the production of full-length transcripts. The function of P-TEFb is regulated through interactions with many other transcription factors and the amount of active P-TEFb in the cell is precisely regulated through a reversible inhibition by the small cellular RNA, 7SK.

Historical Perspective

It has been clear for some time that much of the regulation of prokaryotic gene expression is accomplished by controlling the elongation potential of RNA polymerase, but only recently has it become apparent that eukaryotic gene expression is similarly regulated. Early work in John Lis's lab showed that there was an RNA polymerase II molecule engaged in transcription of the major *Drosophila* heat shock gene, HSP70, under conditions where no full-length mRNAs were being produced. Only after heat shock were the polymerases allowed to continue elongation and produce mRNAs. Similarly, regulated premature termination of transcription was seen on a number of viral genes, most notably from human immuno-deficiency virus (HIV), as well as on cellular genes, including oncogenes such as c-myc and c-fos. In the early 1990s, *in vitro* systems were beginning to reproduce the elongation control process and ultimately these

systems allowed the identification and purification of some of the specific factors involved. The current model of RNA polymerase II elongation control has the look and feel of models of prokaryotic antitermination, but there are important differences in mechanistic details.

Positive and Negative Elongation Factors

Elongation Maintenance Factors

During transcription, RNA polymerase II is aided by specific factors as it encounters numerous blocks to elongation inherent in a template sequence. During normal transcription *in vivo*, each nucleotide addition to an RNA chain requires ~50 ms on average. However, *in vitro* RNA polymerase II may pause for seconds or minutes at specific sites in the absence of any accessory factors. TFIIF, a factor required for RNA polymerase II initiation, can also decrease the time that the polymerase spends at pause sites. This has the effect of increasing the overall elongation rate because, for the most part, the elongation rate is determined by how long the polymerase stops at the strongest pause sites. It has been hypothesized that the mechanism utilized by TFIIF involves an interaction-induced change in the polymerase from the paused conformation to the elongation competent form. Other factors that belong in this same class are eleven-nineteen lysine-rich leukemia and elongin, but TFIIF has the most dramatic elongation stimulatory activity. The elongation factor S-II functions by a completely different mechanism. At some sites a fraction of RNA polymerase II molecules fall into an arrested conformation from which they cannot escape unaided. At these sites the 3' end of the nascent transcript is removed from the active site of the polymerase as the polymerase backslides along the template while maintaining an RNA:DNA hybrid. When this happens, the polymerase may remain engaged for hours or days without extending the transcript. S-II stimulates an intrinsic ribonuclease activity of the polymerase that removes the unpaired 3' end of the transcript and puts the new 3' end in register with the active site of the polymerase. This reactivates elongation and the polymerase has a second chance to pass the arrest site. Eventually, all polymerases can pass the arrest site, although many may require several rounds of S-II-mediated transcript cleavage. Through the combined function of both classes of factors, the elongation rate of RNA polymerase II is maintained between 1000 and 1500 nucleotides per minute.

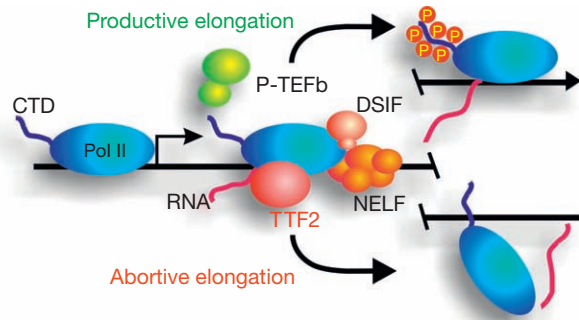


Figure 1 RNA polymerase II elongation control. After initiation, pol II enters abortive elongation by default because of the action of N-TEFs, such as DSIF and NELF. Phosphorylation of the CTD of the large subunit of RNA polymerase II by P-TEFb causes the transition into productive elongation. If P-TEFb does not act, transcription is prematurely terminated by TTF2.

Elongation Control

In addition to the factors that maintain efficient elongation, there is another set of factors responsible for regulating the efficiency of initiated polymerases to reach the 3' end of genes. This process has been termed 'RNA polymerase II elongation control' and is best viewed as an obligatory event mediated by the default action of negative factors, N-TEFs, and the regulated action of a positive factor, P-TEFb (Figure 1). The process was demonstrated *in vitro* by first showing that the transcription inhibitor DRB could block the production of long, but not short, transcripts. Elongation control was further characterized by the separation of N-TEFs, which direct polymerases to generate only short promoter proximal transcripts, from a positive factor, ultimately identified as P-TEFb, which reverses the action of the N-TEFs. A combination of the action of the N-TEFs and the activity of transcription termination factor 2 (TTF2) results in a process called 'abortive elongation' characterized by the premature termination of transcripts before the mature 3' end of the gene is reached. P-TEFb plays a key role in that it is responsible for the transition into productive elongation. Stated in another way, the ability to produce mRNAs is directly linked to the activity of P-TEFb. Elongation control machinery exists in all eukaryotic species that have been examined and the current evidence suggests that P-TEFb is required for transcription of most genes.

N-TEFs and Termination Factors

The evidence gathered so far points to the existence of two factors (DSIF and NELF) that play a negative role in elongation control. DSIF, the DRBIF, was discovered as a protein required for DRB sensitive transcription *in vitro*. It is comprised of two subunits with strong sequence similarity to the yeast proteins SPT4 and SPT5. DSIF has almost no function on RNA polymerase II alone. The NELF also has no effect on RNA polymerase II alone, but the combination of NELF and DSIF is able to slow the elongation rate and cause the polymerase to reside longer at pause sites. In this way, NELF and DSIF act in exactly the opposite way as TFIIF, which reduces pause time. NELF is comprised of four subunits, including a

putative RNA-binding protein called RD. A defined elongation control system comprised of isolated RNA polymerase II elongation complexes (NELF, DSIF, and P-TEFb) displays many of the properties of elongation control seen when transcription is carried out in nuclear extract. DSIF is involved in controlling transcription from the HSP70 locus in *Drosophila*, but the generality of its function has not been demonstrated. Although removal of DSIF in *Caenorhabditis elegans* using RNA interference techniques relieved most of the requirements of P-TEFb at a heat shock locus, similar results were not found in expression of several other genes. These facts along with the fact that homologs of NELF subunits have not been identified in many organisms suggest that other negative factors, which operate at a wide variety of genes, remain to be discovered.

So far only one transcription termination factor, TTF2 (formerly and uninformatively called factor 2), has been identified. It is a member of the SWI2/SNF2 family of proteins that are generally involved in disrupting protein nucleic acid interactions. It binds to both single- and double-strand (ds) DNA and has a strong dsDNA-dependent adenosine triphosphatase activity that is required for termination activity. The termination activity of TTF2 is inhibited by TFIIF, and while the mechanism of this inhibition is unknown, it provides a possible method to control the activity of TTF2 during transcription. The role of TTF2 seems to operate outside of the elongation control process in that its negative effect is not directly reversed by P-TEFb. However, a role for the factor in abortive elongation (premature termination) or in normal termination associated with mature 3' end formation has not been ruled out.

P-TEFb

Subunits and Activity

P-TEFb is a cyclin-dependent kinase comprised of Cdk9 and one of several cyclin subunits. Three genes in humans – T1, T2, and K – encode cyclin subunits that can associate with and activate Cdk9. Two cyclin T genes have been found in *C. elegans*, but only one in *Drosophila*. Supporting a broad role in gene expression, knockout of Cdk9 in *C. elegans* causes loss of gene expression and death at the same early stage as knockout of one of the subunits of RNA polymerase II. The same lethal phenotype is observed when both cyclin Ts are knocked out. The carboxyl terminal domain (CTD) of the large subunit of RNA polymerase II is phosphorylated by P-TEFb, and this domain has been shown to be required for the function of P-TEFb in the transition into productive elongation. The CTD is comprised of multiple repeats of the heptapeptide YSPTSPS, and P-TEFb predominately phosphorylates the second serine. P-TEFb had been shown to have broad substrate specificity *in vitro* and will phosphorylate the large subunit of both DSIF and TFIIF. Although the CTD is phosphorylated by P-TEFb and is required for the function of P-TEFb in transcription, there may be other important functional phosphorylation targets. The effect of DRB on transcription is explained by the fact that it inhibits the kinase activity of P-TEFb. All P-TEFb inhibitors reduce the production of mRNAs *in vivo* and are lethal at high concentrations. Flavopiridol, the most potent P-TEFb inhibitor found, is in clinical trials against cancer.

Recruitment of P-TEFb

P-TEFb interacts in a functional way with a number of important transcription factors. Cyclin T1 interacts with the HIV transactivator Tat, and P-TEFb containing cyclin T1 can be recruited to elongation complexes on the HIV template through the interaction of Tat and cyclin T1 with a stem-loop structure that forms in the nascent HIV transcript called 'transactivation response (TAR)'. This recruitment of P-TEFb activates transcription of the HIV genome. A cellular factor, CIITA, which is involved in activating transcription of the major histocompatibility complex (MHC) class II genes, has also been shown to function through interaction with P-TEFb. That interaction is blocked in the presence of Tat suggesting that the binding site is similar for the viral and cellular protein and providing a mechanism for inhibition of MHC class II gene expression during HIV infection. Other transcription factors and enhancer binding proteins – including nuclear factor kappa B, c-myc, major Cdk9-interacting elongation factor, and the androgen receptor – have been demonstrated to various degrees to have functional interactions with P-TEFb.

Recruitment of P-TEFb through artificial targeting to promoter DNA or to nascent RNA has revealed several important aspects of its function. First, when a protein is expressed in which Cdk9 or cyclin T1 or T2 is tethered to the DNA-binding domain of a yeast protein (Gal4) and the binding sites for that protein are placed close to a human promoter, transcription is stimulated from that promoter. Besides the core promoter elements, SP1 binding sites are also required to see an effect of the recruited P-TEFb in this system. This is consistent with the model for P-TEFb function because P-TEFb only affects polymerases that have already initiated, and SP1 increases the initiation efficiency. Expression of Cdk9, cyclin T1, T2, or K tethered to an RNA-binding domain (e.g., HIV rev) activates transcription of a gene containing the appropriate RNA element. In both tethering experiments, the untethered component of P-TEFb is also recruited due to the inherent interaction between P-TEFb subunits. Cyclin K does not work when tethered to DNA elements because it lacks a CTD present on both cyclin T1 and T2 that plays another role by interacting with the CTD of the large subunit of RNA polymerase II.

Control of P-TEFb by 7SK

Very recently, the laboratories of Olivier Bensaude and Qiang Zhou simultaneously discovered the small cellular RNA, 7SK, in association with a fraction of P-TEFb in cells. The large complex containing P-TEFb and 7SK was relatively inactive compared to the free P-TEFb (**Figure 2**). Addition of small amounts of compounds that inhibit elongation by RNA polymerase II causes a reduction in the large form of P-TEFb *in vivo* and a concomitant increase in the small active form. Evidently, a cell will try to compensate for reduced production of mRNAs by activating more P-TEFb. The change in large to small form of P-TEFb can occur within minutes and is reversible. The significance of the large form of P-TEFb was demonstrated using a mouse model of cardiac hypertrophy. Heart cells that were stimulated to grow in size had an increase in P-TEFb activity, but no increase in the amount of P-TEFb. It was shown that all hypertrophic signals tested caused a release of 7SK. The signal transduction pathways involved in hypertrophy of cardiac cells at least partially routes through the elongation control pathway.

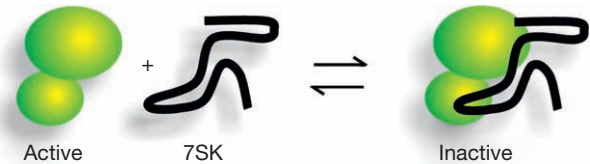


Figure 2 Control of P-TEFb by 7SK. The amount of P-TEFb activity in the cell can be quickly adjusted by the reversible association of 7SK RNA. The large form with 7SK bound is inactive.

Integration of Elongation Control and Gene Expression

Elongation control may play a very significant role in controlling eukaryotic gene expression. Most RNA polymerase II molecules found transcribing human genes require the prior function of P-TEFb. Recruitment of P-TEFb by transcription factors allows specific genes to be targeted, and both the targeted and the untargeted functions of P-TEFb may allow global regulation of mRNA production. Regulating the conversion of the large to small P-TEFb form by signal transduction pathways is another global means to regulate mRNA levels. Overall, the combined action of negative and then positive factors results in a kinetic delay of the progression of polymerases down the gene. This may provide a window of opportunity for RNA-processing machinery to function or associate with the elongating polymerase, ensuring a truly productive elongation event resulting in a complete, functional mRNA. The human capping enzyme is functionally coupled to transcription and guanylates the nascent transcript before P-TEFb acts. However, some components of the polyadenylation and splicing machinery have been shown to associate with the phosphorylated CTD. Perhaps the processing machinery is functionally coupled to transcription through interactions provided by elongation control and scans the nascent transcript for processing sites as it is being made. Many mechanistic details of how RNA polymerase II elongation control is accomplished and the ramifications of the process on subsequent events remain to be determined.

See also: **Molecular Biology:** RNA Polymerase II and Its General Transcription Factors; **Signaling:** Tachykinin/Substance P/Neurokinin-1 Receptors.

Further Reading

- Marshall NF and Price DH (1992) Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Molecular and Cellular Biology* 12: 2078–2090.
- Nguyen VT, Kiss T, Michels AA, and Bensaude O (2001) 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* 414: 322–325.
- Price DH (2000) P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Molecular and Cellular Biology* 20: 2629–2634.
- Sano M, Abdellatif M, Oh H, et al. (2002) Activation and function of cyclin T-Cdk9 (positive transcription elongation factor-b) in cardiac muscle-cell hypertrophy. *Nature Medicine* 8: 1310–1317.
- Shim EY, Walker AK, Shi Y, and Blackwell TK (2002) CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the *C. elegans* embryo. *Genes and Development* 16: 2135–2146.
- Zhu Y, Pe'ery T, Peng J, et al. (1997) Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation *in vitro*. *Genes and Development* 11: 2622–2632.

RNA Polymerase Reaction in Bacteria

S-H Jun, B A Warner, and K S Murakami, The Pennsylvania State University, University Park, PA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Abortive initiation The process during transcription initiation in which RNA polymerase makes and releases short RNA transcripts. As the RNA transcripts clash with the $\sigma 3.2$ -loop near the active site, the short RNA transcripts are released, and the process begins again.

Closed complex For transcription of genes, RNA polymerase is recruited to the promoter DNA to form the initial RNA polymerase and DNA complex. In this form, DNA is entirely double stranded.

Holoenzyme The complete multisubunit RNA polymerase including catalytic core enzyme and promoter DNA recognition σ factor.

Open complex The transcription-ready state complex, which is generated by a significant conformational

change of the closed complex which includes melting of promoter DNA.

Promoter A DNA sequence that initially binds to RNA polymerase and is the site at which the closed complex is formed and subsequently isomerizes to the open complex for transcript initiation.

Secondary channel The channel in RNA polymerase that is the path of substrate NTPs or backtracked RNA.

Trigger loop The loop that recognizes the correct NTP substrate at the active site of RNA polymerase.

It changes its conformation from a disordered loop to two α helices when the correct NTP enters the active site and triggers phosphodiester bond formation.

Transcription: Basic Mechanism

For transcribing RNA based on the DNA sequence, RNAP unwinds a small region of double-stranded DNA (~ 10 nucleotide residues) to the single-stranded form and synthesizes RNA as a complementary sequence of the DNA template (Figures 1 and 2(b)). The unwound region of DNA is called the transcription bubble that contains a DNA–RNA hybrid of ~ 8 base pairs. For synthesis of RNA, nucleotide substrate and catalytically essential divalent cations in addition to the single-stranded template DNA must be accommodated at the RNAP active site. One of four ribonucleotide triphosphates – ATP, GTP, CTP, and UTP – forms a Watson–Crick base pair with a DNA template base and its α -phosphate group is attached by a 3'-hydroxyl of the growing end of the RNA. As a result, a linear RNA is built in the 5' to 3' direction.

For the nucleotidyl transfer reaction (transcription), a two metal-ion catalytic mechanism has been proposed (Figure 2(b)), as the enzyme possesses two divalent catalytic and nucleotide-binding metal cations (Mg^{2+} or Mn^{2+}) chelated by three Asp residues in the absolutely conserved NAEFDGD motif found in all cellular RNAPs. The catalytic metal (Mg^{2+} I in Figure 2(b)) is a Lewis acid, coordinating the RNA 3'-OH lowering its pK_a and facilitating the formation of attacking oxyanion. The nucleotide-binding metal (Mg^{2+} II in Figure 2(b)) is coordinated by the triphosphate of the incoming nucleotide and stabilizes a pentacovalent phosphate intermediate during the nucleotidyl transfer reaction. Both metal ions are proposed to have octahedral coordination at physiological Mg^{2+} concentrations. In bacterial RNAP, the catalytic metal has a higher affinity to the active site and is coordinated at the active site even in the absence of DNA template or substrates. By contrast, the nucleotide-binding metal has a lower affinity and is only present with a nucleotide substrate at physiological Mg^{2+} concentration.

The nucleotide addition is accomplished through the formation of a phosphodiester bond and release of pyrophosphate (PPi) as a by-product. The free energy for this reaction is rather small ($\Delta G = -3.5 \text{ kcal mol}^{-1}$), but additional free energy is provided *in vivo* by rapidly hydrolyzing PPi into two phosphate groups by enzyme pyrophosphatase that makes the entire reaction highly energetically favorable ($\Delta G = -7 \text{ kcal mol}^{-1}$) under physiological condition, which is practically irreversible.

All cellular organisms including bacteria use multisubunit DNA-dependent RNAP for transcribing most of the RNAs in cells. In bacteria, well-characterized RNAPs are from the Gram-negative bacterium *Escherichia coli* and the Gram-positive bacterium *Bacillus subtilis*. For structural studies, RNAPs isolated from the thermophilic bacteria, *Thermus aquaticus* and *Thermoplasma thermophilus*, have been used for X-ray crystallography, but *Thermus* RNAPs are distantly related to *E. coli* and *B. subtilis* RNAPs and also to RNAPs from pathogenic bacteria species.

Transcription Cycle: Initiation

The catalytically competent bacterial core RNAP (subunit composition $\alpha_2\beta\beta'\omega$, with a molecular mass of $\sim 400 \text{ kDa}$) is evolutionarily conserved in sequence, structure, and function from bacteria to humans. The core RNAP subunits are encoded by *rpoA* (for α subunit, MW: 36 kDa), *rpoB* (for β subunit, MW: 150 kDa), *rpoC* (for β' subunit, MW: 155 kDa), and *rpoZ* (for ω subunit, MW: 10 kDa) genes. The overall shape of core RNAP resembles a crab claw. The β and β' subunits form each claw arm, generating a cleft for double-stranded DNA binding. The enzyme active site is located at the bottom of the cleft that coordinates catalytic Mg^{2+} . The architecture around the cleft including the active site is highly conserved among all cellular RNAPs, which indicates that the catalytic mechanism

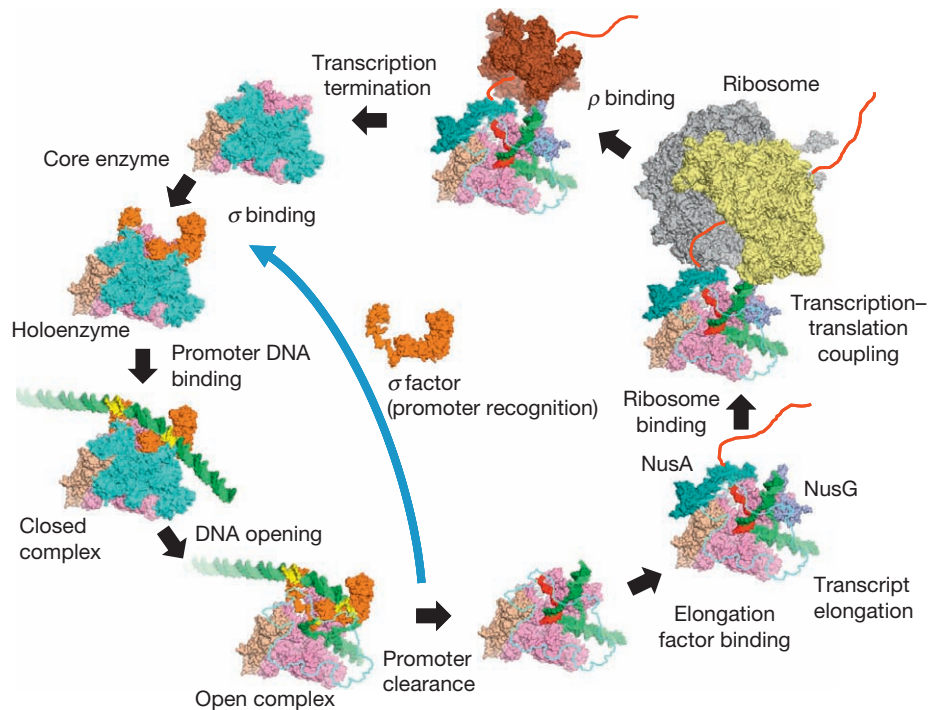


Figure 1 The transcription cycle. RNAP core enzyme (α , light brown; β , cyan; β' , pink; ω , white), σ factor (orange), DNA (template strand, dark green; nontemplate strand, light green; -10 and -35 elements, yellow), and RNA transcript (red) at different stages of the transcription cycle. Transcript elongation factors NusA and NusG, ribosome (30 S, yellow; 50 S, gray) and transcription termination factor ρ are also shown and labeled. From the open complex to transcription termination, β is removed and the outline of β is shown as a cyan line to reveal the inside of RNAP. All structures in this figure are X-ray/NMR structures of *E. coli* proteins or homology models of *E. coli* proteins based on structures from *T. aquaticus*.

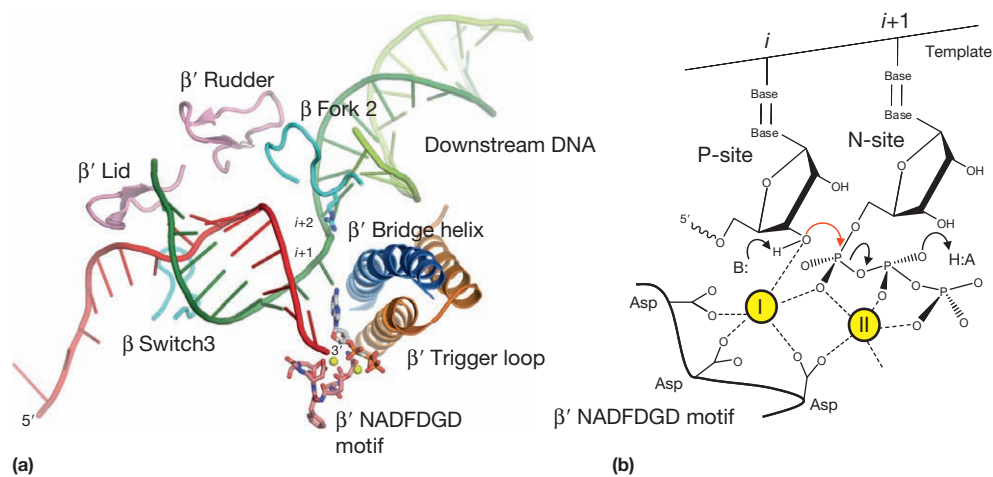


Figure 2 Bacterial RNA polymerase active site and mechanism of catalysis. (a) Active site structure of the bacterial RNA polymerase (PDB: 205J). Template DNA (green), nontemplate DNA (light green), RNA (red), bridge helix (blue), trigger loop (orange), Mg^{2+} ions (yellow), and incoming NTP at N-site (stick model). (b) Mechanism for two- Mg^{2+} -mediated catalysis of nucleotidyl transfer reaction. Two substrates of the active site, P- and N-sites, are shown. NTP is shown in green. The red arrows illustrate the nucleophilic attack of the RNA 3' oxyanion on the NTP α -phosphate and the reaction involves stabilization of a trigonal-bipyramidal transition state by two Mg^{2+} ions.

of RNA synthesis from bacteria to humans is conserved. Although the two largest subunits, β and β' , provide most of the catalytic functions, their assembly is dependent on the presence of two α subunits, which form a platform for assembly.

The core enzyme carries out highly efficient RNA transcription using DNA as a template; however, transcript initiation from specific DNA sequences, called 'promoters', requires an additional DNA sequence recognition protein, σ factor, which binds the core RNAP to form the holoenzyme (Figure 1).

The σ factor transiently associates with the core enzyme for promoter recognition and it dissociates from the core enzyme once the RNAP starts processive transcript elongation. Bacteria encode multiple σ factors (seven in *E. coli*) and the most common *E. coli* σ factor is σ^{70} (encoded by *rpoD* gene) that carries out transcription of housekeeping genes. Other σ factors can substitute housekeeping σ under certain growth and stress conditions and direct RNAP to alternative sets of promoters for optimizing gene expression for adaptation and for cell differentiation such as sporulation in *B. subtilis*.

Most bacterial σ factors belong to a homologous family closely related to *E. coli* σ^{70} , with distinct regions of highly conserved sequence. Group 1 (or primary) σ s, SigA, direct the majority of transcription during log-phase growth. Structural analysis reveals that group 1 σ s are comprised of four flexibly linked domains, $\sigma 1.1$, $\sigma 2$, $\sigma 3$, and $\sigma 4$, containing conserved regions 1.1, 1.2–2.4, 3.0–3.1, and 4.1–4.2, respectively.

In the holoenzyme, the globular domains of σ lay spread out across the upstream face of the RNAP crab claw. The promoter-binding determinants of σ factor, $\sigma 2$ (–10 element) and $\sigma 4$ (–35 element), are solvent exposed, spaced according to their cognate DNA elements. The $\sigma 3$ and $\sigma 4$ domains are separated by 45 Å in the holoenzyme. This distance is spanned by an extended 33-residue linker, comprised primarily of σ region 3.2 (the $\sigma 3.2$ -loop), which loops into the RNAP active site channel, then winds its way out through the RNA exit channel.

The σ and core RNAP interaction is very stable, with a dissociation constant estimated to be $\sim 10^{-9}$ M. Each of the σ domains, as well as the linkers connecting them, interacts with the core RNAP. Nevertheless, when RNAP enters the elongation phase of transcription, σ is generally released. These seemingly contradictory properties are explained by the σ architecture. The simultaneous, independent binding of discrete structural elements of σ ($\sigma 1.1$, $\sigma 2$, $\sigma 3$, $\sigma 3.2$ -loop, and $\sigma 4$) to different parts of the core provide for high-affinity binding, without any one σ -domain/core interaction being particularly stable. Stepwise structural transitions during initiation could induce the dissociation of individual σ -domains, one by one, effecting eventual σ release.

The search for a promoter in the genome occurs through occasional collision and/or sliding of RNAP along DNA. Sliding greatly accelerates the search, which is concluded by establishment of specific contacts between RNAP and a promoter in a relatively unstable closed complex (Figure 1). The RNAP holoenzyme binds to the promoter DNA region from –60 to +20 nucleotides relative to the transcription start site and the σ factor recognizes two 6 bases DNA sequences centered on the positions around –35 and –10 (called the –35 and –10 elements). The consensus sequences of the –35 and –10 elements are conserved in bacteria and their sequences are TTGACA and TATAAT, respectively.

RNAP unwinds DNA from the –10 element downstream past the transcription start site (+1) to form the transcription bubble. Only the single-stranded template strand is directed to the active site and the double-stranded downstream DNA from +5 to roughly +12 is clamped by the RNAP DNA-binding cleft to form the open complex (Figure 1).

Upon formation of the open complex, the RNAP active site takes NTP substrates through the secondary channel and

starts synthesizing RNA. While DNA polymerase requires a DNA or RNA primer for starting DNA synthesis, RNAP is able to initiate RNA synthesis from NTPs in a process called primer-independent or *de novo* transcript initiation. Transcript initiation is the only step in the entire transcription process where two nucleotide substrates are loaded at the active site followed by a nucleotidyl transfer reaction. A short transcript only a few nucleotides in length encounters a $\sigma 3.2$ -loop in its exit path that dissociates RNAs from the active site; this cycle continues several times (abortive initiation). At each cycle, the elongating RNA displaces the $\sigma 3.2$ -loop out of its path. Eventually, the RNA chain that elongates to a length of around 12 nucleotides fills completely the DNA/RNA hybrid and the upstream RNA exit channel under the β subunit flap and thereby displaces σ from its path and ends the abortive initiation.

The displacement of the $\sigma 3.2$ -loop couples the presence of an ~ 12 -nt RNA chain to the initial stages of promoter escape by destabilizing $\sigma 4/\beta$ flap interactions. Release of $\sigma 4$ from the β flap would destabilize the $\sigma 4$ –35 element interactions that allow the RNAP to escape from the promoter and translocate downstream on the DNA for RNA elongation. This transition into RNA elongation does not require the complete release of σ , since other domains of σ bound to RNAP are not incompatible with the paths of the DNA and RNA in the elongation complex (EC). Eventually, complete release of σ results in the EC with core RNAP (Figure 1). In some cases, however, σ remains associated with RNAP in the early, middle, or late stages of transcript elongation, and the timing of σ release is a stochastic process.

Transcription Cycle: Elongation

During transcript elongation, the elongation factors NusA and NusG associate with the EC (Figure 1) and regulate the velocity of RNA synthesis, transcriptional pausing, and termination efficiency. NusA is a multidomain protein with a rod-shaped structure. The NusA N-terminal domain associates with the β flap of RNAP while the C-terminal region contains several RNA-binding motifs and also binds with the α subunit C-terminal domain of RNAP. NusA enhances intrinsic termination as well as transcriptional pausing by interacting with nascent RNA hairpins.

NusG consists of the NusG amino-terminal (NGN) domain and Kyprides–Onzonis–Woese (KOW) motif at the C-terminal domain (hence also called KOW domain), and these two domains fold independently and are connected by a flexible linker. The NGN domain associates with the β' subunit coiled-coil motif of RNAP and regulates the velocity and processivity of the transcript elongation. The NusG–KOW domain plays important roles in interacting with other proteins, for example, it contacts the elongation factor NusE, which is also the ribosomal S10 subunit. This interaction may play an important role in the coupling of transcription and translation. During the transcription of protein-coding genes, the EC is followed by a ribosome (Figure 1) and the overall transcript elongation rate is tightly controlled by the rate of protein translation. This transcription and translation coupling prevents RNAP backtracking and facilitates readthrough of roadblocks on the

DNA template. Both NusA and NusG are essential components of anti-termination complexes for the ribosomal RNA (*rnm*) and phage λ operons, and participate in ρ -dependent transcription termination.

Transcription Cycle: Termination

Transcription termination is the final step of gene expression, which plays an important role in making an end of RNA without affecting unnecessary gene expression from downstream genes. This is particularly important for the 'gene rich' genome of bacteria. There are two types of termination: ρ -dependent termination and ρ -independent intrinsic termination. The ρ -dependent termination requires a hexameric helicase protein ρ , which loads onto nascent RNA strands at *rut* (rho utilization) sites (Figure 1) followed by stimulating an RNA-dependent adenosine triphosphatase (ATPase) activity. Coupling of ATP hydrolysis translocates ρ in the 5' $\rightarrow$ 3' direction along the RNA toward the EC. Eventually, ρ bumps into the RNAP and pulls out RNA from the EC, causing the release of the transcript and dissociation of RNAP from DNA.

The intrinsic terminators contain short self-complementary GC-rich sequences followed by an oligo-A tract on the template DNA sequence. Transcription of such sequences is accompanied by pausing of RNAP due to backtracking, which ensures the formation of a hairpin-like RNA structure. The RNA hairpin destroys the nucleic acid contacts with RNAP, which leads to collapse of the transcription bubble and disintegration of the EC.

Nucleotide Additional Cycle and Proofreading

Bacterial RNAP forms a stable EC with DNA and RNA for processive RNA extension of up to $\sim 10^4$ bases with a rate of 50–100 nucleotides s^{-1} . During transcript elongation, the EC constantly unwinds the double-stranded downstream DNA before the template DNA strand enters the active site. The single-stranded template DNA within the transcription bubble (12–14 bases in length) is positioned at the active site and forms an ~ 8 -nucleotide hybrid with nascent RNA. The template DNA is rewound with the nontemplate DNA strand at the end of the transcription bubble to form the double-stranded upstream DNA.

The RNAP active site, which is buried within a cleft between the β and β' subunits, is connected with the outside through three channels (main, secondary, and RNA exit) to synthesize RNA efficiently. To extend RNA, RNAP moves the 3'-end of RNA to the P-site (primer or product site, which is also referred to as the i site, Figure 2(a)) to make a space for the incoming nucleotide, which is accommodated at the insertion site (N-site, also referred to as the $i + 1$ site) for the nucleotidyl transfer reaction. Downstream DNA is accommodated at the cleft of the main channel and maintains its duplex DNA structure until the $i + 2$ position. The β subunit fork loop 2 plays a role in DNA strand separation as Arg side chain stacks on the $i + 2$ base pair and blocks the passage of the duplex DNA towards the $i + 1$ site (Figure 2(a)). The template DNA strand

forms a sharp kink between the $i + 1$ and $i + 2$ sites to present only one single-stranded DNA base at the $i + 1$ site for each round of the nucleotide addition cycle. The 3'-end of the RNA positions near the catalytic metal and also above a funnel-like secondary channel, which is an entry pore of nucleotide substrates to the active site and 3'-end RNA exit site when RNAP backtracks for transcriptional proofreading. The 5'-end of RNA positions at the RNA exit channel, which is below the β flap domain.

The three protein loops, the lid, the rudder, and the β switch 3, form an interaction network with the upstream edge of the DNA/RNA hybrid (Figure 2(a)). The lid stacks on the upstream base pair and sterically blocks growth of the hybrid in order for RNA to separate from the template DNA. The rudder interacts with DNA separated from the hybrid preventing re-association with the RNA. The β switch 3 interacts with the first displaced RNA base presumably to help the separation of RNA from the hybrid and the re-annealing of the template and nontemplate DNA strands.

RNAP has to operate high-fidelity transcription since transcription errors could produce nonfunctional ribosomes, transfer RNAs (tRNAs), or messenger RNAs (mRNAs) that produce mutant proteins without proper functions. Transcriptional accuracy of bacterial RNAP is relatively high with an estimated error rate of less than 10^{-5} . RNAP secures transcriptional fidelity by using two strategies, substrate selection and proofreading, which remove a mismatched nucleotide from nascent RNA by RNA cleavage activity. RNAP carries out this RNA cleavage reaction by using the same tunable active site that also carries out RNA extension.

To select the correct NTP substrate, RNAP discriminates (1) NTPs from deoxyNTPs (dNTPs) and (2) correct NTPs from incorrect NTPs. To discriminate NTPs from dNTPs, RNAP recognizes the 2'-OH group of the ribose by hydrogen bonding with a side chain of an Asn residue found in the NADFDGD motif (Figure 2(a)). Mutation of this Asn residue increases dNTP incorporation. To discriminate correct NTPs from incorrect NTPs, RNAP carries out two-step reactions that involve an isomerization of the RNAP active site. In the first step, NTP binds to a pre-insertion site for testing Watson–Crick base pairing with the DNA template. Once the Watson–Crick base pairing is established, the NTP is delivered to the insertion site (N-site) by conformational changes of the trigger loop and bridge helix that lead to alignment of the reactive groups including the 3'-OH group of RNA and the triphosphate group of the NTP for the nucleotidyl transfer reaction (Figure 2(a)).

The trigger loop changes its conformation depending on the presence or absence of NTP at the active site, and also correct and incorrect nucleotide at the active site. Without NTP or with an incorrect NTP at the active site, the trigger loop is located farther than 30 Å away from the insertion site. By contrast, with a correct NTP at the insertion site, the trigger loop refolds its structure to form the two α -helices with extensive interactions with the nearby bridge helix and positions in proximity to the insertion site (Figure 2(a)). In this trigger loop conformation, it interacts with all parts of the nucleotide for the alignment of reactive groups for catalysis.

Bacterial RNAP is one of the strongest motor proteins, which converts energy associated with the nucleotidyl transfer reaction into mechanical motion along the DNA template.

Each nucleotide addition cycle couples with the translocation of the DNA/RNA hybrid at the active site that moves a freshly made 3'-end of RNA from the N- to the P-sites for making an empty N-site and brings an unpaired DNA base to the $i + 1$ position for the next round of the nucleotide addition. In the case of cellular RNAPs including bacterial RNAP, the Brownian Ratchet model was proposed for a mechanism of translocation. In this model, the 3'-end of RNA changes its position between pre- and post-translocated states at a speed more rapid than that of nucleotide loading, and the loading of a nucleotide at the N-site drives the RNAP movement for the forward translocation.

When RNAP incorporates a mismatched base in RNA, it slows down and pauses transcript elongation and eventually backtracks RNA to form a frayed RNA 3'-terminal nucleotide. RNAP uses the RNA hydrolysis activity either by an intrinsic or by a Gre factor stimulated RNA cleavage activities for removing one or a few nucleotides of RNA containing a misincorporated nucleotide. This RNA hydrolysis also makes a fresh 3'-OH group at the end of RNA to restart transcript elongation. Gre stimulates RNA hydrolysis activity of RNAP by more than 3000-fold and accesses the RNAP active site through the secondary channel, placing Asp/Glu residues at the RNAP active site for chelating Mg^{2+} for the RNA hydrolytic reaction. Bacterial RNAP has another RNA cleavage activity, which is just reversing the RNA synthesis reaction in the presence of PPi to release NTP from the 3'-end of RNA. This reverse reaction is known as 'pyrophosphorolysis'. For this reaction, the RNA 3'-end has to be positioned at the N-site (pretranslocated register) and RNAP is able to progressively shorten the RNA. The RNAP pyrophosphorolysis reaction requires a higher concentration of PPi that cannot be achieved in physiological conditions and

also nucleotide concentrations in cells are preferential for the RNA synthesis reaction; thus, RNAP cannot use the pyrophospholysis reaction for proofreading *in vivo*.

Transcription Inhibitions by Antibiotics

Transcription is an essential process in bacteria and, as such, is a target for antimicrobial drugs (antibiotics). Bacterial RNAP is an excellent target for antibiotics because (1) it is an essential enzyme for transcription, (2) the bacterial RNAP sequences are highly conserved, allowing for antibiotics to be effective against a wide range of species, and (3) the bacterial RNAP sequences are less conserved with the eukaryotic RNAP sequences, allowing for therapeutic selectivity. In addition, bacterial RNAP is a large enzyme composed of many subunits that increases the likelihood of finding new effective antibiotics.

There are four well-characterized antibiotics, Rifampicin (Rif), Sorangicin (Sor), Streptolydigin (Stl), and Myxopyronin (Myx), which work against bacterial RNAP. Biochemical, genetic, and crystallographic studies of RNAP-antibiotic complexes identified three separate target sites (Rif/Sor, Stl, and Myx sites, Figure 3) and three different mechanisms for inhibiting bacterial RNAP.

One of the best-characterized antibiotics against RNAP is Rifampicin (Rif), which is a member of the Rifamycin family of antibiotics, active against a broad range of Gram-positive and Gram-negative bacterial infections, and in clinical use in the treatment of tuberculosis. The crystal structure of the *T. aquaticus* core RNAP-Rif complex identified the Rif-binding site on the β subunit adjacent to the RNAP active site (Figure 3 (a)), which is not overlapping with the DNA binding or catalytic center

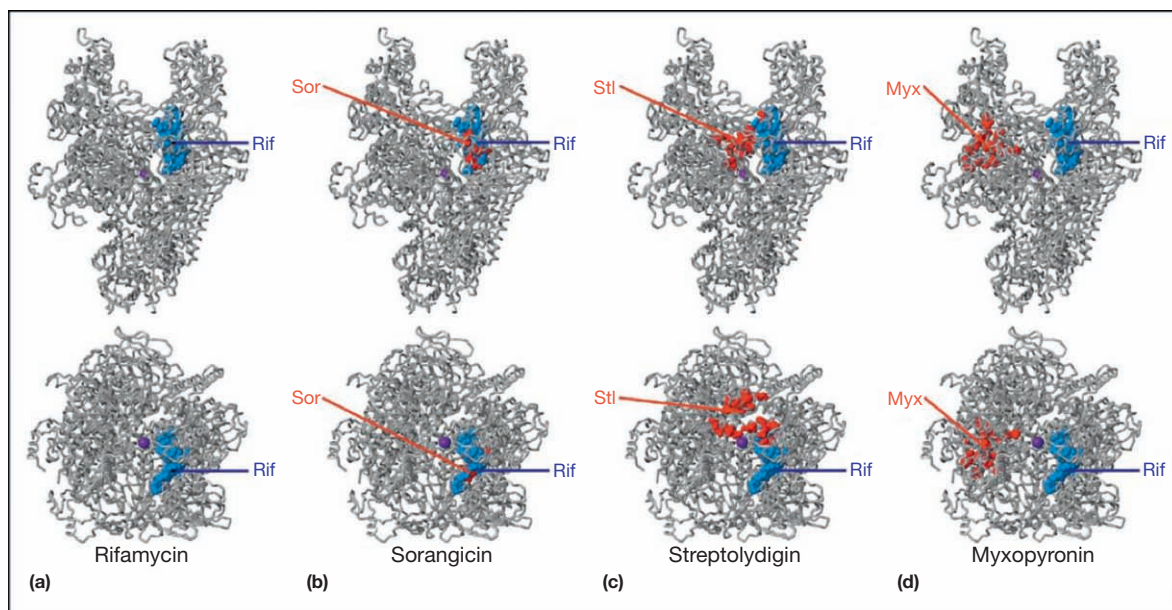


Figure 3 Targets of small-molecule inhibitors of RNAP. (a) Target of rifamycins (Rif). (b) Target of sorangicin (Sor). (c) Target of streptolydigin (Stl). (d) Target of myxopyronin (Myx). Each panel shows two orthogonal views of RNAP. Red, sites of substitutions conferring resistance to the specified inhibitor; blue, sites of substitutions conferring resistance to rifamycins; violet sphere, active-center Mg^{2+} . Adapted from figure 1 of Ho MX, Hudson BP, Das K, Arnold E, and Ebright RH (2009) Structures of RNA polymerase-antibiotic complexes. *Current Opinion in Structural Biology* 19: 715–723, with permission from Elsevier.

but overlaps with the path of nascent RNA. The structure proposed the steric clash model, which is consistent with biochemical observations that Rif inhibits transcript initiation and extension, but does not inhibit elongating RNAP that already has a longer RNA. The Rif-binding site of RNAP is also a binding site for another RNAP inhibitor Sorangicin (Sor) (**Figure 3(b)**), which is a polyketide-derived macrocyclic-polyether product from *Myxobacterium Sorangium cellulosum*. As found for Rif, Sor inhibits transcript initiation by preventing the extension of short RNA (2 ~ 4 nt length). Because of the proximity of their binding sites on RNAP, Rif and Sor show partial cross-resistance; some Rif-resistant mutants are resistant to Sor, and all Sor-resistant mutants are resistant to Rif.

Streptolydigin (Stl) is a polyketide-derived tetramic-acid antibiotic produced by the Actinomycete *Streptomyces griseoflavus*. Stl binds to key elements of the RNAP active site for transcription – the bridge helix and the trigger loop (**Figure 3(c)**) – and inhibits all reactions of RNAP transcription by blocking conformational changes of these elements that occur during transcription. Crystal structures of the RNAP–Stl complexes revealed that the Stl traps the bridge helix and trigger loop in straight and unfolded conformational states, respectively, by direct interactions with these elements. Because the Stl target site is not overlapping with the Rif/Sor target site (**Figures 3(a)** and **3(b)**), Rif and Stl exhibit only minimal cross-resistance.

Myxopyronin (Myx) is a polyketide-derived a-pyrone antibiotic produced by the *Myxobacterium Myxococcus fulvus* Mxf50. Myx binds the RNAP switch region (**Figure 3(d)**), which is a hinge of the RNAP clamp for opening and closing of the DNA-binding channel. Upon binding, Myx inhibits transcript initiation by preventing formation of the promoter open complex. Crystal structures of the RNAP–Myx complexes with biochemical studies proposed two models of inhibition that include the following: Myx inhibits transcription (1) by preventing opening of the RNAP clamp to permit entry and unwinding of DNA (hinge jamming mechanism) or (2) by interfering with interactions between RNAP and the unwound DNA template strand. Because the Myx target site is away from the Rif/Sor target site (**Figures 3(a)** and **3(b)**), Rif and Myx exhibit absolutely no cross-resistance.

Although much work has been completed regarding antibiotics that act by inhibiting RNAP transcription, the escalating problem of antibiotic resistance necessitates further research. The increasing prevalence of bacterial strains resistant to the Rifamycins is making treatment of tuberculosis more difficult. Bacteria are developing various resistance strains by point

mutations in the genes encoding RNAP subunits. This increasing and dangerous problem calls for further research into the modification of existing RNAP-acting drugs as well as the development of novel antibiotics that utilize different methods of inhibiting RNAP.

See also: **Molecular Biology:** Gene Expression in Eukaryotes: RNA Polymerase II Structure; RNA Polymerase I and RNA Polymerase III in Eukaryotes; RNA Polymerase II and Its General Transcription Factors; RNA Polymerase II Elongation Control in Eukaryotes; RNA Polymerase Structure, Bacterial; Sigma Factors; T7 RNA Polymerase.

Further Reading

- Buc H and Strick T (2009) *RNA Polymerases as Molecular Motors*. Cambridge, UK: RSC Pub.
- Darst SA (2001) Bacterial RNA polymerase. *Current Opinion in Structural Biology* 11: 155–162.
- Haugen SP, Ross W, and Gourse RL (2008) Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nature Reviews* 6: 507–519.
- Ho MX, Hudson BP, Das K, Arnold E, and Ebright RH (2009) Structures of RNA polymerase–antibiotic complexes. *Current Opinion in Structural Biology* 19: 715–723.
- Landick R (2006) The regulatory roles and mechanism of transcriptional pausing. *Biochemical Society Transactions* 34: 1062–1066.
- Murakami KS and Darst SA (2003) Bacterial RNA polymerases: The whole story. *Current Opinion in Structural Biology* 13: 31–39.
- Murakami KS, Masuda S, Campbell EA, Muzzin O, and Darst SA (2002) Structural basis of transcription initiation: An RNA polymerase holoenzyme–DNA complex. *Science* 296: 1285–1290.
- Murakami KS, Masuda S, and Darst SA (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* 296: 1280–1284.
- Nudler E (2009) RNA polymerase active center: The molecular engine of transcription. *Annual Review of Biochemistry* 78: 335–361.
- Srivastava A, Talaue M, Liu S, et al. (2011) New target for inhibition of bacterial RNA polymerase: ‘switch region’. *Current Opinion in Microbiology* 14(5): 532–543.
- Sydow JF and Cramer P (2009) RNA polymerase fidelity and transcriptional proofreading. *Current Opinion in Structural Biology* 19: 732–739.
- Vassilyev DG, Vassilyeva MN, Perederina A, Tahirov TH, and Artsimovitch I (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* 448: 157–162.
- Vassilyev DG, Vassilyeva MN, Zhang J, et al. (2007) Structural basis for substrate loading in bacterial RNA polymerase. *Nature* 448: 163–168.
- Villain-Guillot P, Bastide L, Gualtieri M, and Leonetti JP (2007) Progress in targeting bacterial transcription. *Drug Discovery Today* 12: 200–208.
- Zhang G, Campbell EA, Minakhin L, et al. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98: 811–824.

RNA Polymerase Structure, Bacterial

S Borukhov, University of Medicine and Dentistry of New Jersey, School of Osteopathic Medicine, Stratford, NJ, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Michael Anikin, Dmitri Temiakov, and William T. McAllister, volume 3, pp. 781–784, © 2004, Elsevier Inc.

Introduction

Enzymatic Activities of RNA Polymerase

In enzymological terms, RNA polymerase (RNAP) belongs to a nucleotidyl transferase class of enzymes. During RNA polymerization (forward reaction), it catalyzes (in a DNA-template dependent manner) the transfer of ribonucleoside monophosphate moiety from ribonucleotide triphosphate (rNTP) to the 3'-terminal hydroxyl group of the growing chain of RNA product, producing pyrophosphate (diphosphate, PPi) as by-product. In a reverse reaction, pyrophosphorolysis, RNAP transfers ribonucleoside monophosphate moiety from the 3'-terminal nucleotide of RNA back to pyrophosphate resulting in degradation of RNA and formation of rNTPs. In addition, RNAP catalyzes RNase-like hydrolysis of nascent RNA acting as 3'-exoribonuclease and endoribonuclease. In the course of these reactions, RNAP produces shortened nascent RNA with the free 3'-hydroxyl group and ribonucleoside-5'-monophosphate (rNMP) or 2–19-nt-long oligoribonucleotide fragments with 5'-monophosphate, respectively. RNAP also possesses a DNA-helicase activity: it unwinds the double-stranded DNA (dsDNA), separates the DNA strands (DNA melting), and induces both positive and negative DNA supercoiling. Finally, RNAP is a DNA-translocase which acts as a powerful molecular motor. During transcription, it utilizes the energy released from rNTP binding and incorporation into the growing chain of RNA to generate substantial mechanical force and to move unidirectionally along the DNA template. All enzymatic activities of RNAP require the presence of two Mg^{2+} ions (one high-affinity and another low-affinity) at the catalytic center.

Subunit Composition

In bacteria, such as *Escherichia coli* (*Eco*), the catalytically active core enzyme consists of five subunits ($\alpha_2\beta\beta'\omega$): two α (~36 kDa), β (~150 kDa), β' (~155 kDa), and ω (~8.5 kDa), with the total molecular weight of ~380 kDa. On its own, the core enzyme is unable to recognize specific promoter DNA sequences, or melt the DNA and initiate transcription. To carry out these functions, it must bind one of several specificity factors, σ (or σ subunit, 20–70 kDa) to form RNAP holoenzyme ($\alpha_2\beta\beta'\omega\sigma$). Alternative σ factors bind competitively to the core, generating multiple forms of holoenzyme that can utilize different classes of promoters under various growth conditions and in response to stress. The number of σ factors in different bacterial species varies from one in *Mycoplasma genitalie* to seven in *Eco* to 60 in *Streptomyces coelicolor*. Besides σ , other regulatory transcription factors can bind RNAP to modulate its ability to recognize and bind promoters, modify its enzymatic activity, alter the rate of RNA synthesis and its processivity, and control the release of RNA product.

Transcription Cycle

Transcription process can be roughly divided into three major stages: initiation, elongation, and termination. At each of these steps RNAP is targeted by large number of regulatory factors. Initiation is the first and most complex step in transcription. During initiation, the σ^{70} -holoenzyme slides along the DNA scanning for the presence of two conserved hexamer elements of the promoter centered at positions –35 and –10 (with consensus sequences $^{-35}TTGACA^{-30}$ and $^{-12}TATAAT^{-7}$, respectively, relative to the transcription +1 start site) and separated by a $\sim 17 \pm 1$ -bp spacer. In addition to the conserved hexamers (that may differ from consensus), other elements immediately upstream and downstream of the –10 element ($^{-14}TG^{-13}$ extended –10', –15' regions, and 'discriminator' (DSR)) and A/T-rich regions upstream of the –35 element ('UP-element' at –45; –65), all contribute to specific recognition by the σ^{70} holoenzyme. The holoenzyme specifically binds to these elements to form a closed promoter complex (RPC). After several transitional steps the enzyme unwinds the DNA duplex around the –10 region (between nt –12 and +2) creating a 12–14-nt-long transcription bubble, and forms an open promoter complex (RPO). In the presence of rNTPs, RPO transforms into an initiating complex (RPinit) which forms the first phosphodiester bond between rNTPs positioned at +1 and +2 sites and the RNA synthesis commences. At this stage, RNAP exists in the form of a scrunched complex where the upstream DNA–RNAP contacts remain intact; the downstream DNA (from +1 to +15) is pulled into the enzyme and transcribed into the RNA, while the transcription bubble expands extruding the excess of ssDNA strands outside of the complex. The RPinit may proceed into either abortive or productive pathways. In the former, the enzyme enters the cycle of abortive initiation when it repetitively synthesizes and releases short RNAs without leaving the promoter. In the latter, the enzyme synthesizes RNA till it reaches a critical length (typically 11–15 nt), of which 8–9 nt are base-paired with the DNA template strand (DNA–RNA hybrid) and 2–6 nt are threaded through the RNA exit channel, escapes from promoter, and enters the elongation stage of transcription. This transition leads simultaneously to the loss of RNAP–promoter DNA contacts, σ dissociation, and formation of a highly stable and processive ternary elongation complex (EC).

Throughout elongation stage, the size of transcription bubble in EC remains constant at $\sim 12 \pm 1$ nt, and the size of RNA/DNA hybrid is maintained at 8–9 bp. EC can transcribe DNA over long distances (>10 000 bp) without dissociation and release of RNA product. However, the monotonous forward movement of EC can be interrupted by pauses which occur through at least two pathways. The first (type I) is induced by RNAP interaction with the nascent RNA secondary structures (such as RNA hairpins). This leads to distortion in RNA/DNA hybrid and temporary disengagement of RNA

3'-end from the active site. The second (type II) results from various roadblocks to RNAP translocation caused by DNA-binding proteins, misincorporated substrates, DNA lesions, and certain DNA sequences. During type II pausing, RNAP may slide back by 2–20 nt relative to the RNA 3' end (back-track). This results in extrusion of RNA 3'-terminus through the secondary channel of RNAP (see below) and concomitant repositioning of the catalytic center from the 3'-end to an internal site in RNA. Both types of pausing play important regulatory roles in transcription and in synchronization of transcription and translation. Under certain conditions, pausing could also lead to transcription arrest or dissociation of EC.

Transcription termination occurs when EC encounters a termination signal – a 20–35-nt-long G/C-rich RNA sequence of dyad symmetry that forms stem-loop structure immediately followed by a 7–9-nt-long stretch of Us. During termination, RNAP releases the nascent transcript and dissociates from the DNA template, after which it can rebind σ factor and start a new round of transcription. Under certain conditions, transcription termination can also be induced by termination factors ρ and Mfd.

Structure of RNAP Core Enzyme

Eco RNAP is the best characterized bacterial enzyme; however, its structure – determined by cryo-electron microscopy – has a

relatively low resolution of ~ 15 Å. The medium- and high-resolution structures of bacterial RNAP have been obtained by X-ray crystallography using thermophilic organisms *Thermus aquaticus* (*Taq*) and *Thermus thermophilus* (*Tth*). These include: the structures of *Taq* core (3.3-Å resolution); *Taq* holoenzyme (4 Å); *Tth* holoenzyme (2.6 Å); *Taq* and *Tth* RNAP complexes with several small molecule inhibitors and antibiotics (2.6–3.5 Å); the binary complex of *Taq* holoenzyme with promoter DNA fragment (6.5 Å); and the EC of *Tth* core with synthetic DNA/RNA scaffold (2.5 Å). In addition, high-resolution structures (1.8–2.9 Å) of individual domains of *Esc* RNAP subunits β' , α and σ , and their complexes with DNA and regulatory transcription factors were also reported. Together with information gained from a wide range of biochemical, biophysical, and genetic studies, these data refine our understanding of bacterial RNAP structure–function and provide a comprehensive view of transcription process and its regulation.

The structure of a typical bacterial RNAP, such as the *Taq*/*Tth* core enzyme, represents an asymmetric irregularly shaped ellipsoid with a deep cleft (internal channel) in the middle of the molecule, resembling a crab's claw. The overall dimension is about 155 Å in length (from the back to the tip of the claw), 115 Å in height (from the top to the bottom pincer), and 110 Å in width (alongside the channel). The top pincer of the claw is made up mostly of the β subunit, and the bottom pincer mostly of β' (Figure 1). The pincers are joined at the base of the claw by the N-terminal domains (NTDs) of asymmetrically

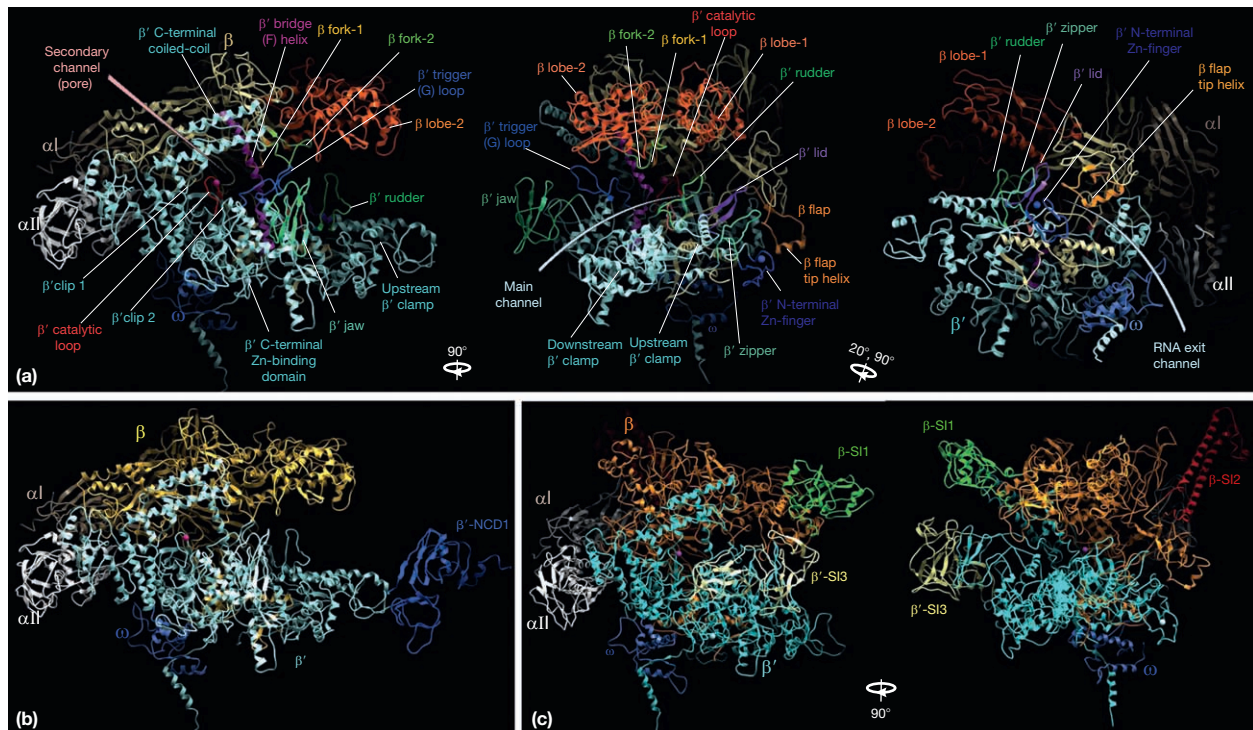


Figure 1 Structural overview of RNAP core. (a) Structure of *Taq* RNAP core shown in three rotational views, using Molsoft ICM Browser Pro program. Left panel, secondary channel view; middle panel, main channel view; right panel, RNA exit channel view. The structure is represented as lightly colored ribbons (α I, light gray; α II, olive; β , light yellow; β' , light cyan; ω , light blue) with structural elements emphasized in full color and indicated. Mg^{2+} ion is shown as small magenta sphere. The structures of β' -trigger loop, β' -rudder and β' N-terminal Zn-finger domains are modeled using the structure of *Tth* holoenzyme. The β' -NCD1 (G164-S449) and the α CTDs are not shown. (b) Secondary channel view of *Taq* RNAP core structure with β' -NCD1 colored blue. (c) Two rotational views of the modeled structure of *Eco* RNAP core (colored as above, except: β , light orange; β' , cyan; and nonconserved domains β SI1, β SI2, and β SI3 colored bright green, red and light yellow, respectively). Left panel, secondary channel view and right panel, main channel view.

placed α -dimer serving as a platform for RNAP assembly. α I-NTD binds only the β subunit, whereas α II-NTD binds mostly β' subunit. The ω subunit wraps around the β' C-terminus near the bottom pincer, serving as a β' chaperone.

RNAP Channels

The large internal space between the pincers is partitioned into three channels: the main channel (20–27 Å in diameter) and two minor channels (10–12 Å in diameter) that branch off from the primary channel to form the downstream-facing ‘secondary channel’ and the upstream-facing ‘RNA exit channel’.

The main (or primary) channel accommodates 13–15 bp of the downstream DNA duplex, 12–14 nt of the melted region of DNA (the transcription bubble), and up to 8–9 nt of DNA/RNA hybrid. In the holoenzyme, the main channel also accommodates the bulk of σ (domains σ 1– σ 3). The back wall of the primary channel – the most conserved part of the RNAP structure – contains a combination of β -sheet and α -helical elements of β and β' that comprise two juxtaposed six- β -stranded double- ψ β -barrel (DPBB) domains forming the active site cleft (see Table 1 for list of β/β' functionally relevant structural domains and elements). The DPBB domain of β' carries the catalytic loop with a triad of aspartates holding essential Mg^{2+} ions, whereas the DPBB domain of β harbors basic residues of the substrate-binding site (see below). The top and bottom walls of the channel are formed by β and β' clamps, respectively. The β clamp consists of an upstream-facing lobe-1 and a downstream-facing lobe-2 with two conserved loops, β fork-1 and β fork-2, protruding into the channel toward the active center. The downstream portion of the β' clamp is made of the C-terminal Zn-binding domain surrounded by groups of conserved α -helical elements that form the cradle for the downstream dsDNA. These include: two dsDNA-binding three-helix bundles, a long α -helical bridge (F-bridge or bridge helix) connecting the β and β' clamps near the active center and two adjacent α -helices of the flexible trigger loop (G-loop). Together with the four-strand β -sheet jaw domain, the trigger loop and bridge helix form the wall separating the primary and secondary channels (Table 1, Figures 1(a) and 1(b)). The upstream portion of the β' clamp consists of an extended N-terminal coiled-coil element (the major docking site for σ and elongation factors NusG and RfaH) with three bulging out loops: rudder, lid, and switch-2, all projecting into the main channel.

The secondary channel serves as an entry/exit pore for NTP substrates and by-products of RNA synthesis: the pyrophosphate, hydrolyzed 3'-terminal nucleotides (NMPs), and short oligoribonucleotides. It also provides an exit pathway for the backtracked RNA 3'-terminus and for abortive transcripts during initiation. The channel, which is made entirely of β' , has a funnel-like shape with wider opening on the outside surface of RNAP and a narrow opening near the catalytic center. This shape allows specialized secondary channel transcription factors to anchor at the entrance of the channel and reach through the channel to the RNAP active center to modulate its activity. The secondary channel's rim is made of the β' C-terminal coiled-coil element which serves as the major docking site for such factors as GreA, GreB, DksA, TraR, RNK, and Gfh1. The back wall of the channel consists of the extended loop adjacent

Table 1 Main structural elements of bacterial RNAP

Domains and structural elements	Residues (<i>T. thermophilus</i>)
β Subunit	
β DPBB domain	668–698; 832–879; 969–1004
β Clamp lobe-1 domain	18–142; 332–397
β Clamp lobe-2 domain	143–331
β Flap domain	701–832
β Lobe-2 pincer loop	151–182
β Lobe-1 loop	387–397
β Fork-1 helix	397–411
β Fork-2	412–441
β Fork-1	442–454
β RNA-binding helix	554–568
β Flap tip helix	768–779
β Switch-3	998–1005
β Helix-loop-helix (HLH)	1050–1080
β' Subunit	
Downstream β' clamp domain	1–131; 455–605; 1444–1468
β' NCD1 domain	132–454
β' DPBB domain	621–646; 694–769
Upstream β' clamp domain	1093–1443
β' Zipper	22–48
β' N-terminal Zn-finger domain	49–83
β' Downstream clamp N-terminal HTL loop	105–113
β' Downstream clamp helix	486–496
β' Rudder	581–603
β' Lid	525–538
β' Upstream clamp N-terminal coiled coil	540–620
β' Switch-2	606–620
β' Exit channel rim helix	664–680
β' Catalytic loop	737–744
β' Secondary channel rim C-terminal coiled-coil	958–1015
β' Clip-1	1020–1035
β' Bridge (F) helix	1066–1103
β' C-terminal Zn-binding domain	1105–1116; 1155–1189
β' Clip-2	1353–1365
β' Trigger (G) loop	1233–1257
β' Jaw	1268–1330
β' Downstream clamp C-terminal HTL loop	1434–1444

to the bridge helix (F-loop), the helix-loop-helix (HLH) element clip-1 and a bundle of short α -helices connected by unstructured loops. The side wall is formed by the HLH element clip-2, the bridge helix, the trigger loop, and the jaw (Table 1, Figure 1).

The RNA exit channel has a curved shape which helps to accommodate the C-terminal domains of σ (σ 3–4 linker and σ 4) in the holoenzyme, and serves as an outlet for RNA product in ECs. The channel is also involved in RNA/DNA hybrid strand separation, and interactions with RNA hairpins in paused ECs and in transcription termination. The exit channel walls are mostly made of the upstream portions of β and β' pincers. The double-twisted loop switch-3 and the C-terminal HLH elements of β , together with the four-helix bundle just upstream of the main channel β -barrel of β' , comprise the back wall. The top wall consists of a flexible β flap domain,

whereas the bottom wall is made of the β' rudder, the lid, and the zipper elements (Table 1, Figure 1(c)). The outside rim of the channel is formed by the β flap tip helix, the β' N-terminal Zn-finger domain, and the β' exit channel rim helix. The C-terminal α -helical domain of α I, α I-CTD (connected to α I NTDs by flexible linker), may also contribute in forming the channel's rim.

Nonconserved Domains

Despite their overall similarity, bacterial RNAPs are structurally distinct, in part due to the presence of different large non-conserved regions residing in their β and β' subunits. These domains are dispensable for RNAP basic functions; however, they play important regulatory roles in transcription. For instance, the *Taq/Tth* RNAP contains a 283-residue domain in the nonconserved portion of β' present in the upstream β' clamp (*Taq/Tth* β' -NCD1). The crystal structure of this domain comprises five repeats of sandwich-barrel hybrid motifs (SBHMs) visible in *Tth* RNAP structure as an extended part of the lower β' pincer (Table 1, Figure 1(d)). The main function of β' -NCD1 is to stabilize core- σ 70 interactions (see below). Unlike *Taq/Tth*, the *Eco* β' has just one SBH motif which contributes to selective binding of alternative σ factors. Instead, *Eco* β' harbors a 188-residue insertion in the conserved G-loop (trigger) element (*Eco* domain β' -SI3), which is absent in *Taq/Tth* RNAP. The crystal structure of this flexible domain reveals two repeats of SBHM (Figure 1(e)) that can be modeled into the cryo-EM map of *Eco* RNAP–DNA promoter initiation complex (see Figure 3). *Eco* β' -SI3 is required for interaction with transcript cleavage factors GreA and GreB; it also strongly affects RNAP's intrinsic nucleolytic activity, transcription fidelity, and ability to pause, arrest, and terminate RNA synthesis. On the other hand, *Eco* β contains two domains that are absent in the *Taq/Tth* β : a 115-residue element (β -SI1) inserted into lobe-2 between conserved regions B and C and a 99-residue segment (β -SI2) inserted into lobe-1 between conserved regions G and H. The high-resolution structures of these domains have been recently solved and modeled into the cryo-EM map of *Eco* RNAP (Figure 1(e)). *Eco* β -SI1 contributes to efficient promoter escape by RNAP; it is also targeted by the bacteriophage T4 termination factor Alc, which selectively induces premature transcription termination on *E. coli* DNA. The role of β -SI2 is currently unknown.

Conformational Flexibility

The structural organization of RNAP can be viewed as a combination of stationary (or fixed) and mobile modules. The fixed core module comprises two α NTDs, ω and the bulk of β and β' subunits that carry the catalytic site and form the back wall of all three channels. The mobile modules include elements associated with binding nucleic acids: the upstream and the downstream portions of β' pincer (the N-terminal σ -binding clamp with rudder and switch-2, and the C-terminal dsDNA-binding β' clamp, respectively); the top β pincer (β lobe-1 and lobe-2 carrying β fork-1 and -2); parts of the active center (β' bridge-helix and trigger-loop); and parts of the RNA exit channel (β flap, β' N-terminal Zn-finger, β' zipper, β' lid). These mobile modules confer considerable conformational

flexibility to RNAP structure: the swinging motion of the β and β' clamps, inferred from comparing the structures of *Taq* and *Eco* core enzymes, results in the opening of the claws by ~ 25 Å. This opening is required for initial σ -subunit binding and to allow the template DNA strand to enter the primary channel and reach the catalytic cleft during transcription initiation. It is also important for σ release during transition to the elongation stage of transcription, and for EC dissociation and transcript release during termination. The subsequent closing of the clamp may help RNAP to tightly hold RNA–DNA hybrid in position during processive elongation.

Structure of RNAP Holoenzyme

The high-resolution structure of bacterial holoenzyme is currently available only with *Tth* σ 70, the major housekeeping σ factor. The *Tth* holoenzyme maintains its overall crab claw-like shape since most of σ is bound on the core surface, except for a short central segment of σ (residues 313–342), which is buried in the core molecule (Figures 2(a)–2(d)). The rest of σ is wedged between the upper (β) and lower (β') pincers at the upstream side of the core enzyme, creating a wall that partially blocks the opening of the primary channel. Transition from core to holoenzyme is accompanied by changes in the positions of several structural domains of core, by 2 to 12 Å. The most noticeable are in the β flap domain (especially β flap tip helix by ~ 12 Å); β lobes-1 and -2, α I-NTD and the β' main channel clamp (~ 4 –6 Å). The RNA exit channel is almost completely blocked by the presence of σ .

Structure of σ

Tth σ 70 and all σ 70-like factors share four regions of sequence homology (1–4), which are further divided into subregions (Figure 2(e)). All conserved regions have been implicated in interactions with core and/or DNA. The visible portion of *Tth* σ 70 resembles a curly bracket-shaped structure. It consists of three α -helical domains σ 2, σ 3, and σ 4, and the linker domain σ 3–4. Domain σ 2 comprises four helix–turn–helix (HTH) motifs, encompassing conserved regions 1.2–3.0 and the large nonconserved segment between regions 1.2 and 2.1. σ 3, comprising three α -helical bundle, contains regions 3.0–3.1. The linker domain σ 3–4 is an extended – mostly unfolded – ~ 30 -residue-long hairpin loop, which includes region 3.2. σ 4 is formed by four α -helices arranged as two HTH motifs, and includes conserved regions 4.1 and 4.2 (Figures 2(d) and 2(e)).

σ –Core Interactions

Domains σ 2– σ 4 are located on the enzyme's surface, wedged between the upper (β) and lower (β') pincers at the upstream opening of the primary channel (Figures 2(a)–2(d)). The hairpin loop of σ 3–4 threads through the primary channel reaching the catalytic pocket and comes out from the RNA exit channel. In the holoenzyme, the σ 2, σ 3, and σ 4 domains are situated at an optimal distance to contact the -10 , extended -10 region, and -35 elements of the promoter DNA in RPC, respectively. The location of domain σ 1 containing the

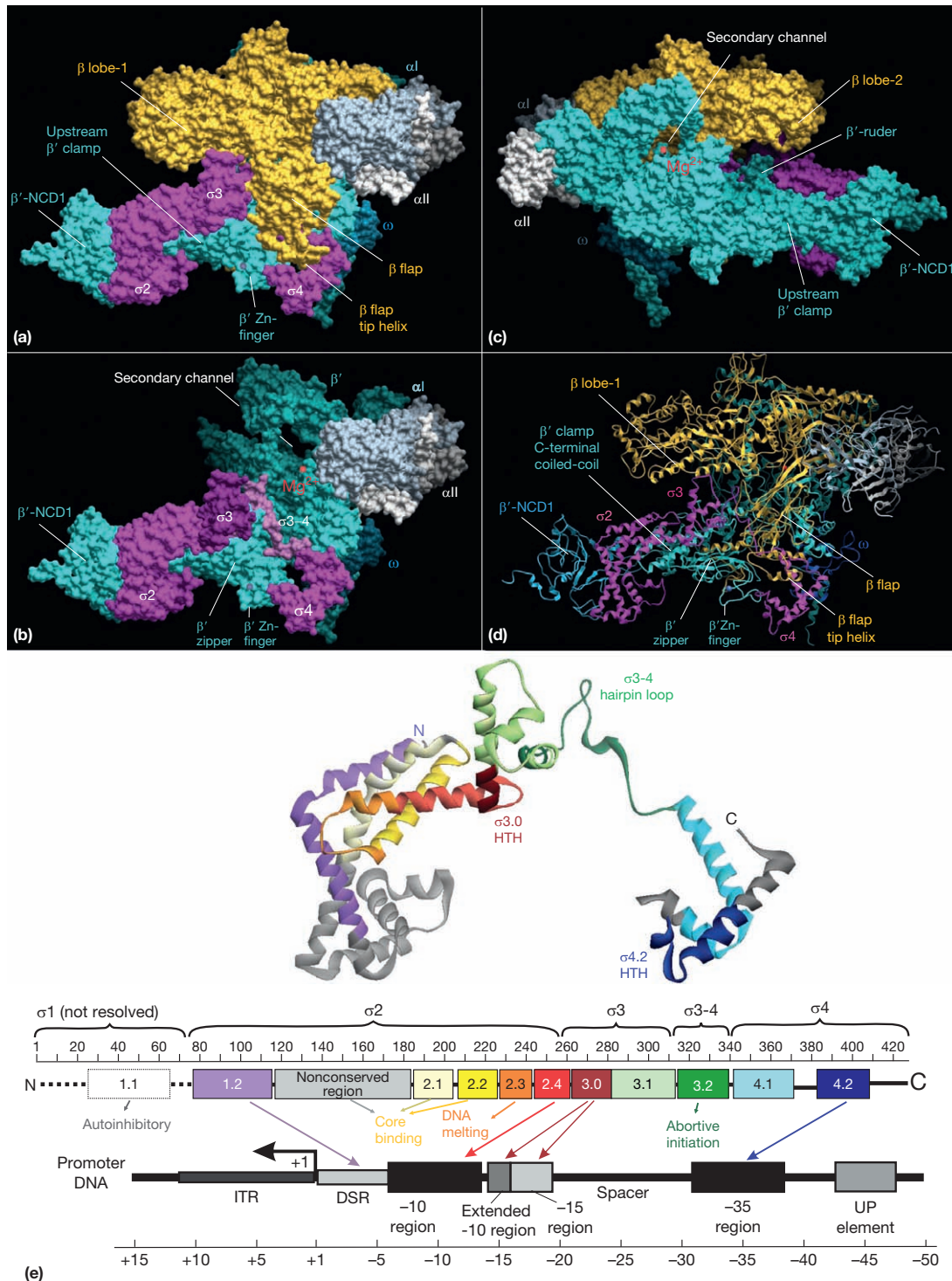


Figure 2 Structural overview of *Tth* RNAP holoenzyme. The structures are shown as molecular surface (a–c) and ribbons (d, e) views using Accelrys DS Visualizer program, with color coding as follows: α I, light gray; α II, slate gray; β , yellow; β' , cyan, σ , magenta, ω , dark cyan. Locations of functionally relevant domains and structural elements are indicated. The Zn^{2+} ion bound in the β' Zn-finger domain is shown as small blue sphere and catalytic Mg^{2+} ion is shown as red sphere. The N-terminal domain of σ carrying region 1.1 missing from the structure is not shown. (a, b, d) The back (RNA-exit channel) view of the holoenzyme. (c) The front (secondary channel) view of holoenzyme is obtained from (a) by $\sim 180^\circ$ rotation about the y -axis. (b) The view of holoenzyme as in (a) with β removed to reveal the inside of the RNA exit channel and σ 3–4 (colored light magenta) buried in it. The location of all visible σ domains is indicated. (e) The structural and functional organization of σ . Top panel is a ribbon view of σ 70 from *Tth* holoenzyme structure shown in (c). Colored regions correspond to the evolutionarily conserved domains of σ as shown in the functional map of σ 70 below. Bottom panel is a linear representation of σ polypeptide with structural domains and conserved regions shown as numbered and color-coded boxes. Underneath is a diagram of DNA promoter regions and interactions made by DNA-binding domains of σ .

unresolved N-terminal portion of σ (residues 1–73), with poorly conserved region 1.1, in the structure of *Tth* holoenzyme is currently unknown. Biochemical and biophysical data suggest that $\sigma 1.1$ possesses autoinhibitory and regulatory functions. In free σ , $\sigma 1.1$ interacts with $\sigma 4$ and blocks σ factor from binding to DNA. In the holoenzyme, $\sigma 1.1$ facilitates the initial association of σ with core and modulates the efficiency of transcription initiation at some promoters.

The σ -core interaction interface covers an extended surface area with multiple cooperative contacts between discrete domains of σ and different parts of the core. This is likely to be the reason for the high stability of the σ -core association ($K_D = 10^{-9}$ M). However, most of these contacts are relatively weak and are distributed over a large area. This explains why different σ factors successfully compete with $\sigma 70$ for binding to core. The strongest interactions are observed between conserved regions 2.1 and 2.2 of $\sigma 2$ and the N-terminal coiled-

coil domain (β' 540–585) of the β' main channel clamp – the major σ docking site (Figure 2(d)). Multiple contacts of β' NCD1 with σ 1.2 and the nonconserved region of σ 2 also contribute to more stable *Tth* σ 70 binding. Additional, although weaker, interactions occur between β flap and σ 4 [53], and between σ 3 and β lobe-1 (Figure 2(d)). In the presence of specific activators, σ 4 also interacts with α -CTD.

All reported crystal structures of RNAP core and holoenzyme still lack a few elements, the most notable of which is an ~80-residue-long α CTD with a ~14-residue flexible linker connecting it to the α NTD. However, the structures of *Eco* α CTD in complex with *Eco* transcriptional activator CAP and DNA are now available, and its position in RNAP is modeled using cryo-electron microscopy structures of *Eco* RNAP–open promoter DNA complex (Figure 3). The α I— and α II— CTDs recognize and bind the upstream (UP) promoter element and serve as targets for many transcriptional activators.

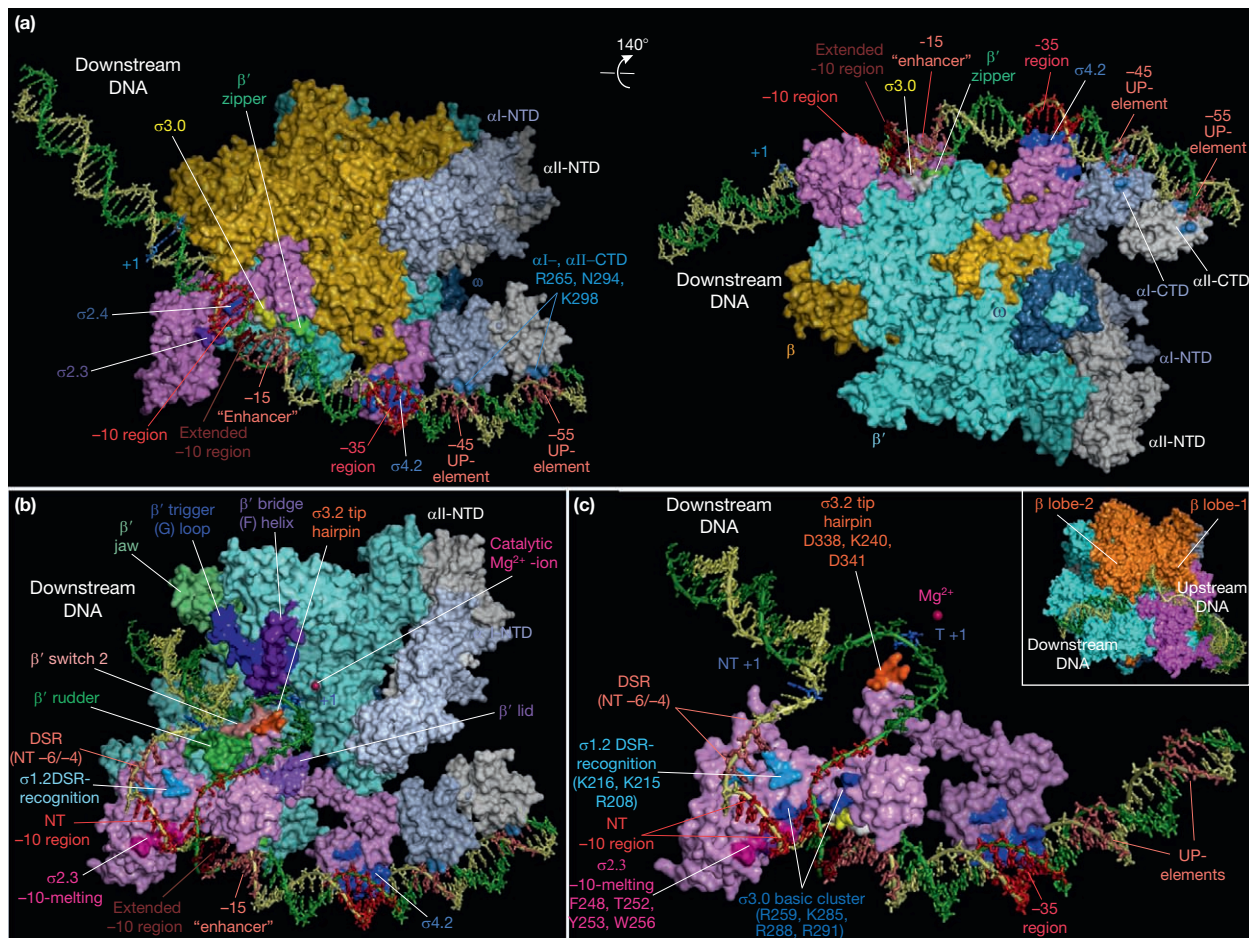


Figure 3 RNAP holoenzyme-promoter DNA complexes. Structural models of *Taq* RPC (a) and RPO (b, c) with Plac promoter DNA (endpoints -65; +25) and α I- and α II-CTDs are shown as molecular surfaces generated by Accelrys DS. The color coding are the same as in **Figure 2**, except for σ which is colored pink. Structural elements of σ , α CTDs and β' involved in interactions with DNA are indicated by different colors. DNA is shown in stick representation with T and NT strands colored dark green and light yellow, respectively, except for promoter elements colored as indicated. Small magenta balls indicate the position of catalytic Mg^{2+} ions. The β' -NCD1 is not shown. (a) Model of *Taq* RPC. Left panel, the RNA exit channel view of RPC (same as in **Figure 2(a)**). Right panel, the bottom (rim of the exit channel) view of RPC, obtained by $\sim 140^\circ$ rotation of the left view about the x-axis. (b) Model of *Taq* RPO is shown with β removed to reveal the inside of the main channel and to show the interactions of σ and β' elements with transcription bubble and downstream DNA. The view is similar to the left view shown in (a) but rotated by $\sim 60^\circ$ about the x-axis. (c) Zoomed-in detailed view of (b) showing σ -promoter DNA interactions. The inset panel shows the full view of RPO and location of β lobe-1 and -2 that close the main channel and block the access to transcription bubble. The view is obtained by rotation of (b) by 90° about y-axis and tilted backward by 25° .

Structure of Holoenzyme–Promoter DNA Complexes

The high-resolution structures of RNAP–DNA binary complexes are yet unavailable. However, the model structures of the RNAP–closed promoter complex (RPC) and the RNAP–open promoter complex (RPO) were reconstructed from low-resolution (~ 15 – $20\text{\AA}$) cryo-electron microscopy studies of *Eco* holoenzyme–*lacUV5* promoter open DNA complexes, and from crystallographic studies of *Eco* CAP– α CTD–DNA complex, *Taq* $\sigma 4$ domain in complex with a DNA fragment containing –35 element (–26 to –37), and of *Taq* holoenzyme binary complex with fork-junction promoter DNA (containing dsDNA from –12 to –45, and ssDNA from –11 to –7) which partially mimics the RPO. The resulting models of RPC and RPO are consistent with most of the biochemical, biophysical, and genetic data accumulated in the last 20 years.

Closed Promoter Complex (RPC)

In the modeled structure of *Taq* RPC, the promoter dsDNA is located on the surface of holoenzyme, outside the RNAP main channel, bound mostly by σ (Figure 3(a)). In RPC the dsDNA is shielded by RNAP from position –55 to –6. Downstream of position –5 dsDNA does not interact with RNAP. All major DNA sequence-specific contacts in RPC, conserved –10, extended –10, and –35 elements of the promoter, are mediated by the $\sigma 2$, $\sigma 3$, and $\sigma 4$ (regions 2.2–2.4, 3.0, and 4.2, respectively) through polar and van der Waals contacts (Figures 2(e) and 3(a)). Specific σ –DNA contacts are summarized in Table 2. Depending on the length of the spacer and the presence of noncanonical sequence elements, the $\sigma 3$ region 3.0 and β' zipper (residues R35, T36, and L37) may contribute to recognition of noncanonical –15 enhancer element, whereas α CTDs (residues R265, N294, and K298) contact A/T-rich sequences at positions from –45 to –65 and up to –90. Because dsDNA is not fully encircled by RNAP in RPC, and because DNA contact area is relatively small and most of the protein–DNA interactions are weak, the RPC is intrinsically unstable. It is sensitive to salt ($>0.2\text{ M NaCl}$) and competitors such as heparin, and can easily dissociate. Though weak, these interactions provide initial promoter recognition by RNAP and increase promoter occupancy. They also cause significant distortion in DNA structure, thereby facilitating further steps in transcription initiation: DNA melting, strand separation, and template strand insertion into the active site cleft. In the RPC the RNAP-bound DNA is

bent or kinked at three places: around –25 (by $\sim 8^\circ$ to accommodate variable spacer length), at the –35 (by $\sim 36^\circ$ induced by insertion of $\sigma 4$ HTH motif into the major groove), and at further upstream of –45 induced by α CTD–DNA minor groove interactions. The DNA bending at –35 aids in proper binding of upstream DNA by α CTD and upstream transcription activators.

Open Promoter Complex (RPO)

The proposed structure of RPO includes both the dsDNA and the ssDNA covering positions from –60 to +25. Unlike RPC, in RPO both strands of DNA up to +20 position are fully enclosed inside the RNAP main channel (Figures 3(a)–3(c)). The RNAP interactions with the upstream portion of dsDNA (from –60 to –17) are similar to that observed for RPC; however, at –16 the DNA trajectory changes due to a sharp bent by 37° toward the RNAP. At position –11 the template (T) and nontemplate (NT) DNA strands separate, and enter different paths for ~ 15 downstream nucleotides until they reanneal at position +3, thus creating the transcription bubble.

Thermal fluctuation and free energy accumulated in the holoenzyme are believed to be sufficient for DNA melting and transcription bubble formation. The initial DNA melting starts at the A/T bp at position –11. Conserved aromatic residues located on the surface of σ region 2.3, *Tth* F248, Y253, W256, W257, and T252 are positioned near the dsDNA–ssDNA junction of the transcription bubble. The nucleotide bases at –8/–9 and –9/–10 interact with F248 and Y253, respectively, whereas W256 and/or W257 stack on the exposed face of the –12 bp, forming the upstream edge of the transcription bubble (see Table 2). Also, W256 is responsible for capturing the exposed, or flipped A base at the key position –11 of the NT-strand, thereby facilitating DNA melting. The DNA NT-strand (from –2 to +4) further continues its path in a groove formed between β lobe-1 and lobe-2 modules (Figures 3b and 3(c)). The conserved $\sigma 1.2$ basic residues (K215, K216, and R208) are implicated in interactions with DNA at positions from –6 to –4 of the DSR region. These interactions stabilize the RPO formation and accelerate transcription initiation at some promoters. Another cluster of conserved basic residues (R259, K285, R288, and R291) of $\sigma 2.4$ and $\sigma 3.0$ is proposed to pull the DNA T-strand (from –7 to +3), through electrostatic interactions, into a groove formed by $\sigma 3$, β' lid, and the β' rudder (Figure 3(c)). The DNA is then placed into the main channel between the active site wall and $\sigma 3$ –4 hairpin loop. Conserved charged

Table 2 Summary of σ –DNA contacts

DNA position (T, NT) ^a	<i>Taq</i> $\sigma 70$ residues	$\sigma 70$ Region (domain)
Major groove at –12A, –12T (base specific)	Q260, N 263	2.4 ($\sigma 2$)
–13, –14, –15 (phosphate backbone)	R237, K241	2.3 ($\sigma 2$)
–13A, –13T (base specific)	E281	3.0 ($\sigma 3$)
Major groove at –14T, –15G (base specific)	H278, E281	3.0 ($\sigma 3$)
–17, –16, –15, –14 (phosphate backbone and base specific)	R274, V277, H278, E281	3.0 ($\sigma 3$)
–31G, –30T (base specific)	R409	4.2 ($\sigma 4$)
–32T (base specific)	R413	4.2 ($\sigma 4$)
–33C (base specific)	E410	4.2 ($\sigma 4$)
–35T (base specific)	R411, Q414	4.2 ($\sigma 4$)
–31, –32, –33; –35, –36 (phosphate/ribose backbone)	R413, R387, L398, E399, R379, T408	4.2 ($\sigma 4$)

^aRoman text is DNA position T, and italic text is DNA position NT.

residues at the tip of the hairpin participate in juxtaposing DNA +1 position to the catalytic center and assist in correct placement of the initiator rNTP (Figures 3(b) and 3(c), Table 2). Simultaneously, the dsDNA downstream of +3 to +12 is brought inside the downstream DNA binding clamp between the β' jaw, β lobe-2, and the downstream β' clamp. The β' switch-2, a small flexible loop residing in the upstream β' clamp in the middle of the main channel cavity, controls the binding of the downstream part of the unwound DNA T-strand in the active site cleft. Switch-2 functions as a hinge mediating opening and closing of the β' clamp, and plays a critical role in downstream propagation of transcription bubble during formation of RPo. It serves as a primary target for the action of antibiotic Myxopyronin (Myx) which binds to β' switch-2 and prevents its refolding. Thereby, Myx interferes with opening of the active center cleft, and blocks entry and unwinding of the downstream DNA. Finally, β lobes-1 and -2 move toward β' clamps by 6–8 Å, which leads to substantial reduction of the main channel opening. This closing hampers DNA dissociation and prevents the collapse of the transcription bubble. Thus, a stable, competitor-resistant and initiation-competent RNAP–promoter DNA complex is formed.

Structure of EC

Several high-resolution structures of bacterial ECs are currently available: (1) the 2.5-Å-resolution structure of *Tth* RNAP core in complex with a synthetic nucleic acid scaffold containing 14-bp-long downstream dsDNA, 9-bp-long RNA/DNA hybrid, and 7-nt-long ssRNA (*Tth* EC); (2) the 3-Å-resolution structures of *Tth* EC co-crystallized with nonhydrolyzable substrate analog, adenosine-5'-[(α,β)-methyleno]-triphosphate (AMPcPP), and with AMPcPP plus antibiotic streptolydigin (Stl); and (3) the 4.1-Å-resolution structure of *Tth* EC with a comparable RNA/DNA scaffold but containing an additional 24-nt-long RNA hairpin, co-crystallized with transcription factor Gfh1. These structures collectively offer a detailed view of the bacterial EC architecture, and help elucidate the mechanisms of RNAP catalysis, that is, substrate binding, incorporation, and enzyme translocation. The structures also provide insights on how the interactions between various structural elements of RNAP, the nucleic acids, substrates, and regulatory molecules determine EC's stability, high processivity, and response to signals encoded in DNA and RNA. Together with the knowledge gained from the structural studies of eukaryotic ECs we have greatly expanded our understanding of the transcription process carried out by multisubunit RNAPs.

Structural Organization of EC and Protein–DNA/RNA Interactions

The structures of *Tth* EC show that 13 bp of the B-form dsDNA resides in the downstream DNA-binding cavity, constricted between the downstream β' clamp, the β' jaw and β lobe-2 (Figure 4(a)). The downstream dsDNA makes very few direct polar contacts with RNAP. Most of the RNAP–DNA interactions including electrostatic and hydrophobic are weak and nonspecific (Table 3); this endows EC with fast translocation and high processivity. The T-strand DNA makes a sharp $\sim 90^\circ$ kink at the junction between the downstream dsDNA and

RNA/DNA hybrid. The kink is induced by the protein barrier in the form of the β' bridge helix and β fork-1 and -2 (Figures 4(a) and 4(b)). The β' rudder loop residing between the dsDNA and DNA/RNA hybrid stabilizes their position and conformation via direct electrostatic contacts (Table 3).

The 9-bp RNA/DNA hybrid is placed at the back wall of the main channel, snugly packed between the upstream β' clamp, β' rudder, and β lobe-1. The hybrid is surrounded by the most conserved structural elements in RNAP including β' DPBB domain with the Mg^{2+} -binding catalytic loop, β' switch-2, β DPBB domain with substrate-binding loops, β switch-3, β RNA-binding helix, β fork-1 with fork-1 helix, and β fork-2 (Figure 4(c)). The latter four elements of β -subunit bind antibiotics Rifampicin (Rif) and Sorangicin (Sor), which strongly inhibit the early steps in RNA synthesis. They act by blocking the path of the growing transcript beyond the length of 2–3 nt, thereby locking RNAP in abortive initiation complex. There are numerous of mixed-mode polar and van der Waals contacts between the phosphate backbone of RNA/DNA hybrid and the conserved basic and hydrophobic residues of the hybrid-binding pocket, respectively (Table 3, Figure 4(c)). These interactions provide an optimal combination of repulsion and attraction forces that ensures a sterically tight placement of the hybrid in the active site cleft, without excessive binding of the nucleic acids. The β fork-2 participates in dsDNA melting by stacking R422 on base pair at $i+2$ register and inducing DNA strand separation (Figures 4(b)–4(d)).

The nascent RNA is separated from the T-strand DNA by the β' lid which sterically blocks the expansion of RNA/DNA duplex beyond 9 bp and fixes the upstream edge of transcription bubble. The β switch-3 also participates in RNA separation by trapping the first displaced RNA nucleotide. The 7-nt-long ssRNA is threaded through the RNA exit channel making weak interaction with the surrounding upstream elements: the β' zipper, β' N-terminal Zn-finger, β' rim helix, and β flap elements (see Table 3, Figures 4(c) and 4(d)).

Compared to the structures of core or holoenzyme, the β' clamp modules in EC are shifted toward β lobes by ~ 4 –6 Å. The closing motion of the clamps substantially reduces the main channel aperture and blocks the access to the RNAP active center. This implies that the unobstructed secondary channel serves as the main conduit for substrates and by-products of RNA synthesis. The observed closure of the clamps also explains the remarkable stability of the EC, for example, its resistance to salt (up to 2 M NaCl) or polyanion DNA competitors (150 $\mu\text{g ml}^{-1}$ heparin).

Structural Rearrangements at the Active Center

The structures of *Tth* ECs are captured in the post-translocated state: the 3'-terminal RNA nucleotide, base-paired with the $i-1$ template DNA base, resides in the active site coordinated by one of the two catalytic Mg^{2+} -ions (Figure 4(e)). The unpaired $i+1$ template DNA base is stacked between the $i-1$ nucleotide of the template and the β' bridge helix, and is positioned in the active site for base pairing with the incoming substrate rNTP (Figure 4(e)). Since the $i+2$ register is base-paired, the substrate can bind only to a single base-specific site, $i+1$. The structure of *Tth* ECs with the nonhydrolyzable substrate analog AMPcPP is consistent with this view (Figures 4(e) and 4(f)).

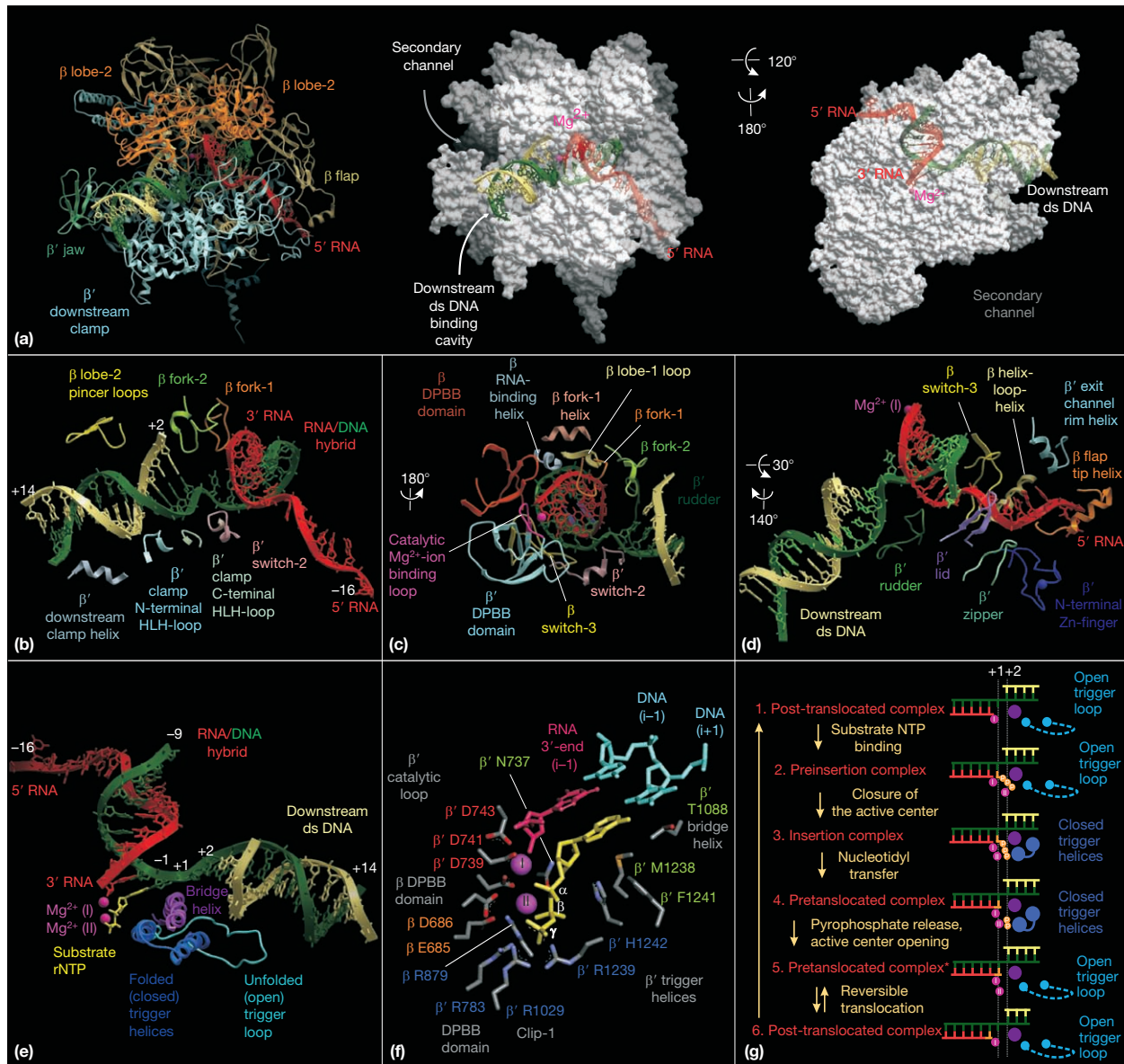


Figure 4 Structure of EC and implications for transcriptional mechanisms. (a) Overall structure of *T7h* EC showing the location and trajectory of nucleic acids. Left panel, the main channel view of EC in ribbon representation. The color code is same as in **Figure 1(a)**. RNA, red; DNA T strand, dark green; DNA NT strand, light yellow. Middle and right panels are the molecular surface views of EC (white) with nascent RNA and RNA/DNA hybrid (shown as colored semitransparent ribbons with sticks) completely buried inside RNAP main channel. Views of the left and middle panels are the same. View in the right panel is obtained by rotation of the middle panel as indicated. (b, c, and d) Close-up view of the RNAP structural elements involved in interaction with the downstream dsDNA (b), RNA/DNA hybrid (c), and nascent RNA (d). Views in (c) and (d) are obtained by rotation of view shown in (b) about vertical and horizontal axis, as indicated. (e) Nucleic acid scaffold and mobile structural elements at the active center. The view is obtained by rotating the view in (b) around y -axis by $\sim 180^\circ$ and tilting forward by $\sim 90^\circ$. The nucleic acids end points and positions of DNA T strand bases at registers -1, +1, and +2 are indicated. Bridge helix and trigger loop are represented as colored ribbons. Trigger loop is shown in two alternative conformations: as folded helices (visible in *T7h* EC-AMPCPP structure), and as unfolded loop (visible in holoenzyme structure), colored blue and cyan, respectively. Substrate AMPcPP bound at the insertion site, yellow sticks; catalytic Mg^{2+} ions, magenta spheres; bridge helix, magenta ribbon. (f) The close-up view of the key components of the catalytic center substrate-insertion site. The view is derived by tilting the view shown in (e) forward by $\sim 20^\circ$. The DNA, RNA, and AMPcPP are represented as cyan, magenta and yellow stick, respectively. Four groups of residues are shown: those that coordinate Mg -I and Mg -II directly (red) or indirectly (orange), those that participate in binding and proper orientation of α -, β -, and γ -phosphates of rNTP (blue), and those responsible for recognition of correct rNTP (light green). (g) Diagram showing the sequential steps of nucleotide addition cycle (NAC). DNA and RNA are depicted as bars, colored coded as in (a-e). Trigger loop, cyan circles connected by long dashed line; trigger helix, blue circles connected by short line; bridge helix, purple circle; Mg^{2+} ions, small magenta circles; rNTP, short orange bar with three small circles (phosphates).

Table 3 Summary of protein–DNA/RNA contacts in EC

	<i>RNA or DNA position (T, NT)^a</i>	<i>Tth β or β' residues</i>	<i>RNAP structural element</i>
Nascent ssRNA	–16 Base	β E770, L773	β Flap tip helix
	–16/–15 Phosphate	β' K671	β' Exit channel rim helix
	–14/–13/–12 Phosphates	β E764	β Flap tip helix
	–12/–11 Phosphate	β Y1013	β Helix–loop–helix
RNA/DNA hybrid	–10/–9 Phosphate	β' R601	β' Rudder
	–9 Base, –9 base	β' A536, V530	β' Lid
	–9/–8 Phosphate	β' R534	β' Lid
	–9/–8 Phosphate, minor groove	β R388	β Lobe-1 loop
	–8, –9 Minor groove	β' R598	β' Rudder
	–8/–7 Phosphate, major groove	β' Q611	β' Switch-2
	–7/–6/–5 Phosphate, minor groove	β Q390, Q393	β Lobe-1 loop
	–5/–4 Phosphate	β N448, R409	β Fork-1
	–4/–3 Phosphate	β R409, R405	β Fork-1 helix
	–4/–3 Phosphate, minor groove	βE1002*, V1001*	β Switch-3
	–3/–2 Phosphate	β Q567, N563	β Rifampicin-binding helix
	–3/–2 Phosphate	β K846	β DPBB domain
	–3/–2 Phosphate	β' R628, R622	β' DPBB domain
	–3/–2 Phosphate	β Q1030, R1031	β Switch-3
	–2/–1 Phosphate	β K846	β DPBB domain
	–2/–1 Phosphate	β' R704	β' DPBB domain
	–2/–1 Phosphate	β Q1030, R1031	β Switch-3
	–2/–1 Phosphate	β' R621	β' DPBB domain
Downstream dsDNA	–1/+1 Phosphate	β' R615	β' Switch-2
	–1/+1 Phosphate	β' R1096	β' Bridge (F) helix
	+1/+2 Phosphate, base	β' G1092, A1089	β' Bridge (F) helix
	+1/+2 Phosphate	β' K610	β' Switch-2
	+2, +2 Base	β R422	β Fork-2
	+2/+3 Phosphate	β' R1096, Y1093	β' Bridge (F) helix
	+3/+4 Phosphate	β' Y1093	β' Bridge (F) helix
	+3/+4 Phosphate	β' Q1441	β' Clamp C-terminal HLH-loop
	+4/+5 Phosphate	β' N1442	β' Clamp C-terminal HLH-loop
	+4/+5 Phosphate	β' R586	β' Rudder
	+4/+5/+6 Phosphate	β' K1266	β' Jaw
	+5/+6 Phosphate	β' K106	β' Clamp N-terminal HLH-loop
	+8/+9 Phosphate, minor groove	β' V108*	β' Clamp N-terminal HLH-loop
	+12/+13 Phosphate, minor groove	β' R486, A487*	β' Downstream clamp helix
	+13/+14/+15 Phosphate	β' R488	β' Downstream clamp helix

^aUnderlined text is RNA; roman text is DNA position T, and italic text is DNA position NT.

AMPcPP binds to EC at the insertion site of the catalytic center where it stacks on the *i*–1 RNA base and forms a base pair with the *i*+1 acceptor template, while its phosphates coordinate two essential Mg²⁺ ions that play a key role in RNA catalysis (Figures 4(e) and 4(f)). In the *Tth* EC–AMPcPP complex the substrate binding induces a refolding of the partially disordered trigger loop (an open conformation observed in the holoenzyme structure) into two α -helices (a closed catalytically active intermediate). Together with the adjacent bridge helix present in straight α -helical conformation, this creates a transient three-helical bundle (Figure 4(e)), which constricts the secondary channel and hinders substrate dissociation. This process is common to all multisubunit RNAPs. Other changes in the EC structure include minor repositioning of the β fork-2 toward the active site (by ~ 1 Å), and a shift of the β' bridge helix (by ~ 1.5 Å) to stabilize the β' trigger helices.

It should be noted that in addition to straight, continuously helical conformation, β' bridge helix can adopt alternative

conformations where its central part has a kink or is flipped out, as observed in the structures of *Taq* core, *Tth* holoenzyme, and *Tth* EC–Gfh1 complex. However, the functional significance of these conformations is unclear. Similarly, in some structures of *Tth* EC as well as yeast RNAP (Pol II EC), the trigger loop displays several partially folded conformations, which appear to be intermediate between fully folded α -helices and unfolded loop, as depicted in (Figure 4(e)). It was proposed that the intermediate conformations of trigger loop (and bridge helix) play an important role for transcription pausing, termination and for the action of transcription factors binding in the secondary channel (such as GreA/B, DksA, and Gfh1).

Structure of the Catalytic Site

At the center of the catalytic site are two essential Mg²⁺ ions (high-affinity Mg-I and low-affinity Mg-II) that play a key

role in RNA catalysis. The two Mg^{2+} ions are surrounded by four highly conserved structural elements of RNAP: the stationary β' DPBB domain with several conserved adjacent loops and helices including the catalytic loop; the stationary β DPBB domain; and the mobile β' bridge helix and β' trigger loop. The high-affinity Mg-I is stably bound in RNAP by a triad of Asp in the β' catalytic loop (D739, D741, and D743) (Figure 4(f)). It is primarily responsible for activation of RNA 3'-terminal OH group and for coordination of the α -phosphate of AMPcPP in the insertion site. The low-affinity Mg-II is held at the insertion site of *Tth* EC, proximal to the location of Mg-I (at a distance of ~ 3.9 Å), mostly by D739 of the β' catalytic loop via direct interactions. Also, residues E685 and D686 of the β DPBB domain and D741 of the β' catalytic loop may contribute to stabilization of Mg-II through additional water-mediated contacts. Mg-II coordinates all three rNTP phosphates providing their optimal spatial alignment for nucleophilic attack of activated 3'-OH group on the α -phosphate. Additionally, the β - and γ -phosphates of rNTP (or pyrophosphate) are stabilized by contacts with basic residues of the β' trigger loop (R1239), β' DPBB domain (R783), β' clip-1 loop (R1029), and β DPBB domain (R879) via hydrogen bonding and electrostatic interactions (Figure 4(f)). The H1242 of the β' trigger loop contacts α -phosphate of rNTP (during RNA synthesis) or the phosphate between RNA $i-1$ and $i+1$ nucleotides (during pyrophosphorolysis and hydrolysis). This interaction greatly facilitates the nucleophilic attack of the 3'-OH group (or the OH^- ion or the water molecule) and is essential for all enzymatic reactions catalyzed by RNAP, that is, RNA synthesis, pyrophosphorolysis, and nucleolytic hydrolysis. To ensure incorporation of correct nucleotides into the RNA, M1238/F1241 residues of the β' trigger loop and T1088 of the β' bridge helix make stacking and hydrophobic/van der Waals contacts with rNTP base and DNA base at $i+1$ position, respectively, while N737 and R704 of the β' catalytic loop establish hydrogen bonds with 2'- and 3'-OH groups of the ribose ring of NTP (Figure 4(f)).

Tth EC-AMPcPP-Stl complex captures AMPcPP in an intermediate preinsertion site in the active center. In the structure, Stl resides in the pocket formed between the β' bridge helix/trigger helix bundle, β fork-2, and the downstream dsDNA ($i+1/i+4$). The Stl binding leads to displacement of the trigger helix from the insertion site, inducing its unfolded (open) conformation. In the absence of the folded trigger helices, the AMPcPP is found in the preinsertion site. It still maintains proper base-pairing with $i+1$ DNA base; however, positions of the α -, β -, and γ -phosphates together with the bound Mg-II ion are shifted by 2–2.5 Å away from Mg-I to a distal location (at a distance of ~ 6.4 Å), due to a $\sim 35^\circ$ rotation around the C4'-C5' bond of the ribose. These observations suggest that AMPcPP binds to EC first in the preinsertion site in an inactive substrate configuration, and, then, in response to trigger helix folding, it isomerizes into a catalytically competent insertion site.

Structural studies revealed that by blocking the folding of the trigger helix, Stl interferes with one of the key steps of nucleotide addition cycle (NAC), thereby inhibiting RNA synthesis. A similar mechanism of action is ascribed to another antibiotic, Tagetitoxin (phytotoxin), which also binds at the base of the secondary channel near the active center and inhibits conformational cycling of trigger loop/trigger helix.

NAC and the Mechanism of Translocation

The high-resolution structures of substrate-bound *Tth* EC offer two key findings:

1. The mobile β' trigger loop/trigger helix is the second essential component of the catalytic center (after the main Mg^{2+} -binding β' catalytic loop). It alternates between the substrate-induced catalytically active α -helical conformation (trigger helix) and the inactive unfolded conformation (trigger loop).
2. In the presence of Mg^{2+} -ion, rNTP can occupy two alternative sites in EC – the catalytically active insertion site stabilized by the trigger helix, and the inactive preinsertion site.

These findings led to a model of NAC and RNAP translocation as illustrated in Figure 4(g). At the beginning of NAC, the EC exists in thermodynamic equilibrium, oscillating between the pretranslocated and posttranslocated states. The rNTP- Mg^{2+} first binds to EC and stabilizes it in a posttranslocated state acting as a ratchet (Figure 4(g), step 1). The binding occurs in a template-dependent manner at the $i+1$ register in the preinsertion site, when the β' trigger loop is in an open, unfolded conformation. Mg-II, coordinated by the three phosphates of rNTP, is placed at a distal position relative to Mg-I, and the preinsertion inactive intermediate complex is formed (step 2). Next, rNTP induces the folding of the trigger loop into a helix, displacing the β' bridge helix and β fork-2, and shifting the position of Mg-II/rNTP triphosphate to a site proximal to Mg-I. This leads to isomerization of EC into the catalytically competent, closed insertion state (step 3). After completion of the catalytic nucleotidyl transfer reaction and transcript extension, the EC transforms into closed, pretranslocated state (step 4). Next, the pyrophosphate-Mg-II complex is released from EC causing destabilization and unfolding of the β' trigger helix. This results in transition of the EC to the pretranslocated state with an open conformation of the trigger loop (step 5). In the last step of NAC, the EC undergoes reversible translocation and returns to the open, posttranslocated ground-state EC capable of binding new rNTP (step 6).

Structural, biochemical, and biophysical data indicate that the energy of NTP hydrolysis is not directly used for EC translocation during elongation. RNAP moves as a Brownian ratchet machine driven forward only by the energy gained from binding the correct substrate. The concerted action of rNTP binding at the catalytic site, the β' trigger loop folding/unfolding and the step-wise transcript extension makes RNAP movement unidirectional. The correct incoming rNTP works as a stationary pawl in the Brownian ratchet preventing RNAP from slipping backward, whereas the β' trigger loop together with the bridge helix and β fork-2 act as a reciprocating pawl, pushing RNAP forward relative to the DNA/RNA scaffold. This system of two pawls couples the forward motion of RNAP with efficient incorporation of correct substrate without falling out of register, nucleotide skipping, or RNA slippage.

Future Prospects

In the past 5 years a great progress has been made in our understanding of the structure and function of bacterial

RNAP, especially its complexes with nucleic acids in the presence of substrates and regulatory molecules. However, our knowledge of RNAP mechanics is still fragmented. Many important details concerning the mechanisms of transcription initiation and termination and their regulation remain to be elucidated. The wish list of RNAP structures that would help address many long-standing questions include complete high-resolution structures of RPc, RPo and their intermediates, EC in pretranslocated state with resolved nontemplate ssDNA and upstream dsDNA, ECs locked in paused and arrested (backtracked) states, termination and anti-termination complexes and their intermediates, and RNAP complexes together with transcription factors, regulatory RNAs, antibiotics, and other small molecules that modulate RNAP activity.

See also: **Molecular Biology: RNA Polymerase II Elongation Control in Eukaryotes**; **T7 RNA Polymerase**.

Further Reading

- Borukhov S and Nudler E (2008) RNA polymerase: The vehicle of transcription. *Trends in Microbiology* 16: 126–134.
- Ho MX, Hudson BP, Das K, et al. (2009) Structures of RNA polymerase–antibiotic complexes. *Current Opinion in Structural Biology* 19: 715–723.
- Hudson BP, Quispe J, Lara-Gonzalez S, et al. (2009) Three-dimensional EM structure of an intact activator-dependent transcription initiation complex. *Proceedings of the National Academy of Sciences of the United States of America* 106: 19830–19835.
- Landick R (2006) The regulatory roles and mechanism of transcriptional pausing. *Biochemical Society Transactions* 34: 1062–1066.
- Murakami KS and Darst SA (2003) Bacterial RNA polymerases: The whole story. *Current Opinion in Structural Biology* 13: 31–39.
- Murakami KS, Masuda S, Campbell EA, et al. (2002) Structural basis of transcription initiation: An RNA polymerase holoenzyme–DNA complex. *Science* 296: 1285–1290.
- Opalka N, Brown J, Lane WJ, et al. (2010) Complete structural model of *Escherichia coli* RNA polymerase from a hybrid approach. *PLoS Biology* 8: e1000483.
- Revyakin A, Liu C, Ebright RH, and Strick TR (2006) Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* 314: 1139–1143.
- Saecker RM, Record MT Jr., and deHaseth PL (2011) Pathway of prokaryotic transcription initiation: Promoter binding, isomerization to initiation-competent open complexes and initiation of RNA synthesis. *Journal of Molecular Biology* 412: 754–771.
- Svetlov V and Nudler E (2009) Macromolecular micromovements: How RNA polymerase translocates. *Current Opinion in Structural Biology* 19: 701–707.
- Tagami S, Sekine S, Kumerevel T, et al. (2010) Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein. *Nature* 468: 978–984.
- Vassilyev DG (2009) Elongation by RNA polymerase: A race through roadblocks. *Current Opinion in Structural Biology* 19: 691–700.
- Vassilyev DG, Sekine S, Laptenko O, et al. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* 417: 712–719.
- Vassilyev DG, Vassilyeva MN, Perederina A, et al. (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* 448: 157–162.
- Vassilyev DG, Vassilyeva MN, Zhang J, et al. (2007) Structural basis for substrate loading in bacterial RNA polymerase. *Nature* 448: 163–168.
- Zhang G, Campbell EA, Minakhin L, et al. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98: 811–824.

RNA Splicing

M S Jurica, University of California Santa Cruz, Santa Cruz, CA, USA

G A Roybal, University of California San Francisco, San Francisco, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Branch point adenosine A specific adenosine in an intron, the 2' hydroxyl of which forms a 2'-5' phosphodiester linkage with the 5' base of the intron during the first chemical step of splicing.

Exon Regions of pre-mRNAs that can be retained in mature messenger RNAs (mRNAs) following splicing.

Introns Noncoding regions of pre-mRNAs that are excised by the splicing process.

Phosphodiester transfer The chemical process in which one phosphodiester bond is broken and another is formed simultaneously.

Polypyrimidine tract The region between the branch point adenosine and 3' splice site, which is rich in cytosine and uridine residues.

Pre-mRNA The primary transcript of a gene synthesized by RNA polymerase II. The primary transcript extends from transcription initiation site to beyond the 3' end of the mature mRNA. It is subject to several processing events including splicing.

Ribozyme An RNA molecule that is capable of catalyzing a chemical reaction.

Spliced-leader A short RNA transcript that is spliced to the 5' end of pre-mRNA transcripts in several species of trypanosomes.

3' splice site The boundary between an exon and the end of an intron.

5' splice site The boundary between an exon and the beginning of an intron.

The study of RNA splicing has impacted several fields of biological research. The discovery of a variety of forms of splicing, including alternative splicing, trans-splicing, and self-splicing sequences, has provided new insights into the increasingly complicated web of gene regulation and the evolution of genomes. Investigations into the molecular mechanisms of the splicing of eukaryotic genes have revealed a complex cellular machinery responsible for the process known as 'splicing'. Furthermore, because splicing is important to the expression of nearly all of human genes, a large number of genetic diseases are linked to splicing. As our understanding of the process advances further, opportunities for developing therapeutic interventions for those diseases will likely arise.

Observation of Split Genes Lead to the Discovery of Splicing

In 1977, Phillip Sharp and Richard Roberts independently reported that eukaryal messenger RNAs (mRNA) do not contain long stretches of sequence that are present in the encoding DNA. These sequences were called 'intervening sequences' and later renamed 'introns'. At first glance, the presence of introns in genes presented a conundrum to the understanding of gene expression at that time, which was encompassed by the central dogma of molecular biology. These inconsistencies were resolved when it was discovered that intron sequences were precisely removed from precursor RNA transcripts (pre-mRNA) while at the same time the flanking sequences, now called 'exons', were joined together into a continuous mRNA before export to the cytoplasm for translation. It is now widely recognized that many eukaryotic genes are not strictly continuous stretches of protein-coding sequence forming intact open reading frames, but rather, that the coding sequences are

interrupted by intron sequence with no obvious function. The intron sequence is precisely removed from RNA transcripts of the genes by the process of pre-mRNA splicing, which results in creation of mRNA with intact open reading frames (Figure 1). Splicing, therefore, is an essential step in the flow of genetic information in eukaryotic cells.

The Chemistry of Pre-mRNA Splicing

Understanding of the chemical mechanism of pre-mRNA splicing was primarily derived from the use of an *in vitro* splicing assay in which a simplified pre-mRNA substrate could be biochemically analyzed. This *in vitro* assay includes a radiolabeled synthetic RNA that is incubated in an extract from the nuclei of the HeLa cell line. Under the appropriate conditions, the splicing machinery within the HeLa nuclear extract recognizes the single intron in the RNA and catalyzes its removal while joining the two flanking exons together. With this system, the basic chemistry of splicing was elucidated to consist of two coordinated phosphodiester transfer reactions (Figure 2). First, the 2' hydroxyl of a specific adenosine residue in the intron attacks at the boundary of the upstream exon and intron, breaking the phosphodiester bond at that site, while creating a new 2'-5' phosphodiester bond between the adenosine and the 5' end of the intron. The reaction loops the intron around the adenosine creating a branched structure at the adenosine called a 'lariat'. Following this first reaction, the newly exposed 3' hydroxyl of the upstream exon attacks at the boundary of the intron and downstream exon. This event creates a normal phosphodiester linkage between the exon sequences and releases the intron in lariat form. The two sites of cleavage at the ends of the intron were termed the '5' and 3' splice sites' and the internal adenosine involved in the first chemical step is known as the 'branch point'.

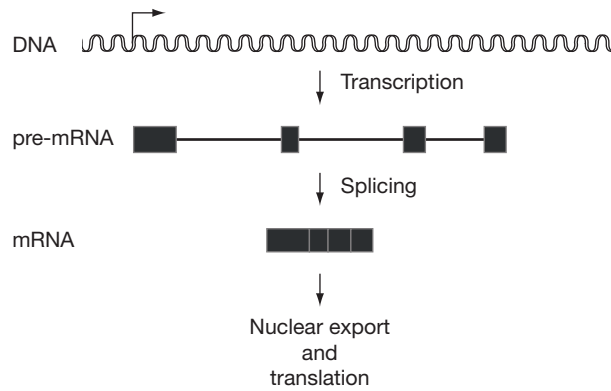


Figure 1 The central dogma of molecular biology explains the transfer of genetic information from DNA → RNA → protein. In the nucleus of a eukaryotic cell, DNA is transcribed to produce pre-messenger RNA (pre-mRNA). By the process of splicing, introns (depicted lines) are removed and exons (depicted as boxes) joined to create messenger RNA (mRNA). The mRNA is then exported to the cytoplasm and translated into protein.

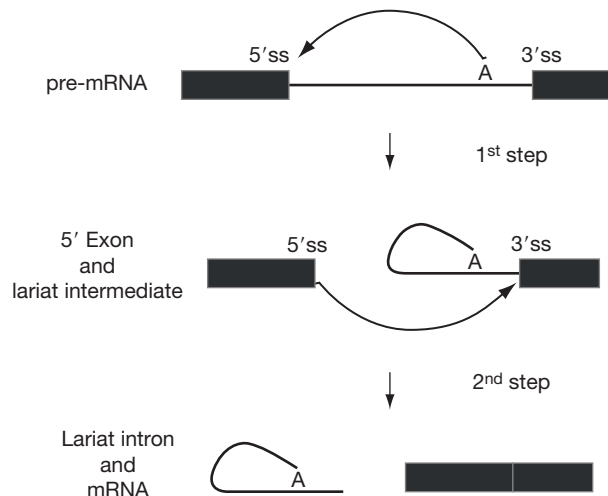


Figure 2 Splicing occurs through two coordinated phosphodiester transfer reactions. In the first chemical step of splicing the 2'-hydroxyl of an adenosine residue (A) within the intron attacks the 5' splice site (5' ss). This reaction produces the intermediate species of pre-mRNA splicing: cleaved 5' exon and lariat intermediate. In the second chemical step, the 3'-hydroxyl of the cleaved 5' exon attacks the 3' splice site (3' ss). This reaction joins the exons to create mRNA and releases the intron in lariat form.

Splicing Consensus Sequences Designate Introns

The 5' and 3' splice sites as well as the branch point adenosine in the intron are specified by sequence elements within the intron and exons (Figure 3). These sequence elements are evolutionarily conserved from yeast to human. For the majority of metazoan introns the 5' splice site is a partly degenerate sequence, although more conserved in the yeast *Saccharomyces cerevisiae*, which has been an important model organism for splicing studies. Within the 5' splice site sequence, the GU

dinucleotide at the beginning of the intron is invariant from yeast to metazoans. The 3' splice site of the major class of metazoan introns is also degenerate, with an invariant AG dinucleotide at the 3' end of the intron. The adenosine at the branch point is contained within a heptamer sequence that is located near the 3' splice site. Again, the sequence is highly conserved in yeast introns and is more degenerate in metazoans. The branch point sequence is usually separated from the 3' splice site by a region composed mainly of C and U residues known as the 'polypyrimidine tract'. The metazoan polypyrimidine tract varies significantly in length and pyrimidine content, while in yeast the tract is composed primarily of U residues and distance between the branch point and 3' splice site is constrained to about 39 nucleotides. There is evidence that the identity of the polypyrimidine tract can influence splice site usage.

As increased numbers of introns have been sequenced, a second class of intron consensus sequences was identified. This class, termed 'minor class,' is found in both vertebrate and invertebrate metazoans. In contrast to the conserved GU-AG sequences that flank the vast majority of introns described above, the 5' and 3' splice sites have, respectively, AU-AC sequences.

An important point to note about splicing consensus sequences is that they alone do not contain enough information to specify splice sites in the context of the vast amount of intron sequence found in metazoan genes. Additional sequences that flank the splice sites also play a role, although more often in an intron-specific fashion. These different layers of information for defining splice sites echo the structure of eukaryotic RNA Pol II promoters, which contain shared core features and gene-specific regulatory elements.

Exon-Intron Architecture of Eukaryotic Genes

The variation in conservation of splicing sequences between yeast and humans is primarily attributed to the differences in the average length and number of introns in yeast and human genes. Only 283 of the ~6000 genes of *Saccharomyces cerevisiae* contain introns, whereas nearly 95% of human genes are predicted to contain introns. Yeast introns on average range from 100 to 400 nucleotides and, with few exceptions, there is only one intron, and therefore two exons, per intron-containing gene. Typically, yeast introns are located in the 5' end of the gene and are removed contrascriptionally. Although less than 5% of yeast genes contain introns, it is of note that these intron-containing genes represent some of the cell's most highly expressed transcripts.

Human genes contain multiple introns (on the order of 8 per gene) that average 11 000 nucleotides, however, with significant variation in length. By contrast, the average length of human exons is ~200 nucleotides. Splicing of human genes is also likely to be a co-transcriptional process, with the first introns emerging from RNA polymerase being removed first. This coupling likely reinforces the ligation of upstream exons to downstream exons such that the order of exons in the RNA transcript is preserved in the mRNA. With alternative splicing, which will be discussed in a later section, exons can be skipped, but exon order is never rearranged.

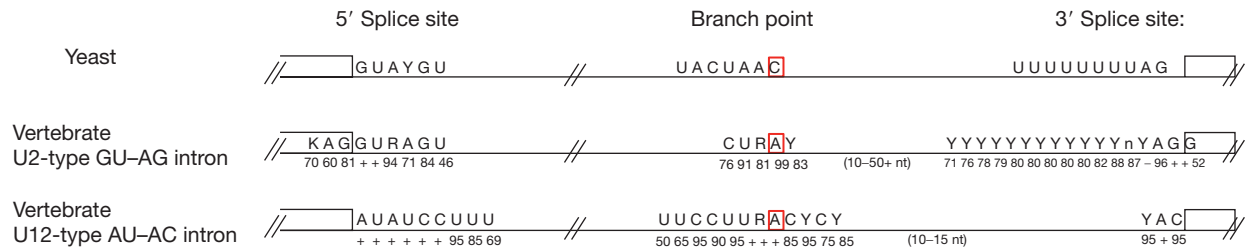


Figure 3 Conserved sequences define the splice sites and branch point adenosine of introns. For yeast, these sequences are nearly universally conserved and the distance between the branch point adenosine (red box) and 3' splice site is usually close to 25 nucleotides. In vertebrates, two types of introns are observed: U2 type and U12 type. For these introns, the splicing sequences are more variable. Shown below each sequence is the percent identity for a given nucleotide. (+) indicates nearly 100% conservation. (n) stands for any nucleotide. (Y) stands for pyrimidine, which is either U or C. (R) stands for purine, which is either A or G. (K) stands for either U or G.

In both yeast and humans, the minimal information content of the splice site consensus sequences means that additional factors must play a role in the regulation of splice site choice. However, the weaker conservation of human introns makes splicing even more dependent on additional regulatory factors, allowing for more complex splicing control. This control is important to the extra levels of regulation that arise with the process of alternative splicing in humans. There is growing evidence that an important step in human splice site recognition is the identification of exons and several factors have been shown to bind specifically to exons and participate in recruitment of the splicing machinery.

The Spliceosome Is the Machinery Responsible for Pre-mRNA Splicing

Eukaryotic cells have evolved an elaborate complex, known as the 'spliceosome', to carry out the two chemical steps of the splicing reaction at each intron. The spliceosome is arguably the most complicated machinery in the cell. It assembles on each intron from well over 100 components on each intron in a highly dynamic manner. The key building blocks of the spliceosome are five ribonucleoprotein (RNP) complexes that each contains a U-rich small nuclear RNA (snRNA), shared core components, and several specific proteins. The complexes, called 'snRNPs' or 'snurps,' are named for their snRNA component: U1, U2, U4, U5, and U6. In addition to the snRNPs, a complex associated with the Prp19 protein (NTC for nineteen complex) and a myriad of other proteins are found in spliceosomes. Together these components play critical roles in identifying conserved *cis*-RNA sequences, assembling the catalytic spliceosome and facilitating splicing chemistry. Given the dynamic complexity of the splicing machinery, a detailed understanding of the mechanics, regulation, and three-dimensional organization of the spliceosome remains elusive.

The roles of several components of the spliceosome have been studied extensively. The snRNAs of the spliceosome, in particular, were shown to have central functions in recognizing splice sites and positioning them in the catalytic core of the complex. U1 snRNA forms base pairs with the 5' splice site consensus early in spliceosome assembly. U2 snRNA base-pairs with the branch point sequence to form a short helix from which the branch point adenosine is bulged, likely to

help position its 2' hydroxyl as the nucleophile for the first step of chemistry. U2 makes extensive base pairs with part of U6 snRNA, while another region of U6 replaces U1 snRNA base pairs with the 5' splice site during catalytic activation of the spliceosome. This intricate network of RNA interactions along with critical roles for specific sequences in the snRNAs for splicing has provided fodder for the hypothesis that snRNAs serve as the catalyst of splicing.

As previously described, a minor class of introns exists with different splice site consensus sequences. These are spliced by the minor spliceosome, which shares many of the same components as the major spliceosome, including U5 snRNP. However, instead of U1, U2, U4, and U6 snRNAs, the minor spliceosome makes use of U11, U12, U4-atac, and U6-atac snRNAs, which differ primarily in the sequences that base-pair with the 5' splice site and branch point.

The functions of some proteins in the spliceosome are understood. Much of this understanding comes from genetic and biochemical studies in the yeast *Saccharomyces cerevisiae* where mutations in spliceosome components can be generated and analyzed. For example, a number of RNA-dependent ATPase proteins have roles in rearranging the structures of the snRNAs as the spliceosome is assembled and activated for catalysis. Because of their important roles, mutations in these proteins result in the accumulation of pre-mRNA in the cells. However, for most proteins in the complex, their functions are still unknown. In particular, there appear to be additional protein factors that associate with metazoan spliceosomes that do not associate with yeast spliceosomes, but their roles in splicing are unclear. Likely, the additional complexity of multiple intron-containing metazoan genes and alternative splicing require these proteins for regulation.

Alternative Pre-mRNA Splicing

Soon after the discovery of splicing, it was noted that some splice products of particular gene transcripts differed in the use of splice sites such that the resulting mRNAs carried different sequences. This phenomenon was termed 'alternative splicing' and is now widely recognized as a key point of gene regulation in metazoans. Well over half of the human genes contain exons that are differentially skipped or included in the final mRNA in response to cellular cues. Alternative splicing leads to a highly

diverse set of messages and protein products being produced from a limited number of genes and plays a key role in creating the complexity contained in, for example, the human nervous system. This important process is discussed in more detail elsewhere in this encyclopedia.

Self-Splicing Introns and the Discovery of Ribozymes

Soon after the initial discovery of introns, Tom Cech and colleagues found a remarkable intron in *Tetrahymena thermophila* that was spliced even in the absence of protein factors. They showed that the intron was autocatalytic in that it promoted its own removal from its associated transcript. Termed a 'ribozyme', the enzymatic function of the intron depends on its ability to fold into a three-dimensional structure. The splicing reaction differs from pre-mRNA splicing in the first step of chemistry where a cofactor guanosine nucleotide is bound and positioned by the intron to attack at the 5' splice site. Exon ligation follows with the free 5' exon attacking at the 3' splice site. With group I intron splicing, the intron is released in linear form and capped with the exogenous guanosine (Figure 4). The intron that Cech characterized and evolutionarily related family members found primarily in bacteria and the organellar genes of plants, protists, and fungi are known as 'group I introns'. The sequence conservation among these introns can be primarily attributed to preservation of secondary structure and tertiary interactions that are important for maintaining the active structure of the ribozyme (Figure 5(a)).

Another structurally distinct family of self-splicing introns (group II) has also been characterized. Like group I, these introns are also found primarily in the organellar genes of protists, but have also been identified in bacteria. Also, like group I, the main selective pressure for group II intron sequence

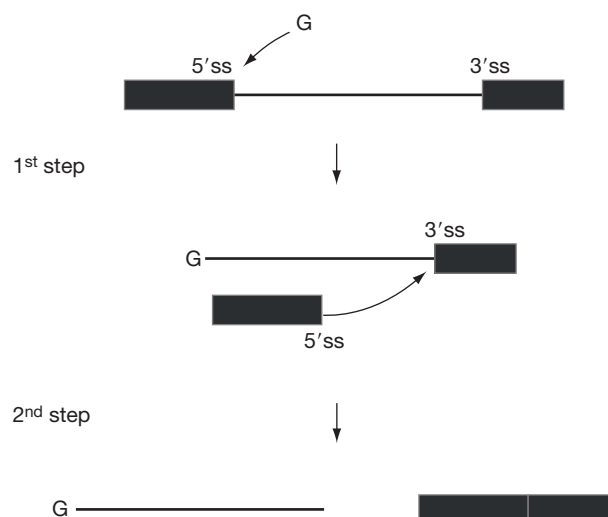


Figure 4 Splicing mechanism for group I introns. The first chemical step of group I intron removal differs from pre-mRNA splicing wherein the 3'-hydroxyl of an exogenous guanosine nucleoside attacks the 5' splice site. Similar to pre-mRNA splicing, the free 5' exon attacks the 3' splice site in the second chemical step. The released intron structure is linear and is capped by the exogenous guanosine.

is preservation of secondary structure and key tertiary interactions that provide the proper stereochemical environment for the two transesterification reactions of splicing (Figure 5(b)). The mechanism of group II intron excision is identical to the splicing of eukaryotic spliceosome introns. A specific adenosine residue in the intron is bulged from a short helix and serves as the nucleophile of the first step of chemistry and produces a lariat intron intermediate. It has also been noted that sequences in group II introns that are essential for catalysis are present in U6 snRNA where they are also important for splicing. The splice site sequences of group II introns (5' GUGYG and 3' AY) also show similarity to spliceosomal introns. Based on these similarities, it is reasoned that group II introns are the evolutionary precursor of the eukaryotic spliceosome.

Although capable of self-splicing *in vitro*, many group I and group II introns require protein cofactors for efficient splicing in cells. In some cases, these proteins are host organism proteins, and in other cases the proteins (termed 'maturases') are encoded by open reading frames located within loop sequences of the intron.

Trans-Splicing

In addition to *cis*-splicing events (transesterification reactions occurring from splice sites within a single transcript), examples of *trans*-splicing events are increasingly reported. A first example is found in trypanosomes, some protists, and some neochordates. In these organisms a short RNA transcript, called the 'spliced leader', is added to the 5' end of protein-coding mRNAs (Figure 6). This addition is accomplished by splicing

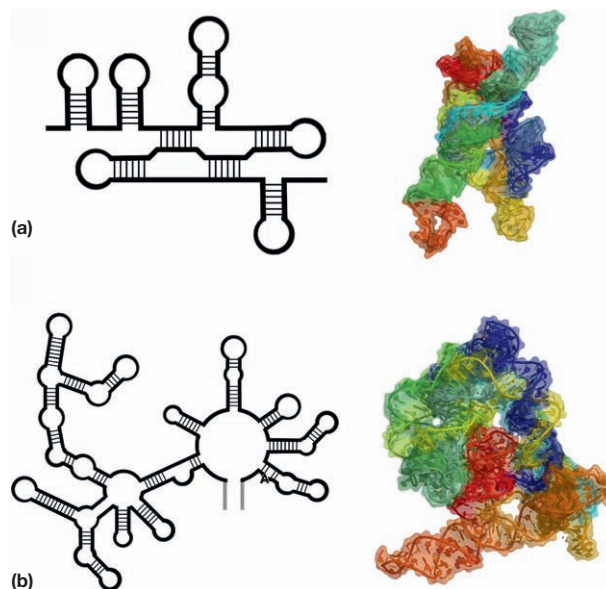


Figure 5 Group I and group II introns have conserved secondary structures and complex tertiary structures. (a) Stereotypical secondary structure of group I introns is schemitized on the left and the three-dimensional structure of a group I intron from an *Azoar* bacterial species (PDB: 1ZZN). (b) Stereotypical secondary structure of group II introns is schemitized on the left and the three-dimensional structure of a group II intron from an *Oceanobacillus* bacterial species (PDB: 3IGI).

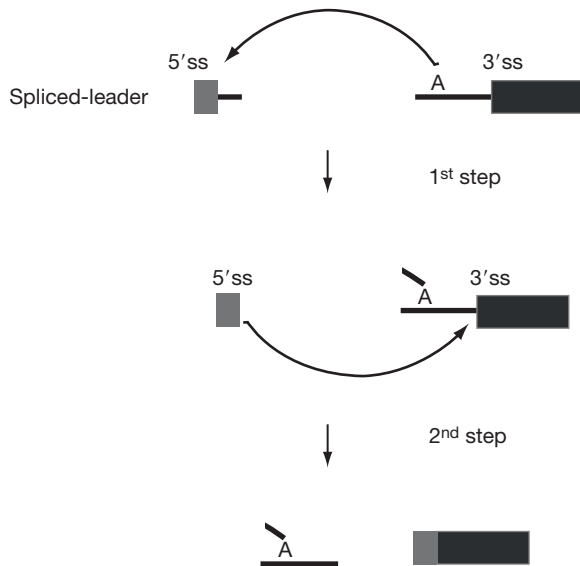


Figure 6 Mechanism of spliced-leader RNA trans-splicing. The spliced-leader RNA sequence contains a 5' splice site that is attacked by the 2'-hydroxyl of a branch point adenosine in the 5' region of a pre-mRNA known as an 'outtron'. The 3'-OH of the cleaved spliced-leader then attacks at the 3' splice site within the pre-mRNA molecule to join the spliced-leader to the pre-mRNA. The outtron is released as a Y-shaped molecule rather than a lariat structure because of its discontinuous structure.

between a 5' splice site in the spliced leader and a 3' splice site at the boundary of 5' region of the pre-mRNA known as an 'outtron' and the following exon. The spliceosomal U snRNPs are responsible for trans-splicing, with the exception of U1 snRNP. The spliced-leader RNA itself along with protein cofactors likely carries out the analogous U1 snRNP function of presenting the 5' splice site to the spliceosome. Again, an adenosine in the outtron upstream of the 3' splice site is the nucleophile for attack in the first step of splicing, although the discontinuous nature of the sequences yield Y-shaped rather than lariat structure. Spliced-leader trans-splicing provides a mechanism to separate individual open reading frames from polycistronic messages that are produced by the host organism and to supply the 5' cap structure required for appropriate translation of mRNA.

Trans-splicing by the spliceosome in other organisms is quite rare. There are a few examples of genes in *Drosophila* with unusual architecture that are resolved by trans-splicing. Included in these examples is the *mod4* gene, which has one exon encoded on the antisense strand that is separately transcribed and joined to the sense transcript by splicing. The *lola* gene shows remarkable trans-splicing between different alleles, such that transcripts from sister chromosomes are joined together. Sporadic reports of other trans-spliced genes have been reported, but it is often not apparent whether these represent biologically programmed events or anomalies of the splicing process.

Another set of well-documented cases of trans-splicing occurs with discontinuous group II introns. For example, the first portion of the intron is located at the end of an RNA gene transcript, while the remaining part of the intron is located at

the beginning of another RNA transcript. These transcripts must come together in order to carry out the splicing reaction, guided in part by base pairing of secondary structural elements of the intron. This type of splicing results in the joining of two exons from separate transcripts. This situation has been found with several organellar genes, and correct splicing is required to generate intact open reading frames for the discontinuous host gene. Given the similarities in splicing mechanism, identity of key nucleotides in the U6 snRNA, and the potential for parts of the group II intron to be expressed in *trans* to the splice sites, a path for evolution of the spliceosome from group II introns has been proposed. Along this path, parts of the group II intron become the snRNAs, which at some point become dependent on proteins for their assembly into a chemically competent catalyst, and eventually introns are fully separated from the splicing machinery.

Evolutionary Implications of Splicing

Data support that self-splicing introns act as progenitors of evolution in that they can behave as mobile genetic elements (thereby creating new genes) and their auto-catalytic activity might be a window back to a world in which RNA alone served as an information repository and carried out the enzymatic functions of life. Although the relationship between the origin of introns and protein-coding genes is not clear, analysis of the increasing number of genome sequences from various organisms indicates that the earliest eukaryotes had introns. A common hypothesis suggests that these arose from mobile group II introns that actively transposed into the genome of the last common eukaryotic ancestor. Indeed, some current group II introns have the ability to retrotranspose using enzymes encoded within the intron.

At first glance, the continued presence of introns in eukaryotic genes begs the question of why these apparently parasitic sequences would be retained through evolution. However, several positive effects have been suggested as possible reasons for introns to be retained and even expanded in the genome. These possibilities include facilitating the creation of new genes by means of rearranging entire exon sequences without disrupting their coding potential. Additionally, splicing provides another point for regulation of gene expression. Alternative splicing appears to be a manifestation of both of these ideas.

Splicing and Human Disease

The spliceosome is charged with two contradictory tasks, in which intron removal must be precise and efficient while still flexible enough to allow for regulation of alternate exon use. Maintaining this balance means that small changes in either sequences in pre-mRNA that direct splicing or the machinery that recognizes these sequences will result in altered gene expression that may result in disease. Several human pathologies are associated with mis-splicing and have been mapped to mutations in *cis*-RNA splicing elements of pre-mRNA or mutations in *trans*-acting factors. Up to 10% of human diseases, the etiology of which is genetic in origin, contain mutations in 5' splice site or 3' splice site consensus sequences.

Examples include atypical cystic fibrosis, neurofibromatosis type 1, and familial dysautonomia. Mutations in sequences that regulate alternative splicing of genes are linked to an increasing number of diseases including cancers, forms of dementia, and Parkinsons. In addition to *cis*-mutation, genetic changes that alter the function of splicing machinery or regulators of alternative splicing are causative for genetic diseases such as retinoblastoma and a range of muscular and neural degenerative diseases. Some examples are spinal muscle atrophy, forms of myotonic dystrophy, and some cerebral ataxias.

If alterations in splicing can cause disease, then it follows that targeting these alterations in patients may yield therapeutic benefits. Indeed, with the disease spinal muscle atrophy, researchers recently reported in a mouse-model the correction of mis-splicing and significant decrease in deleterious symptoms with the introduction of a stable oligonucleotide that targets a mutant regulatory splicing element in the SMN gene. This so-called anti-sense oligotherapy is showing promise in several other splicing disease models. Additionally, a few compounds that target components of the spliceosome or splicing regulators show therapeutic potential. For example, derivatives of a natural product that target a key spliceosome component have been shown to be selectively cytotoxic against a collection of pediatric cancer cell lines. Another promising example is a compound that interferes with the splicing of the HIV transcript. As our understanding of the role of splicing in disease expands, we can expect an increasing arsenal of potential drugs that modulate splicing to mitigate those pathologies.

Summary

When first discovered, introns were suggested to be junk DNA interruptions in protein-coding sequence, and splicing was therefore a chore that cells needed to perform in order to express

their genes. Now we understand that splicing represents an important mechanism for regulating expression of eukaryotic genes. Further exploration of these mechanisms will certainly be important for understanding and intervening in the growing number of human genetic diseases. Looking at the evolution of splicing, it appears that introns likely started as selfish genetic elements that had the surprising ability to self-splice. It has also been suggested that catalytic activity of these introns might be a window back to a world in which RNA was the only enzyme. The study of splicing has opened many fascinating avenues of investigation into molecular biology, evolution, and human health, and one can only hope that the exciting journey will continue.

See also: [Molecular Biology: Alternative Splicing](#).

Further Reading

- Cech TR (1990) Nobel lecture. Self-splicing and enzymatic activity of an intervening sequence RNA from *Tetrahymena*. *Bioscience Reports* 10: 239–261.
- Gingeras TR (2009) Implications of chimaeric non-co-linear transcripts. *Nature* 461: 206–211.
- Gunzl A (2010) The pre-mRNA splicing machinery of trypanosomes: Complex or simplified? *Eukaryotic Cell* 9: 1159–1170.
- Koonin EV (2009) Intron-dominated genomes of early ancestors of eukaryotes. *Journal of Heredity* 100: 618–623.
- Pyle AM (2010) The tertiary structure of group II introns: Implications for biological function and evolution. *Critical Reviews in Biochemistry and Molecular Biology* 45: 215–232.
- Sharp PA (1994) Split genes and RNA splicing. *Cell* 77: 805–815.
- Sharp PA and Burge CB (1997) Classification of introns: U2-type or U12-type. *Cell* 91: 875–879.
- Wahl MC, Will CL, and Luhrmann R (2009) The spliceosome: Design principles of a dynamic RNP machine. *Cell* 136: 701–718.
- Ward AJ and Cooper TA (2010) The pathobiology of splicing. *Journal of Pathology* 220: 152–163.

Roles of Micro-RNAs in Metabolism

Z Wei and G W Wong, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Adipocyte differentiation The process through which adipocyte precursor cells acquire characteristics of mature adipocytes.

Antagomir Antisense oligonucleotide sequence conjugated to cholesterol.

Antisense oligonucleotides Single strands of DNA or RNA that are complementary to a chosen messenger RNA sequence.

Knockdown The reduction of gene expression achieved by certain techniques, such as RNA interference.

Knockout The deletion of gene(s) from the genome of an organism.

β -Cell A type of cell found in the islets of Langerhans in pancreas that produces insulin.

Introduction

Transcriptional regulation of gene expression represents one of several important and fundamental mechanisms governing energy metabolism. There are many different classes of RNAs, one of which is called micro-RNAs (miRNA). miRNAs are small, noncoding RNAs ~19–22 nucleotides in length. They regulate gene expression post-transcriptionally by base-pairing with their cognate target sequences located at the 3'-untranslated regions (3'-UTRs) of protein-coding mRNAs, leading to cleavage or translational inhibition of the target mRNAs. This miRNA-based mechanism of regulating gene expression is evolutionarily conserved from plants to humans. Abundant and ubiquitously expressed, this class of small RNAs has emerged as an important tool kit employed by cells to fine-tune the expression levels of protein-coding genes. As many as 721, 579, and 325 miRNAs have been discovered in the human, mouse, and rat genomes, respectively; annotations of these miRNAs have been deposited in a public miRNA repository database, miRBase. Estimates suggest that ~1–3% of the human and mouse genomes code for miRNA genes, and as much as 30% of the protein-coding genes are regulated by miRNAs in higher eukaryotic species. Therefore, it is not surprising that miRNAs have been shown to play wide-ranging roles in many biological processes, including development, cell differentiation, apoptosis, and immune response. Rapid progress has also been made recently to understand the role of miRNAs in metabolism. Recent studies suggest that miRNAs participate in metabolic processes, including insulin secretion, pancreatic islet cell development, adipocyte differentiation, and insulin signaling, in a variety of tissues. The discovery of miRNAs and their roles in regulating metabolism has added a new dimension to our understanding of the complex metabolic circuitry that governs energy homeostasis.

miRNA Biogenesis

Approximately 720 miRNA coding regions are scattered throughout the human genome. Some miRNA genes exist as independent transcriptional units and are often located in

regions of the genome devoid of protein-coding genes. Other miRNA genes are located in the intronic regions of protein-coding genes. The biogenesis of miRNA starts with transcription initiated by RNA polymerase II or III, generating a long primary miRNA (pri-miRNA) transcript containing a 5'-cap structure and a 3'-poly(A) tail (Figure 1). The complementary region of the pri-miRNA forms a hairpin loop. The loop is recognized by the miRNA processing machinery, composed of Drosha and DGCR8 (DiGeorge syndrome critical region protein 8), which cleaves the structure at the bottom of the stem loop to generate an ~70-nucleotide precursor miRNA (pre-miRNA) transcript. The pre-miRNAs are then exported from the nucleus by the nuclear export machinery (exportin 5 and Ran-GTP). Once in the cytoplasm, the pre-miRNAs are further processed by Dicer, an RNase III-type enzyme, to generate ~22-nucleotide miRNA duplexes. These RNA duplexes are subsequently incorporated into the RNA-induced silencing complex (RISC). One strand of the RNA duplex is discarded while the mature miRNA strand remains in the RISC. The miRNAs then mediate their effects on gene expression by base-pairing with complementary sequences in the 3'-UTR of the target mRNAs, resulting in silencing of the target mRNAs via Argonaute-dependent mRNA cleavage and/or translational repression. Translationally repressed mRNAs are either degraded by the mRNA decay pathway or stored away in a specialized intracellular granule compartment termed the 'P-body'.

miRNA in Pancreatic β -Cell Biology

The evidence for miRNA functions in pancreatic β -cells came from the cloning of small RNAs found in the mouse insulinoma cell line, MIN6. The most abundant miRNA in MIN6 cells is miR-375, which has been shown to negatively regulate glucose-stimulated insulin secretion. Overexpression of miR-375 leads to decreased glucose-stimulated insulin release, while loss of function of miR-375 increases insulin release from pancreatic β -cells. miR-375 regulates insulin secretion partly by repressing the expression of myotrophin (Mtpn), a protein important for intracellular vesicle transport in neurons. Indeed, overexpression of miR-375 in cells decreases the expression of

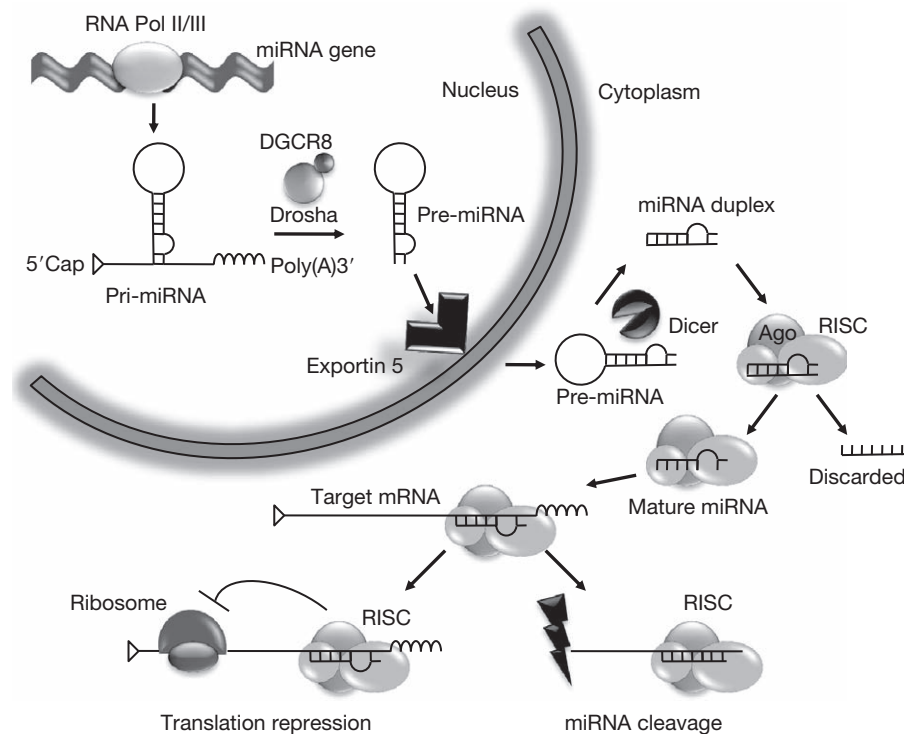


Figure 1 The biogenesis of miRNA.

Mtpn. miR-375 binds directly to the 3'-UTR region of the Mtpn gene, indicating that Mtpn is a direct target of miR-375. Importantly, knockdown of Mtpn mRNA levels by small interfering RNAs (siRNA) also inhibits glucose-stimulated insulin release, consistent with the inhibitory effect of miR-375 on insulin secretion.

miR-9 also negatively regulates insulin secretion, by repressing the expression of a transcriptional repressor, Onecut2 (OC2). One of the genes repressed by OC2 is the guanosine triphosphatase (GTPase) effector, granuphilin/Slp4. Granuphilin is associated with secretory granules and, upon glucose stimulation, negatively regulates the exocytosis of insulin-containing granules from the β -cells. Therefore, when the expression of OC2 is inhibited by miR-9, the expression levels of granuphilin/Slp4 increase. Increased granuphilin/Slp4 results in decreased glucose-stimulated insulin secretion from the β -cells. Because OC2 is a direct target of miR-9, silencing OC2 expression by RNAi mimics the effects of miR-9 on the expression of granuphilin/Slp4 and insulin-containing granules exocytosis.

miRNAs also play a role in the development and proliferation of insulin-producing pancreatic β -cells. For example, miR-124a is abundantly expressed in β -cells and inhibits β -cell development. It acts by targeting Foxa2, a critical transcription factor that regulates β -cell development. Overexpression of miR-124a inhibits the expression of Foxa2 and other essential genes (e.g., Pdx-1, Kir6.2, and Sur-1) involved in β -cell development and function. Knockdown of miR-124a levels in cells abolishes the inhibitory effects on Foxa2 and other targets. In contrast to miR-124, miR-375 positively promotes the development and/or proliferation of β -cells. Mice in which the miR-375 gene has been ablated show significant increases in the numbers of glucagon-producing α -cells, and a decrease in

the numbers of insulin-producing β -cells in their pancreatic islets. These studies support a role for miR-375 in regulating different endocrine cell types in the pancreas.

miRNA in Lipid Metabolism

A study demonstrating how a specific miRNA, miR-14, regulates lipid metabolism in the fruit fly, *Drosophila melanogaster*, first implicated miRNAs in this process. miR-14 negatively regulates intracellular triglyceride levels in the fat body (analogous to mammalian adipose tissue) of *Drosophila*. Ablating the miR-14 gene in *Drosophila* increases triacylglycerol and diacylglycerol levels; however, increasing miR-14 expression lowers these levels. Similarly, miR-278 has been shown to regulate fat body metabolism in *Drosophila*. Loss of function of miR-278 results in decreased fat body size and a corresponding reduction in whole-body triglyceride-to-protein ratio in flies. Interestingly, miR-278-deficient flies also suffer from hyperglycemia despite increased levels of insulin-like peptide, suggesting the presence of insulin resistance in these flies.

Although mammals do not have the equivalent of *Drosophila* miR-14 or miR-278, miRNAs are involved in regulating mammalian lipid metabolism. For example, miR-143 promotes adipogenesis. The expression of miR-143 increases during the course of adipogenesis in tissue culture, as well as in mice. Knockdown of miR-143 levels in human preadipocytes by antisense oligonucleotides causes a significant decrease in adipocyte differentiation. Conversely, overexpression of miR-143 in murine 3T3-L1 preadipocytes accelerates differentiation into mature adipocytes. Similarly, miR-103 promotes adipocyte differentiation. Overexpressing miR-103 in 3T3-L1

preadipocytes leads to increased expression of many adipocyte-specific markers, such as peroxisome proliferator-activated receptor- γ (PPAR- γ) and fatty acid binding protein 4 (FABP4), concurrent with an early increase in triglyceride accumulation during the course of adipocyte differentiation in culture. Two additional miRNA families, miR-200 and miR-17-92, also positively regulate the process of adipocyte development. Overexpression of the miR-200 or miR-17-92 clusters accelerates the conversion of mouse ST2 marrow stromal cells or 3T3-L1 fibroblasts into mature adipocytes, respectively. In contrast, one particular miRNA, miR-27, opposes adipocyte differentiation by targeting and inhibiting the expression of two key transcription factors that drive adipogenesis, PPAR- γ and CCAAT/enhancer binding protein- α (C/EBP- α). These recent studies highlight the intersection of transcription- and miRNA-based mechanisms in controlling adipocyte differentiation.

Unlike in adipocytes, liver miRNA expression is dominated by a single miRNA, miR-122, which exists in $\sim 50\,000$ copies per cell. Abolishment of miR-122 expression in mice using antisense oligonucleotides conjugated to cholesterol (called antagomirs) results in the lowering of plasma cholesterol levels. This effect is mediated, in part, by the miR-122-dependent repression of cholesterol biosynthetic genes in liver, such as the HMG-CoA reductase. A second study, using a similar knockdown approach against miR-122, demonstrated a reduction in fatty acid synthesis and an increase in fatty acid oxidation in primary hepatocytes. These metabolic alterations are accompanied by an increase in adenosine monophosphate (AMP)-activated kinase (AMPK) levels, consistent with the known function of AMPK in promoting fatty acid oxidation while inhibiting fatty acid synthesis in liver.

miRNA in Glucose Metabolism and Insulin Signaling

miRNAs can not only regulate glucose metabolism by affecting the magnitude of insulin secretion from pancreatic β -cells, but also modulate insulin signaling and action in insulin-responsive tissues. Studies comparing miRNA expression profiles in healthy versus type 2 diabetic rats identified two miR-29 family members, miR-29a and miR-29b, that are upregulated in muscle, adipose, and liver of diabetic animals. Overexpression of miR-29 in 3T3-L1 adipocytes suppresses insulin-stimulated glucose uptake, an effect likely due to inhibition of protein kinase B/Akt activation. Consistent with its upregulation in diabetic animals, miR-29 expression is also increased in adipocytes cultured in high glucose and high insulin-containing media. This observation suggests a possible involvement of miR-29 in the development of insulin resistance; however, the mechanism behind the role of miR-29 in insulin resistance is not known. Although insulin-induced gene 1 (Insig1) and caveolin 2 (Cav2) were identified as miR-29 targets, their roles in insulin signaling and glucose metabolism remain to be clarified.

Other miRNAs are also implicated in controlling the insulin receptor signaling pathway. However, the direct effects of these miRNAs on glucose metabolism remain to be established. For example, miR-145 can directly target and suppress the expression of insulin receptor substrate 1 (IRS1), a key component of the insulin receptor signaling pathway. Overexpression of miR-145 in colon cancer cells leads to growth arrest, a

phenotype reminiscent of IRS1 knockdown by siRNA in these cells. Another miRNA cluster, miR-183-96-182, targets multiple components of the insulin receptor signaling pathway, including IRS1, Rasa1, and Grb2. These observations suggest that for a given signaling pathway, multiple miRNAs can target different components of the signaling cascade, effectively shutting it off.

miRNA in the Treatment of Metabolic Diseases

The implication of specific miRNAs in metabolic diseases (e.g., miR-29 in diabetes) has raised the hope that targeting these miRNAs may be a viable therapeutic approach to treat metabolic disorders. Various methods and tools are now available to modulate the expression of specific miRNAs *in vivo*. A gain of function can be achieved *in vivo* using viral-based methods to overexpress a synthetic form of any given miRNA in a desired tissue. A loss of function can be achieved *in vivo* by the delivery of antisense oligonucleotides specific to a target miRNA. In general, these antisense oligonucleotides are chemically modified to increase their stability and binding specificity. Several proof-of-principle studies have successfully demonstrated the feasibility and effectiveness of using adenovirus and lentivirus to deliver miRNA genes in mice. In addition to viral-based strategies, synthetic RNAs (e.g., antisense oligonucleotides) can be successfully delivered to tissues *in vivo* by conjugating synthetic RNAs to cholesterol and encapsulating them within the high- or low-density lipoproteins. However, this approach is limited because the encapsulated synthetic RNAs can be best delivered to only a few tissues such as the liver, intestine, and kidney. Perhaps a bigger problem with targeting miRNA to treat diseases lies in the specificity of the miRNA therapy. First, one miRNA can target several distinct genes and one gene can be targeted by multiple distinct miRNAs. It is not desirable to suppress those miRNAs that are not directly involved in disease pathogenesis. Thus, the potential off-target effects pose a significant challenge to miRNA-based therapy. Second, most of the miRNAs exhibit spatial, temporal, and tissue-specific expression patterns, characteristics necessary to their proper function. Therefore, a generic overexpression of miRNAs may not achieve the desired therapeutic effects. Whether miRNA-based therapeutic agents can be successfully deployed to treat metabolic diseases in clinical settings remains an open question.

Conclusion

miRNAs comprise a diverse family of small noncoding RNAs that play important roles in fine-tuning gene expression and have been implicated in many fundamental physiologic processes. The roles of miRNA in metabolism (Table 1) are still being uncovered, and many miRNAs potentially important to this process are yet to be characterized. The immediate goals in the miRNA community are to validate the authenticity of the predicted miRNA genes *in vivo* and to clone new miRNA genes, establish their function, and identify their mRNA targets. While miRNA-based therapy is an attractive alternative to treat certain metabolic diseases, the questionable specificity of

Table 1 The role of miRNAs in metabolism

miRNA	Metabolic regulation	Target tissues/cell lines	Target genes
miR375	Insulin secretion; α - and β -cell proliferation; β -cell development	Pancreas	Mtpn, Usp1
miR9	Insulin secretion	Pancreas	OC2
miR-124a	Development and proliferation of β -cell	Pancreas	Foxa2, Rab27A
miR-14	Lipid metabolism in fat body	Fat body (<i>Drosophila</i>)	Unknown
miR-278	Lipid metabolism in fat body; insulin resistance	Fat body (<i>Drosophila</i>)	Expanded
miR-143	Adipogenesis	Human preadipocytes; 3T3-L1 cells	ERK5/MAPK7
miR-103	Adipogenesis	3T3-L1	Unknown
miR-200	Adipogenesis	Mouse ST2 cells	Unknown
miR-17-92	Adipogenesis	3T3-L1 cells	Rb2/p130
miR-27	Adipogenesis	3T3-L1 cells	PPAR- γ and C/EBP- α
miR-122	Cholesterol and fatty acid metabolism	Liver	HMG-CoA reductase
miR-29	Diabetes, insulin resistance	Muscle, adipose, and liver	Insig1 and Cav2
miR-145	Insulin signaling; growth	Colon cancer cells	IRS1
miR-183-96-182	Insulin signaling		IRS1, Rasa1, and Grb2

miRNA-based therapy and the challenge of delivering miRNA or anti-miRNA molecules to the desired target tissues *in vivo* will be the key issues that need to be addressed in the near future.

See also: **Molecular Biology:** MicroRNAs in Eukaryotes.

Further Reading

- Baroukh N, Ravier MA, Loder MK, et al. (2007) MicroRNA-124a regulates Foxa2 expression and intracellular signaling in pancreatic beta-cell lines. *Journal of Biological Chemistry* 282(27): 19575–19588.
- Bartel DP (2009) MicroRNAs: Target recognition and regulatory functions. *Cell* 136(2): 215–233.
- Esau C, Davis S, Murray SF, et al. (2006) miR-122 regulation of lipid metabolism revealed by *in vivo* antisense targeting. *Cell Metabolism* 3(2): 87–98.
- Esau C, Kang X, Peralta E, et al. (2004) MicroRNA-143 regulates adipocyte differentiation. *Journal of Biological Chemistry* 279(50): 52361–52365.
- Gauthier BR and Wollheim CB (2006) MicroRNAs: 'Ribo-regulators' of glucose homeostasis. *Nature Medicine* 12(1): 36–38.
- He A, Zhu L, Gupta N, Chang Y, and Fang F (2007) Overexpression of mir-29, highly up-regulated in diabetic rats, leads to insulin resistance in 3T3-L1 adipocytes. *Molecular Endocrinology* 21(11): 2785–2794.
- Krützfeldt J, Rajewsky N, Braich R, et al. (2005) Silencing of microRNAs *in vivo* with 'antagomirs'. *Nature* 438(7068): 685–689.
- Lin Q, Gao Z, Alarcon RM, Ye J, and Yun Z (2009) A role of miR-27 in the regulation of adipogenesis. *FEBS Journal* 276(8): 2348–2358.
- Lynn FC (2009) Meta-regulation: MicroRNA regulation of glucose and lipid metabolism. *Trends in Endocrinology and Metabolism* 20(9): 452–459.
- Pandey AK, Agarwal P, Kaur K, and Datta M (2009) MicroRNAs in diabetes: Tiny players in big disease. *Cellular Physiology and Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology* 23(4–6): 221–232.
- Poy MN, Eliasson L, Krützfeldt J, et al. (2004) A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* 432(7014): 226–230.
- Poy MN, Hausser J, Trajkovski M, et al. (2009) miR-375 maintains normal pancreatic alpha- and beta-cell mass. *Proceedings of the National Academy of Sciences of the United States of America* 106(14): 5813–5818.
- Poy MN, Spranger M, and Stoffel M (2009) MicroRNAs and the regulation of glucose and lipid metabolism. *Diabetes, Obesity and Metabolism* 9(supplement 2): 67–73.
- Tang X, Tang G, and Ozcan S (2008) Role of microRNAs in diabetes. *Biochimica et Biophysica Acta* 1779(11): 697–701.
- Teleman AA, Maitra S, and Cohen SM (2006) *Drosophila* lacking microRNA miR-278 are defective in energy homeostasis. *Genes and Development* 20(4): 417–422.
- Wang Q, Li YC, Wang J, et al. (2008) miR-17-92 cluster accelerates adipocyte differentiation by negatively regulating tumor-suppressor Rb2/p130. *Proceedings of the National Academy of Sciences of the United States of America* 105(8): 2889–2894.
- Xie H, Lim B, and Lodish HF (2009) MicroRNAs induced during adipogenesis that accelerate fat cell development are downregulated in obesity. *Diabetes* 58(5): 1050–1057.
- Xie H, Sun L, and Lodish HF (2009) Targeting microRNAs in obesity. *Expert Opinion on Therapeutic Targets* 13(10): 1227–1238.

Ryanodine Receptor Calcium Ion Channels

G Meissner, University of North Carolina, Chapel Hill, NC, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Antony Galione, volume 3, pp. 465–468, © 2004, Elsevier Inc.

Glossary

Ca²⁺-release channel A channel that releases Ca²⁺, when opened, from an intracellular organelle, commonly from the sarcoplasmic reticulum in muscle and the endoplasmic reticulum in other tissues. The Ca²⁺-release channel is a 2200-kDa complex consisting of four large and four small subunits.

Excitation–contraction (EC) coupling A mechanism that refers to the conversion of an electrical stimulus into a cellular response. In muscle, the electrical stimulus

is usually an action potential and the mechanical response is contraction.

FK506 binding protein (FKBP) It is a 12-kDa immunophilin and subunit of the ryanodine receptors, also referred to as 'calstabin'.

Ryanodine receptor Ca²⁺ channel that binds with high affinity the plant alkaloid ryanodine. Ryanodine opens the Ca²⁺ channel at nanomolar concentrations and closes the channel at micromolar concentrations.

Distribution, Isolation, and Structure of RyRs

RyRs have been characterized in a variety of nonvertebrate and vertebrate species, including flies, crustaceans, birds, and amphibians. In mammals, there are three widely expressed RyR isoforms. RyR1 is the dominant isoform in skeletal muscle. RyR2 is found in high levels in cardiac muscle. RyR3 was initially identified in the brain but is expressed in many tissues including smooth muscle and brain.

The mechanism of RyR ion channel function has been most extensively studied in striated muscles. In skeletal and cardiac muscle, an action potential triggers the intracellular release of Ca²⁺ by a mechanism referred to as 'excitation–contraction (EC)' coupling. The two major membrane structures involved in EC coupling are transverse (T-) tubule membrane system of invaginations (and plasmalemma in myocardium) and intracellular Ca²⁺-storing and Ca²⁺-releasing organelle, the sarcoplasmic reticulum (SR), which contains an adenosine triphosphate (ATP)-dependent Ca²⁺ pump (SERCA) and Ca²⁺-release channel. The release channels typically span the narrow gap where the SR and T-tubule (and plasmalemma) are within ~15 nm of each other. The SR Ca²⁺-release channels are also known as 'feet' or 'junctional processes', and as 'ryanodine receptors' (RyRs) to distinguish them from other intracellular Ca²⁺ channels. The plant alkaloid ryanodine binds with high affinity and specificity to RyRs and modifies their single-channel conductance in a highly characteristic way by inducing a long-lasting subconductance state at low concentrations and closed channels at elevated concentrations.

In mammalian skeletal muscle, an action potential initiates L-type Ca²⁺ channel (Ca_v1.1) protein conformational changes that alter the conformation of closely apposed RyR1s by a direct physical interaction. In cardiac muscle, an action potential results in the influx of extracellular Ca²⁺ through Ca_v1.2 channels. Both mechanisms lead to the release of Ca²⁺ from the SR and subsequent muscle contraction. Sequestration of released Ca²⁺ by the SR Ca²⁺-transporting adenosine triphosphatase (ATPase) (SERCA) and extrusion by the Na⁺–Ca²⁺

exchanger restore the myofibrillar Ca²⁺ concentration from 10^{–6}–10^{–5} to ~10^{–7} M, causing muscle to relax.

The mammalian RyRs share ~70% sequence homology, with the greatest homology in the carboxyl-terminal region. In all isoforms, the C-terminal portion of the protein contains the transmembrane domain. Hydropathy analysis suggests between four and 12 transmembrane segments per RyR subunit. More recent studies suggest that six membrane-spanning segments per RyR subunit are sufficient to form a channel. The remaining RyR amino acids form the large catalytic cytoplasmic foot structure. Experimental information on the structure of the RyRs has been mainly obtained by cryo-electron microscopy. Three-dimensional reconstruction of electron micrographs of frozen-hydrated RyR specimen shows dimensions of 29 × 29 × 12 nm for the large cytoplasmic assembly with a 7-nm length and 8-nm diameter for the transmembrane domain (Figure 1). Cryo-electron microscopy reveals conformational changes that accompany channel opening, shows the location of various effectors on the large RyRs, and provides evidence that the transmembrane region of RyR1 has a pore structure similar to the K⁺ channels.

The large size facilitates isolation of functional RyRs. Good results are obtained using SR vesicle preparations highly enriched in RyRs such as those isolated from rabbit skeletal muscle. SR vesicles are solubilized using the detergent CHAPS and RyR1s are purified in one step as 30 S complexes by sucrose density gradient rate centrifugation.

RyR1 Pore Structure

RyRs have large conductances for both monovalent (~750 pS with 250 mM K⁺) and divalent (~150 pS with 50 mM Ca²⁺) cations, while maintaining divalent versus monovalent selectivity (P_{Ca}/P_K~7). Sequence comparison, cryo-electron microscopy, and extensive mutagenesis, and single channel studies indicate that the RyRs have a pore structure similar to that of K⁺ channels whose structure is known (Figure 2). In the

RyRs, the residues between the two C-terminal membrane-spanning segments are lumenally located and have a pore helix and an amino acid motif (GGGIG) that is similar to the selectivity filter motif (TV/IGYG) of K^+ channels. Mutagenesis shows that rings of negative charges in the luminal and cytosolic vestibules maintain the high rates of RyR1 ion fluxes. Several models of RyR ion permeation based on Eyring rate theory, Poisson–Nernst–Planck-density functional theory, and the solution structure of K^+ channels have been described.

Regulation by Ca^{2+} and Endogenous Effectors

The *in vitro* function of RyR ion channels has been primarily studied by measurement of (1) Ca^{2+} fluxes from SR vesicles, (2) single channels incorporated in planar lipid bilayers, and (3) [3H]ryanodine binding. These studies have shown that RyRs are Ca^{2+} -gated channels that are activated by micromolar

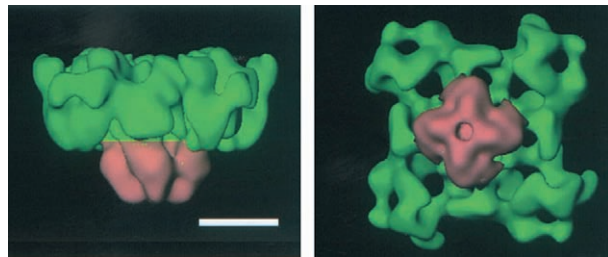


Figure 1 Three-dimensional reconstruction of RyR1. The cytoplasmic (green) and transmembrane (pink) domains are shown. Scale bar, 10 nm. Reproduced from Samso M and Wagenknecht T (1997) Contributions of electron microscopy and single-particle techniques to the determination of the ryanodine receptor three-dimensional structure. *Journal of Structural Biology* 121: 172–180, with permission from Academic Press.

Ca^{2+} concentrations and inhibited by millimolar Ca^{2+} concentrations. Mg^{2+} inhibits RyRs by binding to high-affinity and low-affinity Ca^{2+} -binding sites. ATP complexed with Mg^{2+} modifies regulation of RyRs by Ca^{2+} indicating that it is an allosteric modulator of Ca^{2+} -gated RyR activity. Similarly, monovalent cations and anions affect Ca^{2+} -gated channel activity by binding to specific regulatory sites. In addition to cytosolic Ca^{2+} , SR luminal Ca^{2+} regulates RyRs at several channel sites. Ca^{2+} binds to luminal channel sites and by passing through the channel accesses cytosolic Ca^{2+} activation and inactivation sites.

RyRs and Calmodulin

Calmodulin (CaM) is a small, cytoplasmic Ca^{2+} -binding protein that regulates cellular activities including protein kinases and ion channels. A notable property of the RyRs is their direct interaction with CaM in the absence and presence of Ca^{2+} . Ca^{2+} -free and Ca^{2+} -bound forms of CaM bind to a highly conserved CaM-binding domain of RyRs. Cryo-electron microscopy indicates that Ca^{2+} -free and Ca^{2+} -bound forms of CaM bind to an overlapping region distal to the effector site (ion pore). This suggests that CaM exerts its effects allosterically over relatively long-range interactions within the RyRs. CaM inhibits the three mammalian RyR isoforms at micromolar Ca^{2+} concentrations. At submicromolar Ca^{2+} concentrations, RyR2 is also inhibited by CaM but RyR1 and RyR3 are activated to different extents by CaM. Site-directed mutagenesis in CaM-binding domain alters CaM binding and regulation to different extents, which suggests that other RyR type-specific sites are involved in transducing CaM binding into functional effects.

Crystal structure analysis of the complex between the Ca^{2+} -bound form of CaM and peptide corresponding to the CaM-binding domain of RyR1 (p3614–3643) reveals an interaction

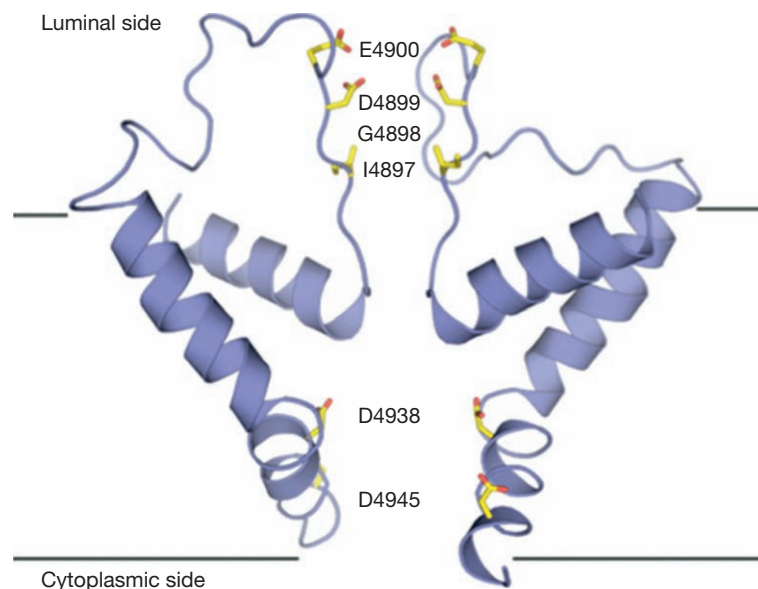


Figure 2 All-atom pore model of RyR1. Two of the four subunits are shown in cartoon representation. Residues whose mutation affects channel function are shown in stick representation. From Ramachandran S, Serohijos AWR, Xu L, Meissner G, and Dokholyan NV (2009) A structural model of the pore-forming region of the skeletal muscle ryanodine receptor (RyR1). *PLoS Computational Biology* 5: 1–10.

of the C-terminal and N-terminal domains of CaM with two hydrophobic anchor residues, W3620 and F3636, respectively, at a novel 1–17 spacing. Nuclear magnetic resonance (NMR) and fluorescence resonance energy transfer (FRET) measurements indicate that the C-terminal domain of CaCaM tightly binds to the N-terminal end of the peptide, and furthermore show that in the complex each domain can interact independently with its own half of the peptide.

Expression of RyR1 mutant deficient in the regulation by CaM in mice suggests that CaM has a modest role in *in vivo* regulation of RyR1. By contrast, CaM inhibition of RyR2 is required for normal cardiac function. Mice with point mutations in the CaM-binding domain of the cardiac ryanodine receptor ion channel have large hearts and die at an early age. Sustained Ca^{2+} transients reveal an abnormal Ca^{2+} handling in homozygous mutant cardiomyocytes.

RyRs and Protein Kinases

Primary sequence analysis suggests the presence of many phosphorylation sites in the large cytoplasmic foot region of the RyRs. Kinases and phosphatases that are part of the RyR1 and RyR2 multiprotein complexes include protein kinase A (PKA), Ca^{2+} /CaM-dependent protein kinase II (CaMKII), and protein phosphatase 1 and protein phosphatase 2A (PP2A). RyR1 has been reported to be phosphorylated at Ser2843. RyR2 is phosphorylated at Ser-2030 by PKA, Ser-2809 by PKA, and CaMKII (corresponding to Ser2843 in RyR1) and Ser-2815 by CaMKII. However, there is evidence for additional phosphorylation sites in RyR1 and RyR2. Phosphorylation of RyRs by protein kinases has been suggested to result in leaky SR Ca^{2+} channels (see below).

RyRs and Reactive Oxygen and Nitrogen Species

The tetrameric mammalian skeletal and cardiac RyRs have 404 and 364 cysteines, with 100 and 89 cysteines per 560 kD subunit, and one and two per FK506-binding protein, respectively. Nearly half of the 404 cysteines in the tetrameric RyR1 channel complex are free in the presence of 5 mM reduced glutathione (GSH) at $\text{pO}_2 \sim 10$ mm Hg, that is, under conditions that simulate pO_2 and glutathione redox state. An increase in oxygen tension from ~ 10 mm Hg to ambient air ($\text{pO}_2 \sim 150$ mm Hg) results in loss of 6–8 free thiols/RyR1 subunit without appreciably changing RyR1 activity. Substitution of GSH with oxidized glutathione (GSSG) at $\text{pO}_2 \sim 10$ mm Hg or ambient air reduces the number of free thiols/RyR1 subunit and results in a large increase in RyR1 activity. RyR2s also have functional thiols that respond to the cells' pO_2 and GSH/GSSG redox state. Regulation of RyR1 by SepN1, a selenoprotein, and RyR1 and RyR2 by plasmalemmal- and SR-associated NAD(P)H oxidases has been described.

Both RyR1 and RyR2 are endogenously S-nitrosylated, suggesting that NO or NO-related species are physiological modulators of skeletal and cardiac muscle EC coupling. Although in normal skeletal and cardiac muscle, the predominant skeletal (nNOS) and cardiac (eNOS) nitric oxide isoforms are targeted to the plasmalemma, a specific association of nNOS

with RyR2 has been described. This suggests that eNOS and nNOS have specific roles in mediating S-nitrosylation and regulation of plasmalemmal $\text{Ca}_v1.2$ and RyR2, respectively.

In *in vitro* studies, NO and NO-related species activate or inhibit RyRs depending on donor concentration, membrane potential, the presence of channel agonists, and other sulfhydryl-modifying reagents. RyR1 is S-nitrosylated and activated by low physiological concentrations of NO at tissue pO_2 (~ 10 mm Hg) but not at ambient air ($\text{pO}_2 \sim 150$ mm Hg). This suggests that RyR1 is an O_2 and NO-sensing molecule. Like RyR1, RyR2 activity is dependent on pO_2 . However, RyR2 is not effectively activated and S-nitrosylated by NO *in vitro*, although it is activated by other endogenous NO-related species.

A number of reactive thiols have been identified. 7-Diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin (CPM) selectively labels seven hyperreactive RyR1 cysteines (Cys-1040, 1303, 2436, 2565, 2606, 2611, and 3635). In another study, nine RyR1 cysteines (Cys-36, 315, 811, 906, 1591, 2326, 2363, 3193, and 3635) are endogenously modified and another three (Cys-253, 1040, and 1303) are modified by exogenous reactive oxygen and nitrogen molecules. RyR1-Cys3635 is specifically S-nitrosylated by NO at low, physiological O_2 concentrations.

Pharmacology of RyRs

The RyRs constitute a rich pharmacological target for controlling SR Ca^{2+} release. Exogenous effectors found to affect RyR function include ryanoids, toxins, caffeine, and other xanthines, anthraquinones, phenol derivatives, dantrolene, local anesthetics, and polycationic and sulfhydryl-reacting reagents.

RyR-Associated Proteins

A large number of junctional SR membrane proteins and cytosolic and SR luminal proteins have been reported to form a macromolecular complex with the RyRs and to modulate their activity. These include kinases, phosphatases, CaM (see above), S100A1, triadin, junctin, JP-45, junctate, juncophilin, homer, sorcin, and calsequestrin.

RyRs and Muscle Disorders

Mutations in RyR1 give rise to muscle diseases, malignant hyperthermia (MH), central core disease (CCD), and multi-minicore disease (MmD). MH is an inherited disease that causes a rapid rise in body temperature and muscle contraction when the affected person receives general anesthesia. MH-linked mutations were initially mapped to hot spots within the NH_2 and central domains; however, more recently identified MH mutations appear to be distributed throughout the entire RyR1 coding sequences including the C00H-terminal domain of the channel. CCD and MmD are congenital diseases associated with a dysfunctional RyR1. In many cases, dominant RyR1 mutations linked to CCD localize to the COOH-terminal domain of the channel. Single-channel measurements

indicate that CCD mutants lose the ability to conduct Ca^{2+} . A CCD mutation in the selectivity filter of RyR1 (I4897T) (Figure 2) provided early evidence regarding the pore structure of the RyRs. Crystal structure of the N-terminal domain has identified the location of several disease-associated mutations in RyR1 and RyR2.

In humans, RyR2 mutations are associated with catecholaminergic polymorphic ventricular tachycardia and arrhythmogenic right ventricular cardiomyopathy. A deficiency in SR luminal cardiac calsequestrin and missense mutation D307H also result in imbalances of Ca^{2+} handling and catecholaminergic polymorphic ventricular tachycardia. Overexpression of CSQ2 results in suppressed SR Ca^{2+} transients and severe cardiac hypertrophy in mice. Abnormal Ca^{2+} handling is also observed in mice that lack or overexpress RyR-associated proteins, triadin, and junctin.

Sustained stimuli associated with stress can lead to maladaptive changes of β -adrenergic receptor signaling, cardiomyocyte death, and heart failure. In failing hearts, RyR2 has been reported to be PKA hyperphosphorylated, depleted of the FKBP12.6 subunit, and to exhibit increased channel activity. Similarly, in skeletal muscle of animal models with heart failure and patients with heart diseases, exercise is linked to hyperphosphorylation and FKBP12 depletion of the skeletal muscle RyR1, increased RyR1 channel activity, and decreased exercise capacity. These findings led Marks and colleagues to propose that PKA hyperphosphorylation and dissociation of the small FKBP subunit from the RyRs result in increased RyR sensitivity to cytosolic Ca^{2+} , referred to as leaky SR Ca^{2+} channels. However, other laboratories have failed to confirm this. More recent studies indicate that CaMKII-mediated RyR2 phosphorylation also results in the formation of leaky SR Ca^{2+} channels.

The small 12- and 12.6-kDa FK506-binding proteins are predominantly associated with RyR1 and RyR2, respectively. They belong to the family of immunophilins, exhibit *cis/trans*-isomerase activity, and their removal is reported to functionally uncouple groups of channels and increase channel activity. Mechanisms implicated in generation of dysfunctional release channels include PKA-mediated phosphorylation, increased S-nitrosylation, oxidation, and loss of PP1, PPA2, and phosphodiesterase 4D from the RyR macromolecular complexes. A 1,4-benzothiazepine derivative, JTV519 (also known as K201), and the more specific derivative S107, have been reported to improve muscle function by stabilizing the RyR–FKBP complex. To define the specificity of RyR1 and RyR2 phosphorylation, mutants were prepared that mimic the dephosphorylated (alanine) or phosphorylated (aspartic acid) forms. In some studies, Ser to Asp substitution increases open channel probability (P_o) and decreases FKBP-binding affinity, whereas mutants with a Ser to Ala substitution exhibit a behavior similar to wild type. In other studies, neither substitution alters channel behavior.

See also: **Bioenergetics:** Arachidonic Acid Regulated Calcium Channel; Calcium-Binding Proteins: Cytosolic (Annexins, Gelsolins, and C_2 -Domain Proteins); ER/SR Calcium Pump: Function.

Further Reading

- Amador FJ, Liu S, Ishiyama N, et al. (2009) Crystal structure of type I ryanodine receptor amino-terminal beta-trefoil domain reveals a disease-associated mutation "hot spot" loop. *Proceedings of the National Academy of Sciences of the United States of America* 106: 11040–11044.
- Bers DM (2001) Cardiac excitation–contraction coupling. *Nature* 415: 198–205.
- Couchonnal LF and Anderson ME (2008) The role of calmodulin kinase II in myocardial physiology and disease. *Physiology* 23: 151–159.
- Franzini-Armstrong C and Protasi F (1997) Ryanodine receptors of striated muscles: A complex channel capable of multiple interactions. *Physiological Reviews* 77: 699–729.
- George CH, Jundi H, Thomas NL, Fry DL, and Lai FA (2007) Ryanodine receptors and ventricular arrhythmias: Emerging trends in mutations, mechanisms, and therapies. *Journal of Molecular Cell Cardiology* 42: 34–50.
- Lai FA, Erickson HP, Rousseau E, Liu QY, and Meissner G (1998) Purification and reconstitution of the calcium release channel from skeletal muscle. *Nature* 331: 315–319.
- Lanner JT, Georgiou DK, Joshi AD, and Hamilton SL (2010) Ryanodine receptors: Structure, expression, molecular details, and function in calcium release. *Cold Spring Harbor Perspectives in Biology* 11: a003996.
- Lehnart Z and Marks AR (2007) Modulation of the ryanodine receptor and intracellular calcium. *Annual Review of Biochemistry* 76: 367–385.
- Meissner G (2004) Molecular regulation of cardiac ryanodine receptor ion channel. *Cell Calcium* 35: 621–628.
- Ramachandran S, Serohijos AWR, Xu L, Meissner G, and Dokholyan NV (2009) A structural model of the pore-forming region of the skeletal muscle ryanodine receptor (RyR1). *PLoS Computational Biology* 5: 1–10.
- Rios E and Pizarro G (1991) Voltage sensor of excitation–contraction coupling in skeletal muscle. *Physiology Reviews* 71: 849–908.
- Robinson R, Carpenter D, Shaw MA, Halsall J, and Hopkins P (2006) Mutations in RYR1 in malignant hyperthermia and central core disease. *Human Mutation* 27: 977–989.
- Samso M, Feng W, Pessah IN, and Allen PD (2009) Corordinated movement of cytoplasmic and transmembrane domains of RyR1 upon gating. *PLoS Biology* 7: e85.
- Samso M and Wagenknecht T (1997) Contributions of electron microscopy and single-particle techniques to the determination of the ryanodine receptor three-dimensional structure. *Journal of Structural Biology* 121: 172–180.
- Stamler JS and Meissner G (2001) Physiology of nitric oxide in skeletal muscle. *Physiological Reviews* 81: 209–237.
- Stowell KM (2008) Malignant hyperthermia: A pharmacogenetic disorder. *Pharmacogenomics* 9: 1657–1672.
- Takeshima H, Nishimura S, Matsumoto T, et al. (1989) Primary structure and expression from complementary DNA of skeletal muscle ryanodine receptor. *Nature* 339: 439–445.
- Treves S, Jungbluth H, Muntoni F, and Zorzato F (2008) Congenital muscle disorders with cores: The ryanodine receptor calcium channel paradigm. *Current Opinion in Pharmacology* 8: 319–326.
- Tung CC, Lobo PA, Kimlicka L, and Van Petegem F (2010) The amino-terminal disease hot spot of ryanodine receptors forms a cytoplasmic vestibule. *Nature* 468: 585–588.

Secretases

R L Henrikson, The Pharmacia Corporation, Kalamazoo, MI, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 4, pp. 7–10, © 2004, Elsevier Inc.

Glossary

Alzheimer's disease (AD) A disease first described by Alois Alzheimer in 1906 characterized by progressive loss of memory and cognition. AD afflicts a major proportion of our aging population and is one of the most serious diseases facing our society today, especially in light of increasing human longevity. The secretases represent important potential therapeutic intervention points in AD treatment.

Proteinases Enzymes that hydrolyze, or split, peptide bonds in protein substrates; they are also referred to as proteolytic enzymes.

Secretase A proteinase identified with respect to its hydrolysis of peptide bonds within a region of a type I integral membrane protein called APP. These cleavages are responsible for liberation, or destruction, of an amyloidogenic peptide of about 40 amino acid residues in length called A β .

β -Amyloid (A β) (The peptide produced from amyloid precursor protein (APP) by the action of β - and γ -secretases. A β shows neurotoxic activity and aggregates to form insoluble deposits seen in the brains of Alzheimer's disease (AD) patients. The α -secretase hydrolyzes a bond within the A β region and releases fragments which do not aggregate.

Secretases are proteolytic enzymes involved in the processing of an integral membrane protein known as amyloid precursor protein, or APP. β -Amyloid (A β) is a neurotoxic and highly aggregative peptide that is excised from APP by secretase action, and that accumulates in the neuritic plaque found in the brains of Alzheimer's disease (AD) patients. The amyloid hypothesis holds that the neuronal dysfunction and clinical manifestation of AD are a consequence of the long-term deposition and accumulation of A β , and that this peptide of 40–42 amino acids is a causative agent of AD. Accordingly, the secretases involved in the liberation, or destruction, of A β are of enormous interest as therapeutic intervention points toward treatment of this dreaded disease.

Background

Proteolytic enzymes play crucial roles in a wide variety of normal and pathological processes in which they display a high order of selectivity for their substrate(s) and the specific peptide bonds hydrolyzed therein. This article concerns secretases, the membrane-associated proteinases that produce, or prevent the formation of, a highly aggregative and toxic peptide called β -amyloid (A β). This A β peptide is removed from a widely distributed and little understood type I integral membrane protein called APP. The apparent causal relationship

between A β and AD has fueled an intense interest in the secretases responsible for its production. Here, we discuss the current understanding of three of the most-studied secretases, α -, β -, and γ -secretases. A schematic representation of the A β region of APP showing the amino acid sequence of A β and the major sites of cleavage for these three secretases is given in [Figure 1](#). A β is produced by the action of β - and γ -secretases, and there is an intense search underway for inhibitors of these enzymes that might serve as drugs in the treatment of AD. The α -secretase cleaves at a site near the middle of A β , and gives rise to fragments of A β that lack the potential for aggregation; therefore, amplification of α -secretase activity might be seen as another approach to AD therapy.

α -Secretase

The activity responsible for cleavage of the Lys¹⁶–Leu¹⁷ bond within the A β region ([Figure 1](#)) is ascribed to α -secretase. This action prevents formation of the 40–42-amino-acid residue A β and leads to release of soluble APP α and the membrane-bound C83-terminal fragment. α -Secretase competes with β -secretase for the APP substrate, but the α -secretase product, soluble APP α (pathway A, [Figure 1](#)), is generated at a level about 20 times that of the sAPP β released by β -secretase (pathway B). Because α -secretase action prevents the

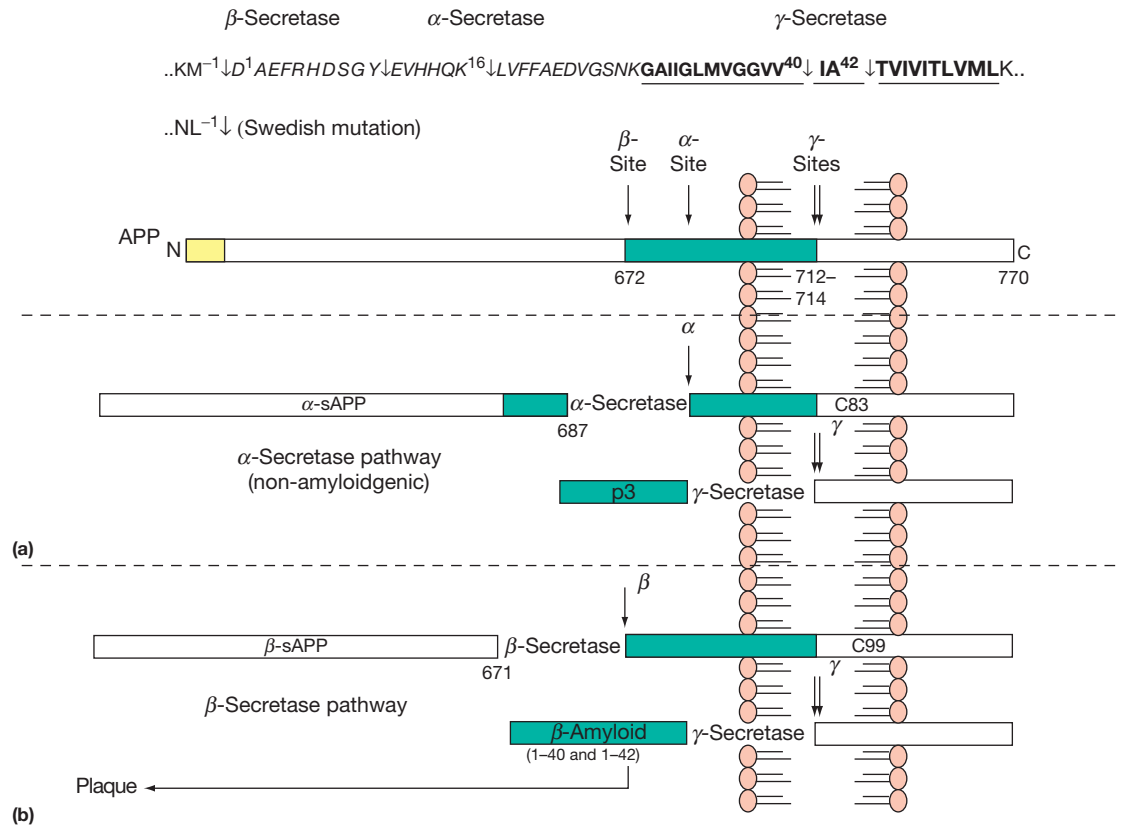


Figure 1 A schematic overview of APP processing by the α -, β -, and γ -secretases. The top panel shows the amino acid sequence of APP upstream of the transmembrane segment (underlined, bold), and encompassing the sequences of A β ₁₋₄₀ and A β ₁₋₄₂ (D¹-V⁴⁰, and D¹-A⁴², respectively). The β -secretase cleaves at D¹ and Y¹⁰; the α -secretase at Lys¹⁶, and the γ -secretase at Val⁴⁰ and/or Ala⁴². Below the sequence is a representation of APP emphasizing its membrane localization and the residue numbers of interest in β - and γ -secretase processing. (a) The non-amyloidogenic α -secretase pathway in which sAPP α and C83 are generated. Subsequent hydrolysis by the γ -secretase produces a p3 peptide that does not form amyloid deposits. (b) The amyloidogenic pathway in which cleavage of APP by the β -secretase to liberate sAPP β and C99 is followed by γ -secretase processing to release β -amyloid peptides (A β ₁₋₄₀ and A β ₁₋₄₂) found in plaque deposits.

formation of the toxic A β peptide, augmentation of this activity could represent a useful strategy in AD treatment, and this has been done experimentally by activators of protein kinase C (PKC) such as phorbol esters and by muscarinic agonists. The specificity of the α -secretase for the Lys¹⁶-↓-Leu¹⁷ cleavage site (Figure 1) appears to be governed by spatial and structural requirements that this bond exists in a local α -helical conformation and be within 12 or 13 amino acids distance from the membrane. α -Secretase has not been identified as any single proteinase, but two members of the ADAM family, ADAM-10 and ADAM-17, are candidate α -secretases (ADAM, a disintegrin and metalloprotease). ADAM-17 is known as a tumor necrosis factor- α -converting enzyme (TACE), and TACE cleaves peptides modeled after the α -secretase site at the Lys¹⁶-↓-Leu position. This was also shown to be the case for ADAM-10; overexpression of this enzyme in a human cell line led to several-fold increase in both basal and PKC-inducible α -secretase activity. As of now, it remains to be proved whether α -secretase activity derives from either or both of these ADAM family metalloproteinases, or whether another as-yet unidentified proteinase carries out this processing of APP.

β -Secretase

The enzyme responsible for cleaving at the amino-terminus of A β is β -secretase (Figure 1). In the mid-1980s, when A β was recognized as a principal component of AD neuritic plaque, an intense search was begun to identify the β -secretase. Finally, in 1999, several independent laboratories published evidence demonstrating that β -secretase is a unique member of the pepsin family of aspartyl proteinases. This structural relationship to a well-characterized and mechanistically defined class of proteases gave enormous impetus to research on β -secretase. The preproenzyme consists of 501 amino acids, with a 21-residue signal peptide, a prosegment of about 39 residues, the catalytic bilobal unit with active site aspartyl residues at positions 93 and 289, a 27-residue transmembrane region, and a 21-residue C-terminal domain. The membrane localization of β -secretase makes it unique among mammalian aspartyl proteases described to date. Another interesting feature of the enzyme is that, unlike pepsin, renin, cathepsin D, and other prototypic members of the aspartyl proteases, it does not appear to require removal of the prosegment as a means of

activation. A furin-like activity is responsible for cleavage in the sequence Arg-Leu-Pro-Arg-↓-Glu²⁵ of the proenzyme, but this does not lead to any remarkable enhancement of activity, at least as is seen in recombinant constructs of pro- β -secretase. β -Secretase has been referred to by a number of designations in the literature, but the term β -site APP cleaving enzyme (BACE) has become most widely adopted. With the discovery of the β -secretase, it was recognized that there was another human homolog of BACE with a transmembrane segment and this has now come to be called BACE2. This may well be a misnomer, as the function of BACE2 is yet to be established, and it is not clear that APP is a normal substrate of this enzyme. At present, BACE2 is not considered to be a secretase.

There is considerable experimental support for the assertion that BACE is, in fact, the β -secretase involved in APP processing. The enzyme is highly expressed in brain, but is also found in other tissues, thus explaining the fact that many cell types can process A β . The use of antisense oligonucleotides to block expression of BACE greatly diminishes production of A β and, conversely, overexpression of BACE in a number of cell lines leads to enhanced A β production. BACE knockout mice show no adverse phenotype, but have dramatically reduced levels of A β . This demonstrates not only that BACE is the true β -site APP processor, but also that its elimination does not pose serious consequences for the animal, a factor of great importance in targeting BACE for inhibition in AD therapy.

Much of the evidence in support of the amyloid hypothesis comes from the observation of mutations near the β - and γ -cleavage sites in APP that influence the production of A β and correlate directly with the onset of AD. One such mutation in APP, which invariably leads to AD in later life, occurs at the β -cleavage site where Lys-Met⁻¹ is changed to Asn-Leu⁻¹ (Figure 1). This so-called Swedish mutation greatly enhances the production of A β , and, as would be expected, β -secretase hydrolyzes the mutated Leu-Asp¹ bond in model peptides ~50 times faster than the wild-type Met-Asp¹ bond. It is important to recognize that BACE cleavage is required for subsequent processing by the γ -secretase; in this sense, a BACE inhibitor will also block γ -secretase. Another BACE cleavage point is indicated in Figure 1 by the arrow at Y¹⁰↓E¹¹; the A β ₁₁₋₄₀ or ₄₂ subsequently liberated by γ -secretase action also forms amyloid deposits and is found in neuritic plaque.

In all respects, therefore, BACE fits the picture expected of β -secretase, and because of its detailed level of characterization and its primary role in A β production, it has become a major target for development of inhibitors as drugs to treat AD. Great strides in this direction have become possible because of the availability of three-dimensional (3-D) structural information on BACE. The crystal structure of BACE complexed with an inhibitor is represented schematically in Figure 2. Homology with the pepsin-like aspartyl proteases is reflected in the similar folding pattern of BACE, with extensive β -sheet organization, and the proximal location of the two aspartyl residues that comprise the catalytic machine for peptide bond cleavage. The C-terminal lobe of the molecule is larger than is customarily seen in the aspartyl proteases, and contains extra elements of structure with as-yet unexplained impact on function. In fact, before the crystal structure was solved, it was thought that this larger C-terminal region might contribute a spacer to distance the catalytic unit from the



Figure 2 Schematic representation of the 3-D structure of the BACE (β -secretase) catalytic unit as determined by X-ray crystallography. Arrows and ribbons designate β -strands and α -helices, respectively. An inhibitor is shown bound in the cleft defined by the amino- (left) and carboxyl- (right) terminal halves of the molecule. The C-terminus of the catalytic unit is marked C to indicate the amino acid residue immediately preceding the transmembrane and cytoplasmic domains of BACE. These latter domains were omitted from the construct that was solved crystallographically. The arrow marks a disulfide bridge, which maintains the C-terminus in close structural association with the body of the catalytic unit. The catalytic entity as depicted sits directly on the membrane surface, thereby restricting its motion relative to protein substrates. Image: Courtesy of Dr. Lin Hong, Oklahoma Medical Research Foundation, Oklahoma City, OK.

membrane and to provide mobility. This appears not to be the case. As denoted by the arrow in Figure 2, there is a critical disulfide bridge linking the C-terminal region just upstream of the transmembrane segment to the body of the molecule. Therefore, the globular BACE molecule is proximal to the membrane surface and is not attached via a mobile stalk that would permit much motion. This steric localization would be expected to limit the repertoire of protein substrates that are accessible to BACE as it resides in the Golgi region. Crystal structures of BACE/inhibitor complexes have revealed much about the nature of protein-ligand interactions, and information regarding the nature of binding sites obtained by this approach will be of critical importance in the design and development of inhibitors that will be effective drugs in the treatment of AD.

γ -Secretase

γ -Secretase activity is produced in a complex of proteins and is yet to be understood in terms of the actual catalytic entity and mechanism of proteolysis. This secretase cleaves bonds in the middle of the APP segment that traverses the membrane (underlined and boldface in Figure 1), and its activity is exhibited subsequent to cleavages at the α - or β -sites. In Figure 1, the γ -secretase cleavage sites are indicated by two arrows. Cleavage at the Val⁴⁰-Ile⁴¹ bond liberates the more abundant 40-amino-acid residue A β (A β ₁₋₄₀). Cleavage at Ala⁴²-Thr⁴³ produces a minor A β species, A β ₁₋₄₂, but one that appears to be much more hydrophobic and aggregative, and it is the 42-residue A β that is believed to be of most significance in AD pathology. As was the case for APP β -site mutations, there

are human APP mutants showing alterations in the vicinity of the γ -site, and these changes, powerfully associated with the onset of AD, lead to higher ratios of $A\beta_{1-42}$.

Central to the notion of the γ -secretase is the presence of presenilins, integral membrane proteins with a mass of ~ 50 kDa. There are a host of presenilin mutations in familial AD (FAD) that are associated with early-onset disease and an increased production of the toxic $A\beta_{1-42}$. This correlation provides strong support for the involvement of presenilin in AD, and its presence in γ -secretase preparations implies that it either is a proteolytic enzyme in its own right, or can contribute to that function in the presence of other proteins. In fact, much remains to be learned about the presenilins; it has been difficult to obtain precise molecular and functional characterization because of their close association with membranes and other proteins in a complex. Modeling studies have predicted a variable number of transmembrane segments (6–8); however, presenilin function is predicated upon processing by an unknown protease to yield a 30-kDa N-terminal fragment (NTF) and a 20-kDa C-terminal fragment (CTF). These accumulate *in vivo* in a 1:1 stoichiometry within high-molecular-weight complexes with a variety of ancillary proteins. Some of the cohort proteins identified in the multimeric presenilin complexes displaying γ -secretase activity include catenins, armadillo-repeat proteins that appear not to be essential for γ -secretase function, and nicastrin. Nicastrin is a type I integral membrane protein with homologs in a variety of organisms, but its function is unknown. It shows intracellular colocalization with presenilin, and is able to bind the NTF and CTF of presenilin as well as the C83 and C99 C-terminal APP substrates of γ -secretase. Interestingly, downregulation of the nicastrin homolog in *Caenorhabditis elegans* gave a phenotype similar to that seen in worms deficient in presenilin and notch. Evidence that nicastrin is essential for γ -secretase cleavage of APP and notch adds to the belief that nicastrin is an important element in presenilin, and γ -secretase function. Efforts to delineate other protein components of γ -secretase complexes and to understand their individual roles in the

enzyme function represent a large current research effort. Recently, two additional proteins associated with the complex have been identified through genetic screening of flies and worms. The *aph-1* gene encodes a protein with seven transmembrane domains, and the *pen-2* gene codes for a small protein passing twice through the membrane. Both of these putative members of the γ -secretase complex are new proteins whose functions, either with respect to secretase activity or in other potential systems, remain to be elucidated.

At present, it is still unclear as to how γ -secretase exerts its function. What is known, however, is that γ -secretase is able to cleave at other peptide bonds in APP near the γ -site in addition to those indicated in [Figure 1](#), and is involved with processing of intramembrane peptide bonds in a variety of additional protein substrates, including notch. This lack of specificity is a major concern in developing drugs for AD targeted to γ -secretase that do not show side effects due to inhibition of processing of these additional, functionally diverse protein substrates.

See also: [Cell Architecture and Function: Metalloproteinases, Matrix](#); [Protein/Enzyme Structure Function and Degradation: Amyloid](#).

Further Reading

- Esler WP and Wolfe MS (2001) A portrait of Alzheimer secretases – new features and familiar faces. *Science* 293: 1449–1454.
- Fortini ME (2002) γ -Secretase-mediated proteolysis in cell-surface-receptor signaling. *Nature Reviews* 3: 673–684.
- Glenner GG and Wong CW (1984) Alzheimer's disease: Initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biophysical Research Communications* 120: 885–890.
- Hendriksen ZJVRB, Nottet HSLM, and Smits HA (2002) Secretases as targets for drug design in Alzheimer's disease. *European Journal of Clinical Investigation* 32: 60–68.
- Sisodia SS and St. George-Hyslop PH (2002) γ -Secretase, notch, $A\beta$ and Alzheimer's disease: Where do the presenilins fit in? *Nature Reviews* 3: 281–290.

Secretory Pathway

S Shikano and K J Colley, University of Illinois, Chicago, IL, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Karen J. Colley, volume 4, pp. 11–16, © 2004, Elsevier Inc.

Glossary

Chaperone A protein that aids in the folding and assembly of other proteins, frequently by preventing the aggregation of folding intermediates.

Cisternal maturation/progression One model of protein transport through the Golgi apparatus that suggests that secretory cargo enters a new cisternae that forms at the *cis* face of the Golgi stack, and that this cisternae and its cargo progresses or matures through the stack by the sequential introduction of Golgi modification enzymes.

Cytoskeleton A cellular scaffold formed from filament proteins in the cell cytoplasm. Microfilaments, intermediate filaments, and microtubules give shape to the cell and, together with motor proteins, participate in the movement of membrane compartments and vesicles.

Glycosylation The modification of proteins and lipids with carbohydrates in the endoplasmic reticulum and Golgi apparatus of the secretory pathway.

Proteasome Cytoplasmic protein complex that mediates the degradation of proteins. Proteins targeted for degradation are usually modified with ubiquitin.

Secretory pathway An intracellular pathway consisting of the endoplasmic reticulum, Golgi apparatus, and associated vesicles that is responsible for the folding, modification, and transport of proteins to the lysosome, plasma membrane, and extracellular space.

Translocon A protein channel in the endoplasmic reticulum membrane that allows the signal-peptide-mediated entry of newly synthesized proteins into the endoplasmic reticulum lumen and the secretory pathway.

Vesicular transport One model of protein transport through the Golgi apparatus that suggests that secretory cargo moves sequentially between stationary Golgi cisternae in transport vesicles and is modified by resident Golgi enzymes in the process.

Introduction

The eukaryotic cell is separated into several functionally distinct, membrane-enclosed compartments (Figure 1). Each compartment contains proteins required to accomplish specific functions. Consequently, each protein must be sorted to its proper location to ensure cell viability. Proteins possess specific signals, either encoded in their amino acid sequences or added as posttranslational modifications, which target them for the various compartments of the cell. The pioneering work of Dr. George Palade provided scientists with their first picture of the functional organization of the mammalian secretory pathway. Later work showed that the secretory pathway acts as a folding, modification, and quality-control system for proteins that function in the endoplasmic reticulum (ER) and Golgi apparatus, and for those that are targeted to the lysosome, plasma membrane, and extracellular space (Figure 1). This article focuses on protein targeting to and within the compartments of the secretory pathway, and how proteins within this pathway function to ensure that correctly folded and modified proteins are delivered to the cell surface and secreted from cells.

Targeting of New Proteins to the Secretory Pathway

What Kinds of Proteins Are Targeted to the Secretory Pathway?

The proteins that are targeted to the secretory pathway can be separated into two groups – those that function in the ER and Golgi to ensure proper protein folding and modification

(i.e., resident proteins) and those that are processed in the ER and Golgi and are transported to later compartments such as the lysosome, plasma membrane, and extracellular space (Figure 1). Each of these proteins not only possesses a signal to enter the secretory pathway but may also have a secondary signal to localize it to a particular organelle within the pathway.

Signals and Mechanisms of Secretory Pathway Entry

The 1999 Nobel prize in physiology or medicine was awarded to Dr. Günter Blobel for his contributions to our understanding of the mechanism of secretory pathway entry. Dr. Blobel and his colleagues found that to enter the secretory pathway, proteins are synthesized with an amino terminal peptide (called the signal peptide) that allows them to cross the membrane of the ER. The signal peptide is a stretch of 13–16 hydrophobic amino acids and is recognized by a complex of proteins and ribonucleic acid called the signal recognition particle (SRP) (Figure 2). As the signal peptide emerges from the ribosome during translation, SRP binds to it and halts translation, and then targets the new protein–ribosome complex to the cytoplasmic face of the ER membrane where it binds to the SRP receptor. Subsequently, the new protein–ribosome complex is released from SRP and its receptor, and transferred to an aqueous membrane channel known as the ‘translocon’. Translation resumes and the new protein is co-translationally transferred through the translocon into the lumen of the ER, where in many cases the signal peptide is cleaved by a specific signal peptidase (Figure 2).

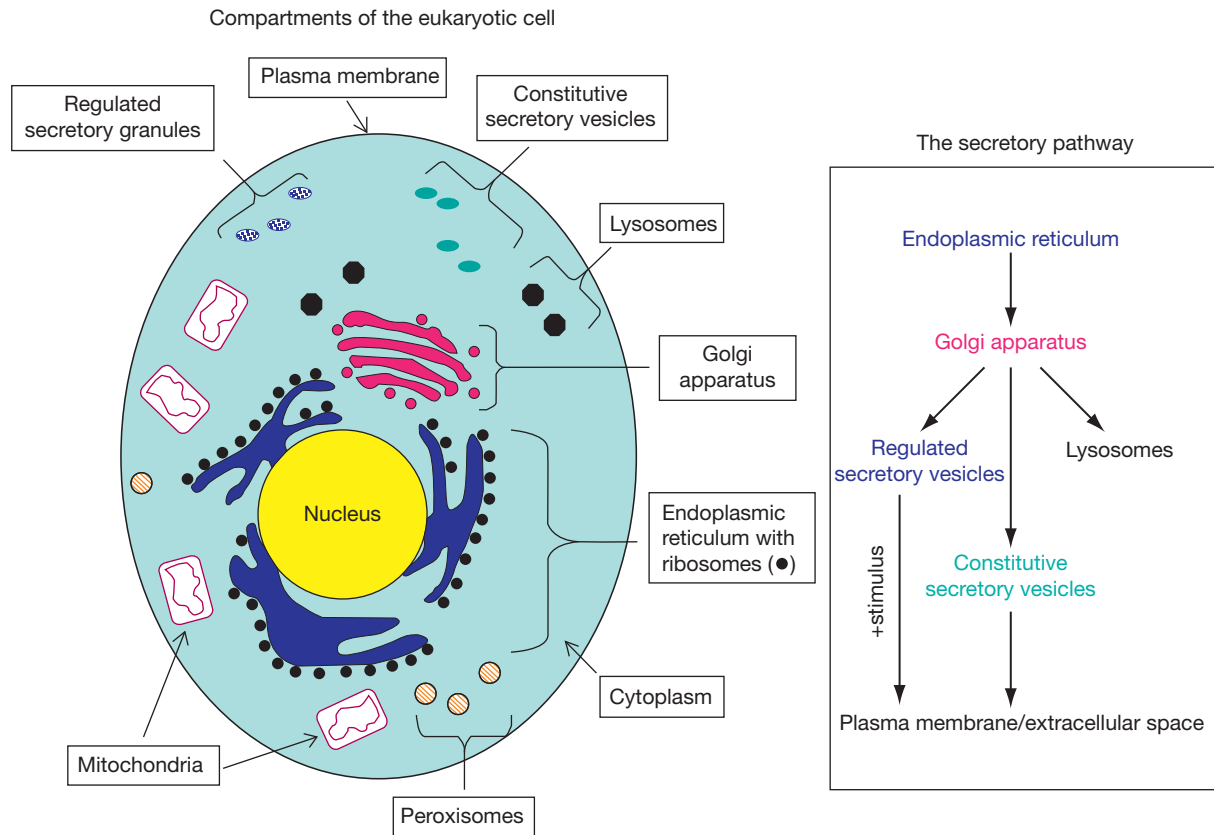


Figure 1 Compartments of mammalian cells and the organization of the secretory pathway. Shown here is diagrammatic representation of the compartments of the mammalian cell. The anterograde flow of membrane and protein traffic in the secretory pathway is shown in the box. Anterograde flow is indicated by arrows. Retrograde flow between the ER and Golgi, endosome/lysosome system and Golgi, and plasma membrane and Golgi does occur, but is not shown.

Soluble and Integral Membrane Proteins

Soluble proteins are completely translocated across the ER membrane into the lumen (Figure 2). These proteins will either remain in the ER, be targeted to another organelle, or be secreted from the cell. Integral membrane proteins that possess one or more hydrophobic membrane-spanning regions will use these sequences to insert into the membrane of the ER and either stay as ER resident transmembrane proteins or be targeted to another cellular membrane. A type I membrane protein has a cleavable signal peptide and a separate hydrophobic stretch of amino acids that acts as a membrane-spanning region. This type of protein has its amino terminus in the lumen of an organelle or the outside of the cell if it reaches the plasma membrane (these are topologically equivalent), and its carboxy terminus in the cytoplasm (Figure 2). By contrast, a type II membrane protein has an uncleavable signal peptide, or signal anchor, that is usually close to the protein's amino terminus and serves a dual function as both signal peptide and a membrane-spanning region. Type II membrane proteins employ a more elaborate insertion mechanism than do type I membrane proteins. For this reason, a type II membrane protein will have its carboxy terminus in the lumen of an organelle or outside the cell if the protein reaches the plasma membrane, and its amino terminus in the cytoplasm (Figure 2).

Other proteins span the membrane several times and are called type III membrane proteins. They can start with either cleavable signal peptides or uncleavable signal anchors and possess variable numbers of hydrophobic membrane-spanning segments.

Protein Folding and Modification in the ER

The Initiation of Protein N-Linked Glycosylation in the ER

As proteins enter the ER lumen, they fold and assemble with the help of chaperone proteins. Many proteins are also co-translationally modified by the addition of carbohydrates to asparagine residues in the process of N-linked glycosylation (Figure 2). A preformed oligosaccharide, consisting of three glucoses, nine mannoses, and two N-acetylglucosamine residues ($\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$), is transferred to accessible asparagine residues in the tripeptide sequence asparagine-X-serine or threonine (X cannot be proline) by the oligosaccharide protein-transferase complex. Subsequent modification by glycosidases (enzymes that remove monosaccharides) and glycosyltransferases (enzymes that add monosaccharides) in the ER and Golgi leads to the remodeling of the N-linked oligosaccharides. These N-linked carbohydrates help proteins fold, protect them from proteolytic degradation, and, in some cases, are critical for

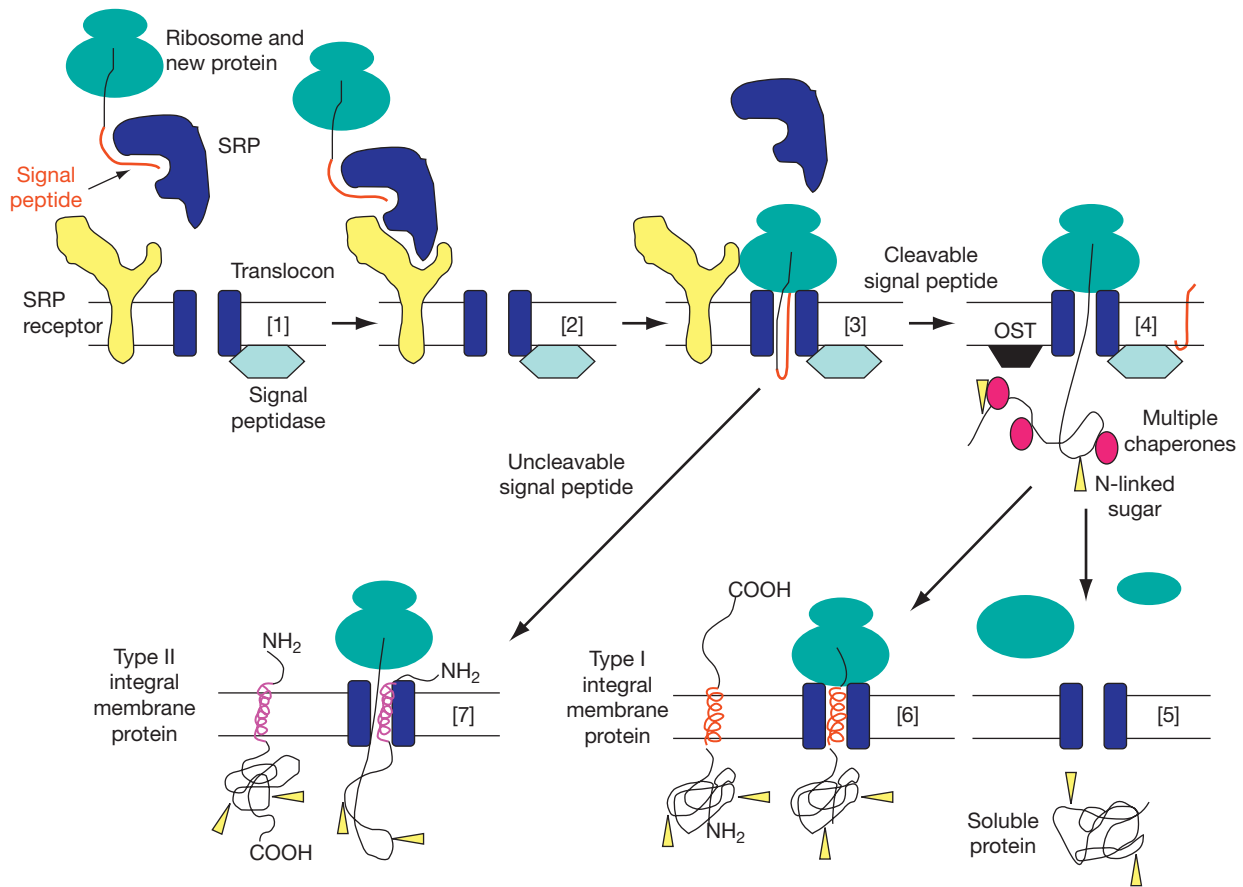


Figure 2 Entry into the secretory pathway. Many proteins are targeted for the secretory pathway by an amino-terminal hydrophobic signal peptide that allows their co-translational translocation across the ER membrane. [1] Signal recognition particle (SRP) recognizes the new protein's signal peptide. [2] The ribosome–new protein–SRP complex interacts with the SRP receptor on the cytoplasmic face of the ER membrane. [3] The ribosome–new protein complex is transferred to the translocon channel, protein synthesis continues, and the protein moves through the aqueous channel. [4] As the new protein enters the lumen of the ER, its signal peptide is cleaved by the signal peptidase, chaperone proteins bind to aid in folding, and oligosaccharide protein transferase complex (OST) transfers oligosaccharides (arrowheads) to asparagine residues in the process of N-linked glycosylation. [5] Soluble proteins lack additional hydrophobic sequences and are translocated through the translocon to complete their folding and modification in the ER lumen. [6] Type I integral membrane proteins have a second hydrophobic sequence that partitions into the lipid bilayer and acts as a membrane-spanning segment. These proteins have their amino termini in the lumen of an organelle or outside the cell when translocated to the plasma membrane and their carboxy-termini in the cell cytoplasm. [7] Unlike proteins with cleavable amino-terminal signal peptides, type II integral membrane proteins have an uncleavable signal anchor that target the protein to the secretory pathway and then partition into the lipid bilayer to act as a membrane spanning segment. These proteins have their amino termini in the cell cytoplasm and their carboxy-termini in the lumen of an organelle or outside the cell. Soluble and integral membrane proteins that enter at the level of the ER need not stay there, and can be transported out of the ER to other locations in the pathway by membrane vesicles (see [Figure 1](#); secretory pathway box).

modulating and mediating protein and cell interactions at the cell surface and in the extracellular space.

Chaperones and the ER Quality Control System

An important function of the ER is to serve as a site of protein folding and quality control. Protein folding in the ER includes the formation of intra-molecular disulfide bonds, prolyl isomerization, and the sequestration of hydrophobic amino acids into the interior of the protein. Protein disulfide bonds are formed as the protein exits the translocon and may at first form incorrectly between cysteine residues close together in the protein's linear amino acid sequence. Thiol-oxidoreductases, such as protein disulfide isomerase (PDI), help to form and reorganize proteins' disulfide bonds into the most energetically

favorable configuration. Different types of chaperones monitor a protein's folding and prevent exit of unfolded and unassembled proteins from the ER. The chaperone BiP, originally identified as an immunoglobulin heavy chain binding protein, interacts with the exposed hydrophobic sequences of folding intermediates of many proteins and prevents their aggregation. Two chaperones called calnexin and calreticulin recognize a monoglucosylated carbohydrate structure ($\text{Glc}_1\text{Man}_9\text{GlcNAc}_2$) that is formed by a special glucosyltransferase that recognizes unfolded or misfolded proteins and adds a single glucose to the $\text{Man}_9\text{GlcNAc}_2$ structure. Proteins that are not folded properly or are not assembled into oligomers with partner subunits are prevented from exiting the ER by chaperone interactions, and can be targeted back across the ER membrane through the translocon into the cytoplasm where they are degraded by

the proteasome complex in a process called ER-associated degradation (ERAD).

Protein Transport Through and Localization in the Secretory Pathway

Vesicular Transport between the ER and Golgi

Proteins move between the ER and Golgi in vesicular carriers. These vesicles are coated with specific sets of cytoplasmic proteins that form the COP-I and COP-II coats. COP-II-coated vesicles move from the ER to the intermediate compartment/*cis*-Golgi, whereas COP-I-coated vesicles move from the Golgi back to the ER and may also mediate transport between Golgi cisternae in both the anterograde (toward the plasma membrane) and retrograde (toward the ER) directions (Figure 3). The process of vesicular transport can be separated into three stages – cargo selection and budding, targeting, and fusion. In the first stage, the COP coats serve to select cargo for exit from a compartment and help to deform the membrane for vesicle budding. They assemble on the membrane with the help of small GTPases called ADP ribosylation factor (ARF; specific for COP-I) and Sar1p (specific for COP-II). After vesicle budding, the hydrolysis of guanosine triphosphate (GTP) by ARF and Sar1p leads to the uncoating of the vesicle. This uncoating reveals other vesicle proteins that are essential for vesicle targeting and fusion. In the second and third stages, tethering

proteins on the transport vesicle and target membrane interact weakly bringing the membranes together. This allows vesicle-associated SNARE proteins (SNAP (soluble NSF attachment protein) receptor, SNARE) and target-membrane-associated SNARE proteins to form complexes. Subsequent conformational changes in the SNARE protein complex bring the membranes together for fusion. Another group of small GTPases called Rabs controls the process of vesicular transport at several levels by recruiting and activating various proteins in the pathway.

Protein Localization in the ER

Proteins involved in protein folding, modification, and quality control must remain in the ER, while proteins destined for the Golgi, lysosome, plasma membrane, or those that are secreted from the cell must exit. Exit from the ER is a selective process that involves cargo receptors that interact with COP-II coat components (Figure 3). It is likely that most resident ER proteins are not selected to exit the ER. It is clear, however, that some resident proteins escape from the ER and are retrieved from the Golgi and intermediate compartment by COP-I vesicles (Figure 3). These proteins have specific amino acid signals that allow their incorporation into COP-I-coated vesicles either by direct interaction with COP-I components or by indirectly interaction with cargo receptors. For example, BiP is a soluble ER chaperone that has the carboxy-terminal four-amino-acid

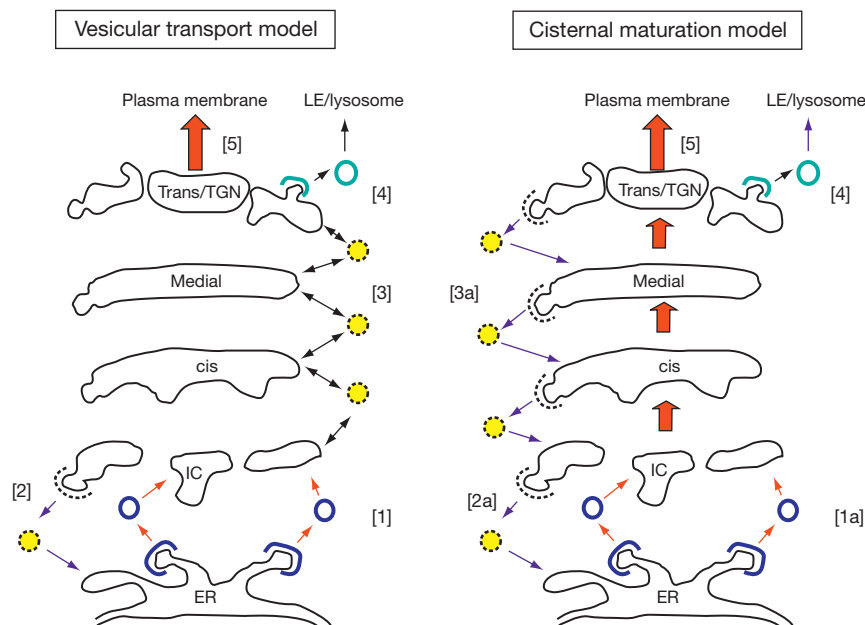


Figure 3 Comparison of two models of protein transport through the Golgi apparatus. In the vesicular transport model, cargo proteins move between the cisternae in vesicles, while Golgi enzymes are retained in their resident cisternae. [1] COP-II-coated vesicles transport new proteins from the ER to the intermediate compartment (IC). [2] Resident ER proteins that escape the ER can be retrieved from the IC in COP-I-coated vesicles. [3] COP-I-coated vesicles also transport anterograde cargo proteins between the Golgi cisternae in both a retrograde and an anterograde fashion (percolating vesicles). In the cisternal maturation model, cargo proteins enter a new cisterna at the *cis* face of the stack and are modified (matured) by resident Golgi enzymes that are continuously transported in a retrograde fashion into the sequentially maturing cisternae. [1a] COP-II-coated vesicles transport new proteins from the ER to the IC where a new *cis* cisterna forms. [2a] Resident ER proteins that escape the ER can be retrieved from the IC in COP-I-coated vesicles. [3a] Golgi enzymes are transported in a retrograde fashion in COP-I-coated vesicles to modify the cargo proteins in earlier cisternae. Mechanisms of protein exit from the TGN are common to both models: [4] clathrin-coated vesicles mediate late endosome (LE)–lysosome transport, while [5] other proteins are secreted in either a regulated or constitutive fashion to the plasma membrane or extracellular space.

sequence lysine–aspartate–glutamate–leucine (KDEL). This KDEL sequence is recognized by the KDEL receptor that mediates incorporation into COP-I vesicles moving from the intermediate compartment back to the ER. The lysine–lysine–X–X sequence (KKXX) is present at the carboxy terminus of some ER integral membrane proteins and it also mediates the retrieval of these proteins via COP-I vesicles.

ER Localization Signals and Protein Quality Control

In addition to carboxy terminal KDEL and KKXX sequences, an arginine–X–arginine signal called the RXR motif mediates the ER retrieval of integral membrane proteins. The RXR motif is often present in the cytoplasmic regions of multimeric proteins that are expressed on the plasma membrane. The RXR motif is thought to be masked during biogenesis and trafficking by various mechanisms such as folding, subunit assembly, protein–protein interactions, and posttranslational modifications that are all required for proteins to function properly at the plasma membrane. Defects in any of these processes can result in the exposure of the RXR motif, which leads to the recognition by COP-I and subsequent retrograde transport to the ER. Thus, the RXR motif serves as a check-point at the ER that ensures the export of proteins with appropriate structures and functions.

Protein Modification in the Golgi

The Golgi apparatus consists of stacks of flattened cisternae that contain enzymes and other proteins involved in the further modification and processing of newly made proteins. It is separated into *cis*- *medial*- and *trans*-cisternae, followed by a meshwork of tubules and vesicles called the *trans*-Golgi network (TGN). The process of N-linked glycosylation is completed through the action of glycosidases and glycosyltransferases localized in specific cisternae. Likewise, the glycosylation of serine and threonine residues (O-linked glycosylation) is accomplished by other glycosyltransferases. Additional modifications also occur in the Golgi. For example, proteins and carbohydrates are sulfated by sulfotransferases and some proteins are phosphorylated on serine and threonine residues by Golgi kinases. In addition, proteins such as digestive enzymes (trypsin, carboxypeptidase) and hormones (insulin) are made as inactive precursors that must be proteolytically processed to their active forms in the late Golgi or post-Golgi compartments.

Transport of Proteins through the Golgi

Currently, there are two different models to explain protein transport through the Golgi (Figure 3). The vesicular transport model proposes that proteins move sequentially between the Golgi cisterna in COP-I-coated vesicles, while the cisternae themselves are stationary. Proteins not retained in the *cis*-Golgi, for example, would be incorporated into coated vesicles and be transported to the *medial*-Golgi, and then to the *trans*-Golgi. Proteins destined for post-Golgi compartments move through successive Golgi cisternae in this fashion, being modified by the resident enzymes in each compartment (Figure 3). In the cisternal maturation or progression model, a new cisterna is formed on *cis* face of the Golgi stack from ER-derived

membrane. This requires both the anterograde transport of newly synthesized proteins from the ER in COP-II-coated vesicles and the retrograde transport of *cis*-Golgi enzymes from the pre-existing *cis* cisterna in COP-I coated vesicles. The new *cis*-cisterna and its contents progressively mature through the stack as resident Golgi enzymes are successively introduced by COP-I-coated vesicles (Figure 3). In the vesicular transport model, the resident Golgi enzymes are retained in the cisternae while the cargo moves in vesicles between the different cisternae. By contrast, in the cisternal maturation/progression model, the resident enzymes are continuously moving in a retrograde fashion, while the anterograde cargo remains in the cisternae. In a budding yeast *Saccharomyces cerevisiae*, recent imaging studies have provided strong evidence for the maturation of cisternae and the cisternal maturation/progression model. Nevertheless, evidence for both mechanisms is compelling, suggesting that they may work in parallel.

Localization of Resident Golgi Enzymes

In the context of the vesicular transport model, Golgi enzymes are retained in specific cisternae. Two mechanisms have been suggested for Golgi protein retention. The bilayer thickness mechanism suggests that the relatively short transmembrane regions of Golgi proteins prevent their incorporation into the wider, cholesterol-rich lipid bilayers of the transport vesicles destined for post-Golgi compartments (such as the plasma membrane), and, as a result, these proteins are retained in the relatively cholesterol-poor Golgi membranes. The oligomerization mechanism predicts that once an enzyme has reached its resident cisterna, it forms homo- or hetero-oligomers that prevent its incorporation into transport vesicles moving to the next compartment. In the context of the cisternal maturation/progression model, resident Golgi enzymes are actively incorporated into COP-I vesicles for retrograde transport to a new cisterna, and one might predict that the cytoplasmic tails of these proteins would interact with COP-I components to allow vesicle incorporation. Interestingly, while the membrane-spanning regions of Golgi resident proteins have been shown to be critical for their localization, evidence continues to accumulate that demonstrates a role for the cytoplasmic sequences of several Golgi proteins in their localization. It is likely that some or all of these mechanisms work together to maintain the steady-state localization of resident Golgi proteins.

Protein Exit from the Golgi and Targeting to Post-Golgi Locations

Protein Exit from the Golgi

Once proteins reach the TGN, they are sorted to post-Golgi compartments that include the lysosome, the plasma membrane, and the extracellular space (Figure 3). Trafficking to the lysosome–endosome system involves clathrin-coated vesicles similar to those that function in the uptake of proteins in endocytosis. By contrast, transport to the plasma membrane, or exocytosis, can occur either constitutively in secretory vesicles/tubules or in a regulated fashion from secretory granules found in specific cell types.

Protein Targeting to the Lysosome

The lysosome is a degradative compartment that contains numerous acid hydrolases that function to digest proteins, lipids, and carbohydrates. The trafficking of the majority of lysosomal enzymes to the lysosome requires mannose 6-phosphate residues on these enzymes' N-linked sugars. The mannose 6-phosphate residues are recognized by receptors in the TGN that mediate the incorporation of the new lysosomal enzymes into clathrin-coated vesicles destined for the late endosome compartment (Figure 3). These clathrin-coated vesicles move from the TGN and fuse with the late endosome, where a decrease in luminal pH causes the lysosomal enzymes to dissociate from the receptors. The enzymes are then transported to the lysosome, while the receptors recycle to the TGN. Some lysosomal membrane proteins that lack the mannose 6-phosphate signal are also trafficked in clathrin-coated vesicles to the lysosome, while others are transported to the cell surface, incorporated into a different set of clathrin-coated vesicles used in the process of endocytosis, and then trafficked to the lysosome via the late endosome.

Constitutive and Regulated Secretion

In many cell types, membrane-associated and soluble proteins move to the plasma membrane constitutively without a requirement for specific signals. Constitutively secreted proteins include receptors, channel proteins, cell adhesion molecules, and soluble extracellular matrix and serum proteins. Other proteins such as hormones and neurotransmitters are targeted to secretory vesicles that are involved in regulated secretion from endocrine and exocrine cells, some types of immune cells, and

neurons. These vesicles remain in a secretion-ready state until extracellular signals that lead to an increase in intracellular calcium levels trigger the exocytosis of their contents.

Microtubule and Motor Proteins in the Secretory Pathway

Vesicular trafficking between membrane compartments requires cytoskeletal filaments and motor proteins. Microtubules emerge from a structure called the microtubule-organizing center (MTOC) where the minus-ends of microtubules are anchored. Motor proteins are a class of molecular motors that are able to transport membrane vesicles by moving along the surface of the cytoskeleton. They are powered by the hydrolysis of ATP and convert chemical energy into mechanical work. For example, several types of kinesin motors transport vesicular cargo toward the microtubule plus-end, whereas cytoplasmic dynein–dynactin transports vesicles toward the minus-end. In addition, certain classes of unconventional myosins convey cargo along actin filaments (Figure 4). For example, the vesicular transport from ER to Golgi along microtubules is dependent on the dynein–dynactin complex, while trafficking in the retrograde pathway from Golgi to ER, as in the recycling of ER-resident proteins, is likely to be kinesin dependent. Kinesins also participate in the trafficking pathways that diverge from the TGN, including transport to late endosomes and sorting to constitutive and regulated secretory vesicles. Recruitment of motor proteins to membrane vesicles is facilitated by various molecules including coat proteins, modular scaffold proteins, transmembrane proteins, small GTPases, and other motor proteins.

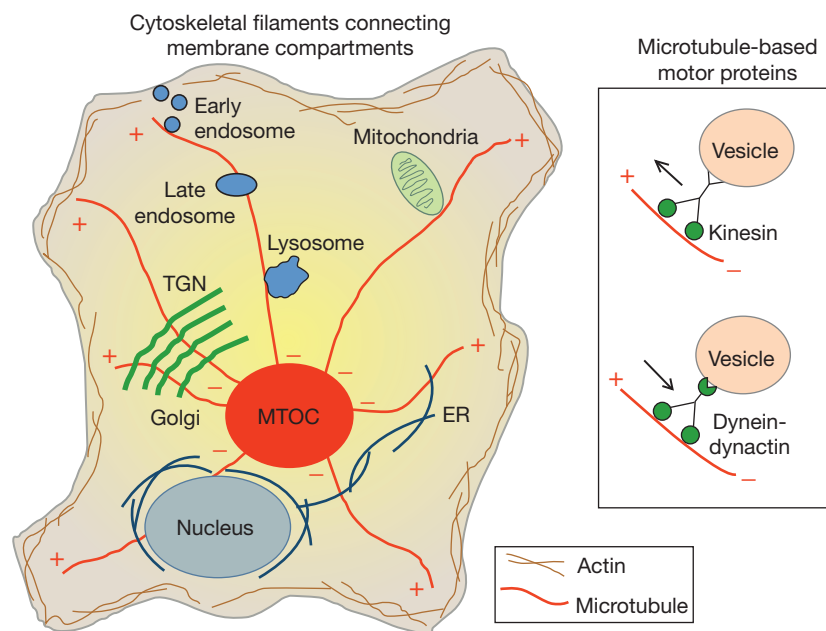


Figure 4 Cytoskeletal filaments connecting the membrane compartments in the secretory pathway. Motor proteins utilizing the cytoskeleton for movement fall into two categories based on their substrates. Some motors such as myosin move along microfilaments through interaction with actin. Other motors such as dynein and kinesin move along microtubules through interaction with tubulin. There are two basic types of microtubule motors: plus-end motors and minus-end motors that walk in different directions along the microtubule cables within the cell. Kinesins typically walk toward the plus end of the microtubule, while the dynein–dynactin complex transports cargo by walking toward the minus end.

See also: **Cell Architecture and Function:** Actin Organization; Dynactin; Dynein; Endoplasmic Reticulum-Associated Protein Degradation; Golgi Complex; Kinesin Superfamily Proteins; Myosin Motors; 26S Proteasome: Structure and Function; **Lipids Carbohydrates Membranes and Membrane Proteins:** Glycoprotein Folding and Processing Reactions; Glycoproteins, N-Linked; N-Linked Glycan-Processing Glucosidases and Mannosidases; **Protein/Enzyme Structure Function and Degradation:** Disulfide Bond Formation; Proteasomes, Overview; Protein Degradation; Protein Folding and Assembly; **Signaling:** Rab Family; Small GTPases.

Further Reading

Alberts B, Johnson A, Lewis J, et al. (eds.) (2007) Intracellular compartments and protein sorting (chapter 12). Intracellular vesicular traffic (chapter 13). In: *Molecular Biology of the Cell*, 5th edn, pp. 695–812. New York, NY: Garland Science.

Anelli T and Sitia R (2008) Protein quality control in the early secretory pathway. *EMBO Journal* 23: 315–327.

Beck R, Rawet M, Wieland FT, and Cassel D (2009) The COPI system: Molecular mechanisms and function. *FEBS Letters* 583: 2701–2709.

Braulke T and Bonifacio JS (2009) Sorting of lysosomal proteins. *Biochimica et Biophysica Acta* 1793: 605–614.

Caviston JP and Holzbaur EL (2009) Microtubule motors at the intersection of trafficking and transport. *Trends in Cell Biology* 16: 530–537.

Dancourt J and Barlowe C (2010) Protein sorting receptors in the early secretory pathway. *Annual Review of Biochemistry* 79: 315–327.

Ellgaard L and Helenius A (2003) Quality control in the endoplasmic reticulum. *Nature Reviews Molecular Cell Biology* 4: 181–191.

Farquhar MG and Palade GE (1998) The Golgi apparatus: 100 years of progress and controversy. *Trends in Cell Biology* 8: 2–10.

Gerdes HH (2008) Membrane traffic in the secretory pathway. *Cellular and Molecular Life Sciences* 65: 2779–2780.

Glick BS and Nakano A (2009) Membrane traffic within the Golgi apparatus. *Annual Review of Cell and Developmental Biology* 25: 113–132.

Lodish H, Berk A, Matsudaira P, et al. (eds.) (2004) *Molecular Cell Biology*, 5th edn., pp. 657–699; 701–742. New York, NY: W. H. Freeman.

Mancias JD and Goldberg J (2005) Exiting the endoplasmic reticulum. *Traffic* 6: 278–285.

Palade G (1975) Intracellular aspects of the process of protein synthesis. *Science* 189: 347–358.

Selenoprotein Synthesis

A Böck, University of Munich, Munich, Germany

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 4, pp. 17–21, © 2004, Elsevier Inc.

Glossary

Elongation factor Helper protein assisting the ribosome in the polypeptide elongation process.

Nonstandard amino acid Amino acid whose insertion is achieved by an expansion of the classical genetic code.

SECIS Selenocysteine insertion sequence of the mRNA which redefines a UGA stop codon, in a sense, codon for the insertion of selenocysteine.

Selenoprotein Protein with one or more selenocysteine residues.

Stop codon A codon signaling chain termination in protein synthesis in the classical genetic code UGA, UAA, or UAG.

tRNA RNA molecule carrying an amino acid at its 3'-end and functioning as an adaptor to incorporate the amino acid into the growing polypeptide chain according to the triplet sequence of the mRNA.

Selenoproteins contain one or more residues of the nonstandard amino acid selenocysteine, which is an analog of cysteine in which a selenol group replaces a thiol. The majority of these proteins catalyzes some oxidation/reduction function in which the selenol of the selenocysteine that is present in the active site of the respective enzyme takes part in the reaction. The advantage of having a selenol instead of a thiol lies in the fact that it confers to these enzymes a higher kinetic efficiency. In some biological systems, selenoproteins may also fulfill a structural role because of their capacity to oligomerize proteins via the formation of diselenide or mixed disulfide–selenide bridges. The biosynthesis of selenoproteins is unique since the incorporation of selenocysteine occurs co-translationally by the ribosome and not posttranslationally. Selenocysteine insertion is DNA encoded, requires the function of a cognate transfer RNA (tRNA) and of a specific translation elongation factor different from elongation factor Tu. Selenocysteine, therefore, has been designated as the 21st amino acid.

Bacterial Selenoprotein Synthesis

The structure and the function of the components involved in selenocysteine biosynthesis have been characterized to a considerable extent in the case of the bacterial system. The process can be divided into three functional steps: (1) the biosynthesis of selenocysteine in the tRNA-bound state; (2) the formation of a complex between elongation factor SelB, guanosine triphosphate (GTP), selenocysteyl-tRNA^{Sec} and the messenger RNA (mRNA); and (3) the decoding event at the ribosome. As far as it is known, though there are some major differences, similarities also exist between bacterial selenoprotein synthesis and the process characteristic of eukarya and archaea.

Selenocysteine Biosynthesis

Figure 1 summarizes the process of selenocysteine biosynthesis as it has been worked out for *Escherichia coli*. It requires the activities of three enzymes, namely, seryl-tRNA synthetase (SerS), selenophosphate synthetase (SelD), and selenocysteine synthase (SelA) plus the specific tRNA (tRNA^{Sec}). Seryl-tRNA synthetase

charges tRNA^{Sec} with L-serine, selenocysteine synthase converts the seryl-tRNA^{Sec} into selenocysteyl-tRNA^{Sec} using selenophosphate as a source for activated selenium. Selenophosphate is provided by selenophosphate synthetase from selenide in an adenosine triphosphate (ATP)-dependent reaction. The genes for these components had been identified with the aid of *E. coli* mutants isolated by Mandrand–Berthelot as being pleiotropically deficient in formate dehydrogenase activities.

tRNA^{Sec}

tRNA^{Sec} (**Figure 2**) is the key molecule of selenoprotein synthesis since it serves both as the adaptor for selenocysteine biosynthesis and for incorporation of the amino acid at the ribosome. It deviates in size, secondary structure, and in normally conserved sequence positions from canonical elongator tRNA species. Due to the elongated extra arm and the one base-pair-extended aminoacyl acceptor arm, tRNA^{Sec} species are the largest members of the elongator tRNA family. All tRNA^{Sec} species identified thus far possess a UCA anticodon, which enables them to pair with UGA stop codons (but only if these are in a special mRNA sequence context). Moreover, tRNA^{Sec} species deviate from canonical elongator tRNA species in sequence positions which are usually invariant and which are involved in the establishment of novel tertiary interactions within the molecule.

On the basis of its serine identity elements, tRNA^{Sec} is charged by the cellular seryl-tRNA synthetase, which also aminoacylates serine-inserting isoacceptors. However, both the affinity and the rate of aminoacylation are diminished in comparison to the charging of tRNA^{Ser}, resulting in an overall 100-fold reduced efficiency.

Selenocysteine synthase

The overall reaction catalyzed by selenocysteine synthase consists in the exchange of the hydroxyl group of the serine moiety of seryl-tRNA^{Sec} by a selenol group (**Figure 1**). The reaction occurs in two steps: first, the amino group of serine forms a Schiff base with the carbonyl of the pyridoxal 5'-phosphate cofactor of selenocysteine synthase leading to the 2,3-elimination of a water molecule and the formation of dehydroalanyl-tRNA^{Sec} and second, nucleophilic addition of reduced selenium to the double

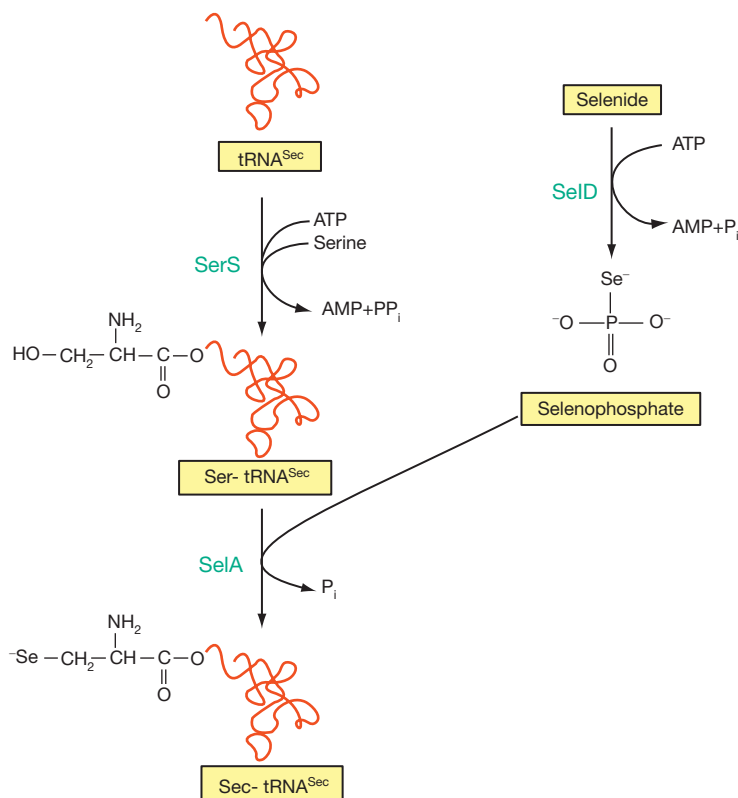


Figure 1 Path of selenocysteine biosynthesis. For explanation, see text.

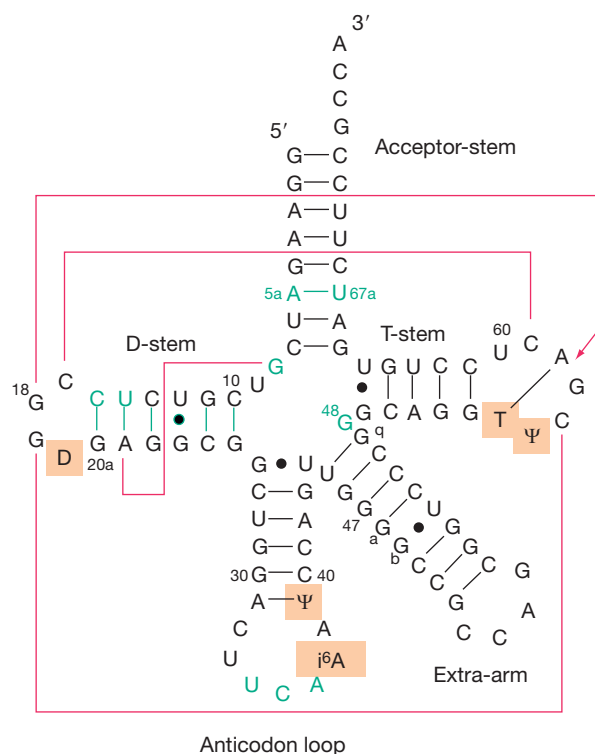


Figure 2 Cloverleaf presentation of the structure of tRNA^{Sec} from *E. coli*. Modified bases are shaded. Tertiary interactions via base pairing are indicated by connecting red lines, and those involving intercalation are denoted by arrows. Bases and base pairings deviating from the consensus are indicated in green.

bond of dehydroalanyl-tRNA^{Sec} from selenophosphate as a donor yields selenocysteyl-tRNA^{Sec}.

Selenocysteine synthase from *E. coli* is a homodecameric enzyme and low-resolution electron microscopy revealed that it is made up of two pentameric rings stacked on top of each other. Each of the two subunits is able to bind one molecule of seryl-tRNA^{Sec}, so the fully loaded enzyme can complex five charged tRNA molecules. As serine isoacceptor tRNAs are not recognized, the tRNA must have determinants for the specific recognition of seryl-tRNA^{Sec} by selenocysteine synthase and antideterminants for the rejection of seryl-tRNA^{Ser} species. The specificity for discrimination of the selenium donor is not as strict since the purified enzyme accepts thiophosphate instead of selenophosphate as a substrate. This results in the formation of cysteyl-tRNA^{Sec}. Therefore, the discrimination between sulfur and selenium must take place at some other step of selenocysteine biosynthesis.

Selenophosphate synthetase

Purified selenocysteine synthase does not exhibit an absolute requirement for selenophosphate as a substrate to convert seryl-tRNA^{Sec} into selenocysteyl-tRNA^{Sec}, since the reaction also occurs in the presence of high concentrations of selenide. Even sulfide is accepted although at a very low efficiency. Therefore, selenophosphate as a substrate may be necessary in one or more of the following three aspects, that is, (1) to discriminate sulfide from selenide, (2) to efficiently use low concentrations of selenium compounds, and (3) to accelerate the reaction rate effected by the activation of the trace element. Indeed, selenophosphate synthetase efficiently discriminates between sulfide and selenide, and thus excludes sulfur from intrusion into the

selenium pathway. Selenophosphate synthetase from *E. coli* is a monomeric enzyme with a unique reaction mechanism since it formally transfers the γ -phosphate of ATP to selenide with the intermediate formation of enzyme-bound adenosine diphosphate (ADP), which is subsequently hydrolyzed into adenosine monophosphate (AMP) and inorganic phosphate.

Formation of the SelB-GTP-selenocysteyl-tRNA-SECIS complex

Elongation factor Tu, which forms a complex with all 20 standard aminoacyl-tRNAs and donates them to the ribosomal A-site, displays only minimal binding affinity for selenocysteyl-tRNA^{Sec}. Consequently, insertion of selenocysteine requires the function of an alternate elongation factor, which is SelB. SelB from *E. coli* is a 70-kDa protein, which contains the sequence elements of elongation factor Tu in the N-terminal two-thirds of the molecule (designated domains I, II, and III) plus a domain IV of about 25 kDa, which can be subdivided into domains IVa and IVb. Domains I, II, and III share their functions with those of elongation factor Tu, namely the binding of guanosine nucleotides and of charged tRNA. An important difference, however, is that they can discriminate between the serylated and the selenocysteylated forms of tRNA^{Sec}. In this way, the insertion of serine instead of selenocysteine, which would lead to an inactive enzyme, is prevented. The structural basis for this discrimination ability has not yet been resolved. A second difference is that the overall affinity for GTP is about 10-fold higher than that for guanosine diphosphate (GDP), which obviates the need for the function of a guanosine nucleotide release factor, since GDP is chemically replaced by GTP. In accordance, the structure of SelB lacks those subdomains, which in elongation factor Tu are responsible for interaction with the GDP release factor EF-Ts. The 25 kDa C-terminal extension of SelB (domain IV) is required for the function in selenoprotein synthesis since its truncation inactivates the molecule. The reason is that subdomain IVb binds to a secondary structure of the mRNA (the SECIS element) coding for selenoprotein synthesis. SelB, thus, is able to form a quaternary complex with GTP and two RNA ligands, namely selenocysteyl-tRNA and the SECIS element (Figure 3). The isolated domains IV or IVb retain the binding capacity for the SECIS element. The formation of the quaternary complex follows random-order kinetics. Another important feature is that

the stability of the complex is increased when both RNA ligands are bound.

The SECIS element is a hairpin structure formed within a section of 39 bases of the selenoprotein mRNA which follows the codon specifying selenocysteine insertion at the 3'-side. The binding of SelB takes place at its apical stem-loop mini-helix of 17 nucleotides. Genetic and structural analysis showed that bases in the loop region plus a bulged-out U in the helix are required for the interaction with SelB. This apical part of the SECIS element is separated by a short unpaired region from a helix at the base of the hairpin. Pairing within this second helix is not essential but it increases the efficiency of selenocysteine insertion. An absolute requirement, however, is that the codon determining selenocysteine insertion lies within a critical distance relative to the binding site of SelB.

Bacterial SECIS elements lie within the reading frame of their selenoprotein mRNAs; they are thus subject to stringent sequence constraints in order to deliver a functional gene product. However, they do not need to be translated since they also function when placed in the 3'-untranslated region at the correct distance to the selenocysteine codon within an upstream reading frame. The sequence constraint (which depends on the protein to be formed) and the requirement for binding of SelB restricts the number of selenoprotein mRNAs to be expressed in a single organism and explains why the vast majority of selenoprotein genes cannot be heterologously expressed unless the cognate SelB gene is coexpressed. Thus, SelB and their SECIS elements are subject to coevolution.

Decoding Event at the Ribosome

In all biological systems analyzed thus far, selenocysteine insertion is directed by the opal stop codon UGA, but only if it is followed by an SECIS element at the correct distance. This violates the dogma that no codon can have more than one meaning within a single cell. The questions to be answered therefore are: (1) what prevents the UGA to be used as a termination signal; and (2) which mechanism interferes with insertion of selenocysteine at ordinary UGA stop codons?

Counteraction of stop at the UGA?

A convincing answer to the question why the selenocysteine-specific UGA codon does not function as an efficient termination signal must await structural information on the decoding complex. It is, however, clear that termination always competes with selenocysteine insertion, especially under conditions when the capacity for decoding the UGA with selenocysteine is a limiting factor. This can be, for example, a surplus of selenoprotein mRNA in relation to the amount of SelB quaternary complex, which forces the ribosome to stall at the UGA. One fact identified to be involved in the suppression of termination is that the base following the UGA at the 3'-side in selenoprotein mRNAs is predominately an A or C, which renders the UGA a weak termination signal. Moreover, the two amino acids preceding selenocysteine in the nascent polypeptide chain are predominantly hydrophobic, which counteracts the dissociation of the nascent polypeptide from the ribosome when translation pauses at a hungry codon present in the A site. However, additional mechanisms must exist, which contribute to the suppression of termination.

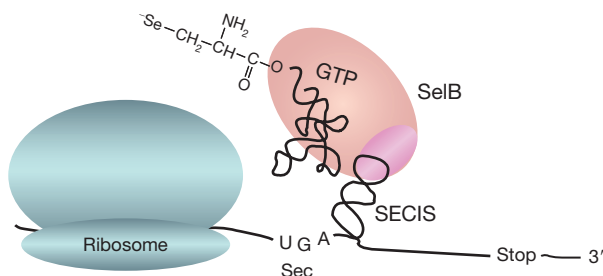


Figure 3 Translation of prokaryotic selenoprotein mRNA. Note that the SECIS element is within the mRNA reading frame and is complexed to domain IVb of SelB carrying selenocysteyl-tRNA^{Sec} and GTP.

Selenocysteine specificity of UGA codons

From the collinearity between the mRNA nucleotide sequence and the amino acid sequence of the translation product, it is clear that UGA determines the position where selenocysteine is to be inserted during translation. The specificity of the UGA, however, is determined by the codon context, that is, by the existence of a SECIS element at the 3'-side. The results of extensive biochemical and biophysical analysis suggest the following scenario for the decoding process: (1) SelB forms the quaternary complex at the mRNA in which the two RNA ligands display cooperativity in their interaction with the protein; (2) in this quaternary complex, SelB attains a conformation compatible for interaction with the ribosome, which then results in stimulation of GTP hydrolysis by SelB, which in turn causes the release of the charged tRNA in the vicinity of the ribosomal A-site; and (3) loss of the tRNA ligand causes the SelB protein to return to a conformation with affinity about 10-fold lower for the SECIS element. As a consequence, the mRNA is released from the protein and freed for the translation of codons downstream of the UGA. The consequence of the complex cascade of reactions is that the efficiency of the decoding of UGA with selenocysteine is lower than that of any of the standard sense codons. It is also reflected by a considerable pause occurring when the ribosome encounters the quaternary complex at the mRNA. In the absence of selenocysteyl-tRNA, binding of SelB alone to the mRNA does not retard the rate of translation.

Archaeal and Eukaryal Selenoprotein Synthesis

tRNA^{Sec} species from archaea and eukarya share several structural similarities with the bacterial counterparts, but they are more related to each other than either one is to bacterial tRNA^{Sec}. There is also considerable sequence similarity among selenophosphate synthetases from all three lines of descent rendering their annotation in genome projects easy. On the other hand, homologs for the bacterial selenocysteine synthase have not been identified yet in any of the genomic sequences from organisms known to synthesize selenoproteins.

Whereas UGA directs selenocysteine insertion in both archaea and eukarya, a fundamental difference is that the SECIS element is not positioned within the reading frame but in the 3'-nontranslated region of the mRNA. SECIS elements from organisms of the three lines of descent differ by sequence and by secondary structure. They may be positioned at different distances from the actual termination codon and/or the selenocysteine-inserting UGA codon, but a critical distance must not be underpassed. It is thought that the selective value for having the SECIS element in the nontranslated region consists in liberating it from the sequence constraint and, thus, allowing the translation of mRNAs with more than one UGA codon specifying selenocysteine insertion. Indeed, proteins

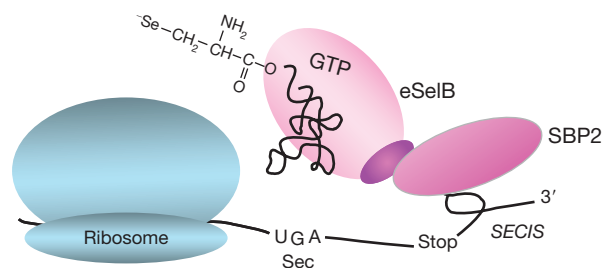


Figure 4 Translation of eukaryal selenoprotein mRNA. Note that the SECIS element is in the 3'-nontranslated region and serves as the binding site for SBP2, which in turn interacts with eukaryal SelB protein.

with up to 17 selenocysteine residues are formed in some eukaryotes and, in one instance, a polypeptide with two such amino acids is synthesized in an archaeon.

Parallel to this deviation in sequence, structure, and position of the SECIS element, there is an alteration of the structure of the archaeal and eukaryal SelB-like translation factors. Domains I, II, and III closely resemble the three homologous domains from the bacterial SelB. However, the C-terminal extension is short, less than 10 kDa, and accordingly, as expected, the archaeal and eukaryal SelB homologs do not bind to their cognate SECIS structures. In eukarya, a second protein fulfills this task, namely SBP2 (SECIS-binding protein 2) (Figure 4). There is evidence that SBP2 interacts with the SelB protein by direct contact in the decoding process. This interaction is stabilized in the presence of selenocysteyl-tRNA^{Sec}. However, the precise function of SBP2 has not yet been resolved.

See also: **Cell Architecture and Function:** Rho GTPases and Actin Cytoskeleton Dynamics; **Molecular Biology:** DNA Mismatch Repair in Disease and Aging.

Further Reading

- Atkins JF, Böck A, Matsufuji S, and Gesteland RF (1999) Dynamics of the genetic code. In: Gesteland RF, Cech TR, and Atkins JF (eds.) *The RNA World*, 2nd edn., pp. 637–673. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Copeland PR, Fletcher JE, Carlson BA, Hattfield DI, and Driscoll DM (2000) A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. *EMBO Journal* 19: 306–314.
- Flohe L, Andreessen JR, Brigelius-Flohe B, Maiorino M, and Ursini F (2000) Selenium, the element of the moon, in life on earth. *Life* 49: 411–420.
- Hattfield DI (ed.) (2001) *Selenium: Its Molecular Biology and Role in Human Health*. New York, NY: Kluwer.
- Krol A (2002) Evolutionarily different RNA motifs and RNA–protein complexes to achieve selenoprotein synthesis. *Biochimie* 84: 765–774.
- Rother M, Resch A, Wilting R, and Böck A (2001) Selenoprotein synthesis in archaea. *BioFactors* 14: 75–83.
- Stadtman TC (1996) Selenocysteine. *Annual Review of Biochemistry* 65: 83–100.

Septins and Cytokinesis

A S Maddox, Université de Montréal, Montréal, QC, Canada

C M Field, Harvard Medical School, Boston, MA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Makoto Kinoshita and Christine M. Field, volume 4, pp. 22–26, © 2004, Elsevier Inc.

Glossary

Cell cortex A specialized layer of cytoplasm beneath the plasma membrane. In animal cells, it contains actin and actin-binding proteins.

Cell cycle The sequence of events by which a cell duplicates its contents and divides into two. There is a network of regulatory proteins that governs the progression through the key events such as DNA replication, formation of the mitotic spindle, and segregation of the chromosome. The cell cycle ends with cytokinesis.

Exocyst complex Complex of conserved proteins, which is an essential part of the exocytotic apparatus.

FH proteins Formin homology proteins are conserved proteins required to assemble some types of actin structures such as actin cables (in yeast), stress fibers, and the contractile ring.

Guanosine di-/tri-phosphate (GDP/GTP) Plays an important role in tubulin stability and microtubule assembly and in signal transduction pathways via small GTPases.

Inositol phospholipids Membrane lipids containing inositol and phosphate(s) that are important in various cell-signaling pathways.

Midbody The thin intercellular bridge of cytoplasm connecting two daughter cells in late cytokinesis.

It contains a tightly packed antiparallel array of microtubules and an electron-dense matrix at its center.

Photobleach The exposure of a fluorescent probe to light such that it is rendered nonfluorescent or 'bleached'. Examining the recovery of fluorescence after photobleaching a tagged protein gives an indication of how dynamic its localization at the bleached structure is.

Polarization fluorescence microscopy Fluorescence imaging of a fluorophore that is rigidly affixed to the protein of interest. Sample is illuminated with polarized light and emission is collected after selection of light with the same axis of polarization as the excitation.

Septation In yeast, the formation of a new cell wall or septum to separate two daughter cells. Septation is separable from cytokinesis.

Small GTPase GTP-binding and hydrolyzing switch proteins that alternate between an active/on state when GTP-bound and an inactive, GDP-bound state.

Cytokinesis

Cytokinesis is the process by which one cell physically separates into two daughter cells at the end of each division cycle. In metazoan cells, the division plane, termed the cell equator, is determined by microtubules of the anaphase spindle. Thus, cytokinesis is temporally and spatially coupled to chromosome segregation, ensuring that each daughter cell receives the correct genomic content. Signals from the anaphase spindle elicit formation of a contractile ring made of actin, myosin-II, and other associated proteins, including septins under the plasma membrane at the cell equator. The cleavage furrow ingresses by a combination of ring contraction driven by myosin-II, and targeted insertion of vesicles near the furrow to supply new plasma membrane. Finally, cytokinesis is completed in a poorly understood process called abscission that involves the disassembly of the contractile ring, compaction of the microtubule midzone, and plasma membrane sealing by targeted membrane deposition and fusion (**Figure 1**).

Septins localize to the contractile ring in all organisms that have been studied. Deletion, mutation, or inhibition of septins typically results in incomplete or abortive cytokinesis, though the severity of the defect varies among organisms. This suggests a conserved function of septins in cytokinesis that is not

required for cleavage plane specification, but is required for optimal furrow ingression and/or completion.

Biochemical and Structural Properties of the Septins

Septins Bind Guanine Nucleotides and form Complexes and Filaments

Sequence analysis shows that all septins have a central globular domain containing conserved motifs found in small guanosine triphosphatases (GTPases), and most septins have a C-terminal predicted coiled-coil region of variable length. On purification, septins are found in large complexes containing multiple septin polypeptides. The septin complex purified from *Drosophila* embryos is composed of three septin polypeptides with a stoichiometry of 2:2:2. Yeast complexes contain a fourth septin subunit, and complexes from mammalian brain are heterogeneous, and may be built from at least six different septin polypeptides. When viewed by negative-stain electron microscopy (EM), a typical septin preparation appears as filaments 7–9-nm thick and of variable lengths. The shortest filament represents the complex itself and is the nonpolar building block from which the longer filaments are formed (**Figures 2(a)** and **2(b)** show a purified yeast complex).

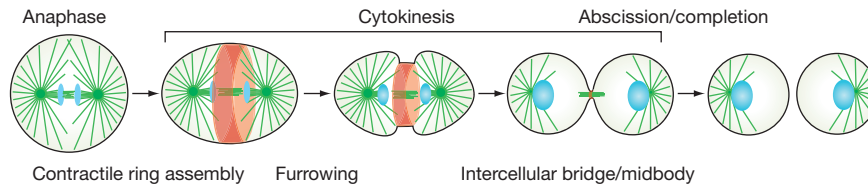


Figure 1 Schematic illustrating the different subprocesses of cytokinesis. DNA is shown in blue, microtubules (MTs) in green, and the contractile ring (CR) in red. The CR assembles and contracts, microtubules become bundled and compacted into the midbody. Cytokinesis is completed by disassembly of the CR and MT structures. Membrane fusion in the cell–cell bridge allows final cell separation.

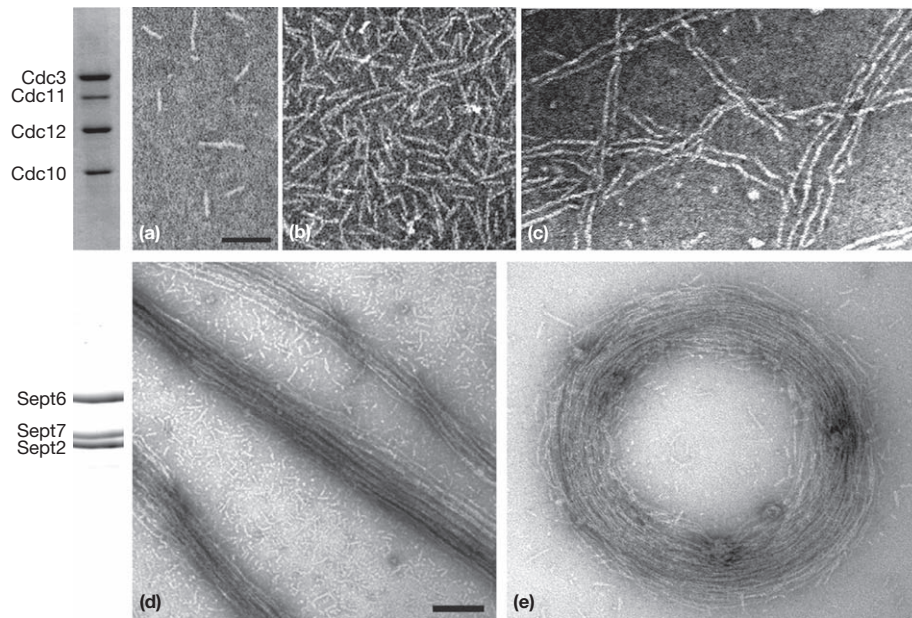


Figure 2 Negative-stain EM of septin structures. (a)–(c) Examples of filamentous structures formed by a four polypeptide septin complex purified from *S. cerevisiae*. (a) Monomers and dimers. (b) Filaments of variable lengths. (c) Long paired filaments (Courtesy of J Frazier). (d) and (e) Higher order structures formed by a three polypeptide recombinant mammalian septin complex. Complexes polymerize into long filaments that bundle (d) and curl up to form rings (e). Coomassie-stained polyacrylamide gel analysis of complexes on the left. Scale = 100 nm. (a) Modified from Byers B and Goetsch L (1976) A highly ordered ring of membrane-associated filaments in budding yeast. *Journal of Cell Biology* 69: 717–721. (e) Reproduced from Kinoshita M, Field CM, Coughlin ML, Straight AF, and Mitchison TJ (2002) Self- and actin-templated assembly of mammalian septins. *Developmental Cell* 3: 791–802.

Purified septin complexes contain tightly bound guanine nucleotide at a level of one molecule per septin polypeptide. The GTP ratio is $\sim 2:1$ for both *Drosophila* embryos and yeast complexes. The role of bound nucleotide in septin biochemistry is still under investigation, and appears to be different from that in Ras superfamily small GTPases, whose function employs rapid nucleotide exchange and hydrolysis. Isolated septin complexes exchange bound nucleotide very slowly, and in yeast, the majority of bound nucleotide does not turn over during one cell cycle. These data suggest that GTP is bound during septin folding or complex assembly, and thereafter is not exchanged, at least on the majority of septins. Thus, bound GTP may play a structural role, analogous to GTP bound to α -tubulin, and not a regulatory role, analogous to nucleotide in β -tubulin or small GTPases.

The guanine nucleotide-binding portion of human SEPT2 was recently expressed in *Escherichia coli* and crystallized as a dimer containing GDP. In addition, the crystal structure of

a three-polypeptide human heterohexameric septin complex containing SEPT2, 6, and 7 has been determined. These X-ray data, combined with EM data on the two polypeptide *Caenorhabditis elegans* heterotetrameric septin complex and a four polypeptide heterooctomeric complex of budding yeast septins, indicate that the unit septin complex in these two systems is nonpolar. This indicates septin filaments polymerized from these complexes are not polar, in contrast to actin and tubulin. However, the orientation of the coiled-coils does provide a lateral polarity to the complex. This could be important in septin filament lateral interactions (see below) or in interactions with membrane.

Septin Filaments can form Higher-Order Assemblies

Septin heterooligomers can assemble into several different higher-order structures *in vitro*. Septin complexes from all organisms studied tend to polymerize end to end to form long filaments the same thickness as the unit complexes.

Filaments assembled from yeast septins tend to associate side by side in pairs a fixed distance apart, suggesting they are cross-bridged by one of the septin polypeptides (Figure 2(c)). Elaborate two-dimensional gauzes composed of septins closely apposed to the plasma membrane have been observed by EM of extracted budding yeast. Such assemblies illustrate how septins could, by creating a neighborhood of even loosely confined alignment or low-affinity-binding sites, collectively have a profound effect on organizing and scaffolding structural and regulator proteins.

Purified mammalian septin filaments tend to assemble into bundles, which under some conditions curl up into rings and coils of $\sim 0.7\ \mu\text{m}$ diameter (Figures 2(d) and 2(e)). This tendency to curve is apparently intrinsic to the septin complex, and may play a role in deforming the plasma membrane in cells. The rings are comparable in size and shape to several septin assemblies in cells, including the yeast bud neck (Figures 3(a)–3(d)) and septin rings formed in mammalian cells under stress (Figure 4(a)). Septin filaments reconstituted *in vitro* can also associate with F-actin bundles via the conserved scaffold protein anillin. This suggests that there is also a close association and possible coalignment of septins, anillin, and F-actin in the contractile ring during cytokinesis. Septins associate with the bundled F-actin in interphase stress fibers of mammalian-cultured cells (Figure 4(b)), but this is via an adaptor other than anillin.

Septins Interact with Inositol Phospholipids

A number of studies have suggested that septins bind directly to lipid bilayers containing inositol lipids, an activity that may be important in septin targeting or in regulating exocytosis. This capacity also suggests that the septins contribute to coupling the plasma membrane to the cortical contractile cytoskeleton during cytokinesis. The questions of how septins target to plasma membranes, and how this affects their functions, are important topics for future study.

Septin Behavior and Function in Cytokinesis

Budding Yeast

Septin proteins were originally identified in budding yeast (*Saccharomyces cerevisiae*) as the protein products of four genes CDC3, CDC10, CDC11, and CDC12. Temperature-sensitive mutations of these genes exhibited hyperpolarized growth and defects in cell wall deposition and cytokinesis. These septin polypeptides localize to the bud neck late in the G1 phase of the *S. cerevisiae* cell cycle, before the localization of other cleavage furrow components. Septin localization and assembly are controlled, at least in part, by the GTPase Cdc42, the master regulator of yeast cell polarity, and are independent of other cytoskeletal components including actin filaments. By EM, septins appear as 10-nm diameter filaments, termed neck filaments that lie circumferentially around the bud neck (Figures 3(a) and 3(b)). By immunofluorescence they can appear as an hourglass-shaped assembly coating the inside of the bud neck or as two rings, one just inside the mother and one on the bud side of the neck (Figure 3(c)). Photobleaching of GFP-tagged septins reveals that septins are quite dynamic before bud emergence, but once they

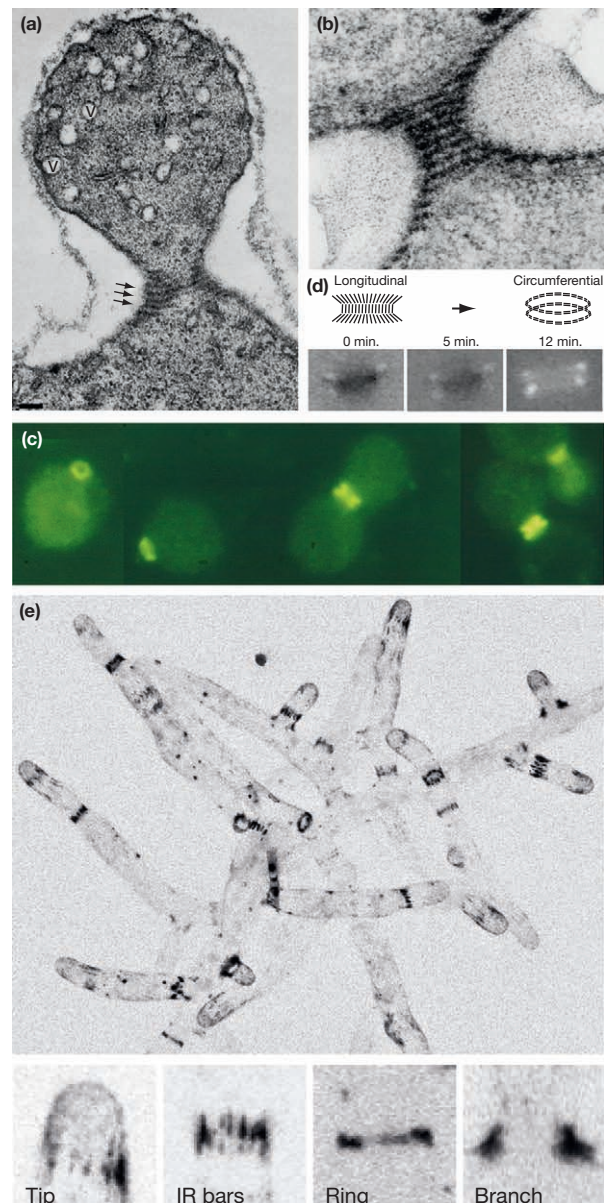


Figure 3 Septin structures/localization in *S. cerevisiae* and *A. gossypii* (a) and (b) Electron micrographs showing grazing sections through an early bud (a) and the neck of a large-budded cell (b) showing the 10-nm neck filaments (arrows in (a)). (c) Yeast at various stages of the cell cycle stained with an antibody against Cdc3p (Courtesy of J Pringle). (d) Polarized fluorescence microscopy reveals that bud neck septin assemblies are dynamic at the onset of cytokinesis, changing orientation by 90° from parallel with the budding axis to circumferential (Courtesy of Alina Vraiboiu). (e) Living *Ashbya gossypii* filamentous fungi expressing Shs1-GFP (Sep7), which localizes to the growing tips, interregion (IR) bars, rings, and branch points. Scale: *Ashbya* hyphae are approximately $5\ \mu\text{m}$ in diameter (Courtesy of Bradley DeMay and Amy Gladfelter). (a) Reproduced from Byers B and Goetsch L (1976) A highly ordered ring of membrane-associated filaments in budding yeast. *Journal of Cell Biology* 69: 717–721. (b) Reproduced from Strathern JN, Jones EW, and Broach JR (eds.) (1981) *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*, pp. 59–96. New York, NY: Cold Spring Harbor Laboratory Press.

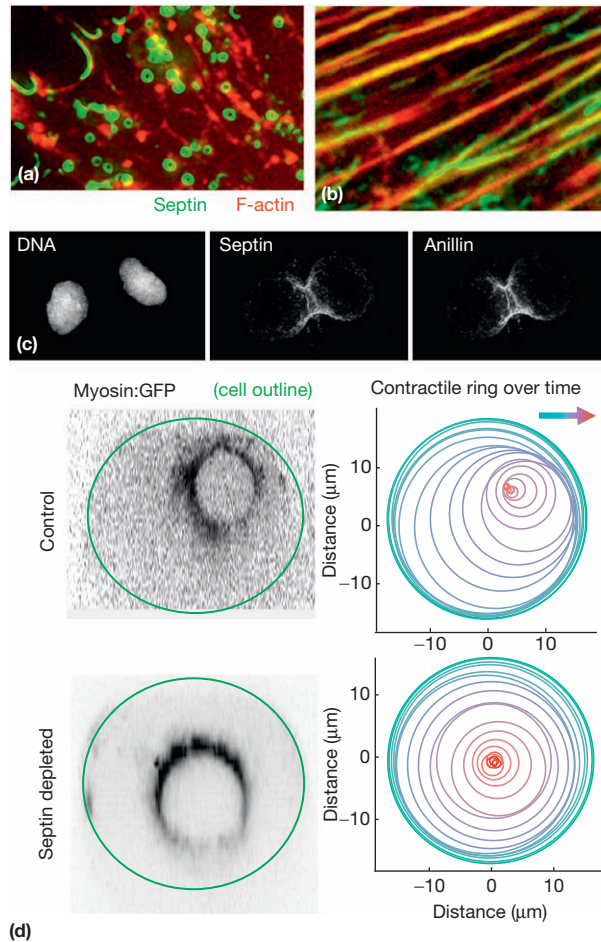


Figure 4 Septin structures, localization, and depletion phenotypes in metazoan cells. (a) and (b) Interphase vertebrate cells costained for septin and filamentous actin. (a) A vertebrate cell treated with a drug that depolymerizes actin filaments. Disruption of F-actin causes Septin to form rings of similar dimensions to those seen by EM *in vitro*. (b) Septin localizes along actin bundles. (c) A vertebrate cell in telophase stained for DNA and with antibodies raised against septin and anillin. Note septins and anillin colocalize in the cleavage furrow (Courtesy of Karen Oegema). (d) *C. elegans* zygotes depleted of the septins complete cytokinesis with approximately normal kinetics, but the contractile ring (visualized using GFP-tagged myosin) closes symmetrically within the division plane (bottom), unlike in controls (top). Left: the division plane of the zygote, viewed down the spindle axis. Right: the location of the contractile ring over time (start = green; end = red). (a) and (b) Reproduced from Kinoshita M, Field CM, Coughlin ML, Straight AF, and Mitchison TJ (2002) Self- and actin-templated assembly of mammalian septins. *Developmental Cell* 3: 791–802.

assemble into neck filaments, their turnover rate is slowed considerably, and they can be considered static. In an elegant demonstration of the dynamicity of bud neck septins, polarization fluorescence microscopy revealed that septin filaments initially lie along the mother-bud axis, and switch upon onset of cytokinesis to be oriented circumferentially (Figure 3(d)).

In budding yeast, septin localization at the bud neck is required for the subsequent recruitment of all other cytokinetic machinery, including a type II myosin heavy chain (Myo1p), its associated light chain (Mlc1p), a formin homology (FH)

protein Bni1p that is probably responsible for nucleating contractile ring F-actin, a Pombe Cdc15 homology (PCH) protein (Hof1p/Cyk2p), and an IQ motif-containing GTPase-activating protein protein (Iqg1p/Cyk1p). All of these proteins have homologs in other eukaryotes that also act in cytokinesis, but it is not clear if their recruitment to the division plane depends on septins in fission yeast or metazoans. Neck filaments are also thought to anchor a chitin synthase complex (Chs3p/4p + Bni4p) responsible for cell wall deposition during cytokinesis.

Budding yeast septins are also required for localization of many other proteins to the bud neck, in addition to the structural cytokinesis machinery mentioned above. A genome-wide screen identified 98 proteins that localize to bud necks, and many of these depend on septins for their localization. Well-characterized examples include the checkpoint kinases, Hsl1p, Gin4p and Kcc4p, a component of the mitotic exit network (MEN), Dbf2/Mob1, and several proteins involved in bud site selection including Bud4.

Septins also act to restrict diffusion of membrane proteins in the plane of the plasma membrane. The neck filaments form a fence that restricts membrane proteins to the bud, and presumably plays an important role in polarizing the yeast cell. If planar septin assemblies indeed bind to lipids of the plasma membrane, this function may be direct, as opposed to being mediated by other proteins that depend on septins for their localization. Such a diffusion barrier function may help explain the roles of septins in dendritic spine morphology and function in mammals, and in *Drosophila* ring canals, contexts in which plasma membrane compartmentalization is essential.

Overall, septins play a central role in the cell biology of budding yeast. This role reflects the importance of the bud neck in cell polarity and cell division, and the function of septins as a scaffold for localizing other proteins to this site, and restricting diffusion through it. In organisms that do not grow by polarized budding, septins may be important, but perhaps their role is not as central to the overall biology of the cell.

Ashbya gossypii

Ashbya is filamentous fungus that grows as a syncytium: because nuclear division is not coordinated with cell division, tens of nuclei exist together throughout the hyphal structure. Throughout the life cycle of *Ashbya*, the fungus elongates, forms spores, and can septate. Distinct septin structures are found, within the same cytoplasm, at the cell cortex in all these contexts, as well as at 'interregions' along the hyphal length (Figure 3(e)). Importantly, septins are not essential genes in *Ashbya*, so mutations and deletions can be made, propagated, and studied. Another characteristic that lends *Ashbya* septin structures to cell biological analysis is that they are larger than those in other fungi. Therefore, this species is likely to be a powerful model system in which to study multiple septin structures, and the interconversion among them, *in vivo*.

Fission Yeast

In fission yeast, *Schizosaccharomyces pombe*, disruption of the septin genes results in a delay in septation (cell-cell

separation), but not severe cytokinetic defects as seen in budding yeast. Septins localize to the division plane late in cytokinesis, on either side of the already-ingressing contractile ring, and as such are not required to recruit actin, myosin, or other contractile ring components. Stability of the septin ring requires the protein Mid2p, which is related to metazoan anillin, suggesting this interaction is conserved. Interestingly, mutations in components of the exocyst, a large complex involved in exocytosis, have a similar septation defect. Thus, in *S. pombe*, the septin scaffold is involved only in the completion stage of cytokinesis, perhaps to target exocytotic vesicles required for membrane fusion or enzymes required for final digestion of the septum.

Metazoa

In animal cells, septin polypeptides are recruited to the equatorial cortex and assembled into the contractile ring at the same time as, but independent of, actin and myosin-II (**Figure 4(c)**). Septin recruitment does depend on anillin. In fact, no known contractile ring proteins in metazoans depend on septins for their localization. Therefore, unlike budding yeast septins that are crucial for assembly of the contractile ring, animal septins act exclusively in its function. Perturbation of septins by gene disruption, RNA interference or microinjection of antiseptin antibodies perturbs cytokinesis in mammalian cells and fly embryos. However, septins are dispensable in some cases. For instance, nematode eggs can complete cytokinesis without septins at early developmental stages, although, furrowing is abnormally symmetric within the division plane (**Figure 4(d)**) and cytokinesis defects manifest in some cell lineages at post-embryonic stages.

In the early *Drosophila* embryo, septins are enriched on the cellularization front, an array of rings which descends among nuclei of the syncytial blastoderm. Septin localization to this structure depends on anillin. Analysis of a panel of anillin point mutants that are predicted to disrupt septin interaction indicates that anillin and septins orchestrate membrane organization during this elaborate event, in which embryo surface area increases 25-fold. How might septins regulate membrane dynamics? Physical and functional interactions have been shown between mammalian septins and syntaxins (soluble NSF attachment protein (SNAP) receptor (SNARE) proteins involved in membrane fusion during secretion) and the exocyst complex (a complex of proteins required for exocytosis). As previously mentioned, exocyst mutants in *S. pombe* have a phenotype similar to that of septin mutants. Thus, septins may

act via these partners to regulate or target vesicle insertion during furrow ingression.

Septins and anillin are also abundant in intracellular bridges between daughter cells following conventional cytokinesis and the stable bridges that form as the result of regulated incomplete cytokinesis in *Drosophila* embryos. This localization suggests that septins and anillin function together late in cytokinesis, perhaps during the complex process of completion/abscission. How might septins function in completion? As stated above, septins may act with syntaxins or the exocyst complex to regulate vesicle insertion at the intercellular bridge during completion. Additionally, septins and anillin may comprise a structure that mechanically stabilizes the fully ingressed membrane after the contractile ring has disassembled.

The differences in requirement for septins in contractile ring ingression and in cytokinesis indicate a divergence in cytokinesis mechanism that we do not understand. Septins may be more important during cytokinesis in some cells than others because of differences in (1) the amount of new membrane required for cytokinesis and (2) the time delay between completing ingression and actually separating the daughter cells.

See also: [Cell Architecture and Function: Cytokinesis.](#)

Further Reading

- Balasubramanian MK, Bi E, and Glotzer M (2004) Comparative analysis of cytokinesis in budding yeast, fission yeast and animal cells. *Current Biology* 14: R806–R806.
- Caudron F and Barral Y (2009) Septins and the lateral compartmentalization of eukaryotic membranes. *Developmental Cell* 16: 493–506.
- DeMay BS, Meseroll RA, Occhipinti P, and Gladfelter AS (2009) Regulation of distinct septin rings in a single cell by Elm1 and Gin4p kinases. *Molecular Biology of the Cell* 20: 2311–2326.
- Field CM, Maddox AS, Pringle JR, and Oegema KF (2008) Septins in the metazoan model systems *Drosophila melanogaster* and *Caenorhabditis elegans*. In: Hall PA, Russell SHE, and Pringle JR (eds.) *The Septins*, pp. 147–168. Chichester: Wiley-Blackwell.
- McMurray MA and Thorner J (2009a) Reuse, replace, recycle. Specificity in subunit inheritance and assembly of higher-order septin structures during mitotic and meiotic division in budding yeast. *Cell Cycle* 8: 195–203.
- McMurray MA and Thorner J (2009b) Septins: Molecular partitioning and the generation of cellular asymmetry. *Cell Division* 4: 18.
- Pan F, Malmberg RL, and Momany M (2007) Analysis of septins across kingdoms reveals orthology and new motifs. *BMC Evolutionary Biology* 1: 103.
- Pollard TD (2010) Mechanics of cytokinesis in eukaryotes. *Current Opinion in Cell Biology* 22(1): 50–56.
- Russell SEH and Hall PA (2005) Do septins have a role in cancer? *British Journal of Cancer* 93: 499–503.
- Weirich CS, Erzberger JP, and Barral Y (2008) The septin family of GTPases: Architecture and dynamics. *Nature Reviews Molecular Cell Biology* 9: 478–489.

Serine/Threonine Phosphatases

T S Ingebritsen, Iowa State University, Ames, IA, USA

Published by Elsevier Inc.

This article is reproduced from the previous edition, volume 4, pp. 27–32, © 2004, Elsevier Inc.

Glossary

Modulator protein Generally a low-molecular-weight, heat-stable protein that alters protein phosphatase activity or substrate specificity.

Protein kinase cascade A series of protein kinases arranged in a linear fashion in a signal transduction pathway such that an upstream protein kinase phosphorylates and activates the immediate downstream kinase.

Protein phosphatase An enzyme whose physiological function is to remove phosphate groups from serine, threonine, or tyrosine residues of proteins.

Targeting subunit A protein that directs a phosphatase to a specific subcellular location or a specific substrate and may also modulate substrate specificity and regulate phosphatase activity.

Transcription factor A protein that binds to a regulatory site on a gene leading to enhanced transcription of the gene.

Phosphorylation of proteins on serine, threonine, and tyrosine residues is a major mechanism for regulating the activity of cell proteins and it plays a central role in virtually all signal transduction pathways in eukaryotes. The steady-state level of phosphorylation of a protein at a particular site depends on the balance of the activities of the protein kinase(s) and protein phosphatase(s) acting on that site. Both protein kinases and protein phosphatases are important targets of cell regulation. This article focuses on the structure, regulation, and function of the two families of protein Ser/Thr phosphatases (phospho-protein phosphatase (PPP) and protein phosphatase Mg^{2+} - or Mn^{2+} -dependent (PPM)). Members of each family are present in all three domains of life (archaea, bacteria, and eukaryotes).

Protein Ser/Thr Phosphatase Catalytic Subunit Families

PPP Family

Members of this family possess a common catalytic core (280 residues) and can be further divided into four subfamilies termed PPP1, PPP2A, PPP2B, and PPP5. Three prokaryotic phosphatases: diadenosine tetraphosphatase, $\Phi 80$ phosphatase, and λ phosphatase exhibit weaker similarity to the PPP family.

PPM Family

Members include PP2C, *Arabidopsis* ABI1 and ABI2, *Arabidopsis* KAPP, *Bacillus subtilis* SpoIIE phosphatase, and pyruvate dehydrogenase phosphatase. The core PPM catalytic domains occur in diverse contexts. For example, the catalytic domain of *Arabidopsis* ABI1 is fused to an EF-hand domain, the *Arabidopsis* kinase-associated protein phosphatase (KAPP) catalytic domain is fused to a kinase-interaction domain. The kinase interaction domain of KAPP associates with a phosphorylated receptor-like protein. The N terminus of *B. subtilis* SpoIIE phosphatase is fused to a domain with 10 membrane-spanning segments.

Biochemical Characterization of Signature Protein Ser/Thr Phosphatases

PP1, PP2A, and PP2B are signature members of the PPP1, PPP2A, and PPP2B subfamilies while PP2C is the signature member of the PPM family. These four enzymes account for the majority of protein Ser/Thr phosphatase activity in cell extracts. The activities of these enzymes can be distinguished in cell extracts based on divalent cation requirements and the effects of physiological and pharmacological inhibitors (Table 1). PP1, PP2A, and PP2C have broad and overlapping substrate specificities whereas the substrate specificity of PP2B is more restricted.

PP2B requires μM Ca^{2+} for activity whereas PP2C requires mM Mg^{2+} . Inhibitor-1 (I_1) and inhibitor-2 (I_2) are PP1 modulator proteins. Okadaic acid is a pharmacological inhibitor of PPP family members. It is a polyether carboxylic acid that is a diarrhetic shellfish poison and a powerful tumor promoter. There are a number of other pharmacological inhibitors of the PPP family (e.g., microcystin, calyculin A, and cantharidin). Okadaic acid, microcystin, and calyculin A inhibit by binding to the phosphatase active site.

Three-Dimensional Structures and Catalytic Mechanism

Three-dimensional (3D) structures have been determined for two members of the PPP family (PP1 and PP2B) and for one member of the PPM family. A surprising finding was that the PPP and PPM families have similar 3D architectures even though the primary structures of the two families are unrelated. The 3D structure of the two families is quite distinct from that of the PTP family of protein Tyr phosphatases.

3D Structures

For both the PPM and PPP families, the core structure consists of a pair of mixed β -sheets that form a β -sandwich structure.

Table 1 Biochemical characterization of signature protein phosphatases

Type	Protein phosphatase	Substrate specificity	Divalent cation requirement	I_1 and I_2 inhibition	Okadaic acid inhibition (IC_{50})
1	PP1	Broad	None	Yes	20 nM
2	PP2A	Broad	None	No	0.2 nM
2	PP2B	Narrow	Ca^{2+}	No	5 μ M
2	PP2C	Broad	Mg^{2+}	No	No effect

PP2B requires μ M Ca^{2+} for activity whereas PP2C requires mM Mg^{2+} . Inhibitor-1 (I_1) and inhibitor-2 (I_2) are PP1 modulator proteins. Okadaic acid is a pharmacological inhibitor of PPP family members. It is a polyether carboxylic acid that is a diarrhetic shell fish poison and a powerful tumor promoter. There are a number of other pharmacological inhibitors of the PPP family (e.g., microcystin, calyculin A, cantharidin). Okadaic acid, microcystin, and calyculin A inhibit by binding to the phosphatase active site.

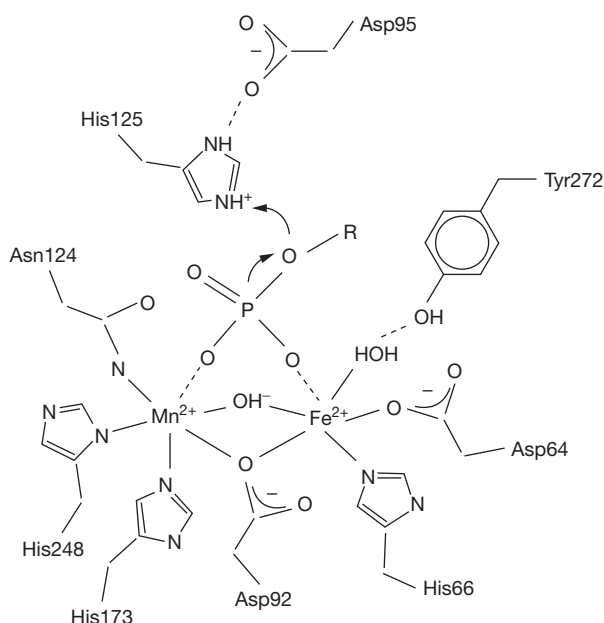


Figure 1 Active site structure and catalytic mechanism of PP1 catalytic subunit. Barford D (1996) Molecular mechanisms of the protein serine/threonine phosphatases. *Trends in Biochemical Sciences* 21: 407–412.

The catalytic sites possess a binuclear metal ion center, which has some similarity to the binuclear metal ion center of purple acid phosphatase (Figure 1). For the PPM family, the two metal ions are Mn^{2+} . In the case of the PPP family, the two metal ions are probably Fe^{2+} and Zn^{2+} , although there is some controversy about whether the second metal ion is Zn^{2+} or Mn^{2+} and also about whether the Fe is in the 2+ or 3+ oxidation state. In both cases, metal ions are coordinated with water molecules.

Catalytic Mechanism

For both the PPP and PPM family, hydrolysis of serine or threonine phosphate esters occurs through a single-step mechanism in which a metal-bound water acts as a nucleophile to attack the phosphorus atom of the substrate phosphate group (Figure 1). The metal ion acts as a Lewis acid to enhance the nucleophilicity of metal-bound water and it also enhances the electrophilicity of the phosphorus atom by coordinating the two oxyanions of the phosphate group. An active site

His sidechain donates a proton to the leaving oxygen of Ser or Thr. This catalytic mechanism is quite different from that used by the PTP family which involves formation of a phospho-enzyme intermediate.

Subunit Structure of PPP Family Members

Members of the PPP family interact with diverse sets of regulatory subunits, which direct the catalytic subunits to specific subcellular locations, alter substrate specificity and/or confer regulatory properties.

Interaction of PP1 with Diverse Regulatory Subunits

The catalytic subunit of PP1 (PP1c) interacts with >50 regulatory subunits, which fall into two classes: targeting subunits and modulator proteins. Targeting subunits direct PP1c to a wide variety of subcellular locations, including glycogen particle, myosin/actin, spliceosomes/RNA, endoplasmic reticulum, proteasomes, nuclear membranes, plasma membranes/cytoskeleton, centrosomes, microtubules, and mitochondria. Targeting subunits also modulate substrate specificities and may regulate phosphatase activity.

Modulators are generally low-molecular-weight, heat-stable proteins that alter PP1 activity or substrate specificity. The activity of some of the modulators (e.g., inhibitor-1, DARPP-32, CPI-17, and G subunit) is regulated through reversible phosphorylation.

Strong binding of many of the regulatory proteins to PP1c is mediated through a short motif termed 'RVxF'. This motif is found in two-thirds of the targeting subunits and one-half of the modulator proteins. The consensus sequence of the motif is: (K/R) x_1 (V/I) x_2 (F/W) where x_1 and x_2 may be any residue except a large hydrophobic residue. In some motifs, x_1 is absent. The RVxF motif binds to a docking site that is remote from the active site (Figure 2(a)). The docking site consists of a hydrophobic groove which binds the two large hydrophobic residues of the RVxF motif and a cluster of negatively charged residues which interact with basic residues at the N-terminal end of the motif.

Additional interactions between the PP1c and the targeting and modulator proteins are thought to strengthen binding and mediate effects on PP1 activity. For example, residues 7–11 of DARPP-32 (KKIQF) bind to the PP1c docking site whereas Thr 34 which is phosphorylated by PKA binds to the active site of the phosphatase.

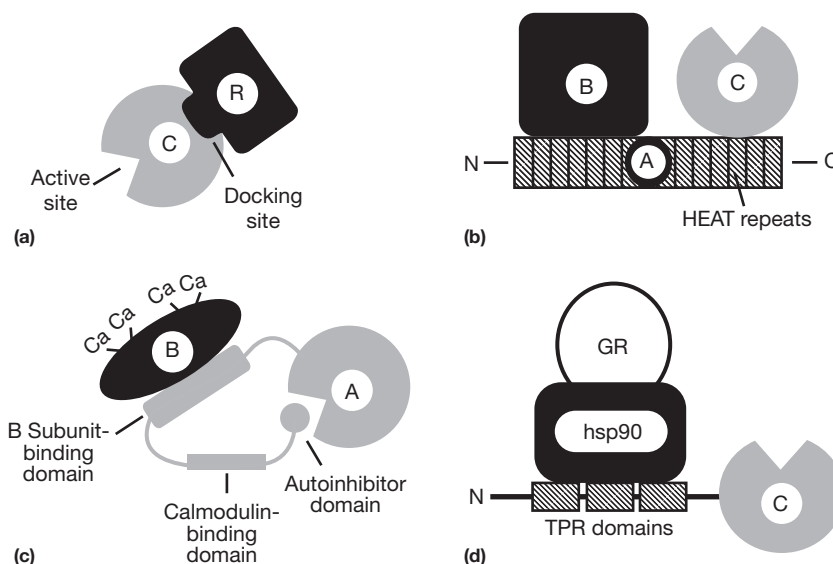


Figure 2 Subunit structure of PPP family members. (a) PP1. C and R designate catalytic and regulatory subunits, respectively. The docking site on the C subunit interacts with the RVxF motif of the regulatory subunit. (b) PP2A. The labels C, B, and A designate the catalytic, regulatory, and adaptor subunits, respectively. The three-dimensional (3D) structure of the A subunit has been determined. Tandem arrays of huntingtin-elongation factor-A subunit-TOR (HEAT) motifs are present in a variety of other proteins. The labels N and C designate the amino and carboxyl termini of the A subunit. (c) PP2B. The labels A and B designate the catalytic and regulatory subunits, respectively. The B subunit has four Ca^{2+} -binding sites. The B subunit-binding, calmodulin-binding and autoinhibitor domains are on a carboxyl-terminal extension of the A subunit. The region from the end of the B subunit-binding domain to the beginning of the autoinhibitor domain is disordered in the absence of calmodulin and thus not visible in the crystal structure. (d) PP5. C and GR designate the catalytic subunit and the glucocorticoid receptor, respectively. TPR domains are characterized by a degenerate 34 amino acid sequence. The label N designates the amino terminus of the C subunit. The 3D structure of the amino-terminal extension of C has been determined in the absence of the core catalytic domain.

Interaction of PP2A with a Diverse Set of Regulatory Subunits

PP2A diversity is also generated by the interaction of a common catalytic (C) subunit with a diverse set of regulatory (B) subunits. However in this case, the regulatory subunits interact with the C subunit indirectly through an adapter subunit (A), thus forming a heterotrimeric phosphatase complex (Figure 2(b)).

Over 15 different B subunits are expressed in a tissue- and developmental-specific manner from four families of genes, termed PR55/B, PR61/B', PR72/B', and B'. Additional gene products are generated through alternate splicing. Functions of the B subunits include regulation of PP2A activity, subcellular targeting, and alteration of substrate specificity.

The A subunit is made up of 15 huntingtin-elongation factor-A subunit-TOR (HEAT) repeats which form an extended and curved structure reminiscent of a hook or the letter C. The catalytic subunit interacts with the C-terminal HEAT repeats (11–15) whereas the B subunits interact with N-terminal repeats (1–10).

Subunit Structure of PP2B

The catalytic (A) subunit of PP2B interacts with two EF-hand-type Ca^{2+} -binding proteins, an integral B subunit, which binds in the absence of calcium, and calmodulin which requires calcium for binding. The A subunit has a regulatory C-terminal extension which has binding sites for the two Ca^{2+} -binding proteins as well as an autoinhibitor site (Figure 2(c)). Binding of Ca^{2+} to the B subunit is absolutely required for phosphatase activity. PP2B activity is further stimulated by Ca^{2+} -calmodulin.

PP5

The catalytic subunit of PP5 has an amino-terminal extension with three tetratricopeptide repeat (TPR) domains. These domains are found in a variety of proteins and act as a scaffold that mediates protein–protein interaction. The TPR domains of PP5 mediate interaction of the phosphatase with the heat shock protein, hsp90, and the glucocorticoid receptor (Figure 2(d)). TPR domains also suppress the catalytic activity of PP5 25-fold.

Examples of Functions and Regulation of Protein Ser/Thr Phosphatases

PPM Family

PPM family members are involved in stress responses in animals, plants, fungi, and prokaryotes. PP2C antagonizes stress-response pathways involving two types of protein kinase cascades: mitogen-activated protein kinase and adenosine monophosphate (AMP)-activated protein kinase. Two other PPM family members (ABI1 and ABI2) play an essential role in the abscisic acid-mediated stress response of plants to water deprivation, and in prokaryotes, SpoIIE is involved in controlling sporulation.

PPM family members also have other cell functions. For example, the pyruvate dehydrogenase phosphate phosphatase is involved in controlling the types of metabolic fuels used in body tissues through a dephosphorylation reaction that activates pyruvate dehydrogenase in mitochondria.

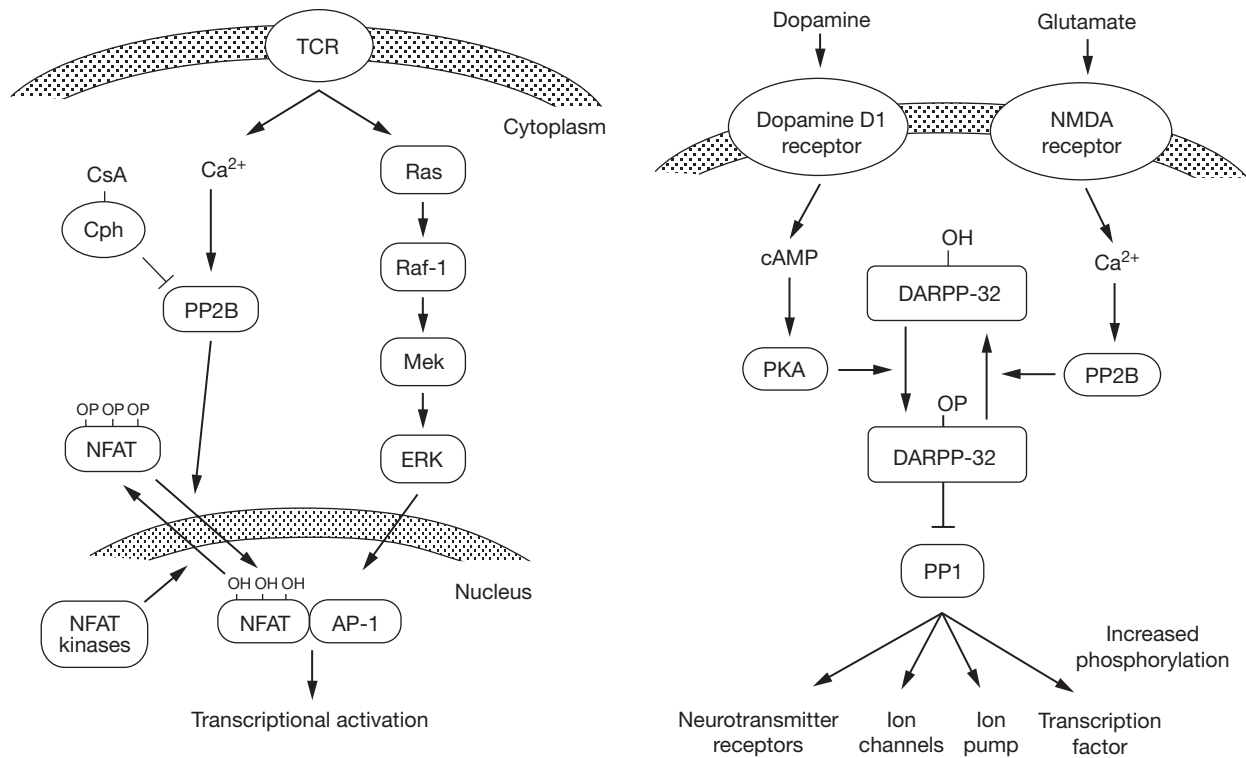


Figure 3 (Left) Role of PP2B in T-cell activation. Raf-1, Mek, and ERK are protein kinases in the Ras signaling pathway. NFAT and AP-1 are transcription factors. NFAT are a family of transcription factors that exist in an inactive, phosphorylated state in the cytoplasm of resting T cells. These proteins are phosphorylated on multiple serine residues in a regulatory region in the amino-terminal half of the molecule. CsA and Cph designate cyclosporin and cyclophilin, respectively. (Right) Central role of DARPP-32 in the regulation of dopaminergic neurons. Inhibition of PP1 is associated with increased activity of NMDA and AMPA glutamate receptors, of L, N, and P type Ca²⁺ ion channels and CREB and with decreased activity of GABA_A receptor, Na⁺ channels, and Na⁺/K⁺ ATPase. There are a variety of other neurotransmitter receptors (opiate, adenosine, VIP) which elevate cAMP in medium spiny neurons. Glutamate acting through AMPA receptors and GABA acting through GABA_A receptors also elevate Ca²⁺ in these neurons.

PPP Family

Regulation of PP1 involving targeting subunits

There are four modes of regulation of PP1c involving targeting subunits: inducible expression of targeting subunits, allosteric regulation through targeting subunits, phosphorylation near the RVxF docking motif, and phosphorylation at sites remote from the docking motif.

The expression of two glycogen-targeting subunits in rat liver, G_L and R5, is decreased by diabetes and starvation and is restored by insulin treatment and refeeding, respectively.

G_L-PP1c is subject to allosteric regulation by glycogen phosphorylase *a* (phosphorylated, active form) which binds to a short segment at the C terminus of G_L with nanomolar affinity. Phosphorylase *a* binding inhibits the glycogen synthase phosphatase activity of G_L-PP1c but has no effect on phosphorylase phosphatase activity of the complex. This helps to coordinate the regulation of glycogen breakdown and synthesis in response to glucagon and perhaps to insulin.

G_M (muscle-specific glycogen targeting), NIPP1 (nuclear targeting), and Neurabin I (actin targeting) are phosphorylated by protein kinase A at sites near the RVxF-docking motif leading to dissociation of PP1c. In the case of G_M-PP1c, this results in decreased activity toward glycogen-bound substrates (glycogen synthase, phosphorylase *a*, and phosphorylase kinase).

The myosin-targeting protein M110 is phosphorylated on sites (Thr 697 and Ser 435) that are distant from the docking motif. The complex of PP1c with M110 is involved in regulating muscle contraction in smooth muscle and nonmuscle cells. Phosphorylation of Ser 435 occurs during mitosis and leads to activation of PP1c and enhanced binding to myosin II.

PP2A

One of the best-documented roles of PP2A is in the regulation of animal growth and development. A role in cell growth was first suggested by the potent inhibition of PP2A by the tumor promoter okadaic acid. Additionally, the β -isoform of the A subunit of PP2A has been identified as a candidate tumor-suppressor gene and the myeloid-leukemia-associated protein SET is a potent inhibitor of PP2A.

PP2A dephosphorylates and inactivates protein kinases involved in growth-regulatory signal transduction pathways (e.g., extracellular signal-regulated protein kinase (ERK) and Mek MAP kinases, protein kinase C, and protein kinase B). Additionally, PP2A is an important cellular target of the SV40 and polyoma DNA tumor viruses. Viral proteins associate with the AC dimer and displace B subunits. This leads to stimulation of growth-related (ERK/Mek) MAP kinase pathways and cell growth.

Genetic approaches in budding and fission yeast as well as in *Drosophila* demonstrate that PP2A has an essential role in regulating the cell cycle. Additionally, PP2A interacts with components of the Wnt signaling cascade, which controls the epithelial–mesenchymal transition during vertebrate development.

Central role of PP2B in T-cell activation

Stimulation of the T-cell receptor leads to the activation of dual signal transduction pathways involving Ca^{2+} and Ras (Figure 3, left). Elevation of intracellular Ca^{2+} results in activation of PP2B, which dephosphorylates nuclear factor of activated T cells (NFAT). This leads to activation of NFAT as a transcription factor and translocation of the NFAT from the cytoplasm to the nucleus. The Ras pathway activates a nuclear transcription factor (AP-1) through a protein kinase cascade. NFAT and AP-1 bind cooperatively to DNA regulatory sites resulting in enhanced transcription of cytokine, cell surface receptor, and transcription factor genes.

PP2B is the site of action for the immune-suppressant drugs, Cyclosporin A and FK506, used to prevent rejection in organ transplant procedures. The use of these drugs has revolutionized organ transplant therapy. The two drugs interact with separate intracellular-binding proteins (cyclophilin and FK506-binding protein) and the resulting complexes bind to and inhibit the activity of PP2B. This in turn inhibits T-cell activation by suppressing transcriptional activation through the NFAT:AP-1 complex.

Similar dual pathways involving PP2B and NFAT family members have been implicated in angiotensin-II-induced cardiac hypertrophy and in the morphogenesis of heart valves.

DARPP-32

DARPP-32 (dopamine and cyclic adenosine 3',5'-monophosphate-regulated phosphoprotein, 32 kDa) is a specific inhibitor of PP1. It is expressed at high concentrations in medium spiny neurons of the neostriatum where it plays a central role in integrating responses to dopamine (acting via the cAMP) and glutamate (acting via Ca^{2+}) in dopaminergic neurons (Figure 3, right). Diseases associated with defects in dopaminergic neurotransmission include Parkinson's disease,

Huntington's disease, attention deficit hyperactivity disorder, and schizophrenia.

Activation of the cAMP pathway in dopaminergic neurons leads to increased phosphorylation of DARPP-32 on Thr 34 and inhibition of PP1. This results in increased phosphorylation of neurotransmitter receptors, voltage-gated ion channels, an electrogenic ion pump (Na^+/K^+ ATPase), and a transcription factor (CREB).

By contrast, glutamate acting through N-methyl-D-aspartate (NMDA) receptors promotes dephosphorylation of brain proteins through activation of a protein phosphatase cascade involving PP2B and PP1. Activation of NMDA receptors elevates Ca^{2+} , which activates PP2B. This leads to DARPP-32 dephosphorylation by PP2B and activation of PP1 through relief of inhibition by DARPP-32.

See also: **Metabolism Vitamins and Hormones:** Pyruvate Kinase; Structure and Regulation of Pyruvate Dehydrogenases; **Protein/Enzyme Structure Function and Degradation:** Allosteric Regulation; **Signaling:** Angiotensin Receptors; Dopamine Receptors; Protein Tyrosine Phosphatases; Src Family of Protein Tyrosine Kinases.

Further Reading

- Barford D (1996) Molecular mechanisms of the protein serine/threonine phosphatases. *Trends in Biochemical Sciences* 21: 407–412.
- Ceulemans H, Stalmans W, and Bollen M (2002) Regulatory-driven functional diversification of protein phosphatase-1 in eukaryotic evolution. *BioEssays* 24: 371–381.
- Cohen P (2002) Protein phosphatase 1 – targeted in many directions. *Journal of Cell Sciences* 115: 241–256.
- Greengard P, Allen P, and Nairn A (1999) Beyond the dopamine receptor: The DARPP-32/protein phosphatase-1 cascade. *Neuron* 23: 435–447.
- Ingebritsen T and Cohen P (1983) Protein phosphatases: Properties and role in cellular regulation. *Science* 221: 331–338.
- Janssens V and Goris J (2001) Protein phosphatase 2A: A highly regulated family of serine/threonine phosphatases implicated in cell growth and signalling. *Biochemical Journal* 353: 417–439.
- Rao A, Luo C, and Hogan P (1997) Transcription factors of the NFAT family: Regulation and function. *Annual Review of Immunology* 15: 707–747.
- Rodriguez P (1998) Protein phosphatase 2C (PP2C) function in higher plants. *Plant Molecular Biology* 38: 919–927.

Serotonin Receptor Signaling

P J Gresch, Vanderbilt University-School of Medicine, Nashville, TN, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Paul J. Gresch and Elaine Sanders-Bush, volume 4, pp. 33–37, © 2004, Elsevier Inc.

Glossary

Autoreceptor A receptor on the neuronal cell body or presynaptic terminal that can regulate its own cell firing and/or neurotransmitter release and synthesis.

Constitutive activity Ability of a receptor to activate second messenger pathways without the binding of an external ligand.

Functional selectivity The ability of a receptor to couple to more than one signal cascade.

Heteroreceptor A presynaptic receptor that regulates the release other than its own ligand.

RNA editing A process whereby the nucleotide sequence of RNA transcripts are chemically altered.

Serotonin Synthesis and Metabolism

5-Hydroxytryptamine (5-HT) is a neurotransmitter in both invertebrates and vertebrates. 5-HT is found in the central nervous system, enterochromaffin cells, gastrointestinal (GI) tract, and platelets. 5-HT is synthesized from the essential amino acid tryptophan by the rate-limiting enzyme tryptophan hydroxylase and a ubiquitous l-aromatic amino acid decarboxylase. 5-HT is released into the synaptic cleft by exocytosis of vesicles in a TTX-sensitive and Ca^{2+} -dependent manner. Inactivation of 5-HT is mediated by reuptake into the presynaptic terminal through a Na^{+} -dependent 5-HT transporter. 5-HT is metabolized into the inactive form, 5-hydroxyindole acetic acid, by the enzymes monoamine oxidase and aldehyde dehydrogenase. Synaptic 5-HT concentrations can fluctuate and be manipulated by environmental influences and pharmacological agents. For instance, restriction of dietary tryptophan or chemical inhibitors of tryptophan hydroxylase reduces brain levels of 5-HT, while selective 5-HT reuptake inhibitors such as escitalopram (Lexapro) and fluoxetine (Prozac) increase the amount of synaptic 5-HT. Released serotonin acts on multiple 5-HT receptors found throughout the body in various tissues including brain, spinal cord, heart, blood, and gut.

Serotonin Receptor Structure and Function

5-HT receptors are classified and characterized by their gene organization, amino acid sequences, pharmacological properties, and second-messenger coupling pathways. With the exception of the 5-HT₃ receptor, the 5-HT receptor family consists of guanine nucleotide triphosphate (GTP)-binding proteins (G-protein)-coupled receptors (GPCRs). The basic protein structure is predicted to contain seven transmembrane regions, three intracellular loops, and three extracellular loops, with the amino terminus being extracellular and carboxy terminus, intracellular (Figure 1(a)). These receptors are linked to their signal transduction pathways through G-proteins. The sequence of events involves the activation of the cell-surface receptor by 5-HT or drugs, binding of receptor and G-protein, which promotes the exchange of bound guanine nucleotide diphosphate for GTP on

the G-protein. The G-protein is comprised of α , β , and γ subunits; the $\beta\gamma$ dimer dissociates when receptor binds and both $G\alpha$ and $G\beta\gamma$ have the ability to interact with effector enzymes. The subunit interactions promote activation or inhibition of adenylate cyclase or activation of phospholipase C (PLC). In turn, the effector enzymes generate second messengers that regulate cellular processes such as Ca^{2+} release and protein kinases and phosphatases. This multistep enzymatic process amplifies receptor signal, and provides the possibility of regulation and cross-talk at multiple levels. We have grouped the numerous 5-HT receptors family by their traditional (primary) G-protein-second-messenger linkage (Table 1), and then the multiple levels of signal diversity generated by RNA splicing, RNA editing, and promiscuity of receptor G-protein activation are discussed.

Serotonin Receptors That Inhibit Adenylyl Cyclase

The five members of the 5-HT₁ receptor family (5-HT_{1A}, 5-HT_{1B}, 5-HT_{1D}, 5-HT_{1E}, and 5-HT_{1F}) and the 5-HT₅ receptor (5-HT_{5A}, 5-HT_{5B} subtype) couple primarily through $G_{\text{zi/o}}$ proteins to inhibit the membrane-bound enzyme, adenylyl cyclase (AC). This inhibition of AC leads to a decrease of cyclic 3'/5'-adenosine monophosphate (cAMP) molecules (Figure 1(b)). The 5-HT₁ receptors are the best characterized of this family. High densities of 5-HT_{1A} receptors are found on the cell bodies of 5-HT neurons in brainstem nuclei, especially the dorsal raphe. In the dorsal raphe, the 5-HT_{1A} receptor functions as an autoreceptor that reduces 5-HT cell firing when activated. The receptor elicits neuronal membrane hyperpolarization by activating G-protein-linked K^{+} channels. In addition, 5-HT_{1A} receptors are located on postsynaptic sites in the hippocampus and other limbic brain regions where they also produce hyperpolarization by opening K^{+} channels. 5-HT_{1A} receptors may have a possible role in anxiety and are targets of a class of antianxiety drugs. The 5-HT_{1A} receptor is also involved in thermoregulation, hormone secretion, and possibly enhancing the effectiveness of antidepressants. Of the other 5-HT₁ receptors, most is known about the 5-HT_{1B} and 5-HT_{1D} subtypes. These receptors are found in basal ganglia and frontal cortex and function as terminal autoreceptors or heteroreceptors that

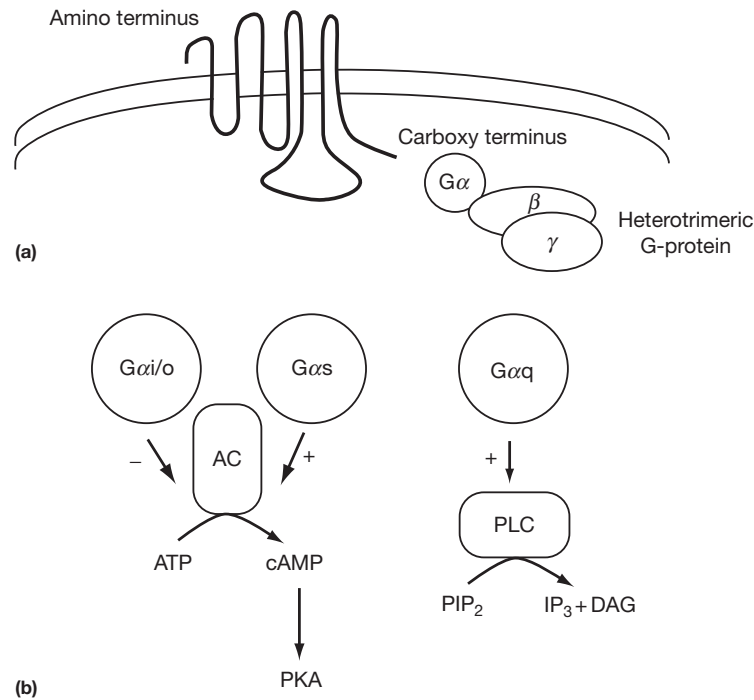


Figure 1 (a) Schematic drawing of the basic components of a G-protein-coupled receptor illustrating the extracellular amino terminus, seven transmembrane domains, three intracellular loops, three extracellular loops, intracellular carboxy terminus, and heterotrimeric G-protein. (b) The primary signaling pathways of the G-protein-coupled receptors. Activation of $G_{\alpha i/o}$ protein inhibits AC resulting in the decrease of cAMP production. Receptors that stimulate AC through $G_{\alpha s}$ results in increased production of cAMP, with subsequent activation of protein kinase A (PKA). Stimulation of PLC β through $G_{\alpha q}$ results in the cleavage of PIP_2 into IP_3 and DAG.

Table 1 Signaling linkages of 5-HT receptors

G-protein	Receptor subtype	Effector linkage
G_i/G_o	1a	Inhibition of adenylate cyclase
	1b	
	1d	
	1e	
	1f	
G_s	5	Activation of adenylate cyclase
	4	
	6	
	7	
G_q/G_{11}	2a	Activation of phospholipase C
	2b	
	2c	
No G-protein	3	Ligand-gated ion channel

modulate neurotransmitter release. 5-HT_{1B/1D} heteroreceptors have been proposed to regulate the release of acetylcholine, glutamate, dopamine, norepinephrine, and γ -aminobutyric acid (GABA). Pharmacological and genomic knockout studies suggest that 5-HT_{1B} receptors are involved in aggressive behavior. 5-HT_{1D} receptors have a role in migraine headaches and many antimigraine drugs target this receptor. Less is known about 5-HT_{1F}, 5-HT_{1E}, and 5-HT₅ receptors primarily due to the lack of specific ligands. 5-HT_{1F} receptors are a target for antimigraine medication and 5-HT₅ receptors appear to have a role in circadian rhythms, mood, and cognition.

Serotonin Receptors Linked to Activation of Adenylate Cyclase

5-HT₄, 5-HT₆, and 5-HT₇ receptors are all coupled to activation of AC. These receptors are linked via the G-protein, G_{α_s} , to AC producing an increase cAMP production (Figure 1(b)). Historically, the first signal transduction pathway to be linked to a 5-HT receptor was stimulation of AC, characterized in mouse collicular neurons. This receptor now known as the 5-HT₄ receptor, which is also found in hippocampus and is involved in learning and memory, has been linked to modulation of release of acetylcholine, dopamine, 5-HT, and GABA. In the peripheral tissues, the 5-HT₄ receptor can influence GI function. The receptor is expressed in both enterochromaffin cells and enteric neurons, where it releases acetylcholine in the ileum, contracts the esophagus and colon, and promotes ion transport in the gut. In addition, the 5-HT₄ receptor elicits cardiac contraction. Due to the paucity of specific ligands less is known about the 5-HT₆ and 5-HT₇ receptors. 5-HT₆ receptors are found in the striatum, amygdala, nucleus accumbens, hippocampus, and cortex. Limited pharmacological data suggest that 5-HT₆ receptors modulate neurotransmitters such as acetylcholine, glutamate, and dopamine. This exciting data support that 5-HT₆ receptors could be a target for cognitive-enhancement therapeutics for Alzheimer's disease and schizophrenia. 5-HT₇ receptors are widely expressed in the brain, with highest expression levels in the thalamus and the hippocampus. The 5-HT₇ receptor may have role in circadian rhythms and thermoregulation. Both 5-HT₆ and 5-HT₇ receptors

have high affinity for many of the atypical antipsychotics leading to speculation of a role for these receptor subtypes in schizophrenia.

Serotonin Receptors Coupled to the Activation of PLC

The 5-HT₂ class of receptor (subtypes 2A, 2B, and 2C) activates the membrane-bound enzyme PLC that catalyzes the degradation of the inositol lipid, phosphatidylinositol 4,5 bisphosphate (PIP₂) with the production of inositol 1,4,5 triphosphate (IP₃) and diacylglycerol (DAG) (Figure 1(b)). IP₃ mobilizes Ca²⁺ from intracellular storage sites; Ca²⁺ then induces multiple responses in the cell including activation of calcium/calmodulin-dependent protein kinase enzymes that phosphorylate protein substrates in the cell. DAG activates another kinase, protein kinase C. The 5-HT₂ receptors are coupled via the G-protein G_{αq} or G_{α11} to activation of PLC.

In the brain, 5-HT_{2A} receptors are found in cerebral cortex, claustrum, basal ganglia, and thalamus. It is thought that hallucinogenic drugs exert their psychotropic action by activating 5-HT_{2A} receptors. Activation of cortical 5-HT_{2A} receptors leads to membrane depolarization and opening of K⁺ channels. In the periphery, 5-HT_{2A} receptors are involved in smooth muscle contraction, platelet aggregation, proliferation of arterial fibroblasts, and arterial vasoconstriction. In addition, 5-HT_{2A} receptors are also found in the spinal cord, where they may have a role in pain. The 5-HT_{2B} receptor was first described from the stomach fundus and later was identified in the gut, heart, kidney, liver, and with moderate expression in the lung and cardiovascular tissues. The most important function of this receptor is during development, where it coordinates the proper formation of structures such as the heart and brain.

The 5-HT_{2C} receptor is found in choroid plexus, where it regulates cerebral spinal fluid production and ion exchange between the cerebral spinal fluid and brain. The 5-HT_{2C} receptor is also found in various brain regions such as cortex, amygdala, basal ganglia, hippocampus, and thalamus. 5-HT_{2C} receptors can modulate mesolimbic dopamine neurotransmission, which may be a therapeutic target for treating substance abuse. 5-HT_{2C} receptors are proposed to be involved in anxiety, obesity, and food intake. Moreover, atypical antipsychotic drugs block the activation of 5-HT_{2A} and 5-HT_{2C} receptors indicating that these receptors may be involved in the pathophysiology of schizophrenia.

The Serotonin 3 Receptor Is a Ligand-Gated Ion Channel

The 5-HT₃ receptor is different from the other 5-HT receptors in that it forms an ion channel that regulates the flux of ions. The structure of receptor is a pentamer, similar to the nicotinic acetylcholine receptor. The receptors are found on neurons in the hippocampus, nucleus tractus solitarius, and area postrema, as well as in the periphery. They are located on pre- and postganglionic autonomic neurons and alter GI tract motility and intestinal secretion. When activated by 5-HT, the receptor triggers rapid depolarization due to an inward current by opening a nonselective cation channel (Na⁺, Ca²⁺ influx, and K⁺ efflux). The receptor is possibly involved in nausea, vomiting, and irritable bowel syndrome.

Alternative Splice Variants of Serotonin Receptors

Functional diversity of proteins can be produced by alternative splicing events. Many of the 5-HT receptor genes contain introns that are subject to alternative splicing with the generation of multiple receptor messenger RNAs (mRNAs) encoding slightly different proteins, referred to as isoforms. Seven carboxy-terminal splice variants of the 5-HT₄ receptor have been described. The most interesting feature of these splice variants is the level of constitutive activity of the receptor, which is markedly increased. Constitutive activity is the ability of a receptor to activate second-messenger pathways spontaneously without the binding of an external ligand. Four carboxy-terminal splice variants of the 5-HT₇ receptor have been identified. The functional consequence of these variants is uncertain. An alternatively spliced variant of the 5-HT_{2C} receptor has been described, which encodes a truncated, nonfunctional protein. More work needs to be done to determine the physiological relevance of these RNA-splicing events; nonetheless it is clear that this process leads to additional diversity in 5-HT receptor signaling.

RNA Editing Produces Multiple Functional 5-HT_{2C} Receptor Isoforms

The 5-HT_{2C} receptor undergoes a unique process termed RNA editing to yield multiple receptor variants. RNA editing is an enzymatic reaction that alters nucleotide sequences of RNA transcripts. For the 5-HT_{2C} receptor, five encoded adenosine residues are converted to inosines by double-stranded RNA adenosine deaminases. In the human 5-HT_{2C} receptor, the adenosines within the predicted intracellular second loop can be converted to inosine at the RNA level resulting in multiple mRNA species with the potential to encode 24 different protein variants (Figure 2). The extensively edited isoforms have different abundances in brain tissue, and the translated proteins exhibit different binding properties, and differential activation of second-messenger systems. For example, 5-HT exhibits a decreased potency when activating the fully edited, valine-glycine-valine (VGV), isoform of the human receptor compared with the unedited, isoleucine-asparagine-isoleucine (INI), form and there is a rightward shift in the dose-response curve for phosphoinositide hydrolysis. In addition, editing can alter the ability of 5-HT_{2C} receptors to couple to multiple G-proteins. For example, the nonedited 5-HT_{2C} receptor functionally couples to G_{αq} and G_{α13}, whereas the edited 5-HT_{2C} receptors has loss coupling to G_{α13}. RNA editing may have clinical significance; studies reveal associations between editing profiles and depressed suicide victims, schizophrenia, anxiety, depression, and memory. However, more work needs to be done to confirm a linkage between these clinical findings and 5-HT_{2C} receptor RNA editing.

Single-Nucleotide Polymorphisms Occur in the Serotonin Receptor Family

Normal genetic variations (single-nucleotide polymorphism, SNP) have been identified in almost all 5-HT receptors.

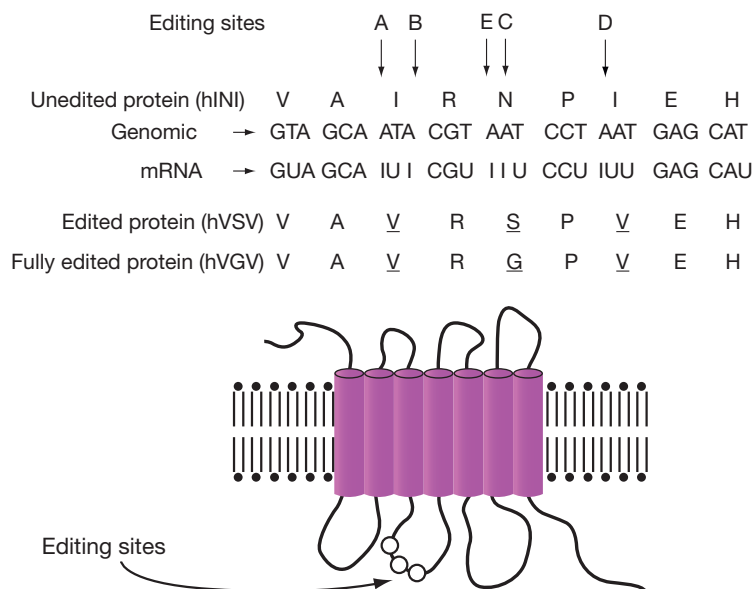


Figure 2 The positions of the editing sites within human 5-HT_{2C} receptors RNA and amino acid sequences are shown for hINI, hVSV, and hVGV edited isoforms of the 5-HT_{2C} receptor. These editing sites are located in the putative second intracellular loop.

Polymorphisms in the coding region of the gene have the potential to alter the receptor's ability to bind ligand, to activate signal transduction pathways, or to adapt to environmental influences. For example, a polymorphism in the amino terminus of the human 5-HT_{1A} receptor attenuates the down-regulation and desensitization produced by the agonist 8-OH-DPAT. A polymorphic variant in the 5-HT_{1B} receptor in the putative third transmembrane domain alters the binding of the antimigraine drug, sumatriptan. Another example is the SNP, H452Y, in the 5-HT_{2A} receptor which appears with the frequency of 9% in the general population and produces a change in the receptor's ability to desensitize after agonist stimulation. This SNP has been associated with attention-deficit hyperactivity disorder and antipsychotic effectiveness. Currently, efforts are being made to link the occurrence of 5-HT receptor polymorphisms with various pathological disorders. Future progress in pharmacogenomics (using genetic information to predict drug response) may potentially lead to better design of serotonergic drugs to reduce side effects and target subpopulations of patients with specific therapies dependent on their genetic profile.

Functional Selectivity and Cross-Talk between 5-HT Receptor Signal Transduction Pathways

Functional selectivity is the ability of a receptor to couple to more than one signal transduction pathway. For example, the 5-HT_{1A} receptor can both inhibit and activate AC. As mentioned above, the primary coupling of this receptor is G_{α_{i/o}} with subsequent inhibition of AC; this has been demonstrated both *in vivo* and in cultured cell systems. However, this receptor has been shown to activate AC, mediated by βγ subunits released from G_{α_{i/o}}, rather than G_{α_s} protein. This activation seems to require high receptor occupancy and high expression

levels in cell-expression systems. In addition, depending on cell type and experimental conditions the 5-HT_{1A} receptor can activate or inhibit PLC. Many studies suggest that the G-protein βγ subunits play a role in such cross-talk between signaling pathways. The 5-HT_{2C} receptor (as well as the 5-HT_{2A} receptor) is another promiscuous receptor that may activate multiple signal transduction pathways. The primary coupling of 5-HT_{2C} receptors is to activate PLC activity via G_{α_q}/G_{α₁₁} proteins to produce IP₃ and DAG. However, this receptor can activate other phospholipases, for example, phospholipase D through G_{α₁₃} proteins and free βγ subunits. Phospholipase D can influence many neuronal functions including endocytosis, exocytosis, vesicle trafficking, and cytoskeletal dynamics. In addition, 5-HT_{2A} and 5-HT_{2C} receptors can activate phospholipase A-2 activity that leads to the release of arachidonic acid and potential formation of bioactive eicosanoids such as the prostaglandins. There is evidence that 5-HT_{2A} and 5-HT_{2C} receptors differentially activate PLC compared to PLA2, dependent on the specific agonists employed. Moreover, there may be differences in the desensitization of these two pathways after repeated agonist treatment. The release of arachidonic acid can result in the activation of K⁺ channels, regulation of neurotransmitter release, and can be a retrograde messenger. Another aspect of functional selectivity is receptor oligomerization, either as homodimers or as heterodimers, which can produce another level of diversity of receptor signaling. It is becoming apparent that specific 5-HT receptors interact with many G-proteins, thereby regulating multiple signal transduction pathways. This diversity may be a target for future therapeutics and interventions.

Summary: Potential Role of Receptor Diversity

There has been much speculation on the origin and significance of the 5-HT receptor diversity. One possible explanation

for the numerous subtypes and variants is that the 5-HT neurotransmitter system emerged early during evolution. Thus, there has been ample time for genetic variation and divergence to occur in the genes encoding for the receptor proteins. Whatever the process of receptor diversity, the multiple receptor signaling has considerable biological significance. For example, inputs to the dorsal raphe nucleus are integrated into global 5-HT release, via both synaptic and volume transmission, which, in turn, can modulate multiple and diverse neuronal functions based on receptor subtype. Thus, one transmitter (5-HT) with multiple receptors/ multiple pathways can introduce levels of complexity, yet specific physiological responses, that are localized to a given brain region or neural system. These responses can be cell-type specific or even cell-compartment specific. Furthermore, a level of fidelity is introduced based on the receptor's relative affinity for 5-HT and localization of 5-HT receptors, synaptic or perisynaptic. The exact nature of this diversity is at present unclear; however, the physiological implications are quite apparent, when considering the diverse role of these receptors in many physiological functions and disease states.

See also: **Molecular Biology:** RNA Editing; **Signaling:** Adenylyl Cyclases; Diacylglycerol Kinases and Phosphatidic Acid Phosphatases; Gi Family of Heterotrimeric G Proteins; G_q Family; Phosphatidylinositol Bisphosphate and Trisphosphate.

Further Reading

- Aghajanian GK and Sanders-Bush E (2002) Serotonin. In: Davis KL, Charney D, Coyle JT, and Nemeroff C (eds.) *Psychopharmacology: The Fifth Generation of Progress*. Philadelphia, PA: Lippincott, Williams, and Wilkins.
- Barnes NM and Sharp T (1999) A review of central 5-HT receptors and their function. *Neuropharmacology* 38: 1083–1152.
- Hoyer D, Clarke DE, Fozard JR, et al. (1994) VII International union of pharmacology classification of receptors for 5-hydroxytryptamine (serotonin). *Pharmacological Reviews* 46: 157–203.
- Millan MJ, Marin P, Bockaert J, and la Cour CM (2008) Signaling at G-protein-coupled serotonin receptors: Recent advances and future research directions. *Trends in Pharmacological Science* 29: 454–464.
- Nichols DE and Nichols CD (2008) Serotonin receptors. *Chemical Reviews* 108: 1614–1641.
- Urban JD, Clarke WP, von Zastrow M, et al. (2007) Functional selectivity and classical concepts of quantitative pharmacology. *Journal of Pharmacology and Experimental Therapeutics* 320: 1–13.

Siglecs

A Varki, University of California, San Diego, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Immunoglobulin superfamily (IgSf) Proteins that have modules homologous to those of antibodies (immunoglobulins). This is an evolutionarily ancient group of proteins whose appearance actually predated the emergence of the immunoglobulins themselves.

I-type lectins Proteins (other than antibodies or T-cell receptors) in which immunoglobulin-like modules mediate binding to glycans (sugar chains).

Sialic acids These acids are a diverse family of nine-carbon acidic sugars that typically occupy a terminal position on glycan chains attached to the cell surface of higher animals of the deuterostome lineage.

Siglecs A major subset of the I-type lectins. Name is based on their defining properties, as sialic acid recognizing IgSf lectins with canonical amino-terminal sequences.

Historical Background and Definition

Sialic acids (Sias) are a family of nine-carbon acidic sugars that typically occupy a terminal position on glycan chains attached to the cell surface of higher animals. The immunoglobulin superfamily (IgSf) is an evolutionarily ancient group of proteins whose appearance predated the emergence of the immunoglobulins themselves. Until the 1990s, it was assumed that IgSf members (other than some antibodies) did not mediate carbohydrate recognition. Independent work on sialoadhesin (Sn, eventually Siglec-1, a protein on certain macrophage subsets) and on CD22 (eventually Siglec-2, a protein on mature B cells) revealed that their first Ig V-set-like domains could mediate Sia recognition. Homologous features of this V-set Ig-like domain and the adjacent C2-set domain then led to the discovery that two other previously cloned molecules – CD33 (eventually Siglec-3) and myelin-associated glycoprotein (MAG, eventually Siglec-4) – also had Sia-binding properties. Following consultation among all researchers working on these proteins, the common name Siglec (suggested by A. Varki) and a numbering system were agreed upon. Criteria for inclusion of IgSf-related proteins as Siglecs are: (1) the ability to recognize sialylated glycans and (2) significant sequence similarity within the N-terminal V-set and adjoining C2-set domains. Evaluation of the human and mouse genomes eventually defined 16 primate and eight mouse molecules that fulfill these criteria. Primates have many more Siglecs than rodents and many of these are rapidly evolving.

Two Broad Subgroups of Siglecs

While Siglecs -1, -2, -4, and -15 appear to be evolutionarily rather conserved, the CD33/Siglec-3-related subgroup (Primate Siglecs -3 and -5–16) are rapidly evolving. Some CD33/Siglec-3-related Siglecs appear to have evolved as hybrids of pre-existing genes and/or by gene conversion. For these reasons, sequence comparisons alone do not allow the conclusive designation of the ortholog status of all genes, and additional features such as gene position and exon structure must be taken

into account. As such issues cannot always be resolved, some mouse Siglecs have been assigned an alphabetical designation.

Common Structural Features

All Siglecs are single-pass type 1 integral membrane proteins with extracellular domains consisting of uniquely similar N-terminal V-set Ig domains, followed by variable numbers of C2-set Ig domains, ranging from 16 in Sn/Siglec-1 to 1 in CD33/Siglec-3. Crystal structures for mouse Siglec-1 and human Siglec-7 indicate (Figure 1) that the V-set immunoglobulin-like fold has several unusual features, including an intra-beta sheet disulfide and a splitting of the standard beta strand G into two shorter strands. These features along with certain key amino acid residues appear to be requirements for Sia recognition. In particular, a conserved arginine residue is involved in a salt bridge with the carboxylate of Sia, in all instances studied to date.

Cell-Type Specific Expression

With the exception of MAG/Siglec-4, Siglec-6, and Siglec-12, expression so far appears to be confined to the hematopoietic and immune systems. Within these systems, each Siglec is expressed in a cell-type specific fashion, suggesting that each may be involved in discrete functions. However, systematic studies of Siglec expression outside the hematopoietic system and during development have not yet been done.

Genomic Organization and Phylogeny

Based on searching for the canonical functional amino acids in the V-set domain of the typical Siglec, there is so far no evidence for Siglec-like molecules in prokaryotes, fungi, or plants, or in animals of the protostome lineage, including organisms for which the complete genome is available. By contrast, it is relatively easy to find Siglec-like V-set domains in many vertebrate taxa. While the conserved Siglecs (-1, -2, -4, and -15) have relatively clear-cut single orthologs that can be identified in

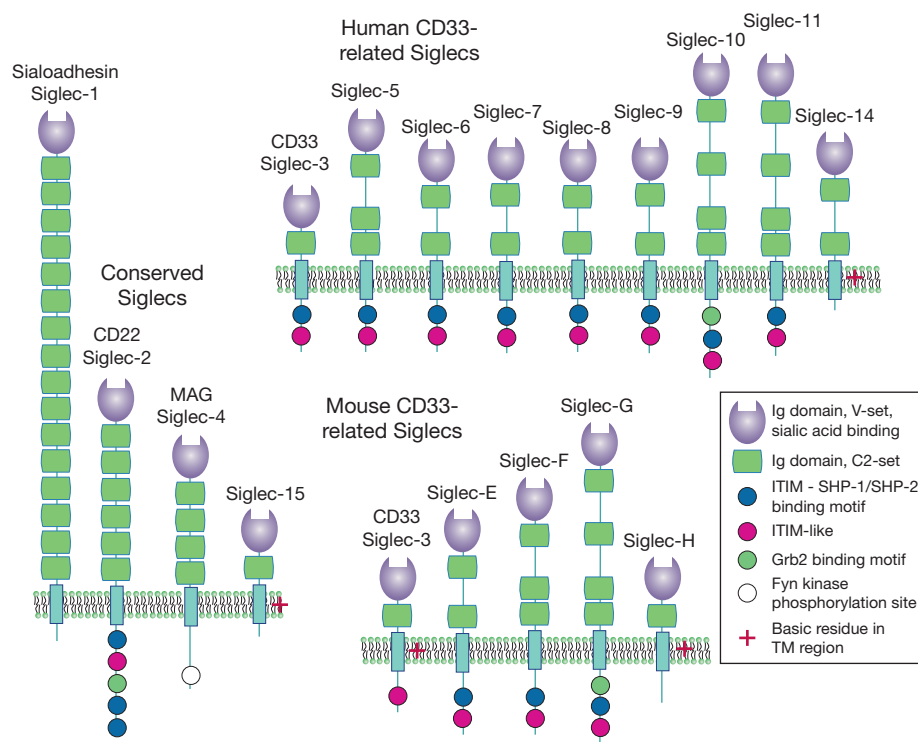


Figure 1 Domain structures of the known Siglecs in humans and mice. There are two main subgroups of Siglecs: one evolutionarily conserved group contains sialoadhesin (Siglec-1), CD22 (Siglec-2), MAG (Siglec-4), and Siglec-15, and the other group contains the rapidly evolving CD33-related Siglecs. In humans, Siglec-12 has lost its arginine residue required for sialic acid binding and Siglec-13 is deleted. ITIM, immunoreceptor tyrosine-based inhibitory motif. The *plus sign* indicates the presence of a charged residue in the transmembrane domain, which has been shown to interact with the immunoreceptor tyrosine-based activatory motif (ITAM)-containing adaptor proteins DAP12 and DAP10. Reproduced with permission from the Consortium of Glycobiology Editors, originally figure 32.1 in Varki A, Cummings RD, Esko JD, et al. (eds.) (2009) *Essentials of Glycobiology*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

most species, the remaining CD33/Siglec-3 related Siglecs appear to have been evolving rapidly. Most of the latter genes are clustered together in an ~500 kb region on human chromosome 19q13.3–13.4 and the syntenic regions of other animals.

Siglec Recognition of Sialic Acids and Their Linkages

The first two Siglecs discovered (Sn/Siglec-1 and CD22/Siglec-2) had strikingly different binding properties for sialosides – with Sn preferring α 2–3 linked targets and CD22 being highly specific for α 2–6 linkages. In the latter case, the binding affinity was in the low micromolar range. MAG/Siglec-4 also has an extended binding site that is even highly specific for the underlying sugar chain. There is also variable preference for certain types of sialic acids, with Sn and MAG not tolerating the common *N*-glycolyl modification of Sias. However, the CD33/Siglec-3-related Siglecs are generally more promiscuous in their preferences for different types and linkages of Sias. Of course, many of the less common linkages and types of sialic acids have not been studied for Siglec recognition. The Golgi enzymes that are potential regulators of Siglec functions are primarily the sialyltransferases and the enzymes which modify sialic acids. Some Siglecs also show preferences for certain macromolecular ligands, for example, CD45 and IgM for CD22/Siglec-2, the mucins CD43, and Muc-1 for

Sn/Siglec-1, CD24 for Siglec-10/H, and specific brain gangliosides for MAG/Siglec-4.

Potential Effects of Neu5Gc Loss on Human Siglec Biology

The most common Sias of mammalian cells are *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc). Humans are a known exception, because of a mutation in cytidine mono-phosphate (CMP)-sialic acid hydroxylase, which occurred after the time (~5–7 Mya) when we shared a common ancestor with great apes. The resulting loss of Neu5Gc and increase in Neu5Ac in humans could have potentially altered the biology of the Siglecs. For example, human cells have a higher density of Sn/Siglec-1 ligands than great apes, the distribution of Sn-positive macrophages in humans is different and a much larger fraction of human macrophages is positive. Other evidence indicates that there are multiple human-specific changes in Siglec biology that may be at least partly related to the loss of Neu5Gc. These include binding changes (in Siglecs -5, -7, -9, -11, -12, and -14); expression pattern changes (in Siglecs -1, -5, -6, and -11); gene conversion (*SIGLEC11*); and deletion or pseudogenization (*SIGLEC13*, *SIGLEC14*, and *SIGLEC16*). Overall, Siglec changes in humans seem to be much more prominent than

in the closely related great apes, suggesting that sialic acid biology was a hot spot in human evolution.

Masking and Unmasking of Siglec-Binding Sites on Cell Surfaces

The initial assumption was that Siglecs were involved in inter-cellular adhesion. However, in most instances, their binding sites appear to be masked by Sias on the same cell surfaces on which they are expressed. Of course, external ligands with high affinity/avidity can still compete for the endogenous masking ligands. There is also some evidence that unmasking can occur under certain conditions, but it is not known if this is biologically relevant. Overall, the significance of Siglec masking is relatively unclear at this time.

Signaling Motifs in Cytosolic Tails

The inhibitory CD33-related Siglecs (Siglecs -3 and -5-12) have conserved tyrosine residues in the cytosolic tails, one of which corresponds to a canonical immunoreceptor tyrosine-based inhibition motif (ITIM). Various *in vitro* manipulations of these receptors indicate that these tyrosines are indeed targets for phosphorylation and that they can modulate signaling events by recruiting certain tyrosine phosphatases such as SHP-1 and SHP-2. However, the true *in vivo* biological functions of these signaling motifs remain somewhat obscure. Another major unresolved question is: what is the connection between extracellular sialic acid recognition and signaling via the cytosolic motifs? Even less is known about the functions of the more recently described activatory Siglecs (Siglecs 13-16), which recruit the DNAX-activating protein (DAP) 12 adaptor along with its canonical immunoreceptor tyrosine-based activating motif (ITAM). It appears likely that these are balancing receptors that generate opposing inhibitory and activatory responses, perhaps in response to sialic acid-expressing pathogens.

Known and Putative Functions of the Siglecs

Various lines of evidence indicate that MAG/Siglec-4 is involved in the maintenance of myelin organization and in the inhibition of neurite outgrowth during regeneration after injury. It is also clear that CD22/Siglec-2 functions as an inhibitory component of the antigen receptor complex of B Cells, and is thus involved in regulating the humoral immune

response. While Sn/Siglec-1 appears to mediate various macrophage adhesion events *in vitro* and *in vivo*, it is still somewhat unclear what exactly the functions of these interactions are. This molecule may also assist macrophages in the phagocytosis of sialic acid-expressing pathogen. Relatively little is known about the functions of CD33-related Siglecs. It appears that these molecules are involved in innate immunity. One hypothesis supported by current data is that Siglecs are sensors for pathogens that have sialylated cell surfaces and/or express extra cellular sialidases. The corresponding endogenous function of the inhibitory CD33-Siglecs appears to be to recognize host Sias as self and dampen unwanted innate immune reactivity. Several findings support this notion. For example, mice deficient in Siglec-F (which is normally found on eosinophils) show hyper-eosinophilia and exaggerated allergic responses, and group B Streptococci engage neutrophil Siglec-9 via their sialylated capsules to dampen innate immune responses. Also, human T and B cells express lower levels of CD33-related Siglecs than in great apes and this correlates with increased responses of the human cells to stimuli. It is suggested (but not proven) that the activatory CD33-related Siglecs (13-16) have opposing effects, via their cytosolic ITAM motifs.

See also: [Lipids Carbohydrates Membranes and Membrane Proteins: Lectins](#); **Signaling:** [Immunoglobulin \(Fc\) Receptors](#).

Further Reading

- Crocker PR, Clark EA, Filbin M, et al. (1998) Siglecs: A family of sialic-acid binding lectins [letter]. *Glycobiology* 8: v2.
- Crocker PR, Mucklow S, Bouckson V, et al. (1994) Sialoadhesin, a macrophage sialic acid binding receptor for haemopoietic cells with 17 immunoglobulin-like domains. *EMBO Journal* 13: 4490-4503.
- Crocker PR, Paulson JC, and Varki A (2007) Siglecs and their roles in the immune system. *Nature Reviews Immunology* 7: 255-266.
- Kelm S, Pelz A, Schauer R, et al. (1994) Sialoadhesin, myelin-associated glycoprotein and CD22 define a new family of sialic acid-dependent adhesion molecules of the immunoglobulin superfamily. *Current Biology* 4: 965-972.
- Powell LD, Sgroi D, Sjöberg ER, Stamenkovic I, and Varki A (1993) Natural ligands of the B cell adhesion molecule CD22beta carry N-linked oligosaccharides with alpha-2, 6-linked sialic acids that are required for recognition. *Journal of Biological Chemistry* 268: 7019-7027.
- Powell LD and Varki A (1995) I-type lectins. *Journal of Biological Chemistry* 270: 14243-14246.
- Varki A (2010) Colloquium paper: Uniquely human evolution of sialic acid genetics and biology. *Proceedings of the National Academy of Sciences of the United States of America* 107(supplement 2): 8939-8946.
- Varki A and Angata T (2006) Siglecs – the major subfamily of I-type lectins. *Glycobiology* 16: 1R-27R.

Sigma Factors

J D Helmann, Cornell University, Ithaca, NY, USA

S R MacLellan, University of New Brunswick, Fredericton, NB, Canada

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by John D. Helmann, volume 4, pp. 41–44, © 2004, Elsevier Inc.

Glossary

Alternative σ A σ -factor that can replace the primary σ -factor in response to a particular stress or signal to activate the expression of a defined target regulon.

Promoter DNA sequences necessary for the accurate initiation of transcription by RNA polymerase.

Regulon The set of genes and operons that are regulated by a common factor.

RNA polymerase Enzyme that copies DNA sequences into the complementary RNA sequence.

Introduction

DNA serves as a storehouse for the genetic information of each organism, but is itself a silent repository. Only upon copying of DNA segments into RNA (transcription), and translation into protein, is the genetic potential of each organism (and each cell type in multicellular organisms) able to be realized. For example, our skin, blood, hair, liver, and lung cells all contain the same DNA, yet they have differentiated into functionally distinct cell types based largely on the selective expression of a small subset of the genes contained within each cell. Similarly, comparative genomics has led to the realization that the evolutionary processes that have led to the divergence of humans and mice, for example, have more to do with changes in 'how' genes are regulated than with 'which' genes are actually present in each organism. These examples highlight the importance of gene regulation: those mechanisms that selectively activate and repress different genes depending on cell type and local environment.

Gene Regulation in Bacteria

Most of the principles of gene regulation were first developed from studies of bacterial cells and, in particular, *Escherichia coli*. The first step in gene expression is the transcription of the genetic code from DNA into an RNA copy that usually serves to direct protein synthesis (translation). RNA polymerase (RNAP) is the enzyme that copies (transcribes) DNA into RNA.

In bacterial cells, most gene regulation occurs at the level of transcription. Genes that are expressed are actively transcribed and translated into their corresponding protein products. Conversely, genes that are silent are not copied into RNA. Thus, factors that act to control the activity of RNAP are central to many processes of gene regulation. These regulators include activator proteins, repressor proteins, and σ -factors. In order to appreciate how these factors can control gene expression, the biochemical properties of RNAP and the overall features of the transcription cycle are discussed first, followed by σ -factors.

Bacterial RNA Polymerase and the Transcription Cycle

In bacteria, RNAP is a multisubunit protein complex and contains both a catalytic core enzyme and a specificity subunit known as σ . The core enzyme (containing the β , β' , two α , and the ω polypeptide chains) is able to catalyze the synthesis of RNA directed by a DNA template. Core enzyme associates reversibly with another polypeptide chain, σ , to form the holoenzyme.

RNAP copies particular DNA segments into RNA chains by first initiating RNA synthesis at a promoter site, elongating the RNA chain, and finally terminating transcription (the transcription cycle). In bacteria, many protein-coding genes are organized into cotranscribed groups called operons. In these cases, each messenger RNA (mRNA) product can direct the synthesis of multiple protein products.

The first step in the transcription cycle is promoter site localization and requires the presence of the σ -subunit: only holoenzyme can recognize promoter sites. Promoter sites are typically characterized by defined DNA sequences located just upstream of the site where RNA synthesis begins (designated the +1 site). For example, sequences centered ~10 bp upstream often have a sequence, designated the '-10 region', similar to 5'TATAAT3'. A different conserved sequence (TTGACA) is usually located between 16 and 18 bp further upstream and is designated the '-35 region'. Together, the -35 and -10 sequences provide the necessary information to allow RNAP to identify a promoter. Specific interactions between the σ -subunit of holoenzyme and the -35 and -10 elements allow RNAP to tightly engage the promoter region. Once bound, RNAP interacts with an extended region of DNA, ultimately contacting the DNA over a region of 60 or more base pairs (from -40 to +20).

Once bound at a promoter site, RNAP melts the DNA over approximately one turn of the helix (from about -10 to +2) to expose the template strand of the DNA. The process of converting the double-stranded promoter region DNA into a locally melted structure (called the open complex) requires the action of σ -factor to initiate the melting process and is then driven by the association of the template strand of the DNA with the RNAP active site (comprised largely of the β - and β' -subunits). The open complex is then poised to initiate

synthesis of an RNA chain. The region of DNA strand separation is often referred to as the transcription bubble, to distinguish it from the otherwise double-stranded DNA in the surrounding regions of the genome.

The process of transcription elongation involves the assembly of ribonucleoside triphosphate precursors into an RNA chain containing the sequence complementary to the template DNA. During transcription elongation, the DNA template is threaded through the active site of RNAP and, as a result, the transcription bubble is propagated down the DNA with melting of the DNA occurring as the DNA enters the active site and reannealing of the template occurring as the DNA exits the RNAP. The complex containing the DNA template, RNAP, and the growing RNA chain is often referred to as the ternary complex. As the growing RNA chain emerges from RNAP, it is thought to displace the σ -subunit, and the process of transcription elongation is continued by the core RNAP subunits. Elongating RNAP sometimes associates with additional protein factors known as elongation factors which can modulate its properties. In general, elongation is highly processive: the RNAP that begins an RNA chain continues synthesis until the chain is terminated.

The process of transcription termination, like that of initiation, is critical for the controlled expression of genetic information. At the end of many transcription units the DNA sequence encodes a G:C-rich stem-loop structure in the RNA chain followed by a short region rich in uridine. This structure functions as a transcription terminator and interacts with the core RNAP leading to dissociation of the ternary complex and release of the terminated RNA chain. In other cases, protein factors such as the rho protein can bind to unstructured regions of RNA and trigger transcription termination. Once released by the termination reaction, the core RNAP must rebound to a σ -subunit to reform holoenzyme and thereby complete the transcription cycle.

Since bacteria lack a nuclear membrane, the processes of transcription and translation can be coupled: ribosomes can bind to the mRNA chain as it emerges from RNAP and immediately initiate protein synthesis. Indeed, along actively transcribed genes several RNAP molecules may be simultaneously engaged in RNA chain elongation with each RNA chain in turn bound by one or more ribosomes engaged in protein synthesis.

Biochemical Properties

As expected for a critical component of the transcription apparatus, σ -factors have been highly conserved through evolution. Indeed, σ -factors from one bacterium can function, at least in some cases, with the core RNAP from distantly related species. In general, all σ -factors share certain defining properties. They all bind to the core RNAP, apparently at a common binding site, to form a holoenzyme. The presence of σ -factor determines the sequence of the promoters that can be bound by the corresponding holoenzyme. The σ -subunit, in those cases that have been studied, also plays a role in the initiation of the DNA-melting step that is an obligatory prelude to transcription initiation. Often, although perhaps not always, σ is released from the ternary complex during the process of RNA chain elongation.

Bacterial σ -factors can be divided, based in part on their protein sequences, into two families: the σ^{70} and the σ^{54} families. All bacteria contain at least one member of the σ^{70} family that is required for the transcription of those genes essential for growth under virtually all conditions (so-called 'housekeeping' functions). This is referred to as the primary (or class 1) σ -factor. In addition, most bacteria contain at least one (and as many as 50 or more) alternative σ -factors (Table 1). These factors typically control genes needed for specialized functions that may only be expressed under particular growth conditions. Examples include genes needed for stress responses, motility, sporulation, uptake or transport of specific nutrients, or antibiotic production.

Most alternative σ -factors are structurally related to the primary (class 1) σ -factors and are therefore considered to be members of the σ^{70} family. Some alternative σ -factors (class 2) are very similar to the primary σ -factors but are, however, dispensable for growth. One example is the *E. coli* σ^S (RpoS) protein that becomes active in stationary phase cells. A much larger class of alternative σ -factors are those more distantly related in sequence to the primary σ -factor, and often lacking one or more conserved regions characteristic of the class 1 and 2 proteins. These alternative σ -factors (classes 3, 4, and 5) control a wide range of physiological processes and are often only active under very specific growth conditions. Prominent among these alternative σ -factors are the numerically abundant

Table 1 σ -Factors in *E. coli* and *B. subtilis*

Organism	σ	Gene	Function
<i>E. coli</i>	σ^{70}	<i>rpoD</i>	Housekeeping genes
	σ^H	<i>rpoH</i>	Heat shock
	σ^E	<i>rpoE</i>	Extreme heat shock, periplasmic stress (ECF ^a)
	σ^F	<i>flaA</i>	Flagellar-based motility
	σ^S	<i>rpoS</i>	Stationary phase adaptations
	σ^N	<i>rpoN</i> , <i>glnF</i>	Nitrogen-regulated genes (σ^{54} family)
	σ^{fecI}	<i>fecI</i>	Ferric citrate uptake (ECF)
	σ^A	<i>sigA</i>	Housekeeping genes
	σ^B	<i>sigB</i>	General stress response
	σ^D	<i>sigD</i>	Flagellar-based motility, autolysins
<i>B. subtilis</i>	σ^E	<i>sigE</i>	Sporulation, early mother cell
	σ^F	<i>sigF</i>	Sporulation, early forespore
	σ^G	<i>sigG</i>	Sporulation, late forespore
	σ^H	<i>sigH</i>	Competence and early sporulation
	σ^K	<i>sigK</i>	Sporulation, late mother cell
	σ^L	<i>sigL</i>	Degradative enzymes (σ^{54} family)
	σ^I	<i>ykoZ</i>	Unknown
	σ^{Xpf}	<i>xpf</i>	PBSX regulation
	$\sigma^{YvrI/Ha}$	<i>yvrI</i> , <i>yvrHa</i>	Two-subunit σ , oxalate decarboxylase expression
	σ^X	<i>sigX</i>	Cell envelope modifications (ECF)
	σ^W	<i>sigW</i>	Antibiotic stress responses (ECF)
	σ^M	<i>sigM</i>	Cell wall stresses (ECF)
	σ^V	<i>sigV</i>	Unknown (ECF)
	σ^Y	<i>sigY</i>	Unknown (ECF)
	σ^Z	<i>sigZ</i>	Unknown (ECF)
	σ^{ylaC}	<i>ylaC</i>	Unknown (ECF)

^aECF indicates a factor that is a member of the large extracytoplasmic function family of alternative σ -factors.

extracytoplasmic function (ECF) σ -factors (class 4 proteins) that appear, in general, to control functions related to the cell surface. Finally, many bacteria have alternative σ -factors with a distinctly different protein sequence that defines the σ^{54} family. Unlike members of the σ^{70} family, the σ^{54} -factors form holoenzymes that almost always require adenosine triphosphate (ATP)-dependent activator proteins to activate their target promoters. The prototype of the σ^{54} -family is the *E. coli* σ^{54} (RpoN) protein that is important for nitrogen regulation of gene expression (Table 1). Similar proteins in other bacteria control a variety of functions.

The division of σ^{70} -family members into different structural classes is based in large part on the comparison of their protein sequences. All primary (class 1) σ -factors (and the closely related class 2 factors) have four conserved regions. Region 1 contains an autoinhibitory domain that prevents the free σ -factor from interacting with DNA. This inhibition is relieved when σ associates with core RNAP. Region 2 contains the key determinants for recognition of the -10 promoter element and for initiation of DNA melting. Region 3 functions as a spacer domain between regions 2 and 4 and is important for transcription from some promoters. Region 4 contains the elements that recognize the -35 promoter element and also serves as a site of contact for some transcription activator proteins. Recent advances in the determination of three-dimensional structures for both σ -factors and holoenzyme have made it clear that these protein sequence regions correspond to folded domains in the σ -structure. The two domains that are required for recognition of promoters, domains 2 and 4, are common to all σ^{70} -family members. By contrast, much of domains 1 and 3 are often deleted in alternative σ -factors.

Conventional σ -factors are single polypeptides but two-subunit forms have been recently described. The T4 phage protein gp55 is a truncated highly divergent σ^{70} sigma factor that along with protein gp33 binds to *E. coli* core polymerase and activates transcription of T4 late genes. gp33 is not obviously similar to σ -factor but binds the flap region of the core β subunit and functions analogously to σ region 4. In *Bacillus subtilis*, two proteins called YvrI and YvrHa (Table 1) also bind core polymerase subunits to activate transcription. Both of these proteins are related to σ -factor and represent the first known two-subunit bacterial σ -factor system.

Roles of σ in Gene Regulation

Just as DNA-binding transcriptional repressors and activators can control the activity of RNAP at promoter sites, alternative σ -factors provide a powerful avenue for the activation of specific sets of genes (the σ -factor regulon) during cell differentiation or in response to various stress conditions. The activity of alternative σ -factors can, in turn, be controlled in numerous ways. In some cases, the σ -factor is only synthesized under very specific conditions. In other cases, the σ -factor is synthesized but is maintained in an inactive state by binding to an inhibitory protein (called an anti- σ -factor) or it is synthesized as an inactive precursor protein that only becomes active when an inhibitory peptide is cleaved by proteolysis. Anti- σ -factors are, in many cases, associated with the cell membrane and thereby able to perceive signals from the external environment. In any

event, once the active σ -factor is synthesized or released, it can compete with other σ -factors for binding to core RNAP and thereby redirect some of the RNAP to its cognate target promoters.

Roles for alternative σ -factors in bacterial gene regulation are legion. For example, alternative σ -factors play key roles in the expression of heat shock genes, iron uptake systems, the developmental program of sporulation (in Gram-positive bacteria), synthesis of flagella and chemotaxis machinery, and various stress responses. A summary of the roles of σ -factors in the two best-studied bacterial systems is presented in Table 1.

Phylogenetic Distribution

σ -Factors are always found associated with bacterial RNAP but have been functionally replaced by other proteins in the more complex RNAP characteristic of eukaryotes and Archaea. However, σ -like factors are found in the chloroplasts of photosynthetic eukaryotes which are descended from a cyanobacterium. In contrast to chloroplasts, the transcription apparatus in present-day mitochondria is most similar to certain bacteriophage-encoded, single-subunit RNAPs and is unlike the multisubunit RNAP of either bacteria or eukarya.

Transcription is a process fundamental to all life forms. Core RNAPs from bacteria, archaea, and eukaryotes are related to one another in terms of sequence and structure prompting searches for evolutionary conserved σ -like proteins among these three domains. In bacteria, σ -factor acts as a specificity factor required for promoter recognition by holoenzyme but proteins homologous to σ -factor have not been identified in either the archaea or eukaryotes (with the exception of chloroplast organelles). In the eukaryotes, the initial transcription complex consists of RNAP (PolII), transcription factor IIB (TFIIB), the TATA-box binding protein TBP, and other general transcription factors. In the archaea, polymerase, the TFIIB homolog TFB, and TBP are sufficient for the initiation of transcription. Recent investigation into the structure of the yeast PolII–TFIIB complex suggests that TFIIB may represent the long-sought analog to σ -factor in eukaryotes and archaea. A region of TFIIB (called B ribbon) contacts the flap region of the β subunit as does region 4 of σ -factor. Another region called B linker binds the β' clamp domain as does region 2 of σ . In both yeast and bacteria, these contacts are required for transcription activation. In terms of sequence, overall structure, and the orientation of contacts between σ /TFIIB and core polymerase, it does not appear that σ -factor and TFIIB are evolutionarily related but may instead represent an example of convergent evolution.

See also: **Molecular Biology:** DNA Modification, Restriction Endonucleases: Type I Enzymes; DNA Polymerase I, Bacterial; RNA Polymerase Reaction in Bacteria; T7 RNA Polymerase; **Signaling:** Dopamine Receptors.

Further Reading

Borukhov S and Nudler E (2003) RNA polymerase holoenzyme: Structure, function and biological implications. *Current Opinion in Microbiology* 6: 93–100.

- Borukhov S and Severinov K (2002) Role of the RNA polymerase sigma subunit in transcription initiation. *Research in Microbiology* 153: 557–562.
- Gruber TM and Gross CA (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annual Review of Microbiology* 57: 441–466.
- Guisbert E, Yura T, Rhodius VA, and Gross CA (2008) Convergence of molecular, modeling, and systems approaches for an understanding of the *Escherichia coli* heat shock response. *Microbiology and Molecular Biology Reviews* 72: 545–554.
- Heinrich J and Wiegert T (2009) Regulated intramembrane proteolysis in the control of extracytoplasmic function sigma factors. *Research in Microbiology* 160: 696–703.
- Helmann JD (2002) The extracytoplasmic function (ECF) sigma factors. *Advances in Microbial Physiology* 46: 47–110.
- Kazmierczak MJ, Wiedmann M, and Boor KJ (2005) Alternative sigma factors and their roles in bacterial virulence. *Microbiology and Molecular Biology Reviews* 69: 527–543.
- Kroos L (2007) The *Bacillus* and *Myxococcus* developmental networks and their transcriptional regulators. *Annual Review of Genetics* 41: 13–39.
- Paget MS and Helmann JD (2003) The sigma70 family of sigma factors. *Genome Biology* 4: 203.
- Sachdeva P, Misra R, Tyagi AK, and Singh S (2010) The sigma factors of *Mycobacterium tuberculosis*: Regulation of the regulators. *FEBS Journal* 277: 605–626.
- Staroń A, Sofia HJ, Dietrich S, et al. (2009) The third pillar of bacterial signal transduction: Classification of the extracytoplasmic function (ECF) sigma factor protein family. *Molecular Microbiology* 74: 557–581.
- Wigneshweraraj S, Bose D, Burrows PC, et al. (2008) Modus operandi of the bacterial RNA polymerase containing the sigma54 promoter-specificity factor. *Molecular Microbiology* 68: 538–546.
- Wosten MM (1998) Eubacterial sigma-factors. *FEMS Microbiology Reviews* 22: 127–150.

Sliding Clamps in DNA Replication: *Escherichia coli* β -Clamp and PCNA Structure

M R Marzahn and L B Bloom, University of Florida, Gainesville, FL, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Eric R. Goedken and John Kuriyan, volume 4, pp. 45–47, © 2004, Elsevier Inc.

Glossary

ATPase A class of enzymes that hydrolyzes adenosine triphosphate (ATP) into adenosine diphosphate (ADP) as an energy source for catalyzing another reaction.

Clamp loader A protein complex that functions to load sliding clamps onto DNA. The prokaryotic clamp loader is gamma complex and the eukaryotic clamp loader is replication factor C (RFC).

DNA polymerase An enzyme whose primary function is to add complementary 2'-deoxyribonucleotides to a DNA strand in a template-directed reaction. In *E. coli*, the primary replicative polymerase is Pol III while the eukaryotic replicative polymerases are DNA polymerases α , δ , and ϵ .

Fluorescence resonance energy transfer (FRET) Also known as resonance energy transfer (RET), this is a mechanism that describes the transfer of energy between two chromophores through nonradiative dipole–dipole coupling.

Processivity In terms of DNA replication, the number of nucleotides added by a DNA polymerase per binding event to the DNA.

Replisome An assembly of proteins and enzymes required for DNA replication.

Sliding clamp A ring-shaped protein complex that serves as a processivity factor for DNA polymerases. The prokaryotic clamp is β -clamp and the eukaryotic clamp is proliferating cell nuclear antigen (PCNA).

Introduction

DNA replication is a process that is essential for both cellular maintenance of the genome and reproduction of this genome from generation to generation. The need to transmit essentially error-free copies of genetic information to succeeding generations has given rise to complex mechanisms for the replication and repair of the genome of even the simplest organisms. Though the specific players within these processes vary from organism to organism, the essential mechanisms are quite similar and in most cases proteins are homologous.

At the most basic level, new DNA is synthesized by DNA polymerases, which are enzymes that incorporate nucleotides to generate a new DNA strand complementary to the original cellular template. These enzymes, along with their accessory proteins that complete the replisome, must possess fidelity and high processivity. Fidelity is achieved both through the base pairing of incoming nucleotides with the DNA template inside the enzyme active site and the proofreading capabilities that allow some DNA polymerases to correct errors. Processivity is achieved by a unique component of the replisome: the sliding clamp. Sliding clamps form ring-shaped complexes that encircle DNA and bind to DNA polymerases. They are fairly well conserved from prokaryotes to eukaryotes and are essential for efficient DNA replication.

Clamp Configuration and Function

Processivity is defined as a polymerase's ability to remain associated with the template and continue adding nucleotides to the growing molecule. It has been shown that polymerases alone, specifically pol III from *Escherichia coli*, can only incorporate tens of nucleotides for each DNA binding event before

dissociating from the template. The addition of a sliding clamp to this reaction confers essentially unlimited processivity to the polymerase, increasing to tens of thousands of nucleotides per binding event.

Solution of the crystal structure of the *E. coli* processivity factor revealed a mechanism by which this protein could so dramatically increase the processivity of the DNA polymerase. Two β monomers form a ring-shaped complex with an inner diameter large enough to encircle duplex DNA. This processivity factor, and others whose structures have since been solved, is now called 'sliding clamps' to illustrate their function. When the polymerase binds a sliding clamp, it is effectively tethered to DNA but capable of moving at rates limited by the rate of DNA synthesis.

Although these clamps vary greatly at the amino acid level for prokaryotes and eukaryotes, structurally they share many characteristics. The clamp from *E. coli*, called the β -clamp and encoded by the *dnaN* gene, is a homodimer of polypeptides each 366 amino acids in length (Figure 1(a)). The sliding clamp in humans and yeast is called proliferating cell nuclear antigen (PCNA). *Saccharomyces cerevisiae* PCNA is a homotrimer of polypeptides each 258 amino acids in length encoded by the *POL30* gene (Figure 1(b)).

Even though these clamps differ in the number of monomers used to construct the active oligomer and differ in amino acid sequences, both clamps are of similar size and shape (Figures 1(c) and 1(d)). As seen in Figure 1, both rings have six globular domains with an inner ring of α -helices and an outer ring of β -sheet clusters. These domains give the structures a pseudo-sixfold axis of symmetry that runs through the center of the clamp.

When the structures are superimposed, the two rings appear to have similar inner and outer diameters (Figure 1(c)) and ring thickness (Figure 1(d)). Both rings are approximately 30 Å

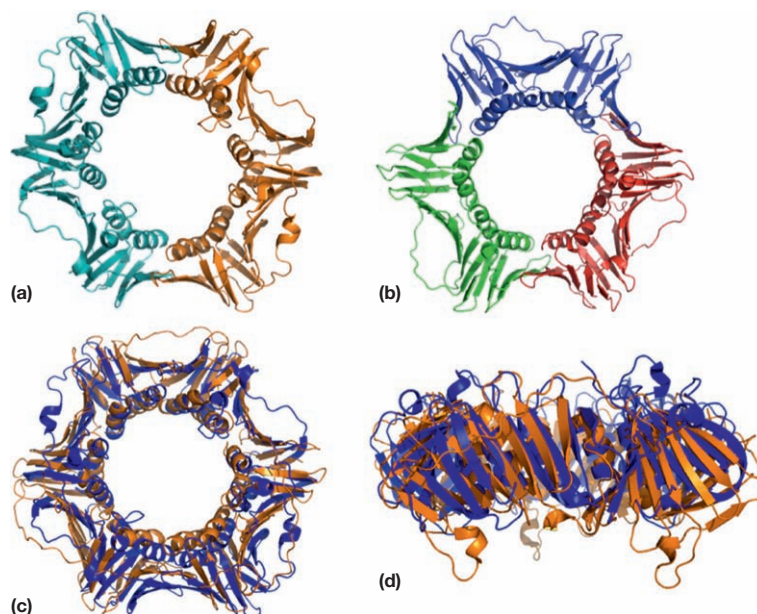


Figure 1 Ribbon diagrams of (a) *E. coli* β -clamp (PDB code 3BEP), (b) *S. cerevisiae* PCNA (PDB code 3K4X). Each monomer is shown in a different color with α -helices represented by coils and β -strands by arrows. A superimposition of the two cartoons in (a) and (b) is shown from a top (c) and side (d) view. In the superimpositions, the β -clamp is shown in blue and PCNA in orange. All images were created using MacPyMol2006.

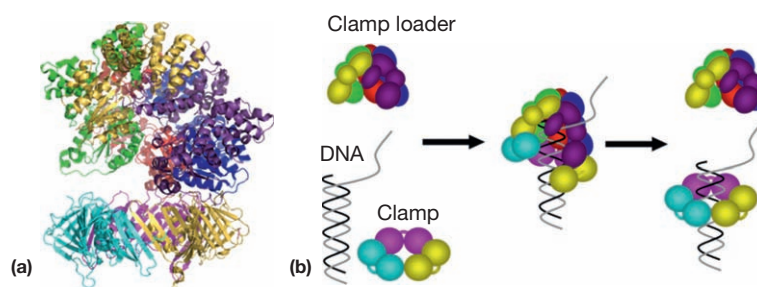


Figure 2 (a) Ribbon diagram of *S. cerevisiae* PCNA bound to RFC (PDB code 1SXJ). In the ribbon diagram, α -helices are represented by coils and β -stands by arrows. Image was created using MacPyMol2006. (b) Schematic diagram of the clamp-loading reaction in the presence of ATP based upon the structure in (a). Corresponding RFC and PCNA subunits are shown in the same colors as the ribbon diagram, while the DNA is represented by a black and gray helix.

in thickness, which would cover almost a helical turn of the DNA duplex when the clamp is present on DNA. The outer diameters differ slightly with β -clamp at about 80 Å and PCNA at around 90 Å (measurements taken using PDB codes 3BEP and 3K4X with MacPyMol2006). The inner diameter of the rings, approximately 40 Å, provides more than enough space to accommodate a DNA helix, which is 20 Å in diameter.

Clamp/Clamp-Loader Interactions

In solution, sliding clamps generally exist as closed ring structures. To encircle DNA and allow for polymerase tethering, the clamp must be opened and loaded onto DNA. In both prokaryotes and eukaryotes, the clamp is positioned onto DNA by a protein belonging to the AAA+ family of adenosine triphosphatases (ATPases) known as the clamp loader. In *E. coli*, the clamp loader is called the gamma or DnaX complex, and in

S. cerevisiae is called replication factor C (RFC). Both complexes contain an oligomer of five core subunits that serve to bind to the clamp, open it, bind to DNA, and release the closed clamp onto the DNA. This release of the clamp onto DNA is accompanied by adenosine triphosphate (ATP) hydrolysis. These subunits are arranged in a ring somewhat like slices of an orange. Each subunit contains three globular domains that are connected by flexible linkers. The C-terminal domains form a tight ring or collar to hold the subunits in the complex. The complex is more open on the side containing the N-terminal domains, creating a space for DNA to bind. A cartoon schematic of this reaction in the presence of ATP is shown in [Figure 2\(b\)](#).

The crystal structure of the *S. cerevisiae* clamp loader RFC bound to PCNA (PDB code 1SXJ) provides some insight into how these proteins interact and allow for loading of the clamp onto DNA. The individual proteins that comprise RFC are arranged in a spiral manner above the center of the PCNA

ring (Figure 2(a)). DNA modeled into the structure suggests that the subunits spiral with the same pitch as the double helix and that this spiral arranges the ATP-binding sites within RFC so that DNA binding can initiate ATP hydrolysis and the release of the clamp from RFC, loading it onto DNA. Additionally, this spiral may also facilitate the opening of the clamp wherein the clamp adopts a lock washer-like conformation (center cartoon of Figure 2(b)) prior to loading onto DNA. Though this open conformation of the clamp/clamp loader complex has not been trapped in a crystal structure, independent studies utilizing fluorescence energy transfer (FRET) and electron microscopy have shown this conformation to exist *in vitro*.

In the crystal structure shown in Figure 2(a), only three subunits of RFC (those colored purple, blue, and red) make contacts with PCNA. PCNA has three hydrophobic pockets on the surface that would allow for protein–protein interactions, and two of these are bound by RFC. The binding interface between RFC and PCNA is similar to that of other proteins bound to clamps in that hydrophobic loops on two of the RFC subunits (colored purple and red in Figure 2) are inserted into the hydrophobic pockets, providing a stable interaction between the two complexes.

Clamp/Polymerase Interactions

After a sliding clamp is loaded onto DNA, a DNA polymerase binds the clamp and this interaction increases the residence time of the polymerase with the DNA template being copied. DNA polymerases interact with the same hydrophobic pockets on the face of the clamp to which the clamp loaders bind. DNA polymerases and clamp loaders contain a conserved sequence motif that typically contains two aromatic residues, often Phe and Tyr, or a hydrophobic and aromatic residue such as Leu and Phe. These residues bind the hydrophobic pocket. In eukaryotes, this sequence motif is referred to as the PCNA-interacting peptide (PIP). Not only do DNA polymerases bind the clamp, other enzymes such as DNA ligase, flap endonuclease, and chromatin assembly factor, required for DNA replication and repair, also interact with the clamp. These proteins also contain the conserved PIP sequence motif.

During replication, specialized high-fidelity DNA polymerases catalyze the bulk of DNA synthesis, but there are times when other DNA polymerases must take over. One such time is when the replicative polymerase encounters

DNA damage and specialized lesion bypass polymerases take over to extend the DNA past the site of damage. There is evidence that these DNA polymerases make contacts with the clamp in addition to those with the hydrophobic pocket. These other interactions may allow multiple DNA polymerases to bind to a clamp so that when a specialized polymerase is needed, it is physically present within the replisome. In eukaryotic cells, posttranslational modifications of PCNA provide an additional level of regulation of interactions with specialized DNA polymerases and DNA repair proteins.

Clamp/DNA Interactions

Biochemical studies, in addition to the unique ring structure and inner diameter described above for β -clamp and PCNA, provide evidence that these oligomers do indeed encircle DNA to create a tethering system to keep the polymerase bound to the DNA template. Though no structure has been solved to date with all components required for clamp loading (clamp/clamp loader/DNA) or DNA replication (clamp/polymerase/DNA) structures of both β -clamp and PCNA have been solved in complexes with DNA (Figures 3(a) and 3(b), respectively). In both structures, the DNA helix appears to be making strong electrostatic contacts with residues in the α -helices in the center of the clamps. However, it is obvious from Figure 3 that the DNA helix makes quite different contacts with the inner surfaces of the β -clamp and PCNA. Furthermore, biochemical analysis using mutagenesis has implicated additional amino acids as participating clamp/DNA interactions. Together, these results suggest that the DNA helix may adopt many conformations within the clamp ring. There are many positively charged residues within α -helices on the inner surfaces of the clamps, which would allow for strong interactions between the clamp and the negatively charged DNA phosphodiester backbone (Figure 3(c)).

Before these clamp–DNA structures were solved, DNA was modeled through the center of the clamp such that the axis of the helix was perpendicular to the plane of the ring. X-ray crystallographic, electron microscopy, and mutagenesis studies have indicated that this assumption is unlikely to be true. Mutagenesis studies have found specific amino acids within the center of the ring that are necessary for clamp–DNA interactions. An archaeal sliding clamp/clamp loader/DNA complex structure has been solved in which the clamp is in an open conformation with the DNA bound near the open

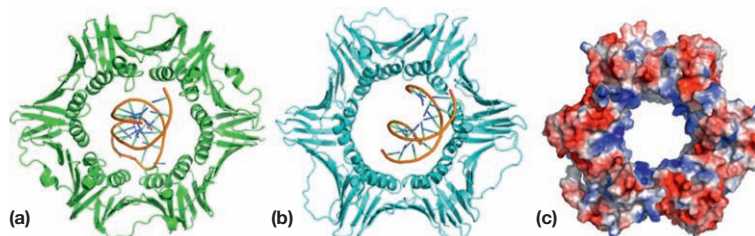


Figure 3 Ribbon diagrams of (a) β -clamp (PDB code 3BEP) and (b) PCNA (PDB code 3K4X) bound to a DNA duplex. In each diagram α -helices are represented by coils and β -strands by arrows. DNA is shown as a sticks with the phosphodiester backbone in orange and the bases modeled as blue and green sticks. PCNA (PDB code 3K4X) surface charge representation (c) is shown in the same orientation as (b). Red indicates negative charge, blue positive charge, and white hydrophobic areas. Images were created using MacPyMol2006.

interface at an angle. In the crystal structures described above, it appears that DNA makes strong electrostatic interactions with the clamp that could potentially inhibit the rate of DNA synthesis when a polymerase is bound. It has been suggested that polymerase binding may alter the angle between the axis of DNA and the clamp so that the clamp can slide freely along DNA, limited only by the rate of synthesis itself. Taken together, the data suggest that the clamp actually functions by making an angle of 20° with the axis of the clamp, which seems to provide not only a strong interaction with the DNA but also a level of regulation in terms of loading the clamp and its subsequent function during replication.

Processivity Factors in Other Species

Though β -clamp (from *E. coli*) and PCNA (from *S. cerevisiae* and *Homo sapiens*) are the most commonly studied sliding clamps, processivity factors have been found and characterized in many other species as well. Some of these factors, such as a clamp involved in eukaryotic DNA damage response and clamps from various species of archaea and bacteriophages, have structures similar to either β -clamp or PCNA while others, such as that utilized by herpes simplex virus (HSV), adopt quite different structures. Though there is little to no sequence homology between them, each of these processivity factors shares some structural and functional characteristics with β -clamp and PCNA.

H. sapiens PCNA

PCNA from *H. sapiens* is a homotrimer of subunits encoded by the PCNA gene with each subunit 261 amino acids in length. *H. sapiens* PCNA has a structure very similar to that of *S. cerevisiae* PCNA, as shown by the superimposition of the two structures in Figure 4(a). Both rings are homotrimers and in addition to similar shape also have almost identical inner and outer diameters, thickness, and surface-charge distribution. In spite of the lack of sequence homology, the three-dimensional structural similarity between these two clamps

has allowed for studies involving the *S. cerevisiae* clamp, which is generally easier to work within the laboratory from a genetic standpoint, to provide insight into the machinery involving the *H. sapiens* sliding clamp.

H. sapiens 9–1–1 Clamp

In *H. sapiens* and yeast, a second clamp, the 9–1–1 complex, has been identified. This clamp is involved in the cellular response to DNA damage (Figure 4(b)). Though this clamp shares structural similarities in size and surface-charge distribution with PCNA (compare the structures in Figures 4(a) and 4(b)), it is a heterotrimer composed of the proteins RAD9 (encoded by the RAD9 gene and 396 amino acids in length), RAD1 (encoded by the RAD1 gene and 286 amino acids in length), and HUS1 (encoded by the HUS1 gene and 283 amino acids in length). The homologous yeast clamps are also heterotrimers, but the structures have not yet been solved. Even though the amino acid sequences and structural organization (homotrimer vs. heterotrimer) of PCNA and the 9–1–1 clamp are radically different, the two adopt very similar tertiary structures.

Archaeal Processivity Factors

Although species within the domain Archaea were once grouped with bacteria under the name archaebacteria, early genome-sequencing projects revealed that at the genomic level these species had more in common with eukaryotes than prokaryotes, specifically with respect to their DNA replication machinery. Some archaea have been found to have two origins of replication within their genome, which is not commonly seen in prokaryotic organisms. Additionally, similarities exist between proteins essential for DNA replication, including the DNA polymerase, helicase, clamp loader, and sliding clamp.

Though β -clamp from *E. coli* is a homodimer, PCNA molecules from archaea adopt different structural assemblies. For example, PCNA from *Haloferax volcanii*, a species of archaea that exists in extreme saline environments like the Dead Sea or the Great Salt Lake, is composed of a homotrimer of subunits (Figure 5(a)). By contrast, PCNA from *Sulfolobus solfataricus*, a species of archaea found in volcanic hot springs that utilize sulfur during metabolism in the absence of an oxygen source, is a heterotrimer (Figure 5(b)). These architectures differ greatly from β -clamp from *E. coli* but *H. volcanii* PCNA is assembled much like *S. cerevisiae* and *H. sapiens* PCNA (compare Figure 5(a) with Figure 4(a)), while *S. solfataricus* is assembled like the 9–1–1 clamp from *H. sapiens* (compare Figure 5(b) with Figure 4(b)).

Even though there are differences between the clamps from prokaryotic, archaeal, and eukaryotic systems with regard to the number and type of subunits, each adopts the same overall shape and size. The crystal structure for PCNA from two additional archaeal species, *Archaeoglobus fulgidus* (PDB code 1RWZ) and *Pyrococcus furiosus* (PDB code 1GE8), has also been solved. These two, not shown in Figure 5, are homotrimers of similar characteristics to PCNA from *S. cerevisiae*.

Additionally, each of these rings show similar charge distribution (*S. cerevisiae* (Figure 3(c)), *H. volcanii* (Figure 5(e)),

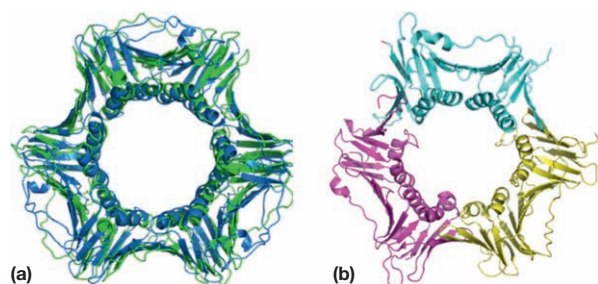


Figure 4 (a) Superimposition of ribbon diagrams for PCNA from *S. cerevisiae* (PDB code 3K4X) and *H. sapiens* (PDB code 1AXC), shown in green and blue, respectively. Ribbon diagram (b) of the checkpoint clamp (known as 9–1–1) from *H. sapiens* with RAD9 shown in pink, RAD1 in yellow, and HUS1 in cyan. The *in vivo* structure of this clamp may adopt a different shape as the Rad9 subunit has been truncated for ease of expression and crystallization. In each diagram α -helices are represented by coils and β -strands by arrows. Images were created using MacPyMol2006.

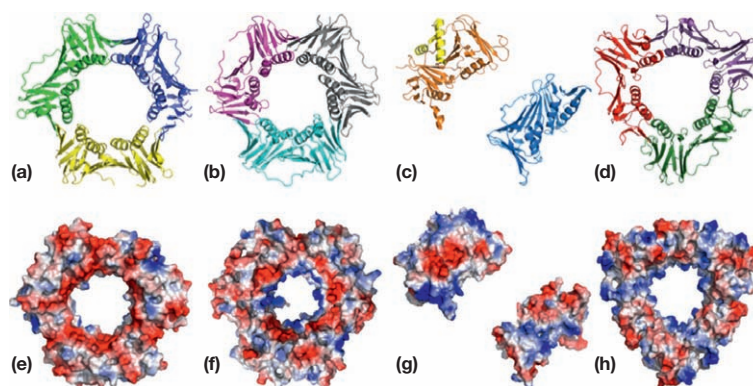


Figure 5 Ribbon diagrams of the PCNA structure from (a) *H. volcanii* (PDB code 3IFV) and (b) *S. solfataricus* (PDB code 2NTI) and the processivity factors (c) UL42 from HSV shown in orange (PDB code 1DML) and a *S. cerevisiae* PCNA monomer shown in blue (PDB code 1PLQ) and (d) gp45 from T4 bacteriophage (PDB code 1CZD). In the ribbon diagrams, each monomer is a different color and α -helices are represented by coils and β -strands by arrows. Surface-charge representations are given below each ribbon diagram (e)–(h) with the protein in the same orientation. For the charge distributions, red indicates negative charge, blue positive charge, and white hydrophobic areas. See text for more details. Images were created using MacPyMol2006.

S. solfataricus (Figure 5(f)) with the inner ring of α -helices having a positive charge that allows for interactions with the negatively charged DNA as it threads through the center of the clamp. The most divergent of the three as far as charge distribution is the *H. volcanii* PCNA. The complex appears to have a much lower concentration of positive charge within the center of the ring (this cluster of positive charge is generally considered to be a hallmark of clamp structure) when compared to other sliding clamps and also has a higher distribution of negative charge throughout the surface of the ring. It has been proposed that bound cations within the solvation shell may serve to coordinate interactions between DNA and the clamp. These differences are likely due to adaptations necessary for survival within a highly saline environment.

Viral Processivity Factors

Most of the processivity factors discovered thus far adopt a ring shape with a pseudo-six-fold axis of symmetry due to domain organization. Interestingly, the processivity factors from some viruses, specifically HSV and Epstein–Barr virus, share some structural features with eukaryotic clamps but have a different mechanism of action. These processivity factors are termed polymerase associated. They do not form ring-shaped protein complexes even though their overall structures are similar to that of a PCNA monomer. The structure of the HSV processivity factor UL42 (orange ribbon diagram) is shown in Figure 5(c) bound to a short segment of the HSV DNA polymerase (yellow ribbon diagram). For the purpose of comparison, the structure has been shown next to that of a *S. cerevisiae* PCNA monomer (blue ribbon diagram). From this figure, it appears that the HSV processivity factor is a two-domain structure with one surface predominantly α -helix and the opposite β -sheet, much like each of the sliding clamp monomers for the structures discussed previously. Though only a small fragment of the associated polymerase has been crystallized, it appears that the polymerase binds to the β -strand domain of UL42 leaving the α -helical surface free to interact with the DNA during replication, as has been observed for sliding clamps. The surface-charge

distribution (Figure 5(g)) also suggests that the α -helices interact with the DNA, as there is a concentration of positive charge on this face of the protein. This structure and biochemical data have shown quite different interactions between DNA and the processivity factor compared to those seen for sliding clamps; however, the overall surface-charge distribution and separation of α -helical and β -strand domains to create unique binding sites remain constant characteristics when compared to typical sliding-clamp processivity factors.

Bacteriophage Processivity Factors

Finally, the structures of processivity factors from bacteriophages have also been solved. The sliding clamps, gp45, from bacteriophage T4 and RB69 adopt a ring structure slightly different from other sliding clamps described thus far. These sliding clamps are homotrimers, like various PCNA molecules, but they lack the pseudo-six-fold axis of symmetry. Instead, they adopt a very strong threefold axis of symmetry that runs through the center of the ring (T4 gp45 shown in Figure 5(d)). In spite of differences in shape, the inner and outer diameter and thickness values are similar to those seen for *S. cerevisiae* PCNA. Additionally, the ring also has the inner ring of positive surface charge that has been observed in other ring structures (Figure 5(h)). Solution studies have revealed that one interface of the bacteriophage T4 clamp is opened, but not far enough to allow the clamp to bind DNA on its own. This clamp must still be loaded onto DNA by the activity of a clamp loader.

Conclusion

Although the replisomes from prokaryotes, archaea, and eukaryotes differ in some of their basic structure and individual components, the sliding clamps that serve to tether the polymerase to the template DNA are quite similar in structure and function. The clamps from *E. coli* and *S. cerevisiae* are almost identical in size and show similar mechanisms for binding to their respective clamp loaders and polymerases and to DNA

templates. Additionally, clamps and processivity factors from other species show much similarity in architecture, size, and shape, and structural features necessary for activity. This conservation suggests that processivity factors, in particular those that form sliding-clamp structures, have evolved as a necessary element for efficient DNA replication.

See also: **Molecular Biology:** DNA Polymerase III, Bacterial; DNA Polymerases: Kinetics and Mechanism; DNA Replication Fork, Bacterial; Eukaryotic DNA Polymerase δ ; Eukaryotic DNA Polymerase α ; Processivity Clamps in DNA Replication.

Further Reading

- Bloom LB (2009) Loading clamps for DNA replication and repair. *DNA Repair* 8: 570–578.
- Bowman GD, O'Donnell M, and Kuriyan J (2004) Structural analysis of a eukaryotic sliding DNA clamp–clamp loader complex. *Nature* 429: 708–709.
- Georgescu RE, Kim SS, Yurieva O, et al. (2008) Structure of a sliding clamp on DNA. *Cell* 132: 43–54.
- Indiani C and O'Donnell M (2006) The replication clamp-loading machine at work in the three domains of life. *Nature Reviews in Molecular and Cellular Biology* 7: 751–761.
- Kong XP, Onrust R, O'Donnell M, and Kuriyan J (1992) Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: A sliding DNA clamp. *Cell* 69: 425–437.
- Lee K-Y and Myung K (2008) PCNA modifications for regulation of post-replication repair pathways. *Molecules and Cells* 26: 5–11.
- Lehman AR (2006) Clubbing together on clamps: The key to translesion synthesis. *DNA Repair* 5: 404–407.
- McHenry CS (2003) Chromosomal replicases as asymmetric dimers: Studies of subunit arrangement and functional consequences. *Molecular Microbiology* 49: 1157–1165.
- Shamoo Y and Steitz T (2000) Building a replisome from interacting pieces: Sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell* 99: 155–166.
- Simonetta KR, Kazmirski SL, Goedken ER, et al. (2009) The mechanism of ATP-dependent primer-template recognition by a clamp loader complex. *Cell* 137: 659–671.
- Wijffels G, Dalrymple B, Kongsuwan K, and Dixon NE (2005) Conservation of eubacterial replicases. *International Union of Biochemistry and Molecular Biology Life* 57: 413–419.
- Winter JA, Christofi P, Morroll S, and Bunting KA (2009) The crystal structure of *Haloflex volcanii* proliferating cell nuclear antigen reveals unique surface charge characteristics due to halophilic adaptation. *BMC Structural Biology* 22: 55.
- Zhuang Z, Yoder BL, Burgers PM, and Benkovic SJ (2006) The structure of a ring-opened proliferating cell nuclear antigen–replication factor C complex revealed by fluorescence energy transfer. *Proceedings of the National Academy of Sciences of the United States of America* 103: 2546–2551.
- Zuccola HJ, Filman DJ, Coen DM, and Hogle JM (2000) The crystal structure of an unusual processivity factor, herpes simplex virus UL42, bound to the C terminus of its cognate polymerase. *Molecular Cell* 5: 267–278.

Small GTPases

K H Pedone and C J Der, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Adam Shutes and Channing J. Der, volume 4, pp. 48–54, © 2004, Elsevier Inc.

Glossary

Cytoskeleton A dynamic network of filaments, microfilaments, microtubules, and intermediate filaments.

Effector proteins Proteins that bind preferentially to the guanosine triphosphate (GTP)-bound form of small guanosine triphosphatases (GTPases) and facilitate the biological response of small GTPases.

GTPase-activating proteins Negative regulatory proteins that accelerate the intrinsic GTP hydrolysis activity of small GTPases.

Guanine nucleotide exchange factors Regulatory protein that accelerates the intrinsic guanosine diphosphate (GDP)/GTP exchange activity of small GTPases.

Guanine nucleotides Amine nucleotide bases composed of a guanine moiety attached to one (guanosine monophosphate, GMP), two (GDP), or three (GTP) phosphate groups.

Guanosine triphosphatases (GTPases) High-affinity GDP-, GTP-binding, and GTP-hydrolyzing proteins that act as molecular switches and timers that

cycle between inactive GDP-bound and active GTP-bound states.

Isoprenylation Posttranslational, covalent modification of proteins at carboxyl-terminal cysteine residues, by either C15 farnesyl or C20 geranylgeranyl isoprenoid lipids.

Kinase An enzyme with phosphorylating activity on either proteins or lipids.

Ras Identified originally as the protein encoded by retrovirus genes responsible for the oncogenic activities of the Harvey and Kirsten murine sarcoma RNA tumor viruses. They represent cellular rat genes transduced into the virus genome from their infected host during infection, leading to formation of rat sarcomas.

Superfamily, family, and subfamily Small GTPases are classified into hierarchical phylogenies on the basis of structural, sequence, and functional similarity between members. Members of the Ras family share ~50% amino acid identity with the four Ras proteins, whereas members of the Rho family, Rab family, and other Ras superfamily GTPases share ~25–30% amino acid identity with Ras proteins.

Identification and Classification of Small Guanosine Triphosphatases

The first and prototypical small guanosine triphosphatases (GTPases) were identified as the proteins encoded by the retrovirus oncogenes of the Harvey and Kirsten rat sarcoma (Ras) viruses (Figure 1). Since these initial discoveries, over 150 mammalian proteins have been identified that show varying degrees of both amino acid sequence and structural similarity with Ras proteins. Small GTPases were identified initially by a variety of approaches, some fortuitously (e.g., Rho), some by yeast genetic studies (e.g., Rabs), and others by nucleic acid sequence homology (e.g., Ral and Rap) or guanosine triphosphate (GTP)-binding activities. *In silico* database searches have identified additional members. These Ras superfamily proteins are further divided into families according to sequence identity and cellular function. The five major branches are the Ras, Rho, Rab, Arf, and Ran families (Figure 2). Small GTPases are well conserved in evolution, with structural and functional homologs of many of the mammalian proteins found in yeast, flies, worms, and plants. Interestingly, no Ras family proteins are found in plants.

In addition to GTPase activity, a feature of many small GTPases is lipid modification important for subcellular localization at membranes. Members of the Ras, Rho, and Rab families are modified by either the C15 farnesyl or C20 geranylgeranyl isoprenoid lipid (Figure 3). Protein prenylation is signaled by C-terminal sequences. For Ras and Rho small

GTPases, this lipid modification occurs at a particular motif, the CAAX tetrapeptide motif, consisting of a cysteine followed by two aliphatic amino acids, and a terminal X residue that determines whether the protein will be modified by farnesyltransferase (X=S, M, Q, and A) or geranylgeranyl transferase-I (X=L, F). Some Ras and Rho family small GTPases also are posttranslationally modified by a C16 palmitate fatty acid on cysteines adjacent to the CAAX motif. Rab proteins possess cysteine-containing C-terminal motifs (Figure 3) that signal geranylgeranyl isoprenoid posttranslational modification by a third prenyltransferase, geranylgeranyltransferase-II. Arf proteins are cotranslationally modified by the C14 myristate fatty acid at the N-terminus. Lipid modification facilitates the association of small GTPases with specific membrane compartments and typically is critical for protein function. Ran and the Rho family members RhoBTB1 and RhoBTB2 are unusual among small GTPases as they do not undergo any lipid modification.

Guanosine Diphosphate/GTP Structure and Regulation

Small GTPase Structure

The structure of the H-Ras protein has been determined in many forms and has established the general rules and requirements for all small GTPases. The H-Ras structure was the first to be solved, and subsequent structural studies of other Ras

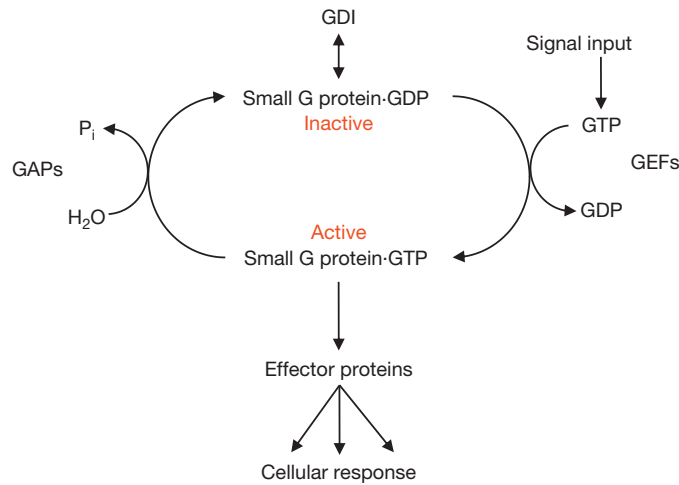


Figure 1 Small GTPases function as GDP/GTP-regulated molecular switches. Intracellular GDP and GTP are at low mM levels, and small GTPases have low pM binding affinities for guanine nucleotides. Hence, small GTPases are persistently guanine nucleotide bound. The intrinsic GDP/GTP exchange activity is accelerated by guanine nucleotide exchange factors (GEFs) to promote formation of the active GTP-bound protein, whereas GTPase-activating proteins (GAPs) stimulate the intrinsic GTP hydrolysis activity leading to inactive, GDP-bound protein and the release of free phosphate (P_i). Guanine nucleotide dissociation inhibitors (GDIs) prevent nucleotide exchange as well as GAP stimulation. Small GTPase activation commonly is mediated by a signal input that triggers GEF function. The GTP-bound GTPase displays higher affinity for downstream effector targets, leading to induction of various cellular responses.

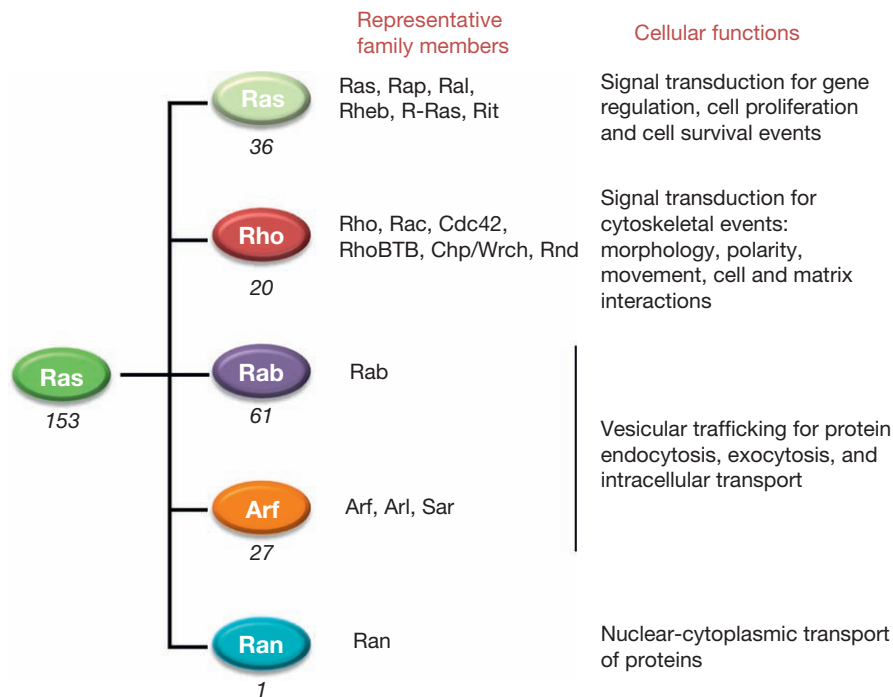


Figure 2 Small GTPases of the Ras superfamily. Five main families comprise the Ras superfamily, and proteins are organized according to sequence and functional similarity. Representative subfamily members are listed for each, with the total number of human family members indicated below the family name.

superfamily members have demonstrated a conserved structural fold (designated the G domain) shared among all GTPases, but with variations in features for different small GTPase family members (Figure 4).

All small GTPases possess a core G domain that is characterized by a set of canonical sequence elements shared with

other GTPases. The G domain contains the GTP-binding pocket and GTP hydrolytic machinery, although some proteins are GTPase deficient due to key sequence divergence in this domain. A significant observation from Ras structural studies was the primary difference between the guanosine diphosphate (GDP)- and GTP-bound forms, which is localized to

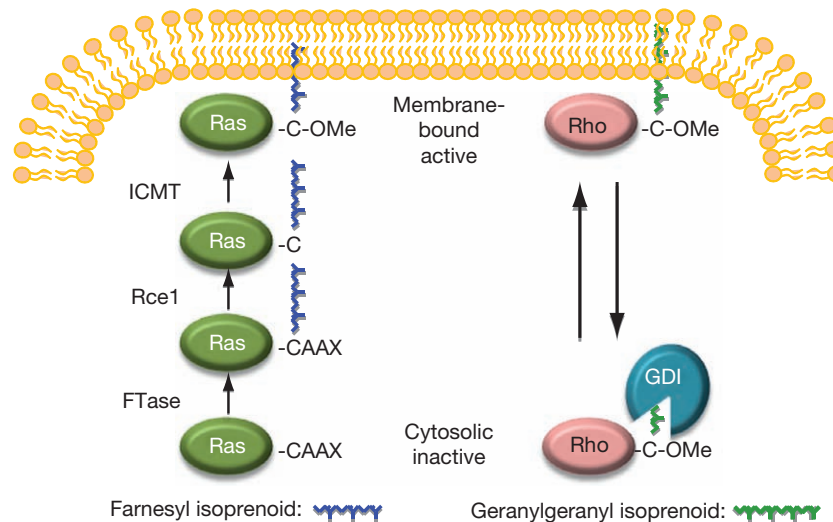


Figure 3 Posttranslational modification by prenylation is important for the function of some small GTPases. Members of the Ras and Rho family of small GTPases terminate with CAAX tetrapeptide sequences that signal a series of posttranslational modifications important for membrane binding and function. For example, Ras proteins are synthesized initially as inactive, cytosolic proteins. Ras proteins are first modified by farnesyltransferase (FTase) which catalyzes the covalent addition of a C15 farnesyl isoprenoid (F) lipid to the cysteine residue of the CAAX motif. This is followed by Rce1-mediated proteolytic removal of the -AAX peptide and Icmt-catalyzed carboxymethylation (O-Me) of the now terminal farnesylated cysteine residue. Rho GTPases undergo the same series of modifications, with the first step, addition of the geranylgeranyl isoprenoid (GG), catalyzed by geranylgeranyltransferase I (GGTase-I). RhoGDIs recognize the prenylated form of Rho GTPases and prevent their association with membranes, thus leading to Rho GTPase inactivation.

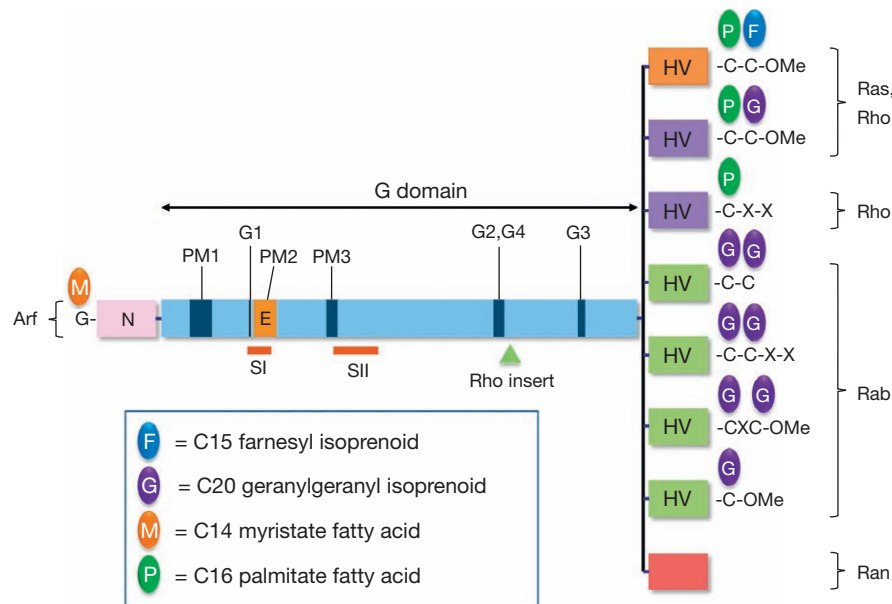


Figure 4 Small GTPase functional domains. Members of the small GTPase family share sequence similarities that define distinct functional domains. The G domain is comprised of a set of five conserved sequence elements involved in binding phosphate/ Mg^{2+} (PM) or the guanine base (G). The switch I and II (SI, SII) regions change in conformation during GDP/GTP cycling and contribute to preferential effector (E) binding to the GTP-bound state. For Arf, the N-terminus changes conformation during GDP/GTP cycling, with GTP-binding exposing the myristate to facilitate membrane binding. By contrast, the C-terminus of Ran changes conformation during GDP/GTP cycling. Ras, Rho, and Rab membrane association is facilitated by lipid modification of C-terminal sequences that include the CAAX motif (for Ras and Rho proteins) as well as sequences upstream of this sequence. Rho GTPases possess additional sequences (Rho insert) not found in other small GTPases. Some small GTPases contain additional N- or C-terminal sequence extensions.

two regions, termed switch I (residues 32–38) and switch II (residues 59–67). Similar switch-I and switch-II conformation changes also have been identified for the GDP- and GTP-bound states of other small GTPases.

Switch I and switch II show sequence divergence between families, although the loop-like structures of each are conserved. In addition to reflecting the nucleotide-bound state of the small GTPase, the switch regions also are involved in the interaction of small GTPases with effectors and regulatory molecules. Sequence divergence in the switch regions of different proteins contributes to the selectivity of small GTPases for various binding partners.

Members of the Rho family of small GTPases possess a unique structural feature that is absent in all other small GTPases. The Rho family proteins contain an insertion of ~13 amino acids, called the insert region, positioned between Ras residues 122 and 123, which form a surface-exposed loop (Figure 4). The primary sequence and secondary conformation of this insert region varies within the Rho family. Although the conformation of the insert region is not influenced by nucleotide binding, there is evidence that it is involved in effector interaction and activation.

Other small GTPases possess additional N- or C-terminal extensions important for function. For example, Arf family proteins contain an N-terminal extension necessary for insertion into and interaction with the membrane, whereas Ran has an elongated C-terminus that is crucial for its function in nuclear transport (Figure 4).

The Small GTPase Cycle and Its Regulation

The 'on' and 'off' switching nature of small GTPases is best described as a GDP/GTP cycle (Figure 1). The intrinsic processes of nucleotide exchange and GTP hydrolysis occur at basal rates characteristic to each small GTPase. Two classes of regulatory proteins act to increase the rate of these processes by several orders of magnitude: guanine nucleotide exchange factors (GEFs) increase the rate of nucleotide exchange, and GTPase-activating proteins (GAPs) stimulate GTP hydrolysis. The Rho and Rab families of small GTPases have additional regulatory protein binding partners called guanine nucleotide dissociation inhibitors (GDIs). Among all small GTPases, evidence indicates that only RhoBTB1 and RhoBTB2 have non-functional GTPase domains, despite sequence homology with the GTPase domains of other Rho family members.

Guanine Nucleotide Exchange Factors

GEFs constitute the primary mechanism for regulated activation of small GTPases and their downstream effects *in vivo*. They are responsible for promoting the exchange of free cytosolic GTP for bound GDP. Exchange is not an active process; rather, GEFs act to facilitate the replacement of one nucleotide with another, and it is the ~10-fold molar excess of cytosolic GTP that favors the binding of GTP rather than GDP. Furthermore, the binding of GTP during the exchange reaction weakens the affinity of the GEF for the small GTPase and induces subsequent dissociation. The binding of GDP does not cause dissociation of the complex.

GEFs have been identified for many small GTPases. They are related by shared sequence of the exchange motif and exhibit selectivity for GTPase binding partners within, but not between, different GTPase families. For example, the two largest classes of GEFs are the CDC25 homology domain-containing proteins which act on Ras family members, and the Dbl homology domain-containing proteins which act on Rho family members. Structurally distinct GEF have been identified for Rab, Arf, and Ran family proteins. As with other smaller classes of GEFs, the structures of the catalytic exchange motifs of RhoGEFs and RasGEFs are unrelated. However, the overall mechanism of exchange appears to be universal; binding of the GEF sufficiently perturbs the phosphate-binding site within the GTPase domain to allow the insertion of GTP and displacement of GDP.

Different GEFs Are Activated by Different Signaling Cascades

The various GEFs identified for Ras and Rho family members overwhelmingly outnumber their respective GTPase binding partners. At least 30 Ras family GEFs have been characterized to date, and more than 80 RhoGEFs have been identified. As a result, many Ras and Rho GTPases can be activated by multiple GEFs. For example, at least 24 Dbl family RhoGEFs have been reported to activate RhoA. The general understanding of these disparities in numbers is that this provides a mechanism whereby divergent signaling cascades may activate the same small GTPase. GEFs that share a common small GTPase substrate often diverge in the sequences that flank the GEF catalytic domain. These sequences commonly contain a large number of different protein–protein and protein–lipid domains or recognition motifs that allow them to respond to a wide variety of signals. In this manner, a common small GTPase can serve as an intracellular convergence point for multiple divergent extracellular cues (Figure 5). By contrast, only a handful of GEFs have been identified for Rab family proteins, the largest family of small GTPases. RabGEFs also possess significant sequence diversity, with at least six distinct families identified. The single Ran small GTPase is regulated by only one RanGEF.

Activation of GEFs occurs by a variety and frequently a combination of mechanisms, including protein–protein interaction, posttranslational modification, second-messenger-induced activation, and membrane recruitment. The classical model of small GTPase activation is that of Ras and its GEF, son of sevenless (Sos). Upon extracellular ligand activation of a receptor protein tyrosine kinase (e.g., the epidermal growth factor receptor, EGFR) and autophosphorylation of specific intracellular tyrosine residues in the cytoplasmic domain of EGFR, an adaptor protein, Grb2, is recruited to specific phosphotyrosine sites via its Src homology 2 (SH2) domain (Figure 5). Grb2, in turn, interacts via Src homology 3 (SH3) domain interaction with specific proline-rich sequences on Sos, and recruits Sos to the membrane to activate resident Ras family proteins.

Membrane localization of the GEF, in proximity with its membrane-bound GTPase targets, is a critical step in the activation process of small GTPases. Scaffolding proteins such as Grb2 represent one mechanism for the targeted recruitment of

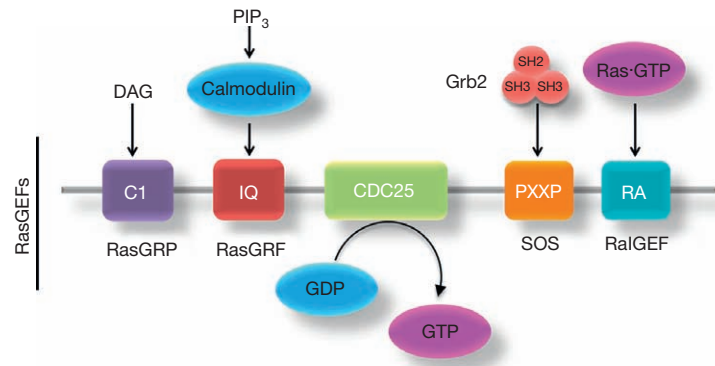


Figure 5 Ras guanine nucleotide exchange factors are activated by diverse upstream signals. Extracellular signals promote Ras activation by stimulating the activity of different RasGEFs. All RasGEFs share a common CDC25 homology domain that catalyzes nucleotide exchange on various Ras family small GTPases (Ras, Rap, R-Ras, and Ral). However, they diverge in sequence flanking the catalytic domain and are characterized by distinct interaction domains/motifs that link their activation to specific upstream signaling events. Examples of various GEF activation mechanisms are shown. RasGEFs can be activated by phospholipase C-catalyzed formation of membrane-bound diacylglycerol (DAG) and DAG binding to the C1 domain of RasGRP1-4; calcium influx and calmodulin binding to the IQ domain of RasGRF; Grb2 adaptor binding to proline-rich (PXXP) motifs in Sos via its tandem SH3 domains; and activated Ras-GTP binding to Ras Association (RA) domains of RalGEFs.

GEFs to the plasma membrane and other intracellular lipid compartments. Transiently produced membrane lipids including diacylglycerol (DAG) and phosphatidylinositol 3,4,5-trisphosphate (PIP_3) are other common mechanisms for controlling localized GEF, and likewise small GTPase, activity within the cell (Figure 4).

An emerging model of GEF activation, which appears to be common to several RhoGEFs of the Dbl family, is the relief of intramolecular interactions that constrain the protein in an inactive conformation. This has been shown best for the Dbl family RhoGEF Vav1 in which phosphorylation leads to the conformational relief of autoinhibition by the N-terminal fragment. Following phosphorylation, the active conformation of Vav1 is available to interact with cognate small Rho GTPases. That N-terminal truncation of sequences upstream of the Dbl homology domain of other RhoGEFs creates constitutively activated variants supports the generality of this auto-inhibitory mechanism. Additional mechanisms of RhoGEF activation include activated Ras binding to a Ras-binding domain in Tiam1 or adenomatous polyposis coli (APC) tumor suppressor binding to Asef.

GTPase-Activating Proteins

The intrinsic hydrolysis rate of the majority of small GTPases is surprisingly slow and sufficiently slow to limit significantly the efficiency of GTPase-mediated signal transduction. In the absence of modulators, signals not only would be transduced too slowly, but also the GTPase 'on' state would persist for an excessive period of time. Whereas GEFs act to accelerate small GTPase activation, GAPs serve to accelerate the hydrolytic activity of the small GTPase and therefore make signaling more efficient and dynamic (Figure 1).

As with GEFs, the number of GAPs which act upon Ras and Rho small GTPases exceeds the number of small GTPase targets. Ras proteins are regulated by six distinct RasGAPs, and there are 67 human RhoGAPs. GAPs for small GTPases are not

conserved in sequence among the small GTPase families and have different mechanisms of action. In general, however, GAPs serve to stabilize the hydrolytic machinery of the small GTPase.

Similar to GEFs, many GAPs are also multidomain proteins. Beyond the GAP catalytic domain, the proteins frequently possess regulatory elements important for interaction with proteins (e.g., Rap1GAP activation by Gi), lipids (e.g., ARAP ArfGAP activation by PIP_2 and PIP_3), and second-messenger molecules (e.g., CAPRI Ras/RapGAP membrane recruitment by Ca^{2+}). Some more complex proteins contain both GEF and GAP domains (e.g., Bcr has both RhoGEF and RhoGAP activity), and some contain multiple domains of the same type but with different selectivity (e.g., Sos has two distinct GEF catalytic exchange domains, a CDC25 homology domain for Ras and a DH domain for Rho proteins).

Guanine Nucleotide Dissociation Inhibitors

GDI have been identified for Rho and Rab proteins. Two distinct negative regulatory functions have been ascribed to GDIs. First, they bind to and mask the isoprenoid lipid modification in the C-terminus of the small GTPase, thus preventing the association of the small GTPase with membranes (Figure 3). Second, in general, GDI binding perturbs GAP and GEF regulation, preventing small GTPase activation. Three RhoGDIs and one RabGDI have been described and can recognize multiple family members. The mechanisms of regulation of GDI activity are unclear but may involve phosphorylation.

Genetic Deregulation of Small GTPases

Normal eukaryotic cells rely on the activity of small GTPases to orchestrate basic biological processes, and the innate mechanisms of GEF, GAP, and GDI regulation, in most cases, maintain a tightly controlled level of GTPase activity sufficient for these

purposes. Genetic deregulation, however, can lead to aberrant GTPase signaling, and, in this manner, overly active small GTPase activity contributes to many human cancers and developmental disorders.

The genetic mechanisms that underlie deregulation of GTPase biology are of a variety of types, the most common being mutational activation of RAS, the most frequently mutated oncogene in human carcinomas (33% of all human cancers). The mutated Ras proteins found in human cancers contain single amino acid substitutions primarily at glycine 12 (G12), glycine 13 (G13), or glutamine 61 (Q61), which render the proteins insensitive to GAP stimulation. These GTPase-deficient mutants are persistently GTP-bound in the absence of upstream stimulation and consequently cause deregulated effector activation. The experimental introduction of analogous mutations at residues corresponding to Ras G12 or Q61 also renders other small GTPases GAP insensitive and constitutively activated. Some wild-type small GTPases, however, possess naturally occurring sequence variation at or near these two residues and are persistently GTP-bound proteins (e.g., RhoE/Rnd3 and Rac1b).

Elevated small GTPase activity in pathological conditions also occurs by indirect genetic mechanisms including altered gene expression of wild-type proteins (e.g., RhoC) or their regulatory proteins, alternative gene splicing (e.g., Rac1b, a constitutively GTP-bound variant of Rac1), and amplified upstream signaling processes. Multiple upstream activators of small GTPases drive oncogenesis by either mutational activation (e.g., EGFR) or protein overexpression (e.g., Vav, EGFR, and P-Rex1).

Genomic loss of negative regulators of small GTPase activity promotes aberrant signal transduction as well, as is the case with the loss of the RhoGAP DLC-1 in liver and other cancers. Paradoxically, increased expression of some ArfGAPs and RhoGDIs has been observed in some cancer types, suggesting that aberrant levels, high or low, of some small GTPase regulators can contribute to pathologically relevant GTPase activity. Thus, while GDP/GTP regulation is the major mode of

small GTPase regulation, for some GTPases, regulation by other mechanisms is also important.

Diverse Biology and Function of Small GTPases

Despite their strong structural and biochemical similarities, small GTPases facilitate a remarkably divergent spectrum of cellular functions. This diversity is accomplished by the input signals that regulate the distinct GAPs and GEFs to modulate GDP/GTP cycling, and by the unique set of effectors that are recognized by the GTP-bound forms of small GTPases within and between families, resulting in many different cellular responses.

Ras Proteins as Signaling Nodes and Regulators of Cell Proliferation

The frequent mutation of Ras proteins in human cancers has made these small GTPases the most intensely studied and best characterized. Recently, mutational activation of Ras has also been found in several human developmental disorders (e.g., Noonan syndrome, Costello syndrome, and cardio-facio-cutaneous syndrome). Ras proteins serve as signaling nodes, where a wide diversity of extracellular signals such as growth factors (epidermal growth factor and platelet-derived growth factor), hormones (insulin), cytokines (interleukin-1), and the extracellular matrix proteins (via integrins) converge on and cause activation of Ras. Activated Ras in turn interacts with and regulates the activities of downstream effectors with highly divergent biochemical functions. Recent reviews have provided detailed discussions of Ras effector utilization. Therefore, we have highlighted various themes.

The Raf serine/threonine kinases are important effectors of Ras and facilitate activation of the extracellular signal regulated kinase (ERK) mitogen-activated protein kinase cascade (Figure 6). Ras induces Raf activation by promoting

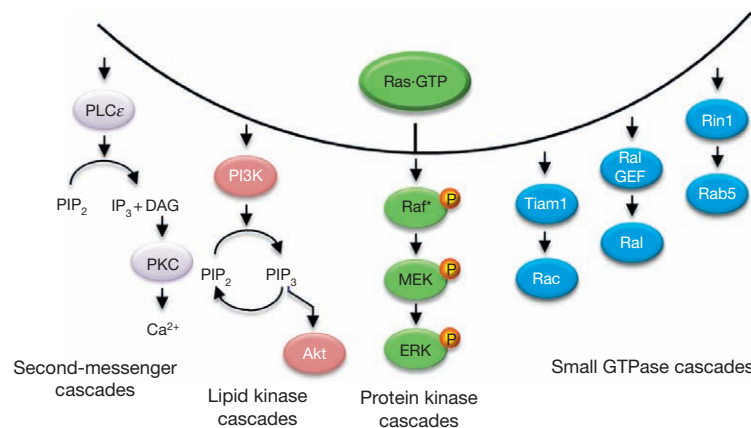


Figure 6 Ras effectors mediate diverse signaling outcomes. Activated, GTP-bound Ras can bind and regulate functionally diverse downstream effector targets. These include activation of a protein kinase cascade (Raf), activation of the PI3K lipid kinase, and activation of PLC epsilon and the production of second messengers (DAG and calcium). Several Ras effectors are GEFs that activate other members of the Ras superfamily, including Rab5 (Rin1-3), Ral (RalGEFs), and Rac (Tiam1). Thus, Ras functions as a signaling node, where diverse signals converge and cause its activation (Figure 5) and induction of diverse downstream cytoplasmic signaling pathways.

association of the normally cytosolic Raf with the plasma membrane where a complex set of events, including protein phosphorylation, activates Raf kinase function. Activated Raf then phosphorylates and activates the MEK1 and MEK2 dual-specificity protein kinases, which phosphorylate and activate ERK1 and ERK2. Activated ERK translocates to the nucleus and phosphorylates a variety of targets that include Ets family transcription factors. The Raf–MEK–ERK kinase cascade contributes significantly to the growth-regulatory functions of Ras. The mutational activation of one Raf protein (B-Raf) in human cancers and the mutational activation of Raf and MEK in developmental disorders support the importance of this effector pathway in promoting the consequences of Ras activation in these diseases.

The second best-characterized class of Ras effectors is the phosphatidylinositol 3-kinases (PI3K), a family of lipid kinases (Figure 6). A major activity of PI3K is the conversion of membrane-associated phosphatidylinositol 4,5-bisphosphate (PIP₂) to phosphatidylinositol 3,4,5-trisphosphate (PIP₃). PIP₃ in turn can regulate the activity of a diverse range of targets, including activation of the Akt serine/threonine kinase and RacGEFs, via PIP₃ interaction with pleckstrin homology domains present in these proteins. Mutational activation of p110 alpha, a catalytic subunit of PI3K, is seen commonly in human cancers.

One significant theme that has emerged is that a number of Ras effectors are GEFs, thereby linking Ras activation with regulation of other small GTPases (Figure 6). First, there is a family of RalGEFs that binds GTP-bound Ras and promotes activation of RalA and RalB, members of the Ras family of small GTPases. Second, the Ras effector Rin1 functions as a GEF for Rab5, thus linking Ras activity with Rab-regulated vesicular trafficking. Third, Ras binds and activates Tiam1 which functions as a GEF for Rac. Fourth, Ras binds phospholipase C epsilon, which also contains a CDC25 homology Ras GEF catalytic domain that activates the Rap small GTPases, members of the Ras family.

While Ras GTPases function as positive regulators of cell proliferation, some small GTPases appear to possess opposing functions and may operate as tumor suppressors rather than oncogene proteins. This includes Rerg, NOEY2/ARH1 and Rig. Thus, while these proteins share significant sequence and biochemical functions with Ras, clearly their effector functions are quite distinct. Additionally, whereas mutational activation of Ras proteins is associated with oncogenesis, it is the loss of gene expression of these Ras-related proteins that accounts for their loss of function in tumor cells.

Rho Family Proteins and Actin Cytoskeletal Organization, Cell Morphology, and Cell Movement

Similar to Ras proteins, Rho family small GTPases also function as signaling nodes activated by diverse extracellular signals. Perhaps the best-characterized cellular function of these proteins is their regulation of actin cytoskeletal organization, which in turn influences cell morphology, cell–matrix and cell–cell interactions, cell polarity, and cell movement. For example, RhoA causes actin stress fiber formation, Rac causes actin accumulation at the leading edge of a motile cell and lamellipodia formation, and Cdc42 promotes actin microspikes and filopodia formation. These changes in actin organization can influence cell shape, cell motility, and cell–cell interactions. Rho family members also participate in signaling events leading to changes in gene transcription and cell-cycle progression.

Small GTPases and Membrane Trafficking

Rab proteins constitute the largest subfamily of small GTPases with more than 60 mammalian members. Rab proteins are involved in the modulation of specific steps in eukaryotic membrane trafficking in the secretory and endocytic pathways. Arf small GTPases are also involved in the regulation of cytoplasmic vesicular trafficking. By contrast, Ran is a regulator of nucleocytoplasmic transport. The regulation of the GDP/GTP cycle is critical for the transport functions of these small GTPases.

See also: **Cell Architecture and Function:** Rho GTPases and Actin Cytoskeleton Dynamics; **Signaling:** Arf Family; G₁₂/G₁₃ Family; Phosphatidylinositol Bisphosphate and Trisphosphate; Rab Family; Ran GTPase; Ras Family.

Further Reading

- Bos JL, Rehmann H, and Wittinghofer A (2007) GEFs and GAPs: Critical elements in the control of small G proteins. *Cell* 129: 865–877.
- Cox AD and Der CJ (2010) Ras history: The saga continues. *Small GTPases* 1: 1–27.
- Jaffe AB and Hall A (2005) Rho GTPases: Biochemistry and biology. *Annual Review of Cell and Developmental Biology* 21: 247–269.
- Stenmark H (2009) Rab GTPases as coordinators of vesicle traffic. *Nature Reviews Molecular Cell Biology* 10: 513–525.
- Wennerberg K, Rossman KL, and Der CJ (2005) The Ras superfamily at a glance. *Journal of Cell Science* 118: 843–846.

Small RNAs in Bacteria

E R Westra, M M Jore, S Al-Attar, S J J Brouns, and J van der Oost, Wageningen University, Wageningen, The Netherlands

© 2013 Elsevier Inc. All rights reserved.

Glossary

cas CRISPR associated.

CRISPR Clustered regularly interspaced short palindromic repeats.

crRNA CRISPR RNA – processed transcript of CRISPR locus.

HGT Horizontal gene transfer – exchange of DNA between different species.

miRNA MicroRNA – specific class of eukaryotic small RNA.

mRNA Messenger RNA – transcript of open reading frame.

NABP Nucleic acid binding protein.

ORF Open reading frame – protein-coding DNA fragment.

pAgo Prokaryotic argonaute – bacterial/archaeal homolog of the eukaryotic key protein of RNA interference.

piRNA Piwi-interacting RNA – specific class of eukaryotic small RNA.

Piwi P-element-induced wimpy testis – subclass of the eukaryotic Argonaute protein.

RAMP Repeat associated mysterious protein – subclass of Cas proteins.

RNAi RNA interference – RNA-dependent pathway that leads to silencing and/or degradation of RNA targets.

RNAP RNA polymerase – protein complex responsible for transcription.

SAM S-adenosyl methionine – metabolite involved in methyl group transfer.

SD Shine Dalgarno sequence – ribosome binding site of mRNA.

siRNA Small interfering RNA – specific class of eukaryotic small RNA.

sRNA Small RNA.

tRNA Transfer RNA – RNA molecule involved in translating a messenger RNA into a protein.

Trp Tryptophan – amino acid.

UTR Untranslated region – part of mRNA transcript that is not translated.

Introduction

Like all living cells, bacteria rely on a myriad of metabolic pathways for the generation of energy and biomass to support maintenance and growth. Bacteria that thrive in ecosystems with fluctuating physical and chemical conditions have evolved control mechanisms that allow them to efficiently adjust their metabolic infrastructure by switching on appropriate biochemical pathways, while switching off others. The regulatory mechanisms responsible for fine-tuning these pathways are of major importance for optimal fitness. Control can be executed at all possible levels: DNA transcription, messenger RNA (mRNA) translation, enzyme activity, protein–protein interactions, or protein localization. Often, regulation acts on more than one level of a pathway. A classical example is the regulation of the tryptophan (Trp) biosynthesis pathway (see [Box 1](#)).

Although the analysis of regulation of cellular processes has often focused on control by proteins, it has become clear during the last decades that small RNA (sRNA) molecules also exert many regulatory functions. During the 30 years since the first discovery that bacterial sRNA molecules regulate plasmid replication, the insight into sRNA pathways has increased exponentially.

Most sRNAs act at the level of transcription and translation of mRNA. These sRNAs can be divided into (1) *cis*-encoded RNA elements that are either part of the transcript they target (riboswitches, usually encoded in the 5′ untranslated region of transcript (UTR)), or derived from the opposite strand (antisense sRNA, with perfect complementarity to the target mRNA) and (2) *trans*-encoded sRNAs that are encoded by distinct genomic loci, acting on mRNA through direct base pairing with imperfect complementarity. *cis*-Acting riboswitches

undergo a conformational change upon temperature increase or substrate binding that affects transcription/translation of the downstream open reading frame (ORF). Both *cis*- and *trans*-encoded antisense sRNAs modulate either mRNA transcription, translation, or mRNA stability. Many sRNAs of these classes mediate downregulation of gene expression, but some can also activate genes.

Other classes of sRNAs specifically bind to proteins (enzymes) and as such affect their functionality (activity). In addition, new classes of sRNAs have recently been discovered that make use of dedicated protein components for exerting their specialized functions. For example, sRNAs derived from

Box 1

To avoid unnecessary energy consumption, activation of the tryptophan (Trp) biosynthesis pathway only occurs when intracellular tryptophan levels are low. Several regulatory mechanisms controlling this pathway have been described. (1) In *Escherichia coli*, tryptophan activates a repressor (TrpR) that negatively regulates expression of the *trp* operon (negative feedback regulation). (2) In *Pseudomonas aeruginosa*, indoleglycerol phosphate, a Trp biosynthetic pathway intermediate, induces TrpI, an activator of transcription of tryptophan biosynthesis genes. Indoleglycerol phosphate levels increase during Trp biosynthesis, giving rise to positive feedback. (3) When tryptophan levels are low, uncharged tRNA^{Trp} accumulate. In *E. coli*, the presence of tandem tryptophan codons in the leader peptide of the *trp* operon causes stalling of ribosomes when uncharged tRNA^{Trp} are relatively abundant. This attenuation permits the formation of an anti-terminator loop, instead of the terminator loop that is present otherwise, thereby allowing for transcription (and translation) of the remainder of the operon to proceed. (4) In *E. coli*, the end product of the pathway (Trp) inhibits the activity of the first enzyme in the biosynthetic pathway (anthranilate synthase), giving rise to negative feedback regulation.

clustered regularly interspaced short palindromic repeat (CRISPR) loci protect the cell against invading viral and plasmid DNA. Although the eukaryotic RNA interference (RNAi) system functions in defense as well, the protein components involved in CRISPR-based defense differ substantially from their eukaryotic counterparts. The core enzyme Argonaute is the only protein component of the RNAi pathway that is conserved in prokaryotes. It has recently been hypothesized that prokaryotic Argonaute (pAgo) is also involved in sRNA (or DNA)-mediated antiviral defense. Altogether, prokaryotic sRNA molecules are involved in a wide range of regulatory processes that will be outlined in more detail below.

cis-Acting sRNAs – Riboswitches

A relatively simple means of regulating gene expression relies on the formation of secondary structures within the mRNA through internal base pairing. These secondary structures may be influenced (switched on/off) by physical or chemical signals as described below, and either interfere with transcription/translation of the mRNA itself or affect the stability of the transcript. Often the secondary structure occludes the Shine–Dalgarno (SD) sequence, hence interfering with ribosome binding and translation initiation.

RNA Thermosensors

In the case of RNA thermosensors, the SD masking structure is usually a regular hairpin that is stable only at lower

temperatures. At elevated temperatures, the helical structure melts, making the SD sequence accessible for ribosome binding and translation initiation (**Figure 1(a)**). This type of regulation often controls genes that are involved in response to sudden temperature increase, such as heat-shock genes, and virulence gene expression after entry in a warm-blooded (mammalian) host. The σ_{32} (RpoH) containing RNA polymerase (RNAP) holoenzyme plays an important role in the *Escherichia coli* heat-shock response. Translation of the *rpoH* mRNA is regulated by a well-studied thermosensor that only allows efficient translation at elevated temperatures. This thermosensor contains four hairpin structures that encompass the entire SD sequence as well as the start codon. The imperfect complementarity in the stems causes enough destabilization for immediate melting of the structure at slightly elevated temperatures, while favoring the hairpin structure at lower temperatures. Introducing stabilizing point mutations in the σ_{32} thermosensor sequence strongly decreases translation rates of the mRNA during heat shock, because it has become unresponsive to temperature change.

Other well-studied examples of thermosensors include the 5'-UTR of the phage λ cIII gene that is involved in deciding whether the lytic or lysogenic pathway of the viral lifecycle will be entered. Another example is the repression of heat-shock gene expression (ROSE) element that is found upstream of several *Bradyrhizobium japonicum* genes. Analogous elements probably exist in other organisms as well. Like the *rpoH* thermosensors, these structures comprise a set of stem-loop structures that include the SD sequence and the AUG start codon. In addition, a certain number of mismatches are required for melting in response to a small increase in temperature.

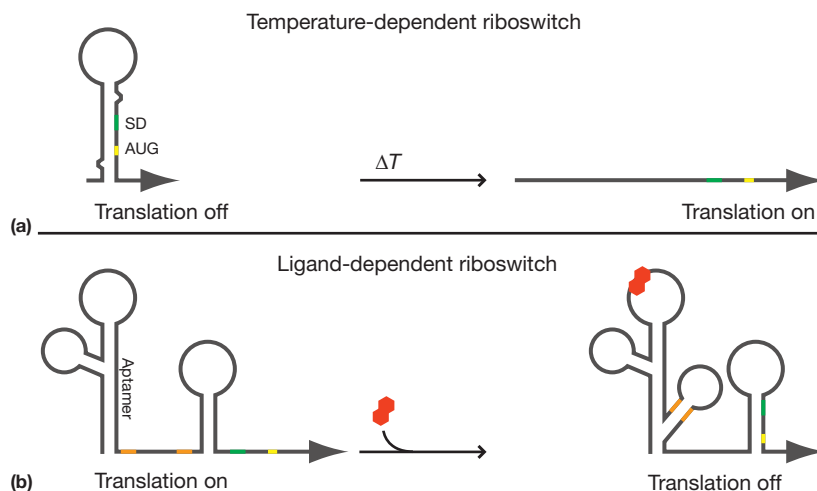


Figure 1 Riboswitches are *cis*-acting regulatory elements that are usually located in the 5' UTR of the mRNA they regulate. Upon elevated temperature (temperature-dependent riboswitch) or ligand binding (ligand-dependent riboswitch), their conformation changes, resulting in alterations in mRNA transcription or translation. Riboswitches serve as an on/off switch, and ligand binding/elevated temperature can either cause inhibition (off-state) or relieve inhibition of mRNA transcription/translation (on-state). In the off-state, either the SD sequence is masked (interference with translation), or a transcriptional terminator stem-loop is formed (interference with transcription). In the examples depicted here, temperature-dependent and ligand-dependent riboswitches that control translation by masking the SD sequence are shown. The secondary structure of a temperature-dependent riboswitch is very sensitive to small increases in temperature, due to the mismatches present in the stem-loop. The melting of the secondary structure allows translation (or transcription when the stem-loop is a transcription terminator) of the downstream gene, since the SD sequence is no longer masked. Ligand-dependent riboswitches contain an aptamer region that specifically binds a ligand, which leads to structural rearrangements in both the aptamer region, as well as the downstream gene expression region. In this example, the structural rearrangements lead to masking of the SD, but ligand-dependent riboswitches are known where the structural rearrangements lead to exposure of the SD sequence, or where transcriptional terminators are formed or lost.

RNA Ligand Sensors

The majority of the riboswitches requires the binding of a ligand, rather than elevated temperature, to induce structural rearrangements. The high number of ligand-dependent riboswitches that have been identified illustrates their importance. These riboswitches consist of a ligand-binding domain and a gene expression domain. The ligand-binding domain, also known as the 'aptamer region' (50 up to several hundreds of nucleotides (nt) in size), functions as a highly specific receptor for metabolites present in the cell. Binding of the metabolite causes formation of alternative secondary structures encompassing the aptamer region and part of the downstream sequence. Generally, structural rearrangements upon ligand binding result in interference with translation (masking of the SD sequence) of the transcript (Figure 1(b)) or with transcription (e.g., formation of a hairpin structure (terminator) that aborts further transcription by RNAP). However, some examples of riboswitches that promote gene expression upon metabolite binding are also known. In these cases, the structural rearrangements lead to the exposure of the SD sequence or the formation of anti-terminator structures (alternative secondary structures that prohibit the formation of the transcription terminator).

A riboswitch makes the mRNA responsive to the availability of nutrients or metabolites inside the cell. This type of regulation is often found in mRNAs that encode proteins involved in the uptake or biosynthesis of the metabolite that it binds. The ligands for a riboswitch can be as simple as a metal ion, or as complex as amino acids and complex cofactors. In some cases, several genes are controlled by the same 5' riboswitch. For example, the S-adenosyl methionine (SAM) riboswitch, which recognizes SAM (important for methylation reactions), is found upstream of a number of *Bacillus subtilis* genes that are involved in methionine and SAM biosynthesis and uptake. When the SAM riboswitch is in the ligand-bound state, expression of the downstream genes is switched off. The structural rearrangements that take place after ligand binding involve both the aptamer and the downstream gene expression domain (Figure 1(b)). As a result, the responsiveness of the riboswitch-containing transcript to the ligand depends not only on the affinity of the aptamer to the ligand but also on the downstream gene expression domain. Although the aptamers in all SAM riboswitches are identical and, thus, have equal affinities for SAM, translation of different genes is not equally responsive to SAM due to differences in the downstream gene expression domain. The mRNAs of the anabolic pathway components are highly sensitive to SAM, so that they are not translated until SAM levels become almost depleted. By contrast, the mRNAs of the components involved in SAM uptake are much less sensitive to SAM, and repression of these genes is relieved while relatively high levels of SAM are still present. This additional layer of regulation causes that SAM uptake systems are switched on first when SAM levels drop. If this has no effect, SAM levels further drop to a level that relieves repression on the mRNA for SAM biosynthesis genes.

A decreased sensitivity can also be accomplished by multiple copies of the same riboswitch, since ligand binding to each aptamer is needed. An example of this is the riboswitch that

binds uncharged transfer RNA (tRNA)^{Thr} to control the *thrZ* gene (threonyl-tRNA synthetase) in *B. subtilis*. Some mRNAs also contain several different 5' riboswitches to allow integration of multiple signals. For example, the mRNA of *metE* (encoding vitamin B₁₂-independent methionine synthase) contains two riboswitches, one of which binds SAM, which is the end product of MetE (feedback control), and the other vitamin B₁₂. The presence of vitamin B₁₂ signifies that a more efficient pathway for SAM biosynthesis can be utilized, which makes use of the vitamin B₁₂-dependent methionine synthase encoded by *metH*.

The generation of transcripts containing a riboswitch that blocks translation may seem a waste of resources compared to regulation at the level of transcription. However, it enables a more rapid response to temperature shock or changes in metabolite levels, which may be worth the expense. To avoid further spoilage of resources due to partial translation of the sequence, the riboswitch is almost always located at the 5'-UTR of the transcript.

cis-Encoded sRNAs – Antisense

Apart from elements in the mRNA itself that regulate transcription or translation through secondary structure formation, gene expression can also be regulated by small antisense RNA molecules that act through base pairing with the target mRNA (Figure 2(a)). The complementarity between the target mRNA and the sRNA is perfect in the case of *cis*-encoded sRNAs, because they are transcribed from the opposite strand of the same locus as the target mRNA. In most cases, *cis*-encoded sRNAs inhibit transcription or translation of their target transcript. Some function as anti-toxin by repressing the translation of a toxic protein. These toxin/anti-toxin systems are often encoded on large conjugative plasmids and ensure that the plasmid is retained in the cell and its progeny, because loss of the short-living anti-toxin causes cell death due to the long-living toxin. These toxin/anti-toxin pairs are sometimes found on the chromosome of bacterial species as well, where they carry out functions that are not yet well understood, possibly being beneficial for the host by inducing slow growth under stress conditions, or providing defense against plasmids with the same toxin/anti-toxin system.

Other *cis*-encoded sRNAs bind polycistronic mRNA transcripts to modulate expression levels of individual genes within the operon. An example of this type of regulation is the base pairing of the antisense sRNA GadY with the *gadXW* mRNA, encoding GadX and GadW, which respectively activate and repress the glutamate-dependent acid response in *E. coli*. The base pairing between GadY and *gadXW* mRNA leads to cleavage between the *gadX* and *gadW* coding sequences and increased *gadX* transcript levels. Other examples are the antisense sRNA-induced transcription termination after *fata* in the *fatDCBAangRT* operon in *Vibrio anguillarum*, and sRNA-induced cleavage between the *isiA* and *isiB* coding sequences in the *isiAB* transcript in *Synechocystis*.

trans-Encoded sRNAs

Similar to *cis*-encoded sRNAs, *trans*-encoded sRNAs regulate mRNA translation and stability. They differ from *cis*-encoded

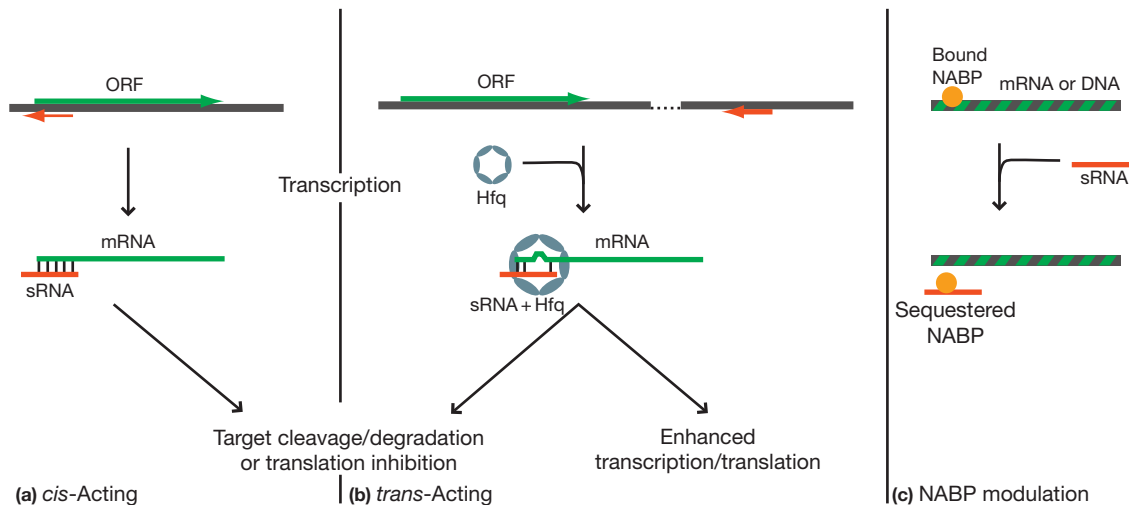


Figure 2 Small noncoding RNA molecules can regulate mRNA stability or translation (antisense RNA), or protein functionality (sequestration). Antisense sRNAs can be either *cis*-encoded or *trans*-encoded. (a) The complementarity between the target mRNA and the sRNA is perfect in the case of *cis*-encoded sRNAs, since they are transcribed from the opposite strand of the same locus as the target mRNA. The outcome of base pairing between the target mRNA and *cis*-encoded sRNA is usually inhibition of translation or degradation of the mRNA, or modulation of the expression of genes in an operon, resulting from sRNA-induced cleavage between ORFs in a polycistronic transcript. (b) *Trans*-encoded sRNAs show only limited complementarity with their target and need the hexameric Hfq to exert their regulatory function. Hfq facilitates binding of *trans*-encoded sRNA to target mRNA. Binding of the sRNA to the target mRNA sequence usually occurs at the 5' end to mask the SD sequence and/or start codon of the mRNA. Although *trans*-encoded sRNAs usually inhibit translation or cause degradation of the mRNA, examples that cause enhanced translation of the mRNA transcript have been described as well. In these cases, the *trans*-encoded sRNAs bind a stem-loop that functions as a transcription terminator (anti-termination), or base pairing induces structural rearrangements in the mRNA that result in exposure of the SD sequence. (c) Some sRNAs are known that modulate the functionality of a nucleic acid-binding protein (NABP). These sRNA molecules function by sequestering the NABP, due to the presence of a motif or structure to which the NABP binds with high affinity.

sRNA in that they are derived from loci other than their targets. Consequently, the base pairing region between the sRNA and the target mRNA contains only partial complementarity. In most cases, the sRNA binds the mRNA sequence at its 5' end to mask the SD sequence and/or start codon (Figure 2(b)). However, *trans*-encoded sRNAs that enhance translation or transcription of the mRNA transcript through an anti-antisense mechanism have been described as well. In this case, base pairing of the sRNA leads to structural rearrangements in the mRNA that result in exposure of the SD sequence or in anti-termination. The best-studied example is the regulation of RpoS, an alternative σ -factor, by the sRNAs DsrA and RprA in *E. coli*. These two sRNAs activate RpoS translation through base pairing with an inhibitory stem of a hairpin present at the 5' end of the RpoS transcript. The genes that are transcribed in an RpoS-dependent manner are important for entry in the stationary phase. The sRNA RprA is transcribed in response to envelope stress, whereas DsrA is transcribed especially at low temperatures.

A key player in regulation by *trans*-encoded sRNA is the protein Hfq, which functions as an RNA chaperone. The hexameric Hfq binds and protects sRNA and is required for the formation of low-complementarity duplexes between the *trans*-encoded sRNA and the target mRNA. In *E. coli* *hfq* mutant strains *trans*-encoded sRNAs fail to exert their regulatory effect. Hfq is probably also involved in melting secondary structures of the target mRNA, or serves as a platform for the sRNA-mRNA interaction only.

A single *trans*-encoded sRNA can have multiple targets, because full complementarity is not required. In some cases

a single sRNA targets multiple genes involved in the same pathway. For example, in *E. coli* the sRNA RyhB is transcribed in response to iron limitation. RyhB targets the mRNA of different operons encoding iron-binding proteins, including *bfr* (bacterioferritin), *sdh* (succinate dehydrogenase), and *sodB* (superoxide dismutase). The Hfq-dependent base pairing between these mRNAs and RyhB leads to RNaseE-dependent degradation of both the mRNA and RyhB. By downregulating the levels of these iron-binding proteins, the availability of iron for essential cellular processes increases.

sRNAs That Bind Protein to Modulate Their Activity

Few examples are known of sRNAs that bind proteins to modulate their activity. These sRNA molecules contain a motif or structure that causes high affinity binding by a specific nucleic acid-binding protein, leading to sequestration of the latter (Figure 2(c)). For example, the σ^{70} -containing RNAP has high affinity for the 6S sRNA due to its conserved secondary structure that is highly reminiscent of the DNA conformation in an open promoter. During exponential growth, the majority of transcription is carried out by RNAP containing σ^{70} . However, when cells enter the stationary phase upon nutrient depletion, alternative sigma factors associate with the core RNAP leading to transcription switching. The 6S sRNA facilitates transcription by alternative sigma factors, since the high affinity binding of σ^{70} to 6S leads to sequestration of the former. As a result, transcription from certain σ^{70} promoters

is inhibited, and instead, transcription mainly takes place from σ^S promoters.

In *E. coli* two sRNAs, CsrB and CsrC, regulate CsrA protein activity in an analogous manner. Dimers of CsrA act as global posttranscriptional regulators of carbon metabolism. These dimers exert their regulatory function by binding target mRNA sequences with GGA motifs in their 5'-UTR. As CsrB and CsrC, each contain multiple GGA motifs, an increase in their levels leads to sequestration of CsrA.

CRISPR-Based Defense against Infection

A more recently discovered class of prokaryotic sRNA molecules is involved in defense against invaders, such as viruses and conjugative plasmids. The hallmark of this defense system are loci of CRISPR with often adjacently located CRISPR-associated (*cas*) genes. CRISPR loci are present in ~40% and ~90% of the sequenced bacterial and archaeal genomes, respectively, and consist of repeating sequences of typically 30 nt that are separated by invader-derived sequences of similar length, called spacers. The presence of a spacer provides specific immunity against infective elements with a complementary sequence. The size of one CRISPR locus ranges from a few to hundreds of spacer/repeat units. CRISPR loci evolve rapidly in natural ecosystems through frequent spacer acquisition (known as 'CRISPR adaptation') and spacer loss. In fact, a CRISPR array can be seen as a record of recent infections that a lineage has overcome.

The CRISPR-based immune response requires transcription of the CRISPR locus into a long pre-crRNA molecule, and subsequent maturation to yield small crRNAs that serve to guide the defense. The protein machinery needed for maturation of the crRNA transcript and for subsequent CRISPR interference is encoded by the *cas* genes. Over 45 families of *cas* genes exist, that are found in nine typical combinations, known as 'subtypes'. Each subtype is named after a representative organism (e.g., *E. coli* K12 has subtype E). A single subtype contains between four to 20 *cas* genes that suffice for an anti-invader response. All CRISPR harboring species contain the core genes *cas1* and *cas2*, and usually a number of sub-type specific *cas* genes. Despite the variation in *cas* gene content between the subtypes, the mechanistic processes that underlie CRISPR-based defense appear to be variations on a theme.

The mechanism of CRISPR-based defense can be divided in three stages: (1) CRISPR adaptation, (2) CRISPR expression, and (3) CRISPR interference (Figure 3(a)). CRISPR adaptation takes place upon phage encounter and results in one or more new spacers being added to the existing CRISPR array. So far, CRISPR adaptation has only been studied in *Streptococcus thermophilus*, a lactic acid bacterium important in the dairy industry. This bacterium contains only four *cas* genes (*csn1*, *csn2*, *cas1*, and *cas2*). Inactivation of the species-specific *csn2* gene results in an inability to integrate new spacers, while gene disruption of *csn1* abolishes CRISPR interference, regardless of the presence of a virus-derived spacer sequence.

In the second stage, CRISPR expression, the array is transcribed into a long pre-crRNA molecule. In *E. coli* K12, eight *cas* genes are present, of which five (*casA*, *casB*, *casC*, *casD*, and

casE) encode proteins that form a multimeric protein complex named 'CRISPR associated complex for antiviral defense' (Cascade). CasE is a metal-independent endoribonuclease needed for the maturation of pre-crRNA. Cleavage by CasE within each repeat sequence generates mature crRNA molecules with the length of a single spacer and a single repeat. A single mature crRNA remains bound by the Cascade complex, forming a ribonucleoprotein complex. In *Pyrococcus furiosus*, maturation of the pre-crRNA is carried out by Cas6. In this species, the crRNAs are further trimmed at their 3' end. Comparison of the crystal structures of CasE from *Thermus thermophilus* and Cas6 from *P. furiosus* reveals that these proteins have a similar architecture. CasE has a catalytic histidine residue, whereas a catalytic triad (His, Tyr, Lys) has been proposed for Cas6. The mature crRNAs in *P. furiosus* are transferred to a Cas protein complex that is composed of six proteins of the so-called repeat-associated mysterious protein (RAMP) family, forming a ribonucleoprotein complex.

During the third stage of CRISPR-based defense, CRISPR interference, the target nucleic acids are recognized and inactivated. Evidence is accumulating that several CRISPR/Cas systems target viral and plasmid DNA. In *E. coli* and in *S. thermophilus*, the presence of virus-derived spacer sequences confers immunity and the complementary target sequence can be located either on the coding or on the template strand, which points towards DNA as the potential target. In *Staphylococcus epidermidis*, the targeting of a noncoding sequence on a plasmid confers immunity, whereas introducing a self-splicing intron in the protospacer, leading to a complementary sequence in the mRNA but not in the DNA, does not provide protection. Although target DNA degradation has not been observed yet, these convincing lines of evidence point toward DNA as the prime target in these CRISPR/Cas subtypes. In addition to the crRNA-loaded Cascade complex from *E. coli*, the predicted nuclease/helicase Cas3 is required for resistance against phage infection, possibly to carry out target DNA cleavage.

Although the aforementioned findings suggested that all CRISPR/Cas systems target DNA directly, a whole new variation on the CRISPR interference theme came from studies done on a system from *P. furiosus*. *In vitro* the RAMP containing Cas protein complex loaded with trimmed crRNA is capable of cleaving complementary single-stranded RNA. The observed cleavage takes place at a well-defined position, 14 nt from the 3' end of the crRNA. This sRNA-guided RNA degradation is mechanistically similar to RNAi-based defense against virus infections in plants, although no evolutionary relation exists between the proteins of CRISPR/Cas and RNAi (see below).

Prokaryotic Argonaute

In RNAi in eukaryotes, three major forms of small regulatory RNAs can be distinguished: microRNAs (miRNAs), small interfering RNAs (siRNAs), and piwi-interacting RNAs (piRNAs). miRNAs are endogenously encoded noncoding RNAs that regulate cellular processes through sequence-specific transcriptional or posttranscriptional gene silencing. Initially, it was thought that siRNAs are exclusively derived from extrachromosomal elements such as viruses, and sequence-specifically

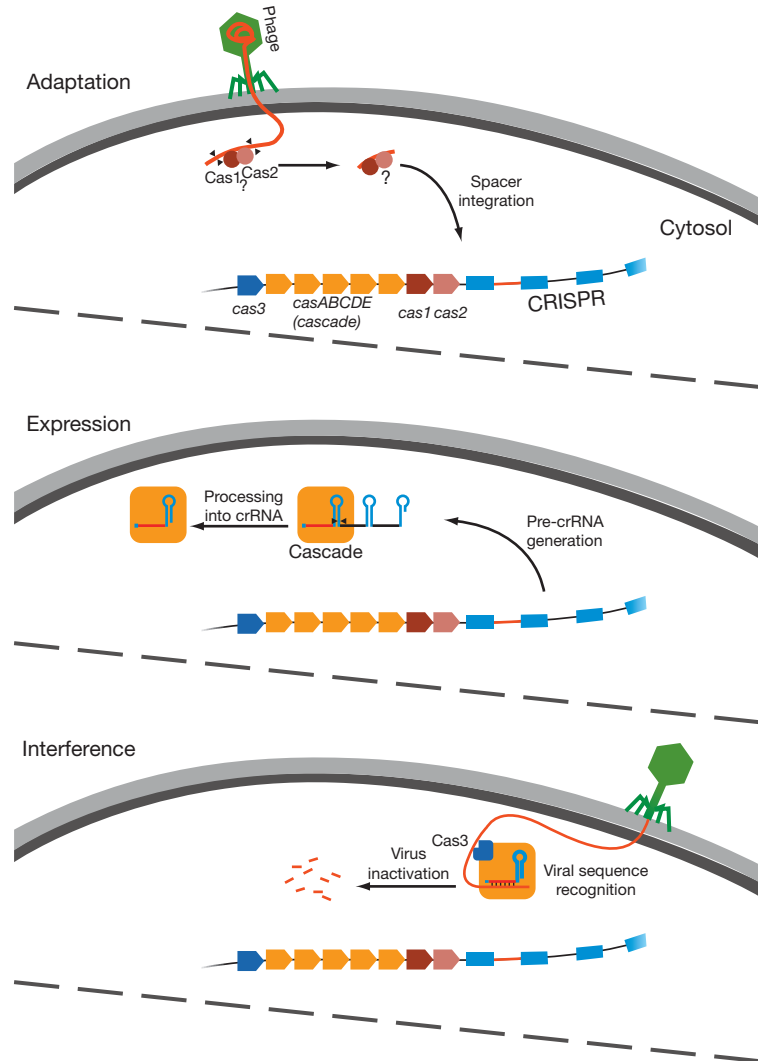


Figure 3 The *Escherichia coli* K12 CRISPR/*cas* locus consists of eight *cas* genes, which encode the machinery to carry out each of the three stages of the CRISPR immune response. Upon phage infection, Cas1 and/or Cas2 may mediate the integration of new spacer sequences in the existing CRISPR array. During the subsequent expression stage, the Cascade protein complex is formed and the CRISPR locus is transcribed into a long pre-crRNA molecule. Cascade mediates the maturation of the pre-crRNA transcript by cleavage within the repeat sequence, at the base of each stem-loop. The mature crRNA remains bound to Cascade, and the crRNA-loaded Cascade specifically binds the double-stranded viral DNA during the CRISPR interference stage. In concert with Cas3, the elimination of the virus is accomplished, probably through cleavage of the viral DNA.

suppress these elements. However, the recent discovery of endogenously encoded siRNAs has made the boundaries between miRNAs and siRNAs less well defined. The biogenesis and function of miRNAs and siRNAs are discussed elsewhere in this encyclopedia. Despite their distinct roles and origins, miRNAs and siRNAs share common themes in their mechanistic pathways. Two core proteins involved in both RNAi pathways are Dicer and Argonaute (note that it can be distinct paralogous proteins in one organism that are involved in either pathway). Dicer contains two RNaseIII domains that each cleave one strand of a double-stranded RNA (dsRNA) precursor molecule, generating short dsRNA. One of these strands (the guide) is then transferred to Argonaute, which facilitates base pairing with complementary RNA target sequences. miRNAs generally display imperfect base pairing while siRNAs generally display perfect base pairing with the target transcript.

The P-element-induced wimpy testis (PIWI)-domain of Argonaute, which shares structural homology to RNaseH family proteins, can cleave the complementary target RNA in case of perfect base pairing. Subsequent recycling of the sRNA-protein complex enhances the silencing effect.

Argonaute is the only protein component of RNAi of which homologs can be found in prokaryotes. Prokaryotic *argonaute* genes have been identified in ~80 sequenced genomes. Interestingly, crystal structures and *in vitro* activity assays of these prokaryotic Argonaute proteins provided the RNAi field with a lot of insight in the mechanism of target binding and cleavage. However, their role in prokaryotes remains elusive. It has recently been hypothesized that they are involved in antiviral (and anti-plasmid) defense, which is supported by several observations: (1) in eukaryotes many homologous Argonaute proteins are involved in

antiviral defense, (2) phylogenetic analysis of prokaryotic *argonaute* gene sequences suggests extensive horizontal gene transfer (HGT), which is typically observed for virus defense systems, (3) comparative gene neighborhood analysis revealed enriched association of *argonaute* genes with known phage-defense genes such as restriction endonucleases, (4) in virus-infected *T. thermophilus* cultures, the expression of the *ago* gene, as well as the *cas* genes is enhanced, and (5) Argonaute proteins from *Aquifex aeolicus* and *T. thermophilus* display DNA- and RNA-guided nuclease activity *in vitro*.

Despite its putative role in antiviral defense, the exact mechanism of virus defense remains to be elucidated. Small guide RNAs could direct base pairing and target cleavage analogously to RNAi in eukaryotes. However, prokaryotic Argonaute shows a higher affinity for guide-DNA than for guide-RNA and contains both DNA- and RNA-guided nuclease activity. Despite *in vitro* RNA degradation, the association of Argonaute with putative dsDNA nucleases suggests that DNA is being targeted directly. More insight in this antiviral defense system, which is supposed to be based on small nucleic acids, is to be expected in the near future.

Conclusions

Altogether sRNA molecules have diverse functions in prokaryotes (both bacteria and archaea), ranging from regulation of gene expression, mRNA translation, and protein functionality, to guiding target recognition in the recently discovered CRISPR-based defense system. The presence of Argonaute homologs in prokaryotes suggests the existence of additional host defense pathways utilizing small nucleic acids.

See also: **Bioenergetics:** Transient Receptor Potential Channels; **Metabolism Vitamins and Hormones:** Regulation of Gene

Transcription by Hypoxia-Inducible Factor 1; Roles of Micro-RNAs in Metabolism; **Molecular Biology:** DNA Modification, Restriction Endonucleases: Type I Enzymes; DNA Restriction and Modification: Type II Enzymes; DNA Restriction and Modification: Type III Enzymes; DNA Sequence Recognition by Proteins; Gene Expression in Bacterial Systems: The *trp* Operon and Attenuation; MicroRNAs in Eukaryotes; Riboswitches; Ribozymes and Evolution; Translation Initiation in Bacteria: Factors and Mechanisms; **Protein/Enzyme Structure Function and Degradation:** Allosteric Regulation.

Further Reading

- Babitzke P and Romeo T (2007) CsrB sRNA family: Sequestration of RNA-binding regulatory proteins. *Current Opinion in Microbiology* 10(2): 156–163.
- Barrangou R, Fremaux C, Deveau H, et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819): 1709–1712.
- Brouns SJ, Jore MM, Lundgren M, et al. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321(5891): 960–964.
- Fozo EM, Hemm MR, and Storz G (2008) Small toxic proteins and the antisense RNAs that repress them. *Microbiology and Molecular Biology Reviews* 72(4): 579–589.
- Frohlich KS and Vogel J (2009) Activation of gene expression by small RNA. *Current Opinion in Microbiology* 12(6): 674–682.
- Henkin TM (2008) Riboswitch RNAs: Using RNA to sense cellular metabolism. *Genes and Development* 22(24): 3383–3390.
- Karginov FV and Hannon GJ (2009) The CRISPR system: Small RNA-guided defense in bacteria and archaea. *Molecular Cell* 37(1): 7–19.
- Makarova KS, Wolf YI, van der Oost J, and Koonin EV (2009) Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biology Direct* 4: 29.
- Marraffini LA and Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Reviews Genetics* 11(3): 181–190.
- Narberhaus F, Waldminghaus T, and Chowdhury S (2006) RNA thermometers. *FEMS Microbiology Reviews* 30(1): 3–16.
- van der Oost J, Jore MM, Westra ER, et al. (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends in Biochemical Sciences* 34(8): 401–407.
- Waters LS and Storz G (2009) Regulatory RNAs in bacteria. *Cell* 136(4): 615–628.

Sodium–Calcium Exchanger: Structural Aspects

M Ottolia, D A Nicoll, and K D Philipson, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

C2 domains Domains of proteins involved in plasma membrane targeting. They consist of a β -sandwich motif composed of eight β -strands that coordinates two or three calcium ions.

Cysteine accessibility mutagenesis A technique in which single residues are mutated to cysteine and reactivity to

membrane impermeable sulfhydryl reacting agents is determined.

FLAG antibody An antibody to an epitope with the amino acid sequence DYKDDDDK.

Nuclear magnetic resonance (NMR) and X-ray crystallography Methods for determining the arrangement of atoms in proteins.

Introduction

Ca^{2+} is one of the most abundant metals in living organisms, and Ca^{2+} movements across cell membranes play a central role in cell signaling. Among the plasma membrane proteins transporting Ca^{2+} across the membrane is the Na^{+} – Ca^{2+} exchanger (NCX), which utilizes the electrochemical gradient of Na^{+} to extrude Ca^{2+} from the cell. During each reaction cycle, one Ca^{2+} is moved out of the cell while three Na^{+} are transported into the cytoplasm. This is considered the normal mode of NCX (or forward mode). However, under appropriate Na^{+} and Ca^{2+} gradients NCX can catalyze Ca^{2+} influx (reverse mode).

NCX is a ubiquitous protein found in tissues such as kidney, smooth muscle, and brain. However, NCX is most abundant in the heart where its fundamental role in contractile events has been extensively investigated. A milestone in the studies of NCX was its cloning in 1990 which, in conjunction with the development of the giant patch technique by Hilgemann, has allowed a detailed analysis of both its biophysical properties and physiological function.

The emphasis of this article is on recent developments regarding the structure/function relationships of the cardiac exchanger (NCX1). Excellent books and reviews covering the topics not discussed here are available to the reader interested in learning more about NCX.

The NCX1 Membrane Topology

The topology of NCX1 has been extensively investigated using a combination of cysteine accessibility mutagenesis, epitope mapping, and functional analysis of mutants. NCX1 is composed of nine transmembrane segments (TMSs) with a large, cytoplasmic loop between TMS 5 and 6. Also, a pair of reentrant loops in the regions known as the α -repeats has been proposed. Much of this information comes from cysteine accessibility mutagenesis in which cysteines are introduced at various locations in NCX1. The addition of membrane impermeant sulfhydryl-modifying reagents to the extracellular or intracellular surface then makes it possible to determine on which side of the membrane the cysteine resides. Figure 1 is a cartoon representation of NCX1 with residues that have been mutated to cysteine.

Residues highlighted in red are sensitive to extracellular sulfhydryl modification while residues highlighted in blue are sensitive to intracellular modification. One residue, at position 842, is accessible from either side of the membrane. Epitope mapping experiments have been conducted by first obtaining antibodies that bind to a known region of NCX1 then examining whether the antibody binds at the intracellular or extracellular surface of NCX. In a first report, monoclonal antibodies were raised to the full length NCX1 then mapped to the large hydrophilic loop of NCX1 by screening an expression library composed of fragments of NCX1. A second approach was to generate antibodies against small peptide fragments of NCX1 and a third approach was to introduce the epitope to the FLAG antibody at known sites of NCX1. The locations of the epitopes were mapped to the extracellular surface by direct antibody binding to cells expressing NCX1 or to the intracellular surface if the antibody was only able to bind the cells after permeabilization. In Figure 1, the approximate center of each epitope is circled in blue. Both the cysteine accessibility and epitope mapping techniques support the model for NCX1 shown in Figure 1. In this model, NCX1 is composed of nine α -helical TMS of 17–24 residues in length with a large cytoplasmic loop between TMS 5 and 6 and reentrant loops between TMS 2 and 3 and between 7 and 8. At least one residue in each of the loops connecting TMSs has been mapped to an intracellular or extracellular surface. The sole observation that does not support the model is for epitope mapping at position 37. In that experiment, a FLAG epitope inserted at residue 37 of NCX1 was only accessible after permeabilizing the cell with the detergent Triton X-100. Given other data that show the amino terminus of NCX1 is extracellular, it seems most likely that residue 37 is extracellular but with limited accessibility in the absence of detergent.

NCX1 Helix Packing

In a series of studies, the packing arrangement of TMSs has been examined by sulfhydryl cross-linking. By introducing pairs of cysteines into a cysteine-free NCX1, treating with sulfhydryl cross-linking reagents, and examining mobility shifts on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), it is possible to determine which TMSs are near

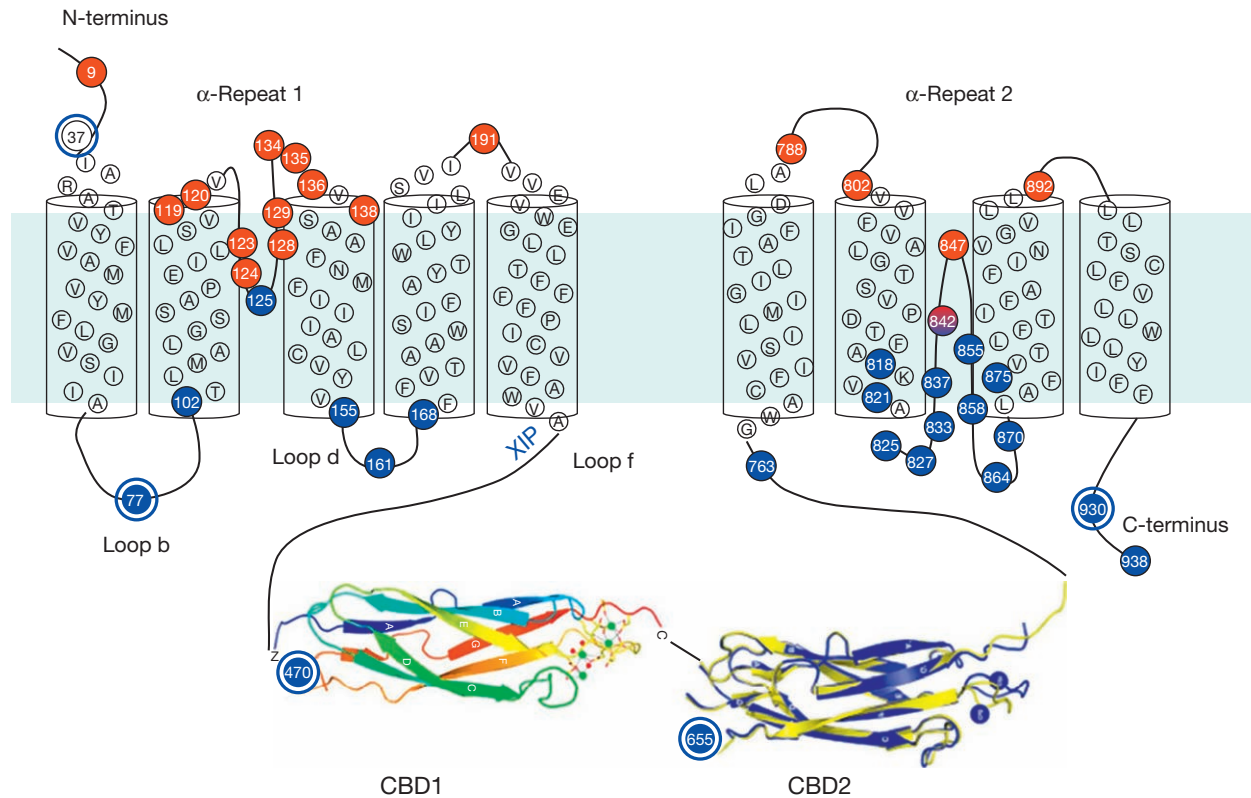


Figure 1 Membrane topology of NCX. Shown are NCX residues and their secondary organization. This model is based on cysteine accessibility and epitope mapping studies. Residue numbers within the red filled circles are accessible to extracellular sulfhydryl reagents while positions indicated by blue filled circles are affected by intracellular application of sulfhydryl reagents. The open blue circles indicate locations of epitopes of NCX. The crystal structures of the exchanger Ca^{2+} -binding domains (CBD1 and CBD2) are also presented. They are shown with an arbitrary orientation since the structure of the arrangement of the two CBDs has not been resolved yet. In the case of CBD2, the Ca^{2+} -free (yellow) and Ca^{2+} -bound (blue) forms are shown. Ca^{2+} ions are indicated in green (CBD1) and blue (CBD2). Reproduced from Nicoll DA, Sawaya M R, Kwon S, et al. (2006) The crystal structure of the primary Ca^{2+} sensor of the $\text{Na}^+/\text{Ca}^{2+}$ exchanger reveals a novel Ca^{2+} binding motif. *Journal of Biological Chemistry* 281(31): 21577–21581 and Besserer GM, Ottolia M, Nicoll DA, et al. (2007) The second Ca^{2+} -binding domain of the $\text{Na}^+-\text{Ca}^{2+}$ exchanger is essential for regulation: Crystal structures and mutational analysis. *Proceedings of the National Academy of Sciences of the United States of America* 104(47): 18467–18472.

one another in the folded protein. Extensive cross-links between the two halves of NCX1 have been observed and a model for helix packing, calculated as the smallest sum of the length of cross-links, is shown in Figure 2(a).

In this model, with the exception of TMS 6, the two halves of NCX are segregated. TMS 2, 3, and 7, which comprise portions of the α -repeats, regions of sequence similarity that are likely the result of a gene duplication event, are opposite one another. Also functionally important residues (e.g., Glu¹¹³, Ala¹⁴⁰, and Asp⁸¹⁴) are directed inward, away from the lipid bilayer (not shown).

NCX1 Dimerization

The first evidence that NCX1 may not exist solely as a monomer came from studies where cyan fluorescent protein (CFP)- and yellow fluorescent protein (YFP)-tagged NCX1 proteins were coexpressed and Ca^{2+} -sensitive interactions between the fluorescent CFP and YFP tags could be observed. Subsequently, we detected cross-linking between certain single cysteine-containing NCX proteins. Thus, NCX can exist as a dimer or

perhaps as a higher-level multimer in the membrane. Neither of these studies was able to determine the biophysical or physiological implications of NCX1 multimers.

Function of NCX Transmembrane Domains

There has been considerable study on the effects of mutations in the TMSs of NCX, especially those TMSs that form the α -repeat segments, TMSs 2, 3, and 7, as well as the putative reentrant loops. In one study, mutations to several residues in each of the putative reentrant loops of the α -repeats altered the affinity of NCX for extracellular, transported Ca^{2+} . In another study, some of the same mutations in the putative α 1-repeat reentrant loop were examined with respect to intracellular ions and found to have altered affinities for intracellular Na^+ but not intracellular Ca^{2+} . Additional mutations in TMS2 and TMS3 cause altered affinity for intracellular Na^+ affinity and/or selectivity but little or no effect on intracellular Ca^{2+} affinity. Particularly, replacement of residues A140 and I147 within TMS3 drastically decreased the sensitivity of NCX for transported Na^+ (see Figures 2(b)–2(d)), indicating an important

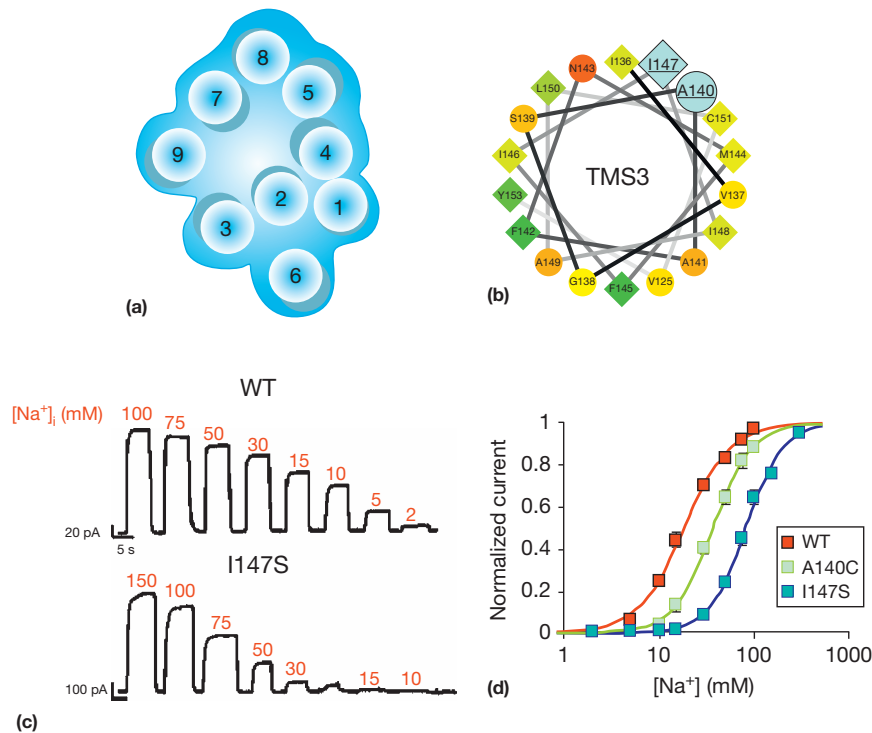


Figure 2 NCX helix packing. Panel (a) shows a top view of the packing arrangements of the NCX transmembrane segments (TMSs). This model is based on extensive crosslinking studies showing proximity of TMSs. Pairs of cysteine were introduced into a cysteineless exchanger and intramolecular cross-linking was detected as changes in electrophoretic mobility. In this model, TMS 2, 3, and 7, which encompass portions of the α -repeats, are opposite to each other. These regions are important for transport of Na^+ and Ca^{2+} ion, as mutagenesis studies reveal. Panel (b) shows that residues Ala¹⁴⁰ and Ile¹⁴⁷ of TMS3 face the same side of the α helix. Replacement of either of these two residues significantly alters the transport properties of NCX. Panel (c) illustrates outward currents recorded in the presence of the indicated Na^+ concentrations for WT and I147S exchangers, while the corresponding Na^+ -dependency curve is shown in panel (d). The normalized current vs. Na^+ concentration for mutant A140C is also presented.

role of these residues in ion transport. Mutations of several hydrophilic residues in TMSs 2, 3, 5, 6, and 7 result in inactive NCXs. TMS5 is especially interesting because the mutated residues correspond to a region with sequence similarity to the Na^+/K^+ adenosine triphosphatase (ATPase).

Structure–Function of NCX Cytoplasmic Domains

Cytoplasmic Loop f

NCX is comprised of 938 amino acids, ~550 of which make a large cytoplasmic loop (loop f) located between transmembrane segments 5 and 6 (see Figure 1). This portion of the exchanger is not required for transport (a mutant lacking most of this loop, $\Delta 240$ –679, still retains exchange activity) but it plays an important role in regulation being the target of factors such as cytosolic Na^+ and Ca^{2+} , phosphatidylinositol 4,5-bisphosphate (PIP_2), anionic lipids, and protons. Insights into the structure of this portion of the exchanger are required to unveil its regulatory properties. Here are summarized the latest findings.

The Exchanger Inhibitory Peptide Region and Na^+ -Dependent Inactivation

At the N-terminus of this loop is found the exchanger inhibitory peptide (XIP) region. This is a stretch of hydrophobic and positive residues (residues 218–239) named after a synthetic

peptide (XIP) with identical sequence which potently inhibits exchanger activity. The endogenous XIP region is involved in several regulatory processes such as Na^+ -dependent inactivation (see Figure 3).

This process describes a slow decay of exchanger current due to the presence of high intracellular Na^+ concentrations. It is thought that the binding of cytosolic Na^+ to transport sites drives the exchanger into a nonfunctional state, without involving alternative Na^+ -binding domains. However, the large cytoplasmic loop is involved in this regulation since electrophysiological data show that replacement of strategic residues within the XIP region alters both the rate and extent of inactivation. This is shown in Figure 3(a).

In addition to being involved in the Na^+ -dependent inactivation process, the XIP region is the target for PIP_2 and anionic long-chain fatty acyl CoA esters (acyl CoAs). Anionic lipids stimulate the exchanger activity by counteracting the Na^+ inactivation, as demonstrated in Figure 3(b).

Despite these advances, the molecular details and physiological roles of Na^+ -dependent inactivation are still unclear. Possibly, the XIP region is in contact with the plasma membrane preventing Na^+ -dependent inactivation. Increases in bound Na^+ could destabilize this interaction and trigger inactivation. Both PIP_2 and acyl-CoAs may stabilize the XIP region in a plasma membrane-bound state. This model is speculative. Additionally, since the Na^+ -dependent inactivation occurs primarily at high Na^+ concentrations, its

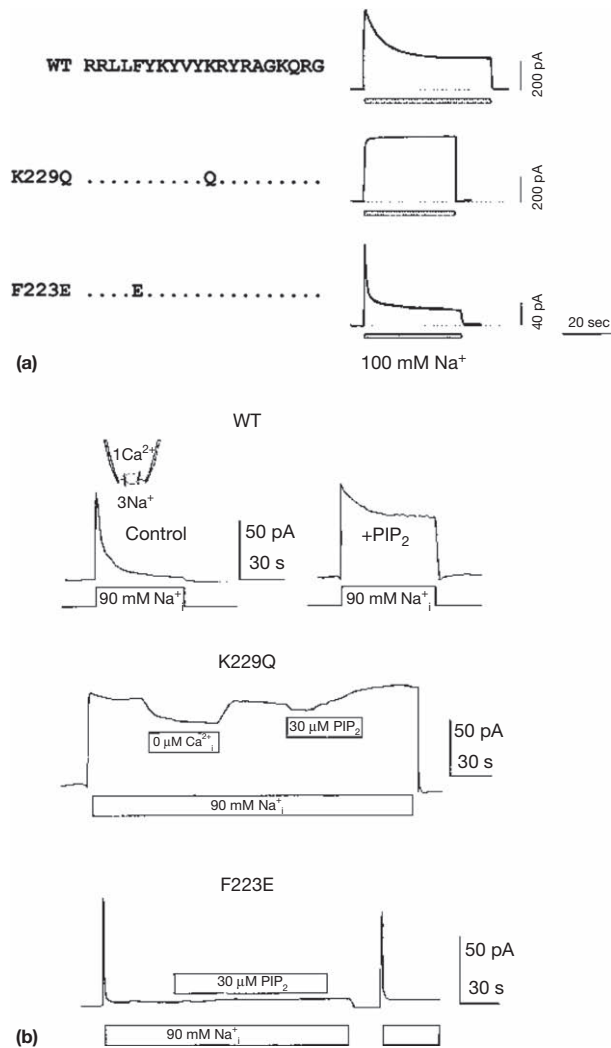


Figure 3 The XIP region is involved in the Na⁺-dependent inactivation and PIP₂ regulation. Panel (a) shows the amino acid sequence of the XIP region for the WT and mutant exchangers and their corresponding outward currents. Note the decay of the current over time due to the presence of high cytoplasmic Na⁺ (Na⁺-dependent inactivation). The inactivation process is abolished in mutant K229Q and enhanced in mutant F223E. Panel (b) shows the effect of PIP₂ on NCX outward currents recorded using the giant patch technique. Application of 30 μM PIP₂ removed the Na⁺-dependent inactivation in the WT exchanger, and had no or very little effect on mutants F223E and K229Q. From these and other experiments, it was concluded that Na⁺-dependent inactivation is required for PIP₂ effects. (a) Modified from Matsuoka S, Nicoll DA, He Z, and Philipson KD (1997) Regulation of cardiac Na⁺-Ca²⁺ exchanger by the endogenous XIP region. *Journal of General Physiology* 109(2): 273–286. (b) Modified from He Z, Feng S, Tong Q, Hilgemann DW, and Philipson KD (2000) Interaction of PIP(2) with the XIP region of the cardiac Na⁺/Ca²⁺ exchanger. *American Journal of Physiology – Cell Physiology* 278(4): C661–C666.

physiological role is unclear. Cytoplasmic Na⁺ concentrations may not reach the Na⁺ levels necessary to induce inactivation of the exchanger. Likewise, the role of PIP₂ in regulating the Na⁺-Ca²⁺ exchanger *in vivo* is unclear. Further investigations are needed to clarify the function of Na⁺-dependent inactivation and its regulation by lipids.

Linker Domain 1

Residues 239–371, immediately following the XIP region, may be important in transducing the movements of the Ca²⁺-binding domains (CBD1 and CBD2; see below) to the transmembrane segments. Supporting this hypothesis are fluorescence resonance energy transfer (FRET) studies showing large conformation changes within linker domain 1 upon binding of Ca²⁺ to CBD1. Still, no detailed functional studies of this portion of exchanger have been conducted leaving the role of these residues unknown.

The Ca²⁺ Binding Domains

The modulation of exchanger activity by cytoplasmic Ca²⁺ is well documented. Since a first observation about 30 years ago by DiPolo, the field has greatly advanced. Regulatory Ca²⁺ promotes exchanger activation and rescues it from the Na⁺-dependent inactivation. In order to stimulate transport activity, cytoplasmic Ca²⁺ binds to two Ca²⁺-binding domains that are distinct from the Ca²⁺ transport site (see Figure 1). These regions follow linker domain 1 and their structures have been resolved by both nuclear magnetic resonance (NMR) and X-ray crystallography. The first Ca²⁺-binding domain (CBD1) encompasses residues 371–501 and the second CBD2 spans amino acids 501–689. Each CBD is organized into a seven-stranded antiparallel β-sandwich with Ca²⁺ ions bound at the head of the sandwich and an unstructured F–G loop at the opposite end. The principal difference of the two CBDs lies in the heads and F–G loops. CBD1 binds four Ca²⁺, while CBD2 only binds two (Figure 4(a)).

The F–G loops of both CBDs are unstructured, but the F–G loop of CBD2 is a site of alternative splicing. It has been proposed that this alternative splicing results in exchangers with different properties but we have found that removing the F–G loops has no effect on the biophysical properties of NCX1. Thus, it seems more likely that the CBD2 F–G loop is involved in tissue-specific localization or interaction with other proteins. The organization of the conjunct CBDs is at present unknown.

The role of these domains in conferring Ca²⁺ sensitivity to the exchanger has been investigated by mutagenesis and electrophysiology. Replacement of residues coordinating Ca²⁺ at sites 3 and 4 of CBD1 significantly decreases the sensitivity of NCX to regulatory Ca²⁺, while a single-point mutation in CBD2 (E516L) eliminates Ca²⁺ regulation (Figures 4(b) and 4(c)). Still, how the binding of Ca²⁺ is triggered into exchanger activation continues to puzzle investigators. The expectation is that binding of Ca²⁺ to CBD1 and CBD2 results in a change of conformation which translates into NCX activation. Indeed, both biochemical and FRET studies indicate that CBD1 moves in response to Ca²⁺. In contrast, the structure of CBD2 appears similar in both the Ca²⁺-free and -bound forms, although elimination of Ca²⁺ coordination by CBD2 deregulates the exchanger. Additional investigations are required to shed light into the mechanism underlying exchanger regulation. Nevertheless, events describing the underlying exchanger regulation have been postulated. In one hypothesis, a twist in the linker region connecting the two CBDs occurs upon Ca²⁺ binding to CBD1, with a resultant activation of the exchanger.

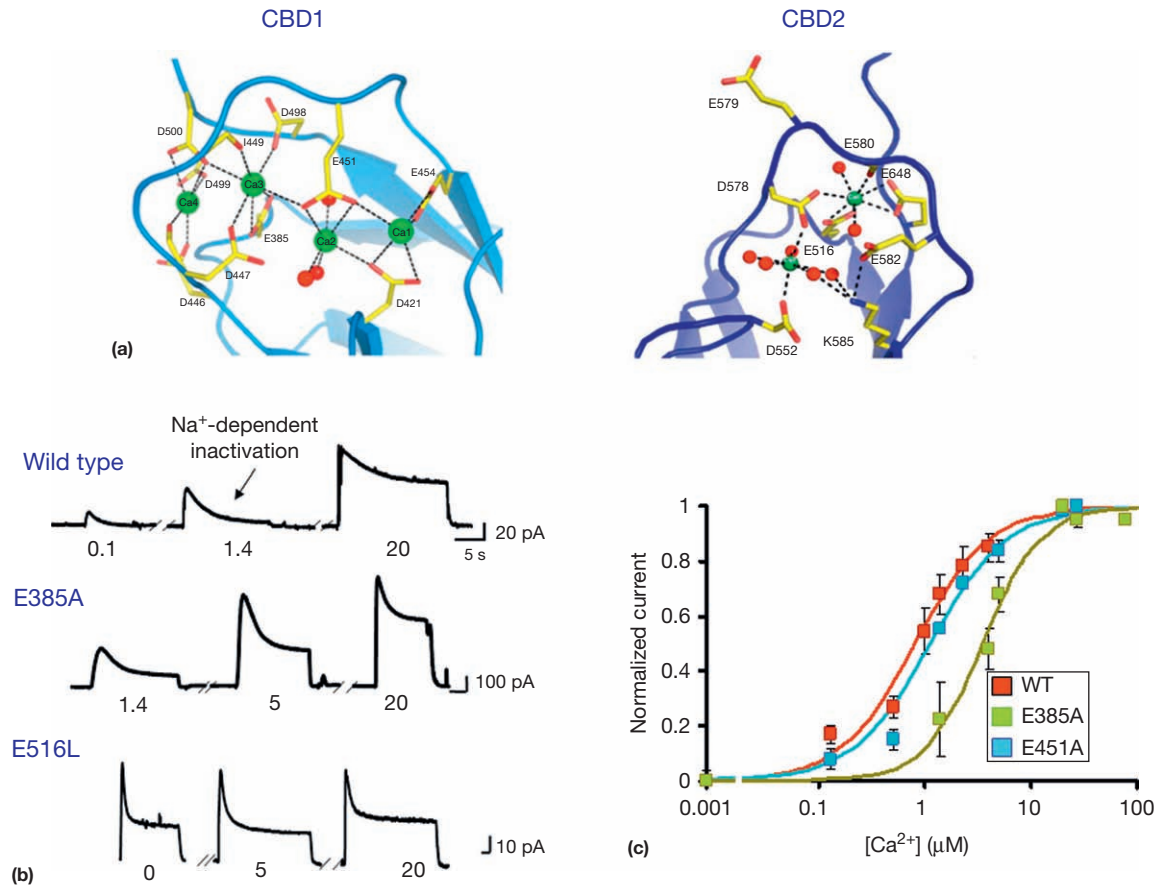


Figure 4 NCX Ca²⁺ regulation. Panel (a) shows the regions of CBD1 and CBD2 coordinating Ca²⁺. Note that CBD1 coordinates four Ca²⁺ and CBD2 only two. Outward currents recorded from WT NCX and exchangers with mutations in CBD1 (E385A) or CBD2 (E516L) are presented in panel (b). Traces were measured at the indicated cytoplasmic Ca²⁺ concentrations. Increases in cytoplasmic Ca²⁺ activate NCX current and decrease the extent of Na⁺-dependent inactivation. However, activation of mutant E385A required higher Ca²⁺ concentrations than WT. Mutation at position 516 caused the loss of Ca²⁺ regulation. These data underline the importance of both CBD1 and CBD2 in the regulation of NCX by cytoplasmic Ca²⁺. Finally, mutagenesis studies indicate that only Ca²⁺ coordinated by sites 3 and 4 of CBD1 are important for exchanger activation. As shown in panel (c), the affinity of WT and E451A (coordinating mainly Ca²⁺ at site 2 of CBD1) were not significantly different. In contrast, exchanger E385A (affecting coordination of Ca²⁺ at site 3) has decreased NCX affinity for cytoplasmic Ca²⁺. Reproduced from Nicoll DA, Sawaya MR, Kwon S, et al. (2006) The crystal structure of the primary Ca²⁺ sensor of the Na⁺/Ca²⁺ exchanger reveals a novel Ca²⁺ binding motif. *Journal of Biological Chemistry* 281(31): 21577–21581, with permission.

This movement correlates with a change in the electrostatic interaction between the two CBDs. Alternatively (or perhaps concurrently), upon binding of Ca²⁺, CBD2 moves with respect to some other region of the protein or the plasma membrane. This behavior would be similar to what is observed in the structurally similar C2 domains which are recruited to the plasma membrane upon Ca²⁺ coordination. Both theories require experimental validation leaving investigations on the regulation of the exchanger by Ca²⁺ open to new discoveries.

Linker Domain 2

Residues 689–768 link the second CBD to transmembrane segment 6. A portion of linker domain 2 was initially thought to be a transmembrane segment but cysteine scanning mutagenesis studies revealed it to be part of the large cytoplasmic loop. There is no information on the function of this region of the exchanger though insertion of green fluorescent protein (GFP) at this location disrupts the targeting of NCX to the

plasma membrane resulting in retention in the endoplasmic reticulum (unpublished observation). Since the insertion of GFP within the large cytoplasmic loop is well tolerated at other sites, this observation suggests a potential role of these residues in NCX trafficking.

Cytoplasmic Loops b and d

Residues 63–101 and 156–167 constitute cytoplasmic loops b and d, respectively. Few mutagenesis studies have been conducted within these regions of the exchanger and information is limited.

In loop b, replacement of Asn 101 with a Cys abolished the sensitivity of the transporter to regulation by cytoplasmic Ca²⁺ and Na⁺, revealing an important role of this loop in exchanger secondary regulation. Similarly, the histidine at position 165 of loop d is involved in the regulatory effects of Na⁺ and Ca²⁺ as mutations at this site drastically alter regulation. How these two loops exert their regulatory properties is unknown.

Interestingly, these loops flank TMSs 2 and 3 which are involved in ion translocation. Possibly, these cytoplasmic regions represent a structural link between ion translocation and regulation.

The C-Terminus

The C-terminus of NCX1 has also been shown to be intracellular. While no role has yet been assigned to this region, we have found that truncations and most insertions in this region result in nonfunctional exchangers. Thus, the C-terminus is likely important, though a specific role for this part of the exchanger remains to be defined.

Structure–Function of NCX Extracellular Domains

Most of the extracellular loops connecting TMSs are predicted to be short. Although amino acid substitutions are fairly well tolerated, we find that insertions into these regions result in nonfunctional exchangers. One exception is the amino terminus, which is about 40 amino acids long and contains a glycosylation site and a pair of cysteines capable of forming disulfide bonds with the extracellular loop connecting TMSs 6 and 7. The amino terminus can tolerate any number of insertions, deletions, and mutations with no apparent effect on NCX activity.

See also: **Bioenergetics: Membrane Transporters: Na⁺/Ca²⁺ Exchangers.**

Further Reading

- Besserer GM, Ottolia M, Nicoll DA, et al. (2007) The second Ca²⁺-binding domain of the Na⁺–Ca²⁺ exchanger is essential for regulation: Crystal structures and mutational analysis. *Proceedings of the National Academy of Sciences of the United States of America* 104(47): 18467–18472.
- DiPolo R and Beuge L (2006) Sodium/calcium exchanger: Influence of metabolic regulation on ion carrier interactions. *Physiological Reviews* 86: 155–203.

- Hilge M, Aelen J, and Vuister GW (2006) Ca²⁺ regulation in the Na⁺/Ca²⁺ exchanger involves two markedly different Ca²⁺ sensors. *Molecular Cell* 22(1): 15–25.
- Hilgemann DW (1990) Regulation and deregulation of cardiac Na⁺–Ca²⁺ exchange in giant excised sarcolemmal membrane patches. *Nature* 344: 242–245.
- Hilgemann DW, Collins A, Cash DP, and Nagel GA (1991) Cardiac Na⁺–Ca²⁺ exchange system in giant membrane patches. *Annals of New York Academy of Science* 639: 126–139.
- Hilgemann DW, Collins A, and Matsuoka S (1992) Steady-state and dynamic properties of cardiac sodium–calcium exchange. Secondary modulation by cytoplasmic calcium and ATP. *Journal of General Physiology* 100(6): 933–961.
- Hilgemann DW, Matsuoka S, Nagel GA, and Collins A (1992) Steady-state and dynamic properties of cardiac sodium–calcium exchange. Sodium-dependent inactivation. *Journal of General Physiology* 100(6): 905–932.
- Iwamoto T, Uehara A, Imanaga I, and Shigekawa M (2000) The Na⁺/Ca²⁺ exchanger NCX1 has oppositely oriented reentrant loop domains that contain conserved aspartic acids whose mutation alters its apparent Ca²⁺ affinity. *Journal of Biological Chemistry* 275(49): 38571–38580.
- Lytton J (2007) Na⁺–Ca²⁺ exchangers: Three mammalian gene families control Ca²⁺ transport. *Biochemical Journal* 406: 365–382.
- Ottolia M, Nicoll DA, and Philipson KD (2005) Mutational analysis of the alpha-1 repeat of the cardiac Na⁺–Ca²⁺ exchanger. *Journal of Biological Chemistry* 280(2): 1061–1069.
- Ottolia M, Nicoll DA, and Philipson KD (2009) Roles of two Ca²⁺-binding domains in regulation of the cardiac Na⁺–Ca²⁺ exchanger. *Journal of Biological Chemistry* 284(47): 32735–32741.
- Ottolia M, Philipson KD, and John S (2004) Conformational changes of the Ca²⁺ regulatory site of the Na⁺–Ca²⁺ exchanger detected by FRET. *Biophysical Journal* 87(2): 899–906.
- Nicoll DA, Longoni S, and Philipson KD (1990) Molecular cloning and functional expression of the cardiac sarcolemmal Na⁺–Ca²⁺ exchanger. *Science* 250(4980): 562–565.
- Nicoll DA, Ottolia M, Lu L, Lu Y, and Philipson KD (1999) A new topological model of the cardiac sarcolemmal Na⁺–Ca²⁺ exchanger. *Journal of Biological Chemistry* 274(2): 910–917.
- Nicoll DA, Sawaya MR, Kwon S, et al. (2006) The crystal structure of the primary Ca²⁺ sensor of the Na⁺/Ca²⁺ exchanger reveals a novel Ca²⁺ binding motif. *Journal of Biological Chemistry* 281(31): 21577–21581.
- Philipson KD, Nicoll DA, Ottolia M, et al. (2002) The Na⁺/Ca²⁺ exchange molecule: An overview. *Annals of New York Academy of Science* 976: 1–10.
- Qiu Z, Nicoll DA, and Philipson KD (2001) Helix packing of functionally important regions of the cardiac Na⁺–Ca²⁺ exchanger. *Journal of Biological Chemistry* 276(1): 194–199.
- Ren X, Nicoll DA, Galang G, and Philipson KD (2008) Intermolecular cross-linking of Na⁺–Ca²⁺ exchanger proteins: Evidence for dimer formation. *Biochemistry* 47(22): 6081–6087.

Somatostatin Receptors

A Schonbrunn, University of Texas Health Science Center – Houston, Houston, TX, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Cortistatins It refers to the short (rat CST-14 and human CST-17) and long (CST-29) biologically active products produced by proteolytic processing of the precursor peptide encoded by the cortistatin gene.

G protein One of a family of related heterotrimeric proteins which binds guanosine triphosphate (GTP) and guanosine diphosphate (GDP). The heterotrimeric forms, which are inactive, become activated at the plasma membrane by agonist-occupied G-protein-coupled receptors (GPCRs) that stimulate the binding of GTP to the G-protein alpha subunit and cause dissociation of the

alpha subunit from the beta-gamma subunit complex. When activated, the G-protein subunits regulate downstream effectors, such as ion channels and enzymes that generate second messengers. G proteins become inactivated by hydrolyzing GTP to GDP, which then permits the reassociation of the alpha subunit with the beta-gamma subunit complex.

Somatostatins Also called somatotropin-release-inhibiting factors. Refer to the 14-amino-acid (SS14) and 28-amino-acid (SS28) peptides which are produced by alternative proteolytic processing from a single 92-amino-acid precursor, called prosomatostatin.

Somatostatin receptors (sst receptors) comprise a family of five-membrane receptors which bind the neuroendocrine peptides somatostatin and cortistatin. These receptors exhibit the seven-transmembrane domain topology typical of G-protein-coupled receptors (GPCRs) and signal primarily through pertussis-toxin-sensitive G proteins. The different sst receptor subtypes are found in specific endocrine and exocrine tissues, including the pituitary and the pancreas, in addition to being widely distributed in the central and peripheral nervous systems and the gastrointestinal tract. They have important physiological roles in inhibiting hormone secretion, particularly the secretion of growth hormone, insulin, glucagon, and gastrin, inhibiting exocrine secretion by the pancreas and stomach, and modulating neuronal excitability and smooth muscle contraction. Many neuroendocrine tumors express high levels of sst receptors and, because of their ability to inhibit tumor-cell secretion and reduce tumor growth, somatostatin analogs are now used to target these receptors both for cancer therapy and diagnosis.

Physiological Sst Receptor Ligands

Somatostatin originally called somatotropin-release-inhibiting factor (SRIF) was discovered serendipitously over 30 years ago while investigators were hunting for the brain peptide responsible for stimulating the release of growth hormone, also called somatotropin. Surprisingly, they found that extracts of the hypothalamus, a specialized brain region which regulates pituitary hormone secretion, inhibited rather than stimulated the secretion of growth hormone. Using this inhibition as an assay, Guillemin and Schally and their co-workers purified the active factor from hypothalamic extracts and identified the 14 amino acid form of somatostatin, a discovery which contributed to their receiving the Nobel prize in 1977 for studies of hormone production in the brain.

Somatostatins are cyclic peptides that exist in two biologically active isoforms, 14 (SS14) and 28 (SS28) amino acids in length, and are produced by alternate proteolytic processing from a common precursor of 116 amino acids generated from

a single gene. These cyclic peptides are synthesized by specific endocrine, gastrointestinal, immune, and neuronal cells, as well as some tumors, and act either locally as paracrine, autocrine, or neuronal modulators or through the blood stream as hormones. Recently, the somatostatin gene has been shown to encode another peptide of 13 residues named neuronostatin, which is structurally unrelated to the two somatostatin isoforms. The variable amounts of neuronostatin and somatostatin produced in different tissues indicate differential and regulated processing of these peptides from the same precursor.

In addition to the peptides derived from the SS gene, the complementary DNA for a somatostatin-related peptide, cortistatin, was discovered in the process of characterizing region-specific rat brain messenger RNAs (mRNAs). Cortistatin was named for its predominant expression in the brain cortex and its ability to depress cortical neuronal activity. Like somatostatin, cortistatin is also synthesized as a precursor and is proteolytically processed into two biologically active products: a short (rat CST14 and human CST-17) and an amino-terminally extended (CST-29) form. Although cortistatin is produced from a different gene than somatostatin, the two peptides have 10 of their carboxy terminal 14 amino acids in common (Figure 1). Cortistatins are not only produced primarily by neurons in the central nervous system (CNS) but also have recently been found in immune cells, lymphoid tissues, bone marrow, and the pancreas.

Because all forms of somatostatin and cortistatin bind to the different sst receptor subtypes (sst's) with similar high affinities it is not surprising that the peptides share many functional properties. However, they also produce some distinct biological effects. For example, intracerebroventricular injection of cortistatin in rats increases slow-wave sleep but not rapid eye movement (REM) sleep whereas somatostatin injection increases REM sleep. Thus, the recent identification of a human cortistatin-selective GPCR that does not bind somatostatin (MrgX2 or Mas-related GPCR member X2, MRGPRX2) fulfills prior predictions for the existence of distinct cortistatin-specific receptors. However, this receptor binds a number of other apparently unrelated neuropeptides, such as proadrenomedullin

N-terminal peptide, so its physiological importance for cortistatin function is unclear. The specific roles of individual sst receptor subtypes as well as cortistatin-specific receptors in mediating particular physiological actions of these peptides remain an area of active investigation.

Biochemical and Molecular Characterization of Sst Receptors

Sst Receptor Subtypes

Five sst receptor genes, located on different chromosomes, have been identified and named *sst1–sst5* in the order of their discovery (Table 1). The coding regions of *sst1*, *sst3*, *sst4*, and *sst5* all occur within a single exon. However, in some mouse and rat tissues, the mRNA for *sst2* undergoes alternative splicing at the 3' end, producing two protein products. The more common *sst2A* variant is the product of the unspliced mRNA whereas the *sst2B* variant contains an alternative exon that encodes a different and shorter receptor carboxy-terminus. In contrast to rodents, the *sst2A* isoform is expressed almost exclusively in humans.

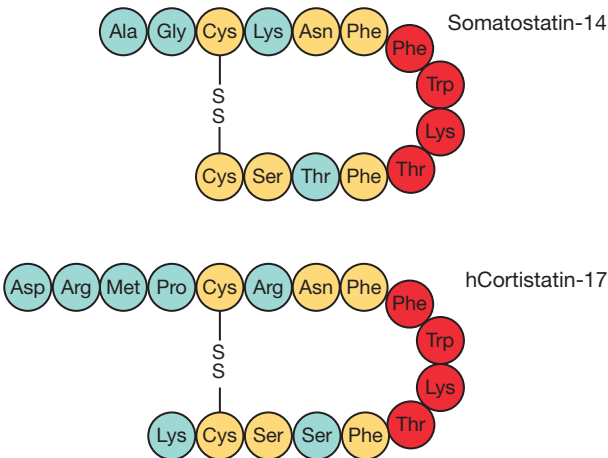


Figure 1 Structure of mammalian somatostatin-14 and human cortistatin-17. The amino acids in the red circles are required for high-affinity binding to somatostatin receptors. The amino acid differences between the two peptides are shown by the blue circles.

Sst receptors belong to the GPCR family and are predicted to contain seven- α -helical-transmembrane domains. Receptors for somatostatin have been cloned from a variety of mammals as well as from several nonmammalian species, including a number of fish. The human sst receptors exhibit between 40% and 57% amino acid identity with each other and sequence identity among the rat, mouse, and human orthologs of each receptor subtype is even higher. The greatest sequence similarity between sst receptor subtypes occurs within the transmembrane domains and hence these domains are thought to be involved in ligand binding. However, some extracellular loop residues have been shown to affect receptor subtype selectivity. Sequence differences in the intra- and extracellular domains appear to be responsible for the unique signaling and trafficking properties of individual sst receptor subtypes.

Based on sequence similarity, sst receptors have been subdivided into two groups: *sst1* and *sst4* receptors form the SRIF1 subgroup and *sst2*, *sst3*, and *sst5* constitute the SRIF2 subgroup. In addition to a closer phylogenetic relationship (Figure 2), members of each group share several functional properties including their affinity for short synthetic somatostatin analogs, such as octreotide and lanreotide, and sensitivity to agonist-induced receptor internalization (Table 1).

The GPCRs most closely related to the sst receptor family are GPR7 and GPR8, which encode receptors for two paralogous brain peptides neuropeptide B (NPB) and neuropeptide W (NPW) (Figure 2). The next most closely related gene family encodes the opioid receptors. The putative cortistatin receptor MrgX2 as well as the Urotensin II receptor (UR2R, GPR14) are more distantly related but share significant sequence identity with the sst receptor family.

Sst Receptor Structure

Sst receptors contain many of the structural features characteristic of GPCRs including consensus sites for asparagine-linked glycosylation within the amino-terminus and multiple Ser and Thr phosphorylation sites in the intracellular loops and carboxy-terminal tail (Table 1). All sst receptor subtypes except *sst3* also contain a conserved Cys residue in the carboxy-terminal region which provides a site for receptor palmitoylation although palmitoylation has yet to be demonstrated biochemically for any of the sst subtypes. However, *sst1*, *sst2*, *sst3*, and *sst5* are known to

Table 1 Properties of human somatostatin receptor subtypes

Property	<i>sst1</i>	<i>sst2A/B</i>	<i>sst3</i>	<i>sst4</i>	<i>sst5</i>
Chomosomal localization	14q13	17q24	22q13.2	20p11.2	16p13.3
NCBI reference sequence	NM_001049.2	NM_001050.2	NM_001051.2	NM_001052.2	NM_001053.3
Amino acids	391	369/356	418	388	364
Asn glycosylation sites	3	4	2	1	3
Receptor glycosylation	+	+	+	–	+
Palmitoylation site	+	+	–	+	+
Receptor phosphorylation	+	+	+	–	nd
SS-stimulated receptor internalization	Slow	Fast	Fast	Slow	Fast
Octreotide/lanreotide affinity	Low	High	Moderate	Low	Moderate
Pasireotide	Moderate	High	High	Low	High

nd, Not determined.

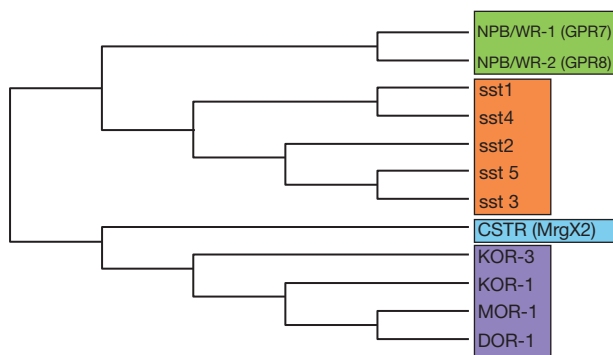


Figure 2 Sequence relationships between somatostatin receptors and closely related peptide receptors. The dendrogram was generated with the Clustal W program using MacVector's default parameters to align the peptide sequences. Human sequences were used in all cases and were obtained either from SwissProt or RefSeq. Somatostatin receptors: hsst1, P30872; hsst2A, P30874; hsst3, P32745; hsst4, P31391; hsst5, P35346. Cortistatin receptor: MrgX2, NP_473371. Opioid receptors: KOR-1, P41145; MOR-1, P35372; DOR-1, P41143; and KOR-3, P41146. Neuropeptide B/W receptors NPBWR-1 or GPR7, P48145; and NPBWR-2 or GPR8, P48146. The plot shows that hsst1 and hsst4 form one subgroup and hsst2, hsst3, and hsst5 form a second subgroup within the somatostatin receptor family.

be glycosylated from the observation that treatment of these receptors with peptide-N-glycosidase F, an enzyme which catalyzes the cleavage of N-glycosidically linked carbohydrate chains from asparagine, causes a large decrease in their apparent molecular weights. Such a shift in molecular weight was not observed for sst4, either because it is not glycosylated or because glycosylation at its single consensus site does not significantly affect its apparent molecular weight.

Receptor phosphorylation, a covalent modification important for GPCR regulation and signaling, has been examined biochemically for some but not all sst receptor subtypes. Somatostatin stimulates the incorporation of $^{32}\text{PO}_4$ into sst1, sst2A, and sst3 receptors and protein kinase C activation also increases phosphorylation of the sst1 and sst2A subtypes. However, phosphorylation of sst4 was not detected following agonist stimulation and sst5 phosphorylation has not been investigated biochemically.

Using phospho-site-specific antibodies seven Ser and Thr residues in the sst2A receptor C-terminus were shown to be phosphorylated within minutes of agonist stimulation. These complex phosphorylation reactions appear to be catalyzed by members of the GPCR kinase family. The sst2A receptor is also phosphorylated on at least two Ser residues by protein kinase C and this phosphorylation is independent of agonist occupancy. Phosphorylation of other sst receptor subtypes has not been studied in detail and we know little about the enzyme mechanisms involved.

Pharmacology of Sst Receptor Subtypes

The native somatostatin peptides exhibit little selectivity among the different receptor subtypes and are rapidly degraded in plasma (half-time <3 min). Thus, analogs with increased metabolic stability and greater receptor specificity have been developed for therapy.

The earliest pharmaceuticals to target sst receptors were short peptides, 6–8 amino acids in length, containing amino acids 7–10 of SS14, the region required for receptor binding (Figure 1). These agonists, including octreotide (SMS201-995), lanreotide (BIM-23014), and seglitide (MK-678), also contain D-amino acid substitutions and carboxy- or amino-terminal modifications to increase their resistance to proteolysis. All three of these truncated peptides bind selectively to the SRIF1 group of receptors, exhibiting subnanomolar affinity for sst2 and nanomolar affinities for sst3 and sst5 (Table 1). Further, octreotide, like somatostatin, does not bind to the MrgX2 cortistatin receptor. Presently, octreotide or lanreotide are used as standard-treatment options for patients with hormone-secreting pituitary adenomas and gastroenteropancreatic tumors and, in various formulations, comprise the only approved pharmaceuticals targeting sst receptors for cancer therapy.

In an effort to generate truly receptor subtype-specific analogs that are even more resistant to proteolytic degradation, Rohrer and co-workers at Merck screened combinatorial libraries to successfully identify the first highly selective, high-affinity, non-peptide agonists for each sst receptor subtype. Such reagents provide the possibility of oral bioavailability although nonpeptide sst agonists are yet to be approved for clinical use. Nonetheless, these specific compounds have proved to be valuable tools for identifying the physiological roles of individual sst receptor subtypes. The development of such nonpeptide somatostatin analogs may not only generate orally available drugs for current indications but also provide new therapies for diseases that require drug access to the CNS.

The recognition that many tumors concomitantly express several sst receptor subtypes has stimulated a search for stable somatostatin analogs that are less receptor selective. The hexapeptide analog SOM230 (Pasireotide) shows quite high affinity for four of the five sst receptor subtypes (Table 1) and is in late-phase clinical development. The nanopeptide analog KE108 appears to be truly universal because it binds to all five sst subtypes with nanomolar affinity. Because of their excellent metabolic stability these broad-spectrum agonists have significant therapeutic potential. Unexpectedly, however, both SOM230 and KE108 behave as biased agonists at the sst2A receptor, the most widely expressed sst receptor subtype in human tumors. Thus, they are full agonists mimicking somatostatin's action to inhibit cyclic adenosine monophosphate (cAMP) production, whereas they fully antagonize sst2A receptor signaling to other pathways, such as ERK phosphorylation and regulation of intracellular calcium. In addition, these analogs are less effective than somatostatin at inducing receptor internalization. The differences between the actions of full agonists, like octreotide, and biased agonists, like SOM230, at the sst2A receptor is likely to have therapeutic consequences, but these remain to be investigated in detail.

The first sst receptor antagonist CYN-154806 was described in 1996 and showed strong selectivity for the sst2 receptor subtype. Since then specific antagonists have been identified for all five sst receptors, although none are used clinically.

In addition to their therapeutic applications, analogs of somatostatin have been developed to localize and stage sst-receptor positive tumors *in situ*. The radiolabeled derivatives used for tumor diagnosis contain a chelating group covalently attached to a stabilized somatostatin peptide, such that its

receptor-binding properties are not compromised. The chelating group is then bound to a short-lived radioisotope, such as the gamma emitter ^{111}In , and injected into patients. After 24 h, accumulation of the radiolabel can be visualized in tumors expressing high levels of sst receptors by gamma camera scintigraphy, thus permitting localization of the tumors and their metastases. The analog in current clinical use is an indium-labeled octreotide derivative, [^{111}In -DTPA-D-Phe¹, Tyr²]octreotide, (OctreoScan) and hence is selective for the SRIF1 group of sst receptors, showing most-sensitive detection of sst2-expressing tumors. Although this analog, therefore, cannot be used to localize tumors expressing sst1 or sst4 receptors, new radiolabeled somatostatin analogs are currently being developed to improve both the sensitivity and breadth of this approach.

Sst Receptor Signaling

Sst receptors regulate a number of diverse signaling effectors, including adenylyl cyclase, phospholipases C and A2, calcium and potassium channels, protein and lipid kinases, and tyrosine and serine/threonine phosphatases. All sst receptors inhibit adenylyl cyclase via pertussis toxin-sensitive G proteins and, thus, decrease intracellular cAMP. However, other signal transduction pathways modulated by sst receptors depend on both the receptor subtype and the environment provided by the target cell, as well as the nature of the agonist.

The mechanism by which sst receptors inhibit secretion in endocrine cells and neurons is understood in some detail. In addition, to reducing intracellular cAMP levels, several sst receptors have been shown to reduce intracellular calcium levels in excitatory cells also via pertussis toxin-sensitive G proteins. The reduction in cytosolic calcium can result either from a stimulation of potassium channels, which hyperpolarizes the cell membrane and thereby decreases calcium influx through voltage-dependent calcium channels, or from direct calcium channel inhibition. The resulting decrease in intracellular cAMP and calcium concentrations together contribute to the inhibitory action of somatostatin on secretion: when either signaling pathway is blocked, the magnitude of somatostatin's inhibitory effect is reduced.

Somatostatin stimulates contraction of intestinal smooth muscle cells by inhibiting adenylyl cyclase and activating phospholipase C- β 3 via the alpha and beta-gamma subunits, respectively, of pertussis toxin-sensitive G proteins. The activated phospholipase catalyzes the hydrolysis of the membrane lipid phosphatidylinositol 4,5-bisphosphate to form the second messengers inositol trisphosphate (IP_3) and diacylglycerol. The binding of IP_3 to receptors on the sarcoplasmic reticulum results in the release of calcium from intracellular stores producing a rise in cytosolic calcium concentrations. The released calcium forms a complex with the protein calmodulin and this complex then activates myosin light-chain kinase (MLCK) to phosphorylate the light chain of myosin leading to smooth muscle contraction. Because protein kinase A (PKA) decreases the sensitivity of MLCK to calcium, somatostatin inhibition of cAMP formation facilitates the calcium effect by reducing the activity of PKA. These pathways are also activated by somatostatin in aortic vascular smooth muscle cells. Interestingly, somatostatin does not activate phospholipase C- β 3 in pituitary cells, even though the levels of this enzyme in the

pituitary appear to be similar to that in smooth muscle cells. The explanation for the signaling differences in these tissues remains to be elucidated, but could be due to differences either in the sst subtypes or the signaling machinery present.

In endothelial cells, somatostatin inhibits cell migration, stress fiber assembly, and cytoskeletal reorganization produced by thrombin and other stimulators. In humans, these effects are mediated by the sst1 receptor and have been implicated in somatostatin's antiangiogenic actions. Although the molecular steps involved remain to be identified, the mechanism includes an unusual, pertussis toxin-independent inhibition of Rho, a low-molecular-mass guanosine triphosphatase which plays a central role in regulating cytoskeletal organization.

The mechanisms by which somatostatin inhibits cell proliferation and stimulates apoptosis have garnered intense interest yet remain poorly understood at least partly because they vary in the different cell types in which they have been examined. In most, although not all, cells somatostatin activates the mitogen-activated protein kinase (MAPK) pathway and increases ERK1/2 phosphorylation by both pertussis toxin-sensitive and -insensitive mechanisms. However, such activation is often observed whether somatostatin inhibits or stimulates cell proliferation. Thus, some of the other effectors regulated by sst receptors must contribute to the final biological response. For example, in pancreatic acinar cells expressing sst2 receptors endogenously, somatostatin-induced growth arrest involves enhanced expression of the cyclin-dependent kinase inhibitor p27Kip which results from inhibition of the phosphatidylinositol-3-kinase (PI3K) pathway. By contrast, in sst2-transfected CHO cells, somatostatin induction of p27Kip appears to be dependent on stimulation rather than inhibition of PI3K. These, and other signaling pathways involved in cell proliferation and apoptosis are currently under investigation in order to elucidate the clinically important actions of somatostatin to inhibit tumor growth.

In summary, most, but not all, signaling by sst receptors involves the pertussis toxin-sensitive Gi/Go family. However, depending on the sst receptor subtype, the cellular environment, and the analog utilized, a spectrum of signaling pathways can be activated. The link between the different effectors regulated by sst receptors and particular biological responses is understood for some but far from all of somatostatin's actions.

Sst Receptor Regulation

Changes in sst receptor responsiveness during somatostatin exposure vary dramatically between tissues, depending both on the response being measured and on the nature of the target cell. In some instances, desensitization occurs within minutes following initiation of somatostatin treatment, whereas in others no desensitization is detected even after years of somatostatin analog exposure. For example, in chick sympathetic neurons, somatostatin inhibition of N-type Ca currents desensitize with a half-time of 3 min. In contrast, desensitization to pharmacologic doses of the somatostatin analog octreotide does not occur for months or years of therapy for pituitary tumors. The molecular basis for such dramatic differences in sst receptor regulation in different cell types remains poorly understood.

Nonetheless, studies in cultured cells have begun to elucidate the mechanisms by which individual sst receptor subtypes are regulated. As for many GPCRs, hormone treatment leads to the rapid phosphorylation of several sst receptor subtypes (Table 1). Interestingly, however, this phosphorylation appears to have different consequences for the different sst receptors. For example, in CHO cells, although both sst1 and sst2A receptors are phosphorylated and desensitized within minutes following somatostatin exposure, only the sst2A receptor is rapidly internalized. As the sites of sst2A receptor phosphorylation have been identified, site-directed mutagenesis has provided insight into the biological importance of the complex pattern of sst2A receptor phosphorylation. In this receptor, Thr phosphorylation sites are required for rapid agonist-stimulated sst2A receptor internalization and arrestin recruitment. By contrast, Ser phosphorylation sites are not required for arrestin recruitment and do not affect receptor endocytosis. However, Ser phosphorylation is important for sst2A receptor desensitization. The manner in which protein kinase C-mediated phosphorylation at specific receptor residues changes sst2A receptor function remains to be determined. The biochemical details and role of receptor phosphorylation in the regulation of sst1, sst3, and sst5 receptors remain to be investigated.

Although large gaps remain in our understanding of sst2A receptor regulation, the outlines of the processes involved are slowly becoming clear. Furthermore, the insights obtained from studies in cultured cells appear to apply to the behavior of sst2A receptors *in vivo*. For example, the sst2A receptor has been shown to be phosphorylated and internalized *in situ* in human tumors. As the molecular components of sst receptor desensitization are characterized, it will be interesting to determine how their functions differ in cells that are sensitive or resistant to somatostatin-analog-induced desensitization.

Conclusions

The five known sst receptors play important physiological roles in the regulation of the endocrine, neuronal, gastrointestinal, and immune systems. In addition, they are recognized to be important targets in the diagnosis and treatment of neuroendocrine tumors.

Although a great deal remains to be learned about many of somatostatin's biological functions, recent studies with knock-out mouse models lacking expression of somatostatin or cortistatin or one of the five sst receptor subtypes have begun to delineate the biological importance of the individual endogenous ligands and their receptors. Interestingly, all these knock-out models are viable and fertile. In fact, they usually exhibit no major overt abnormalities with the exception of the sst4 null mouse, which develops spontaneous seizures. However, specific defects have been described in neuronal, endocrine, and gastrointestinal functions with many of the somatostatin ligand and receptor knockout models. Although these studies have provided substantial insights into the importance of individual sst receptor subtypes, they have also revealed species-specific differences in the functions of these receptors. Fortunately,

studies of sst receptors in different species can now take advantage of the many receptor-specific agonists and antagonists that have become available and are likely to provide additional insight into the normal roles played by this physiologically and clinically important GPCR family. Our increasing molecular understanding of sst receptor signaling and regulation will undoubtedly facilitate the development of novel therapeutic interventions targeting this receptor family.

See also: **Protein/Enzyme Structure Function and Degradation:** Protein Palmitoylation; **Signaling:** G-Protein-Coupled Receptor Kinases and Arrestins.

Further Reading

- Ben-Shlomo A and Melmed S (2010) Pituitary somatostatin receptor signaling. *Trends in Endocrinology and Metabolism* 21: 123–133.
- Broglio F, Papotti M, Muccioli G, and Ghigo E (2007) Brain–gut communication: Cortistatin, somatostatin and ghrelin. *Trends in Endocrinology and Metabolism* 18: 246–251.
- Cescato R, Loesch KA, Waser B, et al. (2010) Agonist-biased signaling at the sst2A receptor: The multi-somatostatin analogs KE108 and SOM230 activate and antagonize distinct signaling pathways. *Molecular Endocrinology* 24: 240–249.
- Colao A, Faggiano A, and Pivonello R (2010) Somatostatin analogues: Treatment of pituitary and neuroendocrine tumors. *Progress in Brain Research* 182: 281–294.
- de Lecea L and Castano JP (2006) Cortistatin: Not just another somatostatin analog. *Nature Clinical Practice* 2: 356–357.
- De Martino MC, Hofland LJ, and Lamberts SW (2010) Somatostatin and somatostatin receptors: From basic concepts to clinical applications. *Progress in Brain Research* 182: 255–280.
- Florio T (2008) Molecular mechanisms of the antiproliferative activity of somatostatin receptors (SSTRs) in neuroendocrine tumors. *Frontiers in Bioscience* 13: 822–840.
- Gahete MD, Cordoba-Chacon J, Duran-Prado M, et al. (2010) Somatostatin and its receptors from fish to mammals. *Annals of the New York Academy of Sciences* 1200: 43–52.
- Kao YJ, Ghosh M, and Schonbrunn A (2011) Ligand-dependent mechanisms of sst2A receptor trafficking: Role of site-specific phosphorylation and receptor activation in the actions of biased somatostatin agonists. *Molecular Endocrinology* 25: 1040–1054.
- Kumar U and Grant M (2010) Somatostatin and somatostatin receptors. *Results and Problems in Cell Differentiation* 50: 137–184.
- Liu Q, Bee MS, and Schonbrunn A (2009) Site specificity of agonist and second messenger-activated kinases for somatostatin receptor subtype 2A (sst2A) phosphorylation. *Molecular Pharmacology* 76: 68–80.
- Reubi JC and Maecke HR (2008) Peptide-based probes for cancer imaging. *Journal of Nuclear Medicine* 49: 1735–1738.
- Schonbrunn A (2008) Selective agonism in somatostatin receptor signaling and regulation. *Molecular and Cellular Endocrinology* 286: 35–39.
- Tulipano G and Schulz S (2007) Novel insights in somatostatin receptor physiology. *European Journal of Endocrinology* 156(Supplement 1): S3–S11.
- Weckbecker G, Lewis I, Albert R, et al. (2003) Opportunities in somatostatin research: Biological, chemical and therapeutic aspects. *Nature Reviews Drug Discovery* 2: 999–1017.
- Wolkenberg SE and Thut CJ (2008) Recent progress in the discovery of selective, non-peptide ligands of somatostatin receptors. *Current Opinion in Drug Discovery and Development* 11: 446–457.
- Zeyda T and Hochgeschwender U (2008) Null mutant mouse models of somatostatin and cortistatin, and their receptors. *Molecular and Cellular Endocrinology* 286: 18–25.

Relevant Websites

<http://www.ncbi.nlm.nih.gov> – NCBI, National Center for Biotechnological Information.

Spastic Paraplegia

E I Rugarli, University of Cologne, Cologne, Germany

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Elena I. Rugarli and Andrea Ballabio, volume 4, pp. 61–66, © 2004, Elsevier Inc.

Glossary

Haploinsufficiency A locus shows haploinsufficiency when the amount of the gene product produced by a single allele is not sufficient to achieve a normal phenotype.

Paraplegia Paralysis of both lower limbs.

Spasticity Muscular weakness associated with increased stiffness and overactive reflexes.

Hereditary spastic paraplegia (HSP) is a group of neurodegenerative diseases characterized by weakness and spasticity of the lower limbs, loss of the vibratory sense, and urinary urgency. In 1983, Harding classified HSP in pure forms, when the symptoms are restricted to lower limbs weakness and spasticity, and in complicated forms, when the clinical picture is associated to a variety of neurological symptoms. These may include cerebral and cerebellar atrophy, optic atrophy, peripheral neuropathy, dementia, retinitis pigmentosa, and amyotrophy. Age of onset is extremely variable, with a few forms arising in childhood, and most cases between 20 and 40 years of age. Usually, the later the onset of the disease, the faster is the progression of its course. Although HSP is an invalidating disease, it never causes shortening of the life span of patients.

Neuropathology

HSP is characterized by the retrograde degeneration of the longest axons of the central nervous system (CNS): those of the crossed corticospinal tracts and of the fasciculus gracilis. In about half of the autopsic cases examined, the spinocerebellar tracts were also involved.

Axons of the corticospinal tracts arise from pyramidal neurons in layer V of the motor cortex and project through the internal capsule to reach the ventral surface of the medulla where they form two elongated swellings, the pyramids. These axons cross the midline at the junction between the bulb and the spinal cord and descend in the contralateral funiculus of the spinal cord. The crossed corticospinal tract conveys voluntary motor impulses to the secondary motor neurons located in the ventral horns of the spinal cord. The fasciculus gracilis is composed of the central branches of axons of pseudounipolar neurons located in dorsal root ganglia and ascend in the most medial part of the dorsal funiculus. These axons transmit deep sensory information from the lower extremities to secondary sensory neurons in the nucleus gracilis.

Characteristically, axonal degeneration in HSP involves preferentially the distal region of the axons. Degeneration of the corticospinal tracts is marked at lumbar levels and less pronounced at cervical levels. In most cases, the neuronal cell bodies are preserved. This pattern of axonal degeneration, also referred to as dying back, is slowly progressive and occurs

in several other conditions following a toxic, metabolic, or genetic insult.

Genetics

HSP is extremely heterogeneous from a genetic point of view. More than 40 different loci (listed SPG1–45) have been mapped at the time of writing. HSP can be transmitted as an X-linked, autosomal dominant, or autosomal recessive trait. In general, autosomal recessive HSPs are associated with a complicated phenotype, whereas autosomal-dominant forms tend to be pure. The molecular bases of the disease are also being unraveled at a quick pace. Nineteen HSP genes have been identified, thus providing clues on the cellular pathways underlying the disease. [Table 1](#) summarizes the current knowledge on the genetic bases of HSP.

Pathogenesis

The identification of several genes involved in HSP is beginning to shed light on the pathogenesis of this disorder. Although the known genes belong to different families and have distinct subcellular localizations, common pathogenic themes have emerged ([Figure 1](#)). The corticospinal neurons have extremely long axons that can reach the length of more than 1 m in humans and contain more than 99% of the cytoplasm of the cell. Development and maintenance of these neurons depend on efficient transport of organelles, molecular cargoes, and synaptic vesicles from the cell body to the distal tip of the axon, and back to the cell body. Transport processes are mediated by motors along cytoskeletal tracks and are coupled to membrane-trafficking events. The vast majority of HSP genes for which we have functional information are involved in different aspects of cell trafficking, such as microtubule (MT) dynamics, anterograde transport on microtubular tracks, endoplasmic reticulum (ER) morphogenesis, or endosomal trafficking. These processes require energy and are supported via coordinated movement of mitochondria along axons. Another group of HSP genes is indeed involved in mitochondrial quality control. Finally, axonal homeostasis strictly depends on myelin integrity, and an emerging category of HSPs is caused by mutations in

Table 1 Genetic bases of HSP

<i>Acronym</i>	<i>MIM</i>	<i>Inheritance</i>	<i>Chromosomal location</i>	<i>Phenotype</i>	<i>Gene product</i>	<i>Proposed function</i>
SPG1/L1CAM	312900	X-linked R	Xq28	Complicated	L1CAM	Development and guidance of the corticospinal tract
SPG2	312920	X-linked R	Xq21–22	Complicated	PLP	Axonal–glial interactions
SPG3A/ATL1	182600	AD	14q11–q21	Pure	atlastin	Endoplasmic reticulum morphogenesis
SPG4/SPAST	182601	AD	2p24–21	Pure	spastin	Microtubule severing
SPG5A/CYP7B1	270800	AR	8p12–q13	Pure	CYP7B1	Cholesterol and neurosteroid metabolism
SPG6/NIPA1	600363	AD	15q11.1	Pure	NIPA1	Membrane trafficking, BMP signaling
SPG7	600146	AR	16q24.3	Pure/complicated	paraplegin	Mitochondrial protease
SPG8/KIAA0196	603563	AD	8q23–24	Pure	Strumpellin/KIAA0196	Spectrin domain
SPG9	601162	AD	10q23–24	Complicated		
SPG10/KIF5A	604187	AD	12q13	Pure	KIF5A	Anterograde motor protein
SPG11	604360	AR	15q13	Pure	Spatacsin	Membrane trafficking?
SPG12	604805	AD	19q13	Pure		
SPG13/HSPD1	605280	AD	2q24	Pure	Hsp60	Mitochondrial chaperon
SPG14	605229	AR	3q27–28	Complicated		
SPG15/ZFYVE26	270700	AR	14q22–q24	Complicated	ZFYVE26(spastizin)	Membrane trafficking?
SPG16	300266	X-linked R	Xq11.2	Complicated		
SPG17/BSCL2	270685	AD	11q12–q14	Complicated	Seipin	Endoplasmic reticulum protein, unknown function
SPG18		AR	8p12–p11.21			
SPG19	607152	AD	9q33–q34	Pure		
SPG20	275900	AR	13q12.3	Complicated	Spartin	Membrane traffic, lipid droplet, mitochondria
SPG21	248900	AR	15q22.31	Complicated	Maspardin	Membrane traffic
SPG23	270750	AR	1q24–q32	Complicated		
SPG24	607584	AR	13q14	Pure		
SPG25	608220	AR	6q23.3–q24.1	Complicated		
SPG26	609195	AR	12p11.1–q14	Complicated		
SPG27	609041	AR	10q22.1–q24.1	Pure		
SPG28	609340	AR	14q21.3–q22.3	Pure		
SPG29	609727	AD	1p31.1–21.1	Complicated		
SPG30	610357	AR	2q37.3	Complicated		
SPG31/REEP1	610250	AD	2p11.2	Pure	REEP1	
SPG32	611252	AR	14q12–q21	Complicated		
SPG34	300750	X-linked R	Xq24–q25	Pure		
SPG35	612319	AR	16q21–q23.1	Complicated	Fatty acid 2-hydroxylase (FA2H)	Incorporation of α -hydroxylated GalCer into myelin
SPG39	612020	AR	19p13.3	Complicated	Neuropathy target esterase	Phospholipid homeostasis
SPG41		AD	11p14.1–p11.2			
SPG42/SLC33A	612539	AD	3q24	Pure	SLC33A1	Acyl-CoA transporter
SPG43		AR	19p13.11–q12	Complicated		
SPG44/GJC2	613206	AR	1q41–q42	Complicated	Connexin 47	Gap junction protein
SPG45	613162	AR	10q24.3–q25.1	Complicated		

proteins important for myelin–axonal interaction. Based on the current data, a tentative classification of known HSP genes in functional groups can be put forward.

Genes Involved in Cellular Trafficking Events

SPG4/spastin

SPG4 is the gene most commonly mutated in AD-HSP (~40% of cases), and its functional characterization clearly emphasizes the importance of cytoskeletal remodeling for axonal homeostasis. *SPG4* encodes spastin, an ATPase belonging to the AAA

(ATPases associated with diverse cellular activities) family. Spastin is involved in MT severing, a process by which long MTs can be fractured into shorter tracks. Spastin forms a ring-shaped hexamer containing a central pore and six radiating arms that dock onto the MTs. Structural data obtained using *Drosophila* spastin put forward a model by which spastin pulls the C-terminus of tubulin through its central pore generating a mechanical force that destabilizes tubulin–tubulin interactions within the MT lattice. MT severing plays an important role during meiosis and mitosis to release MTs from the centrosome and is thought to be essential in neurons, where the capacity of

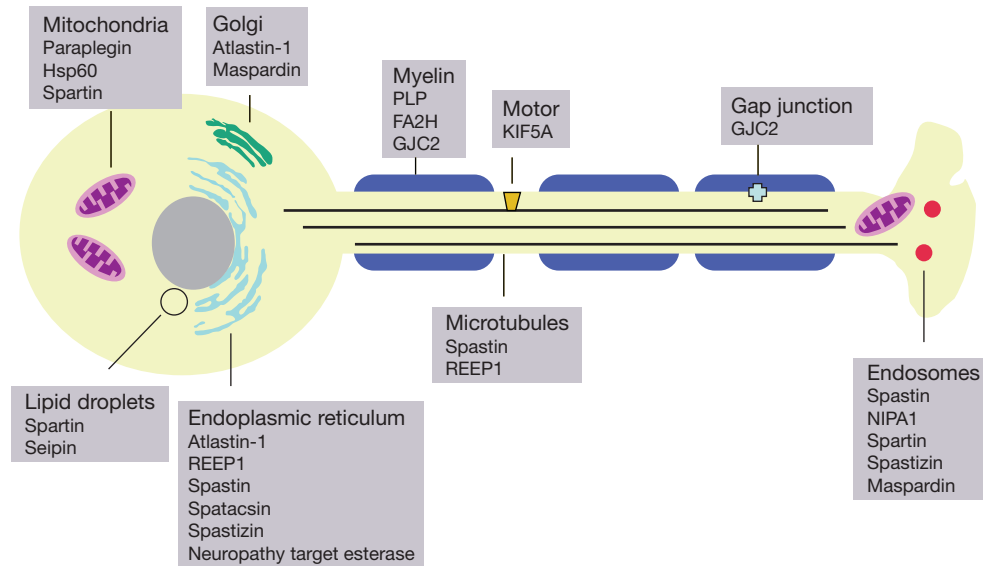


Figure 1 Schematic representation of a motor neuron, with the subcellular localization of the known HSP genes.

an MT to move is strictly related to its length. Spastin is indeed enriched in the centrosome, in the cytokinesis structure midbody, and in growth cones and branching points of neurons, all regions characterized by extensive cytoskeletal remodeling. It is, therefore, an attractive hypothesis that HSP, due to *SPG4* mutations, arises through disturbances of the dynamics of MTs in the long central motor axons.

The relevance of spastin function for axonal development and maintenance has been demonstrated both *in vitro* and *in vivo*. Spastin plays a role in the formation of axonal branches and in axonal development. Spastin knockdown in *Drosophila* by RNA interference (RNAi) caused defects in synaptic growth and neurotransmission due to an aberrantly stabilized MT cytoskeleton. These phenotypes could be ameliorated by administration of vinblastine, an MT-destabilizing drug. A depletion of presynaptic MTs was observed in null and hypomorph deletion spastin mutants. Dramatic defects in motor axon outgrowth and extensive apoptosis were found in the zebrafish, when morpholino antisense oligonucleotides were used to knock down spastin in the developing embryo. Finally, homozygous *Spg4*^{-/-} mice displayed a late-onset motor deficit associated with focal axonal swellings in the lumbar spinal cord with accumulation of organelles and cytoskeletal components, caused by severe reduction of anterograde and retrograde axonal transport.

Spastin MT-severing activity is targeted to specific cell compartments through the ability of spastin to interact with a variety of molecules via its N-terminal region. The N-terminal region of spastin can vary in length, through usage of two AUG codons in the first exon, and alternative splicing of exon 4, thus conferring potential for isoform-specific interactions. Indeed, spastin can be recruited in an isoform-specific pattern to endosomes or the early secretory pathway. The full-length spastin isoform is localized to the early secretory pathway, either on ER domains or on vesicles of the ER-to-Golgi intermediate compartment, while the short form of spastin (lacking the N-terminal 86 amino acids) can be strongly and preferentially recruited to endosomes. Notably, a number of spastin-interacting molecules link

spastin function to membrane-trafficking events. The full-length spastin isoform interacts with atlastin-1, REEP-1, and reticulon-1, three ER morphogens (see below). Spastin also interacts with the endosomal sorting complex required for transport (ESCRT)-III protein CHMP1B through the N-terminal MT interacting and trafficking (MIT) domain. The ESCRT machinery plays important roles in membrane cleavage events during multivesicular body formation and cytokinesis. CHMP1B is required for the recruitment of spastin to the midbody, thus suggesting coordinated MT severing and membrane abscission.

SPG3A/atlastin-1

The second most common gene involved in autosomal-dominant pure HSP (approximately 10% of the cases) has been linked to the *SPG3A* locus and found to encode one member of the atlastin family, atlastin-1. Atlastins belong to the dynamin family of large guanosine triphosphatases (GTPases), but are characterized by two closely spaced transmembrane segments near their C-termini. Atlastin-1 localizes to the ER, ER-to-Golgi intermediate compartment, and to the *cis*-Golgi apparatus and interacts through the transmembrane segments with members of the reticulon and DP1/Yop1p families. Both of these protein families are involved in shaping the tubular ER. Overexpression of a GTPase-inactive version of atlastin-1 or the cellular depletion of atlastin family proteins induces the formation of long unbranched ER tubules, while atlastin-1 overexpression leads to a drastic alteration of the ER morphology. In the *Drosophila* system, it was shown that *trans*-oligomeric complexes of atlastin promote homotypic fusion of ER membranes in a guanosine triphosphate (GTP)-dependent fashion. These studies suggest a model in which the reticulons and Yop1p proteins promote ER membrane tubulation, while a conformational change in atlastin, occurring on GTP hydrolysis, affects branching of the ER tubules by promoting homotypic membrane fusion. It is possible that spastin couples these membrane modeling events to MT regulation events that are also important for ER morphogenesis.

SPG31/REEP1

The hypothesis that ER morphogenesis plays an important role in the pathogenesis of HSP is further emphasized by the implication of the receptor-enhancing protein 1 (REEP1) protein in the third most common autosomal-dominant form of HSP. Although originally identified as a mitochondrial protein, REEP1 is a mammalian homolog of the DP1/Yop1p family. Similar to their yeast counterpart, mammalian REEP proteins are required for ER network formation *in vitro*. REEP1 interacts with the MTs through its C-terminal region and promotes ER alignment along the MT cytoskeleton. Remarkably, REEP1 interacts with both the long spastin isoform and atlastin-1 via hydrophobic hairpin domains in each of these proteins. Thus, at least three proteins involved in pure HSP – spastin, atlastin, and REEP1 – are likely to have a directly related function in ER morphogenesis and its interaction with the cytoskeleton.

SPG10/KIF5A

One of the best examples of the pathogenic role of impaired axonal transport in HSP comes from the identification of mutations in the gene encoding a specific neuronal kinesin heavy chain, KIF5A, in families with an autosomal-dominant pure form of HSP (SPG10). Kinesins are multisubunit complexes that work as motors for anterograde transport (from the cell body to the distal end of the axons) of membranous organelles and other macromolecular cargoes along MTs. KIF5A is abundantly expressed in neurons, where it is found in cell bodies, dendrites, and axons. Most of the KIF5A mutations that have been identified so far are missense mutations located in the kinesin motor domain. *In vitro* transport assays with purified motor domain mutants mixed with wild-type motors to mimic the situation in heterozygous HSP patients revealed that the presence of the mutant motors either slowed moving cargo populations or led to reduced MT binding affinity. In both cases, the gross cargo flux to the synapse might be reduced, either by a simple delay of transport or because of competition of MT-binding deficient motors with wild-type motor for cargo-binding sites.

SPG20/spartin

SPG20 is the gene involved in Troyer syndrome, an autosomal recessive form of HSP complicated by dysarthria, distal amyotrophy, mild developmental delay, and short stature, occurring with high frequency in the Amish population. The SPG20 protein product, spartin, shares homology in the N-terminal region with spastin within the MIT domain (see above). This has led to the hypothesis that spartin may be somehow involved in endosome trafficking. Indeed, endogenous spartin can be recruited to endosomes under some circumstances, such as epidermal growth factor (EGF) stimulation or expression of a dominant negative form of Vps4. Spartin is required for efficient EGF receptor degradation. However, the cell role of spartin appears more complex. A certain amount of endogenous spartin localizes to lipid droplets, and both overexpression and downregulation of spartin increase the formation of lipid droplets. Furthermore, some authors have localized a fraction of spartin on mitochondria. Spartin is monoubiquitinated and interacts with the ubiquitin ligase WW1. A full explanation of spartin role in HSP will require the elucidation of its function at different subcellular sites and the relevance of interaction with the ubiquitin ligase WW1.

SPG6/NIPA1

Mutations in the NIPA1 gene cause autosomal-dominant HSP. NIPA1 encodes for a polytopic integral membrane protein, localized to the cell membrane and to endosomes. One study has suggested that NIPA1 is a functional magnesium transporter. NIPA1 mutations in HSP patients are all missense mutations, which affect trafficking of the protein through the secretory pathway and its trapping in the ER. Studies in mammalian overexpression systems and in *Caenorhabditis elegans* have suggested that the mutations might be pathogenic via induction of ER stress and an unfolded protein response. This mechanism has been, however, questioned by another study. By contrast, a role of NIPA1 in bone morphogenetic protein (BMP) signaling has been proposed. Experiments with the *Drosophila melanogaster* homolog of NIPA1, spichthyin, have demonstrated that spichthyin regulates BMP receptor trafficking. *Drosophila* larvae lacking spichthyin display distal axonal abnormalities characterized by neuromuscular junction overgrowth that can be rescued by genetic manipulations blocking BMP signaling. In mammalian cells, NIPA1 was shown to physically interact with type II BMP receptors and to promote their endocytosis and lysosomal degradation, thus affecting BMP signaling. NIPA1 mutations would be less efficient at promoting degradation of BMP receptors. Remarkably, the same study suggested that spastin and spartin, although with a still unknown mechanism, may also inhibit BMP signaling, thus suggesting a common pathogenic mechanism.

Other HSP Genes with Potential Endosomal Localization

Other HSP genes have been proposed to localize to compartments of the secretory pathway or to play a role in cell trafficking. Mast syndrome is an autosomal recessive complicated form of HSP associated with dementia, present at high frequency in the Old Order Amish population. The causative gene, SPG21, encodes an acid-cluster protein of 33 kDa (ACP33), renamed maspardin, which localizes to the endosomal/*trans*-Golgi network. Maspardin has been found to interact with the aldehyde dehydrogenase ALDH16A1, a component of the large family of ALDH proteins, which catalyze the irreversible oxidation of a wide spectrum of aliphatic and aromatic aldehydes during the metabolism of endogenous and exogenous compounds. Further studies are required to clarify the function of maspardin and the role of this interaction for HSP pathogenesis, but it is worth mentioning that another ALDH protein, ALDH3A2, is involved in Sjögren–Larsson syndrome, a disease characterized by the association of mental retardation, ichthyosis, and spastic paraparesis.

Mutations in spastizin (SPG15) and spatacsin (SPG11) result in complex forms of HSP characterized by thin corpus callosum, maculopathy, and cerebellar signs. Preliminary studies suggest that they are present in different cell compartments, including the ER. Spastizin carries a FYVE domain, a signature for binding to phosphatidylinositol 3-phosphate, a phospholipid found on endosomal membranes.

Proteins Involved in Lipid Metabolism

As already mentioned, two studies have suggested a role of spartin in the formation of lipid droplets. Another HSP protein

involved in lipid droplet biogenesis is seipin (SPG17/BSCL2). Complete loss of function of seipin causes congenital-generalized lypodystrophy type 2, a disease characterized by loss of adipose tissue associated with hypertriglyceridemia, insulin resistance, and diabetes. Yeast seipin is found at the ER lipid droplet junctions and is important for droplet morphology. Two missense mutations that affect glycosylation of seipin have been instead associated to dominant neurological phenotypes, such as HSP, Charcot-Marie-Tooth disease or distal hereditary motor neuropathy type V. The current hypothesis is that these mutations act with a gain-of-function mechanism by inducing an ER stress response.

The CYP7B1 cytochrome P450 enzyme, mutated in SPG5A, hydroxylates oxysterols and neurosteroids, such as dehydroepiandrosterone and pregnenolone. Hydroxylation reduces the biological activity of these substrates and facilitates their conversion to end products easily excreted from the cell bodies. A marked accumulation of 27-hydroxycholesterol was indeed found in both plasma and cerebrospinal fluid in patients with SPG5A, although the pathogenic significance is still unclear.

Neuropathy target esterase (NTE) was originally discovered as the primary target of organophosphorous compounds, poisoning of which causes a delayed paralyzing syndrome with degeneration of motor axons. Subsequent cloning of the gene disclosed that NTE is a member of the patatin-like phospholipase family. NTE is a resident protein of the ER and catalyzes hydrolysis of membrane-associated lipids, as lysophospholipids. It is also implicated in deacylation of phosphatidylcholine. Sustained elevation of phosphatidylcholine was observed in a conditional NTE knockout in which the protein was selectively removed in neurons. This biochemical phenotype was accompanied by progressive axonal swelling and degeneration in sensory and motor spinal tracts, resembling HSP.

Mitochondrial Proteins

SPG7/paraplegin

SPG7 encodes paraplegin, a 795-amino-acid protein highly homologous to two yeast metalloproteases of the AAA family, Yta10p (Afg3p) and Yta12p (Rca1p). Like its yeast homologs, paraplegin has an N-terminal mitochondrial leader sequence and two transmembrane domains that anchor the protein to the inner mitochondrial membrane. Paraplegin is composed of two functional domains: an ATPase domain typical of proteins of the AAA family (ATPases associated with diverse cellular activities) and a proteolytic domain with a consensus metal-binding site.

The yeast Yta10p and Yta12p assemble to form a high-molecular-weight complex of about 1 MDa, the *m*-AAA protease, which is embedded in the inner mitochondrial membrane but exposes proteolytic sites to the mitochondrial matrix (hence the prefix *m* in the name). The *m*-AAA protease ensures protein quality control for inner membrane proteins, by guaranteeing the removal of nonassembled or misfolded proteins, and mediates the processing of specific mitochondrial substrates, such as the mitochondrial ribosome component Mrp132. Yeast cells that lack the *m*-AAA protease are unable to grow on nonfermentable substrates and exhibit deficiencies in the synthesis of mitochondrially encoded polypeptides.

Paraplegin co-assembles with the highly homologous AFG3L2 protein in human mitochondria to form a high-molecular-weight complex very similar to the one described in yeast. The human AFG3L2 protein is also capable of forming functional homo-oligomeric complexes. Remarkably, AFG3L2 is involved in a dominant form of cerebellar ataxia, SCA28. Mammalian *m*-AAA hetero-oligomeric or homo-oligomeric proteases complement the respiratory phenotype of Yta10p-Yta12p-deficient yeast, indicating functional conservation across species. In addition, the mammalian *m*-AAA protease has been implicated in proteolytic maturation of OPA1, which is involved in inner mitochondrial membrane fusion, cristae remodeling, and apoptosis, and is mutated in an autosomal-dominant form of optic atrophy. This opens up the possibility that deficiency of *m*-AAA protease subunits might cause alteration of mitochondrial dynamics, thus affecting the distribution and transport of these organelles.

Paraplegin-deficient mice recapitulate the human disease, being affected by a late-onset progressive distal axonopathy, characterized by axonal swelling and degeneration, due to the accumulation of membranous organelles and neurofilaments, suggesting that axonal transport is impaired. The first ultrastructural pathological sign that can be observed in paraplegin-deficient mice occurs several months before any evidence of axonal swelling and degeneration and consists in the appearance of abnormal mitochondria first in synaptic terminals and then in the distal regions of the long spinal axons that will later on undergo degeneration. Abnormal mitochondria include hypertrophic mitochondria, gigantic mitochondria, and mitochondria with disrupted or abnormal organization of cristae. This mitochondrial phenotype correlates with the onset of a measurable neurological impairment of the paraplegin-deficient mice on the rotarod apparatus and is therefore the primary trigger of the pathological process. Remarkably, mitochondrial abnormalities and axonal degeneration in the peripheral nerve could be partially rescued by intramuscular delivery of paraplegin cDNA via adeno-associated viral vectors.

Biochemical studies on fibroblast cell lines obtained from HSP patients with SPG7 mutations, and on mitochondria from mice with reduced dosage of *m*-AAA protease subunits, found a reduction of complex I activity, probably due to decreased stability of the assembled complex, and an increased sensitivity to oxidative stress, suggesting a potential mechanism for mitochondrial damage. This could be especially important in mitochondria located in nerve terminals where high energetic demands and increased concentration of free radicals are present.

SPG13/HSP60

The Hsp60 protein belongs to a conserved subgroup of chaperone proteins, termed chaperonins, which promote protein folding in the mitochondrial matrix, often in cooperation with the co-chaperonin Hsp10. The first Hsp60 mutation identified in a HSP family results in the V72I substitution. An elegant complementation assay performed in *Escherichia coli* showed that only wild-type Hsp60, but not Hsp60-V72I, together with the co-chaperonin Hsp10, can support the growth of bacteria in which the homologous chromosomal *groESgroEL* chaperonin genes have been deleted, indicating that the V72I mutant is

functionally inactive. It is unclear whether this mutant leads to haploinsufficiency through the formation of functionally inactive mixed chaperonin rings made up by both normal and mutated subunits, or may exert a genuine dominant-negative effect. The frequency of mutations in HSPD1 in HSP is very low. So far, only one other family has been reported in which a different missense mutation (Q461E) caused the disease with low penetrance. A recent study has suggested a role of Hsp60 as genetic modifier for the most common form of HSP due to spastin mutations.

Genes Involved in Myelin–Axonal Interactions

Disruption of intimate glia to axon interactions underlies the axonal degeneration observed in SPG2 patients. SPG2 maps to Xq22 and results from mutations in the proteolipid protein (PLP), one of the major protein components of myelin in the CNS. PLP is a four-helix spanning membrane protein that stabilizes the structure of the CNS myelin, by forming the intraperiod line. SPG2 comprises both pure and complicated forms of HSP and is allelic to the severe hypomyelinating Pelizaeus–Merzbacher disease (PMD). A genotype–phenotype correlation between the nature of PLP mutation and the severity of the clinical picture has been described. In general, the milder phenotypes, such as pure spastic paraplegia, are observed when the gene mutations lead to the formation of protein products that are still able to transverse the secretory pathway and reach the cell surface. Interestingly, these mutants exhibit impaired cholesterol binding and decreased association with lipid raft microdomains. Surprisingly, knockout mice with complete absence of PLP assemble compact myelin sheaths, but subsequently develop widespread axonal swelling and degeneration, most likely secondary to impaired axonal transport. It is still a mystery how the correct expression of PLP in oligodendrocytes provides support for myelinated axons. The future identification of putative signaling molecules would probably lead to a better understanding of this pathological cascade.

The strict relationship between myelination and axonal maintenance is further supported by the finding of mutations in the fatty acid 2-hydroxylase (FA2H) gene in patients with the complicated HSP form SPG35. FA2H is found in oligodendrocytes and is responsible for the first step of incorporation of α -hydroxylated galactosyl ceramides (GalCer) into myelin using free fatty acid as substrates. GalCer represent about 20% of mature myelin lipid. Mice deficient for FA2H display a late-onset axonal degeneration phenotype associated to myelin degeneration.

Recessive mutations in GJC2, the gene encoding the gap junction protein connexin47, cause Pelizaeus–Merzbacher-like disease. However, a recent study has identified one family carrying a novel recessively inherited mutation associated to a milder phenotype of complicated HSP. Gap junctions formed

of connexin47 and connexin43 play an important role for myelin maintenance.

Finally, another HSP gene that could have a role in myelin function is SLC33A1 (SPG42). This gene encodes an acyl-CoA transporter, a multiple transmembrane protein in the ER that transports acyl-CoA from the cytoplasm to the Golgi/ER where it serves as substrate of acetyltransferase that modifies the sialyl residues of gangliosides and glycoproteins. Gangliosides are highly expressed in the brain and play an important role in axon-myelin stability.

Developmental Genes

HSP may result from mutations in genes involved in the development of the corticospinal tracts. This is the case of SPG1, which has been linked to mutations in the cell adhesion molecule L1CAM. Consistently, neuropathological studies found absent or severely reduced pyramids in patients. In this form of HSP, spastic paraplegia begins in the first two decades of life, with delayed acquisition of motor milestones and slow progression of symptoms. Although this form may manifest as pure HSP, it is more often observed in association with a complex disorder, referred to either as mental retardation, adducted thumbs, spasticity and aphasia (MASA) syndrome or as corpus callosum hypoplasia, mental retardation, adducted thumbs, spastic paraplegia, and hydrocephalus (CRASH) syndrome. L1CAM is a transmembrane glycoprotein, which is expressed during development on the surface of long axons and growth cones, including those of the corticospinal tracts. L1 mediates cell adhesion, neurite outgrowth, axon pathfinding, and fasciculation through homophilic and heterophilic binding with a variety of extracellular and transmembrane molecules, and plays an important role in mediating guidance of corticospinal axons through the pyramidal decussation.

See also: **Bioenergetics:** Nuclear Genes Involved in Mitochondrial Function and Biogenesis.

Further Reading

- Dion PA, Daoud H, and Rouleau GA (2009) Genetics of motor neuron disorders; new insights into pathogenic mechanisms. *Nature Reviews Genetics* 10: 769–782.
- Farhan H and Hauri HP (2009) Membrane biogenesis: Networking at the ER with atlastin. *Current Biology* 13: R906–R908.
- Harding AE (1983) Classification of the hereditary ataxias and paraplegias. *Lancet* 21: 1151–1154.
- Martinelli P and Rugarli EI (2010) Emerging roles of mitochondrial proteases in neurodegeneration. *Biochimica et Biophysica Acta* 1797: 1–10.
- Salinas S, Proukakis C, Crosby A, and Warner TT (2008) Hereditary spastic paraplegia: Clinical features and pathogenic mechanisms. *Lancet Neurology* 7: 1127–1138.
- Stevanin G, Ruberg M, and Brice A (2008) Recent advances in the genetics of spastic paraplegia. *Current Neurology and Neuroscience Reports* 8: 198–210.

Spectrophotometric Assays

B Chance, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 4, pp. 67–75, © 2004, Elsevier Inc.

Glossary

Brown converter An electrical chopper developed by the Leeds and Northrup Company that had very little contact potential variation.

Cytochrome oxidase The terminal enzyme of most oxygen-using systems.

Dual-wavelength spectrophotometer A spectrophotometer with time shared to adjacent wavelengths in order to minimize the effect of scattering changes upon absorbance changes, because scattering varies very slowly with wavelength, while cytochrome absorption varies sensitivity with wavelength.

DuBridge electrometer (DU) A remarkable use of the suppressor grid of the pentode to control electron flow with very high input impedance, affording the basis of the Beckman pH meter and spectrophotometer.

Electron tunneling The transfer of electrons between two proteins at a distance without collision of their active sites.

Flavoprotein A widely spread pigment, but in this article the prosthetic group of ketoglutarate and pyruvate dehydrogenase.

Mercury arc A very useful light source for biological studies, which gives light at exactly the correct wavelength for hemoglobin and in some cases cytochrome studies.

Moltano Institute A famous institute directed by David Keilin through the 1930–50s. It became a Mecca for those working with cell respiration.

Near-infrared (NIR) imaging The use of the wavelengths in the red region just at the verge of invisibility between 700 and 900 nm to better penetrate tissue.

Photochemical action spectrum The effect of light upon a biological system often used to activate a carbon monoxide-inhibited cytochrome.

Photomultiplier A highly sensitive light detector used in many spectrophotometers.

Photon migration The phenomenon of photon diffusion through tissues used in great detail recently to image subsurface objects.

Ruby laser One of the early forms of the laser, emitting in the red region.

Scattered light Light that does not proceed directly through tissue.

Spectrophotometer A device that measures the absorbance of materials as a function of wavelength or, in some cases, energy.

Time-resolved spectroscopy (TRS) Spectroscopy using sharp pulses of light to distinguish the scattering from the absorption of tissues.

X-ray absorption spectroscopy (EXAFS) Spectroscopy used to obtain high-resolution structures of Fe and Cu enzymes.

Manual Spectrometers

The rightfully famous physicist R.W. Wood made a high-quality plastic replica grating and incorporated a pair of them in the Coleman spectrophotometer to make a double-grating instrument that was valid from 400 to 700 nm, had a very good f number, and had a resolution independent of wavelength (~ 20 nm). This was an essential element of the kinetic and spectroscopic studies of the enzyme–substrate compounds of catalases and peroxidase, as carried out in the laboratory of Hugo Theorell. Although it was an excellent commercial instrument for laboratory work in a visible region, Cary's quartz monochromator, however, together with Beckman's unreferenced exploitation of DuBridge's ingenious electrometer (of the pH meter as well), made the Beckman DU instrument superior. However, if Charles Chaplin were to make a film of the biochemist similar to 'City Lights', this instrument would serve the purpose. Dark current adjustment, slit-adjustment, cuvette-holder shifting, and balancing to a null by hand were the earmarks of the instrument that most biochemists used from the 1940s to the early 1960s. Note that there were no alternatives. C.C. Yang invented what was called

the 'Yang machine', a very simple and robust instrument used by many in the Johnson Foundation laboratory, which used a vibrating mirror system to send light through a reference cuvette and a measure cuvette, together with dynode feedback, and gave a logarithmic output. Concurrently, Lenart Åckerman in the laboratory of Bo Thorell in Stockholm invented the same machine that also provided spectral scans but that was held off the market for about a decade, so the DU could have an orderly and prosperous senescence.

Dual-Wavelength Technology

The development of dual-wavelength systems began with Glen Millikan, who ingeniously scribed a barrier in the Weston photovoltaic cell and used green and red filters to measure myoglobin spectral changes in the visible region. However, he connected the output to a mirror galvanometer so that the system was intrinsically quite slow and, thus, unsuitable for the rapid measurements required for the flow apparatus. The dual-wavelength principle was continued in the Millikan ear oximeter used so effectively during World War II.

Rapid spectrophotometry required phototubes. The significant experience Britton Chance had acquired with them by inventing an automatic steering device while still in his teens led him to use cesium oxide on silver and then antimony as photocathodes. These were combined with the technology of high-gain direct current amplification from the electrophysiologists Adrian, Gasser, Erlanger, and the Schmitt brothers, who all had to develop an entire field of electronic amplification for studying nerve action potentials. The three-wavelength system used to measure the kinetics of enzyme–substrate compound formation in the Soret region, together with the overall formation of leucomalachite green to malachite green, provided measurements of enzyme–substrate compounds that required a mechanical differential analyzer for the transient solution of the Michaelis–Menten/Briggs–Haldane equations for enzyme kinetics (Figure 1). Indeed, the rapid flow apparatus led the field of fast spectrophotometric determinations, not only of the enzyme–substrate compounds of peroxidase but also, in the hands of Quentin Gibson, for measuring hemoglobin kinetics.

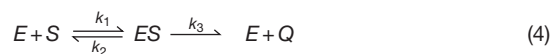
The introduction of the chopper-stabilized amplifiers in systems exploited during World War II by the Group 63 of

Chance and co-workers (Precision Circuits Group at MIT RadLabs) led to a whole new generation of low-level amplifiers and was the basis for many commercial instruments. Furthermore, the basis for the integrated circuit was afforded by the subminiature triodes developed for the proximity fuse, for which the rejection rate was so high that thousands of these diodes suitable for peacetime applications became available and were made in the first step of integrated circuits, precursors of the integrated chips that were soon to come in the postwar era.

Photochemical Action Spectra

One historic experiment, which now seems almost to have been forgotten, was performed by the Schmitt brothers in order to validate the Warburg hypothesis, namely, to obtain the photochemical action spectrum for the propagation of action potentials in nerves. The results did, in fact, coincide with Warburg's hypothesis and provided the much-needed link between prokaryote and eukaryote. The Schmidt brothers received very little credit for their work, without which Warburg's and Keilin's work, based on yeast cells only, would

These reactions are represented by Briggs and Haldane as the bimolecular combination of the enzyme, E , and substrate, S , to form an intermediate compound, ES , followed by a monomolecular decomposition into free enzyme and activated or altered substrate, Q , representative of the products of the 'over-all' enzyme action.



If e is the total molar enzyme concentration, x the molar substance concentration, p the molar concentration of ES , k_1 the second-order rate constant, and k_2 and k_3 the first-order rate constants, then

$$\frac{dp}{dt} = k_1 x(e - p) - (k_2 + k_3)p \quad (5)$$

$$\frac{dx}{dt} = -k_1 x(e - p) + k_2 p \quad (6)$$

These two equations represent the rate of formation of the intermediate compound and the rate of disappearance of the substrate.

The solution of these equations has been already obtained by Briggs and Haldane for the special conditions of the steady state, when

$$p = p_{\max}, \quad \frac{dp}{dt} = 0, \quad \text{and} \quad \frac{p_{\max}}{e} = \frac{x}{x - Km} \quad (7)$$

where

$$Km = \frac{k_2 + k_3}{k_1} = x \frac{(e - p_{\max})}{p_{\max}} \quad (8)$$

A further solution valid during the steady state is obtained by adding eqns 5 and 6,

$$\frac{dx}{dt} = -k_3 p_{\max} \quad (9)$$

where dx/dt is the rate of disappearance of substrate.

Figure 1 Michaelis–Menten theory.

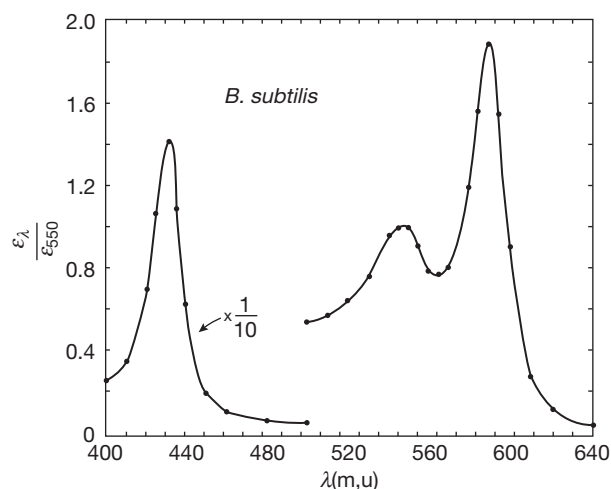


Figure 2 Photochemical action spectrum of cytochrome oxidase as obtained by the LeRoy Castor using the oxygen electrode to measure light-activated respiration. It is seen that this spectrum is much better defined than Warburg's classic spectrum of yeast where mainly the lines of the mercury arc were employed. Here, a continuous monochromated light would be used and the σ bands were clearly delineated as are the Soret bands.

have been less acceptable (Figure 2). At the time, there was no reason to believe that the eukaryotic yeast cell was a model for tissue studies. LaRoy Castor quickly improved the action spectral method and also discovered a heme prosthetic group oxidase (cytochrome *o*).

Tissue Spectroscopy

Glen Millikan was a pioneer in shining light through tissue and recording myoglobin deoxygenation in functional activity. During the same time, there were many microspectroscopic studies by Arvanitaki and Chalazonitis, and also, notably, by the laboratory of Caspersen with his colleague Bo Thorell. Lundegårdh was a pioneer in applying the spectroscopic technique to bundles of plant roots, but was unable to correct for the large changes of scattering due to the osmotic activity of the roots. However, Lou Duysens developed a sensitive single-beam method that he applied with diligence and effectiveness to cytochrome components of green leaves.

A time multiplex dual-wavelength system made possible by simply mounting on top of the Brown converter, or vibrating a small mirror at 60 Hz, allowed time sharing of two wavelengths through suspensions of highly scattered material derived from heart muscle or from liver as intact mitochondria. This dual-wavelength machine was the precursor of the time multiplex systems that have been used for exploiting near-infrared (NIR) spectroscopy.

The Scattered Light Problem

It was recognized early on that the scattered light would confound spectrophotometric measurements of pigments of

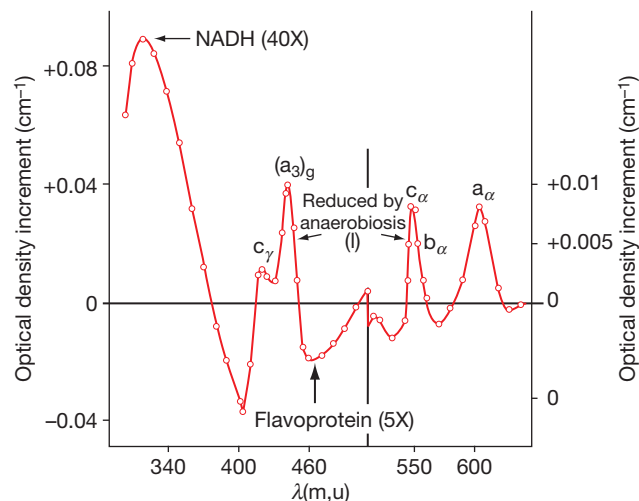


Figure 3 The difference spectrum of intact phosphorylating mitochondria. This was the first difference spectrum for oxidized minus reduced guinea pig liver mitochondria obtained with the dual-wavelength method employing reference wavelengths of 540 nm for the alpha bands and 450 nm for the Soret bands. The incremental absorption with respect to that reference is indicated giving the ratio of the Soret to the alpha bands to be four and the NADH peak to be 40 times that of the alpha peaks. The flavoprotein trough, that is, oxidized flavoprotein is the absorber, is shown at 460 nm. Although the Soret peak of cytochrome oxidase was visible at 444 nm by Keilin's microspectroscope, the flavin trough and the NADH peak (the latter being responsible for the intense autofluorescence of cells) were not available to Keilin in Hartree's heart muscle preparation in which the absorbance of these two components was greatly diminished by the elimination of citric acid cycle enzymes.

tissue. Yet, in 1949, Chance recognized that the scattered light effect would be greatly diminished if the solid angle of the detector system was large instead of small, as in commercial spectrophotometers (for long cuvettes). Thus, instead of a 10-cm-long sample compartment, as used in most of the Beckman/Cary devices, Chance and co-workers placed a large area detector and an 'end-on' photomultiplier almost in contact with the sample, particularly using the time-shared dual-wavelength system, from which very flat baselines were obtained for spectroscopic differences as long as the wavelength difference was not too great. The difference spectrum obtained from guinea pig liver mitochondria is shown in Figure 3 as an example of the data that were obtained with the dual-wavelength spectrophotometer, where the baseline was set at the suspected isosbestic points of 540 and 630 nm. The action spectrum for reduction of the oxidized form of the electron carriers was measured. This spectrum, although validating Keilin's microspectroscopy, clearly showed some anomalies. Cytochrome c_1 was a separate component. A trough due to the disappearance of oxidized flavin and peaks due to appearance of reduced NADH were not identified by Keilin's keen eye, which proved to be the most useful components of the respiratory chain because they were highly fluorescent. But most importantly, they were inhabitants of the mitochondrial matrix space where most of the citric acid cycle reactions were carried out. The instrument was surely very worthwhile.

It had identified not only three components of the respiratory chain that had been previously underestimated or remained completely unobserved, but also, for example, the difference spectrum obtained by Lucille Smith and Aristid Lindenmeyer showed the P450 compound of carbon monoxide, an observation confirmed by G.R. Williams and used by P450 enthusiasts today.

Control of Respiration

No other phenomenon was as important to physiologists as the control of respiration. The topic had received very little attention until Lardy and Wellman's historic paper, which showed that isolated prepared mitochondria very nearly completely ceased respiration in the absence of adenosine diphosphate (ADP) and phosphate but restarted respiration upon addition of ADP and phosphate. This led Chance and Williams to explore the nature of activation of the electron carrier in the respiratory chain. They found that all spectroscopically identifiable components changed their steady-state values when ADP was added to the suspension of resting mitochondria. This phenomenon now provides the basis for all of the metabolic activations studied by magnetic resonance imaging (MRI) as the bulb effect, and for using the saturation of oxyhemoglobin (HbO_2) *in vivo* as a measure of metabolic activity.

At the same time, Lehninger, Chance, and others studied the uptake of calcium into the mitochondrial matrix, and Chance and Williams found that calcium activated the suspensions of mitochondria to the same extent and with the same speed as did ADP suggesting, as recently discussed by Carafoli, the presence of a calcium carrier in the mitochondrial membrane.

Low-Temperature Technology

Perhaps the most secure instrumental data were those obtained at low temperature in liquid nitrogen, where the spectra were more intense. The delineation of the absorption bands, particularly in the bc_1 region, turned out to be essential in showing the two kinds of cytochromes *b*.

Lubbers' Rapid Scan

Tissue absorption and scattering occupied the minds of a number of workers, particularly those of Lubber and Thews, who not only did much theoretical work on scattering but also perfected a galvanometer-driven wavelength scanning device, whose speed far exceeded that of anything else available at the time and which has only been perfected recently by the work of Desr and co-workers.

Photoactivation Studies

Perhaps the greatest advantage of the technique for observing tissue properties was the fact that it was possible to photoactivate the system under spectroscopic examination and observe the photoinduced absorption spectrum, a specialty of the Laboratory of Lou Duysems. This, of course, required a

measure of common mode rejection, at which the dual-wavelength system excelled because the light flashes shared a single detector with an alternating current coupling, allowing the system to reject leakage of not only the photolysis light but also other disturbances that affected the common mode. Furthermore, illumination at subzero temperatures of the solid phase allowed the study of many photoactivated systems containing derivatives of chlorophyll or those rendered photoactivatable by combining with a ligand such as carbon monoxide. Thus, optical measurements in the region of the Soret band and illumination in the region of the alpha band allowed a measure of protection of the dual-wavelength detector system from very strong photolysis light, for example, that provided by intense sources such as the carbon arc. These studies bridged the gap between Otto Warburg's unbridgeable action spectrum experiment on yeast and Keilin and Hartree's microspectroscopical observations of cytochrome oxidase in yeast and in a pigeon heart muscle preparation.

Not only the cytochrome oxidase-CO complex CO but also the splitting and recombination of myoglobin CO compounds were studied in detail, with optical spectroscopy as well as structural methods such as X-ray absorption spectroscopy, with results that make headlines even today for the subject of ligand-docking concomitant with structural changes, confirming the X-ray structural studies in which the E7 histidine was altered by a ligation of myoglobin with CO.

The Ruby Laser and Electron Tunneling

Lacking powerful light sources of short duration, Chance and co-workers relied mainly upon xenon flash lamps, particularly those of Edgerton and co-workers. However, construction of the first ruby laser used in biological studies by Bunkenberg and DeVault opened up a whole new time domain, not only because of the monochromaticity of the laser (which greatly simplified the filter leakages) but also because of the duration of the light pulse (in the microsecond region), which for the time was extraordinarily short. From this study came the totally unexpected observation that the photochemical oxidation of cytochrome *c* by photoactivated chlorophyll of *Chromatium* was temperature independent, not only from room temperature to liquid nitrogen temperature but also eventually at liquid helium temperatures, extending the time range of the spectroscopic method to one of the primary or near-primary results and/or reactions of photosynthesis (Figure 4), and also introducing the concept of electron tunneling in biochemical reactions, a concept pursued by Les Dutton and Harry Gray.

The history of spectroscopy is checkered with unexpected results and meaningful interpretations on the nature of electron transfer reactions. In fact, the study of reactions obtained from photolysis activation of CO compounds in the presence of O_2 caused ligand exchange to occur, especially in the case of CO. Together with Carlo Saronio, Chance discovered a number of intermediate steps involving higher oxidation states of iron mirroring those obtained with peroxidase and H_2O_2 , but much faster and more complex. This was because copper oxidation was involved in the active site of CO, enabling donation of four electrons sequentially for the reduction of oxygen to water without significant amounts of free radical intermediates.

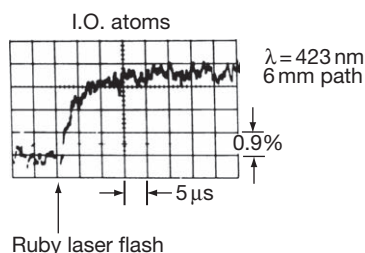


Figure 4 Typical recording of the oxidation of cytochrome *c* of photosynthetic bacteria at low temperatures as activated by a light flash of the ruby laser. Remarkably, this reaction rate was affected very little by the transition from room temperature to liquid nitrogen temperature due to electron tunneling.

This contrasted with the photoactivation of porphyrins in the absence of electron donors, resulting in the creation of singlet oxygen, a process used extensively in photodynamic therapies.

Fluorochromes of Tissues

Spectroscopy of biological fluorochromes, such as NADH and flavoprotein compounds that exhibit strong fluorescence in mitochondria, led Chance and co-workers to study them extensively *in vivo* by a simple method in which excitation was obtained by the very strong 366 and 436 lines of the mercury arc. This yielded emission spectra characteristic of the chromophores in a variety of functional states, characterizing electron transfer through the citric acid cycle as reductant and the cytochrome chain as oxidants, giving essential features of the redox state of the mitochondrial matrix. Interpretations of the importance of metabolic control and thermodynamic principles of the system were derived from these, particularly by R.L. Veech.

Hemoglobin and Cytochrome

It is perhaps ironic that the cell biologists and etymologists of the Molteno Institute focused on the cytochromes of hemoglobin-free cells and tissues and that the real physiology of oxygen delivery to tissues by hemoglobin and its utilization by CO had to wait for more sophisticated methods. Even now, the observation of cytochrome oxidase absorption bands in the presence of physiological concentrations of oxy- and deoxyhemoglobin is fraught with controversy, to the point that no reliable spectroscopic distinction of the copper component of cytochrome oxidase and the overlapping spectra of HbO₂ has been obtained. In fact, nearly all studies find that the so-called cytochrome oxidase Cu changes track those of HbO₂ due to the similarity of their spectroscopic absorption bands in the NIR region, and the great predominance of hemoglobin absorption over that of cytochrome oxidase.

NADH as an Oximeter

It has been shown that the fluorescence of NADH and flavoprotein, when used in a ratiometric manner, can exhibit a

reasonable immunity to changes of hemoglobin concentration, and has afforded standards for the independent changes of hemoglobin and NADH in transient hypoxia and reoxidation. Such independence of the measures of cytochrome and hemoglobin has never been clearly demonstrated for the copper component of cytochrome oxidase, due to the overlapping spectra of HbO₂.

An alternate approach was based upon the fluorescence of the newly discovered NADH and flavoprotein components of the mitochondrial respiratory chain, particularly by using the ratio of these two fluorochromes. Fortunately, NADH fluoresced in the reduced state, while flavoprotein fluoresced in the oxidized state, so their ratio was a measure of the redox state of mitochondria, a signal that was found to be only marginally affected by changes in the oxygenation of hemoglobin in model systems. Using this criterion on the very strong signal of NADH only, it was possible to show that the fluorescence of NADH was unchanged until the oxy/deoxy transition of hemoglobin was almost complete in a system in which functional activity was measured by the photoaction potential of an animal model, removing any doubt about the higher oxygen affinity of cytochrome *in vivo* as compared to hemoglobin (~20 torr). This was a very important milestone for physiologists, who know that the deoxygenation of hemoglobin is very high and the critical *p*O₂ of mitochondrial function is compromised. Thus, the calibration of tissue oximeters in the region of intravenous saturation of hemoglobin (i.e., 20–30%) must be precisely measured to indicate critical tissue hypoxia *in vivo*. This indeed was subsequently validated by measurements of tissue energetics through phosphorus nuclear magnetic resonance (³¹P NMR), particularly by measurements of the phosphocreatine:phosphate ratio.

NIR Spectroscopy

Jobsis used Kramer's technique to measure in the infrared, and developed a technology for the measurement of absorption of the copper component of cytochrome oxidase in the region of 830 nm based upon studies of the cat model and the heads of neonates, which he termed 'transcranial spectroscopy'. He further developed algorithms based upon fluorocarbon-perfused cat brain to give optical pathlengths that were believed to be transferable to the neonate brain and allow a deconvolution of blood volume and saturation changes from those of cytochrome oxidase signals using the full-length light algorithm. Delpy and co-workers avidly followed the lead of Jobsis and developed a close correlation between the decreased concentration of HbO₂ and the so-called copper signal in a number of models, suggesting that the mitochondria in tissues contained a low-affinity cytochrome oxidase and responded to *p*O₂ (in a slice) in a way similar to that of hemoglobin. However, isolation of rat brain mitochondria failed to support this contention. Furthermore, the freeze-trapped hypoxic brain failed to show the absorption band of reduced cytochrome *c* in mild hypoxic stress that caused deoxygenated hemoglobin. In fact, the absorption band of reduced cytochrome *c* was not observed until the band of HbO₂ was no longer detectable, according to the work of Bashford.

Although attempts were made to detect the copper absorption band of hemoglobin in the NIR region, animal studies showed that the fluorescence of NADH and the flavoprotein could be used to detect anoxia in the presence of hemoglobin, particularly when the ratio of the fluorescent-oxidized flavoprotein and the fluorescent-reduced NADH were employed; this value was relatively insensitive to the hemoglobin concentration. In fact, further demonstrations showed that the fluorescence of NADH was unaffected by the deoxygenation of hemoglobin in animal model brain. The NADH fluorescence increased in hypoxia only when the hemoglobin was already almost completely deoxygenated. This observation suggests that measurements of the critical pO_2 in hypoxia require accurate measurement of extreme values of hemoglobin desaturation, at the critical pO_2 for mitochondrial function.

NIR Spectroscopy of Brain and the BOLD Effect Measured by MRI

Much interest in the NIR method is based upon Ogawa's finding that changes in deoxyhemoglobin concentration (changes of the paramagnetic species of deoxyhemoglobin) enhanced water relaxation in the brain. This opened up the field of study of the activation phenomenon in the human brain, in which changes in deoxyhemoglobin levels are measured by NMR and by NIR tissue spectroscopy. The MRI changes are precisely imaged, while the NIR images, although crude, are measures of the rapidity of the changes. But in addition to incremental changes of deoxyhemoglobin, NIR could measure the saturation value of

hemoglobin, which, for reasons involving Beer's law, originates mainly from the arteriolar/capillary/venolar bed. This feature, namely, the value of local oxygen extraction due to incremental changes of mitochondrial functional activity (i.e., localized activation), is not measured by MRI. The two techniques are now widely accepted as indicative of localized brain activation and have afforded the basis for in-depth studies of visual and sensory motor function. But, most importantly, NIR gives an excellent rendition of prefrontal cortex signals without the difficulty of the large water content of the ocular system encountered with NMR (Figure 5). The use of activation images is appropriate to the NIR system, where baseline values may be somewhat variable and difficult to calibrate. The incremental changes of blood volume measured as changes of total hemoglobin, together with the aforementioned oxygen-extraction measure, that is, desaturation of hemoglobin, can be directly related to local metabolic activity, opening up a new field of NIR study of the semi-quantitative nature of the hemoglobin signals.

NIR Imaging

Although the above-mentioned studies used dual-wavelength technology stemming from that of Glenn Millikan, a completely new concept was introduced by the discovery that photon migration through tissues can be modeled by the diffusion equation. Multiple sources and detectors give very reasonable two- and three-dimensional imaging, and the propagation of light through tissue can be quantified by pulse time and phase and amplitude measurements, much as has been the case with measurement of fluorescence.

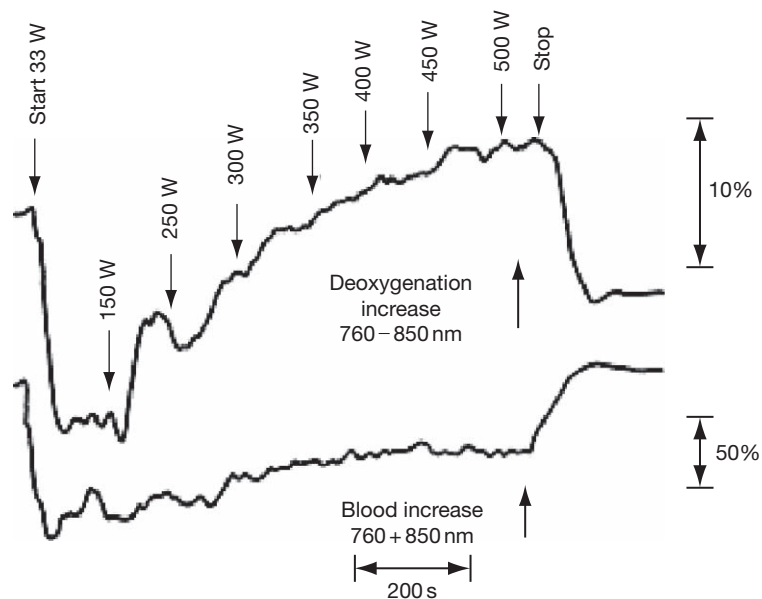


Figure 5 Evidence of activation of metabolism measured in the near-infrared (NIR) by the dual-wavelength method of 760–850 nm (equibestic wavelengths), which show progressive deoxygenation of the quadriceps muscle of a trained athlete during bicycle exercise up to the remarkable level 500 W, at which nearly complete deoxygenation of hemoglobin can occur as measured separately by the sum of signals at these two wavelengths. With appropriate coefficient, the increase of blood flow in the muscle also occurs giving a larger absorbance signal. At this level of exercise, it is probable that myoglobin is not deoxygenated as indicated by separate experiments with animal models.

Photon Migration in Tissues

The discovery that photons migrating through tissue followed the diffusion equation and that the tracks could be simulated by Monte Carlo methods, together with the adaptation of time-correlated single photon counting to the task of measuring propagation times in tissues, opened up an entire field of NIR spectroscopy and imaging. This grew to be a field of medical science in a manner similar to NMR, but differing in the fact that the necessary equipment is not nearly as expensive, so the proliferation of the technique in research laboratories could be much more rapid. Because the profit margins did not match those of NMR, commercial production of NIR imagers has been restricted to two or three companies.

Tissue Optical Properties

Three techniques are outstanding in the measurement of tissue optical properties. The first, and still the foremost, is the pulse time method, in which photon delay is caused by scattering and photon attenuation is caused by absorbers such as hemoglobin, water, and lipid. Because time-resolved spectroscopy (TRS) immediately deconvolutes scattering and absorption by time domain analysis, it is a preferred method. Similarly, variable frequency-modulated light will unravel by Fourier transformation exactly the same quantities as those obtained by TRS. However, the difficulty of stabilizing the phase shifts of electronic systems of variable frequency, together with the limitations of detection response to high-frequency radio waves, has limited this system to approximately 400 MHz. Nevertheless, many instruments have been made in the frequency range of 50–200 MHz that are used for quantifying absorption and scattering in multiwavelength systems and that are capable of measuring hemoglobin saturation with significant precision. Such systems have adopted some cell phone components and are therefore compact and cheap.

The most reliable and most used system modulates the light at very low frequencies in either, or time shares the light sources in a multiplex system, and appears to be the preferred system for many applications. The deconvolution of scattering and absorption can be obtained if sufficient data are taken at various source–detector separations and optical wavelengths to include the scattering variations.

Cancer Detection

In cancer detection, scanning the breast, for example, an activation signal based upon angiogenesis and hypermetabolism is given, causing more blood volume and more deoxygenated hemoglobin to be present. Although this criterion may not be applicable to all cancers, it has given remarkably good scores in one breast cancer study.

Muscle Studies

A series of studies has been based upon activation measurements, in which (e.g., in both muscle and brain) functional activation causes hemodynamic changes due to mobilization of blood flow and saturation changes due to varying metabolic activity (Figure 5). This has been used in muscle not only to evaluate exercise capability, but also to quantify disability due to occlusive disease in the limbs.

Brain Functional Activation

In brain studies, whereas NMR and the optical method both measure hemodynamic activation due to functional activity, the NIR method is unique in that it measures changes of hemoglobin saturation caused by varying degrees of oxygen

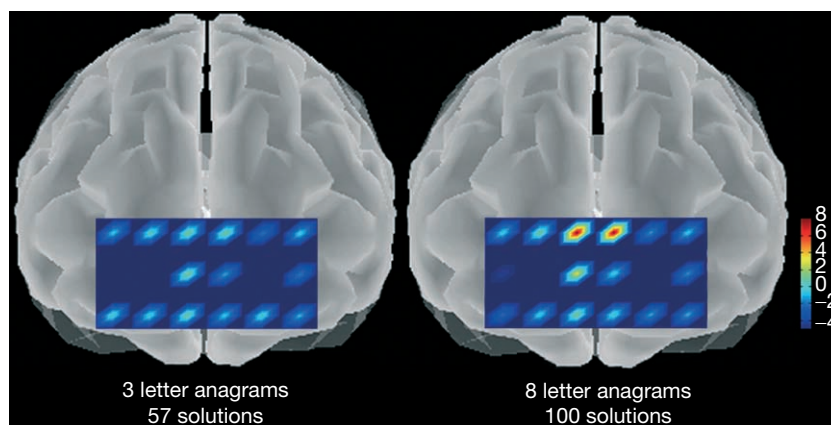


Figure 6 Illustrating the use of the NIR continuous wave dual-wavelength system in localizing the particular voxels in the human forebrain at which brain functional activity is indicated to be increased by the oxygenation changes with respect to the baseline level of 3 letter anagrams. Each colored symbol represents the average of over 100 tests of a particular individual giving an average value of 8 μM oxygenation under this condition of maximal stress. The region observed is the Broadman's 9 and 10 or ear-to-ear and hairline-to-eyebrow region of the projection of the prefrontal cortex. This separation was 4 cm in order to ensure signaling from the prefrontal cortex to accentuate the absorbance changes due to the prefrontal cortex and minimize those that might be associated with tissue layers at smaller depths.

extraction from the capillary blood vessels of the brain. Thus, the convenience and economy of the optical method lends itself to studies of minimally perturbed human subjects, be they adults or neonates (Figure 6).

Brain Studies: The Neuronal Signal

The early studies of David Hill, Tasaki, and Richard Keynes on transparency changes of axons upon stimulation suggest that a functional optical signal could be obtained in animal and human brain. Although the studies of Salzberg did indeed verify the scattering changes in isolated preparations, studies using NIR to seek similar changes in the human brain suggest that only very small and somewhat irreproducible signals can be obtained. For example, the early experiments of Gabriele Gratton have not been duplicated either by himself or by others working in the field (Villringer and Franceschini, and Wolf), in part, due to instrumental difficulties, and in part due to the very small size of the signal. The rise time of the signals appears to be in the range of 100 ms and the amplitude as small as one part in 10^4 or even smaller. However, the relatively robust signals obtained from the hemodynamic and metabolic activations discussed previously show the feasibility of measurements of functional activity in the human brain, particularly in the prefrontal region, and open up a reliable and economical method for human brain studies that predict a vibrant future of the optical methods in human studies.

Summary

The story of the development of optics in biochemistry and biophysics does not end here. In fact, some might say that this is just the beginning of the transferability of optical tomography and optical biopsy to small animals on the one hand and human beings on the other. Perhaps spurred by the interest in online methods for evaluating the growth and recession of cancers under the influence of appropriate drugs in small animals, transferability of these principles to human subjects is becoming important. The optical method is taking its place along with MRI, positron emission tomography, ultrasound, computed tomography, and X-ray mammography as methods for studying pathologies in the human body.

See also: **Bioenergetics:** Cytochrome *bc*₁ Complex (Respiratory Chain Complex III); Cytochrome *c*.

Further Reading

- Carafoli E (2003) Historical review: Mitochondria and calcium: Ups and downs of an unusual relationship. *Trends in Biochemical Sciences* 28: 175–181.
- Chance B (ed.) (1989) *Photon Migration in Tissues*. New York, NY: Plenum.
- Chance B (1991) The optical method. *Annual Review of Biophysics and Biophysical Chemistry* 20: 1–18.
- Keilin D (1966) *The History of Cell Respiration and Cytochrome*. Cambridge: Cambridge University Press.
- Slater E (ed.) (1966) *Flavins and Flavoproteins*. Amsterdam: Elsevier.

Sphingolipid Biosynthesis

A H Merrill Jr., Georgia Institute of Technology, Atlanta, GA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Ceramide An N-acyl-derivative of sphingosine that is both a metabolic intermediate and a cell-signaling molecule. In some cases, the term is applied generically to any N-acyl-sphingoid base.

Glycosphingolipid A compound with a carbohydrate bound to a sphingoid base (and most often, attached to position 1 of an N-acyl-sphingoid base).

Glycosyltransferase An enzyme that transfers a carbohydrate from a donor (usually a UDP-sugar) to an acceptor which, in the case of sphingolipids, is either a ceramide or a carbohydrate chain attached to ceramide.

Mass spectrometry An instrument that separates ionized molecules by their mass to charge ratio (m/z) to allow them to be structurally characterized and quantified, with appropriate internal standards.

Phosphosphingolipid A compound with a phosphate or phosphodiester linked headgroup attached to a sphingoid base (or more often, to position 1 of an N-acyl-sphingoid base).

Sphingoid base The backbone of more complex sphingolipids as well as a cell-signaling molecule. Structurally, a long-chain alkane (or alkene) with an amino at position 2, and (usually) hydroxyl groups at position 1 and 3 plus various alkyl chain lengths, degrees of unsaturation, and additional hydroxyl groups.

Sphingolipidomics The analysis of all of the sphingolipids in a biological system (the sphingolipidome or a logically selected subset that includes all of the subspecies of interest for a particular biological system or phenomenon).

Sphingosine 1-phosphate A bioactive metabolite that serves as an intracellular and an extracellular signal as well as an intermediate of sphingoid base catabolism.

Introduction

Sphingolipids are a complex family of compounds produced by eukaryotes and some prokaryotes and viruses to play structural roles, such as formation of membrane bilayers and rafts, stabilization of lipoproteins, and maintenance of the water barrier of skin, and to serve as intra- and extra-cellular signaling molecules. They share a common structural feature, a so-called sphingoid base backbone that is comprised of an alkyl chain of approximately 18 carbon atoms with one to three hydroxyl groups and one amino group at position 2 (or in one case, at position 1) of the alkyl chain. Sphingoid bases are made *de novo* by mammals and converted into more complex sphingolipids (sphingoid base 1-phosphates, ceramides, phosphosphingolipids, glycosphingolipids and protein adducts). Several diseases result from genetic defects in *de novo* sphingolipid biosynthesis or disruption of the pathway by environmental factors, and many diseases are characterized by abnormal sphingolipid amounts and/or compositions. Therefore, sphingolipid biosynthesis is being explored for its possible role in the etiology of many diseases, such as cancer, and possibly as new targets for drugs and other interventions.

Structures and Nomenclature

Sphingolipids can be divided into several major categories: the sphingoid bases and their simple derivatives, ceramides, and more complex sphingolipids as summarized in Figures 1 and 2.

Sphingoid Bases

The most fundamental feature of these compounds was named sphingosine by J. L. W. Thudichum when he reported its discovery (in 1884) "in commemoration of the many enigmas which it presented to the inquirer." The structure of sphingosine, the major sphingoid base of mammals, is (2*S*,3*R*,4*E*)-2-aminooctadec-4-ene-1,3-diol (it is also called 'D-erythro-sphingosine' and 'E-sphing-4-enine') (Figure 1). The other sphingoid bases found in mammals vary in alkyl chain length and branching, the number and positions of double bonds, the presence of additional hydroxyl groups, and a few other features. Mammals have also been discovered recently to have atypical sphingoid bases that lack the 1-hydroxyl group, as shown for 1-deoxysphinganine in Figure 1.

Sphingoid bases function as intra- and extra-cellular signals and second messengers in the form of free sphingoid bases and derivatives (sphingoid base 1-phosphates and ceramides, and possibly other species) (Figure 1). Free (underivatized) sphingoid bases are usually present in cells in relatively small amounts versus more complex sphingolipids.

Ceramides

Ceramides are fatty acid derivatives of sphingoid bases (Figure 1). The fatty acids are typically saturated or mono-unsaturated with chain lengths from 14 to 26 carbon atoms (or even longer in the special case of skin), and sometimes

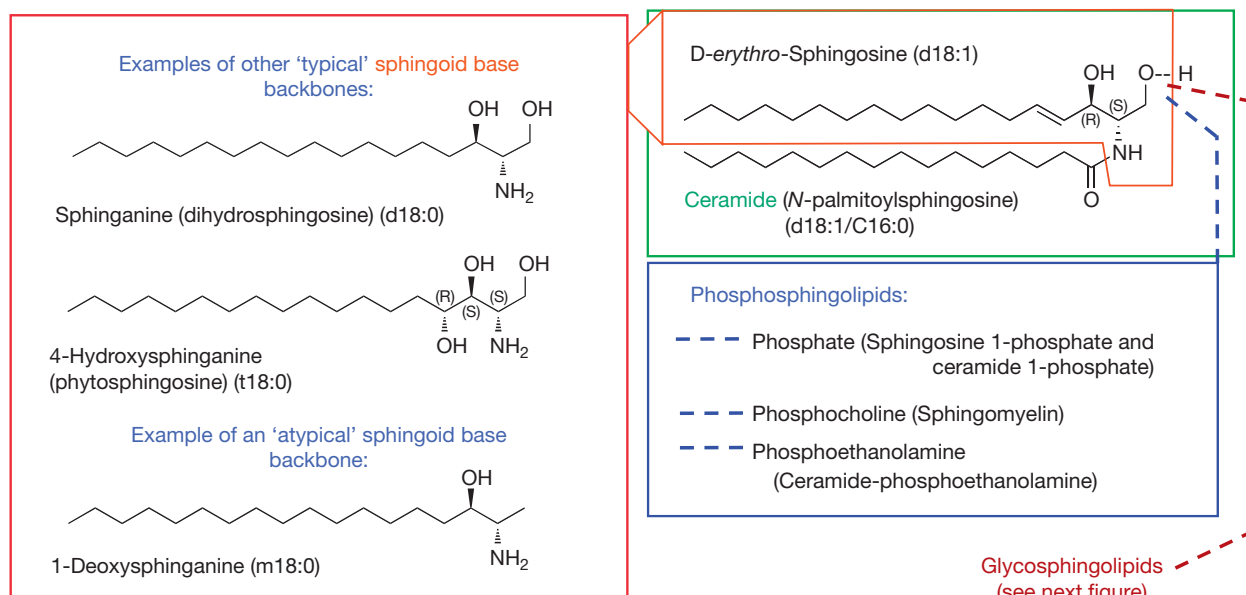


Figure 1 Structures of representative sphingolipids. Shown are several examples of sphingoid bases (sphingosine, sphinganine and 4-hydroxysphinganine, as well as an atypical sphingoid base, 1-deoxysphinganine), a ceramide (N-acyl-sphingosine), and headgroup phosphosphingolipids, such as sphingomyelin (SM), and glycosphingolipids, which are shown in more detail in **Figures 2** and **4**.

have a hydroxyl group on the α - or ω -carbon atom. These structural features favor the segregation of ceramides and some complex sphingolipids into specialized regions of the membrane (called rafts and caveolae) that participate in cell signaling, nutrient transport, and other functions.

Ceramides also serve as second messengers that modulate the biophysical properties of membranes, activate and inhibit proteins that participate in cell signaling (e.g., protein kinases and phosphoprotein phosphatases) and control cell growth, senescence, and programmed cell death (apoptosis). Their biologic activity depends on the type of sphingoid base and fatty acid; for examples, dihydroceramides (i.e., without the 4,5-double bond of the sphingosine backbone) (**Figure 1**) are less potent than ceramides as inducers of apoptosis, but are able to induce the formation of large intracellular vacuoles called autophagosomes.

More Complex Phospho- and Glyco-Sphingolipids

The major phosphosphingolipids of mammals are sphingomyelins (**Figure 2**), followed by much smaller amounts of ceramide 1-phosphate, which participates in cell signaling, and trace levels of ceramide phosphoethanolamines. Glycosphingolipids are classified on the basis of carbohydrate composition. One way to categorize them is by charge: neutral glycosphingolipids contain one or more uncharged sugars such as glucose (abbreviated Glc, hence, glucosylceramide is GlcCer), galactose (Gal), N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc), and fucose (Fuc); whereas, acidic glycosphingolipids contain phosphate or sulfate (the latter being called sulfatides) attached to neutral sugars, or charged sugar residues such as glucuronic acid or sialic acid (which is actually a general name for a family of compounds, for the type in human sphingolipids

is N-acetylneuraminic acid, Neu5Ac) (**Figure 2**). Gangliosides, such as those shown in the figure, are a subcategory of glycosphingolipids with sialic acid, and the number of sialic acid residues is usually denoted with a subscript letter (i.e., mono-, di- or tri-) plus a number (and sometimes additional symbols) reflecting the subspecies within that category, as shown for gangliosides GM1a, GM2, and GM3.

Another nomenclature system, which is becoming more common, is to categorize the glycans by a so-called root structure, as shown in **Figure 2** for the tetrahexose backbone of GM1a (Gal β 1-3GalNAc β 1-4Gal β 1-4Glc β 1Cer). This is called the ganglio- root structure (which is a bit confusing with the term gangliosides referring to sialic acid containing glycosphingolipids). The other root structures for mammalian glycosphingolipids are: lacto- (Lc), Gal β 1-3GlcNAc β 1-3Gal β 1-4Glc β 1-1'Cer; neo-lacto (nLc), Gal β 1-4GlcNAc β 1-3Gal β 1-4Glc β 1-1'Cer; globo (Gb), GalNAc β 1-3Gal α 1-4Gal β 1-4Glc β 1-1'Cer; and isoglobo (iGb), GalNAc β 1-3Gal α 1-3Gal β 1-4Glc β 1-1'Cer; and root structures found in other organisms, such as insects, are mollu (Mu), GalNAc β 1-2Man α 1-3Man β 1-4Glc β 1-1'Cer and arthro (At), GalNAc β 1-4GlcNAc β 1-3Man β 1-4Glc β 1-1'Cer. By this system, the structure is described by the root name plus any substituents, with designation of where they are added by a Roman numeral with a subscript for the hydroxyl to which it is attached (II³ in **Figure 2** for the position of attachment of the Neu5Ac in ganglioside GM1a).

A few glycosphingolipids are still referred to by historically assigned names, such as the blood group antigens (e.g., Lewis x and sialyl-Lewis x). Because these terminologies – and the structures themselves – can be confusing, there is growing usage of symbols to display the glycans, and a generally well-followed convention to depict them is shown in **Figure 2** (e.g., glucose as a light blue circle, glucosamine as a square with the same color, etc.).

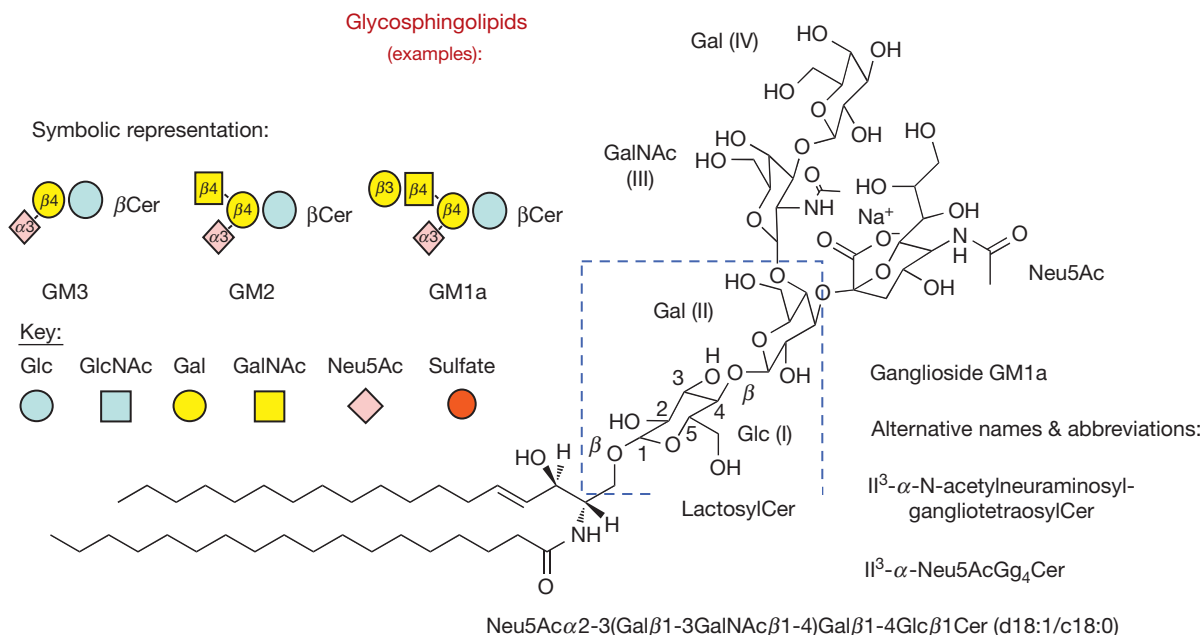


Figure 2 Structure of representative glycosphingolipids. The chemical structure shown is for ganglioside GM1a, but the structure also displays its component parts that represent a monohexosylceramide (glucosylceramide, Glc), a dihexosylceramide (lactosylceramide), and other subcategories of gangliosides (GM3 and GM2), with alternative ways of depicting such compounds (i.e., by explicit structures or symbols for the carbohydrates).

Protein Adducts

Some sphingolipids are covalently attached to protein, for example, ω -hydroxy-ceramides and -glucosylceramides are attached to surface proteins of skin; and, inositol-phosphoceramides are used as membrane anchors for some fungal proteins, in a manner somewhat analogous to the glycosylphosphatidylinositol (GPI)-anchors that are attached to proteins in other eukaryotes.

De novo Biosynthesis of Sphingolipids

Sphingolipid biosynthesis is widespread among eukaryotic cells, a few bacteria, and one recently discovered algal virus, and it appears that new synthesis (i.e., *de novo*) is relied upon more than reutilization of sphingolipids from exogenous sources, such as food. The biosynthetic pathway for such a diverse family of compounds (conservatively estimated to be in the tens of thousands) is obviously complex, however, its fundamental features are summarized in [Figure 3](#) with the steps in the biosynthesis of the lipid backbone at the top one-third of the scheme and most of the early steps in headgroup addition in the lower two-thirds.

Biosynthesis of the Lipid Backbone

The initial step of the pathway is catalyzed by serine palmitoyltransferase (SPT), a pyridoxal 5'-phosphate-dependent heterodimer (of SPTLC1 and SPTLC2, and perhaps additional proteins) that condenses serine and palmitoyl-CoA to

form a 3-keto-intermediate that is rapidly converted to sphinganine by an NAD(P)H-dependent reductase ([Figure 1](#)). The palmitoyl-CoA selectivity of this enzyme is a major determinant of the 18-carbon chain length of the majority of mammalian sphingoid bases, and recent studies indicate that additional SPT isozymes and/or associated proteins can shift the fatty acyl-CoA selectivity to produce other types (for example, elevation of SPTLC3 increases the production of d16:0 sphingoid bases).

SPT has been recently discovered to utilize alanine to produce 1-deoxysphinganine (m18:0) ([Figure 1](#)) and glycine to make 1-(deoxymethyl)sphinganine (m17:0). These atypical sphingoid bases are found *in vivo* and are probably not mistakes, but additional subcategories of sphingolipids with their own biochemical functions. Moreover, there is growing evidence that the neuronal damage in human hereditary sensory neuropathy type I (HSN1), the most common hereditary disorder of peripheral sensory neurons, is due at least in part to an abnormal elevation in 1-deoxysphingoid bases due to mutations in SPT.

Sphinganine is next N-acylated by ceramide synthases (CerS) to produce dihydroceramides using, mostly, the fatty acyl-CoA's shown in [Figure 3](#). The proteins produced from the six known CerS genes vary in fatty acyl-CoA selectivity to allow specification of the chain-length distribution bases on the relative expression levels and the available fatty acyl-CoA's. CerS1 is highly specific for C18:0; CerS2 for very-long-chain fatty acyl-CoA's (ca C22 to C26); CerS3 appears to produce the very-long-chain Cer found in skin ($\geq$ C26); CerS4 utilizes fatty acyl-CoA's with $C20 \pm 2$; CerS5 is fairly specific for C16- and CerS6 utilizes C14- and C16-fatty acyl-CoA. The nature of the ceramide subspecies affects not only the biophysical properties of cell membranes, but also can profoundly alter cell behavior,

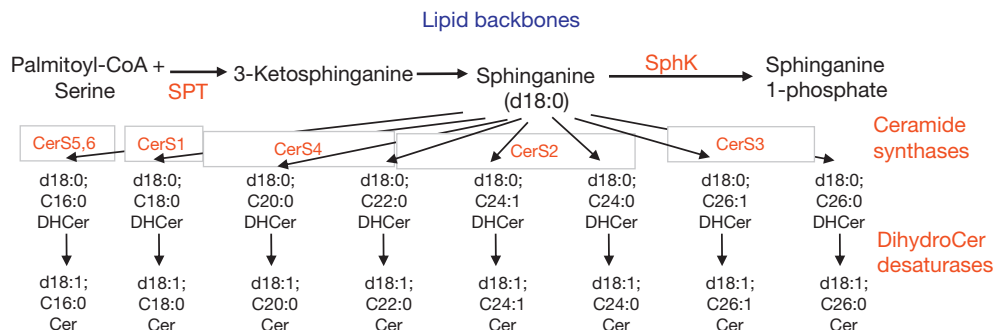


Figure 3 The *de novo* biosynthetic pathway for sphingoid bases and ceramides. Starting with serine palmitoyltransferase (SPT), the upper third of the figure depicts the biosynthesis of sphinganine, which can be phosphorylated by sphingosine kinase (SphK) or N-acylated by ceramide synthases (CerS) to dihydroceramides (d18:0; Cx,y) with the different acyl chains shown by chain length and number of double bonds (Cx,y). Ceramides (d18:1; Cx,y) are made by the dihydroceramide desaturases. Addition of the headgroups is shown in [Figure 4](#).

as exemplified by the finding that CerS1 and C18-ceramides are low in head and neck cancer, and their restoration partially normalizes the behavior of these cells in culture. Sphinganine can also be phosphorylated by sphingosine kinase (SphK) and, to some extent, acylation and phosphorylation are competitive because CerS and SphK reside in the endoplasmic reticulum.

Insertion of double bond(s) into the sphingoid base backbones occurs mainly after formation of dihydroceramide(s) ([Figure 3](#)). For mammals, introduction of the 4,5-double bond is catalyzed by a pyridine nucleotide-dependent desaturase, DES1; and possibly also by DES2, which is mainly responsible for addition of the 4-hydroxyl-group of phytoceramides.

The *de novo* biosynthetic pathway for ceramide is activated by a large number of factors, including bacterial endotoxins, cytokines, cancer chemotherapeutic drugs, and irradiation, and the increased production of ceramide is thought to be responsible for the toxicity of some of these treatments. Ceramide elevation was once thought to be the mechanism for the toxicity for the anti-cancer drug fenretinide (4-hydroxyphenylretinamide); however, later studies established that fenretinide inhibits DES and dihydroceramides are actually elevated, which induces autophagy.

Two useful reagents for studies of *de novo* sphingolipid biosynthesis are myriocin (ISP1), which is a potent inhibitor of SPT, and fumonisins, which inhibit CerS. Ceramide synthase inhibition appears to be the mechanism for the toxicity and carcinogenicity of fumonisins, which are mycotoxins produced by fungi that grow on corn.

Headgroup Addition for Formation of More Complex Sphingolipids

Ceramides in their various forms (i.e., ceramides, dihydroceramides, phytoceramides, etc.) are at a key branchpoint of complex sphingolipid biosynthesis where these intermediates are partitioned into phosphosphingolipids and glycosphingolipids, as shown in the lower portion of [Figure 4](#). The fate of a given intermediate is governed by the relative activities and selectivity of the enzymes at this branchpoint as well as by the subcellular localization of the participants.

Sphingomyelin and other phosphosphingolipids. Sphingomyelins are synthesized by at least two sphingomyelin synthases (SMSs) that transfer phosphorylcholine from phosphatidylcholine to ceramides in the endoplasmic reticulum, Golgi, and plasma membranes. This reversible reaction links glycerolipid and sphingolipid metabolism and signaling because ceramides and diacylglycerols both function as metabolic intermediates and as intracellular second messengers. This may explain why cells produce dihydroceramides as the initial products of *de novo* sphingolipid biosynthesis since that allows a relatively innocuous intermediate to accumulate if later steps in the pathway slow (although, as noted above, dihydroceramides are not inert because their elevation induces autophagy). A structurally related, endoplasmic reticulum enzyme is responsible for making ceramide phosphoethanolamines. Ceramide phosphate is made by phosphorylation of ceramide by ceramide kinase (CERK), which is found in the Golgi.

The transport of ceramide from its site of *de novo* biosynthesis (the endoplasmic reticulum) to the Golgi and plasma membrane is effected through both vesicular trafficking and the ceramide transport protein CERT. Because CERT prefers ceramides with ~C16 fatty acid chain length, downstream enzymes that utilize the ceramide that it has delivered are often enriched in that chain length. CERT is identical to a splicing variant of Goodpasture antigen-binding protein (GPBP), which plays a role in Goodpasture disease, an autoimmune disorder.

Glycosphingolipids. The addition of the carbohydrate headgroups is catalyzed by glycosyltransferases that transfer a specific sugar from the appropriate sugar nucleotide (UDP-Glc, UDP-Gal, CMP-sialic acid, etc.) to ceramide or the nonreducing end of the growing carbohydrate chain attached to ceramide. These enzymes are usually specific for the transferred carbohydrate, the carbohydrate (the type of transfer it will catalyze – i.e., α - or β - and to which particular hydroxyl-group on the glycan) and the general class of acceptor, although in some cases, the glycosyltransferases can add to both glycosphingolipids and glycoprotein chains. The abbreviations for the enzyme names in [Figure 4](#) provide information about the reaction they catalyze; for example: β 3GlcNAcT is a glycosyltransferase that transfers N-acetylglucosamine (GlcNAc) to the 3 position of the acceptor carbohydrate in a beta (β) linkage (a carbohydrate numbering scheme and β -linkage is shown for the glucose of ganglioside GM1a in [Figure 4](#)).

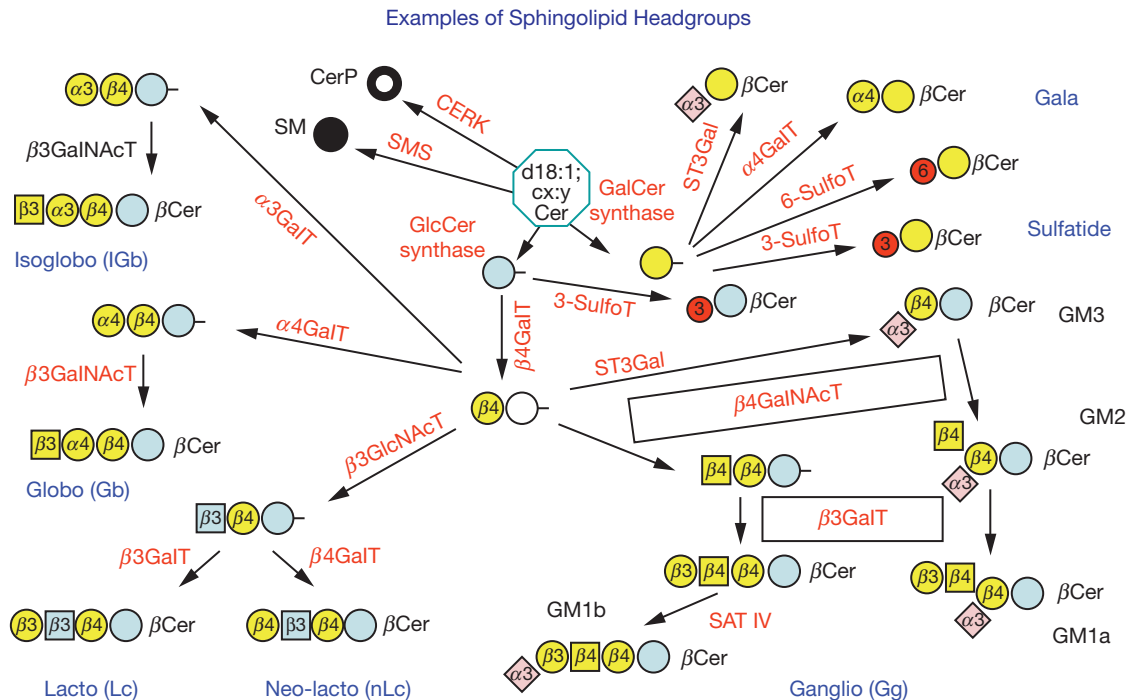


Figure 4 The biosynthetic pathway for complex sphingolipids. The key for this figure is the same as for Figures 1 and 2. Starting with a ceramide backbone (the octagon labeled “d18:1; cx:y”, as an example), the figure shows how ceramides are converted into sphingomyelins (SM) by SM synthases (SMS), phosphorylated to ceramide 1-phosphate (CerP) by ceramide kinase (CERK), or glycosylated to glucosylceramides (GlcCer) or galactosylceramides (GalCer) by the shown synthases. The other abbreviations are standard for the enzyme reactions shown, such as β 4-GalT, for the glycosyltransferase that transfers galactose (Gal) to the 4 position of the acceptor carbohydrate (GlcCer) in a beta (β) linkage. The enzymes that add sialic acid (Neu5Ac) are ST3Gal and SAT IV. Note the manner in which a single enzyme (such as β 4GalNAcT) can act on more than one substrate to make multiple downstream products in a ‘combinatorial’ manner, as depicted in the biosynthetic pathways for gangliosides GM1a and GM1b.

GlcCer and GalCer are the simplest glycosphingolipids, and are synthesized by UDP-Glc (or Gal): ceramide glycosyltransferases (coded for by the *Ugcg* and *Ugt8* genes, respectively). Hence, one determinant of the types of glycosphingolipids made by a given cell type is whether it expresses one or both of these genes. This branchpoint is also regulated by the availability of ceramide because GalCer is made in the lumen of the endoplasmic reticulum whereas GlcCer biosynthesis occurs on the side facing the cytoplasm (where enzymes for *de novo* ceramide biosynthesis are located) and/or Golgi. Therefore, even if the synthase for GalCer is expressed, the rate of synthesis of gala-family glycosphingolipids will depend on whether substantial amounts of ceramide flip from the outer to inner leaflet of the endoplasmic reticulum. Likewise, sulfatide biosynthesis is catalyzed by sulfotransferases inside the endoplasmic reticulum/Golgi and the nature of the products will depend on the type of sulfotransferases expressed (i.e., for 3'- or 5'-sulfo-) and the availability of the substrates (GalCer or other glycosphingolipids as shown in Figure 4) and the activated sulfate donor 3'-phosphoadenosine-5'-phosphosulfate.

The major fate of GlcCer is to traverse the Golgi membrane and be glycosylated to lactosylceramide (LacCer, Gal β 1-4Glc β 1Cer) in the lumen using UDP-Gal. LacCer synthase is thought to be encoded by some of the six members of the β -1,4-galactosyltransferase (β 4GalT) gene family, with β 4GalT-5 having been shown to be essential during early mouse embryogenesis. At this point, LacCer branches toward the root structure families of

glycosphingolipids, as shown in Figure 4, via the stepwise transfer of neutral sugars and sialic acid(s). These branches generally occur in different subcompartments within the Golgi. For example, the sialyltransferase catalyzing the synthesis of a simple ganglioside (ganglioside G_{M3}) is in the *cis* Golgi, whereas enzymes involved in terminal steps of more complex gangliosides are located in the more distal *trans* Golgi network.

Regulation of complex glycosphingolipid biosynthesis involves both transcriptional and posttranscriptional factors. For example, developmentally regulated, tissue selective variations in ganglioside amounts and types in mammalian tissues are under transcriptional control, but the activities of glycosyltransferases can be fine-tuned by posttranslational modification.

Other Metabolic Processes

In recent years, it has become evident that biosynthesis of sphingolipids for a particular cellular function also occurs outside of the endoplasmic reticulum/Golgi network. Some of these involve turnover of more complex sphingolipids with recycling of component part(s). One example of this is the hydrolysis of sphingomyelin to produce ceramide, which is subsequently phosphorylated to ceramide 1-phosphate. Cells also contain metabolites such as sphingosine 1-phosphocholine (lysosphingomyelin) that are involved in cell signaling, but for which a biosynthetic (or catabolic) pathway has not yet been discovered.

Modification of glycosphingolipids occurs at numerous sites, including the nucleus, mitochondria, and plasma membrane, presumably for the purpose of membrane remodeling or signaling. This includes removal of sialic acid(s) from gangliosides by a plasma membrane-associated sialidase (termed 'Neu3') and addition of sialic acid by at least one sialyltransferase (Sial-T2, or GD3 synthase), which has not only been noted at the cell surface, but also appears to exist as an ecto enzyme – that is, on the outer leaflet of the plasma membrane – using CMP-NeuAc available in the extracellular milieu.

This article has not discussed the classic lysosomal turnover of sphingolipids and the associated diseases from genetic defects in many of the steps of these pathways (e.g., Niemann-Pick, Gaucher and Sandhoff's disease, to name just a few). Interesting, some of the enzymes that were formerly thought to be dedicated to sphingolipid catabolism, such as the acidic sphingomyelinase, are now known to have functions in cell signaling and membrane repair. Therefore, they too should be borne in mind in thinking about the metabolic events taking place in the cell.

Sphingolipidomics

The large number and structural complexity of sphingolipids make quantitative analysis of all of the molecular subspecies technically impossible, and analysis of even a modest subset – such as all the ceramides – is difficult if the samples are small, such as cells in culture. However, large numbers of compounds can be analyzed by a combination of gas, liquid, or thin-layer chromatography with tandem mass spectrometry using electrospray or matrix-assisted laser-desorption ionization (MALDI) to convert the compounds of interest into ions that can be analyzed with a substantial degree of structural specificity by tandem mass spectrometry using quadrupole, time-of-flight, ion traps, or Fourier transform ion cyclotron resonance instruments. One of the major barriers to quantitative analysis is the lack of internal standards; however, this is being addressed by the LIPID MAPS Consortium (www.lipidmaps.org) in collaboration with Avanti Polar Lipids. Mass spectrometry can also be used to gain more information about sphingolipid biosynthesis and turnover by using stable isotope precursors, such as [^{13}C]palmitic acid, to label the sphingoid base backbone and N-acyl chains.

One of the most recent developments in sphingolipidomic mass spectrometry is the analysis of sphingolipids in tissue slices, which allows the location to be matched with what is known about the sample from more traditional histological staining. Examples of the application of tissue imaging mass spectrometry (also called MALDI-imaging mass spectrometry when that is the mode of ionization) include determination of the distribution of different gangliosides in specific regions of the brain, and localization of sulfatides in ovarian cancer cells versus stroma in biopsy samples. Although the resolution of tissue imaging mass spectrometry is currently at the histologic level (tens of microns), methods are under development that are using other means of generating ions to achieve sub-micron resolution, which will be of great value in basic and translational studies of these compounds.

Perspectives

For over a century, field of sphingolipidology has mystified scientists with its novelty and complexity, but in large part, what kept it mysterious was the lack of sufficiently sophisticated tools to study their structures, metabolism, and functions. This has changed radically in the last decade with the availability of mass spectrometry, immunohistochemistry, and other analytical tools, and identification of many of the genes for sphingolipid biosynthesis and turnover, allowing creation of cells and knockout mice to probe the functions of specific compounds. All of this new information is daunting in-and-of-itself; however, biosciences in general have recognized that they can only be fully understood by systems analysis – that is, how the multiple components of a living organism fit together, including the sphingolipids.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins: Sphingolipid Catabolism.**

Further Reading

- Bartke N and Hannun YA (2009) Bioactive sphingolipids: Metabolism and function. *Journal of Lipid Research* 50: S91–S96.
- IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) (1998) Nomenclature of glycolipids. Recommendations 1997. *European Journal of Biochemistry* 257: 293–298.
- Kolter T, Proia RL, and Sandhoff K (2002) Combinatorial ganglioside biosynthesis. *Journal of Biological Chemistry* 277: 25859–25862.
- Liu Y, Chen Y, Momin A, et al. (2010) Elevation of sulfatides in ovarian cancer: An integrated transcriptomic and lipidomic analysis including tissue-imaging mass spectrometry. *Molecular Cancer* 9: 186.
- Lopez PH and Schnaar RL (2009) Gangliosides in cell recognition and membrane protein regulation. *Current Opinion in Structural Biology* 19: 549–557.
- Merrill AH Jr. (2008) Sphingolipids. In: Vance DE and Vance JE (eds.) *New Comprehensive Biochemistry: Biochemistry of Lipids, Lipoproteins, and Membranes*, 5th edn., Chapter 13. Amsterdam: Elsevier Science.
- Merrill AH Jr., Stokes TH, Momin A, et al. (2009) Sphingolipidomics: A valuable tool for understanding the roles of sphingolipids in biology and disease. *Journal of Lipid Research* 50: S97–S102.
- Merrill AH Jr., Wang MD, Park M, and Sullards MC (2007) (Glyco)sphingolipidology: An amazing challenge and opportunity for systems biology. *Trends in Biochemical Sciences* 32: 457–468.
- Oskouian B and Saba JD (2010) Cancer treatment strategies targeting sphingolipid metabolism. *Advances in Experimental Medicine and Biology* 688: 185–205.
- Pruett ST, Bushnev A, Hagedorn K, et al. (2008) Biodiversity of sphingoid bases (sphingosines) and related amino alcohols. *Journal of Lipid Research* 49: 1621–1639.
- Tettamanti G, Bassi R, Viani P, and Riboni R (2003) Salvage pathways in glycosphingolipid metabolism. *Biochimie* 85: 423–437.
- Thudichum JLW (1884) *A Treatise on the Chemical Constitution of Brain*. London: Bailliere, Tindall, and Cox.
- Varki A, Cummings RD, Esko JD, et al. (2009) Symbol nomenclature for glycan representation. *Proteomics* 9: 5398–5399.
- Voit EO, Alvarez-Vasquez F, and Hannun YA (2010) Computational analysis of sphingolipid pathway systems. *Advances in Experimental Medicine and Biology* 688: 264–275.
- Yu RK, Nakatani Y, and Yanagisawa M (2009) The role of glycosphingolipid metabolism in the developing brain. *Journal of Lipid Research* 50: S440–S445.

Relevant Website

<http://www.lipidmaps.org> – Lipidomicsgateway.

Sphingolipid Catabolism

A Abe and J A Shayman, University of Michigan, Ann Arbor, MI, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 4, pp. 82–87, © 2004, Elsevier Inc.

Glossary

Cerebroside A glycosphingolipid containing a single carbohydrate, most commonly glucosylceramide and galactosylceramide.

Ganglioside An acidic glycosphingolipid containing one or more sialic acid groups.

Glycosphingolipid A sphingolipid covalently linked with one or more carbohydrates.

Lysosome An organelle bounded by a single membrane bilayer in the cytoplasm in eukaryotic cells and having an internal pH of 4–5. Lysosomes contain several hydrolytic enzymes and serve as the site for the degradation and recycling of cellular metabolites.

Sphingolipid A lipid with a backbone containing a long-chain sphingoid amine.

Sphingomyelin Catabolism

Sphingomyelinase catalyzes the hydrolysis of sphingomyelin to form phosphorylcholine and ceramide. Ceramide may be further metabolized to form other sphingolipids or may function as a signaling molecule. Sphingomyelinases are categorized into four groups: acid sphingomyelinase (aSMase), secreted sphingomyelinase (sSMase), neutral sphingomyelinase (nSMase), and alkaline sphingomyelinase (bSMase).

aSMase

aSMase is a well-characterized sphingomyelinase with an optimal pH of 5 and localized to lysosomes. The enzyme primarily functions in the degradation of sphingomyelin. In humans, a genetic deficiency of lysosomal aSMase results in Niemann–Pick types A and B, autosomal-recessive lipid storage disorders. The enzyme was purified from urine as a 72-kDa monomeric glycoprotein with a 61-kDa polypeptide core. The human aSMase complementary (cDNA) is encoded by the sphingomyelin phosphodiesterase 1 (*SMPD1*) resulting in a 629-amino-acid polypeptide. The *SMPD1* gene is located on chromosome 11. Metabolic labeling studies in cell-line cells transfected with the human aSMase cDNA reveal that a 75-kDa aSMase precursor is processed by extensive posttranslational modification during sorting to the lysosome.

There is evidence that aSMase plays an important role in ceramide formation after stimulation with tumor necrosis factor (TNF) α , lipopolysaccharide, CD95, or treatment with ultraviolet (UV)-A irradiation. Additional stressors, including ischemia and chemotherapeutic drugs, also activate aSMase and are associated with ceramide-induced cell death. However, the requirement of aSMase in apoptosis and differentiation induced by TNF α and Fas ligand is debated. The involvement of aSMase in ceramide-mediated signal transduction may depend on the cell and tissue type and vary with the stimulus.

sSMase

sSMase is the secretory form of the aSMase gene product and is secreted via a Golgi-apparatus-dependent pathway into the

extracellular space. The enzyme is activated in the presence of physiological concentrations of Zn²⁺ and demonstrates maximum activity under acidic conditions. However, the enzyme is able to hydrolyze sphingomyelin from atherogenic lipoproteins, low-density lipoprotein (LDL) extracted from arteriosclerotic lesions, and oxidized at neutral pH. LDL treated with sSMase forms aggregates that are retained on extracellular matrix and stimulates macrophage foam cell formation. Thus, sSMase may serve a normal physiological function in lipoprotein metabolism and a pathological function in atherogenesis. sSMase activity in plasma is higher in individuals with chronic heart failure and type 2 diabetics compared to healthy individuals. The active circulating form of sSMase may be involved in the formation of atherosclerotic plaques.

nSMase

In mammalian cells, nSMase is a membrane-associated protein with an optimal neutral pH. Although the enzyme is expressed ubiquitously in mammalian tissues, the highest activity is predominantly found in the brain. Mammalian tissues express two isoforms of nSMase. The enzyme activity of the higher molecular weight isoform form is Mg²⁺ dependent but that of the lower-molecular-weight isoform is Mg²⁺ independent. The low-molecular-weight isoforms appear to be degradation products of the high-molecular isoforms. The purified enzyme is activated by phosphatidylserine, Mg²⁺, and Mn²⁺, and inhibited by Cu²⁺, Zn²⁺, Ca²⁺, Cd²⁺, Hg²⁺, glutathione, and asialo-ganglioside, GM3.

Recently, three types of nSMases have been identified in mammalian cells at the gene level. nSMase 1 has a molecular mass of 47.5 kDa and is encoded by the *SMPD2* gene. nSMase 1 was cloned from mouse and human based on multiple sequence alignments of bacterial nSMases. The *SMPD2* gene is located on chromosome 6. The product of the *SMPD2* gene is a membrane-bound Mg²⁺-dependent nSMase, is expressed ubiquitously in mammalian tissues, and localized in endoplasmic reticulum, Golgi and/or the nuclear matrix of cells. The natural substrate of nSMase 1 is still uncertain and the enzyme appears not to be involved in ceramide formation after stimulation by TNF α . Recently, nSMase 1 has been shown to

contribute to the induction of ceramide-mediated proapoptotic signaling under heat stress conditions in zebrafish embryonic cells.

A second nSMase (nSMase 2 encoded by the *SMPD3* gene) with a molecular mass of 71 kDa has been cloned based on phylogenetic analysis. The *SMPD3* gene is located on chromosome 16. nSMase 2 is a brain-specific membrane-bound Mg^{2+} -dependent nSMase with a different domain structure and only marginal sequence similarity to other nSMases. nSMase 2 has the basic properties of rat and bovine brain nSMase and is activated in response to TNF α . nSMase 2 co-localizes with Golgi apparatus markers in a number of cell lines and is activated in response to stress-inducing agents including TNF α , staurosporin, C2-ceramide, and H_2O_2 . However, recent work indicates that nSMase 2 may not be involved in cell apoptosis mediated by TNF α or oxidized LDL.

A third nSMase, nSMase 3, is encoded by the *SMPD4* gene and has a molecular mass of 97 kDa. The *SMPD4* gene is located on chromosome 2. nSMase 3 has only slight homology to nSMase 1 and 2. Similar to nSMase2, nSMase3 is Mg^{2+} dependent, is membrane associated, and has an optimal pH of 7. nSMase3 activity is enhanced in the presence of phosphatidylserine and inhibited by scyphostatin. nSMase3 is ubiquitously expressed and localized to the endoplasmic reticulum as well as Golgi apparatus. In addition, nSMase 3 has been demonstrated to be a TNF-responsive nSMase. An additional nSMase candidate that is encoded by a different gene from nSMase 1, 2, and 3 was expression cloned from human DNA. This nSMase cDNA encodes a 397-amino-acid polypeptide. The enzyme is activated by Mg^{2+} , inhibited by Cu^{2+} and glutathione, and recognizes sphingomyelin as a preferred substrate. The deduced amino acid sequence of the enzyme indicates that the enzyme is a membrane-integrated protein and has a significant homology to the death domain of TNF- α and Fas/AP-1 receptors. The overexpression of this recombinant nSMase in human aortic smooth muscle cells results in apoptosis and augments oxidized LDL-induced apoptosis.

bSMase

The enzyme activity of this SMase was initially found in the small intestine. Recently, bSMase was purified from rat intestine and required bile salt for the enzyme activity. This bSMase with molecular mass of 58 kDa is encoded by the *ENPP7* (ectonucleotide pyrophosphatase/phosphodiesterase family member 7) gene, has an alkaline pH optimal, is Mg^{2+} independent, and is inhibited by orthovanadate, adenosine triphosphate (ATP), imidazole, and Zn^{2+} but not by glutathione. bSMase is a single-pass type 1 membrane protein and has a single anchor domain at C-terminus. The *ENPP7* gene is located on chromosome 17. Expression of this enzyme is specific to the intestinal mucosa. The enzyme is mainly localized to the microvillar membrane in small intestine, in endosome-like structures, and in the Golgi complexes of human enterocytes. In addition, the decrease of the enzyme activity has been shown to be associated with colon cancer. Another type of bSMase with an 85-kDa molecular mass is found in human bile. These enzymes appear to function in the catabolism of dietary sphingomyelin.

Ceramide Catabolism

Ceramidase (CDase) catalyzes the hydrolysis of ceramide to fatty acid and sphingosine. Three types of ceramidase have been described based on their pH optima for activity. These include acid ceramidase, neutral ceramidase, and alkaline ceramidase. Sphingosine and its phosphorylated metabolite, sphingosine-1-phosphate, act as potent inhibitors of protein kinase C and potent effectors of cell proliferation and differentiation. CDase may change the balance of ceramide, sphingosine, and sphingosine-1-phosphate within cells in response to various stimuli. CDase may therefore regulate sphingolipid-mediated signaling events.

Acid Ceramidase

Acid ceramidase (aCDase) can be characterized as an *N*-acylsphingosine deacylase with an acidic pH optimum. The enzyme is localized in lysosomal and endosomal compartments. The enzyme functions primarily in the degradation of ceramide. In humans, a genetic deficiency of lysosomal aCDase results in the lysosomal lipid storage disorder known as 'Farber disease'. The enzyme, purified from human urine, is a 55-kDa heterodimeric glycoprotein consisting of two disulfide-linked polypeptide chains of 13 kDa (α) and 40 kDa (β). The human aCDase cDNA is encoded by the *ASHA1* gene, located on chromosome 8, and predicted to encode a protein consisting of 395 amino acids. The 13-kDa (α) and 40-kDa (β) subunits are derived from a common 55-kDa precursor encoded by the full length of aCDase-cDNA. Only the β -subunit is posttranslationally glycosylated during transport to acidic cellular compartments. aCDase activity is enhanced by an activator protein known as 'saposin D'. aCDase activity increases in response to extracellular stimuli. In rat hepatocytes, aCDase activity is bimodally regulated by interleukin (IL)-1 β and is activated by tyrosine phosphorylation. In renal mesangial cells, aCDase is activated by TNF α but inhibited by nitric oxide. Inhibition of aCDase sustains the accumulation of ceramide induced by TNF α . Overexpression of aCDase in L929 cells suppresses TNF α -induced ceramide formation and cell death.

Neutral Ceramidase

In mammals, neutral ceramidase (nCDase) is present in a variety of tissues and cell types. nCDase activity is mainly found in the membrane fractions. nCDase has been purified from rat brain, where the highest activity is found. nCDase is a 90-kDa membrane-bound, non-lysosomal protein. Similar membrane-bound nCDases, purified from mouse liver and rat kidney, are 94 kDa and 112 kDa monomeric glycoproteins, respectively. The enzyme has a broad pH optimum in the neutral to alkaline range and does not require divalent cations for the activity. The enzyme is stimulated by phosphatidic acid and phosphatidylserine. Partial amino acid sequences of each nCDase were used to clone the genes from mouse, rat, and human. The mouse, rat, and human nCDase cDNAs encode 756, 761, and 782 amino acid polypeptides, respectively.

The human nCDase is encoded by the *ASHA2* gene, which is located on chromosome 10, and ubiquitously expressed. The enzyme is found in the plasma membrane when expressed in HEK293 cells. The mouse nCDase is ubiquitously expressed and at high levels in kidney and liver, and is localized to the plasma membrane. The rat nCDase is expressed in kidney, brain, and heart at high levels. In rat kidney, nCDase is mainly localized at the apical membrane of the proximal tubules, distal tubules, and collecting duct. By contrast, liver nCDase are detected in endosome-like organelles within the hepatocytes. nCDase activity is activated by IL-1 β in rat hepatocytes, resulting in a decrease of ceramide-concomitant with an increase of sphingosine. TNF α stimulates and nitric oxide inhibits nCDase activity in renal mesangial cells.

Alkaline Ceramidase

Alkaline ceramidase activity has been described in human, rat, and mouse tissues. Recently, three alkaline ceramidases have been genetically characterized and show similar protein sequences but have no homology to acid or neutral ceramidases. Alkaline ceramidase 1 (alkaline CDase 1) is encoded by the gene of *ACER1*, consisting of 264 (human) and 287 (mouse) amino acids. Alkaline CDase 1 contains putative transmembrane domains. Alkaline CDase 1 is highly and selectively expressed in skin and localized to the endoplasmic reticulum. Alkaline CDase 1 has pH optimum at 8–8.5 and is activated by Ca²⁺. The enzyme activity is inhibited by Zn²⁺, Cu²⁺, Mn²⁺, and sphingosine. The human *ACER1* gene is located on chromosome 19. Alkaline CDase 1 may regulate the levels of ceramide, sphingosine, and sphingosine-1-phosphate. Recently, alkaline CDase 1 has been shown to mediate growth arrest and differentiation of human keratinocytes through sphingosine-1-phosphate.

Human alkaline ceramidase 2 (alkaline CDase 2) is encoded by the *ACRE2* gene, consists of 275 amino acids, and also contains putative transmembrane domains. Alkaline CDase 2 is ubiquitously expressed at low levels but at higher levels in placenta, and is localized to the Golgi apparatus. The *ACRE2* gene is located on chromosome 9. This enzyme contributes to sphingosine-1-phosphate-mediated cell proliferation and survival through the generation of sphingosine-1-phosphate. In addition, alkaline CDase 2/sphingosine pathway regulates β 1 integrin maturation and cell adhesion. The mouse alkaline ceramidase 2 shows an almost identical protein sequence to the human alkaline CDase 2.

The human alkaline ceramidase 3 (alkaline CDase 3) was originally cloned based on sequence homology to the yeast alkaline ceramidase. The enzyme is encoded by the *ACER3* gene and encodes a protein consisting of 267 amino acids. The enzyme has a pH optimum at 9.5 and is activated by Ca²⁺, but is inhibited by Zn²⁺ and sphingosine. The enzyme hydrolyzes phytoceramide preferentially, is ubiquitously expressed at high levels, and, when overexpressed in COS-1 cells, is localized in both the endoplasmic reticulum and Golgi apparatus. The *ACER3* gene is located on chromosome 11. Alkaline CDase 3 may contribute to the clearance of phytoceramide in mammalian cells but not to the generation of sphingosine from ceramide. The mouse

alkaline ceramidase 3 has an almost identical protein sequence and the same cellular localization as the human alkaline ceramidase 3.

Sphingoid Base Catabolism

At the penultimate step of sphingolipid catabolism, the primary hydroxyl group of the sphingoid base is phosphorylated by sphingosine kinase. The phosphorylated sphingoid base is then cleaved between the vicinal carbons with amino and hydroxyl groups by sphingosine-1-phosphate lyase (SPLase). The predominant sphingoid base in mammalian cells is sphingosine. Sphingosine produced from ceramide by ceramidase is phosphorylated by sphingosine kinase and then degraded to form phosphoethanolamine and hexadecanal by SPLase. The sphingosine-1-phosphate, generated during sphingosine catabolism, is not only a catabolic intermediate but also functions as an extracellular and intracellular messenger. Sphingosine-1-phosphate is mitogenic. The extracellular actions of sphingosine-1-phosphate are mediated through the endothelial differentiation gene (EDG) receptor family of G-protein-coupled receptors, now termed 'sphingosine-1-phosphate receptors'.

Sphingosine-1-Phosphate Lyase

SPLase (SGPL1) is a pyridoxal phosphate-dependent member of a class of enzymes known as 'aldehyde lyases' and cleaves 1-phosphorylated sphingoid bases into aliphatic fatty aldehydes and phosphoethanolamine with a neutral pH optimum. SPLase is a ubiquitous enzyme present in multiple species and mammalian tissues. However, platelets lack SPLase activity, suggesting that platelets may be the primary source of circulating sphingosine-1-phosphate. The enzyme is a membrane-bound protein and in rat liver is localized in the endoplasmic reticulum. The catalytic site and other domains essential for SPLase activity face the cytosol. SPLase cDNAs were cloned from mouse and human based on sequence homology to the *Caenorhabditis elegans* and yeast SPLases. Both cDNAs encode proteins of 568 amino acids. Hydropathy analysis indicates the presence of one transmembrane region near the N-terminus. The human *SGPL1* is located on chromosome 10.

Sphingosine-1-phosphate, a bioactive lipid that is a ligand for a family of G-protein-coupled receptors, regulates a variety of physiological functions. Recent studies have shown that SPLase play a role in cell survival, normal development, stress responses, the maintenance of tissue integrity, and immunity. In addition, SPLase has been demonstrated to function as a tumor suppressor and be involved in lymphocyte trafficking. Unlike mammalian and plant cells, in *Leishmania* patatices, SPLase is responsible for the major route for synthesis of phosphatidylethanolamine, an abundant membrane lipid.

Sphingosine-1-phosphate levels in cells during signaling processes seem to be regulated not only by sphingosine kinase and SPLase but also by a general lipid phosphate phosphohydrolase. In fact, several sphingosine-1-phosphate-specific phosphohydrolases have been identified in yeast and mammalian cells.

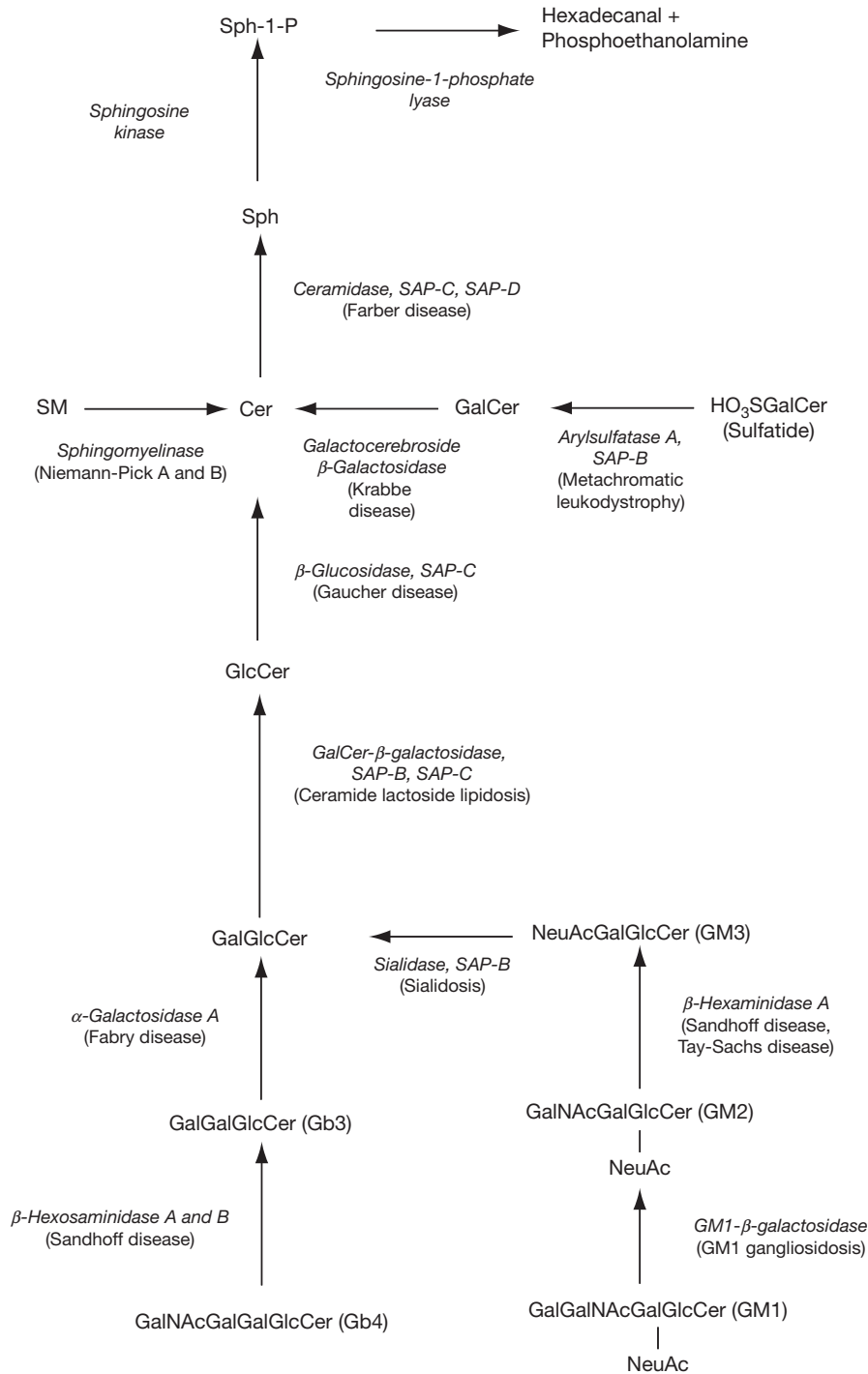


Figure 1 Degradation pathways of sphingolipids. The lysosomal diseases caused by genetic defects in a series of lysosomal enzymes and activator proteins for sphingolipid catabolism are indicated in parentheses. NeuAc, *N*-acetylneuramic acid; Cer, ceramide; Glc, glucose; Gal, galactose; GalNAc, *N*-acetylgalactosamine; Gb, globoside; Gb3, globotriaosyl ceramide; SM, sphingomyelin; HO₃SGalCer, sulfatide; Sph, sphingosine; Sph-1-P, sphingosine-1-phosphate; SAP, sphingolipid activator protein.

Glycosphingolipid Catabolism

Glycosphingolipids are components of cellular membranes and are comprised of one or more sugars linked to ceramide. Glycosphingolipids are tissue and cell-type specific. They form patterns that vary with the cell type, stage of growth and

differentiation, viral transformation, and ontogeny. The biosynthesis of glycosphingolipids occurs within the Golgi apparatus where carbohydrates are added by membrane-associated glycosyltransferases. Glycosphingolipids are subsequently transferred to the plasma membrane. At this site they interact with membrane-associated receptors and enzymes as well as

Table 1 Major sphingolipidoses

<i>Disease</i>	<i>Sphingolipid stored</i>	<i>Defective enzyme</i>	<i>Clinical phenotype</i>
Farber	Ceramide	Acid ceramidase	Painful and deformed joints, subcutaneous nodules, hoarseness
Niemann–Pick A and B	Sphingomyelin	Acid sphingomyelinase	Neurodegeneration, hepatosplenomegaly
Metachromatic leukodystrophy	Sulfatide	Arylsulfatase	Blindness, quadraparesis, seizures
Krabbe	Galactosylceramide	Galactocerebrosidase β -galactosidase	Blindness, spastic paraparesis, dementia
Gaucher	Glucosylceramide	β -Glucocerebrosidase	Hepatosplenomegaly, bone infarctions and fractures
Fabry	Globotriaosylceramide	α -Galactosidase A	Parasthesias, renal failure, cerebrovascular disease
Sandhoff	Gangliosides GM2 and GA2, globoside	Hexosaminidase B	Neurodegeneration
Tay–Sachs	Ganglioside GM2	Hexosaminidase A	Neurodegeneration
GM1 gangliosidosis	Ganglioside GM1	GM1 ganglioside β -galactosidase	Neurodegeneration, skeletal dysplasia, hepatosplenomegaly

with bacterial toxins and adhesion molecules and viruses. Following endocytosis, glycosphingolipids are transported to acidic intracellular compartments where they are degraded from the nonreducing ends by a set of lysosomal exoglycosidases that catalyze the stepwise cleavage of their component carbohydrates.

Glycosphingolipid storage disorders are caused by the deficiency of specific glycosidases that are involved in glycosphingolipid degradation (Figure 1). These deficiencies result in the abnormal accumulation of specific glycosphingolipids both in neural and nonneural tissues (Table 1). These disorders vary in terms of their clinical presentations based on the affected organs.

The lysosomal hydrolases that degrade glycosphingolipids of the plasma membrane are water-soluble enzymes. Their activity is dependent on the presence of water-soluble activator proteins termed 'saposins'. The four saposins (A through D) arise from a common precursor protein, prosaposin. Saposin B is an activator for arylsulfatase A, α -galactosidase A, and β -galactosidase. Saposin C is an activator of acid β -glucosidase. Saposin D is an activator of ceramidase. There is no assigned function for saposin A. Glycosphingolipid storage disorders may also arise from inherited deficiencies of the prosaposin as well as from individual saposins.

Arylsulfatase A

Arylsulfatase A catalyzes the desulfation of 3-O-sulfogalactosyl residues in glycosphingolipids. The enzyme activity requires the presence of saposin B as an activator. The ARSA gene maps to chromosome 22 at the end terminus and encodes a 507-amino-acid precursor protein that undergoes posttranslational processing. In addition to N-linked glycosylation required for lysosomal sorting through the mannose-6 phosphate receptor pathway, there is a unique oxidation that occurs for eukaryotic sulfatases. In human arylsulfatase A, a formylglycine residue is found in place of cysteine 69 and is due to the oxidation of a thiol group to an aldehyde. In addition to sulfatide, arylsulfatase A will cleave sulfate groups

from other naturally occurring glycosphingolipids including lactosylceramide-3-sulfate and psychosine sulfate.

β -Galactosidases

β -Galactosidase (GLB1) catalyzes the degradation of galactosylceramide to galactose and ceramide within the lysosome. It also displays activity against galactosylsphingosine and lactosylceramide. The GLB1 gene is localized to chromosome 3p21.33 and encodes a 677-amino-acid protein. The enzyme is more precisely referred to as GM1 ganglioside β -galactosidase and deficiencies are associated with GM1 gangliosidosis. GM1 ganglioside β -galactosidase has an acidic pH optimum. It is activated by chloride ions. The enzyme is isolated as a large-molecular-weight multimer with monomeric units of 65 kDa. This enzyme is also activated in the presence of saposin B. GM1 ganglioside β -galactosidase reportedly catabolizes ganglioside GM1, GA1, lactosylceramide, and keratan sulfate.

Galactocerebrosidase β -galactosidase (GALC) can be isolated as an 80-kDa polypeptide consisting of 50 and 30 kDa subunits. The pH optimum of the enzyme is acidic. The enzyme is active against galactosylceramides that vary in fatty acid chain length and α -hydroxylation. The enzyme is inhibited by sphingosine, ceramide, and galactose. The human gene for GALC maps to chromosome 14q31. Autosomal recessive mutations are associated with Krabbe disease.

β -Glucocerebrosidase

Human acid β -glucocerebrosidase (glucoceramidase) is a homomeric glycoprotein. The GBA gene localizes to chromosome 1q21 and encodes a mature polypeptide that is 497 amino acids in length. Both saposin A and saposin C activate the enzyme *in vitro*. However, only saposin C deficiency is associated with the clinical manifestations of Gaucher disease. Although saposin interacts directly with the enzyme, the association of glycosphingolipids and anionic phospholipids are required for its activity. The pH optimum for the glucocerebrosidase is 5.5. The enzyme is specific for D-glucosyl and not

1-glucosyl forms of glucosylceramide. Activity is reduced by the absence of the C4–C5 *trans* double bond in sphingosine and is minimally affected by the acyl chain length of the ceramide moiety. The disease alleles in Gaucher disease are commonly missense mutations. This results in the synthesis of β -glucosidases with decreased catalytic activity or stability. These typically lead to the type I form of Gaucher disease that lacks central nervous system manifestations. The neuroopathic forms of Gaucher disease (types II and III) are more commonly associated with complete loss of catalytic activity.

α -Galactosidase A

α -Galactosidase A (GLA) catabolizes glycosphingolipids with terminal α -galactosyl groups. These glycosphingolipids primarily include globotriaosylceramide, but are also present in galabiosylceramide and group B blood antigens. Initial studies on the enzyme activity suggested that there were two isozymes termed ' α -galactosidase A and B'. Subsequent work demonstrated that these are two genetically distinct enzymes. α -Galactosidase B is an α -N-acetylgalactosaminidase. Deficiencies in α -galactosidase B are responsible for Schindler disease. Native GLA has a molecular weight of 101 kDa and represents a homodimer with 49 kDa subunits. The enzyme is heavily glycosylated with asparagine-linked high mannose groups. The pH optimum for the glycosidase is acidic. The enzyme is activated in the presence of saposin B. The locus for the human α -galactosidase A gene (GLA) is found at Xq22.

Hexosaminidases A and B

Ganglioside GM2 is degraded in the lysosome by β -hexosaminidase A. This enzyme removes the terminal N-acetylgalactosaminyl residue. The enzyme requires the presence of an additional protein termed 'GM2 activator' encoded

by the GM2A gene. β -Hexosaminidase A is a heterodimer consisting of an α - and β -subunit. The α -subunit is a 55-kDa polypeptide encoded by the HEXA gene found on chromosome 15q23-24. The β -polypeptide varies between 22 and 30 kDa and is encoded by the HEXB gene found on chromosome 5q13. A second enzyme, β -hexosaminidase B, is a homodimer consisting of two β -subunits. Thus, HEXA gene mutations result in loss of β -hexosaminidase A activity, and HEXB gene mutations result in loss of both β -hexosaminidase A and B activities. The β -hexosaminidase A and B enzymes are posttranslationally glycosylated for sorting to the lysosome through the mannose-6-phosphate receptor pathway. They display acidic pH optima as well.

Inherited mutations in HEXA, HEXB, and GM2A all result in the accumulation of ganglioside GM2 and are associated with Tay–Sachs, Sandhoff, and GM2 activator-deficiency diseases, respectively. β -Hexosaminidase B deficiencies are associated with the accumulation of GA2 as well.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Glycolipid-Dependent Adhesion Processes; Lipid Bilayer Structure; Lipid Rafts; Sphingolipid Biosynthesis; **Signaling:** Lysophospholipid Receptors.

Further Reading

- Hannun YA and Obeid LM (2008) Principles of bioactive lipid signalling: Lessons from sphingolipids. *Nature Reviews in Molecular Cell Biology* 9: 139–150.
- Kolter T and Sandhoff K (2006) Sphingolipid metabolism diseases. *Biochimica et Biophysica Acta* 1758: 2057–2079.
- Pyne NJ, Long JS, Lee SC, et al. (2009) New aspects of sphingosine 1-phosphate signaling in mammalian cells. *Advances in Enzyme Regulation* 49: 214–221.
- Stancevic B and Kolesnick R (2010) Ceramide-rich platforms in transmembrane signaling. *FEBS Letters* 584: 1728–1740.

Src Family of Protein Tyrosine Kinases

J A Cooper, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 4, pp. 93–98, © 2004, Elsevier Inc.

Glossary

C-terminal Src kinase (Csk) Kinase that phosphorylates the C-terminal tyrosine residue present on all Src-related kinases, thus inhibiting Src kinase activity.

Protein tyrosine kinase A protein kinase that phosphorylates tyrosine residues in substrates.

SH2 domain A protein module that binds with high affinity to phosphotyrosine residues contained in a defined peptide sequence, with primary specificity usually being conferred to the three residues C-terminal to the phosphotyrosine.

SH3 domain A protein module that binds to sequences that adopt a left-handed type II polyproline helix, typically with a PXXP core.

The Src (sarcoma) family is a group of protein tyrosine kinases closely related to Src. They are nonreceptor kinases, meaning that they lack a transmembrane domain and lie totally within the confines of the cell, although they are associated with cell membranes. Src family kinase activity is regulated by phosphorylation sites within the kinase-domain activation loop and at the carboxy terminus, and by inhibitory interactions between the kinase domain and two other conserved domains, the SH2 and SH3 scaffolding domains. These latter domains can either inhibit the kinase domain or bind other cell proteins, thus helping localize Src to specific subcellular regions and protein complexes while simultaneously regulating Src kinase activity. There are eight Src family members in vertebrates, with different but overlapping expression patterns and, potentially, distinct functions within a given cell. However, there is also extensive genetic redundancy within the family, so inactivating mutations reveal only a subset of the potential cellular and developmental roles of family members. On the other hand, any of a number of different mutations can activate Src kinases, and the resulting deregulated kinase activity has profound effects on cell biology, including malignant transformation of some cell types and increased differentiation of others. At least one Src-like kinase gene is found in every metazoan, including sponges and cnidarians, but is absent, like other tyrosine kinases, from yeasts, protists, plants, and non-eukaryotes. Src kinases are thus one of the ancestral tyrosine kinases and have basic, fundamental roles in cell biology.

History of Src

The history of Src is a story of firsts, and research on Src has spawned wider areas of research. Src was the first protein tyrosine kinase activity to be identified, as a consequence of its role in malignant transformation. In the early twentieth century, P. Rous demonstrated that a chicken sarcoma could be transmitted from bird to bird by a particle smaller than a bacterium, and the field of tumor virus research was born. Rous' virus (RSV) was later shown to contain separable genetic elements for replication and for cell transformation. RSV also was one of the first viruses shown, by D. Baltimore and

H. Temin, to contain RNA-dependent DNA polymerase (reverse transcriptase) activity, consistent with the idea that transformation was caused by a DNA copy of a viral gene for sarcoma (*src*). D. Stehelin, H. Varmus, and J. Bishop, using nucleic acid probes, showed that a gene related to the viral *src* gene was present in uninfected vertebrate cells and was conserved over evolution. This provided evidence for the cellular origin of viral oncogenes. The viral *src* gene is now generally called *v-src*, while the cellular proto-oncogene is called *c-src*, or just *src*. The role of cellular proto-oncogenes in viral and nonviral cancers of animals and humans is now widely accepted.

In 1977, J. Brugge and R. Erikson found that certain antisera from RSV-infected newborn rabbits precipitated a 60 000 Da protein in addition to the virus structural proteins. The gene encoding this protein was shown to be *v-src*, and the protein was named pp60v-*src* or vSrc. It was soon found that this protein contained, or was associated with, a protein kinase activity that would transfer phosphate to antibody molecules or other model substrate proteins. While this kinase was first thought to phosphorylate threonine residues, T. Hunter and B. Sefton soon found that only tyrosine residues were phosphorylated. Indeed, the first tyrosine kinase to be identified was an activity, now known to be due to cellular Src family kinases, associated with the tumor-causing protein (middle T antigen) of polyoma virus (a DNA tumor virus). Both Src and vSrc are tyrosine kinases, but vSrc is 20+-fold more active than Src due to mutations that interfere with the negative autoregulatory interactions that restrain the activity of Src in normal cells. Different strains and derivatives of RSV have vSrc proteins with different numbers of such activating mutations, which have been very informative in understanding how Src is regulated.

The ability of vSrc to cause malignant transformation is thought to be due to unrestricted phosphorylation of substrate proteins that are normally phosphorylated, in a controlled fashion, by Src and other Src family kinases. In addition, the high kinase activity of vSrc may lead to the phosphorylation of other cell proteins that are not normally phosphorylated by Src, but it seems unlikely that these additional substrates are involved in transformation. Similarly, transformation by polyoma virus involves the deregulation of Src, by physical

association between middle T antigen and Src and several other proteins. The consequent unregulated activity of Src is critical for polyoma virus transformation.

While activated vSrc transforms fibroblasts, it causes other effects in other cell types. In some cases, it can induce differentiation and, in others, increase cell migration or morphology changes. These dramatic effects of activated mutant Src or vSrc led to an expectation that endogenous Src would be vital for many fundamental cellular processes.

Functions of Src in Development

The unique functions of Src were revealed when P. Soriano knocked out the mouse *src* gene. Surprisingly, considering the profound effects of activated v-*src* on vertebrate cells, absence of *src* had remarkably mild effects on development *in utero*. However, some defects were detected after birth. Teeth failed to erupt through the gum, and the cranium grew to form a domed shape. Both of these phenotypes are due to increased bone density, attributable to altered osteoclasts, the macrophage-related blood cells that resorb and remodel bone. Detailed studies of *src* mutant osteoclasts indicate that they differentiate normally but have a reduced ability to form specialized adhesion contacts through which they stick to bone and initiate bone resorption. However, blood platelets and neurons, where Src is highly expressed, are not noticeably altered in *src* mutant mice.

Other Vertebrate Src Family Kinases: Redundant and Specific Functions

The mild effects of *src* deletion are partly due to redundancy with other *src* family genes. The *yes* and *fgr* genes were first identified, as with *src*, as oncogenes in various cancer-causing retroviruses. Upon sequencing, they were found to be related to *src*. *lck*, *hck*, *blk*, *lyn*, and *fyn* were cloned by homology. Of the Src family proteins, Src, Fyn, and Yes are widely expressed in many different cell types in the body, whereas Fgr, Lck, Hck, Blk, and Lyn are restricted to, and are more important in, hematopoietic cells. However, alternative splicing and alternative promoter usage (with different coding exons) create additional diversity. For example, one splice form of Fyn is restricted to hematopoietic cells, while another is more ubiquitous. All the Src family proteins contain canonical features of Src that are critical for regulation and localization: the SH3, SH2, and kinase domains, amino-terminal lipid modification sites, and carboxy terminal and activation loop tyrosine phosphorylation sites. Other subfamilies of nonreceptor protein tyrosine kinases include the Csk, Abl, Fes, Tec/Btk, JAK, Syk, and FAK families, which lack one or more features of Src or contain other distinctive domains. The common features of Src family members allow considerable overlap in function, as shown by targeted knockouts and characterization of multiple mutants.

Individual mutation of *fyn* and *yes*, like *src*, causes only mild effects. For example, *fyn* knockouts have several alterations in the brain, including misorientation of pyramidal neurons in cortex and hippocampus and reduced myelination.

In addition, maturation of *fyn* mutant T lymphocytes is impaired. However, more phenotypes are manifested when *src*, *fyn*, and *yes* are mutated in combination. Double mutants have reduced perinatal survival, and triple knockouts survive poorly beyond mid-gestation (embryonic day 9 in the mouse). These triple mutant embryos resemble embryos mutant for the extracellular matrix protein fibronectin or some integrins, suggesting redundant roles for these three Src relatives in adhesive interactions between cells and the extracellular matrix.

The hematopoietic Src family kinases also have redundant and nonredundant functions. Consistent with their restricted expression, none of these kinases is needed for normal development: mice lacking *hck*, *fgr*, and *fyn* are moderately healthy and fertile, as are those lacking *blk*, *fyn*, and *lyn*. However, some hematopoietic defects are detected. Mutation of *lck* alone blocks thymocyte development at a stage prior to a proliferative burst, with a resulting large decrease in T-cell numbers. Macrophages from *hck* mutant mice have impaired phagocytosis. Combined absence of *fyn* and *lyn* leads to autoimmune disease, absence of *blk*, *fyn*, and *lyn* inhibits late-stage B-cell development, and absence of *hck* and *fgr* leads to decreased resistance to bacterial infection. Blood platelets lacking *src*, *hck*, *fgr*, and *lyn* fail to spread on fibrinogen. Thus, important roles for Src kinases in signaling from immunoreceptors and integrins become evident when multiple Src kinases are absent.

The preceding examples indicate that there is considerable redundancy in the Src family (i.e., one kinase can take the place of another if one is missing). However, whether different family members have distinct functions in a given cell remains unknown, because one kinase may normally be dedicated to a specific function but may perform others if another family member is missing.

Cellular Functions of Src Family Kinases: Signaling Adhesion and Immune Responses

The Src family kinases present in hematopoietic cells are activated by a number of stimuli, including integrin ligands, cytokines, and, notably, stimuli that act via immunoreceptors, such as the T-cell antigen receptor on thymocytes and circulating T lymphocytes, B-cell antigen receptor on pre-B cells, and immunoglobulin (Fc) receptors on macrophages. Cell-based assays confirm essential (but redundant) roles for Src kinases in immunoreceptor responses. Current models envisage weak association of Src kinases with the cytoplasmic tails of subunits associated with the immunoreceptors. Clustering of the receptors, possibly associated with altered proximity to protein tyrosine phosphatases (PTPs) and movement into or out of lipid rafts, activates the Src kinase, allowing phosphorylation of tyrosine-containing activation motifs in the immunoreceptors and recruitment of other SH2-domain-containing nonreceptor tyrosine kinases of the Syk and Tec/Btk families. In this situation, the key Src substrate is the immunoreceptor, but Src also contributes to phosphorylation of downstream signaling molecules involved in calcium mobilization and gene expression.

Src, Fyn, and Yes are also activated by adhesive stimuli and by soluble ligands that act via transmembrane receptors of the tyrosine kinase, G-protein-coupled, cytokine, and other

classes. The substrates that are phosphorylated vary according to the type of stimulus and the other proteins recruited by the respective receptor. For example, integrin stimulation causes Src activation and phosphorylation of another tyrosine kinase, FAK. The exact relationship between FAK activation and Src activation is not clear, but probably FAK autophosphorylates and activates Src, which further phosphorylates FAK. Both proteins contribute to phosphorylation of other proteins associated with the integrin. Functionally, absence of Src kinases alters cytoskeletal dynamics, for example, reducing turnover of focal adhesion complexes and slowing cell migration. However, Src activation by integrins also contributes to tyrosine phosphorylation of signaling molecules responsible for stimulating mitogen-activated protein kinase cascades and regulating gene expression. Src kinase activity is also required for cell proliferation induced by soluble growth factors acting via tyrosine kinase receptors. Dissecting the effects of Src activation by the growth factor receptor from the effects of Src activation by integrins is complex, since many cell-cycle events elicited by growth factors depend on cell adhesion via integrins. Src kinase activity can also regulate protein traffic to and from membranes. In relaying such diverse signals, Src family kinases probably have many important substrates, more than one of which may be necessary for an observed biological response. Src kinases thus represent branch points for activation of multiple signaling pathways.

Regulation of Src Kinase Activity

The combined results of mutational and crystallographic analysis have led to a picture of Src as a machine that integrates many different inputs to regulate its conformation and kinase activity (see Figure 1). There are at least two conformational states, simply thought of as on and off. In the off state, the kinase domain is relatively inactive, and the SH2 and SH3 domains are engaged in low-affinity intramolecular interactions. In the on state, the intramolecular interactions are absent, the kinase domain is at least partially active, and the SH2 and SH3 domains are completely accessible for binding other proteins. Thus, in its on state, Src can also potentially act as a scaffold, independent of its kinase activity. For example, Src may bind one protein through its SH3 domain and another through its SH2 domain, bringing the two proteins into a complex that conveys a signal to the cell. However, the extent to which Src kinases have kinase-independent functions *in vivo* when expressed at normal levels is not yet clear.

Src in the on state is only partially active as a kinase, and the kinase domain becomes fully active only when the activation loop is phosphorylated, a reaction that may be intramolecular but that can also be catalyzed by nearby tyrosine kinases, including other Src molecules. Clustering of partially activated molecules may thus increase activation-loop phosphorylation and lead to full activity. Such clustering may be important in regulation of Src family kinases by integrins and immunoreceptors. Conversely, dephosphorylation of the activation loop would reduce Src activity in the cell.

The on and off states are not stable structures, however. Molecular motions at physiological temperature cause these states to be somewhat flexible, allowing the equilibrium to be

disturbed by interactions with ligands for the SH2 and SH3 domains. As the individual intramolecular interactions stabilizing the off state are each rather weak, reduction of any one interaction leads to a concerted transition to the on state. Thus, increasing the local concentration of either an SH2 or an SH3 ligand would be expected to shift the equilibrium toward the on state. This is thought to occur in some situations in which viral proteins (such as polyoma virus middle T antigen and human immunodeficiency virus Nef) are highly expressed. Particularly potent activators would have binding sites for both the SH2 and SH3 domains. As phosphorylation of substrates by the Src kinase can create binding sites for its SH2 domain, Src can be activated by the products of its own kinase activity. Clustering of Src with substrates containing SH3 domain-binding sites may thus activate Src under certain conditions. Removal of such ligands would allow Src to revert to the off state.

In addition to control by clustering and by phosphorylation of the activation loop, another control is provided by phosphorylation of a tyrosine residue in the carboxy terminus of Src. When phosphorylated, this tyrosine serves as a ligand for the Src SH2 domain, stabilizing the off state. Dephosphorylation of this site allows opening of the off-state structure and hence activation. The carboxy-terminal region plays no apparent role when Src is active, and many mutationally activated forms of Src family kinases, found in various oncogenic retroviruses, have deleted or mutated this region. In the cell, a major source of Src regulation is provided by phosphatases that dephosphorylate the carboxy terminus and kinases that rephosphorylate this site. Genetic evidence suggests that the PTP CD45 is key to activating Lck in antigen-stimulated T cells, and that the PTPs RPTP α and Shp2 are involved in regulating Src in fibroblasts. However, to what extent these PTPs are regulated catalytically, and whether they provide a permissive environment in which Src family kinases can be activated by other means, is not clear.

Rephosphorylation of the carboxy terminus is catalyzed by a dedicated protein kinase, Csk, first identified by M. Okada and H. Nakagawa in brain extracts. Csk has few, if any, substrates other than the carboxy termini of Src family kinases. Genetic disruption of *csk* causes hyperactivation of Src family kinases, with consequent developmental abnormalities and death at mid-gestation stage. In *csk* mutant fibroblasts, Src family kinases are activated and the cells are transformed. A close Csk relative, Chk, may also be involved. In *Caenorhabditis elegans* and *Drosophila melanogaster*, Csk relatives can be recognized by sequence, and they are likely to function as Src inhibitors. Csk is structurally similar to Src family kinases but lacks the activation loop and carboxy-terminal tyrosine phosphorylation sites and an amino-terminal lipid modification site. It is thus not regulated like Src family kinases; its location inside cells, however, is regulated. Recent evidence indicates that Csk can bind to, and is activated by, one or more Src substrates (one named Cbp or PAG) that co-localize with activated Src family kinases in membrane subdomains. When these substrate molecules are phosphorylated, Csk is recruited by its SH2 domain, and the Csk is then brought close to the activated Src family kinases. Csk may also be associated, through its SH3 domain, with PTPs that target the activation loop tyrosine of Src. This mechanism would ensure that the Src family kinases

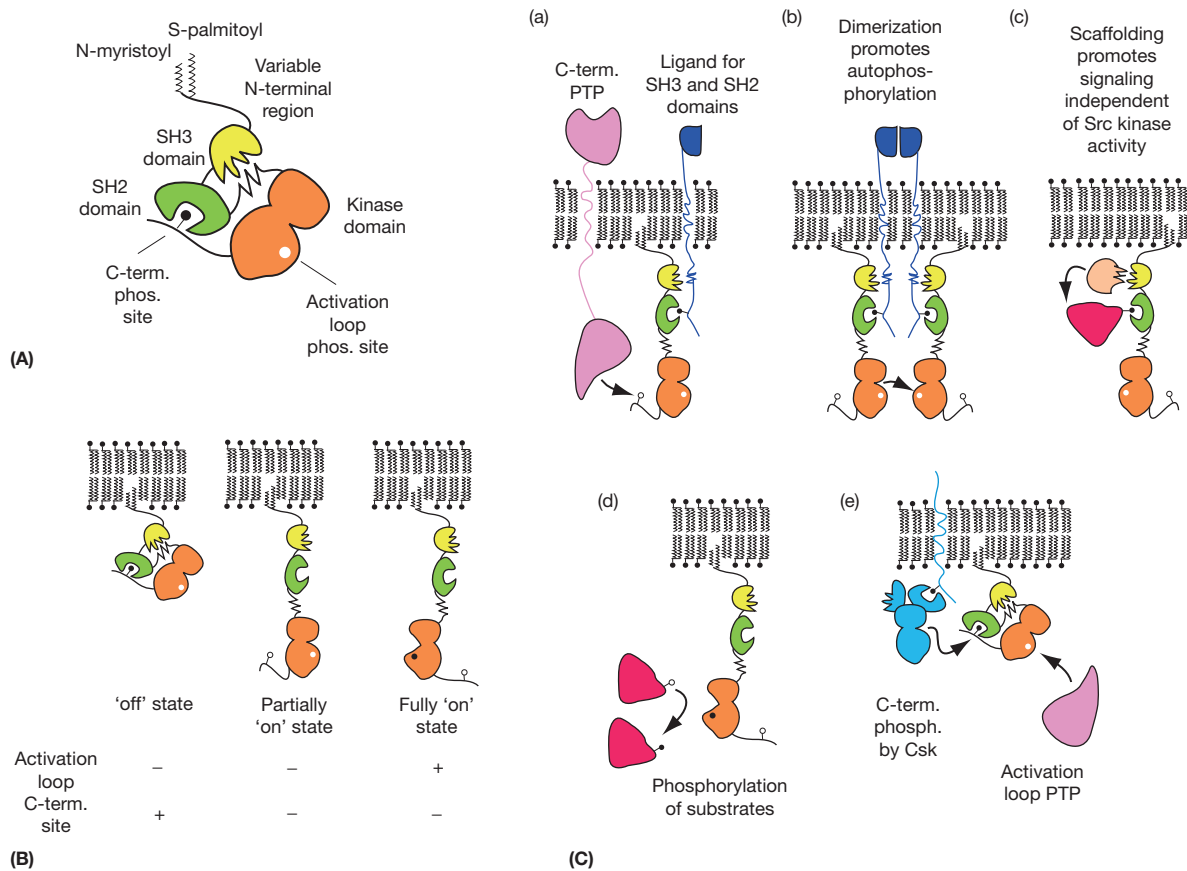


Figure 1 (A) Features of Src family kinases. The N terminus is modified by N-myristoylation and (for most but not all family members) by S-palmitoylation. Conserved structural domains are named SH3, SH2, and kinase (or SH1) domains, proceeding from N terminus to C terminus. In the off-state conformation shown, the SH3 domain is engaged with a linker between the SH2 and kinase domains, and the SH2 domain is engaged with a phosphorylated tyrosine (black circle) near the C terminus. A tyrosine in the activation loop is not phosphorylated (white circle). (B) Different activity states of Src kinases. In both the partially and fully on states, the C-terminal phosphorylation site is dephosphorylated, and the molecule is opened into a flexible conformation with SH3, SH2, and kinase domains available for interactions with other proteins. In the fully on state, the activation-loop phosphorylation site is phosphorylated, the kinase domain undergoes additional conformational changes, and substrates can be phosphorylated with high efficiency. (C) Regulatory interactions of Src kinases with other molecules in the cell. (a) Src can be partially activated by the combined or independent actions of PTPs that dephosphorylate the C-terminal site and protein ligands that can associate with the SH3 and/or SH2 domains. (b) Full activation may occur when partially activated molecules are brought together by clustering of a receptor or by decreased access to a PTP that dephosphorylates the activation loop (not shown). (c) Fully or partially active Src may also act as a scaffold, bringing proteins together but not phosphorylating them. (d) Active Src phosphorylates nearby cell proteins. Partially active Src may also phosphorylate substrates, but less efficiently. Substrates may be associated with the Src SH2 and/or SH3 domains. (e) Src is inactivated by the combined effects of Csk, which phosphorylates the C terminus, PTPs that dephosphorylate the activation loop, and inhibited access to SH2 and SH3 ligands (not shown). Csk may be recruited to Src by interactions with an anchor protein, such as Cbp/PAG. The PTP may also be specifically recruited (not shown). Interactions with regulatory molecules may be affected by movement of Src or the regulatory molecules between different microdomains in the lipid bilayer.

are then turned back off by Csk-catalyzed phosphorylation of their carboxy termini and PTP-catalyzed dephosphorylation of the activation loop.

The activation state of an individual Src kinase molecule in the cell is, thus, the result of a balance between protein–protein interactions and phosphorylation and dephosphorylation of the activation loop and carboxy-terminal tyrosines. However, the activity state (on or off) also affects access by the modifying enzymes. Therefore, there is a complex balance of interactions that allows Src to integrate different inputs. In addition, Src molecules are subject to other posttranslational modifications that regulate activity or localization. In mitosis, Src becomes phosphorylated by Cdk1/Cdc2–cyclin complexes. These serine phosphorylations partially activate Src by reducing the

intramolecular interactions that stabilize the off state. In cells responding to agonists that increase cyclic AMP levels, the cAMP-dependent protein kinase phosphorylates Src and activates it.

Subcellular localization may also be regulated. All Src family kinases are associated with the plasma membrane and intracellular membranes by amino-terminal sequences. The amino-terminal methionine is removed and the next residue (glycine) is subject to N-myristoylation. This modification is cotranslational and not regulated, but the majority of Src family kinases is also subject to reversible, regulated lipid modification by palmitoylation. This modification affects one or more cysteine residues near the amino terminus. Palmitoylation ensures that most Src family kinases (with the exceptions

of Src and Blk) are concentrated in microdomains within the plasma membrane that are enriched in phosphatidylinositol-4,5-bisphosphate, cholesterol, and glycosphingolipids, so-called lipid rafts. These microdomains are also enriched in many signaling molecules, placing the Src kinases close to potentially important substrates and effectors. Cbp/PAG, the aforementioned Csk-recruiting molecule, is also in the lipid rafts. Receptors bearing carboxy-terminal glycosylphosphatidylinositol anchors are localized in the outer leaflet of lipid rafts, and clustering these molecules activates Src family members associated with the inner leaflet. It seems probable that slight changes in protein distribution within lipid rafts will be very important for regulating Src kinase activity and directing phosphorylation activity to the appropriate substrates.

Nonvertebrate Src Kinases and Their Functions

Genetic studies in the fruit fly *D. melanogaster* and nematode *C. elegans*, where there are fewer src genes (two in *D. melanogaster* and one in *C. elegans*), have not revealed distinctive mutant phenotypes. Thus, unlike several other signal transduction pathways, genetic analysis in these organisms has not led the way in understanding how Src is regulated and functions. It is possible that Src proteins in these organisms (as in vertebrates) serve as integrators of many different inputs, and mutations lead to pleiotropic phenotypes. However, one identified function of Src64B in *D. melanogaster* is to regulate cytoskeletal structures in the cells that nurse the developing oocyte, and it does so in concert with a Btk/Tec-related tyrosine kinase. In this regard, it resembles the role of Lck in activating Btk/Tec-related tyrosine kinases in mammalian lymphocytes. Src42A also regulates the actin cytoskeleton in the developing embryo. In *C. elegans*, Src is important in the early embryo for correct cell–cell interactions, which may implicate Src in cytoskeletal organization. Further study in these systems may provide insights into Src regulation and functions in vertebrates.

Summary: Integration of Many Inputs and Regulation of Many Outputs

The many ways to regulate Src family kinases, and the many substrates implicated in different downstream events, position Src family kinases as servants with many masters. In the cell, Src kinases are regulated by the balance of kinases and phosphatases acting at both inhibitory and activating phosphorylation sites, and by proteins that bind to their SH2 and SH3 domains. The proteins that are phosphorylated depend on the subcellular localization and the protein complexes in which the active Src is found, and relay signals to regulate cytoskeletal, membrane, and nuclear events.

See also: **Lipids** Carbohydrates Membranes and Membrane Proteins: Glycosylphosphatidylinositol Anchors; Lipid Rafts; **Molecular Biology:** DNA Polymerases: Reverse Transcriptase Integrase, and Retrovirus Replication; **Protein/Enzyme Structure Function and Degradation:** Protein N-Myristoylation; Protein Palmitoylation; **Signaling:** B-Cell Antigen Receptor; Immunoglobulin (Fc) Receptors; Integrin Signaling; Mitogen-Activated Protein Kinase Family; T-Cell Antigen Receptor.

Further Reading

- Abram CL and Courtneidge SA (2000) Src family tyrosine kinases and growth factor signaling. *Experimental Cell Research* 254: 1–13.
- Blume-Jensen P and Hunter T (2001) Oncogenic kinase signaling. *Nature* 411: 355–365.
- Lowell CA and Soriano P (1996) Knockouts of Src-family kinases: Stiff bones, wimpy T cells, and bad memories. *Genes and Development* 10: 1845–1857.
- Martin GS (2001) The hunting of the Src. *Nature Reviews Molecular Cell Biology* 2: 467–475.
- Sicheri F and Kuriyan J (1997) Structures of Src-family kinases. *Current Opinion in Structural Biology* 7: 777–785.
- Thomas SM and Brugge JS (1997) Cellular functions regulated by Src family kinases. *Annual Review of Cell Development Biology* 13: 513–609.
- Weiss A and Littman DR (1994) Signal transduction by lymphocyte antigen receptors. *Cell* 76: 254–263.

Starvation

O E Owen[†], Southern Illinois University School of Medicine, Springfield, IL, USA

R W Hanson, Case Western Reserve University School of Medicine, Cleveland, OH, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Fuel homeostasis The steady-state maintenance of fuels.

Gluconeogenesis The synthesis of glucose from nonhexose precursors.

Glycogenolysis The breakdown of glycogen.

Ketogenesis The synthesis of acetoacetate and β -hydroxybutyrate and the generation of acetone.

Starvation The deprivation of any or all of the elements for nutrition.

There is a paradoxical relationship between morbid obesity and starvation because most of the reproducible data regarding fasting metabolism were derived from obese humans undergoing prolonged periods of fasting to reduce their body weight. During starvation, the concentration of blood glucose falls and this decrease is paralleled by changes in the concentration of insulin in the serum. Proteolysis, lipolysis, gluconeogenesis, and ketogenesis ensue. During starvation there develops an orchestrated interplay among the organ systems to produce, select, and conserve fuels needed for body metabolism. Steady-state metabolic processes develop after 18 days of total starvation. Urinary excretion of nitrogen declines to 4–5 g d⁻¹, and ammonium becomes the primary nitrogenous excretory product. Blood substrates and hormones are near constant concentrations, and weight losses and energy requirements are comparable from day to day. Glucose oxidation by the brain is suppressed and the majority of energy is provided by the metabolism of β -hydroxybutyrate and acetoacetate derived from the blood. The liver supplies the large quantity of ketone bodies needed to support the metabolism of the brain, aided in part by a decreased rate of ketone body utilization by skeletal muscle. The kidney also limits the amount of ketone bodies that is excreted in the urine in order to maintain near-electrical neutrality with urinary ammonium. The kidney shares the essential but limited role of gluconeogenesis with the liver. Ketogenesis, gluconeogenesis, ammoniogenesis, ureagenesis, and ketonuria are tightly interrelated during prolonged starvation.

Starvation in humans is the deprivation of any or all of the elements necessary for appropriate nutrition. Several extensive review chapters and books have been written regarding the physiological, biochemical, and psychological observations and measurements made during starvation. However, there is little reliable scientific information relating to weight loss, energy requirements, changes in the concentrations of hormones and substrate in the blood, and to organ metabolism from healthy humans during or after prolonged periods of starvation. Nonetheless, it is possible to draw reasonable conclusions regarding the metabolic effects of starvation in humans.

Food deprivation contrasts with conditions in most industrialized nations today, where obesity is epidemic and threatening the health of an increasing proportion of the population. There is a somewhat paradoxical relationship between starvation and obesity in which overweight people often starve to lose weight.

The cells of the various organ systems require a continuous supply of energy to function. The biochemical mechanisms that maintain the constant availability of fuel to support metabolic functioning are part of a complex system of fuel homeostasis. All mammals store energy when food is available and mobilize the stored fuels during fasting.

This article focuses on the metabolic studies of last century that range from fasting overnight to prolonged starvation. It covers topics such as the loss of weight and energy requirements, the release of stored fuels, the conversion of free fatty acids (FFAs), glycerol, and amino acids into ketone bodies and glucose by the liver, the selective utilization of fuels by the brain and muscle, the conservation of fuels by the kidneys and the promotion of gluconeogenesis, and maintenance of acid–base balance.

Background

The resting metabolic rate (RMR) among humans varies widely. However, general principles of bioenergetics for adult humans, who often have body masses varying over a sixfold range, have been defined. First, the RMR increases as body weight increases; however, the energy requirement per unit weight decreases as body weight increases. Thus, as the weight of a human increases from 40 to 160 kg, the RMR per kilogram body weight decreases from 30 to 15 kcal kg⁻¹ per 24 h. Second, based on total body weight, the RMR of men of a given weight is greater than it is for women. This difference disappears when the RMR is based on fat-free mass (FFM), measured by densitometry. Nonetheless, there is a broad range of energy requirements for people of identical gender and body weight, which is independent of body fat content. It is thus understandable why the quantity and nature of fuels oxidized under near-standard conditions among tens of thousands of human subjects show discrepancies for specific body sizes. However, there is general agreement when the fundamental rules of bioenergetics are applied; this is

[†]Deceased.

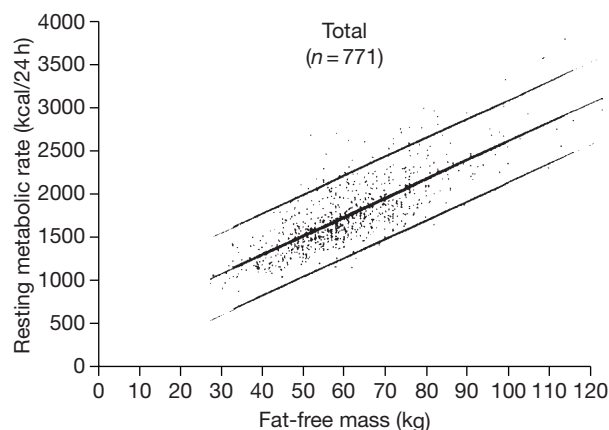


Figure 1 The relationship between fat-free mass, determined by hydrodensitometry, and the resting metabolic rate, as determined by indirect calorimetry, in 771 men and women.

illustrated in **Figure 1** which shows the relationship between FFM and RMR in 771 men and women. It should be easier to understand why these data are diverse when the foregoing principles are applied to the analyses of the studies in **Figure 1**. A major advance in our understanding of the mechanisms of survival occurred with the advent of convenient and accurate methods for measuring β -hydroxybutyrate (β -OHB⁻) and acetoacetate (AcAc⁻). These two water-soluble, four-carbon compounds are synthesized from acetyl coenzyme A (CoA) derived from the hepatic oxidation of long-chain fatty acids. Analytical techniques were developed in the laboratories of HA Krebs and co-workers in England. These fuels, along with acetone, are collectively referred to as ketone bodies and are immensely important for sustaining life during starvation.

Detailed clinical studies of the biology of starvation in humans began after 1965 when G.F. Cahill Jr. and colleagues in the United States and elsewhere advanced the concept of fuel homeostasis in humans.

Obesity and Starvation, and Clinical Features of Starvation

It is somewhat paradoxical to note that mechanisms for maintaining the flux of fuels to provide the energy requirements to sustain life during starvation were derived primarily from obese volunteers. These individuals were recruited from the general population or from the hospitalized patients. Most of the volunteers and patients were housed in closely supervised, general clinical research units in hospitals. Those who underwent prolonged periods of starvation received water and, usually, salt tablets and multivitamins. All were studied after an overnight fast or during 3–63 days of total starvation.

A word about the courageous volunteers who made these important studies possible is appropriate. Prolonged starvation of the length mentioned above (up to 63 days) is a difficult physical and emotional experience and the volunteers required an enormous amount of support to complete these studies. Some developed headaches and postural hypotension during the first few days of total food deprivation.

Everyone was preoccupied with hunger as starvation progressed. Understandably, some individuals became cantankerous and withdrew from the studies, and about one-half would manage to rarely sneak bits of food while under close observation, thus spoiling some of the experimental data. Still, most of the volunteers struggled valiantly to comply with the research protocols. These individuals are unsung heroes for their participation in these studies and are owed a great debt of gratitude.

The compensated state of adult obese humans starving for experimental reasons is in sharp contrast to children, adults, and the elderly dying from inanition during famine or confinement in prisons. In the latter individuals, malnutrition is often complicated by other diseases. Understandably, their behavior varies from hostile selfishness to apathy and stupor. Superimposed bacterial, fungal, verminous, helminthic, and protozoal infections, with diarrhea, cough, and protuberant or contracted abdominal walls, modify the clinical characteristics of starving people. The loss of subcutaneous fat and muscle wasting accentuate the size of joints, especially the knees. The calf muscles are among the first to visibly shrink due to limited mobility, and the buttocks show marked atrophy. Hair is broken and brittle and nails split and become pitted and spoon shaped. The skin is pale from anemia, but has regions of pressure hyperpigmentation, fistulae, and decubitus ulcers. Staring, sunken eyes search to identify anything, and the facial musculature is wasted and sometimes taut. Similar features are seen among patients hospitalized with severe malnutrition. Death occurs when about one-third to one-half of the lean body mass is lost in children, or lean and obese adults. In lean individuals, practically all the body fat is mobilized before the irreversible state of malnutrition occurs.

The data from controlled starvation studies may not mimic many of the findings observed in patients from the harsh environments described above. Nonetheless, the data gathered from research centers provide the best-available and reproducible data regarding starvation, metabolism, and fuel homeostasis. These data were collected largely over a period of about four decades. As would be expected, the interpretation of some of the biomedical data changed with the accumulation of more information. An example of this is the changing attitude toward how the kidneys handle ketone bodies. As will be discussed in detail later, ketone bodies have several metabolic roles and are essential for survival during prolonged starvation.

Metabolic Alteration Related to Starvation: Fed to Fasted to Fed

The Fed State

When food is presented to an average hungry human who has undergone a mere 12-h overnight fast, he/she can readily eat at a rate of 150 times the individual's RMR. Only minor perturbations occur in the concentrations of fuels in the blood but the postprandial concentration of insulin, the major anabolic hormone, can increase 50-fold in obese individuals. It may take several hours to digest, absorb, oxidize, and store the excessive caloric content of the food. Some of the nutrients that are consumed in overabundance are transiently stored as glycogen and intracellular amino acids or protein during the early postabsorptive period. However, the dominant

storage form of energy in the body is triglyceride in adipose tissue. A human can consume about as much as 10 000–12 000 kcal d⁻¹ for many days. If this excessive caloric intake continues over a long period of time, the adipose tissue mass can become huge, accounting for as much as 500 000–1 000 000 kcal of energy present as triglycerides. However, the maximum body fat content is usually limited to about one-half of the total body mass. The lean body mass also increases to support the fat mass. Thus, a 150-kg (330 lb) man or woman can have a triglyceride (fat) mass of 75 kg (165 lb). The stored fat, along with some of the body protein and glycogen, is mobilized during semistarvation or total fasting, usually when approximately 20–50 g of carbohydrate and 15–30 g of protein are eaten. However, during total, prolonged starvation among morbidly obese people, when amino acids are continuously mobilized from vital organs (e.g., the heart), death may occur before the entire fat mass of the body is depleted. Thus, grossly obese humans subjected to prolonged starvation can die with body fat remaining.

The Starved State

In going from the fed state to the starved state, the body shifts from storing to mobilizing fuels. The factors that control the production and utilization of specific fuels change rapidly during this transition. In contrast, starving volunteers enter a near steady state of metabolism after 2–3 weeks of starvation, where day-to-day changes are minimal. During starvation, catabolism is tightly regulated and the supply of fuels from various depots adjusts to meet the body's energy requirements. Hepatic glycogen breakdown is brief (only 600 kcal are available as hepatic glycogen in a 70-kg human), while proteolysis is continuous in order to supply amino acids required for gluconeogenesis, for energy, and to maintain acid–base balance. The triglyceride present in adipose tissue is the major and persistent fuel reserve that supports metabolic process during starvation; it can account for as much as 90–93% of the body's energy needs during periods of prolonged starvation.

Refeeding

During the period of semistarvation, a constant awareness of hunger dominates the thought processes of humans. During refeeding after periods of starvation or undernutrition, emaciated people may ingest as much food as possible, sometimes vomiting because of excessive intake. This craving for food persists until weight gain occurs largely from the deposition of fat in white adipose tissue. Eventually, weight gain plateaus and weight loss may occur until the body weight of an individual reaches an equilibrium that is at or near their weight before semistarvation. On the other hand, after total starvation, the obese volunteers who become outpatients experience mild abdominal pain and may pass fecaliths. During the first 2–3 weeks of refeeding, these individuals routinely retain fluid and become edematous and are encouraged to maintain a low caloric intake until diuresis develops. However, the success rate for preventing the reaccumulation of body fat is low. Some clinical investigators hypothesize that the body possesses a set point or has a body fat lipostat. This is questionable because it is just as likely that individuals eat more than they

need to replace body fat, and gain additional weight predominantly as adipose tissue.

Weight Loss, Body Composition, and Energy Requirements

The measurement of body weight is usually obtained accurately and easily. However, edema can complicate the interpretation of this simple measurement. Measuring body composition and energy requirements has several shortcomings. FFM or active tissue (lean body mass) is a conceptualized mass and difficult to measure accurately. This is because there are differences in the loss of mass among different organ systems in both the lean and obese individuals during starvation. The skeletal muscles and subcutaneous fat lose the greatest quantity of weight. Among organs, the liver and spleen lose at least 50% of their usual mass. The gastrointestinal tract becomes atrophic; the diameter of the lumen of the small intestine becomes reduced to 50% of its normal diameter, and the villae become flattened. Gut motility also decreases. The heart decreases in size and the pulse rate and blood pressure decrease. The loss of bone mass is less than the loss of adipose tissue and skeletal muscle during experimental periods of starvation.

The energy needs of the body are determined by the sum of the metabolic requirements of various tissues (brain, liver, muscle, heart, adipose tissue, spleen, leucocytes, etc.) that depend on the mass and activity of individual organs. During the resting state, there is a 50- to 100-fold variation per unit mass for energy requirements among different tissue types (e.g., brain and adipose tissue). The energy requirements of skeletal muscles can change 12-fold in the transition from the resting state to strenuous exercise.

The proposed metabolic efficiency during starvation has long been claimed in lean people, but it has been difficult to demonstrate in starving, obese humans. Studies by Owen and colleagues differ from those of Leibel and associates in this regard. There is no question that as body size decreases in lean and obese humans, their metabolic requirements also decrease. Nonetheless, metabolic adaptation to starvation, where the energy requirements per unit of measurement decrease out of proportion to changes in unit of measurement, remains a perplexing issue. This is partly due to the method by which the data are calculated or expressed. The non-RMR is more difficult to measure and is more inaccurate than estimates of the resting energy expenditure. Therefore, if there is an adaptive metabolic efficiency induced by food deprivation, it should be demonstrable by the (more) standardized RMR measurement. However, this has not been consistently demonstrated in all studies. Therefore, the increase in efficiency may be more apparent than real.

Thus, there remains an inadequately defined relationship between weight loss and metabolic rate. There are contradictory data as to whether there is a decreased RMR per unit mass in adult men produced by prolonged semistarvation. However, the normal RMR varies widely from low, normal, or high values for humans of identical weight, gender, and age. Therefore, it is not surprising that the literature on RMR for humans, based on the standard of measurement, is inconsistent. For example, differences among the various thyroid hormones,

T_4 , T_3 , and rT_3 , in lean and obese people subjected to protracted food deprivation may influence the RMR. It is well known that after about a week of total starvation, when the diuretic phase of starvation is completed, body weight loss was only 0.32% per day. Later in starvation the RMR of obese, starving volunteers, based on oxygen consumption per kilogram FFM per day, and corrected for urinary nitrogen loss during starvation, is constant. This was truly disheartening for obese people who desperately wanted to lose body fat. It also clearly demonstrates how limited total starvation is as a method for weight reduction for morbidly obese patients. Nonetheless, critical knowledge regarding the metabolic effects of starvation has been obtained from these noble volunteers.

The Nature and Quantity of Fuels Oxidized during Starvation

Indirect calorimetry closely reflects the nature and quantity of the fuels oxidized. The results obtained using this method are influenced by dietary intake, total metabolic requirements, and state of physical activity. When lean individuals fast overnight, their nonprotein respiratory quotient (npRQ) is 0.84. This matches the previous day's intake of 12–20% protein, 40–45% fat, and 40–45% carbohydrate. When obese people eat the same balanced diet, but with a greater caloric intake to match their greater metabolic needs, the npRQ is significantly lower and decreases as body weight increases. Figure 2 shows the trend in the nature and quantity of fuels oxidized after an overnight fast in lean and obese humans. The data show that as weight and body fat increase, the npRQ falls. The greater the RMR, the greater the quantity of fat that is mobilized to meet the energy requirements of fasting humans. Due to body energy demands, obese humans shift into an accelerated rate of fat mobilization. However, this tendency to mobilize and oxidize stored fat becomes more obvious in both lean and obese people as fasting is prolonged. As starvation is extended to 2–3

days and beyond, lipids furnish 90–93% of the resting energy requirements in both lean and obese humans. These changes in the nature and quantities of fuels oxidized by tissues, such as liver, muscle, and brain, act to spare glucose from oxidation; this provides the physiologic basis for insulin resistance when carbohydrate oxidation must be curtailed for survival during food deprivation.

The absolute minimal rates for the major fuel utilization before death occurs have not been defined. The minimum requirement for fat and protein oxidation in $\text{g d}^{-1} \text{kg}^{-1}$ FFM has been approximated as follows: $2.98 \pm 0.21 \text{ g kg}^{-1} \text{ d}^{-1}$ FFM for fat; $0.52 \pm 0.10 \text{ g kg}^{-1} \text{ d}^{-1}$ FFM for protein. Glucose is derived from glyceride-glycerol, amino acids, propionate, and recycled lactate and pyruvate. About $1.91 \pm 1.04 \text{ g kg}^{-1} \text{ d}^{-1}$ FFM are synthesized. However, most of the glucose that is synthesized during starvation is recycled from lactate and pyruvate (the Cori cycle) or made from glycerol, alanine, and glutamine. The quantity of glucose that undergoes oxidation to CO_2 and H_2O is primarily determined by the catabolism of glucose in the central nervous system (CNS) and to a lesser extent by the red blood cells. This quantity is not closely related to the FFM because the size of the brain is unrelated to the FFM. The terminal oxidation of glucose is $40\text{--}45 \text{ g d}^{-1}$, which is derived primarily from gluconeogenesis from amino acids (alanine and glutamine) and glycerol.

In obese humans, during starvation there is no decrease in the metabolic requirement per unit mass, based on measurements of oxygen consumption. However, when the npRQ decreases below 0.7, the theoretical minimum for fat oxidation, there is a difficulty in extrapolating the respiratory exchange rates of O_2 and CO_2 and urinary nitrogen excretion into energy expenditure. The production of CO_2 falls disproportionately to O_2 consumption, creating a mysterious situation characterized by a calculated npRQ of 0.62–0.65. This unusual finding cannot be eliminated by correcting for ketonuria. However, it has been postulated that some of the fatty acids released from triglyceride stores may undergo desaturation before being recycled to storage depots. The process of desaturation consumes oxygen and produces heat, but releases no carbon dioxide. Desaturated FFAs in the blood have been identified in starving animals. If this process occurs in humans, the respiratory quotient (RQ) should rise before death, when unsaturated fatty acids are released and oxidized for energy. The RQ does rise in animals and humans during the premortal period of starvation. This phenomenon has not been fully explained but is usually attributed to the mobilization and oxidation of amino acids. It is just as likely that it is due to oxidation of desaturated fatty acids.

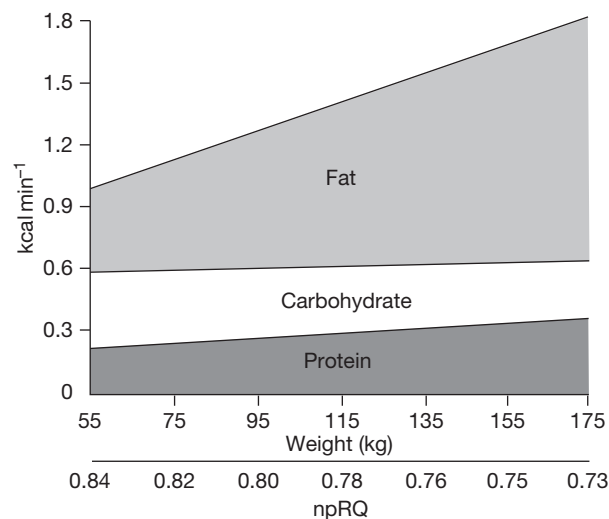


Figure 2 The nature and quantity of fuels oxidized by humans after a 12–14-h fast. Fat oxidation increases as body size increases.

Changes in the Concentration of Substrates and Hormones in the Blood

The concentration of glucose in the blood after an overnight fast is 4.5 mM; this value falls to 3.6 mM after 3 days of starvation. Thereafter, the concentration of glucose in the blood reaches a plateau. Anaerobic glycolysis in tissues such as muscle, brain, red blood cells, and kidney medulla converts glucose into lactate and pyruvate, and the liver extracts these compounds to produce glucose. The concentration of lactate in

the venous blood is less than 1.0 mM, while the concentration of pyruvate is less than 0.1 mM and is constant in the resting state. The concentration of FFA in the plasma rises from 0.6 to 1.4 mM during the first few days of starvation. Thereafter, the concentrations of these fuels remain elevated and relatively constant. The rate of uptake and disposal of FFA is largely determined by its concentration in the blood. Insulin regulates the release of FFA from adipose tissue, primarily by influencing lipolysis, and thus the availability of FFA. The concentration of glycerol in the blood derived from lipolysis is less than 0.1 mM after an overnight fast and rises to 0.15 mM on the third day of starvation. Thereafter, glycerol remains constant because uptake, primarily by the liver, to synthesize glucose matches its release by adipose tissue. The concentration of triglycerides in the blood is less than 1.0 mM and remains constant during fasting.

A characteristic of fasting in humans is the presence of increased concentrations of ketone bodies in the blood (Figure 3) and urine. These water-soluble, short-chained organic acids are synthesized primarily in the liver from the acetyl CoA derived from fatty acid oxidation and serve as alternate fuels for tissues such as the brain. Ketone bodies replace glucose as the dominant energy source for the brain as starvation progresses. Physical activity augments catabolism during starvation and promotes hyperketonemia. No other fuel in the blood changes as markedly as the concentration of ketone bodies during starvation. This is partly due to the low concentration of ketone bodies in the blood during the

postprandial period, after a mixed meal containing adequate carbohydrate. After an overnight fast, lean people have blood AcAc^- and $\beta\text{-OHB}^-$ concentrations of 0.05 mM, while this value is slightly higher in obese individuals. Acetone is virtually absent after an overnight fast, unless the individual is regularly consuming a high-fat diet. There is an exponential rise in the concentration of AcAc^- and $\beta\text{-OHB}^-$ in the blood during starvation, until a new steady level develops. During the first 3 days of fasting, the AcAc^- concentration in the blood increases to 1.5 mM, while $\beta\text{-OHB}^-$ continues to rise to 4.5 mM after day 10 of starvation. AcAc^- plus $\beta\text{-OHB}^-$ reaches a plateau at 6–8 mM after 18 days of total starvation (water, salt, and vitamins were provided). The smell of acetone halitosis becomes evident after 2–3 days of starvation when the blood concentration is 0.25 mM. Acetone slowly increases to 0.35 mM after 21 days of starvation. A minimum estimate of the change in the concentration of ketone bodies in the blood from an overnight fast to 18 or more days of total starvation can be as great as 60-fold. Blood AcAc^- and $\beta\text{-OHB}^-$ are anions and the hydrogen produced during ketogenesis combines with bicarbonate to form CO_2 and water. CO_2 is exhaled, thus decreasing the concentration of bicarbonate in the blood; this creates the typical anion gap of ketonemia. Blood pH also falls appropriately (7.4–7.3), and a mild metabolic acidosis of starvation becomes evident.

The concentration of total amino acids in the plasma declines slowly from an overnight fasting value of 4.6 to 3.7 mM after 40 days of starvation. However, there are four basic patterns of change among the amino acids during total starvation, represented by alanine, glycine, valine, and glutamine. Alanine rapidly decreases to 30% of its overnight concentration. Glycine does the opposite; it rises nearly threefold, before reaching a plateau. The concentration of valine doubles in the plasma during the first 7–10 days of starvation, and then slowly and progressively falls to a value below its overnight fasting value. Venous glutamine, the predominant plasma amino acid, remains relatively stable during 5–6 weeks of fasting. Blood urea nitrogen mimics the concentration of alanine in the blood. The concentration of creatinine in the blood slowly drifts downward, reflecting the decrease in muscle mass.

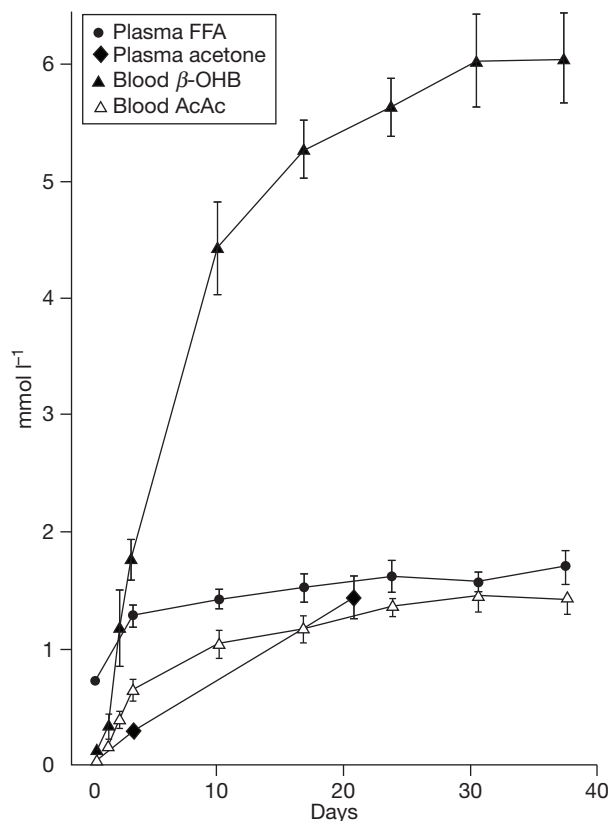


Figure 3 Changes in the concentration of ketone bodies and free fatty acids (FFAs) during starvation.

Hormonal Changes

Insulin is the main hormone that controls the anabolic processes that maintain fuel homeostasis in humans. Its secretion is primarily regulated by the blood glucose concentration, but the levels of amino acid, FFA, and ketone bodies also modulate insulin release by the β -cells of the pancreas. The influence of insulin on the metabolism of the major organs is readily demonstrated in adipose tissue, skeletal muscle, liver, white blood cells, and other tissues. Insulin also inhibits the secretion of glucagon, a significant counterbalance catabolic hormone.

The concentration of insulin in the blood parallels the changes in the levels of glucose. The fall in the concentration of insulin diminishes some of its inhibitory effects on peripheral proteolysis and lipolysis. In the starved state, the concentration of insulin falls, resulting in a decrease in the rate of uptake of glucose, amino acids, and fatty acids in peripheral

tissues, and a subsequent increase in the rates of gluconeogenesis, lipolysis, and proteolysis. After an overnight fast, the concentration of insulin in the portal blood is 2–3 times greater than the peripheral venous concentration. After 2–3 days of starvation, the portal–peripheral insulin concentration gradient is small; however, there is still a high-enough concentration of insulin in the blood to limit the maximal rates of glycogenolysis, gluconeogenesis, and ketogenesis. In addition, insulin has at least two indirect effects on hepatic glucose and ketone body production. Insulin decreases the delivery of gluconeogenic precursors and FFA from the extrahepatic tissue stores to the liver. As the concentration of insulin in the mesenteric blood decreases, the inhibitory effect of insulin on the β -cells of the pancreas curtails the secretion of glucagon. Thus, the low concentration of insulin in the blood promotes glucagon secretion.

Insulin has dual roles in controlling the release of glucose and ketone bodies from the liver (and kidney cortex), amino acid from muscle, and FFA from adipose tissue. Glucagon has the opposite effect. It augments hepatic (and renal) gluconeogenesis and ketogenesis, and promotes peripheral lipolysis and proteolysis. A relatively low blood insulin concentration and relatively high glucagon concentration creates a blood insulin/glucagon ratio that promotes fuel availability. These pancreatic hormones serve collectively to finely balance the fuel needs of various tissues. The catabolic role of glucagon is augmented by catecholamines, glucocorticoids, and growth hormones. As fuel homeostasis is critical for survival, there are multiple levels of hormonal control of this process.

Appetite Regulation

During starvation, there are a number of factors that regulate the control of appetite. Ghrelin, an orexigenic hormone, primarily secreted by the stomach and duodenum, normally increases in the blood before meals and falls after meals. Ghrelin is thought to stimulate hunger, increase body weight, and decrease metabolic rate. It rises with semistarvation but its physiologic roles regarding hunger need further elucidation. Adipose tissue produces a satiety hormone, leptin, which suppresses appetite. The leptin concentration in the blood falls in parallel with insulin during the first 3 days of starvation. As body fat decreases during starvation, the blood leptin concentration decreases. A fall in the levels of leptin in the blood increases the hypothalamic concentration of neuropeptide Y, greatly stimulating appetite. However, the control of appetite is not a simple process, but rather involves the interaction of a number of neuroendocrine systems. There is no single hormonal neuropeptide that controls hunger in a starving human. Prolonged starvation causes exaggerated hunger, a psychosomatic state that dominates the conscious mind.

Insulin plays an important role in the regulation of appetite. Acting together with leptin, insulin circulates at concentrations proportional to the body fat content. Both hormones enter the CNS and bind to specific receptors in neurons involved in controlling food intake. The administration of leptin and insulin directly to the hypothalamus decreases appetite and suppresses food intake. It is probable that leptin is the more important of the two hormones since leptin deficiency, not a lack of insulin, results in obesity. The mechanism

by which these hormones influence the CNS is the subject of intensive study. A detailed discussion of this area is beyond the scope of this article. However, it should be clear that during starvation, as the mass of adipose tissue decreases and the secretion of insulin by the pancreas is depressed due to a lack of food intake, the appetite centers of the brain would be stimulated by the lack of satiety signals such as leptin, insulin, and ghrelin.

Adipose tissue secretes a number of other regulatory peptides, besides leptin, that control fuel deposition. These include adiponectin, resistin, adipisin, acetylation-stimulating protein, angiotensinogen, and cytokines. The most abundant adipose-specific hormone is adiponectin. The concentration of adiponectin in the blood is positively correlated with insulin sensitivity. Adiponectin promotes glucose uptake in muscle and fatty tissue. Its concentration in plasma is inversely related to adipose tissue mass, especially abdominal fat mass. The levels of adiponectin in the blood fall during semi-starvation and total starvation, when glucose intolerance has been documented. Overexpression of the gene for adiponectin in the mouse results in profound obesity in the animal, underlining the importance of this hormone in the control of lipid deposition. Resistin is another protein secreted into the blood by adipose tissue, but it has an effect that is opposite to that of insulin. Its concentration in the blood of humans after prolonged starvation has not been defined. The behavior of adipose tissue cytokines in humans during weight reduction caused by semistarvation is under investigation.

Biochemical Changes during Starvation

Starvation induces a number of specific biochemical changes that permit an individual to survive in the absence of food for a prolonged period of time. These include alterations in glucose synthesis, fatty acid utilization by specific tissues, and the preservation of nitrogen to insure the maintenance of lean body mass. Starvation is a catabolic state in which fuel reserves are mobilized to support metabolic needs. To understand the metabolic adaptations to starvation, it is first necessary to describe the available energy reserves of a human as starvation begins. It has been estimated that a 70-kg human contains 144 000 kcal as fat (largely as triglyceride), 24 000 kcal as protein (lean body mass), and only 900 kcal as carbohydrate (liver and muscle glycogen). The relatively low level of carbohydrate reserve in humans necessitates a number of biochemical adaptations to insure the appropriate energy supply to specific tissues. For example, the brain uses 600 kcal of glucose each day; this means that the total carbohydrate reserve is depleted in about 1 day of starvation! It is clear that mechanisms must exist for the use of fuels other than glucose (fatty acids and ketone bodies) by tissues such as the muscle and the brain, to insure survival during starvation. This process is called fuel sparing.

Control of Fatty Acid Release during Starvation

Since fat is the major caloric reserve in humans, the regulation of FFA mobilization from adipose tissue during starvation is a critical factor for survival. The breakdown of triglyceride via lipolysis is under the control of a number of hormones,

including epinephrine, glucagon, glucocorticoids, and insulin. The concentration of insulin in the blood falls by 50% after 3 days of starvation. At the same time, there is an increase in the concentration of glucagon. This change in the insulin-to-glucagon ratio is a critical factor in the increase in lipolysis and the enhanced rate of hepatic and renal gluconeogenesis that occurs during starvation. Insulin is the major antilipolytic hormone and its decrease is the major factor in insuring the increased availability of the fatty acids needed for energy metabolism during starvation.

Adipose tissue contains a hormone-sensitive lipase that is sensitive to the state of its phosphorylation. An increase in the concentration of cyclic adenosine monophosphate (cAMP) in the tissue (caused by a rise in epinephrine, norepinephrine, and growth hormone, and a drop in the level of insulin) activates protein kinase A (PKA) that in turn phosphorylates, and thus activates the hormone-sensitive lipase; this results in the breakdown of triglyceride and generation of FFA. Insulin counters this process, partly by decreasing the levels of cAMP in the adipose tissue, and by increasing the activity of a phosphoprotein phosphatase that dephosphorylates the hormone-sensitive lipase, thereby inactivating it. There is also evidence that insulin increases the activity of a phosphodiesterase in the adipocyte, causing a fall in the concentration of cAMP in the tissue. The net result of prolonged starvation is an enhanced rate of FFA release from adipose tissue that is used as a fuel by a number of tissues.

About 50% of the plasma FFA mobilized to the liver during starvation is extracted and 50–80% of the extracted fatty acids undergo partial oxidation to synthesize equal quantities of AcAc^- and $\beta\text{-OHB}^-$; most of the remaining quantity of FFA is converted to triglycerides and recycled to adipose tissue as very low density lipoprotein (VLDL). It is interesting that such a large fraction of the FFA released by lipolysis is re-esterified back to triglyceride in adipose tissue or in the liver and other tissues, and returned to the adipose tissue for the resynthesis of triglyceride. Fatty acid recycling via this so-called triglyceride–fatty acid cycle can account for as much as 60% of the fatty acid released after 3–4 days of starvation in humans. The synthesis of triglyceride requires the generation of 3-phosphoglycerol, usually from glucose. During starvation, when glucose is at a premium, the 3-phosphoglycerol is generated from pyruvate, lactate, and amino acids via an abbreviated version of gluconeogenesis termed ‘glyceroneogenesis’. The triglyceride/fatty acid cycle is a futile cycle that uses energy for the synthesis of triglyceride (six molecules of adenosine triphosphate (ATP) per molecule of triglyceride synthesized), so it must have a role in preserving the FFA that was released by adipose tissue to be used later as a fuel by peripheral tissues.

Proteolysis and Amino Acid Metabolism

After a meal containing proteins, amino acids largely escape hepatic extraction and are carried by the blood to extra-hepatic tissues. Insulin promotes the active transport of amino acids into cells, primarily skeletal muscle. Normally, amino acids are in a state of flux; they are precursors for protein synthesis and then appear as free amino acids after protein breakdown. Within the first day of starvation, however, protein catabolism dominates

the metabolic flux. As starvation progresses, relatively more proteolysis occurs and amino acids are mobilized from protein depots. The dominant fate of the carbon skeletons and the amino and amide groups, derived from the breakdown of amino acids in muscle, is conversion to alanine and glutamine that is mobilized from muscle and other depots and transported to liver and kidney to be utilized to synthesize glucose, urea, and ammonia.

Protein catabolism in the splanchnic tissues is somewhat less responsive to insulin than it is in skeletal muscles. In the periphery, the basal concentration of insulin that is present during starvation limits proteolysis. Nonetheless, glucose must be continually synthesized from amino acids by the liver and kidney cortex during starvation; amino acids from muscle protein are a major source of carbon for gluconeogenesis. The first proteins degraded during starvation are the digestive enzymes secreted from the stomach, pancreas, and small intestine. The liver also loses various enzymes needed to process incoming nutrients into plasma protein (e.g., albumin). Thereafter, the largest protein mass of the body, striated muscle, begins to be drained, not only of intracellular proteins, such as enzymes, but also of contractile elements. The disintegrating muscles can easily be seen as skeletal muscle atrophy during prolonged starvation in humans.

The control of proteolysis in muscle is a complex process. Insulin retards proteolysis and enhances protein synthesis. In addition, metabolic acidosis promotes protein breakdown to insure the generation of ammonia to titrate the acidity of the tubular urine. The low plasma insulin concentration and mild metabolic acidosis of starvation are conducive to proteolysis. The best-characterized pathway for protein catabolism is an ATP-independent system of acid proteases (cathepsin) and hydrolases contained in cellular lysosomes. In addition, there are calcium-dependent proteases, as well as cytosolic ATP-dependent and -independent pathways. The most important muscle proteolytic system employs the ubiquitin proteasome pathways.

Amino acids generated from proteolysis undergo deamination and/or deamidation before being further transformed to alanine or glutamine, which efflux from skeletal muscle. Alanine plus glutamine account for 80% of the amino acids released from muscle after prolonged starvation, and are the principal vehicles for transporting nitrogen from peripheral tissues to liver and kidney. These two amino acids together account for only 10% of the composition of skeletal muscle protein, but they provide the majority of the carbon for the synthesis of glucose by the liver and kidney during starvation. Thus, there is a net conversion of other amino acids into alanine and glutamine in the muscle. Alanine is the major nitrogenous compound released from muscle and extracted by the liver for gluconeogenesis and ureagenesis. There is a glucose–alanine cycle between muscle and liver. Glutamine is the other amino acid released in high levels by muscle, but it is extracted from the blood by the small intestine, where it is converted to alanine for subsequent use by the liver, or by the kidney cortex, producing glucose and ammonia. During the catabolism of muscle protein, branched-chain amino acids provide the amino groups to α -ketoglutarate to form glutamate via the enzyme branched-chain aminotransferase; glutamine is derived from glutamate and ammonia by the action of glutamine synthase.

The tricarboxylic acid (TCA) cycle is the key metabolic pathway for energy production: acetyl CoA is oxidized to carbon dioxide in the cycle. During starvation, the TCA cycle has another fundamental role in metabolism. Amino acids are catabolized to 4- (aspartate, asparagine, phenylalanine, tyrosine, methionine, isoleucine, and valine) and 5- (glutamine, glutamate, histidine, proline, and arginine) carbon intermediates that enter the TCA cycle by a process termed anaplerosis. The TCA cycle cannot serve as a carbon sink; therefore, the carbon skeletons that enter the cycle must leave by a process called cataplerosis. Anaplerosis and cataplerosis are reciprocal reaction pathways that must be in balance. These reactions have been reviewed in detail elsewhere in this volume. It is important to note that during starvation more fuels are released from storage depots or synthesized in the liver and kidney than are oxidized to generate energy. About 60% of the fuels that enter the bloodstream are recycled, for example, FFA, glucose, and amino acids, especially alanine and glutamine.

Ketogenesis

The liver not only stores glucose as glycogen, but also converts fuels such as FFA, amino acids, lactate, pyruvate, and glycerol to glucose and ketone bodies and detoxifies the ammonia, which is produced from the deamination of amino acids, by converting it to urea. After an overnight fast, hepatic glycogenolysis, gluconeogenesis, and ketogenesis provide 50% of the total energy-yielding fuels for the body in the resting state. Lipolysis of triglyceride in adipose tissue supplies FFA and glycerol; these substrates become precursors for ketone body (fatty acids) and glucose (glycerol) syntheses. Amino acids released primarily by skeletal muscles complement glycerol as gluconeogenic precursors. As fasting is prolonged, the kidney cortex also contributes to fuel homeostasis by conserving substrates and sharing the gluconeogenic role with the liver.

Ketone bodies are the only fuels synthesized in the body that do not recycle. Unlike FFA, amino acids, and glucose, ketone bodies are either oxidized or excreted in the urine and/or the breath (acetone); only a negligible small amount of acetone can be converted to glucose. β -OHB⁻ and AcAc⁻ are synthesized in liver (and kidney) mitochondria by the following reactions. Two molecules of acetyl CoA condense head to tail to form acetoacetyl CoA; this reaction is catalyzed by acetoacetyl CoA thiolase. Another molecule of acetyl CoA is joined by β -hydroxy- β -methylglutaryl CoA (HMG CoA) synthase to form HMG CoA, also generating a hydronium ion (H⁺). HMG CoA lyase cleaves HMG CoA into AcAc⁻ and acetyl CoA. AcAc⁻ is then reduced to β -OHB⁻ by β -hydroxybutyrate dehydrogenase. Acetone and CO₂ are formed from the nonenzymatic degeneration of AcAc⁻. While the great majority of ketone bodies are synthesized from the acetyl CoA that is derived from the β -oxidation of fatty acids in the mitochondria, a small quantity can be synthesized from ketogenic amino acids during starvation. In addition, acetoacetyl CoA can be formed from FFA and cleaved to AcAc⁻ in the kidneys. The hepatic production of β -OHB⁻ and AcAc⁻ is about equal; however, during hyperketonemia, the concentration of β -OHB⁻ in the blood is 3 times greater than AcAc⁻. This is due to the decreased nicotinamide adenine diphosphate

(NAD)/NADH ratio in the hepatic mitochondria, which favors the reduction of AcAc to β -OHB⁻ and the preferential removal of AcAc⁻ or conversion of AcAc⁻ to β -OHB⁻ by skeletal muscle. The brain removes β -OHB⁻ and AcAc⁻ in a concentration-dependent manner.

The metabolism of ketone bodies during starvation is a critical element in the control of fuel homeostasis in humans. The demonstration that β -OHB⁻ and AcAc⁻ could serve as major fuels for the metabolism of the brain during starvation was a critical factor in evaluating the roles of fatty acid oxidation, amino acid mobilization, glucose conservation, and urinary nitrogen excretion during prolonged starvation. From this research it became clear that the abundance of FFA stored in human tissues provides a substantial reserve for the synthesis of ketone bodies. The metabolism of ketones by the brain during starvation greatly limits the need to use amino acids to make glucose to support the metabolism of this tissue. This is an important example of fuel sparing.

There are few synthetic processes that are quantitatively as large as the daily rate of ketogenesis during starvation. After an overnight fast, hepatic ketogenesis amounts to 10 g d⁻¹. After 2–3 days of starvation, the liver synthesizes over 100–150 g d⁻¹ of ketone bodies. Over this period, the average resting human oxidizes a minimum of 3 g of fat per kg FFM per day. Thus, a large, but not obese adult man, weighing 80 kg, with a body composition of 80% FFM (64 kg) and 20% fat mass (16 kg) oxidizes a minimum of 192 g of fat per day. About 60 g of the FFA derived from the lipid stores undergo β -oxidation in the liver to yield an estimated 113 g d⁻¹ of ketone bodies. The 90% of these water-soluble fuels undergo terminal oxidation, primarily by the brain and muscle. Most of the remainder is excreted in the urine. It is interesting to note that about half of the molecular weight of AcAc⁻ and β -OHB⁻ is oxygen. Thus, if 10–12 g of ketone bodies were excreted in the urine, only 5–6 g of the carbon skeleton would be derived from stored triglycerides. Since β -OHB⁻ and AcAc⁻ are excreted with near equimolar quantities of NH₄⁺, ketosis is an energetically cheap way to excrete nitrogen (the synthesis of urea requires four molecules of ATP per molecule of urea). Finally, for each gram of nitrogen lost in the urine 3.57 g of glucose is synthesized.

Gluconeogenesis

The *de novo* synthesis of glucose from nonhexose precursors (gluconeogenesis) occurs in the liver, kidney cortex, and, perhaps to a minor extent, in the small intestine. The precursor molecules for gluconeogenesis, and the organs involved in this process, change as starvation progresses. After an overnight fast, the liver is the primary, if not the sole, organ that makes a net contribution to glucose in the blood. The kidney extracts and releases glucose after an overnight fast, but its overall contribution is minimal. After 3 days of starvation, renal gluconeogenesis increases and by 10–18 days of total starvation the kidney contributes 50% of the glucose added to the blood. However, during prolonged starvation, total glucose added to the blood from gluconeogenesis is only 50% of what it was after an overnight fast when gluconeogenesis is greatly complemented by hepatic glycogenolysis. In addition, during prolonged starvation, only half of this glucose contributed to the

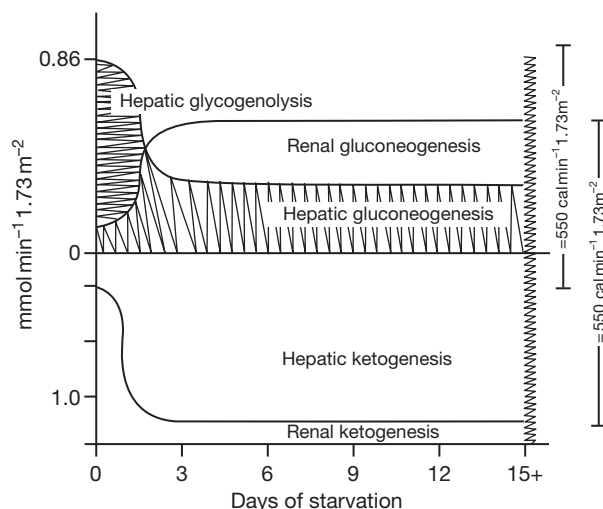


Figure 4 Schematic representation of fuel delivery by the liver and kidney during starvation. Hepatic gluconeogenesis, glycogenolysis and ketogenesis, and renal gluconeogenesis are expressed as $\text{mmol min}^{-1} 1.73 \text{ m}^{-2}$ and converted to calories per 1.73 m^2 body surface area. Hepatic glycogenolysis is rapidly dissipated during the first 1–3 days of fasting. The liver and kidney share the role of gluconeogenesis as starvation progresses. Hepatic ketogenesis becomes a dominant mechanism for supplying fuels to the blood after liver glycogenolysis is curtailed. Renal ketogenesis may contribute to ketonuria, but not to ketonemia.

blood is oxidized to CO_2 and H_2O . This is the result of the sparing of glucose by tissues that use more abundant fuels such as ketone bodies.

A schematic representation of the quantities of ketone bodies and glucose contributed to the blood by the liver and kidneys is shown in **Figure 4**. In general, glycogenolysis decreases as gluconeogenesis increases, and glucose production decreases as ketogenesis increases. An interesting aspect of gluconeogenesis during starvation is the relationship between renal gluconeogenesis and ammoniogenesis. Glutamine produced by the muscle is used by the kidney cortex as the major source of carbon for gluconeogenesis. The amino and amide groups are removed from glutamine to form α -ketoglutarate, which then enters the TCA cycle and is converted to glucose via renal gluconeogenesis. The kidney also has the ability to synthesize glucose from lactate, pyruvate, glycerol, amino acids, and other precursors, but only glutamine has been closely associated with net renal gluconeogenesis.

The decreased rate of glucose production during starvation is linked to the increased ketone body production. Hyperketonemia is accompanied by ketonuria and ammoniogenesis. Renal NH_4 production is coupled to renal gluconeogenesis, while the release of NH_3 by the kidney is linked to hepatic urea production. In summary, gluconeogenesis, ketogenesis, ammoniogenesis, acid–base balance, and ureagenesis are all tightly interdependent during starvation.

Urinary Nitrogen Excretion

Total, daily urinary excretion undergoes an asymptotic, progressive decrease during starvation (**Figure 5**). Humans who consume a diet with 100 g protein (16.0 g nitrogen) excrete

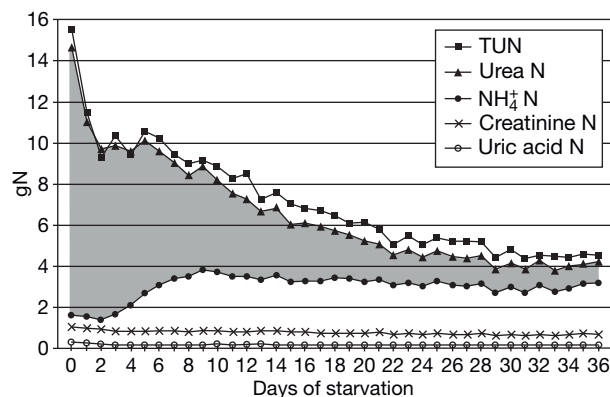


Figure 5 The total quantity and various components of urinary nitrogen excretion during prolonged starvation when five men and five women volunteers were given water, salt, and vitamins. The nitrogen contents of urea, ammonium, creatinine, and uric acid are displayed. TUN, total urinary nitrogen.

15.5 g of nitrogen in the urine. Most of the nitrogen (85%) is excreted as urea. Ammonium, uric acid, and creatinine account for most of the remaining urinary nitrogen. Starvation initially induces an acute decrease in total nitrogen excretion. After 3–6 days of starvation, total urinary nitrogen decreases gradually, so that after 36 days of starvation the nitrogen excretion is 4.6 g d^{-1} . The biggest decrease is in the exponential decay in urea nitrogen excretion. The excretion of ammonium ion (NH_4^+) parallels the excretion of AcAc^- and $\beta\text{-OHB}^-$; it peaks at $120\text{--}140 \text{ mmol d}^{-1}$ at 6–9 days of starvation. After 15–18 days of drinking only water, NH_4^+ overtakes urea as the dominant urinary nitrogen compound. Urinary NH_4^+ gradually decreases as starvation is prolonged. NH_4^+ , AcAc^- , and $\beta\text{-OHB}^-$ urinary excretion rates are sensitive to small quantities of ingested glucose. In early studies of metabolism during starvation, subjects were allowed to drink beverages that contained carbohydrates. This distorted both urea and NH_4^+ urinary excretion. The excretion of creatinine decays in a linear manner that is dependent upon diminishing body muscle mass that occurs as starvation progresses. The output of uric acid in the urine drops acutely during the first 4 days of starvation as ketonuria rises. Thereafter uric acid nitrogen excretion plateaus at 2 mmol d^{-1} (120 mg d^{-1}).

The rate of urea excretion provided the first clue that the concept that brain metabolized only glucose for energy had to be modified. Research in the first half of the twentieth century established that for each gram of urinary nitrogen excreted, 3.5 g of glucose could be synthesized from amino acids. Excluding the glycerol that is present in triglycerides, only trivial quantities of glucose could be derived from triglyceride. The brain requires 100–125 g of glucose or equivalent energy sources daily. Plasma FFAs cannot pass the blood–brain barrier to provide a source of energy. In 1967, it was shown that AcAc^- and $\beta\text{-OHB}^-$ were the primary fuels for brain metabolism during starvation. Therefore, gluconeogenesis is clearly important during starvation, but the quantity of glucose formed is less than the quantity needed to provide energy for the brain. However, it should be pointed out that tissues such as the red blood cells, kidney medulla, and lens of the eye use anaerobic glycolysis as a source of energy. The lactate and pyruvate generated by these tissues are recycled to the liver for the production of glucose.

Fuel Consumption during Starvation

While the liver and kidney release glucose into the blood during starvation, only the liver makes a net contribution of ketone bodies. However, more urinary excretion of AcAc^- and $\beta\text{-OHB}^-$ occurs than is extracted across the renal vascular beds, suggesting that the kidneys synthesize some of the ketone bodies excreted in the urine. Ketonuria augments ammoniogenesis. The deamination of glutamine to form α -ketoglutarate generates two molecules of NH_4^+ , which are released into the tubular urine to maintain electroneutrality and acid–base balance. As mentioned above, the α -ketoglutarate produced in this process enters the TCA cycle via anaplerosis, and is subsequently converted to glucose that is released into the blood. The kidney can also convert lactate, pyruvate, other amino acids, and perhaps glycerol into glucose. Clearly, the kidney shares the role of glucose production with the liver, but the kidneys make a net contribution to blood only after several days of starvation.

Early in starvation, the small intestine may provide gluconeogenic amino acids to the splanchnic (portal) system, but by 3 days of fasting there is no readily detectable release of alanine or glutamine into the portal blood in fasting humans.

After 3 days of starvation, skeletal muscles derive 50% of their energy requirements from ketone bodies and, as fasting progresses, muscle preferentially extracts a small quantity of AcAc^- and may release $\beta\text{-OHB}^-$. At this time in starvation, FFA provides the great majority of fuels for muscle.

After an overnight fast, the brain derives its energy from glucose, consuming $100\text{--}125\text{ g d}^{-1}$. As starvation progresses, blood glucose concentration decreases, and the concentration of AcAc^- and $\beta\text{-OHB}^-$ increases, ketone bodies then become the major fuels for brain metabolism. These water-soluble fuels have access to neurons and can supply two-thirds of the total energy requirements during prolonged starvation.

Concluding Statement

Humans are capable of surviving long periods without food, but there are major limitations in this ability. Death occurs

when about one-half of lean body tissue is consumed to provide energy. Morbidly obese individuals that are subjected to prolonged starvation can deplete the structural proteins of vital organs and die with triglyceride remaining in their adipose tissue. The adaptations that permit humans to survive in the absence of food provide a fascinating look at the underlying biochemical adaptations to starvation.

See also: **Metabolism Vitamins and Hormones:** Amino Acid Metabolism; Adipocyte Fat Mobilization: The Regulatory Roles of Perilipin 1, Adipose Triglyceride Lipase, and Hormone-Sensitive Lipase; Gluconeogenesis; Glycogen Metabolism; Insulin- and Glucagon-Secreting Cells of the Pancreas; Fatty Acid Oxidation; Carnitine and β -Oxidation; **Signaling:** Evolving Concepts of Leptin.

Further Reading

- Balasse EO (1979) Kinetics of ketone body metabolism in fasting humans. *Metabolism* 28(1): 41–50.
- Cahill GF, Jr. (1970) Starvation in man. *New England Journal of Medicine* 282: 668–675.
- Felig P and Bergman M (1995) The endocrine pancreas: Diabetes mellitus. In: Felig P, Baxter JD, and Frohman LA (eds.) *Endocrinology and Metabolism*, 3rd edn., pp. 1107–1250. New York, NY: McGraw-Hill.
- Keys AJ, Brozek J, Henschel A, Mickelsen O, and Taylor HL (1950) *The Biology of Human Starvation*, vols. I and II. Minneapolis, MN: University of Minnesota Press.
- Leibel RM, Rosenbaum M, and Hirsch J (1995) Changes in energy expenditure resulting from altered body weight. *New England Journal of Medicine* 332: 621–628.
- Owen OE, Kalhan SC, and Hanson RW (2002) The key role of anaplerosis and cataplerosis for citric acid cycle function. *Journal of Biological Chemistry* 277(34): 30409–30412.
- Owen OE, Licht JH, and Sapir DG (1981) Renal function and effects of partial rehydration during diabetic ketoacidosis. *Diabetes* 30: 510–518.
- Owen OE, Morgan AP, Kemp HG, et al. (1967) Brain metabolism during fasting. *Journal of Clinical Investigation* 46: 1589–1595.
- Owen OE and Reichard GF, Jr. (1971) Human forearm metabolism during progressive starvation. *Journal of Clinical Investigation* 50: 1536–1545.
- Owen OE, Smalley KJ, D'Alessio DA, et al. (1990) Resting metabolic rate and body composition of achondroplastic dwarfs. *Medicine* 69(1): 56–67.

Steroid/Thyroid Hormone Receptors

R D Ward and N L Weigel, Baylor College of Medicine, Houston, TX, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Ramesh Narayanan and Nancy L. Weigel, volume 4, pp. 111–116, © 2004, Elsevier Inc.

Glossary

Agonists Natural or synthetic ligands that bind to the hormone-binding domain of the receptor and stimulate activity.

Antagonists Ligands that compete with the agonists for binding to the hormone-binding domain, but do not cause activation of the receptor.

Nuclear receptors Ligand-activated transcription factors characterized by conserved zinc-finger motifs in their DNA-binding domains and smaller conserved regions in the hormone-binding domain.

Phosphorylation A posttranslational modification of an amino acid in which a phosphate group is added by a kinase.

Sumoylation A posttranslational modification of an amino acid in which a SUMO protein is added via an enzymatic cascade to modify the function of the protein.

Ubiquitination A posttranslational modification of an amino acid in which one or more ubiquitin monomers are added by the E3 enzyme. Ubiquitin frequently acts as a signal targeting the protein for degradation by proteasomes.

Overview of Nuclear Receptor Ligands and Mechanism of Action

Figure 1 shows the structures of some of the ligands for members of the nuclear receptor family. Estradiol is the primary ligand for both estrogen receptors, estrogen receptor α (ER α) and ER β . Testosterone, which is closely related to estradiol, is one of the two major ligands for the androgen receptor (AR). The ligands for the thyroid hormone receptor (T3 and T4) and the vitamin D receptor (1,25(OH)₂D₃) are also shown. The five major classes of steroids are synthesized from pregnenolone, which is derived from cholesterol through the actions of the cholesterol side-chain cleavage enzyme (P450_{scc}). The synthesis of the hormones is complex, with a number of alternate pathways leading to the same hormone. Testosterone and estradiol are derived from 17 α -hydroxy pregnenolone. Testosterone, the major circulating androgen, is synthesized in the testes of males and in the ovaries of females. It is important for development of the male reproductive tract, fertility, and secondary male characteristics. Estradiol, the major circulating estrogen, is produced in the ovaries of females. Estradiol is important in the female reproductive tract, playing roles in breast and uterine development, fertility, and also in bone and other tissues. Progesterone is synthesized directly from pregnenolone, and the major site of synthesis in females is the ovary. Progesterone, acting through the progesterone receptor (PR), plays important roles in the breast and uterus and in the maintenance of pregnancy. Progesterone is a precursor of corticosterone and aldosterone, which are synthesized in the adrenal glands. The mineralocorticoid, aldosterone, is important for salt retention in the kidney. The glucocorticoid, cortisol, is produced in the adrenals from 17 α -hydroxy pregnenolone and is important for regulation of carbohydrate metabolism; it also plays a role in suppressing immune responses. The secosteroid, 1,25(OH)₂D₃, is derived from cholesterol through a ultraviolet-catalyzed reaction in the epidermis followed by sequential hydroxylations in the liver and kidney. The action of 1,25(OH)₂D₃ is

important for calcium homeostasis and also plays a role in the differentiation of a variety of tissues. Thyroid hormones are produced from tyrosines and iodide in the thyroid gland. Thyroid hormones have multiple actions in regulating metabolism, typically increasing oxidation rates. The hormones are transported through the blood to their sites of action.

Figure 2 depicts the general mechanism of action of nuclear receptors. The ligands are all lipophilic compounds, which enter the cells by passive diffusion. They bind to their cognate intracellular receptors located either in the cytoplasm or the nucleus. The receptors can be divided into two classes – those that do not bind DNA in the absence of hormone (Figure 2(a)) and those that can bind to DNA in the absence of hormone (Figure 2(b)). The regulation of their activities differs somewhat. Classical steroid receptors such as AR, glucocorticoid receptor (GR), PR, mineralocorticoid receptor (MR), and ER belong to the first class; they are maintained in an inactive conformation capable of ligand binding by a complex of chaperone proteins including hsp90 and p23. Whether these complexes are nuclear or cytoplasmic depends upon the receptor. Upon ligand binding during the genomic pathway, the receptor changes its conformation, no longer binding to heat-shock proteins, homodimerizes and, if the unliganded receptor is localized in the cytoplasm, translocates to the nucleus. There, the receptor binds to DNA containing specific sequences called hormone response elements (HREs) that are typically found in the 5' flanking region of target genes. However, some HREs are more than 10 kb upstream and others are found in the introns of target genes. The receptor recruits components of the basal transcription machinery, as well as proteins termed coactivators that perform a variety of functions including histone acetylation that enhance transcription. In addition, these receptors can mediate a rapid, nongenomic activation of signaling pathways as discussed later. Unlike the receptors for classical steroids, thyroid receptors (TRs) are not bound to heat-shock proteins in the absence of hormone and, instead, are bound to the DNA as a heterodimer with the retinoid X receptor (RXR),

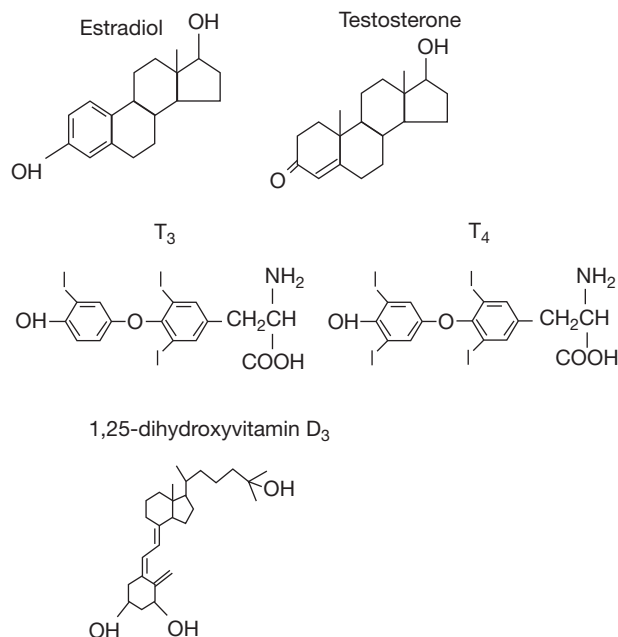


Figure 1 Structure of steroid and thyroid hormones. Estradiol is the ligand for estrogen receptor, testosterone is the ligand for androgen receptor, T₃-3,5,3',5'-tetraiodo-L-thyronine, T₄-thyroxine are the ligands for thyroid receptor, and 1,25-dihydroxyvitamin D₃ is the ligand for the vitamin D receptor.

another member of the steroid/thyroid hormone receptor superfamily (**Figure 2(b)**). In the absence of ligand, TR binds corepressors, which, in turn, bind histone deacetylases resulting in lower levels of histone acetylation and repression of target gene transcription. Hormone binding releases corepressors and promotes binding of coactivators, resulting in increased transcription. Agonist-bound receptors also are known to repress transcription of some target genes although mechanisms for repression are less well defined.

The Steroid/Thyroid Hormone Receptor Superfamily

The steroid/thyroid receptors are the largest family of ligand-activated transcription factors. In addition to the well-characterized steroid, thyroid, retinoid, and vitamin D receptors, the family contains receptors for numerous lipophilic metabolites and xenobiotics. A number of receptor family members are referred to as orphan nuclear receptors. These receptors were not initially identified on the basis that known hormones were regulating their physiological actions through a ligand-binding process, but rather through their sequence similarities to the known receptors. Investigators searched for additional steroid receptors using the conserved regions determined to be common to nuclear receptors in DNA-screening approaches and grouped them based on phylogeny. Although some orphan receptors were later found to be regulated by a ligand binding, for instance, peroxisome proliferator-activator receptor can respond to fatty acid derivatives and certain prostaglandins in addition to an array of synthetic ligands, some orphan receptors such as estrogen-related receptor and TR2 (testes receptor) still have no known ligand. Thus, these orphan receptors may either work independent of ligand or the ligand(s) responsible for their regulation have yet to be discovered.

Receptor Structure

Despite some evolutionary and functional differences, the steroid receptor family members have many similarities especially in their structure. As shown in **Figure 3(a)**, there are multiple domains in the receptors, with all receptors containing domains A–E and only a subset containing the additional F domain.

The N terminus

The N terminus or the A/B domain of the receptor is the least conserved domain among the family members. This region is the most variable in length ranging from a few amino acids to more than 500. This region has an activation function, AF-1, which contributes to the transcriptional activity of the receptor through binding of coactivators. The position of this region within the N terminus differs among the receptors. AF-1 has been defined functionally by deletion analyses, but to date no common structural motifs have been identified in this region.

The DNA-binding domain

The DNA-binding domain (DBD), region C, is important for the binding of receptor to the DNA and is the most highly conserved domain. This region has two type-2 zinc-finger motifs, which are responsible for DNA recognition and dimerization. Each finger is composed of four cysteines that coordinate with one zinc atom. Amino acids in this region also participate in receptor dimerization.

The hinge region

Downstream of the DBD is the hinge region (D), which contains a nuclear localization signal. This is a short lysine-rich region, with a high homology to the simian virus 40 T-antigen nuclear localization signal. Additional functions of this region are receptor specific.

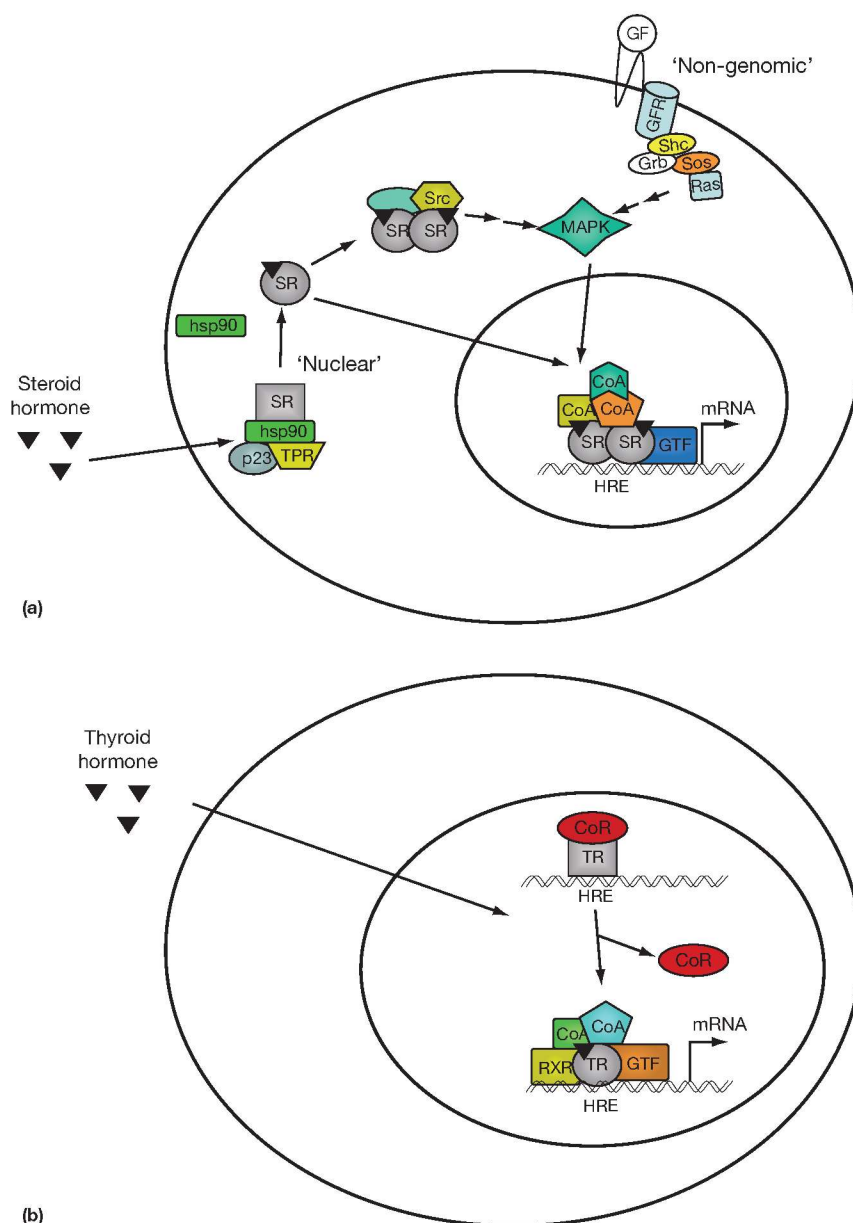


Figure 2 Mechanism of steroid (a) and thyroid (b) hormone action. The two classes are distinguished, in part, by whether they are associated with heat-shock proteins (a) like classical steroid receptors or are bound to DNA in the absence of hormone (b) like thyroid hormone receptor. Classical receptors (a) can bind hormone, dimerize, and translocate to the nucleus in the 'nuclear' pathway. In some cases, they also mediate a rapid activation of signaling pathways in the 'nongenomic' pathway. In both the classical, genomic pathway (a) and the thyroid receptor pathway (b), binding of the agonist causes dissociation of proteins that repress activity and promotes a conformation that induces recruitment of coactivators stimulating transcription of the target gene. SR, steroid receptor; HRE, hormone response element; GTF, general transcription factor; RXR, retinoid X receptor; TR, thyroid receptor; CoA, coactivator; CoR, corepressor; Hsp90, heat shock protein 90; TPR, tetratricopeptide; Src, Src tyrosine kinase; GF, growth factor; GFR, growth factor receptor.

The ligand-binding domain

The ligand-binding domain (E) is essential for the binding of ligand. The primary interaction site for the hsp complex is also in this domain. Also located in this region is the second activation function domain, AF-2, which is responsible for ligand-mediated transcription of target genes. The relative importance of AF-2 and AF-1 in inducing transcription is receptor- and cell-type specific. The structures of the hormone-binding domains

of several receptors have been determined using X-ray diffraction. The hormone-binding domain consists of a series of 12 α -helices. Binding of hormone causes a substantial conformational change in the receptor exposing AF-2 for interactions with coactivators. This domain also contains the strongest dimerization interface in most steroid receptors. The function of the F domain, located at the C terminus of some receptors such as the ER is not well defined.



Figure 3 Receptor structure and DNA binding elements. (a) Shows the common structural features of nuclear receptors using GR as an example. The A/B region contains the AF-1, a region important for transcriptional activation. C is the DNA-binding domain, the most conserved region in the nuclear receptors. D contains a nuclear localization sequence. E contains the hormone-binding domain and second activation function AF-2. Some receptors also contain a C-terminal extension, termed the F domain, whose physiological function is not well described. (b) Shows sequences of consensus hormone response elements. The consensus sequence for a GRE (binds GR, AR, PR) and an ERE (binds ER) are shown. Vitamin D receptor and the thyroid hormone receptors bind to direct repeats separated by three and four nucleotides, respectively. Other receptors bind to direct or inverted repeats with a spacing of 0–6 (n_x indicates that the half-site may be separated by 0–6 nucleotides). GRE, glucocorticoid response element; ERE, estrogen response element; VDRE, vitamin D response element; TRE, thyroid response element; GR, glucocorticoid receptor; AF, activation function.

Receptor Binding to DNA

All of the classical receptors bind to their cognate HREs as dimers. The consensus binding sequence for AR, PR, GR, and MR, shown in **Figure 3(b)**, contains two half-sites separated by three nucleotides with the sites oriented to form a palindrome. ER recognizes a related pair of half-sites with the same spacing and orientation. Each monomer binds to a half-site. The class-II receptors, including vitamin D receptor (VDR) and TR, bind to pairs of half-sites whose sequences are identical to the ER half-site, but whose orientation (direct or inverted repeats) and spacing (0–6 nucleotides) determine the specificity of binding. TR and VDR each heterodimerize with RXR and bind to the 3' end (half of the HRE), whereas RXR binds to the 5' end (half of the response element). Although these sequences represent the consensus binding sites, natural sequences may differ significantly and promoters may contain combinations of HREs as well as individual half-sites all of which contribute to the final activity.

Steroid Receptor Coregulators

When the receptor binds to the DNA, it recruits a series of protein complexes that facilitate recruitment of proteins required for basal transcription of the receptors. Receptors bind proteins or protein complexes that modulate receptor activity through alterations in chromatin structure; these are termed coactivators and corepressors. Coactivators are defined as proteins that interact with the receptors and increase their ability to transactivate the target gene. The mechanism by which

individual coactivators achieve this can vary. More than 100 candidate coactivators have been identified. Although some coactivators function only with steroid receptors and a small subset of other transcription factors, others are used by many transcription factors. The best characterized of the steroid receptor coactivators (SRCs) is the p160 family of coactivators: SRC-1, SRC-2 (GRIP1, TIF2), and SRC-3 (Rac3, AIB1). These bind to the receptor recruiting additional coactivators, including cAMP response element binding (CREB)-binding protein (CBP), CBP-associated factor (p/CAF), and coactivator-associated arginine methyltransferase (CARM-1). CBP/p300, p/CAF and some of the p160 proteins are histone acetyl transferases and their binding increases local histone acetylation. Other coactivators include the DRIP/TRAP (D receptor interacting protein/thyroid hormone receptor-associated protein) complex. Many coactivators interact with AF-2 located in the ligand-binding domain (LBD). In other cases, coactivators interact with the AF-1 region and some interact with both domains. Interactions with AF-2 are typically mediated by LXXLL (L=leucine and X=any amino acid) motifs in the coactivator.

Another class of proteins, termed corepressor, reduces the activation of target gene transcription through interaction with the receptors. The best-characterized nuclear receptor corepressors are nuclear receptor corepressor (NcoR) and silencing mediator of retinoid and thyroid receptors (SMRTs). These proteins bind histone deacetylase complexes and also interact with class II receptors in the absence of hormone resulting in local reductions in histone acetylation. Corepressors do not bind to unliganded steroid receptors. However, steroid receptor antagonists cause changes in the conformation of the hormone-binding domain that induce binding of the corepressors.

Steroid Receptor Agonists and Antagonists

Although steroid receptor family members are important for normal physiological processes, there are a number of instances in which it is desirable to block the actions of selected steroid receptors. These include breast cancer (ER) and prostate cancer (AR). Thus, although natural antagonists of steroid receptor action have not been identified, much effort has been devoted to identifying compounds that will antagonize hormone action. These compounds compete with the natural ligand for binding to the hormone-binding domain of the receptors. Although some antagonists block dissociation from heat-shock protein complexes or destabilize the receptor, most of the antagonists promote dissociation from heat-shock proteins and cause the receptors to bind to DNA. However, the conformation induced by the antagonist differs from that induced by agonist. This prevents recruitment of coactivators to AF-2 and, instead, promotes recruitment of corepressors. In some cases, it is desirable to maintain the activity of a receptor in some tissues while inhibiting activity in other tissues. The most common example of this is the need for tissue-specific regulation of ER activity. Estradiol is important for maintaining bone mass and postmenopausal women frequently develop osteoporosis. However, estradiol can promote uterine cancer and may also be detrimental in breast. Thus, a great deal of effort has been devoted to developing selective estrogen receptor modulators (SERMs) which have tissue-specific agonistic and antagonistic activities.

Tamoxifen (a SERM) has been used in the treatment of breast cancer, but increases the risk of uterine cancer. A newer SERM, raloxifene, is an antagonist in both breast and uterus, but acts as an agonist in bone.

Cross-Talk between Nuclear Receptors and Cell Signaling Pathways

Phosphorylation and other Posttranslational Modifications of Receptors and Coactivators

The nuclear receptors and their coactivators are phosphoproteins; in some cases, enhanced cell signaling is sufficient to induce the transcriptional activity of the receptor in the absence of normal levels of ligand. The ability to be activated by cell-signaling pathways alone is receptor specific, although changes in cell signaling modulate the activity of all of the receptors. Estrogen receptors are activated both by growth factor pathways and by activation of protein kinase A. Other receptors such as GR, require hormone for activity. Recent studies show that there are many mechanisms to regulate receptor activity through posttranslational modification. Examples of posttranslational modification include phosphorylation, sumoylation, and ubiquitination. Phosphorylation of steroid receptors on Ser, Thr, and Tyr residues can occur in the absence of ligand or can be induced by ligand binding and/or kinase signaling to regulate various functions, including protein stability, hormone sensitivity, DNA binding, subcellular localization, and protein interactions. Lysine residues can be modified by ubiquitination, sumoylation, or acetylation and can regulate protein turnover, subcellular localization, DNA-binding, transcriptional potential, or affect chromatin structure that controls the availability of promoter regions to regulatory proteins. Posttranslational modifications of coactivators are also important for regulation of receptor-specific functions.

Functional Interactions between Nuclear Receptors and Signal-Regulated Transcription Factors

In addition to altering transcription through direct binding to DNA, nuclear receptors alter transcription through interactions with other transcription factors. In some cases, these protein-protein interactions enhance transcription, while in others they prevent binding of the transcription factor to its DNA target site in a process referred to as transrepression. For instance, GR can interact with additional transcription factors in order to prevent these proteins from binding to and activating their respective targets. This type of process has been shown to be important for the antiinflammatory actions of GR. In other cases, the receptor binds to the factors on their DNA target sites influencing (either + or –) the transcription of a target gene. GR also can inhibit the transcription of certain genes by direct binding to negative GREs in the promoter regions of these genes. **Figure 4** shows the comparison between two of these pathways for ER. In the upper panel is the classical DNA-binding-dependent induction of transcription. We next depict the ability of ER to induce transcription through interactions with AP-1 complexes. In this instance, both estradiol as well as SERMs will stimulate the activity of AP-1.

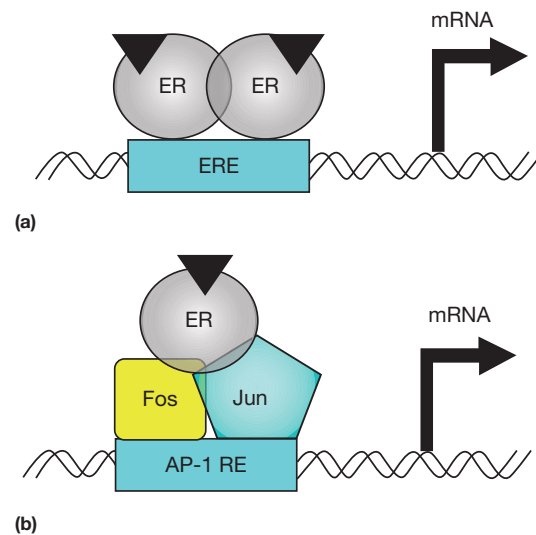


Figure 4 Mechanisms of transcription activation by nuclear receptors. (a) Shows the classical pathway for transcriptional activation with a steroid receptor dimer binding directly to a hormone response element. (b) Shows an alternative mode of activation. In this case, a receptor such as the estrogen receptor binds to another transcription factor in a process referred to as tethering and influences transcription through its interactions with the transcription factor and the recruitment of coregulator proteins to the complex. ER, estrogen receptor; AP-1 RE, AP-1 response element; ERE, estrogen response element.

Nuclear Receptor Stimulation of Cell-Signaling Pathways

Both of the pathways above can be considered genomic pathways in that the nuclear receptor acts by directly altering transcription. Nuclear receptors can also act through stimulating kinase activity, although the final downstream target may be a change in transcription. These actions are rapid (minutes) and are termed nongenomic (**Figure 2**). This process is thought to occur close to the cell membrane and involves the cross-talk of steroid receptors with various adaptor proteins, G proteins, and receptor tyrosine kinases. There is evidence that activation of nuclear receptors can lead to downstream activation of mitogen-activated protein kinase. In some cases, this is through activation of Src kinase and in others through generation of a ligand for a growth factor receptor. There are numerous other examples of these rapid actions. Induction by estradiol of nitric oxide synthase activity in endothelial cells is a rapid response that does not require transcription. Thus, steroid/thyroid hormones alter cellular activities through multiple mechanisms.

See also: **Signaling:** Mitogen-Activated Protein Kinase Family; Thyroid-Stimulating Hormone/Luteinizing Hormone/Follicle-Stimulating Hormone Receptors; Vitamin D Receptor.

Further Reading

Faus H and Haendler B (2006) Post-translational modifications of steroid receptors. *Biomedicine and Pharmacotherapy* 60: 520–528.

- Hammes SR and Levin ER (2007) Extranuclear steroid receptors: Nature and actions. *Endocrine Reviews* 28: 726–741.
- McDonnell DP, Connor CE, Wijayaratne A, Chang CY, and Norris JD (2002) Definition of the molecular and cellular mechanisms underlying the tissue-selective agonist/antagonist activities of selective estrogen receptor modulators. *Recent Progress in Hormone Research* 57: 295–316.
- Tsai MJ and O'Malley BW (1994) Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annual Review of Biochemistry* 63: 451–486.
- Weigel NL and Moore NL (2007) Kinases and protein phosphorylation as regulators of steroid hormone action. *Nuclear Receptor Signalling* 5: e005.
- Whitfield GK, Jurutka PW, Haussler CA, and Haussler MR (1999) Steroid hormone receptors: Evolution, ligands, and molecular basis of biologic function. *Journal of Cellular Biochemistry – Supplement* 32–33: 110–122.
- Wolf IM, Heitzer MD, Grubisha M, and DeFranco DB (2008) Coactivators and nuclear receptor transactivation. *Journal of Cellular Biochemistry* 104: 1580–1586.

Relevant Websites

<http://www.nursa.org> – The Nuclear Receptor Signaling Atlas.

Store-Operated Calcium Channels

A Amcheslavsky, A V Yeromin, A Penna, and M D Cahalan, University of California, Irvine, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

dOrai/*Drosophila* Orai The *Drosophila melanogaster* homolog of ORAI proteins. dOrai subunits oligomerize to form the pore domain of the calcium-release activated calcium (CRAC) channels in *Drosophila melanogaster*.

dStim/*Drosophila* Stim The *D. melanogaster* homolog of STIM proteins. dStim is the endoplasmic reticulum (ER) membrane resident protein that senses decreased ER calcium levels, oligomerizes and translocates toward the plasma membrane (PM) where it physically interacts with and activates dOrai. Together, dOrai and dStim form the CRAC channel in *D. melanogaster*.

ORAI ORAI molecules, named after the three keepers of the gates of heaven in Greek mythology, are conserved transmembrane proteins that form the pore of the CRAC channels. 'ORAI' is used to describe ORAI homologs from any species.

Orai1/Orai2/Orai3 The three mammalian homologs of ORAI proteins. Orai1 subunits oligomerize to form the pore domain of the CRAC channels in mammalian T cells. The three homologs are expressed in other tissues as well.

STIM Stromal interaction molecule, STIM, is a conserved endoplasmic reticulum (ER) membrane resident protein that senses decreased ER calcium levels, oligomerizes and translocates toward the PM where it physically interacts with and activates ORAI oligomers. Together, ORAI and STIM proteins form the CRAC channel. 'STIM' is used to describe STIM homologs from any species.

STIM1/STIM2 The two mammalian homologs of STIM proteins. STIM1 and STIM2 are ER membrane resident proteins that sense decreased ER calcium levels, oligomerize and translocate toward the PM where they physically interact with and activate Orai1/Orai2/Orai3. Together, STIM and ORAI proteins form the CRAC channel.

Introduction

Calcium ions (Ca^{2+}) are vital for cell function. Muscle contraction, synaptic transmission, and exocytosis are just a few examples of physiological processes that cannot occur without Ca^{2+} . Disruption of Ca^{2+} homeostasis can cause apoptosis and cell death. As a second messenger, Ca^{2+} also drives gene expression in many cell types. In lymphocytes, for instance, a sustained increase of cytosolic Ca^{2+} concentration ($[\text{Ca}^{2+}]_i$) is required for cytokine production and cell proliferation through the nuclear factor of activated T cells (NF-AT) transcription factor pathway. Under physiological conditions, receptor stimulation (T-cell receptor (TCR) in T cells, fragment, crystallizable region, receptor (FcR) in mast cells, etc.) induces a cascade of biochemical reactions which result in the production of inositol-1,4,5-trisphosphate (IP_3) by phospholipase C. IP_3 then binds to its receptor in the endoplasmic reticulum (ER) membrane and causes Ca^{2+} release from the ER into the cytosol. But the limited amount of Ca^{2+} contained in the ER allows only a small increase in $[\text{Ca}^{2+}]_i$ that lasts just a few minutes. Ca^{2+} release from the ER is not sufficient for activation. A mechanism for cytosolic Ca^{2+} influx is, therefore, required for a sustained and productive Ca^{2+} signal following receptor engagement.

In 1986, James Putney proposed that depletion of the ER Ca^{2+} store may initiate Ca^{2+} influx across the plasma membrane (PM) in nonexcitable cells through a mechanism called capacitative Ca^{2+} entry or store-operated Ca^{2+} entry (SOCE). Patch-clamp experiments in Jurkat T cells and mast cells, subsequently, revealed a tiny Ca^{2+} -selective current that indeed was activated in response to ER Ca^{2+} store depletion. In experiments, this same current was evoked by six distinct experimental maneuvers that have as their common denominator a

reduction in ER luminal Ca^{2+} concentration: (1) the physiological immunoreceptor stimulus that generates IP_3 , causing Ca^{2+} release through IP_3 receptor channels in the ER membrane; (2) addition of IP_3 to the pipette internal solution to open IP_3 receptor channels; (3) application of Ca^{2+} ionophores such as ionomycin that transport Ca^{2+} across membranes and thereby cause Ca^{2+} release; (4) passive store depletion by dialysis of the cell with Ca^{2+} chelators from the patch pipette (this lowers ER Ca^{2+} gradually as Ca^{2+} leaks from the ER into the cytosol and is then buffered); (5) loading of Ca^{2+} chelators into the ER lumen to reduce the intraluminal-free Ca^{2+} concentration; and (6) application of sarco/ER Ca^{2+} adenosine triphosphatase (SERCA) pump inhibitors such as thapsigargin to reduce intraluminal Ca^{2+} by blocking reuptake into the ER. As the release of Ca^{2+} from the ER store is necessary and sufficient for current activation, this tiny Ca^{2+} current was named Ca^{2+} release-activated Ca^{2+} (CRAC) current. Thus, unlike in excitable cells where Ca^{2+} enters mostly through voltage-gated Ca^{2+} channels or ligand-gated Ca^{2+} channels, in nonexcitable cells and particularly in immune cells the main Ca^{2+} influx pathway is store-operated and mediated by CRAC channels.

The biophysical properties of the CRAC current are well defined. The channel is gated by ER Ca^{2+} store depletion *per se* and carries an inwardly rectifying and highly Ca^{2+} -selective ($P_{\text{Ca}}/P_{\text{Na}} > 1000$) current. Like voltage-gated Ca^{2+} channels, CRAC channels conduct Na^+ only in the complete absence of divalent cations but do not conduct monovalent cations the size of Cs^+ or larger, placing a limit on the pore size of about 3.5 Å. Consistent with the small pore diameter, the single channel conductance is extremely low (21 fS in 20-mM Ca^{2+}). In addition to Ca^{2+} , the channel is able to conduct other divalent cations such as Ba^{2+} or Sr^{2+} , but not Mg^{2+} .

Ca^{2+} influx results in rapid inactivation of the CRAC current (in mammalian but not *Drosophila* cells). This biophysical fingerprint was developed from the late 1980s through the 1990s, but the underlying molecular constituents and the activation mechanism were only recently illuminated. It took much longer than the molecular identification of many channels because two essential components are required. STIM proteins provide the mechanistic link between ER Ca^{2+} store depletion and the PM; and ORAI proteins form the Ca^{2+} -selective pore of the CRAC channel.

What about the transient receptor potential (TRP) channel family, members of which were once among the leading candidates to form the CRAC channel? The biophysical properties of TRP channels vary widely in their gating and cation-selectivity characteristics. Several members of the canonical TRP (TRPC) channel subfamily are activated by signaling pathways that cause Ca^{2+} release from the ER. However, these same signaling pathways also generate messengers such as diacylglycerol which on its own activates some TRPC channels. The diverse-gating characteristics of TRP channels resulted in uncertainty regarding their ability to respond to ER Ca^{2+} store depletion. Moreover, compared to CRAC channels formed from ORAI subunits, TRPC channels are less selective for Ca^{2+} , allowing a significant component of Na^+ current. The discovery of STIM proteins has allowed researchers to test the hypothesis of STIM involvement in TRPC channel activation, and while some data suggest STIM-mediated activation of TRPC channels, it remains controversial.

Identification of STIM and Orai Proteins

In 2005, results of RNA-interference (RNAi) screens revealed STIM proteins as an essential component of the CRAC channel activation mechanism. An initial RNAi candidate screen in *Drosophila* S2 cells sought to identify molecules responsible for thapsigargin-induced Ca^{2+} entry, an experimental way to activate SOCE in intact cells without involving receptor engagement that could trigger potentially irrelevant proximal signaling pathways. Hits were defined by an apparent decrease in SOCE as a result of RNAi treatment for a set of selected genes. *Drosophila* Stim (dStim) was the only hit in this screen; the mammalian homolog STIM1 was then shown to be required for SOCE in T lymphocytes and other human cells. *Drosophila melanogaster* and *Caenorhabditis elegans* have only one Stim protein, whereas two homologs are present in vertebrate genomes: STIM1 and STIM2. An independently conducted RNAi screen identified both STIM1 and STIM2 for SOCE in human HeLa cells. STIM1 had been formerly identified as a molecule required for interaction of pre-B cells with bone marrow stromal cells and given the name stromal interacting molecule, later transformed to 'STIM'. It had also been implicated in tumor suppression. The essential role of dStim in *Drosophila* cells and STIM1 in human cells in mediating SOCE and CRAC current was unambiguously evidenced by Ca^{2+} imaging and whole-cell recording.

STIM proteins are essential mediators of SOCE, but they do not make up the ion-conducting pore of the CRAC channel. Following the success of the candidate-screening approach that

correctly identified a requirement for STIM proteins, the search for the CRAC channel continued by extending RNAi screening to the complete genome of *Drosophila* S2 cells. In 2006, three independent genome-wide screens led to the identification of the previously unheralded *Orai* (formerly known as *olf186-F*, renamed *Orai* and also known as *CRACM*). Only one *Orai* gene is found in the *Drosophila* genome (dOrai), but three mammalian homologs exist: Orai1, Orai2, and Orai3.

The mechanistic roles of STIM and ORAI as the two essential molecular components underlying the CRAC current were clarified by key overexpression and mutagenesis studies. Coexpression of dStim and dOrai, or STIM1 and Orai1, amplified the CRAC current to extremely high levels of functional expression. These studies suggested a requirement for both proteins and nothing else. Mutagenesis of STIM demonstrated its ER Ca^{2+} sensor and messenger functions that trigger signaling from the ER to the PM to activate Ca^{2+} influx. Mutagenesis of ORAI demonstrated its pore-forming function. In addition to activating Ca^{2+} -selective Orai1 channels, available evidence indicates that STIM1 can also activate several Ca^{2+} -permeable TRPC channels to cause SOCE. How does STIM1 convey information regarding ER Ca^{2+} store depletion to its PM partners? Briefly, once the ER luminal Ca^{2+} concentration drops, STIM physically translocates from the bulk ER to specialized junctions at the PM where it interacts directly with ORAI or TRPC subunits and activates channel activity, resulting in SOCE. In this article, we discuss each step of this signaling mechanism.

STIM Localization and Structural Domains

STIM proteins are found in both ER and PMs. However, the ER-resident STIM protein population is predominant and is the one implicated in SOCE. STIM proteins are highly conserved type I membrane proteins, with N-terminal and cytosolic C-terminal domains flanking a single transmembrane segment (Figure 1). The N-terminus in the ER lumen is composed of a signal peptide sequence, two EF-hand domains (one of which is responsible for binding and, thus, sensing ER Ca^{2+}), and a glycosylated sterile α motif (SAM) protein-protein interaction domain. A nuclear magnetic resonance structure of the EF-hand and SAM domains of STIM1 elucidates the Ca^{2+} -binding site of the canonical EF hand 1 that triggers signaling from the ER to the PM following Ca^{2+} store depletion. Adjacent to the hydrophobic membrane-spanning sequence, a cytosolic ezrin-radixin-moesin domain encompasses two coiled-coil domains. The proximal coiled-coil appears to mediate dimerization of STIM under resting conditions. The distal coiled-coil domain plays a key role in activating current through ORAI channels in the PM, functioning as a CRAC-activating domain (CAD). A shorter inactivation domain of STIM1, found directly downstream of the distal coiled coil, is involved in fast Ca^{2+} -dependent inactivation. The more distal STIM C-termini vary among classes of organisms. Invertebrate STIM proteins lack the serine/proline- and lysine-rich domains that may help to anchor STIM1 to negatively charged phospholipids in the PM following Ca^{2+} store depletion.

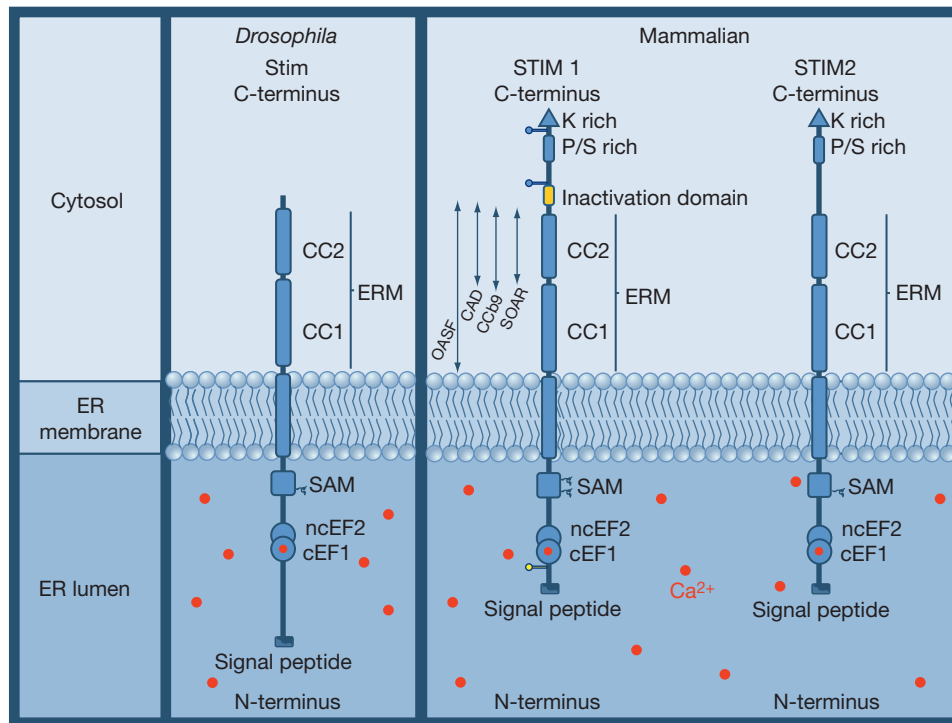


Figure 1 Functional domains of *Drosophila* Stim and mammalian STIM1 and STIM2. STIM proteins implicated in store-operated calcium entry reside in the ER membrane. *Drosophila* Stim has a longer N-terminus, and shorter C-terminus, than the mammalian homologs STIM1 and STIM2. The N-termini of all STIM proteins include a single signal peptide domain, two EF-hand domains and a sterile α motif (SAM) domain. Of the two EF hands, the canonical EF1 (cEF1) domain binds Ca^{2+} (●) when ER stores are filled, and releases Ca^{2+} upon ER store depletion. N-linked glycosylation sites within the SAM domain are indicated as ⋈. Just downstream of the single transmembrane domain is found a conserved ezrin–radixin–moesin (ERM) domain made up of two coiled-coil (CC) domains: a proximal domain that mediates dimerization and a more distal domain that contains a CRAC-activating domain (CAD, overlapping with segments defined by other groups and designated OASF, Orai-activating Stim fragment; CCb9, coiled-coil domain-containing region b9; and SOAR, Stim Orai-activating region). Recent studies on STIM1 have also demonstrated an inactivation domain (■) following the coiled-coil domains. The downstream domains of the STIM C-termini vary; only mammalian homologs contain proline/serine- and lysine-rich domains. Functionally significant sites of posttranslational modification are indicated by ball and stick moieties: a conserved cysteine modification (S-glutathionylation of C56) and two STIM1-specific phosphorylation sites (S486 and S668).

STIM Proteins as ER Calcium Sensors

Ca^{2+} unbinding from STIM is the molecular switch that triggers SOCE following depletion of the ER Ca^{2+} store. The presence of an N-terminal EF-hand domain suggested a possible function of STIM proteins in sensing changes in ER luminal Ca^{2+} concentration. In fact, when the EF hand of STIM1 was mutated in such a way as to prevent it from binding Ca^{2+} , the CRAC channel was shown to become activated as cells exhibited constitutive Ca^{2+} influx, bypassing the normal requirement for ER Ca^{2+} store depletion to activate the CRAC channel. The ER luminal Ca^{2+} concentration in a resting cell is $>400 \mu\text{M}$, orders of magnitude larger than the cytosolic Ca^{2+} concentration of $\sim 50 \text{ nM}$. The Ca^{2+} concentration gradient across the ER membrane is maintained by continuous activity of the SERCA pump. Experimental data showed that an isolated EF hand-SAM truncation of STIM1 binds Ca^{2+} in a 1:1 ratio, with a dissociation constant in the hundreds of micromolar range. Moreover, measuring the luminal ER Ca^{2+} concentration simultaneously with CRAC current demonstrated a steep relationship (Hill coefficient of 4) between ER Ca^{2+} concentration and PM Ca^{2+} channel activity, which is

initiated when luminal Ca^{2+} falls below $400 \mu\text{M}$. The EF hand of STIM1 binds Ca^{2+} with low affinity when ER Ca^{2+} stores are full and releases Ca^{2+} following ER store depletion, leading through several further steps to Ca^{2+} influx across the PM.

STIM Oligomerization: Signal Initiated

Conformational changes in the EF-SAM domain trigger STIM signaling following Ca^{2+} store depletion. Under resting cellular conditions, STIM proteins form dimers that are stabilized by interactions between the membrane-proximal coiled-coil domains. Upon ER Ca^{2+} store depletion, STIM dimers oligomerize; this step is mediated by the N-terminal SAM domains adjacent to the EF-hands that trigger conformational changes when Ca^{2+} unbinds. Notably, chemically induced oligomerization at the N-termini of chimeric STIM1 proteins suffices to send STIM1 to the PM and to activate CRAC current, independent of ER Ca^{2+} store depletion. STIM oligomerization is, thus, a crucial step in CRAC channel activation, causing STIM1 translocation to ER–PM junctions where the ER membrane and PM are closely apposed.

STIM Translocation to the PM: Signal Transmitted

STIM itself senses and then relays the message of ER store depletion directly to PM-resident ORAI channels. Following ER Ca^{2+} store depletion, STIM proteins oligomerize and translocate while still embedded in the ER membrane, forming clusters known as 'puncta' at junctions that are closely apposed to the PM. Both native and fluorescently tagged STIM proteins were shown to migrate to the PM only following Ca^{2+} store depletion. However, mutant Stim or STIM1 proteins lacking the ability to bind Ca^{2+} not only caused constitutive activation of CRAC channels but also were shown to be translocated to the PM even though ER Ca^{2+} stores remained filled. These key studies validated the ER Ca^{2+} sensing and messenger functions of STIM. Förster resonance energy transfer microscopy and electron microscopy have revealed that STIM1 migrates close enough to the PM (within 20 nm) to contact ORAI channel-forming subunits directly. In small cells, oligomerization and translocation of STIM molecules requires 10–30 s following ER Ca^{2+} store depletion and is reversible upon refilling of the ER Ca^{2+} store. There is still much to learn about the subcellular pathway by which STIM proteins migrate from the bulk ER toward the PM as Ca^{2+} -unbound oligomers and back again when the ER Ca^{2+} store is refilled. The ER structure does not appear to change dramatically following store depletion, yet STIM1 migrates directionally and reversibly from bulk ER to predefined junctions with the PM when Ca^{2+} stores are depleted. STIM1 binds to EB-1 at the growing plus end of microtubules, but the role of microtubules in localization and transport of STIM1 to ER–PM junctions needs to be clarified.

STIM Signal Turned On or Off by Posttranslational Modification

Reactive oxygen species (ROS) have long been implicated in aberrant Ca^{2+} signaling through a variety of mechanisms. In addition to Ca^{2+} sensing, STIM proteins may function as cellular redox sensors to initiate Ca^{2+} influx under conditions of oxidative stress induced by ROS. A recent study indicates that the STIM signal from ER to the PM can be turned 'on' by posttranslational modification initiated by cellular oxidants at a conserved cysteine residue (C56 in human STIM1) close to the Ca^{2+} -sensing EF hand of STIM1. A covalent S-glutathionylation reaction at this residue appears to reduce the Ca^{2+} affinity of the adjacent EF hand, resulting in constitutive activation of Ca^{2+} influx through CRAC channels. Moreover, mutation of C56 to alanine replicates the store-independent Ca^{2+} phenotype, highlighting the importance of this residue for Ca^{2+} sensing. Constitutive activation of the STIM1-signaling mechanism may constitute an important and potentially dangerous mechanism by which STIM1 posttranslational modification under conditions of oxidative stress can lead to inappropriate Ca^{2+} signals that result in mitochondrial overload and cell death.

In contrast, the STIM signal can be turned 'off' by a very different posttranslational modification. STIM1 has numerous sites of phosphorylation by serine–threonine kinases; some of these play a role in the previously documented cell-cycle dependence of CRAC channel activity. During M phase, when

cells are rounded up and dividing, the usual experimental measures to deplete the ER Ca^{2+} store fail to activate CRAC current. Suppression is due to phosphorylation during mitosis of STIM1 at two key serine residues (S486 adjacent to CAD and S668 near the poly-lysine C-terminus in human STIM1), resulting in a failure of STIM1 to translocate to the membrane. This uncoupling of ER Ca^{2+} store depletion and SOCE could result from inhibition of STIM1 oligomerization or interference with undetermined molecular interactions that help guide STIM1 to ER–PM junctions.

STIM-Mediated Channel Clustering and Subunit Organization: Signal Received

How does STIM trigger SOC channel activation, and how is the signal modulated and turned off? In addition to serving as the ER Ca^{2+} sensor and the messenger to the PM, STIM organizes ORAI or TRPC subunits into mirror-image molecular clusters at ER–PM junctions and activates Ca^{2+} influx. The newly formed coclusters of messenger STIM and corresponding ORAI channel proteins colocalize precisely with sites of Ca^{2+} influx. This induced aggregation of ORAI (or TRPC) channels concentrates local Ca^{2+} signaling at the particular regions of the cell where activator STIM molecules and ORAI channels cocluster. In T cells, this takes place at the immunological synapse formed with the antigen-presenting cell (APC). Engagement of TCRs at the leading edge of a T cell with antigen bound to major histocompatibility complex on the APC initiates signaling that leads to IP_3 generation and Ca^{2+} release from the ER. Following ER Ca^{2+} store depletion STIM1 translocates preferentially to the immune synapse (some also distributes in caps at the rear end of the cell), and concentrates Orai1 locally such that Ca^{2+} influx into the T cell is maximal immediately adjacent to the synapse. Thus, in at least this circumstance, STIM1 promotes Ca^{2+} influx in a highly specialized subcellular-signaling platform where kinases, adapter proteins, K^+ channels, and other signaling molecules are also concentrated.

At the molecular level, the CAD domain of STIM1 binds directly to the Orai1 protein. STIM1 plays a dual role in Orai1 channel activation, promoting subunit assembly into tetramers and inducing a conformational change in the Orai1 tetramer that opens the Ca^{2+} -selective pore. The functional unit of the CRAC channel is unambiguously a tetramer of dOrai (or Orai1) subunits, but there is disagreement about the state of dOrai (dimers) and Orai1 (tetramers) in resting cells. Available evidence points to a stoichiometry of two STIM1 molecules per tetramer of Orai1 in the active state. STIM-mediated dimerization of ORAI dimers is a unique mechanism of channel subunit assembly mediated by the activator protein.

Opening ORAI: Signal Delivered

How exactly does STIM binding open ORAI tetramers? Abundant evidence indicates that the cytosolic C-terminus of STIM1 functions as the effector domain to open ORAI through a physical electrostatic interaction between the C-terminal distal coiled-coil domain of STIM1 and the coiled-coil domains present in the cytosolic C-termini of ORAI proteins. Indeed, the

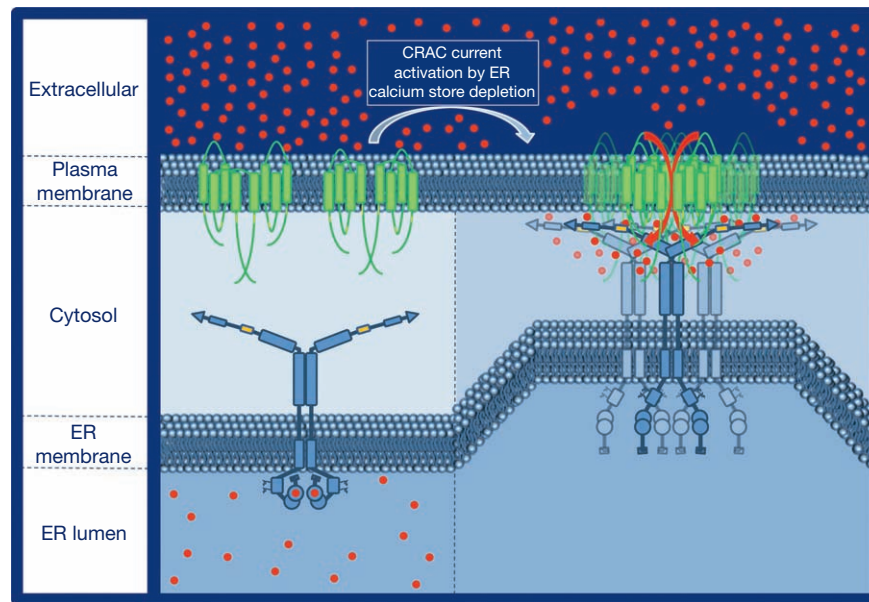


Figure 2 STIM1 translocates to ER–PM junctions to interact with Orai1 and activate SOCE after ER Ca^{2+} store depletion. Background shading indicates Ca^{2+} concentrations from ~ 100 nM in cytosol (light blue) to 2 mM extracellular (dark blue). Under resting conditions (left), Ca^{2+} (●) is bound to the EF hand of STIM1 (shown as a dimer) and Orai1 subunits (■, shown as dimers) are nonconducting. Inactivation domains are indicated for STIM1 (□) and Orai1 (yellow lines in the TM2–TM3 loop and a calmodulin-binding domain (CBD, ▢) adjacent to TM1). Following ER Ca^{2+} store depletion, Ca^{2+} unbinds from STIM, STIM proteins oligomerize and migrate to ER–PM junctions, and induce accumulation of ORAI channels as tetramers. One STIM dimer is shown binding to one ORAI tetramer to induce Ca^{2+} influx. Oligomeric STIM proteins are present at puncta and the binding stoichiometry may vary.

truncated C-terminal portion of STIM1 or dStim is sufficient, when expressed as a cytosolic protein, to constitutively activate native CRAC current, as well as overexpressed ORAI or TRPC channels, bypassing the normal requirement for Ca^{2+} store depletion. Recent studies have narrowed down the minimal activating sequence of STIM1 to approximately 100 amino acids spanning the distal coiled-coil domain. Known by several acronyms and here termed the CAD; overlapping with fragments termed SOAR, OASF or CCb9; see Figure 2), this fragment of STIM1 was shown to activate the Orai1 current and to bind directly to both the N- and C- termini of purified Orai1 protein. Positively charged residues within CAD likely interact with negatively charged ones in the C-terminus of Orai1 as an essential aspect of channel gating. A recent study confirmed that only STIM and ORAI are required for channel activation, as recombinant STIM1 activates Ca^{2+} flux in yeast vesicles expressing human Orai1 in the absence of other proteins related to SOC channel function. However, the STIM–ORAI interaction may be modulated by other protein partners. For example, the EF-hand protein CRACR2A (CRAC regulator 2A) has been identified as a novel cytosolic Ca^{2+} sensor that forms a ternary complex with Orai1 and STIM1, stabilizing CRAC channels in T cells. Elucidating the exact molecular mechanism by which STIM activates ORAI (and TRPC) channels remains a challenging problem.

Inactivation and Store Refilling: Signal Dampened and Terminated

SOCE is normally terminated by the reuptake of Ca^{2+} into the ER, which may take a few minutes. However, on the

millisecond timescale, Ca^{2+} influx through the CRAC channel results in inactivation of CRAC current, thus providing fast and local negative feedback. Inactivation results in smaller, but not completely diminished, Ca^{2+} currents. Long before the molecular identity of the CRAC complex was known, researchers predicted the existence of a Ca^{2+} -binding site within nanometers of the PM that would bind Ca^{2+} and, thus, induce fast Ca^{2+} -dependent inactivation. Recent studies identified structural components of STIM1 and Orai1 that play a role in mediating fast inactivation. For one, an inactivation domain of STIM1 has been defined. Directly downstream of the CAD domain, a short stretch of STIM1 bears several negative charges that are required for fast, Ca^{2+} -dependent inactivation. Two intracellular parts of Orai1 have also become implicated. A cytosolic loop between the second and third transmembrane segments of Orai1 may occlude the intracellular side of the pore in order to inactivate CRAC current. Also, a calmodulin-binding domain at the N-terminal of Orai1, very close to the PM, binds calmodulin in the presence of Ca^{2+} . Perhaps incoming Ca^{2+} ions bind to this anchored calmodulin to initiate fast Ca^{2+} -dependent inactivation. The CRAC current completely subsides following refilling of the ER Ca^{2+} store, the breakdown of STIM1 puncta, and STIM1 moving out of the ER–PM junctions. This process may be accelerated when $[\text{Ca}^{2+}]_i$ is elevated.

STIM1 and TRPC Channels

The search for the molecular determinants of the CRAC channel took many years and several candidates were proposed, mostly the TRP channel family members. Moreover, the biophysical properties of store-operated current recorded in various cell

types were different. For example, unlike classical CRAC current in immune cells, the TRPC channels are relatively nonselective, have high single-channel conductance, and exhibit straighter current–voltage relationships. With the discovery of STIM1 as an ER Ca^{2+} sensor the problem started to resolve. Indeed, if STIM1 interacts with Orai1, can it also interact with TRPC channels? One group showed that STIM1 can activate TRPC1, C2, C3, C4, C5, and C6 (but not C7) via electrostatic interaction. Two conserved negatively charged aspartates of TRPC1 ($^{639}\text{DD}^{640}$) or TRPC3 ($^{697}\text{DD}^{698}$) interact with a positively charged lysines $^{684}\text{KK}^{685}$ located in the polybasic tail of STIM1 resulting in channel activation. Also, the interaction with STIM1 facilitates TRPC1 localization into lipid raft compartments of the PM, switching the receptor-operated mode of the channel to the store-operated one. Direct interaction of TRPC channels with Orai1 is widely discussed in the literature, although it remains unclear and requires further investigation.

STIM in Health and Disease

Genetic studies in knockout mice and human immunodeficiencies are providing new insights into the physiological importance of the STIM family of proteins. Suppression of STIM1 expression, and to a lesser extent STIM2, has devastating effects *in vivo*. *Stim1*^{−/−} mice perished *in utero* or soon after birth of respiratory failure, although early embryos appeared to develop normally. *Stim2*-deficient animals survived after birth but for only a short period of time of 4–5 weeks. The physiological importance of STIM1 and STIM2 for immune cells function has been extensively confirmed at the cellular level using cells isolated from mice lacking STIM1, STIM2, or both. In primary T cells, deletion of *Stim1* resulted in a profound reduction of SOCE, and CRAC currents were nearly abolished. This phenotype is less pronounced in *Stim2*-deficient T cells, yet effects on cytokine production were profound in both *Stim1*- and *Stim2*-knockout T cells. Mice with a T-cell-selective deletion of both *Stim1* and *Stim2* developed a lymphoproliferative disorder due to impaired function of regulatory T cells. Mast cells derived from *Stim1* knockout mice also display a reduced SOCE leading to a strong impairment of acute degranulation and cytokine production by these cells. In heterozygous *Stim1*^{+/-} mice anaphylactic responses were mildly reduced suggesting that the CRAC channel is essential for mast cell function *in vivo* and is a potential drug target for the treatment of asthma. A role of STIM1 in platelets relating to cerebrovascular function has also been described. Platelets from *Stim1*-deficient mice have a markedly reduced agonist-induced SOCE and their activation and aggregation was reduced. Of potential clinical significance, bone-marrow chimeric mice with *Stim1* deficiency in hematopoietic cells, including *Stim1*-deficient platelets, are significantly protected from arterial thrombosis and ischemic brain infarction but have only a mild bleeding time prolongation. On the other hand, mice expressing a constitutively active EF-hand mutant of STIM1 had macrothrombocytopenia and an associated bleeding disorder. STIM1 gain of function increased the resting basal intracellular Ca^{2+} levels in platelets causing premature platelet activation and bleeding. Finally, macrophages lacking STIM1 cannot activate FcγR-induced Ca^{2+} entry and phagocytosis. As a direct consequence, *Stim1* deficiency results in resistance to experimental immune thrombocytopenia

and anaphylaxis, autoimmune hemolytic anemia, and acute pneumonitis.

Consistent with the ubiquitous expression of STIM proteins and their involvement in a broad array of cellular processes, *Stim* gene knockouts in mice lead to phenotypic manifestations outside the immune system. Neonatal lethality in *Stim1*^{−/−} mice is associated with growth retardation and a severe skeletal myopathy with muscle that fatigues rapidly. Collectively, these studies show the instrumental role of the STIM-mediated Ca^{2+} signaling in T lymphocytes, mast cells, skeletal muscle, platelets, and macrophages for these cells to accomplish their function *in vivo*.

Valuable insights into STIM function in animal models have recently been extended to the characterization of an immune deficiency syndrome due to mutation of STIM1 identified in a human family showing how aberrant STIM-mediated SOCE can translate into human disorders. This first described human ‘STIM-opathy’ arises from a frameshift nonsense mutation in exon 3 of *Stim1* resulting in a premature STOP codon and termination of STIM1 protein. STIM1 protein is undetectable in these patients resulting in a complex syndrome of combined immunodeficiency with infections and autoimmunity associated with hepatosplenomegaly, hemolytic anemia, thrombocytopenia, and reduced numbers of regulatory T cells. In addition, patients exhibited muscular hypotonia and an enamel dentition defect, similar to patients with immunodeficiency due to mutations in Orai1 or mouse models lacking STIM1 or Orai1.

Conclusion

SOC channels formed of ORAI or TRPC subunits are activated by the Ca^{2+} -sensing STIM proteins. The major properties of STIM were discussed: ER Ca^{2+} sensing followed by oligomerization, translocation to the ER–PM junctions, organization of ORAI (or TRPC) channels into PM clusters, and activation of SOC channels. Besides molecular mechanisms of SOC channel activation, the physiological importance of STIM was highlighted: *Stim1* null mice die *in utero* and patients with mutation in human STIM1 resulting in premature STOP codon manifest the severe combined immunodeficiency syndrome and muscular and enamel dentition disorders. Finally, additional functions of STIM, not discussed here, have been identified recently, including the activation of adenylyl cyclase resulting in store-operated cyclic adenosine monophosphate production, ER remodeling through EB1 binding, and inhibition of voltage-gated Ca^{2+} channels. Despite rapid progress in the SOC channel field, several key mechanistic questions remain. How does STIM translocate to the PM? What causes ORAI clustering? How does STIM open both ORAI and TRPC channels? Moreover, of biomedical significance, is STIM1 involved in tumorigenesis, and will it be an effective drug target for immunosuppression? Over the past 5 years, the pace of research on STIM and ORAI proteins has been extremely rapid.

See also: **Bioenergetics:** ORAI Store-Operated Calcium Channel; Calcium-Modulated Proteins (EF-Hand); **Signaling:** Phospholipase C; **Bioenergetics:** Inositol Trisphosphate and Calcium Signaling; IP₃ Receptors; Transient Receptor Potential Channels.

Further Reading

- Cahalan MD (2009) Stimulating store-operated Ca^{2+} entry. *Nature Cell Biology* 11(6): 669–677.
- Cahalan MD and Chandy KG (2009) The functional network of ion channels in T lymphocytes. *Immunological Reviews* 231(1): 59–87.
- Feske S, Picard C, and Fischer A (2010) Immunodeficiency due to mutations in ORAI1 and STIM1. *Clinical Immunology* 135(2): 169–182.
- Hogan PG, Lewis RS, and Rao A (2010) Molecular basis of calcium signaling in lymphocytes: STIM and ORAI. *Annual Review in Immunology* 28: 491–533.
- Liou J, Kim ML, Heo WD, et al. (2005) STIM is a Ca^{2+} sensor essential for Ca^{2+} -store-depletion-triggered Ca^{2+} influx. *Current Biology* 15(13): 1235–1241.
- Luik RM, Wang B, Prakriya M, Wu MM, and Lewis RS (2008) Oligomerization of STIM1 couples ER calcium depletion to CRAC channel activation. *Nature* 454(7203): 538–542.
- Luik RM, Wu MM, Buchanan J, and Lewis RS (2006) The elementary unit of store-operated Ca^{2+} entry: Local activation of CRAC channels by STIM1 at ER-plasma membrane junctions. *Journal of Cell Biology* 174(6): 815–825.
- Penna A, Demuro A, Yeromin AV, et al. (2008) The CRAC channel consists of a tetramer formed by Stim-induced dimerization of Orai dimers. *Nature* 456(7218): 116–120.
- Picard C, McCarl CA, Papolos A, et al. (2009) STIM1 mutation associated with a syndrome of immunodeficiency and autoimmunity. *New England Journal of Medicine* 360(19): 1971–1980.
- Roos J, DiGregorio PJ, Yeromin AV, et al. (2005) STIM1, an essential and conserved component of store-operated Ca^{2+} channel function. *Journal of Cell Biology* 169(3): 435–445.
- Schindl R, Muik M, Fahrner M, et al. (2009) Recent progress on STIM1 domains controlling Orai activation. *Cell Calcium* 46(4): 227–232.
- Yeromin AV, Zhang SL, Jiang W, et al. (2006) Molecular identification of the CRAC channel by altered ion selectivity in a mutant of Orai. *Nature* 443(7108): 226–229.
- Zhang SL, Yu Y, Roos J, et al. (2005) STIM1 is a Ca^{2+} sensor that activates CRAC channels and migrates from the Ca^{2+} store to the plasma membrane. *Nature* 437(7060): 902–905.
- Zhou Y, Meraner P, Kwon HT, et al. (2010) STIM1 gates the store-operated calcium channel ORAI1 *in vitro*. *Nature Structural and Molecular Biology* 17(1): 112–116.

Structure and Regulation of Pyruvate Dehydrogenases

J L S Milne, National Institutes of Health, Bethesda, MD, USA

Published by Elsevier Inc.

Glossary

Allosteric regulation Modulation of enzymatic activity by noncovalent binding of a specific metabolite at a location other than the enzyme active site.

Cryo-electron microscopy Electron microscopic method to image at cryogenic temperatures frozen-hydrated and unstained molecules or macromolecular complexes. These entities often may not be amenable to analysis using X-ray crystallography or nuclear resonance spectroscopy. Projection views of molecules in the same orientation relative to the electron beam are averaged together; then, distinct projection views are computationally oriented and aligned to create a three-dimensional reconstruction.

Feedback inhibition A post-translational control mechanism in which the end product of a biochemical pathway accumulates and interacts with the regulatory control enzyme of the pathway to inhibit its activity.

Intermediary metabolism The enzyme-catalyzed reactions occurring within cells that are required to extract chemical

energy from nutrient molecules and use it to synthesize and assemble cellular components.

Nuclear magnetic resonance (NMR) Spectroscopic method to determine the three-dimensional structure of molecules, typically in suspension, by placing the sample in a strong magnetic field and using intermittent exposure to radiofrequency waves that cause absorption and emission of energy arising from changes in the spin states of the nuclei of atoms.

Oxidation Loss of electrons from an ion or molecule.

Prosthetic group A metal ion or other organic compound, other than amino acid residues that binds covalently to a protein and is essential for its biological function.

Reduction Gain of electron by an ion or molecule.

X-ray crystallography Determination of three-dimensional atomic structures by analysis of the X-ray diffraction patterns that result when X-rays pass through a well-ordered crystalline array of molecules.

Introduction

Pyruvate dehydrogenases (PDHs) represent a cornerstone in cellular energy metabolism, linking glycolysis and the metabolism of branched chain amino acids to the citric acid cycle and lipogenesis. They belong to the family of 2-oxo acid dehydrogenases, and catalyze the irreversible oxidative decarboxylation of pyruvate to yield acetyl-coenzyme A (CoA), CO₂, and a reducing equivalent NADH. Acetyl-CoA feeds into the citric acid cycle to produce adenosine triphosphate (ATP) and additional reducing equivalents, which are used to generate additional ATP when energy is required during oxidative phosphorylation. Alternatively, acetyl-CoA is used during the synthesis of fatty acids when fuel storage is possible. The requirement of both bacterial and eukaryotic cells to maintain appropriate energy balance to control the tremendous flux of carbon units through these pathways necessitates stringent regulation of PDH activity as fuel sources peak and wane. This is accomplished through small metabolites, through transcriptional regulation of constituent enzymes, and, in higher organisms, through associated kinases and phosphatases that regulate the phosphorylation state of the enzyme complex in a tissue-specific manner. X-ray crystallographic and nuclear magnetic resonance (NMR) analyses of individual subcomponents, combined with cryo-electron microscopy of the intact multienzyme complexes, have begun to yield new insights into the mechanism and regulation of these fascinating cellular machines.

Components of Pyruvate Dehydrogenases

PDHs are large complexes with sizes ranging from 4 to 10 MDa and diameters ranging from 300 to 500 Å. Each complex contains multiple copies of three essential enzymes – E1, E2, and E3 – organized into complexes with octahedral or icosahedral symmetry, depending upon the organism. Multiple cofactors are critical to its function: thiamin diphosphate (ThDP), CoA, flavin adenine dinucleotide (FAD), and nicotinamide adenine dinucleotide (NAD), which are derived from vitamins thiamin, pantothenate, riboflavin, and niacin, respectively, and lipoic acid, used to form a covalently linked prosthetic group. The hallmark of PDH is a mobile swinging arm domain that orchestrates substrate channeling of covalently bound reaction intermediates produced during the multistep reaction to couple the active sites of the E1, E2, and E3 enzymes. Such structural integration serves to enhance reaction rates of a complex reaction while minimizing undesirable side reactions with other cellular metabolites, such as free lipoic acid.

The PDH of *Bacillus stearothermophilus* is the simplest representative of the icosahedral PDHs that are found in Gram-positive bacteria and in the mitochondrion of eukaryotic cells. It contains 60 copies of E2, a dihydrolipoyl acetyltransferase of 46 kDa; 42–48 copies of E1, a ThDP-dependent pyruvate decarboxylase that is an $\alpha_2\beta_2$ heterotetramer of 154 kDa; and 6–12 copies of E3, a FAD-dependent dihydrolipoyl dehydrogenase that is a homodimer of 98 kDa. E2 is a multifunctional protein

that has three domains. At the two ends of the molecule are a C-terminal, 27-kDa catalytic domain, and an N-terminal, 9-kDa lipoyl-binding domain, with a lipoyl lysine prosthetic group protruding from the apex of a tight β -turn. Between these two domains is a 4 kDa peripheral subunit-binding domain to which E1 or E3 bind tightly and mutually exclusively at partially overlapping sites. The three domains are each separated by a stretch of amino acids that are thought to form flexible linkers. Sixty copies of the E2 enzyme self-associate through the catalytic domain to form the icosahedral core of the complex that is decorated on the periphery by E1 and E3 enzymes, randomly distributed on the surface of the complex (Figure 1(a)). PDHs of Gram-negative bacteria have an analogous architecture although they consist of an octahedral assembly of 24 E2 subunits that have three lipoyl domains and bind E1 and E3 enzymes, both of which are α_2 homodimers.

Mammalian icosahedral PDHs are more complicated than their bacterial counterparts. In addition to the E2 molecules, their core scaffold contains between 12 and 20 copies of

E3-binding protein (E3BP), a polypeptide that has ~40% sequence identity to E2 but lacks the active site histidine amino acid residue that is essential for acetyltransferase activity of E2. E3BP is present in at least one of every four copies of E2 in the pentameric face of the dodecahedron. The mammalian E1 is an $\alpha_2\beta_2$ tetramer as in the case of *B. stearothermophilus*. E1 and E3 of mammalian PDHs bind selectively to the peripheral subunit-binding domain of E2 or E3BP, with affinities ranging from ~20 to 40 nM, respectively, and at occupancy levels that are lower than those seen in *B. stearothermophilus*. Mammalian PDH complexes also have two N-terminal lipoyl domains in E2. The outermost lipoyl domain acts as the swinging arm while the innermost associates noncovalently with regulatory kinases and phosphatases. These latter enzymes are generally present in low copy numbers, and modulate PDH activity by reversible phosphorylation of the E1 enzyme at three distinct sites.

Reaction Mechanism of PDH

Extensive biochemical analysis and X-ray crystallographic and NMR imaging studies on the individual components of PDH provide a wealth of information about the elegant series of reactions, shown in Figure 1(b), that underlie the conversion of pyruvate to acetyl-CoA and NADH, and the regeneration of the enzyme complex for subsequent rounds of catalysis.

E1 Reaction

E1, the rate-limiting enzyme, has two chemically equivalent active sites. These function by a flip-flop mechanism in which the active sites interact with each other while simultaneously performing different steps in the sequence of reactions catalyzed by E1. Each active site contains a bound ThDP that is held in a deep cleft by specific interactions with amino acid residues on both the α - and β -subunits of E1 and by a Mg^{2+} ion which forms a bridge between the protein and the pyrophosphate region of ThDP. The cofactor is positioned into a 'V-conformation' that places the N4' amino group of the pyrimidine ring adjacent to the C2 proton of the thiazolium ring. A proton associated with N1 of ThDP is transferred to a conserved glutamate residue on E1, permitting the extraction of a proton from the acidic C2 carbon by N4 to generate a carbanion intermediate. This highly reactive anion performs a nucleophilic attack on the ketone C2 carbon of pyruvate to yield a α -lactyl-ThDP intermediate that releases carbon dioxide to produce enamine-ThDP. Reductive acetyl transfer then occurs between this intermediate and the lipoyl-lysine prosthetic group of E2. The lipoate S6 position is first protonated by a basic amino residue in the active site of E1. A second carbanion is subsequently generated and the acetyl group and two electrons are transferred to the swinging arm of E2 by nucleophilic attack of the enamine on the S8 position of the lipoate. Finally, transfer of a proton to the N4' atom of ThDP resets the E1 active site to enable further catalysis.

E2 Reaction

X-ray crystallographic analysis of the octahedral form of the E2 catalytic core domain indicates that each monomer is tightly

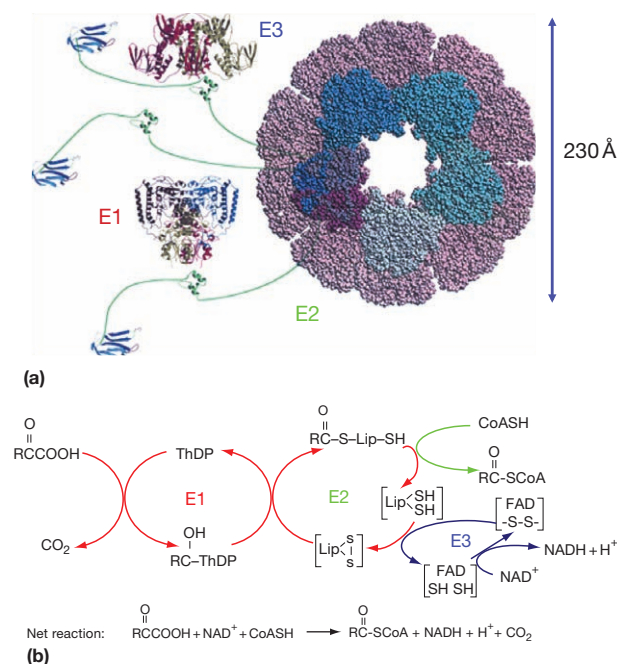


Figure 1 Components and reaction mechanism of pyruvate dehydrogenase (PDH). (a) Schematic model of the icosahedral PDH. One E2 catalytic trimer of the icosahedral inner core is highlighted with its cognate outer lipoyl domains (blue), and peripheral subunit-binding domains (green) bound to an E3 homodimer (top) or to an $\alpha_2\beta_2$ E1 heterotetramer (bottom). The atomic structures of the three E2 domains and the E3 homodimer are from *B. stearothermophilus*, while the E1 complex is from *Pseudomonas putida*. The green linker regions are unstructured. (b) Diagram of the reaction mechanism of a 2-oxo acid dehydrogenase multienzyme complex. R is CH_3 in the case of the PDHs; ThDP, thiamin diphosphate; Lip, lipoic acid. (a) From Perham RN (2000) Swinging arms and swinging domains in multifunctional enzymes: Catalytic machines for multistep reactions. *Annual Review of Biochemistry* 69: 961–1004. (b) Reproduced from Milne JLS, Shi D, Rosenthal PB, et al. (2002) Molecular architecture and mechanism of an icosahedral pyruvate dehydrogenase complex: A multifunctional catalytic machine. *EMBO Journal* 21: 5587–5598, with permission.

associated with two other subunits related by threefold rotational symmetry. Like chloramphenicol acetyltransferase, a trimeric enzyme, each active site of E2 is located in a long 30 Å channel formed at the subunit interface that has a central conserved His amino acid residue that acts as a general base and two entrances for substrates. CoASH, the reduced form of CoA, enters from the interior side of the complex during disruption of a salt bridge between an Asp/Asn dyad found in the ground state of the enzyme. The acetylated lipoyl group, transported ~50–100 Å from the active site of E1 by the swinging arm, enters from the external side of the inner E2 core. Rotation of the Asn side chain positions it to hydrogen-bond with the central His residue, increasing the nucleophilicity of the His N ϵ 2 atom, which draws the hydrogen atom from the reactive sulfur of CoA, positioned 3.2 Å away. This reactive group then attacks the reactive S8 atom of the acetyl substrate and the tetrahedral intermediate is stabilized by a single hydrogen bond donated by a hydroxyl group of an adjacent Ser residue. Rearrangement of this intermediate completes the *trans*-esterification reaction in which the –SH group of CoA replaces the S-acetyl group of E2 to form acetyl-CoA and the reduced lipoamide group, both of which exit the active site cleft. Rotation of the Asn side chain to its ground state position completes the cycle.

E3 Reaction

E3 promotes the transfer of two hydrogen atoms from the reduced lipoamide group of E2 to the FAD prosthetic group of E3, restoring the oxidized form of the lipoyl-lysine group of E2 and ultimately leading to the production of NADH and a free proton. Each E3 monomer has two active sites and four distinct domains: a FAD-binding domain, an NAD-binding domain, as well as central and interface domains. Each active site is formed at the junction of the two subunits with binding pockets for NAD and dihydrolipoamide located on opposite sides of a centrally located flavin, which binds to each chain in an extended conformation to intersect the channel. The redox center of the enzyme consists of a reactive intramolecular disulfide bridge that couples with the flavin rings. Upon dihydrolipoamide binding, the first reductive half reaction occurs whereby electrons are transferred from dihydrolipoamide to open the reactive disulfide bridge, permitting one of the cysteine residues to form a charge transfer complex with the flavin. This half reaction is initiated by abstracting a proton from the substrate by a conserved histidine residue that interacts directly with the bound flavin. The FADH₂ form of E3 is required for binding of NAD⁺, which orients so that its nicotinamide rings are closely juxtaposed with the isoalloxazine rings of the flavin. The second oxidative half reaction then occurs, in which the reduced FADH₂ group transfers a hydride ion to NAD⁺ giving rise to a free proton and NADH, which enters into the respiratory chain to generate ATP. The reactive disulfide bridge of E3 is also regenerated during this half reaction.

Lipoyl Domain in Substrate Channeling and Active Site Coupling

Lipoyl domains in PDH complexes are highly conserved, flattened β -barrel motifs that attach the lipoyl group to the N⁶

amino group of an E2 lysine residue located at the apex of a tight β turn. This architecture creates a long flexible arm, containing the hydrocarbon chains of lipoate and of the lysine residue, that further increases the reach of the linker swinging arm and facilitates entry into the active site channels of E1, E2, and E3. The latter two enzymes have low substrate specificity for lipoate, and are able to function catalytically using free lipoate or other lipoate derivatives. The activity of E1, in contrast, is more than four orders of magnitude higher in the presence of the lipoyl domain corresponding to its cognate E2 as compared with free lipoamide, lipoylated peptides derived from the lipoyl domain, or with lipoyl domains from other E2 species. The lipoyl domain must be tethered to the complex for this effect, resulting in high (millimolar range) local concentrations around the E2 core. Mapping of specific interactions between E1 and its cognate lipoyl domain of E2 using NMR analysis indicates that a significant number of amino acid residues in the lipoyl domain interact with E1, provided both from a prominent surface loop near the lipoyl lysine and amino acids that form the lipoyl lysine β turn residues adjacent to the prosthetic group. Mutation of these sites impairs the E1 reaction. Far fewer interactions occur between the lipoyl domain and E2 and E3. Stringent control of the E1 is an effective mechanism to promote substrate channeling, ensuring that the acyl transfer reaction occurs efficiently without interference from nonproductive substrates such as lipoate. Although there is less specificity in the interactions between E2 and the lipoyl domains, the amino acids on the surface of the lipoyl domain tend to be acidic, while those on the external face of the E2 catalytic domain trimer tend to have patches of basic residues. Thus, nonspecific electrostatic interactions may help guide the trajectory of the lipoyl domain to the active site of E2. Specific interaction of the lipoyl domain with regulatory kinases and phosphatases in eukaryotic cells is also critical for modulation of PDH activity (see below).

Regulation of the PDH

The critical role of PDH at a central branch point in cellular metabolism is reflected by the multiple layers of regulation, often combined in a highly cell-specific manner, that control the precise flux of carbon between pyruvate and acetyl-CoA in response to cellular bioenergetic and biosynthetic requirements. In both bacterial and eukaryotic cells, rapid modulation of PDH activity can be accomplished by small metabolites and ions interacting directly with the complex, while long-term regulation occurs by transcriptional control of one or more of the constituent enzymes. Tissue-specific modulation of the phosphorylation state of PDH complexes by kinases and phosphatases is a key mode of regulation in many eukaryotic cells, and is especially well characterized in mammalian cells. This type of regulation is absent in bacteria. The kinases and phosphatases are themselves subject to allosteric and transcriptional regulation.

Biological Significance of PDH Regulation

PDH regulation in mammals effectively integrates the intermediary metabolism of glucose, amino acids, and lipids under a variety of nutritional and physiological states. For example,

suppression of PDH activity occurs upon ingestion of a high-fat/low-carbohydrate diet or during starvation, which conserves precursor compounds (pyruvate, lactate, and alanine) that are used to synthesize glucose and which causes most body tissues to rely on fatty acids or ketones for fuel. This spares the limited glucose supply for preferential use by neuronal tissues including the brain. Conversely, muscle movement and increased cardiac output during exercise require activation of the PDH complex, as does fatty acid or sterol biosynthesis, which use acetyl-CoA as a substrate. Dysfunction of PDH has long been appreciated in maple syrup disease, lactic acidosis, and primary biliary cirrhosis as well as, more recently, in diabetes, cancer, heart disease, and neurodegeneration. The etiology of many of these diseases arises from specific genetic mutations in E1, E2, or E3, while others result from abnormal regulatory control of PDH by the kinases and phosphatases.

Small Metabolite Regulation of PDH

PDH activity is modulated by ratios of ATP/ADP, NADP/NAD⁺, and acetyl-CoA/CoA, and by phosphoenolpyruvate, pyruvate, and ions, that act as indicators of the metabolic state of the cell. Inhibition of the complex occurs when the cell is in an energy-rich metabolic state, reflected by high concentrations of ATP, NADH, and acetyl-CoA. When the concentration of these metabolites declines, and levels of AMP, CoA, and NAD⁺ rise, activation of the complex occurs. Atomic X-ray structures of the E2 catalytic core and the E3 dihydrolipoyl dehydrogenase reveal that acetyl-CoA and NADH are product inhibitors, because they compete for the binding sites of CoA and NAD⁺, respectively. Other metabolites influence the phosphorylation state of the PDH as discussed below.

PDK Regulation of PDH

There are four PDH kinase isoforms (PDK1–4) that regulate the activity of PDH by modulating the phosphorylation state of E1. These are not related to cytoplasmic Ser/Thr/Tyr kinases, but are distantly related to protein histidine kinases. Upon activation, these kinases undergo autophosphorylation on a conserved histidine residue before transfer of the high-energy phosphate to the acceptor substrate protein. There is considerable variation in the kinetic parameters and regulation of the different PDKs, as well as in their amino acid sequences, although each isoform is highly conserved between mammalian species. PDK1 is present in heart, skeletal muscle, and pancreatic islets. PDK2 is ubiquitously expressed in mammalian tissues and is particularly sensitive to metabolic signals for energy demand and fuel source. It is expressed at higher levels during starvation and diabetes. PDK3 has the highest specific activity of all of the PDKs. It is typically expressed at low levels in most tissues except in testis, kidney, and brain. PDK4, which is expressed at higher levels during starvation and diabetes, is normally expressed at low levels except in heart, skeletal muscle, and pancreatic islets and has the highest basal activity among the PDKs. Insulin represses PDK4 expression, but loss of this regulation leads to elevated blood glucose levels present in insulin-unresponsive patients. Thus, multiple isoforms are expressed in specific tissues simultaneously. Different isoforms can form functional heterodimers that have different kinetic

and regulatory properties than those of the parent homodimers. These add to the complexity of regulation, especially when taken together with differential expression of the PDH phosphatases PDP1 and PDP2. Hormonal regulation and a number of disease states, other than diabetes, can also alter PDK expression significantly. For example, PDK3 and PDK1 are overexpressed in hypoxic cancer cells by the actions of the transcription factor hypoxia-inducible factor-1 (HIF-1). This shifts cellular metabolism to generate high levels of lactate through the activation of glycolysis and the reduction of mitochondrial oxidative pathways (Warburg effect). The PDPs often show opposite patterns of expression to those of the PDKs, thus further amplifying the regulatory effects on PDH activity. Both the PDKs and PDPs are attractive targets for drug design. A number of compounds, including the nonhydrolyzable pyruvate analog dichloroacetate (DCA), have been considered as potential therapeutic agents for diabetes, lactic acidosis, myocardial ischemia, and cancer.

E1 activity can be inhibited by phosphorylation of any of three target Ser residues on its α -subunit. The PDK isoforms show significant variation in their specificity toward individual phosphorylation sites: each can phosphorylate Ser 264 (site 1) and Ser271 (site 2), albeit with different rates, but only PDK1 modifies Ser203 (site 3). These phosphorylation sites are located on two adjacent loops that are close to the active site of E1. Structural studies of human E1 indicate that phosphorylation of site 1, which blocks the majority of E1 activity, disorders two loops adjacent to the E1 active site and disrupts the interaction of ThDP and the E2-bound lipoyl group with their respective binding sites on E1 (**Figure 2(a)**). Phosphorylation of sites 2 and 3 occurs more slowly and is thought to hinder reactivation of E1, primarily by PDP2.

The influence of small molecule and ion regulators on the kinases is complex, although many work by modulating interactions of the enzyme with the inner lipoyl domains of E2. NADH and acetyl-CoA cause the reduction and acetylation of the E2-associated lipoyl groups through reversal of the reactions normally catalyzed by E3 and E2. The modified lipoyl domains (L2) can bind each PDH kinase isoform with a gradation of affinities PDK3 > PDK1 ~ PDK2 > PDK4. Binding to L2 situates each kinase in close proximity to the E1 substrate, and modestly increases activity of PDK2 and PDK3. PDK3 alone is activated strongly by L2 domains associated with the native E2 core or free in solution, indicating that it is also allosterically activated by L2. PDK1 and PDK4, with high intrinsic basal activities, show marginal effect of L2 association. However, for all isoforms, this interaction may be crucial for the movement of limited numbers of kinases around the E2 core. Studies on PDK2 indicate that in the presence of potassium phosphate buffer, ADP and pyruvate (or DCA) slow ADP dissociation and strengthen product inhibition while markedly reducing L2 binding and shifting the dimer to an inactive tetrameric state. Similar effects are seen in PDK1 and PDK4, but not in PDK3, which is not regulated by pyruvate.

X-ray crystallographic analysis of multiple PDK isoforms in the absence or presence of key regulatory compounds provides a satisfying glimpse into the alterations in active site geometry that underlie the action of many of these regulators. Each monomer has two domains of similar size. The N-terminal regulatory domain is a four-helix bundle that binds the lipoyl

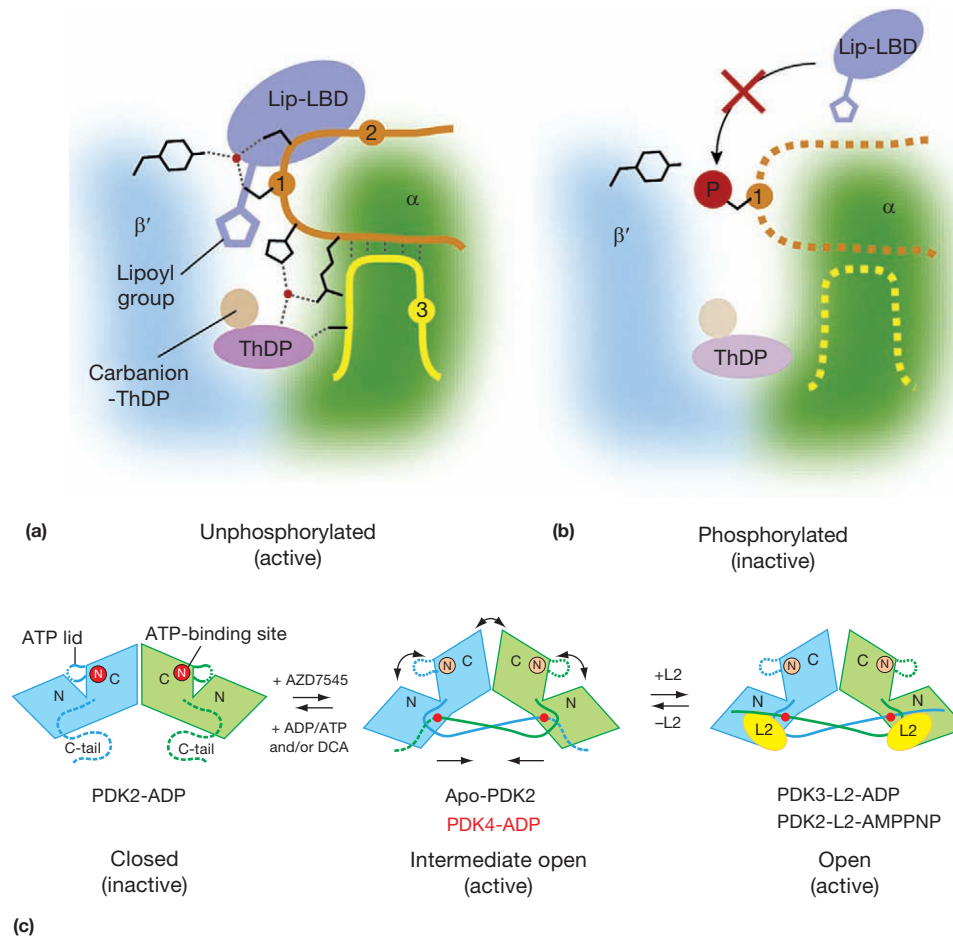


Figure 2 Inactive and active conformations of E1 and PDK. (a) Binding of ThDP (magenta) to the active site of E1 induces ordering of the phosphorylation loops A (orange line) and B (yellow line) which enables the decarboxylation of pyruvate (tan) and the reductive acetylation of the lipoyl group attached to the lipoyl-binding domain, Lip-LBD. Numbered circles represent the three phosphorylation sites; hydrogen bonds are shown by dotted black lines; a water molecule is indicated by a red dot; the E1 α - and β -subunit are green and cyan, respectively. (b) Phosphorylation of site 1 (Ser 264) in human apo-E1 blocks ordering of the phosphorylation loops, reduces the binding of ThDP, abolishes the hydrogen-bonding network that is required to order the loops and blocks access of the lipoyl substrate, resulting in inactivation of E1 catalysis. Disordered phosphorylation loops are indicated by yellow and orange dashed lines. (c) Closed (left), intermediate open (middle), and open (right) conformations of PDK isoforms. In the presence of high-affinity ADP binding (red circled N), PDK2 has a closed active site cleft, with fully disordered C-terminal tails. Apo-PDK2 or PDK4-ADP structures are in an active, intermediate open conformation which has an open active site cleft with low-affinity nucleotide-binding sites (pink circled N) and C-terminal tails that interact with the other subunit using the DW motif anchor sites (red circles), although the terminal amino acids of the tail remain disordered. The active open conformation evident in the PDK3-L2-ADP and PDK2-L2-AMPPNP structures has an open active site and fully ordered C-terminal tails which cross-anchor the subunits and help form the L2-binding domain. Molecules that induce transitions between the three states are indicated above and below the arrows. AZD7545 is a dihydrolipoamide mimetic. The PDK subunits are shown in blue and green, with the N- and C-terminal domains indicated. Solid and dashed lines indicate ordered and disordered regions of the protein, respectively. L2 is shown in yellow. (a, b) Reproduced from Kato M, Wynn RM, Chuang JL, et al. (2008) Structural basis for inactivation of the human pyruvate dehydrogenase complex by phosphorylation: Role of disordered phosphorylation loops. *Structure* 16: 1849–1859, with permission. (c) From Wynn R, Kato M, Chuang JL, Tso S-C, and Chuang DT (2008) Pyruvate dehydrogenase kinase-4 structures reveal a metastable open conformation fostering robust core-free basal activity. *Journal of Biological Chemistry* 283: 25305–25315.

group of L2, pyruvate, and other regulatory ligands. The C-terminal catalytic domain contains α -helical and β -sheet motifs, an ADP/ATP nucleotide-binding site where phosphoryl transfer occurs, a large dimerization interface that results from head-to-tail interactions of the β -sheets of the two subunits, and a C-terminal tail that modulates the switch from inactive to active states of the kinase. The active site of each monomer localizes to the interface between the regulatory and catalytic domains, where ATP/ADP binds to an ATP-binding fold unique to the GHKL superfamily of ATPase/kinases. This fold has an

associated structural loop element called an ATP lid, whose conformation is regulated by ATP hydrolysis, as well as by regulatory signals initiated and transmitted from distal regions of the kinase.

Multiple conformations of the enzyme have been identified (Figure 2(b)). In the inactive closed conformation, the C-terminal tail is disordered and does not interact with the other subunit. The ATP lid interacts with the N-terminal domain, preventing ADP dissociation to maintain product inhibition. L2 binding is required to obtain the active state in which the N-proximal region of the C-terminal tail containing

a conserved Asp-Trp (DW) motif of one subunit wrap around the other subunit and also interact with specific amino acids of the L2 domain. The C-terminal amino acid residues of one subunit are an integral component of the lipoyl-binding domain of the other subunit. These interactions widen both the active site cleft and the dimer interface by destabilizing the interaction of the ATP lid with the N-terminal domain. In turn, this facilitates exchange of ATP for bound ADP and likely also increases the accessibility of target serine residues of E1 to the γ -phosphate of ATP, which is transferred to the substrate serine once it has been activated by a conserved amino acid that acts as a general base during the phosphoryl transfer reaction. Other structural rearrangements have also been observed. For example, in PDK2, there is ordering of a loop that forms the back wall of the E1 binding site, which encompasses both the lipoamide binding pocket and adjacent amino acids that specifically interact with the L2 domain itself. An intermediate active state of PDK4 has been crystallized in the presence of ADP that has an open active site cleft and partially ordered C-terminal tails. The effective removal of ADP by PDK4 in this state requires the DW motif; its loss ablates the robust core-independent basal activity of PDK4.

An important mechanism by which signals are transmitted between the nucleotide-binding site and the L2-binding domain may be provided by a long central rod located near the two sites. This rod is composed of two short helices that link the two sites. A central pyruvate/DCA-binding site sits between the two helices (Figure 3). The geometry of the active site is influenced by its nucleotide occupancy, pyruvate, and L2 binding, and, surprisingly, by K^+ ions. Analysis of human PDK2 with bound DCA indicates that the β -phosphate of ADP chelates K^+ , together with backbone carbonyl groups of three amino acids, and an oxygen atom of a Ser hydroxyl group, to enhance ADP inhibition. Binding of lipoyl domain shifts the arrangement of the participating amino acid residues, disrupts the K^+ binding site, and enhances dissociation of ADP from the active site. Binding of K^+ ions at other sites on the protein stabilizes the L2-binding site and facilitates cross-talk

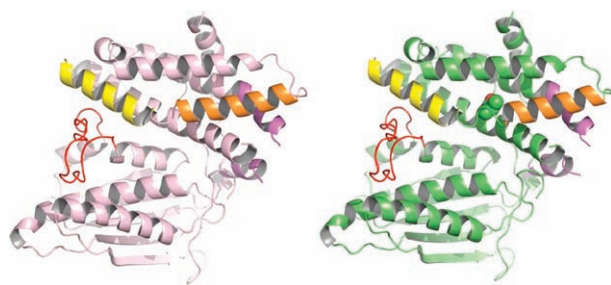


Figure 3 Binding of the pyruvate analog DCA induces conformational changes in PDK1. Apo-PDK1 (pink ribbon) undergoes small movements of helices 6 (orange) and 7 (yellow), and unwinding of a short linker region upon binding of DCA (green space-filling model). DCA-bound PDK1 is shown as a green ribbon. Movement of this helical rod may mediate cross-talk between the nucleotide-binding site and its associated ATP lid (red), the lipoyl-binding site (magenta), and the pyruvate/DCA-binding site. Reproduced from Kato M, Li J, Chuang JL, and Chuang DT (2007) Distinct structural mechanisms for inhibition of pyruvate dehydrogenase kinase isoforms by AZD7545, dichloroacetate, and radicicol. *Structure* 15: 992–1004, with permission.

between the nucleotide- and L2-binding sites. Although much less is known about their regulation, ions are also important regulators of PDP1 and PDP2. Both forms are activated by increasing concentrations of free intra-mitochondrial Mg^{2+} or Mn^{2+} that occur in response to declining levels of ATP. Increases in free mitochondrial Ca^{2+} activate PDP1 in tissues with hormonal or exercise-dependent increases in mitochondrial Ca^{2+} . The ion serves to bridge the interaction of PDP1 with L2 domain of E2, thereby stimulating phosphatase activity. The activities of PDK and PDP2 are regulated in reciprocal fashion by inositol phosphoglycans, compounds that chelate Mn^{2+} and Zn^{2+} metal ions and that are generated in response to stimulation by insulin.

Architecture of PDHs by Cryo-Electron Microscopy

While visualization and extensive biochemical analysis of the individual PDH components have yielded detailed information about the mechanism and regulation of the complex, additional possibilities for how the three enzymes integrate to form a highly efficient multifunctional catalytic machine have been provided by cryo-electron microscopic analysis. Two-dimensional projection images of the complex reveal the general structural arrangement of the central core decorated by peripheral subunits (Figure 4(a)) but the three-dimensional reconstruction of complexes from bacteria, yeast, and mammalian E2 cores in the absence or presence of peripheral E1 or E3 subunits indicates a surprising diversity in detailed structural arrangements of each type of complex.

In PDH subcomplexes of *B. stearothermophilus*, in which the E2 core is fully decorated with E1 or E3, the outer shell of peripheral subunits localizes ~ 75 – 100 Å above the E2 catalytic core (Figures 4(b) and 4(d)). This structural architecture may increase productive substrate channeling because the active sites of E1, E2, and E3 each face the annular gap. The swinging lipoyl domain may also preferentially localize to the gap, increasing its effective concentration, at least when the number of peripheral subunits bound to the complex is high. Three-dimensional snapshots of individual complexes obtained using cryo-electron tomography indicate that the peripheral enzymes localize well above the core, even in lightly decorated complexes although their arrangement deviates from the strict icosahedral symmetry present in the inner core. Similar architectural features are evident in the octahedral complexes present in *Escherichia coli*. In *B. stearothermophilus*, lateral movements of the subunits above the core can extend the effective reach of the swinging arm since a lipoyl domain associated with a specific E2 monomer potentially can contact up to 12 peripheral subunits. Radial movements of the outer enzymes toward the core, possibly through the transient formation of a type VI turn in the inner semiflexible linker, may maintain the peripheral subunits in close proximity to each other and the core to facilitate catalysis. This is particularly relevant for E3 given that few enzymes populate each complex. Yeast PDH also maintains E3 close to the core, albeit by more direct interactions with the pentagonal face of the inner E2 core.

Mammalian PDHs have different constraints than their bacterial counterparts: while E3 is more plentiful and potentially more symmetrically distributed around the E2 core, the

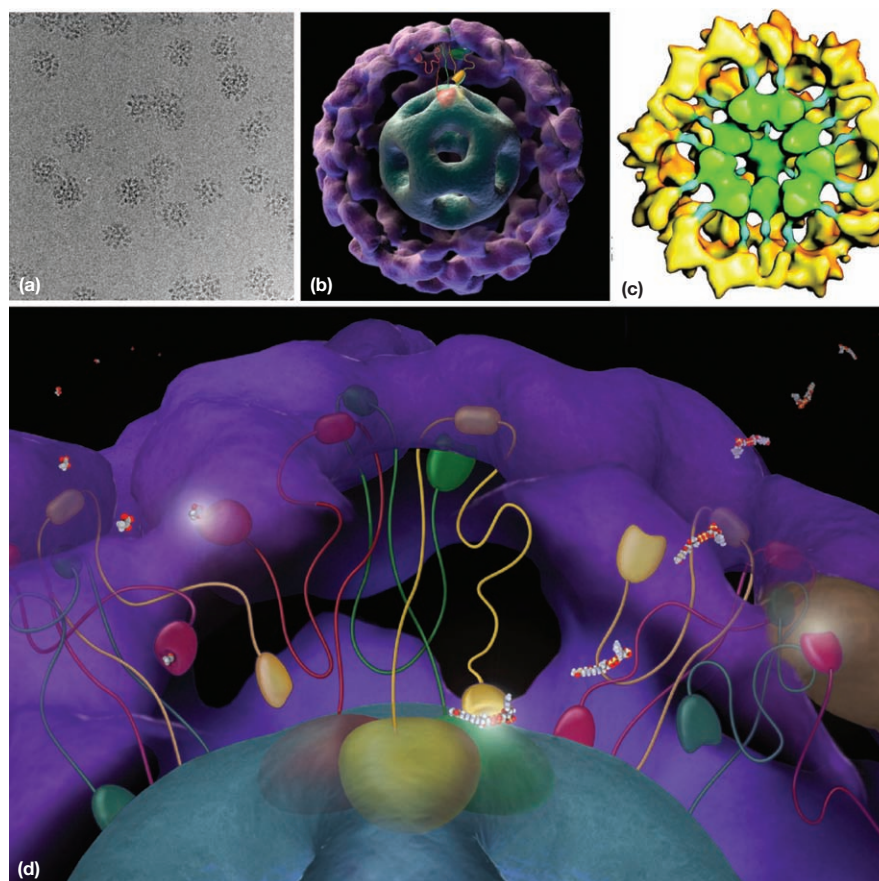


Figure 4 Architecture of icosahedral pyruvate dehydrogenases (PDHs) from cryo-electron microscopic analysis. (a) Cryo-electron microscopic images of a PDH subcomplex from *B. stearothermophilus* consisting of E2 icosahedral core decorated with 60 E1 enzymes are used to reconstruct a three-dimensional model (b). Sectioned schematic view of this model highlights three E2 molecules (colored red, green, and yellow) and shows the movement of the swinging E2 lipoyl domain in the annular region between the inner core (cyan) of E2 molecules and the outer shell of E1 molecules (purple). (c) Sectioned view of a three-dimensional reconstruction of the bovine PDH showing the E2 inner core (green), outer E1 shell (yellow), and rigid inner E2 linkers (blue). (d) Active site coupling between E1, E2, and E3 mediates the synthesis of acetyl-CoA from pyruvate. An acetyl group from pyruvate is covalently attached to the lipoyl domain at the active site of E1 (left-hand side) and is transported across the annular gap to an E2 active site (center) where it is transferred to CoA. The dithiolane ring of the lipoyl domain is then regenerated at an E3 active site (far right-hand side) present in the outer protein shell of the complex. The PDH complexes shown in panels a, b, and c are ~ 450 Å in diameter. (a, b) Reproduced from Milne JLS, Shi D, Rosenthal PB, et al. (2002) Molecular architecture and mechanism of an icosahedral pyruvate dehydrogenase complex: A multifunctional catalytic machine. *EMBO Journal* 21: 5587–5598, with permission. (c) Reproduced from Zhou H, McCarthy DB, O'Connor CM, Reed LJ, and Stoops JK (2001) The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes. *Proceedings of the National Academy of Sciences of the United States of America* 98: 14802–14807, with permission.

kinases and phosphatases that bind to the inner lipoyl domain prior to interaction with E1 must have clear accessibility to the interior of the complex. The inner linker that joins the inner E2 domain and the peripheral subunit-binding domain is more ordered in the region adjacent to the core. Upon E1 binding, the linker is thought to become even more rigid, since density for this region can be visualized in bovine PDH complexes (Figure 4(c)). The E2 monomers also integrate into a relatively dynamic central icosahedron that exhibits variation in size. These modifications likely enhance active site coupling in annular gaps that are more crowded than those found in bacteria, and promote interactions between E1 and the regulatory kinases and phosphatases. In addition, while both bacterial and eukaryotic enzymes are thought to transfer acetyl groups between adjacent lipoyl domains, the eukaryotic complexes additionally could require the movement of the

regulatory enzymes accomplished through coordinated interactions with both E1 and the lipoyl domains of E2 in a hand-over-hand transfer mechanism.

Conclusion

The biochemical and structural studies of PDH components over the last six decades have led to extraordinary insights into the functioning of one of the central players in cellular metabolism. Our present knowledge of PDH mechanism thus ranges from atomic resolution details of substrate interactions and channeling at the active sites of E1, E2, and E3 to the higher-order architecture and regulation by kinases and phosphatases. These studies have also produced fascinating glimpses of how PDH function is altered in response to changes in its chemical

environment and how transcriptional control of PDH and PDK genes *in vivo* may be used to modulate function. Integrating the knowledge derived from these analyses to predict the dynamic structural and functional aspects of PDH in the complex milieu of intact cells in healthy and diseased states is undoubtedly an exciting challenge for the future.

Acknowledgment

This research was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, USA.

See also: [Metabolism Vitamins and Hormones: Biochemistry of Thiamine and Thiamine Phosphate Compounds](#); [Carbohydrate Metabolism in the Central Nervous System](#); [Coenzyme A](#); [Fatty Acid Oxidation](#); [Fatty Acid Structure and Synthesis](#).

Further Reading

- Ciszak EM, Korotchkina LG, Dominiak PM, Sidhu S, and Patel MS (2003) Structural basis for flip-flop action of thiamin pyrophosphate-dependent enzymes revealed by human pyruvate dehydrogenase. *Journal of Biological Chemistry* 278: 21240–21246.
- Green T, Grigorian A, Klyuyeva A, et al. (2008) Structural and functional insights into the molecular mechanisms responsible for the regulation of pyruvate dehydrogenase kinase 2. *Journal of Biological Chemistry* 283: 15789–15798.
- Kato M, Li J, Chuang JL, and Chuang DT (2007) Distinct structural mechanisms for inhibition of pyruvate dehydrogenase kinase isoforms by AZD7545, dichloroacetate, and radicicol. *Structure* 15: 992–1004.
- Kato M, Wynn RM, Chuang JL, et al. (2008) Structural basis for inactivation of the human pyruvate dehydrogenase complex by phosphorylation: Role of disordered phosphorylation loops. *Structure* 16: 1849–1859.
- Lengyel JS, Stott KM, Wu X, et al. (2008) Extended polypeptide linkers establish the spatial architecture of a pyruvate dehydrogenase multienzyme complex. *Structure* 16: 93–103.
- Mattevi A, Obmolova G, Kalk KH, van Berkel WJH, and Hol WGJ (1993) Three-dimensional Structure of Lipoamide Dehydrogenase from *Pseudomonas fluorescens* at 2.8 Å resolution. Analysis of redox and thermostability properties. *Journal of Molecular Biology* 230: 1200–1215.
- Mattevi A, Obmolova G, Schulze E, et al. (1992) Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex. *Science* 255: 1544–1550.
- Milne JLS, Shi D, Rosenthal PB, et al. (2002) Molecular architecture and mechanism of an icosahedral pyruvate dehydrogenase complex: A multifunctional catalytic machine. *EMBO Journal* 21: 5587–5598.
- Patel MS and Korotchkina LG (2003) The biochemistry of the pyruvate dehydrogenase complex. *Biochemistry and Molecular Biology Education* 31: 5–15.
- Pei XY, Titman CM, Frank RAW, Leeper FJ, and Luisi BF (2008) Snapshots of catalysis in the E1 subunit of the pyruvate dehydrogenase multienzyme complex. *Structure* 16: 1860–1872.
- Perham RN (2000) Swinging arms and swinging domains in multifunctional enzymes: Catalytic machines for multistep reactions. *Annual Review of Biochemistry* 69: 961–1004.
- Roche TE and Hiromasa Y (2007) Pyruvate dehydrogenase kinase regulatory mechanisms and inhibition in treating diabetes, heart ischemia and cancer. *Cellular and Molecular Life Sciences* 64: 830–849.
- Stott KM, Yusuf AM, Perham RN, and Jones DD (2009) A surface loop directs conformational switching of a lipoyl domain between a folded and a novel misfolded structure. *Structure* 17: 1117–1127.
- Sugden MC and Holness MJ (1994) Interactive regulation of the pyruvate dehydrogenase complex and the carnitine palmitoyltransferase system. *FASEB Journal* 8: 54–61.
- Wynn RM, Kato M, Chuang JL, Tso S-C, and Chuang DT (2008) Pyruvate dehydrogenase kinase-4 structures reveal a metastable open conformation fostering robust core-free basal activity. *Journal of Biological Chemistry* 283: 25305–25315.
- Zhou H, McCarthy DB, O'Connor CM, Reed LJ, and Stoops JK (2001) The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes. *Proceedings of the National Academy of Sciences of the United States of America* 98: 14802–14807.

Structure–Function of the Cytochrome b_6f Complex of Oxygenic Photosynthesis

W A Cramer, Purdue University, West Lafayette, IN, USA

E Yamashita, Osaka University, Yamada-oka Suita, Japan

D Baniulis, Lithuanian Research Centre for Agriculture and Forestry, Kaunas, Lithuania

J Whitelegge, University of California, Los Angeles, CA, USA

S S Hasan, Purdue University, West Lafayette, IN, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

b_n , Q_n and b_p , Q_p Hemes, quinones bound on electrochemically negative or positive sides of membrane, the chloroplast stroma and lumen, respectively.

Cyclic electron transport Electron transport around photosystem I via plastoquinone and the cyt b_6f -complex which provides extra ATP.

Photosystem The total of pigments assembled around a reaction center.

Q-cycle A mechanism of proton translocating quinol oxidation by cyt bc-complexes in respiratory and photosynthetic electron transport, first formulated by Peter Mitchell in 1975.

Thylakoid Coined by Wilhelm Menke to denominate the structure of the inner membrane system of the chloroplast which in cross-section appears to be built from stacked little bags, but their inner space is connected in the third dimension.

Function

The oxygenic photosynthetic electron transport chain embedded in a lipid membrane transduces the light energy absorbed and transferred from light-harvesting pigments and reaction center complexes (~ 1.8 electron volts ≈ 40 kcal mol $^{-1}$ per red light photon) to free energy stored in the high-energy bond of adenosine triphosphate (ATP) and reducing equivalents stored in pyridine nucleotide (nicotinamide adenine dinucleotide phosphate (reduced), NADPH). The 220-kDa hetero-oligomeric cytochrome b_6f complex dimer is one of three multisubunit integral membrane protein complexes in the photosynthetic electron transport chain, transferring electrons from the H $_2$ O electron donor to the nicotinamide adenine dinucleotide phosphate (NADP) $^{+}$ acceptor. Through the coupling in the b_6f complex of electron transport to proton translocation, a trans-membrane proton electrochemical potential gradient ($\Delta\tilde{\mu}_{H^{+}}$) is generated as the intermediate high-energy intermediate to provide the free energy, ΔG_{ATP} , required for ATP production, discharging the proton gradient through the ATP synthase (Figure 1). With as many as four protons/ATP the ΔG_{ATP} is equal to the number of protons, n , transferred through the ATP synthase per ATP synthesized, that is, $\Delta G_{ATP} \approx n\Delta\tilde{\mu}_{H^{+}}$.

The other two hetero-oligomeric protein complexes in the oxygenic photosynthetic electron transport chain in chloroplast thylakoid or photosynthetic microbial membranes are the dimeric photosystem II (PSII) and trimeric photosystem I (PSI) reaction center complexes. The b_6f complex functions between the two reaction center complexes in the linear electron transport chain from H $_2$ O to NADP $^{+}$ (Figure 1). The complex accepts electrons and protons from the PSII complex via a pool in the membrane bilayer of lipophilic plastoquinone/ol, PQ/PQH $_2$, whose chemical structure consists of a substituted benzene ring connected to a carbon chain of nine isoprenoid groups, that is, 45 carbons. The electron affinity of PQ/PQH $_2$ ($E_{m7} = +100$ mV) is intermediate between that of the bound quinone, Q_A , in the PSII reaction center

and the two iron–two sulfur [2Fe–2S] cluster ($E_{m7} = +300$ mV) of the Rieske iron–sulfur subunit (ISP) in the b_6f complex. Transfer of one electron from PQH $_2$ to the [2Fe–2S] cluster, which generates the neutral semiquinone radical, PQH $^{\bullet}$, is coupled to transfer of a proton to a histidine ligand of the cluster, and then to the p-side aqueous phase, thereby contributing to the $\Delta\tilde{\mu}_{H^{+}}$. The [2Fe–2S] cluster donates an electron to the heme of cytochrome f (f , feuille, leaf) and to the PSI complex through the soluble copper-containing plastocyanin or cytochrome c_6 (Figure 1). The detailed pathways of electron and proton transfer responsible for electrogenic proton translocation and generation of the $\Delta\tilde{\mu}_{H^{+}}$ are centered on a Q cycle, discussed below, and shown in Figure 4.

Crystal Structures

The photosynthetic electron transport/proton translocation complexes shown in Figure 1 are integral membrane–protein complexes that can only be isolated and purified after extraction in detergent from the host membrane. The detergent(s) that allow purification of a given membrane–protein complex in an active and crystallizable state must be determined for each protein complex. Maltoside detergents have been used to purify the complex. In the initial stages of the crystallization studies of the complex, mass spectroscopy (electrospray) was essential in the development of the purification. The moderately thermophilic cyanobacteria, *Mastigocladus laminosus* and *Nostoc* sp. PCC 7120, and the green alga, *Chlamydomonas reinhardtii*, have been the sources for crystallization and crystallography studies of the b_6f complex. The purification of the complex from *C. reinhardtii* utilized affinity tagging/chromatography. A major gap in the structure studies of the b_6f complex is the absence of a structure from a plant source.

Each monomer of the b_6f complex contains 13 membrane-spanning α -helices in which seven prosthetic groups are embedded. The two monomers are arranged with a C2

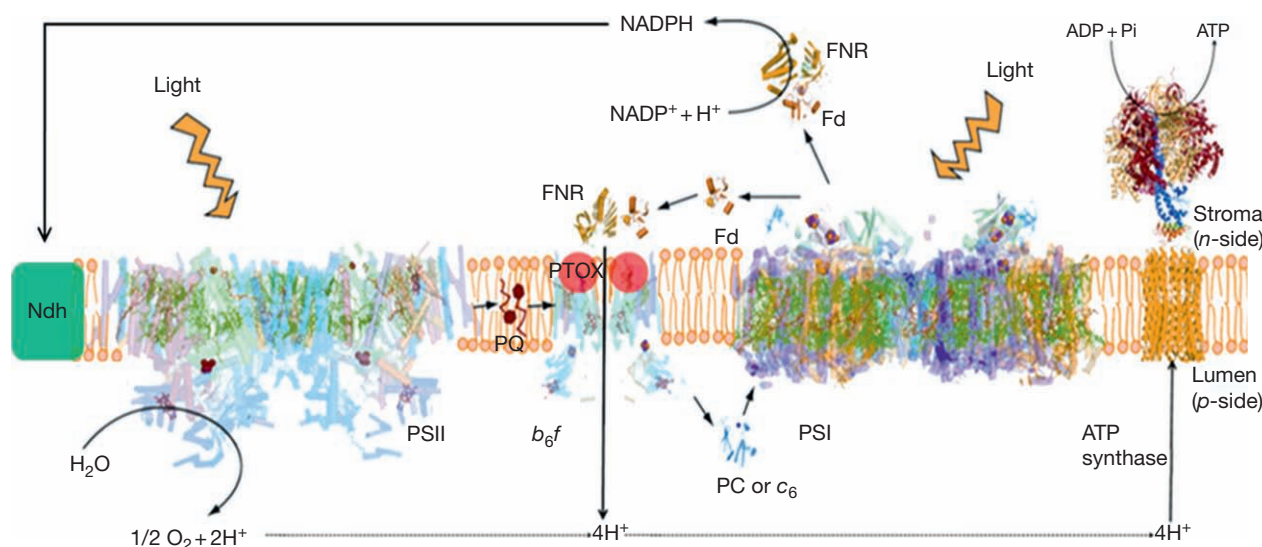


Figure 1 The electron transport/proton translocation chain of oxygenic photosynthesis showing structures of the membrane-embedded multisubunit energy-transducing complexes, except for the NDH-1 complex, not yet solved. Left to right, NDH-1, hetero-dimeric photosystem II reaction center, homo-dimeric cytochrome *b₆f* complex, trimeric photosystem I reaction center complex, and ATP synthase. Maximum $H^+/2e^-$ ratio = 2 for PSII and 4 for the *b₆f* complex; proton stoichiometry of protons utilized for ATP synthesis to ATP synthesized, H^+/ATP , = 4.

symmetry that is relevant for function. A schematic of the complex is shown in a stick model that shows the position of the seven prosthetic groups, which is a dimer with C2 symmetry (Figure 2). The monomer contains eight trans-membrane α -helical subunits, four ‘large’ subunits (cytochromes *f* and *b₆*, the Rieske iron–sulfur [2Fe–2S] protein (ISP), and subunit IV, with molecular weights between 18 and 32 kDa), and four small hydrophobic subunits (petG, L, M, and N), 3–4 kDa, each punctuated by 1–2 positive residues on the n-side, which can be characterized as hydrophobic sticks and span the membrane with one trans-membrane α -helix. These four small subunits form a picket fence-like structure on the periphery of each monomer. The molecular weights of the subunits of the *M. lamosus* complex, and the midpoint oxidation–reduction potentials of the redox prosthetic groups determined *in situ* (in membranes), are summarized (Table 1).

Prior to determination of the structure of the *b₆f* complex, the soluble domains of cytochrome *f* and the Rieske [2Fe–2S] protein were solved. It is of interest that both p-side (lumen) redox proteins are predominantly β -strand, as is the p-side plastocyanin. This secondary structure for cytochrome *f* is virtually unique for a c-type cytochrome. In this respect, and also in its primary amino acid sequence, cytochrome *f* and cytochrome *c₁* of the *bc₁* complex are completely different proteins. The only common feature is the characteristic five-residue Cys–X–Y–Cys–His heme-binding sequence that is diagnostic of c-type cytochromes. The 217 070 and 214 298 molecular weight (without prosthetic groups) *b₆f* complex from *M. lamosus* and *Nostoc*, respectively, is a symmetric dimer.

A salient feature of the structure, in which the 24-kDa cytochrome *b* and 18-kDa subunit IV forms the core, is the set of seven tightly bound prosthetic groups, of which five are bound redox groups (four hemes, cyt *f*, two noncovalently bound *b*-hemes and one covalently bound *c* heme, the unique heme *c_n* (Figure 3)) bound to the cyt *b* polypeptide. Together

with the [2Fe–2S] cluster, these groups define a Q-cycle pathway for quinol oxidation (Figure 4), in which the transfer of the two electrons derived from the oxidation of PQH₂ is split into low- and high-potential chains, respectively, transferring an electron (1) through the low-potential branch consisting of the two *b*-hemes ($E_{m7} \approx -0.05$ to -0.10 V) across the complex to a quinone, PQ_n, bound on the n-side of the complex; (2) through a high-potential chain consisting of the [2Fe–2S] cluster ($E_{m7} \approx +0.3$ V), the high covalently bound heme of cytochrome *f* ($E_{m7} \approx +0.35$ V), to the PSI reaction center via a soluble heme or copper protein (Figure 4). With no electrochemical restraining forces, the result of the oxidation of two PQH₂ molecules would be the transfer of 2 H⁺ to the p-side aqueous phase for each electron transferred to the high potential chain, that is, $H^+/e = 2$.

Important consequences of the C2 symmetry of the dimer

Although a C2 symmetry is not surprising for a dimeric protein, it has important consequences for the binding and passage of the quinone molecules with their extended isoprenoid tail: (1) Quinones bound on either the p- or n-side will have much less steric interference as a result of their position on the opposite side of the complex. (2) The trajectory of the quinone across the complex, from n- to p-side, or p- to n-, will occur on one lateral surface of the complex, thus avoiding possible hindrance imposed by fatty acyl chains in the inter-monomer cavity.

The Quinone-Exchange Cavity

The space between the two monomers is ~ 30 Å high $\times$ 15 deep and 25 Å wide (Figure 1). The walls of the two cavities on opposite sides of the dimer are formed by transmembrane helices (TMHs) from each monomer: the ‘C, D’ TMH of the cytochrome *b* subunit, the ‘F’ TMH from the subunit IV subunit from one monomer, and the ‘A, E,’ and ISP TMH from

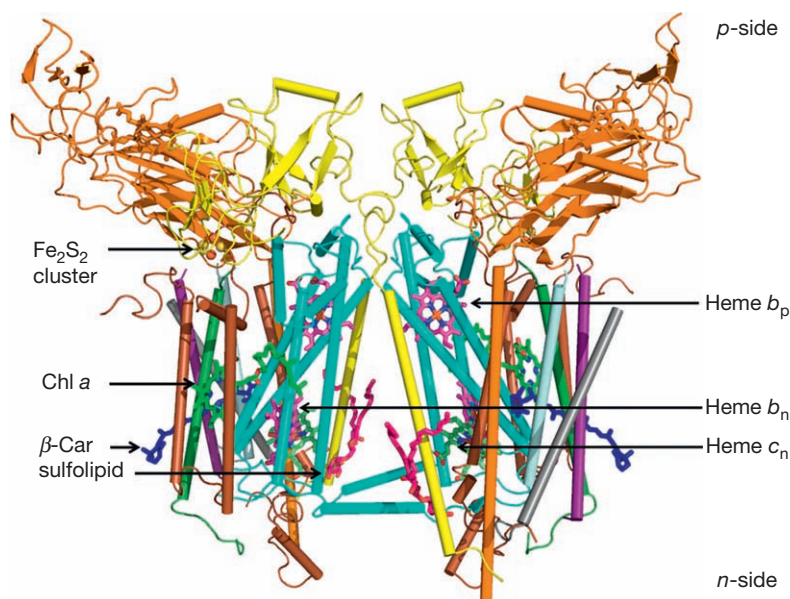


Figure 2 Stick diagram of the symmetric dimeric cytochrome *b₆f* complex, with 13 membrane-spanning α -helices per monomer, derived from the 3.0 Å crystal structure of the cyanobacterium *Nostoc* sp. PCC 7120 (pdb, 2ZT9). Color code: p-side of complex: cyt *f* (*petA*), brown; Rieske ISP (*petA*), yellow; cyt *b* (*petB*), blue; subunit IV (*petD*), brown; petG, L, M, N, gray, green, brown. The bound Chl *a* and β -carotene are shown in dark green and blue. The core of the complex consists of (i) the 24-kDa cytochrome *b* with four trans-membrane helices, the 18-kDa subunit IV (3 TMH), and the single TMH of the Rieske [2Fe–2S] subunit that spans the complex from the soluble domain containing the [2Fe–2S] cluster on the p-side of one monomer across the complex to the n-side of the other. The complex contains seven tightly bound prosthetic groups, of which five are bound redox groups: four hemes (cytochrome *b* in the p-side soluble domain), two noncovalently bound *b*-hemes, *b_p* and *b_n*, that span the complex and one covalently bound *c* heme, the unique heme *c_n*. There are two additional nonredox prosthetic groups, one chlorophyll *a* and one β -carotene whose function is not clear, but which define one of the differences between the *b₆f* and cytochrome *bc₁* complexes. β -Carotene quenches the excited triplet state of the chlorophyll even though the two pigment molecules are separated by 14 Å and can thereby protect the protein network against damage from singlet oxygen. The distances between the prosthetic groups in the *M. lamosus* and *Nostoc* structures are summarized (Table 2).

Quinone-exchange cavity. The cavity between the two monomers is ~30 Å high, 15 Å deep, and 25 Å wide. The boundaries of each of the cavities, which are on opposite sides of the dimer, are formed by trans-membrane helices from each monomer: the C, D TMH of the cytochrome *b* subunit, which contains TMH A–D. The cavity walls are also lined by the F TMH from subunit IV, from one monomer, and the A, E, and ISP TMH from the other. The cavity is probably filled with disordered lipid acyl chains. The structure obtained with a co-crystallized n-side quinone analog inhibitor (pdb: 2E75, 2E76) suggests that the unique heme *c_n* (Figure 3), together with heme *b_n* to which it is electronically coupled, serves as the n-side quinone reductase. The p-side soluble domains of cytochrome *f* and the ISP are predominantly β -strand, as is the p-side plastocyanin. The β -structure for cytochrome *f* is virtually unique for a *c*-type cytochrome. In this respect, and also in its primary amino acid sequence, cytochrome *f* and cytochrome *c₁* of the cytochrome *bc₁* complex are completely different proteins, the only common feature being the characteristic 5-residue Cys–X–Y–Cys–His heme-binding sequence diagnostic of *c*-type cytochromes. Seven lipid-binding sites are found in the structures (not shown), one natural site occupied by an n-side sulfolipid, two sites occupied by DOPC lipid that was used for crystallization and presumably displaced natural lipid, and four sites occupied by UDM detergent molecules.

the other. The true dimensions of the cavity are probably smaller than stated above because the cavity is filled to a large extent with lipid acyl chains. This has not been documented in the crystal structures of the *b₆f* complex, which show seven lipid-binding sites per monomer, but is inferred from the 11 lipids including some acyl chain structures that are resolved in the structure of the yeast *bc₁* complex.

The Unique Heme *c_n*

The presence of heme *c_n* was first detected in intact cells (*in vivo*) by sensitive difference visible light spectrophotometry. Subsequently, in studies with purified *b₆f* complex, heme *c_n* was shown by perpendicular and parallel mode electron paramagnetic resonance (EPR) to be electronically coupled to heme *b_n*, the heme on the n-side of the complex, to be present in *b₆f* complex from spinach thylakoids from which a crystal

structure has not yet been determined, the cyanobacterial *b₆f* complex, and the green alga, *C. reinhardtii*. The crystal structure shows that the distance between *c_n* heme Fe and propionate of heme *b_n* is ~4.0° Å, bridged by an OH[−] ion or H₂O. The proximity between hemes *b_n* and *c_n* is reflected in the strong electronic coupling between the two hemes, as shown by orthogonal and parallel mode EPR spectra of the isolated complex with *g* values as large as 12. Heme *c_n* does not contain any amino acid side chain as an axial ligand (Figure 3). This structural feature makes it a unique heme in all of biochemistry.

Structure: Rotational Isomers of Heme *b_p*

Somewhat remarkably, as seen by the positions of the methyl and vinyl side chains of the heme, the heme plane of heme *b_p* in the *M. lamosus* *b₆f* complex is rotated by 180° about the normal to the membrane plane relative to this heme in

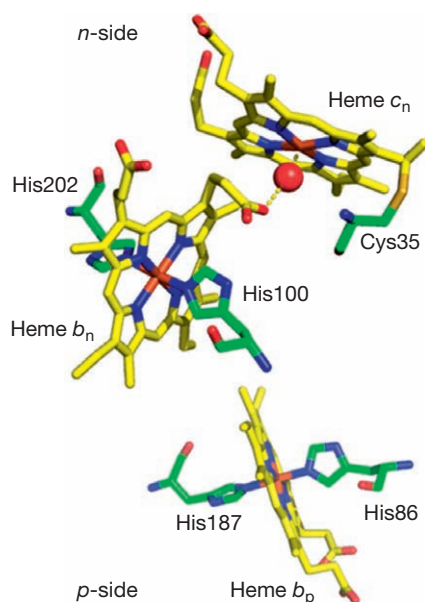
Table 1 Molecular weights and midpoint oxidation–reduction potentials (E_{m7}), values associated with the subunits of the *M. lamosus* cyanobacterial cytochrome *b₆f* complex

Subunit	M_W (Da)	Midpoint redox potential of prosthetic group E_{m7} (mV)
Cyt <i>f</i> (heme) [PetA]	32 273	+350–380
Cyt <i>b₆</i> (two hemes) [PetB]	24 712	(–50 (<i>b_n</i>); –50 to –100 (<i>b_p</i>)) ^a , +100 (heme <i>c_n</i>) ^b
ISP (2Fe–2S)[PetC]	19 295	+295 (pH 6.5) to +320
sulV [Pet D]	17 528	–
PetG	4 058	–
PetM	3 841	–
PetL	3 530	–
PetN	3 304	–

The uncertainty in the E_{m7} values is $\approx \pm 20$ mV.

^aBased on redox titration *in situ*.

^bBased on redox titration *in vitro* with isolated *b₆f* complex.

**Figure 3** Structure and environment of heme *c_n*; arrangement of hemes *b_p*, *b_n*, and *c_n*. The two bis-histidine coordinated *b*-hemes are separated by 7–8 Å edge–edge (pdb: 2E74, 2ZT9) along the direction normal to the membrane plane. Heme *c_n* is positioned on the edge of the intermonomer quinone exchange cavity, with an OH[–] bridging across 4 Å the heme *b_n* and the Fe atom of heme *c_n* (Table 2). The heme is covalently linked via a single thioether group to Cys35 on the n-side of the A TMH of the cyt *b* subunit and, as shown by EPR spectra, electronically coupled to heme *b_n*.

the complex in the *Nostoc* cyanobacterium or in the *C. reinhardtii* alga. The consequences of these rotameric states are not understood.

Structure: Posttranslational Modification of the Iron–Sulfur Protein

Mass spectroscopic analysis has shown that the Rieske protein in the *Nostoc* complex is acetylated at the N-terminus, a

posttranslational modification that is unprecedented in cyanobacterial membranes, and in cytochrome *bc* complexes from any source.

The Chlorophyll and β -Carotene Nonredox Prosthetic Groups

There are two additional prosthetic groups bound to the complex, a chlorophyll *a* and a β -carotene. The β -carotene is able to quench the excited triplet state of the chlorophyll even though the two pigment molecules are separated by 14 Å. The β -carotene can thereby protect the protein network against damage from singlet oxygen. Given this elaborate protection system, the function of the chlorophyll, although unknown, must be important. The prosthetic groups are organized around the central quinone-exchange cavity (Figure 2).

Amino Acid Sequences of Subunits

Amino acid sequences of subunits are as follows: (1) *Mastigocladus laminosus b₆f* complex: (i) gene sequences (*petA*, *B*, *C*, *D* genes), (NCBI accession numbers AY390356 and AY390357); (ii) amino acid sequences in PDB accession code 2E74; (2) *Nostoc* sp. PCC 7120 *b₆f* complex: (i) sequence of entire genome including cyt *b₆*, 215 amino acids (aa); apocytochrome *f*, 333 aa; Rieske iron–sulfur protein, 178 aa; subunit IV, 160 aa; *petG*, 37 aa; *petL*, 31 aa; *petM*, 34 aa; *petN*, 29 aa, in PDB, 2ZT9; (ii) *Chlamydomonas reinhardtii b₆f* complex (PDB, 1Q90).

Function

The second p-side turnover of PQH₂ can transfer a second electron to PQ_n or the plasto-semiquinone, PQ[•], on the n-side of the complex, regenerating PQH₂. Thus, the *b₆f* complex functions as a plastoquinol: plastocyanin/cytochrome *c* oxidoreductase on the electrochemically positive (p) side of the membrane and as a plastoquinone reductase on the electrochemically negative (n) side (Figure 4). These transfer events are analogous to those in the *bc₁* complex of the respiratory chain and photosynthetic bacteria. As noted above, the PQ/PQH₂ cycle in the *b₆f* complex differs in several respects from the Q-cycle model initially proposed for regeneration of ubiquinol through function of the cytochrome *bc₁* complex in the respiratory chain and photosynthetic bacteria: (1) the presence of heme *c_n*, which is inferred from crystal structure data to be the PQ reductase, and, from EPR spectra in the presence of NO₂, the PQ oxidase; (2) the electronic coupling between hemes *b_n* and *c_n*, which implies that quinol reduction need not necessarily proceed through a semiquinone intermediate, but perhaps cooperatively through hemes *b_n* and *c_n*, providing protection against the formation of superoxide ion and reactive O₂ species; and (3) the presence of ferredoxin: NADP⁺ reductase (FNR) bound to purified plant (spinach) *b₆f* complex implies that electrons needed for the reduction of PQ to PQH₂ can also be supplied by cyclic electron transport, which can transfer electrons back to the *b₆f* complex and the central electron transport chain via ferredoxin from the reducing side of PSI. Coupling of electron and proton transfer generates a maximum H⁺/e[–] ratio = 2 for proton translocation across the complex from the n- to the p-side aqueous phase.

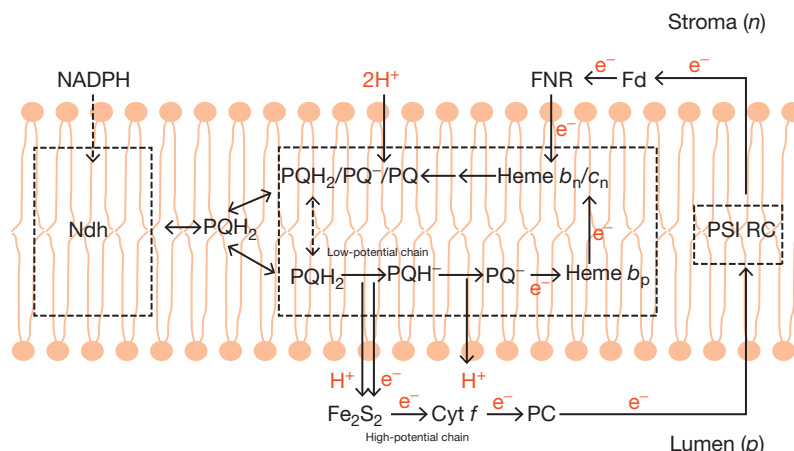


Figure 4 A Q cycle model for a bifurcated pathway of electron transfer into a high (ISP → cyt *f* → PC → PSI) and low-potential (PQH₂ → heme *b_p* → heme *b_n* → PQ_n) chain, resulting in the transfer of as many as 2H⁺ per electron transferred to the p-side aqueous phase. The NDH (ndh-1) complex, which has not yet been completely purified, is shown as another pathway for reduction of the intra-membrane PQ/PH₂ pool.

Rate-limiting step, activity, and inhibitors

Under conditions of sufficient substrates for carbon fixation, the proton-coupled oxidation of plastoquinol by the Rieske [2Fe–2S] ISP is the rate-limiting step, 1–10 ms, of the electron transport reactions of oxygenic photosynthesis. This step is readily assayed through the kinetics of reduction of cytochrome *f*, the electron acceptor of the ISP. High *in vitro* activity of the isolated *b₆f* complex, ~200 electrons cyt *f*^{–1}, s^{–1}, utilizing the plastoquinone analog, decyl-plastoquinone, reduced by borohydride as the electron/proton donor, and plastocyanin oxidized by ferricyanide as electron acceptor, is an important prerequisite for successful crystallization of the isolated *b₆f* complex. A poor electron-transfer rate implies a damaged or deficient complex. The concentration of *b₆f* complex is assayed routinely through the concentration of cytochrome *f*, determined by the amplitude of the hydroquinone-reduced minus ferricyanide-oxidized difference absorbance spectrum, using a molar extinction coefficient of $2.6 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$.

Certain quinol analogs can act as inhibitors of electron transfer in the complex: (1) *p*-side – 2,5-dimethyl, 6-isopropyl benzoquinone (DBMIB), stigmatellin or a tridecyl derivative on the p-side of the complex at the site of quinol oxidation by the ISP; (2) *n*-side – 2-n-nonyl-4-hydroxyquinoline-N-oxide (NQNO) at the site of heme *c_n* that faces the quinone-exchange cavity. NQNO is a much better n-side inhibitor of the *b₆f* complex than is antimycin.

Function: intra-monomer versus inter-monomer electron transfer. The existence of the dimeric structure of cytochrome *bc* complexes can be rationalized on the basis of structure–function considerations involving shielding of hydrophobic surfaces that define the trans-membrane corridor for passage of the lipophilic quinone/quinol. The existence of the dimer raises the question of how it is used for trans-membrane electron transfer, that is, intra-monomer versus inter-monomer pathway. On the basis of the shorter intra-monomer heme distances (Table 2), the intra-monomer heme *b_p*–*b_n* pathway would be favored over a pathway involving p-side cross-over between the two hemes *b_p*.

Table 2 Distances between prosthetic groups of the *b₆f* complex from *M. lamosus* and *Nostoc* sp. PCC 7120

Prosthetic group	Distance (Å)	
	Fe–Fe	Edge–Edge
[2Fe–2S]–Cyt <i>f</i>	29	25, 26
[2Fe–2S]–heme <i>b_p</i>	28	24
Heme <i>b_p</i> –heme <i>b_n</i>	20.6, 20.7	7.4, 8.2
Heme <i>b_p</i> –heme <i>b_p</i>	22.0, 22.1	12.9, 12.7
Heme <i>b_n</i> –heme <i>b_n</i>	35	29
Heme <i>b_n</i> –heme <i>c_n</i>	9.6, 9.7	4 (<i>c_n</i> Fe – <i>b_n</i> propionate)

Distances are very similar in the complex from *C. reinhardtii*.

Evolution

Problems of interest in the evolution of cytochrome *bc* complexes are: (1) the common cytochrome *b* section of the core of *b₆f* and *bc₁* complexes. The conservation of the 4 helix bundle and the trans-membrane portion of the two heme-binding histidine residues in helices 2 and 4, of the heme-binding sites and the distribution of hydrophobicity in the cytochrome *b₆* subunit, and the N-terminal heme-binding half of the cyt *b* polypeptide of the *bc₁* complex implies a common evolutionary origin of the hydrophobic core of *bc* complexes. As evolution led to the addition of different subunits, the *bc₁* and *b₆f* structures diverged in directions parallel and perpendicular to the plane of the membrane. The similarity of evolved function, as an acceptor of electrons from the [2Fe–2S] cluster of the iron–sulfur protein, and almost complete dissimilarity of amino acid sequence and structure of cytochromes *f* and *c₁*, implies convergent evolution of the cyt *f/c* subunits. (2) The comparative evolution of *b₆f* and *bc₁* complexes and the appearance of the heme *c_n*, unique to the *b₆f* complexes, appear to coincide with the emergence of plastoquinone-containing *bc* complexes. (3) The high degree of similarity between the structures of the *b₆f* complex in cyanobacteria and the green alga, *C. reinhardtii*. The crystal structure of the *b₆f* complex at

3.0 Å resolution is essentially identical in the two cyanobacteria that have been analyzed and in the green alga, *C. reinhardtii*, whose appearance on the Earth was approximately 10⁹ years later than that of the cyanobacteria. In both structures, there are eight polypeptide subunits, four (*petA*, B, C, and D) large (M_w = 18–31 kDa) and four (*petG*, L, M, and N) small (M_w = 3.3–4.1 kDa) subunits (Table 1). The root mean square deviation (RMSD) between the 3.0-Å crystal structures from the cyanobacteria *M. laminosus* and *Nostoc* structures, relative to that from the *C. reinhardtii*, are: (1) for the Rieske [2Fe–2S] protein, *M. laminosus* native versus *C. reinhardtii* with the inhibitor, TDS, 1.25 Å; (2) *M. laminosus* TDS structure versus *C. reinhardtii* TDS, 1.06 Å; (3) without the Rieske iron–sulfur protein, *M. laminosus* native structure versus *C. reinhardtii* TDS, 0.98 Å, and *M. laminosus*-TDS structure versus *C. reinhardtii*-TDS, 0.99 Å.

Acknowledgments

The research of the authors that forms the basis of this article was supported by NIH GM 38323 (WAC) and NIH P50-GM08499 (WAC & JW).

See also: **Bioenergetics:** Cytochrome *bc₁* Complex (Respiratory Chain Complex III); Dynamic Behavior of Photosystem II Light Harvesting; *F_X*, *F_A*, and *F_B* Iron–Sulfur Clusters in Type I Photosynthetic Reaction Centers; Photosystem II: Assembly and Turnover of the D1 Protein; Photosystem II: Water Oxidation, Overview.

Further Reading

- Allen JF (2004) Cytochrome *b₆f*: Structure for signaling and vectorial metabolism. *Trends in Plant Science* 9: 130–137.
- Alric J, Pierre Y, Picot D, Lavergne J, and Rappaport F (2005) Spectral and redox characterization of the heme *c₁* of the cytochrome *b₆f* complex. *Proceedings of the National Academy of Sciences of the United States of America* 102: 15860–15865.
- Baniulis D, Yamashita E, Whitelegge JP, et al. (2009) Structure-function and stability of the cyanobacterial cytochrome *b₆f* complex from *Nostoc* sp. PCC 7120. *Journal of Biological Chemistry* 284: 9861–9869.
- Blankenship RE (2002) *Molecular Mechanisms of Photosynthesis*, chs. 7, 8, 11, p. 321. Williston, VT: Blackwell Science.
- Cooley JW, Nitschke W, and Kramer DM (2008) The cytochrome *bc₁* and related *bc* complexes, the Rieske/cytochrome *b* complex as the functional core of a central electron/proton transfer complex. In: Hunter CN, Daldal F, Thurnauer M, and Beatty JT (eds.) *The Purple Photosynthetic Bacteria*, vol. 28, pp. 451–473. Dordrecht: Springer-Verlag.
- Cramer WA, Baniulis D, Yamashita E, et al. (2008) Structure, spectroscopy, and function of the cytochrome *b₆f* complex: Heme *c_n* and *n*-side electron and proton transfer reactions. In: Fromme P (ed.) *Photosynthetic Protein Complexes. A Structural Approach*, pp. 155–179. Weinheim [BC]: Wiley-VCH.
- Cramer WA and Knaff DB (1991) *Energy Transduction in Biological Membranes*, chs. 1, 2, 5, and 6, p. 579. Springer Study Edition [paperback, ISBN 0-387-97533-0].
- Cramer WA, Savikhin S, Yan J, and Yamashita E (2010) The enigmatic chlorophyll *a* molecule in the cytochrome *b₆f* complex. In: Rebeiz C (ed.) *The Chloroplast*, ch. 6. Springer-Verlag.
- Cramer WA, Zhang H, Yan J, Kurisu G, and Smith JL (2006) Trans-membrane traffic in the cytochrome *b₆f* complex. *Annual Reviews of Biochemistry* 75: 769–790.
- Hauska G (2004) The isolation of a functional cytochrome *b₆f* complex: From lucky encounter to rewarding experiences. *Photosynthetic Research* 80(1–3): 277–291.
- Joliot P and Joliot A (1988) The low-potential electron-transfer chain in the cytochrome *bf* complex. *Biochimica et Biophysica Acta* 933: 319–333.
- Kallas T (1994) The cytochrome *b₆f* complex. In: Bryant DA (ed.) *The Molecular Biology of Cyanobacteria*, pp. 259–317. Dordrecht: Kluwer Academic Publishers.
- Kurisu G, Zhang H, Smith JL, and Cramer WA (2003) Structure and mechanism of the cytochrome *b₆f* complex of oxygenic photosynthesis: Tuning the cavity. *Science* 302: 1009–1014.
- Lavergne J (1983) Membrane potential-dependent reduction of cytochrome *b₆* in an algal mutant lacking photosystem I centers. *Biochimica et Biophysica Acta* 725: 25–33.
- Nicholls DG and Ferguson SJ (2002) *Bioenergetics*, 3rd edn., chs. 1, 3, 4, and 6, p. 297. Amsterdam: Academic Press.
- Nitschke W, van Lis R, Schoepp-Cothenet B, and Baymann F (2010) The “green” phylogenetic clade of Rieske/cyt *b* complexes. *Photosynthetic Research* 104: 347–355.
- Soriano GM, Smith JL, and WA Cramer (2001) Structure–function of the cytochrome *b₆f* complex of oxygenic photosynthesis. In: Messerschmidt A, Huber R, Wieghardt K, and Poulos T (eds.) *Handbook of Metalloproteins*, Vol. 1, pp. 172–181. London: Wiley.
- Stroebel D, Choquet Y, Popot J-L, and Picot D (2003) An atypical heme in the cytochrome *b₆f* complex. *Nature* 426: 413–418.
- Whitelegge J, Zhang H, Taylor R, and Cramer WA (2002) Full subunit coverage liquid chromatography electrospray-ionization mass spectrometry (LCMS+) of an oligomeric membrane protein complex: The cytochrome *b₆f* complex from spinach and the cyanobacterium, *Mastigocladus laminosus*. *Molecular and Cellular Proteomics* 1: 816–827.
- Widger WR, Cramer WA, Herrmann RG, and Trebst A (1984) Sequence homology and structural similarity between cytochrome *b* of mitochondrial complex III and the chloroplast *b₆f* complex: Position of the cytochrome *b* hemes in the membrane. *Proceedings of the National Academy of Sciences of the United States of America* 81: 674–678.
- Yamashita E, Zhang H, and Cramer WA (2007) Structure of the cytochrome *b₆f* complex: Quinone analogue inhibitors as ligands of heme *c_n*. *Journal of Molecular Biology* 370: 39–52.
- Yan J, Kurisu G, and Cramer WA (2006) Structure of the cytochrome *b₆f* complex: Binding site and intra-protein transfer of the quinone analogue inhibitor 2, 5-dibromo-3-methyl-6-isopropyl-*p*-benzoquinone. *Proceedings of the National Academy of Sciences of the United States of America* 103: 67–74.
- Zhang H, Kurisu G, Smith JL, and Cramer WA (2003) A defined protein–detergent–lipid complex for crystallization of integral membrane proteins: The cytochrome *b₆f* complex of oxygenic photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* 100: 5160–5163.

Structure of P-Type Adenosine Triphosphatases

I Dach and P Nissen, Aarhus University, Aarhus, Denmark

© 2013 Elsevier Inc. All rights reserved.

Glossary

A-domain Actuator domain that is involved in the dephosphorylation of the pump.

Cardiotonic steroids Inhibitors of the Na^+, K^+ -adenosine triphosphatase (ATPase) used in the treatment of congestive heart failure.

M-domain Membrane domain containing binding sites for substrate ions.

N-domain Nucleotide-binding domain with a conserved nucleotide-binding residue.

P-domain Highly conserved phosphorylation domain containing the phosphorylation site.

Post-Alberts scheme A reaction scheme describing the transport process as cyclical changes between two states

with different affinities for adenosine triphosphate (ATP) and substrates.

Proton pump inhibitors Inhibitors of the gastric H^+, K^+ -ATPase commonly used in the treatment of gastric diseases such as ulcers.

P-type ATPases Family of membrane-bound ion pumps utilizing ATP and characterized by phosphorylation of a conserved amino acid residue during their functional cycle.

Thapsigargin An effective inhibitor of the sarcoplasmic reticulum Ca^{2+} -pump whose derivatives are currently tested for the treatment of prostate cancer.

P-Type Adenosine Triphosphatase Family

P-type adenosine triphosphatase (ATPases) constitute a large family of evolutionary related membrane proteins that can be found in virtually all living organisms, and are essential housekeeping enzymes in eukaryotes. Members of the family include ion pumps that mediate the thermodynamic uphill movement of cations against their concentration gradients across biological membranes by the use of energy derived from adenosine triphosphate (ATP) hydrolysis. The electrochemical gradients created and maintained by P-type ATPases are essential for survival as they are involved in a number of vital processes mediated by channels and secondary transporters. These include, for example, regulation and maintenance of action potentials in nervous tissues, secretion and absorption of solutes, and control of cell volume, ionic homeostasis and pH (Na^+, K^+ -ATPase), acidification of the gastric juice (H^+, K^+ -ATPase) or relaxation of muscles, and Ca^{2+} -dependent signal transduction (Ca^{2+} -ATPase). Many members of the P-type ATPase family have been implicated in the onset of various diseases such as neurodegenerative disorders arising from the mutations in brain Na^+, K^+ -ATPase or Menkes and Wilson diseases resulting from mutations in copper-transporting ATPases. Consequently, P-type ATPases are clinically important targets; compounds inhibiting the cardiac Na^+, K^+ -ATPase are used in the treatment of congestive heart failure, and inhibitors of the gastric H^+, K^+ -ATPase are the main agents used in the treatment of dyspeptic conditions such as gastric ulcers. Furthermore, P-type ATPases encompass promising targets for a broad range of novel drugs – the sarcoplasmic reticulum Ca^{2+} -ATPase is being developed as a target in the treatment of prostatic cancer, and P-type H^+ -ATPases are attractive targets for novel antifungal agents.

According to analysis of sequence and substrate specificities, P-type ATPases are divided into five subfamilies. While most of the pumps transport cations with a substrate specificity

ranging from monovalent cations such as H^+ , Na^+ , and K^+ (P-type ATPase subfamily II and III) to heavy metals such as Cu^{2+} , Ag^{2+} , Zn^{2+} , Cd^{2+} , or Pb^{2+} (subfamily I), members of the IV (and possibly also V) subfamily are involved in flipping of phospholipids and thus play a role in the maintenance of an asymmetric nature of biological membranes. The number of subunits as well as the size of P-type ATPases vary depending on the subfamily, with the molecular mass ranging from 65 to 150 kDa.

The transport process performed by P-type ATPases follows the post-Alberts reaction scheme that is based on cyclical changes between two main conformational regimes (denoted E1 and E2) that differ in their affinities for ATP and substrates. The term 'P-type ATPases' is used to indicate that a key-conserved residue is phosphorylated and forms an acid-stable aspartyl phosphoanhydride intermediate during their functional cycle.

Structural Studies of P-Type ATPases

Since the Nobel laureate Jens C Skou discovered the Na^+, K^+ -ATPase in 1957, a great amount of work has been done to explore the structural and functional characteristics of the P-type ATPases, with the sarcoplasmic reticulum Ca^{2+} -ATPase from rabbit skeletal muscle (SERCA 1a) currently being the most-studied family member. Using X-ray crystallography, atomic structures representing almost all relevant intermediates in its reaction cycle have been determined at a reasonable resolution, allowing us to understand structural aspects of the transport mechanism. In the case of other pumps such as the Na^+, K^+ -ATPase and the plasma membrane H^+ -ATPase, three-dimensional structures of one of the states of the reaction cycle are available, and biochemical analysis and mutational studies are included to investigate the transport mechanism of these pumps. Moreover, because there is a significant degree of similarity in the overall structure and reaction mechanisms of all

P-type ATPases, structures of different conformations of the Ca^{2+} -ATPase are used as a template for computer-assisted homology modeling for the whole P-type ATPase family.

Sarcoplasmic Reticulum Calcium Pump – A Model Member of the P-Type ATPase Family

The sarcoplasmic reticulum Ca^{2+} -ATPase pumps Ca^{2+} ions from the cytosol of skeletal muscle cells into the sarcoplasmic reticulum store. This gives rise to steep Ca^{2+} gradients that enable fast calcium-mediated signaling and the reuptake of cytoplasmic Ca^{2+} in muscle upon the termination of a contraction event. A low cytosolic concentration of Ca^{2+} maintained by the Ca^{2+} -ATPase is also necessary for other cellular processes, including neurotransmission, neurosecretion, and egg fertilization.

The overall structure of the Ca^{2+} -ATPase consists of four well-defined protein domains connected by linkers in a complex topology (Figure 1). These domains, named according to their function or position, are the M-domain (membrane domain) and three cytosolic domains – P for phosphorylation

domain, N for nucleotide-binding domain, and A for actuator domain. The P-domain contains the phosphorylation site of the pump and its sequence is highly conserved throughout the P-type ATPase family. It is composed of two parts – a central N-terminal part with the phosphorylation motif featuring a conserved aspartate residue (Asp 351 in SERCA 1a) and a C-terminal part that surrounds the phosphorylation motif as two hemispherical subdomains denoted P-I and P-II. The P-II subdomain includes an important and well-conserved signature motif present in all P-type ATPases, encompassing another aspartic residue (Asp 703) that along with the phosphorylatable Asp351 coordinates a Mg^{2+} ion during phosphorylation. Two short hinges link to the N-domain inserted in sequence between the N- and C-terminal parts of the P-domain. The N-domain resembles a top-hat and contains a nucleotide-binding site which is highly conserved among all members of the P-type ATPase family. Upon binding of ATP, the adenine base binds to a groove at the lower end of the top hat while the triphosphate part of ATP adopts a bent configuration located outside this groove and cross-linking the N- and P-domain. The smallest of the three cytoplasmic domains – the A-domain – constitutes a central part rich in beta sheets that is capped by a helical

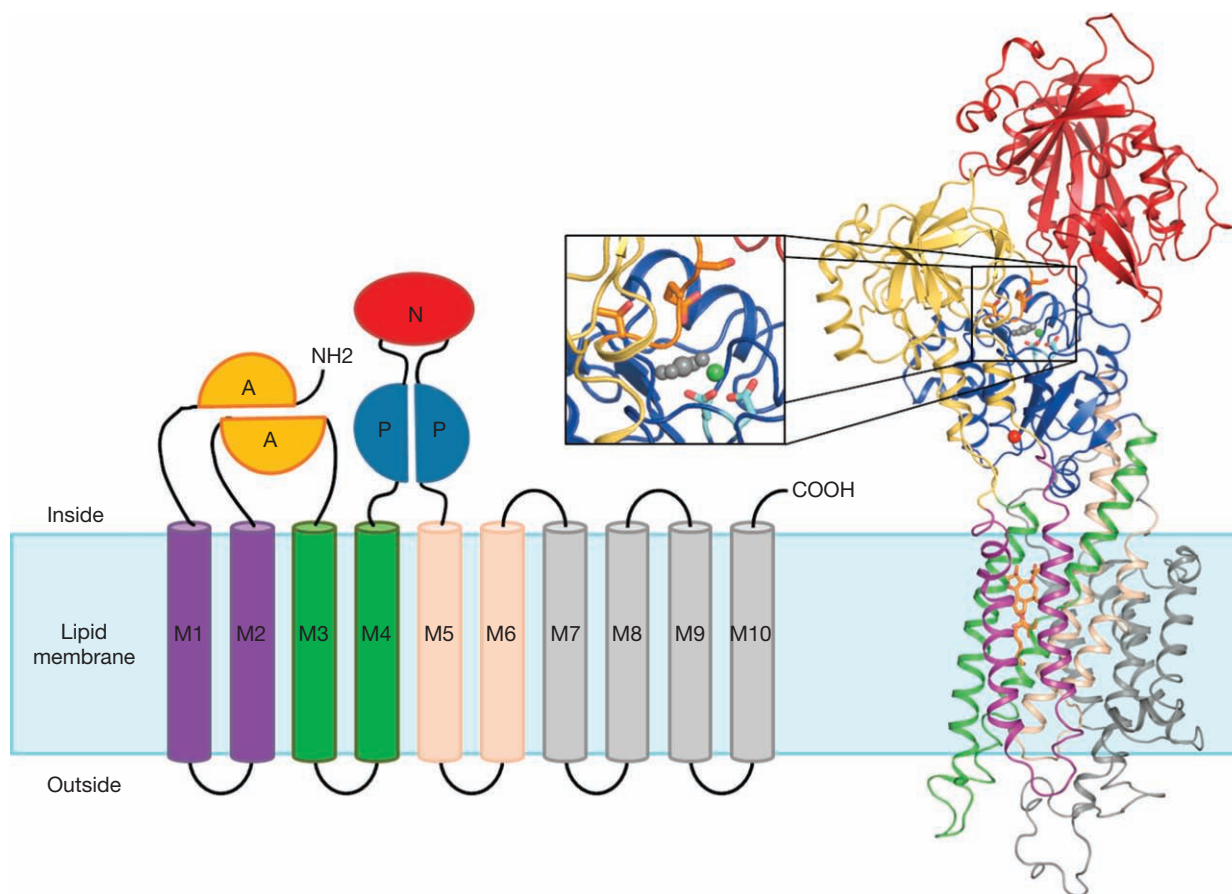


Figure 1 Schematic representation of the structure of the sarcoplasmic reticulum calcium pump (left) and its crystal structure (right) with an inset showing the residues involved in phosphorylation of the pump. A-domain is shown in yellow, P-domain in blue, N-domain in red, transmembrane segments M1-M2 in purple, M3-M4 in green, M5-M6 in wheat and M7-M10 in gray. Amino acid residues forming the TGES motif are highlighted in orange and residues Asp 351 and Asp 701 in cyan. Mg^{2+} and K^{+} ions are depicted as green and red spheres, respectively. The crystal structure is stabilized by aluminium fluoride and thapsigargin shown in gray and orange, respectively. From Olesen C, Sørensen TLM, Nielsen RC, Møller JV, and Nissen P. (2004) Dephosphorylation of the calcium pump coupled to counter occlusion. *Science* 306: 2251–2255.

N-terminal part. The A-domain plays an essential role in dephosphorylation of the pump during its reaction cycle, as it contains several important residues for interactions with other domains including the conserved TGES sequence motif (Thr 181, Gly 182, Glu 183, and Ser 184) that interacts with the phosphorylated Asp 351 residue of the P-domain and catalyzes the hydrolysis of the aspartyl phosphoanhydride. During the reaction cycle, rotation of the A-domain enables the interaction of all three cytosolic domains and actuates the ion transduction coupled to ATP phosphorylation and dephosphorylation. The linkage of the A-domain to the first two transmembrane helices of the M-domain is crucial for the opening and closing of the Ca^{2+} -binding sites. The M-domain consists of 10 transmembrane helices (denoted M1–M10) forming a cylindrical structure in the sarcoplasmic reticulum membrane where it is surrounded by a ring of an irregular bilayer formed primarily by phospholipids. In the E1 state of the pump reaction cycle, two Ca^{2+} ions are coordinated inside the binding cavity between transmembrane helices M4, M5, M6, and M8 with the help of four side-chain carboxyl groups and main-chain carbonyl groups, forming an electroneutral complex that is later exposed to the extracellular/luminal site of the membrane.

Apart from the two binding sites for Ca^{2+} ions, the M-domain possesses several binding cavities for compounds that can inhibit the calcium pump by blocking the opening of its luminal gate for ions. The most powerful inhibitor of the Ca^{2+} -ATPase described so far is thapsigargin, a sesquiterpene lactone derived from a Mediterranean plant *Thapsia garganica*, whose derivatives, able to selectively target cancer cells, show promise in the treatment of slowly growing prostate cancer cells that are resistant to other cancer therapies. Inducing only a few changes in the structure, thapsigargin binds to a cavity between helices M3, M4, M5, and M7, immobilizing them and thus trapping the pump almost irreversibly in the E2 state that prevents calcium binding and thus the E2 to E1 transition (Figure 1). Another group of inhibitors, represented by cyclopiazonic acid and 2,5-di-(*t*-butyl)-dihydroxybenzene, block a cavity formed by helices M1, M2, M3, and M4, presumed to represent the cytoplasmic access channel for calcium leading to an important glutamate residue Glu 309, which acts as a carrier of Ca^{2+} moving up and down with the M4 helix as the pump changes its conformational state.

Structural Changes during the Calcium Pump Reaction Cycle

Conformational changes in the pump structure during the transport cycle are a result of coordinated movements of transmembrane helices connecting the phosphorylation site at the P-domain with the cation-binding sites in the M-domain. These changes, induced by ion and ATP binding, are enabled by the energy derived from the two-step ATP hydrolysis at the phosphorylation site. While significant rearrangement occurs in the M-domain during the transport cycle, each of the P-, N-, and A-domains move almost as rigid bodies. Importantly, all steps in the reaction cycle are reversible.

The Ca^{2+} transport cycle of the calcium pump starts with the binding of two Ca^{2+} ions from the cytoplasm to the E2 state of the ATPase activating the phosphorylation site. In the

ATP-bound E2 state, the A-domain is rotated toward the phosphorylation site and interacts closely with the P- and N-domains. The binding of two Ca^{2+} ions from the cytoplasmic side of the membrane exerts a translational movement of the M4 helix resulting in a displacement of the A-domain, which enables a close interaction between the P- and N-domains cross-linked by ATP now interacting with the Mg^{2+} site and Asp 351 residue. This allows for phosphorylation of the enzyme forming the so-called adenosine diphosphate (ADP)-sensitive $\text{Ca}_2\text{E1}\sim\text{P}$ high-energy intermediate state which is accompanied by the occlusion of bound calcium at high-affinity sites. Phosphorylation of the ATPase unleashes the N-domain from the P-domain and induces a subsequent rotation of the A-domain, interacting now with the phosphorylated P-domain and manipulating the transmembrane segments M1 through M4 in formation of the ADP-insensitive E2P state. This, in turn, forces these helices to spread out and separate from segments M5 and M6 exposing a large water-filled opening, which provides a luminal exit pathway leading from the two, now deformed, Ca^{2+} binding sites in the membrane. Ca^{2+} ions are now released to the lumen from low-affinity sites that consequently adopt a high affinity for counter-transported protons (i.e., changing the pK_a to a higher value). Counter-cation binding stimulates the last step in the reaction cycle, namely the return of the pump to the E2 state by dephosphorylation. This reaction is mediated by hydrolytic cleavage of the aspartyl phosphoanhydride, enabled by a close interaction between the A- and P-domains that results in docking of the A-domain's TGES motif into the phosphorylation site. Here the Glu183 of the TGES motif activates a water molecule for in-line nucleophilic attack on the phosphorylation site. A total of two to three protons (depending on the pH and other environmental conditions) are counter-transported from the luminal to the cytosolic side of the membrane in this process, being released from the E2 state prior to the binding of Ca^{2+} . With two Ca^{2+} ions bound to the pump and no nucleotides present, the ATPase adapts a $\text{Ca}_2\text{E1}$ conformation that is characterized by a rather loose interaction between the three cytosolic domains, resulting from a counter-clockwise rotation of the A-domain around the P-domain upon the coordination of Ca^{2+} . It remains unsolved if this structural state of the pump is of functional importance.

The pump reaction cycle can be regulated on many levels, for example, by the presence of monovalent cations and by high concentrations of ATP or Ca^{2+} . As an example, the dephosphorylation reaction of the E2P state is facilitated by binding of a monovalent cation such as K^+ or Na^+ at the C-terminal site of the P-domain, which helps to stabilize the pump and causes a 20-fold increase of the dephosphorylation rate. Furthermore, high levels of ATP stimulate all partial reactions of the cycle by a so-called modulatory effect. Structures with modulatory ATP show that the nucleotide destabilizes interactions that characterize one state, enabling the transition to a subsequent state (Figure 2).

The Sodium–Potassium Pump

The Na^+,K^+ -ATPase maintains sodium and potassium gradients across the plasma membrane of animal cells by

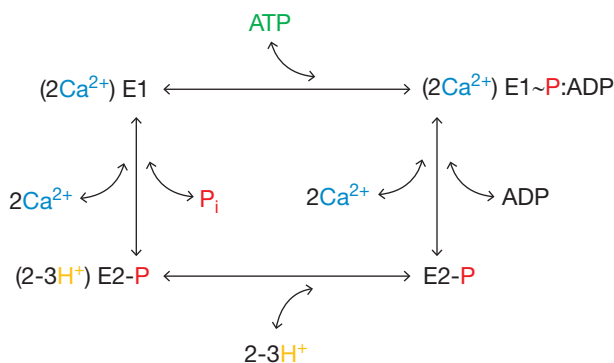


Figure 2 Schematic overview of the functional cycle of the sarcoplasmic reticulum Ca^{2+} -ATPase.

transporting three Na^+ ions out of the cell in exchange for two extracellular K^+ ions. These gradients are necessary for energization of the resting potential, cell signaling, regulation of cell volume, osmotic activity and pH, and for cellular uptake of ions, nutrients, and neurotransmitters. Mutations in the brain Na^+, K^+ -ATPase lead to neurodegenerative disorders such as the migraine subtype familial hemiplegic migraine type 2 and the movement disorder familial rapid-onset dystonia parkinsonism.

Unlike the Ca^{2+} -pump, the Na^+, K^+ -ATPase is composed of three subunits denoted by α , β , and γ (Figure 3). The catalytic α -subunit performs ATP hydrolysis together with Na^+ and K^+ binding and its structure is homologous to the structure of the Ca^{2+} -ATPase, containing three cytoplasmic domains P, A, and N, and a membrane-embedded M-domain with 10 transmembrane spanning helices comprising the binding sites for Na^+ and K^+ . The β -subunit, consisting of one transmembrane-spanning helix and a large extracellular glycosylated domain, is a unique feature of the K^+ -counter-transporting-P-type ATPases – the Na^+, K^+ -ATPase and H^+, K^+ -ATPase. It is directly associated with the transmembrane helices M7 and M10 of the α -subunit covering the extracellular loops of the M5–M10 segment like a lid. The β -subunit is essential for the pump's activity as it is required for binding and occlusion of K^+ ions, for modulation of the pump's catalytic properties, and for stability of the catalytic α -subunit as well as its maturation and trafficking to the plasma membrane. Furthermore, there is evidence that in many cell types the $\alpha\beta$ -subunit complex is associated with a small tissue-specific regulatory γ -subunit from the FXYD protein family. This third subunit is not required for the function of the pump, but can affect its conformational state and the affinity for Na^+ , leading to a modulation of the Na^+, K^+ -ATPase activity. While the transmembrane part of the γ -subunit is within the interaction distance of the α -subunit's M-domain, its extracellular site containing the conserved FXYD motif moves in between the α - and β -subunits.

In humans, four different isoforms of the α -subunit and three isoforms of the β -subunit have been identified, denoted $\alpha 1$ – $\alpha 4$ and $\beta 1$ – $\beta 3$, respectively. These isoforms have different physiological roles, display different voltage sensitivities, and are tissue specific. While there are only very few differences between the four α isoforms in the transmembrane region, they

differ in surface-exposed regions of the A- and N-domains, which may reflect their characteristic interactions with other tissue-specific cellular components, for example, during trafficking. The most abundant form of the enzyme is the $\alpha 1\beta 1$ isoform, expressed in most tissue types. The $\alpha 2$ isoform is found mainly in the skeletal muscle, and also in the brain, heart, and eyes. The $\alpha 3$ isoform is mostly found in the brain and isoform $\alpha 4$ is expressed only in testis/sperm cells.

The overall architecture of the catalytic subunit of the Na^+, K^+ -ATPase is almost identical to the structure of the Ca^{2+} -ATPase, except for a slight shift in the N-domain. Also the conserved residues in the three cytoplasmic domains playing a role in phosphorylation and dephosphorylation are found in nearly the same positions in both pumps – for example, Asp 351, an aspartate residue that becomes phosphorylated during the SERCA 1a reaction cycle, corresponds to an Asp 369 residue in the Na^+, K^+ -pump. While the structure of the P-domain is highly similar in both pumps, the A- and N-domains of the Na^+, K^+ -ATPase are a few surface-loops residues shorter than in the Ca^{2+} -ATPase. As in the Ca^{2+} -ATPase, the M4 and M6 helices of the α -subunit of the Na^+, K^+ -ATPase are unwound in the middle, thereby making space for an ion-binding pocket exposing main chain carbonyls between transmembrane helices M4, M5, M6, and M8. Structural analyses clearly indicate that Na^+ is released on the extracellular side in exchange for K^+ being counter-transported to the cytoplasm through the same occlusion cavity, that is, as suggested by the consecutive transport model. Surprisingly, the structural resemblance of the α -subunit of the Na^+, K^+ -ATPase to the Ca^{2+} -ATPase is very high even in the cation-binding pocket, with only a few differences in the amino acid residues involved in the binding of ions, suggesting that subtle differences in the positions of the side chains and water molecules are contributing to the cation selectivity of the pump. In the membrane domain of the Na^+, K^+ -ATPase, significant differences from the Ca^{2+} -ATPase were observed in helices M7 and M10, and a unique aspect of the Na^+, K^+ -ATPase is the third Na^+ -binding site, whose exact location still remains unresolved.

The transport cycle of the Na^+, K^+ -ATPase resembles that of the calcium pump, starting with the binding of three intracellular Na^+ ions from the cytoplasmic side of the membrane to the E1 form that enables the phosphorylation of Asp 369 by ATP. This leads to a conformational change resulting in the E2P state, accompanied by the release of the three Na^+ ions into the extracellular environment. Subsequently, binding of two K^+ ions from the extracellular side triggers the dephosphorylation reaction, followed by a transition from the E2 to the E1 state that is accompanied by the release of the two K^+ ions into the cytoplasmic side of the membrane. The Na^+ and K^+ ions become temporarily occluded inside the binding pocket during the transport cycle.

The reaction cycle of the Na^+, K^+ -ATPase can be regulated by changes in the Na^+ and K^+ concentrations, through interactions with proteins from the FXYD family, and by hormonal regulation. Additionally, the C-terminus of the α -subunit has been shown to participate in Na^+, K^+ -ATPase regulation by controlling the affinity for Na^+ on both sides of the membrane and interacting with the third Na^+ -binding site.

The most widely studied inhibitors of the Na^+, K^+ -ATPase reaction cycle are cardiotonic steroids (also called cardiac

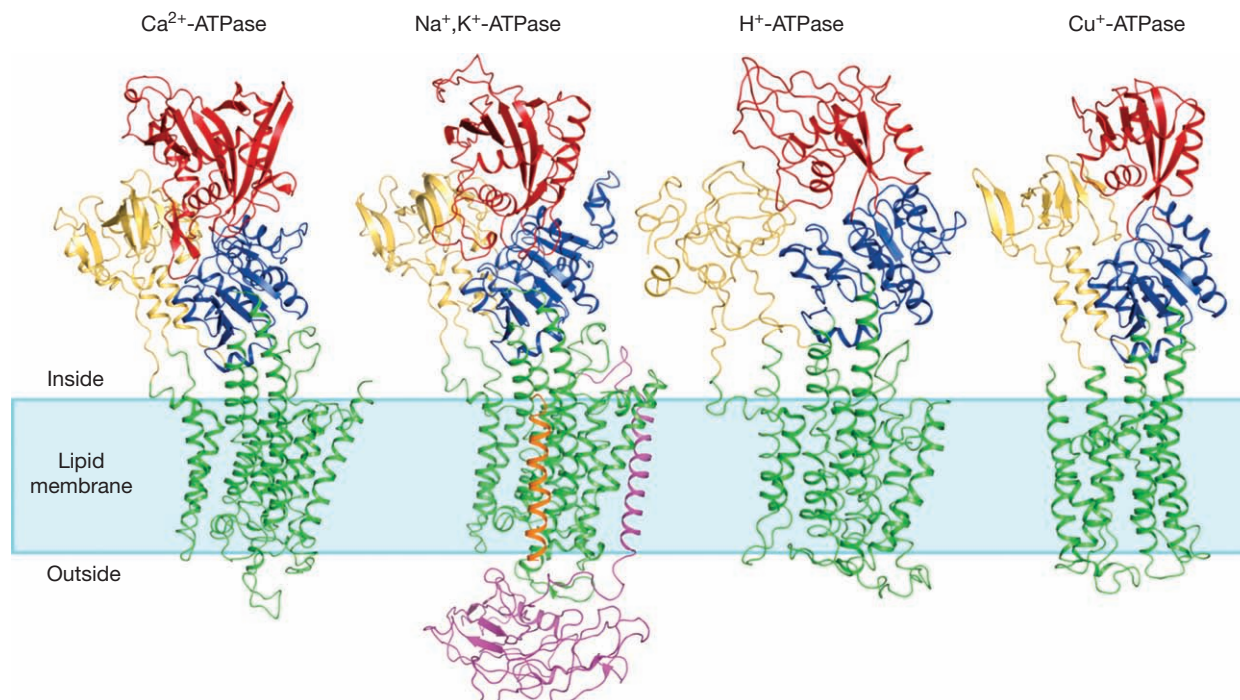


Figure 3 Crystal structures of the rabbit sarcoplasmic reticulum Ca^{2+} -ATPase, pig Na^+, K^+ -ATPase, H^+ -ATPase from *Arabidopsis thaliana* and Cu^+ -ATPase from *Legionella pneumophila*. Transmembrane helices are depicted in green, N-domain in red, P-domain in blue, and A-domain in yellow. The β -subunit and the γ -subunit of the Na^+, K^+ -ATPase is shown in purple and orange, respectively. From Olesen C, Sørensen TLM, Nielsen RC, Møller JV, and Nissen P, (2004) Dephosphorylation of the calcium pump coupled to counter occlusion. *Science* 306: 2251–2255; Morth JP, Pedersen BP, Toustrup-Jensen MS, et al. (2007) Crystal structure of the sodium–potassium pump. *Nature* 450: 1043–1049; Pedersen BP, Buch-Pedersen MJ, Morth JP, Palmgren MG, and Nissen P, (2007) Crystal structure of the plasma membrane proton pump. *Nature* 450: 1111–1114; and Gordon P, Liu X-Y, Skjærringe T, et al. (2011) Crystal structure of a copper-transporting PIB-type ATPase. *Nature* 475: 59–64.

glycosides), naturally found in many plants and secreted as toxins by several species of toads, snakes, or fireflies. These compounds have been used in the treatment of congestive heart failure for more than two centuries as they are able to cause a local increase in the intracellular Na^+ concentration, followed by a slowdown of the secondary active transport of Ca^{2+} out of the cell and a consequent increase of the contractility of the cardiac muscle. Recently, compounds derived from cardiotonic steroids have attracted attention also as anti-cancer drugs for breast cancer treatment. It has been shown that cardiotonic steroids interact with the extracellular part of the Na^+, K^+ -ATPase (preferably with its E2P conformation), most likely preventing the re-entry and occlusion of K^+ ions by blocking the extracellular cation exit pathway and stabilizing an E2P-like conformation that antagonizes K^+ binding.

The Gastric Proton–Potassium Pump

The gastric H^+, K^+ -ATPase is located in the secretory network within the gastric parietal cells of the stomach. It transports protons from the cytoplasm to the extracellular side of the membrane in exchange for K^+ ions, causing an acidification of the stomach that improves digestion and functions as a defensive mechanism against invasive bacteria. Interestingly, the gastric pump is the only member of the P-type

ATPase family with a variable counter-transport stoichiometry per ATP that is changing from 2:2 at pH 6.1 to 1:1 at a pH below 3.

Similarly to the overall structure of the Na^+, K^+ -ATPase, the H^+, K^+ -pump is composed of two subunits – a catalytic α -subunit with 10 transmembrane helices and a β -subunit with one transmembrane segment and six or seven N-linked glycosylation sites on the extracellular side that are required for assembly, maturation, and proper sorting of the pump. Due to the lack of a high-resolution crystal structure of the H^+, K^+ -ATPase, the exact location of the β -subunit remains elusive. It was proposed that the functional form of the gastric H^+, K^+ -ATPase is an oligomer composed of two α - and two β -subunits, where one $\alpha\beta$ -complex is in the E1 state and the second one in the E2 state.

In contrast to the Na^+, K^+ -ATPase, the N-terminal tail of the β -subunit of the gastric pump is in direct contact with the P-domain of the catalytic α -subunit. This may play a role in stabilizing the enzyme conformation and in preventing the reverse reaction of the transport cycle (i.e., the E2P to E1P transition), which enables the generation of large H^+ gradients across the gastric parietal cell membrane. Similarly to the Ca^{2+} - and Na^+, K^+ -ATPase, the ion-binding cavity of the H^+, K^+ -ATPase can be found between transmembrane segments M4, M5, M6, and M8. A lysine residue (Lys 791) on the fifth transmembrane segment seems to be important for the ion specificity of the pump.

The transport cycle of the H^+,K^+ -ATPase is very similar to that of the Ca^{2+} -ATPase, with the E1P to E2P transition being accompanied by the release of H^+ into the extracellular space (allowed by the reorientation of M4 and M6 helices) and by the binding of K^+ ions. Conversely, during the E2P to E1 transition, K^+ ions are transported into the cytoplasmic side of the membrane where new protons become bound to the pump, accompanied by two magnesium ions. There is evidence that the structural changes during the H^+,K^+ -ATPase transport cycle are somewhat smaller than changes during the reaction cycle of the Ca^{2+} -ATPase.

The H^+,K^+ -ATPase is directly involved in the onset of several diseases, including peptic ulcers and gastroesophageal reflux diseases. The most widely used agents in the treatment of these conditions are sulphur-containing benzimidazole derivatives forming a group of so-called proton pump inhibitors, represented by omeprazole as the most prominent member. These compounds are able to bind to the E2 conformation of the pump from the extracellular side of the membrane and, upon their activation by acidic pH, form disulphide bonds with cysteine thiol groups of the H^+,K^+ -ATPase. This results in the formation of an irreversibly inhibited complex as the pump remains trapped in an inactive E2P state where the re-entry of K^+ -ions from the luminal side of the membrane is blocked. Another group of the gastric pump inhibitors is formed by the potassium-competitive acid blockers, consisting of the Schering compound (SCH28080) and its derivatives. These compounds are able to noncovalently bind to the pump from the luminal side of the membrane in a reversible way, and act without the need of activation and thus in a much faster way than the proton pump inhibitors.

The Plasma Membrane Proton Pump

The plasma membrane H^+ -ATPase is found only in plants and fungi, where it acts as a functional analog of the Na^+ , K^+ -ATPase, energizing the plasma membrane potential and the secondary active transport systems. Although the Ca^{2+} -ATPase and the proton pump share only a low sequence homology, their overall structure is remarkably similar (Figure 3). The N-domain of the H^+ -ATPase is slightly smaller than that of the calcium pump, and analogously to the Asp 351 residue in the Ca^{2+} -ATPase, an Asp 329 becomes phosphorylated during the pump transport cycle. In contrast to other P-type ATPases, no counter-transport of ions has been observed in the H^+ -ATPase. On the basis of a crystal structure of the proton pump from *Arabidopsis thaliana* [8], a mechanism for proton binding has been proposed, consisting of a proton-binding site buried between the transmembrane segments M2, M4, and M6 featuring also an aqueous cavity. The proton specificity of the H^+ -ATPase is maintained by a protonated aspartate paired with an asparagine residue (Asp 684 and Asn 106) from the membrane domain.

Heavy Metal P-Type ATPases

P-type ATPases from subfamily I are involved in the homeostasis of heavy metals such as copper or zinc, and can also transport nonphysiological substrates such as Ag^+ , Cd^{2+} , or

Pb^{2+} . In contrast to other members of the P-type ATPase family, heavy metal pumps have only six to eight transmembrane helices and a smaller A-domain (Figure 3). Their characteristic feature is the presence of several cytosolic and transmembrane N- or C-terminal metal-binding domains that are required for phosphorylation and account for the substrate specificity, trafficking of the pump, and regulation of its functional cycle. Copper transporting P-type ATPases (subgroup IB-1) are the most widely studied heavy metal ATPases as their mutation impairs copper homeostasis and results in the onset of several neurodegenerative disorders such as Parkinson's, Alzheimer's, and prion-related diseases or the genetically inherited Menkes and Wilson diseases.

Flippases

Members of the P-type ATPase subfamily IV are called flippases because they are involved in the translocation of phospholipids (such as phosphatidylserine and phosphatidylethanolamine) from the exoplasmic to the cytoplasmic leaflet of membrane bilayers. This maintains the physiologically important asymmetric distribution of lipids across the membrane, necessary, for example, for the formation of new lipid vesicles and for membrane trafficking. While the substrate-binding domain of other P-type ATPases contains cation-interacting amino acid residues, transmembrane helices M4 and M6 in flippases have hydrophobic amino acid residues that can interact with hydrophobic moieties of phospholipids. Mutations in human flippases are associated with severe liver diseases such as the Byler disease.

See also: **Bioenergetics:** ATP Synthesis: Mitochondrial Cyanide-Resistant Terminal Oxidases; P-Type Pumps: Na^+ , K^+ -ATPase; Respiratory Chain and Adenosine Triphosphate Synthase.

Further Reading

- Axelsen KB and Palmgren MG (1998) Evolution of substrate specificities in the P-type ATPase superfamily. *Journal of Molecular Evolution* 46: 84–101.
- Gordon P, Liu X-Y, Skjærvinge T, et al. (2011) Crystal structure of a copper-transporting PIB-type ATPase. *Nature* 475: 59–64.
- Jensen AML, Sørensen TLM, Olesen C, Møller JV, and Nissen P (2006) Modulatory and catalytic modes of ATP binding by the calcium pump. *EMBO Journal* 25: 2305–2314.
- Morth JP, Pedersen BP, Toustrup-Jensen MS, et al. (2007) Crystal structure of the sodium-potassium pump. *Nature* 450: 1043–1049.
- Olesen C, Picard M, Winther AML, et al. (2007) The structural basis of calcium transport by the calcium pump. *Nature* 450: 1036–1042.
- Olesen C, Sørensen TLM, Nielsen RC, Møller JV, and Nissen P (2004) Dephosphorylation of the calcium pump coupled to counter occlusion. *Science* 306: 2251–2255.
- Pedersen BP, Buch-Pedersen MJ, Morth JP, Palmgren MG, and Nissen P (2007) Crystal structure of the plasma membrane proton pump. *Nature* 450: 1111–1114.
- Shin JM, Munson K, Vagin O, and Sachs G (2009) The gastric HK-ATPase: Structure, function, and inhibition. *European Journal of Physiology* 457: 609–622.
- Toyoshima C, Nakasako M, Nomura H, and Ogawa H (2000) Crystal structure of the calcium pump of sarcoplasmic reticulum at a 2.6 Å resolution. *Nature* 405: 647–655.
- Yatime L, Buch-Pedersen MJ, Musgaard M, et al. (2009) P-type ATPases as drug targets: Tools for medicine and science. *Biochimica et Biophysica Acta* 1787: 207–220.

Structure of Respiratory Complex I

J M Berrisford and L A Sazanov, Medical Research Council Mitochondrial Biology Unit, Cambridge, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Iron–sulfur clusters Ensembles of iron and sulfur which can be oxidized or reduced through the donation or acceptance of electrons. Within complex I, only two forms are found: either 4Fe–4S or 2Fe–2S.

Quinone (Q) A hydrophobic molecule containing a conjugated ring structure at one end and a tail from several isoprenoid units. These compounds act as electron

acceptors/donors between several of the respiratory complexes.

Reactive oxygen species (ROS) Highly reactive oxygen molecules due to the presence of an unpaired electron. These molecules have a role in signaling and also cause damage to a cell through reactions with biological molecules.

Thermus thermophilus A Gram-negative extremely thermophilic bacteria which has optimal growth at around 70 °C.

Introduction

Complex I (NADH:ubiquinone oxidoreductase, EC 1.6.5.3) is the first enzyme of the mitochondrial and bacterial respiratory chains. It catalyzes the transfer of two electrons from nicotinamide adenine dinucleotide reduced form (NADH) to quinone (Q), coupled to the translocation of four protons (current consensus value) across either the mitochondrial inner or bacterial cytoplasmic membrane according to the equation: $\text{NADH} + \text{Q} + \text{H}^+ + 4\text{H}^+_{\text{in}} \rightarrow \text{NAD}^+ + \text{QH}_2 + 4\text{H}^+_{\text{out}}$ (where ‘in’ corresponds to the mitochondrial matrix or the bacterial cytoplasm and ‘out’ corresponds to the mitochondrial or the bacterial intermembrane space). This produces nicotinamide adenine dinucleotide (NAD^+), which is used to regenerate NADH, and quinol, which is used as a substrate by other complexes in the respiratory chain. In completing this reaction, complex I contributes approximately 40% of the proton-motive force required for the generation of adenosine triphosphate (ATP) by ATP synthase. Different quinones are used as electron acceptors by complex I from different species. The mitochondrial enzyme uses ubiquinone, complex I from *Thermus thermophilus* uses menaquinone, and the complex in *Escherichia coli* uses either ubiquinone or menaquinone, depending on the growth conditions. Mutations in various complex I subunits lead to many human neurodegenerative diseases. Complex I has also been implicated as a source of reactive oxygen species (ROS) within the mitochondria. This is caused by electron leak during complex I turnover and ROS production increases when the enzyme is inhibited. ROS can damage mitochondrial DNA and may be one of the causes of aging. The sporadic form of Parkinson’s disease may, in part, be caused by an increase in ROS due to malfunctioning complex I.

Overall Organization

Mitochondrial complex I is one of the largest known membrane complexes. Bovine complex I consists of 45 different subunits with a combined mass of about 980 kDa. The prokaryotic enzyme is much simpler, with a total mass of around 550 kDa, and consists of about 14 subunits conserved from

bacteria to humans. Both the mitochondrial and prokaryotic enzymes contain a core of eight iron–sulfur clusters (six 4Fe–4S and two 2Fe–2S) and a noncovalently bound flavin mononucleotide (FMN). Based on the conservation of core subunits and prosthetic groups, the prokaryotic enzyme can be considered a minimal model of the larger eukaryotic enzyme. The overall structure of the enzyme has been analyzed in various species by electron microscopy, revealing that complex I is L shaped and consists of two arms. The hydrophilic peripheral arm contains conserved subunits Nqo1–6 and 9 (*T. thermophilus* nomenclature, which will be used throughout for simplicity) and the hydrophobic membrane arm contains Nqo7–8 and Nqo10–14. The membrane arm is embedded within the lipid bilayer and the hydrophilic arm protrudes almost perpendicular from the membrane into the bacterial cytoplasm (or mitochondrial matrix). The NADH-binding site, FMN, and all iron–sulfur clusters are found within the hydrophilic arm. The nomenclature of equivalent subunits in bovine and *E. coli* complex I is shown in Table 1, as are the locations of the prosthetic groups.

The X-ray crystal structure of the hydrophilic arm of complex I from *T. thermophilus* has been determined at 3.1 Å resolution, in both the oxidized and reduced, NADH-bound, forms. These structures revealed the organization of different subunits and identified the spatial location of the redox centers (Figure 1). In *T. thermophilus*, the hydrophilic arm contains eight subunits: Nqo1–6, Nqo9 and, unexpectedly, an additional subunit, which is not part of the *Nqo* operon, Nqo15.

Overall, the hydrophilic domain of *T. thermophilus* complex I is Y shaped, which is consistent with recent electron microscopy (EM) reconstructions of the complex from *Yarrowia lipolytica*, *Aquifex aeolicus*, and bovine, suggesting a high level of structural conservation within this enzyme. Subunits Nqo4 and Nqo6 form the bottom of the structure, where they interface with the hydrophobic domain and form part of the quinone-binding site. The NADH/FMN-binding site is located in subunit Nqo1 at the top of the enzyme, some 100 Å from Nqo4 and Nqo6. Electron transfer between the two substrate-binding sites occurs via the chain of iron–sulfur clusters (Figure 1). The complex appears to have evolved from several smaller subcomplexes, or building blocks. Subunit Nqo2 is related

to thioredoxins, Nqo9 to ferredoxins, the N-terminus of Nqo3 to FeFe-hydrogenases, the C-terminus of Nqo3 to molybdopterin-containing enzymes, and Nqo4 and Nqo6 to NiFe-hydrogenases. Larger building blocks can be traced to NAD⁺-reducing hydrogenases (subunits Nqo1–3) and membrane-bound hydrogenases (subunits Nqo4–6 and 9 plus most of the membrane domain subunits).

The lack of structural information on the hydrophobic domain means less is known about its organization. Fragmentation studies have shown that directly below Nqo4 and Nqo6

of the hydrophilic arm, subunits Nqo7, Nqo8, Nqo10, and Nqo11 form a bundle which has been implicated in quinone binding. Subunits Nqo12, Nqo13, and Nqo14 are toward the distal end of the hydrophobic domain, with Nqo12 the furthest away from the hydrophilic arm. These subunits share sequence similarity with Na⁺ or K⁺/H⁺ antiporters of the Mrp family. This homology suggests that these three subunits are involved in pumping protons across the membrane. As Nqo12, Nqo13, and Nqo14 are some distance away from the quinone-binding site, conformational changes are likely to be necessary during catalysis to link electron transfer down the hydrophilic arm to proton pumping in the distal end of the hydrophobic arm.

Table 1 A list of the conserved subunits present in *T. thermophilus* and *E. coli* complex I and their equivalents in bovine complex I (the prosthetic groups present in each subunit are also listed, where an N prefix designates an iron–sulfur cluster)

<i>T. thermophilus</i>	<i>E. coli</i>	Bovine	Prosthetic group
Nqo1	NuoF	51 kDa	FMN, N3
Nqo2	NuoE	24 kDa	N1a
Nqo3	NuoG	75 kDa	N1b, N4, N5, and in some species N7
Nqo4	NuoD – fused with NuoC	49 kDa	
Nqo5	NuoC – fused with NuoD	30 kDa	
Nqo6	NuoB	PSST	N6a, N6b
Nqo7	NuoA	ND3	
Nqo8	NuoH	ND1	
Nqo9	NuoI	TYKY	N2
Nqo10	NuoJ	ND6	
Nqo11	NuoK	ND4L	
Nqo12	NuoL	ND5	
Nqo13	NuoM	ND4	
Nqo14	NuoN	ND2	

Hydrophilic Domain and Redox Centers

Structures of the hydrophilic arm revealed the arrangement of all the known redox centers within the enzyme: the FMN and nine iron–sulfur clusters. (There are proposals that the membrane domain contains one or two bound quinone molecules acting as cofactors, but this idea has not been confirmed experimentally.) Electrons are passed down the iron–sulfur cluster chain from the substrate NADH to a quinone molecule in the vicinity of cluster N2. The arrangement of the clusters suggests a likely electron path, as the edge-to-edge distance between the redox centers should be within 14 Å for biologically significant rates of electron transfer. Seven of the Fe–S clusters form a connecting redox chain between the FMN and the quinone-binding site (Figure 1). Two electrons enter the complex from one NADH molecule, which has a midpoint redox potential of –320 mV at pH7, and are transferred as a hydride to FMN (midpoint potential –340 mV), which resides in subunit Nqo1. One electron is then transferred from FMN to a nearby

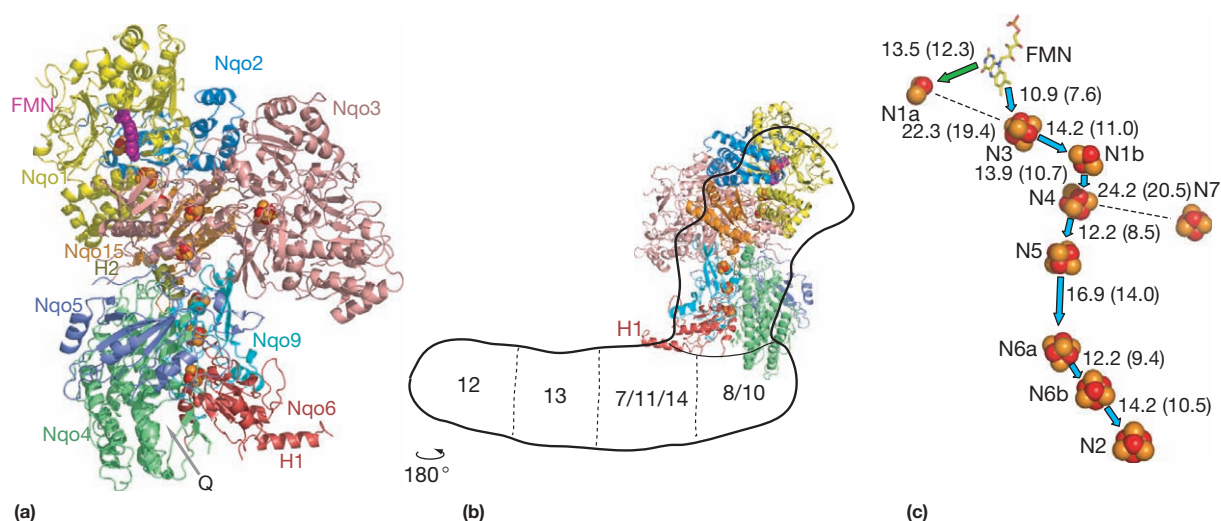


Figure 1 An overview of the structure (PDB 2FUG). (a) Side view, with the membrane arm likely to be beneath and extending to the right, in the direction of helix H1. Each subunit is colored differently; FMN is shown as magenta spheres, Fe atoms as red spheres and S atoms as yellow spheres. Helix H2 is shown in olive. (b) A view rotated by 180°, showing a model of a likely mode of attachment of the peripheral domain to the membrane domain. The outline of intact *E. coli* complex I comes from single particle analysis, and the model of subunit arrangement is derived from fragmentation studies. Numbers indicate hydrophobic Nqo subunits. (c) Arrangement of redox centers. The overall orientation is similar to that in (a). The main pathway of electron transport is indicated by blue arrows and a diversion to cluster N1a by a green arrow. The distances between the centers given in Å were calculated both center-to-center and edge-to-edge (shown in parentheses).

tetranuclear cluster N3, also located in Nqo1. From N3 the electron is passed to the binuclear cluster N1b and then on to tetranuclear clusters N4 and then N5, all coordinated within Nqo3. From N5 the electrons are transferred to the two tetranuclear clusters in Nqo9, N6a, and N6b, before being passed to the final tetranuclear cluster, N2, within Nqo6. N2 has the highest midpoint potential of the clusters (~ -100 mV) and passes its electron onto the quinone substrate, ubiquinone (+110 mV) or menaquinone (-80 mV), depending on the organism. Most of other electron paramagnetic resonance (EPR)-visible clusters in complex I have potentials around -250 mV, but several clusters are EPR silent and may have potentials below -400 mV. Despite its length, this redox chain is very efficient – EPR experiments on freeze-quenched samples have shown that first electron from NADH arrives at N2 within $90 \mu\text{s}$, close to the theoretical limit. Although the exact site of quinone binding is still under debate, cluster N2 lies only $10 \text{ \AA}$ from a likely quinone-binding channel formed between Nqo6 and the adjacent Nqo4 subunit (Figure 1). An additional tetranuclear cluster N7 is present in Nqo3 in *T. thermophilus* but with $20 \text{ \AA}$ between this cluster and its closest neighbor, N4, it is too removed from the redox chain to be part of the electron transfer pathway. As cluster N7 is also only conserved in a small number of bacterial species, it may represent an evolutionary remnant from molybdopterin-containing enzymes.

Binuclear cluster N1a, within subunit Nqo2, is fully conserved but it does not form part of the main redox chain (Figure 1). However, it is within electron transfer distance from the flavin ($12 \text{ \AA}$). This cluster has a midpoint potential lower than the other clusters at -370 mV (in bovine enzyme, the best studied example). This would make it unlikely to accept the first electron from fully reduced FMNH_2 , but it may accept an electron from flavosemiquinone, with its lower potential, after the first electron from FMNH_2 is donated to N3. When the downstream electron transfer chain is ready to accept the second electron (probably after quinone reduction), the electron on N1a could then be returned to the FMN moiety and then passed to N3 and down the redox chain. This mechanism would preclude a flavosemiquinone intermediate from existing for any significant period of time, as the two electrons from NADH could be passed from FMN to N1a and N3 almost simultaneously. This would reduce the prevalence of a reaction of the flavosemiquinone intermediate with oxygen leading to damaging ROS production at the solvent-exposed flavin site. As the likely rate-limiting step of the complex I reaction is the reduction of quinone, most iron-sulfur clusters would remain reduced during catalysis, which may promote flavosemiquinone production. This would make N1a's role as an electron storage cluster essential, as it is shielded from the solvent.

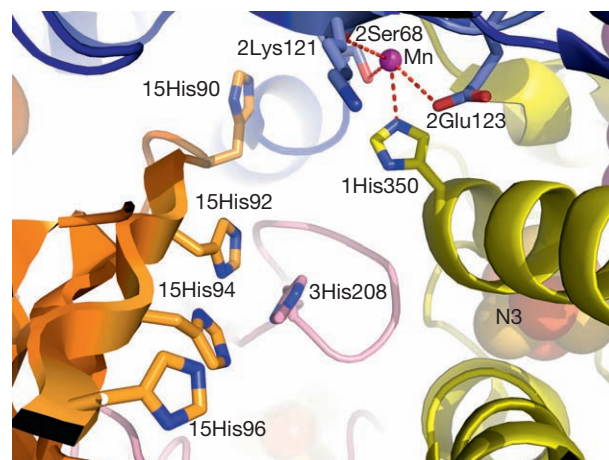
Frataxin-Like Subunit in Complex I

An unexpected finding from the structure of the hydrophilic domain of *T. thermophilus* complex I was the presence of an additional subunit, Nqo15. This protein is not part of the *Nqo* operon and so far its homologs have only been identified in a few extremophilic prokaryotic species. One possible

role for this subunit could be to stabilize the structure of the hydrophilic arm, as it is rather narrow in the middle and Nqo15 interacts with many subunits in this area. Unexpectedly, the fold of Nqo15 is nearly identical to that of frataxin, despite very low sequence homology ($\sim 11\%$ identity). Frataxin is involved in the biogenesis of iron-sulfur clusters in mitochondria and bacteria, although its exact role is still under debate. It interacts with the Isc machinery, possibly by delivering iron, and may have a role in iron storage. The *T. thermophilus* genome contains no other, sequential homolog to frataxin, but the Isc genes are present. As Nqo15 is not incorporated into the *Nqo* operon, this protein might have a double role in *T. thermophilus* – as part of complex I and as frataxin analog in the cytoplasm.

The Nqo15 and frataxin fold consists of a six-stranded antiparallel β -sheet with two α -helices on one face of the β -sheet. Nqo15 interacts with the rest of the complex mainly via the other exposed face of the β -sheet. This face of the β -sheet is also suggested to be involved in interactions of frataxin with its protein partners. *T. thermophilus* complex I provides the first example of the frataxin-like protein permanently bound to its partner and so may illustrate a mode of such interaction.

Frataxins contain several fully conserved acidic residues on the edge of the first β strand and α -helix and these residues are proposed to be involved in iron binding. Some of these residues are not present within Nqo15, suggesting it may not bind metal ions in the same manner. However, Nqo15 has a series of four histidine residues along one face of the fifth β strand. Within *T. thermophilus* complex I, these histidines face a channel formed between subunits Nqo1, 2, 3, and 15 (Figure 2). The side chains of several conserved histidines and glutamates from Nqo1–3 also face into this channel, forming at least two potential metal-binding sites. Divalent cations, including manganese, a close mimic of iron, were identified as bound inside this channel in the hydrophilic arm structures



(Figure 2). As the end of this channel lies in close proximity to the iron–sulfur clusters N1a and N3 (~ 10 Å), this leads to the intriguing possibility that Nqo15 may be involved in iron binding, akin to frataxin, and have a role in the regeneration of these clusters. They are the nearest to the FMN, a known source of ROS, and so may be the first to be damaged during turnover.

NADH Binding

The determination of the structures of the ligand-free oxidized and NADH-bound reduced forms of *T. thermophilus* complex I hydrophilic domain allowed comparisons to be made between the structures. The substrate NADH binds in the cavity within subunit Nqo1 at the top of the hydrophilic arm, adjacent to the FMN moiety (Figure 3). Complex I does not contain a classical Rossmann nucleotide-binding fold and instead has a modified fold which binds both FMN and NADH through the incorporation of an additional glycine-rich loop. The FMN/NADH-binding pocket is highly conserved with the residues involved in interactions with both of the cofactors retained between species as varied as bacteria and humans.

Within the binding pocket, the nicotinamide ring of NADH stacks against the exposed face of the isoalloxazine ring of the FMN moiety, in a manner similar to that seen in other nucleotide-binding flavoenzymes (Figure 3). This positions the B-face of the nicotinamide ring against the *re*-face of the isoalloxazine ring allowing the 4B hydrogen of NADH to transfer as a hydride to the N5 atom of FMN. The transfer of the hydride is fast in complex I, and the distance between these two atoms is slightly shorter than usual, at 3.2 Å. The short distance between the two nucleotides may be caused by the close proximity of the C β atom of Nqo1 Glu97, which is within van der

Waal's contact distance with the C4N atom of the bound NADH, forcing the two nucleotides together.

Upon binding of NADH, a shift of around 1.5 Å in the loop formed by Nqo1 residues 202–207 toward the bound NADH is observed. This positions the side chain of Phe205 1.7 Å closer to the adenine ring, promoting a stacking interaction, and a hydrogen bond forms between the side chain of Lys202 and the phosphate moiety of the NADH molecule. Further stacking interactions are formed with Phe70 and Phe78. Complex I, compared to other dehydrogenases, is unusual in its ability to use deamino-NADH as a substrate. This is due to the relatively open NADH-binding pocket in which N6A of the adenine ring (which is an oxygen in deamino-NADH) forms no interactions with the protein.

Overall, there are no large-scale changes in the rest of the structure upon reduction by NADH. However, reproducible shifts of around 1 Å can be seen in a 4-helix bundle in Nqo4 (residues 110–297) and helices H1 and H2 in Nqo6 (residues 18–35 and 143–160). Due to the proximity of these structural elements to the membrane domain, these movements may be propagated into the membrane arm and form part of the catalytic cycle, eventually resulting in proton pumping. Also, the number of close contacts between subunits is decreased by about 7–8% upon reduction, indicating slight loosening of the structure. This is consistent with the decreasing number of cross-links between subunits, observed upon reduction of complex I from several species. Without the structure of the whole complex it is difficult to draw firm mechanistic conclusions. However, a similar degree of change in cross-linking patterns upon reduction was observed both for the intact complex and for the isolated hydrophilic domain. As the middle part of the hydrophilic domain (between NADH and quinone-binding sites) does not undergo any conformational changes, it is likely that observed shifts of Nqo4/Nqo6 helices are driven by changes in the redox state of the nearby clusters (N2 and N6a/b).

A surprising result was seen when comparing the coordination of the terminal iron–sulfur cluster N2 in the oxidized and reduced crystal structures. N2 is coordinated by an unusual tandem cysteine motif consisting of consecutive Cys45 and Cys46 from Nqo6, with Cys111 and Cys140 completing the ligation (Figure 4). The tandem cysteine motif is fully conserved in complex I and so is likely to be important for its function. The structure revealed strained geometry in the side chains of these residues. In the oxidized form of the enzyme, clear electron density was observed for both the side chains of Cys45 and Cys46 (Figure 4(a)). Upon full reduction of the hydrophilic domain by NADH, the electron density for the side chain of Cys46 disappears, suggesting that this residue has disconnected from the cluster (Figure 4(b)). However, when crystals of the hydrophilic domain were partially reduced with dithionite, the electron density for the side chain of the neighboring cysteine, Cys45, was absent and the side chain of Cys46 had clear connecting density to the cluster (Figure 4(c)). Similar coordination was observed in crystals reduced by NADH but exposed to air for longer period before cryo-cooling – it is likely that in both cases some of the clusters upstream were oxidized, while N2 remained reduced due to its higher potential (as observed by EPR under similar conditions). Thus, upon reduction of N2 one of the tandem cysteines is disconnected,

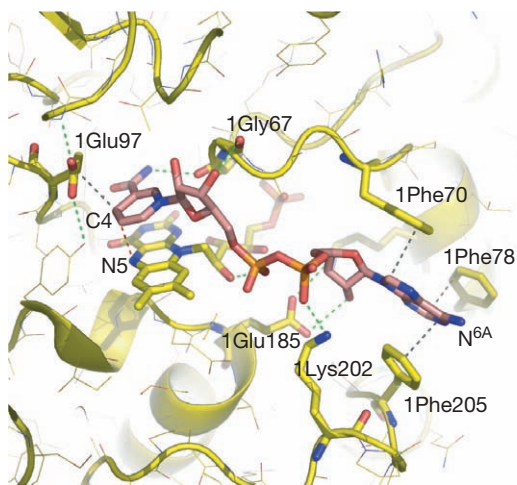


Figure 3 The NADH- and FMN-binding site. The site is viewed from the solvent-exposed side. FMN and residues involved in NADH binding are shown as sticks with carbon in yellow and NADH with carbon in salmon. Hydrogen bonds are shown as dotted lines in green, stacking interactions in gray, and hydride transfer path in red. Van der Waal's contact between 1Glu97 and C4N of NADH is shown in gray. Prefixes before residue names indicate the subunit number.

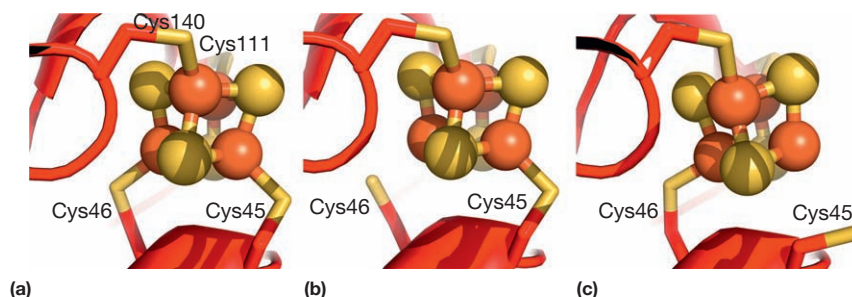


Figure 4 The coordination of iron–sulfur cluster N2 within Nqo6 by Cys45, Cys46, Cys111, and Cys140. The side chains of the adjacent residues Cys45 and Cys46 both ligate the cluster in the oxidized state (a), causing their side chains to have a strained geometry. In the structure of the NADH-reduced state (fully reduced redox chain), the side chain of Cys46 appears disconnected (b), whereas in the structure of the dithionite-reduced state (redox chain partially reduced, including cluster N2), the side chain of Cys45 is disconnected (c).

the choice being dependent on the redox state of neighboring clusters (N6a/b). It is likely that the disconnected cysteine side chains become protonated and that N2 is instead ligated by a solvent molecule (water or OH, akin to aconitase). This is consistent with the well-known redox-Bohr effect – the midpoint redox potential of cluster N2 is pH dependent, that is, it (or nearby side chain) becomes protonated upon reduction. These results suggest that the coordination of N2 is changed when it is reduced and it is possible to envisage that Cys45 and Cys46 are becoming sequentially disconnected from N2 and protonated, before passing their protons on to the bound quinone molecule. The possible mechanism may involve both protonation/de-protonation of tandem cysteines and conformational coupling, propagating movements of the 4-helix bundle in Nqo4 and helices H1 and H2 in Nqo6 into the membrane domain. The unusual coordination by tandem cysteines may also play a role in these conformational changes by providing protein flexibility in the vicinity of cluster N2.

Mutations in Complex I Subunits

Numerous mutations in complex I have been implicated in human diseases. The determination of the structure of the hydrophilic arm from *T. thermophilus* enabled mapping of these mutations onto the structure and allowed identification of their possible role in catalysis. In general, disease-causing mutations are in regions around the substrate-binding pockets, iron–sulfur clusters, or the interface with the hydrophobic domain. The mutations lead to a variety of phenotypes which can be classified as resulting in reduced NADH dehydrogenase activity, reduced complex I levels in the cell, instability of complex I, change in affinity to quinone or sensitivity to quinone-site inhibitors, or altering the properties of iron–sulfur clusters. All of these effects reduce the efficiency of complex I. One such example is the mutations of Y204C and C206G in the 51 kDa subunit in humans (Nqo1 in *T. thermophilus*). These mutations result in only 18% of complex I activity remaining and cause hypotonia, ataxia, and infant death. The mutations can be mapped to Tyr180 and Cys182 in Nqo1, which are both adjacent to the nucleotide-binding sites. In fact, Tyr180 forms a hydrogen bond with Glu97, which is involved in promoting interactions between

NADH and FMN (Figure 3). Mutations in either of these residues would then be likely to have an effect in binding of the substrate NADH to the enzyme. Another example is the mutation of Leu231 to a valine in the human 75 kDa subunit (Nqo3), where only 25% of complex I activity remains, leading to Leigh syndrome and infant death. Leu231 is equivalent to Leu235 in Nqo3 and is directly facing cluster N4, so any mutation in this region is likely to evoke either structural or redox changes in the cluster.

Supernumerary Subunits in Eukaryotic Complex I

Bovine complex I has been shown to contain 45 subunits, with the seven core hydrophobic subunits encoded within the mitochondria and all other subunits encoded in the nucleus. Apart from the core 14 subunits, which have equivalents in the bacterial enzyme, little is known about the roles of the supernumerary subunits in higher eukaryotic complex I. No additional prosthetic groups have been detected in the supernumerary subunits; however, several of these subunits have been implicated in diseases in humans. For example, in human complex I, a splice mutation in NDUFA11 causes encephalomyopathy, and a 5 base pair (bp) duplication eliminating a phosphorylation site in NDUF54 resulted in Leigh-like symptoms due to a deficiency in complex I. Several subunits are homologous to other proteins with known functions, but it is not clear what role they play in complex I. Many smaller subunits classify as single transmembrane domain proteins and may assist in assembly and stability of the membrane domain. The total number of these so-called supernumerary subunits was only finalized recently and further work is required to understand how these subunits affect the function of complex I.

Conclusion

The recent work on the hydrophilic domain from *T. thermophilus* has provided the first structural information on the organization of eight core subunits and nine iron–sulfur clusters. It has identified the electron transfer pathway down a series of seven of the iron–sulfur clusters from the FMN site to a quinone site at the interface with the hydrophobic domain. Analyses of the

structures of the reduced NADH-bound enzyme have defined the mode of NADH binding and revealed unexpected changes in the coordination of the terminal iron–sulfur cluster N2, which are likely implicated in the mechanism. Comparisons of the sequences of complex I have allowed several disease-causing mutations in the human enzyme to be explained through analysis of the structure. However, the structure of the hydrophilic arm is only part of the story and we wait with bated breath for the full structure to become available and provide further insight into the mechanism of this remarkably complex enzyme.

See also: **Bioenergetics:** Bioenergetics: General Definition of Principles; Cytochrome *bc*₁ Complex (Respiratory Chain Complex III); Cytochrome Oxidases, Bacterial; F-ATPases; Iron–Sulfur Proteins; Proton Pumping in the Respiratory Chain; Respiratory Chain and Adenosine Triphosphate Synthase; Respiratory Chain Complex I; Respiratory Chain Complex II and Succinate: Quinone Oxidoreductases; Respiratory Chain Complex IV.

Further Reading

- Berrisford JM and Sazanov LA (2009) Structural basis for the mechanism of respiratory complex I. *Journal of Biological Chemistry* 284: 29773–29783.
- Brandt U (2006) Energy converting NADH:quinone oxidoreductase (complex I). *Annual Review of Biochemistry* 75: 69–92.
- Clason T, Ruiz T, Schagger H, et al. (2010) The structure of eukaryotic and prokaryotic complex I. *Journal of Structural Biology* 169: 81–88.
- Friedrich T (2001) Complex I: A chimaera of a redox and conformation-driven proton pump? *Journal of Bioenergetics and Biomembranes* 33: 169–177.
- Hirst J (2010) Towards the molecular mechanism of respiratory complex I. *Biochemical Journal* 425: 327–339.
- Ohnishi T (1998) Iron–sulfur clusters/semiquinones in complex I. *Biochimica et Biophysica Acta* 1364: 186–206.
- Sazanov LA (2007) Respiratory complex I: Mechanistic and structural insights provided by the crystal structure of the hydrophilic domain. *Biochemistry* 46: 2275–2288.
- Sazanov LA and Hinchliffe P (2006) Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*. *Science* 311: 1430–1436.
- Vinogradov AD and Grivennikova VG (2001) The mitochondrial complex I: Progress in understanding of catalytic properties. *IUBMB Life* 52: 129–134.
- Walker JE (1992) The NADH–ubiquinone oxidoreductase (complex I) of respiratory chains. *Quarterly Reviews of Biophysics* 25: 253–324.
- Yagi T and Matsuno-Yagi A (2003) The proton-translocating NADH–quinone oxidoreductase in the respiratory chain: The secret unlocked. *Biochemistry* 42: 2266–2274.

Substrate Binding, Catalysis, and Product Release

W W Cleland, University of Wisconsin, Madison, WI, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Haldane relationship An equation relating the equilibrium constant to the kinetic constants of an enzymatic reaction.

Isotope effect The effect of isotopic substitution, expressed as the ratio of the parameter for the light isotope to that for the heavy one.

Michaelis constant The concentration of a substrate that gives half of the maximum velocity of an enzymatic

reaction. It is the apparent dissociation constant in the steady state.

Substrate A molecule that is converted to a product during an enzyme-catalyzed reaction.

Turnover number The number of substrate molecules converted to product per enzyme molecule per unit time. The units are usually reciprocal seconds.

Enzymes catalyze reactions by forming complexes with their substrate(s), accelerating the chemical reaction by lowering the activation energy for the reaction, and then releasing the product(s). The enzyme has to have sufficient affinity for a substrate to form a complex with it at the physiological concentration of this molecule. Conversely, however, too great an affinity for the substrate usually results in too great an affinity for the product, so that product release becomes rate limiting. The maximum rate of an enzyme-catalyzed reaction is limited, in fact, by the relationship between the equilibrium constant and the kinetic constants called the Haldane relationship, and one element of this restriction is the affinity of the substrate.

The Haldane Relationship

As mentioned above, the Haldane relationship allows calculation of the maximum rate of an enzyme-catalyzed reaction. For the simple conversion of substrate A into product P, the initial rate of the forward reaction in the absence of P is

$$v_f = -d[A]/dt = V_1[A]/(K_a + [A]) \quad [1]$$

and that in the back reaction is

$$v_r = -d[P]/dt = V_2[P]/(K_p + [P]) \quad [2]$$

where V_1 and V_2 are maximum velocities in forward and reverse reaction, respectively, and K_a and K_p the Michaelis constants for A and P (i.e., apparent dissociation constants in the steady state).

The Haldane for this mechanism is the ratio of the V/K values in forward and reverse directions:

$$K_{eq} = [P]_{eq}/[A]_{eq} = (V_1/K_a)/(V_2/K_p) = V_1K_p/(V_2K_a) \quad [3]$$

With units of reciprocal time, is the apparent first-order rate constant for reaction at low substrate concentration, so eqn [3] is analogous to the one for a nonenzymatic reaction where K_{eq} would equal k_1/k_2 , that is, the ratio of rate constants in forward and reverse directions.

To optimize the turnover number of an enzyme (V_1/E_t) in this case, also known as k_{cat} , with units of reciprocal time, we can raise the two V/K values in constant ratio until one

V_1/K_a value (with units of $M^{-1}s^{-1}$) reaches the limit set by diffusion. V_1/K_a (or k_{cat}/K) is a second-order rate constant for productive combination of enzyme and substrate and subsequent reaction to give product, and the combination of enzyme and substrate can only go as fast as these two can diffuse together. This limit is $\sim 10^9 M^{-1}s^{-1}$ for small substrates, and somewhat less for larger ones like adenosine triphosphate.

Once the V_1/K_a value in eqn [3] is as high as it can go, the only way to increase V_1/E_t is to increase K_a as well. But K_a cannot exceed the physiological level of A, or the enzyme will not be at least half saturated. Thus, V_1/E_t has an upper limit, and enzymes that are part of metabolic pathways are optimized to operate at this limit. The actual turnover numbers vary, depending on K_{eq} and the physiological level of the substrate, but evolution has brought these enzymes to the limit set by the Haldane relationship. On the other hand, enzymes that are involved in control, as opposed to ones in metabolic pathways, often sacrifice speed for control characteristics, and do not have turnover numbers at the Haldane limit.

Catalysis

The catalytic process on an enzyme starts once the substrate is bound. As mentioned above, the enzyme has to have sufficient affinity to bind the substrate at its physiological concentration. The enzyme then has to lower the activation energy of the reaction sufficiently for it to go at a rate approaching the maximum given by the Haldane relationship. It is commonly said that the enzyme accomplishes this by binding the transition state structure more tightly than the ground state of the substrate, but this is really a definition of catalysis. Molecules resembling the transition state structure are often bound much more tightly than the substrate (see entry on transition state analogs). This is either because the geometry of the transition state is a better fit, or includes inherently tighter interactions (stronger hydrogen bonds, for example). Alternatively, the differential binding results because the substrate is destabilized when it is bound (sterically deformed, placed in an unfavorable electrostatic environment, etc.), but this destabilization is relieved in the transition state. Of course,

destabilization in one part of the substrate must be matched by tight binding in another part, or the substrate will not have sufficient affinity for the enzyme at its physiological level. Orotidine monophosphate (OMP) decarboxylase is a classic example, where a substrate analog with $-\text{COO}^-$ replaced by $-\text{O}^-$ binds 10^5 -fold more tightly to the enzyme. The difference is ascribed to charge repulsion between the carboxyl group of the substrate and an aspartate on the enzyme.

Free-Energy Profiles

The energetics of what happens during an enzymatic reaction is illustrated by a free-energy profile. **Figure 1** shows the free-energy profile at equilibrium for the nonenzymatic reaction of A to P. The activation energy, $\Delta G^\ddagger$, corresponds to the height of the barrier, and ΔG for the reaction ($= -RT \ln K_{\text{eq}}$) is the difference between the levels marked A and P. For a corresponding enzymatic reaction, one must decide what levels of A and P to use. **Figure 2** shows an equilibrium free-energy profile where the concentration of A is above its dissociation constant, so that the formation of EA has an equilibrium constant greater than unity, but where P is less than its dissociation constant, so that EP dissociation is favorable. $\Delta G^\ddagger$ is given by the height of the barrier above the level of EA, not E + A. The turnover number (V/E_t or k_{cat}) is determined by $\Delta G^\ddagger$.

If one picks different levels of A or P, the difference between E + A and EA, or E + P and EP, will differ as shown in **Figure 3**. The dissociation level of A (K_{ia}) is convenient, but then the level of P is determined by K_{eq} if an equilibrium free-energy profile is plotted, and will not usually match K_{ip} , its dissociation constant. Free-energy profiles may, of course, be plotted for nonequilibrium concentrations of reactants, but in any case it is critical to state the concentrations of reactions that are assumed in plotting the profile. When there are two or more substrates or products involved in the reaction, one must pick

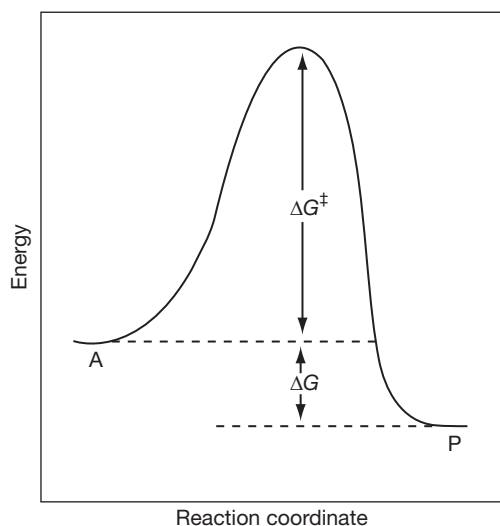


Figure 1 Free-energy profile at equilibrium for an uncatalyzed reaction. $\Delta G^\ddagger$ is the activation energy and $\Delta G = -RT \ln K_{\text{eq}}$, the energy difference between product and substrate.

suitable levels of all reactants (and state them!) in order to plot a free-energy profile.

So far, we have considered only simple free-energy profiles with a single transition state connecting EA and EP. In practice, enzymes have different conformations at the various stages of the catalytic cycle. Thus, they have open forms where the reactants are free to bind and dissociate, and closed forms in which catalysis takes place. The chemistry may also take place in more than one step, so that intermediates are present. **Figure 4** shows a profile where a conformation change converts the open EA form to a closed, precatalytic EA^* form. This undergoes a catalytic reaction to give an EI intermediate complex, which then is further converted to the closed EP^* form. A final conformation change gives the open EP form, from which P can dissociate.

Rate-Limiting Step for V/K

The two independent kinetic constants for an enzymatic reaction are the maximum velocity, V , and the ratio of V and the Michaelis constant, V/K . There is a separate V/K for each substrate. Each of these parameters varies separately with pH, ionic strength, temperature, and the concentrations of other substrates, products, or inhibitors. The Michaelis constant, K , while it measures the apparent dissociation constant of the

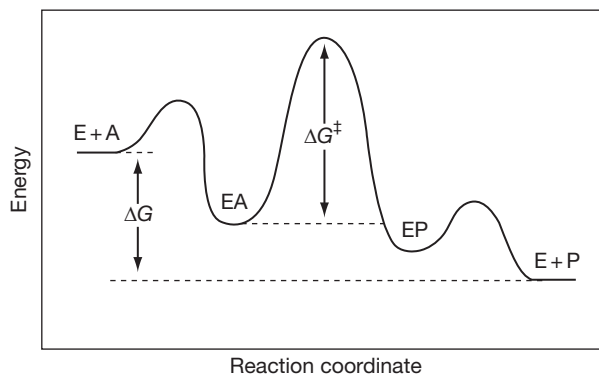


Figure 2 Free-energy profile at equilibrium for an enzymatic reaction. The symbols have the same meaning as in **Figure 1**.

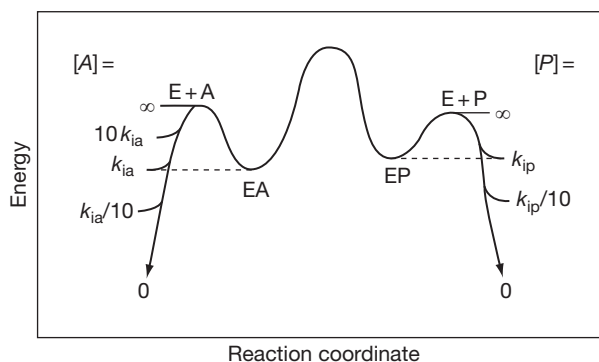


Figure 3 Free-energy profile showing the difference in the levels of (E + A) or (E + P) as the concentrations of A or P are changed. K_{ia} and K_{ip} are the dissociation constants of A from EA and of P from EP.

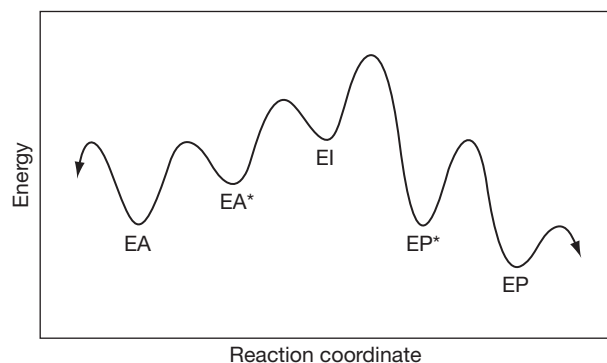


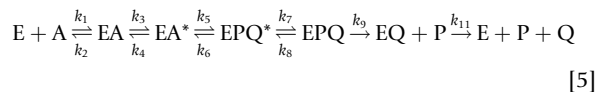
Figure 4 Free-energy profile showing conformation changes in the EA and EP complexes, as well as an intermediate between the two activated complexes. The levels of (E + A) and (E + P) are not shown.

substrate in the steady state, is not an independent constant, but just the ratio of V and V/K . Thus, when one speaks of 'rate-limiting steps' one must specify whether one is referring to V or V/K .

The rate-limiting step for V/K is the one with the highest barrier in the free-energy profile. V/K involves the combination of enzyme and substrate and reaction through the first irreversible step, which normally is release of the first product. Later steps that involve further conformation changes in the enzyme and release of other products do not affect V/K . Thus, the slow release of nicotinamide adenine dinucleotide from dehydrogenases, which often limits the maximum velocities of these enzymes, has no effect on V/K . As a result, the V/K value often is limited more by the chemistry of the reaction than V , and isotope effects on the chemistry are more fully expressed. Except for some slow mutants, full rate limitation by the chemical reaction is unusual, however, and the conformation changes that precede and follow the chemical reaction are usually partly rate limiting. The equation for an isotope effect on V/K is

$$^x(V/K) = (^xk + c_f + ^xK_{eq}c_r)/(1 + c_f + c_r) \quad [4]$$

where x defines the isotope effect (D, T, ^{13}C , ^{15}N , ^{18}O) and the leading superscript indicates the ratio of the parameters for light and heavy isotopes. xk is the intrinsic isotope effect on the chemical step and $^xK_{eq}$ the equilibrium isotope effect on the reaction. The constants c_f and c_r are forward and reverse commitments to catalysis, respectively, and are the ratio of the rate constant for the isotope-sensitive step to the net rate constant for release from the enzyme of either the substrate (for c_f) or the first product (for c_r). For the simple mechanism



$$c_f = (k_5/k_4)(1 + k_3/k_2) \quad c_r = (k_6/k_7)(1 + k_8/k_9) \quad [6]$$

The commitments, thus, consist of partition ratios of intermediates and these ratios correspond to the differences in barrier heights in a free-energy profile. In the profile shown in Figure 4, c_f and c_r will be small if the isotope effect measured involves reaction of EI to EP* because the barrier between EI

and EP* is the highest one. But if the isotope effect measured is on reaction of EA* to EI, c_f will be small, but c_r will be large. Thus, one will see largely the equilibrium isotope effect, as the EA* to EI step approaches equilibrium.

In the definition of c_f in eqn [5], k_5/k_4 is the internal part of the commitment (c_{f-in}) and $k_3k_5/(k_2k_4)$ is the external part (c_{f-ex}). A sticky substrate is one where the net rate constant for reaction of the initial collision complex to give products is faster than the rate constant for dissociation (k_2). This ratio is called the stickiness ratio (S_r) and is given by $S_r = c_{f-ex}/(1 + c_{f-in} + c_r)$ in terms of the parameters of eqn [5].

Rate-Limiting Steps for V

The rate-limiting step for the maximum velocity depends on the definition of rate limiting. One definition is the 'slowest' step, or the one with the highest individual barrier in the forward direction. Another definition is the 'least-conductive' step, which is the one that sees the highest total barrier to reach an irreversible step. A third definition is the most-sensitive step. This is the one that causes the greatest percentage change in the turnover number for a given percent change in the forward rate constant. An isotope effect on this step is more fully expressed than one on any other step. The isotope effect on V for mechanism 3 is

$$^xV = (^xk + c_{vf} + ^xK_{eq}c_r)/(1 + c_{vf} + c_r) \quad [7]$$

where xk , $^xK_{eq}$, and c_r have the same meaning as in eqn [5], but

$$c_{vf} = [k_3k_5/(k_3 + k_4)][1/k_3 + (1/k_7)(1 + k_8/k_9) + 1/k_9 + 1/k_{11}] \quad [8]$$

Note that c_{vf} , unlike c_f , does not consist of partition ratios. Rather k_5 , reduced by the factor $k_3/(k_3 + k_4)$, is compared to k_3 , and to each of the net rate constants after the chemical step. A low value of k_{11} , for example, can result in a large value of c_{vf} , and a greatly reduced expression of the isotope effect on V .

The slowest, least-conducting, and most-sensitive steps may be the same in an enzymatic reaction, but need not be, so one must define what one means when using the term 'rate limiting' for the maximum velocity.

See also: [Bioenergetics: Metabolite Channeling: Creatine Kinase Microcompartments](#); [Protein/Enzyme Structure Function and Degradation: Enzyme Inhibitors; Enzyme Kinetics; Enzyme Reaction Mechanisms: Stereochemistry; Kinetic Isotope Effects; Low-Barrier Hydrogen Bonds](#).

Further Reading

- Cleland WW (1982) An analysis of haldane relationships. *Methods in Enzymology* 87: 366–369.
- Cleland WW (1986) Enzyme kinetics as a tool for determination of enzyme mechanisms. In: Bernasconi CF (ed.) *Investigation of Rates and Mechanisms of Reactions*, 4th edn., part 1, pp. 791–870. New York, NY: Wiley.
- Cleland WW and Northrop DB (1999) Energetics of substrate binding, catalysis, and product release. *Methods in Enzymology* 308: 3–27.

Sugar Nucleotide Transporters

C B Hirschberg, Boston University School of Dental Medicine, Boston, MA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Glycosylation The covalent addition of sugars to each other or to proteins and lipids.

Golgi apparatus A membrane-enveloped organelle in the cytoplasm of all eukaryotes composed of several cisternae and vesicles; virtually all membrane-bound and secreted

proteins of eukaryotes traverse through this organelle en route to their final location within or outside the cell.

Nucleotide sugar Phosphodiester, usually between nucleoside 5'-diphosphates and carbon 1 of the sugar.

Phenotype The visual or physical characteristics of individual cells or a multicellular organism.

A Requirement for Sugar Nucleotide Transport into the Golgi Apparatus Lumen

Approximately half of all proteins in eukaryotic cells are secreted or membrane bound. These proteins are processed through the secretory pathway where 80% undergo posttranslational modifications such as glycosylation, sulfation, and phosphorylation in the lumen of the Golgi apparatus. Intact nucleotide sugars, nucleotide-sulfate, and adenosine 5' triphosphate (ATP), the substrates for these reactions, enter the lumen of the Golgi apparatus via specific transporter proteins (Figure 1). These are multi-transmembrane spanning proteins, which function as antiporters with the corresponding nucleoside monophosphates and thereby concentrate the nucleotide derivatives in the Golgi lumen relative to their concentration in the cytosol (Figure 1). Evidence for this mechanism has been obtained through biochemical studies with rat liver-derived Golgi vesicles as well as with different species of yeast. In the latter, gene disruption of a Golgi luminal guanosine 5' diphosphate (GDP)ase, which generates guanosine 5' monophosphate (GMP) (the antiporter molecule coupled to entry of GDP-mannose into the yeast Golgi lumen), results in severe undermannosylation of proteins and lipids *in vivo*, reduced transport of the GDP-mannose into Golgi vesicles *in vitro*, as well as in a failure to make hyphae in *Candida albicans*.

Mutants in Nucleotide Sugar Transport – Biochemical and Developmental Phenotypes Including Diseases

Mutant mammalian cells grown in tissue culture have been described as having 95% reduced transport of specific nucleotide sugars into their Golgi apparatus, while their proteins and lipids have a drastic reduction in the specific sugar transported by the corresponding nucleotide derivative transporters. Multicellular organisms such as *Caenorhabditis elegans*, *Drosophila melanogaster*, plants, and humans that have mutations in the above proteins also have distinct morphological phenotypes affecting development of limbs, wings, brain, etc. Studies of these mutants have shown that some of the transporters can translocate more than one nucleotide sugar, whereas other transporters retain high substrate specificity such as differentiating between sugar epimers.

Mutants in nucleotide sugar transport have also been described in *Leishmania donovani* and in yeasts such as *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, and *C. albicans*.

Structure of Golgi Nucleotide Sugar Transporters

So far the primary amino acid sequence, together with the substrate specificity of approximately two dozen Golgi membrane nucleotide sugar transporters, has been determined. In most cases, this was done by correcting the phenotypes of mutant cells or mutant multicellular organisms with libraries containing wild-type DNA or cDNA from homologous or heterologous organisms. In most cases, after phenotypic correction, the specific DNA (or cDNA) was expressed in yeast or mammalian cells, Golgi vesicles were isolated, and transport of different nucleotide sugars assayed *in vitro*. Yeast is a particularly attractive system for heterologous expression of nucleotide sugar transporter genes, because it has few endogenous nucleotide sugar transport activities. This approach led to the identification of multisubstrate nucleotide sugar transporters from *C. elegans*. In some instances, a particular cDNA was also tested for its ability to correct the phenotype of mutant cells where the substrate specificity of the defective nucleotide sugar is known.

The fact that the above approaches for determining the substrate specificity of particular transporters have worked is based on the following points:

1. the transporters' DNA (or cDNA) can express the proteins in a heterologous system;
2. the targeting signals allowing the protein to localize to the Golgi apparatus are functional in different cell types; and
3. the targeted transporter is active once it reaches the Golgi apparatus of the recipient cell and can therefore affect phenotypic correction.

To date, the relationship of amino acid sequences and substrate specificities of nucleotide sugar transporters can be summarized as follows:

1. substrate specificities of nucleotides sugar transporters cannot be inferred even when amino acid sequences are ~50% identical;
2. only sequence identities of over 80% allow accurate predictions of substrate specificity; and

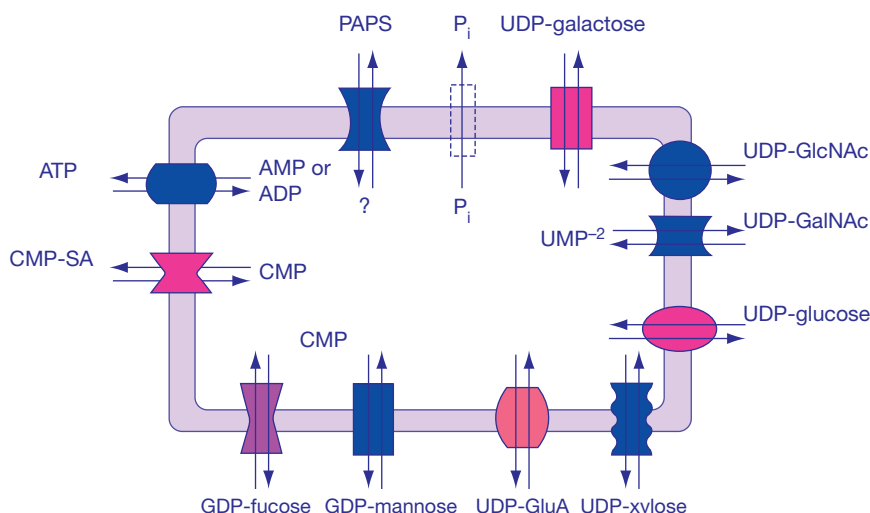


Figure 1 Golgi membrane nucleotide sugar, adenosine 3' phosphate 5' phosphosulfate (PAPS), and ATP transporters have been detected in mammals, yeast, protozoa, nematodes, insects, and plants. They are antiporters with the corresponding nucleoside monophosphate, except for PAPS, where the antiporter is not known, and for ATP, where it is either adenosine 5' monophosphate (AMP) or adenosine 5' diphosphate (ADP) or both.

- transporters from different organisms that have the same substrate specificity may have only 20% overall amino acid sequence identity.

Due to the above reasons the reader is cautioned that annotations about the identity (substrate specificity) of nucleotide sugar transporters in databases should be looked at very carefully!

Topography of Nucleotide Sugar Transporters in the Golgi Membrane

So far the only detailed topographic study of a nucleotide sugar transporter has been done with that for cytidine 5' monophosphate (CMP)-sialic acid in mammalian cells. Using a combination of tagged epitopes and immunocytochemistry, the transporter was found to have 10 transmembrane domains with both its amino and carboxy termini facing the cytosol. A study on the topography of the *K. lactis* uridine 5' diphosphate (UDP)-acetylglucosamine transporter using membrane accessibility of a peptide antibody and amino- and carboxy termini epitope tags also found that both termini face the cytosol and that this transporter has either six or eight transmembrane domains. It is important to stress that none of the existing algorithms in the literature are designed specifically to address the issue of protein topography in the Golgi apparatus membrane, which is somewhat thinner than the plasma membrane; most algorithms are designed for predicting topography in the plasma membrane.

Regulation of Macromolecular Glycosylation by Nucleotide Sugar Transporters

Studies with mutant mammalian cells in nucleotide sugar transporters grown in tissue culture as well as with fibroblasts

and lymphoblasts from patients with leukocyte adhesion deficiency (LAD) syndrome II showed that decreased concentrations of nucleotide sugar in the Golgi lumen result in selective hypoglycosylation of proteins rather than across the board reduction. Thus, mutant Madin-Darby canine kidney (MDCK) cells which have a 95% reduction in transport of UDP-galactose into their Golgi lumen, have a corresponding decrease in galactosylation of bulk glycoproteins, glycolipids, and keratan sulfate but not in heparan and chondroitin sulfates. LAD II patients appear to have normal O-fucosylation of Notch protein, whereas N-fucosylation of bulk proteins is drastically reduced. Although the detailed mechanism for these biochemical phenotypes is not understood, an attractive hypothesis is that when supply of nucleotide sugars is limited due to impairment in their transport, glycosylation reactions with lower K_m 's would proceed, whereas those with higher K_m 's would not.

See also: **Cell Architecture and Function: Golgi Complex;**
Lipids Carbohydrates Membranes and Membrane Proteins:
Glycosylation, Congenital Disorders of.

Further Reading

- Caffaro CE and Hirschberg CB (2006) Nucleotide sugar transporters of the Golgi apparatus: From basic science to diseases. *Accounts of Chemical Research* 39: 805–812.
- Hirschberg CB (2001) Golgi apparatus nucleotide sugar transport and leukocyte adhesion deficiency II. *Journal of Clinical Investigation* 108: 3–6.
- Hirschberg CB, Robbins PW, and Abeijon C (1998) Transporters of nucleotide sugars, ATP and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus. *Annual Review of Biochemistry* 67: 49–69.
- Varki A, Cummings R, Esko J, et al. (2009) *Essentials of Glycobiology*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Superoxide Dismutase

I Fridovich, Duke University Medical Center, Durham, NC, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Apo enzyme Enzyme whose prosthetic group has been removed.

Cu, Zn SOD Superoxide dismutases with copper and zinc at the active site.

Fe SOD Superoxide dismutases with iron at the active site.

HO· The hydroxyl radical.

Homodimer Protein composed of two identical subunits.

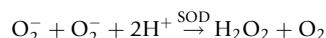
Homotetramer Protein composed of four identical subunits.

Mn SOD Superoxide dismutases with manganese at an active site.

O₂⁻ The superoxide anion radical.

Introduction

Molecular oxygen (O₂), while essential for the life of aerobes, is potentially toxic. This toxicity is due to the propensity of O₂ for reduction by a univalent pathway. This facile univalent pathway of O₂ reduction generates intermediates that lie between one O₂ and its four electron reduction products – two molecules of water – and it is the reactivity of these intermediates that is responsible for the toxicity of O₂. In order of their production, these intermediates are the superoxide anion radical (O₂⁻), hydrogen peroxide (H₂O₂), and the hydroxyl radical (HO·). Superoxide dismutases (SODs) catalyze the conversion of O₂⁻ into H₂O₂ plus O₂, that is,



and, in so doing, provide an important defense. H₂O₂, in turn, is eliminated by the catalases and peroxidases. The concerted action of the SODs with the catalases and peroxidases prevents the formation of the very reactive HO·. This is the case because HO· production *in vivo* requires both O₂⁻ and H₂O₂. O₂⁻ oxidizes the [4Fe-4S] clusters of dehydratases, such as the aconitases, causing release of Fe (II); the freed Fe (II) then reduces H₂O₂ to OH⁻ plus HO· in a reaction known as ‘the Fenton reaction’.

The SOD Family of Enzymes

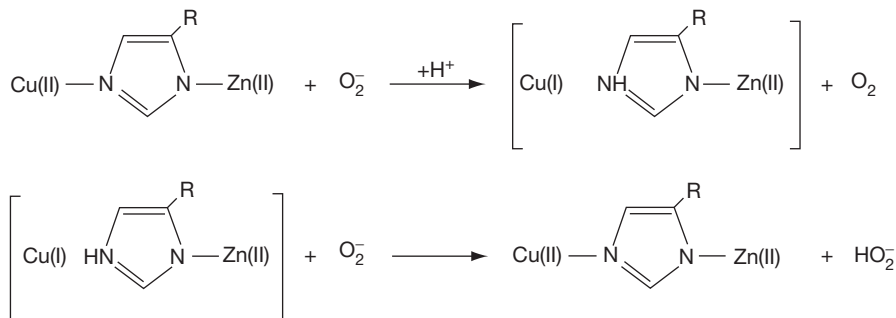
The first truly photosynthetic organisms were the cyanobacteria, and the oxygen they produced oxygenated a previously anaerobic biosphere. This imposed a common selection

pressure on what must have been a varied anaerobic biota. It is not surprising that several different SODs were then evolved to deal with the toxicity of the accumulating O₂ and that these persist to this day. Thus, there are SODs based on Cu (II) plus Zn (II), Mn (III), Fe (III), and Ni (II), all of which catalyze the dismutation of O₂⁻ into H₂O₂ + O₂.

All SODs work by a similar mechanism in which the metal at the active site is reduced by one O₂⁻ and then reoxidized by the next O₂⁻. The active site metal, thus, acts as a mediator passing an electron from one O₂⁻ to the next. In this way, the electrostatic repulsion, which would prevent close approach of one O₂⁻ to another, is bypassed by the SODs. All the SODs are very efficient catalysts and operate at close to the theoretical diffusion limit. Their rate constants for interaction with O₂⁻ are ~3 × 10⁹ M⁻¹ s⁻¹.

The Cu, Zn SODs

These SODs have been found in the cytosols of eukaryotic cells, the intermembrane space of mitochondria, the periplasm of Gram-negative bacteria, and in chloroplasts. The Cu (II) is linked to the Zn (II) at the active site by the imidazolate moiety of a histidine residue and this bridging imidazolate functions as a proton supply during the catalytic cycle. Thus, when the Cu (II) is reduced by O₂⁻, the Cu (I) releases the imidazolate that then protonates to imidazole. When the Cu (I) is reoxidized to Cu (II), the imidazole provides a proton to the nascent O₂⁻ converting it to HO₂⁻, as the bridging to the Zn (II) is reestablished. The HO₂⁻ that leaves the active site protonates in the water to H₂O₂. This mechanism can be depicted as follows:



Both of these half-reactions of the catalytic cycle proceed with a rate constant of $\sim 3 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$. The enzyme surface contains charged groups that provide an electrostatic funnel guiding the O_2^- toward the active site crevice. This electrostatic facilitation contributes to the efficiency of these enzymes.

The Cu, Zn SODs found in eukaryotic cytosols are homodimeric proteins with subunit weights of $\sim 16 \text{ kDa}$. The corresponding enzyme from the periplasm of *Escherichia coli* is monomeric. There is an extracellular Cu, Zn SOD found in higher animals. It is usually homotetrameric and glycosylated and has a subunit weight of $\sim 23 \text{ kDa}$.

The Mn SODs

These enzymes, whether from the matrix of mitochondria or bacteria, exhibit marked sequence similarity, reflecting a close evolutionary history and supporting an endosymbiotic origin for mitochondria. The bacterial enzyme is usually a homodimer, whereas the mitochondrial enzyme is a homotetramer. The subunit weight is $\sim 23 \text{ kDa}$. Some bacterial Mn SODs are tetrameric. This is the case in *Cryptococcus neoformans*. In *E. coli*, the biosynthesis of Mn SOD is under the control of the soxRS regulon, which coordinately upregulates the expression of a number of genes in response to O_2^- . The constitutively expressed SOX R protein is transcriptionally inactive in its reduced form. It can be oxidized by O_2^- and then activates the expression of the SOX S protein, which in turn activates all the genes in the regulon. Thus, Mn SOD is not measurable in extracts of anaerobically grown *E. coli*, but exposure of cultures to aerobic conditions elicits production of Mn SOD. Increasing production of O_2^- , by raising $p\text{O}_2$, or by adding compounds such as viologens, which can mediate enhanced production of O_2^- , increases the level of Mn SOD. It has been possible to force *E. coli* to produce Mn SOD to 7% of its soluble protein by aerobic exposure to the viologen paraquat.

The nectar of tobacco flowers has been found to contain a stable Mn-protein named nectarin that appears to be a Mn SOD.

The Fe SODs

These SODs, which are highly homologous to the Mn SODs, are found in bacteria and in plants. Although usually homodimeric, homotetrameric Fe SOD has been found in *Rhodococcus bronchialis* and *Mycobacterium tuberculosis*. The Fe SOD of *E. coli* is constitutive and thus is found even in anaerobically grown cells. It can thus be viewed as a standby defense against O_2^- , which is always maintained to protect in the event of a sudden exposure to O_2 .

It is possible to reversibly remove the metals from SODs. The apo enzymes so prepared are inactive but can be reactivated by restoring the active site metal. It is striking that, in spite of sequence and structural homologies, the Fe SOD is inactive when made to contain Mn in place of Fe and the Mn SOD is inactive when similarly reconstituted with the non-native metal Fe. There are, however, SODs that are active with either Mn or Fe at their active sites. These SODs, termed

'cambialistic SODs', are usually found in anaerobes such as *Propionibacterium shermanii* or *Streptococcus mutans*. These organisms produce the SOD with iron when grown anaerobically; however, when exposed to low levels of O_2 , produce the same SOD but with Mn. It is clear that even anaerobes have need of a defense against the O_2^- during transient exposure to O_2 . Ferrous salts are soluble and thus available anaerobically, but become insoluble when oxidized to the ferric state by O_2 . Manganous salts remain soluble under both conditions – hence, the wisdom of a cambialistic SOD for an anaerobe. An iron SOD has also been reported in the protozoan *Tetrahymena pyriformis*.

Deletions and Consequences

Escherichia coli

Mutants of *E. coli* lacking both the Mn SOD and the Fe SOD (SodA, SodB) have been prepared. Anaerobically, they grow as well as the parental strain. However, aerobically they grow slowly, exhibit a high rate of mutagenesis, and require branched-chain, sulfur-containing, and aromatic amino acids. The null mutants are also hypersensitive toward paraquat. All of these phenotypic deficits disappear when a functional gene encoding any active SOD is inserted. The slow growth of the null mutant can be understood in terms of the energy devoted to repairing or replacing oxidatively damaged molecules and to the inactivation of such O_2^- -sensitive enzymes as aconitase. The high rate of O_2 -dependent mutagenesis reflects oxidative DNA damage, and repair, which is error prone. The nutritional auxotrophies exhibited by the null mutant have several explanations. The need for branched amino acids arises because the penultimate step in the biosynthesis of these amino acids is catalyzed by a dihydroxy acid dehydratase that contains a [4Fe-4S] cluster that is rapidly oxidized, and inactivated, by O_2^- . The need for sulfur-containing amino acids evidently reflects leakiness of the cell envelope toward sulfite, which is an intermediate on the pathway from sulfate to cysteine. Finally, the aromatic amino acid auxotrophy reflects the inability to make erythrose-4-phosphate, which is one of the starting materials on the pathway leading to these amino acids. Erythrose-4-phosphate, in turn, depends on the sequential actions of transketolase and transaldolase, and the transketolase intermediate 1,2-dihydroxyethylthiamine pyrophosphate is rapidly oxidized by O_2^- . The varied explanations for these phenotypic deficits reflect the variety of damage that can be caused by O_2^- and the multiple targets that are protected by the SODs.

Yeast

Mutants of the yeast *Saccharomyces cerevisia* lacking either the cytosolic Cu, Zn SOD, or the mitochondrial Mn SOD have been prepared. Their phenotypic deficits, which are all oxygen dependent, include slow growth, hypersensitivity toward paraquat or quinones, vacuolar fragmentation, sensitivity toward oxygen, and, in the case of the Cu, Zn SOD-null, lysine auxotrophy. All of these problems can be explained on the basis of O_2^- damage, as was the case for the *E. coli* SOD-null mutants.

Mice

Murine mutants lacking Mn SOD are severely affected and the homozygotes that survive to term are low birth weight and only survive for 4–14 days. Additional phenotypic deficits that have been attributed to subnormal levels of the MnSOD, such as that seen in heterozygotes, include anemia, neuronal degeneration, cardiomyopathy, increased oxidative damage to mitochondrial DNA, diminished levels of glutathione and aconitase, and increased susceptibility to reperfusion injury. Their problems can be seen as a deficit of mitochondrial functions and that in turn can be explained on the basis of damage to mitochondrial components by O_2^- .

In contrast to the embryonic and neonatal fatality imposed by Mn SOD deletion, the homozygous Cu, Zn SOD-null mice were apparently normal, while young, under laboratory conditions. They are, however, reported to be relatively intolerant of neuronal injury and prone to develop hepatoma at >9 months of age. Other debilities seen as these mice age include shortened life span, hearing loss, motor neuron axonopathy, and retinal degeneration. It seems likely that there is some redundancy in scavenging O_2^- in the cytosol, so that the lack of Cu, Zn SOD may be partially compensated by upregulation of another O_2^- scavenger, perhaps a superoxide reductase, similar to that found in anaerobes. The sum of the problems exposed by the loss of SOD activities, and that can be corrected by restoration of SODs, reflects the multiple damages that can be inflicted by O_2^- .

SOD and Amyotrophic Lateral Sclerosis

Amyotrophic lateral sclerosis (ALS) or Lou Gherig's disease is a late-onset, progressive, and fatal paralytic disease due to death of motor neurons. There are sporadic and familial types of ALS and these are not distinguishable on the basis of clinical symptoms. Approximately 20% of the familial cases of ALS have been associated with point mutations causing single amino acid replacements in Cu, Zn SOD. Transgenic mice expressing the wild-type human Cu, Zn SOD are normal, but those expressing one of the ALS-associated mutant Cu, Zn SODs develop progressive paralysis at ~3 months of age. Thus, the mutations in the abundant SOD cause the death of motor neurons because of some gain of function, rather than because of a loss of SOD activity. This is certainly the case because most of the familial ALS-associated mutant SODs retain full SOD

activity; the transgenic mice retain the active murine Cu, Zn SOD; the disease is genetically dominant, that is, heterozygotes develop the disease; and, finally, Cu, Zn SOD knockout mice do not develop ALS. It appears that the mutant SODs all exhibit decreased stability, particularly while in their metal-free apo forms, and are prone to aggregation. The accumulation of protein aggregates may be particularly damaging to motor neurons because such aggregates interfere with transport within their very long axons. These transgenic mice provide a useful model of the human disease and are being used to explore both the toxic gain of function of the mutant Cu, Zn SOD and treatment modalities.

Mimics

O_2^- has been found to play causative roles in numerous inflammatory diseases and reperfusion injuries. For this reason, non-enzymic catalysts of the dismutation of O_2^- are being explored as pharmaceutical agents and appear to have promise.

See also: **Bioenergetics:** Iron–Sulfur Proteins; Mitochondria: Source and Target of Free Radicals; **Molecular Biology:** DNA Oxidation.

Further Reading

- Cozzolino M, Pesares MG, Amori I, et al. (2009) Oligomerization of mutant SOD1 in mitochondria of motoneuronal cells drives mitochondrial damage and cell toxicity. *Antioxidants and Redox Signaling* 11: 1547–1558.
- Furukawa Y, Ronngen R, Denq H-X, and Siddique T (2006) Disulfide cross-linked protein represents a significant fraction of ALS-associated Cu, ZnSOD aggregates in spinal cords of model mice. *Proceedings of the National Academy of Sciences of the United States of America* 103: 7148–7153.
- Hassan HM (1997) Cytotoxicity of oxyradicals and the evolution of superoxide dismutases. In: Clerch LB and Massaro DJ (eds.) *Oxygen, Gene Expression, and Cellular Function*, Lung Biology in Health and Disease, vol. 105, pp. 27–47. New York: Marcel Dekker.
- Maier CM and Chan RH (2002) Role of superoxide dismutases in oxidative damage and neurodegenerative disorders. *Neuroscientist* 8: 323–334.
- McCord JM (2002) Superoxide dismutases in aging and disease. *Methods in Enzymology* 349: 331–341.
- Melov S (2002) Therapeutics against mitochondrial oxidative stress in animal models of aging. *Annals of the New York Academy of Sciences* 959: 330–340.
- Storz G and Imlay JA (1999) Oxidative stress. *Current Opinion in Microbiology* 2: 188–194.

T7 RNA Polymerase

R Sousa, University of Texas Health Science Center, San Antonio, TX, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 4, pp. 147–151, © 2004, Elsevier Inc.

Glossary

Downstream The direction in which an RNA polymerase moves along the DNA during transcription.

Primer A DNA or RNA molecule, typically short, that is extended by a DNA polymerase during DNA replication.

Promoter A DNA from which an RNA polymerase initiates transcription.

Template strand When a nucleic acid directs the synthesis of DNA or RNA, the template strand selects – by Watson–Crick base pairing – the nucleotides incorporated into the newly synthesized molecule.

Transcription The synthesis of RNA using a DNA template.

DNA-Directed RNA Polymerases

The enzymes which synthesize nucleic acids – the polymerases – are functionally defined on the basis of whether they use DNA or RNA as a template, and whether they synthesize RNA or DNA. The DNA-directed RNA polymerases comprise two large families. One includes the large multi-subunit cellular RNA polymerases which synthesize the messenger and ribosomal RNAs of all cells. The other family includes the mitochondrial and chloroplast RNA polymerases, and the RNA polymerases encoded by a variety of bacteriophage. These are simpler, typically single-subunit, enzymes. The best-characterized representative of this family is the RNA polymerase of the T7 bacteriophage.

T7 RNA Polymerase

Structure

Three-dimensional structures of T7 RNA polymerase in isolation and in complexes with nucleic acids and nucleotide triphosphate (NTP) substrates have been determined by X-ray crystallography. The polymerase is comprised of an N-terminal domain (amino acids 1–312), and a C-terminal domain (amino acids 313–883). The C-terminal domain is further subdivided into thumb (amino acids 330–410), palm (amino acids 411–448, 532–540, 788–838), and fingers (amino acids 541–737, 771–778) subdomains, which are so designated because together they form a structure similar in shape to a cupped right hand. Nucleic acids bind within the large cleft in this structure (Figure 1).

Structure–Function Relationships

The C-terminal domain contains the active site where RNA synthesis occurs. Phosphodiester bond formation is catalyzed by two Mg^{2+} ions complexed by two aspartates (D537, D812) of the palm subdomain (Figure 2). The thumb subdomain makes interactions with the RNA that stabilize the transcription complex, while the fingers subdomain binds the template strand and contains residues which bind the NTP and which discriminate ribo-NTPs from deoxyribo-NTPs. Residues 739–770 of the C-terminal domain form a promoter recognition loop which makes sequence-specific interactions with the –7 to –11 base pairs (bp) of the promoter. The C-terminal domain also contains the binding site for the regulatory factor T7 lysozyme, and residues 839–883 form a C-terminal extension which functions in the allosteric mechanism by which T7 lysozyme regulates transcriptional activity. The N-terminal domain is involved in nascent RNA binding, in sequence-specific binding to the promoter, and in separating (opening) the two strands of the promoter during initiation so as to expose one strand for templating RNA synthesis. Residues 93–101 of the N-terminal domain are rich in positively charged amino acids and make specific interactions with the AT-rich –13 to –17 bp of the T7 promoter, whereas residues 232–242 form a β -hairpin which inserts between the two DNA strands to open the promoter.

Similarities to Other Polymerases

T7 RNA polymerase shares extensive sequence similarity with mitochondrial and chloroplast RNA polymerases and with

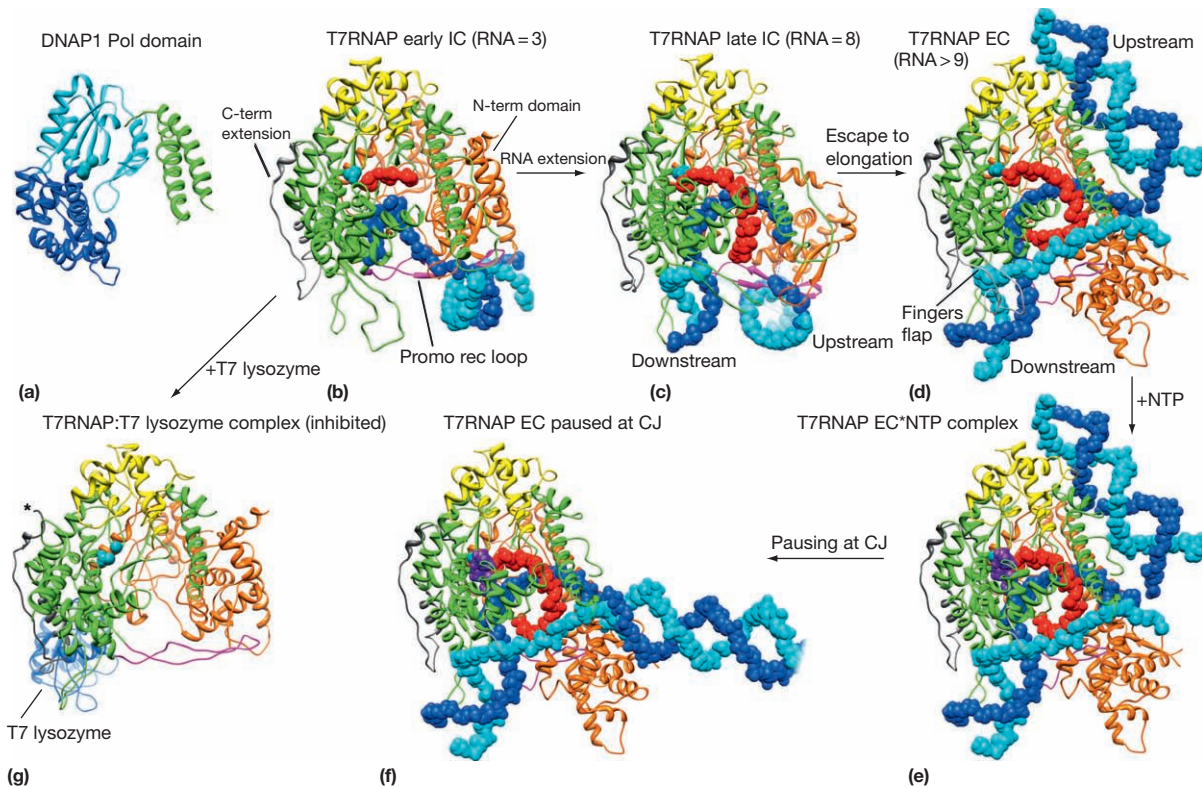


Figure 1 Structural organization of T7 RNA polymerase and its conformational changes. (a) Ribbons model of the polymerase domain of DNA pol 1 (pdb 1KLN) with its thumb subdomain (amino acids 546–656) in green, its palm subdomain (amino acids 657–707, 854–928) in cyan, and its fingers subdomain (amino acids 708–853) in blue. The active site center is indicated by a space-filling representation of the aspartate 882 side chain (in dark cyan). (b) Ribbons model of a T7 RNA polymerase initial transcription complex with a 3-nt RNA (pdb 1QLN). The portion of T7 RNA polymerase homologous to the polymerase domain of DNA pol I (amino acids 330–448, 532–737, 771–838) is in green. This T7 RNA polymerase domain is accessorized with an N-terminal domain (amino acids 1–329; orange), a domain of unknown function (amino acids 449–531; yellow), the promoter recognition loop (amino acids 738–770; magenta), and a C-terminal extension (amino acids 839–883; gray). The center of the active site is indicated by a space-filling representation of side chain of aspartate 812 in dark cyan. The template and nontemplate strands of the promoter DNA are blue and cyan, respectively, and the RNA is red. (c) Ribbons model of a T7 RNA polymerase initial transcription complex with an 8-nt RNA (pdb 3E3J). Relative to the structure in (b), the N-terminal domain and promoter recognition loop have rotated away from the rest of the polymerase to make room for the growing RNA:DNA hybrid while maintaining their interaction with the upstream, base-paired part of the promoter. (d) Ribbons model of a T7 RNA polymerase elongation complex with a 10-nt RNA (pdb 1MSW). Relative to the structure in (c), the N-terminal domain has taken on a completely new fold and the upstream DNA has moved to a new location on the polymerase surface. (e) Model of a T7 RNA elongation complex with NTP (purple and in space-filling representation) bound (pdb 1S76). Relative to the structure in (d), the fingers subdomain (green) has rotated toward the NTP and the fingers flap (amino acids 590–612; colored light gray in (d)) has moved away from the downstream DNA and assumed a disordered conformation. (f) Model of the elongation complex paused at the T7 concatemer junction (CJ). Relative to the structure in (e), the upstream DNA has swung from one site on the polymerase to establish an interaction with a different site. (g) Model of T7RNA polymerase bound to the regulatory inhibitor, T7 lysozyme (in blue). Relative to the other T7 structures shown, the T7 lysozyme-bound polymerase has its C-terminus (indicated by an asterisk) pulled out of the active site and exposed to solvent.

other bacteriophage-encoded RNA polymerases. T7 RNA polymerase is also structurally similar to the DNA-directed DNA polymerases of the pol I/pol α family, to the RNA-directed DNA polymerases (reverse transcriptases), and to the RNA-directed RNA polymerases. However, unlike the extensive sequence similarities between the mitochondrial, chloroplast, and phage RNA polymerases, the identifiable sequence similarities between T7 RNA polymerase and the DNA polymerases, reverse transcriptases, and RNA-directed RNA polymerases are limited to a few well-conserved amino acids in a small number of sequence motifs. The structural similarity suggested by this limited sequence similarity is confirmed by comparison of the three-dimensional structures of these enzymes (Figure 1).

This comparison also reveals that functions specific to particular classes of polymerase are incorporated by accretion of structurally dissimilar domains or modules onto a structurally conserved core. For example, DNA polymerase I and T7 RNA polymerase both exhibit structurally similar thumb, palm, and fingers subdomains which together form a polymerase domain with the core function of processive template-directed nucleic acid synthesis. In T7 RNA polymerase, this conserved core is accessorized with modules that include the N-terminal domain, which is involved in the RNA polymerase-specific functions of promoter recognition and opening, RNA binding and displacement, and transcription bubble resolution and transcription termination. They also include the promoter

recognition loop, which is responsible for sequence-specific binding of the T7 promoter, and the C-terminal extension, which is involved in regulation by T7 lysozyme. None of these modules have structural or functional counterparts in DNAP I which, instead, is accessorized with unrelated domains that carry out the exonucleolytic and proofreading activities required for an enzyme involved in DNA repair and highly accurate genome replication (Figure 3).

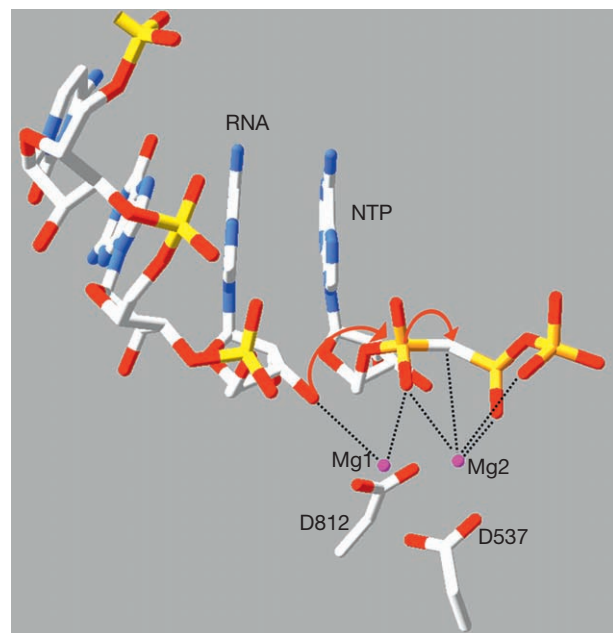


Figure 2 Bond formation in nucleic acid synthesis. The 3'-OH of the nucleotide at the end of the DNA primer or RNA attacks the α -phosphate of the NTP. A new bond is formed, and the β - and γ -phosphates of the NTP are lost as pyrophosphate. The reaction is catalyzed by two metal ions which stabilize both the negative charge on the 3'-OH to enhance its nucleophilicity and the negatively charged pentacovalent α -phosphate intermediate which forms during the transition state. The two metal ions are bound by the side chains of two aspartic acids (D537 and D812 in T7 RNA polymerase) which are found in the active sites of all homologous polymerases.

T7 RNA Polymerase: Transcription Reaction and Conformational Change

Promoter Structure, Recognition, and Opening

The T7 promoter is 23 bp in length and has a tri-partite structure (Figure 4). The -17 to -6 sequence is important for specific binding of the polymerase via interactions between residues 746, 748, 756, and 758 and the -7 to -11 bp, and between residues 93–101 and the -13 to -17 bp (bases are numbered relative to the transcription start site at $+1$). The -17 to -6 bp remain duplex during transcription initiation. The -4 to -1 TATA sequence facilitates promoter opening, which begins at the -4 bp and extends downstream, driven by imposition of a sharp ($\sim 50^\circ$) bend in the promoter when polymerase binds, and by insertion of β -hairpin (residues 231–242) between the DNA strands. The $+1$ to $+6$ initially transcribed sequence enhances the efficiency of the initial transcription reaction. Seven class III T7 promoters occur in the T7 genome and exhibit a perfect match to the sequence shown in Figure 4. There are also 16 less active class II T7 promoters in the T7 genome which typically exhibit a small number of base-pair differences from the class III consensus.

Initial Transcription

Once bound, the polymerase rapidly opens the promoter but the promoter closes again even more quickly unless initiating NTPs bind to stabilize the open complex. Binding of initiating NTPs is followed by their catalytic joining to form a dinucleotide RNA which can then be further extended. When this nascent RNA is small (<12 nucleotides (nt)), the transcription complex is unstable and short RNAs are frequently released from the complex. Once an RNA is released, the polymerase reinitiates, usually without releasing the promoter. For this reason this initial phase of transcription, which is also characteristic of multi-subunit cellular RNAPs, is also referred to as abortive transcription. Throughout this initial phase of transcription, the polymerase retains the specific promoter interactions made with the -17 to -6 bp in the initial binding step. Transcription to $+8$ is achieved by threading the template strand through the active site and by compacting (scrunching) it within the template-binding cleft, as well as by conformational changes that rotate the N-terminal domain away from the C-terminal domain so as

Motif designation:	T/DxxGR	A	B	C	
DNA-directed polymerases					
DNA polymerases (Pol 1-like, pol α -like)	hT-GR	Dh-hEh	Khh---hYG	h-D	
RNA polymerases (phage, mitochondrial, plastid)	hDhRGRhY	Ph-D-C-GhQHh	R-h-K+-VMTh-YG	hHDSFGT	
RNA-directed polymerases					
DNA polymerases		hDh---h-h	h-h--hQG-SP	YHDDhh	Gh-h. ---K h-hLGH
RNA polymerases		Dh---hD	SG---h	hh-GDD-hh	G-h---K
Motif designation:		A	B'	C	D E

Figure 3 Patterns of sequence motif conservation in nucleic acid polymerases. The letter 'h' indicates a hydrophobic residue, '+' is a positively charged residue, '-' is any residue, and '●' is a sequence gap. Invariant amino acids are in bold face. In T7 RNA polymerase the two invariant aspartic acids of motifs A and C correspond to D537 and D812, respectively, whereas the invariant arginine of motif T/DxxGR and the invariant lysine and tyrosine of motif B correspond to, respectively, R425, K631, and Y639.

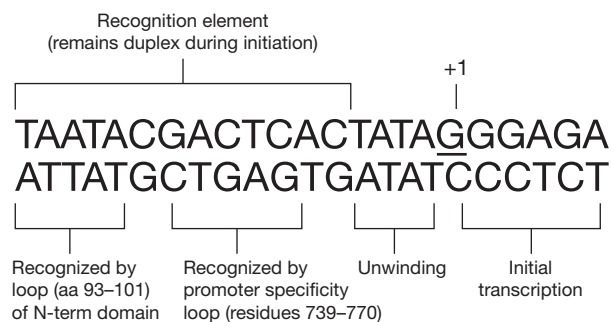


Figure 4 Structure of the T7 RNA polymerase class III promoter with the functions of different parts of the promoter indicated.

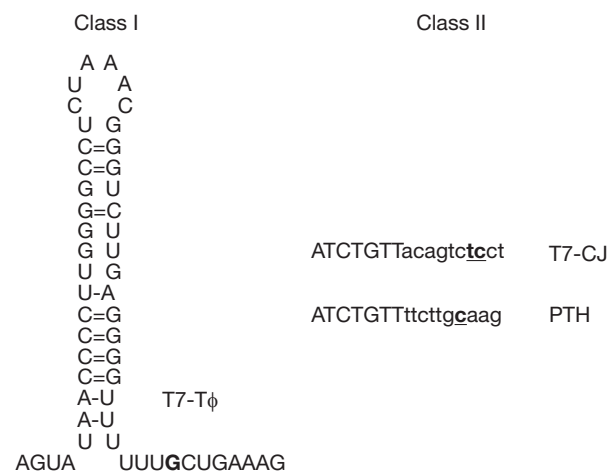


Figure 5 (Left) RNA structure of the T7 RNA polymerase T_ϕ class I terminator. Termination occurs at the underlined nucleotide. (Right) Nontemplate strand DNA sequence of the class II T7 RNA polymerase concatemer junction (CJ) pause/terminator site.

to expand the size of the template-binding cleft to accommodate the growing RNA:DNA hybrid (**Figure 1(c)**).

Promoter Release and Elongation

When the RNA reaches 9 nt in length, the polymerase can release the promoter in a step that involves a large conformational change in which the promoter recognition loop and N-terminal domain move apart to break up the promoter-binding surface created jointly by these two elements (**Figure 1(d)**). This conformational change also refolds the N-terminal domain, leading to the formation of a tunnel through which the emerging RNA passes. The result of all these changes is a stable, processive elongation complex with a ~ 10 base transcription bubble and an ~ 8 -bp RNA–DNA hybrid, which can synthesize transcripts thousands of nucleotide in length at rates of $200\text{--}300\text{ nt s}^{-1}$. During elongation, movement of the complex occurs by a thermal (Brownian) ratchet mechanism, in which the elongation complex slides rapidly and reversibly around the 3'-end of the RNA until the binding of a complementary NTP in the active site induces closing of the fingers subdomain (**Figure 1(e)**) and locks the complex down long enough for an NMP to be added to the growing

chain. Upon release of the pyrophosphate, the fingers subdomain opens and the cycle of reversible sliding of the elongation complex around the 3'-end of the RNA begins again.

Pausing and Termination

Certain DNA sequences can act as pause or terminator sites and interrupt the progress of the elongation complex. Terminators designated 'class I' contain a sequence which can form a G:C-rich hairpin when transcribed into RNA with a U-rich element immediately downstream (**Figure 5**). Formation of a hairpin in the RNA is expected to disrupt interactions made between the polymerase and single-stranded RNA 8–14 nt away from the RNA 3'-end and to disrupt part of the RNA:DNA hybrid. Disruption of these interactions weakens the association of the RNA with the elongation complex, and the U-rich nature of the remaining RNA:DNA base pairs further weakens the RNA's association with the complex, leading to release of the RNA and transcription termination. A class I terminator appears between genes 10 and 11 in the T7 genome, where it is important for attenuating transcription of downstream genes.

A sequence, AUCUGUU, found uniquely at the junction between the concatemerized T7 genomes generated during T7 DNA replication has also been found to cause the elongation complex to pause and to be an effective terminator when coupled with a downstream U-rich element. This sequence has therefore been designated the concatemer junction (CJ) element or class II terminator. Unlike class I terminators, the CJ element exhibits no obvious potential for forming secondary structure in the RNA, though it has been suggested that the CJ RNA can form a divalent metal-binding site that could partially disrupt CJ RNA base pairing with the template DNA. This could account for the observation that the upstream half of the transcription bubble re-anneals at the CJ element leading, in turn, to a dramatic change in the DNA architecture of the elongation complex, because this re-annealing disrupts a bend in the DNA located at the upstream edge of the transcription bubble (**Figure 1(f)**). This large change in the structure of the elongation complex may mediate not only its pausing, but also be the signal that recruits the T7 DNA packaging machinery to the CJ element, as pausing of the T7 RNA polymerase elongation complex at the CJ is required for the cutting of the genome concatemers into individual genome lengths and packaging into viral capsids. Sequences identical to the T7 CJ occur in related phage at similar positions in their genomes and also cause pausing by the polymerases encoded by those phage, implying that this is a conserved mechanism among phage that encode a T7 RNA polymerase homolog.

Regulation

During the T7 life cycle, the transcriptional activity of T7 RNA polymerase is regulated by T7 lysozyme, which binds to the polymerase and inhibits transcription initiation. Inhibition is allosteric: lysozyme binding changes the conformation of the C-terminal extension (amino acids 839–883) of the polymerase (**Figure 1(g)**). This conformational change weakens the affinity of the polymerase for NTPs and short RNAs. This, in turn, reduces the rate of transcription initiation at limiting NTP

concentrations, and decreases the efficiency of progression through initial transcription by increasing the frequency with which short RNAs are released from the complex during abortive transcription. These inhibitory effects are greater for class II promoters, which, relative to class III promoters, display intrinsically higher rates of RNA release during initial transcription and require higher NTP concentrations to achieve high rates of initiation. Class III promoters drive transcription of genes which encode proteins, such as phage coat proteins, that are required late in infection, whereas class II promoters drive transcription of genes which encode proteins required during the early and middle stages of phage infection, such as proteins involved in replication of the phage DNA. Thus, in a T7-infected *Escherichia coli*, the accumulation of T7 lysozyme late in the phage life cycle leads to a disproportionate decrease in transcription from class II promoters and to an increase in the production of the late phage proteins required for assembly of the mature phage particles. As T7 lysozyme is itself encoded by a gene transcribed from a class II promoter, an autoinhibitory feedback loop is created which ensures that repression by T7 lysozyme is kept within an appropriate range.

Priming DNA Replication

In addition to its primary function of transcribing the T7 phage genes, T7 RNA polymerase also primes rightward replication of T7 DNA. Priming occurs within an A-T-rich region immediately downstream of two T7 promoters (dubbed 1.1a and 1.1b) which are located at the T7 origin of replication. The mechanism by which the RNA primer initiated at these promoters is transferred from T7 RNA polymerase to T7 DNA polymerase is not understood.

Conformational Changes during the Transcription Reaction

The number of functionally distinct conformational isomers revealed for T7 RNA polymerase is probably greater than has been seen with any other well-characterized macromolecule. Relative to the crystal structure of the isolated polymerase, the early initiation complex (RNA < 5 nt in length; **Figure 1(b)**) exhibits significant ordering of loops and elements involved in nucleic acid interactions. The early initiation complex then changes significantly as the RNA is extended beyond 5 bases in length as the N-terminal domain rotates away from the C-terminal domain to accommodate the growing RNA:DNA hybrid and form the late initiation complex (**Figure 1(c)**). The late initiation complex is even more extensively remodeled by refolding of the N-terminal domain to form the elongation complex (**Figure 1(d)**). During elongation, every cycle of nucleotide addition is characterized by closing of the fingers subdomain and an order to disorder transition in the fingers flap (amino acids 590–612) coincident with NTP binding (**Figure 1(e)**). Regulatory inhibition by T7 lysozyme involves an allosteric change in which lysozyme binding causes the well-conserved, hydrophobic and functionally critical C-terminus of the polymerase to be extruded from the active site (**Figure 1(g)**). Finally, pausing at the CJ or class II terminator element involves partial transcription bubble collapse and a drastic change in elongation complex architecture (**Figure 1(h)**). This extensive conformational plasticity could be a consequence of selective

pressure on the T7 phage to maintain a small genome size to allow for rapid replication, so that maximum functionality must be packed into every protein. Whether this adaptive explanation is correct or not, these observations reveal the extraordinary degree of conformational variation that can be accessed by a macromolecule in executing its function.

T7 RNA Polymerase: Applications

The stringent promoter specificity and robust transcriptional activity of T7 RNA polymerase have been taken advantage of to overexpress proteins *in vivo* and to synthesize RNAs *in vitro*. In the most widely used embodiment of the former application, the gene encoding T7 RNA polymerase is placed under the control of an inducible promoter and is then stably integrated into the genome of an *E. coli* cell. A plasmid carrying the gene of interest under the control of a T7 promoter is then introduced into *E. coli*. When the gene encoding the T7 RNA polymerase is induced, the expressed T7 RNA polymerase transcribes the gene of interest at a high level, resulting in overproduction of the desired gene product. Similar approaches are used to overexpress proteins in eukaryotic cells. Synthesis of specific RNAs *in vitro* is done by using purified T7 RNA polymerase and templates in which a sequence of interest is placed downstream of a T7 promoter. Mutant T7 RNA polymerases and special reaction conditions have been developed that allow utilization of noncanonical NTP substrates containing, for example, a fluorine rather than a hydroxyl group at the ribose 2'-position. Such modifications can endow the RNAs with desired properties, such as resistance to RNA degrading enzymes in therapeutic applications. Such RNAs, synthesized with natural or modified NTPs, are commonly used for a wide variety of research purposes and as diagnostic and therapeutic reagents.

See also: [Molecular Biology: DNA Polymerase I, Bacterial](#); [DNA Polymerases: Kinetics and Mechanism](#); [DNA Replication, Mitochondrial](#); [HIV-1 Reverse Transcriptase Structures](#); [RNA Polymerase Structure, Bacterial](#).

Further Reading

- Cermakian N, Ikeda TM, Miramontes P, et al. (1997) On the evolution of the single-subunit RNA polymerases. *Journal of Molecular Evolution* 45(6): 671–681.
- Milligan JF, Groebe DR, Witherell GW, and Uhlenbeck OC (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Research* 15: 8783–8798.
- Mooney RA, Artsimovitch I, and Landick R (1998) Information processing by RNA polymerase: Recognition of regulatory signals during RNA chain elongation. *Journal of Bacteriology* 180: 3265–3275.
- Sousa R (1996) Structural and mechanistic relationships between nucleic acid polymerases. *Trends in Biochemical Sciences* 21: 186–190.
- Steitz TA (2006) Visualizing polynucleotide polymerase machines at work. *EMBO Journal* 25(15): 3458–3468.
- Steitz TA (2009) The structural changes of T7 RNA polymerase from transcription initiation to elongation. *Current Opinion in Structural Biology* 19(6): 683–690.
- Studier FW, Rosenberg AH, Dunn JJ, and Dubendorff JW (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology* 185: 60–89.

Tachykinin/Substance P/Neurokinin-1 Receptors

L E Goldsmith and M M Kwatra, Duke University Medical Center, Durham, NC, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Mark D. Richardson and Madan M. Kwatra, volume 4, pp. 152–157, © 2004, Elsevier Inc.

Glossary

Carboxyl tail The hydrophilic portion of a G-protein-coupled receptor (GPCR) that follows the final transmembrane-spanning domain (TM7), extending from the cytosolic face of the plasma membrane and terminating in the cytosolic compartment of the cell.

G-protein-coupled receptor (GPCR) Integral plasma membrane proteins having seven hydrophobic, transmembrane-spanning domains (designated TM1 through 7). GPCRs pass signals across the plasma membrane by transducing stimuli from extracellular signaling molecules to intracellular G proteins. G proteins, in turn, activate intracellular signaling cascades.

Phospholipase An enzyme that can hydrolyze the phosphodiester bond in phospholipids and produce soluble inositol lipids and membrane-bound diacylglycerol; type C phospholipases specifically use phosphoinositides as substrates.

Preprotachykinins The genes containing all the DNA sequence information needed to synthesize tachykinins. The actual production of tachykinins requires alternative splicing of messenger RNA (mRNA) transcripts derived from preprotachykinin, and posttranslational modification of peptides generated from the mRNA transcripts.

Tachykinin A 10–11-amino-acid peptide, having the amino acid sequence F-X-G-L-M-NH₂ at the C-terminal end, where X is a hydrophobic amino acid residue. Tachykinins are released from neurons and act as neurotransmitters.

Tachykinins

Tachykinins are peptides of 10–11 amino acids having the motif F-X-G-L-M-NH₂ at the C terminus; the –NH₂ group indicates that the C-terminal amino acid is amidated, a post-translational modification necessary for biological activity.

There are currently four known mammalian tachykinins: substance P (SP), neurokinin A (NKA), neurokinin B (NKB), and hemokinin 1 (HK-1). As illustrated in [Figure 1](#), these tachykinins are generated from three genes: preprotachykinin-A (PPT-A), preprotachykinin-B (PPT-B), and preprotachykinin-C (PPT-C). Alternative splicing of the PPT-A messenger RNA (mRNA) transcripts yields four products: α -PPT-A, β -PPT-A, γ -PPT-A, and δ -PPT-A. α -PPT-A and δ -PPT-A encode only SP, while β -PPT-A and γ -PPT-A encode both SP and NKA. PPT-B yields only one product, NKB, and PPT-C encodes HK-1. The products of PPT-A transcripts (SP and NKA) are found in the central nervous system (CNS) and across a range of peripheral tissues, while a product of PPT-B (NKB) is found only in the CNS. PPT-C is widely distributed in many peripheral organ systems, but is not detected in the CNS.

The actions of mammalian tachykinins are mediated by three receptors: neurokinin-1 receptor (NK1R), neurokinin-2 (NK2) receptor, and neurokinin-3 (NK3) receptor. These receptors have the following binding preferences: NK1R binds in the order SP>NKA>NKB; NK2 binds in the order NKA>NKB>SP; and NK3 binds in the order NKB>NKA>SP. Although SP, NKA, and NKB bind the tachykinin receptors with different affinities, they are all full agonists at all three receptors.

Mammalian Tachykinin Function

Substance P

SP is present in the outer laminae of the dorsal horn of spinal cord and in many areas of the brain, which is consistent with

SP acting as a neurotransmitter. The hypothesis that SP has a role in pain transmission stems from its presence in primary afferent nerve fibers of the spinal cord, such as C-fibers of laminae I and II of the dorsal horn. Efforts aimed at developing NK1R antagonists to treat pain have met with limited success, causing some doubt about the involvement of SP in pain transmission. However, recent experiments performed in genetically engineered mice, which lack expression of PPT-A, have given new support for the role of SP in pain. While mice lacking PPT-A display normal responses to a variety of painful stimuli, their responses become blunted as the stimulus increases in intensity. However, above a certain pain threshold, mice lacking PPT-A seem to display responses that are similar to control mice. These results suggest that SP is indeed a pain neurotransmitter, and it functions within a discrete window of pain intensities. Other recent studies in mice lacking NK1R also support the conclusion that SP plays a role in pain transmission.

SP also influences blood pressure. Intravenous injection of tachykinins in dogs and rabbits produces a strong hypotensive effect on mean arterial blood pressure, with SP showing much higher potency than NKA, NKB, or nonmammalian tachykinins. In human volunteers, intravenous SP causes a drop in diastolic, but not systolic, blood pressure, whereas NKA has no effect on either. Interestingly, the hypotensive effect is not diminished by autonomic blocking agents, such as atropine, atenolol, or prazosin. This observation suggests that SP acts directly on vascular smooth muscle.

Substance P/Neurokinin A

The PPT-A mRNA transcripts that produce NKA (β - and γ -PPT-A) also produce SP ([Figure 1](#)); thus, SP and NKA are cosynthesized and coreleased from neurons. This suggests that NKA does not act alone; however, NKA can act as the primary signaling peptide in situations where the ratio of NKA to SP

Human tachykinins			
Gene	mRNA	Product	Amino acid sequence
PPT A	α -PPT A	Substance P	RPKPQQFFGLM-NH ₂
	β -PPT A	Substance P Neurokinin A	HKTDSFVGLM-NH ₂
	γ -PPT A	Substance P Neurokinin A	
	δ -PPT A	Substance P	
PPT B		Neurokinin B	DMHDFVGLM-NH ₂
PPT C		Hemokinin 1	TGKASQFFGLM-NH ₂

Figure 1 Synthesis of mammalian tachykinins. The mammalian tachykinins are generated from three genes (PPT-A, -B, and -C). Amino acid sequences of the human tachykinins are shown; shaded boxes highlight the F-X-G-L-M-NH₂ motif.

favors NKA, and where NK2 (the preferred receptor for NKA) is the predominant receptor in the target tissue. NKA is recognized as a very potent bronchoconstrictor, more potent than SP or NKB. Blockade of nerve-evoked bronchoconstriction by an antiserum having high affinity for NKA indicates that NKA has a more important role than SP in bronchoconstriction. Furthermore, experiments using NK2-selective antagonists indicate a predominance of NK2 over NK1R in the mediation of noncholinergic bronchoconstriction responses. NKA is also more potent than SP in decreasing tracheal vascular resistance. Recent data indicate that NKA, like SP, binds to NK1R with high affinity; thus, the effects of NKA may potentially be mediated by either NK1 or NK2 receptor. In the CNS, which does not contain NK2 receptors, NKA produces biological effects similar to SP by interacting with NK1R.

Neurokinin B

Although SP and NKA frequently have overlapping tissue distribution, the distribution of NKB is distinct from that of SP and NKA. For example, NKB is present in spinal cord laminae III, while SP and NKA are abundant in laminae I and II and absent in laminae III. NKB is most abundant in the hypothalamus and in the intestines. The physiological effects of NKB include contraction of rat portal vein and inhibition of gastric acid secretion. Interestingly, NKB potently decreases alcohol intake by a strain of alcohol-preferring rats. Also of potential clinical importance, NKB dilates the vasculature in human placenta and is thought to have an important role in the development of preeclampsia.

Hemokinin 1

Discovered in 2000, the tachykinin HK-1 is unique in that it is not of neuronal origin. HK-1 was first cloned from mouse B-lymphocytes; subsequent cloning of human HK-1 showed that it differs from mouse HK-1 in 5 out of 11 amino acid residues.

Expression of human HK-1 has since been demonstrated in heart, skeletal muscle, thyroid, and skin. Weaker expression is detected in liver, lung, stomach, testes, placenta, and prostate. Receptor-binding assays in cultured cells and tissues show that HK-1 has a strong binding preference for NK1R over NK2 or NK3 receptors, and it is considered a full agonist of NK1R.

HK-1 is similar to SP in many physiological effects, often with a potency that is very close to that of SP. Intravenous injection of HK-1 or SP into guinea pigs produces a dose-dependent hypotensive effect, with HK-1 and SP decreasing diastolic blood pressure at similar potencies of 0.2 nmol kg⁻¹ and 0.1 nmol kg⁻¹, respectively. HK-1 and SP also induce salivary secretion in rats at essentially equal potencies. Studies are underway in several laboratories to develop a more complete understanding of the function of HK-1.

Neurokinin-1 Receptor

NK1R and the other tachykinin receptors belong to the G-protein-coupled receptor (GPCR) superfamily of receptors, which are characterized by the presence of seven hydrophobic transmembrane domains, designated TM1 through 7 (Figure 2). Cloning of human NK1R from glioblastoma cells yields two forms of the receptor. One clone, with a stop codon after amino acid 407, encodes a full-length receptor and the other, with a stop codon after amino acid 311 (Δ 311, Figure 2), encodes a truncated form of the receptor; the truncated form has also been cloned from guinea pig nervous system.

Full-Length NK1R

NK1R has three extracellular (EC) and three intracellular (IC) hydrophilic loops, a fourth IC loop formed by the palmitoylation of a cysteine residue, and in the case of the full-length receptor, a carboxyl tail. The deduced amino acid sequences of full-length rat and human NK1R reveal that these receptors are

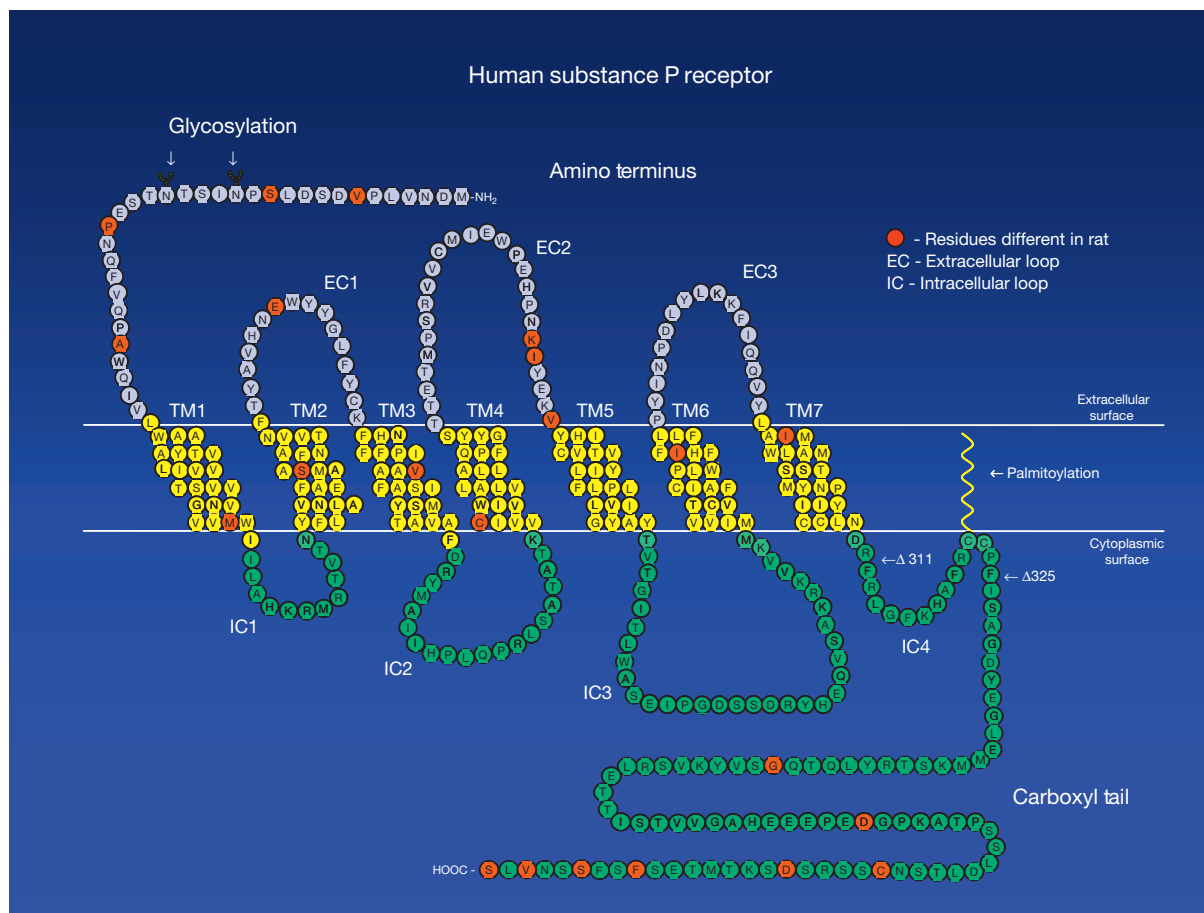


Figure 2 Two-dimensional representation of the human NK1R. The amino acid sequence of human NK1R is shown. Seven putative transmembrane regions are labeled TM1–7, three extracellular loops are labeled EC1 through 3, and four intracellular loops are labeled IC1 through 4. Shaded circles show where amino acid residues in rat NK1R differ from human NK1R. Arrows indicate the termination site of the short form of NK1R at Arg311, and truncation site of the mutant rat NK1R at Phe325.

407 amino acids long and share 95% amino acid identity; there are only 22 amino acids that differ between rat and human NK1R (see shaded circles in [Figure 2](#)).

Although rat and human NK1R are very similar at the amino acid level, they interact differently with SP and several nonpeptide antagonists of NK1R. For example, SP binds to rat NK1R with a K_d of 3 nM, whereas it binds to human NK1R with a higher affinity characterized by a K_d of 0.7 nM. The difference between rat and human NK1R pharmacology becomes even more striking for some of the nonpeptide antagonists. CP-96345, the first nonpeptide antagonist of NK1R to be synthesized, has 100-fold higher affinity for human NK1R than for rat NK1R, and the nonpeptide antagonist RP67580 has higher affinity for rat NK1R than for human NK1R.

As rat and human NK1R differ in only a few amino acid residues in the TM domains (a region believed to be involved in ligand binding), several groups have identified the amino acids responsible for differences in their pharmacology. These groups found that residue 290 in TM7 and the residues located on the second EC loop contribute to the selectivity of CP-96345 for human NK1R. Histidine 197 in TM5 is involved in the binding of CP-96345, whereas histidine 265 in TM6 is involved

in RP67580 binding. These mutagenesis studies not only identify the residues involved in conferring species-dependent pharmacology to NK1R, but also show that ligand binding to the receptor involves the TM domain as well as amino terminus and EC loops.

NK1R activation in most cells stimulates phospholipase C, which hydrolyzes membrane phosphoinositides into two second messengers: inositol triphosphate (IP_3) and diacylglycerol. These molecules stimulate intracellular calcium release and protein kinase C (PKC) activation, respectively. Studies using U373 human glioblastoma cells, in which NK1R occurs naturally, implicate several key molecules in NK1R signaling, including mitogen-activated protein (MAP) kinases ERK1/2, MAP kinase p38, nuclear transcription factor kappa B ($NF-\kappa B$), and epidermal growth factor receptor (EGFR). Of these molecules, ERK1/2 mediates NK1R-stimulated proliferation of U373 MG cells through nonreceptor tyrosine kinase Src. A recent study found that stimulation of NK1R leads to increased phosphorylation and activation of Akt, or protein kinase B, through total involvement of Src and PI-3-kinase, and partial involvement of EGFR. In addition, blockade of NK1R resulted in apoptosis of glioblastoma cells. These findings emphasize the relevance of the

NK1R signaling pathways to understanding the mechanism of glioblastoma growth.

Truncated NK1R

Functional studies in which the truncated (NK1R Δ 311) and full-length forms of NK1R were expressed in *Xenopus* oocytes revealed that the truncated receptor is 100-fold less active than full-length NK1R. Ligand binding analyses of the full-length and truncated forms of NK1R expressed in COS cells reveal that the truncated form binds SP with much lower affinity. More recent studies on full-length and truncated NK1R expressed in HEK 293 cells revealed several differences in signaling through these receptors, including differences in time course of ERK1/2 activation. Interestingly, truncated NK1R, but not the full-length NK1R, expressed in normal breast cells resulted in a transformed phenotype. Furthermore, the expression of truncated NK1R, but not the full-length, was controlled by NF- κ B.

Role of the Carboxyl Tail in NK1R Function

Potentially, the full-length and truncated NK1 receptors could mediate SP signaling differently in different tissues due to the presence or absence of the carboxyl tail; thus, it is important to understand the function of the carboxyl tail in NK1R biology.

One function of NK1R in which the carboxyl tail plays an important role is receptor desensitization. Like many other GPCRs, NK1R undergoes agonist-induced desensitization, a phenomenon in which the responsiveness of the receptor diminishes in spite of the continued presence of the stimulus. Studies performed by Dr. Lefkowitz at Duke University Medical Center and by several other investigators have shown that agonist-specific or homologous desensitization of GPCRs proceeds through two cytosolic proteins, G protein-coupled receptor kinases (GRKs) and β -arrestins. GRKs phosphorylate agonist-occupied receptor at serine and threonine residues on the carboxyl tail and IC loops, and this phosphorylation primes the receptor to bind β -arrestins. The binding of β -arrestin to the intracellular face of the receptor disrupts interactions between the receptor and G protein, resulting in a loss of signaling (Figure 3). β -Arrestin binding also directs receptor internalization through clathrin-coated pits.

As the carboxyl tail plays a key role in GPCR desensitization, its absence would be expected to impair the ability of truncated NK1R to desensitize. However, studies examining the role of the C terminus in rat NK1R desensitization have yielded varying results. It has been reported that a C-terminally truncated mutant of rat NK1R Δ 325 does not desensitize fully, while other studies have reported no loss of desensitization in the same truncation mutant. We had also examined the effect of C-terminal truncation on the desensitization of human

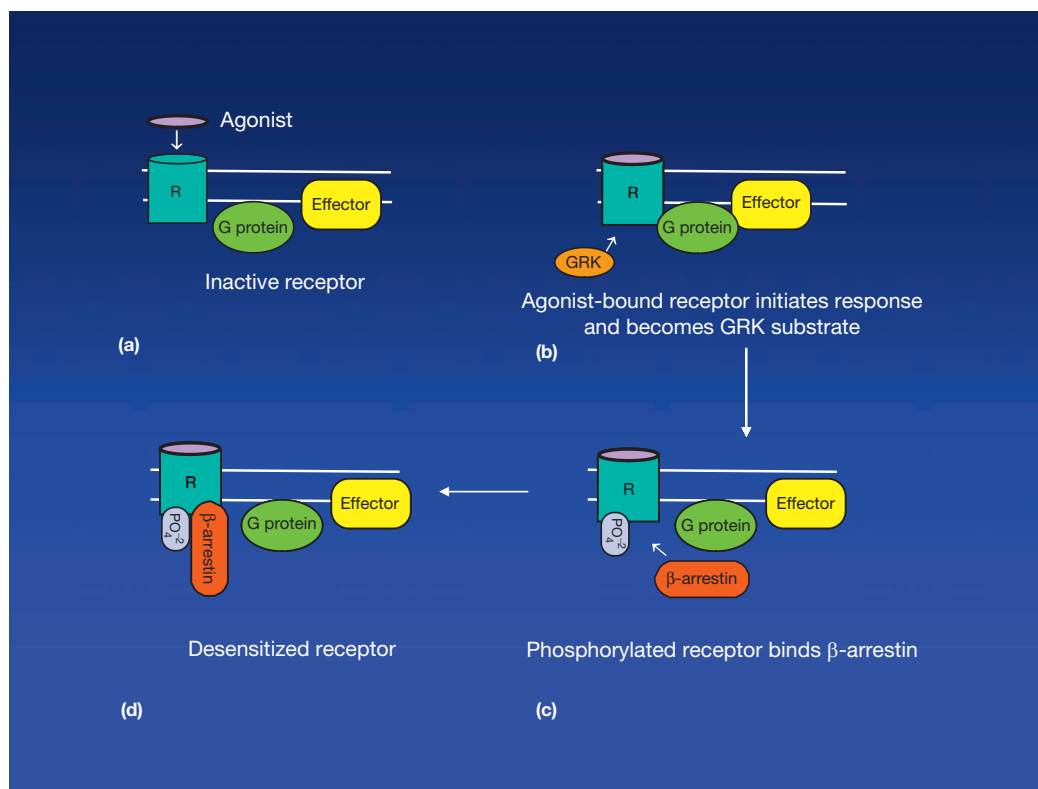


Figure 3 Model of homologous GPCR desensitization. (a) In the resting state, receptor, G protein and effector are inactive. (b) Occupancy of the receptor by agonist enables it to catalyze G-protein activation, leading to activation of the effector by G protein. Receptor occupancy also makes it a target for phosphorylation by GRK. (c) Phosphorylation diminishes the ability of the receptor to activate G protein; the effector then reverts to the inactive state. Phosphorylation also facilitates the association of β -arrestin with the receptor. (d) Association of β -arrestin with phosphorylated receptor prevents any further G-protein activations even though agonist is present.

NK1R by preparing a truncated mutant that ended at amino acid 325 (NK1R Δ 325). Full-length and truncated (Δ 325) human NK1R desensitize similarly, using two independent readouts of receptor signaling, intracellular IP₃ accumulation and subcellular redistribution of fluorescently tagged PKC- β II in live cells. However, truncated and full-length NK1R differ, in GRK-catalyzed phosphorylation and in their ability to form endocytic vesicles with two distinct β -arrestins, β -arrestin 1 and 2. Truncated NK1R, unlike full-length NK1R, does not undergo phosphorylation following exposure to SP. Also, full-length NK1R forms endocytic vesicles equally well with either β -arrestin 1 or β -arrestin 2, while truncated NK1R does not form any endocytic vesicles with β -arrestin 1 and forms fewer vesicles with β -arrestin 2. These data indicate that full-length and truncated NK1R are likely to behave differently during endocytosis and receptor recycling.

The Role of NK1R in Human Disease

While SP and NK1R have been studied for several decades, renewed interest in NK1R biology was generated following the report in 1998 that NK1R antagonists have antidepressant activity. This discovery generated so much enthusiasm that a commentary titled 'Reward for persistence in substance P research' by Claes Wahlestedt appeared in *Science* along with a timeline highlighting important events throughout the history of SP research (Figure 4). The first NK1R antagonist was introduced into clinical medicine as an antiemetic, and active research continues to investigate the role of NK1R in the pathophysiology of several diseases, including emesis, depression, and glioblastoma. Recent preliminary research has also implicated NK1R in HIV, fibromyalgia, and lung inflammation.

NK1R in Emesis

An important role for NK1R in emesis is indicated by clinical trials showing that the nonpeptide NK1R antagonist, MK-869 (also called aprepitant), is effective against prevention of both chemotherapy-induced nausea and vomiting (CINV) and post-operative nausea and vomiting (PONV). NK1R influences emesis through its role as a key regulator of the emetic reflex arc in the brainstem. Clinical trial investigators noted that during the acute phase of CINV (the first 24 h), aprepitant was only effective in 37% of patients, while the standard therapy, ondansetron, was effective in 57% of patients. By contrast, during the delayed phase of CINV (the succeeding 6 days)

aprepitant was effective in 72% of patients while ondansetron was effective in only 7% of patients. A later clinical trial compared standard therapy of ondansetron and dexamethasone with standard therapy plus aprepitant. This study found statistically significant improvement in percentages of patients achieving complete response, as indicated by no emesis or use of rescue therapy, in the overall study period (72.7% vs. 52.3%), acute phase (89.2% vs. 78.1%), and delayed phase (75.4% vs. 55.8%). Taken together, these two studies suggest that acute and delayed phase nausea may involve different molecular mechanisms and that SP is primarily active in the delayed phase of CINV.

Another clinical trial tested a triple combination of aprepitant, granisetron, and dexamethasone for suppression of CINV. In this trial, 87% of patients receiving the triple combination reported that nausea did not affect their ability to perform daily functions, in comparison to 67% of patients receiving only granisetron and dexamethasone. In 2003, aprepitant received Food and Drug Administration (FDA) approval for use in combination with other antiemetic drugs to prevent CINV, which is a particular problem in high-dose cisplatin therapy, followed by approval for combination therapy to prevent PONV in 2006. This result points to NK1R as a promising target protein for future efforts to understand and control nausea. Indeed, several NK1R antagonists are currently in development for treating CINV, including casopitant, which has completed phase 3 clinical trials.

NK1R in Depression and Affective Disorders

Substantial research implicates SP in the regulation of depression, anxiety, stress, and other affective disorders, and the distribution of SP and NK1R in the brain is consistent with this role. For instance, SP is the predominant tachykinin in human brain, and the expression of SP is particularly concentrated in areas of the brain that control either affective behavior or stress responses, such as the hypothalamus, amygdala, habenula, periaqueductal gray, and dorsal raphe nucleus. Likewise, NK1R is the most highly expressed tachykinin receptor in the brain and is also concentrated in areas that are important for affective behavior.

Initial evidence of a role for SP in depression was found in a randomized, double-blind, placebo-controlled study of patients with moderate to severe major depression, which showed that the NK1R antagonist, aprepitant, has antidepressant activity equivalent to paroxetine, a current standard therapy. However, pursuant phase 3 trials were unable to confirm early efficacy

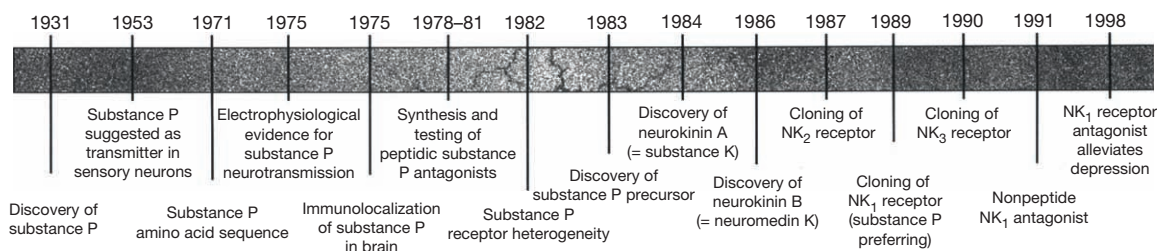


Figure 4 A timeline of landmark events in substance P research. Reproduced from Wahlestedt C (1998) Rewards for Persistence in Substance P Research. *Science* 281: 1624–1625.

signals and do not support further development of aprepitant in patients with major depression or high anxiety. Regardless, the improved tolerability profile expected from NK1R antagonists continues to drive research in this therapeutic area, given the medication adherence and discontinuation issues caused by significant side effects associated with standard of care antidepressant treatments. Specifically, a second NK1R antagonist, L759274, has shown positive results in clinical trials, further supporting a role for NK1R in depression. Future efforts may study the therapeutic potential of dual antagonists targeting both NK1R and NK2R for more complete blockade of SP. Furthermore, attenuation of stress-activated pathways by NK1R antagonists in preclinical models suggests therapeutic potential in stress-related disorders.

NK1R in Glioblastomas

Glioblastomas, classified as grade IV astrocytomas, are among the most aggressive and frequently occurring primary brain tumors in adults. Patients with glioblastomas have an extremely poor prognosis: with a 5-year survival rate of 1%, the currently available treatments of surgery, radiation therapy, and chemotherapy are inadequate. Current research focuses on understanding glioblastoma biology and identifying molecular targets for blocking tumor growth. SP has been observed to have a growth-promoting effect in a variety of cell types, and its role in glioblastoma growth is currently being studied. One study noted NK1R expression in 9 out of 12 astrocytomas, and 10 out of 10 glioblastomas. Furthermore, the density of NK1R correlates with the degree of malignancy, with glioblastomas expressing more receptors than astrocytomas. At the molecular level, NK1R stimulation in U373 MG human glioblastoma cells increases mitogenesis, cell proliferation, and release of interleukin-6 (IL-6). NK1R-dependent release of IL-6 is noteworthy because IL-6 has been implicated in the progression of gliomas. Consistent with an important role in glioblastoma biology, NK1R antagonists are reported to inhibit the growth of glioblastomas in nude mice. These findings, along with results from our laboratory indicating the presence of a constitutively active form of NK1R and NK1R-mediated apoptosis of glioblastoma cells, suggest a therapeutic potential for NK1R antagonists in glioblastoma management.

A diagnostic or therapeutic role has been implicated for NK1R in other carcinomas as well, including breast, pancreatic,

and lung. Preclinical studies suggest that NK1R overexpression in these tissues may indicate and facilitate tumor metastasis. Thus, the potential for the clinical introduction of NK1R antagonists to cancer treatment will remain an active area of research.

See also: [Signaling: G-Protein-Coupled Receptor Kinases and Arrestins; Neurotransmitter Transporters; Phospholipase C.](#)

Further Reading

- Akazawa T, Kwatra SG, Goldsmith LE, et al. (2009) A constitutively active form of neurokinin 1 receptor and neurokinin 1 receptor-mediated apoptosis in glioblastomas. *Journal of Neurochemistry* 109: 1079–1086.
- De Felipe C, Herrero JF, O'Brien JA, et al. (1998) Altered nociception, analgesia and aggression in mice lacking the receptor for substance P. *Nature* 392: 394–397.
- Hökfelt T, Pernow B, and Wahren J (2001) Substance P: A pioneer amongst neuropeptides. *Internal Medicine Journal* 249: 27–40.
- Holmes A, Heilig M, Rupniak NMJ, Steckler T, and Griebel G (2003) Neuropeptide systems as novel therapeutic targets for depression and anxiety disorders. *Trends in Pharmacological Sciences* 24: 580–588.
- Kohout TA and Lefkowitz RJ (2003) Regulation of G protein-coupled receptor kinases and arrestins during receptor desensitization. *Molecular Pharmacology* 63: 9–18.
- Lai JP, Lai S, Tuluc SF, et al. (2008) Differences in the length of the carboxyl terminus mediate functional properties of neurokinin-1 receptor. *Proceedings of the National Academy of Sciences of the United States of America* 105: 12605–12610.
- Maggi CA (2000) Principles of tachykinergic co-transmission in the peripheral and enteric nervous system. *Regulatory Peptides* 93: 53–64.
- Mantyh PW (2002) Neurobiology of substance P and the NK1 receptor. *Journal of Clinical Psychiatry* 63: 6–10.
- Palma C and Maggi CA (2000) The role of tachykinins via NK₁ receptors in progression of human gliomas. *Life Sciences* 67: 985–1001.
- Patel HJ, Ramkissoon SH, Patel PS, and Rameshwar P (2005) Transformation of breast cells by truncated neurokinin-1 receptor is secondary to activation by preprotachykinin-A peptides. *Proceedings of the National Academy of Sciences of the United States of America* 102: 17436–17441.
- Ramkissoon SH, Patel PS, Taborga M, and Rameshwar P (2007) Nuclear factor-kappa B is central to the expression of truncated neurokinin-1 receptor in breast cancer: Implication for breast cancer cell quiescence within bone marrow stroma. *Cancer Research* 67: 1653–1659.
- Rittenberg CN (2002) A new class of antiemetic agents on the horizon. *Clinical Journal of Oncology Nursing* 6: 103–104.
- Severini C, Improta G, Falconieri-Erspamer G, Salvadori S, and Erspamer V (2002) The tachykinin peptide family. *Pharmacological Reviews* 54: 286–322.
- Wahlestedt C (1998) Substance P and related neuropeptides. *Science* 28: 1624–1625.
- Zimmer A, Zimmer AM, Baffi J, et al. (1998) Hypoalgesia in mice with a targeted deletion of the tachykinin 1 gene. *Proceedings of the National Academy of Sciences of the United States of America* 95: 2630–2635.

Taste Receptors

J D Boughter Jr., University of Tennessee Health Science Center, Memphis, TN, USA

S D Munger, University of Maryland School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Taste quality The perceptual quality attributed to a taste stimulus, such as salty, sour, sweet, bitter, or umami.

Taste receptor cell Elongate, spindle-shaped cell in taste buds that expresses taste receptors and/or transduction components.

Taste receptor Membrane protein that interacts with taste stimuli, leading to activation of a transduction cascade, receptor cell depolarization, and transmitter release onto afferent nerve fibers.

Transduction The process, often involving a biochemical cascade, by which one type of signal (e.g., a taste stimulus) is converted to another type (e.g., depolarization of the TRC).

Taste Receptor Cells

The initial events in taste processing occur in taste buds, structures found in the epithelia of the tongue, palate, larynx, and epiglottis of mammals, which contain approximately 50–100 cells of neuroepithelial origin (**Figure 1**). Taste buds contain morphologically distinct cell types, including elongate, spindle-shaped cells that express a variety of identified transduction-related proteins (including taste receptors) and therefore likely function as taste receptor cells (TRCs). TRCs possess apical processes with microvilli that protrude into the oral cavity and interact with taste stimuli. These microvilli contain taste receptors and taste-responsive ion channels. Tight junctions between the cells in a taste bud restrict the access of most stimuli to the apical membranes of the cells. TRCs synapse with afferent special sensory fibers from one of three cranial nerves (VII, IX, and X), and taste information is subsequently relayed through brainstem nuclei to forebrain taste areas or to local circuits controlling oromotor reflexes.

Multiple signaling cascades in TRCs have been described, and there are fundamental differences in the transduction pathways associated with each stimulus quality. For example, salt and acid stimuli directly permeate or activate apical ion channels, leading to TRC depolarization, while umami-, sweet-, and bitter-tasting stimuli are thought to predominantly activate taste receptors that can initiate an intracellular cascade of signaling molecules. Receptors for bitter, sweet, umami, salty, and sour taste appear to be expressed in nonoverlapping populations of TRCs. Stimulus-dependent release of adenosine triphosphate (ATP) from at least some TRCs leads to direct activation of purinergic receptors present on afferent nerve fibers, and may be a primary mechanism by which taste information is transmitted from taste bud to sensory nerves. Recent findings also support the potential for information processing within taste buds via several paracrine and autocrine signaling mechanisms.

G-Protein-Coupled Taste Receptors

The largest gene superfamily in the mammalian genome encodes the group of proteins known as G-protein-coupled

receptors (GPCRs). GPCRs play critical roles in a variety of cellular functions, including neurotransmitter and hormonal signaling and the detection of sensory stimuli. Although they are quite diverse in amino acid sequence, all GPCRs share common structural and functional features. Their most striking structural motif is their seven helical transmembrane domains; that is, the GPCR polypeptide passes seven times through the plasma membrane, leaving an extracellular amino terminus and intracellular carboxy terminus. GPCRs also share a common signaling mechanism, as is evidenced by their name: activation of a GPCR by its ligand initiates an intracellular signaling cascade via the stimulation of an associated heterotrimeric guanosine triphosphate (GTP)-binding protein (G protein). Most chemosensory receptors are GPCRs, including odorant receptors, vomeronasal receptors, and taste receptors. In the taste system, two families of GPCR-type taste receptors (T1Rs and T2Rs) detect sweet-, umami-, and bitter-tasting stimuli.

T1R Receptors

Sweet and umami are appetitive taste qualities that drive intake of food. Diverse compounds evoke sweet taste, including some sugars, D-amino acids and proteins, as well as numerous artificial sweeteners. By contrast, umami taste is evoked by a small number of L-amino acids and 5'-ribonucleotides found in meats and other foods. The receptors for sweet and umami taste are called T1Rs, and function as heterodimeric complexes. Specifically, T1R1 and T1R3 together form a functional receptor for umami-tasting stimuli (e.g., L-amino acids) but are unresponsive to sweet-tasting stimuli, while T1R2 and T1R3 together comprise a receptor for sugars, artificial sweeteners, sweet proteins, and D-amino acids. The native T1R1 and T1R2 subunits are co-expressed with T1R3 in different subsets of TRCs. Thus, T1R3 is a common subunit for receptors with different ligand specificities. Furthermore, these ligand specificities largely depend on the whether the receptor complex contains T1R1 or T1R2.

T1R receptors are class C GPCRs, a group that also includes receptors for glutamate, γ -amino butyric acid and calcium. Class C GPCRs exhibit a large extracellular domain at the amino terminus that comprises approximately 50% of the protein length, a membrane-spanning domain that includes

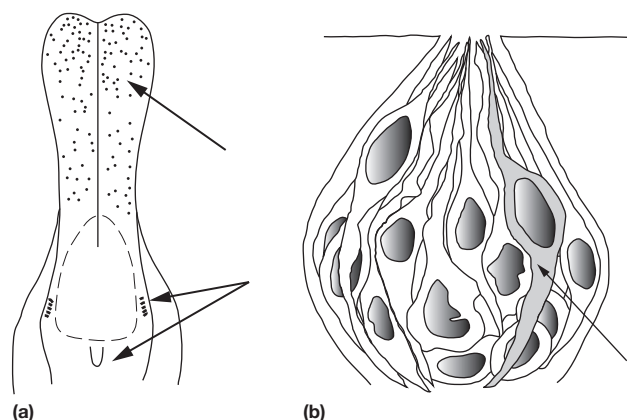


Figure 1 Taste receptor cells. (a) Taste buds are specialized sensory structures located in the epithelia of the mammalian tongue (arrows indicate approximate locations of taste buds), as well as palate, larynx, and epiglottis (not shown). (b) Schematic of a taste bud. Taste buds contain 50–100 cells of morphologically distinct types, including elongate, spindle-shaped cells (arrow) that express taste receptors.

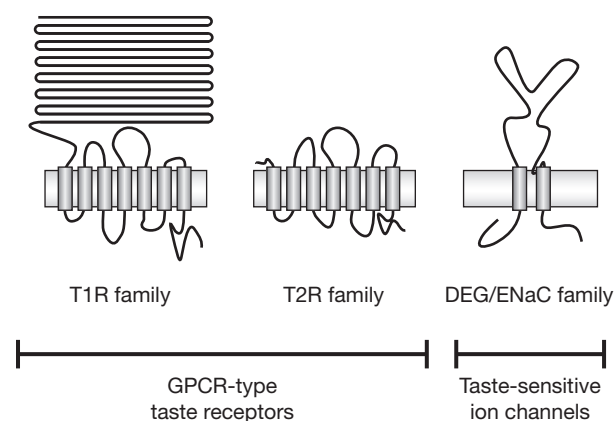


Figure 2 The proposed structures of taste receptor proteins. Amino acid and sweet- and bitter-tasting stimuli are detected by GPCR-type taste receptors. Salt and acid stimuli activate TRCs through a GPCR-independent mechanism; they directly permeate or modulate ion channels, some of which may belong to the DEG/ENaC family.

seven transmembrane helices, and a linker domain joining the two (Figure 2). The amino-terminal domain contains the principal site of ligand binding in class C GPCRs. In the case of T1Rs, this domain contains the binding site(s) for natural sugars, amino acids, and some artificial sweeteners (e.g., sucralose and aspartame). However, T1Rs also contain several other binding sites (e.g., in the transmembrane domain of T1R3) that can bind selected compounds such as the artificial sweetener cyclamate and the sweet taste inhibitor lactisole.

In the gustatory system, T1Rs are found on subsets of taste cells that are linked to a single perceptual quality. For example, T1R2 and T1R3 are co-expressed only in taste cells that elicit a sensation of sweetness, while T1R1 and T1R3 are expressed together in a separate TRC subpopulation that elicit an umami sensation when activated. However, it is now clear that T1Rs (and T2Rs; see below) are found in numerous tissues

outside the gustatory system, including the stomach, intestine, pancreas, and brain. In the gastrointestinal system, T1Rs and associated signaling molecules have been implicated in the regulation of glucose assimilation and glucose-dependent hormonal secretion from enteroendocrine cells. Thus, it appears that mammals use T1Rs as sugar and amino acid sensors throughout the body.

T2R Receptors

A second class of GPCRs, the T2Rs, also function as taste receptors. T2Rs respond to a large number of bitter-tasting compounds. Mammals have a large number of T2R genes (humans have approximately 25 functional T2R genes as well as several apparent pseudogenes). T2Rs loosely resemble class A GPCRs (the same class as mammalian odorant receptors and the light-sensitive receptor rhodopsin), which are characterized by a short amino terminus (Figure 2). Many class A GPCRs bind their ligands within a pocket defined by several of the transmembrane helices and/or the nearby extracellular loops. It has been proposed that both the transmembrane helices and extracellular loops are involved in ligand binding.

Ligands for most human T2Rs have been identified. Individual T2Rs each recognize multiple bitter-tasting stimuli, although the degree of ligand selectivity of each T2R can vary considerably. For example, hT2R14 responds to a structurally diverse set of stimuli, whereas just two agonists activate the recently deorphanized receptor hT2R50. Some T2R receptors, such as hT2R38 and hT2R16, appear to recognize common structural elements in chemically related compounds. Finally, specific compounds have been shown to activate multiple receptors, although sometimes with different efficacies.

The bitter receptor gene family in vertebrates comprises a relatively conserved family size (~15–37 functional genes) with multiple orthologs among mammalian species. This gene family also exhibits a fairly high rate of nonsynonymous mutations, especially in the portions of the genes that code for extracellular domains. It is thought that haplotype frequencies of T2R genes are strongly affected by natural selection, and are moreover associated with factors such as dietary preference and eating behavior.

T2Rs, like T1Rs, are also found in various tissues where they may act as sensors of diverse chemicals. For example, both bitter-tasting compounds and substances produced by pathogenic bacteria can activate T2R-expressing sensory cells in the nasal cavity and promote respiratory reflexes. However, it remains unclear if T2Rs mediate aversive signals in all tissues. Indeed, T2Rs are co-expressed with T1Rs in gut enteroendocrine cells, and may promote similar hormonal responses.

Ion Channels as Taste Receptors

Taste cells rely on the actions of a variety of ion channels to support stimulus-induced changes in membrane polarization and the synaptic release of neurotransmitter. However, several ion-channel species can be thought of as receptors for taste stimuli: the channels interact directly with the taste stimulus and participate in the initial stage of taste transduction. Such channels play a central role in the detection of

salty- and sour-tasting stimuli. In rodents, a large portion of salt taste is dependent on the influx of Na^+ through the epithelial sodium channel ENaC. The ENaCs, members of the degenerin (DEG)/ENaC superfamily of ion channels, are strongly Na^+ selective. DEG/ENaC family members display a common topology: they have two transmembrane domains, a large extracellular loop and a small pore-forming loop (Figure 2). They function as hetero-oligomeric complexes. Data supporting a role for ENaCs in sodium taste are compelling. However, ENaC-independent mechanisms are also implicated in mammalian salt taste, especially for nonsodium salts such as KCl.

Sour (acid) taste stimuli (i.e., H^+) have been proposed to depolarize TRCs in vertebrates via several different mechanisms, including by activating cation channels or by directly permeating ion channels. PKD2L1, a polycystic-kidney-disease-like ion channel expressed in sour-sensitive TRCs has been suggested as a candidate sour taste receptor. However, evidence supporting a primary role in sour taste remains equivocal. Other candidates for sour taste receptors include members of a family of proton-gated ion channels related to the superfamily of DEG/ENaC sodium channels family, the acid-sensing ion channels (ASICs).

Transduction Cascades

The mechanisms of taste transduction are still not fully understood. The G-protein subunit α -gustducin is implicated in sweet-, umami-, and bitter-taste transduction, although other

G proteins are also involved. TRCs express many second-messenger components common to GPCR-coupled cascades such as cyclic nucleotide and phosphoinositide signaling systems. The effector enzyme phospholipase C $\beta 2$, the second messenger inositol 1,4,5-trisphosphate (IP_3), the IP_3 receptor type 3, and the Ca^{2+} -sensitive transient receptor potential channel TRPM5 all play important roles in the transduction of sweet-, umami-, and bitter-taste stimuli. The role of cyclic nucleotide signaling is less clear, but may contribute to adaptation. The influx of sodium ions via TRPM5 may contribute to the receptor potential and/or mediate neurotransmitter secretion onto afferent nerve fibers, which communicate taste information to the central nervous system.

See also: [Cell Architecture and Function: Olfactory Receptors.](#)

Further Reading

- Bachmanov AA and Beauchamp GK (2007) Taste receptor genes. *Annual Review of Nutrition* 27: 389–414.
- Behrens M and Meyerhof W (2009) Mammalian bitter taste perception. *Results and Problems in Cell Differentiation* 47: 203–220.
- Boughter JD, Jr. and Bachmanov AA (2007) Behavioral genetics and taste. *BMC Neuroscience* 8: S3.
- Vigues S, Dotson CD, and Munger SD (2009) The receptor basis of sweet taste in mammals. *Results and Problems in Cell Differentiation* 47: 187–202.
- Yarmolinsky DA, Zuker CS, and Ryba NJ (2009) Common sense about taste: From mammals to insects. *Cell* 139: 234–244.

T-Cell Antigen Receptor

A L Szymczak-Workman, L L Jones, C S Guy, and D A A Vignali, St. Jude Children's Research Hospital, Memphis, TN, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Andrea L. Szymczak and Dario A. A. Vignali, volume 4, pp. 162–168, © 2004, Elsevier Inc.

Glossary

CD3 Complex of polypeptides containing three dimers: $\epsilon\gamma$ heterodimers, $\epsilon\delta$ heterodimers, and, most frequently, $\zeta\zeta$ homodimer (CD247). It is associated with the T-cell antigen receptor (TCR) through charged transmembrane residues and is involved in transducing signals into the T cell upon TCR:peptide–MHC interaction.

Complementarity-determining region (CDR) Areas in the variable regions of antibody and TCR genes. In the TCR, the CDR regions contact the peptide and major histocompatibility complex (MHC) molecule on antigen presenting cells.

Immunoglobulin superfamily Group of proteins that contain immunoglobulin-fold domains of ~100 amino

acids folded into two β -pleated sheets and stabilized by a central disulfide bond. Included in the family are MHC molecules, TCRs, and a number of CD antigens.

Major histocompatibility complex (MHC) A complex of polymorphic genes that encode histocompatibility antigens termed 'H2' in the mouse and 'HLA' in humans. Two main classes of MHC antigens are found as surface glycoproteins on antigen-presenting cells that bind and present peptides to T cells.

T-cell antigen receptor (TCR) Heterodimer of $\text{TCR}\alpha\beta$ or $\text{TCR}\gamma\delta$ expressed on the surface of T cells that is associated with the CD3 complex. The TCR binds to peptide–MHC molecules.

The T-cell antigen receptor (TCR) is a highly organized, multi-molecular complex found exclusively on T and NK T cells. The majority of TCRs are composed of $\text{TCR}\alpha$ and $\text{TCR}\beta$ chains, while a small percentage contain $\text{TCR}\gamma$ and $\text{TCR}\delta$ chains. The TCR chains heterodimerize and are associated with the invariant chains of the CD3 complex, $\text{CD3}\epsilon\gamma\delta$ and CD247 (often referred to as the 'zeta chain', $\text{TCR}\zeta$ or $\text{CD3}\zeta$). Each chain is essential for TCR surface expression and consequently T-cell development and function. This complex mediates the development of T cells that are able to mount a specific immune response against a wide variety of foreign antigens and pathogenic organisms. This unique capability is due to the unusual genomic organization and rearrangement of TCR genes, which generates a vast number of TCRs that recognize almost any antigenic peptide in the context of major histocompatibility (MHC) molecules.

TCR:CD3 Complex: Genes, Proteins, and the Receptor Complex

T-Cell Antigen Receptor

T cells can only recognize and bind antigens when they are presented by MHC molecules, a process called 'MHC restriction'. Initially, the TCR was difficult to study because it was membrane bound and could only recognize antigen in the context of MHC. The first TCR complexes isolated were composed of a disulfide-linked heterodimer of $\text{TCR}\alpha$ and $\text{TCR}\beta$. Monoclonal antibodies against this complex were either specific for a particular T-cell clone or nonspecific, indicating that the TCR chains were similar to the immunoglobulin (Ig) molecules, containing both variable (V) and constant (C) regions. While the vast majority (>95%) of T cells had $\text{TCR}\alpha\beta$ heterodimers, another type of TCR containing a $\text{TCR}\gamma$ and $\text{TCR}\delta$ heterodimer was later identified. The TCR chains, in particular,

the variable regions, are responsible for recognizing and binding to peptide–MHC molecules presented on the surface of antigen-presenting cells (APCs).

Genes: organization and rearrangement

As with antibody molecules, T cells must be able to recognize a wide variety of pathogens; therefore, a diverse repertoire of TCR must be generated that recognize such antigens. Many similarities exist between the Igs and the TCR. The four TCR loci (α , β , γ , and δ) have a germ-line organization similar to that of Ig. The α - and γ -chains are produced by rearrangement of V and J segments, while the β - and δ -chains are produced by rearrangement of V, D, and J segments (Figure 1). The segments for $\text{TCR}\delta$ are located between the $V\alpha$ and $J\alpha$ segments. When the $\text{TCR}\alpha$ gene is rearranged, the $\text{TCR}\delta$ segments are deleted, thereby ensuring that the T cell does not express $\text{TCR}\gamma\delta$ and $\text{TCR}\alpha\beta$ at the same time. The basic domain structure is evolutionarily conserved and is believed to have arisen through gene duplication. Organization of the TCR genes in the mouse and human is similar but varies in the numbers of segments (Table 1).

The TCR genes encode areas of hypervariability similar to the complementarity-determining regions (CDRs) of antibody molecules. Embedded within the V regions are CDR1 and CDR2. There is an additional area of hypervariability (HV4) that does not appear to interact with antigen and is therefore not considered a CDR (Figure 1). CDR3 is a result of VJ or VDJ joining. This intentionally imprecise process frequently results in the addition or deletion of nucleotides, adding significantly to the variability in this region (N region diversity). Occasionally, a nonproductive rearrangement of the first gene locus occurs due to an out-of-frame rearrangement or insertion of a stop codon. This induces the rearrangement of the TCR locus on the other chromosome.

Rearrangement occurs through the action of recombination-activating genes (RAGs) 1 and 2 during T-cell development. Flanking each gene segment in the germ-line DNA are

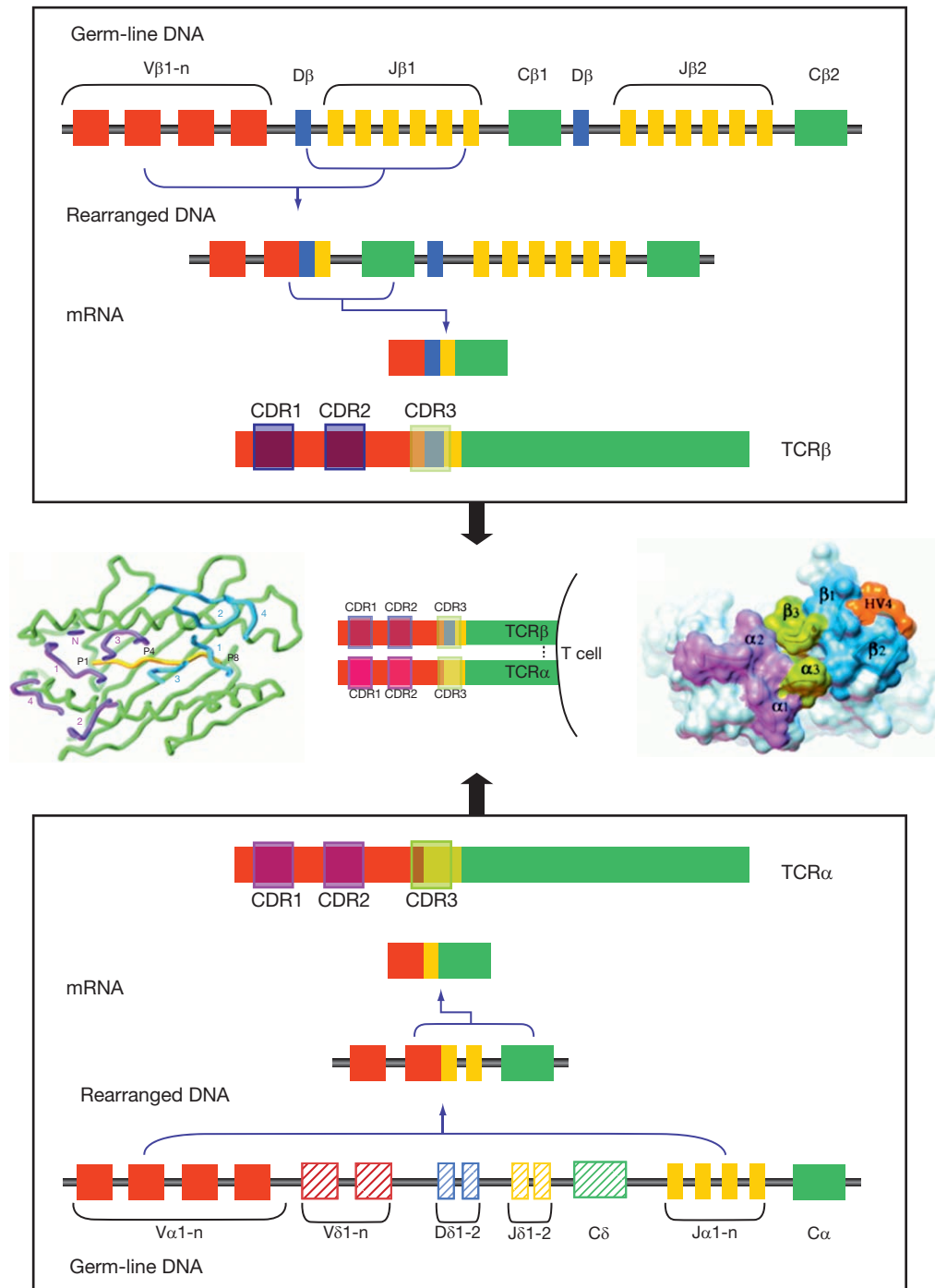


Figure 1 Rearrangement of TCR $\alpha\beta$ genes. Example of TCR α (bottom panel) and TCR β (top panel) rearrangement. TCR α undergoes V to J rearrangement while TCR β undergoes D to J followed by V to DJ rearrangement. Primary transcripts of rearranged DNA are processed to mRNA that encode the mature TCR $\alpha\beta$ chains expressed on the T-cell surface (center panel, middle). Areas encoding the CDR regions are boxed. TCR δ segments in the TCR α locus are shown in stripes. The center left panel shows a view into the peptide-MHC combining site with peptide in yellow and MHC in green superimposed with the CDR regions (CDR1, 2, and 3; HV4) of TCR α (purple) and TCR β (blue). CDR1 of V α and V β and V β CDR3 are clearly positioned along the central axis of the peptide. N, the N-terminal residue of V α . Reprinted from Garcia KC, Degano M, Pease LR, et al. (1998) Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. *Science* 279: 1166–1172. The center right panel shows the surface of the TCR-binding site. The surface of the loop trace of the V α CDRs 1 and 2 are purple; CDRs 1 and 2 of TCR β are blue; V α and V β CDR3s are green; and the hypervariable region of TCR β (HV4) is orange. Reprinted from Garcia KC, Degano M, Stanfield RL, et al. (1996) An $\alpha\beta$ T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* 274: 209–219.

Table 1 Chromosomal location and numbers of TCR gene segments

	Human				Mouse			
	<i>TCRα(14)</i>	<i>TCRβ(7)</i>	<i>TCRγ(7)</i>	<i>TCRδ(14)</i>	<i>TCRα(14)</i>	<i>TCRβ(6)</i>	<i>TCRγ(13)</i>	<i>TCRδ(14)</i>
V	54	76	14	3	98	35	7	6
D	—	2	—	3	—	2	—	2
J	61	14	5	4	60	14	4	2
C	1	2	2	1	1	7	4	1

Total number of gene segments for the human and mouse TCR. Numbers in parenthesis indicate the chromosome on which the gene is located. Genetic information acquired from the international ImMunoGeneTics information system (www.imgt.org).

Table 2 Sequence diversity in TCR and Ig genes

	<i>Ig</i>		<i>TCRαβ</i>		<i>TCRγδ</i>	
	<i>H</i>	<i>κ</i>	<i>α</i>	<i>β</i>	<i>γ</i>	<i>δ</i>
V segments	250-1000	250	50	25	7	10
D segments	10	0	0	2	0	2
Ds read in all frames	Rarely	—	—	Often	—	Often
N-region addition	V-D, D-J	None	V-J	V-D, D-J	V-J	V-D1, D1-D2, D1-J
J segments	4	4	50	12	2	2
V region combinations	62,500-250,000		1250		70	
Junctional combinations	~10 ¹¹		~10 ¹⁵		~10 ¹⁸	

With permission, from the *Annual Review of Biochemistry*, Volume 59 ©1990 by Annual Reviews www.annualreviews.org

conserved recognition signal sequences. The RAG enzymes recognize these heptamer and nonamer sequences and catalyze the VJ and VDJ joining with a mechanism analogous to that used to rearrange the Ig locus. In addition, secondary rearrangements termed 'receptor editing' or 'revision' can occur in both B and T cells. The role of receptor editing and revision in T cells is still unclear, but it has been suggested to contribute to peripheral T-cell tolerance.

Although many similarities exist between Ig and TCR, there is one important distinction: TCR genes do not undergo somatic hypermutation and thus there is no affinity maturation. While the number of TCR V region segments is dramatically reduced compared to the Ig genes, the number of J region gene segments is greater, increasing the potential diversity at the V-J interface (Table 2). This is likely due to differences in the nature of the antigen recognized by TCR and Ig.

TCR structure and recognition of peptide-MHC complexes

The extracellular domain of the TCR is heterodimeric, with each chain consisting of a membrane-proximal constant domain (Cα and Cβ) and a membrane-distal variable domain (Vα and Vβ), formed by the recombination events described above. Structural analyses of TCRs have shown that the variable domains pair via a conserved hydrophobic core, while the constant domains pair via a highly polar interface, with a skewed distribution of acidic residues in Cα and basic residues in Cβ. In addition, there is a conserved disulfide bond between Cα and Cβ near the membrane (Figure 2). Each C region contains a connecting peptide and a transmembrane domain consisting of several conserved positively charged residues that mediate interaction with the CD3 complex. A very short cytoplasmic tail (5–12 amino acids) is also present at the C-terminal end of both constant domains.

TCRs recognize antigenic peptides in the context of the MHC complex. These complexes are bound to the membrane of antigen-presenting cells, and consist of an anti-parallel β-sheet with α-helices lining the sides, forming a peptide-binding groove (Figure 2). The specificity of a TCR for a particular peptide-MHC complex is determined by the CDRs of the Vα and Vβ domains. The germ-line-encoded CDR2 regions interact primarily with the MHC α-helices, while the highly variable CDR3s are nearly always centered over the antigenic peptide. The CDR1 can interact with the peptide, MHC, or both, depending on the specific interaction. These interactions place the TCR at a conserved diagonal orientation, with the Vα domain interacting with the C-terminus of the peptide and the Vβ domain interaction with the N-terminus, although the angle of this interaction is variable. General hydrophobic interactions between CDR2 residues and the MHC helices are often conserved, but there appear to be no absolute requirements for interactions between specific residues that can account for MHC-restricted binding of TCRs. In addition, there are no significant differences in the binding orientation of MHC class I- and class II-restricted TCRs.

Comparison of TCR-pMHC complex structures with those of unligated TCRs have shown that the CDR loops are flexible, suggesting that conformational adjustment is required during pMHC binding. Most CDR loop adjustments occur in the CDR3s and are in the range of 3–6 Å, although 11 Å is the largest shift observed. Smaller adjustments are often observed in the CDR1s, particularly the CDR1α. The effects of CDR flexibility can account for the typically low affinity ($K_D \sim 1\text{--}100 \mu\text{M}$) of TCRs for their pMHC ligands, which is often the result of a slow association rate. In addition, TCR-pMHC binding is often an entropically unfavorable process, suggesting that the CDRs may become more rigid upon pMHC binding.

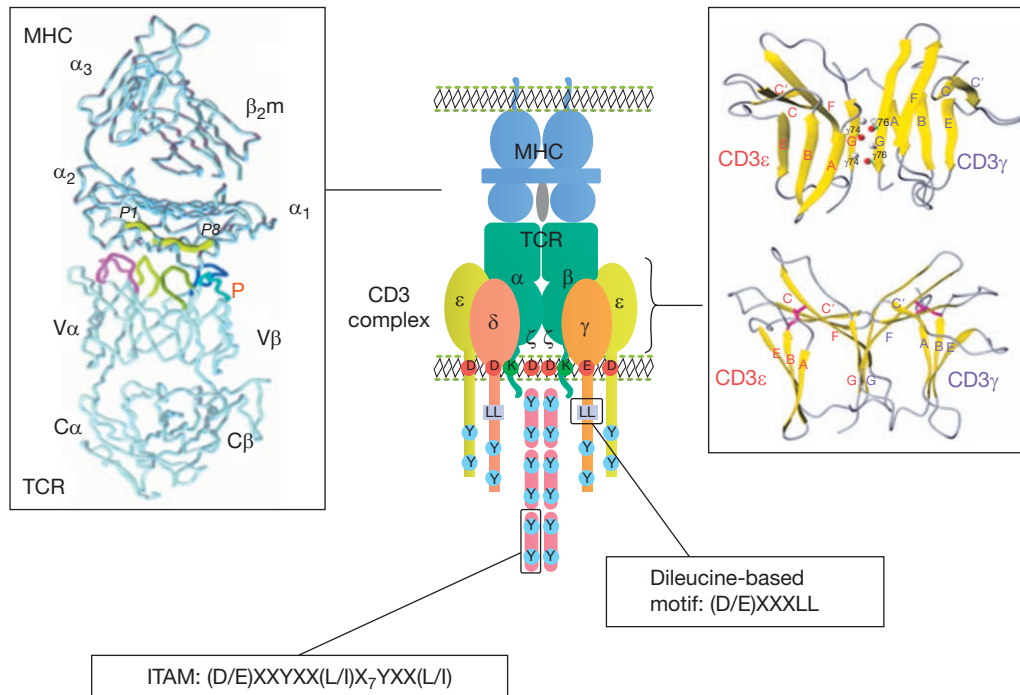


Figure 2 TCR:CD3 complex and its interaction with peptide-MHC. Schematic of the presumed stoichiometry of the TCR:CD3 complex. Highlighted are charged transmembrane residues, the ITAM sequence, and the dileucine-based motif. In the upper left panel is a backbone structure of TCR: peptide-MHC complex. TCR is at the bottom and the MHC is on top. The peptide (P1–8) is shown as a large tube in yellow. V α CDRs 1 and 2 are in light and dark purple, respectively; α HV4 in white; V β CDRs 1 and 2, in light and dark blue, respectively; β HV4 in orange; and V α and V β CDR3s in light and dark yellow, respectively. Reprinted from Garcia KC, Degano M, Stanfield RL, et al. (1996) An $\alpha\beta$ T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* 274: 209–219. In the upper right panel is a ribbon diagram of CD3 γ . The beta strands are in yellow and the text is colored red for CD3 ϵ and blue for CD3 γ . Three pairs of atoms involved in the hydrogen bond formation are designated with amide protons in gray and carbonyl oxygen atoms in red. Two disulfide-linked cysteine residues are shown. Reproduced from Sun ZJ, Kim KS, Wagner G, and Reinherz EL (2001) Mechanisms contributing to T cell receptor signaling and assembly revealed by the solution structure of an ectodomain fragment of the CD3 γ heterodimer. *Cell* 105: 913–923. In the lower left and right panels are the ITAM and dileucine-based consensus motifs, respectively.

Invariant Chains: CD3, CD247, and Pre-T α

CD3 and CD247

The TCR is associated with the invariant chains of the CD3 complex: CD3 ϵ , CD3 γ , and CD3 δ are members of the Ig superfamily with one extracellular Ig domain followed by a transmembrane domain and cytoplasmic tail of ~40 amino acids in length. The zeta family of molecules is unique and has a short extracellular sequence and long cytoplasmic tail. This family consists of CD247 (otherwise known as 'CD3 ζ ' or 'TCR ζ '), its splice variant η , and the γ -chain of the Fc receptor (Fc γ). The CD247 ζ - and η -chains are alternatively spliced gene products and are identical except for the carboxyl-terminal region of the cytoplasmic tail (113 and 155 amino acids long for ζ and η , respectively).

Analysis of CD3 and CD247 knockout mice demonstrates their importance in TCR expression and T-cell development and function. Mice lacking CD247 (CD3 ζ) and CD3 γ exhibit a substantial block in early stages of T-cell development, whereas mice lacking CD3 δ develop a block at a later stage. CD3 ϵ knockout mice display a complete arrest in T-cell development, likely due to its requirement to form heterodimers with CD3 γ and CD3 δ .

The CD3 and CD247 chains contain a number of amino acid residues and motifs that are important for assembly, signal

transduction, and regulation of cell surface expression (Figure 2). The transmembrane regions contain negatively charged amino acids that interact with the positively charged residues of TCR $\alpha\beta$ (CD3 δ , CD3 ϵ , and CD3 ζ -Asp; CD3 γ -Glu). The cytoplasmic tails contain immuno-receptor tyrosine-based activation motifs (ITAMs) that when phosphorylated provide docking sites for SH2 domain-containing proteins, which are important for downstream signal transduction. Each CD3 $\epsilon\gamma\delta$ chain contains one ITAM while CD247 (CD3 ζ) contains three. CD3 δ and CD3 γ also contain a dileucine-based motif. Both the YXXL sequence in the ITAM and the dileucine sequence have been shown to play a role in internalization and downmodulation of many types of receptors from the cell surface. It has been suggested that these motifs are also utilized for TCR transport. A number of studies using T-cell lines *in vitro* have shown a role for the CD3 γ dileucine-based motif in TCR downmodulation; however, its role *in vivo* remains to be defined.

Pre-T α

During T-cell development in the thymus, TCR $\alpha\beta$ precursors express a pre-TCR. The pre-TCR is formed by the functionally rearranged TCR β chain, the CD3 complex, and a surrogate TCR α chain, pre-TCR α (pT α). There are two notable differences between pT α and TCR α . First, pT α only has one invariant

extracellular Ig domain. Second, the cytoplasmic tail of pT α is longer and contains two potential serine and threonine phosphorylation sites and a potential SH3 domain-binding motif. The requirements for surface expression of pT α are still unclear. However, it is clear that a number of signaling events are initiated through the pre-TCR. Pre-TCR signaling confirms the successful rearrangement of TCR β and induces the suppression of further TCR β locus rearrangement, a process called 'allelic exclusion'. Rearrangement of the TCR α locus then occurs leading to progression of T-cell development.

TCR:CD3 Complex

Expression of the TCR on the cell surface requires all six chains of the complex. It is generally accepted that the complex consists of four dimers: TCR $\alpha\beta$ or $\gamma\delta$, heterodimers of CD3 $\epsilon\gamma$ and CD3 $\epsilon\delta$, and a zeta family dimer. The majority of TCR:CD3 complexes (80–90%) contain a CD247 ζ homodimer. Heterodimers of ζ - η or ζ -FcR γ or homodimers of FcR γ have been observed in a small percentage of TCR complexes. An organizing principle has been proposed in which a single negatively charged amino acid in the TCR dimer interacts with two positively charged residues in the CD3 $\epsilon\gamma$, CD3 $\epsilon\delta$, and CD247 ζ dimers. In this model, the TCR α lysine interacts with the CD3 $\epsilon\delta$ dimer, the TCR β lysine interacts with the CD3 $\epsilon\gamma$ dimer and the TCR α arginine interacts with the CD247 ζ dimer.

Role of the TCR in T Cell Biology and Signaling

T-Cell Development

T cells that exit the thymus have been through a selection process based largely on TCR affinity. T cells with TCR that recognize self-peptide–MHC too strongly are negatively selected and deleted. T cells that have too low an affinity to productively interact with self-MHC are ignored and ultimately die of neglect. However, T cells that display a moderate affinity for the self-MHC molecules are positively selected and allowed to exit into the periphery.

Cell Biology

It is known that each of the six protein chains is required for correct assembly and surface expression of the TCR:CD3 complex. Once expressed on the cell surface, the TCR:CD3 complex is constitutively internalized and recycled back to the cell surface via the endosomal network. Once ligated by peptide–MHC complexes, the TCR is downmodulated from the cell surface and diverted to lysosomes for degradation.

Assembly and surface expression of the TCR:CD3 complex

Assembly of the complex is a highly ordered process that takes place in the endoplasmic reticulum (ER). Most of the evidence for assembly order supports a model in which CD3 $\epsilon\gamma$ and CD3 $\epsilon\delta$ heterodimerize followed by sequential addition of TCR α and TCR β chains and the ζ_2 homodimer (or heterodimer). Once the correct stoichiometry is achieved, the intact complex is transported from the ER. It has been shown that the ζ -chain can exit the ER independently of the rest of the

receptor complex and remain in the Golgi complex. CD3 γ and CD3 ϵ have been reported to partly exist as heterodimers in association with calnexin while the TCR α and β -chains have been found to associate with calreticulin. It has been suggested that these molecules may serve as chaperones to prevent the transport of partial complexes. A number of residues are present within the chains that serve as ER retention signals or to target partial complexes to lysosomes for degradation, thereby ensuring that only complete TCR:CD3 complexes are expressed on the cell surface. The charged residues in the transmembrane regions increase their susceptibility to ER degradation. A number of additional residues have been described as possible ER retention or degradation signals; however, a detailed mechanism of complex assembly, including specific interactions between individual chains, remains to be defined.

Internalization and downmodulation of the TCR:CD3 complex

The TCR:CD3 complex is constitutively internalized and recycled back to the cell surface. Although the exact mechanism of this process has not yet been defined, it is likely to involve the interaction of the CD3/CD247 molecules with adaptor protein (AP) complexes that are associated with clathrin at the plasma membrane and intracellular recycling vesicles of the endosomal network. The dileucine-based motif in CD3 γ has been shown to be capable of binding to a member of the AP family of complexes, AP-2. In addition to dileucine based motifs, AP-2 can recognize YXXL-based sequences in a number of receptor systems. Indeed, there is evidence that TCR internalization requires both dileucine and tyrosine-based motifs within the CD3 chains.

Upon ligation with peptide–MHC complexes, the TCR:CD3 complex is downmodulated from the cell surface and recycling is prevented. While the exact mechanism of this important process is unknown, it may involve two related E3 ubiquitin ligases, c-Cbl and Cbl-b, as T cells from mice lacking both proteins fail to downmodulate their TCR following ligation. While both internalization and downmodulation are hallmarks of TCR biology, their physiological importance and function remain to be determined.

Signaling through the TCR

Initiation of T-cell activation occurs when the TCR recognizes peptide–MHC complexes. TCR $\alpha\beta$ consists of the ligand-binding unit while the CD3 complex transduces signals into the T cell. Although initiation of TCR triggering remains unclear, comparative structural studies have offered some evidence for a ligand-induced conformational change model. Structural and fluorescence studies have shown that binding of TCR to an agonist pMHC ligand induces a conformational change within the A–B loop of the TCR C α domain that is reversible upon disengagement of the TCR. Movement of this loop could affect the orientation of CD3 subunits relative to the TCR and initiate signaling. NMR and FRET studies have shown that the cytoplasmic domain of CD3 ϵ is inserted into the plasma membrane in live, nonactivated T cells. Thus, TCR triggering must result in dissociation of the ITAM from the membrane to render the tyrosine residues accessible to kinases. Complete structures of the TCR–CD3 complex would be useful to confirm these models.

Clustering of TCR:peptide-MHC complexes brings in the coreceptors CD8 or CD4 which bind to MHC class I and II molecules, respectively. Both coreceptors are associated with Src-related protein tyrosine kinase (PTK) $p56^{lck}$, while another PTK, $p59^{fyn}$, interacts with the CD3 complex. In resting T cells, these kinases are inactive due to the interaction of the C-terminal phosphotyrosine residues binding to the N-terminal SH2 domain. This intramolecular interaction prevents substrate access to the kinase (SH1) domain. T cell:APC interaction induces the removal of these inhibitory phosphates by the transmembrane phosphatase CD45. Cross-phosphorylation of active site tyrosine residues further potentiates $p56^{lck}$ and $p59^{fyn}$ kinase activity and results in the phosphorylation of the ITAM tyrosine residues in the CD3/CD247 cytoplasmic tails. Phosphorylation of both ITAM tyrosine residues is required for docking of a specialized PTK, zeta-associated protein-70 (ZAP-70), which has two tandem SH2 domains. ZAP-70 kinase activity is further potentiated through phosphorylation by $p56^{lck}$ and $p59^{fyn}$.

Activated ZAP-70 initiates a number of signaling pathways. A key target of ZAP-70 is the raft-resident linker for activated T cells (LAT) which is heavily phosphorylated and recruits a wide variety of downstream signaling molecules. Phosphorylation of phospholipase $C\gamma_1$ ($PLC\gamma_1$) leads to the production of potent second messengers, diacylglycerol (DAG) and inositol triphosphate (IP_3), whose actions lead to protein kinase C (PKC) activation and NF- κ B nuclear translocation as well as Ca^{2+} release and nuclear factor of activated T cells (NFAT) translocation. These events lead to the transcription of genes required for T-cell proliferation and interleukin-2 (IL-2) production. Signaling through ZAP-70 also initiates activation of the Ras pathway and the MAPK signaling cascade which also results in upregulation of genes required for proliferation. A wide array of additional signaling molecules and APs have been shown to contribute to the signaling cascade initiated following TCR ligation and have been reviewed extensively elsewhere.

Advanced microscopic analyses confirmed the biochemical data highlighted above, and have demonstrated that the initiation and regulation of TCR-induced signaling is facilitated by the immunological synapse (IS); a macromolecular structure comprised of TCR:pMHC, CD4/CD8 co-receptors, the co-stimulatory molecule CD28 and its ligands, and adhesion molecules, such as ICAM-1. TCR-induced signaling is initiated in TCR:pMHC microclusters, wherein a single TCR:pMHC complex is sufficient to activate a signaling cascade leading to an increase in intracellular Ca^{2+} levels. Microclusters are subsequently transported laterally toward a central supramolecular activation cluster (cSMAC) which is surrounded by a ring of adhesion molecules, the peripheral supramolecular activation cluster (pSMAC), which are necessary for providing the initial stop signal upon antigenic encounter and which contribute to cytoskeletal remodeling necessary for T cell:APC interactions (Figure 3). Early observations of this prototypical IS comprised of cSMAC and pSMAC regions suggested that it provided the macromolecular organization needed to amplify and sustain TCR:pMHC signals, as evidenced by the accumulation of PKC θ within the cSMAC. However, prevention of microcluster signaling following cSMAC formation has suggested that TCR:pMHC microclusters are responsible for TCR-induced

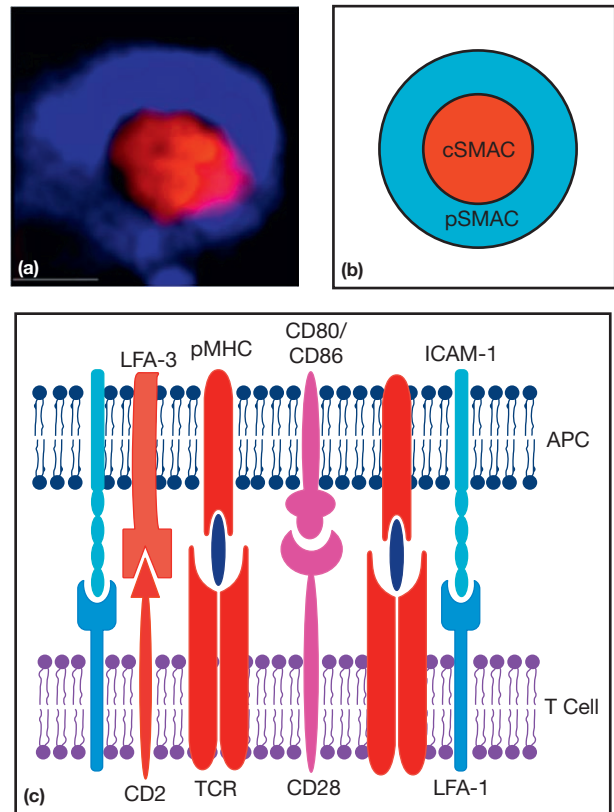


Figure 3 The morphology and molecular composition of the immunological synapse. (a) MHC-specific $CD4^+$ T cells can be stimulated using synthetic lipid bilayers containing MHC molecules, loaded with fluorescently labeled MHC peptide (pMHC; red), and the adhesion molecule ICAM-1 (blue). Schematic representations (b) and (c) of the immunological synapse include its prototypical bullseye comprised of distinct regions including the cSMAC and pSMAC, which are recognized as distinct areas of signal initiation, amplification, and termination. Reproduced from Sun ZJ, Kim KS, Wagner G, and Reinherz EL (2001) Mechanisms contributing to T cell receptor signaling and assembly revealed by the solution structure of an ectodomain fragment of the $CD3\gamma$ heterodimer. *Cell* 105: 913–923, with permission from Elsevier.

signaling, while the role of the cSMAC is to regulate the strength of TCR signal in part by facilitating TCR internalization and degradation. Indeed, in the mature IS tyrosine phosphorylated signaling proteins, such as ZAP70, LAT and SLP76, are restricted to the pSMAC. Thus, activation of the TCR directs a cell signaling paradigm comprised of distinct biophysical and biochemical events which are tightly coordinated for optimal propagation of signal and T-cell activation. However, it is currently unclear whether there are functional redundancies between CD3 subunits of the TCR:CD3 complex, or if particular CD3 chains or ITAMs are required for mediating distinct aspects of TCR-induced signaling.

See also: **Signaling:** Immunoglobulin (Fc) Receptors; Mitogen-Activated Protein Kinase Family; Protein Kinase C Family; Src Family of Protein Tyrosine Kinases.

Further Reading

- Allison TJ and Garboczi DN (2001) Structure of $\gamma\delta$ T cell receptors and their recognition of non-peptide antigens. *Molecular Immunology* 38: 1051–1061.
- Call ME, Pyrdol J, Wiedmann M, and Wucherpfennig KW (2002) The organizing principle in the formation of the T cell receptor-CD3 complex. *Cell* 111: 967–979.
- Davis MM (1990) T cell receptor gene diversity and selection. *Annual Review of Biochemistry* 59: 475–496.
- Fooksman DR, Vardhana S, Vasiliver-Shamis G, et al. (2010) Functional anatomy of T cell activation and synapse formation. *Annual Review of Immunology* 28: 1–27.
- Garcia KC, Degano M, Pease LR, et al. (1998) Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. *Science* 279: 1166–1172.
- Garcia KC, Degano M, Stanfield RL, et al. (1996) An $\alpha\beta$ T cell receptor structure and its orientation in the TCR-MHC complex. *Science* 274: 209–219.
- Germain RN and Stefanova I (1999) The dynamics of T cell receptor signaling: Complex orchestration and the key roles of tempo and cooperation. *Annual Review of Immunology* 17: 467–522.
- Glusman G, Rowen L, Lee I, et al. (2001) Comparative genomic of the human and mouse T cell receptor loci. *Immunity* 15: 337–349.
- Goldsby RA, Kindt TJ, and Osborne BA (eds.) (2000) *Kuby Immunology*. New York: W H Freeman.
- Irvine DJ, Purbhoo M, Krogsgaard M, and Davis MM (2002) Direct observation of ligand recognition by T cells. *Nature* 419: 845–849.
- Kruisbeek AM, Haks MC, and Carleton M (2000) Branching out to gain control: How the pre-TCR is linked to multiple functions. *Immunology Today* 21: 637–644.
- Rudolph MG, Stanfield RL, and Wilson IA (2006) How TCRs bind MHCs, peptides and co-receptors. *Annual Review of Immunology* 24: 419–466.
- Samelson LE, Harford JB, and Klausner RD (1985) Identification of the components of the murine T cell antigen receptor complex. *Cell* 43: 223–231.
- Sun ZJ, Kim KS, Wagner G, and Reinherz EL (2001) Mechanisms contributing to T cell receptor signaling and assembly revealed by the solution structure of an ectodomain fragment of the CD3 γ heterodimer. *Cell* 105: 913–923.
- Valitutti S, Muller S, Cella M, Padovan E, and Lanzavecchia A (1995) Serial triggering of many T-cell receptors by a few peptide-MHC complexes. *Nature* 375: 148–151.

T-Cell Receptor Signaling to NF- κ B

R R McCully and J L Pomerantz, Johns Hopkins University School of Medicine, Baltimore, MD, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Adaptive immune response Also known as acquired immunity, it refers to antigen-specific defense mechanisms that take several days to become protective and are designed to react with and remove a specific antigen.

Antigen Any substance capable of eliciting an adaptive immune response is referred to as an antigen.

Autoimmunity The failure of an organism to recognize its own constituent parts as self, which allows an immune response against its own cells and tissues.

Costimulatory receptor A secondary receptor on the surface of a lymphocyte that can influence the development of an effective immune response in addition to the antigen-specific signal from antigen receptors.

Immunological synapse The cell–cell interface between an antigen-presenting cell and a T lymphocyte.

Inflammation A type of nonspecific immune response that serves to recruit other immune components to the site of injury or infection, and often involves redness, warmth, swelling, and pain.

Lymphoma A cancer that originates in a lymphocyte and presents as a solid tumor of lymphoid cells.

Transcription factor A protein that binds to specific DNA sequences to control the transcription of genetic information from DNA into mRNA.

Ubiquitin A 76-amino-acid regulatory protein that can be conjugated to other proteins by specific enzymatic machinery and can mark proteins for destruction or serve as an activating signal.

The Antigen Receptor and the Immunological Synapse

The generation of the adaptive immune response begins with the recognition of the antigen by specialized receptors on T and B lymphocytes. The T-cell receptor (TCR) recognizes peptide antigen displayed by specialized surveillance cells of the immune system known as antigen presenting cells (APCs), such as B cells, macrophages, and dendritic cells. APCs process antigens and present them on the major histocompatibility complex (MHC) for recognition by the TCR. A costimulatory receptor on T cells, known as CD28, is also required for nuclear factor kappa B (NF- κ B) activation. CD28 recognizes B7 molecules on APCs in an important step for the activation of naive T cells. If the TCR binds the peptide/MHC complex in the absence of the costimulatory interaction between CD28 and B7, the cells will fail to fully activate and become anergic and can no longer mount an effective immune response (Figure 1).

The association of many proximal signaling molecules with the TCR occurs at the site of contact between the T cell and APC, known as the immunological synapse. Within the synapse, supramolecular activation clusters (SMACs) provide a spatial and temporal framework for the recruitment and activation of pathway components. There are three subregions of the SMAC that form concentric rings, each defined by the localization of specific molecules. The innermost region is called the central SMAC (cSMAC), and is characterized by the presence of the TCR and several kinases. The cSMAC is surrounded by the peripheral SMAC (pSMAC), which is characterized by the presence of the integrin adhesion molecule LFA-1. The outermost region is the distal SMAC (dSMAC), and contains the tyrosine phosphatase CD45, which is thought to be excluded from the inner regions of the SMAC in order to maintain optimal activation of the receptor-proximal components. The formation of the immunological synapse is

important for the sustained activation of T cells, and may also be important for downstream cellular responses, such as the directional secretion of pro-inflammatory cytokines.

Receptor-Proximal Signaling

Upon antigen binding by the receptor, tyrosine residues within the immunoreceptor tyrosine-based activation motif (ITAMs) on the cytoplasmic tails of various receptor components become phosphorylated. CD3 γ , CD3 δ , and CD3 ϵ each has a single ITAM, and CD3 ζ has 3 ITAMs, each of which has two tyrosine residues that become rapidly phosphorylated by several members of the Src family of tyrosine kinases, which include Lck and Fyn. Lck and Fyn associate with CD45, which removes phosphates from inhibitory tyrosine residues to mediate full activation of these kinases. ITAM tyrosine phosphorylation leads to the recruitment and activation of ZAP-70, a member of the Syk family of tyrosine kinases, via its SH2 domain. ZAP-70 can phosphorylate and activate the TEC family of tyrosine kinases, including ITK and RLK, as well as the transmembrane adaptor protein LAT (linker of activated T cells). LAT tyrosine phosphorylation leads to the recruitment of multiple downstream SH2-containing proteins, including Grb2, PI3K, GADS, and PLC γ 1. A key molecule activated by the costimulatory receptor is the guanine nucleotide-exchange factor Vav, which is tyrosine phosphorylated and membrane localized after its interaction with the adapter molecule Grb2 downstream of CD28 engagement. Vav is also activated, albeit weakly, downstream of TCR engagement. It has been proposed that Vav acts to integrate TCR and CD28 signaling, thereby robustly activating multiple downstream pathways. Costimulatory signaling is also required for the activation of PI3K, which is recruited to and activated by CD28.

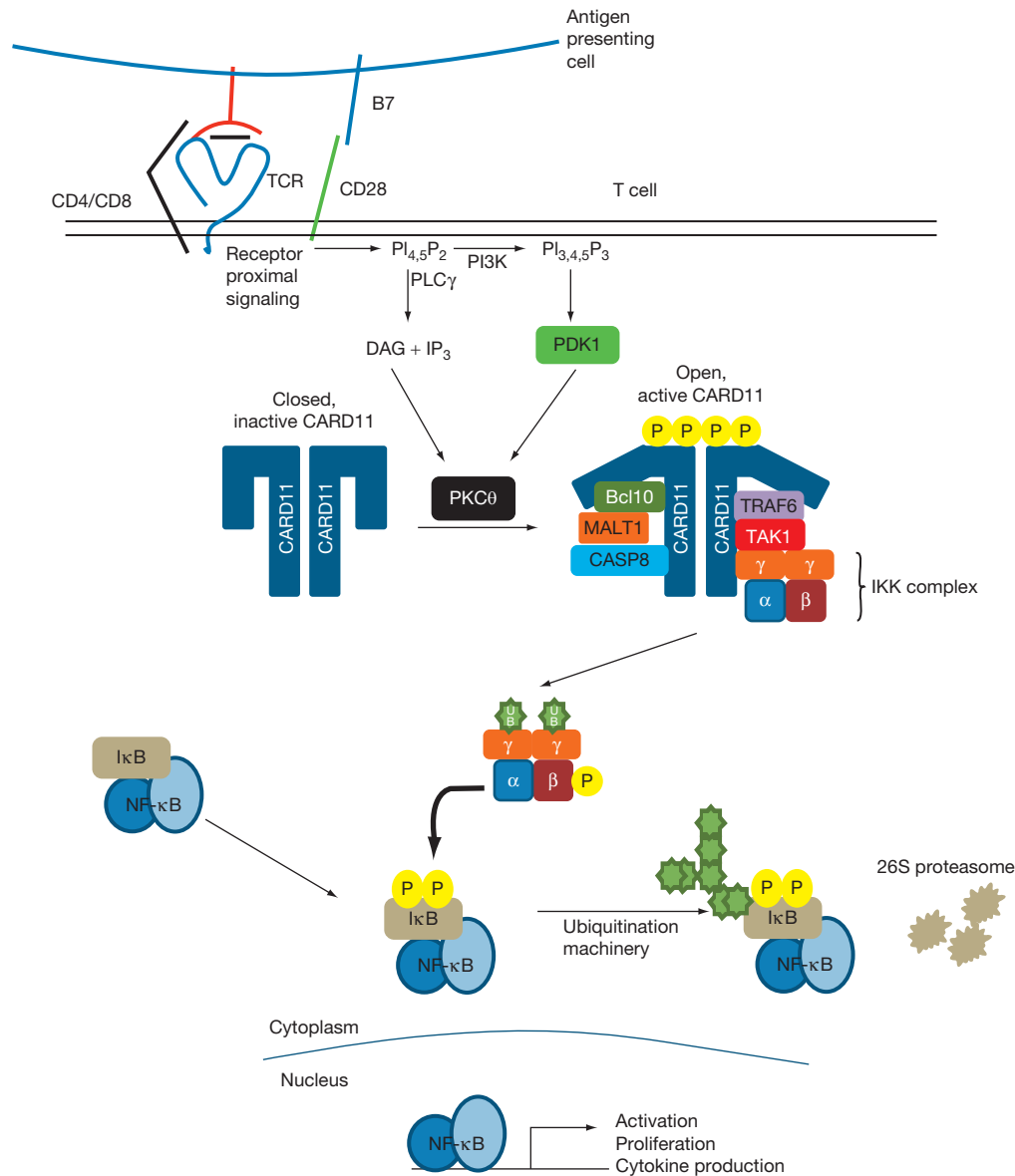


Figure 1 T-cell receptor signaling to NF- κ B. Please see text for details.

Second Messengers

The active antigen receptor signaling complex next activates two important signaling proteins, PLC γ and PI3K, which generate important second messengers. PLC γ and PI3K both act upon the membrane lipid $PI(4,5)P_2$ in the generation of second messengers. PLC γ hydrolyzes $PI(4,5)P_2$ into IP $_3$ and DAG, and PI3K phosphorylates $PI(4,5)P_2$ into $PI(3,4,5)P_3$. The IP $_3$ produced can then bind receptors on the endoplasmic reticulum, triggering the release of another key second messenger, Ca $^{2+}$. These second messengers, in turn, can activate multiple downstream pathways, including the Ras/ERK pathway, the PI3K/Akt pathway, and pathways leading to the activation of multiple transcription factors, including NFAT, AP-1, and NF- κ B. For the sake of simplicity, this article focuses on just the signaling pathway that leads to the activation of NF- κ B.

Downstream Activators

The generation of the second messenger DAG is a key upstream event in the activation of the NF- κ B pathway. DAG binds to and activates protein kinase C (PKC) family members, including PKC- θ in T cells. PKCs also require a threonine phosphorylation on their activation loops for full activity. The kinase phosphoinositide-dependent kinase 1 (PDK1) phosphorylates all known PKCs, including PKC- θ , and PDK1 itself is activated by PI3K. In the antigen receptor pathway, a critical phosphorylation target of PKC- θ is the scaffold protein CARD11/CARMA1 (caspase-recruitment domain containing protein 11/caspase recruitment domain (CARD) membrane-associated guanylate kinase (MAGUK) protein 1). A small fraction of CARD11 is localized, in an SH3-domain-dependent manner, at the plasma membrane prior to antigen receptor activation.

Upon TCR engagement, a significant fraction of CARD11 is recruited to the plasma membrane and is localized near the site of receptor activation. A CARD11 mutant in which a residue in the SH3 domain, L808, is mutated to proline does not localize to the plasma membrane and is unable to mediate antigen receptor to NF- κ B signaling, indicating that localization to the membrane is a critical step in the pathway. It is not clear how CARD11 is recruited to the plasma membrane, although ADAP1 (adhesion- and degranulation-promoting adapter protein) and PDK1 have been shown to contribute to CARD11's recruitment to lipid rafts, which are semirigid cholesterol- and sphingosine-rich membrane subdomains where many antigen receptor components and proximal effectors have been found to concentrate. The recruitment to the plasma membrane of several pathway cofactors is dependent upon CARD11, including Bcl10 (B-cell CLL lymphoma 10) and the IKK complex, though it is not clear that plasma membrane recruitment of these cofactors is required for their activity. CARD11 is phosphorylated in a signal-dependent manner at multiple serine residues, including S564, S577, and S654 within its inhibitory domain (ID). The ID has been shown to inhibit CARD11 activity by binding to other signaling domains of the protein, likely blocking their ability to bind downstream cofactors by steric occlusion. Phosphorylation of the CARD11 ID by PKC- θ , and likely other protein kinases, leads to a derepression of its inhibitory activity, allowing multiple signaling cofactors to bind. These cofactors include Bcl10, mucosa-associated lymphoid tissue 1 (MALT1), TNF receptor associated-factor 6 (TRAF6), Caspase-8, TGF activated kinase 1 (TAK1), and the I κ B kinase (IKK) complex. It is thought that the formation of this multi-protein complex is a critical step in the activation of the IKK complex.

IKK Complex Activation

The precise mechanism of IKK activation is not yet fully understood, though several studies have contributed key insights into the potential mechanism. The IKK complex is composed of three subunits: the catalytically active IKK α and IKK β subunits, and the regulatory IKK γ /NEMO (NF- κ B essential modulator) subunit. The formation of the CARD11-containing signaling complex leads to the recruitment of TRAF6 or TRAF2, which are E3 ubiquitin ligases that are thought to be activated upon binding to MALT1. It seems likely that K63-linked polyubiquitin chains play an important role in the activation of the IKK complex, but there is disparate evidence for the way in which these chains are responsible for activation. One study suggests that TRAF6 adds K63-linked polyubiquitin chains to MALT1, and ubiquitinated MALT1 then acts as a docking site for NEMO. A second study suggests that TRAF6 binding to MALT1 causes MALT1 to act as an E3 ubiquitin ligase, inducing the K63-linked polyubiquitination of NEMO itself. NEMO binding to K63-ubiquitinated Bcl10 is also a critical step in the activation of the IKK complex. The E3 ubiquitin ligase responsible for K63-linked Bcl10 ubiquitination has not yet been determined, though the ubiquitination seems to be dependent upon MALT1.

In addition to K63-ubiquitin chain activation, phosphorylation is also important for activity of the IKK complex.

The phosphorylation of the IKK β subunit occurs in a TAK1-dependent manner. IKK β is phosphorylated on serines 177 and 181 in a required step for NF- κ B activation downstream of antigen receptor signaling and other canonical NF- κ B activating pathways. TAK1 has been shown to associate with and phosphorylate IKK β *in vitro*, and chemical inhibition or genetic deletion of TAK1 in cell lines blocks IKK β phosphorylation and NF- κ B activation. Cell-type specific knockout of TAK1 in mice has only partially confirmed the role of TAK1 in this pathway. Deletion of TAK1 in T cells impairs development and NF- κ B activation.

In addition to its adapter and ubiquitin-ligase functions, MALT1's proteolytic activity has been shown to be important for optimal NF- κ B activation. Mutations at MALT1's catalytic site greatly reduce its ability to activate NF- κ B when overexpressed, and chemical inhibition of MALT1's proteolytic activity inhibits TCR-induced NF- κ B activation. Its proteolytic targets include Bcl10 and the negative regulator A20. The cleavage of five amino acids from the C-terminus of Bcl10 by MALT1 is not required for NF- κ B activation, but rather for integrin-mediated T-cell adhesion. The second MALT1 proteolytic target identified is A20, a known negative regulator of other NF- κ B-activating pathways. It is possible that the cleavage of A20 downregulates its inhibitory activity, leading to enhanced NF- κ B activation. A20 has been shown to have both ubiquitin-adding and -removing abilities, and can add both K48- and K63-linked polyubiquitin chains to target proteins. A20 has been shown to promote the removal of K63-linked ubiquitin chains from MALT1 in the TCR pathway, which likely attenuates the interaction between MALT1 and the IKK complex, leading to a downregulation of sustained IKK activity. A20 itself is degraded in a proteasome-dependent manner in an early step of pathway activation, likely in order to allow efficient IKK activation by MALT1. A20 may promote the deubiquitination of multiple pathway cofactors, as TRAF6 and RIP1 are targets in other pathways, and A20 has been shown to bind to IKK γ /NEMO.

In addition to the proteolytic activity of MALT1, the proteolytic activity of Caspase-8 is thought to be required for IKK activation. Caspase-8 has been shown to inducibly associate with CARD11 and TRAF6 upon antigen receptor cross-linking, though its proteolytic target has not yet been identified. Using fluorescent inhibitors, it has been shown that the catalytic activity of Caspase-8 in this pathway is largely restricted to the plasma membrane and lipid rafts, where many pathway components concentrate during signaling.

Interestingly, IKK activity is also important for the downregulation of NF- κ B activity. IKK β phosphorylates Bcl10 in response to TCR signaling, and this phosphorylation leads to the ubiquitination and degradation of Bcl10, thereby reducing the cellular concentration of this required pathway component. NEDD4, Itch, and cIAP2 mediate Bcl10's ubiquitination and subsequent degradation. It is not yet clear how ubiquitinated Bcl10 is degraded, as there is conflicting evidence suggesting that it may be degraded by either the lysosome or 26S proteasome. The role of Bcl10's degradation is likely distinct from the MALT1-mediated proteolytic cleavage of Bcl10, which is a proactivation signal, as discussed above. The protein levels of other pathway cofactors are also downregulated as a means to prevent

sustained pathway activation. The ubiquitin-mediated degradation of ZAP70, Lck, PLC γ 1, PKC θ , and CARD11 has been shown to be involved in turning off TCR signaling.

NF- κ B Activation

The role of the activated IKK complex is to phosphorylate I κ Bs (inhibitors of NF- κ B), which retain NF- κ B in the cytosol to prevent gene transcription by masking nuclear localization signals (NLSs) on NF- κ B. There are nine members of the I κ B family, which can be subdivided into three groups: classical, unusual, and NF- κ B precursors. The classical group, which is the best characterized, is responsible for NF- κ B inhibition in the antigen receptor pathway, and consists of I κ B α , I κ B β , I κ B γ , and I κ B ϵ . The unusual I κ Bs are I κ B ζ , I κ BNS, and Bcl-3. The unprocessed p105 and p100 precursors of NF- κ B subunits p50 and p52, respectively, can act as I κ Bs under certain circumstances. I κ Bs contain five to seven ankyrin repeats that bind to NF- κ B. The IKK β -mediated phosphorylation of I κ Bs at two N-terminal serines signals for their K48-linked polyubiquitination by SCF-family ubiquitin ligase members β TrcP1 and β TrcP2. This polyubiquitination leads to the targeting of I κ Bs to the 26S proteasome for proteolytic degradation. The degradation of these inhibitors leads to the unmasking of NLSs on NF- κ B, allowing NF- κ B to translocate to the nucleus, where it can bind to DNA enhancer elements and activate the transcription of many genes, including proinflammatory cytokines, cytotoxic and anti-apoptotic molecules, and proliferative genes. NF- κ B also activates the transcription of the NFKBIA gene, whose product is I κ B α , thereby forming a negative feedback loop that can turn off NF- κ B activity by binding NF- κ B subunits in the nucleus and relocalizing them to the cytosol.

The NF- κ B family of transcription factors consists of five members, each containing a conserved Rel homology domain (RHD) at its N-terminus: RelA/p65, RelB, c-Rel, p105/p50, and p100/p52. The RHD mediates dimerization, interaction with I κ Bs, and sequence-specific binding to DNA. RelA, RelB, and c-Rel contain a transactivation domain at their C-termini that allow them to activate gene transcription, whereas p50 and p52 lack the ability to activate gene transcription on their own. NF- κ Bs form homo- or hetero-dimers that are held in an inactive state in the cytoplasm by I κ Bs, and can only activate gene transcription when I κ Bs have been degraded.

Pathway Dysregulation: Cancer and Autoimmunity/Inflammation

The mutation or misregulation of several antigen receptor signaling pathway components has been implicated in the

development of several types of lymphomas. Chromosomal translocations that lead to the overexpression of MALT1 and Bcl10 have been linked to MALT1 lymphoma. A chromosomal translocation resulting in the fusion of API2 and MALT1, whose product is known as API2-MALT1, has also been found to be a robust activator of NF- κ B and is associated with the development of MALT lymphomas. At least three genes in this pathway have been shown to be required for the survival of the activated B-cell type of diffuse large B-cell lymphomas (ABC DLBCLs), which is dependent upon constitutive NF- κ B activity for their survival. Activating mutations in CARD11 and the CD79a and CD79b subunits of the B-cell receptors have been found in ABC DLBCL, though it remains to be shown how these mutations lead to the development of the transformed phenotype. In addition, loss-of-function mutations in the deubiquitinating enzyme A20 have been found in a significant fraction ABC DLBCL samples in one study.

This pathway is also an important potential therapeutic target for the treatment of autoimmunity and inflammation, as well as transplant anti-rejection drugs. For example, CARD11 has been shown to be critical for the development of allergic airway inflammation in a mouse model of asthma. The mutations of genes involved in the downregulation of NF- κ B signaling, such as the E3 ubiquitin ligase Itch, are also involved in the development of autoimmunity. A mouse in which Itch is mutated develops spontaneous skin inflammation. The processes involving CARD11 are especially interesting targets for the treatment of cancers and immune disorders, because CARD11 expression is cell-type restricted and is pathway specific, whereas most other pathway components are involved in the activation of multiple pathways leading to NF- κ B in many tissue types. Selective targeting of CARD11 may thus lead to fewer side effects than approaches aimed at inhibiting kinases such as PKC family members or IKKs, or other components of the signaling pathway.

See also: **Cell Architecture and Function:** 26S Proteasome: Structure and Function; **Signaling:** B-Cell Antigen Receptor; Nuclear Factor Kappa B; Phospholipase C; Protein Kinase C Family; T-Cell Antigen Receptor.

Further Reading

- Blonska M and Lin X (2009) CARMA1-mediated NF- κ B and JNK activation in lymphocytes. *Immunological Reviews* 228: 199–211.
- Hailfinger S, Rebeaud F, and Thome M (2009) Adapter and enzymatic functions of proteases in T-cell activation. *Immunological Reviews* 232: 334–347.
- Paul WE (2008) *Fundamental Immunology*, 6th edn. Philadelphia, PA: Lippincott Williams and Wilkins.
- Schulze-Luehrmann J and Ghosh S (2006) Antigen-receptor signaling to nuclear factor κ B. *Immunity* 25: 701–715.

Telomeres: Maintenance and Replication

A Bianchi, University of Sussex, Brighton, UK

D Shore, University of Geneva, Geneva, Switzerland

© 2013 Elsevier Inc. All rights reserved.

Glossary

Telomerase A nuclear-encoded ribonucleoprotein complex homologous to viral reverse transcriptases that adds single-stranded telomeric repeats to the 3' ends of duplex DNA using its own RNA template; it is responsible for telomere maintenance in most eukaryotic organisms.

Telomere The DNA–protein complex at the extreme ends of linear eukaryotic chromosomes that orchestrates telomere replication and capping.

Telomere-associated sequence (TAS) The highly variable, often repeated DNA sequences found

immediately internal to the terminal simple telomere repeat sequences.

Telomere capping The ability of the telomere DNA–protein complex to prevent chromosome end degradation or fusion (DNA joining) reactions.

Telomeric repeats Simple, short (usually 5–8, but as long as 25 nucleotides) DNA repeat sequences (T₂AG₃ in vertebrates and minor variants in many other eukaryotes) that constitute the DNA component of telomeres. Synthesized by the telomerase enzyme, with the TG-rich sequence extending toward the 3' DNA end.

All linear eukaryotic chromosomes terminate in a specialized nucleoprotein structure, the telomere. Telomeres perform at least two essential functions: they provide (1) a protective cap on chromosome ends that prevents their degradation or deleterious fusion and (2) a special mechanism for replicating the DNA at chromosome ends. In most organisms, telomeres are comprised of a tandem array of simple DNA repeats to which a large set of protein factors are bound. The telomeric DNA repeats are generated by a specialized reverse transcriptase, called telomerase, which uses an endogenous RNA template. Defects in the maintenance of telomeric DNA, for example, through inactivation of the telomerase enzyme, lead to the progressive loss of telomeric repeats and their bound factors, which eventually causes a catastrophic uncapping of telomeres that results in the fusion of chromosome ends. The telomere complex is thus essential to ensure genome stability.

The Telomeric Complex: A Specialized Nucleoprotein Structure at Chromosome Ends

Pioneering studies in the fruit fly and in maize, carried out respectively by H Muller and B McClintock in the early twentieth century, first identified the telomere as a special genetic entity that protects chromosome ends from degradation and fusion, a property absent in DNA ends resulting from random chromosomal breakage. A wealth of subsequent genetic and biochemical studies has led to the understanding that telomeres are specialized nucleoprotein complexes composed of a large number of protein factors assembled onto telomeric DNA.

Telomeric DNA

In most eukaryotes, the terminal DNA sequences at each chromosome are composed of variable-length arrays of simple tandem repeats, first recognized by Blackburn and Gall in the ciliate *Tetrahymena thermophila*. These telomeric repeats are

typically 5–8 nucleotides long, but can be up to 25 base pairs (bp; **Figure 1**), and usually have a higher G content in the DNA strand that runs with a 5'-to-3' polarity toward the end of the chromosome. In all species, the majority of telomeric repeats appear to be in double-stranded form and their total length varies from a few tens of nucleotides in some species of ciliates (e.g., 28 bp of duplex telomeric repeats in *Euplotes*) to tens of thousands of bp in some strains of laboratory mice. In both budding and fission yeasts, telomeres are about 300 bp in length, whereas in human cells telomeres are generally 6–12 kbp long. With the exception of several ciliate species in which the length of telomeric DNA is fixed, in most species the length of each telomere is variable both from individual to individual and during the life of the organism. In many species, it has been shown that telomeres terminate in a single-stranded overhang of the G-rich strand. This overhang can range from only a few nucleotides long (in some ciliates) to about 100–200 bp in mammals, and is believed to represent a general feature of telomeres. G-rich telomeric overhangs *in vitro* can adopt a variety of intra- and intermolecular paired structures, involving most commonly four strands interacting through Hoogsteen-type G–G base-pairing (G-quartets). Mounting *in vivo* evidence for telomeric G-quartets points to a still unclear role for these structures in telomere replication. In mammals, some ciliates, and trypanosomes, telomeric DNA appears to fold back in a structure, the t-loop, where the single-stranded overhang is tucked into the duplex portion of the telomeric tract, by base pairing with the C-rich strand and thus displacing a short portion of the G-strand.

In most organisms, the short telomeric repeats described above are preceded by less conserved DNA elements, called subtelomeric repeats or telomere-associated sequences (TASs). These elements are of variable lengths but are generally at least a few hundred bp long. In any given species, several classes of these repeats can be present, and elements of each class are associated with a particular chromosome with great variation in number and organization. The assembly of these subtelomeric repeats is highly dynamic and subject to active

Vertebrates	
<i>Homo, Mus, Rattus, Gallus</i>	TTAGGG
Arthropodes	
Insects	
<i>Bombix mori</i>	TTAGG
Nematodes	
<i>Ascaris</i>	TTAGGC
<i>Caenorhabditis elegans</i>	TTAGGC
Plants	
<i>Arabidopsis thaliana</i>	TTTAGGG
Green algae	
<i>Chlamydomonas</i>	TTTTAGGG
Fungi	
Ascomycota	
Hemiascomycetes	
<i>Saccharomyces cerevisiae</i>	T(G) ₁₋₃
<i>Kluyveromyces lactis</i>	ACGGATTGATTAGGTATGTGGTGT
Archeascomycetes	
<i>Schizosaccharomyces pombe</i>	TTACAGG(G) ₀₋₄
Euascomycetes	
<i>Neurospora</i>	TTAGGG
Protists	
Alveolata	
<i>Plasmodium</i>	TT[T/C]AGGG
Ciliates	
<i>Tetrahymena</i>	TTGGGG
<i>Oxytricha</i>	TTTTGGGG
<i>Euplotes</i>	TTTTGGGG
Diplomonadida	
<i>Giardia</i>	TAGGG
Euglenozoa	
<i>Trypanosoma</i>	TTAGGG

Figure 1 Representative list of telomeric repeats in several eukaryotic organisms, including some of the most extensively studied ones with regard to telomere biology.

recombination, resulting in a large variation of subtelomeric regions between strains of yeast or between human individuals. TASs do not appear to carry out an essential telomeric function, as both human and yeast chromosomes that are devoid of them are replicated and segregated properly through both mitosis and meiosis.

Telomerase

DNA polymerases replicate DNA uniquely in a 5'-to-3' direction. As a consequence, as the DNA replication bubble moves along the DNA, the two strands are replicated differently: one strand (the 'leading' strand) is replicated in the same direction as the movement of the replication fork, while the other strand ('lagging' strand) is replicated backward in small installments, each one initiating from an RNA primer laid down by the DNA polymerase 1/primase enzyme (Figure 2(a)). Due to the requirement for an initiating RNA primer, which is later removed, replication of linear DNA molecules by the conventional DNA replication machinery would result in a terminal gap in the lagging strand end at the telomere (Figure 2(a)). The presence of overhangs at telomeres could, in principle, mask this problem at the lagging strand, but a terminal gap would

then occur at the strand replicated by leading strand synthesis after processing of the end to generate the single-stranded overhang (Figure 2(b)). Thus, in the absence of a specialized mechanism to replicate the ends, loss of genetic material is expected to occur at each cell division. The most general solution to this problem in eukaryotes is represented by the specialized replication of telomeric repeats by the telomerase enzyme.

Greider and Blackburn first identified telomerase in the ciliate *Tetrahymena thermophila* based on its ability to add telomeric repeats to a telomeric DNA primer in an *in vitro* reaction. This discovery, together with the demonstration by Blackburn and Szostak that telomeric DNA repeats added onto linear DNA fragments contributed to their stability inside cells, resulted in the award of the 2009 Nobel prize in physiology or medicine to Blackburn, Greider, and Szostak. The telomerase enzyme is in fact a ribonucleoprotein, composed of a protein and an RNA moiety, both of which are required for catalytic function (Figure 2(c)). Isolation of the protein component from the ciliate *Euplotes* by Lingner and Cech in 1997 revealed that telomerase shares extensive homology with reverse transcriptases. Thus, telomerase is a specialized reverse transcriptase that utilizes an endogenous RNA template for the synthesis of telomeric repeats. The enzyme appears to be capable

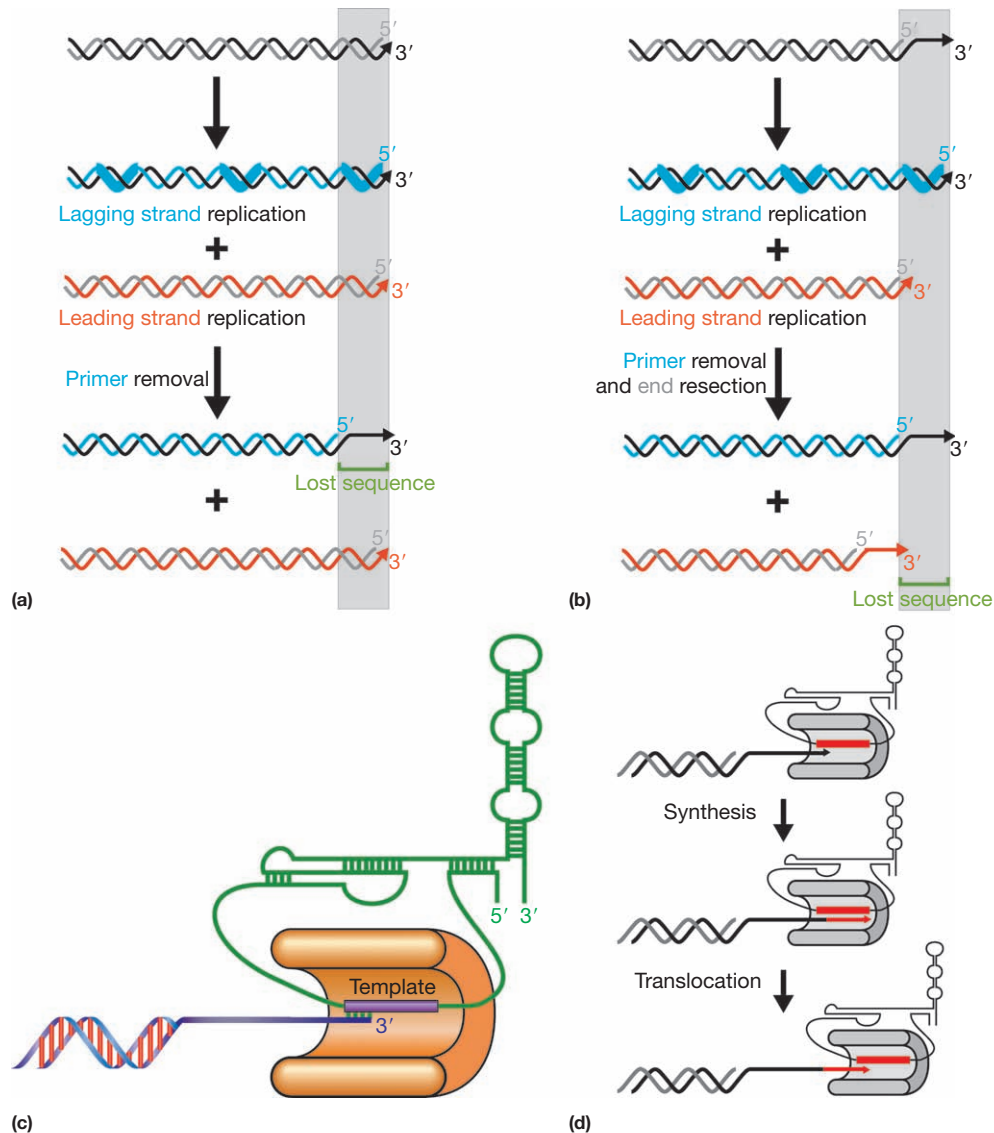


Figure 2 (a) Schematic view of the end replication problem of a linear DNA molecule. The RNA primers laid down by the primase enzyme for the synthesis of the so-called 'lagging' strand are indicated by thicker blue lines. After removal of the primers, replication of a double-stranded DNA molecule would result in a single-stranded terminal gap in the DNA strand replicated by lagging strand synthesis. (b) Schematic view of the end replication problem at a telomere terminating in a single-stranded overhang. Exonucleolytic resection of the ends is hypothesized to generate the overhangs after replication. In this scenario, a terminal gap would be created in the strand replicated by leading strand synthesis. (c) Representation of the structure of the core telomerase enzyme. The protein, based on its homology to known viral reverse transcriptases, is expected to be folded in a structure with a central cleft (palm) bearing the active site and flanked by two domains (thumb and fingers). The template region of the RNA is placed in this cleft, and the secondary structure of the RNA represented here is derived from ciliated protozoa species and appears to be largely conserved in higher eukaryotes. (d) Representation of the processive mechanism of action of telomerase, where a round of synthesis of telomeric repeats is followed by translocation of the enzyme on the DNA and a subsequent new round of synthesis.

of synthesizing several repeats on a particular substrate through repeated steps of elongation and translocation (Figure 2(d)) and, like reverse transcriptases, appears to be a dimer.

Telomere Proteins

Telomeric repeats serve as binding sites for telomeric DNA-binding proteins, which act as a scaffold for the recruitment of additional protein factors to telomeres, including telomerase (Figure 3). The analysis of telomeric proteins in a wide

range of organisms has revealed important similarities, though significant differences have also emerged.

The major double-stranded DNA-binding activity in budding yeast is Rap1, a protein also involved in transcriptional regulation at many other nontelomeric chromosomal sites. *Saccharomyces cerevisiae* Rap1 binds directly to yeast telomeric repeats through two Myb-like DNA-binding domains. Rap1 is instrumental in recruiting to yeast telomeres a set of proteins involved in telomere length regulation and capping (the Rap1 interacting factors 1(Rif1) and Rif2) and in the assembly of a

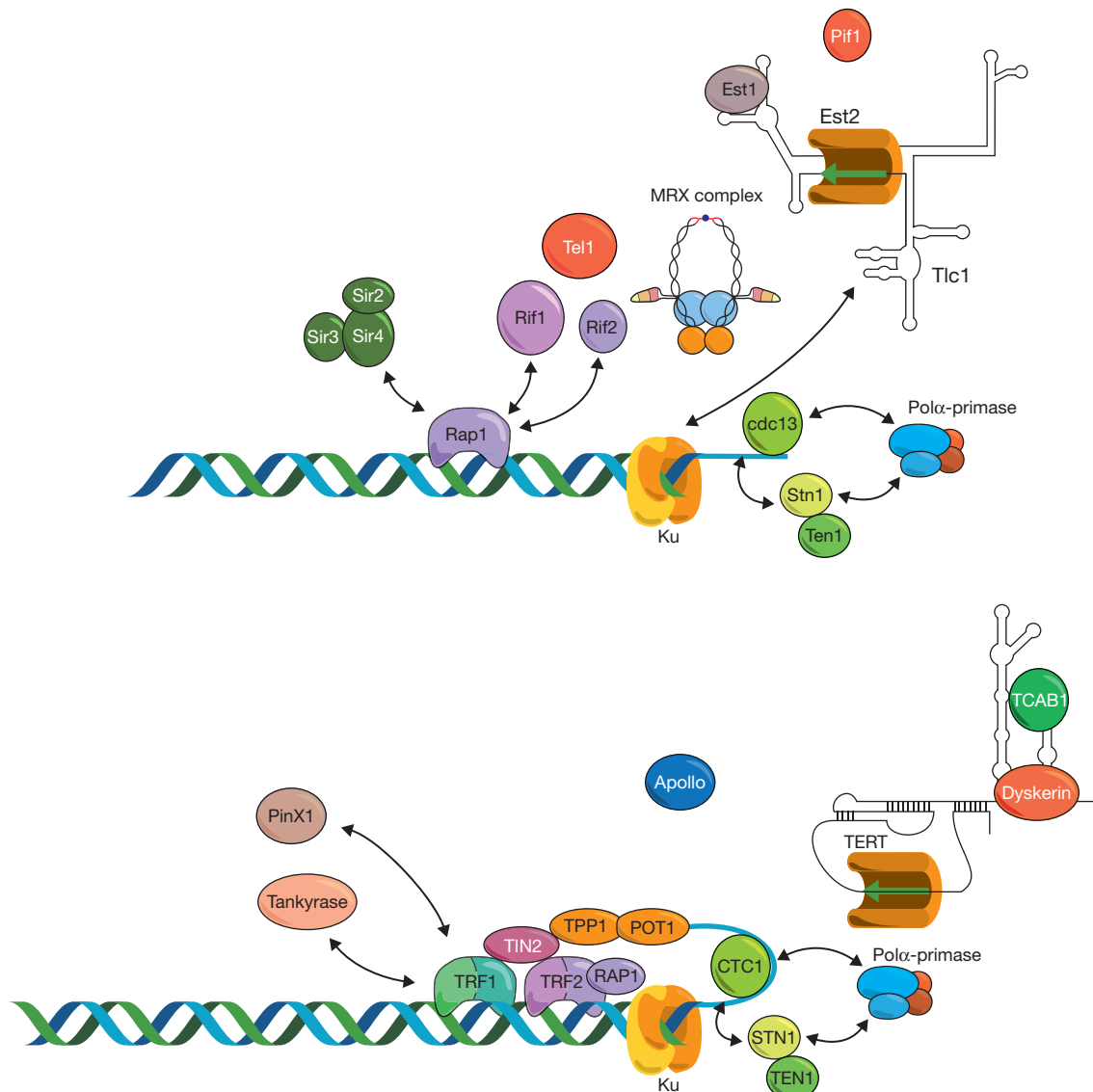


Figure 3 Representation of some of the characterized components of the budding yeast and human telomeric complexes. All proteins depicted have been shown either to affect telomere behavior and/or to interact with other telomere components. The fold-back t-loop structure at human telomeres is not indicated in this diagram.

complex (Sir2, 3, and 4 proteins) that results in the transcriptional repression of genes near telomeres (telomere position effect, or TPE). In contrast, mammalian and fission yeast Rap1 appears to be recruited to telomeres indirectly, through the binding to a different class of double-stranded DNA-binding factors, TRF1 and TRF2 in vertebrates and Taz1 in fission yeast. These proteins contain a single Myb repeat that is essential for DNA binding, and bind as dimers. They share among themselves a large domain (TRF homology (TRFH)) that is responsible for homodimerization. Rap1 is recruited to mammalian and fission yeast telomeres through binding with TRF2 and Taz1, respectively. TRF1 recruits the bridging factor TIN2, together with TRF2, as well as tankyrase, a telomeric poly (adenosine diphosphate (ADP)-ribose) polymerase. TIN2, in turn, is responsible for recruiting TPP1 and the single-stranded DNA-binding protein POT1, both of which are

telomerase regulators. The six mammalian telomere-binding proteins, TRF1/2, RAP1, TIN2, TPP1, and POT1, constitute a stable subcomplex that together serves to cap telomere ends, and is referred to as 'shelterin' (see below).

In addition to double-stranded DNA-binding activities, single-stranded DNA-binding factors also appear to be an essential and general feature of telomeres. Overhang binding proteins were initially characterized in ciliates, where they form tenacious salt-resistant complexes with the single-stranded DNA termini. In budding yeast the Cdc13 protein is bound to the single-stranded overhangs created in S-phase in this organism and it carries out essential functions in telomere replication, through interaction with the telomerase and DNA replication complex, and in telomere protection, through interactions with the Stn1 and Ten1 proteins (as part of the Cdc13/Stn1/Ten1 or CST complex). In fission yeast and

mammals, POT1, an ortholog of the ciliate end factors, is involved in telomere protection in fission yeast and in telomere length regulation in mammals. Both Cdc13 and Pot1 bind DNA through an oligonucleotide/oligosaccharide binding (OB) domain. Recently, Cdc13 homologs (CTC1) have been identified in plants, fission yeast, and mammals, as have homologs of Stn1 and Ten1, suggesting that many eukaryotes may use multiple strategies to protect telomeric single-stranded DNA ends.

An additional factor, present at telomeres from yeast to humans, is the DNA repair protein Ku, which might bind to telomeres through its non-sequence-specific DNA end-binding activity. However, protein-protein interactions between Ku and Sir4, and between Ku and TRF2, have been described in yeast and humans, respectively.

Telomeric RNA

It has long been known that telomeric regions in a variety of organisms are in a heterochromatic state that is repressive of the transcriptional activity of nearby genes. In recent years, however, the surprising discovery has been made that telomeric repeats are actively transcribed and that RNA corresponding to the G-rich strand (but not the C-rich strand) accumulates in mammalian cells (Telomeric Repeat-containing RNA (TERRA)). This RNA is associated with telomeres and appears to carry out essential functions in regulating telomere protection and replication.

Telomere Replication

Telomeres are normally maintained through the action of the telomerase enzyme. Although telomerase activity in an *in vitro* assay requires only the reverse transcriptase-like protein motif and the RNA component, it is now clear that the action of the enzyme at telomeres is subject to complex regulation *in cis* at each individual telomere.

Positive Regulation of the Telomerase Enzyme

In budding yeast, a set of genes is required for telomerase activity *in vivo* in addition to the catalytic core components Est2 and Tlc1. These include Est1, which appears to bridge an interaction between telomerase and Cdc13 required to recruit and activate the enzyme at the telomere, and telomerase-associated Est3, whose mode of action is unclear. The combined action of the DNA damage checkpoint kinases Tel1 and Mec1 is also necessary for telomerase activity in budding yeast, possibly by promoting a structural transition in the telomere complex from a closed to an open state that is accessible to telomerase. A similar requirement exists in fission yeast for the Tel1 and Mec1 homologs Tel1 and Rad3. In mammalian cells, a POT1-TPP1 complex may play a direct role in telomerase recruitment. Significantly, budding yeast telomerase action is strictly limited to the S-phase of the cell cycle and requires active DNA replication.

A Negative Regulatory Loop Stabilizes Telomere Length

In cells expressing telomerase, telomere length is maintained around a constant average value that is species specific. Evidence in a variety of organisms has led to the proposal that telomere length is controlled by a negative feedback, protein-counting mechanism. In this scenario, the telomere length regulation machinery somehow senses the number of double-stranded DNA-binding factors that are associated with the telomere at any given time. An increase in the number of these negative regulators results in the inhibition of telomerase activity *in cis* at the telomere (Figure 4). The inhibitory function of the DNA-binding factors (Rap1 in budding yeast, Taz1 in fission yeast, and TRF1/2 in mammals) is apparently mediated by a set of interacting factors (Rif1 and Rif2 in budding yeast, TIN2, TPP1, and POT1 in humans). The precise mechanism of this proposed *cis* inhibition of telomerase is unknown, but might, at least in some organisms, involve formation of the t-loop structure described above.

Telomerase-Independent Maintenance of Telomeres

In the absence of telomerase, alternative pathways for the maintenance of telomeric repeats exist that are based on homologous recombination. In budding yeast, telomere-associated sequences can contribute to this recombinational process. In mammals, several tumor cell lines have been shown to maintain their telomeres without detectable telomerase activity through a mechanism named ALT, for alternative lengthening of telomeres. Finally, the fruit fly *Drosophila melanogaster* dispenses altogether with short telomeric repeats and forms telomeres through the repeated insertion of telomere-specific transposable elements (HeT-A and TART) onto the ends of its chromosomes. In the fruit fly, telomere protection requires the action of proteins involved in heterochromatin formation, and similar mechanisms might be activated in other organisms upon loss of telomerase activity.

Telomere Capping

Telomeres have evolved multiple mechanisms to protect chromosome DNA ends from deleterious degradation of fusion events and to prevent chromosome ends from eliciting a DNA-damage (checkpoint) response. In budding yeast, the CST complex plays a critical role in inhibiting 5' end resection and Mec1-dependent checkpoint activation in G2/M phases. Rif2 also acts to block resection and appears to inhibit Tel1 (ATM) association to telomeres together with Rif1. Rap1, acting through Rif1/2 and/or independently, blocks telomere-telomere fusion by nonhomologous end joining (NHEJ). In mammals, telomere protection is carried out by the shelterin complex (TRF1/2, RAP1, TIN2, TPP1, and POT1), with Rap1 being the only component conserved in budding yeast. The POT1 and TRF2 components of shelterin are dedicated to inhibition of ATM-related (ATR) and ATM action *in cis* at telomeres, respectively, thus preventing either G1/S or G2/M cell-cycle arrest, and subsequent senescence or apoptosis. These two shelterin components also help to block NHEJ and homology-dependent recombination at telomeres, the latter of which appears to be a specialized

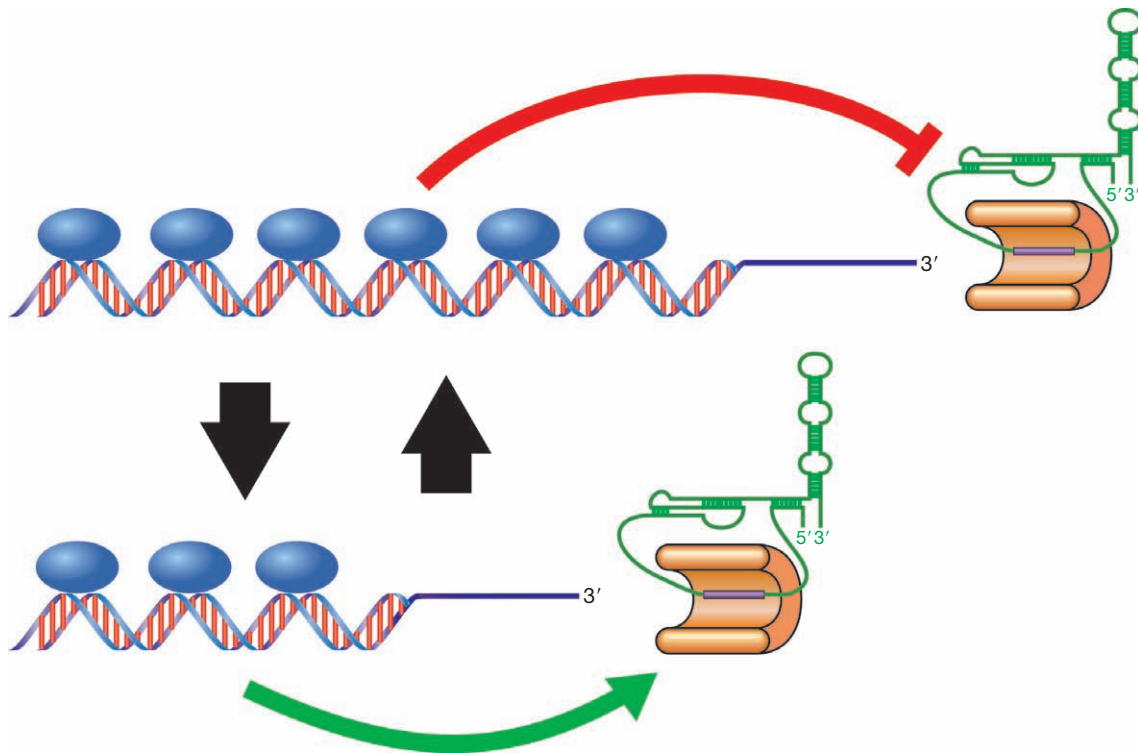


Figure 4 Depiction of the protein counting mechanism for telomere length regulation. In this model, longer telomeres have higher numbers of inhibitory factors bound, resulting in repression of telomerase activity *in cis*. Diminished action by telomerase results in telomere shortening, loss of the bound inhibitor, and telomerase activation. In this manner, telomeres are maintained in a state of dynamic equilibrium. The organism in which our understanding of the telomeric complex is the most complete, with regard to both the identification of protein components and their function, is the budding yeast *Saccharomyces cerevisiae* (Figure 3).

function of mammalian RAP1. Finally, the fold-back t-loop DNA structure may play a key role in capping.

Telomerase Repression in Somatic Cells

Implications for Replicative Senescence in Human Disease

Mice engineered to lack the gene encoding telomerase RNA (mTRT) exhibit a gradual (approximately sixth generation) loss of fertility, defects in cell proliferation, and an increase in the frequency of end-to-end chromosome fusions. These phenotypes result from telomere repeat erosion and consequent loss of telomere function. Since active telomerase enzyme is primarily detected in mouse germline cells, these data indicate that telomerase-dependent telomere length re-setting in the germline is essential for long-term survival in the mouse, and by inference in all mammals. In humans, the gene encoding the catalytic subunit of telomerase is also repressed in most somatic cells, causing telomeres to undergo progressive shortening with each cell division. This process leads to cellular senescence in culture and can be reversed, at least in some primary cell lines, by the ectopic expression of an active telomerase gene in the senescing cells. These observations in cultured primary cells have led some researchers to propose that telomerase repression in humans, and consequent telomere erosion, acts as a kind of mitotic clock that determines aging of the whole organism.

The results in the telomerase RNA knockout models have been corroborated by the identification of an increased array of

genetic defects in components of telomerase or in shelterin that result in age-related disorders. The most common of these, dyskeratosis congenita (DC), leads to abnormalities in skin pigmentation and in the oral mucosa, and to nail dystrophy. The most severe outcomes of DC are pulmonary fibrosis, bone marrow failure, and cancer. These clinical manifestations suggest a defect in maintaining the proliferative capacity of adult tissues. In agreement with the idea that telomerase is important in guaranteeing the replicative potential of mammalian cells, a large fraction of DC patients has mutations in either telomerase or the telomerase RNA, or in dyskerin, a telomerase-associated protein. A smaller percentage of affected individuals have mutations in additional telomeric genes, such as the shelterin component TIN2, or the TCAB1 protein, which is required for the proper nuclear localization of telomerase. A second familial syndrome, Hoyeraal–Hreidarsson syndrome, has even more severe phenotypes and is also caused by mutations in dyskerin, telomerase, and TIN2, as well as in the gene coding for the protein Apollo, which has an important role in telomere replication and the processing of telomere single-stranded overhangs.

Telomere Maintenance and Genome Integrity: Telomerase Inhibition as a Tumor Suppressor Mechanism

An attractive view is that the telomere mitotic clock in humans might represent a critical barrier against malignant transformation. Several experiments suggest that either normal telomere

erosion or uncapping caused by mutational alteration of key telomere proteins will induce a DNA-damage checkpoint and apoptosis (programmed cell death). This mechanism may place a limit on continued cell proliferation, which is a prerequisite for tumor formation. However, telomere uncapping also leads to telomere–telomere fusions and rampant genomic instability, which is thought to be a driving force in oncogenesis. Telomere dysfunction might therefore represent an important step in the early stages of tumorigenesis, particularly in cells with compromised checkpoint (DNA damage surveillance) function. The proper regulation of telomerase and telomere capping thus appear to be remarkably critical cellular functions.

Telomerase and Stem Cells

Consistent with its proposed essential function in continued cell proliferation, telomerase activity is also a characteristic feature of stem cells. A wealth of genetic evidence, from both knockout mice lacking telomerase activity and humans with reduced telomerase function, indicates that telomerase plays a key role in stem cell maintenance. In addition to its ability to counteract telomere erosion during cell division, telomerase may also play a role in the stimulation of quiescent stem cells through direct transcriptional effects on Wnt/beta-catenin pathway target genes. Finally, a number of (largely correlative) studies are beginning to reveal effects of stress, at the organismal level, on telomerase activity and telomere length that may have direct consequences for tissue regeneration and aging. These and other studies indicate that telomeres and telomerase, in addition to tumor suppressor links, may have

critical functions in regulatory networks controlling stem cell proliferation and organismal aging.

See also: [Molecular Biology: DNA Mismatch Repair and Homologous Recombination](#); [DNA Mismatch Repair and the DNA Damage Response](#); [DNA Polymerases: Kinetics and Mechanism](#); [DNA Replication Fork, Eukaryotic](#); [Eukaryotic DNA Polymerase \$\delta\$](#) ; [Nonhomologous End Joining in Eukaryotes](#); [Nuclear Organization, Chromatin Structure, and Gene Silencing](#); [Processivity Clamps in DNA Replication](#).

Further Reading

- Cesare AJ and Reddel RR (2010) Alternative lengthening of telomeres: Models, mechanisms and implications. *Nature Reviews Genetics* 11: 319–330.
- de Lange T (2009) How telomeres solve the end-protection problem. *Science* 326: 948–952.
- de Lange T, Lundblad V, and Blackburn E (eds.) (2006) *Telomeres*, 2nd edn., Cold Spring Harbor Monograph Series 45. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Gilson E and Geli V (2007) How telomeres are replicated. *Nature Reviews Molecular Cell Biology* 8: 825–838.
- McEachern MJ and Haber JE (2006) Break-induced replication and recombinational telomere elongation in yeast. *Annual Review of Biochemistry* 75: 111–135.
- Sahin E and Depinho RA (2010) Linking functional decline of telomeres, mitochondria and stem cells during ageing. *Nature* 464: 520–528.
- Shay JW and Wright WE (2010) Telomeres and telomerase in normal and cancer stem cells. *FEBS Letters* 584: 3819–3825.
- Shore D and Bianchi A (2009) Telomere length regulation: Coupling DNA end processing to feedback regulation of telomerase. *EMBO Journal* 28: 2309–2322.
- Wellinger RJ (2010) When the caps fall off: Responses to telomere uncapping in yeast. *FEBS Letters* 584: 3734–3740.

Thyroid-Stimulating Hormone/Luteinizing Hormone/Follicle-Stimulating Hormone Receptors

D L Segaloff and M Ascoli, The University of Iowa Carver College of Medicine, Iowa City, IA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Deborah L. Segaloff, Dario Mizrahi, and Mario Ascoli, volume 4, pp. 180–186, © 2004, Elsevier Inc.

Glossary

Chorionic gonadotropin (CG) A hormone secreted by the placenta that is nearly identical to the pituitary hormone, luteinizing hormone (LH), and is, therefore, recognized by the LH receptor; also, choriogonadotropin.

Follicle-stimulating hormone (FSH) A pituitary hormone recognized by the FSH receptor; also, follitropin.

Glycoprotein hormones The structurally related hormones thyroid-stimulating hormone (TSH), LH, CG, and FSH.

The glycoprotein hormone receptors are the LH, FSH, and TSH receptors.

Gonadotropins LH, CG, and FSH; members of the glycoprotein hormone family that act upon the gonads. The gonadotropin receptors are the LH and FSH receptors.

G-protein-coupled receptors (GPCRs) A class of membrane receptors that contain seven transmembrane regions and mediate their actions through the activation of G proteins.

Luteinizing hormone (LH) A pituitary hormone recognized by the LH receptor; also, lutropin.

Thyroid-stimulating hormone (TSH) A pituitary hormone recognized by the TSH receptor; also, thyrotropin.

Expression and Physiological Roles of the Glycoprotein Hormone Receptors

The Thyroid-Stimulating Hormone Receptor

TSHR is expressed primarily in the follicular cells of the thyroid. In more recent years, it has also been detected in lymphocytes, thymus, pituitary, testis, kidney, brain, adipose tissue and fibroblasts, heart, and bone.

In response to thyroid-stimulating hormone (TSH), the TSH receptor (TSHR) stimulates the synthesis and secretion of thyroid hormone by the follicular cells of the thyroid. Autoantibodies to the TSHR that are stimulatory are present in individuals with Grave's disease and stimulate the TSHR, causing excessive and unregulated secretion of thyroid hormone. Conversely, inhibitory autoantibodies to the TSHR are found in individuals with Hashimoto's thyroiditis. These antibodies bind to the TSHR and inhibit the binding of TSH, thus blocking the synthesis and secretion of thyroid hormones.

The TSHR present in the adipose and connective tissue of the orbit of the eye is thought to play a role in the development of exophthalmos found in individuals with Grave's disease. Recent studies have also suggested that TSHR present in bone plays a role in bone remodeling. The physiological roles of the TSHR in other nonthyroid tissues are not yet known.

The Luteinizing Hormone Receptor

The luteinizing hormone receptor (LHR) is expressed primarily in the ovaries and the testes. Within the ovary, the LHR is present on theca and interstitial cells and on mature granulosa cells. After ovulation, the granulosa and theca cells of the ruptured follicle differentiate into the luteal cells, and express higher levels of LHR. In males, LHR is expressed on the Leydig cells. In both males and females, nongonadal expression of LHR has also been reported in the reproductive tract and many other tissues.

In the nonpregnant postpubertal female, ovarian theca cells respond to luteinizing hormone (LH) with increased synthesis of androgens, which are used as substrates for estrogen production by follicle-stimulating hormone (FSH)-stimulated granulosa cells. Mature granulosa cells respond to LH with increased progesterone production. In response to the ovulatory surge of LH, the ovarian LH receptors mediate ovulation. If pregnancy ensues, then the LHR of the corpus luteum responds to placental chorionic gonadotropin (CG) with increased progesterone synthesis. As such, the LHR is essential for the maintenance of pregnancy, particularly during the first trimester.

In males, the testicular LHR plays an important physiological role during fetal development. Maternal human chorionic gonadotrophin (hCG) stimulates the fetal Leydig cells to synthesize testosterone, which is required for the differentiation of the external male genitalia and for the descent of the testes into the scrotum. Testosterone levels in boys decrease after birth and the LHR remains unstimulated until the time of puberty. After puberty, the testicular LHR responds to pituitary LH with increased testosterone synthesis.

The physiological roles of nongonadal LHR are not known. It should be pointed out, though, that the only functional consequences detected due to loss-of-function or gain-of-function mutations of the LHR described in males or females have been restricted to abnormalities in reproductive physiology.

The Follicle-Stimulating Hormone Receptor

The follicle-stimulating hormone receptor (FSHR) is expressed in ovarian granulosa cells and in Sertoli cells within the seminiferous tubules of the testes. In postpubertal females, the FSHR mediates follicular growth and controls estrogen synthesis. In postpubertal males, pituitary FSH facilitates spermatogenesis by stimulating the Sertoli cells that are adjacent to the developing sperm to synthesize and secrete components

needed for spermatogenesis. Although it is accepted that optimal spermatogenesis requires the actions of FSH, it is controversial as to whether FSH is essential for this process.

Structural Organization of the Glycoprotein Hormone Receptors

Serpentine Regions

The glycoprotein hormone receptors contain the seven membrane-spanning regions prototypical of the superfamily of GPCRs. Residues that are conserved within the transmembrane (TM) domains of the family A GPCRs are also generally conserved in the glycoprotein hormone receptors. High-resolution crystal structures of rhodopin and other family A GPCRs have provided templates for creating models of the TM regions of the glycoprotein hormone receptors, thus permitting investigators to predict the interhelical interactions maintaining the receptors in their inactive states. The crystal structures also revealed the presence of an eighth α -helix that extends from TM7 and lies parallel to the inner face of the plasma membrane. Helix 8 extends until a palmitoylated cysteine residue and serves to anchor the helix to the plasma membrane. This cysteine is conserved as a single residue or as a pair in the glycoprotein hormone receptors. An alignment of the human TSHR, LHR, and FSHR is shown in [Figure 1](#). The amino acid identity is greatest between the glycoprotein hormone receptors in the transmembrane regions and helix VIII.

Extracellular Domains

A unique feature of the glycoprotein hormone receptors is their relatively large (i.e., 300–400 amino acids) extracellular domains. This is the receptor domain that is responsible for the selective recognition and high-affinity binding of each of the glycoprotein hormones. The extracellular domains are N-glycosylated and a fully conserved tyrosine residue has been shown to be sulfated in the TSHR. The TSHR has the largest extracellular domain and is clipped once the receptor is inserted at the plasma membrane. This proteolytic cleavage results in the formation of an α -subunit containing a portion of the N-terminal extracellular domain and a β -subunit containing the remaining of the N-terminal extracellular domain and the transmembrane and C-terminal domains. Although the α - and β -subunits are initially bound to each other by disulfide bonds, these are reduced and the α -subunit is released from the membrane-bound β -subunit.

The extracellular domains of the glycoprotein hormone receptors consist of a series of leucine-rich repeats (LRRs) followed by a cysteine-rich hinge (linker) region that connects the LRR region to the serpentine domain. The N-terminal portion of the extracellular domain, containing the first LRR, also corresponds to a cysteine-rich region (see [Figure 1](#)). The LRR domain comprises the hormone-binding region of the receptor and the hinge region appears to be necessary for signal transduction. High-resolution crystal structures have been obtained for the hormone-binding domain of the FSHR complexed with FSH and for the TSHR complexed to a TSHR antibody. Each LRR is formed by a β -strand and an α -helix joined by short loops and positioned in a nearly antiparallel

orientation. The tandem arrays of the LRRs form a horseshoe-like structure with consecutive β -strands forming a parallel β -sheet at the concave surface of the horseshoe. FSH is bound to the receptor through contact points with the β -sheets present at the concave surface. The juxtaposition of the hormone-bound FSHR_{HB} relative to the serpentine domain is not known with certainty due to the absence of the hinge region in the structural studies. However, in light of mutagenesis studies, the structural data suggest an orientation that positions the hormone α subunit toward the serpentine domain of the receptor.

Homology cloning and data mining have uncovered additional GPCRs with large extracellular domains containing a variable number of LRRs. Four of these (designated LGR4–8) are found in mammals. Interestingly, the ligands for LGR7 and LGR8, relaxin and/or Leydig cell insulin-like peptide, are structurally homologous to insulin rather than to the glycoprotein hormones.

Dimerization/Oligomerization of the Glycoprotein Hormone Receptors

Each of the glycoprotein hormone receptors can be detected as dimeric and oligomeric complexes in living cells under basal conditions. The propensity for self-association does not appear to be correlated with the activation state of the receptor. Furthermore, the LHR and FSHR have been shown to form dimers and oligomers within the endoplasmic reticulum, suggesting that the organization of the receptors into multimeric complexes is a constitutive process that initiates shortly after biosynthesis and that the receptors are transported to the plasma membrane already self-associated. Both the serpentine as well as the extracellular domains contribute to the dimerization/oligomerization.

Genomic Organization of the Glycoprotein Hormone Receptors

The genes for these three receptors have a similar structure. The LHR gene has 11 exons and the TSHR and FSHR genes each have 10 exons. The last exon codes for a small portion of the N-terminal extracellular domain and the entire serpentine region, including the C-terminal cytoplasmic tail. The other exons are small and they are spliced together to form most of the N-terminal extracellular domain.

The TSHR, LHR, and FSHR genes map to human chromosomes 14q31, 2p21, and 2p21–16, respectively.

Activation of the Glycoprotein Hormone Receptors

G Proteins and Cellular Pathways Activated by the Glycoprotein Hormone Receptors

The identity of the G proteins activated by the FSHR has not been directly examined. The TSHR has been shown to activate all four families of G proteins and the LHR activates at least three or perhaps all four families as well. All of these receptors activate the cyclic adenosine monophosphate (cAMP) and inositol phosphate/diacylglycerol signaling cascades. For each of the glycoprotein hormones, the half-maximal effective

functions of their target cells. The functional consequences of the activation of the inositol phosphate/diacylglycerol pathway are not fully understood. In addition to controlling the differentiated functions of target cells, the glycoprotein hormones also control their proliferation. The three glycoprotein hormones can each activate the extracellular regulated kinase 1/2 cascade (ERK1/2) in their target cells and, at least in the case of the LHR and the TSHR, this activation seems to be

mediated by cAMP. Activation of ERK1/2 also requires the epidermal growth factor (EGF) receptor, but how this is integrated with the cAMP pathway is not well understood and may be target-cell-specific.

Mechanism of Activation

Although it is clear that binding specificity is provided by the large extracellular domains of the receptors and the β -subunits of the hormones, it is not yet clear how the binding of hormone to the extracellular domain leads to receptor activation. Three general models have been proposed. In one model, hormone binding is thought to induce a conformational change in the extracellular domain of the receptor, thus allowing it to interact with and activate the serpentine region. In another model, the common α -subunit of the glycoprotein hormones is thought to directly interact with and activate the serpentine region of the receptor. This model is consistent with that inferred from structural studies of FSH-FSHR_{HB}. The last model proposes that the extracellular domain constrains a constitutive activity intrinsic to the serpentine region and that hormone binding to the extracellular domain relieves this inhibition. This latter model has received experimental support when tested with the TSHR.

Ultimately, agonist binding or constitutively activating mutations are predicted to lead to conformational changes within the serpentine domain that stabilize it in an active conformation. Molecular dynamics simulations based on homology models of the serpentine domains have compared the structures predicted in wild-type receptors and those containing activating mutations. These studies suggest that activating mutations cause an increase in the solvent accessible surface area of the cytoplasmic ends of helices 3 and 6, concomitant with the predicted loss of a salt bridge that links these regions in the resting state.

Human Diseases Associated with the Glycoprotein Hormone Receptors

Autoantibodies to the TSHR

In contrast to the LHR and FSHR, the TSHR can serve as an antigen, eliciting, in some individuals, autoantibodies to the TSHR. Stimulatory autoantibodies to the TSHR cause Grave's disease, a form of hyperthyroidism, whereas blocking autoantibodies (antibodies that bind to the TSHR and prevent the binding of TSH, but do not activate the receptor) cause hypothyroidism.

Naturally Occurring Mutations of the Human Glycoprotein Hormone Receptor Genes

Both activating (gain-of-function) and inactivating (loss-of-function) mutations of the glycoprotein hormone receptor genes have been identified in patients with particular endocrine disorders. Activating mutations of these receptors result in the constitutive activation and unregulated elevation of intracellular cAMP in the absence of bound agonist. Most activating mutations have been found in the serpentine regions of the receptors; however, activating mutations of the TSHR have also been found in the extracellular domain. Germline

mutations resulting in constitutive activation of the glycoprotein hormone receptors are inherited in an autosomal dominant manner.

Loss-of-function mutations refer to a number of different types of mutations that ultimately cause decreased target-cell responsiveness to hormone. These include mutations that impair coupling to G protein, hormone binding, and/or cell surface expression. The majority of loss-of-function mutations of the glycoprotein hormone receptors result in decreased cell-surface expression of the mutant receptor due to the retention of the misfolded mutant in the endoplasmic reticulum. Therefore, loss-of-function mutations have been identified in all regions of the glycoprotein hormone receptors. Mutations of the glycoprotein hormone receptors resulting in a loss of function are inherited in an autosomal recessive manner.

TSHR naturally occurring mutations

A large number of activating mutations of the TSHR have been identified in individuals with autonomously functioning thyroid adenomas (somatic mutations) or autosomal-dominant nonautoimmune hyperthyroidism (germline mutations).

Many loss-of-function TSHR mutations have been identified in individuals with nonautoimmune congenital hypothyroidism, where the degree of hypothyroidism correlates with the extent of target-cell unresponsiveness to TSH.

LHR naturally occurring mutations

Numerous activating mutations of the LHR have been identified in boys with gonadotropin-independent precocious puberty (testotoxicosis). In these cases, the constitutively active human luteinizing hormone receptors (hLHRs) inappropriately stimulate testosterone production in Leydig cells under conditions of prepubertal levels of LH. Only one LHR mutation, D578H, has been found to also cause Leydig cell adenomas. Unlike the other LHR-activating mutations, D578H mutations have been somatic rather than germline in origin. Although it was initially postulated that D578H may activate a pathway(s) different from those activated by the other constitutively active LHRs, such differences have not been observed. Therefore, as with the TSHR, tumor formation by D578H is most likely due to its somatic origin.

Many different loss-of-function LHR mutations have been identified in genetic males with disorders of sexual differentiation, ranging from micropenis (partial loss of function) or pseudohermaphroditism (severe loss of function). In the latter case, the individual presents with female external genitalia but fails to undergo breast development or menstruation at the time of puberty. Female siblings with homozygous or compound heterozygous loss-of-function LHR mutations are infertile.

FSHR naturally occurring mutations

Only one naturally occurring activating mutation of the FSHR in males has been reported. This mutation was found in a hypophysectomized male receiving testosterone supplementation, who exhibited normal spermatogenesis. Studies of this mutant in cell culture, however, have led to disparate results on whether or not it is truly activating. A few activating mutations of the FSHR have been reported in females with spontaneous ovarian hyperstimulation syndrome (OHSS) associated with pregnancy. The OHSS is not due to the constitutive activity of the FSHR, but rather to the fact that constitutively activating

mutations of the FSHR are more readily activated by supramaximal concentrations of hCG than the wild-type FSHR. The increased promiscuous activation by other glycoprotein hormones associated with constitutive activation is unique to the FSHR, as it is not observed with activating mutations of the TSHR or LHR. Activating FSHR mutations have not been found in granulosa cell tumors examined thus far. Nor have they been found associated with nonzygotic twinning. This may be because, in humans, the FSHR exhibits much lower degrees of mutation-induced constitutive activity than the LHR and TSHR.

The first loss-of-function FSHR mutation was identified in a population of Finnish women with primary amenorrhea due to ovarian failure. Cells expressing the recombinant form of this FSHR mutant display no responsiveness to FSH. Surprisingly, men homozygous for this mutation were not infertile, although their fertility may have been impaired somewhat. Therefore, although the FSHR plays an important role in spermatogenesis, these observations raised ongoing debates regarding its absolute requirement. Since then, other loss-of-function mutations of the FSHR have been identified in women with varying degrees of FSH resistance.

See also: **Bioenergetics:** Chloroplasts; **Cell Architecture and Function:** Nuclear Lamina.

Further Reading

- Ascoli M, Fanelli F, and Segaloff DL (2002) The lutropin/choriogonadotropin receptor, a 2002 perspective. *Endocrine Reviews* 23: 141–174.
- Davies TF, Ando T, Lin R-Y, Tomer Y, and Latif R (2005) Thyrotropin receptor-associated diseases: From adenomata to Graves disease. *The Journal of Clinical Investigation* 115: 1972–1983.
- Fan QF and Hendrickson W (2007) Structure of human follicle-stimulating hormone in complex with its receptor. *Nature* 433: 269–277.
- Latronico A and Segaloff DL (1999) Naturally occurring mutations of the luteinizing-hormone receptor: Lessons learned about reproductive physiology and G protein-coupled receptors. *American Journal of Human Genetics* 65: 949–958.
- Paschke R and Ludgate M (1997) The thyrotropin receptor in thyroid diseases. *New England Journal of Medicine* 337: 1675–1681.
- Persma D, Verhoeft-Post M, Berns EM, and Themmen AP (2007) LH receptor gene mutations and polymorphisms: An overview. *Molecular and Cellular Endocrinology* 260–262: 282–286.
- Rapoport B, Chazenbalk GD, Jaume JC, and McLachlan SM (1998) The thyrotropin (TSH) receptor: Interaction with TSH and autoantibodies. *Endocrine Reviews* 19: 673–716.
- Sanders J, Chirgadze DY, Sanders P, et al. (2007) Crystal structure of the TSH receptor in complex with a thyroid-stimulating autoantibody. *Thyroid* 17: 395–410.
- Simoni M, Gromoll J, and Nieschlag E (1997) The follicle-stimulating hormone receptor: Biochemistry, molecular biology, physiology, and pathophysiology. *Endocrine Reviews* 18: 739–773.
- Themmen APN and Huhtaniemi IT (2000) Mutations of gonadotropins and gonadotropin receptors: Elucidating the physiology and pathophysiology of pituitary–gonadal function. *Endocrine Reviews* 21: 551–583.
- Vassart G, Pardo L, and Costagliola S (2004) A molecular dissection of the glycoprotein hormone receptors. *Trends In Biochemical Sciences* 29: 119–126.

Tight Junctions

S Tsukita, Osaka University, Osaka, Japan

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Shoichiro Tsukita, volume 4, pp. 187–189, © 2004, Elsevier Inc.

Glossary

Apical surface The free surface of epithelial cells facing the luminal space or external environment.

Endothelial cells Thin, flattened cells of mesoblastic origin that are arranged in a single layer that lines the blood vessels and certain body cavities (e.g., those of the heart).

Epithelial cells Closely packed cells arranged in one or more layers that cover the outer surfaces of the body or line internal cavities or tubes except the blood vessels, heart, and serous cavities.

Freeze-fracture replica electron microscopy An electron microscopic method that uses metal replicas to visualize the interior of the cell membranes. This technique provides a convenient way for visualizing the distribution of large integral membrane proteins as intramembranous particles on the plane of a membrane.

PDZ domain Protein–protein interaction domain first described in proteins PSD-95, DLG, and ZO-1.

The Src homology region 3 (SH3) domain A domain commonly found in various signaling proteins.

The tight junction (TJ) (zonula occludens: belt-like TJ) is one mode of cell-to-cell adhesion in vertebrate epithelial and endothelial cellular sheets, where it is located at the most apical part of their lateral membranes. It has a major role in preserving the separation of fluid compartments that have different molecular compositions, which is of great importance for the development and maintenance of multicellular organisms. These compartments are delineated by various cellular sheets, which primarily function as barriers for maintaining their distinct internal environment. Within these sheets, individual cells are mechanically linked to form adherens junctions (AJs), which give the sheet its structural integrity, while the intercellular space between adjacent cells is sealed by TJs to prevent the free diffusion of solutes.

Structure

Electron Microscopic Image

In ultrathin section electron microscopy, TJs appear as a zone where plasma membranes of neighboring cells focally make tenacious contact (Figure 1). In freeze-fracture electron microscopy, they are seen as a continuous, anastomosing network of intramembranous particle strands (TJ strands or fibrils) and complementary grooves.

Three-Dimensional Image

Detailed electron microscopy has provided insight into the three-dimensional structure of TJs (Figure 2). Each TJ strand is associated laterally with another TJ strand in apposing membranes of adjacent cells to form paired TJ strands (kissing points), eliminating any intercellular space. Sometimes ion leaky permeability is formed in TJ strands paracellularly.

Integral Membrane Proteins

Several types of integral membrane proteins are localized at TJs including claudins, TJ-associated marvel domain-containing proteins (TAMPs), and immunoglobulin superfamily proteins

such as junctional adhesion molecules (JAMs), coxsackie and adenovirus receptor (CAR), and endothelial cell-selective adhesion molecules (ESAMs) (Figures 1(c) and 3). Claudins, which have a molecular mass of ~23 kDa, are the major constituents of TJ strands and bear four transmembrane domains. In mice and humans, claudins comprise a multigene family consisting of at least 27 members. When each claudin species is overexpressed in mutant mouse fibroblasts lacking endogenous claudin, several species of claudin molecules polymerize within the plasma membranes to reconstitute paired TJ strands. Since many distinct species of claudins are co-expressed in individual cells, similar copolymerization is thought to naturally occur by heterogeneous combinations of claudins. Claudin molecules may adhere to each other in a homotypic and/or heterotypic manner between adjacent TJ strands.

TAMPs are from a recently defined protein family consisting of occludin, tricellulin, and MarvelD3, which contain marvel-domain (MAL and related proteins for vesicle trafficking and membrane link). Occludin, the first discovered TAMP, has a molecular mass of ~65 kDa and bears four transmembrane domains, although none has any sequence similarity with claudin. Occludin is incorporated into TJ strands *in situ*, but unlike claudin is not necessary for TJ formation. Tricellulin consists of four transmembrane domain-containing integral membrane proteins and is especially concentrated in tricellular TJs. MarvelD3 is the most recently discovered TAMP. Its role is still unclear, but it also appears to regulate TJ barrier function.

Several immunoglobulin superfamily proteins such as JAMs, CAR, and ESAM are also found in TJs. The four JAMs, JAM-A, -B, -C, and -D, associate laterally with TJ strands, but do not constitute the strands. These proteins contain an extracellular region with two immunoglobulin (Ig)-like domains, single transmembrane domains, and short intracellular tails that have a PDZ binding motif.

Peripheral Membrane Proteins

Claudins are packed densely within TJ strands with other membrane proteins, especially at the carboxyl-terminal end.

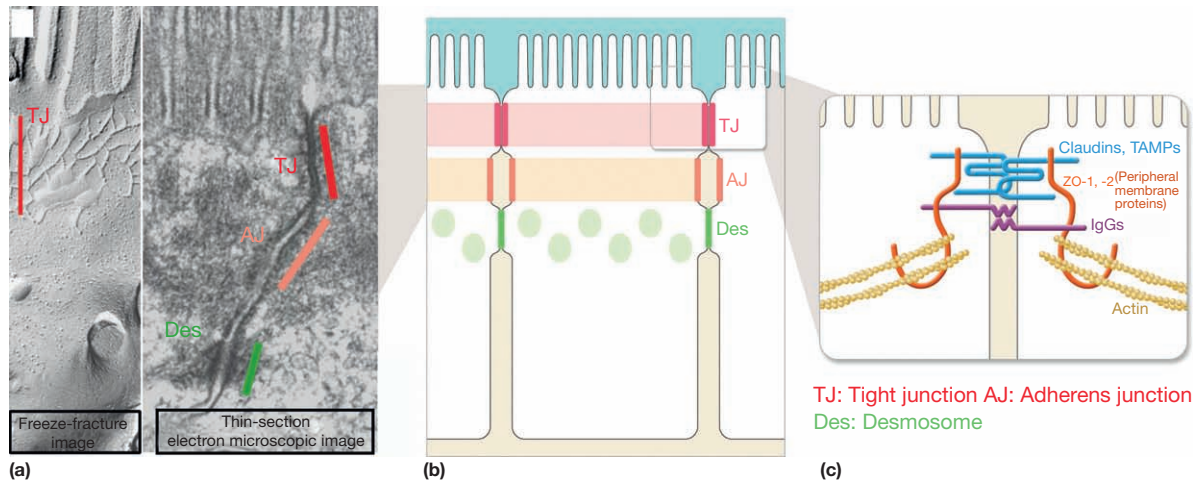


Figure 1 Structure of tight junctions. (a) Electron micrograph of the junctional complex in mouse intestinal epithelial cells. (b) Schematic drawing of simple epithelial cells. The junctional complex is located at the most apical part of the lateral membranes (encircled). (c) Schematic drawing of the molecular architecture of the tight junction. Claudins are tight junctional membrane proteins that form tight junction strands and ultimately the paracellular barrier. TAMPs are tight junctional membrane proteins that help regulate the tight junctional paracellular barrier function. Immunoglobulin superfamily membrane proteins of tight junctions. TJ, tight junction (encircled); AJ, adherens junction; Des, desmosome. Scale = 200 nm (b) and 50 nm (d).

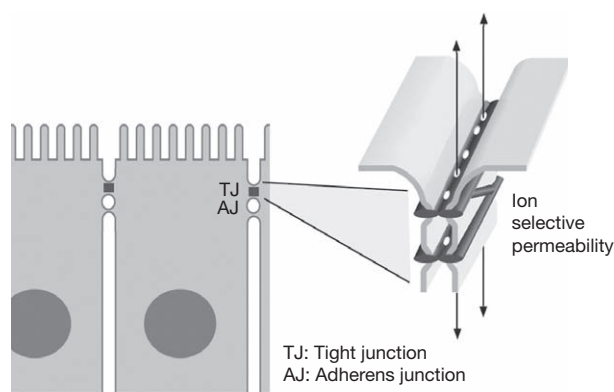


Figure 2 Three-dimensional image of tight junction strands.

Interestingly, the carboxyl terminal mostly consists of valine, which suggests this is where the TJ binds to PDZ proteins. Therefore, the cytoplasmic surface of a TJ strand can be regarded as a magnetic bar that strongly attracts and recruits several PDZ-containing proteins such as ZO-1/-2/-3, Pals-1, and MUPP-1, all of which are involved in either cell signaling or the cytoskeletal structure. Beginning from the NH_2 -terminus, ZO-1/-2/-3 have three PDZ domains (PDZ1–3), one SH3 domain, and one guanylate kinase-like domain, indicating they are membrane-associated guanylate kinase-like homologs. It is the PDZ1 domain that binds directly to the COOH termini of claudins. Recent studies on ZO-1/-2/-3 have revealed that ZO-1/-2, but not ZO-3, are required for claudins to polymerize and form TJ, and to give the TJ its paracellular barrier function.

In addition, several other PDZ-containing proteins are recruited to the cytoplasmic surface of TJ strands including MAGI proteins and AF-6/Afadin. Moreover, the TJ plaque contains scaffolding proteins that lack PDZ domains and proteins that appear to be primarily involved in regulating signaling at

the TJ, rather than proteins that provide basic structural support. These include several adaptor/signaling proteins such as cingulin and JACOP/paracingulin, the angiomin family proteins angiomin, JEAP and MASCOT, signaling proteins and regulators of small guanosine triphosphatases (GTPases; WNK4), phosphatase (PTEN), Rho GEFs (GEF-H1 and p114RhoGEF), and others (sympleskin). Polarity complex proteins including aPKC, PAR-3, PAR-6, PATJ, and PALS-1 are also integrated into TJs. In addition to these resident proteins, proteins including ZONAB, cdk-4, ZO-2, PAR-6, and huASH1 reportedly localize to both the nucleus and the TJ, and tend to regulate cell proliferation, gene expression, and cell differentiation. Thus, the TJ functions as a supramolecular complex.

Functions

Barrier Function

Claudins are directly involved in the formation and barrier function of TJ strands in epithelia that regulate the movement of not only solutes but also pathogens such as bacteria. Furthermore, TJs are not simple barriers; they show ion and size selectivity, a phenomenon known as 'permeability'. The barrier function varies with the cell type and physiological requirements. Most data suggest the first extracellular loop of the claudin protein creates an electrostatic selectivity filter that controls the overall resistance and charge selectivity of the junction's small pores. Charged amino acid side chains in the claudins minimize the permeability of similarly charged ions in solution without affecting the permeability of oppositely charged ions. Such regulated and diversified permeability is required for dynamically maintaining the environment of each compartment. To explain how materials transport across TJs, aqueous pores (or paracellular channels) have been postulated to exist within paired TJ strands. It is now widely believed that the combination and mixing ratios of claudins

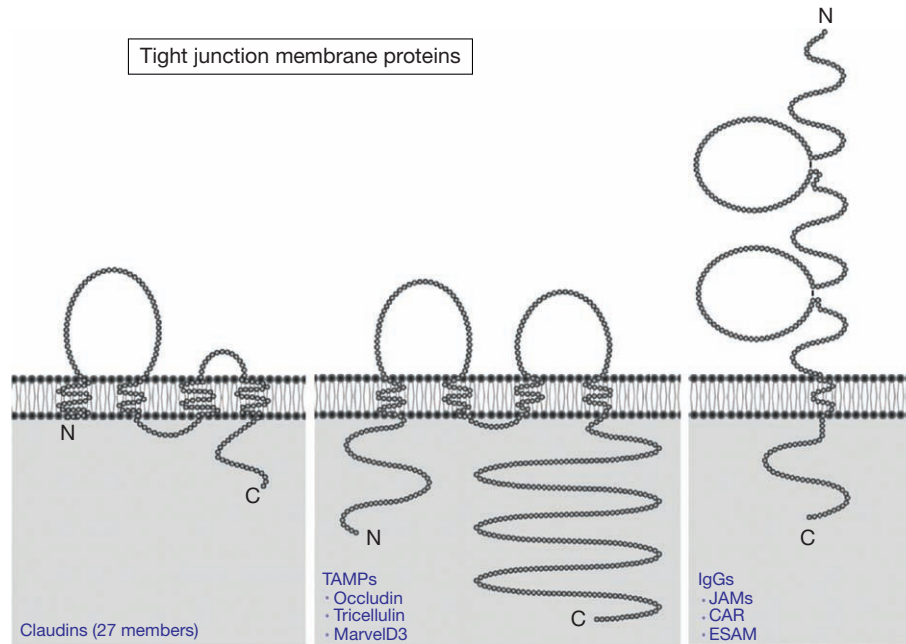


Figure 3 Integral membrane proteins localized at tight junctions. Membrane folding models of mouse claudins, TAMPs, and immunoglobulin superfamily proteins.

within individual paired TJ strands determine the tightness and selectivity of these aqueous pores.

Fence Function

TJ strands are heteropolymers of integral membrane proteins and claudins that continuously encircle the top of individual epithelial/endothelial cells to delineate the border between the apical and basolateral membrane domains. Therefore, it is believed that TJ strands act as a fence, limiting the lateral diffusion of lipids and proteins. Indeed, fluorescently labeled lipids inserted into the outer leaflet of the apical membrane of cultured epithelial cells remain there. In contrast, the importance of TJs in the asymmetric distribution of integral membrane proteins remains controversial. In this respect, ZO-1/-2-deficient cell studies have suggested that even when TJs disappear, the fence function of some proteins is maintained. Thus, the fence function of TJ is under question and it is updated that TJs specifically function as a paracellular barrier.

Claudin Gene Recombinant Mice and Human Diseases

Several claudin genes have been deleted from mice for mutation studies. The resulting changes in permselectivity indicate a permselective barrier role for claudin-1, -2, -5, -11, -14, and -15. On the other hand, deletions of claudin-16 and -19 lead to a phenocopy of their respective human disease mutants and are more easily explained by ion selectivity defects.

Claudin-1 knockout mice show a drastic phenotype of dehydration by destruction of the skin barrier for water. Deletion of claudin-5 resulted in an abnormal blood-brain barrier. Abnormal ion-selective permeability has been seen in the kidneys of claudin-2- or claudin-7-knockout mice. Claudin-11 or -19 induce defects in saltatory conduction due

to mis-sealing nerves. The direct role of claudin in biological homeostasis has been revealed in claudin-2 and -15 knockout mice, which showed a large deviation from ionic biological homeostasis in the intestine and an abnormal transport of nutrients resulting in malabsorption. Furthermore, these abnormal properties resemble those seen in Crohn's disease and ulcerative colitis.

Mice with other knocked-out TJ proteins have also been reported. The knockout of JAMs, for example, showed abnormal immune systems, while knockouts of ZO-1,-2 had lethal embryonic phenotypes. These studies show the importance of tight junctional proteins in disparate biological systems.

Other Possible Functions

Various types of proteins including PDZ-containing proteins are recruited to the cytoplasmic surface of TJ strands to form a huge macromolecular complex. This macromolecular complex is expected to play central roles in the intracellular signaling of epithelial cells by regulating their proliferation and morphogenesis by contact inhibition of cell growth and epithelial cell polarization, respectively.

See also: Lipids Carbohydrates Membranes and Membrane Proteins: Ion Channel Protein Superfamily; Multidrug Resistance Membrane Proteins.

Further Reading

Furuse M, Fujita K, Hiiiragi T, Fujimoto K, and Tsukita S (1998) Claudin-1 and -2: Novel integral membrane proteins localizing at tight junctions with no sequence similarity to occludin. *Journal of Cell Biology* 141: 1539–1550.

- Furuse M, Hata M, Furuse K, et al. (2002) Claudin-based tight junctions are crucial for the mammalian epidermal barrier: A lesson from claudin-1-deficient mice. *Journal of Cell Biology* 156: 1099–1111.
- Mineta K, Yamamoto Y, Yamazaki Y, et al. (2011) Predicted expansion of the claudin multigene family. *FEBS Letters* 585: 606–612.
- Muto S, Hata M, Taniguchi J, et al. (2010) Claudin-2-deficient mice are defective in the leaky and cation-selective paracellular permeability properties of renal proximal tubules. *Proceedings of the National Academy of Sciences of the United States of America* 107: 8011–8016.
- Shen L, Weber CR, Raleigh DR, Yu D, and Turner JR (2010) Tight junction pore and leak pathways: A dynamic duo. *Annual Review of Physiology* 73: 283–309.
- Tamura A, Hayashi H, Imasato M, et al. (2011) Loss of claudin-15, but not claudin-2, causes Na^+ deficiency and glucose malabsorption in mouse small intestine. *Gastroenterology* 140: 913–923.
- Tsukita S, Furuse M, and Itoh M (2001) Multifunctional strands in tight junctions. *Nature Reviews Molecular Cell Biology* 2: 285–293.
- Tsukita S, Yamazaki Y, Katsuno T, Tamura A, and Tsukita S (2008) Tight junction-based epithelial microenvironment and cell proliferation. *Oncogene* 27: 6930–6938.
- Tsukita Sh and Tsukita Sa (1989) Isolation of cell-to-cell adherens junction from rat liver. *Journal of Cell Biology* 108: 31–41.
- Umeda K, Ikenouchi J, Katahira-Tayama S, et al. (2006) ZO-1 and ZO-2 independently determine where claudins are polymerized in tight-junction strand formation. *Cell* 126: 741–754.
- Van Itallie CM and Anderson JM (2006) Claudins and epithelial paracellular transport. *Annual Review of Physiology* 68: 403–429.
- Yamazaki Y, Umeda K, Wada M, et al. (2006) ZO-1- and ZO-2-dependent integration of myosin-2 to epithelial zonula adherens. *Molecular Biology of the Cell* 19: 3801–3811.

Toll-Like Receptors

H Kumar, Indian Institute of Science Education and Research (IISER), Bhopal, India; World Premier International Immunology Frontier Research Center, Osaka, Japan; Osaka University, Osaka, Japan

O Takeuchi and S Akira, World Premier International Immunology Frontier Research Center, Osaka, Japan; Osaka University, Osaka, Japan

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Himanshu Kumar, Kiyoshi Takeda, and Shizuo Akira, volume 4, pp. 190–194, © 2004, Elsevier Inc.

Glossary

Adaptive immunity Adaptive immunity is also known as specific immunity. It is induced by infectious agents through B and T lymphocytes and characterized by specificity and memory for distinct macromolecules of the pathogen. This immunity has the ability to respond more vigorously to repeated infection of same pathogen.

Chemokines Chemokines are a large family of structurally homologous proteins which stimulate and/or regulate leukocyte movement or trafficking from the blood to tissues.

Cytokines Cytokines are proteins secreted by various cell types (immune and nonimmune cells) that mediate inflammatory and immune reactions. Cytokines are principal mediators of communication between cells of the immune system.

Innate immunity Rapid clearances of the pathogen from the host by the mechanisms that exist before infection are known as innate immunity which essentially is not affected

by the repeated infections. The components of innate immunity primarily consist of two components – the cellular component such as epithelial barriers, phagocytic cells (neutrophils, macrophages), and natural killer cells and the soluble component such as the complement system and cytokines. The regulated and coordinated activities of these components result in eradication of most of microbial pathogen.

Type I interferons Type I interferons (IFNs) are family of antiviral cytokines produced by nucleated immune and nonimmune cells such as macrophages, fibroblast cells, etc., and consist of several structurally related IFN- α and a single IFN- β . Type I IFNs primarily inhibit viral replication. In addition to inhibition of virus replication, type I IFNs also increase the lytic potential of natural killer cells, increase expression of class I MHC molecules on virus-infected cells, and stimulate the development of antigen-specific B and Th1 cells responses.

Introduction

Infection of pathogens rapidly activates various components of host immune system. In metazoan, immune system broadly categorized into two components innate immune system and adaptive immune system. Innate immune system (which is also known as natural or native immune system) is first line of host defense system and consists of innate immune cells such as macrophages, dendritic cells (DCs), neutrophils, etc. These cells express the families of germline-encoded receptors known as pattern recognition receptors (PRRs) which recognize several classes of conserved signature molecules such as proteins, lipid, sugars, nucleic acid, lipo-proteins, glyco-proteins, and glyco-lipids derived from pathogen. These signature molecules are known as pathogen-associated molecular patterns (PAMPs). Recognition of PAMPs by PRRs activate series of biochemical signaling events in innate immune cells to induce inflammatory responses via production of cytokines and chemokines for clearance of the pathogen. The innate immune cells are also critical for the development of pathogen-specific adaptive immunity by instructing lymphocytes.

In last one decade, innate immunobiology was the one of most-studied branch of immunobiology. Till date, several families of PRRs such as Toll-like receptors (TLRs), RIG-I-like receptors (RLRs), NOD-like receptors (NLRs), and C-type

lectin receptors (CLRs) were discovered which play a pivotal role in immune surveillance against bacterial, viral, fungal, and protozoal infection. Among these receptors, TLRs are most extensively studied innate receptor. TLRs in human and mouse were discovered after the discovery of toll protein in *Drosophila* which plays a crucial role in immunity against fungus infection. After discovery of toll in *Drosophila*, homologs of toll protein were identified in several species; in human and mouse 10 and 12 TLR members have been identified, respectively. First nine members of both human and mouse TLR are conserved and recognize same PAMPs except mouse TLR8. The PAMP for mouse TLR8 is unknown and mouse TLR10 has stop codon and nonfunctional. Human TLR10, mouse TLR12, and TLR13 are not well characterized.

TLR PAMPs

TLRs recognize several classes of PAMPs such as protein (flagellin recognized by TLR5; hemagglutinin protein recognized by TLR2), sugar polymer (zymosan recognized by TLR2-TLR6 complex; mannan recognized by TLR4), lipid conjugated with sugar (lipid A, recognized by TLR4-MD2 complex; lipoarabinomannan recognized by TLR2), lipid conjugated with protein or peptide (diacyl lipopeptides recognized by TLR2-TLR6 complex, triacyl lipopeptides recognized by TLR1-TLR2), and

nucleic acid (ssRNA recognized by TLR7 or TLR8; dsRNA recognized by TLR3 and CpG motif of DNA recognized by TLR9) derived from pathogen. Nucleic acid TLR sensors (TLR3, TLR7, TLR8, and TLR9) mainly express and recognize their respective PAMPs in the endosomal or endolysosomal compartment of cells, whereas other PAMPs are recognized on surface of the cells. Several TLRs such as TLR2 recognize the several PAMPs in association with other TLR member such as TLR1 or TLR6. On the other hand, some TLRs such as TLR4 associate with non-TLR member MD2 for the recognition of PAMPs, whereas other TLRs such as TLR3 form homodimers for recognition of their PAMPs. The PAMPs and the pathogen recognized by TLRs are explained in great detail in several reviews and shown in Table 1. In addition to pathogen-derived molecules, TLRs are also reported to recognize endogenous host molecules. TLR2 and/or TLR4 recognize high-mobility group box 1 (HMGB1) proteins, hyaluronic acid, fibronectin, surfactant protein A, and biglycan and induce inflammatory cytokine, and this production was completely abrogated in mice lacking these receptors, although we need to carefully interpret the results considering the possible contamination of bacterial components. In addition to this ligand, endosomal-localized TLRs, TLR7 and TLR9 recognize ribonucleoprotein and chromatin–DNA complexes released from dead cells and have been implicated in the development of systemic autoimmune diseases.

Structure of TLRs

TLRs are type I membrane glycoprotein (the N-terminus of this protein faces extracellular space or endosomal lumen and the C-terminus faces cytoplasm) and consist of ectodomain composed of leucine-rich repeats (LRRs), intermediate domain (spanning cell membrane), and cytoplasmic Toll/interleukin-1 receptor (TIR) domain. Cytoplasmic TIR domain consists of a five-strand β -sheet surrounded by five α -helices. Ectodomain contains several LRRs and each LRR contains 20–30 amino acid. Recently, the crystal structure of ectodomain of TLR1–TLR2, TLR3 and TLR4-MD2 with their PAMP were reported. TLR1 and TLR2 exist as a monomer; however, incorporation of synthetic ligand (Pam3CSK4, containing three lipid chains) induces formation of stable heterodimer of TLR1–TLR2 and activate downstream signaling. The co-crystal of TLR1–TLR2–Pam3CSK4 shows that two out of three lipid chains of Pam3CSK4 were embedded into a hydrophobic pocket in TLR2, whereas remaining one lipid chain was embedded into hydrophobic pocket of TLR1. Both of the pockets are located between the centers of the convex region, suggesting that internal convex surfaces rather than outer concave surface of TLR1–TLR2 heterodimer are important for PAMP recognition.

The crystal structure TLR3 was also reported by two independent research groups; they showed that LRR of TLR3 has horseshoe-shaped solenoid structure. Furthermore, the co-crystal

Table 1 TLR PAMPs and immune responses

TLR	Cellular localization of TLR	TLR PAMPs	Effector immune responses
TLR1/2	Cell surface	Triacyl lipopeptides (Bacteria and Mycobacteria)	Induce inflammatory cytokines
TLR2	Cell surface	Peptidoglycan (Gram-positive bacteria), LAM (Mycobacteria), Hemagglutinin (Measles virus), Phospholipomannan (<i>Candida</i>), Glycosylphosphatidyl inositol mucin (<i>Trypanosoma</i>), Zymosan (<i>Saccharomyces</i>)	Induce inflammatory cytokines, type I IFNs induced inflammatory monocytes after virus infection
TLR3	Endosome	ssRNA virus (WNV), dsRNA virus (Reovirus), RSV, MCMV	Induce inflammatory cytokines and type I IFNs
TLR4	Cell surface	LPS (Gram-negative bacteria), Mannan (<i>Candida</i>), Glycoinositolphospholipids (<i>Trypanosoma</i>), Envelope proteins (RSV and MMTV), Endogenous oxidized phospholipids produced after H5N1 Avian influenza virus infection	Induce inflammatory cytokines and type I IFNs
TLR5	Cell surface	Flagellin (Flagellated bacteria)	Induce inflammatory cytokines, induce differentiation of B cells and naïve T cells to IgA producing plasma cells and Th1 and Th17, respectively
TLR6/2	Cell surface	Diacyl lipopeptides (Mycoplasma), LTA (Streptococcus)	Induce inflammatory cytokines
TLR7	Endolysosome	ssRNA viruses (VSV, Influenza virus), purine analog compounds (imidazoquinolines). RNA from bacteria (<i>Group B Streptococcus</i>)	Induce inflammatory cytokines and type I IFNs
Human TLR8	Endolysosome	ssRNA from RNA virus, purine analog compounds (imidazoquinolines)	Induce inflammatory cytokines and type I IFNs
TLR9	Endolysosome	dsDNA viruses (HSV, MCMV), CpG motifs from bacteria and viruses, Hemozoin (Plasmodium)	Induce inflammatory cytokines and type I IFNs
Mouse TLR11	Cell surface	Uropathogenic bacteria, profilin-like molecule (<i>Toxoplasma gondii</i>)	Induce inflammatory cytokines

LAM, lipoarabinomannan; WNV, West Nile virus; RSV, respiratory syncytial virus; MCMV, murine cytomegalovirus; MMTV, mouse mammary tumor virus; LTA, lipoteichoic acid; VSV, vesicular stomatitis virus; HSV, herpes simplex virus; CpG, cytidine-phosphate-guanosine; IFNs, interferons; ds, double stranded; ss, single stranded; and Ig, immunoglobulin.

structure of TLR3 and its PAMP (46 base pair (BP) double-stranded (ds) RNA) reveal that dsRNA binds with lateral side of the convex surface of TLR3 ectodomain. The first three LRRs at the N-terminus and 19–21 LRRs at C-terminus of LRRs which are rich in positive charge amino acids are essential for the interaction with the sugar-phosphate backbones of dsRNA and it is also required for dimerization and subsequent activation of downstream signaling.

In addition to TLR1–TLR2 heterodimer and TLR3 homodimer with their PAMPs, crystal structure of TLR4 with their agonist (lipid A, an endotoxic component of lipopolysaccharide (LPS), consists of phosphorylated diglucosamine and six acyl chains) or antagonist (Eritoran, consisting of phosphorylated diglucosamine and four acyl chains) has been reported. These studies revealed that MD2 forms an hydrophobic pocket where five acyl chains of lipid A are embedded (in case of Eritoran, all four acyl chains are embedded) and remaining one acyl chain are hanging out side of MD2 and interacts with hydrophobic surface of TLR4 which stabilize the complex for the activation of downstream signaling. Collectively, for effective TLR4-mediated signaling, two copies of TLR4, MD2, and LPS are required. Taken together, all of these studies show the unique horseshoe or 'm'-shape-like architecture and provide the better picture for PAMP-induced TLR dimerization which brings TIR domain to close proximity and induce TIR–TIR domain dimerization for activation of downstream signaling through the recruitment of various signaling adaptors.

TLR Signaling Pathway

TLR signaling mediated by the recruitment of mainly four TIR domain-containing adaptors, namely myeloid differentiation primary response gene 88 (MyD88), TIR-containing adaptor protein (TIRAP)/MyD88-adaptor-like (Mal), TIR-containing adaptor inducing IFN β (TRIF), and TRIF-related adaptor molecule (TRAM). TLR1, TLR2, TLR4, TLR5, TLR6, and TLR11 (mouse) express on cell surface. On the other hand, TLR3, TLR7, TLR8 (human), and TLR9 express in endolysosome. All TLRs except TLR3 recruit MyD88 after PAMP ligation and initiate MyD88-dependent signaling pathway. MyD88-dependent signaling pathway primarily induces inflammatory responses via activation of a transcription factor nuclear factor kappa B (NF κ B) and mitogen-activated protein (MAP) kinases such as extracellular signal-regulated kinase (ERK)1/2, p38, and c-jun N-terminal kinase (JNK). Some TLRs also activate interferon-regulatory factors (IRFs) in DCs and plasmacytoid DCs (pDCs). TLR1, TLR2, TLR4, and TLR6 expressing on cell surface require additional TIR domain-containing adaptor protein, TIRAP, which serves as a bridging protein between TLRs and MyD88. TLR3 recruits TRIF, after ligation with TLR3 PAMP (poly IC), and initiates TRIF-dependent signaling for induction of inflammatory cytokines and type I interferons via NF κ B, MAP kinases, and IRFs activation, respectively. Stimulation of DCs and macrophages with TLR4 PAMP recruits both MyD88 and TRIF via TIRAP and TRAM, respectively, to initiate TRIF-dependent and MyD88-dependent signaling for the induction of inflammatory cytokines and type I interferons. To understand the complexity of TLR signaling, it is categorized into two components, MyD88- and TRIF-dependent signaling (Figures 1 and 2).

MyD88-Dependent Signaling Pathway

Conventional DCs (cDCs) and macrophages stimulated with TLR1/2, TLR2/6, TLR5, TLR11, or TLR4 PAMPs initiate MyD88-dependent signaling through recruitment of MyD88 to the TIR domain of TLR. Upon MyD88 recruitment, interleukin-1 receptor-associated kinase (IRAK) family protein kinases are sequentially activated and recruited (first IRAK4, then IRAK1, and finally IRAK2 is recruited) for the activation of TNF receptor-associated factor (TRAF)6. TRAF6 is an E3 ubiquitin ligase and forms a complex with ubiquitin-conjugating enzymes for polyubiquitination at K (lysine) 63 linkage and unconjugated polyubiquitin chain for the activation of transforming growth factor- β -activated kinase (TAK)1 and TAK1-binding protein (TAB). The activated TAK1 finally activates NF κ B and activator protein (AP)-1 transcription factors through the I κ B kinase (IKK) complex and MAP kinases, respectively, for the transcription of inflammatory cytokines (Figure 1).

Plasmacytoid DCs (pDCs) stimulated with TLR7, TLR8 (human), and TLR9 PAMP induce large amount of type I interferons (an antiviral cytokine). This stimulation leads to the transportation of PAMP to the lumen of endosome; however, in the lumen of endosome PAMP neither binds nor activates the signaling pathway through TLR9 until these TLRs are cleaved or processed. For the processing of TLR9, the endosomes are fused with lysosome to form endolysosome where lysosomal enzymes (cathepsins and endopeptidase) activate TLR9 after proteolysis. The activated TLR9 binds with PAMP and initiates MyD88-dependent signaling pathway for the induction of inflammatory cytokine and type I interferons. Although it is not clear, whether TLR7 also undergoes to the similar processing steps, it has been shown that acidification of endosome is essential for the PAMP recognition by TLR7. UNC93B1 is an endoplasmic reticulum (ER)-localized membrane protein and required for the TLR3-, TLR7-, TLR8-, and TLR9-mediated immune responses because UNC93B1-mutant mice show complete abrogation of cytokines after stimulation with their respective PAMPs. Furthermore; in pDCs, UNC93B1 is required for the transportation of TLR7 and TLR9 from ER to endolysosome after stimulation with TLR7 and TLR9 PAMP. Type I interferons primarily induce through the activation of IRF7 after interaction with MyD88 via IRAKs (IRAK4 and IRAK1), TRAF3, and TRAF6. Moreover, studies reveal that IKK α is also critical for IRF7 activation. In addition of type I interferons, pDCs also induce inflammatory cytokines through MyD88-dependent signaling pathway as explained above. cDCs stimulated with TLR7 and TLR9 ligand induce type I interferons via IRF-1 but require IRF7 (Figure 2).

TRIF-Dependent Signaling Pathway

Conventional DCs and macrophages stimulated with TLR3 and TLR4 PAMPs initiate a TRIF-dependent signaling pathway for the induction of inflammatory cytokines and type I interferons. The TRIF-dependent pathway activates NF κ B and MAP kinases by two different signaling pathways via different domains of TRIF; TRAF6 interacts through N-terminus of TRIF to activate NF κ B and MAP kinases, whereas C-terminus of TRIF harboring receptor-interacting protein (RIP) homotypic interaction motif (RHIM) domain interacts with RIP1 and

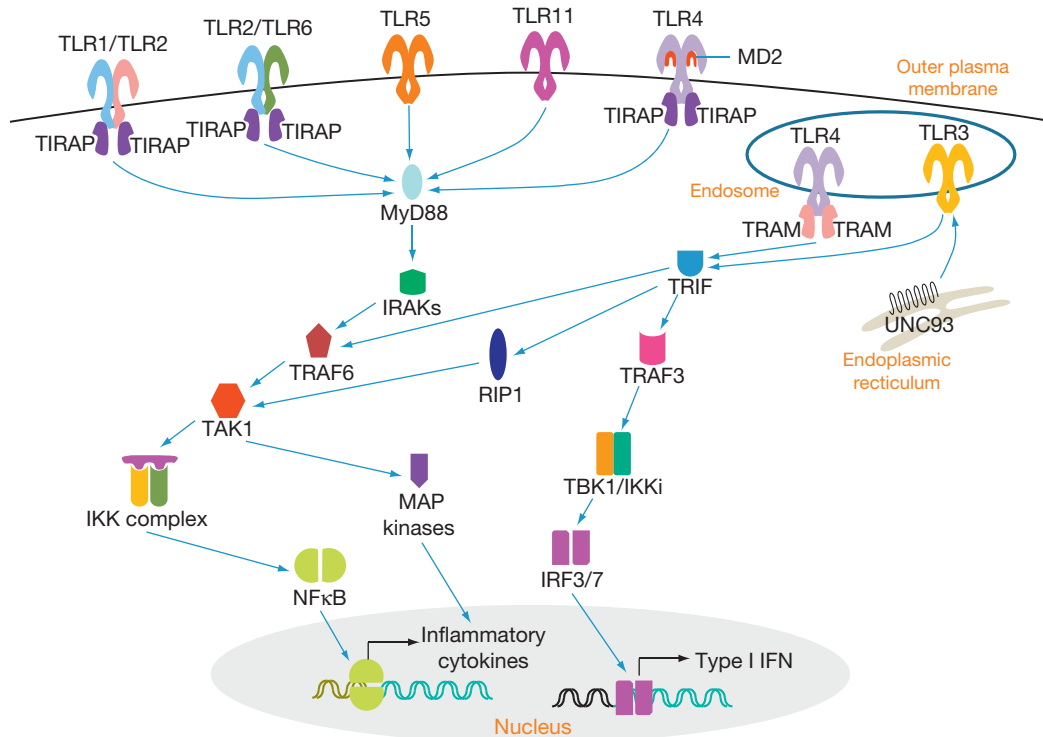


Figure 1 TLR signaling in conventional dendritic cells and macrophages. Ligation of PAMP to the TLR initiates TLR signaling pathways in dendritic cells and macrophages. TLR1, TLR2, TLR4 (TLR4 complex with MD2 to form functional receptor), TLR5, TLR6, and TLR11 recognize the PAMP on cell surface and recruit MyD88 to initiate MyD88-dependent signaling pathway, whereas TLR3 recognizes the PAMP in the endosome (TLR3 translocated to the endosome by transporter protein UNC93B1) and recruits TRIF to initiate TRIF-dependent signaling pathway. TLR1, TLR2, TLR4, and TLR6 recruit MyD88 via additional protein known as TIRAP. TLR4 recruits TRIF via additional protein TRAM after translocation into the endosome. Therefore, TLR4 initiates both MyD88- and TRIF-dependent signaling pathway. In the MyD88-dependent signaling pathway, MyD88 interacts sequentially with different IRAK family proteins which activate TRAF6 and this complex activates TAK1. The activated TAK1 activates NF κ B (via the IKK complex) and MAP kinases. In TRIF-dependent pathway, different domains of TRIF interact with TRAF6 and RIP1 to initiate two independent signaling pathways which subsequently culminate to the activation of NF κ B and MAP kinases. TRIF-dependent pathway also induces type I interferons through activation of IKK-related kinases, TBK1/IKKi, which activate IRF3/7, and TRAF3 serves as a linking protein between TRIF and TBK1/IKKi. The activated NF κ B and MAP kinases finally induce transcription of inflammatory cytokines, whereas activated IRF3/7 induces transcription of family of type I interferons.

RIP3. Two additional proteins, TNF receptor-associated death domain (TRADD) and Fas-associated death domain protein (FADD), form a complex with RIP1, and this complex activates NF κ B. The TRIF-dependent pathway also induces type I interferons through activation of IRF3/7. These IRFs are activated by IKK-related kinase, TANK-binding kinase (TBK)1, and IKKi (IKK ϵ) via TRAF3 which serves as a linker between TRIF and TBK1 and/or IKKi. The activated IRFs are translocated to the nucleus and initiate the transcription of type I interferons (Figure 1).

TLRs and Infectious Diseases

TLRs play a pivotal role in innate immune defense; however, contribution of human TLRs in innate host defense in natural infection was not well studied until recently. Patients with MyD88 and IRAK4 mutation were identified and studied; these patients show unresponsiveness to most of TLR PAMPs except TLR3 PAMP. These patients also show increased susceptibility to pyrogenic bacteria, whereas TLR3-deficient patients are prone

to herpes simplex virus (HSV)-1 infection; however, the mechanism for HSV-1 susceptibility is not clear. Furthermore, patients with UNC93B1 deficiency show unresponsive to TLR3, TLR7, TLR8, and TLR9 PAMP. Collectively, these studies suggest that mutations in the genes encoding TLRs and downstream signal transducing molecules influence innate immune responses and increase susceptibility to many infectious diseases.

TLR Agonist/Antagonist Used as Vaccine Adjuvant or Drug

Several TLR PAMPs have been used as a nonspecific stimulator of immune system or vaccine adjuvant. LPS could be used as a vaccine adjuvant; however, due to its high toxicity in human, its use was restricted as an adjuvant. To remove the toxicity of LPS, LPS derivative monophosphoryl lipid A/trehalose dicorynomycolate was developed which has promising vaccine adjuvant properties against infectious disease and hay fever. By contrast, TLR4 antagonists eritoran and resatorvid were introduced for the treatment of sepsis; currently, these

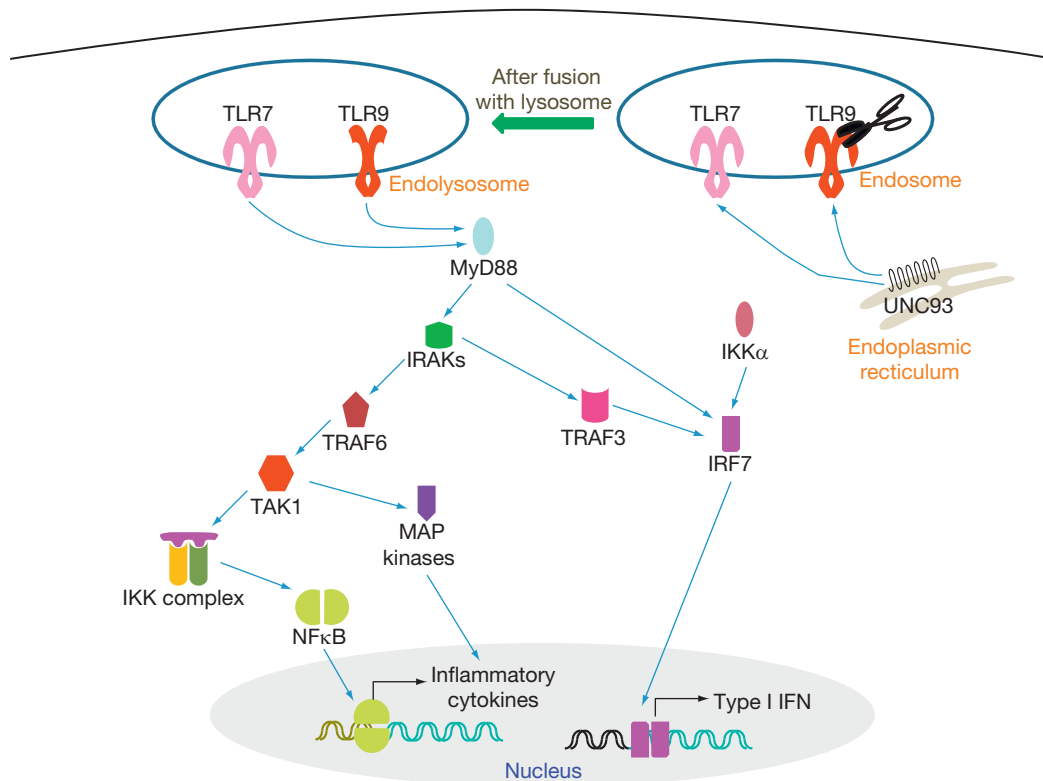


Figure 2 TLR signaling in plasmacytoid dendritic cells. Stimulation of pDCs with TLR7 and TLR9 PAMP causes transportation of PAMP to the endosome (cognate TLR7 and TLR9 were translocated to the endosome by transporter protein UNC93B1). Endosome containing cognate TLR9 undergoes cleavage (activation) after fusion with lysosome to form endolysosome. The cleaved form of TLR9 recognizes PAMP which activates MyD88-dependent signaling pathway for the induction of inflammatory cytokine (same as explained in [Figure 1](#)). For induction of type I interferons, MyD88 associates with IRF7 via the IRAK family of proteins and TRAF3. IKK α also activates IRF7. The activated IRF7 is translocated to the nucleus to induce transcription of type I interferons.

antagonists are in phase I clinical trial. Poly IC (a TLR3 PAMP) and CpG-ODN (a TLR9 PAMP) are also known to be used as a vaccine adjuvant which induces proinflammatory cytokines and type I interferons and also activates strong T-cell responses. Furthermore, these vaccine adjuvants were also proved to be effective in cancer immunotherapy. Imiquimoid and isatoribine, TLR7 agonists are beneficial for the treatment of papilloma-induced genital wart and HCV infection, respectively.

Future Perspective

Incredible progress has been made after the discovery of the TLR in understanding structures of several TLR PAMP–TLR complexes, TLR signaling pathway, as well as regulation of TLR-mediated immune responses. However, it is still unclear that how various innate receptors collaborate with each other and execute complex immune responses via activation of various transcription factors to defend array of infectious diseases. Furthermore, additional studies are required to understand pathogen and TLR interactions in dynamic *in vivo* condition and manipulation of TLR signaling pathways by small molecules to treat various infectious and autoimmune diseases.

See also: [Cell Architecture and Function: Phagocytosis; Signaling: Interferon Receptors; Nuclear Factor Kappa B.](#)

Further Reading

- Bustamante J, Boisson-Dupuis S, Jouanguy E, et al. (2008) Novel primary immunodeficiencies revealed by the investigation of paediatric infectious diseases. *Current Opinion in Immunology* 20: 39–48.
- Carpenter S and O'Neill LA (2009) Recent insights into the structure of Toll-like receptors and post-translational modifications of their associated signalling proteins. *Biochemistry Journal* 422: 1–10.
- Coban C, Ishii KJ, and Akira S (2009) Immune interventions of human diseases through Toll-like receptors. *Advances in Experimental Medicine and Biology* 655: 63–80.
- Ishii KJ, Koyama S, Nakagawa A, Coban C, and Akira S (2008) Host innate immune receptors and beyond: Making sense of microbial infections. *Cell Host and Microbe* 3: 352–363.
- Iwasaki A and Medzhitov R (2010) Regulation of adaptive immunity by the innate immune system. *Science* 327: 291–295.
- Jin MS and Lee JO (2008) Structures of TLR–ligand complexes. *Current Opinion in Immunology* 20: 414–419.
- Jin MS and Lee JO (2008) Structures of the Toll-like receptor family and its ligand complexes. *Immunity* 29: 182–191.
- Kawai T and Akira S (2009) The roles of TLRs, RLRs and NLRs in pathogen recognition. *International Immunology* 21: 317–337.

- Kawai T and Akira S (2010) The role of pattern-recognition receptors in innate immunity: Update on Toll-like receptors. *Nature Immunology* 11: 373–384.
- Krishnan J, Lee G, and Choi S (2009) Drugs targeting Toll-like receptors. *Archives of Pharmacol Research* 32: 1485–1502.
- Kumagai Y and Akira S (2010) Identification and functions of pattern-recognition receptors. *Journal of Allergy and Clinical Immunology* 125: 985–992.
- Kumar H, Kawai T, and Akira S (2009) Pathogen recognition in the innate immune response. *Biochemistry Journal* 420: 1–16.
- Kumar H, Kawai T, and Akira S (2009) Toll-like receptors and innate immunity. *Biochemical and Biophysical Research Communications* 388: 621–625.
- Takeuchi O and Akira S (2010) Pattern recognition receptors and inflammation. *Cell* 140: 805–820.
- van de Vosse E, van Dissel JT, and Ottenhoff TH (2009) Genetic deficiencies of innate immune signalling in human infectious disease. *Lancet Infectious Diseases* 9: 688–698.

Transcription-Coupled DNA Repair Overview

S Tornaletti, University of Florida, Gainesville, FL, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Isabel Mellon, volume 4, pp. 204–208, © 2004, Elsevier Inc.

Glossary

Cyclobutane pyrimidine dimer The covalent linkage of two adjacent pyrimidines in DNA produced by exposure to ultraviolet light.

Glycosylase An enzyme that cleaves the N-glycosylic bond between a damaged base or inappropriate base and the deoxyribose sugar.

Transcription bubble The unwound DNA structure produced by RNA polymerase in the elongation mode of RNA synthesis.

Nucleotide Excision Repair and Its Subpathways

Several DNA repair pathways have evolved in all organisms to prevent genomic instability generated by persistence of DNA damage. Nucleotide excision DNA repair (NER) is the most versatile excision repair pathway, which recognizes and removes a wide variety of structurally unrelated lesions from DNA, including sunlight-induced cyclobutane pyrimidine dimers (CPDs) and 6–4 photoproducts, cigarette-smoke-induced benzo[a]pyrene DNA adducts, and lesions formed by chemotherapeutic agents such as cisplatin. DNA lesions repaired by NER share the common feature of causing DNA distortions. NER is conserved throughout evolution. In *Escherichia coli*, six proteins are sufficient to carry out NER, while in mammalian cells more than 30 are required. NER involves several steps, including damage recognition, unwinding of the DNA at the site of the lesion, dual incision 5' and 3' to the damage containing strand, removal of a short oligonucleotide containing the lesion, DNA synthesis using, as a template, the complementary undamaged strand, and ligation. Deficiencies in NER result in rare autosomal recessive human disorders such as xeroderma pigmentosum (XP), Cockayne syndrome (CS), UV-sensitive syndrome, and trichothiodystrophy.

NER can be divided into two sub-pathways: global genomic repair (GGR) that recognizes and repairs lesions throughout the genome and transcription-coupled repair (TCR) that specifically operates on lesions located in transcribed regions of the genome (Figure 1).

DNA Damage Recognition in TCR and GGR

GGR and TCR differ in their mode of damage recognition; all subsequent steps are common to both repair pathways. In human cells, the heterodimer XPC/HR23B (XPC, xeroderma pigmentosum group C) appears to be the major damage recognition factor in GGR. The DNA damage-binding protein (DDB) is additionally required for GGR of CPDs, the most common DNA photoproduct generated by exposure to sunlight, so that XPC can be more efficiently recruited to the damaged site. A recently reported crystal structure of a CPD-containing DNA bound to yeast Rad4, a homolog of human XPC, has shed some light on the DNA damage recognition step in GGR. It was shown that both the CPD and the complementary undamaged nucleotides on the opposite DNA strand

are flipped out of the DNA helix. Furthermore, that the CPD lesion is not in contact with the Rad4 protein, which instead recognizes the two flipped-out nucleotides opposite the CPD in the undamaged DNA strand. These findings suggest that both the localized structural change of the double helix at the site of DNA damage and the flipping out of the two base pairs (2 bp) opposite the lesion by the DNA damage recognition protein Rad4 are two key elements in the mechanism of damage recognition in GGR.

In TCR, the initial step of damage recognition is represented by the arrest of RNA polymerase at the site of DNA damage. This is followed by recruitment of TCR-specific factors at the site of arrest and subsequent removal of the lesion by NER enzymes. DNA lesions repaired by TCR include sunlight-induced CPDs and (6–4)PPs, cisplatin DNA cross-links, and DNA adducts formed by benzopyrene-diol-epoxide (Table 1).

Transcription-Coupled DNA Repair

Historically, the initial observation was that UV-induced CPDs in an expressed gene in Chinese hamster ovary (CHO) cells were much more efficiently repaired than those in a silent sequence downstream. Then it was shown that this efficient repair was due to the preferential repair of CPD in the transcribed DNA strand.

For some lesions, like CPDs, TCR results in more rapid repair of the transcribed strands compared to the nontranscribed strands of expressed genes. This strand bias has usually been taken as the operational definition of TCR. However, for other lesions, like the more distorting (6–4) photoproducts, which are very efficiently repaired by GGR, the effect of TCR may not be seen as a differential rate of repair between the two DNA strands.

Several observations indicate that initiation of TCR requires that RNA polymerase is in elongation mode. For example, TCR of the *lac operon* of *E. coli* can only be detected when the operon is induced. In mammalian cells, treatment with α -amanitin, a specific inhibitor of eukaryotic RNA polymerase II (RNAP II), abolishes the preferential repair of CPDs in expressed genes. In yeast, with temperature sensitive mutations in the gene encoding a subunit of RNAP II, TCR is eliminated at the nonpermissive temperature.

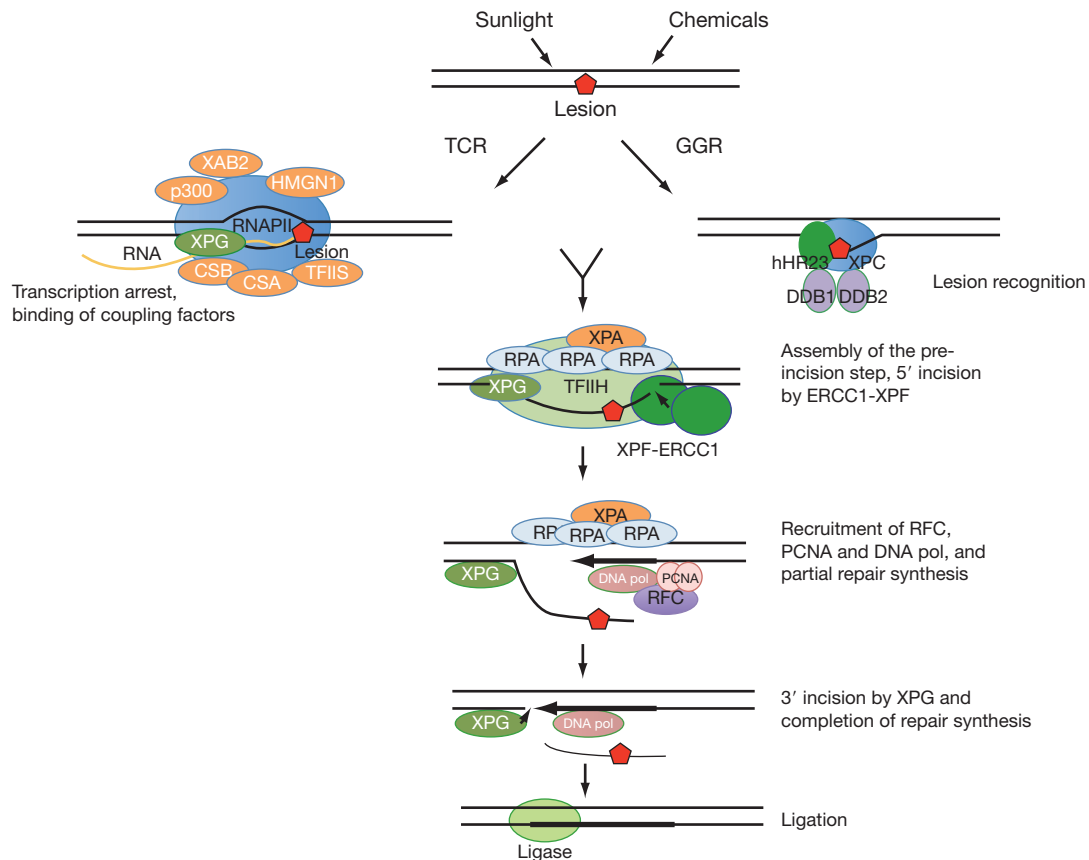


Figure 1 Nucleotide excision DNA repair and its subpathways. Nucleotide excision repair is the most versatile DNA repair pathway, dedicated to removal of several different structurally unrelated DNA lesions generated by environmental exposure. Two sub-pathways of nucleotide excision repair have been described: global genome repair (GGR), which removes DNA lesions throughout the genome, and transcription-coupled DNA repair, specific for transcription blocking-DNA lesions located in transcribed strands of expressed genes. GGR and TCR only differ in their mode of damage recognition, all subsequent steps are common to both repair pathways.

TCR does not vary in efficiency, through the normal cell cycle in mammalian cells, for a gene that is expressed continuously throughout the cycle. It is maintained in several differentiated cell types where the GGR pathway is markedly attenuated. TCR does not appear to be inducible by UV. However, in human cells, efficient global genomic repair of CPDs requires activation of the p53 tumor suppressor. Furthermore, growing evidence indicates that arrested transcription (e. g., at CPDs) provides a sensitive signal for the activation and stabilization of p53.

Although TCR was originally documented for DNA damage induced by sunlight, it has also been reported that oxidative damage can be repaired in a transcription-dependent manner, although this conclusion has been the subject of significant debate in the repair field. Recent evidence suggests that it is the presence of RNAP II complexes arrested at abasic sites in transcribed strands that may trigger TCR through recruitment of the NER machinery at the site of transcription arrest.

RNA Polymerase Arrest and Initiation of TCR

The original and still current TCR model proposes that RNA polymerase arrested at a lesion directs repair enzymes to the

transcribed strand of an active gene. This model assumes that the polymerase must be removed from the site of the lesion to provide access of the repair complex to the lesion site and to allow reannealing of the DNA strands to form a proper substrate for repair. In *E. coli*, the *mfd* gene product, encoding a 130-kDa monomeric protein, participates in this process. Using an *in vitro* system with purified proteins, it was shown that the Mfd protein can promote the release of the RNA polymerase and of the incomplete transcript from the DNA template and target components of NER to the site of transcription blockage. Recent evidence has shown that this is accomplished by a forward movement of the polymerase promoted by Mfd in an ATP-dependent fashion until the polymerase dissociates from the DNA. Forward movement is accompanied by a progressive rewinding of the upstream end of the single-stranded-transcription bubble within the transcription complex until the bubble is sufficiently rewound to release the RNAP and the nascent transcript from the DNA, making the lesion accessible for repair.

In mammalian cells, the *ERCC6/CSB* gene product may be such a coupling factor, although the reactions are probably more complex than those in *E. coli*. The cockayne syndrome group B (CSB) gene encodes a 168-kDa protein that is related

Table 1 Correlation between transcription-coupled DNA repair of a lesion *in vivo* and RNA polymerase arrest *in vitro*

DNA lesion	TCR	Transcription arrest	RNA polymerase
Cyclobutane dimer	+	+	T7, <i>E. coli</i> , yeast, rat liver, human
(6–4)Photoproduct	+	+	Human
Cisplatin	+	+	T7, <i>E. coli</i> , yeast, wheat germ
intranstrand cross-link		+	Rat liver
		+	Human
Psoralen monoadduct	–	+	T7, human
Psoralen interstrand cross-link	+	+	T7, <i>E. coli</i> , human
Benzopyrene-diol-epoxide	+	+	T7, human
Acetylaminofluorene	+	+	T7, rat liver
8-Oxoguanine	±	±	T7, <i>E. coli</i> , rat liver, mouse
Thymine glycol	–	–	Rat liver, human
Abasic sites	ND	+	T7, rat liver
2-Ribonolactone	ND	+	T3, SP6, rat liver

ND, not defined.

to the SWI/SNF family of ATP-dependent chromatin remodeling factors. SWI/SNF family members such as SWI2/SNF2 are DNA helicases. The CSB protein, like most members of this family, displays DNA-dependent ATPase and DNA-binding activity, but not helicase activity. Similarly, both the bacterial and yeast counterparts of CSB, Mfd and Rad26, are also DNA-dependent ATPases. CSB has also nucleosome remodeling activity and binds to core histone proteins *in vitro*. Furthermore, when added to RNA polymerase arrested at a CPD, CSB can stimulate transcription elongation by addition of one nucleotide (1 nt) to the nascent transcript. It remains unclear whether the polymerase is ubiquitinated and degraded, translocated away from the site of damage without dissociating from the template, or simply stalled at the site of damage while repair of the lesion occurs, as recently suggested. Once the template is repaired, the arrested RNA polymerase might then resume elongation without releasing the incomplete transcript.

Following the initiation of transcription, the template DNA strand is preferentially repaired in the region just upstream of the point at which RNAP II clears the promoter and releases TFIIF in normal cells, in CSA and CSB cells, and in the yeast homolog of CSB. Preferential repair of the transcribed strand beyond the point of TFIIF release requires CSA and CSB or *rad26*, in mammalian cells or yeast, respectively. This suggests that CSB and *rad26* might have a role in re-recruiting TFIIF to the repair complex when the RNAP II is arrested at a lesion while in the elongation mode. In support of this model, it was shown that the RNAP II/CSB complex can interact with subunits of TFIIF *in vitro*. Utilizing another *in vitro* system consisting of an artificial transcription bubble-like structure, it was found that CSB and xeroderma pigmentosum group G (XPG) can bind *in vitro* in a cooperative manner to RNAP II arrested at a cisplatin intrastrand cross-link. Furthermore, the lesion

became accessible to XPG incision only in the presence of wild-type TFIIF and required ATP hydrolysis. It was speculated that ATP hydrolysis by TFIIF would promote a conformational change of the RNA polymerase that would render the damaged DNA accessible for incision by XPG. *In vitro* reconstituted systems were very informative to determine the order of assembly of NER factors at arrested RNAP II complexes. These studies have shown that TFIIF is recruited first and that replication protein A (RPA) binds as soon as single-stranded DNA is formed. Xeroderma pigmentosum group A (XPA) is then recruited to the damage site followed by XPG and xeroderma pigmentosum group F (XPF). In this reconstituted system, CSB appeared to be required for the incision activity. *In vivo* studies revealed that following UV irradiation, RNAP II complexes arrested at UV damage sites are associated with CSB. In agreement with current TCR models, the presence of CSB was also necessary to attract NER factors specific for TCR at sites of RNAP II arrest.

The CSA gene product, which is also required for TCR, is a member of the WD-40 repeat family of proteins that is implicated in protein–protein interactions, but its role in TCR is not known. It was recently shown that CSA is part of an E3-ubiquitin ligase complex consisting of DDB1, Cullin4A, ROC1/Rbx1, and the COP9 signalosome (CSN). The association of the CSA-DDB1/CSN complex with the UV-stalled RNAP complex is dependent on functional CSB protein. Recent evidence indicates that the CSA complex is responsible for CSB ubiquitination and degradation following UV irradiation, establishing a functional link between CSA and CSB proteins.

Two other factors, XAB2 and HMG1, have also been identified and play a role in TCR. XAB2 is a tetratricopeptide repeat (TPR)-containing protein involved in pre-mRNA splicing and transcription and might function as a scaffold for protein complex formation in TCR. HMG1 is a nucleosome-binding protein that was recently shown to interact with UV-stalled RNAP II complexes in a CS-dependent manner and may be required for nucleosome removal behind the arrested polymerase so that the RNAP II can translocate backward from the lesion.

Transcription Arrest at DNA Lesions

To understand the potential roles of the various gene products implicated in the initial steps of TCR, several studies have characterized the properties of the transcription complex when it encounters a lesion in its path. The analysis of different types of arrested complexes has helped to understand how an arrested RNA polymerase may signal the repair proteins to initiate a repair event. In support of the original TCR model proposing that RNAP II arrest at a lesion is a prerequisite for initiation of TCR, a strong correlation between transcription arrest by a lesion *in vitro* and TCR of that lesion *in vivo* has been found in most cases analyzed (Table 1).

Cyclobutane Pyrimidine Dimers

CPDs are the most abundant DNA lesions produced after DNA exposure to ultraviolet light. TCR of CPDs was first

documented in Phil Hanawalt's laboratory in UV-irradiated CHO cells, followed by the finding that the efficient repair of CPDs was due to the preferential repair in the transcribed strand. Using a reconstituted *in vitro* transcription system with RNAP II and initiation factors, it was shown that a single CPD in the transcribed strand is a complete block to mammalian RNAP II progression. In addition, the arrested complex was very stable and could be translocated from the site of damage through the transcript cleavage reaction mediated by elongation factor TFIIS, thus restoring the association of the 3' end of the transcript with the catalytic site of the polymerase after arrest. This result supports a model of TCR in which the backward translocation of the polymerase from the site of arrest may be required for subsequent access to the lesion by repair proteins. In further support of this model, chromatin immunoprecipitation studies have recently shown that, following UV irradiation, TFIIS is associated with the arrested RNAP II.

With the resolution of the crystal structure of *Saccharomyces cerevisiae* RNAP II arrested at a T<>T CPD located in the template strand, new details on the mechanism of transcription arrest at a CPD have been revealed. These studies have shown that RNAP II can correctly incorporate adenosine monophosphate (AMP) opposite the 3' thymine of the CPD, but incorporation of the next nucleotide, opposite the 5' thymine of the CPD is a very slow process, that can only occur if uridine monophosphate (UMP) is misincorporated. This, in turn, causes RNA polymerase to become arrested. Arrest at a CPD is the result of a weak RNA:DNA hybrid formed at the 3' end of the transcript. A stable interaction in the last base pair is necessary to properly orient the transcript 3' end with the incoming nucleotide to form a phosphodiester linkage with the nascent transcript. If the hybrid is weak, bond formation is delayed, allowing the polymerase time to translocate upstream on the template. It is likely that the structural change in the double helix caused by formation of a CPD, consisting of unwinding by $\sim 15^\circ$ and bending of at least 7° relative to the B form may be sufficient to affect formation of the RNA:DNA hybrid, and this in turn may shift the equilibrium from nucleotide addition toward arrest.

Cisplatin-Induced Intrastrand Cross-Links

cis-Diamminedichloroplatinum(II) (cisplatin) is a widely used antitumor drug that preferentially reacts with purine bases to form DNA cross-links. These adducts may impose a more serious problem for an elongating RNA polymerase compared to a CPD, since they cause substantial unwinding and bending of the DNA helix.

It was shown that *cis*-Pt-induced intrastrand cross-links are absolute blocks to RNA polymerase progression. The arrested RNAP II complex is stable, as indicated by the ability of elongation factor TFIIS to induce transcript cleavage. It is likely that the structural changes induced by the *cis*-Pt-induced intrastrand cross-links would affect the formation and/or the stability of the RNA:DNA hybrid, an essential component of the elongation complex.

More insights into the mechanism of RNA polymerase arrest at *cis*-Pt-induced intrastrand cross-link were obtained with the resolution of the crystal structure of a *S. cerevisiae*

RNAP II arrested at this lesion. Different from the mechanism of transcription arrest at a CPD, a *cis*-Pt-induced intrastrand cross-link blocks RNAP transcription because of a translocation barrier that prevents the adduct to enter the catalytic site. This occurs because the two adjacent bases in the *cis*-Pt-induced intrastrand cross-link are covalently linked and, as a result, cannot twist by approximately 90° as required for the incoming base to enter the polymerase active site.

Oxidative DNA Damage

Several *in vitro* studies aimed at characterizing the behavior of 8-oxoG and thymine glycol (Tg) on RNAP transcription have shown various levels of RNAP bypass at these lesions. Elongation factors have been recently shown to have a significant role in modulating readthrough past these lesions, suggesting that the differences in the *in vitro* transcription results that are previously observed may have derived from differences in the cell extract used, and/or on the presence of elongation factors among components of the transcription system. Elongin and CSB, which affect the rate of RNAP II elongation, but not TFIIS, which help arrested complexes to bypass transcriptional blocks, were shown to stimulate bypass of Tg, whereas Elongin, CSB, and TFIIS could all enhance readthrough of an 8-oxoG lesion. Furthermore, transcription past these lesions caused misincorporation opposite the lesion, suggesting that elongation factors, by promoting readthrough past 8-oxoG or Tg, could contribute to the transcriptional mutagenesis induced by these lesions.

If 8-oxoG and Tg do not represent a significant barrier to RNAP transcription *in vitro*, how can they elicit TCR *in vivo*? It was initially proposed that RNAP pausing at 8-oxoG could be sufficient to initiate TCR. Another possibility is that intermediates in the repair of oxidative damage such as natural abasic sites may be sufficient to elicit TCR. *In vitro* studies aimed at addressing this question have shown that abasic sites, 2-deoxyribonolactone, an oxidized form of the apurinic/apyrimidinic (AP) site, and its caged precursor, when located in the transcribed strand of template DNA, were strong blocks to T7 RNAP and mammalian RNAP II. Based on the current TCR model postulating that only lesions that block RNAP will be subject to TCR, these findings suggest that the abasic site and its oxidative derivative may be sufficient to initiate TCR *in vivo*. In agreement with this prediction, recent findings indicate that AP sites in transcribed strands trigger TCR *in vivo*.

Perspective

After more than 20 years since the discovery of transcription-coupled DNA repair, we are starting to have a better understanding of the steps leading from arrest of transcription at the site of damage to repair of the lesion and transcription restart. However, we still do not know how the coupling between transcription arrest and initiation of TCR occurs. In addition, it is unclear what determines whether a lesion will block transcription or which features of an RNAP arrested at a lesion make it a substrate for repair with respect to a polymerase arrested at a natural arrest site or at a noncanonical DNA structure.

The development of novel approaches for studying TCR will help to address some of these challenging questions.

See also: **Molecular Biology:** DNA Base Excision Repair Pathways; DNA Damage: Alkylation; DNA Glycosylases: Mechanisms; DNA Oxidation; Gene Expression in Eukaryotes: RNA Polymerase II Structure; Nucleotide Excision Repair, Bacterial: The UvrABCD System; Nucleotide Excision Repair: Biology; RNA Polymerase I and RNA Polymerase III in Eukaryotes; RNA Polymerase Structure, Bacterial; T7 RNA Polymerase.

Further Reading

- Brueckner F, Hennecke U, Carell T, and Cramer P (2007) CPD damage recognition by transcribing RNA polymerase II. *Science* 315: 859–862.
- Fousteri M, Vermeulen W, van Zeeland AA, and Mullenders LH (2006) Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II *in vivo*. *Molecular Cell* 23: 471–482.
- Friedberg EC, Walker GC, Siede W, et al. (2006) *DNA Damage and Mutagenesis*. Washington, DC: ASM Press.
- Hanawalt PC and Spivak G (2008) Transcription-coupled DNA repair: Two decades of progress and surprises. *Nature Reviews Molecular Cell Biology* 9: 958–970.
- Mellon I, Spivak G, and Hanawalt PC (1987) Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* 51: 241–249.
- Min JH and Pavletich NP (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* 449: 570–575.
- Nouspikel T (2009) DNA repair in mammalian cells: Nucleotide excision repair: Variations on versatility. *Cellular and Molecular Life Sciences* 66: 994–1009.
- Park J-S, Marr MT, and Roberts J (2002) *E. coli* transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell* 109: 757–767.
- Sarker AH, Tsutakawa SE, Kostek S, et al. (2005) Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH: Insights for transcription-coupled repair and Cockayne syndrome. *Molecular Cell* 20: 187–198.
- Savery N (2007) The molecular mechanism of transcription-coupled repair. *Trends in Microbiology* 15: 326–333.
- Saxowsky TT and Doetsch PW (2006) RNA polymerase encounters with DNA damage: Transcription-coupled repair or transcriptional mutagenesis? *Chemical Reviews* 106: 474–488.
- Scicchitano D (2005) Transcription past DNA adducts derived from polycyclic aromatic hydrocarbons. *Mutation Research* 577: 146–154.
- Svejstrup JQ (2007) Contending with transcriptional arrest during RNAP II transcript elongation. *Trends in Biochemical Sciences* 32: 165–171.
- Tornaletti S (2009) Transcription-coupled DNA repair: Directing your effort where it's most needed. *Cellular and Molecular Life Sciences* 66: 1010–1020.
- Tornaletti S and Hanawalt PC (1999) Effect of DNA lesions on transcription elongation. *Biochimie* 81: 139–146.

Transcription Termination

T J Santangelo, Ohio State University, Columbus, OH, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Backtracking Reverse translocation of RNAP along the DNA template. The RNA is re-threaded through the enzyme with the 3' end emerging into the secondary channel and an internal phosphodiester linkage in the active site.

Polarity Loss of transcription of a downstream gene in an operon due to transcription termination within an upstream gene in the same operon; transcription termination in the upstream gene is usually a result of failed or stalled translation, allowing access of Rho to the RNA and RNAP.

RNA/DNA hybrid An RNA and a DNA strand paired in an A-form 8 or 9 bp double helix with the main channel of RNA polymerase.

Termination Release of the nascent transcript from RNA polymerase; it is unclear if transcript release and enzyme dissociation are necessarily coupled.

Transcription elongation complex An active complex of RNA polymerase, its nascent transcript, and the associated DNA template.

Introduction and Background

RNA synthesis by DNA-dependent RNA polymerases (RNAPs) is processive, requiring a single enzyme molecule to transcribe the full length of a gene regardless of the length. The requirement for RNAP to remain resolutely associated with the DNA template through multiple kilobases necessitates an extremely stable transcription elongation complex that can transcribe through different sequences and protein-bound DNA templates. Despite this stability, cells must be able to halt RNA synthesis after transcription of a complete gene or operon, and stop any RNAP that has initiated transcription aberrantly. Failure to terminate transcription of an upstream gene could allow regulation-independent expression of downstream genes, and synthesis of untranslated or antisense transcripts with detrimental consequences; aberrant transcription is particularly problematic for the gene-dense chromosomes common to Bacteria and Archaea. Two general mechanisms have evolved to efficiently disrupt transcription elongation complexes that release the RNA transcript and recycle RNAP for further rounds of transcription.

Multi-subunit RNAPs from each domain share a near identical core structure that envelopes an 8- or 9-bp RNA:DNA hybrid within a tight-fitting pocket (Figure 1). High-resolution crystal structures and a wealth of biochemical data from many different RNAPs demonstrate that hydrogen bonding within the hybrid and contacts between the enclosed nucleic acids and RNAP provide stability to transcription elongation complexes. Despite similar transcription elongation complex architecture, RNAPs from different domains, and each of the eukaryotic RNAPs, respond to different termination signals and factors, suggesting that several mechanisms of transcription termination are possible, or that a diverse set of factors and sequences use a common mechanism to disrupt the complex. Conserved elongation factors (i.e., NusG and NusA) modify RNAP activities and add an additional level of regulation to the elongation–termination decision. The mechanistic details of transcript release are best understood in Bacteria, although some features are shared in each domain. This article focuses on transcription termination and its regulation in

Bacteria, with relevant comments to bring attention to similarities and differences in Archaea and Eukarya.

Bacterial Transcription Termination

Two general mechanisms of transcription termination have been characterized in both biochemical reactions with purified components and through genetic experiments. The first, termed 'intrinsic termination', requires only defined template sequences and the resultant RNA transcripts to disrupt the transcription elongation complex. The second, factor-dependent termination, requires the activity of an additional protein that interacts with the transcription elongation complex.

Intrinsic Termination

Intrinsic terminators are composed of two essential elements: (1) a ~10–20 bp region of dyad symmetry in the DNA which is transcribed into an RNA sequence capable of forming an RNA–RNA duplex or RNA hairpin structure and (2) a T-rich sequence that directs the synthesis of a uridine-rich sequence in the RNA immediately following the RNA hairpin (Figure 2). Base pairing within the stem of the hairpin, the presence of the uridine-rich segment, and the spacing of these elements are critical to termination efficiency; flanking sequences can also influence the efficacy of specific terminators. Transcript release occurs 8–10 bp downstream of the 3' base of the hairpin, typically near the end of the uridine-rich sequence.

A combination of factors leads to transcript release, although the mechanics of RNA release may occur via several mechanisms. The transcription elongation complex is weakened by sequence context alone; the hybrid at the normal position of termination is predominantly rU:dA, providing only the weakest hydrogen bonding of all possible combinations of nucleic acids. The uridine-rich sequence serves a second role, to pause RNAP at the position of termination, and any sufficiently uridine-rich sequence has been shown to direct RNAP to pause. The paused, weakened transcription elongation complex is now positioned immediately adjacent and downstream of the region capable of

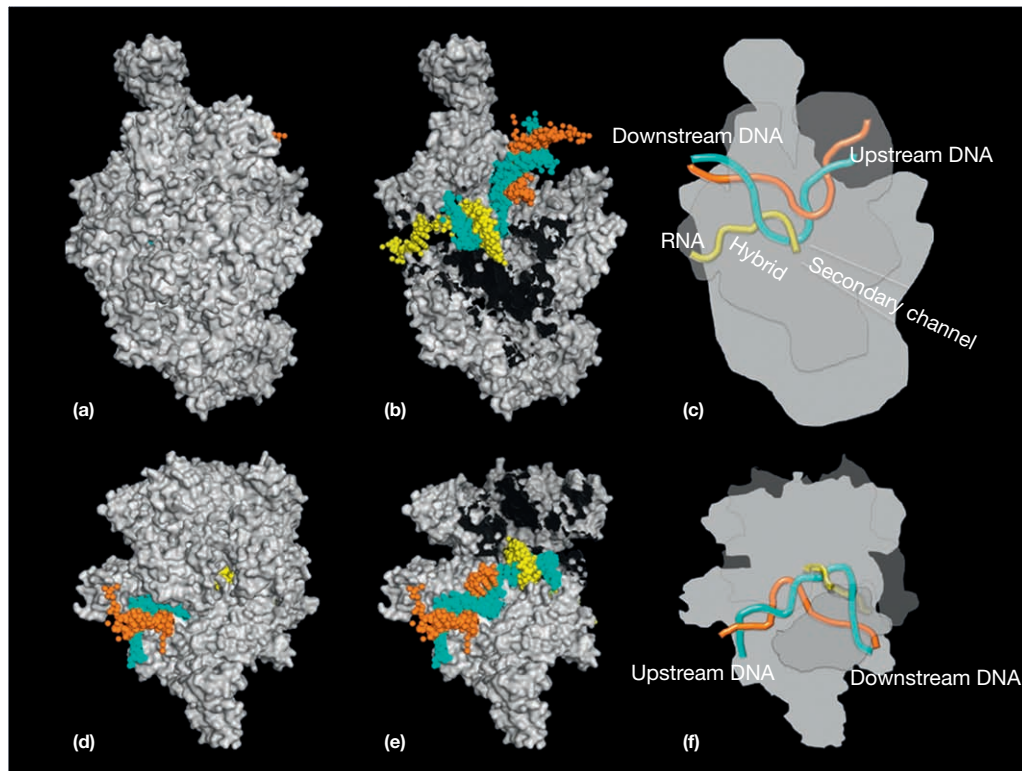


Figure 1 The bacterial transcription elongation complex. Upper panels (top view), with RNAP movement from left to right. Lower panels (front view), with RNAP movement right to left. RNAP (gray), RNA (yellow), template strand (cyan), nontemplate strand (orange). Panels (a)/(d) – surface representation of the bacterial transcription elongation complex. The encapsulated nucleic acids are fully enclosed within the complex. Panels (b)/(e) – as (a)/(d), respectively, with one RNAP subunit (b) removed to reveal the interior of the elongation complex and the path of the enveloped nucleic acids. Panels (c)/(f) – cartoon depiction of the bacterial transcription elongation complex. Sections of RNAP are shown partially transparent to show the hidden nucleic acids.



Figure 2 An intrinsic terminator. The DNA sequence of the t_{82} intrinsic terminator is shown in black, with regions of dyad symmetry shown in bold and underscored by arrows. Asterisks mark the known positions of RNA release. The resultant RNA is shown in red, with base-pairing indicated, and the known 3' terminal bases shown in bold.

hairpin formation, and RNAP interaction with the RNA hairpin results in transcript release.

The role of hairpin formation and its effects on the elongation complex are debated, although most models predict that hairpin formation disrupts the upstream bases of the hybrid, triggering collapse of the transcription elongation complex and

RNA release. RNAP can be pushed forward by hairpin formation, and in the absence of continued synthesis, such movement unwinds the hybrid and eliminates RNAP-hybrid interactions necessary for complex stability. Hairpin formation has also been shown to direct changes in RNAP structure. Allosteric models of transcription termination predict relatively

large structural changes in RNAP that reduce the stability of the complex, hinder further synthesis, and allow dissociation of the complex.

Intrinsic termination also occurs in the Archaea and Eukarya, although this differs in detail from bacterial intrinsic termination. Archaeal RNAPs and eukaryotic RNA polymerase I or III (Pol I or Pol III) do not require an RNA hairpin for intrinsic termination, although each enzyme does terminate transcription, *in vivo* and *in vitro*, in response to oligoT-rich template sequences. For these RNAPs the exact sequence of the oligo-T-rich region, its flanking sequences, and general sequence context dictate termination position and efficiency. Evidence suggests that oligo-T-rich sequences first evolved as termination sequences, and remain so for many RNAPs, with RNA hairpins adding specificity and precision to bacterial intrinsic termination. Eukaryotic RNAP polymerase II (Pol II) is notably resistant to bacterial intrinsic termination sequences and appears completely dependent on auxiliary proteins for transcription termination.

Factor-Dependent Termination

Many genes and operons lack easily identifiable intrinsic termination sequences. Genetic and biochemical evidence demonstrates that termination of their transcription relies on proteins that disrupt the transcription elongation complex. The activity of these factors is also generally employed for all transcription units, regardless of whether an intrinsic termination signal is also present.

Rho-mediated termination

Rho is an RNA-dependent ATPase encoded in most bacterial genomes, but not in Archaea or Eukarya. Rho is a surveillance factor that monitors the normal coupling of bacterial transcription and translation. Rho accesses transcription elongation complexes by binding to ribosome-free RNA, wrapping ~60 nts of RNA around a ring-shaped hexameric structure. Since RNAP elongates at ~50–100 nucleotides per second, Rho provides a rapid response to the uncoupling of transcription and translation. Its mechanism of action is to translocate along the RNA and physically disrupt the elongation complex. Rho normally targets transcription elongation complexes that have reached the end of genes or operons, and thus are uncoupled from translation via ribosome release at the stop codon, but Rho can disrupt any elongation complex that is not immediately and consistently trailed by a ribosome, and is therefore responsible for polarity.

The major function of Rho may be to suppress futile transcription of host genes and limit expression of any introduced phage or foreign genes. Rho-dependent termination also is the natural mechanism of release at certain sites, likely serving as the sole mechanism to terminate transcription of many operons. Properties that favor Rho action are the absence of translation, a cytidine-rich transcript, and the absence of secondary structure. These characteristics are sufficiently common for Rho to engage any elongation complex that is not closely followed by translating ribosome. Release of a nascent transcript is, in principle, possible at all promoter-distal locations, but release is normally only seen at sites that direct RNAP to pause, likely allowing Rho to catch up to the elongation complex. Strong

secondary structures in the RNA can impede Rho binding and activity. Rho action is stimulated by NusG and limited by NusA (see below).

The accepted mechanism of Rho activity invokes movement along the RNA until RNAP is reached, then pulling the RNA out of the elongation complex; alternatively, RNAP may be pushed forward, or allosterically modified to release the transcript. Recent work suggests that Rho may interact with RNAP throughout the transcription cycle *in vitro*, and thus be poised to halt transcription immediately in the absence of translation; stoichiometric details of this interaction are not known, nor has this proposed mechanism been evaluated *in vivo*.

The absence of Rho in eukaryotes is not surprising given the physical separation of transcription and translation. Archaea are prokaryotes with transcription and translation coupled, and decoupling leads to polarity, but the protein(s) responsible are yet to be biochemically characterized.

Mfd and transcription-coupled DNA repair

Repair of highly transcribed DNA is more rapid than noncoding or poorly transcribed DNA. Furthermore, the template strand is repaired faster than the nontemplate strand of the same gene. DNA damage on the template strand stalls RNAP, and so a single lesion in an essential gene could prove deadly. RNAP stalled at the position of the DNA lesion serves as a marker of that lesion and recruits a factor (Mfd) that initiates RNAP removal and DNA repair. Mfd is a DNA-dependent helicase that binds directly to RNAP and rewinds the DNA duplex at the upstream edge of the transcription elongation complex, injecting torque that collapses the transcription bubble, moves RNAP forward, and ultimately disrupts the elongation complex. If RNAP is only stalled or backtracked, Mfd activity can rescue the elongation complex to an active configuration. When continued elongation is blocked by a DNA lesion, the elongation complex is removed, and Mfd then recruits the UvrABC complex to excise the lesion and initiate DNA repair. Like Rho, Mfd can terminate transcription of almost any elongation complex.

Transcription-coupled DNA repair also occurs in Eukarya and requires the transcription-repair coupling factor. To date, there is no experimental evidence for transcription-coupled DNA repair in the Archaea.

Elongation Factors

NusA and NusG are nearly universally conserved in Bacteria and homologs of both are found in almost all Archaea and Eukaryotes. NusG in particular is well represented in all three domains, consistent with an early evolution and the necessary role of this general elongation factor. Both factors bind directly to RNAP, both influence the elongation rate of RNAP, alone and in combination, and both are normally essential in Bacteria. In *E. coli*, removal of remnant phage sequences from the genome can relieve this essentiality; NusG (Spt5), but not NusA, is essential in Archaea and Eukaryotes.

In a purified system, NusA generally slows elongation, enhances pausing, and increases intrinsic termination efficiency. NusA has RNA-binding domains, and may interact with RNA hairpins and enhance interactions between RNA hairpins and RNAP. In simple terms, NusG has the opposite

functions: it enhances elongation and limits pausing. It should be noted that NusG proteins from different Bacteria have been characterized with different activities in a purified system. NusA and NusG can influence RNAP activity at the same time, and although their effects on RNAP activity balance to some degree, each still exerts discrete effects on RNAP activity during elongation. NusA and NusG also influence Rho activity. Strong Rho-dependent termination sites are pause sites but the effects of NusA and NusG on Rho function are opposite to those predicted based on their effects on pausing. NusA promotes pausing but inhibits Rho function, whereas NusG inhibits pausing but promotes Rho function. RNAP pausing is only one component underlying the efficiency of Rho-dependent termination, and the interplay of NusA and Rho activity is emphasized by the ability of a NusA defective cell to survive if Rho is also defective.

Biochemical and structural information regarding the binding sites of NusG and NusA on RNAP have bolstered the role of these factors in modifying RNAP activity, and demonstrated the universal nature of these interactions in each domain. NusG and NusG-related proteins bind to RNAP across the main channel, interacting with the nontemplate DNA strand of the elongation complex, bridging the clamp and lobe domains of RNAP, and effectively enclosing the DNA within the RNAP-factor complex. This enclosure may provide additional stability to the elongation complex and promote continued elongation. NusA interacts with the C-terminal domain of the alpha subunit of bacterial RNAPs and can directly bind to the core body of RNAP near the exit channel for the nascent RNA. NusA, unlike NusG, does not appear to directly enclose or provide stabilizing interactions between RNAP and the enveloped nucleic acids.

Regulated Termination and Antitermination

Efficient transcription termination is often solely thought of as a necessity for controlled expression in compact genomes, but regulation of upstream transcription termination is often employed to regulate downstream gene expression. As with termination, antitermination can be based primarily on sequence information (intrinsic) or reliant on protein factors.

Site-Specific Antitermination: Attenuators and Riboswitches

Upstream intrinsic termination is used to regulate expression of many operons through mechanisms that alternatively permit or prevent transcription termination in a leader sequence or initially translated region (Figure 3). Alternative RNA structures (e.g., a Riboswitch or Ribosensor) in the nascent transcript dictate whether transcription continues or is terminated based on the formation or lack of formation of a terminator structure in the RNA. The RNA fold that forms is influenced in most examples by small molecules (i.e., metabolites, cations, and tRNAs) that directly bind to the RNA and affect folding. Both positive and negative transcription-based riboswitches have been described. Classical riboswitches rely on interactions with small molecules to dictate RNA structure, although examples of protein-influenced riboswitches are known.

As a term, attenuation is most often reserved for translation-coupled regulation of intrinsic termination, whereby positioning of a trailing ribosome influences formation of a downstream terminator. The classical example of such regulation is found in the *trp* operon, where stalling of the ribosome at tryptophan codons, due to limited availability of charged aminoacyl-tRNA^{Trp}, determines which sequences of the transcript can hybridize, and thus whether a terminator hairpin can be formed. Examples of riboswitches and attenuators have been described for Eukaryotes. However, although the prokaryotic Archaea appear poised to take advantage of these systems, there is no evidence for such regulation.

Termination and Antitermination Factors

Transcript-specific regulatory factors, such as the lambdaoid phage N and Q proteins, the bacterial RfaH protein, and the phage HK022 Nun protein, regulate expression of those genes by altering the termination capacity of RNAP. Specific RNA sequences within the nascent transcript can also form structures that interact with and change the termination potential of RNAP. In general, these factors are recruited to particular transcribing RNAPs via *cis*-acting nucleic acid sequences. Exact

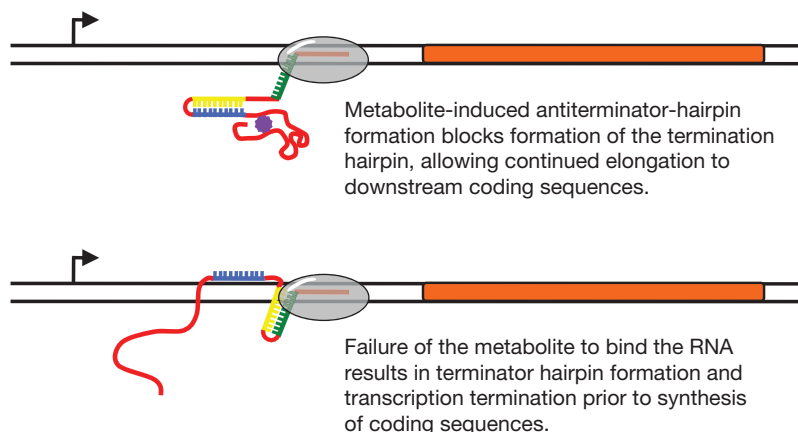


Figure 3 A termination-based Riboswitch. Both panels depict a transcription unit (arrow marks the promoter; DNA in black; RNA in red with sections of sequence highlighted in green, blue, and yellow; RNAP is shown as a gray translucent sphere) with a long leader sequence encoding a Riboswitch upstream of a gene (orange). A fold of the nascent transcript is influenced by binding, or lack thereof, of a small molecule (purple star). Hybridization of the yellow and green RNA sequences results in formation of an intrinsic terminator.

details of the mechanisms and resultant modifications to the elongation complex stability are only just beginning to be described. λ N and HK022 Nun proteins bind to essentially the same RNA structures, but have opposite effects: λ N, in combination with other host factors including NusA and NusG, interacts with RNAP to stimulate elongation through downstream terminators and promotes phage gene expression. By contrast, binding of HK022 Nun protein to this same structure blocks continued elongation, thus preventing superinfection of its host by another lambdoid phage that requires N function for growth. Nun-modified complexes are halted, but not disrupted by Nun action, and thus Nun is not a release factor. Nun-arrested complexes require the activity of the Mfd protein to efficiently release RNAP and the nascent transcript.

RfaH and λ Q proteins require *cis*-acting DNA sequences to pause RNAP at discrete sites and present unique complexes for interaction. Both factors attach to the elongation complex, becoming integral components of a now terminator-resistant complex. λ Q protein interacts with RNAP at a promoter-proximal pause and likely remodels the active site of RNAP and strengthens interactions between RNAP and the hybrid to resist the action of a terminator. RfaH, a homolog of NusG, regulates expression of select operons by interacting with RNAP facilitated by a regulatory pause site (*ops* sequence) that stimulates expression of distal genes. RfaH interactions with RNAP mimic those of NusG, providing stability to the complex by further enveloping the nucleic acids and forming a bridge across the main channel of RNAP.

See also: **Molecular Biology:** Gene Expression in Eukaryotes: RNA Polymerase II Structure; Ribozymes and Evolution; RNA Polymerase I and RNA Polymerase III in Eukaryotes; RNA Polymerase II and Its General Transcription Factors; RNA Polymerase II Elongation Control in Eukaryotes; Transcription-Coupled DNA Repair Overview.

Further Reading

Artsimovitch I and Landick R (2002) The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. *Cell* 109: 193–203.

- Burns CM, Richardson LV, and Richardson JP (1998) Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in *Escherichia coli*. *Journal of Molecular Biology* 278: 307–316.
- Cardinale CJ, Washburn RS, Tadigotla VR, et al. (2008) Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science* 320: 935–958.
- Epshtein V, Cardinale CJ, Ruckenstein AE, Borukhov S, and Nudler E (2007) An allosteric path to transcription termination. *Molecular Cell* 28: 991–1001.
- Epshtein V, Dutta D, Wade J, and Nudler E (2010) An allosteric mechanism of Rho-dependent transcription termination. *Nature* 463: 245–249.
- Gusarov I and Nudler E (1999) The mechanism of intrinsic transcription termination. *Molecular Cell* 3: 495–504.
- Klein BJ, Bose D, Baker KJ, et al. (2011) RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proceedings of the National Academy of Sciences of the United States of America* 108: 546–550.
- Komissarova N, Becker J, Solter S, Kireeva M, and Kashlev M (2002) Shortening of RNA: DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. *Molecular Cell* 10: 1151–1162.
- Nudler E and Gottesman ME (2002) Transcription termination and anti-termination in *E. coli*. *Genes Cells* 7: 755–768.
- Park JS, Marr MT, and Roberts JW (2002) *E. coli* transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell* 109: 757–767.
- Santangelo TJ, Cubonova L, Matsumi R, et al. (2008) Polarity in archaeal operon transcription in *Thermococcus kodakaraensis*. *Journal of Bacteriology* 190: 2244–2248.
- Santangelo TJ, Cubonova L, Skinner KM, and Reeve JN (2009) Archaeal intrinsic transcription termination *in vivo*. *Journal of Bacteriology* 191: 7102–7108.
- Santangelo TJ, Mooney RA, Landick R, and Roberts JW (2003) RNA polymerase mutations that impair conversion to a termination-resistant complex by Q antiterminator proteins. *Genes and Development* 17: 1281–1292.
- Santangelo TJ and Roberts JW (2004) Forward translocation is the natural pathway of RNA release at an intrinsic terminator. *Molecular Cell* 14: 117–126.
- Selby CP and Sancar A (1993) Transcription-repair coupling and mutation frequency decline. *Journal of Bacteriology* 175: 7509–7514.
- Skordalakes E and Berger JM (2003) Structure of the Rho transcription terminator: Mechanism of mRNA recognition and helicase loading. *Cell* 114: 135–146.
- Smith AM, Fuchs RT, Grundy FJ, and Henkin TM (2010) Riboswitch RNAs: Regulation of gene expression by direct monitoring of a physiological signal. *RNA Biology* 7: 104–110.
- Toulkhouonov I, Artsimovitch I, and Landick R (2001) Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science* 292: 730–733.
- Weisberg RA and Gottesman ME (1999) Processive antitermination. *Journal of Bacteriology* 181: 359–367.
- Yarnell WS and Roberts JW (1999) Mechanism of intrinsic transcription termination and antitermination. *Science* 284: 611–615.

Transient Receptor Potential Channels

I S Ambudkar, X Liu, and B Paria, National Institute of Dental and Craniofacial Research, Bethesda, MD, USA

Published by Elsevier Inc.

Glossary

Calcium channels Protein(s) that mediate passive, electrochemical gradient-driven, flux of calcium.

Ca²⁺ signaling Events that initiate elemental changes in cellular levels of Ca²⁺ which are then decoded into a physiological response.

Channelopathy Disease directly linked to channel structure and function.

Human disease Disease related to disruptions in human systems that translate into dysfunction.

Ligands Small molecule or other reagents that gate ion channels.

Neurotransmitter Chemical messengers secreted by nerve endings.

Introduction to Transient Receptor Potential Channels

The transient receptor potential (TRP) superfamily of cation channels represents a remarkable group of ion channel proteins that display a wide array of ion permeabilities and diverse activation mechanisms. TRP channels have been shown to play critical roles in the physiological function of cells as well as in pathophysiology and disease. This versatile family of ion channels respond to all major classes of external stimuli, including sensory (temperature, mechanical, and osmolarity), chemical ligands (tastants, flavors, noxious chemicals, and natural products) as well as endogenous stimuli (such as neurotransmitters and growth factors) which activate plasma membrane G-protein- as well as tyrosine kinase-coupled receptors. Some TRPs are directly activated by the stimuli, while others are activated by downstream signaling components that are generated in response to the stimuli. The TRP superfamily consists of seven subfamilies: TRPC (TRP-canonical), TRPV (TRP-vanilloid), TRPM (TRP-melastatin), TRPN (TRP-no mechanoreceptor potential C (NOMPC)) and TRPA (TRP-ankyrin), and two related channel types, TRPP (TRP-polycystin) and TRPML (TRP-mucolipin). This article discusses the major mammalian TRP channel subfamilies, TRPC, TRPV, and TRPM, with relatively more focus on the function and regulation of TRPC channels (Figure 1).

Major Features of TRP Channel Structure

TRPC channel subfamily consists of seven members, TRPC1–TRPC7 (TRPC2 is a pseudogene in humans), TRPV family of six members (TRPV1–TRPV6), while TRPM family has eight members (TRPM1–TRPM8). TRP channels were originally predicted to have six putative transmembrane (TM) domains with a pore domain between the fifth (S5) and sixth (S6) TMs. Channel topology was confirmed for several TRP channels and it has been suggested that, like other 6-TM channels, TRPs probably form a tetrameric quaternary structure, where each subunit contributes to the channel pore. Some TRP channels display voltage dependency in their gating. Furthermore, gating efficiency in many TRPs can be modulated by other factors, for example, temperature and pH exert simultaneous effects of

TRPV1 activity, while temperature and tastant together can affect TRPM5 activity. Amino acid sequences flanking the pore in TRP channels are most strongly conserved across the TRP channel family. S5, S6, and the proximal C-terminal tail (TRP domain) are strikingly similar even in disparate TRP channels. A homologous block of ~25 intracellular residues immediately C-terminal to S6 is loosely conserved in all TRP mammalian subfamilies except TRPA and TRPP and has been referred to as the TRP domain. This domain is composed of a highly conserved six-amino acid sequence (six amino acids (EWKFAR) in TRPC channels) as well as a proline-rich sequence. Recently, residues within the TRP domain were demonstrated to be required for PIP₂ binding and regulation of channel gating of TRPM8 and TRPV5 channels. However, phosphatidylinositol 4,5-bisphosphate (PIP₂) stimulates some TRP channels (TRPV5, TRPM5, TRPM7, and TRPM8) while inhibiting others (TRPV1). How exactly PIP₂ regulates these channels is not yet known. Ca²⁺ and calmodulin (CaM) also regulated a number of TRP channels. TRPCs and a number of other TRPs have one or more binding sites for CaM as well as Ca²⁺. These have been reported to be involved in activation, inactivation, as well as channel trafficking.

The TRPC and TRPV proteins contain three to four ankyrin repeats in their N-terminal domain as well as a coiled-coiled region. The former has been proposed to play a role in interaction of the channel with scaffolding and regulatory proteins, while the CC-domain might be involved in TRP channel multimerization. Consistent with the presence of putative protein-interacting domains, TRP channels are assembled in a multiprotein signaling complex. Importantly, some key Ca²⁺ signaling proteins are associated with TRPC channels, such as inositol trisphosphate receptor (IP3R), plasma membrane Ca²⁺ ATPase (PMCA), sarco/endoplasmic reticulum Ca²⁺-ATPase (SERCA), G proteins, phospholipase C (PLC), Homer, and Na(+)/H(+)-exchanger regulatory factor-1/Ezrin-radixin-moesin-binding phosphoprotein-50 (NHERF/EBP-50). The TRP-associated proteins range from scaffolding, to trafficking, and regulatory. In addition, putative protein kinase A (PKA), and C (PKC) phosphorylation sites, phosphatidylinositol 3-kinase (PI3K) and Src-homology-2 (SH2)-recognition domains (YXXM motifs) and putative PDZ interaction domains have also been identified in several TRPs. Further studies are

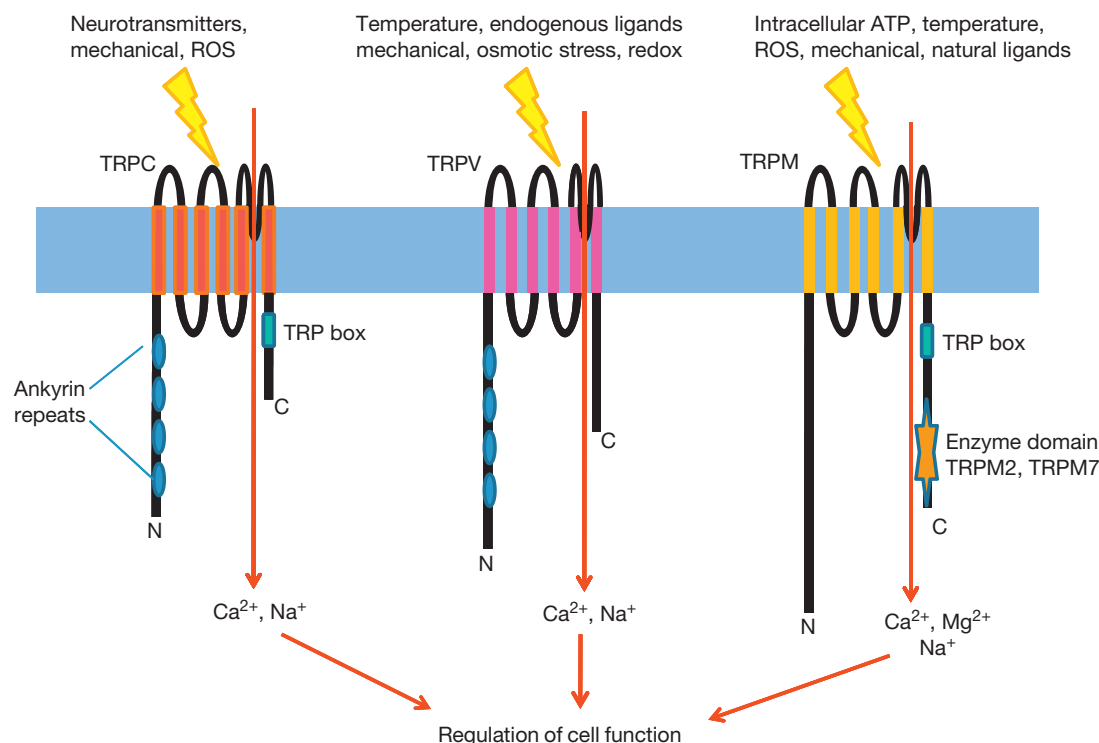


Figure 1 TRP channel structure and activation.

required to clearly understand the exact role of these domains and accessory proteins in channel function.

TRPV subfamily members share ~25% amino acid identity to other TRP proteins, mainly in the pore region. The localization of the pore loop has been confirmed by mutational analyses of residues within the pore for TRPV5 and TRPV6, TRPV1, and TRPV4. Recent studies demonstrate the importance of ANK repeats for the tetramerization of TRPV6 and TRPV5 ion channels; however, they could also mediate interaction of TRPVs with other proteins or the cytoskeleton. TRPV proteins display distinct modes of activation which result from the channel ability to respond to ambient and higher temperatures as well as osmolarity. The TRPV family has been subdivided on the basis of structure and function, TRPV1/TRPV2, TRPV3, TRPV4, and TRPV5/6. TRPV1, 2, 3, and 4 are modestly permeable for Ca^{2+} . TRPV5 and 6 are the only highly Ca^{2+} selective channels in the TRPV family. Like other TRPs, TRPVs are shown to have a number of proteins associated with them, including CaM. TRPV5 interacting protein, namely S100A10 (p11, calpactin light chain, or annexin 2 light chain), was identified. Both TRPV5 and TRPV6 contain a conserved binding site for S100A10 binding localized to the intracellular C-terminal region shortly downstream of the last TM domain (S6). Deletion of this PDZ binding motif within TRPV5 leads to greatly diminished surface expression and function.

The TRPM subfamily of ion channels comprises eight members divided into three main groups: TRPM1/3, TRPM4/5, and TRPM6/7, which regulate diverse cell functions, for example, taste detection, Mg^{2+} homeostasis, cell proliferation, and cold sensing. TRPM4 and TRPM5 are the only Ca^{2+} -impermeable TRP channels identified to date, while TRPM3, TRPM2, and

TRPM8 are Ca^{2+} -permeable cation channels with low Ca^{2+} selectivity. TRPM6 and TRPM7 are relatively highly permeable for divalent cations, including Mg^{2+} . The functional properties of TRPM1 have not been reliably characterized. TRPM channels possess low homology with other TRP subfamilies. All TRPM channels possess a TRPM-homology region which is not found in other known proteins. The biological significance of this region remains unknown. A feature unique to the TRPM family of ion channels is the presence of C-terminal enzymatic domains in three of its family members, TRPM2, TRPM6, and TRPM7. In addition, TRPM2 has a C-terminal domain with homology nucleoside diphosphate linked to X (NUDIX), an enzyme with adenosine diphosphate (ADP)-ribose hydrolase activity. TRPM6 and TRPM7 are not closely related to TRPM2 and are also enzymes, as both contain an eEF2 α -kinase family Ser/Thr kinase domain in their C-terminal region. The exact function of the enzyme domains in these channels is not yet known.

Activation of TRP Channels

Channel classification is not based on activation mechanisms but rather on sequence homology. Members of the TRPV and TRPM subfamilies are activated via temperature sensation, while mechanical or osmotic stimuli also activate some TRPVs. Some TRPs are constitutively active channels and are modulated via regulated trafficking. Other various activation mechanisms include activation of free radicals (TRPM2 and TRPC3), changes in $[\text{H}^{+}]$ (TRPVs), and changes in PIP_2 (implied for a number of TRP channels). Most similarity in

activation mechanisms is seen in the regulation of TRPCs, which are all activated in response to stimulation of PLC, which lead to PIP₂ hydrolysis. The notable feature of some TRP channels is their capacity to be activated by multiple modalities. Good examples of these are members of TRPV and TRPM subfamilies, although some TRPCs appear to display these properties as well.

The major mechanisms of activation of TRP channels can be summarized as follows:

1. *Receptor stimulation.* G-protein-coupled receptors (GPCRs) and receptor tyrosine kinases that activate PLCs can modulate TRP channel activity. This has been demonstrated extensively in the case of TRPCs. Some TRPCs (TRPC4 and TRPC5) appear to be regulated by PIP₂ *per se*, while others (TRPC3, TRPC6, and TRPC7) are activated by the hydrolysis product, diacyl glycerol, although the exact mechanism involved in this is not yet clear. TRPC1, TRPC3, and TRPC4 have also been reported to be activated via depletion of intracellular Ca²⁺ store.
2. *Ligand activation.* Ligands that regulate activation of TRP channels include small organic molecules, including synthetic compounds and natural products, endogenous lipids or products of lipid metabolism, purine nucleotides and their metabolites, or ions such as [Ca²⁺]_i, H⁺, and [Mg²⁺]_i.
3. *Direct activation.* Changes in ambient temperature are strongly coupled to the gating of TRPV1–TRPV3 and TRPM8. Other putative direct activators include mechanical and osmotic stimuli. Heating and cell swelling may also act indirectly to activate TRP channels through generation of second messengers (e.g., TRPV4 activation by hypotonic cell swelling is associated with the generation of arachidonic acid) or other as-yet-unidentified mechanisms.
4. *Constitutive activation.* Some TRPs display constitutive activation. These channels are regulated by trafficking or other mechanisms including protein kinases, Ca²⁺/CaM, as well direct effects of Ca²⁺ or Mg²⁺ (the latter suppresses TRPM7).

TRP Channel Function

Sensory Functions

TRPV, TRPM, TRPA subfamilies play critical roles in sensory modalities, such as touch, hearing, taste, olfaction, vision, and thermal sensation. A number of mammalian TRP channels are activated by temperature changes and play a key role in animal behavior and response to ambient temperature. TRPV1 and TRPV2 are sensors for very warm (>43 °C) and very hot (>52 °C) temperatures, respectively, whereas TRPV3 (>30–39 °C) and TRPV4 (~25–34 °C) contribute to sensing moderate temperatures. TRPM8 appears to function as a cold sensor. The role for TRPA1 in cold sensing is controversial. Most thermoTRPs are expressed in neurons involved in thermosensation. However, TRPV3 and TRPV4 are also expressed in keratinocytes and are activated by temperature increase in these cells. It has been proposed that keratinocytes might release adenosine triphosphate (ATP) that acts in an autocrine fashion to activate residing neurons in the skin. High-temperature sensing TRPs have also been shown to be involved in pain sensation and it is believed that the sensing mechanism at the channel level is similar in both cases.

Many TRP channels bind to and are regulated by various natural product ligands. For example, TRPV1 binds to capsaicin, a compound which elicits the sensation of heat. In addition to displaying impaired responses to heat and vanilloid compounds, TRPV1^{-/-} mice display minimal thermal hyperalgesia. On the other hand, TRPA1 accounts for response to chemical irritants, such as mustard oils, while TRPM5 is a taste transduction channel which broadly senses taste associated with sweet, bitter, and amino acids. Interestingly, TRPM5 contributes to the enhanced perception of sweetness at higher temperatures. TRPC2 functions in pheromone detection and response. Male mice lacking the channel fail to show the normal aggressive behavior toward intruding males.

Other sensory functions served by TRP channels include mechanosensation. TRPV4 is activated by hypotonicity, although it is not believed that the channel is directly activated by membrane stretch, rather arachidonic acid generated in response to osmotic stress might be the activator of the channel. However, TRPV4 serves a sensing function and is required for regulatory volume decrease.

Other Functions Served by TRP Channels

TRP channels play important roles in the secretion and absorption of fluids, electrolytes, and hormones in vertebrates and invertebrates. Further, they are involved in critical cellular processes such as embryonic development, neuronal growth, endothelial migration and proliferation, and cell volume regulation. Two TRPM channels have been suggested to be involved in insulin secretion, TRPM2 and TRPM4. TRPV1 is required for urinary bladder function and mutations in TRPV1 result in an increase in urinary bladder capacity and decreased voiding. The exact mechanism by which TRPV1 regulates urinary bladder function is not clear. There is definitive evidence that TRP channels are required for renal function. TRPV5 mutations are associated with abnormal renal function in humans. Deletion of TRPV5 in mouse causes hypercalciuria due to a defect in Ca²⁺ reabsorption through the distal convoluted tubule of the kidney, where TRPV5 is localized. TRPV5 is also expressed in osteoclasts and contributes to transcellular Ca²⁺ transport during osteoclastic bone resorption. TRPV5 and the highly related TRPV6 are the most Ca²⁺-selective channels among the mammalian TRPs and are involved in absorption of Ca²⁺ in epithelial tissues such as kidney and gut. TRPM7 appears to be involved in embryonic development likely by regulating growth factor-mediated signaling pathways. TRPC channel function and regulation are discussed below.

TRPC Channel Function

Based on amino acid similarities, the mammalian TRPCs can be grouped into four subgroups: TRPC1, TRPC2, TRPC3/6/7, and TRPC4/5. Some TRPCs (such as TRPC1) are expressed ubiquitously. TRPCs have been shown to assemble as both homomeric and heteromeric complexes. Heteromers can be formed by association of TRPC1 with TRPC4 and TRPC5. Likewise, the TRPC3/6/7 subfamily can form heteromers among themselves, although TRPC1/TRPC3 heteromeric channels have been reported. The properties of heteromers may be different from those of the homotetramers. However, presently

there is poor evidence for endogenous TRPC heteromers or the possible functional consequences of such channel diversity.

TRPCs contribute to changes in $[Ca^{2+}]_i$ by acting as Ca^{2+} entry channels in the plasma membrane directly or by changing the membrane potential which can modulate the driving force for other channels. TRPCs are basically nonselective Ca^{2+} -permeable cation channels with variable Ca^{2+}/Na^+ permeability ratios. In excitable cells, Na^+ entry through TRPC channels would lead to membrane depolarization, activation of voltage-gated Na^+ and Ca^{2+} channels, and additional Ca^{2+} entry. In nonexcitable cells, Na^+ entry through TRPCs would make membrane potential more positive, alter cell volume, and affect other Na^+ -dependent cellular processes. Such effects have been reported for TRPC3, TRPC4, and TRPC6. In both excitable and nonexcitable cells, the activation of TRPCs results in increases in intracellular Ca^{2+} concentrations ($[Ca^{2+}]_i$) and Ca^{2+} -induced signaling cascade. Below are listed some of the reported functions of TRPC channels.

In receptor and store-operated calcium entry

TRPC channels are activated downstream of PLC. Therefore, these channels were first suggested to be the channels underlying I_{CRAC} . However, this was ruled out based on patch-clamp experiments which revealed mammalian TRPCs are nonselective cation channels which conduct Na^+ and K^+ as well as Ca^{2+} in most cases; thus, the channels appear to be distinct from Ca^{2+} release-activated Ca^{2+} (CRAC) channels. A substantial body of work demonstrates that TRPCs (TRPC1, TRPC3, and TRPC4) can also contribute to store-operated Ca^{2+} entry (SOCE), at least in some cell types. These data primarily include studies where knockdown of TRPC expression resulted in a reduction in SOCE. However, there is also extensive literature indicating that some TRPC channels do not operate as store-regulated channels *per se* but contribute to SOCE via secondary mechanisms. Indeed, the TRPC3/TRPC6/TRPC7 subfamily channels are activated by the second-messenger diacylglycerol (DAG), a product of PLC-dependent hydrolysis of PIP_2 . In addition, recent studies demonstrate that TRPC4 and TRPC5 can be directly activated by Ca^{2+} and levels of PIP_2 . While TRPC1, TRPC3, and TRPC4 activation is less defined, increasing number of reports now suggest that TRPCs might be regulated by STIM1, the endoplasmic reticulum (ER) calcium sensor that regulates SOCE. The strongest data available in this regard are for TRPC1. It has been shown that TRPC1 interacts with stromal interaction molecule 1 (STIM1) following store depletion and that this interaction occurs in the periphery of the cell. Further, an important study demonstrated that TRPC1 is gated by STIM1 through electrostatic interactions between the polybasic domain of STIM1 and conserved aspartate residues in TRPCs, in this case TRPC1 and TRPC3. In fact, TRPC1 has been suggested to be the link that mediates STIM1-dependent regulation of other TRPC channels. According to this model, TRPCs that interact to form heteromeric channels with TRPC1 would also exhibit STIM1-dependent activation. Another intriguing observation that remains unqualified is that TRPC1 activation is dependent on the function of Orai1, the major CRAC channel component. This functional or direct coupling between Orai1 and TRPC channels remains to be elucidated. Together, these new findings suggest that TRPC channel

regulation is more complex than previously proposed and could involve several components and levels of regulation.

In neurons

TRPC channels play roles in neurite outgrowth and growth cone guidance in the developing nervous system. TRPC3 was suggested as a candidate for regulation of growth cone guidance by brain-derived neurotrophic factor (BDNF) because it is activated in the developing mammalian brain by BDNF which was subsequently supported by studies with knockdown of TRPC3 expression. TRPC5 is involved in neurite extension. TRPC channels also function postsynaptically in response to the activation of metabotropic receptors. For example, release of gamma-aminobutyric acid (GABA) from the dendrites was greatly reduced in TRPC4^{-/-} mice. In contrast, TRPC1 appears to underlie an excitatory postsynaptic current that is generated in response to metabotropic glutamate receptor stimulation in Purkinje cells.

In vascular endothelial and smooth muscle cells

TRP channels exert several different effects on vascular function, including endothelial permeability, responses to oxidative stress, myogenic tone, cellular proliferative activity, migration, and thermoregulation. TRPCs (including TRPC1, TRPC3, TRPC4, and TRPC6) act as SOC entry or receptor-operated calcium (ROC) entry channels and regulate vascular permeability. Influx of Ca^{2+} into vascular endothelial cells causes the release of vasoactive agents, which affect vascular smooth muscle. Although many TRP channels are expressed in the endothelium, the data are strongest for TRPC4. Mice lacking TRPC4 are deficient for an IP_3 -induced Ca^{2+} influx current and display greatly reduced vasorelaxation. In contrast, knockout of TRPC6 increases vascular smooth muscle contractility and blood pressure due to an increase in the expression of TRPC3, which has constitutive activity under these conditions. In addition, TRPC can have an important role in vascular proliferation and remodeling. TRPC6 plays an important role in vascular endothelial growth factor (VEGF)-mediated angiogenesis, while TRPC4 participates in vascular remodeling. Remodeling can be possibly regulated via two mechanisms: proliferative factors may activate TRP channels, with the consequent increase in intracellular Ca^{2+} , thus modulating the signal transduction pathways and leading to remodeling; second, Ca^{2+} influx through TRP channel may stimulate endothelial cells to produce and release proliferative factors such as VEGF and platelet-derived growth factor (PDGF), which then subsequently induce vascular remodeling.

In exocrine gland cells

Salivary secretion is Ca^{2+} dependent, and increases in $[Ca^{2+}]_i$ are associated with increased secretion. Adenovirus-mediated overexpression of human TRPC1 in rat submandibular glands results in a fivefold increase in muscarinic-stimulated fluid secretion and enhanced Ca^{2+} entry response to carbachol and thapsigargin. This was confirmed by studies with TRPC1^{-/-} mice which displayed severe loss muscarinic-stimulated salivary fluid secretion together with decreased agonist- and thapsigargin-stimulated Ca^{2+} entry. Similarly, TRPC3^{-/-} mice displayed reduced SOCE in pancreatic acinar cells, which was linked to a reduction in the severity of pancreatitis.

This suggests a possible involvement of TRPC3 in normal physiology as well as pathophysiology of the exocrine pancreas.

TRP Channels and Human Diseases

Given the diversity in the properties and regulation of TRP channels, it is not unexpected that they appear to have major roles in human disease. While in many cases the effect of alterations in TRP channel function or expression leads to downstream changes in cell signaling and function, several TRP channelopathies have been identified in which mutations in TRP channels have been shown to be causative in the disease. We will summarize these here.

TRPC channel genes have been linked to cardiovascular and pulmonary diseases, inflammation, cancer, skin diseases, neurological diseases, (diabetic) kidney diseases, and proliferative diseases via dysregulation of the cell cycle. The only TRPC-related channelopathy described so far is focal and segmental glomerulosclerosis (FSGS), which has been linked to TRPC6 mutation. FSGS is functionally characterized by proteinuria and progressive decline in renal function of the late-onset type. Since the TRPC6 mutations in FSGS patients are gain of function, it is suggested that enhanced Ca^{2+} entry through TRPC6 may constitute the pathogenic trigger that initiates cell death or compromises the permeability barrier. Activation of nuclear factor of activated T (NFAT) cells could also play an important role in FSGS.

TRPV channelopathies have been linked to skeletal dysplasias. Mutations in *Trpv4* cause a spectrum of skeletal dysplasias, which are lethal in some cases. Spondylometaphyseal dysplasia (SMD) Kozlowski type (SMDK) has been linked to missense mutations in *Trpv4*. These mutations lead to alterations in the basal activity of TRPV4 channels. Scapuloperoneal spinal muscular atrophy (SPSMA) and Charcot-Marie-Tooth disease type 2C (CMT2C; also known as hereditary motor and sensory neuropathy type 2) are phenotypically heterogeneous neurodegenerative disorders in which function of TRPV4 might be anticipated. Four TRPV4 mutations have been identified in this case although how these impair the TRPV4 function is not yet known.

TRPM1, or melastatin, is a putative tumor suppressor protein in melanoma cells. TRPM1 expression decreases in melanoma cells, resulting in progression of malignant melanoma. More recently, around 14 mutations in the *Trpm1* gene are linked to autosomal recessive congenital stationary night blindness (CSNB), a clinically and genetically heterogeneous group of retinal disorders characterized by nonprogressive impaired night vision and variable decreased visual acuity. All patients with recessive *Trpm1* mutations display myopia, reduced central vision, nystagmus, and electroretinographic (ERG) evidence of ON bipolar cell dysfunction. Mutations in *Trpm2* and *Trpm7* genes cause Guamanian amyotrophic lateral sclerosis (ALS-G) and parkinsonism-associated dementia (PDC), as a result two related neurodegenerative disorders. Further, both TRPM2 and TRPM7 are directly implicated in neuronal cell death pathways by sensing oxidative stress (TRPM2) and being crucial for cell viability in neurodegenerative diseases, including Alzheimer's disease and amyotrophic lateral sclerosis. Indeed, a subset of ALS-G and PDC are

heterozygotes for a missense mutation in the *Trpm2* and *Trpm7* genes.

TRPM6, like TRPM7, is involved in Mg^{2+} homeostasis. Mutations in *Trpm6* gene have been associated with hypomagnesemia with secondary hypocalcaemia (HSH) or HOMG. HSH is an autosomal recessive disorder that is characterized by very low Mg^{2+} and Ca^{2+} levels in the serum. This disease is diagnosed during the first 6 months of life due to characteristic neurological symptoms. The primary defect in HSH is an impaired intestinal Mg^{2+} reabsorption. A decreased renal/intestinal Mg^{2+} reabsorption and the resulting decrease in serum Mg^{2+} levels seem to lower parathyroid hormone (PTH) output by the parathyroid gland. A decrease of both PTH and serum Ca^{2+} levels (secondary hypocalcemia) results in convulsions and spasms in early infancy that, if left untreated, can lead to mental retardations or death.

A mutation in the *Trpm4* gene causes autosomal dominant progressive familial heart block type I, a progressive cardiac bundle branch disease in the His-Purkinje system. This has been associated with deSUMOylation of the channel, impairing its endocytosis, thus leading to the elevated channel density in the plasma membrane. The increase in plasma membrane TRPM4 channel very likely causes depolarization of the conduction system and the heart block.

Conclusion

Since their identification almost a decade ago, TRP channels have emerged as regulators of diverse and critical cellular and physiological functions. In addition to those discussed in this chapter, TRP channels are expressed and function in a great variety of multicellular organisms, including worms, fruit flies, zebrafish, and even yeast, indicating a strong evolutionary significance in these channels. Most importantly, based on the wide range of cellular functions performed by mammalian TRP channels, these channels appear to be involved from most basic to highly specialized cellular processes. Further, impact of disruption of channel function has now been associated with the progression of disease and in some instances established as the causative factor in the disease process. Therefore, TRP channels are exciting targets for drug development for the control of pain, asthma, and neurodegenerative diseases. Future studies hold exciting new prospects for further understanding the function and regulation of these remarkable versatile cation channels.

See also: [Molecular Biology: Gene Expression in Bacterial Systems: The *trp* Operon and Attenuation.](#)

Further Reading

- Abramowitz J and Birnbaumer L (2009) Physiology and pathophysiology of canonical transient receptor potential channels. *FASEB Journal* 23(2): 297–328.
- Ambudkar IS (2007) TRPC1: A core component of store-operated calcium channels. *Biochemical Society Transactions* 35(Pt 1): 96–100.
- Ambudkar IS, Bandyopadhyay BC, Liu X, et al. (2006) Functional organization of TRPC- Ca^{2+} channels and regulation of calcium microdomains. *Cell Calcium* 40(5–6): 495–504.

- Ambudkar IS, Ong HL, Liu X, Bandyopadhyay BC, and Cheng KT (2007) TRPC1: The link between functionally distinct store-operated calcium channels. *Cell Calcium* 42(2): 213–223.
- Birnbaumer L (2009) The TRPC class of ion channels: A critical review of their roles in slow, sustained increases in intracellular Ca^{2+} concentrations. *Annual Review of Pharmacology and Toxicology* 49: 395–426.
- Kiselyov K, Wang X, Shin DM, Zang W, and Muallem S (2006) Calcium signaling complexes in microdomains of polarized secretory cells. *Cell Calcium* 40(5–6): 451–459.
- Lee KP, Yuan JP, Hong JH, et al. (2010) An endoplasmic reticulum/plasma membrane junction: STIM1/Orai1/TRPCs. *FEBS Letters* 584(10): 2022–2027.
- Montell C (2005) The TRP superfamily of cation channels. *Science's Signal Transduction Knowledge Environment* 2005(272): re3.
- Montell C, Birnbaumer L, and Flockerzi V (2002) The TRP channels, a remarkable functional family. *Cell* 108: 595–598.
- Nilius B, Owsianik G, Voets T, and Peters JA (2007) Transient receptor potential cation channels in disease. *Physiological Reviews* 87(1): 165–217 (review).
- Owsianik G, Talavera K, Voets T, and Nilius B (2006) Permeation and selectivity of TRP channels. *Annual Review of Physiology* 68: 685–717.
- Yuan JP, Kim MS, Zeng W, et al. (2009) TRPC channels as STIM1-regulated SOCs. *Channels (Austin)* 3(4): 221–225.

Translation Elongation in Bacteria

S C Blanchard, Weill Cornell Medical College, New York, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Aminoacyl-tRNA tRNA enzymatically esterified with an amino acid at its 3'-CCA end.

Codon The three nucleotide units of information within mRNA that is recognized by the anticodon of transfer RNA molecules on the ribosome to bring specific amino acids into the ribosome.

Decoding center/decoding site The region of the small subunit of the ribosome responsible for recognizing shape-specific features of the tRNA anticodon-mRNA codon minihelix during tRNA selection.

Elongation factor-G (EF-G) The conserved protein molecule responsible for catalyzing the process of directional substrate translocation on the ribosome during the elongation phase of protein synthesis. Also responsible for aspects of the ribosome recycling mechanism.

Elongation factor-Ts The protein molecule responsible for catalyzing nucleotide release from EF-Tu; EF-Ts is a guanosine nucleotide exchange factor (GEF).

Elongation factor-Tu (EF-Tu) The conserved protein molecule responsible for escorting tRNA substrates into the A site of the ribosome during the elongation phase of protein synthesis.

GTPase activating center (GAC) The site within the large subunit of the ribosome responsible for specific interactions with translation factors that triggers GTP hydrolysis.

Guanosine diphosphate (GDP) One of the small-molecule by-products of GTP hydrolysis.

Guanosine triphosphate (GTP) A small molecule that serves as the energy source for multiple translation reactions.

Messenger RNA (mRNA) Ribonucleic acid template molecules read by the ribosome in a directional fashion to synthesize specific protein products.

Peptidyltransferase center (PTC) The site within the large subunit of the ribosome responsible for catalyzing peptide bond formation.

Ribosome The enzyme responsible for carrying out protein synthesis in the cell.

Ribozyme A structured molecule composed of ribonucleic acid that is capable of catalyzing specific chemical reactions.

Transfer RNA (tRNA) The substrate of the ribosome, a 25 000 Da, L-shaped adapter molecule composed of ribonucleic acid that brings specific amino acids to the translating particle through base-pairing interactions with the mRNA codon sequence.

Translocation The directional movement of tRNA and mRNA substrates with respect to the ribosome.

tRNA selection The process by which specific aminoacyl-tRNA molecules are selected by the ribosome according to the mRNA codon sequence.

Introduction

The 'central dogma of molecular biology' proposed by Francis Crick in 1958 specifies that genetic information within cells flows from DNA to RNA and then to protein. This process of information transfer is carried out by two, highly conserved macromolecular assemblies, whose regulation is central to the gene expression mechanism.

Transcription, carried out by RNA polymerase, converts double-stranded, chromosomal DNA into single-stranded messenger RNA (mRNA). Translation, carried out by the ribosome, converts the triplet codon sequence of mRNA into a specific polypeptide. Both RNA polymerase and the ribosome serve as integration points for the regulation of gene expression in the cell. The balance of these enzymes' synthesis activities and the various degradation activities in the cell responsible for mRNA and protein turnover give rise to the transcriptome and the proteome, respectively.

While RNA polymerase and the ribosome are physically distinct macromolecular assemblies, the paradigms of transcriptional and translational control are in many ways analogous. Both enzymes are directed to: (1) localize to an appropriate start site of synthesis on a structured nucleic acid template; (2) processively synthesize product by unwinding and transiting

the template strand in discrete steps by pairing individual substrate molecules with the segments of the template strand – the process of elongation; and (3) correctly terminate synthesis upon completion of a full-length product, and not before.

Although it is generally accepted that transcription and translation are principally regulated by controlling initiation, it is increasingly clear that the synthesis phases of both processes serve as important points of control in the gene expression mechanism. Mechanisms that enable control of these molecular machines allow cells to respond to intra- and extracellular queues through changes in the repertoire of mRNAs and proteins that are synthesized. The capacity to promptly regulate active synthesis reactions likely enables cells to respond rapidly to environmental queues and to diversify the nature and spatial distribution of the proteome in ways that are advantageous to cell growth. This chapter focuses on the synthesis phase of translation catalyzed by the ribosome, the elongation phase of protein synthesis.

Bacterial Ribosome Assembly

The bacterial ribosome is a massive ribonucleoprotein (RNA-protein) complex (approximately 2.5 MDa) composed of

distinct gene products, including three RNA molecules and 57 proteins. Ribosome biogenesis occurs in a region within the cell nucleus called the nucleolus. This complex, multistep process is driven by transcription of the conserved ribosome DNA operon by RNA polymerase I. Assembly of the ribosome commences on the nascent ribosomal RNA (rRNA) transcript through a relatively slow process involving a battery of enzymes including chaperones, helicases, and RNAases as well as factors that post-transcriptionally modify the rRNA. Actively growing laboratory strains of bacteria divide in 30–60 min and contain approximately 10 000–20 000 ribosomes. Given steady-state conditions, the rate of ribosome biogenesis can therefore be estimated at roughly 250 ribosomes per minute. To meet such demands, most rapidly dividing strains of bacteria typically contain duplicated copies of the conserved rRNA operon.

The assembled and fully processed ribosome is a two-subunit RNA–protein enzyme. In bacteria, the individual subunits are commonly referred to as 30S and 50S particles, which correspond to their sedimentation velocities. Each subunit is composed of approximately 35% protein and 65% rRNA, the latter of which contains approximately 5–10 unique posttranscriptional modifications whose precise function remains unclear. The locations of modified nucleotides within functional centers of the large and small subunit rRNA suggest some degree of contribution to core functions of the ribosome during translation.

Bacterial Ribosome Composition

The small, 30S, subunit contains a single RNA molecule of approximately 1500 nucleotides (16S rRNA) and 21 proteins. The large, 50S, subunit contains two RNA molecules totaling approximately 3000 nucleotides (~120 of 5S rRNA and ~2900 of 23S rRNA) as well as approximately 36 proteins. While the functional ribosome is an amalgam of both large and small subunits, a 70S particle in bacteria, biochemical studies have attributed distinct and separable activities to each subunit.

The small subunit is principally responsible for interactions with the mRNA template, which encodes the instructions for which protein to synthesize, as well as interactions with the anticodon stem-loop regions of ribosome-associated tRNA molecules. The site where mRNA codon–tRNA anticodon interactions are formed is referred to as the decoding site.

The large ribosomal subunit is responsible for the catalysis of peptide bond formation, an enzymatic activity that covalently links amino acids carried by adjacently bound tRNA substrates into a polypeptide chain. Peptide bond formation on the large subunit occurs within the peptidyltransferase center (PTC). This site contains two highly conserved rRNA elements which form base-pairing interactions with the universally conserved 3′-CCA ends located at the tips of the acceptor stem region of tRNA molecules. The large subunit also contains the GTPase activating center (GAC) that triggers GTP hydrolysis within the ribosome-associated translation factors including elongation factors (EF)-Tu and (EF)-G, initiation factor (IF)-2, as well as release factor (RF)-3.

Bacterial Ribosome Architecture

Following the identification of the ribosome as the site of protein synthesis within the cell in the mid- to late 1950s – an achievement that contributed to the Nobel prize in physiology and medicine awarded to Albert Claude, Christian de Duve, and George Emil Palade – attempts to characterize its physical makeup and structure quickly ensued. Negative stain electron microscopy, immunoelectron microscopy, and cryoelectron microscopy made many early contributions in the area, including the first overall topological maps of the 70S particle. The critical roles electron microscopy played over the years include, among other things, the anthropomorphized nomenclature used to describe the various structural features of the small and large subunits (discussed below).

Cryoelectron microscopy methods continue to have an enormous impact on the translation field to this day by providing structural snapshots of the changing ribosome architecture during initiation and synthesis reactions. Such achievements have served as an important starting point for atomic resolution structure determination efforts.

The first atomic resolution structures of the ribosomal subunits, revealed in 2000, culminated more than three decades of previous structural, biochemical, and genetic research into the mechanism of translation ([Figure 1\(a\)](#)). Three of the pioneers in this area, whose labs contributed to making these breakthroughs possible, were awarded the Nobel prize in chemistry in 2009: Drs. Ada Yonath, Venkatraman (Venki) Ramakrishnan, and Thomas Steitz. The efforts of these scientists capitalized on widely recognized contributions of many other leaders within the field.

The large, 50S, subunit is roughly globular in nature, and approximately 250 Å in diameter, but contains three stalk-like protrusions called the L1 stalk, the L12 stalk, and the central protuberance that extend away from the particle's core. In the crown view orientation of the ribosome, the L1 and L12 stalks – so called because they contain the ribosomal proteins L1 and L12 – extend away from the large subunit core and its central protuberance toward the lagging and leading edges of the particle, respectively. The highly compacted core of the 50S particle is principally composed of 23S rRNA. Ribosomal proteins distribute evenly throughout the subunit such that their globular domains are positioned near the solvent-exposed surfaces. Long, positively charged, C- and N-terminal extensions penetrate deep into the core to stabilize tertiary interactions between 23S rRNA segments. The site responsible for peptide bond formation, the PTC, is located near the center of the 50S particle. The unusual resistance of PTC activities to protease digestion and extraction procedures, combined with structural data revealing that it is principally composed of RNA, has led most to conclude that the ribosome is a ribozyme.

The 30S subunit is relatively flat (approximately 250 Å long, 50 Å thick, and 180 Å wide) and flexible by comparison. The small subunit is composed of three, roughly globular domains, referred to as the head, body, and platform, that are connected by a narrow neck-like extension composed of 16S rRNA. mRNA binds at the interface of these three domains, encircling this neck element. In so doing, the small subunit holds a segment of approximately 18 nucleotides – six tandem mRNA

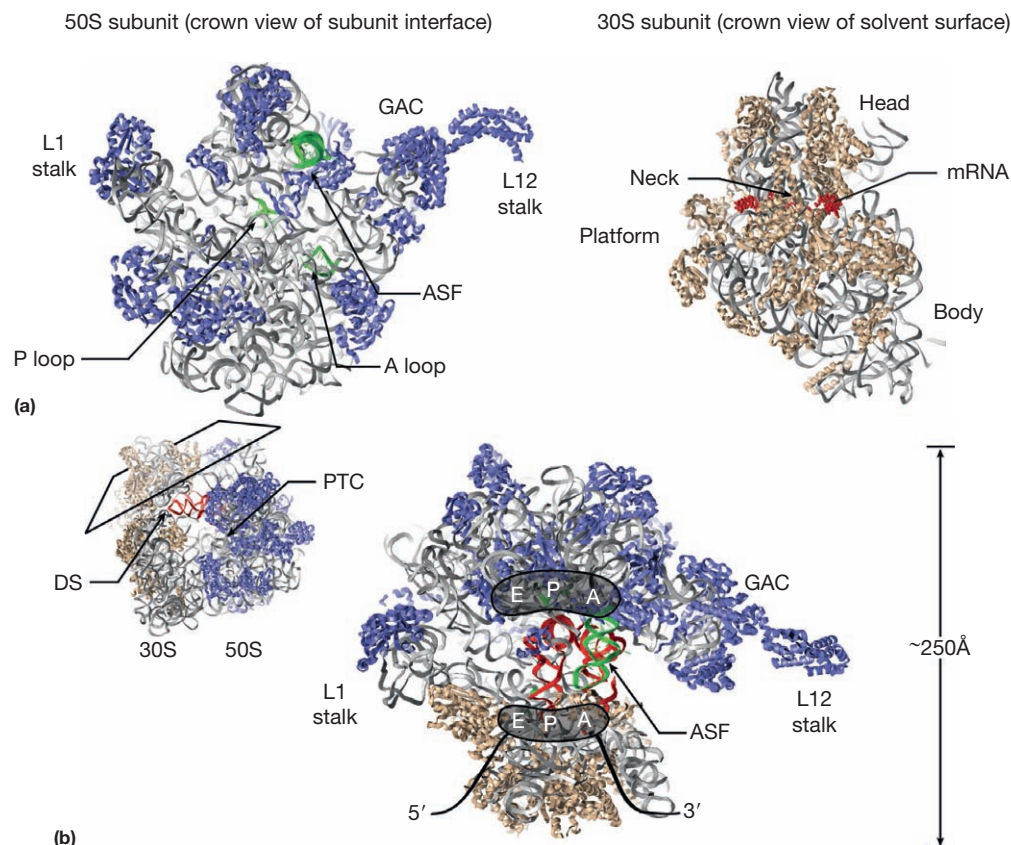


Figure 1 Structures of the 30S and 50S ribosomal subunits as well as the functionally active 70S particle. Distinct views of three-dimensional, atomic models of the bacterial ribosome. (a) The large (50S) and small (30S) subunits of the bacterial ribosome shown in crown view with landmark structural features indicated. Ribosomal proteins are shown in blue and tan in the large and small subunit, respectively; in both subunits, rRNA is shown in gray ribbons. (b) The 70S pre-translocation ribosome complex shown in a cross-section view (as indicated) looking down the subunit interface to reveal the tRNA substrate-binding channel comprising the A, P, and E sites where A- and P-site tRNAs are shown in their classical positions (red). Landmark structural features, including the decoding site (DS), the peptidyltransferase center (PTC), L1 stalk, L12 stalk, the GTPase activating center (GAC), the A-site finger (ASF), and mRNA track are indicated (PDB accession codes 2QAL, 2QAM, 1VS9, and 1Q86).

codons –exposed to the solvent so as to facilitate its interactions with tRNA substrates.

The translationally active, 70S particle, is formed by the joining of large and small subunits through a relatively flat, RNA-rich interface such that the head and central protuberance domains of the small and large subunit juxtapose. High-resolution structures of the 70S particle, emerging as early as 2005 from the group of Dr. Harry Noller and colleagues, provided the first molecular insights into the RNA–RNA and RNA–protein bridges mediating 70S complex formation. Later studies, including work from the groups of Noller, Yusupov, Ramakrishnan, and Cate, further contributed to revealing the molecular basis of the three tRNA-binding sites within the 70S particle. These sites are commonly referred to as the aminoacyl (A), peptidyl (P), and exit (E) sites based on their preferred binding of aminoacyl, peptidyl, and deacylated tRNA, respectively.

tRNA substrates orient within a solvent-exposed channel running the length of the 70S particle in a manner that is roughly perpendicular to the subunit interface. In so doing, the L-shaped, 25 kDa tRNA molecules are positioned such that their anticodon domains can form Watson–Crick base-pairing interactions with mRNA on the small subunit while

their acceptor stem, and single-stranded 3′-CCA residues can interact with conserved elements within the large subunit PTC (Figure 1(b)).

The Mechanism of Bacterial Protein Synthesis

In both prokaryotic and eukaryotic systems, the process of protein synthesis by the ribosome is carried out in four principal stages: initiation, elongation, termination, and recycling (Figure 2). Many key mechanistic considerations for the process of translation are conserved among species, including the size, structure, and functional contributions of each of the translation components. Although initiation is generally considered the rate-determining step in translation, each phase can be exquisitely regulated in the cell through changes in the cellular milieu and/or changes in the composition of the translation machinery.

Initiation

Initiation localizes the large and small subunits of the ribosome, along with a specialized, *N*-formylmethionine-charged

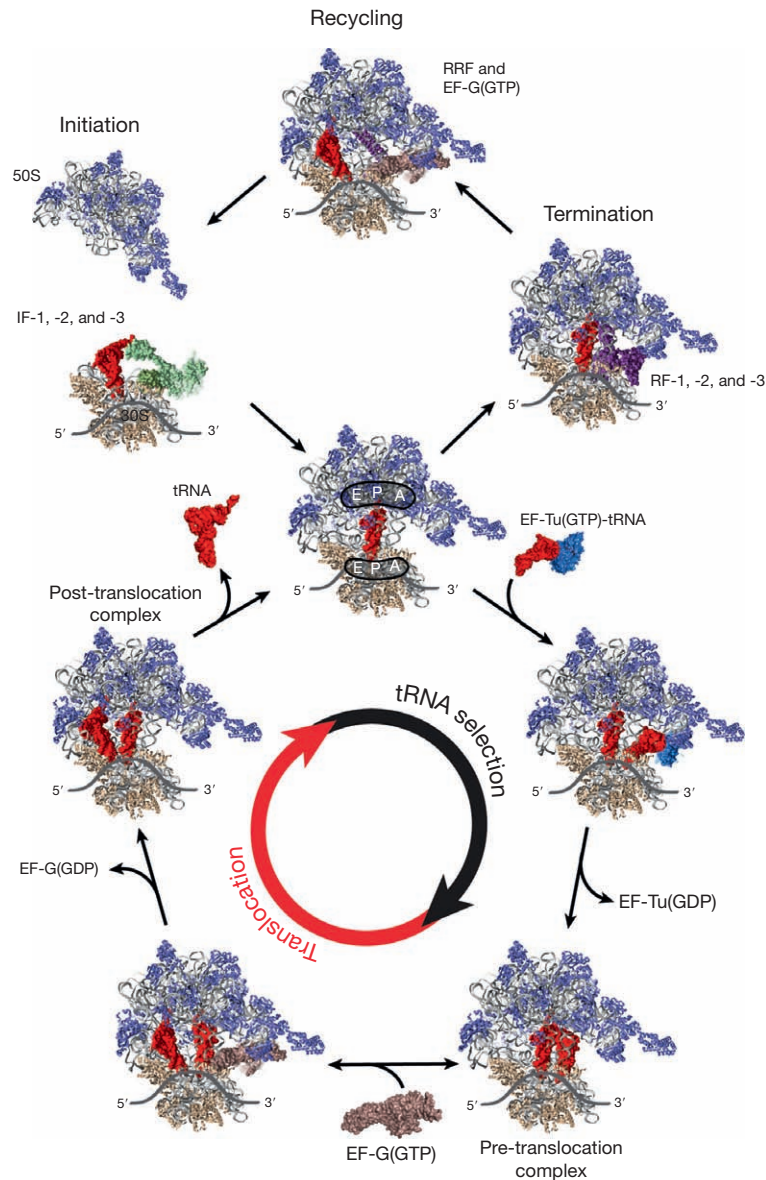


Figure 2 Schematic representation of the four principal phases of translation. This highly simplified diagram illustrates initiation, elongation, termination, and recycling steps, as well as the translation factors responsible for facilitating these reactions.

initiator tRNA substrate ($fMet-tRNA^{fMet}$) to the start site of protein synthesis. This site is typically located within the 5'-end of the mRNA template. As elaborated elsewhere in this edition, the initiation process entails a two-dimensional scan of the mRNA sequence, wherein the ribosome, aided by IFs and GTP hydrolysis, pairs initiator tRNA with the start codon, typically an AUG sequence. Start-site selection establishes the reading frame of protein synthesis and is therefore a key determinant of overall translational fidelity. The use of alternative cryptic start sites increases the diversity of protein products from a single mRNA transcript. Completion of the initiation process results in a post-initiation ribosomal complex in which the subunits are joined and $fMet-tRNA^{fMet}$ binds to the mRNA start codon within the P site of the ribosome. In this complex, the universally conserved, 3'-CCA end of the $tRNA^{fMet}$ acceptor stem forms two Watson-Crick base pairs with the P loop of

23S rRNA, a conserved RNA element within the large subunit PTC. Concurrently, the $tRNA^{fMet}$ anticodon stem loop forms three Watson-Crick base pairs with the mRNA codon within the small subunit-decoding region. Once the formation of this complex is complete, the elongation phase of protein synthesis begins.

Elongation, the Synthesis Phase of Translation

During the elongation phase of protein synthesis, the ribosome translates the mRNA open reading frame (ORF) into protein by pairing specific aminoacylated-tRNA molecules ($aa-tRNA$) with the mRNA codon sequence. There are 44 distinct tRNA genes in the bacterial genome. Each species of tRNA in the cell is enzymatically charged with one of the 20 essential amino acids through the action of a diverse family of proteins

collectively referred to as aminoacyl-tRNA synthetases (aaRSs). The charging of tRNA is an ATP-dependent reaction resulting in the covalent linkage of the carboxylic acid moiety of an amino acid to the terminal 2' or 3' hydroxyl group of the universally conserved 3'-CCA sequence within the tRNA acceptor stem, forming a labile ester bond between the tRNA and amino acid. The specificity of aminoacylation, determined by identity elements within each tRNA molecule, is a critical determinant of translational fidelity. The overall fidelity of tRNA aminoacylation (approximately 1 error in 10^6 – 10^7) is ensured by proofreading mechanisms within specific aaRS proteins, wherein the ester linkage of incorrectly charged tRNAs is specifically hydrolyzed. Each of the 44 aa-tRNA substrates is used by the ribosome to decode one or more of the 61 sense codons. The remaining three mRNA codons are called nonsense codons, which instruct the ribosome to terminate protein synthesis. Nonsense codons are specifically recognized by protein factors that release the synthesized protein from the ribosome by water-catalyzed hydrolysis of the tRNA–amino acid ester linkage.

Elongation, the protein synthesis phase of translation, is driven by repetitive cycles of aa-tRNA selection, peptide bond formation, and the directional translocation of tRNA and mRNA substrates. As detailed below, such processes are dominated by tRNA and factor-binding events at the leading edge of the translating particle mediated by ribosomal stalk protein L12 of the large subunit and factor-catalyzed GTP hydrolysis on the ribosome. The mechanism of elongation and the origins of fidelity in this complex process have received much attention over the past several decades by many research teams. Current models of aa-tRNA selection and translocation stem principally from rapid stopped-flow fluorescence and quenched-flow measurements pioneered by the groups of Wintermeyer, Rodnina, and colleagues, which have been later supplemented by bulk and single-molecule fluorescence and fluorescence resonance energy transfer measurements. Aminoacyl-tRNA molecules are selected at the A site of the ribosome, located at the leading edge of the 70S particle, according to the mRNA template sequence through a multistep, kinetically driven mechanism.

Unlike other processive enzymes in the cell, such as DNA polymerase, the ribosome lacks the capacity to proofread mistakes after the process of substrate selection is complete. Thus, the mechanism of tRNA selection is critical in maintaining the genetic code, the observed correspondence between the mRNA ORF and protein sequence. As for initiation, the process of aa-tRNA selection is driven by Watson–Crick base-pairing interactions between the tRNA anticodon and the mRNA codon and is a key determinant of translational fidelity. The observed fidelity of aa-tRNA selection, estimated through a variety of both *in vivo* and *in vitro* measurements, as 1 error in 10^2 – 10^4 codons, ensures approximately one mistake per protein synthesized or less. Consistent with a careful balance of thermodynamic and kinetic driving forces in the tRNA selection mechanism, both experimental and theoretical evidence suggest that higher fidelity in this process can be achieved at the expense of slower rates of synthesis.

Although the process of tRNA selection is driven by the base pairing between the mRNA codon and tRNA anticodon, the observed fidelity of aa-tRNA selection far exceeds that which can be achieved on the basis of thermodynamic

considerations. Differences in the base-pairing stabilities of correct (cognate) and incorrect (near- and noncognate) aa-tRNAs with the mRNA codon – in some cases as little as a single hydrogen bond – may provide up to one to two orders of magnitude of selectivity, one to two orders of magnitude below what is empirically observed. Research into the origins of fidelity during selection has revealed that selectivity is rooted in the multistep nature of the process. EF-Tu, the conserved GTPase that binds the acceptor stem of aa-tRNA with high affinity, also contributes to the fidelity mechanism. GTP hydrolysis by EF-Tu occurs on the ribosome during the selection process, through a mechanism that is allosterically regulated by the nature of the mRNA codon–tRNA anticodon interaction. Rapid and efficient GTP hydrolysis selectively occurs when the correct, cognate tRNA is recognized through shape-selective elements of the small subunit of the ribosome. Insights into the allosteric GTP hydrolysis triggering mechanism were revealed first through genetic studies and later through structural investigations beginning with cryoelectron microscopy studies led by the Frank and Spahn groups and subsequently by X-ray crystallographic studies out of the Ramakrishnan lab. Such efforts suggest that structural distortions in the entering tRNA molecule as well as EF-Tu are precipitated by conformational changes in the small subunit that occur upon codon–anticodon recognition. tRNA bending and EF-Tu contacts with both large and small subunits participate in the GTP hydrolysis mechanism by allowing the catalytic histidine in EF-Tu to activate water for in-line attack of the gamma-phosphate of GTP. Subsequent release of inorganic phosphate and dissociation of EF-Tu(GDP) from the ribosome allows the aminoacylated, 3'-CCA end of tRNA to enter the PTC and participate in peptide bond formation. Following peptide bond formation, the nascent peptide is covalently linked to tRNA residing in the A site, a complex that is competent for EF-G-catalyzed translocation.

EF-G specifically recognizes the pre-translocation ribosome and operates through a GTP-dependent mechanism to drive the directional movement of tRNA and mRNA substrates with respect to the ribosome. Fidelity during this step relates to the precision with which substrate translocation occurs. In order to maintain the proper reading frame of translation, the ribosome must move with respect to the mRNA by a single codon during translocation. Movement by a nonintegral number of codons compromises the reading frame. Programmed mis-translocation events, including +1 and –1 frameshifting, have been shown to play important roles in the regulation of both housekeeping genes as well as the translation of viral genomes.

Although still an open area of research, EF-G's recognition of the pre-translocation complex is thought to be enabled by a number of specific conformational changes within the ribosome and related movements of 3'-CCA ends of both A- and P-site tRNA with respect to the large subunit. Such processes include a global motion of the small subunit with respect to the large, orthogonal swivel-like motions of the small subunit head domain in the direction of translocation and rearrangements of the A-site finger (ASF), an element of rRNA within the small subunit that lies above and between the A- and P-site tRNA-binding sites. These structural processes facilitate the formation of so-called, hybrid tRNA configurations that allow EF-G to engage the large subunit GAC to catalyze GTP hydrolysis and enter the A site of the small subunit. The formation of

hybrid pre-translocation complex configurations, and their role in the translocation mechanism, has been extensively investigated. Consistent with the notion that the ribosome possesses multiple conformational degrees of freedom, and that the tRNA molecule itself is flexible in nature, numerous hybrid pre-complex configurations likely exist. Thus far, two distinct hybrid configurations of the pre-translocation complex have been rigorously characterized through biophysical and structural means.

While much remains to be delineated about the nature of hybrid pre-translocation complex configurations and their role in subsequent substrate movement (translocation) at a molecular level, the preponderance of evidence suggests that dynamic processes within the pre-translocation complex play a critical role in the translocation mechanism. Current mechanistic models of translocation suggest that EF-G cannot productively engage the ribosome until it spontaneously achieves a configuration in which both A- and P-site tRNAs have adopted hybrid positions, wherein their 3'-CCA ends partially translocate with respect to the large subunit. Of course, such movements are associated with related conformational changes within the ribosome itself. Conformational changes within EF-G, facilitated by GTP hydrolysis, are also likely to play a central role in the translocation mechanism. Such changes are also thought to contribute to the factor's release from the ribosome, preparing the particle for additional elongation steps. Repetitive cycles of tRNA selection and translocation drive the ribosome directionally through the mRNA ORF sequence from the point of translation initiation to termination. In each elongation cycle, the nascent polypeptide is extended by a single amino acid.

Termination and Recycling Complete the Process of Protein Synthesis

When the ribosome reaches a nonsense mRNA codon, one that does not have a cognate tRNA pair, a release factor (either RF-1 or RF-2 depending on the specific stop codon encountered) binds to the A site. Release factor binding enables water to enter the PTC to hydrolyze the mRNA-encoded polypeptide from P-site tRNA, liberating it from the translating particle. A third factor, RF-3, acts to release the bound RF-1 or RF-2 through a process that requires GTP hydrolysis. Ribosomal subunits, tRNA and mRNA, are then separated from each other and prepared for subsequent translation reactions through the action of EF-G and ribosome recycling factor (RRF). Like other translation processes, recycling is mediated by conformational changes in the ribosome as well as energy expenditure. Efficient recycling ensures that the ribosome does not inappropriately reinitiate on downstream ORFs to complete the process of protein synthesis.

Fidelity and Regulation of Translation during Elongation Reactions

As described in previous sections, the elongation phase of protein synthesis serves as a critical determinant of the rate and fidelity of protein synthesis. Correspondingly, regulation of elongation reactions can have a dramatic impact on gene expression and the physical composition of the proteome.

Mechanisms that enable specific control of translation elongation reactions may provide a means to rapidly start or stop the synthesis of specific proteins, or modify the nature of the expressed protein. Known examples of such regulation include programmed alterations of translation elongation *via* the formation of thermodynamically stable secondary structures within the mRNA template. These can either halt the ribosome's transit through the mRNA ORF or induce frameshifting events that alter the reading frame and thus the protein product synthesized.

Examples of global translation elongation regulation are well documented. Such events include the stringent response, triggered by metabolite starvation, resulting in an overabundance of uncharged tRNA molecules that can bind the A site of the translating ribosome. Cells possess the means to detect such events – putatively through the binding of specific factors to the ribosome – resulting in the generation of small-molecule signaling compounds which downregulate ribosome biogenesis and the components of protein synthesis while simultaneously upregulating pathways for metabolite scavenging and synthesis.

The process of translation elongation can also be regulated by controlling the process of nucleotide exchange on EF-Tu. This reaction, catalyzed by elongation factor-Ts (EF-Ts), enables EF-Tu to release GDP after dissociating from the ribosome following the completion of tRNA selection, readying EF-Tu to again bind GTP and aa-tRNA. The activities of EF-Ts, a so-called guanosine nucleotide exchange factor (GEF), are essential to the cell as EF-Tu in its GDP-bound form is unable to bind aa-tRNA. EF-Tu also has an approximately 500-fold higher affinity for GDP than GTP. EF-Ts can be regulated by any number of means including protein–protein interactions that sequester and inactivate it, proteolysis, or post-translational modification.

Elongation reactions can also be regulated by the binding of small-molecule compounds to the translating ribosome. The most dramatic and widely understood examples of this means of control relate to antibiotic agents that are used in research and in the clinic to attenuate the process of translation. According to some estimates, roughly 40–50% of all known antibiotic compounds bind directly to the ribosome. Most of these compounds modify translation elongation reactions in some manner – some act as competitive inhibitors, others block specific conformational events required for proper elongation steps, while some actually promote steps that cause the ribosome to make errors in either tRNA selection, translocation, or termination. These latter compounds have the propensity to specifically or nonspecifically alter the proteome. It is not yet clear whether similar strategies are used by cells to control the gene expression mechanism in some specific manner. The capacity to alter ribosome function through the binding of chemically diverse small-molecule compounds, through both direct and allosteric mechanisms, portends many future discoveries that will shed light on new regulation mechanisms important for gene expression.

Conclusion

Our capacity to measure and quantify mRNA translation using new technologies will be critical to expanding the scope and

breadth of insights into translational control. Here, an emphasis on delineating the precise relationship between the expressed transcriptome and the proteome with increasing sensitivity is paramount. Today, there is no adequate way to do so. The recent emergence of ribosome profiling as well as single-molecule techniques that allow ribosome position on an mRNA to be directly monitored promise to make important contributions in this area that are sure to strengthen our understanding and expand our knowledge of the nature of gene expression mechanisms and the true complexity of bacterial and human proteomes.

See also: [Molecular Biology: Catalysis by RNA, Structural Themes: Group I Introns; Nucleolus, Overview; Ribosome Assembly; Ribosome Structure; Translation Initiation in Eukaryotes: Factors and Mechanisms.](#)

Further Reading

- Berk V and Cate JH (2007) Insights into protein biosynthesis from structures of bacterial ribosomes. *Current Opinion in Structural Biology* 17: 302–309.
- Blanchard SC, Cooperman BS, and Wilson DN (2010) Probing translation with small-molecule inhibitors. *Chemistry and Biology* 17: 633–645.
- Herbert TP and Proud CG (2007) Regulation of translational elongation and the cotranslational protein targeting pathway. In: Mathews MB, Sonenberg N, and Hershey JWB (eds.) *Translational Control in Biology and Medicine*, pp. 601–624. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Korostelev A, Ermolenko DN, and Noller HF (2008) Structural dynamics of the ribosome. *Current Opinion in Structural Biology* 12: 674–683.
- Kozak M (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* 361: 13–37.
- Munro JB, Sanbonmatsu KY, Spahn CM, and Blanchard SC (2009) Navigating the ribosome's metastable energy landscape. *Trends in Biochemical Sciences* 34: 390–400.
- Ogle JM and Ramakrishnan V (2005) Structural insights into translational fidelity. *Annual Review of Biochemistry* 74: 129–177.
- Ramakrishnan V (2008) What we have learned from ribosome structures. *Biochemical Society Transactions* 36: 567–574.
- Wilson DN and Nierhaus KH (2007) The weird and wonderful world of bacterial ribosome regulation. *Critical Reviews in Biochemistry and Molecular Biology* 42: 187–219.

Translation Initiation in Bacteria: Factors and Mechanisms

L Brandi and C O Gualerzi, University of Camerino, Camerino, Italy

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Cynthia L. Pon and Claudio O. Gualerzi, volume 4, pp. 230–236, © 2004, Elsevier Inc.

Glossary

Aminoacyl-tRNA site (A-site) The binding site on the ribosome occupied by the tRNA carrying the next amino acid to be added to a growing polypeptide chain.

Peptidyl-tRNA site (P-site) The binding site on a ribosome occupied by the tRNA carrying a growing peptide chain.

Watson–Crick base pairing Association of two complementary nucleotides in a DNA or RNA molecule stabilized by hydrogen bonding between their base components.

Wobbling interaction Ability of a tRNA to recognize more than one codon by unusual (non-G-C, non-A-U) pairing with the third base of a codon.

Properties, Aminoacylation, and Formylation of the Initiator Transfer RNA

In bacteria, protein synthesis begins with a special brand of transfer RNA (tRNA), the initiator tRNA^{fMet}. Apart from being recognized by the same Met-tRNA synthetase as the elongator tRNA^{met} (the main recognition element being the CAU anticodon), the initiator tRNA is endowed with a number of structural and functional properties which make this particular RNA one of a kind. The main features distinguishing the initiator tRNA from elongator tRNAs are the three consecutive G–C pairs in its anticodon stem, which favor its interaction with the ribosomal P-site, the purine 11–pyrimidine 24 base pairing in the D-stem and the unpaired bases 1 (the 5′-end) and 72 which allow the recognition of Met-tRNA^{fMet} by the transformylase. It should be noted, however, that mutants lacking the transformylase are viable, albeit displaying a highly reduced growth rate. The initiation factor-2 (IF2)–fMet-tRNA interaction requires just a few bases of the 3′ acceptor end of the molecule. The fMet-tRNA is recognized by initiation factor IF2 by virtue of the chemical nature of the methionine, and largely relies on the presence of a formyl group blocking the α-NH₂ group of the amino acid. Indeed, fMet-ACCAAC has almost the same affinity for IF2 as the intact fMet-tRNA. IF2 can form a complex with fMet-tRNA off the ribosome but it has been demonstrated that this complex is nonproductive and that fMet-tRNA is recruited by ribosome-bound IF2.

The Translation Initiation Region of Prokaryotic Messenger RNAs

In bacteria, there are essentially three types of messenger RNAs (mRNAs) that are present in different proportions in different types of bacteria: (1) leadered with Shine–Dalgarno (SD) sequence; (2) leadered without SD sequence; and (3) leaderless (Figure 1). The most-studied and better-known mRNAs are those whose translation initiation region (TIR) contains, upstream of the initiation triplet, a 5′ untranslated region (5′ UTR) of variable length and structure, which generally includes a purine-rich SD sequence complementary to the

anti-SD sequence at the 3′-end of 16S ribosomal RNA (rRNA). The SD sequence and the initiation triplet are separated by a spacer of variable length (typically 5–9 non-structurogenic nucleotides). Also, the sequence at the 3′ side of the initiation triplet, the downstream region (DR), can have an influence on translational efficiency. Highly expressed mRNAs seem to have a bias in favor of an AAA triplet as the second codon and a CA repeat sequence within the first 15 codons was also shown to contribute to highly efficient initiation (Figure 1(a)). Both second and third types of mRNA do not contain the SD sequence either because this is not found in the 5′UTR leader (Figure 1(b)) or because the mRNA does not contain any 5′UTR and begins with the initiation triplet at their 5′-end (Figure 1(c)).

Concerning the initiation triplet, this is the codon from which the ribosomes begin to translate and which sets, at least provisionally, the reading frame of the template (a reading frameshift may occur during elongation). Aside from the leaderless mRNAs, which always begin with an AUG triplet, the other mRNAs may have different initiation triplets. In fact, whereas the most frequently used initiation triplet is AUG, ~10% of the templates begin with rarer yet canonical codons such as GUG and UUG, and, even less frequently, with noncanonical triplets such as AUU, AUC, and AUA. However, regardless of its nature, the initiation triplet is always decoded by the initiator fMet-tRNA. The reason for this initiation codon degeneracy is not always clear, but two facts should be recalled. First, AUG is the only triplet giving rise to a best-fit Watson–Crick base pairing with the anticodon (CAU) of the initiator tRNA, while all other start codons are expected to yield either 3′ or 5′ wobbling interactions with the same anticodon. Second, if not resulting from neutral mutations, the rare initiation triplets must be considered (or suspected to be) important regulatory signals capable of controlling timing and/or efficiency of translation. It is not by chance that, as described below, initiation factor IF3 can discriminate against translation beginning at noncanonical start codons; the best-known example of translation regulation based on the use of a rare codon is the translational auto-repression of IF3 mRNA which begins with an AUU initiation codon.

An important element that contributes to the efficiency of translation initiation is the secondary and tertiary RNA

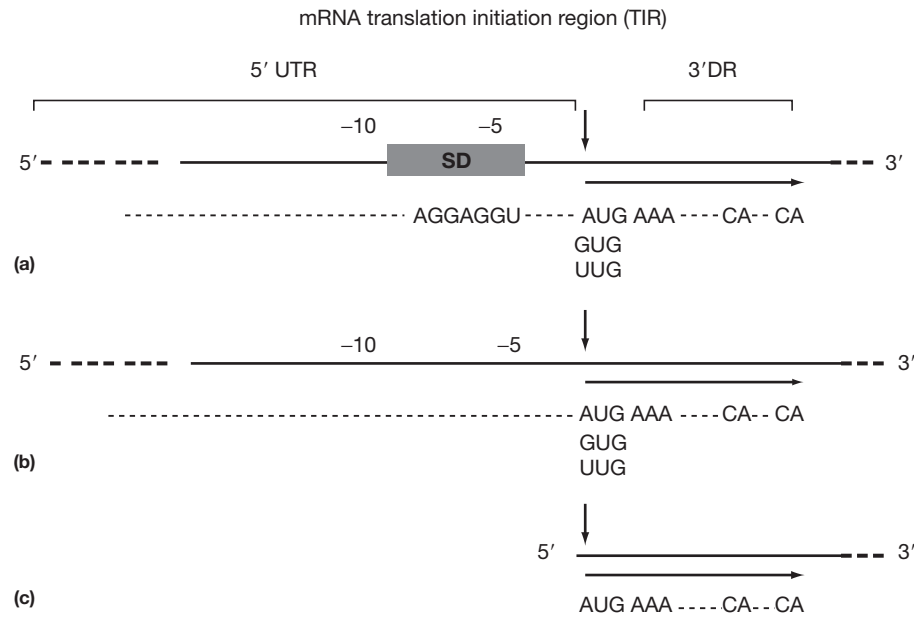


Figure 1 Schematic representation of the features characterizing the translation initiation region (TIR) of bacterial mRNAs. (a) Leadered, SD-containing mRNA. The arrows indicate the beginning of the coding region with three of the most common initiation triplets. To the left of the start point is the 5' untranslated region (5' UTR) with the approximate location of the Shine–Dalgarno sequence (SD) whose theoretically most extended consensus sequence is indicated below. To the right of the start point is the 3' downstream region (3'DR) with the most favorable second codon AAA and the repeated CA dinucleotides acting as translational enhancers in some mRNAs. (b) Leadered, SD-lacking mRNA. The characteristics of the mRNA are the same as those described above but for the absence of the SD sequence in the 5'UTR. (c) Leaderless mRNA. Leaderless mRNA completely lacks the 5'UTR and begins obligatorily with a pAUG initiation codon.

structure of the TIRs, which may favor or disfavor formation of the 30S initiation complex and very often may influence mRNA stability. In general, mRNAs, whose TIRs are devoid of secondary structure, tend to be translated with greater efficiency, although it is the optimal combination of all the aforementioned TIR elements that can maximize translation of mRNA.

Initiation Complex Formation

Aside from the leaderless mRNAs, whose translation is believed to start with monomeric 70S ribosomes, the translation of all other (leadered) mRNAs begins with the formation of a complex containing the 30S ribosomal subunit. Although the structural elements of the mRNA, such as the SD sequence of the TIRs, when present, may contribute to the thermodynamic stability of the productive 30S–mRNA interaction, thereby favoring the selection of the correct start site of the mRNA, the initiation site is selected kinetically by the 30S ribosomal subunit with the aid of the initiator fMet-tRNA whose anticodon base pairs with the initiation codon of the mRNA to form a ternary complex, the 30S initiation complex (30S IC) comprised of the 30S ribosomal subunit, fMet-tRNA, and mRNA. Both conventional and fast kinetic analyses have contributed to the elucidation of the mechanistic aspects of the translation initiation process. The steps leading to the formation of the 30S IC are schematically represented in [Figure 2\(a\)](#), while those leading to the formation of a productive 70S initiation complex (70S IC) are illustrated in [Figure 2\(b\)](#). As seen from

[Figure 2\(a\)](#), the three initiation factors bind to the 30S ribosomal subunit (steps 1, 2, 3) which contains a full complement of initiation factors when it recruits the TIR of the mRNA and the initiator tRNA. These ligands bind to the 30S subunit in stochastic order (steps 4, 5' and 5, 4') forming first two binary complexes and then an unstable pre-ternary complex in which both mRNA and fMet-tRNA are 30S-bound without interacting with each other. A first-order isomerization of this pre-ternary complex kinetically controlled by the three initiation factors (step 6), possibly followed by locking (step 7), causes the mRNA start codon to base pair in the P-site of 30S with the anticodon of fMet-tRNA to yield a 30S IC. The specific role played by the individual initiation factors in this as well as in other steps of initiation will be described later in the article. The 70S IC is generated by the joining of the large (50S) ribosomal subunit (step 8), a process which induces a conformational change in the 30S subunit and thereby causes the ejection of IF3 (step 9) and IF1 (step 10). The intrinsic guanosine triphosphatase (GTPase) activity of IF2 is also triggered in this step to generate an IF2–GDP complex (GDP, guanosine diphosphate) and inorganic phosphate. The latter is then released from the ribosome in a step (step 11) which possibly entails also the dissociation of IF2–GDP. The dissociation of IF2 leaves fMet-tRNA in the ribosomal P-site with the acceptor end near the peptidyl transferase center of the 50S subunit. Subsequently, the binding and the adjustment of the cognate EF-Tu-GTP-aa-tRNA to the ribosomal A-site (step 12) lead to the formation of the initiation dipeptide with the P-site-bound fMet-tRNA. The formation of the initiation dipeptide is a fairly slow process compared to the rate of transpeptidation during elongation.

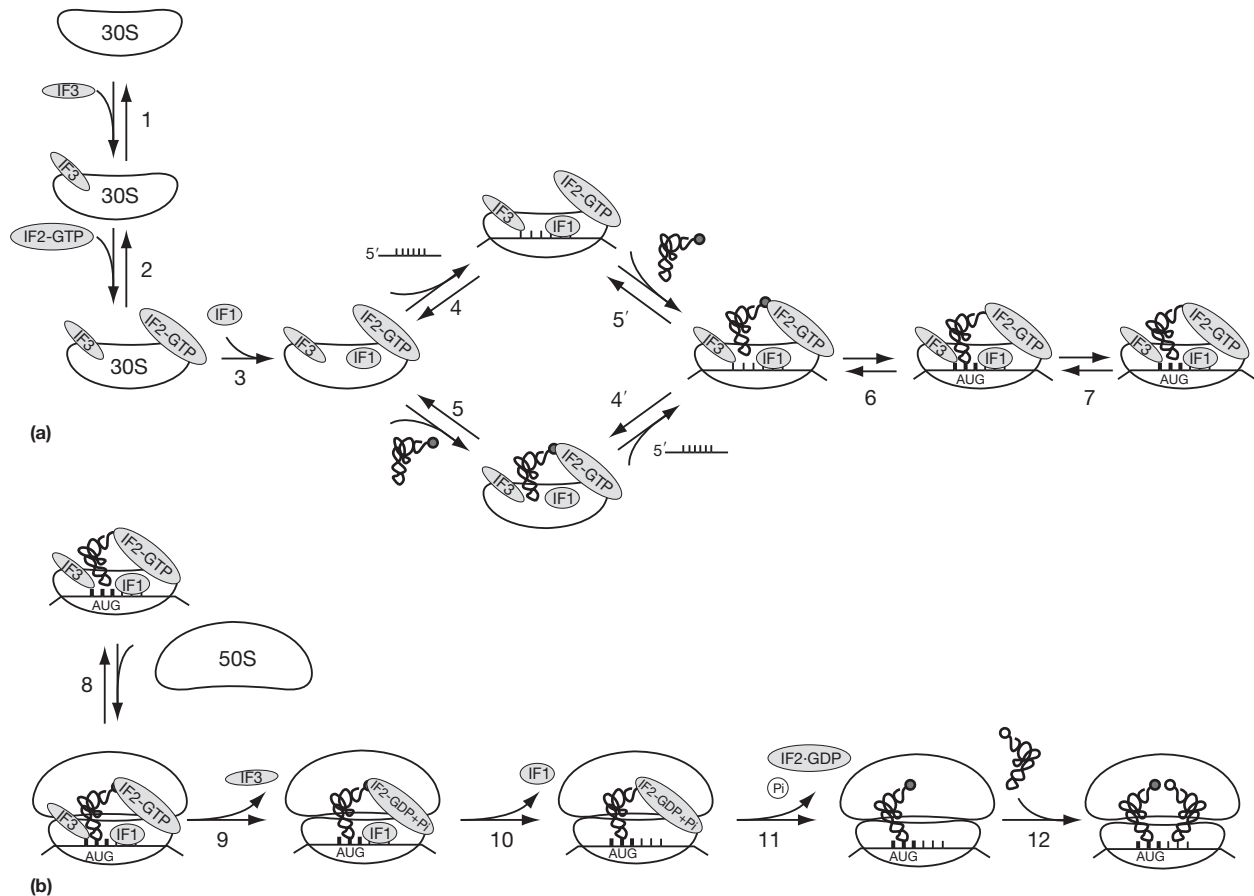


Figure 2 Translation initiation pathway in bacteria. Scheme depicting the steps leading to the formation of (a) the 30S initiation complex and (b) the 70S initiation complex as determined through traditional and fast kinetics analyses. Further details can be found in the text.

The structures of both 30S IC and 70S IC have been obtained from three-dimensional (3D) reconstructions of the cryo-electron microscopic images and are shown in [Figure 3](#).

Structure and Function of the Initiation Factors

Overall, the three initiation factors ensure both speed and accuracy in the initiation phase of protein synthesis. The expression of the genes encoding the initiation factors is regulated so that the relative level of these proteins with respect to each other remains constantly equimolar and their level remains constant at approximately 15% of the ribosomes. Only cold shock causes an imbalance of this ratio because while synthesis and assembly of the ribosomes slows down the genes encoding the factors, which are among the cold-shock genes, are actively expressed.

The structural properties and specific roles of the initiation factors are outlined next.

IF1

Structure

This protein, encoded by *infA*, consists of ~70 amino acids (71 in *Escherichia coli*) and its structure consists of a rigid

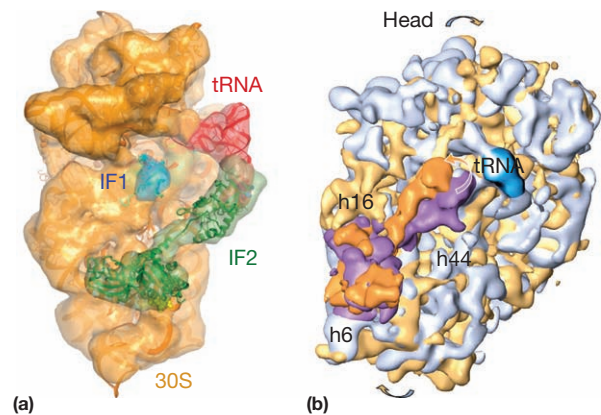


Figure 3 Structure of the ribosomal translation initiation complexes obtained by cryo-EM. (a) Positions of P-site bound fMet-tRNA, IF2, and IF1 in the 30S initiation complex and (b) position occupied by the acceptor end of fMet-tRNA and IF2 before and after GTP hydrolysis in the 70S initiation complex.

five-stranded β -barrel from which protrude the short, disordered, and highly flexible N- and C-terminal tails ([Figure 4 \(a\)](#)). This structural motif, known as the 'OB fold', is characteristic of a class of proteins that interact with oligosaccharides or with oligonucleotides and RNA, as is the case of IF1.

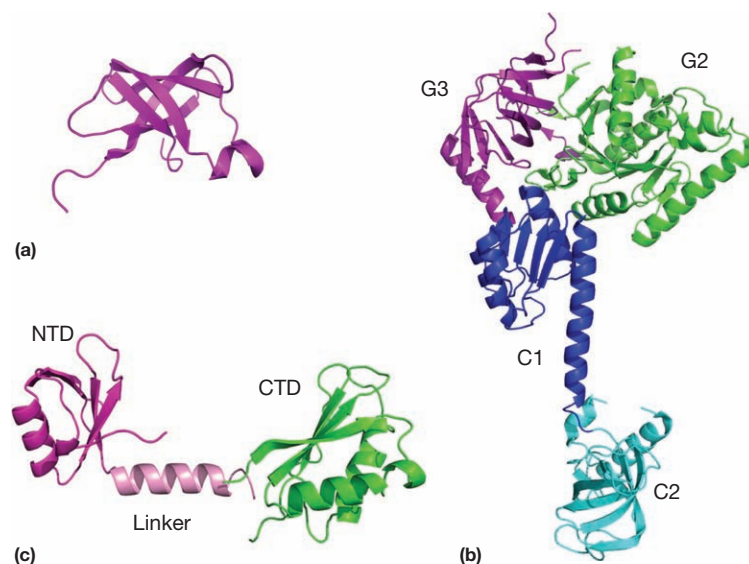


Figure 4 Three-dimensional structures of the initiation factors. (a) β -Barrel structure of *E. coli* IF1 in solution as determined by NMR spectroscopy; (b) predicted modular structure of bacterial IF2 based on the crystal structure of its archaeal homolog (eIF5b) and on the NMR solution structure of *Bacillus stearothermophilus* IF2C1 and IF2C2. The structures of the N-terminal domain and subdomain G1 are not known and these parts of the molecule are not shown. The nomenclature of the domains and subdomains is that accepted for the bacterial IF2; (c) crystallographic structure of *B. stearothermophilus* IF3 displaying the N domain (NTD), the linker, and the multifunctional C-domain (CTD).

Topographical localization and function

IF1 binds to the 30S ribosomal subunit mainly through electrostatic interactions involving the positively charged surface of the protein and the phosphate backbone of specific regions of 16S rRNA. Several lines of evidence, including crystallographic data, show that IF1 binds in the A-site of the 30S ribosomal subunit (Figure 3(a)), more precisely in the cleft between ribosomal protein S12, the 530 loop, and helix 44 of 16S ribosomal RNA (rRNA). IF1 establishes contacts with two functionally important bases (A1492 and A1493) belonging to this rRNA helix. There is also compelling evidence that IF1 binding induces several localized and long-range changes of the 30S structure, inducing an overall rotation of head, platform, and shoulder of the 30S subunit toward the A-site. It is likely that it is through the induction of these conformational changes in the small ribosomal subunit that IF1 increases the rate of 30S initiation complex formation and increases the rates of association/dissociation of ribosomal subunits, the latter activity being probably very important in favoring the ribosome dissociation activity of IF3 (Figure 5(c)).

In light of its A-site localization, it has been suggested that IF1 might block the premature access of aminoacyl-tRNA to the A-site during 30S initiation complex formation, thereby contributing to the fidelity of translation initiation. Furthermore, recent data have shown that the translation initiation fidelity function of IF3 strictly requires the presence of IF1 and that the effect of IF1 is counteracted by aminoglycoside antibiotics such as streptomycin. Finally, it is well established that IF1 and IF2 mutually influence their respective interactions with the ribosome and IF1 has been regarded as a potential modulator of IF2 recycling on and off the ribosomes. These effects are likely indirect and conformationally mediated by the ribosome since,

unlike the situation existing between the archaeal and eukaryal homologs of IF1 and IF2, a direct contact between these factors on the ribosomal surface is unlikely to occur in bacteria.

IF1 is dissociated from the ribosome after IF3, during the transition from 30S IC to 70S IC (Figure 2(b), step 10).

IF2

Structure

Bacterial IF2, encoded by *infB*, is the largest initiation factor (890 residues in *E. coli*) and consists of three major parts: (1) a variable N-terminal region; (2) a highly conserved 40 kDa G-domain consisting of three subdomains, G1, G2, and G3; and (3) the C-terminal domain (25 kDa) which consists of subdomains C1 and C2. So far no 3D structure has been directly determined for the whole bacterial IF2 molecule, and structural information concerning this factor relies on the nuclear magnetic resonance (NMR) structure of subdomains G2, C1, and C2 of *Bacillus stearothermophilus* IF2 and on the crystal structure of *Methanobacterium thermoautotrophicum* eIF5b (Figure 4(b)), the archaeal homolog of IF2. The overall shape of the archaeal factor resembles a chalice whose cup is constituted by G2, G3, and N-terminal part of C1, the stem by the C-terminal part of C1 and the foot by C2. It is likely that the structures of the archaeal and bacterial factors are similar but the functions of these two proteins have little in common. The hydrophilic, positively charged, and flexible N-domain is not present in the archaeal protein and in some bacterial IF2 molecules, and is dispensable for translation both *in vitro* and *in vivo* in *E. coli*. Its role is that of anchoring the factor to the 30S ribosomal subunit until the acceptor end of an fMet-tRNA molecule is captured by the C2 domain and recruited to the ribosome. Subdomain G2 is highly homologous to equivalent domains

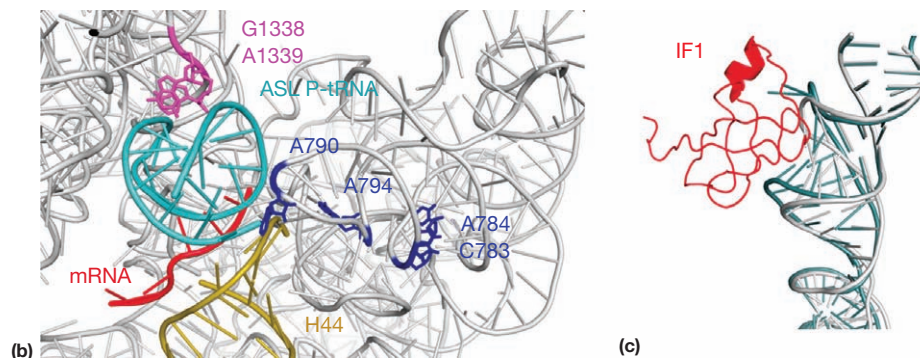
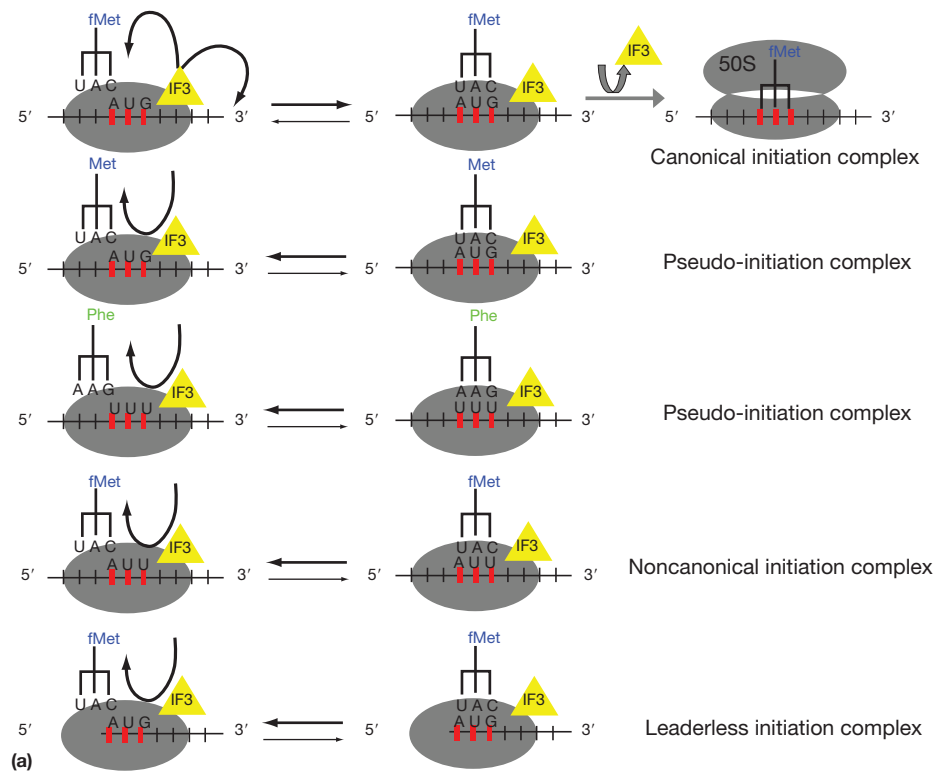


Figure 5 Role and mechanism of the initiation factors in determining translation initiation fidelity. (a) Complexes recognized as correct and incorrect by IF3. The figure presents a scheme of the various types of complexes containing 30S ribosomal subunit, template, and aminoacyl-tRNAs, which are subjected to positive or negative kinetic discrimination by IF3; (b) rRNA structure around the molecular gate separating the P-site from the E-site on the 30S subunit. Since IF3 affects the 30S conformation in the h44 helix and at or near bases A790 and G1338/A1339 present on either side of the gate, it is believed to open the P/E gate, thereby favoring the dissociation of incorrectly bound aminoacyl-tRNAs by allowing it to drift from the P- to the E-site; (c) the IF1-induced conformational change of helix 44 which is believed to play a role in allowing IF3 to perform its fidelity function. Further details may be found in the text.

of other GTP/GDP-binding proteins such as elongation factors EF-Tu and EF-G. This subdomain binds GTP/GDP, contains the GTPase catalytic center, and is also responsible for most of the IF2–50S subunit interaction. G2 consists of an eight-stranded β -sheet flanked by six α -helices and a 3_{10} helix. It also contains the four conserved sequence elements characteristic of GTP-binding proteins (G1/P loop, which participates in phosphate binding, the G3 and G4 loops, forming the walls of a hydrophobic pocket where the guanine moiety of GTP or GDP

is bound and G4). The switch 2 region of this subdomain is centrally located in the cup of the chalice where it makes extensive contacts to subdomains G3 and C1 and can therefore cause conformational changes resulting in profound global rearrangements of the molecule that could reach as far as 90 Å to the C-terminal region of IF2. The G3 domain is responsible, together with C1, for the IF2–30S subunit interaction. Its β -barrel structure is very similar to that of subdomain C2. Three β -strands of this domain interact closely with G2 in the vicinity of the switch

2 region. A 17-residue-long α -helix connects G3 to subdomain C1 which is characterized by a unique α - β - α sandwich fold consisting of a four-stranded parallel β -sheet flanked on both sides by two α -helices. In the archaeal aIF5b, the last 40-Å-long α -helix (H12) of C1 extends from the cup to subdomain C2, thus forming the stem of the chalice. However, several lines of evidence indicate that such a long, continuous α -helix is not present in bacterial IF2. Subdomain C2 is responsible for the specific recognition and binding of the acceptor end of fMet-tRNA and is endowed with a structure which also consists of a β -barrel fold, similar to that of G3 and to domain II of EF-Tu and EF-G.

Topographical localization and function

IF2 is the only one of the three initiation factors displaying a specific and fairly high affinity for both 30S and 50S ribosomal subunits. Its binding to the 30S ribosomal subunit is initially mediated by an interaction between the ribosomal particle and established by the N-terminal domain of the factor. This is followed by a functional interaction, established by subdomain G3, which is stimulated by IF1; this interaction is also favored when GTP is bound to IF2, whereas it is inhibited by the metabolic alarmone ppGpp, whose concentration increases when the cells are nutritionally starved. The interaction of IF2 with the isolated 50S subunit, which involves helix 89, the sarcin-ricin domain (SRD), and the L11/Thiostrepton-binding region of 23S rRNA, and partly overlaps that of EF-G and EF-Tu, is sufficient to elicit its GTPase activity.

The localization of IF2 on both 30S subunits and 70S monomers has been elucidated by cryo-electron microscopy (Figures 3(a) and 3(b)). The topographical studies clearly demonstrate that both the position on the ribosome and the conformation of IF2 change during the subsequent steps of the translation initiation pathway.

The main function of IF2 is that of recognizing and recruiting the initiator fMet-tRNA to the P-site of the 30S ribosomal subunit, contributing to the formation of the 30S IC. Subsequently, IF2 stimulates the association of the 50S subunit with the 30S IC and ensures the correct positioning of fMet-tRNA in the ribosomal P-site of the 70S IC promoting the first transpeptidation reaction (formation of the initiation dipeptide). The function of the ribosome-dependent GTP hydrolysis by IF2 is difficult to pin down, but it appears that removal of the γ -Pi after GTP hydrolysis is necessary to allow IF2 to assume a conformation compatible with the dissociation of the IF2-fMet-tRNA interaction and ultimately favor the dissociation of IF2 from the ribosome.

The fact that either GTP (under optimal growth conditions) or ppGpp (during nutritional stress) can bind to subdomain G2 of IF2 to give rise to an active and an inactive form of the molecule, respectively, and that the substrate of IF2 is fMet-tRNA whose synthesis involves molecules involved in the one-carbon metabolism of the cell, raises the interesting possibility that IF2 might behave like a sensor of the metabolic state of the cell. Thus, in addition to and because of its roles in translation initiation, IF2 may function as a global regulator, linking translational activity to the transcriptional control of stable RNA synthesis by adjusting the translational rate of the cell as a function of the allowable growth rate.

IF3

Structure

IF3 is a medium-sized protein (180 amino acids in *E. coli*) encoded by *infC*. Its structure consists of two domains (N-terminal domain (NTD) and C-terminal domain (CTD)) of approximately equal mass connected by a long lysine-rich linker which, according to NMR data, should be at least partially unstructured and flexible (Figure 4(c)). The NTD contains a globular α/β fold consisting of a single α -helix, packed against a mixed four-stranded β -sheet. The CTD possesses a two-layered α/β sandwich fold, comprising a four-stranded β -sheet which is packed against two parallel α -helices in a $\beta\alpha\beta\alpha\beta\beta$ topology. The fold of the CTD of IF3 is similar to that found in many eukaryotic RNA-binding proteins (such as U1A) and indeed this domain interacts with the 30S subunit via a protein-RNA interaction involving primarily structural elements of this domain such as strands β -7 and β -9 that contain consensus RNP motifs and two loops (L7 and L8). The two domains of IF3 interact, independently of each other, with different sites of the 30S subunit and do not interact with one another on the ribosomal surface. Several lines of evidence demonstrate that IF3 remains bound to the 30S IC until the association of the latter with the 50S subunit promotes the dissociation of the factor. It is noteworthy in this connection that the isolated CTD of IF3 not only binds to 30S subunit but is also capable of performing all the other known functions of the intact molecule while isolated NTD has no autonomous function. However, since the affinity of the CTD of IF3 for the 30S subunits is approximately two orders of magnitude lower than that of the intact molecule and since secondary contacts established by the NTD stabilize the interaction, it has been suggested that the two-domain structure is required to modulate binding and release of IF3 to and from the 30S subunit. The conformational change of the ribosome induced by subunit association would push the platform and head of the 30S subunit away from each other, thereby widening the gap between the NTD- and the CTD-binding sites. The loss of the stabilizing interaction established by the NTD would then facilitate the dissociation of IF3 from the 30S subunit while the 30S-50S inter-subunit bridges are formed.

Topographical localization and function

Whereas there is good agreement that the CTD is localized on the platform of the 30S ribosomal subunit, the ribosomal localization of the NTD is more controversial, although it seems likely that its binding site is located somewhere near the P-site of the particle.

Binding of IF3 to the 30S ribosomal subunit interferes with subunit association, thereby shifting to the left the equilibrium $30S + 50S \rightleftharpoons 70S$. In turn, this increases the pool of free 30S amenable to initiate a new round of translation. Concerning the other functions of IF3, contrary to earlier interpretations, this factor is not required for mRNA binding to the ribosome, but promotes instead a repositioning of 30S-bound mRNA, shifting it from the stand by site to the P-decoding site of the subunit. Furthermore, IF3 increases the on-rate of pre-ternary complex isomerization, which leads to 30S initiation complex formation (Figure 2(a), step 6) and thereby stimulates overall mRNA translation. However, since IF3 also

increases the off-rate of the isomerization, it can promote initiation fidelity favoring the dissociation of both aminoacyl-tRNA and mRNA from incorrectly formed 30S complexes. In fact, IF3 promotes the dissociation of the 30S IC with different rates depending on whether the 30S complex is a canonical, a noncanonical, a leaderless, or a pseudo initiation complex and is in kinetic competition with the association of the same 30S IC with the 50S ribosomal subunit to yield a 70S IC (**Figure 2(b)**). Indeed, it has been shown that whenever the 30S IC is non-canonical or non-best-fit, both the association between the ribosomal subunits and the dissociation of IF3 are much slower. The 30S ICs recognized and accepted as correct by IF3 and those kinetically discriminated against as being incorrect or noncanonical are schematically represented in **Figure 5(a)**. As to the molecular mechanism by which non-best-fit complexes are dissociated, it is clear that this is mediated by an IF3-induced conformation change of the 30S ribosomal subunit which affects the molecular gate separating the P-site from the E-site. The opening of this gate by IF3 would cause the improperly bound aminoacyl-tRNA to drift from the P- to the E-site and dissociate (**Figure 5(b)**). The dissociation would also be favored by the conformational changes caused by IF3 and IF1 on helix44 of 16S rRNA (**Figures 5(b)** and **5(c)**).

See also: **Molecular Biology:** Messenger RNA Degradation in Bacteria; Pre-tRNA and Pre-rRNA Processing in Bacteria; T7 RNA Polymerase.

Further Reading

- Boelens R and Gualerzi CO (2002) Structure and function of bacterial initiation factors. *Current Protein and Peptide Science* 3: 107–119.
- Gualerzi CO, Brandi L, Caserta E, et al. (2001) Initiation factors in the early events of mRNA translation in bacteria. *Cold Spring Harbor Symposium on Quantitative Biology* 66: 363–376.
- Gualerzi CO, Fabbretti A, Brandi L, Milon P, and Pon CL (2010) Role of the initiation factors in mRNA start site selection and fMet-tRNA recruitment by bacterial ribosomes. *Israel Journal of Chemistry* 50: 80–94.
- Laursen BS, Sorensen HP, Mortensen KK, and Sperling-Petersen HU (2005) Initiation of protein synthesis in bacteria. *Microbiology and Molecular Biology Reviews* 69: 101–123.
- Moll I, Grill S, Gualerzi CO, and Blasi U (2002) Leaderless mRNAs in bacteria: Surprises in ribosomal recruitment and translational control. *Molecular Microbiology* 43: 239–246.
- RajBhandary UL and Chow CM (1995) Initiator tRNAs and initiation of protein synthesis. In: Söll D and RajBhandary UL (eds.) *tRNA-Structure, Biosynthesis, and Function*, pp. 511–528. Washington, DC: ASM Press.
- Ramakrishnan V (2002) Ribosome structure and the mechanism of translation. *Cell* 108: 557–572.

Translation Initiation in Eukaryotes: Factors and Mechanisms

T V Pestova, Moscow State University, Moscow, Russia

C U T Hellen, State University of New York Downstate Medical Center, Brooklyn, NY, USA

Published by Elsevier Inc.

This article is reproduced from the previous edition, volume 4, pp. 237–241, © 2004, Elsevier Inc.

Glossary

Eukaryotic initiation factor A protein that acts in one or more steps in the process of translation initiation.

Initiator transfer RNA (tRNA) The anticodon of initiator transfer RNA is complementary to the AUG initiation codon and its structural properties differentiate it from tRNAs that decode AUG triplets during elongation. Initiator tRNA is activated by covalent linkage to methionine to form methionyl-tRNA, which is the substrate used by ribosomes to initiate protein synthesis.

Messenger RNA (mRNA) The RNA template that is translated by ribosomes to synthesize a protein.

Its coding sequence comprises consecutive triplet codons that are decoded by base pairing with the anticodon of aminoacylated tRNAs, and that therefore determine the amino acid sequence of the resulting protein.

Ribosome The complex macromolecule that catalyzes mRNA template-directed protein synthesis. Its two subunits both consist of ribosomal RNA and proteins. The 40S subunit binds mRNA and the anticodon end of tRNA; the 60S subunit aligns the aminoacyl ends of tRNAs and catalyzes peptide bond formation.

Translation initiation, the first stage in protein synthesis, is the process of assembly of large (60S) and small (40S) ribosomal subunits to form an 80S ribosome containing initiator transfer RNA (tRNA) (Met-tRNA^{Met}) that is base paired to the initiation codon of a messenger RNA (mRNA) in the ribosomal peptidyl (P) site. This process is mediated by at least 11 eukaryotic initiation factors (eIFs) and proceeds via the sequential formation of intermediate complexes. Initiation is the rate-limiting step of translation, and is a major focus for pathways that regulate gene expression.

The Structure of Eukaryotic mRNAs

Nearly all eukaryotic mRNAs have a 5'-terminal 7-methylguanosine (m⁷G) cap and a 3'-terminal poly(A) tail that synergistically enhance the efficiency of translation initiation. Most eukaryotic mRNAs contain a single major open reading frame (ORF) that is translated into protein, and initiation usually begins at the first AUG triplet from the 5'-end of an mRNA, which follows a 5' leader that is <100 nucleotide long. An AUG triplet can be bypassed if its context deviates from the optimal sequence GCC(A/G)CCAUGG (in which the initiation codon is underlined and the nucleotides in bold have the greatest influence), if it occurs very close to the 5'-end of an mRNA, or if the 5' leader has little secondary structure. Stable structures in the 5' leader reduce initiation efficiency whereas stable structures downstream of an AUG codon can enhance initiation at it.

The Mechanism of Translation Initiation

Translation initiation on most eukaryotic mRNAs begins with binding of Met-tRNA^{Met} to a 40S subunit, followed by ribosomal attachment at the 5'-end of an mRNA, scanning to the initiation codon and joining with a 60S subunit to form an 80S ribosome. Initiation is mediated by at least 11 eIFs

(Table 1), many of which act at multiple stages in this process (Figure 1). Initiation on a few mRNAs occurs by noncanonical mechanisms, of which the most common is 5'-end independent internal ribosomal entry.

Dissociation of Ribosomes into Free 40S and 60S Subunits

Initiator tRNA and mRNA initially bind to the 40S subunit rather than to the 80S ribosome. However, association of 40S and 60S subunits to form empty 80S ribosomes is favored under ionic conditions in the cytoplasm, and a mechanism to maintain a pool of free subunits is therefore a prerequisite for initiation. eIF1A and eIF3 shift the equilibrium between ribosomes and their subunits toward dissociation. eIF3 dissociates 80S ribosomes and, with eIF1A, prevents subunit reassociation by binding directly to free 40S subunits.

Recruitment of Initiator tRNA to the 40S Ribosomal Subunit

The initiation codon is decoded by a unique initiator methionyl-tRNA. Its sequence and structural features distinguish it from the methionyl-tRNAs that decode AUG triplets during translation elongation. These features enable eIF2 to select Met-tRNA^{Met} from the cytoplasmic pool of aminoacylated and deacylated initiator and elongator tRNAs, and also exclude Met-tRNA^{Met} from translation elongation. eIF2 is a stable heterotrimeric protein consisting of α -, β -, and γ -subunits that binds guanosine triphosphate (GTP) and Met-tRNA^{Met} to form a ternary complex. Binding of the ternary complex to a 40S subunit is strongly stabilized by eIF1A and eIF3 and yields a 43S preinitiation complex. eIF2's activity in binding Met-tRNA^{Met} is regulated; its affinity for Met-tRNA^{Met} is enhanced by prior binding of GTP whereas hydrolysis of eIF2-bound GTP toward the end

Table 1 Mammalian initiation factors

<i>Factor</i>	<i>Subunits</i>	<i>Mass (kDa)^a</i>	<i>Functions</i>
eIF1		13	Promotes scanning and the fidelity of initiation codon recognition
eIF1A		17	Ribosome antiassociation; stabilizes Met-tRNA _i binding to 40S subunit; promotes scanning
eIF2	α, β, γ	36, 39, 52	Binds GTP and Met-tRNA _i to 40S subunit; GTPase
eIF2B	$\alpha, \beta, \gamma, \delta, \epsilon$	34, 39, 50, 58, 80	Guanine-nucleotide exchange factor for eIF2
eIF3	α -I	167, 105, 99, 64, 52, 38, 35, 40, 37, 29, 25, 67	Ribosome dissociation and ribosome subunit antiassociation; stabilizes binding of the eIF2/GTP/Met-tRNA _i complex to 40S subunit; required for ribosomal binding to mRNA and scanning on the 5' leader
eIF4A		44	ATP-dependent RNA helicase/RNA-dependent ATPase
eIF4B		69	mRNA-binding cofactor for eIF4A
eIF4E		25	m ⁷ G cap-binding protein
eIF4F, eIF4E, eIF4A, eIF4G		25, 44, 176	Cap-binding complex comprising eIFs 4A, 4E, and 4G
eIF4G		176	Binds and coordinates the functions of mRNA, PABP, and eIFs 3, 4A, 4E
eIF4H		25	mRNA-binding cofactor for eIF4A
eIF5		49	GTPase-activating protein specific for eIF2
eIF5A		17	May enhance first cycle of translation elongation
eIF5B		139	GTPase; ribosome subunit joining
PABP		70	Binds the 3' poly(A) tail and promotes ribosomal binding to mRNA

^aMasses (kDa) correspond to those of human proteins and, where appropriate, to the largest isoform.

of each initiation cycle yields eIF2^γ guanosine diphosphate (GDP) which dissociates from Met-tRNA_i^{Met}, leaving it in the peptidyl (P) site of the 40S subunit. Biochemical and genetic data suggest that eIF2's γ -subunit binds Met-tRNA_i^{Met} and GTP, but despite eIF2^γ having the sequence motifs and structure characteristic of a conventional GTP-binding protein, eIF2 does not have an intrinsic guanosine triphosphatase (GTPase) activity. Hydrolysis of eIF2-bound GTP is induced by eIF5, a GTPase-activating protein specific for eIF2.

Attachment of 43S Preinitiation Complexes to mRNA

Binding of initiator tRNA to form a 43S complex is an obligatory first step before a 40S subunit can bind mRNA, which for most mRNAs is a 5' end-dependent process. Attachment of 43S complexes is enhanced synergistically by the 5'-terminal cap, the 3' poly(A) tail, and the factors that associate with these structures. The m⁷G cap is bound by eIF4E, the cap-binding subunit of the heterotrimeric eIF4F, which also contains eIF4A and eIF4G subunits. eIF4A is an adenosine triphosphate (ATP)-dependent RNA helicase that cycles in and out of the eIF4F complex. eIF4G is a large polypeptide that binds mRNA and eIF4E, eIF4A, and eIF3 and the cytoplasmic poly(A)-binding protein PABP, thereby coordinating and in some instances enhancing their activities. eIF4B and the less abundant eIF4H enhance the helicase activities of eIF4A and of eIF4F; eIF4B enhances but is not essential for ribosomal attachment to mRNA. The 3' poly(A) tail's influence is dependent on PABP which binds to it, and synergism between the cap and the 3' poly(A) tail depends on the bridging interaction of eIF4G with the m⁷G cap/eIF4E and 3' poly(A)/PABP complexes.

The eIF4F complex is required on most mRNAs to unwind the cap-proximal region to prepare it for attachment of 43S complexes. Recent studies have shown that 43S complexes can bind directly to an mRNA in the absence of eIF4F if the 5'-terminal region is completely unstructured. Ribosomal attachment likely involves protein-protein interactions between

the eIF3 component of 43S complexes and both eIF4G and eIF4B as well as direct binding of eIF3, eIF4G, and the 40S subunit to the mRNA. ATP-dependent restructuring of mRNA and ribosomal attachment to the unwound region of the 5' leader are probably coordinated by interaction of the eIF4G component of the eIF4F/PABP/mRNA complex with eIF3.

Ribosomal Scanning on the 5'-Leader

Following attachment to the 5'-terminal region of an mRNA, a 43S complex scans downstream until it locates the initiation codon, which is usually the first AUG triplet from the 5'-end. Scanning consists of ribosomal movement on the 5'-leader and inspection of it to identify the initiation codon. Recent studies have shown that ribosomal complexes containing only eIF1, eIF2, and eIF3 are intrinsically capable of ATP-independent movement on a 5'-leader if it is unstructured; eIFs 1, 1A, 4A, 4B, and 4F all contribute to the processivity of scanning. eIFs 4A, 4B, and 4F, and ATP are involved in restructuring mRNA to permit scanning and are essential for scanning if the 5'-leader contains even weak-secondary structure. Many details of the mechanism of scanning remain poorly understood, including the identity of the set of factors associated with the scanning complex, the mechanism by which helicase-mediated unwinding of structured 5'-leaders is coupled to ribosomal movement, the stage at which eIF4F dissociates from the cap and from the 43S complex and the mechanism of action of eIF1 and eIF1A.

Selection of the Initiation Codon

The scanning 43S complex stops when it encounters an AUG triplet that it recognizes as an initiation codon. This is primarily determined by base pairing between it and the anticodon of initiator tRNA, but in higher eukaryotes the sequence flanking an AUG triplet also influences its selection. Genetic analyses in yeast of mutants that permit initiation at an UUG triplet have

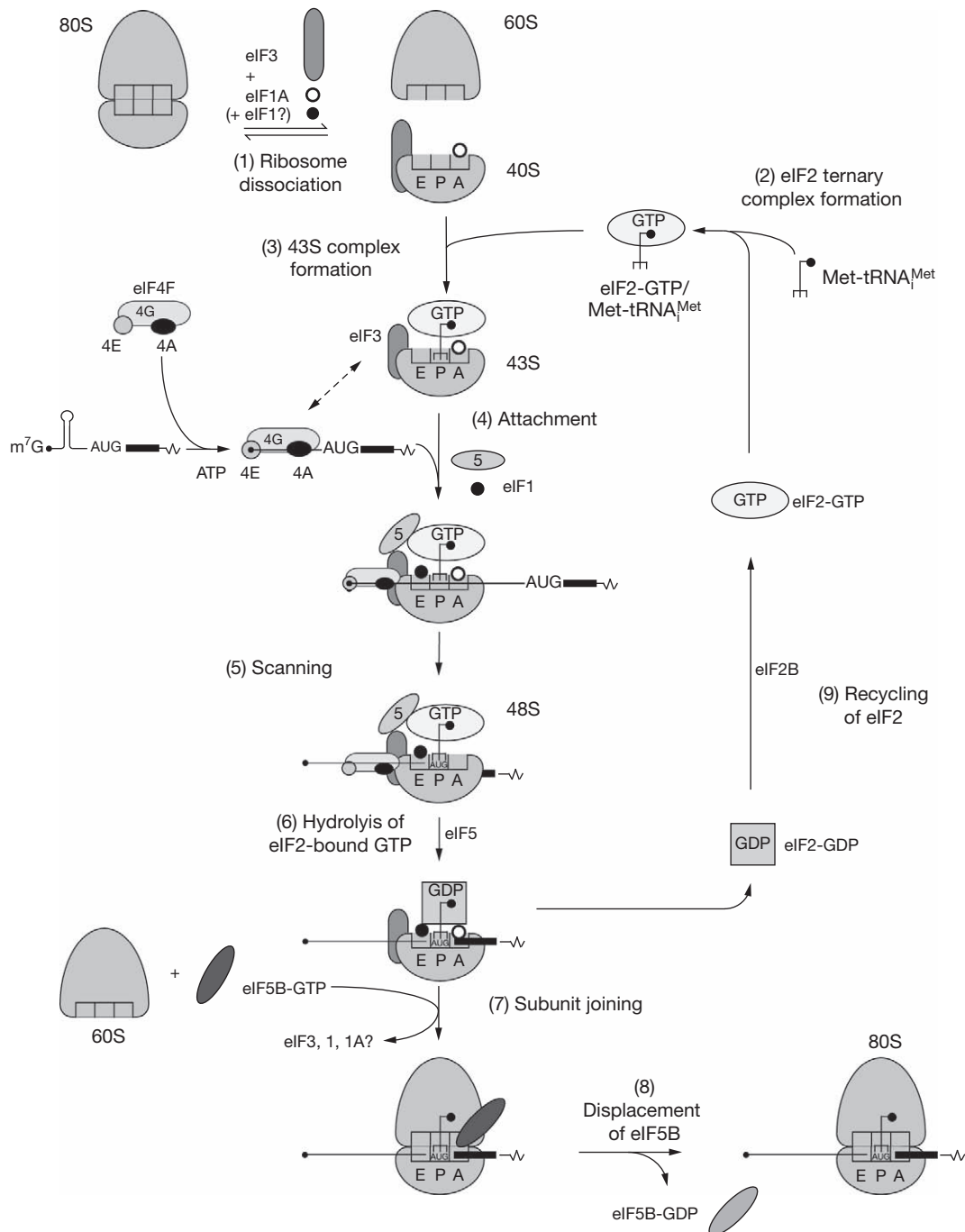


Figure 1 Schematic model of the pathway of 80S initiation complex formation on a model capped eukaryotic mRNA. eIF1A and eIF3 promote dissociation of the 80S ribosome into 40S and 60S subunits (Step 1). eIF2 binds aminoacylated initiator tRNA (Met-tRNA^{Met}) and GTP to form a ternary complex (Step 2). Binding of the ternary complex to a 40S subunit is stabilized by eIF1, eIF1A, and eIF3 to form a 43S complex (Step 3). The eIF4F complex binds to the 5'-terminal m⁷G cap of an mRNA and with associated cofactors creates an unstructured cap-proximal site to which the 43S complex binds (Step 4). This step may be enhanced by the poly(A)-binding protein associated with the mRNA's 3' poly(A) tail (not shown). The dashed, double-headed arrow indicates interactions that promote attachment of the 43S complex to mRNA, such as those of the eIF3 component of the 43S complex with the eIF4G subunit of cap-bound eIF4F and probably with the mRNA. The 43S complex scans downstream on the 5'-leader until it reaches the initiation codon (Step 5), which is base paired to the anticodon of Met-tRNA^{Met} in the resulting 48S complex. Hydrolysis of eIF2-bound GTP is triggered by eIF5, and probably releases of eIF2-GDP (Step 6). The stages at which eIF5 joins and eIF1, eIF1A, and eIF3 are released from the 40S subunit are not known. The GTP-bound form of eIF5B mediates joining of a 60S subunit to the resulting complex (Step 7). eIF5B-GDP is released after hydrolysis of eIF5B-bound GTP, which is induced by both ribosomal subunits. The resulting 80S ribosome is then able to begin protein synthesis (Step 8). eIF2B recycles inactive eIF2-GDP to active eIF2-GTP, which can then bind Met-tRNA^{Met} again (Step 9).

indicated that eIF1, eIF2, and eIF5 all influence start site selection. Recent studies have shown that eIF1 plays the principal role in maintaining the fidelity of initiation codon selection, for example, by destabilizing complexes aberrantly arrested at non-AUG triplets or assembled on AUG triplets that have a poor context.

Displacement of eIF2 from the 48S Complex

Establishment of base pairing between the initiation codon and the anticodon of Met-tRNA_i^{Met} stimulates eIF5-induced hydrolysis of eIF2-bound GTP in ribosomal complexes. eIF2-GDP does not bind Met-tRNA_i^{Met} or the 40S subunit and is released from the 48S complex, leaving Met-tRNA_i^{Met} base paired to the initiation codon. eIF5-induced hydrolysis of eIF2-bound GTP is involved in conversion of the scanning ribosomal complex to a complex that is arrested at the initiation codon. Mutations in eIF5 or eIF2 that alter the rate of this reaction therefore influence selection of the initiation codon. eIF2 has a much higher affinity for GDP than GTP, and GDP has a slow off-rate from eIF2, so eIF2B, a guanine nucleotide exchange factor specific for eIF2 is required to regenerate active eIF2-GTP from inactive eIF2-GDP.

60S Subunit Joining to Form an 80S Ribosome

The release of eIF2-GDP from 48S complexes that is induced by eIF5 is not sufficient to permit 60S subunits to bind to the 40S subunit/*big*/Met-tRNA_i^{Met}/*big*/factor complex assembled at the initiation codon. Subunit joining also requires eIF5B, which is a homolog of the prokaryotic initiation factor 2 and, like it, has a GTPase activity that is maximally stimulated by large and small ribosomal subunits. Subunit joining leads to the release of all initiation factors from the 40S subunit and leaves Met-tRNA_i^{Met} in the ribosomal P site. The GTP-bound form of eIF5B is active in promoting subunit joining but does not dissociate from the ribosome and instead blocks the ribosomal A site. GTP hydrolysis leads to the release of eIF5B-GDP from the 80S ribosome, so that it is able to begin polypeptide synthesis. Translation initiation therefore requires hydrolysis of two GTP molecules in reactions catalyzed by eIF2/eIF5 and eIF5B, respectively. The initial round of elongation may be enhanced by eIF5A.

Regulation of Translation Initiation

The intrinsic efficiency of translation initiation on an mRNA is determined by properties such as the degree of secondary structure in the 5'-leader and the sequence context of the initiation codon. Differences in these properties account for many of the differences in the relative levels of translation of different mRNAs. Translation initiation is also regulated either selectively on a single species or on a small subset of mRNAs, or globally on all mRNAs, in order to integrate protein synthesis with physiological demands. Translation of a single species of mRNA or of a related group of mRNAs can be

repressed by binding of a protein to a specific site on the mRNA in a manner that interferes with initiation. For example, the iron regulatory protein that binds to the 5' leader of ferritin mRNA sterically prevents it from binding to the 43S complex. More commonly, translation is regulated at a more global level by alteration of the activities of initiation factors, either as a result of phosphorylation or by binding to regulatory proteins. Phosphorylation of eIF2 α causes eIF2 to bind to and inhibit eIF2B, ultimately reducing formation of the eIF2/GTP/Met-tRNA_i^{Met} complex and thus downregulating initiation globally. Initiation on most mRNAs is cap mediated and is downregulated by a reduction in the level of active eIF4F, which can occur either by proteolysis of eIF4G (e.g. during some viral infections) or by disruption of eIF4E's interaction with eIF4G by eIF4E-binding proteins that compete with eIF4G for binding to eIF4E.

Initiation of Translation by Internal Ribosomal Entry

Translation on most eukaryotic mRNAs is initiated by end-dependent ribosomal scanning, but initiation on a few viral mRNAs is end independent and is instead mediated by an internal ribosomal entry site (IRES) that promotes binding of the 40S subunit to an internal site in the mRNA without scanning from the 5'-end. IRESs are in general large and contain significant secondary structure. IRESs from a single virus family are similar, but the size and structure of IRESs from unrelated virus families differ greatly from each other. Three groups of IRESs have been characterized in detail; each mediates initiation by a different mechanism but all involve direct, noncanonical interactions of the IRES with canonical components of the translation apparatus. They all have simpler initiation factor requirements than cap-mediated initiation, and therefore escape some mechanisms that regulate that process. IRESs have been identified in several cellular mRNAs, but little is known of the mechanisms by which they promote initiation.

See also: [Molecular Biology: mRNA Polyadenylation in Eukaryotes; Pre-tRNA and Pre-rRNA Processing in Bacteria; Ribosome Assembly; Ribosome Structure.](#)

Further Reading

- Dever TE (2002) Gene-specific regulation by general translation factors. *Cell* 108: 545–556.
- Gingras A-C, Raught B, and Sonenberg N (1999) eIF4 initiation factors: Effectors of mRNA recruitment to ribosomes and regulators of translation. *Annual Review of Biochemistry* 68: 913–963.
- Hellen CUT and Sarnow P (2001) Internal ribosomal entry sites in eukaryotic mRNA molecules. *Genes and Development* 15: 1593–1612.
- Kozak M (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. *Journal of Biological Chemistry* 266: 19867–19870.
- Pestova TV and Kolupaeva VG (2002) The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes and Development* 16: 2906–2922.
- Sonenberg N, Hershey JWB, and Mathews MB (2000) *Translational Control of Gene Expression*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Tricarboxylic Acid Cycle

R L Veech, National Institutes of Health, Rockville, MD, USA

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 4, pp. 256–262, © 2004, Elsevier Inc.

Glossary

TCA cycle The tricarboxylic acid cycle.

ΔG° The standard free energy of a reaction, a constant where the products and reactants are 1 M in activity or more practically concentration.

$\Delta G'$ The free energy of a reaction that is not constant but varies with the ratio of activity of concentration of

products divided by reactions and is defined by the formalism

$$\Delta G' = \Delta G^\circ + RT \ln \frac{[\text{products}]}{[\text{reactants}]}$$

The Discoverer of the Cycle

The tricarboxylic acid (TCA) cycle was first correctly formulated by Han Krebs, who was born on 25 August 1900 in Hildesheim. He was the son of a surgeon, Georg, and mother Alma, who was to die of depression when Krebs was 19. After completing gymnasium in Hildesheim in September 1918, Krebs joined the German army signal corps until the armistice of 11 November 1918. After Completing his training in internal medicine at Gottingen, Freiburg, Munich, and Berlin in December 1925, Krebs looked for ways to obtain biochemical training. The time spent in the clinics had convinced him that it was necessary to pursue medical research. Through the influence of Bruno Mendel, who served with Krebs in the Third Medical Clinic in Berlin, and who was an acquaintance of Albert Einstein, professor of Physics at the University of Berlin, who was in turn a frequent dinner guest at the home of Otto Warburg's father, Mendel heard that Warburg was looking for an assistant and recommended Krebs. This circle of influence allowed Krebs to obtain his first paid position as an assistant to Otto Warburg at the Kaiser Wilhelm Institut für Biologie at Berlin-Dahlem (see [Figure 1](#)). This stint was to change Krebs from a physician to a biochemist. Of Warburg, Krebs wrote (1981):

Otto Warburg was the most remarkable person I have ever been closely associated with. Remarkable as a scientific genius of the highest caliber, as a highly independent, penetrating thinker, as an eccentric who shaped his life with determination and without fear, according to his own ideas and ideals.

The Kaiser Wilhelm Institute was the leading scientific research facility of its time. In addition to Warburg, its faculty included Emil Fischer, Richard Willstater, Max von Laue, Fritz Haber, Otto Hahn, Lisa Meitner, and Michael Polanyi. Otto Meyerhof, a former student of Warburg, was working in the same building and elucidated the glycolytic pathway responsible for the breakdown of glucose to lactate. Students of biochemistry who passed through Kaiser Wilhelm included, among others, Fritz Lipmann and Severo Ochoa, both of whom were to play a critical role in showing the role of coenzyme A in the TCA cycle. In addition to the discipline

and rigor required for scientific research, Krebs learned from Warburg the techniques he had developed: use of tissue slices, spectrophotometry and manometry. These techniques, most notably manometry, were fundamentally important in leading Krebs to formulate the concept of a metabolic cycle, first the urea cycle and later the TCA cycle. It is important to recognize in understanding Krebs' work that the Warburg dictum "methods are everything" was dominant. There was no theoretical master plan. Only during his later work on the thermodynamics linking of cellular redox and phosphorylation states would Krebs remark in relation to that work, "This is the first time I ever thought you could predict in biology." In his early work, Krebs was a thoroughly empiricist. He could measure the rates of reaction. He added various potential participants in a metabolic pathway, and if they increased the overall rate, then he concluded they were part of the pathway. Krebs' training is of central importance to understand the cycle. As his biographer Holmes has pointed out, "Hans Krebs took 31 years to become a well-trained independent scientific investigator, and only 9 months to make one of the most significant discoveries of his generation in his chosen field."

History of the Discovery of the TCA Cycle

After leaving Warburg's lab in 1931, Krebs obtained a position as chief medical resident in the university hospital at Freiburg, and was in charge of caring for 44 patients. With his Warburg manometer and the enzyme urease, he could measure the production of urea in liver slices. In his off-time working by himself with the assistance of a medical student, Kurt Henseleit, within 1 year, he had formulated a salts solution to mimic the composition of plasma, Krebs-Henseleit saline, the basis for modern tissue culture fluids, and had worked out the ornithine cycle responsible for the hepatic synthesis of urea from ammonia. This established for the first time that cyclic, not linear, pathways could play an important role in central metabolic processes. The process of this discovery is the subject of a detailed history dissecting the inventive processes entailed.

One of the great unsolved problems was to determine the reactions whereby foodstuffs undergo combustion to produce



Figure 1 Hans Krebs at the age of 68 in his laboratory at Oxford standing next to his beloved Warburg apparatus, which he learned to use in the laboratory of Otto Warburg and which was the major instrument used in his discovery of both the urea and the citric acid cycles. The square box in the background is a power supply for a Zeiss PMQ spectrophotometer. The spectrophotometer was first devised by Otto Warburg who also first discovered pyridine nucleotide and observed their spectral shift during oxidation/reduction. This spectral change formed the basis of many of the assays Krebs undertook in his later years. With these simple instruments and a devoted staff he worked happily in this building at the Radcliffe Infirmary until he was 81, when he died in a hospital ward in the same building in 1981 after a 2-week illness.

CO₂ and water. It was known that for glucose to undergo these reactions, it must first be broken down to pyruvate in the glycolytic pathway elucidated by Otto Meyerhof. Although in retrospect, it seems logical to assume that the idea of metabolic cycles would have been foremost in Krebs' mind when he began his investigation into the oxidative combustion of foodstuffs, a careful analysis of his writing and notebooks gives no support for this idea. Rather, his formulation of the TCA cycle was the result of the empirical methods learned in Warburg's lab. From Albert Szent-Gyorgyi, he had pigeon breast muscle minces which preserved intact, the then unknown mitochondria containing the enzymes of the TCA cycle, and could oxidatively decompose glucose to CO₂. From Warburg he had a manometer with which he could measure rates of CO₂ production. Again, he added possible intermediates and observed whether the rates of reactions increased. As he put it in his autobiography:

One way of tackling the problem of the intermediate steps of the combustion process is to test which substances, apart from carbohydrate and fat, burn most readily. The logic is that if a substance is an

intermediate then it must readily undergo combustion: if it proves to be non-combustible, it cannot be an intermediate.

In two sentences was his simple plan: the same as he had used in elucidating the urea cycle. There was no grand plan. In choosing the intermediates to be added, he was guided by others working on the same general problem. Earlier Szent-Gyorgyi had shown that addition of small amounts of succinate, fumarate, malate, or oxaloacetate could catalytically increase the rate of oxidation by pigeon breast muscle minces. Szent-Gyorgyi interpreted his findings as indicating that these compounds were not intermediates in a pathway, but rather that they served as hydrogen carriers transporting the H⁺ and electrons from foodstuffs to cytochromes. By 1937, experiments by Martius and Knoop, Krebs' teacher at Freiburg, had shown in liver that citrate was converted to aconitate, then isocitrate and α -ketoglutarate, which in turn was known to be converted to succinate. It was also from Thunberg's work that malonate could inhibit the conversion of succinate to fumarate. This presented a way for Krebs to separate the reactions of Carl Martius and Franz Knoop from those of Albert Szent-Gyorgyi. Krebs then hit on the crucial question:

So I asked myself whether perhaps oxaloacetate, together with a substance derived from foodstuffs, might combine to form citrate again, after the manner of a cycle.

Krebs began his 'Results' section on the discovery of the tricarboxylic acid cycle with the experiments or data showing the catalytic effects of citrate on respiration in pigeon breast minces. His experimental results confirmed his hypothesis. In his mind this cycle could explain the complete combustion of foodstuffs to CO₂ and water.

Into the Wilderness and Back Again

After his triumph in discovering the urea cycle in 1932 and having been invited by Max Plank to discuss his findings, in the same year, on 19 June 1933, he was forced into exile from his own homeland, to which he and his father were devoted and whose army he had served, for reasons of racial identity which he had rejected. He found refuge in Hopkins Biochemistry Laboratory at Cambridge, dependent upon the kindness of strangers.

Triumph would again be followed by rejection, after Krebs' formulation of the TCA cycle. This time, however, the rejection would not be by the Nazi's who took over his homeland, but rejection would come from his scientific colleagues. A revolutionary theory always has its critics. Krebs' tricarboxylic acid cycle, published in 1937 after being rejected by *Nature*, was no exception. By 1940, Earl Evans, who had been trained in Krebs' lab at Sheffield, and Harland Wood, one of America's most distinguished biochemists then working at Iowa State, produced evidence that when C-labeled CO₂ condensed with pyruvate to form oxaloacetate it yielded α -ketoglutarate with the entire radioactivity confined to the carboxyl group adjacent to the carbonyl group. Wood suggested that this radioactive evidence was compatible with the condensation of pyruvate with oxaloacetate forming aconitate, then isocitrate

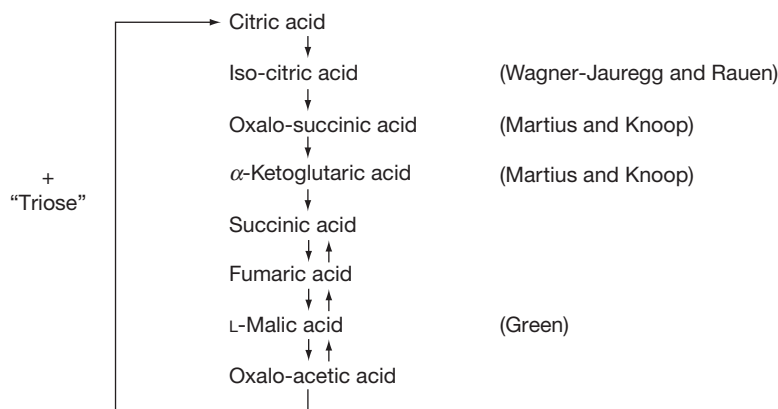


Figure 2 The original schematic of the citric acid cycle as it appeared in an article by Krebs and Johnson in 1937, after having rejected by the editor of *Nature*. Krebs attempted to define the reversible and irreversible reactions of the cycle as he knew them at the time, but which would not be accepted as correct today. The nature of the triose combining with oxaloacetate was not known until the identification of acetyl CoA by Lipmann and Ochoa in the 1950s. The essential nature of the cycle as drawn in 1937 is however, correct to this day.

and finally α -ketoglutarate at a rapid rate. If the formation of citrate from aconitate were only a slow side reaction, then this would account for the radioactive findings. For 7 years this reasoning was accepted and Krebs changed the name of his cycle from the more euphonious citric acid cycle (see [Figure 2](#)) to the more cumbersome TCA cycle, which it bears today to accommodate the generally accepted view in the biochemical community.

Krebs confined his work to practical nutritional problems for wartime Britain, which was literally starving from the depredations of Nazi submarine attacks. Then in 1948, Alexander Ogston published a short paper in *Nature* pointing out that the three-point attachment of the symmetrical citrate molecule to the citrate synthase enzyme would confer optical properties to the product of that chemically symmetrical molecule. The citric acid cycle was back in business as Krebs pointed out in his Harvey Lecture delivered on 17 March 1949. The final confirmation of the validity of Krebs' original postulate of the citric acid cycle came from Fritz Lipmann's study of the acetylation of sulfanilamides where he identified the so-called active acetate as acetyl coenzyme A. Soon thereafter, Stern, Ochoa, and Lynen showed that the intermediate through which pyruvate enters the TCA cycle was acetyl CoA which then combines with oxaloacetate to form citrate, confirming Krebs' original formulation of the cycle.

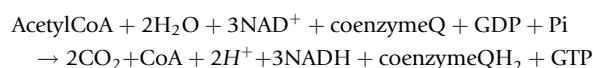
These workers also demonstrated that not only carbohydrate, but also fats and ketone bodies form acetyl CoA. This established that the Krebs citric acid cycle not only accounted for the complete conversion of carbohydrate to CO_2 and water, but also that this cycle was the terminal pathway of the degradation of all foodstuffs, including fats and amino acids (see [Figure 3](#)). It may have been with some irony that Krebs entitled his Nobel prize speech of 11 December 1953, 'the citric acid cycle' rather than the 'TCA cycle' as it was called after the radioactive data were mistakenly interpreted to indicate that aconitate, and not citrate, was the condensation product of carbohydrate degradation. Subsequently, students of biochemistry would certainly have found the original name less fearsome than the tricarboxylic acid cycle, a name that evolved from erroneous interpretation of experimental results and was perpetuated by Krebs inherent modesty.

The TCA Cycle Today

It is now accepted that the TCA cycle is not only the metabolic pathway which accounts for the complete combustion of the product of glycolysis, pyruvate, to CO_2 and H_2O that is the pathway which accounts for the complete combustion of all foodstuffs, carbohydrates, fats, and amino acids in heterotrophic organisms. Much about the control of flux through the TCA cycle remains unknown even today. One persistent question is how cyclic pathways appear to defeat the laws of thermodynamics. It is easy to understand how a linear metabolic pathway, such as glycolysis, uses the chemical energy of the reactants to proceed from A to E. It is not so clear, however, how a pathway which goes from A to A accomplishes this feat in an apparent violation of thermodynamic imperatives.

The eight enzymes catalyzing the reactions of the TCA cycle and their equilibrium constants are listed in [Table 1](#).

The overall reaction of the TCA cycle is therefore



with sum standard free energies, ΔG° at pH 7 of $-11.7 \text{ kcal mol}^{-1}$. The sum of standard free energies of the reactions of the cycle is negative. While the sum of the standard free energies of the reactions of the cycle are negative in the direction of the cycle written, there is no complete information on the actual $\Delta G'$ of the cycle *in vivo*. The actual $\Delta G'$ of a reaction is, of course, a sum of the standard free energy, ΔG° , and a term representing the actual ratio of the concentration of products over reactants.

Without knowledge of the ratio of concentrations in mitochondria of the products over the reactants of the various eight steps of the TCA cycle, no definitive statement about the reversibility of any step can be made, although most biochemical textbooks assign one or another TCA cycle as irreversible.

The synthesis of citrate from acetyl CoA and oxaloacetate, by citrate synthase, has the largest standard free energy of any reaction within the cycle and might therefore qualify for the irreversible designation. However, the distinguished

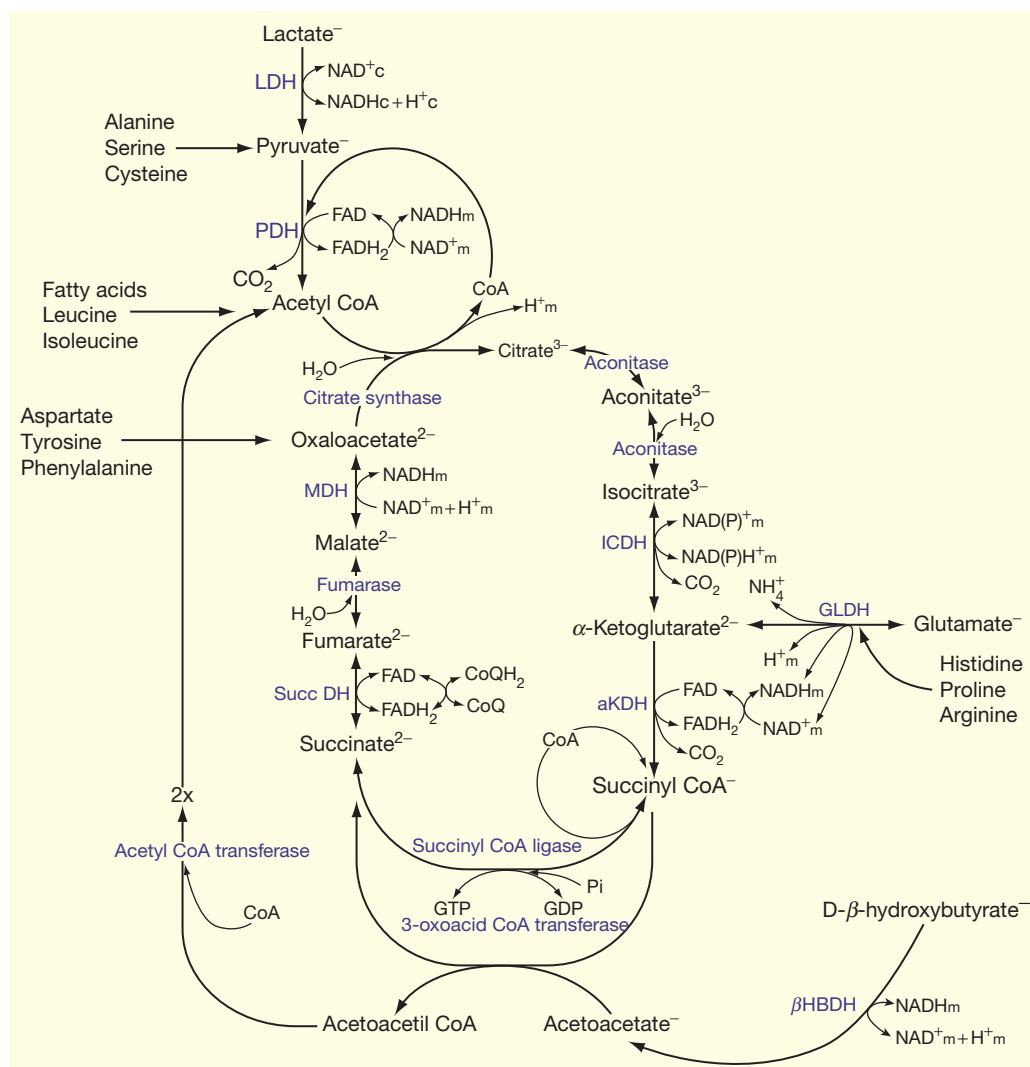


Figure 3 The TCA cycle is now recognized as the central pathway for the complete degradation of the major foodstuffs, including carbohydrate, lipid, and proteins. It also captures the reducing equivalents which are then transferred to the electron transport system of mitochondria where the energy of their redox reactions generate the mitochondrial proton gradient used to power ATP synthesis.

Table 1 The eight enzymes catalyzing the reactions of the TCA cycle and their equilibrium constants

Enzyme	EC number	Reaction	K_{eq} at pH 7.0	ΔG° at pH7 (kcal mol ⁻¹)
Citrate synthase	4.1.3.8	Acetyl CoA + oxaloacetate + H ₂ O → citrate + CoA	2.24×10^6	-9
Aconitase	4.2.1.3	Citrate ↔ isocitrate	18	-1.8
Isocitrate dehydrogenase	1.1.1.41	Isocitrate + NAD(P) ⁺ ↔ α-ketoglutarate + CO ₂ + NAD(P)H	1.17 M	-0.01
α-Ketoglutarate dehydrogenase	Enzyme complex	α-Ketoglutarate + NAD ⁺ + CoA → succinyl CoA + CO ₂ + NADH	1.17×10^3	-4.4
Succinyl CoA ligase	6.2.1.4	Succinyl CoA + GDP + Pi ↔ succinate + GTP + CoA	3.5	-0.08
Succinate dehydrogenase	1.3.5.1	Succinate + ubiquinone ↔ fumarate + ubiquinol	6.5	-1.2
Fumarase	4.2.1.2	Fumarate + H ₂ O ↔ L-malate	4.4	-0.9
Malate dehydrogenase	1.1.1.37	Malate + NAD ⁺ ↔ oxaloacetate + NADH	2.86×10^{-5}	6.5

biochemist, Paul Srere, who spent much of his career in biochemistry studying the kinetics of citrate enzymes, argued forcefully that the citrate synthase reaction achieved near equilibrium *in vivo*. Krebs himself argued that both the NAD- and

NADP-dependent mitochondrial isocitrate dehydrogenases were in near equilibrium with the free mitochondrial [NAD⁺]/[NADH] pool. A near-equilibrium reaction is hardly compatible with irreversibility. The reaction catalyzed by the

α -ketoglutarate dehydrogenase multienzyme complex could well be irreversible, but specific knowledge allowing one to come to a firm conclusion is lacking.

What is clear is that the disposal, of the reducing power produced in the reactions of the TCA cycle, mainly in the mitochondrial electron transport system critically affects the extent and possibly the direction of the reactions of the cycle. Within the last decade, it has been established that the TCA cycle exists in the earliest surviving life forms, the archaeobacteria which developed when Earth was covered in a reducing atmosphere of CO_2 and NH_3 and no O_2 . Archaeobacteria contain a TCA cycle, which was proposed to serve as a primary anaplerotic source of metabolites and amino acids, not as the primary pathway for the degradation of foodstuffs in heterotrophs. When life forms were developing, in the absence of O_2 , the direction of the TCA cycle would have been reversed.

Whether one accepts this speculation or not, it emphasizes the intimate relationship between the TCA cycle and the production of reducing power in the form of reduced pyridine nucleotide, which serves as the primary substrate for the electron transport system. When formulated in 1937, Krebs was primarily concerned with the stoichiometry of reactions accounting for the transformation of foodstuffs into CO_2 . In 1936, Otto Warburg had just discovered NADP in red cells, so the central role of these essential cofactors was not appreciated until much later when Krebs played a crucial role in defining the redox potential of the mitochondrial $[\text{NAD}^+]/[\text{NADH}]$ ratio using principles first defined Holzer, Lynen, Bucher, and Klingenberg in defining the free cytoplasmic $[\text{NAD}^+]/[\text{NADH}]$. Krebs and his colleagues then went on to show that the redox potentials of all of pyridine nucleotides were related to one another in a coherent manner. The free cytosolic NAD couple was about -0.19 V , and hence could accept reducing equivalents from glycolysis, the free mitochondrial NAD couple was a more negative -0.28 V , providing the energy necessary for mitochondrial ATP generation, while the free cytosolic NADP system was the most negative at -0.41 V , required for the completion of the reductive synthesis of fats. While it was recognized at the time that the pyridine nucleotides were related to the free $[\text{ATP}]/[\text{ADP}][\text{Pi}]$ ratio, subtleties in the effects of free $[\text{Mg}^{2+}]$ complicated this relationship and it took another 10 years to accurately determine the equilibrium constants required to quantitatively define that relationship.

In addition to accounting for the chemical decomposition of foodstuffs, it is now recognized that the TCA cycle produces reducing power in the form of three reduced pyridine nucleotides and one reduced coenzyme Q within the mitochondria. Each of these cofactors contains two electrons. Those contained in NADH feed into the mitochondrial

electron transport system at complex I, the NADH dehydrogenase multienzyme complex. The reducing equivalents contained in reduced coenzyme Q, generated in the conversion of succinate to fumarate feed into the electron transport chain at a site with a higher redox potential than that of the mitochondrial $[\text{NAD}^+]/[\text{NADH}]$ couple. Approximately four protons are ejected from mitochondria, at each site creating a $\Delta G [\text{H}^+]$ cytosol/mitochondria, which provides the energy approximately equivalent to that of the $\Delta G'$ of ATP hydrolysis in a near equilibrium reaction catalyzed by the F1 ATPase.

It was originally thought that the TCA cycle was so fundamental that any defects in the cycle would be fatal. It is now recognized that because of its central role, not only in the degradation of foodstuffs, the TCA cycle plays a central role in generating the redox potentials not only of the mitochondrial NAD couple but also of the mitochondrial Q couple. The difference in the redox potential of these two couple, in turn, sets the magnitude of the energy of the proton gradient between the mitochondrial and cytosolic phases of the cell and thus determines the energy realized when the ATP formed by the mitochondria is hydrolyzed. The amount of realizable energy formed is a function of the amount and type of the substrate being utilized by mitochondrial TCA cycle.

See also: [Metabolism Vitamins and Hormones: Coenzyme A; Glycolysis Overview.](#)

Further Reading

- Holmes FL (1991) *Hans Krebs, the Formation of a Scientific Life, 1900 to 1933*. Oxford: Oxford University Press.
- Holmes FL (1993) *Hans Krebs, Architect of Intermediary Metabolism, 1933 to 1937*. Oxford: Oxford University Press.
- Krebs HA (1954) The citric acid cycle. In: Liljestrand G (ed.) *Les Prix Nobel en 1953*, pp. 139–150. Stockholm; Kungl: Boktryckeriet; P.A. Norstedt and Soner.
- Krebs HA (1981) *Hans Krebs: Reminiscences and Reflections*. Oxford: Clarendon.
- Krebs HA and Henseleit K (1932) Untersuchungen über die Hamstoffbildung in Tierkörper. *Hoppe-Seyler's Zeitschrift für physiologische Chemie* 210: 33–66.
- Krebs HA and Johnson WA (1937) The role of citric acid in intermediate metabolism in animal tissue. *Enzymologia* 4: 148–156.
- Krebs HA and Veech RL (1969) Pyridine nucleotide interrelations. In: Papa S, Tager JM, Quagliariello E, and Slaten EC (eds.) *The Energy Level and Metabolic Control in Mitochondria*, pp. 329–382. Bari: Adriatica Editrice.
- Morowitz HJ, Kostelnik JD, Yang J, and Cody GD (2000) The origin of intermediary metabolism. *Proceedings of the National Academy of Sciences of the United States of America* 97: 7704–7708.
- Ogston AG (1948) Mechanism of fixation of carbon dioxide in the Krebs cycle. *Nature* 162: 963.
- Wood HG, Werkman CH, Hemingway A, and Nier AO (1941) Interpretation of experiments on metabolic processes using isotopic tracer elements. *Journal of Biological Chemistry* 139: 483.

tRNA Synthetases

R W Alexander, Wake Forest University, Winston-Salem, NC, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Aminoacyl-tRNA synthetases The family of enzymes that catalyze attachment of an amino acid to its cognate tRNA.

Cytokine A protein that acts as a signaling molecule, typically to activate the immune system.

Genetic code The correspondence of trinucleotide codons with amino acids used in protein biosynthesis.

Ribosome The ribonucleoprotein complex that decodes messenger RNA and catalyzes peptide bond formation.

Translation Biosynthesis of polypeptides from amino acids according to a messenger RNA template.

tRNA The small structured RNA that serves as adapter between a messenger RNA trinucleotide codon and its encoded amino acid, delivering the amino acid to the growing polypeptide chain at the ribosome.

tRNA Synthetases

Protein biosynthesis at the ribosome results in the conversion of nucleic acid genetic information into the polypeptides essential for cellular function. Peptide bond formation requires numerous protein and nucleotide cofactors in addition to the ribosome itself, and is tightly regulated to ensure accurate translation. Transfer RNA (tRNA) synthetases are key players in this synthetic pathway. Each enzyme in the family attaches a specific amino acid to its corresponding tRNA, thus defining the rules of the genetic code. Organisms use up to 20 tRNA synthetases to generate aminoacyl-tRNAs, one for each of the 20 standard amino acids. While all tRNA synthetases carry out the same fundamental chemistry, they are distinguished by their structure, regulation, and often by novel functions outside their canonical role in protein biosynthesis.

Enzymatic Reaction

tRNA aminoacylation occurs as a two-step mechanism (Scheme 1), in which both steps are catalyzed at a single enzymatic active site. In the first step, the amino acid carboxylate attacks the α -phosphate of adenosine triphosphate (ATP); synthesis of an aminoacyl-adenylate (AA-AMP) occurs with release of inorganic pyrophosphate (PP_i). The second step is transfer of the activated amino acid to either the 2'-OH or 3'-OH on the terminal adenosine of the accepting tRNA. Most tRNA synthetases catalyze AA-AMP synthesis in the absence of tRNA; exceptions are arginyl-, glutaminyl-, and glutamyl-tRNA synthetases and class I lysyl-tRNA synthetase. Some tRNA synthetases have a second editing active site that hydrolyzes mis-activated aminoacyl adenylate or mis-aminoacylated tRNA.

Accuracy

The accuracy of protein synthesis has been estimated at 1 error in 10 000 codons translated. While the codon:anticodon interaction at the ribosome is by necessity specific, the major contribution to translational accuracy arises from the specificity of tRNA aminoacylation catalyzed by the tRNA synthetases.

tRNAs

Each tRNA synthetase has a set of isoaccepting tRNA molecules that are used to translate the 61 sense codons of the genetic code. tRNAs are typically 76 nucleotides long, and have conserved secondary and tertiary structure but few conserved nucleotides. All tRNAs have the sequence cytidine-cytidine-adenosine (CCA) at their single-stranded 3'-termini and several post-transcriptionally modified bases, including ribothymidine (T) and pseudouridine (ψ) in the conserved trinucleotide T ψ C. Other nucleotides are not conserved with respect to identity, but co-vary to preserve the cloverleaf secondary structure of tRNA (Figure 1). The tertiary folded structure of tRNA is an L-shaped molecule with the amino acid accepting site and trinucleotide anticodon at the two ends of the L. During protein synthesis, the tRNA anticodon forms base pairs with its complementary mRNA codon and thus delivers its attached amino acid to the growing polypeptide chain. tRNA synthetases recognize their cognate tRNAs through identity elements typically located in the acceptor stem and anticodon loop. Some identity elements are so dominant that transplantation of one or a few nucleotides is sufficient to confer aminoacylation specificity on a noncognate or near-cognate tRNA. For example, introducing the G3:U70 wobble pair of the tRNA^{Ala} acceptor stem into other tRNAs can confer alanine acceptance, while transplanting the cytidine-adenosine-uridine (CAU) anticodon of tRNA^{Met} onto tRNA^{Val} converts the latter into a substrate for methionyl-tRNA synthetase.

Amino Acids

Selectivity of the correct amino acid is limited by potential binding interactions between each amino acid and its cognate enzyme. For some amino acids, H-bonding and hydrophobic interactions can increase binding specificity. The biggest challenge for selectivity occurs for amino acids smaller than the cognate substrate. For example, valine is one methylene group smaller than isoleucine, and fits into the active site of isoleucyl-tRNA synthetase. Mis-activation of noncognate amino acids or mis-aminoacylation of tRNA is observed for many tRNA synthetases *in vitro* and can lead to error-prone translation and cellular dysfunction if not corrected by an editing mechanism.

1. $E + AA + ATP \rightleftharpoons E-AA-AMP + PP_i$
2. $E-AA-AMP + tRNA \rightleftharpoons AA-tRNA + AMP + E$

Scheme 1 Catalysis by tRNA synthetases. In step 1, each tRNA synthetase binds its cognate amino acid (AA) and a molecule of ATP and catalyzes the condensation of aminoacyl adenylate (AA-AMP), releasing inorganic pyrophosphate (PP_i). In step 2, the activated amino acid of AA-AMP is transferred to the cognate tRNA with release of adenosine monophosphate (AMP).

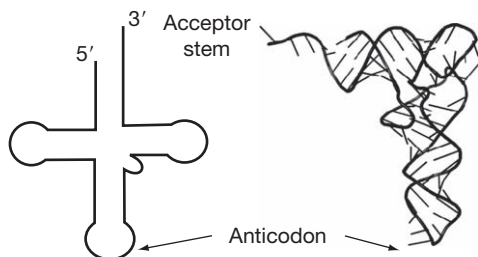


Figure 1 tRNA secondary and tertiary structure. On the left is the secondary cloverleaf structure of tRNA, on the right is the three-dimensional L-shaped tRNA that is functional in aminoacylation by tRNA synthetases and in protein biosynthesis at the ribosome (yeast tRNA^{Phe}, PDB 1TRAJ).

Editing

Several synthetases have evolved additional protein domains that actively hydrolyze aminoacyl adenylate or aminoacyl-tRNA to prevent translational errors and recycle substrates. These activities are termed pre-transfer and post-transfer editing, respectively (**Scheme 2**). In addition to the integrated protein domains of tRNA synthetases that catalyze these editing reactions, isolated polypeptides have been identified that carry out editing reactions in isolation from tRNA synthetases. Examples are YbaK and AlaX proteins that efficiently deacylate misacylated tRNA^{Pro} and tRNA^{Ala}, respectively.

Two Classes of Enzymes

With the determination of tRNA synthetase primary sequences and crystal structures, the organization of the enzymes into two distinct classes became apparent. The two classes are distinguished by mutually exclusive sequence and structure motifs in their active site domains that also influence catalytic differences. The amino acids Arg, Cys, Gln, Glu, Ile, Leu, Met, Trp, Tyr, and Val are attached to their cognate tRNAs by class I enzymes. Class II enzymes attach amino acids Ala, Asp, Asn, Gly, His, Lys, Phe, Pro, Thr, and Ser to their corresponding tRNAs.

Structures

Representative structures of all 20 tRNA synthetases have been determined by X-ray crystallography. Class I synthetases contain signature motifs that include the consensus sequences His-Ile-Gly-His (HIGH) and Lys-Met-Ser-Lys-Ser (KMSKS); their catalytic core is built around a Rossmann dinucleotide binding fold. Class II synthetases contain degenerate sequence motifs that assemble into an antiparallel β -sheet structure.

- $$AA'-AMP + E \rightarrow AA' + AMP + E \text{ Pre-transfer editing}$$
- $$AA'-tRNA + E \rightarrow AA' + tRNA + E \text{ Post-transfer editing}$$

Scheme 2 Editing by tRNA synthetases. Pre-transfer editing is the enzyme-catalyzed hydrolysis of mis-acylated amino acid (AA'), while post-transfer editing is hydrolysis of noncognate amino acid (AA') from cognate tRNA.

Many tRNA synthetases can also be crystallized in the presence of their cognate tRNA. Class I enzymes approach tRNA from the acceptor stem minor groove side, while class II synthetases bind the major groove side of the tRNA acceptor stem. Because of the way class I enzymes interact with their tRNAs, an unwinding of the acceptor stem is needed for the tRNA 3'-end to reach the activated amino acid in the catalytic site (**Figure 2**).

Mechanisms

The functional consequence of the differences in structural orientation is that class I synthetases attach the amino acid first to the 2'-OH of the tRNA terminal adenosine before the amino acid is transferred to the 3'-OH for ribosomal protein synthesis; class II synthetases directly attach the amino acid to the 3'-OH. The tRNA synthetase class organization is maintained across evolution with the exception of lysyl-tRNA synthetase, which exists as either a class I or class II enzyme depending on the organism.

Alternative Pathways and Enzymes

Transamidation Reactions

While we can describe the 20 canonical tRNA synthetases, enzymes to directly aminoacylate tRNAs with asparagine and glutamine are missing in many organisms, including most bacteria, archae, and eukaryotic organelles. Routes to generate Asn-tRNA^{Asn} and Gln-tRNA^{Gln} needed for protein biosynthesis are indirect, relying on aminoacylation of these tRNAs with aspartate and glutamate by nondiscriminating aspartyl- and glutamyl-tRNA synthetases, respectively. Instead of undergoing hydrolysis, however, the mismatched tRNAs are substrates for tRNA-dependent amidotransferases. The transamidation step requires ATP and an amide donor (asparagine, glutamine, or ammonia). In addition to providing correctly aminoacylated tRNA for protein biosynthesis, the transamidation pathway appears to be the sole route to asparagine biosynthesis in some organisms.

Expansion of the Genetic Code

The standard genetic code encompasses 20 amino acids encoded by 61 sense codons; however, natural expansions have been observed and synthetic incorporation of unnatural amino acids has been accomplished. The 21st amino acid, selenocysteine (**Figure 3**), is inserted at uridine-guanosine-adenosine (UGA) stop codons using an elegant set of adaptations to the ribosomal machinery. The amino acid serine is first attached to a unique selenocysteinyl tRNA (tRNA^{Sec}) by seryl-tRNA synthetase. Chemical conversion of serine to selenocysteine is accomplished by selenocysteine synthase. Over 20 selenoproteins have been

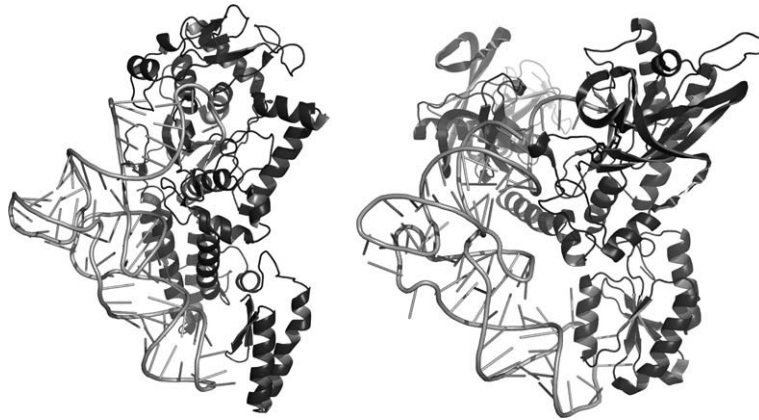


Figure 2 tRNA synthetase structures. Left, cysteinyl-tRNA synthetase, a class I tRNA synthetase (PDB 1U0B). The bound tRNA^{Cys} acceptor stem is unwound to reach into the enzyme active site. Right, threonyl-tRNA synthetase, a class II tRNA synthetase (PDB 1QF6). The N-terminal editing domain of this enzyme is located behind the tRNA acceptor stem in this view.

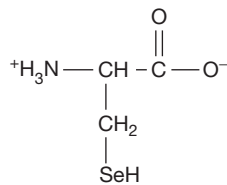


Figure 3 Selenocysteine. The 21st amino acid is attached to its tRNA^{Sec} in an indirect aminoacylation pathway utilizing seryl-tRNA synthetase.

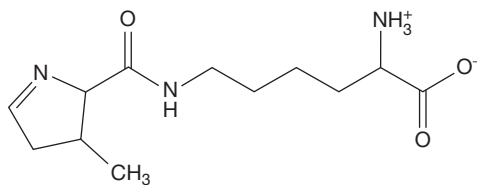


Figure 4 Pyrrolysine. Found in methanogenic bacteria, pyrrolysine is the 22nd genetically encoded amino acid and is attached to its cognate tRNA by a dedicated tRNA synthetase.

identified in humans, several of which are classified as antioxidant enzymes. Selenium is an essential trace nutrient for humans because of its incorporation into proteins as selenocysteine. A 22nd amino acid, pyrrolysine (Figure 4), has been identified in methanogenic organisms, and is cotranslationally inserted at uridine-adenosine-guanosine (UAG) stop codons using a dedicated pyrrolysyl-tRNA synthetase and tRNA^{Pyl}. As the ribosome does not appear to have a proofreading activity to identify mis-aminoacylated tRNAs, peptides or proteins containing unnatural amino acids can be generated by modifying the active site of tRNA synthetases to allow these amino acids. Such unnatural amino acids can add new functionality to proteins, allowing the introduction of novel fluorophores, reactive groups, or heavy atoms.

Novel Functions

In recent years, it has become increasingly apparent that tRNA synthetases have several functions in addition to their

housekeeping role in protein biosynthesis. These additional activities tend to be specific for a given synthetase and even a particular organism. Mitochondrial leucyl-tRNA synthetase from yeast and tyrosyl-tRNA synthetase from *Neurospora crassa* facilitate group I intron splicing. *Escherichia coli* threonyl-tRNA synthetase regulates its own expression by binding its mRNA leader sequence. Human tyrosyl-tRNA synthetase is cleaved by proteolysis into two fragments that act as cytokines in regulating blood vessel formation, while a fragment of human tryptophanyl-tRNA synthetase inhibits blood vessel formation. Some noncanonical activities depend on interactions with non-synthetase proteins, such as for human lysyl-tRNA synthetase, which interacts with the human immunodeficiency virus 1 (HIV-1) gag protein to assemble active virus particles. The human bifunctional glutamyl-prolyl-tRNA synthetase binds accessory proteins to suppress the translation of mRNAs involved in the immune response.

Synthetases in Disease and Medicine

Disease Correlations

The most common confirmed involvement of tRNA synthetases in disease is a disorder termed 'anti-synthetase syndrome', in which individuals develop antibodies against one of six members of the enzyme family; symptoms include muscle inflammation, interstitial lung disease, and arthritis. It has been determined that up to one-third of patients exhibiting chronic inflammatory muscle or skin and muscle disorders have antibodies against their own tRNA synthetases. The means by which anti-synthetase syndrome develops remains unknown. Other links between tRNA synthetases and disease are increasingly being identified. Examples include defects in glycyl-, tyrosyl-, and alanyl-tRNA synthetases in peripheral neuropathies such as Charcot-Marie-Tooth syndrome and ataxia. Several tRNA synthetases are also dysregulated in cancers; for example, the activity of methionyl-tRNA synthetase is increased in human colon cancer. A specific mutation in human mitochondrial leucyl-tRNA synthetase seems linked with type 2 diabetes mellitus, although the mechanism is not yet known.

Inhibitors as Drugs

The essential cellular function of tRNA synthetases makes this family an attractive target for discovery and development of anti-infective agents; numerous academic, biotechnology, and pharmaceutical laboratories are currently investigating tRNA synthetase inhibitors. The occurrence of natural products inhibiting tRNA synthetases lends credibility to the search for novel anti-infectives. In wide use is the anti-bacterial compound mupirocin, secreted by *Pseudomonas fluorescens*, which inhibits isoleucyl-tRNA synthetase by mimicking the isoleucine adenylylate. Borrelidin, named for its discovered action against *Borrelia* species, is secreted by *Streptomyces rochei* and inhibits eukaryotic threonyl-tRNA synthetase. More recently, a benzoxaborole compound has been shown to trap tRNA^{Leu} in the leucyl-tRNA synthetase editing site, inactivating this enzyme in a novel mechanistic manner. The primary challenge in developing anti-synthetase compounds for use other than as topical agents is identifying species-specific structural or mechanistic differences in order to target the pathogen enzyme with minimal effect on host function.

See also: **Metabolism Vitamins and Hormones:** Amino Acid Metabolism; **Molecular Biology:** Eukaryotic Protein Biosynthesis:

The Elongation Cycle; Messenger RNA Processing in Eukaryotes; Pre-tRNA and Pre-rRNA Processing in Bacteria; Ribosome Structure; RNA Editing.

Further Reading

- Ibba M and Söll D (2000) Aminoacyl-tRNA synthesis. *Annual Review of Biochemistry* 69: 617–650.
- Ibba M, Francklyn C, and Cusack S (2005) *The Aminoacyl-tRNA Synthetases*. Georgetown, TX: Landes Biosciences.
- Krzycki JA (2005) The direct genetic encoding of pyrrolysine. *Current Opinion in Microbiology* 8: 706–712.
- Ling J, Reynolds N, and Ibba M (2009) Aminoacyl-tRNA synthesis and translational quality control. *Annual Review of Microbiology* 63: 61–78.
- Liu CC and Schultz PG (2010) Adding new chemistries to the genetic code. *Annual Review of Biochemistry* 79: 413–444.
- Nelson DL and Cox MM (2009) *Lehninger's Principles of Biochemistry*, 5th edn. New York: W. H. Freeman.
- Park SG, Schimmel P, and Kim S (2008) Aminoacyl tRNA synthetases and their connections to disease. *Proceedings of the National Academy of Sciences of the United States of America* 105: 11043–11049.

Troponin

I Ohtsuki, The Jikei University School of Medicine, Tokyo, Japan

S Morimoto, Kyushu University Graduate School of Medicine, Fukuoka, Japan

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Takeyuki Wakabayashi, Setsuro Ebashi, volume 1, pp. 250–255, © 2004, Elsevier Inc.

Glossary

Tropomyosin A fibrous coiled-coil protein of about 40 nm in length (MW ~70 kDa). This protein forms head-to-tail filaments and binds to both troponin and actin.

Troponin isoforms Troponin isoforms from fast and slow skeletal and cardiac muscles are encoded by different genes, except that troponin C from slow skeletal and cardiac

muscles are encoded by the same gene. The cardiac isoform of troponin is widely used clinically as a biochemical marker of myocardial injury. A specific immunoassay for cardiac isoforms of troponins T and I released into blood from injured myocardium is the standard method for detecting myocardial injury, including acute myocardial infarction.

Early Studies

In 1963, a protein factor, which was extracted from minced rabbit skeletal muscle at an extremely low ionic strength condition, was found to confer Ca^{2+} sensitivity to the contractile interaction between myosin and actin. The protein factor was called 'native tropomyosin', because several properties of the factor were similar to those of tropomyosin, which had been isolated from muscular tissue about two decades earlier. However, the factor was soon revealed to be a complex of tropomyosin and a novel protein, named 'troponin'. Troponin was shown to work as a Ca^{2+} receptor for the Ca^{2+} regulation of the contractile interaction between myosin and actin in both skeletal and cardiac muscles. In the absence of Ca^{2+} , troponin, in association with tropomyosin, suppresses the contractile interaction, and, when the Ca^{2+} concentration increases to a micromolar level, this suppression is reversed through the binding of Ca^{2+} to troponin. An immunoelectron microscopic study demonstrated that troponin is distributed at regular intervals of 38 nm along the entire length of thin filaments. This periodic distribution of troponin together with other related findings led to the proposal of the thin filament model, in which two head-to-tail filaments of fibrous tropomyosin molecules run along the grooves of actin double strands and troponin attaches to the specific region of each tropomyosin. The model visualized the essential regulatory pathways over which the action of Ca^{2+} on troponin is mediated to actin molecules through tropomyosin. These studies revealed the molecular basis for the Ca^{2+} -regulatory mechanisms of muscle contraction.

Troponin Subunits

Troponin is a complex of three different subunits: troponins I, C, and T. Troponin I, together with tropomyosin, inhibits the contractile interaction between actin and myosin. Troponin C binds Ca^{2+} , and troponin T binds to tropomyosin. The Ca^{2+} regulation of the contractile interaction takes place in the presence of all three subunits of troponin and tropomyosin.

Troponin I

Troponin I inhibits the contractile interaction between myosin and actin in the presence of tropomyosin. Through this inhibitory activity of troponin I, troponin–tropomyosin suppresses the contractile interaction in the absence of Ca^{2+} . The inhibition by troponin I is extremely weak in the absence of tropomyosin. The inhibitory activity of troponin I was first shown to be largely preserved in the cyanogen bromide fragment CN4 (inhibitory peptide region; residues 96–116) of rabbit skeletal troponin I with 181 residues. Studies with synthetic peptides later demonstrated that the region of residues 105–114 (minimum inhibitory peptide), called the 'inhibitory region' in this article, is essential for the inhibitory activity, and that the region of residues 140–148 (second actin-binding region) is also necessary for full inhibition. The N-terminal region of troponin I adjacent to the inhibitory region binds to troponin C and forms a coiled-coil structure with troponin T, whereas the C-terminal region of troponin I interacts with actin–tropomyosin and binds to troponin C in the presence of Ca^{2+} . The Ca^{2+} -dependent troponin C-binding region of troponin I is located in the helical region, called the 'switch region', adjacent to the C-terminus of the inhibitory region.

Troponin I from cardiac muscle has an additional N-terminal extension of about 30 residues. Two adjacent Ser residues in the N-terminal extension are phosphorylated by protein kinase A (PKA). The N-terminal extension interacts with the N-domain of troponin C but, when phosphorylated by PKA, dissociates from troponin C, decreasing the Ca^{2+} affinity of troponin C as well as the Ca^{2+} sensitivity of contraction. The PKA-dependent phosphorylation of the N-terminal extension is involved in the acceleration of the diastolic phase induced by β -adrenergic agonists during rhythmic contraction of cardiac muscle.

Troponin C

Troponin C is a Ca^{2+} -binding subunit of troponin, which forms a complex with troponin I and troponin T. This subunit neutralizes the suppression of the contractile interaction

between myosin and actin induced by troponin I-tropomyosin, regardless of the Ca^{2+} -concentration (neutralizing action).

Troponin C has four Ca^{2+} -binding sites, termed sites I, II, III, and VI from the N-terminus, each of which consists of a Ca^{2+} -coordinating loop rich in acidic residues flanked with two α -helical segments (E-F-hand motif). X-ray crystallographic analyses have shown that troponin C forms a dumbbell-shaped structure, in which two globular domains, the N-domain and the C-domain, are linked by a central helix. Sites I and II in the N-domain specifically bind Ca^{2+} with low affinity and function as the Ca^{2+} receptor for the Ca^{2+} regulation of contraction (Ca^{2+} -specific sites), whereas sites III and IV in the C-domain bind both Ca^{2+} and Mg^{2+} with high affinity (Ca-Mg sites) and are always occupied by Ca^{2+} or Mg^{2+} under usual intracellular conditions.

In the absence of Ca^{2+} , the region patched with hydrophobic residues in the N-domain is closed within the molecule (closed conformation). The binding of Ca^{2+} to the sites in the N-domain induces the hydrophobic patch to open on the surface (open conformation) and to form a complex with the helical switch region of troponin I. In cardiac troponin C, site I is lost by residue replacement in the coordinating loop. The hydrophobic patch in the N-domain of cardiac troponin C keeps the closed conformation even in the presence of Ca^{2+} and a complex formation with the switch region is required to reach the open conformation. The C-domain of troponin C, together with troponins T and I, forms a structural core in the troponin complex.

Troponin T

Troponin T is a tropomyosin-binding subunit of troponin. This peptide is the largest of the three troponin subunits and interacts with both troponin I and troponin C. Troponin T is not directly involved in the Ca^{2+} -regulatory interactions in the troponin complex, but the presence of troponin T, in addition to troponins C and I, is required for the Ca^{2+} -regulated contractile interaction to take place. The regulatory role of troponin T is to confer the Ca^{2+} sensitivity to the neutralizing effect of troponin C.

Troponin T is split into two subfragments, called 'troponin T_1 (or TnT1)' and 'troponin T_2 (or TnT2)', by treatment with chymotrypsin. In rabbit skeletal troponin T, troponin T_1 is a N-terminal subfragment of 158 residues and binds to tropomyosin through the helical region of about 80 residues, whereas troponin T_2 , a C-terminal subfragment with 101 residues, binds to tropomyosin through its C-terminal end region and interacts with troponins I and C. The Ca^{2+} -regulatory property of troponin T is retained solely in the troponin T_2 region. Within the structural core of the ternary troponin complex, the region of 47 residues of troponin T_2 forms a coiled coil with the N-terminal region of troponin I and stably binds to the C-domain of troponin C. The troponin T_1 and T_2 regions in each molecule are axially arranged on the Z-line and the filament-top sides, respectively, along the thin filament (Figure 1).

Cardiac troponin T is about 30 residues longer than skeletal troponin T at the N-terminus. In the case of human cardiac troponin T, four isoforms are produced by the alternative

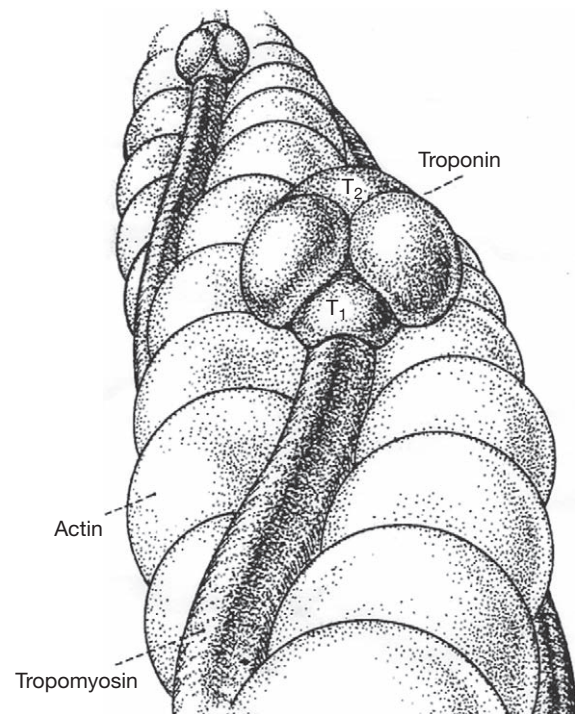


Figure 1 Molecular arrangement of troponin T in the thin filament. The T_1 and T_2 regions of troponin T are located at the near side (Z-line side) and far side (filament-top side), respectively. Reproduced from Ohtsuki I (1980) Functional organization of the troponin-tropomyosin system. In: Ebashi S, et al. (eds.) *Muscle Contraction: Its Regulatory Mechanisms*, pp. 237–250. Tokyo: Japan Scientific Society Press; Berlin Heidelberg, New York: Springer-Verlag, with permission from Japan Scientific Society Press and Springer Science+Business Media.

splicing of exons in this N-terminal extension. The Ca^{2+} -sensitivity of contraction is affected differently by these isoforms of troponin T.

Regulatory Function of Troponin

The essential process of Ca^{2+} regulation in the ternary complex of troponin consists of two complementary interactions: an inhibitory interaction of troponin I with actin-tropomyosin and a Ca^{2+} -induced interaction between troponin I and troponin C. In the absence of Ca^{2+} , the contractile interaction is suppressed by the inhibitory action of troponin I and, in the presence of Ca^{2+} , the suppression by troponin I is reversed through the Ca^{2+} -induced interaction of troponin C with troponin I.

Troponin I alone, in association with tropomyosin, suppresses the interaction between myosin and actin. The suppression by troponin I-tropomyosin is neutralized by troponin C regardless of the Ca^{2+} concentration, although a complex of troponin I-C binds to actin-tropomyosin in the absence of Ca^{2+} but not in the presence of Ca^{2+} . The presence of troponin T, in addition to troponins C and I, is required for the suppression of the contractile interaction through the inhibitory action of troponin I within the ternary troponin complex, in the absence of Ca^{2+} . According to the X-ray crystallographic

analysis of the core domain of troponin, the N-terminal region of troponin I and the C-terminal region of troponin T (troponin T₂) form a stable coiled-coil structure, whose C-terminal ends are adjacent to both the inhibitory region of troponin I and the C-terminus of troponin T₂ that binds to tropomyosin. This close apposition of the inhibitory region of troponin I and the C-terminus of troponin T₂ suggests that the inhibitory activity of troponin I in the absence of Ca²⁺ is protected by both the stable coiled-coil formation with troponin T₂ and the binding of troponin T₂ to tropomyosin, against the neutralizing action of troponin C. Troponin T₁, by binding of its helical region to tropomyosin, would contribute to the overall arrangement of the core domain along the thin filament. There is also evidence indicating that tropomyosin enhances the affinity for troponin I of one of the seven actins in the 38-nm period, which is located at the specific region of each tropomyosin. The inhibitory effect of troponin I on the actin, which probably has a high affinity for troponin I, is mediated to the other actin molecules via tropomyosin.

In the presence of Ca²⁺, the hydrophobic patch region in the regulatory N-domain of troponin C becomes open on the surface of the molecule. The Ca²⁺-bound N-domain with the open conformation then binds to the switch region of troponin I and induces the inhibitory region of troponin I, which is adjacent to the N-terminus of switch region, to dissociate from actin and consequently to trigger the contractile interaction of actin with myosin.

To explore the physiological properties of troponin, it is critically important to remove and replace each subunit of troponin in myofibrils or skinned fibers and to compare the mode of Ca²⁺ regulation with that of the original one. Examinations along this line have been partly realized by treating skinned muscle fibers or myofibrils with excess amounts of troponin T. The added troponin T expels the intrinsic ternary troponin complex by binding to tropomyosin. Through overall depletion of troponins C and I after treatment with troponin T, the force of skinned fibers is activated at a level slightly lower than the maximum force of intact fibers regardless of the Ca²⁺ concentration, and is suppressed by the reconstitution with troponin I. The original Ca²⁺-sensitive force generation is fully restored by the subsequent reconstitution with troponin C. These findings indicate that the force generation of skinned fibers is suppressed by the inhibitory activity of troponin I and de-suppressed through the binding of Ca²⁺ to troponin C. These findings are also consistent with the findings that troponin elevates the maximum level of Ca²⁺-activated tropomyosin-actomyosin adenosine triphosphatase (ATPase) activity. This activating effect of troponin is a minor mechanism in vertebrate striated muscle, but it is a principal mechanism of troponin from scallop muscle. The particular property of scallop troponin is to bind Ca²⁺ solely by the site IV in the C-domain of troponin C, which is not the regulatory site but the structural site in vertebrate troponin C. The activating action of human cardiac troponin is also known to be enhanced by two missense mutations of troponin T causing hypertrophic cardiomyopathy (HCM), which are located in the structural core of the troponin complex.

The above considerations also indicate that all three subunits of troponin within skinned or permeabilized muscle fibers can be selectively replaced by the added troponin

subunits, through treatment with excess troponin T or troponin CIT complex. The exchangeability of the binding of troponin T to tropomyosin in the thin filament has been used to incorporate various isoforms or recombinant mutants of troponin into skinned fibers. Through this procedure, troponin was shown to be involved in the length-dependent activation mechanisms of force development in cardiac muscle. Extensive functional analyses have been made of the large number of human cardiac troponin mutations associated with inherited cardiomyopathies.

Troponin and Genetic Disorders

Genetic analysis has shown that inherited cardiomyopathies are associated with the mutations in genes for sarcomeric and cytoskeletal proteins, including cardiac troponin. Nearly one hundred mutations in the genes for cardiac troponin isoforms have been reported to cause familial cardiomyopathies. HCM is characterized by an asymmetric left ventricular hypertrophy and impaired diastolic function. Dilated cardiomyopathy (DCM) is characterized by ventricular dilation and systolic dysfunction. Restrictive cardiomyopathy (RCM) is the least common of the inherited cardiomyopathies and is characterized by impaired diastolic filling of the left ventricle, normal or near-normal systolic function, and normal ventricular wall thickness. These disorders are associated with a high incidence of heart failure and sudden death.

Troponin T Mutations

The cardiac troponin T mutations causing HCM were the first troponin mutations identified in inherited cardiomyopathies. Functional analyses of the mutations were performed with the troponin-exchange technique in skinned cardiac muscle. In 1998, missense mutations of troponin T (Ile79Asn, Arg92Gln) were found to increase the Ca²⁺ sensitivity of isometric force but not to affect the maximum level and cooperativity. Fifteen troponin T mutations causing HCM were subsequently reported to have the Ca²⁺-sensitizing effects on the skinned fiber force generation and myofibrillar or actomyosin ATPase activity. By contrast, a deletion mutation of troponin T (Δ Lys210)-causing DCM was found to decrease the Ca²⁺ sensitivity of contraction without causing any significant changes in the maximum level or cooperativity, and other mutations of troponin T-causing DCM have also been reported to decrease the Ca²⁺ sensitivity of force generation and ATPase activity of myofibrils or actomyosin (Figure 2).

These functional studies of the troponin T mutations emphasize the importance of the changes in the Ca²⁺ sensitivity as the critical functional consequences in determining the phenotypes of HCM and DCM associated with troponin mutations. In cardiac muscle, the Ca²⁺-activated force achieved during twitch contraction is usually less than half the maximum, and consequently alterations of the Ca²⁺ sensitivity greatly influence the force generation in both the systolic and diastolic phases. The increased Ca²⁺ sensitivity enhances systolic contraction and delays diastolic relaxation, whereas the decreased Ca²⁺ sensitivity reduces systolic contraction. These changes are consistent with the distinct phenotypes in HCM or

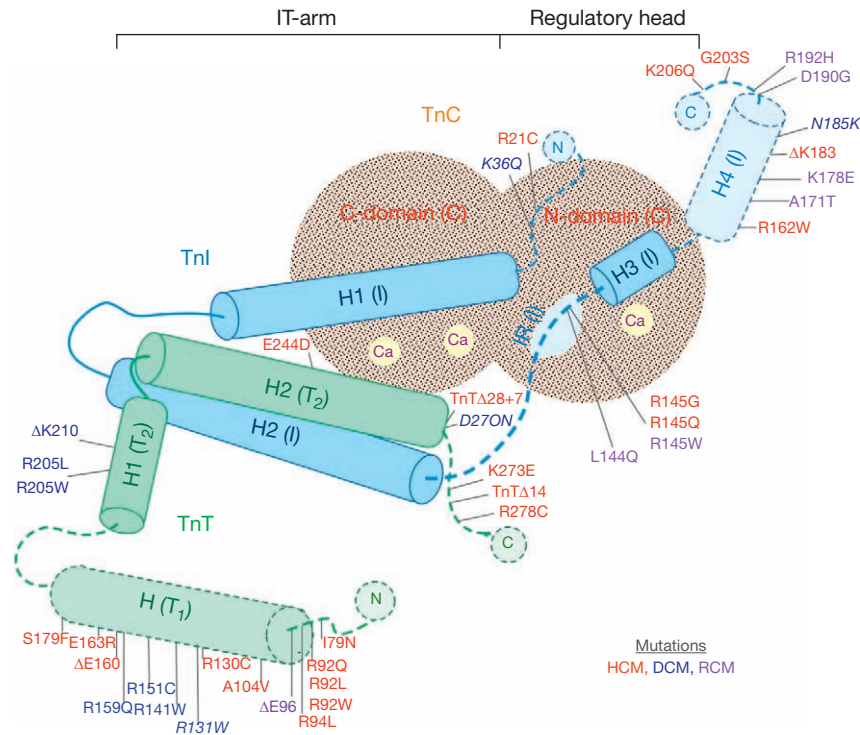


Figure 2 Distribution of mutations of troponins T and I causing HCM, DCM, and RCM in human cardiac troponin. The diagrammatic illustration of Ca^{2+} -bound cardiac troponin in the figure is based on the X-ray crystallographic structure of the core domain of the human cardiac troponin CIT₂-complex. Troponin subunits are indicated by different colors: green (troponin I, TnI), sky blue (troponin T, TnT), and pink (troponin C, TnC). The core domain is composed of two subdomains: an IT arm and a regulatory head (indicated by solid lines). The IT arm is the structural core of troponin which is composed of the C-domain of troponin C, troponin I (H1(I), H2(I)), and troponin T₂ (H1(T₂), H2(T₂)), whereas the regulatory head is a complex of the N-domain of troponin C and the H3(I) switch helix of troponin I. The inhibitory region (IR), which is not visualized in the crystal structure, is indicated by a thick broken line between the two helical regions, H2(I) and H3(I). Other regions of troponin subunits outside the core domain are indicated by thin broken lines. The positions of mutations of troponins T and I, of which the Ca^{2+} -sensitizing and -desensitizing actions have been demonstrated on the skinned fiber force and myofibrillar or actomyosin ATPase activity, are indicated by different colors; red (HCM), dark blue (DCM), and purple (RCM). The mutations, whose effects were examined solely on actomyosin or acto-S1 ATPase, are indicated by *italic letters*. Adapted from Ohtsuki I and Morimoto S (2008) Troponin: Regulatory function and disorders. *Biochemical and Biophysical Research Communications* 369: 62–73, with permission from Elsevier.

DCM: HCM is characterized by impaired diastolic function accompanied by ventricular hypertrophy, whereas DCM is characterized by reduced systolic function and ventricular wall dilation. The essential importance of the decreased Ca^{2+} sensitivity for the pathogenesis of DCM was demonstrated in the knock-in mouse model of DCM caused by the Lys210-deletion mutation of troponin T.

Troponin I Mutations

Mutations of cardiac troponin I causing HCM, DCM, and RCM have been reported. The Ca^{2+} -sensitizing and Ca^{2+} -desensitizing effects, which had been shown by mutations of troponin T causing HCM and DCM, respectively, are also demonstrated by the mutations of troponin I. Analysis of seven examined mutations causing HCM has shown the Ca^{2+} -sensitizing effects on the skinned fiber contraction and myofibrillar or actomyosin ATPase activity. Two missense mutations causing DCM have been reported to decrease the Ca^{2+} sensitivity of actomyosin ATPase activity (Figure 2).

RCM is an uncommon disorder, and six mutations of troponin I associated with RCM (Leu144Gln, Arg145Trp, Ala171Thr, Lys178Glu, Asp190Gly, and Arg192His) were reported.

Although RCM is characterized by restrictive filling and little or no ventricular hypertrophy, these mutations increase the Ca^{2+} sensitivity of skinned fiber contraction and myofibrillar or actomyosin ATPase activity, as do mutations causing HCM. The extent of Ca^{2+} sensitization due to mutations causing RCM is, however, greater than that due to mutations causing HCM. The Ca^{2+} -sensitizing effects due to most mutations causing RCM are accompanied by both elevated minimum force levels and the decreased maximum force generation, whereas the Ca^{2+} -sensitizing effects of mutations causing HCM are accompanied only by elevated resting forces. Most mutations causing RCM and HCM are located in the inhibitory region and its C-terminal side sequence (Figure 2). These findings suggest that common mechanisms exist for the pathogenesis of RCM and HCM. Consistent with this view, a mutation (Asp190Gly) was found to be associated with the mixed appearance of RCM and HCM in a large family. Two mutations

(Lys178Glu and Arg192His) were *de novo*, and two mutations (Leu144Gln and Ala171Thr) were identified only in the probands. Three mutations (Leu144Gln, Arg145Trp, and Asp190Gly) were later identified in patients who had HCM with a restrictive phenotype.

Troponin C Mutations

The Ca^{2+} -sensitizing effects on skinned fiber force generation and actomyosin ATPase activity have been documented by four mutations of troponin C causing HCM, which are distributed in both the N- and C-domains. A mutation causing DCM, located at the C-terminus, was reported to have no significant Ca^{2+} -desensitizing effect on the skinned fiber force generation but was shown to decrease the Ca^{2+} sensitivity of acto-S1 ATPase activity.

Knock-In Mouse Model of DCM Caused by Troponin T Mutation

Knock-in mice, in which three base pairs coding for the amino acid residue Lys210 in cardiac troponin T were deleted from the endogenous cardiac troponin T gene, have demonstrated the critical role of myofilament Ca^{2+} sensitivity in the pathogenesis of DCM. Skinned cardiac muscle fibers from these mutant mice showed a significant decrease in the Ca^{2+} sensitivity of force generation. However, intact cardiac muscle fibers from mutant mice did not show any significant decrease in force-generating capability, which was expected to be caused by the decreased Ca^{2+} sensitivity of myofilaments. Analysis with fura-2, a Ca^{2+} -indicator, revealed a significant increase in the peak amplitude of the Ca^{2+} transient in cardiomyocytes of mutant mice and significant increases in the phosphorylation of ryanodine receptor-2 and of phospholamban catalyzed by a cyclic adenosine monophosphate (cAMP)-dependent protein kinase. These findings suggest that the Ca^{2+} transient is augmented through increased cAMP levels to compensate for the decreased myofilament Ca^{2+} sensitivity and to maintain the force-generating capability of cardiomyocytes. Nevertheless, the mutant mice had markedly enlarged hearts and extremely high rates of premature sudden death due to fatal arrhythmia, closely mimicking the phenotypes of human DCM. These outcomes indicate that the decreased myofilament Ca^{2+} sensitivity leads to global remodeling of the heart to compensate for an inevitable decrease in cardiac function with sacrifice of the electrophysiological stability. As expected, oral administration of a Ca^{2+} sensitizer, known to increase the myofilament Ca^{2+} sensitivity in cardiac muscle, was found to markedly reduce the heart size of mutant mice and almost completely prevent the sudden death.

Troponin Mutations and Regulatory Function

Within Ca^{2+} -bound human cardiac troponin, the core domain consists of two subdomains, an IT-arm and a regulatory head (Figure 2). In the regulatory head, the Ca^{2+} -bound N-domain of troponin C forms a complex with the short helical switch region, H3(I), of troponin I. The inhibitory region of troponin I, which is adjacent to the switch helix region, is located near the core domain and is separated from actin, in the presence of

Ca^{2+} . In the IT-arm region, the C-domain of troponin C binds to both troponin I helix, H1(I), and the troponin T₂ helix, H2(T₂), and, the coiled-coil backbone is formed by troponin T₂ helix, H2(T₂), and troponin I helix, H2(I). The IT-arm region is connected to tropomyosin through the two regions of troponin T: the helical region of troponin T₁ and the C-terminal end region of troponin T₂.

Most mutations of troponin T and troponin I are outside the core domain. Their locations strongly indicate that these mutations affect the Ca^{2+} sensitivity of contraction mainly through interactions with tropomyosin and actin, rather than through conformational or interaction changes within the core domain. This view is consistent with the fact that only a small number of mutations causing HCM or DCM have been reported in the genes for the cardiac troponin C molecule, which is located within the core domain. Troponin T mutations are distributed in both the troponin T₁ and troponin T₂ regions, where mutations causing HCM would induce Ca^{2+} -sensitizing effects by impairing the interactions with tropomyosin and where mutations causing DCM would decrease Ca^{2+} sensitivity by modulating the interactions with tropomyosin and actin. Most troponin I mutations are located in the inhibitory region or the C-terminal regulatory region. Mutations of troponin I causing either HCM or RCM would impair the inhibitory interaction with actin–tropomyosin or modulate the neutralizing interaction with troponin C or both, thereby increasing the Ca^{2+} sensitivity of contraction, which is accompanied by the changes in minimum or maximum force levels or both.

See also: **Bioenergetics: Calcium-Modulated Proteins (EF-Hand); Protein/Enzyme Structure Function and Degradation: Protein N-Myristoylation.**

Further Reading

- Du CK, Morimoto S, Nishii K, et al. (2007) Knock-in mouse model of dilated cardiomyopathy caused by troponin mutation. *Circulation Research* 101: 185–194.
- Ebashi S and Ohtsuki I (eds.) (2007) *Regulatory Mechanisms of Striated Muscle Contraction. Advances in Experimental Medicine and Biology*, vol. 592. Tokyo: Springer.
- Kobayashi T and Solaro RJ (2005) Calcium, thin filaments, and the integrative biology of cardiac contractility. *Annual Review of Physiology* 67: 39–67.
- Morimoto S (2008) Sarcomeric proteins and inherited cardiomyopathies. *Cardiovascular Research* 77: 659–666.
- Morimoto S, Lu Q-W, Harada K, et al. (2002) Ca^{2+} -desensitizing effect of a deletion mutation delta-K210 in cardiac troponin T that causes familial dilated cardiomyopathy. *Proceedings of the National Academy of Sciences of the United States of America* 99: 913–918.
- Ohtsuki I, Maruyama K, and Ebashi S (1986) Regulatory and cytoskeletal proteins of vertebrate skeletal muscle. *Advances in Protein Chemistry* 38: 1–68.
- Ohtsuki I and Morimoto S (2008) Troponin: Regulatory function and disorders. *Biochemical and Biophysical Research Communications* 369: 62–73.
- Perry SV (1999) Troponin I: Inhibitor or facilitator. *Molecular and Cellular Biochemistry* 190: 9–32.
- Takeda S, Yamashita A, Maeda K, and Maeda Y (2003) Structure of the core domain of human cardiac troponin in the Ca^{2+} -saturated form. *Nature* 424: 35–41.
- Vinogradova MV, Stone DB, Malanina GG, et al. (2005) Ca^{2+} -regulated structural changes in troponin. *Proceedings of the National Academy of Sciences of the United States of America* 102: 5038–5043.

Tubulin and Its Isoforms

E Nogales, University of California, Berkeley, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Dynamic instability Nonequilibrium behavior of microtubules by which they can stochastically switch between phases of growth and shrinkage. It originates from the hydrolysis of guanosine triphosphate (GTP) in β -tubulin and can be regulated by the interaction of tubulin/microtubules with cellular factors and antimetabolic agents.

Microtubules Cytoskeletal polymers made of $\alpha\beta$ -tubulin essential for cell transport and cell division. They are polar,

dynamic, and are regulated through the cell cycle by their interaction with stabilizers and depolymerizers.

$\alpha\beta$ -Tubulin dimer Essential, highly conserved protein dimer present in all eukaryotes that self-assembles forming microtubules. It is the target of antimetabolic drugs with anticancer potential.

γ -Tubulin Tubulin isoform most abundant at microtubule-organizing centers where it is involved in microtubule nucleation. It forms higher-order complexes with associated proteins.

Tubulin is an $\alpha\beta$ dimeric protein that self-assembles into microtubules and is present in all eukaryotes. Tubulin is highly conserved across species, reflecting the sequence constraints imposed by microtubule structure and function. Both α - and β -subunits exist in numerous isotypic forms and undergo a variety of posttranslational modifications. Tubulin assembly and disassembly, which are linked to guanosine triphosphate (GTP) hydrolysis, make the microtubule network dynamic in both time and space to accommodate the needs of the cell during the cell cycle. Purified tubulin retains its self-assembling capabilities, allowing the biochemical and biophysical characterization of the microtubule polymerization and depolymerization processes. A variety of proteins interact with tubulin in the cell, affecting the stability of microtubules and their function. Numerous ligands bind to tubulin and influence its assembly properties, which include several drugs that have proved to have anticancer properties.

γ -Tubulin is a rarer tubulin isoform involved in the nucleation of microtubules at microtubule-organizing centers (MTOCs). Recently, additional tubulin isoforms of yet ill-defined function have been identified.

$\alpha\beta$ -Tubulin

Physical Properties

General

Tubulin α - and β -subunits have molecular weights of ~ 50 kDa and are 36–42% identical and 63% homologous. Both tubulin subunits bind guanine nucleotides. The binding to α -tubulin at the N-site is nonexchangeable, while the binding to β -tubulin at the E-site is exchangeable. Magnesium increases the affinity of the β -subunit for GTP with respect to guanosine diphosphate (GDP). Nucleotide in oligomeric tubulin or in microtubules does not exchange with the solution, except for terminal subunits at microtubule ends.

Neuronal cells are particularly rich sources of tubulin because microtubules are required for axonal transport. Neural tissue contains sufficient tubulin to allow tubulin purification by repeated cycles of 37°-induced assembly and 0°-induced

disassembly, with intervening centrifugation to alternately pellet microtubules or impurities. The yield of tubulin from 1 kg of brain and yeast is ~ 150 and ~ 5 mg, respectively. The assembly of purified tubulin can be assayed by light scattering, X-ray scattering, centrifugation, and electron microscopy.

Isotypes and posttranslational modifications

Tubulin exists in different isotypic forms, the biological significance of which is still a matter of debate. The number of tubulin isotypes increases with the organism complexity. Although yeast has only two α - and one β -tubulin isotypes, higher eukaryotes have up to eight β - and six α -tubulin isotypes. Certain isotypes have been found to be tissue specific, and differential expression of β -tubulin isotypes has been observed during the cell cycle. Some of these isotypes have been shown *in vitro* to have different relative stabilities, and such differences seem important for the response of the cell to antitubulin, anticancer drugs. The majority of differences between isotypes localizes within the last 15 residues of the sequences, a region that has been identified as important in the interaction of microtubules with microtubule-associated proteins (MAPs), pointing to a possible relevance for the functionality of microtubules in the cell.

Both tubulin subunits can be extensively altered by posttranslational modification, including dephosphorylation/tyrosination, acetylation/deacetylation, phosphorylation, polyglutamylation, and polyglycylation. All of these modifications, except for the acetylation of α -tubulin at Lys 40, occur at the divergent, highly charged C-terminal end of α - and β -tubulin. As for the different tubulin isotypes, the functionality of the posttranslational modifications is still a matter of debate. Certain modifications were initially identified as causing microtubule stability, and had been later postulated as the effect, not the cause of microtubule stability.

Structure

The structure of tubulin was initially solved by electron crystallography of zinc-induced two-dimensional tubulin sheets stabilized with the anticancer drug taxol. α - and β -tubulin have basically the same secondary structure, each being made of a

core of two β -sheets surrounded by helices as shown in **Figure 1**. The N-terminal domain forms a Rossmann fold with the nucleotide-binding site at the C-terminal end of six parallel strands that are surrounded by five helices. An intermediate domain containing the binding site of taxol and colchicine is formed by a mixed β -sheet of four strands and five surrounding helices. The C-terminal domain contains two long helices that overlap the previous two domains and constitute the outside crest of the protofilaments in the microtubule to which motor molecules bind. The nucleotide sits at the interface between subunits along the protofilament. The N-site in α -tubulin is buried within the dimer, while the E-site in β -tubulin is partially exposed in the dimer but occluded in microtubules.

The X-ray crystallographic structure of tubulin in a nonassembled state has been obtained while bound to a fragment of the depolymerizer protein RB3 and bound to a variety of depolymerizing drugs (e.g., colchicine and vinca alkaloids). Comparison of the straight tubulin dimer in the assembled sheets used for electron crystallography and these curved, depolymerized tubulin shows a conformational change in each monomer subunit that correspond to a hinge motion between the N-terminal and intermediate domains, as well as a displacement of the H7 helix. Lower-resolution structures have been proposed for structural mimics of the intermediates in the process of assembly (sheets) and disassembly (peels). How much differences in curvature in tubulin reflect nucleotide state (GTP vs. GDP), assembly state (assembled vs. disassemble), or a mixture, are still a matter of debate.



Figure 1 Ribbon diagram for the structure of $\alpha\beta$ -tubulin corresponding to a view from the inside of the microtubule with the plus end at the top. β -Tubulin (top) is bound to GDP, while α -tubulin (bottom) is bound to GTP. The color scheme highlights the three domains in the structure of each monomer: green for N-terminal, nucleotide-binding domain (helix H7 or core helix is shown in lime), second or intermediate domain in blue, and C-terminal domain in purple.

Tubulin homologs have now been found in eubacteria and archaeobacteria, including the ubiquitous FtsZ, essential for cytokinesis, and plasmid-encoded proteins involved in plasmid segregation (e.g., TubZ).

Synthesis and folding

Tubulin synthesis in cells is regulated by a process in which an increased subunit concentration leads to specific degradation of β -tubulin messenger RNA. In animal cells, there is a mechanism to assure equivalent synthesis of α - and β -subunits. For its correct folding tubulin requires the cytosolic chaperonin CCT (chaperonin containing TCP1; also referred to as TriC, TCP1, or Ct-cpn60), a hetero-oligomer formed by eight different subunits assembled into a hexadecamer of two double rings. Folding by CCT requires cycles of binding and full release, each cycle consuming one adenosine triphosphate by the chaperonin. In addition, tubulin requires additional chaperonin cofactors that bind sequentially to α - and β -monomers and are necessary for the formation of the tubulin heterodimer. Folding cofactors are also important in regulating α/β -tubulin ratios.

Tubulin Polymerization

Microtubule assembly and structure

The tubulin sequence and structure contain the information required for its self-assembly into polar, dynamic microtubules, which, in turn, interact with a variety of cellular factors. Tubulin dimers bind head to tail, making linear protofilaments, which associate in a parallel fashion giving rise to a polar microtubule. In a cell, the so-called minus end of a microtubule, capped by α -subunits, is attached to the centrosome where γ -tubulin and related proteins nucleate microtubules. The more dynamic plus end, capped by β -subunits, binds the kinetochore in mitosis.

The orientation of the tubulin subunits in the microtubule is such that the C-terminal helices form the crest of the protofilaments on the outside surface, making them an essential part of the binding site for motor proteins (kinesins and dyneins) and MAPs. The taxol-binding site is on the inside surface of the microtubule, and close to lateral interactions. The lateral contact between protofilaments is dominated by the interaction of the so-called M-loop in the second domain with loop H1-S2 and helix H3 in the N-terminal, nucleotide-binding domain.

The plus end of the microtubule is crowned by β -tubulin subunits exposing their nucleotide end to the solution, while the minus end is crowned by α -subunits exposing their catalytic end. When a dimer is added to a plus end, its catalytic end contacts the E-site nucleotide of the previous subunit forming the interface that should bring about hydrolysis.

Dynamic instability and GTP cap model

Microtubules are highly dynamic and can switch stochastically between growing and shrinking phases, both *in vivo* and *in vitro*. This nonequilibrium behavior, known as dynamic instability, is based on the binding and hydrolysis of GTP by tubulin subunits. Only dimers with GTP in their E-site can polymerize, but following polymerization this nucleotide is hydrolyzed by interactions with the previous tubulin dimer and then becomes nonexchangeable. In the GTP cap model,

the unstable body of the microtubule made of GDP-tubulin is stabilized by a layer of tubulin subunits at the ends that still retain their GTP. When this cap is stochastically lost, the microtubule rapidly depolymerizes. Depolymerization may occur by weakening of lateral contacts at the ends, and the consequent release of the constrained GDP subunits into a curved, lower energy, conformational state.

Microtubule assembly and stability are further modified in the cell by interaction with cellular factors that stabilize or destabilize microtubules at different points in the cell or at different stages in the cell cycle. Measurement of microtubule dynamics *in vitro* and *in vivo* by differential interference contrast (DIC) or fluorescence microscopy yields the rate constants for addition of tubulin-GTP and loss of tubulin-GDP subunits at the two microtubule ends, as well as the rates of catastrophe (switch from growth to shrinkage) and rescue (switch from shrinkage to growth). A variety of drugs bind to tubulin and affect its polymerization. Microtubule-depolymerizing drugs, such as colchicine, nocodazole, vinca alkaloids, or podophyllotoxin, have a much higher affinity for the dimer than for tubulin in microtubules, so that disassembly is caused by their mass action effect. At substoichiometric concentrations, tubulin-drug subunits bind to the microtubule ends and form caps that dramatically modify microtubule dynamics. Microtubule-stabilizing drugs such as taxoids, epothilones, or discodermolide act by binding preferentially to the polymerized form of tubulin.

An alternative microtubule behavior is treadmilling, a net flow of subunits from the plus to the minus end without a significant change in microtubule length.

Antimitotic tubulin ligands

Recent years have seen the discovery of numerous tubulin ligands with antimitotic properties and anticancer potential. Concerning their binding site these agents can be classified into three main groups: those that bind tubulin at the colchicine-binding site; those that bind it at the vinblastine site; and those that bind it at the taxol site. Functionally, these antimitotic ligands can be separated into two classes: those that inhibit microtubule assembly (e.g., the colchicine and vinblastine families) and those that promote microtubule assembly and stabilization (e.g., the taxol family). In spite of these differences, the main action of all these agents seems to be to cause mitotic arrest by inhibiting normal dynamic instability at very low concentrations.

The colchicine-binding site is located at the monomer-monomer interface within the dimer, in agreement with a model of colchicine action in which binding and stabilizing a bent conformation of the dimer inhibits its polymerization. Vinblastine-like compounds bind at the longitudinal interface between dimers, also stabilizing a bent protofilament structure. Taxol binds to β -tubulin, near the M-loop and the site of lateral interactions, in a hydrophobic pocket that in α -tubulin is occupied by eight extra residues (see Figure 2). New microtubule-stabilizing agents with promise for cancer treatment, such as epothilones, discodermolide, eleutherobin, and the sarcodictins, while very different in structure, all seem to compete with taxol for the binding to a common site. It has been proposed that taxol stabilizes lateral contacts, or alternatively, that it acts as a bridge that holds the N-terminal and

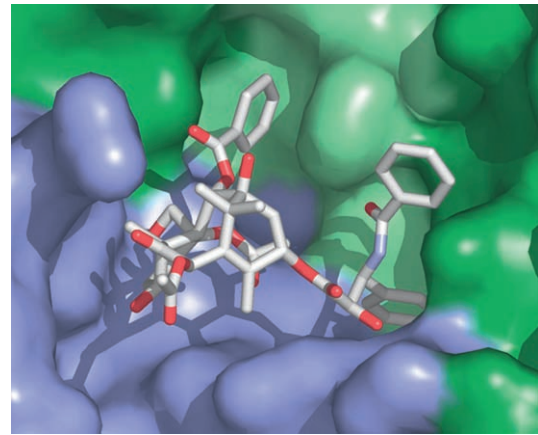


Figure 2 Anticancer drug taxol at its tubulin-binding site. Tubulin is shown on a surface representation with the same color scheme as in Figure 1. Both the N-terminal and intermediate domain are part of the taxol-binding pocket.

second domains in a relative orientation that favors longitudinal contacts between subunits.

γ -Tubulin and microtubule nucleation

Crucial to their dynamic behavior and function is the nucleation of microtubules in the cell at MTOCs, to which they are attached by their minus ends. An essential role in microtubule nucleation is played by γ -tubulin, a protein with high homology to α/β -tubulins that localizes at MTOCs. γ -Tubulin forms ring structures that serve as templates for microtubule growth. The direct involvement of γ -tubulin in microtubule nucleation has been demonstrated *in vitro* using purified γ -tubulin-containing ring complex (γ TuRC) containing at least seven different proteins. There are two main models in the literature. A model based on the shape and size of the γ -TuRC is that the ring forms the first helical turn of the growing microtubule, serving as a template for longitudinal interaction with the tubulin $\alpha\beta$ dimers. An alternative model, based on the similarity of rings structures formed by $\alpha\beta$ -tubulin, γ -tubulin, and FtsZ, is that γ -tubulin forms a protofilament-like structure by longitudinal self-association that then serves as a template for lateral interaction with $\alpha\beta$ -tubulin protofilaments. The X-ray crystallography structure of γ -tubulin shows lateral contacts between proteins similar to those in microtubules, lending support to the ring model.

Rarer tubulin isoforms: δ -, ϵ -, ζ -, and η -tubulins

Four new tubulin isoforms have recently been discovered. The δ -tubulin was identified in *Chlamydomonas* mutants having abnormal basal bodies (these are microtubule structures at the base of cilia and flagella structurally similar to the centrioles in the centrosome at MTOCs). Human δ -tubulin was subsequently found in the human genome database, and shown to localize to the centrosome, where it partially colocalizes with γ -tubulin.

The ϵ -tubulin was identified from the human genome database on the basis of sequence similarity to other tubulins. Like δ -tubulin, ϵ -tubulin localizes to the centrosome, but in a cell-cycle-dependent manner: in cells with duplicated centrosomes ϵ -tubulin localizes only with the old centrosome.

Even rarer, ζ -tubulin has so far only been found on kinetoplastid protozoa where it localizes to the basal body, while η -tubulin has been found in paramecium where it may interact with γ TuRC.

Tubulins and the cell

The involvement of the microtubule cytoskeleton in a large number of essential and diverse functions requires both reliability and flexibility from the system at the expense of biochemical and structural complexity. Dynamic instability is an inherent property of microtubules, built into the $\alpha\beta$ -tubulin structure. The spatial and temporal organization of the microtubule network in the cell is obtained through the regulation of dynamic instability by an increasing number of factors that fine-tune the behavior of the microtubule system to accommodate the requirements of the cell. Regulation may happen at many different stages, via transcription of different tubulin isotypes, the control of tubulin monomer folding, the formation of functional dimers, the posttranslational modification of tubulin subunits, the nucleation of microtubules, or the interaction of microtubules with numerous associated factors, especially those that bind to and affect the behavior of microtubule ends. Tubulin-binding drugs can dramatically disrupt the finely tuned behavior of microtubules. Finally, while γ -tubulin is known to be essential for microtubule nucleation, additional tubulin isoforms, only recently discovered, have yet ill-defined functions.

See also: **Cell Architecture and Function:** Kinesins as Microtubule Disassembly Enzymes; Microtubule-Associated Proteins; **Protein/Enzyme Structure Function and Degradation:** Chaperonins.

Further Reading

- Aldaz H, Rice LM, Stearns T, and Agard DA (2005) Insights into microtubule nucleation from the crystal structure of human γ -tubulin. *Nature* 435: 523–527.
- Desai A and Mitchison TJ (1997) Microtubule polymerization dynamics. *Annual Review of Cell and Development Biology* 13: 83–117.
- Downing KH (2000) Structural basis for the interaction of tubulin with proteins and drugs that affect microtubule dynamics. *Annual Review of Cell and Development Biology* 16: 89–111.
- Lewis SA, Tian G, and Cowan NJ (1997) The α - and β -tubulin folding pathways. *Trends in Cell Biology* 7: 479–484.
- Lowe J, Li H, Downing KH, and Nogales E (2001) Refined structure of alpha beta-tubulin at 3.5 Å resolution. *Journal of Molecular Biology* 313: 1045–1057.
- Ludueña RF (1998) Multiple forms of tubulin: Different gene products and covalent modifications. *International Review of Cytology* 178: 207–275.
- Mitchison T and Krischner M (1984) Dynamic instability of microtubule growth. *Nature* 312: 237–242.
- Nogales E (2000) Structural insights into microtubule function. *Annual Review of Biochemistry* 69: 277–302.
- Nogales E, Whittaker M, Milligan RA, and Downing KH (1999) *Cell* 96: 79–88.
- Nogales E, Wolf SG, and Downing KH (1998) Structure of the $\alpha\beta$ tubulin dimer by electron crystallography. *Nature* 391: 199–203.
- Ravelli RB, Gigant B, Curmi PA, et al. (2004) Insight into tubulin regulation from a complex with colchicine and a stathmine-like domain. *Nature* 428: 198–202.

Tumor Necrosis Factor Receptors

T C Cheung, University of California, San Diego, CA, USA

C F Ware, Sanford|Burnham Medical Research Institute, La Jolla, CA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Karen G. Potter and Carl F. Ware, volume 4, pp. 277–283, © 2004, Elsevier Inc.

Glossary

Apoptosis An orderly process of programmed cell death whereby cellular machinery undergoes changes leading to death.

Autoimmunity An immune response against self-antigens.

Cytokine Proteins made by cells that bind specific receptors that affect the behavior of other cells.

Signal transduction Term used to describe the processes inside cells used to respond to changes in their environment.

Transcription factor A protein that binds DNA to control the transcription of genes involved in a large number of normal cellular and organismal processes.

Features of Tumor Necrosis Factor Receptors

Structure

Tumor necrosis factor receptors (TNFRs) are identified by a highly conserved, cysteine-rich domain (CRD) in the extracellular portion of the protein that binds its specific ligand(s) (Figure 1). The CRD may contain six cysteines that form three disulfide bonds typically recognized by the signature sequence motif CxxCxxC. There are more than 29 members of the cellular TNFR family in mammals, and several variants of TNFRs encoded by herpesviruses and poxviruses. The cellular receptors are primarily type I transmembrane proteins (extracellular N terminus). The cytoplasmic sequence motifs and signaling properties further define the TNFR family into three subsets. The first group includes receptors that contain a death domain (DD), such as TNFR1, Fas, TRAILR1 and 2, and DR3. The DD signaling motif is important for activation of the proteolytic enzymes caspases family that regulates apoptotic cell death. A second subset includes TNFR2, LT β R, HVEM, CD27, 4-1BB, OX40, AITR, and CD30. Their cytoplasmic tails contain binding sites for the TNFR-associated factors (TRAFs) that are important for downstream nuclear factor kappa B (NF- κ B) and activator protein 1 (AP-1) activation. The third group is secreted receptors, which lack transmembrane and cytoplasmic domains, and includes DcR3 and OPG. The soluble receptors are thought to serve as decoys competing with signaling receptors for their cognate ligands.

Given the predominant role of the TNFR family in regulating immunity, this suggests that the evolution of the receptors in this family arose coincident with the evolution of adaptive immunity, also found exclusively in vertebrates. The size of the TNFR superfamily appears to have grown in a large part by gene duplication as many of the TNFR genes are linked to discrete loci reflecting their evolutionary derivation. Perhaps most obvious are those TNFRs found on chromosome (Ch) 12p13 (TNFR1, LT β R, and CD27), which likely underwent duplication and translocation events giving rise to the larger locus of TNFR on Ch 1p36 (TNFR2, HVEM, OX40, CD30, AITR, 4-1BB, and DR3) (Figure 2(a)).

Most members of the TNFRSF have ligands related to TNF; however, some other TNFRs, including neurotrophin p75

receptor (p75 or NGFR), TROY (or Taj), RELT, and DR6, bind ligands unrelated to TNF (Figure 2(b)). The p75 neurotrophin receptor is a functional component within the p75/NgR1/LINGO-1 complex, and can bind to all four neurotrophins: nerve growth factor (NGF), brain derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3), and neurotrophin-4 (NT-4). The p75 subunit serves as a signal transducing receptor for the complex, and its signaling plays an important role in the development of sensory neurons and the repair of nerve damage. Recently, TROY, another orphan TNFRSF, was identified as an alternative component of the p75/NgR1/LINGO-1 complex. TROY is capable of forming a functional receptor complex with NgR1 and LINGO-1 to activate RhoA guanosine triphosphatase (GTPase) in the presence of myelin inhibitors. The herpesvirus entry mediator (HVEM; TNFRSF14) not only binds two TNF-related ligands, LIGHT and LT α , but also engages the Ig superfamily members, B and T lymphocyte attenuator (BTLA) and CD160. HVEM can serve as both a receptor and a ligand.

Another subtle branch in the TNFR family tree are those receptors that engage BAFF, the B-cell survival factor, and related ligands APRIL and TWEAK (Figure 2(b)). These TNFR superfamily members have a single CRD and in the case of BAFFR only two disulfide bonds. Functional divergence is evident in the role some TNFRs play in distinct physiological processes including bone formation (Osteoprotegerin-RANK-RANKL/TRANCE), ectodermal (Ectodermal dysplasin EDA-EDAR), and angiogenesis (TWEAK-Fn14).

Expression

Expression patterns of the receptors are complex. Several members of the TNFR superfamily are expressed on cells of the immune system; for example, BAFFR expression is exclusive to B lymphocytes. Other receptors are found in hair follicles (e.g., EDAR and TROY) or the nervous system (e.g., NGFR); yet others have very broad constitutive tissue expression patterns, such as TNFR1 and LT β R. The inducible expression of some TNFRs, such as OX40, reflects the constitutive expression of its ligand, OX40L. For example, TNFR2 is absent on naive T cells

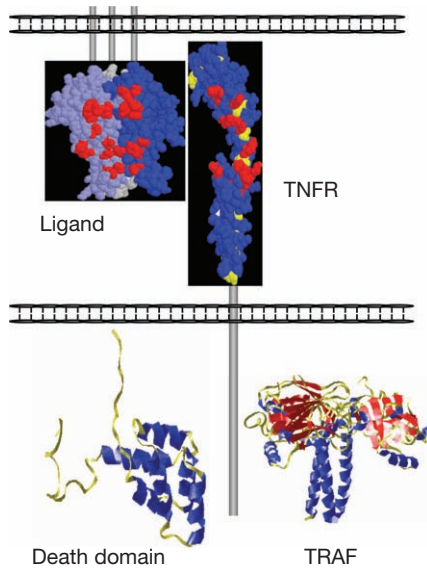


Figure 1 Structural features of TNFR family. Space-filling model of a trimeric TNFR1 ligand, LT α , as it would exist as a membrane-anchored ligand, and a single TNFR1 rotated 180° to reveal contact residues. Also shown are crystal structures of the cytoplasmic DD found in some TNFRs and the TRAF that interacts with TNFR-containing TRAF-binding sites.

but is rapidly upregulated upon activation of T cells. By contrast, HVEM is highly expressed on naive T cells, but shows diminished expression upon T-cell activation. Regulation of receptor expression can also be achieved through the generation of soluble receptors; soluble TNFRs can be generated by proteolytic processing (CD27, CD30, CD40, TNFR1, and TNFR2), alternative splicing of membrane forms (Fas and 4-1BB), or encoded in the genome without a transmembrane region (OPG, DcR3, and TRAIL-R3).

Ligand Binding

TNF ligands are characterized as type II (intracellular N terminus) transmembrane proteins containing a TNF homology domain (THD) in their C terminus (Figure 1). The THD assembles into trimeric proteins; in most cases, the ligands form homotrimers, but LT α and LT β form heterotrimers, and so does BAFF and APRIL. The TNF ligands are active in their membrane forms where cell-to-cell contact is required to initiate signaling; however, some ligands can be proteolytically cleaved from the surface creating soluble cytokines that affect cells distally.

The TNF ligand superfamily currently includes 20 proteins that are able to pair off with one or more TNF receptors. In fact, some ligands display overlapping receptor recognition; for example, LIGHT and LT α 1 β 2 bind LT β R; LT α and LIGHT bind HVEM; BAFF binds TACI, BCMA, and BAFFR. The ligand-receptor pairing seems promiscuous, although comparative studies using mice deficient in a specific TNF ligand or receptor suggest that each cytokine-receptor system has a unique role in physiology.

Furthermore, some TNFRSF members have non-TNF superfamily interacting partners. For example, HVEM interacts with

BTLA or CD160, and these interactions are important for the modulation of coinhibitory signaling for T-cell activation and proliferation (Figure 3(a)). In contrast to the TNFSF ligands, BTLA and CD160 are a type I transmembrane protein with a single Ig-type domain. Interestingly, both BTLA and CD160 also serve as activating ligands for HVEM. Binding of BTLA or CD160 to the first CRD region (CRD1) of HVEM induced HVEM-dependent NF- κ B activation, demonstrating bidirectional signaling between HVEM and BTLA in cells interacting *in trans*. In addition, BTLA and CD160 can self-associate as homo-oligomeric complex on the cell surface, providing a basis for oligomerizing HVEM that leads to TRAF molecules recruitment and activation of NF- κ B RelA. These results highlight the complexity of the TNF-Ig superfamily co-signaling networks.

HVEM interacts with BTLA, CD160, or HSV gD *in cis* within the same membrane, *cis* interactions (Figure 3(b)). The HVEM-BTLA/CD160/gD *cis*-complexes may play important roles in modulating signal transduction among immune cells. Flow cytometric analysis demonstrated that *cis*-association between HVEM and BTLA is the predominant complex expressed on the surface of naive human and mouse T cells, and the formation of HVEM-BTLA *cis*-complex inhibited HVEM-dependent NF- κ B activation. Furthermore, the HVEM-BTLA/CD160/gD *cis*-complex competitively inhibits HVEM *trans*-signaling by all its cellular ligands, providing a mechanism maintaining T cell homeostasis.

TNFR Signaling

Each TNF-related ligand has three receptor binding sites that can cluster together two or three cell-surface receptors, juxtapositioning the cytoplasmic tails of the receptors to initiate signal transduction. Recruitment of specialized signaling molecules (adaptors) to the cytoplasmic domain occurs following receptor clustering. Propagation of TNFR signals occurs through two distinct classes of cytoplasmic adaptor proteins: TRAF or DD molecules.

TRAF-Dependent Signaling

There are six numerically named TRAF proteins (TRAF1–6) that were discovered based on their functions in TNFR-induced signaling; however, TRAF6 and TRAF3 are also key players in signal transduction initiated by the interleukin-1 receptor and the toll-like receptor (TLR) superfamily. TRAF adaptor proteins are a small family of RING finger proteins that play a crucial role in propagating signal transduction leading to the activation of latent transcription factors (Figure 1). The RING domains of TRAF are characteristic of ubiquitin E3 ligases, which has been convincingly demonstrated for TRAF6 and TRAF2. The other TRAF may modify the enzymatic and substrate specificity of the ubiquitination reaction. TRAF can bind directly to the cytoplasmic tails of ligated TNFR. The binding site found in CD40 or the LT β R is a peptide sequence, PXQXI/S or IPEEGD, that fits into a groove in the TRAF homology domain. The binding site in TRAF accommodates a variety of motifs. TNFRs with a DD bind TRAF indirectly via other adaptor proteins, such as TNFR-associated DD (TRADD).

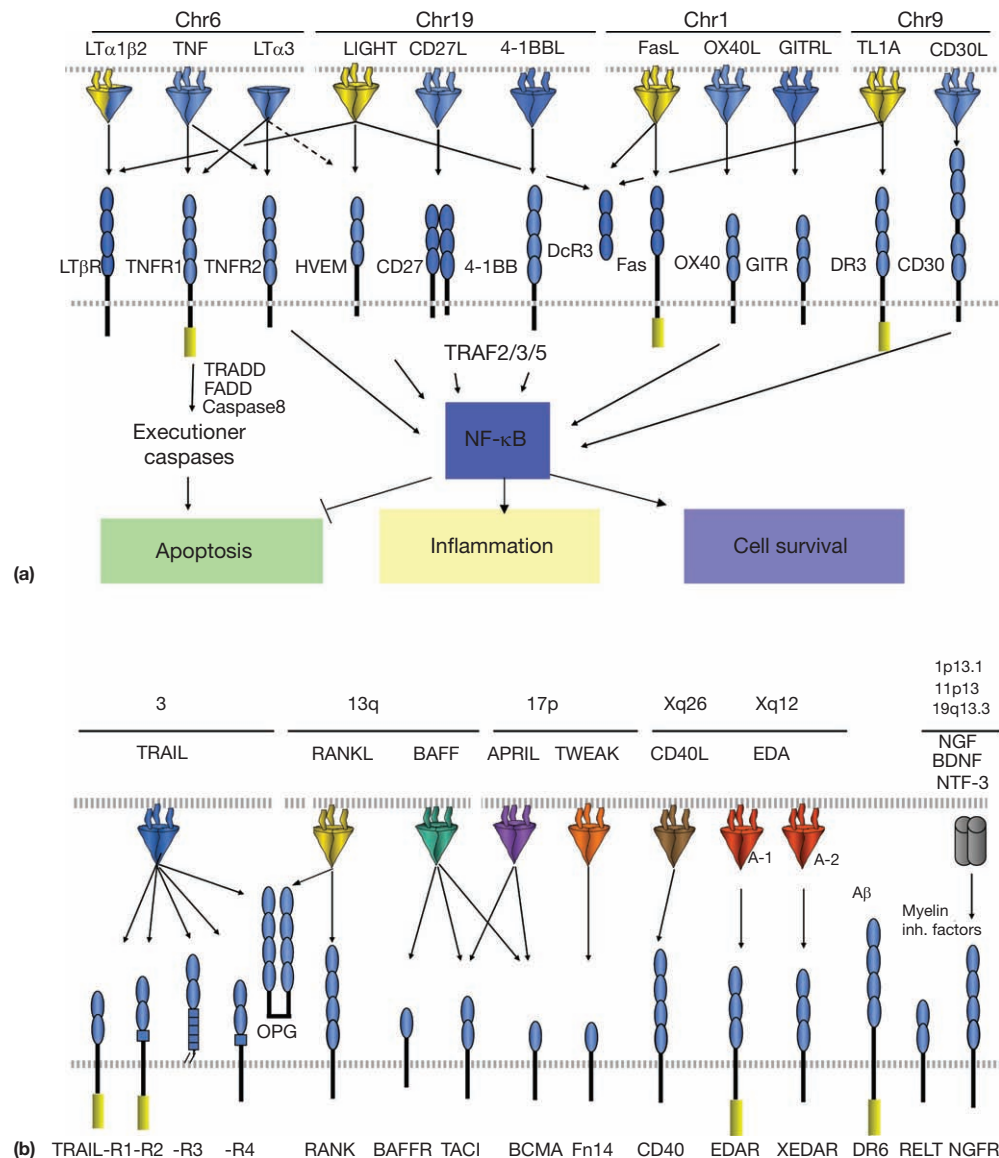


Figure 2 The TNF–TNFR superfamily grouped on the basis of chromosome (Ch) localization of TNFRs. (a) TNFR and TNF ligands are shown with arrows connecting ligand–receptor pairs. CRDs are shown as small ovals. DD is denoted as a rectangular box in the cytoplasmic tail of appropriate receptors. The signaling and effector functions induced by either DD-containing or TRAF-binding TNFR are shown. The panel (b) is the same as (a), with remaining TNFRs grouped on the basis of chromosome localization and ligand binding. Reproduced from Ware CF (2003) The TNF Superfamily. *Cytokine Growth Factor Review* 14(3–4): 181–184.

In the best-studied scenario, TRAF2, TRAF3, and cIAP, another E3 ligase, form a complex that is required to limit the enzyme, NF- κ B-inducing kinase (NIK), in the cell. All three proteins are required to ubiquitinate NIK leading to its continued degradation. The activation of TNFR halts the action of this multisubunit ubiquitin E3 ligase, which allows NIK to accumulate and propagate signaling to IKK α . Functionally, binding of TRAF to TNFR culminates in the activation of transcription factors that act to regulate gene expression. For example, NF- κ B is a small family of latent transcription factors that induces the expression of a large variety of genes involved in inflammatory and immune responses (Figure 4). Two distinct forms of NF- κ B are recognized: NF- κ B1 (RelA, p65) and NF- κ B2 (p100/p52), and each pair with itself or other proteins

(e.g., p50, RelB, and cRel) that form active transcription factors. NF- κ B1 is held in the cytosol by an inhibitor protein (I κ B) while an inhibitory domain controls the cytosolic localization of NF- κ B2. The inhibitory proteins are phosphorylated, ubiquitinated, and degraded in response to activation signals coming from diverse sources that all target the I κ B protein kinase complex (IKK).

Degradation of the inhibitor allows NF- κ B to localize to the nucleus and activate transcription. The NF- κ B1 and B2 pathways induce distinct sets of genes: NF- κ B1 regulates the expression of proinflammatory chemokines and adhesion molecules, while the NF- κ B2 pathway controls the expression of distinct chemokines and cytokines involved in lymphoid organogenesis. BAFFR is only able to activate the NF- κ B2

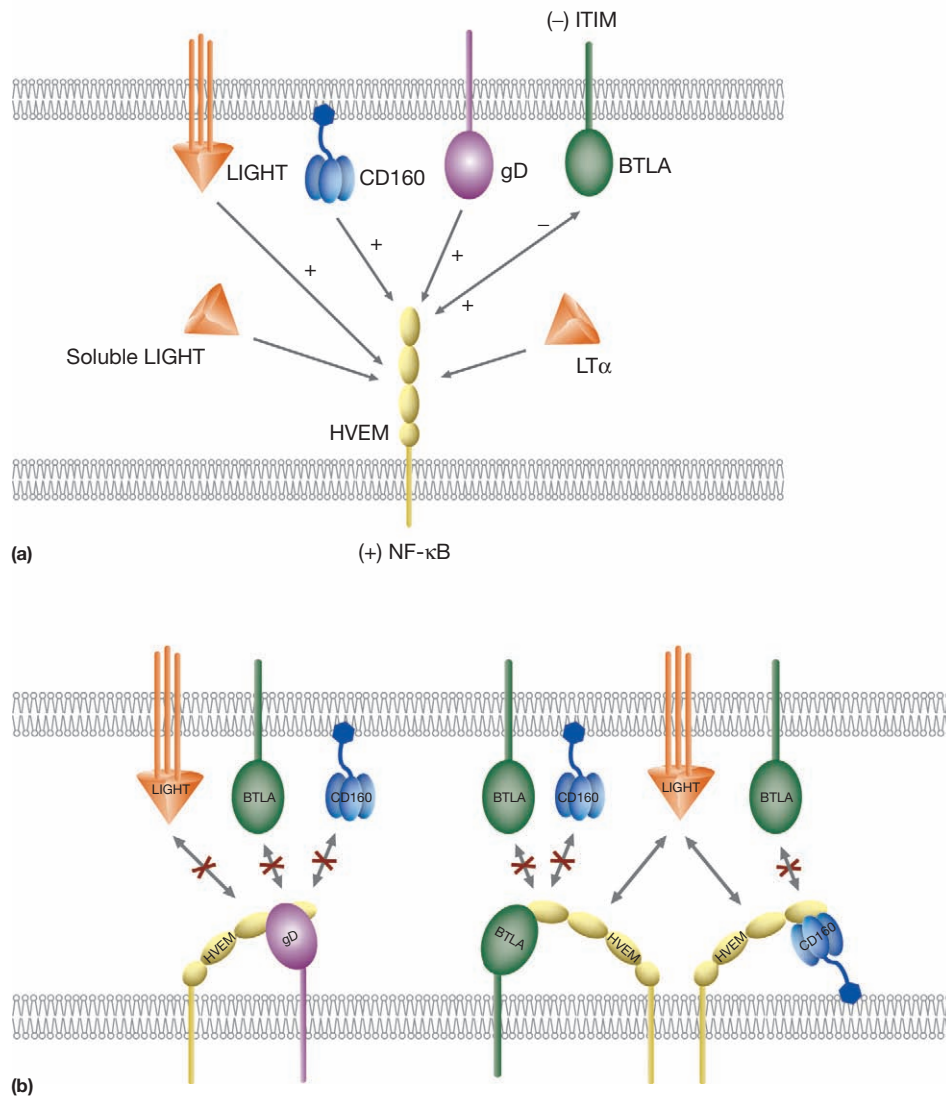


Figure 3 Schematic illustrations of molecular interactions between HVEM and its ligands. (a) Canonical and unconventional ligands of HVEM. LIGHT and LT α are positive regulators and the canonical ligands of HVEM. Ligation of LIGHT or LT α to HVEM activates HVEM-dependent NF- κ B signaling. BTLA, CD160, and HSV gD serve as unconventional ligands for HVEM. Ligation of CD160 or gD to HVEM induces NF- κ B activation. Bidirectional signaling occurs between HVEM and BTLA. HVEM-mediated BTLA signaling induces tyrosine phosphorylation of the ITIM in BTLA, providing an inhibitory signal to T cells. BTLA, CD160, and gD bind HVEM at the CRD1, but LIGHT and LT α interact with HVEM at the CRD2 and 3. The binding of gD to HVEM blocks HVEM binding to LIGHT or BTLA (not shown in diagram). (b) Formation of HVEM-BTLA/CD160/gD *cis*-complexes. *Cis*-interaction between HVEM, BTLA, CD160, and HSV gD inhibits HVEM binding to its ligands *in trans*.

pathway, suggesting that this pathway is crucial for the expression of survival genes specifically important for B lymphocytes. By contrast, TNFR1 is unable to activate NF- κ B2, which may account for its strong proinflammatory action.

Death Receptor Signaling

A number of TNFRs, including TNFR1 and Fas, are termed 'death' receptors because of their ability to induce apoptotic cell death. Their cytoplasmic tails contain a region of ~ 80 amino acids that fold into six α -helices, termed the 'DD'. The DD serves as a protein interaction motif to recruit signaling adapters such as FADD, TRADD, and RIP that recruit proteases of the caspase family, caspase 8.

In the simplest scheme known to activate the cell death machinery, the adaptor FADD is recruited to ligated Fas initiating formation of the death-inducing signaling complex (DISC). Procaspase 8 is recruited to the DISC through a second interaction motif contained in FADD, termed the 'death effector domain (DED)', that converts procaspase 8 into an active enzyme that in turn activates downstream effector caspases (e.g., caspase 3) resulting in the cleavage of critical cellular substrates leading to programmed cell death (apoptosis). Similarly, TNFR1 activates apoptosis although it requires the adaptor protein TRADD to facilitate FADD recruitment, and is also able to recruit another adaptor protein RIP that also plays a role in procaspase 8 activation. The DISC is regulated by two forms of another protein known as FLIP, which can switch the

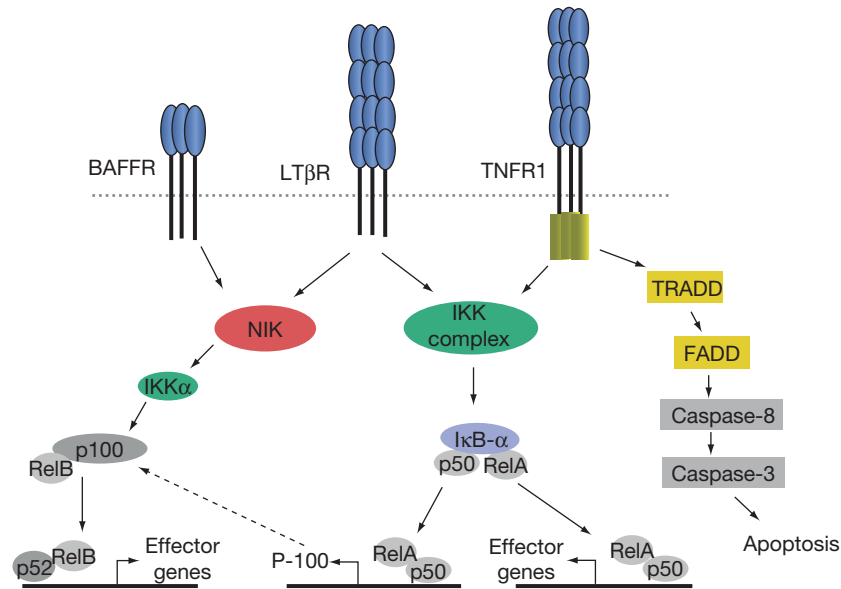


Figure 4 Mechanisms of signaling by TNFRs. Upon ligation of LT β R, two NF- κ B pathways are induced. The first leads to induction of the NF- κ B1 pathway and the activation of IKK β and RelA, which control expression of inflammatory genes. The second pathway results in the activation of NF- κ B2 and the processing of p100 to p52 following the activation of NIK and IKK α , leading to the transcription of genes implicated in secondary lymphoid organogenesis and homeostasis. BAFFR only activates the NF- κ B2 pathway, whereas TNFR1 only activates the NF- κ B1 pathway. TNFR1 also contains a DD to signal caspase activation and apoptosis.

DISC between activating caspases or NF- κ B, which in turn regulates genes that block apoptosis, such as the inhibitors of apoptosis (IAP) that promote cell survival.

Effector Functions of the TNFR Family

Despite similarities in structure and signaling mechanisms, TNFRs function in diverse, and often opposing, roles in immune physiology. As discussed above, several reasons that explain their functional diversity include differences in ligand binding, tissue and cellular expression, regulation of expression, and association with signaling adaptor proteins.

Inflammation

A role for TNF family members in host defense and inflammation was first appreciated when it was realized that TNF is identical to a factor termed 'cachectin', a protein known to cause fever and wasting. Inflammatory cells, such as macrophages, induce TNF when specialized innate immune receptors called TLR recognize molecular patterns on microbial pathogens. Binding of TNF to its receptors, TNFR1 or TNFR2, induces a variety of molecules that are crucial for initiating the acute inflammatory response. For example, TNF signaling induces increased expression of adhesion molecules on endothelial cells and secretion of chemokines to promote inflammatory cell migration to the site of infection. TNF induces nitric oxide synthetase producing NO, a diffusible gas with profound relaxing effects on the cardiovascular system. TNF can also activate macrophages and neutrophils to increase their phagocytic capacity and their ability to secrete tissue-degrading enzymes into tissues. Although inflammation is crucial to prevent

infection, it must also be minimized after the pathogenic challenge subsides in order to prevent uncontrolled damage to fragile host tissues. As such, uncontrolled TNF signaling can cause chronic inflammation and wasting, or acutely released septic shock. Drugs that block TNF signaling, such as soluble TNFR decoy receptors, have been developed to prevent inflammatory diseases such as rheumatoid arthritis.

Apoptosis

Cell death by apoptosis is one of the major functions of DD-containing TNFRs. Fas, for example, mediates cell death for varied purposes. First, Fas and its ligand FasL cooperate as major effectors, along with the perforin/granzyme pathway, to trigger killing of pathogen-infected cells by cytotoxic T lymphocytes (CTLs) or by natural killer (NK) cells. In a second context, Fas is required to eliminate excessive CTLs and other effector cells. Fas is induced on long-term antigen-activated lymphocytes to elicit their removal and minimize the excessive expansion of T lymphocytes that can occur in response to T-cell activation in the space-restricted lymphoid environment. In fact, two naturally occurring mutations in mice encoding the gene *lymphoproliferative disorder (lpr)*, *lpr* or *lpr^{cg}*, that affect the expression or the biological activity of the Fas receptor, respectively, demonstrate the important role of Fas in lymphocyte homeostasis. Both of these mutant mice develop an autoimmune phenotype characterized by massive lymphadenopathy, splenomegaly, and autoantibody production.

Cell Survival

BAFFR, BCMA, and TACI are predominantly expressed on B cells and represent a good example of TNFRs that play a role in cell survival. All three receptors bind the TNF-related ligand

BAFF, produced by macrophages and dendritic cells, while BCMA and TACI can also bind the related ligand APRIL. BAFF is required in the maintenance and differentiation of mature B cells in peripheral tissues by supporting B-cell survival and maturation presumably through activation of anti-apoptotic genes. The mechanism for enhanced B-cell survival by BAFF may be partly due to the stimulation of the pro-survival protein bcl-2 and a decrease in the expression of the pro-apoptotic factors. A crucial role for BAFF in BAFF-mediated B-cell survival is evidenced by the fact that neither BCMA- nor TACI-deficient mice have defects in B-cell maturation, whereas A/WySnJ mice, containing a naturally occurring mutation in BAFFR, show defects in B cell maturation. Overexpression of BAFF in mice can result in autoimmune symptoms; similarly, patients with systemic lupus erythematosus (SLE), rheumatoid arthritis, and Sjogren's syndrome show elevated levels of BAFF in the blood. Other TNFRs also play roles in the survival and development of various tissues including the TRANCE system in bone, EDAR in ectodermal tissues, and NGFR in neurons.

Co-stimulation

T-lymphocyte activation is also regulated by several TNFRs where they act as costimulatory signals to positively or negatively influence the proliferation of antigen-stimulated T cells. OX40, CD27, 4-1BB, and HVEM, expressed on CD4⁺ and CD8⁺ T cells, promote T-cell proliferation and cytokine production in response to their respective ligands expressed on dendritic cells. For example, mice deficient in the OX40–OX40L interaction show impaired T-cell proliferative capabilities and cytokine production in response to viruses such as lymphocytic choriomeningitis virus (LCMV) or influenza virus, and the ability to generate or sustain memory T cells is diminished. HVEM through interaction with BTLA or CD160 limits naive T-cell activation and dendritic cell differentiation. Thus, TNFR family members can contribute to immune tolerance.

In B cells, CD40 is a prominent costimulatory receptor where it interacts with its ligand, CD40L, expressed in T cells to initiate T-cell-dependent B-cell antibody responses. Mutations in CD40 or its ligand affect a number of immune functions including Ig class switching; in humans, X-linked hyper-IgM syndrome is due to defects in CD40L structure and function manifesting as an inability to switch IgM to IgG classes.

Organogenesis and Development

A number of TNFRs play roles in the organization of tissues. The lymphotoxin system mediated by LT β R, and its ligand LT $\alpha\beta$, plays a crucial role in lymph node development. Mice deficient in either receptor or ligand completely lack lymph nodes and Peyer's patches. RANK and its ligand RANKL mediate osteoclast differentiation and therefore are important for maintaining normal bone homeostasis. Osteoporosis, a condition of bone thinning, can be caused by overexpression of the soluble RANKL decoy receptor, OPG, which is induced by estrogen and prevents normal bone formation. RANK–RANKL signaling also supports the development of mammary gland

alveolar buds important for lactation, and consequently an absence of RANK induces accelerated apoptosis of mammary epithelial precursors. In fact, mice or humans with deficiencies in EDAR or EDA lack primary hair follicles and sweat glands. NT p75 receptor (p75 or NGFR) is important for the development of sensory neurons; acting alone it induces apoptosis, but upon association with other NGF receptors it mediates the differentiation and survival of neurons. p75-deficient mice develop cutaneous sensorineuronal defects.

Viral Targeting of TNFR

Since several TNFRs and their ligands play a crucial role in mediating immune defenses against invading pathogens and control death and survival fates for cells, it is not surprising that viruses have evolved mechanisms that directly disrupt communication mediated by the TNF family in order to subvert immune effector mechanisms. In fact, viruses, when viewed collectively, have targeted virtually every step of the TNFR signaling pathway, from ligand binding to activation of apoptosis. For example, some viruses, including cytomegalovirus (CMV), herpes simplex virus (HSV), adenovirus, and HIV, induce the expression of FasL and/or TRAIL in the host cell resulting in the probable killing of recruited immune effector cells that express the cognate TNFR death receptor. By contrast, adenovirus downregulates the proapoptotic receptors Fas and TRAILR from the surface of infected cells, thereby preventing the killing of the host cell harboring the virus. Another mechanism of immune evasion employed by poxviruses is the expression of soluble, secreted orthologs of TNFR2 or CD30 that inhibit the interaction of their respective ligands with their cellular receptors. In some cases, viruses and TNFR signaling have been adapted to function cooperatively. For instance, signaling through the LT β R or TNFR1 can prevent the replication of human CMV in fibroblasts without inducing death of the infected cell thereby allowing continued existence of CMV within the host.

See also: **Bioenergetics:** Cell Death by Apoptosis and Necrosis; **Cell Architecture and Function:** Bcl2 Family: Cell Death Enhancers and Inhibitors; **Signaling:** Chemokine Receptors; Mitogen-Activated Protein Kinase Family; Nuclear Factor Kappa B; Toll-Like Receptors

Further Reading

- Benedict CA, Banks TA, and Ware CF (2003) Death and survival: Viral regulation of TNF signaling pathways. *Current Opinion in Immunology* 15: 59–65.
- Bodmer J, Schneider P, and Tschopp J (2002) The molecular architecture of the TNF superfamily. *Trends in Biochemical Sciences* 27: 19–26.
- Chung JY, Park YC, Ye H, and Wu H (2002) All TRAFs are not created equal: Common and distinct molecular mechanisms of TRAF-mediated signal transduction. *Journal of Cell Science* 115: 679–688.
- Dejardin E, Droin NM, Dehase M, et al. (2002) The lymphotoxin- β receptor induces different patterns of gene expression via two NF κ B pathways. *Immunity* 17: 1–11.
- Locksley RM, Killeen N, and Lenardo MJ (2001) The TNF and TNF receptor superfamilies: Integrating mammalian biology. *Cell* 104: 487–501.
- Orlinick JR and Chao MV (1998) TNF-related ligands and their receptors. *Cellular Signalling* 10: 543–551.

Two-Hybrid Protein–Protein Interactions

I Serebriiskii, S Peri, and E A Golemis, Fox Chase Cancer Center, Philadelphia, PA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Bait A term used to describe a DNA-binding domain–protein X fusion, used as a probe in the yeast two-hybrid system (e.g., to screen a library).

Interaction map A schematic showing a network of protein–protein interactions involving one or more proteins of interest.

Interactome A term describing the complete network of protein–protein interactions occurring within an organism.

Prey A term used to describe a transcriptional–activation domain–protein Y fusion, which interacts with a bait.

Two-hybrid system A system in which the interaction of two hybrid proteins (created so that (1) one is a fusion between a DNA-binding domain and protein X, and the second is a fusion between a transcriptional–activation domain and protein Y and (2) X and Y normally interact) in a host organism such as yeast or bacteria causes the activation of one or more easily detectable reporter genes.

The Yeast Two-Hybrid System

The purpose of the two-hybrid system is to provide a genetic system that is capable of scoring the degree of physical interaction between two proteins of interest. To this end, a simple organism such as yeast (*Saccharomyces cerevisiae*) or bacteria (*Escherichia coli*) is engineered such that the interaction of two proteins generates a detectable signal that can be used either to gauge the interaction affinity of two defined proteins or to identify interacting partners for one defined protein by screening an expression library.

The first working example of a two-hybrid system was described in 1989, by Fields and Song. A schematic showing the system paradigm is shown in [Figure 1](#). There are three minimal components for a functioning two-hybrid system. The first component is an expression construct, which synthesizes a first protein of interest as a fusion with a DNA-binding domain (DBD) that binds a short DNA sequence of defined specificity (hybrid 1: frequently termed the ‘bait’). The second component is an expression construct that is used to express either a second defined protein or random clones from a complementary DNA (cDNA) or genomic library, as a fusion with a transcriptional–activation domain (AD) (hybrid 2: frequently termed the ‘prey’). The third component is a reporter cassette, which consists of the DNA-binding site for the first hybrid protein in the context of a minimal promoter, upstream of the coding sequence for an easily scored reporter gene.

For the yeast two-hybrid system to work, there are two preconditions. First, the bait must not be independently functional as a transcriptional activator, such that yeast containing only the bait and the reporter produce no reporter signal. Second, the bait and the prey must be capable of localizing to the cell nucleus; hence, proteins containing signal sequences that efficiently target them to the plasma membrane, or other non-nuclear compartments, are not generally acceptable as baits. However, for proteins that meet these preconditions, the interaction between the bait and the prey brings the transcriptional–AD encoded in the prey to the promoter sequence bound by the bait and activates the transcription of the reporter gene.

A number of groups have built systems that exploit this two-hybrid paradigm in yeast, with several of these systems in common use. The most common DBDs used in these yeast two-hybrid systems are derived from the yeast transcription factor Gal4p or from the bacterial repressor protein LexA. A variety of different AD sequences have been used, among them the AD from Gal4p; the AD from the viral protein VP16; or ‘artificial’ activation sequences encoded from selected fragments of the *E. coli* genome.

Two categories of reporter genes have been used. One category is colorimetric, exemplified by the bacterial lacZ gene. Activation of a lacZ reporter causes yeast colonies to turn blue when plated on medium containing an appropriate substrate, such as X-gal, or can be assessed quantitatively over a dynamic range in excess of 1000-fold in liquid assays using a spectrophotometer or luminometer. The second category of reporter provides an active selection for yeast growth based on the strength of interaction between two proteins. These reporters include genes encoding critical components of yeast metabolic pathways for amino acid synthesis, such that expression of the gene is required for yeast to grow on medium lacking the relevant amino acid (e.g., show an auxotrophic requirement). *HIS3*, required for growth on medium lacking histidine, and *LEU2*, required for growth on medium lacking leucine, have been most commonly used. In general, two-hybrid systems now generally incorporate both colorimetric and auxotrophic reporters, both responsive to the same bait. The use of at least two different reporter genes greatly minimizes the background of false positive interactions in the system, and is essential for library screening applications, wherein excess of 10^6 clones must commonly be analyzed.

Because of the ease of manipulation provided by a genetic system, a yeast two-hybrid system provides a facile means of exploring the interaction of two known partner proteins. This can be done in a predetermined manner, by inserting targeted mutations or deletions into the AD-fused partner, or based on blind selection, for example, by using polymerase chain reaction (PCR) mutagenesis or DNA sonication to develop a randomized library derived from the AD partner. Each mutant derivative can be readily screened for interaction

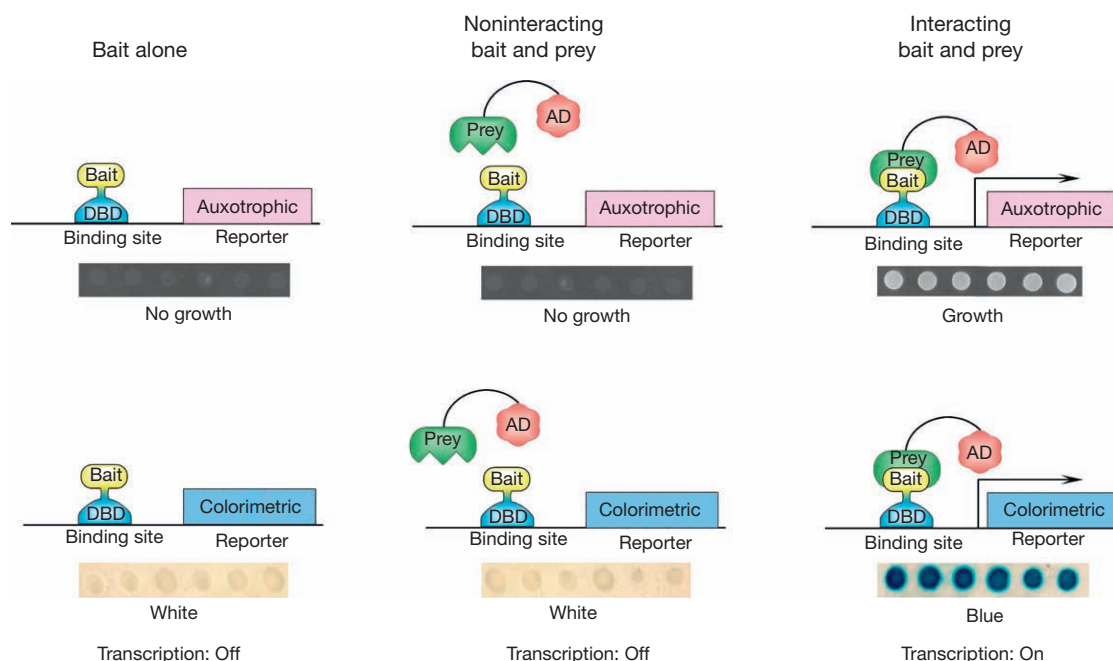


Figure 1 Schematic of yeast two-hybrid components, and profile of reporter activation. Details of system components are described in the text. Three situations are shown: (1) yeast containing only a bait and two reporters, left; (2) yeast containing bait, prey that does not interact with bait, and reporters, center; and (3) interacting bait and prey, and reporters, right. For each of these three situations, with a typical auxotrophic and colorimetric reporter, six independent colonies of yeast are dotted on plates with selective medium: expected results (growth/no growth, or color/no color) are shown as panel inserts below each schematic.

phenotype. However, the most important use of the yeast two-hybrid system throughout the 1990s has clearly been for the purpose of screening libraries to identify novel interacting partners for proteins of interest. In contrast to other means of identifying interacting proteins, such as biochemical copurification, two-hybrid screening is extremely inexpensive, requiring only engineered strains of yeast and standard microbiological media. It is rapid, providing the potential of completing a screen from start to finish in less than a month. It is effective at yielding returns for a significant percentage of bait proteins (estimated at >50%). Finally, following completion of a successful library screen, plasmids encoding the novel interacting preys can easily be isolated from yeast, avoiding the need for protein sequencing or mass spectrometric analysis required by biochemical methodologies. It is not an exaggeration to note that recent broad adaptation of two-hybrid screening methodologies across the scientific community (with the average number of the two-hybrid related publications in PubMed around ~770 in the years 1999–2009, and up to ~21 000 interactions year⁻¹) was a major contributing source of elucidation for key cell-signaling pathways.

Advanced Applications

Following the validation of the basic two-hybrid system as a useful and reliable tool for study of protein–protein interactions, a number of investigators have explored the potential of the system to study interactions between proteins and partner proteins under additional selective constraints. A first step in this direction was the demonstration that it was possible to identify preys that interacted with baits contingent upon the

modification of the bait by a coexpressed kinase. A second point of interest was the feasibility of identifying proteins that interacted not with a single protein, but with a complex formed by a bait and one or more simultaneously expressed co-baits. A number of investigators have used a two-hybrid approach to identify proteins that interact as constituents of ternary or even quaternary complexes.

Another application of interest has been to use the two-hybrid system to identify and characterize interactions between proteins and nonprotein (e.g., small molecules/drugs and RNAs). In each case, the derivative system has been shown to function at the level of library screening. While screening for RNAs has become relatively broadly adapted as a research tool, screening for drugs that regulate protein–protein interactions was outcompeted by more robust alternative technologies. The two-hybrid system has also been used by a number of investigators to identify peptide aptamers capable of binding and regulating the functions of specific bait proteins. Peptides isolated based on two-hybrid selection have been found capable of regulating diverse functions of their cognate baits, and may provide the basis for design of peptidomimetics or small molecule derivatives of therapeutic value. This is currently an area of active investigation.

Finally, efforts have been made to improve ability of the two-hybrid system for evaluating the selectivity of protein interactions. In these approaches, the two-hybrid system is parallelized, using two discrete sets of baits and reporters within a single strain of yeast. In one example, bait 1 (a fusion to LexA) directs the expression of the *lacZ* and *LEU2* reporters, while bait 2 (a fusion to cI) directs the expression of *gusA* (colorimetric) and *LYS2* (auxotrophic) reporters. In a situation

where a prey initially interacts with both baits 1 and 2, turning on all four reporters, mutagenesis of the prey can be coupled with selective screening for activation of the *lacZ* and *LEU2*, versus the *gusA* and *LYS2*, reporters, providing a convenient means of identifying altered specificity mutants that can be used to deconvolute signal-transduction pathways.

Alternative Two-Hybrid Systems

Although the yeast two-hybrid system has found many useful applications, it has not been appropriate in all cases. Toward the goal of identifying small molecules or mutations that disrupt, rather than permit, protein–protein interactions, a reverse two-hybrid system has been developed. In this system, activation of a reporter gene is toxic, providing selection pressure for clones that have lost protein–protein interactions. For proteins that strongly activate transcription in yeast, and hence cannot be used to create baits, one solution has been to move the two-hybrid components to an RNA polymerase III-based reporter system, eliminating transcriptional activity. Other investigators have generated two-hybrid-like systems in bacteria, where eukaryotic transcriptional-activation sequences are generally ineffective. For proteins that are membrane associated, or cannot efficiently be nuclear localized, several methods for assessing protein interaction at membranes or in the cytoplasm have been developed. In some techniques, interaction of the bait and prey reconstitutes an intracellular signaling pathway (e.g., the Sos and Ras recruitment systems). In another approach, called the split-ubiquitin membrane yeast two-hybrid system, interaction of a bait and prey at the cell membrane results in the reconstitution of functional ubiquitin fused to LexA-VP16 transcription factor, which is subsequently cleaved and migrates to the nucleus, activating standard yeast two-hybrid colorimetric and auxotrophic reporters. This approach combines advantages of membrane-based interaction detection with the ability to use well-validated screening modalities.

Role in Proteomics

Two hybrid protein–protein interaction systems are important elements of strategies to analyze the complete protein–protein interactions of proteomes. To date, extensive protein interaction maps have been constructed for representatives of all major groups of living organisms, including animals, plants, fungi, bacteria, and viruses (as well as virus–host interactions). In the case of model organisms, advances in screening methodologies including low- and high-throughput screening

methods contributed to large-scale identification of binary interactions. Concerted efforts in yeast (*S. cerevisiae*), worm (*Caenorhabditis elegans*), fruit fly (*Drosophila melanogaster*), and *Homo sapiens* revealed tens of thousands of protein–protein interactions. These efforts provide a useful complement to simultaneous efforts to analyze protein interactions using protein complementation assay and tandem affinity purification/mass spectrometry. Among these techniques, each having inherent advantages and constraints, the yeast two-hybrid system appears superior at identifying binary protein–protein interactions for nontransmembrane proteins. Interactions being detected through these methods are being integrated into databases including transcriptional profiles, genetic interaction data, and protein localization data for the same proteins, providing a fundamental tool for the emerging field of systems biology. It is likely that continuing application of the two-hybrid protein–protein interaction will remain a source of scientific insights for many years to come.

See also: Protein/Enzyme Structure Function and Degradation: Phage Display for Protein Binding.

Further Reading

- Bernstein DS, Buter N, Stumpf C, and Wickens M (2002) Analyzing mRNA–protein complexes using a yeast three-hybrid system. *Methods* 26: 123–141.
- Fashena SJ, Serebriiskii IG, and Golemis EA (2000) The continued evolution of hybrid screening approaches in yeast: How to outwit different baits with different preys. *Gene* 250: 1–14.
- Fields S (2009) Interactive learning: Lessons from two hybrids over two decades. *Proteomics* 9: 5209–5213.
- Fields S and Song O (1989) A novel genetic system to detect protein–protein interaction. *Nature* 340: 245–246.
- Ideker T, Tan K, and Uetz P (2005) Visualization and integration of protein–protein interactions. In: Golemis EA (ed.) *Protein Interactions*, 2nd edn., pp. 839–856. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Kittanakom S, Chuk M, Wong V, et al. (2009) Analysis of membrane protein complexes using the split-ubiquitin membrane yeast two-hybrid (MYTH) system. *Methods in Molecular Biology* 548: 247–271.
- Kotova E, Coleman T, and Serebriiskii I (2009) Two-hybrid dual bait system. *Current Protocols in Molecular Biology* 20(7). Hoboken, NJ: Wiley.
- Li S, Armstrong CM, Bertin N, et al. (2004) A map of the interactome network of the metazoan *C. elegans*. *Science* 303: 540–543.
- Ratushny V and Golemis E (2008) Resolving the network of cell signaling pathways using the evolving yeast two-hybrid system. *Biotechniques* 44: 655–662.
- Serebriiskii IG, Fang R, Latypova E, et al. (2005) A combined yeast/bacteria two-hybrid system: Development and evaluation. *Molecular and Cellular Proteomics* 4: 819–826.
- Suter B, Kittanakom S, and Stagljar I (2008) Two-hybrid technologies in proteomics research. *Current Opinion in Biotechnology* 19: 316–323.

Tyrosine Sulfation

D Corbeil and W B Huttner, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 4, pp. 294–298, © 2004, Elsevier Inc.

Glossary

PAPS 3'-Phosphoadenosine 5'-phosphosulfate, sulfate donor in the sulfate transfer reaction. PAPS has been known to be the activated form of sulfate and acts as cosubstrate for the sulfation of a wide variety of substances, including proteins.

trans-Golgi network The last station of the Golgi complex. This site is a major branching point of vesicular transport and the origin of two principal pathways of protein secretion: the regulated and constitutive pathways.

The O-sulfation of tyrosine residues of membrane and secretory proteins that transit through the secretory pathway of eukaryotic cells is a ubiquitous posttranslational modification conserved in all multicellular organisms. Tyrosine sulfation is catalyzed by tyrosylprotein sulfotransferase (TPST) isoenzymes, which are integral membrane proteins of the *trans*-Golgi network. Tyrosine sulfation has been shown to be important for protein–protein interactions occurring in diverse biological processes, ranging from the receptor binding of regulatory peptides to the interaction of viral envelope proteins with the cell surface.

Tyrosine-Sulfated Proteins

Occurrence

Sulfation is one of the most abundant posttranslational modifications of tyrosine, since up to 1% of the tyrosine residues of total protein in an organism can be sulfated. Tyrosine is the only amino acid residue in proteins known to undergo sulfation. Following the first description of a sulfated-tyrosine residue in a peptide derived from fibrinogen by Bettelheim in 1954, it has been known since 1982 that tyrosine-sulfated proteins occur in all animals, from lower invertebrates up to humans. Tyrosine-sulfated proteins also exist in the plant kingdom, for example, in the green alga *Volvox*, one of the earliest truly multicellular organisms, or in higher plants such as rice. While occurring throughout metazoan evolution, tyrosine-sulfated proteins appear to be absent in unicellular eukaryotes and prokaryotes, implicating this posttranslational modification in some aspect of multicellularity.

In a given animal, tyrosine-sulfated proteins have been observed in all tissues examined. Each tissue appears to contain a characteristic set of tyrosine-sulfated proteins, suggesting that proteins with tissue-specific expression are major targets for tyrosine sulfation. In cell culture, tyrosine sulfation of proteins has been detected in all primary cultures and cell lines investigated, including various secretory cells, epithelial cells, fibroblasts, neuronal cells, and cells of the immune system.

Tyrosine-sulfated proteins can be identified by various methods, including labeling using radioactive sulfate followed by tyrosine sulfate analysis of a given protein. In line with the intracellular localization of TPST in the *trans*-Golgi network

(Figures 1(a) and 1(b)), all known tyrosine-sulfated proteins are either secretory or plasma membrane proteins. Reviews with comprehensive lists of tyrosine-sulfated proteins have been published, and several of these proteins have been shown to play important biological roles.

Structural Determinants of Tyrosine Sulfation

The recognition, by TPST, of the tyrosine residue to be sulfated in a secretory protein or the extracellular domain of a membrane protein requires the presence of certain structural features. These have been deduced from the comparison of identified tyrosine sulfation sites and the *in vitro* tyrosine sulfation of synthetic peptides. Although no strict consensus sequence for tyrosine sulfation exists, all sequences exhibit the presence of acidic amino acid residues in the vicinity, that is, positions -5 (N-terminal) to $+5$ (C-terminal), of the sulfated tyrosine residue. A particularly critical position appears to be amino-terminal (-1) to the tyrosine. Turn-inducing amino acids (P, G) are also frequently present. The few examples of tyrosine-sulfation sites lacking proline and glycine are located near the N- or C-terminus of the protein and/or contain several of the three other amino acid residues (D, S, N) with significant turn-conformational potential. Finally, another characteristic feature common to most identified tyrosine-sulfation sites is the absence of cysteine residues or potential N-glycosylation sites (NXS or NXT) in the vicinity (positions -7 to $+7$). In either case, the presence of a disulfide bridge or N-linked oligosaccharides is likely to prevent sulfation of nearby tyrosine residues due to steric hindrance.

Tyrosylprotein Sulfotransferase (EC 2.8.2.20)

The Tyrosine-Sulfation Reaction

The sulfate transfer reaction to tyrosine residues is catalyzed by TPST, first described in 1983, and used as a sulfate donor, the cosubstrate 3'-phosphoadenosine 5'-phosphosulfate (PAPS) (Figure 1(b)). The transfer is thought to occur in an ordered reaction mechanism that involves the following sequential steps: (1) cosubstrate PAPS binding; (2) substrate binding;

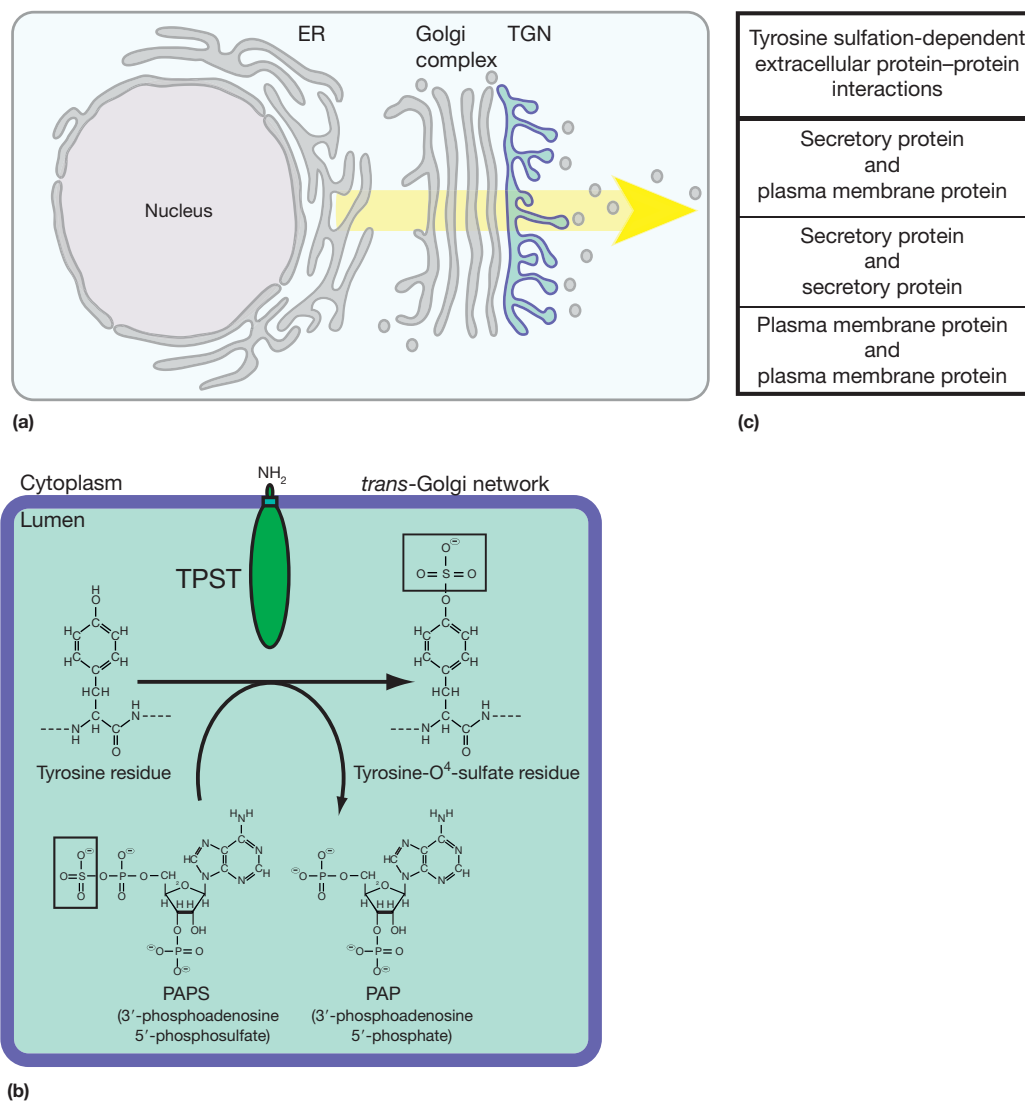


Figure 1 (a) Tyrosine sulfation occurs in the *trans*-Golgi network (TGN). (b) TPST, a type II transmembrane protein, catalyzes the transfer of the sulfate group (boxes) from the co-substrate PAPS to tyrosine residues of secretory proteins and ectodomains of membrane proteins passing through the lumen of the *trans*-Golgi network. (c) The three principal types of tyrosine sulfation-dependent protein-protein interactions (for details, see text).

(3) sulfate transfer; (4) tyrosine-sulfated product release; and (5) PAP release. TPST activity can be detected by incubation of appropriate membrane preparations with [³⁵S]PAPS using either endogenous or exogenous protein substrate. Biologically, tyrosine sulfation appears to be an irreversible event *in vivo* due to the lack of a sulfatase capable of catalyzing the desulfation of tyrosine-sulfated proteins under physiological conditions.

Properties of TPST

TPST is an integral membrane protein residing in the *trans*-Golgi network (Figures 1(a) and 1(b)). The TPST protein can be identified in membrane preparations by modification after substrate cross-linking (MSC) labeling, which is based on the cross-linking of a substrate peptide to TPST followed by intramolecular [³⁵S] sulfate transfer from the cosubstrate PAPS. Various nonionic detergents can be used to solubilize TPST from membranes.

TPST purified from bovine adrenal medulla is a 50–54-kDa sialoglycoprotein with an apparent S-value of 6. Its pH optimum is between 6.0 and 6.5, which is in line with the slightly acidic pH of the *trans*-Golgi network. The catalytic activity of the enzyme, toward endogenously as well as exogenously added proteins, is stimulated by divalent cations (Mg²⁺, Mn²⁺), and fluoride ions are required for maximal activity. The *K_m* for the cosubstrate PAPS is 1.4 μM. Like most sulfo-transferases, TPST is inhibited by PAP. *In vitro*, the activity of the enzyme is also inhibited by certain lipids such as sphingosine, but the physiological relevance of this observation is unknown. Synthetic inhibitors of TPST (with IC₅₀ values of 30–40 μM) have been generated by a combinatorial target-guided ligand assembly technique.

The *K_m* for most peptides (8–13 amino acid residues) with a single tyrosine sulfation site is in the range of 10–100 μM. It is likely that similar values hold true for individual tyrosine-sulfation sites in proteins. Interestingly, *K_m* values for synthetic

substrates containing multiple tyrosine-sulfation sites are considerably lower. Given that several proteins exist with multiple adjacent sulfation sites, for example, cionin, heparin cofactor, and prepro-cholecystokinin, the increased affinity of TPST for such substrates might be of physiological significance, for example, by promoting stoichiometric sulfation. A remarkable example of the sequential sulfation of four tyrosine residues within a short stretch of amino acids is the N-terminal domain of the CC-chemokine receptor 5 (CCR5), which is a coreceptor for HIV.

Molecular Cloning and Membrane Topology of TPST

Distinct TPST isoenzymes, anticipated from the differential substrate specificities of various TPST preparations, exist in the human and mouse genome. Each contains two TPSTs, TPST-1 (see Swiss-Prot Database, Accession No. 060507) and TPST-2 (see Swiss-Prot Database, Accession No. 060704), which were first characterized at the molecular level in 1998. The human *TPST-1* (7q11) and *TPST-2* (22q12.1) genes encode for 370- and 377-amino acid proteins, respectively, that share an overall 65% identity. The murine *tpst* genes are located on chromosome 5, and the corresponding proteins show ~95% identity to their human counterpart. Both TPST isoenzymes are predicted to have a type II membrane topology, with a very short NH₂-terminal cytoplasmic domain (eight residues) and the bulk of the polypeptide, which is responsible for the catalytic activity, being located in the Golgi lumen (see Figure 1). Although overall, TPST isoenzymes display a low degree of amino acid identity to other cytosolic and Golgi-associated sulfotransferases, most of the residues involved in PAPS binding, as deduced from comparison with the crystal structure of estrogen sulfotransferase, are conserved. The luminal domain of either isoenzyme also contains two potential N-glycosylation sites, consistent with presence of N-linked glycans in the TPST protein.

In agreement with the occurrence of tyrosine sulfation in all metazoan species, complementary DNAs (cDNAs) that predict proteins obviously related to mammalian TPSTs are found in various vertebrates and invertebrates, including fish (GenBank Accession No. B1846282), frog (GenBank Accession No. BG815875), chicken (GenBank Accession Nos. BU142429, BU217098), fly (GenBank Accession No. AY124548), and worm (GenBank Accession No. NM_067245).

TPST-1 versus TPST-2

TPST-1 and TPST-2 transcripts are found in all tissues examined, for example, brain, heart, skeletal muscle, gut, kidney, liver, lung, and leukocytes. Although ubiquitously expressed, the tissue distribution of TPST-1 and TPST-2 are not identical. Recombinant TPST-1 and TPST-2 show similar, but not identical, activities toward certain small peptide substrates, consistent with some functional redundancy as well as a certain degree of differential substrate specificity.

TPST-1- and TPST-2-Deficient Mice

TPST-1- and TPST-2-deficient mice, generated by targeted disruption of the *tpst-1* and *tpst-2* genes, also point to functional

redundancy between the two isoenzymes, as either line of knockout mice is viable. Although TPST-1^{-/-} animals appear normal, their body weight is reduced about 5% and an increase in postimplantation fetal death is observed, suggesting that unidentified proteins involved in regulation of body weight and reproductive physiology require tyrosine sulfation for optimal function. TPST-2-deficient mice show a transient delay in growth during postnatal development. In addition, the males, but not females, appear to be infertile. Together, these observations are consistent with the idea that TPST-1 and TPST-2 have distinct, but partially overlapping, physiological roles.

Physiological Function and Medical Relevance

For most tyrosine-sulfated proteins, the physiological function of this posttranslational modification is presently unknown. With regard to the cases in which the biological role of tyrosine sulfation of a particular protein has been elucidated, the common denominator has emerged that tyrosine sulfation promotes extracellular protein-protein interactions. In line with the identification of numerous secretory and plasma membrane proteins that are tyrosine sulfated, paradigmatic examples exist showing that tyrosine sulfation promotes the interaction between (1) a secretory and a plasma membrane protein, (2) two secretory proteins, or (3) two plasma membrane proteins (Figure 1(c)).

The regulatory peptide cholecystokinin is a classical example where tyrosine sulfation of a secretory protein dramatically promotes its interaction with a plasma membrane protein, that is, its cell surface receptor. Thus, sulfated cholecystokinin is 260 times more potent than its unsulfated form. Of the several examples of tyrosine sulfation promoting the interaction between two secretory proteins, the case of the binding of the tyrosine-sulfated blood coagulation factor VIII to von Willebrand factor is particularly intriguing, as it also documents the medical relevance of this posttranslational modification. Humans with a mutation in the critical tyrosine residue of factor VIII that is sulfated and involved in its binding to von Willebrand factor are afflicted with hemophilia A.

An example of tyrosine sulfation promoting the interaction between two plasma membrane proteins is the important role of this posttranslational modification for the high-affinity binding of leukocyte-associated P-selectin glycoprotein ligand (PSGL)-1 to P-selectin on activated endothelial cells. This crucial interaction initiates adhesion of leukocytes to the vascular wall during inflammation. Tyrosine sulfation also occurs in seven-transmembrane-segment chemokine receptors, for example, CCR5. Under physiological conditions, these plasma membrane proteins play a central role in chemokine signaling pathway through G proteins. Remarkably, human and simian immunodeficiency viruses use CCR5 as a co-receptor, together with CD4, to mediate their attachment to the host cell membrane. Specifically, sulfation of tyrosine residues in the CCR5 N-terminal domain has been shown to be critical for the interaction of this protein with human immunodeficiency virus (HIV) envelope glycoprotein gp120, leading to HIV infection. Thus, the design of tyrosine-sulfated peptide competitors – mimicking HIV gp120-binding sites – could turn out to be the basis for new therapeutic compounds that will block HIV

cellular entry. These examples highlight the medical relevance of protein tyrosine sulfation.

See also: **Cell Architecture and Function:** Golgi Complex; **Lipids Carbohydrates Membranes and Membrane Proteins:** Free Oligosaccharides: Formation and Degradation (Free Oligosaccharides Related to N-Linked Glycans); Secretory Pathway.

Further Reading

- Beisswanger R, Corbeil D, Vannier C, et al. (1998) Existence of distinct tyrosylprotein sulfotransferase genes: Molecular characterization of tyrosylprotein sulfotransferase-2. *Proceedings of the National Academy of Sciences of the United States of America* 95: 11134–11139.
- Bettelheim FR (1954) Tyrosine-O-sulfate in a peptide from fibrinogen. *Journal of the American Chemical Society* 76: 2838–2839.
- Corbeil D, Thiele C, and Huttner WB (2005) Tyrosine O-sulfation. *Current Protocols in Protein Science* chapter 14; unit 14.7.1–14.7.14.
- Huttner WB (1982) Sulphation of tyrosine residues – a widespread modification of proteins. *Nature (London)* 299: 273–276.
- Huttner WB (1984) Determination and occurrence of tyrosine O-sulfate in proteins. *Methods in Enzymology* 107: 200–223.
- Huttner WB and Baeuerle PA (1988) Protein sulfation on tyrosine. *Modern Cell Biology* 6: 97–140.
- Huttner WB, Niehrs C, and Vannier C (1991) Bind or bleed. *Current Biology* 1: 309–310.
- Kehoe JW and Bertozzi CR (2000) Tyrosine sulfation: A modulator of extracellular protein–protein interactions. *Chemical Biology* 7: R57–R61.
- Kehoe JW, Maly DJ, Verdugo DE, et al. (2002) Tyrosylprotein sulfotransferase inhibitors generated by combinatorial target-guided ligand assembly. *Bioorganic and Medicinal Chemistry Letters* 12: 329–332.
- Komori R, Amano Y, Ogawa-Ohnishi M, and Matsubayashi Y (2009) Identification of tyrosylprotein sulfotransferase in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 106: 15067–15072.
- Marcello MR, Jia W, Leary JA, Moore KL, and Evans JP (2011) Lack of tyrosylprotein sulfotransferase-2 activity results in altered sperm-egg interactions and loss of ADAM3 and ADAM6 in epididymal sperm. *Journal of Biological Chemistry* 286: 13060–13070.
- Moore KL (2003) The biology and enzymology of protein tyrosine O-sulfation. *Journal of Biological Chemistry* 278: 24243–24246.
- Moore KL (2009) Protein tyrosine sulfation: A critical posttranslation modification in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America* 106: 14741–14742.
- Niehrs C, Beisswanger R, and Huttner WB (1994) Protein tyrosine sulfation, 1993 – an update. *Chemico-Biological Interactions* 92: 257–271.
- Ouyang Y-B, Crawley JTB, Aston CE, and Moore KL (2002) Reduced body weight and increased postimplantation fetal death in tyrosylprotein sulfotransferase-1-deficient mice. *Journal of Biological Chemistry* 277: 23781–23787.
- Sherry DM, Murray AR, Kanan Y, et al. (2010) Lack of protein-tyrosine sulfation disrupts photoreceptor outer segment morphogenesis, retinal function and retinal anatomy. *European Journal of Neuroscience* 32: 1461–1472.
- Stone MJ, Chuang S, Hou X, Shoham M, and Zhu JZ (2009) Tyrosine sulfation: An increasingly recognised post-translational modification of secreted proteins. *New Biotechnology* 25: 299–317.

Ubiquitin-Like Proteins

L-S Lu, Texas Heart Institute at St. Luke's Episcopal Hospital, Houston, TX, USA

E T H Yeh, The University of Texas MD Anderson Cancer Center and Texas Heart Institute at St. Luke's Episcopal Hospital, Houston, TX, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Edward T.H. Yeh, volume 4, pp. 304–307, © 2004, Elsevier Inc.

Glossary

De-UBLylation Ubiquitin-like protein (UBL) proteases-mediated removal of conjugated UBL from substrate.

E1, E2, E3 Components of UBL conjugation machinery, with nomenclature derived from ubiquitin conjugation pathway. E1 activates UBL; E2 transfers the activated UBL either directly to the substrate or to an E3, which facilitates covalent bond formation between the terminal diglycine motif and acceptor cysteine of the substrate.

UBL-binding motif A protein motif that noncovalently interacts with conjugated UBL. Such interaction confers regulatory function and thus proteins with UBL-binding motif usually act as effectors of UBLylation pathway.

UBLylation The process of protein covalent modification by a specific UBL. For example, SUMOylation and NEDDylation refers to small ubiquitin-like modifier (SUMO)-mediated and neural precursor cell expressed, developmentally down regulated 8 (NEDD8)-mediated covalent modification respectively.

Introduction

Ubiquitin is a posttranslational modifier. The elaborate biochemical machinery for ubiquitin conjugation has its root in various metabolite/cofactor biosynthesis pathways in prokaryotes and is further diversified to attach ubiquitin-like proteins (UBLs) to their targets. Since the identification of interferon-induced 15-kDa protein (Isg15) as the first UBL in 1992, now there are 11 proteins (small ubiquitin-like modifier (SUMO)1, SUMO2, SUMO3, SUMO4, neural precursor cell expressed, developmentally down regulated 8 (NEDD8), Isg15, Atg8, Atg12, Fat10, Ufm1, and Urm1) in the eukaryotic UBL family. These proteins are defined not by sequence similarity but instead by the presence of a conserved β -grasp fold in their structures. A β -grasp fold is an ancient structure that can be found in *Escherichia Coli* cofactor synthesis proteins MoaD and ThiS. It consists of a stranded beta-sheet, an alpha-helix, and a C-terminally extruded tail with a di-glycine motif, which is crucial for covalent conjugation to substrates. A general scheme for UBL modification involves activation, conjugation, ligation, and deconjugation (see [Figure 1](#)). Some UBLs are capable of forming polymeric branching chains just like ubiquitin. Monomeric and polymeric chains of UBL add an extensive interacting surface to the target molecule and therefore are capable of regulating enzymatic activity, protein complex formation, subcellular distribution, and protein stability.

Experimentally verified UBL targets are involved in diverse cellular processes that can be found in every facet of cell biology. With the multitude of UBL signals, a specialized set of protein domains co-evolves with UBLs to specifically recognize UBL-modified proteins and ensures the UBL codes are correctly interpreted. Proteins with such UBL-binding domains may bind to monomeric/polymeric UBLs with various affinities and, therefore, render UBL-mediated protein complex formation highly flexible in response to different biological cues. Intriguingly, UBLs not only exist as standalone proteins but are also found as structural domains in more complex polypeptides. They act like built-in UBL modification to mediate interaction with UBL-binding domains and generally are not subject to enzymatic processing and cleavage.

Activation, Conjugation, Ligation, and Deconjugation

The dynamics of UBL attachment is regulated by four groups of enzymes: E1, the activating enzyme; E2, the conjugation enzyme; E3, the ligation enzyme; and UBL proteases (see [Table 1](#)). Most UBLs are synthesized as inactive precursors. They first have to be processed by specific UBL proteases or hydrolases to remove C-terminal tail and expose the di-glycine motif. E1 will adenylate the exposed glycine at the cost of adenosine triphosphate (ATP) hydrolysis. This is the only

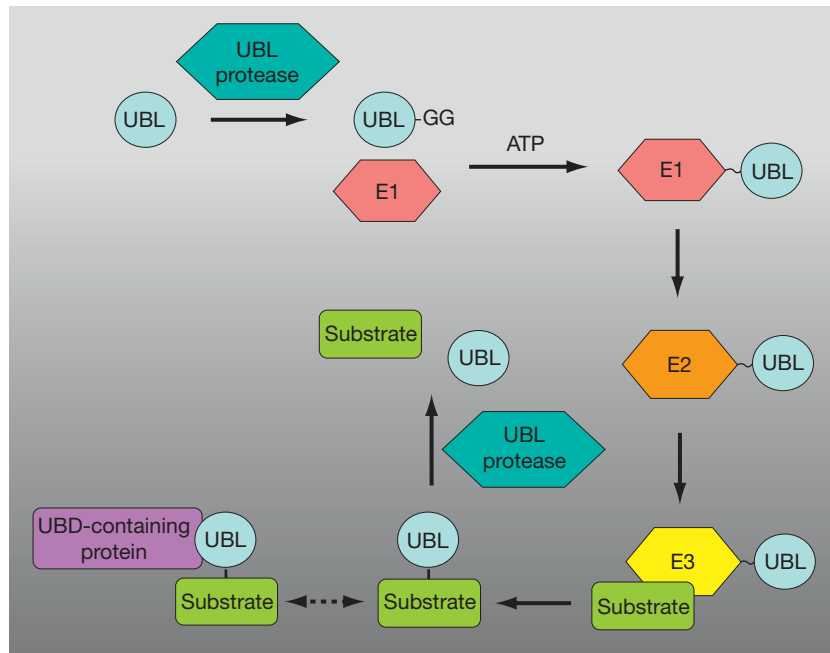


Figure 1 A general scheme for protein UBLylation dynamics.

Table 1 An overview of the UBL conjugation/deconjugation machinery

UBL	E1	E2	E3	UBL proteases
SUMO1 SUMO2 SUMO3	Aos1/Uba2	Ubc9	PIAS1, PIAS2a, PIAS2b, PIAS3, PIAS4, Pc2, RanBP2, MMS21, TOPORS	SEN1, SEN2, SEN3, SEN5, SEN6, SEN7
NEDD8	Uba3/Nae1	Ubc12, Ube2F	Rbx1, Dcn1, Mdm2, Cbl	Den1, CSN, Usp21, ataxin-3, UCH-L3, UCH-L1
ISG15	Uba7	UbcH8	Herc5, Efp	Ubp43
ATG12	Atg7	Atg10		
ATG8	Atg7	Atg3	Atg12/Atg5	Atg4
FAT10	Uba6			
UFM1	Uba5	Ufc1	Ufl1	UfSP1, UfSP2
URM1	Uba4			

step that consumes ATP in the UBL modification reaction. The UBL~adenosine monophosphate (AMP) intermediate is then attacked by a cysteine in the E1 catalytic core to form a high-energy thioester bond with the UBL. Following that, E1~UBL binds to an E2 through an ubiquitin-fold domain on E1 and transfers UBL to the catalytic cysteine of E2. While E2~UBL in many cases is sufficient to relay UBL to the substrate, the ligation is always facilitated by an E3 which brings in the substrate, binds to E2~UBL intermediate, and catalyzes the isopeptide bond formation between the carboxyl group of terminal glycine on UBL and the ϵ -amino group of acceptor lysine on the substrate. Similar to ubiquitin E3, really interesting new gene (RING) domain or homologous to the E6-associated protein carboxyl terminus (HECT) domain in many UBL E3s are the major effectors for ligating UBL.

The transfer of UBL between loaded E1, E2, and E3 is driven by the release of inorganic pyrophosphate and AMP. In its free status, E1 has low affinity for E2. The binding to E2 only takes place with E1~UBL intermediate, which allosterically inhibits E2 binding to E3. On the other hand, free E3 has a high

affinity only for loaded E2 but not for free one. The differential affinity between free/UBL-loaded enzymes and its downstream cognates account for the unidirectional transfer of UBL between E1, E2, and E3.

To deconjugate UBL from its substrate, a UBL protease will bind the UBL-modified substrate and catalytic cysteine of the protease will attack the isopeptide bond to cleave UBL from substrate and form a protease~UBL intermediate. At the end, hydrolysis of the thioester bond will release UBL from the protease. Most UBL-modified substrates only account for a small portion of the steady-state pool, and it is generally accepted that UBL proteases are a crucial factor in determining steady-state level of UBL-modified substrate.

Small Ubiquitin-Like Modifier

SUMOs are the best characterized UBL. Since its discovery in 1996, hundreds of proteins have been experimentally identified to be modified by SUMO. SUMO genes are highly

conserved among yeasts, worms, flies, and vertebrates. Their homologs are also found in plants and archae. There are four SUMO genes in the mammalian genome. SUMO1 is a nuclear protein with 101 amino acids and shares 18% sequence similarity to ubiquitin. SUMO2 and SUMO3, which are 97% identical before their C-terminal di-glycine motif, are also nuclear proteins and only share 50% sequence similarity with SUMO1. While SUMO1, SUMO2, and SUMO3 are ubiquitously expressed, SUMO4 is preferentially expressed in immune cells and pancreatic islets. Although genetic evidence suggests that SUMO4 M55V polymorphism confers susceptibility to type I diabetes, it remains controversial whether or not SUMO4 can be processed into mature UBL and conjugate target proteins. SUMO1, SUMO2, and SUMO3 attach to their targets as a monomer. In addition, SUMO2 and SUMO3 are found to form branching chains just like ubiquitin through Lys11 linkage. SUMO1 also forms polymeric chain via Lys7, Lys16, and Lys17 *in vitro*. However, there is no concrete evidence for polymeric SUMO1 chains *in vivo*. Instead, SUMO1 may conjugate to polymeric SUMO2/3 chain and act as a cap to stop its extension.

The E1 for SUMO conjugation is a heterodimer composed of Aos1 and Uba2. Endogenous regulation of SUMO E1 is not yet understood. However, an avian adenovirus protein Gam1 targets Aos1 for ubiquitination and degradation through its SOCS domain via Cul2/5-Elongin B/C-Roc1 ubiquitin E3 ligase complex. As a consequence, Uba2 also degrades and the loss of functional E1 leads to impaired global SUMOylation. Ubc9 is the only E2 for SUMO. Sequence analysis of the first few SUMO substrates has revealed a common SUMO acceptor motif ψ -K-X-D/E, where ψ stands for aromatic amino acids and X stands for any amino acid. Over a decade, the rule is confirmed in many other molecules and is the only established acceptor consensus among UBLs. However, not all lysines in this motif are SUMO acceptors, and in many cases SUMO conjugates to a cryptic lysine that cannot be predicted based on protein-sequence analysis.

The E3 for SUMO features the presence of RING-type plant homeo domain (PHD) finger. The largest collection of SUMO E3 is the protein inhibitors of activated STAT (signal transducers and activators of transcription) (PIAS) family members, in addition to Pc2, RanBP2, Mms21, Rhes, and Topors. Interestingly, the PHD finger of Trim25 can function as an intramolecular SUMO E3. Although Aos1/Uba2 and Ubc9 are sufficient for SUMOylation, SUMO E3s facilitate the reaction, result in functional compartmentalization, and usually are the targets for regulating SUMOylation balance.

A functional SUMO pathway is essential for life. Unbiased analysis of heat-shocked cells suggests that SUMO regulates transcriptional repression, chromatin regulation, DNA damage response, lipid metabolism, and cell-cycle progression. How does SUMO achieve such versatility? The first clue comes from the long list of SUMO targets. These proteins tend to cluster in functionally relevant categories, and therefore imply the crucial regulatory role of SUMO in those biological processes. In addition to its far-reaching footprints, molecular consequences of protein SUMOylation also diversify. SUMOylation may render stronger interaction with DNA and thus promote chromatin binding. SUMOylation may alter the components of the transcriptional complex and switch on/off transcription by

differential recruitment of transcriptional cofactors. SUMOylation also interacts with other posttranslational modifications. For example, nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha (I κ B α) can be modified by SUMO at Lys21, which is also an ubiquitin acceptor. Lys21 SUMOylation is sufficient to inhibit tumor necrosis factor- α (TNF α)-induced ubiquitination and degradation of I κ B α , and thus blocks subsequent nuclear factor kappa B (NF κ B) activation. Another example comes from the SUMOylation of core histones H2B, H3, and H4 in yeast. SUMOylation of H2B and H4 reversely correlate to acetylation at galactokinase (GAL1) promoters during altered carbon supply. Meanwhile both ubc9 and Δ siz1 Δ siz2 mutants, which cannot undergo normal SUMOylation due to the lack of either E2 or E3, display enhanced acetylation of gene-associated H3. On the contrary, SUMOylation can be attenuated by phosphorylation. In the case of Elk1, mitogen-activated protein (MAP) kinase-mediated phosphorylation of Ser383 blocks SUMOylation, which is required for its maximal transcriptional repression.

In addition to covalent modification, SUMOs are also capable of noncovalent interaction with SUMO interaction motif (SIM). SIM, with a consensus of (V/I/L)-(V/I/L)-X-(V/I/L) or (V/I/L)-X-(V/I/L)-(V/I/L) and usually a flanking acidic residue, binds to the second β -strand and the first α -helix of SUMO. The terminal di-glycine of SUMO is not involved in SUMO-SIM interaction. Such interaction is one to two orders stronger than that between ubiquitin and ubiquitin-binding domains. SIM-containing protein binds and exerts its function depending on the SUMOylation status of its partner, and in this sense the SUMO-SIM pair provides a specific decoding mechanism of protein SUMOylation marks. A good example is RING finger protein 4 (RNF4), a SUMO-targeted ubiquitin ligase. It has four SIMs and one RING-type PHD finger which catalyzes ubiquitin conjugation. RNF4 only binds promyelocytic leukemia protein (PML), a crucial protein in the pathogenesis of acute promyelocytic leukemia, after arsenic-induced SUMOylation. As a consequence, PML is ubiquitinated by RNF4 on its poly-SUMO tags and degrades in a proteasome-dependent manner. This is believed to be one of the therapeutic mechanisms of arsenic in treating acute promyelocytic leukemia.

Considering the large number of SUMO targets, it is surprising that six isopeptidase – sentrin/SUMO specific proteases (SENP) – in the mammalian genome are sufficient for keeping proper SUMO status of hundreds of proteins. SENPs are cysteine peptidase and all of them contain a C48 protease domain in the C-terminal. The six SENPs can be divided into three subfamilies according to the substrate specificity, sequence homology, and cellular localization. SENP1 and SENP2 are mammalian ortholog of the yeast Ulp1 and both readily deconjugate SUMO1 and SUMO2. SENP3 and SENP5 are nucleolar proteins with a preference for deconjugating SUMO2/3. SENP6 and SENP7 have a split C48 protease domain in the C-terminal. While the deSUMOylating activities of these two enzymes are controversial, they are capable of editing polymeric SUMO chains. Early embryonic lethality of mice deficient of either SENP1 or SENP2 suggests that these SENPs perform nonredundant functions. Accessibility of SENP to its substrate may dictate paralog-selective SUMOylation in cells. For example, SUMO1-conjugated RanGAP1 forms a more stable complex with nucleoporin RanBP2 compared to its SUMO2 conjugates and this

prevents binding of the yeast SENP Ulp1. As a result, RanGAP1 is preferentially SUMOylated by SUMO1 *in vivo*.

Neural Precursor Cell Expressed, Developmentally Downregulated 8

NEDD8, also known as related to ubiquitin 1, or Rub1, in budding yeast is the closest kin of ubiquitin in the UBL family (60% sequence identity). The 81-amino-acid protein was identified in 1992 with a subtraction cloning approach from mouse-neural precursor cell complementary DNA (cDNA) library and its function as a UBL was first characterized in 1997 in our laboratory. It is ubiquitously expressed, evolutionarily conserved, and is essential for cell viability. NEDD8 is a crucial regulator of ubiquitin E3 ligase complexes. Dysregulated NEDDylation has been observed in tumors and neurodegenerative diseases, implying its correlation with the onset and progression of human diseases.

NEDD8 is synthesized as inactive precursor with a stretch of four glycines in the C-terminus. NEDD8 protease SENP8 (Den1) and hydrolase UchL3 are both capable of exposing the Gly76 for conjugation. The E1 for NEDD8, like the SUMO E1, is a heterodimer composed of Uba3 and Nae1. Despite the similarity, NEDD8 conjugating machinery is devoted for processing NEDD8 but not ubiquitin. Several mechanisms account for such specificity. First of all, Uba3/Nae1 is capable of discriminating NEDD8 from ubiquitin due to a strategically positioned basic residue in the UBL binding pocket of Uba3, which repels Arg72 in ubiquitin but not Ala72 in NEDD8. Second, the NEDD8 E2 Ubc12 has a unique N-terminal sequence that is required for Uba3/Nae1 binding, which prevents mischarging of ubiquitin-specific E2s to NEDD8 E1. In addition, Ubc12 has a smaller cysteine-binding pocket and a unique groove which accommodates N-terminal extrusion of NEDD8. These structural features allow Ubc12 to discriminate NEDD8 from ubiquitin. Recently, Ube2F is found to be the second NEDD8 E2. Ubc12 and Ube2F form a complex with different E3 (Rbx1 and Rbx2, respectively), neddylation distinct substrates, and result in selective ubiquitin E3 ligase complexes activation.

Rbx1, a common component of Skp1-Cul1-F-box-protein (SCF) and Elongin B/C, Cullen-2, VHL-containing (ECV) ubiquitin E3 ligase complexes, is the first reported NEDD8 E3. It harbors a RING-finger domain which not only approximates ubiquitin E2 to the substrate but also promotes cullin neddylation *in vitro*. The second NEDD8 E3, Dcn1, contains a potentiating neddylation (PONY) domain that is required for neddylation activity *in vitro* and *in vivo*. Dcn1 may act as a scaffold to facilitate interaction between substrate and NEDD8-charged Ubc12. It is interesting that *in vitro* reconstitution of NEDDylation requires the presence of both Dcn1 and Rbx1, which indicates potential synergy between the two molecules. In addition to Dcn1 and Rbx1, ubiquitin E3 ligases Mdm2 and Cbl are also reported to promote NEDDylation in an E3-like fashion.

All of the cullin family members are NEDD8 substrates. Cullins are structural scaffold for Cullin-RING ubiquitin ligase, and neddylation of cullins are essential to bring in E2 to promote the ubiquitination and degradation of their cognate

target cyclins, which are crucial regulators for cell cycle progression. Human Cul-1 is a major component of SCF ubiquitin E3 ligase that catalyzes the ubiquitination of I κ B α , β -catenin, and p27^{Kip1}; NEDDylation of Cul-1 is essential for the ubiquitin ligase activity. Molecular consequences of NEDDylation may involve changes in protein conformation. In the case of SCF, NEDDylation of Cul-1 exposes the RING finger of Rbx1 and therefore facilitates ubiquitin ligation to the proximal substrate. In another case, NEDDylation of cullins allosterically inhibits the binding of Cdn1, which is an inhibitor of cullin E3 complex, and activates SCF ubiquitin ligase. NEDDylation can also recruit new binding partners. For example, epidermal growth factor receptor (EGFR) is a noncullin NEDD8 substrate. It can be neddylation by Ubc12-loaded Cbl at the cytoplasmic domain upon ligand binding, and then recruits Cbl precomplexed with ubiquitin E2 UbcH7 to EGFR. Subsequent ubiquitylation and degradation of EGFR ensures rapid receptor desensitization. Additional noncullin NEDD8 substrates include p53, p73, Mdm2, VHL, L11, and other ribosomal proteins. NEDD8 has been reported to form polymeric chains *in vitro*. The significance of poly-NEDD8 chain *in vivo* remains unknown.

COP9 signalosome (CSN) is the best-characterized NEDD8 isopeptidase. CSN is a protein complex that interacts with Cul-1-containing SCF-type E3 ubiquitin ligases and promotes the cleavage of NEDD8 from neddylation Cul-1. The Jab1/Mpr/Pad1 N-terminal (MPN) domain metalloenzyme motif in Csn5 subunit is responsible for the de-neddylation activity of CSN. Thus, CSN binds to SCF-type E3 ubiquitin ligases and probably regulates their activity through the de-neddylation of conjugated Cul-1. SENP8/Den1, a cysteine protease in the SENP family, is another NEDD8-specific isopeptidase. Unlike CSN, SENP8/Den1 is weak in de-neddylation cullins. SENP8/Den1 deletion results in accumulation of NEDD8 conjugates; removing one allele of SENP8/Den1 genetically counteracts the developmental lethality in NEDD8 hypomorphic mutants. Therefore, SENP8/Den1 may act as the major de-neddylase for noncullin NEDD8 substrates. In addition to CSN and Den1, NEDD8 conjugates can also be cleaved by ubiquitin-NEDD8 dual specific proteases such as Usp21, ataxin-3, PfUch54, Uch-L3, and Uch-L1.

Regulation of NEDD8 pathway is more complicated than simply keeping balanced conjugation/deconjugation. Nub1 (NEDD8 ultimate buster 1), a 66-kDa NEDD8-interacting protein with a N-terminal ubiquitin-like domain, has been shown to direct NEDD8 and its conjugates to 19S subunit of proteasome and results in their degradation. NEDD8 pathway can also be manipulated pharmacologically. MLN4924 is a suicide inhibitor for NEDD8 E1 Nae1/Uba3 heterodimer and suppresses tumor cell growth. It is currently under phase II clinical trial for treating acute myelocytic leukemia and may become the first UBL-targeting compound for human cancer treatment.

Interferon-Induced 15-kDa Protein

Isg15 was first identified as a 15-kDa protein induced by interferon treatment in 1979. The gene actually encodes a 17-kDa protein with two tandem ubiquitin-like domains and was found to form high molecular weight conjugates in 1992.

Unlike SUMO and NEDD8, Isg15 is only present in vertebrates and these homologs share limited sequence similarity. Isg15 protein is detected in lymphoid tissues, striated and smooth muscles, receptive uterine endometrium, epithelial cells, and neurons. In mammals, Isg15 is an important regulator of antiviral defense. Despite the lack of a classical signal peptide and a membrane receptor, Isg15 is secreted and may act in an autocrine/paracrine fashion to modulate interferon production from CD3+ lymphocytes and subsequent NK cell proliferation and activation. While Isg15 knockout mice are viable and fertile, they are more vulnerable to viral infections. Interestingly, the antiviral property seems to be dependent on its UBL property rather than its cytokine property since reconstitution of nonconjugatable Isg15 cannot rescue the virus susceptibility.

Uba7 is the only E1 dedicated to Isg15 activation. This protein is a monomeric E1. It has been reported that influenza virus protein NS1B specifically interacts with Uba7 and inhibits its E1 activity. The cognate E2 for Uba7 is UbcH8. There are only a few E3 that have been identified as Isg15 E3 including Herc5 and Efp. The fact that Isg15, Uba7, and UbcH8 are all upregulated after interferon treatment may explain enhanced protein ISGylation after type I interferon (IFN- α and IFN- β) stimulation. It is noteworthy that some cell lines (such as K562 and 293T cells) do not respond to interferons with detectable protein ISGylation. In the case of K562 cells, it is ascribed to the defective expression of Uba7. So far Ubp43 is the only one Isg15-specific protease that has been identified. An *in vitro* screen of ubiquitin-specific proteases has identified Usp13 as a dual-specificity protease that is capable of processing ubiquitin and Isg15. However, *in vivo* validation of this finding is still pending.

Experimentally verified Isg15 targets include Stat1, Jak1, Erk1, PKR, Mx1, RIG-1, Ubc13, and influenza protein NS1A. ISGylation of these proteins do not lead to proteosomal degradation nor does it lead to altered subcellular localization. ISGylation of an ubiquitin E2 Ubc13 blocks its ability to form thioester bond with ubiquitin and therefore abolishes subsequent Lys63-linked polyubiquitination. Enhanced ISGylation of Jak1 and Stat1 in Ubp43-deficient bone marrow cells are associated with persistent Stat1 phosphorylation and enhanced expression of interferon-stimulated genes. As a result, bone marrow cells die soon after interferon treatment. However, whether ISGylation of Jak1/Stat1 accounts for the phenotype observed in Ubp43-deficient mice was challenged by the observation that compound deletion of Isg15/Ubp43 does not rescue the Ubp43 deletion phenotypes.

Atg8 and Atg12

Autophagy is a tightly regulated catabolic process that turns over proteins and organelles with lysosomal degradation machinery in a eukaryote cell. There are approximately 30 proteins for the autophagy pathway in yeast. Among them are two UBLs, Atg8 and Atg12, and their conjugation system components. Both UBL pathways are required for efficient autophagy as deletion of any component leads to autophagy defect. Atg8 and Atg12 share little sequence similarity. However, structurally they both have ubiquitin-like folds. Both

Atg12 and Atg8 conjugation systems are conserved in mammalian cells and they act in a similar manner.

Atg12 is synthesized as mature UBL and does not require to be processed prior to E1 activation. The E1 for Atg12 is Atg7, which forms a thioester bond between its Cys507 and Gly186 of Atg12. Atg10 is the only E2 for Atg12 conjugation pathway. The transfer of Atg12 from Atg10 to its sole substrate Atg5 does not need an E3 and is constitutively active. Atg12–Atg5 conjugate localizes to the outer membrane of and promotes the expansion of autophagosome. Atg12 conjugation seems to be irreversible since there is no isopeptidase to cleave Atg12 from Atg5. Atg12-modified Atg5 interacts with coiled coil protein Atg16 and forms multimeric complex to direct Atg8-PE to tightly associate with the expanding autophagosome membrane.

The other UBL, Atg8, modifies phosphatidylethanolamine (PE, a lipid moiety on autophagosome membrane). It is synthesized as a precursor as most UBLs do. The protease Atg4 cleaves and exposes Gly116 of Atg8 so that it can conjugate to Cys 507 of Atg7 (the same E1 and the same cysteine residue as used in Atg12 conjugation pathway) for activation. Atg7 is a noncanonical E1 because, unlike the ubiquitin and SUMO E1, it forms a homodimer and the resulting homodimeric adenylation domains are likely to bind two UBL molecules simultaneously. This is more close to the prokaryotic E1 antecedent MoeB and ThiF. Atg7 also stands out among UBL E1s due to its capability to activate two unrelated UBL, Atg12 and Atg8, and specifically transfer these UBL to their cognate E2. The structural basis for such duality is not known. Atg8 is then transferred to the E2 Atg3 and then forms an amide bond between PE and Atg8. Atg12 interacts with Atg3 and such interaction allows Atg12–Atg5 complex to promote Atg8 modification of PE. Although Atg12–Atg5, in this sense, acts like an E3 but the complex lacks HECT or RING domains which is the hallmark of most UBL E3. Atg8–PE is the structural determinant of autophagosome membrane expansion, and quantitative fluorimetry analysis suggests a positive correlation between Atg8 abundance and the size of autophagosome. It is also capable of joining the ends of expanding lipid bilayer and finalizes autophagosome formation. On completion of autophagosome formation, Atg8-PE will be cleaved by the protease Atg4. Atg4-mediated recycling of Atg8 is crucial for efficient autophagy since reintroduction of processed Atg8 into Δ Atg4 mutant is capable of binding to PE, but cannot rescue hampered autophagosome formation upon nutrient depletion.

Fat10

Fat10, also known as diubiquitin or ubiquitin D, is originally identified as a UBL-encoding gene in the human leukocyte antigen F (HLA-F) locus. The molecule consists of two ubiquitin-like domains in tandem arrangement and bears a conserved di-glycine motif at its C-terminus. It is expressed in mature B cells, dendritic cells, B lymphoblasts, thymus, tonsil, pancreatic islets, mammary gland epithelia, and small intestine. Fat10 is capable of covalent conjugation as all UBLs do, but the conjugation machinery is poorly characterized. Fat10 is synthesized in mature form and forms thioester bond with Uba6, which is a monomeric E1 and also surprisingly activates ubiquitin. Uba6 does not load Fat10 onto its cognate ubiquitin E2

including Ubc3, Ubc5, Ubc13, and E2-25K. The structural basis for such specificity is unknown. The other members of the Fat10 pathway, including E2, E3, protease, and the substrate are yet to be identified.

Current knowledge on the function of Fat10 is limited. Like Isg15, Fat10 is an interferon-responsive gene and it can also be induced by tumor necrosis factor alpha. Strong upregulation of Fat10 has been observed in hepatocellular carcinoma, colorectal carcinoma, and HIV-infected renal tubular epithelial cells, which might be due to the proinflammatory microenvironment in these samples. Although Fat10 knockout mice are viable, fertile, and grossly normal, they are more vulnerable to endotoxin-induced sepsis. These evidences imply that Fat10 is involved in innate immunity.

The best-characterized molecular property of Fat10 is its role in the control of protein quantity and quality. Linearly fused Fat10 serves as an ubiquitin-independent degron (the amino acid sequence that tags the protein for degradation) to promote protein degradation at a comparable kinetics to their ubiquitin-fused counterparts. Fat10-dependent degradation is not affected in cells defective in ubiquitin activation, or in cells expressing lysine-mutated Fat10 which prevents ubiquitin conjugation. Noncovalent interaction between Fat10 and NUB1L, which associates with the 26S proteasome, is required for Fat10-fused protein to be processed by proteasome

in vitro and greatly facilitates the degradation of Fat10 monomer and its conjugates *in vivo*. It is interesting to note that both Fat10 and NUB1L are interferon-responsive genes and overexpression of Fat10 but not its conjugation-defective mutant leads to apoptosis. Fat10 also promotes aggresome formation in case of proteasome failure through its interaction with histone deacetylase (HDAC6), which is involved in the transportation of ubiquitylated misfolded protein to aggresome for sequestration and removal. Other functions of Fat10 include its roles in regulating mitosis through its interaction with mitotic spindle assembly checkpoint protein (MAD2) and mediating NF- κ B signal transduction.

See also: [Protein/Enzyme Structure Function and Degradation: Ubiquitin System.](#)

Further Reading

- Ohsumi Y (2001) Molecular dissection of autophagy: Two ubiquitin-like systems. *Nature Reviews Molecular Cell Biology* 2: 211–216.
- Ritchie KJ and Zhang DE (2004) Isg15: The immunological kin of ubiquitin. *Seminars in Cell and Developmental Biology* 15: 237–246.
- Yeh ETH (2009) SUMOylation and De-SUMOylation: Wrestling with life's processes. *Journal of Biological Chemistry* 284: 8223–8227.

Ubiquitin System

A Ciechanover and M H Glickman, Israel Institute of Technology, Haifa, Israel

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 4, pp. 299–303, © 2004, Elsevier Inc.

Glossary

Proteolysis/degradation Hydrolysis of a protein which is a heteropolymer of amino acids to short peptides, which contain a few amino acids, and/or to single amino acids.

Ubiquitination Covalent modification of a protein by the small protein ubiquitin.

In the ubiquitin system, a target substrate is modified by ubiquitin or a ubiquitin-like protein. In most cases, proteins are modified by multiple moieties of ubiquitin, generating a polyubiquitin chain. This modification leads to their degradation by the 26S proteasome complex. Modification by a single moiety of ubiquitin can target proteins for degradation in the lysosome/vacuole. Modification by ubiquitin-like proteins serves nonproteolytic functions. Conjugation of ubiquitin is carried out by a modular cascade of enzymes, specific to each substrate. Ubiquitination of cellular proteins is a highly complex, temporally controlled, and tightly regulated process that targets in a specific manner thousands of cellular proteins and plays major roles in a variety of basic pathways, such as cell division, differentiation, and quality control. Not surprisingly, aberrations in the system underlie the pathogenesis of many diseases, certain malignancies, and neurodegenerative disorders. Mechanism-based drugs are currently being developed.

Mechanisms of Ubiquitination and Degradation

Ubiquitination

Degradation of a protein via the ubiquitin–proteasome pathway involves two discrete and successive steps: (1) tagging of the substrate by covalent attachment of multiple ubiquitin molecules; and (2) degradation of the tagged protein by the 26S proteasome complex with release of free and reusable ubiquitin. This last process is mediated by ubiquitin-recycling isopeptidases (deubiquitinating enzymes). Conjugation of ubiquitin, a highly evolutionarily conserved 76-residue polypeptide, to the protein substrate proceeds via a three-step cascade mechanism. Initially, the ubiquitin-activating enzyme, E1, activates ubiquitin in an adenosine triphosphate-requiring reaction to generate a high-energy thiol ester intermediate, E1-S ~ ubiquitin. One of several E2 enzymes (ubiquitin-carrier proteins or ubiquitin-conjugating enzymes) transfers the activated ubiquitin from E1, via an additional high-energy thiol ester intermediate, E2-S ~ ubiquitin, to the substrate that is specifically bound to a member of the ubiquitin-protein ligase family, E3. There are a number of different classes of E3 enzymes. For the homologous to the E6-AP C-terminus (HECT) domain E3s, the ubiquitin is transferred once again from the E2 enzyme to an active site Cys residue on the E3, to generate a third high-energy thiol ester intermediate,

ubiquitin-S ~ E3, prior to its transfer to the ligase-bound substrate. RING finger-containing E3s catalyze direct transfer of the activated ubiquitin moiety to the E3-bound substrate.

E3s catalyze the last step in the conjugation process: covalent attachment of ubiquitin to the substrate. The ubiquitin molecule is generally transferred to an ϵ -NH₂ group of an internal lysine residue in the substrate to generate a covalent isopeptide bond. In some cases, however, ubiquitin is conjugated to the N-terminal amino group of the substrate. By successively adding activated ubiquitin moieties to internal lysine residues on the previously conjugated ubiquitin molecule, a polyubiquitin chain is synthesized. The chain is recognized by the downstream 26S proteasome complex. Thus, E3s play a key role in the ubiquitin-mediated proteolytic cascade since they serve as the specific recognition factors of the system. In certain cases the first ubiquitin moiety is conjugated to the substrate by one E3, while chain elongation is catalyzed by a different ligase often termed E4. Modification by a single moiety of ubiquitin is catalyzed via an identical mechanism and set of enzymes. The specific enzymes that catalyze modification by ubiquitin-like proteins are somewhat different, though they utilize a similar mechanism.

Degradation

Degradation of polyubiquitinated substrates is carried out by a large protease complex, the 26S proteasome that does not recognize nonmodified substrates. In one established case, that of the polyamine-synthesizing enzyme ornithine decarboxylase, the proteasome recognizes and degrades the substrate without prior ubiquitination. The proteasome is a multicatalytic protease that degrades polyubiquitinated proteins to short peptides. It is composed of two subcomplexes: a 20S core particle (CP) that carries the catalytic activity and a regulatory 19S regulatory particle (RP). The 20S CP is a barrel-shaped structure composed of four stacked rings, two identical outer α -rings and two identical inner β -rings. The eukaryotic α - and β -rings are composed each of seven distinct subunits, giving the 20S complex the general structure of $\alpha_{1-7}\beta_{1-7}\beta_{1-7}\alpha_{1-7}$. The catalytic sites are localized to some of the β -subunits. Each extremity of the 20S barrel can be capped by a 19S RP. One important function of the 19S RP is to recognize ubiquitinated proteins and other potential substrates of the proteasome.

A ubiquitin-binding subunit of the 19S RP has indeed been identified; however, its importance and mode of action have not been discerned. A second function of the 19S RP is to open an orifice in the α -ring that will allow entry of the substrate into the proteolytic chamber. Also, since a folded protein would not be able to fit through the narrow proteasomal channel, it is assumed that the 19S particle unfolds substrates and inserts them into the 20S CP. Both the channel-opening function and the unfolding of the substrate require metabolic energy, and indeed, the 19S RP contains six different adenosine triphosphatase subunits. Following degradation of the substrate, short peptides derived from the substrate are released, as well as reusable ubiquitin. Proteasomal degradation is not always complete. In some cases, the proteasome, rather than completely destroying its target, processes the ubiquitinated substrate precisely, releasing a truncated product. In the case of the nuclear factor kappa B (NF- κ B) transcriptional regulator, an active subunit (p50 or p52) is thus released from a longer inactive precursor (p105 or p100). For a general scheme of the ubiquitin system and its multiple functions, see Figure 1.

Substrate Recognition

A major unresolved question is: How does the system achieve its high specificity and selectivity? Why are certain proteins extremely stable in the cell, while others are extremely short-lived? Why are certain proteins degraded only at a particular time point during the cell cycle or only following specific

extracellular stimuli, yet they are stable under most other conditions? It appears that specificity of the ubiquitin system is determined by two distinct and unrelated groups of proteins: E3s and ancillary proteins. First, within the ubiquitin system, substrates must be specifically recognized by an appropriate E3 as a prerequisite to their ubiquitination. In most cases, however, substrates are not recognized in a constitutive manner, and in some cases they are not recognized directly by the E3. In some instances, the E3 must be switched on by undergoing posttranslational modification in order to yield an active form that recognizes the substrate. In many other cases, it is the substrate that must undergo a certain change that renders it susceptible for recognition. The stability of additional proteins depends on association with ancillary proteins such as molecular chaperones that act as recognition elements in *trans* and serve as a link to the appropriate ligase. Others, such as certain transcription factors, have to dissociate from the specific DNA sequence to which they bind in order to be recognized by the system. Stability of yet other proteins depends on oligomerization. Thus, in addition to the E3s themselves, modifying enzymes (such as kinases), ancillary proteins, or DNA sequences to which substrates bind also play an important role in the recognition process.

Functions and Substrates of the Ubiquitin System

Ubiquitin-mediated proteolysis of a variety of cellular proteins plays an important role in many basic cellular processes.

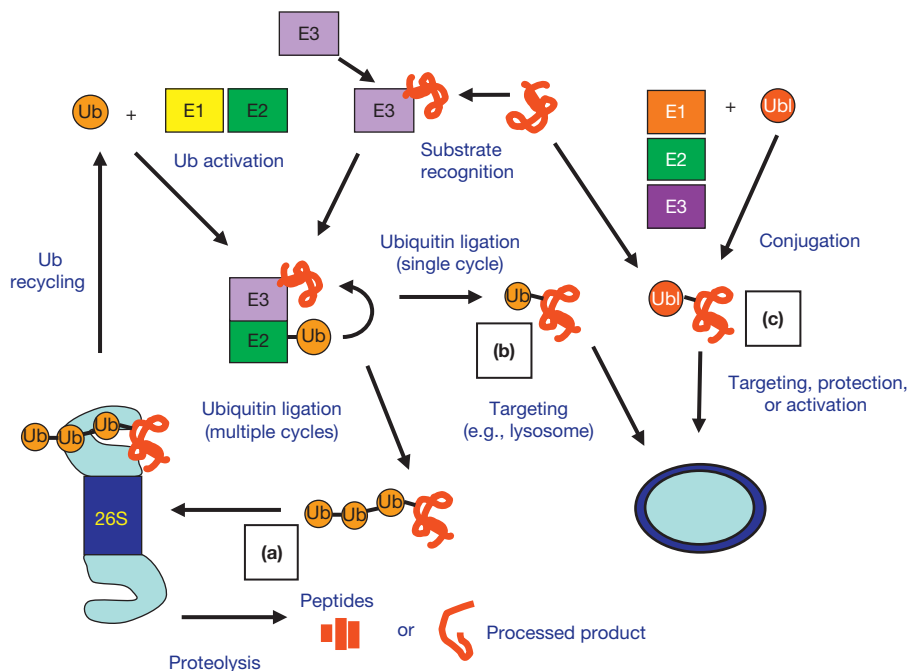


Figure 1 The ubiquitin system for protein modification. Ubiquitin (Ub) is activated by the three enzymes E1, E2, and E3, and the active ubiquitin moiety is transferred to the target substrate (red). Substrates can be either modified by a polyubiquitin chain (a) and targeted to the 26S proteasome complex for degradation (or processing), or modified by a single ubiquitin moiety, monoubiquitinated (b) and targeted to other organelles such as the lysosome (wherein it is degraded). Ubiquitin-like proteins (Ubl) are activated by E1, E2, and E3, which are similar but not identical to the enzymes that activate ubiquitin. Modification by Ubl (c) serves a variety of nonproteolytic functions such as routing of proteins to subcellular organelles (e.g., nucleus), protecting them from ubiquitination, or activating them.

Among these are regulation of cell cycle and division, differentiation and development, involvement in the cellular response to stress and extracellular effectors, morphogenesis of neuronal networks, modulation of cell-surface receptors, ion channels and the secretory pathway, DNA repair, transcriptional regulation, transcriptional silencing, long-term memory, circadian rhythms, regulation of the immune and inflammatory responses, and biogenesis of organelles. The list of cellular proteins that are targeted by ubiquitin is growing rapidly. Among them are cell-cycle regulators such as cyclins, cyclin-dependent kinase inhibitors, and proteins involved in sister chromatid separation, tumor suppressors, transcriptional activators, and their inhibitors. Cell-surface receptors and endoplasmic reticulum (ER) proteins are also targeted by the system. Finally, mutated and denatured/misfolded proteins are recognized specifically and removed efficiently. In this capacity, the system is a key player in the cellular quality control and defense mechanisms. The products of the proteasome can play an important role in the immune response. In the case of degradation of foreign proteins – such as those of viral origin – the resulting short peptides are presented by major histocompatibility complex class I molecules to the cytotoxic T cell that lyse the presenting cell.

Regulation of the Ubiquitin System

The ubiquitin–proteasome pathway can be regulated at the level of ubiquitination or at the level of proteasome activity. Since conjugation and proteasomal degradation is required for a multitude of cellular functions, regulation must be delicately and specifically tuned. In a few cases, general rather than specific, components of the pathway can be modulated by physiological signals. For example, upregulation of the pathway is observed during massive degradation of skeletal muscle proteins that occurs under normal fasting, but also under pathological conditions such as cancer cachexia, severe sepsis, metabolic acidosis, or following denervation. In most cases, however, regulation is specific and the target substrates are recognized by specific ligases that bind to defined motifs. The targeting motif can be a single amino acid residue (e.g., the N-terminal residue) or a sequence (the destruction box in cyclins) or a domain (such as a hydrophobic patch) that is not normally exposed. In other cases, the motif is a posttranslational modification such as phosphorylation that is generated in response to cell needs or external signals. Phosphorylation can occur either on the substrate or on the ligase.

Ubiquitin-Like Proteins

Both enzymes and substrates of the ubiquitin system have been found to be modified by ubiquitin-like proteins. Modification by ubiquitin-like proteins occurs only once. In the case of enzymes, modification affects their activity. For example, the modification by the ubiquitin-like protein NEDD8 enhances the activity of a certain class of E3 ligases. Conjugation of NEDD8 to one of the components of the ligase complex (Cullin) increases its affinity to other components of the conjugation machinery. In the case of substrates, modification can

affect their availability to the ubiquitination/degradation machinery and consequently cellular stability. For example, in the case of I κ B ζ , the inhibitor of the transcriptional regulator NF- κ B, modification by small ubiquitin-like modifier (SUMO)-1 was shown to protect the substrate from ubiquitination. In a completely different case, SUMOylation of RanGAP1 targets the protein to its final subcellular destination in the nuclear pore complex.

Ubiquitination and Pathogenesis of Human Diseases

Diseases

While inactivation of a major enzyme such as E1 is obviously lethal, mutations or acquired changes in enzymes or in recognition motifs in substrates that do not affect vital pathways or that affect the involved process only partially may result in a broad array of diseases. The pathological states associated with the ubiquitin system can be classified into two groups: those that result from (1) loss of function – mutation in a ubiquitin system enzyme or target substrate that results in stabilization of certain proteins and (2) gain of function – abnormal or accelerated degradation of the protein target.

Alterations in ubiquitination and deubiquitination reactions have been directly implicated in the etiology of many malignancies. In general, cancers can result from stabilization of oncoproteins, or destabilization of tumor-suppressor gene products. Some of the natural substrates for degradation by the proteasome are oncoproteins that, if not properly removed from the cell, can promote cancer. For instance, ubiquitin targets N-myc, c-myc, c-fos, c-jun, Src, and the adenovirus E1A proteins. Destabilization of tumor-suppressor proteins, such as p53 and p27, has also been implicated in the pathogenesis of malignancies.

In one fascinating case, that of uterine cervical carcinoma, the level of the tumor-suppressor protein p53 is extremely low. Most of these malignancies are caused by high-risk strains of the human papillomavirus (HPV). Detailed studies have shown that the suppressor is targeted for ubiquitin-mediated degradation by the virally encoded oncoprotein E6. Degradation is mediated by the native HECT domain E3 enzyme E6-AP, where E6 serves as an ancillary protein that allows recognition of p53 *in trans*. E6-AP will not recognize p53 in the absence of E6. E6 associates with both the ubiquitin ligase and the target substrate and brings them to the necessary proximity that is assumed to allow catalysis of conjugation to occur. Removal of the suppressor by the oncoprotein is probably an important mechanism used by the virus to transform cells.

Accumulation of ubiquitin conjugates and/or inclusion bodies associated with ubiquitin, proteasome, and certain disease-characteristic proteins have been reported in a broad array of chronic neurodegenerative diseases, such as the neurofibrillary tangles of Alzheimer's disease, brainstem Lewy bodies – the neuropathological hallmark in Parkinson's disease, and nuclear inclusions in CAG repeat expansion (poly-glutamine expansion) disorders such as Huntington's disease. However, in all these cases, a direct pathogenetic linkage to aberrations in the ubiquitin system has not been established. One factor that complicates the establishment of such linkage is the realization that many of these diseases, such as Alzheimer's and

Parkinson's, are not defined clinical entities, but rather syndromes with different etiologies. Accumulation of ubiquitin conjugates in Lewy inclusion bodies in many of these cases may be secondary, and reflects unsuccessful attempts by the ubiquitin and proteasomal machineries to remove damaged/abnormal proteins. While the initial hypothesis was that inclusion bodies are generated because of the inherent tendency of the abnormal proteins to associate with one another and aggregate, it is now thought that the process may be more complex and involves active cellular machineries, including inhibition of the ubiquitin system by the aggregated proteins. This aggregation of brain proteins into defined lesions is emerging as a common, but poorly understood mechanistic theme in many sporadic and hereditary neurodegenerative disorders.

The case of Parkinson's disease highlights the complexity of the involvement of the ubiquitin system in the pathogenesis of neurodegeneration. Aberrations in several proteins such as mutations in α -synuclein, an important neuronal protein, or in the deubiquitinating enzyme UCH-L1, have been described that link the ubiquitin system to the pathogenesis of the disease. One important player in the pathogenesis of Parkinson's disease is Parkin which is a RING-finger E3. Mutations in the gene appear to be responsible for the pathogenesis of autosomal recessive juvenile parkinsonism (AR-JP), one of the most common familial forms of Parkinson's disease. Parkin ubiquitinates and promotes the degradation of several substrates. It is possible that aberration in the degradation of one of these substrates that leads to its accumulation is neurotoxic and underlies the pathogenesis of AR-JP.

The cystic fibrosis (CF) gene encodes the CF transmembrane regulator (CFTR) that is a chloride channel. Only a small fraction of the protein matures to the cell surface, whereas most of it is degraded from the ER by the ubiquitin system prior to its maturation. One frequent mutation in the channel is $\Delta F508$. The mutation leads to an autosomal recessive inherited multisystem disorder characterized by chronic obstruction of airways and severe maldigestion due to exocrine pancreatic dysfunction. Despite normal ion channel function, CFTR $\Delta F508$ does not reach the cell surface at all, and is retained in the ER from which it is degraded. It is possible that the rapid and efficient degradation results in complete lack of

cell-surface expression of the $\Delta F508$ protein, and therefore contributes to the pathogenesis of the disease.

Drug Development

Because of the central role the ubiquitin system plays in such a broad array of basic cellular processes, development of drugs that modulate the activity of the system may be difficult. Inhibition of enzymes common to the entire pathway, such as the proteasome, may affect many processes nonspecifically, although a narrow window between beneficial effects and toxicity can be identified for a short-term treatment. Recent experimental evidence strongly suggests that such inhibitors may indeed be beneficial in certain pathologies, such as in multiple myeloma, a malignancy that affects the bone marrow.

A completely different approach to drug development can be the development of small molecules that interfere with the activity of specific E3s. For example, peptides that bind specifically to HPV-E6 can prevent its association with p53 and interfere with its targeting by E6-AP. Such peptides were able to induce p53 in HPV-transformed cells with subsequent reversal of certain malignant characteristics and induction of apoptosis.

See also: [Cell Architecture and Function: 26S Proteasome: Structure and Function](#); [Protein/Enzyme Structure Function and Degradation: Proteasomes, Overview](#); [Protein Degradation](#); [Ubiquitin-Like Proteins](#).

Further Reading

- Glickman MH and Ciechanover A (2002) The ubiquitin–proteasome proteolytic pathway: Destruction for the sake of construction. *Physiological Reviews* 82: 373–428.
- Goldberg AL, Elledge SJ, and Wade J (2001) The cellular chamber of doom. *Scientific American* 68–73.
- Hilt W and Wolf DH (2000) *Proteasomes: The World of Regulatory Proteolysis*. Georgetown, TX: Eureka.com/LANDES BIOSCIENCE Publishing Company.
- Weissman AM (2001) Themes and variations on ubiquitylation. *Nature Reviews Molecular Cell Biology* 2: 169–179.

UmuC D Lesion Bypass DNA Polymerase V

L A Hawver and P J Beuning, Northeastern University, Boston, MA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Zvi Livneh, volume 4, pp. 308–312, © 2004, Elsevier Inc.

Glossary

Base substitution mutation A change in DNA in which one DNA base is replaced with another. A purine substituted with a purine or a pyrimidine substituted with a pyrimidine are examples of transition mutations. A purine substituted with a pyrimidine or vice versa are examples of transversion mutations.

Co-protease A protein that facilitates the proteolytic action of another enzyme. For example, RecA facilitates the autocleavage of LexA to initiate the SOS response.

Frameshift mutation A change in DNA that alters the reading frame, which involves either deletion or insertion of one or more nucleotides. Addition or insertion of any number of nucleotides that is not a multiple of three will result in frameshift mutagenesis.

Homologous recombination RecA of *Escherichia coli* binds to ssDNA and aligns it with a homologous sequence in double-stranded DNA, therefore promoting the exchange of the double-stranded DNA piece to the single-stranded DNA and a new double strand is formed.

Lesion Damaged DNA, usually a modified base resulting from chemical or environmental stressors.

Mutagenesis Formation of a mutation, or the insertion of an incorrect base opposite a base or lesion in the template DNA strand.

Translesion synthesis (TLS) The process of copying damaged DNA templates. Depending on the nature of the damage and the polymerase utilized, TLS may be accurate or error prone.

Weigle reactivation Survival of UV-exposed λ bacteriophage is increased when the infected *E. coli* is pre-irradiated with UV light.

The SOS Response: Transcriptional Regulation of Pol V

The SOS response is a cellular response to DNA damage and is named after the universal distress call SOS – ‘save our ship’ or ‘save our souls.’ Upon DNA damage, the replicative DNA polymerase, pol III, is generally unable to copy noncanonical or damaged DNA and so becomes uncoupled from DNA helicase. This leads to the formation of a region of single-stranded DNA (ssDNA), on which the RecA protein polymerizes, a process that requires adenosine triphosphate (ATP), to form the RecA/ssDNA nucleoprotein filament, also known as RecA* (Figure 1). The RecA protein has multiple roles in responding to DNA damage, including in a DNA repair pathway known as homologous recombination repair. The *lexA* gene product, LexA, serves as the repressor of the SOS genes, binding to a consensus regulatory sequence known as the ‘SOS box’. LexA is a dimeric protein consisting of a DNA-binding domain and a protease domain. RecA*, acting as a co-protease, facilitates autocleavage of LexA, a reaction which is carried out by the catalytic dyad Ser119–Lys156 of LexA, at its Ala84–Gly85 bond. LexA cleavage prevents dimerization and results in the exposure of a protease targeting signal on LexA, which leads to degradation of LexA and therefore derepression of the SOS genes. In the SOS response, the expression of at least 57 genes is induced, allowing for the expression of the *umuC* and *umuD* genes among many others, including those involved in DNA repair and regulation of the cell cycle. The *umuD* and *umuC* genes are organized in an operon that is regulated by a single LexA box located upstream of the *umuD* gene. The start codon of *umuC* overlaps by 1 nt with the stop codon of *umuD*. The expression of these genes is coordinately regulated;

however, the *umuC* gene product is present at lower steady-state levels than the *umuD* gene products. UmuC is present at about 15 molecules per cell prior to SOS induction. Once SOS is induced and UmuC is generated, this number increases to about 200 molecules per cell. UmuD is present at about 200 molecules per cell and increases to approximately 2400 molecules per cell after SOS induction.

Not all genes are expressed to their highest levels at the moment that LexA is cleaved. The *umuDC* genes are expressed about 20–50 min after LexA cleavage. This delay gives time for other, accurate modes of DNA damage repair, such as nucleotide excision repair, to occur before potentially mutagenic translesion synthesis (TLS) is utilized. Full-length UmuD together with UmuC have a role in delaying the resumption of DNA replication after DNA damage. This function is beneficial to the cell in that it allows time for accurate repair processes to take place. UmuD and UmuC at elevated levels cause a cold sensitive growth phenotype that is not well understood, but appears to correlate with inhibition of DNA replication. Full-length UmuD also prevents mutagenesis, as cells expressing a noncleavable variant of UmuD are nonmutable.

Regulation of UmuC Function by Cleavage of UmuD

Similar to the mechanism of LexA cleavage, UmuD₂ undergoes slow autocleavage via interaction with RecA* to give UmuD'₂ (Figure 2). The exceptionally slow cleavage reaction of UmuD₂ is likely important for the cell to delay the use of a potentially mutagenic process until necessary. The UmuD'₂ form of the protein is required for UmuC to be active in TLS, forming pol V

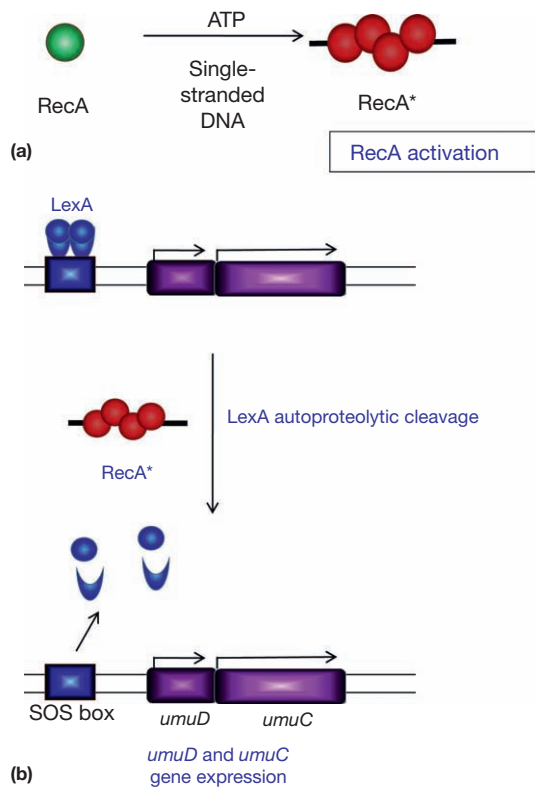


Figure 1 (a) When RecA binds to single-stranded DNA (ssDNA) in the presence of ATP, it is activated as RecA*. (b) This activated form of RecA serves as a coprotease in the autocleavage of LexA, which is the repressor of the SOS genes. Once LexA is cleaved, at least 57 SOS genes are expressed, and proteins such as UmuC and UmuD are translated.

as the UmuD'₂C complex. The UmuD active site is composed of the Ser60–Lys97 catalytic dyad, and cleaves between residues Cys24 and Gly25. Mixing a noncleavable UmuD active site variant with a noncleavable UmuD cleavage site variant results in UmuD cleavage, demonstrating that UmuD'₂ can cleave in *trans*. That is, the active site of one monomer cleaves the arm of its partner in the dimer. There is currently no high-resolution structure of UmuD'₂. The solution NMR and X-ray structures of UmuD'₂ that have been determined are quite different, suggesting that UmuD'₂ is a highly flexible and dynamic protein. This also seems to be true of full-length UmuD'₂. Although UmuD'₂ is a very tight dimer, it readily exchanges monomers, so that if two different UmuD'₂ variants are combined, they form heterodimers. The UmuDD' heterodimer has been found to be the most stable dimeric form, which has implications for protein half-life in the cell, discussed below.

umuDC Role in Mutagenesis

Mutagenesis and its effects were studied long before UmuD'₂C was determined to be a DNA polymerase. In 1953, Jean Weigle discovered that survival of UV-irradiated bacteriophage λ increased if the *Escherichia coli* host cell that it infected was also UV-irradiated prior to infection. This phenomenon, named

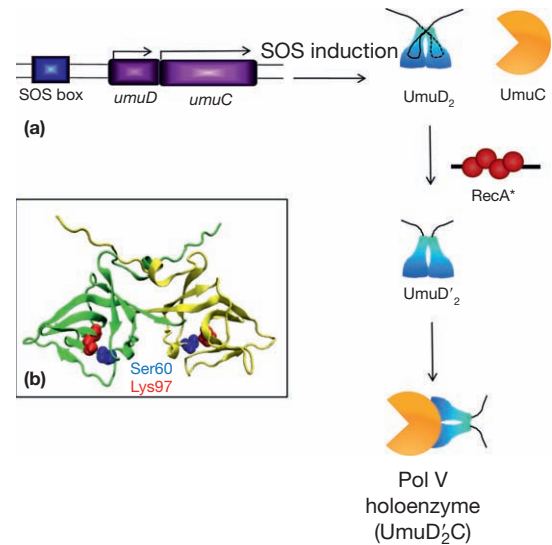


Figure 2 (a) After expression, UmuD'₂ undergoes an autocleavage reaction facilitated by RecA*, similar to that of LexA. UmuD'₂ is cleaved to form UmuD'₂, the active form in pol V. UmuD'₂C is then formed, which is the active DNA polymerase V in *E. coli*. (b) The X-ray crystal structure of UmuD'₂, with each monomer shown as green or yellow and the active site Ser and Lys residues highlighted as blue and red spheres, respectively (PDB ID: 1AY9).

Weigle reactivation, was strong evidence that there was a pathway in *E. coli* that responds to DNA damaged by UV radiation. In 1967, Evelyn Witkin suggested that this hypothesized pathway may be controlled by a repressor that is inactivated by UV radiation. Witkin also suggested that the repressor becomes active once again when the pathway is complete. Miroslav Radman named this pathway that is induced by DNA damage the 'SOS repair' pathway, later known as the SOS response. The SOS response and in particular the functions of the *umuDC* genes were characterized genetically for well over 20 years, before it was discovered that UmuD'₂C is a translesion DNA polymerase.

The *umuDC* genes were so named because *E. coli* strains lacking these genes were unmutable when treated with UV light or other DNA damaging agents. For years the SOS response was referred to as a 'mutagenic repair' response, because while enhanced cell survival was observed, such survival came at the cost of mutations. It is now appreciated that mutagenesis is an active process facilitated by the TLS activity of UmuD'₂C. Mutations generated by the *umuDC* gene products may confer a selective advantage on cells and, therefore, may be a driving force in bacterial evolution. Deletion of the *umuDC* genes actually has little effect on the ability of *E. coli* to survive most DNA damaging treatments. Remarkably, cells harboring a noncleavable variant of UmuD (Ser60Ala), besides being nonmutable, are extremely sensitive to UV light. Thus, the presence of only full-length UmuD'₂ (with UmuC) that cannot be converted to UmuD'₂ is more detrimental to a cell than the complete absence of the *umuDC* genes. This is thought to be related to the functions of UmuD'₂ and UmuC in inhibiting DNA replication, but the exact mechanism is unknown.

The Discovery of Pol V

Originally, it was thought that UmuC and UmuD were accessory factors that somehow modified the replicative polymerase pol III to allow it to bypass lesions in DNA. The difficulty in purifying active, soluble UmuC alone or in complex with UmuD₂ contributed to the persistence of this model. In 1999, the purification of UmuC was reported by two different labs. Zvi Livneh and researchers purified UmuC fused to maltose binding protein (MBP), while Myron Goodman and researchers, working with the group of Roger Woodgate, purified UmuC in a complex with UmuD₂. This allowed the Livneh group to determine that UmuC alone possesses polymerase, but not TLS activity, and that the presence of UmuD₂ is required for UmuC to be active in TLS. Both groups used different DNA substrates as well, with the Livneh group using circular, gapped DNA while the Goodman group used linear DNA. Differences in protein preparation and the substrates used likely led to differing conclusions about the importance in TLS of accessory proteins such as single-stranded DNA-binding protein (SSB), the β processivity clamp, and the γ clamp loader, as well as the role of RecA in these reactions.

Mechanism of Translesion Synthesis by Pol V

It was originally thought that a RecA/ssDNA nucleoprotein filament was necessary for TLS by pol V. It is also clear that a stable RecA filament on ssDNA would serve as a steric block and not permit synthesis by pol V, so it was proposed that pol V removes RecA from the DNA template to allow TLS. However, it was soon discovered that RecA could stimulate pol V under conditions that were not suitable for filament formation. In addition to its roles in SOS induction and UmuD cleavage, RecA has another role in TLS, namely, in the direct stimulation of pol V. RecA, paired with a nucleotide cofactor such as ATP, stimulates TLS by pol V without the presence of the nucleoprotein filament at the TLS site. The pol V mutasome (pol V Mut) consists of pol V (UmuD₂C), RecA, and ATP, and is the active form of pol V. It has been observed that a filament is required to activate RecA, so it is proposed that the RecA/ssDNA nucleoprotein filament transfers a RecA monomer and an ATP cofactor from the 3' end of the filament to UmuD₂C. In addition to the pol V mutasome, interactions with the β clamp stimulate pol V activity and SSB enhances TLS *in vitro*.

The current model of TLS holds that the replicative DNA polymerase encounters a lesion and is unable to copy past it (Figure 3). The formation of a RecA/ssDNA nucleoprotein filament may facilitate targeting of pol V to lesions, although there are many unanswered questions about how TLS polymerases are recruited to damaged DNA, if indeed they are, and the mechanism of polymerase switching at sites of DNA damage. Pol V synthesizes a TLS patch opposite and past the lesion. This patch must be large enough so that when the replicative DNA polymerase resumes synthesis, its proofreading exonuclease activity is not activated by the presence of the lesion. If the TLS patch is too small, the replicative DNA polymerase could become stalled at the lesion in a futile

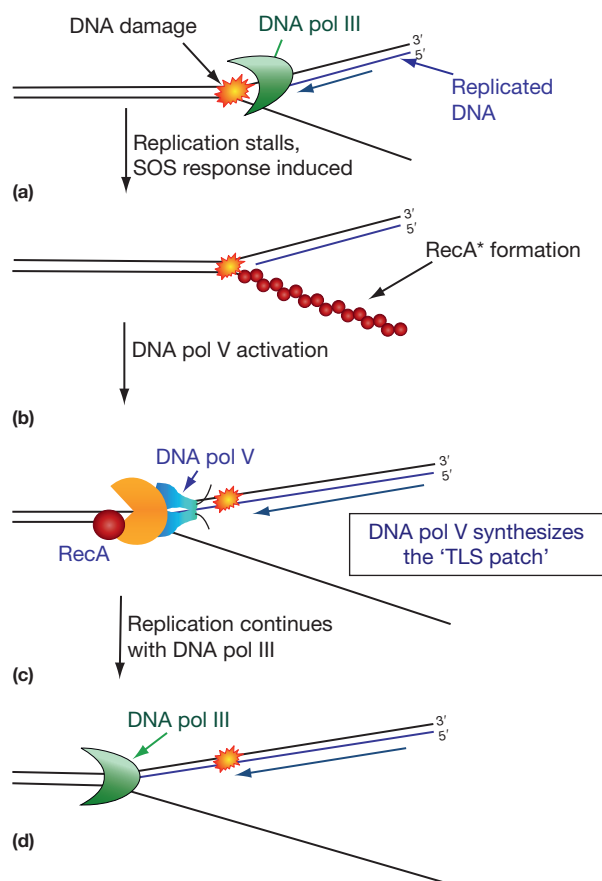


Figure 3 Upon DNA damage, the *E. coli* replicative DNA polymerase III (pol III) cannot continue replication. The polymerase stalls (a), causing RecA to polymerize on the single-stranded DNA (b). Polymerase V (pol V) is induced by the SOS response and is able to bypass the lesion (c). Once pol V bypasses the lesion, it continues for about 6–8 nt and then dissociates, allowing pol III to resume replication of the DNA (d).

cycle of nucleotide insertion and excision. Pol V needs to synthesize at least 5 nt past the lesion to allow pol III to resume efficient replication.

Specificity, Fidelity, and Processivity of Pol V

Pol V is specific for bypass of UV photoproducts such as thymine–thymine (T–T) *cis-syn* cyclobutane pyrimidine dimers (CPDs) and T–T (6–4) photoproducts, as well as abasic sites (Figure 4). Pol V inserts two A bases opposite T–T CPDs, and so is accurate when copying that lesion. However, pol V is prone to inserting guanine (G) opposite the 3' T of T–T (6–4) photoproducts. This is considered the mutagenic signature of pol V due to the fact that normally, adenine (A) is inserted across from T. Moreover, this correlates with the mutation spectrum observed *in vivo* in SOS mutagenesis. When copying abasic sites, pol V inserts mostly A, following the general 'A-rule' of DNA polymerases, as well as G. Pol V also has a role in bypass of *N*²-benzo[a]pyrene–dG as well as certain oxidation products of G.

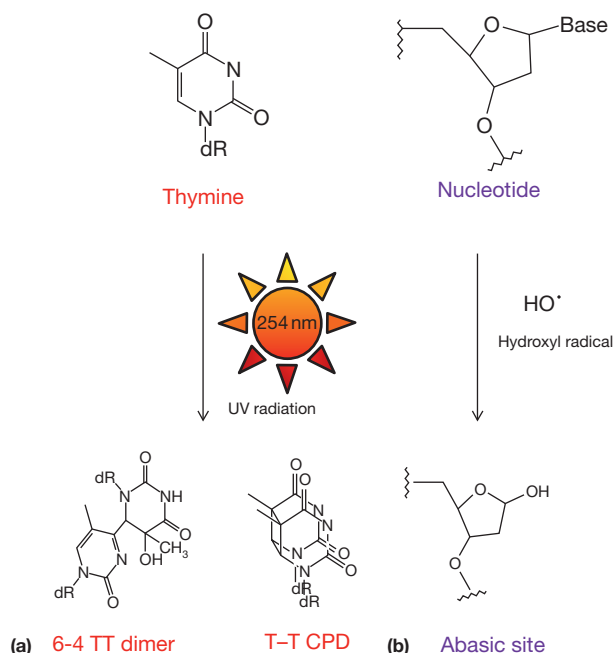


Figure 4 Environmental mutagens such as (a) UV radiation and chemical mutagens like hydroxyl radicals (b), among many other agents, contribute to DNA damage. Common UV photoproducts are modified thymine bases, such as thymine–thymine 6–4 photoproducts and *cis-syn* thymine–thymine cyclobutane pyrimidine dimers (CPDs). Hydroxyl radicals, as well as other agents including ionizing radiation, can lead to the formation of abasic sites. Abasic sites are also generated spontaneously. Pol V bypasses all of the lesions shown.

Pol V has low fidelity when copying undamaged DNA, with error rates of 10^{-3} – 10^{-4} compared to pol III, which is 100–1000-fold more accurate than pol V. Y family polymerases are characterized by low processivity; pol V inserts about 6–8 nt prior to dissociation from the DNA. The processivity of pol V is increased only modestly by the β processivity clamp.

Characteristics of Y Family Polymerases

A major reason that Y family polymerases were not identified as polymerases despite years of genetic characterization is because they possess little sequence homology to DNA polymerases from other families. While many high-resolution structures of Y family DNA polymerases have been determined, there is currently no such structure of pol V. Remarkably, Y family polymerases adopt the classical DNA polymerase right hand conformation, although with some differences. The palm region contains the active site of the polymerase where the incoming nucleotide is bound and incorporated into the primer DNA strand. The fingers and thumb regions bind the template and primer DNA strands. A little finger domain unique to the Y family polymerases makes contacts with the DNA that may facilitate specific lesion bypass and processivity.

In general, replicative DNA polymerases ensure exquisite accuracy through a combination of mechanisms. They have active sites that enforce correct base pairing on the incipient base pair due to their tight fit and relatively closed

conformation. Y family polymerases have open, solvent-accessible active sites, which allow bulky lesions to be bound in the active site and are less selective when binding undamaged DNA. Replicative polymerases have an α -helix, the so-called O helix, that ensures the fidelity of the newly forming base pair. Y family polymerases are clearly missing the O helix and thus lack this important contribution to specificity. Replicative polymerases have either intrinsic or associated 3' to 5' exonuclease domains that proofread the newly formed base pair and can remove the new primer terminus if necessary. Y family polymerases lack intrinsic proofreading and thus are better able to bypass lesions without interference from proofreading, but also exhibit lower fidelity when copying undamaged DNA.

Regulation

In addition to transcriptional regulation via the SOS response and posttranslational regulation by cleavage of UmuD₂, pol V activity is regulated by interactions with a number of other proteins. Pol V is activated by RecA and requires RecA–ATP in order to sustain activity. Once pol V Mut synthesizes past the lesion and dissociates, it is no longer active and must again be activated by another RecA–ATP complex. This is a possible mechanism to limit the potentially mutagenic activity of pol V. Pol V TLS activity is stimulated by the β processivity clamp, and interactions with the β clamp are critical for UV-induced mutagenesis *in vivo*. These interactions may be important for polymerase switching or recruitment of pol V to sites of DNA damage.

Both UmuD₂ and UmuD'₂ interact with the α polymerase, β processivity, and ϵ proofreading subunits of DNA polymerase III. UmuD₂ interacts more strongly with the β clamp than UmuD'₂ does, although both interactions are detectable. It is thought that interaction of UmuD₂ with the β clamp may facilitate inhibition of replication, and that cleavage of UmuD₂ to form UmuD'₂ may release the inhibition. UmuD'₂ interacts more strongly with α than UmuD₂ does, but the significance of this interaction is currently unknown.

UmuD, UmuD', and UmuC are all regulated by proteolysis. UmuC and UmuD₂ are both substrates for the Lon protease. UmuD serves as an adaptor to deliver its partner, whether UmuD or UmuD', to the ClpXP protease for degradation. Given that the UmuDD' heterodimer is the most stable dimeric form, this may be a way for the cell to decrease the levels of UmuD', and therefore pol V, present.

Bacterial Homologs

E. coli has another member of the Y family, DinB or polymerase IV, that also copies damaged DNA but with different specificity than pol V. DinB is known to bypass N²–dG adducts, such as N²–furfuryl–dG and N²–benzo[a]pyrene–dG with high accuracy. Both UmuC and DinB are regulated by the SOS response to DNA damage. DinB, particularly at high levels, causes –1 frameshift mutagenesis, an activity that can be suppressed by UmuD₂ and RecA. The eukaryotic homolog of DinB is pol κ , and it is the only Y family DNA polymerase specifically conserved throughout all domains of life.

Homologs of *umuC* and *umuD*, found in many but not all bacteria, share significant sequence homology, similar functionality, and similar organization in operons. Homologs of UmuD tend to share the same cleavage site and to cleave in a similar manner. Examples of these are *impAB*, *samAB*, and *mucAB*. MucA and MucB are found on native conjugational plasmids. MucA cleaves itself to form MucA', much like UmuD cleavage to form UmuD'. It is said that MucA' and MucB are much stronger mutators due to the efficiency of MucA cleavage compared to UmuD. MucA' and MucB form DNA polymerase RI, which requires RecA and SSB for lesion bypass. Notably, the *mucAB* genes are found on pKM101, a plasmid added to the *Salmonella* spp. strains used in the Ames mutagenicity test.

Eukaryotic Homologs

Y family polymerases are conserved throughout all domains of life. Humans have four members of the Y family, specifically Rev1, pol ι, pol κ, and pol η. It is possible for DNA polymerases other than those in the Y family to perform TLS, with the B family polymerase pol ζ being one notable example. The yeast *Saccharomyces cerevisiae* lacks an identifiable gene for pol κ, although pol κ is present in other yeast species. Rev1 promotes mutagenesis in both organisms mentioned above, as well as others. Interestingly, the *REV1* gene is required for TLS past the (6–4) thymine–thymine UV photoproduct, even though the Rev1 polymerase cannot bypass the lesion. The catalytic activity of Rev1 is not required for all TLS events for which the protein itself is required, suggesting that Rev1 has a scaffolding role in TLS. Pol κ is the eukaryotic homolog of DinB and exhibits similar lesion specificity. Pol ι uses Hoogsteen base pairing with damaged template bases in an unusual mode of DNA synthesis.

Pol η is a functional homolog of pol V. Pol η is known to bypass T–T CPD dimers in an error-free manner. It is most well known for the consequences of loss of its activity. Humans deficient in pol η develop xeroderma pigmentosum variant (XPV), which is a genetic disorder causing skin to be extremely vulnerable to UV radiation from the sun, leading to a high incidence of skin lesions and skin cancer. The disease stems from an increase in UV mutagenesis due to other polymerases substituting for pol η in a potentially error-prone manner.

Other forms of XPVs stem from defects in nucleotide excision repair, an accurate repair pathway.

Conclusions

DNA damage is ubiquitous and multiple pathways exist to respond to such damage. The widespread conservation of Y family DNA polymerases indicates their importance for all organisms. Most Y family polymerases, especially *E. coli* pol V, are potentially mutagenic and therefore potentially harmful. Therefore, while their functions are clearly important, proper regulation is also crucial.

See also: **Molecular Biology:** DNA Base Excision Repair Pathways; DNA Damage: Alkylation; DNA Mismatch Repair and Homologous Recombination; DNA Polymerase I, Bacterial; DNA Polymerase III, Bacterial; DNA Polymerases: Kinetics and Mechanism; LexA Regulatory System; Nucleotide Excision Repair, Bacterial: The UvrABC System; XPV Polymerase and the Bypass of Ultraviolet DNA Damage.

Further Reading

- Friedberg EC, Walker GC, Siede W, et al. (2006) *DNA Repair and Mutagenesis*, 2nd edn. Washington, DC: ASM.
- Jarosz DF, Beuning PJ, Cohen SE, and Walker GC (2007) Y-family DNA polymerases in *Escherichia coli*. *Trends in Microbiology* 15: 70–77.
- Nohmi T (2006) Environmental stress and lesion-bypass DNA polymerases. *Annual Reviews in Microbiology* 60: 231–253.
- Patel M, Jiang Q, Woodgate R, Cox MM, and Goodman MF (2010) A new model for SOS-induced mutagenesis: How RecA protein activates DNA polymerase V. *Critical Reviews in Biochemistry and Molecular Biology* 45: 171–184.
- Radman M (1974) Phenomenology of an inducible mutagenic DNA repair pathway in *Escherichia coli*: SOS repair hypothesis. In: Prakash L, Sherman F, Miller MW, Lawrence CW, and Taber HW (eds.) *Molecular and Environmental Aspects of Mutagenesis*, pp. 128–142. Springfield, IL: Charles C. Thomas.
- Schlacher K and Goodman MF (2007) Lessons from 50 years of SOS DNA-damage-induced mutagenesis. *Nature Reviews in Molecular and Cell Biology* 8: 587–594.

Relevant Website

<http://www.ecoliwiki.net> – EcoliWiki.

Uncoupling Proteins

D Ricquier and F Bouillaud, Centre National de la Recherche Scientifique and Université Paris Descartes, Paris, France

© 2013 Elsevier Inc. All rights reserved.

Glossary

Mitochondria These organelles are the sites of respiration and oxidative phosphorylation in all animal and higher plant tissues as well as in protozoa, fungi, and aerobically grown yeasts. They are ~2–30 µm long and 0.5–10 µm wide.

Mitochondrial transporters The uncoupling proteins (UCPs) actually belong to a family of membranous mitochondrial transporters including the adenosine diphosphate/adenosine triphosphate (ADP/ATP) translocator, the phosphate carrier, the oxoglutarate carrier, the citrate carrier, and the acylcarnitine carrier. All these

transporters play an important role in the metabolic dialog between cytosol and mitochondria. They all derive from the same ancestor gene.

Respiration coupling/uncoupling In coupled mitochondria, respiration rate determines ADP phosphorylation rate and ATP synthesis determines respiration rate. When uncoupling occurs (due to chemical uncouplers or UCP), the respiration rate increases sharply since the control by ADP phosphorylation does not limit respiration. In such a situation, oxidation energy is dissipated as heat.

Introduction

Uncoupling proteins (UCPs) are mitochondrial transporters present in the inner membrane of mitochondria. They belong to the family of anion mitochondrial carriers including adenine nucleotide transporters, phosphate carrier, and other transporters. The term ‘uncoupling protein’ was originally given to UCP1 which is uniquely present in mitochondria of brown adipocytes, the thermogenic cells devoted to maintenance of body temperature in mammals. In these cells, UCP1 acts as a proton carrier creating a shunt between complexes of respiratory chain and the adenosine triphosphate (ATP)-synthase. Purine nucleotide inhibits UCP1, whereas fatty acids activate it. Activation of UCP1 stimulates respiration and the uncoupling process results in a futile cycle and dissipation of oxidation energy into heat. UCP2 is ubiquitous and highly expressed in lymphoid system and macrophages and also in gut, lung, and brain. UCP3 is mainly expressed in skeletal muscles. In comparison to the established uncoupling and thermogenic activities of UCP1, UCP2, and UCP3 rather appear to be involved in the limitation of free radicals levels in cells than in physiological uncoupling and thermogenesis. These novel UCPs are also involved in sparing glucose utilization and consequently in the control of fatty acids and glutamine metabolism. The mechanism of the protonophoric activity of UCP1 is still controversial since it has been proposed that UCP1 transports fatty acid anions and catalyzes proton transport by fatty acid cycling. Quinones and superoxide ions may also activate the UCPs.

The Mitochondria and the Coupling of Respiration to ADP Phosphorylation

Cellular respiration, the reactions of the citric acid cycle, fatty acid oxidation, and several steps of urea synthesis and steroid hormone synthesis and gluconeogenesis take place in specialized cellular organelles, the mitochondria.

Mitochondria

In addition to oxidative phosphorylation and metabolic pathways, mitochondria are involved in thermogenesis, radical production, calcium homeostasis, apoptosis, and protein synthesis. Mitochondria contain two compartments bounded by inner and outer membranes. The outer membrane is permeable to many small metabolites whereas the permeability of the inner membrane is controlled in order to maintain the high electrochemical gradient created by the mitochondrial respiratory chain, which is necessary for energy conservation and ATP synthesis in mitochondria. The inner membrane transports anion substrates such as adenosine diphosphate (ADP), ATP, phosphate, oxoglutarate, citrate, glutamate, and malate.

Coupling of Respiration to ATP Synthesis

It has long been known that respiration and mitochondrial ATP synthesis are coupled. The observation that respiration rate increased when mitochondria synthesized more ATP led to the concept of respiratory control by ADP phosphorylation. In fact, there is a link between mitochondrial ATP synthesis and cellular ATP demand by a feedback mechanism. In agreement with Mitchell's theory, it was demonstrated that the mitochondrial electrochemical proton gradient, generated as electrons are passed down the respiratory chain, is the primary source for cellular ATP synthesis. In this way, several complexes of the respiratory chain pump protons outside of the inner membrane during reoxidation of coenzymes and generate a proton gradient, which is consumed by the reactions of ATP synthesis. Proton leak represents another mechanism consuming the mitochondrial proton gradient. Mitchell's theory predicted that any proton leak noncoupled to ATP synthesis would represent an uncoupling of respiration and result in thermogenesis. An excellent example of such an uncoupling of respiration to ADP phosphorylation is represented by the mitochondrial UCP of brown adipocytes (UCP1) which dissipates energy of substrate oxidation as heat.

Brown Adipose Tissue and UCP1: History of a True Respiration Uncoupling

Maintenance of body temperature in a cold environment or at birth requires thermogenesis. Another situation requiring thermogenesis is arousal from hibernation. The two regulatory thermogenic processes are shivering and metabolic thermogenesis also referred to as nonshivering thermogenesis. The largest part of nonshivering thermogenesis in small mammals is achieved in the brown adipose tissue (BAT).

Brown Adipose Tissue

BAT is easily found in all small mammals and in the newborn of larger mammals, such as humans. Recent data based on the use of positron emission tomography-SCAN analysis of glucose uptake have demonstrated that BAT is present in human adults. BAT is located in specific body areas near large blood vessels and consists of brown adipocytes which are distinct from white adipocytes. The brown adipocytes are characterized by numerous mitochondria containing a highly developed inner membrane (Figure 1). Activation of thermogenesis in BAT occurs in newborns, rodents exposed to cold, and in animals emerging from hibernation. It is commanded by the central nervous system and the orthosympathetic fibers innervating each brown adipocyte. The norepinephrine

released by these fibers binds to adrenergic receptors on the surface of the brown adipocytes. The later steps of the activation of thermogenesis in brown adipocytes are production of cyclic adenosine monophosphate, activation of lipolysis, and oxidation of fatty acids by the numerous mitochondria. Released fatty acids stimulate brown adipocyte respiration and heat production.

Respiration Uncoupling Is the Thermogenic Mechanism in Brown Adipocytes

In most cell types such as muscular fibers, ATP utilization stimulates ADP phosphorylation and respiration (Figure 2). Original observations made in 1967 revealed that the high respiratory rate in brown adipocyte mitochondria was not controlled by mitochondrial ADP phosphorylation suggesting that energy from substrate oxidation was dissipated into heat instead of being converted into ATP (Figure 2). In line with Mitchell's chemiosmotic theory, if proton leakage is activated by a signal such as free fatty acids in the case of the mitochondria of brown adipocytes, the mitochondrial membrane potential is decreased and concomitantly, respiration is activated. Since, in addition, the activated proton leakage is not linked to ADP phosphorylation, respiration is uncoupled from ATP synthesis and oxidation energy is dissipated in the form of heat.

The UCP of Brown Adipocyte Mitochondria

The protein responsible for proton leakage within the inner mitochondrial membrane and for respiratory uncoupling was identified in 1976–77, was purified, and its complementary DNA (cDNA) was cloned. It has an apparent molecular weight of 33 000 and was originally called UCP. This protein is known as UCP1 since the discovery of UCP2 in 1997. UCP1 is abundant in the inner membrane of the mitochondria of brown adipocytes and is specific to these cells. It short-circuits the proton circuit between complexes of the respiratory chain and the mitochondrial ATP-synthase, which is the main consumer of the proton gradient (Figure 3). When there is no need for thermogenesis, purine nucleotides bound to UCP1 inhibit its activity. Upon activation of thermogenesis by the sympathetic nervous system, fatty acids overcome the inhibitory effect of nucleotides and activate UCP1. Heterologous expression of UCP1 in yeasts or mammalian cell lines induces respiration uncoupling. The protonophoric activity of UCP1 has been reconstituted in liposomes where it can be inhibited by nucleotides and activated by free fatty acids. The mechanism of action of UCP1 is still controversial: some scientists believe that it is a true proton transporter, while others assert that it returns anionic fatty acids to the intermembrane space, after they have crossed the membrane in protonated form. The observation that *Ucp1^{-/-}* mice were unable to maintain body temperature in the cold proved that the UCP1-induced uncoupling of respiration was responsible for cold-activated nonshivering thermogenesis of rodents. In agreement with this observation, ectopic expression of UCP1 in skeletal muscles of transgenic mice promotes substrate oxidation in muscles and resistance to obesity and type 2 diabetes. UCP1 functions as a dimer and each monomer is made of six transmembrane

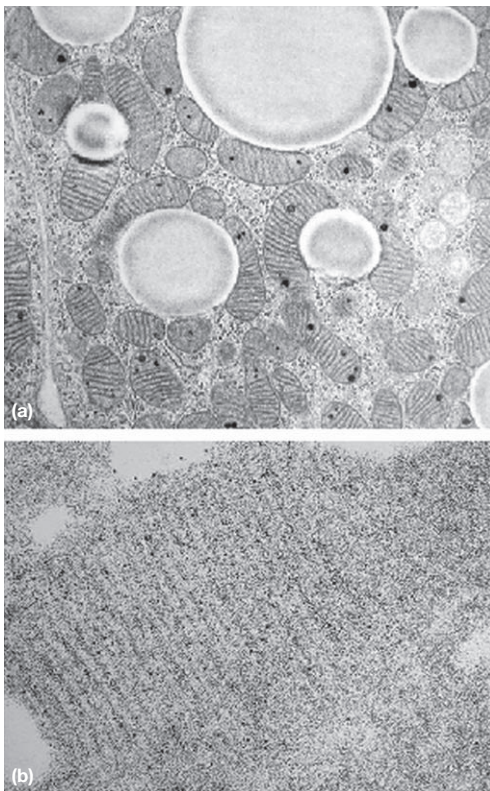


Figure 1 Histology of brown adipocyte. (a) The cytosol of brown adipocytes is characterized by numerous mitochondria and lipid droplets. (b) Magnification of a brown adipocyte mitochondrion showing parallel cristae and UCP1 detected using antibodies (black dots). Figure kindly provided by Dr. Saverio Cinti, University of Ancona.

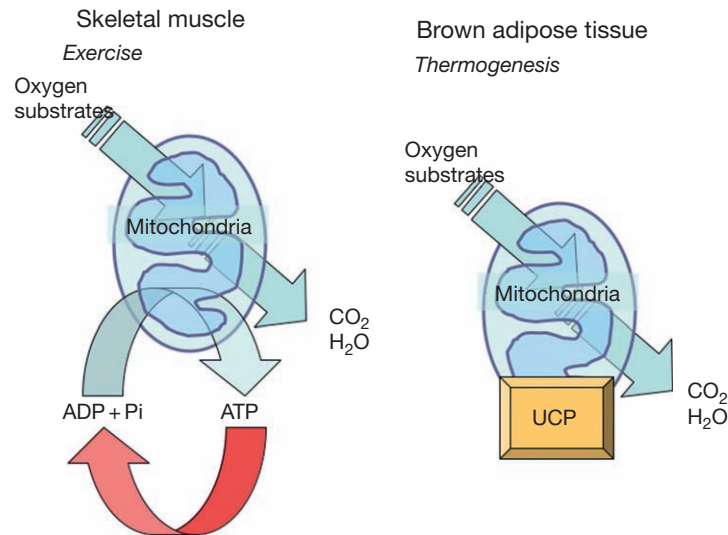


Figure 2 Comparison of skeletal muscle and brown adipose tissue mitochondria. (a) Mitochondria are cellular organelles converting the redox energy liberated during oxidation of organic substrates into ATP, a molecule containing energy under a form readily usable by most enzymes working to maintain cell structure and integrity or to perform work such as mechanical work in muscle. The cellular need for ATP controls oxidation rate by mitochondria. (b) In brown adipocyte, a specific uncoupling protein referred to as UCP (precisely UCP1) is present in a large amount in the inner membrane. UCP1 is able to fool mitochondria which accelerates their oxidation rate whereas no ATP is produced. UCP1 removes the control of energy expenditure by ATP demand and therefore oxidation rate and energy expenditure increase under the form of heat. This process is used in newborns, during arousal from hibernation, and in young and adult mammals exposed to cold when muscle thermogenesis would not be sufficient.

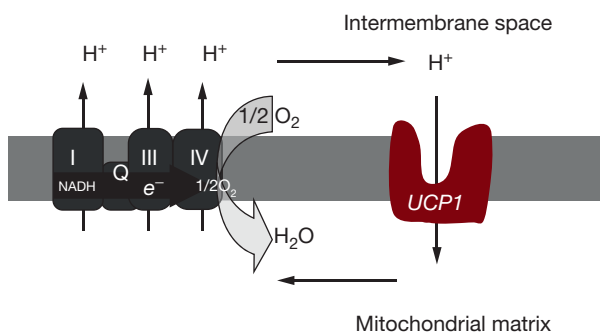


Figure 3 UCP1-induced proton cycling. Q (coenzyme Q) refers to complex II, which is the succinate dehydrogenase. This enzyme directly reduces coenzyme Q but has no proton pumping activity contrary to complexes I, III, and IV. UCP is inserted in the mitochondrial inner membrane where a multienzymatic complex called the 'respiratory chain' made of complexes I–IV is also present. The respiratory chain reoxidizes reduced coenzymes and electrons are driven to oxygen. This oxido-reduction step liberates energy, which is used to generate an electrochemical gradient of protons across the inner membrane. This gradient is normally consumed by the ATP-synthase, which phosphorylates ADP. UCP1 transports protons passively and makes possible a futile cycle of protons across the inner membrane leading to increased energy expenditure. This schema illustrates the situation encountered in brown adipocytes of mammals where a large amount of respiratory chains as well as a large amount of UCP1 are present. Activation of the futile cycling increases considerably energy expenditure and thus heat production in these thermogenic cells. In other cells where homologs of the UCP1 are expressed at a much lower level, this pathway would represent a minor contributor to energy expenditure, but might be of importance to avoid oxidative damage.

fragments (Figure 4). Moreover, UCP1, the ADP/ATP translocator, and other mitochondrial anion carriers derive from the same ancestral gene and probably share a similar structure. They all have a triplicated structure (Figure 4).

The Novel UCPs

Given their specific role in thermogenesis, it has always seemed logical that brown adipocytes be equipped with an original mechanism, partial and regulatable uncoupling of respiration, brought into play by a specific protein, UCP1, which induces proton leakage. In fact, it is known that mitochondrial respiration is always accompanied by heat production as it is imperfectly coupled to ADP phosphorylation in all types of cells. To explain this incomplete coupling of respiration and the energy loss mechanism, some authors have invoked slippage of the respiratory chains, while others referred to proton leaks. It has been estimated that proton leaks from the inner membrane of mitochondria of hepatocytes and myocytes could explain about 20% of the body's basal metabolism.

UCP2: A Homolog of the Brown Fat UCP Present in Various Tissues

A clone corresponding to a protein with 59% sequence identity with the UCP of brown fat was isolated. This new UCP, called UCP2, was also identified in humans. In contrast with UCP1, the messenger RNA (mRNA) of UCP2 is present in almost all tissues and many cell types, such as spleen, thymus, digestive

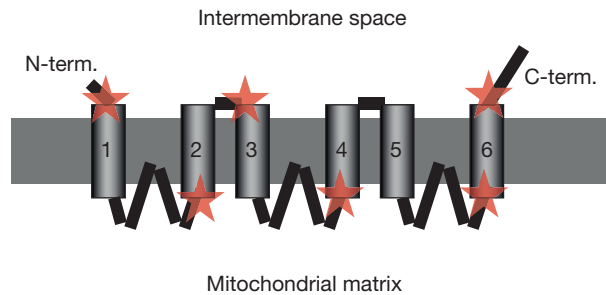


Figure 4 Model of insertion of UCP in the membrane. To modelize the organization of such proteins, indirect methods such as prediction of secondary structure by computers can be used. Experimentally, restricted domains of the protein in the membrane (indicated by stars) were explored using specific antibodies. The UCPs are made of six transmembrane segments (probably alpha helices) and three hydrophilic loops stemming from the matricial side of the membrane. An internal repetition in the structural arrangement of the protein suggests that the protein (made of 300 amino acid residues) is derived from the ancient triplication and divergence of a 100-amino-acid domain (made of two helices and one loop). This model fits with the structure of the mitochondrial ADP-ATP carrier recently identified by crystallization and X-ray diffraction analysis by Pebay-Peroula and Brandolin teams.

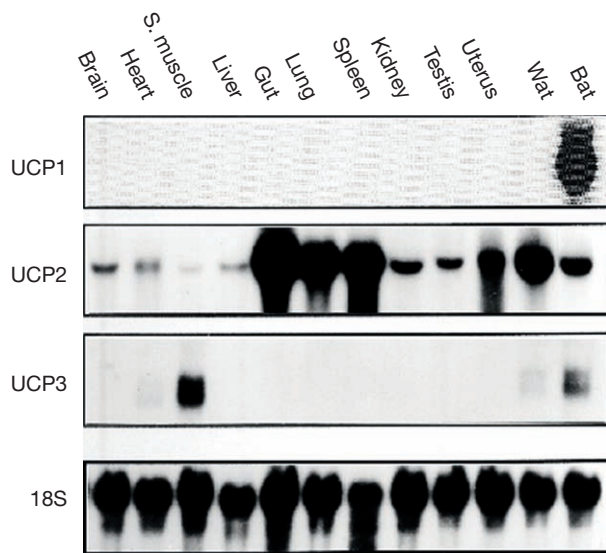


Figure 5 Tissue distribution of UCP1, UCP2, and UCP3 RNAs. The data correspond to mouse tissues. A very similar picture was obtained when analyzing human tissues. 18S refers to ribosomal RNA.

tract, lung, brain, adipose tissues, skeletal muscle, heart, adipocytes, myocytes, and macrophages (Figure 5). Expressed in yeasts, murine UCP2 decreased the potential of the mitochondrial membrane, raised the respiration rate, and reduced sensitivity to uncouplers: UCP2 therefore was able to uncouple respiration and appeared to be a second mitochondrial UCP. However, recent data do not support the idea that this uncoupling activity of UCP2 takes place *in vivo* and other transport activities might well be more relevant to physiology. The UCP2 gene is located on chromosome 7 of the mouse and chromosome 11 of humans, near a region linked to hyperinsulinism and obesity. The expression of UCP2 mRNA was measured in

obesity-prone and obesity-resistant mice given a lipid-rich diet: the obesity-resistant mice overexpressed UCP2 mRNA in their adipose tissue. These findings, together with the function of the protein and the chromosomal location of its gene, led to propose a role for UCP2 in diet-induced thermogenesis. However, this hypothesis was not validated in *Ucp2*^{-/-} mice.

UCP3: Another UCP Homolog Predominantly Present in Skeletal Muscle

cDNAs corresponding to UCP3, a protein homologous to UCP1 and UCP2, were cloned soon after the discovery of UCP2. The amino acid sequence of UCP3 is 72% identical to that of UCP2 and 57% identical to that of UCP1. UCP3 mRNA is predominantly expressed in the skeletal muscles of humans, mice, and rats (Figure 5). Human and murine UCP2 and UCP3 genes are juxtaposed on the same chromosome.

Plant and Bird UCPs

Functional studies of isolated plant mitochondria suggested that plants contain mitochondrial UCPs. Following this observation, a cDNA from a new UCP was isolated from a potato cDNA library. The sequence of the corresponding protein, called 'stUCP', is 44% identical to that of UCP1 and 47% identical to that of UCP2. The mRNA of stUCP is present in most plant organs. The most surprising result was that stUCP mRNA was markedly induced in the leaves of plants exposed to a temperature of 4 °C. Cold, therefore, induced stUCP as it induces UCP1 in the BAT of animals. However, plant and animal UCPs may have different functions. More recently, a chicken UCP termed 'avUCP' was characterized. This gene is only expressed in skeletal muscles of birds and ducks and is induced upon exposure to the cold. Therefore, avUCP seems to be involved in regulatory thermogenesis. Other proteins referred to as UCP4 and BMCP1/UCP5 were characterized. However, their exact relationship to UCP1, UCP2, and UCP3 will require further work.

Role and Function of UCPs other than UCP1: A Role in Controlling the Level of Reactive Oxygen Species and a Role in Glucose Metabolism

Reactive oxygen species

Guided by the fact that UCP2 is expressed at high level in the immune system, a role for UCP2 in immunity was searched for. In fact, *Ucp2*^{-/-} mice are more resistant to infection by a parasite. This phenotype was explained by the fact that disruption of the *Ucp2* gene provoked an elevation of reactive oxygen species (ROS) level that facilitated the killing of the pathogens. Mice lacking the *Ucp3* gene also produced more oxygen radicals in myocytes. Therefore, it appears that UCP2 and UCP3 contribute to limit the level of ROS in cells. Such an activity may be explained by a mild uncoupling activity of the proteins since a small decrease of the mitochondrial membrane potential opposes production of superoxide ion in mitochondria (Figure 6). The observation that UCP2 protects against atherosclerosis highlights its ability to counteract radical synthesis.

Numerous questions remain unanswered, notably concerning the exact catalytic activity of each UCP, and the nature

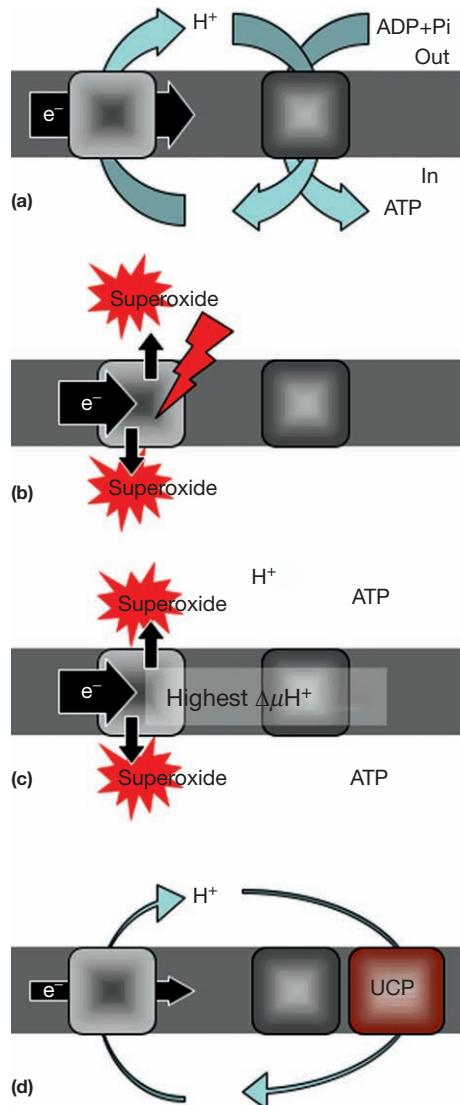


Figure 6 Role of UCP and superoxide generation by mitochondria. (a) Mitochondria phosphorylating ATP: a proton gradient generator (the respiratory chain, left square) uses the energy liberated by the movement of electrons to pump protons through the membrane (gray zone) towards the intermembrane space. A proton gradient consumer (the ATP-synthase, right square) phosphorylates ADP coming from outside into ATP, that is exported afterwards. Therefore, a proton cycle across the inner membrane couples oxidation to ATP synthesis (and *vice versa*). (b, c) When oxidation is impaired (b) or slows down (c), an excess of electrons aiming to travel across the respiratory complexes cannot reach oxygen (to form water). Single electrons react with molecular oxygen and form superoxide anion ($O_2^{\bullet-}$). It occurs either when the respiratory chain is poisoned (red arrow in panel b) or under physiological circumstances when the use of ATP by the cell is weak, ATP level rises, and ADP level decreases (panel c). In that case, the membrane potential rises to its highest value which impedes proton pumping, slows down reoxidation, and therefore increases superoxide production. (d) In tissues other than the thermogenic brown adipose tissue, the UCPs may induce a controlled leak of protons across the inner membrane that allows reoxidation to occur and maintain the membrane potential below the threshold inducing superoxide generation. The mechanism of proton transport (proton channeling or fatty acid cycling), as well as the identity of regulators of UCPs inside the cells are still controversial. Fatty acids would be

of the endogenous ligands of UCP2 and UCP3. These proteins may simply translocate protons. As said above for UCP1, it has been proposed that all the UCPs are active as fatty acid cyclers through the membrane. According to this later hypothesis, protonated fatty acids cross the membrane and release a proton on the matricial side; then, the UCP facilitates the translocation of the anionic form of fatty acids. High loads of UCP2 or UCP3 can uncouple respiration in yeast or mammalian cell, but it is not certain whether the low physiological levels of these proteins may be sufficient to induce a net uncoupling of respiration *in vivo*. The divergent observations of the regulation of the activities of UCP2 or UCP3 by fatty acids or nucleotides will require further studies. Interestingly, it has been proposed that UCP2 and UCP3 could export fatty acid anion under conditions of elevated fatty acid oxidation.

Metabolic activity of UCP2 and UCP3

Conflicting data regarding the association of UCP2 or UCP3 genetic polymorphisms to body mass index, susceptibility to obesity, resting metabolic rate, metabolic efficiency, fat oxidation, insulin resistance, and susceptibility to gain fat with age were obtained. However, UCP2 appears to act as a negative regulator of insulin secretion. Moreover, mice overexpressing a large amount of human UCP3 in skeletal muscle weigh less, have a decreased amount of adipose tissue, and an increased resting oxygen consumption. Recent data suggest that UCP2 and UCP3 favor glucose sparing and oxidation of alternate substrates (glutamine and fatty acids) in cells; this would be explained rather by the transport of a metabolite (pyruvate?) out of mitochondria than by uncoupling.

UCP1, UCP2, and UCP3, Conclusions and Perspectives

UCP1 has a well-demonstrated uncoupling activity and an essential role in maintenance of body temperature in small rodents exposed to the cold, but the exact biochemical and physiological roles of UCP2 and UCP3 remain to be further identified. Certain data support a true activity of these UCPs in respiration uncoupling and substrate oxidation, other data do not agree with a role for these UCPs in controlling adiposity. Several reports support a role for UCP2 and UCP3 in decreasing the level of ROS in cells. While UCP1 is present at a high level, an open question is that of the exact amount of UCP2 and UCP3 in cells since whatever is the exact intrinsic ability of these proteins to uncouple respiration from ATP synthesis, a certain amount of the proteins will be required to achieve a physiological uncoupling. Finally, the possible contribution of a UCP to the hypermetabolic syndrome (known as Luft's disease) previously described in patients remains to be investigated.

cofactors of the proton transport whereas nucleotides are inhibitors. The membrane potential itself is a variable influencing heavily UCP1 activity. It has been proposed that superoxide anion and quinones directly activate proton transport by UCPs.

See also: **Bioenergetics:** ATP in Plant Mitochondria: Substrates, Inhibitors, and Uncouplers; Luft's Disease; Mitochondrial Membranes, Structural Organization; Quinones; Respiratory Chain and Adenosine Triphosphate Synthase.

Further Reading

- Boss O, Hagen T, and Lowell BB (2000) Uncoupling proteins 2 and 3. Potential regulators of mitochondrial energy metabolism. *Diabetes* 49: 143–156.
- Cannon B and Nedergaard J (1985) The biochemistry of an inefficient tissue: Brown adipose tissue. *Essays in Biochemistry* 20: 110–164.
- Echtay KS, Winkler E, and Klingenberg M (2000) Coenzyme Q is an obligatory cofactor for uncoupling protein function. *Nature* 408: 609–613.
- Enerbäck S, Jacobsson A, Simpson EM, et al. (1997) Mice lacking mitochondrial uncoupling protein are cold-sensitive but not obese. *Nature* 387: 90–94.
- Garlid KD and Jaburek M (1998) The mechanism of proton transport mediated by mitochondrial uncoupling proteins. *FEBS Letters* 438: 10–14.
- Himms-Hagen J and Ricquier D (1998) Brown adipose tissue. In: Bray G, Bouchard C, and James WPT (eds.) *Handbook of Obesity*, pp. 415–441. New York, NY: Marcel Dekker.
- Ledesma M, García de Lacoba M, and Rial E (2002) The mitochondrial uncoupling proteins. *Genome Biology* 3(12): Epub November 29, 2002.
- Nedergaard J, Bengtsson T, and Cannon B (2007) Unexpected evidence for active brown adipose tissue in adult humans. *American Journal of Physiology* 293: E444–E452.
- Nicholls DG and Locke RM (1984) Thermogenic mechanisms in brown fat. *Physiological Reviews* 64: 1–64.
- Pecqueur C, Alves-Guerra C, Ricquier D, and Bouillaud F (2009) UCP2, a metabolic sensor coupling glucose oxidation to mitochondrial metabolism? *IUBMB Life* 61: 762–767.
- Ricquier D and Bouillaud F (2000) The uncoupling protein homologues: UCP1, UCP2, UCP3, StUCP & AtUCP. *Biochemical Journal* 345(Pt 2): 161–719.
- Skulachev VP (1988) Uncoupling: New approaches to an old problem of bioenergetics. *Biochimica et Biophysica Acta* 1363: 100–124.
- Stuart JA, Cadenas S, Jekabsons MB, Roussel D, and Brand MD (2001) Mitochondrial proton leak and the uncoupling protein 1 homologues. *Biochimica et Biophysica Acta* 1504: 144–158.

Unfolded Protein Responses

D Ron, University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Cambridge, UK

© 2013 Elsevier Inc. All rights reserved.

Glossary

Chaperone A protein that assists folding and assembly of other proteins and prevents their aggregation in the unfolded state.

Endoplasmic reticulum A membrane-bound organelle, present in all nucleated cells, which serves as the entry point for secreted proteins.

Proteotoxicity A hypothetical property of unfolded and misfolded proteins, whereby peptide domains that are illegitimately exposed on the surface perturb cellular function.

The unfolded protein response (UPR) is a transcriptional and translational response to the accumulation of unfolded proteins in the endoplasmic reticulum (ER). The UPR is mediated by highly conserved signaling pathways that are activated by imbalance between the load of unfolded (or misfolded) ER client proteins and the capacity of the organelle to process this load. Collectively, these pathways restore equilibrium to the protein-folding environment in the organelle by increasing the expression of genes that enhance nearly all aspects of ER function and by transiently repressing the biosynthesis of new client proteins. Interest in the UPR has been stimulated by the realization that postsynthetic protein-processing constitutes an important step in gene expression and that protein misfolding plays an important role in human disease.

ER Function and ER Stress

The ER: A Protein-Processing Machine

Proteins destined for secretion and membrane insertion are translocated across the ER membrane in an unfolded state. In the ER lumen, they undergo chaperone-assisted folding, a variety of organelle-specific posttranslational covalent modifications and often chaperone-assisted assembly into oligomeric structures. Once properly folded and assembled, most ER client proteins are packaged into vesicles that are transported to more distal sites in the secretory pathway. Proteins that fail to attain their proper folded and oligomeric conformation are retained in the ER by continued binding to chaperones and are ultimately translocated from the organelle to the cytoplasm for proteasomal degradation, in a process known as ER-associated protein degradation.

To fulfill these various functions, the ER is endowed with a unique complement of proteins. These include enzymes for posttranslational covalent modifications (e.g., N-linked glycosylation and disulfide bond formation), chaperones that assist in folding and assembly steps, translocation channels and transporters involved in transmembrane traffic, and various components involved in the postassembly steps of client protein egress from the organelle. The buildup of these components defines the capacity of the organelle to handle its client proteins. In the late 1980s, Sambrook, Gething, and

their colleagues discovered that manipulations that interfere with the function of various aspects of the ER client protein-handling machinery, and thereby perturb protein folding in the ER, selectively upregulate the expression of genes that encode components of that machinery. The extent of this transcriptional response was not fully appreciated at the time; however, its selectivity was noted, hence its name: the ER UPR.

It was imagined that the cell possessed means to monitor the load of client proteins presented to its ER and responded to an increase in such load by upregulating the capacity of its ER to process unfolded client proteins. An additional clue to the workings of this response was provided by the seminal observation that forced overexpression of an ER chaperone, BiP (also known as GRP78) markedly suppressed the activity of the UPR. Thus, at least one target gene of the UPR (BiP) is able to exert negative feedback on the entire response. BiP overexpression does not restore function to a challenged ER; this requires the coordinate expression of numerous UPR target genes. However, BiP, which is a member of the highly conserved HSP70/DnaJ family of chaperones, promiscuously recognizes hydrophobic stretches of amino acids. These are normally incorporated into the cores of properly folded polypeptides and assembled oligomeric complexes, but remain exposed on the surface of unfolded, misfolded, or unassembled proteins. The ability of BiP to suppress the UPR suggested that it might be doing so by masking a stress signal generated by many different unfolded and misfolded proteins. This phenomenon was assigned the heuristic term ER stress and its level reflects the balance between client protein load and the capacity of the organelle to process that load.

Physiological and Pathological ER Stress

To the extent that cell types vary in the load of secreted proteins that they are called upon to produce, they are subject to widely different levels of physiological ER stress. This explains enhanced activity of the UPR in various professional secretory cells, such as pancreatic cells of vertebrates or intestinal cells of the nematode, *Caenorhabditis elegans*. ER stress also occurs when a mutation in an abundantly expressed ER client protein renders that protein especially difficult to fold. An example is

provided by various degenerative diseases affecting myelinated neurons in which mutations in a component of the myelin sheath (an ER client protein) cause the protein to misfold and induce high levels of ER stress which, over time, destroys the myelin-producing cell. It is important to emphasize that most mutations that impede folding of ER client proteins do not cause measurable ER stress. However, because they diminish expression of the properly folded protein, such mutations may deprive the organism of the latter's beneficial actions. This genetic mechanism underlies such serious human diseases as cystic fibrosis, familial hypercholesterolemia, and hemophilia. Nonetheless, the level of expression of the mutant protein is not enough to globally challenge ER function in the cell that produces it and the phenotypic expression of the mutation reflects the lack of an important protein, rather than the production of a toxic one.

The above constitute client protein-driven ER stress; however, ER stress may also initiate from impaired function of the organelle, which occurs in cells deprived of energy sources or oxygen, or in cells exposed to certain ER-specific toxins. We do not understand, in detail, how ER stress contributes to further organelle dysfunction and ultimately cell death. However, a framework for thinking about this has recently emerged with the realization that unfolded and misfolded proteins may disrupt the cellular machinery by interacting promiscuously and illegitimately with essential cellular components. According to this theory, proteotoxicity is normally held in check by the chaperones that bind such potentially toxic protein interfaces and let go only once the latter have been buried in the hydrophobic cores of the properly folded client protein. ER stress (by definition) challenges the capacity of the chaperones and may permit illegitimate protein interfaces to emerge. The ability of chaperone overexpression to suppress ER stress signaling suggests that the need to prevent proteotoxicity is an important driving force in evolution of the UPR. The unifying feature of ER stress need not be the presence of toxic moieties on the surface of every unfolded or misfolded protein. The common feature of diseases of protein folding might instead be the exhaustion of a protective chaperone reserve, which normally suppresses the potential proteotoxicity of certain (possibly normal) folding intermediates of ER client proteins.

The Yeast UPR

Inositol Requiring 1: A Prototype of Transmembrane Stress Signaling

Early studies on the UPR were carried out in the yeast, *Saccharomyces cerevisiae*, an organism that lends itself well to forward genetic screens. To screen for mutations affecting the UPR, the regulatory region of yeast BiP gene (KAR2) was fused to a reporter and mutant yeast with suppressed activity of this reporter was sought. The first gene, thus, identified was inositol requiring 1 (IRE1), so-called because its loss of function had been previously noted to result in inositol auxotrophy. IRE1 encodes a transmembrane ER-resident protein with an N-terminal domain residing in the ER lumen and a C-terminal domain that is exposed on the cytoplasmic side.

The membrane topology and subcellular localization of IRE1 immediately suggest a mechanism for transmitting information on the state of the ER (topologically equivalent to the extracellular space) to the cell's interior. The luminal domain somehow senses ER stress, conveying the signal across the ER membrane to the cytoplasmic domain, which broadcasts it to the nucleus, turning 'on' UPR target gene expression (Figure 1).

IRE1's C-terminal, cytoplasmic, effector domain, is a protein kinase and undergoes autophosphorylation when activated by ER stress. This suggested that IRE1 might function like other transmembrane receptors that are also protein kinases and convey their signal by phosphorylating downstream targets, often through a kinase relay. However, other than IRE1 itself no substrates for the aforementioned kinase have been identified to date.

Unconventional Splicing of HAC1 Messenger RNA

The clues to understanding propagation of the UPR signal, downstream of IRE1 again came from yeast genetics. Mutations in two additional genes were noted to block the yeast UPR. One of these, HAC1, encodes a transcription factor, which binds to and activates the promoters of the primary target genes of the yeast UPR, the second, more mysteriously turned out to be RLG1, which encodes a multifunctional enzyme involved in transfer RNA (tRNA) splicing. ER stress promotes accumulation of HAC1 protein and this is blocked by mutations in IRE1 and RLG1. The underlying mechanisms, which turned out to be full of surprise, were revealed in a series of brilliant experiments carried out in the Walter and Mori laboratories.

The regulated step in HAC1 expression entails the removal of a 242-base internal segment at the 3' end of the mature messenger RNA (mRNA). The enzyme which precisely cuts the mRNA at the 5' and 3' ends of this segment turned out to be IRE1 itself, whose highly sequence-specific endoribonucleolytic activity is subordinate to its kinase activity by phosphorylation-dependent nucleotide binding and dimerization of IRE1's composite kinase and RNase domain. The two ends of the severed HAC1 mRNA are held together by extensive base pairing and are subsequently religated by the protein product of RLG1. IRE1 is the master regulator of this unconventional mRNA-splicing reaction, because its kinase and endoribonuclease are activated by ER stress. The removal of the aforementioned internal segment of the HAC1 mRNA derepresses HAC1 translation and the protein encoded by the processed mRNA is more stable and transactivates its target genes more potently than that encoded by the unprocessed mRNA.

Mutations in IRE1 and HAC1 similarly abolish most signaling in the yeast UPR. This was convincingly demonstrated by expression profiling which showed that the vast majority of the genes induced by exposing yeast to toxins that promote ER stress were no longer activated in IRE1 or HAC1 mutant strains. Thus, the yeast UPR consists of a linear pathway in which HAC1 is an essential target of IRE1 and together the two proteins are required for the activation of most known UPR target genes.

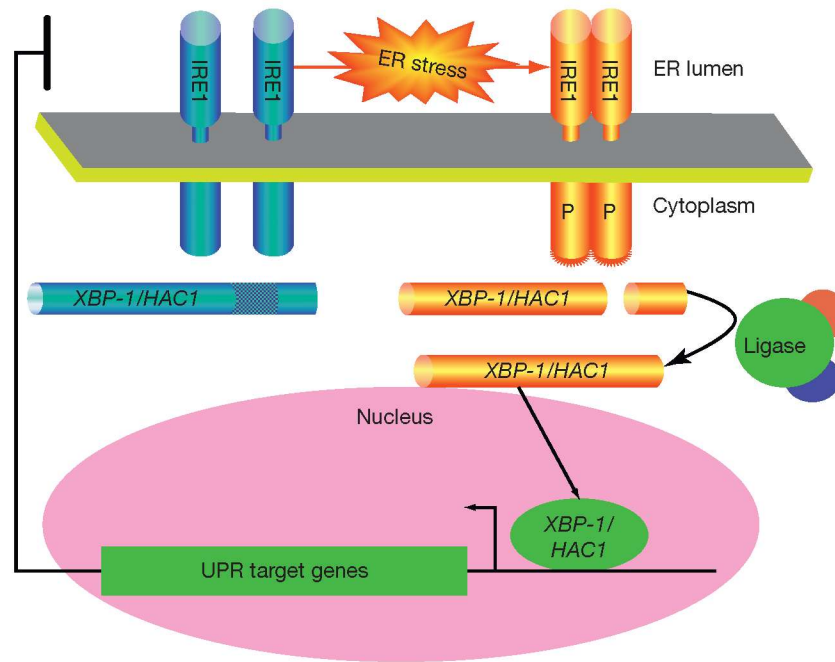


Figure 1 IRE1 and unconventional splicing of its target mRNA signal the unfolded protein response. ER stress leads to oligomerization and *trans*-autophosphorylation of IRE1 (indicated by the 'P' on its cytoplasmic effector domain). Phosphorylation unmasks the effector function of IRE1 (cartooned by the red bristles on the cytoplasmic domain), which consists of the endoribonucleolytic processing of its target mRNA, XBP-1 in metazoans, and HAC1 in yeast. The two ends of the cleaved mRNA are joined together by a ligase. The unconventionally spliced XBP-1/HAC1 mRNA is more efficiently translated than the unprocessed mRNA and encodes a protein that is more stable and more effective at *trans*-activation of UPR target genes in the nucleus.

Diversification in the Metazoan UPR

Conservation of the IRE1 Pathway

IRE1 is conserved in all eukaryotes as is its ability to cleave a unique substrate RNA. The counterpart of HAC1 in metazoans is the XBP1 mRNA whose unconventional splicing mechanism is conserved in many (though not all) details between yeast and mammals. Less conserved, however, is the role of the IRE1 pathway in the yeast and metazoan UPR. Whereas in the former, IRE1 and HAC1 are absolutely required for activating nearly all UPR target genes; in mammals, IRE1 and XBP-1 are dispensable for activation of all but a very small set of UPR targets. This is all the more remarkable when one considers that most UPR target genes are conserved between yeast and mammals, it is just that the latter have found alternative means to couple their expression to ER stress. *C. elegans* occupies an intermediate position between yeast and mammals, in that most of its identifiable UPR targets are partially dependent on IRE1 and XBP-1, but can, to some extent, also be activated independently of that pathway.

Despite this redundancy in the mammalian UPR, both IRE1 and XBP-1 are essential for embryonic development (mammals have two IRE1 genes, a nonessential beta isoform restricted in its expression to the intestinal and bronchial epithelium and a broadly expressed alpha isoform which is essential). The reason(s) for the embryonic lethality of IRE1 α and XBP-1-null animals are not fully understood, but it is hypothesized that in mammals the pathway may have diverted to specify the capacity for especially high levels of protein

secretion. This hypothesis, which is clearly in need of further experimental support, suggests that specification of a secretory cell fate activates a conserved stress pathway that drives a developmental process, namely acquisition of an apparatus required by professional secretory cells for high-capacity protein secretion. This idea is further supported by the observation that many UPR target genes function far downstream in the secretory pathway rendering it unlikely that such genes merely relieve the stressed ER of its load of unfolded client proteins.

ATF6 and Intramembrane Proteolysis in the Metazoan UPR

The identification of ATF6 by Mori and colleagues confirmed the predicted redundancy in the metazoan UPR. Like IRE1, ATF6 is also an ER-localized transmembrane protein. However, its cytoplasmic effector domain consists of a transcription factor that activates UPR target genes directly. In its membrane-bound form, ATF6 is inert, as it cannot reach the nucleus. Activation involves regulated intramembrane proteolysis, which liberates the transcription factor part of ATF6 from the ER membrane under conditions of ER stress (Figure 2). The next surprise came when Brown, Goldstein, Prywes, and colleagues identified the proteases involved in this highly regulated event. These turned out to be the same proteases that process and activate the membrane-bound transcription factor, SREBP, which regulates genes involved in sterol and fatty acid biosynthesis and assimilation.

There are two ATF6 genes in mammals and knockout of both isoforms is lethal. Furthermore, cells compromised in

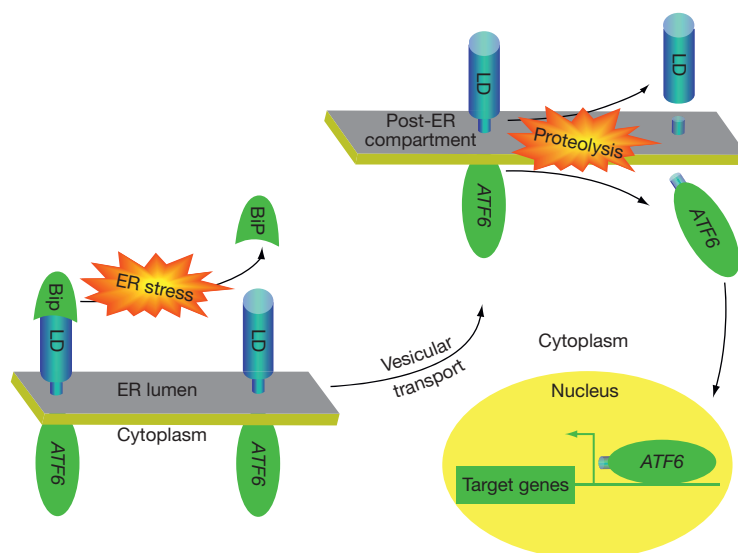


Figure 2 ATF6 activation by regulated intramembrane proteolysis. Membrane tethering inactivates the transcription factor ATF6, whose effector domain is prevented from accessing its target genes in the nucleus. BiP binding to the luminal domain of ATF6 retains the protein in the ER compartment. However, under conditions of ER stress, BiP dissociates from ATF6, which then migrates (presumably by vesicular transport) to a post-ER compartment that contains the proteases that liberate the effector domain to signal in the UPR.

ATF6 function have important defects in UPR target gene expression. It seems likely, therefore, that mammalian ATF6 has picked up some of the role performed by IRE1 and XBP-1/HAC1 in simpler organisms. The similarities between ATF6 and the SREBPs highlight another remarkable feature of the UPR. IRE1 and HAC1 mutant yeast are unable to synthesize adequate amounts of the membrane phospholipid precursor, inositol. Although the details remain to be worked out, insufficiency of membrane lipid components also signals through the yeast UPR and enzymes involved in phospholipid metabolism are targets of IRE1 and HAC1. In metazoans, this aspect of the UPR apparently has been split off and relegated to dedicated transcription factors, the SREBPs that are activated by insufficiency of lipid components. However, the ancient link between the client protein-based UPR and membrane lipid insufficiency signaling has remained in the form of shared machinery for proteolytic activation of ATF6 and SREBP.

Translational Control in the UPR

PERK Couples ER Stress to eIF2 α Phosphorylation and Translational Repression

In addition to activated gene expression, ER stress results in a reduction in protein synthesis. Translational repression is an active process limiting the influx of client proteins into the stressed ER and, thus, serves as a counterpart to the gene-expression program, which increases the organelle's capacity to process client proteins. Brostrom and colleagues and Kaufman and colleagues noted, early on, that translational repression by ER stress is associated with phosphorylation of the α -subunit of translation initiation factor 2 (eIF2 α) on serine 51. This phosphorylation site is conserved in all eukaryotes and serves to regulate translation initiation in diverse stressful conditions. Trimeric eIF2, in complex with guanosine

triphosphate (GTP), recruits the amino-acylated initiator methionyl-tRNA to the small ribosomal subunit, allowing translation initiation. Recognition of an AUG initiation codon on the mRNA leads to hydrolysis of GTP to guanosine diphosphate (GDP) and dissociation of the ribosome-eIF2 complex. To participate in another round of translation initiation, the GDP bound to eIF2 must be exchanged to GTP. The enzyme catalyzing this exchange reaction (eIF2B) is inhibited by phosphorylated eIF2.

Distinct eIF2 α kinases were known to be activated by distinct stress signals; PKR by double-stranded RNA in viral infection and GCN2 by uncharged tRNAs in amino acid starvation. Our group noted the existence of a predicted transmembrane protein in the *C. elegans* genome, with an N-terminal 'extracellular' domain similar to IRE1 and an intracellular domain similar to PKR and GCN2, known eIF2 α kinases. This protein, which is conserved in metazoans (and absent from yeast) and to which we gave the name PERK (PKR-like ER Kinase), couples ER stress to eIF2 α phosphorylation and is essential to translational regulation by ER stress. Cells lacking PERK are hypersensitive to ER stress and in mammals, PERK activity is especially important for preservation of cells engaged in high levels of secretion. Thus, humans with PERK mutations and mice lacking the gene, develop diabetes mellitus and skeletal defects attributed to dysfunction of insulin- and collagen-secreting cells in the endocrine pancreas and bone, respectively.

Gene Expression, eIF2 α Phosphorylation and Integration of Signals in the UPR, and Other Stress Pathways

In addition to limiting the load of client protein on the stressed ER (by inhibiting translation of most mRNA), eIF2 α phosphorylation paradoxically activates the translation of the mRNA encoding the transcription factor ATF4 (and likely that of other regulatory proteins that remain to be discovered,

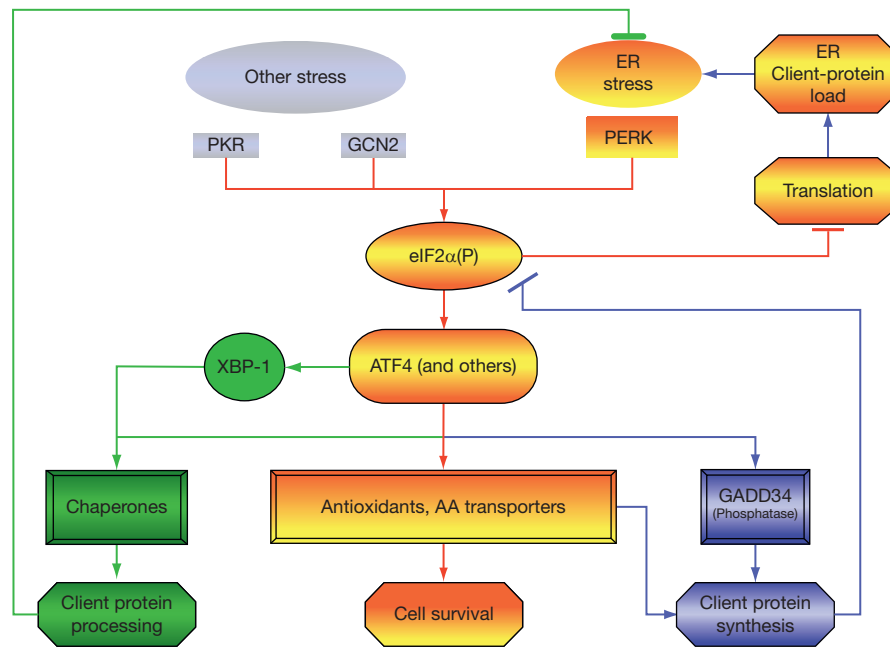


Figure 3 The eIF2 α phosphorylation-dependent integrated stress response. ER stress and other stresses activate eIF2 α kinases. Phosphorylated eIF2 inhibits translation initiation, alleviating ER client protein load, while at the same time activating the transcription factor ATF4 (and other, yet to be identified effectors). These activate a gene-expression program with several components: the eIF2 α phosphatase GADD34 promotes translational recovery and resumed client protein synthesis. Genes involved in antioxidant responses and amino acid import and assimilation promote cell survival and client protein synthesis. PERK signaling also functions synergistically with other arms of the UPR (e.g., by transcriptional activation of the IRE1 target gene XBP-1) to enhance ER client protein processing.

Figure 3). PERK, thus, plays an important role in upregulating target gene expression in the UPR. Expression profiling of ER-stressed wild-type and PERK $^{-/-}$ cells reveals a broad role for PERK in UPR-target gene expression, with mild impairment in most target genes and severe impairment in a smaller group. Among the UPR target genes that are strongly impaired in PERK $^{-/-}$ cells are genes involved in amino acid transport and metabolism and a gene, *GADD34*, which encodes a phosphatase that dephosphorylates eIF2 α and terminates signaling by PERK. The aforementioned strong PERK target genes play an important role in maintaining translation, by providing building blocks for secreted protein synthesis and by promoting eIF2 α dephosphorylation. Thus, PERK and eIF2 α phosphorylation appear to have a dual role in the UPR. Early in the response they protect the stressed cells from ER overload, whereas later they promote conditions required for synthesis of secreted proteins.

A considerable overlap exists between genes activated by ER stress and genes activated by other stressful conditions that promote eIF2 α phosphorylation. Thus, eIF2 α phosphorylation integrates signaling in several stress pathways. The extent of this integrated stress response is not fully known, but it appears to control important aspects of cellular and organismal metabolism.

A Common Mechanism for Sensing ER Stress

The identification of three different stress sensors in the UPR has allowed a comparison between their mode of activation. As mentioned earlier, overexpression of BiP and other ER

chaperones had been noted to repress signaling in the UPR. In the equilibrated ER, all three stress transducers, IRE1, ATF6, and PERK, are found in a simple complex with BiP, which binds on the luminal side to their stress-sensing domains. These complexes are rapidly dissociated under conditions of ER stress. Furthermore, dissociation precedes activation of the stress transducers. BiP dissociation promotes oligomerization of IRE1 and PERK, which, in turn, leads to *trans*-autophosphorylation and activation of downstream signaling. The crystal structure of yeast and human IRE1 luminal domain supports the prominence of dimerization in its activation and also suggest the possibility that direct binding of unfolded proteins may contribute to the stability of the dimer.

The chain of events in the case of ATF6 activation is more complex, but conceptually similar. BiP binding retains ATF6 in the ER, segregated from the proteases that release its transcription factor portion, as these are localized to a post-ER compartment. BiP dissociation liberates ATF6 to migrate to the protease-containing compartment, a migration that presumably takes place by vesicular transport. Thus, dispensable molecules of BiP, which are not engaged in chaperoning ER client proteins actively repress signaling by all three known transducers of the UPR. It would, thus, appear that the recognition of unfolded and misfolded proteins is relegated to professional chaperones and the stress transducers passively monitor the degree to which the latter are engaged by their clients. This arrangement of the upstream steps of the UPR mirrors that of the cytoplasmic heat shock response, in which unfolded proteins in the cytoplasm activate genes implicated in folding and degrading cytoplasmic proteins.

See also: **Cell Architecture and Function:** Endoplasmic Reticulum-Associated Protein Degradation; **Protein/Enzyme Structure Function and Degradation:** Chaperonins; Regulated Intramembrane Proteolysis.

Further Reading

Lee AH and Glimcher LH (2009) Intersection of the unfolded protein response and hepatic lipid metabolism. *Cellular and Molecular Life Sciences* 66: 2835–2850.

Ron D and Harding HP (2007) eIF2 α phosphorylation in cellular stress-responses and disease. In: Hershey JWB, Matthews MB, and Sonenberg N (eds.) *Translational Control in Biology and Medicine*, ch. 13, pp. 345–368. New York, NY: Cold Spring Harbor Press.

Ron D and Walter P (2007) Signal integration in the endoplasmic reticulum unfolded protein response. *Nature Reviews Molecular Cell Biology* 8: 519–529.

Schroder M and Kaufman RJ (2005) The mammalian unfolded protein response. *Annual Review of Biochemistry* 74: 739–789.

Urea Cycle: Disease Aspects

C Ficicioglu and M Yudkoff, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Hyperammonemia The accumulation of ammonia in the blood to an untoward concentration; the result of a congenital deficiency of a constituent enzyme of the urea cycle.

Urea The major end product of the breakdown of proteins and amino acids.

Urea cycle The sequence of reactions by which the liver forms urea from ammonia.

Urea cycle defects Inherited deficiencies of one of the enzymes of the urea cycle.

The breakdown of proteins and other nitrogen-containing molecules results in the production of free ammonia. High ammonia concentrations are toxic to the brain. The primary function of the urea cycle (Figure 1) is to detoxify ammonia by converting it to urea. The urea cycle includes five enzymes: (1) carbamyl phosphate synthetase (CPS); (2) ornithine transcarbamylase (OTC); (3) argininosuccinate synthetase (AS); (4) 4-argininosuccinate lyase (AL); and (5) arginase. In addition to these constitutive enzymes, *N*-acetylglutamate synthetase is needed to produce *N*-acetylglutamate, an obligatory effector of the CPS enzyme. With the exception of OTC deficiency, which is X-linked, all have an autosomal recessive mode of inheritance.

Although individual deficiencies of these enzymes are rare, the overall prevalence of urea cycle disorders is about 1/30 000 live births. AS, AL, and arginase enzyme deficiencies can be detected using tandem mass spectrometry in newborn screening. Recently, some newborn screening programs have included OTC deficiency in their screening panel, using the glutamine/citrulline ratio as a marker of disease.

Inborn Errors of Metabolism Causing Hyperammonemia

Urea cycle disorders are the most common cause of hyperammonemia, but other inborn errors of metabolism, especially organic acidemias, can cause a comparable derangement (Table 1); the pathogenesis for hyperammonemia in some of these conditions is not fully understood.

Clinical Manifestations of Hyperammonemia

Most neonates present in a characteristic manner, regardless of genetic etiology. The affected infant is normal at birth but presents with a sepsis-like picture by 24–48 h of life. The baby ceases feeding, begins vomiting, and displays lethargy that quickly progresses to coma. Physical examination may reveal hepatomegaly, in addition to the neurologic signs of coma. Severe cases manifest increased intracranial pressure with a bulging fontanel and dilated pupils. Seizures are common.

Infants and older children with acute hyperammonemia may present with vomiting and neurologic abnormalities such as ataxia, mental confusion, agitation, irritability, and

combativeness. These manifestations usually occur during a high-protein intake or catabolic stress such as infections.

Routine laboratory tests are often normal. The blood urea tends to be very low. Hyperammonemia may stimulate the brain-stem respiratory center, leading to respiratory alkalosis. There may be mild increases in serum transaminases (alanine transaminase (ALT) and aspartate aminotransferase (AST)), since ammonia can cause swelling of hepatic mitochondria. In infants with organic acidemias, hyperammonemia is commonly associated with high anion gap acidosis and ketonuria. Newborn infants with hyperammonemia are often misdiagnosed as having sepsis and they may die without a correct diagnosis. Brain ultrasound or computed tomography (CT) imaging may reveal edema. There are no specific autopsy findings that would allow the diagnosis of urea cycle disorder or organic acidemias. It is imperative to measure plasma ammonia, amino acids, acylcarnitine levels, and urine organic acid levels in any ill infant whose clinical manifestations cannot be explained by sepsis.

Diagnosis

Hyperammonemia is common to all disorders of the urea cycle. Unfortunately, the measurement of ammonia can be difficult with false elevations noted frequently. Blood should be obtained from a free-flowing source (usually a vein) and rapidly chilled to avoid artificial production from blood cells. Each clinical laboratory should establish its own normal values for blood ammonia; some variation is common. In children and adults, the plasma ammonia concentration is usually less than 35 $\mu\text{mol l}^{-1}$. Blood concentrations in healthy newborn infants often can be as high as 100 $\mu\text{mol l}^{-1}$ in a full-term baby and 150 $\mu\text{mol l}^{-1}$ in a premature infant. The plasma ammonia concentration in the ill infant is usually more than 200 $\mu\text{mol l}^{-1}$.

An approach to the differential diagnosis of hyperammonemia in the newborn infant is illustrated in Figure 2. The age of onset of symptoms and the presence or absence of acidosis help physicians to arrive at a diagnosis of hyperammonemia. Plasma amino acids reveal abnormalities that may help the diagnosis of different urea cycle disorders. In patients with deficiencies of CPS, OTC, or *N*-acetylglutamate (NAG) synthetase, frequent findings are elevations in plasma glutamine and alanine with concurrent decrements in citrulline and arginine.

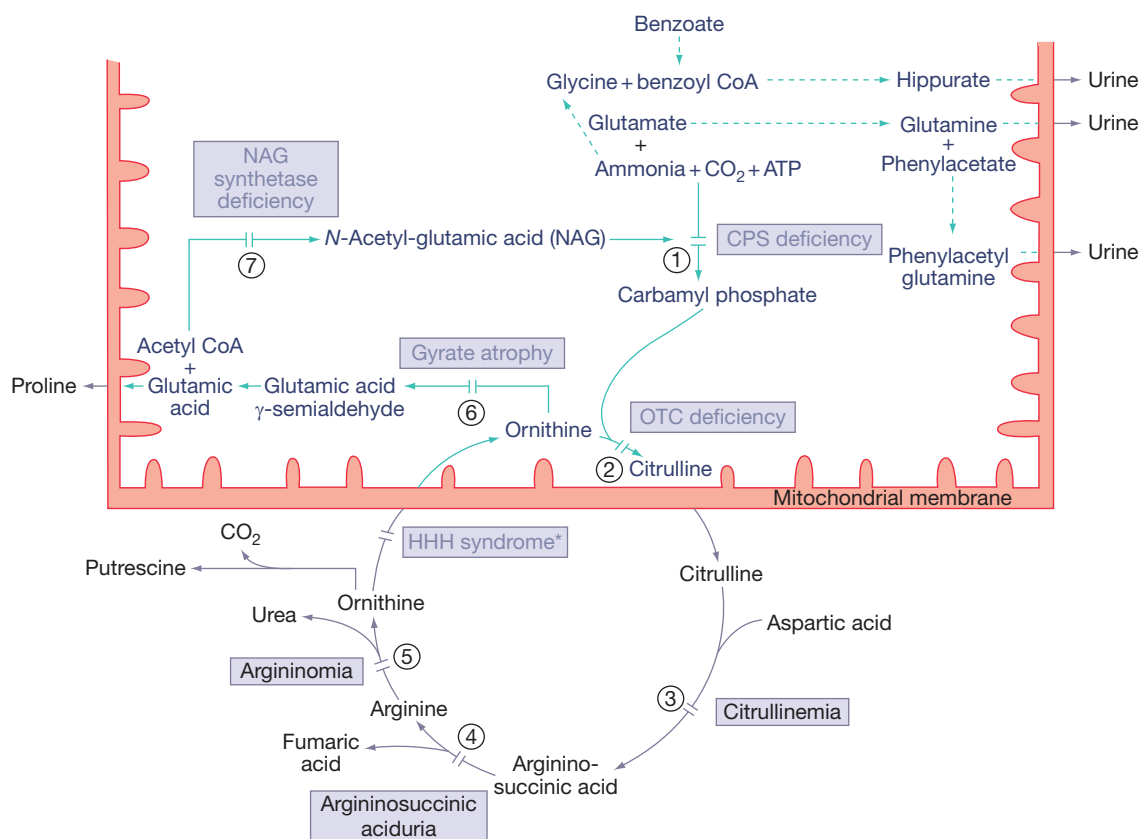


Figure 1 The urea cycle. Reactions occurring in the mitochondria are depicted in purple. Reactions shown with interrupted arrows are the alternate pathways for the disposal of ammonia. Enzymes: (1) carbamyl phosphate synthetase (CPS), (2) ornithine transcarbamylase (OTC), (3) argininosuccinic acid synthetase (AS), (4) argininosuccinic acid lyase (AL), (5) arginase, (6) ornithine 5-aminotransferase, and (7) *N*-acetylglutamate (NAG) synthetase. * HHH syndrome, hyperammonemia–hyperornithinemia–homocitrullinemia.

Table 1 Inborn errors of metabolism causing hyperammonemia

Urea cycle disorders

- *N*-Acetylglutamate synthetase
- Carbamyl phosphate synthetase (CPS)
- Ornithine transcarbamylase (OTC)
- Argininosuccinate synthetase (AS)
- Argininosuccinate lyase (AL)
- Arginase

Organic acidemias

- Propionic acidemia
- Methylmalonic acidemia
- Multiple carboxylase deficiency
- 3-Hydroxy-3-methylglutaric aciduria

Fatty acid oxidation defects

- Medium-chain acyl-CoA dehydrogenase deficiency
- Multiple acyl-CoA dehydrogenase deficiency

Other inborn errors of metabolism

- Lysinuric protein intolerance
- Hyperammonemia–hyperornithinemia–homocitrullinemia (HHH) syndrome
- Pyruvate carboxylase deficiency
- Congenital hyperinsulinism with hyperammonemia
- Transient hyperammonemia of the newborn

These disorders cannot be differentiated from one another on the basis of the plasma amino acid levels alone. A marked increase in urinary orotic acid in patients with OTC deficiency differentiates this defect from CPS deficiency. Differentiation between mutations in CPS versus NAG synthetase may require an assay of the respective enzymes. Clinical improvement occurring after oral administration of carbamylglutamate, however, may suggest NAG synthetase deficiency. Patients with deficiencies of AS, AL, or arginase have marked increases in the plasma levels of citrulline, argininosuccinic acid, or arginine, respectively. Indeed, the combination of hyperammonemia and marked hypercitrullinemia or argininosuccinic acidemia is virtually pathognomonic for these disorders. Elevated plasma levels of arginine point to the diagnosis of arginase deficiency.

Management of Acute Hyperammonemia

Patients with hyperammonemia should be treated promptly and vigorously. The severity of neurologic sequelae is dependent on the duration of hyperammonemia. Serious neurologic damage is very likely in infants with severe elevations in blood ammonia ($>300 \mu\text{mol l}^{-1}$) for more than 12 h.

Treatment lowers the blood ammonia level in two ways: (1) removal of ammonia N in a form other than urea and

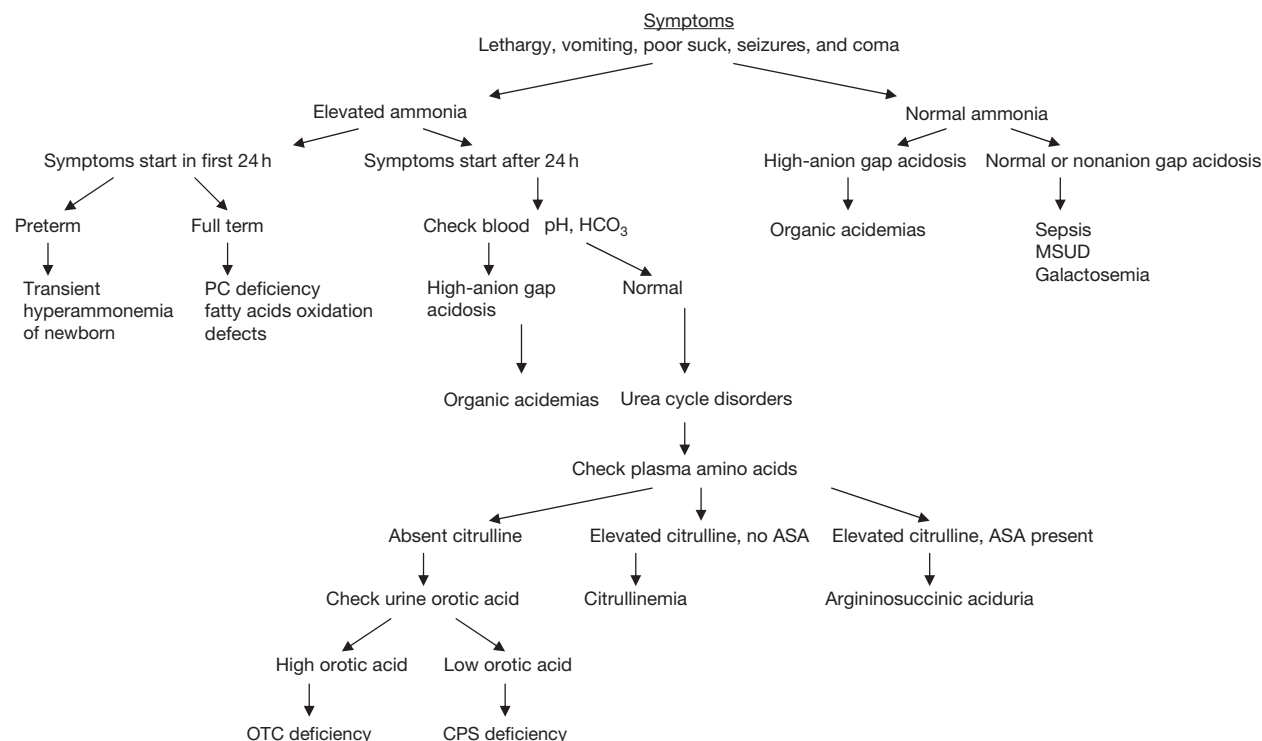


Figure 2 Clinical approach to a newborn infant with symptomatic hyperammonemia.

(2) minimizing tissue protein catabolism (and, hence, ammonia production) by providing adequate calories and essential amino acids. Initially, fluids, electrolytes, glucose (10–20%), and lipids ($1\text{--}3\text{ g kg}^{-1}\text{ 24 h}^{-1}$) should be infused intravenously together. Once it is clear that blood ammonia is declining, it may be possible to add minimal amounts of protein ($0.25\text{ g kg}^{-1}\text{ d}^{-1}$), preferably as essential amino acids, to the infusate. As soon as the clinical condition of the patient allows, oral or gastric feeding with a low-protein formula ($0.5\text{--}1.0\text{ g kg}^{-1}\text{ 24 h}^{-1}$) should be started. It is essential to restrict dietary protein to the amount tolerated without causing hyperammonemia. Some authorities recommend a special diet that provides 50% of prescribed protein as essential amino acids.

Compounds such as sodium benzoate and phenylacetate lower blood ammonia by forming complexes with glycine and glutamine, respectively. The resulting conjugates are quickly excreted into urine, thereby eliminating waste N. Benzoate reacts with glycine to yield hippurate and phenylacetate conjugates with glutamine to produce phenylacetylglutamine (Figure 1). Thus, administration of phenylacetate leads to removal of 2 mol of ammonia N. Benzoate and phenylacetate now are available for both oral and intravenous use.

Arginine is a nonessential amino acid in healthy individuals, who form it in the urea cycle, but it is essential in most patients with urea cycle disorders (except arginase deficiency), in whom supplementation supplies the urea cycle with ornithine and NAG (see Figure 1). In patients with citrullinemia, 1 mol of arginine (as ornithine) reacts with 1 mol of ammonia (as carbamyl phosphate) to form citrulline. In patients with argininosuccinic acidemia, 2 mol of ammonia (as carbamyl phosphate and aspartate) react with citrulline to form argininosuccinic acid.

Citrulline and argininosuccinic acid are far less toxic and more readily excreted by the kidneys than is ammonia. Patients with OTC deficiency benefit from supplementation with citrulline ($150\text{--}200\text{ mg kg}^{-1}\text{ 24 h}^{-1}$) because 1 mol of citrulline reacts with 1 mol of ammonia (as aspartic acid) to form arginine. Administration of arginine or citrulline is contraindicated in patients with arginase deficiency. Arginase deficiency is a rare condition in which acute hyperammonemia rarely occurs as a presenting sign. In patients whose hyperammonemia is secondary to organic acidemias, treatment with arginine is not indicated because no beneficial effect from such therapy can be expected. In a newborn infant with hyperammonemia, arginine should be used until the diagnosis of urea cycle disorders is established.

Sodium benzoate, sodium phenylacetate, and arginine can be administered together for maximal therapeutic effect. A priming dose of these compounds is followed by continuous infusion until recovery from the acute hyperammonemic crisis (see Table 2). Both benzoate and phenylacetate are usually supplied as concentrated solutions and should be diluted in 10% glucose for intravenous use. The recommended therapeutic doses of both compounds deliver a substantial amount of sodium to the patient that should be calculated as part of the daily sodium requirement. A commercial, parenteral preparation of sodium benzoate plus sodium phenylacetate is available. Benzoate and phenylacetate should be used with caution in newborns with hyperbilirubinemia because they may displace bilirubin from albumin, although no documented case of kernicterus has yet been reported in infants so treated.

Hemodialysis should be started as an emergency procedure if the plasma ammonia level does not significantly decrease

Table 2 Treatment of acute hyperammonemia in an infant

Drug	Loading dose/infused over 1–2 h	Maintenance dose/continuous infusion–24 h
Sodium benzoate ^a	250 mg kg ⁻¹ (5.5 g m ⁻²)	250 mg kg ⁻¹ d ⁻¹ (5.5 g m ⁻²)
Sodium phenylacetate ^a	250 mg kg ⁻¹ (5.5 g m ⁻²)	250 mg kg ⁻¹ d ⁻¹ (5.5 g m ⁻²)
Arginine hydrochloride ^b (10%)	210–660 mg kg ⁻¹ (4.0–12.0 g m ⁻²)	210–660 mg kg ⁻¹ d ⁻¹ 4.0–12.0 g m ⁻²)

^aThese compounds are usually prepared as a 1–2% solution (10% glucose) for intravenous use; sodium from these drugs should be included as part of the daily sodium requirement.

^bThe higher dose is recommended in the treatment of patients with citrullinemia and argininosuccinic aciduria; arginine is not recommended in patients with arginase deficiency and in those for whom hyperammonemia is secondary to organic acidemia.

within 8 h. Exchange transfusion or peritoneal dialysis has little effect (ammonia clearance is about 10% of that of hemodialysis). Although hemodialysis clearly is the most effective measure, it is technically difficult to perform and may not be readily available. There is usually a dramatic decrease in the plasma ammonia level within a few hours of hemodialysis, and, in most patients, the plasma ammonia returns to normal within 48 h. In a patient whose hyperammonemia is due to an organic acidemia, hemodialysis effectively removes both the offending organic acid and ammonia from the body.

To minimize the possible production of ammonia by intestinal bacteria, oral administration of neomycin limits growth of intestinal bacteria that can produce ammonia. Oral lactulose acidifies the intestinal lumen, thereby reducing the diffusion of ammonia across the intestinal epithelium. Neither compound has been used extensively to treat acute hyperammonemia caused by inborn errors of metabolism in newborn infants.

There may be considerable lag between the normalization of ammonia and an improvement in the neurologic status of the patient. Several days may be needed before the infant becomes fully alert.

General Aspects of Long-Term Therapy

The goal of therapy is to provide a low-protein diet, arginine, and energy to promote growth and development, while preventing recurrent hyperammonemic crises. In general, all patients, regardless of the enzymatic defect, require some degree of protein restriction (1–2 g kg⁻¹ 24 h⁻¹). In patients with defects in the urea cycle, chronic administration of benzoate (250–500 mg kg⁻¹ 24 h⁻¹), phenylacetate (250–500 mg kg⁻¹ 24 h⁻¹), and arginine (200–400 mg kg⁻¹ 24 h⁻¹) or citrulline (in patients with OTC deficiency, 200–400 mg kg⁻¹ 24 h⁻¹) is effective in maintaining blood ammonia levels within the normal range. Phenylbutyrate may be used in place of phenylacetate because the patient and the family may not accept the latter owing to its offensive odor. A commercial preparation of the compound is available for oral use.

These compounds have been used during pregnancy without obvious teratogenic effect but the experience is still quite limited.

Growth parameters, especially head circumference, and nutritional indices (blood albumin, pre-albumin, pH, electrolytes, amino acids, zinc, and selenium) should be followed closely. Long-term care of these patients is best achieved by a team of experienced professionals (metabolic specialist, primary care physician, nutritionist, neurologist, and developmental pediatrician). Patients with urea cycle defects may develop a skin rash

resembling acrodermatitis enteropathica if they develop essential amino-acid deficiency, especially that of arginine, due to over-restriction of dietary protein. Catabolic states (infections and fasting) triggering hyperammonemia should be avoided or treated vigorously. In patients with urea cycle defects, acute hyperammonemic attacks may be precipitated by valproate administration. The use of this anticonvulsant should be avoided.

CPS and NAG Synthetase Deficiencies

Human CPS is the rate-limiting enzyme catalyzing the first step of ureagenesis. NAG is required for CPS activation. Patients with deficiencies of NAG or CPS can present as a devastating disease in neonates or as a more insidious late-onset condition. The severity of symptoms and its timing varies with the degree of enzyme deficiency and the metabolic environment. In near-complete deficiency, symptoms most commonly appear during the first few days of life when patients develop signs and symptoms of hyperammonemia such as poor feeding, vomiting, lethargy, increased intra-cranial pressure, convulsion, and coma. Late forms (as late as 32 years of age) may present as an acute episode of hyperammonemia (lethargy, vomiting, headache, disorientation, ataxia, seizures, and psychosis) in a seemingly normal individual and the onset of symptoms is often associated with increase in dietary protein intake, infection, stress, or exposure to medications such as valproate or steroids. Coma and death may occur during these episodes. Diagnostic confusion with migraine is frequent. Intermediate forms with mental retardation and chronic subclinical hyperammonemia interspersed with bouts of acute hyperammonemia have also been observed.

Laboratory findings include hyperammonemia. The plasma amino-acid analysis commonly shows a marked increase of glutamine and alanine with relatively low levels of citrulline and arginine. Urinary orotic acid is usually low or may be absent (see Figure 2). Patients with the late-onset form may have normal ammonia, plasma amino-acids, and urinary orotic acid levels when well, a finding that may confound diagnosis.

Treatment of acute hyperammonemic attacks (Table 2) and long-term therapy are discussed above. Patients with NAG synthetase deficiency benefit from oral administration of carbamylglutamate. Although deficiency of NAG synthetase is rare in North America, it is very important to differentiate this entity from CPS deficiency by assay of the enzyme activities in liver biopsy specimens or by gene sequencing.

Both CPS and NAG synthetase deficiencies are inherited as autosomal recessive traits. Both enzymes are normally present in liver and intestine. The CPS gene is mapped to chromosome

2q35. Several disease-causing mutations have been reported but genotype/phenotype correlations were not observed. The NAG synthetase gene is mapped to chromosome 17q21.31, and disease-causing mutations were reported in different families. The prevalence of these conditions is not known.

OTC Deficiency

It is the most common form of all the urea cycle disorders, comprising about 40% of all cases and is transmitted as an X-linked partially dominant trait. The clinical phenotype in affected males as well as heterozygous females shows a spectrum of severity ranging from neonatal hyperammonemic coma to asymptomatic adults.

Clinical manifestations in a male newborn usually are poor feeding, vomiting, lethargy, convulsions, and coma occurring in the first few days of life. Milder forms, which are commonly seen in heterozygous females and in some affected males, characteristically have episodic manifestations that may occur at any age (usually after infancy). Patients usually have episodes of hyperammonemia, and present with vomiting and neurologic abnormalities such as ataxia, mental confusion, agitation, and combativeness, which are separated by periods of wellness. These episodes usually occur after ingestion of a high-protein diet, exposure to medications such as valproate or steroids, or as a result of a catabolic state such as infection or fasting. Hyperammonemic coma, cerebral edema, and death may occur during one of these attacks. Mild-to-moderate mental retardation is common. Mild cerebral dysfunction may be present in asymptomatic female carriers. Gallstones have been seen in the survivors; the mechanism remains unclear.

The laboratory findings during the acute attack are hyperammonemia, and markedly elevated levels of glutamine and alanine, and decreased levels of citrulline and arginine. The blood level of urea (blood urea nitrogen (BUN)) is usually low. Elevation of the plasma concentrations of glutamine and alanine is secondary to hyperammonemia. A marked increase in the urinary excretion of orotic acid differentiates this condition from CPS deficiency (see [Figure 2](#)). In the mild form, patients may not have any abnormal laboratory findings. The diagnosis can be confirmed by performing an assay of enzyme activity that is normally present only in the liver or by mutational analysis of the OTC gene. About 20% of affected patients demonstrate a normal sequence, perhaps because the mutation involves an intron or a leader sequence. There are no prevalent mutations in the OTC gene and essentially every affected family has its own 'private' mutation. Prenatal diagnosis has been achieved by measuring the enzyme activity in the fetal liver tissue or by analysis of DNA studies in amniocytes or in chorionic villi samples. Obligate female carriers may be detected either by oral protein loading, which increases plasma ammonia and urinary orotic acid levels, or the allopurinol loading test, which increases the excretion of urinary orotic acid. A careful family history may reveal important clues that raise the possibility of OTC deficiency. A history of migraine or protein aversion is common in maternal female relatives of the proband. Indeed, careful scrutiny of the family history may reveal a pattern of unexplained deaths in male newborns in the maternal lineage.

Treatment of acute hyperammonemic attacks and the long-term therapy of the condition are discussed above. Citrulline is preferred in place of arginine in patients with OTC deficiency. Patients with the severe form of disease may need early liver transplantation to prevent recurrent metabolic crisis and severe mental retardation.

The OTC gene has been mapped to chromosome Xp21.1. Many disease-causing mutations have been identified. Identifying the deleterious mutations helps to assess the severity of the disorder and select the most appropriate therapy. For example, patients with the most severe mutations, which result in profound enzyme deficiency, may need early liver transplant to prevent severe mental retardation or death. Mothers of affected infants are expected to be carriers of the mutant gene unless a *de novo* mutation has occurred. Gonadal mosaicism can explain the cause of having multiple affected offsprings in a mother with normal genotype.

Citrullinemia

Two clinically and genetically distinct forms of citrullinemia have been identified: citrullinemia type-I (classical form) is due to the deficiency of the AS enzyme and citrullinemia type-II (adult form) is due to deficiency of a mitochondrial transport protein named Citrin.

AS Deficiency (Citrullinemia Type I)

Citrullinemia type I is caused by the deficiency of AS (see [Figure 1](#)). The clinical phenotype varies based on the residual enzyme activity. The severe or neonatal form, which is most common, usually appears in the first few days of life. Affected infants typically appear normal at birth, but, as ammonia builds up in the body, they experience a progressive lethargy, poor feeding, vomiting, seizures, and loss of consciousness. In the subacute or mild form, clinical findings such as failure to thrive, frequent vomiting, intense headaches, ataxia, and developmental delay develop after 1 year of age.

Episodes of hyperammonemia, triggered by an intercurrent catabolic state, may bring the diagnosis to light.

Laboratory findings are similar to those found in patients with OTC deficiency, except that the plasma citrulline concentration is markedly elevated (50–100 times normal) in patients with citrullinemia type I (see [Figure 2](#)). Urinary excretion of orotic acid is moderately increased. The diagnosis is confirmed by assay of enzyme activity in cultured fibroblasts or by mutation analysis. Prenatal diagnosis is possible by assay of enzyme activity in amniotic fluid or by mutation analysis of cultured chorionic villus cells.

Treatment of acute hyperammonemic attacks and the long-term therapy of the condition are outlined earlier in this article. Although prognosis is poor for symptomatic neonates, patients with the mild disease usually do well on a protein-restricted diet in conjunction with sodium benzoate, phenylbutyrate, and arginine therapy. Mild-to-moderate mental deficiency is a common sequela, even in a well-treated patient. Liver transplant should be considered in patients with the severe form of this disease.

Citrullinemia is inherited as an autosomal recessive manner. The gene is located on chromosome 9q34.1. Several

disease-causing mutations have been identified in different families. The majority of patients are compound heterozygotes for two different alleles. The prevalence of the condition is not known. The recent introduction of neonatal screening for urea cycle defects has disclosed affected patients, who are ostensibly asymptomatic, even with ingestion of a regular diet. It is unclear at present whether stringent dietary protein restriction is necessary in these infants. Long-term follow-up is needed.

Citrin Deficiency (Citrullinemia Type II)

Citrin is a mitochondrial transport protein encoded by a gene (*SLC25A13*) located on chromosome 7q21.3. One of the functions of this protein is to transport aspartate from mitochondria to the cytoplasm; aspartate is required to convert citrulline to argininosuccinic acid (see [Figure 1](#)). If aspartate is unavailable to the cytoplasm, urea will not be formed at a normal rate and citrulline will accumulate. Patients with citrin deficiency have decreased AS activity in the liver but they do not have any mutations/deletion in their AS gene. It is postulated that citrin deficiency or its mutated gene interferes with translation of the messenger RNA for AS. The condition initially was reported almost exclusively among Japanese individuals but a few non-Japanese patients have also been identified. There are two phenotypes of citrin deficiency: (1) neonatal intrahepatic cholestasis (citrullinemia type II–neonatal form) and (2) citrullinemia type II (adult form–CTLN2). The pathogenesis of citrullinemia type II (neonatal and adult forms) remains enigmatic.

Neonatal intrahepatic cholestasis (citrullinemia type II–neonatal form)

Clinical symptoms and signs usually start in the first 3 months of age and disappear by age 1 year with appropriate treatment. The findings are intrahepatic cholestasis, diffuse fatty liver and parenchymal cellular infiltration associated with hepatic fibrosis, growth retardation, marked hypoproteinemia, increased prothrombin time and partial thromboplastin times, hemolytic anemia, and hepatomegaly. Patients may have elevated galactose detected through newborn screening but all enzymes involved in galactose metabolism are normal. The reason for hypergalactosemia is probably liver dysfunction. Although the condition is almost exclusively seen in Japan, the diagnosis should be considered in cases of unexplained neonatal hepatitis with cholestasis. Data on the long-term prognosis and the natural history of the condition are limited; development into the adult form of the condition (see later) after several years of seemingly asymptomatic hiatus has been observed.

Citrullinemia type II, adult form (adult-onset citrullinemia, citrullinemia type II–mild form)

Citrullinemia type II is an adult-onset disease, which starts suddenly in a previously normal individual at around 20–50 years of age, and manifests with recurring episodes of hyperammonemia, neuropsychiatric symptoms including nocturnal delirium, aggression, irritability, hyperactivity, delusions, disorientation, aberrant behavior, tremors, and frank psychosis. Patients usually have hyperammonemia and moderately elevated citrulline. Patients who recover from the first episode will have recurrent attacks and most will die within a few years of diagnosis, mainly from cerebral edema. Pancreatitis,

hyperlipidemia, and hepatoma are major complications among the survivors. Many individuals with citrullinemia type II have a strong preference for protein-rich and/or lipid-rich foods, and an aversion to carbohydrate-rich foods. Medical treatment has been mostly ineffective for prevention of future attacks. Indeed, some have speculated that the administration of large amounts of glucose might even prove deleterious, because the citrin transporter is important to the metabolism of glucose. Liver transplantation is the most effective therapy.

Several disease-causing mutations of the gene have been identified in affected families in Japan. Although the frequency of an abnormal gene is quite high in Japan (1:20 000 homozygosity), the clinical condition has a frequency of only 1:100 000. This indicates that a substantial number of homozygous individuals remain asymptomatic. Only a few non-Japanese patients have been identified.

AL Deficiency (Argininosuccinic Aciduria)

There is considerable phenotypic variability. The most severe cases present in the neonatal period with lethargy, vomiting, poor suck, hypothermia, hepatomegaly, and progressive encephalopathy. Mortality is high and neurologic damage is frequent in survivors. The disease may also be manifested later in childhood (subacute or late form) with an acute encephalopathy that often is precipitated by infection or high-protein intake. Patients may also display chronic symptoms such as episodic vomiting, mental status changes, lethargy, and behavioral abnormalities associated with hyperammonemia. Mild-to-moderate developmental delay and seizures can be seen even in patients without a documented history of hyperammonemia. Abnormalities of the hair characterized by dryness and brittleness are of special diagnostic value (trichorrhexis nodosa). Gallstones have been seen in some of the survivors.

Laboratory findings include hyperammonemia, moderate elevations in liver enzymes, nonspecific increases in plasma levels of glutamine and alanine, moderate increase in plasma levels of citrulline, and marked increase in plasma levels of argininosuccinic acid (see [Figure 2](#)). In most amino acid analyzers, argininosuccinic acid appears as series of anhydrides within the isoleucine or methionine region, which may cause confusion in the diagnosis. Argininosuccinic acid can also be found in large amounts in urine and spinal fluid. The levels in the spinal fluid are usually higher than those in plasma. The enzyme is normally present in red blood cells, the liver, and cultured fibroblasts. Prenatal diagnosis is possible by measurement of enzyme activity in cultured amniotic cells or by identification of the mutant gene. Argininosuccinic acid is also elevated in the amniotic fluid of affected fetuses. This deficiency is inherited as an autosomal recessive trait. Several mutations were reported throughout the gene, but there were no genotype/phenotype correlations.

Treatment of acute hyperammonemic attacks and the long-term therapy of the condition are outlined earlier in this article. Mental retardation, persistent hepatomegaly with mild increases in liver enzymes, and bleeding tendencies due to abnormal clotting factors are the common sequelae of the disease. Liver transplantation should be considered in patients with the severe form of the disease.

Arginase Deficiency (Hyperargininemia)

There are two genetically distinct arginases in humans. One is cytosolic (A1) and is expressed in the liver and erythrocytes, and the other (A2) is found in renal and brain mitochondria. The gene for the cytosolic enzyme, which is deficient in patients with arginase deficiency, is mapped to chromosome 6q23. The role of the mitochondrial enzyme is not well understood; its activity increases in patients with argininemia but has no protective effect. Several disease-causing mutations have been identified in different families. Arginase deficiency is inherited as an autosomal recessive trait.

The clinical findings are nonspecific and are quite different from those of other urea cycle enzyme defects. The patients usually remain asymptomatic in the first few months or, sometimes, years of life. A progressive spastic diplegia, choreoathetotic movements, and loss of developmental milestones in a previously normal infant are usually the initial signs and symptoms. Mental retardation is progressive; seizures are common, but episodes of severe hyperammonemia are not usually seen. Hepatomegaly may be present. The acute neonatal form with intractable seizures, cerebral edema, and death has also been reported.

Laboratory findings include three- to fourfold elevation of the plasma and cerebrospinal fluid (CSF) arginine concentration (see [Figure 2](#)). Urinary orotic acid is moderately increased. Plasma ammonia levels may be normal or mildly elevated. Urinary excretions of arginine, lysine, cystine, and ornithine are usually increased, but normal levels have also been noted. Therefore, determination of amino acids in plasma is a critical step in the diagnosis of argininemia. The diagnosis is confirmed by assaying arginase activity in red blood cells or mutation analysis. Patients can be detected with newborn screening.

The mainstay of treatment is a low-protein diet. Administration of a synthetic protein made of essential amino acids usually results in a dramatic decrease in plasma arginine concentration and an improvement in neurologic abnormalities. The composition of the diet and the daily intake of protein should be monitored by frequent plasma amino-acid determinations. Sodium benzoate or sodium phenylbutyrate ($250\text{--}375\text{ mg kg}^{-1}\text{ 24 h}^{-1}$) is also effective in controlling hyperammonemia, when present; lowering of plasma arginine levels has been noted with this treatment. Mental retardation is a common sequela of the condition.

Transient Hyperammonemia of the Newborn

Transient hyperammonemia of unknown cause may be a relatively common variety of neonatal hyperammonemia. Most affected infants are premature and have mild respiratory distress syndrome. Hyperammonemic coma may develop within 24 h of life, and the infant may succumb to the disease if treatment is not started immediately. Laboratory studies reveal marked hyperammonemia, and plasma levels may be as high as $4000\text{ }\mu\text{mol l}^{-1}$. Treatment should be initiated promptly and continued vigorously (see earlier). Recovery without sequelae is common, and hyperammonemia does not recur even with a normal protein diet.

Gyrate atrophy of the retina and choroid

Ornithine is one of the intermediate metabolites of the urea cycle that is not incorporated into natural proteins. Rather, it is generated in the cytosol from arginine and must be transported into the mitochondria, where it is used as a substrate for the enzyme OTC to form citrulline. Excess ornithine is catabolized by two enzymes, ornithine 5-aminotransferase, which is a mitochondrial enzyme and converts ornithine to a proline precursor, and ornithine decarboxylase, which resides in the cytosol and converts ornithine to putrescine (see [Figure 1](#)).

The deficiency of the enzyme ornithine 5-aminotransferase causes gyrate atrophy of the retina and choroid. It is a rare, autosomal recessive disorder (see [Figure 1](#)). About 30% of the reported cases are from Finland. The initial clinical findings are myopia and reduction of peripheral vision and in some patients, reduction of night vision in the first decade of life. Almost all patients demonstrate impaired peripheral vision by age 10 years. Loss of visual function progresses to complete blindness by the fourth decade of life. Atrophic lesions in the retina resemble cerebral gyri. These patients usually have normal intelligence; however, early degenerative and atrophic brain changes and abnormal electroencephalographs (EEGs) were reported in some patients.

There are markedly elevated levels (10- to 15-fold) of ornithine in plasma and other body fluids. The goal of treatment is to reduce high ornithine levels. Pyridoxine therapy may decrease plasma ornithine levels and is recommended for all newly diagnosed patients at a dose of $300\text{--}600\text{ mg d}^{-1}$ for a 2-week trial. Arginine (precursor of ornithine) restriction also helps to decrease ornithine levels. On average, protein foods contain 4–6% arginine. A low-protein diet may be supplemented with essential amino acids. Severe arginine restriction may cause hyperammonemia. The gene for ornithine 5-aminotransferase is mapped to chromosome 10q26. Many disease-causing mutations have been identified in different families.

Hyperammonemia–Hyperornithinemia–Homocitrullinemia Syndrome

Hyperammonemia–hyperornithinemia–homocitrullinemia (HHH) syndrome is a rare autosomal recessive disorder in which there is defective transport of ornithine from the cytosol into the mitochondria, resulting in accumulation of ornithine in the cytosol, and its deficiency in mitochondria. The intramitochondrial deficiency results in condensation of carbamoylphosphate with lysine rather than ornithine (the usual co-reactant in the ornithine transcarbamylase reaction) and production of homocitrulline ([Figure 1](#)). Intramitochondrial ornithine deficiency, due to the transport defect, results in a functional deficiency of OTC and OAT activities. There is a wide spectrum of clinical manifestations. Hyperammonemia may develop shortly after birth or may be delayed until adulthood. Acute episodes of hyperammonemia manifest as refusal to feed, vomiting, and lethargy; coma may occur during infancy. Progressive neurologic signs, such as lower limb weakness, increased deep tendon reflexes, spasticity, clonus, seizures, and varying degrees of psychomotor retardation may develop if the condition remains undiagnosed. Coagulopathy, especially factor VII and

X deficiencies, has been reported in some patients. No clinical ocular findings have been observed in patients with HHH.

The triad of hyperornithinemia, homocitrullinuria, and hyperammonemia is a pathognomonic laboratory finding. Plasma ornithine level is 3–10 times higher than normal and tends to be lower than that seen in patients with gyrate atrophy of the retina and choroid.

Restriction of protein intake improves hyperammonemia. Ornithine supplementation may produce clinical improvement in some patients. The gene for this disorder (*SLC25A15*) is located on chromosome 13q14. Several mutant alleles have been reported in the literature. The F188 deletion is the most common mutation in patients of French Canadian ancestry. No genotype and phenotype correlation has been observed.

See also: **Metabolism Vitamins and Hormones: Amino Acid Metabolism; Branched-Chain Amino Acids.**

Further Reading

- Bachmann C (2003) Inherited hyperammonemias. In: Blau N, Duran M, Blaskovics ME, et al. (eds.) *Physician's Guide to the Laboratory Diagnosis of Metabolic Diseases*, 2nd edn., pp. 261–276. Berlin: Springer.
- Bachmann C (2003) Outcome and survival of 88 patients with urea cycle disorders: A retrospective evaluation. *European Journal of Pediatrics* 162: 410–416.
- Batshaw ML (1984) Hyperammonemia. *Current Problems in Pediatrics* 14: 1–69.
- Brusilow SW (1995) Urea cycle disorders: Clinical paradigm of hyperammonemic encephalopathy. *Progress in Liver Diseases* 13: 293–309.
- Brusilow SW and Horwich AL (2001) Urea cycle enzymes. In: Scriver C, Beaudet A, Valle D, and Sly W (eds.) *Metabolic and Molecular Bases of Inherited Disease*, 8th edn., pp. 1909–1963. New York, NY: McGraw Hill.
- Cederbaum SD, Moedjono SJ, Shaw KN, et al. (1982) Treatment of hyperargininemia due to arginase deficiency with a chemically defined diet. *Journal of Inherited Metabolic Disease* 5: 95–99.
- Ficiocioglu C, Mandell R, and Shih V (2009) Argininosuccinate lyase deficiency: Long-term outcome of 13 patients detected by newborn screening. *Molecular Genetics and Metabolism* 98(3): 273–277.
- Gao HZ, Kobayashi K, Tabata A, et al. (2003) Identification of 16 novel mutations in the argininosuccinate synthetase gene and genotype–phenotype correlation in 3s8 classical citrullinemia patients. *Human Mutation* 22(1): 24–34.
- Ito T, Sumi S, Kidouchi K, et al. (2003) Allopurinol challenge tests performed before and after living-related donor liver transplantation in citrullinaemia. *Journal of Inherited Metabolic Disease* 26: 87–88.
- Saheki T, Kobayashi K, Iijima M, et al. (2004) Adult-onset type II citrullinemia and idiopathic neonatal hepatitis caused by citrin deficiency: Involvement of the aspartate glutamate carrier for urea synthesis and maintenance of the urea cycle. *Molecular Genetics and Metabolism* 81: S20–S26.
- Shih V and Ficiocioglu C (2000) Genotype and phenotype findings in the hyperornithinemia–hyperammonemia–homocitrullinuria (HHH) syndrome. *Journal of Inherited Metabolic Disease* 23(supplement 1): 72.
- Tuchman M, Lee B, Lichter-Konecki U, et al. (2008) Cross-sectional multicenter study of patients with urea cycle disorders in the United States. *Molecular Genetics and Metabolism* 94: 397–402.
- Tuchman M, Matsuda I, Munnich A, et al. (1995) Proportions of spontaneous mutations in males and females with ornithine transcarbamylase deficiency. *American Journal of Medical Genetics* 55: 67–70.
- Valle D and Kaiser-Kupfer M (1982) Gyrate atrophy of the choroid and retina. *Progress in Clinical and Biological Research* 82: 123–134.
- Whittington PF, Alonso EM, Boyle JT, et al. (1998) Liver transplantation for the treatment of urea cycle disorders. *Journal of Inherited Metabolic Disease* 21(supplement 1): 112–118.

Relevant Websites

- <http://www.ammonul.com> – Sodium phenylacetate and sodium benzoate injection 10%/10% product.
- <http://www.buphenyl.com> – The FDA-approved oral medication for the chronic adjunctive management of hyperammonemia in patients with urea cycle disorders (UCDs).

V

Vacuoles

C J Stefan, W M Henne, and S D Emr, Cornell University, Ithaca, NY, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Christopher J. Stefan and Scott D. Emr, volume 4, pp. 331–336, © 2004, Elsevier Inc.

Glossary

Active transport Use of energy to transport a substance, often across membranes from an area where it is in lower concentration to an area where it is in higher concentration. Membrane proteins called transporters perform active transport.

Endocytosis Ingestion of particulate matter or fluid by phagocytosis or pinocytosis; that is, bringing material into a cell by invagination of its surface membrane and then pinching off the invaginated portion to form a vesicle, endosome, or vacuole.

Hydrolases Enzymes that act as catalysts in the cleavage of covalent bonds with accompanying addition of water. Lipases are hydrolases that break down fatty acids and other lipids. Proteases specifically digest peptide bonds.

Multi-vesicular body (MVB) A specific subset of late endosomes defined as containing intra-luminal vesicles

(ILVs) created by the endosomal sorting complexes required for transport (ESCRT) pathway. Transmembrane receptors destined for degradation are packaged into the ILVs, and when the MVBs undergo fusion with the vacuole, they are degraded.

Vacuole A large, membrane-bound cytoplasmic organelle that functions in ingestion, digestion, excretion, and storage of water, sugars, proteins, lipids, ions, and other materials. Vacuoles and vacuole-related organelles are found in fungal, plant, protozoan, and mammalian cells. Most plant cells have a single vacuole that takes up much of the cell and helps maintain the shape/turgor of the cell.

Vesicular transport Trafficking of proteins and lipids from one cellular compartment to another by means of membrane-enclosed intermediates such as endosomes or organelle fragments.

Identification/Discovery of Vacuoles and Vacuolar Constituents

Vacuoles were first described in early microscopy studies on plant and protozoan cells in the mid-nineteenth century. Felix Dujardin conceived the term ‘vacuole’ in 1835 to describe large translucent regions in cells surrounded by a membrane that were devoid of cytoplasm. Subsequent studies in the late nineteenth century revealed that vacuoles were not simply devoid, but serve as storage sites for water, sugars, and salts. In 1955, Christian de Duve coined the term ‘lysosomes’ to describe vacuole-like compartments containing acid hydrolases in mammalian cells. Shortly after, independent cell morphological studies undertaken by the Palade and Novikoff laboratories revealed that lysosomes appear as organelles with varying diameters (0.1–1.0 μm) and with heterogeneous morphologies, containing electron-dense cores and in some cases internal vesicles. In the yeast *Saccharomyces cerevisiae*, vacuoles are larger with respect to cell size, but display similar characteristics in terms of electron density and intraluminal vesicles, when observed under certain conditions (Figures 1(a) and 1(b)).

Biochemical studies performed by de Duve and colleagues initially identified lysosomes as organelles of highly buoyant centrifugal properties that exhibited hydrolytic activities. The hydrolases within isolated lysosomes demonstrated latent activity, dependent on the integrity of their surrounding membrane. Specifically, hydrolase activity was retained within the limiting vacuole/lysosome membrane, restricted from the cell lysate unless the limiting membrane was disrupted. Since vacuoles/lysosomes are morphologically heterogeneous, biochemical studies have formed the current definition of vacuoles/lysosomes as membrane-bound acidic organelles that contain mature acid-dependent hydrolases and certain integral membrane glycoproteins (termed LAMPs, lysosomal-associated membrane proteins). Vacuoles lack other proteins that distinguish them from endosomes (compartments that deliver material to vacuoles), such as biosynthetic sorting receptors (such as the mannose 6-phosphate receptor and sortilin) and recycling cell-surface receptors (such as transferrin).

As mentioned, vacuoles/lysosomes are membrane compartments that serve as a major degradative compartment in eukaryotic cells. Both endogenous and exogenous molecules can be delivered to vacuoles through biosynthetic and

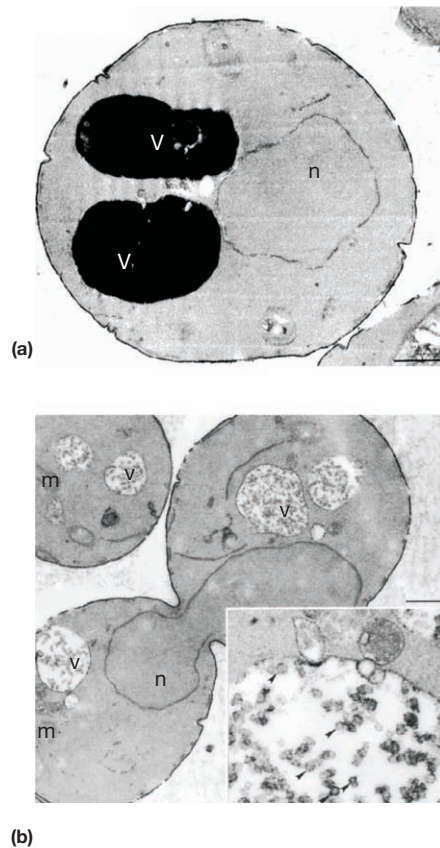


Figure 1 Ultrastructural analysis of vacuole morphology in wild-type (a) and hydrolase-deficient (b) yeast cells reveals electron-dense cores and intraluminal membrane vesicles, respectively. (a) Wild-type cells were grown to mid-log phase, fixed, and visualized by electron microscopy. Vacuoles appear as dark, electron-dense organelles due to the accumulation of cellular proteins, lipids, and other molecules. (b) In cells lacking vacuolar ATPase activity (*vma4Δ* mutant cells), vacuoles appear translucent by electron microscopy and contain intraluminal vesicles. (Inset) High magnification image of vacuolar vesicles (arrowheads). v, vacuoles; n, nucleus; m, mitochondria. Scale = 0.1 μm . Reproduced from Wurmser A and Emr SD (1998) Phosphoinositide signaling and turnover: PtdIns(3)P, a regulator of membrane traffic, is transported to the vacuole and degraded by a process that requires luminal vacuolar hydrolase activities. *EMBO Journal* 17: 4930–4942, copyright 1998.

endocytic pathways. In addition, vacuoles can degrade cytoplasmic proteins via a process termed ‘autophagy’. The degradative function of these organelles is carried out by numerous acid-dependent hydrolases (e.g., proteases, lipases, and glycosidases) contained within its lumen (Figure 2). The outer, limiting membrane of vacuoles/lysosomes also contains a set of highly glycosylated, vacuolar/lysosomal-associated membrane proteins (LAMPs), many with unknown functions (Figure 2). Additional vacuolar membrane proteins mediate transport of ions, amino acids, and other solutes across the vacuolar membrane and maintain an acidic luminal pH in the range of 4.5–6.5 (Figure 2).

Vacuole Function in Yeast

The vacuole of the yeast *S. cerevisiae* plays important roles in protein degradation, pH regulation, ion homeostasis, nutrient storage, and stress responses. Degradation is mediated by a large number of hydrolases that reside within the vacuole (proteases, lipases, and glycosidases; Figure 2) with various substrate specificities. Thus, it is a major site for degradation of cellular proteins, lipids, and even whole organelles.

A remarkable feature of vacuoles is their acidification. In yeast, a specific vacuolar proton pump ATPase (V-ATPase) mediates this process. Biochemical and genetic approaches have identified 14 subunits of the V-ATPase (encoded by the VMA genes for vacuolar membrane ATPase). These 14 subunits comprise two functional subcomplexes: the peripheral V_1 subcomplex, responsible for ATP hydrolysis, and the V_0 subcomplex, responsible for proton translocation across the membrane (Figure 2). Cells lacking a functional V-ATPase are viable, but require an intact endocytic pathway to the vacuole and fail to grow on media with neutral pH, indicating that vacuole acidification is an essential feature. Consistent with this, the electrochemical gradient generated by the V-ATPase is necessary to drive several other transport systems described below (Figure 2).

In addition to its hydrolytic functions, the vacuole serves as a storage compartment for several ions, nutrients, and metabolites (Figure 2). For example, the Vcx1 protein acts as a high-capacity $\text{H}^+/\text{Ca}^{2+}$ exchanger in the regulation of vacuolar calcium flux (Figure 2). The vacuole also stores both phosphate and polyphosphate (Figure 2). Since hydrolysis of phosphate leads to the release of protons, a role for polyphosphate in pH homeostasis has been suggested. Polyphosphate is proposed to function as a cation counter-ion as well. In addition, the vacuole membrane possesses ABC family transporters that sequester toxic metals and metabolites away from the cytoplasm (Figure 2). Basic and neutral amino acids are also stored inside the vacuole at high concentrations. For example, arginine accumulates in vacuoles at concentrations an order of magnitude greater than in the surrounding media (Figure 2). Consistent with this, several H^+ /amino acid antiporter systems have been identified in yeast. Pools of amino acids stored in the vacuole may be utilized during nitrogen starvation.

Vacuole Biogenesis and Transport Pathways to the Vacuole in Yeast

To carry out the numerous functions that vacuoles perform, several membrane trafficking pathways deliver cellular components to vacuoles. Model genetic systems such as the yeast *S. cerevisiae* have been instrumental in identifying both proteins and lipids that constitute components of vacuoles and the machinery involved in transport to vacuoles. Importantly, many of these factors and trafficking pathways are conserved in mammalian cells.

At least eight pathways have been identified in vacuole biogenesis and vacuolar protein trafficking in yeast: endocytosis from the cell surface, two distinct biosynthetic pathways from the Golgi (the vacuolar protein sorting (Vps) and AP-3

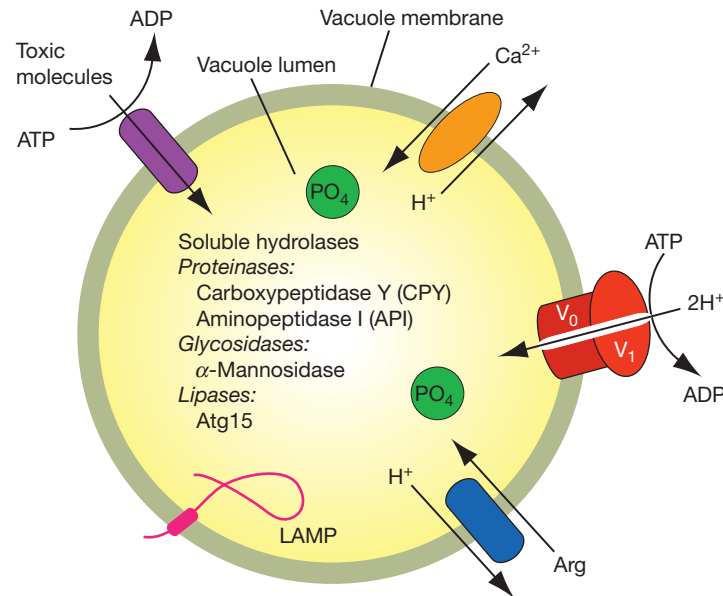


Figure 2 Schematic overview of vacuolar functions and components. The vacuole lumen possesses several hydrolase activities and accumulates ions, such as phosphate (PO_4^-) and polyphosphates. Lysosomal membranes also contain integral transmembrane proteins termed LAMPs. The vacuolar ATPase (V_0/V_1) functions as the primary proton pump that generates an electrochemical gradient used to drive other vacuole membrane transport systems, such as arginine (Arg) and divalent cation (Ca^{2+}) anti-transporters. Toxic metals and metabolites are sequestered in the vacuole by ABC family transporters.

pathways), multivesicular body (MVB) sorting, the cytoplasm to vacuole pathway (Cvt), macro-autophagy, micro-autophagy, and vacuole inheritance during cell division (Figure 3(a)). The endocytic pathway is essential for regulating levels of cell-surface signaling receptors and removal of damaged transmembrane proteins and lipids. The endocytic and Vps pathway intersect at a late endosomal compartment. Components of the Vps sorting pathway are directly involved in vacuolar biogenesis, morphology, and function, as discussed in greater detail below. Alkaline phosphatase (ALP) travels from the Golgi to the vacuole along the AP-3 sorting pathway that is partially independent of the Vps and endocytic pathways. Proteins such as aminopeptidase I (API) are delivered from the cytoplasm to the vacuole through pathways involving products of the *ATG* (autophagy) genes. In micro-autophagy, the vacuole membrane invaginates to engulf cytoplasmic material, including entire organelles such as peroxisomes. Vacuolar segregation into dividing daughter cells of budding yeast requires *VAC* (vacuolar inheritance) genes, some of which also participate in the Vps pathway. Previous studies have also identified several distinct factors involved in vacuolar inheritance.

The Vps Pathway

Carboxypeptidase Y (CPY) is the prototype of a subset of proteins that traffic from the Golgi to the vacuole via an endosomal intermediate (Figure 3(a)). This pathway depends on the function of over 60 *VPS* gene products, mutations in which result in the mis-sorting of CPY to the secretory pathway, abnormal vacuole morphology, and, in some cases, abnormal endosome morphology. In a manner analogous to receptor-mediated sorting in mammalian cells, the CPY receptor Vps10 (the prototype of the

sortilin family) binds CPY at the Golgi via a targeting signal. Sorting of this receptor–ligand complex into vesicles destined for endosomes requires the coat protein clathrin and accessory factors including the AP-1 clathrin adaptor complex, the Gga proteins, and a dynamin family guanosine triphosphatase (GTPase) (Vps1). The class D Vps proteins are involved in CPY vesicle targeting and fusion with endosomes, such as Vps21 (Rab5 homolog), Vps9 (Rab GEF), Vac1 (EEA1), Vps45 (Sec1 homolog), Pep12 (t-soluble NSF attachment protein (SNAP) receptor (t-SNARE)), the class C core vacuole–endosome transport (CORVET) complex (consisting of class D and C Vps proteins), and Sec18 (NSF). The Vps34 PI 3-kinase also contributes to endosomal targeting and maturation by the generation of the phosphoinositide lipid PI3P that recruits effector proteins such as the Fab1, YGLO23, Vps27, and EEA1 (FYVE) domain-containing protein Vac1. At the endosome, Vps10 dissociates from CPY and recycles to the Golgi in a process requiring the retromer complex (consisting of Vps29, Vps26, Vps35, Vps5, and Vps17). Interestingly, PI3P also mediates endosomal recycling through the recruitment of the PX domain-containing proteins Vps5 and Vps17. Mutants defective in retromer function mis-localize Vps10 to the vacuole membrane and secrete CPY, as Vps10 becomes limiting for subsequent sorting reactions at the Golgi.

Another set of Vps proteins, the class E Vps proteins, form the endosomal sorting complexes required for transport (ESCRT) complexes necessary for sorting transmembrane proteins at the endosome and the formation of multi-vesicular bodies (MVBs). Recent attention has been paid to this pathway because of the elucidation of the ESCRT protein complexes, which are necessary for the downregulation of internalized cell-surface signaling receptors and damaged plasma membrane proteins destined for vacuolar degradation (Figure 4).

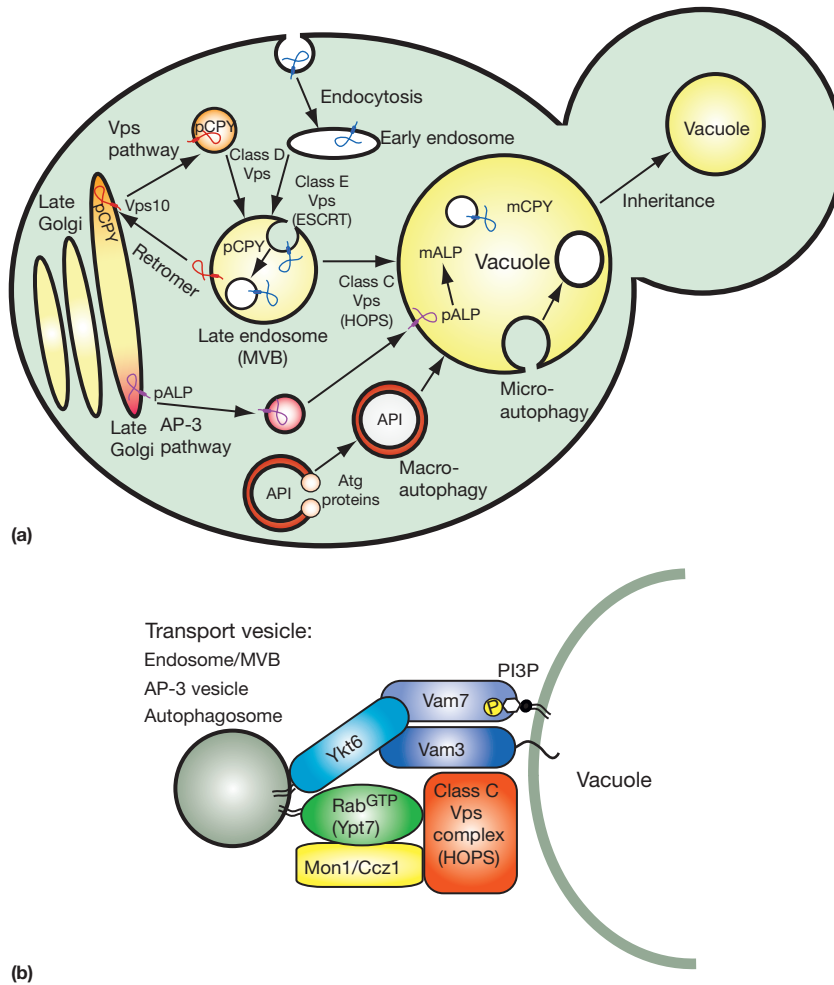


Figure 3 (a) Model representation of the transport pathways to the yeast vacuole. Endocytosis of cell-surface proteins is followed by transport to early endosomes and then late endosomes. The Vps pathway sorts proteins (Vps10 and precursor CPY) from the late Golgi to the late endosome. At the late endosome, ubiquitinated cargo is sequestered into intraluminal vesicles by the ESCRT pathway to form multivesicular bodies (MVBs), while other proteins, such as the Vps10 sorting receptor, are recycled to the Golgi via the retromer pathway. The AP-3 pathway functions as a Golgi to vacuole route that is independent of the late endosome/MVB. In the Cvt pathway, cytoplasmic proteins are enclosed within double membrane structures that subsequently fuse with and deliver their material into the vacuole. Macro-autophagy is similar to the Cvt pathway but is induced primarily under starvation conditions. In micro-autophagy, the vacuole membrane invaginates to mediate the uptake of cytoplasmic material. (b) Schematic diagram of proteins that function in the final docking/fusion step of various cargo vesicles to the vacuole. SNARE proteins (Vam3, Vam7, and Ykt6) are represented by rounded rectangles. Components of the class C Vps complex (Vps11, 16, 18, 33, 39, and 41) that mediate SNARE pairing are shown. Tethering is also mediated by the Rab-like GTPase, Ypt7, and the Mon1/Ccz1 protein complex.

The ESCRT complexes function sequentially on the endosomal surface as discrete modules (elegantly numbered ESCRTs 0, I, II, and III) to bind, cluster, and envelope ubiquitinated transmembrane proteins into the intra-luminal vesicles that bud into the MVB. Cells lacking ESCRT function accumulate abnormal/aberrant endosomes containing both biosynthetic cargoes such as CPY and endocytosed transmembrane proteins. The abnormal/exaggerated endosomes observed in these mutant cells form in part due to impaired MVB vesicle budding into the lumen of the endosome.

Fusion of late endosomes/MVBs with the vacuole requires another set of proteins including Ypt7 (Rab7 homolog) and SNARE proteins (Vti1, Ykt6, Nyv1, Vam7, and Vam3). The

class C Vps protein complex (also termed homotypic vacuolar fusion and protein sorting (HOPS)) consisting of Vps18, Vps11, Vps16, Vps33 (Sec1 homolog), Vps41, and Vps39 (Rab GEF) is required for this final fusion step as well (Figure 3(b)). The Mon1/Ccz1 protein complex has been shown to bind the Rab GTPase Ypt7 in the progression of endosomal maturation and vacuolar targeting. The Vps34 PI 3-kinase also contributes to this fusion step by the generation of PI3P that recruits effector proteins such as the PX domain-containing protein Vam7, and thus PI3P mediates several steps involved in endosomal sorting, maturation, and fusion in route to the vacuole. Mutants with defects in the class C Vps gene products accumulate numerous endosomal intermediates and MVBs that fail to

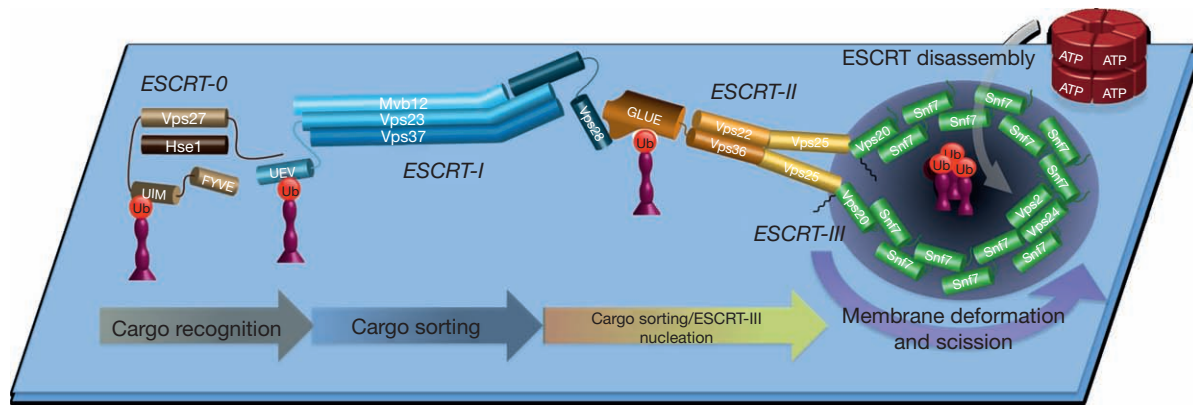


Figure 4 The ESCRT complexes mediate delivery of ubiquitinated cargo to the vacuole/lysosome. Schematic of ESCRT-0, -I, -II, and -III complexes and Vps4 on the late endosome surface. For simplicity, the yeast names for the ESCRT subunits have been used. Once at the endosome, ubiquitinated cargo is recognized by a ubiquitin-interacting motif in ESCRT-0 subunit Vps27. ESCRT-I and -II also bind ubiquitin directly and function as cargo sorting and clustering agents. ESCRT-II also initiates ESCRT-III assembly by a Vps25–Vps20 interaction. Snf7 polymerization is associated with membrane deformation, allowing ESCRT-III to envelope cargo and package it into intra-luminal vesicles that bud into the endosome interior. Finally, Vps2 recruits the AAA ATPase Vps4, which can disassemble the ESCRT machinery for a new round of cargo sequestration. Adapted with permission from Elsevier Limited: *Cell* 137: 181–182, copyright 2009.

fuse with the vacuole. Accordingly, mutants defective in components of the vacuolar fusion machinery display highly fragmented vacuoles and can sometimes even lack vacuoles entirely.

The AP-3 Pathway

ALP is an integral membrane protein that traffics from the Golgi to the vacuole independently of endosome/MVB compartments that transport CPY and endocytic cargoes (Figure 3(a)). A specific adaptor complex, termed AP-3, mediates sorting of ALP into vesicles at the Golgi. However, following formation, fusion of ALP cargo vesicles with the vacuole is dependent on the class C Vps complex, Ypt7, Vam7, and Vam3. Thus, the AP-3 and Vps pathways converge upon common docking/fusion machinery at the vacuole (Figure 3(b)). The mechanism for the convergence, and the reason why AP-3 vesicles fuse with the vacuole, but not with endosomes containing CPY, is still unclear.

Cytoplasm to Vacuole Transport and Macro-Autophagy

Autophagy is a trafficking pathway to vacuoles regulated by changes in nutrient availability. During starvation-induced macro-autophagy, cytoplasmic material is first sequestered in double-membrane vesicles called autophagosomes and then subsequently delivered to the vacuole (Figure 3(a)). While autophagy is induced, the cytoplasm-to-vacuole targeting (Cvt) pathway constitutively packages the hydrolases aminopeptidase I (API) and α -mannosidase into autophagosomes for delivery to vacuoles. Autophagosomes are targeted to, and fuse with, the vacuole by the same machinery that mediates endosome-vacuole fusion (Figure 3(b)). Inside the vacuole, the lipase Atg15 is responsible for autophagosome turnover (Figure 2). Atg15 may also participate in the turnover of intra-luminal vesicles formed by the ESCRT-mediated MVB pathway. In mammalian cells, fusion of autophagosomes with lysosomes results in intermediate compartments termed ‘auto-lysosomes’.

Vacuole/Lysosome Functions in Higher Eukaryotes

The Contractile Vacuole in Protozoa

Protozoa living in freshwater exist in a hypotonic environment. Water flows across their plasma membrane since their cytosol is hypertonic to the environment. To adapt, many protozoa have an organelle, the contractile vacuole complex (CVC), which collects and expels excess water. Previous work shows that CVCs are composed of a central vacuole with radial arm-like extensions. The extensions are often divided into separate bundles of tubules and contain proton-translocating V-ATPase enzymes that provide an electrochemical gradient necessary for fluid collection. The membrane of the central compartment lacks V-ATPases, but can expand into a reservoir for fluid storage, and is capable of fusing with the plasma membrane. It is then that the central compartment undergoes contraction, resulting in fluid expulsion. The exact mechanism of this contraction is unknown, but it has been proposed that in *Dictyostelium* a class of membrane sculpting proteins, the mental retardation GTPase-activating proteins (GAPs) (MEGAPs), polymerizes on the contractile vacuole and use their membrane bending F-BAR domains to squeeze water out of the organelle.

Lysosome-Related Organelles

The mammalian vacuole-like lysosome also is a site for delivery of cellular materials targeted for degradation, as it is the terminal compartment of the endocytic, biosynthetic, phagocytic, and autophagy pathways in mammalian cells. However, the idea that vacuoles/lysosomes are simply degradative sites has evolved. This is partly due to the identification of lysosome-like compartments that perform additional cellular functions, such as lysosomes that secrete their contents after fusion with the plasma membrane. Many properties of vacuoles/lysosomes are shared with this group of cell type-specific lysosome-related organelles, which include melanosomes (pigment granules), lytic granules, platelet-dense granules, basophil granules,

Table 1 Disorders associated with lysosomal dysfunction

<i>Disease</i>	<i>Defective protein(s)</i>	<i>Manifestations and lysosomal defects</i>
<i>Examples of inherited lysosomal storage diseases</i>		
I-cell disease	GlcNAc-phosphotransferase	Multiple lysosomal enzymes are secreted. Cells are highly vacuolated and contain numerous dense inclusion bodies
Tay–Sachs disease	β -Hexosaminidase	Neuronal GM2 gangliosidosis
Pompe's disease	α -Glucosidase	Glycogen storage disease type II
Gaucher's disease	β -Glucosidase	Sphingolipid metabolism defects
Niemann–Pick disease type A	Sphingomyelinase	Sphingolipid and cholesterol metabolism defects. Accumulation of lipid-storing foam cells. In the brain, neurons are highly vacuolated
Hermansky–Pudlack syndrome	AP-3 complex subunits, HOPS complex subunits	Defects in lysosome-related organelle biogenesis, particularly melanosomes and dense granules
Danon disease	LAMP2	Chaperone-mediated autophagy and macro-autophagy defects
Osteopetrosis	CLC7, OSTM1	H ⁺ /Cl [−] anti-transporter defects. Lysosomal acidification defects
<i>Disease</i>	<i>Defective protein(s)</i>	<i>Manifestations and endo-lysosomal defects</i>
<i>Neurodegenerative disorders associated with lysosomal defects</i>		
Alzheimer's disease	APP, PS, SORL-1, Cathepsin D	Autophagy defects, accumulation of dense inclusion bodies
Parkinson disease	α -Synuclein, hVps24 (ESCRT-III subunit)	Autophagy defects, MVB/ESCRT pathway defects
Huntington disease	Htt, hVps24 (ESCRT-III subunit)	Autophagy defects, MVB/ESCRT pathway defects
Amyotrophic lateral sclerosis (ALS)	Rab5, hVps2 (ESCRT-III subunit)	Endosomal and MVB sorting defects
Frontotemporal dementia (FTD-3)	hVps2 (ESCRT-III subunit)	Endosomal and MVB sorting defects, accumulation of aberrant autophagosomes
Charcot–Marie–Tooth disease 2B	Rab7 (Ypt7 ortholog)	Endosomal sorting defects
Niemann–Pick disease type C	NPC1, NPC2	Sphingolipid and cholesterol transport defects. Endosomal and autophagy defects
Batten disease	CLN-1, CLN-2, CLN-3, CLN-7	Neuronal lipofuscinosis (NCL). Impaired lysosomal degradation and autophagy defects

nuetrophil granules, and dendrocyte major histocompatibility complex class II compartments (involved in antigen presentation). However, in addition to lysosomal proteins, these organelles contain cell type-specific components that are responsible for their specialized functions.

Pathogen-Containing Vacuoles

The uptake of foreign objects by macrophages often provides an important line of immune defense. In addition, autophagy is a key pathway in the innate immune response, as auto-lysosomes are the sites for degradation of intracellular pathogens. For some intracellular pathogens (including viruses, bacteria, and protozoa), however, phagocytosis represents an opportunity to gain protected access within a host cell. Other types of pathogens, including *Salmonella* and *Shigella*, actively invade host cells by delivering effectors into the host cell that leads to their direct uptake into intracellular vacuoles. Regardless of the mode of entry, the resulting intracellular vacuole that contains the parasite undergoes a maturation process, involving numerous membrane trafficking events, as well as transmembrane transport of nutrients. Maturation of the vacuole yields a unique intracellular environment where the parasites are not only provided with essential nutrients, but also protected from destruction by the host. Interestingly, many bacterial pathogens encode phosphoinositide-binding proteins as well as regulatory proteins to modulate PI3P levels

on pathogen-containing vacuoles, thus impairing fusion with host cell lysosomes.

Diseases Associated with Lysosomal Disorders

Several human genetic disorders are associated with defects in vacuolar/lysosomal hydrolase function and transport, such as I-cell, Tay–Sachs', Pompe's, Gaucher's disease, and Niemann–Pick disease types A and B (Table 1). I-cell disease is manifested by the inappropriate targeting/transport of multiple lysosomal enzymes. Cells of these patients become highly vacuolated and contain numerous dense inclusion bodies. Patients with I-cell disease display severe clinical defects including skeletal and neurological defects, delayed growth and psychomotor development, and often death before age 5. Abnormalities in the AP-3 pathway and traffic to lysosome-related organelles have also been observed in human genetic diseases such as the Chediak–Higashi and Hermansky–Pudlak syndromes (Table 1). Moreover, lysosomes and auto-lysosomes are sites of degradation for mis-folded or damaged transmembrane and cytoplasmic proteins. As such, defects in both the ESCRT and autophagy pathways have been implicated in neurodegenerative diseases (Table 1). The similarity of genes affected in these lysosomal diseases to genes involved in vacuolar transport in yeast further demonstrates the importance of understanding the molecular machinery involved in the biogenesis and function of vacuoles, lysosomes, and lysosome-related organelles.

Further studies on vacuole biogenesis will likely shed light on additional aspects of vesicular transport in the endosomal, lysosomal, and secretory systems.

See also: **Bioenergetics:** Phosphatidylinositol-3-Phosphate; V-ATPases; **Cell Architecture and Function:** Autophagy in Fungi and Mammals; Phagocytosis; **Lipids Carbohydrates Membranes and Membrane Proteins:** Endocytosis; **Signaling:** Rab Family.

Further Reading

- Allen RD and Naitoh Y (2002) Osmoregulation and contractile vacuoles of protozoa. *International Review of Cytology* 215: 351–394.
- Bakowski MA, Braun V, and Brumell JH (2008) Salmonella-containing vacuoles: Directing traffic and nesting to grow. *Traffic* 9: 2022–2031.
- Ballabio A and Gieselmann V (2009) Lysosomal disorders: From storage to cellular damage. *Biochimica et Biophysica Acta* 1793: 684–696.
- Becker B (2007) Function and evolution of the vacuolar compartment in green algae and land plants (Viridiplantae). *International Review of Cytology* 264: 1–24.
- Bowers K and Stevens TH (2005) Protein transport from the late Golgi to the vacuole in the yeast *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta* 1744: 438–454.
- Braulke T and Bonifacino JS (2009) Sorting of lysosomal proteins. *Biochimica et Biophysica Acta* 1793: 605–614.
- Campoy E and Colombo MI (2009) Autophagy in intracellular bacterial infection. *Biochimica et Biophysica Acta* 1793: 1465–1477.
- de Duve C (2005) The lysosome turns fifty. *Nature Cell Biology* 7: 847–849.
- Jones EW (2002) Vacuolar proteases and proteolytic artifacts in *Saccharomyces cerevisiae*. In: Guthrie C and Fink G (eds.) *Methods in Enzymology, Guide to Yeast Genetics and Molecular and Cell Biology, Part C*, vol. 351, pp. 127–150. San Diego, CA: Academic Press.
- Klionsky DJ, Herman PK, and Emr SD (1990) The fungal vacuole: Composition, function, and biogenesis. *Microbiology and Molecular Biology Reviews* 54: 266–292.
- Kornfeld S and Mellman I (1989) The biogenesis of lysosomes. *Annual Review of Cell Biology* 5: 483–525.
- Li SC and Kane PM (2009) The yeast lysosome-like vacuole: Endpoint and crossroads. *Biochimica et Biophysica Acta* 1793: 650–663.
- Marty F (1999) Plant vacuoles. *Plant Cell* 11: 587–600.
- Nixon RA, Yang DS, and Lee JH (2008) Neurodegenerative lysosomal disorders: A continuum from development to late age. *Autophagy* 4: 590–599.
- Saftig P and Klumperman J (2009) Lysosome biogenesis and lysosomal membrane proteins: Trafficking meets function. *Nature Reviews Molecular Cell Biology* 10: 623–635.
- Saksena S, Sun J, Chu T, and Emr SD (2007) ESCRTing proteins in the endocytic pathway. *Trends in Biochemical Sciences* 32: 561–573.
- Weisman LS (2006) Organelles on the move: Insights from yeast vacuole inheritance. *Nature Reviews Molecular Cell Biology* 7: 243–252.
- Wickner W and Schekman R (2008) Membrane fusion. *Nature Structural and Molecular Biology* 15: 658–664.
- Yang Z and Klionsky DJ (2009) An overview of the molecular mechanism of autophag. *Current Topics in Microbiology and Immunology* 335: 1–32.

Relevant Websites

- <http://www.ibioseminars.org> — A library of lectures on vesicular membrane trafficking, intracellular pathogen trafficking, and the innate immune response.
- <http://www.endocytosis.org> — Researching endocytic mechanisms.
- <http://www.yeastgenome.org> — *Saccharomyces* Genome Database (SAC).

Vascular Endothelial Growth Factor Receptors

K A Thomas, Drew University, Madison, NJ, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Angiogenesis Growth of new blood vessels from existing vessels.

Heparan sulfate proteoglycan (HSPG) A membrane or extracellular matrix-anchored sulfated glycoprotein.

Lymphangiogenesis Growth of new lymphatic vessels.

sVEGFR Alternatively spliced soluble vascular endothelial growth factor receptor (VEGFR) missing its transmembrane sequence and intracellular tyrosine kinase.

Vasculogenesis Embryonic *de novo* organization of blood vessels.

Vascular endothelial growth factor (VEGF) An extracellular soluble dimeric protein that binds and activates one or more VEGFRs.

VEGFR Full-length membrane-spanning VEGF receptor containing an intracellular tyrosine kinase.

Vascular Endothelial Growth Factor Receptor Genes

Evolution

The three vascular endothelial growth factor receptors (VEGFRs) are denoted VEGFR-1, VEGFR-2, and VEGFR-3, or Flt-1, KDR/Flk-1, and Flt-4, respectively. Each VEGFR gene encodes a protein composed of seven extracellular immunoglobulin (Ig)-like domains, a short transmembrane-spanning polypeptide, and an intracellular portion containing tyrosine kinase enzymatic activity. The VEGFRs are most closely related to the hematopoietic receptor tyrosine kinases c-kit, c-fms and Flt3, and to the platelet-derived growth factor receptor (PDGFR)- α and - β , each of which contains five extracellular Ig-like domains and an intracellular tyrosine kinase. The VEGFR-1, -2 and -3 genes are clustered with the hematopoietic receptors and PDGFRs on chromosomes 13, 4, and 5, respectively, consistent with divergence from a common ancestral receptor tyrosine kinase gene.

Structure

Each of the three VEGFRs is encoded by 30 exons. The corresponding exon-coding regions in each gene are of similar size although some of the introns that separate them vary substantially in length resulting in total genomic DNAs ranging from ~190 kb for VEGFR-1 to 47 kb and 46 kb for VEGFR-2 and -3, respectively. In all three genes, exon 1 encodes the secretory leader sequence, exons 2–15 span the extracellular region, exon 16 corresponds to the transmembrane-spanning polypeptide, and exons 17–30 encode the cytoplasmic region including the tyrosine kinase containing a kinase-insert-domain encoded by exon 21.

Gene Expression

Cellular differentiation status and responses to extracellular signals influence the expression of each of the VEGFR genes. Multiple transcription-factor-DNA-binding site consensus sequences are present within approximately the first 1 kb 5' of

the translational start site. Control regions within the first intron might also either augment or inhibit transcription.

VEGFR-1

Subsets of vascular endothelial cells, monocytes, dendritic cell precursors, glial cells, and smooth muscle cells express VEGFR-1. The basal promoter and 5' sequences that control VEGFR-1 endothelial cell-selective expression contain Ets, Sp1, Egr-1, and CRE transcription-factor-binding site consensus sequences, several of which have been shown to be functional. In addition, a single hypoxic response element is located within the first 1 kb 5' of the transcriptional start site, consistent with the observation that the transcription of VEGFR-1 is increased by hypoxia. Transcription is initiated downstream of a TATA box basal transcription-factor-binding site to generate a 7.5–8 kb messenger RNA (mRNA).

VEGFR-2

VEGFR-2 is expressed primarily by vascular endothelial cells although a few other cell types including subsets of neurons and their precursors can also express this receptor. Endothelial cell expression has been mapped to a region within the first 150 bp 5' of the transcriptional start site that contains multiple Sp1, Ap-2, and NF- κ B sites. *In vivo* expression of the corresponding murine VEGFR-2 gene has been studied using transgenic mice. Blood-vessel targeted gene expression in developing mouse embryos is controlled by sequences not only in the region 5' of the translational start site but also within the first intron. The human gene, which does not contain a TATA box, is transcribed as a 7-kb mRNA.

VEGFR-3

VEGFR-3 is expressed on lymphatic endothelial cells and the tips of growing capillaries. Several consensus transcription factor sequences have been recognized in the VEGFR-3 gene 5' of the translational initiation site and a 1.6-kb 5' region has been

reported to drive lymphatic expression. Full-length VEGFR-3 mRNA is transcribed as a 5.8-kb mRNA.

Protein Structure

The primary translation products of human VEGFR-1, -2, and -3 genes are 1338, 1356, and 1363 amino acid residue proteins, respectively, each containing a short N-terminal secretory leader sequence that is removed during posttranslational processing. The ~150-kDa protein is glycosylated at multiple sites on the extracellular sequence to generate mature proteins of 185–230 kDa. Partial proteolysis of VEGFR-3 generates an amino-terminal 70-kDa polypeptide disulfide linked to the remaining 120–125-kDa transmembrane protein.

Extracellular Domains

On the basis of sequence homology, the extracellular regions are predicted to be composed of seven Ig-like folding domains, denoted D1–D7 in the amino to carboxyl terminal direction. Crystal structures of the VEGFR-2 D2 and D3 domains complexed with VEGF-C (Figure 1), the isolated D7 domain, and the D2 domain of VEGFR-1 bound either to VEGF-A or to PlGF confirm their Ig-like folding patterns. These β -sandwich Ig-like domains are each composed of two antiparallel β -sheets enclosing a hydrophobic core containing a buried disulfide bridge.

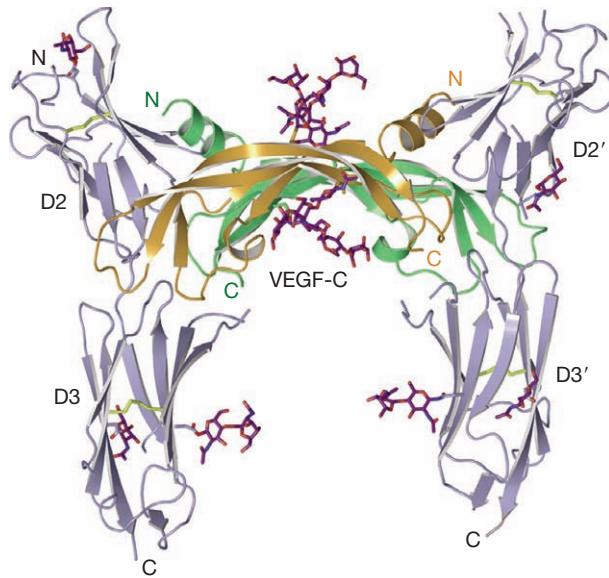


Figure 1 Structure of the complex between a VEGF-C dimer and domains D2 and D3 of two VEGFR-2 receptor fragments. The antiparallel VEGF-A subunits are gold and green and the VEGFR-2 domains are light blue with disulfide bonds and carbohydrate chains shown in yellow and purple, respectively. The amino (N) and carboxyl (C) terminal ends of each subunit are labeled and the secondary structures are illustrated as arrows pointing toward the C-terminal ends. Reproduced from Leppänen V-M, Prota AE, Jeltsch M, et al. (2010) Structural determinants of growth factor binding and specificity by VEGF receptor 2. *Proceedings of the National Academy of Sciences of the United States of America* 107: 2425–2430, with permission from National Academy of Sciences.

Transmembrane and Intracellular Domains

Single-pass membrane-spanning sequences in each of the VEGFRs consist of short, approximately 22–25-amino-acid hydrophobic sequences. This is followed by a juxtamembrane domain, a two-domain tyrosine kinase and a C-terminal polypeptide sequence. The structure of the VEGFR-2 tyrosine kinase, shown in Figure 2, contains N- and C-terminal domains consisting primarily of β -strands and α -helices, respectively. The catalytic site is located at the interface of these two domains denoted by the adenosine diphosphate (ADP) modeled into the structure. It is flanked by the disordered activation-binding loop, the glycine-rich nucleotide-binding loop, and the catalytic loop that can participate in enzyme activation and substrate binding. The VEGFRs each contain a 65–70-amino-acid residue kinase-insert domain within the N-terminal region of the C-terminal kinase domain, which is deleted in the crystallized VEGFR-2 kinase. A C-terminally truncated form of VEGFR-3, missing 65 amino residues, is generated by alternative splicing.

Ligand Binding

VEGF Structure and Receptor Binding

The mammalian VEGF ligands are a family of five homologous dimeric glycoproteins, denoted VEGF-A, VEGF-B, VEGF-C, VEGF-D, and placental growth factor (PlGF). In addition, heterodimers containing VEGF-A and PlGF have been identified. Several functional VEGF homologs are either expressed by

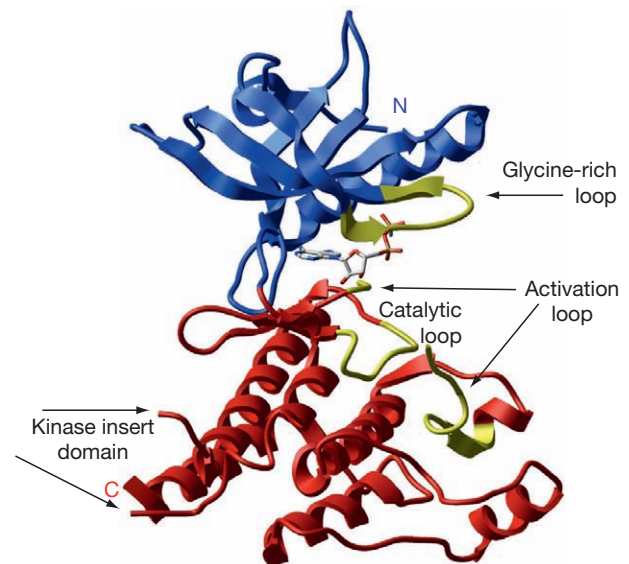


Figure 2 Structure of the VEGFR-2 tyrosine kinase. The N- and C-terminal lobes are blue and red, respectively. The active site region between these two domains is denoted by the location of ADP modeled into it. The glycine-rich, catalytic, and activation loop disordered ends are shown in yellow. The ends of the largely deleted kinase-insert domain are also shown and labeled. Reproduced from Harada S and Thomas KA (2002) Vascular endothelial growth factors. In: Bilezikian JP, Raisz LG, and Rodan GA (eds.) *Principles of Bone Biology*, 2nd edn. vol. 2, pp. 883–902. San Diego, CA: Academic Press, with permission from Academic Press/Elsevier.

viruses or present in snake venoms, collectively denoted VEGF-E and VEGF-F, respectively. Diversity exists among these viral and snake venom VEGFs; some only bind VEGFR-2 whereas others also bind VEGFR-1. The crystal structures of VEGF-A, VEGF-B, VEGF-C, and PlGF homodimers show each ~120-amino-acid monomer to contain a central solvent-exposed antiparallel four-stranded β sheet with a cystine knot motif at one end composed of three intrachain disulfide bonds and one cysteine that forms an interchain disulfide bond with a cysteine at the opposite end of the adjacent antiparallel monomer. As shown in Figure 1, an N-terminal helix extends from the cystine knot end of each subunit and lies over the opposite end of the other subunit to form a small hydrophobic core.

Homodimeric VEGF ligands bind with pM affinity to one or more of the receptor homodimers as shown in Figure 3, in which dashed lines denote proteolytic removal of N- and C-terminal sequences of VEGF-C and -D required for increased affinity to VEGFR-2. Ligand binding promotes receptor dimerization. As seen in Figure 1, each of the two ligand dimeric subunits participate in binding to each of the receptor monomers. The loops between β -strands of both VEGF-C subunits interact with not only the D2 but also the D3 domain of each VEGFR-2 subunit. Similarly, the crystal structures of homodimeric VEGF-A and PlGF complexed at each end to the D2 domain of VEGFR-1 also show that this receptor domain simultaneously binds both ligand subunits. Additional stabilization might be conferred by weaker interactions between the two membrane-proximal D7 and probably the central D4 Ig-like domains in some or all of the VEGFRs, as illustrated in Figure 4. However, D7 interactions are not required for ligand-mediated receptor dimerization, consistent with observed formation of sVEGFR dimers.

Co-receptors

VEGF-A, VEGF-B, and PlGF are subject to alternative splicing that can incorporate C-terminal polycationic regions. These

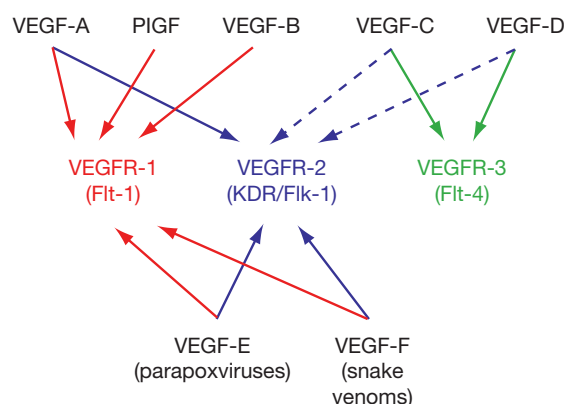


Figure 3 Receptor–ligand binding selectivity. Normal tissue homodimeric VEGFs are listed in the top row and viral VEGF-Es and snake venom VEGF-Fs are present in the bottom row with color-coded arrows pointing toward the high-affinity VEGFRs that they bind. Dashed arrows from VEGF-C and -D to VEGFR-2 indicate high-affinity binding following proteolytic processing to remove N- and C-terminal polypeptides. Some viral VEGF-Es and snake venom VEGF-Fs bind only VEGFR-2, whereas others also bind VEGFR-1.

positively charged sequences promote binding to negatively charged soluble heparin and heparan sulfate proteoglycans (HSPGs) anchored to cellular membranes, extracellular matrix, and basement membranes. Although HSPGs do not induce VEGF signal transduction, they might promote VEGF activity by partitioning the ligands to cellular surfaces thereby increasing their local concentration in the vicinity of the high affinity receptors and by binding both the ligand and receptor. The VEGF/VEGFR complex may also associate with the cell-surface membrane-anchored proteins neuropilin-1, expressed in arteries, and neuropilin-2 found in venous and lymphatic vessels. Neuropilins appear to simultaneously bind the last few amino acid residues of the C-terminal tail of VEGF and heparan sulfate, which also binds the heparin-binding domain of VEGFs and VEGFRs allowing the formation of a VEGF/VEGFR/HSPG/neuropilin complex.

Signal Transduction

Receptor Tyrosine Kinase Activation

Receptor dimerization is necessary but not sufficient for the initiation of signaling. Formation of a pair of salt bridges between the receptor membrane-proximal D7 domains of a dimerized receptor is also required for transphosphorylation of each intracellular VEGFR tyrosine kinase by its dimeric partner as shown schematically in Figure 4. Interactions between adjacent D7 domains and perhaps the membrane-spanning sequences within receptor dimers presumably facilitate proper proximity and orientation of the intracellular tyrosine kinase domains required for mutual transphosphorylation. The initial transphosphorylation of tyrosine residues 1054 and 1059 on the tyrosine kinase activation loop of VEGFR-2 decreases the K_M of the enzyme for adenosine triphosphate (ATP) and for peptide substrates without altering K_{cat} , the effective maximal rate of enzyme activity. This increased affinity for substrates could reflect the movement of the phosphorylated activation loop to provide better substrate access to their binding sites. Activation loop tyrosines exist on VEGFR-1 and -3, so might function in a similar manner. Once activated, several other VEGFR tyrosine residues become phosphorylated. VEGF-induced receptor tyrosine phosphorylation is rapid, reaching maximal levels by 2–5 min, followed within 30 min by receptor internalization.

Activation of VEGF receptors can be inhibited by secreted soluble forms of VEGFR-1 (sVEGFR-1 or sFlt-1) and VEGFR-2 (sVEGFR-2), each of which consists of the six N-terminal Ig-like domains, as illustrated in Figure 4, and are generated by alternatively splicing. These soluble receptors can sequester or ‘trap’ VEGF ligands in a dimeric soluble receptor complex so they are not available to bind full-length transmembrane VEGFRs. In addition, they may dimerize with full-length membrane-spanning VEGFRs, thereby preventing receptor kinase transphosphorylation and activation thus acting in a dominant negative manner.

Recruitment of Signaling Proteins

Once the receptor tyrosine kinase enzyme is activated, additional tyrosines in the intracellular juxtamembrane region, the

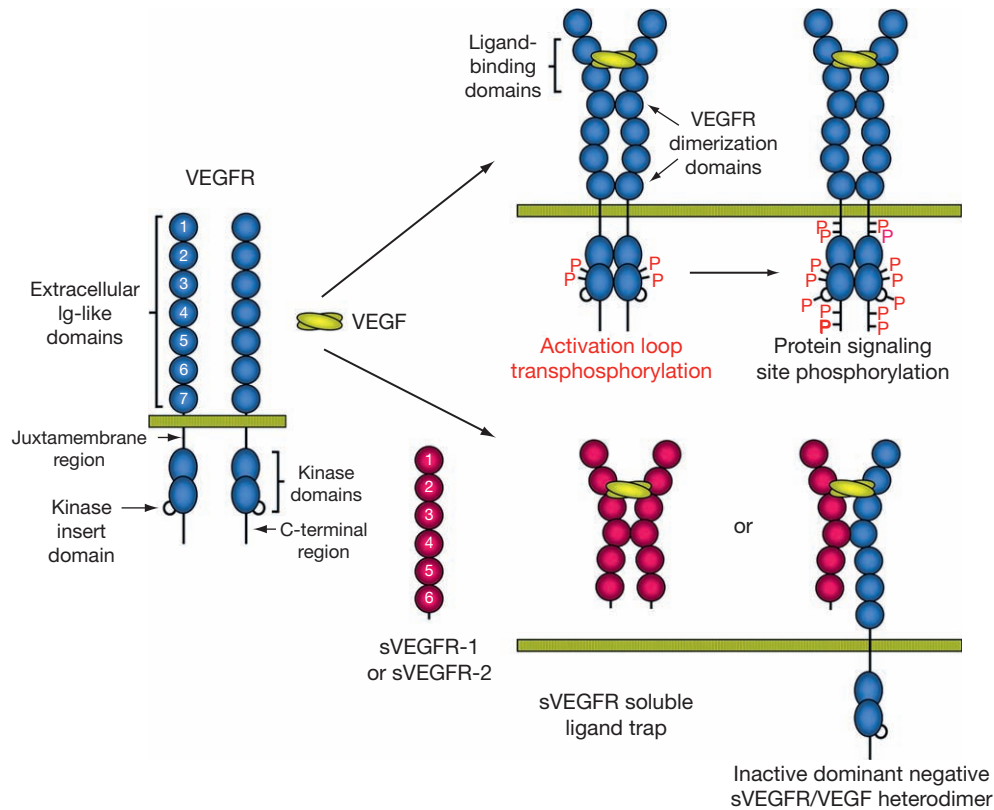


Figure 4 VEGFR activation and inhibition by sVEGFR. The structures of VEGFRs contain seven extracellular Ig-like domains, a juxtamembrane region, the two-domain core tyrosine kinase and kinase insert domain, and a C-terminal tail (labeled on the left). Binding of dimeric VEGF to Ig-like domains 2 and 3 induces receptor dimerization, which can be aided by interactions between receptor domains 7 and probably 4 (shown in the upper right). This promotes phosphorylation of two kinase activation loop tyrosines, denoted by the red P labels, and enzymatic activation in VEGFR-2 and probably in VEGFR-1 and -3. Additional tyrosines in the juxtamembrane region, kinase-insert loop, and C-terminal tail are phosphorylated and bind signaling proteins that trigger several VEGF functions. The sVEGFRs can inhibit VEGFR activation by sequestering ligand and by forming dominant negative dimeric complexes with the full-length membrane-spanning receptors (lower right).

kinase insert domain, and the C-terminal tail can be phosphorylated, as shown schematically in Figure 4. Which subsets of tyrosines are phosphorylated on a particular VEGFR subunit can be influenced by whether it is in a homodimeric or heterodimeric receptor complex, which, in turn, influences the spectrum of downstream signaling events and biologic responses that are initiated. Phosphorylated receptor tyrosine residues can serve as recognition sequences for binding one or more signal transduction proteins including phospholipase $C\gamma$ ($PLC\gamma$) and several adapter proteins such as Grb2, Nck, Crk, Sck, Shc, Shb, and Vrap/TSAd that can bridge the receptors to signaling enzymes such as phosphatidylinositol 3-kinase (PI3K) and src.

Signal Transduction Cascades

Some of these phosphorylated tyrosine residues and their binding proteins have been linked to specific functions, including proliferation, migration, cell survival, and vascular permeability. For example, phosphorylation of VEGFR-2 Tyr₁₁₇₅ in the C-terminal tail is required for the efficient binding of $PLC\gamma$. The phosphorylation of VEGFR-2-associated $PLC\gamma$ activates its enzymatic activity catalyzing the hydrolysis of membrane-anchored phosphatidylinositol 4,5-bisphosphate to generate

inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG). Water-soluble IP_3 activates an endoplasmic reticulum ion channel that releases Ca^{2+} , activating several enzymes including endothelial cell nitric oxide synthase, which promotes vascular permeability, and promoting the translocation of specific protein kinase C (PKC) isoforms to the membrane where they are activated by binding membrane-associated DAG. Activated PKC then activates the mitogen-activated protein kinase (MAPKK; also known as MEK), which, in turn, activates p42/44 MAPK that then enters the nucleus to modulate transcription. Additional pathways, mediated through PI3K, raf, src, and p38, activate cell migration, proliferation, survival, and vascular permeability.

Biologic Activities

Development

Each of the three VEGFRs are crucial for embryonic development as revealed by mouse homozygous gene knockouts that are each embryonically lethal. Blood vessels exist in VEGFR-1 knockout mice but are disorganized, which appears to be a consequence of modified hematopoietic cell fate leading to increased numbers of endothelial cell progenitor hemangioblasts that

alter vessel pattern formation. Surprisingly, knockout of the VEGFR-1 gene segment encoding the tyrosine kinase but retention of the extracellular and membrane-spanning regions leads to normal development of blood vessels and survival in approximately 50% of mice. This result is consistent with the possibility that one of the functions of VEGFR-1, which binds VEGF-A with ~10-fold higher affinity than VEGFR-2, might be to sequester low levels of VEGF, thereby buffering the VEGFR-2 receptor from inappropriate activation by low levels of VEGF. By contrast, the VEGFR-2 knockout is virtually devoid of vascular endothelial cells, consistent with its crucial role as a mediator of endothelial cell mitosis and survival. VEGFR-3 gene knockout mice exhibit vasculogenesis and angiogenesis in early developing embryos but large vessels appear abnormal with defective lumens leading to fluid accumulation in the pericardial cavity and cardiovascular failure. Therefore, at least during early embryogenesis, before lymphatic vessels develop from post-capillary venules, VEGFR-3 plays a crucial role in vascular development.

Neovascularization

In adults, activation of VEGFR-2 can mediate the survival, migration, and mitosis of vascular endothelial cells culminating in microvascular growth, termed 'angiogenesis'. Ligands such as VEGF-A and the viral VEGF-Es that bind VEGFR-2 are also sufficient to drive angiogenesis and the growth of hematopoietic progenitor cells. Antibodies to VEGFR-2 and enzyme inhibitors of the VEGFR-2 tyrosine kinase inhibit angiogenesis and the resulting growth of tumors in mice. VEGFR-3, which can form homodimers and heterodimers with VEGFR-2, is expressed in angiogenic sprouts and stimulation of VEGFR-3 can augment VEGF-induced angiogenesis, whereas VEGFR-3 antibodies suppress angiogenic sprouting and vascular network formation. By contrast, the VEGFR-1-selective ligand PlGF appears to promote the stabilization of existing microvessels by promoting pericyte coverage, maturation of precapillary arterioles by stimulation of smooth muscle cell growth, and development of larger collateral vessels, termed 'arteriogenesis', in both ischemic myocardium and limb muscles resulting in increased perfusion. Activation of VEGFR-1 by its selective ligands PlGF and VEGF-B can also induce vascular endothelial cell expression of proteins such as tissue factor, matrix metalloproteinases, urokinase, plasminogen activator inhibitor-1, hepatocyte growth factor, and pigment epithelium-derived factor that can modulate vascular growth.

The directional growth of new capillary sprouts is driven by VEGF-A gradients stabilized at least in part by binding to extracellular matrix components such as heparan proteoglycans. In addition, growing sprouts are guided away from the parental vessel by its local expression of sVEGFR-1, which can sequester VEGF-A forming an active angiogenic corridor perpendicular to the preexisting vessel, thus inhibiting formation of nonproductive microvascular loops back to the parental vessel. Natural expression of sVEGFR-1 in cornea is also necessary for the maintenance of corneal avascularity and clarity. Expression of transfected sVEGFR-1 by tumor cells severely inhibits their growth *in vivo* but not *in vitro*, consistent with an antiangiogenic mechanism. The VEGFR-2/VEGF-A system can promote tumor angiogenesis and growth and has been the focus of therapeutic anti-angiogenesis efforts. Bevacizumab, an

inhibitory monoclonal antibody that binds VEGF-A, has been approved for use as an adjunct to cancer chemotherapy.

Lymphangiogenesis

The lymphatic system, which develops during embryogenesis from the vascular venous system, drains excess extracellular fluids and transports antigen-presenting immune cells. VEGFR-3 is expressed by lymphatic endothelial cells and its activation supports their migration, proliferation, and survival. Consistent with these cellular activities, the VEGFR-3 ligands VEGF-C and -D can promote lymphangiogenesis, the growth of lymphatic capillaries. Homozygous gene knockouts of either VEGFR-3 or VEGF-C demonstrate that they are both required for the development of a functional lymphatic system. Naturally occurring human heterozygous VEGFR-3 missense mutations with inactive tyrosine kinases are linked to lymphedema, a genetic disease in which fluid accumulates in tissues as a consequence of deficient lymphatic function. Endogenous corneal expression of the naturally occurring soluble receptor sVEGFR-2 antagonizes VEGF-C and inhibits corneal lymphatic development, which prevents the trafficking of antigen-presenting cells and maintains immunologic privilege of the cornea *in vivo*. Therefore, the establishment of an optically clear and immunologically privileged cornea is achieved by endogenous expression of both sVEGFR-1 and sVEGFR-2. Surprisingly, sVEGFR-2 does not efficiently inhibit angiogenesis, which might indicate that it is functioning by formation of heterodimers with VEGFR-3. Heterodimerization of full-length VEGFR-2 and VEGFR-3 has been observed in lymphatic endothelial cells and results in phosphorylation of a different subset of receptor tyrosine phosphorylation sites compared to the corresponding homodimeric receptors. In animal models, VEGF-C and VEGF-D stimulate lymphatic metastasis and their expression in human tumors is correlated with cancer metastatic spread to lymph nodes and a poor clinical prognosis.

Nonmitogenic Functions

In adults, activation of VEGFR-2 induces hypotension and vascular permeability that can promote edema. By contrast, low circulating levels of the unbound VEGFR-2 ligands VEGF-A and PlGF and elevated levels of sVEGFR-1, which can sequester these ligands, is associated with preeclampsia, a condition manifest by hypertension and edema that occurs in 5–7% of pregnancies and causes increased fetal mortality. VEGF-A also promotes the synthesis of prostacyclin, a potent vasodilator and inhibitor of platelet aggregation, through formation of VEGFR-1/VEGFR-2 heterodimers but not the corresponding receptor homodimers.

The VEGFR system is also associated with inflammation. Activation of VEGFR-1 can induce the migration of monocytes and drive expression of monocyte and macrophage tissue factor, monocyte chemoattractant protein-1, and tumor necrosis factor- α . Inhibitory anti-VEGFR-1 antibodies suppress the mobilization of bone marrow myeloid progenitors into the circulation, the infiltration of leukocytes into inflamed tissues, and the resulting development of atherosclerotic and autoimmune arthritic lesions in mice.

All three VEGFRs have been found in the nervous system along with their ligands. In addition to their neurovascular

effects, the VEGF receptors and their ligands can promote neurogenesis, neuronal migration, neurite growth, and survival in a variety of cell culture and animal models. The VEGF/VEGFR system can also support the proliferation and function of a variety of glial cells including astrocytes, myelinating Schwann cells, and microglia.

Summary

The VEGFRs are composed of three homologous membrane-spanning tyrosine kinases that on ligand-induced dimerization not only mediate the development, growth, and function of blood and lymphatic vessels but also appear to be involved in certain aspects of the immune and nervous systems. Alternatively spliced inhibitory soluble receptors corresponding to two of the VEGFRs appear to provide endogenous specific inhibition of these activities. The inappropriate activation or inhibition of the VEGFRs and their ligands can have pathologic consequences making them targets for therapeutic intervention.

See also: **Lipids Carbohydrates Membranes and Membrane Proteins:** Proteoglycans; **Metabolism Vitamins and Hormones:** Biochemistry of Hematopoiesis; Regulation of Gene Transcription by Hypoxia-Inducible Factor 1; **Molecular Biology:** Alternative Splicing; **Signaling:** Inositol Phosphate Kinases and Phosphatases; Mitogen-Activated Protein Kinase Family; Nitric Oxide Signaling; Phospholipase C; Protein Kinase C Family; Protein Tyrosine Phosphatases; Src Family of Protein Tyrosine Kinases.

Further Reading

- Albuquerque RJC, Hayashi T, Cho WG, et al. (2009) Alternatively spliced vascular endothelial growth factor receptor-2 is an essential endogenous inhibitor of lymphatic vessel growth. *Nature Medicine* 15: 1023–1030.
- Ambati BK, Nozaki M, Singh N, et al. (2006) Corneal avascularity is due to soluble VEGF receptor-1. *Nature* 443: 993–997.
- Autiero M, Waltenberger J, Communi D, et al. (2003) Role of PlGF in the intra- and intermolecular cross talk between the VEGF receptors Flt1 and Flk1. *Nature Medicine* 9: 936–943.
- Chappell JC, Taylor SM, Ferrara N, and Bautch VL (2009) Local guidance of emerging vessel sprouts requires soluble Flt-1. *Developmental Cell* 17: 377–386.
- Dixelius J, Makinen T, Wirzenius M, et al. (2003) Ligand-induced vascular endothelial growth factor receptor-3 (VEGFR-3) heterodimerization with VEGFR-2 in primary lymphatic endothelial cells regulates tyrosine phosphorylation sites. *Journal of Biological Chemistry* 278: 40973–40979.
- Ferrara N, Gerber H-P, and LeCouter J (2003) The biology of VEGF and its receptors. *Nature Medicine* 9: 669–676.
- Ferrara N, Hillan KJ, Gerber H-P, and Novotny W (2004) Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nature Reviews Drug Discovery* 3: 391–400.
- Harada S and Thomas KA (2002) Vascular endothelial growth factors. In: Bilezikian JP, Raisz LG, and Rodan GA (eds.) *Principles of Bone Biology*, 2nd edn., vol. 2, pp. 883–902. San Diego: Academic Press.
- Leppänen V-M, Prota AE, Jeltsch M, et al. (2010) Structural determinants of growth factor binding and specificity by VEGF receptor 2. *Proceedings of the National Academy of Sciences of the United States of America* 107: 2425–2430.
- Lohela M, Bry M, Tammela T, and Alitalo K (2009) VEGFs and receptors involved in angiogenesis versus lymphangiogenesis. *Current Opinion in Cell Biology* 21: 154–165.
- Luttun A, Tjwa M, Moons L, et al. (2002) Revascularization of ischemic tissues by PlGF treatment, and inhibition of tumor angiogenesis, arthritis and atherosclerosis by anti-Flt-1. *Nature Medicine* 8: 831–840.
- McTigue MA, Wickersham JA, Pinko C, et al. (1999) Crystal structure of the kinase domain of human vascular endothelial growth factor receptor 2: A key enzyme in angiogenesis. *Structure* 7: 319–330.
- Neaogoe P-E, Lemieux C, and Sirois MG (2005) Vascular endothelial growth factor (VEGF)-A₁₆₅-induced prostacyclin synthesis requires the activation of VEGF receptor-1 and -2 heterodimer. *Journal of Biological Chemistry* 280: 9904–9912.
- Olsson A-K, Dimberg A, Kreuger J, and Claesson-Welsh L (2006) VEGF receptor signaling – in control of vascular function. *Nature Reviews Molecular Cell Biology* 7: 359–371.
- Ruiz de Almodovar C, Lambrechts D, Mazzone M, and Carmeliet P (2009) Role and therapeutic potential of VEGF in the nervous system. *Physiological Reviews* 89: 607–648.
- Tammela T, Zarkada G, Wallgard E, et al. (2008) Blocking VEGFR-3 suppresses angiogenic sprouting and vascular network formation. *Nature* 454: 656–660.
- Vander Kooi CW, Jusino MA, Perman B, et al. (2007) Structural basis for ligand and heparin binding to neuropilin B domains. *Proceedings of the National Academy of Sciences of the United States of America* 104: 6152–6157.
- Weismann C, Fuh G, Christinger HW, et al. (1997) Crystal structure at 1.7 Å resolution of VEGF in complex with domain 2 of the Flt-1 receptor. *Cell* 91: 695–704.
- Yamazaki Y, Matsunaga Y, Tokunaga Y, et al. (2009) Snake venom vascular endothelial growth factors (VEGF-Fs) exclusively vary their structures and functions among species. *Journal of Biological Chemistry* 284: 9885–9891.
- Yang Y, Xie P, Opatowsky Y, and Schlessinger J (2010) Direct contacts between extracellular membrane-proximal domains are required for VEGF receptor activation and cell signaling. *Proceedings of the National Academy of Sciences of the United States of America* 107: 1906–1911.

Relevant Websites

- <http://www.ncbi.nlm.nih.gov> – NCBI GenBank DNA and protein sequences of VEGFRs and VEGFs.
- <http://www.ncbi.nlm.nih.gov/projects/genome/guide/human> – NCBI Human Genome Resources.
- <http://www.rcsb.org> – Protein Data Bank: An Information Portal to Biological Macromolecular Structures.

Vasopressin/Oxytocin Receptor Family

M J Brownstein, Azevan Pharmaceuticals, Bethlehem, PA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Baroreceptor Receptor in the walls of the heart or blood vessels that is stimulated by alterations in pressure.

Dieresis Increased urine excretion.

Diuretic An agent that increases urine excretion.

Hypothalamus The part of the brain that regulates the endocrine and autonomic and autonomic nervous systems, controlling water balance, blood pressure, body temperature, growth, and sexual function.

Lipogenesis Formation of body fat.

Magnocellular neurons Large neurons in the hypothalamus that manufacture vasopressin or oxytocin.

Natriuresis Sodium excretion in the urine.

Pressin An agent that increases blood pressure. In the context of this review, pressins are vasopressin-like peptides with nine amino acids, having a basic residue (arginine or lysine in the eighth position).

Tocin An agent that promotes childbirth by causing uterine contractions; an oxytocin-like peptide lacking arginine or lysine in the eighth position.

The Structures of Pressins and Tocins

For the last billion years, the peptides in vasopressin/oxytocin (VP/OT) family have retained certain structural features. With the exception of the hydrins (partially processed relatives of vasotocin, which are physiologically active in some amphibia) and Ci-VP (CFFRDCSNMDWYR, an unusual, divergent family member made by sea squirts), they all have nine amino acids and a C-terminal amide. Cysteine (C) is invariably present in positions 1 and 6, and the two cysteine residues form a cystine bridge, creating conformationally constrained, cyclic peptides. Except for Ci-VP and seritocin (serine5 and isoleucine8-OT), which was reported in a single species, *Bufo regularis*, all family members have asparagines (N) in position 5, and proline (P) and glycine (G) in positions 7 and 9, respectively. VP-like peptides (pressins) are characterized by the presence of arginine (R) or lysine (K) in position 8, while OT-like molecules (tocins) have leucine (L), isoleucine (I), glutamine (Q), valine (V), or threonine (T) there. All of the invertebrate peptides except for the locust diuretic hormone, which has not been detected in other insect species, have an R instead of Q, N, or S in position 4. Thus, the primordial peptide that gave rise to the molecules we know today should have looked rather similar to one of the conopressins or to vasotocin, the likely ancestor of all of the vertebrate hormones. (The molecules detected in *Hydra* with antivasopressin antibodies share the C-terminal PRG-amide motif with VP, but are otherwise unrelated.)

In addition to the peptides shown in Table 1 and Figure 1, it is worth noting that the major metabolite of VP (pGlu4 and cystine6-AVP) mobilizes calcium in some cells in the central nervous system, and has been suggested to have a unique, but as yet unidentified, receptor of its own.

Peptide Biosynthesis

The peptides in the VP/OT family are synthesized as parts of precursor proteins. The VP precursor has the following structural features:

signal peptide-CYFNQNCPRGG-K*R*neurophysin-copeptin

The signal peptide is removed cotranslationally in the endoplasmic reticulum, liberating the N terminus of the peptide. Then, after the precursor is packaged into vesicles by the Golgi apparatus, it is cleaved at the asterisked sites by a calcium-dependent Lys-Arg endoprotease. The remaining basic amino acids are removed by a carboxypeptidase B-like enzyme, leaving a glycine (G) on the carboxy terminus. This glycine serves as the donor of the amide group, which is generated by the sequential actions of a peptidyl-glycine monooxygenase and a peptidyl-hydroxyglycine lyase. Neurophysin is thought to serve as an internal chaperone, promoting proper folding and cystine-bridge formation. Copeptin, a glycosylated species, is part of the VP, but not the OT precursor. Its function is unknown, but since it is produced and released along with VP, its levels in the blood reflect those of the peptide, and it is a convenient biomarker for VP secretion. The genes encoding the VP and OT precursors are adjacent to one another on the same chromosome. In mammals, they are oriented tail to tail, and share regulatory elements. The gene elements responsible for cell-specific expression of the two peptides remain to be determined.

The Functions of VP and OT

VP/OT-like peptides play different roles in the various species where they are found. They are very commonly involved in water homeostasis and/or reproductive function. In humans, VP is an antidiuretic hormone. It is released from the posterior pituitary into the blood stream when the osmolarity (solute content) of the blood increases, and it acts on V2 receptors in the kidney to cause water retention. V2 receptors are Gs-coupled and increase intracellular cyclic AMP. This, in turn, causes preformed aquaporin (AQP) 2 channels to be inserted into the luminal surfaces of cells in the collecting ducts of the kidney. Water enters these channels, traverses the cells, exits through AQP 3 and 4 channels on their interstitial faces, and enters the bloodstream. Mutations in the genes encoding VP, the V2 receptor, or AQP2 cause diabetes insipidus – an inability to concentrate urine, resulting in high urine volumes and a

Table 1 Vasopressin, oxytocin, and related peptides

<i>Oxytocin-like peptides (tocins)</i>		
Oxytocin (OT)	CYIQNCPLG-amide	Mammals, Pacific ratfish
Mesotocin (MT)	CYIQNCPIG-amide	Nonmammalian tetrapods, marsupials, lungfish
[Phe2]Mesotocin (FMT)	CFIQNCPIG-amide	Australian lungfish
Seritocin (ST)	CYIQSCPIG-amide	Dryness-resistant African toad
Isotocin (IT)	CYISNCPIG-amide	Bony fish
Aspartocin (AspT)	CYINNCPLG-amide	Spiny dogfish
Asvatocin (AsvT)	CYINNCPLG-amide	Spotted dogfish
Glumitocin (GT)	CYISNCPLG-amide	Rays
Phasvatocin (PhT)	CYFNCPVG-amide	Spotted dogfish
Valitocin (ValT)	CYIQNCPLG-amide	Spiny dogfish
Annetocin (AnT)	CFVRNCPTG-amide	Earthworm
Cephalotocin (CT)	CYFRNCPIG-amide	Octopus
<i>Vasopressin-like peptides (pressins)</i>		
A-Vasopressin (AVP)	CYFQNCPRG-amide	Most mammals
L-Vasopressin (LVP)	CYFQNCPLG-amide	Pig, marsupials
Phenylpressin (PP)	CYFQNCPRG-amide	Marsupials
Vasotocin (VT)	CYIQNCPRG-amide	Non-mammalian vertebrates
Hydrin 2 (H2)	CYIQNCPLG	Frogs, toads
A-conopressin (ACP)	CIIRNCPLG-amide	Snail
L-conopressin (LCP)	CFIRNCPLG-amide	Snail, sea hare, leech
Locust diuretic hormone (LDH)	CLITNCPLG-amide	Locust (but not fruitfly), Red flour beetle, Jewel wasp
<i>Putative ancestor</i>	<i>C(F/Y)I(Q/R)NCP(R/K)G-amide</i>	

Abbreviations: C (cysteine), F (phenylalanine), G (glycine), I (isoleucine), K (lysine), N (asparagine), P (proline), Q (glutamine), R (arginine), S (serine), T (threonine).

need to consume large amounts of fluid each day. Since the V2 receptor gene is on the X chromosome, female carriers pass X-linked nephrogenic diabetes insipidus along to their sons, but not their daughters.

Nonosmotic as well as osmotic stimuli can trigger VP secretion. Osmoreceptors on cells in the anterior hypothalamus, and baroreceptors in the left ventricle of the heart, aortic arch, and carotid sinus monitor osmolality and blood pressure, respectively, and participate in controlling the firing of VP-producing cells during phenomena like hemorrhage.

In addition to V2 receptors, VP acts on V1a and V1b receptors. The former are found on blood vessels and are responsible for VP's pressor (blood pressure increasing) activity. In fact, VP and terlipressin, a V1a specific agonist, are used to treat catecholamine-resistant vasodilatory shock. Since certain vascular beds are especially sensitive to VP, including those in the skin and uterus, it has also been suggested that V1a receptor antagonists might be useful for treating vasospasm in Raynaud's disease and dysmenorrhea (menstrual cramps) which may result from excessive constriction of vessels in the digits and uterus, respectively.

V1b receptors are found on corticotrophs (adrenocorticotrophic hormone (ACTH) or ACTH-producing cells) in

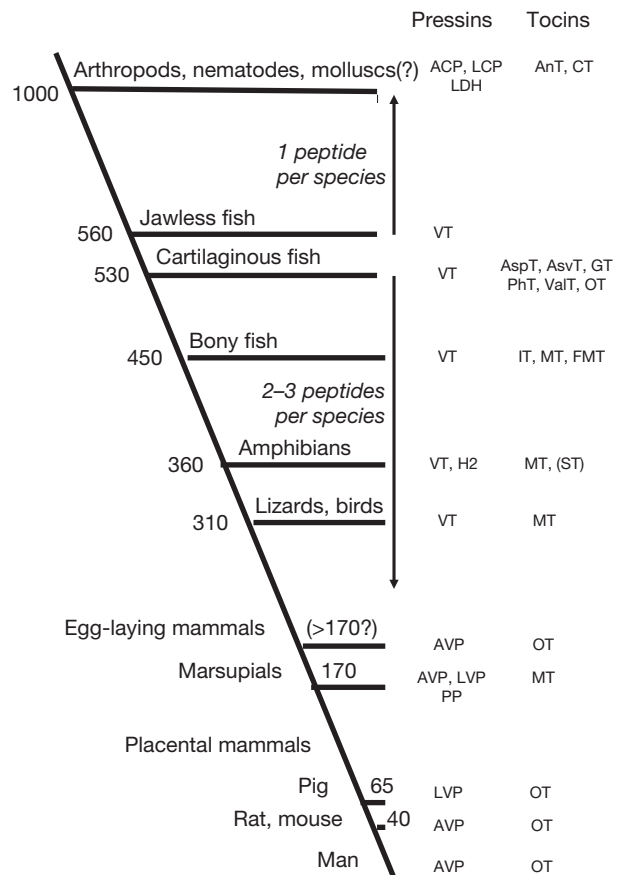


Figure 1 Evolution of vasopressin-related (pressin) and oxytocin-related (tocin) peptides. The timescale (millions of years ago) was taken from Hedges (2002). Note that invertebrates and jawless fish make a single pressin or tocin per species, and that vertebrates make two or, in the case of marsupials, three such peptides – one or two pressins and one tocin. It appears that vasotocin was the ancestor of the vertebrate hormones, and it continued to be synthesized by all vertebrates except mammals, in which it was replaced by arginine or lysine vasopressin. It is not clear why cartilaginous fish make so many different tocins; lungfish make MT or FMT, and all of the remaining bony fish make IT. Based on the structures of the peptide precursors, it has been argued that lungfish are related to tetrapods (e.g., toads) and that bony fish form a separate lineage. Amphibians, lizards, birds, and marsupials make MT; egg-laying and placental mammals differ in the hormones they produce. Marsupials appear to have branched away as suggested by other analyses. It is worth noting that the receptors for the various pressins and tocins shown in this figure evolved in parallel with them. The abbreviations used in this figure are defined in Table 1.

the anterior pituitary. In addition to being made by magnocellular neurons, VP is also synthesized by small cells in the paraventricular nucleus of the hypothalamus. These same cells also make corticotropin-releasing hormone (CRH), and their axons terminate on portal vessels in the external zone of the median eminence. VP and CRH are released into these vessels and transported to the anterior pituitary, where they act in concert to stimulate ACTH secretion. ACTH, in turn, causes the adrenal to release glucocorticoids. Thus, VP is important in mediating the body's response to stress.

V1a and V1b, but not V2, receptors are found in the brain, and the central effects of VP agonists and antagonists are mediated by these receptors. These effects vary from species to species, and learning their roles in humans should be facilitated by the recent development of specific agonists and antagonists that can be used for imaging and clinical studies/trials. It appears that VP may mediate behavioral reactions to stress, aggressive and affiliative behavior, juvenile recognition, and parenting in rodents – especially males. In addition, V1a antagonists appear to block the vomiting associated with motion sickness in pigs and ferrets.

OT-producing mammals make a single receptor for this peptide. Like the V1a and V1b receptors, it is Gq-coupled, and activates phospholipase C (PLC), increasing intracellular calcium. The VSSV/I sequence, which is unique to V1a, V1b, and OT receptors, may participate in PCL coupling.

OT receptors increase dramatically in the pregnant uterus as term approaches. Despite the fact that it has been used for decades to induce uterine contractions, studies of knockout animals have shown that OT is not essential for parturition. This may explain why OT receptor antagonists are not always effective in treating preterm labor, especially when there is an intrauterine infection present. There is no doubt, however, that the pulsatile release of OT from the pituitary in response to cervical dilation and vaginal stimulation facilitates the expulsion of the fetus.

OT is required for milk ejection. Mechanical stimulation of pressure-sensitive receptors in the nipple of the breast by the nursing infant results in activation of magnocellular neurons in the hypothalamus and release of pulses of OT into the bloodstream. The hormone causes breast myoepithelial cells to contract, increasing intramammary pressure and forcing milk into the ducts. In the absence of OT, milk cannot be let down, and the infant will starve if it is not provided an alternative source of food.

In addition to its roles in parturition and lactation, OT appears to affect maternal and social behaviors. Based on the latter, its use in autism is being explored. In addition, it stimulates lipogenesis to compensate for lipid loss in the milk (via an action on insulin secretion) and may participate in regulating salt and water balance. While OT causes natriuresis in rats, it is not clear that this is the case in humans.

VP and OT Receptors

As expected from the fact that their ligands are similar, VP and OT receptors are structurally related. They are members of the rhodopsin superfamily, and have seven α -helical membrane-spanning domains connected to one another by intracellular and extracellular loops. The N terminus of each receptor faces the outside of the cell; the C terminus is cytoplasmic. The intracellular loops and C-terminal tail of the receptors interact with trimeric G proteins, coupling agonist binding to activation of second messenger systems. More than 40 VP/OT receptors found in species ranging from snails to humans have been cloned and sequenced. The primary sequences of some of these are shown in [Figure 2](#). It is remarkable that from the beginning of its first transmembrane domain (TM1) to the end of its seventh one (TM7), the snail conopressin receptor 2 is

43% identical in amino acid sequence to the human V1a, V1b, and OT receptors and the white sucker fish vasotocin receptor. Unlike the vertebrate proteins, however, the conopressin receptor responds equally well to lysine8- and isoleucine8-conopressin (an OT-like synthetic analog of lysine-conopressin). Duplication of a relatively promiscuous receptor of this sort might have permitted trial-and-error evolution of functionally distinct pressors and tocins in vertebrates.

A number of amino acids are conserved among all of the receptors in [Figure 2](#). Some of these residues are found in most G-protein-coupled receptors (GPCRs). Among them, the arginine (R2) in the DRY motif, found just beneath TM3, is thought to dwell in a pocket formed by polar residues in TMs 1, 2, and 7 when the receptor is in its inactive state. Hormone binding dislodges this arginine from its polar pocket, exposing G-protein docking sites on the cytoplasmic loops.

The cysteines in extracellular loops 1 and 2 (C1 and C2, respectively) are also highly conserved among rhodopsin-like receptors. They form a cystine bridge that links these loops, stabilizing the conformation of the receptors. The pair of cysteines (C3 and its neighbor) located 15 aa's below TM7 in the cytoplasmic tails of most VP/OT receptors are palmitoylated, and this is thought to anchor their C termini to the plasma membrane and play a role (in the case of the V2 receptor) in beta-arrestin binding, which leads to efficient receptor endocytosis and ERK1/ERK2 activation. Like other members of the rhodopsin superfamily, VP and OT receptors appear to be glycosylated on their N termini and regulated by phosphorylation of their intracellular domains.

A number of attempts have been made to model the binding of VP, OT, and vasotocin to their receptors. The models are fundamentally similar in the sense that they all predict that the peptide hormones fill a cleft located in the upper third of the barrel formed by the seven membrane-spanning α -helices. The hydrophobic amino acids that comprise the cyclic portion of the peptides (cysteine1, tyrosine2, isoleucine or phenylalanine3, and cysteine6) appear to reside in a hydrophobic pocket formed by aromatic residues on helices 3, 5, 6, and 7. The more polar amino acids (glutamine4 and asparagine5), and the amidated C terminus of the hormones occupy a hydrophilic region formed by residues on helices 2, 3, and 4. More specifically, residues that are conserved in the N-terminal domain and TMs 2, 3, 4, and 6 in most, if not all, of the VP/OT receptors (labeled R1, Q1, Q2, K, Q3, Q4, and Q5 in [Figure 2](#)) have been shown to be important for high-affinity binding, even though their predicted interactions with specific amino acids in the peptide hormones vary from model to model. Parsimony dictates that residues conserved among the various VP/OT peptides should interact with conserved domains in the receptors, but perhaps things are not this simple.

One amino acid appears to be crucial for peptide agonist selectivity. This residue (marked with a V in [Figure 2](#)) is found in the first extracellular loop, and it interacts with the eighth amino acid of the peptide hormones (arginine and leucine in AVP and OT, respectively). The aspartic acid or tyrosine residues found at V in the human V2 and V1a/V1b receptors, respectively, are responsible for their marked preferences for AVP versus OT. The phenylalanine in this position in the OT receptor accounts for its modest preference for OT over AVP.

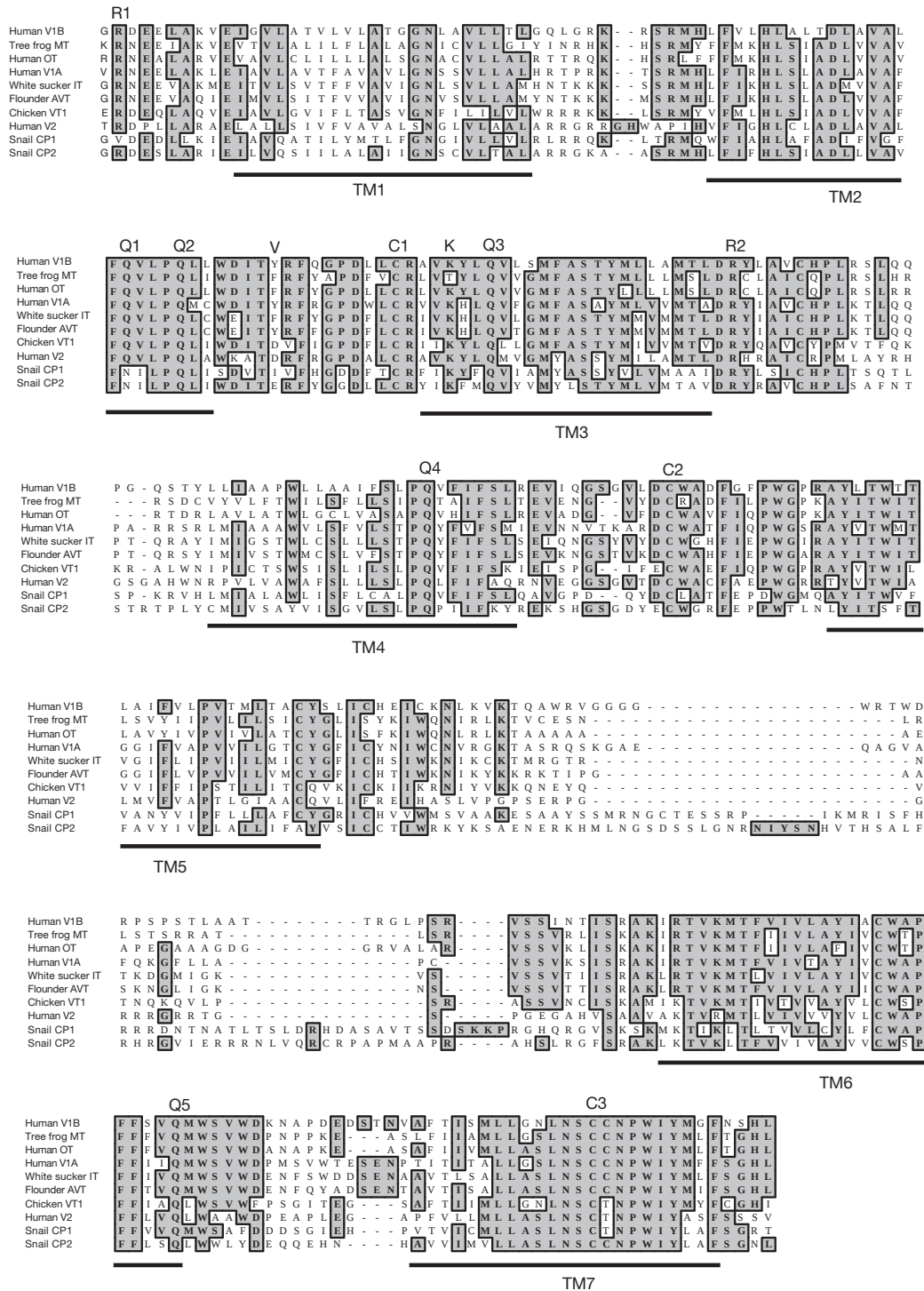


Figure 2 Alignment of vasopressin and oxytocin receptors and selected relatives. To save space, the extracellular N termini and intracellular C termini have been removed. They are quite divergent, but it is remarkable that the transmembrane domains (TMs), the first two extracellular loops (linking TMs 2 and 3, and TMs 4 and 5), and portions of intracellular loops 2 and 3 (linking TMs 3 and 4, and TMs 5 and 6), have remained so similar throughout the course of their evolution. Specific residues in these domains are responsible for ligand binding and selectivity (see text), and other motifs are important for signal transduction. The variable intracellular portions of the receptors allow them to interact with specific G proteins.

VP and OT antagonist-binding sites appear to be different from the ones where the peptide hormones bind, but to overlap them partially. Recently, a peptide (CRAVKY) derived from the sequence of the second extracellular loop of the V2 receptor has been shown to act as a pharmacologically permissive, noncompetitive antagonist. It is an allosteric modulator of receptor function and exhibits functional selectivity – that is, it blocks VP-driven PGI₂ production and cremasteric vasorelaxation, but not camp production or beta-arrestin recruitment. Agents like this that affect one, but not all, second messenger systems triggered by GPCRs may have improved safety or specificity compared with more promiscuous antagonists.

See also: **Metabolism Vitamins and Hormones:** Amino Acid Metabolism; **Signaling:** Phospholipase C.

Further Reading

- Hedges BS (2002) The origin and evolution of model organisms. *Nature Reviews Genetics* 3: 838–849.
- Insel TR (2010) The challenge of translation in social neuroscience: A review of oxytocin, vasopressin, and affiliative behavior. *Neuron* 65: 768–779.
- Neumann ID and Landgraf R (eds.) (2008) *Advances in Vasopressin and Oxytocin – From Genes to Behavior to Disease. Progress in Brain Research*, vol. 170. New York: Elsevier.

V-ATPases

J Capecchi and M Forgac, Tufts University School of Medicine, Boston, MA, USA

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Michael Forgac, volume 4, pp. 349–353, © 2004, Elsevier Inc.

Glossary

Access channel An aqueous channel that allows protons to reach buried carboxyl groups in the center of the membrane from one side of the membrane or the other.

Osteopetrosis A genetic disease in humans associated with the inability to degrade bone, one cause of which is a defect in the vacuolar proton translocating adenosine triphosphatase (V-ATPase) of osteoclasts.

Receptor-mediated endocytosis The process by which cells take up specific ligands from their environment (such as low-density lipoprotein) via cell-surface receptors.

V-ATPase Vacuolar proton translocating ATPase that carries out active proton transport from the cytoplasmic to the noncytoplasmic side of the membrane driven by energy released upon hydrolysis of adenosine triphosphate (ATP).

Vacuolar (H⁺)-Adenosine Triphosphatase Function

Function of Intracellular Vacuolar (H⁺)-ATPases

Vacuolar (H⁺)-adenosine triphosphatase (ATPases) (V-ATPases) have been identified in many intracellular compartments, including endosomes, lysosomes, Golgi-derived vesicles, and secretory vesicles. V-ATPases within endosomal compartments are important for the process of receptor-mediated endocytosis (Figure 1). During receptor-mediated endocytosis, cells take up ligands (such as the cholesterol-carrying complex, low-density lipoprotein, or LDL) from their environment by binding them to receptors on the cell surface and clustering these receptors in specialized regions of the plasma membrane (clathrin-coated pits) which then invaginate into the cell. Following this internalization, the ligand–receptor complexes are exposed to a low pH within the early (sorting) endosome that causes the internalized ligand to dissociate from its receptor. This dissociation allows the receptor to recycle to the plasma membrane (where it is reutilized) and the ligand to proceed to the lysosome, where it is degraded. The low pH within the endosome is generated by the V-ATPase.

Acidification of endosomes is also important in the formation of carrier vesicles that carry the released ligands from early to late endosomal compartments and in the delivery of newly synthesized lysosomal enzymes from the Golgi to lysosomes. The latter process involves the binding of these enzymes to mannose-6-phosphate receptors in the *trans*-Golgi followed by their delivery to an endosomal compartment. Within this compartment, the low pH created by the V-ATPases causes dissociation of the lysosomal enzymes from their receptors, allowing delivery of the enzymes to the lysosome and recycling of the receptors to the *trans*-Golgi. Finally, endosomal acidification is involved in the entry of certain envelope viruses (such as influenza virus) and toxins (such as diphtheria and anthrax toxins) into cells. These viruses and toxins bind to the surface of cells and are internalized by the process of endocytosis. Upon exposure to a low pH, the virus coat fuses with the endosomal membrane, releasing the viral nucleic acid into the cytoplasm of the host cell. In the case of internalized toxins, the low pH induces a conformational change in a portion of the

toxin which creates a pore in the endosomal membrane through which the cytotoxic portion of the toxin enters the cell. Endosomal acidification is, therefore, essential in the process by which these viruses and toxins infect and kill cells.

Lysosomes are the major compartment in which degradation of proteins and other macromolecules occurs in cells. The lysosomal enzymes responsible for this degradation all require an acidic environment to be active. This acidic environment is created by the V-ATPases. Secretory vesicles, such as synaptic vesicles, are also acidic compartments. Synaptic vesicles are located at the synaptic terminal of nerve cells and release neurotransmitters (that chemically trigger the next nerve cell) by fusion with the plasma membrane. Neurotransmitters become concentrated within synaptic vesicles by transport proteins within the synaptic vesicle membrane that utilize either the proton gradient or the membrane potential generated by the V-ATPases to drive uptake of the transmitter. Finally, beyond their role in acidification, V-ATPases have been suggested to play a direct role in membrane fusion.

Function of Plasma Membrane V-ATPases

Plasma membrane V-ATPases play an important role in a number of normal and disease processes. In alpha-intercalated cells in the kidney, V-ATPases are located in the apical membrane where they pump protons into the urine, thus helping to control the pH of the blood. A genetic defect in this pump leads to a disease called renal tubule acidosis, in which the kidney is unable to secrete sufficient acid. V-ATPases are also present in the plasma membrane of osteoclasts, which are cells that function in degradation of bone. These cells are essential during development to facilitate bone remodeling. Plasma membrane V-ATPases in osteoclasts create an acidic extracellular environment that is necessary for bone degradation to occur. A genetic defect in the V-ATPase in osteoclasts leads to the human disease autosomal recessive osteopetrosis, in which the inability to degrade bone leads to severe skeletal defects and perturbations in Ca²⁺ homeostasis.

Plasma membrane V-ATPases in macrophages and neutrophils have been shown to help maintain a neutral internal

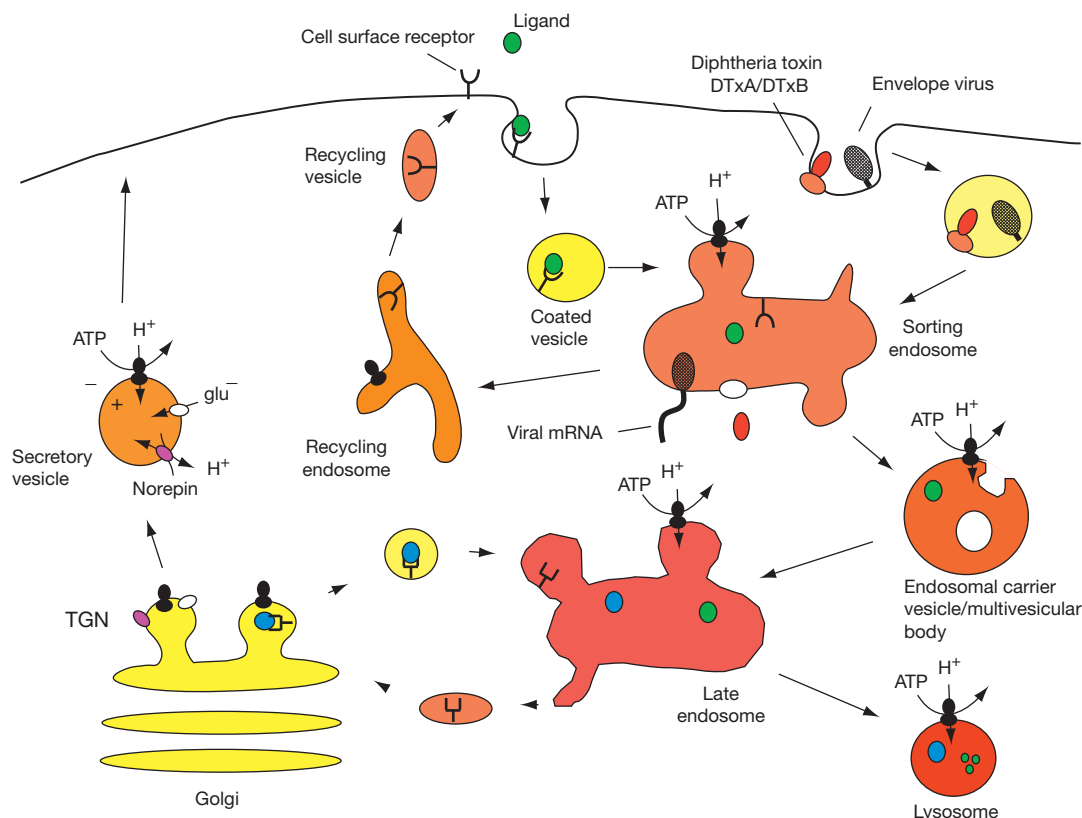


Figure 1 Function of intracellular V-ATPases. Acidification of early (sorting) endosomes by the V-ATPase facilitates receptor recycling following endocytic uptake and formation of endosomal carrier vesicles. Recycling M6P receptors to the *trans*-Golgi is also dependent on a low luminal pH. V-ATPase activity is required to drive neurotransmitter uptake by secretory vesicles and facilitates protein degradation in lysosomes. Finally, endosomal acidification facilitates the entry of the cytotoxic portions of viruses (such as influenza virus) and toxins (such as diphtheria toxin) into cells (see text).

pH under conditions of severe acid load. In the vas deferens, V-ATPases create a low pH environment necessary for sperm development. V-ATPases in the plasma membrane of tumor cells have also been proposed to function in tumor invasion by providing an acidic extracellular environment necessary for secreted lysosomal enzymes to degrade extracellular matrix. There is, thus, interest in V-ATPase inhibitors as potential anticancer drugs. Finally, V-ATPases in intestinal cells in insects create a membrane potential across the apical membrane that is used to drive potassium transport into the gut.

V-ATPase Structure

The V-ATPase is a large complex composed of 14 different subunits. These subunits are arranged into two separate domains termed V_1 and V_0 (Figure 2). The V_1 domain is made up entirely of subunits that are peripheral to the membrane (i.e., not membrane embedded). This domain has a molecular mass of ~720 kDa and contains eight different subunits (subunits A–H) of molecular mass 70–13 kDa (Table 1). The V_1 domain is responsible for hydrolysis of ATP, which occurs on catalytic sites located on the three copies of subunit A. There are, therefore, three catalytic nucleotide-binding sites per complex. The B subunits (which are also present in three copies per

complex) can also bind nucleotides, but these sites are referred to as ‘noncatalytic’ sites, since they do not actually hydrolyze ATP. The function of these sites is not known, but they may play a role in controlling the activity of the V-ATPase. The A and B subunits are arranged in a hexamer, like the segments of an orange, with alternating A and B subunits. ATP is hydrolyzed sequentially at each of the three catalytic sites. The other subunits in the V_1 domain (subunits C–H) function to connect the V_1 domain to the V_0 domain and are discussed here.

The V_0 domain in yeast is composed of six different subunits (subunits a, d, c, c', c'', and e) of molecular mass 100–9 kDa. All of the subunits in the V_0 domain except subunit d are embedded in the membrane and, thus, require detergent for solubilization. The V_0 complex has a molecular mass of 260 kDa and is responsible for transport of protons across the membrane. This proton transport only occurs when the V_1 domain is attached to V_0 and is driven by the hydrolysis of ATP in V_1 . Three of the subunits in the V_0 domain are called ‘proteolipid subunits’ (c, c', and c'') because they are so hydrophobic that they can be extracted from the membrane using organic solvent mixtures, such as chloroform:methanol. These subunits are almost completely embedded in the membrane and the polypeptide chain of each one crosses the membrane 4–5 times (these are called ‘transmembrane segments’). Buried in the middle of one of the transmembrane segments of each of

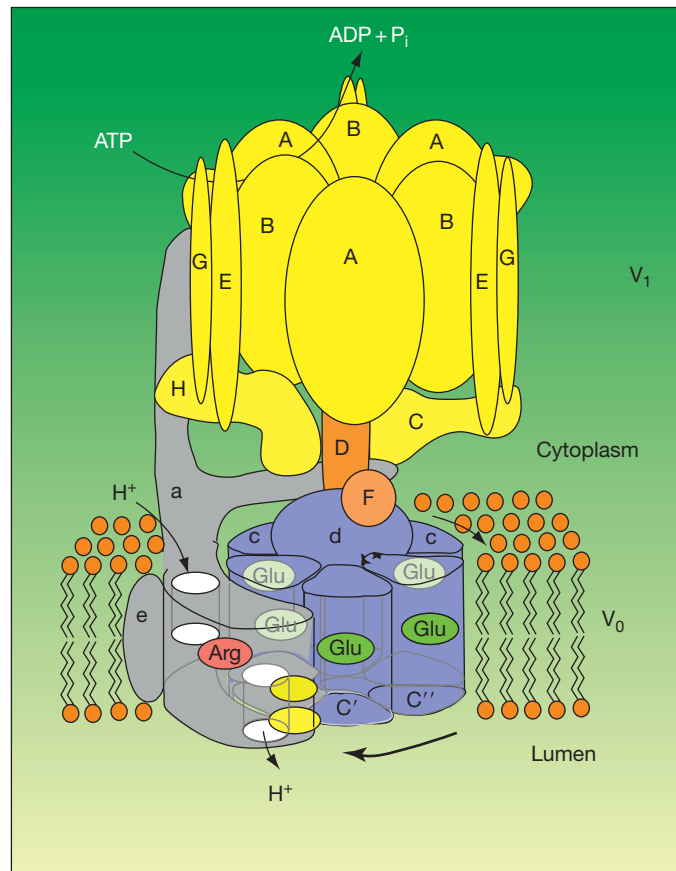


Figure 2 Structure of the V-ATPases. The V-ATPases contain two domains. The V_1 domain is responsible for ATP hydrolysis and the V_0 domain carries out proton transport across the membrane. Like the F-ATPases, the V-ATPases operate by a rotary mechanism in which ATP hydrolysis in V_1 drives rotation of a central stalk which is connected to a ring of proteolipid subunits in V_0 . It is movement of the proteolipid ring relative to subunit a that drives proton transport (see text). An Arginine (Arg) residue on subunit a and glutamic acid (Glu) residues on the ring of proteolipid subunits are crucial for proton translocation and are shown in red and green, respectively. Charged residues on subunit a that contribute to proton-conducting hemi-channels are shown in yellow.

Table 1 Subunit composition of the V-ATPase

Domain	Subunit	Gene (yeast)	M_r (kDa)	Function/location
V_1	A	<i>VMA1</i>	69	Catalytic ATP-binding site
	B	<i>VMA2</i>	58	Noncatalytic ATP-binding site
	C	<i>VMA5</i>	44	Peripheral stalk, regulation of reversible dissociation
	D	<i>VMA8</i>	29	Central stalk
	E	<i>VMA4</i>	26	Peripheral stalk
	F	<i>VMA7</i>	14	Central stalk
	G	<i>VMA10</i>	13	Peripheral stalk
	H	<i>VMA13</i>	54	Peripheral stalk, inhibition of free V_1 ATP hydrolysis
V_0	a	<i>VPH1/STV1</i>	100	Proton translocation, targeting
	d	<i>VMA6</i>	40	Cytoplasmic side
	c	<i>VMA3</i>	17	Proton translocation, bafilomycin-binding site
	c'	<i>VMA11</i>	17	Proton translocation
	c''	<i>VMA16</i>	23	Proton translocation
	e	<i>VMA9</i>	9	Unknown

the proteolipid subunits is a single essential glutamic acid residue, which is reversibly protonated and deprotonated during proton transport by the V-ATPases. Like the A and B subunits in the V_1 domain, the proteolipid subunits in the V_0

domain form a ring, with four copies of subunit c and one copy each of subunit c' and c''. The specific V-ATPase inhibitor bafilomycin has been shown to bind to the proteolipid subunits of the V_0 domain.

In addition to the proteolipid subunits, the a and e subunits of V_0 are also embedded in the membrane. Subunit a is 100 kDa and is made up of two domains. The carboxyl-terminal half of the molecule contains eight transmembrane segments while the amino-terminal half is a hydrophilic domain that is present on the cytoplasmic side of the membrane. Like the proteolipid subunits, subunit a also contains amino acid residues that are essential for proton transport. In particular, there is a positively charged arginine residue near the middle of the seventh transmembrane segment of subunit a that is absolutely required for proton transport by the V-ATPases. Subunit e is a small (9 kDa) highly hydrophobic protein of unknown function. The function of subunit d, which is tightly bound to the V_0 domain but is not embedded in the membrane, is thought to be in bridging the central stalk of V_1 (see below) with the proteolipid ring of V_0 . It should be noted that in mammals there is no gene encoding subunit c', but that the mammalian V_0 domain contains an additional accessory subunit (Ac45) of unknown function.

The V_1 and V_0 domains are connected by a single central stalk and multiple peripheral stalks. The central stalk is composed of the subunits D and F, whereas the peripheral stalks are composed of subunits C, E, G, H, and the soluble domain of subunit a. The function of these stalks is described below.

Mechanism of ATP-Driven Proton Transport by the V-ATPases

The V-ATPases operate by a rotary mechanism, similar to that demonstrated for the F-ATPases (or ATP synthases), which are enzyme complexes present in mitochondria, chloroplasts, and bacteria that function in the reverse direction (that is in proton-driven ATP synthesis). For the V-ATPases, ATP hydrolysis in the V_1 domain drives rotation of the central stalk (containing subunits D and F), which in turn drives rotation of the ring of proteolipid subunits relative to subunit a in the V_0 domain. Subunit a is held fixed relative to the A_3B_3 hexamer of V_1 by the peripheral stalks (or stators), composed of subunits C, E, G, H, and the soluble domain of subunit a. Recent data indicate there are three peripheral stalks, each of which contain a heterodimer of subunits E and G. It is rotation of the ring of proteolipid subunits relative to subunit a that drives active transport of protons from the cytoplasmic to the luminal side of the membrane. A proton enters from the cytoplasmic side of the membrane via a cytoplasmic access channel in subunit a and protonates a buried carboxyl group on one of the proteolipid subunits. ATP hydrolysis in V_1 forces rotation of the proteolipid ring in the plane of the membrane such that the protonated carboxyl group on the proteolipid subunit reaches a second access channel in subunit a that leads to the luminal side of the membrane. Interaction between this carboxyl group on the proteolipid subunit and the buried arginine residue of subunit a (which is positively charged) forces the proton off of the proteolipid subunit into the luminal access channel, where it can be released to the luminal side of the membrane, thus completing the transport cycle. In this way, the rotary motion driven by hydrolysis of ATP is converted into unidirectional transport of protons across the membrane.

Regulation of V-ATPase Activity *In Vivo*

The activity of V-ATPases in different membranes in the cell is known to be regulated such that the pH of different intracellular compartments and the degree of proton transport across the plasma membrane are carefully controlled, but the mechanisms employed in regulating V-ATPase activity in cells are still being elucidated. One important mechanism of regulation involves reversible dissociation of the V-ATPase complex into its component V_1 and V_0 domains. In yeast, dissociation occurs in response to removal of glucose from the media, probably as a way to preserve cellular energy stores. This process has been shown to be regulated by the Ras/cAMP/protein kinase A pathway. Reversible dissociation has also been demonstrated to occur in insects and in mammalian cells. In dendritic cells, activation of antigen processing increases V-ATPase assembly on the lysosomal membrane as a way to stimulate antigen processing, which depends upon the activity of proteases that have optimal activity at low pH. A second proposed regulatory mechanism involves the formation of a disulfide bond between conserved cysteine residues located at the catalytic site on the subunit A. Differences in the efficiency of coupling of ATP hydrolysis and proton transport have also been observed for V-ATPases containing different isoforms of subunit a and likely help to explain the difference in pH of different intracellular compartments.

Changes in the density of V-ATPases in different cellular membranes have also been proposed as a means of controlling proton transport. This has been shown to occur in intercalated cells in the kidney, where exposure to a low pH causes the fusion of intracellular vesicles containing the V-ATPase with the plasma membrane, thus increasing proton transport out of the cell into the renal fluid. Regulated exocytosis of V-ATPases has also been demonstrated in epididymal clear cells. Differential targeting of V-ATPases to different cellular membranes appears to be controlled by isoforms of the a subunit. Thus, the a3 isoform is able to target the V-ATPase to the plasma membrane in osteoclasts, whereas the a4 isoform targets the V-ATPase to the intercalated cell plasma membrane. It is mutations in these isoforms that lead to the human diseases osteopetrosis and renal tubule acidosis mentioned earlier. Isoforms have now been identified in many of the V-ATPase subunits in mammalian cells, and these have been shown to be expressed in both tissue- and organelle-specific manner. This has led to the expectation that specific inhibitors can be identified that are selective in their ability to inhibit particular V-ATPase complexes, which may in turn lead to the development of drugs for the treatment of diseases such as osteoporosis and tumor metastasis.

See also: [Lipids Carbohydrates Membranes and Membrane Proteins: Lipid Rafts.](#)

Further Reading

- Bond S and Forgac M (2008) The Ras/cAMP/protein kinase A pathway regulates glucose dependent assembly of the vacuolar (H^+)-ATPase in yeast. *Journal of Biological Chemistry* 283: 36513–36521.
- Forgac M (2007) Vacuolar ATPases: Rotary proton pumps in physiology and pathophysiology. *Nature Reviews. Molecular Cell Biology* 8: 917–929.

- Hinton A, Sennoune SR, Bond S, et al. (2009) Function of a subunit isoforms of the V-ATPase in pH homeostasis and *in vitro* invasion of MDA-MB231 human breast cancer cells. *Journal of Biological Chemistry* 284: 16400–16408.
- Hurtado-Lorenzo A, Skinner M, El Annan J, et al. (2006) V-ATPase interacts with ARNO and Arf6 in early endosomes and regulates the protein degradative pathway. *Nature Cell Biology* 8: 124–136.
- Lu M, Ammar D, Ives H, Albrecht F, and Gluck SL (2007) Physical interaction between aldolase and vacuolar H⁺-ATPase is essential for the assembly and activity of the proton pump. *Journal of Biological Chemistry* 282: 24495–24503.
- Zhang Z, Zheng Y, Mazon H, et al. (2008) Structure of the yeast vacuolar ATPase. *Journal of Biological Chemistry* 283: 35983–35995.

Vitamin A (Retinoids)

J L Napoli, University of California, Berkeley, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

k_{cat}/K_m A measure of the efficiency of an enzyme for its substrate from dividing the rate of turnover of the substrate by the enzyme by the Michaelis constant.

K_d (equilibrium dissociation constant) A measure of the affinity of a protein for its ligand. Lower numbers indicate higher affinity.

Retinoid-binding protein Proteins that bind specific retinoids with high affinity, several of them with $K_d \leq 10$ nM.

Short-chain dehydrogenase/reductase (SDR) A gene family consisting of ~50 mammalian members in the range of 25–35 kDa that uses pyridine nucleotide cofactors to dehydrogenate or reduce steroids, retinoids, prostanoids, and intermediates in lipid metabolism.

Retinoids and Their Functions

The term retinoid refers to naturally occurring and synthetic compounds with vitamin A activity: the term vitamin A denotes the compound all-*trans*-retinol (atROH). All vertebrates require vitamin A for vision, fertility, embryogenesis, regulation of energy balance, and growth. atROH does not itself fulfill the physiological functions of vitamin A; rather it acts as precursor for biosynthesis of metabolites directly responsible for producing vitamin A activity. These atROH metabolites include 11-*cis*-retinal, the cofactor that covalently binds with opsin to form the visual pigment rhodopsin; all-*trans*-retinoic acid (atRA), the humoral effector of the nonvisual, systemic functions attributed to vitamin A; and 9-*cis*-RA (9cRA), a regulator of glucose-stimulated insulin secretion. These systemic functions include: controlling the differentiation programs of epithelia and of stem cells in the skin, nervous system, bone, immune system, hemopoietic system; acting as a tumor suppressor; regulating apoptosis; regulating energy balance, adiposity, and glucose/insulin tolerance. Nonvisual effects of retinoids also include memory formation, immune responsiveness, stress adaptation, cell fate determination, and suppression of carcinogenesis. atRA and 9cRA activate three nuclear receptors, retinoic acid receptor (RAR) α , β , and γ , and thereby control the expression of hundreds of genes. atRA also binds with PPAR δ/β . Both RAR and PPAR heterodimerize with RXR. CRABPII delivers atRA to RAR, whereas FABP5 delivers atRA to PPAR δ/β . atRA and 9cRA also function through nongenomic mechanisms.

The Chemistry of atROH Generation, Storage, and Metabolic Activation

Carotenoids with at least one β -ionone ring, especially β -carotene, provide the major retinoid precursors in most diets. Oxidative central cleavage by a plasma-membrane-associated, but soluble, carotene 15,15'-monooxygenase (CMO1) produces all-*trans*-retinal – a transient intermediate that occurs in very low concentrations outside of the neural retina. Microsomal retinal reductases (RRDs), members of the short-chain dehydrogenase/reductase (SDR) gene family, catalyze reduction of all-*trans*-retinal into atROH, which then

undergoes esterification with fatty acid moieties, predominantly palmitate, into all-*trans*-retinyl esters (atREs) by microsomal lecithin:retinol acyltransferase (LRAT). Chylomicra transport atRE into circulation and deliver some to adipose, whereas chylomicra remnants deliver most to liver, where ultimately most REs are stored in stellate cells.

In the visual cycle, atREs undergo concerted hydrolysis isomerization into 11-*cis*-retinol by RPE65-isomerohydrolase (IMH), followed by dehydrogenation into 11-*cis*-retinal by SDR. 11-*cis*-Retinal forms a Schiff's base with a lysine residue in the protein opsin to form rhodopsin. When light strikes the neural retina, 11-*cis*-retinal in rhodopsin undergoes *cis*- to *trans*-isomerization, causing a conformation change that initiates nerve impulses through activation of G-proteins, and releases the newly formed all-*trans*-retinal. Reduction of all-*trans*-retinal into atROH (by microsomal SDR) and re-esterification (by LRAT) into atRE completes the visual cycle.

Activation of atROH into atRA uses the same intermediate used in the visual cycle, all-*trans*-retinal, but relies on a metabolically distinct route. atROH, either from blood or from hydrolysis of retinyl esters by microsomal retinyl ester hydrolase (REH), undergoes reversible dehydrogenation into all-*trans*-retinal, catalyzed by microsomal retinol dehydrogenases (RDHs), members of the SDR gene family. In contrast to the comparatively high concentrations of retinal that allows the visual cycle to function, all-*trans*-retinal concentrations during atRA biosynthesis are kept low (50–200 nmol g⁻¹) by reduction (back-reaction of RDH and reduction by microsomal and peroxisomal reductases), and by irreversible dehydrogenation into atRA, catalyzed by soluble retinal dehydrogenases (Raldh), members of the aldehyde dehydrogenase (ALDH) gene family. atRA isomers occur *in vivo*, such as 13-*cis*-RA and 9,13-di-*cis*-RA, but their significance and source (s) remain unclear. Recently, the issue of physiological occurrence of 9cRA as a bona fide hormone has been resolved with its analytically rigorous identification in the pancreas, and the inability to detect it (<0.05 pmol g⁻¹) in a variety of other retinoid target tissues.

These straightforward reactions offer complex opportunities for physiological regulation, owing to compartmentalization, distinct enzymes catalyzing each direction of chemically reversible reactions (e.g., dehydrogenation/reduction of retinol/retinal; esterification/hydrolysis of atROH/atRE), cell-distinct expression patterns, and multiple homologs/paralogs that

Table 1 Physiologically important Rdh

Mouse	Rat	Human
Rdh1	Rdh2, Rdh2_rat, Rdh7 (Rodh2, Rodh1/3)	Rdh16 (RDH-E, RoDH4, NP_0036992)
Rdh10	Rdh10	Rdh10
Dhrs9	Dhrs9 (eRoLDH2)	Dhrs9 (retSDR8, RDHL, Rdh-TBE, hRoDH-E ₂ , 3 α -HSD)

Rows depict orthologs. Columns depict homologs. Names in parenthesis indicate first and/or alternative designations.

catalyze several reactions. For example, at least four reductases have been identified, which belong to the SDR gene family, Rrd (peroxisomal) and Rdh10 outside of the eye, and retSDR and Rdh12, in the neural retina. At least three Rdh's (SDRs) have been identified in the rat, mouse, and human (Table 1). Four Rdh's have been identified in human, rat, and mouse: Rdh1, 2, 3, 4 (aka ALDH1A1, 1A2, 1A6, and 8A1). These constitute a complex enzyme system for absorbing and storing vitamin A, maintaining atRA homeostasis, and recycling vitamin A in the visual cycle.

Retinoid-Binding Proteins and Retinoid Homeostasis

Processing of dietary retinoids and retinoid precursors and biogenesis of active retinoids relies on serum and cellular chaperones for efficient and specific retinoid use, as demonstrated by studies *in vitro* and the consequences of gene knockouts and/or naturally occurring mutations.

Liver and other tissues synthesize a retinol-binding protein, RBP (aka Rbp4), a member of the lipocalin gene family. atROH egress from liver requires sRBP, and sRBP-atROH represents the major form of vitamin A in serum. sRBP circulates as a complex with a tetramer of transthyretin, which protects the ~20 kDa sRBP from kidney filtration. Recent work shows that some cells have a cell-surface receptor, Stra6, which recognizes sRBP. The sRBP-null mouse seems phenotypically normal when fed a diet that contains copious amounts of vitamin A, except for impaired vision after weaning. Feeding a diet with lower, but normally adequate amounts of vitamin A, prevents reproduction in the sRBP-null mouse. Although the eye relies on sRBP for efficient atROH uptake, atROH obtained from postprandial lipoprotein delivery can substitute if the animals are fed diets with copious amounts of vitamin A. Interestingly, atRA serum levels increase in the sRBP null mouse even during high vitamin A diets, indicating that serum delivery of atRA to tissues helps compensate for impaired atROH delivery.

Binding proteins channel retinoid intermediates through the series of reactions that constitute the visual cycle (Figure 1). Mice null in the retinal pigment protein RPE65-IMH, for example, cannot produce 11-*cis*-retinoids, consistent with an RPE65-IMH function of binding the hydrophobic atRE ($K_d \sim 20$ pM), and catalyze acyl hydrolysis and C11 isomerization. RPE65 belongs to the same gene family as the carotenoid-metabolizing enzyme CMO1 (the mouse proteins

have only 37% amino acid identity, however), suggesting a gene family devoted to transport/metabolism of hydrophobic substances. The RPE65-IMH product, 11-*cis*-retinol, undergoes sequestration in the RPE by the cytosolic cellular retinal-binding protein (CRALBP), a member of the gene family that includes the α -tocopherol transfer protein. CRALBP facilitates dehydrogenation of 11-*cis*-retinol into 11-*cis*-retinal, and drives forward the *trans*- to *cis*-isomerization. Mutations in human CRALBP cause night blindness and photoreceptor degeneration. RDH4 and RDH5, respectively, represent the murine and human 11-*cis*-retinol dehydrogenase, which interacts with and/or accepts substrate from CRALBP. Mutations in RDH5 are associated with the rare, autosomal recessive disease fundus albipunctatus, that is, night blindness from delayed photopigment regeneration. Lack of total blindness in the case of the RDH5 mutation indicates that SDRs in addition to RDH4/5 contribute to 11-*cis*-retinal biosynthesis, such as Rdh11. 11-*cis*-Retinal transverses the interphotoreceptor matrix by an unknown mechanism and covalently binds with opsin to form rhodopsin in the rod outer segment (ROS). After the action of light, an adenosine triphosphate (ATP)-dependent transporter, ABCR, facilitates release of all-*trans*-retinal from rhodopsin. Another SDR, Rdh12 (plus other SDRs) then reduce all-*trans*-retinal into atROH, which migrates back through the interphotoreceptor space to the RPE.

Outside of the vision cycle, several cellular retinoid-binding proteins (CRBPs) occur, other than those mentioned above with respect to the visual cycle. These include: CRBPI, CRBP II, CRBP III, CRABPI, and CRABPII (Table 2), which are ~15 kDa globular proteins of the intracellular lipid-binding protein (iLBP) gene family that includes the various fatty acid-binding proteins. All vertebrates express all five, with well-conserved amino acid sequences among orthologs. All are high affinity (except CRBP III), soluble, and specific for their ligands. CRBPI binds atROH, and closely related compounds such as 3,4-didehydro-atROH and all-*trans*-retinal to a lesser extent, and discriminates against atRA. CRABPI binds atRA, metabolites such as 4-OH-atRA, and discriminates against *cis*-RAs and atROH.

The cellular retinol/retinal-binding proteins enclose atROH and all-*trans*-retinal with the hydroxyl/aldehyde function, crucial to metabolic activation, sheltered inside. CRBPI and II confer selective advantage to vertebrates by enhancing efficiency of storing and transporting vitamin A, and limiting its catabolism. Vitamin A absorption and biosynthesis in the intestine relies on CRBP II. CRBP II null mice pups die within 24 h after birth, when delivered by dams fed a diet marginal in vitamin A content. CRBP II contributes ~1% of the soluble protein to the intestinal enterocyte – an indication of a mass-action function to sequester newly synthesized (from carotenoids) all-*trans*-retinal or newly absorbed dietary atROH. *In vitro*, all-*trans*-retinal bound with CRBP II undergoes reduction readily, but neither it nor bound atROH undergoes efficient dehydrogenation. This would limit production of atRA, and other metabolism, from the bolus of all-*trans*-retinal produced during carotenoid cleavage. In the intestine, the esterifying enzyme LRAT recognizes atROH bound with CRBP II to produce atRE for incorporation into chylomicra. Thus, CRBP II likely aids atROH uptake and chaperones the products of

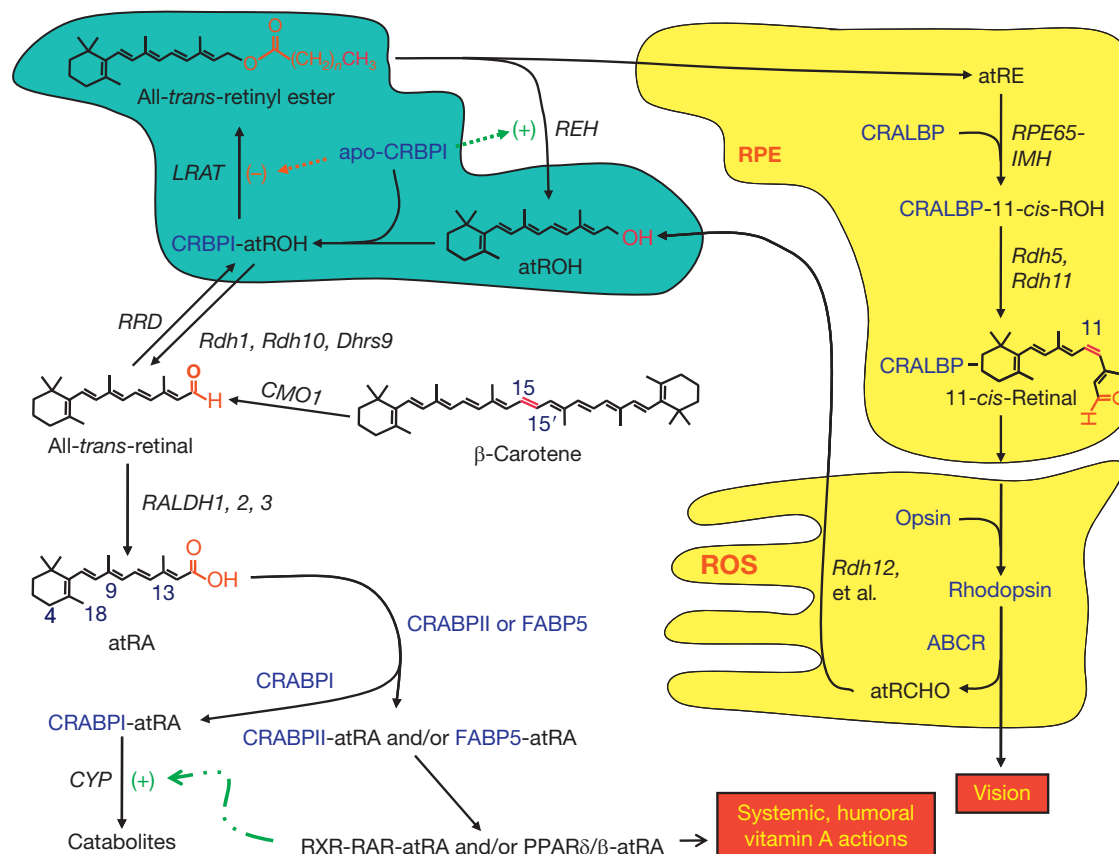


Figure 1 Major paths of retinoid metabolism. The white background designates reactions of retinol homeostasis and activation into atRA in the liver and extrahepatic vitamin A target tissues. The light gray background depicts reactions that occur both in extraocular tissues and in the retinal pigment epithelium (RPE) of the neural retina. The yellow background designates reactions that occur in the RPE and in the rod outer segments (ROSs). Numbers indicate the positions of *trans*- to *cis*-isomerization (C11), the positions of hydroxylations that occur during catabolism of atRA (C4 and C18), and the positions of atRA isomers of physiological interest (C9, C13). The dotted lines indicate the direct actions of apo-CRBPI in regulating atRE hydrolysis and atROH esterification. The dashed-dotted line indicates transcriptional induction of cytochromes P-450 (CYP) by atRA. For simplicity, CRBPII and III are not shown. As these are intracellular reactions, sRBP and the plasma membrane sRBP receptor, Stra6, are not shown. The biosynthetic path of 9cRA is not certain, but SDRs can convert 9-*cis*-retinol into 9-*cis*-retinal, and Raldh can convert 9-*cis*-retinal into 9cRA *in vitro*.

Table 2 Major extraocular retinoid-binding proteins

Retinoid-binding protein	Ligand(s)	K_d (nM)	Postembryonic distribution
CRBPI (cellular retinol-binding protein, type I)	atROH, 9cROH	~10	Nearly ubiquitous (low in intestine)
CRBPII (cellular retinol-binding protein, type II)	all- <i>trans</i> -retinal	10–50	Intestine, neonatal liver
CRBPIII (cellular retinol-binding protein, type III)	atROH	10–50	Heart, muscle, adipose, mammary
CRABPI (cellular retinoic acid-binding protein, type I)	all- <i>trans</i> -retinal	~100	Widespread
CRABPII (cellular retinoic acid-binding protein, type II)	atRA	0.4	Limited (e.g., skin, uterus, and ovary)
	atRA	2	

carotenoid metabolism down the pathway to atRE to enhance efficiency of retinoid recovery from the diet.

Extra-intestinal vitamin A uptake and storage relies on CRBPI. CRBPI-null mice seem morphologically normal, but eliminate atRE sixfold faster than wild-type mice, and may sequester/esterify atROH less efficiently. CRBPI-null mice also have impaired glucose tolerance. Clearly, efficient use of atROH *in vivo* depends on this chaperone. Strikingly, CRBPI has a K_d value for atROH far lower than the atROH

concentration in tissues (0.2–30 μ M), and CRBPI concentrations exceed atROH concentrations, where measured. The law of mass action predicts from these data that nonesterified atROH would occur nearly exclusively bound with CRBPI, in the absence of alternative high-affinity or high-capacity acceptors. Nonspecific alternatives to CRBPI exert limited influence *in vivo*, because isolation of CRBPI from liver by traditional, time-consuming biochemical techniques produces largely holo-protein. Evidently, the capacity of membranes and other

potential acceptors to sequester atROH does not overwhelm the ability of CRBPI to sequester atROH. Like CRBP, LRAT can access atROH bound with CRBPI to produce atRE. Ultimately, liver stellate cells accumulate most of the atREs. CRBPI seems necessary for retinoid transfer from hepatocytes to stellate cells, because the CRBPI-null mouse does not accumulate atRE in stellate cells.

atROH sequestering within CRBPI, the need to store retinol as RE and the need for atRA biosynthesis during specific times at specific loci, suggests that transfer of atROH for metabolism might depend on relationships between metabolizing enzymes and CRBPI. Like CRBP, CRBPI allows esterification of bound atROH, but in contrast to CRBP, CRBPI also allows dehydrogenation of atROH. The CRBPI-atROH complex shows Michaelis-Menton relationships with atRE formation by LRAT and atROH dehydrogenation by RDH. This relationship with the microsomal RDH is maintained even with changes in the CRBP/atROH ratio (provided the CRBPI concentration exceeds the atROH concentration). This eliminates a free diffusion mechanism of transfer from the complex to select enzymes. Specific crosslinking of holo-CRBPI with both RDH and LRAT confirms close proximity of CRBPI and these two enzymes. In addition, a single mutation in an exterior residue of CRBP (L35A) reduces the V_m of atROH dehydrogenation by microsomes, but does not alter the K_m , or the K_d of atROH binding to CRBPI, consistent with conservation of exterior residues that aid transfer of atROH from CRBPI to enzymes. Obviously, atRE and atRA biosynthesis *in vivo* occurs in the absence of CRBPI, as indicated by the lack of morphological pathology in the CRBPI-null mouse, and their ability to sequester RE, albeit inefficiently. This was predicted by the experiments *in vitro*, which showed that neither RDH nor LRAT require presentation of atROH by CRBPI. Not surprisingly, the enzymes' active sites recognize their substrates in the absence of CRBPI. CRBPI operates as a chaperone, which restricts atROH metabolism to select enzymes, and seems required only for efficient atROH use *in vivo*. The ability of LRAT and RDH to access retinol from CRBPI addresses the issue of how atROH would undergo efficient metabolism in the face of limited diffusion from the binding protein.

Knockouts of *Rdh1* and *Rdh10* have phenotypes related to dysfunctions of retinoid-regulated genes and retinoid action. The *Rdh1*-null mouse is up to 11% longer and 37% heavier than wild type, and the *Rdh10*-null mouse dies during embryogenesis because of impaired brain development. Both have disturbances in retinoid metabolism.

RALDHs catalyze the irreversible conversion of atRCHO (all-*trans*-retinal) into atRA in the presence of CRBPI, and can also use atRCHO generated *in situ* from CRBPI-atROH and RDH, or in cells presented with atROH and transfected with RDH and RALDH. In the rat, RALDH1 and RALDH2 have differing, but overlapping, expression patterns, and respond differently to changes in atROH status. This suggests a purpose for multiple RALDH, that is, precise control over atROH use and atRA generation. The RALDH1-null mouse remains fertile and healthy, when fed a diet copious in vitamin A, but may have decreased ability to produce atRA in the liver, and seems to resist weight gain. The RALDH2-null mouse dies *in utero* at midgestation, demonstrating its unique contribution to atRA biosynthesis at a specific time embryogenesis. The situation

may differ in the adult, as testes express RALDH2 strongly, but RALDH1 prevails outside of the testis. The RALDH3-null mouse dies during suckling from an obstruction in the nose. Apparently, RALDH can compensate for each other after critical developmental milestones.

A CRBP^{III} has been detected in limited tissues – some express little or no CRBPI or CRBP^{II} – but has not been detected in other retinoid target tissues, such as liver, kidney, and brain. CRBP^{III} seems to bind about equally well with atROH, 9-*cis*-retinol and 13-*cis*-retinol, but with much lower affinity, either CRBP or CRBP^{II}. In adipose CRBP^{II} functions in lipid metabolism. Humans express yet another CRBP, originally referred to as CRBP^{III}, but distinct from mouse CRBP^{III}, and therefore CRBP^{IV}. CRBP^{IV} mRNA is much more abundant in human liver and intestine than CRBP^{II} mRNA, but the mouse does not encode a complete CRBP^{IV} gene. CRBP^{IV} binds atROH with a K_d value of ~ 60 nM, and does not bind *cis*-isomers. The precise functions of CRBP^{III} and CRBP^{IV} have not been clarified, and unlike CRBPI and CRABP, their endogenous ligands have not been established.

Knockouts of the atRA-binding proteins, CRABPI and CRABP^{II}, have not clarified their functions. Doubly null mice have a nearly fourfold higher rate of death from unknown causes by 6 weeks after birth than wild type, but the survivors appear normal, when fed a diet copious in vitamin A (chow diet), with one exception. The doubly null mouse as well as the CRABP^{II}-only null mouse show 83% and 45%, respectively, incidence of a small outgrowth anomaly on the postaxial side of digit 5, predominantly in the forelimbs. Mice doubly null in CRABPI and CRABP^{II} do not exhibit enhanced sensitivity to exogenous atRA, suggesting that the binding proteins do not protect against atRA toxicity. In contrast to CRBPI, both CRABPI and CRABP^{II} allow projection of the β -ionone ring of their ligand. Significantly, the first reactions of atRA degradation occur at these comparatively accessible sites, that is, C4 and C18. Presenting atRA to microsomes bound with CRABPI enhances kinetic efficiency (k_{cat}/K_m) of catabolism sevenfold. There seems to be little doubt that CRABPI sequesters atRA: delivering the sequestered atRA for efficient catabolism seems a logical mechanism to discharge the ligand without releasing it back into the cell. CRABPI tends to express in cells different from those that express CRBPI and CRABP^{II}, suggesting restriction of atRA function. CRABP^{II}, but not CRABPI, can deliver atRA to RAR in the nucleus, via a transfer that does not require free diffusion. This would complete the chaperoning of atROH on its journey from atRE hydrolysis to atRA biogenesis and nuclear localization.

Metabolism of Pharmacological Retinoid Concentrations

Many other enzymes, including medium-chain alcohol dehydrogenases (ADHs), and aldo-keto reductases, metabolize retinoids *in vitro*, which seems to confirm the need for evolution of CRBP to protect the sparse and valuable vitamin A from clearance as a xenobiotic. These enzymes do not access atROH bound with CRBP. The ADH1-null deer mouse, a natural mutant, and mice made null in ADH1, ADHIV, or ADHIII or doubly null in ADH1 and ADHIV do not: (1) present with

phenotypes resulting from inadequate vitamin A; (2) show compensatory responses in retinoid metabolism; (3) show changes in retinoid-regulated genes; and (4) have disrupted metabolism of physiological amounts of retinoids. Dosing toxic amounts of atROH (50 mg kg⁻¹ or ~300-fold higher than the recommended daily intake) to mice produces a serum atRA concentration ~1600-fold higher than normal. In the ADHI-null mouse, this is reduced to ~200-fold higher than normal. These data show that huge amounts of retinol can overwhelm physiological chaperones and provide a mechanism for retinol toxicity, as modest increases in atRA lead to retinoid toxicity.

Other Naturally Occurring Retinoids

Discrete loci, such as skin and the chick limb bud, synthesize 3,4-didehydro-atRA. 3,4-Didehydro-atRA binds RARs with affinity similar to that of atRA. The purpose of creating a signaling molecule that functions as atRA in specialized loci that also biosynthesize atRA has not been clarified, but subtle differences in ligand structure can have significant differences on the conformation of their receptors. These differences can lead to recruitment of different cofactors and/or binding of different response elements. This is illustrated by the development of synthetic retinoids that act as RAR agonist, antagonists, or inverse agonists. It is reasonable to infer that atRA and 3,4-didehydro-atRA regulate different genes, even though both bind RAR.

Although 9cRA was reported as a physiological ligand of RXR, it is undetectable *in vivo* in the vitamin A-target tissues assayed, except for pancreas. It, therefore, is unlikely to serve as a universal ligand for RXR, but it does modulate glucose-stimulated insulin secretion.

Control of Vitamin A Homeostasis

The primary regulators of retinoid homeostasis appear to be apo-CRBPI and atRA. apo-CRBPI inhibits LRAT and stimulates retinyl ester hydrolysis. Thus, the direction of flux into or out of atRE would reflect the ratio apo-CRBPI/olo-CRBPI, which would reflect the atROH status of the cell. atRA may serve as a signal to liver to release atROH into the serum. Little has been revealed about humoral regulation of atRA biosynthesis, but estrogen and prostaglandin E increase and decrease, respectively, atRA biosynthesis in cultured cells. The metabolism of atRA limits its activity; conversely, inhibitors of atRA metabolism enhance atRA potency. atRA induces its own oxidative metabolism into 5,6-epoxy-atRA, 18-hydroxy-atRA, and 4-hydroxy-atRA, through inducing cytochrome P-450s (CYPs). CYP26A1, B1, C1, and CYP2C39 (mouse only) appear to have major functions in the catalysis of atRA catabolism. CYP26A1-null mice die in mid- to late gestation with serious morphological defects. The precise function(s) of CYP26B1, 26C1, and 2C39 have not been clarified, but 26B1 and 2C39, like 26A1,

are induced by atRA. Many other CYPs reportedly catabolize atRA (e.g., CYP1A1/2, CYP2A6, CYP2C8/9, CYP2E1, and CYP3A4/5), but atRA does not induce these isoforms and most have inefficient kinetics with atRA *in vitro*.

It is notable that PPARδ/β, an atRA activated transcription factor, induces transcription of CMO1, CRBPI, CRBP2, LRAT, and RARβ, indicating regulation of vitamin A homeostasis by atRA, and suggesting correlation between vitamin A homeostasis and lipid metabolism.

Several xenobiotics, including ethanol and polychlorinated aromatic hydrocarbons, reduce atRE stores, likely through enhancing atRA catabolism by inducing CYP: the polychlorinated hydrocarbons, such as dioxin, function through the AH receptor to decrease atRE stores.

Clinical Uses of Retinoids

Numerous studies have correlated vitamin A insufficiency in laboratory animals with increased incidence of spontaneous and carcinogen-induced cancer. Chemopreventive trials in humans show some promise for retinoids in actinic keratoses, oral premalignant lesions, laryngeal leukoplakia, and cervical dysplasia. The FDA has approved retinoids for acute promyelocytic leukemia and in non-life-threatening diseases such as cystic acne and psoriasis. Retinoids also provide the active ingredients in agents to treat sun/age damaged skin, through regulating transcription, differentiation, and cell function in the skin.

The WHO recognizes vitamin A deficiency as a mortality factor for childhood measles. Two large doses (60 000 REQ each) of a water-soluble vitamin A formulation given to children on each of 2 days decrease the risk of death from measles 81% in areas of prevalent vitamin A deficiency.

See also: **Signaling:** Retinoic Acid Receptors; Steroid/Thyroid Hormone Receptors.

Further Reading

- Fields AL, Soprano DR, and Soprano KJ (2007) Retinoids in biological control and cancer. *Journal of Cellular Biochemistry* 102: 886–898.
- Maden M (2007) Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nature Reviews Neuroscience* 8: 755–765.
- Moro JR, Iwata M, and von Andriano UH (2008) Vitamin effects on the immune system: Vitamins A and D take centre stage. *Nature Reviews Immunology* 8: 685–698.
- Napoli JL (2000) Retinoic acid: Its biosynthesis and metabolism. *Progress in Nucleic Acids Research* 63: 139–188.
- Newcomer ME (1995) Retinoid-binding proteins: Structural determinants important for function. *FASEB Journal* 9: 229–239.
- Noy N (2007) Ligand specificity of nuclear hormone receptors: Sifting through promiscuity. *Biochemistry* 46: 13461–13467.
- Travis GH, Golczak M, Moise AR, and Palczewski K (2007) Diseases caused by defects in the visual cycle: Retinoids as potential therapeutic agents. *Annual Reviews of Pharmacology and Toxicology* 47: 469–512.

Vitamin C

F M Steinberg and R B Rucker, University of California, Davis, CA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Adequate intake (AI) One of the four terms used to define dietary reference intakes (DRIs), which are reference values that are quantitative estimates of nutrient intakes to be used for planning and assessing diets for healthy people. The AI is used when the RDA for a nutrient is not available. It is an RDI that is based on observed or experimentally determined approximations of nutrient intake by a group of healthy people.

Amidation A complex reaction, which in some hormones results in a C-terminal protein modification. A C-terminus consensus sequence is required. The consensus is glycine, followed by two basic amino acids (Arg or Lys). The end product is an amide of glycine ($-\text{NH}-\text{CH}_2-\text{CO}-\text{NH}_2$) with loss of the basic amino acid residues.

Anthocyanins Naturally occurring compounds that impart color to fruits, vegetables, and plants. They also have antioxidant and insecticidal properties.

Carnitine Essential component of fatty acid transport into and out of mitochondria for their eventual oxidation. Carnitine is derived from the amino acid lysine following extensive modification.

Complement A term originally used to refer to the heat labile factor in serum that causes immune cytolysis, the lysis of antibody-coated cells. The term now refers to the entire functionally related system comprising at least 20 distinct serum proteins that is the effector immune cytolysis and other biologic functions immune system-related functions.

Dioxygenase An oxidoreductase that incorporates two atoms of oxygen (from one molecule of O_2) into the (reduced) substrate.

Fenton reaction The formation of $\text{OH}^\bullet$, OH^- , and Fe^{3+} from the nonenzymatic reaction of Fe^{2+} with H_2O_2 .

Gibberellins Plant hormones that stimulate growth in the stem and leaves, and trigger the germination of the seed.

Mehler reaction Describes the photoreduction of oxygen in chloroplasts by photosystem I in plants yielding $\text{O}_2^{\bullet-}$.

Monoxygenase Oxidoreductases that induce the incorporation of one atom of oxygen from O_2 into the substance being oxidized.

Recommended dietary allowance (RDA) The average daily dietary intake level that is sufficient to meet the nutrient requirement of nearly all (97–98%) healthy individuals in a group.

Redox Oxidation–reduction reaction in which electrons are removed from one molecule or atom and given to another molecule or atom.

Transporter A class of transmembrane protein that allows substances to cross plasma membranes far faster than would be possible by diffusion alone. A major class of transport proteins that expend energy to move substances (called active transport) are the transport ATPases.

Violaxanthin Photopigment involved in photoprotection in plants. When light energy absorbed by plants becomes excessive (relative to the capacity of photosynthesis), the xanthophyll (a chemical classification designation), violaxanthin is reversibly modified by violaxanthin de-epoxidase as a protective function in plants.

Ascorbic acid (vitamin C) is one of the most important redox cofactors in plant and animal systems. Although most animals make sufficient ascorbic acid; for some, ascorbic acid is a true vitamin because of an inability to carry out sufficient synthesis or production. This is true for humans and other higher primates, a small number of other mammals (e.g., guinea pigs and bats), and some species of birds and some fish. A deficiency causes the disease scurvy in humans, which results in a range of pathologic symptoms, because of defects in ascorbic-acid-specific enzymatic steps and processes. Scurvy in its most severe stages impacts collagen-related supporting structures and can lead to skin lesions and bleeding from the mucous membranes. In advanced scurvy, suppurating wounds, loss of teeth, and psychological changes may occur. Scurvy was first described by Hippocrates (c. 460 BC–c. 380 BC). A Scottish surgeon in the British Royal Navy, James Lind, was among the first to recognize scurvy as a nutritional disease. In early literature, scurvy is sometimes referred to as Barlow's disease,

Moeller's disease, and Cheadle's disease. Ascorbic acid was isolated in 1933 and synthesized in 1934.

Ascorbate synthesis occurs as a glycogenolysis-dependent process. In most plants and animals, ascorbic acid is derived as a product from the direct oxidation of glucose, galactose, and mannose (in plants). Steps in the reaction are glucose or galactose $\rightarrow$ UDP-D-glucuronic acid $\rightarrow$ glucuronic acid/glucuronolactone $\rightarrow$ gulono-1,4-lactone $\rightarrow$ ascorbic acid. In animals, a key enzyme in this process is L-gulonolactone oxidase (GLO, EC 1.1.3.8), which catalyzes the terminal step in ascorbic acid synthesis, that is, gulono-1,4-lactone is oxidized to ascorbic acid. In keeping with the evolutionary links to glucose metabolism, L-gulonolactone oxidase resides in the kidney of most birds and reptiles, and during the course of evolution was transferred to the liver of mammals. In animals that produce insufficient ascorbic acid, mutations in the L-gulonolactone oxidase gene occurred 50–60 million years ago in animals that produce insufficient ascorbic acid. Accordingly, a dietary source of ascorbic

acid is needed in these animals. There is a possibility that minor alternative pathways exist for ascorbic acid synthesis.

Chemistry

The chemical designation for ascorbic acid is 2-oxo-L-theohexono-4-lactone-2,3-enediol. Ascorbic acid is a near planar five-member ring with two chiral centers that resolves into the four stereoisomers. It has a molecular mass of $176.12 \text{ g mol}^{-1}$, a density of 1.65 g cm^{-3} , is very soluble in water ($>30 \text{ g per 100 ml}$), and variably soluble in most short-chain alcohols and glycols ($1\text{--}5 \text{ g per 100 ml}$). The oxidized form of ascorbic acid, dehydroascorbic acid, retains vitamin C activity and can exist as a hydrated hemi-ketal. Crystalline dehydro-L-ascorbic acid can exist as a dimer. The most important chemical property of ascorbic acid is the reversible oxidation of ascorbic acid to semidehydro-L-ascorbic acid and to dehydroascorbic acid (Figure 1).

In addition to facilitating reduction-oxidation reactions, ascorbic acid has the ability to form relatively stable free radical intermediates. Ascorbic acid can act as a free radical scavenger in reactions involving reactive oxidant species (ROS). In this regard, the rate constants for the generation of ascorbate radicals vary considerably, but often dictate rapid radical formation (e.g., $10^5 - 10^{10} k_{\text{obs}}$, unit $\text{M}^{-1}\text{s}^{-1}$). Once formed, ascorbate (Asc) radicals decay relatively slowly by a process of disproportionation ($2 \text{ Asc}^{\bullet-} + \text{H}^+ \rightarrow \text{AscH}^- + \text{DHA}$). The ability to form free radical intermediates can significantly delay or prevent free radical-initiated oxidations. Ascorbic acid readily scavenges reactive oxygen and nitrogen species, such as superoxide, hydroperoxyl, peroxynitrite, and nitroxide radicals. $\text{Asc}[\text{H}^{\bullet-}]$ donates a hydrogen atom ($\text{H}^{\bullet}$ or $\text{H}^+ + \text{e}^-$) to an oxidizing radical to produce the resonance-stabilized tricarbonyl ascorbate free radical. $\text{Asc}[\text{H}^{\bullet}]$ has a pK_a of -0.86 ; thus, it is not protonated. Further, the unpaired electron of $\text{Asc}[-]$ resides in the π -system that includes the tricarbonyl moiety of

ascorbate. It is a weakly oxidizing and reducing radical. Due to its π -character, $\text{Asc}[-]$ does not react with oxygen to form peroxy radicals capable of damaging oxidations. It is relatively unreactive with a one-electron reduction potential of only $+282 \text{ MV}$, and as such is considered a terminal, small-molecule antioxidant.

Ascorbic acid is often associated with the protection of lipid, DNA, and proteins from oxidants. As examples, when peroxy radicals are generated in plasma, vitamin C is consumed faster than other antioxidants (e.g., uric acid, bilirubins, and vitamin E). Ascorbic acid is 10^3 times more reactive than a polyunsaturated fatty acid in reacting with peroxy radicals. In contrast, ascorbic acid can be viewed as a pro-oxidant under aerobic condition when metals capable of redox ($\text{Fe}^{2+} \leftrightarrow \text{Fe}^{3+}$; $\text{Cu}^+ \leftrightarrow \text{Cu}^{2+}$) are also present. Metals, such as iron and copper in their reduced states, are effective Fenton catalysts.

With regard to other chemical properties, the acidity of ascorbic acid is due to the low pK_a of the proton on oxygen-3. In addition, ascorbic acid is not very stable in aqueous media, wherein it can decay within a few hours or even minutes at high pH (>10.0). By contrast, ascorbic acid is relatively stable in blood (a day or more), or if stored at acid pH (<3.0) or below -20°C (often weeks to months).

Nutritional and Biochemical Importance

Ascorbic acid functions in many mono- and dioxygenases to maintain metals in a reduced state. For example, mono- and dioxygenases usually contain copper or iron as redox cofactors, respectively. As an additional characteristic, dioxygenases require α -ketoglutarate and O_2 as cosubstrates in reactions, while mono-oxygenases require only O_2 . Important reactions and processes that require ascorbic acid include:

- Norepinephrine synthesis – dopamine- β -hydroxylase, EC 1.14.17.1, is the final and rate-limiting step in the synthesis

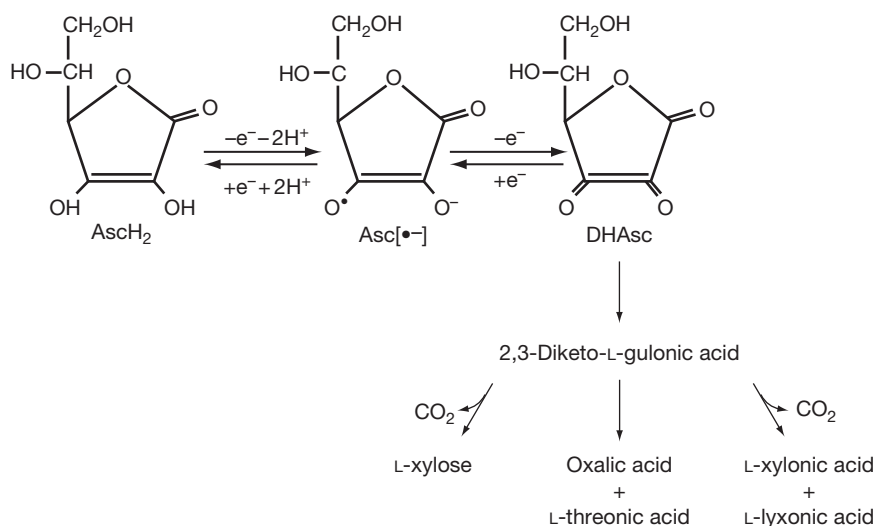


Figure 1 Chemical forms of ascorbic acid, major metabolites and degradation products.

of norepinephrine and requires copper and ascorbic acid for activity.

- Hormone activation – posttranslational steps involving α -amidations activate many hormones and hormone-releasing factors. Examples of hormone activation include melanotropins, calcitonin, releasing factors for growth hormone, corticotrophin and thyrotropin, pro-ACTH, vasopressin, oxytocin, cholecystokinin, and gastrin. Petidylglycine α -amidating mono-oxygenase (EC 1.14.17.3), the enzyme that carries out α -amidation, is found in secretory granules of neuroendocrine cells.
- Carnitine biosynthesis – ascorbate is a cofactor for two of the hydroxylation steps in the pathway of carnitine biosynthesis, gamma-butyrobetaine hydroxylase, and epsilon-N-trimethyllysine hydroxylase. Ascorbate deficiency results in as much as 50% decrease in carnitine in heart and skeletal muscle in animals that require ascorbic acid.
- Prolyl and lysyl hydroxylations – ascorbic acid is an important cofactor in the hydroxylation of proline and lysine amino acid residues. In scurvy, poor wound healing, bruising, and osteopenic abnormalities can result from the under-hydroxylation of collagens. The various types of collagens are distinguished by their relative amounts of hydroxyproline and hydroxylysine. Hydroxylation of proline and lysine is important in collagen assembly and maturation into fibers. Because collagen is the principal connective tissue protein in the body, it is essential to all phases of normal growth, development, and repair. There are also other proteins that are not defined as collagens, which contain hydroxyproline, for example, the subcomponent C1q of complement, acetylcholine esterase, pulmonary surfactant proteins, and proteins that function as cell-surface receptors. In plants, prolyl hydroxylase hydroxylates proline residues in cell wall hydroxyproline-rich glycoproteins required for cell division and expansion. Approximately, 50 hydroxyproline-containing proteins have been identified in viruses and bacteria, wherein ascorbic acid may also play a role as cofactor.

Ascorbic acid, along with glutathione, is an important non-enzymatic antioxidant/reductant in the water-soluble compartment of cells. Glutathione is L- γ -glutamyl-L-cysteine-glycine (GSH). The antioxidant actions of glutathione and ascorbate are closely linked and involve mechanisms in which decreased glutathione may even stimulate ascorbic acid synthesis in those animals that can produce it. During postnatal development, a rapid change occurs as animals adapt from a relatively hypoxic to a relatively hyperoxic environment. In this regard, one of glutathione's many functions is to keep ascorbic acid in a reduced form. An adequate ascorbic acid intake is particularly important in newborns and neonates, who do not yet have the potential to synthesize ascorbate. In adult animals that can make ascorbic acid, a reduction in glutathione levels can lead to a rapid increase in liver dehydroascorbic acid. A deficiency of both ascorbic acid and glutathione can cause pathologic changes to liver and other organs.

The mechanisms for ascorbic acid regeneration are complex. Both nonenzymatic and enzymatic processes have been proposed. In plants, dehydroascorbate reductase (DHAR) catalyzes the regeneration of ascorbic acid from its oxidized state and serves as an important regulator of Asc recycling,

particularly in plants. In animal cells, DHAR is also recognized as the intrinsic activity of thioltransferases (glutaredoxins) and protein disulfide isomerase. These enzymes catalyze the glutathione-dependent two-electron regeneration of ascorbic acid. Their importance is observed in studies of GSH-deficient rodents. GSH deficiency in newborn animals resulted in decreased tissue ascorbic acid and increased dehydroascorbate-to-ascorbate ratios. Administration of ascorbic acid daily to GSH-deficient animals decreased animal mortality and cell damage from oxygen stress. In this regard, a cellular role is proposed for dehydroascorbate in the oxidation of nascent protein dithiols to disulfides catalyzed in the endoplasmic reticulum compartment by protein disulfide isomerase. In chromaffin vesicles and neurohypophyseal secretory vesicles, semidehydroascorbate can be reduced by cytochrome b_{561} to regenerate intravesicular ascorbate. Cytochrome b_{561} , a transmembrane protein, is reduced in turn by an extravesicular electron donor, most probably cytosolic GSH or ascorbic acid. In the erythrocytes of some mammals, cytochrome b_{561} can also reduce extracellular monodehydroascorbate by using intracellular ascorbate (Asc) as an electron donor.

In plant tissues, ascorbic acid reaches a concentration of over 20 mM in chloroplasts and occurs in all cell compartments, including the cell wall. It has a major role in photosynthesis, acting in the Mehler peroxidase reaction with ascorbate peroxidase to regulate the redox state of photosynthetic electron carriers and as a cofactor for violaxanthin de-epoxidase, an enzyme involved in xanthophyll cycle-mediated photoprotection. Ascorbic acid also serves as a cofactor in metabolic pathways important to ethylene, gibberellins, and anthocyanins. There is a rapid response of ascorbate peroxidase expression to (photo)-oxidative stress. Of interest, the ascorbate-deficient vtc 1 mutant of *Arabidopsis thaliana* is very sensitive to ozone, ultraviolet-B radiation, and other forms of oxidant stress.

Cellular Regulation of Ascorbic Acid

Specific nonoverlapping transport proteins mediate the transport of ascorbic acid across biological membranes. In animals, dehydroascorbic acid uptake occurs by the facilitated-diffusion glucose transporters (GLUT1, -3, and -4), but under physiological conditions these transporters do not appear to play a major role in the uptake of dehydroascorbic acid due to the high concentrations of glucose that effectively block influx. In contrast, L-ascorbic acid enters cells via Na^+ -dependent systems. Two isoforms of these transporters have recently been cloned from human and rat sources, which differ in their tissue distribution. Sodium-ascorbate co-transporters (SVCT)1 and SVCT2 import reduced ascorbate across plasma membrane. SVCT2 is involved in vitamin C transport in almost every tissue, except red blood cells which lose SVCT proteins during maturation. Cell accumulation of ascorbic acid occurs, because of cellular dehydroascorbate reduction systems, noted above, that are capable of rapidly generating ascorbic acid. In addition to cell regulation, ascorbate reduction systems are also important in reducing the ascorbic acid radical. As indicated, the ascorbic acid radical is relatively stable. An accessory enzymatic system is needed to reduce the potential of transient accumulation of the ascorbate radical. Excess ascorbate radicals may

initiate free radical cascade reaction or nonspecific oxidations. In plants, nicotinamide adenine dinucleotide (NADH)-monodehydroascorbate reductase (EC 1.6.5.4) has evolved to maintain ascorbic acid in its reduced form. NADH-monodehydroascorbate reductase plays a major role in stress-related responses in plants. In animal tissues, glutathione dehydroascorbate reductase (EC 1.8.5.1) serves this purpose.

Digestion, Absorption, and Nutritional Requirements

When ascorbic acid is consumed in the diet, the ileum and jejunum are major sites of ascorbic acid absorption. The bioavailability of vitamin C is dose dependent. In humans, saturation of transport occurs with dosages of 200–400 mg d⁻¹. Approximately, 70% of a 500 mg dose is absorbed. However, much of the absorbed dose (>50%) is not metabolized and excreted in the urine. With a dose of 1250 mg, only 50% of the dose absorbed and most (>85%) of the absorbed dose is excreted. Vitamin C is not protein bound and is eliminated with an elimination half-life of 10–12 h. In Western populations' plasma, vitamin C concentrations range from 54 to 91 μmol l⁻¹. Some tissues can accumulate as much as 100 times the level of ascorbic acid in blood (e.g., adrenal glands, pituitary, thymus, corpus luteum, and retina). Tissue concentrations range from micromolar to millimolar amounts, that is, high tissue levels of vitamin C are tolerated in mammalian systems. However, there is also evidence that accelerated metabolism or disposal of ascorbic acid occurs after prolonged supplementation of high doses of ascorbic acid, that is, ascorbic acid levels in tissues are maintained by metabolically controlled or regulated processes.

Sources and Detection

A number of approaches are available for ascorbic acid detection. The most common approach involves titration with oxidizing agents. Ascorbic acid absorbs strongly in UV light, the basis for some spectrophotometric assays. The redox properties of ascorbic acid allow for electrochemical detection or detection from interacting with redox-sensitive chromophores and dyes. Enzymatic methods (using ascorbate oxidase) and isotope ratio mass spectrometry are also used for detection and quantitation of ascorbic acid following separation by ion exchange, absorption, or partition chromatographic approaches. Typical food and tissue values are given in Table 1.

Defining Ascorbic Acid Status

Most mammals that require ascorbic acid meet their requirements at intakes of 20–80 mg per 1000 kcal (4.184 MJ). This amount corresponds to what may be extrapolated from synthetic rates in animals that produce ascorbic acid. Such calculations, however, require knowledge of the amount of glucose directed through the direct oxidative pathway and eventually to L-glulonolactone oxidase, allometric scaling approaches, and diurnal fluxes in L-glulonolactone oxidase activity. Regrettably, this is seldom done, which can lead to errors, such as suggesting theoretical intakes of gram quantities of ascorbic acid per day for humans.

Table 1 Vitamin C in selected foods and tissues

<i>Sources of vitamin C</i>	<i>Mg of ascorbic acid per 100 g of wet weight or edible portion</i>
<i>Animal products</i>	
Cow's milk	0.5–2
Human milk	3–6
Beef, pork, veal	2–10
Beef, pork Liver	20–30
Liver, chicken	15–20
Oysters (raw)	30
Kidney, chicken	6–8
Heart, chicken	5
Crab muscle, lobster	1–4
Shrimp	2–4
<i>Fruits</i>	
Apple	3–30
Banana	8–16
Blackberry	8–10
Cherry	15–30
Currant, red	20–50
Currant, black	150–200
Grapefruit	30–70
Kiwi fruit	80–90
Lemon, orange	40–50
Lime	30–45
Melon	9–60
Strawberry	59–70
Pineapple	15–25
Rose hips	250–800
<i>Vegetables</i>	
Beans, various	10–15
Broccoli	70–90
Brussels sprouts	100–120
Cabbage	30–70
Carrot	5–10
Cucumber	6–8
Cauliflower	50–70
Eggplant	15–20
Chive	40–50
Kale	70–100
Onion	10–15
Parsley	90–130
Peas	8–12
Potato	4–30
Pumpkin	15
Radish	25
Spinach	35–40
Tomato	15–20
<i>Condiments</i>	
Chicory	30–40
Coriander (spice)	90
Garlic	15–25
Horseradish	50
Lettuce, various	10–30
Leek	15
Parsley	200–300
Pepper, various	150–200

The requirements for vitamin C in humans (given as the range for female–male and expressed as the recommended dietary allowance (RDA) or adequate uptake (AI)) are as follows: infants, 40 mg d⁻¹ (AI based on what is present in

human milk); children (3–13 years), 15–45 mg d⁻¹ (extrapolated from the RDA for adults); adolescence, 45–65 mg d⁻¹; adults (>19 years), 75–90 mg d⁻¹; lactating mothers, 115–120 mg d⁻¹. In keeping with these estimates, recent clinical investigations have noted vitamin C deficiency in 2–7% of individuals surveyed (often defined as individuals meeting less than two-thirds of the RDA). A serum concentration of 11.4 mmol l⁻¹ is indicative of a vitamin C deficiency and clinical symptoms of scurvy. With regard to marginal vitamin C status based on plasma ascorbic acid values (e.g., <20–28 µg dl⁻¹), the prevalence ranges from 13% to 22% in the US population. Individuals who smoke are more likely to have marginal vitamin C status compared to nonsmokers. Several studies suggest that smokers require over 200 mg vitamin C daily in order to maintain plasma vitamin C concentrations at a level equivalent to that of nonsmokers. Genetic variations in transporters or enzymes involved in ascorbate metabolism can also affect functions or steady-state concentration of the vitamin. Consequently, there is potential for individuals with gene differences to have higher requirements for vitamin C. This is an area for future research.

Summary

Ascorbic acid usually carries out redox reactions by mechanisms dependent upon free-radical processes. Ascorbate metabolism is linked to the metabolism of glutathione. Ascorbic acid is required in animals that lack or have mutations in the gene for L-gulonolactone oxidase. Ascorbic deficiency results in reduced mono- and dioxygenase activities. The consequences of severe deficiency are profound, since growth, extracellular matrix, and hormonal regulation are impaired. Recent data suggest that optimal intakes of ascorbic acid, based on a range of criteria, should be 75–90 mg d⁻¹ for adult humans, and possibly higher in some circumstances.

See also: Protein/Enzyme Structure Function and Degradation: Peptide Amidation.

Further Reading

- Hathcock JN, Azzi A, Blumberg J, et al. (2005) Vitamins E and C are safe across a broad range of intakes. *American Journal of Clinical Nutrition* 81: 736–745.
- Johnson C, Steinberg F, and Rucker RB (2007) Ascorbic acid. In: Zemleni J, Rucker RB, Suttie J, McCormick D, and MacLand L (eds.) *Handbook of Vitamins*, 2nd edn., pp. 489–520. Boca Raton, FL: CRC Press.
- Levine M and Eck P (2009) Vitamin C: Working on the x-axis. *American Journal of Clinical Nutrition* 90: 1121–1123.
- Linster CL and Van Schaftingen E (2007) Vitamin C. Biosynthesis, recycling and degradation in mammals. *FEBS Journal* 274: 1–22.
- Mahan DC, Ching S, and Dabrowski K (2004) Developmental aspects and factors influencing the synthesis and status of ascorbic acid in the pig. *Annual Review of Nutrition* 24: 79–103.
- National Academy Press (2000) *Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids – A Report of the Panel on Dietary Antioxidants and Related Compounds, Subcommittees on Upper Reference Levels of Nutrients and Interpretation and Uses of Dietary Reference Intakes, and the Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine*. Washington, DC: National Academy Press.
- Noctor G and Foyer CH (1998) Ascorbic acid and glutathione: Keeping active oxygen under control. *Annual Review of Plant Physiology and Plant Molecular Biology* 49: 249–279.
- Rucker RB (2007) Allometric scaling, metabolic body size and interspecies comparisons of basal nutritional requirements. *Journal of Animal Physiology and Animal Nutrition (Berlin)* 91: 148–156.
- Schleicher RL, Carroll MD, Ford ES, and Lacher DA (2009) Serum vitamin C and the prevalence of vitamin C deficiency in the United States: 2003–2004 National Health and Nutrition Examination Survey (NHANES). *American Journal of Clinical Nutrition* 90: 1252–1263.
- Wilson JX (2002) The physiological role of dehydroascorbic acid. *FEBS Letters* 527: 5–9.
- Wilson JX (2005) Regulation of vitamin C transport. *Annual Review of Nutrition* 25: 1–21.

Vitamin D

H F DeLuca and M Clagett-Dame, University of Wisconsin–Madison, Madison, WI, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Coactivator A protein that participates with transcription factors to enhance receptor activity in transcription.

Convulsive tetany A disease precipitated by either low blood calcium or low blood magnesium, resulting in a severe convulsive state. It is lethal unless immediately treated with calcium and/or magnesium.

Corepressor A protein that participates with transcription factors to decrease target gene expression.

7-Dehydrocholesterol An intermediate in cholesterol biosynthesis and precursor of vitamin D₃.

Dovonex® A drug marketed by Leo Pharmaceutical containing an analog of 1,25-(OH)₂D₃ and used to treat psoriasis by topical administration.

Ergosterol A plant-derived sterol that is converted to vitamin D₂ upon ultraviolet irradiation.

Hepatocytes Cells comprising approximately 30% of the liver that carry out many of the liver's functions.

Keratinocytes Cells of the skin that produce the keratin proteins and contribute to the skin barrier.

Macrophage A type of white blood cell from the myeloid cell lines that is responsible for engulfing and destroying foreign materials.

Nongenomic Activities that are carried out not involving gene expression.

Osteoblasts Cells of the skeleton that carry out synthesis of new bone and that mediate regulation of bone activities.

Previtamin D An isomer of vitamin D formed during the irradiation process for producing vitamin D; it is in equilibrium with vitamin D.

Proximal convoluted tubule cells Cells lining the nephron leading from Bowman's capsule to the loop of Henle in the kidney.

RANKL A signal protein produced by osteoblasts and stromal cells in response to stimulants such as 1,25-(OH)₂D₃ and parathyroid hormone, and a ligand for the receptor activator for nuclear factor, kappa B, protein also known as the tumor necrosis factor-related activation-induced cytokine (TRANCE) or osteoprotegerin ligand.

Transcription factor A protein that controls whether genes are transcribed or not.

Vitamin D-dependent rickets Results from a mutation in either the 1 α -hydroxylase of vitamin D (type I) or the vitamin D receptor (VDR) (type II) that may respond to high levels of vitamin D depending on the site of mutation.

Vitamin D hormone The biologically active form of vitamin D₃ also referred to as 1 α ,25-dihydroxyvitamin D₃, 1,25-(OH)₂D₃, 1 α ,25-dihydroxycholecalciferol, and calcitriol.

Vitamin D-resistant rickets A disease in which rickets fails to respond to normal doses of vitamin D.

Chemical Identity, Natural Occurrence, and Production

The vitamin Ds are secosterols in which the B-ring of the steroid nucleus is replaced by a diene bridge and an exocyclic methylene group. The two most common forms are vitamin D₂ (ergocalciferol) and vitamin D₃ (cholecalciferol). The structures of these two nutritionally important compounds are illustrated in [Figure 1](#).

Although it is classified as a vitamin because of the nature of the discovery process, vitamin D in actual fact should not be considered a true vitamin for the following reasons. First and foremost, vitamin D utilized by higher organisms can be formed in the epidermis of skin by photolysis of 7-dehydrocholesterol, an intermediate in cholesterol biosynthesis. Second, vitamin D is nearly absent from the food supply. Vitamin D is found in fish liver oils, some fatty fish, and in egg yolk but is not found in virtually all plant materials, in skeletal meats, seeds, fruits, and vegetables. In fact, very little is found in an expected source, milk. We consider milk an important supply of vitamin D in many countries, primarily because it is fortified by the addition of vitamin D.

The chemical process whereby 7-dehydrocholesterol is converted to vitamin D₃ is well known. The 5,7-diene portion of either ergosterol or 7-dehydrocholesterol will absorb 280–310-nm ultraviolet light. This causes an isomerization to

form a compound, previtamin D₃, that in itself is biologically inactive and remains in skin. Previtamin D is in equilibrium with vitamin D₃ and, as the vitamin D₃ is formed, it is transported from the skin to the circulation. The plant sterol, ergosterol, is chemically converted to vitamin D₂ by an identical photolysis process. For many years, vitamin D₂ was the major synthetic form of vitamin D used in vitamin capsules, animal feeds, and in the fortification of foods. Because of improved synthetic chemical methods, vitamin D₃ has become the major form used for both fortification in the human food supply and nutrition of domestic animals. The photolysis of 7-dehydrocholesterol is quite efficient in skin, and summer sun exposure for 5–10 min will provide the daily requirement for vitamin D (20–30 $\mu\text{g d}^{-1}$) for optimal bone health. Winter sun is unable to induce production of vitamin D because of the angle of the sunlight and the filtration of the critical wavelengths that cause vitamin D production. It is generally accepted that skin production of vitamin D₃ is quantitatively the most important source of the vitamin for human health.

Conversion of Vitamin D to Its Hormonal Form

The vitamin D₃ that is produced in skin or that is absorbed from the intestine from vitamin pills or fortified foods is

biologically inactive as such. The body must process vitamin D to form a final vitamin D hormone that is believed to carry out all of the known functions of vitamin D. Vitamin D₃ is first hydroxylated, largely by the liver hepatocytes to form the circulating form of vitamin D₃, 25-hydroxyvitamin D₃ (25-OH-D₃). Interestingly, this compound is also biologically inactive at physiologic levels and must be further modified before function. The second step occurs in the proximal convoluted tubule cells of the kidney where a 1 α -hydroxyl group is added to the molecule to produce the final vitamin D hormone, 1 α ,25-dihydroxyvitamin D₃ (1,25-(OH)₂D₃). The reaction steps that result in this transformation are shown in Figure 2. At least two enzymes are known to hydroxylate vitamin D in the 25-position: a mitochondrial enzyme (CYP27A1) and a microsomal enzyme (CYP2R1), both of which have been cloned. The microsomal enzyme is likely the physiologically relevant 25-hydroxylase. Although a report of a human mutation in the CYP2R1 has appeared, this mutation does not prevent 25-hydroxylation, arguing that more than one enzyme can carry out this step. The 25-hydroxylation step appears largely unregulated. However, the second step occurring in the proximal convoluted tubule cells of the kidney is highly regulated. This 1 α -hydroxylase, also known as CYP27B1, has also been cloned and natural human mutants of this enzyme are now well known. These mutations result in a disease called vitamin D-dependent rickets type I, in which children present with vitamin D-deficiency disease, that is, rickets, hypocalcemia, and hypophosphatemia, even though

they are naturally irradiated with sunlight or are provided vitamin D as a supplement. These children are cured by the administration of physiologic amounts of the natural vitamin D hormone, 1,25-(OH)₂D₃, establishing clearly the two-step activation process of vitamin D for humans.

The Role of 1,25-(OH)₂D₃ in Calcium, Phosphorus, and Bone Metabolism

The basis for discovery of vitamin D in the first place is the disease rickets in children. This disease is also accompanied by hypocalcemia (low blood calcium) that causes convulsive tetany. In adults, the deficiency of vitamin D causes the disease, osteomalacia. The diseases, rickets and osteomalacia, result from a failure to mineralize the organic matrix of newly synthesized bone. Early investigations to understand how vitamin D functions were, of course, directed toward the formation of mineralized bone. However, it is now abundantly clear that 1,25-(OH)₂D₃ does not act directly on the mineralization process in the healing of rickets or osteomalacia but rather acts on intestine, kidney, and bone to provide calcium and phosphorus in the plasma required for the mineralization of skeleton and prevention of the convulsive disease, hypocalcemic tetany. The essence of vitamin D action, therefore, in curing rickets and osteomalacia is the elevation of plasma calcium and phosphorus to supersaturating levels that are required for normal mineralization. The vitamin D hormone acts to provide this calcium primarily by stimulating the enterocyte of the small intestine to absorb calcium as well as phosphate by an independent mechanism both of which require metabolic energy.

Because of the constant need for calcium to support normal neuromuscular function, calcium must be available in the plasma even when it is not available in the diet. The vitamin D hormone, that is, 1,25-(OH)₂D₃, acting together with parathyroid hormone will cause the mobilization of calcium from the skeleton into the plasma (bone resorption). Thus, the skeleton serves not only as a structural function but also as a reservoir of calcium and phosphorus. A final point is that 1,25-(OH)₂D₃ works in concert with the parathyroid hormone to induce the distal nephron of the kidney to reabsorb the last 1% of the filtered load of calcium into the plasma compartment. These systems then work to maintain plasma calcium and phosphate at levels that support both bone formation and

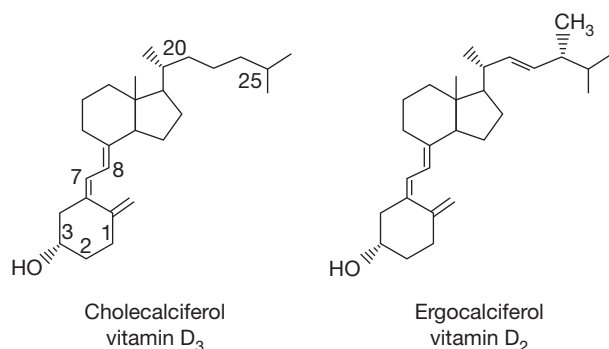


Figure 1 The nutritionally important forms of vitamin D. Vitamin D₃ is the form manufactured in skin and vitamin D₂ is the form produced by the irradiation of ergosterol. In mammals, both are equally active and appear to be metabolized by almost identical routes.

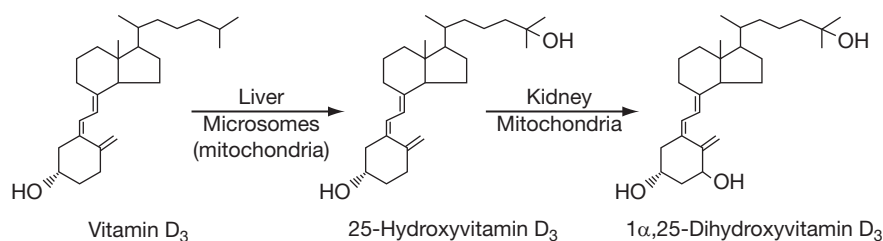


Figure 2 The required metabolism of vitamin D for function. Shown here is the metabolism of vitamin D₃ to the major blood form of vitamin D, 25-hydroxyvitamin D₃, and its subsequent metabolism in the kidney to form the final vitamin D hormone, 1 α ,25-dihydroxyvitamin D₃. It is this form of vitamin D₃ that carries out all of the known functions of the vitamin and its production is very tightly regulated. Vitamin D₂, not shown here, is metabolized in an almost identical fashion to its active form, 1 α ,25-dihydroxyvitamin D₂.

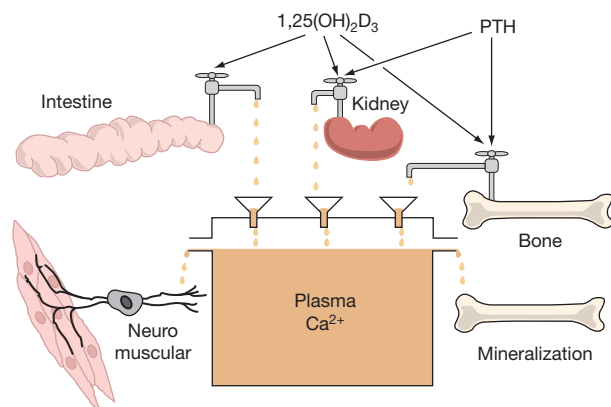


Figure 3 A diagrammatic representation of how the vitamin D hormone ($1,25\text{-(OH)}_2\text{D}_3$) functions to maintain plasma calcium at supersaturating levels which are required for the prevention of hypocalcemic tetany (neuromuscular) and for the prevention of rickets and osteomalacia (mineralization of bone).

prevent hypocalcemic tetany. A diagram representing these functions is shown in [Figure 3](#).

The Vitamin D Endocrine System

One of the most elegantly regulated substances in the body is plasma calcium. The parathyroid glands continuously monitor plasma calcium levels, and a G-protein-coupled calcium-sensing receptor therein responds when calcium falls even slightly below normal. This leads to the secretion of parathyroid hormone, a peptide hormone, that is the body's signal requesting calcium. It binds to the entire nephron of the kidney and to the osteoblasts of bone and, among its functions in the kidney, it stimulates the 1α -hydroxylase (CYP27B1) to produce the vitamin D hormone. The vitamin D hormone then acts on the bone, kidney, and intestine as described above to mobilize calcium to raise blood calcium into the normal range, clearing the set point of the parathyroids shutting down parathyroid secretion. Overshoot of calcium results in the C-cells of the thyroid secreting calcitonin, a 32-amino-acid peptide hormone that blocks calcium mobilization from the skeleton. Calcitonin also has a stimulatory effect on the 1α -hydroxylase to produce the vitamin D hormone in small amounts to support its non-calcemic activities described in the following.

The Vitamin D Receptor

$1,25\text{-(OH)}_2\text{D}_3$ is a true steroid hormone that acts through a nuclear receptor called the vitamin D receptor (VDR). The human receptor is a 427-amino-acid protein that serves as a ligand-dependent transcription factor. It is a member of the nuclear receptor superfamily of receptors. Other members of this superfamily include the retinoic acid, thyroid, estrogen, androgen, glucocorticoid, and other steroid hormone receptors. VDR is the smallest of the steroid hormone receptors and largely acts in a fashion quite similar to other class II nuclear

receptors. Unlike other nuclear receptors, there is only a single VDR in the target tissues. No subtypes are known and it is this receptor through which all known vitamin D actions occur. Natural mutants of the human VDR are known and produce a disease called vitamin D-dependent rickets type II. In this disease, high blood levels of $1,25\text{-(OH)}_2\text{D}_3$ occur in the face of severe rickets and severe hypocalcemia and hypophosphatemia. VDR-null mutant mice have been produced and are available as experimental tools, and these mice largely phenocopy the most severely affected patients. The vitamin D-dependent rickets type II patients vary in disease severity depending upon where in the VDR the mutation has occurred. If there is a complete absence of a functional receptor, then very brittle vitamin D-resistant rickets occurs and can only be treated by the infusion of calcium and phosphorus into the blood stream to mineralize the skeleton, but many of the other functions of vitamin D go unmet. The fact that VDR-null mutant mice are approximately normal at birth shows that vitamin D does not play a major role in embryonic development but functions primarily after parturition.

There are reports of nongenomic functions of vitamin D. These have been largely cellular-based observations which have yet to be shown to occur *in vivo*. The administration of large amounts of $1,25\text{-(OH)}_2\text{D}_3$ to VDR-null mutant mice produces no significant phenotype, suggesting that the only mechanism of action of the vitamin D hormone is through its receptor under physiologic circumstances.

Molecular Mechanism

Vitamin D-response elements (VDREs) are found in the promoter region of target genes and more recently have been located at sites quite distant from the promoter. These VDREs are hexameric repeats (although the sequences are usually not exact repeats) separated by three nonspecified bases. Although response elements with two repeats are most common, more complex response elements have been described. The VDR must heterodimerize with another transcription factor, the retinoid X receptor (RXR), in order to bind to the responsive elements. The RXR binds to the 5' repeat and the VDR to the 3' repeat. [Figure 4](#) provides a cartoon to illustrate what is currently known about the mechanism of action of the vitamin D hormone working through its receptor to regulate expression of target genes. As with other class II nuclear receptors, it is anticipated that the unliganded receptor may be retained in an inactive state by being bound to a corepressor complex. This has not been specifically shown for the VDR but is known for other class II nuclear receptors. Binding of the ligand would cause a release of the corepressor. The receptor would then be free to interact with a coactivator protein that is believed to be involved in chromatin remodeling and to link the complex bound at the hormone response element to the basal transcription machinery. Among the known coactivators believed to interact directly with the VDR are steroid receptor coactivator (SRC) 1, 2, 3 and vitamin D-interacting protein 205 (DRIP 205). Likely, other participating coactivators will be discovered. A phosphorylation does occur on serine 205 of the VDR but whether this is of functional importance remains to be

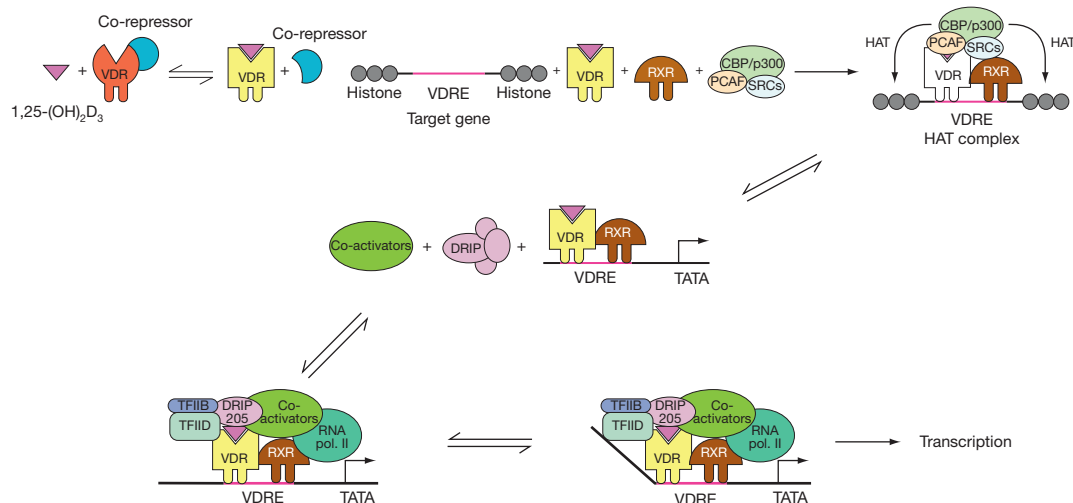


Figure 4 A cartoon representing the molecular mechanism of action of the vitamin D hormone (1,25-(OH)₂D₃) in regulating transcription of target genes in the target tissues. Binding of the ligand, that is, 1,25-(OH)₂D₃ to the VDR/RXR, causes a conformation change of the VDR enabling binding of a coactivator complex leading to chromatin remodeling and enhanced gene transcription. Several proteins are required to form the transcription complex, all of which have not yet been identified. There is a bending of the DNA and a phosphorylation on serine-205 but their exact role in either suppression or activation of transcription remains unknown. VDRE, vitamin D hormone response element; TATA, a consensus sequence of thymine/adenine found in the promoter region of most genes that seems important in determining accurately the position at which transcription is initiated; TFIIB, transcription factor IIB; TFIID, transcription factor IID; RNAPII, ribonucleic acid polymerase II.

determined. More recently, additional VDRE sites have been located on target genes that act as enhancers of gene transcription. These enhancers are believed to act in a three-dimensional sense to recruit and increase effectiveness of transcription machinery. The mechanism whereby the VDR–RXR heterodimer regulates gene suppression is less well understood.

Metabolism and Degradation

The vitamin D hormone is degraded specifically by an enzyme it induces. This enzyme is the CYP24A1 (24-hydroxylase) which carries out a series of reactions on the side chain of the vitamin D hormone and its precursor, 25-OH-D₃. It first causes hydroxylation of the 24R-position, followed by oxidation of that same hydroxyl to a ketone. The next step again carried out by the same enzyme is a 23-hydroxylation, followed by cleavage and oxidation, giving rise to the biliary excretion product called calcitroic acid in the case of 1,25-(OH)₂D₃. Calcitroic acid is biologically inactive. 25-OH-D₃ is believed to be converted in analogous steps to the corresponding C-23 acid or 1-deoxycalcitroic acid. Thus, the vitamin D hormone, an extremely potent biological substance, ensures its limited activity by inducing its own destruction. Other metabolism of 1,25-(OH)₂D₃ and 25-OH-D₃ is known, including the formation of the 26,23-lactone derivative, 26-hydroxylation; however, quantitatively the most important reaction regulating the level of vitamin D hormone is carried out by the CYP24A1. The CYP24A1 is regulated in a negative fashion by parathyroid hormone causing an instability of the messenger RNA (mRNA) encoding for that enzyme. Thus, the parathyroid hormone not only stimulates the 1 α -hydroxylase but also stimulates the destruction of the enzyme that destroys the

hormone, thereby ensuring large amounts of the vitamin D hormone are available to carry out its calcium functions.

New and Nonclassical Roles of the Vitamin D Hormone

With the discovery of the VDR came the observation that it is found in tissues not previously appreciated as target tissues of vitamin D action. The discovery of the VDR in the intestine, kidney tubules, and osteoblasts was not unexpected. However, its presence in the parathyroid gland, the islet cells of the pancreas, the immune cells such as T-cells and macrophages, keratinocytes, and other cells made it possible that the vitamin D hormone had important functions besides raising blood calcium and phosphate and mineralizing the skeleton. It is now clear that an important function of the vitamin D hormone is to suppress the growth of parathyroid glands and to suppress production of the parathyroid hormone. Physicians have taken advantage of this basic function when using 1,25-(OH)₂D₃ and its analogs to treat secondary hyperparathyroidism in patients with kidney failure.

Other important VDR sites include the promyelocytes which are the precursors of the monocyte. The vitamin D hormone stimulates differentiation of the promyelocyte to the monocyte. It also stimulates RANKL production by osteoblasts and stromal cells. RANKL catalyzes the formation of giant osteoclasts that are responsible for bone resorption. It also activates resting or inactive osteoclasts. This results in bone resorption, the first step in bone remodeling. 1,25-(OH)₂D₃ also induces new bone formation by the osteoblast, thus playing an important role in both arms of the bone remodeling system.

The role of the vitamin D hormone in the keratinocyte also has therapeutic importance. The vitamin D hormone

added to the keratinocytes in cultures has been shown to inhibit cell proliferation and promote differentiation. Thus, $1,25\text{-(OH)}_2\text{D}_3$ and analogs have been successfully used in the topical treatment of psoriasis, a disease of hyperproliferation of the keratinocytes.

Finally, attention must be focused on the immune system where the vitamin D hormone is now known to cause important immunomodulatory effects. The vitamin D hormone can be used to block certain animal models of autoimmune disease. For example, it can be used to prevent experimental autoimmune encephalomyelitis (EAE), a model of multiple sclerosis in mice. However, the level of $1,25\text{-(OH)}_2\text{D}_3$ required also induces an abnormal rise in serum calcium. $1,25\text{-(OH)}_2\text{D}_3$ can also prevent diabetes in the nonobese diabetic mouse model of type I diabetes. It can be used to prevent disease in animal models of rheumatoid arthritis, systemic lupus, and inflammatory bowel disease. Although there is a great deal of effort being expended, the exact role of the vitamin D hormone in the immune system remains unclear.

Therapeutic Uses of $1\alpha,25\text{-(OH)}_2\text{D}_3$ and Its Analogs

As pointed out above, besides the treatment of vitamin D-dependent rickets type 1 and vitamin D-resistant rickets of many types, the vitamin D hormone and its analogs are certainly central actors in the treatment of bone disease secondary to kidney failure. This is primarily by suppression of parathyroid proliferation and by suppression of parathyroid hormone production. $1,25\text{-(OH)}_2\text{D}_3$ and its analog, $1\alpha\text{-OH-D}_3$, have been successfully used in certain parts of the world to treat postmenopausal and age-related osteoporosis, giving a modest rise in bone mass along with a clear reduction in fracture rate. The danger of hypercalcemia has limited their use in the treatment of this disease. However, newer vitamin D analogs with greater bone anabolic activity than $1,25\text{-(OH)}_2\text{D}_3$ are currently under development and may make it possible to safely treat osteoporosis in the future.

An analog of $1,25\text{-(OH)}_2\text{D}_3$ called Dovonex[®] is currently marketed for the topical treatment of the hyperproliferative skin disorder, psoriasis. Improved analogs will likely appear for the treatment of psoriasis. There are potential new uses for vitamin D analogs in the autoimmune diseases as described

above or in the treatment of malignant cancer growth. The vitamin D hormone has been found to suppress growth of malignant cells *in vitro* and to cause their differentiation into functional cells. Thus, the use of vitamin D compounds as a potential treatment of neoplastic disease has been exciting but it has not yet yielded a therapeutic agent. The major problem with the use of $1,25\text{-(OH)}_2\text{D}_3$ and many of its analogs for the treatment of disease is that its primary purpose is to raise blood calcium and phosphorus, which means that hypercalcemia is a major side effect that must be eliminated before an analog can be used to treat noncalcemic diseases. One approach to this is the synthesis of analogs that retain the ability to suppress cancerous growth but do not raise blood calcium. New analogs with these desirable characteristics are currently under basic study, but are not yet available for pharmaceutical use.

See also: **Bioenergetics:** Calcium-Buffering Proteins: Calbindin; Calcium-Sensing Receptor (CaSR); ER/SR Calcium Pump: Structure; **Metabolism Vitamins and Hormones:** Diabetes; **Protein/Enzyme Structure Function and Degradation:** Zinc Fingers; **Signaling:** Parathyroid Hormone/Parathyroid Hormone-Related Protein Receptor; Steroid/Thyroid Hormone Receptors; Vitamin D Receptor.

Further Reading

- Darwish HM and DeLuca HF (1996) Recent advances in the molecular biology of vitamin D action. In: Cohn WE and Moldave K (eds.) *Progress in Nucleic Acid Research and Molecular Biology*, pp. 321–344. San Diego, CA: Academic Press.
- DeLuca HF (1974) Vitamin D: The vitamin and the hormone. *Federation Proceedings* 33: 2211–2219.
- DeLuca HF (1997) Historical overview. In: Feldman D, Glorieux FH, and Pike JW (eds.) *Vitamin D*, vol. 1, pp. 3–11. San Diego, CA: Academic Press.
- DeLuca HF (2008) Evolution of our understanding of vitamin D. *Nutrition Reviews* 66(supplement 2): S73–S87.
- Feldman D, Glorieux FH, and Pike JW (eds.) (1997) *Vitamin D*, 1285pp. San Diego, CA: Academic Press.
- Holick MF (2005) Photobiology of vitamin D. In: Feldman D, Pike J, and Glorieux FH (eds.) *Vitamin D*, vol. 1, pp. 37–45. San Diego, CA: Elsevier Academic Press.
- Jones G, Strugnell SA, and DeLuca HF (1998) Current understanding of the molecular actions of vitamin D. *Physiological Reviews* 78: 1193–1231.
- Plum LA and DeLuca HF (2010) The functional metabolism and molecular biology of vitamin D action. *Clinical Reviews in Bone Mineral and Metabolism* 61–97.

Vitamin D Receptor

D R Dowd and P N MacDonald, Case Western Reserve University, Cleveland, OH, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Comodulatory proteins Proteins that interact directly with the nuclear receptor and alter the transcriptional regulatory activity of the receptor. Coactivators enhance the activity of the receptor, and corepressors attenuate the transcriptional regulatory activity of the receptor.

Histone acetyltransferase (HAT) Enzymes that add acetyl groups to the positively charged lysine residues on core histones, effectively negating the positive charge. Histone hyperacetylation is correlated with areas of active transcription, while hypoacetylation is correlated with nontranscribed promoters.

1,25-Dihydroxyvitamin D₃ (1,25(OH)₂D₃) The active form of vitamin D in mammals.

Preinitiation complex (PIC) The complex of basic transcription factors that are necessary and sufficient for the

initiation of transcription to occur. In addition to RNA polymerase II, the PIC consists of transcription factor IIA (TFIIA), TFIIB, TFIIE, TFIIIF, and TATA-binding protein or TBP.

Retinoid X receptor (RXR) A member of the nuclear receptor superfamily of transcription factors that serves as a common heterodimeric partner for many of the class II nuclear receptors, including retinoic acid receptor, vitamin D receptor (VDR), thyroid hormone receptor, and peroxisome proliferator-activating receptor.

Vitamin D responsive elements (VDRE) Specific sequences of DNA that the VDR–RXR heterodimer selectively recognize and bind. This binding event is one of the initial steps in the mechanism through which 1,25-dihydroxyvitamin D₃ and the VDR regulate gene transcription.

Introduction

Vitamin D₃, or cholecalciferol, was discovered nearly a century ago as a micronutrient that is essential for normal skeletal development and for maintaining bone integrity. 1,25(OH)₂D₃ is the bioactive, hormonal form of vitamin D. However, vitamin D₃ is more appropriately classified as a hormone because it is produced in the body in response to serum calcium levels and it functions through a mechanism that is analogous to other steroid hormones. 1,25(OH)₂D₃ is generated by two sequential hydroxylations of vitamin D₃ in response to hypocalcemia and elevated parathyroid hormone. The predominant role of 1,25(OH)₂D₃ is to enhance the intestinal absorption of dietary calcium and phosphorus. 1,25(OH)₂D₃ also acts on mineral-regulating target tissues such as intestine, bone, kidney, and parathyroid glands to maintain normal calcium and mineral homeostasis. 1,25(OH)₂D₃also directly affects skeletal bone remodeling by causing osteoblasts to terminally differentiate into osteocytes and deposit calcified matrix. 1,25(OH)₂D₃also promotes the differentiation of precursor cells into mature osteoclasts which function to resorb bone and maintain appropriate bone remodeling. Thus, it is through an integrated series of diverse effects that vitamin D is thought to preserve and maintain the integrity of the bony tissues. To highlight this role, vitamin D₃ deficiency leads to rickets in children and osteomalacia in adults. The vitamin D endocrine system is also involved in a number of other important physiological processes including blood pressure regulation, immune function, mammary gland development, hair follicle cycling, and protection against chemical and ultraviolet light-induced skin tumorigenesis.

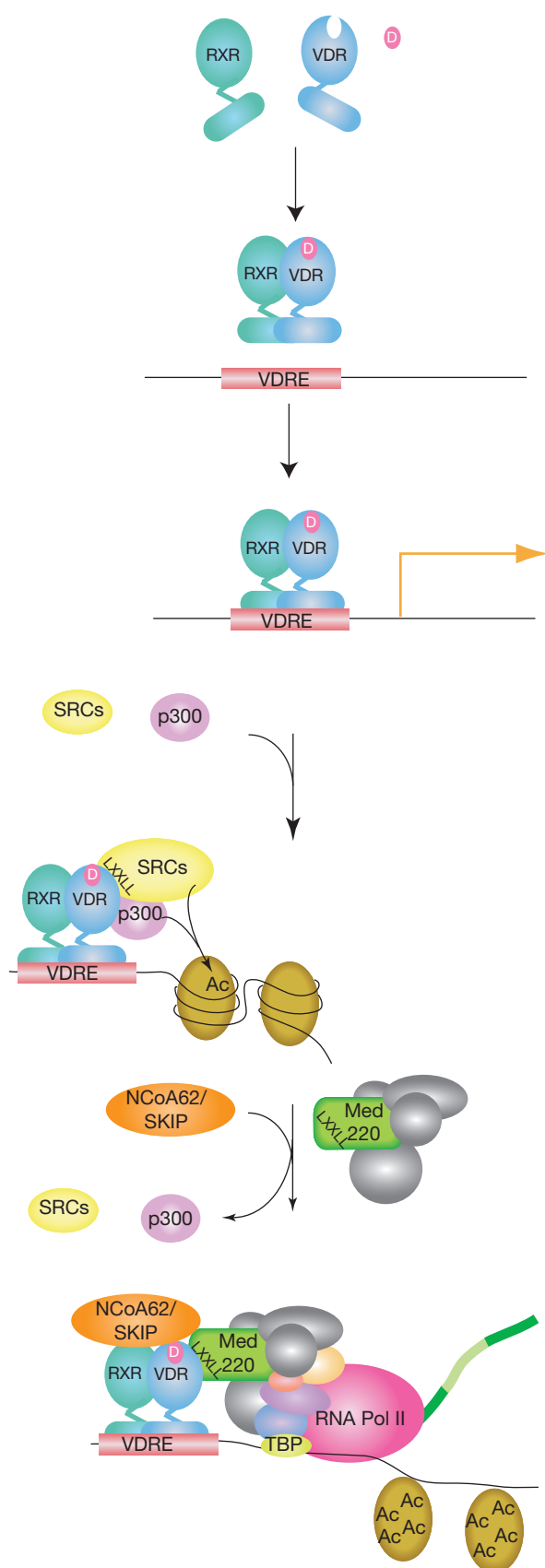
The biological effects of 1,25(OH)₂D₃ are mediated through a soluble receptor protein termed the vitamin D receptor (VDR), a member of the steroid receptor or nuclear receptor (NR)

superfamily. The VDR binds 1,25(OH)₂D₃ with high affinity and high selectivity. In the target cell, the interaction of the 1,25(OH)₂D₃ hormone with VDR initiates a complex cascade of molecular events culminating in alterations in the rate of transcription of specific genes or gene networks (Figure 1). Central to this mechanism is the requisite interaction of VDR with retinoid X receptor (RXR) to form a heterodimeric complex. The dimer binds to specific DNA sequence elements (vitamin D-responsive elements (VDREs)) in vitamin D-responsive target genes and ultimately influences the rate of RNA polymerase II-mediated transcription. An important concept in this mechanism is that protein–protein interactions between the VDR–RXR heterodimer, transcriptional comodulatory proteins and the transcription machinery are essential for the mechanism of vitamin D-mediated gene expression. This article discusses the molecular biology of the VDR with a focus on the macromolecular interactions that are required for the transcriptional regulatory activity of the VDR.

Functional Domains of the VDR

The N-Terminal DNA-Binding Domain

Receptors in the NR superfamily generally have two transcriptional activation domains, termed activation function (AF) 1 and 2, which are required for the receptor to function as a ligand-activated transcription factor. The AF-2 domain exists at the extreme COOH terminus, while the AF-1 domain is located in the N-terminal region (or A/B domain) of the receptors (Figure 2). The AF-1 domain is a constitutive (i.e., hormone-independent) activation domain. In the VDR, the N-terminal A/B domain is truncated compared to other NRs and, thus, an analogous constitutive AF-1 domain is lacking in the VDR.



For VDR to regulate transcription of target genes, it must recognize and bind to DNA in the promoter regions of vitamin D-responsive genes. It does so through a specialized DNA-binding domain (DBD) located near the amino terminus of the receptor (Figure 3). This DBD is required for the VDR to bind selectively and with high affinity to specific DNA sequences termed VDREs. The minimal DBD of the VDR that mediates VDR–DNA interactions resides between amino acid residues 22 and 113 in the human sequence. There are nine cysteine residues within the DBD that are conserved throughout the family members. The first eight of these cysteines (counting from the N terminus) tetrahedrally coordinate two zinc atoms to form two zinc-finger DNA-binding motifs. Mutation of the first eight of the nine cysteine residues to serines eliminates VDR binding to both nonspecific and specific DNA sequences and eliminates VDR-mediated transactivation. Indeed, inactivating mutations within the DBD of the human VDR are responsible for the rare-inherited disorder termed vitamin D-resistant rickets type II. These patients express a nonfunctional VDR that cannot bind to DNA, and they exhibit classical symptoms of vitamin D deficiency that are not corrected by providing an external source of vitamin D.

Ligand-Binding Domain

The large carboxyl-terminal ligand-binding domain (LBD) of VDR is organized into 12 α -helices. As the name suggests, the LBD is responsible for high-affinity binding of the 1,25(OH)₂D₃ ligand, exhibiting equilibrium binding constants of the order of 10^{-10} – 10^{-11} M. Both the 1-hydroxyl and 25-hydroxyl moieties are crucial for efficient recognition by VDR, and the nonhydroxylated, inactive parent vitamin D₃ compound does not bind appreciably to the VDR.

In addition to hormone binding, the LBD is required for several other aspects of receptor function, particularly in mediating protein–protein interactions. One important protein–protein contact is the heterodimerization of VDR with the RXR. As mentioned above, VDR–RXR heterodimer formation is generally required for high-affinity interaction of the receptor with VDREs and an extensive heterodimerization interface that coalesces around helices 10 and 11 in the LBD of VDR mediate protein–protein contacts with RXRs.

The LBD also mediates association of VDR with comodulatory proteins. Many of these interactions occur at the extreme C-terminus of the receptor, where the AF-2 domain is located (Figure 2). The AF-2 domain is highly conserved throughout the hormone receptor superfamily and its main structural

Figure 1 Mechanism of action of the vitamin D receptor. VDR binds to its ligand, 1,25(OH)₂D₃, heterodimerizes with RXR, and binds to a DR-3 type VDRE in the promoter of 1,25(OH)₂D₃-responsive genes. VDR occupies the proximal or 3' half-site, while RXR occupies the distal or 5' half-site. VDR/RXR then recruits coactivators such as SRCs and p300/CBP. The HAT activity of these coactivators loosens the chromatin structure, allowing for a more transcriptionally permissive environment. SRCs and p300/CBP dissociate, allowing for binding of other coregulatory proteins including and the Mediator-D multimeric complex. This latter complex is thought to recruit RNA Pol II and the core transcriptional machinery to initiate active transcription of the target gene.

feature is that of an amphipathic α -helix. Removing 25 amino acids from the C-terminus of hVDR ($\Delta 403$ –427), which contains the AF-2 domain, results in a complete loss of $1,25(\text{OH})_2\text{D}_3$ /VDR-activated transcription. This loss of function is not due to altered binding of RXR, VDRE, or hormone and the mutant receptor was appropriately targeted to the cell nucleus. Thus, the AF-2 domain of VDR, corresponding to helix 12, plays a central role in $1,25(\text{OH})_2\text{D}_3$ -activated transcription mediated by VDR.

Molecular Mechanism of Transcriptional Control by VDR

VDR Interaction with VDREs

VDR modulates transcription by binding to specific VDREs, in the promoter regions of responsive genes. VDREs from a variety of target genes have been identified and on this basis, a consensus VDRE may be generally described as an imperfect direct repeat of a core hexanucleotide sequence, G/A G G T G/C A , with a spacer region of three nucleotides separating each half element (also termed DR-3 for direct repeat with a three nucleotide spacer). The mechanism of VDR binding to VDREs is reflected in the direct repeat nature of the element. For example, purified VDR alone does not bind to a VDRE with high affinity. Rather, RXR is required for high-affinity binding of

VDR to VDREs. The asymmetry of the directly repeating motif causes the VDR–RXR heterodimer to bind to the VDRE with a defined polarity, with RXR occupying the 5' half-site and VDR occupying the 3' half-site.

One important role for the $1,25(\text{OH})_2\text{D}_3$ ligand in the transactivation process is to dramatically enhance both the formation of the VDR–RXR heterodimer and the subsequent interaction of the heterodimer with the VDRE (Figure 1). Upon ligand binding, there is a clear $1,25(\text{OH})_2\text{D}_3$ -dependent decrease in the formation of VDR homodimers and a concomitant increase in VDR heterodimerization with RXR. These ligand-induced changes in VDR–RXR interactions are likely due to altered conformations of the VDR that disrupts weak homodimers of unliganded VDR and promotes liganded VDR heterodimerization with RXR. The interaction of VDR and RXR generates a heterodimeric complex that is highly competent to bind DR-3-like VDREs and subsequently affect the transcriptional process. Thus, the VDR–RXR heterodimer is the primary active complex involved in controlling DR-3 VDRE-driven promoters. Interestingly, VDREs may function over very long distances. A concept is emerging in which VDREs exist at sites that are extremely distal (>50 kb) to the transcriptional start site and at sites both upstream and downstream of the start site.

The natural VDREs identified thus far provide only a snapshot of the DNA sequences that mediate the transcriptional effects of the VDR. Variations on the DR-3 motif for VDREs have been identified in the elements that mediate vitamin D responsiveness in the calbindin D_{9k} and calbindin D_{28k} genes. Moreover, several synthetic elements with large spacer regions and inverted arrangements can mediate vitamin D responsiveness under certain conditions. It is likely that the affinity of VDR for these atypical elements may vary from that of the classic DR-3 motif adding yet another level of regulatory complexity to the process of VDR-mediated gene expression.

Other variations of the VDRE can lead to the specific downregulation of gene expression. Both the rat PTH promoter

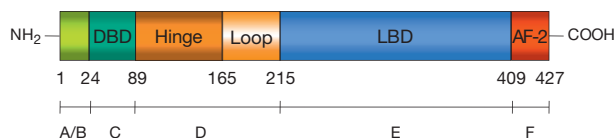


Figure 2 Functional domains of the VDR. A/B, amino terminal region; DBD, DNA-binding domain containing two zinc-finger motifs; LBD, ligand-binding domain; AF-2, activation function-2 domain encompassing helix 12.

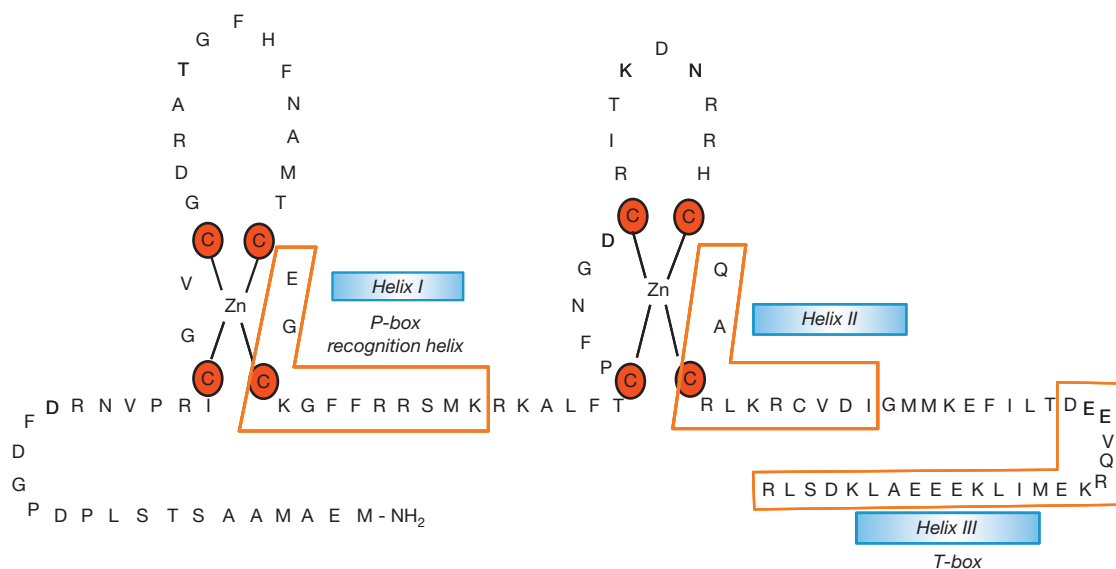


Figure 3 Schematic of the DNA-binding domain of the VDR. The cysteines responsible for coordinating the zinc atoms are shown in red.

and the human PTHrP promoter contain negative VDREs (nVDREs) which closely resemble the consensus VDRE sequence. These nVDREs are bound by either VDR homodimers or by VDR/RXR heterodimers and subsequently down-regulated. A second type of nVDRE, composed of E-box type motifs (5'-CATCTG-3'), was identified in the promoters of the human CYP27B1 gene, the human PTH gene, and the human PTHrP gene. These nVDREs are bound by a helix-loop-helix transcription factor known as VDR-interacting repressor and are transcriptionally active in the absence of $1,25(\text{OH})_2\text{D}_3$. When $1,25(\text{OH})_2\text{D}_3$ is absent, VDR recruits CBP/p300 HAT activity to promote transcription. However, in the presence of $1,25(\text{OH})_2\text{D}_3$, liganded VDR/RXR heterodimers associate with the nVDRE-bound VDR-interacting repressor, resulting in the dissociation of the CBP/p300 HAT coactivator and the recruitment of histone deacetylases and NCoR/SMRT corepressors. Thus, the chromatin is subsequently remodeled and transcription inhibited.

Communication between VDR and the Transcriptional Machinery

The regulation of VDR-mediated transcription involves a complex series of macromolecular interactions occurring in a temporally coordinated fashion (Figure 1). Other transcriptional components that associate with the liganded VDR/RXR heterodimer may be classified into two general categories: general transcription factors and the comodulatory proteins. The interaction of VDR with the first group results in direct contacts with the preinitiation complex (PIC), which may facilitate assembly or recruitment of the PIC and thereby stimulate transcription by RNA Pol II. Transactivator interaction with the PIC may occur via adapter proteins such as the TATA-binding protein-associated factors or via direct interaction with other core transcription factors. In this regard, TFIIB may be a key target because it is known to interact directly with VDR and augment vitamin D-activated transcription in transient gene-expression studies. Indeed, differences in the activities of NH₂-terminal VDR isoforms are attributed to differential interactions with TFIIB, thus, supporting an important role for the VDR-TFIIB interaction in determining the overall transactivation potential of the liganded VDR.

In addition to direct contacts with the general transcription machinery, the liganded VDR is also linked to the transcriptional PIC by the NR comodulatory factors. The general functional properties of the NR comodulators are their ability to interact with NRs and control their transcriptional responsiveness to ligand. NR comodulators are classified either as coactivators or corepressors, and they aid in the induction or repression, respectively, of ligand/receptor-mediated transcription. NR coactivators may function as macromolecular bridges between the receptor and the transcriptional machinery that aid in the assembly or promote the stability of the PIC. Moreover, several coactivator proteins either possess intrinsic histone acetyltransferase (HAT) activities or recruit other proteins that possess HAT activity (e.g., CBP/p300). Acetylation of histones near the promoter presumably results in a loosening of chromatin structure and greater accessibility of the promoter for the transcriptional machinery (Figure 1). The best-characterized coactivators are the steroid receptor

coactivator (SRC) family of NR coactivators that includes three members at present: SRC-1 (NCoA-1), SRC-2 (GRIP-1, TIF2, NCoA-2), and SRC-3 (pCIP, RAC3, ACTR, AIB-1, TRAM-1).

Structural studies of related receptors provide insight into the mechanism of ligand-induced interaction of VDR with SRC coactivators. It is hypothesized that the $1,25(\text{OH})_2\text{D}_3$ ligand promotes coactivator interaction by inducing a repositioning of the AF-2 activation helix, or helix H12 (Figure 4). In the unliganded state, the AF-2 domain (helix H12) projects out away from the globular core of the LBD, while in the liganded state the AF-2 domain is folded over onto the LBD globular core domain. One outcome of helix H12 folding is the creation of a platform or protein interaction surface through which coactivator proteins such as SRCs effectively dock with the VDR. The domain in the coactivator proteins, which is responsible for this interaction is termed the NR or NR box and is an amphipathic α -helix composed of the consensus core sequence LXXLL. This interaction is essential for transactivation mediated by the VDR.

A large multiprotein complex called Mediator D is another coactivator required for VDR-mediated transcription (Figure 1). Mediator D is also known as VDR-interacting protein (DRIP), thyroid receptor-activating protein complex, and the mammalian mediator complex. The Mediator D complex is composed of at least 10 different proteins anchored by Med220, which interacts directly with ligand-activated VDR/RXR heterodimers through one of two LXXLL motifs. DRIP is essential for *in vitro* VDR-mediated transcription from chromatinized templates. However, as the DRIP complex does not contain any SRCs and does not possess HAT activity, it appears to potentiate NR-mediated transcription through distinct mechanisms. In particular, DRIP directly recruits the RNA polymerase II holoenzyme to $1,25(\text{OH})_2\text{D}_3$ -activated VDR, indicating that DRIP may serve as a bridge between VDR and the core transcriptional machinery. Indeed, chromatin immuno-precipitation studies show that DRIPs enter transcriptional complexes at a later time than SRC coactivators, supporting the nonredundant roles of these coactivator classes in NR-mediated transactivation.

Numerous other classes of comodulators have been described that interact through distinct mechanisms and impinge on VDR-mediated transcriptional regulation at proximal and distal steps in the process. One example is NCoA62/SKIP, a unique VDR coactivator protein that lacks LXXLL motifs and that selectively interacts with the VDR-RXR heterodimer in an AF2-independent manner. NCoA62/SKIP enters the transcriptional regulatory complex following SRCs and the mediator complex. Indeed, its close association with the RNA spliceosome complex and functional studies indicate a role for

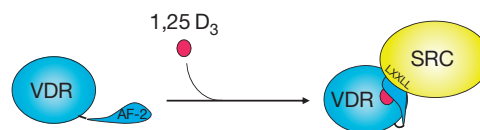


Figure 4 The vitamin D receptor undergoes a conformational change upon binding hormone. The VDR binds ligand and this induces a conformational change in the AF-2 domain to trap the ligand in the binding pocket. This change also creates a hydrophobic cleft or surface on VDR that LXXLL motifs in coactivator proteins use for docking to the VDR.

NCoA62/SKIP in the splicing of nascent RNA transcripts that result from the process of VDR-activated transcription. Thus, NCoA62/SKIP represents a class of coactivators that may function at distal steps in the mechanism of NR-mediated transcription by coupling two global processes in the nucleus, transcription and RNA processing. The combined actions of the variety of distinct classes of comodulatory proteins ensure efficient and rapid responses of the cell to vitamin D and VDR-mediated transcription regulation.

Summary

Research spanning the past 25 years reveals that $1,25(\text{OH})_2\text{D}_3$ -mediated transcription is more complex than the simple binding of the receptor to DNA and the recruitment of RNA Pol II to initiate transcription. VDR/RXR-activated transcription involves complex interactions that may occur in a spatially distinct and temporally coordinated fashion to increase the rate at which $1,25(\text{OH})_2\text{D}_3$ -responsive genes are transcribed and at which the resulting RNA transcript is processed. A model for VDR-mediated transcription is proposed which incorporates numerous properties of VDR that were discussed in this section (Figure 1). The initial event in this model is high-affinity binding of the $1,25(\text{OH})_2\text{D}_3$ ligand to the VDR. Ligand binding induces VDR/RXR heterodimerization and the heterodimer specifically binds VDREs in the promoter regions of vitamin D-responsive genes. The heterodimer then recruits coactivator molecules that acetylate core histones to make the DNA more accessible to the transcriptional machinery. Meanwhile, other coactivators and the PIC are recruited to the site and activated transcription proceeds. Understanding the complex interplay that occurs between these various factors is crucial to unraveling the complexities of activated or repressed transcription mediated by vitamin D and the VDR.

See also: **Bioenergetics:** Calcium-Buffering Proteins; ER Luminal Proteins; **Metabolism Vitamins and Hormones:** Vitamin D; **Protein/Enzyme Structure Function and Degradation:** Zinc Fingers; **Signaling:** Steroid/Thyroid Hormone Receptors.

Further Reading

- Auboeuf D, Batsché E, Dutertre M, Muchardt C, and O'Malley BW (2007) Coregulators: Transducing signal from transcription to alternative splicing. *Trends in Endocrinology and Metabolism* 18: 122–129.
- Bouillon R, Carmeliet G, Verlinden L, et al. (2008) Vitamin D and human health: Lessons from vitamin D receptor null mice. *Endocrine Reviews* 29: 726–776.
- Carlberg C and Seuter S (2009) A genomic perspective on vitamin D signaling. *Anticancer Research* 29: 3485–3493.
- DeLuca HF (2008) Evolution of our understanding of vitamin D. *Nutrition Reviews* 66: S73–S87.
- Feldman D, Glorieux FH, and Pike JW (eds.) (2005) *Vitamin D*, 2nd edn. San Diego, CA: Academic Press.
- Hausssler MR, Hausssler CA, Bartik L, et al. (2008) Vitamin D receptor: Molecular signaling and actions of nutritional ligands in disease prevention. *Nutrition Reviews* 66: S98–S112.
- Jurutka PW, Whitfield GK, Hsieh JC, et al. (2001) Molecular nature of the vitamin D receptor and its role in regulation of gene expression. *Reviews in Endocrine and Metabolic Disorders* 2: 203–216.
- Lonard DM, Lanz RB, and O'Malley BW (2007) Nuclear receptor coregulators and human disease. *Endocrine Reviews* 28: 575–587.
- Lonard DM and O'Malley BW (2006) The expanding cosmos of nuclear receptor coactivators. *Cell* 125: 411–414.
- Norman AW (2006) Minireview: Vitamin D receptor: New assignments for an already busy receptor. *Endocrinology* 147: 5542–5548.
- Pike JW, Zella LA, Meyer MB, Fretz JA, and Kim S (2007) Molecular actions of $1,25$ -dihydroxyvitamin D₃ on genes involved in calcium homeostasis. *Journal of Bone and Mineral Research* 22(supplement 2): V16–V19.
- Prosser DE and Jones G (2004) Enzymes involved in the activation and inactivation of vitamin D. *Trends in Biochemical Sciences* 29: 664–673.
- Sutton AL and MacDonald PN (2003) Vitamin D: More than a 'bone-a-fide' hormone. *Molecular Endocrinology* 17: 777–791.

Vitamin E

J Atkinson, Brock University, St. Catharines, ON, Canada
D Manor, Western Reserve University, Cleveland, OH, USA
R Parker, Cornell University, Ithaca, NY, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

RRR- α -tocopherol The only stereoisomer of α -tocopherol occurring in nature. When α -tocopherol is made by chemical synthesis, it contains not only the RRR-stereoisomer, but also a racemic mixture (called all-*rac*- α -tocopherol) of eight stereoisomers. Synthetic α -tocopherol is mainly used in dietary supplements and in fortified foods.

Tocopherol transfer protein (TTP) A protein expressed primarily in liver and to a lesser degree in the brain and placenta. TTP binds RRR- α -tocopherol with high affinity and

selectivity, and is responsible for the enrichment of plasma with this form of the vitamin.

Vitamin E The term denoting a group of eight neutral lipids of plant origin, including four tocopherols and tocotrienols. α -Tocopherol is the most prominent form of vitamin E, which is considered not only essential in humans and animals, but also the major functional lipid-soluble antioxidant in humans, protecting against oxidative damage to lipids in circulation and in cell membranes.

Vitamin E: Definition and Structures

Vitamin E refers to a group of structurally related plant-derived natural products comprised of tocopherols and tocotrienols. The four tocopherols and four tocotrienols (together called tocols) differ from each other by the number and placement of methyl groups on a parent 6-hydroxychroman (**Figure 1**). The naturally occurring tocopherols have 2*R*,4'*R*,8'*R* (RRR) stereochemistry, while the tocotrienols have only the one center of *R*-stereochemistry at C2. By contrast, synthetic α -tocopherol consists of a racemic mixture of eight stereoisomers (four enantiomeric pairs: RRR-SSS, RSR-SRS, RRS-SSR, and RSS-SRR) arising from the three chiral centers at positions C2, C4', and C8'' and is designated as all-*rac*- α -tocopherol (or *dl*- α -tocopherol). The tocotrienols have three sites of unsaturation on the side chain at carbons 3', 7', and 11'.

Biological Activity

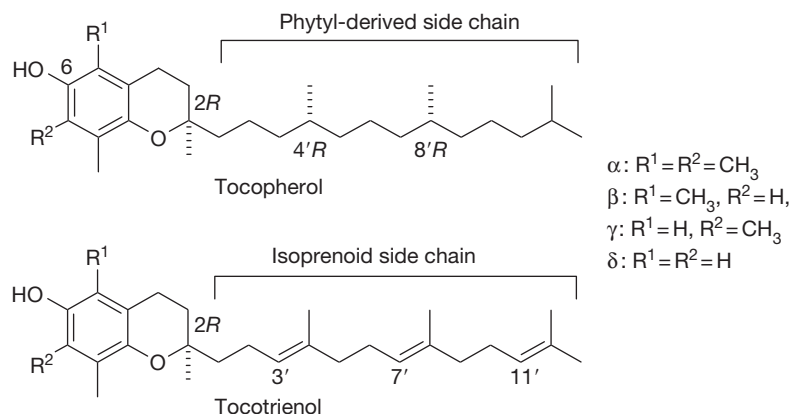
In the 1950s, it was recognized by A.L. Tappel's group that vitamin E is the body's major lipid-soluble antioxidant. The physicochemical basis of this activity was later described in detail by K.U. Ingold and G. Burton, E. Niki, and K. Fukuzawa. It is now accepted that vitamin E inhibits the free radical-chain peroxidation of polyunsaturated lipids in membranes and lipoproteins. The lipids that are most susceptible to peroxidation are those containing *bis*-allylic methylenes, CH₂-groups between two *cis*-double bonds in fatty acid structures (**Figure 2**). The greater the number of these sites, the more prone the lipid is to peroxidation. Thus, the polyunsaturated linolenic, arachidonic, and docosahexaenoic acids are examples of lipids sensitive to peroxidation. All natural forms and synthetic stereoisomers of vitamin E exhibit to varying degrees the ability to inhibit lipid peroxidation as chain-breaking antioxidants. Antioxidant activities of tocols are greater in solution than in lipid membranes, but the rank order of antioxidant activities remains $\alpha > \beta \approx \gamma > \delta$ for both substrate presentations.

It is worth noting that all forms of tocopherol are excellent antioxidants *in vitro*, but that α -tocopherol is superior. α -Tocopherol is approximately twice as efficient a radical quencher as γ - and β -tocopherol, and about 5 times better than δ -tocopherol in chemical measures of antioxidant activity. These findings likely have little relevance to *in vivo* biological activity, as selective distribution and metabolism of dietary tocols heavily favor the retention of α -tocopherol in tissues.

The nature of the initiating radical R• in **Figure 2** is often considered to be reactive oxygen and nitrogen species, but the rates of H-atom abstraction from susceptible lipids by many of these species are very slow. *In vivo*, the most relevant species is most likely the protonated form of superoxide (the perhydroxyl radical HOO•) whose acid dissociation constant (pK_a) is ~ 4.9 and thus is still present ($\sim 0.5\%$) at physiological pH. The fact that HOO• is charge neutral also facilitates its access to the hydrophobic portion of membranes. The role of spontaneously formed lipid hydroperoxides is known to be critical in chemical systems and may also play a role *in vivo*.

The peroxidation of membrane lipids is nonreversible. Lipid hydroperoxides and their breakdown products such as 4-hydroxynonenal and malondialdehyde are known to react with and modify the activity of cell proteins. The selective loss of polyunsaturated fatty acids in phospholipids can also change membrane structure and lower membrane fluidity – critical elements of membrane function. Each molecule that suffers oxidation will eventually be removed from membranes by the action of peroxide reductases, lipases, and esterases.

Most of the research explaining the antioxidant activity of tocols has been performed *in vitro*, often in organic solvents, or in lipid dispersions that may not be perfect models for the biological membranes where vitamin E is found *in vivo*. Biological membranes are now known to be heterogeneous, with portions enriched in certain lipid types, including the so-called lipid rafts rich in saturated lipids, sphingomyelin, and cholesterol, and other portions rich in polyunsaturated lipids where tocopherol has been observed to concentrate. This heterogeneity is difficult to reproduce *in vitro* and



The four naturally occurring forms of tocopherols and tocotrienols

Figure 1 The molecular structures of vitamin E. Tocopherols and tocotrienols exist in four different forms (α -, β -, γ -, and δ -forms) that differ by the extent and position of methyl groups on the chromanol ring. Tocopherol is shown in its naturally occurring *RRR*-configuration.

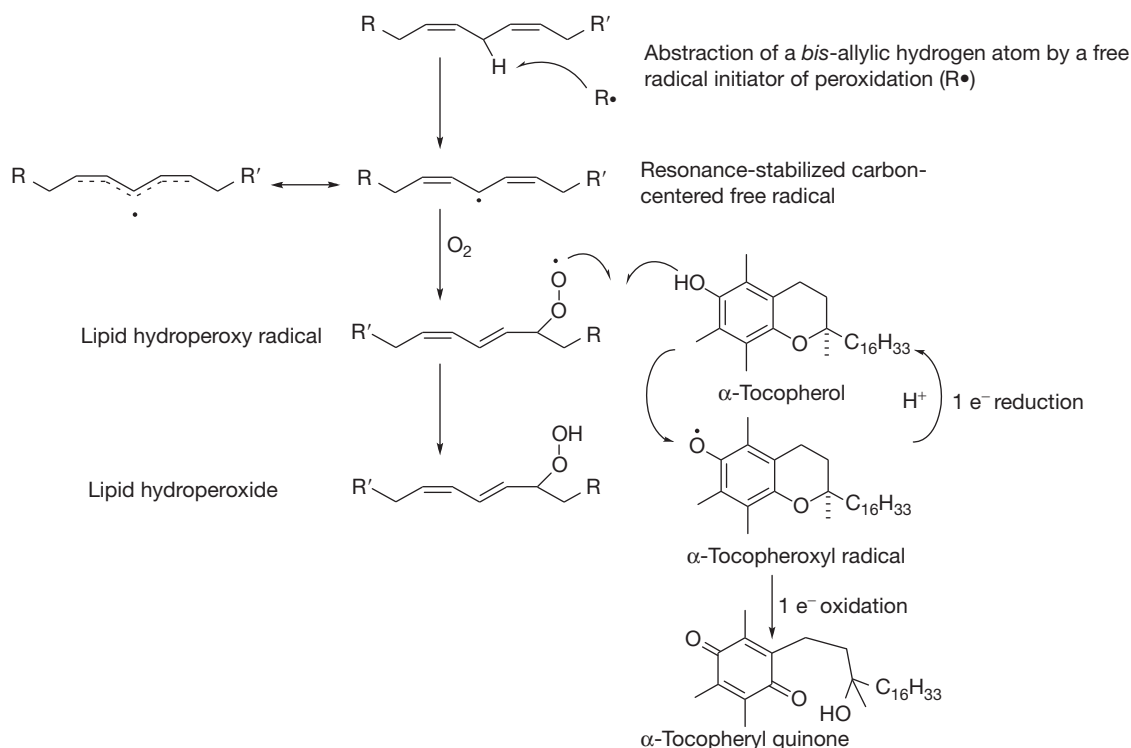


Figure 2 Simplified scheme illustrating the initiation of lipid peroxidation at *bis*-allylic centers on polyunsaturated lipids by a free radical $R^\bullet$, the formation of lipid peroxy radicals (that can abstract an H-atom and continue the chain reaction), and the radical chain termination of a lipid peroxy radical by tocopherol. The α -tocopheroxyl radical can be recycled to α -tocopherol by a one-electron reduction and the reducing agent *in vivo* is thought to be ascorbate. Further one-electron oxidation of the tocopheroxyl radical produces tocopheryl quinone.

complicates kinetic models. Because of technical issues in following short-lived radical species, chemical evidence for the action of vitamin E as a radical chain terminating antioxidant *in vivo* is indirect, relying on the analysis of lipid oxidation products such as 4-hydroxynonenal or malondialdehyde. Earlier products of polyunsaturated lipid peroxidation such as the F_2 -isoprostanes (rather than breakdown products such as

malondialdehyde) are now considered more reliable measures of oxidative stress.

Other biological activities have been reported for tocopherols (mainly α - and γ -tocopherols) and tocotrienols (α - and δ -forms) that may or may not be dependent on their action as antioxidants, including the regulation of cell cycle, apoptosis, and gene expression.

Vitamin E in Health and Disease

The first biological activity ascribed to vitamin E is that of a fertility factor that restored reproductive function to diet-restricted female rodents (hence the name tocopherol; from the Greek 'tocos' (offspring) and 'pherein' (to bring forth)). This protective action was also used to establish that RRR- α -tocopherol is the most biologically active member of the vitamin E family. In the decades since, the critical roles that vitamin E plays in many other biological processes have been established. In accordance with the function of vitamin E as an important antioxidant, the requirement for tocopherol is primarily evident from pathologies that involve a significant redox imbalance. It is important to note that since dietary deficiency of vitamin E is extremely rare in humans, our appreciation of its role in health relies primarily on studies in animal models.

Central Nervous System

The clearest manifestation of vitamin E deficiency is in disturbances of the nervous system. Human carriers of mutations in the tocopherol transfer protein (TTP, see below) present ataxia with vitamin E deficiency (AVED) – a debilitating disorder characterized by progressive ataxia (loss of movement coordination) and low serum vitamin E levels. The neurological damage appears to originate from loss of Purkinjee cells in the cerebellum. If treated at an early stage, AVED symptoms can be normalized by supplementation with high doses of vitamin E. Genetic animal models of vitamin E deficiency (i.e., TTP^{-/-} mice) display neurological symptoms similar to those of human AVED patients that can be prevented by timely supplementation with α -tocopherol. Adequate vitamin E status is also thought to be important for the prevention and treatment of other neurological disorders accompanied by an oxidative stress component, such as Parkinson's disease and Alzheimer's disease, and supplementation with α -tocopherol is often recommended to patients affected with these disorders.

Coronary Heart Disease

Atherosclerosis is a critical component in the etiology of coronary heart disease (CHD), the leading cause of death in the US. The lipid oxidation theory holds that the formation of atherosclerotic plaques is triggered and precipitated by oxidation of circulating lipids. This rationale popularized supplementation of at-risk patients with vitamin E in the 1980s and 1990s as a preventive measure against CHD. Since clinical trials failed to demonstrate a significant benefit for supplementation with vitamin E, α -tocopherol is no longer considered an effective treatment for CHD patients.

Cancer

Since elevation in markers of oxidative stress is seen in many malignant diseases, vitamin E sufficiency is thought to have an important chemopreventive value. In support of this notion, supplementation with vitamin E was shown to bring about profound clinical improvement in genetic animal models of

cancer (e.g., hepatocellular carcinoma and prostate cancer). However, clinical trials to date have not provided consistent support for this claim.

Fertility

Despite the original description of vitamin E as a fertility factor in rodents, the relevance of α -tocopherol to human reproduction has not been determined.

A Vitamin E Paradox?

One of the most intriguing features of vitamin E is its cyclic emergence and disappearance from research texts and pharmacopeias as an effective agent in the treatment of human disease. These oscillations in the biomedical community likely reflect conflicting published reports regarding the vitamin's health-promoting value in different experimental settings. On the one hand, laboratory studies repeatedly demonstrated that tocopherol elicits clear protective actions in isolated cells, tissues, and animal models. On the other hand, controlled intervention trials assessing clinical outcomes in humans provided inconsistent results at best. These seemingly contradicting observations gave rise to general confusion regarding the actual therapeutic value of vitamin E adequacy and supplementation. A clue to this apparent paradox may be found in observations indicating that vitamin E status is a stable individual phenotype under genetic control. Thus, it is possible and even likely that individual differences (such as single-nucleotide polymorphisms) in the genes that regulate vitamin E status may determine the individual biochemical and clinical responsiveness to vitamin E. Future investigations into the interrelations between individual genotype, vitamin E status, and disease risk may provide critical tools that may facilitate reevaluating clinical trial data and obtaining important insight into this apparent paradox.

Dietary Sources and Recommended Intake of Vitamin E

Vitamin E is ubiquitous in foods of both plant and animal origin. Food sources with the highest vitamin E concentrations are the seed (vegetable) oils, in which tocopherols and occasionally tocotrienols are extensively deposited presumably to prevent autooxidation of polyunsaturated lipids. However, vegetable oils differ markedly in the type of vitamin E deposited, with some containing mostly α -tocopherol (such as sunflower with ~95% α -tocopherol and 5% β -tocopherol) and others containing mostly other forms (such as soybean with ~13:3:62:22% of α : β : γ : δ -tocopherol, respectively). Interestingly, the oilseed palm and annatto grown in tropical regions are unique in that a large proportion of vitamin E exists as tocotrienols. Fats from animal origin such as butter and lard contain predominantly α -tocopherol due to the selective accumulation of that form of vitamin E. Thus, the amount and types of oils consumed by an individual determine their intake of different vitamin E forms. While fruits and vegetables contain much lower concentrations of vitamin E than lipid-rich products, some, such as tomatoes and tomato products, are major dietary sources due to the prevalence of their

consumption. The 2000 edition of the Dietary Reference Intakes (IOM) summarized data suggesting that most adults in the US and Canada probably consume the Dietary Reference Intakes (DRI) for vitamin E, that is, the equivalent of 15 mg RRR- α -tocopherol per day, but this value will vary widely with fat intake and persons on fat-restricted diets likely consume significantly less than the DRI.

The latest US governmental recommendations regarding vitamin E were published in 2000 by the Food and Nutrition Board convened by the Institute of Medicine of the National Institutes of Health. As in the past, the panel relied almost exclusively upon the *ex vivo* data of Horwitt and co-workers from the early 1960s involving hydrogen peroxide-induced hemolysis of red blood cells obtained from vitamin E-deficient men before and after repletion with varying amounts of α -tocopherol. The 2000 panel re-interpreted the data to conclude that 12 mg of RRR- α -tocopherol per day was required to yield a plasma concentration ($12 \mu\text{mol l}^{-1}$) sufficient to prevent *in vitro* red blood cell (RBC) hemolysis under the test conditions employed, and therefore the estimated average requirement (EAR) was considered as 12 mg RRR- α -tocopherol per day. This value was increased by the customary two standard deviations (assumed to be 10%) to yield the current recommended dietary allowance (RDA) of 15 mg d^{-1} for adults. No differences are associated with gender or pregnancy, but the RDA for lactating women was set at 19 mg d^{-1} due to losses of vitamin E in breast milk. These RDA values generally represent an increase of 3 mg d^{-1} over the previous set of recommendations. A second deviation from the previous recommendations was that forms of vitamin E other than α -tocopherol (e.g., γ -tocopherol) were to be disregarded in terms of meeting the recommended dietary allowance; previous editions considered γ -tocopherol as one-tenth as bioactive as α -tocopherol.

The relationship between α -tocopherol intake and plasma α -tocopherol concentration is linear within the range of intakes possible from average diet ($0\text{--}17 \text{ mg d}^{-1}$). At higher intakes, using supplements, this relationship becomes severely nonlinear such that a doubling of plasma concentration from 25 to $50 \mu\text{mol l}^{-1}$ requires increasing intake by at least 10-fold, and concentrations much above $60 \mu\text{mol l}^{-1}$ are not possible even with doses near 1 g d^{-1} . The reasons for this nonlinearity are not known, but are likely to involve multiple phenomena.

The natural (RRR) and synthetic (racemic mixture) forms of α -tocopherol exhibit different bioavailabilities in animals and humans. The synthetic form is 70% as active in the rat gestation-resorption assay. A similar difference is seen in humans comparing their pharmacokinetics in plasma following single doses administered separately, that is, under non-competitive conditions. The difference is attributed to the higher affinity of RRR- α -tocopherol to the liver tocopherol transfer protein, such that this form is more efficiently secreted by the liver into the bloodstream.

The current upper limit (UL) for vitamin E intake was set by the US Institute of Medicine panel at $1000 \text{ mg } \alpha$ -tocopherol per day for adults, a sum of natural and synthetic forms. Establishment of this value was based on several factors, including evidence of hemorrhagic effects seen in some clinical trials involving vitamin E supplementation. However, while the human observations were seen as evidence for concern, the data quality was insufficiently quantitative for derivation of

a UL. Therefore, the UL was based on studies of hemorrhagic effects in rats, from which a lowest observed adverse effect level (LOAEF) of $500 \text{ mg per kg body weight per day}$ was determined. An uncertainty factor, consisting of several subfactors including extrapolation from animal to human, was estimated at 36, and the ULL of $14 \text{ mg kg}^{-1} \text{ d}^{-1}$ (1000 mg d^{-1} for adults) was obtained by dividing the LOAEF by the uncertainty factor. The mechanism by which high doses of vitamin E interferes with blood clotting remains unknown.

Transport of Vitamin E (DM)

Absorption

Like most dietary lipids, tocopherols are micellized in the intestinal lumen prior to uptake by intestinal epithelial cells. In the enterocyte, the microsomal triglyceride transfer protein (MTTP) facilitates the incorporation of vitamin E into apoB48-containing chylomicra, which are secreted into the lymph. Recent studies indicate the presence of an additional secretion route, involving active transport through an ABC-family transporter to high-density lipoprotein (HDL) particles. In these early absorptive stages (within $\sim 2 \text{ h}$ after ingestion), all members of the vitamin E family are handled with similar efficiency, such that relative composition of the different vitamin E species in chylomicra mirrors the dietary composition.

Hepatic Processing

The liver serves as the primary site for regulation of body-wide levels and content of vitamin E. Following uptake of chylomicra remnants by parenchymal cells in the liver, the hepatic TTP targets vitamin E to a specific secretory pathway out of the hepatocyte. Secretion of vitamin E from hepatocytes occurs through a unique pathway, independent from the route for assembly and secretion of nascent very low density lipoproteins (VLDLs). Furthermore, secretion from the hepatocytes takes place through an adenosine triphosphate (ATP)-driven system, including the ABCA1 transporter that is defective in Tangier disease patients. Following secretion, the vitamin associates with VLDL particles, which deliver it to extrahepatic tissues. The structure of TTP's ligand-binding pocket dictates a high degree of selectivity, such that the protein binds α -tocopherol at a significantly higher affinity compared to other members of the vitamin E family. TTP's ligand preference, together with the selective degradation of non- α vitamers by CYP4F2 (see below), comprise a remarkably efficient discriminatory system that results in the selective enrichment of human tissues with α -tocopherol.

Circulation and Delivery

In circulation, α -tocopherol is exclusively associated with lipoprotein particles. The plasma phospholipid transfer protein (PLTP) plays an important role in the exchange of vitamin E between HDL and LDL, such that in steady state the vitamin is distributed between these two particle populations. Uptake of α -tocopherol into extrahepatic cells is mediated by a number of parallel, possibly overlapping routes, including receptor-mediated endocytosis through the LDL receptor and uptake

from HDL through the scavenger receptor SR-BI. In addition, endothelial lipoprotein lipase mediates direct exchange of α -tocopherol between lipoprotein particles and the surface of target cells.

The profound hydrophobicity of α -tocopherol (calculated $\log P \sim 9$) raises the attractive hypothesis that specific transport proteins are needed for adequate intracellular distribution of the vitamin in peripheral tissues. An α -tocopherol-binding activity was purified from heart tissue of several species, but the protein has not been identified or cloned. Thus, to date, the presence of an intracellular tocopherol binding in extrahepatic tissues is yet to be determined. The exception is the tocopherol transfer protein, which in addition to liver, is expressed in the brain and the uterus. The modes of action and physiological consequences of TTP's presence in these tissues await future examinations. It is tempting to speculate, however, that brain-TTP is responsible for the efficient enrichment of the central nervous system with tocopherol, and for the remarkable resistance of this tissue to dietary depletion.

Metabolism of Vitamin E

At the present time, there is only one known pathway of metabolism of vitamin E. This pathway, termed the tocopherol- ω -hydroxylase pathway, is initiated by an oxidation (hydroxylation) of one of the two terminal methyl groups of the hydrophobic side chain of tocopherols and tocotrienols (Figure 3). This reaction is catalyzed in the human by

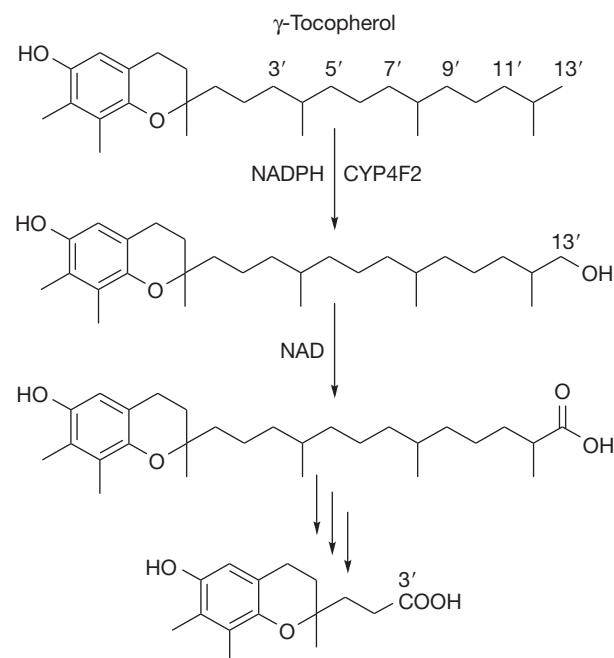


Figure 3 The ω -hydroxylase pathway of vitamin E metabolism, illustrated with γ -tocopherol as substrate. The first step is catalyzed by cytochrome P450 4F2 (CYP4F2), requiring NADPH as electron donor. The final product is the 3'-carboxychromanol, which is excreted in the urine as its glucuronide conjugate. The same pathway is responsible for metabolism of both tocopherols and tocotrienols.

cytochrome P450 4F2, and takes place predominantly in the endoplasmic reticulum of the liver. To date, no other enzymes exhibiting vitamin E ω -hydroxylase activity have been identified. The hydroxylated vitamin E then undergoes a second oxidation to the corresponding ω -carboxylated form by an as-yet unidentified microsomal dehydrogenase. The oxidized side chain then undergoes a series of two- or three-carbon truncations analogous to the oxidation of branched-chain fatty acids, ultimately yielding the 3'- and 5'-carboxychromanol metabolites that are excreted in the urine. Despite the considerable water solubility of these metabolites, they are excreted largely, if not entirely, as glucuronide conjugates. Both conjugated and nonconjugated 3'- and 5'-carboxychromanol metabolites circulate in the bloodstream, suggesting that additional conjugation may take place in the kidney. The site of glucuronidation appears to be the phenol hydroxyl group, meaning that such metabolites do not have antioxidant activity in their conjugated state.

The ω -hydroxylation pathway is of significance due to the remarkable substrate preference for vitamin E forms other than α -tocopherol. α -Tocopherol is metabolized 3–15-fold slower than other tocopherols. Natural and synthetic α -tocopherol are metabolized at similarly low rates. Additionally, tocotrienols are metabolized more rapidly (2–20-fold) than the corresponding tocopherols. Since absorption of the various forms of vitamin E is similar, it is likely that this pattern of substrate discrimination exhibited by CYP4F2 plays an important role in the highly selective tissue accumulation of α -tocopherol, observed in essentially all complex organisms. An additional contribution to this selectivity is made by the hepatic TTP, which selectively destines α -tocopherol for secretion from the liver into the bloodstream. Both TTP and ω -hydroxylase underlie the remarkable difference between the vitamin E composition of the diet, which can favor γ -tocopherol over α -tocopherol by 4:1, and that in human blood and tissues, in which the same ratio is typically about 1:15.

Tocopherol quinones are thought to be generated at slow rates during the antioxidant action of vitamin E. Although the glucuronide conjugate of the 3'-carboxychromanol metabolite of α -tocopherol quinone has been identified in urine, the extent to which its presence reflects artifactual oxidation during analysis versus actual side chain truncation of α -tocopherol quinone *in vivo* remains unclear.

Expression of the tocopherol/tocotrienol- ω -oxidation pathway appears widespread in the animal kingdom, including the fruit fly, *Drosophila melanogaster*. As *Drosophila* does not have a sequence ortholog of cytochrome P450 4F2, another P450 enzyme apparently exists in this species that promotes the selective accumulation of α -tocopherol. Interestingly, the *D. melanogaster* genome does not encode a tocopherol transfer protein ortholog, suggesting that the selective accumulation of α -tocopherol exhibited by this species arose prior to the evolution of TTP, whose function appears to be vitamin E retention. The selective pressure(s) underlying the evolution of the hydroxylase and TTP, and their persistence throughout the animal kingdom, remain a mystery. There are two relatively common nonsynonymous single-nucleotide polymorphisms in the human cytochrome P450 4F2 gene, but their effects on vitamin E status have not yet been determined.

See also: **Bioenergetics:** Cytochrome P-450; Friedreich's Ataxia; Mitochondria: Source and Target of Free Radicals; Reactive Oxygen and Nitrogen Species: Interactions with Mitochondria and Pathophysiology; **Lipids Carbohydrates Membranes and Membrane Proteins:** Lipid Bilayer Structure; Lipid Rafts; **Metabolism Vitamins and Hormones:** Vitamin C.

Further Reading

- Atkinson J, Epand RF, and Epand RM (2008) Tocopherols and tocotrienols in membranes: A critical review. *Free Radical Biology and Medicine* 44: 739.
- Burton GW and Ingold KU (1986) Vitamin E: Applications of the principles of physical organic chemistry to the exploration of its structure and function. *Accounts of Chemical Research* 19: 194.
- Dix TA and Aikens J (1993) Mechanisms and biological relevance of lipid peroxidation initiation. *Chemical Research in Toxicology* 6: 2–18.
- Food and Nutrition Board of the Institute of Medicine (2000) *Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium and Carotenoids*, pp. 186–283. Washington, DC: National Academy Press.
- Fukuzawa K (2008) Dynamics of lipid peroxidation and antioxidation of alpha-tocopherol in membranes. *Journal of Nutritional Science and Vitaminology (Tokyo)* 54: 273.
- Horwitt M (1960) Vitamin E and lipid metabolism in man. *American Journal of Clinical Nutrition* 8: 451–461.
- Horwitt M, Century B, and Zeman A (1963) Erythrocyte survival time and reticulocyte levels after tocopherol depletion in man. *American Journal of Clinical Nutrition* 12: 99–106.
- Manor D and Morley S (2007) The alpha-tocopherol transfer protein. *Vitamins and Hormones* 76: 45–65 (review).
- Ouahchi K, Arita M, Kayden H, et al. (1995) Ataxia with isolated vitamin E deficiency is caused by mutations in the alpha-tocopherol transfer protein. *Nature Genetics* 9(2): 141–145.
- Parker RS (2002) Role of cytochrome P450-mediated metabolism in the bioavailability of vitamin E. *Free Radical Research* 36(supplement 1): 5–7.
- Roxborough HE, Burton GW, and Kelly FJ (2000) Inter- and intra-individual variation in plasma and red blood cell vitamin E after supplementation. *Free Radical Research* 33(4): 437–445.
- Sontag TJ and Parker RS (2007) Influence of major structural features of tocopherols and tocotrienols on their omega-oxidation by tocopherol-omega-hydroxylase. *Journal of Lipid Research* 48(5): 1090–1098.
- Traber MG, Sokol RJ, Kohlschütter A, et al. (1993) Impaired discrimination between stereoisomers of alpha-tocopherol in patients with familial isolated vitamin E deficiency. *Journal of Lipid Research* 34(2): 201–210.

Vitamin K: Biochemistry, Metabolism, and Nutritional Aspects

J W Suttie, University of Wisconsin–Madison, Wisconsin, WI, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

γ -Carboxyglutamic acid An amino acid found in a limited number of proteins, formed by a posttranslational modification of Glu residues.

γ -Glutamyl carboxylase The enzyme involved in the formation of Gla residues; also called vitamin K-dependent carboxylase.

Vitamin K-Dependent Proteins

Prothrombin (clotting factor II), the zymogen of the plasma procoagulant thrombin, was the first protein shown to be dependent on vitamin K for its synthesis, and was also the first protein demonstrated to contain γ -carboxyglutamyl (Gla) residues. Plasma clotting factors VII, IX, and X were all initially identified in patients with hereditary bleeding disorders and were subsequently shown to be vitamin K dependent. Until the mid-1970s, these four vitamin K-dependent clotting factors were the only proteins known to require this vitamin for their synthesis. The amino-terminal, Gla domains of these four vitamin K-dependent procoagulants are very homologous, and the 10–13 Gla residues in each are in essentially the same position as in prothrombin.

Following the discovery of Gla, three more Gla-containing plasma proteins with similar homology were discovered. Protein C and protein S are involved in an anticoagulant rather than a procoagulant role in normal hemostasis, and the seventh Gla-containing plasma protein (protein Z) also has an anticoagulant function under some conditions. As these proteins play a crucial role in hemostasis, they have been extensively studied, and the complementary DNA (cDNA) and genomic organization of each of them are well documented.

The discovery of Gla residues in the plasma vitamin K-dependent proteins led to a search for other proteins with this modification, and a 49-residue protein that contained three Gla residues, called osteocalcin (OC), was isolated from bone. It has little structural homology to the vitamin K-dependent plasma proteins. Although it is the second-most-abundant protein in bone, its function is not clearly defined. A second low-molecular-weight (79-residue) protein, matrix Gla protein (MGP), contains five Gla residues and was also first isolated from bone; it is also synthesized in cartilage and many other soft tissues. The function of these two proteins has not been clearly defined, but it has been shown that OC gene knockout mice develop more dense bones, and MGP knockout mice die from spontaneous calcification of arteries and cartilage. Two additional Gla-containing bone proteins, periostin and Gla-rich protein, have been more recently identified. A limited number of other mammalian proteins have been found to contain Gla residues: Gas 6, a ligand for the tyrosine kinase Ax1, and four integral membrane Gla proteins (PRGP-1, PRGP-2, TMG-3, and TMG-4). A large number of Gla-containing toxic venom peptides are secreted by marine *Conus* snails and are found in some snake venoms. The vitamin

K-dependent carboxylase has been cloned from a number of vertebrates, the *Conus* snail, a tunicate, and *Drosophila*, indicating the ancient evolutionary origin of this posttranslational modification.

Biochemical Role of Vitamin K

In the early 1960s, a biologically inactive form of prothrombin was demonstrated to be present in the plasma of patients treated with oral anticoagulants. Subsequent studies with hypoprothrombinemic rats were consistent with the presence of a hepatic precursor protein pool that was rapidly being synthesized and which could be converted to prothrombin by a posttranslational modification. Characterization of the abnormal prothrombin isolated from the plasma of cows fed with the anticoagulant dicoumarol led directly to an understanding of the metabolic role of vitamin K. This protein lacked the specific calcium-binding sites present in normal prothrombin. Acidic peptides obtained by proteolytic enzyme digestion of prothrombin were shown to contain Gla residues, but Gla residues could not be obtained by proteolysis of the abnormal prothrombin.

The Vitamin K-Dependent Carboxylase

In 1975, it was demonstrated that crude rat liver microsomal preparation contained an enzymatic activity (the vitamin K-dependent carboxylase) that promoted a vitamin K-dependent incorporation of $\text{H}^{14}\text{CO}_3^-$ into Gla residues of endogenous precursors of vitamin K-dependent proteins present in these preparations. Detergent-solubilized microsomal preparations retained this activity, and small peptides containing adjacent Glu–Glu sequences were found to be substrates for the enzyme. A general understanding of the properties of this unique enzyme was gained from studies utilizing this crude enzyme preparation, and these data have been adequately reviewed. The vitamin K-dependent carboxylation reaction does not require adenosine triphosphate, and the energy to drive this carboxylation reaction is derived from the oxidation of the reduced, hydronaphthoquinone form of vitamin K (vitamin KH_2) by O_2 to form vitamin K-2,3-epoxide (Figure 1). The lack of a requirement for biotin and studies of the $\text{CO}_2/\text{HCO}_3^-$ requirement indicate that carbon dioxide rather than HCO_3^- is the active species in the carboxylation reaction. Active substrates for the enzyme are 2-methyl-1,4-naphthoquinones

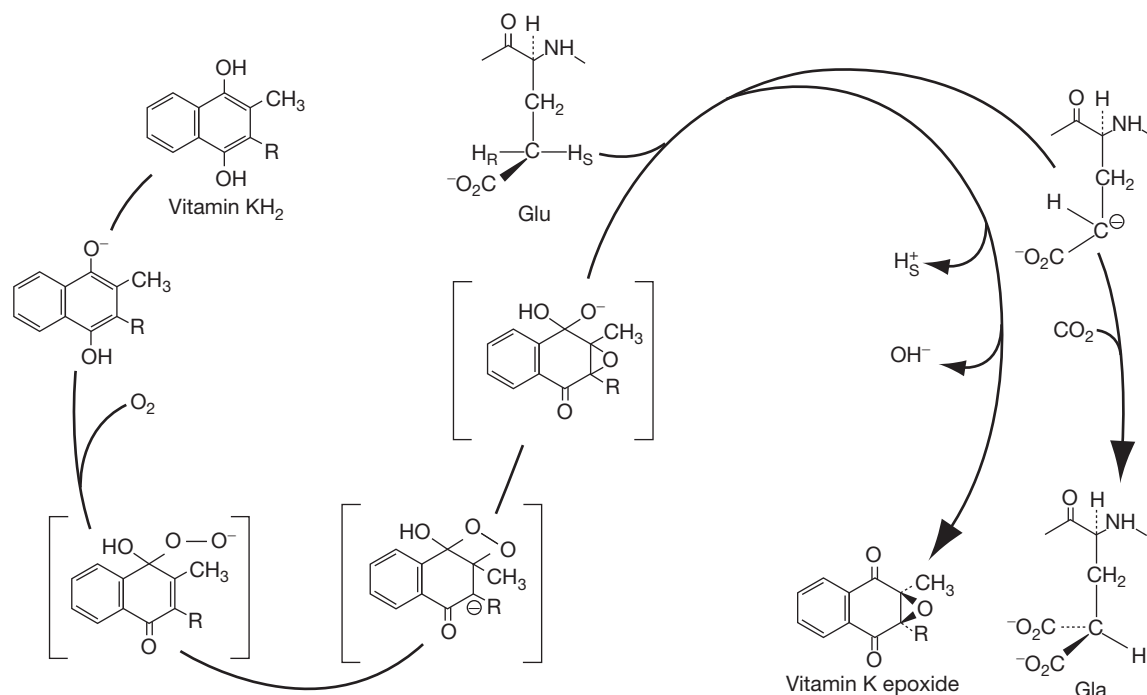


Figure 1 The vitamin K-dependent, γ -glutamyl carboxylase. The available data support an interaction of O₂ with vitamin KH₂, the reduced (hydronaphthoquinone) form of vitamin K, to form an oxygenated intermediate which is sufficiently basic to abstract the γ -hydrogen of the glutamyl residue. The products of this reaction are vitamin K-2,3-epoxide and a glutamyl carbanion. Attack of CO₂ on the carbanion leads to the formation of a Gla residue. The bracketed peroxy, dioxetane, and alkoxide intermediates have not been identified in the enzyme-catalyzed reaction but are postulated based on model organic reactions, and the available data are consistent with their presence.

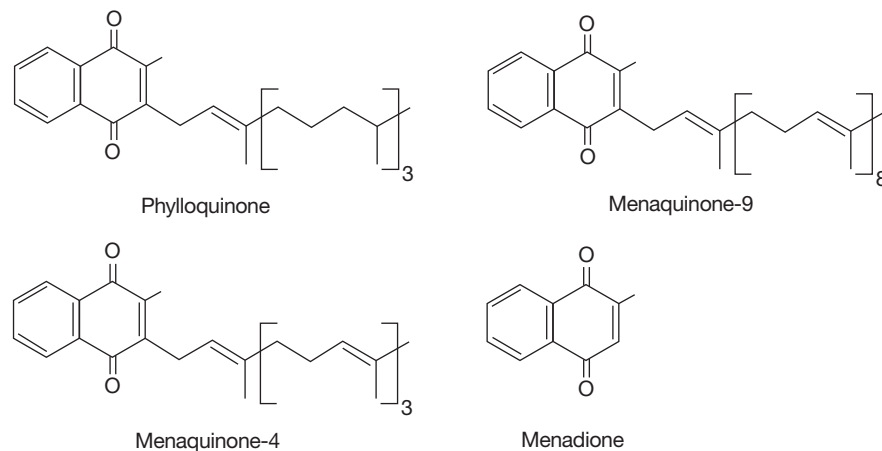


Figure 2 Structures of vitamin K active compounds. Phylloquinone (vitamin K₁) synthesized in plants is the main dietary form of vitamin K. Menaquinone-9 is a prominent member of a series of menaquinones (vitamin K₂) produced by intestinal bacteria, and menadione, vitamin K₃, is a synthetic compound that can be converted to menaquinone-4 by animal tissues.

substituted at the 3-position with a rather hydrophobic group. Although some differences in biological activity can be measured, phylloquinone, the plant form of the vitamin, and the predominant bacterial menaquinones are all effective substrates (Figure 2).

Only a small fraction of the proteins secreted by the liver to the plasma is vitamin K dependent, so an efficient mechanism for recognizing the precursors of the vitamin K-dependent proteins is essential. The primary gene products of the vitamin

K-dependent proteins contain a very homologous domain between the amino terminus of the mature protein and the signal sequence. This propeptide region appears to be both a docking or recognition site for the enzyme and a modulator of the activity of the enzyme by decreasing the apparent K_m of the Glu site substrate. The role of vitamin K in the overall reaction catalyzed by the enzyme is to abstract the hydrogen on the γ -carbon of the glutamyl residue to allow attack of CO₂ at this position. A number of studies that utilized substrates

tritiated at the γ -carbon of each Glu residue have been used to define the action and stoichiometry involved and have shown that it was an apparent equivalent stoichiometry between vitamin K-2,3-epoxide formation and Gla formation. The mechanism by which epoxide formation is coupled to γ -hydrogen abstraction is key to a complete understanding of the role of vitamin K. The association between epoxide formation, Gla formation, and γ -C-H bond cleavage has been studied; the reaction efficiency defined as the ratio of Gla residues formed to γ -C-H bonds cleaved has been shown to be independent of Glu substrate concentrations, and to approach unity at high CO_2 concentrations. Identification of an intermediate chemical form of vitamin K, which could be sufficiently basic to abstract the γ -hydrogen of the glutamyl residue, has been a challenge. A possible candidate is the one first proposed by Dowd, who suggested that an initial attack of O_2 at the naphthoquinone carbonyl carbon adjacent to the methyl group results in the formation of a dioxetane ring that generates an alkoxide intermediate. This intermediate is hypothesized to be the strong base that abstracts the γ -methylene hydrogen and leaves a carbanion that can interact with CO_2 . This pathway leads to the possibility that a second atom of molecular oxygen can be incorporated into the carbonyl group of the epoxide product as well as the epoxide oxygen, and this partial dioxygenase activity has been demonstrated. Although the general scheme shown in Figure 1 is consistent with all of

the available data, the mechanism remains a hypothesis at this time.

Progress in purifying this enzyme was slow. Utilization of bound propeptide as an affinity column aided progress in this field, and the enzyme was eventually purified to near homogeneity and cloned. The carboxylase is a unique 758-amino-acid residue protein with a sequence suggestive of an integral membrane protein with a number of membrane-spanning domains in the N terminus, and a C-terminal domain located in the lumen of the endoplasmic reticulum. It has been demonstrated that the multiple Glu sites on the substrate for this enzyme are carboxylated processively as they are bound to the enzyme via their propeptide, while the Gla domain undergoes intramolecular movement to reposition each Glu for catalysis, and that release of the carboxylated substrate is the rate-limiting step in the reaction.

Metabolic Interconversion of Vitamin K

Gla residues are not metabolized but are excreted in the urine. The amount of Gla excreted by an adult human is in the range of $50 \mu\text{mol d}^{-1}$, so a similar amount must be formed each day. As the dietary requirement of vitamin K is only $\sim 0.2 \mu\text{mol d}^{-1}$, and tissue stores are very low, it is clear that the vitamin K 2,3-epoxide generated by the carboxylase can be actively recycled by the pathway shown in Figure 3. The vitamin

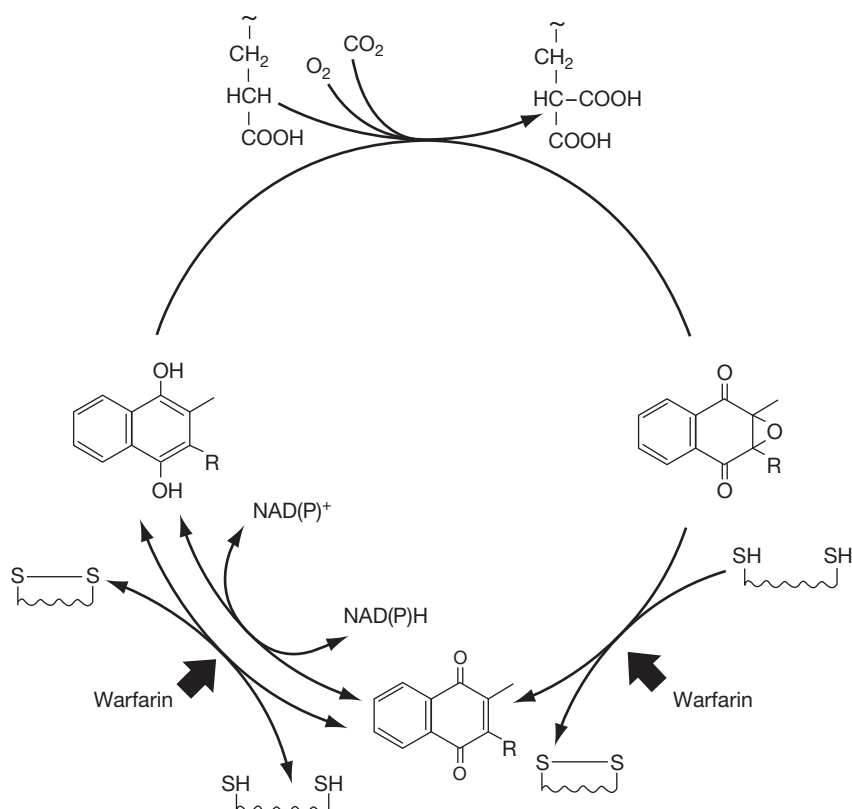


Figure 3 Tissue metabolism of vitamin K. Vitamin K epoxide formed in the carboxylation reaction is reduced to the quinone form of the vitamin by a warfarin-sensitive pathway, vitamin K epoxide reductase, that is driven by a reduced dithiol. The naphthoquinone form of the vitamin can be reduced to the hydronaphthoquinone form either by the same warfarin-sensitive dithiol-driven reductase or by one or more of the hepatic NADH or NADPH-lined quinone reductases which are less sensitive to warfarin.

K-epoxide reductase is capable of reducing both the epoxide to the naphthoquinone and naphthoquinone to the hydro-naphthoquinone, and this is the enzyme which is blocked by the common oral anticoagulant warfarin. In addition to the epoxide reductase, the quinone and hydronaphthoquinone forms of the vitamin can also be interconverted by a number of reduced form of nicotinamide adenine dinucleotide phosphate (NAD(P)H)-linked reductases including one that appears to be a microsomal-bound form of the extensively studied liver DT-diaphorase activity. The epoxide reductase utilizes a sulfhydryl compound as a reductant *in vitro*, but the physiological reductant has not been identified.

A second metabolic pathway utilizing phyloquinone has been discovered more recently. A number of extrahepatic tissues such as brain, salivary gland, and pancreas of rats and humans fed with phyloquinone have been found to contain a much higher concentration of menatetrenone (MK-4) than of phyloquinone. Tissue formation of MK-4 from phyloquinone fed to germ-free rats and the observation that cultured kidney cells can convert phyloquinone to MK-4 have now demonstrated that bacterial action is not involved in the conversion and that numerous cell types have the ability to carry out this transformation. Details of the mechanism of this conversion are lacking, but it seems unlikely that a metabolic pathway leading to MK-4 has evolved unless there is some specific role for this vitamin. This role is unlikely to involve the vitamin K-dependent carboxylase, as phyloquinone and MK-4 have similar activity as a substrate for this enzymatic activity, and is more likely to act as a ligand for various receptors regulating metabolism.

Nutritional Aspects of Vitamin K

Reports of uncomplicated adult deficiencies of vitamin K are extremely rare, and most diets contain an adequate amount ($\sim 100 \mu\text{g d}^{-1}$) of vitamin K. Low lipid intake or the impaired lipid absorption resulting from the lack of bile salts will adversely affect vitamin K absorption, and depression of the vitamin K-dependent coagulation factors has frequently been

reported in various malabsorption syndromes. This classic example of a human vitamin K deficiency is that of early vitamin K deficiency bleeding occurring during the first week of life in healthy-appearing neonates. Low placental transfer of phyloquinone, low clotting factor levels, a sterile gut, and the low vitamin K content of breast milk all contribute to the disease. Although the incidence is low, the mortality rate from intracranial bleeding is high, and prevention by oral or parenteral administration of vitamin K immediately following birth is the standard cure. The presence of large amounts of OC in bone and reports that low vitamin K intake or increased amounts of circulating under- γ -carboxylated OC are correlated with low bone mineral density or fracture rate have focused significant attention on a possible role for vitamin K in bone health. A number of controlled studies of phyloquinone supplementation have, however, failed to demonstrate that additional intake of the vitamin will decrease bone fracture rate or increase bone mineral density. MGP, which has been shown to present ectopic calcification in animal models, is also being studied to determine if it plays a role in controlling cardiovascular disease.

See also: **Bioenergetics: Quinones.**

Further Reading

- Berkner KL (2005) The vitamin K-dependent carboxylase. *Annual Review of Nutrition* 25: 127–149.
- Booth SL (2009) Roles for vitamin K beyond coagulation. *Annual Review of Nutrition* 29: 89–110.
- Gundberg C (2009) Vitamin K and bone: Past, present, and future. *Journal of Bone and Mineral Research* 24: 980–982.
- Presnell SR and Stafford DW (2002) The vitamin K-dependent carboxylase. *Journal of Thrombosis and Haemostasis* 87: 937–946.
- Shearer MJ and Newman P (2008) Metabolism and cell biology of vitamin K. *Journal of Thrombosis and Haemostasis* 100: 530–547.
- Suttie JW (2007) Vitamin K. In: Zempleni J, Rucker RB, McCormick DB, and Suttie JW (eds.) *Handbook of Vitamins*, pp. 111–152. Boca Raton, FL: CRC Press, Taylor and Francis Group.

Voltage-Dependent K⁺ Channels

R Latorre, Universidad de Valparaíso, Valparaíso, Chile

F J Morera, University of California, Los Angeles, Los Angeles, CA, USA

C Zaelzer, Montreal General Hospital, Montreal, QC, Canada

© 2013 Elsevier Inc. All rights reserved.

This article is a revision of the previous edition article by Ramón Latorre and Francisco J. Morera, volume 4, pp. 399–404, © 2004, Elsevier Inc.

Glossary

Delayed rectifier A K⁺ channel that changes the membrane conductance with a delay after a depolarizing voltage step.

Gating The opening or closing of a channel in response to some stimulus.

Inward rectifier A K⁺ channel that opens when the membrane is hyperpolarized.

Long QT syndrome An inherited cardiac arrhythmia.

Missense mutation A mutation in which an incorrect amino acid is incorporated into the protein.

Voltage sensor A channel structure able to detect changes in membrane potential. In voltage-dependent ion channels, the voltage-sensing elements are located in the S4 segment.

Structure and Diversity

Voltage-dependent K⁺ channels are members of the voltage-dependent cation channel family, which includes the voltage-dependent Na⁺, Ca²⁺, and K⁺ channels. The first voltage-dependent K⁺ channel was cloned in 1987 from a fruit fly, *Drosophila*, based on a mutation that causes the *Shaker* phenotype. After that, they have been identified in an immense number of organisms including prokaryotes and eukaryotes.

The 78 members of the K⁺ channel superfamily can be divided into four structural types based on their mode of activation and the number of their transmembrane segments: (1) voltage-gated six-transmembrane K⁺ channels (K_v), (2) voltage- and Ca²⁺-activated seven-transmembrane K⁺ channels (K_{Ca}), (3) Inward rectifying two-transmembrane K⁺ channels (K_{ir}), and (4) two-pore four-transmembrane K⁺ channels (K_{2P}). Only the K_v and K_{Ca} families have channel-forming proteins containing intrinsic voltage sensors (see below).

All the voltage-dependent K⁺ channels are tetrameric assemblies. In the K_v channels, each subunit consists of six hydrophobic segments, S1–S6. The primary sequence of these hydrophobic segments shows similarities between all of the K_v channels, including the voltage-sensor domain (VSD, S1–S4) and the ion-conduction pore domain (S5–S6) (**Figure 1(a)**).

All K⁺ channels have a consensus amino acid sequence inside the pore region, TVGYGD, dubbed the ‘signature sequence’. The residues, TVGYG, line the selectivity filter (in some K⁺ channels, the tyrosine (Y) residue is replaced by phenylalanine (F), e.g., in the ether-a-go-go (eag) channel discussed later on). In the selectivity filter, carbonyl oxygen atoms are directed toward the pore to coordinate dehydrated K⁺ ions. Hydrophobic chains of valine and tyrosine directed toward the hydrophobic core surrounding the filter stabilize the main chain.

Gating in the K_v channels is conferred through the attachment of gating domains to the pore. The basic function of these gating domains is to perform mechanical work on the ion-conduction pore to change its conformation between closed

and opened states. In voltage-dependent channels, a VSD (S1–S4) is present on each subunit. Thus, a voltage sensor converts energy stored in the membrane electric field into mechanical work. There is strong evidence that the positive charges contained in S4 are the voltage-sensing elements. Thus, ion-channel gating is essentially an electromechanical coupling between a gating unit and a pore unit.

The crystal structure of a mammalian voltage-dependent K⁺ channel (K_v1.2) had initially been resolved at 2.9 Å and further improved to 2.4 Å using a chimeric K_v1.2–K_v2.1 channel. In the latter case, the channel was crystallized in complex with lipids. These structures showed that the helices of the ion-conduction pore (S5–S6) related to the helices of the VSD (S1–S4) in a special way. The VSD of one subunit is located near the pore domain of an adjacent subunit (**Figure 1(b)**). The connection between the pore and the VSD is made by the S4–S5 linker helix, which runs parallel to the intracellular membrane surface.

Voltage-dependent K⁺ channels are ubiquitously distributed in different cells and tissues, therefore, K⁺ channel diversity is of great importance in determining the variety of electrical responses produced by cells when subjected to stimuli. The possible mechanisms that originate the immense voltage-dependent K⁺ channel diversity are: (1) multiple genes; (2) alternative splicing; (3) formation of heteromultimeric channels; and (4) coexpression with regulatory β-subunits.

Voltage-dependent K⁺ channel diversity seems to have accompanied animal cells through evolution. Some K⁺ channels are the most conserved proteins in eukaryotes. Potassium channels became fundamental to animal cell physiology very early in evolution, and through voltage-dependent K⁺ channel control of membrane potential they couple the cell inner dynamics to the outer environment. A differential expression of voltage-dependent K⁺ channel messenger RNAs (mRNAs) is found to accompany the events of animal development and also occurs in adult animal tissues.

The diversity of voltage-dependent K⁺ channel genes is shown in the dendrogram shown in **Figure 2**. The dendrogram shows the major classes of voltage-dependent K⁺ channels. The ramification of voltage-dependent K⁺ channels was

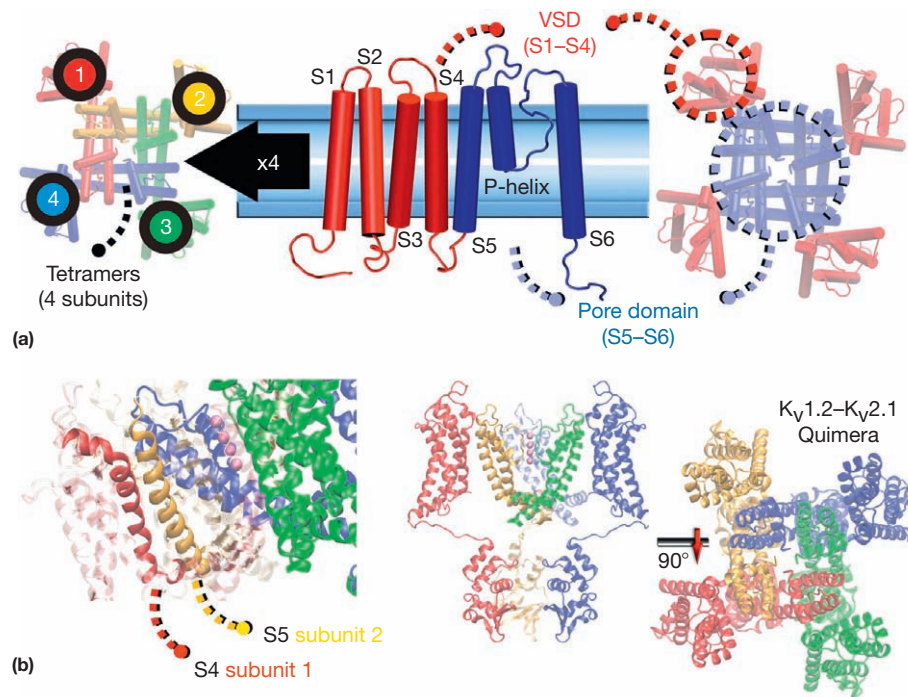


Figure 1 Structural characteristics of voltage-dependent potassium channels. (a) The figure in the center shows a six-transmembrane segment secondary structure of an archetypical voltage-dependent potassium channel. Cylinders represent α -helix domains crossing the bilayer connected by loops. The voltage-sensor domain (VSD) is defined as the structure composed of S1–S4 (red cylinders). The channel pore consists of segment S5 and S6 and a short α -helix denominated the pore helix (P-helix) (blue cylinders). The sequence GYGD is located in the loop that connects the P-helix with S6. The functional channel is a tetramer (left) and the voltage sensor and the channel pore are depicted in the tetramer in the right figure. (b) Crystal structure of the chimerical channel $K_v1.2$ – $K_v2.1$ (access number 2R9R). The left side of the figure shows an amplification of the transmembrane segments S4 (ribbon in red) belonging to the subunit labeled as 1, and S5 (ribbon in yellow) belonging to the contiguous subunit, labeled as 2, showing that in the quaternary structure these segments are spatially related. The central figure shows the lateral structure of the channel. Part of the channel structure has been removed in order to leave the conduction pathway exposed. Four K⁺ ions (purple spheres) are shown coordinated by sites in the selectivity filter. A 90° rotation of the structure shown in the central figure allows us to appreciate the external aspect of the channel.

constructed by carrying out sequence alignments. To compare the amino acid sequences of the different voltage-dependent K⁺ channels, the sequences were aligned using at least 300 amino acids from six hydrophobic segment (S1–S6) channels and their respective linkers. Branch lengths do not correspond to time; however, branching is expected to preserve evolutionary relationships.

Physiological Function

Voltage-Gated Six-Transmembrane K⁺ Channels

The vertebrate fast voltage-gated K⁺ delayed rectifier or, K_v channels, fall into a variety of families (K_v1 – K_v12) according to their amino acid sequence similarity (Figure 2). A channel from the fruit fly *Drosophila* was the first K⁺ channel to be cloned (*Shaker* and denominated in vertebrates as $K_v1.1$, the first number denoting subfamily and the second, order of discovery). The identification started with mutant flies called shakers, which shake their legs while under ether anesthesia. These flies lack a fast-transient K⁺ current in presynaptic terminals, so repolarization of the action potential is delayed and the release of neurotransmitter from a single action potential becomes enormous.

K_v channels are found in a wide range of different tissues and have many different functions. For example, $K_v3.1$ channels are found in some neurons that are specialized to fire very short action potentials at high rates, such as those of the auditory system; $K_v1.3$ channels found in leukocytes play a role in inflammatory response; $K_v1.2$ channels are found in smooth muscle and participate in contractile tone regulation; and $K_v4.2$ channels expressed in myocardium prolong cardiac action potential.

The K_v7 subfamily is also known as KCNQ channels. The first member of the KCNQ subfamily to be cloned, *KCNQ1* ($K_v7.1$), was isolated because mutations in this gene give rise to the most common form of long QT syndrome, LQT1. *KCNQ1* encodes the major subunit of the cardiac I_{Ks} channel, which is involved in the repolarization of ventricular action potential. *KCNQ1* mRNA is expressed strongly in the heart, with lower levels in the pancreas, kidney, lung, placenta, and ear. Exploiting their homology to *KCNQ1*, other members of the K_v7 subfamily were subsequently cloned. It has been suggested that *KCNQ2* ($K_v7.2$) and *KCNQ3* ($K_v7.3$) form a heteromer, an idea which is consistent with their overlapping tissue distribution and studies where antibodies directed against *KCNQ2* are able to coimmunoprecipitate *KCNQ3* and vice versa. The heteromeric *KCNQ2/KCNQ3* channel corresponds to the

M-channel found in neurons. They are widely distributed throughout the brain and sympathetic and dorsal root ganglia, and expressed at high levels in hippocampus, chordate nucleus, and amygdala. KCNQ4 (K_v7.4) is expressed in outer hair cells and neurons of the auditory system and vascular smooth muscle. K_v5, K_v6, K_v8, and K_v9 channels are not functional alone; they coassemble with K_v2 subunits and modify their function.

Defects in K_v channels' function may have profound physiological effects. Linkage studies have identified the gene responsible for episodic ataxia type-1. This gene KCNA1 encodes the K_v1.1 channel, which is expressed in synaptic terminals and dendrites in brain neurons. Moreover, a variant in the promoter of the gene KCNA3 (encoded for K_v1.3 channel) is associated with impaired glucose tolerance and lower insulin sensitivity. A missense mutation in the gene KCNC3 (K_v3.3) causes spinocerebellar ataxia 13. Mutations in KCNQ2 and KCNQ3 give rise to a form of idiopathic epilepsy known as benign familial neonatal convulsions. Loss-of-function mutations in KCNQ4 cause deafness autosomal-dominant 2a.

Another important subgroup of K_v channels was first cloned in mutant fruit flies, which shook their legs like they

were 'go-go dancing' when anesthetized with ether. These flies were carriers of a mutated gene for a voltage-gated K⁺ known as *eag*. A family of *eag*-like genes was isolated from *Drosophila*, known as *eag*, *erg*, and *elk*. The human *eag*-related gene (human ether-a-go-go-related gene (HERG) or K_v11.1) was cloned from a human hippocampal complementary DNA library. HERG was strongly expressed in the heart. LQT2, a mutation on this gene, impairs the repolarization of the cardiac action potential, causing a form of long QT syndrome. Pharmacological blocking of HERG K⁺ channels reduces I_{Kr} and can prolong ventricular repolarization. This is shown by a lengthening of QT interval on the surface electrocardiogram – QT prolongation is a well-known risk factor for the occurrence of ventricular tachycardia.

Plant K_v-Like Channels: Inward and Outward Rectifiers

Potassium is a major factor in resistance to drought, salinity, and fungal disease in plants, as plant cells utilize inorganic ions, especially K⁺, as major osmotically active solutes to drive changes in cell volume. Voltage-dependent inward K⁺ channels (K channel from *Arabidopsis thaliana* (KAT)1/2,

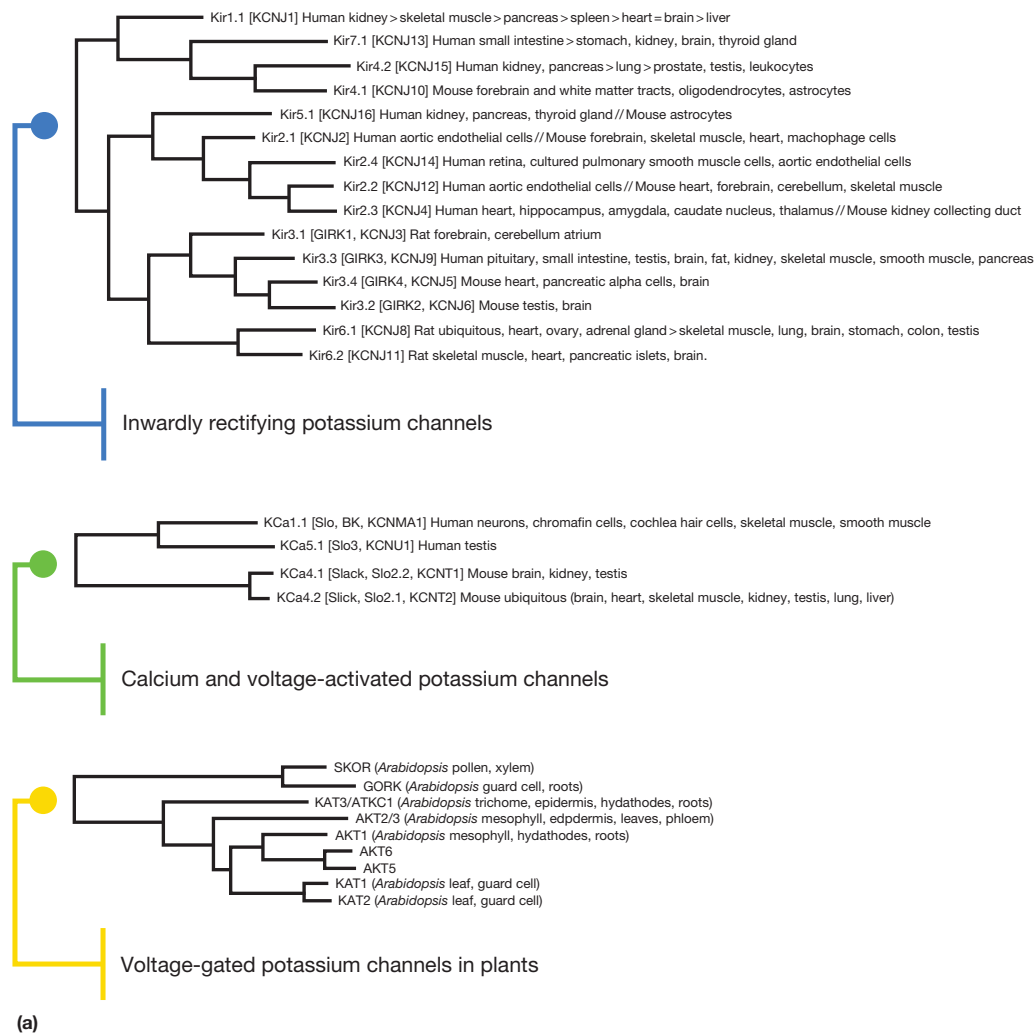


Figure 2 (Continued)

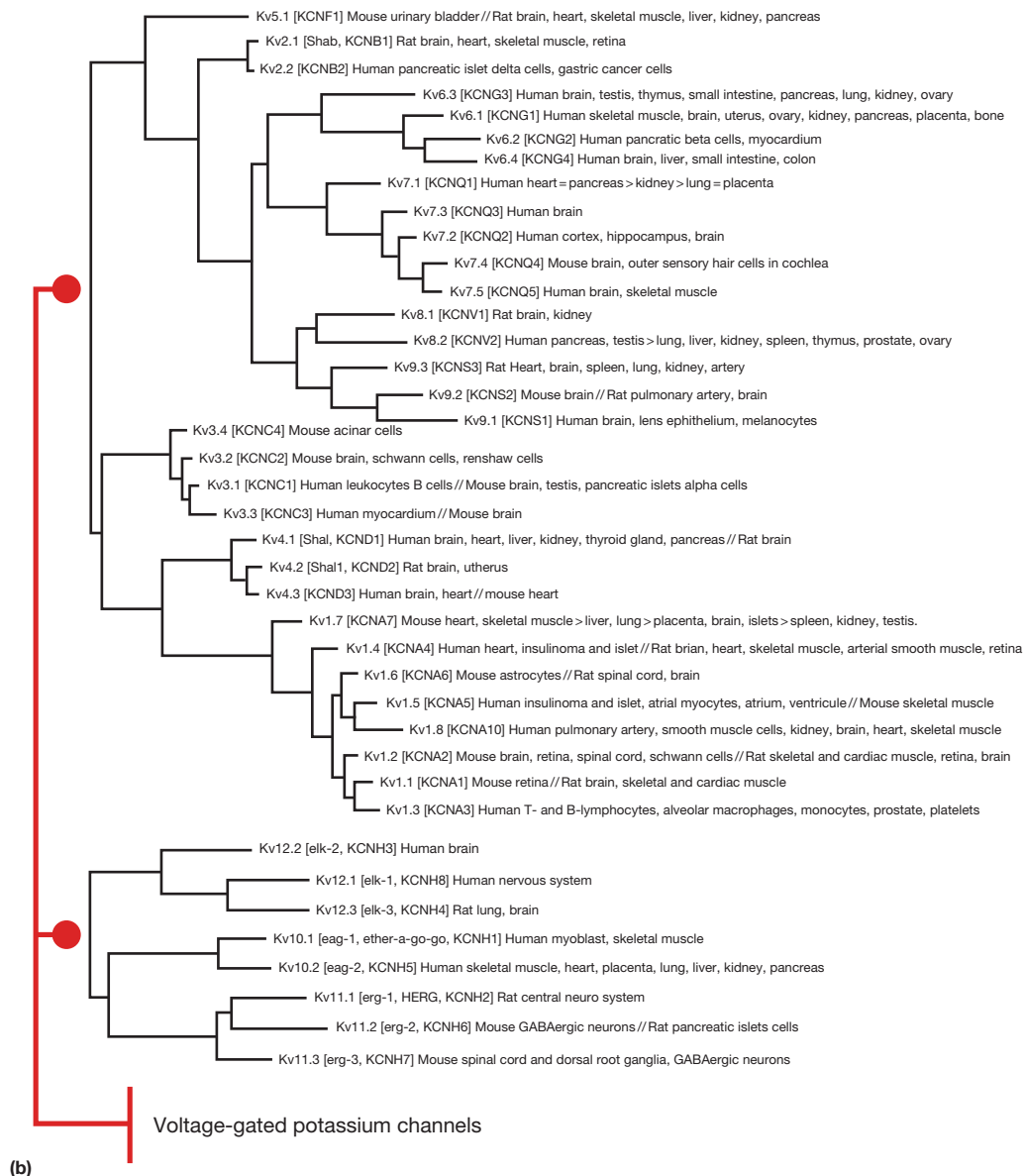


Figure 2 Major classes of voltage-dependent K⁺ channels and tissue localization. (a) Amino acid sequence alignments for the inward rectifier, K_{ir} family, the human KCa1/4/5 group family, and the voltage-gated K⁺ channels in plants were created using CLUSTALW. Analysis by maximum parsimony using PAUP* resulted in the above two first trees, the third one was created using Phylogeny.fr and PhyML 3.0. International Union of Pharmacology and HUGO Gene Nomenclature Committee names of the genes are shown together with their tissue localization in human, mouse, or rat organism for the two first trees. The *Arabidopsis* Information Resource (TAIR) was used for cataloging the genes of voltage-gated K⁺ channels in *Arabidopsis thaliana*. Tissue channel expression together with the corresponding gene is shown. The more similar the sequences for a pair of genes, the shorter the path connecting them. (b) Phylogenetic tree for the K_v1–K_v9 families and K_v10–K_v12: amino acid sequence alignments of the human channel K_v proteins were created using CLUSTALW, and analysis by maximum parsimony using PAUP* resulted in unrooted trees comprising the K_v1–K_v6 and K_v8–K_v9 families and K_v10–K_v12. Sequences of K_v7.1–K_v7.5, K_v6.4, and K_v8.2 were incorporated into the existing tree topology using a combination of maximum parsimony and neighbor-joining analysis. Only the hydrophobic cores (S1–S6) were used for the analysis. The IUPHAR and HGNC names are shown together with the localization in human, mouse, or rat organism. The more similar the sequences for a pair of genes, the shorter the path connecting them.

Arabidopsis thaliana K channel (AKT1/2/3) rectifiers mediate potassium uptake in plants. KAT1 and AKT1 were the first inward rectifier K⁺ channels cloned from the plant *Arabidopsis thaliana*. They are tetramers formed by subunits containing six transmembrane domains with a positively charged fourth (S4) transmembrane segment, such as *Shaker* channels. The KAT1

channel is expressed in guard cells where it regulates the aperture of the stomata and the vascular system of the stem. AKT1 is present in guard cells and in the root cortex and plays an important role in K⁺ uptake from the soil. Voltage-dependent outward rectifiers, on the other hand, are represented by guard cell outward rectifier K channel (GORK) and stomatal potassium

outward rectifier channel (SKOR), present in different tissues of the root. GORK is also present in guard cells regulating the stomata closing. SKOR is expressed within the xylem parenchyma of the root where it facilitates K⁺ loading into the xylem.

Voltage- and Ca²⁺-Activated Seven-Transmembrane K⁺ Channels

A change in the intracellular calcium concentration is a ubiquitous signaling mechanism that is frequently coupled with changes in membrane potential. These signals are integrated by the activation of a heterogeneous family of potassium channels, Ca²⁺-activated potassium channels (K_{Ca}), whose open probability is increased by the elevation of cytosolic calcium to cause membrane hyperpolarization. They are found in a wide range of different tissues and in virtually all multicellular organisms. In nerve cells, they play a key role in controlling the action potential waveform and in regulating cell excitability. In secretory epithelia, they are also involved in K⁺ secretion. K_{Ca} channels contribute to the repolarization and after-hyperpolarization of vertebrate neurons. As this article is related to a voltage-dependent K⁺ channel only, the importance of one K_{Ca} channel family code by a single gene called *slowpoke* (*slo*) is discussed in this section.

Expression of *Slo* in mammalian cells or *Xenopus* oocytes induces the appearance of a highly K⁺-selective, Ca²⁺-activated, and voltage-dependent channel also known as MaxiK or BK (big K) because of its large single-channel conductance (>200 pS in 100-mM symmetrical K⁺). In contrast to its other voltage-dependent K⁺ channel congeners, the *Slo* protein has an extra hydrophobic segment at the NH₂ terminus called S0. Thus, the NH₂ terminus of the protein is placed in the extracellular side of the membrane. The intracellular COOH-terminal comprises about two-thirds of the protein containing four hydrophobic segments, several alternative splicing sites, and a stretch of negatively charged amino acids (aspartates) dubbed the 'Ca²⁺ bowl'. The Ca²⁺ bowl has emerged as a good candidate for the high-affinity Ca²⁺-binding site. The cytoplasmic domain of each *Slo* 1 subunit can be divided into two separate structures, both with sequence homology to the regulatory domain for K⁺ conductance (RCK) domains and was proposed that *Slo*1 channels would also contain a gating ring formed by eight RCK domains from the four *Slo*1 subunits.

Slo channels control a large variety of physiological processes, including smooth muscle tone, neurosecretion, and hearing. Despite being coded by a single gene, the diversity of *Slo* channels is great. Regulatory β -subunits, splicing, and metabolic regulation create this diversity which is fundamental to the adequate function of many tissues. Except in heart myocytes, *Slo* channels are almost ubiquitously expressed

among mammalian tissues, and play a variety of roles. *Slo* channels serve as negative feedback regulators of membrane potential and intracellular calcium concentration, which are important in a number of physiological processes. In smooth muscle, local Ca²⁺ increases, called Ca²⁺ sparks, generate spontaneous transient outward currents, which are generated by *Slo* channels. This hyperpolarizes the membrane causing muscle relaxation. In chromaffin cells, *Slo* channels are important for rapid termination of the action potential. In the cochlea of turtles, frogs, and chicks, gradients of spliced *Slo* channels, and β -subunits through interplay with voltage-dependent Ca²⁺ channels, they are able to generate the different oscillatory responses in the hair cells.

In agreement with their physiological importance, malfunction of BK channels (α - or β -subunits) can lead to epilepsy, motor impairment, noise-induced hearing loss, hypertension, urinary incontinence, overactive urinary bladder, and asthma.

See also: Bioenergetics: Voltage-Gated Ca²⁺ Channels; Lipids Carbohydrates Membranes and Membrane Proteins: Ion Channel Protein Superfamily.

Further Reading

- Ahern CA and Horn R (2004) Stirring up controversy with a voltage sensor paddle. *Trends in Neuroscience* 27(6): 303–307.
- Bezaniilla F (2008) Ion channels: From conductance to structure. *Neuron* 60(3): 456–468.
- Chérel I (2004) Regulation of K⁺ channel activities in plants: From physiological to molecular aspects. *Journal of Experimental Botany* 55(396): 337–351.
- Cui J, Yang H, and Lee US (2009) Molecular mechanism of BK channel activation. *Cellular and Molecular Life Science* 66: 852–875.
- Dreyer I and Blatt MR (2009) What makes a gate? The ins and outs of K_v-like K⁺ channels in plants. *Trends in Plant Science* 14(7): 383–390.
- Goldstein SA, Bockenhauer D, O'Kelly I, and Zilberger N (2001) Potassium leak channels and the KCNK family of two-P-domains subunits. *Nature Reviews Neuroscience* 2: 175–184.
- Hille B (2001) *Ion Channels of Excitable Membranes*. Sunderland, MA: Sinauer Associates.
- Latorre R and Brauchi S (2006) Large conductance Ca²⁺-activated K⁺ (BK) channel: Activation by Ca²⁺ and voltage. *Biological Research* 39(3): 385–401.
- Long SB, Cambell EB, and MacKinnon R (2005) Crystal structure of a mammalian voltage-dependent Shaker family K⁺ channel. *Science* 309: 897–903.
- Long SB, Cambell EB, and MacKinnon R (2005) Voltage sensor of K_v1.2: Structural basis of electromechanical coupling. *Science* 309: 903–908.
- Long SB, Tao X, Cambell EB, and MacKinnon R (2007) Atomic structure of a voltage-dependent K⁺ channel in a lipid membrane-like environment. *Nature* 450: 376–382.
- Tombola F, Pathak MM, and Isacoff EY (2006) How does voltage open an ion channel? *Annual Review of Cell and Developmental Biology* 22: 23–52.
- Torres YP, Morera FJ, Carvacho I, and Latorre R (2007) A marriage of convenience: Beta-subunits and voltage-dependent K⁺ channels. *Journal of Biological Chemistry* 282(34): 24485–24489.
- Wulff H, Castle NA, and Pardo LA (2009) Voltage-gated potassium channels as therapeutic targets. *Nature Review Drugs Discovery* 8: 982–1001.

Voltage-Gated Ca^{2+} Channels

H Reuter, University of Bern, Bern, Switzerland

© 2013 Elsevier Inc. All rights reserved.

Glossary

apoCaM Ca^{2+} -free calmodulin.

Electrical activity Results from ion flow through channels in membranes.

Electrochemical gradient Driving force for ions due to differences in concentrations and electrical potentials across membranes.

G proteins Small intracellular proteins involved in signaling pathways.

Oligomeric proteins Composition of several protein subunits.

Protein kinases Enzymes that catalyze phosphate transfer from adenosine triphosphate (ATP) to proteins; protein kinases are activated by distinct second messengers.

Second messengers Intracellular molecules formed after receptor activation and involved in signaling (e.g., cyclic-adenosine monophosphate (AMP), diacylglycerol).

Signaling cascade Information transfer within a cell.

Ion channels are integral, pore-forming transmembrane proteins that serve to pass ions across biological membranes. The channels differ in their ion selectivity, that is, each type allows certain ion species to permeate through the open channel pore. Conformational changes of the proteins leading to openings or closings of the channels are called 'gating'. It is achieved either by binding of ligands to special sites at the channels (ligand-gated ion channels) or by changes in membrane potential (voltage-gated ion channels). Calcium (Ca^{2+}) channels are voltage gated and have a high selectivity for Ca^{2+} ions. Opening of these channels by membrane depolarization leads to an influx of Ca^{2+} into the cell. The resulting increase in the intracellular Ca^{2+} concentration initiates essential physiological functions, such as cardiac contraction, secretion of hormones and neurotransmitters, gene transcription, or repetitive electrical activity of neurons and the heart. Most of these functions result from binding of Ca^{2+} to specific intracellular proteins (e.g., calmodulin, troponin C, etc.) which subsequently mediate the Ca^{2+} -dependent cellular responses. The diversity of voltage-gated Ca^{2+} channels in surface membranes of different cells is important for initiation of specific signaling cascades within the cells. The various types of Ca^{2+} channels are encoded by different genes and show tissue-specific expression. They are targets for toxins and for drugs used for therapeutic purposes.

Molecular Properties

Structures

Voltage-gated Ca^{2+} channels are oligomeric protein complexes composed of the voltage-sensitive, pore-forming α_1 -subunits, together with auxiliary disulfide-linked $\alpha_2\delta$ -subunits and intracellular β -subunits (Figure 1). An additional transmembrane γ -subunit is associated with $\alpha_1\text{S}$ in skeletal muscles. The diversity of Ca^{2+} channel types not only results from different genes coding for the different channel subunits (10 for α_1 , 4 for $\alpha_2\delta$, and 4 for β), but also from multiple splice variants of the subunits. The α_1 -subunits consist of four homologous domains (I–IV), each being composed of six transmembrane α -helical segments (S1–S6). The N and C termini are located

intracellularly, and the domains are linked together by intracellular loops. In the folded structures of α_1 -subunits, hairpin loops between transmembrane segments S5 and S6 constitute the channel pore. Segment S4 contains highly conserved, positively charged amino acids that act as voltage-sensitive sensors during gating of the channels. The $\alpha_2\delta$ -subunits are involved in the regulation of the channels' gating properties (voltage-dependent inactivation) and, like $\alpha_2\delta$ -subunits, play a role in the functional expression of the channels in cell surface membranes.

Selectivity

Ca^{2+} channels are selective pathways for the movement of Ca^{2+} ions down their electrochemical gradient into the cell. The pore-forming structures of the channel contain four negatively charged glutamate residues, each being in a homologous position on the four domains. The resulting ring of negative charges facing the pore coordinates binding of an entering Ca^{2+} ion. A second Ca^{2+} ion that enters the channel reduces the binding affinity of the first one by electrostatic repulsion, thus kicking it off from its binding site and letting it move down its electrochemical gradient. Thus, the selectivity of the channel for Ca^{2+} is critically dependent on the correctly placed negative charges of glutamate in the pore.

Calmodulin Binding

Two forms of autoregulation, Ca^{2+} -dependent inactivation and facilitation of $\text{Ca}_v1.2$ channels (Table 1), involve calmodulin (CaM). The subunit $\alpha_1\text{C}$ of $\text{Ca}_v1.2$ contains a consensus CaM-binding isoleucine-glutamine (IQ)-motif in its cytoplasmic C-terminal tail. The IQ-motif consists of 12 amino acids. Extensive mutations within this sequence revealed structural determinants for the two opposing forms of channel regulation. Two short stretches of amino acids upstream of the IQ-motif seem to be sites for constitutive, Ca^{2+} -independent binding of apoCaM. The IQ-motif binds CaM only in the presence of Ca^{2+} . It represents the regulatory site for Ca^{2+} -dependent inactivation, while facilitation may require additional activation of Ca/CaMkinaseII.

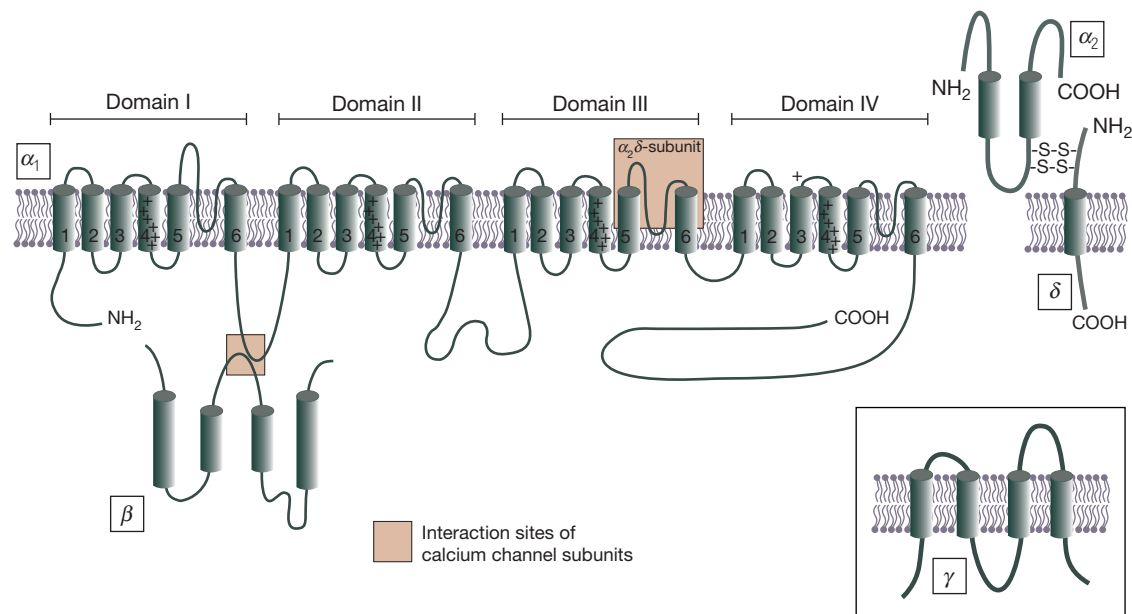


Figure 1 Oligomeric subunit complex of voltage-gated Ca²⁺ channels. The subunit composition consists of the pore-forming α_1 -subunit with its four homologous domains, the intracellular β -subunit, and the disulfide-linked α_2 -subunit. The sites of interaction between the subunits are indicated by the squares. An extra γ -subunit is present in skeletal muscle $\text{Ca}_v1.1$ complex (inset). Kindly prepared by Dr. R. D. Zuehlke, Bern.

Table 1 Ca²⁺ channels

Channel types ^a	Pore-forming α -subunits ^b	Tissue distribution	Pharmacology (blockers)
HVA			
Ca _v 1.1 (L)	$\alpha_1\text{S}$	Skeletal muscle	DHPs
Ca _v 1.2	$\alpha_1\text{C}$	Heart, smooth muscle, CNS	
Ca _v 1.3	$\alpha_1\text{D}$	CNS, kidney, cochlea, glands	
Ca _v 1.4	$\alpha_1\text{F}$	Retina	
Ca _v 2.1 (P/Q)	$\alpha_1\text{A}$	CNS	ω -Agatoxin
Ca _v 2.2 (N)	$\alpha_1\text{B}$	CNS	ω -Conotoxin.GVIA
Ca _v 2.3 (R)	$\alpha_1\text{E}$	CNS, heart, glands	
LVA			
Ca _v 3.1 (T)	$\alpha_1\text{G}$	CNS, sinus node	Mibefradil
Ca _v 3.2	$\alpha_1\text{H}$	CNS, heart, kidney, liver	
Ca _v 3.3	$\alpha_1\text{I}$	CNS	

^aChannel types: HVA, high-voltage-activated channels; LVA, low-voltage-activated channels; letters in brackets refer to the nomenclature of currents carried through the different channel types.

^bPore-forming α_1 -subunits encoded by identified genes have multiple splice variants (not shown); CNS, central nervous system; DHPs, dihydropyridines.

Phosphorylation

Ca²⁺ channel function can be modulated by second messenger-activated protein kinases (PKs) such as cyclic adenosine monophosphate (AMP)-dependent PKA, diacylglycerol-dependent PKC, or Ca/calmodulin-dependent kinases. Although the exact molecular mechanisms of these modulatory pathways are not yet fully understood, they do depend on

phosphorylation by kinases of specific amino acid residues of the Ca²⁺ channel protein. Multiple intracellular sites of phosphorylation have been identified on α_2 - and β -subunits. However, which phosphorylation sites are necessary and sufficient for different forms of modulation of Ca²⁺ channel function by the various protein kinases has not been clearly resolved.

G Proteins

Ca²⁺ channels of the Ca_v2 family (Table 1) that are involved in neurotransmitter and hormone secretion are inhibited by G proteins. There is excellent evidence that the rapid inhibition of the channels results from direct binding of the $\beta\gamma$ -subunits of G proteins to an identified segment of the intracellular loop between domains I and II (Figure 1), but other sequences in the C and N termini of the α_2 -subunits may also be important. Rapid inhibition of the channels by this mechanism does not require formation of soluble intracellular messengers. The inhibition results from a membrane-delimited pathway leading to the release of the $\beta\gamma$ -subunit of the G-protein complex.

Ca²⁺-Currents

Physiology

Depolarization of cell membranes during excitation leads to spontaneous openings (activation) and closings (inactivation) of Ca²⁺ channels. At a given voltage, stochastic openings of multiple channels produce a considerable flow of Ca²⁺ ions into a cell and can thus be measured as an electrical current. The currents are recorded by electrophysiological methods, at a level of resolution where even the current through an individual channel can be measured (patch-clamp method). This allows detailed quantitative analyses of the kinetic properties

(gating) and selectivities of the channels. Currents through different types of Ca^{2+} channels have traditionally been classified as L-, P/Q-, N-, R-, and T-type (Table 1). This classification resulted from functional and pharmacological analyses before the channels were cloned and sequenced. A general overview of the distribution of the channels in different tissues is summarized in Table 1. The specific expression of different Ca^{2+} channel types in different tissues is of great physiological and pharmacological importance. The localization of individual channels in association with other cellular constituents governs specific signaling pathways. For example, P/Q- and N-type channels are directly linked to the soluble *N*-ethylmaleimide sensitive fusion protein attachment protein receptor (SNARE) protein complex in neural and endocrine cells that is essential for secretion of neurotransmitters and hormones. L-type channels that are predominant in cardiac cells are in close proximity to intracellular Ca^{2+} stores. When Ca^{2+} ions move through the channels into these cells during excitation, they release more Ca^{2+} from the stores which eventually leads to contraction of the heart. Thus, changes in the intracellular Ca^{2+} -concentration are coding for important information of diverse cellular functions. Auto-regulatory mechanisms, by which the channels inactivate in voltage- and Ca^{2+} -dependent manners, prevent Ca^{2+} overload of the cells. Ca^{2+} -dependent inactivation has been studied most extensively in L-type Ca^{2+} channels, but may also apply to other channel types. During the flow of Ca^{2+} ions through the pore, they bind to constitutively tethered apoCaM at the cytoplasmic C terminus of the channel. Occupancy of CaM by Ca^{2+} triggers binding of its C-terminal lobe to the nearby IQ-motif and accelerates inactivation of the

channel, possibly by disinhibiting voltage-dependent inactivation (Figure 2).

Modulation

Modulation of Ca^{2+} channel activity by neurotransmitters and hormones has become a major field of research. The upregulation of cardiac Ca^{2+} current ($\text{Ca}_v1.2$) by epinephrine was the first example of modulation of an ion channel by a neurotransmitter. It has been shown to be the result of activation of the cyclic AMP pathway via β_1 -adrenoceptors and subsequent phosphorylation by protein kinase A of amino acid (serine) residues at intracellular sites of the channel. This causes enhanced probability of openings of the channels, resulting in largely increased Ca^{2+} currents. Other examples of modulation mainly concern N- and P/Q-type Ca^{2+} channels ($\text{Ca}_v2.1$ and $\text{Ca}_v2.2$; Table 1). They are inhibited by neurotransmitters, such as norepinephrine. Release of G-protein $\beta\gamma$ -subunits by neurotransmitters shifts channel activation into a reluctant state, thus causing a reduction of Ca^{2+} currents during depolarization. These are just two examples of modulation of Ca^{2+} channel activities. There are several others that are not included under the scope of this article.

Pharmacology

Ca^{2+} channels are important targets for toxins and drugs. Although a full discussion of their effects is beyond the scope of this article, a few of them are listed in Table 1. Dihydropyridines (e.g., nifedipine and nimodipine) are relatively specific inhibitors of the Ca_v1 channel family, mainly $\text{Ca}_v1.2$ type.

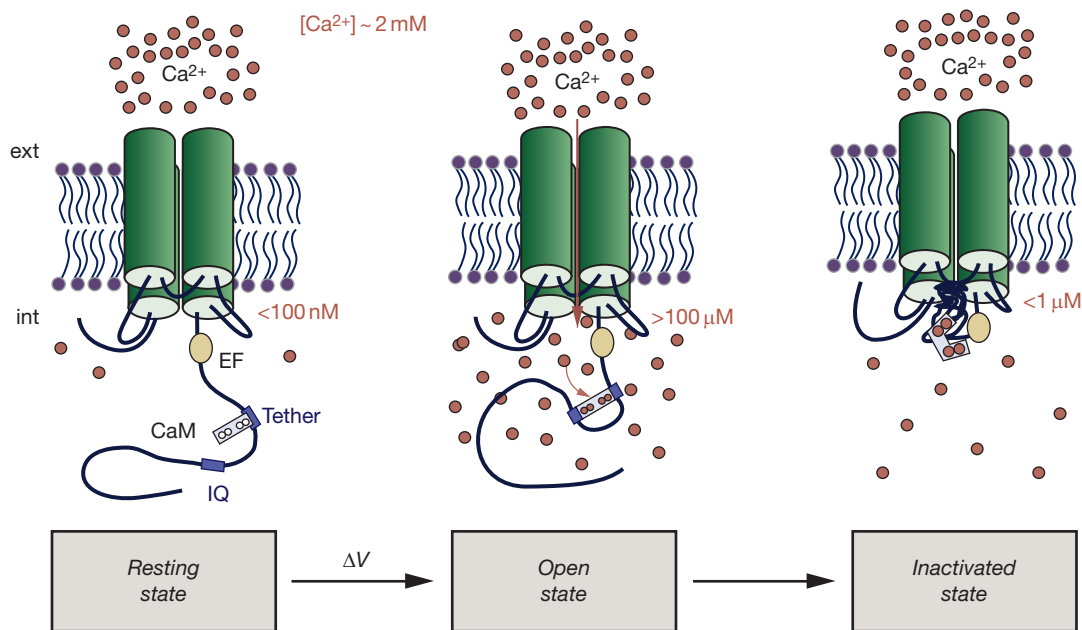


Figure 2 Model of voltage- and Ca^{2+} -dependent gating of a Ca^{2+} channel. The model shows a schematic folding of the $\alpha 1$ -subunit in Figure 1. At membrane resting potential the pore is closed (resting state); upon depolarization (ΔV) the pore opens and Ca^{2+} ions move across the membrane down their electrochemical gradient (open state); when the intracellular Ca^{2+} -concentration increases, Ca^{2+} binds to tethered calmodulin (CaM) which then attaches itself to the regulatory IQ-motif and promotes closure of the pore (inactivated state). Kindly prepared by Dr. R. D. Zuehlke, Bern.

Other drugs acting on these channels are the phenylalkylamines verapamil and the benzothiazepine diltiazem. They are grouped together as Ca-channel blockers and are therapeutically used for the treatment of cardiovascular diseases, notably hypertension and angina pectoris. The toxins have no, or very limited, therapeutic use. In some cases of otherwise untreatable pain, the N-type channel blocker ω -conotoxin GVIA has been successfully applied. Because of unwanted side effects, the T-type channel blocker mibefradil has been withdrawn from therapeutic use, although it was quite effective in the treatment of essential hypertension.

Channelopathies

Diseases resulting from mutations of genes encoding calcium channels are beginning to be identified. Dominantly inherited night blindness seems to result from a mutation of the gene encoding the retinal Ca^{2+} channel $\alpha 1\text{F}$. Several neurological disorders (e.g., familial hemiplegic migraine and episodic ataxias) are associated with mutations in the $\alpha 1\text{A}$ gene. This field of research is rapidly evolving and will hopefully lead to better therapeutic possibilities.

See also: **Bioenergetics:** Calcium Oscillations; Calcium Transport in Mitochondria; Calcium/Calmodulin-Dependent Protein Kinase II; Intracellular Calcium Channels: NAADP⁺-Modulated; **Lipids Carbohydrates Membranes and Membrane Proteins:** Ion Channel Protein Superfamily.

Further Reading

- Catterall WA (2000) Structure and regulation of voltage-gated Ca^{2+} channels. *Annual Review of Cell and Development Biology* 16: 521–555.
- Hille B (2001) *Ionic Channels of Excitable Membranes*. Sunderland, MA: Sinauer.
- McDonald TF, Pelzer S, Trautwein W, and Pelzer DJ (1994) Regulation and modulation of calcium channels in cardiac, skeletal, and smooth muscle cells. *Physiological Reviews* 74: 365–507.
- Perez-Reyes E (2003) Molecular physiology of low-voltage-activated T-type calcium channels. *Physiological Reviews* 83: 117–161.
- Reuter H (1996) Diversity and function of presynaptic calcium channels in the brain. *Current Opinion in Neurobiology* 6: 331–337.
- Saimi Y and Kung C (2002) Calmodulin as an ion channel subunit. *Annual Review Of Physiology* 64: 289–311.
- Zamponi G (ed.) (2005) *Voltage-Gated Calcium Channels*. New York: Kluwer Academic/Plenum.

Voltage-Gated Sodium Channels: Structure, Function, and Pathophysiology

W A Catterall, University of Washington, Seattle, WA, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Action potential The electrical signal conducted by nerve and muscle cells.

Gating charge A charged amino acid residue in a voltage-gated ion channel whose movement causes gating current and initiates a conformational change that opens or closes the pore.

Gating current A capacitive electrical current caused by movement of charged amino acid residues (the gating charges) in transmembrane segment S4 of voltage-gated ion channels across the membrane.

Ion channels Membrane proteins that form a pore through which ions can cross the membrane.

Ion channelopathy A disease caused by mutation of an ion channel.

Membrane potential The voltage across the plasma membrane of the cell caused by the combination of ion gradients generated by ion pumps and selective ion permeability generated by ion channels.

Voltage sensor The structure within a voltage-gated ion channel that senses and responds to the membrane potential.

Voltage-gated sodium channels initiate action potentials in neurons and other excitable cells, and they are responsible for propagation of action potentials along nerve and muscle fibers. They are complexes of a large pore-forming α -subunit and smaller β -subunits. Multiple genes encode sodium channel subunits, and the distinct sodium channel subtypes have subtle differences in functional properties, differential expression in excitable cells, and differential distribution in subcellular compartments. These differences in function and localization contribute to the specialized functional roles of sodium channels in neuronal physiology and pharmacology. Mutations that alter sodium channel expression and/or function cause several inherited diseases.

Sodium Channel Subunit Structure

Na^+ channel proteins in mammalian brain are complexes of a 260-kD α -subunit in association with one or two auxiliary β -subunits ($\beta 1$ – $\beta 4$) of 33–36 kD in size. The primary sequence predicts that the sodium channel α -subunit folds into four internally repeated domains (I–IV), each of which contains six α -helical transmembrane segments (S1–S6) (Figure 1). In each domain, the S1 through S4 segments serve as the voltage-sensing module, and the S5 and S6 segments and the reentrant P loop between them that folds into the transmembrane region of the channel between them serve as the pore-forming module. A large extracellular loop connects the S5 or S6 transmembrane segments to the P loop in each domain, whereas the other extracellular loops are small. Large intracellular loops link the four homologous domains, and the large N-terminal and C-terminal domains also contribute substantially to the mass of the internal face of sodium channels. This view of sodium channel architecture originally derived from hydrophobicity analysis has been largely confirmed by biochemical, electrophysiological, and structural experiments.

Sodium Channel Genes

Sodium channels are the founding members of the ion channel superfamily that includes voltage-gated Ca^{2+} channels, transient receptor potential (TRP) channels, voltage-gated, inward rectifying, and two-pore-domain K^+ channels, and cyclic nucleotide-regulated (CNG) and hyperpolarization-activated, cyclic nucleotide-gated (HCN) channels. In evolution, the four-domain sodium channel was last among the voltage-gated ion channels to appear, and it is only found in multicellular organisms. It is thought that sodium channels evolved by two rounds of gene duplication from ancestral single-domain bacterial sodium channels like NaChBac found in *Bacillus halodurans*. Voltage-gated sodium channel genes are present in a wide variety of metazoan species including fly, leech, squid, and jellyfish. The biophysical properties, pharmacology, gene organization, and even intron-splice sites of these invertebrate sodium channels are largely similar to the mammalian sodium channels, adding further support for the idea that the primordial sodium channel was established before the evolution of vertebrates.

The human, mouse, and rat sodium channel genes and proteins are the best characterized to date. Ten related sodium channel genes are found in vertebrates (Figure 2). Genes encoding sodium channels $\text{Na}_v 1.1$, $\text{Na}_v 1.2$, $\text{Na}_v 1.3$, and $\text{Na}_v 1.7$ are localized on chromosome 2 in human and mouse, and these channels share similarities in sequence, biophysical characteristics, block by nanomolar concentrations of tetrodotoxin (TTX), and broad expression in neurons. A second cluster of genes encoding $\text{Na}_v 1.5$, $\text{Na}_v 1.8$, and $\text{Na}_v 1.9$ channels is localized to human chromosome 3p21–24 and to chromosome 3 in mouse. Although they are more than 75% identical in sequence to the group of channels on chromosome 2, these sodium channels all contain amino acid substitutions that confer varying degrees of resistance to the pore blocker TTX. In $\text{Na}_v 1.5$, the principal cardiac isoform, a single amino acid change from phenylalanine to cysteine in the pore region of domain I is responsible for 200-fold reduction in TTX

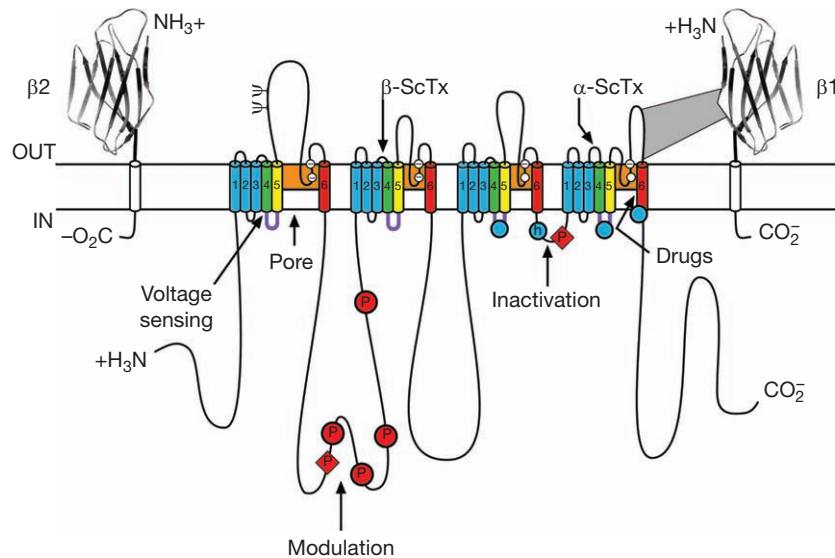


Figure 1 Transmembrane organization of sodium channel subunits. The primary structures of the subunits of the voltage-gated ion channels are illustrated as transmembrane folding diagrams. Cylinders represent probable alpha helical segments. Bold lines represent the polypeptide chains of each subunit with length approximately proportional to the number of amino acid residues in the brain sodium channel subtypes. The extracellular protein domains of the $\beta 1$ and $\beta 2$ subunits are shown as immunoglobulin-like folds. Ψ , sites of probable N-linked glycosylation; P, sites of demonstrated phosphorylation by PKA (circles) and PKC (diamonds); shaded, pore-lining S5-P-S6 segments; white circles, the outer (EEDD) and inner (DEKA) rings of amino residues that form the ion selectivity filter and the tetrodotoxin binding site; ++, S4 voltage sensors; h in blue circle, inactivation particle in the inactivation gate loop; blue circles, sites implicated in forming the inactivation gate receptor. Sites of binding of α - and β -scorpion toxins and a site of interaction between α and $\beta 1$ subunits are also shown. The colors in the transmembrane segments are correlated with the 3D model in Figure 3. Blue and green, the voltage-sensing module composed of S1–S4; yellow, orange, and red, the pore-forming module composed of S5, P, and S6.

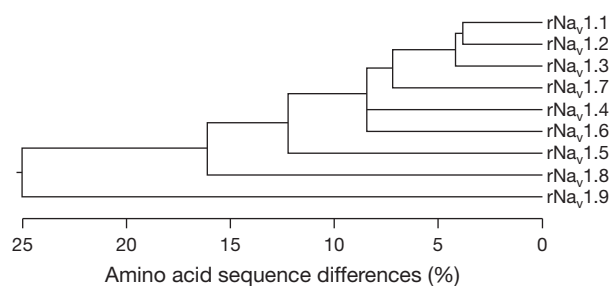


Figure 2 Amino acid sequence similarity and phylogenetic relationships of voltage-gated sodium channel α subunits. Phylogenetic relationships of rat sodium channel sequences $\text{Na}_v1.1$ – $\text{Na}_v1.9$ and Na_x . To perform the analysis, the amino acid sequences for all of the isoforms were aligned and identical residues identified using Clustal W. Divergent portions of the terminal regions and the cytoplasmic loops between domains I–II and II–III were excluded from the analysis. The tree was rooted by including the invertebrate sodium channel sequences.

sensitivity compared to those channels on chromosome 2. At the identical position in $\text{Na}_v1.8$ and $\text{Na}_v1.9$, the amino acid residue is serine, and this change results in even greater resistance to TTX. These two channels are preferentially expressed in peripheral sensory neurons. In comparison to the sodium channels on chromosomes 2 and 3, $\text{Na}_v1.4$ which is expressed in skeletal muscle and $\text{Na}_v1.6$ which is highly abundant in central nervous system (CNS) have greater than 85% sequence identity and similar functional properties, including TTX sensitivity in the nanomolar concentration range. $\text{Na}_v1.4$ and $\text{Na}_v1.6$ are located on human chromosome

11 (mouse chromosome 17) and human chromosome 15 (mouse chromosome 12), respectively.

A tenth sodium channel, Na_x , whose gene is located near the sodium channels of chromosome 2, is evolutionarily more distant. Key differences in functionally important regions of voltage sensor and inactivation gate and lack of functional expression of voltage-gated sodium currents in heterologous cells suggest that Na_x may not function as a voltage-dependent sodium channel. Consistent with this conclusion, targeted deletion of the Na_x gene in mice causes functional deficits in sensing plasma salt levels but not in electrical excitability.

The $\text{Na}_v\beta$ subunits are encoded by four distinct genes in mammals. The $\beta 1$ and $\beta 3$ subunits associate noncovalently with the α subunits, whereas $\beta 2$ and $\beta 4$ are covalently linked by a disulfide bond. All four β subunits have a single transmembrane segment, a large N-terminal extracellular domain, and a small intracellular C-terminal domain (Figure 1). The $\text{Na}_v\beta$ subunits are dual-function proteins. Co-expression of $\text{Na}_v\beta$ subunits alters the kinetics and voltage dependence of activation and inactivation of sodium channel complexes. Moreover, their extracellular domain forms an immunoglobulin-like fold, similar to many cell-adhesion molecules, and the β subunits are involved in interactions of sodium channels with proteins in the extracellular matrix, with other cell-adhesion molecules, and with intracellular regulatory proteins like protein kinases and phosphoprotein phosphatases. The importance of the functions of the $\text{Na}_v\beta$ subunits is underscored by the strong effects of deletion of the genes encoding $\text{Na}_v\beta 1$ and $\text{Na}_v\beta 2$, which include impairments of myelination and axonal conduction, epileptic seizures, and premature death.

Expression and Localization of Sodium Channel Subtypes

The sodium channel subtypes are differentially expressed in tissues and differentially localized within individual cells. $\text{Na}_v1.3$ is highly expressed in fetal nervous tissues, whereas $\text{Na}_v1.1$, $\text{Na}_v1.2$, and $\text{Na}_v1.6$ are abundant in juvenile and adult CNS. Generally, $\text{Na}_v1.1$ and $\text{Na}_v1.3$ are primarily localized in neuronal cell bodies where they may control the neuronal excitability through integration of synaptic impulses to set the threshold for action potential generation and propagation to the dendritic and axonal compartments. $\text{Na}_v1.2$ is primarily localized in unmyelinated axons. During development, $\text{Na}_v1.6$ replaces $\text{Na}_v1.2$ in maturing nodes of Ranvier in myelinated axons. $\text{Na}_v1.1$, $\text{Na}_v1.2$, and $\text{Na}_v1.6$ are also expressed in the peripheral nervous system. However, the three isoforms that have been cloned from sympathetic and dorsal root ganglion neurons, $\text{Na}_v1.7$, $\text{Na}_v1.8$, and $\text{Na}_v1.9$ are more abundant and are the principal sodium channels in the peripheral nervous system. The $\text{Na}_v1.7$ channel is localized in axons, where it functions in initiating and conducting the action potential. More restricted expression patterns are observed for $\text{Na}_v1.8$ and $\text{Na}_v1.9$; these channels are highly expressed in small sensory neurons of dorsal root ganglia and trigeminal ganglia where they have a key role in nociception. $\text{Na}_v1.4$ and $\text{Na}_v1.5$ are the primary muscle sodium channels that control the excitability of the skeletal and cardiac myocytes, respectively. $\text{Na}_v1.5$ is transiently expressed in developing skeletal muscle but is replaced by $\text{Na}_v1.4$ in the adult.

Molecular Basis of Sodium Channel Function

Classical work by Hodgkin and Huxley defined the three key features of sodium channels: (1) voltage-dependent activation, (2) rapid inactivation, and (3) selective ion conductance. Building on this foundation, recent structure–function studies employing molecular, biochemical, structural, and electrophysiological techniques have provided clear understanding of the molecular basis of sodium channel function. The narrow outer pore is formed by the re-entrant P loops between transmembrane segments S5 and S6 of each domain (**Figure 1**). Mutation of a set of four residues in analogous positions in each domain (aspartate in domain I, glutamate in domain II, lysine in domain III, and alanine in domain IV, DEKA) to glutamates confers calcium selectivity, indicating that the side chains of these amino acid residues are likely to interact with sodium ions as they are conducted through the ion selectivity filter of the pore.

Similar to other voltage-gated ion channels, the voltage dependence of activation of sodium channels derives from outward movement of gating charges as a consequence of depolarization of the membrane and reduction of membrane electric field. The S4 segments of each homologous domain serve as the primary voltage sensors for activation. They contain repeated motifs of a positively charged amino acid residue followed by two hydrophobic residues creating a transmembrane spiral of positive charges. Upon depolarization, outward movement and rotation of S4 is thought to initiate a conformational change that opens the sodium channel pore. This sliding helix model is

supported by strong evidence. For example, neutralization of the positively charged residues in S4 reduce the voltage dependence of gating. The outward and rotational gating movement of S4 segment has been detected directly by reaction of substituted cysteine residues in S4 segments with extracellular sulfhydryl reagents following channel activation and by analysis of the movement of fluorescent probes incorporated into these substituted cysteine residues. A structural model of the voltage-sensing and pore-forming modules of sodium channels in the resting and activated states is illustrated in **Figure 3**, and movies of the gating movements of the voltage sensor and pore modules during activation and opening of the sodium channel are presented online (**Movies 1 and 2**; see section **Relevant Website**).

Fast inactivation of the sodium channel is a critical process that occurs within milliseconds of channel opening. The generally accepted model of this process involves a conserved inactivation gate formed by the intracellular loop connecting domains III and IV, which serves as a hinged lid that binds to the intracellular end of the pore and blocks it (**Figure 1** and see below). Intracellular perfusion of proteases or intracellular application of antibodies directed to this loop prevents fast inactivation. The latch of this fast inactivation gate is formed by three key hydrophobic residues, IFM, and an adjacent threonine (T). Mutations of these amino acid residues destabilize the inactivated state, and peptides harboring the IFM motif can restore inactivation to sodium channels having mutated inactivation gates. The closed inactivation gate is thought to make multiple interactions with hydrophobic amino acid residues near the intracellular mouth of the pore that may constitute the inactivation gate receptor. Scanning mutagenesis experiments implicate hydrophobic residues in intracellular S4–S5 loops of domains III and IV, as well as the intracellular end of the S6 transmembrane segment in domain IV in forming the inactivation gate receptor.

Three-Dimensional Structure of Sodium Channels

The three-dimensional (3D) structure of the sodium channel is not yet completely known, but some limited structural information has emerged. Structural determination at 19 Å resolution by cryo-electron microscopy and image reconstruction techniques shows sodium channels as having a bell shape from a side view, a four-fold symmetry of transmembrane masses as viewed from above the membrane, and large intracellular and extracellular masses through which several inlets and outlets allow aqueous access (**Figure 3(a)**). Unexpectedly, this image of sodium channels reveals a central pore that does not directly connect the intracellular and extracellular spaces, instead splitting into four branches near the membrane surface. Moreover, it suggests the presence of four peripherally located narrow transmembrane pores, which may function as gating pores for voltage-sensor movement as described below.

The only high-resolution structure of sodium channels is the solution structure of the inactivation gate in the intracellular loop connecting domains III and IV, as determined by nuclear magnetic resonance (NMR) analysis of this intracellular loop expressed in bacteria (**Figure 3(b)**). This structural analysis reveals a rigid alpha helical structure, preceded by

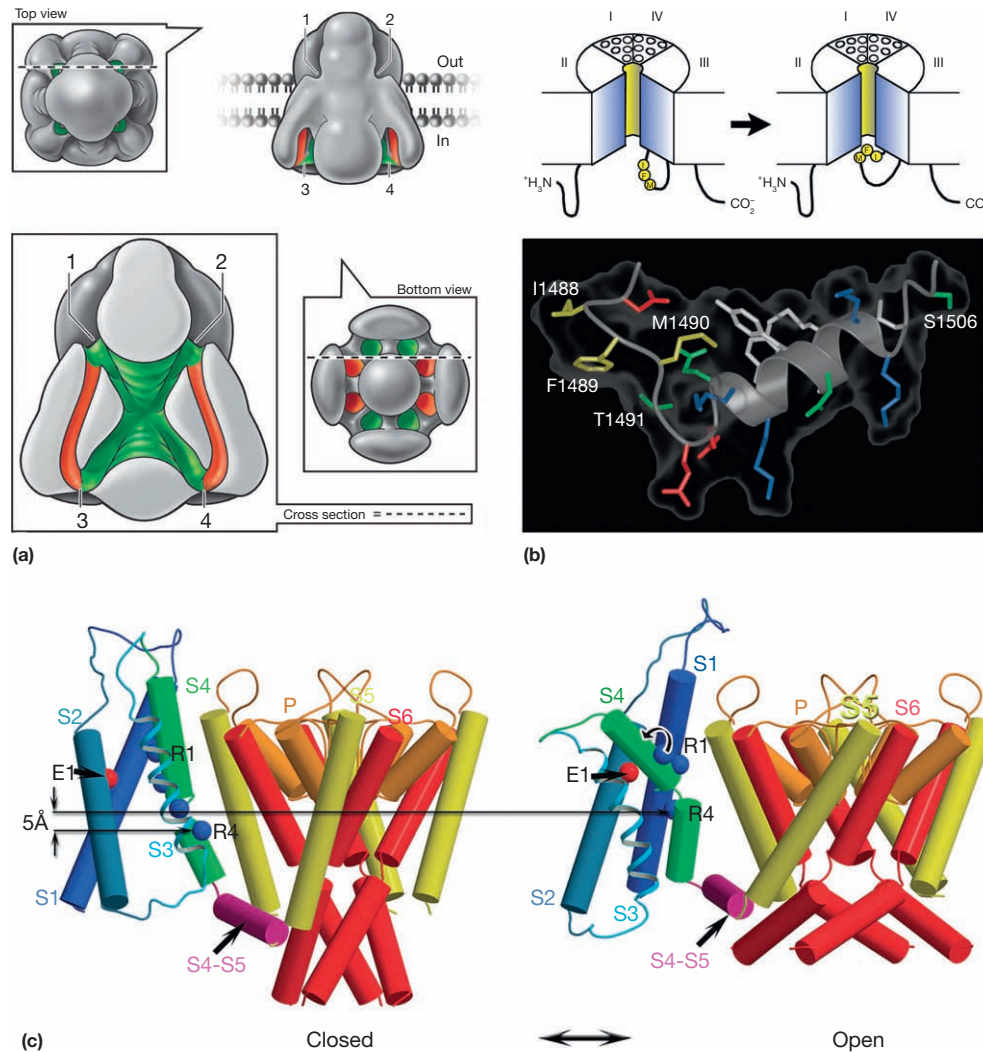


Figure 3 3D structure of sodium channels and a related potassium channel. (a) The 3D structure of the Na_v channel α subunit at 20 Å resolution, compiled from electron micrograph reconstructions. (b) (Top) The hinged-lid mechanism. The intracellular loop connecting domains III and IV of the sodium channel is depicted as forming a hinged lid with the critical phenylalanine (Phe1489) within the IFM motif shown occluding the mouth of the pore during the inactivation process. The circles represent the transmembrane helices. (Bottom) 3D structure of the central segment of the inactivation gate as determined by multi-dimensional NMR. Isoleucine 1488, phenylalanine 1489, and methionine 1490 (IFM) are illustrated in yellow. Threonine 1491, which is important for inactivation, and serine 1506, which is a site of phosphorylation and modulation by protein kinase C, are also indicated. (c) K_v1.2 models of the closed and open states shown in cylinder representation. Only a single voltage-sensing module is shown attached to the tetramer of the pore-forming module for clarity. Transmembrane segments S1 through S6, and P-loop colored by rainbow scheme from blue to red. The S4–S5 linker colored by purple color. Approximate positions of the alpha carbon atoms of the first and fourth gating charge carrying arginines in the S4 (labeled as R1 and R4 and colored in blue) and E226 in the S2 (labeled as E1 and colored in red) shown in sphere representation. All intracellular and extracellular loops, except for the S4–S5 linker, are represented by curved lines for simplicity. Vertical translation of the R4 between the closed and open states along the membrane normal vector and relative to the plane of the membrane indicated by arrows. The S3 represented by ribbon to show clearly the positions of the gating-charge-carrying arginines in the S4 segment. (a) Adapted from Sato C, Ueno Y, Asai K, et al. (2001) The voltage-sensitive sodium channel is a bell-shaped molecule with several cavities. *Nature* 409: 1047–1051. (b) Adapted from Catterall WA (2000) From ionic currents to molecular mechanisms: The structure and function of voltage-gated sodium channels. *Neuron* 26: 13–25.

two turns that position the key hydrophobic sequence motif IFMT such that it can interact with and block the inner mouth of the pore.

The global fold of the sodium channel pore region is predicted to be similar to that of voltage-gated potassium channels (Figure 3(c)). The structure of the open state of this channel has been determined by X-ray crystallography, and the structure of the closed state has been determined by structural

modeling using the Rosetta method. This structural view shows the S5 and S6 region of sodium channel arrayed in four-fold symmetry around the aqueous ion-conduction pathway (Figure 3(c), red, tan, and yellow). The selectivity filter of the pore at its narrow extracellular end is formed by the P loops (tan), whereas the wider vestibule in the center of the pore is a narrow pore formed mostly by the S6 segments (red). A single voltage-sensor module is illustrated for simplicity (Figure 3(c),

blue and green). In the resting (closed) state of the channel, the S4 segment is in an inward position. Activation of the voltage sensor by depolarization causes outward and rotational movement of the S4 segment and complementary movement of the S1, S2, and S3 segments around it (see [Movies 1](#) and [2](#) at the PNAS website; see section [Relevant Website](#)). These coupled molecular movements translocate the gating charges in the S4 segment from exposure to the intracellular milieu to exposure to the extracellular milieu, as required for voltage-dependent gating.

Comparison of the primary structures of the auxiliary β subunits to those of other proteins reveals a close structural relationship to the family of proteins that contain immunoglobulin-like folds, which include many cell-adhesion molecules. The extracellular domains of these type-I single-membrane-spanning proteins are predicted to fold in a similar manner as myelin protein P0, whose immunoglobulin-like fold is known to be formed by a sandwich of two beta sheets held together by hydrophobic interactions ([Figure 1](#)). Myelin P0 protein is a cell-adhesion molecule involved in tight wrapping of myelin sheets, and many related cell-adhesion molecules with extracellular Ig-folds and a single membrane-spanning segment are involved in cell–cell interactions among neurons and glia. As expected from their structure, $\text{Na}_v\beta$ subunits interact with extracellular matrix molecules, other cell-adhesion molecules, and intracellular cytoskeletal proteins and signaling proteins. These interactions are thought to localize and stabilize sodium channels in specific subcellular compartments such as axon hillocks and nodes of Ranvier where they are present in high density and to bring crucial signaling molecules to the sodium channel to regulate it.

Sodium Channel Pharmacology

Sodium channels are the molecular targets for drugs used in prevention of acute pain in dentistry and surgery and in treatment of cardiac arrhythmias, epilepsy, and the manic phase of bipolar disorder. Sodium channel-blocking drugs are also in development for treatment of chronic pain. Local anesthetics bind to a specific receptor within the pore of sodium channels, formed by the S6 segments in domains I, III, and IV ([Figure 4](#)). Their binding blocks ion movement through the pore and stabilizes the inactivated state of sodium channels. Antiarrhythmic drugs and anti-epileptic drugs share similar, overlapping receptor sites. Complete block of sodium channels would be lethal. However, these drugs selectively block sodium channels in depolarized and/or rapidly firing cells, such as axons carrying high-intensity pain information and rapidly firing nerve and cardiac muscle cells that drive epileptic seizures or cardiac arrhythmias. This selective block arises because the drugs can reach their binding site in the pore of the sodium channel more rapidly when the pore is repetitively opened and they bind with high affinity to inactivated sodium channels that are generated in rapidly firing or depolarized cells. This use-dependent action of the sodium channel-blocking drugs is essential for their therapeutic efficacy.

In addition to these important pharmaceutical agents, sodium channels are the molecular targets for a large number of neurotoxins that paralyze prey by preventing neuromuscular function ([Table 1](#)). These toxins act at five or more distinct

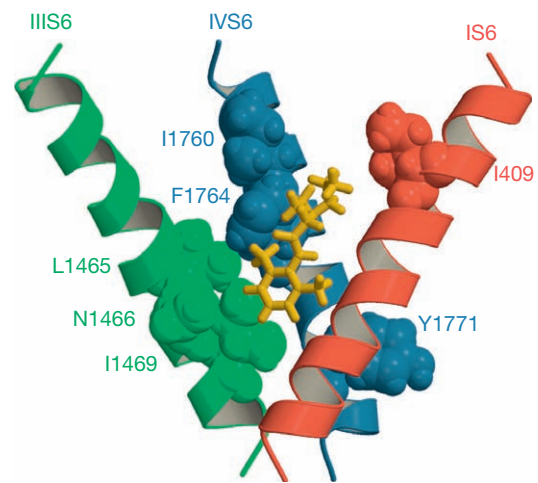


Figure 4 Model of etidocaine binding to the local anesthetic receptor site formed by transmembrane segments IS6, IIIS6, and IVS6 of the $\text{Na}_v1.2$ channel. 3D model of proposed orientation of amino acid residues within the Na^+ channel pore with respect to the local anesthetic etidocaine. Only transmembrane segments IS6 (red), IIIS6 (green), and IVS6 (blue) are shown. Residues important for etidocaine binding are shown in space-filling representation.

Table 1 Receptor sites on sodium channels

Receptor site	Toxin or drug	Location (domain/segment)
Neurotoxin receptor site 1	Tetrodotoxin Saxitoxin	I/P, II/P, III/P, IV/P I/P, II/P, III/P, IV/P
Neurotoxin receptor site 2	μ -Conotoxin Veratridine Batrachotoxin Grayanotoxin	I/S6, IV/S6
Neurotoxin receptor site 3	α -Scorpion toxins Sea anemone toxins	I/S5–IS6, IV/S1–S2, IV/S3–S4 IV/S3–S4
Neurotoxin receptor site 4	β -Scorpion toxins	II/S1–S2, II/S3–S4
Neurotoxin receptor site 5	Brevetoxins	I/S6, IV/S5
Neurotoxin receptor site 6	Ciguatoxins δ -Conotoxins	IV/S3–S4
Local anesthetic receptor site	Local anesthetic drugs Antiarrhythmic drugs Antiepileptic drugs	I/S6, III/S6, IV/S6

receptor sites and either block the pore of the channel or alter the kinetics or voltage dependence of gating. The pore blockers TTX and saxitoxin bind to neurotoxin receptor site 1, which is formed by the P loops in the four domains. Alkaloids and other organic toxins from plants, frogs, and other species bind at neurotoxin receptor site 2 and cause persistent activation of sodium channels, which depolarizes nerves and muscles and causes repetitive firing and paralysis. Polypeptide toxins from scorpions and spiders bind at neurotoxin receptors sites 3 and 4 and either slow inactivation or enhance activation of sodium channels, which also causes repetitive firing and paralysis. Brevetoxins and ciguatoxins, which are large polyether

Table 2 Major inherited diseases of sodium channels

Disease	Gene	Channel
Generalized epilepsy with febrile seizures plus	SCN1A	Na _v 1.1
Severe myoclonic epilepsy in infancy	SCN1A	Na _v 1.1
Familial hemiplegic migraine type III	SCN1A	Na _v 1.1
Benign familial neonatal-infantile seizures	SCN2A	Na _v 1.2
Hypokalemic periodic paralysis type II	SCN4A	Na _v 1.4
K ⁺ -sensitive normokalemic periodic paralysis	SCN4A	Na _v 1.4
Hyperkalemic periodic paralysis	SCN4A	Na _v 1.4
Paramyotonia congenita	SCN4A	Na _v 1.4
Long QT syndrome type III	SCN5A	Na _v 1.5
Brugada syndrome	SCN5A	Na _v 1.5
Erythromelalgia	SCN9A	Na _v 1.7
Paroxysmal extreme pain disorder	SCN9A	Na _v 1.7
Congenital indifference to pain	SCN9A	Na _v 1.7

compounds made by marine dinoflagellates, bind to neurotoxin receptor site 5 and cause persistent activation of sodium channels. The diversity of organisms that have developed toxins that target sodium channels is remarkable, highlighting the physiological importance of these channels in electrical signaling and neuromuscular function.

Sodium Channelopathies

Mutations that alter sodium channel function lead to human diseases of hyperexcitability (Table 2). Multiple mutations of Na_v1.4 that yield hyperactive skeletal muscle sodium channels have been shown to cause hyperkalemic periodic paralysis and paramyotonia congenita. By contrast, mutations in S4 gating charges that cause reduced function of these channels and induce a sodium leak through their mutant voltage sensor domains (termed 'gating pore current') cause hypokalemic periodic paralysis. Mutations of Na_v1.5 that impair inactivation or decrease expression and/or activation lead respectively to inherited long QT syndrome type 3 and Brugada syndrome, in which the risk of fatal cardiac arrhythmias is greatly increased.

Familial febrile seizures, generalized epilepsy with febrile seizure plus (GEFS⁺) type II, and severe myoclonic epilepsy of Infancy (SMEI) all result from mutations of Na_v1.1 channels. These mutations selectively impair electrical excitability and action potential firing in inhibitory neurons in the brain and thereby cause hyperexcitability and epilepsy by disinhibition of excitatory neuron activity. A spectrum of severity of these loss-of-function mutations causes the spectrum of severity of these disease syndromes: SMEI >> GEFS⁺ >> familial febrile seizures. By contrast, gain-of-function mutations in Na_v1.1 channels cause familial hemiplegic migraine type-II. Benign familial neonatal-infantile seizures are caused by mutations in Na_v1.2 channels, which reduce cell surface expression and decrease functional activity of these channels.

The sodium channels expressed in the peripheral nervous system also are the targets for mutations that cause inherited

disease syndromes. Mutations that cause different functional effects on Na_v1.7 channels are responsible for three distinct disease syndromes. Mutations that enhance activation of Na_v1.7 channels cause inherited erythromelalgia, characterized episodes of burning pain, erythema, and mild swelling in the hands and feet triggered by mild warmth or exercise, whereas mutations that impair fast inactivation cause paroxysmal extreme pain disorder, characterized by intense rectal pain. By contrast, loss-of-function mutations in Na_v1.7 channels cause congenital insensitivity to pain, which is characterized by complete loss of the sensation of pain without loss of response to other sensory modalities.

The physiological requirement for intact sodium channel function is further underscored by mouse knockout models, in which specific ablation of Na_v1.1, Na_v1.2, Na_v1.5, and Na_v1.6 lead to premature lethality. By contrast, mutant mice lacking Na_v1.7 or Na_v1.8 are viable and have defects in nociceptive response. These and other mouse models that target specific sodium channel genes should prove to be effective in delineating the significance of sodium channel function and its regulation in physiological contexts.

See also: Lipids Carbohydrates Membranes and Membrane Proteins: Ion Channel Protein Superfamily.

Further Reading

- Armstrong CM (1981) Sodium channel and gating currents. *Physiological Reviews* 61: 644–682.
- Cannon SC (2006) Pathomechanisms in channelopathies of skeletal muscle and brain. *Annual Review of Neuroscience* 29: 387–415.
- Catterall WA (2000) From ionic currents to molecular mechanisms: The structure and function of voltage-gated sodium channels. *Neuron* 26: 13–25.
- Catterall WA, Goldin AL, and Waxman SG (2005) International Union of Pharmacology. XLVII. Nomenclature and structure–function relationships of voltage-gated sodium channels. *Pharmacological Reviews* 57: 397–409.
- Goldin AL (2002) Evolution of voltage-gated Na⁺ channels. *Journal of Experimental Biology* 205: 575–584.
- Hille B (2001) *Ionic Channels of Excitable Membranes*, 3rd edn. Sunderland, MA: Sinauer Associates.
- Hodgkin AL and Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology* 117: 500–544.
- Sato C, Ueno Y, Asai K, et al. (2001) The voltage-sensitive sodium channel is a bell-shaped molecule with several cavities. *Nature* 409: 1047–1051.
- Tombola F, Pathak MM, and Isacoff EY (2006) How does voltage open an ion channel? *Annual Review of Cell and Developmental Biology* 22: 23–52.
- Yarov-Yarovoy V, Baker D, and Catterall WA (2006) Voltage sensor conformations in the open and closed states in ROSETTA structural models of K⁺ channels. *Proceedings of the National Academy of Sciences of the United States of America* 103: 7292–7297.

Relevant Website

<http://www.pnas.org/cgi/content/full/0602350103/DC1> – PNAS; Movies 1 and 2.

von Hippel–Lindau (VHL) Protein

R C Conaway and J W Conaway, Stowers Institute for Medical Research, Kansas City, MO, USA

© 2013 Elsevier Inc. All rights reserved.

Glossary

Dioxygenase Enzyme that catalyzes the insertion of both oxygens of molecular oxygen, O₂, into substrate molecules.

Erythropoietin An acidic glycoprotein hormone that regulates red-cell production by promoting erythroid differentiation and the initiation of hemoglobin synthesis.

Hemangioblastoma A highly vascularized tumor composed of capillaries and stromal cells.

Hemangioma A benign lesion originating from blood vessels.

Pheochromocytoma A tumor of the adrenal gland.

Retinal angioma A benign lesion found in the retina that originates from blood vessels.

Vascular endothelial growth factor A growth factor for vascular endothelial cells that promotes angiogenesis.

The von Hippel–Lindau (VHL) protein is a tumor suppressor. Mutations in the VHL protein can give rise to tumors of multiple organ systems, including the central nervous system, the endocrine system, and the kidney. The VHL protein functions as a subunit of a multiprotein ubiquitin ligase that negatively regulates expression of a large collection of hypoxia-inducible genes controlled by hypoxia-inducible transcription factors (HIFs). The VHL ubiquitin ligase prevents inappropriate expression of these hypoxia-inducible genes when cells are grown in a plentiful supply of oxygen by targeting HIFs for rapid ubiquitylation and degradation by the proteasome.

Clinical Consequences of VHL Mutations

Mutations in the *VHL* gene are found in a variety of human diseases, including VHL disease, sporadic clear cell renal carcinoma, sporadic hemangioblastoma, and congenital polycythemia. VHL disease is an autosomal dominant familial cancer syndrome. In VHL kindreds, inactivation of both copies of the *VHL* gene can give rise to a variety of highly vascularized tumors, including clear cell renal carcinoma, cerebellar hemangioblastoma and hemangioma, retinal angioma, and pheochromocytoma. *VHL* mutations are also responsible for the majority of cases of sporadic clear cell renal carcinoma and for sporadic hemangioblastoma. In addition, *VHL* mutations can also give rise to autosomal recessive familial erythrocytosis 2, a disorder characterized by elevated levels of expression of the hypoxia-inducible *erythropoietin* gene.

Function of the VHL Protein

The VHL protein is present in eukaryotes from worms to mammals, where it performs an evolutionarily conserved function as a subunit of a multiprotein ubiquitin ligase that negatively regulates expression of a large number of hypoxia-inducible genes controlled by HIFs. The expression of these hypoxia-inducible genes is repressed in normal cells grown in a plentiful supply of oxygen (normoxic conditions), but is strongly induced in cells starved for oxygen (hypoxic conditions). The VHL ubiquitin ligase prevents the inappropriate expression of these hypoxia-inducible genes when cells are grown in

normoxic conditions by targeting HIFs for rapid ubiquitylation and degradation by the proteasome. When cells are starved for oxygen, ubiquitylation of HIFs by the VHL ubiquitin ligase is blocked. Under these hypoxic conditions, cellular HIF levels rise, and HIFs enter the nucleus, bind to hypoxia response elements (HREs) in the promoters of hypoxia-inducible genes, and activate their transcription and expression.

The cellular levels of HIFs correlate well with the levels of expression of hypoxia-inducible genes in both normoxic and hypoxic conditions (Figure 1). In normal cells growing under normoxic conditions, HIFs are present in cells at low levels, and expression of hypoxic-inducible genes is barely detectable. Under hypoxic conditions, HIFs are present at relatively high levels, and expression of hypoxia-inducible genes is robust. Under both normoxic and hypoxic conditions in cells lacking a functional VHL protein, HIFs are present in cells at high levels, and maximal expression of hypoxia-inducible genes is observed. Inappropriate overexpression of hypoxia-inducible genes in cells lacking a functional VHL protein may contribute to the pathology of VHL-associated diseases. Among these are the genes encoding vascular endothelial growth factor (VEGF), transforming growth factor- α (TGF α), and erythropoietin (EPO). VEGF is a secreted protein that appears to play a major role in the vascularization of VHL tumors by promoting ingrowth into tumors of blood vessels needed to carry oxygen and nutrients to the tumor cells. TGF α is also a secreted protein that appears to function as an autocrine growth factor for some types of VHL tumor cells and may be responsible for promoting their uncontrolled proliferation. An elevated EPO level in the serum of polycythemic patients is a hallmark of the disease.

Structure of the VHL Ubiquitin Ligase

Purification of the VHL ubiquitin ligase revealed that it is composed of multiple proteins. In addition to the VHL protein, the VHL ubiquitin ligase includes the elongin B and C proteins, cullin protein family member Cul2, and RING finger protein Rbx1 (Figure 2). In the VHL ubiquitin ligase, the VHL protein functions to recruit HIFs for ubiquitylation by interacting directly with them. The elongin B and C proteins bind stably to each other to form an adaptor that links the VHL protein to a Cul2-Rbx1 heterodimeric module. The Cul2-Rbx1 module functions



XPV Polymerase and the Bypass of Ultraviolet DNA Damage

A R Lehmann, University of Sussex, Brighton, UK

© 2013 Elsevier Inc. All rights reserved.

This article is reproduced from the previous edition, volume 4, pp. 419–421, © 2004, Elsevier Inc.

Glossary

DNA polymerase Enzymes which are able to synthesize new DNA molecules using a DNA template.

Nuclear localization signal A sequence of amino acids in a protein that results in its localization in the nucleus.

Nucleotide excision repair The enzymatic pathway by which many types of DNA damage are first recognized, then

cut out of the DNA, and the excised region replaced with new DNA.

Ubiquitination A posttranslational modification of a protein in which a molecule of the 76-amino-acid polypeptide ubiquitin is covalently attached to a target protein.

Zinc finger A sequence of amino acids which forms a finger-like structure and binds a zinc atom.

Xeroderma Pigmentosum Variants

All cells sustain many kinds of DNA damage. This can be generated either endogenously from hydrolysis and oxidation reactions or exogenously from exposure to carcinogens. One of the most prevalent DNA-damaging agents is ultraviolet (UV) radiation from the sun, which generates photoproducts in the DNA of cells in exposed areas of the skin. The most important photoproducts, namely the cyclobutane pyrimidine dimer (CPD) and the pyrimidine 6-4 pyrimidone photoproduct (6-4 PP), are removed from cellular DNA by the process of nucleotide excision repair (NER). The importance of this process can be seen from the features of patients with xeroderma pigmentosum (XP). XP patients show a wide variety of skin abnormalities on sun-exposed areas including freckling, hypo- and hyperpigmentation, and ultimately multiple skin cancers. The incidence of skin cancer in XP individuals has been estimated to be 1000-fold higher than that in unaffected individuals. In the majority of XP patients, this hypersensitivity to the effects of sunlight results from a genetic deficiency in NER. In approximately 20% of XP individuals, however, NER is normal. These so-called XP variants (XPVs) are deficient in their ability to produce intact daughter DNA strands during DNA replication after exposure of the cells to UV irradiation. In other words, they are defective in their ability to replicate DNA damaged by UV light.

Pol η and the Y-Family of DNA Polymerases

Although the cellular deficiency in XPV was identified in 1975, the molecular basis for the defect was not discovered until

1999. The gene defective in XPV cells was found to encode a new DNA polymerase designated DNA polymerase η or pol η . Pol η is a protein of 713 amino acids (aa) with no sequence similarity to previously discovered DNA polymerases, but it does share sequence similarity with several other DNA polymerases that were discovered at about the same time. This new group of polymerases were designated the 'Y-family' and they are involved in translesion synthesis (TLS) past different types of damage. They show sequence conservation over a region of about 400 aa, which contains the catalytic domain and usually forms the N-terminal part of the protein. Y-family polymerases have a C-terminal extension, which is not conserved between family members.

Catalytic Activity

DNA polymerases that carry out the replication of undamaged DNA are very efficient and accurate, but they are blocked by most types of DNA damage, including both major UV photoproducts. By contrast, pol η is a rather poor polymerase on undamaged DNA. It dissociates after incorporating a few nucleotides and it is rather error prone. However, pol η , in contrast to all other eukaryotic polymerases, is able to carry out TLS past a T–T CPD with the same efficiency as past undamaged Ts. Furthermore, in most instances, the correct bases, namely two As, are inserted opposite the T's of the CPD. The reason for these properties became clear when the catalytic domains of several members of the Y-family (including yeast pol η) were crystallized and their three-dimensional (3D) structures solved. Despite the lack of sequence similarity between classical and Y-family polymerases, the 3D structures did have

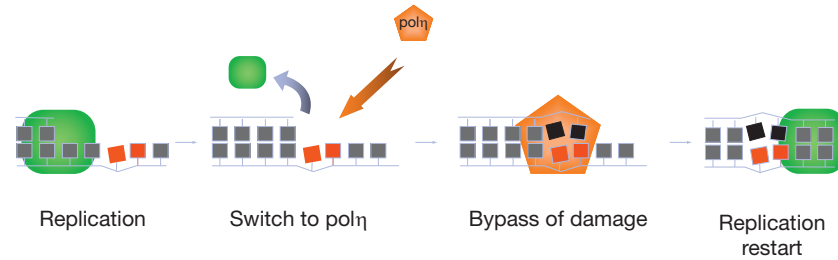


Figure 1 Translesion synthesis by pol η . Replicative DNA synthesis is blocked by a CPD (indicated by red blocks). This leads to a switch to pol η (step 1), which replicates past the CPD (step 2) and then dissociates, so that normal replication can restart (step 3).

the three subdomains, designated finger, thumb, and palm, found in all other DNA polymerases, with the Y-family having an extra subdomain designated pad or little finger. However, whereas the classical polymerases form a closed tight structure, which only permits normal bases into the active site, pol η has a much more open structure, with two important consequences. First, it has a relatively low stringency, so that incorrect bases can be incorporated rather easily. Second, the open structure is able to accommodate the CPD in the template strand and the structure ensures that, in general, the correct bases are inserted opposite the pyrimidines in the CPD. Thus, pol η is tailor-made for replicating past CPDs. It can insert the appropriate bases opposite the damage and extend from these bases, but then it will soon dissociate from the template, so that the more-accurate replicative polymerases can take over again. A simple mechanistic scheme for TLS is shown schematically in Figure 1.

Apart from the CPD, pol η is able to replicate past some other types of damage *in vitro*, but in most cases, it is much less efficient with other types of damaged templates than with undamaged or CPD-containing DNA. Recent data suggest that it may be involved in bypassing other lesions *in vivo*, such as those generated by cisplatin.

The importance of pol η is seen in the severe clinical characteristics of XPV patients and in the very high UV-induced mutation frequency in XPV cells. The UV-mutability of XPV cells is several fold higher than that of normal cells and the frequency of all types of mutations is increased. This indicates that in the absence of pol η , some other enzyme is able to function, but the replacement is less efficient and makes more mistakes.

Localization and Interactions

The catalytic domain of pol η is contained within the first 430 aa. What then is the function of the C-terminal 280 aa? DNA replication takes place during the S phase of the cell cycle in replication factories inside the cell nucleus. These factories can be visualized either by immunofluorescence with an antibody to a replication protein, or by tagging proteins involved in DNA replication with the green fluorescent protein (GFP). The factories appear as bright nuclear foci when viewed by fluorescent microscopy. Pol η localizes in replication factories during the S phase, so that it is always on hand in case the replication machinery encounters DNA damage (Figure 2). It is the C-terminal 120 aa of pol η (aa 595–713) that is responsible for this localization into replication factories. This

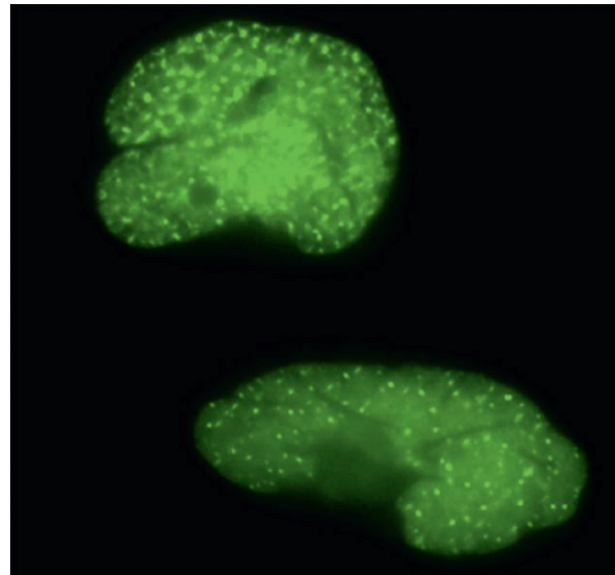


Figure 2 Localization of pol η in replication foci. Pol η , tagged with GFP, localizes into nuclear foci during DNA replication.

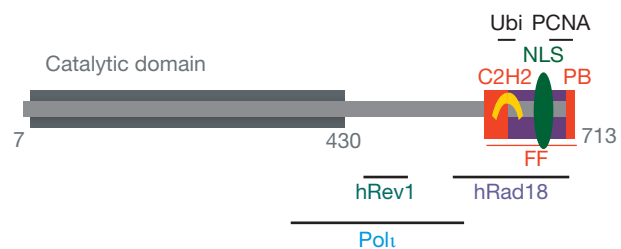


Figure 3 Domain structure of pol η . C2H2, zinc-finger UBZ motif; NLS, nuclear localization signal; PB, PCNA-binding PIP box motif; FF, foci formation; Ubi, ubiquitin. Bars represent regions involved in interactions with the indicated proteins. hRev1 and pol δ are other Y-family polymerases, hRad18 is a E3 ubiquitin ligase that mono-ubiquitinates PCNA when replication forks are stalled at sites of DNA damage.

C-terminal domain contains three important elements, all of which are needed for localization into foci and for efficient TLS. These are (1) a PIP box that is involved in binding to the replication accessory protein, proliferating cell nuclear antigen (PCNA); (2) a bipartite nuclear localization sequence that is

required for transporting pol η into the nucleus but is also part of the PCNA-interacting region; and (3) a C₂H₂ zinc finger that forms a UBZ ubiquitin-binding motif (UBZ, ubiquitin-binding zinc finger). The function of aa 430–595 in pol η has not yet been determined, but it is involved in interactions with other proteins. The domain structure and regions of interaction of pol η with other proteins are summarized in [Figure 3](#).

The interaction of pol η with the protein hRad18 is also important for localization into foci. It is thought that hRad18 is involved in transporting pol η into replication factories. The interaction of pol η with its cousin pol ι is not required for localization of pol η but it does facilitate the localization of pol ι , which is also found in replication foci in S-phase cells. The interaction with Rev1 is discussed below.

Interaction with Ubiquitinated PCNA and the Polymerase Switch

PCNA is a homotrimeric ring, which acts as a sliding clamp to increase the processivity of all DNA polymerases. When DNA replication is blocked by damage, PCNA becomes ubiquitinated. This ubiquitination increases the affinity of PCNA for Y-family polymerases, because they all have ubiquitin-binding domains in their C-terminal regions. In the case of pol η , this is the UBZ motif. Thus, all three elements in the C-terminus of pol η stimulate its binding to ubiquitinated PCNA. This facilitates its ability to displace the replicative polymerase from the template–primer and, if the damage is a CPD, to carry out TLS past the damaged bases.

Mutations in XPV Patients

Mutations have been identified in the pol η gene in some 60 patients. Most of the mutations result in truncations within the catalytic subunit and would not be expected to result in any functional pol η . This demonstrates that pol η is not an essential protein, because its absence is compatible with life. Several missense mutations within the catalytic domain have also been identified and they are predicted either to inhibit binding to DNA or to interfere with the 3D structure of the protein. A small number of mutations do not affect the catalytic domain, but truncate the protein close to the C-terminus. These mutant proteins would be expected to be fully active, but to lack the localization/interaction domain in the C-terminus, which is required for transporting the polymerase into the nucleus and for its interaction with ubiquitinated PCNA at the stalled fork. There is no clear relationship between the site or type of

mutation and the severity of the clinical features, even though the severity of features differs widely between patients.

Replication Past Other Types of UV Damage

Although the CPD is the major UV photoproduct, the 6-4 PP is also formed in substantial quantities (about one-third the frequency of the CPD). No single polymerase has been found that is able to carry out TLS past a 6-4 PP, which produces a much greater distortion in the DNA than a CPD. However, two polymerases acting in concert are able to accomplish this *in vitro*. Insertion of the nucleotides opposite the 6-4 PP can be carried out either by pol η or by its paralog, pol ι , but neither of these can extend further from the incorporated nucleotides. However, once nucleotides have been inserted opposite the damaged bases, they can be extended by DNA polymerase ζ . So far, direct evidence for this dual polymerase mechanism has only been obtained *in vitro*. It is likely that this mechanism is also used *in vivo*, and, indeed, DNA polymerase ζ is required for TLS past 6-4 PPs inside cells, probably for the extension step. As yet, it is not known which polymerase carries out the insertion step. The interaction of pol η with Rev1 (see [Figure 3](#)) has been proposed to be important for the two-stage insertion–extension mode of TLS. Rev1 interacts via a sequence in its extreme C-terminus with the other three Y-family polymerases as well as with the Rev7 subunit of pol ζ . It is suggested that Rev1 can thereby facilitate the handover from the inserting polymerase (Y-family) to the extender (pol ζ).

See also: **Molecular Biology:** DNA Polymerases: Kinetics and Mechanism; DNA Replication Fork, Eukaryotic; Sliding Clamps in DNA Replication: *Escherichia coli* β -Clamp and PCNA Structure; UmuC D Lesion Bypass DNA Polymerase V.

Further Reading

- Guo C, Kosarek-Stancel JN, Tang TS, and Friedberg EC (2009) Y-family DNA polymerases in mammalian cells. *Cellular and Molecular Life Sciences* 66: 2363–2381.
- Lehmann AR, Niimi A, Ogi T, et al. (2007) Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair (Amsterdam)* 6: 891–899.
- McCulloch SD, Kokoska RJ, Masutani C, et al. (2004) Preferential *cis*–*syn* thymine dimer bypass by DNA polymerase η occurs with biased fidelity. *Nature* 428: 97–100.
- Prakash S, Johnson RE, and Prakash L (2005) Eukaryotic translesion synthesis DNA polymerases: Specificity of structure and function. *Annual Review of Biochemistry* 74: 317–353.
- Yang W and Woodgate R (2007) What a difference a decade makes: Insights into translesion DNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America* 104: 15591–15598.

Zinc Fingers

M Isalan, Centre for Genomic Regulation (CRG), UPF, Barcelona, Spain

© 2013 Elsevier Inc. All rights reserved.

Glossary

Classical finger A zinc finger of the TFIIIA type, where a zinc ion is coordinated by two cysteines and two histidines (Cys₂-His₂ or C2H2).

Protein motif A small structural element of a protein with a characteristic fold or shape.

TFIIIA A transcription factor protein from *Xenopus laevis*, in which zinc fingers were originally discovered.

Transcription factor A protein involved in gene regulation that helps to activate or repress the expression of a particular gene.

Zinc finger A small structural protein motif consisting of peptide chain folded around a central chelated (coordinated) zinc ion.

Zinc fingers are small, compact protein subunits, with structured peptide chains folding around chelated zinc ions. Functionally, these subunits carry out a wide variety of tasks within cells by providing stable structural scaffolds and driving critical binding interactions, especially among proteins, DNA, and RNA. Zinc fingers are particularly important in gene regulation, where many proteins employ them to bind DNA in a sequence-specific manner, to activate or inhibit particular genes. Although many families of zinc fingers exist, each with a characteristic fold or structure, perhaps the best studied, is the classical (Cys₂-His₂ or C2H2) type. Classical zinc fingers are extremely versatile, as can be seen by their abundance in nature: over 700 proteins contain such domains in the human genome alone, making them the second most common human protein motif. Zinc fingers may be seen as convenient molecular building blocks, often used in clusters, whose modular nature allows them to be conveniently duplicated, altered, and shuffled by evolution. In a biotechnology context, zinc fingers may be re-engineered to have novel binding properties with numerous applications in both research and medicine.

Zinc Finger Structures and Families

Classical Zinc Fingers

History

The zinc finger motif was discovered in 1985 by Miller, McLachlan, and Klug in a protein responsible for controlling gene expression – the *Xenopus* transcription factor IIIA (TFIIIA). Originally, nine novel consecutive repeats were found, each ~30 amino acids in length. Since every repeat had a conserved pair of cysteines and histidines at defined positions, Klug proposed that

these residues could chelate a zinc ion, stabilizing an independent structural domain, the ‘zinc finger’. Notably, the term finger was derived from the finger-like appearance of the sequences when drawn schematically around a zinc ion, as in [Figure 1\(a\)](#). The repeats found in TFIIIA are now used as a definition of the classical zinc finger, where individual fingers conform to the archetypal consensus sequence shown in [Figure 2](#).

Secondary structure and hydrophobic core

Structural studies revealed that the classical Cys₂-His₂ finger is a simple and compact unit based on an antiparallel β -hairpin packed against an α -helix ([Figure 1\(b\)](#)). In addition to chelating zinc, classical fingers are stabilized at the core by a hydrophobic cluster of residues, converging from three fixed positions (see [Figures 1](#) and [2](#)). Moreover, this hydrophobic triad is highly conserved, often consisting of tyrosine, phenylalanine, and leucine. By contrast, zinc fingers may withstand extensive amino acid substitutions at other positions, without significantly altering the overall structure.

The role of zinc

Small protein domains are inherently unstable and, even though the $\beta\beta\alpha$ fold found in zinc fingers can form around hydrophobic cores alone, the stability of the fold is dramatically improved by zinc chelation. Such stability cannot be obtained using the disulfide bridge – a common extracellular alternative – because of the highly reducing environment inside cells. Consequently, intracellular proteins have solved the problem of stabilizing small domains by utilizing metal-binding sites. Zinc is ideally suited for this purpose as it has a single oxidation state, a fixed and distinguishable ionic radius, and can accommodate both nitrogen and sulfur ligands.

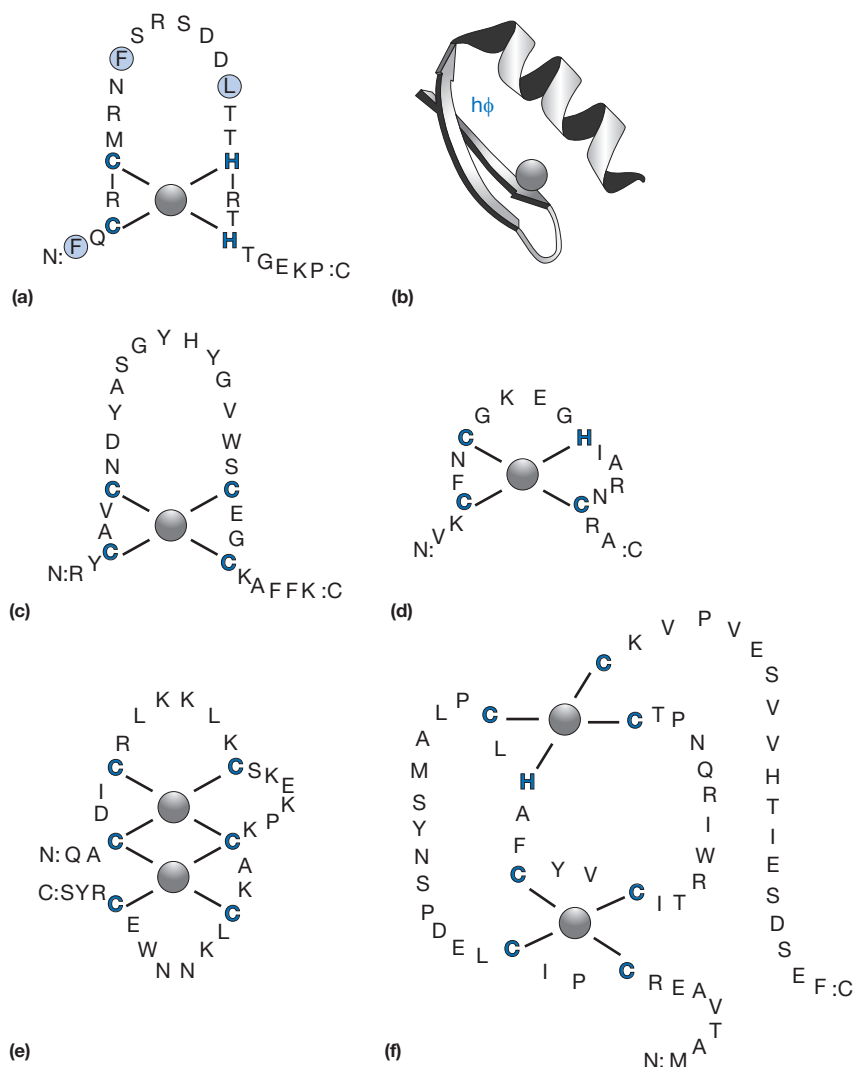


Figure 1 Examples of zinc finger families and folds. (a) A classical Cys₂-His₂ zinc finger is drawn schematically as a string of amino acids (1-letter abbreviations), with cysteine (C) and histidine (H) pairs chelating the central zinc ion. Three circled residues (F, F, L) come together in the structure to form a hydrophobic core. The finger shown is Finger 2 from the mouse transcription factor Zif268. (b) Representation of the actual $\beta\alpha$ structure of a classical Cys₂-His₂ zinc finger. Note the position of the hydrophobic core ($h\phi$). (c) A Cys₄ steroid-receptor zinc finger from the estrogen receptor. Note that this protein actually contains two contiguous Cys₄ zinc finger repeats, forming a single structural unit. (d) A Cys₃-His zinc finger from a viral nucleocapsid protein. (e) A yeast GAL4 domain. (f) The ring finger fold.

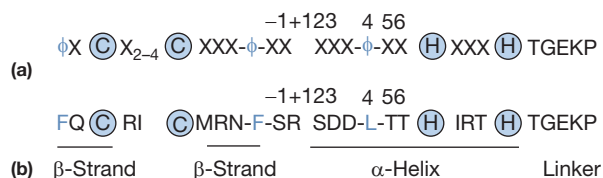


Figure 2 Zinc finger amino-acid sequences. (a) Consensus sequence of a classical Cys₂-His₂ zinc finger, where X is any amino acid and ϕ represents hydrophobic residues (typically Y, F, and L). Zinc-chelating cysteine and histidine residues are circled. α -Helical positions (numbered relative to position +1) are indicated above the sequence. (b) The sequence of Finger 2 of Zif268, showing conformity with the consensus sequence.

Evolution

Zinc fingers may have evolved originally from peptides, which maintained a simple $\beta\alpha$ structure (similar to that shown in Figure 1(b)) independently of zinc. In any case, the $\beta\alpha$ fold is one of the smallest possible units of protein tertiary structure, and may therefore be one of the oldest. Zinc fingers are abundant in nature and the huge variety found in various organisms, from yeast to human, may in part be due to their modular nature. In this regard, not only are zinc fingers independently folding structural building blocks, but they also are often encoded by single exons. Indeed, exon shuffling and gene duplication may account for why zinc fingers have spread so widely during evolution.

Genomic abundance

The recent initial drafts of the human genome have revealed the extent to which classical zinc fingers are used by cells: there are 706 genes containing one or more classical zinc fingers in humans, totaling several thousand fingers, and making them the second most common human protein family. It is remarkable that only immunoglobulin domains are more abundant in humans, and indeed only slightly more so, being found in 765 genes. By contrast, the *Drosophila melanogaster* fly has 357 classical zinc finger genes, actually making these the most common protein family in this organism. Despite the fact that zinc finger usage varies from organism to organism, these motifs are clearly a widespread feature in all eukaryotic cells.

Nonclassical Zinc Fingers

Although zinc-binding proteins have been known for many years, the usual definition of zinc fingers maintains that they are distinct from other zinc-binding protein subregions by virtue of being autonomously folding, functional mini-domains. Following the discovery of the classical Cys₂-His₂ zinc finger, many other classes of zinc finger have emerged. Although these form a diverse collection, often with radically different folding topologies, these proteins are often referred to in the literature as zinc fingers. Therefore, some well-studied examples are outlined here below.

Cys₄ zinc fingers

One of the best-characterized nonclassical zinc fingers is the Cys₄ type, including the steroid-hormone receptor family. The estrogen receptor (Figure 1(c)) was the first protein of this type to be elucidated structurally and the gene contains two Cys₄ zinc finger repeats, forming a single structural unit that is quite unlike the independent classical zinc fingers of TFIIIA. Functionally, the steroid receptors form dimers, which bind hormones such as estrogen, and then mediate activation of gene expression through sequence-specific DNA-binding interactions in the promoter regions of target genes. Further examples of proteins containing Cys₄ finger motifs include other nuclear hormone receptors, such as the glucocorticoid and retinoic acid receptors. Many Cys₄ proteins function as transcription factors, such as the elongation factor TFIIIS and the adenoviral E1A transactivator.

Cys₃-His zinc fingers

The Cys₃-His zinc finger is another variant defined by its characteristic combination of cysteine and histidine residues. For example, the nucleocapsid protein, found in certain viruses such as HIV, contains Cys₃-His fingers, with various nucleic-acid-binding properties that facilitate hybridization and the viral life cycle (Figure 1(d)). Other examples of Cys₃-His proteins include MetRS and the first finger in the LIM domain.

Multi-zinc proteins

Moving beyond the simpler zinc finger domains, which fold around a single zinc ion, there exist larger more complex folds, each employing two or more zinc ions. Examples include the yeast GAL4 domain (Figure 1(e)) and the ring finger (Figure 1(f)). Note that although these larger domains are

not quite as prevalent as the single-zinc modules, they nonetheless form quite abundant families. For example, the ring finger family ranks as the 15th most abundant domain in the human genome, with 210 members, while another fold, the PHD finger, ranks 34th with 84 members.

Zinc Finger Functions

Interactions with DNA and RNA

Since their discovery, zinc fingers have been associated with binding nucleic acids, namely, DNA and RNA. Indeed, the archetypal zinc finger protein, TFIIIA, is bifunctional, having distinct modes of binding which allow it to bind both to the DNA internal control region of 5S RNA genes and to the transcribed RNA gene product. In this sense, zinc fingers are master gene regulators, with the potential to influence both transcription and translation.

Structure of zinc finger bound to DNA

The majority of our knowledge of zinc finger functions has focused on their ability to bind specific DNA sequences. The first structure of zinc fingers bound to DNA, that of the DNA-binding domain of the mouse transcription factor Zif268, was solved by Carl Pabo and colleagues in 1991. The structure revealed that each of the three classical zinc fingers occupies the major groove, wrapping around the DNA for almost one turn of the double helix (Figure 3(a)). The base-specific contacts occur primarily from consecutive turns of the zinc finger α -helices to the guanine-rich strand of the DNA.

DNA-recognition helix

Essentially, classical zinc fingers use the α -helix as a reading head to recognize specific DNA sequences. Contacts from amino acids consist primarily of electrostatic interactions and hydrogen bonds, although hydrophobic interactions also play a role. Conventionally, the amino acids in the α -helix are numbered relative to a defined first position – position +1 (see Figures 2 and 3). Since positions –1, +3, and +6 are all to be found on the external face of the helix, these can form a binding surface that is especially useful in DNA recognition because the spacing of the residues matches the register of the base pairs. It should be noted, however, that contacts from other positions, especially position +2, are also important in DNA recognition and that not all zinc finger–DNA interactions are as regular as the canonical structure of Zif268.

DNA-recognition code

The most striking feature of the canonical Zif268 structure is the presence of regularly spaced one-to-one contacts between amino acids and DNA bases, from the same fixed helical positions in each finger (Figure 3(b)). By studying mutations at these positions, it has emerged that classical zinc fingers are a special case in terms of protein–DNA interactions. Unlike for most DNA-binding proteins, these positions do seem to fit a residue-base code-like pattern. Positions –1, +2, +3, and +6 in the α -helix are particularly important (Figure 3(b)), and some major residue-base associations are outlined in Figure 3(c).

As a note of caution, it is important to mention that the code is not absolute, and is certainly an oversimplification

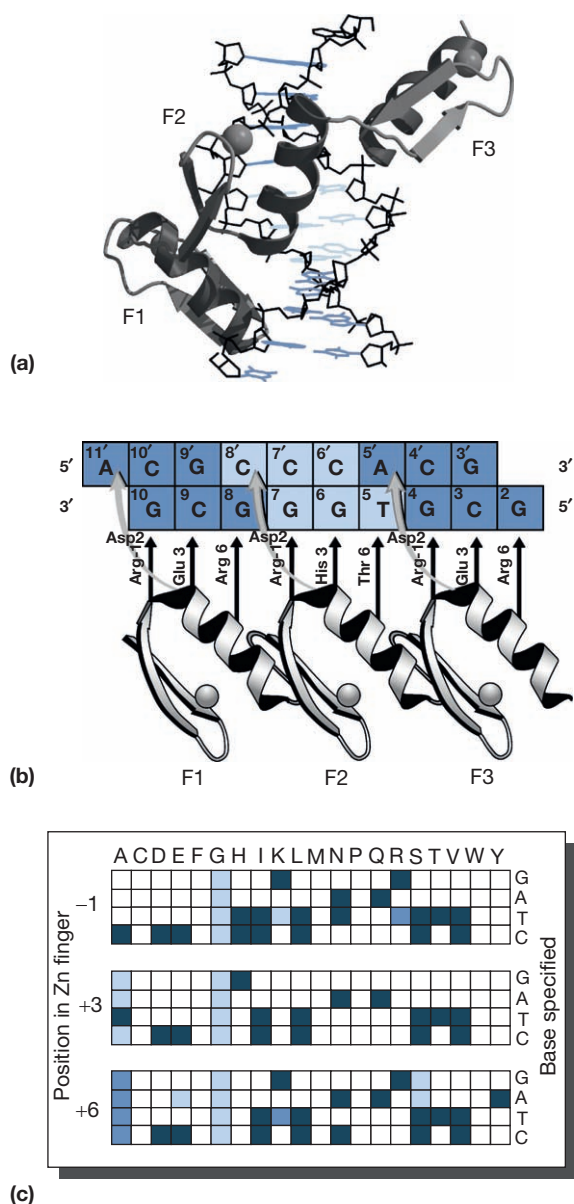


Figure 3 Interactions between zinc fingers and DNA. (a) Side view of Zif268 binding DNA. The three zinc fingers (F1–F3) bind the major groove of the DNA (b) Schematic diagram of a simple model of recognition between the three zinc fingers of Zif268 and the triplet sub-sites of an optimized DNA-binding site. Arrows indicate contacts by recognition residues with bases in each DNA strand. Note that most contacts are to the G-rich strand. Since the N-terminal finger contacts the 3'-end of the DNA and the C-terminal finger the 5'-end, binding is said to be antiparallel. (c) Zinc finger DNA-recognition code. The grid summarizes common associations between amino acid residues (A–Y) and the bases they face in a classical zinc finger DNA-binding site. Darker shading of boxes indicates stronger associations. Note that contacts from overlapping C-terminal zinc fingers (especially from position 2) can significantly alter the coding specificity shown here for position +6. (a) From Pavletich NP and Pabo CO (1991) Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252: 809–817.

of the interplay between different residues in the protein and between factors in the DNA structure, such as base stacking and unwinding. Nonetheless, thinking about zinc finger DNA interactions in this way can give a valuable approximation into the possible binding modes of a given zinc finger protein.

RNA recognition

Unlike the double-helical scaffold of DNA, RNA has much more complex folds, with every RNA presenting a unique binding surface. Nonetheless, zinc fingers have the versatility to accommodate such structures, with contacts from all over the finger framework. Like DNA recognition, electrostatic interactions and hydrogen bonding play a major role, with amino acids contacting either RNA bases or the phosphodiester backbone. Note that in RNA recognition, there is a greater potential for hydrophobic interactions, with aromatic residues such as Tyr, Phe, and Trp stacking around extruding bases.

Non-Nucleic Acid Interactions

Protein–protein interactions

Zinc fingers are not limited to controlling gene expression, but play an important cellular role in mediating specific protein–protein interactions as well. Whereas nucleic acid recognition usually requires clusters of fingers, cytoplasmic proteins often contain only one or two fingers. This is possibly because a single protein-binding finger has sufficient affinity to bind effectively, given the more variable binding surface presented by other proteins.

Lipid binding

Recently, the known binding repertoire of zinc fingers has been extended to include lipid recognition. This is exemplified by the FYVE zinc finger family, which is found in a wide range of eukaryotic cells. These nonclassical fingers have a complex structure consisting of multiple loops, two twin-stranded β -sheets, and an α -helix stabilized by a pair of zinc ions. Functionally, the domains bind to phosphatidylinositol 3-phosphate (PI3P) in cellular membranes, commonly found on endosomes, and mediate further recruitment of a range of proteins.

Zinc Fingers in Biotechnology

Classical zinc fingers are the ideal scaffold for protein engineering. They are small and versatile, withstanding extensive mutations without destroying their basic structure. Moreover, individual finger modules may be linked in tandem repeats to allow recognition of extended target surfaces such as DNA grooves. Novel DNA-binding zinc fingers have been routinely engineered for over 15 years and are beginning to demonstrate their utility in biotechnical applications. By fusing with suitable effector domains, effective transcription activators and repressors can be synthesized that recognize specific DNA sequences, typically 9–18 bp long. These constructs can upregulate or downregulate specific genes in a cell and have been used in a wide variety of contexts, including antiviral agents and angiogenesis-promoting factors. Zinc finger nucleases are the latest technological application and have been used to cut and repair genes inside cells and organisms.

See also: **Bioenergetics:** Indicators of Intracellular Calcium; **Metabolism Vitamins and Hormones:** Vitamin E; **Molecular Biology:** DNA Mismatch Repair and the DNA Damage Response; Gene Expression in Bacterial Systems: The *trp* Operon and Attenuation; Nonhomologous Recombination: Bacterial Transposons; **Protein/Enzyme Structure Function and Degradation:** Aminopeptidases; Calpain; **Signaling:** Cyclic GMP Phosphodiesterases; T-Cell Antigen Receptor.

Further Reading

- Beerli RR and Barbas CF III (2002) Engineering polydactyl zinc-finger transcription factors. *Nature Biotechnology* 20: 135–141.
- Berg JM and Shi Y (1996) The galvanization of biology: A growing appreciation for the roles of zinc. *Science* 271: 1081–1085.

- Carroll D (2008) Progress and prospects: Zinc-finger nucleases as gene therapy agents. *Gene Therapy* 15: 1463–1468.
- Choo Y and Isalan M (2000) Advances in zinc finger engineering. *Current Opinion in Structural Biology* 10: 411–416.
- Klug A and Schwabe JW (1995) Protein motifs 5. Zinc fingers. *FASEB Journal* 9: 597–604.
- Lander ES, Linton LM, Birren B, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.
- Mackay JP and Crossley M (1998) Zinc fingers are sticking together. *Trends in Biochemical Sciences* 23: 1–4.
- Pabo CO, Peisach E, and Grant RA (2001) Design and selection of novel Cys₂His₂ zinc finger proteins. *Annual Review in Biochemistry* 70: 313–340.

Relevant Website

<http://www.zincfingers.org> – The Zinc Finger Consortium.

INDEX

This index is in letter-by-letter order, whereby hyphens and spaces within index headings are ignored in the alphabetization, and it is arranged in set-out style, with a maximum of four levels of heading. Location references refer to the volume number, in bold, followed by the page number. Page numbers suffixed by *f* or *t* refer to figures or tables respectively.

A

- AAA (ATPases associated with various cellular activities) proteins, 1:1–6
- AAA+ superfamily, 1:3–5, 1:3*t*, 2:21, 2:123–124
- basic concepts, 1:1
- cellular functions
 - membrane-bound metalloproteases, 1:6
 - membrane fusion, 1:5–6
 - proteasome-dependent degradation, 1:5
 - vesicle transport, 1:5–6
- classifications, 1:3–5, 1:3*t*, 1:4*f*
- DNA polymerase III holoenzyme, 2:95
- dyneins, 2:184–187
 - axonemal dyneins
 - binding characteristics, 2:185*t*, 2:186–187
 - inner arm dyneins, 2:185*t*, 2:187
 - outer arm dyneins, 2:185*t*, 2:187
 - characteristics and classifications, 2:184, 2:185*t*
 - cytoplasmic dyneins
 - cytoplasmic dynein 1, 2:185–186, 2:185*t*, 2:186*f*
 - cytoplasmic dynein 2 (IFT dynein), 2:185*t*, 2:186
 - dynein heavy chain, 2:184–185, 2:185*t*, 2:186*f*
 - dynein intermediate chain, 2:185*t*, 2:186*f*
 - dynein light chain, 2:185*t*, 2:186*f*
 - research summary, 1:6
 - ring-shaped helicases, 2:21, 2:22*t*, 2:23*f*
 - structure and functions, 1:1–2, 1:2*f*, 1:3*f*
- α/β -hydrolase domain-containing 5 (ABHD5), 1:49–51, 1:50*f*
- AarA protein, 4:51*f*, 4:54
- ABCD1 gene, 3:415–416
- Acacia senegal*, 2:466
- Acanthamoeba castellanii*, 1:12, 1:37*f*, 3:551
- acetaldehyde dehydrogenase, 4:74*f*
- acetate
 - acetate kinase, 4:74*f*
 - acetate thiokinase, 4:74*f*
 - biomass-produced renewable hydrogen, 4:75*f*
 - fermentation reactions, 2:204–206, 2:205*t*, 2:206*f*
- Acetivibrio cellulolyticus*, 4:73–74
- acetoacetate
 - anaplerotic pathways, 1:106*f*
 - cerebral metabolism, 1:371–373
 - starvation, 4:298–299, 4:302, 4:302*f*, 4:303–304, 4:305
- Acetobacterium woodii*, 1:487*t*
- acetogenesis
 - bacterial reductive reactions, 1:487*t*
 - cobalamin, 1:159–160
- acetone, 4:302, 4:302*f*
- acetylcholine, 1:131, 1:423–424, 2:596, 2:719–720, 2:720*f*, 3:215
- acetylcholine carbachol, 2:497*t*
- acetyl-coenzyme A (CoA)
 - acetyl-coenzyme A carboxylase (ACC)
 - body weight regulation, 4:57*f*, 4:58*f*
 - carbohydrate-responsive element
 - binding protein (ChREBP)
 - function, 1:376–377, 1:377*f*
 - cellular fatty acid metabolism regulation, 4:56
 - fatty acid biosynthesis, 2:275–279, 2:276*f*, 2:286
 - insulin metabolism, 2:606–607
 - ischemic heart diseases, 3:68, 3:69*f*, 3:70*f*
 - lipogenic gene expression, 2:287*f*
 - liver pyruvate kinase (LPK) catalysis, 1:374
- anaplerotic pathways, 1:106*f*, 1:108*f*, 1:109*f*
- branched-chain amino acids (BCAAs), 1:244, 1:245*f*
- cardiac metabolism, 3:65–66, 3:66*f*, 3:68*f*
- citric acid cycle, 4:93*f*
- dark fermentation, 4:74*f*
- fatty acid biosynthesis
 - energy metabolism regulation, 2:286
 - fatty acid synthesis pathway, 2:275–279, 2:276*f*
 - lipogenic gene expression, 2:287*f*
 - pyruvate dehydrogenase regulation, 2:285–286
- gluconeogenesis, 2:386*f*
- historical research, 1:539
- ketone body synthesis, 4:298–299
- liver pyruvate kinase (LPK) catalysis, 1:374
- methyl transfer reactions, 1:160
- mitochondrial β -oxidation, 1:386–387
- mitochondrial calcium (Ca^{2+}) transport, 3:124*f*
- mitochondrial DNA (mtDNA), 1:210
- pyruvate dehydrogenases (PDHs), 1:207*f*, 3:124*f*, 4:321, 4:324
- thiamine diphosphate (ThDP), 1:207*f*
- tricarboxylic acid cycle, 4:439*f*
- urea cycle, 4:495*f*
- N*-acetyl cysteine, 2:333
- N*-acetylgalactosamine (O-GalNAc), 2:453, 3:654
- N*-acetylglucosamine (GlcNAc)
 - β 1,2-*N*-acetylglucosaminyltransferase (β 2GnT), 2:454*f*
 - β 1,3-*N*-acetylglucosaminyltransferase (β 3GnT), 2:453, 2:454*f*
 - classifications, 2:453
 - free oligosaccharide (fOS) generation, 2:329, 2:330*f*
 - Glc₃Man₉GlcNAc₂ glycan, 3:257, 3:258*f*, 3:259*t*
 - glycosylphosphatidylinositol (GPI)-anchored proteins, 2:476, 2:477*f*
 - hexosamine biosynthetic pathway, 3:351*f*
 - Man₅GlcNAc₂ glycan, 2:329, 2:330, 2:330*f*, 3:259*t*
 - Man₈GlcNAc₂ glycan, 3:258*f*, 3:259*t*
 - Man₉GlcNAc₂ glycan, 3:259*t*
 - N*-linked glycosylation, 2:470, 2:470*f*, 2:471*t*, 4:204–205
 - O-GlcNAc biosynthesis, 3:350–353
 - background information, 3:350
 - O-GlcNAcase, 3:351, 3:353*f*
 - O-GlcNAc transferase (OGT), 3:350–351, 3:352*f*
 - physiological functions, 3:352–353
 - post-translational modifications, 3:350

- N*-acetylglucosamine (GlcNAc) (*Continued*)
 O-linked glycosylation, 2:466f
 phosphotransferase system (PTS), 2:387
 structural characteristics, 3:654
 UDP-GlcNAc biosynthesis, 3:350, 3:351f
- N*-acetylglutamate, 1:95
- N*-acetylglutamate (NAG) synthetase,
 4:494–495, 4:495f, 4:495t, 4:497–498
- N*-acetylglutamic acid (NAG), 4:495f
- N*-acetylglucosamine, 2:701–703
- N*-acetylneuraminic acid (Neu5Ac),
 4:230–231
- acid ceramidase (aCDase), 4:288
- Acidianus*, 3:34
- Acidianus ambivalens*, 4:92–93
- Acidobacteria*, 2:161
- acid sphingomyelinase (aSMase), 4:287
- aconitase, 4:439f, 4:439t
- acquired immunodeficiency syndrome (AIDS)
 aspartic proteases, 1:139
 research background, 2:101
 RNA editing, 4:157
- actin, 1:12–17, 1:18–26
 actin depolymerizing factor (ADF)/cofilin (ADF-H) domain
 adseverin, 1:23–24
 CapG, 1:23
 characteristics, 1:20–21
 fragmin, 1:23
 gelsolin, 1:28–29, 1:22f, 1:23–24, 1:277–278, 1:277f, 1:278f
 profilin, 1:27–28, 1:28f, 1:29–31, 1:30f
 scinderin, 1:23–24
 severin, 1:23, 3:438–439
 single-domain ADF/cofilin family, 1:21–22, 1:22f
 six-ADF-H-domain proteins, 1:23–24
 three-ADF-H-domain proteins, 1:23
 two-ADF-H-domain twinfilin family, 1:22–23
 villin, 1:23–24, 1:277–278
- actin filament formation
 actin cytoskeleton, 1:430–431, 1:431f
 assembly/disassembly, 1:12–14, 1:13f
 cell migration, 1:433f
 formation mechanisms, 1:18–19, 1:19f
- actin-related proteins (Arps), 1:36–41
 Arp1 protein, 1:37f, 1:38t, 1:39, 1:39f, 2:175, 2:175f
 Arp2 protein, 1:37f, 1:38t
 Arp3 protein, 1:37f, 1:38t
 Arp4 protein, 1:37f, 1:38t, 1:40
 Arp5 protein, 1:37f, 1:38t, 1:40
 Arp6 protein, 1:37f, 1:38t, 1:40
 Arp7 protein, 1:37f, 1:38t, 1:40
 Arp8 protein, 1:37f, 1:38t, 1:40
 Arp9 protein, 1:37f, 1:38t, 1:40
 Arp10 protein, 1:37f, 1:38t
 Arp11 protein, 1:37f, 2:175, 2:175f
 Arp2/3 complex
 actin organization, 1:29–31, 1:30f
 downstream targets, 2:237f, 2:238, 4:113, 4:114f
 eukaryotic Arps, 1:37f, 1:38t, 1:39–40, 1:39f
 leading-edge dynamics, 1:29, 1:20, 1:21f
 phagocytosis, 3:438–439
 prokaryotic Arps, 1:39f
 chromatin-associated Arps, 1:40
 chromatin-associated conventional actin, 1:40
 domain structures, 1:36–37, 1:39f
 eukaryotic Arps
 Arp1 protein, 1:39, 1:39f
 Arp4 protein, 1:40
 Arp5 protein, 1:40
 Arp6 protein, 1:40
 Arp7 protein, 1:40
 Arp8 protein, 1:40
 Arp9 protein, 1:40
 Arp2/3 complex, 1:39–40, 1:39f
 chromatin-associated Arps, 1:40
 chromatin-associated conventional actin, 1:40
 classifications, 1:36, 1:37f, 1:38t
 conventional actin, 1:38–40, 1:39f
 cytoskeletal Arps, 1:38–40
 prokaryotic Arps, 1:36, 1:37–38, 1:39f
- assembly/disassembly
 actin-binding proteins, 1:14, 1:14f, 1:16f
 ADF/cofilin, 1:14f, 1:15, 1:27–28, 1:28f
 basic concepts, 1:12–14, 1:13f
 capping proteins, 1:14f, 1:15
 cellular aspects, 1:15–16, 1:16f
 polymerization conditions, 1:14
 profilin, 1:14f, 1:15
 sequestering proteins, 1:14, 1:14f
 small molecule reagents, 1:16–17, 1:16t
- cadherin signaling, 1:263f
- capping proteins
 actin-filament dynamics, 1:30f, 1:31–32, 1:32f, 1:21f
 assembly/disassembly, 1:14f, 1:15
 barbed-end-capping proteins
 actin depolymerizing factor (ADF)/cofilin (ADF-H) domain, 1:20–21
 actin organization, 1:28f, 1:29–31, 1:30f, 1:31f
 CapZ, 1:31–32, 1:32f, 1:21f, 1:22f, 1:25, 2:174, 2:175f
 tensin, 1:25
 characteristics, 1:25–26
 phagocytosis, 3:438–439
- pointed-end-capping proteins
 actin organization, 1:28f, 1:29–31, 1:30f, 1:31f
 leimodin, 1:21f, 1:25
 spectrin/band 3, 1:34–35, 1:34f
 spectrin/band 4.1, 1:34–35, 1:34f, 1:26
 tropomodulin, 1:22f, 1:25
 Wiskott–Aldrich syndrome protein (WASP) family, 1:20, 1:21f, 1:26, 3:438–439
- characteristics and functional role, 1:27
 chemoattractant signaling, 2:236–238, 2:237f
 conventional actin, 1:38–40, 1:39f
 cyclase-associated protein (CAP), 3:438–439
- cytokinesis, 4:214
- cytoskeleton dynamics
 bipolar thick filaments (BTfs), 1:623–624
 contractile networks, 1:624–625, 1:625f
 focal adhesions, 2:320, 2:320f
 phospholipase D (PLD), 3:475–476
 regulation mechanisms, 1:623–624, 1:623f
 Rho guanosine triphosphatases (Rho GTPases), 1:430–431, 1:431f, 4:111, 4:248
 schematic diagram, 1:432f
 small guanosine phosphatases (GTPases), 4:248
- dynactin subfamilies, 2:175
- enabled/vasodilator-stimulated phosphoprotein (Ena/VASP), 1:28–29, 1:29f, 1:19–20
- endocytosis, 2:198f, 2:199
- F-actin, 1:18, 3:219–221, 3:221f, 3:438–439, 4:216, 4:217f
- Fascin, 1:29f, 3:243, 3:244f
- filament structures, 1:12, 1:13f
- focal adhesions, 2:320, 2:320f
- functional role, 1:12, 1:18–19
- G-actin, 1:18
- Golgi apparatus, 4:208, 4:208f
- G protein-coupled receptor kinases (GRKs), 2:485t
- integrin adhesion network, 2:320f
- isoforms, 1:12, 1:13t
- leading-edge dynamics, 1:28f, 1:29, 1:19–20, 1:21f
- mechano-chemical characteristics, 3:221f
- microtubule-actin cross-linking factor (MACF), 3:105f, 3:107
- monomer structures, 1:12, 1:13f
- myosin kinetic cycle, 3:219–221, 3:221f
- neuronal growth cone, 1:19–20, 1:20f
- organization, 1:27–35
 characteristics and functional role, 1:27
 contraction
 basic concepts, 1:29–31, 1:30f, 1:31f
 nonmuscle cells, 1:33–34, 1:33f
 smooth muscle contraction, 1:32–33
 striated muscle contraction, 1:31–32, 1:32f
- protrusion
 actin depolymerizing factor (ADF)/cofilin (ADF-H) domain, 1:27–28, 1:28f
 actin filament bundles, 1:28–29, 1:29f
 basic concepts, 1:27–28, 1:28f
 branched networks, 1:29
 filopodia elongation, 1:28–29, 1:29f
- shape determination
 actin–spectrin membrane network, 1:34–35, 1:34f
 basic concepts, 1:34
 elastic networks, 1:34–35, 1:34f
 stiff bundles, 1:35
- phagocytosis, 3:438–439, 3:438f
- pheromone receptors, 3:443f
- plectin, 2:625–627

- pointed-end binding and branching, 1:28f, 1:29, 1:20, 1:21f
- SCAR/WAVE complex, 3:438f
- severing proteins
- actin depolymerizing factor (ADF)/cofilin (ADF-H) domain, 1:28–29, 1:29f, 1:20–21
 - actin-interacting protein 1 (Aip1), 1:22f, 1:24, 3:438–439
 - coronins, 1:20f, 1:24, 3:438f
 - cyclase-associated protein (CAP), 1:24, 3:438–439
 - formins, 1:29–31, 1:30f, 1:21f, 1:24
 - molecules interacting with CasL (Micals), 1:24
- actinin, 1:31–32, 1:32f, 1:319t, 1:623, 2:485t, 2:610–611
- Actinobacteria*
- Dam methylation, 2:43
 - proteasome-dependent degradation, 1:5
 - ribosomal RNA (rRNA), 3:559
 - type I DNA topoisomerase, 2:161
- Action in Diabetes and Vascular Disease (ADVANCE), 3:423
- Action to Control Cardiovascular Risk in Diabetes (ACCORD), 3:423
- activated protein C (APC), 3:586f
- activation function 1/2 (AF-1/2), 4:540–541, 4:542f, 4:543f
- activation transcription factor 6 (ATF6), 4:51f, 4:52
- activator protein 1 (AP-1) transcription factor, 3:411f, 4:398
- actinin, 1:183–184
- actomyosin, 2:236–238, 2:237f
- acute intermittent porphyria, 2:542t
- acute myelocytic leukemia (AML), 1:276–277
- acute promyelocytic leukemia (APL), 1:276–277
- acyl-coenzyme A
- acyl-coenzyme A dehydrogenase, 1:385f, 1:386, 1:386–387, 2:14
 - acyl-coenzyme A oxidase (AOX), 2:287f
 - acyl-coenzyme A synthetase (ACS), 2:279
 - acyl-S-coenzyme A (CoA) derivatives, 1:539–540, 1:540f
 - carnitine transesterification, 1:387, 1:385
 - CoA thioesters, 1:539, 2:290
 - fatty acid oxidation
 - cardiac metabolism, 3:66–67, 3:66f, 3:69f, 3:70f
 - cellular uptake and activation, 2:281
 - malonyl-CoA, 4:56
 - mitochondria, 2:282f
 - peroxisomes, 2:283–284, 3:414–415
 - unsaturated fatty acids, 2:282–283, 2:283f
 - fatty acyl-coenzyme A (CoA) synthetases, 2:290–295
 - ACSF proteins, 2:294
 - activation reactions, 2:290, 2:291f
 - amino acid sequence motifs, 2:290–291, 2:291f, 2:294–295
 - Bubblegum ACS family (ACSBG), 2:294
 - characteristics and occurrence, 2:290–294
 - human ACS families, 2:290–294, 2:292t
 - long-chain acyl-CoA synthetase (LCAS), 1:385f, 2:292–293
 - medium-chain acyl-CoA synthetase (MCAS), 2:292
 - nomenclature, 2:292, 2:292t
 - peroxisome proliferator-activated receptor alpha (PPAR α), 3:419–421, 3:420f, 3:422f
 - research summary, 2:295
 - short-chain acyl-CoA synthetase (SCAS), 2:292
 - structure–function relationship, 2:291f, 2:294–295
 - very long-chain acyl-CoA synthetase (VLCAS), 2:293–294
- flavins, 2:311
- gluconeogenesis, 2:381, 2:382f
- insulin resistance, 2:607
- lipogenic gene expression, 2:287f
- acyl-protein thioesterase 1 (APT1), 3:646
- adalimumab, 3:431
- ADAM family metalloproteinases, 4:199–200
- adaptilins, 1:483f
- adducin, 1:34–35, 1:34f
- adducted thumb-clubfoot syndrome, 2:471t
- adenine
- adenine nucleotides, 2:497t
 - adenine nucleotide translocator (ANT), 3:163
 - DNA adenine methylation identification (Dam-ID)
 - general discussion, 2:749
 - nuclear structure and dynamics, 2:750
 - DNA alkylation sites and repair mechanisms, 2:10t
 - oxidized adenine lesions, 2:78, 2:78f
 - riboswitches, 4:137
 - UvrABCD endonuclease, 3:330t
- adenosine
- adenosine deaminases
 - enzyme inhibitors, 2:214f
 - RNA editing, 4:155–156, 4:156f
 - DARPP-32, 4:222f, 4:223
 - molecular structure, 2:310f
 - phage libraries, 3:432–433, 3:433t, 3:434t
- adenosine diphosphate (ADP)
- actin filament formation, 1:12–14, 1:13f, 1:14f, 1:18–19, 1:19f, 1:21f
 - adenosine diphosphate (ADP) phosphorylation, 1:491, 2:204, 2:205f, 4:482
 - adenosine diphosphate (ADP) ribose (ADPribose), 1:173f, 1:174f, 1:176–177, 1:176f, 1:177f
 - adenosine diphosphate ribosylation factors (ARF), 1:134–136
 - action mechanisms, 1:134–135, 1:135f
 - Arf-related proteins, 1:134
 - molecular characteristics, 1:134
 - phospholipase D (PLD), 3:472–473, 3:474–475, 3:474f
 - regulation mechanisms, 1:135–136, 1:135f
 - small guanosine phosphatases (GTPases), 4:242, 4:243f
- adenosine thiamine triphosphate (AThTP) biosynthesis, 1:205f, 1:206
- adenosine triphosphate (ATP) synthesis, 4:482
- anti-mitochondrial antibodies (AMAs), 3:117
- carboxyl methylation, 3:607
- cardiac metabolism, 3:65–66, 3:66f, 3:68f
- citric acid cycle, 4:93f
- cyclic adenosine diphosphate (ADP)-ribose (cADPr)
- calcium ion (Ca²⁺) signaling, 1:331–336
 - basic concepts, 1:331–332
 - calcium ion (Ca²⁺) stores, 1:332
 - fertilization-induced signaling, 1:333–334, 1:333f
 - research summary, 1:336
 - sea urchin egg fertilization, 1:333–334, 1:333f
 - signaling mechanisms, 1:332–333
 - starfish egg fertilization, 1:334–335, 1:334f, 1:335f
 - invertebrates, 2:637–638
 - metabolic pathways, 2:638–639
- cytokinins, 1:627–628
- DNA-strand transferases, 4:37
- glucokinase, 2:546–547, 2:546f
- gluconeogenesis, 2:382f, 2:385f
- glycogen synthesis, 2:425, 2:426f, 2:430f
- glycolysis, 2:443, 2:444f, 2:445
- hypoxia inducible factor (HIF), 4:67
- kinesins, 2:672, 2:673f
- mitochondria
- mitochondrial ATP-dependent potassium ion (K⁺) channel (KATP), 3:551
 - mitochondrial calcium (Ca²⁺) transport, 3:118, 3:119f
 - mitochondrial CK (MtCK) isoenzyme, 3:82–84, 3:83f
 - mitochondrial membranes, 4:482
 - plant mitochondrial electron transport chain, 1:145, 1:146f
 - regulation mechanisms, 3:169–170
- myosin, 3:219–221, 3:221f
- P2Y receptors, 3:375–378
- domain structures, 3:376
 - extracellular nucleotide signaling, 3:375–376
 - functional classification, 3:376–377, 3:376t
 - functional role, 3:375
 - lysophosphatidic acid (LPA) receptors, 2:763t
 - pharmacological classification, 3:377
 - physiological functions, 3:377–378
- phosphofructokinase-2/fructose biphosphatase-2 (PFK-2/FBPase-2), 3:458f
- photosynthetic metabolic pathways, 3:498f
- pyruvate dehydrogenases (PDHs), 4:324
- respiration control, 4:276, 4:482
- selenoprotein synthesis, 4:211–212
- single-strand DNA (ssDNA) translocation, 3:264, 3:264f
- sugar nucleotide transporters, 4:351f

- adenosine diphosphate (ADP) (*Continued*)
 thiamine diphosphate (ThDP)
 biosynthesis, 1:205, 1:205f
 thiamine triphosphate (ThTP)
 biosynthesis, 1:205, 1:205f
 uncoupling proteins (UCPs), 4:482
 vacuole transport pathways, 4:504f, 4:506f
 vascular endothelial growth factor
 receptors (VEGFRs), 4:510, 4:510f
- adenosine monophosphate (AMP)
 adrenergic receptors, 1:59
 aminoacyl-adenylate adenosine
 monophosphate(AA-AMP), 4:442,
 4:441
- AMP-activated protein kinase (AMPK)
 body weight regulation, 4:58f
 cellular fatty acid metabolism regulation,
 4:56
 diabetes treatments
 metformin, 1:657
 weight loss and exercise, 1:657
 fatty acid metabolism regulation, 2:286
 fatty acid synthesis pathway, 2:275–279,
 2:276f
 insulin signal transduction, 2:604f,
 2:605
 ischemic-induced metabolic signaling
 alterations, 3:67–69, 3:69f
 lipogenic gene expression, 2:287f
 microRNA (miRNA)-based regulation,
 4:193
 anaplerotic pathways, 1:106f
 bisubstrate mimics, 2:210, 2:212f
 carbohydrate-responsive element binding
 protein (ChREBP), 1:375–376, 1:376f
 cardiac metabolism, 3:65–66, 3:66f, 3:68f
 cyclic adenosine monophosphate (cAMP)
 adenylyl cyclases (ACs), 1:42–46
 calcium regulation, 1:45, 1:45f
 characteristics and functional role,
 1:567
 functional role, 1:42
 gene families, 1:42–44, 1:43f, 1:43t
 glutamate receptors, 2:714–716,
 2:715t
 G proteins, 1:42, 1:44f, 1:45, 2:377,
 2:378f, 2:517–518, 4:225t
 intracellular second messenger
 systems, 1:42, 1:43f
 physiological functions, 1:45–46,
 1:46t
 plasma membrane components, 1:43f
 research background, 1:42
 serotonin receptors, 4:224–225,
 4:225t
 somatostatin (sst) receptors, 4:265
 adipocyte differentiation processes,
 1:56
 binding domains, 2:697f
 calcium ion (Ca^{2+}) signaling system,
 2:719–720, 2:720f, 3:294–295
 carboxyl methylation, 3:608
 cardiac excitation–contraction (E–C)
 coupling, 1:381
 cell signaling, 3:407
 characteristics and functional role, 1:567
- chemical structure, 2:696f
 class II PDEs
 catalytic activity, 1:572, 1:572f
 characteristics, 1:572
 regulatory mechanisms, 1:572, 1:572f
 class I PDEs, 1:567–568, 1:568f
 CREB (cyclic adenosine monophosphate
 (cAMP) responsive binding protein)
 adipocyte development, 1:53f, 1:55
 calcium ion (Ca^{2+}) signaling system,
 2:719–720, 2:720f, 3:294–295
 coactivator CREB-binding protein
 (CBP), 3:294–295
 CREB shut-off pathway, 3:295, 3:295f
 gene expression regulation, 1:311,
 1:313, 1:313f
 glycogen synthase kinase-3 (GSK3),
 2:436t
 hypoxia inducible factor (HIF), 4:67
 insulin signal transduction,
 2:605–606
 intracellular calcium ion (Ca^{2+})
 waves, 2:646–647
 neurogenesis, 1:197f, 1:199–200,
 1:200f
 neurotrophins, 3:241–243, 3:242f
 nuclear calcium, 1:313
 cyclic adenosine monophosphate
 (cAMP) response element (CRE),
 3:293, 3:293f
 cyclic nucleotide-dependent protein
 kinases, 1:574, 1:575f
 DARPP-32, 4:222f, 4:223
 distribution, 1:567
 GABA_A receptor, 2:351–352
 glycoprotein hormone receptors,
 4:388–390
 G-protein-coupled receptors (GPCRs),
 4:224–225, 4:225f
 G_s family proteins, 2:517–518
 insect biochemistry, 2:592–593
 insulin-secreting cells (β -cells), 2:597f
 liver PK (LPK) *in vitro* regulation, 3:719
 macroautophagy regulation, 1:150t
 molecular structure, 1:568f
 olfactory receptors, 3:349, 3:349f
 Perilipin 1 (Plin1), 1:47–48
 phosphodiesterases (PDEs), 1:580–585
 characteristics, 1:580
 clinical applications, 1:584
 family-specific inhibitors, 1:584
 G_i family functions, 2:379
 homeostatic regulation, 1:582–583
 inherited disorders, 1:584
 intracellular effectors, 1:583
 molecular diversity, 1:580–581
 research summary, 1:584
 signaling pathways, 1:574, 1:575f
 signal integrators, 1:583–584
 spatially organized signalling
 networks and microdomains,
 1:582
 structure and functions, 1:581–582,
 1:581f
 phosphotransferase system (PTS),
 2:387–388
- PKA-mediated lipolysis, 1:49–51, 1:50f,
 4:303–304
 protein N-myristoylation, 3:641–642
 research summary, 1:572
 serotonin receptors, 4:224–225, 4:225f
 signaling pathways, 1:575f
 somatostatin (sst) receptors, 4:265
 starvation effects, 4:303–304
 structural domains, 1:574–575, 1:576f
 cyclobutane pyrimidine dimers (CPDs),
 4:405
 cytokinins, 1:627–628
 enzyme–substrate interactions, 2:223–224,
 2:224f
 firefly luciferase reaction, 1:464, 1:464f
 hypothalamic malonyl-CoA, 4:58f
 intercellular calcium ion (Ca^{2+}) waves,
 2:616
 liver pyruvate kinase (LPK) catalysis, 1:374
 protein biotinylation, 1:227f
 selenoprotein synthesis, 4:211–212
 sugar nucleotide transporters, 4:351f
 thiamine diphosphate (ThDP)
 biosynthesis, 1:205, 1:205f
 thiamine triphosphate (ThTP)
 biosynthesis, 1:205, 1:205f
 tRNA aminoacylation, 4:442, 4:441
 adenosine thiochrome diphosphate
 (AThcDP), 1:204f
 adenosine thiochrome triphosphate
 (AThcTP), 1:204f
 adenosine-to-inosine (A-to-I) RNA editing,
 4:154–155, 4:155f, 4:157f
 adenosine triphosphate (ATP)
 AAA (ATPases associated with various
 cellular activities) proteins, 1:1–6
 AAA+ superfamily, 1:3–5, 1:3t, 2:21
 basic concepts, 1:1
 cellular functions
 membrane-bound metalloproteases,
 1:6
 membrane fusion, 1:5–6
 proteasome-dependent degradation,
 1:5
 vesicle transport, 1:5–6
 classifications, 1:3–5, 1:3t, 1:4f
 DNA polymerase III holoenzyme, 2:95
 dyneins, 2:184–187
 axonemal dyneins, 2:185t, 2:186–187
 characteristics and classifications,
 2:184, 2:185t
 cytoplasmic dynein 1, 2:185–186,
 2:185t, 2:186f
 cytoplasmic dynein 2 (IFT dynein),
 2:185t, 2:186
 dynein heavy chain, 2:184–185,
 2:185t, 2:186f
 dynein intermediate chain, 2:185t,
 2:186f
 dynein light chain, 2:185t, 2:186f
 inner arm dyneins, 2:185t, 2:187
 outer arm dyneins, 2:185t, 2:187
 research summary, 1:6
 ring-shaped helicases, 2:21, 2:22t, 2:23f
 structure and functions, 1:1–2, 1:2f, 1:3f
 vacuole transport pathways, 4:506f

- actin filament formation, 1:12–14, 1:13f, 1:14f, 1:18–19, 1:19f, 1:21f
- adenosine thiamine triphosphate (AThTP) biosynthesis, 1:205f, 1:206
- adenosine triphosphatase (ATPase)
- cardiac excitation–contraction (E–C) coupling, 1:380f, 1:381, 1:382f
 - copper-transporting ATPases (copper pumps), 3:670–675
 - Atox-1 copper chaperone complex, 3:674
 - ATP7A/ATP7B, 3:670–671, 3:671f, 3:674
 - biological functions, 3:670–672
 - characteristics and functional role, 4:333f, 4:334
 - copper homeostasis, 3:670–671, 3:671f
 - cytosolic catalytic domains, 3:673
 - domain structures, 3:672–673, 3:672f
 - functional role, 3:673
 - fungi, 3:671–672
 - human health functions, 3:670–671
 - mammalian copper pump regulation, 3:674
 - molecular structure, 3:672–674, 3:672f
 - N-terminal metal-binding domain (N-MBD), 3:674
 - occurrences, 3:670
 - plants, 3:671–672
 - prokaryotes, 3:671–672
 - transmembrane domains, 3:672–673, 3:672f, 3:673–674
- cytoskeletal motors, 1:634
- DNA mismatch repair (MMR)
- MutL family proteins, 2:68–69, 2:69f
 - MutS family proteins, 2:68
- DNA-strand transferases, 4:37
- endocytic–lysosomal pathways, 3:618–619
- gastric proton–potassium (H^+ – K^+)-ATPase, 3:676–680
- acid secretion, 3:679–680
 - α -subunit, 3:676, 3:677f
 - β -subunit, 3:677, 3:677f
 - characteristics, 3:676, 4:333–334, 4:333f
 - conformational changes, 3:677f, 3:678
 - functional residues, 3:678–679
 - kinetic mechanisms, 3:678
 - oligomeric structure, 3:677–678
- membrane transport
- co-transport (symport), 3:51
 - counter-transport (antiport), 3:51
 - plasma membrane calcium (Ca^{2+})-ATPase, 3:525–528
 - plasma membrane proton (H^+)-ATPases, 3:688–692
 - primary active transport, 3:51
 - sodium–potassium (Na^+ – K^+)-ATPase pump, 3:681–687
 - mitochondrial DNA (mtDNA), 3:133f
 - PET genes, 3:307–309, 3:308f
- plasma membrane calcium (Ca^{2+})-ATPase pump, 3:525–528
- alternative splicing, 3:526–527, 3:527f
- calcium ion (Ca^{2+}) pump reaction cycle, 4:331, 4:332f
- domain structures, 3:525–526, 3:526f
- genetic manipulations, 3:528–529
- isoforms, 3:526–527, 3:527f
- mechanoelectrical transduction (MET) channels, 3:528, 3:528f
- pathology, 3:528
- physiological functions, 3:528, 3:528f
- research summary, 3:529
- tissue distribution, 3:526–527
- plasma membrane proton (H^+)-ATPases, 3:688–692
- biogenesis, 3:690, 3:690f
 - domain structures, 3:688–689
 - functional role, 3:688, 3:689f
 - P-type ATPases, 3:688
 - reaction mechanisms, 3:689–690, 3:689f
 - regulation mechanisms, 3:690–691
 - research background, 3:688
 - research summary, 3:691
- protein degradation
- bacterial adenosine triphosphate (ATP)-dependent proteases, 3:623–624
 - endocytic–lysosomal pathways, 3:618–619
- P-type adenosine triphosphatase (ATPase), 4:329–334
- calcium ion (Ca^{2+}) pump reaction cycle, 4:331, 4:332f
 - characteristics and functional role, 4:329
 - copper-transporting ATPases (copper pumps), 3:670–675, 4:333f, 4:334
 - family members, 4:329
 - flippase proteins, 4:334
 - gastric proton–potassium (H^+ – K^+)-ATPase, 4:333–334, 4:333f
 - heavy metal P-type ATPases, 4:333f, 4:334
 - plasma membrane calcium (Ca^{2+})-ATPase, 3:525–528
 - plasma membrane proton (H^+)-ATPases, 3:688–692, 4:333f, 4:334
 - sarcoendoplasmic reticulum Ca^{2+} -ATPase (SERCA), 3:684f, 4:330–331, 4:330f
 - sodium–potassium (Na^+ – K^+)-ATPase pump, 3:681–687, 3:688–692, 4:331–333, 4:333f
 - structural characteristics, 4:329–330, 4:330f
- sarcoendoplasmic reticulum Ca^{2+} -ATPase (SERCA), 2:226–230, 2:231–235
- ATP-binding site, 2:232–233, 2:233f
 - calcium ion (Ca^{2+}) binding and catalytic activation, 2:228
- calcium ion (Ca^{2+}) pump reaction cycle, 4:331, 4:332f
- calcium oscillations, 1:324f
- cardiac excitation–contraction (E–C) coupling, 1:380f, 1:381
- catalytic and transport cycles, 2:226–228, 2:227f
- coupling mechanism, 2:228–229
- domain structures, 3:684f, 4:330–331, 4:330f
- $E1 \bullet 2Ca^{2+} \rightarrow E1P$ transition/occluded state, 2:232f, 2:233–234, 2:233f
- $E1P \rightarrow E2P$ transition/
- aspartylphosphate hydrolysis/lumenal gate closure, 2:232f, 2:233f, 2:234–235, 2:234f
- $E1P \rightarrow E2P$ transition/ Ca^{2+} release, 2:232f, 2:233f, 2:234, 2:234f, 3:684f, 3:689–690, 3:689f
- $E2 \rightarrow E1 \rightarrow E1 \bullet 2Ca^{2+}$ state, 2:232f, 2:233, 2:233f, 4:330f
- free-energy use, 2:228
- functional role, 4:330–331
- heart failure, 1:381–382, 1:382f
- homeostasis control mechanisms, 1:290, 1:310
- hydrolytic cleavage, 2:228–229
- inhibitors, 2:229
- intercellular calcium ion (Ca^{2+}) waves, 2:614–615, 2:614f
- intracellular calcium ion (Ca^{2+}) stores, 1:127–128
- ion pumping, 2:233–235, 4:330–331
- nucleotide binding, 2:228
- phosphorylation–calcium ion (Ca^{2+}) binding interconversion, 2:229
- phosphoryl transfer, 2:228–229
- physiological considerations, 2:229
- research summary, 2:235, 3:688–689
- ryanodine receptors (RyRs), 4:195
- steady-state behavior, 2:226
- store-operated calcium (SOC) channels, 3:359–360, 3:360f, 4:314–315
- structural characteristics, 2:226–228, 2:227f, 2:231, 2:232f, 4:333f
- substrate-induced conformational fit, 2:228
- transmembrane calcium ion (Ca^{2+}) binding sites, 2:231, 2:232f
- secondary energy transducers, 3:34
- sliding clamp/clamp-loader interactions, 2:114–115, 2:115f, 2:116t, 3:576, 3:577f, 4:237–238, 4:237f
- sodium–potassium (Na^+ – K^+)-ATPase pump, 3:681–687, 3:688–692
- α subunits, 3:683–684, 3:684f
 - β subunits, 3:683f, 3:684–685
 - cardiac glycosides, 3:686–687, 3:686f, 3:687f
 - cation-pumping mechanism, 3:683–684, 3:684f
 - characteristics and functional role, 3:681, 4:331–333
 - domain structures, 3:683–684, 3:683f, 4:333f

- adenosine triphosphate (ATP) (*Continued*)
 FXD proteins, 3:681, 3:682–685, 3:682*t*, 3:683*f*
 physiological functions, 3:685–686
 reaction mechanisms, 3:683, 3:682
 structural characteristics, 3:682–685, 3:683*f*, 3:684*f*
 subunits and isoforms, 3:681, 3:682*t*, 3:683*f*
 20S proteasome, 3:597
 vacuolar (H⁺)-ATPases, 4:520–524
 functional role, 4:520–521
in vivo activity regulation, 4:523
 intracellular V-ATPases, 4:520, 4:521*f*
 plasma membrane V-ATPases, 4:520–521
 structural characteristics, 4:521–523, 4:522*f*
 subunit composition, 4:522*t*
 transport mechanisms, 4:523
- adenosine triphosphate (ATP) synthase, 2:269–274
 anaerobic respiration
 H⁺ translocation, 2:208, 2:208*f*
 proton potential, 2:205*f*, 2:206–208
 bacterial reductive reactions, 1:488*f*
 bioenergetic costs, 2:272
 bioenergetic reactions, 1:486–487, 1:487*f*, 3:127
 characteristics and functional role, 2:269
 citric acid cycle, 4:93*f*
 cytochrome *b₆f* complex, 4:335, 4:336*f*
 domain structures, 2:269–271, 2:270*f*, 2:271*f*
 electron transfer chain, 1:468, 1:470
 F₁ domain, 2:271–272, 2:273*f*
 mitochondrial membranes, 3:147
 occurrences and functional role, 4:85–86
 oxygenic photosynthesis, 2:335
 photosynthetic reaction centers, 3:701*f*
 regulatory mechanisms, 2:273–274
 research summary, 2:274
 structural characteristics
 F₀ component, 4:86
 F₁ component, 4:86
 general characteristics, 4:85–86
 schematic diagram, 4:86*f*
 synthesis mechanisms, 2:271–272, 2:273*f*, 4:86
- anti-mitochondrial antibodies (AMAs), 3:117
- archaeal energy transduction, 3:32–33, 3:33*f*
- ATP-dependent nucleosome remodeling
 chromodomain helicase DNA binding (CHD)/Mi-2 family, 4:60*f*, 4:64
 INO80 (inositol requiring 80) family, 4:60*f*, 4:64–65, 4:65*f*
 ISWI family, 4:60*f*, 4:64
 remodeling enzymes, 4:63–65
 schematic diagram, 4:65*f*
 SWItching defective/Sucose
 NonFermenting (SWI/SNF) family, 1:314, 4:60*f*, 4:63–64, 4:403–404
 Swr1-dependent deposition of H2A.
 Z-H2B dimers, 4:64–65, 4:65*f*
- ATP hydrolysis
 actin-related proteins (Arps), 1:36, 1:39*f*
 DNA-strand transferases, 4:38–39
 reverse gyrase mechanism, 2:159–160
 type I DNA restriction and modification enzymes, 2:74, 2:75*f*
- ATP synthesis, 1:141–144
 background information, 1:141
 bioenergetic reactions, 1:486–487, 1:487*f*, 1:488*f*
 chemiosmotic theory, 1:467–472
 ATP synthase, 1:468, 1:470
 basic concepts, 1:468–472, 1:468*f*
 electron transfer chain, 1:468
 future research outlook, 1:472
 historical research, 1:467
 ion-impermeable membranes, 1:468, 1:468*f*, 1:471*f*
 mitochondrial permeability transition, 3:162
 proton circuit, 1:470–471
 proton gradient coupling, 1:470
 protonmotive electron transfer chain, 1:469–470, 1:469*f*, 1:470*f*
 protonmotive force (PMF), 1:468, 1:468*f*, 1:471*f*
 research background, 1:467
 stoichiometric conditions, 1:470–471
 transport systems, 1:471–472, 1:471*f*
 vectorial redox loop, 1:468–469, 1:469*f*
 chemolithotrophs, 1:488*f*, 1:490–491
 cytochrome *b₆f* complex, 4:335, 4:336*f*
 glycine decarboxylase, 1:142–143
 inhibitors, 1:143, 1:143*t*
 plant mitochondrial electron transport chain, 1:141–142, 1:142*f*, 1:145, 1:146*f*
 substrates, 1:142–143
 uncoupling processes
 alternative NAD(P)H dehydrogenases, 1:144
 alternative oxidases (AOXs), 1:144
 chemical uncouplers, 1:143
 nonphosphorylating respiration, 1:142*f*, 1:144
 uncoupling proteins (UCPs), 1:142*f*, 1:144, 4:482
- biotin-mediated carboxyl transfer reactions, 1:222
- bisubstrate mimics, 2:210, 2:212*f*
- calcium ion (Ca²⁺) binding, 1:282
- carbohydrate-responsive element binding protein (ChREBP), 1:375–376, 1:376–377, 1:376*f*, 1:377*f*
- carboxyl methylation, 3:607
- cardiac metabolism, 3:65–66, 3:66*f*, 3:68*f*, 3:70–71
- cellular respiration, 3:407, 3:408*f*
- cerebral metabolism, 1:369*f*, 1:371–373
- chlorophyll, 3:544
- chloroplast signaling, 1:511–512, 1:512*f*
- creatine kinase/phosphocreatine circuit, 3:81, 3:82*f*
- cytidine diphosphate (CDP)–ethanolamine pathway, 3:480
- cytochrome *b_c1* complex, 1:593, 1:601
- cytokinins, 1:627–628
- cytoskeletal motors, 1:632
- cytosolic creatine kinase, 3:81–82
- DNA ligases
 covalent DNA–AMP covalent intermediate formation, 2:28, 2:33, 2:34*f*, 2:35–37, 2:36*f*, 2:38*f*
 covalent ligase–AMP covalent intermediate formation, 2:28, 2:29*f*, 2:33–35, 2:35*f*, 2:38*f*
 domain structure and organization, 2:33, 2:34*f*
 functional role, 2:33, 2:34*f*
 phosphodiester bond formation, 2:28, 2:29*f*
 research background, 2:28
- DNA mismatch repair (MMR)
 DNA polymerase III holoenzyme, 2:95
 MutH pathway, 2:56, 2:56*f*
 MutL pathway, 2:55–56, 2:56*f*
 MutS pathway, 2:55, 2:56*f*, 2:60, 2:61*f*, 2:68
- DNA polymerases
 DnaX complex, 2:95
 Y-family polymerases, 4:479
- dopamine receptors, 2:169
- enzyme–substrate interactions, 2:223–224, 2:224*f*
- fatty acid oxidation, 2:281
- fatty acid synthesis pathway, 2:275–279, 2:276*f*, 2:291*f*
- fatty acyl-coenzyme A (CoA) synthetases, 2:290–295
 ACSF proteins, 2:294
 activation reactions, 2:290, 2:291*f*
 amino acid sequence motifs, 2:290–291, 2:291*f*, 2:294–295
 Bubblegum ACS family (ACSBG), 2:294
 characteristics and occurrence, 2:290–294
 human ACS families, 2:290–294, 2:292*t*
 long-chain acyl-CoA synthetase (LCAS), 1:385*f*, 2:292–293
 medium-chain acyl-CoA synthetase (MCAS), 2:292
 nomenclature, 2:292, 2:292*t*
 research summary, 2:295
 short-chain acyl-CoA synthetase (SCAS), 2:292
 structure–function relationship, 2:291*f*, 2:294–295
 very long-chain acyl-CoA synthetase (VLCAS), 2:293–294
- fermentation reactions, 2:205, 2:205*t*
- firefly luciferase reaction, 1:464, 1:464*f*
- flippase proteins
 aminophospholipid flippases, 2:314–315, 2:315*f*, 2:316*t*
 flippase-mediated transbilayer movement, 2:314, 2:315*f*
 floppases, 2:314–315, 2:315*f*
 scramblases, 2:314–315, 2:315*f*
- functional role, 1:141
- glucokinase, 2:546–547, 2:546*f*
- gluconeogenesis, 2:382*f*, 2:385*f*

- glycerolipid synthesis, 2:415
- glycogen synthesis, 2:430f
- glycolysis, 1:374, 2:443, 2:444f, 2:445
- glycoprotein folding processes, 2:448–449
- Golgi apparatus, 2:482
- GroEL–GroES protein-folding cycle, 1:457–458, 1:458f
- hypothalamic malonyl-CoA, 4:58f
- hypoxia inducible factor (HIF), 4:67
- insulin receptors (IRs), 2:602
- insulin-secreting cells (β -cells), 2:597f
- intercellular calcium ion (Ca^{2+}) waves
- gap junction (GJ) communication, 2:615, 2:615f
- initiation mechanisms, 2:613–614
- paracrine communication, 2:615–616, 2:615f
- spatial characteristics, 2:616
- intracellular calcium ion (Ca^{2+}) stores, 2:482, 2:576f
- inwardly rectifying potassium channels (K_{ir}), 2:650
- kinesins, 2:672, 2:673f
- light harvesting proteins, 2:178–179, 2:179f
- mitochondria
- historical research, 3:306–307
- mitochondrial ATP-dependent potassium ion (K^+) channel (KATP), 3:551, 3:551f
- mitochondrial β -oxidation, 1:386–387
- mitochondrial calcium (Ca^{2+}) transport, 3:118, 3:119f
- mitochondrial channels, 3:127
- mitochondrial CK (MtCK) isoenzyme, 3:82–84, 3:83f
- mitochondrial dynamics, 3:135–136
- regulation mechanisms, 3:169–170
- mTOR complexes, 3:196
- muscle metabolism, 1:94
- MutS complexes, 2:69f
- myocardial ischemia, 3:108–109, 3:109f
- myosin, 3:219–221, 3:221f
- N-ethylmaleimide-sensitive factor (NSF), 1:3t, 1:5, 3:37
- neuronal death, 1:424
- nucleotide excision repair (NER) pathways, 3:329f
- origin recognition complex (ORC), 2:111–112, 2:112f
- P2Y receptors, 3:375–378
- domain structures, 3:376
- extracellular nucleotide signaling, 3:375–376
- functional classification, 3:376–377, 3:376t
- functional role, 3:375
- lysophosphatidic acid (LPA) receptors, 2:763t
- pharmacological classification, 3:377
- physiological functions, 3:377–378
- peroxisome proliferator-activated receptor alpha (PPAR α), 3:419–421, 3:420f
- phosphofructokinase-2/fructose biphosphatase-2 (PFK-2/FBPase-2), 3:456–457, 3:458f
- phospholipase D (PLD), 3:472–473
- photosynthesis
- electron transport mechanisms, 1:511–512, 1:512f, 3:493–494, 3:494f
- thylakoid membranes, 3:492, 3:494f
- photosynthetic carbon dioxide (CO_2) fixation, 3:497–502
- C4 pathways, 3:498
- Calvin cycle enzyme targets, 3:500–501
- carbon pathway, 3:497, 3:499f
- carboxylation phase, 3:497–498, 3:499f
- Crassulacean acid metabolism (CAM) pathway, 3:498
- enzymatic regulation
- flux control, 3:500
- irreversible enzymes, 3:500
- feedback regulation, 3:501
- genetic modification effects, 3:500–501
- light-regulated energy and power, 3:497, 3:498f
- metabolic regulation, 3:499–500
- reduction phase, 3:498, 3:499f
- regeneration phase, 3:499, 3:499f
- whole plants, 3:501
- physiological hepatocellular bile salt transport, 1:170
- protein biotinylation, 1:226
- purple bacteria, 3:693–695
- pyruvate dehydrogenases (PDHs), 4:321, 4:324
- pyruvate kinase (PK), 3:719
- respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:87–91
- background information, 4:98
- bacterial reductive reactions, 1:488f
- basic concepts, 4:87, 4:341
- disease connections
- general discussion, 4:90–91
- mitochondrial mutations, 4:90–91
- neurotoxins, 4:91
- nuclear mutations, 4:91
- pathological mechanisms, 4:91
- evolutionary development
- general discussion, 4:88–90
- N-module, 4:89
- origins, 4:89–90
- P-module, 4:89
- Q-module, 4:88–89
- variants, 4:89–90
- functional role
- active–deactive transition, 4:90
- electron transfer mechanisms, 3:661–662, 3:662f, 3:663, 4:83–84, 4:90
- general discussion, 4:90
- inhibitors, 4:90
- proton pumping, 3:661–662, 3:662f, 3:663, 4:90
- redox centers, 4:90
- sodium ion (Na^+) pumping, 4:90
- PET genes, 3:308f
- plant mitochondrial electron transport chain, 1:145, 1:146f
- protonmotive electron transfer chain, 1:470f
- schematic diagram, 4:84f, 4:99f
- structural characteristics, 4:341–346
- accessory subunits, 4:88, 4:89t
- central subunits, 4:87–88, 4:88f, 4:89t
- cysteine motif, 4:344–345, 4:345f
- flavin mononucleotide (FMN) binding, 4:344–345, 4:344f
- frataxin-like subunit, 4:343–344, 4:343f
- hydrophilic domain, 4:342–343, 4:342f
- iron–sulfur clusters, 4:87–88, 4:89t, 4:342–343, 4:342f, 4:345f
- multistage assembly pathway, 4:88
- mutations, 4:345
- NADH binding, 4:344–345, 4:344f, 4:345f
- overall structure, 4:341–342
- redox centers, 4:87–88, 4:89t, 4:342–343, 4:342f
- research summary, 4:345–346
- shape, 4:87, 4:88f
- subunits, 4:342t
- supernumerary subunits, 4:345
- selenoprotein synthesis, 4:211–212
- single-strand DNA (ssDNA) translocation, 3:264, 3:264f
- skeletal muscle, 1:108
- starvation effects, 4:303–304
- sugar nucleotide transporters, 4:350, 4:351f
- sugar phosphorylation, 2:543
- synthesis reactions, 1:218
- thiamine diphosphate (ThDP) biosynthesis, 1:205, 1:205f
- thiamine triphosphate (ThTP) biosynthesis, 1:205, 1:205f
- TRiC/CCT eukaryotic chaperonin, 1:458–459
- tRNA aminoacylation, 4:442, 4:441
- 26S proteasome, 3:595–600
- basic concepts, 3:590–592, 3:591f
- cell regulation
- chaperones, 3:599
- E3 ubiquitin ligase, 3:599
- substrate degradation, 3:599
- ubiquitin proteasome pathway (UIPP), 3:599
- characteristics, 3:595
- enzyme mechanism, 3:591, 3:592f
- 19S regulatory complex
- assembly processes, 3:598
- base complex, 3:597f, 3:598
- basic concepts, 3:590–592
- lid complex, 3:597f, 3:598
- regulatory mechanisms, 3:598–599
- structural characteristics, 3:597–598, 3:597f
- physiological importance, 3:592
- research summary, 3:599–600
- structural characteristics, 3:597f
- tissue and subcellular distribution, 3:591–592
- type I DNA restriction and modification enzymes, 2:74, 2:75f

- adenosine triphosphate (ATP) (*Continued*)
 type II DNA topoisomerase, 2:164, 2:164f, 2:165f
 type III DNA restriction and modification enzymes, 2:136, 2:137t, 2:139f
 vacuole transport pathways, 4:504f, 4:506f
 voltage-dependent anion-selective channels (VDACs), 3:161
See also ATP-binding cassette (ABC) transport systems
- adenosylcobalamin, 4:137
- S-adenosyl-L-methionine (AdoMet), 2:130, 2:136, 2:137f, 2:137t, 4:137
- adenosyl-pentaphosphate-adenosine (AP₅A), 2:210, 2:212f
- adenovirus, 4:459
- adenylate cyclase, 2:172, 2:597f, 2:695, 3:349f, 3:646f
- adenylate translocator (ANT), 3:82–84, 3:83f
- adenylosuccinate, 1:106f
- adenylyl cyclases (ACs), 1:42–46
 calcium regulation, 1:45, 1:45f
 characteristics and functional role, 1:567
 functional role, 1:42
 gene families, 1:42–44, 1:43f, 1:43t
 glutamate receptors, 2:714–716, 2:715t
 G proteins
 G_i family, 1:42, 1:44f, 1:45, 2:377, 2:378f, 4:225t
 G_s family, 1:42, 1:44f, 1:45, 2:517–518
 regulation mechanisms, 1:44f, 1:45
 intracellular second messenger systems, 1:42, 1:43f
 physiological functions, 1:45–46, 1:46t
 plasma membrane components, 1:43f
 research background, 1:42
 serotonin receptors, 4:224–225, 4:225t
 somatostatin (sst) receptors, 4:265
- adherens junctions, 1:255, 1:261, 1:262f, 4:392, 4:393f
- adhesion sites
 calcium-sensing receptor (CaSR), 1:329–330
- desmosomes, 1:636–643
 assembly and maintenance, 1:639
 desmosomal cadherins, 1:255–256, 1:256f, 1:259t, 1:261, 1:637, 1:638t
 disease connections, 1:638t
 functional role
 cell–cell adhesion, 1:638t, 1:639
 morphogenetic functions, 1:639
 signaling functions, 1:639–640
 molecular components
 armadillo family members, 1:637–638, 1:638t
 desmosomal cadherins, 1:637, 1:638t
 plakins, 1:638–639, 1:638t
 structural characteristics, 1:636–639, 1:637f
 occurrences and functional role, 1:636–640
 ultrastructure, 1:636, 1:637f
- hemidesmosomes, 1:636–643
 assembly and maintenance, 1:642
 disease connections, 1:641t
 functional role, 1:643
- cell–substrate adhesive function, 1:642
 morphogenetic functions, 1:642
 transmembrane signaling function, 1:642–643
 wound healing, 1:642–643
- molecular components
 α6β4 integrin heterodimer, 1:640f, 1:641, 1:641t, 1:642–643
 BPAG1 protein, 1:641, 1:641t
 CD151, 1:642
 laminin-332, 1:641t, 1:642
 plectin, 1:641–642, 1:641t
 structural characteristics, 1:640–642, 1:640f
 type XVII collagen, 1:640f, 1:641, 1:641t
 ultrastructure, 1:640, 1:640f
- adipogenesis, 1:52–56
- adipocytes
 development process
 C/EBPα, 1:56
 C/EBP homologous protein (CHOP-10), 1:55–56
 cell culture/*in vivo* models, 1:52
 differentiation processes, 1:55f, 1:56, 2:259, 2:259f
 lineage commitment, 1:52–54, 1:53f
 peroxisome proliferator-activated receptor gamma (PPARγ), 1:56, 3:421–423
 sterol regulatory element binding protein (SREBP), 1:55f, 1:56
 Wnt pathway, 1:53–54, 1:53f
- differentiation processes
 basic principles, 1:56, 2:259, 2:259f
 induction processes, 1:54
 mitotic clonal expansion, 1:54, 1:55f
 terminal differentiation, 1:54–55, 1:54f
 transcriptional control, 1:55, 1:55f
- fat mobilization, 1:47–51
- adipose triglyceride lipase (ATGL), 1:49–51, 1:50f
 hormone-sensitive lipase (HSL), 1:48–49, 1:49f, 1:50f
 intracellular neutral lipid storage droplets, 1:47
 lipolysis regulation, 1:48
 Perilipin 1 (Plin1), 1:47–48, 1:48f, 1:48t, 1:49f, 1:50f
 Plin protein family functional role, 1:51
 functional role, 1:52, 1:53f
- intracellular neutral lipid storage droplets
 adipose triglyceride lipase (ATGL), 1:49–51, 1:50f
 animal cells, 1:47
 general discussion, 1:47
 hormone-sensitive lipase (HSL), 1:48–49, 1:49f, 1:50f
 lipolysis regulation, 1:48, 1:50f
 liquid droplets (LDs), 1:47
 Perilipin 1 (Plin1), 1:47–48, 1:48f, 1:48t, 1:49f
- Perilipin/Plin protein family, 1:47, 1:48f, 1:48t
 Plin protein family functional role, 1:51
 microRNA (miRNA)-based regulation, 4:192–193, 4:194t
 origins, 1:52
 transcriptional control
 basic principles, 1:55, 1:55f
 C/EBPβ expression, 1:55
 CREB (cyclic adenosine monophosphate (cAMP) regulatory binding protein), 1:55
 basic concepts, 1:52
 leptin, 3:76
- adiponectin, 3:72–75
 adiponectin receptors, 3:72–73
 appetite regulation, 4:303
 characterization, 3:72
 circulation regulation, 3:72, 3:73t
 physiological functions
 anti-atherogenic effects, 3:74
 anti-inflammatory effects, 3:74
 general discussion, 3:73–74
 insulin sensitivity effect, 3:74
 vascular endothelium, 3:74
 schematic structure, 3:73f
 therapeutic target considerations, 3:74
- adipose tissue
 acyl-coenzyme A synthetase
 long-chain acyl-CoA synthetase (LCAS), 2:292–293
 very long-chain acyl-CoA synthetase (VLCAS), 2:293–294
 adiposity and satiety signal integration, 3:79f
 appetite regulation, 4:303
 aquaporin water channels, 1:124t
- brown adipose tissue (BAT)
 adenosine diphosphate (ADP)
 phosphorylation and respiration, 4:483–484, 4:484f
 characteristics and functional role, 4:483
 G_s family proteins, 2:518–519
 histology, 4:483f
 leptin, 2:258–259
 thermogenesis, 4:483–484, 4:484f
 uncoupling proteins (UCPs), 4:483–484, 4:484f, 4:485f
- carbohydrate-responsive element binding protein (ChREBP) distribution, 1:374
- energy homeostasis, 3:77f, 3:78f
 energy storage, 4:299–300
 G_s family phenotypes, 2:521t
 insulin metabolism, 2:603f, 2:606–607
 insulin resistance, 2:606–607
 leptin, 2:256–261
 adiposity and satiety signal integration, 3:79f
 appetite regulation, 4:303
 brown adipose tissue (BAT), 2:258–259
 characteristics and functional role, 2:256
 deficiency-related diseases, 2:256
 energy homeostasis, 3:77f, 3:78f
 energy intake–energy expenditure relationship, 2:257–258, 2:257f

- expression regulation, 2:259, 2:259f
 historical research, 2:256, 2:257f, 3:76
 homeostatic regulation, 2:257–258, 2:257f, 2:258f
 hypothalamic malonyl-CoA, 4:57, 4:58f
 leptin resistance, 2:260–261, 2:260f
 obesity syndrome, 2:256
 uncoupling proteins (UCPs), 2:258–259
 starvation effects, 4:303–304
 white adipose tissue (WAT), 1:583, 2:259, 2:518–519
- adipose triglyceride lipase (ATGL), 1:49–51, 1:50f
- adipsin, 4:303
- AdoMet (S-adenosylmethionine)
 brassinosteroid biosynthesis, 1:251
 carboxyl methylation, 3:607, 3:608f, 3:609f
 DNA alkylation damage, 2:9
 [FeFe]-hydrogenases, 4:72
 iron–sulfur clusters (FeS), 2:660, 2:661f
 riboswitches, 4:137, 4:138f, 4:251
- ADPribose transferase (ART), 1:176–177, 1:177f
- adrenal glands, 2:292–293
- adrenaline
See epinephrine
- adrenergic receptors, 1:57–60
 action mechanisms, 1:57–58
 characteristics and functional role, 1:57
 classifications
 alpha-1 adrenergic receptors
 general characteristics, 1:58
 pharmacological/molecular characteristics, 1:58
 schematic diagram, 1:58f
 alpha-2 adrenergic receptors
 general characteristics, 1:58
 pharmacological/molecular characteristics, 1:58–59
 schematic diagram, 1:58f
 beta adrenergic receptors
 feedback regulation, 1:578–579, 1:578f
 general characteristics, 1:59
 pharmacological/molecular characteristics, 1:59
 schematic diagram, 1:58f
 characteristics, 1:57–58
- epinephrine
 branched-chain amino acid (BCAA) regulation, 1:245–246
 characteristics and functional role, 1:57
 starvation effects, 4:303–304
- G_q family proteins, 2:496–497, 2:497t
- norepinephrine
 ascorbic acid, 4:531–532
 characteristics and functional role, 1:57
 polymorphisms, 1:60
 regulation mechanisms, 1:59
 signal transduction pathways, 1:59
- adrenochrome, 3:727, 3:727f
- adrenocortical steroidogenesis, 3:25–26, 3:25t
- adrenocorticotrophic hormone (ACTH)
 calcium-sensing receptor (CaSR), 1:329–330
- functional role, 3:24–25, 3:25t, 3:26f
- adrenodoxin, 2:301
- adrenomedullin, 1:265–269
 binding site distribution, 1:268
 characteristics, 1:265, 1:271
 molecular structure
 amino acid ring structure, 1:271f
 amylin receptors, 1:268
 calcitonin receptor-like receptor (CLR), 1:266, 1:267f
 cellular signaling, 1:268
 CLR–RAMP2/CLR–RAMP3 association, 1:266–267
 receptor activity-modifying proteins (RAMPs), 1:266, 1:267f
 receptor component protein (RCP), 1:268
 pathophysiology, 1:268–269
 pharmacological characteristics
 adrenomedullin receptor-selective drugs, 1:265, 1:266f
 nonpeptide antagonists, 1:265, 1:266f
 receptor expression changes, 1:268
 structural characteristics, 1:265, 1:266f
- adseverin, 1:23–24, 1:277–278
- advanced glycation end products (AGEs)
 AGE inhibitors and breakers, 2:409
 AGE receptors, 2:407–409
 aging effects, 2:409, 2:410f
 alternative pathways, 2:407, 2:409f
 basic concepts, 2:405
 characteristics, 2:407, 2:408f
 diabetes, 2:409–410, 2:410f
 disease effects, 2:410
 RAGE (receptor for advanced glycation end products), 2:407–409
- advanced lipoxidation end products (ALEs), 2:407, 2:409f
- Aedes aegypti*
 blood meal metabolism, 2:591, 2:591f
 Dengue fever, 2:591
 20-hydroxyecdysone (20E), 2:593
 urea synthesis, 2:591–592, 2:592f
- Aegilops tauschii*, 3:284f
- Aequorea victoria*, 2:577
- aequorin, 1:319t, 2:577, 2:577f
- Aeromonas*, 2:43
- affective disorders, 4:364–365
- affinity tags, 1:61–65
 functional roles, 1:61
 fusion protein construction
 affinity–fusion gene constructs, 1:61
 general discussion, 1:61
 linker peptide composition, 1:61
 production and immobilization processes, 1:61–62
 selection criteria, 1:61
- system characteristics
 antibody–epitope interaction affinity tags, 1:64
 basic concepts, 1:64–65
 polyamino acid–matrix interaction
 affinity tags, 1:62–64
 protein–ligand interaction affinity tags, 1:62
- protein–protein interaction affinity tags, 1:64–65
- aggrecan, 3:91t
- aging/age-related diseases, 2:79
- agouti receptor antagonist, 3:24–25, 3:25t, 3:26f
- agouti-related protein (AgRP)
 body weight regulation, 3:77, 3:77f, 3:78f
 energy homeostasis, 3:26
 functional role, 3:24–25, 3:25t, 3:26f
 insulin metabolism, 2:603f, 2:606
 leptin expression, 2:257–258
- agrin, 3:657t, 3:658
- Agrobacterium tumefaciens*, 1:491t, 1:631, 3:331f
- Akt (serine–threonine active kinase protein)
 arrestins, 2:488t
 calcium/calmodulin-dependent protein kinase II (CaMKII), 1:297–303
 autophosphorylation
 autonomous activity, 1:299
 calmodulin (CaM) trapping, 1:300
 neuronal activity monitoring, 1:301, 1:301f
 structure and functions, 1:299, 1:300f
 substrate anchoring, 1:300
 threonine (Thr³⁰⁵/Thr³⁰⁶)
 autophosphorylation, 1:300–301
 cardiac regulation, 1:302
 characteristics and functional role, 1:297
 domain structures, 1:298–299, 1:299f
 intracellular localization, 1:297–298, 1:298f
 long-term potentiation, 1:302
 memory and learning regulation, 1:302
 structure and functions, 1:302–303
 substrates, 1:297–298, 1:298f
- chemoattractant signaling, 2:236–238, 2:237f
- chemokine receptors, 1:480–482, 1:482f
- fatty acid metabolism–cancer relationship, 2:277
- focal adhesions, 2:321
- functional role, 3:463–464
- glycogen synthase kinase-3 (GSK3), 2:434, 2:435f
- G protein-coupled receptor kinases (GRKs), 2:485t
- hematopoietin receptors, 2:528f
- insulin signal transduction, 2:604–605, 2:604f
- integrin signaling, 2:611
- liver regeneration, 1:193–195
- mitochondrial permeability transition pore, 3:165–166
- MSP polymerization-activating kinase (MPAK), 3:3t, 3:4, 3:4f
- mTOR complexes, 1:408, 3:195–196
- neurogenesis, 1:199, 1:200f
- neurokinin-1 receptor (NK1R), 4:362–363
- neurotrophins, 3:241–243, 3:242f
- phosphatidylinositol-3,4,5-trisphosphate (PtdIns-3,4,5-P₃), 3:450
- platelet-derived growth factor (PDGF) receptors, 3:541f
- plectin, 2:625–627

- Akt (serine–threonine active kinase protein) (*Continued*)
 protein kinase C (PKC), 3:637–638
 small guanosine phosphatases (GTPases), 4:247f
 somatostatin (sst) receptors, 4:265
 voltage-dependent potassium ion (K^+) channels, 4:557–559, 4:558f
See also protein kinase B (PKB)
- alanin, 3:717–718, 3:717f
- alanine
 alanine transaminase (ALT), 4:494
 anaplerotic pathways, 1:106f
 classifications, 1:91, 1:92t
 elastin, 2:192
 gluconeogenesis, 2:382f, 2:385f
 insect biochemistry, 2:590–591
 muscle metabolism, 1:94f
 phage libraries, 3:434t
 starvation effects, 4:302, 4:304–305
 tricarboxylic acid cycle, 4:439f
- alanine aminotransferase (ALT), 3:565
- Albright hereditary osteodystrophy (AHO), 2:519–520
- Alcaligenes*, 3:29–30
- Alcaligenes eutrophus*, 1:490t, 3:31–32
- alcohol, 2:423
- alcohol dehydrogenases (ADHs), 2:222, 2:223f, 4:74f, 4:528–529
- aldehyde dehydrogenase (ALDH), 4:525
- aldolase, 3:499f, 3:500
- aldoses, 2:409f
- aldosterone, 4:308–309
- Alexander's disease, 2:632t
- algae, 1:66–70
 alternative oxidases (AOXs), 1:145–146
 anoxygenic photosynthesis, 2:209
 antenna complexes, 3:484, 3:484f
 ferredoxins (Fd), 2:296
 green algae, 1:66–70
 anoxygenic photosynthesis, 2:209
 dyneins, 2:184
 ferredoxins (Fd), 2:296
 nonphotochemical quenching (NPQ), 3:486
 plastocyanin (PC), 3:530
 reaction centers, 3:492
 light harvesting complexes, 2:726
 nonphotochemical quenching (NPQ), 3:486
 photosynthesis, 1:513, 1:513f, 3:495
 plastocyanin (PC), 3:530
 pterin distribution, 3:666
 reaction centers, 2:335, 3:492
- aliphatic amino acids, 1:92t
- alkaline ceramidase, 4:289
- alkaline phosphatase, 3:564, 3:565, 4:503–506, 4:505f
- alkaline sphingomyelinase (bSMase), 4:288
- 1-O-alkyl-2-acetyl-sn-glycero-3-phosphocholine, 3:533
- alkyladenine DNA glycosylase (AAG), 2:16–17, 2:20
- alkyl-dihydroxyacetone phosphate (alkyl-DHAP), 3:416–417
- O⁶-alkylguanine-DNA alkyltransferase (AGT), 2:10t, 2:12–13
- allergic proteinases, 1:592
- Allium cepa*, 3:21f
- allograft inflammatory factor 1, 1:319t
- allolactose, 2:696–697, 2:696f
- allopurinol, 2:312
- allosteric regulation, 1:71–74
 allosteric proteins
 aspartate transcarbamoylase (ATCase), 1:73
 cell-surface receptors, 1:74
 GroEL (growth E locus, large gene, bacterial group I chaperonin), 1:74
 protein structures, 1:73
 ribonucleotide reductases (RRs), 1:73–74
- cooperative interactions
 general discussion, 1:71–73
 Hill *n* plot, 1:71
 KNF (sequential) model, 1:71–73, 1:72f
 MWC (symmetry) model, 1:71–73, 1:72f
 oxygen–hemoglobin (O_2 –Hb) binding, 1:71, 1:72f
 two-state models
 basic principles, 1:71–73, 1:72f
 modifications, 1:73
- GABA_A receptors, 2:348f, 2:349
- hemoglobin (Hb), 2:534, 2:534f
- all-trans-retinoic acid (atRA)
 chemical and metabolic reactions, 4:525–526, 4:526t
 clinical uses, 4:529
 functional roles, 4:525
 homeostasis control mechanisms, 4:529
 naturally occurring retinoids, 4:529
 pharmacological retinoid concentrations, 4:528–529
 retinoid-binding proteins (RBPs), 4:526–528, 4:527f, 4:527t
 retinoid homeostasis, 4:526–528
- all-trans-retinol (atROH), 4:525–529
 chemical and metabolic reactions, 4:525–526, 4:526t
 clinical uses, 4:529
 functional roles, 4:525
 homeostasis control mechanisms, 4:529
 naturally occurring retinoids, 4:529
 pharmacological retinoid concentrations, 4:528–529
 retinoid-binding proteins (RBPs), 4:526–528, 4:527f, 4:527t
 retinoid homeostasis, 4:526–528
- all-trans-retinyl esters (atREs), 4:525–526
- alpha-1 adrenergic receptors
 general characteristics, 1:58
 pharmacological/molecular characteristics, 1:58
 schematic diagram, 1:58f
- alpha-2 adrenergic receptors
 general characteristics, 1:58
 pharmacological/molecular characteristics, 1:58–59
 schematic diagram, 1:58f
- α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA)
 functional role, 2:714–716, 2:715t
 kinesin superfamily proteins (KIFs), 2:681–684, 2:685f
 molecular structure, 2:716t, 2:717–718, 2:717f, 2:718f
 schematic diagram, 1:425f
- α-galactosidase A, 4:292
- α-internexin, 2:632t, 2:634, 3:235
- α-melanocyte stimulating hormone (α-MSH), 3:77, 3:77f
- alpha-methylacyl-CoA racemase (AMACR), 2:279
- alpha-oxidation
 fatty acid oxidation
 mitochondria, 2:281–282, 2:282f
 peroxisomes, 2:283–284
 unsaturated fatty acids, 2:282–283, 2:283f
 fatty acid synthesis pathway, 2:291f
- alternative oxidases (AOXs)
 plant mitochondria
 basic concepts, 1:145–146
 biochemical regulation, 1:147–148, 1:147f
 physiological functions, 1:148
 structural characteristics, 1:146–147, 1:147f
 uncoupling processes, 1:144
- alternative splicing, 1:75–80, 1:81–86
- auxiliary sequence elements, 1:76–77, 1:78f
- background information, 1:75
- CD45 gene, 1:79–80, 1:79f
- Drosophila melanogaster*
 double-sex (*dsx*) gene regulation, 1:85–86
 exon organization and splicing patterns, 1:81–82, 1:82f
 fruitless (*fru*) gene regulation, 1:85–86
 research summary, 1:86
 Sxl autoregulation, 1:82–84, 1:85f
 transformer (*tra*) gene regulation, 1:82, 1:85–86, 1:85f
- exonic splicing enhancers (ESEs), 1:76, 1:78f, 3:62, 3:62f
- exonic splicing silencers (ESSs), 3:62f
- exon/intron boundaries, 1:75, 1:76f, 1:397, 1:398f
- fine-tuning determinants, 1:78–79
- hnRNP protein family, 1:76–77, 1:78f, 1:79–80, 1:79f
- intronic splicing enhancers (ISEs), 1:76, 1:78f, 3:62, 3:62f
- intronic splicing silencers (ISSs), 3:62f
- non-*Drosophila* species, 1:86
- post-transcriptional modifications, 1:78
- pre-messenger RNAs (pre-mRNAs), 1:75, 1:76f, 3:60–61, 4:187–188
- regulation mechanisms, 1:77–78, 1:78f
- research summary, 1:86
- small nuclear RNP (snRNP) complexes, 1:75, 1:76f, 1:78f
- splicing patterns, 1:75–76, 1:77f
- splicing silencer sequences, 1:76–77, 1:78f
- Alu genes, 4:160

- Alzheimer's disease, 1:473–479
- amyloid plaques
- amyloid β -protein ($A\beta$), 1:286, 1:473–475, 1:474f, 1:476f, 2:481, 4:199, 4:200f
 - amyloid fibril formation, 1:474–475, 1:475f, 1:476f
 - amyloid precursor protein (APP), 1:473–475
 - caspases, 1:394
 - characteristics, 1:100–101, 1:101t, 1:473, 1:474f
- amyloid precursor protein (APP)
- amyloid β -protein ($A\beta$) formation, 1:286, 1:474f
 - caspases, 1:394
 - characteristics, 1:473–475
 - γ -secretase-based mediation, 4:51f
 - genetic relationships
 - mutation effects, 1:474, 1:474f
 - presenilins, 1:474, 4:201–202
 - proteolysis, 1:474
 - process mechanisms, 4:199, 4:200f
- aspartic proteases, 1:140
- biomarkers, 1:477
- calbindin- D_{28k} , 1:286
- caspase knockouts, 1:392t
- characteristics, 1:473
- chemokine receptors, 1:484–485
- detection approaches, 1:477
- exosomes and microvesicles (EMVs), 2:264
- glutamate receptors, 2:720
- glycogen synthase kinase-3 (GSK3), 2:437
- Golgi-localized endoproteases, 2:481
- inositol trisphosphate (IP_3)/calcium ion (Ca^{2+}) signaling system, 2:588
- lysosomal dysfunction, 4:507t
- mitochondrial DNA (mtDNA), 1:211f
- neurofibrillary tangles, 1:473, 1:474f
- nicotinamide adenine dinucleotide (NAD^+ /NADH) treatments, 1:177
- O-GlcNAc biosynthesis, 3:352
- pathologic changes, 1:473
- pathophysiology
- amyloid β -protein ($A\beta$)-induced mitochondrial fission, 1:478
 - calcium concentrations, 1:478
 - change analyses, 1:478
 - cholesterol, 1:478
 - copper-binding domains, 1:478
- prion diseases, 3:574–575, 3:575t
- protein folding defects, 3:575t, 3:630
- research summary, 1:478
- risk factors, 1:477
- secretases, 4:199–202
- α -secretase, 4:199–200, 4:200f
 - β -secretase, 4:200–201, 4:200f, 4:201f
 - γ -secretase, 4:200f, 4:201–202
- tau proteins
- basic concepts, 1:477–478
 - isoforms, 1:476f
 - neurogenerative disorders, 1:477t
 - tau mutations, 1:475–477
- tetrahydrobiopterin deficiency, 3:669
- thiamine deficiency, 1:208–209
- transient receptor potential (TRP) channels, 4:416
- ubiquitin–proteasome pathway, 3:592, 4:475–476
- vitamin E, 4:547
- Amadori compounds, 2:405, 2:406f, 2:409f
- amaranthus lectin, 2:703f
- ameloblastin, 3:91t
- amelogenin, 3:91t
- amides, 2:408f
- amines
- advanced glycation end products (AGEs), 2:408f
 - amine oxidases (AOs), 1:87–90
 - benzylamine oxidases, 1:89
 - classifications, 1:87
 - copper amine oxidase, 1:88–89
 - diamine oxidases, 1:89
 - hydrogen peroxide (H_2O_2), 1:89–90
 - lysyl oxidase, 1:89
 - monoamine oxidase, 3:185–187
 - aggressive behaviors, 3:187
 - anti-mitochondrial antibodies (AMAs), 3:116
 - catalyzed reactions, 3:185, 3:186f
 - catecholamines, 1:57
 - cellular localization, 3:185, 3:186f
 - encoding processes, 3:185
 - enzyme substrates, 3:185–186, 3:186f, 3:187f
 - flavins, 2:312, 3:185
 - functional role, 1:87
 - inhibitors, 3:186–187
 - pharmacological characteristics, 3:186–187, 3:187f
 - physiological functions, 1:87
 - reaction mechanisms, 1:87
 - structural characteristics, 1:88f, 3:186f
 - plasma amine oxidase, 1:89
 - polyamine oxidase, 1:88, 1:87–88
 - semicarbazide-sensitive amine oxidases, 1:89
 - structural characteristics, 1:88f
 - tissue-bound amine oxidase, 1:89
 - 5-amino-4-imidazolecarboxamide ribonucleotide (AICAR) transformylase, 2:325, 2:325f
 - aminoacetaldehyde, 3:230f
 - amino acids
 - acyl-coenzyme A synthetase, 2:290–291, 2:291f, 2:294–295
 - amino acid radicals, 3:408–409, 3:408f
 - anaplerotic pathways, 1:106f
 - calcium-sensing receptor (CaSR), 1:327
 - citric acid cycle, 1:105
 - classifications, 1:91, 1:92t
 - collagens
 - signature (Gly-XY-) $_n$ sequence
 - basic concepts, 1:545
 - evolutionary origins, 1:548
 - schematic diagram, 1:546f, 1:547f
 - self-assembly control, 1:546
 - cytochrome b_6f complex, 4:338
 - cytochrome c
 - phylogenetic relationships, 1:600–601, 1:600f
 - sequence studies, 1:600
 - digestion and absorption, 2:355f
 - DNA alkylation sites and repair mechanisms, 2:10t
 - elastin
 - composition, 2:192
 - sequence motifs, 2:192
 - essential amino acids, 2:355f
 - O-fucose, 2:454f, 2:455f
 - gamma-carboxyglutamic acid (Gla), 3:588–589, 3:589f
 - O-glucose, 2:454f
 - glycation processes, 2:406
 - glycine receptors (GlyRs), 2:423
 - glycoprotein hormone receptors, 4:389f
 - insect biochemistry, 2:590–591
 - insulin-secreting cells (β -cells), 2:598–600
 - intestinal absorption, 1:92
 - kinesin superfamily proteins (KIFs), 2:680
 - macroautophagy regulation, 1:150t
 - metabolic reactions
 - carbon skeleton, 1:92
 - cellular transport, 1:92–93
 - decarboxylation reactions, 1:92
 - functional roles, 1:93t
 - general discussion, 1:91–92
 - hormonal control
 - anabolic hormones, 1:95, 1:150t
 - catabolic hormones, 1:95–96, 1:96f
 - inter-organ exchanges, 1:93–94
 - intestinal metabolism, 1:92
 - life cycle considerations, 1:96
 - liver metabolism, 1:93
 - muscle metabolism, 1:93, 1:94f
 - nitrogen elimination
 - acid–base homeostasis, 1:95
 - ammoniaogenesis, 1:95
 - general discussion, 1:94
 - ureagenesis, 1:94–95
 - oxidative deamination, 1:92
 - transamination reactions, 1:91–92
 - metalloproteases, 3:87–88
 - molecular structure, 1:91
 - pancreatic endocrine cells, 2:596
 - phage libraries, 3:434t
 - photosynthetic metabolic pathways, 3:498f, 3:499f
 - polyamino acid–matrix interaction affinity tags, 1:62–64
 - pyrrolysine, 4:442–443, 4:443f
 - retinal and choroidal gyrate atrophy, 4:500
 - riboswitches, 4:137
 - selenocysteine, 4:442–443, 4:443f
 - starvation effects, 4:302, 4:304–305
 - sugar nucleotide transporters, 4:350–351
 - tRNA aminoacylation, 4:441
 - tRNA synthetases
 - amino acid selectivity, 4:441
 - tRNA aminoacylation, 4:442
 - vasopressin/oxytocin family, 4:515
 - aminoacyl-adenylate adenosine monophosphate (AA-AMP), 4:442, 4:441
 - aminoacyl-transfer RNA (aa-tRNA)
 - peptidyl transferase center (PTC), 4:131–132, 4:133f

- aminoacyl-transfer RNA (aa-tRNA)
(*Continued*)
protein biosynthesis elongation
 basic concepts, 4:128–129
 elongation phase, 4:122–123, 4:123f,
 4:124f, 4:131, 4:421–423
 schematic diagram, 4:129f
ribosomal subunits, 4:130f, 4:131
structural characteristics, 4:134f
tRNA synthetases, 4:441
5-aminolevulinic acid (ALA), 1:501,
 2:539–540, 2:540f, 3:544, 3:545f, 3:547,
 3:548f
aminolevulinic acid (ALA) synthase (ALAS),
 2:539–540, 2:540f, 2:541–542, 2:542t
aminopeptidases, 1:97–99
 Ape1 protein, 1:151, 1:151f, 2:4f, 2:5t, 2:6,
 2:7
 characteristics, 1:97
 methionine aminopeptidases (MetAPs)
 general discussion, 1:97–99
 isoform functions, 1:98–99
 metal use, 1:98
 structural organization, 1:98, 1:98f
 substrate specificity, 1:98
 N-terminal cotranslational processing,
 1:97–99
 physiological functions, 1:97
 vacuole transport pathways, 1:151,
 4:503–506, 4:504f, 4:505f
aminophospholipid flippases, 2:314–315,
 2:315f, 2:316t
amino-terminal transactivation domain
 (TAD-N), 4:67, 4:68f, 4:70
aminotransferases, 3:564
Amish infantile epilepsy, 2:471t
Amish lethal microcephaly, 1:209
ammonia
 insect biochemistry, 2:591–592, 2:592f
 starvation-based renal ammonia
 formation, 1:109, 4:298, 4:305–306
 urea cycle disorders, 4:494–501
 arginase deficiency (hyperargininemia),
 4:500
 argininosuccinate lyase (AL) deficiency
 (argininosuccinic aciduria),
 4:499
 carbamyl phosphate synthetase (CPS)/
 N-acetylglutamate (NAG)
 synthetase deficiencies, 4:497–498
 citrin deficiency
 adult-onset citrullinemia, 4:499
 characteristics, 4:499
 neonatal intrahepatic cholestasis,
 4:499
 citrullinemia
 argininosuccinate synthetase (AS)
 deficiency, 4:498–499
 citrin deficiency, 4:499
 hyperammonemia
 clinical manifestations, 4:494
 diagnosis, 4:494–495, 4:496f
 inborn errors, 4:494, 4:495t
 long-term therapy considerations,
 4:497
 treatments, 4:495–497, 4:497t
 hyperammonemia–hyperornithinemia–
 homocitrullinemia (HHH)
 syndrome, 4:495f, 4:495t,
 4:500–501
 N-acetylglutamate (NAG) synthetase,
 4:494–495, 4:495f, 4:495t,
 4:497–498
 ornithine transcarbamylase (OTC)
 deficiency, 4:498
 retinal and choroidal gyrate atrophy,
 4:500
 transient hyperammonemia, 4:500
 urinary nitrogen excretion, 4:306f
ammoniaogenesis, 1:95, 4:298, 4:305–306
AMPA (α -amino-3-hydroxy-5-methyl-4-
isoxazolepropionic acid)
AMPA receptors
 excitotoxicity, 1:425f
 functional role, 2:714–716, 2:715t
 kinesin superfamily proteins (KIFs),
 2:681–684, 2:685f
 molecular structure, 2:716t, 2:717–718,
 2:717f, 2:718f
 serine–threonine phosphatases, 4:222f
 functional role, 2:714–716, 2:715t
 molecular structure, 2:716t, 2:717–718,
 2:717f, 2:718f
 schematic diagram, 1:425f
AMP-activated protein kinase (AMPK)
 body weight regulation, 4:58f
 cellular fatty acid metabolism regulation,
 4:56
 diabetes treatments
 metformin, 1:657
 weight loss and exercise, 1:657
 fatty acid metabolism regulation, 2:286
 fatty acid synthesis pathway, 2:275–279,
 2:276f
 insulin signal transduction, 2:604f, 2:605
 ischemic-induced metabolic signaling
 alterations, 3:67–69, 3:69f
 lipogenic gene expression, 2:287f
 microRNA (miRNA)-based regulation,
 4:193
amphibians
 cytochrome *c*, 1:600, 1:600f
 pterin distribution, 3:666
 vasopressin/oxytocin family, 4:516f
amsacrine, 2:167f
amylin, 1:118, 1:268, 1:270–271, 3:78–79
amyloid, 1:100–104
 amyloid β -protein (A β)
 Alzheimer's disease, 1:101t, 1:102,
 1:473–475, 1:474f, 1:476f, 2:481,
 4:199
 amyloid precursor protein (APP)
 processing, 4:199, 4:200f
 calbindin-D_{28k}, 1:286
 caspases, 1:394
 yeast and fungal prions, 3:567–568,
 3:568f
 amyloid enhancement factor (AEF), 1:102
 amyloid precursor protein (APP)
 amyloid β -protein (A β) formation,
 1:286, 1:474f
 caspases, 1:394
 characteristics, 1:473–475
 γ -secretase-based mediation, 4:51f, 4:53
 genetic relationships
 mutation effects, 1:474, 1:474f
 presenilins, 1:474, 4:201–202
 proteolysis, 1:474
 process mechanisms, 4:199, 4:200f
 regulated intramembrane proteolysis
 (RIP), 4:51f, 4:53
 secretases, 1:474, 1:474f
 amyloid proteins, 1:103
 amyloid-related diseases, 1:101t, 1:102,
 1:473, 1:474f, 3:575t, 3:630
 basic concepts, 1:100
 definition, 1:100
 evolutionary development, 1:102
 formation mechanisms
 fibril formation, 3:630
 growth processes, 1:103
 molecular factors, 1:103
 physiological factors, 1:102
 historical research, 1:100–101
 microscopic image, 1:101f
 prion protein (PrP), 1:101t, 1:102
 structural characteristics, 1:103–104,
 1:104f
 toxicity mechanisms, 1:102
 yeast and fungal prions, 3:567–568, 3:568f
amyloplasts, 1:509–510, 1:509f
amyotrophic lateral sclerosis (ALS)
 intermediate filaments, 2:632t, 3:236
 lysosomal dysfunction, 4:507t
 protein folding defects, 3:630
 superoxide dismutases (SODs), 4:354
 transient receptor potential (TRP)
 channels, 4:416
Anabaena spp, 2:296–297, 2:299
anabolic hormones, 1:95, 1:150t
Anacystis, 3:32
anaerobic bacteria, 2:204–209
 adenosine diphosphate (ADP)
 phosphorylation, 2:204, 2:205f
 anaerobic respiration, 2:206–208, 2:207f,
 2:208f
 anoxygenic photosynthesis, 2:209
 catabolic reactions, 2:204
 dicarboxylic acid decarboxylation, 2:208,
 2:208f
 electrogenic transport, 2:208–209, 2:209f
 electron transport phosphorylation,
 2:206–208, 2:207f, 2:208f
 fermentation reactions, 2:204–206, 2:205t,
 2:206f
 H⁺ potential generation, 2:208–209,
 2:209f
 H⁺ translocation, 2:206–208, 2:208f
 proton potential, 2:204, 2:205f, 2:206–208
 redox reactions, 2:206–208, 2:207f, 2:208f
 Na⁺ translocation, 2:206–208, 2:208f
anandamide
 deactivation, 2:195
 intracellular hydrolysis, 2:195
 physiological regulation, 2:194
 synthesis, 2:194, 2:195f
 transport mechanisms, 2:195
anaphase I, 3:19f, 3:20

- anaphase-promoting complex (APC), 1:419–420, 1:419f, 1:420f
- anaphylatoxin, 2:497t
- anaplerotic pathways, 1:105–110
- assessment methods
- carbon-13 (^{13}C) nuclear magnetic spectroscopy, 1:107, 1:108f
 - citric acid cycle pool size changes, 1:107
 - protein expression and activity, 1:107, 1:109f
- citric acid cycle, 1:105, 1:106f
- drainage pathways, 1:107
- environmental stress effects
- diabetic ketoacidosis, 1:109–110, 1:109f
 - exercise and skeletal muscles, 1:108
 - general discussion, 1:110
 - insulin secretion, 1:109–110, 1:109f
 - ischemia, 1:109, 1:109f
 - long-chain fatty acid oxidation defects and myopathies, 1:110
 - small intestine-based glutamine metabolism, 1:108–109
 - starvation-based renal ammonia formation, 1:109
 - substrate-level phosphorylation, 1:109
- research summary, 1:110
- substrates
- aspartate, 1:105, 1:106f
 - assessment methods
 - carbon-13 (^{13}C) nuclear magnetic spectroscopy, 1:107, 1:108f
 - citric acid cycle pool size changes, 1:107
 - protein expression and activity, 1:107, 1:109f - characteristics, 1:105, 1:106f
 - drainage pathways, 1:107
 - fumarate, 1:106, 1:106f
 - glutamate, 1:106, 1:106f, 1:108f, 1:109f
 - α -ketoglutarate, 1:106, 1:106f, 1:108f
 - malate, 1:105, 1:106f
 - oxaloacetate, 1:105, 1:106f, 1:108f, 1:109f
 - pyruvate, 1:105, 1:106f, 1:109f
 - succinyl-coenzyme A (CoA), 1:106, 1:106f, 1:109f
 - tricarboxylic acid cycle, 4:305
- Andersen disease, 2:427t, 2:431t, 2:432
- androgen receptors, 4:308–309, 4:309f
- anemia, 3:564–565
- anesthetic drugs, 2:423, 4:568t
- Angelman syndrome, 1:215–216
- angiogenesis
- c-fes* proto-oncogene (Fes), 1:452–453
 - chemokine receptors, 1:484
 - nuclear factor kappa B (NF- κ B), 3:304t
- angiotensin, 1:111–114
- angiotensin 1-7 (Ang 1-7)
 - Ang (1-7) receptor, 1:114
 - physiological functions, 1:114
 - structure and formation, 1:111, 1:112f
 - angiotensin converting enzyme (ACE)
 - inhibitors, 1:658
 - angiotensin IV (Ang IV)
 - physiological functions, 1:114
 - structure and formation, 1:111, 1:112f, 1:112t
- angiotensin receptors
- Ang I-binding proteins, 1:111
 - angiotensin II (Ang II)
 - amino-acid sequence, 1:112f
 - AT₁ signaling, 1:112f, 1:112t, 1:113, 1:114
 - AT₂ signaling, 1:112f, 1:112t, 1:113, 1:114
 - AT₃ signaling, 1:114
 - physiological functions, 1:114
 - receptor subtypes, 1:111–113, 1:112t
- angiotensin receptor blockers (ARBs), 1:658
- AT₂ signaling
- amino-acid sequence, 1:112f
 - basic concepts, 1:113
 - G-protein-coupled receptors (GPCRs), 1:112t
 - in vitro* studies, 1:113
 - in vivo* studies, 1:113–114
 - physiological functions, 1:114
- G_q family proteins, 2:496–497, 2:497t
- background and characteristics, 1:111
- G-protein-coupled receptors (GPCRs), 1:112t, 2:344t, 2:345f
- structure and formation
- angiotensin 1-7 (Ang 1-7), 1:111, 1:112f
 - angiotensin I (Ang I), 1:111, 1:112f
 - angiotensin II (Ang II), 1:111, 1:112f, 1:112t
 - angiotensin IV (Ang IV), 1:111, 1:112f, 1:112t
- angiotensinogen
- appetite regulation, 4:303
 - physiological functions, 1:114
 - structure and formation, 1:111, 1:112f
- Anguilla anguilla*, 2:702t
- anillin, 1:623, 4:216, 4:217f, 4:218
- animal products, 4:533t
- animals
- lectins, 2:704t
 - microRNA (miRNA) biogenesis, 3:99–100, 3:99f
- anionic glycerophospholipids, 2:415–416, 2:416f
- ankyrin
- actin shape determination, 1:34–35, 1:34f
 - ankyrin repeat-rich membrane spanning adaptor (ARMS), 3:242, 3:242f
- annelids, 3:538
- annetocin, 4:516t
- annexins, 1:275–277, 1:276f
- Anolis*, 3:666
- Anopheles albimanus*, 2:591
- Anopheles gambiae*, 2:591
- Anopheles stephensi*, 2:591
- anoxigenic phototrophs, 4:79
- Antarctic midge, 2:590–591
- anterior pharynx-1, 4:53
- anthramycin, 3:330t
- antiarrhythmic drugs, 4:568t
- antibiotic inhibition, 4:171–172, 4:171f
- antibody–epitope interaction affinity tags, 1:64
- antidiuretic hormones, 1:46t
- antiepileptic drugs, 4:568t
- antigen presenting cells (APCs), 4:373–374, 4:374f, 4:376
- antimitochondrial antibodies, 3:561–562
- antimycin, 1:143t, 1:597f, 1:598
- antiobesity drugs, 3:27
- antioxidants, 2:77, 2:333, 3:174–175
- antiport, 3:51
- anti-synthetase syndrome, 4:443
- antithrombin, 3:586f, 3:587
- antitrypsin, 3:623t
- aorta–gonad–mesonephros (AGM) region, 1:187, 1:188f
- Aos1, 4:468t, 4:469
- β -oxidation
- See alpha-oxidation
- Ape1 protein, 1:151, 1:151f, 2:4f, 2:5t, 2:6, 2:7
- AP endonucleases, 2:3–4, 2:4f, 2:5t
- APEX1/Apex1, 2:5t
- Apicomplexa*, 2:301
- Aplysia californica*, 1:299
- apocrine glands, 2:669t
- apolipoproteins
- apolipoprotein A1, 1:101t
 - apolipoprotein A (apoA), 3:91t
 - apolipoprotein A-IV (APO-IV), 1:116
 - apolipoprotein B (apoB), 4:154, 4:156
 - Apolipoprotein B Editing Catalytic subunit 1, APOBEC-1, 4:156, 4:156f
- apoptosis, 1:423–425
- Bcl2 family, 1:165–166
 - Bcl2 homology (BH) domains, 1:165, 1:166f
 - BH3-only proteins, 1:165, 1:166f
 - Caenorhabditis elegans* programmed cell death, 1:165
 - cell death inhibitors, 1:165, 1:389f, 1:390f
 - family classifications, 1:165, 1:166f
 - therapeutic uses, 1:165–166
- calcium ion (Ca²⁺)
- brain ischemia, 1:423–424
 - excitotoxicity, 1:423–424, 1:425f
 - neurodegenerative diseases, 1:424–425
 - neuronal death, 1:424
 - signal executors, 1:423
 - signaling mechanisms, 1:423
 - signaling subroutines, 1:425
- caspases, 1:388–396
- action mechanisms, 1:391
 - action sequences, 1:389f, 1:391–392, 1:393f
 - activation mechanisms, 1:389, 1:389f, 1:393f
 - apoptotic pathways, 1:389f
 - background information, 1:388
 - Baculovirus IAP repeat (BIR) proteins, 1:393
- caspase 1
- caspase knockouts, 1:392t
 - classifications, 1:388–389
 - inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 - phylogenetic relationships, 1:390f
 - therapeutic outlook, 1:395f

apoptosis (*Continued*)

- caspase 2
 - action sequences, 1:391–392
 - caspase knockouts, 1:392*t*
 - classifications, 1:388–389
 - inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
 - phylogenetic relationships, 1:390*f*
 - sequential activation, 1:393*f*
- caspase 3
 - action sequences, 1:391–392
 - caspase knockouts, 1:392*t*
 - classifications, 1:388–389
 - inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
 - phylogenetic relationships, 1:390*f*
 - regulatory mechanisms, 1:390*f*
 - sequential activation, 1:393*f*
 - therapeutic outlook, 1:395*f*
- caspase 4
 - classifications, 1:388–389
 - inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
 - phylogenetic relationships, 1:390*f*
- caspase 5
 - classifications, 1:388–389
 - inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
 - phylogenetic relationships, 1:390*f*
- caspase 6
 - action sequences, 1:391–392
 - caspase knockouts, 1:392*t*
 - classifications, 1:388–389
 - inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
 - phylogenetic relationships, 1:390*f*
 - regulatory mechanisms, 1:390*f*
 - sequential activation, 1:393*f*
 - therapeutic outlook, 1:395*f*
- caspase 7
 - action sequences, 1:391–392
 - caspase knockouts, 1:392*t*
 - classifications, 1:388–389
 - inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
 - phylogenetic relationships, 1:390*f*
 - regulatory mechanisms, 1:390*f*
 - sequential activation, 1:393*f*
- caspase 8
 - activation mechanisms, 1:394
 - caspase knockouts, 1:392*t*, 1:393
 - characteristics, 1:388–389
 - downstream activators, 1:393, 1:394*f*, 4:377–378
 - inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
 - phylogenetic relationships, 1:390*f*
 - regulatory mechanisms, 1:390*f*
 - sequential activation, 1:393*f*
- caspase 9
 - action sequences, 1:391–392
 - caspase knockouts, 1:392*t*
 - classifications, 1:388–389
 - inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
 - phylogenetic relationships, 1:390*f*
- regulatory mechanisms, 1:390*f*
- sequential activation, 1:393*f*
- therapeutic outlook, 1:395*f*
- caspase 10
 - action sequences, 1:391–392
 - classifications, 1:388–389
 - inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
 - phylogenetic relationships, 1:390*f*
 - sequential activation, 1:393*f*
- caspase 11, 1:392*t*
- caspase 12
 - classifications, 1:388–389
 - phylogenetic relationships, 1:390*f*
- caspase knockouts, 1:392*t*, 1:393, 1:394*f*
- cysteine proteases, 1:591–592
- effector caspases, 1:389*f*, 1:391–392
- historical research, 1:388–389
- inhibitors of apoptosis proteins (IAPs), 1:390*f*, 1:391*t*, 1:392–393, 1:392*t*, 1:590
- pathological implications, 1:394
- regulatory mechanisms, 1:389*f*, 1:390*f*
- signaling mechanisms, 4:457–458, 4:458*f*
- substrate generation, 1:391*t*, 1:392
- substrate recognition, 1:389–390, 1:391*t*, 1:393*f*
- therapeutic outlook, 1:394–395, 1:395*f*
- cathepsins, 1:591–592
- characteristics, 1:423
- cytochrome *c*, 1:601
- DNA damage checkpoints proteins, 1:415
- DNA mismatch repair (MMR)
 - damage-induced cell-cycle growth arrest, 2:52, 2:52*f*
 - tumorigenesis, 2:53, 2:53*f*
- intermediate filaments, 2:636
- mitochondrial inner membrane, 3:148
- mitochondrial outer membrane, 3:161
- mitochondrion-dependent apoptosis, 3:174
- p75 neurotrophin receptor (p75NTR), 3:243–246, 3:244*f*, 3:245*f*
- ribosome biogenesis, 4:121
- tumor necrosis factor receptors (TNFRs), 4:454, 4:456*f*, 4:457–458
- apoptosis signal-regulated kinase-1 (ASK-1), 3:411*f*, 3:412
- apples, 4:533*t*
- APRIL receptor protein, 4:454, 4:456*f*
- aptamers, 4:136
- apurinic/aprimidinic (AP) site, 2:16–17, 2:82
- aquaglyceroporins, 1:123–125, 1:124*f*
- aquaporin water channels, 1:120–126
- aquaglyceroporins, 1:123–125, 1:124*f*
- aquaporin-1 (AQP1)
 - arginine vasopressin (AVP) regulation, 1:123, 1:123*f*
 - functional expression, 1:121*f*
 - historical research, 1:120
 - membrane topology, 1:122*f*
 - physiological functions, 1:121–123, 1:123*f*
- sagittal cross-section schematic diagram, 1:122*f*
- structural characteristics, 1:120–121, 1:122*f*
- aquaporin-2 (AQP2), 1:123, 1:123*f*
- distribution and permeability, 1:123–125, 1:124*t*
- evolutionary relationships, 1:124*f*
- mammalian aquaporin homologs, 1:123–125, 1:124*t*
- research summary, 1:125
- water transport mechanisms, 1:120, 1:121*f*
- aquatic photosynthesis, 3:495
- Aquifex*, 4:146*f*
- Aquifex aeolicus*
 - Cho phylogenetic tree, 3:331*f*
 - copper-transporting ATPases (copper pumps), 3:671–672, 3:672–673
 - respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:341–342
 - small RNAs (sRNAs), 4:254–255
 - thioredoxin-type ferredoxins, 2:297
 - transfer RNA (tRNA), 3:555
 - uvrC* phylogenetic tree, 3:331*f*
- Aquifex fulgidis*, 3:672–673
- Arabidopsis*, 1:125, 3:100
- Arabidopsis thaliana*
 - actin-related proteins (Arps), 1:37*f*
 - acyl-coenzyme A synthetase families, 2:292
 - alternative oxidases (AOXs), 1:146–147
 - ascorbic acid, 4:532
 - aspartic proteases, 1:137, 1:138*f*
 - base excision repair (BER), 2:5*t*
 - brassinosteroid biosynthesis, 1:251
 - copper-transporting ATPases (copper pumps), 3:671–672
 - cytochrome P-450 (CYP), 1:611
 - EF-hand proteins, 1:319*t*
 - glutathione peroxidases, 2:402*f*
 - inwardly rectifying potassium channels (K_{ir}), 4:557–559, 4:558*f*
 - matrix metalloproteinases (MMPs), 3:90
 - multidrug resistance (MDR) membrane proteins, 3:214
 - N-myristoyltransferase (Nmt), 3:642–643
 - O-linked glycosylation, 2:466, 2:467*f*
 - phospholipase D (PLD), 3:474
 - plasma membrane proton (H^+)-ATPases, 3:688, 4:333*f*, 4:334
 - protein-coding genes, 1:75
 - retrotransposons, 3:283, 3:284*f*
 - riboswitches, 4:137–138
 - sarcoendoplasmic reticulum Ca^{2+} ATPase (SERCA), 4:333*f*
 - signal transduction
 - characteristics, 1:252–254
 - chloroplast protein kinase network, 1:514
 - downstream activators, 1:254
 - receptor kinases, 1:253–254, 1:253*f*
- arabinogalactan proteins (AGPs)
 - Arabidopsis thaliana* glycoproteins, 2:467*f*
 - immunohistochemistry, 2:467
 - multigene families, 2:467
 - O-linked glycosylation, 2:465, 2:466*f*

- structural studies, 2:467
- arabinose, 2:466f
- arabinose permease (AraE), 2:388
- arachidonic acid (AA)
- arachidonic acid regulated Ca^{2+} (ARC) channels, 1:127–133
 - characteristics, 1:128–129, 1:128f
 - functional roles, 1:130–131
 - importance, 1:131–133
 - molecular characteristics and activation, 1:129–130
 - ORAI protein family, 1:130, 1:131f, 1:132f
 - reciprocal regulation, 1:129
 - research summary, 1:131–133
 - stromal interacting molecule 1 (STIM1), 1:129–130, 1:130f, 1:132f
 - structural characteristics, 1:130, 1:131f, 1:132f
- biosynthetic pathways, 2:188–189
- biosynthetic relationships, 3:582f
- platelet-activating factor receptor (PAFR), 3:536, 3:536f
- unsaturated fatty acids, 2:288, 2:288f
- arachidonylethanolamide, 2:195f
- 2-arachidonoyl glyceryl ether
- deactivation, 2:195
 - intracellular hydrolysis, 2:195
 - physiological regulation, 2:194–195
 - synthesis, 2:194, 2:195f
 - transport mechanisms, 2:195
- Arachis hypogaea*, 2:702t
- archaea
- chaperonins, 1:456–457
 - chemolithotrophy, 1:486–491
 - membrane-associated energy transduction, 2:204–209, 3:28–35
 - adenosine diphosphate (ADP) phosphorylation, 2:204, 2:205f
 - background information, 3:28
 - catabolic reactions, 2:204
 - energy-transducing functions
 - anaerobic respiration, 2:206–208, 2:207f, 2:208f, 3:29, 3:29t, 3:30f, 3:31f
 - basic concepts, 3:29–30
 - denitrification, 3:29–30
 - dicarboxylic acid decarboxylation, 2:208, 2:208f
 - electrogenic transport, 2:208–209, 2:209f
 - electron transport phosphorylation, 2:206–208, 2:207f, 2:208f
 - fermentations, 3:29
 - ferric iron reduction, 3:30
 - fumarate respiration, 3:30, 3:30f
 - H^+ potential generation, 2:208–209, 2:209f
 - H^+ translocation, 2:206–208, 2:208f
 - redox reactions, 2:206–208, 2:207f, 2:208f
 - schematic diagram, 3:29f, 3:31f
 - Na^+ translocation, 2:206–208, 2:208f
 - sulfur respiration, 3:30 - evolutionary development, 3:34
 - oxygen respiration
 - aerobic chemolithotrophs, 3:31–32
 - aerobic organotrophs, 3:30–31, 3:31f
 - plasma membranes, 3:28
 - proton potential, 2:204, 2:205f, 2:206–208
 - secondary energy transducers, 3:34
 - special mechanisms
 - light utilization, 3:33–34, 3:34f
 - methanogenesis, 3:32–33, 3:33f
 - methanogens, 3:32–34
 - respiratory complexes, 3:34 - reductive reactions, 1:486
 - retrotransposons, 3:283, 3:284f
 - ribozyme distribution, 4:146f
 - RNA polymerases, 2:366t
 - selenoprotein synthesis, 4:213
 - type I DNA topoisomerase, 2:158t, 2:160f
 - type IIB DNA topoisomerase, 2:166
- archaeobacteria
- proteasomes, 3:590, 3:591f
 - type I DNA topoisomerase, 2:158t
- Archaeoglobus fulgidus*, 1:4, 3:671–672, 4:89, 4:92–93, 4:239
- arcuate nucleus (ARC)
- adiposity and satiety signal integration, 3:79f
 - Agouti-related protein (AgRP), 3:77, 3:77f, 3:78f
 - functional role, 3:76
 - leptin expression, 2:257–258, 2:260
 - melanocortin neurons, 3:77
 - neuropeptide Y (NPY), 3:77, 3:77f, 3:78f
- ArdA antirestriction protein, 2:75
- Arf family, 1:134–136
- action mechanisms, 1:134–135, 1:135f
 - Arf-related proteins, 1:134
 - molecular characteristics, 1:134
 - phospholipase D (PLD), 3:472–473, 3:474–475, 3:474f
 - regulation mechanisms, 1:135–136, 1:135f
 - small guanosine phosphatases (GTPases), 4:242, 4:243f
- ArgA Protein, 1:561
- arginase, 1:94–95, 4:494, 4:495f, 4:495t, 4:500
- arginine
- acyl-coenzyme A synthetase, 2:291f
 - advanced glycation end products (AGEs), 2:408f
 - anaplerotic pathways, 1:106f
 - L-arginine, 1:343f
 - arginine fingers, 1:2, 1:3f
 - arginine–glycine–aspartic acid (RGD) peptide, 1:437–438, 1:439f, 2:608
 - citrullination, 4:62
 - classifications, 1:91, 1:92t
 - functional roles, 1:93t
 - glycation pathways, 2:406, 2:406f
 - glycine–phenylalanine–hydroxyproline–glycine–glutamic acid–arginine (GFOGER) sequence, 1:427t, 1:428
 - glycine-serine-arginine (GSR) domain, 3:105f, 3:107
 - hyperammonemia, 4:495–497, 4:497t
 - methylation, 4:60–61
 - mitochondrial DNA (mtDNA), 3:134t
 - N^{ϵ} -hydroxy-L-arginine (NHA), 3:253–254, 3:254f
 - nitric oxide synthase (NOS), 1:343f, 3:254f
 - phage libraries, 3:434t
 - proline–histidine–serine–arginine–asparagine (PHSRN) sequence, 1:426
 - retinal and choroidal gyrate atrophy, 4:500
 - serine–arginine (SR) protein family, 1:76
 - tricarboxylic acid cycle, 4:439f
 - urea cycle, 4:495f
 - vasopressin/oxytocin family, 4:516t
- arginine–glutamic acid–aspartic acid–valine (REDV) sequence, 1:427t
- arginine–glycine–aspartic acid (RGD) sequence, 1:426, 1:427t
- argininosuccinate lyase (AL), 1:94–95, 4:494, 4:495f, 4:495t, 4:499
- argininosuccinate synthetase (AS), 1:94–95, 4:494, 4:495f, 4:495t, 4:498–499
- argininosuccinic acid, 4:495f
- argininosuccinic aciduria, 4:496f, 4:499
- ARGONAUTE (AGO) protein, 3:99–100, 3:99f, 4:253–255
- argyrophilic grain dementia, 1:477t
- ariboflavinosis, 2:375–376, 2:375f
- armadillo family members, 1:637–638, 1:638t
- ARNT (aryl hydrocarbon receptor nuclear translocator), 4:571
- aromatic amino acids, 1:92t, 3:667, 3:668f
- aromatic glycosides, 3:229, 3:230f
- Arp1 protein, 1:37f, 1:38t, 1:39, 1:39f, 2:175, 2:175f
- Arp2 protein, 1:37f, 1:38t
- Arp3 protein, 1:37f, 1:38t
- Arp4 protein, 1:37f, 1:38t, 1:40
- Arp5 protein, 1:37f, 1:38t, 1:40
- Arp6 protein, 1:37f, 1:38t, 1:40
- Arp7 protein, 1:37f, 1:38t, 1:40
- Arp8 protein, 1:37f, 1:38t, 1:40
- Arp9 protein, 1:37f, 1:38t, 1:40
- Arp10 protein, 1:37f, 1:38t
- Arp11 protein, 1:37f, 2:175, 2:175f
- Arp2/3 complex
- actin organization, 1:29–31, 1:30f
 - downstream targets, 2:237f, 2:238, 4:113, 4:114f
 - eukaryotic Arps, 1:37f, 1:38t, 1:39–40, 1:39f
 - leading-edge dynamics, 1:29, 1:20, 1:21f
 - phagocytosis, 3:438–439
 - prokaryotic Arps, 1:39f
- arrestins
- characteristics, 2:487–488
 - domain structures, 2:489–490, 2:489f
 - functional role, 2:488t
 - G protein-coupled receptor kinases (GRKs), 2:485t
 - G-protein-coupled receptors (GPCRs), 1:496f
 - intracellular trafficking, 1:483f
 - olfactory receptors, 3:349f
 - receptor trafficking, 2:488t, 2:489f, 2:490–491

- arrestins (*Continued*)
 signaling scaffolds, 2:488t, 2:491
 spatially organized signalling networks and microdomains, 1:582
 structural characteristics, 2:489–490, 2:489f
 subfamilies, 2:487–488
- arrhythmia incidence, 1:382–383
- arrhythmogenic right ventricular cardiomyopathy (ARVC), 1:638t
- arsenic-based bacterial reductive reactions, 1:487t
- Artemis deficiency, 3:271, 3:272, 3:272f, 3:273t
- arthritis, 3:304–305
- Arthrobacter luteus*, 2:132t
- arylsulfatase A, 4:291
- Ascaris lumbricoides*, 1:140, 3:666
- Ascaris suum*, 3:1, 3:2f, 4:92–93
- ascorbate peroxidase, 3:407–409
- ascorbic acid, 4:530–534
 antioxidant actions, 2:77
 background information, 4:530
 biochemical importance, 4:531–532
 cellular regulation, 4:532–533
 chemical properties, 4:531, 4:531f
 detection methods, 4:533
 digestion and absorption, 4:533
 nutritional importance, 4:531–532
 nutritional requirements, 4:533
 peptidylglycine α -hydroxylating monooxygenase (PHM), 3:396
 recommended dietary allowance (RDA), 4:533–534
 research summary, 4:534
 sources, 4:533t
- Ashbya gossypii*, 4:216f, 4:217
- asparagine
 anaplerotic pathways, 1:106f
 asparagine–proline–alanine (NPA) motif, 1:120–121
 classifications, 1:92t
 glycosylation, 2:449–450, 2:449f, 2:450f
 phage libraries, 3:434t
 proline–histidine–serine–arginine–asparagine (PHSRN) sequence, 1:426
 vasopressin/oxytocin family, 4:516t
- asparagine-linked glycans
 See N-linked glycoproteins
- asparagus peas, 2:702t
- aspartocin, 4:516t
- aspartate
 acyl-coenzyme A synthetase, 2:291f
 anaplerotic pathways, 1:105, 1:106f
 aspartate aminotransferase (AST), 3:565, 4:494
 aspartate transcarbamoylase (ATCase), 1:73
 classifications, 1:92t
 DDE transposases, 3:277–278, 3:278f
de novo biosynthetic pathways, 1:175f
 gluconeogenesis, 2:382f
 insect biochemistry, 2:590–591, 2:591–592
 metalloproteases, 3:87, 3:87f
 phage libraries, 3:434t
- synthetic pathways, 3:712, 3:713f, 3:714f
- tricarboxylic acid cycle, 4:439f
- aspartic acid
 arginine–glycine–aspartic acid (RGD) peptide, 1:437–438, 1:439f, 2:608
 leucine–aspartic acid–valine (LDV) peptide, 2:608
 urea cycle, 4:495f
 WD (tryptophan (W)–aspartic acid (D) dipeptide) domain, 1:24
- aspartic proteases, 1:137–140
 disease processes
 acquired immunodeficiency syndrome (AIDS), 1:139
 Alzheimer's disease, 1:140
 cancer, 1:139
 fungal infections, 1:139–140
 general discussion, 1:139
 hypertension, 1:139
 malaria, 1:139
 genomic sequences, 1:137, 1:138f
 historical research, 1:137
 inhibitors
 naturally occurring inhibitors
 general discussion, 1:140
 IA-3, 1:140
 pepstatin, 1:140
 PI-3, 1:140
 synthetic inhibitors, 1:140
 structural characteristics
 catalytic residues, 1:138–139
 flap structure, 1:140
 overall structure, 1:138, 1:138f
 proenzyme structure, 1:139
 research background, 1:137–139
- Aspergillus nidulans*
 actin-related proteins (Arps), 1:37f
 mitochondrial DNA (mtDNA), 3:133t
- aspirin, 3:305, 3:305t
- astaxanthin, 1:502f
- Asterias amurensis*, 2:544t
- Asterina pectinifera*, 1:334–335
- asthma, 3:304–305, 3:304t
- astrocytes, 1:371–373, 2:643, 2:643f
- Astropecten aranciatus*, 1:334–335, 2:641f
- asvatocin, 4:516t
- ataxia
 ataxia telangiectasia, 3:273
 ataxia telangiectasia-mutated and Rad3-related (ATR) pathway, 1:411t, 1:412–413, 1:413f, 1:414, 2:242f, 2:243
 ataxia telangiectasia mutated (ATM) kinase, 1:411, 1:411t, 1:412f, 1:413–414, 1:413f, 1:415–416, 2:31, 3:269, 3:273t
 ataxia with vitamin E deficiency (AVED), 4:547
 hyperammonemia, 4:494
 polymerase α -primase (Pol α), 2:243–244
 tRNA synthetases, 4:443
- ataxin-3, 4:468t
- ATF6 genes, 2:524f, 2:525, 4:490–491, 4:491f
- Atg8 and Atg12 conjugation systems, 1:151–153, 1:152t, 4:468t, 4:471
- atherosclerosis, 4:547
- atlastin-1, 4:269, 4:269f
- atovaquone, 1:598
- Atox-1 copper chaperone complex, 3:674
- ATP-binding cassette (ABC) transport systems, 1:7–11
 classification, 2:389
 components, 2:389–390
 disease connections, 1:8–9
 DNA mismatch repair (MMR), 2:60
 functional role, 1:8–9, 2:390
 maltodextrin/maltose transport (MalEFGK₂), 2:390, 2:390f
 multidrug exporters and reversal agents, 1:10–11
 multidrug resistance (MDR) membrane proteins, 3:210–214
 ABCG2 (MXR, BCRP), 3:213, 3:214
 BSEP, 3:213, 3:214
 homologous transporters, 3:213
 MDR1
 clinical relevance, 3:210–211
 functional role, 3:213
 high affinity–substrate relationships, 3:213
 physiological functions, 3:213
 polyspecific transport cycle, 3:211–213, 3:212f
 research background, 3:210
 structural characteristics, 3:211–213, 3:211f
 MDR3, 3:213, 3:214
 MRP1 (ABCC1), 3:213, 3:214
 physiological functions
 ABCG2 (MXR, BCRP), 3:214
 ABCG5 and ABCG8, 3:214
 advantages, 3:214
 BSEP, 3:214
 homologous transporters, 3:214
 MDR1, 3:213
 MDR3, 3:214
 MRP1 (ABCC1), 3:214
 research summary, 3:214
 nucleotide binding domains (NBDs), 1:7, 1:8f
 occurrences, 1:7
 phylogeny, 1:7–8
 research summary, 1:11
 structural and biochemical mechanisms, 1:9–10, 1:9f
 structural characteristics, 2:389–390, 2:390f
 substrate specificity, 1:10–11
 topology, 1:7, 1:8f
 transmembrane domains, 1:7, 1:8f
 transport mechanisms, 1:10, 1:10f
- atractylate, 3:163
- atrial natriuretic peptide (ANP), 1:101t, 3:225, 3:226f
- atrial septal defect, 3:223t
- att attachment sites, 1:555–556, 1:558f, 1:560f
- attention-deficit hyperactivity disorder (ADHD), 3:239–240
- attenuators, 4:410
- atypical cadherins, 1:256, 1:256f
- atypical hemolytic uremic syndrome, 3:657t

- AUG triplets, 4:420–421
aurora B/Ipl1 protein, 1:418
autism, 3:669
Autographa californica, 1:393
autoimmune chronic hepatitis, 3:562
autoimmune diseases, 3:561–562
autoimmune treatments, 4:379
autonomously replicating sequences (ARS), 2:109
autophagosomes, 1:149, 1:150f
autophagy, 1:149–154
 endocytic–lysosomal pathways, 3:618, 3:618f
 eukaryotes, 1:153
 macromolecular turnover, 1:153
 mammals
 macroautophagy
 autophagosome formation, 1:149, 1:150f
 basic concepts, 1:149–150, 1:150f
 phagophore maturation, 1:149, 1:150f
 physiological functions, 1:150
 regulatory mechanisms, 1:149, 1:150t
 microautophagy, 1:150
 mTOR complexes, 1:150t, 1:152–153, 3:198, 3:450
 phosphatidylinositol-3-phosphate (PtdIns (3)P), 3:450
 vacuoles, 4:505f, 4:506, 4:506f
 VPS34/Vps34p gene complex, 3:450
 yeasts
 macroautophagy
 Atg proteins, 1:151–153, 1:152t
 cargo selection mechanisms, 1:153
 cytoplasm-to-vacuole targeting (Cvt) pathway, 1:151, 1:151f
 fusion and breakdown mechanisms, 1:153
 model systems, 1:150–151
 molecular mechanisms, 1:151–153, 1:152t
 phagophore assembly site (PAS), 1:152
 regulatory mechanisms, 1:152–153
 microautophagy
 mitophagy, 1:153
 nonselective microautophagy, 1:153
 pexophagy, 1:153
autosomal-dominant hypoparathyroidism (ADH), 1:326, 1:329f
autosomal recessive GPI anchor deficiency, 2:471t
autotrophs, 1:488t
avian leukosis virus (ALV), 4:156–157
avidin, 1:64–65, 1:227
Axin complex, 2:434–435, 2:436f, 2:436t
axonemal dyneins
 binding characteristics, 2:185t, 2:186–187
 inner arm dyneins, 2:185t, 2:187
 outer arm dyneins, 2:185t, 2:187
azide, 1:143t
azo dyes, 2:574–577
Azorhizobium caulinodans, 3:547
Azotobacter, 1:247f
Azotobacter vinelandii, 2:297, 2:301
Azotobacter vinlandii, 1:487t
- B**
Bacillus halodurans, 4:564–565
Bacillus anthracis, 3:331f
Bacillus brevis, 2:132t
Bacillus cereus
 prokaryotic HLR 3-methyladenine (3mA) DNA glycolases, 2:11f, 2:12
 3-methyladenine (3mA) DNA glycolase II, 2:11, 2:11f
 type III DNA restriction and modification enzymes, 2:137t
Bacillus circulans, 1:62
Bacillus coagulans, 2:132t
Bacillus halodurans, 3:331f
Bacillus licheniformis, 4:92
Bacillus megaterium, 2:314
Bacillus stearothermophilus
 DNA glycosylases, 2:18f
 DNA polymerase I, 2:87
 initiation factors, 4:428–430, 4:428f
 MutY repair pathway, 2:58
 PcrA helicase
 characteristics and functional role, 3:261
 DNA unwinding
 active, rolling or hand-over-hand model, 3:264–265, 3:265f
 active versus passive mechanisms, 3:264
 dimeric inch-worm models, 3:265, 3:265f
 inch-worm mechanisms, 3:264, 3:264f, 3:265f
 protein oligomerization, 3:263
 unwinding mechanisms, 3:263–264
 research summary, 3:265
 single-strand DNA (ssDNA)
 translocation, 3:263, 3:264, 3:264f
 structural characteristics
 crystal structures, 3:262, 3:262f
 open and closed confirmations, 3:262–263
 primary structures, 3:261–262
 translocase-based protein displacement, 3:265
 pyruvate dehydrogenases (PDHs), 4:321–322, 4:322f, 4:327f
 type II restriction-modification enzymes, 2:132t
Bacillus subtilis
 Cho phylogenetic tree, 3:331f
 cis-acting riboswitches, 4:251
 DNA polymerase III holoenzyme, 2:92–95, 2:93f
 glmS riboswitch, 4:141
 hexameric enzyme helicases, 2:22
 LexA regulatory protein, 2:710
 membrane-associated energy transduction, 3:30
 messenger RNA (mRNA) decay
 3'-to-5' exoribonucleases, 3:53–55, 3:54f
 5'-to-3' exoribonucleases, 3:56
 model comparisons, 3:54f, 3:56
 polynucleotide phosphorylase (PNPase), 3:53–55, 3:54f, 3:55f
 ribonuclease II (RNase II), 3:53–55, 3:54f
 ribonuclease J1 (RNase J1), 3:56
 ribonuclease J (RNase J), 3:55–56
 ribonuclease Y (RNase Y), 3:56
 non-homologous end joining (NHEJ), 3:266
 photochemical action spectra, 4:275f
 pterin distribution, 3:666
 recombination-dependent DNA replication (RDR), 4:24
 ribosomal RNA (rRNA)
 defective processing consequences, 3:559
 maturation processes
 5S rRNA, 3:557f, 3:559
 16S rRNA, 3:557f, 3:558–559
 transcript structure, 3:556, 3:557f
 23S rRNA, 3:557f, 3:559
 RNA polymerases, 4:167
 sigma(σ)-factors, 4:167–169, 4:233–234, 4:233t
 SpoIVFB protein, 4:51f, 4:53
 succinate:quinone oxidoreductases (SQORs), 4:92
 transfer RNA (tRNA)
 maturation processes, 3:554, 3:555f
 3'-end–RNase process pathway, 3:555f, 3:556
 trp operon
 RNA-binding attenuation protein (TRAP)
 anti-TRAP (AT), 2:363f, 2:364
 bacterial occurrences, 2:364
 characteristics, 2:363, 2:363f
 general discussion, 2:362–364
 hairpin formation, 2:363, 2:363f
 leader transcript, 2:362, 2:362f
 timing factors, 2:363
 tRNA^{Trp}, 2:363f, 2:364
 uvrC phylogenetic tree, 3:331f
bacteria
 anaerobic bacteria, 2:204–209
 adenosine diphosphate (ADP)
 phosphorylation, 2:204, 2:205f
 anaerobic respiration, 2:206–208, 2:207f, 2:208f
 anoxygenic photosynthesis, 2:209
 catabolic reactions, 2:204
 dicarboxylic acid decarboxylation, 2:208, 2:208f
 electrogenic transport, 2:208–209, 2:209f
 electron transport phosphorylation, 2:206–208, 2:207f, 2:208f
 fermentation reactions, 2:204–206, 2:205t, 2:206f
 H⁺ potential generation, 2:208–209, 2:209f
 H⁺ translocation, 2:206–208, 2:208f
 proton potential, 2:204, 2:205f, 2:206–208
 redox reactions, 2:206–208, 2:207f, 2:208f
 Na⁺ translocation, 2:206–208, 2:208f

- bacteria (*Continued*)
- bacterial nucleoid, 3:367–370
 - domain structures, 3:370
 - functional role, 3:370
 - nucleoid-associated proteins (NAPs)
 - characteristics and functional role, 3:367, 3:368f, 3:368t
 - DNA supercoiling, 3:368–369, 3:369f
 - gene expression regulation, 3:367–368
 - topoisomerases, 3:369–370
 - physiological characteristics, 3:370
 - bacterial peroxidases, 3:407–409
 - bioenergetic reactions
 - ATP synthesis, 1:486–487, 1:487f
 - growth categories, 1:488
 - reductants, 1:487, 1:487f, 1:488f, 1:488t
 - terminal electron acceptor (TEA), 1:487, 1:487t
 - calpains, 1:354f
 - chemolithotrophs
 - ATP synthesis, 1:488f, 1:490–491
 - carbon sources, 1:491–492
 - characteristics, 1:488–489
 - ecological characteristics, 1:492
 - electrochemical gradient generation, 1:488f, 1:490–491
 - growth-supporting reductant oxidation
 - carbon monoxide-oxidizing bacteria, 1:490, 1:490t
 - hydrogen-oxidizing bacteria, 1:490, 1:490t
 - metal-oxidizing bacteria, 1:490, 1:491t
 - methanotrophic bacteria, 1:489–490, 1:489t
 - nitrifying bacteria, 1:489, 1:489t
 - sulfur-oxidizing bacteria, 1:490, 1:490t
 - growth yields, 1:492
 - nicotinamide adenine dinucleotide (NAD⁺/NADH) production, 1:491
 - periplasmic release, 1:491
 - proton-related cytoplasmic utilization, 1:491
 - reductive reactions, 1:487t
 - substrate-level phosphorylation, 1:491
 - chemolithotrophy, 1:486–491
 - coenzyme Q, 3:723–724, 3:724f
 - conservative site-specific recombination, 1:555–561
 - basic concepts, 1:1
 - biological characteristics, 1:555–556
 - integration and excision reactions, 1:555–556, 1:556f, 1:560f
 - recombinases
 - reaction mechanisms, 1:556
 - resolvase/invertase mechanisms, 1:558–559, 1:559f
 - serine recombinases, 1:558–559, 1:559f
 - tyrosine recombinases, 1:556–558, 1:557f, 1:558f
 - recombination control mechanisms
 - outcome control, 1:6–7, 1:560f
 - site polarity, 1:559
 - research summary, 1:561
 - cytochrome *bc*₁ complex, 1:593, 1:594f
 - cytochrome *c* oxidases, 1:602–606
 - characteristics, 1:602
 - chemical processes, 1:602, 1:603f
 - experimental research, 1:606
 - green bacteria, 2:508–509, 2:508f
 - heme-copper oxidases, 1:603f, 1:604t, 1:606
 - oxygen binding and reduction, 1:605, 1:605f
 - protein composition, 4:100
 - proton pumping, 1:605–606
 - proton-transfer pathways, 1:603f, 1:604–605, 1:605f
 - reaction cycle, 1:605f
 - structural characteristics, 1:602–604, 1:603f, 1:604t, 4:98, 4:99f
 - Dam methylation, 2:43
 - DNA polymerase I, 2:87–91
 - domain structures, 2:87–88, 2:88f
 - 5' exonuclease, 2:90–91, 2:91f
 - O-helix, 2:88–89, 2:88f
 - phosphoryl transfer reactions, 2:89, 2:89f
 - reaction mechanisms, 2:89, 2:89f
 - reaction specificity
 - base pairing, 2:89–90
 - sugar specificity, 2:90
 - research background, 2:87
 - structural domains, 2:87
 - substrate binding, 2:88–89, 2:88f
 - 3'–5' exonuclease, 2:90
 - DNA replication, 2:114–117
 - DnaA functional domains, 2:121–122, 2:122t
 - leading- and lagging-strand DNA synthesis, 2:92, 2:93f, 2:94f, 2:114, 2:115f, 2:118, 2:119f
 - oriC* origins-based replication, 2:121–125
 - biochemical mechanisms, 2:121–122
 - characteristics, 2:117
 - DnaA–*oriC* complex formation, 2:122–123
 - DnaB–DnaC complex, 2:124
 - DnaB helicase, 2:124
 - DnaC helicase, 2:123–124
 - DNA sequence motifs, 2:121–122, 2:122f
 - factor for inversion stimulation (Fis), 2:122, 2:122f
 - integration host factor (IHF) binding sites, 2:122, 2:122f
 - loading process, 2:124
 - primer foundation, 2:124
 - replisome assembly process, 2:122–123
 - research summary, 2:124–125
 - protein components
 - DNA ligases, 2:115f, 2:116, 2:116t
 - DNA polymerase holoenzymes, 2:92–95, 2:114–115, 2:115f, 2:116t
 - 5'–3' nuclease proteins, 2:115f, 2:116, 2:116t
 - primosomes, 2:115–116, 2:115f, 2:116t
 - requirements, 2:114–116
 - single-strand DNA (ssDNA)-binding proteins, 2:93–94, 2:94f, 2:115f, 2:116, 2:116t
 - replication fork (RF)
 - DNA polymerase holoenzymes, 2:92–95
 - lagging-strand DNA synthesis, 2:92, 2:93f, 2:94f, 2:115f, 2:116–117
 - leading-strand DNA synthesis, 2:92, 2:93f, 2:94f, 2:116, 2:117
 - replisome assembly process
 - characteristics, 2:122–123
 - DiaA protein stabilization, 2:123
 - DnaA-dependent unwinding, 2:123
 - DnaA–*oriC* complex formation, 2:122–123
 - HU protein stabilization, 2:123
 - trombone replication model, 2:114, 2:115f
 - gene transcription regulation
 - antibiotic inhibition, 4:171–172, 4:171f
 - basic concepts, 4:232
 - nucleotide addition cycle (NAC), 4:170–171
 - proofreading processes, 4:170–171
 - sigma(σ)-factors, 4:167–169, 4:168f, 4:233–234, 4:233t
 - transcription cycle
 - basic concepts, 4:168f
 - elongation phase, 4:168f, 4:169–170
 - initiation phase, 4:167–169, 4:168f
 - Rho-mediated termination, 4:168f, 4:170
 - sigma(σ)-factors, 4:232–233
 - termination phase, 4:168f, 4:170
 - green bacteria
 - anoxygenic photosynthesis, 3:32
 - antenna complexes, 3:32
 - bacteriochlorophyll (BChl)
 - biosynthesis, 2:503–504, 2:506f
 - characteristics and classifications, 2:501–503
 - chlorosomes, 2:513–516
 - bacteriochlorophyll (BChl), 2:513, 2:515f
 - carotenoids, 2:514
 - characteristics, 2:502f, 2:504–507
 - energy transfer mechanisms, 2:515–516
 - evolutionary considerations, 2:516
 - homologs, 2:514t
 - molecular structure, 2:514f
 - organizational models, 2:514–515, 2:516f
 - phylogenetic relationships, 2:513
 - proteins, 2:514
 - size and internal structure, 2:513, 2:515f
 - spectral properties, 2:514, 2:515f
 - cytochrome *c*
 - characteristics, 2:510–512
 - cytochrome *bc*₁ complex, 2:511–512, 2:511f
 - cytochrome *c*, 2:510–512
 - cytochrome *c*-554/555, 2:511, 2:511f

- cytochrome *c*, 2:511, 2:511f
 flavocytochrome *c*, 2:511
 newly discovered cytochromes, 2:512
 SoxAX, 2:511, 2:511f
 cytochromes, 2:510–512
 filamentous anoxygenic phototrophic (FAP) bacteria, 2:510
 general discussion, 2:510
 green sulfur bacteria, 2:510, 2:511f
 structural characteristics, 2:511f
 electron transfer chains, 2:508–509, 2:508f
 evolutionary development, 3:34
 Fenna–Matthews–Olson protein, 2:504f, 2:505f, 2:507
 filamentous anoxygenic phototrophic (FAP) bacteria
 alternative complex III, 2:512
 characteristics, 2:501–503, 2:510
 cytochrome *c*-554, 2:512
 cytochromes, 2:510
 electron transfer chains, 2:508–509, 2:508f
 reaction centers, 2:335, 2:507–508, 2:508f
 uncharacterized cytochromes, 2:512
 general discussion, 3:32
 green filamentous bacteria, 2:513, 2:515f
 green sulfur bacteria
 characteristics and classifications, 2:501–503
 chlorosomes, 2:513, 2:515f
 cytochromes, 2:510, 2:511f
 F_x cluster, 2:338
 light harvesting complexes, 2:726
 PscB protein, 2:339–340, 2:339f
 reaction centers, 2:335, 2:337–338
 light harvesting complexes, 2:726
 reaction centers, 2:507–508, 2:508f
 hyperthermophilic eubacteria, 2:161–162
 iron–sulfur (FeS) protein biogenesis
 hydrogenases, 2:663
 iron–molybdenum cofactor (FeMo-co), 2:659f, 2:663
 ISC (iron–sulfur cluster) system
 assembly machinery, 2:662, 2:662f
 nitrogenases, 2:663
 nitrogen fixation (NIF) system, 2:663
 sulfur mobilization machinery, 2:662–663
 lectins, 2:701, 2:702f, 2:702t, 2:704t
 membrane-associated energy transduction, 2:204–209, 3:28–35
 adenosine diphosphate (ADP) phosphorylation, 2:204, 2:205f
 background information, 3:28
 catabolic reactions, 2:204
 energy-transducing functions
 anaerobic respiration, 2:206–208, 2:207f, 2:208f, 3:29, 3:29t, 3:30f, 3:31f
 basic concepts, 3:29–30
 denitrification, 3:29–30
 dicarboxylic acid decarboxylation, 2:208, 2:208f
 electrogenic transport, 2:208–209, 2:209f
 electron transport phosphorylation, 2:206–208, 2:207f, 2:208f
 fermentations, 3:29
 ferric iron reduction, 3:30
 fumarate respiration, 3:30, 3:30f
 H⁺ potential generation, 2:208–209, 2:209f
 H⁺ translocation, 2:206–208, 2:208f
 redox reactions, 2:206–208, 2:207f, 2:208f
 schematic diagram, 3:29f, 3:31f
 Na⁺ translocation, 2:206–208, 2:208f
 sulfur respiration, 3:30
 evolutionary development, 3:34
 oxygen respiration
 aerobic chemolithotrophs, 3:31–32
 aerobic organotrophs, 3:30–31, 3:31f
 phototrophic bacteria
 anoxygenic photosynthesis, 2:209, 2:335, 3:32, 3:693–694, 3:694f, 4:81
 antenna complexes, 3:32, 3:484, 3:484f
 evolutionary development, 3:34
 general discussion, 3:32
 oxygenic photosynthesis, 3:32
 plasma membranes, 3:28
 proton potential, 2:204, 2:205f, 2:206–208
 secondary energy transducers, 3:34
 mesophilic/psychrophilic eubacteria, 2:161
 messenger RNA (mRNA) decay, 3:52–58
 basic concepts, 3:52
 gene expression regulation, 3:56–57, 3:57f
 model comparisons, 3:54f, 3:56
 ribonucleases
 degradosome complex, 3:53
 3'-to-5' exoribonucleases, 3:53–55, 3:54f
 5'-to-3' exoribonucleases, 3:56
 3' exoribonucleases, 3:56
 oligonucleotides, 3:56
 phosphorolysis versus hydrolysis, 3:53–55, 3:55f
 poly(A) polymerase (PAP)-mediated decay, 3:54f, 3:55
 polynucleotide phosphorylase (PNPase), 3:53–55, 3:54f, 3:55f, 3:56–57, 3:57f
 ribonuclease E (RNase E), 3:52–53, 3:53t, 3:54f, 3:56–57, 3:57f
 ribonuclease G (RNase G), 3:53
 ribonuclease III (RNase III), 3:56–57, 3:57f
 ribonuclease J1 (RNase J1), 3:56
 ribonuclease J (RNase J), 3:55–56
 ribonuclease Y (RNase Y), 3:56
 small RNAs (sRNAs), 3:57, 3:57f, 4:249–250
 tmRNA system, 3:58
 translation processes, 3:57–58
 metabolic pathways, 4:249–250
 multidrug resistance (MDR) membrane proteins, 3:214
 non-homologous end joining (NHEJ)
 action mechanisms, 3:267–268, 3:267f
 characteristics, 3:266
 functional role, 3:268
 Ku protein, 3:266–267
 ligase D (LigD), 3:267
 viral infections, 3:268
 nuclear factor kappa B (NF-κB), 3:303t
 periplasmic electron-transport systems, 3:400–406
 complex-independent electron transfer, 3:404
 formate dehydrogenase, 3:402–403, 3:403f
 hydrogenases, 3:403
 iron (Fe(III)) respiration, 3:405
 Mitchellian redox loop mechanism, 3:400–404, 3:402f
 multiheme *c*-type cytochromes, 3:404–405, 3:404f
 nitrate reductase A, 3:403–404, 3:403f
 periplasmic nitrate reductase (Nap), 3:403f, 3:404
 respiratory chain complex III (ubiquinol cytochrome *c* reductase), 3:400, 3:401f
 synthesis, 3:405
 phosphatidylinositol-3-phosphate (PtdIns (3)P), 3:449
 phospholipase D (PLD), 3:473
 phototrophic bacteria
 anoxygenic photosynthesis, 2:209, 2:335, 3:32, 3:693–694, 3:694f
 antenna complexes, 3:32, 3:484, 3:484f
 evolutionary development, 3:34
 general discussion, 3:32
 oxygenic photosynthesis, 3:32, 4:81
 respiration, 4:79–82
 electron transport chains, 4:79, 4:80f
 genetic regulatory networks, 4:80–81, 4:80f
 metabolic characteristics, 4:79
 oxygenic photosynthesis, 4:81
 respiratory chain interactions, 4:79–80
 protein degradation
 adenosine triphosphate (ATP)-dependent proteases, 3:623–624
 functional role, 3:623
 substrate selection, 3:624
 pterin distribution, 3:666
 purple bacteria
 anoxygenic photosynthesis, 2:209, 3:32
 antenna complexes, 3:32
 electron acceptors and donors, 3:693–699
 cytochrome *cbh3* oxidase, 3:694f, 3:697, 3:697t, 4:79, 4:80f
 cytochromes, 3:695–696
 hydrogen (H), 3:698
 nicotinamide adenine dinucleotide dehydrogenase, 3:696
 nitrogen compounds, 3:698, 3:698t
 oxidases, 3:697–698

bacteria (*Continued*)

- photosynthetic reaction centers, 3:696–697, 3:696f
- quinol oxidase, 3:694–695, 3:694f, 3:695f, 3:697–698
- respiratory chain complex III (ubiquinol cytochrome *c* reductase), 3:695
- succinate dehydrogenase, 3:694f, 3:696
- sulfur compounds, 3:698, 3:698t
- electron transport chain
 - anoxygenic photosynthesis, 3:32, 3:693–694, 3:694f
 - membrane transport, 3:693–695
 - RegA/RegB regulation, 3:695, 3:695f
 - respiration, 3:694–695, 3:694f
- evolutionary development, 3:34
- general discussion, 3:32
- metabolic characteristics, 3:693
- oxygenic photosynthesis, 3:32
- photosynthetic reaction centers, 3:700–707
 - anoxygenic photosynthesis, 3:693–694, 3:694f
 - characteristics, 3:700
 - cytochrome *c*₂ oxidation, 3:706–707, 3:706f
 - intramolecular electron transfer, 3:705
 - kinetic mechanisms, 3:703
 - molecular characterization, 3:700–702, 3:702f
 - mutagenesis, 3:705
 - oxidation–reduction potentials, 3:703
 - photosystem II (PSII) complex, 3:707
 - quinone reduction and protonation, 3:704f, 3:706
 - substrate analog and inhibitor complexes, 3:703–705, 3:704f
 - subunit structure and functions, 3:700–702, 3:701f
 - X-ray crystallographic structure, 3:702–703
- recombination-dependent DNA replication (RDR), 4:24
- reductive reactions, 1:486, 1:487t
- regulated intramembrane proteolysis (RIP)
 - cAD1 precursor, 4:51f, 4:52–53
 - RseA, 4:51f, 4:52
 - SpolVFB protein, 4:51f, 4:53
- retrotransposons, 3:283, 3:284f
- ribonucleic acid (RNA)
 - ribosomal RNA (rRNA), 3:554–560
 - Bacillus subtilis*, 3:558–559
 - characteristics, 3:554
 - defective processing consequences, 3:559
 - Escherichia coli*, 3:557–558
 - maturation processes, 3:556, 3:557f
 - proteobacteria, 3:559
 - quality control mechanisms, 3:559–560
 - transcript structure, 3:556, 3:557f
 - RNA polymerases, 2:366t
 - transfer RNA (tRNA), 3:554–560
 - characteristics, 3:554
 - defective processing consequences, 3:559
 - 5'-end–RNase P process pathway, 3:555, 3:555f
 - maturation processes, 3:554, 3:555f
 - quality control mechanisms, 3:559–560
 - structural characteristics, 3:554–555
 - 3'-end–RNase process pathway, 3:555–556, 3:555f
- RNA-binding attenuation protein (TRAP)-mediated transcription attenuation, 2:364
- RNA polymerases, 4:167–172, 4:173–184
 - antibiotic inhibition, 4:171–172, 4:171f
 - core enzyme structure
 - characteristics, 4:174–176, 4:174f
 - conformational flexibility, 4:176
 - domain structures, 4:175–176, 4:175t
 - nonconserved domains, 4:174f, 4:176
 - RNAP channels, 4:175–176
 - elongation complex structure
 - active center rearrangements, 4:180–182, 4:181f
 - catalytic sites, 4:181f, 4:182–183
 - general characteristics, 4:180–183
 - nucleotide addition cycle (NAC), 4:181f, 4:183
 - protein–DNA/RNA interactions, 4:180, 4:181f, 4:182t
 - structural organization, 4:180, 4:181f
 - translocation mechanisms, 4:181f, 4:183
 - enzymatic activities, 4:173
 - future research outlook, 4:183–184
- holoenzyme-promoter DNA complexes
 - closed promoter complex (R_Pc), 4:178f, 4:179, 4:179t
 - open promoter complex (R_Po), 4:178f, 4:179–180, 4:179t
 - structural characteristics, 4:178f, 4:179–180
- holoenzyme structure, 4:176–178, 4:177f
- nucleotide addition cycle (NAC), 4:170–171
- proofreading processes, 4:170–171
- subunit composition, 4:173
- transcription cycle
 - basic concepts, 4:167, 4:168f
 - elongation phase, 4:168f, 4:169–170, 4:173–174
 - initiation phase, 4:167–169, 4:168f, 4:173–174
 - Rho-mediated termination, 4:168f, 4:170
 - sigma(σ)-factors, 4:167–169, 4:168f
 - termination phase, 4:168f, 4:170, 4:173–174
- Tth* σ70 holoenzyme
 - core interactions, 4:176–178, 4:178f
 - domain structures, 4:176–178, 4:177f
 - general characteristics, 4:176–178
 - structural characteristics, 4:176, 4:177f, 4:178f
- selenoprotein synthesis
 - anticodon loop, 4:210, 4:211f
 - general discussion, 4:210–213
 - selenocysteine biosynthesis
 - basic concepts, 4:210–212
 - biosynthetic pathway, 4:211f
 - selenocysteine synthase (SelA), 4:210–211, 4:211–212, 4:211f
 - selenocysteine synthase (SelB)–guanosine triphosphate (GTP)–selenocysteyl-tRNA–selenocysteine insertion sequence (SECIS) element formation, 4:212, 4:213f
 - transfer RNA (tRNA)^{Sec}, 4:210, 4:211f
 - UGA codons
 - ribosomal-level decoding, 4:212–213
 - selenocysteine insertion sequence (SECIS) element, 4:213
 - selenocysteine specificity, 4:213
 - stop codons, 4:212
- site-specific recombination systems, 1:1
- small RNAs (sRNAs), 4:249–255
 - characteristics, 4:249–250
 - cis-acting riboswitches
 - characteristics, 4:250–251
 - RNA ligand sensors, 4:250f, 4:251
 - RNA thermosensors, 4:250, 4:250f
 - cis-encoded RNAs, 4:251, 4:252f
 - nucleic acid-binding protein (NABP), 4:252–253, 4:252f
 - protein modulation, 4:252–253
 - research summary, 4:255
 - trans-encoded RNAs, 4:251–252, 4:252f
- sugar transport, 2:387–390
 - ATP-binding cassette (ABC) transport systems
 - classification, 2:389
 - components, 2:389–390
 - functional role, 2:390
 - maltodextrin/maltose transport (MalEFGK₂), 2:390, 2:390f
 - structural characteristics, 2:389–390, 2:390f
 - background information, 2:387
 - electrochemical cation-gradient-driven symporters
 - classification, 2:388
 - functional role, 2:388
 - lactose permease (LacY), 2:388–389
 - melibiose permease (MelB), 2:389
 - sodium/galactose transport (SgIT), 2:389
 - structural characteristics, 2:388, 2:389f
 - phosphotransferase system (PTS)
 - classification, 2:387
 - components, 2:387, 2:388f
 - functional role, 2:387, 2:388f
 - preferential glucose utilization, 2:387–388
- transcription termination
 - characteristics, 4:407–410
 - elongation complex, 4:408f
 - factor-dependent termination
 - antitermination, 4:410

- elongation factors, 4:168*f*, 4:169–170, 4:409–410
- general discussion, 4:409
- Mfd-mediated termination, 4:409
- regulated termination, 4:410
- Rho-mediated termination, 4:168*f*, 4:170, 4:409
- transcription-coupled DNA repair, 4:409
- hairpin formation, 4:407–409, 4:408*f*
- intrinsic termination, 4:407–409, 4:408*f*
- translation elongation, 4:418–424
 - fidelity, 4:423
 - 50S ribosomal subunit, 4:419–420, 4:420*f*
 - peptidyltransferase center (PTC), 4:419, 4:420*f*
 - protein biosynthesis elongation
 - characteristics, 4:420–423
 - elongation phase, 4:421–423, 4:421*f*
 - initiation phase, 4:420–421, 4:421*f*
 - recycling phase, 4:421*f*, 4:423
 - termination phase, 4:421*f*, 4:423
 - regulation mechanisms, 4:423
 - research summary, 4:423–424
 - ribosome architecture, 4:419–420, 4:420*f*
 - ribosome assembly, 4:418–419
 - ribosome composition, 4:419
 - 70S ribosomal particle, 4:419–420, 4:420*f*
 - 30S ribosomal subunit, 4:419–420, 4:420*f*
 - transcription and translation processes, 4:418
- translation initiation, 4:425–431
 - basic concepts, 4:425
 - initiation complex formation, 4:426–427
 - initiation factors
 - initiation factor 1 (IF1), 4:427–428, 4:427*f*, 4:428*f*
 - initiation factor 2 (IF2), 4:427*f*, 4:428–430
 - initiation factor 3 (IF3), 4:427*f*, 4:428*f*, 4:429*f*, 4:430
 - structure and functions, 4:427–431
 - 70S initiation complex, 4:130*f*, 4:426–427, 4:427*f*
 - 30S initiation complex, 4:130*f*, 4:132*f*, 4:426–427, 4:427*f*, 4:429*f*
 - translation initiation region (TIR), 4:425–426, 4:426*f*
- tryptophan biosynthesis pathway, 4:249*b*
- type I DNA topoisomerase, 2:158*t*
- type II DNA topoisomerase, 2:164–165
- UvrABCD system, 3:328–336
 - bacteria–human conservation model, 3:336
 - damage recognition mechanisms
 - general discussion, 3:333–334
 - padlock model, 3:334–335
 - UvrA₂–DNA complex, 3:333
 - UvrA₂B₂ complex in solution, 3:333
 - UvrA₂–DNA binding, 3:333
 - UvrA dimerization, 3:333
 - UvrB–DNA intermediate complex
 - formation, 3:333–334
 - damage sensing mechanisms
 - basic concepts, 3:334–335, 3:334*f*
 - DNA polymerase I, 3:335
 - dual incision step, 3:334
 - LexA regulatory protein, 2:709–710
 - postincision steps, 3:334
 - UvrB–DNA complex, 3:334–335, 3:334*f*
 - UvrD, 3:334
 - DNA mismatch repair (MMR), 2:45–46
 - protein properties
 - UvrA, 3:330–332, 3:332*f*
 - UvrB, 3:332, 3:332*f*
 - UvrC, 3:332, 3:333*f*
 - substrate ranges, 3:330*t*
 - transcription-coupled repair (TCR), 3:335–336, 3:335*f*
 - UvrABCD endonuclease, 3:328, 3:329*f*, 3:330*t*
 - uvr* genes
 - Cho, 3:330, 3:331*f*
 - historical research, 3:328–330
 - uvrA*, 3:328, 3:329*f*
 - uvrB*, 3:329*f*, 3:330
 - uvrC*, 3:329*f*, 3:330, 3:331*f*
 - Y-family polymerases, 4:480–481
- bacterial luciferase, 1:464, 1:465*f*
- bacteriochlorophyll (BChl)
 - applications, 1:501
 - characteristics, 2:513, 2:515*f*
 - chlorophyll biosynthesis, 1:501
 - chlorosomes, 2:502*f*, 2:504–507
 - Fenna–Matthews–Olson protein, 2:504*f*, 2:505*f*, 2:507
 - green bacteria, 2:337–338, 2:501–503, 2:502*f*, 2:503–504
 - heliobacteria, 2:337–338
 - homologs, 2:514*t*
 - molecular structure, 1:499, 1:499*f*, 2:503*f*, 2:514*f*
 - photosynthetic reaction centers
 - electron acceptors and donors, 3:696–697, 3:696*f*
 - green bacteria, 2:501–503, 2:504*f*, 2:505*f*
 - molecular characterization, 3:700–702, 3:702*f*
 - subunit structure and functions, 3:701*f*
 - type I reaction centers, 2:337–338
 - spectral properties, 1:499–500, 1:500*f*
- bacteriophages
 - bacteriophage λ , 4:410–411
 - LexA regulatory protein, 2:710
 - phage display, 3:431–436
 - display vectors
 - basic concepts, 3:431–432
 - phagemid vectors, 3:432, 3:432*f*
 - phage vectors, 3:432, 3:432*f*
 - polyvalent/monovalent displays, 3:431–432
 - phage-binding analysis, 3:435, 3:435*f*
 - phage-binding selections
 - affinity-based selection, 3:433–435, 3:435*f*
 - binding and elution, 3:433–435
 - preparation methods, 3:433
 - phage libraries
 - characteristics, 3:432–433
 - diversity library design, 3:433, 3:434*t*
 - sequence generation, 3:432–433, 3:433*t*, 3:434*t*
 - research applications, 3:431
 - phage T4, 2:21, 2:22*t*, 2:25–26, 2:114–115, 2:116*t*, 3:577*f*
 - phage T7
 - DNA ligases, 2:29*f*, 2:34*f*
 - DNA polymerase holoenzymes, 2:114–115, 2:116*t*
 - DNA polymerase I, 2:87
 - helicases, 2:21, 2:22*t*, 2:25–26
 - nucleoside triphosphatase (NTPase), 2:22–24
 - recombination-dependent DNA
 - replication (RDR), 4:23–24
 - ring-shaped helicases, 2:21, 2:22*t*
 - site-specific recombination systems, 1:1
 - sliding clamp processivity, 4:240, 4:240*f*
 - transposable bacteriophage (Mu), 3:276*f*, 3:277
- bacteriopheophytins, 3:696–697, 3:696*f*, 3:700, 3:701*f*, 3:702*f*
- bacteriorhodopsin (BR), 3:33–34, 3:34*f*
- Bacteroides*, 2:161
- Bacteroides cellulosolvens*, 4:73–74
- Bacteroides fragilis*, 1:555–556
- Baculovirus IAP repeat (BIR) proteins, 1:393
- BAFF receptor protein, 4:454, 4:456*f*, 4:458–459
- bafilomycin, 1:150*t*
- BAK1 (BRI ASSOCIATED RECEPTOR KINASE 1), 1:253*f*, 1:254
- Bak family, 1:165, 1:166*f*
- Baller–Gerold syndrome (BGS), 4:46
- bananas, 4:533*t*
- BAPTA (1,2-bis(*o*-aminophenoxy)ethane-N, N',N'-tetraacetate), 2:574–577
- barbiturates, 2:723–724
- Bardet–Biedl syndrome, 1:456–457
- Barlow's disease, 4:530
- Barr bodies, 3:314
- Barretia burgdoferi*, 3:331*f*
- Barrier-to-autointegration factor (BAF), 3:314, 3:315*f*
- basal forebrain, 2:172–173
- basal ganglia, 2:172–173
- basic helix-loop-helix (bHLH) domain
 - calcium sensors, 1:311
 - carbohydrate-responsive element binding protein (ChREBP), 1:374–375, 1:375*f*
 - gene expression regulation, 1:314
 - myogenic regulatory factors (MRFs), 1:180–182, 1:181*f*
 - transcription factors, 1:184, 1:184*f*, 1:314
- basophils, 1:188*f*, 1:189
- batrachotoxin, 4:568*t*
- Batten disease, 4:507*t*
- Bax family, 1:165, 1:166*f*
- BbvCI enzyme, 2:131, 2:132*t*
- bc*₁ complex, 1:593–598

*bc*₁ complex (Continued)

action mechanisms
 inhibitors, 1:598
 proton conduction pathways, 1:597–598, 1:597f
 protonmotive Q cycle, 1:596–597, 1:597f, 3:663–664, 3:664f
 anoxygenic photosynthesis, 3:694f
 background information, 4:98
 basic concepts, 1:593
 biochemical free radical chain reactions, 3:170f, 3:172f, 3:173f
 cytochrome *c* interactions, 1:601, 1:601f
 electron transfer complexes, 3:695, 4:93f
 facultative phototrophic respiration, 4:79, 4:80f
 functional role, 1:593, 1:594f, 4:84
 green sulfur bacteria, 1:601f, 2:511f
 mitochondrial ROS, 4:19f, 4:20
 oxygen respiration, 3:30–31
 periplasmic electron-transport systems, 3:400, 3:401f
PET genes, 3:308f
 photosynthesis/respiration, 3:695
 photosynthetic reaction centers, 3:693–694, 3:694f, 3:701f
 plant mitochondrial electron transport chain, 1:145, 1:146f
 protein subunits, 1:593–594
 protonmotive electron transfer chain, 1:470–471, 1:470f
 quinone pool functions, 3:695f
 respiratory electron transport chain, 3:661–662, 3:662f, 3:663–664, 3:664f, 3:694–695, 3:694f, 4:79, 4:80f
 schematic diagram, 4:84f, 4:99f
 spectroscopic properties, 1:594–596, 1:595f
 structural characteristics, 1:596, 1:596f
 thermodynamic properties, 1:595–596, 1:595f

B cells
 B-cell antigen receptors (BCRs), 1:161–164
 antibody–antigen interactions, 1:162
 class switching, 1:163
 diversity mechanisms, 1:161–162
 domain structures, 1:161, 1:162f
 downstream signaling events, 1:162–163, 1:163f
 early signaling events, 1:162, 1:163f
 functional role, 1:161
 gene transcription regulation, 1:162–163, 1:163f
 immune responses, 1:163
 immunoglobulin α (Ig α), 1:161, 1:162, 1:163f
 immunoglobulin β (Ig β), 1:161, 1:162, 1:163f
 immunoreceptor tyrosine-based activated motifs (ITAMs), 1:162, 1:162f, 1:163f
 research summary, 1:163
 signaling pathways, 1:162
 somatic hypermutation, 1:163
 DNA mismatch repair (MMR), 2:64–65
 G₁₂/G₁₃ family functions, 2:345t
 hematopoiesis, 1:188f

immunoglobulin Fc receptors (FcRs), 2:569–571
 lectins, 2:705
 nicotinic acid adenine dinucleotide phosphate (NAADP)-induced calcium ion (Ca²⁺) release, 2:638, 2:638t
 protein tyrosine phosphatase (PTP), 3:649–651, 3:651t
 tumor necrosis factor receptors (TNFRs), 4:459
 V(D)J recombination, 3:272, 3:272f
 BceSI enzyme, 2:137t
 Bcgl enzyme, 2:131, 2:132t, 2:133–134
 Bcl10 family, 4:377–378
 Bcl2 family, 1:165–166
 Bcl2 homology (BH) domains, 1:165, 1:166f
 BH3-only proteins, 1:165, 1:166f, 1:389f
 biochemical free radical chain reactions, 3:173f
Caenorhabditis elegans programmed cell death, 1:165
 cell death inhibitors, 1:165, 1:389f, 1:390f
 family classifications, 1:165, 1:166f
 mitochondrial permeability transition pore, 3:165–166
 mitochondrion-dependent apoptosis, 3:174
 therapeutic uses, 1:165–166
 BCS1 clade, 1:3t, 1:4, 1:4f
 B-DNA (antiparallel right-handed double helix), 2:142
 beans, 4:533t
 Becker muscular dystrophy, 2:471t, 2:474–475
 beef, 4:533t
 beetles, 4:516t
 beets, 2:638t
Belgica antarctica, 2:590–591
 Benson–Calvin cycle
 chloroplasts, 1:507, 1:511–512, 1:512f
 photosynthetic carbon dioxide (CO₂) fixation, 3:494, 3:494f, 3:495f
 benzo[a]pyrene, 3:330t
 benzoate
 medium-chain acyl-CoA synthetase (MCAS), 2:292
 urea cycle, 4:495f
 benzodiazapine, 2:347, 2:348f, 2:722, 2:723–724, 2:725, 2:725f
 benzoquinones
 catecholamine-derived *o*-benzoquinones, 3:727, 3:727f
 chemical characteristics, 3:722–723
 coenzyme Q
 biosynthesis, 3:726f, 3:727
 chemical characteristics, 3:723–724
 functional role
 antioxidant actions, 3:725–727
 electron transfer mechanisms, 3:724–725, 3:724f, 3:725f
 plasma membrane oxidoreductase, 3:725
 Q-cycle, 1:511–512, 3:725, 3:725f
 homologs and analogs, 3:723–724, 3:723t

benzosemiquinones, 3:722–723
 benzyladenine, 1:627, 1:628f
 benzylamine, 3:187f
 benzylamine oxidases, 1:89
 beriberi, 1:202, 1:208
 BES1/BZR1 (bri1-ethylmethane sulphonate suppressor 1/brassinazole-resistant 1), 1:253f, 1:254
 beta adrenergic receptors
 feedback regulation, 1:578–579, 1:578f
 general characteristics, 1:59
 pharmacological/molecular characteristics, 1:59
 schematic diagram, 1:58f
 β -alanin, 3:717–718, 3:717f
 β -aminoisobutyric acid, 3:717–718, 3:717f
 β -arrestins, 4:363–364, 4:363f
 β -carbolines, 2:723–724, 2:724f
 β -carotene
 antioxidant actions, 2:77
 cytochrome *b₆* complex, 4:337f, 4:338
 molecular structure, 1:502f
 β -catenins, 1:53–54, 1:53f
 β -galactosidases, 2:77, 4:291
 β -globin genes, 2:109
 β -glucocerebrosidase, 4:291–292
 β -helix proteins, 2:176
 β -hydroxybutyrate
 cerebral metabolism, 1:371–373
 mitochondrial permeability transition pore, 3:164
 starvation, 4:298–299, 4:302, 4:302f, 4:303–304, 4:305
 β -methylcrotonoyl-coenzyme A (CoA) carboxylase, 1:221, 2:386
 β -oxidation
 fatty acid synthesis pathway, 2:276f, 2:291f
 gluconeogenesis, 2:381
 mitochondrial acyl-CoA/CoA ratio regulation, 1:387
 mitochondrial β -oxidation
 carnitine-dependent cellular uptake and activation, 1:384–385, 1:385f
 mitochondrial carnitine system, 1:385–386
 reaction pathways, 1:386, 1:386–387
 peroxisome proliferator-activated receptor alpha (PPAR α), 3:419–421, 3:420f
 thiamine diphosphate (ThDP), 1:207f
 β -site APP cleaving enzyme (BACE), 4:200–201, 4:201f
 β sliding clamp, 2:54–56, 2:55t, 2:56f
 betaglycan, 1:232, 1:232t
 betaine, 2:722, 2:723f
 betaine homocysteine methyltransferase (BHMT), 2:325f, 2:326
 [BETA] prion, 3:566
 Bethlem myopathy, 3:167
 BIBN4096BS nonpeptide antagonist, 1:265, 1:266f
 bicarbonate, 1:46t, 3:514
 bicuculline, 2:723–724
 BIG protein family, 1:135–136, 1:135f
 bilateral adrenal hyperplasia, 1:584
 bile acids
 cholesterol biosynthetic pathway, 1:517f

- digestion and absorption, 2:358
 peroxisomal metabolic functions, 3:416
 very long-chain acyl-CoA synthetase (VLCAS), 2:293
- bile canaliculi, 1:190–191, 1:191f
- bile salts, 1:167–171
- biochemistry
 bile salt pool, 1:169–170
 biosynthetic pathway, 1:168–169, 1:168f
 characteristics, 1:167–170
 chemical components, 1:169–170
 chemical properties, 1:169
 transport and metabolism regulation, 1:169f
- functional roles, 1:167
- hepatic transport
 characteristics, 1:170
 cholestasis, 1:170
 physiological hepatocellular bile salt transport, 1:170
- therapeutic utilization, 1:170
- bilirubin, 3:547–549, 3:548f
- bi-molecular fluorescence complementation (BiFC), 2:749, 2:753f, 2:754
- bimolecular nucleophilic substitution (S_N2), 2:9
- BIN2 (BRASSINOSTEROID INSENSITIVE 2) kinase, 1:253f, 1:254
- binding immunoglobulin protein (BiP), 1:291t, 1:292–293
- biocuration, 3:616
- biocytin, 1:226
- bioenergetics
 bacteria
 ATP synthesis, 1:486–487, 1:487f
 growth categories, 1:488
 reductants, 1:487, 1:487f, 1:488f, 1:488t
 terminal electron acceptor (TEA), 1:487, 1:487t
 basic concepts, 1:217–220
 general discussion, 1:217
 mitochondrial respiration rate, 1:219–220
 proton circuit, 1:219–220, 1:219f
 thermodynamics
 adenosine triphosphate (ATP) synthesis, 1:218
 bioenergetic interconversions, 1:219
 Gibbs free energy, 1:217–218, 1:218f
 ion electrochemical potential differences, 1:219
 open systems versus closed systems, 1:217–219
 photons, 1:219
 redox reactions, 1:218–219
 glycerolipid synthesis, 2:415
 mitochondrial channels, 3:127
- bioflavonoids, 2:167–168
- BIOGRID database, 3:615t
- bioluminescence, 1:461–466
 definition, 1:461
 intracellular calcium ion (Ca^{2+}) indicators, 2:577, 2:577f
- luciferases
 bacterial luciferase, 1:464, 1:465f
 firefly luciferase, 1:462–465, 1:464f, 2:290–291
Vibrio fischeri, 1:465
- biomass-produced renewable hydrogen, 4:72–75
 background information, 4:72
 challenges, 4:74–75
 future research areas, 4:75
 organisms and fermentation pathways
 dark fermentation, 4:73–74, 4:74f, 4:75f
 photofermentation, 4:74, 4:75f
 process mechanisms, 4:74–75, 4:75f
 redox enzymes
 characteristics, 4:72–73
 hydrogenases (H_2 ases)
 [FeFe]-hydrogenases, 4:72, 4:73f, 4:74f
 [NiFe]-hydrogenases, 4:72–73, 4:73f, 4:74f
 nitrogenases, 4:73
 substrates, 4:73
- bioppterin
 biosynthesis, 3:666–667
 structural characteristics, 3:668f
- biotin, 1:221–225
 biotin acceptor peptides (BAPs), 1:64–65
 biotin domains
 characteristics, 1:222–223
 sequence alignment, 1:223f
 structural characteristics, 1:223–224, 1:223f
 biotin enzymes, 1:221–222, 1:222f
 biotin-mediated carboxyl transfer reactions, 1:222f
 biotin protein ligase (BPL)
 biotin-inducible repressor (BirA), 1:224, 1:224f
 biotinylated proteins, 1:226
 reaction mechanisms, 1:224
 characteristics, 1:221
 protein biotinylation, 1:226–228
 analytical and clinical applications, 1:226
 basic concepts, 1:226
 biosynthesis, 1:226
 biotinylated antibodies, 1:227–228
 chemical synthesis, 1:227, 1:227f
 medical importance, 1:226–227
 research background, 1:221
- birds
 cytochrome *c*, 1:600, 1:600f
 uncoupling proteins (UCPs), 4:485
- bismonoacylglycerolphosphate (BMP), 2:415
 9, 10-bis(phenylethynyl)anthracene, 1:463f
 1,3-bisphosphoglycerate (1,3-BPG), 2:445f, 3:499f
 bisphosphonates, 1:518f, 1:519–520
 bitter taste receptors, 4:367
 BK channel, 4:559
 BK11 (BR11 KINASE INHIBITOR 1), 1:253f, 1:254
 blackberries, 4:533t
Blaps cribrosa, 3:21f
Blastochloris viridis, 1:501, 2:512, 3:696
 bleomycin, 2:2
- blk genes/Blk proteins, 4:294
- Blomstrand's chondrodysplasia, 3:384, 3:389
- blood
 blood-clotting proteases, 3:585–589
 coagulation cascade
 attenuation mechanisms, 3:587
 extrinsic pathway, 3:585, 3:586f
 factor XIII–factor XIIIa conversion, 3:586f, 3:587
 fibrin polymerization, 3:586f, 3:587
 intrinsic pathway, 3:585, 3:586f
 platelet activation, 3:586f, 3:587
 thrombin, 3:586f, 3:587
 structural and functional characteristics
 gamma-carboxyglutamic acid (Gla) domains, 3:588–589, 3:589f
 trypsin/trypsinogen, 3:587–588, 3:588f
 thrombin formation, 3:585, 3:586f
 blood group O antigen, 1:2f
 blood substitutes, 2:535
 hyperammonemia, 4:494–495
 lectin-based blood typing, 2:704–705
 starvation, 4:298
- blood-brain barrier (BBB)
 leptin, 3:76
 MDRI inhibitors, 3:213
 thiamine transport mechanisms, 1:205
- Bloom's syndrome
- BLM protein
 binding proteins, 4:45
 characteristics, 4:45, 4:45f
Saccharomyces cerevisiae Sgs1, 4:45–46
 characteristics, 4:45–46
 DNA secondary structures, 2:147
 hexameric enzyme helicases, 2:26
 mutation effects, 4:44
- blue fluorescent protein (BFP), 2:577–578
- Bohr effect, 2:534
- Bombyx mori*, 3:348–349
- bone, 1:285
- Bonella viridis*, 1:501
- bone morphogenetic proteins (BMPs)
 cAMP responsive element-binding protein 3 (CREB3), 4:52
 cardiac muscle development, 1:183–184, 1:184f
 extracellular matrix (ECM) formation, 1:444
 liver regeneration, 1:192
 neurogenesis, 1:197f, 1:198–199
 signaling pathways, 1:229–239
 Bmp ligands
 Bmp2 and Bmp4, 1:234–235, 1:235t
 circulating Bmp ligands, 1:235–236
 expression regulation, 1:229
 extracellular regulation, 1:229–230, 1:236
 superfamilies, 1:229, 1:230f
 vascular development and disease, 1:234–236, 1:235t
 downstream signaling pathways
 Bmp-activated Smads, 1:238
 Smad4, 1:238
 Smad6, 1:238
 Smad-independent signaling, 1:239

bone morphogenetic proteins (BMPs)

(Continued)

- vascular development and disease, 1:238–239
- extracellular regulation
 - Bmp μ /Cv2 expression, 1:236
 - Bmp ligands, 1:229–230
 - Gremlin expression, 1:236
 - matrix Gla protein (MGP), 1:236
 - vascular development and disease, 1:236
- research summary, 1:239
- signaling co-receptors
 - Alk1, 1:235*t*, 1:237
 - Alk3, 1:235*t*, 1:237
 - betaglycan, 1:232, 1:232*t*
 - Bmpr2, 1:235*t*, 1:236–237
 - characteristics and functional role, 1:231–232, 1:232*t*
 - endoglin, 1:232, 1:232*t*, 1:235*t*, 1:237–238
 - RGM/DRAGON family, 1:232, 1:232*t*
 - vascular development and disease, 1:236–238
- signaling receptors
 - Alk1, 1:235*t*, 1:237
 - Alk3, 1:235*t*, 1:237
 - Bmpr2, 1:235*t*, 1:236–237
 - characteristics and functional role, 1:230–231
 - co-receptors, 1:231–232, 1:232*t*
 - endoglin, 1:235*t*, 1:237–238
 - ligand–receptor interactions, 1:231, 1:231*t*
 - Smad-independent signaling, 1:233–234, 1:239
 - Smad signaling, 1:232–233
 - vascular development and disease, 1:236–238
- Smad signaling
 - characteristics and functional role, 1:232–233
 - nontranscriptional Smad-dependent responses, 1:233
 - phosphatidylinositol-3-phosphate (PtdIns(3)P), 3:450
 - regulation mechanisms, 1:233
 - Smad-dependent transcriptional responses, 1:233
- vascular development and disease
 - Bmp ligands, 1:234–236, 1:235*t*
 - downstream signaling pathways, 1:238–239
 - extracellular regulation, 1:236
 - general discussion, 1:234–239
 - receptors and co-receptors, 1:235*t*, 1:236–238
 - 10T1/2 cell lineage, 1:52–54, 1:53*f*
 - transforming growth factor beta (TGF- β), 1:620
- Bordetella*, 3:285*f*
- Bordetella pertussis*, 2:379*t*, 3:646*f*
- Borrelia*, 4:444
- borrelidin, 4:444
- bortezomib, 3:305, 3:305*t*
- Bos taurus*
 - base excision repair (BER), 2:5*t*
 - N-myristoyltransferase (Nmt), 3:642–643
 - Boveri–Cadenas reaction, 3:169–170
 - bovine complex I, 4:89*t*, 4:341–342, 4:342*t*, 4:345
 - bovine heart mitochondria, 1:604*t*
 - bovine papilloma virus (BPV) E1, 2:21, 2:22*t*, 2:23*f*, 2:25–26
 - bovine spongiform encephalopathy (BSE), 3:574–575, 3:575*t*, 3:630
 - BPAG1/BPAG2 proteins, 1:641, 1:641*t*
 - BPAG1e protein, 1:640*f*, 1:641, 1:641*t*
 - bradykinin, 1:240–243
 - angiotensin receptors, 1:113
 - B1 receptors
 - characteristics and functional role, 1:240–241
 - induction regulation, 1:241
 - knockout mice, 1:241
 - regulation mechanisms, 1:241
 - B2 receptors
 - characteristics and functional role, 1:241–242
 - regulation mechanisms, 1:242
 - signal transduction, 1:242, 1:242*f*
 - B3–B5 receptors, 1:242
 - formation mechanisms, 1:240, 1:241*f*
 - G-protein-coupled receptors (GPCRs), 2:344*t*
 - G α family proteins, 2:496–497, 2:497*t*
 - kallikrein–kininogen–kinin system, 1:240, 1:241*f*
 - nitric oxide synthase (NOS), 1:343–346
 - receptor antagonists, 1:242–243
 - second messenger pathways, 1:43*f*
 - smooth muscle expression, 1:242
- Bradyrhizobium japonicum*, 4:250
- brain
 - acyl-coenzyme A synthetase
 - ACSF proteins, 2:294
 - Bubbligum ACS family (ACSBG), 2:294
 - long-chain acyl-CoA synthetase (LCAS), 2:292–293
 - short-chain acyl-CoA synthetase (SCAS), 2:292
 - very long-chain acyl-CoA synthetase (VLCAS), 2:293–294
 - aquaporin water channels, 1:124*t*
 - carbohydrate-responsive element binding protein (ChREBP) distribution, 1:374
 - cyclic nucleotide-gated (CNG) ion channels, 1:586–587
 - glucose metabolism, 1:368–373
 - cellular metabolic pathways, 1:371–373
 - glucose transporters (GLUTs), 1:369–371
 - metabolic pathways, 1:368, 1:369*f*, 1:370*f*
 - monocarboxylate transporters (MCTs), 1:370*f*, 1:371
 - research summary, 1:373
 - G α family phenotypes, 2:521*t*
 - insulin metabolism, 2:603*f*, 2:606–607
 - ischemia, 1:423–424
 - mammalian glucose transport, 2:392*t*
 - phosphofructokinase-2/fructose bisphosphatase-2 (PFK-2/FBPase-2)-activated phosphorylation, 3:457–458
 - proteinase-activated receptors (PARs), 3:605*t*
- brain-derived neurotrophic factor (BDNF)
 - body weight regulation, 3:77–78, 3:78*f*
 - calbindin-D $_{28k}$, 1:287
 - calcium ion (Ca $^{2+}$) signaling, 3:295
 - functional role, 3:241, 3:242*f*, 3:244*f*
 - neurogenesis, 1:197*f*, 1:198
 - transient receptor potential (TRP) channels, 4:415
- brain-expressed-X-linked 1 (Bex1), 3:243, 3:244*f*
- brain natriuretic peptide (BNP), 3:226, 3:226*f*
- brain research
 - brain diseases, 1:101*t*, 1:102
 - near infrared (NIR) spectroscopy
 - brain functional activation, 4:279–280, 4:279*f*
 - metabolic activity, 4:278, 4:278*f*
 - neuronal signals, 4:280
- branched-chain 2-oxo acid dehydrogenase (BCOAD), 1:206, 1:207*f*
- branched-chain amino acids (BCAAs), 1:244–249
 - brain neuropathology, 1:248
 - branched-chain α -ketoacid dehydrogenase complex (BCKDC)
 - dietary and hormonal regulation, 1:247–248
 - macromolecular organization, 1:246–247, 1:247*f*, 1:247*t*
 - maple syrup urine disease (MSUD), 1:244
 - supermolecular BCAA metabolon, 1:248
 - degradative pathways, 1:244, 1:245*f*
 - interorgan relationships, 1:244–245
 - leucine-based nutritional signals, 1:246, 1:246*f*
 - muscle metabolism, 1:93, 1:94*f*
 - oxidation regulation, 1:245–246
 - supermolecular BCAA metabolon, 1:248
- branched-chain ketoacid-dehydrogenase (BCKA)
 - brain neuropathology, 1:248
 - degradative pathways, 1:244, 1:245*f*
 - interorgan relationships, 1:244–245
 - maple syrup urine disease (MSUD), 1:244, 1:245*f*
 - muscle metabolism, 1:93, 1:94*f*
 - oxidation regulation, 1:245–246
- Brassica napus*, 1:250
- brassinolide
 - brassinosteroid biosynthesis, 1:251, 1:252*f*
 - chemical structure, 1:251*f*
- BRASSINOSTEROID-INSENSITIVE 1 (BRI1), 1:253–254, 1:253*f*
- brassinosteroids (BRs), 1:250–254
 - background and characteristics, 1:250
 - biosynthetic pathway
 - campesterol-to-brassinolide pathway, 1:251, 1:252*f*

cycloartenol-to-campesterol pathway, 1:251
 general discussion, 1:250–251
 mevalonic acid-to-cycloartenol pathway, 1:250–251
 chemical structure, 1:250, 1:251f
 endogenous levels, 1:250
 occurrences, 1:250
 physiological characteristics, 1:252
 research background, 1:250
 signal transduction
 characteristics, 1:252–254
 downstream activators, 1:254
 receptor kinases, 1:253–254, 1:253f
 BRCA1 C-terminal (BRCT) domains, 1:411t, 1:412f, 1:414
 BRCT motifs, 2:29f, 2:31, 2:34f
 break-and-copy models, 4:23
 break-induced replication (BIR), 2:247, 4:29
 breakpoint-cluster region protein (Bcr), 1:451
 breast cancer, 2:305, 4:311–312
 breast cancer type 1 susceptibility protein (BRCA1), 3:272
 brefeldin A (BFA), 1:135–136, 1:135f
 bremanolide, 3:27
 BRENDA database, 3:615t
 brevetoxins, 4:568t
 brevicin, 3:657t, 3:658
 Brg1/hBRM-associated factors (BAFs), 4:63
 bridge helix, 4:174f, 4:175, 4:175t
 Briggs–Haldane equation, 2:217
 brimonidine, 1:58–59
 BRL44408, 1:58–59
 BRM-associated proteins (BAPs), 4:63
Brocadia anammoxidans, 1:489t
 broccoli, 4:533t
 bromo-adjacent homology (BAH) domain, 3:8–9, 3:8f
 bromocriptine, 1:658, 2:173
 bromodeoxyuridine (BrdU), 1:197
 bromodomains, 4:60
 bromopalmitate, 3:647
 brown adipose tissue (BAT)
 adenosine diphosphate (ADP)
 phosphorylation and respiration, 4:483–484, 4:484f
 characteristics and functional role, 4:483
 G_s family proteins, 2:518–519
 histology, 4:483f
 leptin, 2:258–259
 thermogenesis, 4:483–484, 4:484f
 uncoupling proteins (UCPs), 4:483–484, 4:484f, 4:485f
Brucella, 2:43
Brucella suis, 3:331f
 Brussels sprouts, 4:533t
 Bruton's tyrosine kinase (Btk), 3:465–466
 BSK1 (BR SIGNALING KINASE 1), 1:253f, 1:254
 BSU1 (BRI1 SUPPRESSOR 1), 1:253f, 1:254
 Btk family kinases, 1:162, 1:163f
 B-type (brain) natriuretic peptide (BNP), 3:226, 3:226f
 BUB family genes, 1:417, 1:420–421, 1:420f
 See also mitotic checkpoint

BUBR1 protein kinase, 1:417, 1:418–419, 1:419f
 See also mitotic checkpoint
 budding yeasts, 4:216–217, 4:216f
 bullous impetigo, 1:638t
 bullous pemphigoid antigen 1 (BPAG1), 2:624–630
 characteristics, 2:628–629
 deficiency-related diseases, 2:627t, 2:629, 2:629t
 domain structures, 2:625f, 2:626f, 2:628–629
 expression regulation, 2:628
 gene locus, 2:625f, 2:628
 hemidesmosome function, 1:641, 1:641t
 integrin signaling, 2:610–611
 isoform diversity, 2:626f, 2:628
 microtubule-binding domain, 3:105f, 3:107
 molecular interactions, 2:629
 subcellular localization, 2:628
 Burkitt's lymphoma, 3:353, 4:161
 butandiol, 2:204–206
 butanol, 2:204–206
 butyrate, 2:204–206, 2:205t, 2:206f, 4:75f
 butyric acid, 2:204–206, 2:205t
 butyrobetaine, 1:384, 1:385
 Byler disease, 4:334

C

C2-domain proteins, 1:278–280, 1:279f
 C-4a-hydroperoxide, 2:310f, 2:312
 C-4a-hydroxide, 2:310f, 2:312
 C4 pathways
 δ-aminolevulinic acid (ALA) formation, 2:539–540, 2:540f, 3:544
 photosynthetic carbon dioxide (CO₂)-concentrating cycle, 3:495, 3:495f
 photosynthetic carbon dioxide (CO₂) fixation, 3:498
 C5 pathways, 2:539–540, 2:540f, 3:544
 C75, 2:278
 CAAX motif, 4:242, 4:244f
 Cab45/stromal cell derived factor 4, 1:291t, 1:295
 cabbage, 4:533t
 cachexia, 3:27
 cAD1 precursor, 4:51f, 4:52–53
 cadherins, 1:255–260
 adhesion regulation
 cytoplasmic domains
 adaptor proteins, 1:257, 1:258f
 adhesion complex turnover, 1:257–258
 cytoskeletal linkage, 1:257
 signaling functions, 1:258
 structural characteristics, 1:256f
 extracellular domains
 sorting mechanisms, 1:257
 structural characteristics, 1:256–257, 1:256f
 general discussion, 1:256–258
 cadherin signaling, 1:261–264
 action mechanisms
 G₁₂/G₁₃ family, 2:344, 2:345f
 general discussion, 1:262–264
 phosphorylation/dephosphorylation, 1:264, 1:264f
 Rho family GTPase signaling, 1:262–263, 1:263f
 Wnt/β-catenin signaling pathway, 1:263–264, 1:263f, 2:434–435, 2:436f
 functional role
 cadherin level influences, 1:261–262
 cell–cell interactions, 1:261, 1:444–445
 general discussion, 1:261–262
 inside-out signaling, 1:262
 signaling property variations, 1:262
 cell adhesion
 cell–cell interactions, 1:444–445
 cytoplasmic domains
 adaptor proteins, 1:257, 1:258f
 adhesion complex turnover, 1:257–258
 cytoskeletal linkage, 1:257
 signaling functions, 1:258
 structural characteristics, 1:256f
 extracellular domains
 sorting mechanisms, 1:257
 structural characteristics, 1:256–257, 1:256f
 characteristics
 adherens junctions, 1:255, 1:261, 1:262f
 classic cadherins, 1:255, 1:256f, 1:261, 1:262f
 desmosomal cadherins, 1:255–256, 1:256f, 1:259t, 1:261
 E-cadherin, 1:255–256, 1:258–259, 1:259t, 1:261, 1:444–445
 N-cadherin, 1:255–256, 1:258–259, 1:259t, 1:261, 1:444
 P-cadherin, 1:255–256, 1:261
 protocadherins, 1:256, 1:256f, 1:261
 T-cadherin, 1:261
 functional role
 cell and tissue morphogenesis, 1:258–259, 1:259t
 cell migration, 1:444–445
 disease connections, 1:259–260
 general discussion, 1:255–256
 plakoglobin, 1:257
 subfamilies
 atypical cadherins, 1:256, 1:256f
 classic cadherins, 1:255, 1:256f, 1:261, 1:262f
 desmosomal cadherins, 1:255–256, 1:256f, 1:261
 protocadherins, 1:256, 1:256f, 1:261
 cadmium (Cd²⁺)-adenosine triphosphatase (ATPase), 4:334
Caenorhabditis elegans
 actin-interacting protein 1 (Aip1), 1:24
 actin-related proteins (Arps), 1:37f
 acyl-coenzyme A synthetase families, 2:292
 adenosine-to-inosine (A-to-I) RNA editing, 4:154–155
 alkyltransferases, 2:12–13
 AP endonucleases, 2:3–4

Caenorhabditis elegans (Continued)

- aquaporin water channels, 1:125
- aspartic proteases, 1:137, 1:138f
- base excision repair (BER), 2:3–4, 2:5t
- calpains, 1:359
- carbohydrate-responsive element binding protein (ChREBP) distribution, 1:374
- centromere region, 1:447t, 1:448
- cobalamin-dependent methionine synthase, 1:159
- crossover pathways, 2:560
- Csk protein kinase, 4:295–296
- cysteine proteases, 1:591–592
- DNA mismatch repair (MMR), 2:52
- focal adhesion kinase (FAK) family, 2:266f, 2:267–268
- free oligosaccharides (fOSs), 2:331
- glycine receptors (GlyRs), 2:421
- G protein-coupled receptor kinases (GRKs), 2:483
- G-protein-signaling regulators (RGSs), 2:492
- hypoxia inducible factor (HIF), 4:70–71
- IP₃ receptors (IP₃R), 2:653–655
- kinesin superfamily proteins (KIFs), 2:679–681, 2:682f
- matrix metalloproteinases (MMPs), 3:90
- microRNA (miRNA), 3:98–99
- mitochondrial ATP-dependent potassium ion (K⁺) channel (KATP), 3:551
- nicotinamide adenine dinucleotide transhydrogenase, 3:247
- olfactory receptors, 3:346
- ORAI protein family, 4:315
- P2Y receptors, 3:376–377
- parathyroid hormone (PTH)/PTH-related protein (PTHrP) receptor (PPR), 3:384
- peptide amidation, 3:398
- presenilins, 4:202
- programmed cell death, 1:165
- retrotransposons, 3:283
- septin complexes, 4:214–215, 4:217f
- sperm motility, 3:1–2, 3:2f
- Src (sarcoma) family kinases, 4:297
- stromal-interacting molecule (STIM) protein family, 4:315
- sugar nucleotide transporters, 4:350
- unfolded protein response (UPR), 4:488–489
- café-au-lait lesions, 2:519
- caffeine
 - ataxia telangiectasia-mutated (ATM) kinase inhibitors, 3:274
 - UvrABCD endonuclease, 3:330t
- Cajal (coiled) bodies, 3:299–300, 3:300f
- calbindin, 1:284–289
 - calbindin-D_{99k}
 - characterization, 1:287
 - functional role
 - intestines, 1:288
 - reproductive tissues, 1:288
 - knockout mice, 1:288
 - regulation mechanisms, 1:288
 - research summary, 1:288
 - tissue distribution, 1:287, 1:288t
 - calbindin-D_{28k}
 - characterization, 1:284
 - domain structures, 1:285f
 - functional role
 - bone, 1:285
 - intestines, 1:284
 - kidneys, 1:285
 - nervous tissue, 1:286
 - pancreas, 1:285
 - reproductive tissues, 1:285–286
 - knockout mice, 1:287
 - regulation mechanisms
 - 1 α ,25-dihydroxyvitamin D₃ (1,25-(OH)₂D₃), 1:286
 - steroid hormones, 1:286–287
 - research summary, 1:287
 - tissue distribution, 1:284, 1:285t
- characteristics and functional role, 1:284
- EF-hand proteins, 1:319t
- neurogenesis, 1:197f
- research summary, 1:288
- calcilytics, 1:329f, 1:330
- calcimimetics, 1:329f, 1:330
- calcineurin, 1:337–341
 - calcium ion (Ca²⁺) release-activated Ca²⁺ channels (CRAC), 2:587–588, 3:360f
 - calcium ion (Ca²⁺) signaling, 2:720f, 3:293f, 3:294, 3:360f
 - CREB (cyclic adenosine monophosphate (cAMP) responsive binding protein), 1:313
 - domain structures, 1:337, 1:338f
 - EF-hand proteins, 1:319t
 - enzymatic properties
 - calcium/calmodulin-dependent stimulation, 1:337–338
 - enzymatic assays, 1:338
 - substrate specificity, 1:338–339
- functional role
 - gene expression regulation, 1:312, 1:339–340, 1:340f
 - neuronal functions, 1:340
- regulatory mechanisms
 - calcium ion (Ca²⁺) regulation, 1:339
 - regulators of calcineurin (RCANs), 1:339
- Stim1 protein, 3:360f
- subunits and isoforms, 1:337
- calcistorin, 1:291t, 1:293
- calcitonin peptide family, 1:270–274
 - adrenomedullin, 1:265–269
 - binding site distribution, 1:268
 - characteristics, 1:265, 1:271
 - molecular structure, 1:271f
 - amino acid ring structure, 1:271f
 - amylin receptors, 1:268
 - calcitonin receptor-like receptor (CLR), 1:266, 1:267f
 - cellular signaling, 1:268
 - CLR–RAMP2/CLR–RAMP3 association, 1:266–267
 - receptor activity-modifying proteins (RAMPs), 1:266, 1:267f
 - receptor component protein (RCP), 1:268
 - pathophysiology, 1:268–269
 - pharmacological characteristics
 - adrenomedullin receptor-selective drugs, 1:265, 1:266f
 - nonpeptide antagonists, 1:265, 1:266f
 - receptor expression changes, 1:268
 - structural characteristics, 1:265, 1:266f
 - calcitonin receptor-like receptor (CLR)
 - binding site distribution, 1:268
 - cellular signaling, 1:268
 - characteristics and expression, 1:273
 - CLR–RAMP1 association, 1:266, 1:273, 1:273f, 1:273t
 - CLR–RAMP2/CLR–RAMP3 association, 1:266–267, 1:273, 1:273f, 1:273t
 - molecular structure, 1:266, 1:267f
 - pathophysiology, 1:268–269
 - receptor expression changes, 1:268
 - calcitonin receptors (CTRs)
 - calcium homeostasis, 1:328–329
 - distribution, 1:272
 - isoforms, 1:272
 - protein structure, 1:272, 1:272f
 - regulatory mechanisms, 1:272
 - signaling pathways, 1:272
 - calcitonin receptor-stimulating peptide (CRSP), 1:271, 1:271f
 - characteristics, 1:270–271
 - intermedin, 1:271, 1:271f
 - molecular structure, 1:271f
 - peptide amidation, 3:396
 - receptor activity-modifying proteins (RAMPs)
 - binding site distribution, 1:268
 - cellular signaling, 1:268
 - characteristics, 1:273–274
 - CLR–RAMP1 association, 1:266, 1:273, 1:273f, 1:273t
 - CLR–RAMP2/CLR–RAMP3 association, 1:266–267, 1:273, 1:273f, 1:273t
 - molecular structure, 1:266, 1:267f
 - pathophysiology, 1:268–269
 - receptor expression changes, 1:268

- calcium (Ca)
- adenylyl cyclases (ACs), 1:45, 1:45f
 - Alzheimer's disease, 1:478
 - calcieneurin, 1:337–341
 - calcium ion (Ca^{2+}) signaling, 2:720f, 3:293f, 3:294, 3:360f
 - domain structures, 1:337, 1:338f
 - enzymatic properties
 - calcium/calmodulin-dependent stimulation, 1:337–338
 - enzymatic assays, 1:338
 - substrate specificity, 1:338–339
 - functional role
 - gene expression regulation, 1:312, 1:339–340, 1:340f
 - neuronal functions, 1:340
 - regulatory mechanisms
 - calcium ion (Ca^{2+}) regulation, 1:339
 - regulators of calcineurin (RCANs), 1:339
 - subunits and isoforms, 1:337
 - calcium/calmodulin-dependent protein kinase II (CaMKII), 1:297–303
 - autophosphorylation
 - autonomous activity, 1:299
 - calmodulin (CaM) trapping, 1:300
 - neuronal activity monitoring, 1:301, 1:301f
 - structure and functions, 1:299, 1:300f
 - substrate anchoring, 1:300
 - threonine ($\text{Thr}^{305}/\text{Thr}^{306}$)
 - autophosphorylation, 1:300–301
 - cardiac regulation, 1:302
 - characteristics and functional role, 1:297
 - domain structures, 1:298–299, 1:299f
 - intracellular localization, 1:297–298, 1:298f
 - long-term potentiation, 1:302
 - memory and learning regulation, 1:302
 - neurogenesis, 1:199
 - structure and functions, 1:302–303
 - substrates, 1:297–298, 1:298f
 - calcium channel-associated transcriptional regulator (CCAT), 1:314
 - calcium-dependent potassium channels, 3:551–552, 3:551f
 - calcium homeostasis
 - calcium ion (Ca^{2+}) buffering
 - endoplasmic reticulum proteins, 1:290
 - calcium-sensing receptor (CaSR), 1:328–329, 1:329f
 - sodium ion/calcium ion ($\text{Na}^+/\text{Ca}^{2+}$) exchanger, 3:44–48
 - calcium ion (Ca^{2+})-binding domains, 4:257f, 4:259–260, 4:260f
 - cellular calcium ion (Ca^{2+})
 - homeostasis, 1:290, 3:44
 - cytoplasmic domain
 - structure–function relationship, 4:258–261, 4:259f
 - functional role, 3:47, 3:47f, 4:256
 - gene families, 3:45, 3:45f, 3:46f
 - historical research, 3:44
 - NCLX protein superfamily, 3:125
 - NCX1 dimerization, 4:257–258
 - NCX1 helix packing, 4:256–257, 4:258f
 - NCX1 membrane topology, 4:256, 4:257f
 - NCX extracellular domain
 - structure–function relationship, 4:261
 - NCX/NCKX expression, 1:380f, 1:381–382, 1:382f, 3:45–46, 3:45f, 3:46f
 - phosphatidylinositol-4,5-bisphosphate (PIP_2), 2:746–747
 - phototransduction mechanisms, 3:489–490, 3:489f
 - physiological functions, 3:48
 - structural characteristics, 3:45f, 3:46–47, 4:256–261
 - topology model, 3:45f
 - transmembrane domain function, 4:257–258, 4:258f
 - transmembrane segments, 4:256, 4:258f
 - calcium ion (Ca^{2+}), 1:281–283
 - aqueous calcium ion, 1:281
 - arachidonic acid regulated Ca^{2+} (ARC) channels, 1:127–133
 - characteristics, 1:128–129, 1:128f
 - functional roles, 1:130–131
 - importance, 1:131–133
 - molecular characteristics and activation, 1:129–130
 - ORAI protein family, 1:130, 1:131f, 1:132f
 - reciprocal regulation, 1:129
 - research summary, 1:131–133
 - stromal interacting molecule 1 (STIM1), 1:129–130, 1:130f, 1:132f
 - structural characteristics, 1:130, 1:131f, 1:132f
 - binding characteristics
 - binding strengths, 1:281
 - biological salts and ligand complexes, 1:282
 - selectivity, 1:282
 - biological fitness, 1:282–283, 1:283t
 - calbindin- $\text{D}_{28\text{K}}$
 - bone, 1:285
 - kidneys, 1:285
 - nervous tissue, 1:286
 - pancreas, 1:285
 - regulation mechanisms, 1:286–287
 - reproductive tissues, 1:285–286
 - calcineurin regulation, 1:339
 - calcium-binding proteins, 1:275–280
 - annexins, 1:275–277, 1:276f
 - C2-domain proteins, 1:278–280, 1:279f
 - gelsolin, 1:277–278, 1:277f, 1:278f
 - calcium ion (Ca^{2+}) buffering
 - endoplasmic reticulum proteins, 1:290–296
 - background information, 1:290
 - binding immunoglobulin protein (BiP), 1:291t, 1:292–293
 - calreticulin (CRT), 1:291t, 1:292
 - calsequestrin, 1:291t, 1:294
 - Class I binding proteins, 1:291–294, 1:291t
 - Class II binding proteins, 1:291t, 1:295
 - Class I/II binding proteins, 1:291t, 1:294–295
 - endoplasmic reticulum (ER) lumen, 1:291, 1:291t
 - ERcalreticulin/PDI, 1:291t, 1:293
 - ERp44, 1:291t, 1:294
 - ERp57, 1:291t, 1:294
 - ERp72, 1:291t, 1:293
 - glucose-regulated protein (Grp94), 1:291t, 1:292
 - homeostasis control, 1:290
 - P5/PDI, 1:291t, 1:293–294
 - PDI-like proteins, 1:291t, 1:293–294
 - protein disulfide isomerase (PDI), 1:291t, 1:293
 - Stim1 protein, 1:291t, 1:295, 3:360–362, 3:361f
 - Stim2 protein, 3:362
 - calcium ion (Ca^{2+})–calmodulin (CaM)-dependent protein kinases (CaMKs), 1:304–309
 - broad-specificity calcium ion (Ca^{2+})–calmodulin (CaM)-dependent protein kinases (CaMKs), 1:306–308
 - calcium ion (Ca^{2+})–calmodulin (CaM)-dependent protein kinase I (CaMKI), 1:305f, 1:306, 1:311
 - calcium ion (Ca^{2+})–calmodulin (CaM)-dependent protein kinase II (CaMKII), 1:297–303, 1:305f, 1:306–307, 1:308f, 1:311, 1:381
 - calcium ion (Ca^{2+})–calmodulin (CaM)-dependent protein kinase IV (CaMKIV), 1:305f, 1:307, 1:308f, 1:311
 - calcium ion (Ca^{2+})–calmodulin (CaM)-dependent protein kinase kinase (CaMKK), 1:305f, 1:307–308, 1:308f
 - calcium ion (Ca^{2+}) signaling
 - modulation, 1:304, 1:305f, 2:719–720, 2:720f
 - calcium sensors, 1:311
 - domain structures, 1:305f
 - eukaryotic elongation factor 2 kinase (eEF-2K), 1:305f, 1:306
 - gene expression regulation, 1:311
 - general discussion, 1:304
 - myosin light chain kinase (MLCK), 1:305–306, 1:305f
 - narrow-specificity calcium ion (Ca^{2+})–calmodulin (CaM)-dependent protein kinases (CaMKs), 1:305–306
 - neurotransmitter transporters, 3:239
 - phosphorylase kinase (PhK), 1:305f, 1:306
 - research summary, 1:308
 - calcium ion (Ca^{2+})–calmodulin (CaM) regulation

- calcium (Ca) (*Continued*)
- autoinhibitory loop, 1:344, 1:344*t*, 1:345*f*
 - basic concepts, 1:343–346
 - calcium ion (Ca^{2+}) signaling system, 2:719–720, 2:720*f*
 - calcium oscillations, 1:325
 - calcium sensors, 1:311
 - C-terminal extension, 1:344–346, 1:344*t*, 1:345*f*
 - nitric oxide signaling, 3:253–256
 - nitric oxide synthase (NOS)–calmodulin (CaM) binding site, 1:345*f*
 - plasma membrane calcium (Ca^{+})-ATPase pump, 3:525–526, 3:526*f*
 - reductase domain, 1:345*f*
 - structural elements, 1:344*t*, 1:345*f*
 - calcium ion (Ca^{2+}) release-activated Ca^{2+} channels (CRAC)
 - arachidonic acid regulated Ca^{2+} (ARC) channels, 1:127–128, 1:128*f*, 1:129
 - cell proliferation regulation, 2:587–588
 - functional role, 3:359–360, 3:360*f*
 - calcium ion (Ca^{2+}) signaling system
 - brain-derived neurotrophic factor (BDNF), 3:295
 - calcium oscillations, 1:322–325
 - c-fos* gene, 3:293
 - CREB shut-off pathway, 3:295, 3:295*f*
 - cyclic adenosine monophosphate (cAMP) response element (CRE), 3:293, 3:293*f*
 - dendritic/cytoplasmic calcium signals, 3:293*f*, 3:294
 - downstream regulatory element
 - antagonist modulator (DREAM), 3:293, 3:295
 - extrasynaptic NMDA receptors, 3:295, 3:295*f*
 - gene expression regulation, 3:293
 - historical research, 2:653
 - N-methyl-D-aspartate (NMDA), 3:292, 3:295, 3:295*f*
 - nuclear calcium signal, 1:310–311, 3:294–295
 - nuclear factor of activated T cells (NFAT), 3:293, 3:293*f*, 3:294, 3:360*f*
 - serum response factor (SRF), 3:293, 3:293*f*
 - spatial signaling, 3:292–293, 3:293*f*
 - submembranous calcium signals, 3:294
 - calcium-modulated proteins, 1:316–321
 - calcium coordination, 1:316–317
 - EF-hand proteins, 1:317–321, 1:317*f*, 1:318*f*, 1:319*t*
 - functional role, 1:316
 - calcium oscillations, 1:322–325
 - characteristics, 1:322, 1:323*f*
 - decoding methods, 1:325
 - functional role and significance, 1:323–325, 1:324*f*
 - signaling pathways, 1:322–323, 1:324*f*
 - calcium-sensing receptor (CaSR), 1:329–330
 - calcium signal mediators
 - calcium sensors, 1:311
 - protein kinases, 1:311
 - protein phosphatases, 1:312
 - spatial signaling, 3:292–293, 3:293*f*
 - calpains, 1:281, 1:358*f*
 - cell death
 - brain ischemia, 1:423–424
 - excitotoxicity, 1:423–424, 1:425*f*
 - neurodegenerative diseases, 1:424–425
 - neuronal death, 1:424
 - signal executors, 1:423
 - signaling mechanisms, 1:423
 - signaling subroutines, 1:425
 - cerebral metabolism, 1:371–373
 - characteristics, 1:281
 - concentrations, 1:282
 - cyclic adenosine diphosphate (ADP)-ribose (cADPr), 1:331–336
 - basic concepts, 1:331–332
 - calcium ion (Ca^{2+}) stores, 1:332
 - fertilization-induced signaling, 1:333–334, 1:333*f*
 - research summary, 1:336
 - sea urchin egg fertilization, 1:333–334, 1:333*f*
 - signaling mechanisms, 1:332–333
 - starfish egg fertilization, 1:334–335, 1:334*f*, 1:335*f*
 - cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:564, 1:564*f*
 - entry mechanisms, 1:127–128
 - functional role, 1:281, 2:712
 - gene expression regulation, 1:310–315
 - calcium ion (Ca^{2+})-dependent enzymes, 1:314–315
 - calcium sensors, 1:311
 - calcium signal mediators, 1:311–312, 3:292–293, 3:293*f*
 - c-fos* gene, 3:293
 - chromatin-associated proteins, 1:314–315
 - cyclic adenosine monophosphate (cAMP) response element (CRE), 3:293, 3:293*f*
 - downstream regulatory element
 - antagonist modulator (DREAM), 3:293, 3:295
 - final effectors, 1:312–315
 - high-mobility group (HMG) proteins, 1:314–315
 - histones, 1:314–315
 - homeostasis control mechanisms, 1:310
 - nuclear calcium signal, 1:310–311, 3:294–295
 - nuclear factor of activated T cells (NFAT), 3:293, 3:293*f*, 3:294
 - protein kinases, 1:311
 - protein phosphatases, 1:312
 - research summary, 1:315
 - serum response factor (SRF), 3:293, 3:293*f*
 - spatial signaling, 3:292–293, 3:293*f*
 - transcription factors, 1:312–314, 1:313*f*
 - glutamate receptors, 2:712–721
 - AMPA receptors, 1:425*f*, 2:714–716, 2:715*t*, 2:716*t*, 2:717–718, 2:717*f*, 2:718*f*
 - calcium ion (Ca^{2+}) signaling system, 2:719–720, 2:720*f*
 - kainic acid receptors (KARs), 2:718
 - molecular structure, 2:716–717, 2:716*t*, 2:717*f*, 2:718*f*
 - neurodegenerative diseases, 2:720
 - N-methyl-D-aspartate (NMDA), 2:714–716, 2:715*t*, 2:717*f*, 2:718–719, 2:720*f*
 - physiological/pharmacological properties, 2:714–716, 2:715*t*
 - protein structure, 2:717–718, 2:718*f*
 - transcription regulation, 2:719–720, 2:720*f*
 - homeostasis control mechanisms, 1:290, 1:310
 - inositol trisphosphate (IP_3)/calcium ion (Ca^{2+}) signaling system, 2:583–589
 - activation mechanisms, 2:586, 2:587*f*
 - calcium oscillations, 1:322–323, 1:324*f*
 - cardiac cell control, 2:588
 - cell proliferation regulation, 2:587–588
 - cellular control mechanisms, 2:587–588, 2:588*f*
 - characteristics, 2:583
 - domain structures, 2:586, 2:587*f*
 - formation mechanisms, 2:583, 2:584*f*
 - functional role, 2:584, 2:588*f*, 3:359–360, 3:360*f*
 - inositol trisphosphate (IP_3) receptors, 2:584*f*, 2:585*f*, 2:586, 2:587*f*
 - metabolic pathways, 2:583–584, 2:585*f*
 - neuronal calcium ion (Ca^{2+}) signaling, 2:588
 - proteinase-activated receptors (PARs), 3:604
 - receptor agonists, 2:585*f*, 2:586
 - receptor antagonists, 2:586
 - receptor modulation, 2:586–587
 - spatiotemporal patterns, 2:584–587, 2:585*f*
 - insoluble salts, 1:281
 - intercellular calcium ion (Ca^{2+}) waves, 2:613–618
 - active waves, 2:616–617
 - characteristics and functional role, 2:613
 - entry mechanisms, 1:127–128
 - in vivo* conditions, 2:617–618
 - initiation mechanisms, 2:613–614
 - messenger regeneration, 2:617
 - nuclear calcium signal, 1:310–311
 - passive waves, 2:616–617

- phase waves, 2:617
 wave propagation mechanisms, 2:614–616
- intracellular calcium ion (Ca^{2+}) binding proteins, 1:275–280
 annexins, 1:275–277, 1:276f
 C2-domain proteins, 1:278–280, 1:279f
 gelsolin, 1:277–278, 1:277f, 1:278f
 intracellular calcium ion (Ca^{2+}) stores
 adenosine triphosphate (ATP), 1:127–128, 2:482
 anandamide synthesis, 2:194
 2-arachidonoyl glyceryl ether synthesis, 2:194–195
 cyclic adenosine diphosphate (ADP)-ribose (cADPr), 1:332
 functional role, 3:359–360, 3:360f
 G_q family proteins, 2:497–498, 2:498f
 indicators, 2:574–578
 nicotinic acid adenine dinucleotide phosphate (NAADP), 1:332
 nitric oxide signaling, 3:253–256
 sarcoendoplasmic reticulum Ca^{2+} ATPase (SERCA), 1:127–128
 store-operated Ca^{2+} entry (SOCE), 3:364, 3:365f
- intracellular calcium ion (Ca^{2+}) waves, 2:640–647
 basic concepts, 2:640–644, 2:641f
 calcium-induced calcium release (CICR), 2:644–645
 physiological significance, 2:646–647
 propagation mechanisms, 2:644–646
 research summary, 2:647
- intracellular signaling pathways
 calcium oscillations, 1:322–325
 dendrites, 2:713
 functional role, 2:712
 glutamate receptors, 2:714–719
 neuronal stimulation, 2:712–714, 2:713f, 2:714f
 voltage-gated calcium channels, 2:713, 2:714f, 2:720f
 voltage-gated sodium channels, 2:712–713, 2:714f
- messenger functions, 1:282–283, 1:283f
 mitochondrial calcium (Ca^{2+}) transport, 1:347–352
 affinity considerations, 3:123–124
 basic concepts, 1:347
 calcium (Ca^{2+}) uptake, 1:348–350, 1:350f, 3:119t, 3:120–121, 3:120f, 3:121–122, 3:121f
 cell stimulation, 1:347, 3:118–126
 conductivity states, 3:124–125, 3:124f, 3:125f
 disease connections, 3:125–126
 endoplasmic reticulum (ER), 1:348
 functional roles, 1:348–350, 1:350f
 historical research, 1:347, 3:118–126
 homeostasis control mechanisms, 3:124–125, 3:124f, 3:125f
 inner membrane channels, 1:351
 molecular characterization, 3:125
 outer membrane channels, 1:350–351
- permeability transition pore (PTP), 3:124–125, 3:124f
 pH effects, 3:118–120, 3:120f
 release leg, 3:122–123, 3:122f, 3:123f
 research background, 1:348
 research summary, 1:351
 respiratory rate, 3:118–120, 3:119f
 reversible-to-irreversible damage transition, 3:109–110
 schematic diagram, 1:349f, 1:350f
 sodium ion/calcium ion ($\text{Na}^+/\text{Ca}^{2+}$) exchanger, 1:424, 3:125–126
- multiple opioid receptors, 3:357–358
 neurodegenerative diseases, 1:424–425
 nicotinic acid adenine dinucleotide phosphate (NAADP), 2:637–639
 calcium ion (Ca^{2+}) signaling, 1:331–336
 fertilization-induced signaling, 1:333–334, 1:333f
 invertebrates, 2:637–638, 2:638t
 metabolic pathways, 2:638–639
 research summary, 1:336
 sea urchin egg fertilization, 1:333–334, 1:333f
 starfish egg fertilization, 1:334–335, 1:334f, 1:335f
 structural characteristics, 2:637, 2:638f
 two-pore channels, 1:335–336
 vertebrates, 2:638, 2:638t
- nitric oxide synthase (NOS)
 calcium ion (Ca^{2+})–calmodulin (CaM) regulation, 1:343–346, 1:344t, 1:345f, 3:253–254, 3:525–526
 calmodulin (CaM)–nitric oxide synthase (NOS) interactions, 1:305f, 1:342–343
 nitric oxide signaling, 3:253–256
 plasma membrane calcium (Ca^{+})-ATPase pump, 3:525–526
- phosphatidylinositol-4,5-bisphosphate (PIP_2), 2:746–747
 plasma membrane calcium (Ca^{+})-ATPase pump, 3:525–528
 alternative splicing, 3:526–527, 3:527f
- calcium ion (Ca^{2+}) pump reaction cycle, 4:331, 4:332f
 domain structures, 3:525–526, 3:526f
 genetic manipulations, 3:528–529
 isoforms, 3:526–527, 3:527f
 mechanoelectrical transduction (MET) channels, 3:528, 3:528f
 pathology, 3:528
 physiological functions, 3:528, 3:528f
 research summary, 3:529
 tissue distribution, 3:526–527
- rates of exchange, 1:282
 research background, 1:127–128
 sarcoendoplasmic reticulum Ca^{2+} ATPase (SERCA), 2:226–230, 2:231–235
 ATP-binding site, 2:232–233, 2:233f
 calcium ion (Ca^{2+}) binding and catalytic activation, 2:228
- calcium ion (Ca^{2+}) pump reaction cycle, 4:331, 4:332f
 calcium oscillations, 1:324f
 cardiac excitation–contraction (E–C) coupling, 1:380f, 1:381
 catalytic and transport cycles, 2:226–228, 2:227f
 coupling mechanism, 2:228–229
 domain structures, 3:684f, 4:330–331, 4:330f
 $\text{E1} \bullet 2\text{Ca}^{2+} \rightarrow \text{E1P}$ transition/occluded state, 2:232f, 2:233–234, 2:233f
 $\text{E1P} \rightarrow \text{E2P}$ transition/
 aspartylphosphate hydrolysis/
 luminal gate closure, 2:232f, 2:233f, 2:234–235, 2:234f
 $\text{E1P} \rightarrow \text{E2P}$ transition/ Ca^{2+} release, 2:232f, 2:233f, 2:234, 2:234f, 3:684f, 3:689–690, 3:689f
 $\text{E2} \rightarrow \text{E1} \rightarrow \text{E1} \bullet 2\text{Ca}^{2+}$ state, 2:232f, 2:233, 2:233f, 4:330f
 free-energy use, 2:228
 functional role, 4:330–331
 heart failure, 1:381–382, 1:382f
 homeostasis control mechanisms, 1:290, 1:310
 hydrolytic cleavage, 2:228–229
 inhibitors, 2:229
 intercellular calcium ion (Ca^{2+}) waves, 2:614–615, 2:614f
 intracellular calcium ion (Ca^{2+}) stores, 1:127–128
 ion pumping, 2:233–235, 4:330–331
 nucleotide binding, 2:228
 phosphorylation–calcium ion (Ca^{2+}) binding interconversion, 2:229
 phosphoryl transfer, 2:228–229
 physiological considerations, 2:229
 research summary, 2:235, 3:688–689
 ryanodine receptors (RyRs), 4:195
 steady-state behavior, 2:226
 store-operated calcium (SOC) channels, 3:359–360, 3:360f, 4:314–315
 structural characteristics, 2:226–228, 2:227f, 2:231, 2:232f, 4:333f
 substrate-induced conformational fit, 2:228
 transmembrane calcium ion (Ca^{2+}) binding sites, 2:231, 2:232f
- sodium ion/calcium ion ($\text{Na}^+/\text{Ca}^{2+}$) exchanger, 3:44–48
 calcium ion (Ca^{2+})-binding domains, 4:257f, 4:259–260, 4:260f
 cellular calcium ion (Ca^{2+})
 homeostasis, 1:290, 3:44
 cytoplasmic domain
 structure–function relationship, 4:258–261, 4:259f
 functional role, 3:47, 3:47f, 4:256
 gene families, 3:45, 3:45f, 3:46f
 historical research, 3:44
 mitochondrial calcium (Ca^{2+}) transport, 3:125–126
 NCLX protein superfamily, 3:125
 NCX1 dimerization, 4:257–258

- calcium (Ca) (*Continued*)
- NCX1 helix packing, 4:256–257, 4:258f
 - NCX1 membrane topology, 4:256, 4:257f
 - NCX extracellular domain
 - structure–function relationship, 4:261
 - NCX/NCKX expression, 1:380f, 1:381–382, 1:382f, 3:45–46, 3:45f, 3:46f
 - phosphatidylinositol-4,5-bisphosphate (PIP₂), 2:746–747
 - phototransduction mechanisms, 3:489–490, 3:489f
 - physiological functions, 3:48
 - structural characteristics, 3:45f, 3:46–47, 4:256–261
 - topology model, 3:45f
 - transmembrane domain function, 4:257–258, 4:258f
 - transmembrane segments, 4:256, 4:258f
- soluble complex ions, 1:281
- store-operated calcium (SOC) channels, 4:314–320
- Ca²⁺ release-activated Ca²⁺ (CRAC)
 - current, 4:314–315, 4:317–318, 4:318f
 - calcium oscillations, 1:323
 - functional role, 3:359–360, 3:360f, 4:314–315, 4:415
 - inositol trisphosphate (IP₃), 4:314–315
 - insulin-secreting cells (β-cells), 2:597f
 - ORAI protein family, 3:359–366, 4:315, 4:317
 - research summary, 4:319
 - store-operated Ca²⁺ entry (SOCE), 4:314–315
 - stromal-interacting molecule (STIM)
 - protein family, 4:315, 4:415
 - transient receptor potential (TRP)
 - channels, 4:315, 4:318–319, 4:415
- stromal-interacting molecule (STIM)
- protein family
 - calcium oscillations, 1:323
 - cytoplasmic domains, 3:362, 3:362f
 - domain structures, 3:361f, 4:315, 4:316f
 - endoplasmic reticulum (ER) calcium
 - sensors, 3:360–362, 3:361f, 4:316
 - functional role, 3:360–362, 4:315, 4:415
 - localizations, 4:315
 - ORAI tetramers, 4:317–318, 4:318f
 - physiological functions, 4:319
 - posttranslational modification, 4:317
 - research summary, 4:319
 - signal inactivation, 4:318
 - signal initiated oligomerization, 3:360–362, 3:361f, 4:316
 - signal transmitted translocation, 4:317
 - sterile α motif (SAM), 3:360–362, 3:361f, 4:315, 4:316f
 - STIM-mediated channel clustering, 4:317
 - transient receptor potential (TRP)
 - channels, 4:318–319
- structural characteristics, 1:282
- transcription factors
- basic helix-loop-helix (bHLH)
 - domain, 1:314
 - calcium channel-associated
 - transcriptional regulator (CCAT), 1:314
 - CREB (cyclic adenosine monophosphate (cAMP) responsive binding protein), 1:313
 - downstream regulatory element
 - antagonist modulator (DREAM), 1:311, 1:313–314, 1:314f
 - myocyte enhancer factor 2 (MEF2), 1:312, 1:313f
 - nuclear factor kappa B (NF-κB), 1:313
 - nuclear factor of activated T cells (NFAT), 1:312, 1:313f
- troponin, 4:445–449
- dilated cardiomyopathy (DCM), 4:447–449, 4:448f
 - genetic disorders, 4:447–449
 - hypertrophic cardiomyopathy (HCM), 4:447–449, 4:448f
 - inherited cardiomyopathies, 4:447–449
 - knock-in mouse model, 4:449
 - mutations, 4:447–448, 4:448f
 - regulatory functions, 4:446–447, 4:448f, 4:449
 - research background, 4:445
 - restrictive cardiomyopathy (RCM), 4:447–449, 4:448f
 - troponin C, 2:578, 4:445–446, 4:448f, 4:449
 - troponin I, 4:445, 4:448–449, 4:448f
 - troponin T, 4:446, 4:446f, 4:447–448, 4:448f
- voltage- and Ca²⁺-activated seven-transmembrane K⁺ channels (K_{Ca}), 4:558f, 4:559
- voltage-gated calcium channels, 4:560–563
- calcium ion (Ca²⁺) currents, 4:561–563, 4:562f
 - calcium oscillations, 1:322–325
 - calmodulin (CaM) binding, 4:560
 - channel characteristics, 4:561t
 - disease connections, 4:563
 - domain structures, 4:560, 4:561f
 - functional role, 4:560
 - G proteins, 4:561
 - high-voltage-activated channels (HVA), 4:561t
 - isoleucine-glutamine (IQ) motif, 4:560
 - low-voltage-activated channels (LVA), 4:561t
 - molecular properties, 4:560–561
 - phosphorylation, 4:561
 - selectivity, 4:560
- calcium-modulated proteins, 1:316–321
- calcium coordination
- binding kinetics, 1:316
 - pentagonal bipyramidal
 - conformation, 1:316
 - temporal buffering, 1:316–317
- EF-hand proteins
- calmodulin (CaM), 1:318f, 1:319t, 1:321
 - domain structures, 1:318–321, 1:318f
 - evolutionary development, 1:317–318, 1:318f
 - stromal-interacting molecule (STIM)
 - protein family, 3:361–362, 3:361f
 - structural characteristics, 1:317, 1:317f, 1:318f
 - subfamilies, 1:318–321, 1:319t
 - functional role, 1:316
- calcium-sensing receptor (CaSR), 1:326–330
- biochemical characteristics
- agonists, 1:328
 - amino acids, 1:327
 - cell surface dimers, 1:327
 - C-tail, 1:327, 1:327f
 - extracellular domains, 1:326–328, 1:327f
 - extracellular loops, 1:327f
 - functionally important mutations, 1:328
 - intracellular loops, 1:327, 1:327f
 - intracellular signaling pathways, 1:328
 - N-linked glycosylation, 1:327
 - protein kinase C (PKC) feedback
 - regulation, 1:328, 2:583, 2:584f
 - related receptors, 1:326–328
 - structural characteristics, 1:326–328, 1:327f
- calcium homeostasis
- bone, 1:329
 - kidneys and intestines, 1:329
 - parathyroid, 1:329
 - regulatory role, 1:328–329, 1:329f
- functional role
- calcium homeostasis, 1:328–329, 1:329f
 - importance, 1:326
 - nonhomeostatic functions, 1:329–330
 - pharmaceutical considerations
 - calcilytics, 1:329f, 1:330
 - calcimimetics, 1:329f, 1:330
 - general discussion, 1:330
- calcium signaling, 3:292–295
- background information, 3:292
 - brain-derived neurotrophic factor (BDNF), 3:295
 - CREB shut-off pathway, 3:295, 3:295f
 - dendritic/cytoplasmic calcium signals, 3:293f, 3:294
 - extrasynaptic NMDA receptors, 3:295, 3:295f

- gene expression regulation
c-fos gene, 3:293
 cyclic adenosine monophosphate (cAMP) response element (CRE), 3:293, 3:293f
 downstream regulatory element
 antagonist modulator (DREAM), 3:293, 3:295
 nuclear calcium signal, 1:310–311, 3:294–295
 nuclear factor of activated T cells (NFAT), 3:293, 3:293f, 3:294
 serum response factor (SRF), 3:293, 3:293f
 historical research, 2:653
 N-methyl-D-aspartate (NMDA), 3:292, 3:295, 3:295f
 nuclear calcium signal, 1:310–311, 3:294–295
 spatial signaling, 3:292–293, 3:293f
 submembranous calcium signals, 3:294
- calpains, 1:353–361
 characteristics, 1:353, 1:354f
 historical research, 1:353–354
 human calpain homologs, 1:353–354, 1:355t, 1:357f
 m-calpains, 1:353–354, 1:357f, 1:358f
 μ -calpains, 1:353–354, 1:357f
 nomenclature, 1:353–354
 research summary, 1:360–361
 structural characteristics, 1:354f, 1:357f
 structure and functions, 1:354–359, 1:357f, 1:358f
 super-family characteristics
 classifications, 1:354f, 1:359
 structure and functions, 1:354f, 1:359–360
- cardiac excitation–contraction (E–C)
 coupling, 1:379–383
 arrhythmia incidence, 1:382–383
 basic concepts
 action mechanisms, 1:379–381
 action potential (AP), 1:379, 1:380f, 4:566
 cytosolic calcium, 1:379–381, 1:380f
 L-type calcium channels, 1:379–381, 1:380f
 myocytes, 1:379, 1:380f
 myofilaments, 1:380–381, 1:380f
 relaxation mechanisms, 1:381
 delayed after-depolarization (DAD), 1:382–383, 1:382f
 early after-depolarization (EAD), 1:382–383
 heart failure, 1:381–382, 1:382f
 transverse tubule (T-tubule), 1:379–381, 1:380f, 1:382f
 characteristics, 1:281
 intracellular calcium ion (Ca^{2+}), 2:574–578
 adenosine triphosphate (ATP), 1:127–128, 2:482
 anandamide synthesis, 2:194
 2-arachidonoyl glyceryl ether synthesis, 2:194–195
- cyclic adenosine diphosphate (ADP)-ribose (cADPr), 1:332
- indicators
 bioluminescent probes, 2:577, 2:577f
 chemical indicators, 2:574–577, 2:575t, 2:576f
 functional role, 2:574
 genetically encoded indicators, 2:575t, 2:577
 green fluorescent protein (GFP), 2:577–578
 research summary, 2:578–579
- nicotinic acid adenine dinucleotide phosphate (NAADP), 1:332
- nitric oxide signaling, 3:253–256
- sarcoendoplasmic reticulum Ca^{2+} ATPase (SERCA), 1:127–128
- troponin, 4:445–449
 dilated cardiomyopathy (DCM), 4:447–449, 4:448f
 genetic disorders, 4:447–449
 hypertrophic cardiomyopathy (HCM), 4:447–449, 4:448f
 inherited cardiomyopathies, 4:447–449
 knock-in mouse model, 4:449
 mutations, 4:447–448, 4:448f
 regulatory functions, 4:446–447, 4:448f, 4:449
 research background, 4:445
 restrictive cardiomyopathy (RCM), 4:447–449, 4:448f
 troponin C, 2:578, 4:445–446, 4:448f, 4:449
 troponin I, 4:445, 4:448–449, 4:448f
 troponin T, 4:446, 4:446f, 4:447–448, 4:448f
- L-type calcium channels, 1:379–381, 1:380f
- manganese calcium ($\text{Mn}_4\text{Ca}^{2+}$) clusters, 3:514
- mitochondria
 calcium (Ca^{2+}) transport, 1:347–352
 affinity considerations, 3:123–124
 basic concepts, 1:347
 calcium (Ca^{2+}) uptake, 1:348–350, 1:350f, 3:119t, 3:120–121, 3:120f, 3:121–122, 3:121f
 cell stimulation, 1:349f
 conductivity states, 3:124–125, 3:124f, 3:125f
 disease connections, 3:125–126
 endoplasmic reticulum (ER), 1:348
 functional roles, 1:348–350, 1:350f
 historical research, 1:347, 3:118–126
 homeostasis control mechanisms, 3:124–125, 3:124f, 3:125f
 inner membrane channels, 1:351
 molecular characterization, 3:125
 outer membrane channels, 1:350–351
 permeability transition pore (PTP), 3:124–125, 3:124f
 pH effects, 3:118–120, 3:120f
 release leg, 3:122–123, 3:122f, 3:123f
 research background, 1:348
 research summary, 1:351
- respiratory rate, 3:118–120, 3:119f
 reversible-to-irreversible damage
 transition, 3:109–110
 schematic diagram, 1:349f, 1:350f
 sodium ion/calcium ion ($\text{Na}^+/\text{Ca}^{2+}$) exchanger, 1:424, 3:125–126
 reversible-to-irreversible damage
 transition, 3:109–110
- myosin regulation, 3:221–222
- photosystem II (PSII) complex, 3:514
- second messenger pathways, 1:43f
- synapse-to-nucleus communication, 3:292–295
 background information, 3:292
 brain-derived neurotrophic factor (BDNF), 3:295
 CREB shut-off pathway, 3:295, 3:295f
 dendritic/cytoplasmic calcium signals, 3:293f, 3:294
 extrasynaptic NMDA receptors, 3:295, 3:295f
 gene expression regulation
c-fos gene, 3:293
 cyclic adenosine monophosphate (cAMP) response element (CRE), 3:293, 3:293f
 downstream regulatory element
 antagonist modulator (DREAM), 3:293, 3:295
 nuclear calcium signal, 1:310–311, 3:294–295
 nuclear factor of activated T cells (NFAT), 3:293, 3:293f, 3:294
 serum response factor (SRF), 3:293, 3:293f
 N-methyl-D-aspartate (NMDA), 3:292, 3:295, 3:295f
 nuclear calcium signal, 1:310–311, 3:294–295
 spatial signaling, 3:292–293, 3:293f
 submembranous calcium signals, 3:294
 voltage-dependent calcium channels (VDCCs), 2:597f
 voltage-gated calcium channels, 4:560–563
 calcium/calmodulin-dependent protein kinase II (CaMKII), 1:297–298, 1:298f
 calcium channel-associated
 transcriptional regulator (CCAT), 1:314
 calcium ion (Ca^{2+}) currents
 channelopathies, 4:563
 physiological functions, 4:562–563, 4:562f
 channel characteristics, 4:561t
 disease connections, 4:563
 functional role, 2:648–649, 4:560
 glutamate receptors, 2:713, 2:720f
 G proteins, 4:561
 high-voltage-activated channels (HVA), 4:561t
 intracellular signaling pathways, 2:714f
 low-voltage-activated channels (LVA), 4:561t
 molecular properties
 calmodulin (CaM) binding, 4:560

- calcium (Ca) (*Continued*)
 domain structures, 4:560, 4:561f
 isoleucine-glutamine (IQ) motif, 4:560
 selectivity, 4:560
 phosphorylation, 4:561
 schematic diagram, 2:649f
 transport modulation, 2:746–747
- Calcium green, 2:575t, 2:576–577
- calcyphosine, 1:319t
- Caldalkalibacillus thermarum*, 2:273–274
- caldariella quinone, 4:94f
- calerythrin, 1:319t
- calflagin, 1:319t
- calicheamicin, 2:2
- calmodulin (CaM)
 adenylyl cyclases (ACs), 1:45, 1:45f
 calcium/calmodulin-dependent protein kinase II (CaMKII), 1:297–303
 autophosphorylation
 autonomous activity, 1:299
 calmodulin (CaM) trapping, 1:300
 neuronal activity monitoring, 1:301, 1:301f
 structure and functions, 1:299, 1:300f
 substrate anchoring, 1:300
 threonine (Thr³⁰⁵/Thr³⁰⁶)
 autophosphorylation, 1:300–301
 cardiac regulation, 1:302
 characteristics and functional role, 1:297
 domain structures, 1:298–299, 1:299f
 intracellular localization, 1:297–298, 1:298f
 long-term potentiation, 1:302
 memory and learning regulation, 1:302
 neurogenesis, 1:199
 structure and functions, 1:302–303
 substrates, 1:297–298, 1:298f
- calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinases (CaMKs), 1:304–309
 broad-specificity calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinases (CaMKs)
 calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinase I (CaMKI), 1:305f, 1:306, 1:311
 calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinase II (CaMKII), 1:305f, 1:306–307, 1:308f, 1:311, 1:381
 calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinase IV (CaMKIV), 1:305f, 1:307, 1:308f, 1:311
 calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinase kinase (CaMKK), 1:305f, 1:307–308, 1:308f
 calcium/calmodulin-dependent protein kinase II (CaMKII), 1:297–303, 1:311
 calcium ion (Ca²⁺) signaling
 modulation, 1:304, 1:305f, 2:719–720, 2:720f
 calcium sensors, 1:311
- domain structures, 1:305f
- eukaryotic elongation factor 2 kinase (eEF-2K), 1:305f, 1:306
- gene expression regulation, 1:311
- general discussion, 1:304
- myosin light chain kinase (MLCK), 1:305–306, 1:305f
- narrow-specificity calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinases (CaMKs), 1:305–306
- neurotransmitter transporters, 3:239
- phosphorylase kinase (PhK), 1:305f, 1:306
- research summary, 1:308
- calcium ion (Ca²⁺)–calmodulin (CaM) regulation
 autoinhibitory loop, 1:344, 1:344t, 1:345f
 basic concepts, 1:343–346
 calcium ion (Ca²⁺) signaling system, 2:719–720, 2:720f
 calcium oscillations, 1:325
 calcium sensors, 1:311
 C-terminal extension, 1:344–346, 1:344t, 1:345f
 nitric oxide signaling, 3:253–256
 nitric oxide synthase (NOS)–calmodulin (CaM) binding site, 1:345f
 plasma membrane calcium (Ca⁺)-ATPase pump, 3:525–526, 3:526f
 reductase domain, 1:345f
 structural elements, 1:344t, 1:345f
 calmodulin-binding peptide (CalBP), 1:64–65
- cyclic adenosine diphosphate (ADP)-ribose (cADPr), 1:333–334
- EF-hand proteins, 1:318f, 1:319t, 1:321
- green fluorescent protein (GFP) indicators, 2:578
- guanine-nucleotide-exchange factors (GEFs), 4:246f
- myosin regulation, 3:221–222
- nitric oxide synthase (NOS)
 calcium ion (Ca²⁺)–calmodulin (CaM) regulation
 autoinhibitory loop, 1:344, 1:344t, 1:345f
 basic concepts, 1:343–346
 catalytic reactions, 3:253–254
 C-terminal extension, 1:344–346, 1:344t, 1:345f
 nitric oxide synthase (NOS)–calmodulin (CaM) binding site, 1:345f
 plasma membrane calcium (Ca⁺)-ATPase pump, 3:525–526
 reductase domain, 1:345f
 structural elements, 1:344t, 1:345f
 calmodulin (CaM)–nitric oxide synthase (NOS) interactions, 1:305f, 1:342–343
 electron transfer mechanisms, 1:342, 1:343f, 3:253–254
 plectin, 2:625–627
- ryanodine receptors (RyRs), 4:196
- transcription factors, 1:313f
- voltage-gated calcium channels, 4:560
- calnexin (CNX)
 ER-associated degradation (ERAD), 2:201–202
 glycoprotein folding processes, 2:448–449, 2:451–452, 2:451f, 4:205–206
 lectins, 2:702t, 2:704t
 N-linked glycans processing, 3:259t
 N-linked glycosylation, 2:460f, 2:463
- calorie restrictions, 1:177
- calpains, 1:353–361
 characteristics, 1:353, 1:354f
 EF-hand proteins, 1:319t
 focal adhesions, 2:320f, 2:321
 historical research, 1:353–354
 human calpain homologs, 1:353–354, 1:355t, 1:357f
 m-calpains, 1:353–354, 1:357f, 1:358f
 μ-calpains, 1:353–354, 1:357f
 nomenclature, 1:353–354
 research summary, 1:360–361
 sodium ion/calcium ion (Na⁺/Ca²⁺) exchanger, 1:424, 1:425f
 structural characteristics, 1:354f, 1:357f
 structure and functions, 1:354–359, 1:357f, 1:358f
 super-family characteristics
 calpain homologs
 alternative splicing products (nCL-2 and nCL-2'/CAPN8), 1:354f, 1:357f, 1:359–360
 CAPN10, 1:354f, 1:360
 CAPN1 through CAPN16, 1:354f, 1:355t, 1:357f
 CAPN6, 1:354f, 1:360
 CAPNS1/CAPNS2, 1:354f, 1:355t, 1:357f, 1:360
 Cpl1, 1:354f, 1:360
 defective kernel 1 (DEK1), 1:354f, 1:360
 nCL-4/CAPN9, 1:354f, 1:357f, 1:360
 PalB, 1:354f, 1:357f, 1:360
 skeletal muscle-specific calpain (p94/CANP3), 1:354f, 1:357f, 1:359
 SOLH/CAPN15, 1:354f, 1:355t, 1:357f, 1:360
 structural characteristics, 1:354f
 transform genotypic hermaphrodites (TRA-3), 1:354f, 1:360
 classifications, 1:354f, 1:359
 structure and functions, 1:354f, 1:359–360
- calreticulin (CRT)
 calcium ion (Ca²⁺) buffering endoplasmic reticulum proteins, 1:291t, 1:292, 1:295
 calcium sensors, 1:311
 ER-associated degradation (ERAD), 2:201–202
 glycoprotein folding processes, 2:448–449, 2:451–452, 2:451f, 4:205–206
 lectins, 2:704t
 N-linked glycans processing, 3:259t
 N-linked glycosylation, 2:460f, 2:463

- calretinin, 1:196–197, 1:197f, 1:319t
 calsenilin, 1:319t
 calsensin, 1:319t
 calsequestrin, 1:291t, 1:294
 calsymin, 1:319t
 caltractin, 1:319t
 calumenin, 1:291t, 1:295
 Calvin–Benson cycle, 3:497–502
 C4 pathways, 3:498
 carbon pathway, 3:497, 3:499f
 carboxylation phase, 3:497–498, 3:499f
 chloroplasts, 1:507, 1:511–512, 1:512f
 Crassulacean acid metabolism (CAM)
 pathway, 3:498
 enzymatic regulation
 flux control, 3:500
 irreversible enzymes, 3:500
 enzyme targets, 3:500–501
 feedback regulation, 3:501
 genetic modification effects, 3:500–501
 light-regulated energy and power, 3:497, 3:498f
 metabolic regulation, 3:499–500
 oxygenic photosynthesis, 2:335
 photosynthetic carbon dioxide (CO₂)
 fixation, 3:494, 3:494f, 3:495f
 reduction phase, 3:498, 3:499f
 regeneration phase, 3:499, 3:499f
 whole plants, 3:501
 Cameleons, 2:578
 Camgaros, 2:578
 cAMP (cyclic adenosine monophosphate)
 cAMP responsive element-binding protein
 3 (CREB3), 4:51f, 4:52
 carbohydrate-responsive element binding
 protein (ChREBP), 1:374, 1:375–376, 1:376f
 glycogen synthesis, 2:430f
 neurogenesis, 1:200f
 phosphofructokinase-2/fructose
 biphosphatase-2 (PFK-2/FBPase-2)
 activation, 3:457–458
 campesterol
 brassinosteroid biosynthesis, 1:251
 chemical structure, 1:251f
Campylobacter, 2:43
Campylobacter jejuni
 Cho phylogenetic tree, 3:331f
 reductive reactions, 1:487t
 respiratory chain complex I (NADH:
 ubiquinone oxidoreductase), 4:89–90
 succinate:quinone oxidoreductases
 (SQORs), 4:92–93
 uvrC phylogenetic tree, 3:331f
 canals of Herring, 1:190–191
Canavalia ensiformis, 2:702t
 cancer
 aspartic proteases, 1:139
 base excision repair (BER), 2:7
 cadherins, 1:259–260
 chromosome organization, 3:316
 cytosine methylation, 1:522
 defective NER, 3:340
 exosomes and microvesicles (EMVs), 2:264
 fatty acid metabolism–cancer relationship,
 2:275–280
 acyl-coenzyme A synthetase (ACS),
 2:279
 alpha-methylacyl-CoA racemase
 (AMACR), 2:279
 background information, 2:275
 fatty acid synthase expression
 cancer diagnostics, 2:277–278
 functional role, 2:276–279
 human cancer expression, 2:277,
 2:277f
 inhibition mechanisms, 2:278–279
 therapeutic targets, 2:278
 fatty acid synthesis pathway, 2:275–279,
 2:276f
 fibroblast growth factor receptors (FGFRs)
 characteristics, 2:306–308
 fibroblast growth factor receptor 1
 (FGFR1), 2:306–307
 fibroblast growth factor receptor 2
 (FGFR2), 2:307
 fibroblast growth factor receptor 3
 (FGFR3), 2:307
 fibroblast growth factor receptor 4
 (FGFR4), 2:307–308
 fibroblast growth factors (FGFs)
 expression effects, 2:304–306
 fibroblast growth factor 1 (FGF1), 2:305
 fibroblast growth factor 3 (FGF3), 2:305
 fibroblast growth factor 4 (FGF4), 2:305
 fibroblast growth factor 5 (FGF5), 2:306
 fibroblast growth factor 6 (FGF6), 2:306
 fibroblast growth factor 7 (FGF7), 2:306
 fibroblast growth factor 8 (FGF8), 2:306
 fibroblast growth factor 9 (FGF9), 2:306
 fibroblast growth factor 10 (FGF10),
 2:306
 fibroblast growth factor 19 (FGF19),
 2:306
 fibroblast growth factor 2 (FGF2), 2:305
 glycolysis, 2:446
 in vivo cell migration, 1:442–443, 1:443f
 interferon- γ receptors, 2:621
 liver cancer, 1:193–195
 mucin expression, 3:203–204
 neurokinin-1 receptor (NK1R), 4:365
 nuclear factor kappa B (NF- κ B),
 3:304–305, 3:304t
 O-GlcNAc biosynthesis, 3:352
 phospholipase D (PLD), 3:476–477
 rapamycin (mTOR), 3:198–199
 reactive oxygen species (ROS), 2:79
 replication fork (RF) breakage, 4:27
 ribosome biogenesis, 4:121
 spectrophotometric assays, 4:279
 steroid/thyroid hormone receptors,
 4:311–312
 vitamin E, 4:547
Candida albicans
 aquaporin water channels, 1:125
 aspartic proteases, 1:139–140
 calpains, 1:360
 centromere region, 1:450
 copper-transporting ATPases (copper
 pumps), 3:671–672
 N-myristoyltransferase (Nmt), 3:641,
 3:643–644
 pheromone receptors, 3:441
 sugar nucleotide transporters, 4:350
Candidatus Chloracidobacterium
 thermophilum, 2:501–503, 2:504–507,
 2:504f, 2:505f, 2:508f
 cannabinoids, 2:423
 CapG, 1:23, 1:277–278
 caphalotocin, 4:516t
 capsaicin, 1:423–424
 capsulin, 1:185
 CapZ, 1:31–32, 1:32f, 1:21f, 1:22f, 1:25,
 2:174, 2:175f
 carbamoyl phosphate
 carbamoylphosphate synthase (CPS),
 1:94–95, 4:494, 4:495f, 4:495t,
 4:496f, 4:497–498
 carbamoylphosphate synthetase II
 (CPSase II), 3:712, 3:713f, 3:714f
 insect biochemistry, 2:591–592
 synthetic pathways, 3:712, 3:713f,
 3:714f
 carbohydrates
 carbohydrate chain synthesis, 1:362–367
 automated synthesis
 general discussion, 1:5–6
 programmable one-pot synthesis,
 1:5–6, 1:5f
 solid-phase synthesis, 1:4–5, 1:5f
 challenges, 1:1
 chemical synthesis
 automated synthesis, 1:5–6
 basic concepts, 1:4–6, 1:4f
 bond-formation control, 1:3–4, 1:4f
 solution-phase synthesis, 1:4, 1:5f
 stereochemical control, 1:3–4, 1:4f
 sugar hydroxyl protection, 1:4
 enzymatic synthesis
 basic concepts, 1:2–3
 glycosyltransferase-catalyzed
 reactions, 1:2f, 1:3
 glycosidase-catalyzed reactions, 1:2,
 1:2f
 glycosyltransferase-catalyzed reactions
 basic concepts, 1:3
 nucleotide-sugar recycling, 1:3
 phosphorylation effects, 1:3
 product inhibition, 1:3
 recombinant expression, 1:3
 schematic diagram, 1:2f
 molecular structure, 1:2f
 carbohydrate recognition domain (CRD),
 2:701–703
 carbohydrate response element (CHORE),
 2:287, 3:721
 carbohydrate response factor (CHORF),
 2:287
 carbohydrate-responsive element binding
 protein (ChREBP), 1:374–378
 biological significance, 1:376–377,
 1:377f
 ChREBP^{-/-} mice
 genetic ablation studies, 1:376–377,
 1:377f
 leptin-deficient ob/ob mice intercross
 studies, 1:377–378
 functional role, 1:374–375, 1:375f

carbohydrates (*Continued*)

genetic ablation studies, 1:376–377, 1:377f
 historical research, 1:374
 identification and purification processes, 1:374
 liver pyruvate kinase (LPK) catalysis, 1:374, 1:375f, 1:376–377, 1:376f, 1:377f
 regulation mechanisms, 1:375–376, 1:376f
 research summary, 1:378
 species and tissue distribution, 1:374
 structural characteristics, 1:374–375
 transcriptional activation, 1:374
 cardiac metabolism, 3:65–66, 3:66f, 3:68f
 digestion and absorption, 2:359
 exosomes and microvesicles (EMVs), 2:262
 HUFA-response region with carbohydrate response element (CHORE), 2:288
 lipogenic gene expression, 2:286–287
 liver PK (LPK) *in vitro* regulation, 3:721
 carbon (C)
 carbon-13 (¹³C) nuclear magnetic spectroscopy, 1:107, 1:108f
 carbon dioxide (CO₂)
 fermentation reactions, 2:205, 2:205t
 photosynthetic carbon dioxide (CO₂) fixation, 3:497–502
 aquatic photosynthesis, 3:495
 Benson–Calvin cycle, 3:494, 3:494f, 3:495f
 C4 pathways, 3:495, 3:495f, 3:498
 Calvin cycle enzyme targets, 3:500–501
 carbon pathway, 3:497, 3:499f
 carboxylation phase, 3:497–498, 3:499f
 Crassulacean acid metabolism (CAM) pathway, 3:495–496, 3:498
 enzymatic regulation, 3:500
 feedback regulation, 3:501
 general discussion, 3:494–495
 genetic modification effects, 3:500–501
 light-regulated energy and power, 3:497, 3:498f
 metabolic regulation, 3:499–500
 reduction phase, 3:498, 3:499f
 regeneration phase, 3:499, 3:499f
 whole plants, 3:501
 carbon monoxide (CO), 2:534
 carbonyl cyanide *p*-(trifluoromethoxy) phenylhydrazone (FCCP), 1:143
Carboxydotherrmus hydrogenoformans, 1:490t, 3:331f
 carboxyethyllysine (CEL), 2:407
 carboxylases
 acetyl-coenzyme A carboxylase (ACC)
 body weight regulation, 4:57f, 4:58f
 cellular fatty acid metabolism regulation, 4:56
 fatty acid biosynthesis, 2:286
 insulin metabolism, 2:606–607
 ischemic heart diseases, 3:68, 3:69f, 3:70f

lipogenic gene expression, 2:287f
 biotinylation, 1:226
 fatty acid biosynthesis, 2:286
 pyruvate carboxylase (PC)
 characteristics, 2:381–382
 gluconeogenesis, 2:385f
 historical research, 2:381
 protein biotinylation, 1:226
 carboxylic acid
 acyl-S–coenzyme A (CoA) derivatives, 1:539–540, 1:540f
 fermentation reactions, 2:204–206
 carboxyl methylation, 3:607–610
 bacterial chemotaxis, 3:607–608, 3:608f
 basic concepts, 3:607
 chemotaxis receptor fibers, 3:608f
 eukaryotic signal transduction proteins, 3:608
 heterotrimeric G proteins, 3:608–609, 3:608f
 phosphoprotein phosphatase 2A (PP2A), 3:609–610, 3:609f
 Ras-related proteins, 3:608–609
 carboxyl tail, 4:363–364, 4:363f
 carboxyl terminal domain (CTD), 4:165
 carboxyl-terminal transactivation domain (TAD-C), 4:67, 4:68f, 4:70
 carboxymethyllysine (CML), 2:406f, 2:407, 2:408f, 2:409f
 carboxypeptidase A (CPD A), 3:88
 carboxypeptidase Y (CPD Y), 4:504–506, 4:504f, 4:505f
 carcinogenesis
 aspartic proteases, 1:139
 base excision repair (BER), 2:7
 cadherins, 1:259–260
 chromosome organization, 3:316
 cytosine methylation, 1:522
 defective NER, 3:340
 exosomes and microvesicles (EMVs), 2:264
 fatty acid metabolism–cancer relationship, 2:275–280
 acyl-coenzyme A synthetase (ACS), 2:279
 alpha-methylacyl-CoA racemase (AMACR), 2:279
 background information, 2:275
 fatty acid synthase expression
 cancer diagnostics, 2:277–278
 functional role, 2:276–279
 human cancer expression, 2:277, 2:277f
 inhibition mechanisms, 2:278–279
 therapeutic targets, 2:278
 fatty acid synthesis pathway, 2:275–279, 2:276f
 fibroblast growth factor receptors (FGFRs)
 characteristics, 2:306–308
 fibroblast growth factor receptor 1 (FGFR1), 2:306–307
 fibroblast growth factor receptor 2 (FGFR2), 2:307
 fibroblast growth factor receptor 3 (FGFR3), 2:307
 fibroblast growth factor receptor 4 (FGFR4), 2:307–308

fibroblast growth factors (FGFs)
 expression effects, 2:304–306
 fibroblast growth factor 1 (FGF1), 2:305
 fibroblast growth factor 3 (FGF3), 2:305
 fibroblast growth factor 4 (FGF4), 2:305
 fibroblast growth factor 5 (FGF5), 2:306
 fibroblast growth factor 6 (FGF6), 2:306
 fibroblast growth factor 7 (FGF7), 2:306
 fibroblast growth factor 8 (FGF8), 2:306
 fibroblast growth factor 9 (FGF9), 2:306
 fibroblast growth factor 10 (FGF10), 2:306
 fibroblast growth factor 19 (FGF19), 2:306
 fibroblast growth factor 2 (FGF2), 2:305
 glycolysis, 2:446
in vivo cell migration, 1:442–443, 1:443f
 interferon-γ receptors, 2:621
 liver cancer, 1:193–195
 mucin expression, 3:203–204
 neurokinin-1 receptor (NK1R), 4:365
 nuclear factor kappa B (NF-κB), 3:304–305, 3:304t
 O-GlcNAc biosynthesis, 3:352
 phospholipase D (PLD), 3:476–477
 rapamycin (mTOR), 3:198–199
 reactive oxygen species (ROS), 2:79
 replication fork (RF) breakage, 4:27
 ribosome biogenesis, 4:121
 spectrophotometric assays, 4:279
 steroid/thyroid hormone receptors, 4:311–312
 vitamin E, 4:547
 carcinoma-associated fibroblasts (CAFs), 1:442–443, 1:443f
 CARD11 (caspase-recruitment domain containing protein 11), 1:391–392, 4:377–378, 4:377f
 cardiac excitation–contraction (E–C)
 coupling, 1:379–383
 arrhythmia incidence, 1:382–383
 basic concepts
 action mechanisms, 1:379–381
 action potential (AP), 1:379, 1:380f, 4:566
 cytosolic calcium, 1:379–381, 1:380f
 L-type calcium channels, 1:379–381, 1:380f
 myocytes, 1:379, 1:380f
 myofilaments, 1:380–381, 1:380f
 relaxation mechanisms, 1:381
 delayed after-depolarization (DAD), 1:382–383, 1:382f
 early after-depolarization (EAD), 1:382–383
 heart failure, 1:381–382, 1:382f
 transverse tubule (T-tubule), 1:379–381, 1:380f, 1:382f
 cardiac glycosides, 3:686–687, 3:686f, 3:687f
 cardiac muscles, 1:179–186
 development process
 cardiogenesis, 1:183, 1:183f, 1:184f
 microRNA (miRNA), 1:185
 transcription factors, 1:183–185, 1:184f
 research summary, 1:186
 stem cells, 1:186

- structural characteristics, 1:179, 1:180f
cardiac myosin light chain kinase (cMLCK), 1:305–306
cardiac pacemakers, 1:587–588
cardiolipin, 3:113, 3:148, 4:21
cardiomyopathy, 3:223t
Carney complex, 3:223t
carnitine, 1:384–387
 ascorbic acid, 4:531–532
 background information, 1:384
 biosynthesis and homeostasis, 1:384, 1:385
 branched-chain amino acid (BCAA) regulation, 1:245–246
 carnitine acetyltransferase, 1:387
 carnitine:acylcarnitine translocase (CACT), 1:385–386, 1:385f
 carnitine palmitoyltransferase I (CPT I) cardiac metabolism, 3:66–67, 3:69f, 3:70f
 fatty acid oxidation, 1:110, 2:276, 2:276f, 2:281, 2:282f
 lipogenic gene expression, 2:287f, 4:56
 malonyl-CoA
 body weight regulation, 4:57f
 carnitine-dependent cellular uptake and activation, 1:385f
 cellular fatty acid metabolism regulation, 4:56
 functional roles, 2:258–259, 4:56, 4:58
 malonyl-CoA-sensitive carnitine palmitoyltransferase, 1:385–386, 1:385f
 carnitine palmitoyltransferase II (CPT II), 1:385–386, 1:385f, 2:281, 2:282f
functional role
 acyl-coenzyme A transesterification, 1:387, 1:385
 mitochondrial acyl-CoA/CoA ratio regulation, 1:387
 mitochondrial β -oxidation
 cellular uptake and activation, 1:384–385, 1:385f
 mitochondrial carnitine system, 1:385–386
 reaction pathways, 1:386, 1:386–387
 molecular structure, 1:385
carotenoids, 1:498–505
 antioxidant actions, 2:77
 applications, 1:504
 carotenoid monooxygenase 1 (CMO1), 4:525–526, 4:527f
 chemical properties, 1:504
 chlorosomes, 2:514
 cytochrome b_6f complex, 4:337f, 4:338
 functional role, 1:503–504
 green bacteria, 2:504f
 light harvesting processes, 1:498, 2:178–179, 2:726–727, 3:486, 3:492
 metabolism, 1:504
 occurrences, 1:503–504
 orange carotenoid protein (OCP), 3:486
 photosystem I (PSI) complex, 3:503, 3:504f
 phototrophic bacteria, 3:32
 reactive oxygen species (ROS), 1:503–504
 spectral properties, 1:502–503, 1:503f
 structural characteristics, 1:501–502, 1:502f
carpal tunnel syndrome, 1:101t
Carperitide, 3:227
carrots, 4:533t
carvedilol, 1:59
casein, 3:91t
casein kinase (CK2 α), 1:514, 1:514f
caspases
 activation mechanisms, 1:589–590
 apoptosis, 1:388–396
 action mechanisms, 1:391
 action sequences, 1:389f, 1:391–392, 1:393f
 activation mechanisms, 1:389, 1:389f, 1:393f
 apoptotic pathways, 1:389f
 background information, 1:388
 Baculovirus IAP repeat (BIR) proteins, 1:393
 caspase 1
 caspase knockouts, 1:392t
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 therapeutic outlook, 1:395f
 caspase 2
 action sequences, 1:391–392
 caspase knockouts, 1:392t
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 sequential activation, 1:393f
 caspase 3
 action sequences, 1:391–392
 caspase knockouts, 1:392t
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 regulatory mechanisms, 1:390f
 sequential activation, 1:393f
 therapeutic outlook, 1:395f
 caspase 4
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 caspase 5
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 caspase 6
 action sequences, 1:391–392
 caspase knockouts, 1:392t
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 regulatory mechanisms, 1:390f
 sequential activation, 1:393f
 therapeutic outlook, 1:395f
caspase 7
 action sequences, 1:391–392
 caspase knockouts, 1:392t
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 regulatory mechanisms, 1:390f
 sequential activation, 1:393f
caspase 8
 activation mechanisms, 1:394
 caspase knockouts, 1:392t, 1:393
 characteristics, 1:388–389
 downstream activators, 1:393, 1:394f, 4:377–378
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 regulatory mechanisms, 1:390f
 sequential activation, 1:393f
caspase 9
 action sequences, 1:391–392
 caspase knockouts, 1:392t
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 regulatory mechanisms, 1:390f
 sequential activation, 1:393f
 therapeutic outlook, 1:395f
caspase 10
 action sequences, 1:391–392
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 sequential activation, 1:393f
caspase 11, 1:392t
caspase 12
 classifications, 1:388–389
 phylogenetic relationships, 1:390f
caspase knockouts, 1:392t, 1:393, 1:394f
cysteine proteases, 1:591–592
effector caspases, 1:389f, 1:391–392
functional role, 1:591–592
historical research, 1:388–389
inhibitors of apoptosis proteins (IAPs), 1:390f, 1:391t, 1:392–393, 1:392t, 1:590
pathological implications, 1:394
regulatory mechanisms, 1:389f, 1:390f
signaling mechanisms, 4:457–458, 4:458f
substrate generation, 1:391t, 1:392
substrate recognition, 1:389–390, 1:391t, 1:393f
therapeutic outlook, 1:394–395, 1:395f
caspase 1
 caspase knockouts, 1:392t
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 therapeutic outlook, 1:395f

caspases (*Continued*)

caspase 2

- action sequences, 1:391–392
- caspase knockouts, 1:392*t*
- classifications, 1:388–389
- inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
- phylogenetic relationships, 1:390*f*
- sequential activation, 1:393*f*

caspase 3

- action sequences, 1:391–392
- caspase knockouts, 1:392*t*
- classifications, 1:388–389
- inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
- phylogenetic relationships, 1:390*f*
- regulatory mechanisms, 1:390*f*
- sequential activation, 1:393*f*
- therapeutic outlook, 1:395*f*

caspase 4

- classifications, 1:388–389
- inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
- phylogenetic relationships, 1:390*f*

caspase 5

- classifications, 1:388–389
- inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
- phylogenetic relationships, 1:390*f*

caspase 6

- action sequences, 1:391–392
- caspase knockouts, 1:392*t*
- classifications, 1:388–389
- inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
- phylogenetic relationships, 1:390*f*
- regulatory mechanisms, 1:390*f*
- sequential activation, 1:393*f*
- therapeutic outlook, 1:395*f*

caspase 7

- action sequences, 1:391–392
- caspase knockouts, 1:392*t*
- classifications, 1:388–389
- inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
- phylogenetic relationships, 1:390*f*
- regulatory mechanisms, 1:390*f*
- sequential activation, 1:393*f*

caspase 8

- activation mechanisms, 1:394
- caspase knockouts, 1:392*t*, 1:393
- characteristics, 1:388–389
- downstream activators, 1:393, 1:394*f*, 4:377–378
- inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
- phylogenetic relationships, 1:390*f*
- regulatory mechanisms, 1:390*f*
- sequential activation, 1:393*f*

caspase 9

- action sequences, 1:391–392
- caspase knockouts, 1:392*t*
- classifications, 1:388–389
- inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
- phylogenetic relationships, 1:390*f*

- regulatory mechanisms, 1:390*f*
- sequential activation, 1:393*f*
- therapeutic outlook, 1:395*f*

caspase 10

- action sequences, 1:391–392
- classifications, 1:388–389
- inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392*t*
- phylogenetic relationships, 1:390*f*
- sequential activation, 1:393*f*

caspase 11, 1:392*t*

caspase 12

- classifications, 1:388–389
- phylogenetic relationships, 1:390*f*
- caspase knockouts, 1:392*t*, 1:393, 1:394*f*
- classifications, 1:388–389, 1:390*f*
- domain structures, 1:389, 1:393*f*
- downstream activators, 1:393, 1:394*f*, 4:377–378
- historical research, 1:388–389
- inhibitors of apoptosis proteins (IAPs), 1:391*t*, 1:392–393, 1:392*t*, 1:590
- molecular structure, 1:590*f*
- neurotrophins, 3:241–243, 3:242*f*
- phylogenetic relationships, 1:390*f*, 1:589
- signaling mechanisms, 1:394*f*, 1:618*f*, 4:457–458, 4:458*f*
- substrate recognition, 1:389–390, 1:391*t*, 1:393*f*

castanospermine (CSN), 2:459*f*castor beans, 2:702*t*catabolic hormones, 1:95–96, 1:96*f*, 1:150*t*

catabolite receptor protein (CRP)–cAMP complex, 2:387–388

catabolite repressor protein (CRP)

- functional role, 2:698–699
- historical research, 2:695
- schematic diagram, 2:695*f*
- structural characteristics, 2:697*f*, 2:699

catalases

- action mechanisms, 2:536, 2:536*f*
- functional role, 2:536, 3:407
- heme peroxidases, 3:407–409, 3:408*f*
- oxygenases, 2:536
- structural characteristics, 2:536

cataplerotic pathways, 4:305

cataracts, 2:632*t*

catecholamines

- amino acid metabolism, 1:93*t*
- catecholamine-derived *o*-benzoquinones, 3:727, 3:727*f*

epinephrine

- branched-chain amino acid (BCAA) regulation, 1:245–246
- characteristics and functional role, 1:57
- starvation effects, 4:303–304

norepinephrine

- ascorbic acid, 4:531–532
- characteristics and functional role, 1:57
- pancreatic endocrine cells, 2:596
- starvation effects, 4:302–303

catechol-*O*-methyltransferase (COMT), 1:57catenanes, 2:162, 2:162*f*

catenins

- α -catenins
- adaptor proteins, 1:258*f*

- cadherin signaling, 1:261, 1:262*f*
- characteristics, 1:255
- gene deletion effects, 1:259*t*

 β -catenins

- adaptor proteins, 1:258*f*
- adipocyte development, 1:53–54, 1:53*f*
- cadherin signaling, 1:261, 1:262*f*
- characteristics, 1:255
- G₁₂/G₁₃ family functions, 2:345*f*
- gene deletion effects, 1:259*t*
- mucins, 3:206
- neurogenesis, 1:199, 1:200*f*
- phosphorylation/dephosphorylation, 1:264, 1:264*f*
- Wnt/ β -catenin signaling pathway, 1:263–264, 1:263*f*, 2:434–435, 2:436*f*, 3:385–386, 4:386

 γ -catenins, 1:255, 1:258*f*, 1:261, 1:264hakai, 1:255, 1:258*f*p120 catenin, 1:255, 1:258*f*, 1:261, 1:262*f*, 1:264CATH database, 3:615*t*

cathepsins

- cathepsin D, 1:137, 1:139
- molecular structure, 1:590*f*
- phylogenetic relationships, 1:589
- physiological functions
- antigen-li processing, 1:591
- apoptosis, 1:591–592
- bone turnover, 1:591
- lysosomal proteolysis, 1:590–591
- thyroglobulin processing, 1:591

cation-gradient-driven symporters

- classification, 2:388
- functional role, 2:388
- lactose permease (LacY), 2:388–389
- melibiose permease (MelB), 2:389
- sodium/galactose transport (SglT), 2:389
- structural characteristics, 2:388, 2:389*f*

cattle, 3:575*t*cauliflower, 2:638*t*, 4:533*t**Caulobacter crescentus*, 3:331*f*

caveolae, 2:743–744

caveolar endocytosis, 2:198, 2:198*f*

caveolins

- caveolin-1 (Cav-1), 2:278–279
- G protein-coupled receptor kinases (GRKs), 2:485*t*
- nitric oxide synthase (NOS), 1:343–346
- phosphodiesterases (PDEs), 1:581–582, 1:584

Cbl/Cbl-associated protein (CAP) pathways

- insulin receptors (IRs), 2:602–603
- insulin signal transduction, 2:605
- ubiquitin-like proteins (UBLs), 4:468*t*, 4:470

CBM-301106, 3:71*t*CBS Prediction Servers, 3:615*t*

CCA trinucleotide

- transfer RNA (tRNA)
- 5'-end–RNase P process pathway, 3:555, 3:555*f*
- maturation processes, 3:554
- 3'-end–RNase P process pathway
- Bacillus subtilis*, 3:555*f*, 3:556
- Escherichia coli*, 3:555–556, 3:555*f*
- Thermotoga maritima*, 3:555*f*, 3:556

- CD4+/CD8+ cells
 anti-mitochondrial antibodies (AMAs), 3:115
 chemokine receptors, 1:484
 hematopoiesis, 1:187–188
 interferon- γ receptors, 2:621
 primary biliary cirrhosis (PBC), 3:562
 T-cell receptor (TCR) signaling, 4:374, 4:374f, 4:377f
 tumor necrosis factor receptors (TNFRs), 4:459
- CD22 protein, 4:229, 4:230, 4:230f, 4:231
 CD28 protein, 4:376
 CD33 protein, 4:229, 4:230, 4:230f, 4:231
 CD44 protein, 2:262, 2:263f, 3:96
 CD45 gene, 1:79–80, 1:79f, 3:649–651, 3:650f, 3:651t, 4:376
 CD55 protein, 2:263f
 CD95 receptor, 1:390f, 1:392t, 1:394, 4:287
 CD151 protein, 1:642
 Cdc2 protein, 1:402t, 1:624–625
 Cdc6 protein, 1:401, 1:402t, 1:403f, 2:22, 2:108, 2:111, 2:112f
 Cdc7 kinase, 1:401–402, 1:402t, 1:403, 1:403f, 1:404f
 Cdc17 protein, 1:403f
 Cdc18 protein, 1:404f, 2:22
 Cdc20 protein, 1:403f
 Cdc25 protein, 3:648, 3:650f, 3:651f, 4:246f
 Cdc28 protein, 1:403f, 4:374, 4:374f
 Cdc42 protein
 actin cytoskeleton, 1:430–431, 1:431f, 1:626, 2:321–322, 2:321f
 cadherin signaling, 1:262–263, 1:263f
 cell migration, 1:439–440
 chemokine receptors, 1:480–482
 cytokinesis, 1:626
 guanosine triphosphatases (GTPases), 1:440–441, 1:441f, 4:112f, 4:113, 4:113t, 4:114f
 pheromone receptors, 3:441–442, 3:443f
 Rho GTP-binding proteins, 1:440–441, 1:441f
 small guanosine phosphatases (GTPases), 1:453, 4:243f
- Cdc45 protein, 1:401–402, 1:403f, 1:404f
 Cdc48p/p97 protein, 1:3t, 1:4f, 1:5
 Cdc48 protein, 3:599
 Cdk1 protein, 1:624–625
 Cdt1 protein, 1:403f, 2:108, 2:111–112, 2:112f
- C/EBP (Ccaat/enhancer binder protein)
 adipocyte development
 C/EBP homologous protein (CHOP-10), 1:55–56
 lineage commitment, 1:53f
 terminal differentiation, 1:54–55
 transcriptional control, 1:55f
- C/EBP α
 adipocyte development
 microRNA (miRNA)-based regulation, 4:192–193
 terminal differentiation, 1:54–55, 1:54f
 transcriptional control, 1:56
 Wnt pathway, 1:53–54, 1:53f
- granulocyte–monocyte progenitor (GMP), 1:188–189
 hematopoietic stem cells (HSCs), 1:187, 1:188f
- C/EBP β
 adipocyte development
 C/EBP homologous protein (CHOP-10), 1:55–56
 terminal differentiation, 1:54–55, 1:54f
 transcriptional control, 1:56
 transcriptional expression, 1:55
 Wnt pathway, 1:53–54, 1:53f
- granulocyte–monocyte progenitor (GMP), 1:188–189
 C/EBP ϵ , 1:188–189
 glycogen synthase kinase-3 (GSK3), 2:435f, 2:436t
- cell adhesion
 actin cytoskeleton, 1:430–431, 1:431f
 cadherins, 1:255–260
 cytoplasmic domains
 adaptor proteins, 1:257, 1:258f
 adhesion complex turnover, 1:257–258
 cytoskeletal linkage, 1:257
 signaling functions, 1:258
 structural characteristics, 1:256f
- extracellular domains
 sorting mechanisms, 1:257
 structural characteristics, 1:256–257, 1:256f
 functional role, 1:255–256
- subfamilies
 atypical cadherins, 1:256, 1:256f
 classic cadherins, 1:255, 1:256f, 1:261, 1:262f
 desmosomal cadherins, 1:255–256, 1:256f, 1:261
 protocadherins, 1:256, 1:256f, 1:261
- desmosomes, 1:638t, 1:639
 glycosphingolipids (GSLs), 2:439–442
 clustered GSLs, 2:439, 2:440f, 2:441f
 cluster formation, 2:440–441
- L1CAM cell adhesion molecule, 4:272
 polysialylated embryonic form of neural cell adhesion molecule (PSA–NCAM), 1:196–197, 1:197f
- cell cycle, 1:406–409
 basic concepts, 1:406
 cytokinins, 1:629
 DNA damage checkpoints, 1:410–416
 activation mechanisms, 1:410–412, 1:412f
 apoptosis, 1:415
 background information, 1:410
 component proteins, 1:411t
 disease connections, 1:415–416
 G1/S checkpoint, 1:411, 1:412f, 1:415–416
 G2/M checkpoint, 1:412, 1:412f, 1:415–416
 Intra-S checkpoint, 1:411, 1:412f, 1:415–416
 lesion sensors, 1:412, 1:413f
 phases, 1:411f
 phosphoinositide 3-kinase-like kinase (PIKK) family, 1:412–413, 1:413f
 phospho-peptide-binding domains, 1:412f, 1:414
 research summary, 1:416
 senescence, 1:415
 signal transduction, 1:412–415
 spatiotemporal organization, 1:414–415
- E2F transcription factor, 1:406–407, 1:407f, 1:412f
- senescence, 1:415
 signal transduction
 ataxia
 telangiectasia-mutated and Rad3-related (ATR) pathway, 1:411t, 1:413f, 1:414
 ataxia telangiectasia mutated (ATM) kinase, 1:411, 1:411t, 1:412f, 1:413–414, 1:413f, 1:415–416
 BRCA1 C-terminal (BRCT) domains, 1:411t, 1:412f, 1:414
 Chk1 and Chk2 effector kinases, 1:411t, 1:412f, 1:414
 chromatin modifications, 1:413f, 1:414
 forkhead-associated (FHA) binding domains, 1:412f, 1:414
 lesion sensors, 1:412, 1:413f
 phosphoinositide 3-kinase-like kinase (PIKK) family, 1:412–413, 1:413f
 phospho-peptide-binding domains, 1:412f, 1:414
 spatiotemporal organization, 1:414–415
- G₀ phase and G₁ phase
 basic concepts, 1:406
 cell growth regulation, 1:408
 DNA damage checkpoints, 1:410–416
 activation mechanisms, 1:410–412, 1:412f
 apoptosis, 1:415
 ataxia telangiectasia-mutated and Rad3-related (ATR) pathway, 1:411t, 1:413f, 1:414
 ataxia telangiectasia mutated (ATM) kinase, 1:411, 1:411t, 1:412f, 1:413–414, 1:413f, 1:415–416
 background information, 1:410
 BRCA1 C-terminal (BRCT) domains, 1:411t, 1:412f, 1:414
 Chk1 and Chk2 effector kinases, 1:411t, 1:412f, 1:414
 chromatin modifications, 1:413f, 1:414
 component proteins, 1:411t
 disease connections, 1:415–416
 forkhead-associated (FHA) binding domains, 1:412f, 1:414
 G1/S checkpoint, 1:411, 1:412f, 1:415–416
 G2/M checkpoint, 1:412, 1:412f, 1:415–416
 Intra-S checkpoint, 1:411, 1:412f, 1:415–416
 lesion sensors, 1:412, 1:413f
 phases, 1:411f
 phosphoinositide 3-kinase-like kinase (PIKK) family, 1:412–413, 1:413f
 phospho-peptide-binding domains, 1:412f, 1:414
 research summary, 1:416
 senescence, 1:415
 signal transduction, 1:412–415
 spatiotemporal organization, 1:414–415

- cell cycle (*Continued*)
 future research outlook, 1:408
 regulatory mechanisms, 1:406–407, 1:407f
 retinoblastoma protein (pRB), 1:406–407, 1:407f
 stem cell proliferation, 1:408
 rapamycin (mTOR), 3:198
- cell death, 1:423–425
- Bcl2 family, 1:165–166
 Bcl2 homology (BH) domains, 1:165, 1:166f
 BH3-only proteins, 1:165, 1:166f
Caenorhabditis elegans programmed cell death, 1:165
 family classifications, 1:165, 1:166f
 inhibitors, 1:165, 1:389f, 1:390f
 therapeutic uses, 1:165–166
- calcium ion (Ca²⁺)
 brain ischemia, 1:423–424
 excitotoxicity, 1:423–424, 1:425f
 neurodegenerative diseases, 1:424–425
 neuronal death, 1:424
 signal executors, 1:423
 signaling mechanisms, 1:423
 signaling subroutines, 1:425
- caspases, 1:388–396
 action mechanisms, 1:391
 action sequences, 1:389f, 1:391–392, 1:393f
 activation mechanisms, 1:389, 1:389f, 1:393f
 apoptotic pathways, 1:389f
 background information, 1:388
 Baculovirus IAP repeat (BIR) proteins, 1:393
- caspase 1
 caspase knockouts, 1:392t
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 therapeutic outlook, 1:395f
- caspase 2
 action sequences, 1:391–392
 caspase knockouts, 1:392t
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 sequential activation, 1:393f
- caspase 3
 action sequences, 1:391–392
 caspase knockouts, 1:392t
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 regulatory mechanisms, 1:390f
 sequential activation, 1:393f
 therapeutic outlook, 1:395f
- caspase 4
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
- caspase 5
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
- caspase 6
 action sequences, 1:391–392
 caspase knockouts, 1:392t
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 regulatory mechanisms, 1:390f
 sequential activation, 1:393f
 therapeutic outlook, 1:395f
- caspase 7
 action sequences, 1:391–392
 caspase knockouts, 1:392t
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 regulatory mechanisms, 1:390f
 sequential activation, 1:393f
- caspase 8
 activation mechanisms, 1:394
 caspase knockouts, 1:392t, 1:393
 characteristics, 1:388–389
 downstream activators, 1:393, 1:394f, 4:377–378
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 regulatory mechanisms, 1:390f
 sequential activation, 1:393f
- caspase 9
 action sequences, 1:391–392
 caspase knockouts, 1:392t
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 regulatory mechanisms, 1:390f
 sequential activation, 1:393f
 therapeutic outlook, 1:395f
- caspase 10
 action sequences, 1:391–392
 classifications, 1:388–389
 inhibitors of apoptosis proteins (IAPs), 1:391t, 1:392t
 phylogenetic relationships, 1:390f
 sequential activation, 1:393f
- caspase 11, 1:392t
- caspase 12
 classifications, 1:388–389
 phylogenetic relationships, 1:390f
- caspase knockouts, 1:392t, 1:393, 1:394f
- effector caspases, 1:389f, 1:391–392
- functional role, 1:591–592
- historical research, 1:388–389
- inhibitors of apoptosis proteins (IAPs), 1:390f, 1:391t, 1:392–393, 1:392t, 1:590
- pathological implications, 1:394
- regulatory mechanisms, 1:389f, 1:390f
- signaling mechanisms, 4:457–458, 4:458f
- substrate generation, 1:391t, 1:392
- substrate recognition, 1:389–390, 1:391t, 1:393f
- therapeutic outlook, 1:394–395, 1:395f
- characteristics, 1:423
- mitochondrial dynamics
 cristae remodeling, 3:138
 mitochondrial channels, 3:130–131
 mitochondrial fragmentation, 3:138
- cell–matrix interactions, 1:426–429
- extracellular matrix (ECM)
 adhesive motifs, 1:426, 1:427t
 basic concepts, 1:426
 components, 1:426
 signaling pathways, 1:429, 2:608
- integrins
 activation and signaling mechanisms
 conformational changes, 1:428, 2:609–610
 cytoplasmic domains, 1:428–429, 2:609
 focal adhesions, 1:428
 signaling pathways, 1:429, 2:608–612
- cation modulation, 1:428
- cell migration, 1:439f
- ligand binding, 1:427–428
- structural characteristics, 1:427, 1:427t
- cell migration, 1:430–435
- actin filament recycling, 1:431f, 1:434, 1:434f
- basic concepts, 1:430
- cell guidance, 1:434
- extracellular matrix (ECM), 1:436–445
- cell–cell adhesion
 cadherins, 1:444–445
 collective and individual cell migration, 1:444–445, 1:445f
- in vivo* cell migration
 carcinogenesis, 1:442–443, 1:443f
 embryo development, 1:441–442
 immune trafficking, 1:442
 pathological conditions, 1:442, 1:443f
 wound healing, 1:442
- proteases
 degradation processes, 1:443–444
 ECM formation, 1:444
 endostatin, 1:444
 growth factor activation, 1:444
 proteolytic fragments, 1:444
 regulatory mechanisms, 1:444–445
- receptors
 integrins, 1:438–440, 1:439f
 proteoglycans, 1:439–440
 receptor clustering, 1:440, 1:440f
 signaling pathways, 1:439f
- regulatory mechanisms
 focal adhesion kinase (FAK) family, 1:441, 1:441f
 integrins, 1:444–445
 proteases, 1:444–445
 Rho GTP-binding proteins, 1:440–441, 1:441f
- research summary, 1:445
- stages, 1:437f

- three-dimensional (3D) cell migration and invasion, 1:443–444
- two-dimensional (2D) surfaces, 1:436
- in vivo* migration, 1:435
- microtubule formation, 1:434
- movement phases
- adhesion
 - formation mechanisms, 1:431*f*, 1:432–433, 1:432*f*, 1:433*f*
 - polarization, 1:431*f*, 1:433–434, 1:433*f*
 - protrusion, 1:432, 1:432*f*, 1:433*f*
 - cellular fusion proteins, 3:37, 3:38*f*
 - cellular respiration, 3:407, 3:408*f*
 - cellular retinoid-binding proteins (CRBPs)
 - cellular RA-binding protein II (CRABPII), 4:526–528, 4:527*f*, 4:527*t*
 - cellular retinal binding protein (CRALBP), 4:526–528, 4:527*f*, 4:527*t*
 - cellular retinol binding protein II (CRBPII), 4:526–528, 4:527*f*, 4:527*t*
 - cellulose
 - biomass-produced renewable hydrogen, 4:75*f*
 - cellulose-binding domains (CelBDs), 1:62
 - Cenarchaeum symbiosum*, 2:158*t*, 2:161
 - Center for Biological Sequence Analysis, 3:612
 - Center for Information Biology, 3:611–612
 - central core disease (CCD), 4:197–198
 - central nervous system (CNS)
 - chemokine receptors, 1:484–485
 - energy homeostasis, 3:26
 - fast synaptic transmission, 2:395
 - GABA_A receptors, 2:351
 - glutamate receptors, 2:712–721
 - AMPA receptors, 1:425*f*, 2:681–684, 2:714–716, 2:715*t*, 2:716*t*, 2:717–718, 2:717*f*, 2:718*f*
 - calcium ion (Ca²⁺) signaling system, 2:719–720, 2:720*f*
 - metabotropic glutamate (mGlu) receptors
 - functional role, 2:396–397
 - protein structure, 2:717*f*, 2:719
 - molecular structure, 2:716–717, 2:716*t*, 2:717*f*, 2:718*f*
 - neurodegenerative diseases, 2:720
 - physiological/pharmacological properties, 2:714–716, 2:715*t*
 - protein structure
 - AMPA receptors, 2:717–718, 2:718*f*
 - kainic acid receptors (KARs), 2:718
 - metabotropic glutamate (mGlu) receptors, 2:717*f*, 2:719
 - N-methyl-D-aspartate (NMDA), 2:717*f*, 2:718–719
 - transcription regulation, 2:719–720, 2:720*f*
 - glycine receptors (GlyRs), 2:419
 - intermediate filaments, 3:233
 - melanocortin system, 3:26
 - muscarinic acetylcholine receptors, 2:720*f*, 3:217
 - neuromodulation, 2:395
 - vitamin E, 4:547
 - voltage-gated calcium channels, 4:561*t*
 - central supramolecular activation cluster (cSMAC), 4:374, 4:374*f*
 - centrioles, 2:684*f*
 - centromeres, 1:446–450
 - centromere protein E (CENP-E), 1:417, 1:418–419, 1:419*f*
 - chromosome organization, 1:538
 - definition, 1:446
 - domain structures
 - characteristics, 1:446–447
 - heterochromatin domain, 1:446–447, 1:447*t*, 1:449*f*
 - kinetochore domain
 - central domain, 1:446, 1:447*t*
 - characteristics, 1:446, 1:449*f*
 - dynactin, 2:177
 - functional role, 1:447*t*
 - inner kinetochore, 1:446, 1:447*t*
 - outer kinetochore, 1:446, 1:447*t*
 - identity and function
 - as chromatin
 - centromeric histone CENP-A, 1:448–450, 1:449*f*, 1:538
 - characteristics, 1:448–450
 - epigenetic centromeres, 1:448
 - as deoxyribonucleic acid (DNA)
 - characteristics, 1:447–448
 - holocentric centromeres, 1:448
 - multisubunit centromeres, 1:447
 - point centromeres, 1:447, 1:448*f*
 - regional centromeres, 1:447, 1:448*f*
 - Schizosaccharomyces pombe*, 1:447, 1:448*f*
 - as ribonucleic acid (RNA), 1:450
 - meiosis, 3:18, 3:18*f*
 - mitosis, 3:17–18, 3:18*f*, 3:181, 3:182*f*
 - ceramides
 - ceramidase (CDase)
 - acid ceramidase (aCDase), 4:288
 - alkaline ceramidase, 4:289
 - neutral ceramidase (nCDase), 4:288–289
 - reaction mechanisms, 4:288–289
 - structural characteristics, 4:281–282, 4:282*f*, 4:283*f*
 - Ceratitis Capitata tra (Cetra)* gene, 1:86
 - cerebral amyloid angiopathy, 1:101*t*
 - cerebral cortex, 2:172–173, 2:172*t*
 - Cernunnos, 3:271
 - cerulenin, 2:278, 3:647
 - cesium-based bacterial reductive reactions, 1:487*t*
 - c-fes* proto-oncogene (Fes), 1:451–455
 - biological functions
 - angiogenesis, 1:452–453
 - myeloid differentiation, 1:452
 - neovascularization, 1:453
 - domain structures
 - characteristics, 1:451, 1:452*f*
 - F-BAR region, 1:451, 1:452*f*
 - Fes/Fer/CIP4 homology (FCH) domain, 1:451, 1:452*f*
 - kinase domain, 1:452
 - SH2/SH3 domains, 1:451–452, 1:452*f*
 - downstream signaling pathways
 - phosphatidylinositol 3-kinase (PI3K) pathway, 1:454
 - signal transducer and activator of transcription (STAT) signaling, 1:453–454
 - small guanosine phosphatases (GTPases), 1:453
 - expression regulation, 1:451
 - research summary, 1:454
 - substrate proteins, 1:454
 - transcriptional regulation, 1:454
 - c-fos* gene, 3:293
 - c-Fos transcription factor, 2:605–606
 - Cfr10I enzyme, 2:131, 2:132*t*, 2:133
 - chameleons, 1:552*f*
 - Chanarin–Dorfman syndrome (CDS), 1:49
 - Chaos carolinensis*, 3:148
 - chaperone proteins, 2:522–523, 2:523*f*
 - chaperonins, 1:456–460
 - allosteric regulation, 1:74
 - cellular functions, 1:459
 - chaperonin of 60 kDa (Cpn60), 1:456
 - GroEL (growth E locus, large gene, bacterial group I chaperonin)
 - allosteric regulation, 1:74
 - Escherichia coli*, 1:457–458, 1:458*f*, 1:459
 - GroEL–GroES protein-folding cycle, 1:457–458, 1:458*f*
 - ion-mobility spectrometry (IMS), 3:14–15, 3:15*f*
 - occurrences, 1:456
 - structural characteristics, 1:457*f*
 - GroES (growth E locus, small gene, bacterial group I chaperonin cofactor), 1:456, 1:457*f*, 1:458*f*
 - group I chaperonins, 1:456, 1:457*f*
 - group II chaperonins, 1:456–457, 1:457*f*
 - heat shock protein 60 (Hsp60), 1:456
 - heat shock protein 90 (Hsp90), 1:456
 - hereditary spastic paraplegias (HSPs), 4:271–272
 - protein folding processes, 1:456
 - research summary, 1:459–460
 - structural characteristics, 1:456–457, 1:458*f*
 - TRiC/CCT eukaryotic chaperonin, 1:457*f*, 1:458–459
 - tubulin synthesis, 4:451
 - Charcot–Marie–Tooth disease type 2C (CMT2C), 4:6, 4:416
 - Charcot–Marie–Tooth syndrome, 1:215, 2:632*t*, 3:236, 4:6, 4:443, 4:507*t*
 - Cheadle’s disease, 4:530
 - Chediak–Higashi syndrome, 4:507–508
 - chemical odorants, 1:46*t*
 - chemiluminescence, 1:461–466
 - basic principles, 1:461–462, 1:462*f*
 - definition, 1:461
 - light stick applications, 1:462, 1:463*f*
 - tetramethyldioxetane decomposition, 1:462, 1:462*f*
 - chemiosmotic theory, 1:467–472
 - basic concepts
 - ATP synthase, 1:468, 1:470
 - electron transfer chain, 1:468

- chemiosmotic theory (*Continued*)
 ion-impermeable membranes, 1:468, 1:468f, 1:471f
 major postulates, 1:468–472
 mitochondrial permeability transition, 3:162
 proton circuit, 1:470–471
 proton gradient coupling, 1:470
 protonmotive electron transfer chain, 1:469–470, 1:469f, 1:470f
 protonmotive force (PMF), 1:468, 1:468f, 1:471f
 stoichiometric conditions, 1:470–471
 transport systems, 1:471–472, 1:471f
 vectorial redox loop, 1:468–469, 1:469f
 future research outlook, 1:472
 historical research, 1:467
 research background, 1:467
- chemokine receptors, 1:480–485
 desensitization mechanisms, 1:482–483, 1:483f
 functional roles
 angiogenesis, 1:484
 central nervous system (CNS), 1:484–485
 chemotaxis, 1:483–484
 human immunodeficiency virus (HIV-1), 1:484
 tumor growth and metastasis, 1:484
 internalization, 1:483
 intracellular signaling pathways, 1:480–482, 1:482f
 intracellular trafficking, 1:482–483, 1:483f
 ligand binding, 1:480–482, 1:481f, 1:482f
 subfamily classifications, 1:480, 1:481t
- chemokines, 1:619, 1:619f
- chemolithotrophy, 1:486–491
 bacteria
 ATP synthesis, 1:488f, 1:490–491
 carbon sources, 1:491–492
 characteristics, 1:488–489
 ecological characteristics, 1:492
 electrochemical gradient generation, 1:488f, 1:490–491
 growth-supporting reductant oxidation
 carbon monoxide-oxidizing bacteria, 1:490, 1:490t
 hydrogen-oxidizing bacteria, 1:490, 1:490t
 metal-oxidizing bacteria, 1:490, 1:491t
 methanotrophic bacteria, 1:489–490, 1:489t
 nitrifying bacteria, 1:489, 1:489t
 sulfur-oxidizing bacteria, 1:490, 1:490t
 growth yields, 1:492
 nicotinamide adenine dinucleotide (NAD⁺/NADH) production, 1:491
 periplasmic release, 1:491
 proton-related cytoplasmic utilization, 1:491
 reductive reactions, 1:487t
 substrate-level phosphorylation, 1:491
- bioenergetic reactions
 ATP synthesis, 1:486–487, 1:487f
 growth categories, 1:488
 reductants, 1:487, 1:487f, 1:488f, 1:488t
 terminal electron acceptor (TEA), 1:487, 1:487t
 microbe classifications, 1:488, 1:488t
 MutY repair pathway, 2:58
- chemoorganotrophs, 1:488, 1:488t
- chemotaxis, 2:236–240
 characteristics and functional role, 2:236
 chemokine receptors, 1:483–484
 cytoskeletal polarization, 2:238
 mathematical models, 2:238–240, 2:238f
 metrics, 2:239b
 signaling pathways, 2:236–238, 2:237f
- chemotherapy
 cancer cell-based protein expression, 3:210–211
 multidrug resistance (MDR) membrane proteins, 3:210–214
 ABCG2 (MXR, BCRP), 3:213, 3:214
 BSEP, 3:213, 3:214
 homologous transporters, 3:213
 MDR1
 clinical relevance, 3:210–211
 functional role, 3:213
 high affinity–substrate relationships, 3:213
 physiological functions, 3:213
 polyspecific transport cycle, 3:211–213, 3:212f
 research background, 3:210
 structural characteristics, 3:211–213, 3:211f
 MDR3, 3:213, 3:214
 MRP1 (ABCC1), 3:213, 3:214
 physiological functions
 ABCG2 (MXR, BCRP), 3:214
 ABCG5 and ABCG8, 3:214
 advantages, 3:214
 BSEP, 3:214
 homologous transporters, 3:214
 MDR1, 3:213
 MDR3, 3:214
 MRP1 (ABCC1), 3:214
 research summary, 3:214
- chenodeoxycholate, 1:168f, 1:169
- cherries, 4:533t
- chiasmata, 3:18, 3:19f
- chickens
 ascorbic acid content, 4:533t
 color vision, 1:552f
- chicory, 4:533t
- Chi sequence, 4:32
- chitin-binding domain (ChiBD), 1:62
- chives, 4:533t
- Chk2 protein, 3:269, 3:273t
- chlamydiae, 3:559
Chlamydia pneumoniae, 3:331f
Chlamydia psittaci, 3:331f
Chlamydia trachomatis, 3:331f
Chlamydomonas, 1:513, 1:513f, 2:185t, 2:186
Chlamydomonas reinhardtii, 1:66–70, 2:296, 3:486, 4:335–338, 4:339–340
Chlorella, 2:35–37, 2:37f
- chlorin
 spectral properties, 1:499–500, 1:500f
 structural characteristics, 1:499, 1:499f
- chlorine (Cl)
 kinetic isotope effects (KIEs), 2:689–690, 2:689f, 2:690f
 photosystem II (PSII) complex, 3:514
 potassium-chlorine (K-Cl) cotransporter 2 (KCC2), 1:199, 1:200f
 sodium-potassium-chlorine (Na-K-Cl) cotransporter isoform 1 (NKCC1), 1:199, 1:200f
Chlorobaculum tepidum, 2:502f, 2:504–507, 2:504f, 2:505f
Chlorobium, 3:32
 Chlorobium–Flavobacterium group, 3:559
Chlorobium tepidum, 2:510–512, 3:331f
Chloroflexus, 2:209, 3:32
Chloroflexus aurantiacus, 2:503–504, 2:510, 3:700–702
- chlorophototrophs, 2:501–503
- chlorophyll, 1:498–505
 applications, 1:501
 bacteriochlorophyll (BChl)
 characteristics, 2:513, 2:515f
 chlorosomes, 2:502f, 2:504–507
 Fenna–Matthews–Olson protein, 2:504f, 2:505f, 2:507
 green bacteria, 2:337–338, 2:501–503, 2:502f, 2:503–504
 heliobacteria, 2:337–338
 homologs, 2:514t
 molecular structure, 2:503f, 2:514f
 photosynthetic reaction centers
 electron acceptors and donors, 3:696–697, 3:696f
 green bacteria, 2:501–503, 2:504f, 2:505f
 molecular characterization, 3:700–702, 3:702f
 subunit structure and functions, 3:701f
 type I reaction centers, 2:337–338
- biosynthetic pathway, 1:501, 3:545f, 3:546, 3:546f
- catabolic reactions, 3:547–549
 characteristics, 3:544
 chemical properties, 1:501
 cytochrome *b₆* complex, 4:337f, 4:338
 functional role, 1:501
 light harvesting centers (LHCs), 1:500
 light harvesting proteins, 1:498, 2:178–179, 2:179f, 2:181f, 3:482, 3:492
- metabolism, 1:501
 occurrences, 1:501
 photosystem I (PSI) complex, 2:336–337, 2:337f, 3:503, 3:504f, 3:506–507
 phototrophic bacteria, 3:32
 reactive oxygen species (ROS), 1:500, 3:482–484, 3:483f
 spectral properties, 1:499–500, 1:500f
 structural characteristics, 1:499, 1:499f
- chlorophyllide (Chlide), 1:508, 2:503–504, 2:506f
- chloroplasts, 1:506–510

- biogenesis
 basic concepts, 1:508
 etioplasts, 1:508–509, 1:509f
 light requirements, 1:508
 plastids, 1:509–510, 1:509f
- coenzyme Q, 3:723t
- iron–sulfur (FeS) protein biogenesis, 2:663
- organization
 envelope, 1:506–507, 1:507f
 general discussion, 1:506
 grana, 1:511–512, 1:512f
 stroma, 1:507, 1:507f, 1:511–512, 1:512f
 thylakoid system, 1:507–508, 1:507f, 1:508f, 1:511–512, 1:512f
- origins, 1:506
- photosynthesis, 3:492, 3:493f
- photosynthetic electron transport chain, 1:511–515
 basic concepts, 1:511–512, 1:512f
 chloroplast protein kinase network, 1:514, 1:514f
 electron flow pathways, 1:512
 ferredoxin/thioredoxin system, 1:514–515
 plastoquinone pool
 electron flow pathways, 1:511–512, 1:512f
 long-term responses, 1:513–514
 redox state, 1:512–514, 1:514f
 short-term responses, 1:513, 1:513f
 state transitions, 1:513, 1:513f
 thylakoid protein kinase network, 1:514f
 redox control, 1:512
 retrograde signaling pathways
flu mutant, 1:515
Gun mutants, 1:515
 reactive oxygen species (ROS), 1:515
 thylakoid protein kinase network, 1:514, 1:514f
- protonmotive electron transfer chain, 1:470f
- chloroquine, 1:150t
- chlorosomes, 2:513–516
 characteristics
 bacteriochlorophyll (BChl), 2:502f, 2:504–507, 2:513, 2:515f
 carotenoids, 2:514
 proteins, 2:514
 size and internal structure, 2:513, 2:515f
 energy transfer mechanisms, 2:515f
 evolutionary considerations, 2:516
 green sulfur bacteria, 2:501–503, 2:502f
 homologs, 2:514t
 molecular structure, 2:514f
 organizational models, 2:504f, 2:514–515, 2:516f
 phylogenetic relationships, 2:513
 spectral properties, 2:514, 2:515f
- Cho homologs, 3:330, 3:331f
- cholangiocytes, 1:190–191, 1:191f, 1:192f, 1:193–195, 1:194f
- cholate, 1:168f, 1:169
- cholecalciferol (vitamin D₃), 4:535, 4:536f
- cholecystokinin (CCK)
- arachidonic acid regulated Ca²⁺ (ARC) channels, 1:131
- ascorbic acid, 4:531–532
- body weight regulation, 3:78–79, 3:79f
- characteristics and functional role, 1:115
- digestion and absorption, 2:355, 2:355f, 2:357–358
- G_q family proteins, 2:497t
- peptide amidation, 3:396
- cholera
 cholera toxin A subunit (CTA), 1:134–135
 G_s family, 2:519
- cholestasis
 bile salt transport, 1:170
 primary biliary cirrhosis (PBC) treatment, 3:561
- cholesterol, 1:516–520
 Alzheimer's disease, 1:478
 biosynthetic pathway
 general discussion, 1:516–519
 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase regulated degradation, 1:518f, 1:519
 inhibitors
 bisphosphonates, 1:518f, 1:519–520
 statins, 1:519
 mutation effects, 1:519
 schematic diagram, 1:517f
 transcriptional control, 1:517–519, 1:518f
 chemical structure, 1:251f
 digestion and absorption, 2:358
 fatty acyl-coenzyme A (CoA) synthetases, 2:291f
 functional role
 membrane structure, 1:516
 signaling molecule precursor, 1:516
 lipid bilayer structure, 2:733
 lipid rafts, 2:739–740, 2:742f
 microRNA (miRNA)-based regulation, 4:194t
 properties, 1:516
 sources
 biosynthesis, 1:516–519
 diet, 1:516
 lipoproteins, 1:516
- cholesteryl esters (CEs), 1:47, 1:49–51
- cholic acid, 1:645f
- choline, 3:479f, 3:480
- chondrocytes, 2:521t
- chondroitin sulfate (CS), 1:439–440, 3:654–655, 3:655f
- Chordin protein family, 1:229–230
- Chromatiaceae, 3:698t
- chromatin
 actin-related proteins (Arps), 1:40
 centromere identity and function
 centromeric histone CENP-A, 1:448–450, 1:449f, 1:538
 characteristics, 1:448–450
 epigenetic centromeres, 1:448
 chromatin immunoprecipitation (ChIP) analyses, 2:750
 chromatin interaction analysis by paired-end tag sequencing (ChIA-PET), 2:750–751
- 1 α ,25-dihydroxyvitamin D₃ (1,25-(OH)₂D₃), 4:537–538, 4:538f
- DNA damage checkpoints proteins, 1:413f, 1:414
- euchromatin, 1:521
- GAL promoter system, 1:525–529
 chromatin regulation function
 active (galactose-induced) state, 1:527
 inactive states, 1:527–528
 promoter nucleosomes—regulatory factors interactions, 1:528–529
- GAL1–10 gene expression
 basic concepts, 1:525–527
 gene-specific factors, 1:525–526
 hypersensitive region (HR), 1:526–527, 1:526f
 nucleosome positioning, 1:525
 promoter chromatin organization, 1:526–527, 1:526f
 promoter elements, 1:525, 1:526f
 regulatory and transcription factors, 1:525–526
- gene-specific factors
 Gal3p, 1:525–526
 Gal4p, 1:525, 1:528–529
 Gal80p, 1:525, 1:528–529
 global process factors, 1:526
 promoter chromatin organization
 GAL1–10 gene expression, 1:526–527, 1:526f
 regulation function, 1:529
- heterochromatin, 1:521, 3:314–315, 3:315f
- high-throughput chromosome capture (Hi-C) technology, 2:750–751
- methyl-CpG-binding proteins, 1:521–524
 binding domain family
 general discussion, 1:522
 MBD1, 1:523, 1:523f
 MBD2, 1:523, 1:523f
 MBD3, 1:523, 1:523f
 MBD4, 1:523–524, 1:523f, 4:77
 MBD5, 1:523f, 1:524
 MBD6, 1:523f, 1:524
 MeCP2, 1:522–523, 1:523f
- cytosine methylation
 basic concepts, 1:522
 cancer cells, 1:522
 functional role, 1:522
 gene expression regulation, 1:522
 genomic imprinting, 1:522
 X-chromosome inactivation, 1:522
- DNA methylation—chromatin structure
 interdependency, 1:524
 functional role, 1:521
 gene expression regulation, 1:521
 nucleosomes, 1:521, 1:526f, 1:528–529, 1:535–537, 1:536f, 3:314–315, 3:315f
- physical organization, 1:530–534
 background and characteristics, 1:530
- core histones
 chromatin structure, 1:535–537
 gene variants, 1:531–532
 histone H1, 1:531
 nonhistone chromatin proteins (NHCPs), 1:531–532

- chromatin (*Continued*)
- post-translational modifications, 1:530–531
 - dynamic characteristics, 1:534
 - hierarchical classification, 1:532, 1:532t
 - in vivo* organization, 1:534
 - structural characteristics
 - global primary structures, 1:532–534
 - global secondary structures, 1:533, 1:533f
 - local primary structures, 1:532–533
 - local secondary structures, 1:533
 - nucleosome repeat length (NRL), 1:532–534
 - tertiary structures, 1:533–534
 - 30 nm chromatin fiber, 1:533, 1:533f
 - proteomic studies, 3:326f
 - regulation mechanisms, 4:59–66
 - ATP-dependent nucleosome remodeling
 - chromodomain helicase DNA binding (CHD)/Mi-2 family, 4:60f, 4:64
 - INO80 (inositol requiring 80) family, 4:60f, 4:64–65, 4:65f
 - ISWI family, 4:60f, 4:64
 - remodeling enzymes, 4:63–65
 - schematic diagram, 4:65f
 - SWItching defective/Sucrose NonFermenting (SWI/SNF) family, 1:314, 4:60f, 4:63–64, 4:403–404
 - Swr1-dependent deposition of H2A. Z-H2B dimers, 4:64–65, 4:65f
 - histone modifications
 - acetylation, 1:521, 1:522f, 4:60
 - citullination, 4:62
 - covalent modifications, 1:521, 1:522f, 4:59
 - histone variants, 4:60f, 4:62–63
 - methylation, 1:521, 1:522f, 4:60–61
 - phosphorylation, 1:314–315, 1:521, 1:522f, 4:61
 - proteolysis, 4:62
 - sumoylation, 4:62
 - ubiquitylation, 4:61–62
 - Ran protein family, 4:10
 - remodeling mechanisms, 1:37f, 4:66
 - research summary, 4:66
 - S phase regulation, 1:405
 - structural characteristics, 2:749, 3:314–315, 3:315f, 4:59
- Chromatium*, 3:32
- chromium-based bacterial reductive reactions, 1:487t
- Chromobacterium violaceum*, 3:3–4
- chromodomain helicase DNA binding (CHD)/Mi-2 family, 4:60f, 4:64
- chromodomains, 4:60–61
- chromoplasts, 1:509–510, 1:509f
- chromosomes
- chromosome conformation capture (3C), 2:750–751
 - chromosome organization, 1:535–538
 - chromatin structure, 1:535–537, 1:536f, 2:749, 3:315–316, 3:315f
 - mitotic chromosome, 1:537–538
 - nucleosome structure, 1:535–537, 1:536f
 - research summary, 1:538
 - scaffold attachment regions (SARs), 1:537–538
 - three-dimensional structure, 1:537
 - DNA-strand transferases, 4:36f, 4:37
 - dynactin, 2:177
 - high-throughput chromosome capture (Hi-C) technology, 2:750–751
 - LacI/lacO system, 2:751, 2:751f
 - mitotic checkpoint, 1:417–422
 - anaphase-promoting complex (APC), 1:419–420, 1:419f, 1:420f
 - aurora B/Ipl1 protein, 1:418
 - centromere protein E (CENP-E), 1:418–419, 1:419f
 - checkpoint pathways, 1:419
 - checkpoint protein roles, 1:421–422
 - direct inhibitor model, 1:420–421, 1:420f
 - functional role, 1:417
 - kinetochore monitoring, 1:417
 - microtubule attachment monitoring, 1:417–418
 - molecular components, 1:417
 - sequestration model, 1:419–420, 1:419f
 - tension monitoring, 1:417–418
 - unattached kinetochores, 1:422
 - S phase regulation, 1:405
 - telomeres, 4:380–386
 - characteristics and functional role, 4:380
 - chromosome organization, 1:538
 - non-homologous end joining (NHEJ), 3:272, 4:384–385
 - replication processes
 - general discussion, 4:384
 - negative regulation, 4:384, 4:385f
 - positive regulation, 4:384
 - research background, 4:380–384
 - telomerase-independent maintenance, 4:384
 - telomerase repression
 - disease implications, 4:385
 - stem cells, 4:386
 - tumor suppression, 4:385–386
 - telomere capping, 4:384–385
 - telomeric complex
 - telomerase, 4:381–382, 4:382f
 - telomere proteins, 4:382–384, 4:383f
 - telomeric DNA, 4:380–381, 4:381f
 - telomeric RNA, 4:384
 - territories, 3:297, 3:297f
 - chronic wasting disease (CWD), 3:575t
 - chymosin, 1:137
 - chymotrypsin, 2:758–759, 2:758f
 - Cialis, 1:569
 - ciano fluorescent protein (CFP), 2:577–578
 - Cicadella viridis*, 1:125
 - cicatrical pemphigoid, 1:641t
 - ciguatoxin, 4:568t
 - ciliary neurotrophic factor (CNTF), 1:198
 - Ciona intestinalis*, 2:641f, 3:384
 - ciprofloxacin, 2:167, 2:167f
 - cirazoline, 1:58
- cirrhosis
- chronic liver injury repairs, 1:193–195, 1:194f
 - Hedgehog (Hh) pathway reactivation, 1:193–195, 1:194f
 - primary biliary cirrhosis (PBC), 3:561–565
 - anti-mitochondrial antibodies (AMAs), 3:114, 3:115–116
 - background information, 3:561
 - clinical features, 3:563–564
 - diagnosis, 3:564
 - etiology
 - associated autoimmune diseases, 3:562
 - autoantibodies/antimitochondrial antibodies, 3:561–562
 - mitochondrial antigens, 3:562
 - laboratory tests, 3:564
 - pathology
 - developmental stages, 3:563
 - stage III PBC, 3:563f
 - stage II PBC, 3:563f
 - stage I PBC, 3:563f
 - stage IV PBC, 3:564f
 - pathophysiology, 3:562–563
 - prevalence, 3:561
 - symptom treatments
 - anemia, 3:564–565
 - fatigue, 3:564
 - osteoporosis, 3:564–565
 - pruritus, 3:564
 - vitamin malabsorption, 3:564
 - underlying disease process treatments
 - general discussion, 3:565
 - liver transplantation, 3:565
 - medical treatments, 3:565
- cis*-platin, 4:405
- 5 CITEP, 2:106f
- citrate
- anaplerotic pathways, 1:106f, 1:108f, 1:109f
 - body weight regulation, 4:58f
 - carbohydrate-responsive element binding protein (ChREBP) function, 1:376–377, 1:377f
 - citrate synthase, 4:439f, 4:439t
 - citric acid cycle, 4:93f
 - fatty acid synthesis pathway, 2:275–279, 2:276f
 - gluconeogenesis, 2:386f
 - liver pyruvate kinase (LPK) catalysis, 1:374
 - thiamine diphosphate (ThDP), 1:207f
 - tricarboxylic acid cycle, 4:439f
- citric acid
- citric acid cycle
 - anaplerotic pathways, 1:105
 - flavins, 2:311
 - historical research, 4:437–438, 4:438f
 - insect biochemistry, 2:590–591
- citrin deficiency
- adult-onset citrullinemia, 4:499
 - characteristics, 4:499
 - neonatal intrahepatic cholestasis, 4:499
- Citrobacter freundii*, 2:132t
- citrullination, 4:62

- citrulline
 nitric oxide synthase (NOS), 3:254f
 urea cycle, 4:495f
 clathrin-coated pits, 1:516, 2:197
 clathrin-mediated endocytosis
 adaptor proteins, 2:197–198
 basic concepts, 2:197, 2:198f
 dynamin, 2:198
 claudins, 4:392, 4:393f, 4:394, 4:394f
 cleavage and polyadenylation specificity
 factor (CPSF), 3:189–191, 3:190t
 cleavage factor I (CFI), 3:189–191, 3:190t
 cleavage factor II (CFII), 3:189–191, 3:190t
 cleavage stimulation factor (CstF),
 3:189–191, 3:190t
 clonidine, 1:58–59
 Clostridia, 3:559, 4:73–74
Clostridium, 3:30
Clostridium acetobutylicum, 3:331f
Clostridium butyricum, 2:205t
Clostridium pasteurianum, 2:297, 2:565, 4:72
Clostridium thermocellum, 4:73–74, 4:74f
 clustered regularly interspaced short
 palindromic repeat (CRISPR) RNAs
 (crRNAs), 4:253, 4:254f
 CMP-sialic acid (CMP-NeuAc), 1:3
 c-Myb transcription factor, 1:187, 1:188f
 c-Myc transcription factor, 2:605–606
 cnidarians, 3:398
 coactivator CREB-binding protein (CBP),
 3:294–295
 cobalamin, 1:155–160
 absorption and transport, 2:327
 acetogenesis, 1:159–160
 biochemistry, 2:327
 catalytic reactions
 carbon skeleton rearrangement
 reactions, 1:157, 1:157f, 1:158t
 methyl transfer reactions, 1:156f,
 1:158–159, 1:159f
 migration and elimination reactions,
 1:156, 1:157f, 1:158t
 pyridoxal 5'-phosphate (PLP)-
 dependent aminomutase reactions,
 1:157, 1:157f, 1:158t
 radical-mediated rearrangement
 reactions, 1:156, 1:157f, 1:158t,
 2:327
 ribonucleotide triphosphate reductase,
 1:157f, 1:158, 1:158t
 chemical characteristics and structure,
 2:326, 2:326f
 cobalamin cofactor, 1:155–156, 1:156f
 cobalamin-dependent methionine
 synthase, 1:159, 1:159f
 deficiency-related diseases, 2:328
 functional role
 metabolic reactions, 2:327
 methionine synthase, 2:327
 methylmalonyl CoA mutase, 2:327
 gastric secretions, 2:356
 methanogenesis, 1:159–160
 structural characteristics, 1:156f
 cocaine and amphetamine-regulated
 transcript (CART), 2:257–258
 cochlea, 4:561t
 Cockayne syndrome (CS), 3:339–340,
 3:342t, 3:344, 4:403–404
 codon synthesis, 3:432–433, 3:433t, 3:434t,
 4:122–123
 coenzyme A (CoA), 1:539–541
 acyl-S-coenzyme A (CoA) derivatives,
 1:539–540, 1:540f
 enzymatic functions
 acyl carbonyl- and α -carbon atom
 activation mechanisms, 1:539,
 1:540f
 chemical action mechanisms, 1:539
 fatty acid synthase, 1:539
 fatty acids
 fatty acid biosynthesis
 energy metabolism regulation, 2:286
 pyruvate dehydrogenase regulation,
 2:285–286
 fatty acid synthase
 4'-phosphopantetheine (4'-PP), 1:539
 synthesis process, 1:541, 2:285–286
 fatty acyl-coenzyme A (CoA) synthetases,
 2:290–295
 ACSF proteins, 2:294
 activation reactions, 2:290, 2:291f
 amino acid sequence motifs,
 2:290–291, 2:291f, 2:294–295
 Bubbligum ACS family (ACSBG),
 2:294
 characteristics and occurrence,
 2:290–294
 human ACS families, 2:290–294,
 2:292t
 long-chain acyl-CoA synthetase
 (LCAS), 1:385f, 2:292–293
 medium-chain acyl-CoA synthetase
 (MCAS), 2:292
 nomenclature, 2:292, 2:292t
 research summary, 2:295
 short-chain acyl-CoA synthetase
 (SCAS), 2:292
 structure–function relationship,
 2:291f, 2:294–295
 very long-chain acyl-CoA synthetase
 (VLCAS), 2:293–294
 oxidation processes, 2:281
 historical research, 1:539
 metabolic functions, 1:540–541
 β -methylcrotonoyl-coenzyme A (CoA)
 carboxylase, 1:221
 mitochondrial acyl-CoA/CoA ratio
 regulation, 1:387
 molecular structure, 1:540f
 protein biotinylation, 1:226
 coenzyme B₁₂
 biosynthetic pathway, 3:545f, 3:546–547,
 3:546f
 characteristics, 3:544
 coenzyme Q
 biosynthesis, 3:726f, 3:727
 chemical characteristics, 3:723–724
 Friedreich's ataxia (FRDA), 2:333
 functional role
 antioxidant actions, 3:725–727
 electron transfer mechanisms,
 3:723–724, 3:724f, 3:725f
 plasma membrane oxidoreductase,
 3:725
 Q-cycle, 1:511–512, 3:725, 3:725f
 homologs and analogs, 3:723–724, 3:723t
 mitochondrial ROS, 1:210
 PET genes, 3:307–309, 3:308f
 respiratory chain complexes, 4:83
 UCP1-induced proton cycling, 4:484f
 See also respiratory chain complex II
 (succinate coenzyme Q reductase)
 cofilin, 1:14f, 1:15, 2:345f
 cohesins, 2:557, 2:558f
 colchicine
 antimitotic tubulin ligands, 4:452
 macroautophagy regulation, 1:150t
 multidrug resistance (MDR) membrane
 proteins, 3:212f
 collagenase, 1:542–544
 basic concepts
 original definition, 1:542
 revised definition, 1:542–543
 characteristics, 1:543
 classifications, 1:543t
 fibrillar collagenases, 1:543t
 functional roles, 1:543–544
 historical research, 1:542
 matrix metalloproteinases (MMPs),
 3:90–91, 3:91t, 3:92f
 research summary, 1:544
 collagens, 1:545–549
 aging effects, 2:409, 2:410f
 ascorbic acid, 4:531–532
 biosynthesis, 1:547f
 cell–matrix interactions, 1:413f, 1:426,
 1:427t, 1:438–440, 1:444–445, 1:445f
 fibrillar collagens, 1:545–546, 1:546f
 focal adhesion kinase (FAK) family,
 2:265–266
 glycation processes, 2:406
 integrins, 1:427t, 1:438–440, 1:439f
 nonfibrillar collagens, 1:546, 1:546f
 proteoglycans, 3:657t, 3:658
 signature (Gly-X_n-Y-),_n sequence
 basic concepts, 1:545
 biosynthesis, 1:547f
 enzyme functions, 1:548
 evolutionary origins, 1:548
 hydroxyproline content, 1:546–547
 procollagens, 1:547–548
 schematic diagram, 1:546f, 1:547f
 self-assembly control, 1:546
 stromal matrix, 1:437, 1:438f, 1:442–443,
 1:443f
 tumor-associated collagen signatures
 (TACSs), 1:443f
 type XVII collagen, 1:640f, 1:641, 1:641t
 collision-induced dissociation (CID), 3:12,
 3:12f, 3:13–14, 3:14f
 colon
 aquaporin water channels, 1:124t
 long-chain acyl-CoA synthetase (LCAS),
 2:292–293
 microRNA (miRNA)-based regulation,
 4:194t
 proteinase-activated receptors (PARs),
 3:605t

- colon cancer
 DNA mismatch repair (MMR), 2:61–64, 2:62*t*, 2:63*f*
 fatty acid synthase expression, 2:277, 2:277*f*
 tRNA synthetases, 4:443
- color vision, 1:550–554
 definition, 1:550
 evolutionary development, 1:552*f*, 1:553, 1:553*f*
 nerve signals, 1:552–553
 photoreceptors, 1:550, 1:551*f*
 primates, 1:554
 rods and cones
 characteristics, 1:550
 cone opsin genes, 1:551–552, 1:551*f*, 1:552*f*
 cone photopigments, 1:550–551, 1:551*f*
 evolutionary development, 1:553*f*
 schematic diagram, 1:551*f*
- comitin, 3:438–439
- common lymphoid progenitor (CLP), 1:187–188, 1:188*f*
- comparative gene identification-58 (CGI-58), 1:49–51, 1:50*f*
- compound transposons, 3:276, 3:276*f*
- concanavalin A, 2:460–461, 2:703*f*
- concatemer junction (CJ), 4:356*f*, 4:358
- condiments, 4:533*t*
- congenital erythropoietic porphyria, 2:542*t*
- congenital stromal corneal dystrophy, 3:657*t*
- Congo red staining method, 1:100–101, 1:101*f*
- conjugative transposons, 3:276*f*, 3:277
- connexin, 2:618
- conopressin, 4:516*t*
- conotoxins, 4:568*t*
- conservative site-specific recombination, 1:555–561
 basic concepts, 1:1
 biological characteristics, 1:555–556
 integration and excision reactions, 1:555–556, 1:556*f*, 1:560*f*
- recombinases
 reaction mechanisms, 1:556
 resolvase/invertase mechanisms, 1:558–559, 1:559*f*
 serine recombinases, 1:558–559, 1:559*f*
 tyrosine recombinases, 1:556–558, 1:557*f*, 1:558*f*
- recombination control mechanisms
 outcome control, 1:6–7, 1:560*f*
 site polarity, 1:559
 research summary, 1:561
- constitutive secretory vesicles, 4:204*f*
- contractile ring, 4:214, 4:215*f*
- Conus snails, 4:551
- conventional actin, 1:38–40, 1:39*f*
- COP9 signalosome (CSN), 4:468*t*, 4:470
- copper (Cu)
 Alzheimer's disease, 1:478
 copper amine oxidase, 1:88–89
 copper-transporting ATPases (copper pumps), 3:670–675
 Atox-1 copper chaperone complex, 3:674
 ATP7A/ATP7B, 3:670–671, 3:671*f*, 3:674
 biological functions
 copper homeostasis, 3:670–671, 3:671*f*
 fungi, 3:671–672
 human health functions, 3:670–671
 occurrences, 3:670
 plants, 3:671–672
 prokaryotes, 3:671–672
 characteristics and functional role, 4:333*f*, 4:334
 functional role, 3:673
 mammalian copper pump regulation
 kinase-mediated phosphorylation, 3:674
 trafficking pathways, 3:674
 molecular structure
 cytosolic catalytic domains, 3:673
 domain structures, 3:672–673, 3:672*f*
 N-terminal metal-binding domain (N-MBD), 3:674
 transmembrane domains, 3:672–673, 3:672*f*, 3:673–674
 heme-copper centers, 3:664–665
 peptidylglycine α -hydroxylating monooxygenase (PHM)
 basic concepts, 3:397–398
 impaired amidation, 3:399
 pharmacological inhibition, 3:398
 reaction mechanisms, 3:397
 structural characteristics, 3:397, 3:397*f*, 3:398*f*
 plastocyanin (PC), 3:530, 3:531*f*
 superoxide dismutases (SODs), 4:352–353
- coproporphyrin, 2:542*t*
- coproporphyrinogen III, 2:540*f*, 2:541
- coproporphyrinogen oxidase, 2:541, 2:542*t*
- coral tree, 2:702*t*
- core histones
See histones
- coriander, 4:533*t*
- Cori cycle, 2:381
- Cori/Forbes disease, 2:427*t*, 2:431–432, 2:431*t*
- cornea, 1:124*t*
- coronary heart disease (CHD), 4:547
- coronins, 1:22*f*, 1:24, 3:438–439
- corpus callosum hypoplasia, mental retardation, adducted thumbs, spastic paraplegia, and hydrocephalus (CRASH) syndrome, 4:272
- cortactin, 1:30*f*
- corticobasal degeneration, 1:477*t*
- corticotrophin
 ascorbic acid, 4:531–532
 corticotrophin-releasing factor (CRF), 3:78, 3:78*f*
 corticotropin-releasing hormone (CRH), 3:396
- cortisol
 amino acid metabolism, 1:95–96, 1:96*f*
 characteristics and functional role, 4:308–309
- cortistatin, 4:262, 4:263*f*
- cow's milk, 4:533*t*
- Coxiella burnetii*, 3:331*f*
- coxsackie and adenovirus receptor (CAR), 4:392, 4:394*f*
- CpG dinucleotides
 CXXC domain, 1:523, 1:523*f*, 3:8–9, 3:8*f*
 Dnmt3a dimers, 3:6
 methyl-CpG-binding proteins, 1:521–524
 binding domain family
 general discussion, 1:522
 MBD1, 1:523, 1:523*f*
 MBD2, 1:523, 1:523*f*
 MBD3, 1:523, 1:523*f*
 MBD4, 1:523–524, 1:523*f*, 3:9, 3:10*f*, 4:77
 MBD5, 1:523*f*, 1:524
 MBD6, 1:523*f*, 1:524
 MeCP2, 1:522–523, 1:523*f*
- cytosine methylation
 basic concepts, 1:522, 3:6, 3:7*f*
 cancer cells, 1:522
 functional role, 1:522
 gene expression regulation, 1:522
 genomic imprinting, 1:522
 X-chromosome inactivation, 1:522
- DNA methylation—chromatin structure interdependency, 1:524
 functional role, 1:521
 gene expression regulation, 1:521
 glycosylation processes, 3:9, 3:10*f*
- crab, 4:533*t*
- Crassulacean acid metabolism (CAM) pathway, 3:495–496, 3:498
- creatine kinase, 3:81, 3:82*f*
- creatinine, 4:306*f*
- CREB (cyclic adenosine monophosphate (cAMP) responsive binding protein)
 adipocyte development, 1:53*f*, 1:55
 calcium ion (Ca²⁺) signaling system, 2:719–720, 2:720*f*, 3:294–295
 coactivator CREB-binding protein (CBP), 3:294–295
 CREB shut-off pathway, 3:295, 3:295*f*
 gene expression regulation, 1:311, 1:313, 1:313*f*
 glycogen synthase kinase-3 (GSK3), 2:436*t*
 hypoxia inducible factor (HIF), 4:67
 insulin signal transduction, 2:605–606
 intracellular calcium ion (Ca²⁺) waves, 2:646–647
 neurogenesis, 1:197*f*, 1:199–200, 1:200*f*
 neurotrophins, 3:241–243, 3:242*f*
 nuclear calcium, 1:313
- Crepis capillaris*, 3:19*f*
- Cre protein, 1:556–557, 1:557*f*, 1:559
- CREST syndrome, 3:562
- Creutzfeldt-Jakob disease
 amyloid deposits, 1:101*t*, 1:102, 3:630
 epidemiology and diagnosis, 3:575
 scrapie prions, 3:574–575, 3:575*t*
 tau proteins, 1:477*t*
- Cricetulus griseus*, 2:5*t*
- Crigler-Najjar syndrome type I and II, 3:548
- cristae, 2:375–376, 2:375*f*
- Crithidia fasciculata*, 3:551
- Crk protein, 2:266, 2:267*f*
- crocabin, 1:291*t*, 1:295
- Crp proteins, 3:367–368

- cruciforms
background information, 2:146
biological role, 2:146
chemical properties, 2:146
pathology, 2:147, 2:147f
structural characteristics, 2:143f
- Cryptococcus neoformans*
N-myristoyltransferase (Nmt), 3:641, 3:643–644
pheromone receptors, 3:441
superoxide dismutases (SODs), 4:353
- cryptogenic cirrhosis, 2:632t
- CSA/CSB (Cockayne syndrome) protein complex, 3:342t, 3:344, 4:404
- Csk protein kinase, 4:295–296, 4:296f
- Csm protein complexes, 2:514
- C-terminals
adseverin, 1:277–278
calcitonin peptide family, 1:270–271
calcium-sensing receptor (CaSR), 1:326–328, 1:327f
calpains, 1:354–359
carboxyl methylation, 3:608–609, 3:608f
chemokine receptors
ligand binding, 1:480–482, 1:482f
subfamily classifications, 1:480
- C-terminal domain
AAA (ATPases associated with various cellular activities) proteins, 1:1–2, 1:2f
actin depolymerizing factor (ADF)/cofilin (ADF-H) domain, 1:23–24
alternative oxidases (AOXs), 1:146–147, 1:147f
alternative splicing, 1:76
arrestins, 2:489–490, 2:489f
ATP synthase, 2:271–272, 2:273f
bullous pemphigoid antigen 1 (BPAG1), 2:626f, 2:628–629
c-fes proto-oncogene (Fes), 1:451, 1:452f
copper-transporting ATPases (copper pumps), 3:672–673, 3:672f
cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:562–563, 1:563f, 1:564
cyclic nucleotide-dependent protein kinases, 1:576f, 1:577–578
cytoplasmic dynein, 2:185–186, 2:186f
degradosome complex, 3:53
diaphanous autoinhibitory domain (DAD), 1:24
DNA gyrase, 2:164–165, 2:165f
DNA ligases, 2:30–31, 2:33, 2:34f
DNA polymerase III holoenzyme, 2:93
DNA-strand transferases, 4:37–38, 4:37f
elongation cycle, 2:253
glucokinase glucose sensors, 2:601f
glycine receptors (GlyRs), 2:421, 2:422f
glycoprotein hormone receptors, 4:388, 4:389f
hexameric enzyme helicases, 2:23f
initiation factors
initiation factor 1 (IF1), 4:427, 4:428f
initiation factor 2 (IF2), 4:428–430, 4:428f
initiation factor 3 (IF3), 4:428f, 4:430
inositol requiring 1 (IRE1) genes, 4:489, 4:490f
inositol triphosphate (IP₃), 2:586, 2:587f, 2:653–655, 2:654f, 2:655f
intermediate filaments, 3:233
iron–sulfur clusters (FeS), 4:95
keratins, 2:667
kinesin superfamily proteins (KIFs), 2:673f, 2:680
Kyprides–Onzonis–Woese (KOW) motif, 4:169–170
LacI protein, 2:697f, 2:698
LexA regulatory protein, 2:707–708, 2:707f
metabotropic glutamate (mGlu) receptors, 2:396, 2:396f
mitochondrial DNA (mtDNA), 2:7
MutY repair pathway, 2:58
neurofilament triplet proteins (NFTPs), 2:634
nuclear factor kappa B (NF-κB), 3:302
ORAI proteins, 3:362–363, 3:363f, 3:365f
parathyroid hormone (PTH)/PTH-related protein (PTHrP) receptor (PPR), 3:386–387
phosphatidic acid phosphatases (PAPs), 1:661f, 1:662
phosphodiesterases (PDEs), 1:581, 1:581f
photosystem I (PSI) complex, 3:503–505, 3:504f
plasma membrane calcium (Ca²⁺)-ATPase pump, 3:525–526, 3:526f
plasma membrane proton (H⁺)-ATPases, 3:688–689, 3:690–691
platelet-activating factor receptor (PAFR), 3:534–535
platelet-derived growth factor (PDGF), 3:538, 3:539f
plectin, 2:625, 2:626f
polymerase α-primase (Pol α), 2:241–242
polymerase δ (Pol δ), 2:246f
polymerase η (Pol η), 4:573–574, 4:573f
prion domains, 3:566–567, 3:568f
protein kinase B (PKB), 3:463–464
PsaC protein, 2:340, 2:340f
P-type adenosine triphosphatase (ATPase), 4:330–331, 4:330f
Ras proteins, 4:12, 4:13f
replication focus targeting sequence (RFTS), 3:8–9, 3:8f
retinoblastoma protein (pRB), 4:105
RNA polymerase II (Pol II)
characteristics, 2:365
initiation phase, 2:366, 3:63–64, 3:63f
3'-end maturation, 3:189–191, 3:190t
transcription cycle, 2:365, 3:63–64, 3:63f
transcription termination, 2:367, 3:63–64, 3:63f
RNA polymerase domain structures, 4:175–176, 4:175t, 4:177f
ryanodine receptors (RyRs), 4:196
serine recombinases, 1:558–559
sliding clamps, 2:114–115, 2:115f, 3:576, 4:237–238
small guanosine phosphatases (GTPases), 4:242, 4:244f
store-operated Ca²⁺ entry (SOCE), 3:364, 3:365f
stromal-interacting molecule (STIM) protein family, 4:315, 4:316f
T7 RNA polymerase
polymerase sequence similarities, 4:355–357, 4:356f, 4:357f
structural characteristics, 4:355, 4:356f
structure–function relationship, 4:355, 4:357f
tubulin, 4:450–451, 4:451f
type IB DNA topoisomerase, 2:160, 2:160f
type I DNA restriction and modification enzymes, 2:72–73
type I DNA topoisomerase, 2:158, 2:159f
type XVII collagen, 1:641
tyrosine recombinases, 1:556–557, 1:557f
uncoupling proteins (UCPs), 4:483–484, 4:485f
vascular endothelial growth factor receptors (VEGFRs), 4:510, 4:510f
voltage-gated sodium channels, 4:564, 4:565f
xeroderma pigmentosum (XP), 4:574
C-terminal fragments (CTFs), 4:202
ferredoxin (Fd)
ferredoxin-NADP⁺ reductase (FNR)
structure, 2:299, 2:300f
ferredoxin structure, 2:296, 2:297f
iron–sulfur clusters (FeS), 4:95
gelsolin, 1:277–278, 1:277f
G_i family functions, 2:380
glycosylphosphatidylinositol (GPI)-anchored proteins, 2:476–477
histone H1, 1:531
integrin signaling, 2:608
lipid-modified proteins
glycosylphosphatidyl inositol (GPI) anchors, 2:737t, 2:738
palmitoylation, 2:736–737
prenylation, 2:737–738, 2:737t, 3:608–609, 3:608f
major sperm protein (MSP), 3:1–2
microtubule-associated proteins (MAPs), 3:104–105, 3:105f
MutL pathway, 2:55–56
MutS pathway, 2:55
neurofilament triplet proteins (NFTPs), 2:634
nitric oxide synthase (NOS)
calmodulin (CaM)–nitric oxide synthase (NOS) interactions, 1:342–343
C-terminal extension, 1:344–346, 1:344t, 1:345f
olfactory receptors, 3:345–346
peptide amidation, 3:396
pheromone receptors, 3:441–442
presenilins, 4:202

C-terminals (*Continued*)

- proteasome activating nucleotidase (PAN), 3:599
- protein N-myristoylation, 3:642–643, 3:643f
- protein tyrosine phosphatase (PTP), 3:651f
- respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:88
- retinoic acid receptors, 4:109
- Smad signaling, 1:232–233
- sodium ion/calcium ion ($\text{Na}^+/\text{Ca}^{2+}$) exchanger, 3:45f, 3:46–47
- Src (sarcoma) family kinases, 4:295–297, 4:296f
- tachykinins, 4:360–365
 - basic concepts, 4:360
 - hemokinin 1, 4:361, 4:361f
 - neurokinin-1 receptor (NK1R)
 - affective disorders, 4:364–365
 - carboxyl tail, 4:363–364, 4:363f
 - depression, 4:364–365
 - disease connections, 4:364–365, 4:364f
 - emesis, 4:364
 - full-length NK1R, 4:361–363, 4:362f
 - functional role, 4:361, 4:362f
 - glioblastomas, 4:365
 - truncated NK1R, 4:363
 - neurokinin A, 4:360–361, 4:361f, 4:364f
 - neurokinin B, 4:361, 4:361f, 4:364f
 - substance P, 4:360, 4:361f, 4:364f
 - toll-like receptors (TLRs), 4:397–398
 - ubiquitin interaction motif (UIM), 3:598
 - vasopressin/oxytocin family, 4:517–519
- Ctf4 coupling factor, 1:402t, 1:403f
- CTXφ genome, 1:558, 1:558f
- C-type lectin receptors (CLRs), 4:396
- C-type natriuretic peptide (CNP), 3:226, 3:226f
- cucumbers, 4:533t
- cullins, 4:470
- curcumin, 3:305, 3:305t
- currants, 4:533t
- Cushing's syndrome, 1:584
- CXXC domain, 1:523, 1:523f, 3:8–9, 3:8f
- cyan fluorescent proteins (CFPs), 2:751–752
- cyanide, 1:143t, 2:534
- cyanobacteria
 - δ-aminolevulinic acid (ALA) formation, 3:544
 - electron transport pathways, 4:81–82, 4:82f
 - ferredoxin-NADP⁺ reductase (FNR), 2:299
 - iron–sulfur (FeS) protein biogenesis, 2:663
 - light harvesting complexes, 2:726
 - membrane-associated energy transduction
 - anoxygenic photosynthesis, 2:209, 2:335, 3:32
 - antenna complexes, 3:32, 3:484, 3:484f
 - evolutionary development, 3:34
 - general discussion, 3:32
 - oxygenic photosynthesis, 3:32, 4:81
 - nonphotochemical quenching (NPQ), 3:486
 - origins, 1:506
 - photosynthesis, 3:495
 - photosystem I (PSI) complex, 3:507
 - plastocyanin (PC), 3:530
 - PsaC protein, 2:339f, 2:340, 2:340f, 2:341f
 - reaction centers, 2:335, 2:336–337, 2:337f, 3:492
 - redox cofactor arrangement, 2:337f
 - respiratory terminal oxidases (RTOs), 4:81–82
 - ribozyme distribution, 4:146f
 - cyanovirin-N, 2:702t
 - Cyc8p protein, 3:566
 - cyclase-associated protein (CAP), 1:24, 3:438–439
 - cyclic adenosine diphosphate (ADP)-ribose (cADPr)
 - calcium ion (Ca^{2+}) signaling, 1:331–336
 - basic concepts, 1:331–332
 - calcium ion (Ca^{2+}) stores, 1:332
 - fertilization-induced signaling, 1:333–334, 1:333f
 - research summary, 1:336
 - sea urchin egg fertilization, 1:333–334, 1:333f
 - signaling mechanisms, 1:332–333
 - starfish egg fertilization, 1:334–335, 1:334f, 1:335f
 - invertebrates, 2:637–638
 - metabolic pathways, 2:638–639
- cyclic adenosine monophosphate (cAMP)
 - adenyllyl cyclases (ACs), 1:42–46
 - calcium regulation, 1:45, 1:45f
 - characteristics and functional role, 1:567
 - functional role, 1:42
 - gene families, 1:42–44, 1:43f, 1:43t
 - glutamate receptors, 2:714–716, 2:715t
 - G proteins
 - G_i family, 2:377, 2:378f, 4:225t
 - G_sα, 2:517–518
 - regulation mechanisms, 1:44f, 1:45
 - intracellular second messenger systems, 1:42, 1:43f
 - physiological functions, 1:45–46, 1:46t
 - plasma membrane components, 1:43f
 - research background, 1:42
 - serotonin receptors, 4:224–225, 4:225t
 - somatostatin (sst) receptors, 4:265
- adipocyte differentiation processes, 1:56, 2:259, 2:259f
- binding domains, 2:697f
- calcium ion (Ca^{2+}) signaling system, 2:719–720, 2:720f, 3:294–295
- carboxyl methylation, 3:608
- cardiac excitation–contraction (E–C) coupling, 1:381
- cell signaling, 3:407
- characteristics and functional role, 1:567
- chemical structure, 2:696f
- class I PDEs, 1:567–568, 1:568f
- class II PDEs
 - catalytic activity, 1:572, 1:572f
 - characteristics, 1:572
 - regulatory mechanisms, 1:572, 1:572f
- CREB (cyclic adenosine monophosphate (cAMP) responsive binding protein)
 - calcium ion (Ca^{2+}) signaling system, 3:294–295
 - coactivator CREB-binding protein (CBP), 3:294–295
 - CREB shut-off pathway, 3:295, 3:295f
- cyclic adenosine monophosphate (cAMP)
 - receptors (cARs), 1:562–566
 - affinities, 1:562–563, 1:563f
 - desensitization mechanisms, 1:565
 - expression and developmental stages, 1:562–563, 1:563f
 - extracellular levels, 1:562–563, 1:563f
 - gradient sensing, 1:565–566
 - molecular identification, 1:562–563
 - schematic diagram, 1:563f
 - signaling pathways
 - characteristics, 1:563–565
 - G-protein-dependent pathways, 1:563–564, 1:564f
 - G-protein-independent pathways, 1:564–565, 1:564f
 - signaling mechanisms, 1:565
- cyclic adenosine monophosphate (cAMP) response element (CRE), 3:293, 3:293f
- cyclic nucleotide-dependent protein kinases, 1:574
- cyclic nucleotide-regulated cation channels, 1:586–588
 - background information, 1:586
 - structural and physical characteristics, 1:586, 1:587f
- DARPP-32, 4:222f, 4:223
- Dictyostelium discoideum*, 1:562–566
- distribution, 1:567
- dopamine receptors, 2:169, 2:172
- GABA_A receptor, 2:351–352
- gene expression regulation, 1:311
- glycogen synthase kinase-3 (GSK3), 2:434, 2:435f
- glycoprotein hormone receptors, 4:388–390
- G-protein-coupled receptors (GPCRs), 4:224–225, 4:225f
- gradient sensing, 1:565–566
- G_s family proteins, 2:517–518
- insect biochemistry, 2:592–593
- insulin-secreting cells (β-cells), 2:597f
- liver PK (LPK) *in vitro* regulation, 3:719
- macroautophagy regulation, 1:150t
- molecular structure, 1:568f, 1:575f
- multiple opioid receptors, 3:357
- myocardial ischemia, 3:111
- olfactory receptors, 3:349, 3:349f
- Perilipin 1 (Plin1), 1:47–48
- phosphodiesterases (PDEs), 1:580–585
 - characteristics, 1:580
 - clinical applications, 1:584
 - family-specific inhibitors, 1:584
 - G_i family functions, 2:379
 - inherited disorders, 1:584
 - molecular diversity, 1:580–581
 - research summary, 1:584
 - signaling pathways, 1:574, 1:575f
 - structure and functions
 - catalytic domains, 1:581, 1:581f
 - homeostatic regulation, 1:582–583
 - intracellular effectors, 1:583

- regulatory domains, 1:581–582, 1:581f
 signal integrators, 1:583–584
 spatially organized signalling networks and microdomains, 1:582
 phosphotransferase system (PTS), 2:387–388
 PKA-mediated lipolysis, 1:49–51, 1:50f, 4:303–304
 protein N-myristoylation, 3:641–642
 Ras proteins, 4:14–15
 research summary, 1:572
 serotonin receptors, 4:224–225, 4:225f
 signaling pathways, 1:575f
 somatostatin (sst) receptors, 4:265
 starvation effects, 4:303–304
 structural domains, 1:574–575, 1:576f
 cyclic guanosine monophosphate (cGMP) characteristics and functional role, 1:567
 chemoattractant signaling, 2:237f, 2:238
 class I PDEs
 catalytic domains, 1:568–569, 1:568f
 characteristics, 1:567–568, 1:580
 cyclic guanosine monophosphate (cGMP)-hydrolyzing PDEs, 1:569–570
 cyclic nucleotide specificity, 1:569
 domain structures, 1:568f
 dual-specificity PDEs, 1:570–571
 inhibitors, 1:569
 PDE1, 1:570–571, 1:581–582
 PDE2, 1:571, 1:581–582, 1:583–584
 PDE3, 1:571, 1:581–582, 1:583
 PDE5, 1:570, 1:581–582, 3:27
 PDE6, 1:570, 1:581–582
 PDE9, 1:570
 PDE10, 1:571, 1:581–582
 PDE11, 1:571–572, 1:581–582
 class II PDEs
 catalytic activity, 1:572, 1:572f
 characteristics, 1:572
 regulatory mechanisms, 1:572, 1:572f
 cyclic guanosine monophosphate (cGMP)-hydrolyzing PDEs, 1:569–570
 cyclic nucleotide-dependent protein kinases, 1:574
 cyclic nucleotide-regulated cation channels, 1:586–588
 background information, 1:586
 structural and physical characteristics, 1:586, 1:587f
 distribution, 1:567
 inositol trisphosphate (IP₃), 2:586–587
 ion channel proteins, 2:650
 molecular structure, 1:568f, 1:575f
 myocardial ischemia, 3:111
 natriuretic peptides, 3:226f
 nitric oxide synthase (NOS), 3:254f
 phosphodiesterases (PDEs), 1:580–585
 characteristics, 1:580
 clinical applications, 1:584
 family-specific inhibitors, 1:584
 G_i family functions, 2:379
 inherited disorders, 1:584
 molecular diversity, 1:580–581
 research summary, 1:584
 second messenger pathways, 1:43f
 signaling pathways, 1:574, 1:575f
 structure and functions
 catalytic domains, 1:581, 1:581f
 homeostatic regulation, 1:582–583
 intracellular effectors, 1:583
 regulatory domains, 1:581–582, 1:581f
 signal integrators, 1:583–584
 spatially organized signalling networks and microdomains, 1:582
 phototransduction mechanisms, 3:489–490, 3:489f
 platelet-activating factor receptor (PAFR), 3:536, 3:536f
 research summary, 1:572
 second messenger pathways, 1:43f
 signaling pathways, 1:568f, 1:575f
 structural domains, 1:574–575, 1:576f
 cyclic nucleotide-dependent protein kinases, 1:574–579
 binding specificity and affinity, 1:577–578
 cAMP-dependent protein kinase, 1:575–577, 1:576f
 cGMP-dependent protein kinase, 1:577
 cyclic adenosine monophosphate (cAMP), 1:574, 1:575f
 cyclic guanosine monophosphate (cGMP), 1:574, 1:575f
 feed-forward control, 1:578–579, 1:578f
 intracellular receptors, 1:574, 1:575f
 negative feedback, 1:578–579, 1:578f
 structural domains, 1:574–575, 1:576f
 substrate specificity, 1:578
 cyclin-dependent kinases (CDKs)
 cell-cycle-dependent activity, 2:242
 cell growth regulation, 1:408
 DNA damage checkpoints proteins, 1:410–416
 activation mechanisms, 1:410–412, 1:412f
 apoptosis, 1:415
 background information, 1:410
 component proteins, 1:411t
 disease connections, 1:415–416
 G1/S checkpoint, 1:411, 1:412f, 1:415–416
 G2/M checkpoint, 1:412, 1:412f, 1:415–416
 Intra-S checkpoint, 1:411, 1:412f, 1:415–416
 phases, 1:411f
 research summary, 1:416
 senescence, 1:415
 signal transduction
 ataxia telangiectasia-mutated and Rad3-related (ATR) pathway, 1:411t, 1:413f, 1:414
 ataxia telangiectasia mutated (ATM) kinase, 1:411, 1:411t, 1:412f, 1:413–414, 1:413f, 1:415–416
 BRCA1 C-terminal (BRCT) domains, 1:411t, 1:412f, 1:414
 Chk1 and Chk2 effector kinases, 1:411t, 1:412f, 1:414
 chromatin modifications, 1:413f, 1:414
 forkhead-associated (FHA) binding domains, 1:412f, 1:414
 lesion sensors, 1:412, 1:413f
 phosphoinositide 3-kinase-like kinase (PIKK) family, 1:412–413, 1:413f
 phospho-peptide-binding domains, 1:412f, 1:414
 spatiotemporal organization, 1:414–415
 eukaryotic DNA replication, 2:108, 2:112
 functional role, 1:403
 G₀ phase and G₁ phase cell-cycle regulation, 1:406–407, 1:407f
 retinoblastoma protein (pRB), 1:406–407, 1:407f, 4:104
 S phase initiation, 1:401–402, 1:402t, 1:403f, 1:404f, 1:406–409
 stem cell proliferation, 1:408
 cycloartenol
 brassinosteroid biosynthesis, 1:251
 chemical structure, 1:251f
 cyclobutane pyrimidine dimers (CPDs)
 nucleotide excision repair (NER), 3:338, 3:339f, 3:341, 4:402, 4:404–405, 4:404t
 ultraviolet (UV) damage, 4:572, 4:573f
 cyclohexamide, 1:150t
 cyclohexanone monooxygenase, 2:311–312
 cyclooxygenases (COX)
 active sites, 3:581–582
 biological function, 3:373
 biosynthetic pathways, 2:189
 biosynthetic relationships, 3:582f
 dioxygenases, 3:371–372
 nuclear factor kappa B (NF-κB), 3:304t
 cyclophilin (CyP) D, 3:110
 cyclopiazonic acid, 2:229
 CyP-D protein, 3:162–163
Cyphomandra betacea, 3:21f
 cystatin C, 1:101t
 cysteine
 aquaporin-1 (AQP1), 1:120–121
 chemokine receptors, 1:480
 cysteine proteases, 1:589–592
 action mechanisms, 1:589, 1:590f
 calpains, 1:353–361
 characteristics, 1:353, 1:354f
 historical research, 1:353–354
 human calpain homologs, 1:353–354, 1:355t, 1:357f
 m-calpains, 1:353–354, 1:357f, 1:358f
 μ-calpains, 1:353–354, 1:357f
 nomenclature, 1:353–354
 research summary, 1:360–361
 structural characteristics, 1:354f, 1:357f
 structure and functions, 1:354–359, 1:357f, 1:358f
 super-family characteristics, 1:354f, 1:359–360
 control processes
 activation mechanisms, 1:589–590

- cysteine (*Continued*)
 inhibitors, 1:590
 molecular structure, 1:590f
 pathology
 allergenic proteinases, 1:592
 general discussion, 1:592
 parasitic proteinases, 1:592
 phylogenetic relationships, 1:589
 physiological functions
 antigen-li processing, 1:591
 apoptosis, 1:591–592
 bone turnover, 1:591
 lysosomal proteolysis, 1:590–591
 thyroglobulin processing, 1:591
 cysteine-rich domain (CRD), 2:396
 cytochrome *c*, 1:600–601
 disulfide bond formation, 1:664–667, 3:628
 basic concepts, 1:664, 1:665f
 catalytic reactions
 background information, 1:665
 protein disulfide isomerase (PDI), 1:665, 1:665f
 thioredoxin family, 1:665
 cellular environments
 extracellular proteins, 1:666, 1:667f
 regulatory mechanisms, 1:666
 thiol/disulfide redox environments, 1:666
 isomerization processes, 1:664, 1:665f
 molecular structure, 1:664, 1:665f
 oxidative protein folding
 basic principles, 1:664
 formation mechanisms, 1:664, 1:666f
 general discussion, 1:664
 ferredoxin structure, 2:296, 2:297f
 O-fucose, 2:454f, 2:455f
 O-glucose, 2:454f
 glycation processes, 2:406
 G_q family proteins, 2:499
 metalloproteases, 3:87
 N-terminal active sites, 3:596, 3:596f
 phage libraries, 3:434t
 Ras proteins, 4:12, 4:13f
 respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:344–345, 4:345f
 ryanodine receptors (RyRs), 4:197
 selenocysteine biosynthesis, 4:210–212, 4:211f
 tricarboxylic acid cycle, 4:439f
 tumor necrosis factor receptors (TNFRs), 4:454
 vasopressin/oxytocin family, 4:516t
 zinc-finger domains, 4:573–574, 4:573f, 4:575, 4:576f
 cystic fibrosis, 3:630, 4:476
 cytidine
 cytidine diphosphate (CDP)–choline pathway, 3:479f, 3:480
 cytidine diphosphate (CDP)–diacylglycerol (DAG) pathway, 3:478–480, 3:479f
 cytidine diphosphate (CDP)–ethanolamine pathway, 3:479f, 3:480
 cytidine nucleotide formation, 3:715–716, 3:715f
 cytidine triphosphate (CTP), 3:715f
 phage libraries, 3:432–433, 3:433t, 3:434t
 synthetic pathways, 3:713f
 cytidine-to-uridine (C-to-U) RNA editing, 4:155f, 4:156, 4:157f
 cytochalasins, 1:16–17, 1:16t, 1:150t
 cytochromes
 bacteriopheophytins, 3:696–697, 3:696f, 3:701f
 characteristics, 1:599
 cytochrome *a*, 1:596f, 2:537
 cytochrome *aa3* oxidase, 1:603f, 1:604t, 4:79, 4:80f
 cytochrome *b*, 1:142f, 1:596f, 2:508–509, 2:508f, 2:537, 3:32, 3:160
 cytochrome *bc*₁ complex, 1:593–598
 action mechanisms
 inhibitors, 1:598
 proton conduction pathways, 1:597–598, 1:597f
 protonmotive Q cycle, 1:596–597, 1:597f, 3:663–664, 3:664f
 anoxygenic photosynthesis, 3:694f
 background information, 4:98
 basic concepts, 1:593
 biochemical free radical chain reactions, 3:170f, 3:172f, 3:173f
 cytochrome *c* interactions, 1:601, 1:601f
 electron transfer complexes, 3:695, 4:93f
 facultative phototrophic respiration, 4:79, 4:80f
 functional role, 1:593, 1:594f, 4:84
 green sulfur bacteria, 1:601f, 2:511f
 mitochondrial DNA (mtDNA), 3:133f, 3:134t
 mitochondrial ROS, 4:19f, 4:20
 oxygen respiration, 3:30–31
 periplasmic electron-transport systems, 3:400, 3:401f
 PET genes, 3:308f
 photosynthesis/respiration, 3:695
 photosynthetic reaction centers, 3:693–694, 3:694f, 3:701f
 plant mitochondrial electron transport chain, 1:145, 1:146f
 protein subunits, 1:593–594
 protonmotive electron transfer chain, 1:470–471, 1:470f
 quinone pool functions, 3:695f
 respiratory electron transport chain, 3:661–662, 3:662f, 3:663–664, 3:664f, 3:694–695, 3:694f, 4:79, 4:80f
 schematic diagram, 4:84f, 4:99f
 spectroscopic properties, 1:594–596, 1:595f
 structural characteristics, 1:596, 1:596f
 thermodynamic properties, 1:595–596, 1:595f
 cytochrome *b*/f complex
 chloroplast signaling
 electron flow pathways, 1:511–512, 1:512f
 state transitions, 1:513, 1:513f
 thylakoid protein kinase network, 1:514f
 cytochrome *b₆* complex, 4:335–340
 amino acid subunit sequences, 4:338
 C2 symmetry, 4:336
 chlorophyll and carotenoids, 4:337f, 4:338
 crystal structures, 4:335–338, 4:337f
 electron transport chain, 4:335, 4:336f
 evolutionary development, 4:339–340
 functional role, 4:335, 4:336f, 4:338–339, 4:339f
 heme *b* rotational isomers, 4:337–338, 4:338f, 4:339f, 4:339t
 heme *c_v*, 4:337, 4:337f, 4:338f, 4:339t
 inhibitors, 4:339
 midpoint oxidation–reduction potentials, 4:338t
 posttranslational modifications, 4:338
 quinone-exchange cavity, 4:336–337, 4:337f
 rate-limiting steps, 4:339
 Rieske protein, 2:335, 4:335–338, 4:337f, 4:338t, 4:339t
 light harvesting proteins, 2:179f
 photosynthetic processes, 3:32, 3:493–494, 3:493f, 3:494f
 plastocyanin (PC), 3:493–494, 3:494f, 3:530
 cytochrome *c*, 1:599–601
 amino acid sequence studies, 1:600
 anoxygenic photosynthesis, 3:693–694, 3:694f
 apoptosis, 1:601
 bacterial cytochrome *c* oxidases, 1:602–606
 characteristics, 1:602
 chemical processes, 1:602, 1:603f
 experimental research, 1:606
 green bacteria, 2:508–509, 2:508f
 heme–copper oxidases, 1:603f, 1:604t, 1:606
 oxygen binding and reduction, 1:605, 1:605f
 proton pumping, 1:605–606
 proton-transfer pathways, 1:603f, 1:604–605, 1:605f
 reaction cycle, 1:605f
 structural characteristics, 1:602–604, 1:603f, 1:604t
 biochemical free radical chain reactions, 3:170f, 3:172f, 3:173f
 cellular respiration, 3:407, 3:408f
 characteristics, 1:599, 2:537
 cytochrome *bc*₁ complex, 1:593–598
 anoxygenic photosynthesis, 3:694f
 basic concepts, 1:593
 biochemical free radical chain reactions, 3:170f, 3:172f, 3:173f
 cytochrome *c* interactions, 1:601, 1:601f
 electron transfer complexes, 3:695, 4:93f
 facultative phototrophic respiration, 4:79, 4:80f

- functional role, 1:593, 1:594f, 4:84
green sulfur bacteria, 1:601f, 2:511–512, 2:511f
inhibitors, 1:598
mitochondrial DNA (mtDNA), 3:133f, 3:134t
mitochondrial ROS, 4:19f, 4:20
oxygen respiration, 3:30–31
photosynthesis/respiration, 3:695
photosynthetic reaction centers, 3:693–694, 3:694f, 3:701f
protein subunits, 1:593–594
proton conduction pathways, 1:597–598, 1:597f
protonmotive Q cycle, 1:596–597, 1:597f, 3:663–664, 3:664f
quinone pool functions, 3:695f
respiratory chain complex III (ubiquinol cytochrome *c* reductase), 1:601, 1:601f
respiratory electron transport chain, 3:661–662, 3:662f, 3:663–664, 3:664f, 3:694–695, 3:694f, 4:79, 4:80f
schematic diagram, 4:84f
spectroscopic properties, 1:594–596, 1:595f
structural characteristics, 1:596, 1:596f
thermodynamic properties, 1:595–596, 1:595f
cytochrome *c*-554/555, 2:511, 2:511f
cytochrome *c₆*, 3:504–505
cytochrome *c₂*, 2:511, 2:511f
evolutionary development, 1:600–601
ferricytochrome *c* oxidoreductase, 3:663–664, 3:664f
flavocytochrome *c*, 2:511
green sulfur bacteria
characteristics, 2:510–512
cytochrome *bc₁* complex, 2:511–512, 2:511f
cytochrome *c*-554/555, 2:511, 2:511f
cytochrome *c₂*, 2:511, 2:511f
flavocytochrome *c*, 2:511
newly discovered cytochromes, 2:512
SoxAX, 2:511, 2:511f
heme prosthetic group attachment, 1:599–600
hemoglobin (Hb) studies, 4:277
hydroquinone-cytochrome *c* oxidoreductase (QCR), 4:79, 4:80f
mitochondrial inner membrane, 3:148
mitochondrial ROS, 1:212f
multiheme *c*-type cytochromes, 3:404–405, 3:404f
photosynthetic reaction centers, 3:693–694, 3:694f, 3:701f, 3:706–707, 3:706f
photosystem I (PSI) complex, 3:504–505
phylogenetic relationships, 1:600–601, 1:600f
plant mitochondrial electron transport chain, 1:141–142, 1:142f, 1:145, 1:146f
quinone pool functions, 3:695f
research background, 1:599
respiratory chain interactions, 1:601, 1:601f
respiratory electron transport chain, 3:661–662, 3:662f, 3:663–664, 3:664f, 3:694–695, 3:694f, 4:79, 4:80f
SoxAX, 2:511, 2:511f
spectrophotometric assays, 4:276–277, 4:277f
three-dimensional structure, 1:601
cytochrome *cbb3* oxidase, 3:694f, 3:697, 3:697t, 4:79, 4:80f
cytochrome *c* oxidase
bacterial cytochrome *c* oxidases, 1:602–606
characteristics, 1:602
chemical processes, 1:602, 1:603f
experimental research, 1:606
heme–copper oxidases, 1:603f, 1:604t, 1:606
oxygen binding and reduction, 1:605, 1:605f
proton pumping, 1:605–606
proton-transfer pathways, 1:603f, 1:604–605, 1:605f
reaction cycle, 1:605f
structural characteristics, 1:602–604, 1:603f, 1:604t
bacterial reductive reactions, 1:487t, 1:488f
hypoxia inducible factor (HIF), 4:67
respiratory chain complex IV (cytochrome *c* oxidase), 4:98–103
biochemical free radical chain reactions, 3:170f, 3:172f, 3:173f
catalytic cycle, 4:101–103, 4:102f
electron transfer complexes, 4:93f
functional role, 4:84
PET genes, 3:308f
plant mitochondrial electron transport chain, 1:145, 1:146f
prosthetic groups, 4:98–100, 4:99f
protein composition, 4:100
protein structure, 4:100, 4:101f
protonmotive electron transfer chain, 1:470f
proton pumping, 3:664–665, 4:101–103, 4:102f
proton transfer pathways, 4:101, 4:101f
respiratory electron transport chain, 3:661–662, 3:662f
schematic diagram, 4:99f
structural characteristics, 4:100–101
cytochrome *f*
complex formation, 3:531, 3:531f
cytochrome *b₆f* complex, 4:338t, 4:339t
domain structures, 3:530–531
oxygenic photosynthesis, 3:32
plastocyanin (PC), 3:530
cytochrome oxidase
bacterial cytochrome *c* oxidases, 1:602–606
characteristics, 1:602
chemical processes, 1:602, 1:603f
experimental research, 1:606
green bacteria, 2:508–509, 2:508f
heme–copper oxidases, 1:603f, 1:604t, 1:606
oxygen binding and reduction, 1:605, 1:605f
proton pumping, 1:605–606
proton-transfer pathways, 1:603f, 1:604–605, 1:605f
reaction cycle, 1:605f
structural characteristics, 1:602–604, 1:603f, 1:604t
bacterial reductive reactions, 1:488f
cytochrome *c* oxidase
bacterial cytochrome *c* oxidases, 1:602–606, 2:508–509, 2:508f
bacterial reductive reactions, 1:487t, 1:488f
chemical processes, 1:602, 1:603f
experimental research, 1:606
heme–copper oxidases, 1:603f, 1:604t, 1:606
hypoxia inducible factor (HIF), 4:67
oxygen binding and reduction, 1:605, 1:605f
plant mitochondrial electron transport chain, 1:141–142, 1:145, 1:146f
proton pumping, 1:605–606
proton-transfer pathways, 1:603f, 1:604–605, 1:605f
reaction cycle, 1:605f
respiratory chain complex IV (cytochrome *c* oxidase), 4:98–103
structural characteristics, 1:602–604, 1:603f, 1:604t
mitochondrial DNA (mtDNA), 3:133f, 3:134t
photochemical action spectra, 4:275f
plant mitochondrial electron transport chain, 1:142f
ruby laser technology, 4:277f
cytochrome P-450 (CYP), 1:607–612
action mechanisms, 2:537, 2:537f
background information, 1:607
bile salts, 1:168–169, 1:168f, 1:169f
biotechnological applications, 1:611–612
brassinosteroid biosynthesis, 1:251, 1:252f
classifications and characteristics, 1:607–608, 1:608f
CYP4F2, 4:549, 4:549f
cytochrome P-450 monooxygenase (CYP7A1), 1:168–169, 1:168f, 1:169f, 3:374
diversity, 1:611
drug metabolism, 1:611
as drug targets, 1:611
functional role, 2:536–537
general discussion, 2:536–537
hereditary spastic paraplegias (HSPs), 4:271
insect biochemistry, 2:594, 2:594f
leukotriene metabolism, 3:584

- cytochromes (*Continued*)
- mitochondrial-type ferredoxins, 2:297
 - occurrences, 2:536
 - physiological functions, 1:611
 - reaction cycle, 1:609–610, 1:610f
 - reactive oxygen species (ROS), 2:77
 - regulation mechanisms, 1:610–611
 - research areas, 1:608f
 - retinoid metabolic pathways, 4:527f
 - spectral properties, 1:608f
 - structural organization, 1:608–609, 1:609f
 - vitamin A homeostasis control, 4:529
 - vitamin E metabolism, 4:549, 4:549f
 - electron transfer complexes, 3:695–696
 - facultative phototrophic respiration, 4:79, 4:80f
 - functional role, 2:537
 - green bacteria, 2:510–512
 - cytochrome *c*
 - characteristics, 2:510–512
 - cytochrome *bc₁* complex, 2:511–512, 2:511f
 - cytochrome *c*-554/555, 2:511, 2:511f
 - cytochrome *c_v*, 2:511, 2:511f
 - flavocytochrome *c*, 2:511
 - newly discovered cytochromes, 2:512
 - SoxAX, 2:511, 2:511f
 - filamentous anoxygenic phototropic (FAP) bacteria
 - alternative complex III, 2:512
 - characteristics, 2:510
 - cytochrome *c*-554, 2:512
 - uncharacterized cytochromes, 2:512
 - general discussion, 2:510
 - green sulfur bacteria, 2:510, 2:511f
 - structural characteristics, 2:511f
 - respiratory chain complexes, 4:83
 - succinate:quinone oxidoreductases (SQORs), 4:79, 4:80f
 - cytohesins, 1:135
 - cytokeratins, 2:625–627
 - cytokines, 1:613–621
 - adiponectin, 3:74
 - amino acid metabolism, 1:95–96
 - appetite regulation, 4:303
 - background information, 1:613
 - calcineurin-mediated signal transduction, 1:339–340, 1:340f
 - cAMP responsive element-binding protein 3 (CREB3), 4:52
 - chemokines, 1:619, 1:619f
 - chronic liver injury repairs, 1:194f
 - insulin resistance, 2:606–607
 - insulin signal regulation, 2:606
 - interferon receptors, 2:619–623
 - characteristics and functional role, 2:619
 - interferon- α /b receptors
 - antiviral effects, 2:623
 - gene expression, 2:621
 - janus tyrosine kinase (JAK)/signal transducer and activator of transcription (STAT) signaling, 2:622–623, 2:622f
 - ligand binding, 2:622
 - signal transduction, 2:622–623, 2:622f
 - structural characteristics, 2:621–622
 - interferon- γ receptors
 - antibacterial effects, 2:621
 - gene expression, 2:619
 - janus tyrosine kinase (JAK)/signal transducer and activator of transcription (STAT) signaling, 2:620–621, 2:620f
 - ligand binding, 2:620
 - signal transduction, 2:620–621, 2:620f
 - structural characteristics, 2:619–620
 - interleukins
 - B1 receptor regulation, 1:241
 - calcineurin-mediated signal transduction, 1:339–340, 1:340f
 - cAMP responsive element-binding protein 3 (CREB3), 4:52
 - c-fes* proto-oncogene (*Fes*), 1:452
 - general discussion, 1:613–617
 - hematopoietin receptors
 - expression regulation, 2:529
 - signal transduction, 2:529
 - structural characteristics, 2:526–527, 2:527f
 - interleukin-1 β -converting enzyme (ICE), 1:591–592
 - interleukin-1 (IL-1) family, 1:613–614, 1:614f
 - interleukin-1 receptor-associated kinase (IRAK) family protein kinases, 4:398, 4:399f, 4:400f
 - interleukin-2 (IL-2) family, 2:527f, 2:530
 - interleukin-3 (IL-3) family
 - expression regulation, 2:529
 - signal transduction, 2:529
 - structural characteristics, 2:526–527, 2:527f
 - interleukin-5 (IL-5) family
 - expression regulation, 2:529
 - signal transduction, 2:529
 - structural characteristics, 2:526–527
 - interleukin-6 (IL-6) family, 1:614–615, 1:614f, 2:526–527, 2:527f, 2:529–530
 - interleukin-10 (IL-10) family, 1:615, 1:615f
 - interleukin-17 (IL-17) family, 1:615–616, 1:616f
 - nuclear factor kappa B (NF- κ B), 3:303t, 3:304t
 - mucins, 3:206–208, 3:206f
 - neurogenesis, 1:198–199
 - nuclear factor kappa B (NF- κ B), 3:303t
 - suppressor of cytokine signaling-3 (SOCS-3), 2:260, 2:260f, 2:526, 2:528f
 - toll-like receptors (TLRs), 4:396–397, 4:397t
 - transforming growth factor beta (TGF- β), 1:620, 1:620f
 - tumor necrosis factor (TNF), 1:617–619, 1:618f
 - type I cytokines, 1:616, 1:617f
 - type II cytokines, 1:616–617, 1:617f, 1:618f, 2:619
 - cytokinesis, 1:622–626
 - basic concepts, 1:622–626, 1:623f, 4:214, 4:215f
 - contractile networks, 1:624–625, 1:625f
 - mechanical processes, 1:625–626
 - mitosis, 3:182f, 3:183f, 3:184
 - regulation mechanisms
 - actin cytoskeleton, 1:623–624, 1:623f
 - bipolar thick filaments (BTfFs), 1:623–624
 - mitotic spindle, 1:623–624
 - myosin, 1:623–624
 - protein interactions, 1:622–623
 - research summary, 1:626
 - septins, 4:214–218
 - behavior and function
 - Ashbya gossypii*, 4:216f, 4:217
 - budding yeasts, 4:216–217, 4:216f
 - fission yeasts, 4:217–218
 - metazoans, 4:217f, 4:218
 - domain structures
 - filamentous complexes, 4:214–215, 4:215f
 - higher-order assemblies, 4:215–216, 4:216f, 4:217f
 - inositol phospholipids, 4:216
 - functional role, 4:214–218
 - spatiotemporal events, 1:624–625
 - species-specific cell division mechanisms, 1:626
 - cytokinins, 1:627–631
 - background information, 1:627
 - biosynthesis, 1:627–628
 - chemical structure, 1:627, 1:628f
 - functional roles
 - biotechnological applications, 1:631
 - cell cycle, 1:629
 - pathogenicity, 1:631
 - plant development and growth, 1:629–630
 - metabolism
 - catabolic reactions, 1:629
 - conjugations, 1:628–629
 - cytokinin oxidases/dehydrogenases (CKX), 1:629
 - interconversions, 1:628
 - signaling pathways
 - histidine kinases, 1:629, 1:630f
 - signal perception, 1:629
 - signal transduction, 1:629, 1:630f
 - two-component system (TCS) pathway, 1:629, 1:630f
 - transport mechanisms, 1:629
 - cytomegalovirus
 - nuclear factor kappa B (NF- κ B), 3:304t
 - tumor necrosis factor receptors (TNFRs), 4:459
 - cytoplasmic domains
 - cell adhesion
 - adaptor proteins, 1:257, 1:258f
 - adhesion complex turnover, 1:257–258
 - cytoskeletal linkage, 1:257
 - signaling functions, 1:258
 - structural characteristics, 1:256f

- cell–matrix interactions, 1:428–429, 2:609–610
- copper-transporting ATPases (copper pumps), 3:672f
- kinesin superfamily proteins (KIFs), 2:684f
- mucins, 3:201f, 3:202
- ORAI proteins, 3:363f
- ryanodine receptors (RyRs), 4:195
- stromal-interacting molecule (STIM) protein family
- stromal interacting molecule 1 (STIM1), 3:362, 3:362f
 - stromal interacting molecule 2 (STIM2), 3:362
- cytosine
- chemical structure, 4:77f
- cytosine methylation
- basic concepts, 1:522, 3:6, 3:7f
 - cancer cells, 1:522
 - functional role, 1:522
 - gene expression regulation, 1:522
 - genomic imprinting, 1:522
 - X-chromosome inactivation, 1:522
- DNA alkylation sites and repair mechanisms, 2:10t
- 5-methylcytosine (5mC)
- chemical structure, 4:77f
 - deoxyribonucleic acid (DNA), 3:9, 3:9f
 - functional role, 2:16, 4:76
 - mismatch repair (MMR), 1:523f
- cytoskeleton
- actin cytoskeleton
 - bipolar thick filaments (BTfFs), 1:623–624
 - composition, 1:430–431
 - contractile networks, 1:624–625, 1:625f
 - regulation mechanisms, 1:623–624, 1:623f
 - schematic diagram, 1:432f
 - small guanosine phosphatases (GTPases), 4:248 - chemotaxis, 2:238
 - composition
 - actin cytoskeleton, 1:430–431, 1:431f, 4:248
 - intermediate filaments, 1:430, 3:17, 3:233–237
 - microtubules, 1:431f, 1:434 - cytoskeletal walking model, 1:632–635
 - cargo binding, 1:634
 - coordination and gating, 1:634
 - limitations, 1:632–633
 - mechanical performance, 1:634–635, 1:635f
 - regulation mechanisms, 1:633–634
 - research summary, 1:635
 - stepping schemes, 1:632, 1:633f
- filopodia
- actin organization, 1:28–29, 1:29f
 - cell migration, 1:430–431, 1:431f, 1:432f, 1:433f, 1:434f
 - structural characteristics, 1:430–431, 1:431f
- focal adhesions, 2:320
- lamellipodia
- cell migration, 1:430–431, 1:431f, 1:432f, 1:433f, 1:434f
 - structural characteristics, 1:430–431, 1:431f
 - myosin formation, 1:430–431, 1:431f, 1:434f
 - stress fibers, 1:430–431, 1:431f
- cytosol
- AAA (ATPases associated with various cellular activities) proteins, 1:3t
 - acyl-coenzyme A synthetase
 - long-chain acyl-CoA synthetase (LCAS), 2:292–293
 - short-chain acyl-CoA synthetase (SCAS), 2:292 - calcium/calmodulin-dependent protein kinase II (CaMKII), 1:297–298, 1:298f
 - calcium-modulated proteins, 1:316–321
 - calcium coordination
 - binding kinetics, 1:316
 - pentagonal bipyramidal conformation, 1:316
 - temporal buffering, 1:316–317 - EF-hand proteins
 - calmodulin (CaM), 1:318f, 1:319t, 1:321
 - domain structures, 1:318–321, 1:318f
 - evolutionary development, 1:317–318, 1:318f
 - stromal-interacting molecule (STIM) protein family, 3:361–362, 3:361f
 - structural characteristics, 1:317, 1:317f, 1:318f
 - subfamilies, 1:318–321, 1:319t
 - functional role, 1:316
- carbohydrate-responsive element binding protein (ChREBP), 1:377f
- cardiac metabolism, 3:66f, 3:68f, 3:69f, 3:70f
- cytosolic calcium, 1:379–381, 1:380f
- cytosolic creatine kinase, 3:81–82, 3:82f
- endocytosis, 2:199
- free oligosaccharide (fOS) generation, 2:329, 2:330f
- heme biosynthesis, 2:539–540, 2:540f
- iron–sulfur (FeS) protein biogenesis, 2:663–664
- mitochondrial outer membrane, 3:161
- multidrug resistance (MDR) membrane proteins, 3:212f
- prenylation, 2:737–738, 2:737t, 3:608–609
- protein kinase C (PKC), 3:638–639, 3:639f
- thiamine diphosphate (ThDP), 1:207f
- cytosolic regulator of adenyl cyclase (CRAC), 1:563–564
- cytotoxic T lymphocytes, 4:458
- phospholipase D (PLD), 3:473
- Danon disease, 4:507t
- DAN protein family, 1:229–230
- dark fermentation, 4:73–74, 4:74f, 4:75f
- DARPP-32, 4:222f, 4:223
- Daucus carota*, 3:98
- daunorubicin, 2:167f
- DAWDLE protein, 3:99–100
- db/db* mice intercross studies, 2:256
- Dbf4-dependent kinases (DDK), 2:108, 2:112
- dcm* genes, 4:76, 4:77f
- Dcn1*, 4:468t, 4:470
- DEAD-box helicase, 4:151–152, 4:152f
- DEAD box motif, 2:136–137, 2:137f
- deafness, 3:223t
- death domain (DD), 3:243–246, 4:454, 4:455f, 4:456f
- death effector domain (DED), 1:390f
- death-inducing signaling complex (DISC), 1:390f, 4:457–458
- de Broglie equation, 2:688–689, 2:689t
- DEC1 genes, 4:67–68
- decorin, 3:91t
- deep-brain photoreceptors, 3:491
- deer, 3:575t
- degenerin (DEG)/ENaC superfamily, 4:367–368, 4:367f
- Dehalococcoides*, 3:331f
- Dehalospirillum*, 2:207f
- dehydroascorbate reductase (DHAR), 4:532
- 7-dehydrocholesterol, 1:516, 4:535
- dehydrogenation theory, 3:371
- Deinococcus*, 2:161
- Deinococcus radiodurans*
- Cho phylogenetic tree, 3:331f
 - recombination-dependent DNA replication (RDR), 4:24
 - single-strand DNA (ssDNA) translocation, 3:264, 3:264f
 - uvrC* phylogenetic tree, 3:331f
- delavirdine, 2:549
- delayed after-depolarization (DAD), 1:382–383, 1:382f
- δ-globin genes, 2:109
- Delta protein family, 2:455
- dementia, 2:173
- dementia pugilistica, 1:477t
- Den1*, 4:468t, 4:470
- dendrites, 2:713, 2:714f
- dendritic brush, 1:20
- Dengue fever, 2:591
- DENN/MADD adapter protein, 2:687
- dense fibrillar component (DFC), 3:324, 3:325f
- dentate gyrus, 1:200f
- deoxyadenosine triphosphate (dATP), 3:716
- deoxycholate, 1:168f
- deoxycytidine triphosphate (dCTP), 3:716
- deoxymannojirimycin (DMN), 2:459f
- deoxynojirimycin (DN), 2:459f
- deoxynucleoside triphosphate (dNTP), 2:96, 2:98f, 2:99f, 2:103
- deoxyribonolactone, 2:2f, 2:6–7

D

- damage-binding protein (DDB), 4:402
- Danio rerio*
- acyl-coenzyme A synthetase families, 2:292
 - base excision repair (BER), 2:5t

- deoxyribonucleic acid (DNA)
 alkylation damage, 2:9–15
 alkyltransferase-like (ATL) proteins, 2:14, 2:15f
 base excision repair (BER)
 AlkA family, 2:10t, 2:11, 2:11f
 AlkC/AlkD families, 2:10t, 2:11f, 2:12
 basic concepts, 2:10–12
 DNA alkylation sites and repair mechanisms, 2:10t
 glycolase family, 2:10–12, 2:16, 2:17f
 human AAG, 2:10t, 2:11, 2:11f, 2:12f
 MagIII family, 2:10t, 2:11, 2:11f
 oxidative damage repair, 2:79–80
 prokaryotic HLR 3-methyladenine (3mA) DNA glycolases, 2:11f, 2:12
 TAG family, 2:10–11, 2:10t, 2:11f
 3-methyladenine (3mA) DNA glycolase I, 2:10–11, 2:11f
 3-methyladenine (3mA) DNA glycolase II, 2:11, 2:11f
 3-methyladenine (3mA) DNA glycolase III, 2:11, 2:11f
 causal factors, 2:9
 damage types, 2:9, 2:10f, 2:10t
 direct damage reversal proteins
 ABH2/ABH3 family, 2:13–14, 2:13f
 Ada/N-Ada, 2:10t, 2:12, 2:12f
 AidB response, 2:14
 AlkB family, 2:13–14, 2:13f
 O⁶-alkylguanine-DNA alkyltransferase (AGT), 2:10t, 2:12–13
 alkyltransferases, 2:12–13
 C-Ada, 2:10t, 2:12–13, 2:12f
 FTO dioxygenase, 2:13f, 2:14
 general discussion, 2:12–14
 iron (Fe(II))/2OG-dependent dioxygenases, 2:13–14
 lesion bypass polymerases, 2:14
 mismatch repair, 2:14
 Ogt, 2:10t, 2:12–13, 2:12f
 general discussion, 2:9
 nucleotide excision repair (NER), 2:14, 2:15f
 research summary, 2:14–15
 base excision repair (BER), 2:1–7
 alkylation damage
 AlkA family, 2:10t, 2:11, 2:11f
 AlkC/AlkD families, 2:10t, 2:11f, 2:12
 basic concepts, 2:10–12
 DNA alkylation sites and repair mechanisms, 2:10t
 glycolase family, 2:10–12, 2:16, 2:17f
 human AAG, 2:10t, 2:11, 2:11f, 2:12f
 MagIII family, 2:10t, 2:11, 2:11f
 oxidative damage repair, 2:79–80
 prokaryotic HLR 3-methyladenine (3mA) DNA glycolases, 2:11f, 2:12
 TAG family, 2:10–11, 2:10t, 2:11f
 3-methyladenine (3mA) DNA glycolase I, 2:10–11, 2:11f
 3-methyladenine (3mA) DNA glycolase II, 2:11, 2:11f
 3-methyladenine (3mA) DNA glycolase III, 2:11, 2:11f
 apurinic/apyrimidinic (AP) site, 2:82, 2:83f
 basic concepts, 2:1
 B-DNA (antiparallel right-handed double helix), 2:142
 core enzymes
 AP endonucleases, 2:3–4, 2:4f, 2:5t
 DNA glycosylases, 2:3, 2:3t, 2:4f, 2:5t
 DNA ligases, 2:5t, 2:6, 2:30t, 2:31
 DNA polymerases, 2:4f, 2:5
 end-processing enzymes, 2:5
 5'-deoxyribose-5-phosphate lyase, 2:5, 2:6f
 flap endonuclease, 2:5–6
 protein interactions, 2:6
 disease risks, 2:7
 DNA glycosylases, 2:3, 2:3t, 2:4f, 2:5t, 2:10–12, 2:16, 2:17f, 2:57–58, 2:57f
 functional roles, 2:7
 lesion sources and characteristics, 2:1–3, 2:2f, 2:7–8
 long-patch (LP)-BER, 2:4f, 2:5, 2:6–7, 2:7–8, 2:31
 mitochondrial DNA (mtDNA), 2:7
 polymerase β (Pol β), 2:4f, 2:5, 2:82–83, 2:83f, 2:85–86, 2:85f
 polymerase δ (Pol δ), 2:5, 2:247–248
 research summary, 2:7–8
 short-patch repair, 2:31
 calcium ion (Ca²⁺) binding, 1:282
 centromere identity and function
 characteristics, 1:447–448
 holocentric centromeres, 1:448
 multisubunit centromeres, 1:447
 point centromeres, 1:447, 1:448f
 regional centromeres, 1:447, 1:448f
 Schizosaccharomyces pombe, 1:447, 1:448f
 chromatin structure, 3:314–315, 3:315f
 chromosome organization, 1:535–538
 chromatin structure, 1:535–537, 1:536f, 3:315–316, 3:315f
 mitotic chromosome, 1:537–538
 nucleosome structure, 1:535–537, 1:536f
 research summary, 1:538
 scaffold attachment regions (SARs), 1:537–538
 three-dimensional structure, 1:537
 conservative site-specific recombination, 1:555–561
 basic concepts, 1:1
 biological characteristics, 1:555–556
 integration and excision reactions, 1:555–556, 1:556f, 1:560f
 recombinases
 reaction mechanisms, 1:556
 resolvase/invertase mechanisms, 1:558–559, 1:559f
 serine recombinases, 1:558–559, 1:559f
 tyrosine recombinases, 1:556–558, 1:557f, 1:558f
 recombination control mechanisms
 outcome control, 1:6–7, 1:560f
 site polarity, 1:559
 research summary, 1:561
 damage repair
 damage checkpoints, 1:410–416
 activation mechanisms, 1:410–412, 1:412f
 apoptosis, 1:415
 ataxia telangiectasia-mutated and Rad3-related (ATR) pathway, 1:411t, 1:413f, 1:414
 ataxia telangiectasia mutated (ATM) kinase, 1:411, 1:411t, 1:412f, 1:413–414, 1:413f, 1:415–416
 background information, 1:410
 BRCA1 C-terminal (BRCT) domains, 1:411t, 1:412f, 1:414
 Chk1 and Chk2 effector kinases, 1:411t, 1:412f, 1:414
 chromatin modifications, 1:413f, 1:414
 component proteins, 1:411t
 disease connections, 1:415–416
 forkhead-associated (FHA) binding domains, 1:412f, 1:414
 G1/S checkpoint, 1:411, 1:412f, 1:415–416
 G2/M checkpoint, 1:412, 1:412f, 1:415–416
 Intra-S checkpoint, 1:411, 1:412f, 1:415–416
 lesion sensors, 1:412, 1:413f
 phases, 1:411f
 phosphoinositide 3-kinase-like kinase (PIKK) family, 1:412–413, 1:413f
 phospho-peptide-binding domains, 1:412f, 1:414
 research summary, 1:416
 senescence, 1:415
 signal transduction, 1:412–415
 spatiotemporal organization, 1:414–415
 LexA regulatory protein, 2:707f, 2:709–710
 DNA adenine methylation identification (Dam-ID)
 general discussion, 2:749
 nuclear structure and dynamics, 2:750
 DNA-binding domain
 DNA ligases, 2:33, 2:34f, 2:35f, 2:37–39, 2:37f, 2:38f
 domain Ia, 2:34f, 2:35, 2:35f
 lac operon, 2:697f
 retinoic acid receptors, 4:108
 steroid/thyroid hormone receptors, 4:309
 DNA-dependent protein kinase catalytic subunit (DNA-PK_{cs}), 3:269–271, 3:270f, 3:272f
 DNA-directed RNA polymerases, 4:355, 4:357f
 DNA glycosylases, 2:16–20
 alkyladenine DNA glycosylase (AAG), 2:16–17, 2:20
 base excision repair (BER), 2:3, 2:3t, 2:4f, 2:5t, 2:10–12, 2:16, 2:17f, 2:57–58, 2:57f
 bifunctional enzymes, 2:17f

- catalytic reaction mechanisms, 2:18–19, 2:19f
 classifications, 2:16–17, 2:17f
 functional role, 2:16
 MutS–MutY interactions, 2:58
 MutY repair pathway, 2:54, 2:57–58, 2:57f
 nucleotide flipping, 2:17–18, 2:18f
 specificity approaches, 2:19–20
 3-methyladenine (3mA) DNA glycolase I, 2:10–11, 2:11f
 3-methyladenine (3mA) DNA glycolase II, 2:11
 thymine DNA glycosylase (TDG)
 background and characteristics, 4:78
 biochemistry, 4:78
 functional role, 2:16
- DNA gyrase
 characteristics, 3:369t
 functional role, 2:165–166, 3:369
 GyrA/GyrB, 2:164–165, 2:165f, 3:369–370
 mesophilic/psychrophilic eubacteria, 2:161
 supercoils, 3:369f
 transcription and replication processes, 2:162
 type II DNA topoisomerase, 2:164–165, 2:165f
 type IIA DNA topoisomerase, 2:165–166
- DNA ligases, 2:28–32, 2:33–39
 background information, 2:28
 bacterial DNA replication, 2:115f, 2:116, 2:116t
 base excision repair (BER), 2:5t, 2:6
 cellular functions
 DNA replication, 2:31
 double-strand breaks (DSBs), 2:31–32
 excision repair, 2:31
 general discussion, 2:31–32
 mitochondrial DNA metabolism, 2:30t, 2:32
 domain structure and organization, 2:28–30, 2:29f, 2:33, 2:34f
 encoded polypeptides
 characteristics, 2:30–31
 DNA ligase I (LIG1), 2:29f, 2:30, 2:30t, 2:34f
 DNA ligase III (LIG3), 2:29f, 2:30–31, 2:30t, 2:34f
 DNA ligase IV (LIG4), 2:29f, 2:30t, 2:31, 2:34f, 3:271, 3:272f, 3:273t
 end-joining reactions, 2:34f, 2:35–37, 2:38f
 functional role, 2:33, 2:34f
 ligase D (LigD), 3:267
 mismatch repair (MMR)
 Dam methyltransferase, 2:41f
 DNA polymerase I, 2:54–56, 2:55t, 2:56f
 excision and repair synthesis, 2:56
 GO system (MutM–MutT–MutY), 2:80f
 polymerase δ (Pol δ), 2:52t
 non-homologous end joining (NHEJ), 3:266, 3:267f, 3:271, 3:272f, 3:273t
- protein clamps, 2:37–39, 2:37f
 reaction mechanisms
 covalent DNA–AMP covalent
 intermediate formation, 2:28, 2:33, 2:34f, 2:35–37, 2:36f, 2:38f
 covalent ligase–AMP covalent
 intermediate formation, 2:28, 2:29f, 2:33–35, 2:35f, 2:38f
 phosphodiester bond formation, 2:28, 2:29f
 research background, 2:28
 research summary, 2:32
- DNA microarrays
 basic concepts, 2:370–371
 data analysis, 2:371
 schematic diagram, 2:371f
- DNA mutation-caused diabetes, 1:653
- DNA photolyase, 2:312
- DNA polymerases
 base excision repair (BER), 2:4f, 2:5
 biological effects, 2:79
 DNA polymerase I, 2:87–91
 base pairing, 2:89–90
 damage sensing mechanisms, 3:335
 domain structures, 2:87–88, 2:88f
 5' exonuclease, 2:90–91, 2:91f
 functional role, 2:54–56, 2:55t, 2:56f
 mismatch repair (MMR), 4:77
 O-helix, 2:88–89, 2:88f
 phosphoryl transfer reactions, 2:89, 2:89f
 reaction mechanisms, 2:89, 2:89f
 research background, 2:87
 structural domains, 2:87
 substrate binding, 2:88–89, 2:88f
 sugar specificity, 2:90
 3'–5' exonuclease, 2:90
 DNA polymerase III holoenzyme, 2:92–95
 α subunit, 2:92–93, 2:93f
 bacterial species, 2:95
 β subunit, 2:93–94, 2:93f
 biological functions, 2:95
 DNA replication, 2:92, 2:93f, 2:94f, 2:707f, 2:709–710
 DnaX complex, 2:95
 leading- and lagging-strand DNA synthesis, 2:94f, 2:95
 mismatch repair (MMR), 2:41f, 2:67, 2:67f, 2:95
 Pol III core, 2:92–93, 2:93f
 Pol III' subassembly, 2:93, 2:93f
 Pol III* subassembly, 2:93–94, 2:93f
 schematic diagram, 2:93f
 subunits and subassemblies, 2:92–95, 2:93f
 τ subunit, 2:93, 2:93f
 DNA polymerase IV, 2:709–710, 4:480–481
 DNA polymerase V, 4:477–481
 bacterial homologs, 4:480–481
 cleavage reactions, 4:477–478, 4:478f
 eukaryotic homologs, 4:481
 fidelity, 4:480
 mutagenesis, 4:478
 processivity, 4:480
- regulatory mechanisms, 4:480
 research background, 4:479
 research summary, 4:481
 SOS response, 2:709–710, 4:477, 4:478f, 4:479f
 specificity, 4:479–480, 4:480f
 transcription regulation, 4:477, 4:478f
 translesion synthesis (TLS), 4:479, 4:479f
- DNA replication, 4:236
 excision and repair synthesis, 2:56
 functional role, 2:55t, 2:56f
 GO system (MutM–MutT–MutY), 2:80f
 polymerase accessory proteins, 2:119
 polymerase α -primase (Pol α), 2:241–244
 background information, 2:108, 2:241
 cell-cycle-dependent activity, 2:242
 characteristics, 2:119
 checkpoint pathways, 2:243–244
 functional role, 2:241–242, 2:242f
 polymerase α -primase (Pol α)–
 helicase relationship, 2:242–243, 2:242f
 S phase initiation proteins, 1:402t
 structural characteristics, 2:241–242
- polymerase β (Pol β), 2:82–86
 background information, 2:82
 biological role, 2:82–83, 2:83f
 C-subdomain, 2:84, 2:85f
 DNA ligases, 2:30t
 domain organization, 2:83–85, 2:83f
 D-subdomain, 2:84, 2:85f
 fidelity, 2:97t
 5'-deoxyribose-5-phosphate lyase, 2:5, 2:6f
 lyase domain, 2:5, 2:83–84, 2:83f, 2:85f
 N-subdomain, 2:84–85, 2:85f
 polymerase β -base excision
 repair–protein interactions, 2:4f, 2:5, 2:85–86
 polymerase domain, 2:84–85
- polymerase δ (Pol δ), 2:245–248
 associated subunits, 2:245–246
 base excision repair (BER), 2:5, 2:247–248
 catalytic subunit, 2:245, 2:246f, 2:246t
 DNA recombination, 2:247
 DNA replication, 2:120, 2:246–247, 2:247f
 eukaryotic DNA replication, 2:120
 functional roles, 2:246
 mismatch repair (MMR), 2:51, 2:52t, 2:67f
 nucleotide excision repair (NER), 2:247–248
 processivity, 2:246
 proliferating cell nuclear antigen (PCNA), 2:246
 S phase initiation proteins, 1:402t
 structural domains, 2:245–246
- polymerase ϵ , 1:402t, 2:5, 2:108, 2:120
 polymerase γ (Pol γ), 2:97t
 polymerase η (Pol η)

- deoxyribonucleic acid (DNA) (*Continued*)
- catalytic activity, 4:572–573, 4:573*f*
 - characteristics, 4:572–574
 - functional homologs, 4:481
 - localizations, 4:573–574, 4:573*f*
 - mutation effects, 4:574
 - proliferating cell nuclear antigen (PCNA), 4:574
 - ubiquitin-binding zinc finger (UBZ), 4:574
- polymerase ι , 2:5, 4:481
- polymerase κ , 4:481
- polymerase λ , 2:5, 2:7, 2:30*t*, 2:32
- polymerization kinetics, 2:96–100
- catalytic reactions, 2:98–99, 2:99*f*
 - deoxynucleoside triphosphate (dNTP), 2:96, 2:98*f*, 2:99*f*
 - exonuclease proofreading, 2:99–100, 2:99*f*
 - fidelity, 2:97*t*, 2:98
 - general discussion, 2:96
 - incorporation measurements, 2:97–98, 2:97*f*, 2:97*t*
 - polymerization sequence, 2:96, 2:97*f*
 - research summary, 2:100
 - selectivity, 2:96–97, 2:97*t*
 - structural determinants, 2:98, 2:98*f*
- replication processes, 4:381–382, 4:382*f*
- reverse transcriptase (RT), 2:101–107
- biochemical characteristics, 2:103–104, 2:103*f*
 - enzymatic functions, 2:102–103
 - functional role, 2:104–105, 2:104*f*, 2:107
 - integrase, 2:105–107, 2:106*f*
 - integration processes, 2:105–107, 2:106*f*
 - research background, 2:101
 - retroviral genomes, 2:102, 2:102*f*
 - retroviral replication, 2:101, 2:105–107
 - ribonuclease H (RNase H) activity, 2:105, 2:106*f*
 - subunit structure, 2:102*f*
- riboswitches, 4:137–138
- sliding clamp/polymerase interactions, 3:577, 3:578*f*, 4:238
- sliding clamps, 2:119
- translesion synthesis (TLS), 2:247, 4:479, 4:479*f*, 4:572–574, 4:573*f*
- Y-family polymerases, 4:477–481
- bacterial homologs, 4:480–481
 - characteristics, 4:480, 4:572–574
 - cleavage reactions, 4:477–478, 4:478*f*
 - eukaryotic homologs, 4:481
 - mutagenesis, 4:478
 - regulatory mechanisms, 4:480
 - research summary, 4:481
 - SOS response, 4:477, 4:478*f*, 4:479*f*
 - transcription regulation, 4:477, 4:478*f*
 - translesion synthesis (TLS), 4:479, 4:479*f*
- DNA polymorphisms, 3:96
- DNA recombination, 4:29–34
- branch migration, 4:32
 - DNA repair processes, 4:29
- helicases
- D-loop disruption, 4:32–33
 - enzyme translocation, 4:30–31, 4:31*t*
 - Holliday junction (HJ) dissolution, 4:33
 - Holliday junction (HJ) resolution, 4:33
 - recombination inhibitors, 4:33–34
 - recombination models, 4:30*f*
 - research summary, 4:34
 - single-strand DNA (ssDNA), 4:31–32
- nucleases
- cleavage processes, 4:31
 - deoxyribonucleic acid (DNA), 4:34
 - D-loop dissociation, 4:32–33
 - enzyme translocation, 4:31*t*
 - recombination models, 4:30*f*
 - research summary, 4:34
 - repair models, 4:29, 4:30*f*
 - single-strand DNA (ssDNA), 4:31–32
 - strand invasion, 4:32
 - translocases, 4:30–31, 4:30*f*, 4:31*t*, 4:32
 - yeast Rad1/Rad10 nucleases, 4:32
- DNA replication, 2:108–113
- background information, 2:108
 - bacterial DNA, 2:114–117
 - DNA ligases, 2:115*f*, 2:116, 2:116*t*
 - DNA polymerase holoenzymes, 2:92–95, 2:114–115, 2:115*f*, 2:116*t*
 - 5'–3' nuclease proteins, 2:115*f*, 2:116, 2:116*t*
 - leading- and lagging-strand DNA synthesis, 2:92, 2:93*f*, 2:94*f*, 2:114, 2:115*f*, 2:116–117, 2:118, 2:119*f*
 - oriC* origins-based replication, 2:117, 2:121–125
 - primosomes, 2:115–116, 2:115*f*, 2:116*t*
 - protein components, 2:114–116
 - replication fork (RF) synthesis, 2:92, 2:93*f*, 2:94*f*, 2:116–117
 - single-strand DNA (ssDNA)-binding proteins, 2:93–94, 2:94*f*, 2:115*f*, 2:116, 2:116*t*
 - trombone replication model, 2:114, 2:115*f*
 - basic concepts, 2:118, 2:119*f*, 4:236
 - cell-cycle engine—coupling replication control mechanisms, 1:404–405, 1:404*f*
 - cyclin-dependent kinase (CDK) inhibitors, 1:404
 - pre-replication complex (preRC)
 - activation, 1:405
 - proteolysis, 1:404
 - dcm* gene life cycle, 4:76, 4:77*f*
 - DnaA functional domains, 2:121–122, 2:122*t*
 - DNA polymerases, 4:236
 - DNA-strand transferases, 4:35–42
 - adenosine triphosphate (ATP) hydrolysis, 4:38–39
 - anti-recombination activity, 4:40
- chromosomal resolution and restoration, 4:36*f*, 4:37
- D-loop assays, 4:30*f*, 4:32, 4:39*f*, 4:41
- filament dissociation, 4:37
- filament formation, 4:35
- filter-binding assays, 4:39*f*, 4:40–41
- functional role, 4:35–37, 4:36*f*
- general discussion, 4:35
- in vitro* assays, 4:39*f*, 4:40–41
- paranemic joint formation, 4:35–36, 4:40–41
- plectonemic joint formation, 4:37
- properties, 4:37–38, 4:37*f*
- recombination mediators, 4:40
- secondary binding sites, 4:39–40
- signaling mechanisms, 4:41
- single molecule studies, 4:41
- structural characteristics, 4:38*f*, 4:39
- three-strand DNA exchange assay, 4:39*f*, 4:41
- translesion synthesis (TLS), 4:41
- future research areas, 2:120
- homologous recombination (HR), 4:23
- inducible stable DNA replication (iSDR), 4:24
- kinase functions
- Cdc7 kinase, 1:403, 1:404*f*
 - cyclin-dependent kinases (CDKs), 1:403, 1:404*f*, 1:406–409
- leading- and lagging-strand DNA synthesis, 2:92, 2:93*f*, 2:94*f*, 2:114, 2:115*f*, 2:246–247, 2:247*f*
- mini-chromosome maintenance (MCM) proteins, 2:118, 2:119*f*
- oriC* origins-based replication, 2:121–125
- biochemical mechanisms, 2:121–122
 - characteristics, 2:117
 - DnaA–*oriC* complex formation, 2:122–123
 - DnaB–DnaC complex, 2:124
 - DnaB helicase, 2:124
 - DnaC helicase, 2:123–124
 - DNA sequence motifs, 2:121–122, 2:122*f*
 - factor for inversion stimulation (Fis), 2:122, 2:122*f*
 - integration host factor (IHF) binding sites, 2:122, 2:122*f*
 - loading process, 2:124
 - primer foundation, 2:124
 - replisome assembly process, 2:122–123
 - research summary, 2:124–125
- origin recognition complex (ORC), 2:108
- basic concepts, 2:118
 - conservation, 2:111
 - discovery, 2:111
 - elongation cycle, 1:402–403
 - functional role, 2:111–112
 - initiation phase, 1:401–402
 - kinase functions, 1:403
 - non-replicative functions, 2:112–113
 - pre-recombination complex
 - formation, 2:111–112, 2:112*f*

- pre-replication complex (preRC)
 - assembly, 1:401, 1:403f
 - regulation mechanisms, 2:112
 - specification mechanisms, 2:110
 - S phase initiation proteins, 1:402t
- origins
 - characteristics, 1:401–403
 - elongation cycle, 1:402–403
 - elongation factors, 1:401
 - initiation factors, 1:401
 - initiation phase, 1:401–402
 - kinase functions, 1:403
 - lower eukaryotes, 2:108–109
 - metazoan origins, 2:109–110, 2:109f
 - origin plasticity, 2:110–111
 - origin recognition complex (ORC), 2:110
 - pre-replication complex (preRC)
 - assembly, 1:401, 1:403f
 - prokaryotes, 2:108–109, 2:109f
 - replicon model, 2:108
 - research methodologies, 2:111
 - schematic diagram, 2:109f
 - specification mechanisms, 2:110
 - S phase, 1:401–405
 - viruses, 2:108–109
- polymerase accessory proteins, 2:119
- polymerase α -primase (Pol α), 1:402t, 2:119
- polymerase δ (Pol δ), 1:402t, 2:120, 2:246–247, 2:247f
- polymerase ϵ , 1:402t, 2:120
- proliferating cell nuclear antigen (PCNA), 1:402t, 2:119
- protein components
 - DNA ligases, 2:115f, 2:116, 2:116t
 - DNA polymerase holoenzymes, 2:92–95, 2:114–115, 2:115f, 2:116t
 - 5′–3′ nuclease proteins, 2:115f, 2:116, 2:116t
 - primosomes, 2:115–116, 2:115f, 2:116t
 - requirements, 2:114–116
 - single-strand DNA (ssDNA)-binding proteins, 2:93–94, 2:94f, 2:115f, 2:116, 2:116t
- recombination-dependent DNA replication (RDR), 4:23–28
 - backup mechanisms, 4:25–26, 4:25f
 - bacterial cells, 4:24
 - bacteriophage studies, 4:23–24, 4:24f
 - break-and-copy models, 4:23
 - cancer predisposition, 4:27
 - displacement loop (D-loop)
 - formation, 4:23–24, 4:24f
 - eukaryotic cells, 4:24–25
 - genome instability, 4:27, 4:27f
 - inducible stable DNA replication (iSDR), 4:24
 - replication fork (RF) breakage, 4:25–26, 4:25f, 4:27, 4:27f
 - restart pathways, 4:26–27, 4:27f
- replication factor C (RFC), 2:120
- replication fork (RF)
 - backup mechanisms, 4:25–26, 4:25f
- bacterial DNA, 2:114–117
- breakage consequences, 4:27, 4:27f
- DNA polymerase holoenzymes, 2:92–95
- eukaryotes, 2:118–120
- functional role, 4:47–49, 4:48f
- lagging-strand DNA synthesis, 2:92, 2:93f, 2:94f, 2:115f, 2:116–117
- leading-strand DNA synthesis, 2:92, 2:93f, 2:94f, 2:116, 2:117
- restart pathways, 4:26–27, 4:27f
- replication protein A (RPA), 1:402t, 2:118–119, 2:119f
- replisome assembly process
 - characteristics, 2:122–123
 - DiaA protein stabilization, 2:123
 - DnaA-dependent unwinding, 2:123
 - DnaA–*oriC* complex formation, 2:122–123
 - HU protein stabilization, 2:123
 - sliding clamps, 3:577–579, 3:579f
- research summary, 2:113
- sliding clamps, 3:576–580, 4:236–241
 - archaeal processivity factor, 4:239–240, 4:240f
 - bacteriophage processivity factor, 4:240, 4:240f
 - basic concepts, 4:236
 - clamp/clamp-loader interactions, 2:114–115, 2:115f, 2:116t, 3:576, 3:577f, 4:237–238, 4:237f
 - clamp configurations, 4:236–237, 4:237f
 - clamp/DNA interactions, 3:577, 3:578f, 4:238–239
 - clamp/polymerase interactions, 3:577, 3:578f, 4:238
 - Escherichia coli* β -clamp, 3:577, 3:578f, 4:236–237, 4:237f, 4:238f
 - future research areas, 3:580
 - Homo sapiens* 9–1–1 clamp, 4:239, 4:239f
 - Homo sapiens* PCNA, 4:239, 4:239f
 - polymerase exchange, 3:577–579, 3:579f
 - polymerase switching, 3:577, 3:578f
 - processivity, 2:119, 3:576–580, 4:236–237
 - proliferating cell nuclear antigen (PCNA), 3:576, 3:577f, 3:578f, 4:236–237, 4:237f, 4:238f
 - replication factor C (RFC), 3:576, 3:577f, 4:237–238, 4:237f
 - research summary, 4:240–241
 - structural characteristics, 3:576, 3:577f
 - translesion synthesis (TLS), 3:577–579, 3:579f
 - viral processivity factor, 4:240
- S phase, 1:401–405
 - Cdc7 kinase, 1:403, 1:404f
 - cell-cycle engine–coupling replication, 1:404–405, 1:404f
 - chromosome structure regulation, 1:405
 - cyclin-dependent kinase (CDK) inhibitors, 1:404
- cyclin-dependent kinases (CDKs), 1:403, 1:404f, 1:406–409
- elongation cycle, 1:402–403
- elongation factors, 1:401
- initiation factors, 1:401
- initiation phase, 1:401–402
- initiation proteins, 1:402t
- kinase functions, 1:403
- origin characteristics, 1:401–403
- pre-replication complex (preRC)
 - activation, 1:405
- pre-replication complex (preRC)
 - assembly, 1:401, 1:403f
- proteolysis, 1:404
- regulation mechanisms, 1:401
- research summary, 1:405
- topoisomerases, 2:120
- trombone replication model, 2:114, 2:115f
- DNA transposons
 - classifications and characteristics, 3:283, 3:284f
 - diversity-generating retroelements (DGRs), 3:284
 - extra-chromosomally-primed (EP) retrotransposons
 - characteristics, 3:284
 - integration site preference, 3:288
 - long terminal repeat (LTR) retrotransposons, 3:284–286, 3:285f, 3:286f, 3:287f
 - open reading frames (ORFs), 3:284–286, 3:286f, 3:287f
 - tyrosine recombinases, 3:285f, 3:287
 - virus-like particles (VLPs), 3:284–286, 3:286f
- functional role
 - genome structure and function, 3:289–290
 - host gene expression, 3:290
 - laboratory and clinical applications, 3:290–291
 - retroviruses, 3:290
- future research areas, 3:291
- integration site preference
 - extra-chromosomally-primed (EP) retrotransposons, 3:288
 - general discussion, 3:288–289
 - target-primed (TP) retrotransposons, 3:288–289
- major element classes, 3:284
- regulation mechanisms, 3:289
- retrotransposons, 3:283–291
- reverse transcriptase (RT), 3:284
- target-primed (TP) retrotransposons
 - characteristics, 3:284
 - integration site preference, 3:288–289
 - long interspersed elements (LINEs), 3:284, 3:285f, 3:287–288
 - organellar retrotransposons, 3:288
 - Penelope-like elements (PLEs), 3:284, 3:285f, 3:288
 - prokaryotic retrotransposons, 3:288
 - R2-type elements, 3:287, 3:287f
 - short interspersed elements (SINEs), 3:285f, 3:287–288

deoxyribonucleic acid (DNA) (*Continued*)

DNA unwinding

active, rolling or hand-over-hand model,
3:264–265, 3:265f

active versus passive mechanisms, 3:264
dimeric inch-worm models, 3:265,
3:265f

DnaA-dependent unwinding, 2:123

DnaA helicase, 2:121–122, 2:122f

hexameric enzyme helicases, 2:25–26,
2:25f

inch-worm mechanisms, 3:264, 3:264f,
3:265f

protein oligomerization, 3:263

unwinding mechanisms, 2:25–26, 2:25f,
3:263–264

double-strand DNA (dsDNA)

double-strand breaks (DSBs)

break-induced replication (BIR),
2:247

cell-cycle-dependent repair, 3:266,
3:269

damage checkpoints, 1:411, 1:412f

Dam methyltransferase, 2:41f

DNA ligases, 2:31–32

enzyme requirements, 3:266, 3:267f

homologous recombination (HR),
2:247, 2:557, 2:559f, 3:266

Ku protein, 3:266–267, 3:267f,
4:31–32

Mut protein family, 2:42f

non-homologous end joining (NHEJ),
3:266, 4:44–45

polymerase δ (Pol δ), 2:247

recombination-dependent DNA
replication (RDR), 4:24–25,
4:31–32

RecQ helicases, 4:44–45, 4:47f

hexameric enzyme helicases, 2:24–25

meiosis, 3:21–22, 3:22f

eukaryotic DNA replication, 2:108–113

background information, 2:108

DNA ligases, 2:31

origin recognition complex (ORC),
2:108

conservation, 2:111

discovery, 2:111

functional role, 2:111–112

non-replicative functions, 2:112–113

pre-recombination complex

formation, 2:111–112, 2:112f

regulation mechanisms, 2:112

specification mechanisms, 2:110

origins

lower eukaryotes, 2:108–109

metazoan origins, 2:109–110,
2:109f

origin plasticity, 2:110–111

origin recognition complex (ORC),
2:110

prokaryotes, 2:108–109, 2:109f

replicon model, 2:108

research methodologies, 2:111

schematic diagram, 2:109f

specification mechanisms, 2:110

viruses, 2:108–109

polymerase α -primase (Pol α),

2:241–244

background information, 2:108,
2:241

cell-cycle-dependent activity, 2:242

checkpoint pathways, 2:243–244

functional role, 2:241–242, 2:242f

polymerase α -primase (Pol α)–
helicase relationship, 2:242–243,
2:242f

structural characteristics, 2:241–242

recombination-dependent DNA

replication (RDR), 4:24–25

replication fork (RF), 2:118–120

basic concepts, 2:118, 2:119f

future research areas, 2:120

mini-chromosome maintenance

(MCM) proteins, 2:118, 2:119f

polymerase accessory proteins, 2:119

polymerase α -primase (Pol α), 2:119

polymerase δ (Pol δ), 2:120

polymerase ϵ , 2:120

proliferating cell nuclear antigen

(PCNA), 1:402t, 2:119

replication factor C (RFC), 2:120

replication protein A (RPA), 1:402t,
2:118–119, 2:119f

sliding clamps, 2:119

topoisomerases, 2:120

research summary, 2:113

functional role, 4:232

GAL promoter system, 1:525–529

chromatin regulation function

active (galactose-induced) state, 1:527

inactive states, 1:527–528

promoter nucleosomes–regulatory
factors interactions, 1:528–529

GAL1–10 gene expression

basic concepts, 1:525–527

gene-specific factors, 1:525–526

hypersensitive region (HR),
1:526–527, 1:526f

nucleosome positioning, 1:525

promoter chromatin organization,
1:526–527, 1:526f

promoter elements, 1:525, 1:526f

regulatory and transcription factors,
1:525–526

gene-specific factors

Gal3p, 1:525–526

Gal4p, 1:525, 1:528–529

Gal80p, 1:525, 1:528–529

global process factors, 1:526

promoter chromatin organization

GAL1–10 gene expression, 1:526–527,
1:526f

regulation function, 1:529

genetic information transfer, 4:185, 4:186f

helicases

bacterial DNA replication, 2:92, 2:93f,

2:94f, 2:115–116, 2:116t,

2:121–122, 2:122f

characteristics and functional role, 2:21,
3:261, 4:43

D-loop dissociation

enzyme disruption, 4:32–33

Holliday junction (HJ) dissolution,

4:30f, 4:33

Holliday junction (HJ) resolution,

4:30f, 4:33

resolvases, 4:30f, 4:32–33

DnaB–DnaC complex, 2:124

DnaB helicase, 2:21, 2:22, 2:22t,

2:25–26, 2:92, 2:93f, 2:94f, 2:116t,
2:124

DnaC helicase, 2:115–116, 2:116t,

2:123–124

DNA recombination

D-loop disruption, 4:32–33

enzyme translocation, 4:30–31, 4:30f,
4:31t

Holliday junction (HJ) dissolution,

4:33

Holliday junction (HJ) resolution,

4:33

recombination inhibitors, 4:33–34

recombination models, 4:30f

research summary, 4:34

single-strand DNA (ssDNA),
4:31–32

eukaryotic DNA replication, 2:118

hexameric enzyme action, 2:21–27

classifications, 2:21, 2:22t

double-strand DNA (dsDNA)

translocation, 2:24–25

loaders/loading process, 2:21–22,
2:22t

nucleoside triphosphatase (NTPase)-
powered translocation, 2:22–24,
2:24f

research summary, 2:26

single-strand DNA (ssDNA)

interactions, 2:21

structural characteristics, 2:21, 2:23f

unwinding mechanisms, 2:25–26,
2:25f

mismatch repair (MMR), 2:67, 2:67f

PcrA helicase

characteristics and functional role,
3:261

crystal structures, 3:262, 3:262f

DNA unwinding, 3:263–264

open and closed conformations,
3:262–263

primary structures, 3:261–262

research summary, 3:265

single-strand DNA (ssDNA)

translocation, 3:263, 3:264,
3:264f

translocase-based protein

displacement, 3:265

RecA helicases, 2:21, 2:22t, 4:32

RecBCD helicases, 4:32

RecB helicases, 4:32

RecD helicases, 4:32

recombination inhibitors, 4:33–34

Rep helicases

characteristics and functional role,
3:261

crystal structures, 3:262, 3:262f

DNA unwinding, 3:263–264

domain structures, 3:263f

- open and closed confirmations, 3:262–263
- primary structures, 3:261–262
- research summary, 3:265
- single-strand DNA (ssDNA)
- translocation, 3:263, 3:264, 3:264f
- translocase-based protein displacement, 3:265
- ring-shaped helicases, 2:21
- SF1 (superfamily 1) helicases, 3:261–265
- characteristics and functional role, 3:261
 - crystal structures, 3:262, 3:262f
 - DNA unwinding, 3:263–264
 - domain structures, 3:263f
 - helicase motifs, 3:263f
 - open and closed confirmations, 3:262–263, 3:263f
 - primary structures, 3:261–262
 - research summary, 3:265
 - single-strand DNA (ssDNA) translocation, 3:263, 3:264, 3:264f
 - translocase-based protein displacement, 3:265
- single-strand DNA (ssDNA), 4:31–32
- UvrD helicases
- characteristics and functional role, 3:261
 - crystal structures, 3:262, 3:262f
 - DNA unwinding, 3:263–264
 - open and closed confirmations, 3:262–263
 - primary structures, 3:261–262
 - research summary, 3:265
 - single-strand DNA (ssDNA) translocation, 3:263, 3:264, 3:264f
 - translocase-based protein displacement, 3:265
- human immunodeficiency virus (HIV-1), 2:554
- hypoxia inducible factor (HIF), 4:67, 4:68f
- matrix metalloproteinases (MMPs), 3:96
- methyltransferases, 2:40–44, 3:6–10
- basic concepts, 3:6, 3:7f
 - cytosine methylation
 - basic concepts, 1:522
 - cancer cells, 1:522
 - functional role, 1:522
 - gene expression regulation, 1:522
 - genomic imprinting, 1:522
 - X-chromosome inactivation, 1:522
- Dam methyltransferase
- altered gene expression, 2:42–43
 - bacterial pathogenesis, 2:43
 - chromosome replication, 2:42
 - dam* mutant strains, 2:40
 - DNA adenine methylation identification (Dam-ID), 2:750
 - gene inactivation, 2:43–44
 - live cell imaging, 2:750
 - microRNA (miRNA), 3:101–102
- mismatch repair, 2:40–42, 2:41f, 2:42f, 2:54
- Mut protein system, 2:40–42, 2:41f, 2:42f
- oriC* region, 2:42, 2:121
- post-transcriptional regulation, 2:43
- properties, 2:40
- transposable elements (TEs), 3:281
- viruses, 2:43–44
- DNA demethylation
- glycosylation processes, 3:9, 3:10f
 - hydroxylation processes, 3:9
- Dnmt1, 3:8–9, 3:8f
- Dnmt3 family
- characteristics, 3:6, 3:7f
 - CpG dinucleotide spacing, 3:6
 - de novo* DNA methylation, 3:6, 3:7f
 - Dnmt3a, 3:6, 3:7f
 - Dnmt3L, 3:6, 3:7f
 - histone H3 lysine 4 (H3K4)/histone H3 lysine 9 (H3K9) methylation, 3:6–8, 3:7f
- 5-methylcytosine (5mC), 3:9, 3:9f
- methyl-CpG-binding proteins, 1:521–524
- binding domain family, 1:522
 - cytosine methylation, 1:522
 - DNA methylation–chromatin structure interdependency, 1:524
 - functional role, 1:521
 - gene expression regulation, 1:521
- research summary, 3:9–10
- SET- and RING-associated (SRA) domain, 3:9, 3:9f
- type II restriction and modification enzymes, 2:134
- type III restriction and modification enzymes, 2:137–138, 2:137t
- UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1), 3:9, 3:9f
- mismatch repair (MMR), 2:51–53, 2:54–59
- apoptosis
- damage-induced cell-cycle growth arrest, 2:52, 2:52f
 - tumorigenesis, 2:53, 2:53f
- basic concepts, 2:51
- B-DNA (antiparallel right-handed double helix), 2:142
- components, 2:52t
- cytotoxic agents, 2:51–52
- damage-induced cell-cycle growth arrest, 2:52, 2:52f
- Dam methyltransferase, 2:40–42, 2:41f, 2:42f
- disease states, 2:60–65
- Lynch syndrome/colorectal cancer, 2:61–64, 2:62t, 2:63f
 - replication fidelity, 2:60
 - research summary, 2:65
- DNA damage signaling interactions, 2:52
- DNA polymerase III holoenzyme, 2:95
- double-strand breaks (DSBs), 2:41, 2:42f, 2:79–80
- downstream steps
- characteristics, 2:69–70
 - excision, 2:69–70
 - resynthesis, 2:70
- eukaryotic MMR proteins, 2:51
- functional role, 2:51–52, 2:52t
- G:T mispair repair
- MBD4-mediated repair, 2:16, 4:77–78
 - thymine DNA glycosylase (TDG), 2:16, 4:78
- very-short patch (VSP) repair, 4:76–77
- homologous recombination (HR), 2:45–50
- break-and-copy models, 4:23
 - DNA polymerase III holoenzyme, 2:95
 - Escherichia coli* anti-recombination activity, 2:47–48, 2:48f
 - evolutionary role, 2:50
 - loss of heterozygosity (LOH), 2:49–50, 2:49f
 - mutation-related carcinomas, 2:49–50, 2:49f
 - Saccharomyces cerevisiae* anti-recombination activity, 2:48–49
 - sequence divergence, 2:48f, 2:49–50, 2:49f
 - very-short patch (VSP) repair, 2:46f
- long-patch mismatch repair
- Dam methyltransferase, 2:54
 - DNA polymerase III holoenzyme, 2:95
 - drug resistance, 2:57
 - excision and repair synthesis, 2:56
 - initiation phase, 2:55–56
 - mismatch repair deficiency, 2:57
 - MutH pathway, 2:56, 2:56f
 - MutL pathway, 2:55–56, 2:56f
 - MutS pathway, 2:55, 2:56f
 - process mechanisms, 2:54–56, 2:56f
 - repair enzymes, 2:54–56, 2:55t, 2:56f
 - specificity, 2:54
- mammals, 2:66–71
- basic concepts, 2:66–68, 2:68f
 - downstream steps, 2:69–70
 - mutation types, 2:67f
 - MutL family proteins, 2:68–69, 2:69f
 - MutS family proteins, 2:68, 2:69f
 - research summary, 2:71
 - strand discrimination, 2:70–71, 2:70f
- mismatch repair systems (MRSs), 2:66–71
- basic concepts, 2:45, 2:46f, 2:66–68, 2:68f
 - downstream steps, 2:69–70
 - Escherichia coli* methyl-directed MRS, 2:45–46
 - eukaryotes, 2:46–47
 - GO system (MutM–MutT–MutY), 2:80, 2:80f
 - long patch mismatch-repair system (LPMRS), 2:46f
 - mammals, 2:66–71
 - mutation types, 2:67f
 - MutL family proteins, 2:68–69, 2:69f

- deoxyribonucleic acid (DNA) (*Continued*)
- Mut protein family, 2:45–46, 2:48f, 2:54–59, 2:66–68, 2:67f
 - MutS family proteins, 2:60–61, 2:61f, 2:68, 2:69f
 - research summary, 2:71
 - sequence divergence, 2:48f, 2:66–68
 - specificity, 2:54
 - strand discrimination, 2:70–71, 2:70f
 - very short patch mismatch-repair system (VSPMRS), 2:46f
 - molecular mechanisms, 2:60–61, 2:61f
 - MutS–MutY interactions, 2:58
 - MutY repair pathway, 2:57–58, 2:57f
 - very-short patch (VSP) repair
 - homologous recombination (HR), 2:46f
 - research background, 4:76
 - vsr* gene biochemistry, 4:76–77
 - vsr* gene mutagenesis, 4:77
 - vsr* gene structure and regulation, 4:76, 4:77f
 - mitochondrial DNA (mtDNA), 3:132–133
 - background information, 3:132
 - base excision repair (BER), 2:7
 - characteristics and functional role, 2:126
 - disease connections
 - epigenomic mutations, 1:215–216
 - mitochondrial DNA (mtDNA), 1:212f, 1:214–215
 - nDNA-mitochondrial gene mutations, 1:215
 - DNA replication, 2:126–129
 - characteristics, 2:127–128
 - future research areas, 2:128
 - mammalian mtDNA, 2:127–128, 2:127f
 - mitochondrial constituents, 2:126
 - proteins, 2:126
 - ribonucleic acid (RNA), 2:126
 - strand-displacement mtDNA replication, 2:127–128, 2:127f
 - yeast systems, 2:128
 - epigenomics, 1:213
 - functional role, 1:210–213
 - future research areas, 3:134
 - gene expression, 3:140–144
 - basic process, 3:140–141, 3:141f
 - encoding processes, 3:140
 - 5'-end–RNase P process pathway, 3:142, 3:142f
 - mosaic mitochondrial genes, 3:142–144, 3:143f
 - RNA splicing, 3:142–144, 3:143f
 - 3'-end–RNase process pathway, 3:142, 3:142f
 - transcription processes, 3:141, 3:141f, 3:142f
 - translation processes, 3:144
 - yeast systems, 3:140, 3:141f
 - genetic studies, 1:213
 - global radiation, 1:214–215, 1:214f
 - historical research, 3:132–133
 - metabolism, 1:212f
 - mitochondrial permeability transition pore, 1:212–213, 1:212f, 3:128t, 3:130–131
 - mutation effects, 1:214–215
 - origins, 1:210–213
 - oxidative phosphorylation (OXPHOS), 1:210–213, 1:212f
 - properties, 3:133
 - reactive oxygen species (ROS), 1:212f, 2:79
 - research summary, 1:216
 - schematic diagram, 1:211f, 1:212f
 - sequence and gene content, 3:133–134, 3:133f, 3:133t, 3:134t
 - mitochondrial DNA replication, 2:126–129
 - characteristics, 2:127–128
 - future research areas, 2:128
 - mammalian mtDNA, 2:127–128, 2:127f
 - mitochondrial constituents
 - mitochondrial DNA (mtDNA), 2:126
 - proteins, 2:126
 - ribonucleic acid (RNA), 2:126
 - strand-displacement mtDNA replication, 2:127–128, 2:127f
 - yeast systems, 2:128
 - multicopy single-stranded DNA (msDNA), 3:284
 - nucleoid-associated proteins (NAPs), 3:367–370
 - characteristics and functional role, 3:367, 3:368f, 3:368t
 - DNA supercoiling, 3:368–369, 3:369f
 - gene expression regulation, 3:367–368
 - topoisomerases, 3:369–370
 - nucleosomic regulation mechanisms
 - ATP-dependent nucleosome remodeling
 - chromodomain helicase DNA binding (CHD)/Mi-2 family, 4:60f, 4:64
 - INO80 (inositol requiring 80) family, 4:60f, 4:64–65, 4:65f
 - ISWI family, 4:60f, 4:64
 - remodeling enzymes, 4:63–65
 - schematic diagram, 4:65f
 - SWItching defective/Sucrose NonFermenting (SWI/SNF) family, 4:60f, 4:63–64
 - Swr1-dependent deposition of H2A. Z–H2B dimers, 4:64–65, 4:65f
 - histone modifications
 - acetylation, 4:60
 - citrullination, 4:62
 - covalent modifications, 4:59
 - histone variants, 4:60f, 4:62–63
 - methylation, 4:60–61
 - phosphorylation, 4:61
 - proteolysis, 4:62
 - sumoylation, 4:62
 - ubiquitylation, 4:61–62
 - remodeling mechanisms, 4:66
 - research summary, 4:66
 - nucleotide excision repair (NER), 3:337–340
 - alkylation damage, 2:14, 2:15f
 - basic concepts, 4:402
 - B-DNA (antiparallel right-handed double helix), 2:142
 - defective NER
 - disease connections, 3:339–340
 - mouse models, 3:340
 - DNA ligases, 2:30t
 - eukaryotes, 3:341–344
 - biological importance, 3:341
 - damage recognition, 3:341–342, 3:342t
 - damage verification, 3:342–343, 3:342t
 - global genome repair (GGR), 3:341–342, 3:342f
 - incision and synthesis processes, 3:342t, 3:343–344, 3:343f
 - NER factors, 3:342t
 - open complex formation, 3:342–343, 3:342t
 - repair pathways, 3:342f
 - transcription-coupled repair (TCR) pathway, 3:337, 3:342f, 3:342t, 3:344
 - eukaryotic genomes, 3:337
 - global genome repair (GGR)
 - damage recognition mechanisms, 4:402
 - schematic diagram, 4:403f
 - kinetic mechanisms, 3:338–339
 - oligonucleotide fragments, 3:337
 - operational principles, 3:337, 3:338f
 - operation modes, 3:337
 - polymerase δ (Pol δ), 2:247–248
 - repair mechanisms, 3:338f
 - repair pathways, 3:329f
 - repair synthesis detection, 3:338, 3:339f
 - substrate specificity, 3:338, 3:339f
 - transcription-coupled repair (TCR) pathway
 - basic concepts, 4:402–403
 - damage recognition mechanisms, 4:402
 - lesion-related transcription arrest, 4:404–405, 4:404t
 - research perspective, 4:405–406
 - RNA polymerase II (Pol II)-based arrest and initiation, 4:403–404, 4:404t
 - schematic diagram, 4:403f
 - ultraviolet (UV) damage, 4:572
 - oxidative damage, 2:77–81
 - biological effects
 - aging/age-related diseases, 2:79
 - carcinogenesis, 2:79
 - cytotoxicity, 2:79
 - mutagenesis, 2:79
 - damage mechanisms
 - deoxyribose oxidation, 2:79
 - 8-oxo-guanine (8-oxoG), 2:78, 2:78f, 4:405
 - formamidopyrimidines, 2:78, 2:78f
 - oxidized adenine lesions, 2:78, 2:78f
 - oxidized pyrimidine lesions, 2:78f, 2:79
 - principal lesions, 2:78f

- reactive oxygen species (ROS), 2:78–79
- thymine glycol, 2:78f, 2:79, 4:405
- lesion sources and characteristics, 2:1–3, 2:2f
- oxidants and antioxidants, 2:77
- repair mechanisms
 - base excision repair (BER), 2:1
 - characteristics, 2:79–80
 - GO system (MutM–MutT–MutY), 2:80, 2:80f
 - transcription-coupled repair (TCR) pathway, 4:405
- platelet-activating factor receptor (PAFR), 3:534–535
- protein–DNA interactions, 2:149–153
 - DNA-binding motifs
 - basic helix–loop–helix (bHLH) motifs, 2:151f, 2:152
 - basic leucine zipper motif, 2:151f, 2:152
 - β -sheet motif, 2:151f, 2:152
 - characteristics, 2:151–152
 - helix–turn–helix (HTH) motif, 2:151–152, 2:151f
 - ribbon–helix–helix motif, 2:151f, 2:152
 - zinc fingers, 2:151f, 2:152
 - protein structural elements
 - α -helices, 2:150, 2:151f
 - β -strands, 2:150–151, 2:151f
 - loops, 2:151
- research summary, 2:152–153
- specificity, 2:149
- structural and chemical properties
 - DNA sequences, 2:149–150
 - importance, 2:149–150
 - major and minor grooves, 2:149, 2:150f
 - nonspecific protein–DNA interactions, 2:150
 - schematic diagram, 2:150f
- recognition helices, 2:697f
- retrotransposons, 3:283–291
 - classifications and characteristics, 3:283, 3:284f
 - diversity-generating retroelements (DGRs), 3:284
 - extra-chromosomally-primed (EP) retrotransposons
 - characteristics, 3:284
 - integration site preference, 3:288
 - long terminal repeat (LTR) retrotransposons, 3:284–286, 3:285f, 3:286f, 3:287f
 - open reading frames (ORFs), 3:284–286, 3:286f, 3:287f
 - tyrosine recombinases, 3:285f, 3:286f
 - virus-like particles (VLPs), 3:284–286, 3:286f
- functional role
 - genome structure and function, 3:289–290
 - host gene expression, 3:290
 - laboratory and clinical applications, 3:290–291
- retroviruses, 3:290
- future research areas, 3:291
- integration site preference
 - extra-chromosomally-primed (EP) retrotransposons, 3:288
 - general discussion, 3:288–289
 - target-primed (TP) retrotransposons, 3:288–289
- major element classes, 3:284
- regulation mechanisms, 3:289
- reverse transcriptase (RT), 3:284
- target-primed (TP) retrotransposons
 - characteristics, 3:284
 - integration site preference, 3:288–289
 - long interspersed elements (LINEs), 3:284, 3:285f, 3:287–288
 - organellar retrotransposons, 3:288
 - Penelope-like elements (PLEs), 3:284, 3:285f, 3:288
 - prokaryotic retrotransposons, 3:288
 - R2-type elements, 3:287, 3:287f
 - short interspersed elements (SINEs), 3:285f, 3:287–288
- secondary structures, 2:142–148
 - B-DNA (antiparallel right-handed double helix), 2:142
 - copy number variation (CNV), 2:147
 - cruciforms/slipped structures
 - background information, 2:146
 - biological role, 2:146
 - chemical properties, 2:146
 - pathology, 2:147, 2:147f
 - structural characteristics, 2:143f
 - molecular biology, 2:148
 - pathology, 2:146–147, 2:147f
 - quadruplex DNA (G4-DNA)
 - background information, 2:144–146
 - biological role, 2:144–146, 2:145f
 - bonding schemes, 2:144f
 - chemical properties, 2:144
 - pharmacological role, 2:146
 - structural characteristics, 2:143f
 - triplex DNA (H-DNA)
 - background information, 2:143–144
 - biological role, 2:144
 - bonding schemes, 2:144f
 - chemical properties, 2:143–144
 - structural characteristics, 2:143f
- Z-DNA
 - background information, 2:142–143
 - biological role, 2:142–143
 - chemical properties, 2:142, 2:143f
- single-strand DNA (ssDNA)
 - bacterial DNA replication, 2:93–94, 2:94f, 2:115f, 2:116, 2:116t
- DNA ligases, 2:30t
- helicases
 - characteristics and functional role, 3:261
 - DNA recombination, 4:31–32
 - DNA unwinding, 3:263–264
 - domain structures, 3:263f
 - helicase motifs, 3:263f
 - hexameric enzyme action, 2:21, 2:23f
 - mismatch repair (MMR), 2:67, 2:67f
- polymerase α -primase (Pol α)–helicase relationship, 2:242–243
- research summary, 3:265
- translocase-based protein displacement, 3:265
- translocation, 3:263, 3:264, 3:264f
- SOS response, 4:477, 4:478f
- sugar nucleotide transporters, 4:350–351
- supercoils, 2:154–156
 - biological effects, 2:156
 - eukaryotic versus prokaryotic cells, 2:161
 - handedness, 2:154–156, 2:155f
 - linking number (*Lk*), 2:154
 - research background, 2:154
 - stress generation mechanisms
 - hyperthermophilic eubacteria, 2:161–162
 - mesophilic/psychrophilic eubacteria, 2:161
 - superhelicity, 2:154, 2:155f, 2:163
 - topoisomerases, 2:156, 2:161, 2:162f, 2:163, 3:368–369, 3:369f
 - transcription and replication processes, 2:162, 2:162f
 - twists and writhes, 2:154, 2:163
- telomeres, 4:380–386
 - characteristics and functional role, 4:380
 - chromosome organization, 1:538
 - non-homologous end joining (NHEJ), 3:272, 4:384–385
 - replication processes
 - general discussion, 4:384
 - negative regulation, 4:384, 4:385f
 - positive regulation, 4:384
 - research background, 4:380–384
 - telomerase-independent maintenance, 4:384
 - telomerase repression
 - disease implications, 4:385
 - stem cells, 4:386
 - tumor suppression, 4:385–386
 - telomere capping, 4:384–385
 - telomeric complex
 - telomerase, 4:381–382, 4:382f
 - telomere proteins, 4:382–384, 4:383f
 - telomeric DNA, 4:380–381, 4:381f
 - telomeric RNA, 4:384
- topoisomerases
 - functional role, 2:120, 2:163–164
- reverse gyrase
 - adenosine triphosphate (ATP) hydrolysis, 2:159–160
 - classifications, 2:158t
 - properties, 2:158
 - protein domains, 2:159f
- supercoils, 2:156, 2:161, 2:162f, 2:163, 3:368–369, 3:369f
- type I DNA topoisomerase, 2:157–162
 - catalytic reactions, 2:157
 - cellular roles, 2:161–162
 - classifications, 2:158, 2:158t
 - DNA supercoiling, 3:369f
 - Escherichia coli*, 3:369t
 - functional role, 2:157, 2:163–164
 - hyperthermophilic eubacteria, 2:161–162

- deoxyribonucleic acid (DNA) (*Continued*)
- mesophilic/psychrophilic eubacteria, 2:161
 - nomenclature, 2:158, 2:158t
 - properties, 2:158
 - replication processes, 2:162, 2:162f
 - supercoils, 2:161, 2:162f
 - transcription processes, 2:162, 2:162f
 - type IA DNA topoisomerase, 2:158–160, 2:158t, 2:159f
 - type IB DNA topoisomerase, 2:158, 2:158t, 2:160–161, 2:160f
 - type II DNA topoisomerase, 2:163–168
 - antibacterial drugs, 2:167, 2:167f
 - anticancer drugs, 2:167, 2:167f
 - archaeal type IIB DNA topoisomerase, 2:166
 - cellular roles, 2:161–162
 - characteristics, 2:164–165
 - dietary topoisomerase II poisons, 2:167–168, 2:167f
 - domain structures, 2:164–165, 2:165f
 - double-stranded DNA passage reaction, 2:164, 2:164f
 - enzyme mechanism, 2:164, 2:164f
 - Escherichia coli*, 3:369t
 - eukaryotic type IIA DNA topoisomerase, 2:166
 - functional role, 2:1
 - prokaryotic type IIA DNA topoisomerase, 2:165–166
 - as therapeutic targets, 2:167–168
 - Top6A/Top6B, 2:165, 2:165f, 3:369–370
 - type IIA DNA topoisomerase, 2:165–166
 - type III DNA topoisomerase
 - Escherichia coli*, 3:369t
 - protein domains, 2:159f
 - sources, 2:158t
 - type IV DNA topoisomerase, 2:164–165, 2:165f, 2:166, 3:369t
 - type V DNA topoisomerase, 2:158t
 - type VI DNA topoisomerase, 2:166
 - topological properties, 2:163
 - type I restriction and modification enzymes, 2:72–75
 - antirestriction reactions, 2:75–76
 - assembly and control mechanisms, 2:74
 - biochemical function, 2:72, 2:73f
 - protein structures
 - hsd* family gene expression, 2:72–74
 - modification (M) subunits, 2:73
 - restriction (R) subunits, 2:74
 - sequence specificity (S) subunits, 2:72–73, 2:73f
 - restriction reactions
 - adenosine triphosphate (ATP) hydrolysis, 2:74, 2:75f
 - DNA binding, 2:74
 - DNA cleavage, 2:75
 - DNA translocation, 2:74–75, 2:75f
 - unmodified DNA recognition, 2:74
 - type II restriction and modification enzymes, 2:130–135
- S-adenosyl-L-methionine (AdoMet), 2:130
- characteristics
 - nonspecific DNA interactions, 2:133
 - protein structures, 2:131
 - recognition sequences, 2:131, 2:132t
 - sequence cleavage mechanisms, 2:132–133
 - classifications, 2:130
 - methyltransferases
 - enzyme characteristics, 2:134
 - modification functions, 2:134
 - new systems research, 2:130
 - organization
 - expression regulation, 2:131
 - modification and restriction genes, 2:130–131
 - research applications, 2:134
 - research background, 2:130
 - subsets
 - background information, 2:133–134
 - type IIB enzymes, 2:132t, 2:133–134
 - type IIE enzymes, 2:132t, 2:133
 - type IIF enzymes, 2:132t, 2:133
 - type IIS enzymes, 2:132t, 2:133
- type III restriction and modification enzymes, 2:136–141
- action mechanisms, 2:138–139, 2:139f
 - S-adenosyl-L-methionine (AdoMet), 2:136, 2:137f, 2:137t
 - cofactor requirements, 2:136
 - domain organization
 - methyltransferase (MTase), 2:137–138, 2:137t
 - modification (M) subunits, 2:137–138, 2:137t
 - restriction subunit, 2:136–137, 2:137f
 - phase variations, 2:139–140
 - restriction activity regulation, 2:136
- ultraviolet (UV) damage
- cyclobutane pyrimidine dimers (CPDs), 4:572–573
 - polymerase η (Pol η), 4:572–574
 - pyrimidine 6–4 pyrimidone photoproduct (6–4 PP), 4:574
 - xeroderma pigmentosum (XP), 4:572
- zinc-finger domains
- DNA-recognition code, 4:577–578, 4:578f
 - DNA-recognition helix, 4:577, 4:578f
 - functional role, 4:577–578
 - RNA recognition, 4:578
 - structural characteristics, 4:577, 4:578f
- deoxyribonucleotide triphosphate (dNTP), 2:88–89, 2:89f
- deoxyribose phosphate (dRP), 2:82, 2:83f
- deoxythymidine diphosphate (dTDP), 2:23–24
- deoxythymidine monophosphate (dTMP), 2:325, 2:325f
- deoxythymidine triphosphate (dTTP), 2:22, 3:716
- deoxyuridine diphosphate (dUDP), 3:715–716, 3:715f
- deoxyuridine monophosphate (dUMP), 2:325, 2:325f, 3:715–716, 3:715f
- depression
- neurokinin-1 receptor (NK1R), 4:364–365
 - tetrahydrobiopterin deficiency, 3:669
- dermatan sulfate (DS), 3:654–655, 3:655f
- Δ -5 desaturase (Δ 5ase), 2:287f, 2:288, 2:288f
- Δ -6 desaturase (Δ 6ase), 2:287f, 2:288, 2:288f
- desmin, 2:632t, 2:634, 3:236
- desmocollins, 1:255–256, 1:261, 1:637, 1:637f
- desmogleins, 1:255–256, 1:261, 1:637, 1:637f
- desmoplakin, 1:257, 1:258f, 1:259t, 1:637–638, 1:637f, 1:638t
- desmoplasia, 1:438
- desmosomes, 1:636–643
- assembly and maintenance, 1:639
 - desmosomal cadherins, 1:255–256, 1:256f, 1:259t, 1:261, 1:637, 1:638t
 - disease connections, 1:638t
 - functional role
 - cell–cell adhesion, 1:638t, 1:639
 - morphogenetic functions, 1:639
 - signaling functions, 1:639–640
 - molecular components
 - armadillo family members, 1:637–638, 1:638t
 - desmosomal cadherins, 1:637, 1:638t
 - plakins, 1:638–639, 1:638t
 - structural characteristics, 1:636–639, 1:637f
 - occurrences and functional role, 1:636–640
 - tight junctions, 4:393f
 - ultrastructure, 1:636, 1:637f
- desmuslin, 2:632t, 2:634
- Desulfovibrio*, 1:487t, 2:207f, 2:562–563, 4:88–89
- Desulfovibrio desulfuricans*
- iron-iron (FeFe) hydrogenases, 2:564f, 2:565
 - periplasmic nitrate reductase (Nap), 3:404
 - reductive reactions, 1:487t
- Desulfovibrio vulgaris*
- biosynthetic pathway, 3:546
 - Cho phylogenetic tree, 3:331f
 - [FeFe]-hydrogenases, 4:72
 - nickel-iron (NiFe) hydrogenases, 2:563f
 - uvrC* phylogenetic tree, 3:331f
- Desulfuromonas*, 2:207f
- DET2 genes, 1:251, 1:252f
- detergent-resistant membranes (DRMs)
- GPI-anchored proteins, 2:743
 - lipid-phase behaviors, 2:742–743
- detergents, 1:644–648
- chemical structure
 - characteristics, 1:644–646
 - head–tail linkages, 1:646
 - hydrophilic groups (headgroups)
 - ionic detergents, 1:644–645, 1:645f
 - nonionic detergents, 1:645–646, 1:645f
 - hydrophobic groups (tails)
 - ring-based hydrocarbons, 1:644, 1:645f
 - simple hydrocarbon chains, 1:644, 1:645f

- definition, 1:644
- membrane studies
- lipid bilayer reconstitution, 1:648
 - membrane proteins, 1:648
 - membrane solubilization, 1:647–648
- micelle structure
- ring-based detergents, 1:647
 - single-tailed detergents, 1:647, 1:647f
- properties
- aggregation number, 1:646–647
 - critical micelle concentration (CMC), 1:646, 1:646f
 - temperature effects, 1:647
- research summary, 1:648
- ring-based detergents, 1:647
- single-tailed detergents, 1:647, 1:647f
- deubiquitylating enzymes (DUBs), 3:597–598, 3:620
- deuterium (2H), 2:688
- DFak56 domain, 2:266t, 2:267–268
- DiaA protein, 2:123
- diabetes, 1:649–658
- acute complications
 - ketoacidosis, 1:654
 - nonketotic hyperosmolar coma, 1:654–655 - adiponectin, 3:73t, 3:74
 - amyloid deposits, 1:101t
 - anaplerotic pathways, 1:109–110, 1:109f
 - basic concepts, 1:649
 - β-cell genetic defects
 - maturity-onset diabetes of the young (MODY), 1:652–653
 - mitochondrial DNA mutations, 1:653
 - neonatal diabetes, 1:653
 - treatments and medications, 1:658
- branched-chain amino acid (BCAA) regulation, 1:245–246
- calpains, 1:360
- chronic complications
- advanced glycation end products (AGEs), 1:655–656, 2:409–410, 2:410f
 - general discussion, 1:656–657
 - hexosamine pathway, 1:656
 - macrovascular disease, 1:655
 - microvascular disease, 1:655–656
 - oxidative stress, 1:656–657
 - polyol pathway, 1:656
 - protein kinase C (PKC) activation, 1:656
- diabetic ketoacidosis, 1:109–110, 1:109f
- etiology
- general discussion, 1:649–652
 - insulin resistance
 - basic concepts, 1:650, 1:650f
 - insulin secretory defects, 1:652
 - molecular mechanisms, 1:650–652, 1:651f - miscellaneous types, 1:652
 - type 1 diabetes mellitus, 1:649
 - type 2 diabetes mellitus
 - calpains, 1:360
 - insulin resistance, 1:650, 1:650f
 - insulin secretory defects, 1:652
 - prevalence, 1:652
- gestational diabetes, 1:654
- gluconeogenesis, 2:383–384
- glycated hemoglobin, 2:405–406
- glycogen metabolism, 2:426–427
- glycogen synthase kinase-3 (GSK3), 2:437–438
- insulin resistance
- basic concepts, 2:606–607
 - etiology
 - basic concepts, 1:650, 1:650f
 - insulin secretory defects, 1:652
 - molecular mechanisms, 1:650–652, 1:651f - gestational diabetes, 1:654
 - insulin receptor mutations, 1:653
 - metabolic syndrome (syndrome X), 1:654
 - microRNA (miRNA)-based regulation, 4:194t
 - prediabetes, 1:654
 - secondary diabetes mellitus, 1:653
- leptin deficiency, 2:256
- Maillard reaction, 2:409–410, 2:410f
- maturity-onset diabetes of the young (MODY), 1:652–653
- metabolic syndrome (syndrome X), 1:654
- neonatal diabetes, 1:653
- nuclear factor kappa B (NF-κB), 3:304t
- O-GlcNAc biosynthesis, 3:352
- peroxisome proliferator-activated receptor alpha (PPARα), 3:421
- prediabetes, 1:654
- protein folding defects, 3:575t
- pyruvate dehydrogenase kinase (PDK) regulation, 4:324–326
- secondary diabetes mellitus, 1:653
- treatments and medications
- α-glucosidase inhibitors, 1:657–658
 - beta-cell therapies, 2:600–601
 - bromocriptine, 1:658
 - complication-related treatments, 1:658
 - general discussion, 1:658
 - incretin hormones, 1:658
 - insulin secretagogues, 1:657
 - metformin, 1:657
 - thiazolidinedione (TZD), 1:657, 3:418–419, 3:423
 - weight loss and exercise, 1:657
- tRNA synthetases, 4:443
- type 2 diabetes mellitus
- beta-cell therapies, 2:600–601
 - etiology
 - calpains, 1:360
 - insulin resistance, 1:650, 1:650f
 - insulin secretory defects, 1:652
 - prevalence, 1:652 - tRNA synthetases, 4:443
- diacid amino acids, 1:92t
- diacylglyceride (DAG), 1:49–51
- diacylglycerol (DAG), 1:659–663
- adrenergic receptors, 1:59
 - arachidonic acid regulated Ca²⁺ (ARC) channels, 1:128–129
 - 2-arachidonoyl glyceryl ether synthesis, 2:195f
 - B2 receptors, 1:242f
- B-cell antigen receptors (BCRs), 1:162–163, 1:163f
- calcium ion (Ca²⁺) signaling, 2:719–720, 2:720f
- cytidine diphosphate (CDP)-diacylglycerol (DAG) pathway, 3:478–480, 3:479f
- diabetes, 1:656
- diacylglycerol kinases (DGKs)
- compartmentalization, 1:660–661
 - EF-hand proteins, 1:319t
 - gene deletion studies, 1:661–662
 - isoforms, 1:659, 1:661f
 - regulatory mechanisms, 1:659–660
 - research summary, 1:663
 - signaling protein interactions, 1:661
 - structural domains, 1:659, 1:661f
- enzymatic pathways, 1:660f
- fatty acyl-coenzyme A (CoA) synthetases, 2:291f
- functional role, 2:414–415
- G₁₂/G₁₃ family functions, 2:345f
- G_{βγ} subunit functions, 1:480–482, 1:482f, 2:377, 4:265
- gene expression regulation, 1:311
- glycoprotein hormone receptors, 4:388–390
- G-protein-coupled receptors (GPCRs), 4:225f
- G_q family proteins, 2:496, 2:497–498, 2:498f
- guanine-nucleotide-exchange factors (GEFs), 4:246f
- historical research, 3:637
- insulin resistance, 2:607
- insulin-secreting cells (β-cells), 2:597f
- IP₃ receptors (IP₃R), 2:654f
- neurokinin-1 receptor (NK1R), 4:362–363
- neurotrophins, 3:241–243, 3:242f
- neutral zwitterionic glycerolipids, 2:416–417, 2:416f
- nuclear calcium signal, 1:310–311
- phagocytosis regulation, 3:439–440
- phosphatidic acid (PA)
- enzymatic pathways, 1:660f
 - phosphatidic acid phosphatases (PAPs), 1:662
 - enzymatic pathways, 1:659, 1:660f, 1:661f, 1:662
 - phosphatidic acid (PA), 1:662
 - physiological functions, 1:662–663
 - research summary, 1:663
 - signaling events, 1:662
 - signaling events, 1:662
 - synthesis, 2:416f
- phosphatidylinositol-4,5-bisphosphate (PIP₂), 3:447–448, 3:459
- phosphoinositides, 3:459, 3:460f
- phospholipase C (PLC), 2:654f, 3:467–468, 3:468f
- phospholipase D (PLD), 3:473–474
- platelet-activating factor receptor (PAFR), 3:536, 3:536f
- proteinase-activated receptors (PARs), 3:604
- protein kinase C (PKC), 3:638–639, 3:639f

- diacylglycerol (DAG) (*Continued*)
 research summary, 1:663
 second messenger pathways, 1:43f
 serotonin receptors, 4:225f, 4:226
 small guanosine phosphatases (GTPases), 4:247f
 somatostatin (sst) receptors, 4:265
 store-operated calcium (SOC) channels, 3:359–360, 3:360f
 structural characteristics, 2:412–414, 2:414f
 T-cell receptors (TCRs), 4:374
 vascular endothelial growth factor receptors (VEGFRs), 4:512
- diakinesis, 3:19f, 3:20
- diamine oxidases, 1:89
- diamminedichloroplatinum(II), 4:405
- Diamond-Blackfan anaemia, 4:121
- diaphanous autoinhibitory domain (DAD), 1:24
- diaphanous inhibitory domain (DID), 1:24
- diarylpyrimidine (DAPY) NNRTIs, 2:551
- diastrophic dysplasia achondrogenesis, 2:471t
- diazepam, 2:723–724, 2:724f
- dibasic amino acids, 1:92t
- dicalcin, 1:319t
- dicarboxylic acid decarboxylation, 2:208, 2:208f
- DICER-LIKE1 (DCL1) protein, 3:99–100, 3:99f, 4:253–255
- dichloroacetate, 3:71t
- 5,6-dichlororibofuranosyl benzamidazole (DRB) sensitivity inducing factor (DSIF), 4:164, 4:165f
- Dictyostelium*, 1:23, 1:319t, 1:567
- Dictyostelium discoideum*
 actin-related proteins (Arps), 1:37f
 characteristics, 1:562, 1:563f
 chemotaxis, 2:236–240
 characteristics and functional role, 2:236
 cytoskeletal polarization, 2:238
 mathematical models, 2:238–240, 2:238f
 metrics, 2:239b
 signaling pathways, 2:236–238, 2:237f
- class II PDEs, 1:572
- class III PDEs, 1:580
- cyclic adenosine monophosphate (cAMP), 1:562–566
- cyclic adenosine monophosphate (cAMP) receptors (cARs)
 affinities, 1:562–563, 1:563f
 desensitization mechanisms, 1:565
 expression and developmental stages, 1:562–563, 1:563f
 extracellular levels, 1:562–563, 1:563f
 G-protein-dependent pathways, 1:563–564, 1:564f
 G-protein-independent pathways, 1:564–565, 1:564f
 gradient sensing, 1:565–566
 molecular identification, 1:562–563
 schematic diagram, 1:563f
 signaling mechanisms, 1:565
- signaling pathways, 1:563–565, 1:564f
 functional role, 1:562, 1:563f
 cytokinesis, 1:624–625
 extra-chromosomally-primed (EP) retrotransposons, 3:285f
 intracellular neutral lipid storage droplets, 1:47, 1:48f
 phagocytosis, 3:437
- Diels-Alderase ribozymes, 4:144
- dietary carbohydrates, 2:286–287
- difluoromethylornithine (DFMO), 2:214f, 2:215
- digalactosyldiacylglycerol (DGDG), 1:508
- digitalis, 3:686–687
- Digitalis purpurea*, 3:686–687
- diglucosyl C₃₀-carotenoid, 1:502f
- digoxin, 3:686–687, 3:686f
- 7,8-dihydro-8-oxo-guanine (8-oxoG), 2:54
- dihydrobiopterin, 3:666–667, 3:668f
- dihydrofolate reductase (DHFR), 2:109–110, 2:325, 2:325f
- 24,25-dihydrolanosterol, 1:518f, 1:519
- dihydrolipoyl acetyltransferase, 4:321–322
- dihydrolipoyl dehydrogenase, 4:321–322, 4:324
- dihydroorotase (DHOase), 3:712, 3:713f, 3:714f
- dihydroorotate dehydrogenase (DHODH), 3:712, 3:713f, 3:714f
- dihydropyrimidine dehydrogenase (DPD), 3:717–718, 3:717f
- dihydrothymine, 3:330t
- dihydroxyacetone phosphate (DHAP), 2:415, 2:416f, 2:445f, 3:416–417, 3:499f
- dihydroxybenzene, 3:722–723
- dihydroxyphenylalanine (L-DOPA), 1:57, 2:173
- dihydrozeatin, 1:627, 1:628f
- dilated cardiomyopathy (DCM), 1:638t, 2:632t, 4:447–449, 4:448f
- dilated myocarditis, 3:117
- dimethylallyl pyrophosphate (DMAPP), 1:627–628
- 3,3-dimethyldioxirane (DMDO), 1:4f
- dimethylthiosulfonium triflate (DMTST), 1:4f
- dimethylsulfoxide (DMSO), 4:79, 4:80f
- dimethylsulfoxide reductase, 1:487t
- DinB homolog, 2:709–710, 4:480–481
- diol dehydratase, 1:156, 1:158t
- dioxygenases, 3:371–372
- dipeptidylpeptidase-4 (DPP-4), 1:116, 1:658
- 9, 10-diphenylanthracene, 1:463f
- 2,3-diphosphoglycerate (2,3-DPG), 2:534
- diplonema, 3:19, 3:19f, 3:22f
- disrupted-in-schizophrenia 1 (DISC1) protein, 1:197f, 1:199, 1:200f
- disulfide bond formation, 1:664–667
 basic concepts, 1:664, 1:665f
 catalytic reactions
 background information, 1:665
 protein disulfide isomerase (PDI), 1:665, 1:665f
 thioredoxin family, 1:665
 cellular environments
- extracellular proteins, 1:666, 1:667f
 regulatory mechanisms, 1:666
 thiol/disulfide redox environments, 1:666
- isomerization processes, 1:664, 1:665f
- molecular structure, 1:664, 1:665f
- oxidative protein folding
 basic principles, 1:664
 formation mechanisms, 1:664, 1:666f
 general discussion, 1:664
- disulfiram, 3:398
- ditercalinium complex, 3:330t
- diubiquitin, 4:471–472
- diversity-generating retroelements (DGRs), 3:284
- Dmc1 recombinase, 2:558, 2:559f, 4:37f
- dnaA* gene, 2:42
- DnaA helicase
 DNA unwinding, 2:121–122, 2:122f
 functional domains, 2:121–122, 2:122t
 replisome assembly process
 characteristics, 2:122–123
 DiaA protein stabilization, 2:123
 DnaA-dependent unwinding, 2:123
 DnaA-oriC complex formation, 2:122–123
 HU protein stabilization, 2:123
- DnaB–DnaC complex, 2:124
- DnaB helicase, 2:21, 2:22, 2:22t, 2:25–26, 2:92, 2:93f, 2:94f, 2:116t, 2:124
- DnaC helicase, 2:115–116, 2:116t, 2:123–124
- DNA Data Bank of Japan, 3:611–612, 3:615t
- DnaG helicase, 2:92, 2:93f, 2:116t
- DnaX complex, 2:95
- dnaX* gene, 2:95
- DOCK family proteins, 1:441f, 1:480–482, 4:111–112, 4:112f
- docosahexenoic acid, 2:288, 2:288f
- dogfish, 4:516t
- dolichols
 cholesterol biosynthetic pathway, 1:516, 1:517f
 dolichol phosphate (Dol-P), 2:457, 2:458f
 N-linked glycosylation, 2:457, 2:458f
- dolichyl pyrophosphate (Dol-P-P), 2:315–316, 2:315f, 2:316t
- Donnan potential, 3:161
- dopamine
 DARPP-32, 4:222f, 4:223
 dopamine β-hydroxylase, 1:57, 2:259, 2:259f
 dopamine receptors, 2:169–173
 calcium ion (Ca²⁺) signaling, 2:720f
 D1 and D2 receptor-selective drugs, 2:171
 receptor distribution, 2:172–173
 receptor signaling
 general discussion, 2:171–172
 G protein coupling, 2:171, 2:172t
 heterotrimeric G proteins, 2:171
 signaling pathways, 2:172
 receptor subtypes
 characteristics, 2:172t
 D1 and D2 subtypes, 2:169, 2:170f
 DNA sequences, 2:169–171, 2:170f

- general discussion, 2:169–171
 molecular cloning, 2:169–171, 2:170f
 receptor variants, 2:172t, 2:173
 research background, 2:169
 structural characteristics, 2:170f, 2:171
 therapeutic uses, 2:173
 monoamine oxidase, 3:187f
 synthetic pathways, 1:57
 transport mechanisms, 3:238–239
 doublecortin (DCX), 1:196–197, 1:197f
 double Holliday junction (DHJ), 2:558, 2:559f, 3:22f, 4:29, 4:30f, 4:46, 4:47f, 4:48f
double-sex (dsx) gene, 1:82, 1:82f, 1:85–86
 double-stranded RNAs, 3:98, 4:155–156, 4:156f
 Dovonex®, 4:539
 downstream regulatory element antagonist modulator (DREAM), 1:311, 1:313–314, 1:314f, 2:485t, 3:293, 3:295
 Down syndrome
 meiotic defects, 3:23
 tau proteins, 1:477t
 doxorubicin, 2:167, 2:167f
 Dpb11 loading factor, 1:401–402, 1:402t, 1:403f, 1:404f
 DPPY motif, 2:137–138, 2:137f
 DRIP complex, 4:543
Drosophila melanogaster
 actin-related proteins (Arps), 1:37f
 adenosine-to-inosine (A-to-I) RNA editing, 4:154–155
 alternative splicing-based sex
 determination regulation, 1:81–86
 aquaporin water channels, 1:125
 aspartic proteases, 1:137, 1:138f
 base excision repair (BER), 2:5t
 Bubblegum ACS family (ACSBG), 2:294
 calpains, 1:354f, 1:359
 carbohydrate-responsive element binding protein (ChREBP) distribution, 1:374
 cardiac muscle development, 1:183–185
 centromere region, 1:447t, 1:448f
 chromodomain helicase DNA binding (CHD)/Mi-2 family, 4:64
 classical zinc fingers, 4:577
 Csk protein kinase, 4:295–296
 DNA replication, 2:109–110
 EF-hand proteins, 1:319t
 endocrine regulation
 juvenile hormone (JH), 2:594
 prothoracicotropic hormone (PTTH), 2:594, 2:594f
 20-hydroxyecdysone (20E), 2:593, 2:593f
 focal adhesion kinase (FAK) family, 2:266f, 2:267–268
 G protein-coupled receptor kinases (GRKs), 2:483
 hexokinases, 2:544
 intracellular neutral lipid storage droplets, 1:47, 1:48f, 1:51
 IP₃ receptors (IP₃R), 2:653–655
 kinesins, 2:674
 kinesin superfamily proteins (KIFs), 2:679–681, 2:682f
 LacI-lamin protein, 3:315–316, 3:316f
 matrix metalloproteinases (MMPs), 3:90
 metabolomic research, 2:590
 microRNA (miRNA)-based lipid
 metabolism regulation, 4:192–193, 4:194t
 multiple-domain proteins, 3:628f
 olfactory receptors, 3:346
 ORAI protein family, 4:315
 origin recognition complex (ORC), 2:111
 P2Y receptors, 3:376–377
 parathyroid hormone (PTH)/PTH-related protein (PTHrP) receptor (PPR), 3:384
 peptide amidation, 3:398
 phospholipase D (PLD), 3:473–474
 platelet-derived growth factor (PDGF), 3:538
 protein palmitoylation, 3:645–646
 pterin distribution, 3:666
 recombination-dependent DNA
 replication (RDR), 4:25
 RecQ helicases, 4:43–44, 4:44f
 regulated intramembrane proteolysis (RIP)
 rhomboid-based mediation, 4:51f, 4:54
 sterol regulatory element binding protein (SREBP), 4:52
 retrotransposons, 3:283, 3:284f
 septin complexes, 4:214–215, 4:218
 sex determination
 double-sex (dsx) gene regulation, 1:85–86
 exon organization and splicing patterns, 1:81–82, 1:82f
 fruitless (fru) gene regulation, 1:85–86
 regulatory hierarchy, 1:81, 1:82f
 research summary, 1:86
 Sxl autoregulation, 1:82–84, 1:85f
 transformer (tra) gene regulation, 1:82, 1:85–86, 1:85f
 small RNAs (sRNAs), 3:98
 Src (sarcoma) family kinases, 4:297
 stromal-interacting molecule (STIM)
 protein family, 4:315, 4:316f
 sugar nucleotide transporters, 4:350
 SWItching defective/Sucrose NonFermenting (SWI/SNF) family, 4:63–64
 vitamin E metabolism, 4:549
 voltage-gated six-transmembrane K⁺ channels, 4:556–557
Drosophila simulans, 3:285f
Drosophila virilis, 3:285f
 dsRNA-binding motifs, 3:98, 4:155–156, 4:156f
 dual-specificity tyrosine phosphorylation-regulated kinase 2 (DYRK2), 1:312, 1:313f
 Duchenne muscular dystrophy, 2:471t, 2:474–475
 DUF 283 domain, 3:99–100
 dUTP pyrophosphatase (dUTPase), 2:2
 DWARF genes, 1:251
 dwarfism, 3:657t, 3:658
dwarf genes, 1:251, 1:252f
 dynein, 2:174–177
 binding activities
 cargo binding, 1:634, 2:175f
 cellular membranes, 2:177
 chromosomal kinetochores, 2:177
 cytoplasmic structures, 2:177
 general discussion, 2:176
 microtubule binding, 2:175f, 2:176–177, 2:176f
 motor binding, 2:175f, 2:176f, 2:177
 composition, 2:174, 2:175f
 functional roles
 dynein–dynactin complex
 adapter functions, 2:174–175
 dynein-based motility, 1:37f, 1:38t, 2:174
 processivity, 2:175
 dynein-independent functions, 2:175
 structural characteristics, 2:174, 2:175f
 subunit structure and functions
 actin-related/actin-binding proteins, 2:175
 β-helix proteins, 2:176
 dynamitin, 2:175–176, 2:175f
 elongated α-helical proteins, 2:175–176
 p150^{Glued protein}, 2:175–176, 2:175f, 2:176f
 p24 protein, 2:175–176, 2:175f
 p25 protein, 2:175f, 2:176
 p27 protein, 2:175f, 2:176
 p62 protein, 2:175f, 2:176
 RING/LIM domain protein, 2:175f, 2:176
 structural subfamilies, 2:175
 dynamin
 chemokine receptors, 1:483f
 clathrin-mediated endocytosis, 2:198, 2:198f
 dynamin related protein 1 (Drp1), 3:136–137, 3:137f
 nitric oxide synthase (NOS), 1:343–346
 dynamitin, 2:175–176, 2:175f
 dyneins, 2:184–187
 axonemal dyneins
 binding characteristics, 2:185t, 2:186–187
 inner arm dyneins, 2:185t, 2:187
 outer arm dyneins, 2:185t, 2:187
 characteristics and classifications, 2:184, 2:185t
 cytoplasmic dyneins
 cytoplasmic dynein 1, 2:185–186, 2:185t, 2:186f
 cytoplasmic dynein 2 (IFT dynein), 2:185t, 2:186
 kinesin superfamily proteins (KIFs), 2:684f
 cytoskeletal walking model, 1:632, 1:633f
 dynein–dynactin complex, 1:37f, 1:38t, 2:174, 4:208, 4:208f
 dynein heavy chain, 2:184–185, 2:185t, 2:186f
 dynein intermediate chain, 2:185t, 2:186f
 dynein light chain, 2:185t, 2:186f
 dysfunctional mitochondria, 3:173–174
 dyskeratosis congenita (DC), 4:121, 4:385
 dyskerin, 4:385
dystonia musculorum (dt), 2:628–629, 2:629t
 α-dystroglycan, 2:454f, 2:456, 2:474–475

dystrophic epidermolysis bullosa (DEB),
1:641t
dystrophin
 dystrophin–glycoprotein complex (DGC),
 2:474–475
 pectin, 2:625–627

E

- E2F transcription factor, 1:406–407, 1:407f,
1:412f, 4:106–107
early after-depolarization (EAD), 1:382–383
early-onset (type 1a) psoriasis, 1:638t
earthworms, 4:516t
EBF transcription factor, 1:187–188, 1:188f
eccrine glands, 2:669t
ecdysone
 ecdysone receptor (ECR), 2:593, 2:593f
 insects, 2:590, 2:592–593
echinoderms, 2:637–638, 2:638t
ecm29p proteasome activator, 3:592–593
EcoP15I enzyme, 2:136, 2:137t
EcoP1I enzyme, 2:136, 2:137t
EcoRII enzyme, 2:132t, 2:133
Ectothiorhodospiraceae, 3:698t
editosome complex, 4:156
eEF1A, 2:254
eels, 2:702t
efavirenz, 2:549
EF-hand proteins
 calmodulin (CaM), 1:318f, 1:319t, 1:321
 domain structures, 1:318–321, 1:318f
 endoplasmic reticulum (ER) lumen,
 1:291t, 1:295
 evolutionary development, 1:317–318,
 1:318f
 stromal-interacting molecule (STIM)
 protein family, 3:361–362, 3:361f
 structural characteristics, 1:317, 1:317f,
 1:318f
 subfamilies, 1:318–321, 1:319t
 temporal buffering, 1:316–317
egg fertilization, 2:640, 2:641f
eggNOG database, 3:615t
eggplant, 4:533t
Egr-1 transcription factor, 1:188–189
eicosanoid receptors, 2:188–191
 action mechanisms, 2:189
 biosynthetic pathways
 arachidonic acid (AA), 2:188–189
 cyclooxygenase (COX) pathway, 2:189
 epoxygenase pathway, 2:189
 lipoxygenase pathway, 2:188–189
 characteristics, 2:188
 lipoxygenase metabolite receptors
 characteristics, 2:189
 5-oxo-EET receptors, 2:189
 leukotriene receptors, 2:189
 lipoxin receptors, 2:189
 peroxisome proliferator-activated receptors
 (PPARs), 2:191
 prostaglandin receptors, 2:189–190,
 2:190f
8-oxo-guanine (8-oxoG)
 biological effects
 aging/age-related diseases, 2:79
 carcinogenesis, 2:79
 cytotoxicity, 2:79
 DNA glycosylases, 2:3t
 GO system (MutM–MutT–MutY), 2:80,
 2:80f
 molecular structure, 2:2f, 2:78f
 MutY repair pathway, 2:57–58, 2:57f
 oxidative damage, 2:1, 2:54, 2:78, 4:405
Eimeria tenella, 3:248f
elastin, 2:192–193
 amino acid composition, 2:192
 amino acid sequence motifs, 2:192
 characteristics and functional role, 2:192
 gene characterization
 gene promoter, 2:193
 gene structure, 2:193
 gene expression
 mitochondrial RNA (mRNA), 2:193
 protein isoforms, 2:193
 gene expression regulation
 characteristics, 2:193
 posttranscriptional regulation, 2:193
 transcriptional regulation, 2:193
 insoluble elastin, 2:192
 matrix metalloproteinases (MMPs), 3:91t
 soluble elastin, 2:192–193
elderberries, 2:702t
electrochemical cation-gradient-driven
 symporters
 classification, 2:388
 functional role, 2:388
 lactose permease (LacY), 2:388–389
 melibiose permease (MelB), 2:389
 sodium/galactose transport (SgIT), 2:389
 structural characteristics, 2:388, 2:389f
electron microscopy, 4:129–130
electron-nuclear double resonance (ENDOR)
 spectroscopy, 2:661–662
electron paramagnetic resonance (EPR),
2:660–661
electron transfer flavoprotein (ETF), 4:19f,
4:20
electron-transfer proteins, 2:311
electron-transferring flavoprotein (ETF)–
 ubiquinone oxidoreductase, 4:85
electron transport phosphorylation,
2:206–208, 2:207f, 2:208f
electrospray ionization analysis, 4:286
11S regulators (REG α , REG β , REG γ)/PA28s,
3:592, 3:593f
Elhers–Danlos syndrome, 2:471t, 2:475
elk, 3:575t
elongated α -helical proteins, 2:175–176
elongation cycle
 DNA synthesis, 1:402–403
 elongation factors
 bacterial transcription termination,
 4:168f, 4:169–170, 4:409–410
 DNA replication, 1:401
 elongation factor G (EF-G), 4:122–127
 bacterial translation elongation,
 4:420–421, 4:421f
 EF-G–ribosome interactions, 4:123f,
 4:125–126, 4:126f, 4:128–129,
 4:129f, 4:421f
 research summary, 4:126
 ribosome recycling, 4:126
 translocation cycle, 4:125, 4:124f
 EF-Tu–ribosome interactions, 4:123f,
 4:128–129, 4:129f, 4:421f
 guanosine triphosphate (GTP)
 hydrolysis, 4:123–125, 4:124f
 initial selection phase, 4:123, 4:124f
 research summary, 4:126
 substrate proofreading, 4:125
NusA/NusG elongation factors, 4:168f,
4:169–170, 4:409–410
protein biosynthesis elongation,
2:249–255
 aminoacyl (A) sites, 2:249–251, 2:250f,
 2:251f, 2:253f, 2:254f, 4:122–123
 background information, 2:249,
 4:122–123
 components, 2:250t
 eEF1A elongation factor, 2:249–251,
 2:250f, 2:250t, 2:252, 2:253–254,
 2:253f, 2:254f
 eEF2 elongation factor, 2:252, 2:254f
 eEF3 elongation factor, 2:253
 eEFSec elongation factor, 2:253
 elongation cycle chemistry, 2:251–252,
 2:251f
 elongation factor G (EF-G), 4:122–127
 EF-G–ribosome interactions, 4:123f,
 4:125–126, 4:126f
 research summary, 4:126
 ribosome recycling, 4:126
 translocation cycle, 4:125
 elongation factor Tu (EF-Tu), 4:122–127
 EF-Tu–ribosome interactions, 4:123f
 guanosine triphosphate (GTP)
 hydrolysis, 4:123–125, 4:124f
 initial selection phase, 4:123, 4:124f
 research summary, 4:126
 substrate proofreading, 4:125
entry (F) sites, 2:252, 2:253f
exit (E) sites, 2:252, 2:253f, 4:122–123
molecular mimicry, 2:252
Nobel-winning research, 2:254
peptidyl (P) sites, 2:249–251, 2:250f,
2:251f, 2:253f, 2:254f, 4:122–123
regulation mechanisms, 2:253–254
structural model, 2:254f
elongation initiation factor 2B (eIF2B),
2:604f, 2:605
elongin, 4:570–571, 4:571f
elvitegravir, 2:106f
Emanuel syndrome, 2:147, 2:147f
Embden–Meyerhof–Parnas pathway,
2:204–206, 2:205t
Embden–Meyerhof pathway, 4:74, 4:74f
embryo implantation
 implantation process, 3:205, 3:206f
 mucins, 3:205–208
 background information, 3:205
 expression and transcriptional control,
 3:206–208, 3:208f
 functional role, 3:207t, 3:208–209

- research summary, 3:209
- transmembrane proteins, 3:207t
- transport and metabolism regulation, 3:208–209
- research summary, 3:209
- Emery–Dreifuss muscular dystrophy, 3:313
- enabled/vasodilator-stimulated phosphoprotein (Ena/VASP), 1:28–29, 1:29–31, 1:29f, 1:30f, 1:19–20, 2:320–321, 2:320f
- enamelysin, 3:91t, 3:92f, 3:93
- endo- β -N-acetylglucosaminidase (ENGase), 2:330, 2:330f
- endocannabinoids, 2:194–196
 - anandamide
 - deactivation, 2:195
 - intracellular hydrolysis, 2:195
 - physiological regulation, 2:194
 - synthesis, 2:194, 2:195f
 - transport mechanisms, 2:195
- 2-arachidonoyl glyceryl ether
 - deactivation, 2:195
 - intracellular hydrolysis, 2:195
 - physiological regulation, 2:194–195
 - synthesis, 2:194, 2:195f
 - transport mechanisms, 2:195
- cannabinoid receptors, 2:195–196
- characteristics, 2:194
- functional role, 2:196
- endocytosis, 2:197–199
 - basic concepts, 3:437
 - caveolar endocytosis, 2:198, 2:198f
 - clathrin-independent endocytosis, 2:198
 - clathrin-mediated endocytosis
 - adaptor proteins, 2:197–198, 2:198f
 - basic concepts, 2:197, 2:198f
 - dynamins, 2:198, 2:198f
 - functional role, 2:197
 - internalized membrane fate, 2:198f, 2:199
 - noncoated endocytosis, 2:198f, 2:199
 - phagocytosis
 - basic concepts, 2:198f, 2:199
 - characteristics, 2:197
 - phosphatidic acid (PA), 3:476
 - pinocytosis
 - characteristics, 2:197, 2:198f
 - clathrin-independent endocytosis, 2:198
 - macropinocytosis, 2:198f, 2:199
 - proteoglycans, 3:659–660
 - Rho guanosine triphosphatases (Rho GTPases), 4:114
- endoglin, 1:232, 1:232t, 1:235t
- Endo III superfamily, 2:16–17
- endonucleases
 - AP endonucleases, 2:3–4, 2:4f, 2:5t
 - base excision repair (BER), 2:3–4, 2:4f, 2:5t
 - crossover pathways, 2:560
 - DNA mismatch repair (MMR), 2:54–56, 2:55t, 2:56f, 2:57–58, 2:57f
 - flap endonuclease, 2:5–6, 2:91f
 - Nei repair pathway, 2:57–58, 2:57f
 - oxidative damage repair, 2:80
 - type II restriction and modification enzymes, 2:130–135
 - S-adenosyl-L-methionine (AdoMet), 2:130
- characteristics
 - nonspecific DNA interactions, 2:133
 - protein structures, 2:131
 - recognition sequences, 2:131, 2:132t
 - sequence cleavage mechanisms, 2:132–133
- classifications, 2:130
- methyltransferases
 - enzyme characteristics, 2:134
 - modification functions, 2:134
- new systems research, 2:130
- organization
 - expression regulation, 2:131
 - modification and restriction genes, 2:130–131
- research applications, 2:134
- research background, 2:130
- subsets
 - background information, 2:133–134
 - type IIB enzymes, 2:132t, 2:133–134
 - type IIE enzymes, 2:132t, 2:133
 - type IIF enzymes, 2:132t, 2:133
 - type IIS enzymes, 2:132t, 2:133
- endoplasmic reticulum (ER)
 - acyl-coenzyme A synthetase
 - long-chain acyl-CoA synthetase (LCAS), 2:292–293
 - very long-chain acyl-CoA synthetase (VLCAS), 2:293–294
- autophagy, 1:149
- calcium (Ca^{2+}) transport, 1:348
- calcium ion (Ca^{2+}) buffering proteins, 1:290–296
 - background information, 1:290
- Class I binding proteins
 - binding immunoglobulin protein (BiP), 1:291t, 1:292–293
 - calreticulin (CRT), 1:291t, 1:292
 - calsequestrin, 1:291t, 1:294
 - characteristics, 1:291–294
 - ERcalcistorin/PDI, 1:291t, 1:293
 - ERp44, 1:291t, 1:294
 - ERp57, 1:291t, 1:294
 - ERp72, 1:291t, 1:293
 - glucose-regulated protein (Grp94), 1:291t, 1:292
 - P5/PDI, 1:291t, 1:293–294
 - PDI-like proteins, 1:291t, 1:293–294
 - protein disulfide isomerase (PDI), 1:291t, 1:293
- Class I/II binding proteins
 - characteristics, 1:294–295
 - Stim1 protein, 1:291t, 1:295, 3:360–362, 3:361f
 - Stim2 protein, 3:362
 - store operated Ca^{2+} entry (SOCE), 1:294–295
- endoplasmic reticulum (ER) lumen
 - characteristics, 1:291
 - Class I binding proteins, 1:291–294, 1:291t
 - Class II binding proteins, 1:291t, 1:295
 - Class I/II binding proteins, 1:291t, 1:294–295
 - functional role, 1:295–296
- homeostasis control, 1:290
- cholesterol biosynthetic pathway, 1:517–519, 1:518f
- class I α -mannosidases
 - endoplasmic reticulum (ER) α -mannosidase I, 2:201, 3:258, 3:258f, 3:259t
 - ER degradation enhancing α -mannosidase-like protein (EDEM), 2:201, 3:259, 3:259t
- cytoplasmic leaflets, 2:314, 2:316t
- ER-associated degradation (ERAD), 2:200–203
 - action mechanisms, 2:202f
 - basic concepts, 2:200
 - BiP/protein disulfide isomerase (BiP/PDI) system, 2:201–202, 2:202f
 - calnexin/calreticulin cycle, 2:201–202
 - chaperone networks, 2:201–202
 - class I α -mannosidases, 3:259, 3:259t
 - disease connections, 2:201
 - glucosidases, 2:201
 - mannose-binding lectins, 2:202, 2:202f
 - plasma membrane proton (H^{+})-ATPases, 3:690
 - protein synthesis and folding, 2:200–201
 - ubiquitin proteasome pathway (UPP), 2:525, 3:599
- flippase-mediated transbilayer movement, 2:314, 2:315f
- free oligosaccharide (fOS) generation, 2:329, 2:330f
- gluconeogenesis, 2:381
- glycogen synthesis, 2:428–429, 2:430f
- glycoprotein
 - folding processes, 2:448–452
 - basic principles, 2:448–449
 - chaperones, 4:205–206
 - general discussion, 2:448
 - N-glycan processing reactions, 2:200–201, 2:449–450, 2:449f, 2:450f, 4:204–205
 - quality control mechanisms, 2:329–330, 2:451–452, 2:451f, 4:205–206
 - UDP-Glc: glycoprotein glucosyltransferase (UGGT), 2:200–201, 2:449f, 2:450, 2:450f, 2:451f, 2:457, 2:458f, 2:460f, 2:461f
- glycosidase digestion, 2:457–458, 2:459f
- N-linked glycoproteins
 - congenital disorders, 2:470, 2:471t
 - flippase-mediated transbilayer movement, 2:314, 2:315f
 - folding processes, 2:449–450, 2:449f, 2:450f
 - glycosidase digestion, 2:457–458, 2:459f
 - glycosylation processes, 2:457, 2:458f
 - structural remodeling, 2:457, 2:459f, 2:460f, 2:461f
- glycosylation processes
 - N-linked glycosylation
 - congenital disorders, 2:470, 2:471t

- endoplasmic reticulum (ER) (*Continued*)
 free oligosaccharide (fOS) generation, 2:329, 2:330f
 protein synthesis and folding, 2:457, 2:458f, 4:204–205, 4:205f
 O-linked glycosylation, 2:465
 G protein methylation, 3:608–609
 G_q family proteins, 2:498f
 hereditary spastic paraplegias (HSPs), 4:267–270, 4:269f
 inositol trisphosphate (IP₃), 2:583, 2:584f, 2:585f, 2:587f
 insulin secretory defects, 1:652
 intercellular calcium ion (Ca²⁺) waves, 2:614–615, 2:614f
 intracellular calcium ion (Ca²⁺) stores, 1:127–128, 2:574, 2:576f
 IP₃ receptors (IP₃R), 2:654f
 lumen
 calcium ion (Ca²⁺) buffering proteins
 binding immunoglobulin protein (BiP), 1:291t, 1:292–293
 calreticulin (CRT), 1:291t, 1:292
 calsequestrin, 1:291t, 1:294
 characteristics, 1:291
 Class I binding proteins, 1:291–294, 1:291t
 Class II binding proteins, 1:291t, 1:295
 Class I/II binding proteins, 1:291t, 1:294–295
 ERcalcistorin/PDI, 1:291t, 1:293
 ERp44, 1:291t, 1:294
 ERp57, 1:291t, 1:294
 ERp72, 1:291t, 1:293
 functional role, 1:295–296
 glucose-regulated protein (Grp94), 1:291t, 1:292
 P5/PDI, 1:291t, 1:293–294
 PDI-like proteins, 1:291t, 1:293–294
 protein disulfide isomerase (PDI), 1:291t, 1:293
 Stim1 protein, 1:291t, 1:295, 3:360–362, 3:361f
 Stim2 protein, 3:362
 store-operated Ca²⁺ entry (SOCE), 3:359–360, 3:360f, 3:364, 3:365f
 membrane biogenesis-related bilayer expansion, 2:314
 mitochondrial dynamics, 3:137–138, 3:137f
 N-linked glycans processing, 2:201, 3:257
 phosphatidylinositol 4-kinase (PI4K) pathway, 3:459–460
 phospholipase C (PLC), 2:654f, 3:467–468, 3:468f
 protein degradation, 3:622, 3:623t
 protein palmitoylation, 3:645–646
 proteoglycans, 3:659–660, 3:659f
 receptor antagonists, 1:242–243
 sarcoendoplasmic reticulum Ca²⁺ ATPase (SERCA), 2:226–230, 2:231–235
 ATP-binding site, 2:232–233, 2:233f
 calcium ion (Ca²⁺) pump reaction cycle, 4:331, 4:332f
 calcium oscillations, 1:324f
 cardiac excitation–contraction (E–C) coupling, 1:380f, 1:381
 catalytic and transport cycles, 2:226–228, 2:227f
 coupling mechanism
 calcium ion (Ca²⁺) binding and catalytic activation, 2:228
 free-energy use, 2:228
 hydrolytic cleavage, 2:228–229
 nucleotide binding, 2:228
 phosphorylation–calcium ion (Ca²⁺) binding interconversion, 2:229
 phosphoryl transfer, 2:228–229
 substrate-induced conformational fit, 2:228
 domain structures, 3:684f, 4:330–331, 4:330f
 functional role, 4:330–331
 heart failure, 1:381–382, 1:382f
 homeostasis control mechanisms, 1:290, 1:310
 inhibitors, 2:229
 intercellular calcium ion (Ca²⁺) waves, 2:614–615, 2:614f
 intracellular calcium ion (Ca²⁺) stores, 1:127–128
 ion pumping
 E1●2Ca²⁺→E1P transition/occluded state, 2:232f, 2:233–234, 2:233f
 E1P→E2P transition/
 aspartylphosphate hydrolysis/
 luminal gate closure, 2:232f, 2:233f, 2:234–235, 2:234f
 E1P→E2P transition/Ca²⁺ release, 2:232f, 2:233f, 2:234, 2:234f, 3:684f, 3:689–690, 3:689f
 E2→E1→E1●2Ca²⁺ state, 2:232f, 2:233, 2:233f, 4:330f
 physiological considerations, 2:229
 research summary, 2:235, 3:688–689
 ryanodine receptors (RyRs), 4:195
 steady-state behavior, 2:226
 store-operated calcium (SOC) channels, 3:359–360, 3:360f, 4:314–315
 structural characteristics, 2:226–228, 2:227f, 2:231, 2:232f, 4:333f
 transmembrane calcium ion (Ca²⁺) binding sites, 2:231, 2:232f
 secretory pathway, 4:203–209
 compartmental organization, 4:203, 4:204f
 microtubule-based motor proteins, 4:208, 4:208f
 plasma membrane proton (H⁺)-ATPases, 3:690, 3:690f
 protein folding and modification
 chaperones, 4:205–206
 N-linked glycosylation, 4:204–205, 4:205f
 quality control mechanisms, 4:205–206
 targeted proteins
 characteristics, 4:203
 constitutive secretion, 4:208
 entry mechanisms, 4:203, 4:205f
 integral membrane proteins, 4:204, 4:205f
 protein exit, 4:207
 regulated secretion, 4:208
 signal recognition particles (SRPs), 4:203, 4:205f
 soluble proteins, 4:204, 4:205f
 transport mechanisms
 protein localization, 4:206–207, 4:206f
 retinoid X receptor (RXR), 4:207
 vesicular transport, 4:206, 4:206f
 yeasts, 3:690, 3:690f
 store-operated calcium (SOC) channels, 4:314–315, 4:318f
 stromal interacting molecule 1 (STIM1), 1:129–130
 T-cell receptors (TCRs), 4:373
 unfolded protein response (UPR), 4:488–493
 basic concepts, 4:488
 chaperones, 4:488, 4:492f
 ER-associated degradation (ERAD), 2:524f, 2:525
 functional roles, 4:488
 health monitoring, 2:524–525, 2:524f
 metazoans
 ATF6 gene activation, 4:490–491, 4:491f
 inositol requiring 1 (IRE1) gene conservation, 4:490
 intramembrane proteolysis, 4:490–491
 physiological and pathological stresses
 ATF6 gene activation, 2:524–525, 2:524f, 4:490–491, 4:491f
 basic concepts, 4:488–489
 gene expression regulation, 4:491–492
 HAC1 gene expression, 2:524–525, 2:524f, 4:489, 4:490f
 inositol requiring 1 (IRE1) genes, 2:524–525, 2:524f, 4:489, 4:490f
 intramembrane proteolysis, 4:490–491, 4:491f
 stress sensors, 4:492
 translation initiation factor 2 (eIF2 α) phosphorylation, 4:491, 4:492f
 XBP-1 gene expression, 2:524–525, 2:524f
 translational control
 gene expression regulation, 4:491–492
 PERK (PKR-like ER kinase), 2:524–525, 2:524f, 4:491, 4:492f
 stress sensors, 4:492
 translation initiation factor 2 (eIF2 α) phosphorylation, 2:524–525, 2:524f, 4:491, 4:492f
 yeasts
 HAC1 gene expression, 4:489, 4:490f
 inositol requiring 1 (IRE1) genes, 4:489, 4:490f
 endoribonucleases
 ribonuclease E (RNase E), 3:52–53, 3:53t, 3:54f, 3:56–57, 3:57f

- transfer RNA (tRNA), 3:555, 3:555f
 See also ribosomal RNA (rRNA); transfer RNA (tRNA)
- endosomal sorting complexes required for transport (ESCRT) pathway, 1:624–625, 4:504–505, 4:505f, 4:506f
- endostatin, 1:444
- endothelial cell-selective adhesion molecules (ESAMs), 4:392, 4:394f
- endothelial differentiation gene (EDG) family, 2:762
- endothelin, 2:344t, 2:345f, 2:496–497, 2:497t
- energy homeostasis, 3:76–79
 body weight regulation
 adiposity and satiety signal integration
 cholecystokinin (CCK), 3:78–79, 3:79f
 general discussion, 3:79
 ghrelin, 3:79
 schematic diagram, 3:79f
 arcuate nucleus (ARC)
 adiposity and satiety signal integration, 3:79f
 Agouti-related protein (AgRP), 3:77, 3:77f, 3:78f
 functional role, 3:76
 melanocortin neurons, 3:77
 neuropeptide Y (NPY), 3:77, 3:77f, 3:78f
 background information, 3:76
 hypothalamic neuronal systems
 arcuate nucleus (ARC), 3:76
 lateral hypothalamic area (LHA), 3:78
 paraventricular nucleus (PVN), 3:78, 3:79f
 ventromedial hypothalamus (VMH), 3:77–78, 3:78f
 malonyl-CoA, 4:56–57
 model diagram, 3:77f
 research summary, 3:79
 historical research, 3:76
- energy metabolism
 microRNA (miRNA) transcription regulation, 4:191
- myocardial ischemia
 basic principles, 3:67, 3:68f
 ischemic-induced metabolic signaling alterations, 3:67–69, 3:69f
 metabolic-based treatment approaches, 3:71, 3:71t
 reperfused ischemic heart
 basic principles, 3:69–70, 3:70f
 intracellular acidosis/cardiac efficiency, 3:70–71
 research summary, 3:71
- pyruvate dehydrogenases (PDHs), 4:321–328
 active site coupling, 4:323
 characteristics, 4:321
 components, 4:321–322, 4:322f
 cryo-electron microscopic architectural analyses, 4:326–327, 4:327f
 lipoyl domains, 4:323, 4:325f, 4:327f
 reaction mechanisms
 E1 enzyme reaction, 4:322, 4:322f, 4:325f, 4:327f
 E2 enzyme reaction, 4:322–323, 4:322f, 4:327f
 E3 enzyme reaction, 4:322f, 4:323, 4:327f
 general discussion, 4:322–323
 regulation mechanisms
 basic principles, 4:323–326
 biological significance, 4:323–324
 PDH kinase (PDK) isoforms, 4:324–326, 4:325f, 4:326f
 small metabolite regulation, 4:324
 research summary, 4:328
 substrate channeling, 4:323
- enoyl-CoA hydratase (EH), 1:385f, 1:386, 1:386–387
- Ensembl and Ensembl Genomes, 3:615t
- entactin, 3:91t
- Entamoeba histolytica*, 1:592, 2:702t
- Enterobacter aerogenes*, 4:73–74
- Enterococcus faecalis*
 cAD1 precursor pheromone, 4:51f, 4:52–53
 Cho phylogenetic tree, 3:331f
 DNA ligases, 2:35f
uvrC phylogenetic tree, 3:331f
- Enterococcus hirae*, 3:671–672
- enterostatin, 1:118
- ENTH domain, 3:467–468
- Entner–Doudoroff pathway, 2:204–206, 2:205t
- Entomopoxvirinae*, 1:319t
- Entrez database, 3:615t
- enzyme-catalyzed reactions, 4:347–349
 basic concepts, 4:347
 free-energy profiles, 4:348, 4:348f, 4:349f
 Haldane relationship, 4:347
 process mechanisms, 4:347–348
 rate-limiting steps
 Michaelis constant (V/K), 4:348–349
 velocity (V), 4:349
- Enzyme Commission (EC), 3:614
- ENZYME Database, 3:615t
- enzyme I (EI), 2:387, 2:388f
- enzyme IIA^{Clu}, 2:387, 2:388f
- enzyme II (EII), 2:387, 2:388f
- enzyme inhibitors, 2:210–215
 mechanism-based inhibitors
 anti-metabolite compounds, 2:215
 characteristics, 2:215
 ornithine decarboxylase inhibition, 2:214f
 polyamine synthesis, 2:215
 noncatalytic site inhibitors
 catalytic site binding prevention, 2:212–213
 characteristics, 2:213–215
 conformational change prevention, 2:213–215
 subunit formation prevention, 2:213
- reversible catalytic-site inhibitors
 bisubstrate mimics, 2:210, 2:212f
 reactant and product analogs, 2:210, 2:211f, 2:212f
 transition-state analogs
 characteristics, 2:210–211
 design considerations, 2:211, 2:213f
 enzymatic transition states, 2:211, 2:213f
 natural product transition-state analogs, 2:211, 2:214f
- enzyme kinetics, 2:216–220
 multireactant enzymes
 Cleland's slope rules, 2:220
 intercept rules, 2:219f, 2:220
 kinetic mechanisms, 2:218, 2:219f
 velocity equations, 2:218–220, 2:220f
- unireactant enzymes
 Briggs–Haldane equation, 2:217
 Henri's equation, 2:216–218
 historical research, 2:216–218
 k_{cat}/K_m ratio, 2:217
 Michaelis–Menten equation, 2:216–217
 rapid equilibrium assumption, 2:216–217
 reversible reactions, 2:217–218
 steady-state assumption, 2:217
 velocity curve plot, 2:218, 2:218f
- enzyme-linked immunosorbent assay (ELISA)
 fatty acid synthase expression, 2:277–278
 phage-binding experiments, 3:433–435, 3:435f
- eosinophil peroxidase (EPO), 3:407–409, 3:408f
- eosinophils, 1:188f, 1:189
- epidermal growth factor (EGF)
 angiotensin receptors, 1:113
 blood-clotting proteases, 3:588–589, 3:589f
 endocytosis, 2:199
 epidermal growth factor (EGF) receptors, 1:637, 3:476
 O-fucose, 2:453, 2:454f
 O-glucose, 2:453–454, 2:454f
 G protein-coupled receptor kinases (GRKs), 2:485t
 integrin signaling, 2:611
 macroautophagy regulation, 1:150t
 mucin domains, 3:201f, 3:202
 mucins, 3:206
 neurokinin-1 receptor (NK1R), 4:362–363
 nuclear factor kappa B (NF- κ B), 3:303t
 proteolysis, 1:444
 rhomboid-based mediation, 4:54
- epidermolysis bullosa acquisita, 1:641t
- epidermolysis bullosa simplex (EBS), 2:627–628, 2:627t, 2:670t
- epidermolysis bullosa with muscular dystrophy (EB-MD), 1:641t
- epidermolytic hyperkeratosis, 2:670t
- epigallocatechin gallate (EGCG), 2:167–168, 2:167f
- epigenetic centromeres, 1:448
- epigenetics
 methyl-CpG-binding proteins, 1:521–524
 binding domain family
 general discussion, 1:522
 MBD1, 1:523, 1:523f
 MBD2, 1:523, 1:523f
 MBD3, 1:523, 1:523f

- epigenetics (*Continued*)
- MBD4, 1:523–524, 1:523f, 4:77
 - MBD5, 1:523f, 1:524
 - MBD6, 1:523f, 1:524
 - MeCP2, 1:522–523, 1:523f
 - cytosine methylation
 - basic concepts, 1:522
 - cancer cells, 1:522
 - functional role, 1:522
 - gene expression regulation, 1:522
 - genomic imprinting, 1:522
 - X-chromosome inactivation, 1:522
 - DNA methylation–chromatin structure
 - interdependency, 1:524
 - functional role, 1:521
 - gene expression regulation, 1:521
 - epilepsy, 4:569, 4:569t
 - epilysin, 3:91t, 3:92f, 3:93
 - epimerase, 3:499f
 - epinephrine
 - adenyl cyclases (ACs), 1:46t
 - branched-chain amino acid (BCAA)
 - regulation, 1:245–246
 - catecholamine-derived *o*-benzoquinones, 3:727, 3:727f
 - characteristics and functional role, 1:57
 - G_q family proteins, 2:497t
 - second messenger pathways, 1:43f
 - starvation effects, 4:303–304
 - epithelial–mesenchymal transition (EMT), 1:259–260, 2:304–306
 - epoxygenase pathway, 2:189
 - Epstein–Barr virus
 - Epstein–Barr virus (EBV) transcription factor EBNA1, 2:110, 3:623t
 - nuclear factor kappa B (NF- κ B), 3:303t, 3:304–305
 - RNA polymerase III (Pol III), 4:161
 - Epstein syndrome, 3:223t
 - ErbB-4, 4:51f, 4:53
 - ERC-55 protein, 1:291t, 1:295
 - ERcalcistatin/PDI, 1:291t, 1:293
 - ERCC1–XPF complex, 3:342f, 3:342t, 3:343–344, 3:343f
 - ERGIC-53, 2:703f, 2:704
 - ergocalciferol (vitamin D₂), 4:535, 4:536f
 - ERp44, 1:291t, 1:294
 - ERp57, 1:291t, 1:294
 - ERp72, 1:291t, 1:293
 - Erwinia*, 2:43
 - Erythrina corallodendron*, 2:702t
 - erythritol, 2:204–206
 - erythropoiesis, 1:187, 1:188f
 - erythropoietic protoporphyria, 2:542, 2:542t
 - erythropoietin (Epo)
 - c-fes* proto-oncogene (Fes), 1:452
 - expression regulation, 2:527
 - signaling pathways, 2:527–528, 2:528f
 - structural characteristics, 2:526–527, 2:527f
 - von Hippel–Lindau protein (VHL), 4:570
 - erythrose-4-phosphate (E4P)
 - pentose phosphate pathway
 - carbon fate, 3:394f
 - chemical balance, 3:394f
 - metabolic reactions, 3:392
 - nonoxidative reactions, 3:391, 3:393f
 - photosynthetic metabolic pathways, 3:499f
 - Escherichia coli*
 - adenosine diphosphate ribosylation factors (ARF), 1:134–135
 - alkyltransferase-like (ATL) proteins, 2:14
 - alkyltransferases
 - Ada/N-Ada, 2:10t, 2:12, 2:12f
 - AidB response, 2:14
 - AlkB family, 2:13–14, 2:13f
 - C-Ada, 2:10t, 2:12f
 - Ogt, 2:10t, 2:12–13, 2:12f
 - amyloid deposits, 1:102
 - AP endonucleases, 2:3–4
 - base excision repair (BER), 2:3–4, 2:5t
 - biotin
 - biotin carboxyl carrier protein (BCCP), 1:222, 1:223f
 - biotin-inducible repressor (BirA), 1:224, 1:224f
 - domain structure, 1:223–224, 1:223f
 - sequence alignment, 1:223f
 - calcineurin, 1:337
 - calpains, 1:354f
 - chemotaxis receptor fibers, 3:607–608
 - Cho phylogenetic tree, 3:331f
 - chromosome replication, 2:154
 - clustered regularly interspaced short palindromic repeat (CRISPR) RNAs (crRNAs), 4:253, 4:254f
 - coenzyme Q, 3:723t
 - Dam methyltransferase, 2:40–44
 - altered gene expression, 2:42–43
 - bacterial pathogenesis, 2:43
 - chromosome replication, 2:42
 - dam* mutant strains, 2:40
 - gene inactivation, 2:43–44
 - mismatch repair, 2:40–42, 2:41f, 2:42f, 2:54
 - Mut protein system, 2:40–42, 2:41f, 2:42f
 - oriC* region, 2:42
 - post-transcriptional regulation, 2:43
 - properties, 2:40
 - viruses, 2:43–44
- dicarboxylic acid decarboxylation, 2:208
- DNA glycosylases
 - 3-methyladenine (3mA) DNA glycolase
 - I, 2:10–11, 2:11f
 - 3-methyladenine (3mA) DNA glycolase
 - II, 2:11, 2:11f
- DNA gyrase, 2:165f
- DNA ligases, 2:35–36, 2:36f, 2:37f
- DNA polymerase I, 2:87–91
 - domain structures, 2:87–88, 2:88f
 - 5' exonuclease, 2:90–91, 2:91f
 - O-helix, 2:88–89, 2:88f
 - phosphoryl transfer reactions, 2:89, 2:89f
 - reaction mechanisms, 2:89, 2:89f
 - reaction specificity
 - base pairing, 2:89–90
 - sugar specificity, 2:90
 - research background, 2:87
 - structural domains, 2:87
 - substrate binding, 2:88–89, 2:88f
- 3'–5' exonuclease, 2:90
- DNA replication, 2:114–117
 - DnaA functional domains, 2:121–122, 2:122t
 - Escherichia coli* β -clamp, 2:94f, 2:114–115, 2:116t, 3:576, 3:577f, 4:236–237, 4:237f, 4:238f
 - Escherichia coli* γ complex, 3:576, 3:577f
 - eukaryotic DNA replication, 2:108–109
 - leading- and lagging-strand DNA
 - synthesis, 2:92, 2:93f, 2:94f, 2:114, 2:115f
 - oriC* origins-based replication, 2:121–125
 - biochemical mechanisms, 2:121–122
 - characteristics, 2:117
 - DnaA–*oriC* complex formation, 2:122–123
 - DnaB–DnaC complex, 2:124
 - DnaB helicase, 2:124
 - DnaC helicase, 2:123–124
 - DNA sequence motifs, 2:121–122, 2:122f
 - factor for inversion stimulation (Fis), 2:122, 2:122f
 - integration host factor (IHF) binding sites, 2:122, 2:122f
 - loading process, 2:124
 - primer foundation, 2:124
 - replisome assembly process, 2:122–123
 - research summary, 2:124–125
 - protein components
 - DNA ligases, 2:115f, 2:116, 2:116t
 - DNA polymerase holoenzymes, 2:92–95, 2:114–115, 2:115f, 2:116t
 - 5'–3' nuclease proteins, 2:115f, 2:116, 2:116t
 - primosomes, 2:115–116, 2:115f, 2:116t
 - requirements, 2:114–116
 - single-strand DNA (ssDNA)-binding proteins, 2:93–94, 2:94f, 2:115f, 2:116, 2:116t
 - replication fork (RF)
 - DNA polymerase holoenzymes, 2:92–95
 - lagging-strand DNA synthesis, 2:92, 2:93f, 2:94f, 2:115f, 2:116–117
 - leading-strand DNA synthesis, 2:92, 2:93f, 2:94f, 2:116, 2:117
 - replisome assembly process
 - characteristics, 2:122–123
 - DiaA protein stabilization, 2:123
 - DnaA-dependent unwinding, 2:123
 - DnaA–*oriC* complex formation, 2:122–123
 - HU protein stabilization, 2:123
 - trombone replication model, 2:114, 2:115f
- DNA topoisomerases, 3:368–369, 3:369t
- Escherichia coli* β -clamp, 3:577, 3:578f, 4:236–237, 4:237f, 4:238f
- extra-chromosomally-primed (EP)
 - retrotransposons, 3:285f

- fermentation reactions, 2:205*t*
 ferredoxin-NADP⁺ reductase (FNR), 2:301
 fumarate respiration, 2:207*f*, 2:208*f*
 GroEL (growth E locus, large gene, bacterial group I chaperonin), 1:457–458, 1:458*f*, 1:459
 heme biosynthesis, 3:547
 heme-copper oxidases, 1:604*t*
 LacI-lamin protein, 3:315, 3:316*f*
lac operon
 catabolite repressor protein (CRP)
 functional role, 2:698–699
 historical research, 2:695
 schematic diagram, 2:695*f*
 structural characteristics, 2:697*f*, 2:699
 characteristics, 2:695–696
 historical research, 2:694–695
 lactose repressor protein (LacI)
 functional role, 2:696–697
 looped structures, 2:698, 2:698*f*
 pseudo-operator sequences, 2:698, 2:698*f*
 structural characteristics, 2:697–698, 2:697*f*
 research errors, 2:699
 schematic diagram, 2:695*f*
 transcription-coupled repair (TCR), 4:402
 lactose permease (LacY), 2:389*f*
 lectins, 2:702*f*, 2:702*t*
 LexA regulatory protein
 gene expression regulation, 2:706
 regulon characteristics, 2:706–707, 2:710
 repression mechanisms, 2:707–708, 2:707*f*
 SOS response, 2:706
 structural characteristics, 2:707*f*, 2:708*f*
 maltose-binding protein (MalBP), 1:62, 2:389–390
 membrane-associated energy transduction
 ferric iron reduction, 3:30
 oxygen respiration, 3:30–31
 messenger RNA (mRNA) decay
 degradosome complex, 3:53
 poly(A) polymerase (PAP)-mediated decay, 3:54*f*, 3:55
 ribonuclease E (RNase E), 3:52–53, 3:53*t*, 3:54*f*, 3:56–57, 3:57*f*
 ribonuclease G (RNase G), 3:53
 metalloproteases, 3:87–88
 mismatch repair systems (MRSs)
 anti-recombination activity, 2:47–48, 2:48*f*
 components, 2:42
 GO system (MutM–MutT–MutY), 2:80, 2:80*f*
 long-patch mismatch repair
 Dam methyltransferase, 2:54
 DNA polymerase III holoenzyme, 2:95
 drug resistance, 2:57
 excision and repair synthesis, 2:56
 initiation phase, 2:55–56
 mismatch repair deficiency, 2:57
 MutH pathway, 2:56, 2:56*f*
 MutL pathway, 2:55–56, 2:56*f*
 MutS pathway, 2:55, 2:56*f*
 process mechanisms, 2:54–56, 2:56*f*
 repair enzymes, 2:54–56, 2:55*t*, 2:56*f*
 specificity, 2:54
 methyl-directed MRS, 2:45–46
 Mut protein family, 2:66–68, 2:67*f*
 MutS–MutY interactions, 2:58
 MutY repair pathway, 2:57–58, 2:57*f*
 specificity, 2:54
 very-short patch (VSP) repair
 research background, 4:76
 vsr gene biochemistry, 4:76–77
 vsr gene mutagenesis, 4:77
 vsr gene structure and regulation, 4:76, 4:77*f*
 mitochondrial-type ferredoxins, 2:297
 multidrug resistance (MDR) membrane proteins, 3:214
 nickel-iron (NiFe) hydrogenases, 4:72–73
 nicotinamide adenine dinucleotide transhydrogenase, 3:247, 3:248*f*, 3:250, 3:250*f*
 non-homologous end joining (NHEJ), 3:266
 nucleoid-associated proteins (NAPs)
 characteristics and functional role, 3:367, 3:368*f*, 3:368*t*
 DNA supercoiling, 3:368–369, 3:369*f*
 gene expression regulation, 3:367–368
 topoisomerases, 3:369–370
 phage display, 3:431–436
 display vectors
 basic concepts, 3:431–432
 phagemid vectors, 3:432, 3:432*f*
 phage vectors, 3:432, 3:432*f*
 polyvalent/monovalent displays, 3:431–432
 phage-binding analysis, 3:435, 3:435*f*
 phage-binding selections
 affinity-based selection, 3:433–435, 3:435*f*
 binding and elution, 3:433–435
 preparation methods, 3:433
 phage libraries
 characteristics, 3:432–433
 diversity library design, 3:433, 3:434*t*
 sequence generation, 3:432–433, 3:433*t*, 3:434*t*
 research applications, 3:431
 phosphotransferase system (PTS), 2:387, 2:388*f*
 protein N-myristoylation, 3:642
 pterin distribution, 3:666
 pyruvate dehydrogenases (PDHs), 1:206–207, 4:326
 pyruvate kinase (PK), 3:719, 3:720*f*
 recombination-dependent DNA replication (RDR), 4:24
 RecQ helicases, 4:43–44, 4:44*f*
 reductive reactions, 1:487*t*
 Rep helicases
 characteristics and functional role, 3:261
 DNA unwinding
 active, rolling or hand-over-hand model, 3:264–265, 3:265*f*
 active versus passive mechanisms, 3:264
 dimeric inch-worm models, 3:265, 3:265*f*
 inch-worm mechanisms, 3:264, 3:264*f*, 3:265*f*
 protein oligomerization, 3:263
 unwinding mechanisms, 3:263–264
 research summary, 3:265
 single-strand DNA (ssDNA)
 translocation, 3:263, 3:264, 3:264*f*
 structural characteristics
 crystal structures, 3:262, 3:262*f*
 domain structures, 3:263*f*
 open and closed confirmations, 3:262–263
 primary structures, 3:261–262
 translocase-based protein displacement, 3:265
 respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:87–91
 basic concepts, 4:87, 4:341
 disease connections
 general discussion, 4:90–91
 mitochondrial mutations, 4:90–91
 neurotoxins, 4:91
 nuclear mutations, 4:91
 pathological mechanisms, 4:91
 evolutionary development
 general discussion, 4:88–90
 N-module, 4:89
 origins, 4:89–90
 P-module, 4:89
 Q-module, 4:88–89
 variants, 4:89–90
 functional role
 active–inactive transition, 4:90
 electron transfer mechanisms, 4:90
 general discussion, 4:90
 inhibitors, 4:90
 proton pumping, 4:90
 redox centers, 4:90
 sodium ion (Na⁺) pumping, 4:90
 structural characteristics
 accessory subunits, 4:88, 4:89*t*
 central subunits, 4:87–88, 4:88*f*, 4:89*t*
 cysteine motif, 4:344–345, 4:345*f*
 flavin mononucleotide (FMN)
 binding, 4:344–345, 4:344*f*
 frataxin-like subunit, 4:343–344, 4:343*f*
 hydrophilic domain, 4:342–343, 4:342*f*
 iron-sulfur clusters, 4:87–88, 4:89*t*, 4:342–343, 4:342*f*, 4:345*f*
 multistage assembly pathway, 4:88
 mutations, 4:345
 NADH binding, 4:344–345, 4:344*f*, 4:345*f*
 overall structure, 4:341–342
 redox centers, 4:87–88, 4:89*t*, 4:342–343, 4:342*f*
 research summary, 4:345–346
 shape, 4:87, 4:88*f*
 subunits, 4:342*t*
 supernumerary subunits, 4:345

- Escherichia coli* (Continued)
- retrotransposons, 3:283
 - ribosomal RNA (rRNA)
 - defective processing consequences, 3:559
 - maturation processes
 - basic concepts, 3:556
 - 5S rRNA, 3:557f, 3:558, 3:558f
 - schematic diagram, 3:557f
 - 16S rRNA, 3:557, 3:557f
 - transcript structure, 3:556, 3:557f
 - 23S rRNA, 3:557–558, 3:557f
 - riboswitches, 4:137–138
 - RNA polymerases, 4:167, 4:168f
 - RseA, 4:51f
 - selenocysteine biosynthesis, 4:210–212, 4:211f
 - septin complexes, 4:214–215
 - sigma(σ)-factors, 4:167–169, 4:233–234, 4:233t
 - succinate:quinone oxidoreductases (SQORs), 4:92–93, 4:94f
 - superoxide dismutases (SODs), 4:353
 - T7 RNA polymerase, 4:358–359
 - thiamine biosynthesis, 1:203–204
 - threonyl-tRNA synthetase, 4:443
 - transcription-coupled repair (TCR), 4:403–404
 - transfer RNA (tRNA)
 - defective processing consequences, 3:559
 - 5'-end–RNase P process pathway, 3:555, 3:555f, 3:558f
 - maturation processes, 3:554, 3:555f
 - structural characteristics, 3:554–555
 - 3'-end–RNase process pathway, 3:555–556, 3:555f
 - translocation cycle, 2:252
 - TRiC/CCT eukaryotic chaperonin, 1:459
 - trp* operon
 - control mechanisms, 2:360–362, 2:361f
 - research background, 2:360
 - transcription regulation, 2:360–362
 - tryptophan biosynthesis pathway, 4:249b
 - type I DNA restriction and modification enzymes, 2:72–74
 - type I DNA topoisomerase
 - characteristics, 3:369t
 - cleavage/strand-passage domain crystal structure, 2:158–159, 2:159f
 - protein domains, 2:158, 2:159f
 - type II DNA restriction and modification enzymes, 2:132t
 - type IIA DNA topoisomerase, 2:165f
 - type III DNA topoisomerase, 2:159f, 3:369t
 - type IV DNA topoisomerase, 2:165f
 - UmuDC protein/*umuDC* genes, 4:478
 - uvrC* phylogenetic tree, 3:331f
 - uvr* gene research, 3:328–330
- ESCRT protein complexes, 4:504–505, 4:505f, 4:506f
- esmolol, 1:59
- esophagus
 - digestion and absorption, 2:355
 - esophageal cancer, 2:305
- espin, 1:34
- essential amino acids, 2:355f
- esterases, 2:759
- estradiol, 1:286–287, 1:288, 4:308–309, 4:309f
- estrogen receptors
 - characteristics and functional role, 4:308–309
 - estrogen response element (ERE), 4:311f
 - mucins, 3:206–208, 3:206f
 - selective estrogen receptor modulators (SERMs), 4:311–312
- ETF:ubiquinone oxidoreductase (ETF–QOR), 4:19f, 4:20
- ethanol, 2:204–206, 2:205t, 2:206f
- ethanolamine, 3:479f, 3:480
- ethanolamine ammonialyase, 1:158t
- ethenocytosine, 4:77f, 4:78
- ether lipid synthesis, 2:417–418
- ethnicity-dependent obesity, 3:657t
- ethylenediamine tetraacetic acid (EDTA), 3:89
- etioplasts, 1:508–509, 1:509f
- etoposide
 - anticancer drugs, 2:167, 2:167f
 - multidrug resistance (MDR) membrane proteins, 3:212f
- etravirine, 2:549
- eubacteria
 - chaperonins, 1:456–457
 - eubacterial glutamyl-tRNA(Gln) amidotransferase, subunit C homolog (GATC) methyltransferases, 2:40–44
 - altered gene expression, 2:42–43
 - chromosome replication, 2:42
 - dam* mutant strains, 2:40
 - mismatch repair, 2:40–42, 2:41f, 2:42f
 - properties, 2:40
 - hyperthermophilic eubacteria, 2:161–162
 - mesophilic/psychrophilic eubacteria, 2:161
 - proteasomes, 3:590
 - ribozyme distribution, 4:146f
 - type I DNA topoisomerase, 2:158t, 2:160f
- euchromatin, 1:521
- eukaryotes
 - actin-related proteins (Arps)
 - Arp1 protein, 1:39, 1:39f
 - Arp4 protein, 1:40
 - Arp5 protein, 1:40
 - Arp6 protein, 1:40
 - Arp7 protein, 1:40
 - Arp8 protein, 1:40
 - Arp9 protein, 1:40
 - Arp2/3 complex, 1:39–40, 1:39f
 - chromatin-associated Arps, 1:40
 - chromatin-associated conventional actin, 1:40
 - classifications, 1:36, 1:37f, 1:38t
 - conventional actin, 1:38–40, 1:39f
 - cytoskeletal Arps, 1:38–40
 - autophagy, 1:153
 - chaperonins, 1:456–457
 - chemotaxis, 2:236–240
 - characteristics and functional role, 2:236
 - cytoskeletal polarization, 2:238
 - mathematical models, 2:238–240, 2:238f
 - metrics, 2:239b
 - signaling pathways, 2:236–238, 2:237f
- cytochrome *c*, 1:599
- DNA mismatch repair (MMR), 2:46–47
- DNA replication, 2:108–113
 - background information, 2:108
 - DNA ligases, 2:31
 - origin recognition complex (ORC), 2:108
 - conservation, 2:111
 - discovery, 2:111
 - functional role, 2:111–112
 - non-replicative functions, 2:112–113
 - pre-recombination complex
 - formation, 2:111–112, 2:112f
 - regulation mechanisms, 2:112
 - specification mechanisms, 2:110
- origins
 - lower eukaryotes, 2:108–109
 - metazoan origins, 2:109–110, 2:109f
 - origin plasticity, 2:110–111
 - origin recognition complex (ORC), 2:110
 - prokaryotes, 2:108–109, 2:109f
 - replicon model, 2:108
 - research methodologies, 2:111
 - schematic diagram, 2:109f
 - specification mechanisms, 2:110
 - viruses, 2:108–109
- polymerase α -primase (Pol α), 2:241–244
 - background information, 2:108, 2:241
 - cell-cycle-dependent activity, 2:242
 - checkpoint pathways, 2:243–244
 - functional role, 2:241–242, 2:242f
 - polymerase α -primase (Pol α)–helicase relationship, 2:242–243, 2:242f
 - structural characteristics, 2:241–242
- recombination-dependent DNA replication (RDR), 4:24–25
- replication fork (RF), 2:118–120
 - basic concepts, 2:118, 2:119f
 - future research areas, 2:120
 - mini-chromosome maintenance (MCM) proteins, 2:118, 2:119f
 - polymerase accessory proteins, 2:119
 - polymerase α -primase (Pol α), 2:119
 - polymerase δ (Pol δ), 2:120
 - polymerase ϵ , 2:120
 - proliferating cell nuclear antigen (PCNA), 1:402t, 2:119
 - replication factor C (RFC), 2:120
 - replication protein A (RPA), 1:402t, 2:118–119, 2:119f
 - sliding clamps, 2:119
 - topoisomerases, 2:120
 - research summary, 2:113
- glycosylphosphatidylinositol (GPI)-anchored proteins, 2:476–478
- biosynthesis
 - attachment mechanisms, 2:476–477
 - biosynthetic pathways, 2:476, 2:477f

- characteristics, 2:476
functional roles, 2:478
mammalian cells, 2:477–478
parasitic protozoa, 2:477–478
structural characteristics, 2:476, 2:477f
yeasts, 2:477–478
- hexokinases, 2:544, 2:545f
- in vivo* eukaryotic ribosome biogenesis
assembly factors, 4:118–119, 4:120f
neighborhood associations, 4:120
pre-ribosomal intermediates,
4:118–119, 4:119f, 4:120f
ribosomal proteins, 4:120
ribosomal RNA (rRNA) synthesis and
processing, 4:117–118, 4:118f
structural proofreading, 4:119, 4:120f
subunit quality control mechanisms,
4:120–121
- iron–sulfur (FeS) protein biogenesis
chloroplasts, 2:663
cytosolic system, 2:663–664
mitochondrial Fe–S cluster assembly
machinery, 2:663, 2:663f
- microRNA (miRNA), 3:98–102
biogenesis, 3:99–100, 3:99f
biological role, 3:101–102
characteristics, 3:98
functional role, 3:100, 3:101f
historical research, 3:98–99
RNA-induced silencing complex (RISC),
3:98, 3:99f, 3:101f
- multidrug resistance (MDR) membrane
proteins, 3:214
- non-homologous end joining (NHEJ),
3:269–274
accessory proteins, 3:272
basic concepts, 3:269–271, 3:270f
disease connections, 3:273, 3:273t
DNA-dependent protein kinase catalytic
subunit (DNA-PK_{cs}), 3:269–271,
3:270f, 3:272f
enzymatic functions, 3:269–271, 3:270f
inhibitors, 3:273–274
Ku70/80 protein, 3:269–271, 3:270f,
3:272f, 3:273t
telomere maintenance, 3:272,
4:384–385
therapeutic interventions, 3:273–274
V(D)J recombination, 3:272, 3:272f
- nucleotide excision repair (NER),
3:341–344
biological importance, 3:341
damage recognition, 3:341–342, 3:342t
damage verification, 3:342–343, 3:342t
global genome repair (GGR),
3:341–342, 3:342f
incision and synthesis processes, 3:342t,
3:343–344, 3:343f
NER factors, 3:342t
open complex formation, 3:342–343,
3:342t
repair pathways, 3:342f
transcription-coupled repair (TCR)
pathway, 3:337, 3:342f, 3:342t,
3:344
- polymerase α -primase (Pol α), 2:241–244
background information, 2:108, 2:241
cell-cycle-dependent activity, 2:242
characteristics, 2:119
checkpoint pathways, 2:243–244
functional role, 2:241–242, 2:242f
polymerase α -primase (Pol α)–helicase
relationship, 2:242–243, 2:242f
structural characteristics, 2:241–242
- proteasomes
basic concepts, 3:590
Hsp90 proteasome inhibitor, 3:593,
3:593f
inhibitors, 3:593, 3:593f
P131 proteasome inhibitor, 3:593
proteasome activators
characteristics, 3:592–593
ecm29p, 3:592–593
11S regulators (REG α , REG β , REG γ)/
PA28s, 3:592, 3:593f
PA200, 3:592, 3:593f
schematic diagram, 3:593f
research summary, 3:593
- 20S proteasome
assembly processes, 3:596–597
core particle (CP) gate, 3:597
enzyme mechanism and inhibitors,
3:590
immunoproteasome, 3:596
N-terminal active site, 3:596, 3:596f
occurrences, 3:595–597
structural characteristics, 3:590,
3:591f, 3:595–596, 3:596f
tymoproteasome, 3:596
- 26S proteasome
basic concepts, 3:590–592, 3:591f
cell regulation, 3:599
enzyme mechanism, 3:591, 3:592f
19S regulatory complex, 3:590–592,
3:597–598, 3:597f
physiological importance, 3:592
structural characteristics, 3:597f
tissue and subcellular distribution,
3:591–592
- protein biosynthesis elongation,
2:249–255
aminoacyl (A) sites, 2:249–251, 2:250f,
2:251f, 2:253f, 2:254f, 4:122–123
background information, 2:249,
4:122–123
components, 2:250t
eEF1A elongation factor, 2:249–251,
2:250f, 2:250t, 2:252, 2:253–254,
2:253f, 2:254f
eEF2 elongation factor, 2:252, 2:254f
eEF3 elongation factor, 2:253
eEFSec elongation factor, 2:253
elongation cycle chemistry, 2:251–252,
2:251f
elongation factor G (EF-G), 4:122–127
EF-G–ribosome interactions, 4:123f,
4:125–126, 4:126f, 4:128–129,
4:129f
research summary, 4:126
ribosome recycling, 4:126
translocation cycle, 4:125, 4:132–133
elongation factor Tu (EF-Tu), 4:122–127
EF-Tu–ribosome interactions, 4:123f,
4:128–129, 4:129f
guanosine triphosphate (GTP)
hydrolysis, 4:123–125, 4:124f
initial selection phase, 4:123, 4:124f
research summary, 4:126
substrate proofreading, 4:125
entry (E) sites, 2:252, 2:253f
exit (E) sites, 2:252, 2:253f, 4:122–123
molecular mimicry, 2:252
Nobel-winning research, 2:254
peptidyl (P) sites, 2:249–251, 2:250f,
2:251f, 2:253f, 2:254f, 4:122–123
regulation mechanisms, 2:253–254
structural model, 2:254f
- protein-coding genes, 1:75
reactive oxygen species (ROS), 2:77
retrotransposons, 3:283, 3:284f
ribosomal structure, 4:135, 4:135f
ribozyme distribution, 4:146f
RNA polymerase I (Pol I), 2:366t,
2:367–368, 4:159–161
Alu genes, 4:160
disease connections, 4:161
gene transcription regulation, 4:159
molecular characteristics, 4:159
nucleoli, 4:159
retrotransposition, 4:160
transcription apparatus, 4:160, 4:160f
tumor suppression, 4:160–161
- RNA polymerase II (Pol II)
conservation considerations, 2:367–368
C-terminal repeat domain (CTD)
characteristics, 2:365
initiation phase, 2:366, 3:63–64, 3:63f
transcription cycle, 2:365, 3:63–64,
3:63f
transcription termination, 2:367,
3:63–64, 3:63f
- elongation control, 4:164–166
basic concepts, 4:165, 4:165f
carboxyl terminal domain (CTD),
4:165
cyclin subunits, 4:165
5,6-dichlororibofuranosyl
benzamizole (DRB) sensitivity
inducing factor (DSIF), 4:165,
4:165f
elongation maintenance factors, 4:164
gene expression integration, 4:166
historical research, 4:164
negative elongation factor (NELF),
4:165, 4:165f
negative transcription elongation
factors (N-TEFs), 4:164, 4:165,
4:165f
positive transcription elongation
factor (P-TEFb), 4:164,
4:165–166, 4:165f
recruitment processes, 4:166
7SK RNA, 4:166, 4:166f
termination factors, 4:165
transcription initiation factor IIS
(TFIIS), 4:164
transcription initiation factors,
4:164

eukaryotes (*Continued*)

- expression mechanisms
 - elongation phase, 2:366–367, 3:63–64, 3:63f
 - initiation phase, 2:366, 3:63–64, 3:63f
 - transcription cycle, 2:365, 3:63–64, 3:63f
 - transcription termination, 2:367, 3:63–64, 3:63f
 - mobile clamps, 2:365
 - structural characteristics, 2:365, 2:366f, 2:366t
 - RNA polymerase III (Pol III), 2:366t, 2:367–368, 4:159–161
 - Alu genes, 4:160
 - disease connections, 4:161
 - gene transcription regulation, 4:159
 - molecular characteristics, 4:159
 - nucleoli, 4:159
 - retrotransposition, 4:160
 - transcription apparatus, 4:160, 4:160f
 - tumor suppression, 4:160–161
 - RNA processing
 - ancillary RNA-processing signals, 3:62, 3:62f
 - basic concepts, 3:59–60, 3:60f
 - exon definition model, 3:61–62, 3:61f
 - exonic splicing silencers (ESSs), 3:62f
 - gene transcription regulation, 3:63–64, 3:63f
 - intronic splicing silencers (ISSs), 3:62f
 - life cycle, 3:62–63
 - open reading frames (ORFs), 3:60–61
 - research summary, 3:64
 - RNA-binding proteins, 3:59–60, 3:60f, 3:62–63
 - spliceosomes, 3:62–63, 3:63f
 - secretory pathway, 4:203–209
 - compartmental organization, 4:203, 4:204f
 - endoplasmic reticulum-based protein folding and modification
 - chaperones, 4:205–206
 - N-linked glycosylation, 4:204–205, 4:205f
 - quality control mechanisms, 4:205–206
 - microtubule-based motor proteins, 4:208, 4:208f
 - targeted proteins
 - characteristics, 4:203
 - constitutive secretion, 4:208
 - entry mechanisms, 4:203, 4:205f
 - integral membrane proteins, 4:204, 4:205f
 - protein exit, 4:207
 - regulated secretion, 4:208
 - signal recognition particles (SRPs), 4:203, 4:205f
 - soluble proteins, 4:204, 4:205f
 - transport mechanisms
 - Golgi apparatus, 4:206f, 4:207
 - protein localization, 4:206–207, 4:206f
 - retinoid X receptor (RXR), 4:207
 - vesicular transport, 4:206, 4:206f
 - selenoprotein synthesis, 4:213, 4:213f
 - supercoiling mechanisms, 2:161
 - translation initiation, 4:432–435
 - eukaryotic messenger RNA (mRNA)
 - AUG triplets, 4:433–435
 - 80S initiation complex, 4:434f
 - 80S ribosome formation, 4:435
 - eukaryotic initiation factors (eIFs), 4:433t, 4:434f, 4:435
 - 5' leader scanning, 4:433
 - 40S ribosomal subunits, 4:432
 - 43S preinitiation complexes, 4:433, 4:434f
 - 48S initiation complex, 4:435
 - initiation codon selection, 4:433–435
 - initiation mechanisms, 4:432
 - internal ribosomal entry sites (IRESs), 4:435
 - regulation mechanisms, 4:435
 - ribosomal dissociation, 4:432
 - 60S ribosomal subunits, 4:432, 4:435
 - structural characteristics, 4:432
 - transfer RNA (tRNA), 4:432–433
 - transposable elements (TEs), 3:281–282
 - TRiC/CCT eukaryotic chaperonin, 1:457f, 1:458–459
 - type I DNA topoisomerase, 2:158t, 2:160f
 - type II DNA topoisomerase, 2:164–165
 - type IIA DNA topoisomerase, 2:166
 - vacuole/lysosome functions
 - disease connections, 4:507–508, 4:507t
 - lysosome-related organelles, 4:506–507
 - pathogen-containing vacuoles, 4:507
 - protozoan contractile vacuole complex, 4:506
 - Y-family polymerases, 4:481
 - eukaryotic elongation factor 2 kinase (eEF-2K), 1:305f, 1:306
 - eukaryotic initiation factor (eIF) 4E-binding proteins, 2:604f, 2:605, 3:195, 3:196f, 3:197f, 4:433, 4:434f
 - Euphorbia characias*, 1:89
 - European Bioinformatics Institute (EBI), 3:611–612
 - European Molecular Biology Laboratory/ Nucleotide Archive (ENA), 3:611–612
 - eutherians, 1:553f
 - everolimus, 3:198–199
 - excitotoxicity, 1:423–424
 - exercise, 1:247–248
 - exocrine glands
 - exocrine gland cells, 4:415–416
 - melanocortin system, 3:26–27
 - exocytosis, 3:476
 - exons
 - alternative splicing, 1:75, 1:76f, 3:60–61, 4:187–188
 - ancillary RNA-processing signals, 3:62, 3:62f
 - chemical splicing mechanisms, 3:59–60, 3:60f, 4:185, 4:186f
 - exon definition model, 3:61–62, 3:61f
 - exonic splicing enhancers (ESEs), 1:76, 1:78f, 3:62, 3:62f
 - exonic splicing silencers (ESSs), 3:62f
 - exon–intron architecture, 3:142–144, 3:143f, 4:186–187
 - historical research, 4:185
 - mosaic mitochondrial genes, 3:142–144, 3:143f
 - ribozymes, 4:188
 - self-splicing reactions, 1:397, 1:398f, 4:188, 4:188f
 - splicing pathways, 4:186f, 4:187f
 - exonucleases
 - AP endonucleases, 2:3–4
 - DNA mismatch repair (MMR)
 - downstream activators, 2:69–70
 - excision and repair synthesis, 2:56
 - long-patch mismatch repair, 2:56f
 - molecular mechanisms, 2:61f
 - Mut protein family, 2:41f, 2:45–46, 2:54–56, 2:55t, 2:67f, 2:80
 - strand discrimination, 2:67
 - DNA polymerase kinetics, 2:99–100, 2:99f
 - exonuclease I (ExoI), 2:41f, 2:51, 2:52t, 2:53f, 2:61f, 2:69–70
 - exonuclease III, 2:3–4, 2:5t
 - 5' exonuclease, 2:90–91, 2:91f
 - polymerase β (Pol β), 2:84
 - polymerase δ (Pol δ), 2:245, 2:246f, 2:247f
 - 3'–5' exonuclease, 2:90
 - exoribonucleases
 - 3'-to-5' exoribonucleases, 3:53–55, 3:54f, 3:142, 3:142f
 - 5'-to-3' exoribonucleases, 3:56, 3:142, 3:142f
 - 3' exoribonucleases, 3:56
 - oligonucleotides, 3:56
 - transfer RNA (tRNA), 3:555–556, 3:555f
 - See also* ribosomal RNA (rRNA); transfer RNA (tRNA)
 - exosomes and microvesicles (EMVs), 2:262–264
 - biogenesis
 - cellular functions, 2:263
 - plasma membranes, 2:263
 - trans-acting factors, 2:263
 - composition
 - carbohydrates, 2:262
 - lipids, 2:262
 - nucleic acids, 2:262–263
 - proteins, 2:262, 2:263f
 - definition, 2:262
 - disease connections, 2:264
 - immunological signaling, 2:263–264
 - neighboring cell effects, 2:263–264
 - research summary, 2:264
 - structural characteristics, 2:262
- expressed sequence tags (ESTs), 2:369
- extensins
 - Arabidopsis thaliana* glycoproteins, 2:467f
 - multigene families, 2:467
 - structural studies, 2:467
- external monooxygenases, 3:372
- external NADH dehydrogenase, 4:85
- extracellular matrix (ECM), 1:426–429
 - adhesive motifs, 1:426, 1:427t
 - basic concepts, 1:426
 - cell adhesion process, 1:437, 1:437f, 2:318
 - cell–matrix adhesion diversity, 2:321–322, 2:321f
 - cell migration, 1:436–445

- cell–cell adhesion
 cadherins, 1:444–445
 collective and individual cell
 migration, 1:444–445, 1:445f
in vivo cell migration
 carcinogenesis, 1:442–443, 1:443f
 embryo development, 1:441–442
 immune trafficking, 1:442
 pathological conditions, 1:442, 1:443f
 wound healing, 1:442
- proteases
 degradation processes, 1:443–444
 ECM formation, 1:444
 endostatin, 1:444
 growth factor activation, 1:444
 proteolytic fragments, 1:444
 regulatory mechanisms, 1:444–445
- receptors
 integrins, 1:438–440, 1:439f
 proteoglycans, 1:439–440
 receptor clustering, 1:440, 1:440f
 signaling pathways, 1:439f
 regulatory mechanisms
 focal adhesion kinase (FAK) family,
 1:441, 1:441f
 integrins, 1:444–445
 proteases, 1:444–445
 Rho GTP-binding proteins,
 1:440–441, 1:441f
 research summary, 1:445
 stages, 1:437f
 three-dimensional (3D) cell migration
 and invasion, 1:443–444
 two-dimensional (2D) surfaces, 1:436
 components, 1:426
 focal adhesion kinase (FAK) family,
 2:265–266
 integrin–extracellular matrix (ECM)
 contacts, 1:439f, 2:610–611, 2:611f
 matrix metalloproteinases (MMPs), 3:90,
 3:95f
 signaling pathways, 1:429, 2:608
 structure and functions
 basement membranes, 1:437, 1:438f
 biochemical properties, 1:437–438
 cell migration, 1:437
 mechanical properties, 1:437–438
- extracellular proteins
 disulfide bond formation, 1:666, 1:667f
 glycation processes, 2:406
- extracellular signal-regulated kinase (ERK)
 angiotensin receptors, 1:113
 calcium ion (Ca²⁺) signaling system,
 2:719–720, 2:720f
c-fes proto-oncogene (Fes), 1:453
 chemokine receptors, 1:480–482, 1:482f
 cyclic adenosine monophosphate (cAMP)
 receptors (cARs), 1:564, 1:564f
 ERK1/ERK2, 3:165–166, 3:176, 3:177f
 ERK3/ERK4, 3:177f, 3:178
 ERK5, 3:177f, 3:178
 ERK7/ERK8, 3:177f, 3:178
 fibroblast growth factor receptors (FGFRs),
 2:304, 2:305f
 focal adhesion kinase (FAK) family,
 2:265–266
- focal adhesion signaling, 2:321f, 2:322
- glycoprotein hormone receptors,
 4:388–390
- G protein-coupled receptor kinases
 (GRKs), 2:485t
- G_q family proteins, 2:497–498, 2:498f
- integrin signaling, 1:439f, 2:611
- mitochondrial permeability transition
 pore, 3:165–166
- mTOR complexes, 3:195–196
- myocardial ischemia, 3:111
- neurotrophins, 3:241–243, 3:242f
- phylogenetic relationships, 3:177f
- Raf kinases, 4:247–248, 4:247f
- schematic diagram, 3:293f
- serine–threonine phosphatases,
 4:222–223, 4:222f
- signaling cascades, 3:177f
- somatostatin (sst) receptors, 4:265
- submembranous calcium signals, 3:294
- tachykinins, 4:362–363
- toll-like receptors (TLRs), 4:398–399
- transforming growth factor beta (TGF-β),
 1:620f
- extra-chromosomally-primed (EP)
 retrotransposons
 characteristics, 3:284
 integration site preference, 3:288
 long terminal repeat (LTR)
 retrotransposons, 3:284–286, 3:285f,
 3:286f, 3:287f
 open reading frames (ORFs), 3:284–286,
 3:286f, 3:287f
 tyrosine recombinases, 3:285f, 3:286f
 virus-like particles (VLPs), 3:284–286,
 3:286f
- eyes
 aquaporin water channels, 1:124t
 mucin expression, 3:203
- ezzrin–radixin–moesin (ERM) proteins,
 1:28–29, 1:29f, 2:320–321, 2:320f
- ## F
- Fabry disease, 4:291t
- facilitative glucose transporters (GLUTs),
 2:392f
- cerebral metabolism, 1:369–371, 1:370f
 characteristics, 2:394
 classifications, 2:392t
- GLUT1
 ascorbic acid, 4:532–533
 cardiac metabolism, 3:65–66, 3:66f,
 3:68f, 3:69f, 3:70f
 cerebral metabolism, 1:369–371, 1:370f
 characteristics, 2:392–393, 2:392f
 distribution and function, 2:392t
- GLUT2
 carbohydrate-responsive element
 binding protein (ChREBP)
 function, 1:376–377, 1:377f
 characteristics, 2:393
 distribution and function, 2:392t
- GLUT3
 ascorbic acid, 4:532–533
- cerebral metabolism, 1:369–371, 1:370f
 characteristics, 2:393
 distribution and function, 2:392t
- GLUT4
 Akt phosphorylation, 2:604–605
 ascorbic acid, 4:532–533
 cardiac metabolism, 3:65–66, 3:66f,
 3:68f, 3:69f, 3:70f
 cerebral metabolism, 1:369–371, 1:370f
 characteristics, 2:393
 distribution and function, 2:392t
 insulin resistance mechanisms,
 1:651–652, 1:651f
 insulin signal transduction, 2:604f
 proline-rich tyrosine kinase 2 (PYK2),
 2:267
 thiazolidinedione (TZD), 3:421–422,
 3:422f
- GLUT5
 cerebral metabolism, 1:369–371, 1:370f
 characteristics, 2:393
 distribution and function, 2:392t
- GLUT6
 characteristics, 2:393
 distribution and function, 2:392t
- GLUT7
 characteristics, 2:393
 distribution and function, 2:392t
- GLUT8
 characteristics, 2:393–394
 distribution and function, 2:392t
- GLUT9
 characteristics, 2:394
 distribution and function, 2:392t
- GLUT10
 characteristics, 2:394
 distribution and function, 2:392t
- GLUT11
 characteristics, 2:394
 distribution and function, 2:392t
- GLUT12
 characteristics, 2:394
 distribution and function, 2:392t
- GLUT13
 characteristics, 2:394
 distribution and function, 2:392t
- GLUT14
 characteristics, 2:394
 distribution and function, 2:392t
- FACKEL* genes, 1:251
- F-actin, 1:18, 3:219–221, 3:221f, 3:438–439,
 4:216, 4:217f
- factor F₄₃₀
 biosynthetic pathway, 3:545f, 3:546–547,
 3:546f
 characteristics, 3:544
- factor for inversion stimulation (Fis), 2:122,
 2:122f
- familial benign hypocalciuric hypercalcemia
 (FBHH), 1:326
- familial British dementia, 1:101t
- familial glucocorticoid deficiency (FGD),
 3:25–26
- familial hypocalciuric hypercalcemia (FHH),
 1:326, 1:329f
- familial partial lipodystrophy, 2:632t

- familial tumoral calcinosis, 2:471*t*
 Farber syndrome, 4:291*t*
 farnesoid X-receptor (FXR)-dependent
 mechanisms, 1:167, 1:169*f*
 farnesyl diphosphate, 1:516, 1:517*f*
 farnesyl isoprenoid, 4:244*f*
 farnesyl transferase, 2:737–738, 2:737*f*
 farnesyl transferase inhibitors (FTIs), 2:740
 Fas-associated protein with death domain
 (FADD), 1:389, 1:390*f*, 1:394*f*,
 1:617–619, 1:618*f*
 fasciculation and elongation protein zeta-1
 (FEZ1), 1:197*f*, 1:199, 1:200*f*
 Fascin, 1:29*f*, 1:34, 3:243, 3:244*f*
 Fat10 (diubiquitin), 4:468*t*, 4:471–472
 fatty acid 2-hydroxylase (FA2H) gene, 4:272
 fatty acids
 acyl-coenzyme A transesterification,
 1:384–385, 1:385*f*, 2:290, 2:291*f*
 anaplerotic pathways, 1:106*f*
 biosynthesis, 2:285–289
 de novo fatty acid biosynthesis,
 2:285–286
 lipogenic gene expression
 carbohydrate induction, 2:286–287
 highly unsaturated fatty acids
 (HUFAs), 2:287*f*, 2:288
 HUFA-response region with
 carbohydrate response element
 (CHORE), 2:288
 HUFA-response region with insulin
 response element (HUFA-RR/IRE),
 2:287–288
 nutritional regulation, 2:287–288
 omega-6/omega-3 fatty acids, 2:288
 saturated fatty acids
 basic concepts, 2:285–286
 energy metabolism regulation, 2:286
 malonyl-CoA, 2:285–286
 pyruvate dehydrogenase regulation,
 2:285–286
 unsaturated fatty acids
 arachidonic and docosahexenoic acid
 synthesis, 2:288, 2:288*f*
 oleic and palmitoleic acids, 2:288
 branched-chain amino acid (BCAA)
 regulation, 1:245–246
 carbohydrate-responsive element binding
 protein (ChREBP), 1:376, 1:376*f*
 cardiac metabolism, 3:66–67, 3:66*f*
 coenzyme A (CoA)
 peroxisome proliferator-activated
 receptor alpha (PPAR α),
 3:419–421, 3:420*f*
 4'-phosphopantetheine (4'-PP), 1:539
 synthesis process, 1:541
 fatty-acid binding proteins (FABPs), 2:281
 fatty acid metabolism
 AMP-activated protein kinase (AMPK),
 2:286, 4:56
 fatty acid synthase expression
 cancer diagnostics, 2:277–278
 functional role, 2:276–279
 human cancer expression, 2:277,
 2:277*f*
 inhibition mechanisms, 2:278–279
 therapeutic targets, 2:278
 metabolism—cancer relationship,
 2:275–280
 acyl-coenzyme A synthetase (ACS),
 2:279
 alpha-methylacyl-CoA racemase
 (AMACR), 2:279
 background information, 2:275
 fatty acid synthase, 2:276–279
 fatty acid synthesis pathway,
 2:275–279, 2:276*f*
 fatty acid synthase
 carbohydrate-responsive element
 binding protein (ChREBP),
 1:376–377, 1:377*f*
 fatty acid metabolism—cancer
 relationship
 cancer diagnostics, 2:277–278
 functional role, 2:276–279
 human cancer expression, 2:277,
 2:277*f*
 inhibition mechanisms, 2:278–279
 therapeutic targets, 2:278
 fatty acid synthesis pathway, 2:275–279,
 2:276*f*
 lipogenic gene expression, 2:287*f*, 4:56
 liver pyruvate kinase (LPK) catalysis,
 1:374
 malonyl-CoA
 body weight regulation, 4:57, 4:57*f*
 functional roles, 4:56
 metabolite channeling, 3:80
 4'-phosphopantetheine (4'-PP),
 1:539
 synthesis process, 1:541, 2:285–286
 fatty acid synthesis pathway, 2:275–279,
 2:276*f*
 fatty acyl-coenzyme A (CoA) synthetases,
 2:290–295
 ACSF proteins, 2:294
 activation reactions, 2:290, 2:291*f*
 amino acid sequence motifs, 2:290–291,
 2:291*f*, 2:294–295
 Bubbligum ACS family (ACSBG), 2:294
 characteristics and occurrence,
 2:290–294
 human ACS families, 2:290–294, 2:292*t*
 long-chain acyl-CoA synthetase (LCAS),
 1:385*f*, 2:292–293
 medium-chain acyl-CoA synthetase
 (MCAS), 2:292
 nomenclature, 2:292, 2:292*t*
 peroxisome proliferator-activated
 receptor alpha (PPAR α),
 3:419–421, 3:420*f*, 3:422*f*
 research summary, 2:295
 short-chain acyl-CoA synthetase (SCAS),
 2:292
 structure—function relationship, 2:291*f*,
 2:294–295
 very long-chain acyl-CoA synthetase
 (VLCAS), 2:293–294
 free fatty acids (FFAs), 1:50*f*, 2:355*f*,
 4:301–302, 4:302*f*, 4:303–304, 4:305
 gluconeogenesis, 2:384
 insulin metabolism, 2:606–607
 insulin resistance, 1:650, 1:650*f*,
 2:606–607
 insulin-secreting cells (β -cells), 2:598–600
 long-chain fatty acid oxidation defects and
 myopathies, 1:110
 malonyl-CoA
 body weight regulation, 4:56–58
 carnitine palmitoyltransferase 1
 (CPT 1), 2:258–259, 4:56, 4:58
 cellular fatty acid metabolism
 regulation, 4:56
 energy homeostasis, 4:56–57
 fatty acid synthase inhibitors, 4:57,
 4:57*f*
 functional roles, 4:56
 hypothalamic malonyl-CoA, 4:57,
 4:57*f*, 4:58*f*
 hypothalamus, 4:56–57
 nutrient signaling processes, 4:58
 cardiac metabolism, 3:66*f*, 3:68*f*, 3:69*f*,
 3:70*f*
 fatty acid biosynthesis
 energy metabolism regulation, 2:286
 lipogenic gene expression, 2:287*f*
 pyruvate dehydrogenase regulation,
 2:285–286
 fatty acid oxidation, 1:541, 2:283
 microRNA (miRNA)-based regulation,
 4:194*t*
 mitochondrial acyl-CoA/CoA ratio
 regulation, 1:387
 mitochondrial β -oxidation
 carnitine-dependent cellular uptake and
 activation, 1:384–385, 1:385*f*
 mitochondrial carnitine system,
 1:385–386
 reaction pathways, 1:386, 1:386–387
 oxidation processes, 2:281–284
 acyl-coenzyme A
 cardiac metabolism, 3:66–67, 3:66*f*,
 3:69*f*, 3:70*f*
 cellular uptake and activation, 2:281
 malonyl-CoA, 4:56
 mitochondria, 2:282*f*
 peroxisomes, 2:283–284
 unsaturated fatty acids, 2:282–283,
 2:283*f*
 α -oxidation, 2:284
 β -oxidation
 mitochondria, 2:281–282, 2:282*f*,
 3:65–66, 3:66*f*, 3:68*f*, 3:69*f*,
 3:70*f*
 peroxisomes, 2:283–284
 unsaturated fatty acids, 2:282–283,
 2:283*f*
 cardiac metabolism, 3:66–67, 3:66*f*
 cellular uptake and activation, 2:281
 hyperammonemia—hyperornithinemia—
 homocitrullinemia (HHH)
 syndrome, 4:495*t*
 malonyl-CoA, 1:541, 2:283, 3:71
 mitochondria
 β -oxidation, 2:281–282, 2:282*f*
 general discussion, 2:281
 regulation mechanisms, 2:283
 uptake and activation, 2:281, 2:282*f*

- peroxisomal fatty acid β -oxidation, 3:414–415
 ω -oxidation, 2:284
 pancreatic endocrine cells, 2:596
 peroxisome proliferator-activated receptor
 alpha (PPAR α), 3:419–421, 3:420f
 polyunsaturated fatty acids (PUFAs), 1:376
 saturated fatty acids
 biosynthesis
 basic concepts, 2:285–286
 energy metabolism regulation, 2:286
 malonyl-CoA, 2:285–286
 pyruvate dehydrogenase regulation, 2:285–286
 protein palmitoylation, 2:737
 starvation, 4:303–304
 thiamine diphosphate (ThDP), 1:207f
 tricarboxylic acid cycle, 4:439f
 unsaturated fatty acids
 biosynthesis
 arachidonic and docosahexenoic acid
 synthesis, 2:288, 2:288f
 oleic and palmitoleic acid synthesis, 2:288
 protein palmitoylation, 2:737
 FBOX8 protein, 1:136
 Fechtner syndrome, 3:223t
female lethal on the 2nd (fl(2)d) gene, 1:84
 FEN1 protein, 2:4f, 2:5–6
 Fenna–Matthews–Olson protein, 2:501–503, 2:504f, 2:505f, 2:507
 Fenton/Haber-Weiss pathway, 3:169
 FERM domain
 focal adhesion kinase (FAK) family, 2:265–266, 2:266f
 phospholipase C (PLC), 3:467–468
 ferredoxin (Fd), 2:296–298
 characteristics, 2:296
 chloroplast signaling, 1:514–515
 dark fermentation, 4:74f
 ferredoxin-NADP⁺ reductase (FNR), 2:299–302
 catalytic mechanisms
 ferredoxin (Fd) interactions, 2:301, 2:301f
 flavin adenine dinucleotide (FAD/FADH₂), 2:299–301, 2:300f
 NADP(H) interactions, 2:300, 2:300f
 cytochrome *b₆* complex, 4:338–339, 4:339f
 heterotrophic glutathione reductase (GR)-type FNRs, 2:301
 molecular structure, 2:299, 2:300f
 nonphotosynthetic plant-type FNRs
 bacterial/apicomplexan FNRs, 2:301
 root FNRs, 2:301
 photosynthetic processes, 2:296, 2:297f
 physiological functions, 2:296–297, 2:299, 2:300f
 research background, 2:299–301
 subcellular localization, 2:299
 green bacteria, 2:504f, 2:507–508, 2:508f
 iron–sulfur clusters (FeS), 2:297–298, 2:301f, 2:661f, 3:485, 3:505, 3:505f
 mitochondrial-type ferredoxins, 2:297
 molecular structure, 2:296, 2:297f
 photosynthetic processes, 1:511–512, 1:512f, 2:296, 2:297f, 3:32, 3:493–494, 3:494f
 photosystem I (PSI) complex, 3:485, 3:505, 3:505f
 physiological functions, 2:296–297, 2:297f
 plant-type ferredoxins, 2:296
 thioredoxin-type ferredoxins, 1:514–515, 2:297, 2:661f
 ferricytochrome *c* oxidoreductase, 3:663–664, 3:664f
 ferrochelatase, 2:540f, 2:541, 2:542t
Ferropasma acidarmanus, 2:12–13
 fertility, 4:547
 fetal hemoglobin (HbF), 2:534
 fgr genes/Fgr proteins, 4:294
 fibrillar centers, 3:324, 3:325f
 fibrillar proteins, 1:426
 fibrillin, 3:91t
 fibrinogen (FG)
 amyloid deposits, 1:101t
 blood-clotting proteases, 3:586f, 3:587
 cell–matrix interactions, 1:427t
 integrins, 1:413f, 1:427t, 1:438–440
 matrix metalloproteinases (MMPs), 3:91t
 fibroblast growth factors (FGFs), 2:303–308
 cadherin signaling, 1:264
 calbindin-D_{28k}, 1:287
 calcium-sensing receptor (CaSR), 1:329–330
 cancer-related diseases
 epithelial–mesenchymal transition (EMT), 2:304–306
 expression effects, 2:304–306
 fibroblast growth factor 1 (FGF1), 2:305
 fibroblast growth factor 2 (FGF2), 2:305
 fibroblast growth factor 3 (FGF3), 2:305
 fibroblast growth factor 4 (FGF4), 2:305
 fibroblast growth factor 5 (FGF5), 2:306
 fibroblast growth factor 6 (FGF6), 2:306
 fibroblast growth factor 7 (FGF7), 2:306
 fibroblast growth factor 8 (FGF8), 2:306
 fibroblast growth factor 9 (FGF9), 2:306
 fibroblast growth factor 10 (FGF10), 2:306
 fibroblast growth factor 19 (FGF19), 2:306
 cardiac muscle development, 1:183–184, 1:184f
 characteristics and functional role
 general discussion, 2:303
 genes/gene subgroups, 2:303
 physiological functions, 2:303
 proteins, 2:303
 evolutionary relationships, 2:304f
 fibroblast growth factor 1 (FGF1), 2:305
 fibroblast growth factor 2 (FGF2), 1:197f, 1:198–199, 2:305
 fibroblast growth factor 3 (FGF3), 2:305
 fibroblast growth factor 4 (FGF4), 2:305
 fibroblast growth factor 5 (FGF5), 2:306
 fibroblast growth factor 6 (FGF6), 2:306
 fibroblast growth factor 7 (FGF7), 2:306
 fibroblast growth factor 8 (FGF8), 2:306
 fibroblast growth factor 9 (FGF9), 2:306
 fibroblast growth factor 10 (FGF10), 2:306
 fibroblast growth factor 19 (FGF19), 1:167, 2:306
 fibroblast growth factor receptors (FGFRs)
 cancer-related diseases
 characteristics, 2:306–308
 fibroblast growth factor receptor 1 (FGFR1), 2:306–307
 fibroblast growth factor receptor 2 (FGFR2), 2:307
 fibroblast growth factor receptor 3 (FGFR3), 2:307
 fibroblast growth factor receptor 4 (FGFR4), 2:307–308
 genes, 2:303
 immunoglobulin (Ig)-like domains, 2:303–304, 2:304f, 2:305f
 physiological functions, 2:304
 proteins, 2:303–304, 2:304f, 2:305f
 research summary, 2:308
 liver regeneration, 1:192, 1:192f
 neurogenesis, 1:198–199, 1:200f
 parathyroid hormone (PTH), 3:386
 research summary, 2:308
 fibroblasts, 3:105–106, 3:106f
 fibronectin (FN)
 adhesive motifs, 1:426, 1:427t
 cell–matrix interactions, 1:426
 fibronectin type-III (FNIII) motif, 2:619
 focal adhesion kinase (FAK) family, 2:265–266
 integrins, 1:413f, 1:427t, 1:438–440
 matrix metalloproteinases (MMPs), 3:91t
 fibulin, 3:91t
 ficolin, 2:702t
 filamentous anoxygenic phototropic (FAP) bacteria
 alternative complex III, 2:512
 characteristics, 2:510
 cytochrome *c*-554, 2:512
 uncharacterized cytochromes, 2:512
 filamin, 2:320–321, 2:610–611
 filensin, 2:632t, 2:634
 fimbrin, 1:34, 1:319t
 firefly luciferase, 1:462–465, 1:464f, 2:290–291
 Fis1 protein, 3:136–137, 3:137f
 fish
 cytochrome *c*, 1:600, 1:600f
 pterin distribution, 3:666
 vasopressin/oxytocin family, 4:516f, 4:516t
 Fis protein, 3:367–368, 3:368t, 3:370
 fission yeasts, 4:217–218
 5 CITEP, 2:106f
 5-hydroxymethylcytosine (5hmC), 3:9
 5-hydroxytryptamine-2C receptor (5-HT_{2c}R)
 adenosine deaminases, 4:155–156, 4:156f
 adenosine-to-inosine (A-to-I) RNA editing, 4:155
 5-methylcytosine (5mC)
 chemical structure, 4:77f
 deoxyribonucleic acid (DNA), 3:9, 3:9f
 functional role, 2:16, 4:76
 mismatch repair (MMR), 1:523f
 5-oxo-ETE receptors, 2:189

- FK506 binding protein (FKBP) subunits, 3:365, 4:198
- flavins, 2:309–313
- C-4a-hydroperoxide, 2:310f, 2:312
 - C-4a-hydroxide, 2:310f, 2:312
 - chemical properties, 2:309
 - classifications
 - dehydrogenase/electron transferase, 2:311
 - electron-transfer proteins, 2:311
 - oxidases, 2:312
 - oxygenases, 2:311–312
 - pyridine nucleotide oxidoreductases, 2:311
 - definition, 2:309
 - DNA photolyase, 2:312
 - flavin adenine dinucleotide (FAD/FADH₂)
 - branched-chain α -ketoacid dehydrogenase complex (BCKDC), 1:246–247, 1:247t
 - cardiac metabolism, 3:67
 - chemical properties, 2:309, 4:83
 - cobalamin-dependent methionine synthase, 2:327
 - cytochrome P-450 (CYP), 1:608–609, 1:609f
 - enzyme catalysis, 2:691–693, 2:691f, 2:692f, 2:692t
 - fatty acid oxidation, 2:281–282, 2:282f, 3:67
 - ferredoxin-NADP⁺ reductase (FNR), 2:299–301, 2:300f
 - histone modifications, 4:61
 - importance, 2:309
 - kinetic isotope effects (KIEs), 2:691–693, 2:691f, 2:692f, 2:692t
 - mitochondrial β -oxidation, 1:386, 1:386–387
 - mitochondrial ROS, 1:210, 4:19f, 4:20
 - molecular structure, 2:310f
 - nitric oxide synthase (NOS), 1:342, 1:343f, 1:344–346, 1:344t, 1:345f, 3:253–254, 3:668
 - PET genes, 3:308f
 - plant mitochondrial electron transport chain, 1:145, 1:146f
 - pyruvate dehydrogenases (PDHs), 4:321–322, 4:323
 - respiratory chain complexes, 4:83
 - succinate:quinone oxidoreductases (SQORs), 4:95, 4:95f
 - tricarboxylic acid cycle, 4:439f
 - flavin mononucleotide (FMN)
 - biochemical free radical chain reactions, 3:169–170
 - chemical properties, 2:309, 4:83
 - cobalamin-dependent methionine synthase, 2:327
 - cytochrome P-450 (CYP), 1:608–609, 1:609f
 - importance, 2:309
 - mitochondrial ROS, 4:18–20, 4:19f
 - molecular structure, 2:310f
 - nitric oxide synthase (NOS), 1:342, 1:343f, 1:344–346, 1:344t, 1:345f, 3:253–254, 3:668
 - reactive oxygen species (ROS), 2:77
 - respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:341–342, 4:342f, 4:342t, 4:344–345, 4:344f
 - respiratory chain complexes, 4:83
 - riboswitches, 4:137, 4:138f
 - flavohydroquinones, 2:309, 2:310f
 - flavoquinones, 2:309, 2:310f
 - flavosemiquinones, 2:309, 2:310f
 - importance, 2:309
 - medical importance, 2:312–313
 - molecular structure, 2:310f
 - oxidation reactions, 2:309, 2:310f
 - photochemical properties, 2:312
 - redox properties, 2:309
 - respiratory chain complexes, 4:83
 - riboflavin
 - chemical properties, 2:309
 - importance, 2:309
 - molecular structure, 2:310f
 - spectroscopic properties, 2:309–311, 2:311f
 - flavivirus, 3:41
 - Flavobacterium okeanokoites*, 2:132t
 - flavocytochrome *c*, 2:511
 - flavodoxin, 3:505
 - flavoenzymes, 3:116
 - flavohydroquinones, 2:309, 2:310f
 - flavopiridol, 4:165
 - flavoproteins, 4:85
 - flavoquinones, 2:309, 2:310f
 - flavosemiquinones, 2:309, 2:310f
 - flies
 - cytochrome *c*, 1:600, 1:600f
 - multisubunit centromeres, 1:447
 - FLIP inhibiting protein, 1:390f, 1:393–394, 1:394f
 - flippase proteins, 2:314–317
 - aminophospholipid flippases, 2:314–315, 2:315f, 2:316t
 - cytoplasmic leaflets, 2:314, 2:316t
 - dolichyl pyrophosphate (Dol-P-P), 2:315–316, 2:315f, 2:316t
 - flippase-mediated transbilayer movement, 2:314, 2:315f
 - flopases, 2:314–315, 2:315f
 - functional role, 4:334
 - glucosylphosphoryldolichol (Glc-P-Dol), 2:315–316, 2:315f, 2:316t
 - glycosylation, 2:315–316
 - glycosyl phosphatidylinositol (GPI) anchors
 - bacterial cell wall complex, 2:316–317
 - biosynthesis, 2:316–317
 - glucosylceramide (Glc-Cer), 2:316, 2:316t
 - lipid bilayer structure, 2:735
 - mannosylphosphoryldolichol (Man-P-Dol), 2:315–316, 2:315f, 2:316t
 - membrane biogenesis-related bilayer expansion, 2:314
 - protein O- and C-mannosylation, 2:316
 - scramblases, 2:314–315, 2:315f
 - flopases, 2:314–315, 2:315f
 - FLP protein, 1:556–557, 1:559
 - flu* mutant, 1:515
 - Fluo3/4, 2:575t, 2:576–577
 - fluorescein isothiocyanate (FITC), 3:678
 - fluorescence
 - basic principles, 1:461, 1:462f
 - bi-molecular fluorescence
 - complementation (BiFC), 2:749, 2:753f, 2:754
 - fluorescence *in situ* hybridization (FISH), 3:296, 3:297, 3:297f
 - fluorescence resonance energy transfer (FRET)
 - cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:565
 - DNA unwinding, 2:25
 - G-protein-coupled receptors (GPCRs), 1:565
 - green fluorescent protein (GFP)
 - indicators, 2:578
 - membrane fusion, 3:42
 - sliding clamps, 4:237–238
 - intracellular calcium ion (Ca²⁺) indicators, 2:574–577, 2:575t, 2:576f
 - three-dimensional fluorescence *in situ* hybridization (3D-FISH) analysis
 - general discussion, 2:749
 - nuclear structure and dynamics, 2:749–751
 - fluoropalmitate, 3:647
 - focal adhesion kinase (FAK) family, 2:265–268
 - activation and signaling mechanisms, 2:265–266, 2:267f
 - cell migration, 1:439, 1:439f, 1:441, 1:441f
 - cellular responses, 2:266
 - characteristics, 2:265–266, 2:266f
 - domain conservation, 2:266t, 2:267–268
 - expression regulation, 2:265
 - FERM domain, 2:265–266, 2:266f
 - intracellular signaling pathway
 - connections, 2:611
 - invertebrates, 2:267–268
 - localizations, 2:265–266, 2:266f
 - matrix adhesions, 1:439, 1:439f
 - signaling adhesion functions, 4:294–295
 - submembrane plaque, 2:320–321, 2:320f
 - focal adhesions, 2:318–323
 - cell adhesion process, 1:439, 1:439f, 2:318
 - cell–matrix adhesion diversity, 2:321–322, 2:321f
 - characteristics, 2:318
 - cytoskeletal domain, 2:320
 - extracellular matrix (ECM) ligands, 2:320, 2:320f
 - integrins
 - integrin adhesion network, 2:320f
 - signaling pathways, 1:428, 1:439, 1:439f
 - mechanosensitivity, 2:322, 2:322f
 - molecular organization, 2:319f, 2:320, 2:320f
 - nonmuscle cell contraction, 1:33–34
 - signaling pathways, 2:321f, 2:322
 - structural characteristics, 2:318–319, 2:319f
 - submembrane plaque, 2:320–321, 2:320f

- focal and segmental glomerulosclerosis (FSGS), 4:416
- fodrin, 1:319t, 2:625–627
- FOG-1 transcription factor, 1:187
- FokI enzyme, 2:131, 2:132t, 2:133
- folate, 2:324–328
- absorption and transport, 2:324
 - biochemistry, 2:324–326
 - chemical characteristics and structure, 2:324, 2:325f
 - deficiency-related diseases, 2:327
 - functional role
 - methionine synthesis, 2:324–326, 2:325f
 - serine–glycine interconversion, 2:324–325
 - metabolic reactions, 2:325f
 - neural tube defects (NTDs), 2:327
 - nucleotide synthesis, 2:325, 2:325f
- follicle-stimulating hormone receptor (FSHR)
- dimerization/oligomerization, 4:388
 - extracellular domains, 4:388
 - functional role, 4:387–388
 - genomic organization, 4:388
 - G protein activation, 4:388–390
 - naturally occurring mutations, 4:390–391
 - structural organization, 4:389f
- follicular lymphoma, 1:165
- See also* Bcl2 family
- Follistatin protein, 1:229–230
- food deprivation, 4:298
- forkhead-associated (FHA) binding domains, 1:412f, 1:414
- forkhead box class O1 (FOXO1) transcription factor, 2:604–605, 2:604f, 2:607, 3:241–243, 3:242f
- formamidopyrimidines, 2:1, 2:2f, 2:78, 2:78f
- formate
- anaerobic respiration, 2:206, 2:208f
 - biomass-produced renewable hydrogen, 4:75f
 - fermentation reactions, 2:204–206, 2:205t
 - formate dehydrogenase, 3:402–403, 3:403f, 4:93f
- formins, 1:29–31, 1:30f, 1:21f, 1:24
- formyl-coenzyme A (CoA), 1:207f
- 4-*N*-trimethylammonio- 3-hydroxybutanoate
- See* carnitine
- FoxA1/A2 transcription factors, 1:192
- foxglove, 3:686–687
- Fpg protein, 2:57–58, 2:57f, 2:80, 2:80f
- fragmin, 1:23
- free fatty acids (FFAs), 1:50f, 2:355f, 4:301–302, 4:302f, 4:303–304, 4:305
- Freeman–Sheldon syndrome, 3:223t
- free oligosaccharides (FOSs), 2:329–331
- generation mechanisms, 2:329, 2:330f
 - glycoprotein-based folding processes, 2:329–330
 - Golgi-modified free oligosaccharides (FOSs), 2:331
 - research summary, 2:331
 - species specificity, 2:331
 - transport mechanisms, 2:329
- free radicals, 3:168–175
- aging/age-related diseases, 3:174–175
 - biochemical chain reactions, 3:168–169, 3:169f
 - dysfunctional mitochondria, 3:173–174
 - historical research, 3:168–169
 - hydrogen peroxide (H₂O₂)
 - biochemical chain reactions, 3:168–169, 3:169f
 - mitochondrial production, 3:169–170, 3:171f, 3:173f
 - physiological functions, 3:172–173
 - mitochondrion-dependent apoptosis, 3:174
 - nitric oxide (NO), 3:170–172, 3:171f, 3:172–173, 3:172f, 3:173f
 - oxidative stress, 3:173–174
 - peroxynitrite, 3:171f, 3:172f
 - superoxide anion (O₂⁻)
 - biochemical chain reactions, 3:168–169, 3:169f
 - mitochondrial production, 3:169–170, 3:170f, 3:171f, 3:173f
 - physiological functions, 3:172–173
- Fridovich dogma, 3:168–169
- Friedreich's ataxia (FRDA), 2:332–334
- clinical features, 2:332
 - genetic studies
 - frataxin functions, 2:333
 - mapping research, 2:332–333
 - phenotype–genotype correlations, 2:333
 - historical research, 2:332
 - pathology, 2:332
 - prevalence, 2:332
 - therapeutic interventions, 2:333
- Fringe, 2:453, 2:454f
- Frizzled receptors, 2:435, 2:436f
- frogs, 2:638t, 4:516t
- frontotemporal dementia, 4:507t
- FRT recombination site, 1:556–557, 1:559
- fructose
- fructose 1,6-bisphosphate (fructose-1,6-BP)
 - cardiac metabolism, 3:65–66, 3:66f, 3:68f, 3:69f
 - gluconeogenesis, 2:381, 2:382f, 2:383, 2:385f
 - glycolytic pathway, 2:445f
 - photosynthetic metabolic pathways, 3:499–500, 3:499f
 - fructose 6-phosphate (fructose-6-P)
 - carbohydrate-responsive element
 - binding protein (ChREBP), 1:377f
 - cardiac metabolism, 3:65–66, 3:66f, 3:68f, 3:69f
 - enzymology, 3:456–457
 - gluconeogenesis, 2:382f
 - glycolytic pathway, 2:443–444, 2:445f
 - pentose phosphate pathway, 3:391, 3:392f, 3:393f, 3:394f
 - photosynthetic metabolic pathways, 3:498f, 3:499, 3:499f
 - phosphotransferase system (PTS), 2:387
 - sugar phosphorylation, 2:543
- fructoselysine, 2:406f
- fruit flies
- See* *Drosophila melanogaster*
- fruitless (*fru*) gene, 1:82, 1:82f, 1:85–86
- fruits, 4:533t, 4:547–548
- FtsH metalloprotease, 1:3t, 1:4f, 1:6
- FtsK protein, 1:561
- FTY720 phosphate, 2:762
- fucose
- O-fucose
 - carbohydrate modifications, 2:453, 2:454f
 - classifications, 2:453
 - epidermal growth factor (EGF)-like repeats, 2:453, 2:454f
 - notch signaling, 2:455
 - protein O-fucosyltransferase I (Pofut1), 2:453, 2:454f, 2:456
 - thrombospondin type 1 repeats (TSRs), 2:454, 2:455f, 2:456
 - uPA–uPA receptor signaling, 2:455
 - N-linked glycosylation, 2:458–460, 2:459f
 - O-linked glycosylation, 2:466f
- fucosanthin, 1:502f
- Fukuyama-type congenital muscular dystrophy (FCMD), 2:471t, 2:474–475
- fumarase, 4:439t
- fumarate
- anaerobic respiration, 2:206, 2:207f, 2:208f
 - anaplerotic pathways, 1:106, 1:106f
 - citric acid cycle, 4:438f
 - fumarate reductase, 1:487t
 - plant mitochondrial electron transport chain, 1:145, 1:146f
 - quinol:fumarate reductases (QFRs), 4:92, 4:93f, 4:94f
 - tricarboxylic acid cycle, 4:439f
- fumaric acid, 4:495f
- fungi
- alternative oxidases (AOXs), 1:145–146
 - calpains, 1:354f
 - copper-transporting ATPases (copper pumps), 3:671–672
 - fungal infections, 1:139–140
 - fungal peroxidases, 3:407–409
 - mitochondrial DNA (mtDNA), 3:133t
 - plasma membrane proton (H⁺)-ATPases, 3:688–692
 - biogenesis, 3:690, 3:690f
 - characteristics, 4:333f, 4:334
 - domain structures, 3:688–689
 - functional role, 3:688, 3:689f
 - P-type ATPases, 3:688
 - reaction mechanisms, 3:689–690, 3:689f
 - regulation mechanisms, 3:690–691
 - research background, 3:688
 - research summary, 3:691
- prions, 3:566–570
- disease states, 3:568–569
 - functional domains, 3:566–567, 3:568f
 - generation mechanisms, 3:569
 - genetic criteria, 3:566, 3:567f
 - known prions, 3:566
 - prion domains, 3:566–567, 3:568f
 - research summary, 3:569–570
 - self-propagating amyloidoses, 3:567–568, 3:568f

fungi (*Continued*)

- protein biosynthesis elongation, 2:253
- septins, 4:216f, 4:217
- funiculosin, 1:597f, 1:598
- Fura group, 2:575t, 2:576, 2:576f
- Furaptra, 2:575t, 2:576
- fusicoccin, 3:691
- FXGXG motif, 2:137–138, 2:137f
- fyn genes/Fyn proteins, 4:294
- FYVE domain, 3:448, 3:449t

G

- G●U wobble pairs, 1:398f, 1:399–400, 1:399f, 1:400f
- G40P helicase, 2:21, 2:22t
- G4-DNA
 - background information, 2:144–146
 - biological role, 2:144–146, 2:145f
 - bonding schemes, 2:144f
 - chemical properties, 2:143f, 2:144
 - pharmacological role, 2:146
- G-actin, 1:18
- gadolinium (Gd), 1:328
- Gal4p transcription factor, 4:460
- galactitol, 2:387
- galactose
 - chemical structure, 2:696f
 - β1,4-galactosyltransferase (b4GalT), 2:454f
 - N-linked glycosylation, 2:458–460, 2:459f, 2:470, 2:470f, 2:471t
 - O-linked glycosylation, 2:453, 2:454f, 2:466f
 - structural characteristics, 3:654
 - sugar phosphorylation, 2:543
- galactosemia, 4:496f
- galactose permease (GalP), 2:388
- galactosidases
 - α-galactosidase A, 4:292
 - β-galactosidase, 2:77, 4:291
- galanin, 2:344t
- Galanthus nivalis*, 2:702t
- galectins, 2:702t, 2:703f, 2:704t, 3:91t
- gallamine, 3:215
- gall bladder, 2:355f
- Gallus gallus*, 1:37f
- GAL promoter system, 1:525–529
 - GAL1–10 gene expression
 - basic concepts, 1:525–527
 - gene-specific factors
 - Gal3p, 1:525–526
 - Gal4p, 1:525, 1:528–529
 - Gal80p, 1:525, 1:528–529
 - global process factors, 1:526
 - hypersensitive region (HR), 1:526–527, 1:526f
 - nucleosome positioning, 1:525
 - promoter chromatin organization, 1:526–527, 1:526f
 - promoter elements, 1:525, 1:526f
 - regulatory and transcription factors, 1:525–526
 - promoter chromatin organization
 - GAL1–10 gene expression, 1:526–527, 1:526f

- regulation function
 - active (galactose-induced) state, 1:527
 - inactive states, 1:527–528
 - promoter nucleosomes–regulatory factors interactions, 1:528–529
- galvanometer-driven wavelength scanning, 4:276
- gametogenesis, 3:20
- gamma-aminobutyric acid (GABA), 2:722–725
 - agonist- and drug-binding sites, 2:725, 2:725f
 - amino acid metabolism, 1:93t
 - biosynthesis, 2:722, 2:723f
 - body weight regulation, 3:77
 - degradation pathways, 2:722, 2:723f
 - disease states, 2:725
 - functional role, 2:347
- GABA_A receptor, 2:347–350
 - biochemistry
 - benzodiazepine-binding sites, 2:722, 2:725, 2:725f
 - molecular biology, 2:722–723, 2:724f, 2:724t
 - functional domains
 - allosteric regulation, 2:348f, 2:349
 - assembly processes, 2:348f, 2:349
 - ion channels, 2:348f, 2:349
 - ligand-binding sites, 2:348f, 2:349
 - functional role, 2:347
 - molecular structure
 - anatomical localization, 2:348
 - associated proteins, 2:349
 - functional heterogeneity, 2:348
 - heteropentamers, 2:347, 2:348f
 - subcellular localization, 2:349
 - subunits and isoforms, 2:347, 2:348f
 - synaptic plasticity, 2:349
 - transgenic mice, 2:348–349
- GABA_B receptor, 2:351–353
 - agonists, 2:352–353, 2:352f
 - allosteric regulation, 2:352–353, 2:352f
 - antagonists, 2:353, 2:353f
 - structure and functions, 2:351–352, 2:352f
- glutamic acid decarboxylase (GAD), 2:347
- glycine receptors (GlyRs), 2:419, 2:420, 2:420f
- GTPase-activating proteins (GAPs), 4:13f
- inhibitory synaptic transmission, 2:395
- neurogenesis, 1:199, 1:200f
- physiological functions
 - modulation, 2:723–724, 2:724f
 - in situ* studies, 2:723
 - transgenic mice, 2:724
 - in vitro* studies, 2:724
- serine–threonine phosphatases, 4:222f
- structural characteristics, 2:725, 2:725f
- synthesis
 - GABA transporters (GATs), 2:722, 2:723f
 - transport mechanisms, 2:722, 2:723f
- transport mechanisms, 3:238–239
- γ-carboxylglutamic acid (Gla)
 - amino acid metabolism, 1:93t
- blood-clotting proteases, 3:588–589, 3:589f

- vitamin K-dependent proteins, 4:551, 4:552f
- γδ resolvasin, 1:6–7
- gamma glutamyl transpeptidase (GGT), 3:564
- gangliosides, 4:282, 4:282f
- gap junction (GJ) communication, 2:615, 2:615f
- garden snails, 2:702t
- garlic, 4:533t
- Gas2-related (GAR) domain, 2:628–629, 3:105f, 3:107
- gastic inhibitory peptide (GIP), 2:596
- gastric inhibitory peptide (GIP), 1:117, 1:272, 1:658
- gastricisin, 1:137, 1:138f
- gastric proton–potassium (H⁺–K⁺)-ATPase, 3:676–680
 - acid secretion, 3:679–680
 - characteristics, 3:676, 4:333–334, 4:333f
 - functional residues, 3:678–679
 - structural characteristics
 - α-subunit, 3:676, 3:677f
 - β-subunit, 3:677, 3:677f
 - conformational changes, 3:677f, 3:678
 - kinetic mechanisms, 3:678
 - oligomeric structure, 3:677–678
- gastrin
 - ascorbic acid, 4:531–532
 - calcitonin receptors (CTRs), 1:272
 - calcium-sensing receptor (CaSR), 1:329–330
 - gastrin-releasing peptide (GRP), 3:399
 - peptide amidation, 3:396
- gastrointestinal tract
 - digestion and absorption, 2:354–359
 - basic concepts, 2:354
 - colon nutrient absorption, 2:359
 - esophageal transport, 2:355
 - gastric functions
 - food transit, 2:355, 2:355f
 - gastic lipase, 2:356
 - intra gastric pH, 2:356
 - micronutrient absorption, 2:356
 - protein digestion, 2:356
 - secretory and digestive functions, 2:355–356
 - oral cavity functions, 2:354
- small intestine
 - bile acid secretion and absorption, 2:358
 - carbohydrate digestion and absorption, 2:359
 - cholesterol, 2:358
 - intestinal pH, 2:357
 - lipid digestion and absorption, 2:358
 - nutrient transit, 2:356–357
 - pancreatic exocrine secretion, 2:357–358, 2:357f
 - phospholipids, 2:358
 - protein digestion and absorption, 2:358–359
 - triglyceride (TG), 2:358
- fibroblast growth factors (FGFs), 2:306
- mucins, 3:203

- GATA-1 zinc finger transcription factor, 1:187, 1:188f
- GATA-2 zinc finger transcription factor, 1:187, 1:188f
- GATA-3 zinc finger transcription factor, 1:187–188, 1:188f
- GATA-4 zinc finger transcription factor, 1:184
- GATA-6 zinc finger transcription factor, 1:184f
- GATC sequence
- Dam methyltransferase
 - altered gene expression, 2:42–43
 - chromosome replication, 2:42
 - dam* mutant strains, 2:40
 - mismatch repair, 2:40–42, 2:41f, 2:42f
 - oriC* region, 2:121
 - properties, 2:40
- Gaucher's disease, 4:291–292, 4:291t, 4:507–508, 4:507t
- G-box-binding factor (GBF), 1:564–565, 1:564f
- GDP-fucose (GDP-Fuc), 1:3
- GDP-mannose (GDP-Man), 1:3
- gelatinases, 3:91t, 3:92–93, 3:92f
- gelsolin
 - actin depolymerizing factor (ADF)/cofilin (ADF-H) domain, 1:22f, 1:23–24, 1:277–278
 - amyloid deposits, 1:101t
 - characteristics and functional role, 1:28–29, 1:277–278, 1:277f, 1:278f
- GenBank, 3:615t
- GENE3D database, 3:615t
- Gene Ontology (GO), 3:614, 3:615t
- gene-specific upstream activation sequence (UAS_c)
 - See GAL promoter system
- genetic disorders, 4:569, 4:569t
- genistein, 2:167–168, 2:167f
- genome-wide gene expression analysis, 2:369–374
 - basic concepts, 2:369
- DNA microarrays
 - basic concepts, 2:370–371
 - data analysis, 2:371
 - schematic diagram, 2:371f
- expressed sequence tags (ESTs), 2:369
- next-generation sequencing (NGS) technologies
 - analytical tools, 2:374
 - second-generation sequencing (SGS), 2:371–373, 2:372f
 - third-generation sequencing (TGS), 2:372f, 2:373–374
- serial analysis of gene expression (SAGE), 2:140, 2:369–370, 2:370f
- Geobacter sulfureducens*, 3:331f
- gephyrin, 2:421
- geranylgeranyl diphosphate
 - carotenoid metabolism, 1:504
 - cholesterol biosynthetic pathway, 1:516, 1:517f
 - lipid-modified proteins, 2:737–738
- geranylgeranyl/geranylgeranyl transferases, 2:737–738, 2:737f, 4:242, 4:244f
- geranylgeranyl transferase inhibitors (GGTIs), 2:740
- Gerstmann–Sträussler disease, 1:477t
- gestational diabetes, 1:654
- Gfi-1 zinc finger transcription factor, 1:188–189
- ghrelin
 - adiposity and satiety signal integration, 3:79
 - appetite regulation, 4:303
 - body weight regulation, 3:77, 3:77f
 - functional role, 1:118–119
 - hypothalamic malonyl-CoA, 4:57, 4:58f
 - starvation effects, 4:303
- ghrelin O-acyltransferase (GOAT), 1:118–119
- giant mitochondria, 2:375–376, 2:375f
- Giardia*, 1:134
- Giardia intestinalis*, 2:674
- Giardia lamblia*, 1:37f
- Gibbs free energy, 1:217–218, 1:218f
- Gilberts syndrome, 3:548
- Gin invertase, 1:560–561
- GJC2 gene, 4:272
- Gleason grade, 2:277–278
- glia
 - cerebral metabolism, 1:369, 1:370f, 1:371–373
 - glial fibrillary acidic protein (GFAP), 1:196–197, 1:197f, 2:625–627, 2:632t, 2:634, 3:233
- GLI-family transcription factors
 - liver regeneration, 1:193
 - neurogenesis, 1:199
- glioblastomas, 2:306, 4:362–363, 4:365
- glmS* riboswitch, 4:138f, 4:141
- global genome repair (GGR)
 - damage recognition mechanisms, 4:402
 - nucleotide excision repair (NER)
 - damage recognition, 3:341–342, 3:342f
 - NER factors, 3:342t
 - repair pathways, 3:342f
 - schematic diagram, 4:403f
- global inhibition (LEGI) model, 2:238f, 2:239–240
- globin genes, 2:109
- globo H antigen, 1:2f
- glucagon
 - amino acid metabolism, 1:95–96, 1:96f
 - branched-chain amino acid (BCAA) regulation, 1:245–246
 - calcitonin receptors (CTRs), 1:272
 - carbohydrate-responsive element binding protein (ChREBP), 1:375–376, 1:376f
 - fatty acid synthesis pathway, 2:275–279, 2:276f
 - functional role, 1:117–118
 - starvation effects, 4:302–303, 4:303–304
- glucagon-like peptide-1 (GLP-1), 1:115, 1:116–117, 1:658, 2:596, 2:597f, 3:78–79, 3:79f
- glucoceramidase, 4:291–292
- glucocorticoids
 - adrenocortical steroidogenesis, 3:25–26
 - branched-chain α -ketoacid dehydrogenase complex (BCKDC), 1:247–248
 - glucocorticoid receptor (GR), 4:311f
- glucocorticoid response element (GRE), 4:311f
- gluconeogenesis, 2:384
- leptin regulation, 2:259
- macroautophagy regulation, 1:150t
- nuclear factor kappa B (NF- κ B), 3:305, 3:305t
- starvation effects, 4:302–303, 4:303–304
- glucokinase
 - catalytic reaction mechanisms, 2:543
 - classifications, 2:544
 - evolutionary development, 2:544, 2:545f
 - glucokinase disease, 2:600
 - glucokinase glucose sensor concept, 2:596–598
 - insulin-secreting cells (β -cells), 2:547
 - isozymic diversity, 2:543–544, 2:544t
 - liver glucokinase/glucokinase regulatory protein (GCKR), 2:547
 - sigmoidal kinetics, 2:546–547, 2:546f
 - structural characteristics, 2:544–546, 2:545f
 - sugar phosphorylation, 2:543
- gluconeogenesis
 - carbohydrate-responsive element binding protein (ChREBP), 1:375–376
 - citric acid cycle, 1:105
 - glucagon, 1:117–118
 - importance, 2:386
 - insulin metabolism, 2:603f
 - insulin signal transduction, 2:604f
 - physiological controls
 - dietary effects, 2:383–384
 - fatty acid oxidation, 2:384
 - metabolic pathways, 2:384
 - redox balance, 2:384
 - regulation mechanisms
 - allosteric enzymatic regulation, 2:385, 2:385f
 - covalent modification, 2:385
 - gene expression alterations, 2:385
 - mass action enzymatic regulation, 2:384
 - phosphorylation, 2:385
 - substrate delivery, 2:384
 - regulatory enzymes
 - fructose 1,6-bisphosphate (fructose-1,6-BP), 2:383
 - general discussion, 2:381–383
 - glucose-6-phosphate (glucose-6-P), 2:383
 - phosphoenolpyruvate carboxykinase (PEPCK), 2:382–383
 - pyruvate carboxylase (PC), 2:381–382
 - renal gluconeogenesis, 2:384, 2:385, 2:386f
 - starvation, 4:298, 4:302–303, 4:305–306, 4:306f
- glucosamine-6-phosphate (GlcN6P), 4:137, 4:141
- glucose
 - antioxidant actions, 2:77
 - ascorbic acid, 4:532–533
 - body weight regulation, 4:58f
 - carbohydrate-responsive element binding protein (ChREBP), 1:374–375
 - chemical structure, 2:696f

glucose (*Continued*)

energy metabolism

cardiac metabolism, 3:65–66, 3:66f, 3:68f

cerebral metabolism, 1:368–373

cellular metabolic pathways, 1:371–373

glucose transporters (GLUTs), 1:369–371

metabolic pathways, 1:368, 1:369f, 1:370f

monocarboxylate transporters (MCTs), 1:370f, 1:371

research summary, 1:373

fatty acid synthesis pathway, 2:275–279, 2:276f

fermentation reactions, 2:204–206

free oligosaccharide (fOS) generation, 2:330f

gluconeogenesis, 2:381–386

basic concepts, 2:381, 2:382f

historical research, 2:381

O-glucose

carbohydrate modifications, 2:453–454, 2:454f

classifications, 2:453

epidermal growth factor (EGF)-like repeats, 2:453–454, 2:454f

notch signaling, 2:455

glucose-1-phosphate (glucose-1-P)

glycogen degradation pathways, 2:425, 2:426f, 2:428–429, 2:430f

photosynthetic metabolic pathways, 3:498f

glucose-6-phosphate (glucose-6-P)

carbohydrate-responsive element

binding protein (ChREBP), 1:377f

cardiac metabolism, 3:65–66, 3:66f, 3:68f, 3:69f, 3:70f

cerebral metabolism, 1:369f

characteristics, 2:383

glucokinase, 2:546–547, 2:546f

gluconeogenesis, 2:381, 2:383

glycogen degradation pathways, 2:425, 2:426f, 2:428–429, 2:430f

glycogen storage diseases

glycolytic pathway, 2:443–444, 2:445f

Type Ia glycogen storage disease, 2:431, 2:431t

Type Ib glycogen storage disease, 2:431, 2:431t

hexokinases

hexokinase IV, 2:546–547, 2:546f

isozymic diversity, 2:543–544, 2:544t

mammalian low- K_m hexokinases, 2:546

historical research, 3:391

pentose phosphate pathway

basic concepts, 3:392f

hexose monophosphate shunt, 3:391

nonoxidative segments, 3:391, 3:393f

operation modes, 3:391

oxidative segments, 3:391, 3:392f

quantitation methods, 3:392, 3:394f

reaction mechanisms, 3:391

regulation mechanisms, 3:391–392

phosphotransferase system (PTS), 2:388f

photosynthetic metabolic pathways, 3:498f

thiamine diphosphate (ThDP), 1:207f

glucose phosphate isomerase (GPI), 2:445f

glucose-regulated protein (Grp94), 1:291t, 1:292

glucose-stimulated insulin release (GSIR), 2:596–601

alpha-cell function, 2:600

amino acids, 2:598–600

augmentation pathway (APW), 2:598, 2:598f

beta-cell therapies, 2:600–601, 2:601f

fatty acids, 2:598–600

functional role, 2:596

GDH or K^+ channel-linked

hyperinsulinemic hypoglycemia, 2:600

GK activators (GKAs), 2:597f,

2:600–601, 2:601f

glucokinase disease, 2:600

glucokinase glucose sensor concept,

2:596–598, 2:597f, 2:598f

insulin release stimulation, 2:599f

K^+ channel inhibitors and activators (KCI and KCOs), 2:597f,

2:600–601

neuroendocrine modifications, 2:600

triggering pathway (TPW), 2:598, 2:598f

β 1,3-glucosyltransferase (β 3GlcT), 2:454f

GLUT4

Akt phosphorylation, 2:604–605

ascorbic acid, 4:532–533

cardiac metabolism, 3:65–66, 3:66f, 3:68f, 3:69f, 3:70f

cerebral metabolism, 1:369–371, 1:370f

characteristics, 2:393

distribution and function, 2:392t

insulin resistance mechanisms,

1:651–652, 1:651f

insulin signal transduction, 2:604f

proline-rich tyrosine kinase 2 (PYK2), 2:267

thiazolidinedione (TZD), 3:421–422, 3:422f

glycation pathways, 2:406f

glycolysis, 2:443, 2:444f, 3:65–66, 3:66f, 3:68f, 3:69f

hypothalamic malonyl-CoA, 4:58f

insect biochemistry, 2:590–591

insulin signal transduction, 2:604f

lac operon, 2:694–695

lipogenic gene expression, 2:286–287

mammalian glucose transport,

2:391–394

facilitative glucose transporters (GLUTs), 2:392f

cerebral metabolism, 1:369–371,

1:370f

characteristics, 2:394

classifications, 2:392t

GLUT1, 1:369–371, 1:370f,

2:392–393, 2:392f, 2:392t,

3:65–66, 3:66f, 3:68f, 3:69f,

3:70f, 4:532–533

GLUT2, 1:376–377, 1:377f, 2:392t, 2:393

GLUT3, 1:369–371, 1:370f, 2:392t, 2:393, 4:532–533

GLUT4, 1:651–652, 1:651f, 2:392t, 2:393, 2:604–605, 2:604f,

3:65–66, 3:66f, 3:68f, 3:69f, 3:70f, 4:532–533

GLUT5, 1:369–371, 1:370f, 2:392t, 2:393

GLUT6, 2:392t, 2:393

GLUT7, 2:392t, 2:393

GLUT8, 2:392t, 2:393–394

GLUT9, 2:392t, 2:394

GLUT10, 2:392t, 2:394

GLUT11, 2:392t, 2:394

GLUT12, 2:392t, 2:394

GLUT13, 2:392t, 2:394

GLUT14, 2:392t, 2:394

sodium-dependent glucose transporters (SGLTs), 2:391, 2:392f

metalloproteases, 3:87, 3:87f

microRNA (miRNA)-based regulation, 4:193, 4:194t

N-linked glycosylation, 2:460f, 2:470, 2:470f, 4:204–205

O-linked glycosylation, 2:466f

phosphotransferase system (PTS),

2:387–388

proline-rich tyrosine kinase 2 (PYK2), 2:267

reaction rates, 2:406–407

starvation effects, 4:301–302, 4:307

sugar phosphorylation, 2:543

thiamine diphosphate (ThDP), 1:207f

uncoupling proteins (UCPs), 4:486

See also diabetes

glucosidases

 α -glucosidases

degradation pathways, 2:425, 2:426f

functional role, 3:257

α -glucosidase I/ α -glucosidase II, 3:257, 3:258f, 3:259t

N-linked glycans processing, 3:257–260

ER-associated degradation (ERAD), 2:201

Glc₃Man₉GlcNAc₂ glycan, 3:257, 3:258f, 3:259t

α -glucosidase I/ α -glucosidase II, 3:257, 3:258f, 3:259t

Man₅GlcNAc₂ glycan, 3:257–259, 3:259t

Man₈GlcNAc₂ glycan, 3:258f, 3:259t

Man₉GlcNAc₂ glycan, 3:259t

processing pathways, 3:258f

glucosylceramide (Glc-Cer), 2:316, 2:316t

glucosylphosphoryldolichol (Glc-P-Dol), 2:315–316, 2:315f, 2:316t

O-glucosyltransferase (Poglut), 2:453–454, 2:454f

glucuronic acid, 3:654

glutitocin, 4:516t

GLUT1

ascorbic acid, 4:532–533

cardiac metabolism, 3:65–66, 3:66f, 3:68f, 3:69f, 3:70f

cerebral metabolism, 1:369–371, 1:370f

- characteristics, 2:392–393, 2:392*f*
 digestion and absorption, 2:359
 distribution and function, 2:392*t*
- GLUT2
 carbohydrate-responsive element binding protein (ChREBP) function, 1:376–377, 1:377*f*
 characteristics, 2:393
 digestion and absorption, 2:359
 distribution and function, 2:392*t*
- GLUT3
 ascorbic acid, 4:532–533
 cerebral metabolism, 1:369–371, 1:370*f*
 characteristics, 2:393
 digestion and absorption, 2:359
 distribution and function, 2:392*t*
- GLUT4
 Akt phosphorylation, 2:604–605
 ascorbic acid, 4:532–533
 cardiac metabolism, 3:65–66, 3:66*f*, 3:68*f*, 3:69*f*, 3:70*f*
 cerebral metabolism, 1:369–371, 1:370*f*
 characteristics, 2:393
 digestion and absorption, 2:359
 distribution and function, 2:392*t*
 insulin resistance mechanisms, 1:651–652, 1:651*f*
 insulin signal transduction, 2:604*f*
 proline-rich tyrosine kinase 2 (PYK2), 2:267
- GLUT5
 cerebral metabolism, 1:369–371, 1:370*f*
 characteristics, 2:393
 digestion and absorption, 2:359
 distribution and function, 2:392*t*
- GLUT6
 characteristics, 2:393
 distribution and function, 2:392*t*
- GLUT7
 characteristics, 2:393
 distribution and function, 2:392*t*
- GLUT8
 characteristics, 2:393–394
 distribution and function, 2:392*t*
- GLUT9
 characteristics, 2:394
 distribution and function, 2:392*t*
- GLUT10
 characteristics, 2:394
 distribution and function, 2:392*t*
- GLUT11
 characteristics, 2:394
 distribution and function, 2:392*t*
- GLUT12
 characteristics, 2:394
 distribution and function, 2:392*t*
- GLUT13
 characteristics, 2:394
 distribution and function, 2:392*t*
- GLUT14
 characteristics, 2:394
 distribution and function, 2:392*t*
- glutamate
 anaplerotic pathways, 1:106, 1:106*f*, 1:108*f*, 1:109*f*
 branched-chain amino acids (BCAAs), 1:248
 classifications, 1:92*t*
 cytidine nucleotide formation, 3:715*f*
 DDE transposases, 3:277–278, 3:278*f*
 excitotoxicity, 1:423–424
 functional roles, 1:93*t*
 gamma-aminobutyric acid (GABA)
 synthesis, 2:722, 2:723*f*
 glutamate dehydrogenase (GDH), 1:246, 2:598–600
 glutamate-gated ion channel (GluR-B)
 adenosine deaminases, 4:155–156, 4:156*f*
 adenosine-to-inosine (A-to-I) RNA editing, 4:154–155
 research background, 4:154
 glutamate mutase, 1:158*t*
 glutamate receptors, 2:712–721
 AMPA receptors, 1:425*f*, 2:681–684, 2:714–716, 2:715*t*, 2:716*t*, 2:717–718, 2:717*f*, 2:718*f*
 AMPA-type glutamate receptor (GluR2), 2:681–684, 2:685*f*
 calcium ion (Ca^{2+}) signaling system, 2:719–720, 2:720*f*
 glutamate receptor interacting protein 1 (GRIP1), 2:681–684, 2:685*f*
 ionotropic-glutamate (iGlu) receptors, 2:395, 2:714–716, 2:715*t*, 2:716*t*, 2:717*f*
 metabotropic glutamate (mGlu) receptors, 2:395–398
 characteristics, 2:716*t*
 fast excitatory synaptic transmission, 2:395
 functional role, 2:396–397, 2:714–716, 2:715*t*
 gene families, 2:395–396, 2:396*f*
 molecular structure, 2:717*f*
 neuromodulation, 2:395
 protein structure, 2:717*f*, 2:719
 research background, 2:395
 schematic diagram, 2:396*f*
 therapeutic potential, 2:397
 molecular structure, 2:716–717, 2:716*t*, 2:717*f*, 2:718*f*
 neurodegenerative diseases, 2:720
 physiological/pharmacological properties, 2:714–716, 2:715*t*
 protein structure
 AMPA receptors, 2:717–718, 2:718*f*
 kainic acid receptors (KARs), 2:718
 metabotropic glutamate (mGlu) receptors, 2:717*f*, 2:719
 N-methyl-D-aspartate (NMDA), 2:717*f*, 2:718–719
 transcription regulation, 2:719–720, 2:720*f*
 glycine receptors (GlyRs), 2:423
 G_q family proteins, 2:497*t*
 ionotropic-glutamate (iGlu) receptors, 2:395
 leucine–glutamate cycle, 1:248
 metabotropic glutamate (mGlu) receptors, 2:395–398
 fast excitatory synaptic transmission, 2:395
 functional role, 2:396–397
 gene families, 2:395–396, 2:396*f*
 neuromodulation, 2:395
 research background, 2:395
 schematic diagram, 2:396*f*
 therapeutic potential, 2:397
 phage libraries, 3:434*t*
 photosynthetic metabolic pathways, 3:497–498, 3:498*f*
 renal gluconeogenesis, 2:385, 2:386*f*
 thymidine nucleotide formation, 3:715*f*
 tricarboxylic acid cycle, 4:439*f*
 urea cycle, 4:495*f*
- glutamic acid
 calcium ion (Ca^{2+}) signaling system, 2:712
 glutamic acid decarboxylase (GAD), 2:347
 glycine–phenylalanine–hydroxyproline–glycine–glutamic acid–arginine (GFOGER) sequence, 1:427*t*, 1:428
 urea cycle, 4:495*f*
- glutamine
 anaplerotic pathways, 1:106*f*, 1:108–109
 branched-chain amino acids (BCAAs), 1:248
 classifications, 1:91, 1:92*t*
 functional roles, 1:93*t*
 isoleucine-glutamine (IQ) motif, 4:560
 muscle metabolism, 1:94*f*
 phage libraries, 3:434*t*
 renal gluconeogenesis, 2:385, 2:386*f*
 starvation effects, 4:302, 4:304–305
 synthetic pathways, 3:712, 3:713*f*, 3:714*f*
 urea cycle, 4:495*f*
 vasopressin/oxytocin family, 4:516*t*
- glutamyl-tRNA synthetase, 4:441
 glutamyl-1-semialdehyde (GSA), 2:539–540, 2:540*f*
 glutamyl-prolyl-tRNA synthetase, 4:443
 glutamyl-tRNA synthetase, 2:539–540, 2:540*f*, 4:441
- glutathione
 antioxidant actions, 2:77
 glutathione peroxidases, 2:399–404
 biological functions, 2:403–404
 catalytic reaction mechanisms, 2:399, 3:409–411, 3:410*f*, 3:411*f*
 cellular respiration, 4:407
 future research areas, 2:404
 historical research, 2:399–400
 homologs, 2:402*f*
 kinetic mechanisms, 2:400–401, 2:400*f*, 2:401*f*
 mitochondrial ROS, 1:212, 1:212*f*
 research summary, 2:404
 structural characteristics, 2:401–403, 2:402*f*, 2:403*f*
 substrate specificity, 2:401
 glutathione reductase, 3:411–412, 3:411*f*, 3:715*f*
 glutathione S-transferase (GST)
 antioxidant actions, 4:532
 functional role, 3:411*f*, 4:532
 protein–ligand interaction affinity tags, 1:62

glutathione (*Continued*)

heterotrophic glutathione reductase (GR)-
type FNRs, 2:301
iron-sulfur (FeS) protein biogenesis,
2:663–664
pyridine nucleotide oxidoreductases, 2:311
ryanodine receptors (RyRs), 4:197
glybenclamide, 3:551
glycation, 2:405–411
advanced glycation end products (AGEs)
AGE inhibitors and breakers, 2:409
AGE receptors, 2:407–409
aging effects, 2:409, 2:410f
alternative pathways, 2:407, 2:409f
basic concepts, 2:405
characteristics, 2:407, 2:408f
diabetes, 1:655–656, 2:409–410, 2:410f
disease effects, 2:410
RAGE (receptor for advanced glycation
end products), 2:407–409
alternative pathways, 2:407, 2:409f
basic concepts, 2:405, 2:406f
blood proteins
amino acids, 2:406
extracellular proteins, 2:406
glucose reactivity, 2:406–407
glycated hemoglobin
characteristics, 2:405
functional role, 2:405–406
plasma proteins, 2:406
glycooxidation, 2:407
Maillard reaction, 2:407
glyceraldehyde 3-phosphate (G-3-P)
gluconeogenesis, 2:382f, 2:384
glyceraldehyde-3-phosphate
dehydrogenase (G3PDH)
glycolytic pathway, 2:445f
metabolic regulation, 3:499–500
metabolite channeling, 3:80
myocardial ischemia, 3:67, 3:108–109
glycolysis
anaerobic glycolysis, 2:445f
glycolytic pathway, 2:443–444, 2:445f
pentose phosphate pathway, 3:391, 3:392,
3:393f, 3:394f
photosynthetic carbon dioxide (CO₂)
fixation
carbon pathway, 3:497, 3:499f
metabolic regulation, 3:499–500
regeneration phase, 3:499
thiamine diphosphate (ThDP), 1:207f
glycerate-3-phosphate (C3), 3:494f, 3:495f
glycerate-3-phosphate-kinase, 3:499f
glycerol
fermentation reactions, 2:204–206
glycerol-3-phosphate acyltransferase
(GPAT), 2:415
glycerol-3-phosphate dehydrogenase,
1:319t, 4:85
glycerol dehydratase, 1:156, 1:158t
glycerol phosphate dehydrogenase
(GPDH), 3:80, 4:19f, 4:20
lipolysis regulation, 1:50f
glycerolipids, 2:412–418
background information, 2:412
functional role, 2:414–415

glycerolipid receptors

intracellular receptors, 2:764
plasma membrane receptors, 2:762–764
phosphatidylcholine (PtdCho), 2:412,
2:413f, 2:414–415, 2:416f, 2:417f
phosphatidylethanolamine (PtdEtn),
2:412, 2:413f, 2:414–415, 2:416f
phosphatidylglycerol (PtdGro), 2:412,
2:413f, 2:414–415, 2:416f
phosphatidylinositol (PtdIns)
biosynthetic pathways, 2:416f
functional role, 2:414–415
phosphatidylinositol-3,4,5-triphosphate
(PIP₃), 3:447–451
binding proteins, 3:450
biosynthetic pathways, 3:449–450
cellular targets, 3:449f, 3:450–451
characteristics, 3:447
exosomes and microvesicles (EMVs),
2:262
phagocytosis regulation, 3:439–440
phosphatase and tensin (PTEN)
homolog, 1:563–564, 2:237f,
2:581f, 3:195–196, 3:439–440,
3:448f, 3:450
phosphatases, 3:450
SHIP1/SHIP2 phosphatases,
2:236–237, 3:439–440, 3:448f,
3:450
phosphatidylinositol-4,5-bisphosphate
(PIP₂), 3:447–451
actin depolymerizing factor (ADF)/
cofilin (ADF-H) domain, 1:22,
1:22f
affinity and sensitivity, 2:747
analytical methodologies, 2:746t
biosynthetic pathways, 3:447–448
cellular targets, 3:449, 3:449f
characteristics, 3:447
chemoattractant signaling,
2:236–238, 2:237f
cyclic adenosine monophosphate
(cAMP) receptors (cARs),
1:563–564
exosomes and microvesicles (EMVs),
2:262
functional role, 2:745–746, 2:748
glycogen synthase kinase-3 (GSK3),
2:435f
G protein-coupled receptor kinases
(GRKs), 2:485t
G-protein-coupled receptors (GPCRs),
4:225f
G_q family proteins, 2:497–498,
2:498f
hydrolysis mechanisms, 3:448–449
insulin signal transduction, 2:604,
2:604f
intracellular regulation overlap, 2:747,
2:747f
ion channel activity, 2:746–747,
2:746f, 2:746t
ion channel modulation, 2:746–747,
2:748
metabolic pathways, 3:448f
molecular structure, 3:448f, 3:460f

phospholipase C (PLC), 3:467–468,
3:468f
phospholipase D (PLD), 3:474–475,
3:474f
plectin, 2:625–627
regulation specificity, 2:747–748
research summary, 2:748
serotonin receptors, 4:225f, 4:226
transport modulation, 2:746–747
type I PIP kinases, 3:460–461, 3:460f
type II PIP kinases, 3:460f, 3:461
voltage-sensitive phosphatases, 2:746f
structural characteristics, 2:412, 2:413f
phosphatidylserine (PtdSer), 2:412,
2:413f, 2:414–415, 2:416f
phospholipid remodeling, 2:417f, 2:418
structural characteristics
diacylglycerol (DAG), 2:412–414, 2:414f
ether linked glycerolipids, 2:413, 2:414f
general discussion, 2:412–414
general structural features, 2:413f
sn-glycerol-3-phosphate, 2:412–414,
2:413f
lyso-phospholipids, 2:413–414, 2:415f
phosphatidic acid (PtdOH), 2:412–414,
2:413f
triacylglycerol (TAG), 2:412–414, 2:414f
synthesis process
anionic glycerophospholipids,
2:415–416, 2:416f
bioenergetics, 2:415
ether lipids, 2:417–418
neutral zwitterionic glycerolipids,
2:416–417, 2:416f
phosphatidic acid synthesis, 2:415,
2:416f
uncharged glycerolipids, 2:416–417,
2:416f
glyceroneogenesis, 1:105
glycerophospholipids, 2:733, 2:734f
glycinamide ribonucleotide (GAR)
transformylase, 2:325, 2:325f
glycine
arginine-glycine-aspartic acid (RGD)
peptide, 1:437–438, 1:439f, 2:608
classifications, 1:91, 1:92t
collagens
signature (Gly-X-Y)_n sequence
basic concepts, 1:545
evolutionary origins, 1:548
schematic diagram, 1:546f, 1:547f
self-assembly control, 1:546
cytochrome *c*, 1:600–601
elastin, 2:192
glycine decarboxylase, 1:142–143
glycine-phenylalanine-hydroxyproline-
glycine-glutamic acid-arginine
(GFOGER) sequence, 1:427t, 1:428
glycine receptors (GlyRs), 2:419–424
allosteric modulation
alcohol, 2:423
anesthetics, 2:423
background information, 2:423–424
cannabinoids, 2:423
glutamate, 2:423
ivermectin, 2:423

- neurosteroids, 2:423
- posttranslational modifications, 2:422*f*, 2:423–424
- zinc (Zn), 2:422*f*, 2:423
- characteristics, 2:419
- disease connections
 - animal mutants, 2:424
 - human disorders, 2:424
- functional role
 - central nervous system (CNS), 2:419
 - excitatory function, 2:420
 - general discussion, 2:419
 - long-term synapse plasticity, 2:420
 - mixed GABA and glycinergic synapses, 2:420, 2:420*f*
 - motor control, 2:419, 2:420*f*
 - sensory processing, 2:419, 2:420*f*
 - synaptic versus extrasynaptic inhibition, 2:419–420, 2:420*f*
- glycine receptor (GlyR) channels
 - heterologous expression, 2:421
 - isoforms, 2:421
 - purification and isolation processes, 2:420–421
 - spatial and developmental expression, 2:421
- pharmacological properties
 - agonistic amino acids, 2:423
 - allosteric modulation, 2:423–424
 - channel block antagonism, 2:423
 - strychnine antagonism, 2:421–423
- phylogenetic relationships, 2:422*f*
- principal properties, 2:419, 2:420*f*
- structural characteristics
 - β -subunit formation, 2:420*f*, 2:421
 - domain structures, 2:421, 2:422*f*
 - ion channel function, 2:421, 2:422*f*
 - molecular model, 2:421
 - postsynaptic receptor clustering, 2:420*f*, 2:421
- glycine-serine-arginine (GSR) domain, 3:105*f*, 3:107
- insect biochemistry, 2:590–591
- phage libraries, 3:434*t*
- photosynthetic metabolic pathways, 3:495, 3:497–498, 3:498*f*
- riboswitches, 4:137
- serine–glycine interconversion, 2:324–325
- starvation effects, 4:302
- urea cycle, 4:495*f*
- vasopressin/oxytocin family, 4:516*t*
- Glycine max*
 - lectins, 2:702*t*
 - matrix metalloproteinases (MMPs), 3:90
- glycine–phenylalanine–hydroxyproline–glycine–glutamic acid–arginine (GFOGER) sequence, 1:427*t*, 1:428
- glycoaldehyde, 2:409*f*
- glycogen, 3:111
 - functional roles
 - glucose homeostasis, 2:428
 - metabolic role, 2:425–426, 2:427*f*
 - muscle glycogen, 2:426
 - glycogen synthase kinase-3 (GSK3), 2:434–438
 - adipocyte development
 - transcriptional control, 1:55, 1:55*f*
 - Wnt pathway, 1:53–54, 1:53*f*
 - diabetes, 2:437–438
 - domain structures, 2:434, 2:435*f*
 - inactivation consequences, 2:438
 - insulin signal transduction, 1:198–199, 1:200*f*, 2:604–605, 2:604*f*
 - knockout mice models, 2:438
 - mitochondrial permeability transition pore, 3:165–166
 - myocardial ischemia, 3:111
 - neurogenesis, 1:199
 - neurological diseases, 2:437
 - physiological functions, 2:436*f*, 2:437
 - regulatory mechanisms, 2:434–436, 2:435*f*, 2:436*f*
 - signaling pathways, 1:564–565, 1:564*f*
 - substrates, 2:436–437, 2:436*t*
 - Wnt/ β -catenin signaling pathway, 1:263*f*, 2:434–435, 2:436*f*
- metabolic characteristics
 - degradation pathways, 2:425, 2:426*f*, 2:428–429, 2:430*f*
 - general discussion, 2:425
 - metabolic control mechanisms, 2:429
 - pathway defects and consequences, 2:431*t*, 2:432–433
 - synthetic pathways, 2:425, 2:426*f*, 2:428, 2:430*f*
- related diseases
 - diabetes, 2:426–427
 - glycogen storage diseases, 2:428–433
 - characteristics, 2:426, 2:432–433
 - classifications, 2:427*t*, 2:431*t*
 - genetic mutations, 2:383
 - research summary, 2:433
 - synthetic and degradation pathways, 2:426*f*
- structural characteristics, 2:425, 2:426*f*, 2:428, 2:429*f*
- glycogenolysis/glycogenesis
 - carbohydrate-responsive element binding protein (ChREBP), 1:375–376
 - starvation, 4:302–303
- glycohydrolases, 1:176–177, 1:177*f*
- glycolate, 3:498*f*
- glycolate oxidase, 2:312
- glycolipids
 - chloroplasts, 1:508
 - fatty acyl-coenzyme A (CoA) synthetases, 2:291*f*
 - lipid rafts, 2:741
- N*-glycolylneuraminic acid (Neu5Gc), 4:230–231
- glycolysis, 2:443–447
 - anaerobic glycolysis, 2:444–445, 2:445*f*
 - carbohydrate-responsive element binding protein (ChREBP), 1:375–376, 1:376–377, 1:377*f*
 - cardiac metabolism, 3:65–66, 3:66*f*, 3:68*f*, 3:69*f*, 3:70*f*
 - cerebral metabolism, 1:369*f*, 1:371–373
 - control reactions
 - adenosine triphosphate (ATP) turnover, 2:445
 - general discussion, 2:445–446
 - malignant tumor growth, 2:446
 - oxidation effects, 2:445–446, 2:446*f*
 - creatine kinase/phosphocreatine circuit, 3:81, 3:82*f*
 - functional role
 - adenosine triphosphate (ATP) production, 2:443, 2:444*f*
 - glucose oxidation preparatory pathway, 2:443, 2:444*f*
- glycolytic pathway
 - energy-trapping stage, 2:444, 2:445*f*
 - priming stage, 2:443, 2:445*f*
 - splitting stage, 2:443–444, 2:445*f*
- lactic acidosis, 2:446–447
- pyruvate kinase (PK), 1:374, 3:719
- thiamine diphosphate (ThDP), 1:207*f*
- glycophorin, 1:34–35, 1:34*f*
- glycoproteins
 - cell–matrix interactions, 1:426
 - folding processes, 2:448–452
 - basic principles, 2:448–449
 - chaperones, 4:205–206
 - general discussion, 2:448
 - N*-glycan processing reactions, 2:200–201, 2:449–450, 2:449*f*, 2:450*f*, 4:204–205
 - quality control mechanisms, 2:329–330, 2:451–452, 2:451*f*, 4:205–206
 - UDP-Glc: glycoprotein glucosyltransferase (UGGT), 2:200–201, 2:449*f*, 2:450, 2:450*f*, 2:451*f*, 2:457, 2:458*f*, 2:460*f*, 2:461*f*
- glycoprotein hormone receptors, 4:387–391
 - activation mechanisms
 - binding mechanisms, 4:390
 - cellular pathways, 4:388–390
 - G proteins, 4:388–390
 - disease connections
 - autoantibodies, 4:390
 - naturally occurring mutations, 4:390–391
- follicle-stimulating hormone receptor (FSHR)
 - dimerization/oligomerization, 4:388
 - extracellular domains, 4:388
 - functional role, 4:387–388
 - genomic organization, 4:388
 - G protein activation, 4:388–390
 - naturally occurring mutations, 4:390–391
 - structural organization, 4:389*f*
- genomic organization, 4:388
- luteinizing hormone receptor (LHR)
 - dimerization/oligomerization, 4:388
 - extracellular domains, 4:388
 - functional role, 4:387
 - genomic organization, 4:388
 - G protein activation, 4:388–390
 - naturally occurring mutations, 4:390–391
 - structural organization, 4:389*f*
- naturally occurring mutations
 - characteristics, 4:390–391
 - follicle-stimulating hormone receptor (FSHR), 4:390–391

- glycoproteins (*Continued*)
 luteinizing hormone receptor (LHR), 4:390
 thyroid-stimulating hormone receptor (TSHR), 4:390
 structural organization
 dimerization/oligomerization, 4:388
 extracellular domains, 4:388, 4:389f
 serpentine regions, 4:388
 thyroid-stimulating hormone receptor (TSHR)
 autoantibodies, 4:390
 extracellular domains, 4:388
 functional role, 4:387
 genomic organization, 4:388
 G protein activation, 4:388–390
 naturally occurring mutations, 4:390
 structural organization, 4:389f
lectins, 2:704–705
mucins, 3:200–204, 3:205–208
 background information, 3:205
 characteristics, 3:200
 disease connections
 cancer, 3:203–204
 gastrointestinal disorders, 3:203
 ocular surfaces, 3:203
 pulmonary disorders, 3:203
 domain structures
 adhesion-associated domains, 3:201f, 3:202
 AP-2 domain, 3:201f, 3:202
 Cys subdomain (CS) domain, 3:201f, 3:203
 cystine knot (CK)-like domains, 3:201f, 3:203
 cytoplasmic domains, 3:201f, 3:202
 D-like domains, 3:201f, 3:202–203
 epidermal growth factor (EGF)-like domains, 3:201f, 3:202
 SEA-like domains, 3:201f, 3:202
 secreted mucins, 3:201f, 3:202
 transmembrane domains, 3:201f, 3:202
 embryo implantation process, 3:205, 3:206f
 expression and transcriptional control, 3:206–208, 3:208f
 family classifications
 characteristics, 3:201, 3:201f
 tethered mucins, 3:201
 functional role, 3:200, 3:207t, 3:208–209
 glycosylation
 definition, 3:200–201
 N-linked glycosylation, 3:201
 O-linked glycosylation, 3:200–201
 research summary, 3:209
 transmembrane proteins, 3:207t
 transport and metabolism regulation, 3:208–209
 variable number tandem repeats (VNTRs), 3:201f, 3:202
O-linked glycans, 2:453–456
 characteristics, 2:465, 2:466f
 classifications, 2:453
 O-fucose
 carbohydrate modifications, 2:453, 2:454f
 classifications, 2:453
 epidermal growth factor (EGF)-like repeats, 2:453, 2:454f
 notch signaling, 2:455
 protein O-fucosyltransferase I (Pofut1), 2:453, 2:454f, 2:456
 thrombospondin type 1 repeats (TSRs), 2:454, 2:455f, 2:456
 uPA–uPA receptor signaling, 2:455
 functional role
 background information, 2:455–456
 α -dystroglycan function, 2:454f, 2:456
 notch signaling, 2:455
 uPA–uPA receptor signaling, 2:455
O-glucose
 carbohydrate modifications, 2:453–454, 2:454f
 classifications, 2:453
 epidermal growth factor (EGF)-like repeats, 2:453–454, 2:454f
 notch signaling, 2:455
O-mannose
 carbohydrate modifications, 2:454, 2:454f
 classifications, 2:453
 α -dystroglycan function, 2:454f, 2:456
mucins, 3:206–208
 synthetic pathways, 2:465–466
O-xylose, 2:453
plants, 2:465–468
 functional role analyses
 future research areas, 2:467
 genetic studies, 2:467
 immunohistochemistry, 2:467
 structural studies, 2:467
 glycan classifications
 N-linked glycans, 2:465, 2:466f
 O-linked glycans, 2:465, 2:466f
 multigene families, 2:467
 O-linked glycans
 characteristics, 2:465, 2:466f
 synthetic pathways, 2:465–466
glycosaminoglycans (GAGs)
 congenital disorders, 2:475
 proteoglycans
 chondroitin sulfate (CS)/dermatan sulfate (DS), 3:654–655, 3:655f
 general discussion, 3:654–656
 heparan sulfate (HS)/heparin, 3:655–656, 3:655f
 hyaluronan, 3:654, 3:655f
 keratan sulfate (KS), 3:655f, 3:656
 sugar building blocks, 3:654
glycosidases, 4:504f
glycoside–pentoside–hexuronide:cation symporter family, 2:388
glycosomes, 3:425
glycosphingolipids (GSLs)
 biosynthesis, 4:283f, 4:284
 catabolic reactions
 α -galactosidase A, 4:292
 arylsulfatase A, 4:291
 β -galactosidases, 4:291
 β -glucocerebrosidase, 4:291–292
 hexosaminidases A and B, 4:292
 cell adhesion, 2:439–442
 carbohydrate-to-carbohydrate interaction, 2:440f
 clustered GSLs, 2:439, 2:440f, 2:441f
 trans-interaction, 2:440f
 characteristics and functional role, 4:290–292
 classifications, 4:282, 4:282f
 cluster formation
 glycosynapse assembly, 2:439, 2:440f, 2:441f
 GSL–GSL-binding protein interactions
 human neuroblastoma adhesion (GM1-to-galectin-1 interaction), 2:442
 neuronal axon binding to myelin (ganglioside–siglec-4 interaction), 2:441
 neutrophil/myelogenous leukemia HL60 cell–E-selectin adhesion (fucosyl-poly-LacNAc ganglioside interaction), 2:442
 renal cell carcinoma TOS1 cell–peripheral blood mononuclear cell aggregation (disialoganglioside–siglec-7 interaction), 2:441–442
 GSL-to-GSL *cis*-interaction, 2:439, 2:440f
 GSL-to-GSL interactions
 carcinoma F9 cell autoaggregation (Le^x-to-Le^x interaction), 2:441
 human embryonal carcinoma autoaggregation (Gb4-to-nLc₄/Gb4-to-Gb5 interaction), 2:441
 mouse melanoma B16 cell adhesion to mouse endothelial cells (GM3-to-LacCer/GM3-to-Gg3Cer interaction), 2:440–441
 rainbow trout sperm binding to egg ((KDN)GM3 interaction), 2:441
 lipophilic signal transduction, 2:439
 flippase-mediated transbilayer movement, 2:314, 2:315f
 saposins, 4:291
 storage disorders, 4:290–291, 4:290f, 4:291t
glycosylation
 asparagine, 2:449–450, 2:449f, 2:450f
 congenital disorders, 2:469–475
 clinical features, 2:471t, 2:474
 congenital muscular dystrophies, 2:471t, 2:474–475
 dolichol phosphate (Dol-P), 2:457, 2:458f
 effective therapies, 2:474
 future research areas, 2:475
 glycosaminoglycan (GAG), 2:475
 N-glycosylation process, 2:457, 2:458f, 2:470, 2:470f
 inherited disorders, 2:470–473, 2:471t
 laboratory diagnoses, 2:473
 mutation effects, 2:471t, 2:473–474
 proteoglycan defects, 2:475
 definition, 3:200–201
 lectins, 2:704–705

- N-linked glycosylation
 calcium-sensing receptor (CaSR), 1:327
 endoplasmic reticulum (ER)
 congenital disorders, 2:470, 2:471*t*
 glycosylation processes, 2:457, 2:458*f*
 protein synthesis and folding,
 4:204–205, 4:205*f*
 Golgi apparatus, 2:449–450, 2:449*f*,
 2:450*f*, 2:480, 4:207
 mammalian glucose transport,
 2:392–393
 mucins, 3:201
 plant glycoproteins, 2:465, 2:466*f*
 stromal interacting molecule 1 (STIM1),
 1:130*f*
 O-linked glycosylation, 2:465, 2:466*f*,
 2:480, 3:200–201, 4:207
 sugar chain–protein linkages, 2:469,
 2:470*t*
 sugar nucleotide transporters, 4:351
 glycosyl phosphatidylinositol (GPI)
 flippase proteins
 bacterial cell wall complex, 2:316–317
 biosynthesis, 2:316–317
 flippase-mediated transbilayer
 movement, 2:314, 2:315*f*
 glucosylceramide (Glc-Cer), 2:316,
 2:316*t*
 glycoprotein folding processes, 2:448–449
 glycosylphosphatidyl inositol (GPI)
 anchors, 2:737*f*, 2:737*t*, 2:738, 3:91*t*,
 3:92*f*, 3:93
 lipid rafts, 2:741, 2:742*f*, 2:743
 plant glycoproteins, 2:467, 2:467*f*
 glycosylphosphatidylinositol (GPI)-
 anchored proteins, 2:476–478
 biosynthesis
 attachment mechanisms, 2:476–477
 biosynthetic pathways, 2:476, 2:477*f*
 characteristics, 2:476
 functional roles, 2:478
 mammalian cells, 2:477–478
 parasitic protozoa, 2:477–478
 structural characteristics, 2:476, 2:477*f*
 yeasts, 2:477–478
 glycosyl transferases, 2:459*f*, 2:466, 3:231,
 4:207
 glyoxysomes, 3:425
 GM1 gangliosidosis, 4:291, 4:291*t*
 go-ichi-ni-san (GINS) proteins, 1:401–402,
 1:402*t*, 1:404*f*
 goldfish, 1:552*f*
 Golgi apparatus, 2:479–481
 actin-related proteins (Arps), 1:38*t*
 basic concepts, 2:479
 cholesterol biosynthetic pathway,
 1:517–519, 1:518*f*
 congenital disorders of glycosylation
 (CDG), 2:470–473, 2:471*t*
 flippase-mediated transbilayer movement,
 2:314, 2:315*f*, 2:316*t*
 glucosylceramide (Glc-Cer), 2:316, 2:316*t*
 Golgi α -mannosidase I, 3:258, 3:258*f*,
 3:259*t*
 Golgi α -mannosidase II, 3:258*f*, 3:259,
 3:259*t*, 4:504*f*
 intracellular calcium ion (Ca²⁺) stores,
 2:482, 2:574
 kinesin superfamily proteins (KIFs), 2:684*f*
 lipogenic gene expression, 2:286–287
 morphology
 cell polarity, 2:480
 cisternae, 2:479–480, 2:480*f*
 golgin proteins, 2:479
 microtubule organization, 2:480
 stacked morphology, 2:479, 2:480*f*
 subcellular localization, 2:479–480,
 2:480*f*
 N-linked glycoproteins
 flippase-mediated transbilayer
 movement, 2:314, 2:315*f*
 glycosidase digestion, 2:457–458,
 2:459*f*
 N-linked glycosylation, 2:449–450,
 2:449*f*, 2:450*f*, 2:480
 process mechanisms, 3:257
 structural remodeling, 2:457, 2:459*f*,
 2:460*f*, 2:461*f*
 O-linked glycosylation, 2:466, 2:480,
 4:207
 phosphatidic acid (PA), 3:476
 phosphatidylinositol 4-kinase (PI4K)
 pathway, 3:459–460
 protein processing
 protein glycosylation, 2:480
 proteolytic processing
 Alzheimer's disease, 2:481
 general discussion, 2:480–481
 prohormone processing, 2:480–481
 sterol regulatory element binding
 protein (SREBP), 2:481
 transcription factor regulation, 2:481
 proteoglycans, 3:659–660, 3:659*f*
 research summary, 2:482
 secretory pathway, 4:203–209
 compartmental organization, 4:203,
 4:204*f*
 microtubule-based motor proteins,
 4:208, 4:208*f*
 targeted proteins
 characteristics, 4:203
 constitutive secretion, 4:208
 entry mechanisms, 4:203, 4:205*f*
 integral membrane proteins, 4:204,
 4:205*f*
 protein exit, 4:207
 regulated secretion, 4:208
 signal recognition particles (SRPs),
 4:203, 4:205*f*
 soluble proteins, 4:204, 4:205*f*
 trans-Golgi network (TGN), 2:479–480,
 2:480*f*, 4:207, 4:208, 4:208*f*
 transport mechanisms
 protein localization, 4:206–207,
 4:206*f*
 protein modification, 4:207
 protein transport, 4:206*f*, 4:207
 resident enzyme localization, 4:207
 retinoid X receptor (RXR), 4:207
 vesicular transport, 4:206, 4:206*f*
 sugar nucleotide transporters, 4:350–351
 genetic mutations, 4:350
 macromolecular glycosylation
 regulation, 4:351
 structural characteristics, 4:350–351
 topographic characteristics, 4:351
 transport mechanisms, 4:350, 4:351*f*
 transport mechanisms
 cisternal progression, 2:481
 protein sorting
 basic concepts, 2:481–482
 KDEL receptor, 2:482
 mannose-6-phosphate receptor, 2:482
 protein aggregation, 2:482
 receptor-mediated sorting, 2:481–482
 vesicular transport, 2:481
 vacuole transport pathways, 4:503–506,
 4:505*f*
 gonadotropin-releasing hormone (GnRH)
 G_q family proteins, 2:496–497, 2:497*t*
 peptide amidation, 3:396
 gout, 2:312
 G-protein-coupled receptors (GPCRs)
 adrenergic receptors, 1:57–60
 action mechanisms, 1:57–58
 characteristics and functional role, 1:57
 classifications
 alpha-1 adrenergic receptors, 1:58,
 1:58*f*
 alpha-2 adrenergic receptors, 1:58,
 1:58*f*
 beta adrenergic receptors, 1:58*f*, 1:59
 characteristics, 1:57–58
 epinephrine
 branched-chain amino acid (BCAA)
 regulation, 1:245–246
 characteristics and functional role,
 1:57
 starvation effects, 4:303–304
 norepinephrine
 ascorbic acid, 4:531–532
 characteristics and functional role,
 1:57
 polymorphisms, 1:60
 regulation mechanisms, 1:59
 signal transduction pathways, 1:59
 allosteric proteins, 1:74
 anandamide synthesis, 2:194
 angiotensin, 1:112*t*, 2:344*t*, 2:345*f*
 arrestins, 1:496*f*
 blood-clotting proteases, 3:586*f*, 3:587
 bradykinin, 2:344*t*
 calcitonin receptor-like receptor (CLR),
 1:266, 1:267*f*
 calcitonin receptors (CTRs)
 distribution, 1:272
 isoforms, 1:272
 protein structure, 1:272, 1:272*f*
 regulatory mechanisms, 1:272
 signaling pathways, 1:272
 calcium/calmodulin-dependent protein
 kinase II (CaMKII), 1:297–298, 1:298*f*
 calcium-sensing receptor (CaSR),
 1:326–330
 biochemical characteristics
 agonists, 1:328
 amino acids, 1:327
 cell surface dimers, 1:327

G-protein-coupled receptors (GPCRs)

(Continued)

- C-tail, 1:327, 1:327f
- extracellular domains, 1:326–328, 1:327f
- extracellular loops, 1:327f
- functionally important mutations, 1:328
- intracellular loops, 1:327, 1:327f
- intracellular signaling pathways, 1:328
- N-linked glycosylation, 1:327
- protein kinase C (PKC) feedback regulation, 1:328, 2:583, 2:584f
- related receptors, 1:326–328
- structural characteristics, 1:326–328, 1:327f
- calcium homeostasis
 - bone, 1:329
 - kidneys and intestines, 1:329
 - parathyroid, 1:329
 - regulatory role, 1:328–329, 1:329f
- functional role
 - calcium homeostasis, 1:328–329, 1:329f
 - importance, 1:326
 - nonhomeostatic functions, 1:329–330
- pharmaceutical considerations
 - calcilytics, 1:329f, 1:330
 - calcimimetics, 1:329f, 1:330
 - general discussion, 1:330
- cannabinoid receptors, 2:195–196
- characteristics, 1:493–494
- chemoattractant signaling, 1:496f, 2:236–238, 2:237f
- chemokine receptors, 1:480–485
 - desensitization mechanisms, 1:482–483, 1:483f
 - functional roles
 - angiogenesis, 1:484
 - central nervous system (CNS), 1:484–485
 - chemotaxis, 1:483–484
 - human immunodeficiency virus (HIV-1), 1:484
 - tumor growth and metastasis, 1:484
 - internalization, 1:483
 - intracellular signaling pathways, 1:480–482, 1:482f
 - intracellular trafficking, 1:482–483, 1:483f
 - ligand binding, 1:480–482, 1:481f, 1:482f
 - subfamily classifications, 1:480, 1:481t
- chemotactic peptide receptors
 - characteristics, 1:494
 - ligands, 1:494
- complement receptors
 - C3a receptor, 1:494, 1:495, 1:496f
 - C5a receptor, 1:494, 1:495, 1:496f
 - effector proteins, 1:495, 1:496f
 - G proteins, 1:495, 1:496f
- cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:562–566
 - affinities, 1:562–563, 1:563f
 - desensitization mechanisms, 1:565
 - expression and developmental stages, 1:562–563, 1:563f
 - extracellular levels, 1:562–563, 1:563f
 - gradient sensing, 1:565–566
 - molecular identification, 1:562–563
 - schematic diagram, 1:563f
 - signaling pathways
 - characteristics, 1:563–565
 - G-protein-dependent pathways, 1:563–564, 1:564f
 - G-protein-independent pathways, 1:564–565, 1:564f
 - signaling mechanisms, 1:565
- dopamine receptors, 2:169–173
 - D1 and D2 receptor-selective drugs, 2:171
 - receptor distribution, 2:172–173
 - receptor signaling
 - general discussion, 2:171–172
 - G protein coupling, 2:171, 2:172t
 - heterotrimeric G proteins, 2:171
 - signaling pathways, 2:172
 - receptor subtypes
 - characteristics, 2:172t
 - D1 and D2 subtypes, 2:169, 2:170f
 - DNA sequences, 2:169–171, 2:170f
 - general discussion, 2:169–171
 - molecular cloning, 2:169–171, 2:170f
 - receptor variants, 2:172t, 2:173
 - research background, 2:169
 - structural characteristics, 2:170f, 2:171
 - therapeutic uses, 2:173
- drug discovery, 1:496
- eicosanoid receptors, 2:188–191
 - action mechanisms, 2:189
 - biosynthetic pathways
 - arachidonic acid (AA), 2:188–189
 - cyclooxygenase (COX) pathway, 2:189
 - epoxygenase pathway, 2:189
 - lipoxigenase pathway, 2:188–189
 - characteristics, 2:188
 - lipoxigenase metabolite receptors
 - characteristics, 2:189
 - 5-oxo-EETE receptors, 2:189
 - leukotriene receptors, 2:189
 - lipoxin receptors, 2:189
 - peroxisome proliferator-activated receptors (PPARs), 2:191
 - prostaglandin receptors, 2:189–190, 2:190f
- endothelin, 2:344t, 2:345f
- GABA_A receptors, 2:347
- galanin, 2:344t
- glycogen synthase kinase-3 (GSK3), 2:434, 2:435f
- glycoprotein hormone receptors, 4:387–391
 - activation mechanisms
 - binding mechanisms, 4:390
 - cellular pathways, 4:388–390
 - G proteins, 4:388–390
 - disease connections
 - autoantibodies, 4:390
 - naturally occurring mutations, 4:390–391
- follicle-stimulating hormone receptor (FSHR)
 - dimerization/oligomerization, 4:388
 - extracellular domains, 4:388
 - functional role, 4:387–388
 - genomic organization, 4:388
 - G protein activation, 4:388–390
 - naturally occurring mutations, 4:390–391
 - structural organization, 4:389f
- luteinizing hormone receptor (LHR)
 - dimerization/oligomerization, 4:388
 - extracellular domains, 4:388
 - functional role, 4:387
 - genomic organization, 4:388
 - G protein activation, 4:388–390
 - naturally occurring mutations, 4:390
 - structural organization, 4:389f
- naturally occurring mutations
 - characteristics, 4:390–391
- follicle-stimulating hormone receptor (FSHR), 4:390–391
- luteinizing hormone receptor (LHR), 4:390
- thyroid-stimulating hormone receptor (TSHR), 4:390
 - structural organization
 - dimerization/oligomerization, 4:388
 - extracellular domains, 4:388, 4:389f
 - serpentine regions, 4:388
- thyroid-stimulating hormone receptor (TSHR)
 - autoantibodies, 4:390
 - extracellular domains, 4:388
 - functional role, 4:387
 - genomic organization, 4:388
 - G protein activation, 4:388–390
 - naturally occurring mutations, 4:390
 - structural organization, 4:389f
- G protein-coupled receptor kinases (GRKs), 2:483–491
 - carboxyl tail, 4:363–364, 4:363f
 - disease connections, 2:487
 - G_q family proteins, 2:498
 - olfactory receptors, 3:349, 3:349f
 - phosphorylation, 2:483–486, 2:485t
 - regulation mechanisms, 2:486
 - regulatory functions, 2:483
 - signaling mechanisms, 1:565, 2:345f
 - structural characteristics, 2:484f, 2:486–487
 - subfamilies, 2:483
- heterotrimeric G proteins, 3:439, 3:608–609, 3:608f
- inositol trisphosphate (IP₃)/calcium ion (Ca²⁺) signaling system, 2:583, 2:584f
- lipid-modified proteins, 2:739
- lysophosphatidic acid, 2:344t, 2:345f
- metabotropic glutamate (mGlu) receptors, 2:395–398
 - fast excitatory synaptic transmission, 2:395
 - functional role, 2:396–397

- gene families, 2:395–396, 2:396f
 neuromodulation, 2:395
 research background, 2:395
 schematic diagram, 2:396f
 therapeutic potential, 2:397
- multiple opioid receptors, 3:357
- muscarinic acetylcholine receptors, 3:215–217
 acetylcholine, 3:215
 calcium ion (Ca^{2+}) signaling, 2:720f
 gene families, 3:215, 3:216t
in vivo actions
 central nervous system (CNS), 2:720f, 3:217
 peripheral nervous system, 3:216–217
 signal transduction
 guanosine diphosphate (GDP)
 dissociation, 3:215
 ion channels, 3:216
 mitogen-activated protein kinase (MAPK), 3:216
 second messenger pathways, 3:215–216
 structural characteristics
 coupling regions, 3:216
 ligand binding, 3:216
 transmembrane topology, 3:216
- neurokinin-1 receptor (NK1R)
 carboxyl tail, 4:363–364, 4:363f
 disease connections
 affective disorders, 4:364–365
 depression, 4:364–365
 emesis, 4:364
 glioblastomas, 4:365
 research background, 4:364–365, 4:364f
 full-length NK1R, 4:361–363, 4:362f
 functional role, 4:361, 4:362f
 truncated NK1R, 4:363
- neurotransmitter transporters, 3:239
- olfactory receptors, 3:345–349
 basic concepts, 3:345
 classifications, 3:345–346
 combinatorial versus monospecific
 olfactory processing, 3:347f, 3:348–349, 3:348f
 functional role, 3:345
 gene expression regulation, 3:347, 3:347f
 gene structure/genomic arrangement, 3:346
 information processing pathway, 3:347f
 ligand spectra, 3:347–348
 monogenic and monoallelic expression, 3:346–347, 3:347f
 schematic diagram, 3:346f
 signal transduction, 3:349, 3:349f
- P2Y receptors, 3:375–378
 domain structures, 3:376
 extracellular nucleotide signaling, 3:375–376
 functional classification, 3:376–377, 3:376t
 functional role, 3:375
 lysophosphatidic acid (LPA) receptors, 2:763t
 pharmacological classification, 3:377
 physiological functions, 3:377–378
 palmitoylation, 3:646f
 parathyroid hormone (PTH)/PTH-related protein (PTHrP) receptor (PPR), 3:384–390
 characteristics and functional role, 3:384
 disease connections
 Blomstrand's chondrodysplasia, 3:389–390
 Jansen's metaphyseal chondrodysplasia, 3:389
 research background
 molecular cloning, 3:384, 3:385f
 parathyroid hormone (PTH), 3:384–386
 parathyroid hormone-related protein (PTHrP), 3:386
 research summary, 3:390
 structure–activity relationships
 parathyroid hormone (PTH), 3:386–387
 parathyroid hormone (PTH)/PTH-related protein (PTHrP) receptor (PPR), 3:387–389
 parathyroid hormone-related protein (PTHrP), 3:386–387
- pheromone receptors, 3:444–445
- platelet-activating factor receptor (PAFR), 3:533–537
 general discussion, 3:533
 G_q family proteins, 2:497t
 PAF acetylhydrolase, 3:533–534
 platelet-activating factor (PAF), 3:533–534, 3:534f
 regulation mechanisms, 3:535–536
 signal transduction, 3:536, 3:536f
 structural characteristics, 3:534–535, 3:535f
 transcript structure, 3:534–535, 3:535f
- proteinase-activated receptors (PARs), 3:601–606
 activating and inactivating proteinases, 3:601, 3:602f
 activation mechanisms, 3:601, 3:602f
 agonists, 3:601–602, 3:603t
 antagonists, 3:603–604, 3:603f
 bioassays
 calcium signaling measurement, 3:604
 platelet aggregation, 3:605
 tissue bioassays, 3:604–605
 desensitization mechanisms, 3:602
 detection methods, 3:604
 distribution, 3:604, 3:605t
 domain structures, 3:601, 3:603t
 intermolecular activation, 3:602–603
 internalization, 3:602
 localizations, 3:605t
 PAR-activating peptides (PAR-APs), 3:601–602, 3:602f, 3:603t
 physiological functions, 3:605
 receptor dimerization, 3:602–603
 resensitization, 3:602
 thrombin, 3:601, 3:602f, 3:603t
 receptor regulation, 1:495–496, 1:496f
 schematic diagram, 4:225f
- serotonin receptors, 2:344t, 4:224, 4:225f, 4:225t
 somatostatin (sst) receptors, 4:262–266
 agonists/antagonists, 4:264–265
 biochemical/molecular characterization
 sst receptor structure, 4:263–264
 sst receptor subtypes, 4:263, 4:263t, 4:264f
 characteristics and functional role, 4:262
 desensitization mechanisms, 4:265–266
 isoforms, 4:262–263
 ligand binding, 4:262–263
 regulation mechanisms, 4:265–266
 research summary, 4:266
 signaling pathways, 4:265
 structural characteristics, 4:262–263, 4:263f
 therapeutic uses, 4:264–265
- spingosine-1-phosphate, 2:344t, 2:345f
- substance P, 2:344t
- taste receptors, 4:366–367
- thrombin, 2:344t, 2:345f
- thromboxanes, 2:344t, 2:345f
- thyrotropin, 2:344t
- vasopressin, 2:344t
- Wiskott–Aldrich syndrome protein (WASP) family, 1:29, 1:20, 1:21f, 1:26
See also G proteins
- G proteins
 carboxyl methylation, 3:608–609
 chemoattractant signaling, 2:236–238, 2:237f
 complement receptors, 1:495, 1:496f
 elongation factor G (EF-G), 4:122–127
 EF-G–ribosome interactions, 4:123f, 4:125–126, 4:126f, 4:128–129, 4:129f, 4:421f
 research summary, 4:126
 ribosome recycling, 4:126
 translocation cycle, 4:125, 4:132–133
 elongation factor Tu (EF-Tu), 4:122–127
 EF-Tu/EF-G–ribosome interactions, 4:123f, 4:128–129, 4:129f, 4:421f
 guanosine triphosphate (GTP)
 hydrolysis, 4:123–125, 4:124f, 4:420–421, 4:421f
 initial selection phase, 4:123, 4:124f
 research summary, 4:126
 substrate proofreading, 4:125
- $\text{G}_{12}/\text{G}_{13}$ family, 2:343–346
 cellular functions, 2:343–344
 characteristics, 2:343
 guanine-nucleotide exchange factors (GEFs), 2:344, 2:345f
in vivo functions, 2:344–346, 2:345t
 protein interactions, 2:344, 2:345f
 receptor-mediated activation, 2:343, 2:344t
- G_i family, 2:377–380
 adenylyl cyclases (ACs), 1:42, 1:44f, 1:45, 2:377, 2:378f, 4:225t
 background and characteristics, 2:377
 $\text{G}_{\beta\gamma}$ subunits, 1:480–482, 1:482f, 2:377, 4:265
 G_g protein (gustducin), 2:379t, 2:380

G proteins (*Continued*)

- G_i protein, 2:378–379, 2:378f, 2:379t, 4:225t
- G_o protein, 2:379, 2:379t, 4:225t
- G_t protein (transducin), 2:379, 2:379t
- G_z protein, 2:379, 2:379t
- phospholipase C-β (PLCβ), 3:469–470
- posttranslational modifications, 2:379t, 2:380
- sequence comparisons, 2:378f
- signaling pathways, 2:377, 2:378f
- signal transduction, 2:379
- glycoprotein hormone receptors, 4:388–390
- Golgi proteins, 2:479
- G protein-gated inward rectifying K⁺ (GIRK) channels, 2:494
- G-protein-signaling regulators (RGSs), 2:492–495
 - action mechanisms, 2:493
 - domain structures, 2:492, 2:494f
 - functional role, 2:492, 2:493–494
 - historical research, 2:492
 - mammalian RGS proteins, 2:493t
 - regulation mechanisms, 2:495
 - selectivity, 2:493
 - signal termination, 2:493–494
 - therapeutic uses, 2:495
- G protein transducin, 3:641–642
- G_q family, 2:496–500
 - activation mechanisms, 2:496–497, 2:497t
 - cardiovascular diseases, 2:499
 - cerebellum development, 2:499
 - characteristics and functional role, 2:496
 - complementary roles, 2:499–500
 - downstream effector mechanisms, 2:498
 - family members, 2:496, 2:497f
 - function modulation, 2:499
 - homology, 2:497f
 - phospholipase C-β (PLCβ), 2:496, 2:497–498, 2:498f
 - phylogenetic relationships, 2:497f
 - platelet aggregation, 2:499
 - signal transduction pathways, 2:497–498, 2:498f
- G_s family, 2:517–521
 - adenylyl cyclases (ACs), 1:42, 1:44f, 1:45, 2:517–518
 - characteristics and functional role, 2:517
 - disease connections
 - activating G_sα mutations, 2:519
 - GNAS/Gnas imprinting defects, 2:520
 - inactivating G_sα mutations, 2:519–520
 - posttranslational modifications, 2:519
 - GNAS/Gnas gene organization, 2:517, 2:518f
 - isoforms
 - G_sα, 2:517–518
 - XLαs, 2:518–519
 - knockout mice phenotypes, 2:521t
 - research summary, 2:521
- intracellular second messenger systems, 1:43f
- macroautophagy regulation, 1:149
- phototransduction mechanisms, 3:489f
- platelet-activating factor receptor (PAFR), 3:536, 3:536f
- serotonin receptors, 4:224, 4:225t
- voltage-gated calcium channels, 4:561
- GRAM-like ubiquitin binding in EAP45 (GLUE) domain, 3:449
- grancalcin, 1:319t
- granular component (GC), 3:324, 3:325f
- granulocyte colony stimulating factor (G-CSF)
 - expression regulation, 2:529
 - signaling pathways, 2:529
 - structural characteristics, 2:526–527, 2:527f
- granulocyte–macrophage stimulating factor (GM-CSF)
 - calcineurin-mediated signal transduction, 1:339–340, 1:340f
 - c-fes proto-oncogene (Fes), 1:452
 - hematopoietin receptors
 - expression regulation, 2:529
 - signal transduction, 2:529
 - structural characteristics, 2:526–527, 2:527f
 - type I cytokines, 1:616
- granulocyte–monocyte progenitor (GMP), 1:188–189, 1:188f
- granulopoiesis, 1:188–189, 1:188f
- grapefruit, 4:533t
- grayanotoxin, 4:568t
- green algae, 1:66–70
 - anoxygenic photosynthesis, 2:209
 - dyneins, 2:184
 - ferredoxins (Fd), 2:296
 - nonphotochemical quenching (NPQ), 3:486
 - plastocyanin (PC), 3:530
 - reaction centers, 2:335, 3:492
- green bacteria, 2:501–509
 - anoxygenic photosynthesis, 2:209, 2:335, 3:32
 - antenna complexes, 3:32, 3:484, 3:484f
 - bacteriochlorophyll (BChl) biosynthesis, 2:503–504, 2:506f
 - characteristics and classifications, 2:501–503
 - chlorosomes, 2:513–516
 - characteristics, 2:502f, 2:504–507
 - bacteriochlorophyll (BChl), 2:513, 2:515f
 - carotenoids, 2:514
 - proteins, 2:514
 - size and internal structure, 2:513, 2:515f
 - energy transfer mechanisms, 2:515f
 - evolutionary considerations, 2:516
 - homologs, 2:514t
 - molecular structure, 2:514f
 - organizational models, 2:514–515, 2:516f
 - phylogenetic relationships, 2:513
 - spectral properties, 2:514, 2:515f
- cytochromes, 2:510–512
 - cytochrome c
 - characteristics, 2:510–512
 - cytochrome bc₁ complex, 2:511–512, 2:511f
 - cytochrome c-554/555, 2:511, 2:511f
 - cytochrome c₂, 2:511, 2:511f
 - flavocytochrome c, 2:511
 - newly discovered cytochromes, 2:512
 - SoxAX, 2:511, 2:511f
 - filamentous anoxygenic phototrophic (FAP) bacteria
 - alternative complex III, 2:512
 - characteristics, 2:510
 - cytochrome c-554, 2:512
 - uncharacterized cytochromes, 2:512
 - general discussion, 2:510
 - green sulfur bacteria, 2:510, 2:511f
 - structural characteristics, 2:511f
 - electron transfer chains, 2:508–509, 2:508f
 - evolutionary development, 3:34
 - Fenna–Matthews–Olson protein, 2:504f, 2:505f, 2:507
 - filamentous anoxygenic phototrophic (FAP) bacteria
 - alternative complex III, 2:512
 - characteristics, 2:501–503, 2:510
 - cytochrome c-554, 2:512
 - cytochromes, 2:510
 - electron transfer chains, 2:508–509, 2:508f
 - reaction centers, 2:335, 2:507–508, 2:508f
 - uncharacterized cytochromes, 2:512
 - general discussion, 3:32
 - green filamentous bacteria, 2:513, 2:515f
 - green sulfur bacteria
 - characteristics and classifications, 2:501–503
 - chlorosomes, 2:513, 2:515f
 - cytochromes, 2:510, 2:511f
 - light harvesting complexes, 2:726
 - reaction centers
 - characteristics, 2:335
 - functional role, 2:337–338
 - F_x cluster, 2:338
 - PscB protein, 2:339–340, 2:339f
 - light harvesting complexes, 2:726
 - oxygenic photosynthesis, 3:32
 - reaction centers, 2:507–508, 2:508f
- green fluorescent protein (GFP)
 - bioluminescence, 1:465
 - calcium oscillation decoding, 1:325
 - cell migration, 1:432, 1:433f
 - intracellular calcium ion (Ca²⁺) indicators, 2:577–578
 - polymerase η (Pol η), 4:573–574, 4:573f
 - protein kinase C (PKC) translocation, 3:639, 3:639f
- green tea, 2:167–168, 3:305, 3:305t
- Gremlin expression, 1:236
- Griscelli syndrome, 3:223–224, 3:223t
- GroEL (growth E locus, large gene, bacterial group I chaperonin)
 - allosteric regulation, 1:74
 - Escherichia coli*, 1:457–458, 1:458f, 1:459
 - GroEL–GroES protein-folding cycle, 1:457–458, 1:458f
 - occurrences, 1:456

- structural characteristics, 1:457f
- GroES (growth E locus, small gene, bacterial group I chaperonin cofactor), 1:456, 1:457f, 1:458f
- groovin, 1:319t
- group I/group II introns, 1:397–400, 4:147–153
- alternative splicing, 1:75, 1:76f, 3:60–61, 4:187–188
- characteristics, 4:142, 4:143f, 4:147, 4:148f
- chemical splicing mechanisms, 3:59–60, 3:60f, 4:185, 4:186f
- distribution, 4:146f
- exon–intron architecture, 3:142–144, 3:143f, 4:186–187
- exon/intron boundaries, 1:75, 1:76f, 1:397, 1:398f, 4:149f
- folding processes
- ai5 γ ribozyme folding, 4:151, 4:152f
- autonomous *in vitro* folding, 4:151, 4:152f
- protein-facilitated folding and splicing, 4:151–152, 4:152f
- research background, 4:151–152
- historical research, 4:185
- intronic splicing enhancers (ISEs), 1:76, 1:78f, 3:62, 3:62f
- intronic splicing silencers (ISSs), 3:62f
- research summary, 4:153
- ribozymes, 4:188
- self-splicing reactions, 1:397, 1:398f, 4:188, 4:188f
- spliceosomes, 4:152–153, 4:153f
- splicing pathways, 4:147, 4:149f, 4:186f, 4:187f
- structural characteristics
- domain I, 4:147–148, 4:148f, 4:151f
- domain II, 4:148–149, 4:148f, 4:151f
- domain III, 4:148f, 4:149, 4:151f
- domain IV, 4:148f, 4:149, 4:151f
- domain V, 4:148f, 4:149–150, 4:151f
- domain VI, 4:148f, 4:149f, 4:150, 4:151f
- secondary structures, 1:397, 1:398f
- tertiary structures, 1:397–399, 1:398f
- substrate recognition
- G $\bullet$ U wobble pairs, 1:398f, 1:399–400, 1:399f, 1:400f
- guanosine-binding sites, 1:398f, 1:399, 1:399f, 1:400f
- helix P1/5' splice site recognition, 1:398f, 1:399–400, 1:399f, 1:400f
- P4–P6 domain, 1:398f, 1:399f, 1:400f
- three-dimensional structure, 4:150, 4:151f
- growth arrest and DNA damage inducible 45-beta (Gadd45b), 1:200, 1:200f
- growth factor receptor bound protein 2 (Grb-2)
- c-fes proto-oncogene (Fes), 1:453
- focal adhesion kinase (FAK) family, 2:265–266
- guanine-nucleotide-exchange factors (GEFs), 4:245
- hematopoietin receptors, 2:526, 2:528f
- insulin signal transduction, 2:603–604, 2:605–606
- mucins, 3:207–208
- platelet-derived growth factor (PDGF) receptors, 3:539–540, 3:540f, 3:541f
- steroid/thyroid hormone receptors, 4:310f
- growth hormones (GHs)
- amino acid metabolism, 1:95
- calcitonin receptors (CTRs), 1:272
- calcium-sensing receptor (CaSR), 1:329–330
- growth hormone secretagog receptor (GHS-R 1a), 1:118–119
- nuclear factor kappa B (NF- κ B), 3:303t, 3:304t
- starvation effects, 4:302–303
- Guamanian amyotrophic lateral sclerosis (ALS-G), 4:416
- guanidinium chloride (GuHCl), 3:625
- guanine
- DNA alkylation sites and repair mechanisms, 2:10t
- 8-oxo-guanine (8-oxoG)
- biological effects
- aging/age-related diseases, 2:79
- carcinogenesis, 2:79
- cytotoxicity, 2:79
- DNA glycosylases, 2:3t
- GO system (MutM–MutT–MutY), 2:80, 2:80f
- molecular structure, 2:2f, 2:78f
- MutY repair pathway, 2:57–58, 2:57f
- oxidative damage, 2:1, 2:54, 2:78, 4:405
- guanine-nucleotide exchange factors (GEFs), 2:344, 2:345f, 2:494, 3:439, 4:111–112
- guanine nucleotide-exchange proteins (GEPs), 1:135–136, 1:135f
- riboswitches, 4:137
- UvrABCD endonuclease, 3:330t
- See also G proteins
- guanine-nucleotide-binding proteins (G proteins)
- See G proteins
- guanosine
- group I/group II introns, 1:398f, 1:399, 1:399f, 1:400f
- phage libraries, 3:432–433, 3:433t, 3:434t
- guanosine diphosphate (GDP)
- adenosine diphosphate ribosylation factors (ARF), 1:134, 1:135f
- anaplerotic pathways, 1:106, 1:106f
- carboxyl methylation, 3:608
- citric acid cycle, 1:106, 1:106f
- cyclic guanosine monophosphate (cGMP), 3:536, 3:536f
- elongation factor Tu (EF-Tu), 4:428–430
- GDP-fucose (GDP-Fuc), 1:3
- GDP-mannose (GDP-Man), 1:3
- G $_i$ family, 2:377–380
- background and characteristics, 2:377
- G $_{\beta\gamma}$ subunits, 2:377
- G $_g$ protein (gustducin), 2:379t, 2:380
- G $_i$ protein, 2:378–379, 2:378f, 2:379t
- G $_o$ protein, 2:379, 2:379t
- G $_t$ protein (transducin), 2:379, 2:379t
- G $_z$ protein, 2:379, 2:379t
- posttranslational modifications, 2:379t, 2:380
- sequence comparisons, 2:378f
- signaling pathways, 2:377, 2:378f
- signal transduction, 2:379
- G $_q$ family proteins, 2:498f
- G $_s$ family proteins, 2:517–518
- guanine-nucleotide-exchange factors (GEFs), 4:12–13, 4:13f, 4:246f
- guanosine diphosphate/guanosine triphosphate (GDP/GTP) cycle, 4:12–13, 4:13f, 4:243f, 4:244f
- hematopoietin receptors, 2:528f
- initiation factors, 4:428–430
- mitochondrial ATP-dependent potassium ion (K $^+$) channel (KATP), 3:551
- multiple opioid receptors, 3:357
- muscarinic acetylcholine receptors, 3:215
- pheromone receptors, 3:441–442
- phosphoinositide-dependent protein kinases, 3:464f, 3:465
- platelet-activating factor receptor (PAFR), 3:536, 3:536f
- protein biosynthesis elongation, 2:249–251, 2:250f, 2:253f, 2:254f
- translation initiation complex formation, 4:426–427, 4:427f
- translation initiation factor 2 (eIF2 α) phosphorylation, 4:491
- tricarboxylic acid cycle, 4:439f, 4:439t
- tubulin, 4:450
- guanosine monophosphate (GMP)
- cyclic guanosine monophosphate (cGMP)
- characteristics and functional role, 1:567
- chemoattractant signaling, 2:237f, 2:238
- class II PDEs
- catalytic activity, 1:572, 1:572f
- characteristics, 1:572
- regulatory mechanisms, 1:572, 1:572f
- class I PDEs
- catalytic domains, 1:568–569, 1:568f
- characteristics, 1:567–568, 1:580
- cyclic guanosine monophosphate (cGMP)-hydrolyzing PDEs, 1:569–570
- cyclic nucleotide specificity, 1:569
- domain structures, 1:568f
- dual-specificity PDEs, 1:570–571
- inhibitors, 1:569
- PDE1, 1:570–571, 1:581–582
- PDE2, 1:571, 1:581–582, 1:583–584
- PDE3, 1:571, 1:581–582, 1:583
- PDE5, 1:570, 1:581–582, 3:27
- PDE6, 1:570, 1:581–582
- PDE9, 1:570
- PDE10, 1:571, 1:581–582
- PDE11, 1:571–572, 1:581–582
- cyclic guanosine monophosphate (cGMP)-hydrolyzing PDEs, 1:569–570
- cyclic nucleotide-dependent protein kinases, 1:574, 1:575f
- distribution, 1:567
- inositol trisphosphate (IP $_3$), 2:586–587
- ion channel proteins, 2:650
- molecular structure, 1:568f
- muscarinic acetylcholine receptors, 3:215–216

- guanosine monophosphate (GMP)
(*Continued*)
natriuretic peptides, 3:226f
nitric oxide synthase (NOS), 3:254f
phosphodiesterases (PDEs), 1:580–585
 characteristics, 1:580
 clinical applications, 1:584
 family-specific inhibitors, 1:584
 G_i family functions, 2:379
 homeostatic regulation, 1:582–583
 inherited disorders, 1:584
 intracellular effectors, 1:583
 molecular diversity, 1:580–581
 phototransduction mechanisms, 3:489f
 research summary, 1:584
 second messenger pathways, 1:43f
 signaling pathways, 1:568f, 1:574, 1:575f
 signal integrators, 1:583–584
 spatially organized signalling networks and microdomains, 1:582
 structure and functions, 1:581–582, 1:581f
 phototransduction mechanisms, 3:489–490, 3:489f
 research summary, 1:572
 second messenger pathways, 1:43f
 signaling pathways, 1:568f, 1:575f
 structural domains, 1:574–575, 1:576f
- guanosine triphosphate (GTP)
adenosine diphosphate ribosylation factors (ARF), 1:134, 1:135f, 4:242, 4:243f
anaplerotic pathways, 1:106, 1:106f
cadherin signaling, 1:262–263, 1:263f
carboxyl methylation, 3:608
citric acid cycle, 1:106, 1:106f
cytidine nucleotide formation, 3:716
elongation factor Tu (EF-Tu), 4:123–125, 4:124f, 4:420–421, 4:421f, 4:428–430
G₁₂/G₁₃ family functions, 2:345f
G_i family, 2:377–380
 background and characteristics, 2:377
 G_{βγ} subunits, 2:377
 G₈ protein (gustducin), 2:379t, 2:380
 G_i protein, 2:378–379, 2:378f, 2:379t
 G_o protein, 2:379, 2:379t
 G_t protein (transducin), 2:379, 2:379t
 G_z protein, 2:379, 2:379t
 posttranslational modifications, 2:379t, 2:380
 sequence comparisons, 2:378f
 signaling pathways, 2:377, 2:378f
 signal transduction, 2:379
G_q family proteins, 2:498f
G_s family proteins, 2:517–518
guanine-nucleotide-exchange factors (GEFs), 1:453, 2:345f, 4:12–13, 4:13f, 4:243f, 4:245, 4:246f
guanosine diphosphate/guanosine triphosphate (GDP/GTP) cycle, 4:12–13, 4:13f, 4:243f, 4:244f
guanosine triphosphatases (GTPases)
 cadherin signaling, 1:262–263, 1:263f
 calcium ion (Ca²⁺) signaling, 2:719–720, 2:720f
 filamentous septins, 1:263f, 4:214–215
 GTPase-activating proteins (GAPs), 1:453, 4:12–13, 4:13f, 4:243f, 4:245
 GTPase factor binding sites, 4:134
 integrin signaling, 2:608
 mitochondrial dynamics, 3:136–137
 mTOR complexes, 3:195–196
 muscarinic acetylcholine receptors, 3:215
 neurotrophins, 3:241–243, 3:242f
 phosphatidic acid (PA), 3:475, 3:475f
 phospholipase D (PLD), 3:472–473, 3:474–475, 3:474f
 Rab guanosine triphosphatase (GTPase) activating protein (GAP)
 Golgi proteins, 2:479
 insulin signal transduction, 2:604–605
 small guanosine phosphatases (GTPases), 4:242, 4:243f, 4:244f
 transport pathways, 4:206, 4:504–506, 4:505f
Ras proteins, 2:528f, 4:12
Rho guanosine triphosphatases (Rho GTPases), 4:111–115
 actin, 1:20
 actin cytoskeleton dynamics, 1:430–431, 1:431f, 4:111, 4:248
 Cdc42 protein, 1:440–441, 1:441f, 4:112f, 4:113, 4:113t, 4:114f
 cell–matrix adhesion diversity, 2:321–322, 2:321f
 cell migration, 1:439f, 4:114
 cell proliferation, 4:115
 c-fes proto-oncogene (Fes), 1:451
 chemokine receptors, 1:480–482
 Dbl-related GEFs, 4:111–112, 4:112f
 DOCK family proteins, 1:441f, 1:480–482, 4:111–112, 4:112f
 downstream targets, 4:113–114
 endocytosis, 4:114
 formins, 1:24
 G₁₂/G₁₃ family, 2:343–344
 G_q family proteins, 2:498
 GTPase-activating proteins (GAPs), 2:344, 2:345f, 3:489f, 4:112, 4:112f
 guanine-nucleotide exchange factors (GEFs), 2:344, 2:345f, 2:494, 3:439, 4:111–112
 myosin, 3:222–223
 occurrences, 4:111
 phospholipase C-β (PLCβ), 3:469–470
 phospholipase D (PLD), 3:474f, 3:475–476
 Rac effectors, 1:440–441, 1:441f, 4:112f, 4:113, 4:113t
 Rho effectors, 4:113, 4:113t
 Rho-GDI (guanine-nucleotide-dissociation inhibitors), 3:243, 3:244f, 4:112, 4:112f, 4:243f, 4:244f, 4:245, 4:246
 RhoG protein, 4:112f, 4:113t
 Rho GTP-binding proteins, 1:440–441, 1:441f
 RhoH protein, 4:112f, 4:113t
 Rho-kinase (ROCK), 1:439f, 1:440–441, 1:441f, 1:623–624, 2:237f, 2:238, 4:113, 4:113t
 Rnd2/Rnd3 proteins, 4:112f, 4:113t, 4:114
 secretion, 4:114
 small guanosine phosphatases (GTPases), 4:242, 4:243f, 4:244f
 subfamilies, 4:111, 4:112f
 tumor necrosis factor receptors (TNFRs), 4:454
 vesicle transport, 4:114
small guanosine phosphatases (GTPases), 4:242–248
 actin cytoskeleton, 4:248
 activating proteins, 4:246
 calcium ion (Ca²⁺) signaling, 2:719–720, 2:720f
 cell migration, 1:439–440
 cell proliferation regulation, 4:247–248, 4:247f
 c-fes proto-oncogene (Fes), 1:451, 1:453
 chemokine receptors, 1:483f
 dissociation inhibitors, 4:243f, 4:244f, 4:245, 4:246
 domain structures, 4:2f, 4:242–245, 4:244f
 functional role, 4:243f, 4:244f, 4:247
 genetic deregulation mechanisms, 4:246–247
 guanine-nucleotide-exchange factors (GEFs), 1:453, 4:245
 hemidesmosome function, 1:642–643
 membrane transport, 4:248
 molecular switches, 4:7
 nucleocytoplasmic transport, 4:7, 4:8f
 phosphatidylinositol-3-phosphate (PtdIns(3)P), 3:449
 phospholipase C-ε (PLCε), 3:470
 Rab protein family, 4:1–6
 Ran protein family, 4:7–11
 regulation mechanisms, 4:245
 research background, 4:242
 signaling cascades, 4:245–246, 4:246f
 small GTPase Ran, 3:320–321, 3:321f, 4:7, 4:8f
 transfer RNA (tRNA), 4:123–125
 vacuole transport pathways, 4:504–506
- guanosine triphosphate (GTP)-binding proteins (G proteins)
 chemoattractant signaling, 2:236–238, 2:237f
 dopamine receptors, 2:171
 Golgi proteins, 2:479
 metabotropic glutamate (mGlu) receptors, 2:395
 Ras proteins, 4:12
 hematopoietin receptors, 2:528f
 hydrolysis mechanisms, 4:123–125, 4:124f, 4:420–421, 4:421f
 initiation factors, 4:428–430, 4:428f
 mitochondrial ATP-dependent potassium ion (K⁺) channel (KATP), 3:551
 muscarinic acetylcholine receptors, 3:215

- natriuretic peptides, 3:226f
 nucleotide excision repair (NER) pathways, 3:329f
 pheromone receptors, 3:441–442, 3:443f
 phosphoinositide-dependent protein kinases, 3:464f, 3:465
 platelet-activating factor receptor (PAFR), 3:536, 3:536f
 protein biosynthesis elongation, 2:249–251, 2:250f, 2:251f, 2:253f, 2:254f, 4:122–123
 pterin biosynthesis, 3:666–667, 3:667f
 secretory pathway, 4:206
 selenoprotein synthesis, 4:212
 septins, 4:214–218
 behavior and function
 Ashbya gossypii, 4:216f, 4:217
 budding yeasts, 4:216–217, 4:216f
 fission yeasts, 4:217–218
 metazoans, 4:217f, 4:218
 domain structures
 filamentous complexes, 4:214–215, 4:215f
 higher-order assemblies, 4:215–216, 4:216f, 4:217f
 inositol phospholipids, 4:216
 functional role, 4:214
 serotonin receptors, 4:224
 translation initiation complex formation, 4:426–427, 4:427f
 translation initiation factor 2 (eIF2 α)
 phosphorylation, 4:491
 tricarboxylic acid cycle, 4:439f, 4:439t
 tubulin, 4:450
 guanylate cyclase (sGC), 3:255
 guanylyl cyclase (GC)
 characteristics and functional role, 1:567
 chemotaxis, 2:236–237
 guanylyl-cyclase-linked natriuretic peptide receptors, 3:226f, 3:227
 signaling pathways, 1:568f
 guard cell outward rectifier K channel (GORK), 4:557–559, 4:558f
 L-gulonolactone oxidase (GLO), 4:530–531
 Gun mutants, 1:515
 gustducin, 2:379t, 2:380
 gut hormones, 1:115–119
 anorexigenic hormones
 apolipoprotein A-IV (APO-IV), 1:116
 cholecystokinin (CCK), 1:115
 peptide YY (PYY), 1:115–116
 background information, 1:115
 food intake and obesity, 1:119
 incretin hormones
 characteristics, 1:116–117
 gastric inhibitory peptide (GIP), 1:117
 glucagon-like peptide-1 (GLP-1), 1:116–117
 oxyntomodulin (OXM), 1:117
 orexigenic hormones, 1:118–119
 pancreatic peptides
 amylin, 1:118
 enterostatin, 1:118
 glucagon, 1:117–118
 insulin, 1:118
 pancreatic polypeptide (PP), 1:115–116, 1:117
 GyrA/GyrB, 2:164–165, 2:165f, 3:369–370
- ## H
- H2AX protein, 3:273t
 HAC1 gene expression, 2:524–525, 2:524f, 4:489, 4:490f
Haemophilus, 2:43
Haemophilus haemolyticus, 2:132t
Haemophilus influenzae
 Cho phylogenetic tree, 3:331f
 type III DNA restriction and modification enzymes, 2:137t, 2:139–140
 uvrC phylogenetic tree, 3:331f
 Hailey–Hailey disease, 2:482
 hairpin ribozymes, 4:144, 4:146f
 hairy-related transcription factor (Hrt)
 family, 1:185
 hakai, 1:255, 1:258f
 Haldane relationship, 4:347
Haloarcula, 3:32–34
Halobacterium
 archaeal energy transduction
 methanogens, 3:32–34
 respiratory complexes, 3:34
 secondary energy transducers, 3:34
 Cho phylogenetic tree, 3:331f
 uvrC phylogenetic tree, 3:331f
Halobacterium salinarum
 anoxygenic photosynthesis, 2:209
 light utilization, 2:209, 3:33–34, 3:34f
 respiratory complexes, 3:34
Halococcus, 3:32–34
Haloferrax, 3:32–34
Haloferrax volcanii, 4:239, 4:240f
 haloperoxidases, 2:535–536
 halorhodopsin (HR), 3:33–34, 3:34f
 hammerhead ribozymes, 4:143f, 4:144, 4:146f
 Hand1/2 transcription factor, 1:184, 1:184f
Hansenula anomala, 1:146–147
Hansenula polymorpha, 1:89
 Hap1 transcriptional activator, 3:309
 HASTY protein, 3:99–100
 hck genes/Hck proteins, 4:294
 H-DNA
 background information, 2:143–144
 biological role, 2:144
 bonding schemes, 2:144f
 chemical properties, 2:143–144
 structural characteristics, 2:143f
 heart
 acyl-coenzyme A synthetase
 long-chain acyl-CoA synthetase (LCAS), 2:292–293
 medium-chain acyl-CoA synthetase (MCAS), 2:292
 short-chain acyl-CoA synthetase (SCAS), 2:292
 very long-chain acyl-CoA synthetase (VLCAS), 2:293–294
 aquaporin water channels, 1:124t
 arrhythmia incidence, 1:382–383
 cadherins, 1:259–260
 carbohydrate-responsive element binding protein (ChREBP) distribution, 1:374
 cardiac muscles, 1:179–186
 development process
 cardiogenesis, 1:183, 1:183f, 1:184f
 microRNA (miRNA), 1:185
 transcription factors, 1:183–185, 1:184f
 research summary, 1:186
 stem cells, 1:186
 structural characteristics, 1:179, 1:180f
 delayed after-depolarization (DAD), 1:382–383, 1:382f
 early after-depolarization (EAD), 1:382–383
 energy metabolism
 aerobic heart, 3:65–66, 3:66f
 carbohydrate metabolism, 3:65–66, 3:66f, 3:68f
 fatty acid metabolism, 3:66–67, 3:66f
 ischemic heart, 3:67, 3:68f
 ischemic-induced metabolic signaling alterations, 3:67–69, 3:69f
 optimization approaches, 3:71, 3:71t
 reperfused ischemic heart
 basic principles, 3:69–70, 3:70f
 intracellular acidosis/cardiac efficiency, 3:70–71
 research summary, 3:71
 excitation–contraction (E–C) coupling, 1:379–383
 arrhythmia incidence, 1:382–383
 basic concepts
 action mechanisms, 1:379–381
 action potential (AP), 1:379, 1:380f, 4:566
 cytosolic calcium, 1:379–381, 1:380f
 L-type calcium channels, 1:379–381, 1:380f
 myocytes, 1:379, 1:380f
 myofilaments, 1:380–381, 1:380f
 relaxation mechanisms, 1:381
 delayed after-depolarization (DAD), 1:382–383, 1:382f
 early after-depolarization (EAD), 1:382–383
 heart failure, 1:381–382, 1:382f
 transverse tubule (T-tubule), 1:379–381, 1:380f, 1:382f
 heart disease
 amyloid deposits, 1:101t
 vitamin E, 4:547
 heart failure, 1:381–382, 1:382f
 ischemic heart diseases, 3:65–71
 leukotriene receptors, 3:584
 mammalian glucose transport, 2:392t
 peroxisome proliferator-activated receptor alpha (PPAR α), 3:421
 phosphofructokinase-2/fructose bisphosphatase-2 (PFK-2/FBPase-2)-activated phosphorylation, 3:457–458
 proteinase-activated receptors (PARs), 3:605t

- heart (*Continued*)
 sodium ion/calcium ion ($\text{Na}^+/\text{Ca}^{2+}$)
 exchanger, 3:48
 voltage-gated calcium channels, 4:561*t*
 heat shock protein 40 (Hsp40), 2:522–523,
 2:523*f*, 3:569
 heat shock protein 60 (Hsp60), 1:456,
 4:269*f*, 4:271–272
 heat shock protein 70 (Hsp70), 2:522–523,
 2:523*f*, 3:569
 heat shock protein 90 (Hsp90)
 hypoxia inducible factor (HIF), 4:69
 nitric oxide synthase (NOS), 1:343–346
 proteasome inhibition, 2:345*f*, 3:593,
 3:593*f*
 protein folding processes, 1:456
 steroid/thyroid hormone receptors,
 4:308–309, 4:310*f*
 stress responses, 2:522–523, 2:523*f*
 heat shock protein 104, 3:569
 heat/stress responses, 2:522–525
 heat shock response
 activation mechanisms, 2:522–523,
 2:523*f*
 chaperone proteins, 2:522–523, 2:523*f*
 input conditions, 2:522
 output conditions
 aberrant protein degradation, 2:523
 chaperone genes, 2:523
 environmental stress response,
 2:523–524
 protein repair and preservation, 2:523
 research background, 2:522–524
 stress detection, 2:522–523, 2:523*f*
 unfolded protein response (UPR)
 activation mechanisms, 2:524*f*
 endoplasmic reticulum (ER),
 2:524–525, 2:524*f*
 ER-associated degradation (ERAD),
 2:524*f*, 2:525
 output conditions, 2:525
 stress sensors, 2:524–525, 2:524*f*
 transmembrane signaling, 2:524–525,
 2:524*f*
 unfolded proteins, 2:522, 2:523*f*
 Hedgehog (Hh) pathway
 glycogen synthase kinase-3 (GSK3), 2:436*f*,
 2:437
 liver regeneration
 chronic liver injury repairs, 1:193–195,
 1:194*f*
 fetal development, 1:192, 1:192*f*
 post-partial hepatectomy (PH)
 reactivation, 1:193
 research summary, 1:195
 protein palmitoylation, 2:736–737,
 3:645–646, 3:646*f*
Sex-lethal (Sxl) master-switch gene, 1:82
 HEK293 cells, 1:129, 1:130–131
 helicases
 bacterial DNA replication, 2:92, 2:93*f*,
 2:94*f*, 2:115–116, 2:116*t*, 2:121–122,
 2:122*f*
 characteristics and functional role, 2:21,
 3:261, 4:43
 DEAD-box helicase, 4:151–152, 4:152*f*
 D-loop dissociation
 enzyme disruption, 4:32–33
 Holliday junction (HJ) dissolution,
 4:30*f*, 4:33
 Holliday junction (HJ) resolution, 4:30*f*,
 4:33
 resolvases, 4:30*f*, 4:32–33
 DnaA helicase
 DNA unwinding, 2:121–122, 2:122*f*
 functional domains, 2:121–122, 2:122*t*
 replisome assembly process
 characteristics, 2:122–123
 DiaA protein stabilization, 2:123
 DnaA-dependent unwinding, 2:123
 DnaA-*oriC* complex formation,
 2:122–123
 HU protein stabilization, 2:123
 DnaB–DnaC complex, 2:124
 DnaB helicase, 2:21, 2:22, 2:22*t*, 2:25–26,
 2:92, 2:93*f*, 2:94*f*, 2:116*t*, 2:124
 DnaC helicase, 2:115–116, 2:116*t*,
 2:123–124
 DnaG helicase, 2:92, 2:93*f*, 2:116*t*
 DNA recombination
 D-loop disruption, 4:32–33
 enzyme translocation, 4:30–31, 4:30*f*,
 4:31*t*
 Holliday junction (HJ) dissolution, 4:33
 Holliday junction (HJ) resolution, 4:33
 recombination inhibitors, 4:33–34
 recombination models, 4:30*f*
 research summary, 4:34
 single-strand DNA (ssDNA), 4:31–32
 eukaryotic DNA replication, 2:118
 hexameric enzyme action, 2:21–27
 classifications, 2:21, 2:22*t*
 double-strand DNA (dsDNA)
 translocation, 2:24–25
 loaders/loading process, 2:21–22, 2:22*t*
 nucleoside triphosphatase (NTPase)-
 powered translocation, 2:22–24,
 2:24*f*
 research summary, 2:26
 single-strand DNA (ssDNA) interactions,
 2:21
 structural characteristics, 2:21, 2:23*f*
 unwinding mechanisms, 2:25–26, 2:25*f*
 intron splicing, 4:151–152, 4:152*f*
 mismatch repair (MMR), 2:67, 2:67*f*
 PcrA helicase
 characteristics and functional role, 3:261
 DNA unwinding
 active, rolling or hand-over-hand
 model, 3:264–265, 3:265*f*
 active versus passive mechanisms,
 3:264
 dimeric inch-worm models, 3:265,
 3:265*f*
 inch-worm mechanisms, 3:264,
 3:264*f*, 3:265*f*
 protein oligomerization, 3:263
 unwinding mechanisms, 3:263–264
 research summary, 3:265
 single-strand DNA (ssDNA)
 translocation, 3:263, 3:264, 3:264*f*
 structural characteristics
 crystal structures, 3:262, 3:262*f*
 domain structures, 3:263*f*
 helicase motifs, 3:263*f*
 open and closed confirmations,
 3:262–263, 3:263*f*
 primary structures, 3:261–262
 crystal structures, 3:262, 3:262*f*
 open and closed confirmations,
 3:262–263
 primary structures, 3:261–262
 primary structures, 3:261–262
 translocase-based protein displacement,
 3:265
 polymerase α -primase (Pol α)–helicase
 relationship, 2:241–242, 2:242–243,
 2:242*f*
 RecA helicases, 2:21, 2:22*t*, 4:32
 RecBCD helicases, 4:32
 RecB helicases, 4:32
 RecD helicases, 4:32
 RecJ helicases, 2:41*f*, 2:45–46, 2:54–56,
 2:55*t*, 2:56*f*
 recombination inhibitors, 4:33–34
 Rep helicases
 characteristics and functional role, 2:21,
 3:261
 DNA unwinding
 active, rolling or hand-over-hand
 model, 3:264–265, 3:265*f*
 active versus passive mechanisms,
 3:264
 dimeric inch-worm models, 3:265,
 3:265*f*
 inch-worm mechanisms, 3:264,
 3:264*f*, 3:265*f*
 protein oligomerization, 3:263
 unwinding mechanisms, 3:263–264
 research summary, 3:265
 single-strand DNA (ssDNA)
 translocation, 3:263, 3:264, 3:264*f*
 structural characteristics
 crystal structures, 3:262, 3:262*f*
 domain structures, 3:263*f*
 open and closed confirmations,
 3:262–263
 primary structures, 3:261–262
 translocase-based protein displacement,
 3:265
 ring-shaped helicases, 2:21
 SF1 (superfamily 1) helicases, 3:261–265
 characteristics and functional role, 3:261
 DNA unwinding
 active, rolling or hand-over-hand
 model, 3:264–265, 3:265*f*
 active versus passive mechanisms,
 3:264
 dimeric inch-worm models, 3:265,
 3:265*f*
 inch-worm mechanisms, 3:264,
 3:264*f*, 3:265*f*
 protein oligomerization, 3:263
 unwinding mechanisms, 3:263–264
 research summary, 3:265
 single-strand DNA (ssDNA)
 translocation, 3:263, 3:264,
 3:264*f*
 structural characteristics
 crystal structures, 3:262, 3:262*f*
 domain structures, 3:263*f*
 helicase motifs, 3:263*f*
 open and closed confirmations,
 3:262–263, 3:263*f*
 primary structures, 3:261–262

- translocase-based protein displacement, 3:265
- single-strand DNA (ssDNA), 4:31–32
- UvrD helicases
- characteristics and functional role, 3:261
 - damage sensing mechanisms, 3:334
 - DNA unwinding
 - active, rolling or hand-over-hand model, 3:264–265, 3:265f
 - active versus passive mechanisms, 3:264
 - dimeric inch-worm models, 3:265, 3:265f
 - inch-worm mechanisms, 3:264, 3:264f, 3:265f
 - protein oligomerization, 3:263
 - unwinding mechanisms, 3:263–264
 - research summary, 3:265
 - single-strand DNA (ssDNA)
 - translocation, 3:263, 3:264, 3:264f
 - structural characteristics
 - crystal structures, 3:262, 3:262f
 - open and closed confirmations, 3:262–263
 - primary structures, 3:261–262
 - translocase-based protein displacement, 3:265
- See also* RecQ helicases
- Helicase-SANT-Associated (HSA) domain, 4:63
- Helicobacter pylori*
- Cho phylogenetic tree, 3:331f
 - flavins, 2:313
 - respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:89–90
 - succinate:quinone oxidoreductases (SQORs), 4:92–93
 - 3-methyladenine (3mA) DNA glycolase III, 2:111, 2:111f
 - type III DNA restriction and modification enzymes, 2:139–140
 - uvrC* phylogenetic tree, 3:331f
 - vitamin B₁₂ deficiencies, 2:328
- heliobacteria
- anoxygenic photosynthesis, 2:209, 2:335, 3:32
 - antenna complexes, 3:32
 - evolutionary development, 3:34
 - general discussion, 3:32
 - oxygenic photosynthesis, 3:32
 - reaction centers
 - characteristics, 2:335
 - F_A- and F_B-containing proteins, 2:339, 2:339f
 - functional role, 2:337–338
 - F_x cluster, 2:338
 - PshB protein, 2:339, 2:339f
- Heliothis virescens*, 2:593
- helitrons, 3:282
- helix-hairpin-helix (HhH) domain, 2:16–17, 2:29f, 2:34f, 2:37f
- helix-loop-helix (HLH) domain
- DNA-binding motifs, 2:151f, 2:152
 - hypoxia inducible factor (HIF), 4:67, 4:68f
 - polymerase β (Pol β)
 - C-subdomain, 2:84, 2:85f
 - D-subdomain, 2:84
 - lyase domain, 2:83–84, 2:83f, 2:85f
 - N-subdomain, 2:84–85, 2:85f
 - RNA polymerases, 4:175, 4:175t
- Helix pomatia*, 2:702t
- helix-two-turn-helix (H2TH) superfamily, 2:16–17
- Hematococcus*, 1:504
- hematopoiesis, 1:187–189
- biochemical development
 - basophils, 1:188f, 1:189
 - cellular pathways, 1:188f
 - eosinophils, 1:188f, 1:189
 - erythropoiesis, 1:187, 1:188f
 - granulopoiesis, 1:188–189, 1:188f
 - hematopoietic stem cells (HSCs), 1:187, 1:188f
 - mast cells, 1:188f, 1:189
 - megakaryopoiesis, 1:187, 1:188f
 - monopoiesis, 1:188–189, 1:188f
 - transcription factors, 1:187, 1:188f
- hematopoietic cells
- G_s family phenotypes, 2:521t
 - signaling receptors, 2:743
- hematopoietin receptors, 2:526–531
- characteristics and functional role, 2:526
 - classifications, 2:526–527, 2:527f
 - class I hematopoietin receptors, 2:526–527, 2:527f
 - class II hematopoietin receptors, 2:526–527, 2:530
 - disease connections, 2:530
 - interferons, 2:526–527, 2:530
 - single subunit receptors
 - erythropoietin (Epo)
 - expression regulation, 2:527
 - signaling pathways, 2:527–528, 2:528f
 - structural characteristics, 2:526–527, 2:527f
 - granulocyte colony stimulating factor (G-CSF)
 - expression regulation, 2:529
 - signaling pathways, 2:529
 - structural characteristics, 2:526–527, 2:527f
 - thrombopoietin (Tpo)
 - expression regulation, 2:528
 - signaling pathways, 2:528–529
 - structural characteristics, 2:526–527, 2:527f
- heme peroxidases, 3:407–409, 3:408f
- heme proteins, 2:532–538
- bacterial heme-copper oxidases, 1:602–604, 1:603f, 1:604t, 1:606, 4:98, 4:99f
 - biosynthetic pathway, 2:539–542
 - 5-aminolevulinic acid (ALA) formation, 2:539–540, 2:540f
 - basic principles, 3:546
 - catabolic reactions, 3:547–549, 3:548f
 - porphyric disorders, 2:542, 2:542t, 3:547, 3:548f
 - protoporphyrinogen IX, 2:540f, 2:541
 - protoporphyrinogen oxidase (PPO), 2:540f, 2:541, 2:542t
 - regulation mechanisms, 2:541–542
 - schematic diagram, 2:540f, 3:545f, 3:546f
 - uroporphyrinogen III formation, 2:540–541, 2:540f
- characteristics, 2:532, 3:544
- cytochrome *b₆f* complex, 4:337, 4:337f, 4:338f, 4:339t
- cytochrome *bc₁* complex, 1:596, 1:596f, 1:597f, 3:663–664, 3:664f
- cytochrome *cbh3* oxidase, 3:694f, 3:697, 3:697t, 4:79, 4:80f
- electron carriers
- cytochrome a, 2:537
 - cytochrome b, 1:142f, 2:537
 - cytochrome *c*, 1:599–601, 2:537
 - functional role, 2:537
- enzymes
- catalases
 - action mechanisms, 2:536, 2:536f
 - functional role, 2:536
 - oxygenases, 2:536
 - structural characteristics, 2:536
 - characteristics, 2:535–537
 - cytochrome P-450 (CYP)
 - action mechanisms, 2:537, 2:537f
 - functional role, 2:536–537
 - general discussion, 2:536–537
 - occurrences, 2:536
 - peroxidases
 - action mechanisms, 2:535, 2:535f
 - functional role, 2:535
 - haloperoxidases, 2:535–536
 - reactive oxygen species (ROS), 2:536
 - structural characteristics, 2:535
 - functional role, 2:539, 2:542
- heme a
- cholesterol biosynthetic pathway, 1:516, 1:517f
 - cytochrome *c* oxidase, 4:101f
- heme b
- cytochrome *b₆f* complex, 4:337–338, 4:338f, 4:339f, 4:339t
 - PET* genes, 3:309
- heme *c_m*, 4:337, 4:337f, 4:338f, 4:339t
- heme-copper centers, 3:664–665
- heme/copper-containing terminal oxidases, 4:98, 4:99f
- heme d₁
- biosynthetic pathway, 3:545f, 3:546–547, 3:546f
 - characteristics, 3:544
- nitric oxide synthase (NOS), 1:342, 1:343f, 1:344–346, 1:344t, 1:345f, 3:253–254, 3:668
- oxygen carriers
- basic concepts, 2:532–535
 - hemoglobin (Hb)
 - blood substitutes, 2:535
 - carbon monoxide (CO), 2:534
 - cyanide, 2:534
 - fetal hemoglobin (HbF), 2:534
 - methemoglobin formation, 2:534
 - mutant hemoglobins, 2:534–535
 - oxygen-binding characteristics, 2:534, 2:534f

- heme proteins (*Continued*)
 sickle cell hemoglobin (HbS),
 2:534–535
 structural characteristics, 2:533
 myoglobin (Mb)
 allosteric regulation, 1:71
 carbon monoxide (CO), 2:534
 cyanide, 2:534
 functional role, 2:532
 oxygen-binding characteristics, 2:533,
 2:534f
 structural characteristics, 2:532, 2:533f
PET genes, 3:309
 photosynthetic reaction centers, 3:701f
 photosystem II (PSII) complex, 3:513f,
 3:514
 porphyric disorders, 2:542, 2:542t
 proton pumping, 3:664–665
 regulation mechanisms
 δ -aminolevulinic acid (ALA) function,
 3:547
 non-ALAS enzymatic regulation, 3:547
 structural characteristics, 2:532, 2:533f,
 2:539, 2:540f
 hemicellulose, 4:75f
 hemidesmosomes, 1:636–643
 assembly and maintenance, 1:642
 disease connections, 1:641t
 functional role
 cell–substrate adhesive function, 1:642
 general discussion, 1:643
 morphogenetic functions, 1:642
 transmembrane signaling function,
 1:642–643
 wound healing, 1:642–643
 integrin–extracellular matrix (ECM)
 contacts, 1:439f, 2:611f
 molecular components
 $\alpha\beta 4$ integrin heterodimer, 1:640f,
 1:641, 1:641t, 1:642–643
 BPAG1 protein, 1:641, 1:641t
 CD151, 1:642
 laminin-332, 1:641t, 1:642
 plectin, 1:641–642, 1:641t
 structural characteristics, 1:640–642,
 1:640f
 type XVII collagen, 1:640f, 1:641,
 1:641t
 ultrastructure, 1:640, 1:640f
 hemoglobin (Hb)
 carbon monoxide (CO), 2:534
 cyanide, 2:534
 glycated hemoglobin
 characteristics, 2:405
 functional role, 2:405–406
 hemoglobin mutants
 blood substitutes, 2:535
 fetal hemoglobin (HbF), 2:534
 sickle cell hemoglobin (HbS),
 2:534–535
 methemoglobin formation, 2:534
 nicotinamide adenine dinucleotide
 (NAD⁺/NADH) fluorescence, 4:277
 oxygen-binding characteristics
 allosteric regulation, 2:534, 2:534f
 Bohr effect, 2:534
 2,3-diphosphoglycerate (2,3-DPG),
 2:534
 oxygen–hemoglobin (O₂–Hb) binding,
 1:71, 1:72f
 spectrophotometric assays, 4:277
 structural characteristics, 2:533
 hemokinin 1
 functional role, 4:361
 synthesis, 4:361f
 Henri's equation, 2:216–218
 heparan sulfate (HS)/heparin, 3:655–656,
 3:655f, 4:511
 heparan sulfate proteoglycans (HSPGs),
 1:439–440, 2:303–304, 4:511
 heparin pentasaccharide, 1:2f
 hepatic carcinoma, 1:192f, 1:193–195
 hepatic gluconeogenesis, 2:384, 4:305–306,
 4:306f
 hepatic insulin resistance, 1:650, 1:650f
 hepatic mitochondria, 2:375–376, 2:375f
 hepatitis
 anti-mitochondrial antibodies (AMAs),
 3:116
 hepatitis B virus (HBV), 3:12–13, 3:13f,
 3:303t
 hepatitis C virus (HCV), 4:156–157
 hepatitis delta virus (HDV), 4:156–157
 hepatitis delta virus (HDV) ribozyme,
 4:144, 4:146f
 nuclear factor kappa B (NF- κ B), 3:303t,
 3:304–305
 primary biliary cirrhosis (PBC), 3:562
 hepatocytes
 healthy adult liver, 1:190–191, 1:191f
 intracellular calcium ion (Ca²⁺) waves,
 2:644
 liver regeneration, 1:192f, 1:193–195,
 1:194f
 peroxisome proliferator-activated receptor
 alpha (PPAR α), 3:420
 heptasaccharide intermediate
 (Man₅GlcNAc₂-P-P-Dol), 2:315–316,
 2:315f, 2:316t
 HER2/*neu* gene, 2:278
 herbicides, 3:547
 Herc5, 4:468t, 4:471
 hereditary inclusion body myopathy type II,
 2:471t, 2:475
 hereditary nonpolyposis colorectal cancer
 (HNPCC), 2:61–64, 2:62t, 2:63f
 hereditary spastic paraplegias (HSPs),
 4:267–272
 axonal degeneration, 4:267
 definition, 4:267
 endosomal localization
 cytochrome P-450 (CYP), 4:271
 developmental genes, 4:272
 lipid metabolism, 4:270–271
 maspardin/SPG21, 4:269f, 4:270–272
 mitochondrial proteins
 Hsp60 protein/SPG13, 4:271–272
 paraplegin/SPG7, 4:271
 myelin–axonal interactions, 4:272
 neuropathy target esterase (NTE), 4:271
 seipin, 4:269f, 4:270–271
 genetic bases, 4:267, 4:268t
 neuropathology, 4:267
 pathogenesis
 cellular trafficking genes
 atlastin-1/SPG3A, 4:269, 4:269f
 KIF5A/SPG10, 4:269f, 4:270
 NIPA1/SPG6, 4:269f, 4:270
 receptor-enhancing protein 1
 (REEP1)/SPG31, 4:269f, 4:270
 spartin/SPG20, 4:269f, 4:270
 spastin/SPG4, 4:268–269, 4:269f
 motor neurons, 4:267–270, 4:269f
 Herlitz junctional epidermolysis bullosa
 (JEB), 1:641t
 Hermansky–Pudlack syndrome, 4:507–508,
 4:507t
 herpes simplex virus (HSV), 4:240, 4:240f,
 4:459
 herpesvirus entry mediator (HVEM), 4:454,
 4:455, 4:456f, 4:457f
 Hers disease, 2:427t, 2:431t, 2:432
 heterochromatin, 1:521, 3:296–297, 3:297f,
 3:314–315, 3:315f
 heterocyclic amino acids, 1:92t
 heteropentamers, 2:347, 2:348f
 heterotrimeric G proteins
 carboxyl methylation, 3:608–609, 3:608f
 phagocytosis, 3:439
 heterotrophic glutathione reductase (GR)-
 type FNRs, 2:301
 [Het-s] prion
 disease states, 3:568–569
 genetic properties, 3:566
 prion domains, 3:568f
 self-propagating amyloidoses, 3:567–568
 hexanoate
 branched-chain amino acid (BCAA)
 regulation, 1:245–246
 medium-chain acyl-CoA synthetase
 (MCAS), 2:292
 hexokinase, 2:543–547
 catalytic reaction mechanisms, 2:543
 classifications, 2:544
 evolutionary development, 2:544, 2:545f
 hexokinase IV
 insulin-secreting cells (β -cells), 2:547
 liver glucokinase/glucokinase regulatory
 protein (GCKR), 2:547
 sigmoidal kinetics, 2:546–547, 2:546f
 ischemic-induced metabolic signaling
 alterations, 3:68, 3:69f
 isozymic diversity, 2:543–544, 2:544t
 mammalian low-*K_m* hexokinases
 activity regulation, 2:546
 glucose-6-phosphate (glucose-6-P)
 inhibition, 2:546
 subcellular localization, 2:546
 structural characteristics, 2:544–546,
 2:545f
 sugar phosphorylation, 2:543
 hexosamine biosynthetic pathway, 1:656,
 3:351f
 hexosaminidases A and B, 4:292
 HExxH motif, 4:50–53
 HhaI enzyme, 2:131, 2:132t
 Hhat proteins, 2:736–737
 high-density lipoproteins (HDLs)

- adiponectin, 3:74
 vitamin E, 4:548
 highly unsaturated fatty acids (HUFAs)
 HUFA-response region with carbohydrate response element (CHORE), 2:288
 HUFA-response region with insulin response element (HUFA-RR/IRE), 2:287–288
 lipogenic gene expression, 2:287f, 2:288
 high-mannose-bearing mannose-6-phosphate residues, 2:461f, 2:463–464
 high-mobility group (HMG) proteins, 1:314–315
 high-performance liquid chromatography (HPLC), 2:590–591
 high-voltage-activated channels (HVA), 4:561t
 HinfIII enzyme, 2:136, 2:137t
 Hin invertase, 1:560–561
 hippocampus, 1:196–197, 1:197f, 2:172t
 hippurate, 4:495f
 hisactophilin, 3:438–439
 histamines, 2:496–497, 2:497t
 histidine
 acyl-coenzyme A synthetase, 2:291f
 anaplerotic pathways, 1:106f
 classifications, 1:92t
 cysteine protease action mechanisms, 1:589, 1:590f
 cytochrome *c*, 1:600–601
 glycation processes, 2:406
 histidine kinases, 1:629, 1:630f
 histidine protein (HPr), 2:387, 2:388f
 low barrier hydrogen bonds (LBHBs), 2:758–759, 2:758f
 metalloproteases, 3:87, 3:87f, 4:50–53
 phage libraries, 3:434t
 proline-histidine-serine-arginine-asparagine (PHSRN) sequence, 1:426
 tricarboxylic acid cycle, 4:439f
 zinc-finger domains, 4:573–574, 4:573f, 4:575, 4:576f
 histone-like nucleoid-structuring protein (H-NS), 3:367, 3:368f, 3:368t
 histones
 centromeric histone CENP-A, 1:448–450, 1:449f, 1:538
 core histones
 chromatin structure, 1:535–537
 gene variants
 characteristics, 1:531–532
 histone H1, 1:531
 nonhistone chromatin proteins (NHCPs), 1:531–532
 post-translational modifications, 1:530–531
 de novo DNA methylation, 3:6–8, 3:7f
 gene expression regulation, 1:314–315
 histone acetyltransferase (HAT), 1:311, 1:521, 1:530–531, 2:110, 4:541f, 4:543–544
 histone acetyltransferases (HATs), 4:60
 histone deacetylases (HDACs), 1:181, 1:181f, 1:312, 1:406–407, 1:407f, 1:521, 1:530–531, 2:110, 4:60
 histone H3 lysine 4 (H3K4)/histone H3 lysine 9 (H3K9) methylation, 3:6–8, 3:7f
 histone locus bodies, 3:300–301
 nucleosome regulation
 acetylation, 1:521, 1:522f, 4:60
 citullination, 4:62
 covalent modifications, 1:521, 1:522f, 4:59
 histone variants, 4:60f, 4:62–63
 methylation, 1:521, 1:522f, 4:60–61
 phosphorylation, 1:314–315, 1:521, 1:522f, 4:61
 proteolysis, 4:62
 sumoylation, 4:62
 ubiquitylation, 4:61–62
 nucleosome structure, 1:535–537, 1:536f
hixL/hixR recombination sites, 1:558
 HK022 Nun protein, 4:410–411
 HMGN1 protein, 4:404
 hnRNP protein family, 1:76–77, 1:78f, 1:79–80, 1:79f
 Hodgkin's disease, 4:161
 Holliday junctions (HJs), 1:449–450, 2:242–243, 2:558, 2:559f, 4:29, 4:30f, 4:33
 holocentric centromeres, 1:448
 homologous recombination (HR)
 break-and-copy models, 4:23
 DNA mismatch repair (MMR), 2:45–50
 DNA polymerase III holoenzyme, 2:95
 Escherichia coli anti-recombination activity, 2:47–48, 2:48f
 evolutionary role, 2:50
 loss of heterozygosity (LOH), 2:49–50, 2:49f
 mutation-related carcinomas, 2:49–50, 2:49f
 Saccharomyces cerevisiae anti-recombination activity, 2:48–49
 sequence divergence, 2:48f, 2:49–50, 2:49f
 DNA-strand transferases, 4:35–42
 adenosine triphosphate (ATP)
 hydrolysis, 4:38–39
 anti-recombination activity, 4:40
 chromosomal resolution and restoration, 4:36f, 4:37
 D-loop assays, 4:30f, 4:32, 4:39f, 4:41
 filament dissociation, 4:37
 filament formation, 4:35
 filter-binding assays, 4:39f, 4:40–41
 functional role, 4:35–37, 4:36f
 general discussion, 4:35
 in vitro assays, 4:39f, 4:40–41
 paranemic joint formation, 4:35–36, 4:40–41
 plectonemic joint formation, 4:37
 properties, 4:37–38, 4:37f
 recombination mediators
 nucleation, 4:40
 presynaptic filament formation, 4:40
 Rad51 paralogs, 4:31–32, 4:40
 Rad54 protein function, 4:32, 4:40
 secondary binding sites, 4:39–40
 signaling mechanisms, 4:41
 single molecule studies, 4:41
 structural characteristics, 4:38f, 4:39
 three-strand DNA exchange assay, 4:39f, 4:41
 translesion synthesis (TLS), 4:41
 double-strand break (DSB) repair, 4:45
 meiosis, 2:557–561
 axial elements, 2:560, 2:560f
 basic concepts, 3:21–22
 crossover pathways, 2:560
 crossover versus noncrossover
 chromosomes, 2:558, 2:559f, 3:22, 3:22f
 functional role, 2:557
 gene conversion, 2:558–559, 2:559f
 genetic interference, 2:560
 meiotic recombination checkpoint/
 pachytene checkpoint, 2:560
 mitotic versus meiotic recombination, 2:561
 molecular model
 double Holliday junction (DHJ), 2:558, 2:559f, 3:22, 3:22f
 double-strand breaks (DSBs), 2:247, 2:557, 2:559f, 3:22f, 3:266
 initiation phase, 2:557, 2:559f
 invasion phase, 2:558, 2:559f
 resection phase, 2:557–558, 2:559f
 resolution phase, 2:558, 2:559f
 recombination control mechanisms, 3:21–22, 3:22f
 recombination nodules, 2:560, 2:560f
 single-end invasion (SEI), 3:22
 synaptonemal complex, 2:560, 2:560f, 3:20, 3:21f, 3:22f
 synthesis-dependent strand annealing (SDSA), 2:558, 2:559f, 3:22
 ultrastructural level, 3:20–21
Homo sapiens
 actin-related proteins (Arps), 1:37f
 base excision repair (BER), 2:5t
 centromere region, 1:447t, 1:448f
 extra-chromosomally-primed (EP)
 retrotransposons, 3:285f
 glutathione peroxidases, 2:402f
 N-myristoyltransferase (Nmt), 3:642–643
 RecQ helicases, 4:43–44, 4:44f
 retrotransposons, 3:284f
 sliding clamp processivity
 9–1–1 complex, 4:239, 4:239f
 proliferating cell nuclear antigen (PCNA), 4:239, 4:239f
 hormonal control
 ascorbic acid, 4:531–532
 G_q family proteins, 2:496–497, 2:497t
 hormone response elements (HREs), 4:308–309, 4:310f
 hormone-sensitive lipase (HSL), 1:48–49, 1:49f, 1:50f
 horseradish, 4:533t
 horseradish peroxidases, 3:407–409
 horseshoe crabs, 2:702t
 Hoyeraal–Hreidarsson syndrome, 4:385
hsd gene expression, 2:72–74
 Hsp40 protein, 2:522–523, 2:523f, 3:569
 Hsp60 protein, 1:456, 4:269f, 4:271–272

- Hsp70 protein, 2:522–523, 2:523f, 3:569
- Hsp90 protein
- hypoxia inducible factor (HIF), 4:69
 - nitric oxide synthase (NOS), 1:343–346
 - proteasome inhibition, 2:345f, 3:593, 3:593f
 - protein folding processes, 1:456
 - steroid/thyroid hormone receptors, 4:308–309, 4:310f
 - stress responses, 2:522–523, 2:523f
- Hsp104 protein, 3:569
- HtrA2, 4:51f, 4:54–55
- HUA ENHANCER1 protein, 3:99–100
- human embryonic kidney (HEK293) cells, 1:328
- human galectin-1, 2:703f
- human galectin-7, 2:703f
- human herpes virus 8 (Kaposi sarcoma virus), 4:156–157
- human immunodeficiency virus (HIV-1)
- characteristics, 2:554
 - chemokine receptors, 1:484
 - exosomes and microvesicles (EMVs), 2:264
 - HIV-1 reverse transcriptase (RT), 2:548–553
 - drug design factors, 2:551
 - integrase, 2:105–107, 2:106f
 - non-nucleoside inhibitor-binding pocket (NNIBP), 2:552f
 - non-nucleoside reverse transcriptase inhibitor (NNRTI) resistance, 2:551, 2:552f
 - nucleoside analog reverse transcriptase inhibitor (NRTI) resistance, 2:550–551, 2:550f
 - p51 subunit, 2:548, 2:549f
 - p66 subunit, 2:548, 2:549f
 - ribonuclease H (RNase H), 2:105, 2:551–552
 - structure and functions
 - functional implications, 2:548–550
 - general discussion, 2:548
 - non-nucleoside reverse transcriptase inhibitors (NNRTIs), 2:549
 - schematic diagram, 2:549f
 - HIV protease, 2:554–556
 - definition, 2:554
 - drug-resistant variants
 - HIV reverse transcriptase (RT), 2:555–556
 - research objectives, 2:556
 - structural characteristics, 2:556
 - inhibitor binding
 - binding modes, 2:555, 2:556f
 - research background, 2:555
 - as model system, 2:556
 - structural characteristics
 - Gag/Pol polyproteins, 2:554–555
 - general discussion, 2:554
 - isolation and crystallization, 2:555
 - three-dimensional structure, 2:555f
 - viral protein assembly, 2:554–555
 - lysyl-tRNA synthetase, 4:443
 - nuclear factor kappa B (NF-κB), 3:304–305, 3:304t
- proteins
- Gag and Gag/Pol polyproteins, 2:102, 2:102f, 2:554, 2:555f, 4:443
 - Pol proteins, 2:554
 - research background, 2:101
 - RNA editing, 4:157
 - tumor necrosis factor receptors (TNFRs), 4:459
- human papillomavirus (HPV), 4:475–476
- Human Protein Atlas, 3:615t
- humans
- carbohydrate-responsive element binding protein (ChREBP) distribution, 1:374
 - color vision, 1:552f, 1:554
 - fatty acyl-coenzyme A (CoA) synthetases, 2:294
 - human milk, 4:533t
 - lectins, 2:701, 2:702t
 - mitochondrial DNA (mtDNA), 3:133f, 3:134t
 - multisubunit centromeres, 1:447
 - nicotinic acid adenine dinucleotide phosphate (NAADP)-induced calcium ion (Ca²⁺) release, 2:638t
 - nucleotide excision repair (NER), 3:336, 3:341, 3:342f, 3:342t
 - polymerase δ (Pol δ), 2:246t
 - prion diseases, 3:575t
 - RecQ helicases, 4:43–44, 4:44f
 - retrotransposons, 3:284f
 - ribozyme distribution, 4:146f
 - sliding clamp processivity
 - 9–1–1 complex, 4:239, 4:239f
 - proliferating cell nuclear antigen (PCNA), 4:239, 4:239f
- human stomach cancer transforming factor-1 (hst-1)*, 2:305
- Huntingtin polyglutamine sequence, 1:101t
- Huntington's disease
- amyloid deposits, 1:101t, 1:103
 - caspases, 1:394
 - dopamine receptors, 2:173
 - exosomes and microvesicles (EMVs), 2:264
 - lysosomal dysfunction, 4:507t
 - protein folding defects, 3:575t
 - rapamycin (mTOR), 3:198–199
 - ubiquitin–proteasome pathway, 3:592
- HU protein, 2:123
- Hutchinson–Gilford progeria syndrome, 3:313, 3:313f
- Hyalophora columbia*, 3:21f
- hyaluronan, 3:654, 3:655f
- Hydra*, 3:398
- HYDRA1* genes, 1:251
- Hydra vulgaris*, 3:90
- hydrin, 4:516t
- hydrochloric acid (HCl), 2:356
- hydrogenases, 2:562–567
- classifications and characteristics, 2:562
 - iron (Fe) hydrogenases
 - catalytic mechanisms, 2:564f, 2:566
 - structural characteristics, 2:564f, 2:566
 - iron-iron (FeFe) hydrogenases
 - catalytic cycle, 2:564f, 2:566
 - structural characteristics, 2:564f, 2:565
 - iron–sulfur (FeS) protein biogenesis, 2:663
- nickel-iron (NiFe) hydrogenases
- biomass-produced renewable hydrogen, 4:72–73, 4:73f, 4:74f
 - catalytic cycle, 2:563–565, 2:563f
 - structural characteristics, 2:562–563, 2:563f
- nickel-iron-selenium (NiFeSe) hydrogenases, 2:562–563
- periplasmic electron-transport systems, 3:403
- hydrogen (H)
- biomass-produced renewable hydrogen, 4:72–75
 - background information, 4:72
 - challenges, 4:74–75
 - future research areas, 4:75
 - organisms and fermentation pathways
 - dark fermentation, 4:73–74, 4:74f, 4:75f
 - photofermentation, 4:74, 4:75f
 - process mechanisms, 4:74–75, 4:75f
 - redox enzymes
 - characteristics, 4:72–73
 - [FeFe]-hydrogenases, 4:72, 4:73f, 4:74f
 - [NiFe]-hydrogenases, 4:72–73, 4:73f, 4:74f
 - nitrogenases, 4:73
 - substrates, 4:73
- H⁺ potential generation, 2:208–209, 2:209f
- H⁺ translocation, 2:206–208, 2:208f
- kinetic isotope effects (KIEs), 2:688
- low barrier hydrogen bonds (LBHBs), 2:756–759
- basic concepts, 2:756
 - characteristics, 2:756–757
 - enzymes
 - Δ5-3-ketosteroid isomerase, 2:759
 - serine esterases, 2:759
 - serine proteases, 2:758f
 - physicochemical properties
 - bond lengths, 2:757
 - deuterium fractionation factors, 2:758
 - spectroscopic properties, 2:757–758
 - research background, 2:756
- strong hydrogen bonds
- bond lengths, 2:757
 - deuterium fractionation factors, 2:758
 - serine proteases, 2:758–759
 - spectroscopic properties, 2:757–758
 - structural characteristics, 2:756–757, 2:757f
- very strong (single-well) hydrogen bonds
- bond lengths, 2:757
 - deuterium fractionation factors, 2:758
 - spectroscopic properties, 2:758
 - structural characteristics, 2:757, 2:757f
- weak hydrogen bonds
- bond lengths, 2:757
 - deuterium fractionation factors, 2:758
 - spectroscopic properties, 2:757
 - structural characteristics, 2:756, 2:757f
- purple bacteria electron acceptors and donors, 3:698

- hydrogen peroxide (H₂O₂)
 amine oxidases (AOs), 1:89–90
 biochemical free radical chain reactions, 3:168–169, 3:169f
 glutathione peroxidases, 2:403
 historical research, 3:168–169
 mitochondrial production, 3:169–170, 3:171f, 3:173f
 mitogen-activated protein kinase (MAPK), 3:173
 nuclear factor kappa B (NF-κB), 3:303t
 photosystem I (PSI) complex, 3:485
 physiological functions, 3:172–173
 redox signaling mechanisms, 3:411–412
 regulation mechanisms, 2:77, 3:407, 3:408f
 hydroxynaphthoquinone, 4:551–553, 4:553–554, 4:553f
 hydroperoxides, 2:403–404
 hydroperoxyl radical, 3:168–169
 hydroperoxytetraenoic acids (HPETEs), 2:188–189
 hydroquinone-cytochrome c oxidoreductase (QCR), 4:79, 4:80f
 3-hydroxy-3-methylglutaric aciduria, 4:495t
 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase
 cholesterol biosynthetic pathway, 1:517f, 1:518f
 regulated degradation, 1:519
 3-hydroxyacyl-CoA dehydrogenase (HADH), 1:385f, 1:386, 1:386–387
 2-hydroxyacyl-CoA lyase (HACL1), 1:206, 1:207f
 hydroxyadenine, 2:78, 2:78f
 3-hydroxyanthranilate 3,4-dioxygenase (3HAO), 1:175f
 3-hydroxy anthranilic acid, 1:175f
p-hydroxybenzoate hydroxylase, 2:311–312
 β-hydroxybutyrate (β-OHB)
 cerebral metabolism, 1:371–373
 mitochondrial permeability transition pore, 3:164
 starvation, 4:298–299, 4:302, 4:302f, 4:303–304, 4:305
 hydroxydecanoic acid, 3:551
 hydroxyketones, 1:206–207
 3-Hydroxy-L-kynurenine, 1:175f
 hydroxyl radical, 2:77, 3:168–169, 3:169f, 3:407, 3:408f
 hydroxymethylbilane, 2:540–541, 2:540f, 3:545f, 3:546
 hydroxymethyluracil, 2:2f
 hydroxyproline
 amino acid metabolism, 1:93t
 collagens
 signature (Gly-X_n-Y-)_n sequence
 basic concepts, 1:545
 content amounts, 1:546–547
 evolutionary origins, 1:548
 schematic diagram, 1:546f, 1:547f
 self-assembly control, 1:546
 glycine-phenylalanine-hydroxyproline-glycine-glutamic acid-arginine (GFOGER) sequence, 1:427t, 1:428
 O-linked glycosylation, 2:465, 2:466, 2:466f, 2:467f
 hydroxyquinones, 1:598
 hydroxytetraenoic acids, 2:188–189
 5-hydroxytryptamine (5HT)
See serotonin
 hyperammonemia-hyperornithinemia-homocitrullinemia (HHH) syndrome, 4:495f, 4:495t, 4:500–501
 hyperargininemia, 4:500
 hypercholesterolemia, 1:519
 hyperekplexia, 2:424
 hyperglycemia
 acute complications
 ketoacidosis, 1:654
 nonketotic hyperosmolar coma, 1:654–655
 characteristics, 1:649
 chronic complications
 hexosamine pathway, 1:656
 macrovascular disease, 1:655
 microvascular disease, 1:655–656
 oxidative stress, 1:656–657
 polyol pathway, 1:656
 protein kinase C (PKC) activation, 1:656
 glycogen metabolism, 2:426–427
 insulin resistance, 1:650, 1:650f
 maturity-onset diabetes of the young (MODY), 1:652–653
 neonatal diabetes, 1:653
 treatments and medications
 α-glucosidase inhibitors, 1:657–658
 complication-related treatments, 1:658
 hyperinsulinemia, 2:600–601, 2:606–607, 4:495t
 hyperphenylalaninemia, 3:669
 hyperpigmentation, 2:519
 hypertension, 1:139
 hypertrophic cardiomyopathy (HCM), 4:447–449, 4:448f
 hypocalcemia, 4:536–537
See also vitamin D
 hypomagnesemia with secondary hypocalcaemia (HSH), 4:416
 HYPONASTIC LEAVES1 protein, 3:99–100
 hypothalamus
 body weight regulation
 adiposity and satiety signal integration
 cholecystokinin (CCK), 3:78–79, 3:79f
 general discussion, 3:79
 ghrelin, 3:79
 schematic diagram, 3:79f
 arcuate nucleus (ARC)
 adiposity and satiety signal integration, 3:79f
 Agouti-related protein (AgRP), 3:77, 3:77f, 3:78f
 functional role, 3:76
 melanocortin neurons, 3:77
 neuropeptide Y (NPY), 3:77, 3:77f, 3:78f
 energy homeostasis, 3:76, 3:77f
 malonyl-CoA, 4:56–57
 model diagram, 3:77f
 neuronal systems
 arcuate nucleus (ARC), 3:76
 lateral hypothalamic area (LHA), 3:78
 paraventricular nucleus (PVN), 3:78, 3:79f
 ventromedial hypothalamus (VMH), 3:77–78, 3:78f
 research summary, 3:79
 dopamine receptors, 2:172t
 insulin metabolism, 2:606
 hypotrichosis, 1:638t
 hypotrichosis simplex of the scalp (HTSS), 1:638t
 hypoxanthine, 2:213f
 hypoxia inducible factor (HIF), 4:67–71
 energy metabolism regulation, 4:70–71
 functional roles, 4:67
 gene transcription regulation
 chromatin-modifying histone methylases/demethylases, 4:68
 DNA-binding functions, 4:67, 4:68f
 hypoxia response elements (HREs), 4:67, 4:68f
 microRNA transcriptional regulation, 4:68
 monomerization, 4:68
 NOTCH-dependent transcription, 4:68
 oxygen homeostasis regulation, 4:67
 transcriptional activation and repression responses, 4:67–68
 glycogen synthase kinase-3 (GSK3), 2:436t
 glycolysis, 2:446, 2:446f
 HIF-1α level regulation
 general discussion, 4:69–70
 HIF-1α mRNA levels, 4:69
 HIF-1α protein levels
 acute hypoxic conditions, 4:69
 differential HIF-1α/HIF-2α stability regulation, 4:69
 heat shock protein 90 (HSP90)
 function, 4:69
 HIF-1α expression and activity
 regulation, 4:70
 HIF-1α TAD function regulation, 4:70
 HIF-2α stability regulation, 4:69
 HIF-2α TAD function regulation, 4:70
 intermittent hypoxic conditions, 4:70
 prolonged hypoxic conditions, 4:69
 protein synthesis regulation, 4:69–70
 small ubiquitin-related modifiers (SUMOs), 4:69
 ischemic-induced metabolic signaling alterations, 3:68, 3:69f
 mitochondrial DNA (mtDNA) mutations, 1:215–216
 pyruvate dehydrogenase kinase (PDK)
 regulation, 4:324
 ubiquitin-proteasome degradation pathway, 3:620, 3:623t
 von Hippel-Lindau protein (VHL), 4:570, 4:571, 4:571f

I

IA-3 inhibitor, 1:140
 ibuprofen, 3:305
 I-cell disease, 4:507–508, 4:507t
 ichthyosis bullosa of Siemens, 2:670t

- idarubicin, 2:167f
idebenone, 2:333
I κ B kinase (IKK) complex, 4:378–379
illicicolin, 1:597f, 1:598
imidate, 3:230f
imidazolium salts, 2:408f
imino acids, 1:92t
iminoaspartate, 1:175f
immobilized-metal affinity chromatography (IMAC), 1:62–64
Immucillin-H, 2:213f
immune system
 innate immune system, 4:396
 leukocytes, 1:493
 Src (sarcoma) family kinases, 4:294–295
 toll-like receptors (TLRs), 4:396–397, 4:397t
 vitamin D, 4:538–539
immunoglobulins
 amyloid deposits, 1:101t
 B-cell antigen receptors (BCRs), 1:161, 1:162f
 binding immunoglobulin protein (BiP), 1:291t, 1:292–293
 fibroblast growth factor receptors (FGFRs), 2:303–304, 2:304f, 2:305f
 immunoglobulin α (Ig α), 1:161, 1:162, 1:163f
 immunoglobulin A (IgA)
 endocytosis, 2:199
 IgA pemphigus, 1:638t
 immunoglobulin A (IgA) FcRs, 2:569–571, 2:569f, 2:570f
 primary biliary cirrhosis (PBC), 3:562
 immunoglobulin β (Ig β), 1:161, 1:162, 1:163f
 immunoglobulin Fc receptors (FcRs), 2:568–573
 biological functions, 2:572
 characteristics and functional role, 2:568, 2:569f
 definition, 2:568
 domain structures, 2:570f
 immunoglobulin A (IgA) FcRs, 2:569–571, 2:569f, 2:570f
 immunoglobulin E (IgE) FcRs, 2:569
 immunoglobulin G (IgG) FcRs, 2:569, 2:569f, 2:570f
 immunoglobulin M (IgM) FcRs, 2:569–571
 leukocyte FcRs, 2:568, 2:569f, 2:571
 ligand binding sites, 2:570f, 2:571
 neoglycoproteins, 2:231
 neonatal Fc receptor (FcRn), 2:570f, 2:572
 nonleukocyte FcRs, 2:572
 phagocytosis, 3:437–438
 poly Ig receptor, 2:572
 receptor binding sites, 2:570f, 2:571
 research background, 2:572
 signal transduction complexes, 2:569f, 2:570f, 2:571
 immunoglobulin G (IgG), 1:255, 2:569, 2:569f, 2:570f, 3:437–438, 3:438f, 4:394f
 immunoglobulin superfamily, 4:229, 4:392, 4:394f
 membrane-bound immunoglobulin (mlg), 1:161, 1:162f
 reactive oxygen species (ROS), 2:77
 siglecs, 4:229–231
 binding sites, 4:231
 cell-type specific expression, 4:229
 cytosolic tails, 4:231
 functional roles, 4:230, 4:231
 genomic organization, 4:229–230
 historical research, 4:229
 immunoreceptor tyrosine-based inhibition motifs (ITIMs), 4:231
 N-acetylneuraminic acid (Neu5Ac)/N-glycolylneuraminic acid (Neu5Gc), 4:230–231
 phylogenetic relationships, 4:229–230
 siglec-1, 2:702t, 2:704t
 structural characteristics, 4:229, 4:230f
 subgroups, 4:229
 transposable elements (TEs), 3:281
 V(D)J recombination, 3:272, 3:272f
 immunogold electron microscopy, 1:124
 immunoproteasome, 3:596
 immunoreceptor tyrosine-based activated motifs (ITAMs)
 B-cell antigen receptors (BCRs), 1:162, 1:162f, 1:163f
 immunoglobulin Fc receptors (FcRs), 2:569, 2:569f, 2:571
 receptor–proximal signaling, 4:376
IMMUTANS, 1:146
IMP-IMBA Bioinformatics group, 3:615t
importin β family, 3:320, 3:321f
inclusion body myopathy, 3:223t
incretin hormones
 characteristics, 1:116–117
 diabetes treatments, 1:658
 gastric inhibitory peptide (GIP), 1:117
 glucagon-like peptide-1 (GLP-1), 1:116–117
 oxyntomodulin (OXM), 1:117
Indian hedgehog (Ihh) pathway, 1:193
Indo 1, 2:575–576, 2:575t
indoleamine 2,3-dioxygenase (IDO), 1:175f, 3:373
inducible stable DNA replication (iSDR), 4:24
infantile neuronal ceroid lipofuscinosis (INCL), 3:646
inflammation, 4:458
inflammatory bowel disease
 intermediate filaments, 2:632t
 nuclear factor kappa B (NF- κ B), 3:304–305, 3:304t
influenza viruses, 2:702t, 2:703–704, 2:704t
inherited disorders, 4:569, 4:569t
inhibitor of κ B kinase (IKK), 1:618f, 2:607
injection-localized amyloidosis, 3:575t
innate immune system, 4:396
inner arm dyneins, 2:185t, 2:187
inner nuclear membrane (INM), 2:751–752, 2:752f, 3:314, 3:315f
INO80 (inositol requiring 80) family, 4:60f, 4:64–65, 4:65f
inorganic phosphate, 1:18–19, 2:22–24, 3:163, 3:219–221, 3:221f
inorganic pyrophosphate (PPi), 4:442, 4:167, 4:441
inosine, 2:213f
inositol
 chemical structure, 2:581f
 inositol phosphatases/inositol phosphate kinases, 2:580–582
 inositol ring nomenclature, 2:580, 2:581f
 metabolic interrelationships and nomenclature, 2:581–582, 2:581f
 phosphate-transfer reactions, 2:580, 2:581f
 phosphotransferase system (PTS), 2:581–582
 regulatory input, 2:581f, 2:582
 inositol ring nomenclature, 2:580, 2:581f
 inositol trisphosphate (IP₃)
 B2 receptors, 1:242f
 B-cell antigen receptors (BCRs), 1:162–163, 1:163f
 calcium ion (Ca²⁺) signaling system, 2:583–589
 calcium oscillations, 1:322–323, 1:324f
 cellular control mechanisms
 basic concepts, 2:587–588, 2:588f
 cardiac cells, 2:588
 cell proliferation regulation, 2:587–588
 neuronal calcium ion (Ca²⁺) signaling, 2:588
 proteinase-activated receptors (PARs), 3:604
 characteristics, 2:583
 formation mechanisms, 2:583, 2:584f
 functional role, 2:584, 2:588f, 3:359–360, 3:360f
 metabolic pathways, 2:583–584, 2:585f
 spatiotemporal patterns
 activation mechanisms, 2:586, 2:587f
 basic concepts, 2:584–587, 2:585f
 domain structures, 2:586, 2:587f
 inositol trisphosphate (IP₃) receptors, 2:584f, 2:585f, 2:586, 2:587f
 receptor agonists, 2:585f, 2:586
 receptor antagonists, 2:586
 receptor modulation, 2:586–587
 cyclic adenosine diphosphate (ADP)-ribose (cADPr), 1:334–335
 endoplasmic reticulum (ER), 2:583, 2:584f, 2:585f, 2:587f
 G₁₂/G₁₃ family functions, 2:345f
 G $\beta\gamma$ subunit functions, 1:480–482, 1:482f, 2:377, 4:265
 Golgi apparatus, 2:482
 G-protein-coupled receptors (GPCRs), 4:225f
 G_q family proteins, 2:496, 2:497–498, 2:498f
 insulin-secreting cells (β -cells), 2:597f
 intercellular calcium ion (Ca²⁺) waves, 2:613–618
 active waves, 2:616–617
 characteristics and functional role, 2:613

- in vivo* conditions, 2:617–618
initiation mechanisms, 2:613–614
messenger regeneration
 external messengers, 2:617
 internal messengers, 2:617
nuclear calcium signal, 1:310–311
passive waves, 2:616–617
phase waves, 2:617
wave propagation mechanisms
 intercellular propagation, 2:615–616, 2:615f
 intracellular propagation, 2:614–615, 2:614f
intracellular calcium ion (Ca^{2+}) stores, 2:482
invertebrates, 2:637–638
IP₃ receptors (IP₃R), 2:653–656
 calcium homeostasis, 1:290, 1:310
 domain structures, 2:653–655, 2:654f, 2:655f
 historical research, 2:653
 intercellular calcium ion (Ca^{2+}) wave propagation, 2:614–615, 2:614f, 2:615f
 physiological functions, 2:655–656
 regulatory mechanisms, 2:655
 signal transduction, 2:654f
metabolic pathways, 2:583–584, 2:585f
neurokinin-1 receptor (NK1R), 4:362–363
neurotrophins, 3:241–243, 3:242f
phospholipase C (PLC), 2:654f, 3:467–468, 3:468f
platelet-activating factor receptor (PAFR), 3:536, 3:536f
PtdIns transfer protein (PTIP), 2:583–584, 2:585f
second messenger pathways, 1:43f
serotonin receptors, 4:225f, 4:226
signal transduction pathways, 1:59, 4:512
somatostatin (sst) receptors, 4:265
store-operated calcium (SOC) channels, 4:314–315
taste receptors, 4:368
T-cell receptors (TCRs), 4:374
transient receptor potential (TRP)
 channels, 3:359–360, 3:360f, 3:360t, 4:412–413
insects, 2:590–595
 biochemistry
 metabolomic research, 2:590–591
 mosquitoes
 blood meal metabolism, 2:591, 2:591f
 urea synthesis, 2:591–592, 2:592f
 characteristics, 2:590
 ecdysone, 2:590
 endocrine regulation
 background information, 2:590, 2:594
 ecdysone, 2:592–593
 juvenile hormone (JH), 2:593–594, 2:593f
 prothoracicotropic hormone (PTTH), 2:590, 2:594, 2:594f
 20-hydroxyecdysone (20E), 2:592–593, 2:593f
 juvenile hormone (JH), 2:590
 mitochondrial DNA (mtDNA), 3:134t
 prothoracicotropic hormone (PTTH), 2:590
 pterin distribution, 3:666
 insertion sequences (ISs), 3:275–276, 3:276f
 Insig, 1:517–519, 1:518f
 insulin, 2:602–607
 adiponectin, 3:74
 adiposity and satiety signal integration, 3:79f
 amino acid metabolism, 1:95
 anaplerotic pathways, 1:109–110, 1:109f
 appetite regulation, 4:303
 arachidonic acid regulated Ca^{2+} (ARC) channels, 1:131
 β -cells
 arachidonic acid regulated Ca^{2+} (ARC) channels, 1:129, 1:131
 glucokinase, 2:547
 insulin regulation, 4:302–303
 microRNA (miRNA) regulation, 4:191–192, 4:194t
 branched-chain amino acid (BCAA) regulation, 1:245–246
 calcium-sensing receptor (CaSR), 1:329–330
 characteristics, 2:602
 energy homeostasis, 3:77f, 3:78f
 functional role, 1:118
 historical research, 3:76
 hypothalamic malonyl-CoA, 4:57, 4:58f
 insulin-like growth factor-1 (IGF1) calbindin-D_{28k}, 1:287
 insulin-like growth factor-1 receptor (IGF-1R), 1:198, 1:200f, 2:435f, 2:602, 2:603f
 macroautophagy regulation, 1:150t
 nuclear factor kappa B (NF- κ B), 3:303t
 insulin-regulated amino peptidase (IRAP), 1:114
 insulin resistance
 basic concepts, 1:650, 1:650f, 2:606–607
 gestational diabetes, 1:654
 insulin receptor mutations, 1:653
 insulin secretory defects, 1:652
 metabolic syndrome (syndrome X), 1:654
 microRNA (miRNA)-based regulation, 4:194t
 molecular mechanisms, 1:650–652, 1:651f
 prediabetes, 1:654
 secondary diabetes mellitus, 1:653
 insulin secretagogues, 1:657
 insulin-secreting cells (β -cells), 2:596–601
 alpha-cell function, 2:600
 amino acids, 2:598–600
 augmentation pathway (APW), 2:598, 2:598f
 beta-cell therapies, 2:600–601, 2:601f
 fatty acids, 2:598–600
 functional role, 2:596
 GDH or K^{+} channel-linked hyperinsulinemic hypoglycemia, 2:600
 GK activators (GKAs), 2:597f, 2:600–601, 2:601f
 glucokinase, 2:547, 2:600
 insulin release stimulation, 2:599f
 K^{+} channel inhibitors and activators (KCI and KCOs), 2:597f, 2:600–601
 neuroendocrine modifications, 2:600
 research summary, 2:601
 signaling pathways, 2:597f, 2:598f
 triggering pathway (TPW), 2:598, 2:598f
 leptin regulation, 2:259
 macroautophagy regulation, 1:150t
 metabolic effects
 basic concepts, 2:603f, 2:606–607
 insulin resistance, 2:606–607
 microRNA (miRNA)
 glucose metabolism/insulin signaling, 4:193, 4:194t
 pancreatic β -cell function, 4:191–192, 4:194t
 pyruvate dehydrogenase kinase (PDK) regulation, 4:324–326
 signaling mechanisms
 insulin receptors (IRs), 2:602–603, 2:603f
 insulin receptor substrate (IRS) family characteristics, 1:650–652, 1:651f
 metabolic signaling pathways, 2:603–606, 2:604f
 insulin signal transduction
 Akt phosphorylation, 2:604–605, 2:604f
 Cbl/Cbl-associated protein (CAP) pathways, 2:605
 metabolic signaling pathways, 2:603–606, 2:604f
 mitogen-activated protein kinase (MAPK) pathway, 2:605–606
 phosphatidylinositol 3-kinase (PI3K) pathway, 2:604, 2:604f
 regulation mechanisms, 2:606
 starvation effects, 4:302–303, 4:303–304
 IntAct database, 3:615t
 integral membrane proteins, 4:204, 4:205f, 4:392, 4:394f
 integrase, 2:105–107, 2:106f
 integration host factor (IHF), 2:122, 2:122f, 3:367, 3:368f, 3:368t
 integrins
 $\alpha\beta$ 4 integrin heterodimer, 1:438–439, 1:439f, 1:640f, 1:641, 1:641t, 1:642–643
 activation and signaling mechanisms, 2:608–612
 basic concepts, 2:608
 conformational changes, 1:428, 2:609–610
 cytoplasmic domains, 1:428–429, 2:609
 focal adhesions, 1:428
 inside-out signaling
 characteristics, 2:609–610
 integrin activation, 2:609–610
 physiological functions, 2:609
 schematic diagram, 2:612f
 outside-in signaling
 adhesion sites, 2:610–611
 characteristics, 2:610–611

- integrins (*Continued*)
- cross-talk, 2:611
 - cytoskeletal effects, 2:611
 - intracellular signaling pathway
 - connections, 2:611
 - ligand binding, 2:611
 - schematic diagram, 2:612f
 - research summary, 2:611–612
 - signaling pathways, 1:429, 2:608–612
 - cation modulation, 1:428
 - cell migration, 1:438–440, 1:439f, 1:444–445
 - characteristics and functional role, 2:608
 - extracellular signal-regulated kinase (ERK), 1:439f, 2:611
 - focal adhesion kinase (FAK) family, 2:611
 - inside-out signaling
 - characteristics, 2:609–610
 - integrin activation
 - avidity modulation, 2:610
 - conformational changes, 2:609–610
 - cytoplasmic domains, 2:609
 - extracellular domains, 2:610
 - hematopoietic cells, 2:609
 - integrin–ligand interactions, 2:610
 - membrane-proximal residues, 2:609
 - negative regulation, 2:610
 - protein–protein interactions, 2:609–610, 2:609f
 - tail-binding proteins, 2:610
 - tissues, 2:609
 - transmembrane domains, 2:609–610
 - schematic diagram, 2:612f
 - integrin adhesion network, 2:320f
 - integrin-associated protein (IAP), 2:611
 - integrin-linked kinase (ILK), 2:320–321, 2:320f, 2:611
 - c-jun N terminal kinase (JNK) pathway, 2:611
 - ligand binding, 1:427–428
 - outside-in signaling
 - adhesion sites
 - formation mechanisms, 2:611
 - integrin–extracellular matrix (ECM)
 - contacts, 1:439f, 2:610–611, 2:611f
 - regulatory mechanisms, 2:611
 - characteristics, 2:610–611
 - cross-talk, 2:611
 - cytoskeletal effects, 2:611
 - intracellular signaling pathway
 - connections, 2:611
 - ligand binding, 2:611
 - schematic diagram, 2:612f
 - p21-associated kinase (PAK), 2:611
 - Ras proteins, 2:611
 - receptor family, 1:438–440, 1:439f
 - research summary, 2:611–612
 - structural characteristics, 1:427, 1:427t
 - Syc family kinases, 2:611
 - integrin integrase (IntI), 1:555–556
 - IntEnz database, 3:615t
 - intercellular calcium ion (Ca²⁺) waves, 2:613–618
 - active waves, 2:616–617
 - characteristics and functional role, 2:613
 - entry mechanisms, 1:127–128
 - in vivo* conditions, 2:617–618
 - initiation mechanisms, 2:613–614
 - messenger regeneration
 - external messengers, 2:617
 - internal messengers, 2:617
 - nuclear calcium signal, 1:310–311
 - passive waves, 2:616–617
 - phase waves, 2:617
 - wave propagation mechanisms
 - intercellular propagation
 - characteristics, 2:615–616
 - gap junction (GJ) communication, 2:615, 2:615f
 - paracrine communication, 2:615–616, 2:615f
 - spatial characteristics, 2:616
 - intracellular propagation, 2:614–615, 2:614f
- interferon
 - characteristics and functional role, 2:619
 - hematopoietin receptors, 2:526–527, 2:527f, 2:530
 - interferon receptors, 2:619–623
 - characteristics and functional role, 2:619
 - interferon- α/β receptors
 - antiviral effects, 2:623
 - gene expression, 2:621
 - janus tyrosine kinase (JAK)/signal transducer and activator of transcription (STAT) signaling, 2:622–623, 2:622f
 - ligand binding, 2:622
 - signal transduction, 2:622–623, 2:622f
 - structural characteristics, 2:621–622
 - interferon- γ receptors
 - antibacterial effects, 2:621
 - gene expression, 2:619
 - janus tyrosine kinase (JAK)/signal transducer and activator of transcription (STAT) signaling, 2:620–621, 2:620f
 - ligand binding, 2:620
 - signal transduction, 2:620–621, 2:620f
 - structural characteristics, 2:619–620
 - 20S proteasome, 3:596
 - Isg15 (interferon-induced 15 kDa protein), 4:468t, 4:470–471
 - nuclear factor kappa B (NF- κ B), 3:304t
 - signaling pathways, 1:616–617, 1:617f, 1:618f
 - toll-like receptors (TLRs), 4:398–399, 4:399f, 4:400f
- interleukins
 - B1 receptor regulation, 1:241
 - calcineurin-mediated signal transduction, 1:339–340, 1:340f
 - cAMP responsive element-binding protein 3 (CREB3), 4:52
 - c-fes* proto-oncogene (*Fes*), 1:452
 - general discussion, 1:613–617
 - hematopoietin receptors
 - expression regulation, 2:529
 - signal transduction, 2:529
 - structural characteristics, 2:526–527, 2:527f
 - interleukin-1 β -converting enzyme (ICE), 1:591–592
 - interleukin-1 (IL-1) family, 1:613–614, 1:614f
 - interleukin-1 receptor-associated kinase (IRAK) family protein kinases, 4:398, 4:399f, 4:400f
 - interleukin-2 (IL-2) family, 2:527f, 2:530
 - interleukin-3 (IL-3) family
 - expression regulation, 2:529
 - signal transduction, 2:529
 - structural characteristics, 2:526–527, 2:527f
 - interleukin-5 (IL-5) family
 - expression regulation, 2:529
 - signal transduction, 2:529
 - structural characteristics, 2:526–527
 - interleukin-6 (IL-6) family, 1:614–615, 1:614f, 2:526–527, 2:527f, 2:529–530
 - interleukin-10 (IL-10) family, 1:615, 1:615f
 - interleukin-17 (IL-17) family, 1:615–616, 1:616f
 - nuclear factor kappa B (NF- κ B), 3:303t, 3:304t
 - intermediate filaments, 2:631–636, 3:233–237
 - α -internexin, 2:632t, 2:634, 3:235
 - assembly and structure, 2:631–632, 2:633f
 - characteristics, 2:631, 3:233
 - cytoskeleton composition, 1:430
 - desmin, 2:632t, 2:634, 3:236
 - desmuslin, 2:632t, 2:634
 - disease connections, 2:635–636, 3:236
 - domain structures, 2:631, 2:633f
 - family members, 2:632t
 - filensin, 2:632t, 2:634
 - functional role
 - apoptosis, 2:636
 - cell proliferation regulation, 2:636
 - laminopathies, 2:635f, 2:636
 - spatial organization, 2:636
 - stress resistance, 2:635, 2:635f
 - glial fibrillary acidic protein (GFAP), 2:632t, 2:634
 - keratins, 2:632t, 2:634
 - lamins, 2:632t, 2:633f, 2:634, 3:296, 3:297f, 3:310, 3:311f
 - meiosis, 3:17
 - nestin, 2:632t, 2:634, 3:236
 - neurofilament triplet proteins (NFTPs), 2:632t, 2:633f, 2:634, 3:233–235, 3:234f
 - paranemin, 2:632t, 2:634
 - peripherin, 2:632t, 2:634, 3:236
 - phakinin, 2:632t, 2:634
 - regulation mechanisms, 2:632
 - superfamily classifications, 2:632–634, 2:632t
 - syncollin, 2:632t, 2:634
 - synemin, 2:632t, 2:634
 - vimentin, 1:47, 2:625–627, 2:632t, 2:634, 3:233
 - intermedin, 1:271, 1:271f

- intermembrane space-targeting signals (ITs), 3:632
- internal monooxygenases, 3:372
- International Molecular Exchange (IMEx) consortium, 3:612
- International Society for Biocuration, 3:616
- interphase I, 3:19f, 3:20
- InterPro database, 3:615t
- intestines
- acyl-coenzyme A synthetase
 - long-chain acyl-CoA synthetase (LCAS), 2:292–293
 - very long-chain acyl-CoA synthetase (VLCAS), 2:293–294
 - amino acid metabolism, 1:92
 - aquaporin water channels, 1:124t
 - calbindin-D_{09k}, 1:288
 - calbindin-D_{28k}, 1:284
 - calcium-sensing receptor (CaSR), 1:329
 - carbohydrate-responsive element binding protein (ChREBP) distribution, 1:374
 - digestion and absorption, 2:355f
 - mammalian glucose transport, 2:392t
 - proteinase-activated receptors (PARs), 3:605t
 - vitamin D, 4:538–539
- intracellular calcium ion (Ca²⁺), 2:574–578
- adenosine triphosphate (ATP), 1:127–128, 2:482
- anandamide synthesis, 2:194
- 2-arachidonoyl glyceryl ether synthesis, 2:194–195
- cyclic adenosine diphosphate (ADP)-ribose (cADPr), 1:332
- indicators
- bioluminescent probes, 2:577, 2:577f
 - chemical indicators, 2:574–577, 2:575t, 2:576f
 - functional role, 2:574
 - genetically encoded indicators, 2:575t, 2:577
 - green fluorescent protein (GFP), 2:577–578
 - research summary, 2:578–579
- intracellular calcium ion (Ca²⁺) waves, 2:640–647
- basic concepts, 2:640–644, 2:641f
 - calcium-induced calcium release (CICR), 2:644–645
 - physiological significance, 2:646–647
 - propagation mechanisms, 2:644–646
 - research summary, 2:647
- nicotinic acid adenine dinucleotide phosphate (NAADP), 1:332
- sarcoendoplasmic reticulum Ca²⁺ ATPase (SERCA), 1:127–128
- introns, 1:397–400, 4:147–153
- alternative splicing, 1:75, 1:76f, 3:60–61, 4:187–188
 - characteristics, 4:142, 4:143f, 4:147, 4:148f
 - chemical splicing mechanisms, 4:185, 4:186f
 - distribution, 4:146f
 - exon–intron architecture, 3:142–144, 3:143f, 4:186–187
 - exon/intron boundaries, 1:75, 1:76f, 1:397, 1:398f, 4:149f
 - folding processes
 - ai5γ ribozyme folding, 4:151, 4:152f
 - autonomous *in vitro* folding, 4:151, 4:152f
 - protein-facilitated folding and splicing, 4:151–152, 4:152f
 - research background, 4:151–152
 - historical research, 4:185
 - intronic splicing enhancers (ISEs), 1:76, 1:78f, 3:62, 3:62f
 - intronic splicing silencers (ISSs), 3:62f
 - mosaic mitochondrial genes, 3:142–144, 3:143f
 - research summary, 4:153
 - ribozymes, 4:188
 - self-splicing reactions, 1:397, 1:398f, 4:188, 4:188f
 - spliceosomes, 4:152–153, 4:153f
 - splicing pathways, 3:59–60, 3:60f, 4:147, 4:149f, 4:186f, 4:187f
 - structural characteristics
 - domain I, 4:147–148, 4:148f, 4:151f
 - domain II, 4:148–149, 4:148f, 4:151f
 - domain III, 4:148f, 4:149, 4:151f
 - domain IV, 4:148f, 4:149, 4:151f
 - domain V, 4:148f, 4:149–150, 4:151f
 - domain VI, 4:148f, 4:149f, 4:150, 4:151f
 - secondary structures, 1:397, 1:398f
 - tertiary structures, 1:397–399, 1:398f
 - substrate recognition
 - G●U wobble pairs, 1:398f, 1:399–400, 1:399f, 1:400f
 - guanosine-binding sites, 1:398f, 1:399, 1:399f, 1:400f
 - helix P1/5' splice site recognition, 1:398f, 1:399–400, 1:399f, 1:400f
 - P4–P6 domain, 1:398f, 1:399f, 1:400f
 - three-dimensional structure, 4:150, 4:151f
- invadopodia, 1:443–444
- invertebrates
- calpains, 1:354f
 - hexokinases, 2:544t
 - lectins, 2:701, 2:702f, 2:702t, 2:704t
 - nicotinic acid adenine dinucleotide phosphate (NAADP)-induced calcium ion (Ca²⁺) release, 2:637–638, 2:638t
 - Src (sarcoma) family kinases, 4:297
- ion channel protein superfamily, 2:648–652
- cyclic nucleotide-regulated cation channels, 1:586–588
- background information, 1:586
 - cyclic nucleotide-gated (CNG) ion channels
 - background information, 1:586
 - structural and physical characteristics, 1:586–587, 1:587f, 2:650
 - hyperpolarization-activated CNG (HCN) ion channels
 - background information, 1:586
 - structural and physical characteristics, 1:586, 1:587–588, 1:587f, 2:650
 - structural and physical characteristics, 1:586, 1:587f
- degenerin (DEG)/ENaC superfamily, 4:367–368, 4:367f
- dopamine receptors, 2:172
- electrical signals, 2:648
- functional role, 2:651, 2:651f
- GABA_A receptors, 2:347, 2:348f, 2:349
- inwardly rectifying potassium channels (K_{ir}), 2:649f, 2:650, 2:651f, 4:557–559, 4:558f
- muscarinic acetylcholine receptors, 3:216
- phosphatidylinositol-4,5-bisphosphate (PIP₂)
- affinity and sensitivity, 2:747
 - analytical methodologies, 2:746t
 - cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:563–564
 - functional role, 2:748
 - intracellular regulation overlap, 2:747, 2:747f
 - modulation, 2:746–747, 2:748
 - regulation specificity, 2:747–748
 - research summary, 2:748
 - voltage-sensitive phosphatases, 2:746f
- research summary, 2:651–652
- second messenger-gated channels, 2:649–650, 2:649f
- sensory ion channels, 2:649–650
- structural characteristics, 2:651, 2:651f
- taste receptors, 4:367–368, 4:367f
- transient receptor potential (TRP) channels
- structural and physical characteristics, 2:650
 - transport modulation, 2:746–747
- two-pore-motif potassium channels (K_{2P}), 2:649f, 2:650–651
- voltage-dependent calcium channels (VDCCs), 2:597f
- voltage-gated calcium channels, 4:560–563
- calcium/calmodulin-dependent protein kinase II (CaMKII), 1:297–298, 1:298f
 - calcium channel-associated transcriptional regulator (CCAT), 1:314
 - calcium ion (Ca²⁺) currents
 - channelopathies, 4:563
 - physiological functions, 4:562–563, 4:562f
 - calcium oscillations, 1:322–325
 - characteristics, 1:322, 1:323f
 - decoding methods, 1:325
 - functional role and significance, 1:323–325, 1:324f
 - signaling pathways, 1:322–323, 1:324f
 - channel characteristics, 4:561t
 - disease connections, 4:563
 - functional role, 2:648–649, 4:560
 - glutamate receptors, 2:713, 2:720f
 - G proteins, 4:561
 - high-voltage-activated channels (HVA), 4:561t
 - intracellular signaling pathways, 2:714f

ion channel protein superfamily (*Continued*)

- low-voltage-activated channels (LVA), 4:561*t*
- molecular properties
 - calmodulin (CaM) binding, 4:560
 - domain structures, 4:560, 4:561*f*
 - isoleucine-glutamine (IQ) motif, 4:560
 - selectivity, 4:560
 - phosphorylation, 4:561
 - schematic diagram, 2:649*f*
 - transport modulation, 2:746–747
- voltage-gated potassium channels, 1:379, 1:380*f*, 2:649, 2:649*f*, 2:746–747, 2:746*f*, 4:558*f*
- voltage-gated sodium channels, 4:564–569
 - cardiac excitation–contraction (E–C)
 - coupling, 1:379, 1:380*f*
 - evolutionary development, 4:564–565
 - functional role, 2:648–649, 4:564
 - gene characterization, 4:564–565, 4:565*f*
 - inherited disorders, 4:569, 4:569*t*
 - intracellular signaling pathways, 2:712–713, 2:714*f*
 - localizations, 4:566
 - molecular bases, 4:566
 - neurotoxin receptor sites, 4:568–569, 4:568*t*
 - pharmacological functions, 4:568–569, 4:568*f*
 - phylogenetic relationships, 4:565*f*
 - schematic diagram, 2:649*f*
 - structural characteristics, 4:566–568, 4:567*f*
 - subunit structure and functions, 4:564, 4:565*f*
- ionizing radiation
 - non-homologous end joining (NHEJ), 3:273, 3:273*t*
 - reactive oxygen species (ROS), 2:77
- ion-mobility spectrometry (IMS), 3:11, 3:12*f*, 3:14–15
- ionotropic-glutamate (iGlu) receptors, 2:395, 2:714–716, 2:715*t*, 2:716*t*, 2:717*f*
- iron (Fe)
 - bacterial reductive reactions, 1:487*t*
 - iron (Fe) hydrogenases
 - catalytic mechanisms, 2:564*f*, 2:566
 - structural characteristics, 2:564*f*, 2:566
 - iron (Fe(II))/2OG-dependent
 - dioxygenases, 2:13–14
 - iron (Fe(III)) respiration, 3:405
 - iron-iron (FeFe) hydrogenases
 - biomass-produced renewable hydrogen, 4:72, 4:73*f*, 4:74*f*
 - catalytic cycle, 2:564*f*, 2:566
 - structural characteristics, 2:564*f*, 2:565
 - iron protoporphyrin IX, 1:342, 1:343*f*, 2:539, 2:540*f*
 - iron regulatory protein (IRP), 2:661
 - iron–sulfur (FeS) proteins, 2:657–664
 - bacterial biogenesis
 - hydrogenases, 2:663
 - iron–molybdenum cofactor (FeMo-co), 2:659*f*, 2:663
 - ISC (iron–sulfur cluster) system
 - assembly machinery, 2:662, 2:662*f*
 - nitrogenases, 2:663
 - nitrogen fixation (NIF) system, 2:663
 - sulfur mobilization machinery, 2:662–663
 - biogenesis
 - bacteria, 2:662–663, 2:662*f*
 - basic concepts, 2:662
 - eukaryotes, 2:663–664, 2:663*f*
 - catalytic iron–sulfur clusters (FeS)
 - analytical analyses, 2:661–662
 - catalytic reaction mechanisms, 2:660–661, 2:661*f*
 - disulfide bond reduction, 2:660–661, 2:661*f*
 - free-radical reactions, 2:660, 2:661*f*
 - regulatory functions, 2:661
 - sensing functions, 2:661
 - eukaryotic biogenesis
 - chloroplasts, 2:663
 - cytosolic system, 2:663–664
 - mitochondrial Fe–S cluster assembly machinery, 2:663, 2:663*f*
 - functional role
 - analytical analyses, 2:661–662
 - catalytic iron–sulfur clusters (FeS), 2:660–661, 2:661*f*
 - electron transfer, 2:660
 - free-radical reactions, 2:660
 - protein structure stabilization, 2:660
 - reducing powers, 2:660
 - regulatory functions, 2:661
 - sensing functions, 2:661
 - iron–sulfur clusters (FeS)
 - characteristics and functional role, 2:335–336, 2:336*f*
 - electronic structure, 2:657–658
 - F_x and F_b clusters, 2:338–339
 - F_x- and F_b-containing proteins, 2:339–340, 2:339*f*
 - ferredoxins (Fd), 2:297–298, 2:301*f*, 2:661*f*, 3:485, 3:505, 3:505*f*
 - frataxin functions, 2:333
 - F_x cluster, 2:338, 2:338*f*
 - magnetic characteristics, 2:657–658
 - nomenclature, 2:657
 - respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:342–343, 4:342*f*, 4:345*f*
 - respiratory chain complex III, 1:596, 1:596*f*, 3:663–664, 3:664*f*
 - stability, 2:657
 - structural characteristics, 2:657, 2:658*f*
 - succinate:quinone oxidoreductases (SQORs), 4:95
 - type I reaction centers, 2:335–342
 - nomenclature, 2:658
 - respiratory chain complexes, 4:83
 - Rieske protein
 - cytochrome *b₆* complex, 2:335, 4:335–338, 4:337*f*, 4:338*t*, 4:339*t*
 - cytochrome *bc₁* complex, 1:596, 1:596*f*, 3:30–31, 3:663–664, 3:664*f*

- cytochrome *f*, 3:32
- posttranslational modifications, 4:338
- protonmotive Q cycle, 3:663–664, 3:664*f*
- structural characteristics
 - flexibility, 2:659–660
 - higher-order structures, 2:659
 - metallic superclusters, 2:659, 2:659*f*
 - primary structures, 2:658
 - secondary structures, 2:658–659
 - tertiary structures, 2:658–659
- nickel-iron (NiFe) hydrogenases
 - biomass-produced renewable hydrogen, 4:72–73, 4:73*f*, 4:74*f*
 - catalytic cycle, 2:563–565, 2:563*f*
 - structural characteristics, 2:562–563, 2:563*f*
- nickel-iron-selenium (NiFeSe)
 - hydrogenases, 2:562–563
- periplasmic electron-transport systems, 3:405
- protonmotive Q cycle, 3:664*f*
- superoxide dismutases (SODs), 4:353
- ischemic heart diseases, 3:65–71
 - anaplerotic pathways, 1:109, 1:109*f*
 - basic concepts, 3:65
 - energy metabolism
 - basic principles, 3:67, 3:68*f*
 - ischemic-induced metabolic signaling alterations, 3:67–69, 3:69*f*
 - metabolic-based treatment approaches, 3:71, 3:71*t*
 - reperfused ischemic heart
 - basic principles, 3:69–70, 3:70*f*
 - intracellular acidosis/cardiac efficiency, 3:70–71
 - research summary, 3:71
 - nuclear factor kappa B (NF-κB), 3:304*t*
- ischemic post-conditioning (IPoC), 3:111
- ischemic preconditioning (IPC), 3:111
- Isg15 (interferon-induced 15 kDa protein), 4:468*t*, 4:470–471
- Isl1 (Isl1) transcription factor, 1:184*f*, 1:185
- islet amyloid polypeptide (IAPP), 1:101*t*, 1:103
- isoalloxazine, 2:310*f*
- isobutyl-1-methylxanthine (IBMX), 2:597*f*
- isobutyryl-CoA, 1:244, 1:245*f*
- isobutyryl-CoA mutase, 1:158*t*
- isocitrate
 - anaplerotic pathways, 1:106*f*
 - citric acid cycle, 4:93*f*, 4:438*f*
 - isocitrate dehydrogenase, 4:439*t*
 - NAD-dependent isocitric dehydrogenase (NAD-IDH), 3:124*f*
 - tricarboxylic acid cycle, 4:439*f*
- isoleucine
 - anaplerotic pathways, 1:106*f*
 - classifications, 1:91, 1:92*t*
 - degradative pathways, 1:244, 1:245*f*
 - isoleucine-glutamine (IQ) motif, 4:560
 - mitochondrial DNA (mtDNA), 3:134*t*
 - phage libraries, 3:434*t*
 - tricarboxylic acid cycle, 4:439*f*
 - vasopressin/oxytocin family, 4:516*t*

isomerases, 2:759
 isopentenyltransferases (IPTs), 1:627–628
 isoprenylcysteine methyltransferase (ICMT), 3:608–609
 isopropyl β -D-1-thiogalactopyranoside (IPTG), 2:696f, 2:751, 2:751f, 2:753f
 isoproterenol, 1:59
 isotocin, 4:516t
 isovaleryl-CoA, 1:244, 1:245f
 ISWI family, 4:60f, 4:64
 ivermectin, 2:423

J

jack beans, 2:702t
 jackfruit, 2:702t
 Jackson–Lawler disease, 2:670t
 Jadhasson–Lewandowsky disease, 2:670t
 Jansen's metaphyseal chondrodysplasia, 3:384, 3:389
 janus kinase (JAK)
 angiotensin receptors, 1:113
 c-fes proto-oncogene (Fes), 1:452
 chemokine receptors, 1:480–482
 hematopoietin receptors, 2:526, 2:528f
 interferon receptors, 2:620–621, 2:620f, 2:622f
 leptin expression, 2:260f
 neurogenesis, 1:198
 signaling pathways, 1:616, 1:617f, 1:618f
 jasplakinolide, 1:16–17, 1:16t
 jumonji C domain-containing histone demethylase (JHDM) family, 4:61
 junctional adhesion molecules (JAMs), 4:392, 4:394f
 junctional epidermolysis bullosa (JEB), 1:641t
 junctional epidermolysis bullosa with pyloric atresia (JEB-PA), 1:641t
 c-jun N terminal kinase (JNK) pathway
 angiotensin receptors, 1:113
 c-fes proto-oncogene (Fes), 1:453
 functional role, 3:177–178
 glycogen synthase kinase-3 (GSK3), 2:435f, 2:436t
 hemidesmosome function, 1:642–643
 hydrogen peroxide (H_2O_2), 3:173, 3:173f
 insulin resistance, 2:606–607
 insulin signal transduction, 2:605–606
 integrin signaling, 2:611
 mitochondrion-dependent apoptosis, 3:174
 muscarinic acetylcholine receptors, 3:216
 neurogenesis, 1:199
 nitric oxide (NO), 3:173, 3:173f
 p75 neurotrophin receptor (p75NTR), 3:245, 3:245f
 phylogenetic relationships, 3:177f
 signaling pathways, 1:617–619, 1:618f, 1:620f, 3:177f
 toll-like receptors (TLRs), 4:398–399
 juvenile hormone (JH), 2:590, 2:593–594, 2:593f
 juxtaglomerular granule, 2:521t
 juxtamembrane, 3:539f

K

kainic acid (KA), 2:714–716, 2:715t, 2:716t, 2:717f, 2:718
 kale, 4:533t
 kallikrein
 angiotensin receptors, 1:113
 kallikrein–kininogen–kinin system, 1:240, 1:241f
kaposi sarcoma Fgf (kFgf), 2:305
 Kaposi sarcoma virus, 2:305, 4:156–157
 karyopherin β family, 3:320, 3:321f, 4:7, 4:8–9
 KASH domain proteins, 3:314, 3:315f
 KAT1/2 channels, 4:557–559, 4:558f
 katanin, 1:1–2, 1:3t, 1:4f, 3:105
k_{cat}/K_m ratio, 2:217
 K-Cl cotransporter 2 (KCC2), 1:199, 1:200f
 KCNQ channels, 4:556–557
 Kearns–Sayre syndrome, 1:214
 KEGG Database, 3:615t
 keratan sulfate (KS), 3:655f, 3:656
 keratinocytes
 intermediate filaments, 2:636
 keratinocyte growth factor (KGF), 2:307
 vitamin D, 4:538–539
 keratins, 2:665–671
 characteristics, 2:634, 2:665, 2:668f
 classifications, 2:632t, 2:665, 2:666f
 disease connections, 2:670–671, 2:670t
 domain structures, 2:667, 2:668f
 functional role
 signaling pathways, 2:670
 structural support, 2:668f, 2:669–670
 gene expression
 hair follicles, 2:669, 2:669t
 mammalian skin tissue, 2:667–669, 2:668f, 2:669t
 thick epidermis, 2:669, 2:669t
 thin epidermis, 2:667–669, 2:669t
 intracellular regulation, 2:667, 2:668f
 posttranslational modifications, 2:667
 self-assembly control, 2:667
 tissue expression, 2:665–667, 2:666f
 ketoacidosis, 1:654
 3-ketoacyl-CoA thiolase (KT), 1:385f, 1:386, 1:386–387
 ketoamines, 2:405, 2:406f
 α -keto- β -methylisovalerate (KMV), 1:244, 1:245f
 ketogenesis
 amino acids, 1:91
 insulin metabolism, 2:603f, 2:606–607
 starvation, 4:298, 4:302–303, 4:305
 α -ketoglutarate
 anaplerotic pathways
 citric acid cycle, 1:106f, 4:93f
 ischemia, 1:109, 1:109f
 starvation-based renal ammonia formation, 1:109
 substrates, 1:106, 1:108f
 ascorbic acid, 4:531–532
 citric acid cycle, 4:438f, 4:439t
 gluconeogenesis, 2:385, 2:386f
 insect biochemistry, 2:590–591

α -ketoglutarate dehydrogenase, 3:124f, 4:439t
 starvation, 4:307
 tricarboxylic acid cycle, 4:439f
 α -ketoisocaproate (KIC), 1:244, 1:245f, 1:246f
 α -ketoisovalerate (KIV), 1:244, 1:245f
 ketones
 fatty acid synthesis pathway, 2:276f
 ketone bodies
 branched-chain amino acid (BCAA) regulation, 1:245–246
 cerebral metabolism, 1:371–373
 free fatty acids (FFAs), 4:305
 starvation, 4:301–302, 4:302f, 4:305
 synthesis, 4:298–299
 lipogenic gene expression, 2:287f
 renal gluconeogenesis, 2:385, 2:386f, 4:305–306, 4:306f
 ketoses, 2:409f
 Δ 5-3-ketosteroid isomerase, 2:759
 kidneys
 acyl-coenzyme A synthetase
 ACSF proteins, 2:294
 medium-chain acyl-CoA synthetase (MCAS), 2:292
 short-chain acyl-CoA synthetase (SCAS), 2:292
 very long-chain acyl-CoA synthetase (VLCAS), 2:293–294
 amyloid deposits, 1:101t
 anaplerotic pathways, 1:109
 aquaporin water channels, 1:124t
 calbindin-D_{28k}, 1:285
 calcium-sensing receptor (CaSR), 1:329
 carbohydrate-responsive element binding protein (ChREBP) distribution, 1:374
 cyclic nucleotide-gated (CNG) ion channels, 1:586–587
 gluconeogenesis, 2:384, 2:385, 2:386f, 4:305–306, 4:306f
 mammalian glucose transport, 2:392t
 peroxisome proliferator-activated receptor alpha (PPAR α), 3:420–421
 proteinase-activated receptors (PARs), 3:605t
 sodium ion/calcium ion ($\text{Na}^+/\text{Ca}^{2+}$) exchanger, 3:48
 starvation-based renal ammonia formation, 1:109, 4:298, 4:305–306
 vitamin D, 4:538–539
 voltage-gated calcium channels, 4:561t
 KIF5A, 4:269f, 4:270
 kifunensine (KF), 2:459f
 KIN32 domain, 2:266t, 2:267–268
 kinases
 acetate kinase, 4:74f
 A-kinase anchoring protein (AKAP) family, 1:129, 1:575, 1:582, 2:345f
 AMP-activated protein kinase (AMPK)
 body weight regulation, 4:58f
 cellular fatty acid metabolism regulation, 4:56
 diabetes treatments
 metformin, 1:657
 weight loss and exercise, 1:657

kinases (*Continued*)

- fatty acid metabolism regulation, 2:286
- fatty acid synthesis pathway, 2:275–279, 2:276f
- insulin signal transduction, 2:604f, 2:605
- ischemic-induced metabolic signaling alterations, 3:67–69, 3:69f
- lipogenic gene expression, 2:287f
- microRNA (miRNA)-based regulation, 4:193
- angiotensin receptors, 1:113
- apoptosis signal-regulated kinase-1 (ASK-1), 3:411f, 3:412
- ataxia telangiectasia mutated (ATM)
 - kinase, 1:411, 1:411t, 1:412f, 1:413–414, 1:413f, 1:415–416, 2:31, 3:269, 3:273t
- calcium/calmodulin-dependent protein kinase II (CaMKII), 1:297–303
 - autophosphorylation
 - autonomous activity, 1:299
 - calmodulin (CaM) trapping, 1:300
 - neuronal activity monitoring, 1:301, 1:301f
 - structure and functions, 1:299, 1:300f
 - substrate anchoring, 1:300
 - threonine (Thr³⁰⁵/Thr³⁰⁶)
 - autophosphorylation, 1:300–301
 - cardiac regulation, 1:302
 - characteristics and functional role, 1:297
 - domain structures, 1:298–299, 1:299f
 - intracellular localization, 1:297–298, 1:298f
 - long-term potentiation, 1:302
 - memory and learning regulation, 1:302
 - structure and functions, 1:302–303
 - substrates, 1:297–298, 1:298f
- calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinases (CaMKs), 1:304–309
 - broad-specificity calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinases (CaMKs)
 - calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinase I (CaMKI), 1:305f, 1:306, 1:311
 - calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinase II (CaMKII), 1:305f, 1:306–307, 1:308f, 1:311, 1:381
 - calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinase IV (CaMKIV), 1:305f, 1:307, 1:308f, 1:311
 - calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinase kinase (CaMKK), 1:305f, 1:307–308, 1:308f
 - calcium/calmodulin-dependent protein kinase II (CaMKII), 1:297–303, 1:311
 - calcium ion (Ca²⁺) signaling
 - modulation, 1:304, 1:305f, 2:719–720, 2:720f
 - calcium sensors, 1:311
 - domain structures, 1:305f
 - eukaryotic elongation factor 2 kinase (eEF-2K), 1:305f, 1:306
 - gene expression regulation, 1:311
 - general discussion, 1:304
 - myosin light chain kinase (MLCK), 1:305–306, 1:305f
 - narrow-specificity calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinases (CaMKs), 1:305–306
 - neurotransmitter transporters, 3:239
 - phosphorylase kinase (PhK), 1:305f, 1:306
 - research summary, 1:308
- calcium signal mediators, 1:311
- casein kinase (CK2 α), 1:514, 1:514f
- Cdc7 kinase, 1:401–402, 1:402t, 1:403, 1:403f, 1:404f
- ceramide kinase (CERK), 4:283f, 4:284–285
- c-fes proto-oncogene (Fes), 1:452
- Chk1 and Chk2 effector kinases, 1:411t, 1:412f, 1:414
- chloroplast protein kinase network, 1:514, 1:514f
- chloroplast signaling, 1:513, 1:513f
- copper-transporting ATPases (copper pumps), 3:674
- creatine kinase, 3:81, 3:82f
- Csk protein kinase, 4:295–296, 4:296f
- cyclic nucleotide-dependent protein kinases, 1:574–579
 - binding specificity and affinity, 1:577–578
 - cAMP-dependent protein kinase, 1:575–577, 1:576f
 - cGMP-dependent protein kinase, 1:577
 - cyclic adenosine monophosphate (cAMP), 1:574, 1:575f
 - cyclic guanosine monophosphate (cGMP), 1:574, 1:575f
 - feed-forward control, 1:578–579, 1:578f
 - intracellular receptors, 1:574, 1:575f
 - negative feedback, 1:578–579, 1:578f
 - structural domains, 1:574–575, 1:576f
 - substrate specificity, 1:578
- cyclin-dependent kinases (CDKs)
 - cell-cycle-dependent activity, 2:242
 - cell growth regulation, 1:408
 - DNA damage checkpoints proteins, 1:410–416
 - activation mechanisms, 1:410–412, 1:412f
 - apoptosis, 1:415
 - ataxia telangiectasia-mutated and Rad3-related (ATR) pathway, 1:411t, 1:413f, 1:414
 - ataxia telangiectasia mutated (ATM) kinase, 1:411, 1:411t, 1:412f, 1:413–414, 1:413f, 1:415–416
 - background information, 1:410
 - BRCA1 C-terminal (BRCT) domains, 1:411t, 1:412f, 1:414
 - Chk1 and Chk2 effector kinases, 1:411t, 1:412f, 1:414
 - chromatin modifications, 1:413f, 1:414
 - component proteins, 1:411t
 - disease connections, 1:415–416
 - forkhead-associated (FHA) binding domains, 1:412f, 1:414
 - G1/S checkpoint, 1:411, 1:412f, 1:415–416
 - G2/M checkpoint, 1:412, 1:412f, 1:415–416
 - Intra-S checkpoint, 1:411, 1:412f, 1:415–416
 - lesion sensors, 1:412, 1:413f
 - phases, 1:411f
 - phosphoinositide 3-kinase-like kinase (PIKK) family, 1:412–413, 1:413f
 - phospho-peptide-binding domains, 1:412f, 1:414
 - research summary, 1:416
 - senescence, 1:415
 - signal transduction, 1:412–415
 - spatiotemporal organization, 1:414–415
- eukaryotic DNA replication, 2:108, 2:112
- functional role, 1:403
- G₀ phase and G₁ phase cell-cycle regulation, 1:406–407, 1:407f
- retinoblastoma protein (pRB), 1:406–407, 1:407f, 4:104
- S phase initiation, 1:401–402, 1:402t, 1:403f, 1:404f, 1:406–409
- stem cell proliferation, 1:408
- Dbf4-dependent kinases (DDK), 2:108, 2:112
- diacylglycerol kinases (DGKs)
 - compartmentalization, 1:660–661
 - EF-hand proteins, 1:319t
 - gene deletion studies, 1:661–662
 - isoforms, 1:659, 1:661f
 - regulatory mechanisms, 1:659–660
 - research summary, 1:663
 - signaling protein interactions, 1:661
 - structural domains, 1:659, 1:661f
- DNA damage checkpoints proteins, 1:411t
- DNA-dependent protein kinase catalytic subunit (DNA-PK_{cs}), 3:269–271, 3:270f, 3:272f
- eukaryotic elongation factor 2 kinase (eEF-2K), 1:305f, 1:306
- extracellular signal-regulated kinase (ERK)
 - angiotensin receptors, 1:113
 - calcium ion (Ca²⁺) signaling system, 2:719–720, 2:720f
 - c-fes proto-oncogene (Fes), 1:453
 - chemokine receptors, 1:480–482, 1:482f
 - cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:564, 1:564f
 - ERK1/ERK2, 3:165–166, 3:176, 3:177f
 - ERK3/ERK4, 3:177f, 3:178
 - ERK5, 3:177f, 3:178
 - ERK7/ERK8, 3:177f, 3:178
 - fibroblast growth factor receptors (FGFRs), 2:304, 2:305f

- focal adhesion kinase (FAK) family, 2:265–266
- focal adhesion signaling, 2:321*f*, 2:322
- glycoprotein hormone receptors, 4:388–390
- G protein-coupled receptor kinases (GRKs), 2:485*t*
- G_q family proteins, 2:497–498, 2:498*f*
- integrin signaling, 1:439*f*, 2:611
- mitochondrial permeability transition pore, 3:165–166
- mTOR complexes, 3:195–196
- myocardial ischemia, 3:111
- neurotrophins, 3:241–243, 3:242*f*
- phylogenetic relationships, 3:177*f*
- Raf kinases, 4:247–248, 4:247*f*
- schematic diagram, 3:293*f*
- serine–threonine phosphatases, 4:222–223, 4:222*f*
- signaling cascades, 3:177*f*
- somatostatin (sst) receptors, 4:265
- submembranous calcium signals, 3:294
- tachykinins, 4:362–363
- toll-like receptors (TLRs), 4:398–399
- transforming growth factor beta (TGF-β), 1:620*f*
- focal adhesion kinase (FAK) family, 2:265–268
- activation and signaling mechanisms, 2:265–266, 2:267*f*
- cell migration, 1:439, 1:439*f*, 1:441, 1:441*f*
- cellular responses, 2:266
- characteristics, 2:265–266, 2:266*f*
- domain conservation, 2:266*t*, 2:267–268
- expression regulation, 2:265
- FERM domain, 2:265–266, 2:266*f*
- intracellular signaling pathway connections, 2:611
- invertebrates, 2:267–268
- localizations, 2:265–266, 2:266*f*
- matrix adhesions, 1:439, 1:439*f*
- signaling adhesion functions, 4:294–295
- submembrane plaque, 2:320–321, 2:320*f*
- glycogen synthase kinase-3 (GSK3), 2:434–438
- adipocyte development transcriptional control, 1:55, 1:55*f*
- Wnt pathway, 1:53–54, 1:53*f*
- diabetes, 2:437–438
- domain structures, 2:434, 2:435*f*
- inactivation consequences, 2:438
- insulin signal transduction, 1:198–199, 1:200*f*, 2:604–605, 2:604*f*
- knockout mice models, 2:438
- mitochondrial permeability transition pore, 3:165–166
- myocardial ischemia, 3:111
- neurogenesis, 1:199
- neurological diseases, 2:437
- physiological functions, 2:436*f*, 2:437
- regulatory mechanisms, 2:434–436, 2:435*f*, 2:436*f*
- signaling pathways, 1:564–565, 1:564*f*
- substrates, 2:436–437, 2:436*t*
- Wnt/β-catenin signaling pathway, 1:263*f*, 2:434–435, 2:436*f*
- G protein-coupled receptor kinases (GRKs), 2:483–491
- carboxyl tail, 4:363–364, 4:363*f*
- disease connections, 2:487
- G_q family proteins, 2:498
- olfactory receptors, 3:349, 3:349*f*
- phosphorylation, 2:483–486, 2:485*t*
- regulation mechanisms, 2:486
- regulatory functions, 2:483
- signaling mechanisms, 1:565, 2:345*f*
- structural characteristics, 2:484*f*, 2:486–487
- subfamilies, 2:483
- G-protein-signaling regulators (RGSs), 2:493
- histidine kinases, 1:629, 1:630*f*
- IκB kinase (IKK), 1:313, 3:303, 3:304*f*, 3:305, 4:456, 4:458*f*
- inositol phosphate kinases, 2:580–582
- inositol ring nomenclature, 2:580, 2:581*f*
- metabolic interrelationships and nomenclature, 2:581–582, 2:581*f*
- phosphate-transfer reactions, 2:580, 2:581*f*
- phosphotransferase system (PTS), 2:581–582
- regulatory input, 2:581*f*, 2:582
- integrin-linked kinase (ILK), 2:320–321, 2:320*f*, 2:611
- interleukin-1 receptor-associated kinase (IRAK) family protein kinases, 4:398, 4:399*f*, 4:400*f*
- janus kinase (JAK)
- angiotensin receptors, 1:113
- c-*fes* proto-oncogene (Fes), 1:452
- chemokine receptors, 1:480–482
- hematopoietin receptors, 2:526, 2:528*f*
- interferon receptors, 2:620–621, 2:620*f*, 2:622*f*
- leptin expression, 2:260*f*
- neurogenesis, 1:198
- signaling pathways, 1:616, 1:617*f*, 1:618*f*
- c-jun N terminal kinase (JNK) pathway
- angiotensin receptors, 1:113
- c-*fes* proto-oncogene (Fes), 1:453
- functional role, 3:177–178
- glycogen synthase kinase-3 (GSK3), 2:435*f*, 2:436*t*
- hemidesmosome function, 1:642–643
- hydrogen peroxide (H₂O₂), 3:173, 3:173*f*
- insulin resistance, 2:606–607
- insulin signal transduction, 2:605–606
- integrin signaling, 2:611
- mitochondrion-dependent apoptosis, 3:174
- muscarinic acetylcholine receptors, 3:216
- neurogenesis, 1:199
- nitric oxide (NO), 3:173, 3:173*f*
- p75 neurotrophin receptor (p75NTR), 3:245, 3:245*f*
- phylogenetic relationships, 3:177*f*
- signaling cascades, 3:177*f*
- toll-like receptors (TLRs), 4:398–399
- kinase homology domain (KHD), 3:226*f*, 3:227
- mitogen-activated protein kinase (MAPK), 3:176–180
- adipocyte development, 1:55, 1:55*f*
- adrenergic receptors, 1:59
- bile salts, 1:169
- calcium ion (Ca²⁺) signaling system, 2:719–720, 2:720*f*, 3:293*f*, 3:294
- chemokine receptors, 1:480–482, 1:482*f*
- dopamine receptors, 2:172
- extracellular signal-regulated kinase (ERK)
- angiotensin receptors, 1:113
- calcium ion (Ca²⁺) signaling system, 2:719–720, 2:720*f*
- c-*fes* proto-oncogene (Fes), 1:453
- chemokine receptors, 1:480–482, 1:482*f*
- cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:564, 1:564*f*
- ERK1/ERK2, 3:176, 3:177*f*
- ERK3/ERK4, 3:177*f*, 3:178
- ERK5, 3:177*f*, 3:178
- ERK7/ERK8, 3:177*f*, 3:178
- fibroblast growth factor receptors (FGFRs), 2:304, 2:305*f*
- focal adhesion kinase (FAK) family, 2:265–266
- focal adhesion signaling, 2:321*f*, 2:322
- glycoprotein hormone receptors, 4:388–390
- G protein-coupled receptor kinases (GRKs), 2:485*t*
- G_q family proteins, 2:497–498, 2:498*f*
- integrin signaling, 2:611
- mTOR complexes, 3:195–196
- myocardial ischemia, 3:111
- neurotrophins, 3:241–243, 3:242*f*
- phylogenetic relationships, 3:177*f*
- Raf kinases, 4:247–248, 4:247*f*
- schematic diagram, 3:293*f*
- serine–threonine phosphatases, 4:222–223, 4:222*f*
- signaling cascades, 3:177*f*
- somatostatin (sst) receptors, 4:265
- submembranous calcium signals, 3:294
- tachykinins, 4:362–363
- toll-like receptors (TLRs), 4:398–399
- transforming growth factor beta (TGF-β), 1:620*f*
- fibroblast growth factor receptors (FGFRs), 2:304, 2:305*f*
- functional roles, 3:176
- G_{βγ} subunit functions, 2:377
- gene expression regulation, 3:179
- glycogen synthase kinase-3 (GSK3), 2:436

kinases (*Continued*)

G-protein-signaling regulators (RGSs), 2:493

G_q family proteins, 2:497–498, 2:498f

hematopoietin receptors, 2:528f

hemidesmosome function, 1:642–643

hydrogen peroxide (H₂O₂), 3:173

insulin-like growth factor-1 (IGF1), 1:198–199, 1:200f

insulin receptors (IRs), 2:602–603

insulin resistance mechanisms, 1:651, 1:651f

insulin signal regulation, 2:606

insulin signal transduction, 2:605–606

c-jun N terminal kinase (JNK) pathway

angiotensin receptors, 1:113

c-fes proto-oncogene (Fes), 1:453

functional role, 3:177–178

glycogen synthase kinase-3 (GSK3), 2:435f, 2:436t

hydrogen peroxide (H₂O₂), 3:173, 3:173f

insulin resistance, 2:606–607

insulin signal transduction, 2:605–606

integrin signaling, 2:611

mitochondrion-dependent apoptosis, 3:174

muscarinic acetylcholine receptors, 3:216

neurogenesis, 1:199

nitric oxide (NO), 3:173, 3:173f

p75 neurotrophin receptor (p75NTR), 3:245, 3:245f

phylogenetic relationships, 3:177f

signaling cascades, 3:177f

toll-like receptors (TLRs), 4:398–399

mTOR complexes, 1:408, 3:195–196

multiple opioid receptors, 3:357

muscarinic acetylcholine receptors, 3:216

neurogenesis, 1:198–199, 1:200f

neurokinin-1 receptor (NK1R), 4:362–363

neurotrophins, 3:241–243, 3:242f

nitric oxide (NO), 3:173

nuclear signal transmission specificity and fidelity, 3:178–179

p38 kinase

angiotensin receptors, 1:113

chemokine receptors, 1:480–482

functional role, 3:178

G protein-coupled receptor kinases (GRKs), 2:485t

G-protein-signaling regulators (RGSs), 2:493

muscarinic acetylcholine receptors, 3:216

neurokinin-1 receptor (NK1R), 4:362–363

phylogenetic relationships, 3:177f

signaling pathways, 1:618f, 1:620f, 3:177f

pheromone receptors, 3:441–442, 3:443f

phylogenetic relationships, 3:177f

platelet-activating factor receptor (PAFR), 3:536, 3:536f

platelet-derived growth factor (PDGF) receptors, 3:541f

RAGE (receptor for advanced glycation end products), 2:407–409

somatostatin (sst) receptors, 4:265

steroid/thyroid hormone receptors, 4:310f

T-cell receptors (TCRs), 4:374

toll-like receptors (TLRs), 4:398–399

vascular endothelial growth factor receptors (VEGFRs), 4:512

myosin light chain kinase (MLCK)

calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinases (CaMKs), 1:305–306, 1:305f

cell migration, 1:440–441, 1:441f

EF-hand proteins, 1:319t, 1:321

focal adhesions, 2:322, 2:322f

G₁₂/G₁₃ family functions, 2:345f

somatostatin (sst) receptors, 4:265

nuclear factor kappa B (NF-κB)-inducing kinase (NIK), 4:456, 4:458f

p21-associated kinase (PAK), 1:440–441, 1:441f, 2:611, 3:441–442, 4:113

p38 kinase

angiotensin receptors, 1:113

chemokine receptors, 1:480–482

functional role, 3:178

G protein-coupled receptor kinases (GRKs), 2:485t

G-protein-signaling regulators (RGSs), 2:493

muscarinic acetylcholine receptors, 3:216

neurokinin-1 receptor (NK1R), 4:362–363

phylogenetic relationships, 3:177f

signaling pathways, 1:618f, 1:620f, 3:177f

partitioning defective (PAR)/atypical protein-kinase (aPKC) complex, 1:262–263

PERK (PKR-like ER kinase), 2:524–525, 2:524f, 4:491

phosphatidylinositol 3-kinase (PI3K) pathway

angiotensin receptors, 1:113

B-cell antigen receptors (BCRs), 1:162–163, 1:163f

cellular functions, 3:461

c-fes proto-oncogene (Fes), 1:454

chemoattractant signaling, 2:236–238, 2:237f

cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:563–564

diabetes treatments, 1:657

DNA damage checkpoints proteins, 1:411t

focal adhesion kinase (FAK) family, 2:265–266

G_{βγ} subunit functions, 1:480–482, 1:482f, 2:377, 4:265

glycogen synthase kinase-3 (GSK3), 2:435f

hematopoietin receptors, 2:526, 2:528f

hemidesmosome function, 1:642–643

insulin receptors (IRs), 1:198–199, 1:200f, 2:602–603

insulin resistance mechanisms, 1:651, 1:651f

insulin signal regulation, 2:606

insulin signal transduction, 2:604

integrin signaling, 2:611

liver regeneration, 1:193–195

macroautophagy regulation, 1:149

mTOR complexes, 1:408, 3:195–196, 3:197

neurokinin-1 receptor (NK1R), 4:362–363

neurotransmitter transporters, 3:239

neurotrophins, 3:241–243, 3:242f

phagocytosis regulation, 3:439–440

platelet-activating factor receptor (PAFR), 3:536, 3:536f

platelet-derived growth factor (PDGF) receptors

chemotaxis, 3:542

mediated cellular responses, 3:541, 3:541f

signaling pathways, 3:539–540, 3:540f

small guanosine phosphatases (GTPases), 4:247f, 4:248

sodium–potassium (Na⁺–K⁺)-ATPase pump, 3:687

somatostatin (sst) receptors, 4:265

transforming growth factor beta (TGF-β), 1:620f

vacuole transport pathways, 4:504–506

zinc-finger domains, 4:578

phosphatidylinositol-4,5-bisphosphate (PIP₂), 3:448f

phosphatidylinositol 4-kinase (PI4K) pathway

cellular functions, 3:459–460

subfamilies, 3:459

phosphatidylinositol phosphate kinases (PIP kinases)

domain structures, 3:460, 3:460f

type I PIP kinases, 3:460–461, 3:460f

type II PIP kinases, 3:460f, 3:461

type III PIP kinases, 3:460f, 3:461

phosphoinositide 3-kinase-like kinase (PIKK) family, 1:412–413, 1:413f

phosphoinositide-dependent protein kinases, 3:463–466

functional role, 3:463

kinase S6 kinase (S6K), 2:604f, 2:605, 3:464f, 3:465

mTOR (rapamycin), 3:464f, 3:465

phosphoinositide 3-kinase (PI 3-kinase)

lipid signaling pathway, 3:448f, 3:463, 3:464f

phosphoinositide-dependent kinase 1 (PDK1), 3:464–465, 3:464f, 4:377–378

phosphoinositide-dependent kinase 2 (PDK2), 3:464f, 3:465

- protein kinase B (PKB), 3:463–464
 research summary, 3:466
 tyrosine kinases (Tec family), 3:465–466
 phosphorylase kinase (PhK), 1:305f, 1:306
- protein kinase A (PKA)
 adipocyte development, 1:55
 arachidonic acid regulated Ca^{2+} (ARC)
 channels, 1:129
 binding specificity and affinity,
 1:577–578
 carbohydrate-responsive element
 binding protein (ChREBP),
 1:375–376, 1:376f
 characteristics and functional role,
 1:575–577
 cyclic adenosine monophosphate
 (cAMP) receptors (cARs),
 1:563–564, 1:564f
 feed-forward control, 1:578–579, 1:578f
 glycogen synthase kinase-3 (GSK3),
 2:435f
 G protein-coupled receptor kinases
 (GRKs), 2:485t
 insulin-secreting cells (β -cells), 2:597f
 liver PK (LPK) *in vitro* regulation, 3:719
 negative feedback, 1:578–579, 1:578f
 neurogenesis, 1:199
 Perilipin 1 (Plin1), 1:47–48, 1:48f
 phosphofructokinase-2/fructose
 biphosphatase-2 (PFK-2/FBPase-2)
 activation, 3:457–458
 PKA-mediated lipolysis, 1:49–51, 1:50f,
 4:303–304
 plasma membrane calcium (Ca^{+})-
 ATPase pump, 3:525–526
 structural domains, 1:574–575, 1:575f,
 1:576f
 substrate specificity, 1:578
- protein kinase B (PKB)
 cyclic adenosine monophosphate
 (cAMP) receptors (cARs),
 1:563–564, 1:564f
 functional role, 3:463–464
 glycogen synthase kinase-3 (GSK3),
 2:435f
 insulin signal transduction, 2:604
 mTOR complexes, 1:408, 3:195–196
 neurokinin-1 receptor (NK1R),
 4:362–363
 neurotransmitter transporters, 3:239
 phosphatidylinositol-3,4,5-
 trisphosphate (PtdIns-3,4,5- P_3),
 3:450
- protein kinase C (PKC), 3:637–640
 arachidonic acid regulated Ca^{2+} (ARC)
 channels, 1:129
 background information, 3:637
 bile salts, 1:169
 calcium ion (Ca^{2+}) signaling system,
 2:719–720, 2:720f
 calcium oscillation decoding, 1:325
 calcium-sensing receptor (CaSR), 1:328,
 2:583, 2:584f
 chemokine receptors, 1:480–482, 1:482f
 diabetes, 1:656
 domain structures, 3:637, 3:638f
 downregulation, 3:639
 $\text{G}_{12}/\text{G}_{13}$ family functions, 2:345f
 $\text{G}_{\beta\gamma}$ subunit functions, 1:480–482,
 1:482f, 4:265
 gene expression regulation, 1:311
 glycogen synthase kinase-3 (GSK3),
 2:435f
 G protein-coupled receptor kinases
 (GRKs), 2:485t
 G_q family proteins, 2:497–498, 2:498f
 historical research, 3:637
 insulin resistance, 1:650
 insulin-secreting cells (β -cells), 2:597f
 insulin signal transduction, 2:604
 isozyme families, 3:637, 3:638f
 mitochondrial permeability transition
 pore, 3:165–166
 neurogenesis, 1:199
 neurokinin-1 receptor (NK1R),
 4:362–363
 neurotransmitter transporters, 3:239
 neurotrophins, 3:241–243, 3:242f
 nuclear calcium signal, 1:310–311
 phospholipase D (PLD), 3:472–473,
 3:474–475, 3:474f
 phosphorylation, 3:637–638, 3:638f
 plasma membrane calcium (Ca^{+})-
 ATPase pump, 3:525–526
 platelet-activating factor receptor
 (PAFR), 3:536, 3:536f
 platelet-derived growth factor (PDGF)
 receptors, 3:541f
 signaling pathways, 3:639–640
 somatostatin (sst) receptors, 4:265
 store-operated calcium (SOC) channels,
 3:359–360, 3:360f
 subcellular localization, 3:639
 T-cell receptors (TCRs), 4:374,
 4:377–378
 translocation cycle, 3:638–639, 3:639f
 vascular endothelial growth factor
 receptors (VEGFRs), 4:512
- protein kinase G (PKG)
 arachidonic acid regulated Ca^{2+} (ARC)
 channels, 1:129
 binding specificity and affinity,
 1:577–578
 characteristics and functional role, 1:577
 feed-forward control, 1:578–579, 1:578f
 mitochondrial permeability transition
 pore, 3:165–166
 negative feedback, 1:578–579, 1:578f
 structural domains, 1:574–575, 1:575f,
 1:576f
 substrate specificity, 1:578
- protein-tyrosine kinase-linked receptors
 (PTKRs), 2:583, 2:584f
 Raf kinases, 2:528f, 4:247–248, 4:247f
 receptor-interacting protein kinase (RIPK),
 1:393–394, 1:394f
 Reperfusion Injury Salvage Kinase (RISK)
 group, 3:165–166
 Rho-kinase (ROCK), 1:439f, 1:440–441,
 1:441f, 1:623–624, 2:237f, 2:238,
 4:113, 4:113t
 ryanodine receptors (RyRs), 4:197
- S6 kinases, 3:196f, 3:197, 3:197f,
 3:241–243, 3:242f
 sphingosine kinase, 4:283f
 Src (sarcoma) family kinases, 4:293–297
 angiotensin receptors, 1:113
 arrestins, 2:488t, 2:491
 cadherin signaling, 1:264, 1:264f
 calcium ion (Ca^{2+}) signaling system,
 2:719–720, 2:720f
 cell migration, 1:439, 1:439f, 1:441,
 1:441f
 c-fes proto-oncogene (Fes), 1:451, 1:452f
 characteristics, 4:293
 chemokine receptors, 1:480–482, 1:482f
 focal adhesion kinase (FAK) family,
 2:265–266, 2:267f, 2:611,
 4:294–295
 focal adhesions, 2:321
 functional role, 4:294
 group III 5-phosphatases, 3:462
 guanine-nucleotide-exchange factors
 (GEFs), 4:245, 4:246f
 hematopoietin receptors, 2:526, 2:528f
 historical research, 4:293–294
 immune responses, 4:294–295
 invertebrate Src kinases, 4:297
 neurokinin-1 receptor (NK1R),
 4:362–363
 phosphorylation, 4:295–297, 4:296f
 platelet-derived growth factor (PDGF)
 receptors
 chemotaxis, 3:542
 mediated cellular responses, 3:540,
 3:541f
 signaling pathways, 3:539–540,
 3:540f
 protein palmitoylation, 2:738,
 4:296–297, 4:296f
 protein prenylation, 2:737t
 protein tyrosine kinase (PTK), 1:162,
 1:163f, 3:470
 redundant functions, 4:294
 regulation mechanisms, 4:295–297,
 4:296f
 research summary, 4:297
 SH2/SH3 domains, 1:451–452, 1:452f,
 4:294, 4:295–297, 4:296f
 signaling adhesion functions, 4:294–295
 steroid/thyroid hormone receptors,
 4:310f
- Stt1/STN8, 1:514, 1:514f
 stress-activated protein kinase (SAPK),
 3:177–178, 3:177f
 Stt7/STN7, 1:513, 1:513f, 1:514, 1:514f
 Syc family kinases, 2:611
 thylakoid protein kinase network, 1:514,
 1:514f
 tyrosine kinase receptors (Trks)
 glycogen synthase kinase-3 (GSK3),
 2:435f
 messenger RNA (mRNA), 2:602–603
 neurotrophins
 cross-talk, 3:246
 functional role, 3:241
 p75 neurotrophin receptor (p75NTR),
 3:246

kinases (*Continued*)

- signaling pathways, 3:241–243, 3:242f
- phospholipase D (PLD), 3:474, 3:474f
- signaling pathways, 3:539–540, 3:540f
- T-cell receptor (TCR) signaling, 4:376
- vascular endothelial growth factor (VEGF)
 - biological development, 4:512–513
 - evolutionary development, 4:509
 - lymphangiogenesis, 4:513
 - neovascularization, 4:513
 - signal transduction, 4:511, 4:512f
 - structural characteristics, 4:509
 - transmembrane domains, 4:510, 4:510f
- tyrosine kinases
 - calcium ion (Ca^{2+}) signaling system, 2:719–720, 2:720f
 - c-fes* proto-oncogene (Fes), 1:451, 1:452f
 - chemokine receptors, 1:480–482, 1:482f
 - dual-specificity tyrosine phosphorylation-regulated kinase 2 (DYRK2), 1:312, 1:313f
 - fibroblast growth factor receptors (FGFRs), 2:303–304, 2:304f, 2:305f
 - focal adhesions, 2:321
 - neurotransmitter transporters, 3:239
 - phospholipase C (PLC), 3:470
 - proline-rich tyrosine kinase, 2:345f
 - proline-rich tyrosine kinase 2 (PYK2)
 - activation and signaling mechanisms, 2:266–267, 2:267f
 - cellular responses, 2:267
 - characteristics, 2:266–267, 2:266f
 - domain conservation, 2:266t
 - protein–tyrosine kinase-linked receptors (PTKRs), 2:583, 2:584f
 - protein tyrosine kinase (PTK), 1:162, 1:163f, 3:470
 - protein tyrosine phosphatase PTB1B, 2:260f
 - protein tyrosine phosphatase (PTP), 3:411–412, 3:411f, 4:294–295, 4:296f
 - somatostatin (sst) receptors, 4:265
 - Tec family, 1:264, 2:620–621, 3:465–466, 3:539–540, 3:540f
 - ubiquitin–proteasome degradation pathway, 3:623t
- See also* Akt (serine–threonine active kinase protein)
- kindlins, 2:610
- kinesin, 2:679–687
 - characteristics, 2:672
 - cytoskeletal walking model
 - limitations, 1:632–633
 - regulation mechanisms, 1:633–634
 - stepping schemes, 1:632, 1:633f
 - dynein–dynactin complex, 2:175, 4:208, 4:208f
 - functional role, 2:679
 - kinesin superfamily proteins (KIFs)
 - binding consensus sequences, 2:681f
 - cargoes
 - binding regulation, 2:687
 - DENN/MADD adapter protein, 2:687
 - recognition and binding, 2:685f, 2:686–687
 - schematic diagram, 2:684f
 - specificity and redundancy, 2:686
 - transport direction regulation, 2:685f, 2:687
 - C-kinesin
 - Kinesin-14A family, 2:681f, 2:682f, 2:683t, 2:686
 - Kinesin-14B family, 2:681f, 2:682f, 2:683t, 2:686
 - domain structures, 2:680
 - family members, 2:683t
 - functional role, 2:680f, 2:683t
 - intracellular trafficking, 2:679–681, 2:680f
 - KIF1 family, 2:680f, 2:681f, 2:682f, 2:683t, 2:684–685, 2:684f
 - KIF3 family, 2:680f, 2:681f, 2:682f, 2:683t, 2:684f, 2:685–686
 - KIF5 family
 - cargo organelles, 2:684f, 2:685f
 - functional role, 2:681–684, 2:683t
 - intracellular trafficking, 2:680f
 - phylogenetic relationships, 2:682f
 - properties, 2:683t
 - research background, 2:679–681
 - schematic diagram, 2:684f
 - structural characteristics, 2:680f, 2:681f, 2:683t
 - KIF13 family, 2:681f, 2:682f, 2:683t, 2:684f, 2:685, 2:685f
 - KIF14 family, 2:683t
 - KIF17 family, 2:681f, 2:682f, 2:684f, 2:685f, 2:686
 - KIFC1 family, 2:681f, 2:682f, 2:683t, 2:686
 - KIFC2/KIFC3 family, 2:680f, 2:681f, 2:682f, 2:683t, 2:686
 - Kinesin-13 family, 2:683t, 2:686
 - M-kinesin, 2:683t, 2:686
 - N-kinesin
 - Kinesin-1 family, 2:681–684, 2:681f, 2:682f, 2:683t, 2:684f, 2:685f
 - Kinesin-2 family, 2:680f, 2:681f, 2:682f, 2:683t, 2:684f, 2:685–686
 - Kinesin-3 family, 2:680f, 2:681f, 2:682f, 2:683t, 2:684–685, 2:684f, 2:685f
 - Kinesin-4 family, 2:680f, 2:681f, 2:684f, 2:686
 - Kinesin-6 family, 2:681f, 2:686
 - Kinesin-7 family, 2:681f, 2:686
 - Kinesin-8 family, 2:681f, 2:686
 - Kinesin-11 family, 2:681f, 2:686
 - phylogenetic relationships, 2:682f
 - properties, 2:683t
 - research background, 2:679–681
 - schematic diagram, 2:684f
 - structural characteristics, 2:680f, 2:681f, 2:683t
 - Unc104/KIF1 family, 2:681f, 2:682f, 2:683t, 2:684–685
- microtubule disassembly enzymes, 2:672–678
- background information, 2:672
- domain structures, 2:673f
- family classifications, 2:672, 2:673f
- Kinesin-8 family
 - biophysical properties, 2:674–675, 2:675f
 - domain structures, 2:673f
 - localizations, 2:676–677, 2:676f
 - physiological functions, 2:675–678, 2:677f
 - regulation mechanisms, 2:672, 2:673f
- Kinesin-13 family
 - biophysical properties, 2:674–675, 2:675f
 - domain structures, 2:673f
 - localizations, 2:676–677, 2:676f
 - physiological functions, 2:675–678, 2:677f
 - regulation mechanisms, 2:672, 2:673f
- Kip3, 2:673f, 2:674
- MCAK/Kif2C
 - biophysical properties, 2:674–675, 2:675f
 - domain structures, 2:673f
 - localizations, 2:676–677, 2:676f
 - physiological functions, 2:675–678, 2:677f
 - regulation mechanisms, 2:672, 2:673f
- mitotic kinesin-like protein-1 (MKLP-1), 1:623–624
- secretory pathways, 4:208, 4:208f
- kinetic isotope effects (KIEs), 2:688–693
- activation energy, 2:689–690
- characteristics, 2:688
- chemical bonds, 2:688
- enzyme catalysis
 - characteristics, 2:690–691
 - experimental considerations, 2:691
 - glucose oxidase catalysis, 2:691–693, 2:691f, 2:692f, 2:692t
- harmonic oscillations, 2:688
- physical origin, 2:688
- predicted value deviations, 2:690
- reaction mechanisms, 2:689
- transition state theory, 2:689–690, 2:689f, 2:690f
- wave–particle dualism, 2:688–689, 2:689t
- zero-point energies, 2:690, 2:690f
- kinetin, 1:627, 1:628f
- kinetochores
 - microtubule disassembly enzymes, 2:675–678, 2:676f, 2:677f
 - mitotic checkpoint, 1:417–422
 - anaphase-promoting complex (APC), 1:419–420, 1:419f, 1:420f
 - aurora B/Ipl1 protein, 1:418
 - centromere protein E (CENP-E), 1:418–419, 1:419f
 - checkpoint pathways, 1:419
 - checkpoint protein roles, 1:421–422
 - direct inhibitor model, 1:420–421, 1:420f
 - functional role, 1:417
 - kinetochore monitoring, 1:417

- microtubule attachment monitoring, 1:417–418
 - molecular components, 1:417
 - sequestration model, 1:419–420, 1:419f
 - tension monitoring, 1:417–418
 - unattached kinetochores, 1:422
 - kinin (kallikrein–kininogen–kinin system), 1:240, 1:241f
 - kininogen (kallikrein–kininogen–kinin system), 1:240, 1:241f
 - kirromycin, 4:124f
 - kissing bugs, 2:593–594
 - Kit ligand, 1:452
 - kiwi fruit, 4:533t
 - Klebsiella*, 2:43, 2:208
 - Klebsiella pneumoniae*
 - Cho phylogenetic tree, 3:331f
 - lectins, 2:701–703
 - uvrC* phylogenetic tree, 3:331f
 - Klenow fragments
 - domain structures, 2:87–88, 2:88f
 - 5' exonuclease, 2:90–91, 2:91f
 - O-helix, 2:88–89, 2:88f
 - phosphoryl transfer reactions, 2:89, 2:89f
 - polymerization kinetics, 2:97t
 - reaction mechanisms, 2:89, 2:89f
 - reaction specificity
 - base pairing, 2:89–90
 - sugar specificity, 2:90
 - research background, 2:87
 - structural domains, 2:87
 - substrate binding, 2:88–89, 2:88f
 - 3'–5' exonuclease, 2:90
 - Klinefeleter syndrome, 3:23
 - Kluyveromyces lactis*, 4:350
 - Knallgas bacteria, 3:31–32
 - Knobloch syndrome type I, 3:657t
 - knockout mice
 - bradykinin receptors, 1:241
 - bullous pemphigoid antigen 1 (BPAG1), 2:629, 2:629t
 - calbindin-D_{9k}, 1:288
 - calbindin-D_{28k}, 1:287
 - gamma-aminobutyric acid (GABA) receptors, 2:348–349
 - glycerolipid receptors, 2:762–764
 - glycogen synthase kinase-3 (GSK3), 2:438
 - G_s family phenotypes, 2:521t
 - P2Y receptors, 3:377–378
 - plasma membrane calcium (Ca⁺)-ATPase pump, 3:528–529
 - plectin, 2:627–628, 2:628t
 - sphingolipid receptors, 2:762
 - Krabbe disease, 4:291t
 - Krebs cycle
 - anaplerotic pathways, 1:106f
 - electron flow pathways, 4:93f
 - research background, 4:437–438, 4:439f
 - thiamine diphosphate (ThDP), 1:206, 1:207f
 - Krebs, Hans
 - citric acid cycle, 4:437–438, 4:438f
 - photograph, 4:437f
 - research background, 4:436
 - urea cycle research, 4:436–437
 - Kruppel-associated box-associated protein-1 (KAP-1), 1:451
 - Ku70/80 protein, 3:269–271, 3:270f, 3:272f, 3:273t
 - Kupffer cells, 1:190–191, 1:191f
 - Ku protein, 3:266–267, 3:267f, 4:31–32
 - kuru disease, 3:574–575, 3:630
 - kynurenine-3-hydrolase (K3H), 1:175f
 - kynurenine formamidase (KFase), 1:175f
 - Kyprides–Onzonis–Woese (KOW) motif, 4:169–170
- L**
- L1CAM cell adhesion molecule, 4:272
 - LacI/lacO system
 - chromosomal domains, 2:751, 2:751f
 - forced domain repositioning, 2:751–752, 2:752f
 - future research outlook, 2:752–754, 2:753f
 - inner nuclear membrane (INM), 2:751–752, 2:752f
 - lac* operon, 2:694–700
 - catabolite repressor protein (CRP)
 - functional role, 2:698–699
 - historical research, 2:695
 - schematic diagram, 2:695f
 - structural characteristics, 2:697f, 2:699
 - characteristics, 2:695–696
 - complex organisms, 2:699
 - historical research, 2:694–695
 - lactose repressor protein (LacI)
 - functional role, 2:696–697
 - looped structures, 2:698, 2:698f
 - pseudo-operator sequences, 2:698, 2:698f
 - structural characteristics, 2:697–698, 2:697f
 - live cell imaging
 - chromosomal domains, 2:751, 2:751f
 - forced domain repositioning, 2:751–752, 2:752f
 - future research outlook, 2:752–754, 2:753f
 - inner nuclear membrane (INM), 2:751–752, 2:752f
 - research errors, 2:699
 - schematic diagram, 2:695f
 - transcription-coupled repair (TCR), 4:402
 - lacrimal glands, 1:124t
 - lactate
 - anaplerotic pathways, 1:106f
 - cerebral metabolism, 1:369f, 1:371–373
 - fermentation reactions, 2:204–206, 2:206f, 4:75f
 - gluconeogenesis, 2:381, 2:382f, 2:385f
 - glycolysis
 - adenosine triphosphate (ATP) production, 2:443
 - anaerobic glycolysis, 2:444–445, 2:445f
 - glycolytic pathway, 2:445f
 - oxidation effects, 2:445–446, 2:446f
 - schematic diagram, 2:444f
 - lactate dehydrogenase (LDH)
 - cardiac metabolism, 3:65–66, 3:66f, 3:68f, 3:70f
 - dark fermentation, 4:74f
 - dehydrogenation mechanisms, 1:174–175, 1:176f
 - structural characteristics, 1:176f
 - tricarboxylic acid cycle, 4:439f
 - lactic acid, 2:204–206, 2:205t
 - lactic acidosis, 1:211f, 2:446–447
 - Lactobacillus lactis*, 2:205t
 - Lactobacillus leichmannii*, 1:158
 - Lactococcus lactis*
 - Cho phylogenetic tree, 3:331f
 - extra-chromosomally-primed (EP) retrotransposons, 3:285f
 - H⁺ potential generation, 2:208–209
 - multidrug resistance (MDR) membrane proteins, 3:214
 - type III DNA restriction and modification enzymes, 2:137t
 - uvrC* phylogenetic tree, 3:331f
 - lactoperoxidase (LPO), 3:407–409, 3:408f
 - lactose
 - allolactose, 2:696–697, 2:696f
 - chemical structure, 2:696f
 - lac* operon
 - catabolite repressor protein (CRP)
 - functional role, 2:698–699
 - historical research, 2:695
 - schematic diagram, 2:695f
 - structural characteristics, 2:697f, 2:699
 - characteristics, 2:695–696
 - complex organisms, 2:699
 - historical research, 2:694–695
 - lactose repressor protein (LacI)
 - functional role, 2:696–697
 - looped structures, 2:698, 2:698f
 - pseudo-operator sequences, 2:698, 2:698f
 - structural characteristics, 2:697–698, 2:697f
 - live cell imaging
 - chromosomal domains, 2:751, 2:751f
 - forced domain repositioning, 2:751–752, 2:752f
 - future research outlook, 2:752–754, 2:753f
 - inner nuclear membrane (INM), 2:751–752, 2:752f
 - research errors, 2:699
 - schematic diagram, 2:695f
 - transcription-coupled repair (TCR), 4:402
 - lectins, 2:701–703
 - lactose permease (LacY)
 - crystal structure, 2:389f
 - functional role, 2:388–389
 - phosphotransferase system (PTS), 2:387–388, 2:388f
 - transport functions, 2:696
 - α-lactyl-ThDP, 1:206–207, 1:208f
 - lacZ reporter gene, 4:460
 - Lafora disease, 2:426, 2:427t, 3:651–652
 - Laforin, 3:650f, 3:651–652, 3:651t

- lamellipodia
 actin organization
 contraction, 1:29–31, 1:30f
 protrusion, 1:28–29, 1:29f
 group II 5-phosphatases, 3:462
 type I PIP kinases, 3:460–461
- laminin (LM)
 cell–matrix interactions, 1:426, 1:427t
 integrins, 1:427t, 1:438–440, 1:439f
 laminin-332, 1:641t, 1:642
 matrix metalloproteinases (MMPs), 3:91t, 3:96
- lamins, 2:632t, 2:633f, 2:634, 3:296, 3:297f, 3:310, 3:311f
- lamivudine triphosphate, 2:550, 2:550f
- lamprey, 1:600, 1:600f
- lanosterol
 chemical structure, 1:251f
 cholesterol biosynthetic pathway, 1:517f
- lanreotide, 4:263t, 4:264
- last common universal ancestor (LUCA), 4:144, 4:146f
- lateral hypothalamic area (LHA), 3:78, 3:78f
- latrunculins, 1:16–17, 1:16t
- layilin, 2:320
- LC8 family, 2:185–186, 2:185t, 2:186f, 2:187
- Lck genes/Lck proteins, 4:294
- LDG motifs, 4:50–53
- leadered Shine–Dalgarno (SD) sequence, 4:425–426, 4:426f
- leaderless Shine–Dalgarno (SD) sequence, 4:425–426, 4:426f
- lead (Pb²⁺)-adenosine triphosphatase (ATPase), 4:334
- Leber's hereditary optic neuropathy (LHON), 1:211f, 1:214, 4:90–91
- Leber's hereditary optic neuropathy (LHON) plus dystonia (LDYT), 1:211f, 4:90–91
- lecithin:retinol acyltransferase (LRAT), 4:525–526, 4:527f
- lectins, 2:701–705
 applications, 2:704–705
 characteristics and occurrence, 2:701, 2:702t
 functional role, 2:703–704, 2:704t
 mannose-binding lectins, 2:202, 2:202f, 2:702f, 2:702t, 2:703–704, 2:704t
 molecular properties, 2:701–703, 2:703f
 N-linked glycans processing, 3:259–260, 3:259t
- siglecs, 4:229–231
 binding sites, 4:231
 cell-type specific expression, 4:229
 cytosolic tails, 4:231
 functional roles, 4:230, 4:231
 genomic organization, 4:229–230
 historical research, 4:229
 immunoreceptor tyrosine-based inhibition motifs (ITIMs), 4:231
 N-acetylneuraminic acid (Neu5Ac)/N-glycolylneuraminic acid (Neu5Gc), 4:230–231
 phylogenetic relationships, 4:229–230
 siglec-1, 2:702t, 2:704t
 structural characteristics, 4:229, 4:230f
 subgroups, 4:229
 structural characteristics, 2:702f
- leeks, 4:533t
- Legionella pneumophila*, 4:333f
- legumains, 1:589
- Leigh disease (subacute necrotizing encephalomyelopathy), 1:209, 1:215
- leimodin, 1:21f, 1:25
- Leishmania*
 glycosylphosphatidylinositol (GPI)-anchored proteins, 2:476, 2:478
 multidrug resistance (MDR) membrane proteins, 3:214
Leishmania donovani, 2:158t, 4:350
- LEM domain proteins, 3:314, 3:315f
- lemons, 4:533t
- Lens esculenta*, 1:89
- lentil lectin, 2:703f
- leptin, 2:256–261
 adiposity and satiety signal integration, 3:79f
 appetite regulation, 4:303
 brown adipose tissue (BAT), 2:258–259
 characteristics and functional role, 2:256
 deficiency-related diseases, 2:256
 energy homeostasis, 3:77f, 3:78f
 energy intake–energy expenditure relationship, 2:257–258, 2:257f
 expression regulation, 2:259, 2:259f
 historical research, 2:256, 2:257f, 3:76
 homeostatic regulation, 2:257–258, 2:257f, 2:258f
 hypothalamic malonyl-CoA, 4:57, 4:58f
 leptin resistance, 2:260–261, 2:260f
 obesity syndrome, 2:256
 peroxisome proliferator-activated receptor gamma (PPAR γ), 3:422
 starvation effects, 4:303
 uncoupling proteins (UCPs), 2:258–259
- leptonema, 3:19, 3:19f, 3:22f
- Leptothrix* sp., 1:491t
- lettuce, 4:533t
- leucine
 classifications, 1:91
 degradative pathways, 1:244, 1:245f
 elastin, 2:192
 leucine–aspartic acid–valine (LDV) sequence, 1:427t, 2:608
 leucine–glutamate cycle, 1:248
 leucine-responsive regulatory protein (Lrp), 3:367, 3:368t
 leucine zipper domains
 carbohydrate-responsive element binding protein (ChREBP), 1:374–375, 1:375f
 CREB (cyclic adenosine monophosphate (cAMP) responsive binding protein), 1:313
 DNA sequences, 2:151f, 2:152
 G₁₂/G₁₃ family functions, 2:345f
 hematopoiesis, 1:188–189
 mitochondrial DNA (mtDNA), 3:134t
 nutritional signals, 1:246, 1:246f
 phage libraries, 3:434t
 tricarboxylic acid cycle, 4:439f
- Leuconostoc*, 2:205t
- leucopterin, 3:667f
- leucyl-tRNA synthetase, 4:443
- leukemia inhibitory factor (LIF), 1:198
- leukocyte adhesion deficiency (LAD) syndrome II, 4:351
- leukocytes
 aquaporin water channels, 1:124t
 immune system response, 1:493
 leukocyte FcRs, 2:568, 2:569f, 2:571
 long-chain acyl-CoA synthetase (LCAS), 2:292–293
- leukotrienes, 3:581–584
 B2 receptors, 1:242f
 biological activity, 3:584
 conversion pathways, 3:583, 3:583f
 5-lipoxygenase, 3:583–584, 3:583f
 5-lipoxygenase activating protein (FLAP), 3:584
 G_q family proteins, 2:496–497, 2:497t
 leukotriene A₄ (LTA₄), 3:583, 3:583f
 leukotriene A₄ (LTA₄) hydrolase, 3:584
 leukotriene C₄ (LTC₄) synthase, 3:584
 leukotriene receptors, 2:189
 metabolic reactions, 3:584
 platelet-activating factor receptor (PAFR), 3:536, 3:536f
 receptors, 3:584
- Levinthal paradox, 3:626
- Levitra, 1:569
- levofloxacin, 2:167, 2:167f
- Lewy body disease, 1:473, 4:475–476
- LexA regulatory protein, 2:706–711
 bacteriophages, 2:710
 gene expression regulation, 2:706
 SOS response
 bacterial LexA regulon diversity, 2:710
 cell-cycle checkpoints, 2:710
 conformational changes, 2:708–709, 2:709f
 control mechanisms, 2:710
 DNA damage repair, 2:707f, 2:709–710
 functional role, 2:706
 inactivation induction, 2:708, 2:708f
 mutation effects, 2:710–711
 plasma-encoded genes, 2:710
 regulon characteristics, 2:706–707, 2:710
 repression mechanisms, 2:707–708, 2:707f
 response triggers, 2:708
 self-cleavage mechanisms, 2:708–709, 2:709f
 signaling pathways, 2:708, 2:708f
 transcription regulation, 4:477, 4:478f
 structural characteristics, 2:707f, 2:708f
 two-hybrid protein–protein interaction systems, 4:460
- ligand-gated ion channels (LGICs)
 allosteric proteins, 1:74
 calcium/calmodulin-dependent protein kinase II (CaMKII), 1:297–298, 1:298f
 glycine receptors (GlyRs), 2:419–424
 allosteric modulation
 alcohol, 2:423
 anesthetics, 2:423
 background information, 2:423–424
 cannabinoids, 2:423
 glutamate, 2:423

- ivermectin, 2:423
- neurosteroids, 2:423
- posttranslational modifications, 2:422*f*, 2:423–424
- zinc (Zn), 2:422*f*, 2:423
- characteristics, 2:419
- disease connections
 - animal mutants, 2:424
 - human disorders, 2:424
- functional role
 - central nervous system (CNS), 2:419
 - excitatory function, 2:420
 - general discussion, 2:419
 - long-term synapse plasticity, 2:420
 - mixed GABA and glycinergic synapses, 2:420, 2:420*f*
 - motor control, 2:419, 2:420*f*
 - sensory processing, 2:419, 2:420*f*
 - synaptic versus extrasynaptic inhibition, 2:419–420, 2:420*f*
- glycine receptor (GlyR) channels
 - heterologous expression, 2:421
 - isoforms, 2:421
 - purification and isolation processes, 2:420–421
 - spatial and developmental expression, 2:421
- pharmacological properties
 - agonistic amino acids, 2:423
 - allosteric modulation, 2:423–424
 - channel block antagonism, 2:423
 - strychnine antagonism, 2:421–423
- phylogenetic relationships, 2:422*f*
- principal properties, 2:419, 2:420*f*
- structural characteristics
 - β -subunit formation, 2:420*f*, 2:421
 - domain structures, 2:421, 2:422*f*
 - ion channel function, 2:421, 2:422*f*
 - molecular model, 2:421
 - postsynaptic receptor clustering, 2:420*f*, 2:421
- serotonin receptors, 4:226
- ligand-operated membrane channels, 2:722–725
- ligase ribozymes, 4:144, 4:145*f*
- light-emitting compounds
 - See bioluminescence; chemiluminescence
- light harvesting, 2:726–728
 - higher plant light-harvesting antenna
 - basic concepts, 2:726–727
 - pigment components, 2:727
 - protein components
 - LHC I organization, 2:727, 3:503–505, 3:505*f*
 - LHC II organization, 2:726–727, 2:727*f*, 3:505*f*
 - Lhca and Lhcb proteins, 2:726, 3:503–505, 3:505*f*
 - structural characteristics, 2:726
 - light-harvesting chlorophyll *a/b*-binding protein complex II (LHCII), 3:508, 3:510*f*
 - photoprotection mechanisms, 2:727–728
 - photosystem I (PSI) complex, 3:503–505, 3:505*f*
 - photosystem II (PSII) complex, 3:512
 - regulation functions, 2:728
 - system heterogeneity, 2:726
- light sticks, 1:462, 1:463*f*
- lignin, 4:75*f*
- limb-girdle muscular dystrophy, 2:471*t*, 2:474–475
- limes, 4:533*t*
- Limulus polyphemus*, 2:702*t*
- lin-4* locus, 3:98–99
- linoleic acid, 2:288, 2:288*f*
- linoleoyl-CoA, 2:282–283, 2:283*f*
- lipases, 2:729–732
 - characteristics, 1:1
 - functional roles, 1:1
 - gastic lipase, 2:356
 - hydrolysis mechanisms, 1:1, 1:2*f*
 - lipolysis regulation
 - cofactors, 1:3
 - interfaces
 - physiochemical importance, 1:2–3, 1:2*f*
 - structural adaptations, 1:3, 1:3*f*
 - regulatory domains, 1:3, 1:3*f*
- lipoprotein lipase, 3:418, 3:422*f*
- phospholipases
 - adrenergic receptors, 1:59
 - biosynthetic relationships, 3:582*f*
 - calcium-sensing receptor (CaSR), 1:328
 - phospholipase A (PLA), 1:242*f*, 1:423, 2:236–237
 - phospholipase C (PLC), 3:467–471
 - catalytic mechanisms, 3:468–469
 - characteristics and functional role, 3:467
 - domain structures, 3:468, 3:469*f*
 - dopamine receptors, 2:172
 - EF-hand proteins, 1:319*t*
 - enzymatic pathways, 1:659, 1:660*f*
 - G₁₂/G₁₃ family functions, 2:345*f*
 - G_{βγ} subunit functions, 1:480–482, 1:482*f*, 4:265
 - glutamate receptors, 2:714–716, 2:715*t*
 - G-protein-coupled receptors (GPCRs), 4:225*f*
 - G_q family proteins, 2:496, 2:497–498, 2:498*f*
 - intercellular calcium ion (Ca²⁺) waves, 2:617
 - IP₃ receptors (IP₃R), 2:654*f*
 - isozyme families, 3:468, 3:469–471, 3:469*f*
 - muscarinic acetylcholine receptors, 3:215–216
 - neurokinin-1 receptor (NK1R), 4:362–363
 - phosphatidylinositol (PtdIns), 3:448–449
 - phospholipase C- β (PLC β), 2:496, 2:497–498, 2:498*f*, 3:469–470, 3:469*f*
 - phospholipase C- δ (PLC δ), 1:279*f*, 3:469*f*, 3:470
 - phospholipase C- ϵ (PLC ϵ), 3:469*f*, 3:470
- phospholipase C- γ 1 (PLC γ 1), 2:305*f*, 3:241–243, 3:242*f*, 3:450, 3:540, 3:540*f*, 3:541*f*, 3:542
- phospholipase C- γ (PLC γ), 3:469*f*, 3:470
- phospholipase C- η (PLC η), 3:469*f*, 3:471
- phospholipase C- ζ (PLC ζ), 3:469*f*, 3:470–471
- physiological functions, 2:654*f*, 3:467–468, 3:468*f*
- platelet-activating factor receptor (PAFR), 3:536, 3:536*f*
- second messenger pathways, 1:43*f*, 1:279*f*
- serotonin receptors, 4:225*f*, 4:226
- somatostatin (sst) receptors, 4:265
- taste receptors, 4:368
- T-cell receptors (TCRs), 4:374
- phospholipase D (PLD), 3:472–477
 - actin cytoskeleton rearrangement, 3:475–476
 - B2 receptors, 1:242*f*
 - bacterial PLD, 3:473
 - cancer treatments, 3:476–477
 - cellular functions, 3:473–477
 - Drosophila melanogaster*, 3:473–474
 - endocytosis, 3:476
 - enzymatic pathways, 1:659, 1:660*f*
 - epidermal growth factor (EGF) receptors, 3:476
 - exocytosis, 3:476
 - functional role, 3:472
 - Golgi apparatus, 3:476
 - mammals, 3:474–475
 - mitochondrial localization signal (mitoPLD), 3:477
 - mTOR (rapamycin), 3:476
 - phosphatidic acid (PA) regulation, 3:475–477
 - phosphatidylcholine hydrolysis, 3:472, 3:473*f*
 - phosphatidylinositol-3,4,5-trisphosphate (PtdIns-3,4,5-P₃), 3:449
 - photoreceptors, 3:473–474
 - plants, 3:474
 - research summary, 3:477
 - signal transduction, 3:476–477
 - structural domains, 3:472–473
 - vesicle transport, 3:476
 - yeast PLD (*spo14*), 3:473
 - zebrafish, 3:473
- substrate specificity, 1:1, 1:2*f*
- lipid-linked oligosaccharides (LLOs), 2:457, 2:458*f*, 2:470, 2:470*f*
- LIPID MAPS Consortium, 4:286
- lipids
 - detergent-resistant membranes (DRMs)
 - GPI-anchored proteins, 2:743
 - lipid-phase behaviors, 2:742–743
- digestion and absorption
 - cholesterol, 2:358
 - phospholipids, 2:358
 - triglyceride (TG), 2:358
- dolichol phosphate (Dol-P), 2:457, 2:458*f*

lipids (*Continued*)

exosomes and microvesicles (EMVs), 2:262
 fatty acyl-coenzyme A (CoA) synthetases, 2:291f
 functional role, 2:745
 glycerolipids, 2:412–418
 background information, 2:412
 functional role, 2:414–415
 phospholipid remodeling, 2:417f, 2:418
 structural characteristics
 diacylglycerol (DAG), 2:412–414, 2:414f
 ether linked glycerolipids, 2:413, 2:414f
 general discussion, 2:412–414
 general structural features, 2:413f
 sn-glycerol-3-phosphate, 2:412–414, 2:413f
 lyso-phospholipids, 2:413–414, 2:415f
 phosphatidic acid (PtdOH), 2:412–414, 2:413f
 triacylglycerol (TAG), 2:412–414, 2:414f
 synthesis process
 anionic glycerophospholipids, 2:415–416, 2:416f
 bioenergetics, 2:415
 ether lipids, 2:417–418
 neutral zwitterionic glycerolipids, 2:416–417, 2:416f
 phosphatidic acid synthesis, 2:415, 2:416f
 uncharged glycerolipids, 2:416–417, 2:416f
 hereditary spastic paraplegias (HSPs), 4:270–271
 lipid bilayer structure, 2:733–735
 bilayer dimensions, 2:733
 chemical structure, 2:733, 2:734f
 form–function relationship, 2:734
 local aqueous environment impacts, 2:735
 membrane proteins, 2:733, 2:734f
 membrane transport, 3:49, 3:50f
 organization and formation, 2:733, 2:734f
 physical states
 basic principles, 2:734
 lateral diffusion influences, 2:735
 lipid conformation, 2:735
 molecular organization, 2:734–735
 molecular organization effects, 2:734–735
 transverse movements, 2:735
 lipid-modified proteins, 2:736–740
 detection methods, 2:738
 fatty acylation
 background information, 2:736–740
 myristoylation, 2:736, 2:737f, 2:737t
 palmitoylation, 2:736–737, 2:737f, 2:737t
 saturated fatty acids, 2:737
 unsaturated fatty acids, 2:737
 glycosylphosphatidyl inositol (GPI)
 anchors, 2:737f, 2:737t, 2:738

inhibitors, 2:740
 intracellular trafficking, 2:739
 membrane binding, 2:738
 membrane raft targeting, 2:739–740
 membrane targeting, 2:738
 prenylation, 2:737–738, 2:737t, 3:608–609, 3:608f
 reversible membrane binding, 2:738–739
 switch mechanisms, 2:738–739
 lipid-phase behaviors, 2:741–742
 lipid rafts, 2:741–744
 functional roles
 caveolin-1, 2:743–744
 general discussion, 2:743
 signaling mechanisms, 2:743
 GPI-anchored proteins, 2:742f, 2:743
 lipid-modified proteins, 2:739–740
 membrane raft targeting, 2:739–740
 physical properties, 2:741, 2:742f
 raft formation, 2:742–743
 targeted proteins, 2:741
 vitamin E actions, 4:545–546, 4:546f
 membrane-associated energy transduction, 3:28
 microRNA (miRNA) regulation, 4:192–193, 4:194t
 peroxisome proliferator-activated receptors (PPARs), 3:418
 phosphatidic acid phosphatases (PAPs), 1:662–663
 phospholipids
 B-cell antigen receptors (BCRs), 1:162–163, 1:163f
 chemical structure, 2:733, 2:734f
 cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:563–564
 digestion and absorption, 2:358
 ether-linked phospholipid biosynthesis, 3:416–417
 exosomes and microvesicles (EMVs), 2:262
 fatty acyl-coenzyme A (CoA) synthetases, 2:291f
 flippase-mediated transbilayer movement, 2:314, 2:315f
 inositol phospholipids, 4:216
 lipid rafts, 2:742, 2:742f
 lipogenic gene expression, 2:287f
 lysophospholipids
 glycerolipid receptors, 2:762–764
 intracellular receptors, 2:762, 2:764
 lysophosphatidic acid, 2:344t, 2:345f
 lysophosphatidic acid (LPA) receptors, 2:762–764, 2:763t
 lysophospholipid receptors, 2:762–764
 plasma membrane receptors, 2:762–764
 remodeling processes, 2:417f, 2:418
 sphingolipid receptors, 2:762
 structural characteristics, 2:413–414, 2:415f
 phosphatidic acid (PtdOH), 2:412–414, 2:413f, 2:415, 2:416f

phosphatidylcholine (PtdCho), 2:314, 2:358, 2:412, 2:413f, 2:414–415, 2:416f, 2:417f, 3:478, 3:479f
 phosphatidylethanolamine (PtdEtn), 2:412, 2:413f, 2:414–415, 2:416f
 phosphatidylglycerol (PtdGro), 2:412, 2:413f, 2:414–415, 2:416f
 phosphatidylinositol-4,5-bisphosphate (PIP₂)
 cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:563–564
 functional role, 2:745–746, 2:748
 ion channel activity, 2:746–747, 2:746f, 2:746t
 research summary, 2:748
 transport modulation, 2:746–747
 phosphatidylinositol (PtdIns)
 biosynthetic pathways, 2:416f
 cellular targets, 3:449, 3:449f
 characteristics, 3:447
 functional role, 2:414–415
 metabolic pathways, 3:448f
 phosphatidylinositol-3,4,5-trisphosphate (PIP₃), 3:447–451
 phosphatidylinositol-4,5-bisphosphate (PIP₂), 2:745–746, 3:447–451
 structural characteristics, 2:412, 2:413f
 phosphatidylserine (PtdSer), 2:412, 2:413f, 2:414–415, 2:416f
 phosphoinositides
 phosphatidylinositol-4,5-bisphosphate (PIP₂), 1:563–564, 2:745–746
 species diversity, 2:745–746
 phospholipid hydroperoxide
 glutathione peroxidase (PHGPx), 3:409, 3:410f
 phospholipid transfer protein (PLTP), 4:548–549
 platelet-activating factor (PAF), 3:533–534, 3:534f
 remodeling processes, 2:417f, 2:418
 vitamin E, 4:548–549
 photosynthetic metabolic pathways, 3:498f
 vitamin E actions, 4:545–546, 4:546f
 zinc-finger domains, 4:578
 lipins, 1:661f, 1:662, 3:420
 lipogenesis
 carbohydrate-responsive element binding protein (ChREBP), 1:376–377, 1:377f
 insulin metabolism, 2:603f
 insulin signal transduction, 2:604f
 lipoic acid, 3:307–309, 3:308f
 lipolysis
 insulin metabolism, 2:603f, 2:606–607
 insulin resistance, 2:606–607
 Perilipin 1 (Plin1), 1:48
 starvation, 4:298
 lipopolysaccharide, 4:287
 lipoproteins
 adiponectin, 3:74
 cholesterol biosynthetic pathway, 1:516, 1:517f
 lipoprotein lipase, 3:418, 3:422f

- matrix metalloproteinases (MMPs), 3:91*t*
 lipoxins, 2:188–189
 lipoxygenase
 biological function, 3:373–374
 biosynthetic pathways, 2:188–189
 lipoxygenase metabolite receptors
 characteristics, 2:189
 5-oxo-EET receptors, 2:189
 leukotriene receptors, 2:189
 lipoxin receptors, 2:189
Listeria innocua, 3:331*f*
Listeria monocytogenes, 3:331*f*
 lithium (Li)
 macroautophagy regulation, 1:150*t*
 NCLX protein superfamily, 3:125
 lithocholate, 1:168*f*
 lithotrophs, 1:488*t*
 LiveBench database, 3:615*t*
 liver
 acyl-coenzyme A synthetase
 long-chain acyl-CoA synthetase (LCAS),
 2:292–293
 medium-chain acyl-CoA synthetase
 (MCAS), 2:292
 very long-chain acyl-CoA synthetase
 (VLCAS), 2:293–294
 amino acid metabolism, 1:93
 aquaporin water channels, 1:124*t*
 ascorbic acid content, 4:533*t*
 carnitine biosynthesis, 1:385
 gluconeogenesis, 2:384, 4:305–306,
 4:306*f*
 glycogen storage diseases, 2:431*t*
 glycogen synthesis, 2:430*f*
 G_s family phenotypes, 2:521*t*
 healthy adult liver characteristics,
 1:190–191, 1:191*f*
 insulin metabolism, 2:603*f*
 liver glucokinase/glucokinase regulatory
 protein (GCKR), 2:547
 liver PK (LPK) *in vitro* regulation
 allosteric regulation, 3:719
 crystallographic studies, 3:719–721,
 3:720*f*
 gene expression regulation
 carbohydrate response element
 (CHORE), 3:721
 glucagon levels, 3:721
 insulin levels, 3:721
 pH effects, 3:719
 phosphorylation/dephosphorylation,
 3:719
 liver pyruvate kinase (LPK)
 carbohydrate-responsive element
 binding protein (ChREBP),
 1:374, 1:375*f*, 1:376–377, 1:376*f*,
 1:377*f*
 liver PK (LPK) *in vitro* regulation,
 3:719–721
 transcriptional activation, 1:374
 mammalian glucose transport, 2:392*t*
 megamitochondria, 2:375–376, 2:375*f*
 microRNA (miRNA)-based regulation,
 4:192–193, 4:194*t*
 peroxisome proliferator-activated receptor
 alpha (PPAR α), 3:420–421
 phosphofructokinase-2/fructose
 bisphosphatase-2 (PFK-2/FBPase-2)-
 activated phosphorylation,
 3:457–458, 3:458*f*
 proteinase-activated receptors (PARs),
 3:605*t*
 regeneration processes, 1:190
 basic concepts, 1:190
 chronic liver injury repairs, 1:193–195,
 1:194*f*
 embryogenesis, 1:192
 Hedgehog (Hh) pathway
 chronic liver injury repairs,
 1:193–195, 1:194*f*
 fetal development, 1:192, 1:192*f*
 post-partial hepatectomy (PH)
 reactivation, 1:193
 research summary, 1:195
 post-partial hepatectomy (PH)
 regeneration, 1:191–192
 research summary, 1:195
 schematic diagram, 1:192*f*
 starvation, 4:298, 4:305
 vitamin E processing, 4:548
 voltage-gated calcium channels, 4:561*t*
 liver diseases
 amyloid deposits, 1:101*t*
 cancer, 1:193–195
 mitochondrial permeability transition
 pore, 3:166
 primary biliary cirrhosis (PBC), 3:561–565
 anti-mitochondrial antibodies (AMAs),
 3:114, 3:115–116
 background information, 3:561
 clinical features, 3:563–564
 diagnosis, 3:564
 etiology
 associated autoimmune diseases, 3:562
 autoantibodies/antimitochondrial
 antibodies, 3:561–562
 mitochondrial antigens, 3:562
 laboratory tests, 3:564
 pathology
 developmental stages, 3:563
 stage III PBC, 3:563*f*
 stage II PBC, 3:563*f*
 stage I PBC, 3:563*f*
 stage IV PBC, 3:564*f*
 pathophysiology, 3:562–563
 prevalence, 3:561
 symptom treatments
 anemia, 3:564–565
 fatigue, 3:564
 osteoporosis, 3:564–565
 pruritus, 3:564
 vitamin malabsorption, 3:564
 underlying disease process treatments
 general discussion, 3:565
 liver transplantation, 3:565
 medical treatments, 3:565
 liver-X receptor (LXR), 2:287–288
 lizards, 4:516*f*
 LlaFI enzyme, 2:137*t*
 lobster, 4:533*t*
 localized autosomal hypotrichosis (LAH1),
 1:638*t*
Locusta migratoria, 3:19*f*
 locust diuretic, 4:516*t*
 long-chain acyl-CoA synthetase (LCAS),
 1:385*f*, 2:292–293
 long interspersed elements (LINEs), 3:284,
 3:285*f*, 3:287–288
 long terminal repeat (LTR) retrotransposons,
 3:284–286, 3:285*f*, 3:286*f*, 3:287*f*
 long-term potentiation, 1:302
 loss of heterozygosity (LOH), 2:49–50, 2:49*f*
Lotus tetragonolobus, 2:702*t*, 2:705
 Lou Gehrig's disease
 See amyotrophic lateral sclerosis (ALS)
 low barrier hydrogen bonds (LBHBs),
 2:756–759
 basic concepts, 2:756
 characteristics, 2:756–757
 enzymes
 Δ 5-3-ketosteroid isomerase, 2:759
 serine esterases, 2:759
 serine proteases, 2:758*f*
 physicochemical properties
 bond lengths, 2:757
 deuterium fractionation factors, 2:758
 spectroscopic properties, 2:757–758
 research background, 2:756
 strong hydrogen bonds
 bond lengths, 2:757
 deuterium fractionation factors, 2:758
 serine proteases, 2:758–759
 spectroscopic properties, 2:757–758
 structural characteristics, 2:756–757,
 2:757*f*
 very strong (single-well) hydrogen bonds
 bond lengths, 2:757
 deuterium fractionation factors, 2:758
 spectroscopic properties, 2:758
 structural characteristics, 2:757, 2:757*f*
 weak hydrogen bonds
 bond lengths, 2:757
 deuterium fractionation factors, 2:758
 spectroscopic properties, 2:757
 structural characteristics, 2:756, 2:757*f*
 low-density lipoproteins (LDLs)
 adiponectin, 3:74
 cholesterol biosynthetic pathway, 1:518*f*
 endocytosis
 adaptor proteins, 2:197–198, 2:198*f*
 clathrin-mediated endocytosis, 2:197,
 2:198*f*, 2:199
 peroxisome proliferator-activated receptor
 gamma (PPAR γ), 3:422–423
 secreted sphingomyelinase (sSMase),
 4:287
 statins, 1:518*f*, 1:519
 vitamin E, 4:548–549
 lowest observed adverse effect level (LOAEF),
 4:548
 low molecular weight phosphatase (LMW-
 PTP), 3:540*f*, 3:542, 3:543
 low-voltage-activated channels (LVA), 4:561*t*
loxP recombination site, 1:556–557, 1:557*f*,
 1:559
 L-type calcium channels, 1:379–381, 1:380*f*
 Lubbers' rapid scan, 4:276
 luciferases, 1:462–465, 1:464*f*, 2:290–291

- Luft's disease, 2:760–761
 clinical features, 2:760
 definition, 2:760
 muscle biopsy results
 biochemistry, 2:761
 genetic studies, 2:761
 morphology, 2:760–761, 2:760f
 polarographic studies, 2:761, 2:761f
 therapies, 2:760
 lungfish, 4:516t
 lungs
 aquaporin water channels, 1:124t
 mucin expression, 3:203
 proteinase-activated receptors (PARs), 3:605t
 short-chain acyl-CoA synthetase (SCAS), 2:292
 lutein, 2:178–179, 2:179f, 2:726–727
 luteinizing hormone receptor (LHR)
 dimerization/oligomerization, 4:388
 extracellular domains, 4:388
 functional role, 4:387
 genomic organization, 4:388
 G protein activation, 4:388–390
 naturally occurring mutations, 4:390
 structural organization, 4:389f
Lycaon pictus, 1:319t
 lycopen, 1:502f
Lycopersicon esculentum, 3:688
Lymnaea stagnalis
 glycine receptors (GlyRs), 2:421
 peptide amidation, 3:398
 lymphatic filariasis, 2:591
 lymphocytes, 1:190–191, 1:191f
 lymphocytic choriomeningitis virus (LCMV), 4:156–157
 lymphoid-enhancing factor/T-cell factor (Lef/Tcf), 1:53
 lymphoid–myeloid progenitor (LMP), 1:187–188, 1:188f
 lymphomas, 4:379
 Lynch syndrome, 2:61–64, 2:62t, 2:63f
 lyn genes/Lyn proteins, 1:162, 1:163f, 4:294
 lysine
 acetylation, 4:60
 acyl-coenzyme A synthetase, 2:291f
 advanced glycation end products (AGEs), 2:408f
 ascorbic acid, 4:531–532
 biotin enzymes, 1:222–223, 1:222f, 1:223f
 carnitine, 1:384
 classifications, 1:92t
 cytochrome c, 1:600–601
 glycation pathways, 2:406, 2:406f
 LexA regulatory protein, 2:708–709, 2:709f
 lysine 5,6-aminomutase, 1:158t
 lysine (K)-specific demethylase LSD1/KDM1 family, 4:61
 methylation, 4:60–61
 neurofilament triplet proteins (NFTPs), 3:233–235, 3:234f
 phage libraries, 3:434t
 riboswitches, 4:137
 vasopressin/oxytocin family, 4:516t
 lysine–aspartate–glutamate–leucine (KDEL) sequence, 4:206–207
 lysine–glutamine–alanine–glycine–aspartic acid–valine (KQAGDV) sequence, 1:427t
 lysine–lysine–X–X (KKXX) sequence, 4:206–207
 lysophospholipids
 lysophosphatidic acid, 2:344t, 2:345f
 lysophospholipid receptors, 2:762–764
 glycerolipid receptors
 intracellular receptors, 2:764
 plasma membrane receptors, 2:762–764
 lysophosphatidic acid (LPA) receptors, 2:762–764, 2:763t
 sphingolipid receptors
 intracellular receptors, 2:762
 plasma membrane receptors, 2:762
 remodeling processes, 2:417f, 2:418
 structural characteristics, 2:413–414, 2:415f
 lysosomes, 4:502–508
 characteristics, 3:617–619
 characteristics and functional role, 4:502–503, 4:503f, 4:504f
 disease connections, 4:507–508, 4:507t
 endocytic–lysosomal pathways
 functional role
 general discussion, 3:617–619
 lysosomal enzymes, 3:618–619
 major histocompatibility complex (MHC), 3:619
 substrate uptake, 3:618, 3:618f
 eukaryotic functions
 disease connections, 4:507–508, 4:507t
 lysosome-related organelles, 4:506–507
 pathogen-containing vacuoles, 4:507
 protozoan contractile vacuole complex, 4:506
 free oligosaccharide (fOS) generation, 2:330f
 glycogen synthesis, 2:430f
 historical research, 4:502–503
 lysosomal-associated membrane proteins (LAMPs), 1:483f, 4:502–503, 4:504f
 lysosomal storage diseases, 3:624
 secretory pathway, 4:203–209
 compartmental organization, 4:203, 4:204f
 microtubule-based motor proteins, 4:208, 4:208f
 targeted proteins
 characteristics, 4:203
 constitutive secretion, 4:208
 entry mechanisms, 4:203, 4:205f
 integral membrane proteins, 4:204, 4:205f
 protein exit, 4:207
 regulated secretion, 4:208
 signal recognition particles (SRPs), 4:203, 4:205f
 soluble proteins, 4:204, 4:205f
 transport mechanisms, 4:206f, 4:208
 transport mechanisms
 protein localization, 4:206–207, 4:206f
 retinoid X receptor (RXR), 4:207
 vesicular transport, 4:206, 4:206f
 lysozymes, 1:101t
 lysyl oxidase, 1:89, 2:192
 lysyl-tRNA synthetase, 4:441, 4:443
 lysyl tyrosyl quinone, 1:89
Lytechinus pictus, 3:538
Lytechinus variegatus, 2:266f
- ## M
- macroautophagy
 mammals
 autophagosome formation, 1:149, 1:150f
 basic concepts, 1:149–150, 1:150f
 phagophore maturation, 1:149, 1:150f
 physiological functions, 1:150
 regulatory mechanisms, 1:149, 1:150t
 yeasts, 4:505f, 4:506, 4:506f
 macrophage mannose receptor (MMR), 2:702t, 2:703–704, 2:704t
 macrophages
 vacuole/lysosome functions, 4:507
 vitamin D, 4:538–539
 macropinocytosis, 2:198f, 2:199
 macrovascular disease, 1:655
 macular corneal dystrophy, 2:471t, 2:475
 mad cow disease, 1:101t, 1:102, 3:630
 MAD family genes, 1:417, 1:419–420, 1:419f, 1:420–421, 1:420f
 See also mitotic checkpoint
 Mag Fura 2, 2:575t, 2:576
 magnetic circular dichroism (MCD), 2:661–662
 magnetic resonance imaging (MRI), 4:278
 Maillard reaction
 aging effects, 2:409, 2:410f
 basic concepts, 2:405
 characteristics, 2:407
 diabetes, 2:409–410, 2:410f
 disease effects, 2:410
 maize, 3:284f
 major facilitator superfamilies (MFS), 2:388
 major histocompatibility complex (MHC)
 antigen receptors, 4:376
 endocytic–lysosomal pathways, 3:619
 exosomes and microvesicles (EMVs), 2:262, 2:263f
 immunoproteasome, 3:621–622
 interferon- γ receptors, 2:621
 olfactory receptors, 3:346
 T-cell receptors (TCRs)
 MHC restrictions, 4:369–371
 peptide–MHC complex recognition, 4:371, 4:372f
 20S proteasome, 3:597
 ubiquitin–proteasome degradation pathway, 3:623t
 major late transcription factor (MLTF), 1:374
 major sperm protein (MSP), 3:1–5
 accessory proteins
 MSP fiber protein 1 (MFP1), 3:3t, 3:4, 3:4f
 MSP fiber protein 2 (MFP2), 3:3t, 3:4, 3:4f

- MSP fiber protein 3 (MFP3), 3:3*t*, 3:4–5, 3:4*f*
- MSP-organizing protein (MPOP), 3:3–5, 3:3*t*, 3:4*f*
- MSP polymerization-activating kinase (MPAK), 3:3*t*, 3:4, 3:4*f*
- protein phosphatase 2A (PP2A), 3:3*t*, 3:4*f*, 3:5
- biochemistry, 3:2
- in vitro* motility system, 3:3, 3:3*f*
- molecular biology, 3:2
- nematode sperm motility, 3:1, 3:2*f*
- signaling functions, 3:1–2
- structural characteristics, 3:2–3, 3:2*f*
- malaria
- aspartic proteases, 1:139
- mosquitoes, 2:591
- malate
- anaplerotic pathways, 1:105, 1:106*f*
- citric acid cycle, 4:93*f*, 4:438*f*
- gluconeogenesis, 2:382*f*, 2:386*f*
- malate dehydrogenase, 1:141–142, 1:142*f*, 4:439*t*
- tricarboxylic acid cycle, 4:439*f*
- MalE
- See maltose transport complex
- malectin, 3:259*t*
- MalF
- See maltose transport complex
- MalFGK₂
- See maltose transport complex
- MalG
- See maltose transport complex
- malic acid, 3:496
- malignant hyperthermia (MH), 4:197–198
- malignant tumor growth, 2:446
- MalK
- See maltose transport complex
- malnutrition, 4:299
- malonate, 1:143*t*
- malonyl-CoA
- body weight regulation, 4:56–58
- carnitine palmitoyltransferase I (CPT I), 2:258–259, 4:56, 4:58
- cellular fatty acid metabolism regulation, 4:56
- energy homeostasis, 4:56–57
- fatty acid synthase inhibitors, 4:57, 4:57*f*
- functional roles, 4:56
- hypothalamic malonyl-CoA
- functional role, 4:57, 4:57*f*
- regulation mechanisms, 4:57–58, 4:58*f*
- hypothalamus, 4:56–57
- nutrient signaling processes, 4:58
- carbohydrate-responsive element binding protein (ChREBP), 1:377*f*
- cardiac metabolism, 3:66*f*, 3:68*f*, 3:69*f*, 3:70*f*
- fatty acid biosynthesis
- carnitine-dependent cellular uptake and activation, 1:385*f*
- energy metabolism regulation, 2:286
- fatty acid synthesis pathway, 2:275–279, 2:276*f*
- lipogenic gene expression, 2:287*f*
- pyruvate dehydrogenase regulation, 2:285–286
- fatty acid oxidation, 1:541, 2:283, 3:71
- malonyl CoA decarboxylase (MCD), 2:279, 3:71
- malonyl-CoA-insensitive carnitine palmitoyltransferase, 1:385–386, 1:385*f*
- malonyl-CoA-sensitive carnitine palmitoyltransferase, 1:385–386, 1:385*f*
- maltodextrin/maltose transport (MalEFGK₂), 2:390, 2:390*f*
- maltose-binding protein (MalBP)
- ATP-binding cassette (ABC) transport systems, 2:389–390
- protein–ligand interaction affinity tags, 1:62
- maltose transport complex
- ATP-binding cassette (ABC) transport systems, 2:389–390, 2:390*f*
- phosphotransferase system (PTS), 2:387–388, 2:388*f*
- mammalian hepatic binding protein (HBP), 2:702*t*
- mammals
- autophagy, 1:149–154
- macroautophagy
- autophagosome formation, 1:149, 1:150*f*
- basic concepts, 1:149–150, 1:150*f*
- phagophore maturation, 1:149, 1:150*f*
- physiological functions, 1:150
- regulatory mechanisms, 1:149, 1:150*t*
- microautophagy, 1:150
- calpains, 1:354*f*
- carbohydrate-responsive element binding protein (ChREBP) distribution, 1:374
- color vision, 1:553*f*
- cytochrome *c*, 1:600, 1:600*f*
- DNA damage checkpoints proteins, 1:411*t*
- DNA mismatch repair (MMR), 2:66–71
- basic concepts, 2:66–68, 2:68*f*
- downstream steps
- characteristics, 2:69–70
- excision, 2:69–70
- resynthesis, 2:70
- mutation types, 2:67*f*
- MutL family proteins, 2:68–69, 2:69*f*
- MutS family proteins, 2:68, 2:69*f*
- research summary, 2:71
- strand discrimination, 2:70–71, 2:70*f*
- glucose transport, 2:391–394
- facilitative glucose transporters (GLUTs), 2:392*f*
- cerebral metabolism, 1:369–371, 1:370*f*
- characteristics, 2:394
- classifications, 2:392*t*
- GLUT1, 1:369–371, 1:370*f*, 2:392–393, 2:392*f*, 2:392*t*, 3:65–66, 3:66*f*, 3:68*f*, 3:69*f*, 3:70*f*, 4:532–533
- GLUT2, 1:376–377, 1:377*f*, 2:392*t*, 2:393
- GLUT3, 1:369–371, 1:370*f*, 2:392*t*, 2:393, 4:532–533
- GLUT4, 1:651–652, 1:651*f*, 2:392*t*, 2:393, 2:604–605, 2:604*f*, 3:65–66, 3:66*f*, 3:68*f*, 3:69*f*, 3:70*f*, 4:532–533
- GLUT5, 1:369–371, 1:370*f*, 2:392*t*, 2:393
- GLUT6, 2:392*t*, 2:393
- GLUT7, 2:392*t*, 2:393
- GLUT8, 2:392*t*, 2:393–394
- GLUT9, 2:392*t*, 2:394
- GLUT10, 2:392*t*, 2:394
- GLUT11, 2:392*t*, 2:394
- GLUT12, 2:392*t*, 2:394
- GLUT13, 2:392*t*, 2:394
- GLUT14, 2:392*t*, 2:394
- sodium-dependent glucose transporters (SGLTs), 2:391, 2:392*f*
- glycosylphosphatidylinositol (GPI)-anchored proteins, 2:477–478
- hexokinases, 2:543–544, 2:544*t*, 2:546
- mammalian low-*K_m* hexokinases
- activity regulation, 2:546
- glucose-6-phosphate (glucose-6-P)
- inhibition, 2:546
- subcellular localization, 2:546
- mammalian peroxidases, 3:407–409, 3:408*f*
- mitochondrial DNA (mtDNA), 2:127–128, 2:127*f*, 3:133*t*, 3:134*t*
- nicotinic acid adenine dinucleotide phosphate (NAADP)-induced calcium ion (Ca²⁺) release, 2:638*t*
- phospholipase D (PLD)
- activation mechanisms, 3:474–475, 3:474*f*
- alcohol effects, 3:475
- enzymatic functions, 3:474
- inhibitors, 3:475
- phosphatidic acid (PA)
- activation mechanisms, 3:475
- functional role, 3:475, 3:475*f*
- sensors, 3:475
- pterin distribution, 3:666
- pyruvate dehydrogenases (PDHs), 4:321–322, 4:326–327, 4:327*f*
- regulated intramembrane proteolysis (RIP)
- γ-secretase-based mediation
- amyloid precursor protein (APP), 4:51*f*, 4:53
- ErbB-4, 4:51*f*, 4:53
- Notch, 4:51*f*, 4:53
- rhomboid-based mediation, 4:51*f*
- site-2 protease (S2P)-based mediation
- activation transcription factor 6 (ATF6), 4:51*f*, 4:52
- cAMP responsive element-binding protein 3 (CREB3), 4:51*f*, 4:52
- sterol regulatory element binding protein (SREBP), 4:51–52, 4:51*f*
- vasopressin/oxytocin family, 4:516*f*, 4:516*t*
- mammary glands, 2:292
- mammary tumors, 2:305
- MAN1 domain proteins, 3:314, 3:315*f*
- mandibular acrodysplasia, 2:632*t*

- manganese (Mn)
 bacterial reductive reactions, 1:487t
 photosystem II (PSII) complex
 assembly processes, 3:521, 3:521f
 manganese calcium ($\text{Mn}_4\text{Ca}^{2+}$) clusters, 3:514
 research background, 3:522–523
 structural characteristics, 3:522–523
 protein phosphatase Mg^{2+} - or Mn^{2+} -dependent (PPM), 4:219
 superoxide dismutases (SODs), 4:353
- mannitol
 fermentation reactions, 2:204–206
 insect biochemistry, 2:590–591
 phosphotransferase system (PTS), 2:387
- mannose
 congenital disorders, 2:474
 free oligosaccharide (fOS) generation, 2:330f
 $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ glycan, 3:257, 3:258f, 3:259t
 glycosylphosphatidylinositol (GPI)-anchored proteins, 2:476, 2:477f
 high-mannose-bearing mannose-6-phosphate residues, 2:461f, 2:463–464
 $\text{Man}_5\text{GlcNAc}_2$ glycan, 2:329, 2:330, 2:330f
 O-mannose
 carbohydrate modifications, 2:454, 2:454f
 classifications, 2:453
 α -dystroglycan function, 2:454f, 2:456
 mannose-6-phosphate receptors, 2:482, 2:702t, 2:704t
 mannose-binding lectin A (MBL-A), 2:702f, 2:702t, 2:703–704, 2:704t
 α -mannosidase (Man2C1), 2:330, 2:330f, 2:331
 N-linked glycosylation, 2:458–460, 2:459f, 2:460f, 2:461f, 2:470, 2:470f, 2:471t, 4:204–205
 O-linked glycosylation, 2:466f
- mannosidases
 Man2A2/MAN2A2, 3:259
 α -mannosidase (Man2C1), 2:330
 N-linked glycans processing, 3:257–260
 class I α -mannosidases
 endoplasmic reticulum (ER) α -mannosidase I, 2:201, 3:258, 3:258f, 3:259t
 ER degradation enhancing α -mannosidase-like protein (EDEM), 2:201, 3:259
 Golgi α -mannosidase I, 3:258, 3:258f, 3:259t
 class II α -mannosidases, 3:259
 endo α -mannosidase, 2:201, 3:257, 3:258f
 $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ glycan, 3:257, 3:258f, 3:259t
 Golgi α -mannosidase II, 3:258f, 3:259, 3:259t, 4:504f
 $\text{Man}_5\text{GlcNAc}_2$ glycan, 3:257–259, 3:259t
 $\text{Man}_8\text{GlcNAc}_2$ glycan, 3:258f, 3:259t
 $\text{Man}_9\text{GlcNAc}_2$ glycan, 3:259t
 processing pathways
 basic concepts, 3:257–259
 schematic diagram, 3:258f
 vacuole transport pathways, 4:504f
- mannosylphosphoryldolichol (Man- P-Dol)
 flippase-mediated transbilayer movement, 2:314, 2:315f
 N-linked glycosylation, 2:315–316, 2:315f, 2:316t
 protein O- and C-mannosylation, 2:316
- maple syrup urine disease (MSUD)
 branched-chain α -ketoacid dehydrogenase complex (BCKDC), 1:244, 1:245f
 branched-chain amino acid (BCAA) regulation, 1:245–246
 branched-chain ketoacid-dehydrogenase (BCKA), 1:248
 thiamine deficiency, 1:209
- marinobufagenin, 3:686–687, 3:686f
- Marsipenaues japonicus*, 2:266f
- marsupials
 color vision, 1:553f
 vasopressin/oxytocin family, 4:516f, 4:516t
- MarvelD3 protein, 4:392, 4:394f
- maspardin, 4:269f, 4:270–272
- mass spectrometry
 ion-mobility spectrometry (IMS), 3:11, 3:14–15
 matrix-assisted laser-desorption ionization (MALDI), 4:286
 native protein complexes, 3:11–16
 collision-induced dissociation (CID), 3:12, 3:12f, 3:13–14, 3:14f
 complex structure, 3:13–14
 dynamic protein complexes, 3:13
 future research areas, 3:15
 gas-phase complex structure, 3:14–15
 GroEL (growth E locus, large gene, bacterial group I chaperonin), 3:14–15, 3:15f
 nano-electrospray ionization (nanoESI), 3:11–12, 3:12f
 protein complex composition, 3:12–13, 3:13f
 research methodologies, 3:11–12
 structural biology analyses, 3:11
 tandem (MS/MS) mass spectrometry, 3:12, 3:12f, 3:14f
 time-of-flight (ToF) mass spectrometry, 3:11–12
- mast cells, 1:188f, 1:189
- Mastigocladus laminosus*, 4:335–338, 4:337f, 4:338t, 4:339–340, 4:339t
- matrilysins, 3:91t, 3:92f, 3:93
- matrix-assisted laser-desorption ionization (MALDI), 4:286
- matrix Gla protein (MGP), 1:236, 4:551
- matrix metalloproteinases (MMPs), 3:90–97
 activation mechanisms
 chemical/physical agents, 3:95
 proteinases, 3:95, 3:95f
 zymogens (proMMPs), 3:94
 biological functions, 3:96
 characteristics, 1:543
 characteristics and functional role, 3:90
 classifications, 1:543t
 DNA polymorphisms, 3:96
 domain structures, 3:90, 3:92f, 3:93–94, 3:93f
 enamelysin, 3:91t, 3:92f, 3:93
 endogenous inhibitors
 α -2 macroglobulins, 3:96
 tissue inhibitors of metalloproteinases (TIMPs), 3:96
 epilysin, 3:91t, 3:92f, 3:93
 families/subfamilies
 characteristics, 3:90
 collagenases, 3:90–91, 3:91t, 3:92f
 gelatinases, 3:91t, 3:92–93, 3:92f
 matrilysins, 3:91t, 3:92f, 3:93
 membrane-type MMPs, 3:91t, 3:92f, 3:93
 stromelysins, 3:91t, 3:92f, 3:93
 functional roles, 1:543–544
 genetic disorders, 3:96
 historical research, 1:542
 inhibition mechanisms, 3:96
 metalloelastase, 3:91t, 3:92f, 3:93
 occurrences, 3:90
 peptide bond hydrolysis mechanisms, 3:94, 3:94f
 substrate specificity, 3:95–96
 three-dimensional (3D) cell migration and invasion, 1:443–444
 tissue expression, 3:92f
 tissue inhibitors of metalloproteinases (TIMPs), 1:443–444, 3:93f, 3:95f, 3:96
 vertebrate matrixins, 3:90, 3:91t
See also collagenase
- maturity-onset diabetes of the young (MODY), 1:652–653
- MaxiK channel, 4:559
- May–Hegglin anomaly, 3:223t
- MBD4
See methyl binding domain IV (MBD4); methyl-CpG-binding proteins
- MBOAT (membrane bound O-acyl transferase) family, 2:736–737
- Mcalp protein, 3:566
- [MCA] prion
See prions
- McArdle disease, 2:427t, 2:431t, 2:432
- McCune–Albright syndrome (MAS), 2:519
- mDia effector protein, 1:440–441, 1:441f
- Mdm2, 3:381–382, 4:468t, 4:470
- Mdtra/Mdtra2* genes, 1:86
- mealworms, 1:384
- measles virus, 4:156–157
- medakas, 1:552f
- medium-chain acyl CoA dehydrogenase deficiency, 4:495t
- medium-chain acyl-CoA synthetase (MCAS), 2:292
- Meesman’s corneal dystrophy, 2:670t
- megakaryocyte–erythroid progenitor (MEP), 1:187, 1:188f
- megakaryopoiesis, 1:187, 1:188f
- megamitochondria, 2:375–376, 2:375f
- meiosis, 3:17–23
 basic concepts, 3:17
 chromosome behavior, 2:557, 2:558f
 chromosome cycle, 3:18, 3:18f
 disease connections, 3:23
 DNA sequence divergence, 2:48–49

- gametogenesis, 3:20
- homologous recombination (HR),
2:557–561
- axial elements, 2:560, 2:560f
- basic concepts, 3:21–22
- crossover pathways, 2:560
- crossover versus noncrossover
chromosomes, 2:558, 2:559f, 3:22,
3:22f
- functional role, 2:557
- gene conversion, 2:558–559, 2:559f
- genetic interference, 2:560
- meiotic recombination checkpoint/
pachytene checkpoint, 2:560
- mitotic versus meiotic recombination,
2:561
- molecular model
- double Holliday junction (DHJ),
2:558, 2:559f, 3:22, 3:22f
- double-strand breaks (DSBs), 2:247,
2:557, 2:559f, 3:22f, 3:266
- initiation phase, 2:557, 2:559f
- invasion phase, 2:558, 2:559f
- resection phase, 2:557–558, 2:559f
- resolution phase, 2:558, 2:559f
- recombination control mechanisms,
3:21–22, 3:22f
- recombination nodules, 2:560, 2:560f
- single-end invasion (SEI), 3:22
- synaptonemal complex, 2:560, 2:560f,
3:20, 3:21f, 3:22f
- synthesis-dependent strand annealing
(SDSA), 2:558, 2:559f, 3:22
- ultrastructural level, 3:20–21
- meiosis II, 3:20
- post-prophase I, 3:19f, 3:20
- prophase I, 3:18–20, 3:19f, 3:22f
- recombination-independent mechanisms,
3:22–23
- sexual life cycle, 3:17
- structural characteristics
- loop–axis structure, 3:20, 3:21f
- synaptonemal complex, 2:560, 3:20,
3:21f, 3:22f
- ultrastructural level, 3:20–21
- Melander–Westheimer principle, 2:690
- melanin concentrating hormone (MCH),
3:78, 3:78f
- melanocortin system, 3:24–27
- adrenocortical steroidogenesis, 3:25–26
- adrenocorticotrophic hormone (ACTH),
3:24–25, 3:25t, 3:26f
- agouti receptor antagonist, 3:24–25, 3:25t,
3:26f
- agouti-related protein (AgRP)
- body weight regulation, 3:77, 3:77f,
3:78f
- energy homeostasis, 3:26
- functional role, 3:24–25, 3:25t, 3:26f
- insulin metabolism, 2:603f, 2:606
- leptin expression, 2:257–258
- drug discovery efforts
- synthetic ligands, 3:27
- therapeutic uses
- antiobesity drugs, 3:27
- cachexia, 3:27
- sexual dysfunction, 3:27
- energy homeostasis, 3:26
- exocrine gland function, 3:26–27
- functional role, 3:24, 3:25t, 3:26f
- ligands and receptors, 3:24–25, 3:25t,
3:26f
- melanocortin neurons, 3:77
- melanocortin receptors, 2:520, 3:24–25,
3:25t, 3:26f, 3:77, 3:77f
- melanocyte-stimulating hormone (MSH),
3:24–25
- pigmentation, 3:25, 3:25t, 3:26f
- proopiomelanocortin (POMC)
- energy homeostasis, 3:26, 3:77f
- functional role, 3:24–25, 3:25t, 3:26f
- ghrelin activity, 1:118
- insulin metabolism, 2:603f, 2:606
- leptin expression, 2:257–258, 2:260f
- melanocortin neurons, 3:77
- synthetic ligands, 3:27
- melanoma antigen gene (MAGE) family,
3:243, 3:244f
- melanoma chondroitin sulfate proteoglycan
(MCSP), 1:439–440
- melanotan-II (MT-II), 3:27
- melanotropins, 4:531–532
- melastatin, 4:416
- melibiose permease (MelB)
- functional role, 2:389
- phosphotransferase system (PTS),
2:387–388, 2:388f
- melons, 4:533t
- membrane-bound immunoglobulin (mlg),
1:161, 1:162f
- membrane fusion, 3:36–43
- AAA (ATPases associated with various
cellular activities) proteins, 1:5–6
- analytical methodologies
- conformational changes, 3:42
- functional studies, 3:42
- fusion protein structure, 3:42
- in vitro* reconstitution, 3:43
- protein–protein interactions, 3:42
- single-molecule analysis, 3:43
- single virion entry assays, 3:43
- basic concepts, 3:36, 3:37f
- cellular fusion proteins, 3:37, 3:38f
- functional role, 3:36
- hemifusion, 3:36, 3:37f, 3:38f, 3:39f
- research summary, 3:43
- soluble N-ethylmaleimide-sensitive factor
association protein receptor (SNARE),
3:37, 3:38f
- stalk–pore hypothesis, 3:36, 3:37f
- viral fusion proteins
- characteristics and functional role, 3:38
- class I fusion proteins, 3:39–41, 3:39f
- class II fusion proteins, 3:39f, 3:41
- class III fusion proteins, 3:40f, 3:41–42
- membrane proteins
- flippase proteins, 2:314–317
- aminophospholipid flippases,
2:314–315, 2:315f, 2:316t
- cytoplasmic leaflets, 2:314, 2:316t
- dolichyl pyrophosphate (Dol-P-P),
2:315–316, 2:315f, 2:316t
- flippase-mediated transbilayer
movement, 2:314, 2:315f
- floppases, 2:314–315, 2:315f
- glucosylphosphoryldolichol (Glc-P-
Dol), 2:315–316, 2:315f, 2:316t
- glycosylation, 2:315–316
- glycosyl phosphatidylinositol (GPI)
anchors
- bacterial cell wall complex, 2:316–317
- biosynthesis, 2:316–317
- glucosylceramide (Glc-Cer), 2:316,
2:316t
- lipid bilayer structure, 2:735
- mannosylphosphoryldolichol (Man- P-
Dol), 2:315–316, 2:315f, 2:316t
- membrane biogenesis-related bilayer
expansion, 2:314
- protein O- and C-mannosylation, 2:316
- scramblases, 2:314–315, 2:315f
- lipid bilayer structure, 2:733, 2:734f
- mucins, 3:207t
- membrane transport, 3:49–51
- active transport
- basic concepts, 3:50
- primary active transport, 3:51
- secondary active transport
- basic concepts, 3:51
- co-transport (symport), 3:51
- counter-transport (antiport), 3:51
- basic concepts, 3:49
- carrier-mediated transport, 3:50
- facilitated diffusion, 3:50
- phospholipid bilayer barrier, 3:49, 3:50f
- pore/channel (restricted) diffusion,
3:49–50
- simple diffusion, 3:49
- menadione, 3:727–728, 4:552f
- menaquinol, 2:508f
- menaquinone, 2:206, 2:508f, 3:34, 3:700,
3:702f, 3:727–728, 4:94f, 4:552f
- Menkes disease, 3:670–671
- mental retardation, adducted thumbs,
spasticity and aphasia (MASA)
syndrome, 4:272
- mercury-based bacterial reductive reactions,
1:487t
- MEROPS database, 3:88
- mesenchymal-back to epithelialtype cells
(MET), 1:259–260
- mesenchymal stem cells (MSCs), 1:52, 1:53f
- Mesorhizobium loti*, 3:331f
- mesotocin, 4:516t
- messenger RNA (mRNA)
- alternative splicing, 1:75–80, 1:81–86
- auxiliary sequence elements, 1:76–77,
1:78f
- background information, 1:75
- CD45 gene, 1:79–80, 1:79f
- Drosophila melanogaster*
- double-sex (dsx)* gene regulation,
1:85–86
- exon organization and splicing
patterns, 1:81–82, 1:82f
- fruitless (fru)* gene regulation, 1:85–86
- research summary, 1:86
- Sxl* autoregulation, 1:82–84, 1:85f

messenger RNA (mRNA) (*Continued*)
transformer (tra) gene regulation, 1:82, 1:85–86, 1:85f
 exonic splicing enhancers (ESEs), 1:76, 1:78f, 3:62, 3:62f
 exonic splicing silencers (ESSs), 3:62f
 exon/intron boundaries, 1:75, 1:76f, 1:397, 1:398f
 fine-tuning determinants, 1:78–79
 hnRNP protein family, 1:76–77, 1:78f, 1:79–80, 1:79f
 intronic splicing enhancers (ISEs), 1:76, 1:78f, 3:62, 3:62f
 intronic splicing silencers (ISSs), 3:62f
 non-*Drosophila* species, 1:86
 post-transcriptional modifications, 1:78
 pre-messenger RNAs (pre-mRNAs), 1:75, 1:76f, 3:60–61, 4:187–188
 regulation mechanisms, 1:77–78, 1:78f
 research summary, 1:86
 small nuclear RNP (snRNP) complexes, 1:75, 1:76f, 1:78f
 splicing patterns, 1:75–76, 1:77f
 splicing silencer sequences, 1:76–77, 1:78f
 bacterial decay, 3:52–58
 basic concepts, 3:52
 gene expression regulation, 3:56–57, 3:57f
 model comparisons, 3:54f, 3:56
 ribonucleases
 degradosome complex, 3:53
 3'-to-5' exoribonucleases, 3:53–55, 3:54f
 5'-to-3' exoribonucleases, 3:56
 3' exoribonucleases, 3:56
 oligonucleotides, 3:56
 phosphorolysis versus hydrolysis, 3:53–55, 3:55f
 poly(A) polymerase (PAP)-mediated decay, 3:54f, 3:55
 polynucleotide phosphorylase (PNPase), 3:53–55, 3:54f, 3:55f, 3:56–57, 3:57f
 ribonuclease E (RNase E), 3:52–53, 3:53f, 3:54f, 3:56–57, 3:57f
 ribonuclease G (RNase G), 3:53
 ribonuclease III (RNase III), 3:56–57, 3:57f
 ribonuclease J1 (RNase J1), 3:56
 ribonuclease J (RNase J), 3:55–56
 ribonuclease Y (RNase Y), 3:56
 small RNAs (sRNAs), 3:57, 3:57f, 4:249–250
 tmRNA system, 3:58
 translation processes, 3:57–58
 carbohydrate-responsive element binding protein (ChREBP) distribution, 1:374
 chloroplast-encoded PSII proteins, 3:508
 energy homeostasis, 3:77
 eukaryotic messenger RNA (mRNA), 4:432
 exosomes and microvesicles (EMVs), 2:262–263
 HAC1 gene expression, 4:489, 4:490f
 insulin receptors (IRs), 2:602–603
lacZ mRNA, 2:695f, 2:696–697, 2:699

leptin expression, 2:257–258
 liver regeneration, 1:193
 maturation processes, 3:188–193
 cleavage sites, 3:188–189
 exon definition model, 3:61–62, 3:61f
 gene expression regulation, 3:192
 polyadenylation, 3:60–61, 3:188
 protein complexes, 3:189–191, 3:189f
 regulation mechanisms, 3:191–192
 3'-end maturation, 3:60–61, 3:189–191, 3:189f, 3:190f
 microRNA (miRNA), 3:98
 mitochondrial genomes, 3:140–141, 3:142f
 mucins, 3:206–208
 Perilipin 1 (Plin1), 1:47–48
 platelet-activating factor receptor (PAFR), 3:534–535
 pre-messenger RNAs (pre-mRNAs), 3:188–193
 alternative splicing, 1:75, 1:76f, 3:60–61, 4:187–188
 maturation processes
 cleavage sites, 3:188–189
 exon definition model, 3:61–62, 3:61f
 gene expression regulation, 3:192
 polyadenylation, 3:60–61, 3:188
 protein complexes, 3:189–191, 3:189f
 regulation mechanisms, 3:191–192
 3'-end maturation, 3:60–61, 3:189–191, 3:189f, 3:190f
 RNA splicing
 alternative splicing, 3:60–61, 4:187–188
 ancillary RNA-processing signals, 3:62, 3:62f
 chemical mechanisms, 4:185, 4:186f
 disease connections, 4:189–190
 evolutionary implications, 4:189
 exon definition model, 3:61–62, 3:61f
 exonic splicing silencers (ESSs), 3:62f
 exons, 3:142–144, 3:143f, 4:185, 4:186–187, 4:186f, 4:187f
 5' and 3' splice sites, 3:59–60, 3:60f, 4:186, 4:186f, 4:187f, 4:188f
 genetic information transfer, 4:185, 4:186f
 historical research, 4:185
 intronic splicing silencers (ISSs), 3:62f
 introns, 3:142–144, 3:143f, 4:185, 4:186–187, 4:186f, 4:187f
 life cycle, 3:62–63
 mosaic mitochondrial genes, 3:142–144, 3:143f
 research summary, 4:190
 ribozymes, 4:188
 self-splicing reactions, 4:188, 4:188f
 spliceosomes, 3:62–63, 3:63f, 4:187
 splicing consensus sequences, 4:186, 4:187f
 trans-splicing reactions, 4:188–189, 4:189f
 proteomic studies, 3:326f
 retroviral genomes, 2:102, 2:102f
 ribosomes, 4:116, 4:122–123, 4:128, 4:129f

riboswitches, 4:136–141
 characteristics and functional role, 4:136–137
 functional role
 alternative splicing, 4:139
 conformational switching, 4:140–141, 4:141f
 as drug targets, 4:141
 gene regulation, 4:139, 4:140f
 gene transcription regulation, 4:139, 4:140f
 glmS riboswitch, 4:138f, 4:141
 kinetic versus thermodynamic control, 4:139–140
 small molecule recognition, 4:138–139
 ThiM gene, 4:139, 4:140f, 4:141f
 research background
 antecedent developments, 4:136
 genetic regulation, 4:136
 systematic evolution of ligands by exponential enrichment (SELEX), 4:136
 sequence elements, 4:137
 structures and classes
 aptamer domain structures, 4:137–138, 4:138f
 cognate ligands, 4:137
 domain structures, 4:137
 untranslated regions (UTRs), 4:137
 RNA editing, 4:154, 4:157f
 RNA polymerase I (Pol I), 2:366t, 2:367–368
 RNA polymerase II (Pol II), 2:365–368
 conservation considerations, 2:367–368
 C-terminal repeat domain (CTD)
 characteristics, 2:365
 initiation phase, 2:366, 3:63–64, 3:63f
 transcription cycle, 2:365, 3:63–64, 3:63f
 transcription termination, 2:367, 3:63–64, 3:63f
 expression mechanisms
 elongation phase, 2:366–367, 3:63–64, 3:63f
 initiation phase, 2:366, 3:63–64, 3:63f
 transcription cycle, 2:365, 3:63–64, 3:63f
 transcription termination, 2:367, 3:63–64, 3:63f
 mobile clamps, 2:365
 structural characteristics, 2:365, 2:366f, 2:366t
 RNA polymerase III (Pol III), 2:366t, 2:367–368
 RNA processing, 3:59–64
 ancillary RNA-processing signals, 3:62, 3:62f
 basic concepts, 3:59–60, 3:60f
 exon definition model, 3:61–62, 3:61f
 exonic splicing silencers (ESSs), 3:62f
 gene transcription regulation, 3:63–64, 3:63f
 intronic splicing silencers (ISSs), 3:62f
 life cycle, 3:62–63
 open reading frames (ORFs), 3:60–61

- research summary, 3:64
 RNA-binding proteins, 3:59–60, 3:60f, 3:62–63
 spliceosomes, 3:62–63, 3:63f
 RNA trafficking, 3:323
 thiamine biosynthesis, 1:203–204
 3'-end maturation
 basic concepts, 3:59–60, 3:60f
 cleavage sites, 3:188–189
 gene expression regulation, 3:192
 polyadenylation, 3:188
 protein complexes, 3:189–191, 3:189f, 3:190t
 regulation mechanisms, 3:191–192
 untranslated regions (UTRs), 3:188, 3:192
 transcription and translation processes, 4:418
 translation initiation, 4:425–431
 basic concepts, 4:425
 eukaryotes
 AUG triplets, 4:433–435
 80S initiation complex, 4:434f
 80S ribosome formation, 4:435
 eukaryotic initiation factors (eIFs), 4:433t, 4:434f, 4:435
 5' leader scanning, 4:433
 40S ribosomal subunits, 4:432
 43S preinitiation complexes, 4:433, 4:434f
 48S initiation complex, 4:435
 initiation codon selection, 4:433–435
 initiation mechanisms, 4:432
 internal ribosomal entry sites (IRESs), 4:435
 regulation mechanisms, 4:435
 ribosomal dissociation, 4:432
 60S ribosomal subunits, 4:432, 4:435
 40S ribosomal subunits, 4:432
 initiation complex formation, 4:426–427
 initiation factors
 initiation factor 1 (IF1), 4:427–428, 4:427f, 4:428f
 initiation factor 2 (IF2), 4:427f, 4:428–430
 initiation factor 3 (IF3), 4:427f, 4:428f, 4:429f, 4:430
 structure and functions, 4:427–431
 ribosomal dissociation, 4:432
 70S initiation complex, 4:130f, 4:426–427, 4:427f
 60S ribosomal subunits, 4:432
 30S initiation complex, 4:130f, 4:132f, 4:426–427, 4:427f, 4:429f
 translation initiation region (TIR), 4:425–426, 4:426f
 uncoupling proteins (UCPs), 4:484–485, 4:485f
 metabolic syndrome (syndrome X), 1:654
 metabolite channeling, 3:80–85
 action mechanisms, 3:80
 advantages, 3:80–81, 3:81f
 basic concepts, 3:80
 creatine kinase/phosphocreatine circuit, 3:81, 3:82f
 cytosolic creatine kinase, 3:81–82, 3:82f
 energy-transducing mitochondrial microcompartments, 3:82–84, 3:83f
 glycolytic multienzyme complexes, 3:81
 mitochondrial CK (MtCK) isoenzyme energy-transducing microcompartments, 3:82–84, 3:83f
 intramitochondrial inclusions, 3:84, 3:84f
 low-cellular-energy state, 3:84
 microcompartments, 3:82–84
 mitochondrial permeability transition pore, 3:84
 subcellular microcompartments, 3:80
 metabotropic glutamate (mGlu) receptors, 2:395–398
 characteristics, 2:716t
 fast excitatory synaptic transmission, 2:395
 functional role, 2:396–397, 2:714–716, 2:715t
 gene families, 2:395–396, 2:396f
 molecular structure, 2:717f
 neuromodulation, 2:395
 protein structure, 2:717f, 2:719
 research background, 2:395
 schematic diagram, 2:396f
 therapeutic potential, 2:397
 metachromatic leukodystrophy, 4:291t
 metalloelastase, 3:91t, 3:92f, 3:93
 metalloproteases, 3:86–89
 AAA (ATPases associated with various cellular activities) proteins, 1:3–5, 1:3t, 1:4f, 1:6
 characteristics, 3:86
 FtsH metalloprotease, 1:3t, 1:4f, 1:6
 functional classification, 3:86
 genetic origin, 3:88
 inhibitors
 catalytic chelators, 3:88f, 3:89
 general discussion, 3:88–89
 metal-bound water displacement, 3:88f, 3:89
 metal hydroxide complexes, 3:88f, 3:89
 ligand nature, 3:88
 physiological functions, 3:88
 regulated intramembrane proteolysis (RIP), 4:50–53
 structural classification
 catalytic sites, 3:87, 3:87f
 cocatalytic sites, 3:87–88
 metal sites, 3:86–87, 3:87f, 4:50–53
 research background, 3:86–88
 structural sites, 3:87
Metallospira, 3:34
 metaphase I, 3:19f, 3:20
 metazoans
 alternative oxidases (AOXs), 1:145–146
 conventional actin, 1:38–40
 septins, 4:217f, 4:218
 S phase initiation proteins, 1:402t
 unfolded protein response (UPR)
 ATF6 gene activation, 4:490–491, 4:491f
 inositol requiring 1 (IRE1) gene conservation, 4:490
 intramembrane proteolysis, 4:490–491
 metformin, 1:657
Methanobacterium, 3:32–34
Methanobacterium formicicum, 4:89
Methanobacterium thermoautotrophicum
 Cho phylogenetic tree, 3:331f
 initiation factors, 4:428–430, 4:428f
 uvrC phylogenetic tree, 3:331f
Methanocaldococcus jannaschii, 2:564f, 2:566, 4:50
Methanococcus, 3:32–34
Methanococcus jannaschii, 2:72–73, 3:34
Methanococcus maripaludis, 1:457f
 methanogenesis
 anaerobic respiration, 2:207f
 archaeal energy transduction, 3:32–33, 3:33f
 bacterial reductive reactions, 1:487t
 cobalamin, 1:159–160
Methanobolus, 3:32–34
Methanopyrus kandleri, 2:157, 2:158t, 2:566
Methanosarcina, 3:32–34
Methanosarcina barkeri
 biosynthetic pathway, 3:546
 reductive reactions, 1:487t
 respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:88–89
Methanosarcina mazei, 4:89
Methanothermobacter, 4:146f
Methanothermobacter marburgensis, 1:125
Methanothermus, 3:32–34
Methanotrix, 3:32–34
 methemoglobin, 2:534
 methionine
 anaplerotic pathways, 1:106f
 carnitine, 1:384
 cobalamin-dependent methionine synthase, 1:159, 1:159f, 2:327
 cytochrome *c*, 1:600–601
 folate coenzymes, 2:324–326, 2:325f
 metalloproteases, 3:87–88
 methionine aminopeptidases (MetAPs)
 general discussion, 1:97–99
 isoform functions, 1:98–99
 metal use, 1:98
 structural organization, 1:98, 1:98f
 substrate specificity, 1:98
 methionine sulfoxide reductase, 3:411f
 methylenetetrahydrofolate reductase, 2:312–313
 mitochondrial DNA (mtDNA), 3:134t
 phage libraries, 3:434t
 methoxamine, 1:58
 methoxy-4-hydroxymandelic acid (VMA), 1:57
 methoxy-4-hydroxyphenylglycol (MHPG), 1:57
 methyladenine, 2:1, 2:2f
 methylation, 2:9, 4:60–61
 methyl binding domain IV (MBD4)
 cancer prevention, 4:78
 comparison diagram, 1:523f
 DNA demethylation, 3:9, 3:10f
 DNA glycosylases, 2:16
 functional role, 1:523–524
 mismatch repair (MMR), 1:523f
 specificity, 4:77–78
 structure and biochemistry, 4:77

- α -methylbutyryl-CoA, 1:244, 1:245f
 methyl-CpG-binding proteins, 1:521–524
 binding domain family
 general discussion, 1:522
 MBD1, 1:523, 1:523f
 MBD2, 1:523, 1:523f
 MBD3, 1:523, 1:523f
 MBD4
 cancer prevention, 4:78
 comparison diagram, 1:523f
 DNA demethylation, 3:9, 3:10f
 functional role, 1:523–524
 mismatch repair (MMR), 1:523f
 specificity, 4:77–78
 structure and biochemistry, 4:77
 MBD5, 1:523f, 1:524
 MBD6, 1:523f, 1:524
 MeCP2, 1:522–523, 1:523f
 cytosine methylation
 basic concepts, 1:522
 cancer cells, 1:522
 functional role, 1:522
 gene expression regulation, 1:522
 genomic imprinting, 1:522
 X-chromosome inactivation, 1:522
 DNA methylation–chromatin structure
 interdependency, 1:524
 functional role, 1:521
 gene expression regulation, 1:521
 glycosylation processes, 3:9, 3:10f
 β -methylcrotonoyl-coenzyme A (CoA)
 carboxylase, 1:221
 2-methyleneglutarate mutase, 1:158t
 methylenetetrahydrofolate reductase
 (MTHFR), 2:312–313, 2:325f, 2:326
 methylenetetrahydrofolate acid (THF),
 3:715–716, 3:715f
 methylesterase (MEase), 3:607, 3:609–610,
 3:609f
 O⁶-methylguanine-DNA methyltransferase
 (MGMT), 2:12–13
 methylmalonic acidemia, 4:495t
 methylmalonyl-CoA, 2:208, 2:208f
 methylmalonyl CoA mutase, 1:158t, 2:327
 methylmethanesulfonate (MMS), 2:9, 2:10f,
 4:48–49
 N-methyl-N'-nitro-N-nitrosoguanidine
 (MNNG), 2:9, 2:10f, 2:51–52, 2:52f
 N-methyl-N-nitrosourea (MNU), 2:9, 2:10f
 Methylococcus capsulatus, 1:489t, 3:331f
 Methylomirabilis oxyfera, 1:489t
 methyltransferase (MTase)
 carboxyl methylation, 3:607, 3:609–610
 cytosine methylation
 basic concepts, 1:522, 3:6, 3:7f
 cancer cells, 1:522
 functional role, 1:522
 gene expression regulation, 1:522
 genomic imprinting, 1:522
 X-chromosome inactivation, 1:522
 deoxyribonucleic acid (DNA), 2:40–44,
 3:6–10
 basic concepts, 3:6, 3:7f
 Dam methyltransferase
 altered gene expression, 2:42–43
 bacterial pathogenesis, 2:43
 chromosome replication, 2:42
 dam mutant strains, 2:40
 gene inactivation, 2:43–44
 mismatch repair, 2:40–42, 2:41f, 2:42f
 Mut protein system, 2:40–42, 2:41f,
 2:42f
 oriC region, 2:42
 post-transcriptional regulation, 2:43
 properties, 2:40
 viruses, 2:43–44
 DNA demethylation
 glycosylation processes, 3:9, 3:10f
 hydroxylation processes, 3:9
 Dnmt1, 3:8–9, 3:8f
 Dnmt3 family
 characteristics, 3:6, 3:7f
 CpG dinucleotide spacing, 3:6
 de novo DNA methylation, 3:6, 3:7f
 Dnmt3a, 3:6, 3:7f
 Dnmt3L, 3:6, 3:7f
 histone H3 lysine 4 (H3K4)/histone
 H3 lysine 9 (H3K9) methylation,
 3:6–8, 3:7f
 5-methylcytosine (5mC), 3:9, 3:9f
 research summary, 3:9–10
 SET- and RING-associated (SRA)
 domain, 3:9, 3:9f
 type II restriction and modification
 enzymes, 2:134
 type III restriction and modification
 enzymes, 2:137–138, 2:137t
 UHRF1 (ubiquitin-like, containing PHD
 and RING finger domains 1), 3:9,
 3:9f
 methylurapidil, 1:58
 metoprolol, 1:59
 mevalonic acid (MVA)
 brassinosteroid biosynthesis, 1:251
 3-hydroxy-3-methylglutaryl coenzyme A
 (HMG-CoA) reductase regulated
 degradation, 1:518f, 1:519
 Mevinolin, 1:519
 Mfd protein, 3:335–336, 3:335f, 4:409
 Mgm1, 4:51f, 4:54
 mice
 DNA mismatch repair (MMR), 2:52
 fatty acyl-coenzyme A (CoA) synthetases,
 2:294
 nicotinic acid adenine dinucleotide
 phosphate (NAADP)-induced calcium
 ion (Ca²⁺) release, 2:638t
 superoxide dismutases (SODs), 4:354
 Michaelis complex, 2:758–759, 2:758f
 Michaelis constant (V/K), 4:348–349
 Michaelis-Menten/Briggs-Haldane
 equations, 4:273–274, 4:274f
 Michaelis-Menten equation, 2:216–217
 microarrays
 DNA microarrays
 basic concepts, 2:370–371
 data analysis, 2:371
 schematic diagram, 2:371f
 lectins, 2:704–705
 microautophagy, 1:150
 microbial electrolysis cell (MEC), 4:74–75,
 4:75f
 microfilaments, 3:233
 microglia, 1:369, 1:370f, 1:371–373
 microglobulins, 1:101t
 microRNA (miRNA), 3:98–102, 4:191–194
 biogenesis, 3:99–100, 3:99f, 4:191, 4:192f
 biological role, 3:101–102
 cardiac muscle development, 1:185
 characteristics, 3:98
 exosomes and microvesicles (EMVs),
 2:262–263
 functional role
 gene transcription regulation, 4:68,
 4:191
 glucose metabolism/insulin signaling,
 4:193, 4:194t
 lipid metabolism, 4:192–193, 4:194t
 metabolic disease treatment, 4:193,
 4:194t
 pancreatic β -cell function, 4:191–192,
 4:194t
 RNA-induced silencing complex (RISC),
 3:100, 3:101f
 historical research, 3:98–99
 hypoxia inducible factor (HIF), 4:68
 RNA editing, 4:155
 RNA-induced silencing complex (RISC),
 3:98, 3:99f, 3:101f
 RNA trafficking, 3:322
 skeletal muscle development, 1:182–183
 microsomal triglyceride transfer protein
 (MTTP), 4:548
 microtubule-associated proteins (MAPs),
 3:103–107
 bundle formation, 3:105–106, 3:106f
 classical MAPs
 characteristics, 3:103, 3:104t
 filamentous structures, 3:104f,
 3:105–106
 MAP1A/MAP1B, 3:103–104, 3:104t,
 3:105f, 3:106–107
 MAP2A/MAP2B/MAP2C, 3:104–105,
 3:104t, 3:105f, 3:106–107, 3:106f
 MAP4, 3:104t, 3:105, 3:105f
 microtubule-actin cross-linking factor
 (MACF), 3:105f, 3:107
 microtubule-based motor proteins, 4:208,
 4:208f
 neurite outgrowth, 3:105–106
 neurodegenerative diseases, 3:106–107
 plakin family, 3:105f, 3:107
 regulatory mechanisms, 3:105
 tau proteins, 3:104t, 3:105, 3:105f,
 3:106–107, 3:106f
 transgenic and knockout mice, 3:106–107
 tubulin isoforms, 4:450
 microtubules, 1:430–431, 1:431f, 3:233,
 4:214, 4:215f
 microvascular disease, 1:655–656
 microvesicles
 See exosomes and microvesicles (EMVs)
 microvillar inclusion disease, 3:223t
 microvilli, 1:35
 milk, 4:533t
 Miller syndrome, 3:713–714
 miniature inverted repeat transposable
 elements, 3:276, 3:276f, 3:281

- mini-chromosome maintenance (MCM)
 - proteins
 - characteristics and functional role, 2:118, 2:119f
- eukaryotic DNA replication, 2:108, 2:111–112, 2:112f
- polymerase α -primase (Pol α), 2:242–243, 2:242f
- pre-replication complex (preRC) assembly, 1:401, 1:402t, 1:403f
- recombination-dependent DNA replication (RDR), 4:24–25
- ring-shaped helicases, 2:21, 2:22t, 2:25–26
- MINT database, 3:615t
- miscarriages, 3:23
- mismatch repair (MMR), 2:51–53, 2:54–59
 - apoptosis
 - damage-induced cell-cycle growth arrest, 2:52, 2:52f
 - tumorigenesis, 2:53, 2:53f
 - basic concepts, 2:51
 - B-DNA (antiparallel right-handed double helix), 2:142
 - components, 2:52t
 - cytotoxic agents, 2:51–52
 - damage-induced cell-cycle growth arrest, 2:52, 2:52f
 - Dam methyltransferase, 2:40–42, 2:41f, 2:42f
 - disease states, 2:60–65
 - Lynch syndrome/colorectal cancer, 2:61–64, 2:62t, 2:63f
 - replication fidelity, 2:60
 - research summary, 2:65
 - DNA damage signaling interactions, 2:52
 - DNA polymerase III holoenzyme, 2:95
 - double-strand breaks (DSBs), 2:41, 2:42f, 2:79–80
 - downstream steps
 - characteristics, 2:69–70
 - excision, 2:69–70
 - resynthesis, 2:70
 - eukaryotic MMR proteins, 2:51
 - functional role, 2:51–52, 2:52t
 - G:T mispair repair
 - MBD4-mediated repair, 2:16, 4:77–78
 - thymine DNA glycosylase (TDG), 2:16, 4:78
 - very-short patch (VSP) repair, 4:76–77
 - homologous recombination (HR), 2:45–50
 - break-and-copy models, 4:23
 - DNA polymerase III holoenzyme, 2:95
 - Escherichia coli* anti-recombination activity, 2:47–48, 2:48f
 - evolutionary role, 2:50
 - loss of heterozygosity (LOH), 2:49–50, 2:49f
 - mutation-related carcinomas, 2:49–50, 2:49f
 - Saccharomyces cerevisiae* anti-recombination activity, 2:48–49
 - sequence divergence, 2:48f, 2:49–50, 2:49f
 - very-short patch (VSP) repair, 2:46f
 - long-patch mismatch repair
 - Dam methyltransferase, 2:54
 - DNA polymerase III holoenzyme, 2:95
 - drug resistance, 2:57
 - excision and repair synthesis, 2:56
 - initiation phase, 2:55–56
 - mismatch repair deficiency, 2:57
 - MutH pathway, 2:56, 2:56f
 - MutL pathway, 2:55–56, 2:56f
 - MutS pathway, 2:55, 2:56f
 - process mechanisms, 2:54–56, 2:56f
 - repair enzymes, 2:54–56, 2:55t, 2:56f
 - specificity, 2:54
 - mammals, 2:66–71
 - basic concepts, 2:66–68, 2:68f
 - downstream steps, 2:69–70
 - mutation types, 2:67f
 - MutL family proteins, 2:68–69, 2:69f
 - MutS family proteins, 2:68, 2:69f
 - research summary, 2:71
 - strand discrimination, 2:70–71, 2:70f
 - mismatch repair systems (MRs), 2:66–71
 - basic concepts, 2:45, 2:46f, 2:66–68, 2:68f
 - downstream steps, 2:69–70
 - Escherichia coli* methyl-directed MRS, 2:45–46
 - eukaryotes, 2:46–47
 - GO system (MutM–MutT–MutY), 2:80, 2:80f
 - long patch mismatch-repair system (LPMRS), 2:46f
 - mammals, 2:66–71
 - mutation types, 2:67f
 - MutL family proteins, 2:68–69, 2:69f
 - Mut protein family, 2:45–46, 2:48f, 2:54–59, 2:66–68, 2:67f
 - MutS family proteins, 2:60–61, 2:61f, 2:68, 2:69f
 - research summary, 2:71
 - sequence divergence, 2:48f, 2:66–68
 - specificity, 2:54
 - strand discrimination, 2:70–71, 2:70f
 - very short patch mismatch-repair system (VSPMRS), 2:46f
 - molecular mechanisms, 2:60–61, 2:61f
 - MutS–MutY interactions, 2:58
 - MutY repair pathway, 2:57–58, 2:57f
 - very-short patch (VSP) repair
 - homologous recombination (HR), 2:46f
 - research background, 4:76
 - usr* gene biochemistry, 4:76–77
 - usr* gene mutagenesis, 4:77
 - usr* gene structure and regulation, 4:76, 4:77f
 - mismatch uracil glycosylase (MUG) family, 2:20
 - Mitchellian redox loop mechanism, 1:468–469, 1:469f, 3:400–404, 3:402f
 - Mitchell, Peter, 1:467
 - mitochondria
 - adenosine triphosphate (ATP)
 - historical research, 3:306–307
 - mitochondrial ATP-dependent potassium ion (K^+) channel (KATP), 3:551, 3:551f
 - mitochondrial β -oxidation, 1:386–387
 - mitochondrial calcium (Ca^{2+}) transport, 3:118, 3:119f
 - mitochondrial channels, 3:127
 - mitochondrial CK (MtCK) isoenzyme, 3:82–84, 3:83f
 - mitochondrial dynamics, 3:135–136
 - regulation mechanisms, 3:169–170
 - aging/age-related diseases, 3:174–175
 - amyloid β -protein (A β)-induced mitochondrial fission, 1:478
 - anion selective channels, 3:128t, 3:129–130
 - anti-mitochondrial antibodies (AMAs), 3:113–117
 - anti-M2 AMAs
 - antigen identities, 3:114, 3:114t
 - B-cell epitopes, 3:115
 - bile duct cells, 3:115
 - characteristics, 3:113–116
 - chronic bacterial infections, 3:115
 - cross-reactivity evidence, 3:115
 - enzyme inhibition, 3:114–115
 - immune reaction stimuli, 3:115
 - occurrences, 3:114t
 - potential targets, 3:115
 - primary biliary cirrhosis (PBC), 3:114, 3:115–116
 - urinary tract infections (UTIs), 3:115
 - detection methods, 3:113
 - M-classification scheme
 - anti-M1 AMAs, 3:113, 3:114t
 - anti-M2 AMAs, 3:113–116, 3:114t
 - anti-M3 AMAs, 3:114t, 3:116
 - anti-M4 AMAs, 3:114t, 3:116
 - anti-M5 AMAs, 3:114t, 3:116
 - anti-M6 AMAs, 3:114t, 3:116
 - anti-M7 AMAs, 3:114t, 3:116
 - anti-M8 AMAs, 3:114t, 3:116
 - anti-M9 AMAs, 3:114t, 3:116
 - background information, 3:113–116
 - naturally occurring anti-mitochondrial antibodies, 3:117
 - pathogenic anti-mitochondrial antibodies, 3:117
 - tissue damage responses, 3:117
 - calcium ion (Ca^{2+}) messenger functions, 1:282–283, 1:283f
 - carbohydrate-responsive element binding protein (ChREBP) function, 1:377f
 - cation selective channels, 3:128t, 3:129
 - coenzyme Q, 3:723–724, 3:724f, 4:18–20, 4:19f
 - cytochrome *bc*₁ complex, 1:593, 1:594f
 - cytochrome *c*, 1:599
 - DNA ligases, 2:30t, 2:32
 - DNA mutation-caused diabetes, 1:653
 - DNA replication, 2:126–129
 - characteristics, 2:127–128
 - future research areas, 2:128
 - mammalian mtDNA, 2:127–128, 2:127f
 - mitochondrial constituents
 - mitochondrial DNA (mtDNA), 2:126
 - proteins, 2:126
 - ribonucleic acid (RNA), 2:126
 - strand-displacement mtDNA replication, 2:127–128, 2:127f

- mitochondria (*Continued*)
- yeast systems, 2:128
 - dysfunctional mitochondria, 3:173–174
 - fatty acid biosynthesis
 - lipogenic gene expression, 2:287f
 - pyruvate dehydrogenase regulation, 2:285–286
 - fatty acid oxidation
 - β -oxidation, 2:281–282, 2:282f, 3:65–66, 3:66f, 3:68f, 3:69f, 3:70f
 - general discussion, 2:281
 - regulation mechanisms, 2:283
 - uptake and activation, 2:281, 2:282f
 - fatty acid synthesis pathway, 2:276f
 - ferredoxins (Fd), 2:297
 - free radicals, 3:168–175
 - aging/age-related diseases, 3:174–175
 - biochemical chain reactions, 3:168–169, 3:169f
 - dysfunctional mitochondria, 3:173–174
 - historical research, 3:168–169
 - hydrogen peroxide (H_2O_2)
 - biochemical chain reactions, 3:168–169, 3:169f
 - mitochondrial production, 3:169–170, 3:171f, 3:173f
 - physiological functions, 3:172–173
 - mitochondrion-dependent apoptosis, 3:174
 - nitric oxide (NO), 3:170–172, 3:171f, 3:172–173, 3:172f, 3:173f
 - oxidative stress, 3:173–174
 - peroxynitrite, 3:171f, 3:172f
 - superoxide anion (O_2^-)
 - biochemical chain reactions, 3:168–169, 3:169f
 - mitochondrial production, 3:169–170, 3:170f, 3:171f, 3:173f
 - physiological functions, 3:172–173
 - gene expression, 3:140–144
 - basic process, 3:140–141, 3:141f
 - encoding processes, 3:140
 - 5'-end–RNase P process pathway, 3:142, 3:142f
 - mosaic mitochondrial genes, 3:142–144, 3:143f
 - RNA splicing, 3:142–144, 3:143f
 - 3'-end–RNase P process pathway, 3:142, 3:142f
 - transcription processes, 3:141, 3:141f, 3:142f
 - translation processes, 3:144
 - yeast systems, 3:140, 3:141f
 - glycolysis, 2:443, 2:444f
 - heme biosynthesis, 2:539–540, 2:540f
 - hereditary spastic paraplegias (HSPs), 4:267–270, 4:269f
 - Luft's disease, 2:760–761
 - clinical features, 2:760
 - definition, 2:760
 - muscle biopsy results
 - biochemistry, 2:761
 - genetic studies, 2:761
 - morphology, 2:760–761, 2:760f
 - polarographic studies, 2:761, 2:761f
 - therapies, 2:760
 - megamitochondria, 2:375–376, 2:375f
 - membranes, 3:145–148
 - cristae
 - characteristics, 3:145–146
 - dilation, 3:148
 - functional role, 3:146
 - fusion and fission regulation, 3:147
 - megamitochondria, 2:375–376, 2:375f
 - spontaneous formation, 3:146
 - electrochemical gradient, 3:145–146
 - electron microscopy analyses, 3:145
 - inner membrane
 - apoptosis, 3:148
 - ATP synthase, 3:147
 - calcium (Ca^{2+}) transport, 1:351
 - characteristics, 3:145
 - condensed mitochondria, 3:147
 - cristae, 3:145–146
 - cytochrome c release, 3:148
 - dilation, 3:148
 - dynamic properties, 3:146–147
 - functional role, 3:128–130, 3:128t, 3:130f
 - function-linked changes, 3:147–148
 - fusion and fission regulation, 3:147
 - hydrogen peroxide (H_2O_2), 3:169–170
 - metabolic-based changes, 3:147
 - mitochondrial ATP-dependent potassium ion (K^+) channel (KATP), 3:550
 - orthodox mitochondria, 3:147
 - osmotic adjustments, 3:147
 - oxidative stress, 3:148
 - structure, 3:146, 3:146f
 - intermembrane space, 3:145
 - matrix space, 3:145
 - outer membrane, 3:159–161
 - apoptosis, 3:161
 - calcium (Ca^{2+}) transport, 1:350–351
 - characteristics, 3:145
 - composition, 3:160
 - functional roles, 3:127–128, 3:128t, 3:159
 - importance, 3:161
 - megamitochondria, 2:375–376, 2:375f
 - origin, 3:159
 - structural characteristics, 3:159, 3:160f
 - tomographic reconstruction, 3:145, 3:146f, 3:147f
 - mitochondria-associated membranes (MAMs)
 - acyl-coenzyme A synthetase, 2:292–293
 - long-chain acyl-CoA synthetase (LCAS), 2:292–293
 - very long-chain acyl-CoA synthetase (VLCAS), 2:293–294
 - mitochondrial acyl-CoA/CoA ratio
 - regulation, 1:387
 - mitochondrial β -oxidation
 - carnitine-dependent cellular uptake and activation, 1:384–385, 1:385f
 - mitochondrial carnitine system, 1:385–386
 - reaction pathways, 1:386, 1:386–387
 - mitochondrial calcium (Ca^{2+}) transport, 1:347–352
 - affinity considerations, 3:123–124
 - basic concepts, 1:347
 - calcium (Ca^{2+}) uptake
 - affinity considerations, 3:123–124
 - conductivity states, 3:124–125, 3:124f, 3:125f
 - discovery, 3:120–121
 - disease connections, 3:125–126
 - early research results, 3:121–122
 - functional roles, 1:348–350, 1:350f
 - heart mitochondria, 3:119t
 - homeostasis control mechanisms, 3:124–125, 3:124f, 3:125f
 - molecular characterization, 3:125
 - permeability transition pore (PTP), 3:124–125, 3:124f
 - rat kidney mitochondria, 3:120f, 3:121f
 - rat liver mitochondria, 3:121f
 - reversible-to-irreversible damage transition, 3:109–110
 - sodium ion/calcium ion ($\text{Na}^+/\text{Ca}^{2+}$) exchanger, 3:125–126
 - cell stimulation, 1:349f
 - conductivity states, 3:124–125, 3:124f, 3:125f
 - disease connections, 3:125–126
 - endoplasmic reticulum (ER), 1:348
 - functional roles, 1:348–350, 1:350f
 - historical research, 1:347, 3:118–126
 - homeostasis control mechanisms, 3:124–125, 3:124f, 3:125f
 - inner membrane channels, 1:351
 - molecular characterization, 3:125
 - outer membrane channels, 1:350–351
 - permeability transition pore (PTP), 3:124–125, 3:124f
 - pH effects, 3:118–120, 3:120f
 - release leg, 3:122–123, 3:122f, 3:123f
 - research background, 1:348
 - research summary, 1:351
 - respiratory rate, 3:118–120, 3:119f
 - reversible-to-irreversible damage transition, 3:109–110
 - schematic diagram, 1:349f, 1:350f
 - sodium ion/calcium ion ($\text{Na}^+/\text{Ca}^{2+}$) exchanger, 1:424, 3:125–126
 - mitochondrial channels, 3:127–131
 - bioenergetics, 3:127
 - calcium (Ca^{2+}) transport, 1:350–351
 - cell death mechanisms, 3:130–131
 - characteristics, 3:127
 - functional role
 - anion selective channels, 3:128t, 3:129–130
 - cation selective channels, 3:128t, 3:129
 - inner membrane channels, 3:128–130, 3:128t, 3:130f
 - outer membrane channels, 3:127–128, 3:128t, 3:130f
 - mitochondrial apoptosis-induced channel (MAC), 3:128t, 3:130–131

- mitochondrial outer membrane permeabilization (MOMP), 3:130–131
- mitochondrial CK (MtCK) isoenzyme energy-transducing microcompartments, 3:82–84, 3:83f
- intramitochondrial inclusions, 3:84, 3:84f
- low-cellular-energy state, 3:84
- microcompartments, 3:82–84
- mitochondrial permeability transition pore, 3:84
- mitochondrial DNA (mtDNA), 3:132–133
 - background information, 3:132
 - base excision repair (BER), 2:7
 - characteristics and functional role, 2:126
 - disease connections
 - epigenomic mutations, 1:215–216
 - mitochondrial DNA (mtDNA), 1:212f, 1:214–215
 - nDNA-mitochondrial gene mutations, 1:215
 - DNA replication, 2:126–129
 - characteristics, 2:127–128
 - future research areas, 2:128
 - mammalian mtDNA, 2:127–128, 2:127f
 - mitochondrial constituents, 2:126
 - proteins, 2:126
 - ribonucleic acid (RNA), 2:126
 - strand-displacement mtDNA replication, 2:127–128, 2:127f
 - yeast systems, 2:128
 - epigenomics, 1:213
 - functional role, 1:210–213
 - future research areas, 3:134
 - gene expression, 3:140–144
 - basic process, 3:140–141, 3:141f
 - encoding processes, 3:140
 - 5'-end-RNase P process pathway, 3:142, 3:142f
 - mosaic mitochondrial genes, 3:142–144, 3:143f
 - RNA splicing, 3:142–144, 3:143f
 - 3'-end-RNase process pathway, 3:142, 3:142f
 - transcription processes, 3:141, 3:141f, 3:142f
 - translation processes, 3:144
 - yeast systems, 3:140, 3:141f
 - genetic studies, 1:213
 - global radiation, 1:214–215, 1:214f
 - historical research, 3:132–133, 3:306–307
 - metabolism, 1:212f, 2:30t, 2:32
 - mitochondrial permeability transition pore, 1:212–213, 1:212f, 3:128t, 3:130–131
 - mutation effects, 1:214–215
 - origins, 1:210–213
 - oxidative phosphorylation (OXPHOS), 1:210–213, 1:212f
 - properties, 3:133
 - reactive oxygen species (ROS), 1:212f, 2:79
 - research summary, 1:216
 - schematic diagram, 1:211f, 1:212f
 - sequence and gene content, 3:133–134, 3:133f, 3:133t, 3:134t
- mitochondrial dynamics, 3:135–139
 - cell death
 - cristae remodeling, 3:138
 - mitochondrial channels, 3:130–131
 - mitochondrial fragmentation, 3:138
 - characteristics and functional role, 3:135
 - endoplasmic reticulum (ER) tethering, 3:137–138, 3:137f
 - guanosine triphosphatases (GTPases), 3:136–137
 - mitochondrial fission, 3:136–137, 3:137f
 - mitochondrial fusion, 3:137, 3:137f
 - mitochondrial shape
 - cell death, 3:138
 - cell function, 3:138
 - morphology, 3:135–136, 3:136f
 - research summary, 3:139
 - mitochondrial Fe-S cluster assembly machinery, 2:663, 2:663f
- mitochondrial genome, 1:210–216, 3:140–144
 - disease connections
 - epigenomic mutations, 1:215–216
 - mitochondrial DNA (mtDNA), 1:212f, 1:214–215
 - nDNA-mitochondrial gene mutations, 1:215
 - epigenomics, 1:213
 - functional role, 1:210–213
 - general discussion, 1:210
 - genetic studies, 1:213
 - global radiation, 1:214–215, 1:214f
 - historical research, 3:306–307
 - metabolism, 1:212f
 - mitochondrial permeability transition pore, 1:212–213, 1:212f
 - mutation effects, 1:214–215
 - origins, 1:210–213
 - oxidative phosphorylation (OXPHOS), 1:210–213, 1:212f
 - reactive oxygen species (ROS), 1:212f
 - research summary, 1:216
 - schematic diagram, 1:211f, 1:212f
- mitochondrial permeability transition characteristics and functional role, 3:162
- chemiosmotic theory, 3:162
- mitochondrial permeability transition pore, 3:162–167
 - adenine nucleotide translocator (ANT), 3:163
 - calcium homeostasis, 3:164–165, 3:165f
 - CyP-D protein, 3:162–163
 - disease connections, 3:166–167
 - divalent cations, 3:163
 - kinase signaling pathways, 3:165–166
 - liver diseases, 3:166
 - mitochondrial CK (MtCK) isoenzyme, 3:84
 - mitochondrial DNA (mtDNA), 1:212–213, 1:212f, 3:128t, 3:130–131
- muscle diseases, 3:167
- myocardial ischemia, 3:110–111
- myocardial ischemia–reperfusion, 3:166
- neurological diseases, 3:166
- opening consequences, 3:164
- proton electrochemical gradient, 3:163–164
- reactive oxygen species (ROS), 4:18f
- redox sensitivity, 3:164
- regulatory mechanisms, 3:163–164
- theoretical perspective, 3:162–163
- voltage-dependent anion channel (VDAC), 3:163
- mitochondrial RNA (mRNA), 2:193
- mitochondrial topoisomerase, 2:158t
- mitochondrion-dependent apoptosis, 3:174
- mitophagy, 1:153
- myocardial ischemia, 3:108–112
 - cardiomyocytes, 3:108–109
 - endogenous self-defense mechanisms, 3:111
 - energy metabolism, 3:108–109, 3:109f
 - general discussion, 3:108
 - mitochondrial permeability transition pore, 3:110–111
 - reversible-to-irreversible damage transition, 3:109–110
- paraplegin, 4:271
- PET genes, 3:306–309
 - cytochrome *b₆f* complex, 4:338t
 - disease connections, 3:309
 - functional role
 - essential PET genes, 3:309
 - general discussion, 3:307–309
 - mitochondrial genome, 3:307
 - nonessential PET genes, 3:307–309, 3:308f
 - respiratory chain complexes, 3:308f
 - historical research, 1:226
 - midpoint oxidation–reduction potentials, 4:338t
 - origins, 3:307
 - regulation mechanisms, 3:309
 - Saccharomyces cerevisiae*
 - model systems, 3:307
 - number of proteins, 3:307
- phospholipase D (PLD), 3:477
- photosynthetic metabolic pathways, 3:497–498, 3:498f
- plants, 1:145–148
 - alternative oxidases (AOXs)
 - basic concepts, 1:145–146
 - biochemical regulation, 1:147–148, 1:147f
 - physiological functions, 1:148
 - structural characteristics, 1:146–147, 1:147f
 - cyanide-resistant respiration, 1:145–146
 - mitochondrial DNA (mtDNA), 3:133t
 - mitochondrial electron transport chain, 1:141–142, 1:142f, 1:145, 1:146f
 - potassium ion (K⁺) channels, 3:550–553
 - ATP-dependent potassium ion (K⁺) channel (KATP), 3:551, 3:551f

mitochondria (*Continued*)

- calcium-dependent potassium channels, 3:551–552, 3:551f
- characteristics and functional role, 3:550
- MitoKv1.3, 3:551f, 3:552
- potassium (K^+) cycle, 3:550–551, 3:551f
- research summary, 3:552
- two-pore potassium channel (TASK-3), 3:551f, 3:552
- primary biliary cirrhosis (PBC), 3:562
- protein degradation, 3:622–623, 3:623t
- protein import, 3:632–636
 - background information, 3:632
 - inner membrane
 - general discussion, 3:635–636
 - insertion from matrix, 3:634f, 3:636
 - TIM22 complex, 3:634f, 3:636
 - TIM23 complex, 3:634f, 3:635–636
 - intermembrane space, 3:634–635
 - mitochondrial matrix
 - N-terminal presequences, 3:635
 - protein folding and assembly processes, 3:635
 - TIM23 complex, 3:635
 - targeting signals
 - characteristics, 3:632
 - internal targeting signals, 3:632
 - presequences, 3:632
 - transport mechanisms
 - general discussion, 3:633–634
 - insertion pathways, 3:633–634, 3:634f
 - translocase of the outer membrane (TOM) complex, 3:633, 3:633t, 3:634f
- protonmotive electron transfer chain, 1:470–471, 1:470f, 1:471f
- reactive oxygen species (ROS)
 - ATP-dependent potassium ion (K^+) channel (KATP), 3:551
 - control mechanisms, 4:21
 - functional role, 4:17, 4:18f
 - generation mechanisms
 - basic principles, 4:18–20, 4:19f
 - glycerol phosphate dehydrogenase (GPDH), 4:19f, 4:20
 - pyruvate dehydrogenases (PDHs), 4:19f, 4:20
 - respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:18–20, 4:19f
 - respiratory chain complex II (succinate coenzyme Q reductase), 4:19f, 4:20
 - respiratory chain complex III (ubiquinol cytochrome *c* reductase), 4:19f, 4:20
 - reverse electron transport (RET), 4:19f, 4:20
 - sensing mechanisms, 4:20–21
 - respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:87–91
 - basic concepts, 4:87, 4:341
 - disease connections
 - general discussion, 4:90–91
 - mitochondrial mutations, 4:90–91
 - neurotoxins, 4:91
 - nuclear mutations, 4:91
 - pathological mechanisms, 4:91
 - evolutionary development
 - general discussion, 4:88–90
 - N-module, 4:89
 - origins, 4:89–90
 - P-module, 4:89
 - Q-module, 4:88–89
 - variants, 4:89–90
 - functional role
 - active–deactive transition, 4:90
 - electron transfer mechanisms, 3:661–662, 3:662f, 3:663, 4:90
 - general discussion, 4:90
 - inhibitors, 4:90
 - proton pumping, 3:661–662, 3:662f, 3:663, 4:90
 - redox centers, 4:90
 - sodium ion (Na^+) pumping, 4:90
 - PET genes, 3:308f
 - structural characteristics, 4:341–346
 - accessory subunits, 4:88, 4:89t
 - central subunits, 4:87–88, 4:88f, 4:89t
 - cysteine motif, 4:344–345, 4:345f
 - flavin mononucleotide (FMN) binding, 4:344–345, 4:344f
 - frataxin-like subunit, 4:343–344, 4:343f
 - hydrophilic domain, 4:342–343, 4:342f
 - iron–sulfur clusters, 4:87–88, 4:89t, 4:342–343, 4:342f, 4:345f
 - multistage assembly pathway, 4:88
 - mutations, 4:345
 - NADH binding, 4:344–345, 4:344f, 4:345f
 - overall structure, 4:341–342
 - redox centers, 4:87–88, 4:89t, 4:342–343, 4:342f
 - research summary, 4:345–346
 - shape, 4:87, 4:88f
 - subunits, 4:342t
 - supernumerary subunits, 4:345
- respiratory chain complex IV (cytochrome *c* oxidase), 4:98, 4:99f
- rhomboid-based mediation, 4:51f, 4:54
- ribozyme distribution, 4:146f
- spectrophotometric assays, 4:275–276, 4:275f
- structural characteristics, 3:145, 3:146f
- thiamine diphosphate (ThDP), 1:206, 1:207f
- type I DNA topoisomerase, 2:160f
- uncoupling proteins (UCPs), 4:482–487
 - adenosine diphosphate (ADP) phosphorylation and respiration
 - adenosine triphosphate (ATP) synthesis, 4:482
 - brown adipose tissue (BAT), 4:483–484, 4:484f
 - mitochondrial membranes, 4:482
 - ATP synthesis, 1:142f, 1:144
 - brown adipose tissue (BAT)
 - adenosine diphosphate (ADP) phosphorylation and respiration, 4:483–484, 4:484f
 - characteristics and functional role, 4:483
 - histology, 4:483f
 - thermogenesis, 4:483–484, 4:484f
 - UCP1, 4:483–484, 4:484f, 4:485f
 - characteristics and functional role, 4:482
 - leptin, 2:258–259
 - mitochondrial ROS, 4:21
 - novel UCPs
 - energy loss mechanisms, 4:484–486
 - functional role, 4:485–486, 4:486f
 - glucose metabolism, 4:486
 - messenger RNA (mRNA), 4:484–485, 4:485f
 - plant and bird UCPs, 4:485
 - reactive oxygen species (ROS), 4:485–486, 4:486f
 - skeletal muscle, 4:485, 4:485f
 - UCP2, 4:484–485, 4:485f
 - UCP3, 4:485, 4:485f
 - research summary, 4:486
 - vitamin D conversion processes, 4:535–536, 4:536f
 - voltage-dependent anion-selective channels (VDACs)
 - functional role, 3:161
 - mammalian low- K_m hexokinases, 2:546
 - mitochondrial CK (MtCK) isoenzyme, 3:82–84, 3:83f
 - outer membrane functions, 3:159
 - structural characteristics, 3:160–161, 3:160f
- mitochondrial BCAT (BCATm) knockout mice, 1:246, 1:246f
- mitochondrial DNA (mtDNA)
 - gene expression, 3:140–144
 - basic process, 3:140–141, 3:141f
 - encoding processes, 3:140
 - 5'-end–RNase P process pathway, 3:142, 3:142f
 - mosaic mitochondrial genes, 3:142–144, 3:143f
 - RNA splicing, 3:142–144, 3:143f
 - 3'-end–RNase process pathway, 3:142, 3:142f
 - transcription processes, 3:141, 3:141f, 3:142f
 - translation processes, 3:144
 - yeast systems, 3:140, 3:141f
- mitochondrial encephalomyopathy (MELAS), 1:211f
- mitochondrial fission factor (Mff), 3:136–137
- mitochondrial intermembrane space sorting sequences (MISSs), 3:632
- mitochondrial sterol 27-hydroxylase (CYP27A1)
 - bile salts, 1:168–169, 1:168f
 - vitamin D, 4:535–536
- mitogen-activated protein kinase (MAPK), 3:176–180
- adipocyte development, 1:55, 1:55f
- adrenergic receptors, 1:59

- bile salts, 1:169
calcium ion (Ca^{2+}) signaling system, 2:719–720, 2:720f, 3:293f, 3:294
chemokine receptors, 1:480–482, 1:482f
dopamine receptors, 2:172
extracellular signal-regulated kinase (ERK)
 angiotensin receptors, 1:113
 calcium ion (Ca^{2+}) signaling system, 2:719–720, 2:720f
 c-fes proto-oncogene (*Fes*), 1:453
 chemokine receptors, 1:480–482, 1:482f
 cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:564, 1:564f
 ERK1/ERK2, 3:176, 3:177f
 ERK3/ERK4, 3:177f, 3:178
 ERK5, 3:177f, 3:178
 ERK7/ERK8, 3:177f, 3:178
 fibroblast growth factor receptors (FGFRs), 2:304, 2:305f
 focal adhesion kinase (FAK) family, 2:265–266
 focal adhesion signaling, 2:321f, 2:322
 glycoprotein hormone receptors, 4:388–390
 G protein-coupled receptor kinases (GRKs), 2:485t
 G_q family proteins, 2:497–498, 2:498f
 integrin signaling, 2:611
 mTOR complexes, 3:195–196
 myocardial ischemia, 3:111
 neurotrophins, 3:241–243, 3:242f
 phylogenetic relationships, 3:177f
 Raf kinases, 4:247–248, 4:247f
 schematic diagram, 3:293f
 serine–threonine phosphatases, 4:222–223, 4:222f
 signaling cascades, 3:177f
 somatostatin (*sst*) receptors, 4:265
 submembranous calcium signals, 3:294
 tachykinins, 4:362–363
 toll-like receptors (TLRs), 4:398–399
 transforming growth factor beta ($\text{TGF-}\beta$), 1:620f
fibroblast growth factor receptors (FGFRs), 2:304, 2:305f
functional roles, 3:176
 $\text{G}_{\beta\gamma}$ subunit functions, 2:377
gene expression regulation, 3:179
glycogen synthase kinase-3 (GSK3), 2:436
G-protein-signaling regulators (RGSs), 2:493
 G_q family proteins, 2:497–498, 2:498f
hematopoietin receptors, 2:528f
hemidesmosome function, 1:642–643
hydrogen peroxide (H_2O_2), 3:173
insulin-like growth factor-1 (IGF1), 1:198–199, 1:200f
insulin receptors (IRs), 2:602–603
insulin resistance mechanisms, 1:651, 1:651f
insulin signal regulation, 2:606
insulin signal transduction, 2:605–606
c-jun N terminal kinase (JNK) pathway
 angiotensin receptors, 1:113
 c-fes proto-oncogene (*Fes*), 1:453
 functional role, 3:177–178
 glycogen synthase kinase-3 (GSK3), 2:435f, 2:436t
 hydrogen peroxide (H_2O_2), 3:173, 3:173f
 insulin resistance, 2:606–607
 insulin signal transduction, 2:605–606
 integrin signaling, 2:611
 mitochondrion-dependent apoptosis, 3:174
 muscarinic acetylcholine receptors, 3:216
 neurogenesis, 1:199
 nitric oxide (NO), 3:173, 3:173f
 p75 neurotrophin receptor (p75NTR), 3:245, 3:245f
 phylogenetic relationships, 3:177f
 signaling cascades, 3:177f
 toll-like receptors (TLRs), 4:398–399
mTOR complexes, 3:195–196
multiple opioid receptors, 3:357
muscarinic acetylcholine receptors, 3:216
neurogenesis, 1:198–199, 1:200f
neurokinin-1 receptor (NK1R), 4:362–363
neurotrophins, 3:241–243, 3:242f
nitric oxide (NO), 3:173
nuclear signal transmission specificity and fidelity, 3:178–179
p38 kinase
 angiotensin receptors, 1:113
 chemokine receptors, 1:480–482
 functional role, 3:178
 G protein-coupled receptor kinases (GRKs), 2:485t
 G-protein-signaling regulators (RGSs), 2:493
 muscarinic acetylcholine receptors, 3:216
 neurokinin-1 receptor (NK1R), 4:362–363
 phylogenetic relationships, 3:177f
 signaling pathways, 1:618f, 1:620f, 3:177f
pheromone receptors, 3:441–442, 3:443f
phylogenetic relationships, 3:177f
platelet-activating factor receptor (PAFR), 3:536, 3:536f
platelet-derived growth factor (PDGF) receptors, 3:541f
RAGE (receptor for advanced glycation end products), 2:407–409
somatostatin (*sst*) receptors, 4:265
steroid/thyroid hormone receptors, 4:310f
T-cell receptors (TCRs), 4:374
toll-like receptors (TLRs), 4:398–399
vascular endothelial growth factor receptors (VEGFRs), 4:512
mitophagy, 1:153
mitosis, 3:181–184
 chromosome cycle, 3:17–18, 3:18f
 chromosome organization, 1:537–538
 DNA mismatch repair (MMR), 2:49–50, 2:49f
 nuclear lamina assembly
 characteristics, 3:311–312
 nuclear pore complex positioning, 3:312
 nuclear shape and rigidity, 3:312
 nuclear size, 3:312
phases
 anaphase, 3:182f, 3:183–184, 3:183f
 cytokinesis, 3:182f, 3:183f, 3:184
 metaphase, 3:182f, 3:183, 3:183f
 prometaphase
 micrograph, 3:182f, 3:183f
 microtubule behavior, 3:182–183, 3:182f
 nuclear envelope breakdown, 3:182
 spindle assembly, 3:182–183
prophase
 centromeres, 3:18, 3:18f, 3:181, 3:182f
 centrosome separation, 3:181–182, 3:182f
 chromosome condensation, 3:181, 3:182f
 kinetochore assembly, 3:18, 3:18f, 3:181, 3:182f
 micrograph, 3:183f
 telophase, 3:182f, 3:183f, 3:184
 spindle structure, 1:623–624, 3:181, 3:182f
mitotic checkpoint, 1:417–422
 anaphase-promoting complex (APC), 1:419–420, 1:419f, 1:420f
 aurora B/Ipl1 protein, 1:418
 centromere protein E (CENP-E), 1:418–419, 1:419f
 checkpoint pathways, 1:419
 checkpoint protein roles, 1:421–422
 direct inhibitor model, 1:420–421, 1:420f
 functional role, 1:417
 kinetochore monitoring, 1:417
 microtubule attachment monitoring, 1:417–418
 molecular components, 1:417
 sequestration model, 1:419–420, 1:419f
 tension monitoring, 1:417–418
 unattached kinetochores, 1:422
mitotic kinesin-like protein-1 (MKLP-1), 1:623–624
mitoxantrone, 2:167f
MK0974 nonpeptide antagonist, 1:265, 1:266f
Mlh protein family, 2:46–47, 2:48–49
MLL transcription factor, 1:187, 1:188f
MMS21, 4:468t, 4:469
ModBase database, 3:615t
Moeller's disease, 4:530
molecular motors
 See myosin
molecular oxygen (O_2) toxicity, 4:352
molecules interacting with CasL (Micals), 1:24
molluscus, 3:538
molybdenum (Mo)
 iron–molybdenum cofactor (FeMo-co), 2:659f, 2:663
 pterin metabolism, 3:668–669, 3:669f
 riboswitches, 4:137
monilethrix, 2:670t
monoacylglyceride (MAG), 1:49–51, 3:418
monoamine oxidase, 3:185–187
aggressive behaviors, 3:187

- monoamine oxidase (*Continued*)
 anti-mitochondrial antibodies (AMAs), 3:116
 catalyzed reactions, 3:185, 3:186f
 catecholamines, 1:57
 cellular localization, 3:185, 3:186f
 encoding processes, 3:185
 enzyme substrates, 3:185–186, 3:186f, 3:187f
 flavins, 2:312, 3:185
 functional role, 1:87
 inhibitors, 3:186–187
 pharmacological characteristics, 3:186–187, 3:187f
 physiological functions, 1:87
 reaction mechanisms, 1:87
 structural characteristics, 1:88f, 3:186f
- monocarboxylate transporters (MCTs), 1:370f, 1:371
- monocytes, 2:569
- monogalactosyldiacylglycerol (MGDG), 1:508
- monooxygenases
 ascorbic acid, 4:531–532
 biological function, 3:374
 characteristics, 3:372
 external monooxygenases, 3:372
 internal monooxygenases, 3:372
- monopoiesis, 1:188–189, 1:188f
- monosaccharides
 N-glycan processing reactions, 2:449–450, 2:449f, 2:450f
 sugar phosphorylation, 2:543
- monotremes, 1:553f
- mosquitoes
 blood meal metabolism, 2:591, 2:591f
 juvenile hormone (JH), 2:593–594
 20-hydroxyecdysone (20E), 2:592–593
 urea synthesis, 2:591–592, 2:592f
- Mössbauer spectroscopy, 2:661–662
- [MOT3] prion
See prions
- moths, 1:600, 1:600f
- motivizumab, 3:431
- MPS1 gene, 1:417
- Mrc1 coupling factor, 1:402t, 1:403f
- MRE11 protein, 3:272, 4:31–32
- MRN protein complex, 3:272
- mst-2* gene, 1:82, 1:82f
- MSP1 clade, 1:3t, 1:4, 1:4f
- Mss116p protein, 4:151–152, 4:152f
- mTOR (rapamycin), 3:194–199
 background information, 3:194
 branched-chain amino acids (BCAAs), 1:246f
 cell growth regulation, 1:408
 characteristics and functional role, 3:194–195, 3:195f
 chemokine receptors, 1:482f
 chemotaxis, 2:236–237
 disease connections, 3:198–199
 fatty acid metabolism–cancer relationship, 2:277
 functional role, 3:464f
 insulin signal transduction, 2:604f, 2:605
 leptin regulation, 2:259
 macroautophagy regulation, 1:150t
- mTOR complexes
 cellular functions
 general discussion, 3:196–198
 mTORC 1, 3:196–198
 mTORC 2, 3:198
 characteristics and functional role, 3:194–195, 3:195f, 3:196f
 mTORC 1
 autophagy, 1:150t, 1:152–153, 3:198, 3:450
 cell cycle/cell survival, 3:198
 cellular functions, 3:196–198
 characteristics and functional role, 3:194–195, 3:195f
 gene transcription regulation, 3:198
 protein phosphatases, 3:198
 protein synthesis, 3:196–197, 3:197f
 regulatory mechanisms, 3:195–196
 ribosome biogenesis, 3:197–198
 S6 kinases, 3:196f, 3:197, 3:197f
 substrates, 3:195, 3:196f
 regulatory mechanisms, 3:195–196
 substrates, 3:195, 3:196f
 neurogenesis, 1:199, 1:200f
 phosphatidylinositol-3-phosphate (PtdIns(3)P), 3:450
 phospholipase D (PLD), 3:476
 protein kinase C (PKC), 3:637–638
 Ras proteins, 3:195–196, 4:15f
 research summary, 3:199
- mucins, 3:200–204, 3:205–208
 background information, 3:205
 characteristics, 3:200
 disease connections
 cancer, 3:203–204
 gastrointestinal disorders, 3:203
 ocular surfaces, 3:203
 pulmonary disorders, 3:203
 domain structures
 adhesion-associated domains, 3:201f, 3:202
 AP-2 domain, 3:201f, 3:202
 Cys subdomain (CS) domain, 3:201f, 3:203
 cystine knot (CK)-like domains, 3:201f, 3:203
 cytoplasmic domains, 3:201f, 3:202
 D-like domains, 3:201f, 3:202–203
 epidermal growth factor (EGF)-like domains, 3:201f, 3:202
 SEA-like domains, 3:201f, 3:202
 secreted mucins, 3:201f, 3:202
 transmembrane domains, 3:201f, 3:202
 embryo implantation process, 3:205, 3:206f
 expression and transcriptional control, 3:206–208, 3:208f
 family classifications
 characteristics, 3:201, 3:201f
 tethered mucins, 3:201
 functional role, 3:200, 3:207t, 3:208–209
 glycosylation
 definition, 3:200–201
 N-linked glycosylation, 3:201
 O-linked glycosylation, 3:200–201
 research summary, 3:209
- secreted mucins
 Cys subdomain (CS) domain, 3:201f, 3:203
 cystine knot (CK)-like domains, 3:201f, 3:203
 D-like domains, 3:201f, 3:202–203
 domain structures, 3:201f, 3:202
 transmembrane proteins, 3:207t
 transport and metabolism regulation, 3:208–209
 variable number tandem repeats (VNTRs), 3:201f, 3:202
- mucosa-associated lymphoid tissue 1 (MALT1), 4:377–378
- multicopy single-stranded DNA (msDNA), 3:284
- multidrug resistance (MDR) membrane proteins, 3:210–214
 ABCG2 (MXR, BCRP), 3:213, 3:214
 BSEP, 3:213, 3:214
 homologous transporters, 3:213
 MDR1
 clinical relevance, 3:210–211
 functional role, 3:213
 high affinity–substrate relationships, 3:213
 physiological functions, 3:213
 polyspecific transport cycle, 3:211–213, 3:212f
 research background, 3:210
 structural characteristics, 3:211–213, 3:211f
 MDR3, 3:213, 3:214
 MRP1 (ABCC1), 3:213, 3:214
 physiological functions
 ABCG2 (MXR, BCRP), 3:214
 ABCG5 and ABCG8, 3:214
 advantages, 3:214
 BSEP, 3:214
 homologous transporters, 3:214
 MDR1, 3:213
 MDR3, 3:214
 MRP1 (ABCC1), 3:214
 research summary, 3:214
- multi-heme c-type cytochromes, 3:404–405, 3:404f
- multi-minicore disease (MmD), 4:197–198
- multiple carboxylase deficiency, 4:495t
- multiple epiphyseal dysplasia, 3:657t
- multiple hereditary exostoses (MHE), 2:471t, 2:475
- murexide, 2:574–577
- murine embryonic fibroblasts (MEFs), 1:406–407, 1:407f
- muscarinic acetylcholine receptors, 3:215–217
 acetylcholine, 3:215
 calcium ion (Ca²⁺) signaling, 2:720f
 gene families, 3:215, 3:216t
in vivo actions
 central nervous system (CNS), 2:720f, 3:217
 peripheral nervous system, 3:216–217
 signal transduction
 guanosine diphosphate (GDP) dissociation, 3:215

- ion channels, 3:216
- mitogen-activated protein kinase (MAPK), 3:216
- second messenger pathways, 3:215–216
- structural characteristics
 - coupling regions, 3:216
 - ligand binding, 3:216
 - transmembrane topology, 3:216
- muscle–eye–brain (MEB) disease, 2:471*t*, 2:474–475
- muscles
 - amino acid metabolism, 1:93, 1:94*f*
 - aquaporin water channels, 1:124*t*
 - carbohydrate-responsive element binding protein (ChREBP) distribution, 1:374
 - cardiac muscles, 1:179–186
 - development process
 - cardiogenesis, 1:183, 1:183*f*, 1:184*f*
 - microRNA (miRNA), 1:185
 - transcription factors, 1:183–185, 1:184*f*
 - research summary, 1:186
 - stem cells, 1:186
 - structural characteristics, 1:179, 1:180*f*
 - glycogen activation, 2:426
 - glycogen storage diseases, 2:431*t*
 - glycogen synthesis, 2:430*f*
 - G_s family phenotypes, 2:521*t*
 - leukotriene receptors, 3:584
 - mammalian glucose transport, 2:392*t*
 - microRNA (miRNA)-based regulation, 4:194*t*
 - muscle contraction, 2:641–642
 - pyruvate dehydrogenase kinase (PDK) regulation, 4:324–326
 - skeletal muscle
 - development process
 - microRNA (miRNA), 1:182–183
 - myogenesis, 1:179–180, 1:180*f*, 1:181*f*
 - transcription factors, 1:180–182, 1:181*f*
 - research summary, 1:186
 - stem cells, 1:186
 - spectrophotometric assays, 4:278*f*, 4:279
 - striated muscles, 1:179–186
 - development process
 - cardiogenesis, 1:183, 1:183*f*, 1:184*f*
 - microRNA (miRNA), 1:182–183, 1:185
 - myogenesis, 1:179–180, 1:180*f*, 1:181*f*
 - transcription factors, 1:180–182, 1:181*f*, 1:183–185, 1:184*f*
 - research summary, 1:186
 - stem cells
 - cardiac muscles, 1:186
 - skeletal muscle, 1:186
 - structural characteristics, 1:179, 1:180*f*
 - vascular smooth muscles, 2:345*t*
 - muscular dystrophies, 2:471*t*, 2:474–475, 2:627–628, 2:627*t*, 2:632*t*, 3:167
 - Mus musculus*
 - actin-related proteins (Arps), 1:37*f*
 - base excision repair (BER), 2:5*t*
 - N-myristoyltransferase (Nmt), 3:642–643
 - retrotransposons, 3:284*f*
 - small RNAs (sRNAs), 3:98
 - mutagenic O⁴-methylthymine (O⁴-mT), 2:9, 2:10*t*, 2:12–13
 - mutant hemoglobins
 - fetal hemoglobin (HbF), 2:534
 - sickle cell hemoglobin (HbS), 2:534–535
 - Mut protein family
 - components, 2:52*t*
 - DNA glycosylases, 2:18, 2:18*f*, 2:19*f*, 2:54
 - functional role, 2:52*t*, 2:54–56, 2:55*t*, 2:56*f*
 - mismatch repair (MMR)
 - apoptosis, 2:53, 2:53*f*
 - cytotoxic agents, 2:51–52
 - Dam methyltransferase, 2:40–42, 2:41*f*, 2:42*f*
 - Escherichia coli*, 2:45–46, 2:48*f*, 2:51, 2:54, 2:66–68, 2:67*f*
 - MutL–MutS pathway, 4:76–77
 - MutS–MutY interactions, 2:58
 - MutS pathway, 2:54
 - MutY repair pathway, 2:54, 2:57–58, 2:57*f*
 - specificity, 2:54
 - MutH
 - double-strand breaks (DSBs), 2:42*f*
 - excision and repair synthesis, 2:56
 - functional role, 2:52*t*, 2:54–56, 2:55*t*, 2:56*f*
 - mismatch repair (MMR), 2:40–42, 2:41*f*, 2:56, 2:56*f*, 2:67, 2:67*f*
 - MutL
 - apoptosis, 2:53*f*
 - excision and repair synthesis, 2:56
 - functional role, 2:51, 2:52*t*, 2:54–56, 2:55*t*, 2:56*f*
 - mismatch repair (MMR)
 - Dam methyltransferase, 2:40–42, 2:41*f*
 - functional role, 2:55–56, 2:66–68, 2:67*f*
 - homologous recombination (HR), 2:48*f*
 - long-patch mismatch repair, 2:56*f*
 - mismatch recognition and excision, 2:68–69, 2:69*f*
 - molecular mechanisms, 2:60–61
 - very-short patch (VSP) repair, 4:76–77
 - MutM, 2:16–17, 2:57–58, 2:57*f*, 2:80, 2:80*f*
 - MutS
 - apoptosis, 2:53*f*
 - crossover pathways, 2:560
 - excision and repair synthesis, 2:56
 - functional role, 2:51, 2:52*t*, 2:54–56, 2:55*t*, 2:56*f*
 - mismatch repair (MMR)
 - Dam methyltransferase, 2:40–42, 2:41*f*
 - functional role, 2:55, 2:66–68, 2:67*f*
 - homologous recombination (HR), 2:48*f*
 - long-patch mismatch repair, 2:56*f*
 - Lynch syndrome/colorectal cancer, 2:61–64, 2:62*t*, 2:63*f*
 - mismatch recognition complex, 2:68, 2:69*f*
 - molecular mechanisms, 2:60–61, 2:61*f*
 - MutS–MutY interactions, 2:58
 - very-short patch (VSP) repair, 4:76–77
 - MutT, 2:57–58, 2:57*f*, 2:80, 2:80*f*
 - MutTY, 2:57*f*
 - MutY
 - DNA glycosylases, 2:18, 2:19*f*
 - functional role, 2:54–56, 2:55*t*, 2:56*f*
 - MutS–MutY interactions, 2:58
 - oxidative repair, 2:80, 2:80*f*
 - repair pathways, 2:57–58, 2:57*f*
 - myasthenic syndrome, 3:657*t*
 - Mycobacterium genitalium*, 2:5*t*
 - Mycobacterium leprae*, 3:331*f*
 - Mycobacterium phlei*, 3:372
 - Mycobacterium smegmatis*, 3:268
 - Mycobacterium tuberculosis*
 - acyl-coenzyme A synthetase families, 2:292
 - ATP synthase, 2:273–274
 - Cho phylogenetic tree, 3:331*f*
 - group II chaperonins, 1:456–457
 - heterotrophic glutathione reductase (GR)-type FNRs, 2:301
 - nicotinamide adenine dinucleotide (NAD⁺/NADH), 1:177
 - non-homologous end joining (NHEJ), 3:266
 - phagocytosis, 2:199
 - superoxide dismutases (SODs), 4:353
 - uvrC* phylogenetic tree, 3:331*f*
 - Mycoplasma genitalium*
 - Cho phylogenetic tree, 3:331*f*
 - ribosomal RNA (rRNA), 3:556
 - RNA polymerases, 4:173
 - transfer RNA (tRNA)
 - maturation processes, 3:555*f*
 - structural characteristics, 3:554–555
 - 3'-end–RNase process pathway, 3:555*f*, 3:556
 - uvrC* phylogenetic tree, 3:331*f*
 - Mycoplasma hyapneumoniae*, 3:331*f*
 - Mycoplasma pneumoniae*, 3:331*f*
 - Mycoplasma pubnonis*, 3:331*f*
 - MyD88-dependent signaling pathway, 4:398, 4:399*f*, 4:400*f*
 - myelin-associated glycoprotein (MAG), 3:243, 3:244*f*, 4:229, 4:230, 4:230*f*, 4:231
 - myelin–axonal interactions, 4:272
 - myeloid leukemia, 1:452
 - myeloperoxidase (MPO), 3:407–409, 3:408*f*
 - Myf5, 1:180–182, 1:181*f*
 - myocardial ischemia, 3:65–71
 - basic concepts, 3:65
 - energy metabolism
 - basic principles, 3:67, 3:68*f*
 - ischemic-induced metabolic signaling alterations, 3:67–69, 3:69*f*
 - metabolic-based treatment approaches, 3:71, 3:71*t*
 - reperfused ischemic heart
 - basic principles, 3:69–70, 3:70*f*

- myocardial ischemia (*Continued*)
 intracellular acidosis/cardiac efficiency, 3:70–71
 research summary, 3:71
 mitochondria, 3:108–112
 cardiomyocytes, 3:108–109
 endogenous self-defense mechanisms, 3:111
 energy metabolism, 3:108–109, 3:109f
 general discussion, 3:108
 mitochondrial permeability transition pore, 3:110–111
 reversible-to-irreversible damage transition, 3:109–110
 mitochondrial permeability transition pore, 3:166
 myocardin, 1:184f
 myocarditis, 3:117
 myoclonic epilepsy, 1:211f, 1:214
 myocyte enhancer factor 2 (MEF2), 1:180–182, 1:181f, 1:184f, 1:312, 1:313f
 MyoD, 1:180–182, 1:181f
 myogenic regulatory factors (MRFs), 1:180–182, 1:181f
 myogenin, 1:180–182, 1:181f
 myoglobin (Mb)
 allosteric regulation, 1:71
 carbon monoxide (CO), 2:534
 cyanide, 2:534
 functional role, 2:532
 oxygen-binding characteristics, 2:533, 2:534f
 structural characteristics, 2:532, 2:533f
 MyoR, 1:185
 myosin, 3:218–224
 actin cytoskeleton, 1:430–431, 1:431f, 1:623–624
 actin organization
 contraction, 1:29–31, 1:30f, 1:31f
 protrusion, 1:28–29, 1:29f
 cellular functions, 3:222–223
 cytoskeletal walking model, 1:632–633
 disease connections, 3:223–224, 3:223t
 domain structures, 3:218, 3:219f, 3:220f
 EF-hand proteins, 1:319t
 evolutionary diversity, 3:218–219, 3:220f
 mechano-chemical characteristics, 3:219–221, 3:221f
 microvilli, 1:35
 myosin heavy chain (MHC) proteins, 1:185
 myosin II, 1:623–624, 2:238, 4:214
 myosin light chain kinase (MLCK)
 calcium ion (Ca²⁺)–calmodulin (CaM)-dependent protein kinases (CaMKs), 1:305–306, 1:305f
 cell migration, 1:439f, 1:440–441, 1:441f
 EF-hand proteins, 1:319t, 1:321
 focal adhesions, 2:322, 2:322f
 G₁₂/G₁₃ family functions, 2:345f
 somatostatin (sst) receptors, 4:265
 nonmuscle myosin, 1:33–34, 1:33f
 phylogenetic classification, 3:218–219, 3:220f
 regulation mechanisms, 3:221–222
 smooth muscle myosin light chain kinase (smMLCK), 1:32–33, 1:305–306, 1:305f
 stereocilia, 1:35
 tropomyosin, 1:31–32, 1:32f, 1:34–35, 1:34f, 1:18–19, 1:22f
 myostatin, 1:182
 myotonic dystrophy, 1:477t
 N-myristoylation
 c-fes proto-oncogene (Fes), 1:452–453
 fatty acyl-coenzyme A (CoA) synthetases, 2:291f
 lipid-modified proteins, 2:736, 2:737f, 2:737t
 N-myristoyltransferase (Nmt)
 lipid-modified proteins, 2:736, 2:737f, 2:737t
 prevalence, 3:642–643
 reaction mechanisms, 3:642–643
 structural characteristics, 3:643, 3:643f
 substrates, 3:642f
 therapeutic uses, 3:643–644
 Myxopronin, 4:171, 4:171f
 myxothiazol, 1:597f, 1:598
- N**
- N-(2,4-diamino-6-oxo-5H-pyrimidin-5-yl)formamide, 2:1, 2:2f
 N-acetyl cysteine, 2:333
 N-acetyl-glutamic acid (NAG), 4:495f
 N-acetylneuraminic acid (Neu5Ac), 4:230–231
 Nael enzyme, 2:133
 nails, 2:669t
 Na-K-Cl cotransporter isoform 1 (NKCC1), 1:199, 1:200f
 nano-electrospray ionization (nanoESI), 3:11–12, 3:12f
 Nanog, 2:438
 naphthoquinones, 3:722–723, 3:727–728, 4:553–554, 4:553f
 napsin, 1:138f
 National Center for Biotechnology Information (NCBI), 3:611–612
 native protein complexes, 3:11–16
 characteristics, 3:11
 mass spectrometry
 collision-induced dissociation (CID), 3:12, 3:12f, 3:13–14, 3:14f
 complex structure, 3:13–14
 dynamic protein complexes, 3:13
 future research areas, 3:15
 gas-phase complex structure, 3:14–15
 GroEL (growth E locus, large gene, bacterial group I chaperonin), 3:14–15, 3:15f
 nano-electrospray ionization (nanoESI), 3:11–12, 3:12f
 protein complex composition, 3:12–13, 3:13f
 research methodologies, 3:11–12
 structural biology analyses, 3:11
 natriuretic peptides, 3:225–228
 atrial natriuretic peptide (ANP), 1:101t, 3:225, 3:226f
 B-type (brain) natriuretic peptide (BNP), 3:226, 3:226f
 characteristics and functional role, 3:225
 C-type natriuretic peptide (CNP), 3:226, 3:226f
 natriuretic peptide receptors
 characteristics, 3:226–227
 guanylyl-cyclase-linked receptors, 3:226f, 3:227
 natriuretic peptide clearance receptor, 3:226f, 3:227
 therapeutic uses, 3:227
Natronomonas, 3:32–34
Natronomonas pharaonis, 4:92–93
 natural killer (NK) cells
 hematopoiesis, 1:187–188, 1:188f
 liver regeneration, 1:190–191, 1:191f, 1:194
 tumor necrosis factor receptors (TNFRs), 4:458
 nausea
 dopamine receptors, 2:173
 neurokinin-1 receptor (NK1R), 4:364
 Naxos disease, 1:638t
 NBS1 protein, 3:272, 3:273t
 Nck protein, 2:266
 NCLX protein superfamily, 3:125
 NCX/NCKX expression, 1:380f, 1:381–382, 1:382f, 3:45–46, 3:45f, 3:46f
 near infrared (NIR) spectroscopy
 brain research
 brain functional activation, 4:279–280, 4:279f
 metabolic activity, 4:278, 4:278f
 neuronal signals, 4:280
 cancer detection, 4:279
 historical research, 4:277–278
 multidimensional imaging, 4:278
 muscle studies, 4:278f, 4:279
 photon migration, 4:279
 nebulin, 1:31–32, 1:32f
 Necdin, 3:243, 3:244f
 necrosis, 1:423–425
 calcium ion (Ca²⁺)
 brain ischemia, 1:423–424
 excitotoxicity, 1:423–424, 1:425f
 neurodegenerative diseases, 1:424–425
 neuronal death, 1:424
 signal executors, 1:423
 signaling mechanisms, 1:423
 signaling subroutines, 1:425
 characteristics, 1:423
 nectarin, 4:353
 nectins, 1:255
 Nedd4 protein, 1:279–280
 negative elongation factor (NELF), 4:165, 4:165f
 negative transcription elongation factors (N-TEFs), 4:164, 4:165, 4:165f
 Nei repair pathway, 2:57–58, 2:57f
Neisseria, 2:43
Neisseria gonorrhoeae, 2:137t, 2:139–140
Neisseria meningitidis
 Cho phylogenetic tree, 3:331f

- Dam methylation, 2:43
 type III DNA restriction and modification enzymes, 2:139–140
uvrC phylogenetic tree, 3:331f
 nematode sperm motility, 3:1, 3:2f
Nematostella vectensis, 2:14
 neoglycoproteins, 3:229–232
 applications
 carbohydrate receptor analyses, 3:231
 enzyme substrates, 3:231
 glycoprotein structure–function relationship analyses, 3:231
 isolation affinants, 3:231
 targeting devices, 3:231
 preparation methods
 glycopeptide conjugation, 3:231
 glycosyl transferases, 3:231
 oligosaccharide–protein reductive amination, 3:229, 3:230f
 protein thiol groups, 3:230–231
 synthetic carbohydrate derivatives
 aromatic glycosides, 3:229, 3:230f
 general discussion, 3:229–230
 nonaromatic glycosides, 3:229, 3:230f
 reduced and periodate-oxidized natural oligosaccharides, 3:229–230
 neomycin B, 1:2f, 1:328
 neonatal Fc receptor (FcRn), 2:570f, 2:572
 neonates
 neonatal diabetes, 1:653
 neonatal severe primary hyperparathyroidism (NSHPT), 1:326
 urea cycle disorders
 arginase deficiency (hyperargininemia), 4:500
 argininosuccinate lyase (AL) deficiency (argininosuccinic aciduria), 4:499
 carbamyl phosphate synthetase (CPS)/*N*-acetylglutamate (NAG) synthetase deficiencies, 4:497–498
 citrullinemia
 argininosuccinate synthetase (AS) deficiency, 4:498–499
 citrin deficiency, 4:499
 neonatal intrahepatic cholestasis, 4:499
 hyperammonemia
 clinical manifestations, 4:494
 diagnosis, 4:494–495, 4:496f
 inborn errors of metabolism, 4:495t
 long-term therapy considerations, 4:497
 treatments, 4:495–497, 4:497t
 ornithine transcarbamylase (OTC) deficiency, 4:498
 transient hyperammonemia, 4:500
 neopterin, 3:666–667, 3:667f
 neostriatum, 2:169
 neosynephrine, 1:58
 neovascularization, 1:453
 neoxanthin
 light harvesting process, 2:179f, 2:726–727
 molecular structure, 1:502f
 nephronectin (NN), 1:427t
 nerve growth factor (NGF), 3:241, 3:242f, 3:244f
 nervous tissue, 1:286
 Nesiritide, 3:227
 nesprin, 2:625–627
 nestin, 1:196–197, 1:197f, 2:627–628, 2:628t, 2:632t, 2:634, 3:236
N-ethylmaleimide (NEM), 3:164
N-ethylmaleimide-sensitive factor (NSF), 1:3t, 1:5, 3:37
 neural crest, 2:345t
 neural precursor cell expressed, developmentally downregulated 8 (NEDD8), 4:468t, 4:470, 4:475
 neural tube defects (NTDs), 2:327
 neurocan, 3:657t, 3:658
 neuroendocrine-specific protein of 55 kDa (NESP55), 2:517, 2:518f
 neurofilament proteins, 2:625–627
 neurofilament triplet proteins (NFTPs), 2:632t, 2:633f, 2:634, 3:233–235, 3:234f
 neurogenerative disorders, 1:477t
 neurogenesis, 1:196–200
 basic concepts, 1:196
 dentate gyrus, 1:200f
 influential factors
 extrinsic factors
 basic concepts, 1:199
 cytokines, 1:198–199
 growth factors, 1:198–199
 neurotransmitters, 1:199
 neurotrophins, 1:198–199
 schematic diagram, 1:200f
 intrinsic factors
 cell-autonomous signaling complexes, 1:199
 cell-type specific factors, 1:200
 epigenetic modulators, 1:199–200, 1:200f
 scaffold proteins, 1:197f, 1:199
 schematic diagram, 1:200f
 signaling pathways, 1:200
 transcriptional regulation, 1:199–200, 1:200f
 research summary, 1:200–201
 stages, 1:196–197, 1:197f
 neurogenic muscle weakness, ataxia, retinitis pigmentosa (NARP), 1:211f
 neuroglian, 3:628f
 neurokinin-1 receptor (NK1R)
 carboxyl tail, 4:363–364, 4:363f
 disease connections
 affective disorders, 4:364–365
 depression, 4:364–365
 emesis, 4:364
 glioblastomas, 4:365
 research background, 4:364–365, 4:364f
 full-length NK1R, 4:361–363, 4:362f
 functional role, 4:361, 4:362f
 truncated NK1R, 4:363
 neurokinin A
 functional role, 4:360–361
 research background, 4:364f
 synthesis, 4:361f
 neurokinin B
 functional role, 4:361
 research background, 4:361f
 synthesis, 4:361f
 neurologic abnormalities, 4:494
 neuronal calbindin, 1:286
 neuronal intermediate filaments, 3:233–237
 α -internexin, 3:235
 characteristics, 3:233
 cytoskeleton composition, 1:430
 disease connections, 3:236
 meiosis, 3:17
 nestin, 3:236
 neurofilament triplet proteins (NFTPs), 3:233–235, 3:234f
 peripherin, 3:236
 neurons
 cerebral metabolism, 1:369, 1:370f, 1:371–373
 G₁₂/G₁₃ family functions, 2:345t
 intracellular calcium ion (Ca²⁺) waves, 2:642–643
 neuron-specific neuronal marker (NeuN), 1:196–197, 1:197f
 neuropathy target esterase (NTE), 4:271
 neuropeptide B (NPB), 4:263, 4:264f
 neuropeptide W (NPW), 4:263, 4:264f
 neuropeptide Y (NPY)
 appetite regulation, 4:303
 body weight regulation, 3:77, 3:77f, 3:78f
 insulin metabolism, 2:606
 leptin expression, 2:257–258
 peptide amidation, 3:396
 starvation effects, 4:303
Neurospora, 1:600, 1:600f
Neurospora crassa
 actin-related proteins (Arps), 1:37f
 free oligosaccharides (fOSs), 2:331
 mitochondrial functions, 3:132
 peroxisomal matrix protein import, 3:426–429
 plasma membrane proton (H⁺)-ATPases, 3:688
 respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:88
 tRNA synthetases, 4:443
 Woronin body, 3:425
 neurosteroids, 2:423, 2:723–724
 neurotensin, 2:497t
 neurotoxins, 4:568–569, 4:568t
 neurotransmitter transporters, 3:238–240
 adenylyl cyclases (ACs), 1:46t
 characteristics, 3:238
 disease connections, 3:239–240
 G_q family proteins, 2:496–497, 2:497t
 regulation mechanisms, 3:239
 structure and functions, 3:239
 transport mechanisms
 reuptake and electrogenicity, 3:238–239, 3:239f
 reverse transport, 3:239
 neurotrophins, 3:241–246
 brain-derived neurotrophic factor (BDNF)
 body weight regulation, 3:77–78, 3:78f
 calcium ion (Ca²⁺) signaling, 3:295
 functional role, 3:241, 3:242f, 3:244f
 neurogenesis, 1:197f, 1:198
 cross-talk, 3:246

- neurotrophins (*Continued*)
 functional role, 3:241
 nerve growth factor (NGF), 3:241, 3:242f, 3:244f
 neurogenesis, 1:198–199
 neurotrophin 3 (NT3), 1:287, 3:241, 3:242f, 3:244f
 neurotrophin 4 (NT4), 3:241, 3:242f, 3:244f
 neurotrophin receptor-interacting factor (NRIF), 3:243, 3:244f, 3:245f
 neurotrophin receptor-interacting MAGE homolog (NRAGE), 3:243, 3:244f, 3:245f
 p75 neurotrophin receptor (p75NTR), 3:241, 3:243–246, 3:244f, 3:245f, 4:459
 signaling pathways, 3:241–243, 3:242f
 tumor necrosis factor receptors (TNFRs), 4:454
 tyrosine kinase receptors (Trks)
 cross-talk, 3:246
 functional role, 3:241
 p75 neurotrophin receptor (p75NTR), 3:246
 signaling pathways, 3:241–243, 3:242f
 neutral ceramidase (nCDase), 4:288–289
 neutral lipid storage disease with myopathy (NLSMD), 1:49
 neutral sphingomyelinase (nSMase), 4:287–288
 neutral zwitterionic glycerolipids, 2:416–417, 2:416f
 neutrophils
 chemotaxis, 2:236–240
 characteristics and functional role, 2:236
 cytoskeletal polarization, 2:238
 mathematical models, 2:238–240, 2:238f
 metrics, 2:239b
 signaling pathways, 2:236–238, 2:237f
 immunoglobulin Fc receptors (FcRs), 2:569–571
 nevirapine, 2:549
 newts, 1:552f
 N-formyl-methionyl-tRNA (fMet-tRNA)
 background information, 4:425
 initiation complex formation, 4:426–427, 4:427f
 translation initiation region (TIR), 4:425–426
 N-glycolylneuraminic acid (Neu5Gc), 4:230–231
 NgoAXP enzyme, 2:137t
 N-hydroxysuccinimide ester (NHS), 1:227, 1:227f
 niacin
See nicotinamide adenine dinucleotide (NAD⁺/NADH)
 nicastrin, 4:53
 nickel (Ni)
 nickel-iron (NiFe) hydrogenases
 biomass-produced renewable hydrogen, 4:72–73, 4:73f, 4:74f
 catalytic cycle, 2:563–565, 2:563f
 structural characteristics, 2:562–563, 2:563f
 nickel-iron-selenium (NiFeSe)
 hydrogenases, 2:562–563
Nicotiana plumbaginifolia, 3:688
 nicotinamide adenine dinucleotide (NAD⁺/NADH), 1:172–178
 aging process functions, 1:177
 anoxygenic photosynthesis, 3:693–694, 3:694f
 ascorbic acid, 4:532–533
 bioenergetic reactions, 1:486–487, 1:487f, 1:488f
 calcium ion (Ca²⁺) signaling, 1:331–332
 cardiac excitation–contraction (E–C) coupling, 1:381
 cardiac metabolism, 3:65–66, 3:66f, 3:67, 3:68f
 chemolithotrophs, 1:491
 citric acid cycle, 4:93f
 coenzyme Q, 3:724–725, 3:724f, 4:18–20, 4:19f
 dark fermentation, 4:74f
de novo biosynthetic pathways, 1:172, 1:174f, 1:175f
 dioxygenases, 3:371–372
 disease treatment applications, 1:177
 DNA ligases
 covalent ligase–AMP covalent intermediate formation, 2:28, 2:29f, 2:33–35, 2:35f, 2:38f
 domain structure and organization, 2:33, 2:34f
 functional role, 2:33, 2:34f
 research background, 2:28
 external NADH dehydrogenase, 4:85
 fatty acid oxidation, 2:281–282, 2:282f, 3:65–66, 3:66f, 3:68f
 gluconeogenesis, 2:382f, 2:384
 glycolysis
 adenosine triphosphate (ATP)
 production, 2:443
 anaerobic glycolysis, 2:444–445, 2:445f
 glycolytic pathway, 2:445f
 oxidation effects, 2:445–446, 2:446f
 schematic diagram, 2:444f
 green bacteria, 2:508–509, 2:508f
 historical research, 1:172
 mitochondria
 homeostasis control mechanisms, 3:124f
 mitochondrial β -oxidation, 1:386, 1:386–387
 mitochondrial calcium (Ca²⁺) transport, 3:118, 3:119f
 mitochondrial DNA (mtDNA), 3:133f, 3:134t
 regulation mechanisms, 3:169–170
 NAD-dependent isocitric dehydrogenase (NAD-IDH), 3:124f
 nicotinamide adenine dinucleotide dehydrogenase
 anoxygenic photosynthesis, 3:693–694, 3:694f
 biochemical free radical chain reactions, 3:169–170, 3:171f, 3:172f
 dysfunctional mitochondria, 3:173–174
 electron transfer complexes, 3:696
 plant mitochondrial electron transport chain, 1:145, 1:146f
 quinone pool functions, 3:695f
 respiratory electron transport chain, 3:661–662, 3:662f, 3:663–664, 3:664f, 3:694–695, 3:694f, 4:79, 4:80f
 nicotinamide adenine dinucleotide transhydrogenase, 3:247–252
 coupling mechanism, 3:251
 helix packing, 3:250
 hydride transfer mechanisms, 3:249–250, 3:249f
 membrane topology, 3:250, 3:250f
 physiological functions, 3:251–252
 proton pumping/proton channel, 3:250–251, 3:250f
 structure and functions, 3:247, 3:248f
 substrate-binding domains, 3:247–249, 3:248f
 non-redox functions, 1:176–177, 1:177f
 peroxidases, 3:411–412, 3:411f
PET genes, 3:308f
 phosphatidic acid synthesis, 2:415, 2:416f
 photosynthetic metabolic pathways, 3:498f
 plant mitochondrial electron transport chain, 1:141–142, 1:142f, 1:145, 1:146f
 Preiss–Handler pathway, 1:172, 1:174f
 protonmotive electron transfer chain, 1:470–471, 1:470f
 pyridine nucleotide oxidoreductases, 2:311
 pyruvate dehydrogenases (PDHs)
 carbohydrate-responsive element binding protein (ChREBP), 1:377
 components, 4:321–322
 mitochondrial calcium (Ca²⁺) transport, 3:124f
 reaction mechanisms, 4:323
 regulation mechanisms, 4:324
 quinone pool functions, 3:695f
 redox metabolism, 1:172–175, 1:173f
 redox potentials, 1:218–219
 respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:87–91
 background information, 4:98
 bacterial reductive reactions, 1:488f
 basic concepts, 4:87, 4:341
 disease connections
 general discussion, 4:90–91
 mitochondrial mutations, 4:90–91
 neurotoxins, 4:91
 nuclear mutations, 4:91
 pathological mechanisms, 4:91
 evolutionary development
 general discussion, 4:88–90
 N-module, 4:89
 origins, 4:89–90
 P-module, 4:89
 Q-module, 4:88–89
 variants, 4:89–90
 functional role
 active–inactive transition, 4:90
 electron transfer mechanisms, 3:661–662, 3:662f, 3:663, 4:83–84, 4:90

- general discussion, 4:90
 inhibitors, 4:90
 proton pumping, 3:661–662, 3:662f, 3:663, 4:90
 redox centers, 4:90
 sodium ion (Na⁺) pumping, 4:90
 mitochondrial ROS, 4:18–20, 4:19f
PET genes, 3:308f
 plant mitochondrial electron transport chain, 1:145, 1:146f
 protonmotive electron transfer chain, 1:470f
 reverse electron transport (RET), 4:19f, 4:20
 schematic diagram, 4:84f, 4:99f
 structural characteristics, 4:341–346
 accessory subunits, 4:88, 4:89t
 central subunits, 4:87–88, 4:88f, 4:89t
 cysteine motif, 4:344–345, 4:345f
 flavin mononucleotide (FMN)
 binding, 4:344–345, 4:344f
 frataxin-like subunit, 4:343–344, 4:343f
 hydrophilic domain, 4:342–343, 4:342f
 iron–sulfur clusters, 4:87–88, 4:89t, 4:342–343, 4:342f, 4:345f
 multistage assembly pathway, 4:88
 mutations, 4:345
 NADH binding, 4:344–345, 4:344f, 4:345f
 overall structure, 4:341–342
 redox centers, 4:87–88, 4:89t, 4:342–343, 4:342f
 research summary, 4:345–346
 shape, 4:87, 4:88f
 subunits, 4:342t
 supernumerary subunits, 4:345
 respiratory electron transport chain, 3:661–662, 3:662f, 3:663–664, 3:664f, 3:694–695, 3:694f, 4:79, 4:80f
 Rossmann fold, 1:173, 1:176f
 salvage pathways, 1:172, 1:174f
 spectrophotometric assays
 fluorescence, 4:277
 hemoglobin (Hb) studies, 4:277
 near infrared (NIR) spectroscopy, 4:277–278
 structural characteristics, 1:172, 1:173f
 tightly-bound prosthetic groups, 1:175–176
 tricarboxylic acid cycle, 4:438–440, 4:439f, 4:439t
 nicotinamide adenine dinucleotide phosphate (NADP⁺/NADPH)
 adiponectin, 3:74
 brassinosteroid biosynthesis, 1:251, 1:252f
 calcium ion (Ca²⁺) signaling, 1:323–324, 1:324f, 1:331–332
 carbohydrate-responsive element binding protein (ChREBP), 1:376–377, 1:377f
 cerebral metabolism, 1:369f, 1:371–373
 chlorophyll, 3:544
 chloroplasts, 1:508, 1:511–512, 1:512f
 cobalamin-dependent methionine synthase, 2:327
 cyanobacterial respiration, 4:81, 4:82f
 cytidine nucleotide formation, 3:715f
 cytochrome *b₆* complex, 4:335, 4:336f, 4:338–339, 4:339f
 cytochrome P-450 (CYP), 1:608–609, 1:609f
 de novo biosynthetic pathways, 1:172, 1:174f, 1:175f
 dioxigenases, 3:371–372
 fatty acid oxidation, 2:282–283, 2:283f
 fatty acid synthesis pathway, 2:276–279, 2:276f
 fermentation reactions, 2:204–206
 ferredoxin-NADP⁺ reductase (FNR), 2:299–302
 catalytic mechanisms
 ferredoxin (Fd) interactions, 2:301, 2:301f
 flavin adenine dinucleotide (FAD/FADH₂), 2:299–301, 2:300f
 NADP(H) interactions, 2:300, 2:300f
 cytochrome *b₆* complex, 4:338–339, 4:339f
 heterotrophic glutathione reductase (GR)-type FNRs, 2:301
 molecular structure, 2:299, 2:300f
 nonphotosynthetic plant-type FNRs
 bacterial/apicomplexan FNRs, 2:301
 root FNRs, 2:301
 photosynthetic processes, 2:296, 2:297f
 physiological functions, 2:296–297, 2:299, 2:300f
 research background, 2:299–301
 subcellular localization, 2:299
 folate metabolic cycle, 2:325f
 green sulfur bacteria, 2:511f
 historical research, 1:172
 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase regulated degradation, 1:519
 hyperglycemia, 1:656
 insulin-secreting cells (β-cells), 2:598–600
 light harvesting proteins, 1:511–512, 1:512f, 2:178–179, 2:179f
 mitochondrial DNA (mtDNA), 1:210, 1:212f
 NADPH oxidase 4 (NOX4), 3:110
 nicotinamide adenine dinucleotide transhydrogenase, 3:247–252
 coupling mechanism, 3:251
 helix packing, 3:250
 hydride transfer mechanisms, 3:249–250, 3:249f
 membrane topology, 3:250, 3:250f
 physiological functions, 3:251–252
 proton pumping/proton channel, 3:250–251, 3:250f
 structure and functions, 3:247, 3:248f
 substrate-binding domains, 3:247–249, 3:248f
 nitric oxide synthase (NOS)
 calcium ion (Ca²⁺)–calmodulin (CaM) regulation, 1:343f, 1:344t, 1:345f
 C-terminal extension, 1:344–346
 electron transfer mechanisms, 1:342
 metabolic reactions, 3:668, 3:668f
 nitric oxide signaling, 3:253–254
 non-redox functions, 1:176–177, 1:177f
 oxygenic photosynthesis, 2:335
 pentose phosphate pathway, 3:391–395
 basic concepts, 3:392f
 historical research, 3:391
 operation modes
 hexose monophosphate shunt, 3:391
 metabolic reactions, 3:392
 regulation mechanisms, 3:391–392
 quantitation
 carbon fate, 3:391, 3:394f
 chemical balance, 3:391, 3:394f
 clinical importance, 3:394
 estimations, 3:392–394
 methodologies, 3:392
 reaction mechanisms
 nonoxidative segments, 3:391, 3:393f
 oxidative segments, 3:391, 3:392f
 peroxidases, 3:411–412, 3:411f
 phosphatidic acid synthesis, 2:415, 2:416f
 photosynthetic carbon dioxide (CO₂) fixation, 3:497–502
 C4 pathways, 3:498
 Calvin cycle enzyme targets, 3:500–501
 carbon pathway, 3:497, 3:499f
 carboxylation phase, 3:497–498, 3:499f
 Crassulacean acid metabolism (CAM) pathway, 3:498
 enzymatic regulation
 flux control, 3:500
 irreversible enzymes, 3:500
 feedback regulation, 3:501
 genetic modification effects, 3:500–501
 light-regulated energy and power, 3:497, 3:498f
 metabolic regulation, 3:499–500
 reduction phase, 3:498, 3:499f
 regeneration phase, 3:499, 3:499f
 whole plants, 3:501
 photosynthetic processes, 3:493–494, 3:493f, 3:494f
 phototrophic respiration, 4:79, 4:80f
 phox homology (PX) domain, 3:448–449
 plant mitochondrial electron transport chain, 1:141–142, 1:142f, 1:145, 1:146f
 Preiss–Handler pathway, 1:172, 1:174f
 protonmotive electron transfer chain, 1:470f
 protoporphyrinogen IX, 2:541
 pyridine nucleotide oxidoreductases, 2:311
 reactive oxygen species (ROS), 2:77, 4:18–20
 redox metabolism, 1:172–175, 1:173f
 redox potentials, 1:218–219
 Rossmann fold, 1:173, 1:176f
 salvage pathways, 1:172, 1:174f
 structural characteristics, 1:172, 1:173f
 thymidine nucleotide formation, 3:715f
 tightly-bound prosthetic groups, 1:175–176
 tricarboxylic acid cycle, 4:438–440, 4:439f, 4:439t
 vitamin E metabolism, 4:549, 4:549f

- nicotinamide adenine dinucleotide phosphate (NADP⁺/NADPH) (*Continued*)
 vitamin K metabolic pathways, 4:553–554, 4:553f
- nicotinamide mononucleotide adenylyl transferase enzymes (NMNAT), 1:174f
- nicotinamide mononucleotide (NMN), 1:172, 1:174f, 2:33, 2:34f
- nicotinamide riboside kinases (NRKs), 1:174f
- nicotinic acid adenine dinucleotide (NaAD), 1:172, 1:174f
- nicotinic acid adenine dinucleotide phosphate (NAADP)
 calcium ion (Ca²⁺) release, 2:637–639
 invertebrates, 2:637–638, 2:638t
 metabolic pathways, 2:638–639
 vertebrates, 2:638, 2:638t
 calcium ion (Ca²⁺) signaling, 1:331–336
 basic concepts, 1:331–332
 calcium ion (Ca²⁺) stores, 1:332
 fertilization-induced signaling, 1:333–334, 1:333f
 research summary, 1:336
 sea urchin egg fertilization, 1:333–334, 1:333f
 starfish egg fertilization, 1:334–335, 1:334f, 1:335f
 two-pore channels, 1:335–336
 structural characteristics, 2:637, 2:638f
- nicotinic acid mononucleotide (NaMN), 1:172, 1:174f
- nicotinic acid phosphoribosyltransferase enzyme (NAPRT), 1:174f
- nidogen, 3:91t
- Niemann–Pick disease, 4:291t
 lysosomal dysfunction, 4:507–508, 4:507t
 tau proteins, 1:477t
- niguldipine, 1:58
- Nijmegen breakage syndrome (NBS), 1:415–416
- 9-cis-retinoic acid (9cRA), 3:418–419, 4:525
- 19S regulatory complex
 assembly processes, 3:598
 base complex, 3:597f, 3:598
 basic concepts, 3:590–592
 lid complex, 3:597f, 3:598
 regulatory mechanisms
 characteristics, 3:598–599
 core particle (CP) gate opening, 3:599
 deubiquitylation, 3:598
 substrate recognition, 3:598
 translocation and unfolding processes, 3:599
 structural characteristics, 3:597–598, 3:597f
- N-iodosuccinimide (NIS), 1:4f
- NIPA1, 4:269f, 4:270
- Nitrobacter winogradskyi*, 1:489t
- nitrogen (N)
 insect biochemistry, 2:591–592, 2:592f
 nitrate reductase, 1:487t, 3:403–404, 3:403f, 3:695f
 nitrate respiration, 2:206, 2:207f, 2:208f
 nitric oxide (NO)
 adiponectin, 3:74
 amino acid metabolism, 1:93t, 3:668, 3:668f
 biochemical free radical chain reactions, 3:169f
 biological functions, 3:253, 3:254f
 chemical characteristics and reactions, 3:254–255, 3:255f
 functional role, 4:17
 glutamate receptors, 2:720
 mitochondrial ATP-dependent
 potassium ion (K⁺) channel (KATP), 3:551
 mitochondrial production, 3:170–172, 3:171f, 3:172f, 3:173f
 mitochondrial ROS, 4:21
 mitogen-activated protein kinase (MAPK), 3:173
 nitric oxide reductase, 4:98
 nitric oxide signaling, 3:253–256
 catalytic reactions, 3:254f
 chemical characteristics and reactions, 3:254–255, 3:255f
 components, 3:253
 guanylate cyclase (sGC), 3:255
 non-cyclic guanosine monophosphate (cGMP) signaling pathways, 3:255
 nitric oxide synthase (NOS), 1:342–346
 calcium ion (Ca²⁺)–calmodulin (CaM) regulation, 1:343–346, 1:344t, 1:345f, 3:253–254, 3:525–526
 calmodulin (CaM)–nitric oxide synthase (NOS) interactions, 1:305f, 1:342–343
 characteristics and functional role, 1:342
 cytokine-inducible nitric oxide synthase (iNOS), 1:342–343, 1:344t, 3:253–254
 electron transfer mechanisms, 1:342, 1:343f
 endothelial nitric oxide synthase (eNOS), 1:342–343, 1:344t, 1:345f, 2:345f
 functional role, 4:17, 4:18f
 glutamate receptors, 2:720
 metabolic reactions, 3:668, 3:668f
 mitochondrial production, 3:170–172, 3:171f, 3:173f
 neuronal nitric oxide synthase (nNOS), 1:342–343, 1:344t
 plasma membrane calcium (Ca⁺)-ATPase pump, 3:525–526
 ryanodine receptors (RyRs), 4:197
 nitric oxide synthetase, 3:304t
 physiological functions, 3:172–173
 nitrogenases, 2:663, 4:73
 nitrogen fixation (NIF) system, 2:663
 nitroreductase, 2:313
 nitrous oxide reductase, 1:487t
 peroxynitrite, 2:77
 purple bacteria electron acceptors and donors, 3:698, 3:698t
 reactive nitrogen species (RNS), 4:17–22
 characteristics, 4:18
 functional role, 4:17, 4:18f
 mitochondrial ROS, 4:21
 regulation mechanisms, 2:77
 research summary, 4:21
 ryanodine receptors (RyRs), 4:197
 urinary nitrogen excretion, 4:306, 4:306f
- Nitrosomonas europaea*, 1:488f, 1:489t
- Nitrosopumilus maritimus*, 2:158t, 2:161
- Nkx2.5* gene, 1:183–185, 1:184f
- N-linked glycans
 α -glucosidases, 3:257
 endoplasmic reticulum (ER)
 congenital disorders, 2:470, 2:471t
 free oligosaccharide (fOS) generation, 2:329, 2:330f
 glycosylation processes, 2:457, 2:458f
 protein synthesis and folding, 2:200–201, 4:204–205, 4:205f
 free oligosaccharides (fOSs), 2:329–331
 generation mechanisms, 2:329, 2:330f
 glycoprotein-based folding processes, 2:329–330
 Golgi-modified free oligosaccharides (fOSs), 2:331
 research summary, 2:331
 species specificity, 2:331
 transport mechanisms, 2:329
 functional role, 2:329
 Golgi apparatus, 4:207
 mammalian glucose transport, 2:392–393
 Man₅GlcNAc₂ glycan, 2:329, 2:330, 2:330f, 3:257–259
 plant glycoproteins, 2:465, 2:466f
 processing pathways
 class I α -mannosidases
 endoplasmic reticulum (ER)
 α -mannosidase I, 2:201, 3:258, 3:258f, 3:259t
 ER degradation enhancing
 a-mannosidase-like protein (EDEM), 2:201, 3:259
 Golgi α -mannosidase I, 3:258, 3:258f, 3:259t
 class II α -mannosidases, 3:259
 endo α -mannosidase, 3:257, 3:258f
 Glc₃Man₉GlcNAc₂ glycan, 3:257, 3:258f, 3:259t
 α -glucosidase I/ α -glucosidase II, 3:257, 3:258f, 3:259t
 Golgi α -mannosidase II, 3:258f, 3:259, 3:259t, 4:504f
 Man₅GlcNAc₂ glycan, 3:257–259, 3:259t
 Man₈GlcNAc₂ glycan, 3:258f, 3:259t
 Man₉GlcNAc₂ glycan, 3:259t
 recognition molecules, 3:259–260, 3:259t
 research summary, 3:260
 schematic diagram, 3:258f
- N-linked glycoproteins, 2:457–464
 analytical analyses
 endoglycosidase/exoglycosidase digestion, 2:459f, 2:461
 mutation studies, 2:462
 physicochemical analyses, 2:462
 plant lectin binding, 2:460–461
 radiolabeling, 2:461

- tunicamycin (TN) inhibitors, 2:458f, 2:461–462
- biosynthesis
- dolichol phosphate (Dol-P), 2:457, 2:458f
 - glycosidase digestion, 2:457–458, 2:459f
 - transfer processes, 2:457
- as experimental aids
- glycoprotein purification, 2:462
 - radiolabeling and tagging, 2:462
 - secretory pathway monitoring, 2:462
- flippase proteins
- dolichyl pyrophosphate (Dol-P-P), 2:315–316, 2:315f, 2:316t
 - flippase-mediated transbilayer movement, 2:314, 2:315f
 - glucosylphosphoryldolichol (Glc-P-Dol), 2:315–316, 2:315f, 2:316t
 - glycosylation, 2:315–316
 - mannosylphosphoryldolichol (Man-P-Dol), 2:315–316, 2:315f, 2:316t
- functional roles
- calnexin (CNX), 2:463
 - calreticulin (CRT), 2:463
 - cognate lectins, 2:463–464
 - free oligosaccharide (fOS) generation, 2:329
 - high-mannose-bearing mannose-6-phosphate residues, 2:461f, 2:463–464
 - monoglucosylated oligosaccharides, 2:463
 - plant glycoproteins, 2:465, 2:466f
 - structural remodeling, 2:457, 2:459f, 2:460f, 2:461f
- N-methyl-D-aspartate (NMDA)
- calcium ion (Ca^{2+}) signaling, 2:720f, 3:292, 3:295, 3:295f
 - chemokine receptors, 1:484–485
 - excitotoxicity, 1:423–424, 1:425f
 - extrasynaptic NMDA receptors, 3:295, 3:295f
 - functional role, 2:715t
 - glutamate receptors, 2:714–716, 2:715t
 - glycine receptors (GlyRs), 2:419
 - intracellular calcium ion (Ca^{2+}) waves, 2:644–646
 - metabotropic glutamate (mGlu) receptors, 2:397
 - molecular structure, 2:716t, 2:717f
 - N-methyl-D-aspartate (NMDA) receptors, 2:419
 - proline-rich tyrosine kinase 2 (PYK2), 2:267
 - protein structure, 2:718–719
 - serine–threonine phosphatases, 4:222f, 4:223
- N-myristoyltransferase (Nmt)
- prevalence, 3:642–643
 - reaction mechanisms, 3:642–643
 - structural characteristics, 3:643, 3:643f
 - substrates, 3:642f
 - therapeutic uses, 3:643–644
- Nobel Prize, 2:254
- Nocardia otitidis*, 2:132t
- NOD-like receptors (NLRs), 4:396
- Noggin
- bone morphogenetic proteins (BMPs), 1:229–230
 - neurogenesis, 1:197f, 1:198–199
- Nogo, 3:243, 3:244f
- anoladin ether, 2:195f
- nonaromatic glycosides, 3:229, 3:230f
- noncoated endocytosis, 2:198f, 2:199
- nonheme peroxidases, 3:409–411, 3:410f
- non-Herlitz JEB/generalized atrophic benign epidermolysis bllosa (GABEB), 1:641t
- non-Herlitz junctional epidermolysis bullosa (JEB), 1:641t
- nonhistone chromatin proteins (NHCPs), 1:531–532
- non-homologous end joining (NHEJ), 3:266–268
- bacterial NHEJ repair pathway
- action mechanisms, 3:267–268, 3:267f
 - characteristics, 3:266
 - functional role, 3:268
 - Ku protein, 3:266–267
 - ligase D (LigD), 3:267
 - viral infections, 3:268
- basic concepts, 3:269–271, 3:270f
- chromosome organization, 3:316
- DNA repair processes
- cell-cycle-dependent repair, 3:266, 3:269
 - DNA ligases, 2:30t, 2:32, 3:266, 3:267f
 - double-strand breaks (DSBs), 3:266, 4:29, 4:45
 - enzyme requirements, 3:266, 3:267f
 - Ku protein, 3:266–267, 3:267f, 4:31–32
 - mismatch repair systems (MRs), 2:45
 - nucleases, 4:34
 - yeast Rad1/Rad10 nucleases, 4:32
- enzymatic functions, 3:269–271, 3:270f
- eukaryotic NHEJ repair pathway, 3:269–274
- accessory proteins, 3:272
 - basic concepts, 3:269–271, 3:270f
 - disease connections, 3:273, 3:273t
 - DNA-dependent protein kinase catalytic subunit (DNA-PK_{cs}), 3:269–271, 3:270f, 3:272f
 - enzymatic functions, 3:269–271, 3:270f
 - inhibitors, 3:273–274
 - Ku70/80 protein, 3:269–271, 3:270f, 3:272f, 3:273t
 - telomere maintenance, 3:272, 4:384–385
 - therapeutic interventions, 3:273–274
 - V(D)J recombination, 3:272, 3:272f
 - functional role, 3:268
 - oxidative damage repair, 2:79–80
 - viral infections, 3:268
- non-insulin-dependent diabetes mellitus (NIDDM), 1:360
- nonketotic hyperosmolar coma, 1:654–655
- nonmuscle cell contraction, 1:33–34, 1:33f
- non-nucleoside inhibitor-binding pocket (NNIBP), 2:552f
- non-nucleoside reverse transcriptase inhibitors (NNRTIs), 2:549, 2:551, 2:552f
- nonrod/noncone photoreceptors, 3:490, 3:490f
- nontranscriptional Smad-dependent responses, 1:233
- norepinephrine/noradrenaline
- ascorbic acid, 4:531–532
 - characteristics and functional role, 1:57
 - G_q family proteins, 2:497t
 - transport mechanisms, 3:238–239
- Nostoc* sp. PCC 7120, 4:335–338, 4:337f, 4:339–340, 4:339t
- Notch
- glycogen synthase kinase-3 (GSK3), 2:436t, 2:438
 - hairly-related transcription factor (Hrt) family, 1:185
 - regulated intramembrane proteolysis (RIP), 4:51f, 4:53
 - Sex-lethal* (*Sxl*) master-switch gene, 1:82
- NOTCH-dependent transcription
- hypoxia inducible factor (HIF), 4:68
 - liver regeneration, 1:192f
 - O-linked glycans, 2:455
- Novosphingobium aromaticivorans*, 3:115, 3:561
- Nrf2 system, 1:657, 3:411f
- Nt.BstNBI enzyme, 2:131, 2:132t
- N-terminals
- calcitonin peptide family, 1:270–271
 - calcium-sensing receptor (CaSR), 1:326–328, 1:327f
 - calmodulin (CaM)–nitric oxide synthase (NOS) interactions, 1:342–343
 - calpains, 1:354–359
 - c-fes* proto-oncogene (Fes), 1:452–453
 - chemokine receptors
 - ligand binding, 1:480–482, 1:482f
 - subfamily classifications, 1:480 - ferredoxin (Fd)
 - ferredoxin-NADP⁺ reductase (FNR) structure, 2:299, 2:300f
 - ferredoxin structure, 2:296, 2:297f
 - iron–sulfur clusters (FeS), 4:95 - fibroblast growth factors (FGFs), 2:303
 - GABA_A receptors, 2:347
 - gelsolin, 1:277–278, 1:277f
 - glycosylphosphatidylinositol (GPI)-anchored proteins, 2:476–477
 - histone H1, 1:531
 - integrin signaling, 2:608
 - major sperm protein (MSP), 3:1–2
 - microtubule-associated proteins (MAPs), 3:104–105, 3:105f
 - MutS pathway, 2:55
 - myristoylation, 1:452–453, 2:736
- N-terminal domain
- AAA (ATPases associated with various cellular activities) proteins, 1:1–2, 1:2f
 - actin depolymerizing factor (ADF)/cofilin (ADF-H) domain, 1:23–24
 - adenosine diphosphate ribosylation factors (ARF), 1:134
 - alternative oxidases (AOXs), 1:146–147, 1:147f
 - alternative splicing, 1:76

N-terminals (*Continued*)

- arrestins, 2:489–490, 2:489f
- ATP synthase, 2:271–272, 2:273f
- bullous pemphigoid antigen 1 (BPAG1), 2:626f, 2:628–629
- calcitonin receptors (CTRs), 1:272
- c-fes* proto-oncogene (*Fes*)
 - characteristics, 1:451, 1:452f
 - Fes*/*Fer*/CIP4 homology (*FCH*) domain, 1:451
- copper-transporting ATPases (copper pumps), 3:672–673, 3:672f, 3:674
- cyclic nucleotide-dependent protein kinases, 1:576f, 1:577–578
- cytoplasmic dynein, 2:185–186, 2:186f
- degradosome complex, 3:53
- diaphanous inhibitory domain (*DID*), 1:24
- DNA gyrase, 2:164–165, 2:165f
- DNA ligases, 2:28–30, 2:29f, 2:30–31, 2:33, 2:34f
- DNA-strand transferases, 4:37–38, 4:37f
- dynamitin, 2:176
- elongation cycle, 2:253
- 5'-deoxyribose-5-phosphate lyase, 2:5, 2:6f
- GABA_A receptor biochemistry, 2:724t, 2:725
- GABA_A receptors, 2:348f, 2:349
- GABA_B receptors, 2:351–352
- glucokinase glucose sensors, 2:601f
- glycine receptors (*GlyRs*), 2:421, 2:422f
- glycogen synthase kinase-3 (*GSK3*), 2:434–436, 2:435f
- glycoprotein hormone receptors, 4:388, 4:389f
- hexameric enzyme helicases, 2:23f
- initiation factors
 - initiation factor 1 (*IF1*), 4:427, 4:428f
 - initiation factor 2 (*IF2*), 4:428–430, 4:428f
 - initiation factor 3 (*IF3*), 4:428f, 4:430
- inositol requiring 1 (*IRE1*) genes, 4:489, 4:490f
- inositol trisphosphate (*IP*₃), 2:586, 2:587f, 2:653–655, 2:654f, 2:655f
- intermediate filaments, 3:233
- iron–sulfur clusters (*FeS*), 4:95
- keratins, 2:667
- kinesins, 2:673f
- kinesin superfamily proteins (*KIFs*), 2:680
- Kyprides–Onzonis–Woese (*KOW*) motif, 4:169–170
- LacI protein, 2:697–698, 2:697f
- LexA regulatory protein, 2:707–708, 2:707f
- metabotropic glutamate (*mGlu*) receptors, 2:396, 2:396f
- mitochondrial DNA (*mtDNA*), 2:7
- mTOR complexes, 3:196f
- MutL family proteins, 2:68–69, 2:69f
- nuclear-encoded PSII proteins, 3:508
- nuclear factor kappa B (*NF-κB*), 3:302
- ORAI proteins, 3:362–363, 3:363f, 3:365f
- parathyroid hormone (*PTH*)/*PTH*-related protein (*PTHrP*) receptor (*PPR*), 3:386–387
- phosphatidic acid phosphatases (*PAPs*), 1:661f, 1:662
- phosphodiesterases (*PDEs*), 1:581, 1:581f
- photosystem I (*PSI*) complex, 3:503–505, 3:504f
- plasma membrane calcium (*Ca*⁺)-ATPase pump, 3:525–526, 3:526f
- plasma membrane proton (*H*⁺)-ATPases, 3:688–689, 3:690–691
- platelet-derived growth factor (*PDGF*), 3:538, 3:539f
- plectin, 2:625, 2:626f
- polymerase α -primase (*Pol* α), 2:241–242
- polymerase δ (*Pol* δ), 2:245, 2:246f
- prion domains, 3:566–567, 3:568f
- protein kinase B (*PKB*), 3:463–464
- PsaC protein, 2:340, 2:340f
- P-type adenosine triphosphatase (*ATPase*), 4:330–331, 4:330f
- replication focus targeting sequence (*RFTS*), 3:8–9, 3:8f
- reverse gyrase mechanism, 2:159–160
- RNA polymerases, 4:174f, 4:175–176, 4:177f
- ryanodine receptors (*RyRs*), 4:196
- secreted mucins, 3:202–203
- serine recombinases, 1:558–559
- sliding clamps, 3:576, 4:237–238
- small guanosine phosphatases (*GTPases*), 4:242, 4:244f
- store-operated *Ca*²⁺ entry (*SOCE*), 3:364, 3:365f
- stromal-interacting molecule (*STIM*) protein family, 4:315, 4:316f
- tubulin, 4:450–451, 4:451f
- type IB DNA topoisomerase, 2:160, 2:160f
- type I DNA restriction and modification enzymes, 2:72–73, 2:73f
- type II DNA topoisomerase, 2:164–165, 2:165f
- type XVII collagen, 1:641
- uncoupling proteins (*UCPs*), 4:483–484, 4:485f
- vascular endothelial growth factor receptors (*VEGFRs*), 4:510, 4:510f
- voltage-gated sodium channels, 4:564, 4:565f
- N-terminal fragment (*NTFs*), 4:202
- N-terminal nucleophile (*Ntn*) hydrolase superfamily, 3:596, 3:596f
- olfactory receptors, 3:345–346
- palmitoylation, 2:736–737
- pheromone receptors, 3:441–442
- presenilins, 4:202
- protein N-myristoylation, 3:642–643, 3:643f
- protein tyrosine phosphatase (*PTP*), 3:651f
- respiratory chain complex I (*NADH*:ubiquinone oxidoreductase), 4:88
- retinoic acid receptors, 4:109
- Smad signaling, 1:232–233
- Src (sarcoma) family kinases, 4:295–297, 4:296f
- steroid/thyroid hormone receptors, 4:309
- toll-like receptors (*TLRs*), 4:397–398
- 20S proteasome, 3:596, 3:596f
- vasopressin/oxytocin family, 4:517–519
- vitamin D receptor (*VDR*), 4:540–541, 4:542f
- nuclear envelope
 - post-mitosis assembly
 - characteristics, 3:311–312
 - nuclear pore complex positioning, 3:312
 - nuclear shape and rigidity, 3:312
 - nuclear size, 3:312
 - structural characteristics, 3:296, 3:310, 3:311f
- nuclear export signals (*NEs*), 3:318–320, 4:7, 4:8f
- nuclear factor (erythroid-derived 2)-like 2 (*Nrf2*), 1:657, 3:411f
- nuclear factor kappa B (*NF-κB*), 3:302–305, 4:376–379
 - apoptotic pathways, 1:389f
 - arrestins, 2:488t
 - background and characteristics, 3:302
 - biological functions, 3:303–304, 3:304t
 - calcium oscillations, 1:323–324
 - disease connections, 3:304–305, 3:304t
 - gene expression regulation, 1:311
 - glutathione peroxidases, 3:411f
 - glycogen synthase kinase-3 (*GSK3*), 2:436t, 2:438
 - G protein-coupled receptor kinases (*GRKs*), 2:485t
 - hematopoiesis, 1:188–189
 - I κ B kinase (*IKK*), 1:313, 3:303, 3:304f, 3:305, 4:456, 4:458f
 - inhibitors, 3:305t
 - mucins, 3:206–208, 3:206f, 3:208f
 - neurokinin-1 receptor (*NK1R*), 4:362–363
 - neurotrophins, 3:241–243, 3:242f
 - nuclear calcium, 1:313
 - nuclear factor kappa B (*NF-κB*)-inducing kinase (*NIK*), 4:456, 4:458f
 - pharmacological regulations, 3:305, 3:305t
 - platelet-activating factor receptor (*PAFR*), 3:535, 3:535f
 - protein characteristics and structure, 3:302, 3:303f
 - protein–DNA interactions, 2:151
 - RAGE (receptor for advanced glycation end products), 2:407–409
 - RANK/RANKL receptor proteins, 4:459
 - regulatory mechanisms, 3:302–303, 3:303t, 3:304f
 - signaling pathways, 1:618f
 - signal transduction, 3:303, 3:304f
 - target genes, 3:303–304, 3:304t
 - T-cell receptor (*TCR*) signaling, 4:376–379
 - antigen receptors, 4:376, 4:377f
 - CARD11 (caspase-recruitment domain containing protein 11), 4:377–378, 4:377f
 - chromosomal translocations, 4:379
 - downstream activators, 4:377–378

- I κ B kinase (IKK) complex activation, 4:378–379
 immunological synapses, 4:376
 mucosa-associated lymphoid tissue 1 (MALT1), 4:378–379
 NF- κ B activation, 4:379
 pathway dysregulation, 4:379
 receptor–proximal signaling, 4:376, 4:377f
 second messengers, 4:377
 TGF activated kinase 1 (TAK1), 4:378–379
 TNF receptor associated-factor 6 (TRAF6), 4:378–379
 zeta-associated protein-70 (ZAP-70), 4:374
 toll-like receptors (TLRs), 4:398–399
 transforming growth factor- β -activated kinase (TAK)1, 4:398, 4:399f, 4:400f
 nuclear factor of activated T cells (NFAT)
 calcineurin-mediated signal transduction, 1:312, 1:338–339, 1:339–340, 1:340f
 calcium ion (Ca²⁺) signaling, 1:312, 1:313f, 2:587–588, 2:719–720, 2:720f, 3:293, 3:293f, 3:294, 3:359–360, 3:360f
 calcium oscillations, 1:323–324
 Cyp-D protein, 3:162–163
 glycogen synthase kinase-3 (GSK3), 2:435f, 2:436t
 serine–threonine phosphatases, 4:222f, 4:223
 T-cell receptors (TCRs), 4:374
 nuclear inhibitor of protein phosphatase 1 (NIPP-1), 1:312, 4:222
 nuclear lamina, 3:310–313
 composition, 3:310
 disease connections
 Emery–Dreifuss muscular dystrophy, 3:313
 Hutchinson–Gilford progeria syndrome, 3:313, 3:313f
 laminopathies, 2:635f, 2:636, 3:313
 DNA replication regulation, 3:312
 functional role, 3:311
 gene expression regulation, 3:312
 LacI/*lacO* system, 2:751–752, 2:752f
 lamin-associated proteins, 3:311
 lamin genes, 3:310
 lamins, 2:632t, 2:633f, 2:634, 3:296, 3:297f, 3:310, 3:311f
 live cell imaging, 2:751–752, 2:752f
 post-mitosis assembly
 characteristics, 3:311–312
 nuclear pore complex positioning, 3:312
 nuclear shape and rigidity, 3:312
 nuclear size, 3:312
 postranslational modification
 isoprenylation, 3:312
 phosphorylation, 3:311f, 3:312
 proteolysis, 3:312–313
 structural characteristics, 3:310, 3:311f
 nuclear localization signal (NLS), 2:29f, 2:30, 3:318–320, 4:7, 4:8f
 nuclear magnetic resonance (NMR) spectroscopy, 2:661–662, 3:11
 nuclear organization, 3:314–317
 chromatin structure, 2:749, 3:314–315, 3:315f
 chromosome organization, 3:315–316, 3:315f
 disease impacts, 3:316
 gene expression, 3:315–316, 3:316f
 live cell imaging, 2:749–755
 chromatin interaction analysis by paired-end tag sequencing (ChIA-PET), 2:750–751
 chromosome conformation capture (3C), 2:750–751
 DNA adenine methylation identification (Dam-ID), 2:749, 2:750
 future research outlook, 2:752–754, 2:753f
 high-throughput chromosome capture (Hi-C) technology, 2:750–751
 LacI/*lacO* system
 chromosomal domains, 2:751, 2:751f
 forced domain repositioning, 2:751–752, 2:752f
 future research outlook, 2:752–754, 2:753f
 inner nuclear membrane (INM), 2:751–752, 2:752f
 nuclear dynamics, 2:749
 nuclear zip codes, 2:752–754, 2:753f
 three-dimensional fluorescent *in situ* hybridization (3D-FISH) analysis, 2:749–751
 nuclear envelope, 3:314, 3:315f
 reporter genes, 3:315–316, 3:316f
 research summary, 3:316
 schematic diagram, 3:315f
 nuclear structure, 3:296–301
 Cajal (coiled) bodies, 3:299–300, 3:300f
 chromosome territories, 3:297, 3:297f
 compartmentalization, 3:296
 histone locus bodies, 3:300–301
 live cell imaging, 2:749–755
 nuclear envelope
 post-mitosis assembly
 characteristics, 3:311–312
 nuclear pore complex positioning, 3:312
 nuclear shape and rigidity, 3:312
 nuclear size, 3:312
 structural characteristics, 3:296, 3:310, 3:311f
 nuclear periphery
 lamina/lamin proteins, 3:296, 3:297f
 nuclear envelope, 3:296
 peripheral heterochromatic compartment, 3:296–297, 3:297f
 nuclear pore complexes (NPCs), 3:318–323
 characteristics, 3:318
 nucleocytoplasmic transport, 4:7, 4:8f
 nucleoporin proteins (Nups), 3:318, 3:319f, 3:320t
 post-mitosis assembly, 3:312
 protein trafficking
 characteristics, 3:318–321
 classical NLS-mediated import, 3:321, 3:322f
 import/export cycles, 3:320–321, 3:321f
 nuclear localization signal (NLS), 3:318–320, 4:7, 4:8f
 small GTPase Ran cycle, 3:320–321, 3:321f, 4:7, 4:8f
 targeting signals, 3:318–320
 transport receptors, 3:320
 RNA trafficking
 characteristics, 3:321–323
 messenger RNA (mRNA), 3:323
 microRNA (miRNA), 3:322
 ribosomal RNA (rRNA), 3:322–323
 transfer RNA (tRNA), 3:322
 U small nuclear RNA (snRNA), 3:322
 structural characteristics, 3:318, 3:319f
 transport mechanisms
 protein trafficking, 3:318–321
 RNA trafficking, 3:321–323
 nucleolus, 3:324–327
 amplified nucleoli, 3:324
 clinical significance, 3:324–325
 dense fibrillar component (DFC), 3:324, 3:325f
 fibrillar centers, 3:324, 3:325f
 functional organization, 3:324, 3:325f
 granular component (GC), 3:324, 3:325f
 historical research, 3:324
 missing nucleoli, 3:324
 perinucleolar compartment (PNC), 3:324–325
 phase-contrast micrograph, 3:325f
 plurifunctional nucleolus, 3:325
 proteomic studies, 3:325, 3:326f
 ribosome biogenesis, 4:120f
 signal-recognition particle (SRP), 3:325
 structural characteristics, 3:297–298, 3:298f, 3:324
 paraspeckles, 3:298, 3:299f
 promyelocytic leukemia nuclear bodies (ND 10), 3:298–299, 3:299f
 research summary, 3:301
 speckles/splicing factor compartments, 3:298, 3:298f, 3:299f
 survival of motor neuron (SMN) gems, 3:300–301, 3:300f
 transcription factories, 3:297
 nucleases
 cleavage processes, 4:31
 deoxyribonucleic acid (DNA), 4:34
 D-loop dissociation, 4:32–33
 enzyme translocation, 4:31t
 recombination models, 4:30f
 research summary, 4:34
 nucleic acids
 DNA microarrays
 basic concepts, 2:370–371
 data analysis, 2:371
 schematic diagram, 2:371f
 exosomes and microvesicles (EMVs), 2:262–263
 hexameric enzyme action, 2:21–27
 classifications, 2:21, 2:22t

- nucleic acids (*Continued*)
- double-strand DNA (dsDNA)
 - translocation, 2:24–25
 - loaders/loading process, 2:21–22, 2:22*t*
 - nucleoside triphosphatase (NTPase)-powered translocation, 2:22–24, 2:24*f*
 - research summary, 2:26
 - single-strand DNA (ssDNA) interactions, 2:21
 - structural characteristics, 2:21, 2:23*f*
 - unwinding mechanisms, 2:25–26, 2:25*f*
- nucleobindin, 1:319*t*
- nucleoid-associated proteins (NAPs), 3:367–370
- characteristics and functional role, 3:367, 3:368*f*, 3:368*t*
 - DNA supercoiling, 3:368–369, 3:369*f*
 - gene expression regulation, 3:367–368
 - topoisomerases, 3:369–370
- nucleolus, 3:324–327
- amplified nucleoli, 3:324
 - clinical significance, 3:324–325
 - dense fibrillar component (DFC), 3:324, 3:325*f*
 - fibrillar centers, 3:324, 3:325*f*
 - functional organization, 3:324, 3:325*f*
 - granular component (GC), 3:324, 3:325*f*
 - historical research, 3:324
 - missing nucleoli, 3:324
 - perinucleolar compartment (PNC), 3:324–325
 - phase-contrast micrograph, 3:325*f*
 - plurifunctional nucleolus, 3:325
 - proteomic studies, 3:325, 3:326*f*
 - ribosome biogenesis, 4:120*f*
 - signal-recognition particle (SRP), 3:325
 - structural characteristics, 3:297–298, 3:298*f*, 3:324
- nucleoside analog reverse transcriptase inhibitors (NRTIs), 2:550–551, 2:550*f*
- nucleoside diphosphate (NDP), 1:73–74, 2:22–24
- nucleoside triphosphate (NTP)
- nucleoside triphosphatase (NTPase), 2:22–24, 2:24*f*
 - RNA polymerase II (Pol II), 2:366–367
 - T7 RNA polymerase, 4:355, 4:356*f*, 4:357*f*
- nucleosomes
- chromatin structure, 1:521, 1:526*f*, 1:528–529, 1:535–537, 1:536*f*, 3:314–315, 3:315*f*
 - GAL1–10 gene expression regulation, 1:526*f*, 1:528–529
 - nucleosome positioning, 1:525
 - regulation mechanisms
 - ATP-dependent nucleosome remodeling
 - chromodomain helicase DNA binding (CHD)/Mi-2 family, 4:60*f*, 4:64
 - INO80 (inositol requiring 80) family, 4:60*f*, 4:64–65, 4:65*f*
 - ISWI family, 4:60*f*, 4:64
 - remodeling enzymes, 4:63–65
 - schematic diagram, 4:65*f*
 - SWItching defective/Sucrose NonFermenting (SWI/SNF)
 - family, 1:314, 4:60*f*, 4:63–64, 4:403–404
 - Swr1-dependent deposition of H2A. Z–H2B dimers, 4:64–65, 4:65*f*
 - histone modifications
 - acetylation, 1:521, 1:522*f*, 4:60
 - citrullination, 4:62
 - covalent modifications, 1:521, 1:522*f*, 4:59
 - histone variants, 4:60*f*, 4:62–63
 - methylation, 1:521, 1:522*f*, 4:60–61
 - phosphorylation, 1:314–315, 1:521, 1:522*f*, 4:61
 - proteolysis, 4:62
 - sumoylation, 4:62
 - ubiquitylation, 4:61–62
 - remodeling mechanisms, 4:66
 - research summary, 4:66
 - structural characteristics, 4:59
- nucleotide addition cycle (NAC), 2:367, 4:170–171, 4:181*f*, 4:183
- nucleotide-binding domain (NBD)
- multidrug resistance (MDR) membrane proteins
 - clinical relevance, 3:210–211
 - functional role, 3:213
 - high affinity–substrate relationships, 3:213
 - polyspecific transport cycle, 3:211–213, 3:212*f*
 - research background, 3:210
 - structural characteristics, 3:211–213, 3:211*f*
- nucleotide excision repair (NER), 3:337–340
- alkylation damage, 2:14, 2:15*f*
 - background information, 3:328
 - basic concepts, 4:402
 - B-DNA (antiparallel right-handed double helix), 2:142
 - biological importance, 3:341
 - defective NER
 - disease connections, 3:339–340
 - mouse models, 3:340
 - DNA ligases, 2:30*t*
- eukaryotes, 3:341–344
- biological importance, 3:341
 - damage recognition, 3:341–342, 3:342*t*
 - damage verification, 3:342–343, 3:342*t*
 - global genome repair (GGR), 3:341–342, 3:342*f*
 - incision and synthesis processes, 3:342*t*, 3:343–344, 3:343*f*
 - NER factors, 3:342*t*
 - open complex formation, 3:342–343, 3:342*t*
 - repair pathways, 3:342*f*
 - transcription-coupled repair (TCR)
 - pathway, 3:337, 3:342*f*, 3:342*t*, 3:344
- eukaryotic genomes, 3:337
- global genome repair (GGR)
 - damage recognition mechanisms, 4:402
 - schematic diagram, 4:403*f*
 - kinetic mechanisms, 3:338–339
 - oligonucleotide fragments, 3:337
 - operational principles, 3:337, 3:338*f*
- operation modes, 3:337
- oxidative damage repair, 2:79–80
- repair mechanisms, 3:338*f*
- repair pathways, 3:329*f*
- repair synthesis detection, 3:338, 3:339*f*
- substrate specificity, 3:338, 3:339*f*
- transcription-coupled repair (TCR)
 - pathway
 - basic concepts, 4:402–403
 - damage recognition mechanisms, 4:402
 - lesion-related transcription arrest
 - cisplatin-induced intrastrand cross-links, 4:405
 - cyclobutane pyrimidine dimers (CPDs), 4:404–405, 4:404*t*
 - general discussion, 4:404–405
 - oxidative damage, 4:405
 - research perspective, 4:405–406
 - RNA polymerase II (Pol II)-based arrest and initiation, 4:403–404, 4:404*t*
 - schematic diagram, 4:403*f*
- ultraviolet (UV) damage, 4:572
- UvrABCD system, 3:328–336
- bacteria–human conservation model, 3:336
 - damage recognition mechanisms
 - general discussion, 3:333–334
 - padlock model, 3:334–335
 - UvrA₂–DNA complex, 3:333
 - UvrA₂B₂ complex in solution, 3:333
 - UvrA₂–DNA binding, 3:333
 - UvrA dimerization, 3:333
 - UvrB–DNA intermediate complex formation, 3:333–334
 - damage sensing mechanisms
 - basic concepts, 3:334–335, 3:334*f*
 - DNA polymerase I, 3:335
 - dual incision step, 3:334
 - LexA regulatory protein, 2:709–710
 - postincision steps, 3:334
 - UvrB–DNA complex, 3:334–335, 3:334*f*
 - UvrD, 3:334
 - DNA mismatch repair (MMR), 2:45–46
 - protein properties
 - UvrA, 3:330–332, 3:332*f*
 - UvrB, 3:332, 3:332*f*
 - UvrC, 3:332, 3:333*f*
 - substrate ranges, 3:330*t*
 - transcription-coupled repair (TCR), 3:335–336, 3:335*f*
 - UvrABCD endonuclease, 3:328, 3:329*f*, 3:330*t*
- uvr* genes
- Cho, 3:330, 3:331*f*
 - historical research, 3:328–330
 - uvrA*, 3:328, 3:329*f*
 - uvrB*, 3:329*f*, 3:330
 - uvrC*, 3:329*f*, 3:330, 3:331*f*
- nucleotides
- guanine-nucleotide exchange factors (GEFs), 2:344, 2:345*f*, 2:494, 4:439, 4:111–112
 - nucleotide insertion/deletion editing, 4:157
 - phage libraries, 3:432–433, 3:433*t*, 3:434*t*

nucleotidyltransferase (NTase), 2:33, 2:34f, 2:35f, 2:36f, 2:37f, 2:38f
 nucleus accumbens, 2:172t
 Nun protein, 4:410–411
 NusA/NusG elongation factors, 4:168f, 4:169–170, 4:409–410

O

Obelia geniculata, 2:577
 obelin, 2:577, 2:577f
 obesity
 antiobesity drugs, 3:27
 branched-chain amino acid (BCAA) regulation, 1:245–246
 leptin deficiency, 2:256, 2:257f
 leptin resistance, 2:260–261, 2:260f
 starvation, 4:298
ob/ob mice intercross studies, 2:256, 2:257f
 occludin, 4:392, 4:394f
 Oct4, 2:438
 octanoate, 1:245–246
 [OCT+] prion
 See prions
 octreotide, 4:263t, 4:264
 octyl glucoside, 1:645–646, 1:645f
 ocular cicatricial pemphigoid, 1:641t
 ocular surfaces, 3:203
Oenococcus oeni, 2:208–209
 okadaic acid, 1:150t
 Okazaki fragments
 DNA mismatch repair (MMR), 2:67f, 2:70–71
 DNA replication
 DNA ligases, 2:31
 lagging-strand DNA synthesis, 2:116–117
 leading- and lagging-strand DNA synthesis, 2:92, 2:93f, 2:94f, 2:114, 2:115f, 2:118, 2:119f
 polymerase δ (Pol δ), 2:246–247, 2:247f
 proliferating cell nuclear antigen (PCNA), 2:119
 polymerase α -primase (Pol α), 2:119, 2:241, 2:242–243
 oleic acid, 2:288
 oleosins, 1:47
 oleyl CoA, 3:551
 olfactory system
 cyclic nucleotide-gated (CNG) ion channels, 1:586–587
 dopamine receptors, 2:172t
 olfactory receptors, 3:345–349
 basic concepts, 3:345
 classifications, 3:345–346
 combinatorial versus monospecific olfactory processing, 3:347f, 3:348–349, 3:348f
 functional role, 3:345
 gene expression regulation, 3:347, 3:347f
 gene structure/genomic arrangement, 3:346
 information processing pathway, 3:347f
 ligand spectra, 3:347–348

monogenic and monoallelic expression, 3:346–347, 3:347f
 schematic diagram, 3:346f
 signal transduction, 3:349, 3:349f
 oligodendrocytes
 cerebral metabolism, 1:369, 1:370f, 1:371–373
 oligodendrocyte myelin glycoprotein (OMgp), 3:243, 3:244f
 oligomeric protein assembly, 3:629–630, 3:630f
 oligomycin, 1:143t
 oligonucleotides
 DNA microarrays
 basic concepts, 2:370–371
 data analysis, 2:371
 schematic diagram, 2:371f
 messenger RNA (mRNA), 3:56
 oligonucleotide-binding (OB) domain, 2:33, 2:34f, 2:35f, 2:36f, 2:37f, 2:38f
 phage libraries, 3:432–433, 3:433t, 3:434t
 quadruplex DNA (G4-DNA), 2:146
 triplex-forming oligonucleotides (TFOs), 2:143–144
 oligosaccharides
 carbohydrate chain synthesis, 1:362–367
 automated synthesis
 general discussion, 1:5–6
 programmable one-pot synthesis, 1:5–6, 1:5f
 solid-phase synthesis, 1:4–5, 1:5f
 challenges, 1:1
 chemical synthesis
 automated synthesis, 1:5–6
 basic concepts, 1:4–6, 1:4f
 bond-formation control, 1:3–4, 1:4f
 solution-phase synthesis, 1:4, 1:5f
 stereochemical control, 1:3–4, 1:4f
 sugar hydroxyl protection, 1:4
 enzymatic synthesis
 basic concepts, 1:2–3
 glycosyltransferase-catalyzed reactions, 1:2f, 1:3
 glycosidase-catalyzed reactions, 1:2, 1:2f
 glycosyltransferase-catalyzed reactions
 basic concepts, 1:3
 nucleotide-sugar recycling, 1:3
 phosphorylation effects, 1:3
 product inhibition, 1:3
 recombinant expression, 1:3
 schematic diagram, 1:2f
 molecular structure, 1:2f
 free oligosaccharides (fOSs), 2:329–331
 generation mechanisms, 2:329, 2:330f
 glycoprotein-based folding processes, 2:329–330
 Golgi-modified free oligosaccharides (fOSs), 2:331
 research summary, 2:331
 species specificity, 2:331
 transport mechanisms, 2:329
 mucins, 3:206–208
 N-linked glycans
 Glc₃Man₉GlcNAc₂ glycan, 3:257, 3:258f
 Man₅GlcNAc₂ glycan, 3:257–259, 3:259t

Man₈GlcNAc₂ glycan, 3:258f, 3:259t
 Man₉GlcNAc₂ glycan, 3:259t
 processing pathways, 3:257–260
 class I α -mannosidases, 3:258
 class II α -mannosidases, 3:259
 endo α -mannosidase, 3:257, 3:258f
 endoplasmic reticulum (ER) α -mannosidase I, 3:258, 3:258f, 3:259t
 ER degradation enhancing α -mannosidase-like protein (EDEM), 3:259t
 ER degradation enhancing α -mannosidase-like protein (EDEM), 3:259
 Glc₃Man₉GlcNAc₂ glycan, 3:257, 3:258f
 α -glucosidase I/ α -glucosidase II, 3:257, 3:258f
 Golgi α -mannosidase I, 3:258, 3:258f, 3:259t
 Golgi α -mannosidase II, 3:258f, 3:259
 Man₅GlcNAc₂ glycan, 3:257–259, 3:259t
 Man₈GlcNAc₂ glycan, 3:258f, 3:259t
 Man₉GlcNAc₂ glycan, 3:259t
 recognition molecules, 3:259–260, 3:259t
 research summary, 3:260
 schematic diagram, 3:258f
 N-linked glycoproteins, 2:457–464
 analytical analyses
 endoglycosidase/exoglycosidase digestion, 2:459f, 2:461
 mutation studies, 2:462
 physicochemical analyses, 2:462
 plant lectin binding, 2:460–461
 radiolabeling, 2:461
 tunicamycin (TN) inhibitors, 2:458f, 2:461–462
 biosynthesis
 dolichol phosphate (Dol-P), 2:457, 2:458f
 glycosidase digestion, 2:457–458, 2:459f
 transfer processes, 2:457
 as experimental aids
 glycoprotein purification, 2:462
 radiolabeling and tagging, 2:462
 secretory pathway monitoring, 2:462
 flippase proteins
 dolichyl pyrophosphate (Dol-P-P), 2:315–316, 2:315f, 2:316t
 flippase-mediated transbilayer movement, 2:314, 2:315f
 glucosylphosphoryldolichol (Glc-P-Dol), 2:315–316, 2:315f, 2:316t
 glycosylation, 2:315–316
 mannosylphosphoryldolichol (Man-P-Dol), 2:315–316, 2:315f, 2:316t
 functional roles
 calnexin (CNX), 2:463
 calreticulin (CRT), 2:463
 cognate lectins, 2:463–464

- oligosaccharides (*Continued*)
- high-mannose-bearing mannose-6-phosphate residues, 2:461*f*, 2:463–464
 - monoglucosylated oligosaccharides, 2:463
 - N*-glycosylation process, 2:451–452, 2:451*f*, 2:465, 2:466*f*, 2:470, 2:470*f*
 - plant glycoproteins, 2:465, 2:466*f*
 - structural remodeling, 2:457, 2:459*f*, 2:460*f*, 2:461*f*
 - oligosaccharide-binding (OB) domain, 2:33, 2:34*f*, 2:35*f*, 2:36*f*, 2:37*f*, 2:38*f*
 - oligosaccharide–protein reductive amination, 3:229, 3:230*f*
 - oligosaccharide/proton symporter family, 2:388
 - oligosaccharyl transferase (OST), 2:329, 2:457
 - reduced and periodate-oxidized natural oligosaccharides, 3:229–230
 - RNA polymerases, 3:342*f*
 - O-linked glycans, 2:453–456
 - characteristics, 2:465, 2:466*f*
 - classifications, 2:453
 - O-fucose
 - carbohydrate modifications, 2:453, 2:454*f*
 - classifications, 2:453
 - epidermal growth factor (EGF)-like repeats, 2:453, 2:454*f*
 - notch signaling, 2:455
 - protein O-fucosyltransferase I (Pofut1), 2:453, 2:454*f*, 2:456
 - thrombospondin type 1 repeats (TSRs), 2:454, 2:455*f*, 2:456
 - uPA–uPA receptor signaling, 2:455
 - functional role
 - background information, 2:455–456
 - α -dystroglycan function, 2:454*f*, 2:456
 - notch signaling, 2:455
 - uPA–uPA receptor signaling, 2:455
 - O-glucose
 - carbohydrate modifications, 2:453–454, 2:454*f*
 - classifications, 2:453
 - epidermal growth factor (EGF)-like repeats, 2:453–454, 2:454*f*
 - notch signaling, 2:455
 - O-mannose
 - carbohydrate modifications, 2:454, 2:454*f*
 - classifications, 2:453
 - α -dystroglycan function, 2:454*f*, 2:456
 - mucins, 3:206–208
 - synthetic pathways, 2:465–466
 - O-xylose, 2:453
 - omega-6/omega-3 fatty acids, 2:288
 - omega (ω)-oxidation, 2:291*f*
 - omodysplasia, 3:657*t*
 - 1-methyladenine (1mA), 2:9, 2:10*t*
 - onion, 4:533*t*
 - Op18, 3:105
 - OPA1 protein, 3:137, 3:147, 4:51*f*, 4:54–55
 - open reading frames (ORFs)
 - bacterial translation elongation, 4:421–423
 - extra-chromosomally-primed (EP) retrotransposons, 3:284–286, 3:286*f*, 3:287*f*
 - mitochondrial gene expression, 3:141*f*, 3:142*f*, 3:143*f*
 - pre-messenger RNAs (pre-mRNAs), 3:60–61, 3:143*f*
 - trp* operon, 2:360–362, 2:361*f*
 - opioid receptors, 3:354–358
 - characteristics, 3:354
 - future research areas, 3:357–358
 - historical research, 3:354
 - multiple opioid receptors
 - distribution and pharmacology, 3:355–356, 3:355*t*
 - historical research, 3:354–355
 - molecular mechanisms, 3:357
 - opioid ligand selectivity, 3:356–357, 3:356*t*
 - research summary, 3:357–358
 - second messenger pathways, 1:43*f*
 - opsin
 - characteristics, 3:488
 - color vision, 1:550–551
 - retinoid metabolic pathways, 4:527*f*
 - Optic atrophy 1 (Opa1) protein, 3:137, 3:147, 4:51*f*, 4:54–55
 - ORAI protein family, 3:359–366
 - arachidonic acid regulated Ca^{2+} (ARC) channels, 1:130, 1:131*f*, 1:132*f*
 - calcium ion (Ca^{2+}) release-activated Ca^{2+} channels (CRAC), 3:359–360, 3:360*f*, 3:360*t*, 3:364–365, 3:365*f*
 - calcium oscillations, 1:323
 - disease connections, 3:365
 - ORAI1 channel
 - biophysical properties, 3:360*t*
 - cytoplasmic domains, 3:362, 3:362*f*
 - ion conduction pathways, 3:363–364
 - membrane topology and assembly, 3:362–363, 3:363*f*
 - store-operated calcium (SOC) channels, 3:359–360, 3:360*f*
 - transmembrane domains, 3:362–363, 3:363*f*
 - ORAI2 channel, 3:364
 - ORAI3 channel, 3:364
 - ORAI proteins
 - functional role, 3:362–364
 - ion conduction pathways, 3:363–364
 - isoforms, 3:364
 - membrane topology and assembly, 3:362–363, 3:363*f*
 - transmembrane domains, 3:362–363, 3:363*f*
 - store-operated calcium (SOC) channels
 - disease connections, 3:365
 - functional role, 3:360–364, 4:315
 - ORAI tetramers, 4:317, 4:318*f*
 - signaling pathways, 3:364, 3:365*f*
 - signal modulation, 3:364–365
 - oral cavity, 2:354
 - oral pemphigoid, 1:641*t*
 - oral white sponge nevus, 2:670*t*
 - orange carotenoid protein (OCP), 3:486
 - oranges, 4:533*t*
 - Orc1 protein, 2:112
 - Orc6 protein, 2:112–113
 - Oregon green, 2:575*t*, 2:576–577
 - organellar retrotransposons, 3:288
 - organotrophs, 1:488*t*
 - oriC* region, 2:42, 2:108, 2:117, 2:121
 - origin recognition complex (ORC)
 - background information, 2:108
 - conservation, 2:111
 - discovery, 2:111
 - functional role, 2:111–112
 - non-replicative functions, 2:112–113
 - pre-recombination complex formation, 2:111–112, 2:112*f*
 - regulation mechanisms, 2:112
 - specification mechanisms, 2:110
 - ornithine
 - D-ornithine 4,5-aminomutase, 1:158*t*
 - ornithine carbamoyltransferase (OCT), 1:94–95
 - ornithine decarboxylase inhibition, 2:214*f*
 - ornithine transcarbamylase (OTC), 4:494, 4:495*f*, 4:495*t*, 4:496*f*, 4:498
 - retinal and choroidal gyrate atrophy, 4:500
 - urea cycle, 4:495*f*
 - orogressive subcortical gliosis, 1:477*t*
 - orogressive supranuclear palsy, 1:477*t*
 - orotate phosphoribosyltransferase (OPRTase), 3:712, 3:713*f*, 3:714*f*
 - orotidine decarboxylase (ODCase), 3:712, 3:713*f*, 3:714*f*
 - orphan nuclear receptors, 4:309–311
 - Oryza sativa*
 - actin-related proteins (Arps), 1:37*f*
 - aspartic proteases, 1:138*f*
 - OS-9/XTP3-B, 3:259*t*
 - osteoarthropathies, 1:101*t*
 - osteoblasts
 - G_s family phenotypes, 2:521*t*
 - parathyroid hormone (PTH), 3:385–386
 - osteocalcin (OC), 4:551
 - osteocytes, 3:385–386
 - osteomalacia, 4:536–537
 - osteonectin, 1:319*t*, 3:91*t*
 - osteopetrosis, 4:507*t*
 - osteopontin (OP)
 - cell–matrix interactions, 1:426
 - integrins, 1:427*t*
 - matrix metalloproteinases (MMPs), 3:91*t*
 - osteoporosis
 - bisphosphonates, 1:519–520
 - primary biliary cirrhosis (PBC), 3:564–565
 - osteoprotogerin (OPG), 3:385–386
 - Otsuka Long Evans Tokushima Fatty (OLETF) rat, 1:115
 - ouabain, 3:686–687, 3:686*f*, 3:687*f*
 - outer arm dyneins, 2:185*t*, 2:187
 - outer mitochondrial membrane (OMM), 3:136–137
 - ovaries
 - long-chain acyl-CoA synthetase (LCAS), 2:292–293
 - ovarian cancer, 2:305

ovarian ecdysteroidogenic hormone (OEH), 2:592–593

overcome classical restriction (Ocr), 2:75

OX40 receptor protein, 4:454, 4:456f, 4:459

oxaloacetate

- anaplerotic pathways, 1:105, 1:106f, 1:108f, 1:109f
- citric acid cycle, 4:93f, 4:438f
- decarboxylation, 2:208, 2:209f
- gluconeogenesis, 2:382f, 2:385f, 2:386f
- thiamine diphosphate (ThDP), 1:207f
- tricarboxylic acid cycle, 4:439f

Oxalobacter formigenes, 2:208–209, 2:209f

oxalo-succinic acid, 4:438f

oxidases, 2:312, 3:697–698

oxidative phosphorylation

- coenzyme Q, 3:723–724, 3:724f
- creatine kinase/phosphocreatine circuit, 3:81, 3:82f
- mitochondrial CK (MtCK) isoenzyme, 3:82–84, 3:83f
- mitochondrial inner membrane, 3:146
- proton pumping, 3:661
- respiratory chain complex I (NADH: ubiquinone oxidoreductase), 1:488f, 4:87, 4:91

oxidized guanines, 2:54, 2:57–58, 2:57f

2-oxo-acid dehydrogenase complexes, 3:114

oxoglutarate dehydrogenase (OGDHC), 1:206, 1:207f

oxoguanine glycosylase (OGG1), 2:16–17

2-oxo-L-theohexono-4-lactone-2,3-enediol

See ascorbic acid

oxygenases, 3:371–374

- biological function, 3:373–374
- catalases, 2:536
- characteristics, 3:371
- classifications

 - dioxygenases, 3:371–372
 - monooxygenases

 - biological function, 3:374
 - characteristics, 3:372
 - external monooxygenases, 3:372
 - internal monooxygenases, 3:372

evolutionary development, 3:374

flavins, 2:311–312

historical research, 3:371

oxygen fixation reactions/substrates

- molecular oxygen, 3:372
- oxygen acceptors, 3:372–373
- reaction mechanisms, 3:373

oxygen-dependent degradation domain (ODDD), 4:68f, 4:571

oxygen-evolving complex (OEC), 3:508, 3:510f

oxygenic photosynthesis

- basic concepts, 2:335
- cytochrome *b_L*/f complex

 - cytochrome *b_L*/f complex, 4:335–340
 - amino acid subunit sequences, 4:338
 - C2 symmetry, 4:336
 - chlorophyll and carotenoids, 4:337f, 4:338
 - crystal structures, 4:335–338, 4:337f
 - electron transport chain, 4:335, 4:336f

evolutionary development, 4:339–340

functional role, 4:335, 4:336f, 4:338–339, 4:339f

heme b rotational isomers, 4:337–338, 4:338f, 4:339f, 4:339t

heme c_n, 4:337, 4:337f, 4:338f, 4:339t

inhibitors, 4:339

midpoint oxidation–reduction potentials, 4:338t

posttranslational modifications, 4:338

quinone-exchange cavity, 4:336–337, 4:337f

rate-limiting steps, 4:339

Rieske protein, 2:335, 4:335–338, 4:337f, 4:338t, 4:339t

phototrophic bacteria, 3:32

O⁶-methylguanine (O⁶-mG), 2:9, 2:10t, 2:12–13, 2:54

oxyntomodulin (OXM), 1:115, 1:117

oxysterols, 1:517f

oxytocin (OXY), 4:515–519

- ascorbic acid, 4:531–532
- body weight regulation, 3:78, 3:78f
- evolutionary development, 4:516f
- functional role, 4:515–517
- peptide amidation, 3:396
- peptide biosynthesis, 4:515
- receptors, 4:517–519, 4:518f
- structural characteristics, 4:515
- transmembrane domains, 4:517–519, 4:518f
- types and sources, 4:516t

oysters, 4:533t

ozymetazoline, 1:58–59

P

p16^{INK4} inhibitor, 4:105–106

p21-associated kinase (PAK), 1:440–441, 1:441f, 2:611, 3:441–442, 4:113

p23 protein, 4:310f

p24 protein, 2:175–176, 2:175f

p25 protein, 2:175f, 2:176

p27 protein, 2:175f, 2:176, 3:623t

p38 kinase

- angiotensin receptors, 1:113
- chemokine receptors, 1:480–482
- functional role, 3:178
- G protein-coupled receptor kinases (GRKs), 2:485t
- G-protein-signaling regulators (RGSs), 2:493
- muscarinic acetylcholine receptors, 3:216
- neurokinin-1 receptor (NK1R), 4:362–363
- phylogenetic relationships, 3:177f
- signaling pathways, 1:618f, 1:620f, 3:177f

p53 protein, 3:379–383

- apoptotic pathways, 1:389f, 1:412f, 1:415
- DNA damage checkpoints proteins, 1:412f, 1:415
- DNA repair processes, 3:269, 3:273t
- family classifications

 - characteristics, 3:382

p63/p73 proteins

- development controls, 3:382
- domain structures, 3:382f
- protein–protein interactions, 3:382f
- structural characteristics, 3:382
- therapeutic uses, 3:382–383
- tumor suppression, 3:382

functional domains

- characteristics, 3:379
- DNA-binding domain, 3:379, 3:380f, 3:382f
- oligomerization domain, 3:379, 3:380f, 3:382f
- transactivation domain, 3:379, 3:380f, 3:382f

glutathione peroxidases, 3:411f

glycogen synthase kinase-3 (GSK3), 2:436t

G protein-coupled receptor kinases (GRKs), 2:485t

p53-mediated signaling

- characteristics, 3:379–381
- growth arrest and apoptosis, 3:380–381, 3:380f
- regulated signaling pathways, 3:380f, 3:381

protein–DNA interactions, 2:151

regulation mechanisms

- general discussion, 3:381–382
- posttranslational modifications

 - acetylation, 3:380f, 3:381
 - characteristics, 3:381
 - phosphorylation, 3:380f, 3:381
 - ubiquitination, 3:380f, 3:381

protein–protein interactions, 3:381–382

- stress activation, 3:380f, 3:381

ribosome biogenesis, 4:121

tumor suppression, 2:151, 3:379

ubiquitin–proteasome degradation pathway, 3:623t

p62 protein, 2:175f, 2:176

p63 protein

- development controls, 3:382
- domain structures, 3:382f
- protein–protein interactions, 3:382f
- structural characteristics, 3:382
- therapeutic uses, 3:382–383
- tumor suppression, 3:382

p73 protein

- apoptotic pathways, 1:389f
- development controls, 3:382
- domain structures, 3:382f
- protein–protein interactions, 3:382f
- structural characteristics, 3:382
- therapeutic uses, 3:382–383
- tumor suppression, 3:382

p75 neurotrophin receptor (p75NTR), 3:241, 3:243–246, 3:244f, 3:245f, 4:459

p107 protein, 4:104

p130 protein, 1:480–482, 2:266–267, 4:104

P131 proteasome inhibitor, 3:593

p150^{Glued}_{protein}, 2:175–176, 2:175f, 2:176f

P2Y receptors, 3:375–378

- domain structures, 3:376
- extracellular nucleotide signaling, 3:375–376

- P2Y receptors (*Continued*)
 functional classification, 3:376–377, 3:376t
 functional role, 3:375
 lysophosphatidic acid (LPA) receptors, 2:763t
 pharmacological classification, 3:377
 physiological functions, 3:377–378
- P5/PDI, 1:291t, 1:293–294
- PA200 proteasome activator, 3:592, 3:593f
- pacemakers, 1:587–588
- pachynema, 3:19, 3:19f, 3:22f
- pachyonychia congenita, 2:670t
- pachytene checkpoint, 2:560
- Paenibacillus macerans*, 4:92
- palmitate
 fatty acid synthesis pathway, 2:275–279, 2:276f
 lipid rafts, 2:741
 protein palmitoylation, 3:645
- palmitic acid, 3:645
- palmitoleic acid, 2:288
- palmitoyl acyl transferases (PATs), 2:736–737
- palmitoylation
 fatty acyl-coenzyme A (CoA) synthetases, 2:291f
 G_q family proteins, 2:499
 lipid-modified proteins, 2:736–737, 2:737f, 2:737t
 proteins, 3:645–647
 basic concepts, 2:736–737, 3:645
 covalent modifications, 3:645, 3:646f
 functional role, 3:646–647
 inhibitors, 3:647
 N-palmitoylation, 2:737f, 3:645, 3:646f
 O-palmitoylation, 3:645, 3:646f
 palmitoylthioesterases, 3:646
 palmitoyltransferases, 2:736–737, 3:645–646
 research summary, 3:647
 S-palmitoylation, 2:736–737, 2:737f, 3:645, 3:646f
- palmitoyl CoA
 fatty acid synthesis pathway, 2:275–279, 2:276f
- mitochondrial ATP-dependent potassium ion (K⁺) channel (KATP), 3:551
- protein palmitoylation, 2:736–737, 3:646–647
- palmoplantar keratoderma, 2:670t
- pancreas
 acinar cells, 1:129, 1:130–131
 aquaporin water channels, 1:124t
 β-cells
 arachidonic acid regulated Ca²⁺ (ARC) channels, 1:129, 1:131
 glucokinase, 2:547
 microRNA (miRNA) regulation, 4:191–192, 4:194t
 calbindin-D_{28k}, 1:285
 digestion and absorption, 2:355f, 2:357–358, 2:357f
 energy homeostasis, 3:77f, 3:78f
 exocrine secretion, 2:357–358, 2:357f
- glucose-stimulated insulin release (GSIR), 2:596–601
 α-cell function, 2:600
 amino acids, 2:598–600
 augmentation pathway (APW), 2:598, 2:598f
 β-cell therapies, 2:600–601, 2:601f
 fatty acids, 2:598–600
 functional role, 2:596
 GDH or K⁺ channel-linked hyperinsulinemic hypoglycemia, 2:600
 GK activators (GKAs), 2:597f, 2:600–601, 2:601f
 glucokinase disease, 2:600
 glucokinase glucose sensor concept, 2:596–598, 2:597f, 2:598f
 insulin release stimulation, 2:599f
 K⁺ channel inhibitors and activators (KClIs and KCOs), 2:597f, 2:600–601
 neuroendocrine modifications, 2:600
 triggering pathway (TPW), 2:598, 2:598f
- G_s family phenotypes, 2:521t
- insulin metabolism, 2:603f
- insulin-secreting cells (β-cells), 2:596–601
 α-cell function, 2:600
 amino acids, 2:598–600
 augmentation pathway (APW), 2:598, 2:598f
 β-cell therapies, 2:600–601, 2:601f
 fatty acids, 2:598–600
 functional role, 2:596
 GDH or K⁺ channel-linked hyperinsulinemic hypoglycemia, 2:600
 GK activators (GKAs), 2:597f, 2:600–601, 2:601f
 glucokinase, 2:547, 2:600
 insulin release stimulation, 2:599f
 K⁺ channel inhibitors and activators (KClIs and KCOs), 2:597f, 2:600–601
 neuroendocrine modifications, 2:600
 research summary, 2:601
 signaling pathways, 2:597f, 2:598f
 triggering pathway (TPW), 2:598, 2:598f
- intracellular calcium ion (Ca²⁺) waves, 2:643–644
- mammalian glucose transport, 2:392t
- phosphofructokinase-2/fructose biphosphatase-2 (PFK-2/FBPase-2)-activated phosphorylation, 3:458
- proteinase-activated receptors (PARs), 3:605t
- pyruvate dehydrogenase kinase (PDK) regulation, 4:324–326
- vitamin D, 4:538–539
- pancreatic polypeptide (PP), 1:115–116, 1:117
- pantothenate kinase-associated neurodegeneration, 1:477t
- papain, 1:589
- Papaver somniferum*, 3:354
- Paracentrotus lividus*, 3:90
- Paracoccus*, 2:207f, 3:29–30
- Paracoccus denitrificans*
 anaerobic respiration, 3:30–31, 3:31f
 ATP synthase, 2:273–274
 cytochrome bc₁ complex, 1:593, 1:594f
 heme–copper oxidases, 1:603f, 1:604t
 oxygen binding and reduction, 1:605
 periplasmic electron-transport systems, 3:400, 3:401f
 reductive reactions, 1:487t
 respiratory chain complex IV (cytochrome c oxidase), 4:99f
 structural characteristics, 4:100, 4:101f
 succinate:quinone oxidoreductases (SQORs), 4:92–93
- Paracoccus pantotrophus*, 1:490t
- paracrine communication, 2:615–616, 2:615f
- Paramecium*, 1:567
- paramyxovirus F protein, 3:39f, 3:40
- paranemin, 2:632t, 2:634
- paraneoplastic pemphigus, 1:638t
- paraplegin, 4:271
- parasitic proteinases, 1:592
- parasitic protozoa, 2:477–478
- paraspeckles, 3:298, 3:299f
- parathyroid gland, 4:538–539
- parathyroid hormone (PTH)/PTH-related protein (PTHrP) receptor (PPR), 3:384–390
 calcitonin receptors (CTRs), 1:272
 calcium-sensing receptor (CaSR), 1:328–329, 1:329–330, 1:329f
 characteristics and functional role, 3:384
 disease connections
 Blomstrand's chondrodysplasia, 3:389
 Jansen's metaphyseal chondrodysplasia, 3:389
 parathyroid hormone (PTH)
 bone-forming osteoblasts/osteocytes, 3:385–386
 discovery, 3:384–385
 renal effects, 3:386
 structural characteristics, 3:386f
 structure–activity relationships
 PTH(1–34) functional domains, 3:386–387
 PTH ligands, 3:387
 parathyroid hormone-related protein (PTHrP)
 calcium-sensing receptor (CaSR), 1:329–330
 discovery, 3:386
 genetic manipulations, 3:386
 growth plate actions, 3:386
 PTHrP(1–34) functional domains, 3:386–387
 structural characteristics, 3:386f
 structure–activity relationships, 3:386–387
 research background
 molecular cloning, 3:384, 3:385f
 parathyroid hormone (PTH)
 bone-forming osteoblasts/osteocytes, 3:385–386
 discovery, 3:384–385
 renal effects, 3:386

- structural characteristics, 3:386f
parathyroid hormone-related protein (PTHrP)
discovery, 3:386
genetic manipulations, 3:386
growth plate actions, 3:386
structural characteristics, 3:386f
research summary, 3:390
structure–activity relationships
ligand interaction mechanism
extracellular domain (ECD)
interactions, 3:388
J-domain interactions, 3:388–389
two-site model, 3:387–388, 3:388f
molecular architecture
amino-terminal domain, 3:385f, 3:386f, 3:387
juxtamembrane domain, 3:387
signaling and regulation
carboxyl-terminal tail, 3:389
conformational changes, 3:389
transient receptor potential (TRP)
channels, 4:416
paraventricular nucleus (PVN), 3:78, 3:78f, 3:79f
parietal-eye photoreceptor, 3:490
parkinsonism-associated dementia (PDC), 4:416
Parkinson's disease
amyloid deposits, 1:101t
dopamine receptors, 2:173
exosomes and microvesicles (EMVs), 2:264
intermediate filaments, 2:632t
lysosomal dysfunction, 4:507t
mitochondrial DNA (mtDNA), 1:211f
nicotinamide adenine dinucleotide (NAD⁺/NADH) treatments, 1:177
protein folding defects, 3:575t, 3:630
rapamycin (mTOR), 3:198–199
respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:91
tau proteins, 1:477t
tetrahydrobiopterin deficiency, 3:669
ubiquitin–proteasome pathway, 3:592, 4:475–476
vitamin E, 4:547
PARL-type rhomboids, 4:51f, 4:54
paroxysmal nocturnal haemoglobinuria, 2:471t
parsley, 4:533t
parvalbumin, 1:316–317, 1:319t
pasireotide, 4:263t, 4:264
Pasteur effect, 2:445–446, 2:446f
Pasteurella, 2:43
Pasteurella haemolytica, 2:137t, 2:139–140
Pasteurella multocida, 2:496, 3:331f
pathogens
pathogen-associated molecular patterns (PAMPs)
innate immune system, 4:396
toll-like receptors (TLRs), 4:396–397, 4:397t
protein N-myristoylation, 3:641
vacuole/lysosome functions, 4:507
pattern recognition receptors (PRRs)
innate immune system, 4:396
toll-like receptors (TLRs), 4:396–401
future research areas, 4:400
infectious diseases, 4:399
pathogen-associated molecular patterns (PAMPs), 4:396–397, 4:397t
research background, 4:396
signaling pathways
characteristics, 4:398–399
MyD88-dependent signaling pathway, 4:398, 4:399f, 4:400f
TIR-containing adaptor inducing IFN β (TRIF)-dependent signaling pathway, 4:398–399, 4:399f, 4:400f
structural characteristics, 4:397–398
TIR-containing adaptor protein (TIRAP), 4:398–399, 4:399f, 4:400f
Toll/interleukin-1 receptor (TIR)
domain, 4:397–398
TRIF-related adaptor molecule (TRAM), 4:398–399, 4:399f, 4:400f
vaccine adjuvants, 4:399–400
Pax3 transcription factor, 1:182
Pax7 transcription factor, 1:182
paxillin, 2:321
PAZ domain, 3:99–100
PcrA helicase
characteristics and functional role, 3:261
DNA unwinding
active, rolling or hand-over-hand model, 3:264–265, 3:265f
active versus passive mechanisms, 3:264
dimeric inch-worm models, 3:265, 3:265f
inch-worm mechanisms, 3:264, 3:264f, 3:265f
protein oligomerization, 3:263
unwinding mechanisms, 3:263–264
research summary, 3:265
single-strand DNA (ssDNA) translocation, 3:263, 3:264, 3:264f
structural characteristics
crystal structures, 3:262, 3:262f
open and closed confirmations, 3:262–263
primary structures, 3:261–262
translocase-based protein displacement, 3:265
PDBSum database, 3:615t
pea lectin, 2:704t
peanuts, 2:702t, 2:703f
Pearson's marrow pancreas syndrome, 1:214
peas, 4:533t
peflin, 1:319t
pegvisomant, 3:431
Pelizaeus–Merzbacher disease (PMD), 4:272
pellagra, 1:177
pemphigus, 1:259–260
pemphigus foliaceus (PF), 1:638t, 1:639–640
pemphigus vulgaris (PV), 1:638t, 1:639–640
Penelope-like elements (PLEs), 3:284, 3:285f, 3:288
pentaheme nitrite reductase, 1:487t
pentose phosphate pathway, 3:391–395
basic concepts, 3:392f
carbon fate, 3:391, 3:394f
cerebral metabolism, 1:369f
chemical balance, 3:391, 3:394f
fatty acid metabolism–cancer relationship, 2:277
historical research, 3:391
operation modes
hexose monophosphate shunt, 3:391
metabolic reactions, 3:392
regulation mechanisms, 3:391–392
photosynthetic carbon dioxide (CO₂) fixation, 3:494, 3:494f, 3:495f
quantitation
carbon fate, 3:391, 3:394f
chemical balance, 3:391, 3:394f
clinical importance, 3:394
estimations, 3:392–394
methodologies, 3:392
reaction mechanisms
nonoxidative segments, 3:391, 3:393f
oxidative segments, 3:391, 3:392f
thiamine diphosphate (ThDP), 1:207f
pentosidine, 2:406f, 2:408f
PepA Protein, 1:561
peppers, 4:533t
pepsin, 1:137, 1:138f, 3:88
pepstatin, 1:140
peptidases
aminopeptidases, 1:97–99
characteristics, 1:97
methionine aminopeptidases (MetAPs)
general discussion, 1:97–99
isoform functions, 1:98–99
metal use, 1:98
structural organization, 1:98, 1:98f
substrate specificity, 1:98
N-terminal cotranslational processing, 1:97–99
physiological functions, 1:97
aspartic proteases, 1:137–140
disease processes
acquired immunodeficiency syndrome (AIDS), 1:139
Alzheimer's disease, 1:140
cancer, 1:139
fungal infections, 1:139–140
general discussion, 1:139
hypertension, 1:139
malaria, 1:139
genomic sequences, 1:137, 1:138f
historical research, 1:137
inhibitors
IA-3, 1:140
naturally occurring inhibitors, 1:140
pepstatin, 1:140
PI-3, 1:140
synthetic inhibitors, 1:140
structural characteristics
catalytic residues, 1:138–139
flap structure, 1:140
overall structure, 1:138, 1:138f
proenzyme structure, 1:139
research background, 1:137–139
peptides
calcitonin gene-related peptide (CGRP), 1:265–269
binding site distribution, 1:268

peptides (*Continued*)

- characteristics, 1:265
- molecular structure
 - amylin receptors, 1:268
 - calcitonin receptor-like receptor (CLR), 1:266, 1:267f
 - cellular signaling, 1:268
 - CLR-RAMP1 association, 1:266
 - receptor activity-modifying proteins (RAMPs), 1:266, 1:267f
 - receptor component protein (RCP), 1:268
- pathophysiology, 1:268–269
- pharmacological characteristics
 - CGRP receptor-selective drugs, 1:265, 1:266f
 - nonpeptide antagonists, 1:265, 1:266f
- receptor expression changes, 1:268
- structural characteristics, 1:265, 1:266f
- calcium-sensing receptor (CaSR), 1:329–330
- folding and assembly processes, 3:625–631
 - complex folding reactions
 - disulfide-coupled folding, 3:628
 - general discussion, 3:627–629
 - ligand binding, 3:629
 - molecular chaperones, 3:629
 - multiple-domain proteins, 3:628, 3:628f
 - folding defects, 3:630
 - folding mechanisms and kinetics
 - folding and unfolding rates, 3:626
 - folding intermediates, 3:626–627
 - Levinthal paradox, 3:626
 - transition states, 3:627, 3:627f
 - oligomeric protein assembly, 3:629–630, 3:630f
 - research background, 3:625
 - thermodynamics, 3:625, 3:626f
 - unfolding–refolding transition, 3:626f
- glucagon-like peptide-1 (GLP-1), 1:115, 1:116–117, 1:658, 3:78–79, 3:79f
- natriuretic peptides, 3:225–228
 - atrial natriuretic peptide (ANP), 1:101t, 3:225, 3:226f
 - B-type (brain) natriuretic peptide (BNP), 3:226, 3:226f
 - characteristics and functional role, 3:225
 - C-type natriuretic peptide (CNP), 3:226, 3:226f
 - natriuretic peptide receptors
 - characteristics, 3:226–227
 - guanylyl-cyclase-linked receptors, 3:226f, 3:227
 - natriuretic peptide clearance receptor, 3:226f, 3:227
 - therapeutic uses, 3:227
- neuropeptide Y (NPY)
 - appetite regulation, 4:303
 - body weight regulation, 3:77, 3:77f, 3:78f
 - insulin metabolism, 2:606
 - peptide amidation, 3:396
 - starvation effects, 4:303
- peptide:N-glycanase (PNGase), 2:329, 2:330f, 2:331

- peptide YY (PYY), 1:115–116
- peptidylglycine α -amidating
 - monooxygenase (PAM), 3:396–399
 - biochemistry, 3:396–398, 3:397f
 - C-terminal α -amidation, 3:396
 - dysregulation
 - impaired amidation, 3:399
 - overexpression, 3:399
 - evolutionary relationships, 3:398
 - expression, 3:398
 - impaired amidation
 - impaired copper delivery, 3:399
 - targeted gene disruption, 3:399
 - molecular genetics, 3:398–399
- peptidyl- α -hydroxyglycine α -amidating
 - lyase (PAL), 3:396–398, 3:397f
- peptidylglycine α -hydroxylating
 - monooxygenase (PHM)
 - ascorbic acid, 3:396
 - copper cofactors, 3:397–398, 3:397f, 3:398f
 - essential cofactors, 3:396–398
 - structural characteristics, 3:396–398, 3:397f
 - structural characteristics, 3:396, 3:397f
 - substrates, 3:396
- peptidyl transferase center (PTC), 4:131–132, 4:133f
- phage display, 3:431–436
 - display vectors
 - basic concepts, 3:431–432
 - phagemid vectors, 3:432, 3:432f
 - phage vectors, 3:432, 3:432f
 - polyvalent/monovalent displays, 3:431–432
 - phage-binding analysis, 3:435, 3:435f
 - phage-binding selections
 - affinity-based selection, 3:433–435, 3:435f
 - binding and elution, 3:433–435
 - preparation methods, 3:433
 - phage libraries
 - characteristics, 3:432–433
 - diversity library design, 3:433, 3:434t
 - sequence generation, 3:432–433, 3:433t, 3:434t
 - research applications, 3:431
 - ribosomes, 4:131–132, 4:133f
 - See also* proteins
- peptidylarginine deiminases (PADs), 4:62
- peptidyltransferase center (PTC), 4:419, 4:420f
- PER-ARNT-SIM (PAS) homology domain, 4:67, 4:68f, 4:571
- perhexiline, 3:71t
- Pericams, 2:578
- peridin, 1:502f, 1:503f
- Perilipin/Plin protein family
 - adipocyte fat mobilization, 1:47, 1:48f, 1:48t
 - functional roles, 1:51
- Perilipin 1 (Plin1)
 - characteristics, 1:47–48
 - lipolysis regulation, 1:48, 1:50f
 - major family members, 1:48f, 1:48t

- perinucleolar compartment (PNC), 3:324–325
- periostin, 4:551
- peripheral membrane proteins, 4:392–393
- peripheral nervous system, 3:216–217, 3:233
- peripheral neuropathy, 1:101t
- peripheral supramolecular activation cluster (pSMAC), 4:374, 4:374f
- peripherin, 2:632t, 2:634, 3:236
- periplasmic electron-transport systems, 3:400–406
 - complex-independent electron transfer, 3:404
 - formate dehydrogenase, 3:402–403, 3:403f
 - hydrogenases, 3:403
 - iron (Fe(III)) respiration, 3:405
 - Mitchellian redox loop mechanism, 3:400–404, 3:402f
 - multiheme *c*-type cytochromes, 3:404–405, 3:404f
 - nitrate reductase A, 3:403–404, 3:403f
 - periplasmic nitrate reductase (Nap), 1:487t, 3:403f, 3:404
 - respiratory chain complex III (ubiquinol cytochrome *c* reductase), 3:400, 3:401f
 - synthesis, 3:405
- PERK (PKR-like ER kinase), 2:524–525, 2:524f
- perlecan, 3:91t, 3:657t, 3:658
- permeases
 - functional role, 2:388
- phosphotransferase system (PTS), 2:387–388, 2:388f
- peroxidases, 3:407–412
 - action mechanisms, 2:535, 2:535f
 - biosynthetic relationships, 3:582f
 - characteristics, 3:407–411
 - functional role, 2:535
 - haloperoxidases, 2:535–536
 - heme peroxidases, 3:407–409, 3:408f
 - hydrogen peroxide (H₂O₂) regulation, 3:407, 3:408f, 3:411–412
 - nonheme peroxidases, 3:409–411, 3:410f
 - reactive oxygen species (ROS), 2:536
 - research summary, 3:412
 - structural characteristics, 2:535
- peroxiredoxins, 3:409, 3:410f, 3:411–412, 3:411f
- peroxisomes, 3:425–430
 - alpha-oxidation, 2:283–284
 - biogenesis
 - basic concepts, 3:426, 3:427t
 - budding/scission processes, 3:414
 - de novo* formation, 3:414
 - disease connections, 3:414
 - general discussion, 3:413–414
 - metabolite transporters, 3:413
 - peroxisomal matrix proteins and enzymes, 3:413
 - peroxisomal membrane proteins (PMPs), 3:413
 - characteristics, 3:413, 3:425
 - division and proliferation mechanisms, 3:429
 - eukaryotic cells, 4:204f

- flavins, 2:312
- functional role, 3:413, 3:425
- metabolic functions, 3:413–417
- alkyl-dihydroxyacetone phosphate (alkyl-DHAP), 3:416–417
 - bile acid synthesis, 3:416
 - cellular functions, 3:417
 - ether-linked phospholipid biosynthesis, 3:416–417
 - peroxisomal fatty acid α -oxidation, 3:416
 - peroxisomal fatty acid β -oxidation
 - disease connections, 3:415–416
 - oxidation processes, 3:414–415
 - purpose, 3:414–417
 - Refsum disease, 3:416
 - rhizomelic chondrodysplasia punctata (RCDP), 3:414, 3:417
- origin and distribution, 3:425
- peroxisomal matrix protein import, 3:413, 3:426–429, 3:428f
- peroxisomal membrane protein import, 3:413, 3:429
- peroxisome proliferator-activated receptor alpha (PPAR α)
- action mechanisms, 3:419–421, 3:420f
 - adiponectin, 3:72–73
 - expression, 3:418–419
 - functional roles, 3:418
 - liver-X receptor (LXR), 2:287–288
 - research summary, 3:423–424
 - therapeutic uses, 3:423
- peroxisome proliferator-activated receptor delta (PPAR δ)
- action mechanisms, 3:423
 - eicosanoid receptors, 2:191
 - expression, 3:418–419
 - functional roles, 3:418
 - research summary, 3:423–424
 - therapeutic uses, 3:423
- peroxisome proliferator-activated receptor gamma (PPAR γ)
- action mechanisms, 3:421–423, 3:422f
 - adipocyte development
 - lineage commitment, 1:53f
 - microRNA (miRNA)-based regulation, 4:192–193
 - terminal differentiation, 1:54–55, 1:54f
 - transcriptional control, 1:55f, 1:56, 3:421–423
- adiponectin, 3:73t, 3:74
- diabetes treatments
- thiazolidinedione (TZD), 1:657
 - weight loss and exercise, 1:657
- expression, 3:418–419
- functional roles, 3:418
- hypoxia inducible factor (HIF)
- transcriptional activation and repression responses, 4:67–68
- insulin signal transduction, 2:604–605, 2:604f
- leptin resistance, 2:260–261
- lysophospholipid receptors, 2:763t, 2:764
- mucins, 3:207–208, 3:208f
- research summary, 3:423–424
- therapeutic uses, 3:423
- peroxisome proliferator-activated receptors (PPARs), 3:418–424
- classifications, 3:418–419
 - eicosanoid receptors, 2:189, 2:191
 - functional roles, 3:418
 - research summary, 3:423–424
 - therapeutic uses, 3:423
- peroxisome-related disorders, 3:425–426, 3:426t
- pexophagy, 1:153
- photosynthetic metabolic pathways, 3:497–498, 3:498f
- prostanoids, 3:582–583
- regulation mechanisms, 3:429
- retinoic acid receptors, 3:418–419, 3:420f, 3:422f, 4:525, 4:527f
- thiamine diphosphate (ThDP), 1:207f
- very long-chain acyl-CoA synthetase (VLCAS), 2:293–294
- peroxynitrite, 2:77, 2:720, 3:169f, 3:171f, 3:172f
- peroxypterin, 3:668f
- Peters Plus syndrome, 2:471t
- PET genes, 3:306–309
- cytochrome *b₆* complex, 4:338t
 - disease connections, 3:309
 - functional role
 - essential PET genes, 3:309
 - general discussion, 3:307–309
 - mitochondrial genome, 3:307
 - nonessential PET genes, 3:307–309, 3:308f
 - respiratory chain complexes, 3:308f
 - historical research, 3:306–307
- midpoint oxidation–reduction potentials, 4:338t
- origins, 3:307
- regulation mechanisms, 3:309
- Saccharomyces cerevisiae*
- model systems, 3:307
 - number of proteins, 3:307
- pexophagy, 1:153
- PFAM database, 3:615t
- PhaBI enzyme, 2:137t
- phages
- phage display, 3:431–436
 - display vectors
 - basic concepts, 3:431–432
 - phagemid vectors, 3:432, 3:432f
 - phage vectors, 3:432, 3:432f
 - polyvalent/monovalent displays, 3:431–432
 - phage-binding analysis, 3:435, 3:435f
 - phage-binding selections
 - affinity-based selection, 3:433–435, 3:435f
 - binding and elution, 3:433–435
 - preparation methods, 3:433
 - phage libraries
 - characteristics, 3:432–433
 - diversity library design, 3:433, 3:434t
 - sequence generation, 3:432–433, 3:433t, 3:434t
 - research applications, 3:431
- recombination-dependent DNA replication (RDR), 4:23–24
- site-specific recombination systems, 1:1
- sliding clamp processivity, 4:240, 4:240f
- transposable bacteriophage (Mu), 3:276f, 3:277
- See also bacteriophages
- phage T4, 2:21, 2:22t, 2:25–26, 2:114–115, 2:116t, 3:577f
- phage T7
- DNA ligases, 2:29f, 2:34f
 - DNA polymerase holoenzymes, 2:114–115, 2:116t
 - DNA polymerase I, 2:87
 - helicases, 2:21, 2:22t, 2:25–26
 - nucleoside triphosphatase (NTPase), 2:22–24
- phagocytosis, 3:437–440
- basic concepts, 2:198f, 2:199, 3:437, 3:438f
 - characteristics, 2:197
 - endocytic–lysosomal pathways, 3:618, 3:618f
 - immunoglobulin Fc receptors (FcRs), 3:437–438
 - phosphatidylinositol-3-phosphate (PtdIns (3)P), 3:449
 - receptor recognition, 3:437–438
 - regulation mechanisms
 - actin/actin-binding proteins/actin-regulating proteins, 3:438–439, 3:438f
 - heterotrimeric and small GTP-binding proteins, 3:439
 - phosphoinositides, 3:439–440, 3:440f
- phakinin, 2:632t, 2:634
- phalloidin, 1:16–17, 1:16t
- phasvatocin, 4:516t
- phenoxybenzamine, 1:58
- phentolamine, 1:58
- phenylacetate
- hyperammonemia, 4:495–497, 4:497t
 - urea cycle, 4:495f
- phenylacetyl glutamine, 4:495f
- phenylalanine
- anaplerotic pathways, 1:106f
 - biopterin, 3:667
 - classifications, 1:91, 1:92t
 - functional roles, 1:93t
 - glycine–phenylalanine–hydroxyproline–glycine–glutamic acid–arginine (GFOGER) sequence, 1:427t, 1:428
 - phage libraries, 3:434t
 - tricarboxylic acid cycle, 4:439f
 - vasopressin/oxytocin family, 4:516t
- phenylephrine, 1:58, 2:497t
- phenylethanolamine N-methyltransferase, 1:57
- phenylpropanoids, 3:498f, 3:499f
- phenyppressin, 4:516t
- pheophytin, 3:483–484, 3:483f
- ϕ 12 P4 helicase, 2:22t
- ϕ 29 dsDNA-packaging motor, 2:24–25
- phorbol esters, 3:637
- Phormidium lamosum*, 3:530–531, 3:531f
- Phormidium lapideum*, 3:248f

- G_{bg} subunit functions, 1:480–482, 1:482*f*, 2:377
 glycogen synthase kinase-3 (GSK3), 2:435*f*
 guanine-nucleotide-exchange factors (GEFs), 4:245–246
 insulin signal transduction, 2:604, 2:604*f*
 phagocytosis regulation, 3:439–440
 phosphatase and tensin (PTEN) homolog, 1:563–564, 2:236–238, 2:237*f*, 2:581*f*, 3:195–196, 3:439–440, 3:448*f*, 3:450
 phosphatases, 3:450
 SHIP1/SHIP2 phosphatases, 2:236–237, 3:439–440, 3:448*f*, 3:450
 small guanosine phosphatases (GTPases), 4:247*f*, 4:248
 phosphatidylinositol 3-kinase (PI3K) pathway
 angiotensin receptors, 1:113
 B-cell antigen receptors (BCRs), 1:162–163, 1:163*f*
 cellular functions, 3:461
 c-fes proto-oncogene (Fes), 1:454
 chemoattractant signaling, 2:236–238, 2:237*f*
 cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:563–564
 diabetes treatments, 1:657
 DNA damage checkpoints proteins, 1:411*t*
 focal adhesion kinase (FAK) family, 2:265–266
 G_{bg} subunit functions, 1:480–482, 1:482*f*, 2:377, 4:265
 glycogen synthase kinase-3 (GSK3), 2:435*f*
 G protein-coupled receptor kinases (GRKs), 2:485*t*
 hematopoietin receptors, 2:526, 2:528*f*
 hemidesmosome function, 1:642–643
 insulin receptors (IRs), 1:198–199, 1:200*f*, 2:602–603
 insulin resistance mechanisms, 1:651, 1:651*f*
 insulin signal regulation, 2:606
 insulin signal transduction, 2:604
 liver regeneration, 1:193–195
 macroautophagy regulation, 1:149
 mTOR complexes, 1:408, 3:195–196, 3:197
 neurokinin-1 receptor (NK1R), 4:362–363
 neurotransmitter transporters, 3:239
 neurotrophins, 3:241–243, 3:242*f*
 phagocytosis regulation, 3:439–440
 phosphatidylinositol-3-phosphate (PtdIns (3)P), 3:447–448, 3:448*f*
 platelet-activating factor receptor (PAFR), 3:536, 3:536*f*
 platelet-derived growth factor (PDGF) receptors
 chemotaxis, 3:542
 mediated cellular responses, 3:541, 3:541*f*
 signaling pathways, 3:539–540, 3:540*f*
 small guanosine phosphatases (GTPases), 4:247*f*, 4:248
 sodium–potassium ($Na^+ - K^+$)-ATPase pump, 3:687
 somatostatin (sst) receptors, 4:265
 transforming growth factor beta (TGF- β), 1:620*f*
 vacuole transport pathways, 4:504–506
 zinc-finger domains, 4:578
 phosphatidylinositol-3-phosphate (PtdIns (3)P), 3:447–450
 autophagy control, 3:450
 characteristics, 3:447
 dephosphorylation, 3:448
 FYVE domain, 3:448, 3:449*t*
 molecular targets and cellular functions
 bacterial invasion, 3:449
 endosomal membrane trafficking, 3:449, 3:449*t*
 endosomal system signaling, 3:450
 phagocytosis, 3:449
 phox homology (PX) domain, 3:448–449, 3:449*t*
 protein domains, 3:448–449
 phosphatidylinositol 3-kinase (PI3K) pathway, 3:447–448, 3:448*f*
 phosphoinositide phosphatases, 3:447–448, 3:448*f*
 phosphorylation, 3:448
 rapamycin (mTOR) signaling, 3:450
 synthesis and turnover, 3:447, 3:448*f*
 VPS34/Vps34p gene complex, 3:447–448, 3:448*f*, 3:450
 phosphatidylinositol-4,5-bisphosphate (PIP₂), 3:447–451
 actin depolymerizing factor (ADF)/cofilin (ADF-H) domain, 1:22, 1:22*f*
 affinity and sensitivity, 2:747
 analytical methodologies, 2:746*t*
 biosynthetic pathways, 3:447–448
 cellular targets, 3:449, 3:449*f*
 characteristics, 3:447
 chemoattractant signaling, 2:236–238, 2:237*f*
 cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:563–564
 exosomes and microvesicles (EMVs), 2:262
 functional role, 2:745–746, 2:748
 G_{bg} subunit functions, 1:480–482, 1:482*f*, 4:265
 glycogen synthase kinase-3 (GSK3), 2:435*f*
 G protein-coupled receptor kinases (GRKs), 2:485*t*
 G-protein-coupled receptors (GPCRs), 4:225*f*
 G_q family proteins, 2:497–498, 2:498*f*
 hydrolysis mechanisms, 3:448–449
 insulin signal transduction, 2:604, 2:604*f*
 intracellular regulation overlap, 2:747, 2:747*f*
 ion channel activity, 2:746–747, 2:746*f*, 2:746*t*
 ion channel modulation, 2:746–747, 2:748
 IP₃ receptors (IP₃R), 2:654*f*
 metabolic pathways, 3:448*f*
 molecular structure, 3:448*f*, 3:460*f*
 phospholipase C (PLC), 2:654*f*, 3:467–468, 3:468*f*
 phospholipase D (PLD), 3:474–475, 3:474*f*
 plectin, 2:625–627
 regulation specificity, 2:747–748
 research summary, 2:748
 serotonin receptors, 4:225*f*, 4:226
 small guanosine phosphatases (GTPases), 4:247*f*
 somatostatin (sst) receptors, 4:265
 transient receptor potential (TRP) channels, 3:359–360, 3:360*f*, 3:360*t*, 4:412
 transport modulation, 2:746–747
 voltage-sensitive phosphatases, 2:746*f*
 phosphatidylinositol 4-kinase (PI4K) pathway
 cellular functions, 3:459–460
 subfamilies, 3:459
 phosphatidylinositol (PI), 3:478, 3:479*f*
 phosphatidylserine (PS), 3:478, 3:479*f*
 phosphocreatine, 3:81, 3:82*f*
 phosphodiesterases (PDEs), 1:580–585
 characteristics, 1:580
 class I PDEs
 affinity and selectivity characteristics, 1:569
 catalytic domains, 1:568–569, 1:568*f*
 characteristics, 1:567–568
 cyclic guanosine monophosphate (cGMP)-hydrolyzing PDEs, 1:569–570
 cyclic nucleotide specificity, 1:569
 domain structures, 1:568, 1:568*f*
 dual-specificity PDEs, 1:570–571
 inhibitors, 1:569
 metal requirements, 1:569
 occurrences, 1:567
 PDE1, 1:570–571, 1:581–582
 PDE2, 1:571, 1:581–582, 1:583–584
 PDE3, 1:571, 1:581–582, 1:583
 PDE5, 1:570, 1:581–582, 3:27
 PDE6, 1:570, 1:581–582
 PDE9, 1:570
 PDE10, 1:571, 1:581–582
 PDE11, 1:571–572, 1:581–582
 class II PDEs
 catalytic activity, 1:572, 1:572*f*
 characteristics, 1:572
 regulatory mechanisms, 1:572, 1:572*f*
 clinical applications, 1:584
 cyclic adenosine monophosphate (cAMP) characteristics and functional role, 1:567
 class I PDEs, 1:567–568, 1:568*f*
 class II PDEs
 catalytic activity, 1:572, 1:572*f*
 characteristics, 1:572
 regulatory mechanisms, 1:572, 1:572*f*
 distribution, 1:567
 molecular structure, 1:568*f*
 research summary, 1:572
 cyclic guanosine monophosphate (cGMP) characteristics and functional role, 1:567
 class I PDEs
 catalytic domains, 1:568–569, 1:568*f*
 characteristics, 1:567–568, 1:580

- phosphodiesterases (PDEs) (*Continued*)
- cyclic guanosine monophosphate (cGMP)-hydrolyzing PDEs, 1:569–570
 - cyclic nucleotide specificity, 1:569
 - domain structures, 1:568f
 - dual-specificity PDEs, 1:570–571
 - inhibitors, 1:569
 - PDE1, 1:570–571, 1:581–582
 - PDE2, 1:571, 1:581–582, 1:583–584
 - PDE3, 1:571, 1:581–582, 1:583
 - PDE5, 1:570, 1:581–582, 3:27
 - PDE6, 1:570, 1:581–582
 - PDE9, 1:570
 - PDE10, 1:571, 1:581–582
 - PDE11, 1:571–572, 1:581–582
 - class II PDEs
 - catalytic activity, 1:572, 1:572f
 - characteristics, 1:572
 - regulatory mechanisms, 1:572, 1:572f
 - cyclic guanosine monophosphate (cGMP)-hydrolyzing PDEs, 1:569–570
 - distribution, 1:567
 - molecular structure, 1:568f
 - research summary, 1:572
 - signaling pathways, 1:568f, 1:575f
 - family-specific inhibitors, 1:584
 - G_i family functions, 2:379
 - inherited disorders, 1:584
 - molecular diversity, 1:580–581
 - phototransduction mechanisms, 3:489f
 - research summary, 1:572, 1:584
 - second messenger pathways, 1:43f
 - signaling pathways, 1:574, 1:575f
 - structure and functions
 - catalytic domains, 1:581, 1:581f
 - homeostatic regulation, 1:582–583
 - intracellular effectors, 1:583
 - regulatory domains, 1:581–582, 1:581f
 - signal integrators, 1:583–584
 - spatially organized signalling networks and microdomains, 1:582
 - phosphoenolpyruvate carboxykinase (PEPCK), 2:381, 2:382–383, 2:385f
 - phosphoenolpyruvate (PEP)
 - carbohydrate-responsive element binding protein (ChREBP), 1:377f
 - glycolytic pathway, 2:445f
 - liver pyruvate kinase (LPK) catalysis, 1:374
 - phosphotransferase system (PTS), 2:387, 2:388f
 - photosynthetic metabolic pathways, 3:495
 - pyruvate kinase (PK), 3:719
 - phosphofructokinase
 - carbohydrate-responsive element binding protein (ChREBP), 1:375–376, 1:376f, 1:377f
 - gluconeogenesis, 2:385f
 - myocardial ischemia, 3:108–109
 - 6-phosphofructo-2-kinase/ fructose-2,6-P₂-bisphosphatase (F6P2K/F26Pase), 1:376f, 1:377f
 - phosphofructokinase-1 (PFK1)
 - cardiac metabolism, 3:65–66, 3:66f, 3:68f, 3:69f
 - enzymology, 3:456–457
 - phosphofructokinase-2/fructose bisphosphatase-2 (PFK-2/FBPase-2), 3:456–457
 - enzymology
 - fructose bisphosphatase-2 (FBPase-2), 3:457
 - phosphofructokinase-2 (PFK-2), 3:456–457
 - functional roles, 3:457–458
 - historical research, 3:456
 - isoforms, 3:457
 - phosphorylation activation, 3:457–458, 3:458f
 - structural characteristics, 3:456
 - 2-phosphoglycerate (2-PG), 2:445f
 - 3-phosphoglycerate (3-PG), 2:445f
 - phosphoglycerate kinase (PGK), 2:445f
 - phosphoglycerate mutase (PGM), 2:445f
 - phosphoglyceric acid, 3:497–498, 3:498f, 3:499f
 - phosphoglycolate, 2:2f, 3:498f
 - phosphoinositide 3-kinase-like kinase (PIKK) family, 1:412–413, 1:413f
 - phosphoinositides, 3:459–461
 - exosomes and microvesicles (EMVs), 2:263f
 - nomenclature, 3:459
 - phagocytosis regulation, 3:439–440, 3:440f
 - phosphatidylinositol 4-kinase (PI4K) pathway
 - cellular functions, 3:459–460
 - subfamilies, 3:459
 - phosphatidylinositol phosphatases
 - characteristics, 3:461
 - phosphoinositide 4-phosphatases
 - cellular functions, 3:461
 - characteristics, 3:461
 - phagocytosis regulation, 3:439–440
 - phosphoinositide 5-phosphatases
 - characteristics, 3:461
 - group I 5-phosphatases, 3:462
 - group II 5-phosphatases, 3:462
 - group III 5-phosphatases, 3:462
 - group IV 5-phosphatases, 3:462
 - phosphatidylinositol phosphate kinases (PIP kinases)
 - domain structures, 3:460, 3:460f
 - type I PIP kinases, 3:460–461, 3:460f
 - type II PIP kinases, 3:460f, 3:461
 - type III PIP kinases, 3:460f, 3:461
 - phosphoinositide-dependent protein kinases
 - functional role, 3:463
 - kinase S6 kinase (S6K), 3:464f, 3:465
 - mTOR (rapamycin), 3:464f, 3:465
 - phosphoinositide 3-kinase (PI 3-kinase)
 - lipid signaling pathway, 3:448f, 3:463, 3:464f
 - phosphoinositide-dependent kinase 1 (PDK1), 3:464–465, 3:464f, 4:377–378
 - phosphoinositide-dependent kinase 2 (PDK2), 3:464f, 3:465
 - protein kinase B (PKB), 3:463–464
 - research summary, 3:466
 - tyrosine kinases (Tec family), 3:465–466
 - phosphoinositide phosphatases, 3:447–448, 3:448f
 - structural characteristics, 3:459, 3:460f
 - phosphoketolase pathway, 2:204–206, 2:205t
 - phospholamban (PLB), 1:380f, 1:381, 1:382f
 - phospholipases
 - adrenergic receptors, 1:59
 - 2-arachidonoyl glyceryl ether synthesis, 2:195f
 - biosynthetic relationships, 3:582f
 - calcium-sensing receptor (CaSR), 1:328
 - phosphatidylcholine hydrolysis, 3:473f
 - phospholipase A (PLA), 1:242f, 1:423, 2:236–237
 - phospholipase C (PLC), 3:467–471
 - catalytic mechanisms, 3:468–469
 - characteristics and functional role, 3:467
 - domain structures, 3:468, 3:469f
 - dopamine receptors, 2:172
 - EF-hand proteins, 1:319t
 - enzymatic pathways, 1:659, 1:660f
 - G₁₂/G₁₃ family functions, 2:345f
 - G_{bg} subunit functions, 1:480–482, 1:482f, 4:265
 - glutamate receptors, 2:714–716, 2:715t
 - G-protein-coupled receptors (GPCRs), 4:225f
 - G_q family proteins, 2:496, 2:497–498, 2:498f
 - intercellular calcium ion (Ca²⁺) waves, 2:617
 - IP₃ receptors (IP₃R), 2:654f
 - isozyme families, 3:468, 3:469–471, 3:469f
 - muscarinic acetylcholine receptors, 3:215–216
 - neurokinin-1 receptor (NK1R), 4:362–363
 - phosphatidylinositol (PtdIns), 3:448–449
 - phospholipase C-β (PLCβ), 2:496, 2:497–498, 2:498f, 3:469–470, 3:469f
 - phospholipase C-δ (PLCδ), 1:279f, 3:469f, 3:470
 - phospholipase C-ε (PLCε), 3:469f, 3:470
 - phospholipase C-γ (PLCγ)
 - B-cell antigen receptors (BCRs), 1:162–163, 1:163f
 - characteristics, 3:469f, 3:470
 - phospholipase C-γ1 (PLCγ1), 2:305f, 3:241–243, 3:242f, 3:450, 3:540, 3:540f, 3:541f, 3:542
 - phospholipase C-η (PLCη), 3:469f, 3:471
 - phospholipase C-ζ (PLCζ), 3:469f, 3:470–471
 - physiological functions, 2:654f, 3:467–468, 3:468f
 - platelet-activating factor receptor (PAFR), 3:536, 3:536f
 - second messenger pathways, 1:43f, 1:279f

- serotonin receptors, 4:225f, 4:226
 somatostatin (sst) receptors, 4:265
 taste receptors, 4:368
 T-cell receptors (TCRs), 4:374
 phospholipase D (PLD), 3:472–477
 actin cytoskeleton rearrangement, 3:475–476
 B2 receptors, 1:242f
 cancer treatments, 3:476–477
 cellular functions
 bacterial PLD, 3:473
 Drosophila melanogaster, 3:473–474
 mammals, 3:474–475
 photoreceptors, 3:473–474
 plants, 3:474
 yeast PLD (*spo14*), 3:473
 zebrafish, 3:473
 endocytosis, 3:476
 enzymatic pathways, 1:659, 1:660f
 epidermal growth factor (EGF) receptors, 3:476
 exocytosis, 3:476
 functional role, 3:472
 Golgi apparatus, 3:476
 mammals
 activation mechanisms, 3:474–475, 3:474f
 alcohol effects, 3:475
 enzymatic functions, 3:474
 inhibitors, 3:475
 phosphatidic acid (PA), 3:475, 3:475f
 sensors, 3:475
 mitochondrial localization signal (mitoPLD), 3:477
 mTOR (rapamycin), 3:476
 phosphatidic acid (PA) regulation
 actin cytoskeleton rearrangement, 3:475–476
 endocytosis, 3:476
 epidermal growth factor (EGF) receptors, 3:476
 exocytosis, 3:476
 Golgi apparatus, 3:476
 mTOR (rapamycin), 3:476
 signal transduction, 3:476–477
 vesicle transport, 3:476
 phosphatidylcholine hydrolysis, 3:472
 phosphatidylinositol-3,4,5-triphosphate (PIP₃), 3:449
 research summary, 3:477
 signal transduction, 3:476–477
 structural domains, 3:472–473
 vesicle transport, 3:476
 phospholipids
 anti-mitochondrial antibodies (AMAs), 3:116
 B-cell antigen receptors (BCRs), 1:162–163, 1:163f
 chemical structure, 2:733, 2:734f
 composition, 3:478, 3:479f
 digestion and absorption, 2:358
 ether-linked phospholipid biosynthesis, 3:416–417
 exosomes and microvesicles (EMVs), 2:262
 fatty acyl-coenzyme A (CoA) synthetases, 2:291f
 flippase-mediated transbilayer movement, 2:314, 2:315f
 functional role, 3:478
 inositol phospholipids, 4:216
 lipid rafts, 2:742, 2:742f
 lipogenic gene expression, 2:287f
 lysophospholipids
 lysophosphatidic acid, 2:344t, 2:345f
 lysophospholipid receptors, 2:762–764
 glycerolipid receptors, 2:762–764
 intracellular receptors, 2:762, 2:764
 lysophosphatidic acid (LPA) receptors, 2:762–764, 2:763t
 plasma membrane receptors, 2:762–764
 sphingolipid receptors, 2:762
 remodeling processes, 2:417f, 2:418
 structural characteristics, 2:413–414, 2:415f
 phosphatidic acid (PtdOH), 2:412–414, 2:413f, 2:415, 2:416f
 phosphatidylcholine (PtdCho), 2:314, 2:358, 2:412, 2:413f, 2:414–415, 2:416f, 2:417f, 3:478, 3:479f
 phosphatidylethanolamine (PtdEtn), 2:412, 2:413f, 2:414–415, 2:416f
 phosphatidylglycerol (PtdGro), 2:412, 2:413f, 2:414–415, 2:416f
 phosphatidylinositol-3,4,5-triphosphate (PIP₃), 3:447–451
 binding proteins, 3:450
 biosynthetic pathways, 3:449–450
 cellular targets, 3:449f, 3:450–451
 characteristics, 3:447
 exosomes and microvesicles (EMVs), 2:262
 phagocytosis regulation, 3:439–440
 phosphatase and tensin (PTEN) homolog, 1:563–564, 2:237f, 2:581f, 3:195–196, 3:439–440, 3:448f, 3:450
 phosphatases, 3:450
 SHIP1/SHIP2 phosphatases, 2:236–237, 3:439–440, 3:448f, 3:450
 phosphatidylinositol-3,4,5-trisphosphate (PIP₃), 2:262, 3:447–451
 exosomes and microvesicles (EMVs), 2:262
 phosphatidylinositol-4,5-bisphosphate (PIP₂), 3:447–451
 actin depolymerizing factor (ADF)/cofilin (ADF-H) domain, 1:22, 1:22f
 affinity and sensitivity, 2:747
 analytical methodologies, 2:746t
 biosynthetic pathways, 3:447–448
 cellular targets, 3:449, 3:449f
 characteristics, 3:447
 chemoattractant signaling, 2:236–238, 2:237f
 cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:563–564
 exosomes and microvesicles (EMVs), 2:262
 functional role, 2:745–746, 2:748
 glycogen synthase kinase-3 (GSK3), 2:435f
 G protein-coupled receptor kinases (GRKs), 2:485t
 G-protein-coupled receptors (GPCRs), 4:225f
 G_q family proteins, 2:497–498, 2:498f
 hydrolysis mechanisms, 3:448–449
 insulin signal transduction, 2:604, 2:604f
 intracellular regulation overlap, 2:747, 2:747f
 ion channel activity
 affinity and sensitivity, 2:747
 analytical methodologies, 2:746t
 intracellular regulation overlap, 2:747, 2:747f
 modulation, 2:746–747, 2:748
 regulation specificity, 2:747–748
 voltage-sensitive phosphatases, 2:746f
 ion channel modulation, 2:746–747, 2:748
 metabolic pathways, 3:448f
 molecular structure, 3:448f, 3:460f
 phospholipase C (PLC), 3:467–468, 3:468f
 phospholipase D (PLD), 3:474–475, 3:474f
 plectin, 2:625–627
 regulation specificity, 2:747–748
 research summary, 2:748
 serotonin receptors, 4:225f, 4:226
 transport modulation, 2:746–747
 voltage-sensitive phosphatases, 2:746f
 phosphatidylinositol (PtdIns)
 biosynthetic pathways, 2:416f
 functional role, 2:414–415
 phosphatidylinositol-3,4,5-triphosphate (PIP₃), 3:447–451
 binding proteins, 3:450
 biosynthetic pathways, 3:449–450
 cellular targets, 3:449f, 3:450–451
 characteristics, 3:447
 exosomes and microvesicles (EMVs), 2:262
 phagocytosis regulation, 3:439–440
 phosphatase and tensin (PTEN) homolog, 1:563–564, 2:237f, 2:581f, 3:195–196, 3:439–440, 3:448f, 3:450
 phosphatases, 3:450
 SHIP1/SHIP2 phosphatases, 2:236–237, 3:439–440, 3:448f, 3:450
 phosphatidylinositol-3,4,5-trisphosphate (PIP₃), 3:447–451
 exosomes and microvesicles (EMVs), 2:262
 phosphatidylinositol-4,5-bisphosphate (PIP₂), 3:447–451
 actin depolymerizing factor (ADF)/cofilin (ADF-H) domain, 1:22, 1:22f
 affinity and sensitivity, 2:747
 analytical methodologies, 2:746t
 biosynthetic pathways, 3:447–448

- phospholipase C- γ (PLC γ), 3:469f, 3:470
- phospholipase C- η (PLC η), 3:469f, 3:471
- phospholipase C- ζ (PLC ζ), 3:469f, 3:470–471
- physiological functions, 2:654f, 3:467–468, 3:468f
- platelet-activating factor receptor (PAFR), 3:536, 3:536f
- second messenger pathways, 1:43f, 1:279f
- serotonin receptors, 4:225f, 4:226
- somatostatin (sst) receptors, 4:265
- taste receptors, 4:368
- T-cell receptors (TCRs), 4:374
- phospholipase D (PLD), 3:472–477
- actin cytoskeleton rearrangement, 3:475–476
- B2 receptors, 1:242f
- bacterial PLD, 3:473
- cancer treatments, 3:476–477
- cellular functions, 3:473–477
- Drosophila melanogaster*, 3:473–474
- endocytosis, 3:476
- enzymatic pathways, 1:659, 1:660f
- epidermal growth factor (EGF) receptors, 3:476
- exocytosis, 3:476
- functional role, 3:472
- Golgi apparatus, 3:476
- mammals, 3:474–475
- mitochondrial localization signal (mitoPLD), 3:477
- mTOR (rapamycin), 3:476
- phosphatidic acid (PA) regulation, 3:475–477
- phosphatidylcholine hydrolysis, 3:472
- phosphatidylinositol-3,4,5-trisphosphate (PtdIns-3,4,5-P₃), 3:449
- photoreceptors, 3:473–474
- plants, 3:474
- research summary, 3:477
- signal transduction, 3:476–477
- structural domains, 3:472–473
- vesicle transport, 3:476
- yeast PLD (*spo14*), 3:473
- zebrafish, 3:473
- pyridoxal phosphate, 3:708–711
- aldol-type reactions, 3:709f, 3:710, 3:710t
- carbanionic quinonoid intermediate reaction, 3:708–710, 3:710f
- characteristics and functional role, 3:708
- chemical mechanisms, 3:709f, 3:710f, 3:711
- decarboxylation, 3:709f, 3:710, 3:710t
- mechanism-based enzyme inhibitors, 2:214f
- molecular structure, 3:709f
- pyridoxal phosphate-dependent enzymatic reactions, 3:708–710, 3:709f, 3:710f, 3:710t
- racemization, 3:708–710, 3:709f, 3:710t
- serine–glycine interconversion, 2:324–325
- transamination reactions, 3:708, 3:709f
- sphingosine-1-phosphate lyase (SPLase), 4:289, 4:290f
- phosphorylase
- glycogen degradation pathways, 2:425, 2:428–429
- glycogen metabolic pathways, 2:429, 2:430f
- glycogen storage diseases
- McArdle disease, 2:431t, 2:432
- Type IXa glycogen storage disease, 2:431t, 2:432
- Type IXb glycogen storage disease, 2:431t, 2:432
- Type IXc glycogen storage disease, 2:431t, 2:432–433
- glycogen synthesis, 2:428–429, 2:430f
- glycolytic pathway, 2:443, 2:445f
- phosphorylase kinase (PhK), 1:305f, 1:306
- phosphorylation
- adenosine diphosphate (ADP) phosphorylation, 1:491, 2:204, 2:205f, 4:482
- Akt phosphorylation, 2:604–605, 2:604f
- anaerobic respiration, 2:206–208, 2:207f, 2:208f
- anaplerotic pathways, 1:109
- annexins, 1:276
- ATP synthesis, 1:141–142, 1:142f, 4:482
- basic concepts, 3:607
- branched-chain amino acid (BCAA) regulation, 1:247–248
- cadherin signaling, 1:264, 1:264f
- calmodulin (CaM)–nitric oxide synthase (NOS) interactions, 1:344–346, 1:344t, 1:345f
- carbohydrate-responsive element binding protein (ChREBP), 1:374–375, 1:376f
- cerebral metabolism, 1:370f, 1:371–373
- chloroplast protein kinase network, 1:514, 1:514f
- copper-transporting ATPases (copper pumps), 3:674
- CREB (cyclic adenosine monophosphate (cAMP) responsive binding protein), 1:313
- C-terminal repeat domain (CTD) initiation phase, 2:366, 3:63–64, 3:63f
- transcription cycle, 2:365, 3:63–64, 3:63f
- transcription termination, 2:367, 3:63–64, 3:63f
- cyclic nucleotide-dependent protein kinases, 1:578, 1:578f
- DNA damage checkpoints proteins, 1:412f
- electron transport phosphorylation, 2:206–208, 2:207f, 2:208f
- fibroblast growth factor receptors (FGFRs), 2:305f
- gastric proton–potassium (H⁺–K⁺)-ATPase, 3:677f, 3:678
- G_i family functions, 2:380
- gluconeogenesis, 2:385
- glutamate receptors, 2:719–720
- glycogen synthase kinase-3 (GSK3), 2:434–436, 2:435f
- glycosyltransferase-catalyzed reactions, 1:3
- G protein-coupled receptor kinases (GRKs), 2:483–486, 2:485t
- histone modifications, 1:314–315, 1:521, 1:522f, 4:61
- insulin metabolism, 2:606–607
- insulin receptors (IRs), 2:603–606
- integrin signaling, 2:608
- liver PK (LPK) *in vitro* regulation, 3:719
- mitochondrial calcium (Ca²⁺) transport, 3:120–121, 3:120f, 3:121f
- mTOR complexes, 3:194–195, 3:196f, 3:197f
- mucins, 3:206
- neurofilament triplet proteins (NFTPs), 3:233–235, 3:234f
- neurokinin-1 receptor (NK1R), 4:363–364, 4:363f
- neurotrophins, 3:241–243, 3:242f
- nicotinamide adenine dinucleotide (NAD⁺/NADH), 1:172, 1:174f
- nuclear lamina, 3:311f, 3:312
- oxidative phosphorylation
- coenzyme Q, 3:723–724, 3:724f
- creatine kinase/phosphocreatine circuit, 3:81, 3:82f
- mitochondrial CK (MtCK) isoenzyme, 3:82–84, 3:83f
- respiratory chain complex I (NADH: ubiquinone oxidoreductase), 1:488f, 4:87, 4:91
- phosphate-transfer reactions, 2:580, 2:581f
- phosphofructokinase-2/fructose bisphosphatase-2 (PFK-2/FBPase-2)-activated phosphorylation, 3:457–458, 3:458f
- PKA-mediated lipolysis, 1:49–51, 1:50f, 4:303–304
- plant mitochondrial electron transport chain, 1:141–142, 1:142f
- protein kinase C (PKC), 3:637–638, 3:638f
- protein tyrosine phosphatase (PTP), 3:648, 4:294–295, 4:296f
- pyruvate dehydrogenase kinase (PDK) regulation, 4:324–326, 4:325f
- retinoblastoma protein (pRB), 4:105
- sarcoendoplasmic reticulum Ca²⁺ ATPase (SERCA), 2:228–229
- sodium–potassium (Na⁺–K⁺)-ATPase pump, 3:682–685
- somatostatin (sst) receptors, 4:263–264, 4:263t
- Src (sarcoma) family kinases, 4:295–297, 4:296f
- steroid/thyroid hormone receptors, 4:312
- sugar phosphorylation, 2:543
- translation initiation factor 2 (eIF2 α) phosphorylation, 4:491
- tyrosine, 2:620–621, 2:620f, 2:622f, 4:61
- voltage-gated calcium channels, 4:561
- phosphosphingolipids, 4:282, 4:282f
- phosphotransacetylase, 4:74f

- phosphotransferase system (PTS)
 classification, 2:387
 components, 2:387, 2:388f
 functional role, 2:387, 2:388f
 inositol phosphatases/inositol phosphate kinases, 2:581–582, 2:581f
 preferential glucose utilization, 2:387–388
 phosphotyrosine binding (PTB) domain, 2:603–606
Photinus pyralis, 2:290–291
Photobacterium phosphoreum, 3:666
 photofermentation, 4:74, 4:75f
 photoinhibition, 3:482–487
 basic concepts, 3:484
 charge recombination yield, 3:485
 D1 protein degradation, 3:484–485
 molecular mechanisms, 3:485
 proteases, 3:484
 photolithoautotrophs, 1:488t
 photolithotrophs, 1:488t
 photon migration, 4:279
 photoorganotrophs, 1:488t
 photoreceptors, 3:488–491
 characteristics, 3:488
 deep-brain photoreceptors, 3:491
 nonrod/noncone photoreceptors, 3:490, 3:490f
 parietal-eye photoreceptor, 3:490
 phospholipase D (PLD), 3:473–474
 phototransduction mechanisms, 3:489–490, 3:489f
 pineal photoreceptors, 3:490
 retinal photoreceptors
 characteristics, 3:488
 color vision, 1:550, 1:551f
 cones, 3:488, 3:489f
 rods
 cell body, 3:488, 3:489f
 inner segment, 3:488, 3:489f
 morphology, 3:488, 3:489f
 outer segment, 3:488, 3:489f
 synaptic terminal, 3:488, 3:489f
 photosynthesis, 3:492–496
 adenosine triphosphate (ATP) synthesis, 1:511–512, 1:512f, 3:493–494, 3:494f
 antenna complexes, 3:484f
 basic concepts, 2:335
 carbon dioxide (CO₂)-concentrating mechanisms
 aquatic photosynthesis, 3:495
 C4 pathways, 3:495, 3:495f
 Crassulacean acid metabolism (CAM) pathway, 3:495–496
 carbon dioxide (CO₂) fixation
 Benson–Calvin cycle, 3:494, 3:494f, 3:495f
 general discussion, 3:494–495
 photorespiration, 3:494–495
 electron transport mechanisms, 3:493–494, 3:493f, 3:494f
 light harvesting process, 2:178–183
 antenna complexes, 2:726–727, 3:484, 3:484f
 basic concepts, 2:178–179, 2:179f, 3:482
 biological diversity, 2:182
 chlorophyll and carotenoids, 1:498–505, 3:486, 3:492
 chloroplast signaling
 electron flow pathways, 1:511–512, 1:512f
 state transitions, 1:513, 1:513f
 thylakoid protein kinase network, 1:514f
 energy dissipation, 3:493
 excess light response, 2:182
 general discussion, 2:178
 light harvesting centers (LHCs), 1:500
 nonphotochemical energy dissipation
 energy-dependent quenching (qE), 2:180, 2:181f
 quenched state characteristics, 2:181–182, 2:181f
 quenching site, 2:180
 xanthophyll cycle, 2:180–181, 2:181f, 3:486–487, 3:487f
 nonphotochemical quenching (NPQ)
 algae, 3:486
 basic concepts, 3:485–486
 cyanobacteria, 3:486
 photoinhibition, 3:484
 reaction centers, 1:500, 2:178–179, 2:335, 3:492, 3:512, 3:513f
 reactive oxygen species (ROS), 1:500, 3:482–484, 3:483f, 3:493
 regulatory mechanisms, 2:179
 state transitions, 2:179–180, 2:180f
 structural characteristics, 2:179f
 thylakoid superstructure function, 2:181f, 2:182
 water oxidation, 3:493
 xanthophyll cycle, 3:486–487, 3:487f
 photoinhibition, 3:482–487
 basic concepts, 3:484
 charge recombination yield, 3:485
 D1 protein degradation, 3:484–485
 molecular mechanisms, 3:485
 proteases, 3:484
 photosynthetic structures, 3:492, 3:493f
 protonmotive electron transfer chain, 1:469–470, 1:470f
 type I reaction centers, 2:335–342
 F_A and F_B clusters, 2:338–339
 F_A- and F_B-containing proteins
 heliobacteria, 2:339, 2:339f
 PsaA/PsaB heterodimer, 2:340–341, 2:341f
 PsaC protein, 2:339f, 2:340, 2:340f, 2:341f
 PscB protein, 2:339–340, 2:339f
 PshB protein, 2:339, 2:339f
 tight binding conditions, 2:340–341, 2:341f
 F_x cluster
 heliobacteria, 2:338
 photosystem I (PSI) complex, 2:338, 2:338f
 green sulfur bacteria
 characteristics, 2:335
 functional role, 2:337–338
 F_x cluster, 2:338
 PscB protein, 2:339–340, 2:339f
 heliobacteria
 characteristics, 2:335
 F_A- and F_B-containing proteins, 2:339, 2:339f
 functional role, 2:337–338
 F_x cluster, 2:338
 PshB protein, 2:339, 2:339f
 iron–sulfur clusters (FeS)
 characteristics and functional role, 2:335–336, 2:336f
 F_x cluster, 2:338, 2:338f
 occurrences, 2:335
 photosystem I (PSI) complex, 2:336–337, 2:337f
 redox cofactor arrangement, 2:337f
 photosynthetic carbon dioxide (CO₂) fixation, 3:497–502
 C4 pathways, 3:498
 Calvin cycle enzyme targets, 3:500–501
 carbon pathway, 3:497, 3:499f
 carboxylation phase, 3:497–498, 3:499f
 Crassulacean acid metabolism (CAM) pathway, 3:498
 enzymatic regulation
 flux control, 3:500
 irreversible enzymes, 3:500
 feedback regulation, 3:501
 genetic modification effects, 3:500–501
 light-regulated energy and power, 3:497, 3:498f
 metabolic regulation, 3:499–500
 reduction phase, 3:498, 3:499f
 regeneration phase, 3:499, 3:499f
 whole plants, 3:501
 photosystem I (PSI) complex, 3:503–507
 basic concepts, 3:503
 core-antenna system, 3:506–507
 cytochrome *b₆f* complex, 4:335
 definition, 3:503
 domain structures, 3:531, 3:532f
 electron transfer chain, 3:505–506, 3:506f
 electrostatic interactions, 3:532, 3:532f
 excitation energy transfer
 core-antenna system, 3:506–507
 electron transfer chain, 3:505–506, 3:506f
 hydrophobic interactions, 3:531–532, 3:532f
 light harvesting complexes, 3:503–505, 3:505f
 light-induced damage, 3:485
 protonmotive electron transfer chain, 1:470f
 research summary, 3:507
 structure and functions
 cofactors, 3:503, 3:504f
 ferredoxin (Fd), 3:505, 3:505f
 iron–sulfur clusters (FeS), 3:505, 3:505f
 proteins, 3:503–505, 3:504f, 3:505f
 structural characteristics, 3:503, 3:504f
 type I reaction centers, 2:335–342
 characteristics, 2:336–337, 2:337f
 F_A- and F_B-containing proteins
 heliobacteria, 2:339, 2:339f
 PsaA/PsaB heterodimer, 2:340–341, 2:341f

- PsbC protein, 2:339f, 2:340, 2:340f, 2:341f
 PsbB protein, 2:339–340, 2:339f
 PshB protein, 2:339, 2:339f
 tight binding conditions, 2:340–341, 2:341f
 F_x cluster, 2:338, 2:338f
 iron–sulfur clusters (FeS)
 characteristics and functional role, 2:335–336, 2:336f
 F_x cluster, 2:338, 2:338f
 redox cofactor arrangement, 2:337f
 photosystem II (PSII) complex, 3:508–511
 antenna complexes, 3:484f
 assembly processes
 regulation mechanisms, 3:509
 research background, 3:508
 sequence, 3:508–509
 basic concepts, 3:503, 3:520
 characteristics and functional role, 3:512
 chloroplast–nucleus cooperation
 characteristics, 3:508
 chloroplast-encoded PSII proteins, 3:508
 nuclear-encoded PSII proteins, 3:508
 chloroplast signaling
 basic concepts, 1:511–512
 ferredoxin/thioredoxin system, 1:514–515
 plastoquinone pool
 electron flow pathways, 1:511–512, 1:512f
 long-term responses, 1:513–514
 redox state, 1:512–514, 1:514f
 short-term responses, 1:513, 1:513f
 state transitions, 1:513, 1:513f
 retrograde signaling pathways
 flu mutant, 1:515
 Gun mutants, 1:515
 reactive oxygen species (ROS), 1:515
 thylakoid protein kinase network, 1:514, 1:514f
 cytochrome *b₆*f complex, 4:335, 4:336f
 D1 protein turnover processes
 photoinactivation–damage
 mechanisms, 3:509, 3:510f
 photoinhibition repair cycle
 basic concepts, 3:484
 charge recombination yield, 3:485
 D1 protein degradation, 3:484–485, 3:509–510, 3:510f
 D1 protein synthesis, 3:510, 3:510f
 initiation phase, 3:509, 3:510f
 molecular mechanisms, 3:485
 post-translational assembly and reactivation, 3:510–511, 3:510f
 proteases, 3:484
 singlet oxygen production, 3:484–485
 reactive oxygen species (ROS), 3:482–484, 3:483f
 light harvesting, 2:178–183
 antenna complexes, 2:726–727, 3:484, 3:484f
 basic concepts, 2:178–179, 2:179f, 3:482
 biological diversity, 2:182
 chlorophyll and carotenoids, 3:492
 chloroplast signaling
 electron flow pathways, 1:511–512, 1:512f
 state transitions, 1:513, 1:513f
 thylakoid protein kinase network, 1:514f
 energy dissipation, 3:493
 excess light response, 2:182
 general discussion, 2:178
 nonphotochemical energy dissipation
 energy-dependent quenching (qE), 2:180, 2:181f
 quenched state characteristics, 2:181–182, 2:181f
 quenching site, 2:180
 xanthophyll cycle, 2:180–181, 2:181f, 3:486–487, 3:487f
 nonphotochemical quenching (NPQ)
 algae, 3:486
 basic concepts, 3:485–486
 cyanobacteria, 3:486
 photoinhibition, 3:484
 reaction centers, 2:178–179, 3:492
 reactive oxygen species (ROS), 3:482–484, 3:483f, 3:493
 regulatory mechanisms, 2:179
 state transitions, 2:179–180, 2:180f
 structural characteristics, 2:179f
 thylakoid superstructure function, 2:181f, 2:182
 water oxidation, 3:493
 xanthophyll cycle, 3:486–487, 3:487f
 morphology
 light harvesting complexes, 3:512
 reaction centers, 3:512, 3:513f
 thylakoid membrane, 3:512
 photoinhibition, 3:482–487
 basic concepts, 3:484
 charge recombination yield, 3:485
 D1 protein degradation, 3:484–485
 molecular mechanisms, 3:485
 proteases, 3:484
 protein components, 3:512–519
 PsbA-D1 protein, 3:514–515, 3:515t, 3:516f
 PsbB-CP47 protein, 3:515, 3:515t, 3:516f
 PsbC-CP43 protein, 3:515, 3:515t, 3:516f
 PsbD-D2 protein, 3:515, 3:515t, 3:516f
 PsbE protein, 3:515t, 3:516, 3:516f
 PsbF protein, 3:515t, 3:516, 3:516f
 psb genes, 3:514–519, 3:515t
 PsbG protein, 3:515t, 3:516, 3:516f
 PsbH protein, 3:515t, 3:516–517, 3:516f
 PsbI protein, 3:515t, 3:516f, 3:517
 PsbJ protein, 3:515t, 3:516f, 3:517
 PsbK protein, 3:515t, 3:516f, 3:517
 PsbL protein, 3:515t, 3:516f, 3:517
 PsbM protein, 3:515t, 3:516f, 3:517
 PsbN protein, 3:515t, 3:516f, 3:517
 PsbO protein, 3:515t, 3:516f, 3:517–518
 PsbP protein, 3:515t, 3:516f, 3:518
 PsbQ protein, 3:515t, 3:516f, 3:518
 PsbR protein, 3:515t, 3:516f, 3:518
 PsbS protein, 3:515t, 3:516f, 3:518
 PsbT_c protein, 3:515t, 3:516f, 3:518
 PsbT_n protein, 3:515t, 3:516f, 3:518
 PsbU protein, 3:515t, 3:516f, 3:518
 PsbV protein, 3:515t, 3:516f, 3:518
 PsbW protein, 3:515t, 3:516f, 3:518
 PsbX protein, 3:515t, 3:516f, 3:519
 PsbY protein, 3:515t, 3:516f, 3:519
 PsbZ protein, 3:515t, 3:516f, 3:519
 reaction center core complex, 3:514–519, 3:515t, 3:516f
 research summary, 3:519
 Ycf12 protein, 3:515t, 3:516f, 3:519
 protonmotive electron transfer chain, 1:470f
 purple bacterial photosynthetic reaction centers, 3:707
 redox active cofactors, 3:512–519
 bicarbonate, 3:514
 chloride ions, 3:514
 Chl_{z_{o1}} and Chl_{z_{o2}}, 3:513f, 3:514
 heme proteins, 3:513f, 3:514
 manganese calcium (Mn₄Ca²⁺) clusters, 3:514
 P680, 3:512, 3:513f
 pheophytin α (Pheo_{o1}), 3:512, 3:513f
 plastoquinone (PQ) molecule (Q_A), 3:513f, 3:514
 plastoquinone (PQ) molecule (Q_B), 3:513f, 3:514
 tyrosines Y_z and Y_o, 3:513f, 3:514
 structural characteristics, 3:521f
 thylakoid membranes, 3:508–511
 assembly processes
 regulation mechanisms, 3:509
 research background, 3:508
 sequence, 3:508–509
 chloroplast–nucleus cooperation
 characteristics, 3:508
 chloroplast-encoded PSII proteins, 3:508
 nuclear-encoded PSII proteins, 3:508
 D1 protein turnover processes
 photoinactivation–damage
 mechanisms, 3:509, 3:510f
 photoinhibition repair cycle, 3:509–511, 3:510f
 photoinhibition repair cycle
 D1 protein degradation, 3:509–510, 3:510f
 D1 protein synthesis, 3:510, 3:510f
 initiation phase, 3:509, 3:510f
 post-translational assembly and reactivation, 3:510–511, 3:510f
 water oxidation, 3:520–524
 action mechanisms
 general discussion, 3:523
 oxygen (O₂) formation, 3:523
 proton release, 3:523
 substrate/bulk water molecule exchange, 3:523
 water–water cycle, 3:485
 catalytic centers, 3:523
 manganese clusters
 assembly processes, 3:521, 3:521f
 research background, 3:522–523
 structural characteristics, 3:522–523
 redox reactions, 3:523

- photosystem II (PSII) complex (*Continued*)
 research summary, 3:523–524
 water-splitting enzyme
 basic concepts, 3:520–521
 oxygen (O₂) emission yield, 3:522f
 S states oxidation rates, 3:522
 S states redox potentials, 3:522
 thermodynamics, 3:521–522
- phototrophic bacteria
 anoxygenic photosynthesis, 2:209, 2:335, 3:32
 antenna complexes, 3:32
 chlorophototrophs, 2:501–503
 classifications, 4:79
 evolutionary development, 3:34
 general discussion, 3:32
 metabolic characteristics, 4:79
 oxygenic photosynthesis, 3:32
 respiration, 4:79–82
 cyanobacteria
 electron transport pathways, 4:81–82, 4:82f
 oxygenic photosynthesis, 4:81
 respiratory terminal oxidases (RTOs), 4:81–82
 electron transport chains, 4:79, 4:80f
 genetic regulatory networks, 4:80–81, 4:80f
 metabolic characteristics, 4:79
 oxygenic photosynthesis, 4:81
 respiratory chain interactions, 4:79–80
- phox (PX) domain, 3:448–449, 3:449t, 3:450
- phycobillins, 3:32
- phyloquinone, 3:505–506, 3:506f, 3:727–728, 4:552f
- Physarum*, 1:23
- phytanic acid
 peroxisomal fatty acid α -oxidation, 3:416
 thiamine diphosphate (ThDP), 1:207f
- phytesin, 1:138f
- phytoestrogens, 2:167–168
- phytoplankton, 3:495
- PI-3 inhibitor, 1:140
- PI3 kinase pathway
 See phosphatidylinositol 3-kinase (PI3K) pathway
- PIAS family, 3:207–208, 3:208f, 4:468t, 4:469
- Pichia pastoris*
 calpains, 1:360
 lysyl tyrosyl quinone, 1:89
 peroxisomal matrix protein import, 3:426–429
- Pick disease, 1:477t
- picrotoxinin, 2:423, 2:723–724
- piERICIDIN A, 1:143, 1:143t
- pigmentation, 3:25, 3:25t, 3:26f
- pigments, 3:666
- pigs, 4:516f, 4:516t
- pilV* locus, 1:555–556
- pindolol, 1:59
- pineal photoreceptors, 3:490
- pineapple, 4:533t
- Ping-pong mechanism, 2:218, 2:219–220, 2:219f, 2:220f
- pinocytosis
 basic concepts, 3:437
 characteristics, 2:197, 2:198f
 clathrin-independent endocytosis, 2:198
 endocytic-lysosomal pathways, 3:618, 3:618f
 macropinocytosis, 2:198f, 2:199
- [PIN+] prion
 generation mechanisms, 3:569
 genetic properties, 3:566
 self-propagating amyloidosis, 3:567–568
- Pisum sativum*, 1:89, 3:505f
- pituitary adenylate cyclase-activating polypeptide (PACAP), 3:396
- pituitary gland, 2:172–173, 2:172t
- piwi-interacting RNAs (piRNAs), 4:253–255
- placenta, 2:292–293, 2:392t
- placental growth factor (PLGF), 4:510–511
- plakin family
 bullous pemphigoid antigen 1 (BPAG1), 2:624–630
 characteristics, 2:628–629
 deficiency-related diseases, 2:627t, 2:629, 2:629t
 domain structures, 2:625f, 2:626f, 2:628–629
 expression regulation, 2:628
 gene locus, 2:625f, 2:628
 integrin signaling, 2:610–611
 isoform diversity, 2:626f, 2:628
 microtubule-binding domain, 3:105f, 3:107
 molecular interactions, 2:629
 subcellular localization, 2:628
- desmosomes, 1:636–639
- disease connections, 1:638t
- hemidesmosomes, 1:640f
- microtubule-binding domain, 3:105f, 3:107
- plectin, 2:624–630
 deficiency-related diseases, 2:627–628, 2:627t, 2:628t
 domain structures, 2:625, 2:625f, 2:626f
 expression regulation, 2:624–625
 gene locus, 2:624, 2:625f
 hemidesmosome function, 1:641–642, 1:641t
 integrin signaling, 2:610–611
 isoform diversity, 2:624, 2:626f
 microtubule-binding domain, 3:107
 molecular interactions, 2:625–627
 research background, 2:624–628
 structural characteristics, 1:640f, 1:641t
 subcellular localization, 2:624–625
- plakoglobin, 1:257, 1:259t, 1:636–639
- plakophilin, 1:257, 1:258f, 1:636–639, 1:637f
- plant glycoproteins, 2:465–468
 functional role analyses
 future research areas, 2:467
 genetic studies, 2:467
 immunohistochemistry, 2:467
 structural studies, 2:467
- glycan classifications
 N-linked glycans, 2:465, 2:466f
 O-linked glycans, 2:465, 2:466f
- multigene families, 2:467
- O-linked glycans
 characteristics, 2:465, 2:466f
 synthetic pathways
 glycosyl transferases, 2:466
 hydroxyproline residue arrangement effects, 2:466, 2:466f, 2:467f
 proline hydroxylation, 2:465
- plant lectin binding analyses, 2:460–461
- plants
 antenna complexes, 3:484, 3:484f
 aquaporin water channels, 1:125
- ascorbic acid, 4:532
- ATP synthesis, 1:141–144
 background information, 1:141
 glycine decarboxylase, 1:142–143
 inhibitors, 1:143, 1:143t
 mitochondrial electron transport chain, 1:141–142, 1:142f
 substrates, 1:142–143
- uncoupling processes
 alternative NAD(P)H dehydrogenases, 1:144
 alternative oxidases (AOXs), 1:144
 chemical uncouplers, 1:143
 nonphosphorylating respiration, 1:142f, 1:144
 uncoupling proteins (UCPs), 1:142f, 1:144
- brassinosteroids (BRs), 1:250–254
 background and characteristics, 1:250
- biosynthetic pathway
 campesterol-to-brassinolide pathway, 1:251, 1:252f
 cycloartenol-to-campesterol pathway, 1:251
 general discussion, 1:250–251
 mevalonic acid-to-cycloartenol pathway, 1:250–251
- chemical structure, 1:250, 1:251f
- endogenous levels, 1:250
- occurrences, 1:250
- physiological characteristics, 1:252
- research background, 1:250
- signal transduction
 characteristics, 1:252–254
 downstream activators, 1:254
 receptor kinases, 1:253–254, 1:253f
- calpains, 1:354f
- copper-transporting ATPases (copper pumps), 3:671–672
- cytochrome *c*, 1:600, 1:600f
- cytokinins, 1:627–631
 background information, 1:627
 biosynthesis, 1:627–628
 chemical structure, 1:627, 1:628f
 functional roles
 biotechnological applications, 1:631
 cell cycle, 1:629
 pathogenicity, 1:631
 plant development and growth, 1:629–630
- metabolism
 catabolic reactions, 1:629
 conjugations, 1:628–629
 cytokinin oxidases/dehydrogenases (CKX), 1:629

- interconversions, 1:628
- signaling pathways
 - histidine kinases, 1:629, 1:630f
 - signal perception, 1:629
 - signal transduction, 1:629, 1:630f
 - two-component system (TCS)
 - pathway, 1:629, 1:630f
 - transport mechanisms, 1:629
- ferredoxins (Fd), 2:296
- flavins, 2:312
- glycine decarboxylase, 1:142–143
- glycolate oxidase, 2:312
- higher plant light-harvesting antenna
 - basic concepts, 2:726–727
 - pigment components, 2:727
 - protein components
 - LHC I organization, 2:727, 3:503–505, 3:505f
 - LHC II organization, 2:726–727, 2:727f, 3:505f
 - Lhca and Lhcb proteins, 2:726, 3:503–505, 3:505f
 - structural characteristics, 2:726
- lectins, 2:701, 2:702f, 2:702t, 2:703–704, 2:704t
- light harvesting, 2:178–183
 - antenna complexes, 3:484, 3:484f
 - basic concepts, 2:178–179, 2:179f
 - biological diversity, 2:182
 - chlorophyll and carotenoids, 3:492
 - energy dissipation, 3:493
 - excess light response, 2:182
 - general discussion, 2:178
 - higher plant light-harvesting antenna
 - basic concepts, 2:726–727
 - LHC I organization, 2:727, 3:503–505, 3:505f
 - LHC II organization, 2:726–727, 2:727f, 3:505f
 - Lhca and Lhcb proteins, 2:726, 3:503–505, 3:505f
 - pigment components, 2:727
 - protein components, 2:726, 3:503–505, 3:505f
 - nonphotochemical energy dissipation
 - energy-dependent quenching (qE), 2:180, 2:181f
 - quenched state characteristics, 2:181–182, 2:181f
 - quenching site, 2:180
 - xanthophyll cycle, 2:180–181, 2:181f
 - reaction centers, 2:178–179, 3:492
 - reactive oxygen species (ROS), 3:493
 - regulatory mechanisms, 2:179
 - state transitions, 2:179–180, 2:180f
 - structural characteristics, 2:179f
 - thylakoid superstructure function, 2:181f, 2:182
 - water oxidation, 3:493
- microRNA (miRNA) biogenesis, 3:99–100, 3:99f
- mitochondria, 1:145–148
 - alternative oxidases (AOXs)
 - basic concepts, 1:145–146
 - biochemical regulation, 1:147–148, 1:147f
 - physiological functions, 1:148
 - structural characteristics, 1:146–147, 1:147f
 - cyanide-resistant respiration, 1:145–146
 - mitochondrial DNA (mtDNA), 3:133t
 - mitochondrial electron transport chain, 1:141–142, 1:142f, 1:145, 1:146f
 - multisubunit centromeres, 1:447
 - nicotinic acid adenine dinucleotide phosphate (NAADP)-induced calcium ion (Ca^{2+}) release, 2:638t
 - nonphotochemical quenching (NPQ), 3:485–486
 - nonphotosynthetic plant-type FNRs
 - bacterial/apicomplexan FNRs, 2:301
 - root FNRs, 2:301
 - phospholipase D (PLD), 3:474
 - photosystem I (PSI) complex, 3:503–505, 3:505f
 - plant peroxidases, 3:407–409, 3:408f
 - plasma membrane proton (H^+)-ATPases, 3:688–692
 - biogenesis, 3:690, 3:690f
 - characteristics, 4:333f, 4:334
 - domain structures, 3:688–689
 - functional role, 3:688, 3:689f
 - P-type ATPases, 3:688
 - reaction mechanisms, 3:689–690, 3:689f
 - regulation mechanisms, 3:690–691
 - research background, 3:688
 - research summary, 3:691
 - pterin distribution, 3:666
 - uncoupling proteins (UCPs), 4:485
 - voltage-gated potassium channels, 4:557–559, 4:558f
- plasma amine oxidase, 1:89
- plasmacytoid DCs (pDCs), 4:398–399, 4:400f
- plasma membranes
 - calcium ion (Ca^{2+}) channel oscillations, 1:322–325
 - characteristics, 1:322, 1:323f
 - decoding methods, 1:325
 - functional role and significance, 1:323–325, 1:324f
 - signaling pathways, 1:322–323, 1:324f
- copper-transporting ATPases (copper pumps), 3:670–675
 - Atox-1 copper chaperone complex, 3:674
 - ATP7A/ATP7B, 3:670–671, 3:671f, 3:674
 - biological functions
 - copper homeostasis, 3:670–671, 3:671f
 - fungi, 3:671–672
 - human health functions, 3:670–671
 - occurrences, 3:670
 - plants, 3:671–672
 - prokaryotes, 3:671–672
 - characteristics and functional role, 4:333f, 4:334
 - functional role, 3:673
 - mammalian copper pump regulation
 - kinase-mediated phosphorylation, 3:674
- trafficking pathways, 3:674
- molecular structure
 - cytosolic catalytic domains, 3:673
 - domain structures, 3:672–673, 3:672f
 - N-terminal metal-binding domain (N-MBD), 3:674
 - transmembrane domains, 3:672–673, 3:672f, 3:673–674
- plasma membrane calcium (Ca^{+})-ATPase pump, 3:525–528
 - alternative splicing, 3:526–527, 3:527f
 - calcium ion (Ca^{2+}) pump reaction cycle, 4:331, 4:332f
 - domain structures, 3:525–526, 3:526f
 - genetic manipulations, 3:528–529
 - isoforms, 3:526–527, 3:527f
 - mechanoelectrical transduction (MET)
 - channels, 3:528, 3:528f
 - pathology, 3:528
 - physiological functions, 3:528, 3:528f
 - research summary, 3:529
 - tissue distribution, 3:526–527
- plasma membrane proton (H^+)-ATPases, 3:688–692
 - biogenesis, 3:690, 3:690f
 - characteristics, 4:333f, 4:334
 - domain structures, 3:688–689
 - functional role, 3:688, 3:689f
 - P-type ATPases, 3:688
 - reaction mechanisms, 3:689–690, 3:689f
 - regulation mechanisms, 3:690–691
 - research background, 3:688
 - research summary, 3:691
- plasma membrane V-ATPases, 4:520–521
- secretory pathway, 4:203–209
 - compartmental organization, 4:203, 4:204f
 - microtubule-based motor proteins, 4:208, 4:208f
 - plasma membrane proton (H^+)-ATPases, 3:690, 3:690f
- targeted proteins
 - characteristics, 4:203
 - constitutive secretion, 4:208
 - entry mechanisms, 4:203, 4:205f
 - integral membrane proteins, 4:204, 4:205f
 - protein exit, 4:207
 - regulated secretion, 4:208
 - signal recognition particles (SRPs), 4:203, 4:205f
 - soluble proteins, 4:204, 4:205f
- transport mechanisms
 - protein localization, 4:206–207, 4:206f
 - retinoid X receptor (RXR), 4:207
 - vesicular transport, 4:206, 4:206f
- yeasts, 3:690, 3:690f
- sodium–potassium ($\text{Na}^+ - \text{K}^+$)-ATPase pump, 3:681–687, 3:688–692
 - cardiac glycosides, 3:686–687, 3:686f, 3:687f
 - characteristics and functional role, 3:681, 4:331–333
 - domain structures, 4:333f

- plasma membranes (*Continued*)
 FXD proteins, 3:681, 3:682–685, 3:682*t*, 3:683*f*
 physiological functions, 3:685–686
 reaction mechanisms, 3:683, 3:682
 structural characteristics
 α subunits, 3:683–684, 3:684*f*
 β subunits, 3:683*f*, 3:684–685
 cation-pumping mechanism, 3:683–684, 3:684*f*
 domain structures, 3:683–684, 3:683*f*, 4:333*f*
 FXD proteins, 3:682–685, 3:683*f*
 schematic diagram, 3:684*f*
 subunits and isoforms, 3:683*f*
 subunits and isoforms, 3:681, 3:682*t*, 3:683*f*
 plasmepsin, 1:138*f*
 plasmids, 1:1
 plasmin, 1:443–444
 plasminogen, 3:91*t*
Plasmodium
 glycosylphosphatidylinositol (GPI)-anchored proteins, 2:478
 guanylyl cyclases (GCs), 1:567
 pyrimidine *de novo* synthesis, 3:714
Plasmodium berghei, 1:125
Plasmodium falciparum
 actin-related proteins (Arps), 1:37*f*
 aspartic proteases, 1:139
 EF-hand proteins, 1:319*t*
 protein N-myristoylation, 3:641
 pyrimidine *de novo* synthesis, 3:714
 plasmologens, 2:291*f*
 plastids, 1:509–510, 1:509*f*
 plastid terminal oxidase (PTOX), 1:146
 plastin, 1:34
 plastocyanin (PC), 3:530–532
 background information, 3:530
 cytochrome *f*
 complex formation, 3:531, 3:531*f*
 domain structures, 3:530–531
 electron flow pathways, 1:511–512, 1:512*f*
 photosynthetic processes, 2:179*f*, 3:493–494, 3:494*f*, 3:531–532
 photosystem I (PSI) complex
 domain structures, 3:531, 3:532*f*
 electron carrier proteins, 3:504–505
 electrostatic interactions, 3:532, 3:532*f*
 hydrophobic interactions, 3:531–532, 3:532*f*
 protonmotive electron transfer chain, 1:470*f*
 structural characteristics
 copper center, 3:530, 3:531*f*
 molecular surface
 cyanobacterial plastocyanins, 3:530
 green algae plastocyanins, 3:530
 higher plant plastocyanins, 3:530, 3:531*f*
 plastoquinol, 1:470*f*
 plastoquinone (PQ)
 coenzyme Q, 3:723–724, 3:723*t*
 photosynthesis, 2:179*f*, 3:493–494, 3:494*f*
 photosynthetic electron transport chain
 cytochrome *b₆f* complex, 4:335, 4:336*f*, 4:338–339, 4:339*f*
 electron flow pathways, 1:511–512, 1:512*f*
 long-term responses, 1:513–514
 redox state, 1:512–514, 1:514*f*
 short-term responses, 1:513, 1:513*f*
 state transitions, 1:513, 1:513*f*
 thylakoid protein kinase network, 1:514*f*
 protonmotive electron transfer chain, 1:470*f*
 platelet-activating factor receptor (PAFR), 3:533–537
 general discussion, 3:533
 G_q family proteins, 2:497*t*
 PAF acetylhydrolase, 3:533–534
 platelet-activating factor (PAF), 3:533–534, 3:534*f*
 regulation mechanisms, 3:535–536
 signal transduction, 3:536, 3:536*f*
 structural characteristics, 3:534–535, 3:535*f*
 transcript structure, 3:534–535, 3:535*f*
 platelet-derived growth factor (PDGF), 2:497*t*, 3:538–543
 functional role, 3:538
 G protein-coupled receptor kinases (GRKs), 2:485*t*
 isoforms, 3:538, 3:539*f*
 nuclear factor kappa B (NF- κ B), 3:303*t*
 PDGF receptors
 associated proteins, 3:539–540, 3:540*f*
 binding characteristics, 3:538–539, 3:539*f*
 chemotaxis
 characteristics, 3:542
 low molecular weight phosphatase (LMW-PTP), 3:542
 phosphatidylinositol 3-kinase (PI3K) pathway, 3:542
 phospholipase C- γ 1 (PLC γ 1), 3:542
 Ras-dependent pathways, 3:542
 developmental role, 3:543
 intracellular domains, 3:540*f*
 mediated cellular responses
 chemotaxis, 3:542
 general discussion, 3:540–542
 phosphatidylinositol 3-kinase (PI3K) pathway, 3:541, 3:541*f*
 phospholipase C- γ 1 (PLC γ 1), 3:540, 3:541*f*
 proliferation, 3:540–541, 3:541*f*
 Ras-dependent pathways, 3:540, 3:541*f*
 Src (sarcoma) family kinases, 3:540, 3:541*f*
 negative signal regulation
 degradation, 3:542–543
 general discussion, 3:542–543
 internalization, 3:542–543
 phosphatase-dependent
 dephosphorylation, 3:543
 signaling pathways, 3:539–540
 pleckstrin homology (PH) domain
 chemoattractant signaling, 2:236–238, 2:237*f*
 insulin receptors (IRs), 2:603–606
 phospholipase C (PLC), 3:467–468
 plectin, 2:624–630
 deficiency-related diseases, 2:627–628, 2:627*t*, 2:628*t*
 domain structures, 2:625, 2:625*f*, 2:626*f*
 expression regulation, 2:624–625
 gene locus, 2:624, 2:625*f*
 hemidesmosome function, 1:641–642, 1:641*t*
 integrin signaling, 2:610–611
 isoform diversity, 2:624, 2:626*f*
 microtubule-binding domain, 3:107
 molecular interactions, 2:625–627
 research background, 2:624–628
 structural characteristics, 1:640*f*, 1:641*t*
 subcellular localization, 2:624–625
 plurifunctional nucleolus, 3:325
 Pmel17 melanosome protein, 3:568–569
 podosomes, 2:610–611
Podospira anserina, 3:566, 3:569
 point centromeres, 1:447, 1:448*f*
 pokeweed mitogens (PWMs), 2:705
 polar lipids, 2:733, 2:734*f*
 polcalcin, 1:319*t*
 poly(A) binding protein II (PABPII), 3:189–191, 3:190*t*
 poly(ADPribose) polymerase (PARP), 1:176–177, 1:177*f*
 poly(ADPribose) polymerase-1 (PARP-1), 2:31–32, 2:85–86, 2:686
 polyamine oxidase, 1:88, 1:87–88
 polyamine synthesis, 2:215
 polyamino acid–matrix interaction affinity tags, 1:62–64
 poly(A) polymerase (PAP), 3:189–191, 3:190*t*, 3:559–560
 polybromo-associated BAF (PBAFs), 4:63
 polybromo-associated BAP (PBAP), 4:63
 polycystin, 1:319*t*
 polyethylene glycol (PEG), 1:374
 polymerase chain reaction (PCR) analyses, 2:750
 polymerases
 base excision repair (BER), 2:4*f*, 2:5
 DNA polymerases
 base excision repair (BER), 2:4*f*, 2:5
 biological effects, 2:79
 DNA polymerase I, 2:87–91
 base pairing, 2:89–90
 damage sensing mechanisms, 3:335
 domain structures, 2:87–88, 2:88*f*
 5' exonuclease, 2:90–91, 2:91*f*
 functional role, 2:54–56, 2:55*t*, 2:56*f*
 mismatch repair (MMR), 4:77
 O-helix, 2:88–89, 2:88*f*
 phosphoryl transfer reactions, 2:89, 2:89*f*
 reaction mechanisms, 2:89, 2:89*f*
 research background, 2:87
 structural domains, 2:87
 substrate binding, 2:88–89, 2:88*f*
 sugar specificity, 2:90
 3'–5' exonuclease, 2:90
 DNA polymerase II, 3:577–579, 3:579*f*

- DNA polymerase III holoenzyme, 2:92–95
 α subunit, 2:92–93, 2:93f
 bacterial species, 2:95
 β subunit, 2:93–94, 2:93f
 biological functions, 2:95
 DNA replication, 2:92, 2:93f, 2:707f, 2:709–710, 3:577–579, 3:579f
 DnaX complex, 2:95
 leading- and lagging-strand DNA synthesis, 2:94f, 2:95
 mismatch repair (MMR), 2:41f, 2:67, 2:67f, 2:95
 Pol III core, 2:92–93, 2:93f
 Pol III' subassembly, 2:93, 2:93f
 Pol III* subassembly, 2:93–94, 2:93f
 schematic diagram, 2:93f
 subunits and subassemblies, 2:92–95, 2:93f
 τ subunit, 2:93, 2:93f
- DNA polymerase IV, 2:709–710, 3:577–579, 4:480–481
- DNA polymerase V, 4:477–481
 bacterial homologs, 4:480–481
 cleavage reactions, 4:477–478, 4:478f
 eukaryotic homologs, 4:481
 fidelity, 4:480
 mutagenesis, 4:478
 processivity, 4:480
 regulatory mechanisms, 4:480
 research background, 4:479
 research summary, 4:481
 SOS response, 2:709–710, 4:477, 4:478f, 4:479f
 specificity, 4:479–480, 4:480f
 transcription regulation, 4:477, 4:478f
 translesion synthesis (TLS), 4:479, 4:479f
- DNA replication, 3:577, 3:578f, 4:236
 excision and repair synthesis, 2:56
 functional role, 2:55t, 2:56f
 GO system (MutM–MutT–MutY), 2:80f
 polymerase α -primase (Pol α), 2:241–244
 background information, 2:108, 2:241
 cell-cycle-dependent activity, 2:242
 checkpoint pathways, 2:243–244
 functional role, 2:241–242, 2:242f
 polymerase α -primase (Pol α)–helicase relationship, 2:242–243, 2:242f
 S phase initiation proteins, 1:402t
 structural characteristics, 2:241–242
- polymerase β (Pol β), 2:82–86
 background information, 2:82
 biological role, 2:82–83, 2:83f
 C-subdomain, 2:84, 2:85f
 DNA ligases, 2:30t
 domain organization, 2:83–85, 2:83f
 D-subdomain, 2:84, 2:85f
 fidelity, 2:97t
 5'-deoxyribose-5-phosphate lyase, 2:5, 2:6f
 lyase domain, 2:5, 2:83–84, 2:83f, 2:85f
- N-subdomain, 2:84–85, 2:85f
 polymerase b-base excision repair–protein interactions, 2:4f, 2:5, 2:85–86
 polymerase domain, 2:84–85
- polymerase δ (Pol δ), 2:245–248
 associated subunits, 2:245–246
 base excision repair (BER), 2:5, 2:247–248
 catalytic subunit, 2:245, 2:246f, 2:246t
 DNA recombination, 2:247
 DNA replication, 2:120, 2:246–247, 2:247f
 functional roles, 2:246
 mismatch repair (MMR), 2:51, 2:52t, 2:67f
 nucleotide excision repair (NER), 2:247–248
 processivity, 2:246
 proliferating cell nuclear antigen (PCNA), 2:246
 S phase initiation proteins, 1:402t
 structural domains, 2:245–246
- polymerase ϵ , 1:402t, 2:5, 2:108, 2:120
- polymerase γ (Pol γ), 2:97t
- polymerase η (Pol η)
 catalytic activity, 4:572–573, 4:573f
 characteristics, 4:572–574
 functional homologs, 4:481
 localizations, 4:573–574, 4:573f
 mutation effects, 4:574
 proliferating cell nuclear antigen (PCNA), 4:574
 ubiquitin-binding zinc finger (UBZ), 4:574
- polymerase ι , 2:5, 4:481
- polymerase κ , 4:481
- polymerase λ , 2:5, 2:7, 2:30t, 2:32
- polymerization kinetics, 2:96–100
 catalytic reactions, 2:98–99, 2:99f
 deoxynucleoside triphosphate (dNTP), 2:96, 2:98f, 2:99f
 exonuclease proofreading, 2:99–100, 2:99f
 fidelity, 2:97t, 2:98
 general discussion, 2:96
 incorporation measurements, 2:97–98, 2:97f, 2:97t
 polymerization sequence, 2:96, 2:97f
 research summary, 2:100
 selectivity, 2:96–97, 2:97t
 structural determinants, 2:98, 2:98f
 replication processes, 4:381–382, 4:382f
- reverse transcriptase (RT), 2:101–107
 biochemical characteristics, 2:103–104, 2:103f
 enzymatic functions, 2:102–103
 functional role, 2:104–105, 2:104f, 2:107
 integrase, 2:105–107, 2:106f
 integration processes, 2:105–107, 2:106f
 research background, 2:101
 retroviral genomes, 2:102, 2:102f
 retroviral replication, 2:101, 2:105–107
- retrovirus replication, 2:101
- ribonuclease H (RNase H) activity, 2:105, 2:106f
 subunit structure, 2:102f
- riboswitches, 4:137–138
- sliding clamp/polymerase interactions, 3:577, 3:578f, 4:238
- T7 DNA polymerase, 2:97t, 2:98, 2:98f
- translesion synthesis (TLS), 3:577–579, 3:579f, 4:479, 4:479f, 4:572–574, 4:573f
- Y-family polymerases, 4:477–481
 bacterial homologs, 4:480–481
 characteristics, 4:480, 4:572–574
 cleavage reactions, 4:477–478, 4:478f
 eukaryotic homologs, 4:481
 mutagenesis, 4:478
 regulatory mechanisms, 4:480
 research summary, 4:481
 SOS response, 4:477, 4:478f, 4:479f
 transcription regulation, 4:477, 4:478f
 translesion synthesis (TLS), 4:479, 4:479f
- polymerase accessory proteins, 2:119
- polymerase α -primase (Pol α), 2:241–244
 background information, 2:108, 2:241
 cell-cycle-dependent activity, 2:242
 characteristics, 2:119
 checkpoint pathways, 2:243–244
 DNA replication, 3:577–579, 3:579f
 functional role, 2:241–242, 2:242f
 polymerase α -primase (Pol α)–helicase relationship, 2:242–243, 2:242f
 S phase initiation proteins, 1:402t
 structural characteristics, 2:241–242
- polymerase β (Pol β), 2:82–86
 background information, 2:82
 biological role, 2:82–83, 2:83f
 C-subdomain, 2:84, 2:85f
 DNA ligases, 2:30t
 domain organization, 2:83–85, 2:83f
 D-subdomain, 2:84, 2:85f
 5'-deoxyribose-5-phosphate lyase, 2:5, 2:6f
 lyase domain, 2:5, 2:83–84, 2:83f, 2:85f
 N-subdomain, 2:84–85, 2:85f
 polymerase b-base excision repair–protein interactions, 2:4f, 2:5, 2:85–86
 polymerase domain, 2:84–85
- polymerase δ (Pol δ), 2:245–248
 base excision repair (BER), 2:5, 2:247–248
 DNA recombination, 2:247
 DNA replication, 2:120, 2:246–247, 2:247f
 eukaryotic DNA replication, 2:120
 functional roles, 2:246
 mismatch repair (MMR), 2:51, 2:52t, 2:67f
 nucleotide excision repair (NER), 2:247–248
 processivity, 2:246
 proliferating cell nuclear antigen (PCNA), 2:246
 S phase initiation proteins, 1:402t

- polymerases (*Continued*)
 structural domains
 associated subunits, 2:245–246
 catalytic subunit, 2:245, 2:246f, 2:246t
 polymerase ϵ , 1:402t, 2:5, 2:108, 2:120
 polymerase η (Pol η)
 catalytic activity, 4:572–573, 4:573f
 characteristics, 4:572–574
 functional homologs, 4:481
 localizations, 4:573–574, 4:573f
 mutation effects, 4:574
 proliferating cell nuclear antigen (PCNA), 4:574
 ubiquitin-binding zinc finger (UBZ), 4:574
 polymerase ι , 2:5, 4:481
 polymerase κ , 4:481
 polymerase λ , 2:5, 2:7, 2:30t, 2:32
 reverse transcriptase (RT), 2:101–107
 biochemical characteristics
 action mechanisms, 2:103, 2:103f
 fidelity, 2:96–97, 2:97t, 2:103
 inhibitors, 2:103–104, 2:106f
 processivity, 2:103
 recombination processes, 2:104
 replication errors/mutations, 2:103
 enzymatic functions, 2:102–103
 functional role, 2:104–105, 2:104f, 2:107
 integrase, 2:105–107, 2:106f
 integration processes, 2:105–107, 2:106f
 proteinase-activated receptors (PARs), 3:604
 research background, 2:101
 retroviral genomes, 2:102, 2:102f
 retroviral replication, 2:101, 2:105–107
 ribonuclease H (RNase H) activity, 2:105, 2:106f
 subunit structure, 2:102f
 riboswitches, 4:137–138
 RNA polymerases, 4:173–184
 bacterial gene transcription regulation, 4:167–172
 antibiotic inhibition, 4:171–172, 4:171f
 basic concepts, 4:232
 elongation cycle, 4:168f, 4:232–233
 initiation phase, 4:167–169, 4:168f
 nucleotide addition cycle (NAC), 4:170–171
 proofreading processes, 4:170–171
 sigma(σ)-factors, 4:167–169, 4:168f, 4:233–234, 4:233t
 transcription cycle, 4:167, 4:168f, 4:232–233
 transcription termination, 4:168f, 4:170
 bacterial RNA polymerase, 4:167–172
 bacterial transcription termination
 antitermination, 4:410
 characteristics, 4:407–410
 elongation complex, 4:408f
 elongation factors, 4:168f, 4:169–170, 4:409–410
 factor-dependent termination, 4:409
 hairpin formation, 4:407–409, 4:408f
 intrinsic termination, 4:407–409, 4:408f
 Mfd-mediated termination, 4:409
 regulated termination, 4:410
 Rho-mediated termination, 4:168f, 4:170, 4:409
 transcription-coupled DNA repair, 4:409
 cis-acting riboswitches, 4:250, 4:250f
 core enzyme structure
 characteristics, 4:174–176, 4:174f
 conformational flexibility, 4:176
 domain structures, 4:175–176, 4:175t
 nonconserved domains, 4:174f, 4:176
 RNAP channels, 4:175–176
 DNA-directed RNA polymerases, 4:355, 4:357f
 DNA transcription and replication processes, 2:162f
 elongation complex structure
 active center rearrangements, 4:180–182, 4:181f
 catalytic sites, 4:181f, 4:182–183
 general characteristics, 4:180–183
 nucleotide addition cycle (NAC), 4:181f, 4:183
 protein–DNA/RNA interactions, 4:180, 4:181f, 4:182t
 structural organization, 4:180, 4:181f
 translocation mechanisms, 4:181f, 4:183
 enzymatic activities, 4:173
 future research outlook, 4:183–184
 holoenzyme-promoter DNA complexes
 closed promoter complex (RPC), 4:178f, 4:179, 4:179t
 open promoter complex (RPO), 4:178f, 4:179–180, 4:179t
 structural characteristics, 4:178f, 4:179–180
 holoenzyme structure, 4:176–178, 4:177f
 lac operon, 2:695–696, 2:695f, 4:402
 mitochondrial gene expression, 3:141
 nucleoid-associated proteins (NAPs), 3:367–368
 poly(A) polymerase (PAP)-mediated decay, 3:54f, 3:55
 RNA polymerase I (Pol I), 2:366t, 2:367–368, 4:117–118, 4:407–409
 RNA polymerase II (Pol II), 2:365–368, 4:162–163
 bacterial intrinsic termination, 4:407–409
 conservation considerations, 2:367–368
 core promoters, 4:162–163
 C-terminal repeat domain (CTD), 2:365, 3:63–64, 3:63f
 1 α ,25-dihydroxyvitamin D₃ (1,25-(OH)₂D₃), 4:538f
 elongation phase, 2:366–367, 3:63–64, 3:63f
 encoding processes, 4:163
 GAL promoter system, 1:526f
 general transcription factors (GTFs), 4:163
 in vivo eukaryotic ribosome biogenesis, 4:117–118
 initiation phase, 2:366, 3:63–64, 3:63f, 4:163
 mobile clamps, 2:365
 ribonucleoside triphosphate (NTP), 4:162
 structural characteristics, 2:365, 2:366f, 2:366t
 structure and functions, 4:162–163
 transcription-coupled repair (TCR) pathway, 4:403–404, 4:404t
 transcription cycle, 2:365, 3:63–64, 3:63f
 transcription initiation, 4:163
 transcription termination, 2:367, 3:63–64, 3:63f, 4:163
 vitamin D receptor (VDR), 4:541f, 4:543–544
 RNA polymerase III (Pol III), 2:366t, 2:367–368, 3:322, 4:234, 4:407–409
 sigma(σ)-factors, 4:232–235
 bacterial gene transcription cycle, 4:232–233
 classifications, 4:233–234, 4:233t
 functional role, 4:233t, 4:234
 phylogenetic distribution, 4:234
 subunit composition, 4:173
 T7 RNA polymerase, 4:355–359
 applications, 4:359
 DNA replication priming, 4:359
 initial transcription reaction, 4:356f, 4:357–358
 pausing or terminator sites, 4:356f, 4:358, 4:358f
 polymerase sequence similarities, 4:355–357, 4:356f, 4:357f
 promoter release and elongation, 4:356f, 4:358
 promoter structure, recognition, and opening, 4:357, 4:358f
 regulation mechanisms, 4:356f, 4:358–359
 structural characteristics, 4:355, 4:356f
 structure–function relationship, 4:355, 4:357f
 transcription reaction/conformational change, 4:356f, 4:359
 transcription attenuation, 2:360–362, 2:361f
 transcription-coupled repair (TCR) pathway, 3:342f, 3:344, 4:402–403, 4:404t
 transcription cycle
 basic concepts, 4:167, 4:168f, 4:418
 elongation phase, 4:168f, 4:169–170, 4:173–174
 initiation phase, 4:167–169, 4:168f, 4:173–174
 Rho-mediated termination, 4:168f, 4:170
 sigma(σ)-factors, 4:167–169, 4:168f

- termination phase, 4:168f, 4:170, 4:173–174
- Tth* σ 70 holoenzyme
- core interactions, 4:176–178, 4:178f
 - domain structures, 4:176–178, 4:177f
 - general characteristics, 4:176–178
 - structural characteristics, 4:176, 4:177f, 4:178f
- UvrABCD system, 3:335–336, 3:335f
- sliding clamps, 2:119
- Y-family polymerases, 4:477–481
- bacterial homologs, 4:480–481
 - characteristics, 4:480, 4:572–574
 - cleavage reactions, 4:477–478, 4:478f
 - eukaryotic homologs, 4:481
 - mutagenesis, 4:478
 - regulatory mechanisms, 4:480
 - research summary, 4:481
 - SOS response, 4:477, 4:478f, 4:479f
 - transcription regulation, 4:477, 4:478f
 - translesion synthesis (TLS), 4:479, 4:479f
- polynucleotide phosphorylase (PNPase), 3:53–55, 3:54f, 3:55f, 3:56–57, 3:57f
- polyoma virus, 4:156–157
- polyoxyethylene headgroups, 1:645–646, 1:645f
- polysaccharides, 1:426
- polysialylated embryonic form of neural cell adhesion molecule (PSA–NCAM), 1:196–197, 1:197f
- polyubiquitination, 3:542–543, 3:619–620
- polyunsaturated fatty acids (PUFAs)
- carbohydrate-responsive element binding protein (ChREBP), 1:376
 - mitochondrial β -oxidation, 1:386
- Pompe disease, 2:427t, 2:431, 2:431t, 4:507–508, 4:507t
- ponsin, 2:320–321
- ponticulins, 3:438–439
- Porc proteins, 2:736–737
- pork, 4:533t
- porphobilinogen (PBG), 2:540–541, 2:540f
- porphyria cutanea tarda, 2:542t
- porphyrias, 2:542, 2:542t, 3:547, 3:548f
- porphyrins, 3:544–549
- biosynthetic deficiencies
 - herbicide effects, 3:547
 - porphyrias, 2:542, 2:542t, 3:547, 3:548f
- biosynthetic pathway
- δ -aminolevulinic acid (ALA) formation, 3:544, 3:545f
 - δ -aminolevulinic acid (ALA) formation, 2:540–541, 2:540f
 - chlorophyll biosynthesis, 1:501, 3:545f, 3:546, 3:546f
 - coenzyme B₁₂, 3:546–547, 3:546f
 - factor F₄₃₀, 3:546–547, 3:546f
 - general discussion, 3:544–547
 - heme biosynthesis, 2:539–540, 2:540f, 3:545f, 3:546, 3:546f
 - heme d₁, 3:546–547, 3:546f
 - schematic diagram, 3:545f
 - siroheme, 3:546–547, 3:546f
- uroporphyrinogen III formation, 2:540–541, 2:540f, 3:544–546, 3:545f, 3:546f
- catabolic reactions, 3:547–549, 3:548f
- characteristics, 3:544
- cytochrome *c*, 1:600–601
- regulation mechanisms
- basic concepts, 3:547
 - heme biosynthesis
 - δ -aminolevulinic acid (ALA) function, 3:547
 - non-ALAS enzymatic regulation, 3:547
 - spectral properties, 1:499–500, 1:500f
 - structural characteristics, 1:499, 1:499f
- Porphyromonas gingivalis*, 3:331f
- positive transcription elongation factor (P-TEFb), 4:164, 4:165–166, 4:165f
- post-Albers scheme, 3:683, 3:682
- postsynaptic density (PSD), 1:297–298, 1:298f
- potassium (K)
- gastric proton–potassium (H⁺–K⁺)-ATPase, 3:676–680
 - acid secretion, 3:679–680
 - characteristics, 3:676, 4:333–334, 4:333f
 - functional residues, 3:678–679
 - structural characteristics
 - α -subunit, 3:676, 3:677f
 - β -subunit, 3:677, 3:677f
 - conformational changes, 3:677f, 3:678
 - kinetic mechanisms, 3:678
 - oligomeric structure, 3:677–678
 - inwardly rectifying potassium channels (K_{ir}), 2:649f, 2:650, 2:651f, 4:557–559, 4:558f
 - mitochondrial potassium channels, 3:550–553
 - ATP-dependent potassium ion (K⁺) channel (KATP), 3:551, 3:551f
 - calcium-dependent potassium channels, 3:551–552, 3:551f
 - characteristics and functional role, 3:550–553
 - MitoKv1.3, 3:551f, 3:552
 - potassium (K⁺) cycle, 3:550–551, 3:551f
 - research summary, 3:552
 - two-pore potassium channel (TASK-3), 3:551f, 3:552
 - potassium-chlorine (K-Cl) cotransporter 2 (KCC2), 1:199, 1:200f
 - sodium-potassium-chlorine (Na-K-Cl) cotransporter isoform 1 (NKCC1), 1:199, 1:200f
 - sodium–potassium (Na⁺–K⁺)-ATPase pump, 3:681–687, 3:688–692
 - cardiac glycosides, 3:686–687, 3:686f, 3:687f
 - characteristics and functional role, 3:681, 4:331–333
 - domain structures, 4:333f
 - FXYD proteins, 3:681, 3:682–685, 3:682t, 3:683f
 - physiological functions, 3:685–686
 - reaction mechanisms, 3:683, 3:682
- structural characteristics
- α subunits, 3:683–684, 3:684f
 - β subunits, 3:683f, 3:684–685
 - cation-pumping mechanism, 3:683–684, 3:684f
 - domain structures, 3:683–684, 3:683f, 4:333f
 - FXYD proteins, 3:682–685, 3:683f
 - schematic diagram, 3:684f
 - subunits and isoforms, 3:683f
 - subunits and isoforms, 3:681, 3:682t, 3:683f
- two-pore-motif potassium channels (K_{2P}), 2:649f, 2:650–651
- voltage-dependent potassium ion (K⁺) channels, 4:555–559
- classifications, 4:558f
 - diversity, 4:555
 - physiological functions
 - inwardly rectifying potassium channels (K_{ir}), 4:557–559, 4:558f
 - voltage- and Ca²⁺-activated seven-transmembrane K⁺ channels (K_{Ca}), 4:558f, 4:559
 - voltage-gated six-transmembrane K⁺ channels, 4:556–557, 4:558f
 - structural characteristics, 4:555–556, 4:556f
- voltage-gated potassium channels, 1:379, 1:380f, 2:649, 2:649f, 2:746–747, 2:746f, 4:558f
- potatoes, 4:533t
- poxviral topoisomerase, 2:158t
- PP1 (calcium/calcineurin-dependent type 1 serine/threonine protein phosphatase), 1:312, 4:220, 4:221f, 4:222
- Prader-Willi syndrome, 4:121
- prazosin, 1:58, 2:497t
- Prb1p protein, 3:566
- precatenanes, 2:162, 2:162f
- prediabetes, 1:654
- pregnane X receptor (PXR), 1:168, 1:169f
- Preiss–Handler pathway, 1:172, 1:174f
- pre-messenger RNAs (pre-mRNAs), 3:188–193
- alternative splicing, 1:75, 1:76f, 3:60–61, 4:187–188
- maturation processes
- cleavage sites, 3:188–189
 - exon definition model, 3:61–62, 3:61f
 - gene expression regulation, 3:192
 - polyadenylation, 3:60–61, 3:188
 - protein complexes, 3:189–191, 3:189f
 - regulation mechanisms, 3:191–192
 - 3'-end maturation, 3:60–61, 3:189–191, 3:189f, 3:190t
- RNA splicing
- alternative splicing, 3:60–61, 4:187–188
 - ancillary RNA-processing signals, 3:62, 3:62f
 - chemical mechanisms, 4:185, 4:186f
 - disease connections, 4:189–190
 - evolutionary implications, 4:189
 - exon definition model, 3:61–62, 3:61f
 - exonic splicing silencers (ESSs), 3:62f

- pre-messenger RNAs (pre-mRNAs)
(*Continued*)
exons, 3:142–144, 3:143f, 4:185,
4:186–187, 4:186f, 4:187f
5' and 3' splice sites, 3:59–60, 3:60f,
4:186, 4:186f, 4:187f, 4:188f
genetic information transfer, 4:185,
4:186f
historical research, 4:185
intronic splicing silencers (ISSs), 3:62f
introns, 3:142–144, 3:143f, 4:185,
4:186–187, 4:186f, 4:187f
life cycle, 3:62–63
mosaic mitochondrial genes,
3:142–144, 3:143f
research summary, 4:190
ribozymes, 4:188
self-splicing reactions, 4:188, 4:188f
spliceosomes, 3:62–63, 3:63f, 4:187
splicing consensus sequences, 4:186,
4:187f
trans-splicing reactions, 4:188–189, 4:189f
prenylation, 2:737–738, 2:737t, 3:608–609,
3:608f
prenyltransferase, 4:242, 4:244f
preprotachykinin-A (PPT-A), 4:360, 4:361f
preprotachykinin-B (PPT-B), 4:360, 4:361f
preprotachykinin-C (PPT-C), 4:360, 4:361f
presenilins
amyloid precursor protein (APP), 1:474,
4:201–202
calbindin-D_{28k}, 1:286
 γ -secretase-based mediation, 4:53
Golgi-localized endoproteases, 2:481
PriA protein, 4:24
PRIDE database, 3:615t
primary biliary cirrhosis (PBC), 3:561–565
anti-mitochondrial antibodies (AMAs),
3:114, 3:115–116
background information, 3:561
clinical features, 3:563–564
diagnosis, 3:564
etiologic
associated autoimmune diseases, 3:562
autoantibodies/antimitochondrial
antibodies, 3:561–562
mitochondrial antigens, 3:562
laboratory tests, 3:564
pathology
developmental stages, 3:563
stage III PBC, 3:563f
stage II PBC, 3:563f
stage I PBC, 3:563f
stage IV PBC, 3:564f
pathophysiology, 3:562–563
prevalence, 3:561
symptom treatments
anemia, 3:564–565
fatigue, 3:564
osteoporosis, 3:564–565
pruritus, 3:564
vitamin malabsorption, 3:564
underlying disease process treatments
general discussion, 3:565
liver transplantation, 3:565
medical treatments, 3:565
primates, 1:554
prion protein cerebral amyloid angiopathy,
1:477t
prions, 3:571–575
degradation resistance, 3:571
epidemiology and diagnosis, 3:575
pathophysiology, 3:574
physiological functions, 3:574
prion protein (PrP)
amyloid deposits, 1:101t, 1:102, 3:575t
conformational transitions, 3:572–574,
3:573f
exosomes and microvesicles (EMVs),
2:264
genetic criteria, 3:566
isoforms, 3:571, 3:572t
pathophysiology, 3:574
physiological functions, 3:574
propagation and amplification model,
3:571–572, 3:573f
structural characteristics, 3:572f
therapeutic treatments, 3:575
prion strains, 3:574
protein misfolding diseases, 3:574–575,
3:575t
research background, 3:571
scrapie prions, 3:574–575
synthetic prions, 3:573–574
therapeutic treatments, 3:575
yeasts and fungi, 3:566–570
chaperone functions, 3:569
disease states, 3:568–569
functional domains, 3:566–567,
3:568f
generation mechanisms, 3:569
genetic criteria, 3:566, 3:567f
known prions, 3:566
prion domains, 3:566–567, 3:568f
research summary, 3:569–570
self-propagating amyloidoses,
3:567–568, 3:568f
procarbazine, 2:51–52
profilaggrin, 1:319t
profilin, 1:14f, 1:15, 1:27–28, 1:28f, 1:29–31,
1:30f
progeria, 3:313, 3:313f
progesterone
characteristics and functional role,
4:308–309
progesterone receptor (PR)-binding sites,
3:206–208, 3:206f, 3:208f
programmed cell death
See cell death
progressive osseous heteroplasia (POH),
2:519–520
prokaryotes
actin-related proteins (Arps), 1:36,
1:37–38, 1:39f
anoxygenic photosynthesis, 2:209
copper-transporting ATPases (copper
pumps), 3:671–672
cytochrome *c*, 1:599
DNA replication, 2:108–109, 2:109f
hexokinases, 2:544
membrane-associated energy transduction,
2:204–209, 3:28–35
adenosine diphosphate (ADP)
phosphorylation, 2:204, 2:205f
archaeal energy transduction
light utilization, 3:33–34, 3:34f
methanogenesis, 3:32–33, 3:33f
respiratory complexes, 3:34
background information, 3:28
catabolic reactions, 2:204
energy-transducing functions
anaerobic respiration, 2:206–208,
2:207f, 2:208f, 3:29, 3:29t, 3:30f,
3:31f
basic concepts, 3:29–30
denitrification, 3:29–30
dicarboxylic acid decarboxylation,
2:208, 2:208f
electrogenic transport, 2:208–209,
2:209f
electron transport phosphorylation,
2:206–208, 2:207f, 2:208f
fermentations, 3:29
ferric iron reduction, 3:30
fumarate respiration, 3:30, 3:30f
H⁺ potential generation, 2:208–209,
2:209f
H⁺ translocation, 2:206–208, 2:208f
redox reactions, 2:206–208, 2:207f,
2:208f
schematic diagram, 3:29f, 3:31f
Na⁺ translocation, 2:206–208, 2:208f
sulfur respiration, 3:30
evolutionary development, 3:34
methanogens, 3:32–34
oxygen respiration
aerobic chemolithotrophs, 3:31–32
aerobic organotrophs, 3:30–31, 3:31f
phototrophic bacteria
anoxygenic photosynthesis, 2:209,
2:335, 3:32, 3:693–694, 3:694f,
4:81
antenna complexes, 3:32, 3:484,
3:484f
evolutionary development, 3:34
general discussion, 3:32
oxygenic photosynthesis, 3:32
plasma membranes, 3:28
proton potential, 2:204, 2:205f,
2:206–208
secondary energy transducers, 3:34
non-homologous end joining (NHEJ),
3:267–268, 3:267f
protein biosynthesis elongation
background information, 2:249
components, 2:250t
small RNAs (sRNAs), 4:253–255
supercoils
characteristics, 2:161
stress generation mechanisms
hyperthermophilic eubacteria,
2:161–162
mesophilic/psychrophilic eubacteria,
2:161
translation initiation, 4:425–431
basic concepts, 4:425
initiation complex formation,
4:426–427

- initiation factors
 - initiation factor 1 (IF1), 4:427–428, 4:427f, 4:428f
 - initiation factor 2 (IF2), 4:427f, 4:428–430
 - initiation factor 3 (IF3), 4:427f, 4:428f, 4:429f, 4:430
 - structure and functions, 4:427–431
- 70S initiation complex, 4:130f, 4:426–427, 4:427f
- 30S initiation complex, 4:130f, 4:132f, 4:426–427, 4:427f, 4:429f
- translation initiation region (TIR), 4:425–426, 4:426f
- type IIA DNA topoisomerase
 - DNA gyrase, 2:165–166
- type IV DNA topoisomerase, 2:166
- prolactin, 2:526–527, 2:527f
- proliferating cell nuclear antigen (PCNA)
 - base excision repair (BER), 2:85–86
 - DNA ligases, 2:30, 2:30t
 - DNA methyltransferases (MTases), 3:8–9, 3:8f
 - DNA mismatch repair (MMR)
 - functional role, 2:52t
 - molecular mechanisms, 2:61, 2:61f
 - strand discrimination, 2:70–71, 2:70f
 - DNA replication, 1:402t, 2:119, 3:576, 3:577f
- polymerase δ (Pol δ), 2:246
- polymerase η (Pol η), 4:573–574, 4:573f
- proline
 - acyl-coenzyme A synthetase, 2:291f
 - anaplerotic pathways, 1:106f
 - ascorbic acid, 4:531–532
 - classifications, 1:91, 1:92t
 - collagens
 - signature (Gly–XY–)_n sequence
 - basic concepts, 1:545
 - evolutionary origins, 1:548
 - schematic diagram, 1:546f, 1:547f
 - self-assembly control, 1:546
 - elastin, 2:192
 - functional roles, 1:93t
 - metalloproteases, 3:87–88
 - O-linked glycosylation, 2:465
 - phage libraries, 3:434t
 - proline–histidine–serine–arginine–asparagine (PHSRN) sequence, 1:426
 - proline-rich tyrosine kinase, 2:345f
 - proline-rich tyrosine kinase 2 (PYK2)
 - activation and signaling mechanisms, 2:266–267, 2:267f
 - cellular responses, 2:267
 - characteristics, 2:266–267, 2:266f
 - domain conservation, 2:266t
 - tricarboxylic acid cycle, 4:439f
 - urea cycle, 4:495f
 - vasopressin/oxytocin family, 4:516t
- promyelocytic leukemia nuclear bodies (ND 10), 3:298–299, 3:299f
- proopiomelanocortin (POMC)
 - energy homeostasis, 3:26, 3:77f
 - functional role, 3:24–25, 3:25t, 3:26f
 - ghrelin activity, 1:118
 - insulin metabolism, 2:603f, 2:606
 - leptin expression, 2:257–258, 2:260f
 - melanocortin neurons, 3:77
- propanol, 2:204–206
- propionate
 - anaerobic respiration, 2:208f
 - anaplerotic pathways, 1:106f
 - fermentation reactions, 2:204–206, 2:205t, 2:206f
 - gluconeogenesis, 2:381
- Propionibacterium*, 2:205t
- Propionibacterium freudenreichii* subsp. *shermanii* TC (transcarboxylase), 1:222, 1:223f
- Propionibacterium shermanii*, 4:353
- propionic acid, 2:204–206, 2:205t
- propionic acidemia, 4:495t
- Propionigenium modestum*, 2:208, 2:208f, 3:29
- propionyl-coenzyme A (CoA), 2:327
- propionyl-coenzyme A (CoA) carboxylase (PCC), 1:109f
- proplastids, 1:509–510, 1:509f
- propranolol, 1:59
- pro- σ^K , 4:51f, 4:53
- prostaglandin
 - adenyl cyclases (ACs), 1:46t
 - B2 receptors, 1:242f
 - eicosanoid receptors, 2:189–190, 2:190f
 - G_q family proteins, 2:496–497, 2:497t
 - platelet-activating factor receptor (PAFR), 3:536, 3:536f
 - prostanoid acid, 2:189
- prostanoids, 3:581–584
 - action mechanisms, 3:582–583
 - biosynthetic pathway, 3:581
 - biosynthetic relationships, 3:581, 3:582f
 - catabolic reactions, 3:582
 - dioxygenases, 3:371–372
 - metabolic pathways, 3:582
 - prostaglandin endoperoxide H synthase (PGHS)
 - biosynthetic pathway, 3:581
 - biosynthetic relationships, 3:582f
 - catalysis and inhibition, 3:581–582, 3:582f
 - PGHS-1/PGHS-2 gene expression, 3:582
- prostate
 - fatty acid metabolism–cancer relationship, 2:277–278
 - fibroblast growth factor 6 (FGF6), 2:306
 - mammalian glucose transport, 2:392t
 - proteinase-activated receptors (PARs), 3:605t
 - steroid/thyroid hormone receptors, 4:311–312
- proteases
 - aspartic proteases, 1:137–140
 - disease processes
 - acquired immuno-deficiency syndrome (AIDS), 1:139
 - Alzheimer's disease, 1:140
 - cancer, 1:139
 - fungal infections, 1:139–140
 - general discussion, 1:139
 - hypertension, 1:139
 - malaria, 1:139
 - genomic sequences, 1:137, 1:138f
 - historical research, 1:137
 - naturally occurring inhibitors
 - general discussion, 1:140
 - IA-3, 1:140
 - pepstatin, 1:140
 - PI-3, 1:140
 - structural characteristics
 - catalytic residues, 1:138–139
 - flap structure, 1:140
 - overall structure, 1:138, 1:138f
 - proenzyme structure, 1:139
 - research background, 1:137–139
 - synthetic inhibitors, 1:140
- bacterial adenosine triphosphate (ATP)-dependent proteases, 3:623–624
- blood-clotting proteases, 3:585–589
 - coagulation cascade
 - attenuation mechanisms, 3:587
 - extrinsic pathway, 3:585, 3:586f
 - factor XIII–factor XIIIa conversion, 3:586f, 3:587
 - fibrin polymerization, 3:586f, 3:587
 - intrinsic pathway, 3:585, 3:586f
 - platelet activation, 3:586f, 3:587
 - thrombin, 3:586f, 3:587
 - structural and functional characteristics
 - gamma-carboxyglutamic acid (Gla) domains, 3:588–589, 3:589f
 - trypsin/trypsinogen, 3:587–588, 3:588f
 - thrombin formation, 3:585, 3:586f
- HIV protease, 2:554–556
 - definition, 2:554
 - drug-resistant variants
 - HIV reverse transcriptase (RT), 2:555–556
 - research objectives, 2:556
 - structural characteristics, 2:556
- inhibitor binding
 - binding modes, 2:555, 2:556f
 - research background, 2:555
- as model system, 2:556
- structural characteristics
 - Gag/Pol polyproteins, 2:554–555
 - general discussion, 2:554
 - isolation and crystallization, 2:555
 - three-dimensional structure, 2:555f
 - viral protein assembly, 2:554–555
- metalloproteases, 3:86–89
 - AAA (ATPases associated with various cellular activities) proteins, 1:3–5, 1:3t, 1:4f, 1:6
 - characteristics, 3:86
 - FtsH metalloprotease, 1:3t, 1:4f, 1:6
 - functional classification, 3:86
 - genetic origin, 3:88
 - inhibitors
 - catalytic chelators, 3:88f, 3:89
 - general discussion, 3:88–89
 - metal-bound water displacement, 3:88f, 3:89
 - metal hydroxide complexes, 3:88f, 3:89
 - ligand nature, 3:88
 - physiological functions, 3:88

- proteases (*Continued*)
 regulated intramembrane proteolysis (RIP), 4:50–53
 structural classification
 catalytic sites, 3:87, 3:87f
 cocatalytic sites, 3:87–88
 metal sites, 3:86–87, 3:87f, 4:50–53
 research background, 3:86–88
 structural sites, 3:87
 photoinhibition, 3:484
 protease-activated receptors (PARs), 1:444, 3:586f
 protease calpains, 2:320f, 2:321
 serine proteases, 2:758–759, 2:758f, 3:585–589
 three-dimensional (3D) cell migration and invasion
 degradation processes, 1:443–444
 ECM formation, 1:444
 endostatin, 1:444
 growth factor activation, 1:444
 proteolytic fragments, 1:444
 proteasomes, 3:590–594
 basic concepts, 3:590
 Hsp90 proteasome inhibitor, 3:593, 3:593f
 inhibitors, 3:593, 3:593f
 P131 proteasome inhibitor, 3:593
 proteasome activating nucleotidase (PAN), 3:599, 3:623–624
 proteasome activators
 characteristics, 3:592–593
 ecm29p, 3:592–593
 11S regulators (REG α , REG β , REG γ)/PA28s, 3:592, 3:593f
 PA200, 3:592, 3:593f
 proteasome activating nucleotidase (PAN), 3:599, 3:623–624
 schematic diagram, 3:593f
 research summary, 3:593
 20S proteasome
 assembly processes, 3:596–597
 core particle (CP) gate, 3:597
 enzyme mechanism and inhibitors, 3:590
 immunoproteasome, 3:596
 N-terminal active site, 3:596, 3:596f
 occurrences, 3:595–597
 structural characteristics, 3:590, 3:591f, 3:595–596, 3:596f
 tymoproteasome, 3:596
 26S proteasome, 3:595–600
 basic concepts, 3:590–592, 3:591f
 cell regulation
 chaperones, 3:599
 E3 ubiquitin ligase, 3:599
 ER-associated degradation (ERAD), 2:200, 2:202f
 substrate degradation, 3:599
 ubiquitin proteasome pathway (UPP), 3:599
 characteristics, 3:595
 enzyme mechanism, 3:591, 3:592f
 19S regulatory complex
 assembly processes, 3:598
 base complex, 3:597f, 3:598
 basic concepts, 3:590–592
 lid complex, 3:597f, 3:598
 regulatory mechanisms, 3:598–599
 structural characteristics, 3:597–598, 3:597f
 physiological importance, 3:592
 protein degradation
 characteristics and functional role, 3:620–621, 3:621f
 degradation mechanisms, 3:622
 19S proteasome regulatory particle, 3:621, 3:621f
 20S proteasome core particle, 3:620–621, 3:621f, 4:473–474, 4:474f
 research summary, 3:599–600
 structural characteristics, 3:597f
 tissue and subcellular distribution, 3:591–592
 ubiquitination, 4:473–474, 4:474f
 ubiquitin–proteasome degradation pathway, 3:619–622, 3:619f, 4:473–474, 4:474f
 ubiquitin–proteasome degradation pathway
 degradation mechanisms, 3:619–622, 3:619f, 4:473–474, 4:474f
 disease connections, 3:622
 hybrid proteasome, 3:621–622, 3:621f
 immunoproteasome, 3:621–622
 mitochondria, 3:622–623, 3:623t
 signal transduction, 3:304f
 26S proteasome
 characteristics and functional role, 3:620–621, 3:621f
 degradation mechanisms, 3:622
 19S proteasome regulatory particle, 3:621, 3:621f
 20S proteasome core particle, 3:620–621, 3:621f
 ubiquitin conjugation
 deubiquitination, 3:620
 polyubiquitin chain formation, 3:619–620
 substrate recognition mechanisms, 3:620
 protein arginine methyltransferases (PRMTs), 4:61
 proteinases
 aspartic proteases, 1:137–140
 disease processes
 acquired immunodeficiency syndrome (AIDS), 1:139
 Alzheimer's disease, 1:140
 cancer, 1:139
 fungal infections, 1:139–140
 general discussion, 1:139
 hypertension, 1:139
 malaria, 1:139
 genomic sequences, 1:137, 1:138f
 historical research, 1:137
 inhibitors
 IA-3, 1:140
 naturally occurring inhibitors, 1:140
 pepstatin, 1:140
 PI-3, 1:140
 synthetic inhibitors, 1:140
 structural characteristics
 catalytic residues, 1:138–139
 flap structure, 1:140
 overall structure, 1:138, 1:138f
 proenzyme structure, 1:139
 research background, 1:137–139
 matrix metalloproteinases (MMPs), 3:90–97
 activation mechanisms
 chemical/physical agents, 3:95
 proteinases, 3:95, 3:95f
 zymogens (proMMPs), 3:94
 biological functions, 3:96
 characteristics, 1:543
 characteristics and functional role, 3:90
 classifications, 1:543t
 DNA polymorphisms, 3:96
 domain structures, 3:90, 3:92f, 3:93–94, 3:93f
 enamelysin, 3:91t, 3:92f, 3:93
 endogenous inhibitors
 α -2 macroglobulins, 3:96
 tissue inhibitors of metalloproteinases (TIMPs), 3:96
 epilysin, 3:91t, 3:92f, 3:93
 families/subfamilies
 characteristics, 3:90
 collagenases, 3:90–91, 3:91t, 3:92f
 gelatinases, 3:91t, 3:92–93, 3:92f
 matrilysins, 3:91t, 3:92f, 3:93
 membrane-type MMPs, 3:91t, 3:92f, 3:93
 stromelysins, 3:91t, 3:92f, 3:93
 functional roles, 1:543–544
 genetic disorders, 3:96
 historical research, 1:542
 inhibition mechanisms, 3:96
 metalloelastase, 3:91t, 3:92f, 3:93
 occurrences, 3:90
 peptide bond hydrolysis mechanisms, 3:94, 3:94f
 substrate specificity, 3:95–96
 tissue expression, 3:92f
 tissue inhibitors of metalloproteinases (TIMPs), 3:93f, 3:95f, 3:96
 vertebrate matrixins, 3:90, 3:91t
 proteinase-activated receptors (PARs), 3:601–606
 activating and inactivating proteinases, 3:601, 3:602f
 activation mechanisms, 3:601, 3:602f
 agonists, 3:601–602, 3:603t
 antagonists, 3:603–604, 3:603f
 bioassays
 calcium signaling measurement, 3:604
 platelet aggregation, 3:605
 tissue bioassays, 3:604–605
 desensitization mechanisms, 3:602
 detection methods, 3:604
 distribution, 3:604, 3:605t
 domain structures, 3:601, 3:603t
 intermolecular activation, 3:602–603
 internalization, 3:602
 localizations, 3:605t
 PAR-activating peptides (PAR-APs), 3:601–602, 3:602f, 3:603t

- physiological functions, 3:605
 receptor dimerization, 3:602–603
 resensitization, 3:602
 thrombin, 3:601, 3:602*f*, 3:603*t*
- protein disulfide isomerase (PDI), 1:291*t*, 1:293, 1:665, 1:665*f*, 2:448–449, 4:205–206
- Protein Information Resource (PIR), 3:611–612
- protein kinase A (PKA)
 adipocyte development, 1:55
 binding specificity and affinity, 1:577–578
 carbohydrate-responsive element binding protein (ChREBP), 1:375–376, 1:376*f*
 characteristics and functional role, 1:575–577
 feed-forward control, 1:578–579, 1:578*f*
 glycogen synthase kinase-3 (GSK3), 2:435*f*
 G protein-coupled receptor kinases (GRKs), 2:485*t*
 insulin-secreting cells (β -cells), 2:597*f*
 liver PK (LPK) *in vitro* regulation, 3:719
 negative feedback, 1:578–579, 1:578*f*
 neurogenesis, 1:199
 Perilipin 1 (Plin1), 1:47–48, 1:48*f*
 phosphofructokinase-2/fructose biphosphatase-2 (PFK-2/FBPase-2)
 activation, 3:457–458
 PKA-mediated lipolysis, 1:49–51, 1:50*f*, 4:303–304
 plasma membrane calcium (Ca^{2+})-ATPase pump, 3:525–526
 structural domains, 1:574–575, 1:575*f*, 1:576*f*
 substrate specificity, 1:578
- protein kinase B (PKB)
 functional role, 3:463–464
 glycogen synthase kinase-3 (GSK3), 2:435*f*
 insulin signal transduction, 2:604
- protein kinase C (PKC), 3:637–640
 background information, 3:637
 bile salts, 1:169
 calcium ion (Ca^{2+}) signaling system, 2:719–720, 2:720*f*
 calcium oscillation decoding, 1:325
 calcium-sensing receptor (CaSR), 2:583, 2:584*f*
 chemokine receptors, 1:480–482, 1:482*f*
 diabetes, 1:656
 domain structures, 3:637, 3:638*f*
 downregulation, 3:639
 G_{bg} subunit functions, 1:480–482, 1:482*f*, 4:265
 glycogen synthase kinase-3 (GSK3), 2:435*f*
 G protein-coupled receptor kinases (GRKs), 2:485*t*
 G_q family proteins, 2:497–498, 2:498*f*
 historical research, 3:637
 insulin resistance, 1:650
 insulin-secreting cells (β -cells), 2:597*f*
 insulin signal transduction, 2:604
 isozyme families, 3:637, 3:638*f*
 mitochondrial permeability transition pore, 3:165–166
 neurogenesis, 1:199
 phosphorylation, 3:637–638, 3:638*f*
- plasma membrane calcium (Ca^{2+})-ATPase pump, 3:525–526
- platelet-activating factor receptor (PAFR), 3:536, 3:536*f*
- signaling pathways, 3:639–640
- somatostatin (sst) receptors, 4:265
- store-operated calcium (SOC) channels, 3:359–360, 3:360*f*
- subcellular localization, 3:639
- T-cell receptors (TCRs), 4:374, 4:377–378
- translocation cycle, 3:638–639, 3:639*f*
- protein O-fucosyltransferase I (Pofut1), 2:453, 2:454*f*, 2:456
- protein O-mannosyltransferase (POMT), 2:454, 2:454*f*
- protein phosphatase 1 nuclear-targeting subunit (PNUITS), 1:312
- proteins
- AAA (ATPases associated with various cellular activities) proteins, 1:1–6
- AAA+ superfamily, 1:3–5, 1:3*t*
- basic concepts, 1:1
- cellular functions
- membrane-bound metalloproteases, 1:6
- membrane fusion, 1:5–6
- proteasome-dependent degradation, 1:5
- vesicle transport, 1:5–6
- classifications, 1:3–5, 1:3*t*, 1:4*f*
- dyneins, 2:184–187
- axonemal dyneins, 2:185*t*, 2:186–187
- characteristics and classifications, 2:184, 2:185*t*
- cytoplasmic dynein 1, 2:185–186, 2:185*t*, 2:186*f*
- cytoplasmic dynein 2 (IFT dynein), 2:185*t*, 2:186
- dynein heavy chain, 2:184–185, 2:185*t*, 2:186*f*
- dynein intermediate chain, 2:185*t*, 2:186*f*
- dynein light chain, 2:185*t*, 2:186*f*
- inner arm dyneins, 2:185*t*, 2:187
- outer arm dyneins, 2:185*t*, 2:187
- research summary, 1:6
- structure and functions, 1:1–2, 1:2*f*, 1:3*f*
- affinity tags, 1:61–65
- functional roles, 1:61
- fusion protein construction
- affinity–fusion gene constructs, 1:61
- general discussion, 1:61
- linker peptide composition, 1:61
- production and immobilization processes, 1:61–62
- selection criteria, 1:61
- system characteristics
- antibody–epitope interaction affinity tags, 1:64
- basic concepts, 1:64–65
- polyamino acid–matrix interaction affinity tags, 1:62–64
- protein–ligand interaction affinity tags, 1:62
- protein–protein interaction affinity tags, 1:64–65
- amyloid proteins, 1:103
- Arf-related proteins, 1:134
- biotinylation, 1:226–228
- analytical and clinical applications, 1:226
- basic concepts, 1:226
- biosynthesis, 1:226
- biotinylated antibodies, 1:227–228
- chemical synthesis, 1:227, 1:227*f*
- medical importance, 1:226–227
- calcium/calmodulin-dependent protein kinase II (CaMKII), 1:297–303
- autophosphorylation
- autonomous activity, 1:299
- calmodulin (CaM) trapping, 1:300
- neuronal activity monitoring, 1:301, 1:301*f*
- structure and functions, 1:299, 1:300*f*
- substrate anchoring, 1:300
- threonine ($\text{Thr}^{305}/\text{Thr}^{306}$)
 autophosphorylation, 1:300–301
- cardiac regulation, 1:302
- characteristics and functional role, 1:297
- domain structures, 1:298–299, 1:299*f*
- intracellular localization, 1:297–298, 1:298*f*
- long-term potentiation, 1:302
- memory and learning regulation, 1:302
- structure and functions, 1:302–303
- substrates, 1:297–298, 1:298*f*
- calcium ion (Ca^{2+})–calmodulin (CaM)-dependent protein kinases (CaMKs), 1:304–309
- broad-specificity calcium ion (Ca^{2+})–calmodulin (CaM)-dependent protein kinases (CaMKs)
- calcium ion (Ca^{2+})–calmodulin (CaM)-dependent protein kinase I (CaMKI), 1:305*f*, 1:306, 1:311
- calcium ion (Ca^{2+})–calmodulin (CaM)-dependent protein kinase II (CaMKII), 1:305*f*, 1:306–307, 1:308*f*, 1:311, 1:381
- calcium ion (Ca^{2+})–calmodulin (CaM)-dependent protein kinase IV (CaMKIV), 1:305*f*, 1:307, 1:308*f*, 1:311
- calcium ion (Ca^{2+})–calmodulin (CaM)-dependent protein kinase kinase (CaMKK), 1:305*f*, 1:307–308, 1:308*f*
- calcium/calmodulin-dependent protein kinase II (CaMKII), 1:297–303, 1:311
- calcium ion (Ca^{2+}) signaling modulation, 1:304, 1:305*f*
- calcium sensors, 1:311
- domain structures, 1:305*f*
- eukaryotic elongation factor 2 kinase (eEF-2K), 1:305*f*, 1:306
- gene expression regulation, 1:311
- general discussion, 1:304
- myosin light chain kinase (MLCK), 1:305–306, 1:305*f*

proteins (*Continued*)

- narrow-specificity calcium ion (Ca^{2+})-calmodulin (CaM)-dependent protein kinases (CaMKs), 1:305–306
- neurotransmitter transporters, 3:239
- phosphorylase kinase (PhK), 1:305*f*, 1:306
- research summary, 1:308
- carboxyl methylation, 3:607–610
 - bacterial chemotaxis, 3:607–608, 3:608*f*
 - basic concepts, 3:607
 - chemotaxis receptor fibers, 3:608*f*
 - eukaryotic signal transduction proteins, 3:608
 - heterotrimeric G proteins, 3:608–609, 3:608*f*
 - phosphoprotein phosphatase 2A (PP2A), 3:609–610, 3:609*f*
 - Ras-related proteins, 3:608–609
- chlorosomes, 2:514
- Crp proteins, 3:367–368
- Csk protein kinase, 4:295–296, 4:296*f*
- Csm protein complexes, 2:514
- digestion and absorption, 2:358–359
- exosomes and microvesicles (EMVs), 2:262, 2:263*f*
- fibroblast growth factor receptors (FGFRs), 2:303–304, 2:304*f*, 2:305*f*
- fibroblast growth factors (FGFs), 2:303
- Fis protein, 3:367–368, 3:368*t*, 3:370
- folding and assembly processes, 3:625–631
 - chaperonins, 1:456
 - complex folding reactions
 - disulfide-coupled folding, 3:628
 - general discussion, 3:627–629
 - ligand binding, 3:629
 - molecular chaperones, 3:629
 - multiple-domain proteins, 3:628, 3:628*f*
 - folding defects, 3:630
 - folding mechanisms and kinetics
 - folding and unfolding rates, 3:626
 - folding intermediates, 3:626–627
 - Levinthal paradox, 3:626
 - transition states, 3:627, 3:627*f*
 - oligomeric protein assembly, 3:629–630, 3:630*f*
 - research background, 3:625
 - thermodynamics, 3:625, 3:626*f*
 - unfolding–refolding transition, 3:626*f*
- gastric secretions, 2:356
- heme proteins, 2:532–538
 - characteristics, 2:532
 - cytochromes, 2:537
 - electron carriers, 2:537
- enzymes
 - catalases, 2:536
 - characteristics, 2:535–537
 - cytochrome P-450 (CYP), 2:536–537
 - peroxidases, 2:535
- oxygen carriers
 - basic concepts, 2:532–535
 - hemoglobin (Hb), 2:533
 - myoglobin (Mb), 2:532
- photosystem II (PSII) complex, 3:513*f*, 3:514
 - structural characteristics, 2:532, 2:533*f*
- human immunodeficiency virus (HIV-1) Gag and Gag/Pol polyproteins, 2:102, 2:102*f*, 2:554, 2:555*f*, 4:443
- Pol proteins, 2:554
- integrin-associated protein (IAP), 2:611
- leucine-responsive regulatory protein (Lrp), 3:367, 3:368*t*
- lipid modification, 2:736–740
 - detection methods, 2:738
 - fatty acylation
 - background information, 2:736–740
 - myristoylation, 2:736, 2:737*f*, 2:737*t*
 - palmitoylation, 2:736–737, 2:737*f*, 2:737*t*
 - saturated fatty acids, 2:737
 - unsaturated fatty acids, 2:737
 - glycosylphosphatidyl inositol (GPI) anchors, 2:737*f*, 2:737*t*, 2:738
 - inhibitors, 2:740
 - intracellular trafficking, 2:739
 - membrane binding, 2:738
 - membrane raft targeting, 2:739–740
 - membrane targeting, 2:738
 - prenylation, 2:737–738, 2:737*t*, 3:608–609, 3:608*f*
 - reversible membrane binding, 2:738–739
 - switch mechanisms, 2:738–739
- lipoproteins, 1:516, 1:517*f*, 3:74, 3:418, 3:422*f*
- low barrier hydrogen bonds (LBHBs), 2:756–759
 - basic concepts, 2:756
 - characteristics, 2:756–757
 - enzymes
 - Δ 5-3-ketosteroid isomerase, 2:759
 - serine esterases, 2:759
 - serine proteases, 2:758*f*
 - physicochemical properties
 - bond lengths, 2:757
 - deuterium fractionation factors, 2:758
 - spectroscopic properties, 2:757–758
 - research background, 2:756
 - strong hydrogen bonds
 - bond lengths, 2:757
 - deuterium fractionation factors, 2:758
 - serine proteases, 2:758–759
 - spectroscopic properties, 2:757–758
 - structural characteristics, 2:756–757, 2:757*f*
 - very strong (single-well) hydrogen bonds
 - bond lengths, 2:757
 - deuterium fractionation factors, 2:758
 - spectroscopic properties, 2:758
 - structural characteristics, 2:757, 2:757*f*
 - weak hydrogen bonds
 - bond lengths, 2:757
 - deuterium fractionation factors, 2:758
 - spectroscopic properties, 2:757
 - structural characteristics, 2:756, 2:757*f*
- mitochondria, 2:126
- mitotic checkpoint, 1:421–422
- Mss116p protein, 4:151–152, 4:152*f*
- nuclear factor kappa B (NF- κ B), 3:302, 3:303*f*
- nucleoid-associated proteins (NAPs), 3:367–370
 - characteristics and functional role, 3:367, 3:368*f*, 3:368*t*
 - DNA supercoiling, 3:368–369, 3:369*f*
 - gene expression regulation, 3:367–368
 - topoisomerases, 3:369–370
- palmitoylation, 3:645–647
 - basic concepts, 2:736–737, 3:645
 - covalent modifications, 3:645, 3:646*f*
 - functional role, 3:646–647
 - inhibitors, 3:647
 - N-palmitoylation, 2:737*f*, 3:645, 3:646*f*
 - O-palmitoylation, 3:645, 3:646*f*
 - palmitoylthioesterases, 3:646
 - palmitoyltransferases, 2:736–737, 3:645–646
 - research summary, 3:647
 - S-palmitoylation, 2:736–737, 2:737*f*, 3:645, 3:646*f*
- phage display, 3:431–436
 - display vectors
 - basic concepts, 3:431–432
 - phagemid vectors, 3:432, 3:432*f*
 - phage vectors, 3:432, 3:432*f*
 - polyvalent/monovalent displays, 3:431–432
 - phage-binding analysis, 3:435, 3:435*f*
 - phage-binding selections
 - affinity-based selection, 3:433–435, 3:435*f*
 - binding and elution, 3:433–435
 - preparation methods, 3:433
 - phage libraries
 - characteristics, 3:432–433
 - diversity library design, 3:433, 3:434*t*
 - sequence generation, 3:432–433, 3:433*t*, 3:434*t*
 - research applications, 3:431
- photosystem I (PSI) complex, 3:503–505, 3:504*f*, 3:505*f*
- photosystem II (PSII) complex, 3:512–519
 - PsbA-D1 protein, 3:514–515, 3:515*t*, 3:516*f*
 - PsbB-CP47 protein, 3:515, 3:515*t*, 3:516*f*
 - PsbC-CP43 protein, 3:515, 3:515*t*, 3:516*f*
 - PsbD-D2 protein, 3:515, 3:515*t*, 3:516*f*
 - PsbE protein, 3:515*t*, 3:516, 3:516*f*
 - PsbF protein, 3:515*t*, 3:516, 3:516*f*
 - psb* genes, 3:514–519, 3:515*t*
 - PsbG protein, 3:515*t*, 3:516, 3:516*f*
 - PsbH protein, 3:515*t*, 3:516–517, 3:516*f*
 - PsbI protein, 3:515*t*, 3:516*f*, 3:517
 - PsbJ protein, 3:515*t*, 3:516*f*, 3:517
 - PsbK protein, 3:515*t*, 3:516*f*, 3:517
 - PsbL protein, 3:515*t*, 3:516*f*, 3:517
 - PsbM protein, 3:515*t*, 3:516*f*, 3:517
 - PsbN protein, 3:515*t*, 3:516*f*, 3:517
 - PsbO protein, 3:515*t*, 3:516*f*, 3:517–518

- PsbP protein, 3:515*t*, 3:516*f*, 3:518
 PsbQ protein, 3:515*t*, 3:516*f*, 3:518
 PsbR protein, 3:515*t*, 3:516*f*, 3:518
 PsbS protein, 3:515*t*, 3:516*f*, 3:518
 PsbT_C protein, 3:515*t*, 3:516*f*, 3:518
 PsbT_N protein, 3:515*t*, 3:516*f*, 3:518
 PsbU protein, 3:515*t*, 3:516*f*, 3:518
 PsbV protein, 3:515*t*, 3:516*f*, 3:518
 PsbW protein, 3:515*t*, 3:516*f*, 3:518
 PsbX protein, 3:515*t*, 3:516*f*, 3:519
 PsbY protein, 3:515*t*, 3:516*f*, 3:519
 PsbZ protein, 3:515*t*, 3:516*f*, 3:519
 reaction center core complex,
 3:514–519, 3:515*t*, 3:516*f*
 research summary, 3:519
 Ycf12 protein, 3:515*t*, 3:516*f*, 3:519
 protein biosynthesis elongation,
 2:249–255
 aminoacyl (A) sites, 2:249–251, 2:250*f*,
 2:251*f*, 2:253*f*, 2:254*f*, 4:122–123
 background information, 2:249,
 4:122–123
 components, 2:250*t*
 eEF1A elongation factor, 2:249–251,
 2:250*f*, 2:250*t*, 2:252, 2:253–254,
 2:253*f*, 2:254*f*
 eEF2 elongation factor, 2:252, 2:254*f*
 eEF3 elongation factor, 2:253
 eEFSec elongation factor, 2:253
 elongation cycle chemistry, 2:251–252,
 2:251*f*
 elongation factor G (EF-G), 4:122–127
 EF-G–ribosome interactions, 4:123*f*,
 4:125–126, 4:126*f*, 4:128–129,
 4:129*f*
 research summary, 4:126
 ribosome recycling, 4:126
 translocation cycle, 4:125, 4:132–133
 elongation factor Tu (EF-Tu), 4:122–127
 EF-Tu–ribosome interactions, 4:123*f*,
 4:128–129, 4:129*f*
 guanosine triphosphate (GTP)
 hydrolysis, 4:123–125, 4:124*f*
 initial selection phase, 4:123, 4:124*f*
 research summary, 4:126
 substrate proofreading, 4:125
 entry (F) sites, 2:252, 2:253*f*
 exit (E) sites, 2:252, 2:253*f*, 4:122–123
 molecular mimicry, 2:252
 Nobel-winning research, 2:254
 peptidyl (P) sites, 2:249–251, 2:250*f*,
 2:251*f*, 2:253*f*, 2:254*f*, 4:122–123
 regulation mechanisms, 2:253–254
 structural model, 2:254*f*
 protein data resources, 3:611–616
 genomic sequences
 contextual relationships, 3:611–612
 protein modification and secondary
 structure resources, 3:612
 protein primary sequence resources,
 3:611–612, 3:615*t*
 proteomics and protein interaction
 resources, 3:612, 3:615*t*
 primary sequence data, 3:611, 3:615*t*
 protein function resources
 annotations, 3:613–614
 functional annotation resources, 3:614
 protein expression, interaction, and
 localization, 3:614, 3:615*t*
 resource availability and maintenance
 biocuration, 3:616
 major resources, 3:615*t*
 modes of access, 3:614–615
 online versus print access, 3:615–616
 three-dimensional protein data
 resources
 Protein Data Bank, 3:612–613, 3:615*t*
 protein structure classification
 systems, 3:613, 3:615*t*
 protein three-dimensional models,
 3:613
 protein degradation, 3:617–624
 bacteria
 adenosine triphosphate (ATP)-
 dependent proteases, 3:623–624
 functional role, 3:623
 substrate selection, 3:624
 basic concepts, 3:617
 disease connections, 3:623*t*
 drug development, 3:624
 endocytic–lysosomal pathways
 functional role, 3:617–619
 lysosomal enzymes, 3:618–619
 lysosomes, 3:617–619
 major histocompatibility complex
 (MHC), 3:619
 schematic diagram, 3:618*f*
 substrate uptake, 3:618, 3:618*f*
 endoplasmic reticulum (ER)-associated
 degradation, 3:622, 3:623*t*
 mitochondria, 3:622–623, 3:623*t*
 ubiquitin–proteasome pathways
 degradation mechanisms, 3:619–622,
 3:619*f*
 disease connections, 3:622
 hybrid proteasome, 3:621–622, 3:621*f*
 immunoproteasome, 3:621–622
 mitochondria, 3:622–623, 3:623*t*
 schematic diagram, 3:618*f*
 26S proteasome, 3:620–621, 3:621*f*
 ubiquitin conjugation, 3:619–620
 protein–DNA interactions, 2:149–153
 DNA-binding motifs
 basic helix–loop–helix (bHLH)
 motifs, 2:151*f*, 2:152
 basic leucine zipper motif, 2:151*f*,
 2:152
 β-sheet motif, 2:151*f*, 2:152
 characteristics, 2:151–152
 helix–turn–helix (HTH) motif,
 2:151–152, 2:151*f*
 ribbon–helix–helix motif, 2:151*f*, 2:152
 zinc fingers, 2:151*f*, 2:152
 protein structural elements
 α-helices, 2:150, 2:151*f*
 β-strands, 2:150–151, 2:151*f*
 loops, 2:151
 research summary, 2:152–153
 specificity, 2:149
 structural and chemical properties
 DNA sequences, 2:149–150
 importance, 2:149–150
 major and minor grooves, 2:149,
 2:150*f*
 nonspecific protein–DNA
 interactions, 2:150
 schematic diagram, 2:150*f*
 protein kinases
 A-kinase anchoring protein (AKAP)
 family, 1:129, 1:575, 1:582, 2:345*f*
 AMP-activated protein kinase (AMPK)
 body weight regulation, 4:58*f*
 cellular fatty acid metabolism
 regulation, 4:56
 diabetes treatments, 1:657
 fatty acid metabolism regulation, 2:286
 fatty acid synthesis pathway,
 2:275–279, 2:276*f*
 insulin signal transduction, 2:604*f*,
 2:605
 ischemic-induced metabolic signaling
 alterations, 3:67–69, 3:69*f*
 lipogenic gene expression, 2:287*f*
 microRNA (miRNA)-based regulation,
 4:193
 calcium ion (Ca²⁺)–calmodulin (CaM)-
 dependent protein kinases
 (CaMKs), 1:304–309
 broad-specificity calcium ion (Ca²⁺)–
 calmodulin (CaM)-dependent
 protein kinases (CaMKs),
 1:306–308
 calcium/calmodulin-dependent
 protein kinase II (CaMKII),
 1:297–303, 1:311
 calcium ion (Ca²⁺)–calmodulin
 (CaM)-dependent protein kinase
 I (CaMKI), 1:305*f*, 1:306, 1:311
 calcium ion (Ca²⁺)–calmodulin
 (CaM)-dependent protein kinase
 II (CaMKII), 1:305*f*, 1:306–307,
 1:308*f*, 1:311, 1:381
 calcium ion (Ca²⁺)–calmodulin
 (CaM)-dependent protein kinase
 IV (CaMKIV), 1:305*f*, 1:307,
 1:308*f*, 1:311
 calcium ion (Ca²⁺)–calmodulin
 (CaM)-dependent protein kinase
 kinase (CaIMKK), 1:305*f*,
 1:307–308, 1:308*f*
 calcium ion (Ca²⁺) signaling
 modulation, 1:304, 1:305*f*,
 2:719–720, 2:720*f*
 calcium sensors, 1:311
 domain structures, 1:305*f*
 eukaryotic elongation factor 2 kinase
 (eEF-2K), 1:305*f*, 1:306
 gene expression regulation, 1:311
 general discussion, 1:304
 myosin light chain kinase (MLCK),
 1:305–306, 1:305*f*
 narrow-specificity calcium ion (Ca²⁺)–
 calmodulin (CaM)-dependent
 protein kinases (CaMKs),
 1:305–306
 neurotransmitter transporters, 3:239
 phosphorylase kinase (PhK), 1:305*f*,
 1:306

proteins (*Continued*)

- research summary, 1:308
- calcium signal mediators, 1:311
- casein kinase (CK2 α), 1:514, 1:514*f*
- chloroplast protein kinase network, 1:514, 1:514*f*
- cyclic nucleotide-dependent protein kinases, 1:574–579
 - binding specificity and affinity, 1:577–578
- cAMP-dependent protein kinase, 1:575–577, 1:576*f*
- cGMP-dependent protein kinase, 1:577
- cyclic adenosine monophosphate (cAMP), 1:574, 1:575*f*
- cyclic guanosine monophosphate (cGMP), 1:574, 1:575*f*
- feed-forward control, 1:578–579, 1:578*f*
- intracellular receptors, 1:574, 1:575*f*
- negative feedback, 1:578–579, 1:578*f*
- structural domains, 1:574–575, 1:576*f*
- substrate specificity, 1:578
- DNA-dependent protein kinase catalytic subunit (DNA-PK α), 3:269–271, 3:270*f*, 3:272*f*
- interleukin-1 receptor-associated kinase (IRAK) family protein kinases, 4:398, 4:399*f*, 4:400*f*
- macroautophagy regulation, 1:149
- mitogen-activated protein kinase (MAPK), 3:176–180
 - adipocyte development, 1:55, 1:55*f*
 - adrenergic receptors, 1:59
 - bile salts, 1:169
 - calcium ion (Ca²⁺) signaling system, 2:719–720, 2:720*f*, 3:293*f*, 3:294
 - chemokine receptors, 1:480–482, 1:482*f*
 - dopamine receptors, 2:172
 - extracellular signal-regulated kinase (ERK), 2:497–498, 2:498*f*, 3:176–178, 3:177*f*, 3:293*f*, 3:294, 4:222–223, 4:222*f*
 - fibroblast growth factor receptors (FGFRs), 2:304, 2:305*f*
 - functional roles, 3:176
 - G $\beta\gamma$ subunit functions, 2:377
 - gene expression regulation, 3:179
 - glycogen synthase kinase-3 (GSK3), 2:436
 - G-protein-signaling regulators (RGSs), 2:493
 - G α family proteins, 2:497–498, 2:498*f*
 - hematopoietin receptors, 2:528*f*
 - hemidesmosome function, 1:642–643
 - hydrogen peroxide (H₂O₂), 3:173
 - insulin-like growth factor-1 (IGF1), 1:198–199, 1:200*f*
 - insulin receptors (IRs), 2:602–603
 - insulin resistance mechanisms, 1:651, 1:651*f*
 - insulin signal regulation, 2:606
 - insulin signal transduction, 2:605–606
 - c-jun N terminal kinase (JNK) pathway, 3:177–178, 3:177*f*
 - mTOR complexes, 1:408, 3:195–196
 - muscarinic acetylcholine receptors, 3:216
 - neurogenesis, 1:198–199, 1:200*f*
 - neurokinin-1 receptor (NK1R), 4:362–363
 - neurotrophins, 3:241–243, 3:242*f*
 - nitric oxide (NO), 3:173
 - nuclear signal transmission specificity and fidelity, 3:178–179
 - p38 kinase, 3:177*f*, 3:178
 - pheromone receptors, 3:441–442, 3:443*f*
 - phylogenetic relationships, 3:177*f*
 - platelet-activating factor receptor (PAFR), 3:536, 3:536*f*
 - platelet-derived growth factor (PDGF) receptors, 3:541*f*
 - RAGE (receptor for advanced glycation end products), 2:407–409
 - signaling cascades, 3:177*f*
 - somatostatin (sst) receptors, 4:265
 - steroid/thyroid hormone receptors, 4:310*f*
 - T-cell receptors (TCRs), 4:374
 - vascular endothelial growth factor receptors (VEGFRs), 4:512
 - partitioning defective (PAR)/atypical protein-kinase (aPKC) complex, 1:262–263
 - phosphoinositide-dependent protein kinases, 3:463–466
 - functional role, 3:463
 - kinase S6 kinase (S6K), 2:604*f*, 2:605, 3:464*f*, 3:465
 - mTOR (rapamycin), 3:464*f*, 3:465
 - phosphoinositide 3-kinase (PI 3-kinase) lipid signaling pathway, 3:448*f*, 3:463, 3:464*f*
 - phosphoinositide-dependent kinase 1 (PDK1), 3:464–465, 3:464*f*, 4:377–378
 - phosphoinositide-dependent kinase 2 (PDK2), 3:464*f*, 3:465
 - protein kinase B (PKB), 3:463–464
 - research summary, 3:466
 - tyrosine kinases (Tec family), 3:465–466
 - protein kinase A (PKA)
 - adipocyte development, 1:55
 - arachidonic acid regulated Ca²⁺ (ARC) channels, 1:129
 - binding specificity and affinity, 1:577–578
 - carbohydrate-responsive element binding protein (ChREBP), 1:375–376, 1:376*f*
 - characteristics and functional role, 1:575–577
 - cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:563–564, 1:564*f*
 - feed-forward control, 1:578–579, 1:578*f*
 - glycogen synthase kinase-3 (GSK3), 2:435*f*
 - G protein-coupled receptor kinases (GRKs), 2:485*t*
 - neurogenesis, 1:199
 - Perilipin 1 (Plin1), 1:47–48, 1:48*f*
 - phosphofructokinase-2/fructose biphosphatase-2 (PFK-2/FBPase-2) activation, 3:457–458
 - PKA-mediated lipolysis, 1:49–51, 1:50*f*, 4:303–304
 - plasma membrane calcium (Ca⁺)-ATPase pump, 3:525–526
 - structural domains, 1:574–575, 1:575*f*, 1:576*f*
 - substrate specificity, 1:578
 - protein kinase B (PKB)
 - cyclic adenosine monophosphate (cAMP) receptors (cARs), 1:563–564, 1:564*f*
 - functional role, 3:463–464
 - glycogen synthase kinase-3 (GSK3), 2:435*f*
 - insulin signal transduction, 2:604
 - mTOR complexes, 1:408, 3:195–196
 - neurokinin-1 receptor (NK1R), 4:362–363
 - neurotransmitter transporters, 3:239
 - phosphatidylinositol-3,4,5-trisphosphate (PtdIns-3,4,5-P₃), 3:450
 - protein kinase C (PKC), 3:637–640
 - arachidonic acid regulated Ca²⁺ (ARC) channels, 1:129
 - background information, 3:637
 - bile salts, 1:169
 - calcium ion (Ca²⁺) signaling system, 2:719–720, 2:720*f*
 - calcium oscillation decoding, 1:325
 - calcium-sensing receptor (CaSR), 1:328, 2:583, 2:584*f*
 - chemokine receptors, 1:480–482, 1:482*f*
 - diabetes, 1:656
 - domain structures, 3:637, 3:638*f*
 - downregulation, 3:639
 - G₁₂/G₁₃ family functions, 2:345*f*
 - G $\beta\gamma$ subunit functions, 1:480–482, 1:482*f*, 4:265
 - gene expression regulation, 1:311
 - glycogen synthase kinase-3 (GSK3), 2:435*f*
 - G protein-coupled receptor kinases (GRKs), 2:485*t*
 - G α family proteins, 2:497–498, 2:498*f*
 - historical research, 3:637
 - insulin resistance, 1:650
 - insulin-secreting cells (β -cells), 2:597*f*
 - insulin signal transduction, 2:604
 - isozyme families, 3:637, 3:638*f*

- mitochondrial permeability transition pore, 3:165–166
- neurogenesis, 1:199
- neurokinin-1 receptor (NK1R), 4:362–363
- neurotransmitter transporters, 3:239
- neurotrophins, 3:241–243, 3:242f
- nuclear calcium signal, 1:310–311
- phospholipase D (PLD), 3:472–473, 3:474–475, 3:474f
- phosphorylation, 3:637–638, 3:638f
- plasma membrane calcium (Ca⁺)-ATPase pump, 3:525–526
- platelet-activating factor receptor (PAFR), 3:536, 3:536f
- platelet-derived growth factor (PDGF) receptors, 3:541f
- signaling pathways, 3:639–640
- somatostatin (sst) receptors, 4:265
- store-operated calcium (SOC) channels, 3:359–360, 3:360f
- subcellular localization, 3:639
- T-cell receptors (TCRs), 4:374, 4:377–378
- translocation cycle, 3:638–639, 3:639f
- vascular endothelial growth factor receptors (VEGFRs), 4:512
- protein kinase G (PKG)
- arachidonic acid regulated Ca²⁺ (ARC) channels, 1:129
 - binding specificity and affinity, 1:577–578
 - characteristics and functional role, 1:577
 - feed-forward control, 1:578–579, 1:578f
 - mitochondrial permeability transition pore, 3:165–166
 - negative feedback, 1:578–579, 1:578f
 - structural domains, 1:574–575, 1:575f, 1:576f
 - substrate specificity, 1:578
- protein–tyrosine kinase-linked receptors (PTKRs), 2:583, 2:584f
- protein tyrosine phosphatase (PTP), 3:648–653
- catalytic mechanisms, 3:648, 3:649f
 - CDC25 family, 3:648, 3:650f, 3:651f
 - classic PTPs, 3:648, 3:650f, 3:651f
 - disease connections, 3:649–652
 - dual-specific PTPs (DS-PTPs), 3:648, 3:650f
 - intracellular PTPs, 3:648–649, 3:650f, 3:651f
 - low-molecular-weight PTPs (LMW-PTPs), 3:648–649, 3:650f, 3:651f
 - major sperm protein (MSP), 3:3, 3:3f
 - phosphorylation, 3:648, 4:294–295, 4:296f
 - protein tyrosine phosphatase PTB1B, 2:260f, 3:649
 - receptor-like PTPs (RPTPs), 3:649, 3:650f
 - redox signaling mechanisms, 3:411–412, 3:411f
 - therapeutic uses, 3:652
- ryanodine receptors (RyRs), 4:197
- stress-activated protein kinase (SAPK), 3:177–178, 3:177f
- thylakoid protein kinase network, 1:514, 1:514f
- protein misfolding diseases, 3:574–575, 3:575t
- protein N-myristoylation, 3:641–644
- basic concepts, 3:641
 - N-myristoylproteins
 - N-myristoylation functions, 3:641–642
 - substrate selection, 3:642, 3:642f - N-myristoyltransferase (Nmt)
 - prevalence, 3:642–643
 - reaction mechanisms, 3:642–643
 - structural characteristics, 3:643, 3:643f
 - substrates, 3:642f
 - therapeutic uses, 3:643–644
- therapeutic uses, 3:643–644
- protein phosphatases
- calcium signal mediators, 1:312
 - chemokine receptors, 1:483f
 - mTOR complexes, 3:198
 - nuclear inhibitor of protein phosphatase 1 (NIPP-1), 1:312, 4:222
 - phospho-protein phosphatase (PPP)
 - biochemical characterization, 4:219, 4:220t
 - catalytic mechanisms, 4:220, 4:220f
 - characteristics, 4:219
 - PP1 (calcium/calineurin-dependent type 1 serine/threonine protein phosphatase), 1:312, 4:220, 4:221f, 4:222
 - protein phosphatase 2A (PP2A), 3:3t, 3:4f, 3:5, 4:221, 4:221f, 4:222–223
 - protein phosphatase 2B (PP2B), 4:221, 4:221f, 4:223
 - protein phosphatase 5 (PP5), 4:221, 4:221f
 - regulation mechanisms, 4:222
 - regulatory subunit interactions, 4:220, 4:221f
 - T-cell activation, 4:222f, 4:223
 - three-dimensional structure, 4:219–220, 4:220f, 4:221f
 - PP1 (calcium/calineurin-dependent type 1 serine/threonine protein phosphatase), 1:312, 4:220, 4:221f, 4:222
 - protein phosphatase 1 nuclear-targeting subunit (PNUTS), 1:312
 - protein phosphatase 2A (PP2A), 3:3t, 3:4f, 3:5, 4:221, 4:221f, 4:222–223
 - protein phosphatase 2B (PP2B), 3:3t, 3:4f, 3:5, 4:221, 4:221f, 4:223
 - protein phosphatase 5 (PP5), 4:221, 4:221f
 - protein phosphatase Mg²⁺- or Mn²⁺-dependent (PPM)
 - biochemical characterization, 4:219, 4:220t
 - catalytic mechanisms, 4:220
- cellular functions, 4:221
- characteristics, 4:219
- regulation mechanisms, 4:221
- three-dimensional structure, 4:219–220
- protein tyrosine kinase (PTK), 1:162, 1:163f, 3:470
- regulated intramembrane proteolysis (RIP), 4:50–55
- basic concepts, 4:50
 - γ -secretase-based mediation
 - amyloid precursor protein (APP), 4:51f, 4:53
 - basic concepts, 4:53
 - ErbB-4, 4:51f, 4:53
 - Notch, 4:51f, 4:53 - rhomboid-based mediation, 4:51f, 4:54–55
- signal peptide peptidase (SPP)-based mediation, 4:51f, 4:53–54
- site-2 protease (S2P)-based mediation
- activation transcription factor 6 (ATF6), 4:51f, 4:52
 - animal cells, 4:50–52
 - bacteria, 4:51f, 4:52–53
 - basic concepts, 4:50–53
 - cAD1 precursor, 4:51f, 4:52–53
 - cAMP responsive element-binding protein 3 (CREB3), 4:51f, 4:52
 - HEXxH motif, 4:50–53
 - RseA, 4:51f, 4:52
 - SpoIVFB protein, 4:51f, 4:53
 - sterol regulatory element binding protein (SREBP), 4:51–52, 4:51f
- respiratory chain complex IV (cytochrome *c* oxidase), 4:100
- selenoprotein synthesis, 4:210–213
- archaeal selenoprotein synthesis, 4:213
 - bacterial selenoprotein synthesis
 - anticodon loop, 4:210, 4:211f
 - general discussion, 4:210–213
 - ribosomal-level decoding, 4:212–213 - selenocysteine biosynthesis, 4:210–212, 4:211f
 - selenocysteine specificity, 4:213
 - selenocysteine synthase (SelA), 4:210–211, 4:211–212, 4:211f
 - selenocysteine synthase (SelB)–guanosine triphosphate (GTP)–selenocysteyl-tRNA–selenocysteine insertion sequence (SECIS) element formation, 4:212, 4:213f
 - transfer RNA (tRNA)^{Sec}, 4:210, 4:211f
 - UGA stop codons, 4:212
 - basic concepts, 4:210
 - eukaryal selenoprotein synthesis, 4:213, 4:213f
- serine–threonine phosphatases, 4:219–223
- biochemical characterization, 4:219, 4:220t
 - catalytic mechanisms, 4:220, 4:220f
 - DARPP-32, 4:222f, 4:223
 - disease connections, 4:223
 - nuclear factor of activated T cells (NFAT), 4:222f, 4:223

proteins (*Continued*)

- phospho-protein phosphatase (PPP)
 - biochemical characterization, 4:219, 4:220*t*
 - catalytic mechanisms, 4:220, 4:220*f*
 - characteristics, 4:219
 - PP1 (calcium/calineurin-dependent type 1 serine/threonine protein phosphatase), 1:312, 4:220, 4:221*f*, 4:222
 - protein phosphatase 2A (PP2A), 4:221, 4:221*f*, 4:222–223
 - protein phosphatase 2B (PP2B), 3:3*t*, 3:4*f*, 3:5, 4:221, 4:221*f*, 4:223
 - protein phosphatase 5 (PP5), 4:221, 4:221*f*
 - regulation mechanisms, 4:222
 - regulatory subunit interactions, 4:220, 4:221*f*
 - T-cell activation, 4:222*f*, 4:223
 - three-dimensional structure, 4:219–220, 4:220*f*, 4:221*f*
- protein phosphatase Mg²⁺- or Mn²⁺-dependent (PPM)
 - biochemical characterization, 4:219, 4:220*t*
 - catalytic mechanisms, 4:220
 - cellular functions, 4:221
 - characteristics, 4:219
 - regulation mechanisms, 4:221
 - three-dimensional structure, 4:219–220
- subunit families
 - phospho-protein phosphatase (PPP), 4:219
 - protein phosphatase Mg²⁺- or Mn²⁺-dependent (PPM), 4:219
 - three-dimensional structure, 4:219–220, 4:220*f*
- two-hybrid protein–protein interaction systems, 4:460–462
 - basic concepts, 4:460–462, 4:461*f*
 - proteomic studies, 4:462
 - research applications, 4:461–462
 - reverse two-hybrid systems, 4:462
- ubiquitin-like proteins (UBLs), 4:467–472
 - Atg8 and Atg12 conjugation systems, 1:151–153, 1:152*t*, 4:468*t*, 4:471
 - attachment dynamics, 4:467–468, 4:468*f*, 4:468*t*
 - Fat10 (diubiquitin), 4:468*t*, 4:471–472
 - functional role, 4:467, 4:468*f*, 4:475
 - Isg15 (interferon-induced 15 kDa protein), 4:468*t*, 4:470–471
 - neural precursor cell expressed, developmentally downregulated 8 (NEDD8), 4:468*t*, 4:470, 4:475
 - small ubiquitin-related modifiers (SUMOs), 4:69, 4:468–470, 4:468*t*, 4:475
- uncoupling proteins (UCPs), 4:482–487
 - adenosine diphosphate (ADP) phosphorylation and respiration
 - adenosine triphosphate (ATP) synthesis, 4:482
 - brown adipose tissue (BAT), 4:483–484, 4:484*f*
 - mitochondrial membranes, 4:482
 - ATP synthesis, 1:142*f*, 1:144
 - brown adipose tissue (BAT)
 - adenosine diphosphate (ADP) phosphorylation and respiration, 4:483–484, 4:484*f*
 - characteristics and functional role, 4:483
 - histology, 4:483*f*
 - thermogenesis, 4:483–484, 4:484*f*
 - UCP1, 4:483–484, 4:484*f*, 4:485*f*
 - characteristics and functional role, 4:482
 - leptin, 2:258–259
 - mitochondrial ROS, 4:21
 - novel UCPs
 - energy loss mechanisms, 4:484–486
 - functional role, 4:485–486, 4:486*f*
 - glucose metabolism, 4:486
 - messenger RNA (mRNA), 4:484–485, 4:485*f*
 - plant and bird UCPs, 4:485
 - reactive oxygen species (ROS), 4:485–486, 4:486*f*
 - skeletal muscle, 4:485, 4:485*f*
 - UCP2, 4:484–485, 4:485*f*
 - UCP3, 4:485, 4:485*f*
 - research summary, 4:486
 - unfolded protein response (UPR), 4:488–493
 - basic concepts, 4:488
 - chaperones, 4:488, 4:492*f*
 - functional roles, 4:488
 - metazoans
 - ATF6 gene activation, 4:490–491, 4:491*f*
 - inositol requiring 1 (IRE1) gene conservation, 4:490
 - intramembrane proteolysis, 4:490–491
 - physiological and pathological stresses
 - ATF6 gene activation, 2:524–525, 2:524*f*, 4:490–491, 4:491*f*
 - basic concepts, 4:488–489
 - gene expression regulation, 4:491–492
 - HAC1 gene expression, 2:524–525, 2:524*f*, 4:489, 4:490*f*
 - inositol requiring 1 (IRE1) genes, 2:524–525, 2:524*f*, 4:489, 4:490*f*
 - intramembrane proteolysis, 4:490–491, 4:491*f*
 - stress sensors, 4:492
 - translation initiation factor 2 (eIF2 α) phosphorylation, 4:491, 4:492*f*
 - XBP-1 gene expression, 2:524–525, 2:524*f*
 - translational control
 - gene expression regulation, 4:491–492
 - PERK (PKR-like ER kinase), 2:524–525, 2:524*f*, 4:491, 4:492*f*
 - stress sensors, 4:492
- translation initiation factor 2 (eIF2 α)
 - phosphorylation, 2:524–525, 2:524*f*, 4:491, 4:492*f*
- yeasts
 - HAC1 gene expression, 4:489, 4:490*f*
 - inositol requiring 1 (IRE1) genes, 4:489, 4:490*f*
 - vitamin K-dependent proteins, 4:551
 - vSrc protein, 4:293–294
 - zinc-finger domains, 4:578
 - See also* kinases; peptides
- protein tyrosin phosphatase PTB1B, 2:260
- proteobacteria
 - ribosomal RNA (rRNA), 3:559
 - ribozyme distribution, 4:146*f*
 - transfer RNA (tRNA), 3:556
 - type I DNA topoisomerase, 2:161
- proteoglycans, 3:654–660
 - congenital disorders, 2:475
 - functional role, 1:439–440
 - gene families
 - characteristics, 3:657*t*, 3:658–659
 - membrane-bound proteoglycans
 - general discussion, 3:659
 - glypicans, 3:657*t*, 3:658–659
 - miscellaneous membrane proteoglycans, 3:657*t*, 3:659
 - syndecans, 3:657*t*, 3:658
 - secreted and matrix proteoglycans
 - Aggrecan family, 3:657*t*, 3:658
 - basement membrane proteoglycans, 3:657*t*, 3:658, 3:659*f*
 - Decorin family, 3:656–658, 3:657*t*
 - general discussion, 3:658
 - secretory granule proteoglycans, 3:657*t*, 3:658, 3:659*f*
 - small interstitial proteoglycans, 3:656–658, 3:657*t*
- life cycle
 - cell-surface desulfation, 3:659
 - endocytosis, 3:659–660
 - general discussion, 3:659–660
 - shedding and heparanase cleavage, 3:659, 3:659*f*
- matrix metalloproteinases (MMPs), 3:91*t*
- structural characteristics
 - basement membrane proteoglycans, 1:437, 1:438*f*
 - glycosaminoglycans (GAGs)
 - chondroitin sulfate (CS)/dermatan sulfate (DS), 3:654–655, 3:655*f*
 - general discussion, 3:654–656
 - heparan sulfate (HS)/heparin, 3:655–656, 3:655*f*
 - hyaluronan, 3:654, 3:655*f*
 - keratan sulfate (KS), 3:655*f*, 3:656
 - sugar building blocks, 3:654
 - protein cores, 3:656, 3:657*t*
- proteolysis, 3:312–313, 4:62, 4:298, 4:304–305
- proteolytic enzymes
 - aspartic proteases, 1:137–140
 - disease processes
 - acquired immunodeficiency syndrome (AIDS), 1:139
 - Alzheimer's disease, 1:140

- cancer, 1:139
 fungal infections, 1:139–140
 general discussion, 1:139
 hypertension, 1:139
 malaria, 1:139
 genomic sequences, 1:137, 1:138f
 historical research, 1:137
 inhibitors
 IA-3, 1:140
 naturally occurring inhibitors, 1:140
 pepstatin, 1:140
 PI-3, 1:140
 synthetic inhibitors, 1:140
 structural characteristics
 catalytic residues, 1:138–139
 flap structure, 1:140
 overall structure, 1:138, 1:138f
 proenzyme structure, 1:139
 research background, 1:137–139
 ProteomeXchange consortium, 3:612
 Proteomics Standard Initiative (PSI)
 Common Query Interface, 3:612
 proteomic studies
 mass spectrometry, 3:11–16
 collision-induced dissociation (CID),
 3:12, 3:12f, 3:13–14, 3:14f
 complex structure, 3:13–14
 dynamic protein complexes, 3:13
 future research areas, 3:15
 gas-phase complex structure, 3:14–15
 GroEL (growth E locus, large gene,
 bacterial group I chaperonin),
 3:14–15, 3:15f
 nano-electrospray ionization (nanoESI),
 3:11–12, 3:12f
 protein complex composition, 3:12–13,
 3:13f
 research methodologies, 3:11–12
 structural biology analyses, 3:11
 nucleolus, 3:325, 3:326f
 two-hybrid protein–protein interaction
 systems, 4:462
Proteus vulgaris, 2:132t
 prothoracicotropic hormone (PTTH), 2:590,
 2:594, 2:594f
 prothrombin, 3:585, 3:586f, 4:551
 protium (1H), 2:688
 protocadherins, 1:256, 1:256f, 1:261
 protochlorophyllide (Pchl_{id}), 1:508
 protonmotive electron transfer chain,
 1:469–470, 1:469f, 1:470f
 protonmotive force (PMF), 1:468, 1:468f,
 1:471f
 protonmotive Q cycle, 1:596–597, 1:597f,
 3:663–664, 3:664f
 proton pumping, 3:661–665
 nicotinamide adenine dinucleotide
 transhydrogenase, 3:250–251,
 3:250f
 oxidative phosphorylation, 3:661
 plasma membrane proton (H⁺)-ATPases,
 3:688–692
 biogenesis, 3:690, 3:690f
 domain structures, 3:688–689
 functional role, 3:688, 3:689f
 P-type ATPases, 3:688
 reaction mechanisms, 3:689–690,
 3:689f
 regulation mechanisms, 3:690–691
 research background, 3:688
 research summary, 3:691
 Q cycle, 3:663–664, 3:664f
 redox loop versus proton pump,
 3:662–663, 3:663f
 respiratory chain complex I (NADH:
 ubiquinone oxidoreductase),
 3:661–662, 3:662f, 3:663, 4:90
 respiratory chain complex III (ubiquinol
 cytochrome *c* reductase), 3:663–664,
 3:664f
 respiratory chain complex IV, 3:664–665
 protoporphyrin IX, 2:540f, 3:545f, 3:546,
 3:546f
 protoporphyrinogen IX, 2:540f, 2:541
 protoporphyrinogen oxidase (PPO), 2:540f,
 2:541, 2:542t
 protozoa
 alternative oxidases (AOXs), 1:145–146
 contractile vacuole complex, 4:506
 conventional actin, 1:38–40
 glycosylphosphatidylinositol (GPI)-
 anchored proteins, 2:477–478
 lectins, 2:701, 2:702t
Providencia stuartii
 rhomboid-based mediation, 4:51f, 4:54
 type III DNA restriction and modification
 enzymes, 2:137t
 pruritus, 3:563–564
 pseudohypoparathyroidism type 1A
 (PHP1A), 2:519–520
 pseudoknots, 4:137–138
 pseudolupus syndrome (PLE), 3:116
Pseudomonas, 1:247f, 3:29–30
Pseudomonas aeruginosa
 Cho phylogenetic tree, 3:331f
 cytochrome *c*, 1:599
 lectins, 2:702t
 mucin domains, 3:202
 non-homologous end joining (NHEJ),
 3:266
 tryptophan biosynthesis pathway,
 4:249b
 uvrC phylogenetic tree, 3:331f
Pseudomonas fluorescens
 coenzyme Q, 3:723t
 tRNA synthetases, 4:444
Pseudomonas mendocina, 2:132t
Pseudomonas putida
 Cho phylogenetic tree, 3:331f
 growth-supporting reductant oxidation,
 1:491t
 mitochondrial-type ferredoxins, 2:297
 uvrC phylogenetic tree, 3:331f
Pseudomonas stutzeri, 1:604t
Pseudomonas syringae
 Cho phylogenetic tree, 3:331f
 cytokinins, 1:631
 ribosomal RNA (rRNA), 3:559
 uvrC phylogenetic tree, 3:331f
 pseudopseudohypoparathyroidism (PPHP),
 2:519–520
 pseudouridine, 4:441
 [PSI⁺] prion
 disease states, 3:568–569
 generation mechanisms, 3:569
 genetic properties, 3:566
 self-propagating amyloidoses, 3:567–568
 psoriasis, 1:638t, 4:539
 PstII enzyme, 2:137t
 pteridines/pterins, 3:666–669
 biosynthesis, 3:666–667, 3:667f
 metabolic reactions
 aromatic amino acid hydroxylases,
 3:667, 3:668f
 catecholamine neurotransmitters, 3:3–4,
 3:668f
 molybdopterin-containing enzymes,
 3:668–669, 3:669f
 nitric oxide synthase (NOS), 3:668,
 3:668f
 occurrences, 3:666
 structural characteristics, 3:666, 3:667f
 tetrahydrobiopterin deficiency, 3:669
 P-type adenosine triphosphatase (ATPase),
 1:319t
 PU.1 transcription factor, 1:187, 1:188f
 pumpkins, 4:533t
 purines
 purine nucleoside phosphorylase (PNP),
 2:213f
 purinergic receptors, 2:496–497, 2:497t
 riboswitches, 4:137
 Purkinje cells, 2:499
 purple bacteria
 anoxygenic photosynthesis, 2:209, 2:335,
 3:32, 3:693–694, 3:694f
 antenna complexes, 3:32
 electron acceptors and donors, 3:693–699
 cytochromes, 3:695–696
 hydrogen (H), 3:698
 nicotinamide adenine dinucleotide
 dehydrogenase, 3:696
 nitrogen compounds, 3:698, 3:698t
 photosynthetic reaction centers,
 3:696–697, 3:696f
 respiratory chain complex III (ubiquinol
 cytochrome *c* reductase), 3:695
 respiratory functions
 cytochrome *cbh3* oxidase, 3:694f,
 3:697, 3:697t, 4:79, 4:80f
 oxidases, 3:697–698
 quinol oxidase, 3:694–695, 3:694f,
 3:695f, 3:697–698
 succinate dehydrogenase, 3:694f,
 3:696
 sulfur compounds, 3:698, 3:698t
 electron transport chain
 anoxygenic photosynthesis, 3:32,
 3:693–694, 3:694f
 membrane transport, 3:693–695
 RegA/RegB regulation, 3:695, 3:695f
 respiration, 3:694–695, 3:694f
 evolutionary development, 3:34
 general discussion, 3:32
 metabolic characteristics, 3:693
 oxygenic photosynthesis, 3:32
 photosynthetic reaction centers,
 3:700–707

- purple bacteria (*Continued*)
- anoxygenic photosynthesis, 3:693–694, 3:694f
 - characteristics, 3:700
 - chlorophyll, 1:501
 - cytochrome *c*₂ oxidation, 3:706–707, 3:706f
 - functional derivatives
 - mutagenesis, 3:705
 - substrate analog and inhibitor complexes, 3:703–705, 3:704f
 - functional roles
 - kinetic mechanisms, 3:703
 - oxidation–reduction potentials, 3:703
 - intramolecular electron transfer, 3:705
 - molecular characterization, 3:700–702, 3:702f
 - photosystem II (PSII) complex, 3:707
 - quinone reduction and protonation, 3:704f, 3:706
 - subunit structure and functions, 3:700–702, 3:701f
 - X-ray crystallographic structure, 3:702–703
- putrescine
- insect biochemistry, 2:590–591
 - urea cycle, 4:495f
- pyloric atresia, 2:627–628, 2:627t
- pyridine nucleotide oxidoreductases, 2:311
- pyridoxal, 3:709f
- pyridoxal phosphate, 3:708–711
- aldol-type reactions, 3:709f, 3:710, 3:710t
 - carbanionic quinonoid intermediate reaction, 3:708–710, 3:710f
 - characteristics and functional role, 3:708
 - chemical mechanisms, 3:709f, 3:710f, 3:711
 - decarboxylation, 3:709f, 3:710, 3:710t
 - heme biosynthesis, 2:539–540
 - mechanism-based enzyme inhibitors, 2:214f
 - molecular structure, 3:709f
 - pyridoxal phosphate-dependent enzymatic reactions, 3:708–710, 3:709f, 3:710f, 3:710t
 - racemization, 3:708–710, 3:709f, 3:710t
 - serine–glycine interconversion, 2:324–325
 - transamination reactions, 3:708, 3:709f
- pyridoxamine, 3:709f
- pyridoxamine phosphate, 3:708, 3:709f
- pyridoxine
- molecular structure, 3:709f
 - retinal and choroidal gyrate atrophy, 4:500
 - vitamin B₆, 3:708
- pyridoxine phosphate, 3:709f
- pyrimidine, 3:712–718
- biosynthesis-related interconversions
 - cytidine nucleotide formation, 3:715–716, 3:715f
 - deficiency-related diseases, 3:716
 - regulation mechanisms, 3:716
 - thymidine nucleotide formation, 3:715–716, 3:715f
 - catabolic reactions
 - deficiency-related diseases, 3:718
 - regulation mechanisms, 3:717f, 3:718
 - uracil and thymine degradation, 3:717–718, 3:717f
 - de novo* synthesis
 - deficiency-related diseases, 3:713–714
 - regulation mechanisms, 3:712–713, 3:714f
 - uridine monophosphate (UMP)
 - formation, 3:712, 3:713f
- DNA glycosylases, 2:3t
- oxidized pyrimidine lesions, 2:78f, 2:79
- pyrimidine 6–4 pyrimidone photoproduct (6–4 PP), 4:572, 4:574
- salvage pathways
- deficiency-related diseases, 3:717
 - phosphorylation/dephosphorylation, 3:716
 - regulation mechanisms, 3:717
- UvrABCD endonuclease, 3:330t
- Pyrobaculum*, 3:34
- Pyrococcus furiosus*, 2:35f, 2:38, 3:87–88, 4:239, 4:253
- pyrophosphate, 4:167, 4:173
- pyrroles, 2:408f
- pyrroline, 1:88, 1:87–88
- pyrrolysine, 4:442–443, 4:443f
- pyruvate
- anaplerotic pathways, 1:105, 1:106f, 1:109f
 - branched-chain amino acid (BCAA)
 - regulation, 1:245–246
 - cerebral metabolism, 1:369f, 1:371–373
 - fatty acid synthesis pathway, 2:276f
 - fermentation reactions, 2:204–206, 4:74f
 - gluconeogenesis, 2:381, 2:382f
 - glycolysis
 - adenosine triphosphate (ATP) production, 2:443
 - anaerobic glycolysis, 2:444–445, 2:445f
 - glycolytic pathway, 2:445f
 - oxidation effects, 2:445–446, 2:446f
 - schematic diagram, 2:444f
 - transcriptional activation, 1:374
 - liver pyruvate kinase (LPK)
 - carbohydrate-responsive element binding protein (ChREBP), 1:374, 1:375f, 1:376–377, 1:376f, 1:377f
 - transcriptional activation, 1:374
 - muscle metabolism, 1:94f
 - phosphoenolpyruvate carboxykinase (PEPCK), 2:381, 2:382–383, 2:385f
 - phosphotransferase system (PTS), 2:388f
 - pyruvate analog dichloroacetate (DCA), 4:324, 4:326f
 - pyruvate carboxylase (PC)
 - anaplerotic pathways, 1:105, 1:107, 1:109f, 1:110
 - characteristics, 2:381–382
 - deficiency-related diseases, 1:110
 - gluconeogenesis, 2:385f
 - historical research, 2:381
 - inborn errors of metabolism, 4:495t
 - plant mitochondrial electron transport chain, 1:141–142, 1:142f
 - protein biotinylation, 1:226
 - pyruvate decarboxylation, 1:206–207, 1:208f
- pyruvate dehydrogenases (PDHs), 4:321–328
- active site coupling, 4:323
 - anti-mitochondrial antibodies (AMAs), 3:114
 - carbohydrate-responsive element binding protein (ChREBP), 1:377
 - cardiac metabolism, 3:65–66, 3:66f, 3:68f, 3:69f, 3:70f
 - characteristics, 4:321
 - components, 4:321–322, 4:322f
 - cryo-electron microscopic architectural analyses, 4:326–327, 4:327f
 - hypoxia inducible factor (HIF), 4:70–71
 - lipoyl domains, 4:323, 4:325f, 4:327f
 - metabolite channeling, 3:80
 - mitochondrial calcium (Ca²⁺) transport, 3:124f
 - mitochondrial ROS, 4:19f, 4:20
 - pyruvate dehydrogenase phosphate phosphatase, 3:124f
 - reaction mechanisms
 - E1 enzyme reaction, 4:322, 4:322f, 4:325f, 4:327f
 - E2 enzyme reaction, 4:322–323, 4:322f, 4:327f
 - E3 enzyme reaction, 4:322f, 4:323, 4:327f
 - general discussion, 4:322–323
 - regulation mechanisms
 - basic principles, 4:323–326
 - biological significance, 4:323–324
 - PDH kinase (PDK) isoforms, 4:324–326, 4:325f, 4:326f
 - small metabolite regulation, 4:324
 - research summary, 4:328
 - saturated fatty acid biosynthesis, 2:285–286
 - substrate channeling, 4:323
 - thiamine diphosphate (ThDP), 1:206, 1:207f
- pyruvate kinase (PK), 3:719–721
- glycolytic reactions, 3:719
 - liver PK (LPK) *in vitro* regulation
 - allosteric regulation, 3:719
 - carbohydrate response element (CHORE), 3:721
 - crystallographic studies, 3:719–721, 3:720f
 - gene expression regulation, 3:721
 - glucagon level regulation, 3:721
 - insulin level regulation, 3:721
 - pH effects, 3:719
 - phosphorylation/dephosphorylation, 3:719
 - liver pyruvate kinase (LPK)
 - liver PK (LPK) *in vitro* regulation, 3:719–721
 - transcriptional activation, 1:374
 - PK fetal isoform (PKM2), 3:721
 - structural characteristics, 3:719, 3:720f
 - Warburg effect, 3:721
 - tricarboxylic acid cycle, 4:439f

Q

Q cycle

- coenzyme Q, 1:511–512, 3:725, 3:725f
- cytochrome *b₆f* complex, 4:338–339, 4:339f
- protonmotive electron transfer chain, 1:470f
- proton pumping, 3:663–664, 3:664f
- respiratory chain complex III (ubiquinol cytochrome *c* reductase), 1:596–597, 1:597f, 4:84
- respiratory electron transport chain, 3:400, 3:401f

Q/N-rich domains, 3:566, 3:568f

quadruplex DNA (G4-DNA)

- background information, 2:144–146
- biological role, 2:144–146, 2:145f
- bonding schemes, 2:144f
- chemical properties, 2:144
- pharmacological role, 2:146
- structural characteristics, 2:143f

quiferales, 3:559

Quin 2, 2:574–575, 2:575t

quinhydrone, 3:722–723

quinine, 3:551

quinol, 1:488f, 2:208f, 2:511–512, 3:725f, 4:92, 4:93f

quinolinic acid (QA), 1:175f

quinolones, 2:167, 2:167f

quinol oxidases (QOX)

- cyanobacterial respiration, 4:81–82, 4:82f
- respiratory electron transport chain, 3:694–695, 3:694f, 3:695f, 3:697–698, 4:79, 4:80f

quinones, 3:722–729

- anaerobic respiration, 2:208f
- anoxygenic photosynthesis, 3:693–694, 3:694f

- bacteriopheophytins, 3:700, 3:701f, 3:702f

benzoquinones

catecholamine-derived

- o*-benzoquinones, 3:727, 3:727f

chemical characteristics, 3:722–723

coenzyme Q

- antioxidant actions, 3:725–727
- biosynthesis, 3:726f, 3:727
- chemical characteristics, 3:723–724
- functional role, 3:724–727, 3:724f, 3:725f
- homologs and analogs, 3:723–724, 3:723t
- plasma membrane oxidoreductase, 3:725

caldariella quinone, 4:94f

chemical characteristics, 3:722–723

chloroplast signaling, 1:511–512

cytochrome *b₆f* complex, 4:336–337, 4:337f

hydronaphthoquinone, 4:551–553, 4:553–554, 4:553f

lysyl tyrosyl quinone, 1:89

menadione, 3:727–728, 4:552f

menaquinone, 2:206, 3:34, 3:700, 3:702f, 3:727–728, 4:94f, 4:552f

naphthoquinones, 3:722–723, 3:727–728, 4:553–554, 4:553f

photosynthetic reaction centers, 3:704f, 3:706

photosystem II (PSII) complex, 3:484

phyloquinone, 3:727–728, 4:552f

protonmotive electron transfer chain, 1:469–470

quinone-exchange cavity, 4:336–337, 4:337f

quinone pool, 3:695, 3:695f

respiratory chain complex I (NADH: ubiquinol oxidoreductase), 4:87–91

background information, 4:98

bacterial reductive reactions, 1:488f

basic concepts, 4:87, 4:341

disease connections

- general discussion, 4:90–91
- mitochondrial mutations, 4:90–91
- neurotoxins, 4:91
- nuclear mutations, 4:91
- pathological mechanisms, 4:91

evolutionary development

general discussion, 4:88–90

N-module, 4:89

origins, 4:89–90

P-module, 4:89

Q-module, 4:88–89

variants, 4:89–90

functional role

active–deactive transition, 4:90

- electron transfer mechanisms, 3:661–662, 3:662f, 3:663, 4:83–84, 4:90

general discussion, 4:90

inhibitors, 4:90

proton pumping, 3:661–662, 3:662f, 3:663, 4:90

redox centers, 4:90

sodium ion (Na⁺) pumping, 4:90

mitochondrial ROS, 4:18–20, 4:19f

PET genes, 3:308f

plant mitochondrial electron transport chain, 1:145, 1:146f

protonmotive electron transfer chain, 1:470f

reverse electron transport (RET), 4:19f, 4:20

schematic diagram, 4:84f, 4:99f

structural characteristics, 4:341–346

- accessory subunits, 4:88, 4:89t
- central subunits, 4:87–88, 4:88f, 4:89t
- cysteine motif, 4:344–345, 4:345f
- flavin mononucleotide (FMN) binding, 4:344–345, 4:344f
- frataxin-like subunit, 4:343–344, 4:343f
- hydrophilic domain, 4:342–343, 4:342f
- iron–sulfur clusters, 4:87–88, 4:89t, 4:342–343, 4:342f, 4:345f
- multistage assembly pathway, 4:88
- mutations, 4:345
- NADH binding, 4:344–345, 4:344f, 4:345f
- overall structure, 4:341–342

redox centers, 4:87–88, 4:89t, 4:342–343, 4:342f

research summary, 4:345–346

shape, 4:87, 4:88f

subunits, 4:342t

supernumerary subunits, 4:345

respiratory electron transport chain, 3:661–662, 3:662f, 3:663, 3:694–695, 3:694f

succinate:quinone oxidoreductases (SQORs), 4:92–97

characteristics and functional role, 4:92, 4:93f

classifications

functional classification, 4:92

structural classification, 4:92–93

electron-coupled proton transfer, 4:96–97, 4:97f

electron transfer, 4:96

facultative phototrophic respiration, 4:79, 4:80f

structure and functions

domain structures, 4:93, 4:94f

flavin adenine dinucleotide (FAD/FADH₂), 4:95, 4:95f

iron–sulfur clusters (FeS), 4:95

quinol oxidation/quinone reduction, 4:95–96

relative orientations, 4:93–94

subunit A, 4:95, 4:95f

subunit B, 4:95

subunits C and D, 4:95–96

succinate oxidation/fumarate reduction, 4:95, 4:95f

thermoplasma-quinone (TK), 4:94f

vitamin K, 3:727–728, 3:728f, 4:552f

quinonoid dihydrobiopterin, 3:666–667, 3:668f

R

rabbits, 2:638t

Rab Escort Protein (REP), 2:737–738

Rab guanosine triphosphatase (GTPase) activating protein (GAP)

Golgi proteins, 2:479

insulin signal transduction, 2:604–605

phosphatidylinositol-3-phosphate (PtdIns(3)P), 3:449

small guanosine phosphatases (GTPases), 4:242, 4:243f, 4:244f

transport pathways, 4:206, 4:504–506, 4:505f

Rab protein family, 4:1–6

cellular localization, 4:2, 4:4f, 4:4t

disease connections, 4:4–6

domain effectors, 4:4

domain structures, 4:1–2, 4:2f

evolutionary development, 4:2

functional cycle, 4:2–4, 4:5f

functional role, 4:4t

phylogenetic relationships, 4:3f

posttranslational modifications, 4:2

research background, 4:1

tissue expression, 4:4t

- Rac1 protein, 1:262–263, 1:263f, 1:430–431, 1:431f, 2:321–322, 2:321f, 4:112f, 4:113t
- Rac2 protein, 4:112f, 4:113t
- Rac3 protein, 4:112f, 4:113t
- radiosensitive severe-combined immune deficiency (RS-SCID) syndrome, 3:273, 3:273t
- radishes, 4:533t
- Rad protein complex
- DNA ligases, 2:30, 2:30t
 - DNA-strand transferases, 4:37, 4:37f, 4:38f
 - Homo sapiens* 9–1–1 clamp, 4:239, 4:239f
 - Homo sapiens* PCNA, 4:239f
 - meiotic recombination, 2:557–558, 2:559f
 - non-homologous end joining (NHEJ), 3:272
 - nucleotide excision repair (NER), 3:338f, 3:341–342, 3:342t, 3:343f, 4:404
 - polymerase α -primase (Pol α), 2:243–244
- Rad50 protein, 3:272
- recombination-dependent DNA replication (RDR), 4:24–25, 4:31–32
- Raf kinases, 1:480–482, 1:482f, 2:528f, 4:247–248, 4:247f
- RAG1/RAG2 proteins, 3:272, 3:272f
- RAGE (receptor for advanced glycation end products), 2:407–409
- ragged red fiber disease, 1:211f, 1:214
- Ralstonia eutropha*, 4:89
- raltegravir, 2:106f
- Ramakrishnan, Venkatraman, 2:254
- Raman spectroscopy, 2:661–662
- RanBP2, 4:468t, 4:469
- ranibizumab, 3:431
- RANK/RANKL receptor proteins, 3:385–386, 4:459
- Ran protein family, 4:7–11
- cellular gradient, 4:8, 4:9f, 4:10f
 - chromatin regulation function, 4:10
 - domain structures, 4:244f
 - functional role, 4:7, 4:243f
 - mitotic spindle regulation, 4:9–10
 - molecular switches, 4:7
 - nuclear envelope assembly regulation, 4:9–10
 - nuclear transport, 4:8–9, 4:10f
 - nucleocytoplasmic transport, 4:7, 4:8f
- RanGAP (GTPase-stimulating protein), 4:8, 4:9f, 4:10f
- RanGEF (guanine nucleotide exchange factor), 4:8, 4:9f
- Ranunculus*, 1:509f
- RAPADILINO, 4:46
- rapamycin (mTOR), 3:194–199
- background information, 3:194
 - branched-chain amino acids (BCAAs), 1:246f
 - cell growth regulation, 1:408
 - characteristics and functional role, 3:194–195, 3:195f
 - chemokine receptors, 1:482f
 - chemotaxis, 2:236–237
 - disease connections, 3:198–199
 - fatty acid metabolism–cancer relationship, 2:277
 - functional role, 3:464f
 - insulin signal transduction, 2:604f, 2:605
 - leptin regulation, 2:259
 - macroautophagy regulation, 1:150t
- mTOR complexes
- cellular functions
 - general discussion, 3:196–198
 - mTORC 1, 3:196–198
 - mTORC 2, 3:198 - characteristics and functional role, 3:194–195, 3:195f, 3:196f
- mTORC 1
- autophagy, 1:150t, 1:152–153, 3:198, 3:450
 - cell cycle/cell survival, 3:198
 - cellular functions, 3:196–198
 - characteristics and functional role, 3:194–195, 3:195f
 - gene transcription regulation, 3:198
 - protein phosphatases, 3:198
 - protein synthesis, 3:196–197, 3:197f
 - regulatory mechanisms, 3:195–196
 - ribosome biogenesis, 3:197–198
 - S6 kinases, 3:196f, 3:197, 3:197f
 - substrates, 3:195, 3:196f
 - regulatory mechanisms, 3:195–196
 - substrates, 3:195, 3:196f
- neurogenesis, 1:199, 1:200f
- phosphatidylinositol-3-phosphate (PtdIns (3)P), 3:450
- phospholipase D (PLD), 3:476
- protein kinase C (PKC), 3:637–638
- Ras proteins, 3:195–196, 4:15f
- research summary, 3:199
- Raphanus*, 1:509f
- Ras proteins, 4:12–16
- angiotensin receptors, 1:113
 - calcium ion (Ca^{2+}) signaling, 2:719–720, 2:720f
 - cell migration, 1:439f
 - chemoattractant signaling, 2:236–238, 2:237f
 - chemokine receptors, 1:480–482, 1:482f
 - diacylglycerol kinases (DGKs), 1:661
 - fibroblast growth factor receptors (FGFRs), 2:305f
 - functional role, 3:439
 - G_q family proteins, 2:498
 - GTPase-activating proteins (GAPs), 4:12–13, 4:13f
 - guanine-nucleotide exchange factors (GEFs), 3:439, 4:12–13, 4:13f, 4:243f, 4:245–246, 4:246f
 - guanosine diphosphate/guanosine triphosphate (GDP/GTP) cycle, 4:12–13, 4:13f, 4:243f
 - hematopoietin receptors, 2:528f
 - integrins, 1:439f, 2:611
 - neurotrophins, 3:241–243, 3:242f
 - phosphatidic acid (PA), 3:475, 3:475f
 - phospholipase C (PLC), 3:470
 - physical properties, 4:12, 4:13f
 - platelet-derived growth factor (PDGF) receptors
 - chemotaxis, 3:542
 - Ras-dependent pathways, 3:540, 3:541f, 3:542
 - signaling pathways, 3:539–540, 3:540f
 - protein palmitoylation, 2:737t, 2:738, 3:645–646, 3:646f
 - protein prenylation, 2:737t
- Rab protein family, 4:1–6
- cellular localization, 4:2, 4:4f, 4:4t
 - disease connections, 4:4–6
 - domain effectors, 4:4
 - domain structures, 4:1–2, 4:2f
 - evolutionary development, 4:2
 - functional cycle, 4:2–4, 4:5f
 - functional role, 4:4t
 - phylogenetic relationships, 4:3f
 - posttranslational modifications, 4:2
 - research background, 4:1
 - tissue expression, 4:4t
- Ras mutations, 4:12–13, 4:13f
- serine–threonine phosphatases, 4:222f, 4:223
- signaling cascades, 3:177f
- signaling nodes, 4:247–248, 4:247f
- small guanosine phosphatases (GTPases), 1:453, 4:242, 4:243f, 4:244f
- steroid/thyroid hormone receptors, 4:310f
- subfamilies
- ARHI/DiRas subfamily, 4:15f, 4:16
 - H-Ras subfamily, 4:13–14, 4:15f
 - K-Ras subfamily, 4:13–14, 4:15f
 - N-Ras subfamily, 4:13–14, 4:15f
 - Ral subfamily, 4:14, 4:15f
 - Rap1/Rap2 subfamily, 3:470, 4:14–15, 4:15f, 4:382–384, 4:383f
 - Rerg/RasL11 subfamily, 4:15f, 4:16
 - RGK subfamily branch, 4:15f, 4:16
 - Rheb subfamily, 3:195–196, 4:15, 4:15f
 - Rit/Rin subfamilies, 4:15f, 4:16
 - R-Ras subfamily, 4:14, 4:15f
 - T-cell receptors (TCRs), 4:374
- ratfish, 4:516t
- rats, 2:5t, 2:638t, 3:642–643
- Rattus norvegicus*, 3:642–643
- Rattus rattus*, 2:5t
- rauwolfscine, 1:58–59
- Raynaud's disease, 3:562
- Rbd1, 4:51f, 4:54
- Rbx1, 4:468t, 4:470
- R-coformycin, 2:214f
- reactive nitrogen species (RNS), 4:17–22
- characteristics, 4:18
 - functional role, 4:17, 4:18f
 - mitochondrial ROS, 4:21
 - regulation mechanisms, 2:77
 - research summary, 4:21
 - ryanodine receptors (RyRs), 4:197
- reactive oxygen species (ROS), 4:17–22
- apoptotic pathways, 1:389f
 - ascorbic acid, 4:531
 - basic concepts, 2:77
 - biological effects
 - aging/age-related diseases, 2:79
 - carcinogenesis, 2:79
 - cytotoxicity, 2:79
 - mutagenesis, 2:79

- calcium/calmodulin-dependent protein kinase II (CaMKII), 1:302
- carotenoids, 1:503–504
- characteristics, 4:18
- chlorophyll, 1:500, 3:482–484, 3:483f
- chloroplast signaling, 1:512, 1:515
- diabetes, 1:656–657
- DNA damage, 2:78–79
- functional role, 4:17, 4:18f
- heme proteins, 2:536
- hypoxia inducible factor (HIF), 4:70–71
- metabolic functions, 3:411–412
- mitochondrial CK (MtCK) isoenzyme, 3:83–84, 3:83f
- mitochondrial ROS
- ATP-dependent potassium ion (K^+) channel (KATP), 3:551
 - control mechanisms, 4:21
 - functional role, 4:17, 4:18f
 - generation mechanisms
 - basic principles, 4:18–20, 4:19f
 - glycerol phosphate dehydrogenase (GPDH), 4:19f, 4:20
 - pyruvate dehydrogenases (PDHs), 4:19f, 4:20
 - respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:18–20, 4:19f
 - respiratory chain complex II (succinate coenzyme Q reductase), 4:19f, 4:20
 - respiratory chain complex III (ubiquinol cytochrome *c* reductase), 4:19f, 4:20
 - reverse electron transport (RET), 4:19f, 4:20
- reversible-to-irreversible damage transition, 3:109–110
- sensing mechanisms, 4:20–21
- nuclear factor kappa B (NF- κ B), 3:303t
- photosynthetic processes, 1:500, 3:482–484, 3:483f, 3:493
- potassium (K^+) cycle, 3:550–551
- regulation mechanisms, 3:407, 3:408f
- research summary, 4:21
- ryanodine receptors (RyRs), 4:197
- sources, 2:77
- STIM signaling, 4:317
- uncoupling proteins (UCPs), 4:485–486, 4:486f
- REACTOME database, 3:615t
- RecA helicases, 2:21, 2:22t, 2:23f, 4:32
- RecA protein
- DNA damage repair, 2:707f, 2:709–710
 - DNA-strand transferases, 4:37, 4:37f, 4:38f
 - meiotic recombination, 2:558, 3:21–22
 - mismatch repair (MMR), 2:47–48, 2:48f
 - SOS response, 2:706, 2:708, 2:708f, 4:477f, 4:478f
 - translesion synthesis (TLS), 4:41, 4:479, 4:479f
- RecBCD helicases, 4:32
- RecB helicases, 4:32
- RecD helicases, 4:32
- receptor activity-modifying proteins (RAMPs)
- binding site distribution, 1:268
 - cellular signaling, 1:268
 - characteristics, 1:273–274
 - CLR–RAMP1 association, 1:266, 1:273, 1:273f, 1:273t
 - CLR–RAMP2/CLR–RAMP3 association, 1:266–267, 1:273, 1:273f, 1:273t
 - molecular structure, 1:266, 1:267f
 - pathophysiology, 1:268–269
 - receptor expression changes, 1:268
 - receptor-enhancing protein 1 (REEP1), 4:269f, 4:270
 - receptor for AGE (RAGE), 2:407–409
 - receptor interacting protein 2 (RIP2), 3:243, 3:244f
 - receptor-interacting protein kinase (RIPK), 1:393–394, 1:394f
 - receptor-interacting protein (RIP), 1:617–619, 1:618f, 4:398–399, 4:399f
 - RecJ helicases, 2:41f, 2:45–46, 2:54–56, 2:55t, 2:56f, 2:67, 2:67f
 - Reclinomonas americana*, 3:133–134
 - recoverin, 1:319t, 2:485t
 - RecQ helicases, 4:43–49
 - anti-recombinogenic roles, 4:46, 4:47f
 - Bloom's syndrome
 - BLM protein
 - binding proteins, 4:45
 - characteristics, 4:45, 4:45f
 - Saccharomyces cerevisiae* Sgs1, 4:45–46
 - characteristics, 4:45–46
 - mutation effects, 4:44
 - characteristics and functional role, 4:27, 4:43–44, 4:44f
 - DNA repair processes, 4:32, 4:44–45
 - exonuclease domain, 4:43–44, 4:44f
 - HRDC domain, 4:43–44, 4:44f
 - meiosis, 4:47
 - mitotic cell division, 4:49
 - pro-recombinogenic roles, 4:46–47
 - replication function, 4:47–49, 4:48f
 - research summary, 4:49
 - Rothmund Thomson syndrome (RTS)
 - characteristics, 4:46
 - mutation effects, 4:44
 - RQC domain, 4:43–44, 4:44f
 - Werner's syndrome
 - characteristics, 4:46
 - mutation effects, 4:44
- red algae, 2:726
- red beets, 2:638t
- red cells, 1:124t, 1:125
- red flour beetle, 4:516t
- reductive pentose phosphate pathway
 - See* Calvin–Benson cycle
- Ref-1 factor, 2:7
- RefSeq database, 3:615t
- Refsum disease, 3:416
- RegA/RegB regulation, 3:695, 4:80–81, 4:80f
- regional centromeres, 1:447, 1:448f
- regulated intramembrane proteolysis (RIP), 4:50–55
 - basic concepts, 4:50
- γ -secretase-based mediation
 - amyloid precursor protein (APP), 4:51f, 4:53
- basic concepts, 4:53
- ErbB-4, 4:51f, 4:53
- Notch, 4:51f, 4:53
- rhomboid-based mediation, 4:51f, 4:54–55
- signal peptide peptidase (SPP)-based mediation, 4:51f, 4:53–54
- site-2 protease (S2P)-based mediation
 - animal cells
 - activation transcription factor 6 (ATF6), 4:51f, 4:52
 - basic concepts, 4:50–52
 - cAMP responsive element-binding protein 3 (CREB3), 4:51f, 4:52
 - sterol regulatory element binding protein (SREBP), 4:51–52, 4:51f
- bacteria
 - cAD1 precursor, 4:51f, 4:52–53
 - RseA, 4:51f, 4:52
 - SpoIVFB protein, 4:51f, 4:53
- basic concepts, 4:50–53
- HEXxH motif, 4:50–53
- Rel homology domain (RHD), 3:302
- renal gluconeogenesis, 2:384, 2:385, 2:386f, 4:305–306, 4:306f
- renin, 1:138f, 1:139
- rennin, 1:329–330
- Reperfusion Injury Salvage Kinase (RISK) group, 3:165–166
- Rep helicases
 - characteristics and functional role, 2:21, 3:261
- DNA unwinding
 - active, rolling or hand-over-hand model, 3:264–265, 3:265f
 - active versus passive mechanisms, 3:264
- dimeric inch-worm models, 3:265, 3:265f
- inch-worm mechanisms, 3:264, 3:264f, 3:265f
- protein oligomerization, 3:263
- unwinding mechanisms, 3:263–264
- research summary, 3:265
- single-strand DNA (ssDNA) translocation, 3:263, 3:264, 3:264f
- structural characteristics
- crystal structures, 3:262, 3:262f
 - domain structures, 3:263f
 - open and closed confirmations, 3:262–263
 - primary structures, 3:261–262
- translocase-based protein displacement, 3:265
- replication factor C (RFC), 2:30, 2:30t, 2:60, 2:61f, 2:120, 3:576, 3:577f, 4:237–238, 4:237f
- replication focus targeting sequence (RFTS), 2:29f, 2:30, 2:30t, 3:8–9, 3:8f
- replication fork (RF)
- bacterial DNA replication
 - DNA polymerase holoenzymes, 2:92–95
 - lagging-strand DNA synthesis, 2:92, 2:93f, 2:94f, 2:115f, 2:116–117
 - leading-strand DNA synthesis, 2:92, 2:93f, 2:94f, 2:116, 2:117
 - functional role, 4:47–49, 4:48f

- replication fork (RF) (*Continued*)
 replication fork (RF) breakage
 backup mechanisms, 4:25–26, 4:25f
 breakage consequences, 4:27, 4:27f
 restart pathways, 4:26–27, 4:27f
 replication protein A (RPA)
 Bloom's syndrome, 4:45
 DNA damage checkpoints proteins,
 1:412–413, 1:413f
 DNA mismatch repair (MMR), 2:52t, 2:61,
 2:61f
 DNA replication, 1:402t, 2:118–119,
 2:119f
 nucleotide excision repair (NER), 4:404
 reproductive tissues
 calbindin-D_{09k}, 1:288
 calbindin-D_{28k}, 1:285–286
 reptiles, 1:600, 1:600f
 Research Institute of Molecular Pathology
 Bioinformatics Group, 3:612
 RESID database, 3:615t
 resistin, 4:303
 resolvases, 4:32–33
 resonance Raman spectroscopy, 2:661–662
 respiratory chain complex I (NADH:
 ubiquinone oxidoreductase), 4:87–91
 background information, 4:98
 bacterial reductive reactions, 1:488f
 basic concepts, 4:87, 4:341
 biochemical free radical chain reactions,
 3:170f, 3:172f, 3:173f
 coenzyme Q, 3:724–725, 3:724f, 4:18–20,
 4:19f
 disease connections
 general discussion, 4:90–91
 mitochondrial mutations, 4:90–91
 neurotoxins, 4:91
 nuclear mutations, 4:91
 pathological mechanisms, 4:91
 electron flow pathways, 4:93f
 evolutionary development
 general discussion, 4:88–90
 N-module, 4:89
 origins, 4:89–90
 P-module, 4:89
 Q-module, 4:88–89
 variants, 4:89–90
 functional role
 active–deactive transition, 4:90
 electron transfer mechanisms,
 3:661–662, 3:662f, 3:663, 4:83–84,
 4:90
 general discussion, 4:90
 inhibitors, 4:90
 proton pumping, 3:661–662, 3:662f,
 3:663, 4:90
 redox centers, 4:90
 sodium ion (Na⁺) pumping, 4:90
 inhibitors, 1:143, 1:143t
 mitochondrial ROS, 4:18–20, 4:19f
 PET genes, 3:308f
 plant mitochondrial electron transport
 chain, 1:145, 1:146f
 protonmotive electron transfer chain,
 1:470f
 reverse electron transport (RET), 4:19f, 4:20
 schematic diagram, 4:84f, 4:99f
 structural characteristics, 4:341–346
 accessory subunits, 4:88, 4:89t
 central subunits, 4:87–88, 4:88f, 4:89t
 cysteine motif, 4:344–345, 4:345f
 flavin mononucleotide (FMN) binding,
 4:344–345, 4:344f
 frataxin-like subunit, 4:343–344, 4:343f
 hydrophilic domain, 4:342–343,
 4:342f
 iron–sulfur clusters, 4:87–88, 4:89t,
 4:342–343, 4:342f, 4:345f
 multistage assembly pathway, 4:88
 mutations, 4:345
 NADH binding, 4:344–345, 4:344f,
 4:345f
 overall structure, 4:341–342
 redox centers, 4:87–88, 4:89t,
 4:342–343, 4:342f
 research summary, 4:345–346
 shape, 4:87, 4:88f
 subunits, 4:342t
 supernumerary subunits, 4:345
 respiratory chain complex II (succinate
 coenzyme Q reductase)
 background information, 4:98
 biochemical free radical chain reactions,
 3:170f, 3:172f, 3:173f
 coenzyme Q, 3:724–725, 3:724f, 4:18–20,
 4:19f
 inhibitors, 1:143, 1:143t
 mitochondrial ROS, 4:19f, 4:20
 PET genes, 3:308f
 plant mitochondrial electron transport
 chain, 1:145, 1:146f
 protonmotive electron transfer chain,
 1:470–471, 1:470f
 schematic diagram, 4:84f, 4:99f
 succinate:quinone oxidoreductases
 (SQORs), 4:92–97
 characteristics and functional role, 4:92,
 4:93f
 classifications
 functional classification, 4:92
 structural classification, 4:92–93
 electron-coupled proton transfer,
 4:96–97, 4:97f
 electron transfer, 4:96
 facultative phototrophic respiration,
 4:79, 4:80f
 structure and functions
 domain structures, 4:93, 4:94f
 flavin adenine dinucleotide (FAD/
 FADH₂), 4:95, 4:95f
 iron–sulfur clusters (FeS), 4:95
 quinol oxidation/quinone reduction,
 4:95–96
 relative orientations, 4:93–94
 subunit A, 4:95, 4:95f
 subunit B, 4:95
 subunits C and D, 4:95–96
 succinate oxidation/fumarate
 reduction, 4:95, 4:95f
 UCP1-induced proton cycling, 4:484f
 respiratory chain complexes, 4:83–86
 complex characteristics and functions
- respiratory chain complex I (NADH:
 ubiquinone oxidoreductase),
 4:83–84
 respiratory chain complex III (ubiquinol
 cytochrome *c* reductase), 4:84, 4:84f
 respiratory chain complex IV
 (cytochrome *c* oxidase), 4:84
 schematic diagram, 4:84f
 components
 cytochromes, 4:83
 flavins, 4:83
 general discussion, 4:83
 iron–sulfur (FeS) proteins, 4:83
 ubiquinone, 4:83
 organization, 4:85
 PET genes, 3:307–309, 3:308f
 ubiquinol-oxidizing enzymes, 4:85
 ubiquinone-reducing enzymes
 electron-transferring flavoprotein (ETF)–
 ubiquinone oxidoreductase, 4:85
 external NADH dehydrogenase, 4:85
 general discussion, 4:85
 glycerol-3-phosphate dehydrogenase,
 4:85
 succinate dehydrogenase, 4:85
 respiratory chain complex III (ubiquinol
 cytochrome *c* reductase), 1:593–598
 action mechanisms
 inhibitors, 1:598
 proton conduction pathways,
 1:597–598, 1:597f
 protonmotive Q cycle, 1:596–597,
 1:597f, 3:663–664, 3:664f
 anoxygenic photosynthesis, 3:694f
 background information, 4:98
 basic concepts, 1:593
 biochemical free radical chain reactions,
 3:170f, 3:172f, 3:173f
 coenzyme Q, 3:724–725, 3:724f, 4:18–20,
 4:19f
 cytochrome *c* interactions, 1:601, 1:601f
 electron transfer complexes, 3:695, 4:93f
 facultative phototrophic respiration, 4:79,
 4:80f
 functional role, 1:593, 1:594f, 4:84
 inhibitors, 1:143, 1:143t
 mitochondrial ROS, 4:19f, 4:20
 oxygen respiration, 3:30–31
 periplasmic electron-transport systems,
 3:400, 3:401f
 PET genes, 3:308f
 photosynthesis/respiration, 3:695
 photosynthetic reaction centers,
 3:693–694, 3:694f, 3:701f
 plant mitochondrial electron transport
 chain, 1:145, 1:146f
 protein subunits, 1:593–594
 protonmotive electron transfer chain,
 1:470–471, 1:470f
 quinone pool functions, 3:695f
 respiratory electron transport chain,
 3:661–662, 3:662f, 3:663–664,
 3:664f, 3:694–695, 3:694f, 4:79, 4:80f
 schematic diagram, 4:84f, 4:99f
 spectroscopic properties, 1:594–596,
 1:595f

- structural characteristics, 1:596, 1:596f
thermodynamic properties, 1:595–596, 1:595f
- respiratory chain complex IV (cytochrome *c* oxidase), 4:98–103
background information, 4:98
biochemical free radical chain reactions, 3:170f, 3:172f, 3:173f
catalytic cycle, 4:101–103, 4:102f
electron transfer complexes, 4:93f
functional role, 4:84
PET genes, 3:308f
plant mitochondrial electron transport chain, 1:145, 1:146f
prosthetic groups, 4:98–100, 4:99f
protein composition, 4:100
protonmotive electron transfer chain, 1:470f
proton pumping, 3:664–665, 4:101–103, 4:102f
respiratory electron transport chain, 3:661–662, 3:662f
schematic diagram, 4:99f
structural characteristics
chemical structure, 4:99f
prosthetic groups, 4:98–100
protein structure, 4:100, 4:101f
proton transfer pathways, 4:101, 4:101f
respiratory chain complex V, 1:145, 1:146f, 3:308f, 4:93f, 4:99f
respiratory electron transport chain
basic concepts, 3:400, 3:401f
periplasmic electron-transport systems, 3:400–406
complex-independent electron transfer, 3:404
formate dehydrogenase, 3:402–403, 3:403f
hydrogenases, 3:403
iron (Fe(III)) respiration, 3:405
Mitchellian redox loop mechanism, 3:400–404, 3:406f
multiheme *c*-type cytochromes, 3:404–405, 3:404f
nitrate reductase A, 3:403–404, 3:403f
periplasmic nitrate reductase (Nap), 1:487t, 3:403f, 3:404
respiratory chain complex III (ubiquinol cytochrome *c* reductase), 3:400, 3:401f
synthesis, 3:405
plant mitochondrial electron transport chain, 1:145, 1:146f
protonmotive electron transfer chain, 1:470f
quinol oxidases (QOX), 3:695f, 3:697–698
quinones, 3:663
respiratory chain complex III (ubiquinol cytochrome *c* reductase), 3:661–662, 3:662f, 3:663–664, 3:664f, 3:694–695, 3:694f, 4:79, 4:80f
respiratory chain complex IV (cytochrome *c* oxidase), 3:661–662, 3:662f
respiratory proteins, 1:143, 1:143t
restrictive cardiomyopathy (RCM), 4:447–449, 4:448f
- reticulocalbin, 1:291t, 1:295, 1:319t
- retina
deep-brain photoreceptors, 3:491
evolutionary development, 1:552f, 1:553, 1:553f
nerve signals, 1:552–553
nonrod/noncone photoreceptors, 3:490, 3:490f
parietal-eye photoreceptor, 3:490
photoreceptors, 1:550, 1:551f, 3:488
primates, 1:554
retinal and choroidal gyrate atrophy, 4:500
rods and cones
characteristics, 1:550
cone opsin genes, 1:551–552, 1:551f, 1:552f
cone photopigments, 1:550–551, 1:551f
cyclic nucleotide-gated (CNG) ion channels, 1:586–587
evolutionary development, 1:553f
hyperpolarization-activated CNG (HCN) ion channels, 1:587–588
morphology
cell body, 3:488, 3:489f
cones, 3:488
general discussion, 3:488, 3:489f
inner segment, 3:488, 3:489f
outer segment, 3:488, 3:489f
synaptic terminal, 3:488, 3:489f
phototransduction mechanisms, 3:489–490, 3:489f
schematic diagram, 1:551f
sodium ion/calcium ion ($\text{Na}^+/\text{Ca}^{2+}$) exchanger
functional role, 3:47
gene families, 3:45
historical research, 3:44
NCX/NCKX expression, 3:45–46
physiological functions, 3:48
voltage-gated calcium channels, 4:561t
retinaldehyde, 3:489–490, 3:489f
retinal dehydrogenases (Raldh), 4:525
retinitis pigmentosa, 1:584
retinoblastoma protein (pRB), 4:104–107
cell–cycle regulation, 1:406–407, 1:407f, 4:105
cellular functions, 4:105
E2F transcription factor, 1:406–407, 1:407f, 4:106–107
functional role, 4:104
future research areas, 4:106
 G_0 phase and G_1 phase cell-cycle regulation, 1:406–407, 1:407f
molecular functions, 4:106–107
pRB-related proteins, 4:104
RB1 tumor suppressor gene, 1:406–407, 1:407f, 4:104–105
regulatory pathways, 4:105–106
9-cis-retinoic acid (9cRA), 3:418–419, 4:525
retinoic acid receptors, 4:108–110
background information, 4:108
domain structures, 4:108, 4:109f
functional role, 4:109–110
functional roles, 4:525
hematopoiesis, 1:188–189
naturally occurring retinoids, 4:529
- peroxisome proliferator-activated receptors (PPARs), 3:418–419, 3:420f, 3:422f
retinoic acid receptor α (RAR α), 1:188–189, 4:109–110
retinoic acid receptor β (RAR β), 4:109–110
retinoic acid receptor γ (RAR γ), 4:109–110
retinoic acid response elements (RAREs), 4:109
retinoid metabolic pathways, 4:527f
signaling pathways, 4:108
structure and functions
C-terminus regions, 4:109
DNA-binding domain, 4:108
general discussion, 4:108–109
N-terminus regions, 4:109
retinoic acid (RA)-binding domain, 4:108–109
transcriptional activation, 4:109, 4:110f
retinoids, 4:525–529
chemical and metabolic reactions, 4:525–526, 4:526t
clinical uses, 4:529
functional roles, 4:525
homeostasis control mechanisms, 4:529
naturally occurring retinoids, 4:529
pharmacological retinoid concentrations, 4:528–529
retinoid-binding proteins (RBPs), 4:526–528, 4:527f, 4:527t
retinoid homeostasis, 4:526–528
retinoid X receptor (RXR)
endoplasmic reticulum (ER), 4:207
insulin signal transduction, 2:604–605, 2:604f
mucins, 3:208f
secretory pathway, 4:207
steroid/thyroid hormone action
mechanisms, 4:308–309, 4:310f
vitamin D receptor (VDR), 4:540, 4:541f
vitamin D-response elements (VDREs), 4:537–538, 4:538f, 4:540, 4:541f, 4:542–543
retinol dehydrogenases (RDHs), 4:525, 4:526t, 4:527f
retinyl ester, 1:49–51
retinyl ester hydrolase (REH), 4:525
retrotransposons, 3:283–291
classifications and characteristics, 3:283, 3:284f
diversity-generating retroelements (DGRs), 3:284
extra-chromosomally-primed (EP) retrotransposons
characteristics, 3:284
integration site preference, 3:288
long terminal repeat (LTR)
retrotransposons, 3:284–286, 3:285f, 3:286f, 3:287f
open reading frames (ORFs), 3:284–286, 3:286f, 3:287f
tyrosine recombinases, 3:285f, 3:286f
virus-like particles (VLPs), 3:284–286, 3:286f
functional role
genome structure and function, 3:289–290

- retrotransposons (*Continued*)
- host gene expression, 3:290
 - laboratory and clinical applications, 3:290–291
 - retroviruses, 3:290
 - future research areas, 3:291
 - integration site preference
 - extra-chromosomally-primed (EP) retrotransposons, 3:288
 - general discussion, 3:288–289
 - target-primed (TP) retrotransposons, 3:288–289
 - major element classes, 3:284
 - regulation mechanisms, 3:289
 - reverse transcriptase (RT), 3:284
 - target-primed (TP) retrotransposons
 - characteristics, 3:284
 - integration site preference, 3:288–289
 - long interspersed elements (LINEs), 3:284, 3:285f, 3:287–288
 - organellar retrotransposons, 3:288
 - Penelope-like elements (PLEs), 3:284, 3:285f, 3:288
 - prokaryotic retrotransposons, 3:288
 - R2-type elements, 3:287, 3:287f
 - short interspersed elements (SINEs), 3:285f, 3:287–288
- retroviruses
- retrotransposons, 3:290
 - retroviral genomes, 2:102, 2:102f
 - reverse transcriptase (RT)
 - biochemical characteristics
 - action mechanisms, 2:103, 2:103f
 - fidelity, 2:96–97, 2:97t, 2:103
 - inhibitors, 2:103–104, 2:106f
 - processivity, 2:103
 - recombination processes, 2:104
 - replication errors/mutations, 2:103
 - enzymatic functions, 2:102–103
 - functional role, 2:104–105, 2:104f, 2:107
 - integrase, 2:105–107, 2:106f
 - integration processes, 2:105–107, 2:106f
 - research background, 2:101
 - retroviral replication, 2:101, 2:105–107
 - ribonuclease H (RNase H) activity, 2:105, 2:106f
 - subunit structure, 2:102f
 - transposable elements (TEs), 3:281
- reverse gyrases
- adenosine triphosphate (ATP) hydrolysis, 2:159–160
 - classifications, 2:158t
 - properties, 2:158
 - protein domains, 2:159f
- reverse transcriptase (RT), 2:101–107
- biochemical characteristics
 - action mechanisms, 2:103, 2:103f
 - fidelity, 2:96–97, 2:97t, 2:103
 - inhibitors, 2:103–104, 2:106f
 - processivity, 2:103
 - recombination processes, 2:104
 - replication errors/mutations, 2:103
 - enzymatic functions, 2:102–103
 - functional role, 2:104–105, 2:104f, 2:107
 - integrase, 2:105–107, 2:106f
 - integration processes, 2:105–107, 2:106f
 - proteinase-activated receptors (PARs), 3:604
 - research background, 2:101
 - retrotransposons, 3:284
 - retroviral genomes, 2:102, 2:102f
 - retroviral replication, 2:101, 2:105–107
 - ribonuclease H (RNase H) activity, 2:105, 2:106f
 - subunit structure, 2:102f
- RfaH protein, 4:410–411
- RGM/DRAGON family, 1:232, 1:232t
- rhabdovirus fusion protein, 3:40f, 3:41–42
- rhamnose, 2:466f
- Rheb subfamily, 3:195–196
- rheumatoid arthritis, 3:562
- Rhizobium etli*, 1:319t
- rhizomelic chondrodysplasia punctata (RCDP), 3:414, 3:417
- Rhod family, 2:575t, 2:576–577
- Rhodnius prolixus*, 2:593–594
- Rhodobacter*, 1:593, 1:594f, 3:32
- Rhodobacter capsulatus*, 2:511, 2:511f, 3:700–702, 4:80–81, 4:80f
- Rhodobacter sphaeroides*
- δ -aminolevulinic acid (ALA) formation, 3:544
 - cytochrome *c* oxidases, 1:603f, 1:604t
 - electron transfer complexes, 3:695–696
 - heme–copper oxidases, 1:604t
 - LexA regulatory protein, 2:710
 - photosynthetic reaction centers, 3:700–702, 3:702f, 3:706f
 - respiration, 3:30–31, 4:79, 4:80f
 - structural characteristics, 4:100
- Rhodococcus*, 3:596–597
- Rhodococcus bronchialis*, 4:353
- Rhodococcus fascians*, 1:631
- Rhodopseudomonas*, 3:32
- Rhodopseudomonas palustris*, 3:693
- Rhodopseudomonas viridis*, 2:512, 3:700–702, 3:701f, 3:702f, 3:706f, 3:725f
- rhodopsin
- archaeal energy transduction, 3:33–34, 3:34f
 - characteristics, 3:488
 - G_i family functions, 2:379
 - G protein-coupled receptor kinases (GRKs), 2:483
 - phototransduction mechanisms, 3:489–490, 3:489f
 - retinoid metabolic pathways, 4:527f
- Rhodospirillaceae, 4:74
- Rhodospirillum*, 3:32
- Rhodospirillum rubrum*
- coenzyme Q, 3:723t
 - cytochrome *bc*₁ complex, 1:593–594
 - cytochrome *c*, 1:599
 - nicotinamide adenine dinucleotide transhydrogenase, 3:247, 3:248f
 - photosynthetic reaction centers, 3:700–702
 - reductive reactions, 1:487t
- Rho protein family, 1:623
- actins, 1:24, 1:430–431, 1:431f
 - bacterial transcription termination, 4:168f, 4:170, 4:409
 - cadherin signaling, 1:262–263, 1:263f
 - Cdc42 protein
 - actin cytoskeleton, 1:430–431, 1:431f, 1:626, 2:321–322, 2:321f
 - cadherin signaling, 1:262–263, 1:263f
 - cell migration, 1:439–440
 - chemokine receptors, 1:480–482
 - cytokinesis, 1:626
 - guanosine triphosphatases (GTPases), 1:440–441, 1:441f, 4:112f, 4:113, 4:113t, 4:114f
 - pheromone receptors, 3:441–442, 3:443f
 - Rho GTP-binding proteins, 1:440–441, 1:441f
 - cell–matrix adhesion diversity, 2:321–322, 2:321f
 - chemokine receptors, 1:482f
 - diacylglycerol kinases (DGKs), 1:660–661
 - functional role, 3:439
 - phospholipase D (PLD), 3:472–473, 3:474–475, 3:474f
 - platelet-derived growth factor (PDGF) receptors, 3:541f, 3:542
 - Rac1 protein, 1:262–263, 1:263f, 1:430–431, 1:431f, 2:321–322, 2:321f, 4:112f, 4:113t
 - Rac2 protein, 4:112f, 4:113t
 - Rac3 protein, 4:112f, 4:113t
 - RhoA protein, 1:480–482, 2:237f, 2:238, 3:245, 3:245f, 4:112f, 4:113t
 - RhoB protein, 4:112f, 4:113t
 - RhoBTB protein, 4:112f
 - RhoC protein, 4:112f, 4:113t
 - RhoD protein, 4:112f
 - RhoE protein, 4:112f
 - RhoF protein, 4:112f
 - RhoG protein, 4:112f, 4:113t, 4:114
 - Rho guanosine triphosphatases (Rho GTPases), 4:111–115
 - actin, 1:20
 - actin cytoskeleton dynamics, 1:430–431, 1:431f, 4:111, 4:248
 - cell–matrix adhesion diversity, 2:321–322, 2:321f
 - cellular functions
 - cell migration, 1:439f, 4:114
 - cell proliferation, 4:115
 - endocytosis, 4:114
 - general discussion, 4:114–115
 - secretion, 4:114
 - vesicle transport, 4:114
 - c-fes* proto-oncogene (Fes), 1:451
 - chemokine receptors, 1:480–482
 - Dbl-related GEFs, 4:111–112, 4:112f
 - DOCK family proteins, 1:441f, 1:480–482, 4:111–112, 4:112f
 - downstream targets
 - Cdc42 protein, 4:112f, 4:113, 4:113t, 4:114f
 - general discussion, 4:113–114
 - G_q family proteins, 2:498
 - Rac effectors, 1:440–441, 1:441f, 4:112f, 4:113, 4:113t

- Rho effectors, 4:113, 4:113*t*
 RhoG protein, 4:112*f*
 RhoH protein, 4:112*f*, 4:113*t*
 Rho-kinase (ROCK), 1:439*f*,
 1:440–441, 1:441*f*, 1:623–624,
 2:237*f*, 2:238, 4:113, 4:113*t*
 Rnd2/Rnd3 proteins, 4:112*f*, 4:113*t*,
 4:114
 formins, 1:24
 G₁₂/G₁₃ family, 2:343–344
 G_q family proteins, 2:498
 GTPase-activating proteins (GAPs),
 2:344, 2:345*f*, 3:489*f*, 4:112, 4:112*f*
 guanine-nucleotide exchange factors
 (GEFs), 2:344, 2:345*f*, 2:494, 3:439,
 4:111–112
 myosin, 3:222–223
 occurrences, 4:111
 phospholipase C- β (PLC β), 3:469–470
 phospholipase D (PLD), 3:474*f*,
 3:475–476
 Rho-GDI (guanine-nucleotide-
 dissociation inhibitors), 3:243,
 3:244*f*, 4:112, 4:112*f*, 4:243*f*,
 4:244*f*, 4:245, 4:246
 Rho GTP-binding proteins, 1:440–441,
 1:441*f*
 small guanosine phosphatases
 (GTPases), 4:242, 4:243*f*, 4:244*f*
 subfamilies, 4:111, 4:112*f*
 tumor necrosis factor receptors (TNFRs),
 4:454
 RhoH protein, 4:112*f*, 4:113*t*, 4:114
 RhoJ protein, 4:112*f*
 RhoQ protein, 4:112*f*, 4:113*t*
 RhoU protein, 4:112*f*
 RhoV protein, 4:112*f*
 ring-shaped helicases, 2:21, 2:22*t*, 2:23*f*
 Rnd2/Rnd3 proteins, 4:112*f*, 4:113*t*, 4:114
 transforming growth factor beta (TGF- β),
 1:620*f*
 rho zero mutants, 3:306
 ribbon–helix–helix motif, 2:151*f*, 2:152
 riboflavin
 chemical properties, 2:309
 importance, 2:309
 megamitochondria, 2:375–376, 2:375*f*
 molecular structure, 2:310*f*
 ribonucleases
 degradosome complex, 3:53
 3'-to-5' exoribonucleases, 3:53–55, 3:54*f*
 5'-to-3' exoribonucleases, 3:56
 3' exoribonucleases, 3:56
 oligonucleotides, 3:56
 polynucleotide phosphorylase (PNPase),
 3:53–55, 3:54*f*, 3:55*f*, 3:56–57, 3:57*f*
 ribonuclease E (RNase E), 3:52–53, 3:53*t*,
 3:54*f*, 3:56–57, 3:57*f*
 ribonuclease G (RNase G), 3:53
 ribonuclease H (RNase H), 2:102–103,
 2:102*f*, 2:105, 2:106*f*, 2:548, 2:549*f*,
 2:551–552
 ribonuclease II (RNase II), 3:53–55, 3:54*f*,
 3:55*f*, 3:56–57
 ribonuclease III (RNase III), 3:56–57
 ribonuclease J1 (RNase J1), 3:56
 ribonuclease J (RNase J), 3:55–56
 ribonuclease P (RNase P), 3:140–141
 ribonuclease Y (RNase Y), 3:56
 transfer RNA (tRNA), 3:555, 3:555*f*
See also ribosomal RNA (rRNA); transfer
 RNA (tRNA)
 ribonucleic acid (RNA)
 catalytic RNA
 characteristics, 4:142
 chemical characteristics, 4:142
 classifications, 4:142–144, 4:143*f*
in vitro evolution
 evolutionary strategies, 4:146, 4:146*f*
in vitro continuous evolution, 4:146
in vitro selection, 4:145–146, 4:146*f*
 systematic evolution of ligands by
 exponential enrichment (SELEX),
 4:144–145
 last common universal ancestor (LUCA),
 4:144, 4:146*f*
 origins of life, 4:144
 phylogenetic distribution, 4:144, 4:146*f*
 unnatural ribozymes, 4:144, 4:145*f*
 centromere identity and function, 1:450
 characteristics, 2:126, 4:142
 double-stranded RNAs, 3:98, 4:155–156,
 4:156*f*
 functional role, 2:126
 genetic information transfer, 4:185, 4:186*f*
 human immunodeficiency virus (HIV-1),
 2:102, 2:102*f*, 2:554
in vitro evolution
 evolutionary strategies, 4:146, 4:146*f*
in vitro continuous evolution, 4:146
in vitro selection, 4:145–146, 4:146*f*
 systematic evolution of ligands by
 exponential enrichment (SELEX),
 4:144–145
 introns, 1:397–400
 alternative splicing, 1:75, 1:76*f*, 3:60–61,
 4:187–188
 characteristics, 4:142, 4:143*f*
 chemical splicing mechanisms, 3:59–60,
 3:60*f*, 4:185, 4:186*f*
 distribution, 4:146*f*
 exon–intron architecture, 3:142–144,
 3:143*f*, 4:186–187
 historical research, 4:185
 intronic splicing enhancers (ISEs), 1:76,
 1:78*f*, 3:62, 3:62*f*
 intronic splicing silencers (ISSs), 3:62*f*
 ribozymes, 4:188
 self-splicing reactions, 1:397, 1:398*f*,
 4:188, 4:188*f*
 splicing pathways, 4:186*f*, 4:187*f*
 structural characteristics
 secondary structures, 1:397, 1:398*f*
 tertiary structures, 1:397–399, 1:398*f*
 substrate recognition
 G•U wobble pairs, 1:398*f*,
 1:399–400, 1:399*f*, 1:400*f*
 guanosine-binding sites, 1:398*f*,
 1:399, 1:399*f*, 1:400*f*
 helix P1/5' splice site recognition,
 1:398*f*, 1:399–400, 1:399*f*, 1:400*f*
 P4–P6 domain, 1:398*f*, 1:399*f*, 1:400*f*
 messenger RNA (mRNA)
 alternative splicing, 1:75–80, 1:81–86
 auxiliary sequence elements, 1:76–77,
 1:78*f*
 background information, 1:75
 CD45 gene, 1:79–80, 1:79*f*
double-sex (dsx) gene regulation,
 1:85–86
Drosophila melanogaster, 1:81–82
 exonic splicing enhancers (ESEs),
 1:76, 1:78*f*, 3:62, 3:62*f*
 exonic splicing silencers (ESSs), 3:62*f*
 exon/intron boundaries, 1:75, 1:76*f*,
 1:397, 1:398*f*
 fine-tuning determinants, 1:78–79
fruitless (fru) gene regulation, 1:85–86
 hnRNP protein family, 1:76–77, 1:78*f*,
 1:79–80, 1:79*f*
 intronic splicing enhancers (ISEs),
 1:76, 1:78*f*, 3:62, 3:62*f*
 intronic splicing silencers (ISSs),
 3:62*f*
 non-*Drosophila* species, 1:86
 post-transcriptional modifications,
 1:78
 pre-messenger RNAs (pre-mRNAs),
 1:75, 1:76*f*, 3:60–61, 4:187–188
 regulation mechanisms, 1:77–78,
 1:78*f*
 research summary, 1:86
 small nuclear RNP (snRNP)
 complexes, 1:75, 1:76*f*, 1:78*f*
 splicing patterns, 1:75–76, 1:77*f*
 splicing silencer sequences, 1:76–77,
 1:78*f*
Sxl autoregulation, 1:82–84, 1:85*f*
transformer (tra) gene regulation, 1:82,
 1:85–86, 1:85*f*
 bacterial decay, 3:52–58
 basic concepts, 3:52
 degradosome complex, 3:53
 3'-to-5' exoribonucleases, 3:53–55,
 3:54*f*
 5'-to-3' exoribonucleases, 3:56
 3' exoribonucleases, 3:56
 gene expression regulation, 3:56–57,
 3:57*f*
 model comparisons, 3:54*f*, 3:56
 oligonucleotides, 3:56
 phosphorolysis versus hydrolysis,
 3:53–55, 3:55*f*
 poly(A) polymerase (PAP)-mediated
 decay, 3:54*f*, 3:55
 polynucleotide phosphorylase
 (PNPase), 3:53–55, 3:54*f*, 3:55*f*,
 3:56–57, 3:57*f*
 ribonuclease E (RNase E), 3:52–53,
 3:53*t*, 3:54*f*, 3:56–57, 3:57*f*
 ribonuclease G (RNase G), 3:53
 ribonuclease III (RNase III), 3:56–57,
 3:57*f*
 ribonuclease J1 (RNase J1), 3:56
 ribonuclease J (RNase J), 3:55–56
 ribonuclease Y (RNase Y), 3:56
 small RNAs (sRNAs), 3:57, 3:57*f*,
 4:249–250

- ribonucleic acid (RNA) (*Continued*)
- tmRNA system, 3:58
 - translation processes, 3:57–58
 - carbohydrate-responsive element
 - binding protein (ChREBP)
 - distribution, 1:374
 - chloroplast-encoded PSII proteins, 3:508
 - energy homeostasis, 3:77
 - eukaryotic messenger RNA (mRNA), 4:432
 - exosomes and microvesicles (EMVs), 2:262–263
 - HAC1 gene expression, 4:489, 4:490f
 - insulin receptors (IRs), 2:602–603
 - lacZYA* mRNA, 2:695f, 2:696–697, 2:699
 - leptin expression, 2:257–258
 - liver regeneration, 1:193
 - maturation processes, 3:188–193
 - cleavage sites, 3:188–189
 - exon definition model, 3:61–62, 3:61f
 - gene expression regulation, 3:192
 - polyadenylation, 3:60–61, 3:188
 - protein complexes, 3:189–191, 3:189f
 - regulation mechanisms, 3:191–192
 - 3'-end maturation, 3:60–61, 3:189–191, 3:189f, 3:190t
 - microRNA (miRNA), 3:98
 - mitochondrial genomes, 3:140–141, 3:142f
 - mucins, 3:206–208
 - Perilipin 1 (Plin1), 1:47–48
 - platelet-activating factor receptor (PAFR), 3:534–535
 - pre-messenger RNAs (pre-mRNAs), 3:188–193
 - alternative splicing, 1:75, 1:76f, 3:60–61, 4:187–188
 - ancillary RNA-processing signals, 3:62, 3:62f
 - chemical splicing mechanisms, 4:185, 4:186f
 - cleavage sites, 3:188–189
 - exon definition model, 3:61–62, 3:61f
 - exonic splicing enhancers (ESEs), 3:62
 - exonic splicing silencers (ESSs), 3:62f
 - exons, 3:142–144, 3:143f, 4:185, 4:186–187, 4:186f, 4:187f
 - 5' and 3' splice sites, 3:59–60, 3:60f, 4:186, 4:186f, 4:187f, 4:188f
 - gene expression regulation, 3:192
 - genetic information transfer, 4:185, 4:186f
 - historical research, 4:185
 - intronic splicing silencers (ISSs), 3:62f
 - introns, 3:142–144, 3:143f, 4:185, 4:186–187, 4:186f, 4:187f
 - life cycle, 3:62–63
 - maturation processes, 3:60–61, 3:188
 - protein complexes, 3:189–191, 3:189f
 - regulation mechanisms, 3:191–192
 - ribozymes, 4:188
 - RNA splicing, 3:142–144, 3:143f
 - self-splicing reactions, 4:188, 4:188f
 - spliceosomes, 3:62–63, 3:63f, 4:187
 - splicing consensus sequences, 4:186, 4:187f
 - 3'-end maturation, 3:60–61, 3:189–191, 3:189f, 3:190t
 - trans*-splicing reactions, 4:188–189, 4:189f
 - proteomic studies, 3:326f
 - retroviral genomes, 2:102, 2:102f
 - ribosomes, 4:116, 4:122–123, 4:128, 4:129f
 - riboswitches, 4:136–141
 - alternative splicing, 4:139
 - aptamer domain structures, 4:137–138, 4:138f
 - cognate ligands, 4:137
 - conformational switching, 4:140–141, 4:141f
 - domain structures, 4:137
 - as drug targets, 4:141
 - gene regulation, 4:139, 4:140f
 - gene transcription regulation, 4:139, 4:140f
 - glmS* riboswitch, 4:138f, 4:141
 - kinetic versus thermodynamic control, 4:139–140
 - research background, 4:136–137
 - small molecule recognition, 4:138–139
 - ThiM* gene, 4:139, 4:140f, 4:141f
 - RNA editing, 4:154, 4:157f
 - RNA polymerase I (Pol I), 2:366t, 2:367–368
 - RNA polymerase II (Pol II), 2:365–368
 - conservation considerations, 2:367–368
 - C-terminal repeat domain (CTD), 2:365, 3:63–64, 3:63f
 - elongation phase, 2:366–367, 3:63–64, 3:63f
 - initiation phase, 2:366, 3:63–64, 3:63f
 - mobile clamps, 2:365
 - structural characteristics, 2:365, 2:366f, 2:366t
 - transcription cycle, 2:365, 3:63–64, 3:63f
 - transcription termination, 2:367, 3:63–64, 3:63f
 - RNA polymerase III (Pol III), 2:366t, 2:367–368
 - RNA processing, 3:59–64
 - ancillary RNA-processing signals, 3:62, 3:62f
 - basic concepts, 3:59–60, 3:60f
 - exon definition model, 3:61–62, 3:61f
 - exonic splicing silencers (ESSs), 3:62f
 - gene transcription regulation, 3:63–64, 3:63f
 - intronic splicing silencers (ISSs), 3:62f
 - life cycle, 3:62–63
 - open reading frames (ORFs), 3:60–61
 - research summary, 3:64
 - RNA-binding proteins, 3:59–60, 3:60f, 3:62–63
 - spliceosomes, 3:62–63, 3:63f
 - RNA trafficking, 3:323
 - thiamine biosynthesis, 1:203–204
 - 3'-end maturation processes
 - basic concepts, 3:59–60, 3:60f
 - cleavage sites, 3:188–189
 - gene expression regulation, 3:192
 - polyadenylation, 3:188
 - protein complexes, 3:189–191, 3:189f, 3:190t
 - regulation mechanisms, 3:191–192
 - untranslated regions (UTRs), 3:188, 3:192
 - transcription and translation processes, 4:418
 - translation initiation, 4:425–431
 - AUG triplets, 4:420–421, 4:433–435
 - basic concepts, 4:425
 - 80S initiation complex, 4:434f
 - 80S ribosome formation, 4:435
 - eukaryotic initiation factors (eIFs), 4:433t, 4:434f, 4:435
 - 5' leader scanning, 4:433
 - 40S ribosomal subunits, 4:432
 - 43S preinitiation complexes, 4:433, 4:434f
 - 48S initiation complex, 4:435
 - initiation codon selection, 4:433–435
 - initiation complex formation, 4:426–427
 - initiation factors, 4:427–431
 - initiation mechanisms, 4:432
 - internal ribosomal entry sites (IRESs), 4:435
 - regulation mechanisms, 4:435
 - ribosomal dissociation, 4:432
 - 70S initiation complex, 4:130f, 4:426–427, 4:427f
 - 60S ribosomal subunits, 4:432, 4:435
 - 30S initiation complex, 4:130f, 4:132f, 4:426–427, 4:427f, 4:429f
 - translation initiation region (TIR), 4:425–426, 4:426f
 - uncoupling proteins (UCPs), 4:484–485, 4:485f
 - microRNA (miRNA), 3:98–102, 4:191–194
 - biogenesis, 3:99–100, 3:99f, 4:191, 4:192f
 - biological role, 3:101–102
 - cardiac muscle development, 1:185
 - characteristics, 3:98
 - exosomes and microvesicles (EMVs), 2:262–263
 - functional role
 - gene transcription regulation, 4:68, 4:191
 - glucose metabolism/insulin signaling, 4:193, 4:194t
 - lipid metabolism, 4:192–193, 4:194t
 - metabolic disease treatment, 4:193, 4:194t
 - pancreatic β -cell function, 4:191–192, 4:194t
 - RNA-induced silencing complex (RISC), 3:100, 3:101f
 - historical research, 3:98–99
 - hypoxia inducible factor (HIF), 4:68
 - RNA editing, 4:155
 - RNA-induced silencing complex (RISC), 3:98, 3:99f, 3:101f

- RNA trafficking, 3:322
skeletal muscle development, 1:182–183
mitochondrial RNA (mRNA), 2:193
piwi-interacting RNAs (piRNAs), 4:253–255
retrotransposons, 3:283–291
 classifications and characteristics, 3:283, 3:284f
 diversity-generating retroelements (DGRs), 3:284
 extra-chromosomally-primed (EP)
 retrotransposons
 characteristics, 3:284
 integration site preference, 3:288
 long terminal repeat (LTR)
 retrotransposons, 3:284–286, 3:285f, 3:286f, 3:287f
 open reading frames (ORFs), 3:284–286, 3:286f, 3:287f
 tyrosine recombinases, 3:285f, 3:286f
 virus-like particles (VLPs), 3:284–286, 3:286f
 functional role
 genome structure and function, 3:289–290
 host gene expression, 3:290
 laboratory and clinical applications, 3:290–291
 retroviruses, 3:290
 future research areas, 3:291
 integration site preference
 extra-chromosomally-primed (EP)
 retrotransposons, 3:288
 general discussion, 3:288–289
 target-primed (TP) retrotransposons, 3:288–289
 major element classes, 3:284
 regulation mechanisms, 3:289
 reverse transcriptase (RT), 3:284
 target-primed (TP) retrotransposons
 characteristics, 3:284
 integration site preference, 3:288–289
 long interspersed elements (LINEs), 3:284, 3:285f, 3:287–288
 organellar retrotransposons, 3:288
 Penelope-like elements (PLEs), 3:284, 3:285f, 3:288
 prokaryotic retrotransposons, 3:288
 R2-type elements, 3:287, 3:287f
 short interspersed elements (SINEs), 3:285f, 3:287–288
ribonucleases
 degradosome complex, 3:53
 5'-to-3' exoribonucleases, 3:142f
 3'-to-5' exoribonucleases, 3:142, 3:142f
 5'-to-3' exoribonucleases, 3:142
 3'-to-5' exoribonucleases, 3:53–55, 3:54f
 5'-to-3' exoribonucleases, 3:56
 3' exoribonucleases, 3:56
 oligonucleotides, 3:56
 phosphorolysis versus hydrolysis, 3:53–55, 3:55f
 poly(A) polymerase (PAP)-mediated decay, 3:54f, 3:55
 polynucleotide phosphorylase (PNPase), 3:53–55, 3:54f, 3:55f, 3:56–57, 3:57f
 ribonuclease E (RNase E), 3:52–53, 3:53t, 3:54f, 3:56–57, 3:57f
 ribonuclease G (RNase G), 3:53
 ribonuclease H (RNase H), 2:102–103, 2:102f, 2:105, 2:106f, 2:548, 2:549f, 2:551–552
 ribonuclease III (RNase III), 3:56–57, 3:57f
 ribonuclease J1 (RNase J1), 3:56
 ribonuclease J (RNase J), 3:55–56
 ribonuclease P (RNase P), 3:140–141, 4:142, 4:146f
 ribonuclease Y (RNase Y), 3:56
 ribosomal RNA (rRNA), 3:554–560, 4:116–121
 assembly processes
 basic principles, 4:116–117
 biogenesis–disease relationship, 4:121
 in vivo eukaryotic ribosome biogenesis, 4:117–121, 4:118f
 in vivo studies, 4:116–117
 model systems, 4:116
 Bacillus subtilis
 5S rRNA, 3:557f, 3:559
 16S rRNA, 3:557f, 3:558–559
 23S rRNA, 3:557f, 3:559
 biogenesis–disease relationship, 4:121
 characteristics, 3:554, 4:116
 defective processing consequences, 3:559
 Escherichia coli
 5S rRNA, 3:557f, 3:558, 3:558f
 maturation processes, 3:557–558
 16S rRNA, 3:557, 3:557f
 23S rRNA, 3:557–558, 3:557f
 maturation processes
 Bacillus subtilis, 3:558–559
 basic concepts, 3:556
 Escherichia coli, 3:557–558, 3:557f
 proteobacteria, 3:559
 mitochondrial genomes, 3:140–141, 3:306–307
 mTOR complexes, 3:197–198
 nucleolus, 3:324–327
 amplified nucleoli, 3:324
 clinical significance, 3:324–325
 dense fibrillar component (DFC), 3:324, 3:325f
 fibrillar centers, 3:324, 3:325f
 functional organization, 3:324, 3:325f
 granular component (GC), 3:324, 3:325f
 historical research, 3:324
 missing nucleoli, 3:324
 perinucleolar compartment (PNC), 3:324–325
 phase-contrast micrograph, 3:325f
 plurifunctional nucleolus, 3:325
 proteomic studies, 3:325, 3:326f
 ribosome biogenesis, 4:120f
 signal-recognition particle (SRP), 3:325
 structural characteristics, 3:324
 quality control mechanisms, 3:559–560
 RNA trafficking, 3:322–323
 transcript structure, 3:556, 3:557f
ribosomes
 assembly processes, 4:116–121
 basic principles, 4:116–117
 biogenesis–disease relationship, 4:121
 in vivo eukaryotic ribosome biogenesis, 4:117–121, 4:118f
 in vivo studies, 4:116–117
 model systems, 4:116
 biogenesis–disease relationship, 4:121
 catalytic RNA, 4:144, 4:146f
 characteristics and functional role, 4:116, 4:122–123, 4:128, 4:129f
 elongation factor G (EF-G), 4:122–127
 background information, 4:122–123
 EF-G–ribosome interactions, 4:123f, 4:125–126, 4:126f, 4:128–129, 4:129f
 research summary, 4:126
 ribosome recycling, 4:126
 translocation cycle, 4:125, 4:132–133
 elongation factor Tu (EF-Tu), 4:122–127
 background information, 4:122–123
 EF-Tu–ribosome interactions, 4:123f, 4:128–129, 4:129f
 guanosine triphosphate (GTP)
 hydrolysis, 4:123–125, 4:124f
 initial selection phase, 4:123, 4:124f
 research summary, 4:126
 substrate proofreading, 4:125
 50S initiation complex, 4:130f
 proteomic studies, 3:326f
 ribosomal dissociation, 4:432
 S6 kinases, 3:196f, 3:197, 3:197f, 3:241–243, 3:242f
 70S initiation complex, 4:130f, 4:426–427, 4:427f
 16S rRNA, 4:130–131, 4:132f
 structural characteristics, 4:128–135
 aminoacyl-transfer RNA (aa-tRNA), 4:133–135, 4:134f
 decoding centers, 4:130f, 4:131, 4:132f
 80S ribosome, 4:135, 4:135f
 elongation factor G (EF-G), 4:133–135, 4:134f
 elongation factor P (EF-P), 4:133–135, 4:134f
 elongation factor Tu (EF-Tu), 4:133–135, 4:134f
 eukaryotic ribosomes, 4:135, 4:135f
 general characteristics, 4:130–131
 initiation factors, 4:133–135, 4:134f
 kink turns, 4:131f
 large subunit, 4:131–132
 packing features, 4:131f
 peptide bond formation, 4:131–132, 4:133f
 release factors I and II, 4:133–135, 4:134f
 research background, 4:129–130
 research summary, 4:135

- ribonucleic acid (RNA) (*Continued*)
- ribosome recycling factor, 4:133–135, 4:134f
 - 70S ribosome, 4:132–133, 4:134f
 - small subunit, 4:130f, 4:131
 - subunit organization and nomenclature, 4:129
 - three-dimensional structure, 4:130f
 - translocation cycle, 4:132–133
 - 30S initiation complex, 4:132f, 4:426–427, 4:427f, 4:429f
 - translation initiation complex
 - formation, 4:128–129, 4:129f, 4:426–427
 - 23S rRNA, 4:130–131, 4:131–132, 4:133f
- ribozymes, 4:142–146
- background information, 4:142
 - characteristics, 4:142
 - chemical characteristics, 4:142
 - classifications, 4:142–144, 4:143f
 - group II introns, 4:147, 4:188
 - in vitro* evolution
 - evolutionary strategies, 4:146, 4:146f
 - in vitro* continuous evolution, 4:146
 - in vitro* selection, 4:145–146, 4:146f
 - systematic evolution of ligands by exponential enrichment (SELEX), 4:144–145
 - last common universal ancestor (LUCA), 4:144, 4:146f
 - origins of life, 4:144
 - phylogenetic distribution, 4:144, 4:146f
 - unnatural ribozymes, 4:144, 4:145f
- RNA editing, 4:154–158
- adenosine deaminases, 4:155–156, 4:156f
 - adenosine-to-inosine (A-to-I) editing, 4:154–155, 4:155f, 4:157f
 - Apolipoprotein B Editing Catalytic subunit 1, APOBEC-1, 4:156, 4:156f
 - basic concepts, 4:154
 - cytidine-to-uridine (C-to-U) editing, 4:155f, 4:156, 4:157f
 - dsRNA-binding motifs, 4:155–156, 4:156f
 - editosome complex, 4:156
 - functional role, 4:157–158, 4:157f
 - nucleotide insertion/deletion, 4:157
 - research background, 4:154
 - serotonin receptors, 4:226, 4:227f
 - viral RNAs, 4:156–157
- RNA maturase, 3:143, 3:143f
- RNA polymerases, 4:173–184
- antibiotic inhibition, 4:171–172, 4:171f
 - bacterial gene transcription regulation, 4:167–172
 - antibiotic inhibition, 4:171–172, 4:171f
 - basic concepts, 4:232
 - elongation cycle, 4:168f, 4:232–233
 - initiation phase, 4:167–169, 4:168f
 - nucleotide addition cycle (NAC), 4:170–171
 - proofreading processes, 4:170–171
 - sigma(σ)-factors, 4:167–169, 4:168f, 4:233–234, 4:233t
 - transcription cycle, 4:167, 4:168f
 - transcription termination, 4:168f, 4:170
- bacterial RNA polymerase, 4:167–172
- bacterial transcription termination
- antitermination, 4:410
 - characteristics, 4:407–410
 - elongation complex, 4:408f
 - elongation factors, 4:168f, 4:169–170, 4:409–410
 - factor-dependent termination, 4:409
 - hairpin formation, 4:407–409, 4:408f
 - intrinsic termination, 4:407–409, 4:408f
 - Mfd-mediated termination, 4:409
 - regulated termination, 4:410
 - Rho-mediated termination, 4:168f, 4:170, 4:409
 - transcription-coupled DNA repair, 4:409
- cis*-acting riboswitches, 4:250, 4:250f
- core enzyme structure
- characteristics, 4:174–176, 4:174f
 - conformational flexibility, 4:176
 - domain structures, 4:175–176, 4:175t
 - nonconserved domains, 4:174f, 4:176
 - RNAP channels, 4:175–176
- DNA-directed RNA polymerases, 4:355, 4:357f
- DNA transcription and replication
- processes, 2:162f, 2:707f
- elongation complex structure
- active center rearrangements, 4:180–182, 4:181f
 - catalytic sites, 4:181f, 4:182–183
 - general characteristics, 4:180–183
 - nucleotide addition cycle (NAC), 4:181f, 4:183
 - protein–DNA/RNA interactions, 4:180, 4:181f, 4:182t
 - structural organization, 4:180, 4:181f
 - translocation mechanisms, 4:181f, 4:183
- enzymatic activities, 4:173
- functional role, 4:159
- future research outlook, 4:183–184
- holoenzyme-promoter DNA complexes
- closed promoter complex (RPC), 4:178f, 4:179, 4:179t
 - open promoter complex (RPO), 4:178f, 4:179–180, 4:179t
 - structural characteristics, 4:178f, 4:179–180
- holoenzyme structure, 4:176–178, 4:177f
- lac* operon, 2:695–696, 2:695f, 4:402
- mitochondrial gene expression, 3:141
- nucleoid-associated proteins (NAPs), 3:367–368
- nucleotide addition cycle (NAC), 4:170–171
- poly(A) polymerase (PAP)-mediated decay, 3:54f, 3:55
- proofreading processes, 4:170–171
- RNA polymerase I (Pol I), 2:366t, 2:367–368, 4:159–161, 4:407–409
- Alu genes, 4:160
 - disease connections, 4:161
 - functional role, 4:159
 - gene transcription regulation, 4:159
 - in vivo* eukaryotic ribosome biogenesis, 4:117–118
 - molecular characteristics, 4:159
 - nucleoli, 4:159
 - retrotransposition, 4:160
 - transcription apparatus, 4:160, 4:160f
 - tumor suppression, 4:160–161
- RNA polymerase II (Pol II), 2:365–368, 4:162–163
- bacterial intrinsic termination, 4:407–409
 - conservation considerations, 2:367–368
 - core promoters, 4:162–163
 - C-terminal repeat domain (CTD), 2:365
 - 1 α ,25-dihydroxyvitamin D₃ (1,25-(OH)₂D₃), 4:538f
 - elongation control, 4:164–166
 - elongation phase, 2:366–367
 - encoding processes, 4:163
 - functional role, 4:159
 - GAL promoter system, 1:526f
 - general transcription factors (GTFs), 4:163
 - in vivo* eukaryotic ribosome biogenesis, 4:117–118
 - initiation phase, 2:366, 4:163
 - mobile clamps, 2:365
 - ribonucleoside triphosphate (NTP), 4:162
 - structural characteristics, 2:365, 2:366f, 2:366t, 4:159
 - structure and functions, 4:162–163
 - transcription-coupled repair (TCR) pathway, 4:403–404, 4:404t
 - transcription cycle, 2:365
 - transcription initiation, 4:163
 - transcription termination, 2:367, 4:163
 - vitamin D receptor (VDR), 4:541f, 4:543–544
- RNA polymerase III (Pol III), 2:366t, 2:367–368, 3:322, 4:159–161, 4:234, 4:407–409
- Alu genes, 4:160
 - disease connections, 4:161
 - functional role, 4:159
 - gene transcription regulation, 4:159
 - molecular characteristics, 4:159
 - nucleoli, 4:159
 - retrotransposition, 4:160
 - transcription apparatus, 4:160, 4:160f
 - tumor suppression, 4:160–161
- sigma(σ)-factors, 4:232–235
- bacterial gene transcription cycle, 4:167–169, 4:168f, 4:232–233
 - classifications, 4:233–234, 4:233t
 - functional role, 4:233t, 4:234
 - phylogenetic distribution, 4:234

- subunit composition, 4:173
- T7 RNA polymerase, 4:355–359
- applications, 4:359
 - DNA replication priming, 4:359
 - initial transcription reaction, 4:356f, 4:357–358
 - pausing or terminator sites, 4:356f, 4:358, 4:358f
 - polymerase sequence similarities, 4:355–357, 4:356f, 4:357f
 - promoter release and elongation, 4:356f, 4:358
 - promoter structure, recognition, and opening, 4:357, 4:358f
 - regulation mechanisms, 4:356f, 4:358–359
 - structural characteristics, 4:355, 4:356f
 - structure–function relationship, 4:355, 4:357f
 - transcription reaction/conformational change, 4:356f, 4:359
- transcription attenuation, 2:360–362, 2:361f
- transcription-coupled repair (TCR)
- pathway, 3:342f, 3:344, 4:402–403, 4:404t
- transcription cycle
- basic concepts, 4:167, 4:168f, 4:418
 - elongation phase, 4:168f, 4:169–170, 4:173–174
 - initiation phase, 4:167–169, 4:168f, 4:173–174
 - Rho-mediated termination, 4:168f, 4:170
 - sigma(σ)-factors, 4:167–169, 4:168f
 - termination phase, 4:168f, 4:170, 4:173–174
- transcription termination, 4:407–411
- attenuation mechanisms, 4:410
 - bacteria, 4:407–410
 - basic concepts, 4:407, 4:408f
 - riboswitches, 4:410, 4:410f
 - RNA polymerase II (Pol II), 2:367
 - site-specific antitermination, 4:410, 4:410f
 - termination and antitermination
 - regulatory factors, 4:410–411
- Tth* σ 70 holoenzyme
- core interactions, 4:176–178, 4:178f
 - domain structures, 4:176–178, 4:177f
 - general characteristics, 4:176–178
 - structural characteristics, 4:176, 4:177f, 4:178f
- UvrABCD system, 3:335–336, 3:335f
- RNA primers, 2:243–244
- RNA splicing, 4:185–190
- alternative splicing, 3:60–61, 4:187–188
 - ancillary RNA-processing signals, 3:62, 3:62f
 - chemical mechanisms, 4:185, 4:186f
 - disease connections, 4:189–190
 - evolutionary implications, 4:189
 - exon definition model, 3:61–62, 3:61f
 - exonic splicing silencers (ESSs), 3:62f
 - exons, 3:59–60, 3:60f, 3:142–144, 3:143f, 4:185, 4:186–187, 4:186f, 4:187f
 - 5' and 3' splice sites, 3:59–60, 3:60f, 4:186, 4:186f, 4:187f, 4:188f
 - genetic information transfer, 4:185, 4:186f
 - intronic splicing silencers (ISSs), 3:62f
 - introns, 3:59–60, 3:60f, 3:142–144, 3:143f, 4:185, 4:186–187, 4:186f, 4:187f
 - life cycle, 3:62–63
 - mitochondrial gene expression, 3:142–144, 3:143f
 - research background, 4:185
 - research summary, 4:190
 - ribozymes, 4:188
 - self-splicing reactions, 4:188, 4:188f
 - spliceosomes, 3:62–63, 3:63f, 4:187
 - splicing consensus sequences, 4:186, 4:187f
 - trans*-splicing reactions, 4:188–189, 4:189f
- RNA synthesis, 4:407
- RNA trafficking
- characteristics, 3:321–323
 - messenger RNA (mRNA), 3:323
 - microRNA (miRNA), 3:322
 - ribosomal RNA (rRNA), 3:322–323
 - transfer RNA (tRNA), 3:322
 - U small nuclear RNA (snRNA), 3:322
- RNA World, 4:142, 4:146f
- 7SK RNA, 4:166, 4:166f
- short interfering RNAs (siRNAs), 1:450, 1:453, 3:98, 3:236, 4:253–255
- small nuclear RNAs (snRNAs)
- basic concepts, 3:59–60, 3:60f
 - intronic splicing silencers (ISSs), 3:62f
 - mitochondrial gene expression, 3:142–144
 - RNA-binding proteins, 3:62–63
 - RNA trafficking, 3:322
 - spliceosomes, 4:187
- small RNAs (sRNAs), 4:249–255
- ARGONAUTE (AGO) protein, 4:253–255
 - bacterial decay, 3:57, 3:57f, 4:249–250
 - characteristics, 4:249–250
 - cis*-acting riboswitches
 - characteristics, 4:250–251
 - RNA ligand sensors, 4:250f, 4:251
 - RNA thermosensors, 4:250, 4:250f - cis*-encoded RNAs, 4:251, 4:252f
 - clustered regularly interspaced short palindromic repeat (CRISPR) RNAs (crRNAs), 4:253, 4:254f
 - microRNA (miRNA), 3:98
 - nucleic acid-binding protein (NABP), 4:252–253, 4:252f
 - protein modulation, 4:252–253
 - research summary, 4:255
 - trans*-encoded RNAs, 4:251–252, 4:252f
- telomeric RNA, 4:384
- tmRNA system, 3:58
- transfer RNA (tRNA), 3:554–560
- aminoacyl-transfer RNA (aa-tRNA),
 - peptidyl transferase center (PTC), 4:131–132, 4:133f
 - protein biosynthesis elongation, 4:122–123, 4:123f, 4:124f, 4:128–129, 4:129f, 4:131, 4:421–423
 - ribosomal subunits, 4:130f, 4:131
 - structural characteristics, 4:134f
 - tRNA synthetases, 4:441
- bacterial selenoprotein synthesis, 4:210, 4:211f
- bacterial translation elongation, 4:419–420, 4:420f, 4:421f
- characteristics, 3:554
- defective processing consequences, 3:559
- elongation cycle
- aminoacyl-tRNA, 4:122–123, 4:123f, 4:124f, 4:128–129, 4:129f, 4:131
 - background information, 4:122–123
 - EF-Tu/EF-G–ribosome interactions, 4:123f, 4:125–126, 4:126f, 4:128–129, 4:129f, 4:421f
 - guanosine triphosphate (GTP) hydrolysis, 4:123–125, 4:124f
 - initial selection phase, 4:123, 4:124f, 4:420–421, 4:421f
 - research summary, 4:126
 - ribosome recycling, 4:126
 - substrate proofreading, 4:125
- 5'-to-3' exoribonucleases, 3:142
- 5'-end–RNase P process pathway, 3:142, 3:142f, 3:555, 3:555f
- glutamyl-transfer RNA (tRNA)
- synthetase, 2:539–540, 2:540f
- maturation processes, 3:554, 3:555f
- mitochondrial genomes, 3:306–307
- quality control mechanisms, 3:559–560
- retroviral genomes, 2:102, 2:102f
- reverse transcriptase (RT) activity, 2:104–105, 2:104f
- ribosomes, 4:116, 4:122–123, 4:128, 4:129f
- RNA trafficking, 3:322
- selenocysteyl-tRNA, 4:212
- structural characteristics, 3:554–555
- 3'-end–RNase process pathway
- Bacillus subtilis*, 3:555f, 3:556
 - Escherichia coli*, 3:555–556, 3:555f
 - Mycoplasma genitalium*, 3:555f, 3:556
 - Saccharomyces cerevisiae*, 3:142, 3:142f
 - Thermotoga maritima*, 3:555f, 3:556
- translation initiation, 4:425–431
- basic concepts, 4:425
 - 40S ribosomal subunits, 4:432–433
 - initiation complex formation, 4:426–427
 - initiation factors, 4:427–431
 - ribosomal dissociation, 4:432
 - 70S initiation complex, 4:130f, 4:426–427, 4:427f
 - 60S ribosomal subunits, 4:432
 - 30S initiation complex, 4:130f, 4:132f, 4:426–427, 4:427f, 4:429f
 - translation initiation region (TIR), 4:425–426, 4:426f

ribonucleic acid (RNA) (*Continued*)

- tRNA synthetases, 4:441–444
 - alternative pathways, 4:442–443
 - amino acid selectivity, 4:441
 - aminoacyl-transfer RNA (aa-tRNA), 4:441
 - attachment mechanisms, 4:442
 - classifications, 4:442
 - disease correlations, 4:443
 - editing activities, 4:442, 4:442
 - enzymatic reactions, 4:442, 4:441
 - functional role, 4:441
 - genetic code expansion, 4:442–443, 4:443f
 - glutamyl-tRNA synthetase, 4:441
 - glutamyl-prolyl-tRNA synthetase, 4:443
 - glutamyl-tRNA synthetase, 4:441
 - inhibitors as drugs, 4:444
 - leucyl-tRNA synthetase, 4:443
 - lysyl-tRNA synthetase, 4:441, 4:443
 - novel functions, 4:443
 - structural characteristics, 4:441, 4:442, 4:442f, 4:443f
 - threonyl-tRNA synthetase, 4:443
 - transamidation reactions, 4:442
 - translational accuracy, 4:441–442
 - tryptophanyl-tRNA synthetase, 4:443
 - tyrosyl-tRNA synthetase, 4:443
- tRNA^{Trp}, 2:363f, 2:364
- U small nuclear RNA (snRNA), 3:322
- zinc-finger domains
 - functional role, 4:577–578
 - structural characteristics, 4:577, 4:578f
- ribonucleotide reductases (RRs)
 - allosteric regulation, 1:73–74
 - cytidine nucleotide formation, 3:716
 - glutathione peroxidases, 3:411f
- ribonucleotide triphosphate reductase, 1:157f, 1:158, 1:158t
- ribonucleotide triphosphate (rNTP), 4:173
- ribose-5-phosphate (R-5-P)
 - pentose phosphate pathway
 - carbon fate, 3:394f
 - chemical balance, 3:394f
 - nonoxidative reactions, 3:391, 3:392f, 3:393f
 - photosynthetic metabolic pathways, 3:499f
 - thiamine diphosphate (ThDP), 1:207f
- ribosomal RNA (rRNA), 3:554–560, 4:116–121
 - assembly processes
 - basic principles, 4:116–117
 - biogenesis–disease relationship, 4:121
 - in vivo* eukaryotic ribosome biogenesis
 - assembly factors, 4:118–119, 4:120f
 - neighborhood associations, 4:120
 - pre-ribosomal intermediates, 4:118–119, 4:119f, 4:120f
 - ribosomal proteins, 4:120
 - ribosomal RNA (rRNA) synthesis and processing, 4:117–118, 4:118f
 - structural proofreading, 4:119, 4:120f
 - subunit quality control mechanisms, 4:120–121
 - in vivo* studies, 4:116–117
 - model systems, 4:116
 - biogenesis–disease relationship, 4:121
 - catalytic RNA, 4:144, 4:146f
 - characteristics and functional role, 4:116, 4:122–123, 4:128, 4:129f
 - 50S initiation complex, 4:130f
 - protein biosynthesis elongation
 - background information, 4:122–123
 - elongation factor G (EF-G), 4:122–127
 - EF-G–ribosome interactions, 4:123f, 4:125–126, 4:126f, 4:128–129, 4:129f, 4:421f
 - research summary, 4:126
 - ribosome recycling, 4:126
 - translocation cycle, 4:125, 4:132–133
 - elongation factor Tu (EF-Tu), 4:122–127
 - EF-Tu–ribosome interactions, 4:123f, 4:128–129, 4:129f, 4:421f
 - guanosine triphosphate (GTP) hydrolysis, 4:123–125, 4:124f
 - initial selection phase, 4:123, 4:124f, 4:421f
 - research summary, 4:126
 - substrate proofreading, 4:125
 - proteomic studies, 3:326f
 - ribosomal dissociation, 4:432
 - S6 kinases, 3:196f, 3:197, 3:197f, 3:241–243, 3:242f
 - 70S initiation complex, 4:130f, 4:426–427, 4:427f
 - structural characteristics, 4:128–135
 - aminoacyl-transfer RNA (aa-tRNA), 4:133–135, 4:134f
 - decoding centers, 4:130f, 4:131, 4:132f
 - 80S ribosome, 4:135, 4:135f
 - elongation factor G (EF-G), 4:133–135, 4:134f
 - elongation factor P (EF-P), 4:133–135, 4:134f
 - elongation factor Tu (EF-Tu), 4:133–135, 4:134f
 - eukaryotic ribosomes, 4:135, 4:135f
 - general characteristics, 4:130–131
 - initiation factors, 4:133–135, 4:134f
 - kink turns, 4:131f
 - large subunit, 4:131–132
 - packing features, 4:131f
 - peptide bond formation, 4:131–132, 4:133f
 - release factors I and II, 4:133–135, 4:134f
 - research background, 4:129–130
 - research summary, 4:135
 - ribosome recycling factor, 4:133–135, 4:134f
 - 70S ribosome, 4:132–133, 4:134f
 - small subunit, 4:130f, 4:131
 - subunit organization and nomenclature, 4:129
 - three-dimensional structure, 4:130f
 - translocation cycle, 4:132–133
 - 30S initiation complex, 4:130f, 4:132f, 4:426–427, 4:427f, 4:429f
 - translation initiation complex formation, 4:128–129, 4:129f, 4:426–427
- riboswitches, 4:136–141
 - characteristics and functional role, 4:136–137
 - functional role
 - alternative splicing, 4:139
 - conformational switching, 4:140–141, 4:141f
 - as drug targets, 4:141
 - gene regulation, 4:139, 4:140f

- gene transcription regulation, 4:139, 4:140f
- glmS* riboswitch, 4:138f, 4:141
- kinetic versus thermodynamic control, 4:139–140
- small molecule recognition, 4:138–139
- ThiM* gene, 4:139, 4:140f, 4:141f
- research background
- antecedent developments, 4:136
 - genetic regulation, 4:136
 - systematic evolution of ligands by exponential enrichment (SELEX), 4:136
- sequence elements, 4:137
- small RNA *cis*-acting riboswitches
- characteristics, 4:250–251
 - RNA ligand sensors, 4:250f, 4:251
 - RNA thermosensors, 4:250, 4:250f
- structures and classes
- aptamer domain structures, 4:137–138, 4:138f
 - cognate ligands, 4:137
 - domain structures, 4:137
 - transcription termination, 4:410, 4:410f
 - untranslated regions (UTRs), 4:137
- riboswitch ribozyme, 4:144, 4:146f
- ribothymidine, 4:441
- ribulose 1,5-bisphosphatase (Rubisco)
- Benson–Calvin cycle, 3:494, 3:494f, 3:495f
 - biological function, 3:373–374
 - chloroplasts, 1:507
 - photosynthetic metabolic pathways, 3:497–498, 3:498f, 3:499–500, 3:499f
- ribulose 1,5-bisphosphate (RuBP), 3:497, 3:499, 3:499f
- ribulose-5-phosphate (Ru-5-P)
- pentose phosphate pathway
 - chemical balance, 3:394f
 - oxidative reactions, 3:391, 3:392f, 3:393f
 - photosynthetic metabolic pathways, 3:499f
- ricin, 2:703f, 2:705, 4:124f
- Ricinus communis*, 2:702t
- rickets, 4:536–537
- See also vitamin D
- Rickettsia conorii*, 3:331f
- Rickettsia prowazekii*, 3:331f
- ridaforolimus, 3:198–199
- Rieske protein
- cytochrome *b₆f* complex, 4:335–338
 - cytochrome *bc₁* complex, 1:596, 1:596f, 3:30–31, 3:663–664, 3:664f
 - cytochrome *f*, 3:32
 - protonmotive Q cycle, 3:663–664, 3:664f
- Rifampicin, 4:171, 4:171f
- RIG-I-like receptors (RLRs), 4:396
- RING finger proteins, 4:455–457
- RING/LIM domain protein, 2:175f, 2:176
- ring-shaped helicases, 2:21
- RNA-dependent DNA polymerase
- See reverse transcriptase (RT)
- RNA-induced silencing complex (RISC), 3:98, 3:99f
- Rnq1p protein
- characteristics, 3:566
 - generation mechanisms, 3:569
 - prion domains, 3:568f
- self-propagating amyloidoses, 3:568, 3:568f
- Ro 15–1788, 2:723–724, 2:724f
- roadblock family, 2:185–186, 2:185t, 2:186f, 2:187
- rodents, 2:638t
- rods and cones
- color vision
 - characteristics, 1:550
 - cone opsin genes, 1:551–552, 1:551f, 1:552f
 - cone photopigments, 1:550–551, 1:551f
 - evolutionary development, 1:553f
 - schematic diagram, 1:551f
- cyclic nucleotide-gated (CNG) ion channels, 1:586–587
- hyperpolarization-activated CNG (HCN) ion channels, 1:587–588
- morphology
- cell body, 3:488, 3:489f
 - cones, 3:488
 - general discussion, 3:488, 3:489f
 - inner segment, 3:488, 3:489f
 - outer segment, 3:488, 3:489f
 - synaptic terminal, 3:488, 3:489f
- phototransduction mechanisms, 3:489–490, 3:489f
- sodium ion/calcium ion ($\text{Na}^+/\text{Ca}^{2+}$) exchanger
- functional role, 3:47
 - gene families, 3:45
 - historical research, 3:44
 - NCX/NCKX expression, 3:45–46
 - physiological functions, 3:48
- romiplostim, 3:431
- rose hips, 4:533t
- rosiglitazone, 3:208f
- Rossmann fold, 1:134, 1:173, 1:176f, 3:247–249, 4:450–451
- rosy* (*ry*) mutation, 2:590
- rotenone, 1:143, 1:143t
- Rothmund Thomson syndrome (RTS)
- characteristics, 4:46
 - DNA secondary structures, 2:147
 - mutation effects, 4:44
- roundworms
- See *Caenorhabditis elegans*
- Rous' virus (RSV), 4:293–294
- RRR- α -tocopherol
- See tocopherols
- (*R,R,R*)-QZ59, 3:212f
- RseA, 4:51f, 4:52
- Rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase)
- biological function, 3:373–374
 - chloroplasts, 1:507
 - photosynthetic carbon dioxide (CO_2) fixation
 - Benson–Calvin cycle, 3:494, 3:494f, 3:495f
 - carbon pathway, 3:499f
 - carboxylation phase, 3:497–498
 - light-regulated energy and power, 3:498f
 - metabolic regulation, 3:499–500
- rubrene, 1:463f
- Rubrivivax gelatinosus*, 3:706–707
- Rumi (*O*-glucosyltransferase), 2:453–454, 2:454f
- RUNX1 transcription factor, 1:187–188, 1:188f
- RuvB helicase, 2:22t
- ryanodine receptors (RyRs), 4:195–198
- associated proteins, 4:197
 - calcium homeostasis, 1:290, 1:310
 - calcium ion (Ca^{2+}) regulation, 4:196
 - calcium ion (Ca^{2+}) release, 2:584, 2:584f, 2:637–638
 - calcium ion (Ca^{2+}) stores, 1:332
 - calmodulin (CaM) regulation, 4:196
 - cardiac excitation–contraction (E–C) coupling, 1:380f
 - cytoplasmic domains, 4:195
 - distribution, 4:195
 - domain structures, 4:195
 - EF-hand proteins, 1:319t
 - inositol trisphosphate (IP_3)/calcium ion (Ca^{2+}) signaling system, 2:584, 2:584f
 - intercellular calcium ion (Ca^{2+}) waves, 2:614–615, 2:614f, 2:615f
 - muscular disorders, 4:197–198
 - pharmacological characteristics, 4:197
 - pore structure, 4:195–196, 4:196f
 - protein kinases, 4:197
 - reactive nitrogen species (RNS), 4:197
 - reactive oxygen species (ROS), 4:197
 - sodium ion/calcium ion ($\text{Na}^+/\text{Ca}^{2+}$) exchanger, 1:379–381, 1:380f, 1:381–382, 1:382f
 - transmembrane domains, 4:195

S

- S100 protein family, 1:311
- S6 kinases, 3:196f, 3:197, 3:197f, 3:241–243, 3:242f
- Saccharomyces cerevisiae*
- actin-related proteins (Arps), 1:37f
 - acyl-coenzyme A synthetase families, 2:292, 2:294–295
 - adenosine diphosphate ribosylation factors (ARF), 1:134
 - amyloid deposits, 3:630
 - AP endonucleases, 2:3–4
 - aquaporin water channels, 1:125
 - aspartic proteases, 1:137, 1:138f, 1:140
 - ATP-binding cassette (ABC) transport systems, 1:7
 - base excision repair (BER), 2:3–4, 2:5t
 - calpains, 1:359
 - centromere region, 1:447, 1:447t, 1:448f
 - cisternal maturation, 4:207
 - coenzyme Q, 3:723t
 - copper-transporting ATPases (copper pumps), 3:671–672
 - cytochrome *bc₁* complex, 1:593–594, 1:595f
 - DNA damage checkpoints proteins, 1:411t
 - DNA mismatch repair (MMR), 2:46–47, 2:48–49
 - DNA replication, 1:401, 1:402t, 2:109, 4:236–237

Saccharomyces cerevisiae (Continued)

exon–intron architecture, 3:142–144, 3:143f, 4:186–187
 extra-chromosomally-primed (EP) retrotransposons, 3:285f
 fermentation reactions, 2:205t
 flippase-mediated transbilayer movement, 2:315–316
 free oligosaccharides (fOSs), 2:331
 GAL1–10 gene expression, 1:525
 glutathione peroxidases, 2:402f
 glycosylphosphatidylinositol (GPI)-anchored proteins, 2:478
 G-protein-signaling regulators (RGSs), 2:492
 GroEL (growth E locus, large gene, bacterial group I chaperonin), 1:459
 heterotrophic glutathione reductase (GR)-type FNRs, 2:301
 hexameric enzyme helicases, 2:22
 hexokinases, 2:543–544, 2:544t, 2:545f
 isoprenylcysteine methyltransferase (ICMT), 3:608–609
 kinesin superfamily proteins (KIFs), 2:679–681, 2:682f
 macroautophagy
 Atg proteins, 1:151–153, 1:152t
 cytoplasm-to-vacuole targeting (Cvt) pathway, 1:151, 1:151f
 model systems, 1:150–151
 molecular mechanisms, 1:151–153
 Atg proteins, 1:151–153, 1:152t
 cargo selection, 1:153
 fusion and breakdown, 1:153
 phagophore assembly site (PAS), 1:152
 regulatory mechanisms, 1:152–153
 meiotic loop–axis structure, 3:21f
 microautophagy
 mitophagy, 1:153
 nonselective microautophagy, 1:153
 pexophagy, 1:153
 mitochondrial DNA (mtDNA)
 functional role, 3:132
 gene expression, 3:140–144
 basic process, 3:140–141, 3:141f
 encoding processes, 3:140
 5′-end–RNase P process pathway, 3:142, 3:142f
 model systems, 3:140, 3:141f
 mosaic mitochondrial genes, 3:142–144, 3:143f
 RNA splicing, 3:142–144, 3:143f
 3′-end–RNase process pathway, 3:142, 3:142f
 transcription processes, 3:141, 3:141f, 3:142f
 translation processes, 3:144
 sequence and gene content, 3:133t
 multidrug resistance (MDR) membrane proteins, 3:214
 N-myristoyltransferase (Nmt), 3:642–643, 3:643f
 nucleoporin proteins (Nups), 3:319f, 3:320t
 nucleotide excision repair (NER), 3:341, 3:342f, 3:342t

origin recognition complex (ORC), 2:111
 peptide amidation, 3:398
 peroxisomal matrix protein import, 3:426–429
PET genes
 model systems, 3:307
 number of proteins, 3:307
 pheromone receptors, 3:441–446
 conformational constraints, 3:445
 conjugation functions, 3:441, 3:442f
 G-protein activation, 3:445–446
 ligand binding, 3:444–445
 receptor biogenesis, 3:443–444, 3:444f
 regulatory mechanisms, 3:442–443, 3:443f, 3:444f
 signal transduction pathways, 3:441–442, 3:443f
 spatial gradient sensors, 3:442
 phosphodiesterases (PDEs), 1:572
 phospholipase D (PLD), 3:473
 phospholipid synthesis, 3:478–481
 basic process, 3:478
 cytidine diphosphate (CDP)–choline pathway, 3:479f, 3:480
 cytidine diphosphate (CDP)–diacylglycerol (DAG) pathway, 3:478–480, 3:479f
 cytidine diphosphate (CDP)–ethanolamine pathway, 3:479f, 3:480
 major phospholipids, 3:478, 3:479f
 phosphatidate (PA) synthesis, 3:478
 phosphatidylglycerolphosphate (PGP), 3:480
 regulation mechanisms
 biochemical regulation, 3:481
 general discussion, 3:480
 genetic regulation, 3:480–481
 plasma membrane proton (H⁺)-ATPases, 3:688
 polymerase δ (Pol δ), 1:402t, 2:245–246, 2:246f, 2:246t
 prions, 3:566, 3:568–569
 protein biosynthesis elongation, 2:253
 proteolysis, 4:62
 recombination-dependent DNA replication (RDR), 4:24–25
 RecQ helicases, 4:43–44, 4:44f, 4:45, 4:47–49
 retrotransposons, 3:283, 3:284f
 rhomboid-based mediation, 4:51f
 ribosomal RNA (rRNA)
 pre-ribosomal intermediates, 4:118–119, 4:119f, 4:120f
 synthesis and processing, 4:117–118, 4:118f
 septins, 4:216–217, 4:216f
 sliding clamp processivity
 9–1–1 complex, 4:239, 4:239f
 proliferating cell nuclear antigen (PCNA), 4:236–237, 4:237f, 4:239, 4:239f, 4:240f
 S phase initiation proteins, 1:401, 1:402t
 sugar nucleotide transporters, 4:350
 superoxide dismutases (SODs), 4:353

telomere proteins, 4:382–384
 3-methyladenine (3mA) DNA glycolase II, 2:11
 translocation cycle, 2:252
 TRiC/CCT eukaryotic chaperonin, 1:459
 20S proteasome, 3:591f
 tyrosine recombinases, 1:556–557
 vacuoles
 biogenesis, 4:503–506
 characteristics, 4:502–503
 functional role, 4:503, 4:504f
 transport pathways
 AP-3 pathway, 4:505f, 4:506
 autophagy/macro-autophagy, 4:505f, 4:506, 4:506f
 characteristics, 4:503–506, 4:505f
 Vps pathway, 4:504–506, 4:506f
Saccharomyces chevalieri, 1:125
Saccharomyces pombe, 2:22
Saccharopolyspora, 1:319t
 S-adenosylhomocysteine (SAH), 3:607, 3:608f
 S-adenosylmethionine (SAM)
 bacteriochlorophyll (BChl) biosynthesis, 2:503–504
 brassinosteroid biosynthesis, 1:251
 carboxyl methylation, 3:607, 3:608f, 3:609f
 DNA alkylation damage, 2:9
 DNA methylation, 3:6
 [FeFe]-hydrogenases, 4:72
 iron–sulfur clusters (FeS), 2:660, 2:661f
 protoporphyrinogen IX, 2:541
 riboswitches, 4:137, 4:138f, 4:251
 salicylhydroxamic acid (SHAM), 1:145–146
 salivary glands
 aquaporin water channels, 1:124t
 digestion and absorption, 2:354
 salmeterol, 1:59
Salmonella
 Dam methylation, 2:43
 nuclear factor kappa B (NF- κ B), 3:303t
 vacuole/lysosome functions, 4:507
Salmonella enterica
 Dam methylation, 2:43
 fatty acyl-coenzyme A (CoA) synthetases, 2:294–295
 type I DNA restriction and modification enzymes, 2:72–74
Salmonella typhi
 macropinocytosis, 2:199
 3-methyladenine (3mA) DNA glycolase I, 2:10–11, 2:11f
Salmonella typhimurium
 Cho phylogenetic tree, 3:331f
 mismatch repair (MMR), 2:54
 phosphotransferase system (PTS), 2:387
 site-specific recombination systems, 1:555–556
 type III DNA restriction and modification enzymes, 2:137t
 uvrC phylogenetic tree, 3:331f
 salty taste receptors, 4:367–368
Sambucus nigra, 2:702t
 Sandhoff disease, 4:291t
 saposins, 4:291

- SAR1 proteins, 1:134
- SARA proteins, 3:450
- sarcin-ricin loop (SRL), 4:124f
- sarcoendoplasmic reticulum Ca^{2+} ATPase (SERCA), 2:226–230, 2:231–235
- ATP-binding site, 2:232–233, 2:233f
- calcium ion (Ca^{2+}) pump reaction cycle, 4:331, 4:332f
- calcium oscillations, 1:324f
- cardiac excitation-contraction (E-C) coupling, 1:380f, 1:381
- catalytic and transport cycles, 2:226–228, 2:227f
- coupling mechanism
- calcium ion (Ca^{2+}) binding and catalytic activation, 2:228
 - free-energy use, 2:228
 - hydrolytic cleavage, 2:228–229
 - nucleotide binding, 2:228
 - phosphorylation-calcium ion (Ca^{2+}) binding interconversion, 2:229
 - phosphoryl transfer, 2:228–229
 - substrate-induced conformational fit, 2:228
- domain structures, 3:684f, 4:330–331, 4:330f
- functional role, 4:330–331
- heart failure, 1:381–382, 1:382f
- homeostasis control mechanisms, 1:290, 1:310
- inhibitors, 2:229
- intercellular calcium ion (Ca^{2+}) waves, 2:614–615, 2:614f
- intracellular calcium ion (Ca^{2+}) stores, 1:127–128
- ion pumping
- $\text{E1} \bullet 2\text{Ca}^{2+} \rightarrow \text{E1P}$ transition/occluded state, 2:232f, 2:233–234, 2:233f
 - $\text{E1P} \rightarrow \text{E2P}$ transition/aspartylphosphate hydrolysis/luminal gate closure, 2:232f, 2:233f, 2:234–235, 2:234f
 - $\text{E1P} \rightarrow \text{E2P}$ transition/ Ca^{2+} release, 2:232f, 2:233f, 2:234, 2:234f, 3:684f, 3:689–690, 3:689f
 - $\text{E2} \rightarrow \text{E1} \rightarrow \text{E1} \bullet 2\text{Ca}^{2+}$ state, 2:232f, 2:233, 2:233f, 4:330f
- IP_3 receptors (IP_3R), 1:290
- physiological considerations, 2:229
- plasma membrane proton (H^+)-ATPases, 3:688–689
- research summary, 2:235
- ryanodine receptors (RyRs), 4:195
- steady-state behavior, 2:226
- store-operated calcium (SOC) channels, 3:359–360, 3:360f, 4:314–315
- structural characteristics, 2:226–228, 2:227f, 2:231, 2:232f, 4:333f
- transmembrane calcium ion (Ca^{2+}) binding sites, 2:231, 2:232f
- sarcomere, 1:31–32, 1:32f, 1:179, 1:180f
- saturated fatty acids
- biosynthesis
 - basic concepts, 2:285–286
 - malonyl-CoA
 - energy metabolism regulation, 2:286
 - lipogenic gene expression, 2:287f
 - pyruvate dehydrogenase regulation, 2:285–286
 - pyruvate dehydrogenase regulation, 2:285–286
 - protein palmitoylation, 2:737
- saxitoxins, 4:568t
- SB (+) 273779 nonpeptide antagonist, 1:265, 1:266f
- scaffold proteins, 1:197f, 1:199
- Scap protein complex, 1:517–519, 1:518f
- scapuloperoneal spinal muscular atrophy (SPSMA), 4:416
- SCAR/WAVE complex, 2:237f, 2:238, 3:438f
- Schiff base linkage, 3:711
- Schindler disease, 4:292
- Schistosoma japonicum*, 1:359
- Schistosoma mansoni*
- calpains, 1:359
 - hexokinases, 2:546
 - peptide amidation, 3:398
- schistosomes, 1:354f
- schizophrenia, 2:173
- Schizosaccharomyces pombe*
- actin-related proteins (Arps), 1:37f
 - acyl-coenzyme A synthetase families, 2:292
 - alkyltransferase-like (ATL) proteins, 2:14
 - base excision repair (BER), 2:5t
 - centromere region, 1:447, 1:447t, 1:448f, 1:450
 - crossover pathways, 2:560
 - cytochrome b_c1 complex, 1:593–594
 - cytokinesis, 1:624–625
 - DNA damage checkpoints proteins, 1:411t
 - DNA replication, 1:401, 1:402t, 2:109
 - kinesins, 2:674
 - mitochondrial DNA (mtDNA), 3:133t
 - origin recognition complex (ORC), 2:111
 - pheromone receptors, 3:441
 - plasma membrane proton (H^+)-ATPases, 3:688
 - polymerase δ (Pol δ), 2:245–246, 2:246t
 - RecQ helicases, 4:43–44, 4:44f
 - retrotransposons, 3:283
 - septins, 4:217–218
 - S phase initiation proteins, 1:401, 1:402t
 - 3-methyladenine (3mA) DNA glycolase II, 2:11
- Schneckenbecken dysplasia, 2:471t
- Schwann cell factor 1 (SC1), 3:243, 3:244f
- Schwartz-Jampel syndrome (SJS), 3:657t, 3:658
- scid, 1:617f
- scinderin, 1:23–24
- SCL transcription factor, 1:187, 1:188f
- SCOP database, 3:615t
- scorpion toxins, 4:568t
- scramblases, 2:314–315, 2:315f
- scrapie, 1:102, 3:571, 3:575t, 3:630
- See also prions
- scurvy, 4:530
- sea anemones
- peptide amidation, 3:398
 - toxins, 4:568t
- seafood, 4:533t
- sea snails, 2:638t
- sea squirts, 2:637–638, 2:638t
- sea urchins, 1:333–334, 1:333f, 2:637–638, 2:638t, 3:538
- sebaceous glands, 2:669t
- Sebastian syndrome, 3:223t
- secondary diabetes mellitus, 1:653
- second messenger pathways, 1:279–280
- secosteroids, 4:308–309
- secosterols, 4:535
- secretases, 4:199–202
- α -secretase, 4:199–200, 4:200f
 - amyloid precursor protein (APP)
 - Alzheimer's disease, 4:199
 - genetic relationships, 1:474, 1:474f
 - process mechanisms, 4:199, 4:200f - β -secretase, 4:200–201, 4:200f, 4:201f
 - γ -secretase, 3:245, 3:245f, 4:200f, 4:201–202
- secreted frizzled related protein 1 (sFRP1), 1:193
- secreted frizzled related protein 3 (sFRP3), 1:199, 1:200f
- secreted sphingomyelinase (sSMase), 4:287
- secretin, 1:272, 2:357–358
- secretory pathway, 4:203–209
- compartmental organization, 4:203, 4:204f
 - endoplasmic reticulum-based protein
 - folding and modification
 - chaperones, 4:205–206
 - N-linked glycosylation, 4:204–205, 4:205f
 - quality control mechanisms, 4:205–206 - microtubule-based motor proteins, 4:208, 4:208f
 - targeted proteins
 - characteristics, 4:203
 - constitutive secretion, 4:208
 - entry mechanisms, 4:203, 4:205f
 - integral membrane proteins, 4:204, 4:205f
 - protein exit, 4:207
 - regulated secretion, 4:208
 - signal recognition particles (SRPs), 4:203, 4:205f
 - soluble proteins, 4:204, 4:205f - transport mechanisms
 - protein localization, 4:206–207, 4:206f
 - retinoid X receptor (RXR), 4:207
 - vesicular transport, 4:206, 4:206f
- secretory vesicles, 3:690, 3:690f
- sedoheptulose-1,7-bisphosphatase (SBPase), 3:499–500, 3:499f
- sedoheptulose-1,7-bisphosphate (S1,7BP), 3:499f
- sedohepulo-7-phosphate (S-7-P)
- pentose phosphate pathway
 - carbon fate, 3:394f
 - chemical balance, 3:394f
 - nonoxidative reactions, 3:391, 3:393f - photosynthetic metabolic pathways, 3:499f
 - thiamine diphosphate (ThDP), 1:207f
- seglitide, 4:264
- seipin, 4:269f, 4:270–271
- selectins, 2:701–703, 2:702f, 2:702t, 2:704t
- selective estrogen receptor modulators (SERMs), 4:311–312

- selenium (Se)
 bacterial reductive reactions, 1:487*t*
 Friedreich's ataxia (FRDA), 2:333
 nickel-iron-selenium (NiFeSe)
 hydrogenases, 2:562–563
- selenocysteine
 aminoacylation pathway, 4:443*f*
 glutathione peroxidases, 2:400–401,
 2:400*f*, 3:409–411, 3:410*f*
 selenocysteine synthase (SelA), 4:210–212,
 4:211*f*, 4:442–443
- selenocysteyl-tRNA, 4:212
- selenophosphate synthetase (SelD),
 4:210–212, 4:211*f*
- selenoprotein synthesis, 4:210–213
 archaeal selenoprotein synthesis, 4:213
 bacterial selenoprotein synthesis
 anticodon loop, 4:210, 4:211*f*
 general discussion, 4:210–213
 selenocysteine biosynthesis
 basic concepts, 4:210–212
 biosynthetic pathway, 4:211*f*
 selenocysteine synthase (SelA),
 4:210–211, 4:211–212, 4:211*f*
 selenocysteine synthase
 (SelB)—guanosine triphosphate
 (GTP)—selenocysteyl-
 tRNA—selenocysteine insertion
 sequence (SECIS) element
 formation, 4:212, 4:213*f*
 transfer RNA (tRNA)^{Sec}, 4:210, 4:211*f*
 selenocysteine insertion sequence
 (SECIS) element, 4:213
 UGA codons
 ribosomal-level decoding, 4:212–213
 selenocysteine insertion sequence
 (SECIS) element, 4:213
 selenocysteine specificity, 4:213
 stop codons, 4:212
 basic concepts, 4:210
 eukaryal selenoprotein synthesis, 4:213,
 4:213*f*
- selenocysteine insertion sequence (SECIS)
 element
 eukaryal selenoprotein synthesis, 4:213,
 4:213*f*
- selenocysteine synthase (SelB)—
 guanosine triphosphate (GTP)—
 selenocysteyl-tRNA—selenocysteine
 insertion sequence (SECIS) element
 formation, 4:212, 4:213*f*
- UGA codons, 4:213
- semicarbazide-sensitive amine oxidases, 1:89
- semiquinone, 3:664*f*
- senescence, 1:415
- sensory rhodopsins, 3:33–34, 3:34*f*
- sentrin/SUMO specific proteases (SENPs),
 4:468*t*, 4:469–470
- separase, 1:589
- sepsis
 hyperammonemia, 4:496*f*
 nuclear factor kappa B (NF- κ B), 3:304*t*
- septins, 4:214–218
 behavior and function
 Ashbya gossypii, 4:216*f*, 4:217
 budding yeasts, 4:216–217, 4:216*f*
- fission yeasts, 4:217–218
 metazoans, 4:217*f*, 4:218
- domain structures
 filamentous complexes, 4:214–215,
 4:215*f*
 higher-order assemblies, 4:215–216,
 4:216*f*, 4:217*f*
 inositol phospholipids, 4:216
 functional role, 4:214–218
- SeqA protein, 2:42, 2:117
- serial analysis of gene expression (SAGE)
 basic concepts, 2:369–370, 2:370*f*
 type III DNA restriction and modification
 enzymes, 2:140
- serine
 classifications, 1:92*t*
 DNA ligases, 2:29*f*
 gluconeogenesis, 2:382*f*
 glycine-serine-arginine (GSR) domain,
 3:105*f*, 3:107
 insect biochemistry, 2:590–591
 LexA regulatory protein, 2:708–709, 2:709*f*
 mitochondrial DNA (mtDNA), 3:134*t*
 neurofilament triplet proteins (NFTPs),
 3:233–235, 3:234*f*
 N-terminal active sites, 3:596, 3:596*f*
 O-linked glycosylation, 2:465, 2:466*f*,
 4:207
 phage libraries, 3:434*t*
 phosphorylation, 4:61
 photosynthetic metabolic pathways,
 3:497–498, 3:498*f*
 proline–histidine–serine–arginine–
 asparagine (PHSRN) sequence, 1:426
 recombinases, 1:558–559, 1:559*f*
 serine–arginine (SR) protein family, 1:76
 serine esterases, 2:759
 serine–glycine interconversion,
 2:324–325
 serine hydroxymethyltransferase (SHMT),
 2:324–325
 serine palmitoyltransferase (SPT), 4:283*f*
 serine proteases, 2:758–759, 2:758*f*,
 3:585–589
 serine proteinase, 3:601
 serine–threonine kinase (Akt)
 arrestins, 2:488*t*
 calcium/calmodulin-dependent protein
 kinase II (CaMKII), 1:297–303
 autonomous activity, 1:299
 autophosphorylation, 1:299, 1:300*f*
 calmodulin (CaM) trapping, 1:300
 cardiac regulation, 1:302
 characteristics and functional role,
 1:297
 domain structures, 1:298–299, 1:299*f*
 intracellular localization, 1:297–298,
 1:298*f*
 long-term potentiation, 1:302
 memory and learning regulation,
 1:302
 neuronal activity monitoring, 1:301,
 1:301*f*
 structure and functions, 1:302–303
 substrate anchoring, 1:300
 substrates, 1:297–298, 1:298*f*
- threonine (Thr³⁰⁵/Thr³⁰⁶)
 autophosphorylation, 1:300–301
 chemoattractant signaling, 2:236–238,
 2:237*f*
 chemokine receptors, 1:480–482, 1:482*f*
 fatty acid metabolism—cancer
 relationship, 2:277
 focal adhesions, 2:321
 functional role, 3:463–464
 glycogen synthase kinase-3 (GSK3),
 2:435*f*
 G protein-coupled receptor kinases
 (GRKs), 2:485*t*
 hematopoietin receptors, 2:528*f*
 insulin signal transduction, 2:604–605,
 2:604*f*
 liver regeneration, 1:193–195
 mitochondrial permeability transition
 pore, 3:165–166
 MSP polymerization-activating kinase
 (MPAK), 3:3*t*, 3:4, 3:4*f*
 mTOR complexes, 1:408, 3:195–196
 neurogenesis, 1:199, 1:200*f*
 phosphatidylinositol-3,4,5-
 trisphosphate (PtdIns-3,4,5-P₃),
 3:450
 plectin, 2:625–627
 protein kinase C (PKC), 3:637–638
 somatostatin (sst) receptors, 4:265
 serine–threonine phosphatases,
 4:219–223
 biochemical characterization, 4:219,
 4:220*t*
 catalytic mechanisms, 4:220, 4:220*f*
 DARPP-32, 4:222*f*, 4:223
 disease connections, 4:223
 nuclear factor of activated T cells (NFAT),
 4:222*f*, 4:223
 phospho-protein phosphatase (PPP)
 biochemical characterization, 4:219,
 4:220*t*
 catalytic mechanisms, 4:220, 4:220*f*
 characteristics, 4:219
 PP1 (calcium/calcineurin-dependent
 type 1 serine/threonine protein
 phosphatase), 1:312, 4:220,
 4:221*f*, 4:222
 protein phosphatase 2A (PP2A),
 4:221, 4:221*f*, 4:222–223
 protein phosphatase 2B (PP2B), 3:3*t*,
 3:4*f*, 3:5, 4:221, 4:221*f*, 4:223
 protein phosphatase 5 (PP5), 4:221,
 4:221*f*
 regulation mechanisms, 4:222
 regulatory subunit interactions, 4:220,
 4:221*f*
 T-cell activation, 4:222*f*, 4:223
 three-dimensional structure,
 4:219–220, 4:220*f*, 4:221*f*
 protein phosphatase Mg²⁺- or Mn²⁺-
 dependent (PPM)
 biochemical characterization, 4:219,
 4:220*t*
 catalytic mechanisms, 4:220
 cellular functions, 4:221
 characteristics, 4:219

- regulation mechanisms, 4:221
 three-dimensional structure, 4:219–220
 subunit families
 phospho-protein phosphatase (PPP), 4:219
 protein phosphatase Mg^{2+} - or Mn^{2+} -dependent (PPM), 4:219
 three-dimensional structure, 4:219–220, 4:220f
 site-specific recombination systems, 1:558–559, 1:559f
 small guanosine phosphatases (GTPases), 4:247f
 tricarboxylic acid cycle, 4:439f
 vasopressin/oxytocin family, 4:516t
See also Akt (serine–threonine active kinase protein)
 seritocin, 4:516t
 serotonin, 4:224–228
 adenosine-to-inosine (A-to-I) RNA editing, 4:155
 amino acid metabolism, 1:93t
 G-protein-coupled receptors (GPCRs), 2:344t, 4:224, 4:225f, 4:225t
 G_q family proteins, 2:497t
 monoamine oxidase, 3:187f
 serotonin receptors
 adenylyl cyclase activation, 4:225–226, 4:225t
 adenylyl cyclase inhibition, 4:224–225, 4:225t
 alternative splice variants, 4:226
 functional selectivity, 4:227
 5-hydroxytryptamine-2C receptor (5-HT2cR)
 adenosine deaminases, 4:155–156, 4:156f
 adenosine-to-inosine (A-to-I) RNA editing, 4:155
 RNA editing, 4:226, 4:227f
 ligand-gated ion channels (LGICs), 4:226
 phospholipase C (PLC), 4:225f, 4:226
 receptor diversity functions, 4:227–228
 signal transduction, 4:227
 single-nucleotide polymorphisms, 4:226–227
 structure and functions, 4:224, 4:225f, 4:225t
 synthesis and metabolism, 4:224
 transport mechanisms, 3:238–239
 Serrate/Jagged protein family, 2:455
 SERRATE protein, 3:99–100
 serum amyloid A, 1:101t
 serum amyloid protein (SAP), 2:703f
 serum response factor (SRF), 1:182–183, 1:184f, 3:293, 3:293f
 seryl-tRNA synthetase (SerS), 4:210–212, 4:211f
 SET- and RING-associated (SRA) domain, 3:9, 3:9f
 7-methylguanine (7mG), 2:9, 2:10t
 7SK RNA, 4:166, 4:166f
 seven-transmembrane receptors
 chemokine receptors, 1:480–485
 desensitization mechanisms, 1:482–483, 1:483f
 functional roles
 angiogenesis, 1:484
 central nervous system (CNS), 1:484–485
 chemotaxis, 1:483–484
 human immunodeficiency virus (HIV-1), 1:484
 tumor growth and metastasis, 1:484
 internalization, 1:483
 intracellular signaling pathways, 1:480–482, 1:482f
 intracellular trafficking, 1:482–483, 1:483f
 ligand binding, 1:480–482, 1:481f, 1:482f
 subfamily classifications, 1:480, 1:481t
 dopamine receptors, 2:171
 metabotropic glutamate (mGlu) receptors, 2:395–396
 muscarinic acetylcholine receptors, 3:216
 platelet-activating factor receptor (PAFR), 3:533–537
 general discussion, 3:533
 G_q family proteins, 2:497t
 PAF acetylhydrolase, 3:533–534
 platelet-activating factor (PAF), 3:533–534, 3:534f
 regulation mechanisms, 3:535–536
 signal transduction, 3:536, 3:536f
 structural characteristics, 3:534–535, 3:535f
 transcript structure, 3:534–535, 3:535f
 proteinase-activated receptors (PARs), 3:601–606
 activating and inactivating proteinases, 3:601, 3:602f
 activation mechanisms, 3:601, 3:602f
 agonists, 3:601–602, 3:603t
 antagonists, 3:603–604, 3:603f
 bioassays
 calcium signaling measurement, 3:604
 platelet aggregation, 3:605
 tissue bioassays, 3:604–605
 desensitization mechanisms, 3:602
 detection methods, 3:604
 distribution, 3:604, 3:605t
 domain structures, 3:601, 3:603t
 intermolecular activation, 3:602–603
 internalization, 3:602
 localizations, 3:605t
 PAR-activating peptides (PAR-APs), 3:601–602, 3:602f, 3:603t
 physiological functions, 3:605
 receptor dimerization, 3:602–603
 resensitization, 3:602
 thrombin, 3:601, 3:602f, 3:603t
 severe-combined immune deficiency (SCID) syndrome, 1:616, 3:273, 3:273t, 3:362–364, 3:649–651
 severin, 1:23, 3:438–439
 sex-determining region Y (SRY)-box 2 (Sox2), 1:196–197, 1:197f
Sex-lethal (*Sxl*) master-switch gene
 autoregulation, 1:82–84, 1:85f
 background information, 1:81
 exon organization and splicing patterns, 1:81–82, 1:82f
 regulatory hierarchy, 1:81, 1:82f
 transformer (*tra*) gene regulation, 1:82, 1:85f
 sexual dysfunction drugs, 3:27
 sexual life cycle, 3:17
 See also meiosis
 SF1 (superfamily 1) helicases, 3:261–265
 characteristics and functional role, 3:261
 DNA unwinding
 active, rolling or hand-over-hand model, 3:264–265, 3:265f
 active versus passive mechanisms, 3:264
 dimeric inch-worm models, 3:265, 3:265f
 inch-worm mechanisms, 3:264, 3:264f, 3:265f
 protein oligomerization, 3:263
 unwinding mechanisms, 3:263–264
 research summary, 3:265
 single-strand DNA (ssDNA) translocation, 3:263, 3:264, 3:264f
 structural characteristics
 crystal structures, 3:262, 3:262f
 domain structures, 3:263f
 helicase motifs, 3:263f
 open and closed confirmations, 3:262–263, 3:263f
 primary structures, 3:261–262
 translocase-based protein displacement, 3:265
 Sfil enzyme, 2:131, 2:132t, 2:133
 SH2/SH3 domains, 1:451–452, 1:452f, 3:470, 4:294, 4:295–297, 4:296f
Shaker gene, 4:555–556, 4:556–557
 Shaker K-channels, 1:587–588
 sheep, 3:575t
 shellfish, 4:533t
 shelterin, 4:385
 Shemin pathway, 3:544
Shewanella oneidensis, 1:487t
Shewanella putrefaciens, 1:487t, 3:331f
Shigella, 4:507
 Shine–Dalgarno (SD) sequence, 4:425–426, 4:426f
 short-chain acyl-CoA synthetase (SCAS), 2:292
 short-chain dehydrogenases/reductases (SDRs), 4:525–526
 short interfering RNAs (siRNAs), 1:450, 1:453, 3:98, 3:236, 4:253–255
 short interspersed elements (SINEs), 3:285f, 3:287–288
 SHP-2 (Src homology 2 domain-containing protein tyrosine phosphatase), 1:113, 3:539–540, 3:540f, 3:541f, 3:542, 3:651t
 shrimp, 4:533t
 SHU9119, 3:27
 sialic acid
 O-fucose-based modifications, 2:453, 2:454f
 N-linked glycosylation, 2:458–460, 2:459f, 2:470, 2:470f
 sialyltransferase (SiaT), 2:454f

- sialic acid (*Continued*)
- siglecs, 4:229–231
 - binding sites, 4:231
 - cell-type specific expression, 4:229
 - cytosolic tails, 4:231
 - functional roles, 4:230, 4:231
 - genomic organization, 4:229–230
 - historical research, 4:229
 - immunoreceptor tyrosine-based inhibition motifs (ITIMs), 4:231
 - N*-acetylneuraminic acid (Neu5Ac)/*N*-glycolylneuraminic acid (Neu5Gc), 4:230–231
 - phylogenetic relationships, 4:229–230
 - siglec-1, 2:702*t*, 2:704*t*
 - structural characteristics, 4:229, 4:230*f*
 - subgroups, 4:229
 - sialoadhesin, 4:229, 4:230, 4:230*f*, 4:231
 - sialyl Lewis X antigen, 1:2*f*
 - Sicca syndrome, 3:564
 - sickle cell hemoglobin (HbS), 2:534–535
 - Siemens ichthyosis bullosa, 2:670*t*
 - siglecs, 4:229–231
 - binding sites, 4:231
 - cell-type specific expression, 4:229
 - cytosolic tails, 4:231
 - functional roles, 4:230, 4:231
 - genomic organization, 4:229–230
 - historical research, 4:229
 - immunoreceptor tyrosine-based inhibition motifs (ITIMs), 4:231
 - N*-acetylneuraminic acid (Neu5Ac)/*N*-glycolylneuraminic acid (Neu5Gc), 4:230–231
 - phylogenetic relationships, 4:229–230
 - siglec-1, 2:702*t*, 2:704*t*
 - structural characteristics, 4:229, 4:230*f*
 - subgroups, 4:229
 - signal recognition particles (SRPs), 3:325, 4:203, 4:205*f*
 - signal transducer and activator of transcription (STAT)
 - c-fes* proto-oncogene (*Fes*), 1:452, 1:453–454
 - chemokine receptors, 1:480–482
 - hematopoietin receptors, 2:526, 2:528*f*
 - interferon receptors, 2:620–621, 2:620*f*, 2:622*f*
 - leptin resistance, 2:260–261, 2:260*f*
 - myocardial ischemia, 3:111
 - neurogenesis, 1:198
 - signaling pathways, 1:564–565, 1:564*f*, 1:616, 1:617*f*, 1:618*f*, 3:539–540, 3:540*f*
 - sildenafil, 1:569, 3:27
 - silencing mediator of retinoic acid and thyroid hormone receptors (SMRTs), 1:311
 - silver (Ag^+)-adenosine triphosphatase (ATPase), 4:334
 - Silverman–Handmaker dyssegmental dysplasia (DDSH), 3:657*t*, 3:658
 - simian virus 40 large T-antigen, 2:21, 2:22*t*, 2:25–26
 - Simpson–Golabi–Behmel syndrome, 3:657*t*
 - single-strand annealing (SSA), 4:29
 - single-strand DNA-binding protein (SSB), 2:51, 2:52*t*, 2:54–56, 2:55*t*, 2:56*f*
 - single-well hydrogen bonds
 - bond lengths, 2:757
 - deuterium fractionation factors, 2:758
 - spectroscopic properties, 2:758
 - structural characteristics, 2:757, 2:757*f*
 - Sinorhizobium loti*, 3:331*f*
 - Sinorhizobium meliloti*, 3:559
 - Sin recombinase, 1:560
 - siroheme
 - biosynthetic pathway, 3:545*f*, 3:546–547, 3:546*f*
 - characteristics, 3:544
 - sirolimus, 3:198–199
 - sirtuins, 1:177, 1:177*f*, 3:420
 - sister chromatids
 - meiosis, 3:18
 - mitosis, 3:17–18, 3:181, 3:182*f*
 - sister chromatid exchange (SCE), 2:557, 2:559*f*, 4:45–46
 - site 1 protease (S1P), 1:517–519, 1:518*f*
 - site 2 protease (S2P), 1:517–519, 1:518*f*
 - 6–4 photoproducts (6–4PP), 3:338, 3:339*f*, 3:341
 - Sjögren–Larsson syndrome
 - ALDH protein, 4:270–272
 - leukotriene metabolism, 3:584
 - Sjögren's syndrome, 1:124, 3:562
 - skeletal dysplasias, 4:416
 - skeletal muscle
 - acyl-coenzyme A synthetase
 - long-chain acyl-CoA synthetase (LCAS), 2:292–293
 - medium-chain acyl-CoA synthetase (MCAS), 2:292
 - short-chain acyl-CoA synthetase (SCAS), 2:292
 - very long-chain acyl-CoA synthetase (VLCAS), 2:293–294
 - adenosine triphosphate (ATP) turnover, 1:108
 - anaplerotic pathways, 1:108
 - branched-chain α -ketoacid dehydrogenase complex (BCKDC), 1:247–248
 - branched-chain amino acids (BCAAs), 1:244–245
 - carbohydrate-responsive element binding protein (ChREBP) distribution, 1:374
 - carnitine biosynthesis, 1:385
 - α -dystroglycan function, 2:454*f*, 2:456
 - G_s family phenotypes, 2:521*t*
 - insulin metabolism, 2:603*f*, 2:606–607
 - myosin light chain kinase (MLCK), 1:305–306, 1:305*f*
 - striated muscles, 1:179–186
 - development process
 - microRNA (miRNA), 1:182–183
 - myogenesis, 1:179–180, 1:180*f*, 1:181*f*
 - transcription factors, 1:180–182, 1:181*f*
 - research summary, 1:186
 - stem cells, 1:186
 - structural characteristics, 1:179, 1:180*f*
 - uncoupling proteins (UCPs), 4:485, 4:485*f*
 - voltage-gated calcium channels, 2:648–649, 4:561*t*
 - skin
 - aquaporin water channels, 1:124*t*
 - cadherins, 1:259–260
 - keratins, 2:665–671
 - characteristics, 2:665, 2:667, 2:668*f*
 - classifications, 2:665, 2:666*f*
 - disease connections, 2:670–671, 2:670*t*
 - domain structures, 2:667, 2:668*f*
 - functional role
 - signaling pathways, 2:670
 - structural support, 2:668*f*, 2:669–670
 - gene expression
 - hair follicles, 2:669, 2:669*t*
 - mammalian skin tissue, 2:667–669, 2:668*f*, 2:669*t*
 - thick epidermis, 2:669, 2:669*t*
 - thin epidermis, 2:667–669, 2:669*t*
 - intermediate filaments, 2:632*t*, 2:634
 - intracellular regulation, 2:667, 2:668*f*
 - posttranslational modifications, 2:667
 - self-assembly control, 2:667
 - tissue expression, 2:665–667, 2:666*f*
 - skin fragility ectodermal dysplasia syndrome, 1:638*t*
 - SLC33A1 (SPG42) gene, 4:272
 - Sld2/3 loading factors, 1:402*t*, 1:403*f*, 1:404*f*
 - slime molds, 1:38–40
 - slipped structures
 - background information, 2:146
 - biological role, 2:146
 - chemical properties, 2:146
 - structural characteristics, 2:143*f*
 - slowpoke (slo)* gene, 4:559
 - Smad signaling
 - characteristics and functional role, 1:232–233
 - downstream signaling pathways, 1:238
 - Smad4, 1:238
 - Smad6, 1:238
 - vascular development and disease, 1:238
 - nontranscriptional Smad-dependent responses, 1:233
 - phosphatidylinositol-3-phosphate (PtdIns (3)P), 3:450
 - regulation mechanisms, 1:233
 - Smad-dependent transcriptional responses, 1:233
 - Smad transcription factors, 1:620, 1:620*f*
 - small-angle X-ray scattering (SAXS), 4:140–141, 4:141*f*
 - small GTP-binding proteins, 3:439
 - small guanosine phosphatases (GTPases), 4:242–248
 - actin cytoskeleton, 4:248
 - activating proteins, 4:246
 - calcium ion (Ca^{2+}) signaling, 2:719–720, 2:720*f*
 - cell migration, 1:439–440
 - cell proliferation regulation, 4:247–248, 4:247*f*
 - c-fes* proto-oncogene (*Fes*), 1:451, 1:453
 - chemokine receptors, 1:483*f*
 - dissociation inhibitors, 4:243*f*, 4:244*f*, 4:245, 4:246

- domain structures, 4:242–245, 4:244f
functional role, 4:243f, 4:244f, 4:247
genetic deregulation mechanisms, 4:246–247
guanine-nucleotide-exchange factors (GEFs), 1:453, 4:245
hemidesmosome function, 1:642–643
membrane transport, 4:248
molecular switches, 4:7
nucleocytoplasmic transport, 4:7, 4:8f
phosphatidylinositol-3-phosphate (PtdIns(3)P), 3:449
phospholipase C- ϵ (PLC ϵ), 3:470
Rab protein family, 4:1–6
cellular localization, 4:2, 4:4f, 4:4t
disease connections, 4:4–6
domain effectors, 4:4
domain structures, 4:1–2, 4:2f
evolutionary development, 4:2
functional cycle, 4:2–4, 4:5f
functional role, 4:4t
phylogenetic relationships, 4:3f
posttranslational modifications, 4:2
research background, 4:1
tissue expression, 4:4t
Ran protein family, 4:7–11
cellular gradient, 4:8, 4:9f, 4:10f
chromatin regulation function, 4:10
domain structures, 4:244f
functional role, 4:7, 4:243f
mitotic spindle regulation, 4:9–10
molecular switches, 4:7
nuclear envelope assembly regulation, 4:9–10
nuclear transport, 4:8–9, 4:10f
nucleocytoplasmic transport, 4:7, 4:8f
RanGAP (GTPase-stimulating protein), 4:8, 4:9f, 4:10f
RanGEF (guanine nucleotide exchange factor), 4:8, 4:9f
small GTPase Ran, 3:320–321, 3:321f, 4:7, 4:8f
regulation mechanisms, 4:245
research background, 4:242
signaling cascades, 4:245–246, 4:246f
small intestine
bile acid secretion and absorption, 2:358
carbohydrate digestion and absorption, 2:359
digestion processes, 2:355f
intestinal pH, 2:357
lipid digestion and absorption
cholesterol, 2:358
phospholipids, 2:358
triglyceride (TG), 2:358
nutrient transit, 2:356–357
pancreatic exocrine secretion, 2:357–358, 2:357f
protein digestion and absorption, 2:358–359
small nuclear RNAs (snRNAs)
basic concepts, 3:59–60, 3:60f
intronic splicing silencers (ISSs), 3:62f
mitochondrial gene expression, 3:142–144
RNA-binding proteins, 3:62–63
RNA trafficking, 3:322
spliceosomes, 4:187
small nuclear RNP (snRNP) complexes, 1:75, 1:76f, 1:78f
small RNAs (sRNAs), 4:249–255
ARGONAUTE (AGO) protein, 4:253–255
bacterial decay, 3:57, 3:57f, 4:249–250
characteristics, 4:249–250
cis-acting riboswitches
characteristics, 4:250–251
RNA ligand sensors, 4:250f, 4:251
RNA thermosensors, 4:250, 4:250f
cis-encoded RNAs, 4:251, 4:252f
clustered regularly interspaced short palindromic repeat (CRISPR) RNAs (crRNAs), 4:253, 4:254f
microRNA (miRNA), 3:98
nucleic acid-binding protein (NABP), 4:252–253, 4:252f
protein modulation, 4:252–253
research summary, 4:255
trans-encoded RNAs, 4:251–252, 4:252f
small ubiquitin-related modifiers (SUMOs), 3:191–192, 4:69, 4:468–470, 4:468t, 4:475
smooth muscles
actin organization, 1:32–33
smooth muscle contraction, 1:32–33, 2:642
smooth muscle myosin light chain kinase (smMLCK), 1:32–33, 1:305–306, 1:305f
voltage-gated calcium channels, 4:561t
snails
lectins, 2:702t
vasopressin/oxytocin family, 4:516t
SNARE proteins, 4:206, 4:504–506, 4:505f
Snell's Waltzer mouse, 3:223–224
SNIP1 protein, 3:99–100
snowdrops, 2:702t, 2:704t
SNX proteins, 3:449
SOD enzyme, 4:21
sodium-ascorbate co-transporters (SVCT1/2), 4:532–533
sodium-dependent glucose transporters (SGLTs), 2:391, 2:392f
sodium/galactose transport (SglT), 2:388, 2:389, 2:389f
sodium (Na), 4:161
sodium benzoate, 4:495–497, 4:497t
sodium dodecyl sulfate, 1:644–645, 1:645f
sodium ion/calcium ion (Na⁺/Ca²⁺) exchanger, 3:44–48
calcium ion (Ca²⁺)-binding domains, 4:257f, 4:259–260, 4:260f
calpains, 1:424, 1:425f
cellular calcium ion (Ca²⁺) homeostasis, 1:290, 3:44
cytoplasmic domain structure–function relationship
calcium ion (Ca²⁺)-binding domains, 4:257f, 4:259–260, 4:260f
C-terminus regions, 4:261
cytoplasmic loop f, 4:258
cytoplasmic loops b and d, 4:260–261
exchanger inhibitory peptide region, 4:258–259, 4:259f
linker domain 1, 4:259
linker domain 2, 4:260
Na⁺-dependent inactivation, 4:258–259, 4:259f
excitotoxicity, 1:424, 1:425f
functional role, 3:47, 3:47f, 4:256
gene families, 3:45, 3:45f, 3:46f
historical research, 3:44
mitochondrial calcium (Ca²⁺) transport, 3:125–126
NCLX protein superfamily, 3:125
NCX1 dimerization, 4:257–258
NCX1 helix packing, 4:256–257, 4:258f
NCX1 membrane topology, 4:256, 4:257f
NCX extracellular domain
structure–function relationship, 4:261
NCX/NCKX expression, 1:380f, 1:381–382, 1:382f, 3:45–46, 3:45f, 3:46f
phosphatidylinositol-4,5-bisphosphate (PIP₂), 2:746–747
phototransduction mechanisms, 3:489–490, 3:489f
physiological functions, 3:48
structural characteristics, 3:45f, 3:46–47, 4:256–261
topology model, 3:45f
transmembrane domain function, 4:257–258, 4:258f
transmembrane segments, 4:256, 4:258f
sodium ion (Na⁺) pumping, 4:90
Na⁺ translocation, 2:206–208, 2:208f
sodium phenylacetate, 4:495–497, 4:497t
sodium-potassium-chlorine (Na-K-Cl) cotransporter isoform 1 (NKCC1), 1:199, 1:200f
sodium–potassium (Na⁺–K⁺)-ATPase pump, 3:681–687, 3:688–692
cardiac glycosides, 3:686–687, 3:686f, 3:687f
characteristics and functional role, 3:681, 4:331–333
domain structures, 4:333f
FXYD proteins, 3:681, 3:682–685, 3:682t, 3:683f
physiological functions, 3:685–686
reaction mechanisms, 3:683, 3:682
structural characteristics
 α subunits, 3:683–684, 3:684f
 β subunits, 3:683f, 3:684–685
cation-pumping mechanism, 3:683–684, 3:684f
domain structures, 3:683–684, 3:683f, 4:333f
FXYD proteins, 3:682–685, 3:683f
schematic diagram, 3:684f
subunits and isoforms, 3:683f
subunits and isoforms, 3:681, 3:682t, 3:683f
voltage-gated sodium channels, 4:564–569
cardiac excitation–contraction (E–C) coupling, 1:379, 1:380f
evolutionary development, 4:564–565
functional role, 2:648–649, 4:564

- sodium (Na) (*Continued*)
 gene characterization, 4:564–565, 4:565f
 inherited disorders, 4:569, 4:569t
 intracellular signaling pathways, 2:712–713, 2:714f
 localizations, 4:566
 molecular bases, 4:566
 neurotoxin receptor sites, 4:568–569, 4:568t
 pharmacological functions, 4:568–569, 4:568f
 phylogenetic relationships, 4:565f
 schematic diagram, 2:649f
 structural characteristics, 4:566–568, 4:567f
 subunit structure and functions, 4:564, 4:565f
- soluble frizzled related peptide 1 (sFrp1), 1:193
- soluble N-ethylmaleimide-sensitive factor association protein receptor (SNARE), 3:37, 3:38f
- solute sodium symporter (SSS) family, 2:388
- somatostatin (sst) receptors, 4:262–266
 agonists/antagonists, 4:264–265
 biochemical/molecular characterization
 sst receptor structure, 4:263–264
 sst receptor subtypes, 4:263, 4:263t, 4:264f
 characteristics and functional role, 4:262
 desensitization mechanisms, 4:265–266
 isoforms, 4:262–263
 ligand binding, 4:262–263
 regulation mechanisms, 4:265–266
 research summary, 4:266
 signaling pathways, 4:265
 structural characteristics, 4:262–263, 4:263f
 therapeutic uses, 4:264–265
- somites, 1:179–180, 1:180f
- Sonic hedgehog (Shh) pathway
 cardiac muscle development, 1:183–184, 1:184f
 liver regeneration, 1:192
 neurogenesis, 1:197f, 1:199
- son of sevenless (Sos), 4:245, 4:246f
- Sorangicin, 4:171, 4:171f
- sorbitol
 hyperglycemia, 1:656
 phosphotransferase system (PTS), 2:387
- Sortilin, 3:244
- sorting nexin (SNX) family, 3:449
- Sos proteins, 4:310f
- SOS response
 LexA regulatory protein
 bacterial LexA regulon diversity, 2:710
 cell-cycle checkpoints, 2:710
 conformational changes, 2:708–709, 2:709f
 control mechanisms, 2:710
 DNA damage repair, 2:707f, 2:709–710
 functional role, 2:706
 inactivation induction, 2:708, 2:708f
 mutation effects, 2:710–711
 plasma-encoded genes, 2:710
 regulon characteristics, 2:706–707, 2:710
 repression mechanisms, 2:707–708, 2:707f
 response triggers, 2:708
 self-cleavage mechanisms, 2:708–709, 2:709f
 signaling pathways, 2:708, 2:708f
 transcription regulation, 4:477, 4:478f
 translesion synthesis (TLS), 4:41, 4:479f
- sour taste receptors, 4:367–368
- SoxAX, 2:511
- soybeans, 2:702t
- space of Disse, 1:190–191, 1:191f
- spartin, 4:269f, 4:270
- spasmin, 1:319t
- spastic paraplegias, 4:267–272
 axonal degeneration, 4:267
 definition, 4:267
 endosomal localization
 cytochrome P-450 (CYP), 4:271
 developmental genes, 4:272
 lipid metabolism, 4:270–271
 maspardin/SPG21, 4:269f, 4:270–272
 mitochondrial proteins
 Hsp60 protein/SPG13, 4:271–272
 paraplegin/SPG7, 4:271
 myelin–axonal interactions, 4:272
 neuropathy target esterase (NTE), 4:271
 seipin, 4:269f, 4:270–271
 genetic bases, 4:267, 4:268t
 neuropathology, 4:267
 pathogenesis
 cellular trafficking genes
 atlastin-1/SPG3A, 4:269, 4:269f
 KIF5A/SPG10, 4:269f, 4:270
 NIPA1/SPG6, 4:269f, 4:270
 receptor-enhancing protein 1 (REEP1)/SPG31, 4:269f, 4:270
 spartin/SPG20, 4:269f, 4:270
 spastin/SPG4, 4:268–269, 4:269f
 motor neurons, 4:267–270, 4:269f
 spastin, 1:3t, 1:4, 1:4f, 4:268–269, 4:269f
 spastizin, 4:269f, 4:270
 spatascin, 4:269f, 4:270
 speckles/splicing factor compartments, 3:298, 3:298f, 3:299f
- spectrin
 actin shape determination, 1:34
 actin–spectrin membrane network, 1:34–35, 1:34f
 EF-hand proteins, 1:319t
 plectin, 2:625–627
 spectrin/band 3 actin complex, 1:34–35, 1:34f
 spectrin/band 4.1 actin complex, 1:34–35, 1:34f, 1:26
- spectroscopic methods
 carbon-13 (¹³C) nuclear magnetic spectroscopy, 1:107, 1:108f
 near infrared (NIR) spectroscopy
 brain research
 brain functional activation, 4:279–280, 4:279f
 metabolic activity, 4:278, 4:278f
 neuronal signals, 4:280
 cancer detection, 4:279
 historical research, 4:277–278
 multidimensional imaging, 4:278
 muscle studies, 4:278f, 4:279
 photon migration, 4:279
 spectrophotometric historical research, 4:273–280
 biological fluorochromes, 4:277
 brain functional activation, 4:279–280, 4:279f
 brain research, 4:278
 cancer detection, 4:279
 dual-wavelength systems, 4:273–274, 4:278f, 4:279
 electron tunneling, 4:276–277
 galvanometer-driven wavelength scanning, 4:276
 hemoglobin (Hb) studies, 4:277
 low-temperature technologies, 4:276
 manual spectrometers, 4:273
 Michaelis–Menten/Briggs–Haldane equations, 4:273–274, 4:274f
 multidimensional imaging, 4:278
 muscle studies, 4:278f, 4:279
 near infrared (NIR) spectroscopy, 4:277–278, 4:278f
 neuronal signals, 4:280
 nicotinamide adenine dinucleotide (NAD⁺/NADH), 4:277
 photoactivation studies, 4:276
 photochemical action spectra, 4:274–275, 4:275f
 photon migration, 4:279
 research summary, 4:280
 respiration control, 4:276
 ruby laser technology, 4:276–277, 4:277f
 scattered light problem, 4:275–276, 4:275f
 time-resolved spectroscopy (TRS), 4:279
 tissue optical properties, 4:279
 tissue spectroscopy, 4:275
 time-resolved spectroscopy (TRS), 4:279
- sperm
 major sperm protein (MSP), 3:1–5
 accessory proteins
 MSP fiber protein 1 (MFP1), 3:3t, 3:4, 3:4f
 MSP fiber protein 2 (MFP2), 3:3t, 3:4, 3:4f
 MSP fiber protein 3 (MFP3), 3:3t, 3:4–5, 3:4f
 MSP-organizing protein (MPOP), 3:3–5, 3:3t, 3:4f
 MSP polymerization-activating kinase (MPAK), 3:3t, 3:4, 3:4f
 protein phosphatase 2A (PP2A), 3:3t, 3:4f, 3:5
 biochemistry, 3:2
in vitro motility system, 3:3, 3:3f
 molecular biology, 3:2
 nematode sperm motility, 3:1, 3:2f
 signaling functions, 3:1–2
 structural characteristics, 3:2–3, 3:2f
 mammalian glucose transport, 2:392t

- spermidine, 1:88, 1:87–88
spermine, 1:87–88, 1:88, 1:328
SPG2 gene, 4:272
sphinganine, 4:283f
sphingolipids, 4:281–286
 biosynthesis
 de novo biosynthesis, 4:283, 4:283f
 headgroup additions, 4:283f, 4:284–285
 lipid backbone, 4:282f, 4:283–284, 4:283f
 metabolic processes, 4:285–286
 catabolic reactions, 4:287–292
 ceramidase (CDase)
 acid ceramidase (aCDase), 4:288
 alkaline ceramidase, 4:289
 neutral ceramidase (nCDase), 4:288–289
 reaction mechanisms, 4:288–289
 sphingoid bases
 reaction mechanisms, 4:289
 sphingosine-1-phosphate lyase (SPLase), 4:289, 4:290f
 sphingomyelinase (SMase), 4:287–288
 degradation pathways, 4:290f
 future research areas, 4:286
 general characteristics, 4:281
 lipid rafts, 2:739–740, 2:741, 2:742f
 lysophosphatidic acid (LPA) receptors, 2:763t
 nomenclature, 4:281, 4:282f
 protein adducts, 4:283
 sphingolipidomics, 4:286
 sphingolipid receptors
 intracellular receptors, 2:762
 plasma membrane receptors, 2:762
 structural characteristics
 ceramides, 4:281–282, 4:282f, 4:283f
 chemical structure, 2:733, 2:734f
 glycosphingolipids, 4:282, 4:282f
 phosphosphingolipids, 4:282, 4:282f
 sphingoid bases, 4:281, 4:282f
 thiamine diphosphate (ThDP), 1:207f
 See also glycosphingolipids (GSLs)
sphingomyelin
 catabolic reactions, 4:287–288
 fatty acyl-coenzyme A (CoA) synthetases, 2:291f
sphingomyelinase (SMase)
 acid sphingomyelinase (aSMase), 4:287
 alkaline sphingomyelinase (bSMase), 4:288
 catabolic reactions, 4:287–288
 lipogenic gene expression, 2:287f
 neutral sphingomyelinase (nSMase), 4:287–288
 secreted sphingomyelinase (sSMase), 4:287
sphingomyelin synthases (SMSs), 4:284–285
sphingosin, 4:281, 4:282f
sphingosine-1-phosphate lyase (SPLase), 4:289, 4:290f
sphingosine-1-phosphate (SPP), 2:762, 3:542
sphingosine kinase, 4:283f
sphingosylphosphorylcholine (SPC), 2:762
spinach, 3:530, 3:531f, 4:533t
sphingosine-1-phosphate, 2:344t, 2:345f
Spirulina platensis, 1:501
spitz, 4:51f, 4:54
spleen
 acyl-coenzyme A synthetase
 long-chain acyl-CoA synthetase (LCAS), 2:292–293
 short-chain acyl-CoA synthetase (SCAS), 2:292
 amyloid deposits, 1:101t
 mammalian glucose transport, 2:392t
spliceosomes
 group I/group II introns, 4:152–153, 4:153f
 RNA splicing, 3:62–63, 3:63f, 4:187
Spo11 protein, 2:557, 2:559f, 2:561
SpoIVFB protein, 4:51f, 4:53
spondylo-epimetaphyseal dysplasia, 2:471t, 3:657t, 3:658
spondylometaphyseal dysplasia (SMD)
 Kozlowski type (SMDK), 4:416
Spt-Ada-Gen5- acetyltransferase (SAGA) complex, 1:526f
squalene, 1:517f
squaric acid, 3:230f
squidulin, 1:319t
Src (sarcoma) family kinases, 4:293–297
 angiotensin receptors, 1:113
 arrestins, 2:488t, 2:491
 cadherin signaling, 1:264, 1:264f
 calcium ion (Ca²⁺) signaling system, 2:719–720, 2:720f
 cell migration, 1:439, 1:439f, 1:441, 1:441f
 c-fes proto-oncogene (Fes), 1:451, 1:452f
 characteristics, 4:293
 chemokine receptors, 1:480–482, 1:482f
 focal adhesion kinase (FAK) family, 2:265–266, 2:267f, 2:611, 4:294–295
 focal adhesions, 2:321
 functional role, 4:294
 group III 5-phosphatases, 3:462
 guanine-nucleotide-exchange factors (GEFs), 4:245, 4:246f
 hematopoietin receptors, 2:526, 2:528f
 historical research, 4:293–294
 immune responses, 4:294–295
 invertebrate Src kinases, 4:297
 neurokinin-1 receptor (NK1R), 4:362–363
 phosphorylation, 4:295–297, 4:296f
 platelet-derived growth factor (PDGF) receptors
 chemotaxis, 3:542
 mediated cellular responses, 3:540, 3:541f
 signaling pathways, 3:539–540, 3:540f
 protein palmitoylation, 2:738, 4:296–297, 4:296f
 protein prenylation, 2:737t
 protein tyrosine kinase (PTK), 1:162, 1:163f, 3:470
 redundant functions, 4:294
 regulation mechanisms, 4:295–297, 4:296f
 research summary, 4:297
 SH2/SH3 domains, 1:451–452, 1:452f, 4:294, 4:295–297, 4:296f
 signaling adhesion functions, 4:294–295
steroid/thyroid hormone receptors, 4:310f
stable tubule-only polypeptide (STOP), 3:105
S-tag protein fusion system, 1:65
staphylococcal scalded-skin syndrome (SSSS), 1:638t
Staphylococcus, 3:303t
Staphylococcus aureus
 ATP-binding cassette (ABC) transport systems, 1:7
 Cho phylogenetic tree, 3:331f
 DNA polymerase III holoenzyme, 2:95
 LexA regulatory protein, 2:710–711
 type II restriction-modification enzymes, 2:132t
 uvrC phylogenetic tree, 3:331f
Staphylococcus epidermis, 3:331f, 4:253
starch, 3:498f, 3:499f
starch excess 4 (SEX4), 3:652
starfish, 1:334–335, 1:334f, 1:335f, 2:544t, 2:637–638, 2:638t
Startle disease, 2:424
starvation, 4:298–307
 anaplerotic pathways, 1:109, 4:305
 appetite regulation, 4:303
 basic concepts, 4:298
 biochemical changes
 amino acid metabolism, 4:304–305
 basic concepts, 4:303–306
 fatty acid release control, 4:303–304
 gluconeogenesis, 4:305–306, 4:306f
 ketogenesis, 4:305
 proteolysis, 4:304–305
 tricarboxylic acid cycle, 4:305
 blood substrate and hormone concentration changes, 4:301–302
 body composition, 4:300–301
 branched-chain α -ketoacid dehydrogenase complex (BCKDC), 1:247–248
 branched-chain amino acid (BCAA) regulation, 1:245–246
 clinical features, 4:299
 energy requirements, 4:300–301
 free fatty acids (FFAs), 4:301–302, 4:302f, 4:303–304
 fuel consumption, 4:307
 fuel oxidation characteristics, 4:301, 4:301f
 hormonal changes, 4:302–303
 human adaptations, 4:307
 metabolic alterations
 fed state, 4:299–300
 refeeding state, 4:300
 starved state, 4:300
 obesity–starvation paradox, 4:299
 pyruvate dehydrogenase kinase (PDK) regulation, 4:324–326
 resting metabolic rate–fat-free mass (RMR–FFM) relationship, 4:298–299, 4:299f, 4:300–301
 urinary nitrogen excretion, 4:306, 4:306f
 weight loss, 4:300–301
stathmin, 3:105
statins, 1:517f, 1:519
STAT transcription factors
 See signal transducer and activator of transcription (STAT)

- stearyl-CoA desaturase
 carbohydrate-responsive element binding protein (ChREBP) function, 1:376–377, 1:377f
 lipogenic gene expression, 2:287f, 2:288, 2:288f
- steatocystoma multiplex, 2:670t
- Ste gene complex, 3:441–442, 3:443f, 3:444f
- Steitz, Thomas A, 2:254
- stellate cells, 1:190–191, 1:191f
- stem cells
 cardiac muscles, 1:186
 cyclin-dependent kinases (CDKs), 1:408
 G₀ phase and G₁ phase, 1:408
 skeletal muscle, 1:186
 telomeres, 4:386
- stereochemistry, 2:221–225
 alcohol dehydrogenases (ADHs), 2:222, 2:223f
 enzymatic reactions, 2:221–222, 2:223f
 enzyme–substrate interactions, 2:223–224, 2:224f
 metal–nucleotide interactions, 2:223–224, 2:224f
 parallel reaction mechanisms, 2:224f
 sequential reaction mechanisms, 2:224f
 stereoselectivity, 2:221–222
 stereospecificity, 2:221–222
 steric course, 2:221–222, 2:222–223, 2:223f, 2:224f
 structural terminology, 2:221, 2:222f
- stereocilia, 1:35
- steroid hormones
 calbindin-D_{28k}, 1:286–287
 cholesterol biosynthetic pathway, 1:517f
 mucins, 3:206–208, 3:206f
- steroid hormone receptors, 4:308–313
 action mechanisms, 4:310f
 agonists/antagonists, 4:311–312
 characteristics and functional role, 4:308–309
 coregulators, 4:311
 ligand binding, 4:308–309
 molecular structure, 4:309f
 receptor–signaling pathway cross-talk
 cell signaling pathway stimulation, 4:312–313
 phosphorylation, 4:312
 posttranslational modifications, 4:312
 signal-regulated transcription factors, 4:312, 4:312f
- receptor superfamily
 DNA-binding domain, 4:309
 domain structures, 4:309–311, 4:311f
 hinge region, 4:310
 ligand-binding domain, 4:310–311
 N terminus, 4:309
 orphan nuclear receptors, 4:309–311
 receptor binding to DNA, 4:310–311
 structural characteristics, 4:309–311, 4:311f
- sterol methyl transferase 1 (SMT1), 1:251
- sterol regulatory element binding protein (SREBP)
 adipocyte development, 1:55f, 1:56
- cholesterol biosynthetic pathway, 1:517–519, 1:517f, 1:518f
 Golgi apparatus-based regulation, 2:481
 HUFA-response region with insulin response element (HUFA-RR/IRE), 2:287–288
 insulin metabolism, 2:606–607
 insulin resistance, 2:607
 lipogenic gene expression, 2:286–287, 2:287f
 mTOR complexes, 3:198
 regulated intramembrane proteolysis (RIP), 4:51–52, 4:51f
 unfolded protein response (UPR), 4:490–491
- stigmatellin (Stg), 1:143t, 1:596f, 1:597–598, 1:597f, 3:702f, 3:703–705, 3:704f
- Stokes shift, 1:461
- stomach
 aquaporin water channels, 1:124t
 energy homeostasis, 3:77f
 proteinase-activated receptors (PARs), 3:605t
- stomatal potassium outward rectifier channel (SKOR), 4:557–559, 4:558f
- stop codons, 2:250f
- store-operated calcium (SOC) channels, 2:597f
- strawberries, 4:533t
- Strep-tag protein fusion system, 1:64
- streptavidin, 1:64–65
- Streptococcus mutans*, 4:353
- Streptococcus pneumoniae*, 3:331f
- Streptococcus pyogenes*, 2:95, 3:331f
- Streptococcus thermophilus*, 4:253
- Streptolydigin, 4:171, 4:171f
- Streptomyces*, 1:140
- Streptomyces antibioticus*, 2:211
- Streptomyces coelicolor*, 4:173
- Streptomyces fimbriatus*, 2:132t
- Streptomyces griseoruber*, 2:132t
- Streptomyces rochei*, 4:444
- stress-activated protein kinase (SAPK), 3:177–178, 3:177f
- stress fibers, 1:430–431, 1:431f
- stress responses, 2:522–525
 heat shock response
 activation mechanisms, 2:522–523, 2:523f
 chaperone proteins, 2:522–523, 2:523f
 input conditions, 2:522
 output conditions
 aberrant protein degradation, 2:523
 chaperone genes, 2:523
 environmental stress response, 2:523–524
 protein repair and preservation, 2:523
 research background, 2:522–524
 stress detection, 2:522–523, 2:523f
 unfolded protein response (UPR)
 activation mechanisms, 2:524f
 endoplasmic reticulum (ER), 2:524–525, 2:524f
 ER-associated degradation (ERAD), 2:524f, 2:525
 output conditions, 2:525
- stress sensors, 2:524–525, 2:524f
 transmembrane signaling, 2:524–525, 2:524f
 unfolded proteins, 2:522, 2:523f
- striated muscles, 1:179–186
- actin organization, 1:31–32, 1:32f
- cardiac muscles
 development process
 cardiogenesis, 1:183, 1:183f, 1:184f
 microRNA (miRNA), 1:185
 transcription factors, 1:183–185, 1:184f
 research summary, 1:186
- stem cells, 1:186
- structural characteristics, 1:179, 1:180f
- skeletal muscle
 development process
 microRNA (miRNA), 1:182–183
 myogenesis, 1:179–180, 1:180f, 1:181f
 transcription factors, 1:180–182, 1:181f
 research summary, 1:186
- stem cells, 1:186
- structural characteristics, 1:179, 1:180f
- striate palmoplantar keratoderma (SPPK), 1:638t
- stromal-interacting molecule (STIM) protein family
 calcium oscillations, 1:323
 cytoplasmic domains, 3:362, 3:362f
 disease connections, 3:365
 domain structures, 3:361f, 4:315, 4:316f
 endoplasmic reticulum (ER) calcium sensors, 3:360–362, 3:361f, 4:316
 functional role, 3:360–362, 4:315, 4:415
 localizations, 4:315
 ORAI tetramers, 4:317–318, 4:318f
 physiological functions, 4:319
 posttranslational modification, 4:317
 research summary, 4:319
 signal inactivation, 4:318
 signal initiated oligomerization, 3:360–362, 3:361f, 4:316
 signal transmitted translocation, 4:317
 sterile α motif (SAM), 3:360–362, 3:361f, 4:315, 4:316f
- Stim1 protein, 1:295t, 1:295, 3:360–362, 3:361f
- STIM-mediated channel clustering, 4:317
- store-operated calcium (SOC) channels
 disease connections, 3:365
 signaling pathways, 3:364, 3:365f
 signal modulation, 3:364–365
- stromal interacting molecule 1 (STIM1)
 arachidonic acid regulated Ca²⁺ (ARC) channels, 1:129–130, 1:130f, 1:132f
 calcium oscillations, 1:323
 domain structures, 3:361f
 functional role, 2:587–588, 3:360–362
 store-operated calcium (SOC) channels, 3:360f, 3:360t
- stromal interacting molecule 2 (STIM2), 3:362

- transient receptor potential (TRP)
channels, 4:318–319
- stromal matrix, 1:437, 1:438f, 1:442–443, 1:443f
- stromelysins, 3:91t, 3:92f, 3:93
- strong hydrogen bonds
bond lengths, 2:757
deuterium fractionation factors, 2:758
serine proteases, 2:758–759
spectroscopic properties, 2:757–758
structural characteristics, 2:756–757, 2:757f
- Strongylocentrotus purpuratus*, 1:319t
- Strophanthus gratus*, 3:686–687
- strychnine, 2:421–423
- StyLTI enzyme, 2:137t
- subacute sclerosing panencephalitis, 1:477t
- subgranular zone (SGZ), 1:196
- submembrane plaque, 2:320–321, 2:320f
- substance P
calcitonin gene-related peptide (CGRP), 1:270–271
functional role, 4:360
G-protein-coupled receptors (GPCRs), 2:344t
peptide amidation, 3:396
research background, 4:364f
synthesis, 4:361f
- substantia nigra, 2:169, 2:172–173, 2:172t
- subventricular zone (SVZ), 1:196
- succinate
anaerobic respiration, 2:207f, 2:208f
anaplerotic pathways, 1:106f, 1:109f
citric acid cycle, 4:438f
fermentation reactions, 2:205, 2:206f
gluconeogenesis, 2:386f
mitochondrial ROS, 4:19f, 4:20
plant mitochondrial electron transport chain, 1:145, 1:146f
protonmotive electron transfer chain, 1:470–471, 1:470f
succinate dehydrogenase
electron transport chain
anoxygenic photosynthesis, 3:693–694, 3:694f
electron transfer complexes, 3:694f, 3:696
quinone pool functions, 3:695f
respiration, 3:661–662, 3:662f, 3:694–695, 3:694f, 4:79, 4:80f
tricarboxylic acid cycle, 4:439t
ubiquinone-reducing enzymes, 4:85
UCP1-induced proton cycling, 4:484f
- succinate:quinone oxidoreductases (SQORs), 4:92–97
characteristics and functional role, 4:92, 4:93f
classifications
functional classification, 4:92
structural classification, 4:92–93
electron-coupled proton transfer, 4:96–97, 4:97f
electron transfer, 4:96
facultative phototrophic respiration, 4:79, 4:80f
structure and functions
domain structures, 4:93, 4:94f
flavin adenine dinucleotide (FAD/FADH₂), 4:95, 4:95f
iron–sulfur clusters (FeS), 4:95
quinol oxidation/quinone reduction, 4:95–96
relative orientations, 4:93–94
subunit A, 4:95, 4:95f
subunit B, 4:95
subunits C and D, 4:95–96
succinate oxidation/fumarate reduction, 4:95, 4:95f
thiamine diphosphate (ThDP), 1:207f
tricarboxylic acid cycle, 4:439f
See also respiratory chain complex II (succinate coenzyme Q reductase)
- succinic semialdehyde dehydrogenase (SSADH), 2:722, 2:723f
- succinic semialdehyde (SSA), 2:722, 2:723f
- succinyl-coenzyme A (CoA)
 δ -aminolevulinic acid (ALA) formation, 2:539–540, 2:540f, 3:544
anaerobic respiration, 2:208f
anaplerotic pathways, 1:106, 1:106f, 1:109f
branched-chain amino acids (BCAAs), 1:244, 1:245f
citric acid cycle, 4:93f
gluconeogenesis, 2:381, 2:386f
tricarboxylic acid cycle, 4:439f, 4:439t
- sucrose, 3:498f
- sucrose non-fermenting-1-related protein kinase 1 (SnRK1), 3:498f, 3:501
- sucrose permease (CscB), 2:388
- suf* operon, 2:662–663
- sugar nucleotide transporters, 4:350–351
genetic mutations, 4:350
macromolecular glycosylation regulation, 4:351
structural characteristics, 4:350–351
topographic characteristics, 4:351
transport mechanisms, 4:350, 4:351f
- sugar transport
bacteria, 2:387–390
ATP-binding cassette (ABC) transport systems
classification, 2:389
components, 2:389–390
functional role, 2:390
maltodextrin/maltose transport (MalEFGK₂), 2:390, 2:390f
structural characteristics, 2:389–390, 2:390f
background information, 2:387
electrochemical cation-gradient-driven symporters
classification, 2:388
functional role, 2:388
lactose permease (LacY), 2:388–389
melibiose permease (MelB), 2:389
sodium/galactose transport (SgIT), 2:389
structural characteristics, 2:388, 2:389f
fermentation reactions, 2:204–206
phosphotransferase system (PTS)
classification, 2:387
components, 2:387, 2:388f
functional role, 2:387, 2:388f
preferential glucose utilization, 2:387–388
- sulfenic acid, 3:410, 3:410f
- sulfhydryl, 1:147–148, 1:147f
- sulfinic acid, 3:410–411, 3:410f
- Sulfolobus*, 3:34
- Sulfolobus acidocaldarius*, 4:92–93
- Sulfolobus shibatae*, 2:165f
- Sulfolobus solfataricus*, 2:35f, 2:38, 3:577f, 3:671–672, 4:239, 4:240f
- sulfur (S)
bacterial reductive reactions, 1:487t
disulfide bond formation, 1:664–667
basic concepts, 1:664, 1:665f
catalytic reactions
background information, 1:665
protein disulfide isomerase (PDI), 1:665, 1:665f
thioredoxin family, 1:665
cellular environments
extracellular proteins, 1:666, 1:667f
regulatory mechanisms, 1:666
thiol/disulfide redox environments, 1:666
isomerization processes, 1:664, 1:665f
molecular structure, 1:664, 1:665f
oxidative protein folding
basic principles, 1:664
formation mechanisms, 1:664, 1:666f
general discussion, 1:664
protein folding and assembly processes, 3:628
- iron–sulfur (FeS) proteins, 2:657–664
bacterial biogenesis
hydrogenases, 2:663
iron–molybdenum cofactor (FeMo-co), 2:659f, 2:663
ISC (iron–sulfur cluster) system
assembly machinery, 2:662, 2:662f
nitrogenases, 2:663
nitrogen fixation (NIF) system, 2:663
sulfur mobilization machinery, 2:662–663
- biogenesis
bacteria, 2:662–663, 2:662f
basic concepts, 2:662
eukaryotes, 2:663–664, 2:663f
catalytic iron–sulfur clusters (FeS)
analytical analyses, 2:661–662
catalytic reaction mechanisms, 2:660–661, 2:661f
disulfide bond reduction, 2:660–661, 2:661f
free-radical reactions, 2:660, 2:661f
regulatory functions, 2:661
sensing functions, 2:661
eukaryotic biogenesis
chloroplasts, 2:663
cytosolic system, 2:663–664
mitochondrial Fe–S cluster assembly machinery, 2:663, 2:663f
functional role
analytical analyses, 2:661–662

- sulfur (S) (*Continued*)
- catalytic iron-sulfur clusters (FeS), 2:660–661, 2:661f
 - electron transfer, 2:660
 - free-radical reactions, 2:660
 - protein structure stabilization, 2:660
 - reducing powers, 2:660
 - regulatory functions, 2:661
 - sensing functions, 2:661
 - iron-sulfur clusters (FeS)
 - characteristics and functional role, 2:335–336, 2:336f
 - electronic structure, 2:657–658
 - F_x and F_b clusters, 2:338–339
 - F_x- and F_b-containing proteins, 2:339–340, 2:339f
 - ferredoxins (Fd), 2:297–298, 2:301f, 2:661f, 3:485, 3:505, 3:505f
 - frataxin functions, 2:333
 - F_x cluster, 2:338, 2:338f
 - magnetic characteristics, 2:657–658
 - nomenclature, 2:657
 - respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:342–343, 4:342f, 4:345f
 - respiratory chain complex III, 1:596, 1:596f, 3:663–664, 3:664f
 - stability, 2:657
 - structural characteristics, 2:657, 2:658f
 - succinate:quinone oxidoreductases (SQORs), 4:95
 - type I reaction centers, 2:335–342
 - nomenclature, 2:658
 - respiratory chain complexes, 4:83
 - Rieske protein
 - cytochrome *b₆*f complex, 2:335, 4:335–338, 4:337f, 4:338t, 4:339t
 - cytochrome *bc₁* complex, 1:596, 1:596f, 3:30–31, 3:663–664, 3:664f
 - cytochrome *f*, 3:32
 - posttranslational modifications, 4:338
 - protonmotive Q cycle, 3:663–664, 3:664f
 - structural characteristics
 - flexibility, 2:659–660
 - higher-order structures, 2:659
 - metallic superclusters, 2:659, 2:659f
 - primary structures, 2:658
 - secondary structures, 2:658–659
 - tertiary structures, 2:658–659
 - purple bacteria electron acceptors and donors, 3:698, 3:698t
 - sulfate respiration, 2:206, 2:207f
 - sulfite oxidase, 3:116
- sulindac, 3:305t
- sumoylation, 4:62
- SUN domain proteins, 3:314, 3:315f
- Sup35 protein
 - characteristics, 3:566
 - disease states, 3:568–569
 - generation mechanisms, 3:569
 - prion domains, 3:568f
 - self-propagating amyloidoses, 3:567–568, 3:568f
- superoxide anion (O₂)
- biochemical free radical chain reactions, 3:168–169, 3:169f
 - mitochondrial production, 3:169–170, 3:170f, 3:171f, 3:173f
 - physiological functions, 3:172–173
 - superoxide dismutases (SODs), 4:352–354
 - amyotrophic lateral sclerosis (ALS), 4:354
 - cellular respiration, 3:407, 3:408f
 - deletion consequences
 - Escherichia coli*, 4:353
 - mice, 4:354
 - Saccharomyces cerevisiae*, 4:353
 - enzyme families
 - action mechanisms, 4:352
 - copper SODs, 4:352–353
 - iron SODs, 4:353
 - manganese SODs, 4:353
 - zinc SODs, 4:352–353
 - functional role, 4:352
 - historical research, 3:168–169
 - mitochondrial DNA (mtDNA), 1:212f
 - mitochondrial production, 3:171f, 3:173f
 - nuclear factor kappa B (NF-κB), 3:304t
 - pharmaceutical mimics, 4:354
 - photosystem I (PSI) complex, 3:485
 - prion protein (PrP), 3:574
 - regulation mechanisms, 2:77
 - suppressor of cytokine signaling-3 (SOCS-3), 2:260, 2:260f, 2:526, 2:528f
 - supra-molecular activation clusters (SMACs), 4:376
 - surfactant proteins (SPs), 2:703–704, 2:704t
 - survival of motor neuron (SMN) gems, 3:300–301, 3:300f
 - Su(var)3–9, Enhancer-of-zeste, Trithorax (SET) domain, 4:60–61
 - swainsonine (SW), 2:459f
 - sweat glands, 1:124t, 2:669t
 - sweet taste receptors, 4:366–367
 - Swi1p protein, 3:566
 - swiholide A, 1:16, 1:16t
 - [SWI+] prion
 - See* prions
 - Swiss Institute of Bioinformatics (SIB), 3:611–612
 - SWItching defective/Sucrose NonFermenting (SWI/SNF) family, 1:314, 4:60f, 4:63–64, 4:403–404
 - Syc family kinases, 2:611
 - Syk family kinases, 1:162, 1:163f
 - sympathetic nervous system, 2:258–259
 - sympleskin, 3:189–191, 3:190t
 - sympport, 3:51
 - synaptotagmin I, 1:278, 1:279f
 - syncoilin, 2:632t, 2:634
 - syndecans, 1:439–440, 2:320, 2:320f, 3:91t
 - syndrome X, 1:654
 - Synechococcus*, 3:32
 - Synechococcus elongatus*, 3:531
 - Synechocystis*
 - δ-aminolevulinic acid (ALA) formation, 3:544
 - Cho phylogenetic tree, 3:331f
 - copper-transporting ATPases (copper pumps), 3:671–672
 - LexA regulatory protein, 2:710
 - oxygenic photosynthesis, 3:32
 - respiration, 4:81, 4:82f
 - succinate:quinone oxidoreductases (SQORs), 4:92–93
 - uvrC* phylogenetic tree, 3:331f
- Synechocystis*, 1:506
- synemin, 2:632t, 2:634
- synthesis dependent strand annealing (SDSA), 4:29
- synucleins
 - G protein-coupled receptor kinases (GRKs), 2:485t
 - α-synuclein, 1:101t, 1:103, 3:623t
 - ubiquitin-proteasome degradation pathway, 3:623t
- syphilis, 3:113, 3:114t
- systematic evolution of ligands by exponential enrichment (SELEX), 4:136, 4:144–145
- systemic lupus erythematosus (SLE), 3:113

T

- T7 RNA polymerase, 4:355–359
- applications, 4:359
 - polymerase sequence similarities, 4:355–357, 4:356f, 4:357f
 - structural characteristics, 4:355, 4:356f
 - structure–function relationship, 4:355, 4:357f
 - transcription reaction/conformational change
 - DNA replication priming, 4:359
 - initial transcription reaction, 4:356f, 4:357–358
 - pausing or terminator sites, 4:356f, 4:358, 4:358f
 - process mechanisms, 4:356f, 4:359
 - promoter release and elongation, 4:356f, 4:358
 - promoter structure, recognition, and opening, 4:357, 4:358f
 - regulation mechanisms, 4:356f, 4:358–359
- tachykinins, 4:360–365
- basic concepts, 4:360
 - G_q family proteins, 2:497t
 - hemokinin 1, 4:361, 4:361f
 - neurokinin-1 receptor (NK1R)
 - carboxyl tail, 4:363–364, 4:363f
 - disease connections
 - affective disorders, 4:364–365
 - depression, 4:364–365
 - emesis, 4:364
 - glioblastomas, 4:365
 - research background, 4:364–365, 4:364f
 - full-length NK1R, 4:361–363, 4:362f
 - functional role, 4:361, 4:362f
 - truncated NK1R, 4:363
 - neurokinin A, 4:360–361, 4:361f, 4:364f
 - neurokinin B, 4:361, 4:361f, 4:364f
 - substance P, 4:360, 4:361f, 4:364f
- tadalafil, 1:569
- talin, 2:609f, 2:610–611, 3:438–439

- tamsulosin, 1:58
- tandem (MS/MS) mass spectrometry, 3:12, 3:12f, 3:14f
- tangle-only dementia, 1:477t
- target-primed (TP) retrotransposons
- characteristics, 3:284
 - integration site preference, 3:288–289
 - long interspersed elements (LINEs), 3:284, 3:285f, 3:287–288
 - organellar retrotransposons, 3:288
 - Penelope-like elements (PLEs), 3:284, 3:285f, 3:288
 - prokaryotic retrotransposons, 3:288
 - R2-type elements, 3:287, 3:287f
 - short interspersed elements (SINEs), 3:285f, 3:287–288
- target recognition domain (TRD)
- DNA methyltransferases (MTases), 3:8–9, 3:8f
 - type I DNA restriction and modification enzymes, 2:72–73, 2:73f
 - type III DNA restriction and modification enzymes, 2:137–138
- taste receptors, 4:366–368
- bitter taste receptors, 4:367
 - G-protein-coupled receptors (GPCRs), 4:366–367
 - ion channels, 4:367–368, 4:367f
 - salty taste receptors, 4:367–368
 - sour taste receptors, 4:367–368
 - sweet taste receptors, 4:366–367
 - T1R1 receptors, 4:366–367, 4:367f
 - T2R receptors, 4:367, 4:367f
 - taste buds, 4:366, 4:367f
 - taste receptor cells (TRCs), 4:366, 4:367f
 - transduction cascades, 4:368
 - umami taste receptors, 4:366–367
- Tata, 4:51f, 4:54
- TATA-binding protein (TBP)
- GAL1–10 gene expression, 1:525, 1:526f
 - gene transcription regulation, 4:234
 - RNA polymerase II (Pol II), 4:163
- tau proteins
- basic concepts, 1:477–478
 - bundle formation, 3:106f
 - characteristics, 3:104t, 3:105, 3:105f
 - glycogen synthase kinase-3 (GSK3), 2:436t, 2:437
 - isoforms, 1:476f
 - neurodegenerative diseases, 3:106–107
 - neurogenerative disorders, 1:477t
 - tau mutations, 1:475–477
- Tauri disease, 2:431t, 2:432
- taxol, 4:450–451
- Tay–Sachs disease, 4:291t, 4:507–508, 4:507t
- T-box transcription factor family, 1:184, 1:184f
- T-brain gene-2 (Tbr-2), 1:196–197, 1:197f
- T cells
- anti-mitochondrial antibodies (AMAs), 3:115–116
 - calcineurin-mediated signal transduction, 1:312, 1:339–340, 1:340f
 - calcium ion (Ca^{2+}) release-activated Ca^{2+} channels (CRAC), 1:127–128
 - calcium oscillations, 1:323–324
 - chemokine receptors, 1:484
 - diacylglycerol kinases (DGKs), 1:661–662
 - exosomes and microvesicles (EMVs), 2:264
 - $\text{G}_{12}/\text{G}_{13}$ family functions, 2:345t
 - hematopoiesis, 1:187–188, 1:188f
 - immunoglobulin Fc receptors (FcRs), 2:569–571
 - interferon- γ receptors, 2:621
 - lectins, 2:705
 - liver regeneration, 1:190–191, 1:191f, 1:194
 - nicotinic acid adenine dinucleotide phosphate (NAADP)-induced calcium ion (Ca^{2+}) release, 2:638, 2:638t
 - nuclear factor of activated T cells (NFAT)
 - calcineurin-mediated signal transduction, 1:312, 1:338–339, 1:339–340, 1:340f, 4:222f
 - calcium ion (Ca^{2+}) signaling, 1:312, 1:313f, 2:587–588, 2:719–720, 2:720f, 3:293, 3:293f, 3:294
 - calcium oscillations, 1:323–324
 - glycogen synthase kinase-3 (GSK3), 2:435f, 2:436t
 - serine–threonine phosphatases, 4:223 - T-cell receptors (TCRs), 4:374
 - primary biliary cirrhosis (PBC), 3:562
 - protein tyrosine phosphatase (PTP), 3:649–651, 3:651t
 - serine–threonine phosphatases, 4:222f, 4:223
 - signaling receptors, 2:743
 - stromal-interacting molecule (STIM)
 - protein family, 4:319 - T-cell factor/lymphoid enhancer factor (TCF/LEF) family, 1:263–264, 1:263f
 - T-cell receptors (TCRs), 4:369–375
 - characteristics, 4:369
 - complementarity-determining regions (CDRs), 4:369–371, 4:370f, 4:372f
 - domain structures, 4:371, 4:372f
 - functional role
 - cell biology, 4:373–374
 - T-cell development, 4:373 - genetic organization and rearrangement, 4:369–371, 4:370f, 4:371t
 - inositol triphosphate (IP_3)/calcium ion (Ca^{2+}) signaling system, 2:587–588
 - integrin signaling, 2:611
 - invariant chains
 - CD3/CD247 complexes, 4:372f, 4:373
 - pre-T α complex, 4:372–373
 - TCR:CD3 complex, 4:373 - major histocompatibility complex (MHC), 4:369–371
 - nucleotide excision repair (NER), 3:344
 - peptide–MHC complex recognition, 4:371, 4:372f
 - protein tyrosine phosphatase (PTP), 3:649–651, 3:651t
 - sequence diversity, 4:371t
 - signaling to nuclear factor kappa B (NF- κ B), 4:376–379
 - antigen receptors, 4:376, 4:377f
 - CARD11 (caspase-recruitment domain containing protein 11), 4:377–378, 4:377f - chromosomal translocations, 4:379
 - downstream activators, 4:377–378
 - I κ B kinase (IKK) complex activation, 4:378–379
 - immunological synapses, 4:376
 - mucosa-associated lymphoid tissue 1 (MALT1), 4:378–379
 - NF- κ B activation, 4:379
 - pathway dysregulation, 4:379
 - receptor–proximal signaling, 4:376, 4:377f
 - second messengers, 4:377
 - TGF activated kinase 1 (TAK1), 4:378–379
 - TNF receptor associated-factor 6 (TRAF6), 4:378–379
 - zeta-associated protein-70 (ZAP-70), 4:374
- TCR:CD3 complex
- assembly and surface expression, 4:373
 - characteristics, 4:373
 - internalization/downmodulation, 4:373
 - signal transduction, 4:373–374, 4:374f
 - TCR chains, 4:369–371
 - V(D)J recombination, 3:272, 3:272f
- tumor necrosis factor receptors (TNFRs), 4:458
- V(D)J recombination, 3:272, 3:272f
- vitamin D, 4:538–539
- Wnt/ β -catenin signaling pathway, 1:263–264, 1:263f, 2:434–435, 2:436f
- Tctex family, 2:185–186, 2:185t, 2:186f, 2:187
- technetium-based bacterial reductive reactions, 1:487t
- telomeres, 4:380–386
- characteristics and functional role, 4:380
 - chromosome organization, 1:538
 - non-homologous end joining (NHEJ), 3:272, 4:384–385
 - replication processes
 - general discussion, 4:384
 - negative regulation, 4:384, 4:385f
 - positive regulation, 4:384 - research background, 4:380–384
 - telomerase-independent maintenance, 4:384
 - telomerase repression
 - disease implications, 4:385
 - stem cells, 4:386
 - tumor suppression, 4:385–386 - telomere capping, 4:384–385
 - telomeric complex
 - telomerase, 4:381–382, 4:382f
 - telomere proteins, 4:382–384, 4:383f
 - telomeric DNA, 4:380–381, 4:381f
 - telomeric RNA, 4:384
- telophase I, 3:19f, 3:20
- TEL transcription factor, 1:187, 1:188f
- temozolomide, 2:51–52
- temsirolimus, 3:198–199
- 10T1/2 cell lineage, 1:52–54, 1:53f

- tenascin (TN)
 integrins, 1:427*t*, 1:438–440
 matrix metalloproteinases (MMPs), 3:91*t*
Tenebrio molitor, 1:384
- ten-eleven translocation (TET) proteins, 3:9
- teniposide, 2:167*f*
- tenofovir diphosphate, 2:550, 2:550*f*
- tensin
 barbed-end-capping proteins, 1:25
 phosphatase and tensin homolog (PTEN), 1:565–566, 3:195–196, 3:448*f*, 3:450
 submembrane plaque, 2:320–321, 2:320*f*
- terbutaline, 1:59
- testis
 acyl-coenzyme A synthetase
 Bubblegum ACS family (ACSBG), 2:294
 long-chain acyl-CoA synthetase (LCAS), 2:292–293
 medium-chain acyl-CoA synthetase (MCAS), 2:292
 short-chain acyl-CoA synthetase (SCAS), 2:292
 aquaporin water channels, 1:124*t*
 cyclic nucleotide-gated (CNG) ion channels, 1:586–587
 mammalian glucose transport, 2:392*t*
 proteinase-activated receptors (PARs), 3:605*t*
- testosterone
 characteristics and functional role, 4:308–309
 chemical structure, 1:251*f*, 4:309*f*
 leptin regulation, 2:259
- tetrachlorethene, 2:206, 2:207*f*
- 5-tetradecyloxy-2-furoic acid (TOFA), 2:279
- tetrahydrobiopterin
 biosynthesis, 3:666–667
 deficiency-related diseases, 3:669
 nitric oxide synthase (NOS), 1:342, 1:343*f*, 3:668
 occurrences, 3:666
 structural characteristics, 3:668*f*
- tetrahydroisoquinoline (THIQ), 3:27
- 1-methyl-4-phenyl 1,2,3,6-tetrahydropyridine (MPTP), 3:185–186, 3:187*f*, 4:90, 4:91
- Tetrahymena*
 guanylyl cyclases (GCs), 1:567
 ribozyme distribution, 4:146*f*
- Tetrahymena pyriformis*, 1:319*t*, 4:353
- Tetrahymena thermophila*, 4:143*f*, 4:188, 4:380–381
- tetrakis-(2-pyridylmethyl)ethylenediamine (TPEN), 3:359–360, 3:360*f*
- tetramethyldioxetane, 1:462, 1:462*f*
- 5, 6, 11, 12-tetraphenylnaphthacene, 1:463*f*
- tetrapyrroles, 2:541
 biosynthetic deficiencies
 herbicide effects, 3:547
 porphyrias, 2:542, 2:542*t*, 3:547, 3:548*f*
 biosynthetic pathway
 δ-aminolevulinic acid (ALA) formation, 2:540–541, 2:540*f*, 3:544, 3:545*f*
 chlorophyll biosynthesis, 1:501, 3:545*f*, 3:546, 3:546*f*
 coenzyme B₁₂, 3:546–547, 3:546*f*
 factor F₄₃₀, 3:546–547, 3:546*f*
 general discussion, 3:544–547
 heme biosynthesis, 2:539–540, 2:540*f*, 3:545*f*, 3:546, 3:546*f*
 heme d₁, 3:546–547, 3:546*f*
 schematic diagram, 2:540*f*, 3:545*f*
 siroheme, 3:546–547, 3:546*f*
 uroporphyrinogen III formation, 2:540–541, 2:540*f*, 3:544–546, 3:545*f*, 3:546*f*
 catabolic reactions, 3:547–549, 3:548*f*
 characteristics, 3:544
 regulation mechanisms
 basic concepts, 3:547
 heme biosynthesis
 δ-aminolevulinic acid (ALA) function, 2:539–540, 3:547
 non-ALAS enzymatic regulation, 3:547
- tetraspanins, 2:262, 2:263*f*
- tetratricopeptide (TPR), 4:310*f*
- tetrodotoxin (TTX), 4:564–565, 4:568*t*
- TGF activated kinase 1 (TAK1), 4:377–378
- thapsigargin (TG), 2:229
- thenoyltrifluoroacetone (TTFA), 1:143*t*
- Theorell–Chance mechanisms, 2:218–220, 2:219*f*, 2:220*f*
- Thermoanaerobacter tengcongensis*, 4:72
- Thermochromatium tepidum*, 3:702–703
- thermophiles, 2:87
- Thermoplasma*, 3:34, 3:596–597
- Thermoplasma acidophilum*, 3:591*f*, 3:595–596, 3:596*f*, 4:92
- thermoplasma-quinone (TK), 4:94*f*
- Thermosynechococcus elongatus*, 3:504*f*, 3:505*f*
- Thermotoga*, 4:146*f*
- thermotogales, 3:559
- Thermotoga maritima*
 Cho phylogenetic tree, 3:331*f*
 copper-transporting ATPases (copper pumps), 3:671–672
 DNA alkylation sites and repair mechanisms, 2:10–12
 [FeFe]-hydrogenases, 4:72
 transfer RNA (tRNA)
 maturation processes, 3:555*f*
 3'-end–RNase process pathway, 3:555*f*, 3:556
uvrC phylogenetic tree, 3:331*f*
- Thermus aquaticus*, 2:55, 2:87, 4:167, 4:168*f*, 4:174–176, 4:174*f*
- Thermus filiformis*, 2:29*f*, 2:35*f*, 2:38
- Thermus thermophilus*
 alkyltransferase-like (ATL) proteins, 2:14
 clustered regularly interspaced short palindromic repeat (CRISPR) RNAs (crRNAs), 4:253
 fatty acyl-coenzyme A (CoA) synthetases, 2:294–295
 heme–copper oxidases, 1:604*t*
 respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:87–91
 basic concepts, 4:87, 4:341
 disease connections
 general discussion, 4:90–91
 mitochondrial mutations, 4:90–91
 neurotoxins, 4:91
 nuclear mutations, 4:91
 pathological mechanisms, 4:91
 evolutionary development
 general discussion, 4:88–90
 N-module, 4:89
 origins, 4:89–90
 P-module, 4:89
 Q-module, 4:88–89
 variants, 4:89–90
 functional role
 active–deactive transition, 4:90
 electron transfer mechanisms, 4:90
 general discussion, 4:90
 inhibitors, 4:90
 proton pumping, 3:661–662, 3:662*f*, 4:90
 redox centers, 4:90
 sodium ion (Na⁺) pumping, 4:90
 structural characteristics
 accessory subunits, 4:88, 4:89*t*
 central subunits, 4:87–88, 4:88*f*, 4:89*t*
 cysteine motif, 4:344–345, 4:345*f*
 flavin mononucleotide (FMN)
 binding, 4:344–345, 4:344*f*
 frataxin-like subunit, 4:343–344, 4:343*f*
 hydrophilic domain, 4:342–343, 4:342*f*
 iron–sulfur clusters, 4:87–88, 4:89*t*, 4:342–343, 4:342*f*, 4:345*f*
 multistage assembly pathway, 4:88
 mutations, 4:345
 NADH binding, 4:344–345, 4:344*f*, 4:345*f*
 overall structure, 4:341–342
 redox centers, 4:87–88, 4:89*t*, 4:342–343, 4:342*f*
 research summary, 4:345–346
 shape, 4:87, 4:88*f*
 subunits, 4:342*t*
 supernumerary subunits, 4:345
 ribosomal structure, 4:129–130
 RNA polymerases, 4:167, 4:174–176, 4:174*f*, 4:175*t*
 structural characteristics, 4:100
- thiamine, 1:202–209
 adenosine thiamine triphosphate (AThTP)
 derivative determinations, 1:204, 1:204*f*
 metabolic pathways, 1:205*f*, 1:206
 molecular structure, 1:203*f*
 non-cofactor roles, 1:207–208
 background information, 1:202
 biosynthesis, 1:203–204
 chemical properties, 1:202–203, 1:204*f*
 derivative determinations, 1:204, 1:204*f*
 metabolic pathways
 adenosine thiamine triphosphate (AThTP), 1:205*f*, 1:206
 general discussion, 1:206
 thiamine diphosphate (ThDP), 1:205, 1:205*f*
 thiamine triphosphate (ThTP), 1:205–206, 1:205*f*
 non-cofactor roles, 1:207–208
 related diseases

- beriberi, 1:202, 1:208
 nutritional deficiencies, 1:208–209
 transport defects
 Amish lethal microcephaly, 1:209
 Leigh disease (subacute necrotizing encephalomyelopathy), 1:209
 maple syrup urine disease (MSUD), 1:209
 thiamine-responsive megaloblastic anemia (TRMA), 1:209
 Wernicke–Korsakoff syndrome, 1:202, 1:208–209
 research summary, 1:209
 riboswitches, 1:203–204
 thiamine-binding proteins, 1:206
 thiamine diphosphate (ThDP)
 background information, 1:202
 catalytic reaction mechanisms, 1:206–207, 1:208f
 cofactor role, 1:206, 1:207f
 derivative determinations, 1:204f
 hydrolysis mechanisms, 1:206
 α -lactyl-ThDP, 1:206–207, 1:208f
 metabolic pathways, 1:205, 1:205f
 molecular structure, 1:203f
 prevalence, 1:204
 pyruvate dehydrogenases (PDHs), 4:321–322, 4:325f
 thiamine monophosphate (ThMP)
 biosynthetic pathway, 1:203–204
 derivative determinations, 1:204f
 hydrolysis mechanisms, 1:206
 metabolic pathways, 1:205f
 molecular structure, 1:203f
 riboswitches, 4:138–139
 thiamine pyrophosphate (TPP), 4:137, 4:138–139, 4:140f, 4:141f
 thiamine triphosphate (ThTP)
 derivative determinations, 1:204, 1:204f
 hydrolysis mechanisms, 1:206
 metabolic pathways, 1:205–206, 1:205f
 molecular structure, 1:203f
 non-cofactor roles, 1:207–208
 transport mechanisms, 1:205
 thiazolidinedione (TZD)
 adipocyte development, 3:421–423, 3:422f
 adiponectin, 3:74
 diabetes treatments, 1:657, 3:418–419, 3:423
 thiazolium dipolar ions, 1:206–207, 1:208f
 thidiazuron, 1:627, 1:628f
ThiM gene, 4:139, 4:140f, 4:141f
Thiobacillus, 3:29–30
Thiobacillus denitrificans, 1:490t
Thiobacillus ferrooxidans, 1:491t, 3:31–32
Thiobacillus thiooxidans, 1:490t
Thiobacillus versutus, 1:490t
 thiochrome diphosphate (ThcDP), 1:204f
 thiochrome monophosphate (ThcMP), 1:204f
 thiochrome triphosphate (ThcTP), 1:204f
 thioesters, 1:539, 2:290
 thiogalactosidase transacetylase (LacA), 2:77
 thioglycosides, 3:229
 thiol-oxidoreductases, 4:205–206
 thioredoxin family
 catalytic reactions, 1:665
 chloroplast signaling, 1:514–515
 homologs, 2:403
 hydrogen peroxide (H₂O₂) regulation, 3:411–412
 peroxiredoxins, 3:407, 3:410, 3:411f
 thioredoxin reductase, 3:411–412, 3:411f
 thioredoxin-type ferredoxins, 1:514–515, 2:297, 2:661f
 thoracic aortic aneurysm, 3:223t
 3-hydroxyanthranilate 3,4-dioxygenase (3HAO), 1:175f
 3-methyladenine (3mA), 2:9, 2:10t
 3-methylcytosine (3mC), 2:9
 three-base ribozymes, 4:144
 three-dimensional fluorescent *in situ* hybridization (3D-FISH) analysis
 general discussion, 2:749
 nuclear structure and dynamics, 2:749–751
 threonine
 classifications, 1:92t
 mitochondrial DNA (mtDNA), 3:134t
 N-terminal active sites, 3:596, 3:596f
 O-linked glycosylation, 2:465, 4:207
 phage libraries, 3:434t
 phosphorylation, 4:61
 serine–threonine kinase (Akt)
 arrestins, 2:488t
 calcium/calmodulin-dependent protein kinase II (CaMKII), 1:297–303
 autonomous activity, 1:299
 autophosphorylation, 1:299, 1:300f
 calmodulin (CaM) trapping, 1:300
 cardiac regulation, 1:302
 characteristics and functional role, 1:297
 domain structures, 1:298–299, 1:299f
 intracellular localization, 1:297–298, 1:298f
 long-term potentiation, 1:302
 memory and learning regulation, 1:302
 neuronal activity monitoring, 1:301, 1:301f
 structure and functions, 1:302–303
 substrate anchoring, 1:300
 substrates, 1:297–298, 1:298f
 threonine (Thr³⁰⁵/Thr³⁰⁶)
 autophosphorylation, 1:300–301
 chemoattractant signaling, 2:236–238, 2:237f
 chemokine receptors, 1:480–482, 1:482f
 fatty acid metabolism–cancer relationship, 2:277
 focal adhesions, 2:321
 functional role, 3:463–464
 glycogen synthase kinase-3 (GSK3), 2:435f
 G protein-coupled receptor kinases (GRKs), 2:485t
 hematopoietin receptors, 2:528f
 insulin signal transduction, 2:604–605, 2:604f
 liver regeneration, 1:193–195
 mitochondrial permeability transition pore, 3:165–166
 MSP polymerization-activating kinase (MPAK), 3:3t, 3:4, 3:4f
 mTOR complexes, 1:408, 3:195–196
 neurogenesis, 1:199, 1:200f
 phosphatidylinositol-3,4,5-trisphosphate (PtdIns-3,4,5-P₃), 3:450
 plectin, 2:625–627
 protein kinase C (PKC), 3:637–638
 somatostatin (sst) receptors, 4:265
 serine–threonine phosphatases, 4:219–223
 biochemical characterization, 4:219, 4:220t
 catalytic mechanisms, 4:220, 4:220f
 DARPP-32, 4:222f, 4:223
 disease connections, 4:223
 nuclear factor of activated T cells (NFAT), 4:222f, 4:223
 phospho-protein phosphatase (PPP)
 biochemical characterization, 4:219, 4:220t
 catalytic mechanisms, 4:220, 4:220f
 characteristics, 4:219
 PP1 (calcium/calmodulin-dependent type 1 serine/threonine protein phosphatase), 1:312, 4:220, 4:221f, 4:222
 protein phosphatase 2A (PP2A), 4:221, 4:221f, 4:222–223
 protein phosphatase 2B (PP2B), 3:3t, 3:4f, 3:5, 4:221, 4:221f, 4:223
 protein phosphatase 5 (PP5), 4:221, 4:221f
 regulation mechanisms, 4:222
 regulatory subunit interactions, 4:220, 4:221f
 T-cell activation, 4:222f, 4:223
 three-dimensional structure, 4:219–220, 4:220f, 4:221f
 protein phosphatase Mg²⁺- or Mn²⁺-dependent (PPM)
 biochemical characterization, 4:219, 4:220t
 catalytic mechanisms, 4:220
 cellular functions, 4:221
 characteristics, 4:219
 regulation mechanisms, 4:221
 three-dimensional structure, 4:219–220
 subunit families
 phospho-protein phosphatase (PPP), 4:219
 protein phosphatase Mg²⁺- or Mn²⁺-dependent (PPM), 4:219
 three-dimensional structure, 4:219–220, 4:220f
 small guanosine phosphatases (GTPases), 4:247f
 threonine (Thr³⁰⁵/Thr³⁰⁶)
 autophosphorylation, 1:300–301
 vasopressin/oxytocin family, 4:516t
 See also Akt (serine–threonine active kinase protein)
 threonyl-tRNA synthetase, 4:443

- thrombin
 blood-clotting proteases, 3:585, 3:586f, 3:587, 3:588f
 G-protein-coupled receptors (GPCRs), 2:344t, 2:345f
 G_q family proteins, 2:496–497, 2:497t
 proteinase-activated receptors (PARs), 3:601, 3:602f, 3:603t
 three-dimensional (3D) cell migration and invasion, 1:443–444
- thrombophilia, 3:657t
- thrombopoietin (Tpo)
 expression regulation, 2:528
 signaling pathways, 2:528–529
 structural characteristics, 2:526–527, 2:527f
- thrombospondin (Tsp)
 cell–matrix interactions, 1:426
 integrins, 1:427t
 thrombospondin type 1 repeats (TSRs), 2:454, 2:455f, 2:456
- thromboxanes
 biosynthetic relationships, 2:189, 3:582f
 G-protein-coupled receptors (GPCRs), 2:344t, 2:345f
 G_q family proteins, 2:496–497, 2:497t
 platelet-activating factor receptor (PAFR), 3:536, 3:536f
- thylakoid membranes
 photosynthesis, 3:492, 3:493f, 3:494f
 photosystem II (PSII) complex, 3:508–511
 assembly processes
 regulation mechanisms, 3:509
 research background, 3:508
 sequence, 3:508–509
 characteristics, 3:512
 chloroplast–nucleus cooperation
 characteristics, 3:508
 chloroplast-encoded PSII proteins, 3:508
 nuclear-encoded PSII proteins, 3:508
 D1 protein turnover processes
 photoinactivation–damage
 mechanisms, 3:509, 3:510f
 photoinhibition repair cycle, 3:509–511, 3:510f
 light harvesting proteins, 2:178–179, 2:179f, 2:181f, 2:182
 photoinhibition repair cycle
 D1 protein degradation, 3:509–510, 3:510f
 D1 protein synthesis, 3:510, 3:510f
 initiation phase, 3:509, 3:510f
 post-translational assembly and reactivation, 3:510–511, 3:510f
 protonmotive electron transfer chain, 1:470f
- thylakoid protein kinase network, 1:514, 1:514f
- thymine
 phage libraries, 3:432–433, 3:433t, 3:434t
 synthetic pathways, 3:713f
 thymine nucleotide formation, 3:715–716, 3:715f
- thymidylate synthase (TS), 2:325, 2:325f
- thymine
 chemical structure, 4:77f
 degradation pathways, 3:717–718, 3:717f
 DNA alkylation sites and repair mechanisms, 2:10t
 synthetic pathways, 3:713f
 thymine DNA glycosylase (TDG), 2:19–20
 background and characteristics, 4:78
 biochemistry, 4:78
 functional role, 2:16
 ultraviolet (UV) photoproducts, 4:479–480, 4:480f
 UvrABCD endonuclease, 3:330t
- thymine glycol, 2:1, 2:2f, 2:3t, 2:78f, 2:79, 4:405
- thymosin, 1:15–16
- thymus, 3:605t
- thyroid, 3:605t
- thyroid hormone receptors, 4:308–313
 action mechanisms, 4:310f
 agonists/antagonists, 4:311–312
 characteristics and functional role, 4:308–309
 coregulators, 4:311
 ligand binding, 4:308–309
 molecular structure, 4:309f
 receptor–signaling pathway cross-talk
 cell signaling pathway stimulation, 4:312–313
 phosphorylation, 4:312
 posttranslational modifications, 4:312
 signal-regulated transcription factors, 4:312, 4:312f
 receptor superfamily
 orphan nuclear receptors, 4:309–311
 structural characteristics
 DNA-binding domain, 4:309
 domain structures, 4:309–311, 4:311f
 hinge region, 4:310
 ligand-binding domain, 4:310–311
 N terminus, 4:309
 receptor binding to DNA, 4:310–311
 thyroid response element (TRE), 4:311f
- thyroiditis, 3:562
- thyroid peroxidase (TPO), 3:407–409, 3:408f
- thyroid-stimulating hormone receptor (TSHR)
 disease connections
 autoantibodies, 4:390
 naturally occurring mutations, 4:390
 extracellular domains, 4:388
 functional role, 4:387
 genomic organization, 4:388
 G protein activation, 4:388–390
 structural organization, 4:389f
- thyrotropin
 ascorbic acid, 4:531–532
 G-protein-coupled receptors (GPCRs), 2:344t
 thyrotropin-releasing hormone (TRH)
 body weight regulation, 3:78, 3:78f
 G_q family proteins, 2:497t
 peptide amidation, 3:396
- tiazofurin, 1:177
- tight junctions, 4:392–395
 characteristics, 4:392
- functional roles
 barrier functions, 4:393–394
 claudin knockout mice, 4:394
 fence functions, 4:394
- structural characteristics
 electron microscope images, 4:392, 4:393f
 integral membrane proteins, 4:392, 4:394f
 peripheral membrane proteins, 4:392–393
 three-dimensional images, 4:392, 4:393f
- Tim complex, 3:128–130, 3:128t, 3:130f, 3:159
- time-of-flight (ToF) mass spectrometry, 3:11–12
- time-resolved spectroscopy (TRS), 4:279
- timolol, 1:59
- tissues
 acyl-coenzyme A synthetase
 Bubblegum ACS family (ACSBG), 2:294
 long-chain acyl-CoA synthetase (LCAS), 2:292–293
 medium-chain acyl-CoA synthetase (MCAS), 2:292
 short-chain acyl-CoA synthetase (SCAS), 2:292
 very long-chain acyl-CoA synthetase (VLCAS), 2:293–294
 tissue-bound amine oxidase, 1:89
 tissue transglutaminase, 3:91t
- titin, 1:31–32, 1:32f
- TJ-associated marvel domain-containing proteins (TAMPs), 4:392, 4:393f, 4:394f
- tmRNA system, 3:58
- Tn3 resolvase, 1:6–7
- Tn7 transposon, 3:276f, 3:277
- Tn syndrome, 2:471t
- toads, 4:516t
- Tobacco budworm, 2:593
- tocopherols, 4:545–550
 antioxidant activities, 2:77, 4:545–546, 4:546f
 coenzyme Q, 3:725–727
 dietary sources, 4:547–548
 health and disease considerations
 cancer, 4:547
 central nervous system (CNS), 4:547
 coronary heart disease (CHD), 4:547
 fertility, 4:547
 medical paradox, 4:547
 protective behaviors, 4:547
 RRR- α -tocopherol, 4:547
 hydrophobicity, 4:549
 metabolic characteristics, 4:549, 4:549f
 recommended dietary allowance (RDA), 4:547–548
 structural characteristics, 4:545, 4:546f
 transport mechanisms
 absorption, 4:548
 circulation and delivery processes, 4:548–549
 hepatic processing, 4:548
- tocotrienols, 4:545–550
 metabolic characteristics, 4:549
 structural characteristics, 4:545, 4:546f

- togavirus, 3:41
- toll-like receptors (TLRs), 4:396–401
 future research areas, 4:400
 infectious diseases, 4:399
 innate immune system, 4:396
 pathogen-associated molecular patterns (PAMPs), 4:396–397, 4:397*t*
 research background, 4:396
 signaling pathways
 characteristics, 4:398–399
 MyD88-dependent signaling pathway, 4:398, 4:399*f*, 4:400*f*
 TIR-containing adaptor inducing IFN β (TRIF)-dependent signaling pathway, 4:398–399, 4:399*f*, 4:400*f*
 structural characteristics, 4:397–398
 TIR-containing adaptor protein (TIRAP), 4:398–399, 4:399*f*, 4:400*f*
 Toll/interleukin-1 receptor (TIR) domain, 4:397–398
 TRAF protein family, 4:455–457
 TRIF-related adaptor molecule (TRAM), 4:398–399, 4:399*f*, 4:400*f*
 vaccine adjuvants, 4:399–400
- tomatoes, 4:533*t*, 4:547–548
- Tom complex, 3:127–128, 3:128*t*, 3:130*f*, 3:159
- Top6A/Top6B, 2:165, 2:165*f*, 3:369–370
- topoisomerases
 DNA replication, 2:120
 DNA supercoiling, 2:156, 2:161, 2:162*f*, 2:163, 3:368–369, 3:369*f*
 mitochondrial topoisomerase, 2:158*t*
 nucleoid-associated proteins (NAPs), 3:369–370
 poxviral topoisomerase, 2:158*t*
 type I DNA topoisomerase, 2:157–162
 catalytic reactions, 2:157
 cellular roles
 hyperthermophilic eubacteria, 2:161–162
 mesophilic/psychrophilic eubacteria, 2:161
 replication processes, 2:162, 2:162*f*
 supercoils, 2:161, 2:162*f*
 transcription processes, 2:162, 2:162*f*
 type II DNA topoisomerase, 2:161–162
 classifications, 2:158, 2:158*t*
 DNA supercoiling, 3:369*f*
Escherichia coli, 3:369*t*
 functional role, 2:157, 2:163–164
 nomenclature, 2:158, 2:158*t*
 properties, 2:158
 type IA DNA topoisomerase
 classifications, 2:158
 cleavage/strand-passage domain
 crystal structure, 2:158–159, 2:159*f*
 protein domains, 2:158, 2:159*f*
 reverse gyrase mechanism, 2:159–160
 sources, 2:158*t*
 strand passage enzyme bridging mechanism, 2:159, 2:159*f*
 type IB DNA topoisomerase
 catalytic reactions, 2:160–161
 classifications, 2:158
 crystal structure, 2:160, 2:161*f*
 domain structure, 2:160, 2:160*f*
 sequence conservation, 2:160, 2:160*f*
 sources, 2:158*t*
 strand passage rotational mechanism, 2:161
- type II DNA topoisomerase, 2:163–168
 archaeal type IIB DNA topoisomerase, 2:166
 cellular roles, 2:161–162
 characteristics, 2:164–165
 domain structures, 2:164–165, 2:165*f*
 double-stranded DNA passage reaction, 2:164, 2:164*f*
 enzyme mechanism, 2:164, 2:164*f*
Escherichia coli, 3:369*t*
 functional role, 2:1
 Top6A/Top6B, 2:165, 2:165*f*, 3:369–370
- type IIA DNA topoisomerase
 antibacterial drugs, 2:167, 2:167*f*
 anticancer drugs, 2:167, 2:167*f*
 dietary topoisomerase II poisons, 2:167–168, 2:167*f*
 DNA gyrase, 2:165–166
 domain structures, 2:164–165, 2:165*f*
 eukaryotic type IIA DNA topoisomerase, 2:166
 prokaryotic type IIA DNA topoisomerase, 2:165–166
 as therapeutic targets, 2:167–168
 type IV DNA topoisomerase, 2:166
- type IIB DNA topoisomerase, 2:165, 2:165*f*, 2:166
- type III DNA topoisomerase
Escherichia coli, 3:369*t*
 protein domains, 2:159*f*
 sources, 2:158*t*
 type IIIA DNA topoisomerase, 4:45, 4:48*f*
- type IV DNA topoisomerase, 2:164–165, 2:165*f*, 2:166, 3:369*t*
- type V DNA topoisomerase, 2:158*t*
- type VI DNA topoisomerase, 2:166
- TOPORS, 4:468*t*, 4:469
- Tourette's syndrome, 2:173
- toxic lectins, 2:703*f*, 2:705
- Toxoplasma gondii*
 actin-related proteins (Arps), 1:37*f*
 pyrimidine *de novo* synthesis, 3:714
- trachea
 intracellular calcium ion (Ca²⁺) waves, 2:643
 proteinase-activated receptors (PARs), 3:605*t*
 tracheal epithelial cells, 2:643
- TRAF protein family, 4:455–457, 4:455*f*, 4:456*f*
- transaldolase, 3:393*f*
- transcriptional repression domain (TRD), 1:522–523, 1:523*f*
- transcription attenuation, 2:360–364
 research background, 2:360
trp operon, 2:360
 control mechanisms, 2:360–362, 2:361*f*
 RNA-binding attenuation protein (TRAP)
- anti-TRAP (AT), 2:363*f*, 2:364
- bacterial occurrences, 2:364
- characteristics, 2:363, 2:363*f*
- general discussion, 2:362–364
- hairpin formation, 2:363, 2:363*f*
- leader transcript, 2:362, 2:362*f*
- timing factors, 2:363
- tRNA^{Trp}, 2:363*f*, 2:364
- transcription regulation, 2:360–362
- transcription-coupled repair (TCR) pathway, 4:402–406
- basic concepts, 4:402–403
- damage recognition mechanisms, 4:402
- lesion-related transcription arrest
 cisplatin-induced intrastrand cross-links, 4:405
 cyclobutane pyrimidine dimers (CPDs), 4:404–405, 4:404*t*
 general discussion, 4:404–405
 oxidative damage, 4:405
- nucleotide excision repair (NER)
 basic concepts, 3:337, 3:342*f*, 3:342*t*, 3:344, 4:402–403
 damage recognition mechanisms, 4:402
- lesion-related transcription arrest
 cisplatin-induced intrastrand cross-links, 4:405
 cyclobutane pyrimidine dimers (CPDs), 4:404–405, 4:404*t*
 general discussion, 4:404–405
 oxidative damage, 4:405
- NER factors, 3:342*t*
- repair pathways, 3:342*f*
- research perspective, 4:405–406
- RNA polymerase II (Pol II)-based arrest and initiation, 4:403–404, 4:404*t*
- schematic diagram, 4:403*f*
- UvrABCD system, 3:335–336, 3:335*f*
- research perspective, 4:405–406
- RNA polymerase II (Pol II)-based arrest and initiation, 4:403–404, 4:404*t*
- schematic diagram, 4:403*f*
- transcription-repair coupling factor (TRCF), 3:335–336
- transcription factors
 basic helix-loop-helix (bHLH) domain, 1:184, 1:184*f*, 1:314
 calcium channel-associated transcriptional regulator (CCAT), 1:314
 calcium sensors, 1:311
 cardiac muscle development, 1:183–185, 1:184*f*
 CREB (cyclic adenosine monophosphate (cAMP) responsive binding protein), 1:313
 downstream regulatory element antagonist modulator (DREAM), 1:311, 1:313–314, 1:314*f*
 E2F transcription factor, 4:106–107
 Forkhead (FKHR) transcription factor family, 3:450–451
 GATA-1 zinc finger transcription factor, 1:187, 1:188*f*
 GATA-2 zinc finger transcription factor, 1:187, 1:188*f*

transcription factors (*Continued*)

- GATA-3 zinc finger transcription factor, 1:187–188, 1:188f
- GATA-4 zinc finger transcription factor, 1:184
- GATA-6 zinc finger transcription factor, 1:184f
- G-box-binding factor (GBF), 1:564–565, 1:564f
- Golgi apparatus-based regulation, 2:481
- hairly-related transcription factor (Hrt) family, 1:185
- Isl1 (Isl1) transcription factor, 1:184f, 1:185
- mitochondrial transcription factor (Mtf1p), 3:141, 3:142f
- nuclear calcium
- basic helix-loop-helix (bHLH) domain, 1:314
 - calcium channel-associated transcriptional regulator (CCAT), 1:314
- CREB (cyclic adenosine monophosphate (cAMP) responsive binding protein), 1:313
- downstream regulatory element antagonist modulator (DREAM), 1:311, 1:313–314, 1:314f
- myocyte enhancer factor 2 (MEF2), 1:312, 1:313f
- nuclear factor kappa B (NF-κB), 1:313
- nuclear factor of activated T cells (NFAT), 1:312, 1:313f
- nuclear factor kappa B (NF-κB), 3:302–305, 4:376–379
- background and characteristics, 3:302
 - biological functions, 3:303–304, 3:304t
 - disease connections, 3:304–305, 3:304t
 - hematopoiesis, 1:188–189
 - IκB kinase (IKK), 1:313, 3:303, 3:304f, 3:305, 4:456, 4:458f
 - inhibitors, 3:305t
 - mucins, 3:206–208, 3:206f, 3:208f
 - neurokinin-1 receptor (NK1R), 4:362–363
 - nuclear calcium, 1:313
 - pharmacological regulations, 3:305, 3:305t
 - platelet-activating factor receptor (PAFR), 3:535, 3:535f
 - protein characteristics and structure, 3:302, 3:303f
 - RAGE (receptor for advanced glycation end products), 2:407–409
 - RANK/RANKL receptor proteins, 4:459
 - regulatory mechanisms, 3:302–303, 3:303t, 3:304f
 - signal transduction, 3:303, 3:304f
 - target genes, 3:303–304, 3:304t
 - T-cell receptor (TCR) signaling, 4:376–379
 - antigen receptors, 4:376, 4:377f
 - CARD11 (caspase-recruitment domain containing protein 11), 4:377–378, 4:377f
 - chromosomal translocations, 4:379
 - downstream activators, 4:377–378
 - IκB kinase (IKK) complex activation, 4:378–379
 - immunological synapses, 4:376
 - mucosa-associated lymphoid tissue 1 (MALT1), 4:378–379
 - NF-κB activation, 4:379
 - pathway dysregulation, 4:379
 - receptor–proximal signaling, 4:376, 4:377f
 - second messengers, 4:377
 - TGF activated kinase 1 (TAK1), 4:378–379
 - TNF receptor associated-factor 6 (TRAF6), 4:378–379
 - zeta-associated protein-70 (ZAP-70), 4:374
- nuclear factor of activated T cells (NFAT), 1:338–339
- peroxisome proliferator-activated receptors (PPARs), 3:418–424
- RNA polymerase II (Pol II), 4:163
- signal transducer and activator of transcription (STAT)
- hematopoietin receptors, 2:526, 2:528f
 - interferon receptors, 2:620–621, 2:620f, 2:622f
 - leptin resistance, 2:260–261, 2:260f
 - myocardial ischemia, 3:111
 - neurogenesis, 1:198
 - signaling pathways, 1:564–565, 1:564f, 1:616, 1:617f, 1:618f, 3:539–540, 3:540f
- skeletal muscle development, 1:180–182, 1:181f
- steroid/thyroid hormone receptors, 4:312, 4:312f
- SWItching defective/Sucrose NonFermenting (SWI/SNF) family, 1:314, 4:403–404
- T-box transcription factor family, 1:184, 1:184f
- transcription factor IIB (TFIIB)
- 1α,25-dihydroxyvitamin D₃ (1,25-(OH)₂D₃), 4:538f
 - functional role, 4:163
 - gene transcription regulation, 4:234
 - vitamin D receptor (VDR), 4:543
- transcription factor IID (TFIID)
- 1α,25-dihydroxyvitamin D₃ (1,25-(OH)₂D₃), 4:538f
 - functional role, 4:163
 - 3'-end maturation, 3:189
- transcription factor IIE (TFIIE), 4:163
- transcription factor IIF (TFIIF), 4:163
- transcription factor IIH (TFIIH)
- functional role, 4:163
 - nucleotide excision repair (NER), 3:337, 3:338f, 3:342–343, 3:342f, 3:342t, 4:404
- transcription factor IIS (TFIIS), 4:404–405
- transcription factor IIIA (TFIIIA)
- classical zinc fingers, 4:160, 4:575, 4:576f
 - RNA polymerase III (Pol III), 4:160, 4:160f
- transcription factor IIIB (TFIIIB), 4:160, 4:160f
- transcription factor IIIC (TFIIIC), 4:160, 4:160f
- vitamin D receptor (VDR), 4:544
- transducin, 2:379, 2:379t
- transferrin (Tf)
- endocytosis, 2:197, 2:199
 - glycosylation, 2:473
- transfer RNA (tRNA), 3:554–560
- aminoacyl-transfer RNA (aa-tRNA) peptidyl transferase center (PTC), 4:131–132, 4:133f
 - protein biosynthesis elongation, 4:421–423
 - basic concepts, 4:128–129
 - elongation phase, 4:122–123, 4:123f, 4:124f, 4:131, 4:421–423
 - schematic diagram, 4:129f
 - ribosomal subunits, 4:130f, 4:131
 - structural characteristics, 4:134f
 - tRNA synthetases, 4:441
- bacterial selenoprotein synthesis, 4:210, 4:211f
- bacterial translation elongation, 4:419–420, 4:420f, 4:421f
- characteristics, 3:554
- defective processing consequences, 3:559
- elongation cycle
- aminoacyl-tRNA, 4:122–123, 4:123f, 4:124f, 4:128–129, 4:129f, 4:131
 - background information, 4:122–123
 - EF-Tu/EF-G–ribosome interactions, 4:123f, 4:125–126, 4:126f, 4:128–129, 4:129f, 4:421f
 - guanosine triphosphate (GTP) hydrolysis, 4:123–125, 4:124f
 - initial selection phase, 4:123, 4:124f, 4:420–421, 4:421f
 - research summary, 4:126
 - ribosome recycling, 4:126
 - substrate proofreading, 4:125
- 5'-end–RNase P process pathway, 3:142, 3:142f, 3:555, 3:555f
- glutamyl-transfer RNA (tRNA) synthetase, 2:539–540, 2:540f
- maturation processes, 3:554, 3:555f
- mitochondrial genomes, 3:306–307
- quality control mechanisms, 3:559–560
- retroviral genomes, 2:102, 2:102f
- reverse transcriptase (RT) activity, 2:104–105, 2:104f
- ribosomes, 4:116, 4:122–123, 4:128, 4:129f
- RNA trafficking, 3:322
- selenocysteyl-tRNA, 4:212
- structural characteristics, 3:554–555
- 3'-end–RNase process pathway
- Bacillus subtilis*, 3:555f, 3:556
 - Escherichia coli*, 3:555–556, 3:555f
 - Mycoplasma genitalium*, 3:555f, 3:556
 - Saccharomyces cerevisiae*, 3:142, 3:142f
 - Thermotoga maritima*, 3:555f, 3:556
- translation initiation, 4:425–431
- basic concepts, 4:425
 - 40S ribosomal subunits, 4:432–433

- initiation complex formation, 4:426–427
- initiation factors
- initiation factor 1 (IF1), 4:427–428, 4:427f, 4:428f
 - initiation factor 2 (IF2), 4:427f, 4:428–430
 - initiation factor 3 (IF3), 4:427f, 4:428f, 4:429f, 4:430
 - structure and functions, 4:427–431
- ribosomal dissociation, 4:432
- 70S initiation complex, 4:130f, 4:426–427, 4:427f
- 60S ribosomal subunits, 4:432
- 30S initiation complex, 4:130f, 4:132f, 4:426–427, 4:427f, 4:429f
- translation initiation region (TIR), 4:425–426, 4:426f
- tRNA synthetases, 4:441–444
- alternative pathways
 - genetic code expansion, 4:442–443, 4:443f
 - transamidation reactions, 4:442
 - aminoacyl-transfer RNA (aa-tRNA), 4:441
 - attachment mechanisms, 4:442
 - classifications, 4:442
 - disease correlations, 4:443
 - enzymatic reactions, 4:442, 4:441
 - functional role, 4:441
 - glutamyl-tRNA synthetase, 4:441
 - glutamyl-prolyl-tRNA synthetase, 4:443
 - glutamyl-tRNA synthetase, 4:441
 - inhibitors as drugs, 4:444
 - leucyl-tRNA synthetase, 4:443
 - lysyl-tRNA synthetase, 4:441, 4:443
 - novel functions, 4:443
 - structural characteristics, 4:442, 4:443f
 - threonyl-tRNA synthetase, 4:443
 - translational accuracy
 - amino acid selectivity, 4:441
 - codon:anticodon interaction, 4:441–442
 - editing activities, 4:442, 4:442
 - structural characteristics, 4:441, 4:442f
 - tryptophanyl-tRNA synthetase, 4:443
 - tyrosyl-tRNA synthetase, 4:443
- tRNA^{Trp}, 2:363f, 2:364
- transformer (tra)* gene, 1:81–82, 1:82f, 1:85–86, 1:85f
- transformer (tra)* protein, 1:85
- transforming growth factor alpha (TGF- α), 4:570
- transforming growth factor beta (TGF- β)
- annexins, 1:276–277
 - arrestins, 2:488t
 - bone morphogenetic proteins (BMPs), 1:229, 1:230f
 - cardiac muscle development, 1:183–184, 1:184f
 - cytokines, 1:620, 1:620f
 - liver regeneration, 1:193–195
 - mucins, 3:207–208
 - nuclear factor kappa B (NF- κ B), 3:303t
 - proteolysis, 1:444
 - signaling pathways, 1:229, 1:230f, 1:620, 1:620f
 - skeletal muscle development, 1:182
 - transforming growth factor- β -activated kinase (TAK)1, 4:398, 4:399f, 4:400f
- transgenic mice, 2:348–349, 2:724
- transient hyperammonemia, 4:500
- transient receptor potential (TRP) channels, 4:412–417
- arachidonic acid regulated Ca²⁺ (ARC) channels, 1:128–129, 1:130–131
 - channel activation mechanisms, 4:413–414
 - characteristics, 4:412, 4:413f
 - disease connections, 4:416
 - functional role
 - cellular functions, 4:414
 - fluid secretions, 4:414
 - sensory functions, 4:414
 - TRPC channels
 - basic concepts, 4:414–416
 - brain-derived neurotrophic factor (BDNF), 4:415
 - exocrine gland cells, 4:415–416
 - neurite outgrowth, 4:415
 - smooth muscle cells, 4:415
 - store-operated calcium (SOC) channels, 4:415
 - vascular function, 4:415
 - glutamate receptors, 2:715–716
 - G_q family proteins, 2:496–497, 2:497t
 - research summary, 4:416
 - store-operated calcium (SOC) channels, 1:323, 4:315, 4:318–319
 - structural characteristics, 2:650, 4:412–413, 4:413f
 - transient receptor potential vanilloid-subtype 5 (TRPV5), 1:285
 - transport modulation, 2:746–747
 - TRPC channel subfamily, 3:359–360, 3:360f, 3:360t, 4:412–413, 4:413f, 4:414–416
 - TRPM channel subfamily, 4:413, 4:413f
 - TRPV channel subfamily, 4:413, 4:413f
- transketolase
- carbohydrate-responsive element binding protein (ChREBP), 1:377f
 - pentose phosphate pathway, 3:393f, 3:394
 - photosynthetic carbon dioxide (CO₂) fixation, 3:499f, 3:500
 - thiamine diphosphate (ThDP), 1:206, 1:207f
- translation initiation, 4:425–431, 4:432–435
- basic concepts, 4:425
 - eukaryotic messenger RNA (mRNA)
 - AUG triplets, 4:433–435
 - 80S initiation complex, 4:434f
 - 80S ribosome formation, 4:435
 - eukaryotic initiation factors (eIFs), 4:433t, 4:434f, 4:435
 - 5' leader scanning, 4:433
 - 40S ribosomal subunits, 4:432
 - 43S preinitiation complexes, 4:433, 4:434f
 - 48S initiation complex, 4:435
 - initiation codon selection, 4:433–435
 - initiation mechanisms, 4:432
 - internal ribosomal entry sites (IRESs), 4:435
 - regulation mechanisms, 4:435
 - ribosomal dissociation, 4:432
 - 60S ribosomal subunits, 4:432, 4:435
 - structural characteristics, 4:432
 - transfer RNA (tRNA), 4:432–433
- initiation complex formation, 4:426–427
- initiation factors
- initiation factor 1 (IF1)
 - initiation pathway, 4:427f
 - structure, 4:427, 4:427f, 4:428f
 - topographical localization and function, 4:428, 4:429f
 - initiation factor 2 (IF2)
 - initiation pathway, 4:427f
 - structure, 4:427f, 4:428–430
 - topographical localization and function, 4:430
 - initiation factor 3 (IF3)
 - initiation pathway, 4:427f
 - structure, 4:428f, 4:430
 - topographical localization and function, 4:429f, 4:430–431
 - structure and functions, 4:427–431
- 70S initiation complex, 4:130f, 4:426–427, 4:427f
- 30S initiation complex, 4:130f, 4:132f, 4:426–427, 4:427f, 4:429f
- translation initiation factor 2 (eIF2 α)
- phosphorylation, 4:491
- translation initiation region (TIR), 4:425–426, 4:426f
- translesion synthesis (TLS)
- DNA polymerases, 2:247, 3:579f, 4:572–574, 4:573f
 - DNA-strand transferases, 4:41
 - Y-family polymerases, 4:479, 4:479f
- translocases, 4:30–31, 4:30f, 4:31t, 4:32
- transmembrane domains
- ATP-binding cassette (ABC) transport systems, 1:7, 1:8f
 - copper-transporting ATPases (copper pumps), 3:672–673, 3:672f, 3:673–674
 - integrin activation, 2:609–610
 - mucins, 3:201f, 3:202
 - ORAI protein family, 3:362–363, 3:363f
 - oxytocin (OXY), 4:517–519, 4:518f
 - pheromone receptors, 3:444–445
 - ryanodine receptors (RyRs), 4:195
 - sodium ion/calcium ion (Na⁺/Ca²⁺) exchanger, 4:257–258, 4:258f
 - vascular endothelial growth factor (VEGF), 4:510, 4:510f
 - vasopressin, 4:517–519, 4:518f
- transmembrane proteins, 3:207t
- transmissible spongiform encephalopathies (TSEs), 3:566, 3:568–569
- transposable bacteriophage (Mu), 3:276f, 3:277
- transposable elements (TEs), 3:275–282
- classifications
 - compound transposons, 3:276, 3:276f
 - conjugative transposons, 3:276f, 3:277

- transposable elements (TEs) (*Continued*)
 insertion sequences (ISs), 3:275–276, 3:276f
 miniature inverted repeat transposable elements, 3:276, 3:276f
 Tn7 transposon, 3:276f, 3:277
 transposable bacteriophage (Mu), 3:276f, 3:277
 unit transposons, 3:276–277, 3:276f
 definitions, 3:275–277
 eukaryotic transposons, 3:281–282
 functional role, 3:275
 regulation mechanisms, 3:280–281
 S transposases (S-Tpases), 3:275–277, 3:279, 3:280f
 target choices, 3:279
 transposases (Tpases), 3:275–277
 transposition mechanisms
 DDE transposases, 3:277–278, 3:278f
 general discussion, 3:277–279
 non-DDE transposases
 S transposases (S-Tpases), 3:279, 3:280f
 Y1 transposases (Y1 Tpases), 3:279
 Y2 transposases (Y2 Tpases), 3:279
 Y transposases (Y-Tpases), 3:275–277, 3:279, 3:280f
- transposons
 compound transposons, 3:276, 3:276f
 conjugative transposons, 3:276f, 3:277
 site-specific recombination systems, 1:1
 unit transposons, 3:276–277, 3:276f
See also retrotransposons
- transthyretin, 1:101t
- treadmilling, 1:27–28, 1:28f, 1:18–19
- trehalose 6-phosphate, 2:543–544, 2:544t
- Treponema pallidum*, 3:113, 3:331f
- triacylglycerides (TAGs), 1:47, 1:49–51, 1:50f
- triacylglycerol (TAG)
 fatty acyl-coenzyme A (CoA) synthetases, 2:291f
 functional role, 2:414–415
 peroxisome proliferator-activated receptors (PPARs), 3:418
 phosphatidic acid synthesis, 2:416f
 structural characteristics, 2:412–414, 2:414f
 uncharged glycerolipids, 2:416–417, 2:416f
- tricarboxylic acid (TCA)
 carbohydrate-responsive element binding protein (ChREBP), 1:377f
 cardiac metabolism, 3:65–66, 3:66f, 3:68f, 3:70f
 cerebral metabolism, 1:369f, 1:371–373
 tricarboxylic acid cycle, 4:436–440
 basic concepts, 4:438–440, 4:439f
 catalyzing enzymes, 4:438, 4:439t
 mitochondrial ROS, 1:210, 1:212f
 PET genes, 3:307–309, 3:308f
 research background, 4:437–438
 starvation effects, 4:305
- TRiC/CCT eukaryotic chaperonin, 1:457f, 1:458–459
- tricellulin, 4:392, 4:394f
- trichlorethene, 2:207f
- trichohyalin, 1:319t
- trichothiodystrophy (TTD), 3:339–340
- triethylsilyl triflate (Et₃SiOTf), 1:4f
- triflic acid (TfOH), 1:4f
- triflic anhydride (Tf₂O), 1:4f
- trifunctional protein (TFP), 1:385f, 1:386–387
- trigger loop, 4:174f, 4:175, 4:175t
- triglyceride (TG)
 diabetes, 1:655
 digestion and absorption, 2:358
 energy storage, 4:299–300
 insulin metabolism, 2:606–607
 lipogenic gene expression, 2:287f
 starvation effects, 4:303–304
- trihydroxylated metabolites, 2:188–189
- trimetazidine, 3:71t
- trimethylamine-N-oxide (TMAO), 4:79
- trimethylamine reductase, 1:487t
- trimethylsilyl triflate (TMSOTf), 1:4f
- triose phosphate isomerase (TPI), 2:445f, 3:499f
- triose phosphate (TP)
 photosynthetic carbon dioxide (CO₂) fixation, 3:494f
 photosynthetic metabolic pathways, 3:497, 3:498f
- triplex DNA (H-DNA)
 background information, 2:143–144
 biological role, 2:144
 bonding schemes, 2:144f
 chemical properties, 2:143–144, 2:143f
- Triticum vulgare*, 2:702t
- tritium (3H), 2:688
- TrkB tyrosine kinase (TrkB), 3:77–78
- tropomodulin, 1:34–35, 1:34f, 1:22f, 1:25
- tropomyosin, 1:31–32, 1:32f, 1:34–35, 1:34f, 1:18–19, 1:22f
- tropoin, 4:445–449
 actin organization, 1:31–32, 1:32f
 EF-hand proteins, 1:319t
 genetic disorders
 dilated cardiomyopathy (DCM), 4:447–449, 4:448f
 hypertrophic cardiomyopathy (HCM), 4:447–449, 4:448f
 inherited cardiomyopathies, 4:447–449
 knock-in mouse model, 4:449
 regulatory functions, 4:448f, 4:449
 restrictive cardiomyopathy (RCM), 4:447–449, 4:448f
 troponin C mutations, 4:448f, 4:449
 troponin I mutations, 4:448–449, 4:448f
 troponin T mutations, 4:447–448, 4:448f
 regulatory functions, 4:446–447, 4:448f, 4:449
 research background, 4:445
 subunits
 characteristics, 4:445–446
 troponin C, 2:578, 4:445–446
 troponin I, 4:445
 troponin T, 4:446, 4:446f
- TROY receptor protein, 4:454, 4:456f
- Trp185 binding protein, 1:276, 1:276f
- trp operon
 control mechanisms, 2:360–362, 2:361f
 research background, 2:360
 RNA-binding attenuation protein (TRAP)
 anti-TRAP (AT), 2:363f, 2:364
 bacterial occurrences, 2:364
 characteristics, 2:363, 2:363f
 general discussion, 2:362–364
 hairpin formation, 2:363, 2:363f
 leader transcript, 2:362, 2:362f
 timing factors, 2:363
 tRNA^{Trp}, 2:363f, 2:364
 transcription regulation, 2:360–362
- Trypanosoma*, 2:215
- Trypanosoma brucei*, 1:146, 1:580, 2:158t, 2:476, 2:478
- Trypanosoma cruzi*, 1:580, 3:551
- trypanosomatids, 2:160f
- trypanosomes
 calpains, 1:354f
 guanylyl cyclases (GCs), 1:567
- trypsin, 3:88, 3:587–588, 3:588f, 3:678–679
- trypsinogen, 3:587–588, 3:588f
- tryptophan
 biopterin, 3:667
 biosynthesis pathway, 4:249b
 classifications, 1:92t
 de novo biosynthetic pathways, 1:175f
 dioxygenases, 3:371–372
 functional roles, 1:93t
 mitochondrial DNA (mtDNA), 3:134t
 phage libraries, 3:434t
 trp operon, 2:360–362, 2:361f
 tryptophan pyrrolase, 3:373
 tryptophan synthase, 3:80
 WD (tryptophan (W)-aspartic acid (D) dipeptide) domain, 1:24
- tryptophan 2,3-dioxygenase (TDO), 1:175f
- tryptophanyl-tRNA synthetase, 4:443
- TSC1/TSC2 genes, 3:464f, 3:465
- tsp* gene, 2:42–43
- tuberculosis, 2:301
- tuberin, 4:5–6
- tuberous sclerosis, 3:198–199
- tuberous sclerosis/tuberin protein (TSC1/TSC2) complex, 1:408, 2:277, 2:604f, 2:605, 3:195–196
- tubulin, 4:450–453
 αβ-tubulin
 physical properties
 domain structures, 4:450–451, 4:451f
 folding processes, 4:451
 general characteristics, 4:450
 isotypes, 4:450
 microtubules, 3:233
 posttranslational modifications, 4:450
 synthesis, 4:451
 polymerization
 antimitotic tubulin ligands, 4:452, 4:452f
 dynamic instability, 4:451–452
 GTP cap model, 4:451–452
 microtubule assembly and structure, 4:451
 microtubule nucleation, 4:452
 cellular functions, 4:453
 δ-tubulin, 4:452–453

- ϵ -tubulin, 4:452
 G protein-coupled receptor kinases (GRKs), 2:485*t*
 γ -tubulin, 4:452
 η -tubulin, 4:453
 isoforms, 4:450
 microtubule-associated proteins (MAPs), 4:450
 rare isoforms, 4:452–453
 ζ -tubulin, 4:453
 tumor-associated collagen signatures (TACSs), 1:443*f*
 tumor necrosis factor (TNF)
 acid sphingomyelinase (aSMase), 4:287
 adiponectin, 3:74
 calbindin-D_{28k}, 1:287
 calcineurin-mediated signal transduction, 1:339–340, 1:340*f*
 calpains, 1:358
 caspases, 1:393, 1:394*f*
 glycogen synthase kinase-3 (GSK3), 2:438
 insulin resistance, 2:606–607
 insulin signal regulation, 2:606
 intermediate filaments, 2:636
 liver regeneration, 1:193
 matrix metalloproteinases (MMPs), 3:96
 mucins, 3:206–208, 3:208*f*
 neutral sphingomyelinase (nSMase), 4:287–288
 nuclear factor kappa B (NF- κ B), 3:303*t*, 3:304*t*
 p75 neurotrophin receptor (p75NTR), 3:241, 3:243–246, 3:244*f*, 3:245*f*, 4:459
 peroxisome proliferator-activated receptor gamma (PPAR γ), 3:422
 signaling pathways, 1:617–619, 1:618*f*
 TNF receptor-associated factor (TRAF) 6, 3:243, 3:244*f*, 3:245*f*, 4:377–378, 4:398, 4:399*f*, 4:400*f*
 TNF receptor associated protein 1 (TRAP-1), 3:166
 TNF receptor-associated protein with a death domain (TRADD), 1:394*f*, 1:617–619, 1:618*f*
 tumor necrosis factor receptors (TNFRs), 4:454–458
 effector functions
 apoptosis, 4:458
 cell survival, 4:458–459
 co-stimulation, 4:459
 general discussion, 4:458–459
 inflammation, 4:458
 organogenesis and development, 4:459
 expression regulation, 4:454–455
 ligand binding, 4:455, 4:457*f*
 signaling pathways
 characteristics, 4:455–458
 cytokines, 1:617–619, 1:618*f*
 death receptor signaling, 4:457–458
 TRAF-dependent signaling, 4:455–457, 4:458*f*
 structural characteristics, 4:454, 4:455*f*, 4:456*f*
 viral targeting, 4:459
 tumour necrosis factor α (TNF α)
 converting enzyme (TACE), 3:245, 3:245*f*, 4:51*f*, 4:53
 tumors
 cadherin signaling, 1:261–262
 chemokine receptors, 1:484
 DNA mismatch repair (MMR), 2:61–64, 2:62*t*, 2:63*f*
 interferon- γ receptors, 2:621
 mucin expression, 3:203–204
 tumor suppressors
 c-fes proto-oncogene (Fes), 1:453
 telomeres, 4:385–386
 tumor suppressor p53
 glutathione peroxidases, 3:411*f*
 ribosome biogenesis, 4:121
 tungsten (W), 4:137
 tunicamycin (TN), 2:457, 2:458*f*, 2:461–462, 3:647
 Turner syndrome, 3:23
 TWEAK receptor protein, 4:454, 4:456*f*
 20-hydroxyecdysone (20E), 2:592–593, 2:593*f*
 20S proteasome
 assembly processes, 3:596–597
 core particle (CP) gate, 3:597
 enzyme mechanism and inhibitors, 3:590
 immunoproteasome, 3:596
 N-terminal active site, 3:596, 3:596*f*
 occurrences, 3:595–597
 structural characteristics, 3:590, 3:591*f*, 3:595–596, 3:596*f*
 tymoproteasome, 3:596
 26S proteasome, 3:595–600
 basic concepts, 3:590–592, 3:591*f*
 cell regulation
 chaperones, 3:599
 E3 ubiquitin ligase, 3:599
 ER-associated degradation (ERAD), 2:200, 2:202*f*
 substrate degradation, 3:599
 ubiquitin proteasome pathway (UPP), 3:599
 characteristics, 3:595
 enzyme mechanism, 3:591, 3:592*f*
 19S regulatory complex
 assembly processes, 3:598
 base complex, 3:597*f*, 3:598
 basic concepts, 3:590–592
 lid complex, 3:597*f*, 3:598
 regulatory mechanisms
 characteristics, 3:598–599
 core particle (CP) gate opening, 3:599
 deubiquitylation, 3:598
 substrate recognition, 3:598
 translocation and unfolding processes, 3:599
 structural characteristics, 3:597–598, 3:597*f*
 physiological importance, 3:592
 protein degradation
 characteristics and functional role, 3:620–621, 3:621*f*
 degradation mechanisms, 3:622
 19S proteasome regulatory particle, 3:621, 3:621*f*
 20S proteasome core particle, 3:620–621, 3:621*f*, 4:473–474, 4:474*f*
 research summary, 3:599–600
 structural characteristics, 3:597*f*
 tissue and subcellular distribution, 3:591–592
 ubiquitination, 4:473–474, 4:474*f*
 ubiquitin–proteasome degradation pathway, 3:619–622, 3:619*f*, 4:473–474, 4:474*f*
 TWINKLE helicase, 2:21, 2:22*t*, 2:26
 Twisted gastrulation (Tsg) protein, 1:229–230
 two-base ribozymes, 4:144
 two-hybrid protein–protein interaction systems, 4:460–462
 basic concepts, 4:460–462, 4:461*f*
 proteomic studies, 4:462
 research applications, 4:461–462
 reverse two-hybrid systems, 4:461–462
 two-pore channels, 1:335–336
 tymoproteasome, 3:596
 type I DNA restriction and modification enzymes, 2:72–75
 antirestriction reactions, 2:75–76
 assembly and control mechanisms, 2:74
 biochemical function, 2:72, 2:73*f*
 protein structures
 hsd family gene expression, 2:72–74
 modification (M) subunits, 2:73
 restriction (R) subunits, 2:74
 sequence specificity (S) subunits, 2:72–73, 2:73*f*
 restriction reactions
 adenosine triphosphate (ATP) hydrolysis, 2:74, 2:75*f*
 DNA binding, 2:74
 DNA cleavage, 2:75
 DNA translocation, 2:74–75, 2:75*f*
 unmodified DNA recognition, 2:74
 type I DNA topoisomerase, 2:157–162
 catalytic reactions, 2:157
 cellular roles
 hyperthermophilic eubacteria, 2:161–162
 mesophilic/psychrophilic eubacteria, 2:161
 replication processes, 2:162, 2:162*f*
 supercoils, 2:161, 2:162*f*
 transcription processes, 2:162, 2:162*f*
 type II DNA topoisomerase, 2:161–162
 classifications, 2:158, 2:158*t*
 DNA supercoiling, 3:369*f*
 Escherichia coli, 3:369*t*
 functional role, 2:157, 2:163–164
 nomenclature, 2:158, 2:158*t*
 properties, 2:158
 type IA DNA topoisomerase
 classifications, 2:158
 cleavage/strand-passage domain crystal structure, 2:158–159, 2:159*f*
 protein domains, 2:158, 2:159*f*
 reverse gyrase mechanism, 2:159–160
 sources, 2:158*t*

- type I DNA topoisomerase (*Continued*)
 strand passage enzyme bridging
 mechanism, 2:159, 2:159f
- type IB DNA topoisomerase
 catalytic reactions, 2:160–161
 classifications, 2:158
 crystal structure, 2:160, 2:161f
 domain structure, 2:160, 2:160f
 sequence conservation, 2:160, 2:160f
 sources, 2:158t
 strand passage rotational mechanism,
 2:161
- type II DNA topoisomerase, 2:163–168
 archaeal type IIB DNA topoisomerase,
 2:166
 cellular roles, 2:161–162
 characteristics, 2:164–165
 domain structures, 2:164–165, 2:165f
 double-stranded DNA passage reaction,
 2:164, 2:164f
 enzyme mechanism, 2:164, 2:164f
Escherichia coli, 3:369t
 functional role, 2:1
 Top6A/Top6B, 2:165, 2:165f, 3:369–370
- type IIA DNA topoisomerase
 DNA gyrase, 2:165–166
 domain structures, 2:164–165, 2:165f
 eukaryotic type IIA DNA topoisomerase,
 2:166
 prokaryotic type IIA DNA
 topoisomerase, 2:165–166
 as therapeutic targets
 antibacterial drugs, 2:167, 2:167f
 anticancer drugs, 2:167, 2:167f
 dietary topoisomerase II poisons,
 2:167–168, 2:167f
 functional role, 2:167–168
 type IV DNA topoisomerase, 2:166
- type IIB DNA topoisomerase, 2:165,
 2:165f, 2:166
- type II restriction and modification enzymes,
 2:130–135
 S-adenosyl-L-methionine (AdoMet), 2:130
 characteristics
 nonspecific DNA interactions, 2:133
 protein structures, 2:131
 recognition sequences, 2:131, 2:132t
 sequence cleavage mechanisms,
 2:132–133
 classifications, 2:130
 methyltransferases
 enzyme characteristics, 2:134
 modification functions, 2:134
 new systems research, 2:130
 organization
 expression regulation, 2:131
 modification and restriction genes,
 2:130–131
 research applications, 2:134
 research background, 2:130
 subsets
 background information, 2:133–134
 type IIB enzymes, 2:132t, 2:133–134
 type IIE enzymes, 2:132t, 2:133
 type IIF enzymes, 2:132t, 2:133
 type IIS enzymes, 2:132t, 2:133
- type III DNA restriction and modification
 enzymes, 2:136–141
 action mechanisms, 2:138–139, 2:139f
 S-adenosyl-L-methionine (AdoMet), 2:136,
 2:137f, 2:137t
 cofactor requirements, 2:136
 domain organization
 methyltransferase (MTase), 2:137–138,
 2:137t
 modification (M) subunits, 2:137–138,
 2:137t
 restriction subunit, 2:136–137, 2:137f
 phase variations, 2:139–140
 restriction activity regulation, 2:136
- type III DNA topoisomerase
Escherichia coli, 3:369t
 protein domains, 2:159f
 sources, 2:158t
 type IIIA DNA topoisomerase, 4:45, 4:48f
- type IV DNA topoisomerase, 2:164–165,
 2:165f, 3:369t
- type VI DNA topoisomerase, 2:166
- Type I glycogen storage disease
See von Gierke disease
- Type Ia glycogen storage disease, 2:431,
 2:431t
- Type Ib glycogen storage disease, 2:431,
 2:431t
- Type II glycogen storage disease
See Pompe disease
- Type III glycogen storage disease
See Cori/Forbes disease
- Type IV glycogen storage disease
See Andersen disease
- Type IXa glycogen storage disease, 2:431t,
 2:432
- Type IXb glycogen storage disease, 2:431t,
 2:432
- Type IXc glycogen storage disease, 2:431t,
 2:432–433
- Type V glycogen storage disease
See McArdle disease
- Type VI glycogen storage disease
See Hers disease
- Type VII glycogen storage disease
See Tauri disease
- tyrosine
 anaplerotic pathways, 1:106f
 biopterin, 3:667
 classifications, 1:91, 1:92t
 functional roles, 1:93t
 immunoreceptor tyrosine-based activated
 motifs (ITAMs)
 B-cell antigen receptors (BCRs), 1:162,
 1:162f, 1:163f
 immunoglobulin Fc receptors (FcRs),
 2:569, 2:569f, 2:571
 receptor–proximal signaling, 4:376
 mucins, 3:206
 phage libraries, 3:434t
 phosphorylation, 2:620–621, 2:620f,
 2:622f, 4:61
 platelet-derived growth factor (PDGF)
 receptors, 3:539–540, 3:540f
 protein tyrosine phosphatase (PTP),
 3:648–653
- catalytic mechanisms, 3:648, 3:649f
 disease connections
 abnormal functions, 3:649–652
 CD45 gene, 3:649–651, 3:650f, 3:651t
 gene disruption studies, 3:651t
 Laforin, 3:650f, 3:651–652, 3:651t
 protein tyrosine phosphatase PTB1B,
 3:649
- intracellular PTPs
 CDC25 family, 3:648, 3:650f, 3:651f
 classic PTPs, 3:648, 3:650f, 3:651f
 dual-specific PTPs (DS-PTPs), 3:648,
 3:650f
 low-molecular-weight PTPs (LMW-
 PTPs), 3:648–649, 3:650f, 3:651f
 receptor-like PTPs (RPTPs), 3:649,
 3:650f
 major sperm protein (MSP), 3:3, 3:3f
 phosphorylation, 3:648, 4:294–295,
 4:296f
 protein tyrosine phosphatase PTB1B,
 2:260f, 3:649
 redox signaling mechanisms,
 3:411–412, 3:411f
 therapeutic uses, 3:652
- recombinases
 retrotransposons, 3:285f, 3:286f
 site-specific recombination systems,
 1:556–558, 1:557f, 1:558f
- SHP-2 (Src homology 2 domain-
 containing protein tyrosine
 phosphatase), 1:113, 3:539–540,
 3:540f, 3:541f, 3:542, 3:651t
- site-specific recombination systems,
 1:556–558, 1:557f, 1:558f
- Src (sarcoma) family kinases, 4:293–297
 angiotensin receptors, 1:113
 cadherin signaling, 1:264, 1:264f
 calcium ion (Ca²⁺) signaling system,
 2:719–720, 2:720f
 cell migration, 1:439, 1:439f, 1:441,
 1:441f
 c-fes proto-oncogene (Fes), 1:451, 1:452f
 characteristics, 4:293
 chemokine receptors, 1:480–482, 1:482f
 focal adhesion kinase (FAK) family,
 2:265–266, 2:267f, 2:611,
 4:294–295
 functional role, 4:294
 group III 5-phosphatases, 3:462
 hematopoietin receptors, 2:526, 2:528f
 historical research, 4:293–294
 immune responses, 4:294–295
 invertebrate Src kinases, 4:297
 neurokinin-1 receptor (NK1R),
 4:362–363
 phosphorylation, 4:295–297, 4:296f
 platelet-derived growth factor (PDGF)
 receptors
 chemotaxis, 3:542
 mediated cellular responses, 3:540,
 3:541f
 signaling pathways, 3:539–540,
 3:540f
 protein palmitoylation, 2:738,
 4:296–297, 4:296f

- protein prenylation, 2:737t
protein tyrosine kinase (PTK), 1:162, 1:163f, 3:470
redundant functions, 4:294
regulation mechanisms, 4:295–297, 4:296f
research summary, 4:297
SH2/SH3 domains, 1:451–452, 1:452f, 4:294, 4:295–297, 4:296f
signaling adhesion functions, 4:294–295
steroid/thyroid hormone receptors, 4:310f
synthetic pathways, 1:57
tricarboxylic acid cycle, 4:439f
tyrosine kinase receptors (Trks)
glycogen synthase kinase-3 (GSK3), 2:435f
messenger RNA (mRNA), 2:602–603
neurotrophins
cross-talk, 3:246
functional role, 3:241
p75 neurotrophin receptor (p75NTR), 3:246
signaling pathways, 3:241–243, 3:242f
phospholipase D (PLD), 3:474, 3:474f
signaling pathways, 3:539–540, 3:540f
T-cell receptor (TCR) signaling, 4:376
vascular endothelial growth factor (VEGF)
biological development, 4:512–513
evolutionary development, 4:509
lymphangiogenesis, 4:513
neovascularization, 4:513
signal transduction, 4:511, 4:512f
structural characteristics, 4:509
transmembrane domains, 4:510, 4:510f
tyrosine kinases
calcium ion (Ca^{2+}) signaling system, 2:719–720, 2:720f
c-fes proto-oncogene (Fes), 1:451, 1:452f
chemokine receptors, 1:480–482, 1:482f
dual-specificity tyrosine phosphorylation-regulated kinase 2 (DYRK2), 1:312, 1:313f
fibroblast growth factor receptors (FGFRs), 2:303–304, 2:304f, 2:305f
focal adhesion kinase (FAK) family, 2:265–268
activation and signaling mechanisms, 2:265–266, 2:267f
cellular responses, 2:266
characteristics, 2:265–266, 2:266f
domain conservation, 2:266t, 2:267–268
expression regulation, 2:265
FERM domain, 2:265–266, 2:266f
intracellular signaling pathway connections, 2:611
invertebrates, 2:267–268
localizations, 2:265–266, 2:266f
signaling adhesion functions, 4:294–295
focal adhesions, 2:321
neurotransmitter transporters, 3:239
phospholipase C (PLC), 3:470
proline-rich tyrosine kinase, 2:345f
proline-rich tyrosine kinase 2 (PYK2)
activation and signaling mechanisms, 2:266–267, 2:267f
cellular responses, 2:267
characteristics, 2:266–267, 2:266f
domain conservation, 2:266t
protein-tyrosine kinase-linked receptors (PTKRs), 2:583, 2:584f
protein tyrosine kinase (PTK), 1:162, 1:163f, 3:470
protein tyrosine phosphatase PTB1B, 2:260f
protein tyrosine phosphatase (PTP), 3:411–412, 3:411f, 4:294–295, 4:296f
somatostatin (sst) receptors, 4:265
Tec family, 1:162, 1:163f, 1:264, 2:620–621, 3:465–466, 3:539–540, 3:540f
ubiquitin-proteasome degradation pathway, 3:623t
tyrosine sulfation, 4:463–466
general discussion, 4:463
trans-Golgi network, 4:463, 4:464f
tyrosine-sulfated proteins, 4:463, 4:464f
tyrosylprotein sulfotransferase (TPST)
medical importance, 4:465–466
membrane topology, 4:465
molecular cloning, 4:465
physiological functions, 4:465–466
properties, 4:464–465
TPST-1- and TPST-2-deficient mice, 4:465
TPST-1 versus TPST-2, 4:465
tyrosine-sulfation reaction, 4:463–464, 4:464f
See also vascular endothelial growth factor (VEGF)
tyrosyl-tRNA synthetase, 4:443
tyroxine, 1:93t
- ## U
- U2 auxiliary factor (U2AF), 3:62–63, 3:62f
UAS_e elements
See GAL promoter system
ubichromenols, 3:723–724, 3:723t
ubiquinol (QH₂)
biochemical free radical chain reactions, 3:169f, 3:170f
chemical characteristics, 3:723–724
mitochondrial ROS, 1:210
respiratory chain complexes, 4:85
respiratory chain complex III (ubiquinol cytochrome *c* reductase)
background information, 4:98
iron-sulfur clusters (FeS), 1:596f, 3:663–664, 3:664f
mitochondrial ROS, 4:19f, 4:20
periplasmic electron-transport systems, 3:400, 3:401f
proton conduction pathways, 1:597–598
protonmotive electron transfer chain, 1:470–471, 1:470f
protonmotive Q cycle, 1:597f
schematic diagram, 4:99f
ubiquinol-oxidizing enzymes, 4:85
ubiquinone
archaeal energy transduction, 3:34
bacteriopheophytins, 3:696–697, 3:696f, 3:700, 3:701f, 3:702f
biochemical free radical chain reactions, 3:169f, 3:170f
chemical characteristics, 3:723–724
cholesterol biosynthetic pathway, 1:516, 1:517f
cytochrome *bc*₁ complex, 1:596f, 1:597f, 1:598
ETF:ubiquinone oxidoreductase (ETF-QOR), 4:19f, 4:20
homologs and analogs, 3:723t
mitochondrial ROS, 1:210
plant mitochondrial electron transport chain, 1:141–142, 1:142f, 1:145, 1:146f
reactive oxygen species (ROS), 2:77, 4:18–20
respiratory chain complex I (NADH:ubiquinone oxidoreductase), 4:87–91
background information, 4:98
bacterial reductive reactions, 1:488f
basic concepts, 4:87, 4:341
disease connections
general discussion, 4:90–91
mitochondrial mutations, 4:90–91
neurotoxins, 4:91
nuclear mutations, 4:91
pathological mechanisms, 4:91
evolutionary development
general discussion, 4:88–90
N-module, 4:89
origins, 4:89–90
P-module, 4:89
Q-module, 4:88–89
variants, 4:89–90
functional role
active-deactive transition, 4:90
electron transfer mechanisms, 3:661–662, 3:662f, 3:663, 4:83–84, 4:90
general discussion, 4:90
inhibitors, 4:90
proton pumping, 3:661–662, 3:662f, 3:663, 4:90
redox centers, 4:90
sodium ion (Na^+) pumping, 4:90
mitochondrial ROS, 4:18–20, 4:19f
PET genes, 3:308f
plant mitochondrial electron transport chain, 1:145, 1:146f
protonmotive electron transfer chain, 1:470f
reverse electron transport (RET), 4:19f, 4:20
schematic diagram, 4:84f, 4:99f
structural characteristics, 4:341–346
accessory subunits, 4:88, 4:89t
central subunits, 4:87–88, 4:88f, 4:89t

ubiquinone (*Continued*)

cysteine motif, 4:344–345, 4:345f
 flavin mononucleotide (FMN)
 binding, 4:344–345, 4:344f
 frataxin-like subunit, 4:343–344, 4:343f
 hydrophilic domain, 4:342–343, 4:342f
 iron–sulfur clusters, 4:87–88, 4:89t, 4:342–343, 4:342f, 4:345f
 multistage assembly pathway, 4:88
 mutations, 4:345
 NADH binding, 4:344–345, 4:344f, 4:345f
 overall structure, 4:341–342
 redox centers, 4:87–88, 4:89t, 4:342–343, 4:342f
 research summary, 4:345–346
 shape, 4:87, 4:88f
 subunits, 4:342t
 supernumerary subunits, 4:345
 respiratory chain complexes, 1:145–146, 4:83
 succinate:quinone oxidoreductases (SQORs), 4:94f
 synthetic pathways, 3:712, 3:713f, 3:714f
 ubiquinone-reducing enzymes
 electron-transferring flavoprotein (ETF)–ubiquinone oxidoreductase, 4:85
 external NADH dehydrogenase, 4:85
 glycerol-3-phosphate dehydrogenase, 4:85
 succinate dehydrogenase, 4:85
 ubiquitin, 4:473–476
 cellular functions, 4:474–475
 degradation pathways, 2:525, 3:599, 4:473–474, 4:474f
 deubiquitylating enzymes (DUBs), 3:597–598, 3:620
 disease connections, 4:475–476
 drug development, 4:476
 functional role, 4:467, 4:473
 heat shock response, 2:523
 polyubiquitination, 3:542–543, 3:619–620
 regulation mechanisms, 4:475
 small ubiquitin-related modifiers (SUMOs), 3:191–192, 4:69, 4:468–470, 4:468t
 substrate recognition, 4:474
 T-cell receptor (TCR) signaling, 4:379
 26S proteasome, 3:595–600
 basic concepts, 3:590–592, 3:591f
 cell regulation
 chaperones, 3:599
 E3 ubiquitin ligase, 3:599
 substrate degradation, 3:599
 ubiquitin proteasome pathway (UPP), 3:599
 characteristics, 3:595
 enzyme mechanism, 3:591, 3:592f
 19S regulatory complex
 assembly processes, 3:598
 base complex, 3:597f, 3:598
 basic concepts, 3:590–592
 lid complex, 3:597f, 3:598

regulatory mechanisms, 3:598–599
 structural characteristics, 3:597–598, 3:597f
 physiological importance, 3:592
 protein degradation
 characteristics and functional role, 3:620–621, 3:621f
 degradation mechanisms, 3:622
 19S proteasome regulatory particle, 3:621, 3:621f
 20S proteasome core particle, 3:620–621, 3:621f, 4:473–474, 4:474f
 research summary, 3:599–600
 structural characteristics, 3:597f
 tissue and subcellular distribution, 3:591–592
 ubiquitination, 4:473–474, 4:474f
 ubiquitin–proteasome degradation pathway, 3:619–622, 3:619f, 4:473–474, 4:474f
 ubiquitination, 1:412f, 4:473, 4:474f
 ubiquitin-binding zinc finger (UBZ), 4:573–574, 4:573f
 ubiquitin-like proteins (UBLs), 4:467–472
 Atg8 and Atg12 conjugation systems, 1:151–153, 1:152t, 4:468t, 4:471
 attachment dynamics, 4:467–468, 4:468f, 4:468t
 Fat10 (diubiquitin), 4:468t, 4:471–472
 functional role, 4:467, 4:468f, 4:475
 Isg15 (interferon-induced 15 kDa protein), 4:468t, 4:470–471
 neural precursor cell expressed, developmentally downregulated 8 (NEDD8), 4:468t, 4:470, 4:475
 small ubiquitin-related modifiers (SUMOs), 4:69, 4:468–470, 4:468t, 4:475
 ubiquitin–proteasome degradation pathway
 degradation mechanisms, 3:619–622, 3:619f
 disease connections, 3:622
 hybrid proteasome, 3:621–622, 3:621f
 immunoproteasome, 3:621–622
 mitochondria, 3:622–623, 3:623t
 signal transduction, 3:304f
 26S proteasome
 characteristics and functional role, 3:620–621, 3:621f
 degradation mechanisms, 3:622
 19S proteasome regulatory particle, 3:621, 3:621f
 20S proteasome core particle, 3:620–621, 3:621f
 ubiquitin conjugation
 deubiquitination, 3:620
 polyubiquitin chain formation, 3:619–620
 substrate recognition mechanisms, 3:620
 UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1), 3:9, 3:9f
 von Hippel–Lindau protein (VHL)
 functional role, 4:570, 4:571f

hypoxia inducible factor (HIF)
 regulation, 4:571
 structural characteristics, 4:570–571, 4:571f
 ubiquitylation, 4:61–62
 ubisemiquinone (CoQH)
 biochemical free radical chain reactions, 3:168–169, 3:170f, 3:171f, 3:172f
 chemical characteristics, 3:723–724
 mitochondrial ROS, 1:210
 UCH family, 4:468t
 udenafil, 1:569
 UGA codons
 protein biosynthesis elongation, 2:249–251
 selenocysteine specificity, 4:213
 stop codons, 2:249–251, 4:212
 UHDBT, 1:597f, 1:598
 UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1), 3:9, 3:9f
Ulex europaeus, 2:705
 Ullrich congenital muscular dystrophy, 3:167
 ultrafine bridges (UFBs), 4:49
 ultraspiracle (USP) protein, 2:593, 2:593f
 ultraviolet (UV) radiation
 apoptotic pathways, 1:389f
 DNA damage
 cyclobutane pyrimidine dimers (CPDs), 4:572–573
 polymerase η (Pol η), 4:572–574
 pyrimidine 6–4 pyrimidone photoproduct (6–4 PP), 4:574
 xeroderma pigmentosum (XP), 4:572
 photoproduct generation, 3:338, 3:339f, 3:341
 phototrophic bacteria, 3:32
 UV-DDB complex, 3:341, 3:342f, 3:342t
 umami taste receptors, 4:366–367
 UmuDC protein/*umuDC* genes
 bacterial homologs, 4:480–481
 cleavage reactions, 4:477–478, 4:478f
 DNA polymerase V discovery, 4:479
 mutagenesis, 4:478
 regulatory mechanisms, 4:480
 SOS response, 4:477, 4:478f
 uncharged glycerolipids, 2:416–417, 2:416f
 uncoupling proteins (UCPs), 4:482–487
 adenosine diphosphate (ADP)
 phosphorylation and respiration
 adenosine triphosphate (ATP) synthesis, 4:482
 brown adipose tissue (BAT), 4:483–484, 4:484f
 mitochondrial membranes, 4:482
 ATP synthesis, 1:142f, 1:144, 4:482
 brown adipose tissue (BAT)
 adenosine diphosphate (ADP)
 phosphorylation and respiration, 4:483–484, 4:484f
 characteristics and functional role, 4:483
 histology, 4:483f
 thermogenesis, 4:483–484, 4:484f
 UCP1, 4:483–484, 4:484f, 4:485f
 characteristics and functional role, 4:482
 leptin, 2:258–259
 mitochondrial ROS, 4:21

- novel UCPs
 energy loss mechanisms, 4:484–486
 functional role
 glucose metabolism, 4:486
 reactive oxygen species (ROS), 4:485–486, 4:486f
 messenger RNA (mRNA), 4:484–485, 4:485f
 plant and bird UCPs, 4:485
 skeletal muscle, 4:485, 4:485f
 UCP2, 4:484–485, 4:485f
 UCP3, 4:485, 4:485f
 research summary, 4:486
- undecaprenyl pyrophosphate (Und-P-P), 2:316–317
- unfolded protein response (UPR), 4:488–493
 basic concepts, 4:488
 chaperones, 4:488, 4:492f
 endoplasmic reticulum (ER)
 ER-associated degradation (ERAD), 2:524f, 2:525
 health monitoring, 2:524–525, 2:524f
 functional roles, 4:488
- metazoans
 ATF6 gene activation, 4:490–491, 4:491f
 inositol requiring 1 (IRE1) gene
 conservation, 4:490
 intramembrane proteolysis, 4:490–491
 physiological and pathological stresses
 ATF6 gene activation, 2:524–525, 2:524f, 4:490–491, 4:491f
 basic concepts, 4:488–489
 gene expression regulation, 4:491–492
 HAC1 gene expression, 2:524–525, 2:524f, 4:489, 4:490f
 inositol requiring 1 (IRE1) genes, 2:524–525, 2:524f, 4:489, 4:490f
 intramembrane proteolysis, 4:490–491, 4:491f
 stress sensors, 4:492
 translation initiation factor 2 (eIF2 α)
 phosphorylation, 4:491, 4:492f
 XBP-1 gene expression, 2:524–525, 2:524f
- stress responses, 2:522
- translational control
 gene expression regulation, 4:491–492
 PERK (PKR-like ER kinase), 2:524–525, 2:524f, 4:491, 4:492f
 stress sensors, 4:492
 translation initiation factor 2 (eIF2 α)
 phosphorylation, 2:524–525, 2:524f, 4:491, 4:492f
- yeasts
 HAC1 gene expression, 4:489, 4:490f
 inositol requiring 1 (IRE1) genes, 4:489, 4:490f
- unimolecular nucleophilic substitution (S_N1), 2:9
- UniProt Consortium, 3:611–612
- UniProtKB, 3:615t
- unit transposons, 3:276–277, 3:276f
- unnatural ribozymes, 4:144, 4:145f
- unsaturated fatty acids
 β -oxidation, 2:282–283, 2:283f
 biosynthesis
 arachidonic and docosahexenoic acid
 synthesis, 2:288, 2:288f
 oleic and palmitoleic acid synthesis, 2:288
 protein palmitoylation, 2:737
- untranslated leader sequences (UTLs), 3:144
- untranslated regions (UTRs)
 c-fes proto-oncogene (Fes), 1:454
 5' untranslated region, 1:454, 3:60–61, 4:425–426, 4:426f
 3' downstream region, 4:425–426, 4:426f
 3'-end maturation, 3:188, 3:192
 3' untranslated region, 1:182–183, 3:60–61
- uPA–uPA receptor signaling, 2:455
- uracil
 degradation pathways, 3:717–718, 3:717f
 MBD4-mediated repair, 1:523–524, 1:523f
 mismatch uracil glycosylase (MUG) family, 2:20
 synthetic pathways, 3:713f
 uracil DNA glycosylase (UNG), 2:16, 2:18, 2:19–20
- uranium (U), 1:487t
- Ure2p protein
 characteristics, 3:566
 disease states, 3:568–569
 prion domains, 3:568f
 self-propagating amyloidoses, 3:567–568, 3:568f
- [URE3] prion
 disease states, 3:568–569
 generation mechanisms, 3:569
 genetic properties, 3:566, 3:567f
 self-propagating amyloidoses, 3:567–568
- urea
 insect biochemistry, 2:590–591, 2:591–592, 2:592f
 protein folding and assembly processes, 3:625
- urea cycle
 arginase, 4:494, 4:495f, 4:495t, 4:500
 argininosuccinate lyase (AL), 4:494, 4:495f, 4:495t, 4:499
 argininosuccinate synthetase (AS), 4:494, 4:495f, 4:495t, 4:498–499
 carbamyl phosphate synthetase (CPS), 4:494, 4:495f, 4:495t, 4:496f, 4:497–498
- citrin deficiency
 adult-onset citrullinemia, 4:499
 characteristics, 4:499
 neonatal intrahepatic cholestasis, 4:499
- diseases, 4:494–501
 arginase deficiency
 (hyperargininemia), 4:500
 argininosuccinate lyase (AL)
 deficiency (argininosuccinic aciduria), 4:499
 argininosuccinate synthetase (AS)
 deficiency, 4:498–499
 carbamyl phosphate synthetase (CPS)/N-acetylglutamate (NAG)
 synthetase deficiencies, 4:497–498
- citric deficiency, 4:499
- citrullinemia, 4:498–499
- hyperammonemia, 4:494
 hyperammonemia–
 hyperornithinemia–
 homocitrullinemia (HHH)
 syndrome, 4:495f, 4:495t, 4:500–501
- ornithine transcarbamylase (OTC)
 deficiency, 4:498
- retinal and choroidal gyrate atrophy, 4:500
- transient hyperammonemia, 4:500
- functional role, 4:494, 4:495f
- hyperammonemia
 clinical manifestations, 4:494
 diagnosis, 4:494–495, 4:496f
 inborn errors, 4:494, 4:495t
 long-term therapy considerations, 4:497
 treatments, 4:495–497, 4:497t
- N-acetylglutamate (NAG) synthetase, 4:494–495, 4:495f, 4:495t, 4:497–498
- ornithine transcarbamylase (OTC), 4:494, 4:495f, 4:495t, 4:496f, 4:498
- research background, 4:436–437
- urinary nitrogen excretion, 4:306f
- ureagenesis, 1:94–95
- Ureoplasma ureolyticum*, 3:331f
- uric acid
 antioxidant actions, 2:77
 gout, 2:312
 urinary nitrogen excretion, 4:306f
- uridine, 3:713f
- uridine diphosphate (UDP)
 mitochondrial ATP-dependent potassium ion (K⁺) channel (KATP), 3:551
- P2Y receptors, 3:375–378
 domain structures, 3:376
 extracellular nucleotide signaling, 3:375–376
 functional classification, 3:376–377, 3:376t
 functional role, 3:375
 lysophosphatidic acid (LPA) receptors, 2:763t
 pharmacological classification, 3:377
 physiological functions, 3:377–378
 synthetic pathways, 3:713f
- uridine diphosphate (UDP)-galactose (UDP-Gal), 1:3
- uridine diphosphate (UDP)-Glc:
 glycoprotein glucosyltransferase (UGGT)
 N-linked glycans processing, 3:258f, 3:259t
 N-linked glycosylation, 2:200–201, 2:449f, 2:450, 2:450f, 2:451f, 2:457, 2:458f, 2:460f, 2:461f
- uridine diphosphate (UDP)-glucose
 glycogen synthesis, 2:425, 2:426f, 2:428, 2:430f
 glycosyltransferase-catalyzed reactions, 1:3
 photosynthetic metabolic pathways, 3:498f

- uridine diphosphate (UDP)-glucuronic acid (UDP-GlcUA), 1:3
- uridine diphosphate (UDP)-N-acetylgalactosamine (UDP-GalNAc), 1:3
- uridine diphosphate (UDP)-N-acetylglucosamine (GlcNAc)
glycosyltransferase-catalyzed reactions, 1:3
- glycosylphosphatidylinositol (GPI)-anchored proteins, 2:476, 2:477f
- hexosamine pathway, 1:656
- uridine monophosphate (UMP)
cyclobutane pyrimidine dimers (CPDs), 4:405
- de novo* synthesis, 3:712, 3:713f, 3:714f
- uridine triphosphate (UTP)
glycogen synthesis, 2:430f
- P2Y receptors, 3:375–378
domain structures, 3:376
extracellular nucleotide signaling, 3:375–376
functional classification, 3:376–377, 3:376t
functional role, 3:375
lysophosphatidic acid (LPA) receptors, 2:763t
pharmacological classification, 3:377
physiological functions, 3:377–378
synthetic pathways, 3:712–713, 3:713f, 3:715f
- urinary tract infections (UTIs), 2:306, 3:115
- urine
maple syrup urine disease (MSUD)
branched-chain α -ketoacid
dehydrogenase complex (BCKDC), 1:244, 1:245f
branched-chain amino acid (BCAA)
regulation, 1:245–246
thiamine deficiency, 1:209
urinary nitrogen excretion, 4:306, 4:306f
- uroporphyrinogen decarboxylase, 2:540f, 2:541, 2:542t
- uroporphyrinogen III, 2:540–541, 2:540f, 3:544–546, 3:545f, 3:546f
- ursodeoxycholate (UDC), 1:168f, 1:169, 1:170
- ursodeoxycholic acid (UDCA), 3:561, 3:565
- Usher's syndrome, 3:223–224, 3:223t
- U small nuclear RNA (snRNA), 3:322
- Ustilago maydis*, 3:441
- uterus
long-chain acyl-CoA synthetase (LCAS), 2:292–293
proteinase-activated receptors (PARs), 3:605t
- uterine cervical carcinoma, 4:475–476
- uterine mucins, 3:205–208
background information, 3:205
embryo implantation process, 3:205, 3:206f
expression and transcriptional control, 3:206–208, 3:208f
functional role, 3:207t, 3:208–209
research summary, 3:209
transmembrane proteins, 3:207t
- transport and metabolism regulation, 3:208–209
- utrophin, 2:625–627
- UvrABCD system, 3:328–336
bacteria–human conservation model, 3:336
damage recognition mechanisms
general discussion, 3:333–334
padlock model, 3:334–335
UvrA₂–DNA complex, 3:333
UvrA₂B₂ complex in solution, 3:333
UvrA₂–DNA binding, 3:333
UvrA dimerization, 3:333
UvrB–DNA intermediate complex
formation, 3:333–334
damage sensing mechanisms
basic concepts, 3:334–335, 3:334f
DNA polymerase I, 3:335
dual incision step, 3:334
LexA regulatory protein, 2:709–710
postincision steps, 3:334
UvrB–DNA complex, 3:334–335, 3:334f
UvrD, 3:334
DNA mismatch repair (MMR), 2:41f, 2:45–46, 2:67
functional role, 2:54–56, 2:55t, 2:56f
protein properties
UvrA, 3:330–332, 3:332f
UvrB, 3:332, 3:332f
UvrC, 3:332, 3:333f
substrate ranges, 3:330t
transcription-coupled repair (TCR), 3:335–336, 3:335f
UvrABCD endonuclease, 3:328, 3:329f, 3:330t
- UvrD helicases
characteristics and functional role, 3:261
damage sensing mechanisms, 3:334
DNA unwinding
active, rolling or hand-over-hand
model, 3:264–265, 3:265f
active versus passive mechanisms, 3:264
dimeric inch-worm models, 3:265, 3:265f
inch-worm mechanisms, 3:264, 3:264f, 3:265f
protein oligomerization, 3:263
unwinding mechanisms, 3:263–264
research summary, 3:265
single-strand DNA (ssDNA)
translocation, 3:263, 3:264, 3:264f
structural characteristics
crystal structures, 3:262, 3:262f
open and closed confirmations, 3:262–263
primary structures, 3:261–262
translocase-based protein displacement, 3:265
- uvr* genes
Cho, 3:330, 3:331f
historical research, 3:328–330
uvrA, 3:328, 3:329f
uvrB, 3:329f, 3:330
uvrC, 3:329f, 3:330, 3:331f
UvsX/UvsY proteins, 4:23–24, 4:37, 4:37f
- V**
- vaccine adjuvants, 4:399–400
- vacuolar (H⁺)-ATPases, 4:520–524
functional role
intracellular V-ATPases, 4:520, 4:521f
plasma membrane V-ATPases, 4:520–521
in vivo activity regulation, 4:523
structural characteristics, 4:521–523, 4:522f
subunit composition, 4:522t
transport mechanisms, 4:523
- vacuolar protein sorting 4 (Vps4) protein, 1:3t, 1:5–6
- vacuoles, 4:502–508
characteristics and functional role, 4:502–503, 4:503f, 4:504f
eukaryotic functions
disease connections, 4:507–508, 4:507t
lysosome-related organelles, 4:506–507
pathogen-containing vacuoles, 4:507
protozoan contractile vacuole complex, 4:506
historical research, 4:502–503
yeasts
biogenesis, 4:503–506
characteristics, 4:502–503, 4:503f
functional role, 4:503, 4:504f
transport pathways
AP-3 pathway, 4:505f, 4:506
autophagy/macro-autophagy, 4:505f, 4:506, 4:506f
characteristics, 4:503–506, 4:505f
Vps pathway, 4:504–506, 4:506f
- valine
anaplerotic pathways, 1:106f
classifications, 1:92t
degradative pathways, 1:244, 1:245f
elastin, 2:192
leucine–aspartic acid–valine (LDV)
peptide, 2:608
phage libraries, 3:434t
starvation effects, 4:302
- valitocin, 4:516t
- valproate, 2:292
- Van Hippel Lindau (VHL) substrate model, 1:459
- varafenil, 1:569
- VARICOSE (VCS) gene, 3:100
- variegate porphyria, 2:542t
- vascular endothelial growth factor (VEGF)
cadherin signaling, 1:264
matrix metalloproteinases (MMPs), 3:96
neurogenesis, 1:197f, 1:198, 1:200f
nuclear factor kappa B (NF- κ B), 3:304t
transient receptor potential (TRP)
channels, 4:415
- vascular endothelial growth factor
receptors (VEGFRs), 4:509–514
biological activity
embryonic development, 4:512–513
lymphangiogenesis, 4:513
neovascularization, 4:513
nonmitogenic functions, 4:513–514
evolutionary development, 4:509

- gene expression regulation
 general discussion, 4:509–510
 VEGFR-1, 4:509
 VEGFR-2, 4:509
 VEGFR-3, 4:509–510
- ligand binding
 co-receptors, 4:511
 receptor binding, 4:510–511, 4:511f
 VEGF structure, 4:510–511, 4:511f
- protein structure
 characteristics, 4:510
 extracellular domains, 4:510, 4:510f
 intracellular domains, 4:510, 4:510f
 transmembrane domains, 4:510, 4:510f
- research summary, 4:514
- signal transduction
 receptor tyrosine kinase activation, 4:511, 4:512f
 signaling protein recruitment, 4:511–512
 signal transduction cascades, 4:512
 structural characteristics, 4:509
- von Hippel–Lindau protein (VHL), 4:570
- vascular endothelium
 adiponectin, 3:74
 aquaporin water channels, 1:124t
 blood-clotting proteases, 3:586f, 3:587
 leukotriene receptors, 3:584
- vascular smooth muscles, 2:345t
- vas deferens, 1:124t
- vasoactive intestinal polypeptide (VIP), 1:272
- vasopressin, 4:515–519
 ascorbic acid, 4:531–532
 evolutionary development, 4:516f
 functional role, 4:515–517
 G-protein-coupled receptors (GPCRs), 2:344t
 peptide amidation, 3:396
 peptide biosynthesis, 4:515
 receptors, 4:517–519, 4:518f
 structural characteristics, 4:515
 transmembrane domains, 4:517–519, 4:518f
 types and sources, 4:516t
- vasotocin, 4:516t
- VASP/Ena (EHAH) proteins, 1:28–29, 1:29–31, 1:29f, 1:30f, 1:19–20, 2:320–321, 2:320f
- V(D)J recombination, 3:272, 3:272f
- veal, 4:533t
- vectorial redox loop, 1:468–469, 1:469f
- vegetables, 4:533t, 4:547–548
- venomous peptides, 4:551
- ventromedial hypothalamus (VMH), 2:260, 3:77–78, 3:78f
- venturicidin, 1:143t
- Venus flytrap (VFT) domain, 2:396
- veratridine, 4:568t
- versican, 3:657t, 3:658
- vertebrates
 calpains, 1:354f
 lectins, 2:701, 2:702t, 2:704t
 matrix metalloproteinases (MMPs), 3:90, 3:91t
- nicotinic acid adenine dinucleotide phosphate (NAADP)-induced calcium ion (Ca^{2+}) release, 2:638, 2:638t
- vasopressin/oxytocin family, 4:516t
- very long-chain acyl-CoA dehydrogenase (VLCAD), 1:385f
- very long-chain acyl-CoA synthetase (VLCAS), 2:293–294
- very low density lipoproteins (VLDLs)
 diabetes, 1:655
 insulin metabolism, 2:606–607
 lipogenic gene expression, 2:287f
 peroxisome proliferator-activated receptors (PPARs), 3:418
 starvation effects, 4:303–304
 vitamin E, 4:548
- very-short patch (VSP) repair
 homologous recombination (HR), 2:46f
 research background, 4:76
- vsr* genes
 biochemistry, 4:76–77
 mutagenesis, 4:77
 structure and regulation, 4:76, 4:77f
- very strong (single-well) hydrogen bonds
 bond lengths, 2:757
 deuterium fractionation factors, 2:758
 spectroscopic properties, 2:758
 structural characteristics, 2:757, 2:757f
- Veterans Affairs Diabetes Trial (VADT), 3:423
- Viagra, 1:569
- Vibrio*, 2:43
- Vibrio anguillarum*, 4:251
- Vibrio cholerae*
 adenosine diphosphate ribosylation factors (ARF), 1:134–135
 chemotaxis receptor fibers, 3:607–608
 Cho phylogenetic tree, 3:331f
 G_s family, 2:519
 LexA regulatory protein, 2:710–711
 tyrosine recombinases, 1:558, 1:558f
uvrC phylogenetic tree, 3:331f
- Vibrio fischeri*, 1:465, 1:572
- Vibrio parahaemolyticus*, 2:14, 2:388, 2:389f
- villin, 1:23–24, 1:277–278
- vimentin, 1:47, 2:625–627, 2:632t, 2:634, 3:233
- vinblastin, 1:150t
- vinculin, 2:320–321, 2:320f
- vinexin, 2:320–321
- violaxanthin
 chemical properties, 1:504
 light harvesting process
 energy dissipation, 3:493
 nonphotochemical energy dissipation, 2:181f
 protein components, 2:726–727
 reaction centers, 2:178
 structural characteristics, 2:179f
 xanthophyll cycle, 2:180–181, 3:486–487, 3:487f
 molecular structure, 1:502f
- viral fusion proteins
 characteristics and functional role, 3:38
 class I fusion proteins, 3:39–41, 3:39f
 class II fusion proteins, 3:39f, 3:41
 class III fusion proteins, 3:40f, 3:41–42
- viral RNAs, 4:156–157
- viral src gene (v-src), 4:293–294
- viroldamine, 2:195f
- viruses
 DNA replication, 2:108–109
 hexameric enzyme helicases, 2:26
 lectins, 2:701, 2:702t, 2:704t
 non-homologous end joining (NHEJ), 3:268
 nuclear factor kappa B (NF- κ B), 3:303t, 3:304t
 ribozyme distribution, 4:146f
 sliding clamp processivity, 4:240, 4:240f
See also human immunodeficiency virus (HIV-1)
- vitamin A, 4:525–529
 chemical and metabolic reactions, 4:525–526, 4:526t
 clinical uses, 4:529
 functional roles, 4:525
 homeostasis control mechanisms, 4:529
 naturally occurring retinoids, 4:529
 pharmacological retinoid concentrations, 4:528–529
 retinoid-binding proteins (RBPs), 4:526–528, 4:527f, 4:527t
 retinoid homeostasis, 4:526–528
- vitamin B₁, 1:202–209
 adenosine thiamine triphosphate (AThTP)
 derivative determinations, 1:204, 1:204f
 metabolic pathways, 1:205f, 1:206
 molecular structure, 1:203f
 non-cofactor roles, 1:207–208
 background information, 1:202
 biosynthesis, 1:203–204
 chemical properties, 1:202–203, 1:204f
 derivative determinations, 1:204, 1:204f
 metabolic pathways
 adenosine thiamine triphosphate (AThTP), 1:205f, 1:206
 general discussion, 1:206
 thiamine diphosphate (ThDP), 1:205, 1:205f
 thiamine triphosphate (ThTP), 1:205–206, 1:205f
 molecular structure, 1:203f
 non-cofactor roles, 1:207–208
 related diseases
 beriberi, 1:202, 1:208
 nutritional deficiencies, 1:208–209
 transport defects
 Amish lethal microcephaly, 1:209
 Leigh disease (subacute necrotizing encephalomyelopathy), 1:209
 maple syrup disease, 1:209
 thiamine-responsive megaloblastic anemia (TRMA), 1:209
 Wernicke–Korsakoff syndrome, 1:202, 1:208–209
 research summary, 1:209
 riboswitches, 1:203–204
 thiamine-binding proteins, 1:206
 thiamine diphosphate (ThDP)
 background information, 1:202
 catalytic reaction mechanisms, 1:206–207, 1:208f

- vitamin B₁ (*Continued*)
 cofactor role, 1:206, 1:207f
 derivative determinations, 1:204f
 hydrolysis mechanisms, 1:206
 α -lactyl-ThDP, 1:206–207, 1:208f
 metabolic pathways, 1:205, 1:205f
 molecular structure, 1:203f
 prevalence, 1:204
 pyruvate dehydrogenases (PDHs), 4:321–322, 4:325f
 thiamine monophosphate (ThMP)
 biosynthetic pathway, 1:203–204
 derivative determinations, 1:204f
 hydrolysis mechanisms, 1:206
 metabolic pathways, 1:205f
 molecular structure, 1:203f
 thiamine triphosphate (ThTP)
 derivative determinations, 1:204, 1:204f
 hydrolysis mechanisms, 1:206
 metabolic pathways, 1:205–206, 1:205f
 molecular structure, 1:203f
 non-cofactor roles, 1:207–208
 transport mechanisms, 1:205
 vitamin B₁₂, 1:155–160, 2:324–328
 absorption and transport, 2:327
 acetogenesis, 1:159–160
 biochemistry, 2:327
 catalytic reactions
 carbon skeleton rearrangement reactions, 1:157, 1:157f, 1:158t
 methyl transfer reactions, 1:156f, 1:158–159, 1:159f
 migration and elimination reactions, 1:156, 1:157f, 1:158t
 pyridoxal 5'-phosphate (PLP)-dependent aminomutase reactions, 1:157, 1:157f, 1:158t
 radical-mediated rearrangement reactions, 1:156, 1:157f, 1:158t, 2:327
 ribonucleotide triphosphate reductase, 1:157f, 1:158, 1:158t
 chemical characteristics and structure, 2:326, 2:326f
 cobalamin cofactor, 1:155–156, 1:156f
 cobalamin-dependent methionine synthase, 1:159, 1:159f
 coenzyme B₁₂, 3:544
 deficiency-related diseases, 2:328
 functional role
 metabolic reactions, 2:327
 methionine synthase, 2:327
 methylmalonyl CoA mutase, 2:327
 gastric secretions, 2:356
 methanogenesis, 1:159–160
 structural characteristics, 1:156f
 vitamin B₂
 chemical properties, 2:309
 importance, 2:309
 molecular structure, 2:310f
 vitamin B₆
 characteristics and functional role, 3:708
 molecular structure, 3:709f
 vitamin C, 4:530–534
 background information, 4:530
 biochemical importance, 4:531–532
 cellular regulation, 4:532–533
 chemical properties, 4:531, 4:531f
 detection methods, 4:533
 digestion and absorption, 4:533
 nutritional importance, 4:531–532
 nutritional requirements, 4:533
 peptidylglycine α -hydroxylating monooxygenase (PHM), 3:396
 recommended dietary allowance (RDA), 4:533–534
 research summary, 4:534
 sources, 4:533t
 vitamin D, 4:535–539
 characteristics and functional role, 4:540
 chemical identity, 4:535, 4:536f
 cholesterol biosynthetic pathway, 1:516, 1:517f
 1 α ,25-dihydroxyvitamin D₃ (1,25-(OH)₂D₃)
 action mechanisms, 4:541f
 calbindin-D_{9k}
 intestines, 1:288
 regulation mechanisms, 1:288
 calbindin-D_{28k}
 intestines, 1:284
 kidneys, 1:285
 regulation mechanisms, 1:286
 calcium homeostasis, 1:328–329
 characteristics and functional role, 4:308–309
 chemical structure, 4:309f
 degradation pathways, 4:538
 endocrine regulation, 4:537
 functional roles, 4:538–539
 hormonal conversion process, 4:535–536, 4:536f
 metabolism, 4:538
 molecular mechanisms, 4:537–538, 4:538f
 negative VDREs (nVDREs), 4:542–543
 plasma calcium and phosphorus regulation, 4:536–537, 4:537f
 therapeutic uses, 4:539
 vitamin D receptor (VDR), 4:311, 4:537, 4:543–544, 4:543f
 vitamin D-response elements (VDREs), 4:311f, 4:537–538, 4:538f, 4:542–543
 endocrine regulation, 4:537
 functional roles, 4:538–539
 hormonal conversion process, 4:535–536, 4:536f
 negative VDREs (nVDREs), 4:542–543
 occurrences, 4:535
 production processes, 4:535
 therapeutic uses, 4:539
 vitamin D receptor (VDR), 4:540–544
 1 α ,25-dihydroxyvitamin D₃ (1,25-(OH)₂D₃), 4:311, 4:537, 4:541f
 functional domains
 ligand-binding domain, 4:541–542, 4:542f
 N-terminal DNA-binding domain, 4:540–541, 4:542f
 functional role, 4:540, 4:541f
 hematopoiesis, 1:188–189
 research summary, 4:544
 transcriptional control
 LXXLL motif, 4:541f, 4:543–544, 4:543f
 Mediator D complex, 4:543
 NCoA62/SKIP coactivator protein, 4:543–544
 steroid receptor coactivator (SRC) family, 4:541f, 4:543, 4:543f
 transcriptional machinery interactions, 4:543–544
 vitamin D-response elements (VDREs), 4:542–543
 vitamin E, 4:545–550
 antioxidant activities, 4:545–546, 4:546f
 characteristics, 4:545
 coenzyme Q, 3:725–727
 dietary sources, 4:547–548
 Friedreich's ataxia (FRDA), 2:333
 health and disease considerations
 cancer, 4:547
 central nervous system (CNS), 4:547
 coronary heart disease (CHD), 4:547
 fertility, 4:547
 medical paradox, 4:547
 protective behaviors, 4:547
 RRR- α -tocopherol, 4:547
 hydrophobicity, 4:549
 lipid peroxidation, 4:545–546, 4:546f
 metabolic characteristics, 4:549, 4:549f
 recommended dietary allowance (RDA), 4:547–548
 structural characteristics, 4:545, 4:546f
 transport mechanisms
 absorption, 4:548
 circulation and delivery processes, 4:548–549
 hepatic processing, 4:548
 vitamin H, 1:221–225
 biotin domains
 characteristics, 1:222–223
 sequence alignment, 1:223f
 structural characteristics, 1:223–224, 1:223f
 biotin enzymes, 1:221–222, 1:222f
 biotin-mediated carboxyl transfer reactions, 1:222f
 biotin protein ligase (BPL)
 biotin-inducible repressor (BirA), 1:224, 1:224f
 biotinylated proteins, 1:226
 reaction mechanisms, 1:224
 characteristics, 1:221
 protein biotinylation, 1:226–228
 analytical and clinical applications, 1:226
 basic concepts, 1:226
 biosynthesis, 1:226
 biotinylated antibodies, 1:227–228
 chemical synthesis, 1:227, 1:227f
 medical importance, 1:226–227
 research background, 1:221
 vitamin K, 4:551–554
 biochemical role, 4:551–553
 chemical characteristics, 3:727–728
 deficiency-related diseases, 4:554

- functional role, 3:728, 3:728f
 metabolic interconversion, 4:553–554, 4:553f
 nutritional role, 4:554
 vitamin K-dependent carboxylase, 4:551–553, 4:552f
 vitamin K-dependent proteins, 4:551
 vitamin malabsorption, 3:564
 vitellogenesis, 2:592–593
 vitronectin (VN)
 cell–matrix interactions, 1:413f, 1:426, 1:427t, 1:438–440
 focal adhesion kinase (FAK) family, 2:265–266
 integrins, 1:427t, 1:438–440
 matrix metalloproteinases (MMPs), 3:91t
 voltage-dependent anion-selective channels (VDACs)
 functional role, 3:161
 mammalian low- K_m hexokinases, 2:546
 mitochondrial CK (MtCK) isoenzyme, 3:82–84, 3:83f
 mitochondrial DNA (mtDNA), 1:212f
 mitochondrial outer membrane, 1:350–351, 3:127–128, 3:128t, 3:130f, 3:159
 mitochondrial permeability transition pore, 3:163
 structural characteristics, 3:160–161, 3:160f
 voltage-dependent calcium channels (VDCCs), 2:597f
 vomiting
 dopamine receptors, 2:173
 hyperammonemia, 4:494
 neuromin-1 receptor (NK1R), 4:364
 von Gierke disease, 2:383, 2:427t, 2:431, 2:431t
 von Hippel–Lindau protein (VHL), 4:570–571
 functional role, 4:570, 4:571f
 HIF-1 α level regulation, 4:69
 mutation consequences, 4:570
 ubiquitin ligase
 functional role, 4:570, 4:571f
 hypoxia inducible factor (HIF) regulation, 4:571
 structural characteristics, 4:570–571, 4:571f
 von Willebrand factor (vWF)
 cell–matrix interactions, 1:426, 1:427t
 integrins, 1:427t
 VPS34/Vps34p gene complex, 3:447–448, 3:448f, 3:450
 vSrc protein, 4:293–294
vsr genes
 biochemistry, 4:76–77
 mutagenesis, 4:77
 structure and regulation, 4:76, 4:77f
 VS ribozyme, 4:144, 4:146f
- W**
- Wagner syndrome, 3:657t, 3:658
 Walker A (GX4GKT) motif, 1:1, 1:3f, 1:7, 1:9f, 2:22–24, 2:136–137, 2:137f
 Walker–Warburg syndrome, 2:471t, 2:474–475
 Warburg effect
 fatty acid metabolism–cancer relationship, 2:275
 historical research, 3:391
 malignant tumor growth, 2:446
 photochemical action spectra, 4:274–275, 4:275f
 pyruvate dehydrogenase kinase (PDK) regulation, 4:324
 pyruvate kinase (PK), 3:721
 Warburg, Otto, 4:436
 warfarin, 3:588–589, 4:553–554, 4:553f
 water
 aquaporin water channels, 1:120–126
 aquaglyceroporins, 1:123–125, 1:124f
 aquaporin-1 (AQP1)
 arginine vasopressin (AVP) regulation, 1:123, 1:123f
 functional expression, 1:121f
 historical research, 1:120
 membrane topology, 1:122f
 physiological functions, 1:121–123, 1:123f
 sagittal cross-section schematic diagram, 1:122f
 structural characteristics, 1:120–121, 1:122f
 aquaporin-2 (AQP2), 1:123, 1:123f
 distribution and permeability, 1:123–125, 1:124t
 evolutionary relationships, 1:124f
 mammalian aquaporin homologs, 1:123–125, 1:124t
 research summary, 1:125
 water transport mechanisms, 1:120, 1:121f
 metalloproteases, 3:87–88
 water transport mechanisms, 1:120, 1:121f
 WAVE proteins, 1:440–441, 1:441f, 4:113
 waxes, 2:291f
 WBSCR 14 syndrome, 1:374
 WD (tryptophan (W)-aspartic acid (D) dipeptide) domain, 1:24
 weak hydrogen bonds
 bond lengths, 2:757
 deuterium fractionation factors, 2:758
 spectroscopic properties, 2:757
 structural characteristics, 2:756, 2:757f
 Werner's syndrome
 characteristics, 4:46
 DNA secondary structures, 2:147
 hexameric enzyme helicases, 2:26
 mutation effects, 4:44
 nicotinamide adenine dinucleotide (NAD⁺/NADH), 1:177
 Wernicke–Korsakoff syndrome, 1:202, 1:208–209, 3:394
 wheat
 lectins, 2:702t
 retrotransposons, 3:284f
 wheat germ, 2:704t
- white adipose tissue (WAT), 1:583, 2:259, 2:518–519
 Williams–Beuren syndrome, 1:374
 Wilms' tumor, 3:657t
wilm's tumor associated protein (wtap) gene, 1:84
 Wilson's disease, 3:670–671
 Wiskott–Aldrich syndrome protein (WASP)
 family, 1:29, 1:20, 1:21f, 1:26, 3:438–439, 4:113, 4:114f
 Wnt pathway
 adipocyte development, 1:53–54, 1:53f
 cardiac muscle development, 1:183–184, 1:184f
 fibroblast growth factor 8 (FGF8), 2:306
 liver regeneration, 1:192f
 neurogenesis, 1:197f
 parathyroid hormone (PTH), 3:385–386
 protein palmitoylation, 2:736–737, 3:645–646
 telomeres, 4:386
 Wnt/ β -catenin signaling pathway, 1:263–264, 1:263f, 2:434–435, 2:436f, 3:385–386, 4:386
Wolbachia sp, 3:331f
Wolinella, 2:207f, 2:208f, 3:30
Wolinella succinogenes, 3:30, 4:92–93, 4:94f, 4:95f
 woolly hair syndrome, 1:638t
 Woronin body, 3:425
 wortmannin, 1:150t, 3:274
 wwPDB database, 3:615t
- X**
- XAB2 protein, 4:404
 xanthine oxidase, 2:77, 2:312, 2:590
Xanthomanas campestris, 3:350–351
 xanthophylls, 2:178–179, 2:726–727, 3:486–487, 3:487f
 xanthopterin, 3:667f
 XBP-1 gene expression, 2:524–525, 2:524f
 X-chromosome inactivation, 1:522
Xenopus laevis
 adenosine-to-inosine (A-to-I) RNA editing, 4:154–155
 aquaporin water channels, 1:120, 1:121f, 1:125
 calpains, 1:359
 DNA replication, 2:109–110
 kinesins, 2:674
 origin recognition complex (ORC), 2:111
 RecQ helicases, 4:43–44, 4:44f
 small RNAs (sRNAs), 3:98
 voltage-gated potassium channels, 4:559
Xenopus tropicalis, 2:5t
 XerC/XerD/*dif* system, 1:555–556, 1:558f, 1:561
 xeroderma pigmentosum (XP)
 hexameric enzyme helicases, 2:26
 nucleotide excision repair (NER), 3:339–340, 4:404
 polymerase η (Pol η), 4:481, 4:572–574
 ultraviolet (UV) damage, 4:572
 X-linked adrenoleukodystrophy, 3:415–416

X-linked dominant protoporphyria, 2:542, 2:542t
 X-linked sideroblastic anemia, 2:542t
 XPC protein complex, 3:341–342, 3:342f, 3:342t
 XPG protein complex, 3:342f, 3:342t, 3:343–344, 3:343f
 X-ray absorption spectroscopy, 2:661–662
 X-ray cross-complementing factor-1 (XRCC1), 2:85–86
 X-ray crystallography, 3:11, 4:129–130
 XRCC1 protein, 2:4f, 2:6, 2:30–31, 2:30t
 XRCC4-like factor (XLF), 3:271, 3:273t
 XRCC4 protein, 3:271, 3:272f, 3:273t
Xylella fastidiosa, 3:331f
 xylitol, 2:387
 xylose
 O-linked glycosylation, 2:466f
 structural characteristics, 3:654
 O-xylose, 2:453
 α 1,3-xylosyltransferase (α 3XylT), 2:454f
 xylulose-5-phosphate (Xu-5-P)
 carbohydrate-responsive element binding protein (ChREBP), 1:375–376, 1:376f, 1:377f
 pentose phosphate pathway
 carbon fate, 3:394f
 chemical balance, 3:394f
 nonoxidative reactions, 3:391, 3:393f
 oxidative reactions, 3:391–392
 photosynthetic metabolic pathways, 3:499f
 thiamine diphosphate (ThDP), 1:207f

Y

Yarrowia lipolytica
 peroxisomal matrix protein import, 3:426–429
 respiratory chain complex I (NADH: ubiquinone oxidoreductase), 4:88, 4:89t, 4:341–342
 yeasts
 budding yeasts, 4:31–32, 4:216–217, 4:216f
 calpains, 1:354f
 conventional actin, 1:38–40
 cytochrome *c*, 1:600, 1:600f
 exon–intron architecture, 3:142–144, 3:143f, 4:186–187
 fission yeasts, 4:217–218
 glycosylphosphatidylinositol (GPI)-anchored proteins, 2:477–478
 macroautophagy
 Atg proteins, 1:151–153, 1:152t
 cytoplasm-to-vacuole targeting (Cvt) pathway, 1:151, 1:151f
 model systems, 1:150–151
 molecular mechanisms, 1:151–153
 Atg proteins, 1:151–153, 1:152t
 cargo selection, 1:153
 fusion and breakdown, 1:153
 phagophore assembly site (PAS), 1:152
 regulatory mechanisms, 1:152–153
 microautophagy

mitophagy, 1:153
 nonselective microautophagy, 1:153
 pexophagy, 1:153
 mitochondrial DNA (mtDNA), 3:134t
 DNA replication, 2:128
 gene expression, 3:140–144
 basic process, 3:140–141, 3:141f
 encoding processes, 3:140
 5'-end–RNase P process pathway, 3:142, 3:142f
 model systems, 3:140, 3:141f
 mosaic mitochondrial genes, 3:142–144, 3:143f
 RNA splicing, 3:142–144, 3:143f
 3'-end–RNase process pathway, 3:142, 3:142f
 transcription processes, 3:141, 3:141f, 3:142f
 translation processes, 3:144
 nucleoporin proteins (Nups), 3:319f, 3:320t
 pheromone receptors, 3:441–446
 conformational constraints, 3:445
 conjugation functions, 3:441, 3:442f
 G-protein activation, 3:445–446
 ligand binding, 3:444–445
 receptor biogenesis, 3:443–444, 3:444f
 regulatory mechanisms, 3:442–443, 3:443f, 3:444f
 signal transduction pathways, 3:441–442, 3:443f
 spatial gradient sensors, 3:442
 phospholipase D (PLD), 3:473
 phospholipid synthesis, 3:478–481
 basic process, 3:478
 cytidine diphosphate (CDP)–choline pathway, 3:479f, 3:480
 cytidine diphosphate (CDP)–diacylglycerol (DAG) pathway, 3:478–480, 3:479f
 cytidine diphosphate (CDP)–ethanolamine pathway, 3:479f, 3:480
 major phospholipids, 3:478, 3:479f
 phosphatidate (PA) synthesis, 3:478
 phosphatidylglycerolphosphate (PGP), 3:480
 regulation mechanisms
 biochemical regulation, 3:481
 general discussion, 3:480
 genetic regulation, 3:480–481
 plasma membrane proton (H^+)-ATPases, 3:688
 prions, 3:566–570
 chaperone functions, 3:569
 disease states, 3:568–569
 functional domains, 3:566–567, 3:568f
 generation mechanisms, 3:569
 genetic criteria, 3:566, 3:567f
 known prions, 3:566
 prion domains, 3:566–567, 3:568f
 research summary, 3:569–570
 self-propagating amyloidosis, 3:567–568, 3:568f
 protein biosynthesis elongation, 2:253

recombination-dependent DNA replication (RDR), 4:24–25
 retrotransposons, 3:283, 3:284f
 ribozyme distribution, 4:146f
 secretory pathway, 3:690, 3:690f
 septins
 budding yeasts, 4:216–217, 4:216f
 fission yeasts, 4:217–218
 superoxide dismutases (SODs), 4:353
 two-hybrid protein–protein interaction systems, 4:460–462
 basic concepts, 4:460–462, 4:461f
 proteomic studies, 4:462
 research applications, 4:461–462
 reverse two-hybrid systems, 4:461–462
 unfolded protein response (UPR)
 HAC1 gene expression, 4:489, 4:490f
 inositol requiring 1 (IRE1) genes, 4:489, 4:490f
 vacuoles
 biogenesis, 4:503–506
 characteristics, 4:502–503
 functional role, 4:503, 4:504f
 transport pathways
 AP-3 pathway, 4:505f, 4:506
 autophagy/macro-autophagy, 4:505f, 4:506, 4:506f
 characteristics, 4:503–506, 4:505f
 Vps pathway, 4:504–506, 4:506f
 See also *Saccharomyces cerevisiae*
 yellow fluorescent protein (YFP), 1:465, 2:577–578, 2:751–752, 2:753f, 4:257–258
Yersinia, 2:43
Yersinia enterocolitica, 2:43
Yersinia murine toxin (Ymt), 3:473
Yersinia pestis, 3:331f, 3:473
 yes genes/Yes proteins, 4:294
 Y-family polymerases, 4:477–481
 bacterial homologs, 4:480–481
 characteristics, 4:480, 4:572–574
 cleavage reactions, 4:477–478, 4:478f
 eukaryotic homologs, 4:481
 mutagenesis, 4:478
 regulatory mechanisms, 4:480
 research summary, 4:481
 SOS response, 4:477, 4:478f, 4:479f
 transcription regulation, 4:477, 4:478f
 translesion synthesis (TLS), 4:479, 4:479f
 yohimbine, 1:58–59
 Yonath, Ada E, 2:254

Z

Z-DNA
 background information, 2:142–143
 binding characteristics, 4:155–156, 4:156f
 biological role, 2:142–143
 chemical properties, 2:142, 2:143f
Zea mays, 3:101, 3:284f
 zeatin, 1:627, 1:628f
 zeaxanthin
 light harvesting process
 energy dissipation, 3:493

- nonphotochemical energy dissipation, 2:181f
- reaction centers, 2:178
- structural characteristics, 2:179f
- xanthophyll cycle, 2:180–181, 3:486–487, 3:487f
- molecular structure, 1:502f
- zebrafish
- See Danio rerio*
- Zellweger syndrome
- leukotriene metabolism, 3:584
- peroxisome biogenesis, 3:414
- zeta-associated protein-70 (ZAP-70), 3:649–651, 4:374
- zidovudine (AZT), 2:549f, 2:550–551, 2:550f
- zinc (Zn)
- glycine receptors (GlyRs), 2:422f, 2:423
- matrix metalloproteinases (MMPs), 3:90–97
- activation mechanisms
- chemical/physical agents, 3:95
- proteinases, 3:95, 3:95f
- zymogens (proMMPs), 3:94
- biological functions, 3:96
- characteristics, 1:543
- characteristics and functional role, 3:90
- classifications, 1:543t
- DNA polymorphisms, 3:96
- domain structures, 3:90, 3:92f, 3:93–94, 3:93f
- enamelysin, 3:91t, 3:92f, 3:93
- endogenous inhibitors
- α -2 macroglobulins, 3:96
- tissue inhibitors of metalloproteinases (TIMPs), 3:96
- epilysin, 3:91t, 3:92f, 3:93
- families/subfamilies
- characteristics, 3:90
- collagenases, 3:90–91, 3:91t, 3:92f
- gelatinases, 3:91t, 3:92–93, 3:92f
- matrilysins, 3:91t, 3:92f, 3:93
- membrane-type MMPs, 3:91t, 3:92f, 3:93
- stromelysins, 3:91t, 3:92f, 3:93
- functional roles, 1:543–544
- genetic disorders, 3:96
- historical research, 1:542
- inhibition mechanisms, 3:96
- metalloelastase, 3:91t, 3:92f, 3:93
- occurrences, 3:90
- peptide bond hydrolysis mechanisms, 3:94, 3:94f
- substrate specificity, 3:95–96
- tissue expression, 3:92f
- tissue inhibitors of metalloproteinases (TIMPs), 3:93f, 3:95f, 3:96
- vertebrate matrixins, 3:90, 3:91t
- metalloproteases, 3:86–87, 3:87f, 4:50–53
- P-type adenosine triphosphatase (ATPase), 4:334
- superoxide dismutases (SODs), 4:352–353
- zinc-binding domains
- DNA ligases, 2:34f
- type I DNA topoisomerase, 2:158, 2:159f
- zinc-finger domains, 4:575–579
- basic concepts, 4:575
- biotechnological applications, 4:578
- cardiac muscle development, 1:184
- classical zinc fingers
- amino acid sequence motifs, 4:576f
- biotechnological applications, 4:578
- evolutionary development, 4:576
- genomic abundance, 4:577
- historical research, 4:575
- hydrophobic core, 4:575, 4:576f
- structural characteristics, 4:575, 4:576f
- transcription factor IIIA (TFIIIA), 4:575, 4:576f
- zinc functions, 4:575
- DNA-binding motifs, 2:151f
- DNA ligases, 2:29f, 2:30–31, 2:34f
- DNA–RNA interactions
- DNA-recognition code, 4:577–578, 4:578f
- DNA-recognition helix, 4:577, 4:578f
- functional role, 4:577–578
- RNA recognition, 4:578
- structural characteristics, 4:577, 4:578f
- functional role, 4:577–578
- GATA-1 zinc finger transcription factor, 1:187, 1:188f
- GATA-2 zinc finger transcription factor, 1:187, 1:188f
- GATA-3 zinc finger transcription factor, 1:187–188, 1:188f
- GATA-4 zinc finger transcription factor, 1:184
- GATA-6 zinc finger transcription factor, 1:184f
- hematopoiesis, 1:187, 1:188f
- methionine aminopeptidases (MetAPs), 1:98, 1:98f
- nonclassical zinc fingers
- background information, 4:577
- Cys₃–His zinc fingers, 4:576f, 4:577
- Cys₄ zinc fingers, 4:576f, 4:577
- GAL4 domain, 4:576f, 4:577
- multi-zinc proteins, 4:576f, 4:577
- ring fingers, 4:576f, 4:577
- non-nucleic acid interactions
- biotechnological applications, 4:578
- lipid binding, 4:578
- protein–protein interactions, 4:578
- polymerase η (Pol η), 4:573–574, 4:573f
- retinoic acid receptors, 4:108
- RNA polymerases, 4:174f, 4:175–176, 4:175t
- Zydena, 1:569
- zygonema, 3:19, 3:19f, 3:22f
- Zymomonas mobilis*, 2:205t